

POWER BI OPINTOJEN KESKEYTTÄMISEN EHKÄISYSSÄ

Case: Valkeakosken seudun koulutuskuntayhtymä



Ammattikorkeakoulututkinnon opinnäytetyö

Hämeenlinnan korkeakoulukeskus
Tietojenkäsittelyn koulutusohjelma

syksy, 2019

Anu Halttunen

Tietojenkäsittelyn koulutusohjelma
Hämeenlinnan korkeakoulukeskus

Tekijä	Anu Halttunen	Vuosi 2019
Työn nimi	Power BI opintojen keskeyttämisen ehkäisyssä Case: Valkeakosken seudun koulutuskuntayhtymä	
Työn ohjaaja	Lasse Seppänen	

TIIVISTELMÄ

Opinnäytetyön tavoitteena on selvittää, mitä mahdollisuuksia Microsoftin Power BI-ohjelmisto tarjoaa toimeksiantajan tiedolla johtamisen prosesseihin ja voidaanko ohjelmiston avulla selvittää opiskelijatiedoista opintojen keskeyttämiseen johtavia tekijöitä. Toimeksiantaja on Valkeakosken seudun koulutuskuntayhtymä.

Nykypäivänä uutta dataa syntyy jatkuvasti ja yritysten mahdollisuudet sen hyödyntämiseksi kasvaa, kun keinot analysoida dataa kehittyvät samaa tahtia tietomäärän kasvun kanssa. Historiatietojen raportointi ei enää riitä, vaan analytiikan avulla pyritään löytämään datasta syitä tapahtuneisiin, tekemään ennusteita tulevista ja ohjaamaan toimintaa ihan uudella tavalla. Tiedolla johtamisessa datasta tuotettua tietoa tuodaan osaksi päätöksentekoa ja tieto pyritään jalkauttamaan osaksi yrityskulttuuria ja ihmisten arkea.

Modernit analytiikkatyökalut kattavat koko tiedonkäsittelyn työnkulun datan valmistelusta tiedon visuaaliseen esittämiseen. Tässä opinnäytetyössä tutustuttiin Microsoftin Power BI-ohjelmistoon. Power BI:hin voidaan yhdistää lukuisia eri tietolähteitä, jolloin se voi toimia toimeksiantajan tiedolla johtamisen keskuksena. Koontinäytöt toimivat visuaalisina ja vuorovaikutteisina näkyminä yrityksen mittareihin. Lisäksi Power BI:n analytiikkatyökalujen avulla voidaan selvittää tekijöitä mittareiden määrällisten lukujen takana. Toimeksiantajan opiskelijadatan tiedonlouhinnan tuloksena selviää, että opinnot keskeytetään todennäköisimmin niiden alkuvaiheessa ja kun suorituksia on kertynyt vain vähän.

Avainsanat Data, tiedolla johtaminen, analytiikka

Sivut 42 sivua

Degree Programme in Business Information Technology
Hämeenlinna University Centre

Author	Anu Halttunen	Year 2019
Subject	Utilizing Power BI in Preventing Student Drop-Out Case: Valkeakoski Region Educational Consortium	
Supervisor	Lasse Seppänen	

ABSTRACT

The aim of the thesis was to explore the opportunities Microsoft Power BI offers in the commissioner's knowledge-based management processes and if it is possible to find out reasons behind student drop-out. The commissioner of the thesis is Valkeakoski Region Educational Consortium.

New data is generated constantly, and companies have more and more possibilities to take advantage of it. Ways of analyzing data is evolving at the same pace as new data is generated. Reporting based on only history is not enough. Analytics is needed to describe what happened, make predictions and manage operations. In knowledge-based management information is utilized in decision making. The aim is to implement information as part of corporate culture and employees' everyday life.

Modern analytic tools cover the whole data management process from data preparation to publishing dashboards. In this thesis Microsoft Power BI was explored. Power BI can be connected to several different data sources and in that way, it can act as a center point of the commissioner's knowledge-based management. Dashboards offer visual and interactive views on the company's key performance indicators. Power BI's analytical tools can be applied to find out the factors behind quantitative measures. The research showed that students are most likely to drop out in the early stages of the studies and when they have not completed many competence points.

Keywords Data, knowledge-based management, business analytics

Pages 42 pages

SISÄLLYS

1	JOHDANTO.....	1
2	DATAN MERKITYS LIIKETOIMINNASSA	2
2.1	Datan luokittelu ja kerääminen.....	3
2.2	Tiedolla johtaminen	6
2.3	Strategia datan käsittelyn lähtökohtana	8
2.4	Raportointi vs. analytiikka	10
2.5	Tiedonlouhinta CRISP-DM-mallin mukaisesti.....	12
2.6	Microsoft analytiikka- ja BI-tarjoajien joukossa	15
2.7	Microsoft Power BI.....	17
3	DATA VALKEAKOSKEN SEUDUN KOULUTUSKUNTAYHTYMÄSSÄ	19
3.1	ISO 9001:n mukainen laatujärjestelmä	19
3.2	Mittaaminen ja tiedolla johtaminen	20
3.3	Kehittämiskohteet.....	22
4	TIEDONLOUHINTA OPISKELIJADATASTA	23
4.1	Liiketoiminnan ymmärtäminen.....	23
4.2	Datan ymmärtäminen	24
4.3	Datan valmistelu.....	27
4.4	Mallinnus, tiedonlouhinta	31
4.4.1	Deskriptiivinen mallinnus	31
4.4.2	Diagnostiivinen mallinnus	33
4.4.3	Prediktiivinen mallinnus	35
4.4.4	Preskriptiivinen mallinnus	37
4.5	Arviointi	38
4.6	Tulosten julkaisu.....	39
5	JOHTOPÄÄTÖKSET	40
6	YHTEENVETO	42
	LÄHTEET	43

SANASTO

Analytiikka (Business Analytics)

Liiketoiminnassa analytiikalla tarkoitetaan menetelmiä, joiden avulla datasta saadaan esille päätöksentekoa ja toimenpiteitä tukevaa uutta tietoa, ymmärrystä, korrelaatioita tai ennusteita. Analytiikka on tiedon rikastamista liiketoiminnan päätöksenteon käyttöön esimerkiksi yhdistelemällä tietoja eri lähteistä.

Big Data

Nopeasti kasvavaa ja sisällöltään monimuotoista datamassaa.

Business Intelligence (BI)

Datan hyödyntämistä liiketoiminnassa; liiketoimintatiedon hallintaa.

Cross-Industry Standard Process for Data Mining (CRISP-DM)

Tiedonlouhinnan standardoitu prosessimalli.

Data

Tässä opinnäytetyössä datalla tarkoitetaan mitä tahansa digitaalisessa muodossa olevaa tietoa. Data voi olla määrämuotoista, esimerkiksi numeroita tai tekstiä, tai jäsentelemätöntä tietoa, josta osa on historiatietoa, osa reaaliaikaista ja se sijaitsee hajallaan erilaisissa tietovarastoissa.

Data Analysis Expressions (DAX)

Muun muassa Power BI:ssä käytettävä kaavakieli, jolla voidaan tehdä esim. uusia laskettuja sarakkeita Power BI:n tauluihin. DAX muistuttaa Excelin funktioita.

Master Data

Yrityksen ”kivijalkadata” eli tieto, joka on liiketoiminnan kannalta oleellista, pitkäikäistä ja hitaasti muuttuvaa. Tällaisia tietoja ovat tyypillisimmillään asiakas-, tuote- ja henkilöstötiedot, mutta myös erilliset tunnistetiedot ja luokittelutiedot ja arkikielessä niitä kutsutaan rekistereiksi.

Power BI

Microsoftin raportointi- ja analytiikkatyökalu, joka on osa yritysten Office-pakettia.

SFS-EN ISO 9001:2015

Laadunhallintajärjestelmän rakentamisen ja kehittämisen standardoitu malli.

Teollinen internet (Internet of Things)

Datan keräämistä internetiin kytketyistä laitteista ja koneista.

Tiedolla johtaminen

Johtamis- ja toimintamalli, jossa analysoidua tietoa ja dataa tuodaan osaksi päätöksentekoa.

Tiedonlouhinta (Data Mining)

Datan analysointia algoritmien avulla. Menetelmä, jolla suuresta datamassasta karsitaan oleellinen.

Visuaalinen mittaristo (Dashboard)

Raportointinäköymä. Power BI:ssä puhutaan myös koontinäytöstä, jolla visualisointeja esitetään ja julkaistaan.

1 JOHDANTO

Opinnäytetyön tavoitteena on selvittää, miten analytiikkaa, ja nimenomaan Microsoftin Power BI -ohjelmistoa, voidaan hyödyntää tiedolla johtamisen prosesseissa toisen asteen ammatillisessa koulutusorganisaatiossa. Työn toimeksiantaja on Valkeakosken seudun koulutuskuntayhtymä.

Valkeakosken seudun koulutuskuntayhtymä ylläpitää Etelä-Pirkanmaalla toimivaa Valkeakosken ammattiopistoa, jossa perus-, ammatti- ja erikoisammattitutkintoja opiskelee vuosittain noin 1300 opiskelijaa. Henkilöstöä on noin 110.

Tiedolla johtaminen on keskeinen ja ajankohtainen organisaation kehittämisen teema Valkeakosken seudun koulutuskuntayhtymässä. Tiedolla johtaminen on myös osa koulutuskuntayhtymän laatujärjestelmää, joka on SFS-EN ISO 9001:2015 sertifioitu vuonna 2018. Samana vuonna kuntayhtymä sai kunniamaininnan opetus- ja kulttuuriministeriön lautupalkintokilpailussa (OKM, 2018). Laatujärjestelmän runkona on toimintakäsikirja, joka sisältää kuvauksen toimintajärjestelmästä ja laadunhallinnan periaatteista. Lisäksi toimintakäsikirjaan linkittyy tilastoja, toimintaohjeita ja prosessikuvauksia. Strategiset mittarit kaipaavat nyt tuekseen analytiikkaa.

Kehittämistyön tietoperustana selvitetään datan merkitystä liiketoiminnassa sisältäen tiedolla johtamisen ja analytiikan teoriaa toimeksiantajan näkökulma huomioiden. Lisäksi kuvataan toimeksiantajan laadunhallinnan ja tiedolla johtamisen nykytilanne.

Opinnäytetyö on tyypiltään toiminnallinen ja käytännön osuuden kautta tutkitaan, miten Microsoftin Power BI-ohjelmistoa voidaan hyödyntää osana tiedolla johtamisen käytänteitä. Asiaa tarkastellaan case-tyyppisestitutkimalla opiskelijoiden opintojen keskeyttämiseen johtavia taustatekijöitä. Power BI ei tällä hetkellä ole toimeksiantajalla aktiivisessa käytössä, mutta toiveena on saada opinnäytetyön avulla selkeä käsitys siitä, miten Power BI:a voitaisiin hyödyntää organisaation tiedolla johtamisen välineenä.

Tutkimuskysymykset ovat:

- Mitä mahdollisuuksia Microsoftin Power BI -ohjelmisto tarjoaa toimeksiantajan tiedolla johtamisen prosesseihin?
- Miten Power BI:n avulla voidaan selvittää toimeksiantajan opiskelijatiedoista opintojen keskeyttämiseen johtavia syitä?

2 DATAN MERKITYS LIKETOIMINNASSA

Kielitoimiston sanakirjan (Kielitoimisto, n.d.) mukaan data on ”asian säännönmukainen esitys viestitettävässä tai käsittelykelpoisessa muodossa; tieto, tiedot”. Tässä opinnäytetyössä data on yleisnimitys kaikelle digitaalisessa muodossa olevalle tiedolle.

Datalla on ollut pitkään tärkeä rooli yrityksissä, mutta perinteisesti sitä on hyödynnetty lähinnä liiketoiminnan prosessien mittaamiseen ja hallintaan sekä ennustamiseen ja pitkän aikavälin suunnitteluun. Dataa on tuotettu tutkimusten, kyselyiden ja mittausten avulla ja se on varastoitu erillisiin tietokantoihin. Tallennusmäärät tietokannoissa ovat olleet rajalliset ja tietojärjestelmät tyypillisesti hidastuivat, mitä suurempia määriä dataa oli säilötyä. (Marr, 2016, s.3; Rogers, 2016, s.91)

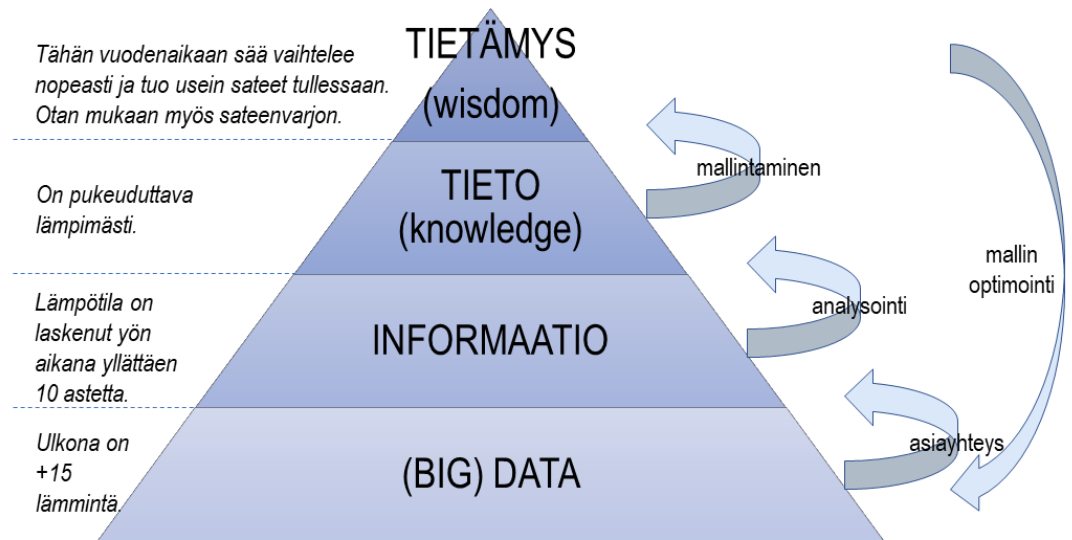
Nykyään datan rooli on muuttunut ja käyttömahdollisuudet ovat rajattomat. Taulukossa 1 vertaillaan tapahtuneita muutoksia strategisten oletusten näkökulmasta. Helpoin osuus tänä päivänä on datan tuottaminen, sillä uutta dataa syntyy jatkuvasti. Nykypäivän digitaalisuuden myötä kaikesta tekemästämme jää datajälkiä, jota on mahdollista kerätä ja analysoida. Data voidaan louhia verkossa tekemistämme toiminnoista, keskusteluista, jakamistamme kuvista ja videoista, erilaisista tunnistimista sekä verkkoon kytketyistä laitteista, jolloin puhutaan termistä Internet of Things (IoT). Esimerkiksi älypuhelimien GPS (Global Positioning System) kertoo, missä olemme, kuinka nopeasti liikumme, millainen sää vallitsee tai jopa kuinka suurella voimakkuudella painamme älypuhelimien näyttöä. (Marr, 2015, s. 64, 2016, s. 2; Rogers, 2016, s. 91)

Taulukko 1. Muutokset siirryttäessä analogisesta ajasta digitaaliseen aikaan (Rogers, 2016, s. 91).

Vanha	Uusi
Datan tuottaminen on kallista.	Dataa tuotetaan jatkuvasti joka paikassa.
Haasteena on datan varastointi ja hallinta.	Haasteena on muuttaa data arvokkaaksi ja hyödynnettävissä olevaksi informaatioksi.
Yritykset voivat hyödyntää vain jäsennettyä dataa.	Jäsentelemätön data on enenevissä määrin käytettävissä ja hyödynnettävissä.
Dataa varastoidaan ja käsitellään erillisissä silloissa.	Datan arvo tulee sen käytettävyydestä sillojen välillä.
Data on prosessien optimoinnin väline.	Data on keskeinen aineeton pääoma, joka tuottaa yritykselle arvoa.

