TUOTANTOTALOUDEN KOULUTUSOHJELMA

Analytiikan hyödyntäminen markkinoinnissa: segmentointi ja asiakkaan arvottaminen

Utilizing analytics in marketing: segmentation and customer valuation

Kandidaatintyö

Topias Suortti
Joel Lahtinen

TIIVISTELMÄ

Tekijä: Topias Suortti, Joel Lahtinen

Työn nimi: Analytiikan hyödyntäminen markkinoissa: segmentointi ja asiakkaan arvottaminen

Vuosi: 2017 **Paikka: Lappeenranta**

Kandidaatintyö. Lappeenrannan teknillinen yliopisto, tuotantotalous.
40 sivua, 10 kuvaa ja 7 taulukkoa.
Tarkastaja: Kalle Elfvengren

Hakusanat: Analytiikka, Data, Liiketoimintatiedon hallinta, Markkinointianalytiikka, Markkinointi, Liiketoiminta-analytiikka
Keywords: Analytics, Data, Business intelligence, Marketing analytics, Marketing, Business analytics

Työn tavoitteena on selvittää, miten data-analytiikkaa voidaan käyttää markkinoinnissa. Kandidaatintyö käsittelee data-analytiikkaa, sen käyttöä ja sen hyödyntämistä markkinoinnissa. Pääpaino työssä on analytiikalla saavutettavista hyödyistä markkinoinnin toimialueella.

Markkinointianalytiikka on osa-alue, joka kuuluu data-analytiikkaan ja se on liiketoimintatiedon hallinan sovelluksen alue. Kandidaatintyö on kirjallisuuskatsaus, jossa yhdistetään eri lähteitä yhdeksi kokonaisuudeksi. Työssä käydään läpi tarpeen määrittämisen ja datan löytämisen kautta analysoinnin avulla tuotaviin etuihin ja sovellutuksiin. Kahden esimerkkicasen avulla havainnollistetaan mahdollisia sovellutustapoja.

Tärkeimpiä tuloksia työssä on se, että jo tällä hetkellä markkinoinnin tehostamiseksi ja kohdentamiseksi data-analytiikka on erittäin tarpeellinen väline. Dataa on tarjolla niin paljon ja tapoja on monenlaisia luoda lisää arvoa yritykselle tästä datasta. Yritysten on suositeltavaa käyttää analyyttisia keinoja kilpailuedun saavuttamiseksi.

SISÄLLYSLUETTELO

1. Johdanto 1
1.1 Taustaa työstä 1
1.2 Työn rajaukset, tavoitteet ja aineisto 2
1.3 Työn rakenne 3
2 Data-analytiikka 4
2.1 Data-analytiikan määritelmä sekä käsitteet 4
2.2 Yleistä data-analytiikasta 6
2.3 Tekstin louhinta 7
2.4 Verkkolouhinta 9
2.5 Datan eri tyypit 10
3 Liiketoiminta-analytiikka ja Liiketoimintatiedon hallinta 12
3.1 Liiketoiminta-analytiikka eli business analytics 12
3.2 Liiketoimintatiedon hallinta eli business intelligence 16
4 Markkinointianalytiikka 17
4.1 Mistä dataa saadaan markkinointiin ja mihin sitä voidaan käyttää? 17
4.2 Segmentointi 19
4.3 Asiakkaan elinkaaren arvon ja ostamisen alttiuden ennustaminen 20
4.4 Kohti tehokkaampaa markkinointia 21
5 Esimerkkitapaukset 23
5.1 RFM-malliin perustuva asiakassegmentointi datan louhinnan avulla 23
5.1.1 Tutkimuksen lähtökohdat 24
5.1.2 Työtavat 25
5.1.3 Tutkimuksen tulokset ja pohdinta 27
5.2 Myynninedistämisen tehokkuuden mittaaminen kävijämäärien avulla. 30
5.2.1 Työn lähtökohdat 30
5.2.2 Työtavat 31
5.2.3 Tutkimuksen tulokset ja pohdinta 32
5.3 Yhteenveto tutkimukista sekä pohdintaa 34
6 Johtopäätökset 36
7 yhteenveto 39
8 Lähteet 41

Lyhenteet

Lyhenne	Sana	Suomennos
BI	Business Intelligence	Liiketoimintatiedon hallinta
BA	Business analytics	Liiketoiminta-analytiikka
IoT	Internet of Things	Esineiden internet
M2M	Machine to machine	Laitteelta laitteelle
MPC	Marketing/Promotion Campaign	Markkinointi- ja/tai promootiokampanja
IT	Information techology	Informaatioteknologia
CRM	Customer relationship management	Asiakassuhteiden hallinta
RFM-Model	Recency, frequency and monetary model	Äskettäisyyteen, toistuvuuteen ja rahalliseen arvoon perustuva malli
ROI	Return of investment	Sijoitetun pääoman tuotto

1. JOHDANTO

1.1 Taustaa työstä

Nopea kehittyminen data-analyysityökaluissa ja -sovelluksissa on jättänyt monet julkisen ja yksityisen sektorin toimijat valmistautumattomaksi muutokseen (Gandomi & Murtaza 2014, 1). IBM arvioi vuonna 2015, että ihmiskunta tuottaa 2.4 triljoonaa bittiä dataa päivittäin, mikä vaatii yhä tehokkaampia ja enemmän käytettyjä keinoja kaiken tarvittavan raakatiedon analysoimiseksi (García et al. 2015, 93). Tässä vaiheessa data-analytiikka astuu kuvaan, jotta erilaiset instituutiot voivat käyttää tarvitsemaansa dataa toimintansa tukemiseen. Data-analytiikan kehitystä kuvaa hyvin se, että vielä 20 vuotta sitten käsitteenä oli käytössä vain datan louhinta ja nyt data-analytiikka tarkoittaa monivaiheista prosessia, jolla tuotetaan materiaalia päätöksenteon tueksi (Runkler 2012, 2). Liiketoiminta-analytiikan trendikkyys huomattiin jo 2010, kun toimialan kasvuprosentin oletettiin kasvavan 8 prosenttiin seuraavaan kolmeen vuoteen (Liberatore & Luo 2010, 2).

Data-analytiikkaa on käytetty pitkään jo erilaisten liiketoimintafunktioiden tukena, kuten esimerkiksi asiakaspalvelussa, rahoituksessa ja human resources (HR)-puolella (Saxena & Srinivasan 2013, 85). Kehittyneet ja koko ajan jalostuvat työkalut, kuten ERP-systeemit ja erilaiset analyysiohjelmat, mahdollistavat tehokkaamman datan hyötykäytön yritysten käytössä ja päätöksenteossa. Tätä kutsutaan liiketoimintatiedon hallinnaksi eli business intelligenceksi. Myös markkinointianalytiikka osana liiketoimintatiedon hallintaa on kehittynyt hiljalleen, vaikka yleensä analytiikka liitetään yrityksissä vahvasti rahoituspuoleen (Rackley 2015, 2). Kuitenkin Rackley vertaa markkinoinnin operointia ilman kunnon analyysityökaluja auton ajamiseen moottoritiellä ilman minkäänlaisia mittareita tai ohjauspaneelia, sillä ilman analytiikkaa yrityksellä ei ole mitään ennakoivaa tapaa saada arvioitua tuloksia tai kehittää sen perusteella toimintaansa, ja ainut indikaattori onnistumiselle on lopullinen tulos operaatiossa (Rackley 2015, 1).

1.2 Työn rajaukset, tavoitteet ja aineisto

Työssä on tarkoitus tuottaa katsaus nykyaikaiseen data-analytiikan toimintaan, käyttökohteisiin ja mahdollisuuksiin markkinoinnin näkökulmasta. Painopiste on data-analytiikassa nimenomaan analyysiosassa ja mitä kaikkea analysoinnilla saadaan aikaan markkinoinnin puolella. Data-analytiikan selvittämisen ja avaamisen jälkeen luodaan katsaus data-analytiikan hyödyntämisestä yleisesti liiketoiminnassa. Teoriaosion lopussa käydään läpi markkinointianalytiikkaa ja sen mahdollisuuksia. Case-esimerkeissä tarkoituksena on vertailla käytettyjä markkinointianalytiikan malleja eri tapauksissa ja pohtia niiden hyötyjä.

Data-analytiikka koostuu kolmesta osasta, jotka ovat: tiedon kerääminen ja raportointi, analysointi ja mallintaminen sekä käyttö päätöksenteon tukena. Käsitteenä data-analytiikka jakautuu työssä useampaan osaan, jotka ovat: big data, liiketoimintatiedon hallinta, liiketoiminta-analytiikka ja markkinointianalytiikka. Näistä työssä tärkein osa-alue on liiketoimintatiedon hyödyntäminen nimenomaan markkinointianalyysityökaluilla, joten keskitymme enemmän siihen, mitä analysoimalla saa datasta irti ja kuinka sitä voi käyttää, eikä siihen, kuinka data-analytiikkaprosessi toimii markkinoinnin puolella kokonaisuudessaan. Työssä luokitellaan ominaisuuksia, joita kunnollisilla analyysityökaluilla tulisi olla ja esimerkit näistä työkaluista tulevat casetöiden ohessa. Keskeiset tutkimuskysymykset ovat:

1. Mitä data-analytiikka on markkinoinnissa?
2. Kuinka data-analytiikkaa voi hyödyntää markkinoinnissa?

Lähteet työhön ovat pääasiassa data-analytiikkaa, liiketoiminta-analytiikkaa, liiketoimintatiedon hallintaa ja markkinointia sekä markkinointianalytiikkaa käsitteleviä kirjoja ja artikkeleita. Pääasiallinen tietokanta, josta kirjoja ja artikkeleita hankittiin, oli Lappeenrannan teknillisen yliopiston kirjaston Finna-palvelu. Hakusanat, joita käytettiin muotoutuivat käsitteen data-analytiikka ympärille ja useat tietolähteet löytyivätkin yhdistelemällä tarvittuja konsepteja, kuten louhintaa, liiketoimintatietoa ja markkinointia data-analytiikan käsitteeseen. Kirjat, sekä data-analytiikan, että markkinoinnin puolelta, tuovat teoriaa ja termistöä työhön. Artikkelit kertovat, kuinka data-analytiikkaa käytetään nykyään ja kuinka sitä sovelletaan käytännössä. Case-esimerkit ovat artikkeleita, joissa pyritään

parantamaan yritysten toimintaa uusilla data-analytiikan ratkaisuilla ja niiden soveltamisella. Työssä käsitellään pääasiassa liiketoimintadatan analysointia ja siitä saatavia johtopäätöksiä, vaikka käsitteeseen kuuluu laajasti muitakin ominaisuuksia.

Markkinointianalytiikan osalta tarkastelukohteita ovat segmentointi ja markkinatutkimus, joita tutkitaan data-analytiikan soveltamisen ehdoin. Tärkeänä osana on tunnistaa oikeat mittarit markkinoinnin tuloksien taustalla ja kerätä sekä analysoida dataa näistä lähteistä. Tutkinta kohdistetaan data-analytiikan tämän hetkisiin käyttökohteisiin markkinoinnin strategisissa päätöksissä sekä tulevaisuuden tuomiin mahdollisuuksiin data-analytiikan soveltamisessa.

1.3 Työn rakenne

Työ koostuu kappaleista, joilla luodaan katsaus data-analytiikkaan ja sen käyttökohteisiin markkinoinnissa. Ensimmäisessä luvussa pureudutaan data-analytiikkaan käsitteenä ja siinä selvitetään termistöä, analytiikkaprosessin kulkua, tärkeintä datan louhintatapaa ja datan eri lähteitä. Toisessa luvussa tuomme esiin liiketoiminassa hyödynnettäviä malleja ja mistä palasista liiketoiminta-analytiikka koostuu sekä kuinka se linkittyy liiketoimintatiedon hallintaan. Liiketoiminnassa hyödyllisiä analytiikan sovellutuksia ja työkaluja käydään läpi niiden mahdollisuuksien kautta sekä selvitetään kuusi analytiikan toimialueen osa-aluetta ja niiden sisältö sekä funktio.

Kolmas luku puretutuu markkinointianalytiikkaan, markkinointidatan lähteisiin ja markkinointianalytiikan työkalujen vaatimuksiin. Lisäksi luvussa perehdytään tarkemmin casetöissä käytettyihin markkinointianalytiikan käyttökohteisiin, jotka ovat segmentointi sekä asiakkaan arvo yritykselle ja nämä linkitetään organisaation arvon luomiseen. Pääpaino on markkinoinnin kehittämisessä analytiikan avulla ja siitä saatavat hyödyt.

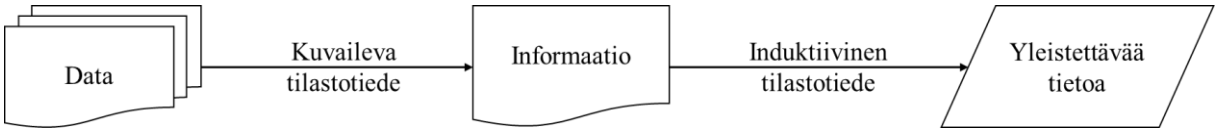
Luvussa 4 tutkimme oikean elämän esimerkkitapauksia siitä, miten data-analytiikkaa pystyy hyödyntämään markkinoinnissa ja miten se voi hyödyttää yrityksiä. Työhön valitut esimerkkitapauksevat käsittelevät segmentointia sekä markkinoinnin tehokkuuden mittaamista data-analyyttisin työkaluin.

2 DATA-ANALYTIIKKA

Tässä kappaleessa käsitellään data-analytiikkaa. Käsittelyssä ei syvennytä itse laskentaan asti, vaan pysytään yleisemmällä tasolla. Ensimmäisessä osiossa keskitytään aihepiirin yleisiin käsitteisiin ja toisessa osiossa käsitellään data-analytiikkaprosessia.

2.1 Data-analytiikan määritelmä sekä käsitteet

Data-analytiikka (data analytics) on yleisesti määritelmänsä mukaan tietokonesovellusten käyttämistä suurten tietokokonaisuuksien analysointiin. Data-analytiikka on sisällöltään monipuolinen käsite, koska siihen sisältyy ominaisuuksia tilastotieteestä, koneoppimisesta sekä operaatioanalyysistä. (Runkler 2012, 2)



Kuva 1. Tilastotieteiden ero. (Cleff 2014, 3)

Karkeasti tilastotiede voidaan jakaa kahteen osaan: kuvailevaan tilastotieteeseen ja induktiiviseen datan analyysiin. Erona näillä kahdella on se, että siinä missä kuvaileva tilastotiede liittyy kaiken datan havainnollistamiseen, niin induktiivinen datan analyysi vaatii tietoa tuottaa tarkoitukseen tyypillinen otos, jonka avulla voidaan luoda yleistävä ja pätevä tulos tutkimuksesta (Kuva 1). Hyvänä esimerkkinä tästä erosta on vaalimielipidemittaukset: sen sijaan, että kysyttäisiin joka ikiseltä äänioikeuden omaavalta henkilöltä, niin kerätään joukko ihmisiä erilaisita taustoista ja tuotetaan tutkimus heidän mielipiteistään. (Cleff 2014, 3-4)

Data-analytiikkaan liittyy läheisesti Big Data, joka on tietoa jonka koko ylittää perinteisten tietokantajärjestelmien prosessointikyvyt (Morabito 2015, 8). Täten on tullut tarve kehittää erilaisia käsittelytapoja suurille datamäärille, joita on useita erilaisia kaupallisia, esimerkiksi MATLAB ja SAS, ja ei-kaupallisia tuotteita kuten R ja WEKA, havainnollistettu taulukossa 1. (Runkler 2012, 3).

