# CLIENT DATA ANALYTICS IN EQUITY SALES & TRADING: DEVELOPING ORGANIZATION'S CAPABILITIES

Lappeenranta–Lahti University of Technology LUT

Industrial Engineering and Management

2024

Roope Nevalainen

Examiners: Docent, post-doctoral researcher Lasse Metso, D.Sc. (Tech)

University Teacher Leena Tynninen

ABSTRACT

Lappeenranta–Lahti University of Technology LUT

LUT School of Engineering Sciences

Industrial Engineering and Management

Roope Nevalainen

**Client Data Analytics in Equity Sales & Trading: Developing Organization's Capabilities**

Master's thesis

2024

103 pages, 15 figures, 11 tables and 5 appendices

Examiners: Docent, post-doctoral researcher Lasse Metso, D.Sc. (Tech)

University Teacher Leena Tynninen

Keywords: Data, data analytics, client analytics, data analytics capabilities, equity brokerage, data architecture

Globally investment banks are massive consumers of data and skilful at leveraging their businesses by using data analytics. One common application of data analytics in global investment banks are client analytics. Globally client data analytics are used for client risk profiling, product recommendations systems, and for client satisfaction analytics. However, global applications of client data analytics cannot be directly applied to Nordic equity brokerage businesses as Nordic investment banking has its special characteristics.

This master's thesis is conducted as qualitative research. It utilized research interviews and constructive research method to recognize differences between global and Nordic equity brokerage businesses, to recognize applications of client data analytics for Nordic equity brokerage business, to explore client data analytics capabilities in target organization, and to provide suggestions to develop Nordic equity brokerage business's client data analytics capabilities. As a result, it found out that the gap between global and Nordic brokerage houses is so large that most of the globally recognized applications of client data analytics cannot be systematically applied for Nordic institutions. Instead, it would be more efficient to develop organization's data analytics capabilities. This research offers suggestions to develop organization's technological, human resources, and organization cultural data analytics capabilities. It also evaluates the direct, semi-direct, and indirect effects of the proposed solutions for the recognized deficiencies.

TIIVISTELMÄ

Lappeenrannan–Lahden teknillinen yliopisto LUT

LUTin insinööritieteiden tiedekunta

Tuotantotalous

Roope Nevalainen

**Asiakasanalytiikka osakevälityksessä: Organisaation kyvykkyyksien kehittäminen**

Tuotantotalouden maisteriohjelma

2024

103 sivua, 15 kuvaa, 11 taulukkoa ja 5 liitettä

Tarkastajat: Dosentti, tutkijatohtori Lasse Metso

Yliopisto-opettaja Leena Tynninen

Avainsanat: Data analytiikka, osakevälitys, asiakasanalytiikka, data analytiikkakyvykkyydet, data arkkitehtuuri

Globaalit investointipankit ovat suuria datan kuluttajia ja taitavia hyödyntämään data-analytiikkaa liiketoimintansa tukena. Yksi yleinen data-analytiikan sovellusalue globaaleissa investointipankeissa on asiakasanalytiikka. Maailmanlaajuisesti asiakasdataa käytetään asiakasriskiprofiilien luomiseen, tuotesuositusjärjestelmiin ja asiakastyytyväisyyden analysointiin. Globaalisti sovelletut asiakasdata-analytiikan sovellukset eivät kuitenkaan ole suoraan sovellettavissa pohjoismaiseen osakevälitysliiketoimintaan, koska pohjoismaisella investointipankkiliiketoiminnalla on omat erityispiirteensä.

Tämä tutkimus on toteutettu laadullisena tutkimuksena. Tutkimusmenetelminä se hyödyntää tutkimushaastatteluja ja konstruktiivista tutkimusotetta. Tutkimuksessa selvitetään globaalien ja pohjoismaisten osakevälitysliiketoimintojen eroavaisuuksia, asiakasdata-analytiikan sovelluksia, ja kohdeorganisaation data-analytiikka kyvykkyyksiä. Se myös tarjoaa ehdotuksia organisaation data-analytiikkakyvykkyyksien kehittämiseksi. Tutkimuksen mukaan ero pohjoismaisten ja globaalien välitystalojen välillä on niin suuri, että monia data-analytiikan sovelluksia ei voida kohdeorganisaatiossa käyttää. Sen sijaan tehokkaampaa olisi kehittää organisaation data-analytiikan kyvykkyyksiä. Tämä tutkimus tarjoaa ehdotuksia organisaation teknologisten, henkilöstöresurssillisten ja organisaatiokulttuurillisten data-analytiikan kyvykkyyksien kehittämiseksi. Se myös tarkastelee ehdotettujen toimien suoria ja epäsuoria vaikutuksia tunnistettujen vajaavaisuuksien kannalta.

**Table of contents**

Abstract

1 Introduction ..... 6
1.1 Background and research problem ..... 6
1.2 Goals, research questions, and scoping ..... 7
1.3 Methods and literature ..... 8
1.4 Structure of the thesis ..... 9
2 Development of Equity Sales & Trading ..... 11
2.1 Brief history of investment banking ..... 11
2.2 What are Equity Sales & Trading? ..... 16
2.3 Development of Equity Sales & Trading ..... 19
3 Client data analytics in investment banking ..... 25
3.1 Definition and applications of client data analytics ..... 25
3.2 Challenges and requirements of client data analytics implementation ..... 30
3.3 Data analytics architecture and data mining applications for Sales & Trading organization ..... 35
4 Applications of client data analytics in Nordic Equity Sales & Trading ..... 46
4.1 Research methodology and materials ..... 46
4.2 Special characteristics of Nordic investment banking ..... 52
4.3 Target organization's capabilities and goals in data analytics ..... 60
4.4 Development of target organization's client data analytics capabilities ..... 69
5 Findings, discussion, and conclusions ..... 81
5.1 Answers for research questions ..... 81
5.2 Discussion and further research ..... 90
References ..... 93

Appendices

Appendix 1. Business environment research interview

Appendix 2. Data analytics capabilities, goals and projects research interview

Appendix 3. Suggestion for target organization's data architecture

Appendix 4. History of investment banking, timeline

Appendix 5. Effects of presented suggestions

# 1 Introduction

This master's thesis aims to recognize use cases, benefits, challenges, and requirements for client data analytics implementation in equity sales & trading business operating inside a Nordic bank. It seeks to recognize special characteristics of Nordic investment banking and to set a baseline for the target organization's client data analytics capabilities. This thesis also seeks to find applicable methods to improve target organization's data analytics capabilities.

The introduction part of this thesis introduces the background of the thesis and the reasons why it was carried out. It discusses about goals, research questions, and the scoping of this thesis, and aims to give reader throughout picture of what this thesis is about. The introduction covers also used research methods and literature along with the thesis's structure.

## 1.1 Background and research problem

The financial industry, particularly investment banking, has been at the forefront of adopting advanced data analytics to enhance operational efficiency and client service. The rapid evolution of data-driven technologies has prompted banks to leverage vast amounts of data to gain a competitive edge. Investment banks utilize data analytics for various purposes, including risk management, customer relationship management, and personalized service delivery. The ability to analyse client data effectively allows banks to predict client behaviour, tailor financial products, and optimize risk assessment protocols. Globally investment banks have been among the largest users of data, and lot of the data analytics efforts has been put to client analytics. Some frequently mentioned applications of client analytics in investment banking are client profiling and segmentation, tailored product offering and -marketing, client satisfaction management and evaluation, and client churn prediction and prevention. Some of the key drivers in data analytics implementation processes has been stagnation of revenues caused by competitive and highly regulated environment. Banks are faced with pressure to identify new revenue streams and improve profit margins. Increasingly complex regulatory requirements demand sophisticated tools to

ensure compliance and manage risks. Technological advancement, especially in big data related technologies have provided banks tools to utilize enormous amounts of data. Missing out on these technologies might have negative impacts on the bank's competitive advantages as most of the leading banks globally have adopted advanced data analytics into their daily operational routines. Despite the clear advantages, successful data analytics implementation is challenging and requires a lot of resources. The success is often dependent on technological capabilities, relevantly skilled employees, and fit organization culture.

The target organization of this thesis is an equity brokerage business operating inside bank located in Nordic region. This thesis's goal is to find out what is the baseline for client data analytics capabilities in the target organization and suggest feasible methods to develop them. Its goal is also to present how client data analytics are being utilized globally, and to evaluate which of those applications could be implemented in target organization.

## 1.2 Goals, research questions, and scoping

The primary goal of this thesis is to act as a preliminary investigation for future client data analytics projects. It aims to figure out common applications of client analytics in investment banking, requirements for successful implementation, challenges often associated with each of the requirements and offer suggestions that are in line with literature and the characteristics of target bank's equity brokerage business.

The research questions of this thesis are:

1.) *How can client data be utilized in Nordic equity sales & trading?*

2.) *How Nordic equity sales & trading organization could develop its client data analytics capabilities?*

The first research question aims to thoroughly explore different ways in which data analytics can be used to improve Nordic equity brokerage house's client relationships management. It also discusses about requirements for successful data analytics implementation and challenges associated with it considering the special characteristics of Nordic operating region. The second research question seeks to explore ways to improve Nordic equity sales & trading organization data analytics capabilities taking current state and characteristics into

attention. Both of these questions are researched by exploring target organization, and then by trying to generalize the findings to consider whole Nordic region.

This thesis is scoped so that it supports the target organization's future development in its data analytics capabilities as well as possible. It explores the directions of development in global equity brokerage industry and aims to recognize factors that are also relevant in Nording operating environment. Considering the title of this thesis, there could be a large variety in scoping. This thesis excludes further research in equity markets returns and trading strategies, and in-depth exploration of regulation. It also excludes product / vendor specific suggestions in data architecture and focuses on high level modelling. Considering the current state of target organization's data analytics capabilities, in-depth technological modelling, such as deciding data attributes, defining data warehouse schemas, creating visualization dashboards, and the analysis work itself are scoped out. Instead, this thesis focuses on creating high level picture of strategy that should be implemented to develop target organization's data analytics capabilities.

## 1.3 Methods and literature

This research is conducted by utilizing literature review, specialist interviews, and constructive research method. Literature review is needed to construct a clear picture of the characteristics and developmental directions of global equity brokerage business as well as the usual applications, requirements, and challenges of the data analytics implementation. The use of literature review as a research method in building sufficient theoretical base for master's thesis is justified as its goal is to give reader throughout picture of the whole topic, explore and assess existing theory, recognize gaps in existing theory and to develop existing theory. Literature review can also be used to recognize the historical development of the theory. (Baumeister & Leary 1997)

The specialist research interviews consist of two different interviews that were held for designated specialists. First interview asks questions on special characteristics of Nordic investment banking and the second focuses on gaining information on target organization's data analytics capabilities. Interviews are held as qualitative interview research, and for experienced professionals working at target organization.

The constructive research method is used to offer target organization solutions in developing its client data analytics capabilities. It explores different challenges in developing target organization's capabilities and tries to find solutions to those. Solutions are presented in time prioritized order. More in-depth description of research methodology and material can be found from chapter 4.1.

## 1.4 Structure of the thesis

The first chapter of this thesis, *Development of Equity Sales & Trading*, discusses about basic principles and evolution of investment banking, development of Equity Sales & Trading including history and direction of future development. It also presents the working principles of equity brokerage business which is often referred to later in this thesis. Second chapter, *Client data analytics in investment banking*, defines client data analytics and discusses about its applications, benefits, requirements, and challenges in context of investment banking and equity brokerage. It also discusses about data architecture and data mining applications that can be used in equity brokerage business's client analytics. Third chapter discusses about special characteristics of investment banking in the Nordics and seeks to find ways in which client data analytics can be implemented in target organization. It also explores applicable ways in which client analytics could be implemented considering the target organization capabilities and seeks to find ways to improve target organization's data analytics capabilities. Third chapter is also the empiric part of this thesis and its finding as are reflected with the theory presented in chapters before. Fourth chapter presents the findings of this thesis and discusses whether they are generalizable for other similar organizations in the same geographical region. Table 1 illustrates the structure of the research with the help of input-output table.

*Table 1: Structure of the research*

| Input | Chapter | Output |
|---|---|---|
| General description of the research and reseach methods | **1. Introduction** | Background, reseach problem, goals, scope, research methods, and structure of the research |
| Literature | **2. Development of Equity Sales & Trading** | Principles and evolution of investment banking, development of Equity Sales & Trading including its history and direction of future development |
| Literature | **3. Client data analytics in investment banking** | Definition of data and client analytics, applications, benefits, requirements, and challenges of client data analytics in context of investment banking and equity brokerage |
| Findings from literature, empiric research material | **4. Applications of client data analytics in Nordic Equity Sales & Trading** | Special characteristics of Nordic investment banking, target organization's data analytics capabilities and goals, development of target organization's data analytics capabilities |
| Findings from the literature and research material | **5. Findings, discussion, and conclusions** | Answers for reseach questions, findings of the research, generalizability of findings, further research |

In the Table 1, input column indicates the primary source of the chapter, and output column the results of the chapter. *Literature* refers to material used in literature review.

# 2 Development of Equity Sales & Trading

Investment banks are financial institutions whose fundamental job is to help companies and governments meet their long-term capital needs. Commonly, an investment bank acts as an intermediary between issuers of and investors in securities. Investment banks may employ their own capital into these activities with an investment purpose. Historically there has been a lot of variances between the scopes and structures of investment banks depending on the country of origin and the period of existence. Some investment banks are narrower on scope and more focused on specialized class of finance whereas most of the bigger ones are well diversified financial institutions offering variety of services such as brokerage, securities dealing, investment advising, commercial banking and insurance. These institutions subsequently mentioned are usually called "universal banks". On the other hand, pure investment banks commonly focus on wholesale or retail placements and are typically associated with brokerage houses or commercial banks. (Berger 2010; Michalczyk 2012; Stowell 2012) In this chapter history of investment banking and Equity Sales & Trading is discussed from the early origins of medieval merchant banking to the post financial crisis times with the whole new regulatory environment.

## 2.1 Brief history of investment banking

Despite that the services associated with investment banking had been emerging within other financial institutions for centuries, the term "Investment Banking" became popular just around a century ago in late nineteenth - early twentieth century, especially in United States. The early roots of investments baking originate from commodities merchants who traded commodities such as grains, spices, silk, metals and so forth on their own account. Some of the earliest observations of this commodities trading appeared in Italian cities such as Siena, Florence, Genova, and other cities that engaged in medieval trade in late 1400's (Kohn 1999; Hwang 2024). First modern type of stocks was traded in Amsterdam in early 1600's. In 1700's the city of Amsterdam arose as a merchant banking centre and a while later London took the lead. Besides trading with their own capital, early merchant bankers used to finance their trades by using a bill of exchange, a written order that binds one part to pay a fixed sum

of money to another party on a demand or a predetermined date (Hargrave 2024). Merchant bankers also leveraged their investments by taking interest-bearing deposits, similar to modern venture capital firms, or by selling commercial paper, a fixed interest yielding notes, to investors and other merchants (Hayes 2024). Eventually, some merchant bankers started to finance their operations by selling longer term securities to outside investors and taking on long-term capital. In addition to interest yielding notes, these securities also included early version of stocks. However, modern shareholder owned private companies remained relatively rare until early nineteenth-century, and the early securities markets traded mostly on government issued securities. The modern large-scale investment banking business itself rose parallel with the increased capital needs of large-scale businesses and the liberation of incorporation (limited liability companies). Investment banking started to take its modern form in mid-1800's to the First World War. First activities from such era entailed underwriting, the process of determining a required fee for the associated risks, and sales of government bonds (Banton 2023). Transaction sizes were typically large compared to the investor's assets and governments demanding large-scale funding, especially in times of wars, dominated among early clients. As the demand for long-term financing kept growing, the demand for wider range of different solutions also increased. Alongside to the debt solutions rose equity financing which traded off some of the cashflow certainty of the debt products and provided voting power to the company of issuance. As the nineteenth century progressed with the expansion of railway networks, advanced steel production and broader industrialization, the increased need for intermediatory between outside investors and the entrepreneurs forced investment bankers to collaborate and to promote the development of secondary markets for securities. This led to a birth of early organized stock markets. The further widening variety of legal structures for investment banking firms broadened with the liberalization of legislation and as gaining limited liability became increasingly easy. In the industrialized countries of the continental Europe, limited liability model performed so well that the biggest banks combining commercial and investment banking had become one of the largest companies of the region by 1870's. In contrast, investment banks in England and United States remained as partnerships during that era. The structures and institutional forms of investment banks varied largely based on the origin of operation and the needs of the economy investment bank operated on. Most investment banks operating in England, as well as those investment banks with close ties to England, preferred specialization in long-term capital needs and left the handling of short-term capital needs for commercial banks. In

countries other than England, especially in United States, Netherlands, and France, most often universal banks with large variety of services for short and long-term capital handling were formed. (Fohlin 2014; Geisst 2012; Michalczyk 2012)

Most significant growth in securitized finance and investment banking outside of the railways took place between 1860's and 1890's as the development of technology, especially in communications, mining, and textiles, scaled the production and increased the need for external capital. In United States the Civil War significantly boosted demand for specialized investment banking services as federal government needed external funding. The increasingly liquid stock markets enable efficient distribution of companies' stocks and led to birth of new method of investment banking called "company promoter". These company promoters would buy an entire join-stock limited liability firm and sell it to the second-hand investors with the help of necessary lawyers, accountants and brokers. Company promoters would earn the spread between the price they paid for the company and the price they got from the second-hand investors. Sometimes promoters would also hold large equity stakes on these corporation and place their own directors and board members. However, this action was far from riskless, as strict stock market regulations, especially in Belgium, kept many firms from trading resulting to illiquid markets. When concretized, this risk indeed forced many investment banking firms out of business. Stock market regulations finally eased in 1867, and particular after the liberation of incorporation in 1873, exchanges grew steadily until the First World War. When the First World War finally hit, it led governments to massive burdens of debt and seized dislocation of internation capital markets for protracted period. The turmoil led to a series of banking and financial crises and in some countries to hyperinflation. In turn governments made interventions and tightened financial regulations in financial sector. Despite the unstable economic environment of 1920's, markets for new issues expanded increasing demand for the services of investment bankers. Eventually, after the booming late 1920's came a crash that led the global economy to depression and gave birth to a new era of tighter financial regulations that lasted the majority of twentieth century. (Fohlin 2014; Geisst 2012)

Already in the late 1920's regulators had started to pay attention to and grow concerns about the power that investment bankers had in industrial economy. The attention was especially drawn to the matter of permitting universal banks to take ownerships in companies and use the voting power permitted by it. In addition, concerns were raised for combining

commercial and investment banking in a fear of mixing retail deposits and equity funding. In the United States this concern was famously handled by splitting commercial and investment banking into two separate categories in the Glass-Steagall Act of 1933 (Federal Reserve History). What made Glass-Steagall Act so influential was that it made some banks to completely change the nature of their business. J.P. Morgan for example became credit and deposit bank, whereas Morgan-Stanley switched solely to investment banking. Bank Holding Company Act further deviated the gap between the services offered by commercial and investment banks and gave Federal Reserve, the central bank of US, more regulatory power and oversight over the whole banking system (Federal Reserve History). Similar to U.S., many European countries such as Belgium, Italy, and Greece also developed their financial regulation in order to protect retail bank customers from conflicts of interest and other risks arising from combining commercial and investment banking. In the other hand, Germany by contrast, took a different approach and imposed significantly less regulation considering corporate governance or financial markets regulation affecting their investment banks (Detzer 2014). During the era between 1960's and 1970's financial institutions began to expand and consolidate by enlarging the scope and scale of their services. This led to a birth of large financial "conglomerates". In many countries that had separated investment and commercial banking by regulation, the gap started to shrink due to pressure caused by increase in global competition. Deregulation led to a situation where most of the largest banks operated practically in all financial sectors, including investment banking, but still had clear distinction between the various components. The US's Glass-Steagall Act started to erode in late 1980's when the Federal Reserve started to allow US's commercial banks to participate in some investment banking activities and was finally repealed in 1999 (Heakal 2024). Even England with its highly specialized investment banks deregulated the industry resulting to the formation of large universal banks and for example legalized proprietary trading. During the late nineteenth century most of the investment banks also changed their legal structure by changing from private partnerships to publicly traded, joint-stock companies with limited liability. However, the era of light regulation lasted only for few decades as the global financial crisis set off a wave of heavy restructuring and re-regulations for the whole industry setting it up for its modern frames. (Fohlin 2014; Michalczyk 2012; Stowell 2012) In this thesis reasons for financial crisis in further depth were not discussed as there are plenty enough of great master's thesis and other academic research written about it.

Post-financial crisis regulations reformed the industry completely. The regulation brought end to "too big to fail" -sized financial conglomerates and forced universal banks to structure themselves in a way that they can be recapitalized without taxpayers' money in case of future crises. Financial firms were made to value their securities more consistent with each other meaning that the value of securities cannot fluctuate too much depending on the method or generosity of the valuation. Regulation also forced financial institutions to be more transparent with their business. In U.S., one of the most reputable regulations legislated was the Dodd-Frank Act that included the Volcker Rule. Regulation made investment banks to separate their proprietary trading operations from the banking business by forbiting banks to own, invest, or sponsor hedge funds, private equity funds, or other proprietary trading operations that could yield profits for the bank. Along with the U.S., European Commission also tightened its financial regulation significantly tackling many problems similar to U.S.'s. (Berman 2023; European Commission 2013; Liaw 2011)

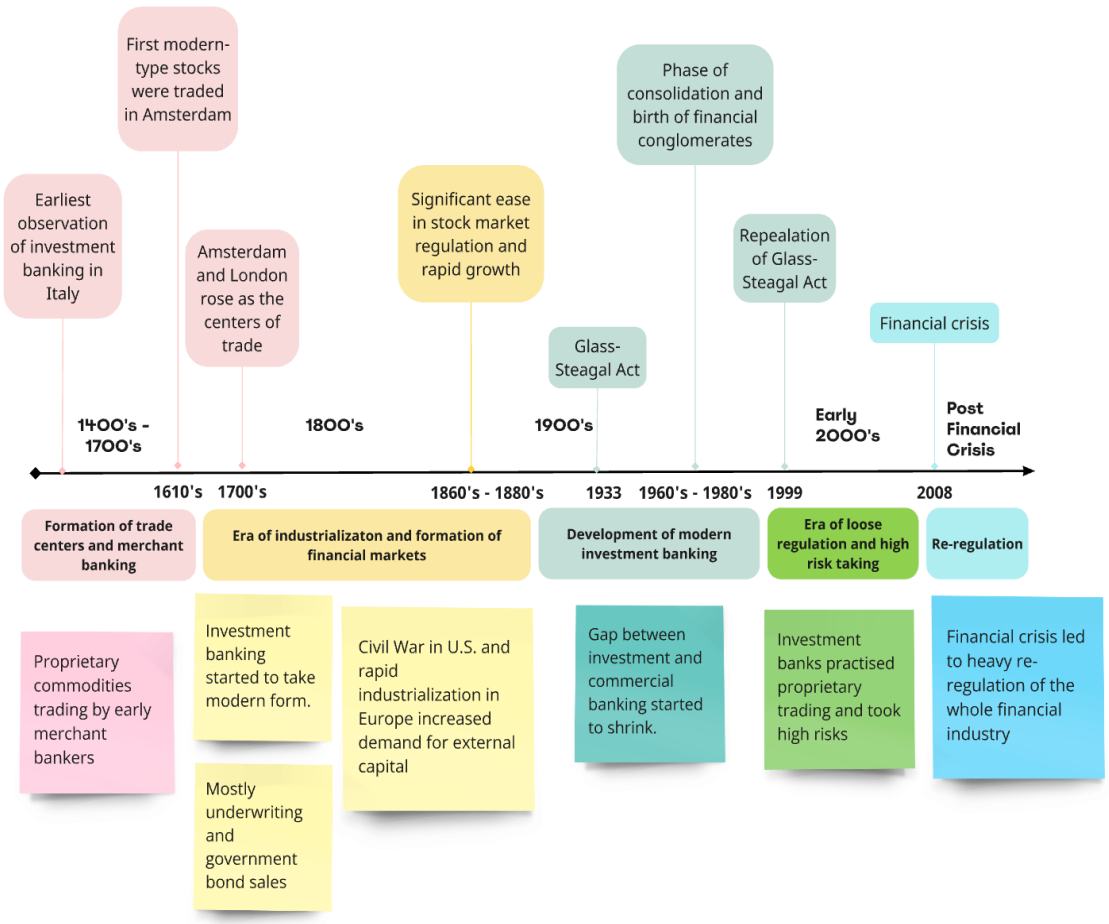



*Figure 1: History of investment banking, timeline. Found also as appendix 4.*

Figure 1. summarizes the timeline for history of investment banking. Colour coding is used to distinguish different time periods.