Uusia, nopeasti kasvavia ja sisällöltään monimuotoisia datamassoja kuvataan usein termillä big data. Toisaalta big datalla viitataan myös uusiin mahdollisuuksiin käsitellä ja tulkita tällaista dataa. Lisäksi big datan määritelmällä viitataan uuteen pilvipalveluiden infrastruktuuriin, jolloin data on helposti yhä useamman yrityksen saatavilla. (Rogers, 2016, s. 98; Salo, 2014, s. 35)

On hyvä muistaa, että data ei yksinään riitä tuottamaan arvoa liiketoiminnalle. Kuva 1 kertoo, kuinka data oikeassa asiayhteydessä tuottaa informaatiota, josta voidaan analysoimalla tuottaa tietoa. Saatu tieto voidaan ymmärtää syvällisesti, kun sen taustalla olevista malleista voidaan tehdä olettamuksia. Tämä synnyttää tietämyksen, jonka ansiosta voidaan tehdä oikeita päätöksiä. Tietämyksen avulla tätä mallia voidaan myös optimoida. (Succi & Coveney, 2019)



Kuva 1. Yksinkertaistettu tietohierarkian kuvaus datan jalostumisesta viisaudeksi ja asian havainnollistaminen säätilan muutosta kuvaavalla esimerkillä. (Listenmaa, Piikkilä, & Aho, 2018; Succi & Coveney, 2019)

Uudet teknologiat mahdollistavat datan varastoimisen ja käsittelyn eri paikoissa sijaitsevien tietokantojen välillä verkkoyhteyksien avulla. Hajautettu tietojenkäsittely tarkoittaa valtavien tietomäärien varastointia pieninä osina useissa eri tietokannoissa ja niiden analysointia eri servereillä, jolloin ne suorittavat vain osia analysointikokonaisuudesta. Tähän tarkoitukseen on tarjolla monenlaisia työkaluja ja palveluita, mikä mahdollistaa datan analysoinnin kenelle tahansa, myös pienen budjetin ja rajallisen IT-tuen omaaville yrityksille. Suurin haaste on datan muuttaminen hyödynnettävään muotoon. Datasta ei tule automaattisesti yrityksen pääomaa vaan myös asenteen yrityksessä tulee muuttua. Datan hyödyntämisen tuoma muutos ei siis pohjimmiltaan koske vain teknologiaa vaan myös yrityksen strategiaa ja uutta tapaa ajatella. (Marr, 2016, s. 3; Rogers, 2016, s. 91)

2.1 Datan luokittelu ja kerääminen

Data voidaan luokitella sijaintinsa perusteella joko yrityksen sisäiseen tai ulkoiseen dataan. Sisäinen data on nimensä mukaisesti kaikkea sitä tietoa, johon yrityksellä itsellään on suora pääsy. Näitä voivat olla erilaiset palautteet, myyntidata, yms. Sisäisen datan perusta on ns. Master Data, joka on

yrityksen liiketoiminnan kannalta oleellista, pitkäikäistä ja hitaasti muuttuvaa tietoa. Näitä kutsutaan tavallisesti rekistereiksi ja tyypillisesti ne sisältävät asiakas-, tuote- ja henkilöstötietoja sekä erilaisia tunnist- ja luokiteltutietoja. Ulkoinen data taas sijaitsee yrityksen ulkopuolella ja se voi olla joko julkista tai yksityistä. Julkinen data on vapaasti saatavilla, mutta yksityinen data on sellaista, joka hankitaan kolmannelta osapuolelta. Ulkoista dataa ovat esimerkiksi säätiedot, väestörekisterin tiedot, sosiaalisen median tiedot, keskustelut, Googlen trendit tai karttapalvelut. (Markkula & Syväniemi, 2015, s. 39; Marr, 2015, s. 62-64)

Sijaintinsa lisäksi data voidaan jakaa joko jäsenneltyyn tai jäsentelemättömään dataan. Jäsennelty data (structured data) on tallennettu määrättyssä muodossa tiettyyn kenttään tiettyssä rekisterissä tai tiedostossa. Tallennuspaikkana on usein relaatiotietokanta tai laskentataulukko, jossa jokaisella kentällä on nimi ja yhteydet kenttien välillä on tarkasti määritelty. Jäsennelty data on helppoa syöttää, varastoida ja analysoida. Kuten taulukossa 1 nähtiin, aiemmin voitiin käyttää ainoastaan jäsenneltyä dataa ja kaikki muu data joko hylättiin kokonaan tai varastoitui paperille ja mappiin. Jäsentelemätön (unstructured) tai puolijäsennelty (semi-structured) data kattaa kaikenlaisen muun datan, jota ei voi osoittaa tiettyyn sarakkeeseen, riviin tai kenttään. Se voi olla muodoltaan tekstiä, päivämääriä, numeroita, kuvia, tiedostoja, videoita, verkkosivuja tai mitä vaan. Puolijäsennellyssä datassa voi olla jonkin verran rakennetta, mutta se ei ole täysin määrämutoista. (Marr, 2015, s. 60-61)

Dataa keräämiseen on useita keinoja, joista taulukko 2 esittelee seitsemän. Sisäinen jäsennelty data on kaikkein helpoimmin kerättävissä ja analysoitavissa sekä yleensä se on edullisinta hankkia, kun taas ulkoisen jäsentelemättömän datan hankinta vaatii yleensä suurimmat ponnistelut. (Marr, 2015, s. 84-85)

Taulukko 2. Seitsemän keinoa hankkia dataa (Marr, 2015, s. 86-87)

HANKINTATAPA	Luokittelu			Sijainti	
	jäsen- nelty	puoli- jäsennelty	jäsentele- mätön	sisäinen	ulkoinen
Muodostettu data (created data)	x	x		x	x
Tuotettu data (provoked data)	x	x		x	x
Liiketahtumadata (transaction data)	x			x	
Mukautettu data (complied data)	x				x
Kokeellinen data (experimental data)	x	x		x	x
Toiminnasta syntyvä data (captured data)			x	x	x
Käyttäjän generoima data (user-generated data)			x	x	x

Muodostettua dataa saadaan, kun ihmisiltä kysytään kysymyksiä tai heille tarjotaan välineitä, joiden avulla he voivat antaa vastauksia. Esimerkkeinä näistä ovat markkinatutkimukset sekä asiakastytyvyisyys- ja muut kyselyt. Kanta-asiakasohjelmat voidaan laskea myös tämän datatyyppin piiriin, sillä liittyessään jäseneksi ihmiset antavat usein vapaaehtoisesti itsestään tietoja yritykselle. (Marr, 2015, s. 86)

Tuotettua dataa varten ihmisiä kutsutaan ilmaisemaan mielipiteitään ja näkemyksiään asioista tai tuotteista. Muun muassa verkkokauppojen ostosten yhteydessä pyydettyt pisteytykset esimerkiksi 1-5 tähdellä ja tuotteiden sanalliset arvioinnit ovat yksi esimerkki tällaisesta datasta. (Marr, 2015, s. 86)

Liiketapahtuma- eli transaktiodataa muodostuu nimensä mukaisesti liiketoiminnan tapahtumissa. Kun asiakas ostaa tuotteen myymälässä tai verkossa, saadaan dataa, joka kertoo mitä, missä ja koska ostettiin. (Marr, 2015, s. 86)

Mukautettua dataa tarjoavat yritykset, joiden liiketoimintana on kerätä valtavia tietomassoja yksittäistä ihmisistä. Tiedot koostuvat esimerkiksi yhteystiedoista, iästä, sukupuolesta, siviilisäädystä, talouden koosta, vakuutus- tai luottokorttitiedoista sekä esimerkiksi millaisen auton henkilö omistaa. Yritykset voivat ostaa mukautettua dataa omiin markkinointitarkoituksiinsa. (Marr, 2015, s. 87; Wilson, 2013)

Kokeellinen data on yhdistelmä muodostettua- ja liiketapahtumadataa. Siinä toteutetaan esimerkiksi tiettyä markkinointia erilaisille asiakasryhmille ja sitten tarkastellaan, mikä niistä tuottaa parhaan tuloksen. (Marr, 2015, s. 87)

Toiminnasta syntyvää dataa kerätään passiivisesti yksilön tai tavaran toiminnan seurauksena yleensä ilman, että ihmiset ovat edes tietoisia tiedonkeruusta. Esimerkkinä voidaan mainita Googlen hakuhistoria tai puhelimen GPS-tiedot. Internet of Things kuuluu tähän kategoriaan. (Marr, 2015, s. 88)

Käyttäjän generoima data on sellaista, jota yksilöt tai yritykset luovat tietoisesti esimerkiksi tekemällä sosiaalisen median päivityksiä, kommentointia ja keskusteluita sekä blogitekstejä. Tällainen data on yleensä varsin jäsentelemätöntä. (Marr, 2015, s. 88)

Datan keräämisen yhteydessä on tärkeää huomioida datan oikeellisuus, läpinäkyvyys ja tietosuoja, jolloin puhutaan datan laadunhallinnasta (Data Quality Management). Oikeellisuuden varmistaminen voi olla vaativaa, mutta se on välttämätöntä luotettavan tiedon saamiseksi, sillä väärä data tuottaa väärää tietoa. Datan laadunhallinnan merkityksen arvellaan korostuvan lähitulevaisuudessa entisestään. Sen lisäksi, että datan tulee lähtökohtaisesti olla oikeaa ja laadukasta, laadunhallintaa tulee tehdä koko

datan elinkaaren ajan. Siihen kuuluu datan puhdistusta, valmistelua, yhä suuremman määrän datan keräämistä sekä datan tehokasta jakelua ja jakelun hallintaa. Tulee varmistaa, että dataa käytetään kaikissa sen elinkaaren vaiheissa oikein. Läpinäkyvyyden näkökulmasta on noudatettava tiedonkeruuta koskevia lainsäädäntöjä, jotka säätelevät esimerkiksi henkilötietojen ja verkossa tapahtuvasta toiminnasta syntyvän datan keräämistä. Tietosuojasäädösten arvellaan kiristyvän entisestään tulevaisuudessa ja datan laadunhallinnalla on tärkeä rooli säädösten noudattamisessa. (Keller, 2019; Markkula & Syväniemi, 2015, s. 56 - 57, 63)

Liiketoiminnan kannalta pelkkä datan kerääminen ja varastointi eivät tuota arvoa yritykselle, vaan datasta tulee arvokasta vasta, kun se on jalostettu tiedoksi. Yrityksen ei ole järkevää kerätä kaikkea saatavilla olevaa tietoa ja rakentaa suuria ja monimutkaisia tietokantoja, vaan suositeltavampaa on rakentaa niin pieni tietokanta kuin mahdollista. Se onnistuu, kun suunnitellaan huolellisesti, mitä dataa oikeasti tarvitaan ja mihin sitä on tarkoitus käyttää. Datan elinkaareen tulee kiinnittää huomiota, sillä esimerkiksi asiakasdata vanhenee viimeistään viiden vuoden kuluessa. Vanhan datan läpikäynti voi tuottaa ”kiva tietää”-tyyppistä tietoa, mutta tulevaisuuden trendien kannalta tieto voi olla merkityksetöntä, eikä siitä ole johtamisen kannalta hyötyä. (Markkula & Syväniemi, 2015, s. 36; Marr, 2015, s. 27 - 28)

2.2 Tiedolla johtaminen

Laadun kannalta tieto on ratkaiseva perusta yrityksen päätöksenteossa ja oikeanlaisen tiedon hankkiminen on nostettu organisaatioissa merkittäväksi tekijäksi. Johtamisen yhteydessä puhutaan termeistä näyttöön perustuva johtaminen (evidence-based management), tietojohdaminen (knowledge management) ja tiedolla johtaminen (knowledge-based management). Ne kertovat tarpeesta vähentää epävarmuutta ja luoda johtamisen tuloksena uudenlaista järkiperäisyyttä organisaatioiden toimintaan. (Listenmaa et al., 2018, s. 5 - 6; Pellinen, 2017, s. 15; Viitala & Jylhä, 2019)

Tietojohdamiselle ja tiedolla johtamiselle on olemassa useita eri määritelmiä ja monesti niistä puhutaan synonyymeina. Selkein eroavuus on se, että tietojohdaminen keskittyy tiedon luomiseen ja jakamiseen, kun taas tiedolla johtamisen painopiste on ihmisten ja organisaatioiden johtamisessa ja tiedon jalkauttamisessa osaksi yrityskulttuuria ja ihmisten arkea. Tiedolla johtamisella tarkoitetaan siis johtamis- ja toimintamallia, jossa tietoa hyödynnetään ja analysoitua tietoa ja dataa tuodaan osaksi päätöksentekoa. Kuva 2 kertoo tiedolla johtamisen lähtökohdista. (Listenmaa et al., 2018, s. 5 - 6; Pellinen, 2017, s. 15; Tietokiri, 2019)

STARTTIKYSYMYKSIÄ TIEDOLLA JOHTAMISEN PROJEKTIIN:

- Mitkä ovat strategian johtamisen ja operatiivisen toiminnan kannalta keskeiset mittarit?
- Kuka tarvitsee tietoa, millä rytmillä ja miten jaettuna?
- Mitä dataa tällä hetkellä varastoidaan, mihin ja missä muodossa, kuka vastaa?
- Miten dataa hyödynnetään ja kuka vastaa datan jalostamisesta informaatioksi?
- Mikä on laatutaso ja miten se varmistetaan jatkossa?
- Mikä on yrityksen Master Data?
- Mikä on tiedon hyödyntämisen toimivin infrastruktuurirakenne?
- Mikä tieto tulee välittömästi päivittää?
- Miten tietoa kerätään jatkossa omista tietolähteistä ja mitä hankitaan ulkoa?
- Kuka vastaa tietosuoja- ja kyberturvallisuuskysymyksistä?
- Mikä tieto on relevanttia ennusteita laadittaessa?

Kuva 2. Kysymyksiä tiedolla johtamisen projektin aloittamiseen (Markkula & Syväniemi, 2015, s. 38)

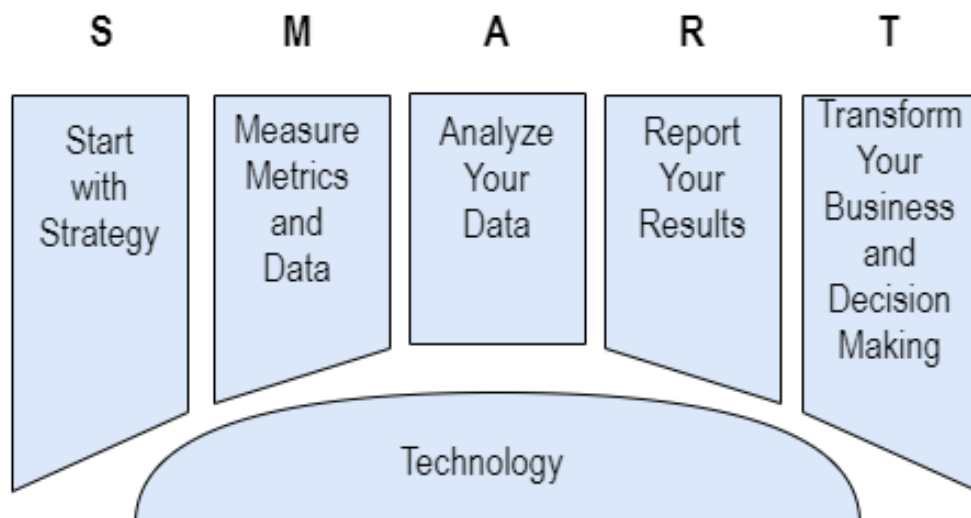
Suurin haaste tiedolla johtamiseen siirryttäessä on sen vaatima kulttuurimuutos, jossa tieto kuuluu kaikille. Tiedolla johtaminen on jatkuvaa vuorovaikutusta toimintaympäristön, organisaation, tiimin ja yksilön välillä. Perinteisistä osasto- tai toimintokohtaisista, yksittäisistä ja erillisistä tiedoista pitää muodostaa yhtenäinen näkymä koko liiketoimintaan ja sen tavoitteisiin ja tuloksiin. Ajantasaisen tiedon tulee tukea prosesseja siellä, missä tietoa tarvitaan. Tiedon esittämisessä käyttäjänäkymä voi muuttua näkökulman mukaan esim. osastokohtaisesti, mutta yhteinen tietopohja kuuluu kaikille. Kun yksilön tiedolla johtamisen kyvykkyyttä, vastuullisuutta, intoa ja sitoutumista kehitetään, tuetaan tiimien sisäistä toimintaa ja tiedolla johtamisen jalkauttamista. Voidaan sanoa, että tiedolla johtamisen muutoksessa on kysymys enemmän informaatio- kuin teknologiahankkeesta. (Listenmaa et al., 2018, s. 21; Syväniemi, 2015)

Kuten aiemmin on todettu, yrityksillä ei tänä päivänä ole puutetta datasta, mutta kaikki data ei ole liiketoiminnan ja päätöksenteon kannalta merkityksellistä. On välttämätöntä erottaa oleellinen tieto epäolennaisesta, varmistaa datan käytettävyys, varastoinnin rakenteet, sen arvoa lisäävä jalostus sekä toimintoihin liittyvät vastuukysymykset. Mikäli lähitulevaisuudessa yritykset ymmärtävät paremmin sen, että kun datalla on omistaja eli vastuu datasta on selkeästi määritelty, voidaan päästä monimutkaisten visuaalisten mittaristojen viidakosta aidosti keskitettyyn ja määrittäisen analytiikan kulttuuriin, joka tukee yrityksen tiedolla johtamista parhaalla mahdollisella tavalla. Jotta tiedolla johtaminen saadaan vietyä teoriasta käytäntöön, tulee kaiken toiminnan lähtökohtana olla yrityksen strategia. Riskinä on muutoin se, että tieto ja analytiikka jäävät irrallisiksi toiminoiksi, jotka eivät todellisuudessa ohjaa arjen päätöksentekoa. (Digia, 2018; Syväniemi, 2015, 2016)

2.3 Strategia datan käsittelyn lähtökohtana

Puhuttaessa big datasta, hehkutetaan usein ulkoisen, jäsentelemättömän datan ylivoimaisuutta, mutta todellisuudessa ilman strategiaa mikään data ei ole toista parempaa, eikä tiedolla johtamista saada vietyä teoriatasolta käytäntöön. Strategian jalkautus toiminnaksi vaatii yhteiseen tietoon ja tavoitteisiin pohjautuvia mittareita. Datan käsittelyn ja tiedolla johtamisen tavoitteena tulee olla tiedon helppokäyttöisyys ja oikein kohdennettu visualisointi. (Marr, 2015, s. 63; Syväniemi, 2015, 2016)

Bernard Marr (2015) esittelee kirjassaan *Big Data: Using SMART Big Data, Analytics and Metrics to Make Better Decision and Improve Performance* kuvan 3 mukaisen viisiportaisen mallin, jonka vaiheiden avulla kaiken kokoiset yritykset voivat ottaa datan hyödyntämisen haltuun riippumatta käytettävissä olevasta budjetista. Mallia voidaan hyödyntää yrityksen tiedolla johtamisen käynnistämisessä.