Taulukko 1. Erilaisia tietojenkäsittelyohjelmia.

Tuote	Kuvaus	Maksullinen	www-osoite
MATLAB	Monikäyttöinen datankäsittelyohjelma, suljettu lähdekoodi	X	https://www.mathworks.com/products/matlab.html
SAS	Monia eri alaohjelmia eri tarkoituksiin, suljettu lähdekoodi	X	https://www.sas.com/fi_fi/home.html
R	Avoimen lähdekoodin tietojenkäsittelyohjelma		https://www.r-project.org/
WEKA	Koneoppimiseen käytetty tietojenkäsittelyohjelma. Avoin lähdekoodi		http://www.cs.waikato.ac.nz/ml/weka/

Big datan piirteisiin kuuluu täyttää neljän V:n sääntö, joista nämä kirjaimet ovat englanniksi volume (volyymi), velocity (nopeus), variety (moninaisuus) ja veracity/value (todenmukaisuus/arvo). Volyymi on käsitteistä helpoin ja selvin, sillä se viittaa vaan datan suureen määrään. Nopeus viittaa nopeuteen, jolla dataa tuotetaan ja vaihdetaan. Moninaisuus tarkoittaa erilaisia datan tyyppejä, joita tuotetaan, kuten esimerkiksi valuuttaa, tekstiä, numeroita ja päivämääriä. Todenmukaisuus tai arvo tarkoittaa sitä, että dataa tuottavat lähteet eivät ole kontrolloituja tai niiden laatua ei tarkasteta, joten on kyettävä toteamaan, mikä datasta on juuri kyseisessä tilanteessa pätevä ja kyetäänkö sieltä löytämään jotain arvokasta. (Shmueli et al. 2016, 7; Mohanty et al. 2015, 3-4)

Liiketoiminta-analytiikka (business analytics) on kvantitatiivisen datan tuottamista luettavassa muodossa päätöksenteon tueksi (Shmueli et al. 2016, 3). Seuraava taso liiketoiminta-analytiikassa on liiketoimintatiedon hallinta (business intelligence) joka tarkoittaa datan visualisointia, raportointia sekä pyrkimystä ymmärtämään mitä on tapahtunut sekä mitä tulee tapahtumaan (Shmueli et al. 2016, 3). Liiketoimintatiedon hallinta on yleisesti sateenvarjotermi, jota käytetään kuvaamaan tekniikoita ja prosesseja, joilla kerätään, varastoidaan ja analysoidaan dataa nimenomaan parempien päätösten tekemisen vuoksi eli liiketoiminta-analytiikkakin on liiketoimintatiedon hallinnan alalaji (Mach-Król et al. 2015, 90; Kobielus 2010, 2). Vaikka liiketoiminta-analytiikkaa ja liiketoimintatiedon hallintaa

määritelläänkin useissa eri lähteissä, niin yhteistä ja selkeää määritelmää ei näille ole vielä tehty, joten eri lähteestä riippuen niiden määritelmä voi poiketa tässä työssä esitetyistä.

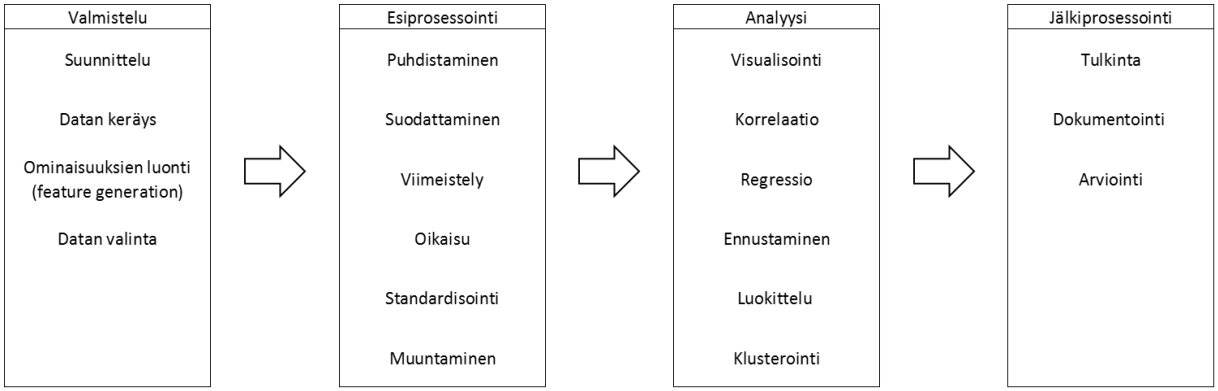
Käyttökelpoisen datan hankintaan on kehitetty termi, datan louhinta (data mining). Datan louhinta on yläkäsite kuvaamaan erilaisia tekniikoita, joilla käydään läpi suuria määriä dataa ja sen ideana on tunnistaa erilaisia kuvioita datajoukosta ja löytää sieltä hankkijalle jotakin mielenkiintoista (Lee 2013, 2). ”Jotakin mielenkiintoista” tarkoittaa tässä tapauksessa tietoa, joka on ihmiselle ymmärrettävää sekä käyttökelpoista (Runkler 2012, 2).

2.2 Yleistä data-analytiikasta

Data-analytiikassa käytettävä data on usein määrältään suurta ja se voi olla peräisin erilaisista lähteistä. Teollisuuden prosesseista saatavaa dataa hankitaan, varastoidaan ja käytetään jatkuvasti enevissä määrin. Datan kerryttämisellä ja käytöllä on pyrkimys auttaa automatisoimaan ja hallitsemaan koko tuotantoteollisuutta: tuotantoa, valmistusta, jakelua, logistiikkaa ja toimitusketjua. (Runkler 2012, 1)

Liiketoiminnasta tuleva data voi olla esimerkiksi peräisin rahoituksesta, myynneistä, markkinoista, markkinoinnista tai asiakkaista (Runkler 2012, 1). Tässä työssä keskitymme markkinoinnista saatavaan sekä siinä käytettävään dataan ja sen hyödyntämiseen.

Data-analytiikkaprosessi voidaan jakaa jaotella useammalla eri tavalla, mutta Runkler on jakanut ne neljään eri vaiheeseen. Ensimmäisessä vaiheessa vaiheessa valmistellaan suunnitellaan projekti, kerätään dataa, luodaan ominaisuudet (feature generation) ja valitaan haluttu data. Ominaisuuksien luomisella tarkoitetaan esim. muuttujien lisäämistä analyysien helpottamiseksi tai turhan datan karsimiseksi. Seuraavassa data vaiheessa puhdistetaan, suodatetaan, viimeistellään, oikaistaan, standardisoidaan ja muunnetaan haluttuun muotoon. Kolmannessa vaiheessa, analyysissä, dataa visuaalisoidaan, tutkitaan korrelaatioita ja regressioita, ennustetaan, luokitellaan ja klusteroidaan sen mukaan, mitä etsitään. Neljännessä prosessin vaiheessa, jälkiprosessoinnissa tulkitaan, dokumentoidaan ja arvioidaan tuloksia. (Runkler 2012, 2)



Kuva 2. Data-analytiikkaprosessin eri vaiheet (Runkler 2012, 3).

Kuva 2 havainnollistaa prosessia. Tämänkaltaisella data-analytiikkaprosessilla eli datan louhinnalla tuotettua tietoa voi käyttää erilaisiin käyttötarkoituksiin, esimerkiksi päätöksenteon tukemiseen tai jopa sen perusteeksi.

Datan louhinta on siis datajoukosta toistuvien kuvioiden ja sitä kautta tärkeän tiedon etsimistä suurista datamassoista. Datan louhinta kuuluu statistiikan ja koneoppimisen eli keinotekoisen älyn yhtymäkohtaan, koska se yhdistää suuren määrän dataa ja valtavasti laskuvoimaa, jolloin kone itsessään voi löytää ilman ihmisen ohjausta mielenkiintoista dataa (Shmueli et al. 2016, 6). Erityyppisistä datan lähteistä on eroavia prosesseja, joilla näistä saa dataa louhittua.

2.3 Tekstin louhinta

Tekstin louhinnalla on kaksi merkitystä: luokitella dokumentti johonkin luokkaan tai klusteroida samanlaisia dokumentteja tai kerätä tarkempaa tietoa itse dokumentista (Shmueli et al. 2016, 452). Tekstidata on myös riippuvaista kontekstistaan ja perinteisiä datan louhinnan menetelmiä on vaikea soveltaa suoraan tekstistä tärkeän informaation ulosmittaamiseen (Aggarwal 2015, 8). Jotta dokumenttia olisi helpompi analysoida, tulisi suorittaa muutoksia, jotta dokumentti saataisiin standardisoitua (Grossmann & Rinderle-Ma 2015, 296). Tämä stardandisointi tarkoittaa merkkijonojen (string) puhdistamista ja neljää muutosta, jotka ovat poisto-operaatiot, välilyöntien poisto ja suurten kirjainten poisto, pysäytyssanojen poistaminen ja turhien liitteiden poisto (Grossmann & Rinderle-Ma 2015, 296). Taulukossa 2 havainnollistetaan dokumentissä tehtäviä muutoksia sen standardoinnissa.

Taulukko 2. Dokumentin muutokset standardoimiseksi (Grossmann & Rinderle-Ma 2015, 296).

	Tekstissä tarkoittaa	Alkuperäinen teksti	Muokattu teksti
Poisto-operaatiot	Välimerkit ja erikoismerkit (esim. @, /) poistetaan	abc.kerho@osoite.fi	”abc kerho osoite fi”
Ensimmäisen vaiheen valkoisten tilojen ja suurten kirjainten poisto	Poistetaan suuret kirjamet ja poistetaan tarvittaessa tulleita välilyöntejä	Tieteellinen tutkimus Suomen Akatemian mielestä…	”tieteellinen tutkimus suomen akatemian mielestä…”
Pysäytyssanojen poisto	Sanat, jotka eivät lisää merkitystä poistetaan esim. artikkelit ja prepositiot	teemme auditiontia koska haluamme luoda lisää arvoa	”teemme auditointia haluamme luoda lisää arvoa”
Turhien liitteiden poisto	Etenkin englannissa turhat lopetukset, kuten ”ed” tai ”s” päätteet	tommy’s data warehouse was full	“tommy data warehouse was full”

Kun tekstidata on standardisoitu kannattaa se muuttaa numeroiksi, kuten binäärikoodiksi tai joksikin muuksi normitetuksi numerosarjaksi, jossa eri sanat saavat vastineensa numeroista, joilloin perinteisiä datan louhintakeinoja voi käyttää myös muutetun tekstidatan louhimiseen (Weiss et al. 2015, 3-4). Myös ilman tätä muutosta numeroiksi saa tekstiä analysoitua, mutta tämä analysointi rajoittuu pitkälti sanojen esiintyvyyden tutkimiseen ja niiden painotukseen (Grossmann & Rinderle-ma 2015, 313), mutta tästäkin ominaisuudesta on paljon hyötyä, sillä yksi tärkeimmistä asioista tekstin louhinnassa etsijällä on tunnistaa tärkeimpiä hakusanoja tulosten löytämiseksi (Aggarwal 2015, 429-430).

Tekstidatan louhinnasta saa parhaiten irti informaation luokittelua ja ennustamista (Weiss et al. 2015, 5). Yhden ainoan dokumentin analysointia kutsutaan luonnollisen kielen prosessoiniksi (natural language processing) ja tämä prosessi auttaa tunnistamaan paremmin luokiteltua tekstidataa (Shmueli et al. 2016, 462). Käytännön käyttötarkoituksena tekstin louhinnalla on esimerkiksi tunnistaa nettisivujen evästeistä arvostelua (luokittelu) ja näiden avulla parantaa palvelua. Ennustamisen esimerkkinä toimii hyvin tietyn käyttäjän kirjoittamien dokumenttien avulla analysointi, mistä tämä pitää ja mitä hän todennäköisesti klikkaa nettisivuilla ja mitä ei (Aggarwal 2015, 615).

2.4 Verkkolouhinta

Internet on mullistamassa tapaa, jolla yritykset kommunikoivat sekä toistensa että asiakkaidensa kanssa ja tuo uusia kanavia, joiden avulla kommunikointi onnistuu (Kohavi et al. 2002, 1-2). Tähän kun lisätään internetin kytkeytyminen melkein kaikkeen muuhunkin elämään, niin tämä on tuottanut uuden tärkeän informaatiolähteen organisaatioille ja yksityshenkilöille: internetkytkennäisen tutkimuksen, jossa datan hankintatapana on verkkolouhinta (Lappas 2008, 179). Verkkolouhinta tarkoittaakin datan louhinnan eri tekniikoiden soveltamista käytännöllisen tai arvokaan tiedon hankkimiseksi internetistä (Lappas 2008, 179-180).

Aggarwal (2015, 589-590) jakaa verkosta saatavan datan kahteen pääasialliseen tyyppiin, jotka ovat internetin sisällön data ja internetin käyttödata. Näistä sisällölliseen dataan kuuluvat dokumenteista saatava data sekä linkittyneisyydestä löytyvä data, joka on eri sivujen linkittyminen solmukohtien kanssa ja näiden välinen vuorovaikutus. Internetin käyttödataan taas kuuluvat internetin lokitallenteet, kuten evästeet ja muut käyttötiedot sekä internetissä tehdyt transaktiot, arviot ja käyttäjien palautteet. Data, jota näistä saadaan käyttöön on usein joko informaatio useista asioista ja niiden suhteista tai henkilökohtaista informaatiota käyttäjistä ja näiden mielipiteistä.

Varsinkin sosiaalinen media osana internetiä antaa verkkolouhinnan avulla paljon irti yksilöistä, heidän tarpeistaan, toiveistaan ja mieltymyksistään (Shmueli et al. 2016, 446). Verkkolouhinta yleensäkin kontribuoi monella tavalla organisaatioille ja se on usein yksi kilpailueduista, jos yritys kykenee tehokkaaseen verkkolouhintaan (Singh 2015, 3). Hyviä esimerkkejä, joihin verkkolouhintaa voi käyttää yrityksessä ovat mm. asiakashallinnan kehittäminen, varastojen ja logistiikan optimointi, virheiden huomaaminen ja ennaltaehkäisy sekä poikkeavaisuuksien estäminen (Singh, 3-7). Verkkolouhintaa voi käyttää myös muuhunkin kuin liiketoimintaan ja Lappas (2008, 184-190) listaakin käyttökohteiksi mm. sähköisen oppimisen, sähköiset julkiset palvelut, sähköisen demokratian kehityksen, helpdeskien parantamisen, sähköiset kirjastot ja turvallisuuden sekä rikostutkinnan kehittämisen.