## 2.2 What are Equity Sales & Trading?

Before diving into the characteristics of Equity Sales & Trading, it is crucial to define the client-vendor relationship of capital markets. In capital markets, this divergence is typically referred to as “buy-side” and “sell-side”. The buy-side refers to firms that buys securities such as pension funds, investment managers and hedge funds. Their performance, at least in longer term, is usually dependent on the performance of those securities bought (or sold). On the other hand, sell-side refers to the firms that sell, trade, or issue securities. Those institutions include investment banks, advisory firms, and corporations. (Powell)

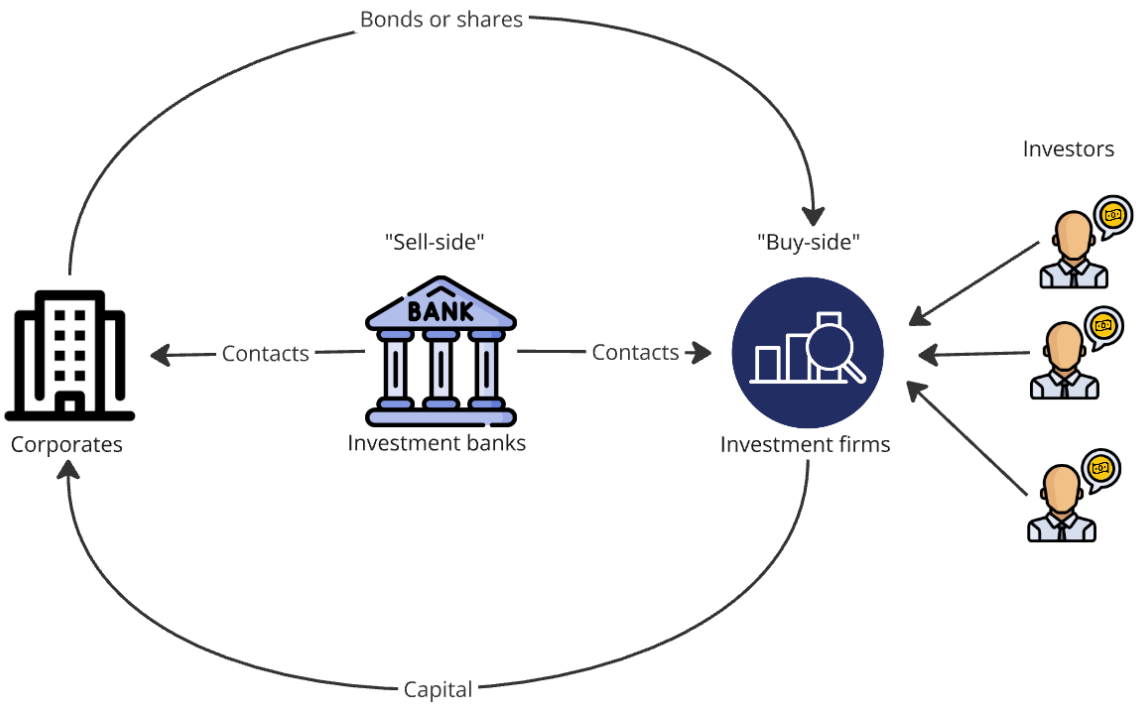



*Figure 2: Relationship of buy-side and sell-side. Following: Powell.*

In practice, the relationship between sell-side and buy-side works as following: Corporates seeks to raise capital and is in contact with investment bank. A specialized investment bank creates necessary financial models to price the share or bond issued and prepares variety of marketing material that is distributed to potential investors, for example buy-side portfolio

managers. Then the buy-side manager decides whether to invest in those securities or not. The relationship between buy-side and sell-side extends also to secondary markets where sell-side acts as intermediary between sellers and buyers of securities. These kinds of firms that focuses on finding matching counterparts in securities traded in secondary markets are called brokerage houses and are part of Sales and Trading departments of investment banks. Unless separately mentioned, in this master's thesis author is referring to a brokerage house operating under a bank when discussing about sales and trading firms. This is important to note as there are different regulatory considerations depending on the financial service that the firm is offering. (Powell)

Equity Sales & Trading is typical division of investment bank which focuses on acting in secondary securities markets offering dealer or/and brokerage services. Dealer is a person who buys or sells securities for their own account and a broker a person who buys or sells securities for their client's account (SEC). Equity Sales focuses on pushing their equity ideas to, and receive equity orders from, their clients. The fundamental role of an Equity Sales is to be the primary link between the sell-side investment bank and the buy-side client. Traditionally they are responsible for filtering, summarising and marketing of equity analysts' research for clients and for the maintenance of client relationships. They communicate with companies' executives and try to gain information their clients might be interested of. Whereas Equity Sales have the client facing functionality, Traders can be seen as having the distinct market facing functionality in buying or selling shares. They are responsible for executing the client's (in case of a brokerage house) orders with the best possible price by either finding the natural counterparty for the trade, or by using digital channels such as electronic stock exchanges and trading algorithms. (Lee & Manochin 2021)

In addition to order handling, equity brokerage houses usually provide buy, sell, and hold recommendations and research for their clients, or in return for charge. Historically research fees have often been bundled with brokerage commissions, but after EU's MiFID II legislation, brokerage houses operating in EU have been made to separate brokerage commissions from research costs when operating with most financial services sector clients (HSBC 2018). In United States research fees are still often bundled with brokerage commissions as it exempts broker-dealers from having to register as investment advisor (CFA Institute 2019). Nowadays most of the brokerage business's earnings are yielded from

the fees and commissions gained from matching the participants of the trade as current regulation forbids short-term proprietary trading for U.S. banks and heavily regulates it for European banks. However, there are some risk-taking methods which are allowed for banks' brokerage houses that might lead to profits or losses. These methods include for example guaranteeing a certain price benchmark such as volume weighted average price (VWAP) or giving price to client upfront. Getting a better price compared to the given price or benchmark leads to profits and other way around. (Michalczyk 2012; Liaw 2011; Leshik & Cralle 2011)

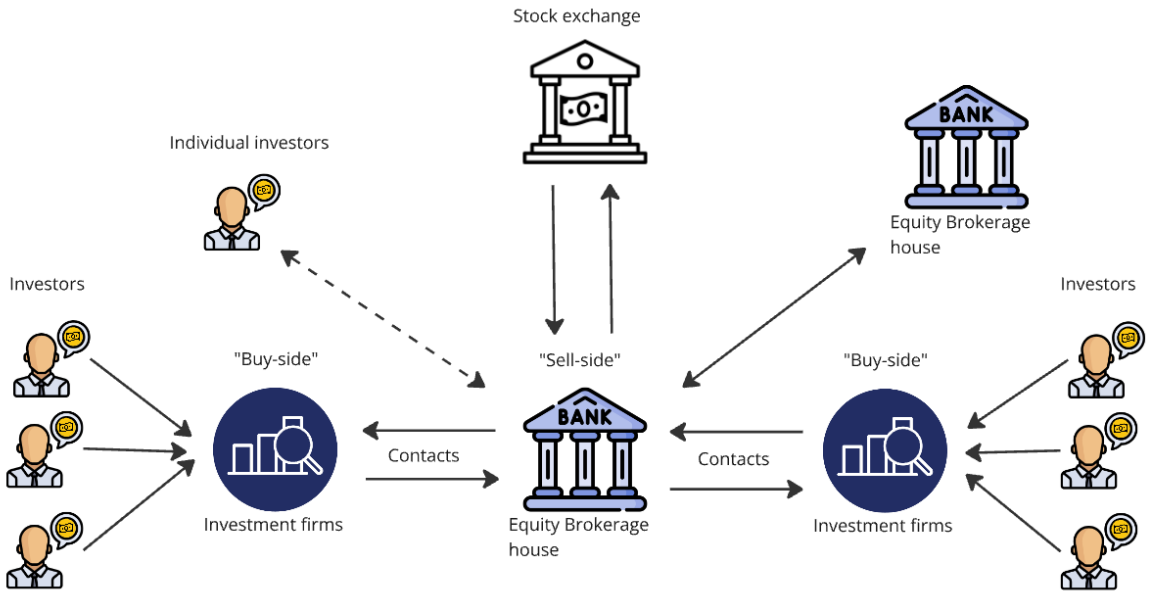



*Figure 3: Equity brokerage business. Following: Michalczyk 2012; Liaw 2011; Leshik & Cralle 2011*

The fundamental goal of equity brokerage houses is to earn commissions by acting as an intermediary between sellers and buyers of shares. The picture above describes the principality of equity brokerage house including its relationship with the buy-side. Brokerage house's buy-side clients are typically institutional investors such as fund management firms, pension funds and hedge funds, other brokerage houses, and individual investors. Arrows indicates the flow of information, money, and shares. The line between brokerage house and individual investors is dashed as some brokerage houses focuses on only serving institutional customers. Buy-side portfolio managers pool money from investors or use their own capital to buy or sell shares. Having an investment bank as an intermediary has some benefits. It gives individual investors access to global equity markets

which would otherwise be hard to acquire and provides more liquidity by having connection to buy-side and advanced tools to use in electronic stock exchanges. Improved liquidity again helps the execution of larger transaction by mitigating implementation shortfall. Implementation shortfall is the difference between asset's theoretical net value and the net value after completing the transaction. Transactions that are too big compared to available supply or demand changes the supply / demand equilibrium leading to worse price originally anticipated. On the other hand, the longer it takes to complete the transaction, the higher is the chance of unwanted market movement. Brokerage houses try to tackle this problem by either trying to find direct counterparty for the transaction that would agree to pay the current market price for whole transaction size or to trade the transaction in electronic stock markets in a way that mitigates the implementation shortfall. However, there are different kind of brokerage house models, and the illustration presented here is only one of a kind. For example, some brokerage houses are solely focused on serving institutional customers and some are more retail focused (CFI). (Downey 2024; Perold 1988) In this thesis brokerage houses that serves primarily institutional customers is mostly being focussed to.

## 2.3 Development of Equity Sales & Trading

Like the whole investment banking industry, Equity Sales & Trading has also gone through significant changes. Back in 2000, Goldman Sachs employed around 600 equity traders who fulfilled the orders of their large clients in its New York's headquarters. By the year of 2017, there were only two equity traders left. Algorithmic trading has taken over Goldman Sachs's equity trading and is supported by 200 computer scientists. Even though the number of cash equity traders in Goldman Sachs haven't been stated in recent years, the direction of the industry is clear. It is also safe to say that the rest of the industry is moving into same direction as it is estimated that almost 45 % of the revenues from cash equities trading industry came from algorithmic trading already during 2017. (Byrnes 2017) Another change driving factor in brokerage industry is the ever-increasing pressure to lower trading costs. Alongside to the traditional full-service brokerage firms has become a variety of new brokerage business models offering significantly cheaper execution options. These extremely low commission brokerage houses are usually only providing execution services to their clients and use highly automated trading strategies to keep their own costs low. This has led to steadily declining per share brokerage commissions and forced traditional

brokerage houses to examine their own business models. (Goldstein et al. 2009; Banton 2024)

Despite revolutionizing the trading industry just as in early 2000's, the first trading algorithm was written by Alfred Jones as early as in 1949 to serve the world's first hedge fund. During the 1960's computers were mostly used to analyse price movements of stocks, and in 1970's to execute block trades, trades with minimum of 10 000 shares or $ 1 million in worth. The actual adoption of algorithmic started during 1980's when group of Morgan Stanley employed scientists begun to used computer algorithms for so called statistical arbitrage. In this strategy the computer algorithm would simultaneously buy one security and sell other in a way that would lock in a return regardless of the market's move. As the computational power kept on growing and the capabilities of trading algorithms started to shape the industry, it was soon discovered that to fully benefit from the seemingly endless opportunities of algorithmic trading, the employees had to own new kind of skills. The new recruits of Wall Street were demanded to have skills in physics, statistics, mathematics and computer science, and a new job title "quantitative analyst" was born. (Leshik & Cralle 2011)

Nowadays trading is more and more data driven and big data has been widely used to recognize patters and trends and to forecast future outcomes. One of the major technological developments affecting modern data analytics has been the ability to analyse unstructured data. This innovation has enlarged the available data sources and enabled the use of images, speech and natural language as a data source. However, as the usage of big data has grown exponentially, big data has started to lose its value. To keep on benefitting from big data, traders are required to create efficient processes to turn large quantities of data into valuable insight. (Akyildirim et al. 2021; Ali, Suwabe & Walsh 2016)

On the buy-side, asset managers use publicly available big data to create models that recognize trends in economies and evaluate companies fundamentally to gain valuable insight. Modern trading algorithms use historical data to optimize their performance. Especially on the sell-side, the sole purpose of trading algorithms is not only to predict price movements in shares but also to offer tools for better execution of client orders and hence can be roughly divided into two categories: execution algorithms and to proprietary trading algorithms. Broker-traders (traders executing client orders) typically use execution trading algorithms to minimize the trading costs and to optimize execution strategy of their choice.

Execution algorithms might for example try to target a certain price benchmark such as volume weighted average price of the day, aim to mitigate implementation shortfall, or to seek liquidity from multiple venues and dark pools. Indeed, according to survey conducted by Greenwich Associates (2019) top three most used execution algorithms were liquidity seeking algorithms (29 % in US, 31 % in Europe), VWAP targeting algorithms (21 % in US, 19 % in Europe), and implementation shortfall mitigating algorithms (16 % in US, 10 % in Europe). Algorithmic low-touch trading, highly automated form of trading, also known as electronic trading, handled approximately 41 % of all buy-side equity order executions in United States during 2018 and was expected to grow to 47 % during next three years. According to J.P. Morgans annual trader survey, in 2024 electronic execution of equities covered already 65 % from all buy-side orders and was expected to keep on growing (J.P. Morgan). (Ali, Suwabe & Walsh 2016; Johnson 2019; Nuti G. et al. 2011)

But what will the future of Equity Sales & Trading look like? JP Morgan conducted their annual trader survey of 2024 surveying over 4000 institutional traders across asset classes. Traders from over 65 countries participated for the survey that was held for the eighth year now. From the technological viewpoint, 65 percent of the traders surveyed saw that artificial intelligence and machine learning will be the most influential in shaping the industry over next three years. 13 % of the surveyed, the second biggest group, believed the API integration to be the most influential. Electronic trading of equities was expected to keep on growing steadily. In 2023 electronic trading of equities covered around 65 % of the executed buy-side volume whereas in 2025 the share was expected to grow to 69 %. Access to liquidity and low brokerage commissions were held in high value. Hence to conclude the survey of 2024, artificial intelligence, and machine learning were expected to be most influential technology wise, electronic trading of equities is expected to continue its rapid growth, and the low brokerage commissions and access to liquidity were highly valued. (J.P. Morgan)

As the future of trading is evolving, the role of equity sales is also forced to adapt to the new landscape of the industry. Report *Future of Sales* (2019) written by PWC discusses about the future of the sales in investment banks and states that the traditional sales functionality is no longer fit for the purpose. The main drivers of the change are changes in client behaviour, new technology, regulatory changes, and focus on returns. Clients are moving traditionally bank offered services such as research and solution capabilities to in-house and

are demanding more connectivity, tech solutions and direct access to banks' traders. Increased adoption of technology is supporting electronic trading facilitation, advanced analytics on both client's and sales' side, and automating manual processes. Regulatory changes are yielding more transparent and automated markets as well as better capture of data from the whole lifecycle of the client. Banks publicly stated financial targets affects all areas of business putting pressure on generating more revenue and reducing costs. (PWC 2019)

The skillset required for capital market salespeople is evolving as client interaction with banks changes. While traditional attributes remain relevant, analytical and entrepreneurial skills are especially addressed in many of the discussions had while creating the PWC's report. Significant changes are evident in transactional sales roles due to the shift towards electronic execution. Banks are seeking salespeople who are quantitatively adept, technically skilled, and can handle coding and connectivity solutions. This has led to younger e-sales teams managing a broader range of clients across asset classes, moving away from traditional relationship management and deep product knowledge. The banks' adoption of advanced Customer Relationship Management (CRM) systems like Salesforce and Microsoft Dynamics has transformed sales management. These platforms allow for sophisticated analytics and workflow optimizations, with leading banks using data analytics for advanced client segmentation and customized service offerings. This trend is expected to continue influencing the industry in the years to come. (PWC 2019)

*Table 2: Development on Equity Sales & Trading*

| Development of Equity Sales and Trading | |
|---|---|
| **Past** | |
| **Trading** | **Sales** |
| • More profits from proprietary trading (before re-regulation and financial crisis)<br>• Higher cost of execution<br>• More high-touch (manual) approach on order execution<br>• Less regulation | • Deep knowledge of products (i.e. shares, management, markets)<br>• High focus on client relationship management<br>• Link between client and trader |
| **Now** | |
| • Highly regulated<br>• Around 65 % of orders handled with electronic execution (highly automated)<br>• Low cost of execution<br>• Very limited risk taking in banks' brokerage houses | • Deep product knowledge still held in high value<br>• Client relationship management still focus point<br>• Many future characterestics alredy apply<br>• The shift toward future requirements is seen as challenge in trying to find fit workforce |
| **Future** | |
| • Increase in Electronic execution of orders<br>• Applications of AI and machine learning<br>• Further developed trading algorithms (enhanced with AI)<br>• Further decreasing cost of execution<br>• More direct connection between traders and buy-side as clients are relying increasingly to execution only services | • New required skill set emphasizing tehcnical and quantitative skills<br>• Broader ranges of clients across asset classes<br>• Less product focused sales and more solution and services based approach<br>• More data centric focus to sales management<br>• Use of data analytics for client segmentation and more customized service offering |

The Table 2 *Development of Equity Sales & Trading* summarizes the main points on the development of Equity Sales and Trading. In the past, especially before financial crisis of 2007-2008 industry looked a lot different. There were significantly less regulation, and a lot of the profits yielded from trading came from immense risk taking. After the re-regulation of post financial crisis, risk taking in banks' brokerage houses has become very limited and proprietary trading has practically ended. During past years and decades, the cost of execution has constantly been decreasing and the share of electronic execution of order flow has been increasing steadily. The same trend is also expected to continue in the future. Artificial intelligence and machine learning are expected to have significant effect on the industry, especially in few upcoming years. Salespeople of investment banks are required to have new skills alongside with the traditional client relationship management and product

knowledge. These skills include technical and quantitative capabilities and wider understanding of customers technical needs. In the future sales are expected to serve larger variety of clients across different asset classes and have more understanding on the technical solution and services that clients demand. Client relationship management is becoming increasingly more data centric and use of advanced customer management tools such as Salesforce is going to be in the centre of this development. This thesis focusses on recognizing applications of data analytics from global literature, evaluating which of those could be applicable for target organization, and developing target organization's capabilities in client data analytics.

# 3 Client data analytics in investment banking

As the current research and professional consensus points out, investment banks are constantly adopting new technologies and innovations to keep on benefiting from global technological development and to preserve their competitive advantages. According to an article *Data Analytics in Investment Banks* written by Iraqi et al. (2021) investment banks have been one of the first and most important users of data analytics. Indeed, the industry has been ranked as a number one most data consuming industry in statistics of 2016.

## 3.1 Definition and applications of client data analytics

Few of the factors further accelerating the adoption of data analytics during the past few years have been the stagnation of revenues, fallen margins, the rising complexity of regulation, and increased competition. With the successful implementation of data analytics investment banks have become better at risk management, customer management and internal monitoring to name a few. Quantitative risk models can detect fraudulent activities by analysing atypical operation from historical patterns and give alerts to appropriate personnel. Analytics can be used to keep track of short- and long-term liquidity in real time, and to run simulation to guarantee that necessary funds are always available. Internal client data enriched with external data sourced from web, social media and news can be used to run credit analytics on clients to analyse their financial situation and to prevent credit losses. Client analytics can be used to better understand the needs of clients and the optimal ways of addressing them. Client data sourced from web, social media, news, research reports, customer management tools and trading data can be used to make predictions on what kind of products and services should be offered to clients and when. Data analytics can be used to large and throughout monitoring where problems and patterns of incidents can be identified through machine learning algorithms making responding them as efficient as possible. However, despite the promising benefits of data analytics, the actual implementation of data analytics initiatives is not as common as one would expect. The reason for is the technical complexity and large scale of the implementation process which itself does not guarantee the increase in bottom line. Also, it is hard to find the relevant data

points from the vast amount of data that could be produced daily, and storing large amounts of un-relevant data is expensive. One way of tackling this problem that is often used is to first identify the business goals and KPIs (key performance indicators), and then start collection necessary data. (Iraqi et al. 2021) This research explores different ways in which client produced data can be used to support equities brokerage business operating in Nordic located bank.

Data analytics has become a buzz word in today's data-driven world and is often used in various contexts and industries. Its definition can sometimes be hard to express, as it encompasses a broad spectrum of disciplines and methodologies. A comprehensive definition provided by CETIS (the Center for Educational Technology and Interoperability Standards) describes data analytics as "the process of developing actionable insights through problem definition and the application of statistical models and analysis against existing and/or simulated future data." This definition underscores the transformative potential of analytics to influence decisions and strategies in real-time scenarios. (Taylor-Sakyi 2016) The core of data analytics lies in its ability to produce actionable insights. Unlike mere data analysis, which often stops at describing or diagnosing past events, analytics aims to equip decision-makers with the knowledge to take informed actions. This distinction highlights the forward-looking nature of analytics, which not only presents historical data but also anticipates future scenarios. For instance, e-commerce giants like Amazon and eBay utilize analytics through A/B testing to understand customer preferences and optimize their user experiences accordingly. These insights directly inform strategic decisions that enhance performance and customer satisfaction. (Shende 2017)

To better understand the scope and application of data analytics, it is essential to recognize its various types. The literature typically identifies four primary types of analytics, each serving different purposes and answering distinct questions. Descriptive Analytics addresses the "what" questions about data. It involves summarizing past data to understand what has happened. This form is often visualized through dashboards and reports that provide a snapshot of metrics and trends over time. Diagnostic Analytics moves deeper and answers the "why" questions. It involves looking at past performances to determine cause and effect, identifying patterns that could explain outcomes. This type often uses techniques like drill-down, data discovery, correlations, and pattern recognition. Predictive Analytics focuses on the future. It forecasts potential outcomes based on historical data and answers questions

such as "what is likely to happen?" This type employs statistical models and forecasting techniques to make educated guesses about future trends and behaviours. Prescriptive Analytics is the most advanced form of data analytics. It suggests possible actions to benefit from predictions and insights and answers "what should be done?" by analysing potential decisions and their impacts on future trends. This type often involves complex algorithms and simulation techniques to predict the effects of different choices. (Zikopoulos et al. 2012)

After defining data analytics itself, defining client data analytics is quite self-explanatory. One definition offered by Forbes Advisor is that *Customer analytics is the process of collecting, validating, interpreting and analysing data to gain valuable insights into your customers' preferences, behaviours and needs*. Definition of client analytics can also be yielded from the vendors perspective. Another way of defining client analytics is to identify it as a systematic examination of client produced data to identify, attract and retain the most profitable customers. (Davis 2024; Murphy 2023) Difference between these two definitions is that the first one aims to gain more insight on the way in which clients could be served better while the second one focuses on analysing the profit potential of each client to the service provider. In this thesis client analytics are approached in a way that aims to create value *to* customers, hence corresponding best to the first definition.