Kuva 3. SMART-malli (Marr, 2015, s. 21)

Strategian määrittely (Start with Strategy) on SMART-mallin ensimmäinen vaihe. Vaiheen tarkoituksena on määritellä huolellisesti kysymykset, joihin halutaan datan käsittelyn avulla löytää vastaukset. Kysymysten huolellinen asettelu auttaa tunnistamaan, millaista dataa todellisuudessa tarvitaan. Monesti tunnistetaan omien tietolähteiden lisäksi tarve uusien datan muotojen käyttöönotolle. (Marr, 2015, s. 25)

Seuraavassa vaiheessa (Measure Metrics and Data) pyritään täsmentämään, millaisen datan ja tietolähteiden avulla voidaan löytää vastaukset ensimmäisessä vaiheessa tuotettuihin kysymyksiin. Apuna tietolähteiden määrittelyyn voi käyttää taulukon 3 mukaista lomaketta. (Marr, 2015, s. 100)

Taulukko 3. Kysymykset, joiden avulla voidaan määritellä datan tarve ja sijainti (Marr, 2015, s. 100)

SMART-kysymys, johon haetaan vastausta:			
Tietolähteet, joiden avulla SMART-kysymykseen vastataan			
- Tietojoukon nimi - Tietotyyppin kuvaus - Tiedon sijainti ja omistajuus: sisäinen/ulkoinen - Datan muoto: jäsenneily vai jäsentelemtön? - Tiedonkeruutapa - Mihin data on tallennettu? - Datamäärän kuvaus - Kuvaile datan nopeus, taa-juus ja viive - Kuvaile datan oikeelli-suus/laatu - Kuinka dataa tullaan analy-soimaan? - Tiedonkeruuseen, tallenta-miseen ja analysointiin liitty-vät kustannukset.	Tietolähde 1	Tietolähde 2	Tietolähde 3

Kolmannessa eli datan analysoinnin (Analyze Your Data) vaiheessa valitaan, mitä analytiikkamenetelmää sovelletaan, jotta vastaus löydetään. Menetelmän valintaan vaikuttaa mm. käytettävissä oleva data eli onko kyseessä tekstiä tai numeroita sisältävää dataa vai jotain muuta, kuten kuvia, videoita tai anturidataa. Analysoinnin toteutuksessa on hyvä muistaa, että tavoitteena on tuottaa liiketoimintaymmärrystä lisäävää tietoa. Jatkuva uuden datan kerääminen on merkityksentöntä, mikäli siitä tuotetulla tiedolla ei tehdä mitään. (Marr, 2015, s. 105)

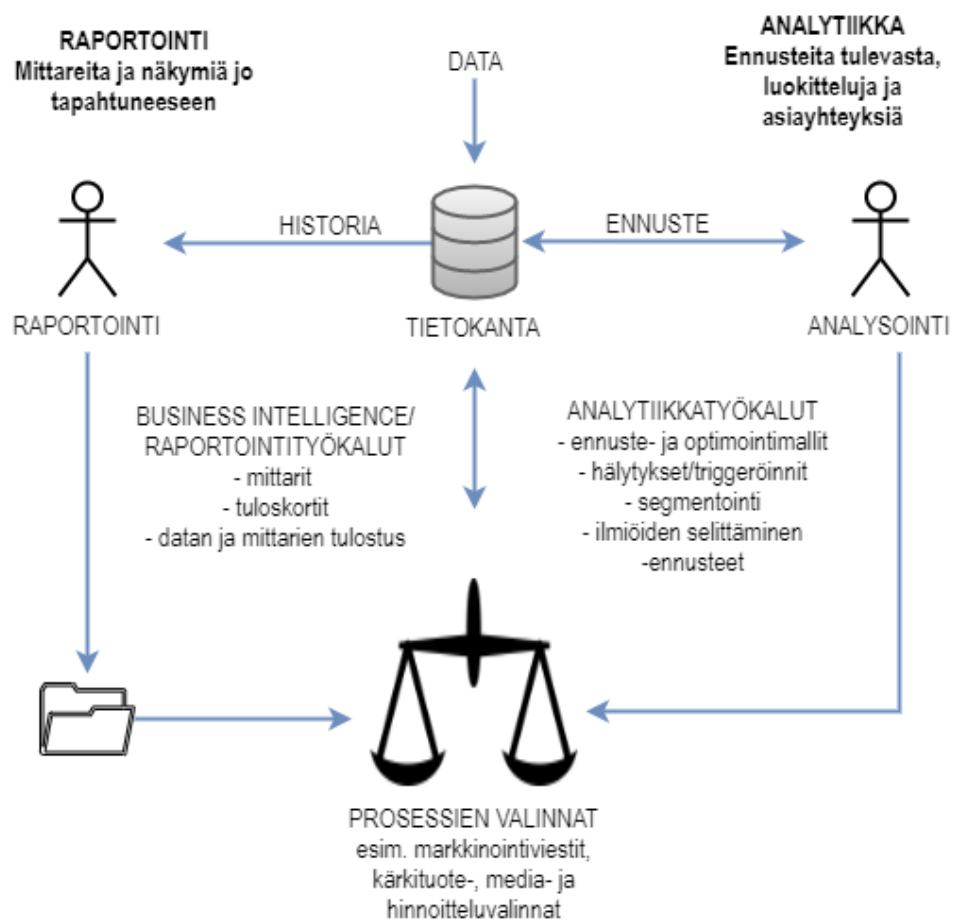
Kun analysointi on tehty, neljäs vaihe on tulosten raportointi (Reporting Results). Tärkeintä on, että oikeat ihmiset saavat oikeassa muodossa tarvitsemansa tiedon. Tulokset tulee esittää ymmärrettävällä tavalla ja mielellään mahdollisimman tiiviisti. Raportin suuri sivumäärä ei itsessään tuota lisäarvoa, vaan joskus jopa vain yksisivuinen visualisointi nostaa parhaiten olennaisen tiedon esiin. Visualisointiin ja oikean kaavion valintaan kannattaa kiinnittää huomiota. Perinteisten pylväs-, viiva ja piirakkadiagrammien rinnalla on tarjolla lukematon määrä kaavioita, joilla voidaan kuvata myös mm. tekstejä, käyttäytymistä, tunteita tai yhteyksiä. Tärkeintä on miettiä, miten visualisointi parhaiten vastaa määriteltyyn kysymykseen. (Marr, 2015, s. 155 - 190)

Viimeinen vaihe SMART-mallissa on liiketoiminnan ja päätöksenteon kehittäminen saadun vastauksen perusteella (Transform Your Business and Decision Making). Löydetyt oivallukset tulee ottaa käyttöön ja rohkaista yrityksen työntekijöitä muuttamaan ajattelutapaansa näyttöön perustuvalla johtamisella. Joitain vastauksia voidaan hyödyntää vain kertaluontoisesti, mutta joidenkin osalta jatkuva seuranta on kannattavaa. (Marr, 2015, s. 229 - 230)

Tiivistetysti SMART-mallin tärkein sanoma on se, että data yksinään ei riitä, vaan tiedolla johtamisen perustana täytyy olla strategiaan perustuva kysymys, johon halutaan vastaus. Toisaalta se muistuttaa siitä, että datan käsittelyn ja tulosten saamisen jälkeen tulee vielä vaihe, jossa sekä liiketoimintaa että ajattelua pyritään tarvittaessa muuttamaan saatujen vastauksen perusteella. Tärkeintä on miettiä, miksi dataa halutaan kerätä ja käsitellä. Kun lähtökohtana on yrityksen strategia eikä data, voidaan keskittyä siihen, mikä on tärkeää eikä siihen, mikä on mahdollista.

2.4 Raportointi vs. analytiikka

Syväniemen (2016) mukaan valitettavan usein tiedolla johtamisella viitataan historiatietoja raportoihin Business Intelligence –järjestelmiin ja erilaisiin seurantadokumentteihin. Nykypäivän kilpailutilanteessa tämä ei riitä, vaan trendejä ja ilmiöitä on ymmärrettävä ja on hallittava ennuste- ja optimointimalleja. Tämän myötä analytiikka on tullut välttämättömäksi osaksi liiketoiminnan arkea. Kuva 4 täsmentää raportoinnin ja analytiikan eroavaisuutta.



Kuva 4. Business Intelligencen ja analytiikan erot (Markkula & Syväniemi, 2015, s. 90)

Perinteisesti analytiikka jaetaan neljään tyyppiin: deskriptiiviseen (descriptive), diagnostiiviseen (diagnostic), prediktiiiviseen (predictive) ja preskriptiiviseen (prescriptive) analytiikkaan. Kuvaileva eli deskriptiivinen analytiikka pyrkii kertomaan, mitä on tapahtunut. Dataan perehdytään tarkemmin käyttämällä esimerkiksi luokittelua tai ääriarvojen hakua. Diagnostiivinen analytiikan pyrkimyksenä on selittää, miksi jokin asia tapahtui. Siinä tarkastellaan korrelaatioita ja pyritään hakemaan riippuvuuksia asioiden välillä. Ennustava eli prediktiiivinen analytiikka rakentaa tilastollisia malleja tai hyödyntää koneoppimisalgoritmeja, joiden avulla selvitetään syvemmällä piilossa olevia yhteyksiä. Toimivaksi todettuja malleja voidaan käyttää tulevien tapahtumien ennustamisessa. Preskriptiivisen eli toimintaa ohjaavan analytiikan tarkoituksena on kertoa, mitä seuraavaksi kannattaisi tehdä. Sen perustana on aiempien analytiikkavaiheiden ennustemallit, joiden avulla voidaan räätälöidä esimerkiksi asiakaskohtaisia tuotesuosituksia tai hinnoitteluja. (Sherman, 2015, s. 16; Sihvo, 2016)

Keinot analysoida dataa kehittyvät samaa tahtia tietomäärän kasvun kanssa. Algoritmit pystyvät jo tunnistamaan ihmisen kuvasta sekä ymmärtämään ja analysoimaan kirjoitetun tekstin sisältöä. Tekstimuotoisen datan analysoinnissa uutta tietoa ja näkemyksiä poimitaan suuren tekstimäärän joukosta. Tekstiä voidaan luokitella, ryhmitellä, tehdä yhteenvetoja, kääntää toiselle kielelle, poimia käsitteitä tai sen sävystä tehdään tunneanalyysiä pääättelemällä, sanotaanko vastaanottajalle positiivisia vai negatiivisia asioita. Tekstit voivat sijaita esimerkiksi tiedostoissa, raporteilla, yrityksen sisäisessä tai ulkoisessa viestinnässä, sähköposteissa, blogeista, verkkosivuilla tai sosiaalisen median päivityksissä. (Marr, 2015, s. 108 - 118)

Tekstin analysoimisen ohella myös puheen analysoinnissa voidaan sisällön lisäksi päätellä tunnetiloja. Myös puheen äänenkorkeutta, intonaatioita eli sävelkulkuja sekä puheen ja taukojen välistä suhdetta analysoidaan. Puheen analysointia ovat lisäksi esimerkiksi äänentunnistus, puheen muuttaminen tekstiksi ja puheohjaus. (Marr, 2015, s. 119 - 121)

Video- ja kuva-analyyseja käytetään esimerkiksi kasvojentunnistuksen avulla tapahtuvaan ihmisten tunnistamiseen. Myös ihmisen käytöstä voidaan seurata ja arvioida videolta esimerkiksi myymälävarkauksien ja muiden rikosten ehkäisyssä. (Marr, 2015, s. 123 - 129)

Webdatan analysointiin on kehitelty omia menetelmiään. Menetelmät voidaan jakaa kolmeen eri kategoriaan. Verkkosisällön tiedonlouhinnan menetelmiä käytetään nimensä mukaisesti hyödyllisen tiedon tuottamiseen verkkosivujen teksti-, kuva-, video-, ääni- yms. sisällöistä. Web-rakenteen louhinta analysoi sivujen hyperlinkkien rakennetta, kuten sitä, miten sivut linkittyvät toisiinsa. Verkkokäytön louhinnan menetelmillä tutkitaan sivustoilla vierailevien kävijöiden navigointipolkuja. (Bošnjak, Grljević, & Bošnjak, 2019; Jokar, Honavar, AgHamirzadeh, & Esfandiari, 2016)

Koneoppimista hyödynnetään parhaillaan monipuolisesti tiedonlouhinnassa ja automatisoidussa analytiikassa, mutta ihmisen roolina on edelleen ollut tulosten tulkinta, johtopäätösten tekeminen ja toimenpiteistä päättäminen. Uskotaan, että tulevaisuudessa kaikki analytiikassa automatisoitavissa oleva automatisoidaan ja ennustetaan, että lähitulevaisuudessa jopa tulosten tulkinnan voisi hoitaa tekoäly, jolloin ihmiset voidaan korvata päätöksentekopisteissä analytiikalla. Kuitenkin edelleen muistutetaan, että analytiikan tehtävänä on olla tiedolla johtamisen työkalu, se ei itsessään ole ratkaisu kaikkeen. (Digia, 2018)

2.5 Tiedonlouhinta CRISP-DM-mallin mukaisesti

Tiedonlouhinta on analysoinnin muoto, jossa datasta etsitään kiinnostavia malleja tai rakenteita, joita käytetään nykytilan selittämiseen ja tulevaisuuden ennustamiseen. Lopputulemana voi olla monimutkaisia visuaalisia mittaristoja (dashboard) tai yksinkertaisia listoja asioista ja muodoltaan se voi olla mitä tahansa sääntöjoukoista kuvaajiin tai yhtälöihin. (Roiger, 2017, s. 5)

Kun tiedonlouhinta toteutetaan systemaattisesti mallin avulla, voidaan minimoida projektien epäonnistuminen. Yksi tunnetuimmista tiedon louhinnan standardoinnin malleista on Cross-Industry Standard Process for Data Mining eli CRIPS-DM-menetelmä, jonka varhaiset tiedonlouhinnan toimijat DaimlerChrysler, SPSS ja NCR kehittivät jo vuonna 1996. Sitä pidetään edelleen sen kokonaisvaltaisuuden ja joustavuuden ansiosta parhaana käytännön mallina analytiikan alalla. Kuten aiemmin esitelty SMART-malli, myös CRISP-DM lähtee liikkeelle strategiasta. Siinä missä SMART-malli toimii yrityksen datankäsittelyn yleisprosessina, CRISP-DM on tarkoitettu nimenomaan tiedonlouhinnan malliksi. (Bahrepour, 2018; Chapman et al., 2000; Markkula & Syväniemi, 2015, s. 95)

CRISP-DM:n periaatteena on purkaa prosessi kuuteen vaiheeseen, jotka on kuvattu taulukossa 4. Jokaisessa vaiheessa on määritelty tehtävät ja niistä saatavat lopputulokset. Prosessi etenee vaihe vaiheelta, mutta on mahdollista palata takaisin edelliseen vaiheeseen. Kyseessä on iteratiivinen prosessi, jossa viimeisen vaiheen jälkeen voidaan aloittaa alusta ja parantaa työtä uudella kierroksella. (Biecek, 2019; Kienzler, 2018)

Taulukko 4. CRISP-DM referenssimallin vaiheet (Chapman et al., 2000; Smart Vision Europe, 2018). Lihavoituna tehtävä ja kursivoituna lopputulos.

VAIHE 1: Liiketoimin- nan ymmär- täminen	VAIHE 2: Datan ymmärtämi- nen	VAIHE 3: Datan valmistelu	VAIHE 4: Mallinnus	VAIHE 5: Arviointi	VAIHE 6: Julkaisu
Määritä lii- ketoiminnal- liset tavoit- teet <i>Taustatiedot</i> <i>Tavoitteet</i> <i>Liiketoimin- nan menes- tyskriteerit</i> Arvioi ti- lanne <i>Resurssit</i> <i>Vaatimukset,</i> <i>oletukset ja</i> <i>rajoitukset</i> <i>Riskit ja eh-</i> <i>dot</i> <i>Termistö</i> <i>Kulut ja hyö-</i> <i>dyt</i> Määritä tie- donlouhin- nan tavoit- teet <i>Louhinnan</i> <i>tavoitteet ja</i> <i>menestyste-</i> <i>kijät</i> Laadi pro- jektisuunni- telma <i>Projektisuun-</i> <i>nitelma</i> <i>Arvio työka-</i> <i>luista ja tek-</i> <i>niikoista</i>	Kerää alus- tava data <i>Raportti alus-</i> <i>tavasta da-</i> <i>tasta</i> Datan ku- vaaminen <i>Raportti da-</i> <i>tan kuvauk-</i> <i>sesta</i> Datan tutki- minen <i>Raportti da-</i> <i>tan tutkimi-</i> <i>sesta</i> Varmista da- tan laatu <i>Raportti da-</i> <i>tan laadusta</i>	Datan va- linta <i>Mitä tiedosta</i> <i>otetaan mu-</i> <i>kaan ja mitä</i> <i>jätetään pois</i> Datan sii- vous <i>Raportti da-</i> <i>tan siivouk-</i> <i>sesta</i> Datan raken- taminen <i>Uusien attri-</i> <i>buuttien luo-</i> <i>minen</i> <i>Uusien tietu-</i> <i>eiden gene-</i> <i>roiminen</i> Datan integ- rointi <i>Tietojen yh-</i> <i>distäminen ja</i> <i>koostaminen</i> <i>useista tau-</i> <i>luista tai</i> <i>tietueista</i> Datan for- matointi <i>Uudelleen</i> <i>muotoillut tie-</i> <i>dot</i>	Valitse mal- linnusmene- telmä <i>Mallinnusme-</i> <i>netelmä</i> <i>Mallinnuksen</i> <i>olettamukset</i> Suunnittele testiversio <i>Testiversio</i> Rakenna malli <i>Parametrien</i> <i>asetukset</i> <i>Mallit</i> <i>Kuvaukset</i> Arvioi mallia <i>Mallin arvio</i> <i>Tarkistetut</i> <i>parametrien</i> <i>asetukset</i>	Arvioi tulok- sia <i>Tiedon-</i> <i>louhinnan ar-</i> <i>viointi suh-</i> <i>teessa liike-</i> <i>toiminnan</i> <i>menestyskri-</i> <i>teereihin</i> Käy pro- sessi läpi <i>Prosessin lä-</i> <i>pikäynti</i> Määrittele seuraavat stepit <i>Lista mahdol-</i> <i>lisista toi-</i> <i>menpiteistä</i> <i>Päätökset</i>	Suunnittele käyttöön- otto <i>Käyttöönotto-</i> <i>suunnitelma</i> Suunnittele mallin seu- ranta ja yllä- pito <i>Seuranta- ja</i> <i>ylläpitosuun-</i> <i>nitelma</i> Laadi loppu- raportti <i>Loppuraportti</i> <i>Tulosten jul-</i> <i>kaisu</i> Arvioi pro- jekti <i>Projektin yh-</i> <i>teenveto</i>

Ensimmäisessä eli liiketoiminnan ymmärtämisen vaiheessa selvitetään taustatekijät, jotka vaikuttavat prosessin etenemiseen ja määritellään, mitä projektin aikana halutaan liiketoiminnan näkökulmasta saavuttaa. Usein valitaan yksi tai useampi käyttötapaus tai kysymys, johon lähdetään prosessin aikana hakemaan ratkaisua. Tavoitteiden asettaminen, projekti-suunnitelman tekeminen ja nykytilanteen arviointi kuuluvat mallin ensimmäiseen vaiheeseen. (Markkula & Syväniemi, 2015, s. 96; Smart Vision Europe, 2018)

Datan ymmärtämisen vaiheessa käytötapaukset validoidaan tarkastelemalla, mitä dataa on käytettävissä ja onko se riittävää. Ymmärrys datasta kerätään kyselyillä, raporteilla ja visualisoinneilla ja vaiheen keskeiset tavoitteet ovat datamuotojen ja -sisältöjen ymmärtäminen ja datan laadun selvittäminen. (Markkula & Syväniemi, 2015, s. 96; Smart Vision Europe, 2018)