2.5 Datan eri tyypit

Dataa tulee nykyään hyvin monesta erilaista lähteestä ja organisaatioden vaikeudeksi tulee tunnistaa ja hyödyntää heitä itseään kiinnostavat ja heihin liittyvät materiaalit (Morabito 2015, 83). Organisaatioiden olisikin hyvä tiedostaa se asia, että kun dataa kerätään ja tuotetaan jatkuvasti eri lähteistä, sitä myös analysoidaan eri tavalla ja sitä varastoidaan erilaisiin käyttötarkoituksiin, jolloin voidaan löytää arvokasta tietoa arvaamattomista tiedoista (Farah 2016, 1-2). Koska erilaisia lähteitä on paljon ja monet yritykset, julkiset toimijat, voittoa tavoittelemattomat organisaatiot, tieteellinen tutkimus sekä yksityishenkilöt jakavat dataa kokoajan keskenään, niin se vaikeuttaa alkuperäisen lähteen löytymistä (Sivarajah et al. 2016, 1). Jotta tätä dataa halutaan käyttää tehokkaasti hyväksi, olisi suotavaa kyetä tunnistamaan sen tyyppi ja lähde (Sivarajah et al. 2016, 1).

Aggarwalin mukaan yksi mielenkiintoisimmista asioista data-analytiikkaprosessissa on tarjolla olevien erilaisten datatyyppien määrä ja hän jakaa nämä datatyypit kahteen laajaan käsitteeseen: riippumattomaan ja riippuvaiseen dataan (Aggarwal 2015, 6). Riippumaton data on yksinkertaista tietoa, jolla on yksinään selvää merkitystä, eikä se muutu toisten ominaisuuksien astuessa kuvaan, kun taas riippuvainen data on usein yksinään vaikeasti selitettävissä ja se muuttuu suhteessa muihin muuttujiin (Aggarwal 2015, 6). Price ja Shanks tutkimuksessaan toteavatkin riippuvaisen datan määrän lisääntyvän koko ajan sekä sen alkuperän muuttuvan yhä monimuotoisemmaksi ja vaikeammaksi ymmärtää (Price & Shanks 2011, 2).

Tämä luokittelu yksinään on kuitenkin aivan liian suppea avaamaan datan eri tyyppejä ja siksi yhdistämme Morabiton ja Hashemin et al. tutkimusten tulokset ja kykenemme luomaan viisi eri tyyppiä, jotka tuottavat dataa: internet ja sosiaalinen media, koneet ja IoT, transaktiot, aistihavainnot ja biometriset tunnisteet ja ihmisen tuottama data (Morabito 2015, 85; Hashem et al. 2014, 4). Näistä juuri nyt pinnalla on vahvasti IoT-sovellukset ja niiden datan tuottamat mahdollisuudet elämisen helpottamisessa, kun niin kutsutut fiksut kaupungit voisivat jo itse ilman ihmistä prosessoida IoTin tuottamaa raakadataa ja esimerkiksi lyhetää työmatkan pituutta (Mohanty et al 2015, 2). Taulukossa 3 havainnollistetaan datan tyypit, niiden lähteet ja esimerkki datan tuottajasta ja käyttökohteesta.

Taulukko 3. Datan tyypit ja lähteet (Morabito 2015, 85-86; Hashem et al. 2014, 4-5).

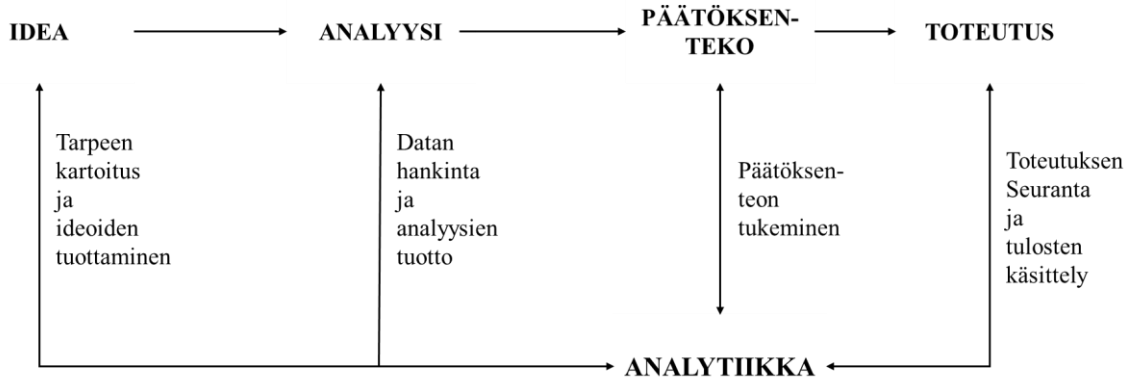
Lähde		Esimerkki lähteestä	Esimerkki käytöstä
Internet ja sosiaalinen media	Nettisivujen evästeet ja muut käyttötiedot	Nettikaupat, Twitter ja julkishallinnon tiedotuskanavat	Kohdennettu markkinointi
Koneet ja IoT (M2M)	Erilaisten systeemien välinen kommunikaatio	Paikkatieto ja RFID-tunnisteet	Ruuhkaraportit GPS-signaalien mukaan
Transaktiot	Vaihdanta ja liiketoimet	Yrityslaskut ja sopimusdokumentit	Yritysanalyysit, markkina-analyysit
Biometriset tunnisteet	Väestötiedot ja tunnistautuminen	DNA, kasvon piirteet ja sormenjäljet	Biometrinen lukon avaaminen esim. iirisskanneri
Ihmisen tuottama data	Erilainen ihmisen välisen kommunikoinnin tuottama data	Sähköpostit, kyselyt, ääniviestit ja tekstiviestit	Henkilökuvan rakentaminen, kohdennettu markkinointi

3 LIIKETOIMINTA-ANALYTIIKKA JA LIIKETOIMINTATIEDON HALLINTA

”Kuinka data-analytiikkaa voi käyttää liiketoiminnan tehostajana?”, on kysymys, johon moni yritys haluaisi selvän vastauksen. Provostin ja Fawcettin (2013, 19-20) mukaan helppo ja suora ohje tähän on muuttaa liiketoimintaongelma pienemmiksi tehtäviksi, joita lähdetään selvittämään tarkentamalla datan louhintaan liittyvät asiakehykset ja tämän jälkeen lähtemällä toteuttamaan data-analytiikkaprosessia. Data-analyysityökaluilla yritykset ulosmittaavat potentiaaliaan sisäisestä datasta ja monilla organisaatioilla tämä onkin yksi tärkeistä kilpailuedun lähteistä (Lee 2013, 2).

3.1 Liiketoiminta-analytiikka eli business analytics

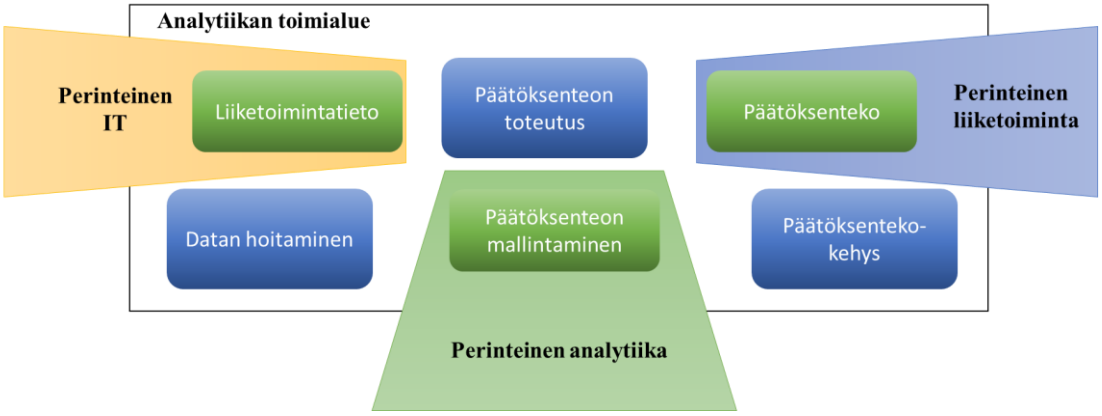
Liiketoiminta-analytiikka on siis datan muokkaamista päätöksentekoa varten. Sitä varten Saxena ja Srinivasan (2013, 1) ovat luoneet viitekehityksen, kuinka tulisi pyrkiä liittämään liiketoiminta-analytiikka pysyväksi osaksi päätöksentekoa ja miten analytiikka tukee päätöksentekoprojektin elinkaarta. Tätä on havainnollistettu kuvassa 3.



Kuva 3. Liiketoiminta-analytiikan soveltaminen projektissa. (Saxena & Srinivasan 2013, 85)

Saxena ja Srinivasan (2013, 6) jakavat liiketoiminta-analytiikan kuuteen osa-alueeseen, jotka sivuavat perinteisiä liiketoiminnan osa-alueita ja osittain linkittyvät myös liiketoimintatiedon hallintaan. Nämä kuusi osa-aluetta ovat liiketoimintatieto, datan hoitaminen, päätöksenteon

toteutus, päätöksenteon mallintaminen, päätöksentekokehys ja päätöksenteko (Saxena & Srinivasan 2013, 6). Kuvassa 4 esitetetään nämä kuusi osa-aluetta ja niiden nivoutuminen perinteisiin liiketoimintoihin, jotka ovat IT, analytiikka ja liiketoiminta.



Kuva 4. Liiketoiminta-analytiikan osa-alueet (Saxena & Srinivasan 2013, 6).

Liiketoimintatieto on sitä empiiristä tietoa, jonka yritys tuottaa pitkälti itse asiakkaiden kanssa vuorovaikutuksessa ja jota se käyttää päivittäisessä liiketoiminnassa (Grossmann & Rinderle-Ma 2015, 2-3). Liiketoimintatietoa tarvitaan, jotta liiketoiminta-analyysia voisi edes aloittaa ja tietoa tuottaa analysoitavaksi jokainen osasto yrityksessä. Tämä datan ulosmittaaminen ja sen analysointiin tarvittavien työkalujen tuottaminen ovat perinteisesti olleet IT-osaston toimenkuvaan kuuluvia (Saxena & Srinivasan 2013, 6).

Kobieluksen mukaan liiketoiminta-analytiikassa käytettävien ennustavan analytiikan työkalujen tulisi täyttää vähintään neljä eri käyttötarkoitusta. Nämä neljän käyttötarkoitusta ovat: käyttää suurta valikoimaa erilaisia datan valmistelutoimintoja, suorittaa hienostuneita tilastollisia mallinnuksia, mahdollistaa ennustettavien mallien ja tuloksien joustavaa visualisaatiota sekä hyödyntää useamman data-alustan mahdollisuuksia. Näiden ominaisuuksien avulla ammattilainen kykenee tuottamaan oikeaan muotoon muokattua informaatiota eteenpäin. (Kobielus et al. 2010, 2-3)

Ongelmaksi muodostuu kuitenkin se, että liiketoiminta-analytiikka tuottaa tietoa organisaatiolle, jolloin se vaatii raa'an datan lisäksi ymmärrystä organisaatiosta ja sen eri osa-alueista (Olszak 2015, 90). Tähän Kobielus (2010, 3-4) on tuonut ratkaisuksi viisi kyvykkyyttä,

jotka kunnollisen työkalun tulisi myös täyttää ja ne ovat: mallintaa useita liiketoimintaskenaarioita, sisällyttää useampia informaatiolähteitä malleihin, käyttää statistiikkaan perustuvia algoritmejä ja lähetymistapoja malleissa, käyttää useita mittareita mallien laadun valvonnassa sekä käyttää usean muuttujan löytö- ja arviontilähestymistapoja.

Datan hoitamista (Data stewardship) voi pitää kokoelmana datan hallintakeinoja, joilla varmistetaan saaminen, varastointi, koostaminen ja tarvittavat toimintamallit datan vapauttamiseen sekä käyttöön (Rosenbaum 2010, 3). Koska datapaketit voivat sisältää ongelmia, jotka voivat tehdä analyysistä harhaanjohtavaa tai täysin väärän, on datan hoitaminen onnistuneen analyysin kannalta erittäin tärkeää (Saxena & Srinivasan 2013, 101-102). Datan hoitamisesta vastaava henkilö on avainasemassa, sillä datapaketti usein muokkautuu tämän henkilön mukaan ja tämä henkilö usein pääsee käsiksi tietoihin, jotka ovat yksityisiä sekä luottamuksellisia, joka muodostaa varsinkin terveydenhuollon puolella yksitysyyskysymyksiä ja -ongelmia (Rosenbaum 2010, 7-8).

Päätöksenteon mallintaminen tähtää analyyttisen liiketoimintamallin luomiseen, joka kykenee tuottamaan tarkkoja vastauksia analyyttisten mallien pohjalta liiketoimintakysymysille (Grossman & Rinderle-Ma 2015, 17). Tämä toiminta on perinteisesti analytiikkaosaston työtä ja heidän vastuullaan on ollut tehdä mallista niin realistinen, että päätöksentekijä kykenevät oppimaan oikean elämän muuttujista mallin muuttujia tarkastelemalla ja manipuloimalla (Saxena & Srinivasan 2013, 31). Mallin luominen vaatii jonkin näkökulman sen luomiseen, kuten esimerkkinä asiakkaan näkökulma, jossa malliin valitaan tekijöitä, jotka houkuttavat asiakkaita ja tämän jälkeen lähdetään mallia luomaan näiden tekijöiden ja niiden välisten suhteiden ehdoilla (Grossmann & Rinderle-Ma 2015, 17-18; Saxena & Srinivasan 2013, 31). Hyvä malli voisi olla vaikka yrityksen pitämän tapahtuman vaikutusta tuotteen menekkiin, jolloin malliin on syötetty tietoja erilaisten potentiaalisten asiakkaiden käyttäytymisestä ja määrästä, jolloin erilaisen tapahtuman pitämisellä olisi erilainen vaikutus menekkiin (Williams 2016, 38).