Then what is all the buzz about in client analytics? According to an article *Analytics: The real-world use of big data in financial service*s written by IBM Institute for Business Value in collaboration with Oxford University (2013) almost half of the companies operating in banking and financial services industries surveyed answered that from all their big data analytics initiatives and goals, client related analytics were their top priority. The IBM's article listed few factors as an explanation for high prioritization of client analytics. Even though the paper is over 10 years old during the writing of this thesis, these factors can be still seen relevant today. According to the paper banks are shifting from being product-centric to customer-centric. This shift requires a deep understanding of customer needs and preferences. By focusing on the customer as the central organizing principle, banks can tailor their operations, technology, and systems around delivering superior customer experiences. Data analytics enables banks to anticipate changes in market conditions and client preferences, ensuring they remain competitive and relevant in a rapidly changing environment. With insights gained from data analytics, banks can develop and offer new products and services that are specifically designed to meet the evolving needs of their

customers. This not only helps in attracting new customers but also in retaining existing ones by enhancing their satisfaction and loyalty. (Tuner et al. 2013)

The various ways in which data analytics can be used to support client management are immense. Some main areas in which investment banks should focus on when implementing client analytics are getting to know whether clients are likely to consume certain products or services, what is the likelihood of clients taking specific actions and the relevancy of those actions, how likely are clients churning or leaving the bank, and what is the lifecycle value of the client. In the context of a bank's in-house equity brokerage house client's propensity to send certain kind of equity orders could be estimated by tracking the equity research they consumer, or by analysing their recent market activity. Client's likelihood to take specific action could be tracked by analysing historical patterns in their actions. Client churn could be forecasted with analysing clients' sentiment toward the services and products they are offered. Client's lifecycle value could be estimated by analysing the historical and estimating future brokerage commissions they have or are likely to pay. To create throughout picture of a client, client data should be sourced from various perspectives such as behavioural, sentimental, interactional and external. Behavioural data defines the behaviour of client. In the brokerage house's context this data could contain records of the previous market activity such as buy and sell orders or research agreements. Sentimental, or in other words attitudinal data could be sourced from email and chat conversations had with customer, or even from social media post published by the client and analysed with the help of sentiment scoring algorithms. Interactional data contains records on the number of interactions that clients have had with the brokerage house. Those interactions might be derived from physical meetings with sales, email, phone and chat interactions or from the visits to bank's web site. External data is data that refers to client's location, workplace and used devices. In this context it might mean for example the clients preferred working habits such as on site, remote, or hybrid work, availability, and devices that client preferably works with. This method of creating throughout picture of the customer is called "Customer 360". (Boobier 2020)

Some application in which the 360 vision of client can be utilized are client risk profiling, product recommendation systems and client satisfaction analytics. Client risk profiling includes assessing the risk a client is willing to take and the performance level they are expecting. There are multiple ways to assess the clients risk profile from which scoring algorithms based on questionaries are often used. Client's risk preferences can also be

assessed based on their past market activity. (Ertek et al. 2012) Product recommendation systems are algorithms that based on other similar clients past and recent behaviour tries to predict what products clients might be interested in now. The rationale in behind is to give all products a scoring on the likely hood of each client being satisfied with the recommendation. The data used in product recommendation systems can also include information about client's risk profile, strategic orientation, and sentimental data gathered from various sources. (Isinkaye et al. 2015) Client satisfaction analytics can be performed based on data sourced from received complaints, email, chat and phone conversations and from online writings such as articles, blog posts or social media. Most used method to analyse client's sentiment is to use machine learning based sentiment analysis. This method can be used with naturally written language, and some often-used algorithms are decision trees, naïve bayes, and neural networks. (Ali, Suwabe, & Walsh 2016)

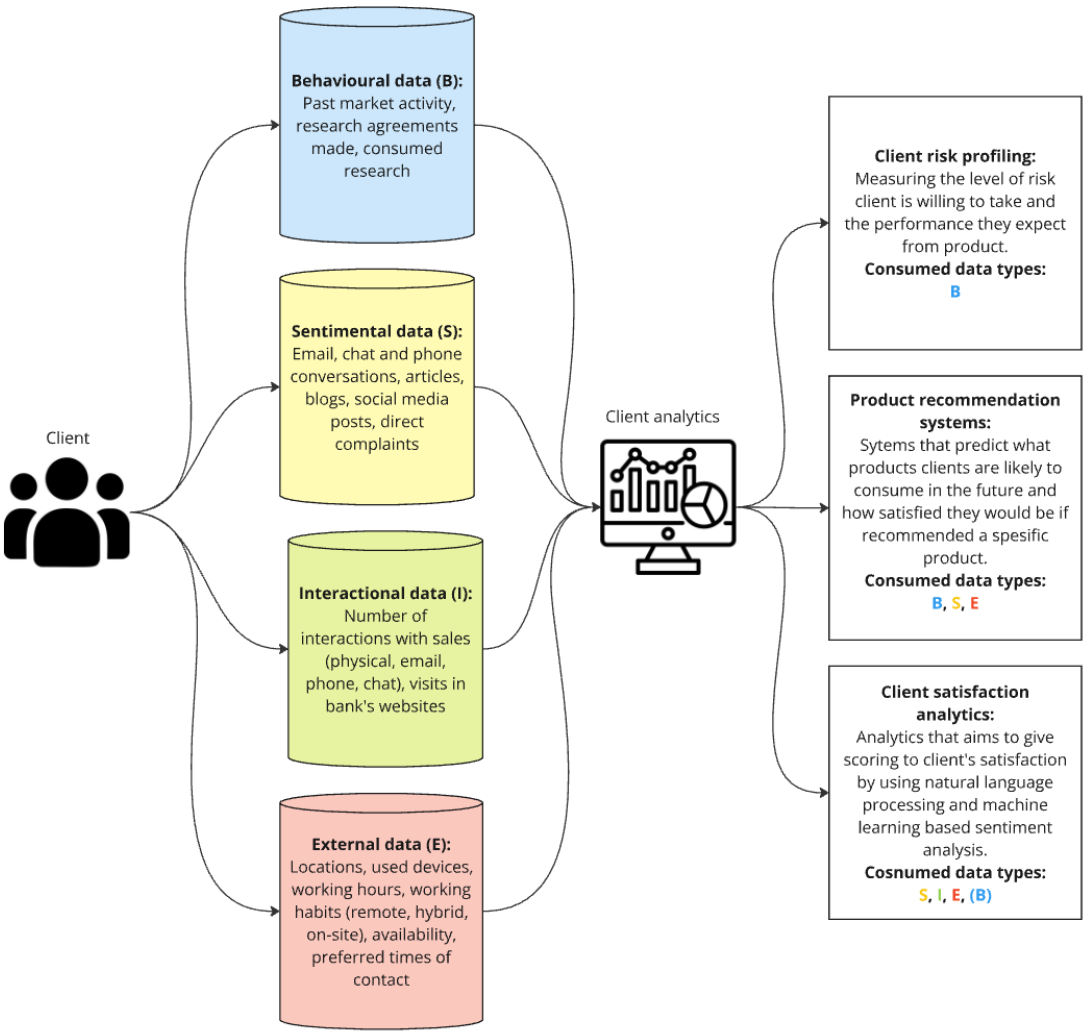



*Figure 4: Applications of client analytics in investment banking.*

Figure 4 presents four different perspectives from which client data should be collected and some ways in which client analytics could be used in investment banking. It also provides authors view on which type of data should be processed to accomplish the analytics goal in hand. In the figure, lines represent the flow of information and colours different perspectives of data.

## 3.2 Challenges and requirements of client data analytics implementation

The promise of significant business value from data analytics depends on establishing a robust information architecture and processes to handle increasing data volume, variety, and velocity. Some of the major challenges in creating value from data is connecting data across organizational silos, creating data analytics supporting organization culture, recruiting workforce with relevant skillset, and building data analytics supporting IT architecture. Especially the problem of separated data silos has been underlined in the banking industry where multiple mergers and acquisitions have created problematic amount of separated data silos across organizations. These challenges can often be over overcome with sophisticated cultural shift towards data-driven organization, investments in skilful workforce and right technology. (Tuner et al. 2013; Iraqi, B. et al. 2023) Other challenges in data-analytics implementation are justifying the cost of new IT equipment despite them being often cheaper that the current database-based solutions, lack of motivation from senior management, limited direction from regulators, and the difficulty of committing for long term project when there are more urgent milestones to hit. (Moody's Analytics 2018) In this chapter requirements for client data analytics are divided into three categories, technological requirements, human resources requirements, and organizational requirements, and challenges associated with the implementation process are assessed.

Implementation of successful data analytics capabilities foundation is complex process which often requires many changes to bank's IT ecosystem (Doko & Miskovski 2019). At the highest level of technological implementation of data analytics is the data architecture. It describes how data is managed throughout the whole data processing including collection, transformation, distribution and consumption. It can be seen as the blueprint for the data and for the way it flows through data storage systems. Well implemented data architecture helps to integrate data across different domains such as geographies and departments and breaks

down data silos without the need to store everything in one place. Modern data architecture solutions often utilized cloud platforms to manage and process data due to their scalability and computing power. Technological requirements of data analytics implementation are dependent on the data architecture model chosen. Modern data architecture designs usually require cloud solutions for storing and processing of data, robust, scalable, and portable data pipelines which supports real-time integration, APIs in applications from which data is sourced to ensure seamless data integration, and business intelligence tools to visualize the data. In practice, this means that the applications from which data is planned to be pulled, needs to have a connection point (API) to enable automated flow of data. Data pulled from various sources needs to be transferred from data source to cloud service, and from cloud service to visualization application through pipelines, requiring designated software. Modern cloud applications, such as Snowflake, can often handle cleaning, validation, standardization, and data mining of the data. Designated business intelligence tools such as PowerBI can be used to visualize the processed data. (Anoshin et al. 2020; IBM a)

Behind every successful implementation of any complex data analytics initiative are the skilful and dedicated employees of the company, sometimes with the help of external consultants. In the context of implementing data-analytics within an investment banking environment, the role of management and organizational structure is essential. The success of data-driven transformations largely relies on the active involvement and sponsorship of top management. Such leadership is crucial as it provides necessary guidance and aligns the interests and expectations of middle managers, who might perceive data analytics either as a competitive threat or hold overly optimistic views about the potential insights it can generate. To facilitate a smooth transition and effective operation, it is recommended that a dedicated data analytics team is established. This team should be tasked with addressing the data needs of the entire organization, ensuring that all business lines have easy access to data and insights, and integrating essential skills and talent to support data-driven initiatives. Furthermore, the implementation process should include the integration of specific roles categorized into two main families: technology-enabler roles and business-impacting roles. Technology-enablers, such as data engineers and developers, are responsible for constructing the necessary infrastructure, designing data pipelines, and ensuring the accessibility of data analytics tools. On the other hand, business-impacting roles, including data scientists, business analysts, and data architects, focus on utilizing the technical resources to analyse data and generate relevant business insights. These roles require a

profound understanding of the business dynamics to effectively convert data into actionable intelligence. The shift towards a data-driven model not only necessitates the introduction of new roles but also requires a significant transformation of existing roles within the organization. This adaptation is essential to accommodate the changing demands of data management and analysis, ensuring that the organization remains competitive and responsive to the insights derived from data analytics. (Iraqi et al. 2023; Kremser & Brunauer 2019; Kutzner, Schoormann & Knackstedt 2018).

To fully benefit from the use of data analytics, whatever the case of use, investments into technology and skilled workforce are not often solely sufficient. Cultural shift towards data-driven organization is often needed to instil the right values to all employees. According to research article written by Iraqi et al. (2023), to fully benefit from data analytics, investment banks need to make sure that their organizational culture is in line with few values. Firstly, all decision making should be data driven. In contrast to traditional organizations where the decision making is solely in the hands of management, in data-driven companies decision making is heavily dependent on the data's insight. This is especially important in an environment where there must be high attention to detail and similarities between past and present situations may be misleading. Decision making in investment banks is often reliant on external and internal data points such as transaction history, market data, and client data hence making feeling-based decision making very risky. Another highly important value in data-driven organization is the seamless access to data and provisioning between business units. Data silos, collections of data that are isolated from an organization and inaccessible to all parts of the organization, are not accepted in data-driven organizations unless there are regulatory causes. Data access should be democratized to all parts of the organization regardless of the ownership of the data and be used by other business lines to enrich their own data and to provide them better insights. Given the strong specialization of different departments within investment banks, data considering markets or clients is usually scattered throughout different teams. Mutualization of this data can be very beneficial for bank's client relationship management and could provide valuable insight on relevancy of offered services. However, there are some regulatory challenges that affects the seamlessness of data access on which will later be returned to. Third important value pointed out in the article is the open discussion and knowledge sharing about data processing. Its main goal is to get everyone to understand the different data processes, and to understand that different pipelines, and data mechanisms are shared with everybody. This value is necessary because

it engages everyone to be involved with the change to data-driven organization and it makes clear that everyone can use, participate and bring new ideas related to use, structure and usage of data. (Iraqi et al. 2023) However, regulatory limitations often set boundaries for use of data across business domains. Clients are allowed to decline banks the use of their data and some information cannot be shared with other departments due to Chinese Wall arrangement in investment banks. In investment banking context, Chinese Wall means an ethical barrier that attempts to prevent conflicts of interest. In practice this might mean for example preventing Sales & Trading getting market sensitive information such as stock downgrades or information regarding mergers and acquisitions before it is public. By having this information in hand, it would be tempting for Sales and Trading to benefit from that information hence perpetrating on insider trading. (Chandani & Mehta 2015; Du 2023)

*Table 3: Requirements and challenges in implementation of data analytics in investment bank*

| Technological requirements | Human resources requirements | Organization cultural requirements |
|---|---|---|
| • Applications that are used as sources of data needs to have APIs<br>• Pipelines are needed to transfer data between source application, processing and storing software, and visualization software<br>• Modern data architectures utilize cloud platforms in processing and storing data<br>• Visualization software is needed to gain insight from data | • Management has to be commited as success relies on active involvement and sponsorship<br>• Overly optimistic expectations and feelings of being threatened has to be managed, especially for middle management<br>• Designated team for data analytics should be lined up<br>• Techonology enabler roles such as developers and data engineers should be hired to desing pipelines and infrastructure and to ensure accessability to data<br>• Business enabler roles such as data scients, business analyst and data architects should be hired to create business insight from data<br>• Transformation of existing roles to support new analytics goals | • Cultural shift towards data-driven organization culture meaning that:<br>• Decisions making should be based on data insight instead of gut feeling of upper management<br>• Seamless connectivity between organizational data silos<br>• Democritized access to data: everyone should be allowed to access, enrich and use data in their own business application within regulatory limits<br>• Knowledge sharing on data processes: everyone should understand the basic data processes and be able to bring up their own ideas to improve data processes |
| **Challenges associated with:**<br>• Data is often fragmented through out various organizational silo<br>• Data can be in different formats and require lot of pre-processing (such as standardization, validation and cleaning) before its ready for use<br>• Data source systems might not have APIs available<br>• Cost of new IT equipment can be hard to justify despite it been cheaper in long-term compared to current | **Challenges associated with:**<br>• It is often challenging to motivate senior management to such change<br>• Talent acquisition is challenging especially in competitive job market<br>• Continuous traning of new and existing staff is resource consuming<br>• Changes in roles might be un-motivating especially if employee is not motivated to new tasks | **Challenges associated with:**<br>• Shift in organization culture is long project and requires full commitment from management<br>• Well established organizations often face significant resistance toward change<br>• Regulatory limitations, especially in investment banking enviroment, might have negative impact on many priciples of data-driven organsational culture. For example, there are limitations in sharing data between private and public business domains |

Table 3 presents conclusions from chapter 3.2. It combines findings from theory with authors own conclusions. Each colour represents category of recognized requirements category.

## 3.3 Data analytics architecture and data mining applications for Sales & Trading organization

As one can imagine, the scope of this topic is gigantic and there are numerous books written about it. To serve its purpose in this thesis, it needs to be scoped so that it is relevant in the context of equity brokerage business. This chapter includes theory of data analytics architectures from parts that are relevant considering the thesis's context and focuses on processing theory of data architecture in high level without going too much into technical details. It also aggregates different sources of theory into one interpretation.

As described in previous paragraphs, technological requirements of data analytics implementation require defining the data sources, data pipelines between applications, tools for processing and storing data, and tools to visualize it. The general pipeline of data analytics consists of data collection and integration, data preparation, data mining, presentation, and knowledge interpretation. Data often comes from various sources and in different formats. Different data source applications need to be integrated to the processing software. Data collection is one of the most important steps in data analytics as the quality of collected data carries significant weight to the insights that can be gained from the data. Subsequent processes of data analytics are likely to become useless if collected data does not describe the overall system or process sufficiently. Raw data collected from various systems is rarely suitable for use as it is. Raw data can have missing values, contain outliers, be in different scale compared to data from other sources, or be in wrong format. Hence, preprocessing of data is often required to make raw data into usable form. Processed data can be enriched with data mining, a process of gaining insight from existing data with the use of statistical analysis and machines learning algorithms. Designated machine learning or artificial intelligence algorithms can be trained using historical data and be used to uncover patterns and other new information that can be used in visualisation and interpretation. Finally, the data that is ready for use can be visualized and interpreted. (Holdsworth 2024; Kalita et al. 2024)

Data can be stored in multiple ways throughout the analytics process. Data lakes are repositories that hold raw, structured or unstructured data collected from various sources and in various formats. Due to their scalable and open formats, data lakes can accommodate all types of data including images, audio files, and social media posts in addition to more

commonly stored data types. These repositories are usually massive in size and the data they contain needs a lot of processing and preparations before it is ready for use. Data lakes are beneficial because they are low-cost option to aggregate data from multiple sources and provide accessibility to data across organization by eliminating data silos. In equity brokerage businesses context, they are also useful because there are regulatory requirements for storing transactions related data (ESMA). (Microsoft Azure, IBM b)

From data lake, necessary data can be transferred into a data warehouse. Data warehouses are centralized storages of data that consolidates raw data from various sources such as data lakes or source applications and stores it in a structured form. They contain data that is in unified form and ready for data mining, artificial intelligence and machine learning. However, data warehouses are difficult and expensive to scale because data often needs to be pre-processed from raw format to a data warehouse specific format. Hence it is important to choose what data is relevant from the business' perspective to warehouse. (Microsoft Azure; IBM b) The process of storing data into data warehouse requires ETL (extract, transform, load) process. ETL process is a set of business rules that are set for cleaning, organizing and preparing of data for storage, data analytics and machine learning. ETL's extraction phase includes extracting or copying data into staging area, an intermediate storage area for temporary storing of extracted data that is used in transformation phase. Transformation phase includes basic data transformation methods such as data cleansing, deduplication and format revision and advanced data transformation methods such as derivation, joining and splitting of data, summarization, and encryption. Basic data transformation methods focus on bettering data quality while advanced data transformation methods focus on making data easier for analytics. Loading phase of ETL moves transformed data from staging area into target data warehouse. Staging area is often cleared after this process is done. (AWS)

From data warehouse, ready-to-use data can be analysed with designated business intelligent tool or further transferred into data marts, focused versions of data warehouse. Data marts contain a smaller subset of data imported from data warehouse and can be used by a smaller part of business such as teams or departments. Construction of data mart is a complex process that requires building a physical database, filling it with data, and deciding access rights and management protocols. Despite the challenging process of building it, it provides a quicker way to gain insights from relevant dataset compared to working with full dataset. In equity

brokerage houses context, data marts can be used to tackle regulatory limitations of sharing the data. Research analysts could have their own data mart build with data necessary to them and contain information that is not allowed to be shared with sales and traders, and vice versa. From data marts, data can be again utilized with business intelligence tools. (IBM c)

The last part of data analytics pipeline is the visualization of data. It is a process in which data is represented in graphical form by using visual elements such as graphs, charts, and maps. With the help of visualization, it is easier to understand trends and patterns in data and detect outliers. It is also excellent way to present data to non-technical audience. The ways in which data can be visualized depends on the type of data and the goal that is wanted to achieve with analytics. Some common methods of visualization are tables, histograms, scatter plots, line, bar, and pie charts, timelines, and Venn diagrams. In this thesis it is not necessary to go more in depth on the properties and benefits of each diagram or chart type, but it is worth mentioning that each of them has their own places in well implemented data visualization. (Miller 2017; Tableau)

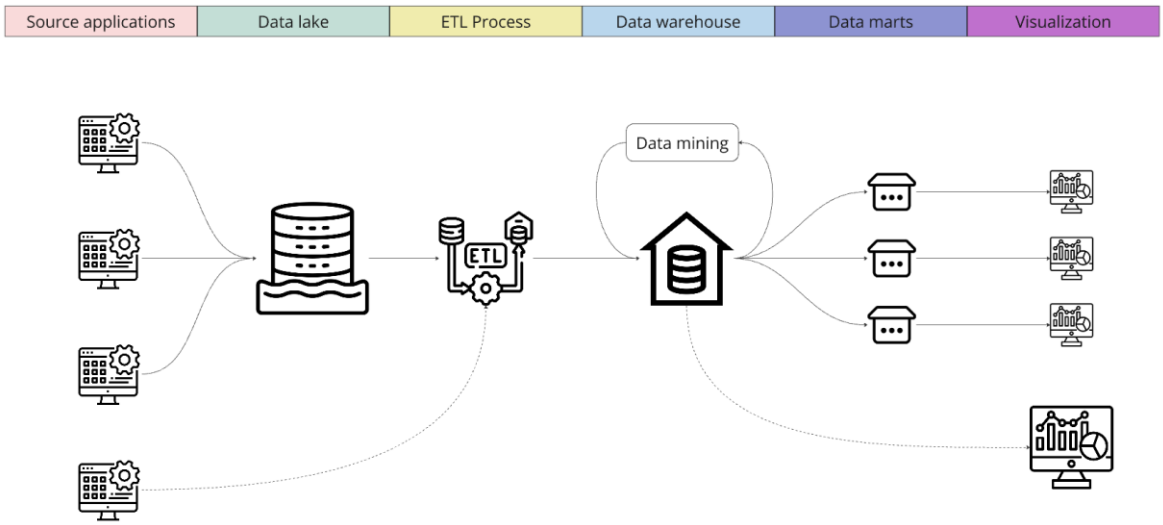



*Figure 5: High level data architecture for equity brokerage house*

Figure 5 presents one plausible data architecture that is based on theory and author's conclusions. It utilizes data lake to eliminate organizational data silos and because it is a low-cost option for storing regulatory adequate data. It includes data warehouse to store business relevant data in a structured form, and data marts to make data analytics more efficient for each team and to manage data access rights from regulatory perspective. Line between source application and ETL process, and line between data warehouse and analytics are dashed to illustrate that data lake or data marts does not have to always intermediate data

flow between source applications and ETL process or data warehouse and analytics. For example, designated data analytics team without regulatory limitations could have access to the whole data warehouse and use it for overall analytics and risk management.

As highlighted in previous chapter, the definition of data analytics underlines its forward-looking nature. Data mining is a form of data science that focuses on discovering knowledge, patterns, and insights from an organized collection of data. It is the most popular technological tool to convert large quantities of data into useful information. The implicit assumption in data mining is that it is possible to find patterns and trends from data that could be projected into the future. For example, a useful pattern to recognize from equity brokerage house's perspective would be "a client who buys stock A and stock B at the same time has 90 % chance of selling stock C." Another example of useful pattern would be to notice if client consumes equity research notes on stock A, has been historically active in the same stock at current price levels, and hence has 80 % chance of buying or selling the stock. With this knowledge in hand, brokerage house's salespeople could target those indicated stocks in their conversations with the client, leading possibly to higher client satisfaction and enhanced brokerage commissions. (Miyan 2017; Maheshwari 2014) There are numerous different data mining methods and algorithms. In this thesis author is focusing on associate rule mining and two classification algorithms, k-mean clustering and decision trees as they have been noted especially beneficial for client analytics in banking context (Hassani et al. 2018).