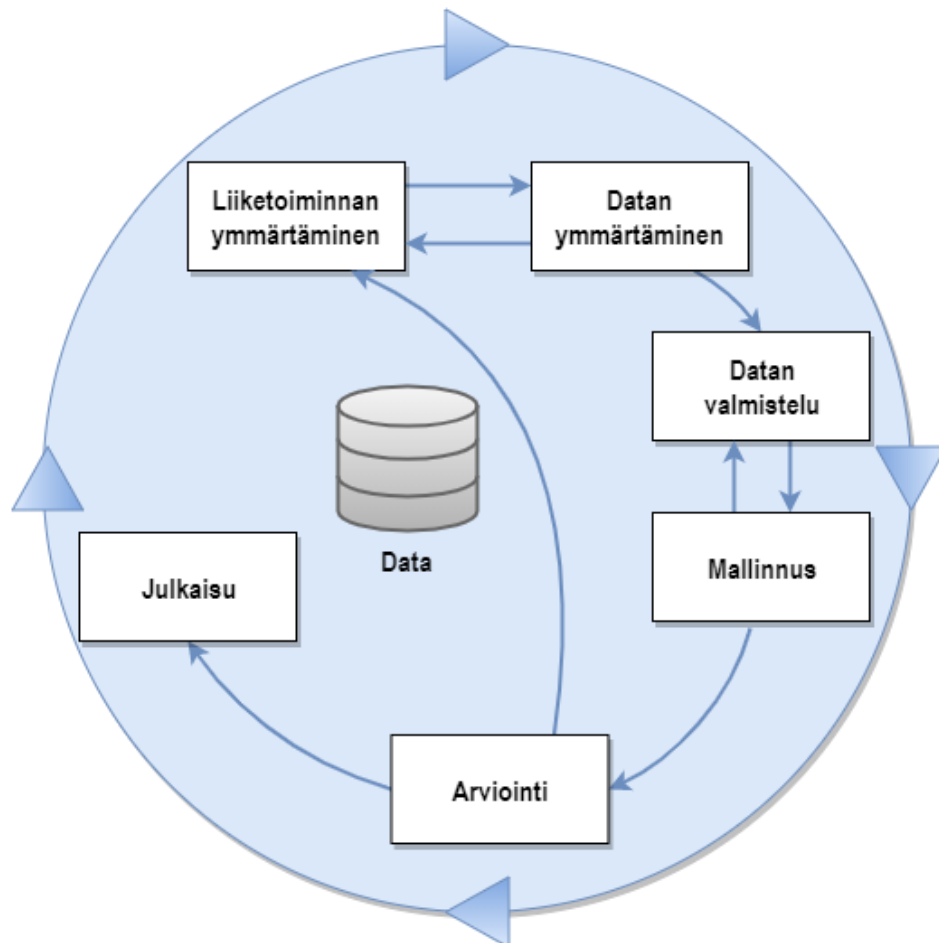
Datan valmisteluvaihe on tulosten oikeellisuuden kannalta keskeinen vaihe ja vie yleensä suurimman osan koko projektin ajasta ja resursseista. Siinä analytiikan työkaluilla dataa käsitellään paremmin mallinnettavaan muotoon. Vaiheeseen liittyy datakokonaisuuksien yhdistämistä, otosten ottamista ja vertailua, tietojen summaamista, uusien muuttujien lisäämistä, tyhjien arvojen poistoa ja testejä. Dataa siis kerätään, siivotaan, yhdistellään ja karsitaan. (Markkula & Syväniemi, 2015, s. 96; Smart Vision Europe, 2018)

Mallinnusvaiheessa analysoidaan valmisteltua dataa tulosten saamiseksi ja tilastollisia ja matemaattisia menetelmiä hyödynnetään liiketoimintaongelman ratkaisemiseksi. Mallinnustekniikan valinta perustuu usein oletuksiin ja mallinnus tehdään iteratiivisesti. Tässä vaiheessa mallia joudutaan tyypillisesti säätämään ja saatetaan palata datan valmisteluvaiheeseen, jotta data on sopivaa valittua mallinnusta varten. (Markkula & Syväniemi, 2015, s. 96; Smart Vision Europe, 2018)

Arviointivaiheessa tarkastellaan saatuja tuloksia ja vertaillaan niitä asetettuihin tavoitteisiin. Mallien oikeellisuus on todettu jo edellisessä vaiheessa, joten tämän vaiheen keskeinen tarkoitus on varmistaa, että tulokset antavat ratkaisun määriteltyyn ongelmaan ja että tavoitteiden sekä saatujen tulosten tarkkuustasovaatimukset täyttyvät. Prosessin onnistumista kokonaisuudessaan voidaan arvioida ja tehdä päätös tulevista toimista. (Markkula & Syväniemi, 2015, s. 96 - 97; Smart Vision Europe, 2018)

Viimeinen vaihe on julkaisuvaihe, jossa prosessin toimivaksi todetut tulokset käyttöön otetaan ja niiden avulla pyritään kehittämään liiketoiminnan operatiivista puolta. Julkaisuvaiheessa tehdään loppuraportti ja suunnitellaan tiedonlouhinnan seuranta ja ylläpito. Siinä tarkastellaan myös, mikä projektissa meni hyvin ja mitä seuraavassa projektissa kannattaa tehdä toisin. (Markkula & Syväniemi, 2015, s. 97; Smart Vision Europe, 2018)

CRISP-DM-mallin iteratiivinen eteneminen esitellään kuvassa 5, jossa edellä kerrottujen vaiheiden välillä voidaan liikkua molempiin suuntiin. Prosessin aikana on mahdollista palata edelliseen vaiheeseen tai jopa takaisin alkuun ja prosessin päättyessä lähdetään mahdollisesti uudelleen alusta liikkeelle.



Kuva 5. CRISP-DM-referenssimallin vaiheet (Chapman et al., 2000)

CRISP-DM-referenssimalli on ollut myöhemmin monen uuden mallin esikuvana. IBM esitteli CRISP-DM:ään pohjautuvan ASUM-DM-mallin (Analytics Solutions Unified Method for Data Mining/Predictive Analytics) vuonna 2015. ASUM-DM-malli etenee CRISP-DM:n mukaisten vaiheiden läpi, mutta tehtäviä ja toimintoja on laajennettu ja eri vaiheisiin on lisätty uusia malleja ja ohjeita. Lisäksi moni analytiikkayritys on kehittänyt omia mallejaan, kuten kansainvälisellä analytiikkaratkaisuja tarjoavalla yrityksellä SAS Institutella on SEMMA-metodi. SEMMA tulee sanoista Sample, Explore, Modify, Model ja Assess eli kokeile, tutki, muokkaa, mallinna ja arvioi. (Bahrepour, 2018; Biecek, 2019; SAS, 2014)

2.6 Microsoft analytiikka- ja BI-tarjoajien joukossa

Analytiikka- ja Business Intelligence-alustoja tarjoavia yrityksiä on lukuisia. Yhdysvaltalainen tutkimus- ja konsultointiyritys Gartner julkaisee säännöllisesti ICT-alan vertailuja, joiden joukossa on kattava analytiikka- ja BI-alustojen vertailu. Vertailussa on mukana modernin analytiikka- ja BI-alustan vaatimukset täyttävät toimittajat. Modernit alustat ovat helppokäyttöisiä työkaluja, jotka kattavat koko analytiikan työkulun datan valmistelusta tiedon visuaaliseen esittämiseen ja ne erottuvat perinteisistä

BI-alustoista sillä, että erilliset tietovarastot ja IT-henkilöstön merkittävä osallistuminen eivät ole välttämättömiä. (Gartner, 2019)

Mukana Gartnerin analytiikka- ja BI-alustojen vertailussa on niin perinteisiä toimittajia kuin uusia tulokkaita. Toimittajia on arvioitu käyttötapausten ja kriittisten ominaisuuksien pohjalta niin it-infrastruktuurin, tiedonhallinnan, analysoinnin ja sisällönhallinnan kuin havaintojen jakamisen näkökulmista. Lisäksi järjestelmien kokonaiskuvaa on arvioitu helppokäyttöisyyden, visuaalisuuden ja hallinnoinnin kannalta. Toimittajat sijoitellaan kuvassa 6 olevaan nelikenttään vertailutulosten perusteella. (Gartner, 2019)



Kuva 6. "Magic Quadrant for Analytics and Business Information Platforms" (Gartner, 2019)

Haastajat (challengers) suoriutuvat hyvin tai voivat hallita suurtakin segmenttiä, mutta eivät täysin tiedosta markkinoiden tulevaisuudensuuntauksia, kun taas johtajat (leaders) ovat erittäin tulevaisuusorientoituneita. Visionääreillä (visionaries) on hyvä ymmärrys siitä, mihin suuntaan markkinat ovat menossa ja heillä voi olla visio markkinoiden muuttamisesta, mutta he eivät kuitenkaan ole kovin menestyneitä. Neljäs

kentän Niche Player-termillä kuvataan pienen segmentin toimijoita, jotka eivät juurikaan innovoi muita toimijoita. (Gartner, n.d.)

Noin puolet vertailun toimittajista sijoittuu nelikentän Niche Player-osioon. Jokaisella on omat vahvuutensa ja monen tuote on arvioitu helppokäyttöiseksi ja analytiikkaominaisuuksiltaan monipuolisiksi, mutta tyypillisesti puutteita on erityisesti alustan toiminnallisuuksissa, asiakastyytyväisyydessä ja markkinatietoisuudessa. Selkeät tulevaisuuden visiot puuttuvat. (Gartner, 2019)

Neljä toimittajaa sijoittui johtajat-osioon. Analytiikka- ja BI-alustojen käyttöönotto on ollut suosittua viime aikoina ja onnistuessaan niiden käyttö laajenee yritysten sisällä, jopa niin voimakkaasti, että niistä muodostuu yritysstandardi. Johtajat-osion toimittajien asiakkaat ovat tyypillisesti erittäin tyytyväisiä ja toimittajat panostavat tuotteensa kehittämiseen. (Gartner, 2019)

Vuonna 2019 Microsoft on ollut 12 vuotta peräkkäin vertailun johtava analytiikka- ja BI -alustojen toimittaja. Microsoftin tuotekartan katsotaan olevan kattava ja näkemyksellinen ja tavoitteena on tuottaa ratkaisu kaikkiin analytiikan tarpeisiin. Käyttö on maailmanlaajuista ja asiakastyytyväisyys korkealla tasolla. (Gartner, 2019; Microsoft, 2019a)

Microsoftin vahvuutena muihin toimijoihin verrattuna on tuotteen edullisuus ja se onkin asettanut muille toimijoille hinnoittelupaineita. Heikkoutena nähdään eroavaisuudet työpöytäsovelluksen ja pilvipalveluna saatavilla olevan sovelluksen välillä. Tietyt ominaisuudet ovat käytettävissä vain toisessa, jolloin ne puuttuvat toisesta. Tämä on aiheuttanut tuotteen kehittämisen hajontaa ja eri ympäristöt seuraavat kehitystä eri tahtia. (Gartner, 2019)

2.7 Microsoft Power BI

Power BI on Microsoftin tarjoama analytiikka-alusta, jossa eri tietolähteistä tuotua dataa voidaan käsitellä ja esittää visuaalisesti ja interaktiivisesti. Power BI on tarjolla pilvipalveluna ja työpöytäsovelluksena. Power BI Desktop on työpöytäsovellus, jolla mm. muodostetaan yhteys tietolähteisiin, yhdistetään ne tietomalliksi, visualisoidaan kuvaajiksi ja muodostetaan julkaistaviksi raporteiksi. Pilvipohjaista versiota kutsutaan Power BI-palveluksi ja se on toiminnoiltaan hieman suppeampi kuin työpöytäversio. Alustaan kuuluu myös mobiiliversio ja Power BI Report Server-palvelu raporttien julkaisua varten. (Gartner, 2019; Microsoft, 2019a)

Power BI Desktopiin voi tuoda dataa yli sadasta erityyppisestä lähteestä. Perinteisten tiedostojen ja yrityksen omien tietovarantojen lisäksi dataa voidaan tuoda mm. lukuisista tietokannoista, online-palveluista, verkkosivuilta, sosiaalisesta mediasta ja pilvipalveluiden tarjoajilta. Tieto voidaan

tuoda joko paikalliseksi kopioksi Power BI:n käsiteltäväksi tai suorana yhteytenä, jossa Power BI lähettää reaaliaikaisesti kyselyjä taustalla olevaan tietolähteeseen. (Aspin, 2018, s. 2; Knight, Knight, Pearson, Quintana, & Powell, 2018, s. 11; Microsoft, 2019b, 2019c)

Datan käsittelyyn on tarjolla laajasti vaihtoehtoja. Yksinkertaisimmillaan dataa siivotaan ja sitä käsitellään tietoja yhdistelemällä, erottelemalla ja uudelleenmuotoilemalla sarakkeissa tai luomalla kokonaan uusien sarakkeita esimerkiksi DAX-kielen funktioiden ja laskutoimitusten avulla. Monimutkaisempia toimintoja varten Power BI:ssa voi hyödyntää R-ohjelmointikieltä, joka on tilastolliseen laskentaan ja grafiikan tuottamiseen tarkoitettu kieli. Power BI:hin tuodut tiedot esitetään tauluina, joiden välille luodaan suhteita. (Aspin, 2018, s. 2; Knight et al., s. 43, 2018; R Foundation, n.d.)

Power BI:n visualisoinnit ovat monipuolisia ja vuorovaikutteisia. Käytettävissä on Excelistä tuttujen taulukoiden ja kaavioiden lisäksi laaja valikoima muita kaavioita sekä karttoja, matriiseja ja analyytiikkatyökaluja. Visualisointeja voidaan helposti suodattaa ja pilkkoa. Power BI -raportit voidaan optimoida mobiililaitteille käyttämällä Power BI Desktopin puhelinasettelu-toimintoa. (Aspin, 2018, s. 3; Powell, 2018, s. 267)

Monella suomalaisella yrityksellä ja kunnalla on käytössään Power BI:llä toteutettuja koonteja, jotka on julkisesti jaettu heidän verkkosivuillaan. Julkishallinnon organisaatioista esimerkiksi Suomen Pankilla (n.d.) on nähtävillä visualisointi suomalaisten lainoihin ja talletuksiin liittyen. Kaupungeista esimerkiksi Tampereen kaupunki (n.d.) seuraa vuoteen 2030 tähtäävää strategiaa ja Kirkkonummen sivuilta (n.d.) löytyy vuosien 2017 ja 2018 ostolaskut Power BI -visualisoinnin avulla. Espoon kaupungin (n.d.) kunnallistekniikan investointeja ja käyttötalousbudjettia pääsee myös tarkastelemaan reaaliajassa.

Power BI nähdään helppokäyttöisenä myös monimutkaisissa analysoinneissa. Tuotetta kehitetään jatkuvasti ja Microsoftin oman vision mukaan Power BI:hin tullaan jatkossakin tekemään parannuksia varsinkin analyytikatoimintoihin. Uusia koneoppimista ja tekoälyä hyödyntäviä ominaisuuksia tuodaan saataville helppossa muodossa myös käyttäjille, joilla ei ole niin vahvaa tietoteknistä osaamista. (Gartner, 2019)

3 DATA VALKEAKOSKEN SEUDUN KOULUTUSKUNTAYHTYMÄSSÄ

Opinnäytetyön toimeksiantaja on Valkeakosken seudun koulutuskuntayhtymä, joka omistaa ja ylläpitää Etelä-Pirkanmaalla toimivaa Valkeakosken ammattiopistoa. Ammattiopistossa opiskelee vuosittain noin 1300 opiskelijaa ja henkilökuntaa on noin 110. Koulutuskuntayhtymän tehtävänä on järjestää ammatillista koulutusta jäsenkuntiensa puolesta. Jäsenkunnat ovat Akaa, Kangasala, Lempäälä, Pälkäne, Urjala ja Valkeakoski. Valkeakosken ammattiopistossa suoritetaan perus-, ammatti- ja erikoisammattitutkintoja useilta eri aloilta sekä lisäksi lyhytkestoisia koulutuksia. Toimintaa ohjaavat useat kunta- ja opetusalan lait ja asetukset.

Tiedolla johtaminen on keskeinen ja ajankohtainen organisaation kehittämisen teema Valkeakosken seudun koulutuskuntayhtymässä. Askeleita mittareiden seurannan automatisointiin on otettu vuonna 2019 opiskelijahallintojärjestelmän tiedonkeruun ja reaaliaikaisen esittämisen käyttöönotolla uudessa raportointijärjestelmässä. Lähitulevaisuuden tavoitteena on se, että tiedolla johtamisen prosessit ovat selkeitä ja tarvittava tieto on löydettävissä mahdollisimman helposti.