Päätöksentekokehys on päätöksen kehyksen ja tarpeen määrittämistä, eli siinä pitää ymmärtää selvästi mikä on se konteksti ja päämäärä, jossa päätös tehdään ja mitä kykyjä organisaatiolla on toteuttaa päätös, jotta oikea kehys päätökselle saadaan luotua (Saxena & Srinivasan 2013,

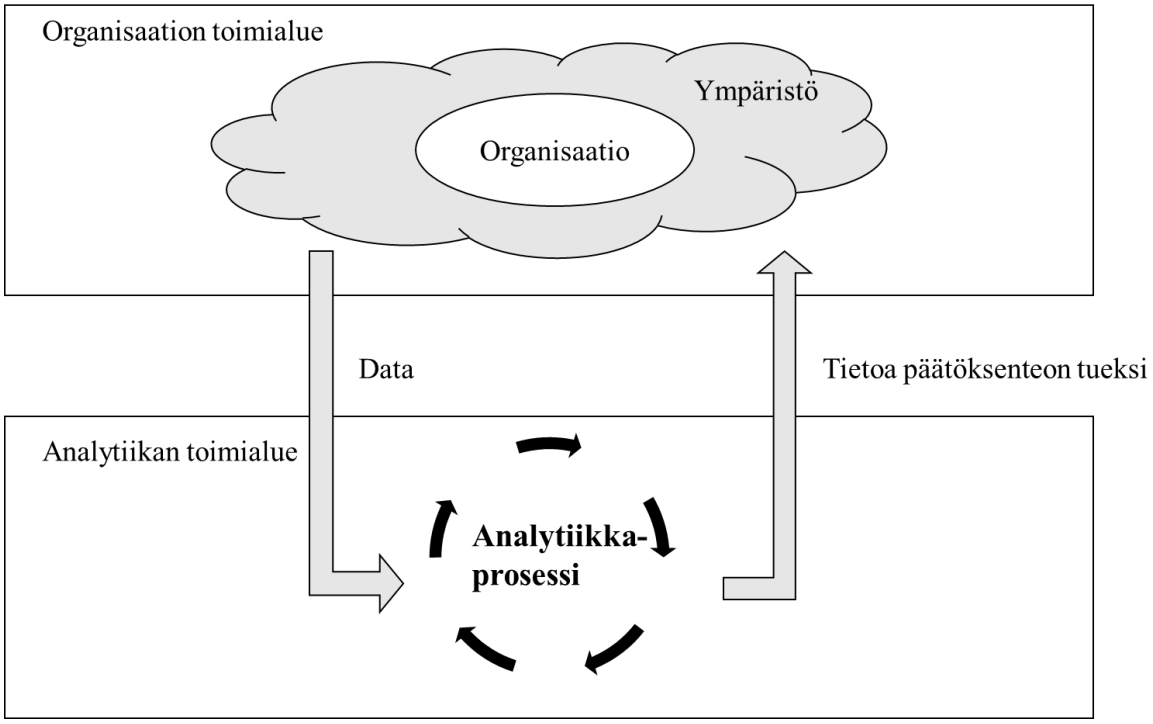
19-21). Tässä vaiheessa määritetään päätöksentekoympäristö, liiketoimintanäkökulma, analyyttisen maalit, datan arvotus ja osa datasta tunnistetaan, jotta kyetään luomaan datalle sopiva maali (Grossmann & Rinderle-Ma 2015, 15-16). Tämä kehys sitten ohjaa mallintajaa kaikissa vaiheissa, jotta mallinnuksesta saadaan toimiva ja heikkosti määritelty kehys voi mahdollisesti tehdä analyyseistä hyödyttömiä (Saxena & Srinivasan 2013, 19).

Päätöksenteko on päätöksen tekemistä mallin antamiin tuloksiin perustuen ja tämä osa ei enää ole puhtaasti liiketoiminta-analytiikkaa vaan päätöksenteko kuuluu analytiikan ja liiketoimintatiedon hallinnan välimaastoon, sillä vaikka päätöksen pohja usein pitkälti perustuu analytiikkaan, niin päätöksentekijä on kokenut ihminen organisaation puolelta, joka kytkee analytiikan tulokset omiin kokemuksiinsa liiketoiminnasta (Saxena & Srinivasan 2013, 6, 67-68). Tämä päätöksentekoprosessi alkaa siis analytiikan puolelta, mutta loppuu liiketoimintatiedon hallinnan puolelle (Delibasic et al. 2015, 10). Päätöksenteossa tärkeää on ymmärtää konteksti, jossa päätös tehdään, käsitellä juuri oikeaa ongelmaa, sytyttää keskustelu ja väitellä tärkeiden sidosryhmien välillä ja karsia pois epärationaalisuudet sekä poliittiset pyrkimykset, jotta päätöksestä saadaan organisaatiolle mahdollisimman tehokas (Saxena & Srinivasan 2013, 67-69).

Päätöksenteon toteutus taas tapahtuu liiketoimintatiedon hallinnan puolella, koska tässä vaiheessa päätös viedään selvästi organisaatioon, jossa tapahtuu toimia tämän päätöksen mukaan ja ne heijastuvat organisaatiosta ympäristöön (Saxena & Srinivasan 2013, 6-7, 79). Kuitenkin tässäkin osassa on analytiikalla tärkeä osuus nimenomaan tuottaa uutta informaatiota, sillä se seuraa ja raportoi mallin toimimista tosielämässä ja tätä kautta uuden datan avulla kyetään päivittämään mallia yhä tarkemmaksi ja tehokkaammaksi (Grossmann & Rinderle-Ma 2015, 20-22). Yhä kertyvän datan avulla näistä analyyttisista malleista voidaan saada hyvinkin toimivia ja tämä tehokas toiminta voi johtaa konkreettisiin markkinavoittoihin (Grossman & Rinderle-Ma 2015, 20-21; Leventhal & Langdell 2013, 7).

3.2 Liiketoimintatiedon hallinta eli business intelligence

Liiketoiminta-analytiikka muuttuu siinä vaiheessa liiketoimintatiedon hallinnaksi, kun analytiikan toimialue yhdistetään organisaation ja sen ympäristön toimialueeseen eli organisaatiolta ja sen ympäristöstä tuotettu tieto analysoinnin jälkeen palautetaan valmiina päätösehdotuksina tai tukimatereaaleina päätöksen tekijöille (Saxena & Srinivasan 2013, 6). Koska laskentateho on yhä halvempaa ja saatavilla, jokaista asiakasta analysoidaan yhä enemmän ja on olemassa yhä enemmän ohjelmistoja, joilla analytiikka voidaan tuottaa, niin organisaatiolle on yhä tärkeämpää harjoittaa tehokasta liiketoimintatiedon hallintaa kilpailukyvyn säilyttämiseksi (Leventhal & Langdell 2013, 1-3). Kuvassa 5 havainnollistetaan liiketoiminta-analytiikan nivoutumista liiketoimintatiedon hallintaan.



Kuva 5. Liiketoiminta-analytiikan nivoutuminen liiketoimintatiedon hallintaan (Saxena & Srinivasan 2013, 6).

Erona myös liiketoiminta-analytiikkaan on se, että liiketoimintatiedon hallinta ei enää itse luo uutta informaatiota, vaan se käyttää jo varastoitua ja vähintäänkin osin käsiteltyä informaatiota, joka luotiin jo analyysivaiheessa (Maisel & Cokins 2014, 1-2). Liiketoimintatiedon hallinnan tärkein asia onkin kehittää liiketoiminnallisia päätöksiä käyttämällä faktaan perustuvaa tietoa

ja systeemejä (Grossmann & Rinderle-Ma 2015, 1). Tästä seuraa se, että liiketoimintatiedon hallinnan tehokkuutta on vaikea mitata, sillä sen tulokset näkyvät suoraan organisaation tuloksissa, mutta jos organisaation tulos on odotettua parempi, niin silloin voidaan katsoa, että liiketoimintatietoa on hallitu oikein.

4 MARKKINOINTIANALYTIIKKA

Markkinointianalytiikassa liiketoiminta-analytiikka ja liiketoimintatiedon hallinta tuodaan markkinoinnin indikaattoreiden ja toiminnan pariin. Markkinointianalytiikan hyödyntäminen yrityksen sisällä onkin prosessi, jossa tunnistamalla oikeat mittarit markkinoinnin suoriutumisesta ja sen markkinointitavoitteen jahtaamisesta, seuraamalla ja analysoimalla näitä muuttujia sekä soveltamalla tulosta yhä uudelleen ja uudelleen saadaan aikaan tehokkaampaa markkinointia (Rackley 2015, 2-3). Usein tarkastelun keskipisteessä markkinoinnissa on kuluttaja, johon markkinoinnilla halutaan vaikuttaa ja tämän potentiaalisen arvon määrittäminen ja hankkiminen yrityksen käyttöön (Green 2012, 1). Koska markkinoiden tutkimukseen ja asiakkaiden käyttäytymiseen on 2000-luvun puolella tullut suuria harppauksia, on yritysten ymmärrettävä, että se data, jota he asiakkaista haluavat ei enää tule kyselyillä suoraan, vaan asiakkaiden käyttäymisen jättämistä jäljistä, jonka joutuu dataviidakosta itse hankkimaan (Hauser 2007, 3).

4.1 Mistä dataa saadaan markkinointiin ja mihin sitä voidaan käyttää?

Koska markkinointi on ihmisten sekä sosiaalisten tarpeiden tunnistamista ja kohtaamista, niin menestyvä yritys kykenee tehokkaasti vastaamaan näihin kahteen asiaan markkinoinnin saralla (Kotler et al. 2009, 45). Analytiikan avulla kyetään tunnistamaan erilaisia kuluttajia, joiden tarpeet ovat tietynlaisia ja heidän ostokäyttäytymistään pystytään ennustamaan, jolloin analytiikka on tunnistanut asiakkaan ja kertonut, kuinka ja milloin tämä kohdataan. Useilla toimijoilla on vieläkin hyvin rajoittunut kuva markkinointianalytiikan mahdollisuuksita, mutta sillä kykenee ymmärtämään syvemmin asiakasta, optimoimaan markkinointikanavia, ymmärtämään ja kehittämään markkinasisältöä, ymmärtämään brändin luomia mielikuvia sekä ennustamaan asiakkaiden käyttäytymistä (Rackley 2015, 36-37).

Rackley (2015, 95) listaa kolmeksi tärkeimmäksi lähteeksi internetistä saatavan datan, CRM-systeemit ja sosiaalisen median. Nykyään internetistä saatava datan määrä on valtava ja markkinoinnin kannalta esimerkiksi nettikaupoissa jokainen käyttäjä jättää jälkensä evästeiden muodossa (Nelson 2013, 3). Sosiaalinen media taas näyttää ihmisistä heidän tietämyksensä, kiinnostuksen kohteet, sosiaaliset verkostot ja usein myös perustiedot, kuten ikä ja sukupuoli (Nelson 2013, 3). CRM-systeemeihin taas saadaan dataa transaktioista, kyselyistä ja fyysisistä vierailuista esimerkiksi kaupassa sekä näihin systeemeihin tallennetaan internetistä ja sosiaalisesta mediasta tullut data (Rackley 2015, 40-41). Tärkeää on, että näistä eri lähteistä saadaan kerättyä niin kutsuttu asiakkaan DNA-profiili, joka tunnistaa ihmisiä internetin normaalien käyttäytymiskuvioiden, tuotekiinnostusten ja selailutavan perusteella (Hauser 2007, 11). Kun tähän vielä lisätään asiakkaan perustiedot, kuten ikä, sukupuoli, asuinpaikka sekä esimerkiksi koulutus ja mahdollinen jo asiakkaasta CRM-systeemiin kerätty data, niin kyetään jo hyvin tarkasti määrittelemään asiakkaan mieltymyksiä, käyttäytymistä ja tärkempänä tarpeita (Hauser 2007, 11).

Markkinointianalytiikan työkalut voidaan jakaa kahteen eri osa-alueeseen, jotka ovat strategiset työkalut ja operaatio- sekä logistiikkatyökalut. Strategisilla työkaluilla pureudutaan liiketoimintatietoon, asiakastietoon sekä asiakkaiden ostokäyttäytymiseen ja ennustavaan analyysiin, kun taas operaatio- ja logistiikkatyökaluilla hoidetaan, testataan ja optimoidaan verkkoläsnäoloa, usean kampanjan suoritusta ja maantieteellistä mallintamista. Näistä kahdesta strategiset työkalut ovat tässä työssä enemmän pinnalla, sillä ne liittyvät suoraan asiakkaaseen ja asiakkaan tuottamaan dataan ja ne ovat juuri niitä työkaluja, joilla erotutaan markkinoinnissa kilpailijoista asiakkaiden näkökulmasta katsottuna. (Rackley 2015, 108-109).

Kun asiakkaasta saatu data analysoidaan, niin sillä on useampi käyttökohde. Yleisesti useat yritykset käyttävät tätä informaatiota asiakkaista segmentoimalla näitä eri tavalla, jolloin tietylle segmentille osataan markkinoida juuri sellaisia asioita, joille on potentiaalista tarvetta (Green 2011, 8). Informaatiolla voidaan myös ennustaa asiakkaan elinikäistä arvoa yritykselle sekä asiakkaan taipumusta eri asioissa, kuten ostamisessa ja kiinnostumisessa, jolloin yritys kykenee oikeaan aikaan kontaktoimaan asiakasta oikealla asialla ja näin sitouttaa tämä yritykseen (Rackley 2015, 111). Asiakkaalle pystytään myös tuottamaan suosituksia yrityksen näkökulmasta, kuten suosittelemaan suurempaa tilauskokoa, myymään toisiaan tukevia

tuotteita tai seuraavaa hyvää ostosta asiakkaalle (Rackley 2015, 111). Taulukossa 4 tiivistetään markkinointidatan lähteet, käyttökohteet ja annetaan esimerkki datan hyödyntämisestä.

Taulukko 4. Markkinointidatan pääasialliset lähteet.

	Mitä saadaan?	Mihin käytetään?	Esimerkki hyödyntämisestä
Internet	Sivujen evästeet ja klikkaukset	Käyttäytymisprofiilin luominen: mitä on etsinyt, mitkä asiat kiinnostavat, potentiaaliset tarpeet	Asiakas on selaillut nettikaupassa kelloja, jolloin näitä markkinoidaan hänelle enemmän
Sosiaalinen media	Perustiedot, kiinnostuksen kohteet ja sosiaaliset verkostot	Henkilöprofiilin luominen: demografiset piirteet, mieltymykset, tuttavapiiri	Asiakas harrastaa vuorikiipeilyä, jolloin kiipeilytavaraa markkinoidaan hänelle enemmän
CRM-systeemit	Ostotapahtumat, reklamaatiot ja muu palaute	Mennyt käyttäytyminen: rahan käyttö ostoissa, ostofrekvenssi, antaako palautetta ja muun informaation varastointi	Asiakas käy kaupassa usein ja käyttää paljon rahaa kerralla, jolloin kauppa keskittyy markkinoinnillaan tehokkaammin asiakkaaseen

4.2 Segmentointi

McKinsey listaa viisi tärkeintä big dataan ja analytiikkaan liittyvää ilmiötä, joilla tuottaa lisää arvoa yritykselle ja yksi näistä on ihmismassan segmentointi kustomoinnin edistämiseksi (Mach-Król et al. 2015, 175). Analytiikan kehityksen myötä on kehittynyt termi massakustomointi, joka tarkoittaa sitä, että tehokkaalla analytiikan käytöllä yritykset kykenevät nyt tuottamaan suurelle joukolle erilaisia ihmisiä, juuri sitä, mitä itse kukin tarvitsee ja haluaa (Morabito 2015, 69). Koska asiakkaat jakavat yhä enemmän ja avoimemmin mielitpiteitään ja mieltymyksiään, niin on tuotteiden ja palveuiden oltava yhä enemmän räätälöityjä, jotta asiakkaiden kiinnostus säilyy (Morabito 2015, 6-7). Kun henkilöstä saatuun informaatioon lisätään tämän yhteydet sosiaalisessa mediassa ja näiden vaikutus henkilöön ja henkilön vaikutus omiin kontakteihin, niin saadaan luotua klustereita, joita analysoimalla kyetään

luomaan toisiaan sivuavia segmenttejä ja asettaa eri henkilöt yhä tarkempiin segmentteihin (Mohanty et al. 2015, 173).