With the help of fitting data mining algorithms banks can enhance their customer relationship management and its potential in acquiring new and retaining existing customers, and in maximizing their lifetime value (Vijay & Woldesamuel 2012). K-Means clustering is a clustering machine learning algorithm designed to find similarities between data based on its set of features. The algorithm aims to construct k number of clusters based on distances between data points. Model is trained using historical data and can handle millions of data rows. In banking context, K-Means clustering is often used to build accurate client profiles, for sufficient client segmentation, conduct client -centric business strategies, targeted marketing and product sales, for client churn prediction and prevention, attracting new clients, and for future service design. (Hassani et al. 2018; Piech 2013)

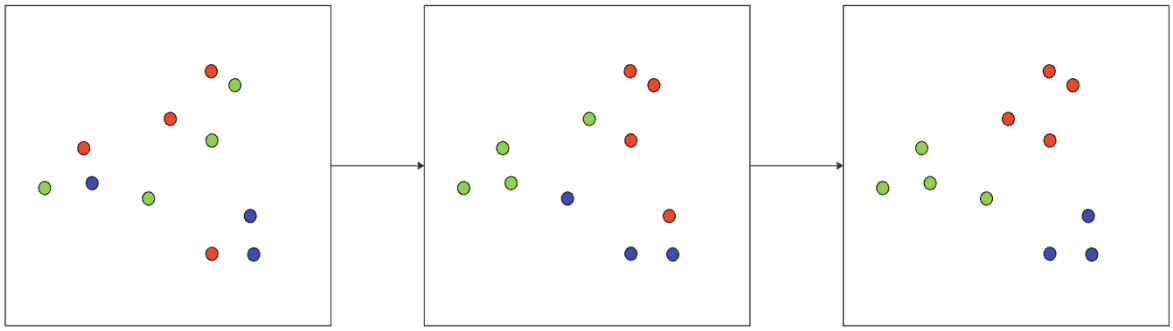


*Figure 6: Formation of k (3) clusters. Following: Hassani et al. 2018; Piech 2013.*

Figure 6 represents the iterative formation of clusters in K – means clustering. After each iteration, the algorithms seek to find better representation of clusters based on the learnings from previous iteration. Even though figure 6 is two-dimensional representation of clustering, K – Means algorithm can handle n dimensional data. Visualization of K – Means is possible only up to three-dimensions and in in case of $n > 3$ dimensions, only mathematical representation is possible. (Piech 2013) In brokerage house's context, each cluster could represent a type of client segmented by their willingness to take risk or preferred investment strategy.

Decision tree is a classification machine learning algorithm which goal is to create a model that predicts the value of target variable by learning set of decision rules based on the data's features. Some benefits of decision trees are that they are quite simple to understand and can be visualized. Decision trees are often used for client profiling, targeted marketing, client churn prediction and prevention, and strategic service design. (Hassani et al. 2018; Scikit Learn a) Decision trees consist of root node, internal node, and leaf node. Root node is the first conditional node of the tree, and it divides tree into two branches based on the outcome. Internal nodes are also conditional nodes that divide the tree further based on the outcome of the condition. Leaf nodes, also called terminal nodes, are nodes from which tree is not further continued as there are no longer conditionality in remaining data points. Leaf nodes can originate from internal or root nodes. (IBM d)

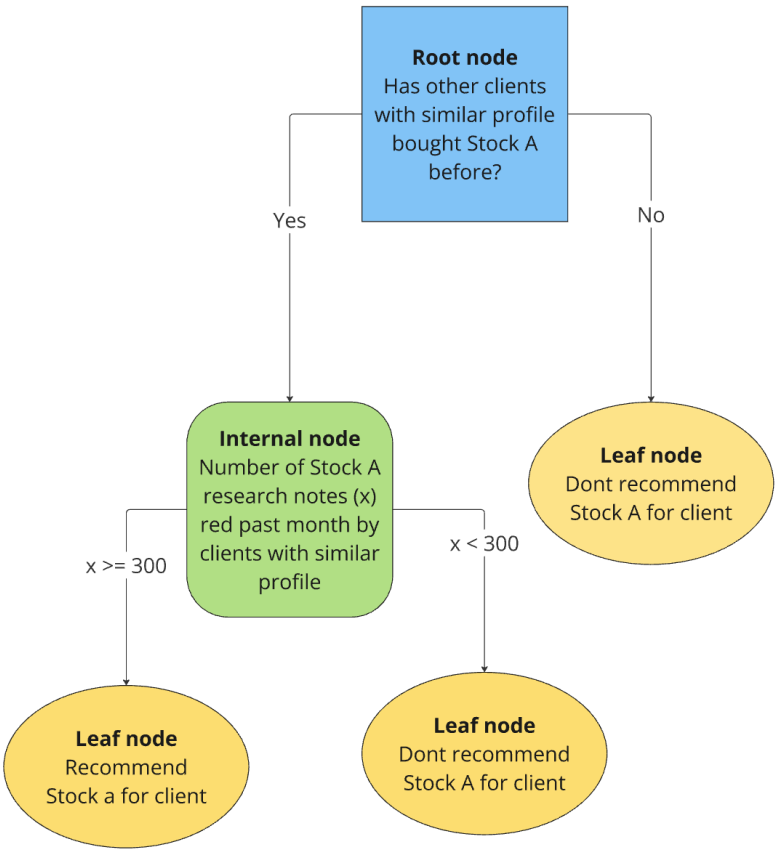



*Figure 7: Visualized decision tree. Following: IBM*

Figure 7 represents the decision tree in its visualized form and is the outcome of the decision tree algorithm. If such tree could be constructed from set of data, it would imply that historically clients have been more likely to buy Stock A if, firstly, other clients with same profile have bought Stock A before, and secondly, if clients with same profile has red 300 or more research notes on total past month from Stock A. Increased likelihood of client being interested in Stock A gives recommendation signal. However, few question are likely to arise. How are the conditions of decision tree and its structure decided? The simplified answer to the question is that decision tree algorithm iterates through every possible splitting of conditions and chooses the splits with lowest amount of uncertainty given the previous conditions. (Roy 2020; Scikit Learn a)

Associate Rule Mining (ARM) is a data mining method which can be used to assess the association between different items. It is also known as Market Basket Analysis as its roots originates from stores trying to identify which items are often bought together. The main goal of this data mining technique is to find out which products are often bought together and use that information for marketing strategies. Grocery stores might for example locate

products that are often bought together near each other to make customer's shopping easier and to improve their sales. In practice, this might mean that customers who are buying pen and paper, have also 75 % chance of buying eraser. Hence, placing pens, paper, and erasers near each other seams practical. (Bhambri 2021) In equity brokerage houses context, ARM can be used to detect behavioural patterns like the case presented earlier in this chapter: "a client who buys stock A and stock B at the same time has 90 % chance of selling stock C."

*Table 4: Applications of data mining in investment bank's Client Relationship Management. Following: Hassani et al. 2018*

| CRM Process | Key techniques | Purposes |
|---|---|---|
| Client profiling and knowledge | K-Means Clustering, Decision Trees | Efficient building of client profiles |
| Client segmentation | K-Means Clustering | Client segmentation, conduction of client-centric strategies |
| Client segmentation | K-Means Clustering | Management and enhancement of client satisfaction |
| Client customization and development | Decision Trees, K-Means Clustering, ARM | Direct and targeted marketing, product cross and up selling |
| Client acquisition and retention | Decision Trees, K-Means Clustering | Client churn prevention and prediction, attraction of potential clients, and strategic future service design |

Table 4 concludes data mining applications for investment bank's Client Relationship Management based on the article *Digitalisation and Big Data Mining in Banking* written by Hassani et al. (2018). However, this table only presents relevant parts of the original source scoping out other than client analytics applications of data mining and data mining methods not included in the three previously reviewed.

The insights gained from data mining are dependent on the functionality of the built model. There are various ways to measure the performance of machine learning algorithms. Classification algorithms' such as Decision Tree's performance can be evaluated by counting numbers of true positive predictions (TP), false positive predictions (FP), true negative predictions (TN), and false negative predictions (FN), and calculation evaluation

metrics such as precision and recall, F-score, and accuracy score. Results of predictions can be presented in confusion matrix, a two-by-two representation of actual classifications compared to predicted ones. To accurately measure the performance of classification data mining methods, it is important to divide the data into two sets, one for training the model and one for testing its performance. Accuracy score is the percentage of correct predictions out of all predictions. However, it might give misleading information in cases where the target variables are heavily skewed towards one or another. For example, if dataset with two classes contains 100 records from which 10 belong to class A and 90 to class b, and if classification model would randomly choose class A or B, its accuracy score on predicting class A would be 90 %. Randomly choosing one out of two options is not predicting, hence 90 % accuracy of the model is highly overstated. To overcome this problem, it is important to proportion true positive predictions to false positive, and true negatives to false negative predictions. Precision can be calculated by dividing number of true positives with the sum of false and true positives. It tells its user the percentage of correctly predicted true values out of all true predicted values. Recall, also known as sensitivity, measures the proportion of true positives predictions out of all actual positive records in the dataset. F-score combines precision and recall by interpreting their harmonic mean. F-score is normalized between 0 and 1, 1 being best and 0 being worst value. (Jain 2024; Rafatirad 2022; Scikit Learn b)

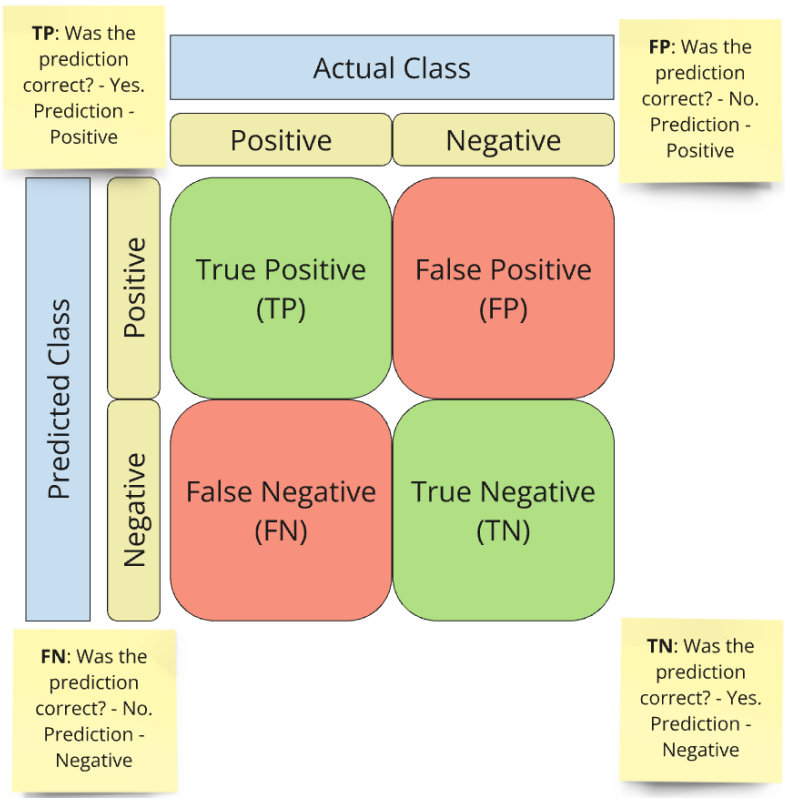



*Figure 8: Confusion Matrix. Following: Jain 2024*

Figure 8 represents a confusion matrix. True positives are values that were predicted to be positive, and they were actually positive. False positives are values that were predicted to be positive but were actually negative. False negatives are values that were predicted to be negative but were actually positive. True negatives are values that were predicted to be negative, and they were actually negative.

Whereas classification algorithms can be evaluated by comparing predictions to actual values, clustering algorithms have their own performance evaluation metrics. One method of evaluating clustering algorithms such as K-Means clustering is to perform Silhouette Analysis. It as a quantitative method that measures how well each point of data fits to its assigned cluster compared to its neighbouring clusters. Value 1 means that the observation is located in the centre of the cluster, while value 0 means that it's located between two clusters. Values below zero indicate that observation is located inside another cluster and is potentially misassigned. Performance of the whole cluster formation can be assessed by taking an average from all silhouette values calculated from the whole dataset. While 2-dimensional representations of clustering can often be assessed visually quite easily, Silhouette analysis is highly important when assessing clustering with higher number of

dimensions. Silhouette coefficient is calculated by taking a mean of the distances between cluster A's centre (cohesion) and its assigned members and the mean of the distances between cluster A's centre and cluster B's members (separation). Cohesion value is then subtracted from separation value and divided by the maximum of those two. Clusters A and B are clusters that are nearest to each other. This process is repeated for every sample in the dataset. (Belyadi & Haghighat 2021; Scikit Learn c)

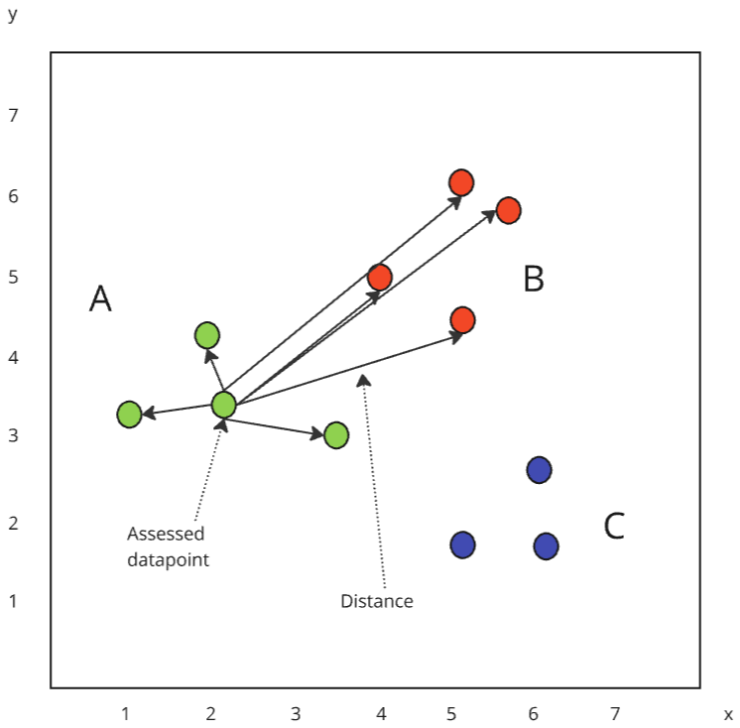



*Figure 9: Silhouette coefficient. Following: Belyadi & Haghighat 2021*

Figure 9 represents calculation of silhouette coefficient for assessed datapoint. Assessed datapoint's distance to every other datapoint in cluster A is calculated and taken average of. Assessed datapoint's distance to every datapoint in cluster B is calculated and taken average of. Finally, the average distance from assessed datapoint to datapoints in cluster B is subtracted from the assessed datapoint's distance to other datapoints in cluster A and divided by the larger of those two averages. The process is repeated for every observation in data set, and the two clusters used are always the ones nearest to each other. In this case clusters A and B are the closest. The silhouette coefficient of whole clustering is the mean of all independent observation's silhouette coefficients.

Performance of associate rule mining is evaluated using tree key metrics: Support, confidence, and lift. Support describes how frequently an item or set of items comes across

in the universe of transactions. In brokerage house's context it might mean that how often is on stock or set of different stocks bought together compared to all transactions. Confidence measures the likelihood of set of items being bought when one of the items in set is also bought. For example, it might measure what is the likely hood of client buying stock A and B simultaneously if stock B is bought. Lift measures the strength of association between two items. It is calculated by dividing support of itemset by the multiple of support of item x and item y. (Herath 2024; Soni et al. 2020)

$$Support = \frac{Frequency(X,Y)}{N} \quad (1)$$

where Frequency (X, Y) means number of observations where both X and Y appears simultaneously within N number of transactions.

$$Confidence = \frac{Frequancy\ (X,Y)}{Frequency(X)} \quad (2)$$

where Frequency (X, Y) means number of observations where both X and Y appears simultaneously and Frequency(X) number of X observations.

$$Lift = \frac{Support}{Support(X) \times Support(Y)} \quad (3)$$

where Support(X) means percentage of X observations within N transactions, Support(Y) percentage of Y observations within N transactions and Support how often X and Y are observed together in the set of N transactions.

Formulas 1, 2, 3 (Herath 2024) presents formulas for support, confidence, and lift. In this example, when calculating support, the occurrence of item X is measured when also Y is observed. However, as said in previous paragraph, measuring occurrence of one item among all transactions is also viable way of calculating support. In that case, support is calculated for the itemset.

# 4 Applications of client data analytics in Nordic Equity Sales & Trading

This chapter starts off the empiric part of this chapter. So far, this thesis has discussed about development of equity sales and trading and client data analytics in investment banking in global perspective. However, Nordic region comes with its special characteristics, and some of the findings discovered in global operating environment might not apply for investment banks operating in Nordic region. Also, different banks have different capabilities. This chapter aims to compile throughout picture on special characteristics of Nordic investment banking and to explore target organizations capabilities considering client data analytics implementation.

## 4.1 Research methodology and materials

This research's empiric part was conducted by utilizing research interviews and constructive research method. The research material was compiled by interviewing target organization's designated specialists with two separate research interviews. The material was used to gain understanding on the target organization's operating environment and its data analytics capabilities. Constructive research method was used to create suggestions to develop target organization's data analytics capabilities.

According to Hirsijärvi et al. (2017a) qualitative research is often conducted with interviews. In this research, interviews were held for experienced specialists and managers who had worked in international environment in sales & trading business. In total, interviewed had over 210 years of cumulative experience, and there were in total 13 specialists who participated. According to a book written by Hyvärinen et al. (2017), interviewing specialists can yield valuable insight as it is expected that they have significant experience and knowledge in the topic that they are being interviewed. To conduct a specialist interview, specialist must be defined. In these research interviews specialist is defined as someone who has had a well-established career in the capital markets industry and depending on the interview, has worked in international environment in Nordic banking sector or been a part of organization's IT development. Most the interviewed had had over two decades of

experience in the industry and all of them worked in managerial or/and senior roles. Interview questions were asked in live meeting, and the answers were unstructured. As also noted in the book, specialist interviews often require well defined questions that require some groundwork. Interviewer might also be required to have some knowledge on the topic as being able to ask appropriate return questions might help to gain more understanding. Compared to quantitative survey research where results can be statistically modelled and set metrics evaluated, qualitative interview research can even change during the interviewing period. New information might raise new questions that can be further introduced in next interviews. In the qualitative interview search answers are open and unstructured and going through them takes more time and effort. (Hyvärinen et al. 2017; Vehkalahti 2014)

The empiric framework of this thesis is based on the research interviews held at the target organization as well as on the other discussions had with target organisation's employees. In qualitative research interview empiric part is in continuous discussion with the theory so that the reader can easily understand the discoveries of the interviews. (Puusa & Juuti 2020) The interviews were conducted as open Microsoft Teams meetings or live meetings depending on the location of interviewed. Interviewed specialists answered questions that were provided to them prior to the meeting. Answers were documented, common conclusions aggregated, and results compared to findings from theory.

The first research interview asked ten questions on special characteristics of Nordic investment banking compared to global landscape. It aimed to recognize developmental factors that were in line with global directions of development and to explore characteristics especially considering client base, demanded services, and technological capabilities. The interview was held for eight experienced specialists and managers with established careers in equity markets industry. In total, participants had over 180 years of cumulative experience. On average, one interview lasted around 38 minutes meaning that in total over five hours were spent interviewing. Interviews were held between September 12. 2024 and October 11. 2024. Questions for the first research interview can be found from appendix 1.

The second research interview aimed to figure out the current state of target organization's data analytics capabilities. It tried to find out the currently on-going initiatives, future aspirations, challenges, employees' skill level, organization's culture towards data analytics, and the effects of regulation. Interview was held for experienced IT and equity markets specialists and managers likely to take part in the target organization's future client data

analytics projects. As noted in this thesis's theoretical part (Table 3), successful implementation of data analytics projects has not only technological requirements but also human recourses and organization cultural requirements. The second interview consists of ten questions that examines the target organization's data analytics capabilities from technological, human resources, and organization cultural perspectives, as well as currently on-going projects, goals, challenges, and competitors' capabilities. It is built while keeping the findings from literature in mind and tries to replicate the framework of Table 3; "Requirements and challenges in implementation of data analytics in investment bank". After concluding the research interview, its findings are reflected with the findings from the theoretical part of this research. In the second research interview, there were 5 participants with around 100 years of cumulative experience. Average length of an interview was around 36 minutes meaning that in total around 3 hours were spent interviewing. Interviews were held between October 23. 2024 and November 14. 2024. Questions for the first research interview can be found from appendix 2.

*Table 5: Summary of the research interviews*

| | First interview | Second interview |
|---|---|---|
| Goals | • To recognize differences between global and Nordic investment banking<br>• To understand the reasons for the differences | • To gain information from target organization's current data analytics capabilities, on-going initiatives, and future aspirations |
| Target participants | Experienced specialists and managers with established careers in equity markets industry | Experienced IT and equity markets specialists and managers likely to take part in the target organization's future client data analytics projects |
| Number of participants / sessions | 8 employees / 8 sessions | 5 employees / 5 sessions |
| Average length of session | 38 minutes | 36 minutes |
| Number of questions | 10 | 10 |
| Research window | September 12. 2024 - October 11. 2024 | October 23. 2024 - November 14. 2024 |

Constructive research method consists of six steps. 1. step is to recognize practically relevant problem. In this research, several deficiencies in target organization's data analytics capabilities were recognized. 2. step is to obtain comprehensive understanding of the problem. This is done by utilizing research interviews and literature review. First research interview has provided clear picture of the differences between global and Nordic investment banking environment and second one drawn a baseline for target organization's data analytics capabilities. 3. step is to design applicable solutions to the problem. In this thesis

solutions are offered for improving principles of data driven organization, accessibility to data, storing data, and data architecture. The top-line challenge that was recognized is that target organization's capabilities are not at the level to implement many of the globally recognized applications of client data analytics. As the main challenges consists of multiple factors, the factors that the author recognized to be the most influential were analysed independently by using steps of constructive research method. 4. step is to demonstrate the solution's feasibility. This can be done by having discussions about the offered solutions with involved employees, or by demonstrating offered solutions in practice. In this thesis solutions are first evaluated in the writing and then discussed with target organization's specialists. 5. step is to link the results to theory and demonstrate practical contribution. This is done by comparing recognized challenges to challenges came across in theory, discussing whether the requirements recognized from theory are realistic considering special characteristics of target organization, and by exploring which application of client analytics can be implemented in target organization compared to those presented in literature. 6. and the last step is to examine whether the results are generalizable. This means for example considering if the special characteristics of target organization also apply for other investment banks operating in Nordic region. (Lehtiranta, Junnonen & Pekuri 2015).

*Table 6: Steps of constructive research method*

| | Input | Output |
|---|---|---|
| **1. Problem recognition** | Research interviews and literature review | Several deficiencies in target organization's data analytics capabilities were recognized when compared to requirements in literature |
| **2. Gaining understanding on the problem and its effects** | Research interviews and literature review | Understanding on the reasons behind the problems and the effects of the problems were recognized |
| **3. Solution designing** | Understanding on the problem, findings from the theory | Four suggestions to improve recognized deficiencies were designed |
| **4. Demonstrating the solution's feasibility** | Discussions with target organization's management. Linking solutions to theory. Author's conclusions | Effects of the designed solutions for recognized deficiencies were evaluated and the feasibility of solutions discussed with management |
| **5. Linking results to theory** | Designed solutions, findings from the research interviews, literature review | Comparison between target organizastion's problems and literature. Assessment of applicability of the requirements based on literature |
| **6. Assessment of the results's generalizability** | Research interviews, literature review, designed solutions | Assessment whether other organization's capabilities could be analysed using same research framework and if the offered solution could benefit other similar organizations |

Table 6 summarizes the steps of constructive research method and its applications in this research. Input column reveals the main sources of information each step consumes. Output column presents results of each step. Horizontal arrow illustrates the direction table should be red.

To ensure trustworthiness of the research, research material is acquired by utilizing appropriate research methods and their best practices. According to Hirsijärvi et. al (2004),

when individual topic is researched enough in-depth, its significant factors and driving phenomena are also understood in general level. This means that the conclusions made from the research material are generalizable also when the target organization is other similar organization. In this research, multiple generalizable factors driving the findings of this research were recognized. According to Eskola & Suoranta (1998), when conducting research interviews to acquire research material, the interviewed should have similar scale of experience, firsthand information on the researched topic, and to be interested and favorable in participating to interviews. In this research, the research interview participants were all very co-operative and favorable in participating interviews. Many of them had also firsthand experience from working in other peer organizations of the target organization. There was no reason to believe that their answers to non-organization-specific questions would have been different if they were employees of other peer organizations. (Hirsjärvi, S. et al. 2004; Eskola & Suoranta 1998) Taking these factors into account, conducting similar research in any of the target organization's peer organizations by using previously introduced methods would be likely to produce somewhat similar results. However, there are organizational differences, and despite common driving factors, varying characteristics are likely to be highlighted in each individual organization.