3.1 ISO 9001:n mukainen laatujärjestelmä

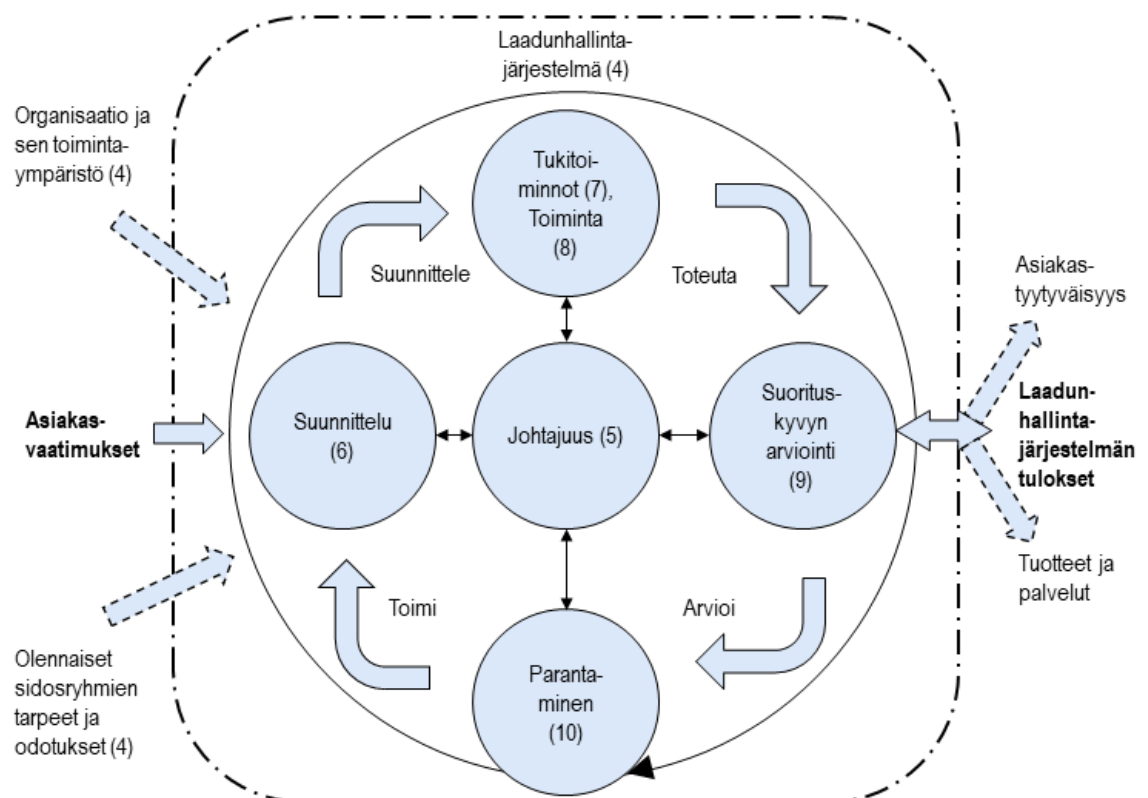
Valkeakosken seudun koulutuskuntayhtymän tiedolla johtamisen perustana ja laatujärjestelmän runkona on yrityksen intranet-sivuilla julkaistu toimintakäsikirja, joka on koonti yrityksen yhteisistä toimintatavoista. Se sisältää kuvaksen toimintajärjestelmästä ja laadunhallinnan periaatteista. Lisäksi toimintakäsikirjaan linkittyy mm. tilastoja, toimintaohjeita ja prosessikuvauksia. Toimintakäsikirjassa on kuvaukset organisaatiosta, toimintaympäristöstä, asiakkaista ja sidosryhmistä. (VAAO intranet, 2018)

Laatujärjestelmä on toteutettu SFS-EN ISO 9001:2015-standardin mukaisesti ja se on sertifioitu vuonna 2018. Samana vuonna kuntayhtymä sai myös kunniainimaininnan opetus- ja kulttuuriministeriön lautupalkintokilpailussa (OKM, 2018), jossa teemana oli opiskelijoiden ja henkilöstön hyvinvointi. Nämä kertovat siitä, että laatuasiat ovat keskeisessä asemassa organisaation toiminnassa.

Laatujärjestelmän perustana oleva ISO 9001-standardi on työkalu yrityksen laadunhallintajärjestelmän rakentamiseen ja kehittämiseen. Standardissa on seitsemän päälukua, joissa määritellään laadunhallintajärjestelmää toteutettaessa huomioitavat toiminnot. Pääluvut ovat organisaation toimintaympäristö, johtajuus, suunnittelu, tukitoiminnot, operatiivinen toiminta, suorituskyvyn arviointi ja parantaminen. (SFS, n.d.)

ISO 9001-standardissa noudatetaan prosessimaista toimintamallia, johon yhdistyy kuvan 7 mukaisesti PDCA-malli (plan, do, check, act eli suunnittele, toteuta, arvioi, toimi) sekä riskiperusteinen ajattelu. PDCA-mallin

suunnitteluvaiheessa (plan) asetetaan tavoitteet järjestelmälle ja sen prosesseille ja määritellään tarvittavat resurssit, joilla tulokset voidaan saavuttaa asiakkaiden vaatimusten ja organisaation oman toimintapolitiikan mukaisesti sekä tunnistetaan ja käsitellään riskit ja mahdollisuudet. Seuraavassa eli toteutus (do) vaiheessa nimensä mukaisesti suunnitelmat toteutetaan. Kolmantena on arviointi (check) vaihe, jossa seurataan ja tarvittaessa mitataan prosesseja, niistä syntyviä tuotteita ja palveluja ja verrataan niitä toimintapolitiikkaan, tavoitteisiin, vaatimuksiin ja suunniteltuihin toimintoihin. Neljännessä vaiheessa, eli toimi (act), ryhdytään tarvittaviin toimenpiteisiin, joilla parannetaan suorituskykyä. (SFS/2015, s. 7 - 8)



Kuva 7. ISO 9001-standardin rakenne PDCA-mallin (suunnittele, toteuta, arvioi, toimi) mukaisesti. Suluissa viittaus standardin kohtaan. (SFS 9001/2015, s. 9)

ISO 9001-sertifikaatti on voimassa kolme vuotta ja voimassaolon edellytyksenä on määräaikaisarviointien suorittaminen ja havaittujen poikkeamien sulkeminen.

3.2 Mittaaminen ja tiedolla johtaminen

Uusi laki ammatillisesta koulutuksesta astui voimaan vuoden 2018 alusta ja sen myötä rahoituksen perusteet muuttuivat. Tämä korosti jatkuvan seurannan ja tiedolla johtamisen merkitystä. Aiemmin tiettyjen tekijöiden seurannan sykliseksi on riittänyt jopa vuositasoinen tarkastelu, kun nyt lähes kaikki mittarit ovat jatkuvasti seurattavia.

Tällä hetkellä mittarit pohjautuvat suureksi osaksi rahoitukseen vaikuttaviin tekijöihin. Mitattavia asioita ovat esimerkiksi kertyneet opiskelijavuodet, suoritettujen tutkintojen ja suoritusten määrä sekä erilaisten palautteiden ja kyselyiden tulokset. Myös talouteen ja henkilöstöön liittyviä mittareita on määritelty.

Mittareiden tietolähteet ovat pääasiassa sisäiset talous-, henkilöstö- ja opiskelijahallinnon järjestelmät ja sisäisesti kerättävät säännölliset kyselyt, mutta myös ulkoisia tietolähteitä on mukana. Opiskelijoilta kerätään valtakunnallisesti palautetta opintojen alku- ja valmistumisvaiheessa ja työpaikalla järjestettävästä koulutuksesta sekä opiskelija että työpaikkaohjaaja vastaavat kyselyyn. Näiden kyselyiden vastaukset ovat nähtävillä julkisesti opetushallinnon tilastopalvelusta.

Tiedolla johtamista tukemaan on hankittu vuonna 2019 raportointijärjestelmä, jonne parhaillaan tuodaan automaattisesti tietoa opiskelijahallintojärjestelmästä tiedonsiirtoa varten perustetun tietokannan kautta. Muiden mittareiden osalta päivitys on toistaiseksi manuaalista. Raportointijärjestelmässä on aloitussivulla tulokorttinäkymä, josta nähdään kaikkien mittareiden nykytilanne kootusti. Jokaisesta mittarista aukeaa uuteen ikkunaan mittarikohtainen seuranta, josta voi avata vielä tarkemman jaotteen eri taustatekijöiden perusteella. Esitystapana on allekkain olevia pylväskaavioita ja suodattimia voi lisätä erillisen ikkunan kautta. Raportointijärjestelmä on koko henkilöstön saatavilla.

Käytettävyyden kannalta tulokorttinäkymä on selkeä, mutta kaikki muu toiminta vaatii useamman klikkauksen ja siirtymisen ikkunasta toiseen. Kaikki olennainen tieto on kuitenkin löydettävissä ja järjestelmän käyttöönotto on tuonut selkeän parannuksen entisiin toimintatapoihin. Kaikkia järjestelmän tarjoamia ominaisuuksia ei vielä ole otettu käyttöön ja kehitystyötä tehdään parhaillaan. Tavoitteena on muidenkin tietolähteiden automatisointi mahdollisuuksien mukaan.

Tiedolla johtamisen prosessissa esimiehet seuraavat oman alansa tilannetta kuukausittain ja tietyiltä osin vähintään kvartaaleittain. Esiin tulevat havainnot tulee tuoda esiin välittömästi. Seuranta dokumentoidaan kvartaaleittain kyselylomakkeen avulla. Esimiestason lisäksi seuranta ja seurannan dokumentointia toteutetaan tulosalueiden ja koko organisaation laajuisesti.

Valkeakosken seudun koulutuskuntayhtymässä on havaittu muitakin tietolähteitä, joita voitaisiin lähitulevaisuudessa hyödyntää. Toiminnan parantamisen menetelmiä ovat sisäiset ja ulkoiset auditoinnit, katselmukset, itsearviointit sekä poikkeaminen havaitseminen ja käsitteleminen sekä niihin kohdistettujen muutosten vaikuttavuuden seuranta. Nämä kaikki tapahtumat dokumentoidaan. Intranetsivuille tallennetaan monenlaisia seurantatietoja, Excel-tiedostoja, palautteita ja asiakirjoja. Tällä hetkellä käytössä ei ole digitaalista arkistoa, mutta Laki julkisen hallinnon

tiedonhallinnasta (906/2019) astuu voimaan vuoden 2020 alussa ja takaa sen, että jatkossa yhä suurempi osa aineistosta tulee olemaan saatavilla sähköisesti. Tämä voi mahdollistaa analytiikan laajemman hyödyntämisen organisaation toiminnassa ja tiedolla johtamisessa tulevaisuudessa.

3.3 Kehittämiskohteet

Raportointijärjestelmä on ollut ensimmäinen askel jatkuvan seurannan automatisoinnissa ja edistämisessä. Tämän opinnäytetyön tarkoituksena on tuottaa taustatietoa tiedolla johtamisen kehittämiseksi edelleen. Haasteena tiedolla johtamisen jalkauttamisessa tiedetään olevan sen, että tieto on edelleen hajallaan monessa paikassa ja sen hyödyntäminen vaatii usein käyttäjältä toimenpiteitä ennen kuin juuri hänen tarvitsema tieto on esillä. Kaiken tiedon tulisi mahdollisuuksien mukaan olla yhdessä paikassa ja sellaisessa muodossa, että käyttäjän ei tarvitse tehdä sen eteen ylimääräisiä ponnistuksia.

Raportointijärjestelmä ei työvälineenä ole kovin visuaalinen ja vuorovai-
kutteinen, mutta se on toimiva johdon tarpeisiin ja tiedon esittämiseen eri tahoille. Heikkoutena voidaan mainita se, että uusien mittareiden, suodat-
timien tai tietoyhteyksien rakentaminen vaatii aina toimenpiteitä järjestel-
mätoimittajalta, joten kehitystyö ei ole täysin omissa käsissä. Toisaalta
kaikkia järjestelmän tarjoamia ominaisuuksia ei ole vielä otettu käyttöön
ja etuna voidaan nähdä se, että kehittämistarpeita tulee järjestelmätoimit-
tajalle myös muilta heidän asiakkailtaan ja niiden myötä työkalun kehittä-
minen on jatkuvaa.

Valkeakosken seudun koulutuskuntayhtymällä on käytössään Microsoftin
ohjelmistopaketti, jossa Power BI on automaattisesti tarjolla ilman erillistä
veloitusta. Tällä hetkellä ohjelmistoa ei vielä juurikaan hyödynnetä. On ym-
märretty, että asioiden ennakoimiseksi ja nopeaan reagointiin varautu-
miseksi analytiikkaa ja tarkempaa taustatekijöiden kartoittamista tarvitaan
raportoinnin tueksi. Tiedetään, että Power BI:hin voidaan yhdistää suoraan
erilaisia tietolähteitä ja ohjelmisto on tunnettu analytiikkaominaisuuksis-
taan. Power BI:n visuaaliset koontinäytöt voivat tukea raportointijärjestel-
män esitystapaa tiedon jakamisessa. Raportointijärjestelmää ei ole tarkoi-
tus korvata Power BI:lla, mutta halutaan selvittää, miten Power BI voi
tuoda lisäarvoa jo kerättävälle datalle. Raportointijärjestelmän taustalle
rakennettu tietokanta voidaan jatkossa yhdistää suoraan Power BI:hin yh-
dyskäytävän avulla, mutta tämä osuus on rajattu opinnäytetyön ulkopuo-
lelle. Ensin selvitetään Power BI:n käyttömahdollisuuksia ja sen pohjalta
suunnitellaan Power BI:n laajempaa hyödyntämistä tiedolla johtamisessa.
Tutustuminen Power BI:hin toteutetaan konkreettisen case-esimerkin
avulla.

4 TIEDONLOUHINTA OPISKELIJADATASTA

Power BI:n analytiikkaominaisuuksiin tutustutaan toimeksiantajan opiskelijahallintojärjestelmän datan avulla. Sekä SMART- että CRISP-DM-mallien mukaan datan käsittelyn lähtökohtana tulee olla liiketoiminnallinen tavoite ja selkeä kysymys, johon halutaan vastauksia. Selvityksen kohteena on opiskelijoiden opintojen keskeyttäminen, sillä keskeyttämisten määrä ja syyt on yksi toimeksiantajan tiedolla johtamisen mittareista. Tieto keskeyttämisten lukumääristä päivittyy raportointijärjestelmään päivittäin ja nyt opiskelijadatasta selvitetään taustatekijöitä opintojen keskeyttämisen taustalla.

Tiedonlouhinnan tuloksia on tarkoitus hyödyntää keskeytysten ennakkoinnissa. Työn edetessä Power BI:n toiminnallisuuksia voidaan verrata toimeksiantajan raportointijärjestelmään. Työn perusteella saadaan tietoa Power BI:n tarjoamista mahdollisuuksista ja voidaan jatkossa kehittää tiedolla johtamisen prosesseja. Opinnäytetyön tietoperustassa on kuvattu datan merkitystä liiketoiminnalle ja toiminnallinen osuus tuo esiin nostettuja asioita käytäntöön.

Tiedonlouhinta etenee taulukossa 4 kuvatun CRISP-DM-mallin vaiheiden mukaisesti. Tällä varmistetaan projektin huolellinen läpivienti ja vähennetään virheiden mahdollisuutta. Tämän kappaleen alaotsikot on nimetty CRISP-DM:n vaiheiden mukaan.

4.1 Liiketoiminnan ymmärtäminen

CRISP-DM:n ensimmäisessä vaiheessa määritetään liiketoiminnalliset tavoitteet, arvioidaan tilanne, määritetään tiedonlouhinnan tavoitteet ja laaditaan projektisuunnitelma. Valkeakosken seudun koulutuskuntayhtymän laatuja järjestelmässä yrityksen perustehtäväksi on määritelty osaajien kouluttaminen työelämään ja yhteiskuntaan ja kansallisena koulutuksen järjestämisen tavoitteena on se, että opiskelijat saavuttavat omat tavoitteensa ja suorittavat tutkinnon. Valitettavasti osa opiskelijoista päätyy syystä tai toisesta opintojen keskeyttämiseen. Tilastokeskuksen (2019) keräämän Suomen virallisen tilaston mukaan 7,4 % nuorille suunnatun ammatillisen koulutuksen opiskelijoista keskeytti opinnot lukuvuonna 2016-2017.

Liiketoiminnallinen tavoite on se, että opintojen keskeyttämisiä tulee mahdollisimman vähän. Keskeytyksiä halutaan ennakoida ja asioihin reagoimiseksi tutkitaan, löydetäänkö keskeyttäneiden opiskelijoiden taustatiedoista yhteneväisyyksiä. Tiedonlouhinnan kysymykseksi asetetaan siksi, miten Power BI:n avulla voidaan selvittää opiskelijahallintojärjestelmään tallennetuista tiedoista opintojen keskeyttämiseen johtavia syitä.

Valkeakosken seudun koulutuskuntayhtymän opiskelijahallintojärjestelmänä on Visma Oy:n toimittama Primus. Ennen datan keräämistä tiedetään, että opiskelijoita on opiskelijahallintojärjestelmässä kolmessa erillisessä ympäristössä. Ennen 1.1.2018 voimaan tullutta lakia ammatillisesta koulutuksesta (531/2017) järjestelmässä oli erilliset ympäristönsä nuorten ja aikuisten ammatilliselle koulutukselle. Kolmas ympäristö otettiin käyttöön, kun uusi lainsäädäntö astui voimaan. Datan kannalta haasteen aiheuttaa eri ympäristöjen väliset rakenteelliset eroavaisuudet ja vaihtelu kirjaustavoissa. Toimintatavat ovat muuttuneet vuosien varrella ja lainsäädännön tuomat tutkintorakenteen muutokset ovat lisänneet datan pirstaleisuutta. Esimerkki perustutkinnon laajuus on vaihtunut opintoviikoista osaamispisteiksi vuoden 2015 lakimuutoksessa (245/2015) ja aikuisten ammatillisessa näyttöperusteisessa aikuiskoulutuksessa opintojen laajuus on ilmoitettu tunteina. Viimeisimmän 2018 lakimuutoksen myötä kaikilla laajuutena on osaamispisteet.

Otannaksi valitaan perustutkintoa suorittavat opiskelijat, jotka ovat aloittaneet opintonsa 1.8.2016 - 31.7.2019. Joukosta rajataan ulkopuolelle ennen 1.1.2018 aloittaneet näyttötutkintoon valmistavan koulutuksen opiskelijat eli ns. aikuiskoulutuksen opiskelijat, sillä heidän kohdalla tutkinnon rakenne ja järjestelmään tehdyt kirjaukset poikkeavat huomattavasti muusta joukosta. Koulutuksen kesto on heillä ollut lyhyempi ja kaikkia taustatietoja ei ole kirjattu. Valitulla otannalla saadaan vertailukelpoisempaa ja luotettavampaa tietoa taustatekijöistä.

Data poimitaan opiskelijahallintojärjestelmästä Excel-tiedostoon, sillä yhdyskäytävän rakentaminen tietokantaan rajattiin projektin ulkopuolelle. Datan pirstaleisuus huomioidaan mahdollisuuksien mukaan opiskelijahallintojärjestelmän hakulausekkeissa, joiden avulla dataa ajetaan Excel-tiedostoihin. Näin vähennetään mahdollisuuksien mukaan datan siivousta Power BI:ssä. Excelissä voidaan vielä käydä läpi ja puhdistaa saatua datamassaa ennen taulukkojen noutoa Power BI:hin. Power BI:ssä data muotoillaan lopulliseen muotoonsa ja tehdään tarvittaessa uusia sarakkeita ja laskentaa kerätystä tiedosta. Lopuksi käytetään joko yhtä tai useampaa visualisointia tulosten esittämiseksi.