Klusteroinnin avulla saadaan jaettua heterogeeninen joukko useaksi homogeeniseksi segmentiksi, kun käytetään edellä mainittuja tietoja hyväksi (Chen et al. 2015, 4). Näille homogeenisille segmenteille pystytään antamaan yhdistäviä piirteitä, kuten esimerkiksi, että segmentti 1 kuluttaa rahallisesti kerralla paljon ja luksustuotteisiin, kun taas segmentti 2 kuluttaa kerralla vähän ja peruselintarvikkeisiin. Tällöin yrityksen kannattaa keskittyä enemmän segmentin 1 asiakkaisiin, sillä pitkäaikaisen asiakassuhteen luominen näiden kanssa tuottaa enemmän arvoa yritykselle, unohtamatta palvella perustasolla myös segmenttiä 2. Tätä kutsutaan asiakkaan arvon määrittämiseksi klusteroinnin avulla (Rackley 2015, 111). Koska kuitenkin ollaan yritysmaailmassa, niin yritykset haluavat linkittää markkinoinnin tehokkuuden johonkin helposti mitattavaan muuttujaan ja Rackleyn (2015, 71) mukaan näistä paras on sijoitetun pääoman tuotto (ROI), koska se luo hyväksyntää yrityksen joka osastolla ja positiviinen ROI oikeuttaa ja helpottaa markkinointibudjetin tekoa.

Toinen yleinen tapa segmenttien luomiseksi klusteroinnin avulla on luoda tietyille tuotteille asiakassegmenttejä eli luoda erilaisia ryhmiä, joille kohdennetaan tietyn tuotteen tai palvelun markkinointia (Rackley 2015, 111). Tämä tapa luo arvoa yritykselle säästämällä turhia markkinointikustannuksia, kun asiakasprofiloinnin avulla kyetään sulkemaan pois tietyt segmentit tuotteelta. Koska kohdennettu markkinointi on huomattavasti tehokkaampaa, niin se heijastuu myös kyseisen markkinointiprojektin sijoitetun pääoman tuottoon. Asiakkaat kykenevät myös auttamaan yritystä itse, jos yrityksellä on tarjota alusta, jossa koneälyn avustuksella asiakas räätälöi itse tarvitsemansa tuotteensa ja palvelunsa, jolloin koneäly kykenee helpottamaan muutaman asiakkaan tiedoilla seuraavan asiakkaan tarpeita, jos niillä on yhtään yhteisiä piirteitä (Morabito 2015, 69).

4.3 Asiakkaan elinkaaren arvon ja ostamisen alttiuden ennustaminen

Asiakkaan elinkaaren arvon arviointi on tärkeää yrityksille siinä vaiheessa, kun lähdetään hankkimaan uusia asiakkaita tai sitouttamaan harvoin käyviä asiakkaita yritykseen tiiviimmin (Kotler et al. 2009, 394). Yrityksen kannattaa tehdä arviontia asiakkaan tuottamasta elinkaaren

arvosta verrattuna kuluihin, joita tulee uuden asiakkaan hankinnasta (Rackley 2015, 111). Asiakkaan DNA-profiilista tärkeimmät muuttujat tässä tapauksessa ovat kuluttamistottumukset, niistä eritoten kuinka usein ja kuinka paljon asiakas on tottunut kuluttamaan (Hauser 2007, 11; Rackley 2015, 111). Ongelmaksi tämän analysin tuottamisessa on siinä, että jos tutkittu potentiaalinen asiakas ei ole asioinut vielä yrityksen kanssa, niin kulutustietoa ja tottumuksia on hyvin vaikea saada. Tällöin dataa voi yrittää saada kolmannilta osapuolilta, mutta ongelmaksi voi muodostua datan sopivuus ja siinä esiintyvien puutteiden määrä (Green 2012, 8-11). Helpoin tapa onkin lähteä sitouttamaan jo muutaman kerran yrityksen palveluita käyttänyttä asiakasta tiiviimpään suhteeseen yrityksen kanssa, jolloin saataisiin koko asiakkaan potentiaalinen ostovoima hyödynnettyä (Bayer & Servan-Schreiber 2011, 9). Loppujen lopuksi, kun yritys kykenee sitouttamaan monta asiakasta, joilla on korkea kulutuselinkaaren arvo, niin tämä taas näkyy markkinointiprojektin positiivisena sijoitetun pääoman tuottona (Rackley 2015, 74).

Potentiaalisen asiakkaan ostamisen alttiuden ennustaminen on yrityksille hyödyllistä, sillä tällaisten ihmisten tunnistaminen helpottaa ensimmäisen ostopäätöksen markkinointia ja sitä kautta uuden asiakkaan sitouttamista yritykseen (Kotler et al. 2009, 380-381). Nämä asiakkaat ovat sellaisia, että he ovat alttiita esimerkiksi klikkaamaan linkkiä sähköpostimainoksessa tai sosiaalisessa mediassa kohdennetun mainonnan linkkiä tai ovat käyttäytymisen perusteella valmiita tekemään ostoksen (Rackley 2015, 111). Tällaista käyttäytymistä on esimerkiksi tietyn tyyppisten kenkien selailu, jolloin asiakas on etsinyt tiettyä merkkiä ja tiettyä väriä kengistä. Tässä antaa markkinoijalle tai algoritmille mahdollisuuden laukaista ostotapahtuma esimerkiksi erityisellä tarjouksella tai kohdistamalla tämän tietyn tuotteen aggressiivisempaa markkinointia mahdolliselle asiakkaalle (Rackley 2015, 111). Kun ensimmäinen ostotapahtuma on käynyt toteen on taas yritys saanut lisää dataa siitä, kuinka tämä tietty ja sen tyyppinen asiakas rationalisoi itselleen ostotapahtuman ja kykenee tämän avulla ennustamaan asiakkaan käyttäytymistä ja tarpeita paremmin.

4.4 Kohti tehokkaampaa markkinointia

Asiakkaan DNA-profiili siis kertoo paljon olennaista asiakkaasta yritykselle ja sen luomista voidaan pitää erittäin kannattavana jokaiselle, joka haluaa tehostaa markkinointiaan. Useat

yritykset, jotka ottavat vasta ensiaskeliaan data-analytiikan soveltamisessa markkinointiin nykyään ymmärtävät ne kolme tärkeintä asiakasdatan lähdettä. Kuitenkin yrityksille pitäisi yhä painottaa entistä enemmän sitä, kuinka paljon asiakkaiden DNA-profiilien luomisella ja analysoinnilla saadaan aikaan parempaa tulosta tehokkaan markkinoinnin takia. Tämä tarkoittaa konkreettisesti enemmän tuloja ja parempaa tulosta. Toki pienillä ja keskisuurilla yrityksillä voi olla resurssipuutteita kunnollisten analytiikkatiimien kokoamiseen, mutta konsultointiapukin todennäköisesti on erittäin hyvä vaihtoehto.

Heterogeenisessä ihmismassassa on aina hyvin paljon eroavaisuuksia eri ihmisten kesken, vaikka tämä otos olisi hankittu vain tietyltä alueelta tai esimerkiksi tietyltä koulutusalalta. Kuten Power (2013, 3) toteaa, niin itse datalla ei sinänsä ole arvoa, ennen kuin se tuodaan johonkin kontekstiin tai sovellutukseen. Siksi segmentointi on avainasemassa markkinointia suunniteltaessa, koska sen avulla kyetään nappaamaan juuri niitä ryhmiä tai niitä henkilöitä, joita yritys haluaa ja tarvitsee. Asiakkaan arvon määritys on toinen tärkeä muuttuja, sillä kun saadaan eri segmenttejä, niin nekin kannattaa arvottaa. Tähän on kaksi tehokasta tapaa, joko eri tuotesegmenttien sisältä etsiä arvokkaimpia asiakkaita ja keskittyä heihin tavanomaista enemmän tai sitten tuottaa segmenttejä asiakkaiden arvon perusteella ja keskittyä tavanomaista enemmän parhaimpiin segmentteihin. Molempien tapojen yhdenaikainen soveltaminen tuo tehokkaimman tavan luoda markkinoinnin avulla arvoa yritykselle.

Yhä kiristyvä kilpailu tulevaisuudessa vaatii lisää tapoja keinoja kilpailuetua muihin nähden ja tehokkuus markkinoinnissa on oiva tapa karsia kuluja. Markkinointianalytiikka kehittyy jatkuvasti kovaa vauhtia ja siihen on olemassa jo useita sovellutuksia, kuten ilmainen R-kieli, josta löytyy useita ohjeita markkinointianalytiikan haasteiden voittamiseen ja vaikka tehokkaan klusteroinnin ja luokittelun tekemiseen (Chapman & Feit 2015, 299). Tästä nopeasta kehityksestä voi tulevaisuudessa seurata markkinointianalytiikassa olevan tämänhetkisen trendin muutos, jossa se muuttuu kilpailuvaltista yhdeksi liiketoimminnan elintärkeistä toiminnoista, jota ilman markkinoinnilla ei ole mitään mahdollisuuksia pärjätä. Tulevaisuudessa valmiit ja pitkään kerätyt tietopankit asiakkaista voivat olla jo niin arvokkaita, että ihan näitä tietoja jakamalla voi pyörittää hyvinkin tuottoisaa ja arvokasta liiketoimintaa.

5 ESIMERKKITAPAUKSET

Olemme valikoineet työhön kaksi kappaletta esimerkkitapauksia siitä, miten analytiikkaa voi hyödyntää markkinoinnin tehokkuuden mittaamisessa sekä päätöksenteon apuvälineenä. Ensimmäinen esimerkkitapaus pureutuu segmentointiin datan avulla kun taas toisessa käytetään ennustamista promootion tehokkuuden mittaamiseen.

5.1 RFM-malliin perustuva asiakassegmentointi datan louhinnan avulla

Case-työ esittelee, miten on mahdollista segmentoida asiakkaat RFM-mallilla datan louhintaa käyttäen. RFM-malli on työkalu asiakassegmentointiin, jossa käytetään muuttujina seuraavia kolmea: Recency, Frequency ja Monetary. Vapaasti suomennettuna nämä ovat äskettäisyys, toistuvuus ja rahallinen arvo. RFM-mallissa pyritään jaottelemaan asiakkaat näiden muuttujien mukaan. Tutkimus on tehty vuonna 2012 Daqing Chenin, Sai Laing Sainin, sekä Kun Guon toimesta ja julkaistu Database Marketing & Customer Strategy Management-lehdessä. (Chen et al. 2012, 197)

Verkkokauppojen ero perinteiseen jälleenmyyntiin on huomattava. Jokaisen asiakkaan ostosprosessia ja muuta toimintaa voidaan seurata tarkasti ja välittömästi. Asiakkaan tilaukseen liittyy usein toimitus-ja laskutusosoite. Lisäksi asiakkailla on käyttäjätilit verkkokauppaan tarpeellisilla yhteys-ja maksutiedoilla. Näiden tietojen avulla verkkojälleenmyyjät pystyvät kohtelemaan asiakkaita yksilöinä ja rakentamaan asiakaskeskeistä liiketoimintatiedon hallintaa. (Chen et al. 2012, 197)

Asiaskeskeissä liiketoimintatiedonhallinnassa herää seuraavanlaisia kysymyksiä:

- Ketkä ovat arvokkaimpia ja vähiten arvokkaita asiakkaita yritykselle? Mitkä ovat heidän yleiset piirteensä?
- Mikä on asiakkaiden käyttäytymismalli? Mitä tuotteita asiakkaat ostavat yleensä yhdessä? Missä järjestyksessä ostotapahtumat tapahtuvat?
- Millainen on myyntikuvio esimerkiksi tuotteilla, alueilla, ja ajalla (viikottainen, kuukausittainen) ja niin edespäin. (Chen et al. 2012, 197)

Edeltävänkaltaisiin kysymyksiin saadaan tietoa datan louhinnan avulla. Datan louhinta on yleinen toimenpide verkkokauppaliiketoiminnassa, ja siihen liitetään usein hyväksi todettuja liiketoimintamittareita, kuten tässä työssä käytetty RFM-malli. Monella suurella verkkokauppayritykset, kuten Amazonilla, datan louhinta on jo yleinen toimenpide ja osa liiketoimintaprosessia luoden asiakaskeskeistä liiketoimintatiedon hallintaa ja tukien asiakaskeskeistä markkinointia. Suurten yritysten hyödyntäessä analytiikkaa, pienemmilta toimijoilta puuttuu tekninen osaaminen tähän. (Chen et al. 2012, 198)

Tutkimuksen tarkoituksena oli tutkia datan louhintatekniikkoja asiakaskeskeisessä liiketoimintatiedonhallinnassa verkkokaupan tapauksessa. Esimerkkinä toimiva verkkokauppayritys on tyypillinen: pieni ja suhteellinen uusi toimija, joka tietää analytiikan ja datan louhinnan tärkeyden nykypäivänä, mutta siltä puuttuu resursseja ja teknistä osaamista. Analyysin päätarkoituksena on auttaa yritystä ymmärtämään paremmin asiakkaitaan paremman yksilöllisellä markkinoinnilla. (Chen et al. 2012, 198)

5.1.1 Tutkimuksen lähtökohdat

Case-yritys on Iso-Britannialainen noin 80 hengen jälleenmyyntiyritys, joka toimii verkossa eikä sillä ole yhtäkään kivijalkaliikettä. Se on perustettu vuonna 1981, päätuotteinaan lahjat. Yritys siirtyi pelkkään verkkokauppaan vuonna 2010, ennen sitä myynti tapahtui katalogien kautta. Siitä lähtien yrityksellä on ollut vakaa määrä asiakkaita Iso-Britanniassa ja Europpassa ja se on tuottanut kahdessa vuodessa paljon dataa asiakkaistaan. (Chen et al. 2012, 198)

Tutkimuksessa käytetty data on ajalta 1.1.2011-31.12.2011. Datassa on 22190 validia kauppaa niistä 4381 validin postinumeron, datassa muuttuja on nimetty PostCode, kanssa. Yhteensä 406 830 kappaletta myytiin. Tästä voi laskea, että keskimääräinen ostos 2011 oli 18.3 esinettä per tilaus. Lisäksi yksittäinen asiakas tilasi verkkokaupalta tuotteita kahden kuukauden välein. Tilausten suuret kappalemäärät viittaa siihen, että tilaajat ovat useimmin organisaatioita kuin yksilöitä.