## 4.2 Special characteristics of Nordic investment banking

As can be noticed from the previous chapters, most of the literature used in this thesis is written in context of biggest global banks. However, the operating landscape for global actors and Nordic banks are different in many ways. To be able to assess which of the globally recognized benefits, challenges, requirements, and applications of client data analytics are applicable to Nordic operating environment, it is necessary to explore the differences between those two. Also, by identifying local differences, resources budgeted for data analytics projects can be allocated better based on the insight. There isn't much research made on the special characteristics of Nordic investment banking. This section of thesis seeks to fulfil the gap in existing theory by conduction a qualitative research interview on the topic. The main points of the aggregated answers are bolded to highlight their importance.

In the first research interview (appendix 1), the operating environment of the biggest Nordic banks' equity brokerage houses were compared to biggest global banks' equity brokerage houses. Past development of Nordic equity brokerage as well as the future directions of development were also assessed. The answers for the first question,

*What directions of development has changed Nordic equity brokerage business the most during past decade?*

highlighted especially two things, MiFID 2 regulation, and increased electronic trading. Some other things mentioned were the increased curiosity of big global banks towards Nordic equity markets, reduced commissions, fragmentation of liquidity to multiple venues, and re-highlighted importance of sales traders. **From the MiFID 2 regulation, the unbundling of commissions and research fees were mentioned to have been especially influential. Electronic trading has emerged as an extremely cheap method of execution and had its influence on overall commission levels.** On the other hand, development of dark pools and multiple new kinds of electronic venues have made liquidity more fragmented and access to it more challenging, hence highlighting importance of sales traders. When compared to literature presented in the previous chapters, findings of this question go along with it without controversy. Increase in electronic trading and regulation were something that were also noted in global literature.

The next question,

*What are likely to be the biggest challenges and opportunities during the next decade?*

discusses about future of equity brokerage business. The answers to this question highlighted two things: **maintaining specialization in Nordic companies and liquidity providing and increasing passive investing and electronic execution.** As long as Nordic brokerage houses maintain their specialization in Nordic companies, can offer liquidity in Nordic shares, and remain close to their clients, they have competitive edge over global players. These competitive advantages can especially be notices in IPO markets (initial public offering). By listing a company for public, banks not only earn fees for listing, but also acquire important insight on the firms' owners, and form personal relationships with the management. However, preserving this edge will always be challenging considering the high level of competition and the shift towards more electronic trading. Increase in passive investing pose a threat for classic brokerage business as passive investments are often

executed electronically with low commissions, and positions held steady for long periods of time. Also, Nordic banks already struggle in the competition of executing large cap orders. High liquidity makes electronic execution possible, especially benefiting global peers. On the other hand, block trades that are too big to be executed electronically in those names are often done between global clients who might also prefer global brokers. Another challenge mentioned were staying technologically relevant to be able to compete with global actors. However, some opportunities might arise from the on-going debate on reducing some of the burdening regulation, and the significant wealth tied to private equity that might unfold at some point.

Nordic banks also face regulatory challenges that are less affecting to their global competitors. The answers for the question,

*What are the biggest regulatory challenges considering Nordic equity brokerage houses vs. global brokerage houses?*

highlighted few things: geographical differences in regulation, Nordic banks' spending on compliance, and the Nordic style of interpreting regulation. Geographical differences in regulations means usually acting according to the strictest regulation. Biggest global banks, especially in the US can focus more on the home market regulation. **Nordic banks must be specialists in both, home market regulation and different global regulations. Compliance is also somewhat scalable meaning that the same sized compliance team could monitor bigger operations as well reducing its relative cost.** Global banks were also said to be better at interpreting regulation for their benefit and at taking regulation related risks as **Nordic banks tend to be more "black and white" in regulatory interpretation.** However, even global banks cannot always avoid regulatory sanctions, and must pay fines for their risk taking. Hence to conclude, when compared to Nordic peers, global banks have a competitive advantage in compliance as they can focus more on home market regulation, are better at interpreting regulation for their benefit, and have relatively smaller compliance costs. Other point made out were that global banks benefit more on the Systematic Internaliser (SI) regulation as they have more order flow.

All the previously discussed answers were mostly in line with the globally recognized trends. However, the next question,

*Could you describe the level of technological adoption in Nordic equity brokerage houses compared to global ones?*

started to bring up more differences in local and global peers. According to most of the interviewed, Nordic banks were lagging, depending on the interviewed, more or less in the level of technological adoption. **Nordic banks' brokerage houses were said to be clearly lagging in in-house technologies and data analytics capabilities.** Significant proportion of interviewed even described locals' (Nordics') technological capabilities being at *elementary* level, or global peers' technological capabilities being *light years ahead*. The interviewed specialists who described the technological gap to be the biggest were also those who had most experience in working in different global enterprises. Especially these specialists underlined the utilization of data and level of automation lagging. On the other hand, some of the interviewed said Nordics to be only slower at adopting new technologies, but to be quite good at sourcing it. Only one of the interviewed didn't recognize there to be much gap. Interesting finding here is that the less recent experience interviewed specialist had in working in a company that could be included in the global peer group, the smaller they described the technological gap to be. The main reason for the gap can be found from massive differences in earnings. Developing new technologies and implementing demanding data analytics projects is relatively more expensive for much smaller Nordic institutions and often pose too big of an operation risk. As software products often scale well, cost of building does not grow along with the amount of end users hence giving larger global universal banks a competitive advantage.

The technological gap between Nordic universal banks' equity brokerage business and their global peers is the first major special characteristic found from this research. As practically all literature used in this thesis are written in context of global banks, many of the applications of data analytics recognized might be way ahead of Nordic peers' technological capabilities. The technological capabilities of the target organization are examined in the next research interview.

When discussing about the differences in competitive landscape, the answers for a question,

*What does the competitive landscape look like for Nordic equity brokerage houses versus global brokerage houses?*

points out few things: global competitors have always been interested in entering Nordic markets, but so far, locals have been able to compete with their specialization in Nordic equity markets and by having close relationships with their clients. **Compared to global peers, Nordic banks know Nordic companies better and are better at finding liquidity in Nordic shares.** However, global banks are better at macro-economic strategies, selling products originated outside of Nordics, and at offering cheap electronic executions services. Another interesting point made out was that **Nordic equity market is "over brokered", meaning that there are too many similar services offering brokers compared to the market's size.** This has led to a situation where level of **seniority in Nordic banks' brokerage houses is very high, and business is often relying on even decades long personal client relationships.** In contrast, big global banks have started to "juniorize" their staff meaning that recent graduates are overweighted in hiring, and often expensive senior employees have been let go in the hopes of cost savings. This has been possible as level of automation has increased and interactions between client and broker has become more systematic. In conclusion, despite Nordic banks being able to compete against global peers in the Nordic equity markets, it would not be possible for Nordic banks to compete against them in global markets. Biggest global universal banks face competition only from each other.

Answers for the question,

*How is equity sales job different in Nordic environment compared to global environment?*

highlighted few things and goes partly along with the previous question's answers. Compared to locals, **global sales are often strongly divided into either specialists or generalist.** Specialist sales strongly specialized in certain sectors or products whereas generalist sales focus more on client relationship management and connecting clients with right specialists. Locally, sales often have more hybrid roles acting as a specialists and client relationship managers simultaneously. Globally sales are also expected to bring more business as the number of clients is practically limitless and the typical client size bigger. This also resonates with the previous question's findings where Nordic equity market is said to be over brokered. **Locally, the number of clients is much more limited**, and the earnings brought in by single sales capped. Locally brokers are often competing from the same large institutional clients hence emphasizing the level of seniority. **Globally there are always**

**new clients to be acquired**. Another point brought out was that globally there are also more sales focusing on selling electronic trading solutions.

This finding is also important in the context of this thesis. **As the number of clients increase, it becomes increasingly hard to recognize behavioural patterns in their action without any data analytics.** On the other hand, **the more there are clients, the better their behavioural patterns can be statistically modelled with data analytics. As the number of clients in Nordic banks' equity brokerage businesses is much smaller than in global peers, some applications of data analytics presented in the literature might not be possible to implement.** Also, as many of the client relationships are based on the long and personal relationship with the client, and the overall number of clients is relatively small, profound statistical modelling of client's actions might not be as beneficial as it is for global peers.

When discussed about the differences in typical research analyst's role, especially one point was highlighted. According to the question's,

*How is equity research analyst's job different in Nordic environment compared to global landscape?*

answers, **the biggest difference in local and global research analyst's job is the sector splits in companies.** Globally research analysts are more specialized in sectors whereas local analysts are looking at companies more across the peer groups. Globally, often only the management teams of the largest companies are personally known whereas it is common for local analysts to directly communicate with even small cap companies' management. Also, as there are less companies in total, the overall coverage of companies is higher, and the local companies are known better.

Globally, there are bigger difference between different kind of clients. Question,

*Is there a difference in clients' equity strategies in the Nordics compared to global landscape?*

discussed about the differences between clients. The answer that kept repeating was that **globally there is much more variety in clients' equity trading strategies.** Whereas local clients are often "long-only" pension funds, globally the universe is much broader. For

example, globally hedge funds are very important type of clients for brokerage businesses, but among Nordic clients, there are practically no hedge funds.

In the context of this thesis this is also important finding as hedge funds and other very actively trading clients also produce lot of data with their trading. **The more there are trades to analyse, the more accurate models can be built.** Without enough trade data to analyse, it becomes impossible to recognize reliable patterns in clients' trading preferences.

The last question of the research interview asked interviewed,

*Is there some another question that we should include into this interview?*

Most of the answers to this question were different in their nature. Some of the interviewed were highlighting the **cultural differences in global and local investment banks**, and some data and technology. One specialist who had worked in multiple global enterprises said that the organization culture in Nordic investment banks is a lot softer and less international. Globally employees get fired all the time and employee turnover is kept high. International investment banks also employee people with multiple different nationalities and backgrounds. **One of the interviewed pondered whether data could be used to leverage block trading.** Another interesting topic discussed about was the transparency in trading. One specialist said that Nordic clients typically appreciate transparency in their trading whereas globally transparency is seen more as a threat.

The findings of this research interview affect the way in which client data analytics should be implemented in Nordic investment banking environment. Past development in areas such as electronic trading and regulation as well as the future directions of development in electronic trading emphasizes the importance of automated data analytics. As increasingly large share of client orders is executed automatically, the collection of information from client's actions is increasingly relying on automated data analytics processes. Development in regulation has made collecting and storing client's trading data mandatory process for every brokerage house under MiFID 2 regulation. Data analytics could offer ways to monetize the data that is anyway going to be stored. In addition to the challenges caused by increase in passive investing and electronic trading also maintaining high level of specialization in Nordic equity markets is determinant in the competition against large global banks. To support high level of specialization, data analytics could be used to enhance block trading by analysing past client orders and indications of interest. As discovered in the

research interview, compliance poses greater relative costs for globally operating Nordic banking institutions compared to global peers. In this context, data analytics and automated data flows could be used to automate reporting, and for monitoring purposes, with a goal of increasing efficiency. Despite Nordic institutions having limited number of clients, data analytics could still be used to support client relationships management. The benefits of quantitatively analysing smaller groups of clients suppresses the expected benefits of data analytics (compared to larger client base of global peers) but is still likely to deepen the understanding of clients' behaviour. Also, as the Nordic clients are already known better by Nordic banks, the findings from data analytics are likely to be understood better by Nordic banks, further increasing the specialization in Nordic clients. The differences in clients' trading strategies are also something to consider when assessing the benefits of client data analytics. As the local clients tend to be more passive compared to global hedge funds and other actively trading client institutions, there might be less available trading data to analyse. The statistical methods that rely strongly on vast amounts of data might be impossible to implement.

However, as the Nordic investment banking has been heavily relying on high level of specialisations and long personal client relationships, and as large-scale data analytics projects are very expensive compared to the size of Nordic investment banks' earnings, the data analytics capabilities of local institutions are not at the level of their global peers'. To be able to realistically benefit from data analytics implementation with limited budget and low baselevel capabilities, resources should be allocated to support specific tasks that increase specialization in Nordic equity markets, and efficiency of compliance. The solutions should also be built in a way that is expandable in the future.

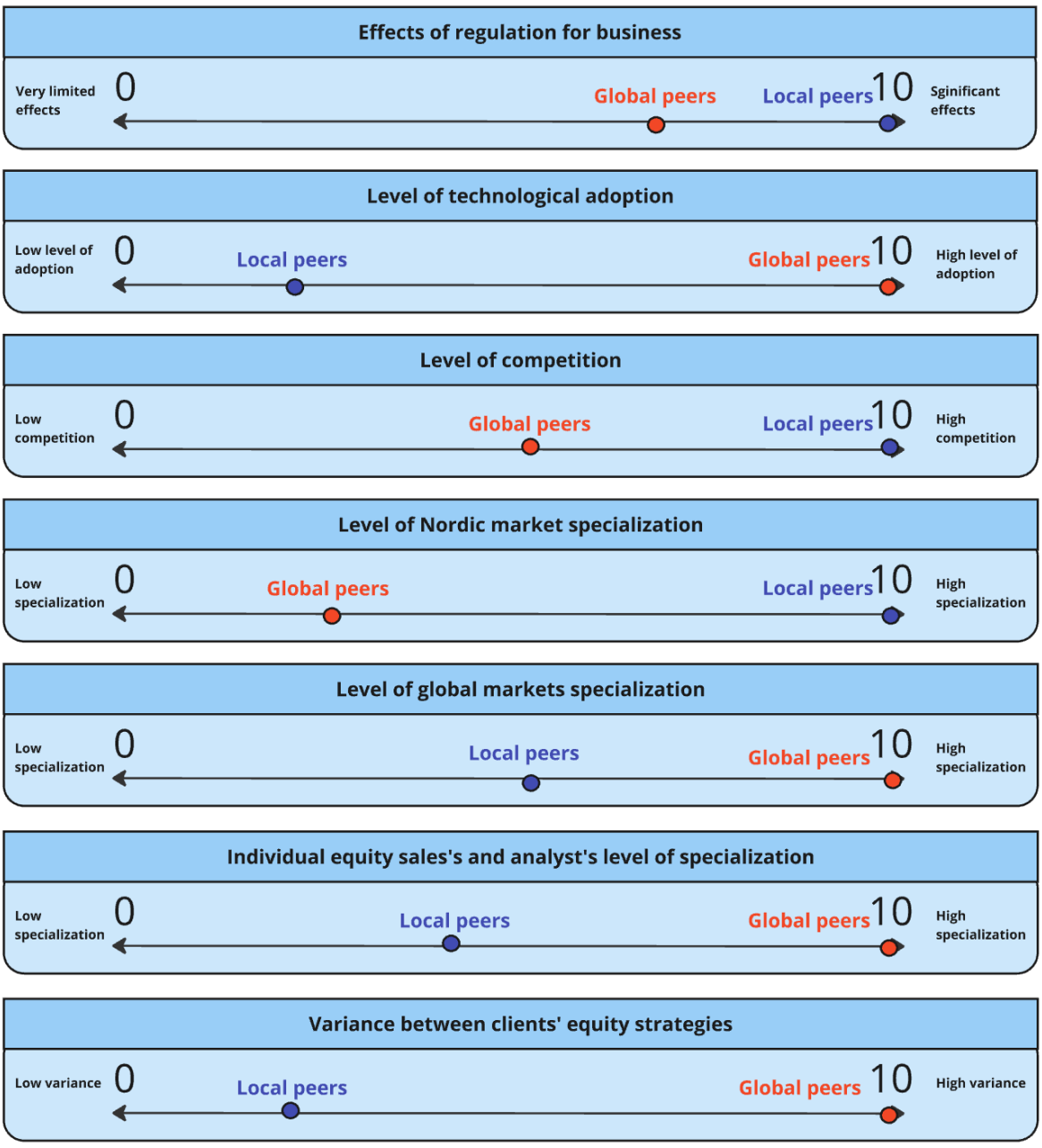



*Figure 10: Differences in Nordic and global peers' operating environment.*

Figure 10 concludes the findings from the first research interview and reflects authors views on them. The figure has been formatted so that either global or local peer group is given the highest value based on the question, and then the other peer group is compared to that.

## 4.3 Target organization's capabilities and goals in data analytics

This chapter examines the technological capabilities of the target organization. It aims to figure out the biggest challenges in data analytics development and the current state of data

analytics capabilities. Each organization has their own technological capabilities and aspirations. If the level of technological capabilities is low, implementing large and complex projects becomes increasingly hard.

The research was conducted the same as way earlier. Five specialists with experience in developing target organization's technological capabilities were interviewed in an open research interview. Some of the interviewed were the end users of the data, making analyses, some focusing on IT development, and some more of project managers. The questions can be found from appendix 2. Main points of each aggregated answers are bolded.

The research interview begun with the question:

*What data analytics processes are currently going on? How are data analytics currently being utilized in target organization's equity brokerage business?*

The answers concluded that **there are multiple "ad hoc" analytics initiatives going on, from which most are done independently from each other.** Some of the on-going projects were analytics in algorithm performance, smart order router quality assurance, market microstructure analytics (analytics for different venues target organization has access to), buyback program analytics (especially to ensure that the algorithms used in those programs executed trades in compliant way), internal monitoring for other quality assurance initiatives and compliance, transaction cost analysis (TCA), and some elementary client analytics. Many of these projects were done independently and required first acquiring the data (which often proved to be complex task only done by few), then processing it, and finally visualizing it. Most of the tasks were directed from the headquarters and the end products were rarely deployed to other offices. According to the interviewed, optimally the end products of these data analytics initiatives should be available for as wide deployment and possible as higher number of end users could spark more ideas of development and benefit from the tools themselves. Only the initial deployment of the project and its management should be done centralized. Another conclusion made was that **most of the systematic analytics initiatives focused to either regulatory purposes or to reporting that was held as industry standard.** Many of these systematic approaches to data related tasks, especially those with regulatory purposes, included only collecting and storing the data. The target organization was also looking into using data platform solution provided by external vendor. However,

this project was advancing slowly due to same commonly recognized challenges brought up later in this interview.

The second question,

*Could you describe target organization's current data analytics infrastructure? What tools and technologies are currently being used in target organization's data analytics processes? Where does the data come from?*

discussed about the technological aspects of collecting, storing and utilizing the data. The conclusion was that **there is not an overall data architecture in target organization.** Its answers also highlighted the **lack of centralized data platform and efficient access to data**. Data was described to be fragmented to different source systems and exists in various structures and formats. Some available data analytics tools were experienced to be too difficult to use for employees in non-tech roles. Only either regulatory-wise mandatory data, or data required by industry standard reports were systematically collected and stored. The process involved mechanically reading a daily trading system produced file, uploading it to SQL database, and monitoring the process to catch errors and log actions. Despite existing, the SQL server was not regularly queried due to its complex, non-user-friendly format. The times, couple of times a month, it is queried, its extracts are often used for regulatory purposes. There are also many recognized data sources systems whose data is not currently utilized in any way. One example of such is CRM.

Data analytics were recognized to have benefits in internal as well as external use. The question,

*What would the target organization want to be able to accomplish? How would data analytics benefit target organization the most?*

discussed about desirable use cases of data analytics. As mentioned before, most of the on-going data analytics initiatives are made for internal use. Some interviewed recognized that data analyses produced for external use could be used as marketing material. However, these pre-trade analytics were estimated to be very hard to implement. **Another highly emphasized use case of data analytics were the different applications of client analytics.** Data analytics were recognized to be very beneficial if they could be used to quickly and efficiently obtaining a view of the business, and to make work of sales traders more efficient. Sources of business and methods of its creation could be analysed as well as the clients'

behaviour, and possible seasonality of actions. Client analytics could also be used to support client relation management by analysing clients' satisfactory towards the service they are receiving and by analysing their interactions with individual sales traders. **The primary end goal of these analytics would be to increase the likelihood of finding trade counterparties, i.e. to enhance block trading.** To accomplish these goals, data should be collected from clients' past behaviour including information on the traded stocks, market levels, trade directions, recurring interests, possible seasonal patterns in trading activities, stock holdings data, and research consumption in various stocks. Some recognized **practical use cases of these data analytics would be efficient, accurate, and timely targeting of flow to all relevant clients, smaller dependency on individual sales traders remembering client interests, and a chance to work with larger house orders when the probability of finding counterparties is increased.** According to one interviewed, this kind of tool has been found to be beneficial in some another asset class's trading desks. Other desired outcomes would be to make accessing data easier for larger variety of stakeholders, creating automated pipelines for collecting and storing wide range of data from internal and external sources, and to create easy-to-use analytics tools that would also benefit employees without technical skills across the organization. Also, following common, pre-defined protocols when initiating new data analytics projects were said to improve the lacking data architecture.

The answers for the question,

*What are target organization's competitors doing? How does target organization compare to them?*

**describes the target organisation to be mostly in-line with its Nordic peers.** However, as already discovered in the previous research interview, target organization tends to be lacking in analytical capabilities when compared to global competitors. Especially data analytics produced for external use are practically non-existent in target organization whereas global competitors skilfully create data analyses from various topics to lure new clients and to serve the existing ones. Some global competitors are even fully dependent on the data with their block trading, flow advertising and market making.

The answers for the question,

*What are the biggest challenges in target organization's data analytics initiatives?*

also highlights the same findings already discovered in the previous research interview. **Management prioritizations and employee skills were said to be the biggest challenges**. Historically, management's mindset towards data analytics and investments in new technology was said to have been the most restrictive factor. **However, during past few years, this area had faced significant improvements.** Other challenges noticed were the ability to commit to data analytics projects for a certain period of time at the time without constant hesitation, and the lack of work time prioritization to data analytics projects with the employees with sufficient skills. According to one interviewed, best option would be if new projects were given predetermined time and budget, and there would be a solid commitment to those decisions. After the predetermined time has come to an end, the results of the project are examined and continuation decided. Bad projects should be killed fast and successful ones developed further.

The next question,

*Could you describe the level of data analytics skills target organization's employees have?*

focused on the technological capabilities of target organization human resources. When interviewed described the technological skillset their colleagues had, few things were pointed out. **Most employees tended to lack technological skills, or their work time was not prioritized for technology related tasks.** Also, employees without strictly technological roles were said to be **quite restricted by the internal guidelines when it came to developing new tools.** High level of seniority were argued to be one affecting factor for the average technological skill level of employees. However, in addition to the employees working in equity brokerage business, the owner bank employs great number of employees working at various technology related jobs. Common digital working channels enable getting help and support across the whole organisation.

In the theoretical part of this thesis, organization's cultural attitude towards data analytics were said to be one factor affecting its data analytics capabilities. A little surprisingly, the question,

*Could you describe the target organization's culture towards data analytics? Are employees open to new innovations, and to possible changes in their daily routines? Do target organization's employees think data analytics could improve its business?*

answers concluded that **most employees saw data analytics and technological development more as an opportunity than threat.** Especially, the attitude of management had gone through significant improvements as some new management had been hired and old let go. However, **before large investments are made, more evidence from positive returns of data analytics are expected.** Also, someone has to take the ownership of the processes and be the driving force to initiate new projects and keep the old ones running. The new tools and processes are expected to be user-friendly so that employees without too many technical skills are also able to benefit from them. Another factor affecting target organization's cultural attitude towards data analytics was said to be the **lack of understanding in the data's potential. Only very few employees actually had experience in working with data.** Despite the overall positive attitude towards data analytics, the lack of experience in working with data and understanding in its realistic potential might lead to situation where some employees feel themselves either threatened or are overly optimistic towards the results.

The question,

*How does regulation affect target organization's data analytics initiatives?*

explored the ways in which regulation affects target organizations data analytics initiatives. The importance of handling client related data in compliant way was highlighted. Data containing client identifying information were interpreted to be allowed to use only if there was a business reason for it. It was also discovered that regulatory reasons for data analytics made getting funding for those projects easier. However, interesting finding was that **none of the five interviewed knew exactly what regulations affected the target organization's data analytics initiatives and how.**

The last question,

*Is there some other question we should consider?*

did not spark too much conversation. However, one interviewed wondered **how much target organization should try to develop in-house, and how much should target organization source from external vendors.** It was admitted that target organization's capabilities as current, and likely in the future as well promotes more the use of external vendors.