Projektin myötä selviää, onko opiskelijahallintojärjestelmän data käyttökelpoista sellaisenaan analysointien tekemiseen ja onko taustatietoa ylipäättään tallennettuna riittävästi. Samalla voidaan pohtia, kannattaako keskeytyksiin liittyviä tekijöitä seurata jatkuvasti, vai onko kyseessä kertaluonteinen analyysi tulevaisuuden toimia hyödyntämään. Projektin perusteella voidaan lopulta päättää, kannattaako yhdyskäytävä tehdä Power BI:n ja tietokannan välille.

4.2 Datan ymmärtäminen

Toisessa CRISP-DM:n vaiheessa kerätään alustava data, kuvataan ja tutkitaan sitä ja varmistetaan datan laatu. Datan valinnassa ja keräämisessä

opiskelijahallintojärjestelmästä käytetään ennakkokäsityksiin perustuvaa harkintaa ja ymmärrystä siitä, mitä tietoa on ylipäättään mahdollista kerätä. Kerättävä data on kuvattu taulukossa 5. Opiskelijoiden tiedot ovat järjestelmässä opiskeluoikeuksittain, ei henkilöittäin. Toisin sanoen yhdestä opiskelijasta voi olla järjestelmässä monta rekisterikorttia, jos hän on opiskellut useamman kerran tai useampaa tutkintoa oppilaitoksessa. Jokaisessa opiskeluoikeudessa on kirjattuna suoritettava tutkinto, päivämäärätiedot opiskelun aloittamisesta ja päättämisestä, tehdyt suoritukset ja opiskelijan taustatiedot. Opiskelijoille kirjataan tuntimerkintöihin läsnä- ja poissaolotietoja sekä mahdollisia järjestyssääntörikkomuksia. Mikäli opinnot keskeytyvät väliaikaisesti, tieto keskeytyksen ajasta kirjataan opiskeluoikeuteen. Tietosuojasyistä opiskelijoiden henkilötietoja ei oteta mukaan eli henkilötunnuksen sijaan data yksilöidään opiskelijaoikeuksittain rekisterikortin ID:n perusteella. Nimitietoa ei tuoda lainkaan, mutta opiskelijahallintojärjestelmän salaama henkilötunnus tuodaan, jotta voidaan tarvittaessa laskea erillisten opiskelijoiden määrä.

Taulukko 5. Opiskelijahallintojärjestelmästä kerättävä data.

Data	Tyyppi	Kuvaus ja ennakoitu käyttötapa
Opiskelijakortin ID	Número (5 merkkiä)	Opiskeluoikeutta yksilöivä numero. Tämän kentän avulla eri taulukoissa oleva tieto voidaan yhdistää.
Salattu henkilötunnus	Teksti (32 merkkiä)	Järjestelmän luoma yksilöllinen merkkijono, jossa on sekä numeroita että tekstiä. Mahdollistaa yksittäisten opiskelijoiden määrän laskemisen.
Syntymäaika	Päivämäärä	Opiskelijan iän laskemiseksi
Sukupuoli	Teksti	
Kotikunta	Teksti	Eroavaisuudet kunnittain
Valintatapa	Teksti	Opiskeluihin hakeutumisen tapa, esim. yhteishaku tai sisäinen siirto
Pohjakoulutus	Teksti	
Tutkinto	Teksti	Suoritettavan tutkinnon nimi.
Aloituspäivä	Päivämäärä	
Valmistumispäivä	Päivämäärä	
Keskeyttämispäivä	Päivämäärä	Analysoinnin tärkeä taustatieto, jonka perusteella voidaan myös päätellä, onko keskeyttäminen tosi vai epätosi.
Keskeytyssyy ID	Kokonaisluku	Keskeyttämisen syykoodin ID, yhdistetään rekisteriin, josta tiedon saa selkokielisenä.
Väliaikaiset keskeytykset	Pvm tai Tosi/epätosi	Tieto opintojen aikana olleista väliaikaisista keskeytyksistä.
Väliaikaiset keskeytyksen syy	Kokonaisluku	Kuten edellä eron syy.
Poissaolojen määrä	Kokonaisluku	Opiskelijan tuntimerkintöihin merkityt poissaolot.
Järjestyssääntörikkomukset	Tosi/epätosi	Opiskelijan tuntimerkintöihin merkityt järjestyssääntörikkomukset.
Suoritettujen tutkinnon osien lukumäärä	Kokonaisluku	Jos halutaan selvittää suoritusten lukumäärä.
Suoritettujen tutkinnon osien osaamispistemäärä	Desimaaliluku	Suoritusten kokonaislaajuus.
Erityinen tuki	Tosi/epätosi	Onko opiskelija tarvinnut opinnoissaan erityistä tukea?

Tiedot sijaitsevat opiskelijahallintojärjestelmässä useassa eri rekisterissä ja Power BI:hin siirrettäviä taulukkoja syntyy yhteensä kahdeksan. Datan poiminnassa tulee huomioida se, että lakimuutokset ja koulutuksen järjestäjän käytänteissä tapahtuva vaihtelut ovat aiheuttaneet kirjaamisteknisiä muutoksia järjestelmässä vuosien varrella. Esimerkkinä kuvasta 8 nähdään, että opiskelijan läsnä-/poissaoloon liittyvät tiedot on tallennettu aiemmin lukuvuosittain taulukkoon ja erilliseen keskeytystaulukkoon, kun taas nyt sama tieto kirjataan yhteen taulukkoon aikajaksoina. Tämä eroavaisuus on mahdollista huomioida opiskelijahallintojärjestelmässä tehtävissä hakulausekkeissa.

Vanha tapa:

Tilapäisen kesk. alkupäiv	Tilapäisen kesk. päättymisp	Tilapäisen kesk. syy
13.08.2012	12.08.2013	Varusmies- tai siviilipalvelus
Eron syy (rekisteristä)		
muu henkilökohtainen syy		
Lukuvuosi	Syksy 20.9.	Kevät 20.1.
2012-2013	Poissa	Poissa
2013-2014	Läsnä	Läsnä

Uusi tapa:

Pvm alkaen	Pvm saakka	Opiskelu-oikeuden tilat	Rahoitusmuoto	Osa-aikaisuus	Keskeytyksen syy	Eron syy
08.08.2018	18.11.2018	Läsnä	1 Valtionosuusrahoitteinen koulutus	100		
19.11.2018	06.01.2019	Väliaikaisesti keskeytynyt	1 Valtionosuusrahoitteinen koulutus	100	Muu henkilökohtainen syy	
07.01.2019	30.01.2019	Läsnä	1 Valtionosuusrahoitteinen koulutus	100		
31.01.2019		Katsotaan eronneeksi	1 Valtionosuusrahoitteinen koulutus	100		Siirtynyt työelämään

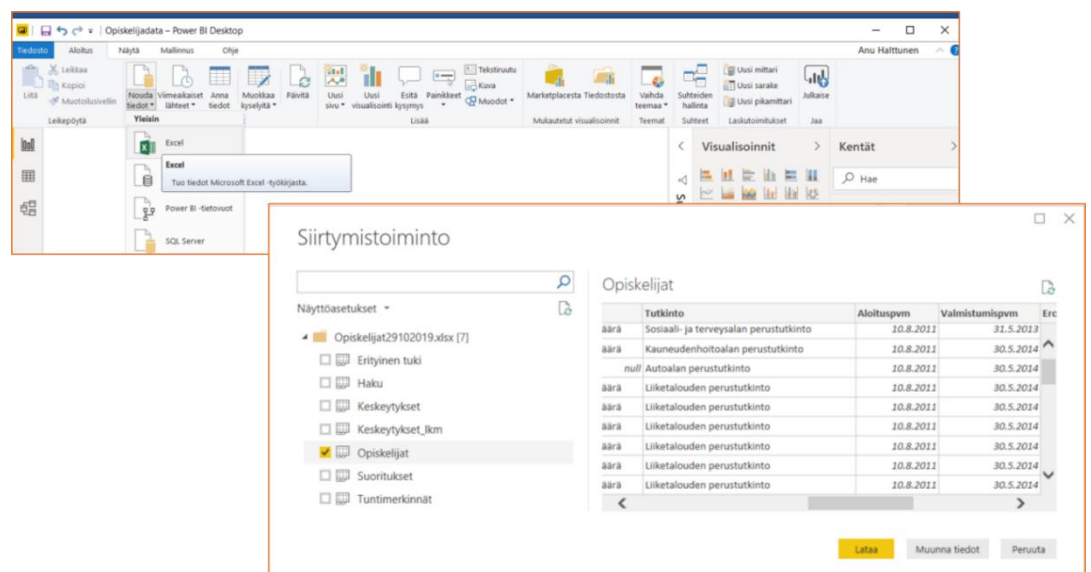
Kuva 8. Opiskelijan läsnäoloon liittyvät tiedot opiskelijahallintojärjestelmässä.

Tulee myös kiinnittää huomiota siihen, että tuntimerkintöjen luokittelut ovat muuttuneet vuosien varrella useasti ja poissaolokoodeja on matkan varrella lisätty ja poistettu. Samaa on tapahtunut keskeytyksen syille ja eri aikakausina on käytetty erilaisia selitteitä. Raporttiin otetaan mukaan kaikki eri vaihtoehdot, mutta on hyvä pitää mielessä, että ohjeistukset poissaolojen merkitsemiseen ja eron syiden kirjaukseen ovat saattaneet muuttua ja voidaan olettaa, että tältä osin dataan tulee suhtautua pienellä varauksella.

Tukitoimiin liittyvien tapahtumien yhteyttä keskeyttämisten määrään olisi mielenkiintoista selvittää, mutta valitettavasti raportoitavaa tietoa ei ole juurikaan opiskelijahallintojärjestelmässä kerättävissä. Ainoastaan päätökset ammatillisen koulutuksen laissa määritellyn HOJKS:in eli henkilökohtainen opetuksen järjestämisestä koskeva suunnitelman (630/1998) ja erityisen tuen suunnitelman (531/2017) laatimisista löytyy merkintä joko yksittäisestä päivämääräkentästä, erityisopiskelijan ryhmää kuvaavasta valikkokentästä tai suunnitelmaa koskevan päätöksen alkamis- ja päättymispäiväkentistä. Koska tieto on näin hajallaan, tarkastellaan edellä mainittuja kenttiä ja päätellään niiden perusteella, onko suunnitelma tehty. Merkitään numerolla 1, mikäli merkinnät viittaavat siihen, että suunnitelma on tehty ja 0, mikäli ei ole tehty. Tämä tieto muutetaan Power BI:ssä tietotyy-piksi tosi/epätosi.

4.3 Datan valmistelu

Datan valmisteluun kuuluu CRISP-DM:n kolmannessa vaiheessa datan valintaa, siivousta, rakentamista, integrointia ja formatointia. Datan valmistelu toteutetaan Power BI:ssä, johon opiskelijahallintojärjestelmästä tuotetut Excelit tuodaan kuvan 9 mukaisesti Nouda tiedot -toiminnolla. Avautuvassa ikkunassa valitaan taulukot, jotka halutaan tuoda Power BI:hin. Tämän jälkeen niitä käsitellään Power Query -näkyymässä. Sarakkeiden tietotyypit tarkistetaan ja muutetaan tarvittaessa sisältöä vastaaviksi. Lisäksi sarakkeiden sisältöä voidaan muokata, poistaa tarpeettomia tai uusia sarakkeita voidaan lisätä.



Kuva 9. Tietojen noutaminen Excelistä ja taulukkojen valitseminen.

Ensivaiheessa kerätyssä datassa on siis mukana opintonsa aloittaneet opiskelijat ajalla 1.8.2016 - 31.7.2019, jolloin otanta on 1311 opiskeluoikeutta. Tarkemmassa datan tutkinnasta ja alustavista visualisoinneista kuitenkin todetaan, että rajausta näin lyhyelle aikajaksolle ei ole perusteltua ja Power BI:n analyysitoimintojen hyödyntämiseksi taustajoukkoa on syytä laajentaa. Lopulliseen aineistoon aikajaksoa lisätään viidellä vuodella ja lopullisen tiedonlouhinnan kohteeksi otetaan 1.8.2011 – 31.10.2019 aloittaneet perustutkinto-opiskelijat, jolloin otanta on 3555 opiskeluoikeutta. Mikäli vain tuoreempia tietoja halutaan tarkastella, voidaan Power BI:n suodattimien ja osoittimien avulla tehdä tarvittaessa aikaan liittyviä rajoituksia.

Aikajakson muutoksen myötä huomataan, että tiedonkeruussa tulee nyt huomioida muutamia lisäseikkoja ja datan valmistelua tulee jatkaa Power BI:ssä. Suoritusten osalta jo aiemmin todettiin, että vuoteen 2015 asti on käytössä ollut opintoviikot ja sen jälkeen osaamispisteet. Ennen vuotta 2015 ammatillisen perustutkinnon laajuus oli 120 opintoviikkoa ja sen jälkeen 180 osaamispistettä. Paremman vertailtavuuden vuoksi tehdään uusi

sarake Suoritukset-taulukkoon, jossa ennen vuotta 2015 aloittaneiden opiskelijoiden suoritukset muunnetaan osaamispisteiksi kertoimella 1,5. Tähän käytetään Power BI:n DAX-funktiota kaavan 1 mukaisesti.

Suoritukset osaamispisteinä = IF(Suoritukset[Osaamispisteitä]=0;
Suoritukset[Opintoviikkoja]*1,5;Suoritukset[Osaamispisteitä]) (1)

Analystointia varten Opiskelijat-taulukkoon lisätään useita uusia sarakkeita. Kaavan 2 mukaisesti lasketaan opiskelijoiden opintojen kokonaiskesto kuukausina aloituspäivästä joko valmistumis- tai keskeyttämispäivään. Mikäli opinnot eivät ole vielä päättyneet, lasketaan kesto raportointiajankohdan päättymispäivään eli 31.10.2019. Kaavassa on kaksi sisäkkäistä JOS-lauseketta, joissa tarkastellaan, onko päivämäärään viittaavat kentät tyhjiä ja sen perusteella lasketaan DATEDIFF-funktiolla kuukausien määrä päivämäärien välissä.

Kesto kuukausina = IF(NOT(ISBLANK(Opiskelijat[Eropvm]));
(DATEDIFF(Opiskelijat[Aloituspvm];Opiskelijat[Eropvm];MONTH));
IF((NOT(ISBLANK(Opiskelijat[Valmistumispvm]));
(DATEDIFF(Opiskelijat[Aloituspvm];Opiskelijat[Valmistumispvm];MONTH));
DATEDIFF(Opiskelijat[Aloituspvm];"31.10.2019";MONTH))) (2)

Vastaavalla logiikalla lasketaan opiskelijan ikä syntymäpäivästä joko valmistumis-, keskeyttämis- tai raportointiajankohdan päättymispäivään kaavan 3 mukaisesti.

Ikä viim läsnäolopvm = IF(NOT(ISBLANK(Opiskelijat[Eropvm]));
(DATEDIFF(Opiskelijat[Syntymäaika];Opiskelijat[Eropvm];YEAR));
IF((NOT(ISBLANK(Opiskelijat[Valmistumispvm]));
(DATEDIFF(Opiskelijat[Syntymäaika];Opiskelijat[Valmistumispvm];YEAR));
DATEDIFF(Opiskelijat[Syntymäaika];"31.10.2019";YEAR))) (3)

Kuvassa 10 lisätään uusi ehdollinen sarake, jossa opiskelijan opintojen tila päätellään valmistumis- ja eropäivämääristä, jolloin lopputulemana on joko Valmistunut, Keskeyttänyt tai Opinnot jatkuu.

Lisää ehdollinen sarake

Lisää ehdollinen sarake, joka on laskettu muista sarakkeista tai arvoista.

Uuden sarakkeen nimi
Tila

Sarakkeen nimi	Operaattori	Arvo	Tuloste
Jos	Eropvm	ei ole yhtä suuri k...	Keskeyttänyt
Muut...	Valmistumispvm	ei ole yhtä suuri k...	Valmistunut

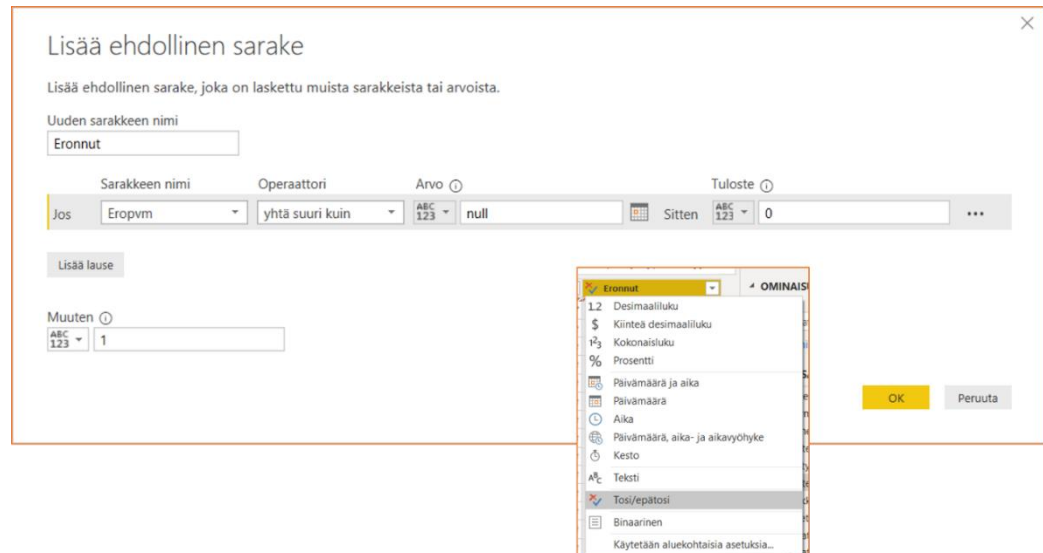
Lisää lause

Muuten
Opinnot jatkuu

OK Peruuta

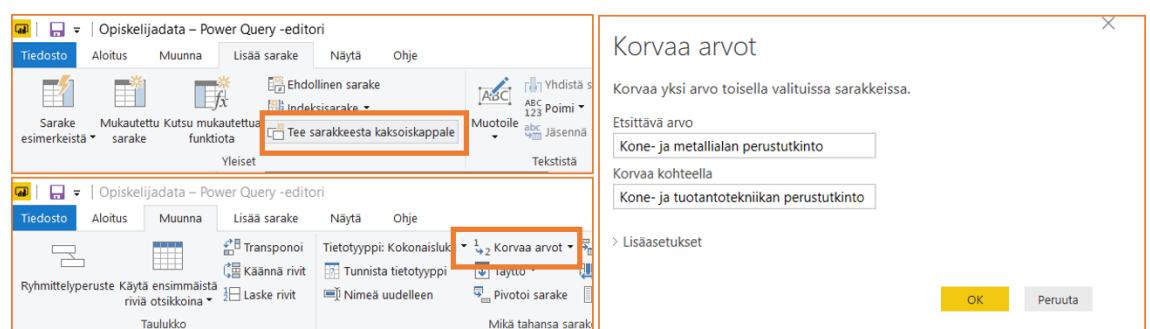
Kuva 10. Opiskelijan opintojen tilan lisääminen ehdollisena sarakkeena.