Postinumero (PostCode) on elintärkeä tieto verkkokaupan pitäjälle, koska sen avulla kykenee yksilöimään asiakkaat. Muita muuttujia asiakaskaupan aineistossa ovat tilausnumero (Invoice), tuotteen numero (StockCode), Kuvaus (Description), Määrä (Quantity), Hinta (Price), tilauspäivä (InvoiceDate), Osoiterivit 1 … 3 osoitteen osille (Addressline 1-3) sekä maa (Country). Tämä on havainnollistettu kuvassa 6. (Chen et al. 2012, 199)

Table 1: Variables in the customer transaction dataset (4381 instances)

Variable name	*Data type*	*Description; typical values and meanings*
Invoice	Nominal	Invoice number; a 6-digit integral number uniquely assigned to each transaction
StockCode	Nominal	Product (item) code; a 5-digit integral number uniquely assigned to each distinct product
Description	Nominal	Product (item) name; CARD I LOVE LONDON
Quantity	Numeric	The quantities of each product (item) per transaction
Price	Numeric	Product price per unit in sterling; £45.23
InvoiceDate	Numeric	The day and time when each transaction was generated; 31/05/2011 15:59
Address Line 1	Nominal	Delivery address line 1; 103 Borough Road
Address Line 2	Nominal	Delivery address line 2; Elephant and Castle
Address Line 3	Nominal	Delivery address line 3; London
PostCode	Nominal	Delivery address postcode, mainly for consumers from the UK; SE1 0AA
Country	Nominal	Delivery address country; England

Kuva 6. Tutkimuksessa käytettävät muuttujat. (Chen et al. 2012, 199)

5.1.2 Työtavat

Tutkimus perustui siis RFM-mallilla segmentointiin. Yrityksen asiakkaat ovat jaettu erilaisiin merkittäviin ryhmiin k-keskiarvoklusteroinnilla. K-keskiarvoklusterointi on eräs yleisimmistä tavoista muokata dataa (Celebi 2015, 8). Jokaisen asiakassegmentin pääpiirteet ovat selvästi määritelty. Analyysi on tuotettu asteittain, ja siihen on käytetty SAS Enterprise Guidea sekä SAS Enterprise Mineriä. (Chen et al. 2012, 198)

Tutkimuksen data-analytiikkaprosessi on oppikirjaesimerkki data-analytiikan työvaiheista. Tutkimus aloitettiin datan esikäsittelyllä, jonka työvaiheet olivat seuraavanlaiset:

1. Tärkeiden muuttujien valinta: työssä valittiin tilausnumero, tuotteen numero, määrä, hinta, tilauspäivä ja postinumero.
2. Muuttujan määrä (Quantity) luominen hinta-ja määrä-muuttujien kertolaskuna.
3. Muuttujan tilauspäivä muuttaminen erillisiksi muuttujiksi: päivämäärä (Date) ja kellonaika (Time).

4. Datan suodattaminen sellaisten rivien poistamiseksi, jossa ei ole muuttujaa postinumero tai postinumero ei ole Yhdistyneiden kuningaskuntien postinumero.
5. Datan lajittelu postinumeron mukaan ja kolmen RFM-mallin, tutkimuksen kannalta tärkeän muuttujienn äskettäisyys (kuukausissa), toistuvuus (postinumeroittain) ja rahallinen arvo luominen postinumeroittain. (Chen et al. 2012, 200)

Tämänkaltaisilla toimenpiteillä luotiin datajoukko tutkimusta varten. Lopulliseen datajoukkoon lisättiin muuttujat ensimmäinen ostos ja ostosten minimi, maksimi ja mediaani postinumeroittain. Lisäksi nimellinen ostaja-muuttuja lisättiin dataan joka kuvastaa postinumeroa. Lopulta datajoukko ladattiin SAS Enterprise Miner 6.2:een analyysia varten. (Chen et al. 2012, 200)

Tutkimuksessa käytetty analyysimenetelmä on siis RFM-malli-pohjainen klusterointianalyysi, jolla selvitetään ovatko asiakkaat segmentoitavissa postinumeroittain RFM-mallilla. K-keskiarvo klusterointi on herkkä suurille muutoksille datassa, kuten tässä tapauksessa jotkin ostokset olivat rahallisesti ja toistuvuudeltaan hyvin paljon eroavaisia normaalista datasta, joten ne eristettiin ja käsiteltiin erillisenä. Lisäksi RFM-mallissa on huomattavaa, että toistuvuus-, äskettäisyys-ja rahallisen arvo-muuttujien arvot poikkeavat suuruusluokaltaan huomattavasti toisistaan, joten ne tulee normalisoida ennen klusterianalyysiä. (Chen et al. 2012, 201)

Taulukko 5 on tiivistelmä klustereista, josta selviää tapausten jakautuminen klustereiden välillä. Huomattavaa on klusterin 3 osuus, sen käsittää lähes puolet asiakkaista.

Taulukko 5. Tapausten jakautuminen eri klustereihin. (Chen et al. 2012, 203)

Klusteri	Klusterin asiakaskoko	Prosenttiosuus
1	527	14,14 %
2	636	17,07 %
3	1748	46,91 %
4	627	16,83 %
5	188	5,05 %

5.1.3 Tutkimuksen tulokset ja pohdinta

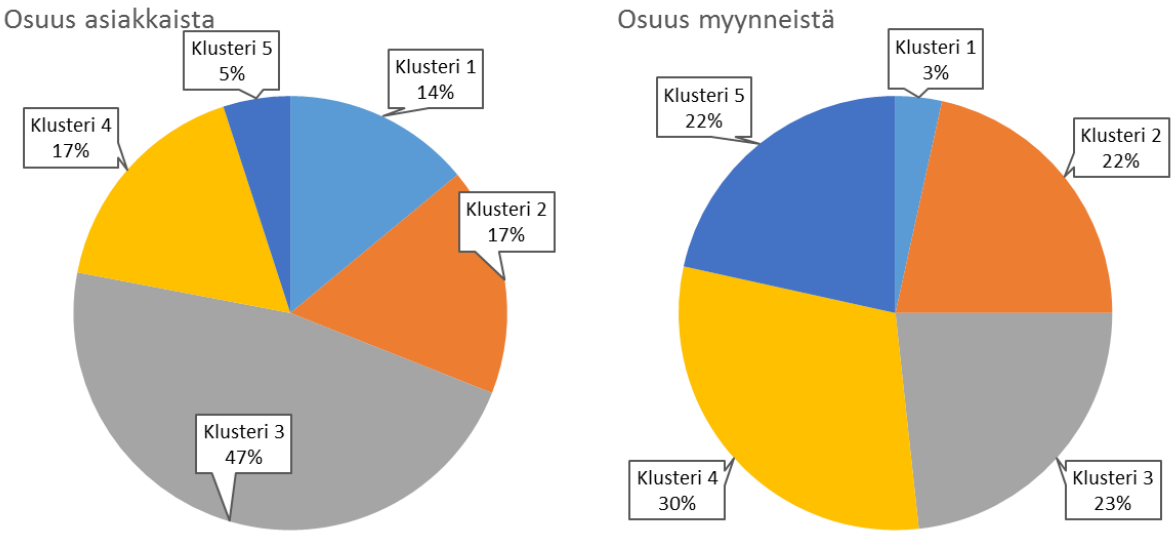
Analyysin avulla saatiin selvitettyä, mitkä asiakaskunnat ovat kaikista kannattavimpia asiakkaita ja ketkä huonoimpia asiakkaita. Tilastollisesti klusteri 1:n asiakkaat ovat huonoiten tuottavia asiakkaita ja heitä on 14.14% asiakaskunnasta. He ostavat harvoiten ja pienimmillä summilla. Tilastollisesti klusteri 5 sisältää parhaat asiakkaat. He ostivat useimmiten sekä suurimmilla summilla. Huolimatta siitä, että klusteri 5 käsitti vain 5.05% asiakkaista, he tuottivat 25.5% myynneistä. Toiseksi tuottavin ryhmä oli klusteri 4, joka oli käsitti 16,83% asiakkaista. He ostivat myös usein ja rahallisesti isolla summalla, mutta silti vähemmän kuin klusteri 5. (Chen et al. 2012, 205)

Klusteri 2 tuotti huomattavasti vähemmän kuin klusterit 4 ja 5, ja sen osuus oli 17,07% asiakkaista. Tämä ryhmä voidaan luokitella matalan äskettäisyyden, korkean toistuvuuden sekä keskiverron rahallisen arvon ryhmäksi. Klusteri 3 on suurin ryhmä, käsittäen lähes puolet asiakkaista. Sillä on järkevä toistuvuuden arvo. Verrattuna klusteriin 2 ja 4, tällä ryhmällä on matalampi mutta silti järkevä rahallinen arvo, ja tämän ryhmä sisältää paljon uusia asiakkaita jotka ovat alkaneet vastikään käyttämään tämän kaupan palveluja. Ryhmä edustaa keskivertokuluttajia ja sillä on täten tietty epävarmuuden taso tuottavuuden mielessä. Pitkällä

tähtäimellä nämä asiakkaat voivat potentiaalisesti olla joko erittäin korkean tai matalan tuottavuuden asiakkaita. Nämä ovat huomattavissa taulukosta 5. (Chen et al. 2012, 205)

Taulukko 6. Tilastoja klustereista. (Chen et al. 2012, 203)

	Minimi	Mediaani	Maksimi
Klusteri 1			
Äskettäisyys	8	9,8	12
Toistuvuus	1	1,3	4
Rahallinen arvo	3,75	361,2	7741,47
First_Purchase	8	11,1	12
Klusteri 2			
Äskettäisyys	4	5,4	7
Toistuvuus	1	2,3	13
Rahallinen arvo	15	586,19	3906,27
First_Purchase	4	7,7	12
Klusteri 3			
Äskettäisyys	0	1,5	3
Toistuvuus	1	2,6	7
Rahallinen arvo	20,8	658,71	4314,72
First_Purchase	0	5,3	12
Klusteri 4			
Äskettäisyys	0	1	5
Toistuvuus	3	8,3	16
Rahallinen arvo	191,17	2425,09	7330,8
First_Purchase	1	1	12
Klusteri 5			
Äskettäisyys	0	0,7	6
Toistuvuus	3	17,7	28
Rahallinen arvo	1641,48	5962,85	13110,02
First_Purchase	0	11,1	12



Kuva 7. Segmentit (vasen kaavio) sekä osuudet myynneistä (oikea kaavio). (Chen et al. 2012, 206)

Kuva 7 tiivistää työssä tehdyn analyysin. 47% asiakkaista ovat tavallisia kuluttajia järkevällä kulutuksella sekä toistuvuudella, 34% keskituottoisesta korkeatuottoisiin asiakkaita, 5% erittäin korkeatuottoisia sekä jäljelle jäävä 14 erittäin matalatuottoisia asiakkaita. Noin 22% asiakkaista tuottivat noin 60% kaikista myynneistä. Kaikenkaikkiaan case-yrityksen liiketoiminta vaikuttaa olevan terveellä pohjalla. (Chen et al. 2012, 206)

Yrityksen arvokkaimpien asiakkaiden tuottaessa yli 60% kokonaismyynneistä vuonna 2011, kun taas vähiten arvokkaat asiakkaat 4%, on jatkossa tärkeää löytää tuotteet, joita eri segmentit ostavat, minkä tuotteiden kanssa niitä ostetaan ja missä järjestyksessä. Yritys voi saada parempaa ymmärrystä kuluttajista tutkimalla segmenttejä ja niiden ostamia tuotteita. Tutkimusta voi tehdä esine- tai tuoteryhmätasolla. Lisäksi voisi olla hyödyllistä tutkia, onko kuluttaja tekemisissä parhaiden ja huonojen asiakasryhmien kanssa. Monet yrityksen asiakkaista ostavat tuotteita paljon kerralla, koska ne ovat organisaatioita. On syytä tutkia myös, millä tavalla organisaatiot ja yksityishenkilöt eroavat kulutukseltaan. Myöskin tutkimalla tuotteiden kausiluonteisuutta sekä kuluttajien maantieteellistä sijaintia on mahdollista saavuttaa yritykselle etua. (Chen et al. 2012, 206)

Kaikenkaikkiaan case-työ tarjoaa hyvä katsauksen siihen, miten asiakaskeskeistä liiketoimintatiedon hallintaa voi luoda datan louhintatekniikoilla verkkojälleenmyyjillä. Segmentoinnin avulla case-yrityksellä on mahdollisuus ymmärtää asiakasryhmien tuottavuutta ja täten käyttää sopivia markkinointistrategioita erilaisille kuluttajille. Analyysin pohjalta työn tekijät arvioivat, että datan louhinnassa kaksi tärkeää vaihetta vievät paljon aikaa: datan valmistelu sekä mallien tulkinta ja niiden arviointi. (Chen et al. 2012, 207)

5.2 Myynninedistämisen tehokkuuden mittaaminen kävijämäärien avulla.

Case-työ esittelee omanlaisensa lähestymistavan myynninedistämisen tehokkuuden mittaamiseen kivijalkaliikkeessä ja se on julkaistu 2016 Journal of Business Researchissa. Tutkimuksen ovat tehneet Epstein, Flores, Goodstein ja Milberg. Työ vertailee toteutuneita kävijämääriä kampanjan aikana ennustettuihin lukuihin ja täten antaa mahdollisuuden arvioida kampanjan tehokkuutta. (Epstein et al. 2016, 4394)

5.2.1 Työn lähtökohdat

Case-työn lähtökohtana on se, että myynninedistämisen tehokkuutta on hankala mitata. Hankaluus johtuu siitä, että on käytännössä erittäin hankalaa erottaa kävijämääristä ne kävijät, jotka tulisivat liikkeeseen ilman mitään myynninedistämistä sekä ne kävijät, jotka tulevat liikkeeseen vasta myynninedistämisen ansiosta. (Epstein et al. 2016, 4394)

Jälleenmyyjät mittaavat MPC-tehokkuutta keräämällä dataa. Tähän soveltuvaa dataa kerryttävät mm. kassakoneet. Kassakoneista saatava data on niinkutsuttua MPP-dataa eli marked point process, suomennettuna merkitty pisteprosessi. Pisteprosessidata on satunnaisia pisteitä tietyssä tilassa, ja merkittyyn pisteprosessiin kuuluu myös tapahtuman tyyppi (Jacobsen 2006, 3). Tässä tapauksessa pisteet ovat ostoksia ja tietty tila on kellonaika, jolloin ostokset tapahtumat. (Epstein et al. 2016, 4394)

Ostotapahtumien lisäksi dataa voi kerätä muillakin tavoilla. Case-työ antaa esimerkkinä Skillup-Chile-yrityksen, joka tarjoaa järjestelmän joka kerää liikkeen ovelle asennetun kameran

kuvasta pisteprosessidataa. Liikkeessä vierailijan muuttuminen maksavaksi asiakkaaksi tarkoittaa pisteprosessin muuttumista merkityksi pisteprosessiksi. (Epstein et al. 2016, 4394)

Perustavanlaatuinen ongelma MPC-tehokkuuden määrittelemisessä on vertailupohjan (baseline) puuttuminen täysin. Kaupan johdon investoidessa MPC:iin on mahdotonta tietää, mikä osa investoinnin tuotosta on juuri kampajan ansiota, ja case-työ pureutuu juuri tähän ongelmaan. (Epstein et al. 2016, 4395)

Tutkimuksessa käytettävä data on peräisin Skillup-Chile-yritykseltä, joka on erikoistunut analytiikkaan. Tutkimuksen datan taustalla oleva MPC on 20% alennus asiakkaille, jotka ovat paikallisen sanomalehden asiakkuusohjelman jäseniä. (Epstein et al. 2016, 4395)