To conclude the findings from the interviews, there were multiple ad hoc analytics processes going on and only few systematic "large-scale" initiatives. Ad hoc analytics focused on various quality assurance, post-trade, and market microstructure analytics, whereas the more systematic approaches focused either on serving regulatory purposes or to fulfil industry standard reporting demands. Initiatives with regulatory purposes often only included collecting and storing the data. The target organization's data infrastructure proved to be non-existing. Data was still mostly extracted manually by few IT employees, pipelines between data source systems and business intelligence tools were rarely built, and only few of the recognized data sources were utilized. Various different formats of data in source systems were also recognized as a challenge. The potential benefits of data analytics were recognized for multiple different use cases. Especially beneficial data analytics were estimated to be in enchaining block trading. By analysing data generated by clients' actions in trading, research consumption, indications of interests, and positioning with different stocks, the possibility of finding trade counterparties might be increased. However, the built tools were expected to be easy-to-use, efficient, and accurate. Compared to the local competitors, target organization was said to be mostly in line. In some areas better, in some worse. Compared to global competitors, target organization, as well as other Nordic peers, were said to be lagging. Especially the capabilities to create pre-trade data analytics that could be sent to clients were described to be very limited compared to global competitors. When discussed about the challenges in developing data analytics capabilities, the lack of resourcing and management prioritization, as well as limited skills of current employees were highlighted. The target organization's cultural attitude towards data analytics were described to be overall positive, but the lack of experience in working with data made it hard for regular employees and management to realistically judge the impact and benefits of data analytics. Regulation was described to impact the way in which client data analytics could be implemented, but the actual effects of regulation was not discovered. It also seemed that there was not an internal consensus on how the regulation related to client data analytics should be interpreted in practice, and partly due to this, this type of projects were not implemented in larger scale. However, many of the interviewed did not see how regulation would stop the target organization in using its own internal client data, but compliance was still often said to be against it. Also, as there wasn't clear alignments in policies regarding the use of internal client data, employees were not encouraged to participate in development

spontaneously as there was a risk that their actions would later be interpreted to be against the company policy or even regulation.

*Table 7: Target organization's data analytics capabilities.*

| Technological capabilities | Human resources capabilities | Organization cultural capabilities |
|---|---|---|
| • Trading system can output data in simple file format<br>• There are different systems that holds data such as trading system, CRM, stock ownership portal, chats and emails, and other market data providing external vendors<br>• SQL server is being used to store some trading system data<br>• Software provided by external vendor is used for post trade analytics (TCA) | • There are some employees, especially younger ones with more technological skills<br>• The owner bank employs great number of employees working at various tech related jobs. Digital communication channels can be used to reach employees across the whole organization | • Management's attitude towards data analytics has improved and there is a motivation to improve that area<br>• Employees are mostly open to new ideas and ready to change their working habits if it improves the business |
| **Challenges associated with:**<br>• Data is fragmented to various source systems and is in different formats<br>• Accessing data if very hard, almost always requiring help of IT employees<br>• Getting data is often a manual task as there are often no pipelines between source systems and analytics software<br>• Only regulatory-wise mandatory, and industry standard reporting data is collected and stored automatically. If there are some other automated analytics projects, they are ad hoc<br>• Some source systems lacks APIs | **Challenges associated with:**<br>• Overall, the target organization lacks in employees with technological skills. This is partly explained with high level of seniority and historically different prioritization in recruitment<br>• There are also some employees with good technological skills, but their work time is prioritized more to non-tech related tasks<br>• The usage of skills were also restricted by internal guidelines that limits the development capabilities of employees working at equity desk<br>• Despite having lot of people working at various tech-related jobs, the challenges is to know who to contact | **Challenges associated with:**<br>• Currently used data analytics tools are not very user-friendly for employees without tech-oriented roles<br>• To keep employees optimistic attitude towards data analytics, new tools has to be easy-to-use, efficient, and accurate<br>• Even while management is motivated to improve data analytics capabilities, more signs of positive returns is required before bigger funding<br>• Employees' very limited experience in working with data makes it hard for them to judge the impact and benefits of data analytics<br>• This might make some employees either overly optimistic on the results, or feeling threatened |

The main conclusion from the second research interview is that the target organization's data analytics capabilities are at elementary level, but not non-existing. The level sets significant limitations to implementing various applications of client data analytics presented at the thesis's theoretical part. Figure 4, *Applications of client analytics in investment banking,* presented the optimal sources of client data and the various applications to which the data analytics could be used. The figure 4 suggested that data should be collected from client's behavioural actions, sentiment, interactions with bank, and other external data sources. Behavioural data included client's past trading activities with bank and equity research consumption. Sentiment data included data mined data from chats, emails, blogs, reviews, and directs feedback, and its intention was to analyse how satisfied clients were with the received services. Interactional data included numbers and types of interactions between the client and bank, and other external data information on client's location, used devices, working hours and habits, availability, etc. In the literature, client data was used to make profiling on clients' risk preferences and expected returns, to build product recommendation systems, and to run client satisfaction analytics. The client profiling consumed solely behavioural data, product recommendation systems behavioural, sentimental, and external data, and satisfaction analytics optimally all four types of data. As can be noticed from research interview's findings, the target organization can systematically only acquire client related data from clients' past trading activities. This means that the target organization's capabilities in running throughout client data analytics are very limited and are affecting the diversity of different data analytics systems that could be built, as well as the accuracy of those. As noticed in the first research interview, the number of clients in Nordic equity brokerage businesses are very limited compared to the number of clients their global peers' have. Many global peers have also run their data analytics for years already and have stored data from longer periods of time. This affects the type of data analytics that could be run as many statistical methods require enough sample size to be accurate. However, the target organization had recognized that there are also other data sources that could be utilized in the future. If those recognized data sources can be exploited in the future, target organization's capabilities in implementing complex client analytics becomes more aligned with requirements presented at Figure 4.

Another significant difference to theory was the lack of organized data architecture. The model presented in the theory included data lake to store all possibly needed data in its raw format, ETL process to turn business-wise relevant data into structured format, data mining

to find patterns from the structured data, data marts to distribute data according to regulation, and visualization software to help making insights from data more efficient. The second research interview found out that target organization does not have a data architecture. Data that was not used stayed at the source systems and accessing it was at best very difficult. There were even some recognized data source systems which capabilities in systematically outputting the stored data was not explored. In some rare situations where data was automatically processed, it was stored in a structured format since the beginning. Data mining was not known to be used in any of the data analytics initiatives. The access to data was controlled by access rights that had to be individually signed to each employee applying to those. Some data visualization tools had been historically used, but none of them were known to be currently in active use. This thesis suggests an alternative data architecture to better answer target organization's characteristics. This model is presented in the next chapter and can also be found from appendices (appendix 3).

## 4.4 Development of target organization's client data analytics capabilities

The final chapter of this thesis's empirical part focuses on offering solutions in developing target organization's data analytics capabilities, especially in area of client data analytics. It utilizes constructive research method as it provides clear framework to offer solutions for practical issues.

When comparing target organization's capabilities, applications and benefits of data analytics to those recognized in literature, multiple differences were noticed. In the literature, organizations that had well established data analytics capabilities were able to collect data from multiple different source systems, process and visualize it, and turn it into valuable insights. All this was done systematically and automatically due to their efficient and well-designed data architectures. Their capabilities were results of long and persistent investments in technological development, employee skillsets, and organizational culture. On the other hand, target organizations capabilities were far from this level. Usable data source systems were very limited, there were deficiencies in employees' technological skills, and data architecture was non-existent. When offering target organization feasible solutions to develop its data analytics capabilities, it is important to keep this large gap in mind. It would not be realistic to directly target these deficiencies, as they are results from much greater

issues. Even though limited data source systems were recognized as a deficiency, only increasing usable data would not help much if there were not automated processes to process the data. Hiring new data analytics professionals alone wouldn't do much if there was not data to utilize, or compliance provided guidelines to point out clear regulatory limits. Also, without first demonstrating positive operational returns from data analytics, management wouldn't likely to be eager to provide funding for these initiatives. Instead, it would be more effective to start laying the groundwork for better environment in which these capabilities could spontaneously start developing. Hence, four suggestions to start building target organization's data analytics foundation are provided.

*Table 8: Suggestions to develop organization's data analytics capabilities*

| Prioritization | Suggestions to improve organization's data analytics capabilities |
|---|---|
| | 1.) create clear alignments in compliance set policies regarding interpretation of data analytics related regulation and make it available for all employees. |
| | 2.) access to data should be made easier and data silos be eliminated. |
| | 3.) simple prototypes of potentially beneficial tools should be built to demonstrate their potential for management to get more funding. |
| | 4.) general and practical discussion about data analytics and their potential should be included into organization level briefings and mandatory and optional education offered for employees. |

1.) compliance should make clear alignment in policies regarding their interpretation of data analytics related regulations and make it available for all employees. 2.) access to data should be made easier and some data silos be eliminated. 3.) Simple prototypes of potentially beneficial tools should be built to demonstrate their potential for management to get more funding. 4.) general and practical discussion about data analytics and their potential should be included into organization level briefings and mandatory and optional education offered for employees.

The recognized challenge that the author estimates to be the most influential is the lack of alignment in client data analytics related policies by compliance. None of the interviewed

specialists knew the regulation or its internal interpretation considering client data analytics. Specialists even had dissenting views on whether there are any compliance issues on processing client produced trading data. The lack of alignment in policies prevented employees from developing target organisation's data analytics capabilities. For example, employees working at different desks faced challenges in building prototypes from potentially beneficial tools as some were restricted by compliance, while others lacked clear guidelines on the permissible use of data and the specific circumstances under which it could be used.

The offered solution for this problem is to allocate time for compliance to interpret the current regulation and to draw clear alignments in policy. These policies should be openly available for every employee and the internal interpretations of them included in organization's mandatory training. Employees interested in developing target organization's data analytics capabilities should understand what actions are permitted and in which circumstances. Also, if something is forbidden, it is important to include the reasoning to help employees gain view on the whole entity. This would lay the groundwork for organization in which employees with sufficient skills and interest could build their own prototypes of tools and explore the available data that might be relevant for their desk's operations. As client analytics are already in daily use in another asset class's trading desks, target organization's compliance could ask for their help in interpreting similar regulation.

The feasibility of this suggestion was also evaluated with discussions had with some of the specialists interviewed in the previous research interview. The challenges related to creating aligned policies were recognized and were said to be partly due to large size and complex nature of the project. According to one interviewed, large and complex projects often tend to lure lot of participants from various backgrounds and with varying motives. If there are unnecessarily many participants involved, decision making tends to get harder as there are many dissenting views. To tackle this, designated project team should be named to be in charge of the decision making and project management. It was also noted that compliance also has its own costs and spending time on creating these alignments requires prioritization.

As pointed out in the literature, regulation has been having increasingly large impact on the whole banking industry. Findings from the first research interview included that Nordic banks have historically been more "black and white" in interpreting regulation. Nordic banks also have to interpret regulation across their whole geographical client base as revenues are

often generated globally. These factors definitely pose challenges in interpreting regulation around data analytics. These challenges in interpreting data analytics related regulation are also likely to be common among other Nordic banks' equity brokerage businesses as their general capabilities in data analytics have been recognized to be lacking compared to global peers. However, it is likely that once first movers starts to implement data analytics based by certain interpretation of regulation, other will follow.

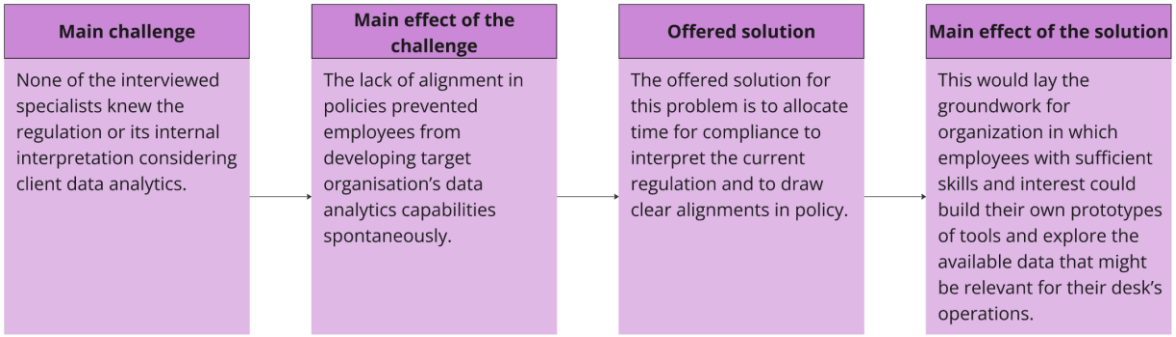



*Figure 11: Effects of suggestion 1*

Figure 11 summarises suggestion 1's effects on the recognized problem. The arrows between tables illustrates the workflow of utilizing constructive research method where the problem is first recognized, information on its effects gained, solution to the problem designed, and finally the solution's feasibility assessed. Figures 13, 14, and 15 follows the structure similar to Figure 11.

The second most important area of development is the accessibility to data and the elimination of data silos. This should be taken into focus after most of the compliance issues are solved. Currently, target organisation's data is often stored in data source systems and not utilized in any way. Data that is actively stored and processed is very hard to access for employees not in IT roles. One of the major reasons why data is still often not utilized is the lack of target organization's skills in transforming data into correctly structured format, as well as the source systems capabilities in automatically extracting data. Another reason why access to data was described to be hard was the target organization's non-user-friendly database. Only employees specialized to pulling data out of it were able to do it. By making the access to data so hard, employees were not encouraged to explore it, and many skills that could have been developed that way were left undeveloped. Many employees who were not directly working with target organization's database were not even aware of the different data that was stored. Despite lacking technological skills to process the data, many of the

senior employees working at target organization's equity desk could gain significant insight from the data once it is visualized. Now many of those insights were left unnoticed.

The offered solution to this problem consists of two parts which can be implemented simultaneously, without prioritizing other.

1.) making access to structured data easier by documentation, and if needed, by restructuring the database to more user-friendly structure,

2.) eliminating data silos by creating centralized data lake for data in its raw format.

Access to target organization's database should be made easier by providing documentation and education. Different users interested to explore the data should be assigned access rights to database based on their role, but still providing as wide accessibility as regulatory-wise possible. By granting access to data and providing necessary support to acquire it, users in different roles would start to gain experience in working with data and become familiar with the basic concepts. After processes around granting data access rights are defined, each team could name one data responsible person responsible for the team's data analytics development. Education sessions for executing simple SQL commands to pull the data from database could be held for the interested, and compliance defined rules for storing and processing data taught. Also, as found out, target organization has a lot of data that is stored in data source systems, and due to various formats and inconsistent accessibility, is not utilized in any way. Despite not being able to be put into structured format that is aligned with target organization's database's format, this data could still benefit some users in the organization. One way to solve this problem would be to use data lake to store all data in its raw format. From the data lake data that is proved to be business-wise relevant could be transformed into unified format and loaded into common database with ETL-process. Data lake should be also documented well so that users can find the datasets they are looking for. Accessibility to data in data lake by various user groups should also be regulated to assure regulation is followed.

By eliminating data silos and making access to data easier, interested employees could start developing their data analytics skills with lower threshold. This would also make it possible for different users to determine what data is relevant for their business and to request ETL process for the particular dataset. Also, storing data from various sources (even from those that cannot be currently processed) creates an option to benefit from the data some point in

the future when the target organization's data analytics capabilities have improved. Centralized approach to storing data also enables capable employees to spontaneously explore, ideate, and develop the target organization capabilities further and develop its data driven organization culture as working with data becomes a part of regular tasks for increasing number of employees. Also, as employees develop their personal data related skills, data analytics starts to yield profits without much additional funding. Optimally, employees interested in running analyses have the ability and skills to pull and process data, and employees with skills to understand the data can make valuable observations from it. As mentioned in the second research interview, there are already some employees with great technological skills and capabilities to work with data. As the access to data is made easier and all of it are stored centrally, these employees could start working with it much more efficiently and start building protypes. However, it is important to remain attentive when granting employees access to data to ensure data related regulations are followed.

The feasibility of this suggestion was also evaluated with discussions had with some of the specialists interviewed in the previous research interviews. Their stand on the solution was that it seems feasible and something to look into. However, it was wondered why solution similar to this were not already implemented before and if there are some undiscovered challenges that makes implementing it impossible. One interviewed wondered if there are enough usable data source systems to make implementing data lake worthwhile. It was also noted that when pulling data from source systems, the access rights to that data might change.

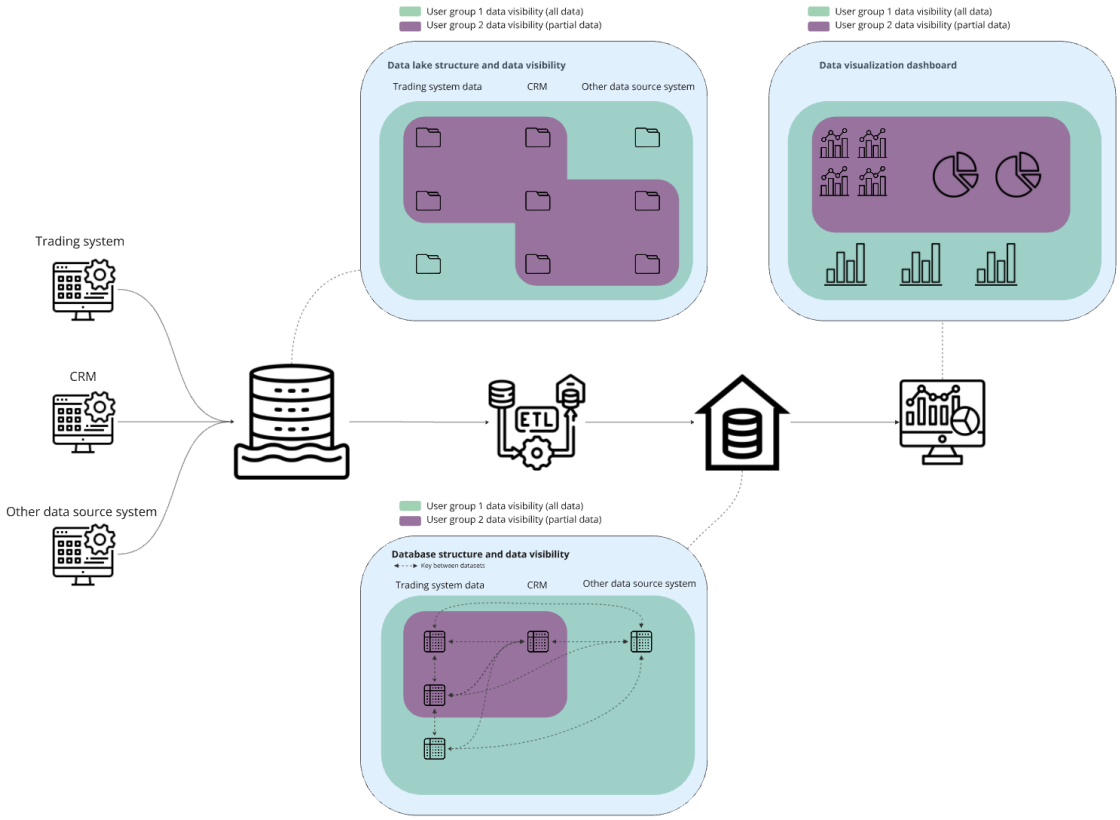



*Figure 12: Suggestion for target organization's data architecture (also as appendix 3)*

The challenge of siloing data was recognized as a challenge in the literature as well. Especially in the banking sector which has gone through significant consolidation, data was often described to be fragmented into various parts of the organizations. Hence, target organization's challenges in this area are commonly shared among competitors. However, it is likely, that the local peers face challenges in this area now that many global competitors have already solved years ago. In the theoretical part of this thesis the conclusion was that some of these challenges could be solved by creating a data architecture presented in Figure 5. It was also concluded that organization seeking to build its data analytics capabilities should hire a complete data analytics team with developers, data engineers, data scientists, data architects, and data analysts. Considering target organization's current data analytics capabilities (Table 7), amount of utilizable data, and management's willingness to provide funding, the solutions concluded in literature are probably unrealistic. Instead, the solution this thesis offers, is that the target organization should offer best possible setting for these capabilities to start developing spontaneously. Access to data is made easier and employees are encouraged to explore the data and educate themselves. Instead of regulating access to data with data marts (as presented in Figure 5), this thesis suggests that it is done with access

rights that are initially defined for data lake. Data visualization is done in common dashboard, but users can only see the charts to which data they have access to. Figure 12 illustrates the data architecture that the target organization should aim for. Green and purple colours indicate user groups with different accesses to data. In the cross-section of data warehouse, dashed lines illustrates the connections between different table. Folder icons in data lake's cross-section illustrate the locations different types of data is stored in. In a data warehouse, table icons are used to describe the similar situation as data is stored in structured format. However, the idea of these cross-sections is not to model the actual structure of data lake, data warehouse, or the visualization dashboard but to highlight the accessibility of data for different user groups, to highlight that not all data is transferred from data lake to data warehouse, and to highlight the common dashboard structure which visibility is conducted from the initial data access rights.

As each organization is different, it is hard to assess whether data architecture exactly like this could be applied in different organizations. However, accessibility to data is one of the key principles in data driven organization, and any other organization trying to develop its data analytics capabilities would benefit from it. Local peers are also likely to face similar challenges in their current capabilities, employee skills, and management prioritization, hence emphasizing the generalization of these suggestions.

| Main challenges | Main effects of the challenges | Offered solutions | Main effect of the solution |
|---|---|---|---|
| Currently, target organisation's data is often stored in data source systems and not utilized in any way. Data that is actively stored and processed is very hard to access for employees not in IT roles. | By making the access to data so hard, employees were not encouraged to explore it and many skills that could have been developed that way were left undeveloped. Working with data is inefficient. | 1.) making access to structured data easier by documentation, and if needed, by restructuring the database to more user-friendly structure, 2.) eliminating data silos by creating centralized data lake for data in its raw format. | Lower threshold for working with data and greater learning opportunities. Better understanding on available data. Option to benefit from currently ununsable data in the future. More efficient working with data. |

*Figure 13: Effects of suggestion 2*

The third most important way of developing target organization's data analytics capabilities is by creating simple prototypes of tools. One recognized major challenge in developing target organization's data analytics capabilities is the lack of funding from management. Without first seeing actual profits from data analytics initiatives, management is cautious in allocating funding for these projects. On the other hand, large profits often require large investments. To stop this spiral, low-cost but effective prototypes of tools should be built to

demonstrate the feasibility and utility of data analytics. After target organization has made its regulatory alignments, made access to structured data more efficient, and started to store its unstructured data centrally, capable employees across target organizations can start building simple prototypes of data analytics tools to support their business. By prototyping, target organization can also develop its skills in many of the required skill areas such as in processing and visualizing data, creating pipelines between systems, and in making observations based on the data. As target organization has put emphasize on technological skills of its new hires, building these prototypes does not necessarily require spending much on new recruitments.

The feasibility of this suggestion was also evaluated with discussions had with some of the specialists interviewed in the previous research interview. Prototyping, especially in order to develop organization's skills were said to be very important and something we should put emphasize on.

The benefits of prototyping for organization's data analytics capabilities were not assessed in this thesis's literature review, and because of that, it is hard to assess whether the target organization would have had similar effects as expected. However, the benefits of prototyping can be tested in practice by building simple versions of applications of data analytics presented in the literature. The literature suggests numerous applications of data analytics from which many can be implemented in varying level. Many of the recognized applications can be implemented at least to some degree without needing to have fully developed data architecture behind. For example, some methods of data mining presented in the literature review could be implemented to simple datasets by only having the data and necessary tools to apply the data mining. Results of these analytics could then be visualized by using some data visualization tool, or even simpler by using Python libraries. The benefits of prototyping in developing organization's data analytics capabilities are likely to be effective method for other similar organizations as well. Even organizations with already well-established capabilities in data analytics can benefit from it by adjusting the level of prototypes to their current level needs.