Analysointityökalujen kohdemuuttujana tulee usein olla tieto, joka on tyyppiä tosi/epätosi. Kuvan 11 Lisää ehdollinen sarake -toiminnolla tämä toteutetaan antamalla tyhjälle eropäivämäärälle arvo 0 ja muussa tapauksessa kenttä saa arvon 1. Tämän jälkeen kyseisen sarakkeen tyyppiä muutetaan Tosi/epätosi. Tästä sarakkeesta nähdään, onko opiskelija keskeyttänyt opintonsa vai ei.



Kuva 11. Lisää ehdollinen sarake -toiminto ja saraketyypin muuttaminen.

Opiskelijat-taulukosta voidaan todeta, että tutkintojen nimet ovat muuttuneet useita kertoja aikajakson aikana. Esimerkiksi liiketalouden perustutkinnosta on tullut liiketoiminnan perustutkinto, kone- ja metallialan perustutkinnosta on tullut kone- ja tuotantotekniikan perustutkinto ja hiusalan perustutkinto ja kauneudenhoitoalan perustutkinto ovat yhdistyneet hius- ja kauneudenhoitoalan perustutkinnoksi. Analysoinnin ja pidemmän aikavälin vertailtavuuden kannalta tutkintonimikkeet muutetaan siihen muotoon kuin ne nykyinsäädännön mukaan ovat. Muutos tehdään Power BI:ssä kopioimalla Tutkinto-kenttä uudeksi kentäksi ja tämän jälkeen Korvaa arvot-toiminnolla muutetaan vanhat tutkintonimikkeet uusiksi kuten kuvan 12 esimerkissä. Jatkossa on suositeltavampaa poimia opiskelijahallintojärjestelmässä tutkintoihin kiinnitetyt kustannuspaikkatiedot, jolloin visualisoinnin voi rakentaa suoraan aloittain.



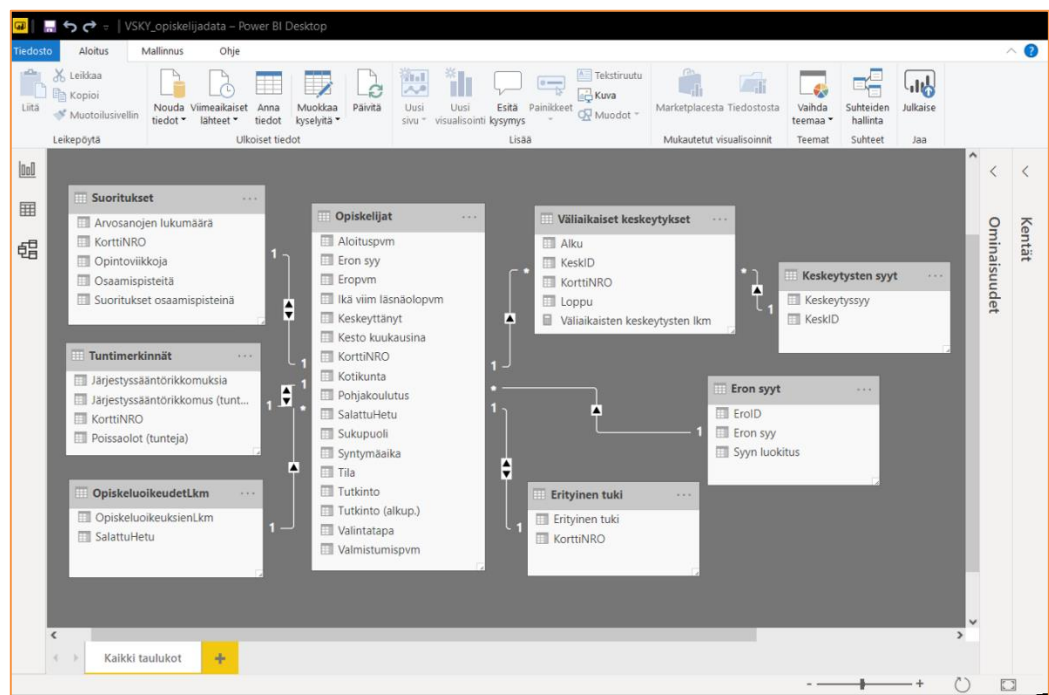
Kuva 12. Sarakkeen kopioiminen uudeksi ja arvojen korvaaminen toisella.

Väliaikaiset keskeytykset -taulukossa jokainen keskeytys on omana rivinä ja siksi taulukkoon tehdään vielä mittari kaavalla 4, jotta voidaan tarkastella väliaikaisten keskeytysten määrää opiskelijakohtaisesti. Funktio laskee määrän aloituspäivämääristä.

$$\text{Väl. kesk.lkm} = \text{COUNT}('Väliaikaiset keskeytykset'[Alku]) \quad (4)$$

Muiden taulukoiden osalta Tuntimerkinnät-taulukossa huomataan, että merkintöjen selitteet ovat muuttuneet huomattavasti aikajakson aikana, joten tietomalliin jätetään vain poissaolojen kokonaistuntimäärä ja muut poistetaan. Järjestyssääntörikkomukset-tuntimerkinnän avulla tehdään uusi tosi/epätosi-tyyppinen sarake tarkasteltavaksi analysointien yhteydessä. Taulukko koskien erityisen tuen suunnitelmia sisältää ainoastaan tiedon, onko opiskelijalla ollut päätös vai ei kenttätyyppillä tosi/epätosi ja opiskeluoikeuksien lukumäärä per opiskelija näkyy omasta taulukostaan.

Lopullinen tietomalli suhteineen nähdään kuvassa 13. Perustiedot opiskeluoikeudesta on Opiskelijat-taulussa, jonka KorttiNRO-kentän avulla taulut yhdistetään toisiinsa. Suorituksiin, tuntimerkintöihin ja erityiseen tukeen liittyvät taulukot ovat yksi yhteen-riippuvuussuhteessa Opiskelijat-taulukoon ja muiden taulukoiden riippuvuus on mallia monen suhde yhteen.



Kuva 13. Kaikki taulukot ja niiden väliset suhteet.

Datan valmistelun yhteydessä havaitaan, että opiskelijoiden tiedot ovat sitä puutteellisempia, mitä vanhempia ne ovat. Havaitaan, että esimerkiksi vuosina 2011-2014 aloittaneita opiskelijoita näyttää aineiston mukaan olevan läsnä vielä lukuisia, vaikka tiedetään, että tämä ei pidä paikkansa. Tarkempi selvitys osoittaa, että opiskelupaikkansa peruneita opiskelijoita on tullut mukaan raportille ja joidenkin osalta päättymispäivämäärä on jäänyt

kirjaamatta. Näiden havaintojen perusteella korjataan virheelliset tiedot ensin opiskelijahallintojärjestelmään ja sen jälkeen data haetaan uudelleen. Tämän jälkeen mallinnus on mahdollista toteuttaa.

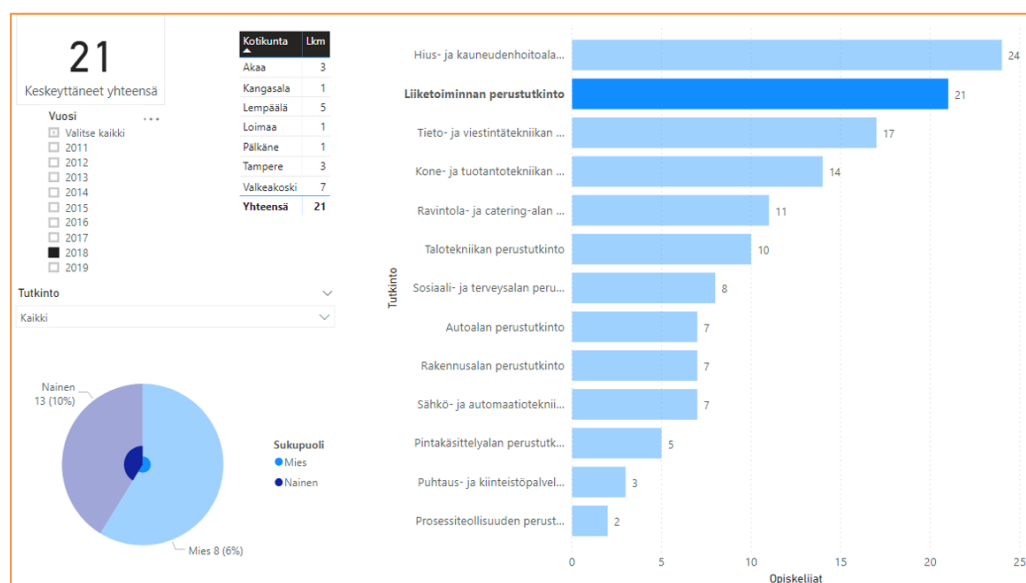
4.4 Mallinnus, tiedonlouhinta

Mallinnus on CRISP-DM:n neljäs vaihe, jossa valitaan mallinnusmenetelmä, suunnitellaan testiversio ja rakennetaan malli. Lopuksi malli arvioidaan. Tässä työssä mallinnusta tehdään neljän analytiikan perustyyppin näkökulmasta ja kaavioita rakennetaan useita, jotta Power BI:n eri ominaisuuksia voidaan monipuolisesti esitellä toimeksiantajalle.

4.4.1 Deskriptiivinen mallinnus

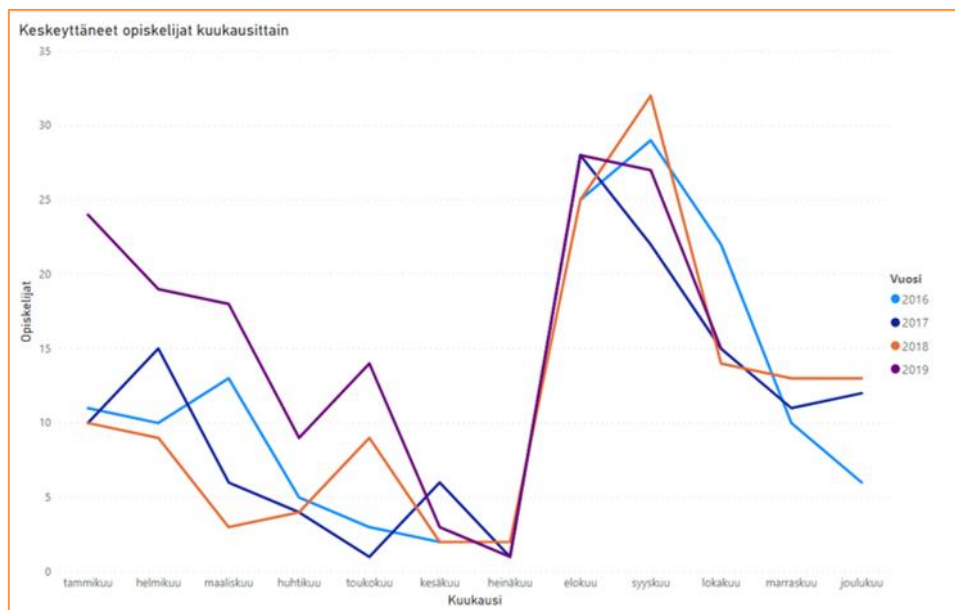
Deskriptiivinen analytiikka kertoo, mitä on tapahtunut. Power BI:n perusvisualisoinnit kuten pylväs-, palkki-, viiva-, ympyrä ja aluekaavioita toimivat hyvin tähän tarkoitukseen. Kuten aiemmin todettiin, Valkeakosken seudun koulutuskuntayhtymän raportointijärjestelmässä tieto on esitelty tuloskorttinäkymällä ja yksittäisillä pylväsdigrammeilla. Power BI:ssä vastaavaa tietoa saadaan sijoiteltua ruudulle siten, että useita erityyppisiä kaavioita on näkyvillä samanaikaisesti.

Esimerkkinä Power BI:n perusvisualisoinneista on kuvassa 14 keskeyttäneiden opiskelijoiden taustatietoja esitettynä taulukolla sekä palkki- ja ympyräkaaviolla. Osittajat ovat tapa suodattaa ruudulla näkyvää tietoa ja ne voivat olla muodoltaan joko luetteloita, alasvetovalikoita tai aikavälin valitsimia. Kuvassa on osittaja vuosiluvulle ja tutkinnolle ja valinnaksi on otettu vuosi 2018. Palkkikaaviossa on klikattu liiketoiminnan perustutkinnon palkkia, jolloin tieto näkyy ruudulla kyseisen tutkinnon mukaan. Muu joukko näkyy taustalla haaleammalla värillä.



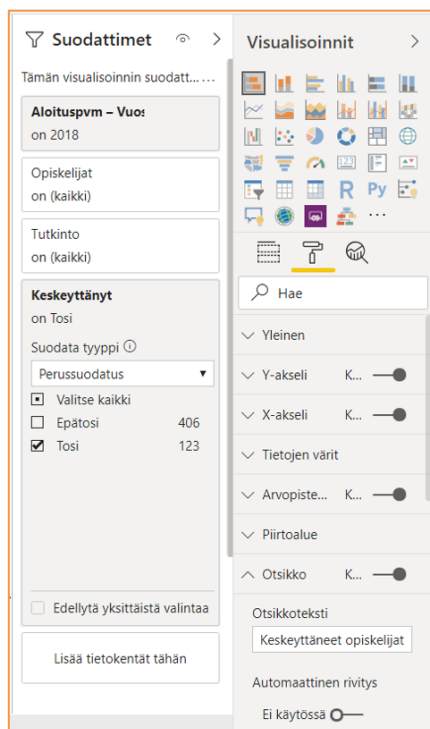
Kuva 14. Keskeytykset liiketoiminnan perustutkinnon mukaan.

Toisena esimerkkinä kuvassa 15 käytetään viivakaaviota, jonka avulla esitetään keskeyttämisten ajoittuminen kalenterivuoden mittaan. Siitä nähdään, että keskeytyksiä on eniten elo-syyskuussa.



Kuva 15. Viivakaavio keskeytyksistä kuukausittain ajalla 2016 - 2019

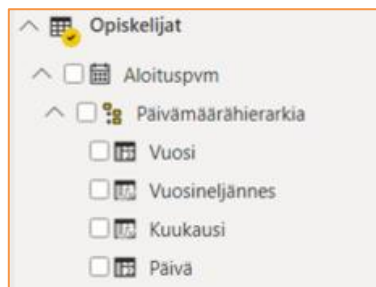
Kaavioiden sisältöä ja ulkoasua voidaan muokata. Suodattimien avulla voidaan määritellä kaaviossa mukana olevaa tietoa ja Muotoile-toiminnolla kaavioin eri osia voidaan muuttaa. Kuvassa 16 kaavioon on otettu mukaan vain keskeyttäneet opiskelijat suodattimella ja siihen on lisätty otsikko.



Kuva 16. Suodattimen lisääminen ja kaavion muotoilu

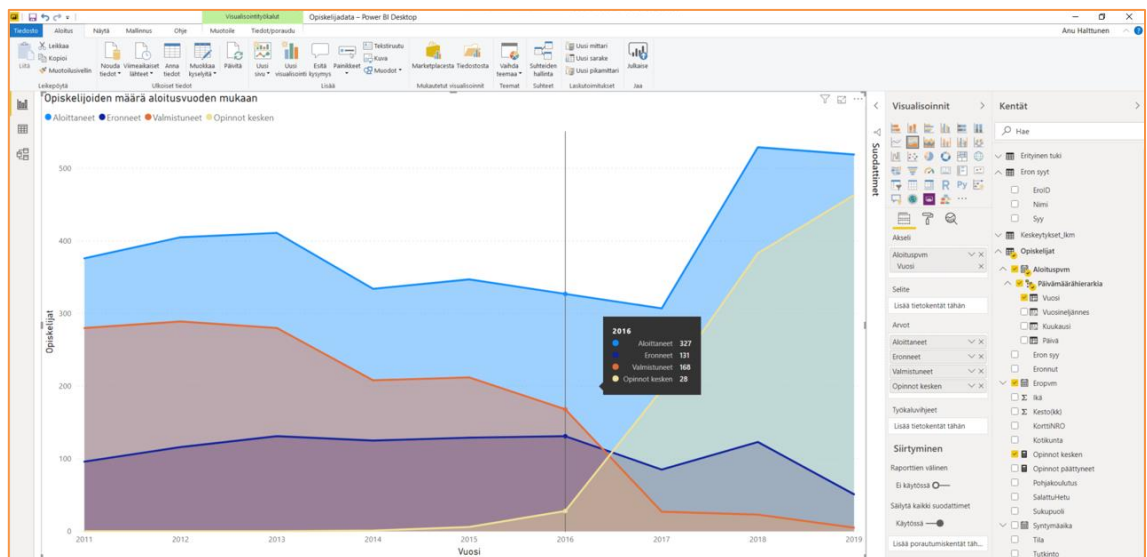
4.4.2 Diagnostiivinen mallinnus

Diagnostiivinen mallinnus etsii syitä, miksi jotain on tapahtunut. Esimerkki diagnostiivisesta mallinnuksesta tehdään aluekaaviolla tarkastelemalla aloittamisvuoden perusteella eronneita ja valmistuneita opiskelijoita. Power BI tunnistaa automattisesti päivämääräkentät ja niihin syntyy päivämäärähierarkioita kuvan 17 mukaisesti. Päivämäärähierarkiaa voidaan hyödyntää visualisointien rakentamisessa.