5.2.2 Työtavat

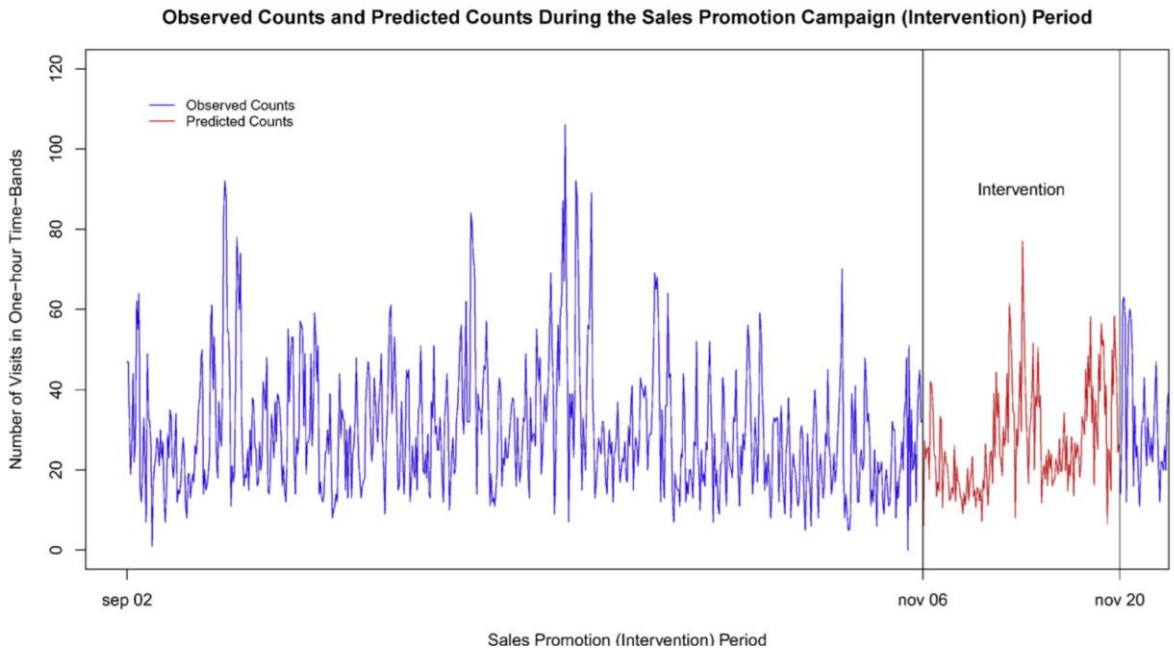
Tutkimuksessa käytettävä data on kävijämäärät tunneittain 2.7.2011-30.11.2011 ja MPC sijoittuu aikavälille 6.11.2011-20.11.2011. Työn lähtötilanteen ongelma on siis vertailudatan puuttuminen, eli paljonko kävijöitä olisi ilman kampanjaa. Tätä varten työssä on luotu vertailupohja aiempien toteutuneiden kävijämäärien pohjalta ennustamalla. (Epstein et al. 2016, 4395)

Työssä on käytetty monia erilaisia laskennallisia malleja. Visualisoinnissa on mallinnettu kävijämääriä ehdollisella Poisson-regressiomallilla havainnoista, jotka ovat tarkastelujakson ulkopuolelta. Mallien sovittamiseen, analyyseihin ja kaavioihin on käytetty R-pakettia. Mallia iteroitiin useamman kerran. (Epstein et al. 2016, 4395)

Vertailupohjan luomiseen Poisson-regression avulla on käytetty seuraavia arvoja:

1. Päivien aukiolotunteja (10-22)
2. Viikonpäiviä (m, tu, … , sa,su)
3. Viikkoja kuukaudessa (1, 2, … , 5)
4. Kuukautta (sep, oct, nov)
5. Indeksiä t tarkastelujaksojen päiville (1, ... , 90). (Epstein et al. 2016, 4395)

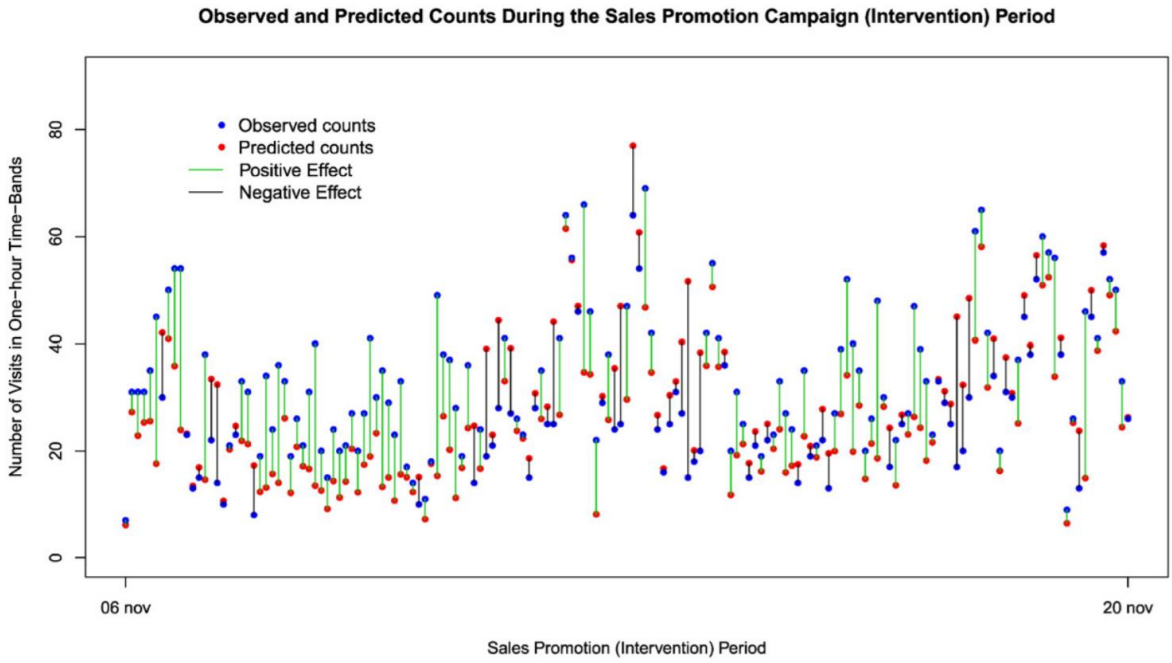
Ensimmäiset 4 ovat tekijöitä, kun taas laskennassa on käytetty t^2 kovarianttina muiden kuin lineaaristen trendien etsintää varten. Tälläisillä matemaattisilla operaatioilla on saatu Seuraavassa kuvassa kuvattua ennakoidut arvot tarkkailujaksolle eli kampanjan ajaksi. Kuvassa 8 punaisella ennustetut arvot, eli vertailupohja, sekä sinisellä toteutuneet muilta ajanjaksoilta kuin kampanjan ajankohdalta. Vertailupohjan luomisen jälkeen työssä testattiin autokorrelaatiota sekä laskettu ennusteita vähennettynä toteutuneella kävijämäärillä ja visualisoitu näitä tuloksia. (Epstein et al. 2016, 4395)



Kuva 8. Ennustetut ja toteutuneet kävijämäärät. Ennustettu ajanjakso välillä 6.11.2011-20.11.2011 (Epstein et al. 2016, 4397)

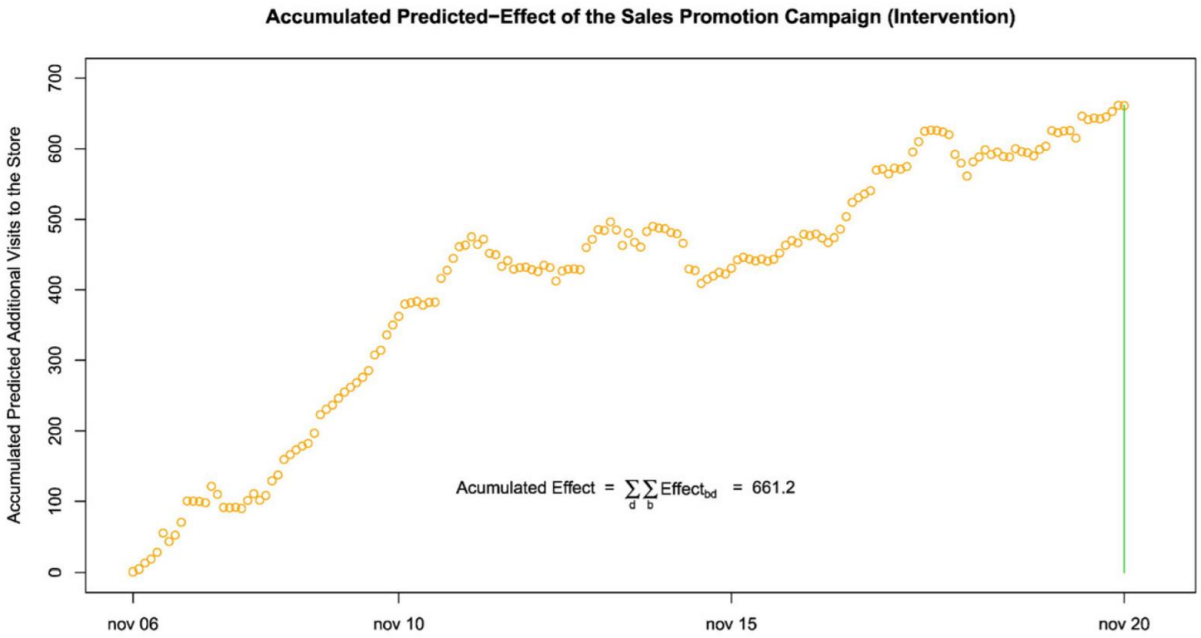
5.2.3 Tutkimuksen tulokset ja pohdinta

Työn tuloksia visualisoitiin monella eri tapaa. Tässä kandidaatintutkielmassa keskitytään vain siihen, kuinka hyvin ennustaminen onnistui, sekä milllaista tulosta kampanja tuotti laskennallisesti. Työn tuloksien analysoinnissa voidaan olettaa, että ennusteet ovat luotettavia. Tämänkaltaisessa projektissa, kuten kaikissa analytiikassa on erittäin tärkeää, että käytettävä data on luotettavaa sekä laskennallisten tulosten on oltava oikeallisia. Muuten vertailu tuottaa vääriä tuloksia.



Kuva 9. Kampanjan vaikutukset. (Epstein et al. 2016, 4398)

Kuvassa 9 on visualisoitu kampanjan vaikutusta kävijämääriin kampanjan aikana. Punaiset pisteet merkitsevät ennustettuja kävijämääriä ilman kampanjaa kun taas siniset toteutuneita kampanjan aikana. Vihreillä viivoilla pisteiden välissä on kuvattu kampanjan positiivista vaikutusta kun taas mustilla negatiivista vaikutusta. Kuvasta voidaan nähdä, että kampanjalla on ollut selvästi enemmän positiivista kuin negatiivista vaikutusta. (Epstein et al. 2016, 4398)



Kuva 10. Akkumuloituva kampanjan vaikutus (Epstein et al. 2016, 4400)

Kuvasta 10 näkee, että kampanjan ennustettu kumulatiivinen vaikutus on 661 ylimääräistä käyntiä aikavälillä 6.11.2011 ja 20.11.2011. Tämä on noin 44 käyntiä päivässä kampanjan ansiosta. Tämä on laskennallisesti todennäköisimmin oikeallisin tulos. (Epstein et al. 2016, 4401)

Kampanjan tehokkuuden laskennassa tulee ottaa huomioon normaalin saavutettavan kassavirran kasvamisen ja kampanjointikustannusten lisäksi myös analytiikasta syntyneet kustannukset.

5.3 Yhteenveto tutkimukista sekä pohdintaa

Case-esimerkit havainnollistivat erinomaisesti miten analytiikkaa sovelletaan oikeassa liike-elämässä, ja varsinkin millaisilla matemaattisilla malleilla ja työkaluilla case-töissä työskenneltiin. Tutkimukset esittelevät hyvin analytiikkaprosessin kulkua, sekä analytiikassa haettavaa lopputulosta: käyttökelpoista tietoa.

Varsinkin ensimmäisessä tutkimuksessa, asiakkaiden segmentoinnissa ostokäyttäytymisen perusteella, herää selkeitä huomioita miten analytiikka yrityksessä olisi mahdollista viedä pidemmälle. Konkreettinen esimerkki tähän on asiakkaan ostokäyttäytymisen selvittämistä ja sen analysointia (Chen et al. 2012, 206). Tämänkaltaisilla toimenpiteillä on mahdollista esimerkiksi verkkokaupassa saada asiakas tekemään heräteostoksia tarjoamalla ehdotuksina tuotteita, joita samankaltaiset ostajat ovat myös ostaneet.

Taulukko 7. Yhteenveto läpikäydyistä case-töistä.

Tutkimus	1.	2.
Tarkoitus	Segmentoida asiakkaat ostoskäyttäytymisen perusteella	Mitata myynninedistämiskampanjan tehokkuutta
Tutkimustapa	Datan käsittely käytettävään muotoon jonka jälkeen matemaattisen mallin soveltaminen. Tämän jälkeen tulosten tulkintaa.	Datan käsittely käytettävään muotoon jonka jälkeen matemaattisen mallin soveltaminen. Tämän jälkeen tulosten tulkintaa.
Matemaattinen malli	K-keskiarvoklusterointi	Poisson-regressio
Käytetty työkalu	SAS Enterprise Guide SAS Enterprise Miner	R-paketti
Riski	Segmentoinnin epätarkkuus	Liian pessimistinen arvio asiakkaista ilman myynninedistämiskampanjaa

Taulukossa 6 on yhteenveto kandidaatintutkielmassa esitellyistä case-töistä. Huomattavaa taulukossa on analytiikkaprosessin (tutkimustavan) samanlaisuus: molemmat ovat oppikirja-esimerkkejä analytiikan soveltamisesta. Käytetyt työkalut eroavat toisistaan SAS-tuotteiden ollessa maksullisia ja R-paketin maksuton.

Molemmissa case-töissä riskin lähde on sama: matemaattinen epätarkkuus. Laskenta voi olla joissakin tapauksissa virheellistä. Tämän takia data-analytiikkaa soveltavan henkilön tulee olla tietoinen oman toimialansa lainalaisuuksista ja täten pystyä tulkitsemaan tuloksia kriittisesti.

6 JOHTOPÄÄTÖKSET

Tutkimuksen tavoitteena oli selvittää, miten data-analytiikkaa voi käyttää hyödyksi markkinoinnissa. Tutkimus toteutettiin kirjallisuuskatsauksena ja tätä varten aineiston hankkimiseen kului runsaasti aikaa. Aineiston avulla data-analytiikan osa-alueita on yhdistetty markkinointiin. Tutkimus rajoittuu hyvin pitkälti data-analytiikkaprosessin analysointiosaan sekä teoriassa että case-esimerkeissä, joten se ei tarjoa kattavaa näkemystä kokonaisvaltaiseen data-analytiikkaan. Lisäksi tutkimusta rajoitti se, että itsellämme ei ollut resursseja suorittaa konkreettista tutkimusta, jollaisia oli case-esimerkeissä tehty. Tutkimuskysymykset tiivistävät työn tavoitteet:

1. Mitä data-analytiikka on markkinoinnissa?

Markkinoinnissa on kohteena asiakas ja yrityksen tavoitteena markkinoinnin avulla täyttää asiakkaan tarve tai luoda asiakkaalle tarve, jonka organisaatio kykenee täyttämään. Data-analytiikka tulee siinä vaiheessa mukaan, kun lähdetään miettimään tehokkaampia tapoja kohdata asiakas ja luomaan tämän avulla lisää arvoa yritykselle. Erilaiset potentiaaliset asiakkaat tuottavat massoittain dataa, jota keräämällä ja käsittelemällä yritykset kykenevät tehostamaan huomattavasti markkinointiprosessiaan. Tehokas markkinointiprosessi taas tulee näkymään suurempana sijoitetun pääoman tuottona, joka antaa markkinonnille lisää tilaa ja luottamusta operoida.