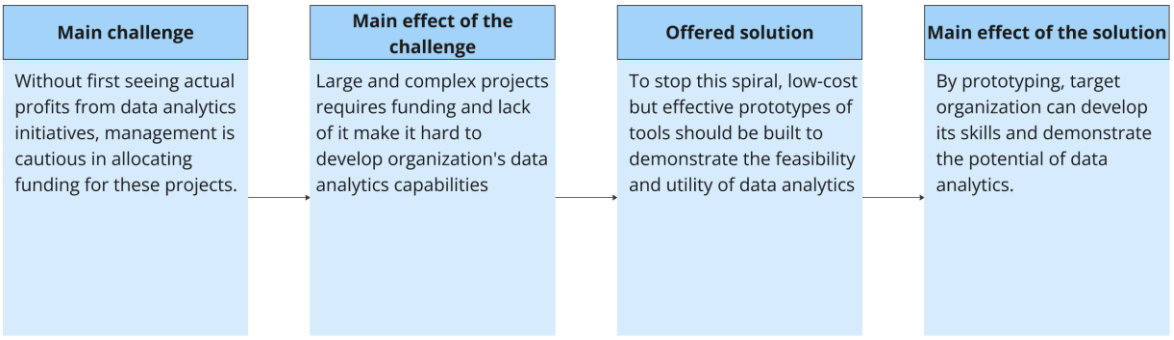



*Figure 14: Effects of suggestion 3*

The fourth suggestion this research provides is to increase the amount of general discussion and education in data analytics in target organization's common meetings and townhalls, mandatory training, and voluntary education. Currently, most employees lacks experience in working with data. Even employees, who might otherwise have relatively good technological skills are often very inexperienced in working with data and lacks understanding in basic terms and principles. This not only makes the communication between end users and developers harder, but also increases the" mystique" behind data analytics. Employees often tend to either under- or overestimate the contribution that is required to do different data analytics related tasks. "Data analytics" is also often seen as an almost meaningless buzz word without any concrete behind. On the other hand, by solely solving the first three challenges presented previously, employees might not become familiar with the changes. To get the full benefits from them, new available methods of working has to be introduced for employees, and education offered for the interested. This way the common terms and principles would slowly start to become familiar to each employee, and the "mystique" behind data could be diminished. On the general townhall meetings discussion about on-going data analytics initiatives could be included. Mandatory training could include education on data analytics related regulation, the capabilities of data analytics, and common terminology and principle of data analytics. Voluntary education could include more advanced content, such as basics of database commands, coding, and data visualization.

The feasibility of this suggestion was also evaluated with discussions had with some of the specialists interviewed in the previous research interview. Their unanimous stand on the suggestion was that it is definitely doable. Some education has to be mandatory in order to participate the whole organization.

In literature review, many skill-related requirements were recognized. Organizations were expected to have skills in technology and business enabler roles, roles responsible for processing and understanding the data. Lack of experience in working with data made it hard for employees to judge whether data analytics could benefit them in any way. When implementing data analytics projects, especially middle management were expected to have realistic expectations of the data analytics' potential. To fulfil these requirements in organization's skills, either new employees must be hired, or existing employees educated. Considering target organization's realities in funding, educating existing employees is probably more affordable and can be done with the help of existing capable employees, or if necessary, with the help of external consultants. Such a suggestion is also likely to be generalizable to other peer organisations as well. High level of seniority in target organization's local peers is likely to correlate with low level of data analytics skills, as it does within the target organization. As target organization's data analytics capabilities is described to be at the same line with its other local peers, it is unlikely that there would have happened a sudden shift in the local peer organization's employee structure.

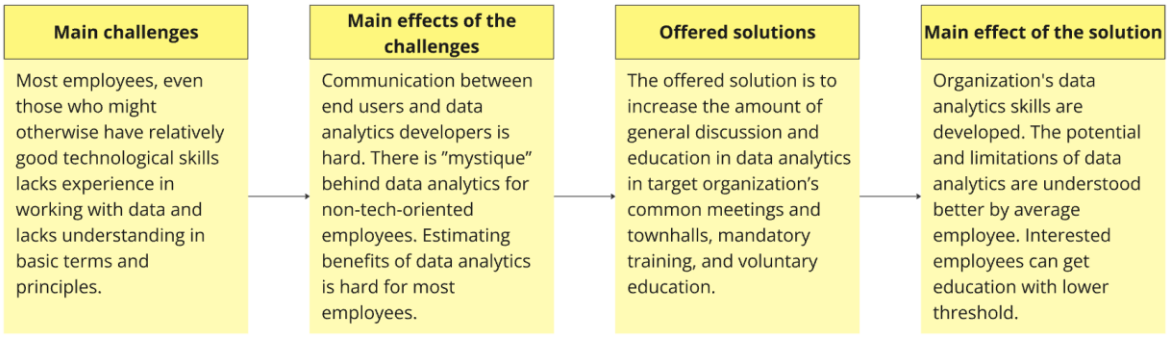



*Figure 15: Effects of suggestion 4*

Currently, there is one data analytics initiative focusing on building simple and easy-to-use prototype of a tool that can be used to support block trading. However, as this tool would utilize client data, its development has not advanced partly due to challenges mentioned previously. As mentioned in the introduction chapters of this thesis, one motive for this thesis is to figure out the challenges in implementing larger scale client data analytics project. Without being able to implement these suggestions, implementing larger scale data analytics projects becomes unlikely to success due to following factors:

**1.)** compliance has to be heavily involved as each step has to be separately evaluated from compliance's perspective. This will take time and resources making the whole process slower.

**2.)** if the database this client analytics tool uses lacks documentation or is built inefficiently from user's perspective, building the back end for the tool becomes dependent on few key employees who knows the database. There might also be performance issues if the database needs to be continuously queried but is not designed keeping this in mind.

**3.)** if the target organization has not been able to develop its basic data analytics skills by building even simple prototypes, it is likely that bigger projects are going to require additional workforce with necessary skills or help of consultants. Both options are expensive and includes their own risks.

**4.)** if the technological skill level is very polarized among employees, (i.e. 20 % of employees are very skilled and 80 % totally lacks skills) it becomes increasingly hard for employees in different roles to understand each other. This not only has negative effects on the organization's culture but also might lead to situation where average employees are not encouraged to take part in technology related tasks as the gap in skills is felt to be so large.

# 5 Findings, discussion, and conclusions

Findings of a research are based on its compiled synthesises and bind together key topics and their results. Findings are also used to present clear answers to the research questions. (Hirsijärvi et al. 2007b) This chapter answers the research questions presented at the introduction and binds the answers to the target organization's context. In this chapter the trustworthiness of the research, as well as the possible future research needs are assessed.

## 5.1 Answers for research questions

The goals of this thesis were to investigate how client data analytics can be used in Nordic equity sales & trading business, to investigate target organization's capabilities in executing large-scale client data analytics projects, and to develop target organization's client data analytics capabilities. The first research question,

*How can client data be utilized in Nordic equity sales & trading?*

was first explored from global perspective, and multiple applications of data analytics were recognized from the literature. Then after special characteristics of Nordic operating environment as well as the target organization's client data analytics capabilities were investigated with research interviews, the applicability of these globally recognized applications of data analytics were evaluated for target organization and its Nordic peers. From global literature it was recognized that client data analytics are for example used for creating client profiles based on their risk preferences, product recommendation systems, and for client satisfaction analytics. Data required to implement these applications demanded large variety of different sources and perspectives of data. From global literature it was also recognized that in order to successfully implement large scale data analytics projects, organization must have technological, human resources, and organization cultural capabilities. The first research interview explored the differences between biggest global equity brokerage businesses and their Nordic peers. Its findings concluded that compared to global institutions, effects of regulation were heavier for Nordic peers, Nordic peers were lacking in technological capabilities, had very limited number of clients, were highly specialized in Nordic markets, and had less variety in its local client's trading strategies.

Some similarities between global and Nordic peers were also recognized. Both were said to face further increasing share of electronic execution and increasing regulation.

As target organization were from technological perspective only able to systematically utilize data from its trading system, many of the recognized application could not fully be implemented. There were also other limitations in target organization's capabilities when compared to recognized requirements in global context. Target organization lacked data architecture to systematically collect, store, process, and visualize data, access to stored data was too hard for employees in non-tech roles, employees lacked demanded skills in data analytics, despite improvements in management prioritization funding was still inadequate, and there was no clear alignment in compliance set policies between different desks. However, despite the challenges, target organization were estimated to be mostly at the same level with its Nordic peers with its data analytics capabilities. Therefore, it is likely that other Nordic peers also face similar challenges with their aspirations in implementing data analytics projects. In this chapter, when discussed about target organization's capabilities, if not separately mentioned, the findings are generalized to consider whole Nordic peer group.

Considering Nordic peer group's special characteristics and likely common aspirations, designed use-cases of client data analytics should be something simple, yet effective. Solution should be able to be built by only using data from trading system, but optionally could be enriched with market data. Cost of building and running the tool should be low. It should focus on further increasing existing competitive advantages such as knowledge in local markets and clients, as global competitors cannot then simply replicate it. Taking these factors into account, equity sales & trading firms operating in Nordic region could use data analytics to further increase their specialization in Nordic markets by creating simple product recommendations systems to enhance their block trading. At simplest, these tools could utilize only trading data that would anyways be stored for compliance purposes and enhance it with current market data. As an output, tool would suggest what stocks clients might be interested of based on their past trading activities and current market insight. Also, considering the target organization's capabilities in implementing data analytics projects, the implemented tools must be very simple and low in costs. Building these kinds of tools is supported by following findings in research interviews and literature: Increased electronic trading decreases direct client interactions and makes knowledge sharing dependent on systems. One key factor giving Nordic institutions a competitive edge over global peers is

their specialization in local markets and long-standing client relationships. Developing tools that integrate client-produced data with Nordic market insights could further enhance this advantage. Management is still often hesitant in providing funding, so cost of implementation should be low. To optimally benefit from such tool, considering limited number of clients in Nordic brokerage houses, Nordic institutions must be able to utilize client data across their whole organization. In practice this means that Nordic equity brokerage houses that have desks across different Nordic countries has to pool their client data into one location. This is also logical as each county's own clients are generally known locally quite well and client data analytics might not create much additional knowledge. Another use-case of data analytics Nordic institutions might benefit from are automated monitoring and reporting for compliance. As discovered from first research interview, cost of compliance is often relatively more expensive for Nordic institutions. By automation, some processes might become more efficient reducing required amount of hired workforce. However, the discovered deficiencies in globally recognized requirements for successful data analytics implementations are so large that instead of focusing on designing an actual tool, this thesis focuses on developing target organization's data analytics capabilities. After the target organization's capabilities have improved, it is much more efficient to start implementing larger scale data analytics initiatives.

The second research question,

*How Nordic equity sales & trading organization could develop its client data analytics capabilities?*

focuses on developing target organization's data analytics capabilities. As previously mentioned, considering the special characteristics of Nordic investment banking and the target organization's technological capabilities relative to its Nordic peers, it is likely that the target organization's capabilities can broadly represent the entire Nordic peer group. Therefore, the answer to this research question is derived from the research interview conducted with the target organization.

The second research interview recognized following deficiencies in the target organization's technological data analytics capabilities. In relation to vast amount and large variety of required data, target organization's available data was not sufficient to implement many globally recognized applications of client data analytics. Target organization lacked

organized data architecture and was not able to use many of its recognized data sources. In many situations data stayed in the source systems as it was in various, often unusable structures and formats, and didn't have pipelines to transfer it into centralized location. Data was also described to be hard to access, and the used database not user friendly. From the solutions offered (Table 8) in the previous chapter, the second suggestion focussed most directly into developing target organization's technological capabilities. The presented data architecture (appendix 3) would enable storing data that cannot currently be utilized into data lake and create an option to benefit from it in the future as target organization's capabilities are enhanced. Data lake would eliminate data silos and make access to data easier. One common data visualization dashboard would make its usage easier as it is centrally developed and documented. Also, more end users might spark more ideas for development. Building prototypes might increase target organization's technological capabilities as other deficiencies in technological capabilities could be spotted and then developed. Other suggested solutions can also indirectly develop target organization's technological capabilities. By building prototypes, benefits of data analytics could be demonstrated, and if the potential is recognized by the management, the amount of funding for technological development might be increased. Also, development in employee's technological skills and organization's data driven culture, are likely to develop its technological capabilities in the long run.

*Table 9: Development of organization's technological capabilities*

| Severity | Development of technological capabilities |
|---|---|
| **Direct** | **2. suggestion**<br>• In addition to storing mandatory data in low-cost format, data lake can be used to store currently unusable data in raw format<br>• Data silos could be eliminated and access to data made easier<br>• Common visualization dashboard would make development and documentation easier<br>• More end-users could spark more ideas for development |
| **Semi-direct** | **3. suggestion**<br>• Building prototypes might help to spot other deficiencies in technological capabilities |
| **Indirect** | **1. suggestion**<br>**2. suggestion**<br>**3. suggestion**<br>**4. suggestion**<br>• Building prototypes might increase funding for technologial development if the potential can be demonstrated for management<br>• By developing organization's data analytics skills and data driven organization culture, its technological skills are also likely to increase. By making access to data easier and storing it centrally, threshold for exploring data can be decreased. |

Table 9 summarizes the effects of suggested solutions for organization's technological capabilities and ranks them by severity. The suggestions that have the most direct effect on the development of technological capabilities are presented higher in the table.

According to second research interview, target organization had following deficiencies in its human resources capabilities. Target organization was described to lack in employees with skills in data analytics. Those employees who had relevant skills in data analytics were often restricted by internal guidelines, and/or their worktime was prioritized elsewhere. Despite target organization being a part of much larger organization where relevant skills exists, finding the right employees with relevant knowledge is often very hard due to enormous headcount parent organization employs. Also, if demanded data analytics skills owning employee is found among parent organization's employees, they almost certainly lack understanding in the target organization's business.

From the offered solutions, all suggestions are aimed to develop target organization's employees' data analytics skills. Suggestion 1. enables interested employees to take part in development by providing them clear boundaries what they are allowed to do. Suggestion 2. makes access to data easier lowering the threshold and required skills to access the data. Suggestion 3. develops employee's skills by participating them into the actual development process. Employees with better technological skills can participate in the technological implementation while employees with better understanding on the business logic can participate in defining the requirements and acting as end-users. Suggestion 4. provides education for employees in all roles.

*Table 10: Development of organization's human resources capabilities*

| Severity | Development of human resources capabilities |
|---|---|
| **Direct** | **4. suggestion**<br>• As there are more education available, and data analytics related education is integrated into organization's mandatory training, employees' skills are likely to develop in long term |
| **Semi-direct** | **1. suggestion**<br>**2. suggestion**<br>**3. suggestion**<br>• 1. suggestion enables employees to participate in development spontaneously when they are aware of their boundaries<br>• 2. suggestion makes access to data easier lowering the threshold and required skills to access the data.<br>• 3. suggestion develops employee's skills by participating them into the actual development process. |
| **Indirect** | **1. suggestion**<br>**2. suggestion**<br>**3. suggestion**<br>**4. suggestion**<br>• All of the provided suggestions are made to participate as many employees as possible to developing organization's skills. The more employees work with data analytics related tasks, the more are organization's data analytics skills likely to develop. |

Table 10 summarizes the effects of suggested solutions for organization's human resources capabilities and ranks them by severity similar to Table 9.

The second research interview revealed that the target organization's culture towards data analytics had already gone through significant improvements. Management has started to prioritize data analytics more, and employees were open to new innovations. However, as target organization's skills and experience in working with data analytics were still at an elementary level, following challenges were recognized. Despite management's new focus on developing target organization's data analytics capabilities, funding they provided were still inadequate. This messaged that there is still hesitation whether data analytics would increase organization's profits in the future. Due to employees' lack of experience in working with data, potential of data analytics were hard to assess for management and regular employees. Some of the currently used tools were too advanced for regular employees to use which might increase the resistance to change as all tools are seen useless or too hard to use.

All of the given suggestions are supposed to improve organization's data driven organization culture as working with data becomes more usual even for employees without completely technology-related roles. 1. suggestion enables working with data for larger group of employees as compliance has set clear boundaries for working with data. 2. second suggestion aims to make access to data easier and working with it more efficient. From data driven organization culture's perspective, the goal of 2. suggestion is to democratize the access to data, lower individual employee's threshold to explore the data, and therefore make it easier for employees to evaluate the potential of data analytics for their own business. 3. suggestion helps more employees to recognize the potential of data analytics as simple prototypes becomes part of daily business processes. Some of these prototypes might eventually be built into actual products if they are recognized to bring significant benefits. 4. suggestions aims to decrease unnecessary "mystique" around data analytics as common terms and principles are educated for every employee.

*Table 11: Development of organization's cultural capabilities*

| Severity | Development of organization cultural capabilities |
|---|---|
| **Direct** | **4. suggestion**<br>• Even though its not realistic that all employees participate actively in developing organization's data analytics capabilities, by offering education and making parts of it mandatory, many basic principles and terms becomes more familiar for all employees. This adds more concrete meaning behind the word "Data analytics" and decreases the unneccessary "mystique". |
| **Semi-direct** | **2. suggestion**<br>**3. suggestion**<br>• 2. suggestion aims to democratize access to data, lower individual employee's threshold to explore the data, and therefore make it easier for employees to evaluate the potential of data analytics for their own business.<br>• 3. suggestion aims to demonstrate the potential of data analytics in practice and therefore support management's commitment. |
| **Indirect** | **1. suggestion**<br>**2. suggestion**<br>**3. suggestion**<br>**4. suggestion**<br>• All of the provided suggestions are made to participate as many employees as possible to working with data, therefore building organization's data driven culture. The more employees work with data analytics related tasks, the more experience they gain. As organization's experience in working with data increases, the full potential of data analytics are likely to be recognized as well. |

Table 11 follows representation similar to Table 9 and Table 10. The aggregated table can be found from appendix 5.

## 5.2 Discussion and further research

The trustworthiness of a research is often evaluated by validity and reliability. Validity measures how well thesis researches the topics that were selected to be researched. Reliability measures the repeatability of the research's findings and the correctness of the source material. When assessing the credibility of research, it is crucial to consider its purpose and target audience, the author's involvement, the methods used for collecting empirical data, the researcher's relationship with the client, the timeframe of the study, and the research methods and reporting employed. (Tuomi & Sarajärvi 2018.)

It is necessary to evaluate author's impartiality in the research, as well as their role and position. During the research researcher worked closely with the target organization to acquire the whole picture, but to best effort, preserved objectivity. However, it is important to note that author's academic background, interests, and skills diverges from the majority of the target organization's employees. This might lead into misinterpretations between author and target organization's representatives. The initial idea of the research were offered by the target organization, but its goals, results, and scope changed majorly as the research advanced. In the beginning, idea was to investigate how client data analytics could be utilized in target organization, to create a tool to analyse trading data, and to investigate whether some patterns from client's behaviour could be recognized. However, while research advanced and more understanding on the special characteristics of Nordic operating environment as well as target organization's capabilities were acquired, the direction of the research was changed. In the research it was recognized that instead of focusing on developing a single tool, much more effective would be to develop target organization's deficient data analytics capabilities. Another reason for the change in research's direction was that target organization's capabilities in data analytics were so far from the requirements recognized from global literature that implementing globally recognized applications of data analytics pretended to be impossible.

In total, 13 interviews were held for two separate research interviews, and there were 11 individual participants. All of the interviewed participants had significant amount of experience from the industry and well-established careers. Some of the interviewed were individual specialists, and some managers managing up to 100 employees. In total individual interviewed had well over 210 years of experience from the industry which supports their

credibility. Similar answers were presented for many of the questions which supports repeatability of the research. Each interviewed were asked whether they agree that the interview is recorded, and their identity, role, or title was not published. The researcher had no other than work related relationships with any of the interviewed. The timeframe of the research interviews were divided for two months as many of the interviewed employees were very busy. However, considering the scope of the research and currently on-going other projects in the target organization, there were no risk that the development in target organization would made the findings of the research obsolete during the research process. Despite the relatively small sample size of the research interviews, its results can be still argued to describe target organization's current state well as most of the answers were highlighting similar themes. Also, considering the cumulative experience of the interviewed, measuring the repeatability of results cannot only be assessed by the number of interviewed. As many of the interviewed have also worked in other Nordic peer organisations and described target organisation's capabilities to be similar, the findings of this research can be to some extent generalized to other similar organisations operating in the same region. However, the situation can change quickly if other organisations decide to start prioritising development of their data analytics capabilities. It also possible that some of the peer organisations have already started to prioritise the development, but the results are not yet recognized by the industry.

As already mentioned in the chapter 4.1 when individual topic is researched in depth, general phenomena are recognized. In this research, especially in the first research interview, common differences between Nordic and global investment banks were recognized (Figure 10). Some of those differences were highly responsible for the recognized technological gap between them. Keeping these findings in mind, other similar Nordic investment banks are likely to face similar challenges in their data analytics capabilities. Target organization's peer organizations could utilize the framework of this thesis by analysing their own technological, human resources, and organization cultural data analytics capabilities, and reflecting them with the requirements recognized from global literature. The four suggestions presented in this research (Table 8) could be utilized in other organizations as well by adjusting them to face organization's current state. Other organizations could also do their own internal evaluation on their data analytics capabilities, and then modify these suggestions to match their personal requirements. On the other hand, if peer organization's data analytics capabilities finds out to be beyond target organization's capabilities, the

literature review with its global applications of client data analytics might prove to be more valuable.

This research recognized areas of improvement and offered solutions to develop target organization's client data analytics capabilities. The offered solutions were evaluated by few of the research interview participants. Their stand on the offered solutions were overall positive and they evaluated suggestions to be practical and feasible. After the target organization's capabilities to implement client data analytics projects becomes sufficient, further research should be done on the actual implementation of various data analytics applications. The data analytics capabilities framework introduced in this research could also be applied to evaluate other asset class's brokerage business's capabilities. As there wasn't existing research on the special characteristics of Nordic investment banking or on their data analytics capabilities, this research aimed to fulfil that gap. However, as Nordic region consists of individual countries with individual histories, cultures, and laws, researching the differences between individual Nordic countries' investment banks might help universal Nordics banks to become more united. This research scoped out investigation on the effects and applications of artificial intelligence (AI) for client data analytics. Nevertheless, the immense number of applications and possibilities are likely to bring AI even closer to client data analytics in the future. Hence, further research should be directed to investigate applications of AI for client data analytics in Nordic investment banks.

# References

Akyildirim, E., Sensoy, A., Gulay, G., Corbet, S., & Salari, H. N. (2021). Big data analytics, order imbalance and the predictability of stock returns. *Journal of Multinational Financial Management*, 62, 100717.

Ali, O., Suwabe, T., & Walsh, D. (2016). The Role of Big Data in Investing. Goldman Sachs Asset Management. Available: https://www.gsam.com/content/gsam/global/en/market-insights/gsam-insights/gsam-perspectives/2016/big-data/gsam-roundtable.html

AWS. What is ETL (Extract Transform Load)? Available: https://aws.amazon.com/what-is/etl/

Anoshin, D., Shirokov, D., & Strok, D. (2020). Jumpstart Snowflake: A Step-by-Step Guide to Modern Cloud Analytics. 1st ed. 2020. [Online]. Berkeley, CA: Apress.

Banton, C. (2024). Discount Broker: Definition, What They Do, Downsides, and Cost. Investopedia. Available: https://www.investopedia.com/terms/d/discountbroker.asp

Banton, C. (2023). Underwriting: Definition and How the Various Types Work. Investopedia. Available: https://www.investopedia.com/terms/u/underwriting.asp

Baumeister, R. & Leary, M. (1997). Writing Narrative Literature Reviews. *Review of General Psychology*.

Belyadi, H. & Haghighat, A. (2021). Machine learning guide for oil and gas using Python: a step-by-step breakdown with data, algorithms, codes, and applications. Amsterdam, The Netherlands: Gulf Professional Publishing.

Berger, A. N., Molyneux, P., & Wilson, J. O. S. (2010). Oxford handbook of banking. 1st ed. Oxford; Oxford University Press.

Berman, N. (2023). What Is the Dodd-Frank Act? Council on Foreign Relations. Available: https://www.cfr.org/backgrounder/what-dodd-frank-act

Boobier, T. (2020). AI and the future of banking. Wiley, West Sussex, England.

Byrnes, N. (2017). As Goldman Embraces Automation, Even the Masters of the Universe Are Threatened. *MIT Technology Review*. Available: https://www.technologyreview.com/2017/02/07/154141/as-goldman-embraces-automation-even-the-masters-of-the-universe-are-threatened/

Chandani A., & Mehta, M. (2015). Banking on Big Data: A Case Study. *Journal of Engineering and Applied Sciences*, vol. 10, 2015.

Corporate Finance Institute. Brokerage. [Cited: 25.6.2024]. Available: https://corporatefinanceinstitute.com/resources/commercial-lending/brokerage/

Davis, L. (2024). What Is Customer Analytics? Forbes Advisor. Available: https://www.forbes.com/advisor/business/customer-analytics/

Detzer, D. (2015). Financial Market Regulation in Germany - Capital Requirements of Financial Institutions. *PSL Quarterly Review*, 68, 57-87.