Konkreettisesti markkinoinnissa data-analytiikka tarjoaa markkinoinnin parissa toimiville henkilöille käsiteltyä dataa, jonka avulla markkinoinnin kohdentaminen eri segmentteihin ja henkilöihin on helpompaa. Pääasiassa klusteroinnin, käyttäytymisen mallintamisen ja klikkauksien mallintamisen keinoin markkinointianalytiikka tuottaa muuttujia, joiden avulla erilaisten piirteiden ja tarpeiden tunnistaminen on mahdollista ja tehokasta. Markkinointianalytiikka käyttää datan tyyppeinä transaktiodataa, ihmisten tuottamaa dataa sekä internetin ja sosiaalisen median dataa. Datan pääasiallisia lähteita taas ovat nettikauppojen, internetsivujen ja sosiaalisen median evästeet ja käyttötiedot sekä CRM-systeemit, johon on tallennettu tiedot asiakkaiden ostoksista, käynneistä ja kaikesta kyselyistä sekä asiakaspalautteesta. Tärkeimmät datan louhintatyypit markkinointianalytiikan puolelta ovat

teksti- ja verkkolouhinta, sillä data on usein tekstityyppistä tai internetistä. Suurena ongelmana organisaatiolle voi tulla oikeanlaisen datan hankkiminen ja sen varsinkin sen hoitaminen, sillä kokonaisuuden kannalta täysin hyödytöntä dataa on valtavasti joka lähteessä ja harhaanjohtavia yksittäisasioita ja -tapauksia, jotka voivat ilman datan kunnollista hoitamista pilata täysin analyysin ja tehdä sen tuloksista harhaanjohtavia ja tuhoisia yritykselle.

Toinen työn alussa määritelty tutkimuskysyms oli seuraavanlainen:

2. Kuinka data-analytiikkaa voi hyödyntää markkinoinnissa?

Toista tutkimuskysymystä, kuinka data-analytiikkaa voi hyödyntää markkinoissa, lähestyimme ensiksi teorian kautta. Tämän jälkeen tutkimme oikean elämän esimerkkitapauksia, joista ensimmäinen käsitteli segmentointia data-analytiikan avulla. Työstä selvisi, että on mahdollista yksilöidä kuluttajia ostotietojen perusteella. Yksilöityjä kuluttajia klusteroitiin segmentteihin data-analytiikan avulla hyödyntäen RFM-mallia, jossa ostotapahtumien äskettäisyys, toistuvuus ja rahallinen arvo olivat mitattavia suureita. Tämänkaltaisesta segmentoinnista selvisi, että todella on olemassa erilaisia asiakasryhmiä ja että segmenttejä pystyy jaotella vielä tarkemmin paremman informaation saamiseksi. Yritys voi käyttää tätä tietoa käyttääkseen erilaisia markkinointistrategioita eri asiakasryhmille, tuoden kustannussäästöjä vähentäen turhaa markkinointia. Työssä huomattiin myös, että analytiikkaa voi ja on jopa kannattavaa viedä pidemmälle selvittääkseen kuluttajien käyttäytymistä tarkemmin, esimerkiksi mitä tuotteita selataan ja missä järjestyksessä liitettynä tietoon minkä asiakasryhmän asiakkaan toimesta.

Toisessa työssä keskityttiin siihen, miten MPC-tehokkuutta eli markkinointi-ja promootiokampanjan tehokkuutta pystyy mittaamaan analytiikan työkaluin. Työssä luotiin olemassaolevasta datasta analytiikan työkaluin vertailupohja, joka kuvastaa asiakaskäyntejä ilman markkinointikampanjaa. Vertailemalla toteutuneita käyntejä vertailupohjaan pystyttiin mittaamaan markkinointi-ja promootiokampanjan tehokkuutta. Tämänkaltainen toiminta antaa hyötyä päätöksentekijälle esimerkiksi tilanteessa, jossa harkitaan markkinointi-ja promootiokanpanjan aloittamista. Päätöksentekijä pystyy tutkia esimerkiksi aiempien

kampanjoiden kuluja verrattuna tuottoihin ja täten saa osviittaa siitä, onko aiemmankaltaisen kampanjan aloittaminen kannattavaa vai ei.

Käytännön esimerkkien kautta huomattiin, kuinka hyvää lisäarvoa analytiikka voi tuottaa yrityksille. Sen soveltamiskohteet ovat käytännössä rajattomat. Lisäksi huomattavaa oli se, että analytiikan työkalujen käyttäminen vaatii huomattavaa matemaattista sekä tietoteknistä osaamista. Tutkimus pureutui yleiselle tasolle markkinointianalytiikassa ja suurten hyötyjen saavuttamiseksi pitäisi pureutua tarkemmin analyyttiseen mallintamiseen, jotta kyettäisiin luomaan näitä dataan pohjautuvia malleja. Myös tärkeimpiä markkinoinnin mittareita olisi tutkittava tarkemmin, jotta kyettäisiin tapauskohtaisesti poimimaan näistä tärkeimmät ja oleellisimmat erilaisten organisaatioiden ja tapahtumien osalta.

7 YHTEENVETO

Data-analytiikka on noussut yhä enemmän pinnalle liike-elämässä ja sen käyttö on viime vuosien aikana normalisoitunut yrityksissä eri osa-alueilla. Nykyään tuotettavan datan määrä on valtavaa ja sen hyödyntäminen on pitkään ollut haaste varsinkin pienemmille ja keskisuurille toimijoille. Kehittyneet työkalut ja ymmärrys data-analytiikan mahdollisuuksista ovat tuoneet kuitenkin data-analytiikkaa yhä enemmän ja enemmän myös muiden kuin suurten yritysten ja kehittyneiden analyyttisten toimijoiden ulottuville sekä käyttöön. Liiketoiminnan osa-alueista analytiikkaa on hyödynnetty pisimpään rahoituksen puolella, mutta se on levinnyt nyt myös muihin toimintoihin, kuten asiakaspalveluun, laskentatoimeen ja markkinointiin.

Data-analytiikka on tietokonesovellusten käyttöä suurten datakokoelmien analysointiin ja se on sekoitus tilastotiedettä, koneoppimista ja operaatioanalyysiä. Moderni data-analytiikka voidaan jakaa useampaan kokonaisuuteen, joihin kuuluvat myös big data -analyysit, liiketoiminta-analytiikka, induktiivinen tilastotiede ja liiketoimintatiedon hallinta. Datan louhinnalla tarkoitetaan käyttökelpoisen datan hankkimista tietomassoista. Data-analytiikkaprosessiin kuuluu neljä vaihetta, jotka ovat valmistelu, esiprosessointi, analysointi ja jälkiprosessointi. Tärkeimmät datan louhintatyypit liittyen markkinointianalytiikkaan ovat tekstin louhinta ja verkkolouhinta. Datan lähteitä tunnistetaan viisi kappaletta ja ne ovat internet ja sosiaalinen media, koneet ja IoT, transaktiot, biometriset tunnisteet ja ihmisen tuottama data.

Liiketoiminta-analytiikka on datan muokkaamista päätöksentekoa ja sen tukemista varten. Päätöksenteon prosessiin kuuluu idea, analyysi, päätöksenteko sekä toteutus ja liiketoiminta-analytiikka tukee jokaista prosessin vaihetta eri tavoin. Liiketoiminta-analytiikasta on tunnistettavissa kuusi aluetta, jotka nivoutuvat liiketoimintaan, analytiikkaan ja informaatioteknologiaan. Nämä kuusi aluetta ovat: liiketoimintatieto, päätöksentekokehys, datan hoitaminen, päätöksenteon mallintaminen, päätöksenteko ja päätöksenteon toteutus. Liiketoiminta-analytiikka muuttuu liiketoimintatiedon hallinnaksi, kun kaikki siellä tehdyt toiminnot tuodaan vuorovaikutukseen organisaation ja sen ympäristön kanssa. Ympäristö ja organisaatio siis tuottavat analytiikkaprosessiin karkeasti dataa ja analytiikkaprosessi palauttaa organisaatiolle tietoa.

Markkinointianalytiikka on liiketoimintatiedon hallinnan sekä liiketoiminta-analytiikan soveltamista markkinoinnissa. Silloin siis mallinnettavat asiat liittyvät markkinoinnin muuttujiin ja indikaattoreihin. Markkinointianalytiikan kolme tärkeintä datan lähdettä ovat internet, sosiaalinen media ja CRM-systeemit, joiden avulla saadaan luotua asiakkaan DNA-profiili, joka sisältää informaatiota asiakkaasta. Tämän informaation avulla organisaatio kykenee tehostamaan omaa markkinointiprosessiaan ja näin luomaan lisää arvoa itselleen. Pääasialliset menetelmät markkinoinnissa ovat segmentointi ja asiakkaan arvon määritteleminen, joita yhdistämällä saadaan luotua kannattavia asiakassegmenttejä, joihin organisaation kannattaa keskittyä. Tulevaisuudessa markkinointianalytiikalla tulee olemaan yhä suurempi painoarvo, sillä hakuammunnalla suoritettu markkinointi on valitettavan tehotonta verrattuna analytiikan avulla toteutettuun, jolloin se voi osoittautua jopa tappiolliseksi.

Data-analytiikka on markkinoinnissa siis tapa tuoda myynteihin, asiakkaisiin, mainoskampanjoihin ja kaikkiin muihin markkinoinnin osa-alueisiin selkeä arvotus, jonka avulla markkinoinnin parissa työskentelevät ihmiset tietävät, että mikä toimii ja mikä ei. Se tarjoaa oikeiden mittareiden ja mallintamisen avulla konkreettisia päätöksentekoa tukevia argumenttejä, jotka helpottavat nimenomaan oikeiden päätösten tekoa. Data-analytiikkaa hyödynnetään markkinoinnissa erilaisten ihmisten segmentoinnissa, ihmisten elinikäisen arvon arvottamisessa ja ostopäätösten tekemisen herkkyydessä. Tärkeää tässä oli se, että ihmisiä yksilöimällä ja sitten löytämällä yhtäläisyyksiä kyetiin tehostamaan markkinointia ja tuottamaan konkreettista arvoa yritykselle.

8 LÄHTEET

Aggarwal, C. (2015), Data mining. Cham: Springer, s.8-615.

Bayer, J. & Servan-Schreiber, E. (2011), Gaining competitive advantage through the analysis of customers' social networks. Journal of Direct, Data and Digital Marketing Practice, 13(2), s.106-118.

Celebi, M. (2015), Partitional clustering algorithms. Heidelberg: Springer, s.8.

Chapman, C. & Feit, E. (2015), R for marketing research and analytics. Springer International Publishing Switzerland, s. 299.

Chen, D., Sain, S. & Guo, K (2012), Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. Database Marketing & Customer Strategy Management, 19(3), s.197-208.

Cleff, T. (2014), Exploratory data analysis in business and economics. Springer, s. 3-4.

Delibašić, B., Hernández, J., Papathanasiou, J., Dargam, F., Zaraté, P., Ribeiro, R., Liu, S. ja Linden, I. (2015), Decision support systems v - big data analytics for decision making. New York: Springer, s. 10.

Epstein, L., Flores, A., Goodstein, R. & Millberg, S. (2016), A new approach to measuring retail promotion effectiveness: A case of store traffic. Journal of Business Research, [verkkolähde] 69(10), s.4394-4401. Saatavilla: http://dx.doi.org/10.1016/j.jbusres.2016.03.062 [Viitattu 20.3.2017].

Gandomi, A. & Haider, M. (2015), Beyond the hype: big data concepts, methods, and analytics. International Journal of Information Management, 35(2), s. 1.

García Márquez, F. & Lev, B. (2015), Advanced business analytics. Springer, s. 93

Green, A. (2012), Understanding the value of customer data.
Journal of Direct, Data and Digital Marketing Practice, 13(3), s. 1-11.

Grossmann, W & Rinderle-Ma, S. (2015), Fundamentals of business intelligence.
Berlin, Heidelberg: Springer Berlin Heidelberg, pp.1-313.

Hashem, I., Yaqoob, I., Anuar, N., Mokhtar, S., Gani, A. & Ullah Khan, S. (2015), The rise of "big data" on cloud computing: review and open research issues. Information Systems, 47, s. 98-115.

Hauser, W. (2007), Marketing analytics: the evolution of marketing research in the twenty-first century. Direct Marketing: An International Journal, Vol. 1 Iss 1 s. 38 - 54

Jacobsen, M. (2006), Point process theory and applications. 1st ed. Boston, Mass.: Birkhäuser, s .3.

Kobielus, J. (2010), The Forrester Wave™: predictive analytics and data mining solutions. For Business Process & Applications Professionals, Q1, s.3.

Kohavi, R., Masand, B., Spiliopoulou, M. & Srivastava J. (2002) Web mining. Kluwer Academic Publishers, Iss 6, s. 5-8.

Kotler, P. (2009), Marketing management. 1st ed. Essex (Gran Bretaña): Pearson Education, s. 45-394.

Lappas, G. (2008), An overview of web mining in societal benefit areas. Online Information Review. Vol. 32, s.179-195.

Lee, P. (2013), Use of data mining in business analytics to support business competitiveness. Review of Business Information Systems (RBIS), 17(2), s. 2.

Leventhal, B. & Langdell, S. (2013), Adding value to business applications with embedded advanced analytics. Journal of Marketing Analytics, 1(2), s. 1-7.

Liberatore, M. J. & Luo, W. (2010), the analytics movement: implications for operations research. Interfaces. Vol. 50, nro. 4, s. 2

Mach-Król, M., M.Olszak, C. & Pełech-Pilichowski, T. (2015), Advances in ICT for business, industry and public sector. Cham: Springer International Publishing. s. 175

Maisel, L. & Cokins, G. (2014), Predictive business analytics. Hoboken: J. Wiley, s.1-2.

Morabito, V. (2015), Big data and analytics. 1st ed. Springer, s.3-200.

Nelson, R. (2013), Measure your marketing. American bar association, Family Advocates iss. 3 vol. 35, s.3.

Power, D. (2013), Using big data for analytics and decision support. MWAIS 2013 proceedings. s. 1-5

Price, R. & Shanks, G. (2011), The impact of data quality tags on decision making outcomes and process. Journal of the Association for Information Systems, 12(4), s. 2.

Provost, F. & Fawcett, T. (2013), Data science for business. Sebastopol, CA: O'Reilly, s. 19-20.

Rackley, J. (2015). Marketing analytics roadmap. Berkeley, CA: Apress. s. 1-111.

Runkler, T. (2012), Data analytics. Wiesbaden: Springer Vieweg , s. 1-3.

Saxena, R. & Srinivasan, A. (2013), Business analytics. Springer, s. 1-85.

Singh, J. (2015), Business intelligence: achieving fineness through data, text and web mining. International Journal of Computer Applications, Vol. 128, s. 1-7.

Sivarajah, U., Kamal, M., Irani, Z. & Weerakkody, V. (2017), Critical analysis of Big Data challenges and analytical methods. Journal of Business Research, 70, s. 263-286.

Shmueli, G., Bruce, P. & Patel, N. (2016), data mining for business analytics: concepts, techniques, and applications. John Wiley & Sons, s. 3

Weiss, S., Indurkhya, N. & Zhang, T. (2015), Fundamentals of predictive text mining. London: Springer London. s. 3-5.