Doko, F. & Miskovski, I. (2019). An Overview of Big Data Analytics in Banking: Approaches, Challenges and Issues. UBT International Conference. Available: https://knowledgecenter.ubt-uni.net/cgi/viewcontent.cgi?article=2443&context=conference

Downey, L. (2024). Brokerage Commission House: What It Is, How It Works. Investopedia. [cited: 24.6.2024]. Available: https://www.investopedia.com/terms/c/commissionhouse.asp

Du, Y. (2023). Chinese Wall. Wall Street Oasis. Available: https://www.wallstreetoasis.com/resources/skills/finance/chinese-wall-definition

European Commission. (2013). A comprehensive EU response to the financial crisis: a strong financial framework for Europe and a banking union for the eurozone. Available: https://ec.europa.eu/commission/presscorner/detail/de/MEMO_13_679

Ertek, G., Kaya, M., Kefeli, C., Onur, Ö., Uzer, K. (2012). Scoring and Predicting Risk Preferences. Behavior Computing. Springer, London. https://doi.org/10.1007/978-1-4471-2969-1_9

Eskola, J. & Suoranta, J. (1998). Johdatus laadulliseen tutkimukseen. Tampere: Vastapaino.

ESMA. Article 16 Organizational requirements. [Cited: 30.7.2024]. Available: https://www.esma.europa.eu/publications-and-data/interactive-single-rulebook/mifid-ii/article-16-organisational-requirements

Federal Reserve History. Banking Act of 1933 (Glass-Steagall) Available: https://www.federalreservehistory.org/essays/glass-steagall-act

Fohlin, C. (2014). A brief history of investment banking from medieval times to the present. In The Oxford Handbook of Banking and Financial History, p.133->.

Geisst, C. R. (2012). Wall Street: a history. Updated ed. Oxford University Press.

Goldstein, M. A., Irvine, P., Kandel, E., & Wiener, Z. (2009). Brokerage Commissions and Institutional Trading Patterns. *The Review of Financial Studies*, 22(12), 5175–5212.

Hargrave, M. (2024). Bill of Exchange Definition: Examples and How It Works. Investopedia. Available: https://www.investopedia.com/terms/b/billofexchange.asp

Hayes, A. (2024). Commercial Paper: Definition, Advantages, and Example. Investopedia. Available: https://www.investopedia.com/terms/c/commercialpaper.asp#:~:text=Commercial%20paper%20is%20an%20unsecured,a%20discount%20from%20face%20value.

Hassani, H., Huang, X., & Silva, E. (2018). Digitalisation and Big Data Mining in Banking. *Big data and cognitive computing*. [Online] 2 (3), 18-.

Heakal, R. (2024). Glass-Steagall Act of 1933: Definition, Effects, and Repeal. Investopedia. (cited: 3.6.2024) Available: https://www.investopedia.com/articles/03/071603.asp#:~:text=The%20Glass%2DSteagall%20Act%20prevented,in%201999%20under%20President%20Clinton.

Herath, S. (2024). Fundamentals of Associate Rule mining. Medium. Available: https://medium.com/image-processing-with-python/fundaments-of-associate-rule-mining-468801ec0a29

Hirsjärvi, S., Remes, P. & Sajavaara, P. (2007). Tutki ja kirjoita. 13. painos. Helsinki, Tammi.

Hirsjärvi, S., Remes, P., & Sajavaara, P. (2004). Tutki ja kirjoita. 10. osin uud. laitos. Helsinki: Tammi.

HSBC. MiFID Framework. (2018). [cited: 17.7.2024]. Available: https://www.gbm.hsbc.com/en-gb/financial-regulations/mifid/the-mifid-framework

Holdsworth, J. (2024). What is data mining? IBM. Available: https://www.ibm.com/topics/data-mining

Hwan, I. (2024). A Brief History of the Stock Market. SoFi Learn. Available: https://www.sofi.com/learn/content/history-of-the-stock-market/

Hyvärinen, M., Nikander, P., Ruusuvuori, J., & Aho, A. L. (2017). Tutkimushaastattelun käsikirja. Tampere: Vastapaino.

IBM a. What is a data architecture? Available: https://www.ibm.com/topics/data-architecture

IBM b. What is data warehouse? Available: https://www.ibm.com/topics/data-warehouse

IBM c. What is data mart? Available: https://www.ibm.com/topics/data-mart

IBM d. What is a decision tree? Available: https://www.ibm.com/topics/decision-trees

Iraqi, B., Benhiba, L., & Janati Idrissi, M. A. (2023). Data-Driven Transformation in Investment Banks. in 2023 IEEE 6th International Conference on Cloud Computing and Artificial Intelligence: Technologies and Applications (CloudTech). IEEE, pp. 01–09.

Isinkaye, O., Folajimi Y., & Ojokoh B. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*. 16. 10.1016/j.eij.2015.06.005.

Jain, A. (2024). A Comprehensive Guide to Performance Metrics in Machine Learning. Medium. Available: https://medium.com/@abhishekjainindore24/a-comprehensive-guide-to-performance-metrics-in-machine-learning-4ae5bd8208ce

J.P. Morgan. (2024). The e-Trading Edit: Insights from the inside. Available: https://www.jpmorgan.com/markets/etrading-trends#introduction

Johnson, R. (2019). Trends in Global Equity Electronic Execution. Bloomberg. Available: https://data.bloomberglp.com/promo/sites/12/607535_Trends-Global-Equity-Electronic-Execution.pdf

Kalita, J. K., Bhattacharyya, D. K., & Roy, S. (2024). Fundamentals of data science: theory and practice. London: Elsevier Science & Technology.

Kohn, M. (1999). The Evolution of Financial Institutions and Markets in Twentieth Century Europe. Aldershot: Scholar Press, pp. 187–208.

Kremser, W., & Brunauer, R. (2019). Do we have a Data Culture? Data Science – Analytics and Applications. Springer Vieweg, Wiesbaden. Available: https://doi.org/10.1007/978-3-658-27495-5_11

Kutzner, K., Schoormann, T., & Knackstedt, R. (2018). DIGITAL TRANSFORMATION IN INFORMATION SYSTEMS RESEARCH: A TAXONOMY-BASED APPROACH TO STRUCTURE THE FIELD. European Conference on Information Systems.

Lee, K., & Manochin, M. (2021). Sell-side equity analysts and equity sales: A study of interaction. *The British Accounting Review*, Volume 53, Issue 5.

Lehtiranta, L., Junnonen, M. & Pekuri, L. (2015). 'The Constructive Research Approach: Problem Solving for Complex Projects', in Designs, Methods and Practices for Research of Project Management. [Online]. Routledge. pp. 125–136.

Leshik, E. A., & Cralle, J. (2011). An introduction to algorithmic trading basic to advanced strategies. 1st ed. Wiley, Chichester, U.K.

Liaw, K. (2011). The Business of Investment Banking: A Comprehensive Overview. John Wiley & Sons, Incorporated, Newark. Available from: ProQuest Ebook Central.

Maheshwari, A. (2014). 'Data Mining', in Business Intelligence and Data Mining. United States: Business Expert Press.

Michalczyk, L. (2012). Market for investment banking products in modern economy. *E-finance*, 8(1), 14–22.

Microsoft Azure. What is a Data Lake? Available: https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-a-data-lake

Miyan, M. (2017). Applications of Data Mining in Banking Sector. *International journal of advanced research in computer science*. 8 (1).

Miller, J. D. (2017). Big data visualization. 1st edition. Birmingham, England; Packt Publishing.

Murphy, T. (2023). Definition: Customer analytics (customer data analytics). TechTarget. Available: https://www.techtarget.com/searchbusinessanalytics/definition/customer-analytics

Moody's Analytics. (2018). Regulatory Big Data: Meeting the Costs & Challenges of Granular Reporting. Available: https://www.moodysanalytics.com/articles/2018/regulatory-big-data-meeting-the-cost-challenges-reporting

Nuti, G., Mirghaemi, M., Treleaven, P., & Yingsaeree, C. (2011). Algorithmic Trading. *Computer (Long Beach, Calif.)*, 44(11), 61–69.

Piech, C. (2013). K Means. [cited: 1.8.2024] Stanford University. Available: https://stanford.edu/~cpiech/cs221/handouts/kmeans.html

Perold, A. F. (1988). The implementation shortfall: Paper versus reality. *Journal of Portfolio Management,* 14(3), 4–9.

Puusa, A. & Juuti, P. (2020). Laadullisen tutkimuksen näkökulmat ja menetelmät. Helsinki, Gaudeamus

Powell, S. Buy Side vs Sell Side. Corporate Finance Institute. [cited: 17.6.2024]. Available: https://corporatefinanceinstitute.com/resources/career/buy-side-vs-sell-side/#:~:text=Buy%2DSide%20vs%20Sell%20Side,%2C%20advisory%20firms%2C%20and%20corporations.

PWC. (2019). Future of Sales: How Investment Banks Must Adapt to a Shifting Sales Landscape. Available: https://www.pwc.co.uk/financial-services/assets/pdf/future-sales-investment-banks-must-adapt-to-shifting-sales-landscape.pdf

Rafatirad, S. (2022). Machine learning for computer scientists and data analysts : from an applied perspective. Cham, Switzerland: Springer.

Roy, A. (2020). A Dive into Decision Trees. Towards Data Science. Available: https://towardsdatascience.com/a-dive-into-decision-trees-a128923c9298

SEC. What is a broker-dealer? [cited: 5.6.2024] Available: https://www.sec.gov/files/oasb-broker-dealer-building-block.pdf

Seth, S. (2023). Basics of Algorithmic Trading: Concepts and Examples. Investopedia. Available: https://www.investopedia.com/articles/active-trading/101014/basics-algorithmic-trading-concepts-and-examples.asp

Shende, V. (2017). Project Analytics to Improve Project and Portfolio Decision Making. PMI Conference 2017. Available: https://www.pmi.org.in/conference2017/pdfs/papers-pdfs/theme-3-rapidly-changing-world/21-Project-Analytics-to-Improve-Project.pdf

Scikit Learn a. Decision Trees. [cited: 1.8.2024]. Available: https://scikit-learn.org/stable/modules/tree.html

Scikit Learn b. f1_score. Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html

Scikit Learn c. silhouette_score. Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html

Soni, A., Saxena, A., & Bajaj, P. (2020) A Methodological Approach for Mining the User Requirements Using Apriori Algorithm. *Journal of cases on information technology*. [Online] 22 (4), 1–30.

Stowell, D. (2012). Investment Banks, Hedge Funds, and Private Equity: The New Paradigm. 2nd ed. Academic Press/Elsevier, Waltham, MA.

Taylor-Sakyi, K. (2016). Big Data: Understanding Big Data. arXiv.org. [Online]

Tableau. What Is Data Visualization? Definition, Examples, And Learning Resources. Available: https://www.tableau.com/learn/articles/data-visualization

Turney, D., Schroeck, M., & Shockley, R. (2013). Analytics: The Real-World Use of Big Data in Financial Services. IBM® Institute for Business Value, and Saïd Business School at the University of Oxford. IBM. Available: https://www.ibm.com/downloads/cas/E4BWZ1PY

Tuomi, J. & Sarajärvi, A. 2018. Laadullinen tutkimus ja sisällönanalyysi. Uudistettu laitos. Helsinki: Kustannusosakeyhtiö Tammi.

Zikopoulos, P. C., Eaton, C., DeRoos, D., Deutsch, T., Lapis, G. (2012). Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. McGraw-Hill, New York, NY.

Vehkalahti, K. (2014) Kyselytutkimuksen mittarit ja menetelmät. Helsinki: Finn Lectura.

Vijay, B. R. M. & Woldesamuel, L. B. (2012). Improving Customer Relationship Management in Electronic Transaction Expansion in Banking Sector by using Data mining techniques. *I-Manager's Journal on Software Engineering*. [Online] 6 (3), 28–43.

Vivek B. (2021). Application of Association Rule Mining in Banking Sector. *International Journal of Economic Perspectives*,15(1),335–343

Wilcox, T., Jin, N., Flach, P., & Thumim, J. (2019) A Big Data platform for smart meter data analytics. *Computers in industry*. [Online] 105250–259.

## Appendices

Appendix 1. Business environment research interview.

1. What is your background in equities business and how many years of experience do you have?
2. What directions of development has changed Nordic equity brokerage business the most during past decade?
3. What are likely to be the biggest challenges and opportunities during the next decade?
4. What are the biggest regulatory challenges considering Nordic equity brokerage houses vs. global brokerage houses?
5. Could you describe the level of technological adoption in Nordic equity brokerage houses compared to global ones?
6. What does the competitive landscape look like for Nordic equity brokerage houses versus global brokerage houses?
7. How is equity sales job different in Nordic environment compared to global environment? (for example, more clients to cover, broader or narrower product offering)
8. How is equity research analyst's job different in Nordic environment compared to global landscape? (for example, less companies could mean more throughout knowledge of all the biggest listed companies and their management)
9. Is there a difference in clients' equity strategies in the Nordics compared to global landscape? (for example, globally more hedge funds and short selling?)
10. Is there some another question that we should include into this interview?

Appendix 2. Data analytics capabilities, goals and projects research interview.

1. What has been your role in developing target organization's IT and data analytics capabilities?

2. What data analytics processes are currently going on? How are data analytics currently utilized in target organization's equity brokerage business?

3. Could you describe target organization's current data analytics infrastructure? What tools and technologies are currently being used in target organization's data analytics processes? Where does the data come from?

4. What would target organizations want to be able to accomplish? How would data analytics benefit the target organization the most?

5. What are target organization's competitors doing? How do target organization compare to them?

6. What are the biggest challenges in target organization's data analytics initiatives?

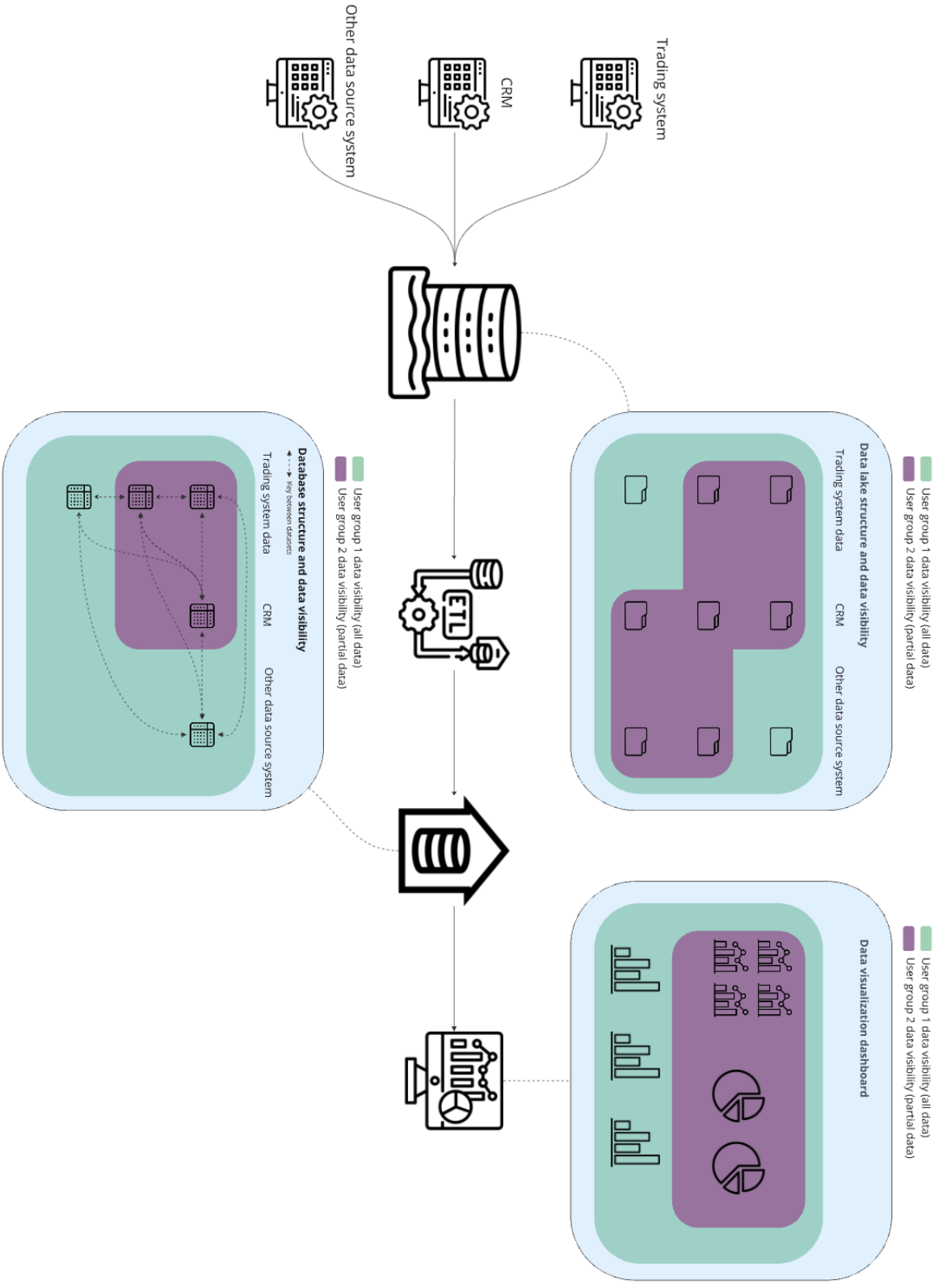
7. Could you describe the level of data analytics skills target organization's employees have?

8. Could you describe the organization culture towards data analytics? Are employees open to new innovations, and to possible changes in their daily routines? Do target organization's employees think data analytics could improve its business?

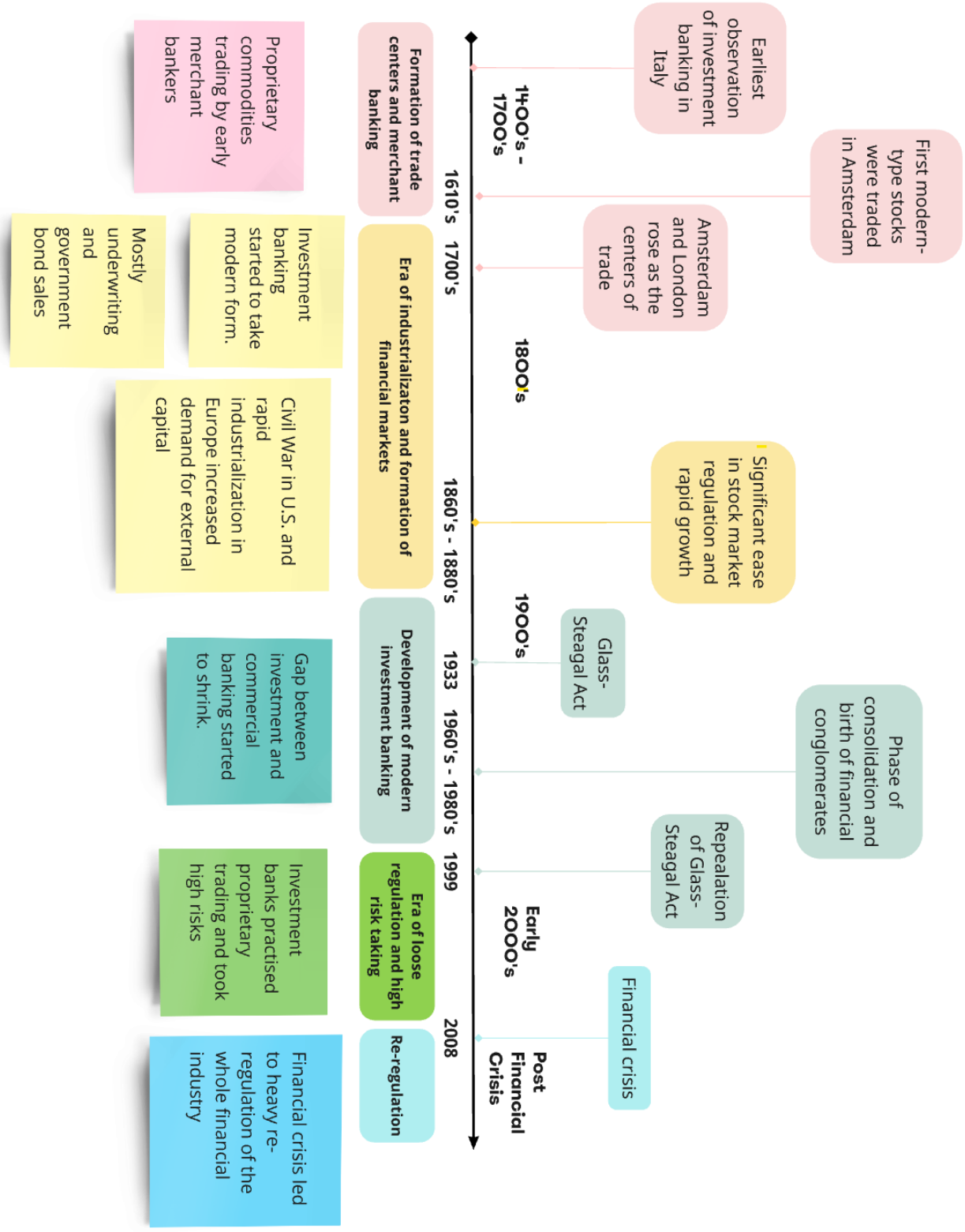
9. How does regulation affect target organization's data analytics initiatives?

10. Is there some other question we should consider?

Appendix 3. Suggestion for target organization's data architecture.



Appendix 4. History of investment banking, timeline.



Appendix 5. Effects of presented suggestions.

| Severity | Development of technological capabilities | Development of human resources capabilities | Development of organization cultural capabilities |
|---|---|---|---|
| **Direct** | **2. suggestion**<br>• In addition to storing mandatory data in low-cost format, data lake can be used to store currently unusable data in raw format<br>• Data silos could be eliminated and access to data made easier<br>• Common visualization dashboard would make development and documentation easier<br>• More end-users could spark more ideas for development | **4. suggestion**<br>• As there are more education available, and data analytics related education is integrated into organization's mandatory training, employees' skills are likely to develop in long term | **4. suggestion**<br>• Even though its not realistic that all employees participate actively in developing organization's data analytics capabilities, by offering education and making parts of it mandatory, many basic principles and terms becomes more familiar for all employees. This adds more concrete meaning behind the word "Data analytics" and decreases the unneccessary "mystique". |
| **Semi-direct** | **3. suggestion**<br>• Building prototypes might help spot other deficiencies in technological capabilities | **1. Suggestion**<br>**2. Suggestion**<br>**3. Suggestion**<br>• 1. suggestion enables employees to participate in development spontaneously when they are aware of their boundaries<br>• 2. Suggestion makes access to data easier lowering the threshold and required skills to access the data.<br>• Suggestion 3. develops employee's skills by participating them into the actual development process. | **2. Suggestion**<br>**3. Suggestion**<br>• 2. suggestion aims to democratize access to data, lower individual employee's threshold to explore the data, and therefore make it easier for employees to evaluate the potential of data analytics for their own business.<br>• 3. suggestion aims to demonstrate the potential of data analytics in practice and therefore support management's commitment. |
| **Indirect** | **1. Suggestion**<br>**2. Suggestion**<br>**3. Suggestion**<br>**4. Suggestion**<br>• Building prototypes might increase funding for technologial development if the potential can be demonstrated for management<br>• By developing organization's data analytics skills and data driven organization culture, its technological skills are also likely to increase. By making access to data easier and storing it centrally, threshold for exploring data can be decreased. | **1. Suggestion**<br>**2. Suggestion**<br>**3. Suggestion**<br>**4. Suggestion**<br>• All of the provided suggestions are made to participate as many employees as possible to developing organization's skills. The more employees work with data analytics related tasks, the more are organization's data analytics skills likely to develop. | **1. Suggestion**<br>**2. Suggestion**<br>**3. Suggestion**<br>**4. Suggestion**<br>• All of the provided suggestions are made to participate as many employees as possible to working with data, therefore building organization's data driven culture. The more employees work with data analytics related tasks, the more experience they gain. As organization's experience in working with data increases, the full potential of data analytics are likely to be recognized as well. |