Kuva 17. Power BI:n luoma päivämäärähierarkia aloituspäivämäärässä.

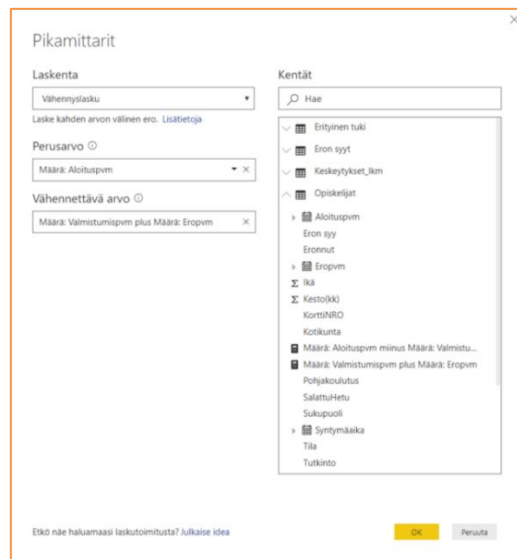
Kuvan 18 aluekaaviossa aloituspäivämäärän mukainen vuosi on asetettu kuvaajan x-akseliksi. Vaaleansinisellä värillä nähdään aloittaneet opiskelijat, tummansinisellä keskeyttäneet ja punaisella valmistuneet opiskelijat. Keskeyttäneet ja valmistuneet näkyvät siis aloitusvuotensa perusteella kuvaajassa.



Kuva 18. Aluekaavio opiskelijoiden määrästä aloitusvuoden mukaisesti

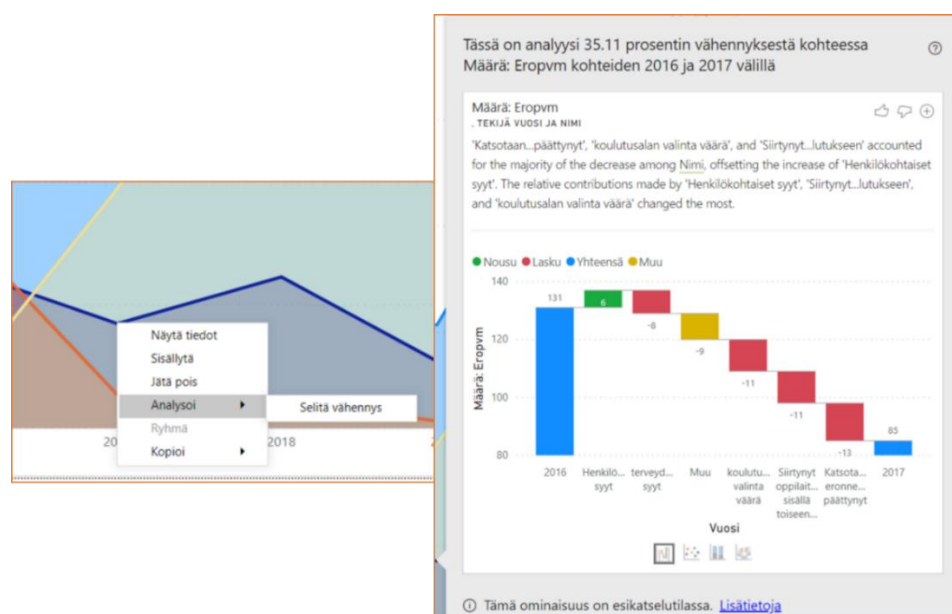
Aluekaaviossa on lisäksi hyödynnetty Pikamittari-toimintoa. Pikamittari tehdään kenttälue telossa hiiren kakkospainikkeella ja avautuvassa ikkunassa valitaan laskentatapa ja arvot, joista laskenta toteutetaan. Arvoja pystyy valitsemaan myös eri taulukoista. Kuvan 19 pikamittarissa aloittaneiden määrästä on vähennetty eronneet ja valmistuneet. Pikamittari on

otettu mukaan aluekaavioon ja nyt keltaisella alueella nähdään opiskelijat, joiden opinnot ovat vielä kesken.



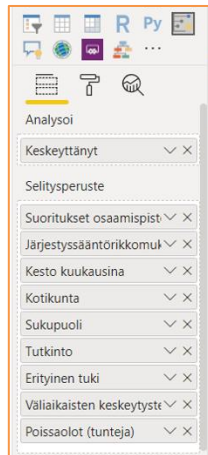
Kuva 19. Pikamittari, jonka avulla lasketaan opiskelijat, joiden opinnot ovat kesken.

Diagnostiiviseksi kuvaajan tekee kuvaajassa itsessään tarjolla olevan analysointimahdollisuuden hyödyntäminen. Sen avulla voidaan etsiä syitä tietyn arvon muutokselle tai poikkeamalle kuvaajassa. Analysointitoiminto on saatavilla monissa kaavioissa ja siihen pääsee käsiksi klikkaamalla hiiren kakkospainiketta halutussa kaavion kohdassa. Kuvassa 20 hiiren kakkospainikkeella on painettu aluekaaviossa vuoden 2017 keskeytymäärän kohdalla, jolloin analysointi etsii syitä arvon vähentymiseen verrattuna vuoteen 2016. Ruudulle aukeaa ehdotuksia kuvaajista, jotka voi halutessaan lisätä sivulle.



Kuva 20. Kaavion analysointityökalu

22 nähtävällä tavalla. Analysoitava tieto tässä tapauksessa on Keskeyttänyt-kentän Tosi-vaihtoehto ja selityspenusteeksi voidaan lisätä mahdollisia taustatekijöitä. Tämän jälkeen automaattinen laskenta tuo ruudulle näkyviin seikat, joilla on suurin vaikutus analysoitavaan asiaan.



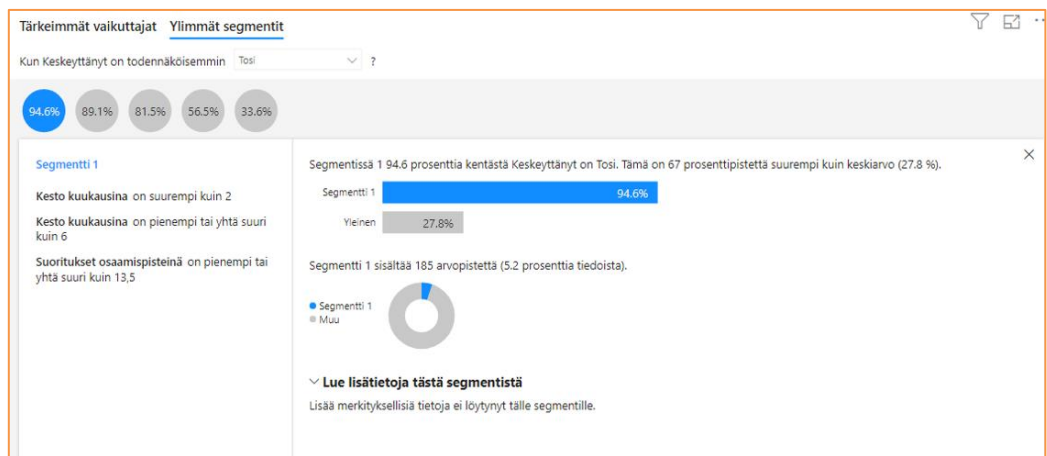
Kuva 22. Tärkeimmät vaikuttajat visualisoinnin rakentaminen

Visualisoinnissa on kaksi välilehteä. Ensimmäisellä välilehdellä nähdään vaikuttavat seikat suuruusjärjestyksessä. Kuvan 23 esimerkissä aikajaksoa ei ole rajattu ja mukana on opiskelijat vuodesta 2011 alkaen. Mitä vähemmän osaamispisteitä on kertynyt ja mitä vähemmän aikaa koulutus on keskeyttänyt, sitä todennäköisempää keskeyttäminen on. Todennäköisyyttä ilmaisevasta ympyrästä klikatessa viereen aukeaa asiaa tarkentava kaavio. Muiksi tärkeiksi vaikuttajiksi laskenta nostaa poissaolojen suuren määrän ja tutkinnoista tieto- ja viestintätekniikan perustutkinto keskeytetään hienommin todennäköisemmin kuin muut. Sukupuoli ei nouse asiaan vaikuttavaksi tekijäksi.



Kuva 23. Tärkeimmät vaikuttajat

Toisella välilehdellä tärkeimmät vaikuttajat on nostettu segmentteihin. Kuvassa 24 nähdään, että tässä analysoinnissa segmenttien selittävinä tekijöinä on ainoastaan koulutuksen kesto ja osaamispisteiden määrä.



Kuva 24. Ylimmät segmentit

Tärkeimpien vaikuttajien visualisointiin voidaan tarvittaessa lisätä osittajia, jos halutaan tarkastella lukuja tarkemmin jonkun tietyn taustamuuttujan osalta. Kun osittajan avulla tehdään rajoituksia aloitusvuoden perusteella, voidaan todeta, että tuloksissa tapahtuu huomattavia muutoksia. Esimerkiksi poissaolojen vaikutus keskeyttämiseen pienenee, mitä uudempiä keskeyttämisiä tarkastellaan. Tämän voi selittää muuttuneet poissaolomerkintöjen käytänteet, joissa poissaolojen sijaan seurataankin läsnäoloja. Eri tutkimukset nousevat eri aikajaksoilla eri tavalla kaaviolle.

Tämä visualisointi soveltuu hyvin keskeytysten taustatekijöiden selvittämiseen, mutta tarkasteltavaan aikajaksoon tulee kiinnittää huomiota. On suositeltavaa koko raportointijakson lisäksi tarkastella vain viimeisimpiä vuosia erikseen. Sen myötä voidaan nähdä, millaisia muutoksia todennäköisyyksissä tapahtuu. Tiedetyt seikat eivät enää viime vuosina aiheuta yhtä todennäköisesti opintojen keskeyttämistä.

4.4.4 Preskriptiivinen mallinnus

Preskriptiivinen analytiikka kertoo, mitä seuraavaksi pitäisi tehdä ja sen perustana on aiemmat analytiikkavaiheet. Kun tarkastellaan kaikkia tehtyä visualisointeja, tuottavat ne jokainen tärkeää tietoa keskeyttämisiin liittyen.

Tulosten jatkojalostamiseksi ja opiskelijoiden keskeyttämisten ennakkointia ja ehkäisyä ajatellen opinnäytetyössä tehtyt visualisoinnit voivat toimia toimeksiantajalla monialaisen käsittelijäryhmän keskustelun pohjana. Visualisointeja voidaan edelleen kehittää ja tarkentaa esiin nousevien havaintojen pohjalta. Toimeksiantajan suunnitelmana on se, että tiedolla johtamisen perusprosessiin lisätään säännölliset analytiikkapajat, joissa

ennalta sovitun teeman pohjustukseksi tehdään tiedonlouhintaa ja tulok-
sista keskustellaan ja niiden perusteella sovitaan seuraavat vaiheet. Ensimmäisen analytiikkapajan aiheena tulee olemaan opintojen keskeyttämisen
ehkäisy, jossa tämän opinnäytetyön tuloksia hyödynnetään.

4.5 Arviointi

Arviointi on CRISP-DM:n viides vaihe, jossa tuloksia arvioidaan ja prosessi
käydään läpi sekä seuraavat vaiheet esitellään. Tässä tiedonlouhinnassa
kaavioiden avulla pystyttiin todentamaan tiettyä säännönmukaisuutta ja
yhtäläisyyksiä keskeyttämisten taustatekijöistä. Opinnot keskeytetään tyy-
pillisesti aivan alkuvaiheessa ja elo-syyskuussa keskeytyksiä tapahtuu eni-
ten. Tämä on loogista, sillä suurin osa opiskelijoista valitaan yhteishaun pe-
rusteella, jolloin suurimmalla osalla opinnot alkavat elokuussa. Opiskelijat
saattavat päästä yhteishaun ensisijaiseen hakukohteeseensa tai opintojen
alkaessa todetaan, että muu ala kiinnostaa enemmän. Opintojen edetessä
poissaolojen määrän kasvu tai opiskelija jääminen väliaikaiselle keskeytyk-
selle kasvattaa riskiä opintojen päättämiseksi. Tutkintojen välillä on pientä
eroavaisuutta keskeyttämisen todennäköisyydessä.

Lähtökohta onnistuneeseen tiedonlouhintaan on huolellinen valmistelu.
Ensin tulee olla selvää, miksi tiedonlouhinta tehdään. Tässä tutkimuksessa
tavoite oli selkeä ja varsinaisen tiedonlouhinnan lisäksi haluttiin tutustua
monipuolisesti Power BI:n ominaisuuksiin. Tämä antoi vapaat kädet kokei-
lemaan ja tutustumaan ennakkoluulottomasti erilaisiin tapoihin työstää
dataa.

Tuloksellinen tiedonlouhinta edellyttää ymmärrystä käsillä olevasta aineis-
tosta. Prosessin aikana havaitaan, että data opiskelijahallintojärjestel-
mässä on erittäin pirstaleista ja vaatii huolellista validointia ennen kuin
analysointia siitä voidaan tehdä. Mikäli opiskelijahallintojärjestelmä ei olisi
tuttu, olisi haasteita aiheuttanut oikeanlaisen aineistoon tuottaminen Ex-
cel-tiedostoon. Puutteet datassa olisi ollut hankalampi havaita ja ylimää-
rää tietoa olisi voinut päätyä aineistoon. Datan valmisteluvaihe on tär-
kein ja aikaa vievin osuus tiedonlouhinnan prosessissa. Huolellinen datan
läpikäyminen ja arvioiminen auttaa seuraavia vaiheita ja tuottaa luotetta-
vampaa tietoa.

Visualisointien tekemisessä pääsee Power BI:ssä nopeasti vauhtiin. Lo-
giikka on Microsoftin tuotteista tuttu, mutta mahdollisuuksia on paljon
enemmän. Visualisointien luomisessa on hyvä miettiä, millaista tietoa da-
tasta haluaa löytää ja kannattaa kokeilla eri kaavioita, jotta selkein vaihto-
ehto löytyy. Testatut analytiikkaominaisuudet yllättivät helppokäyttöisyy-
dellään.

Tiedonlouhinnan vaiheet ovat olleet onnistuneita, mutta niissä on jou-
duttu palaamaan taaksepäin useita kertoja. CRISP-DM malli oli hyvä runko
tekemiselle, jotta eri vaiheet tulee tehtyä huolellisesti.

4.6 Tulosten julkaisu

Viimeisenä CRISP-DM:n vaiheena on käyttöönotto ja mallien seuranta sekä ylläpidon suunnittelu. Tarvittaessa laaditaan loppuraportti ja arvioidaan projekti kokonaisuudessaan.

Koontinäytön jakaminen Power BI:ssä muille kuin tekijälle vaatii erillisen lisenssin. Tämän tiedonlouhinnan tuloksia ei tässä kohdassa julkaista jaettavalla koontinäytöllä, vaan tulokset esitellään toimeksiantajalle. Mikäli halutaan rakentaa automaattisesti päivittyvä koontinäyttö, tulee yhdyskäytävä rakentaa opiskelijahallintojärjestelmään yhteydessä olevaan tietokantaan ja hankkia lisenssejä koontinäyttöjen katseluun tai liittämiseen intranetsivuille. Jos tulokset halutaan nähtäville myös mobiilisti, tulee visualisoinnit muuttaa Power BI:n puhelinasettelu-toiminnolla sopivaan muotoon.

SMART-mallin mukaan datan käsittely ei pääty vielä tulosten julkaisuun, vaan vielä sen jälkeen vastausten pohjalta on liiketoiminnan ja ajattelutavan muututtava. Näkemys sopii hyvin nyt toteutettuun tiedonlouhintaan, jotta analytiikan hyödyntäminen ei jää kertaluontoiseksi kokeiluksi.

Projekti on onnistunut, vaikkakin se vahvisti ennakkokäsityksen lähdejärjestelmän datan pirstaleisuudesta. Tiedonlouhinta osoitti sen, että analytiikkaa on hyvä tuoda mukaan tiedolla johtamisen prosesseihin. Datasta löydettiin todennäköisyyksiä keskeyttämisten taustalla. Opiskelijahallintojärjestelmän dataa voidaan hyödyntää jatkossakin, kunhan datan valmisteluun paneudutaan huolellisesti. Tuoreempi data on luotettavampaa ja ehjempää kuin vanhempi.

5 JOHTOPÄÄTÖKSET

Analysointikeinot kehittyvät datamäärän kasvun myötä ja yrityksillä on yhä enemmän mahdollisuuksia hyödyntää tietoa liiketoiminnan tukena. Tiedolla johtamisessa datasta tuotettua tietoa tuodaan osaksi päätöksentekoa ja tieto pyritään jalkauttamaan osaksi yrityskulttuuria ja ihmisten arkea. Historiatietoihin perustuva raportointi ei enää riitä, vaan analytiikan avulla voidaan löytää syitä tapahtuneisiin, tehdä ennusteita tulevista ja ohjata toimintaa ihan uudella tavalla.

Tiedonlouhinta on analysoinnin muoto, jossa datasta etsitään kiinnostavia malleja tai rakenteita, joita käytetään nykytilan selittämiseen ja tulevaisuuden ennustamiseen. Sekä tiedolla johtamisen että tiedonlouhinnan lähtökohtana tulee olla yrityksen strategia. Kun tavoitteet ja selkeä kysymys on määritelty heti aluksi, voidaan keskittyä siihen, mikä on tärkeää eikä siihen, mikä on mahdollista. Datan valmistelu on tulosten oikeellisuuden kannalta projektissa keskeinen vaihe ja vie yleensä suurimman osan koko projektin ajasta ja resursseista. Huolellinen datan laadunhallinta varmistaa, että datan oikeellisuus, läpinäkyvyys ja tietosuoja toteutuu datan koko elinkaaren ajan. Tulevaisuudessa analytiikan arvellaan automatisoituvan entisestään, mikä voi mahdollistaa jopa ihmisen korvaamisen päätöksentekopisteessä.

Microsoft Power BI voi toimia toimeksiantajan raportointityökalun tukena tiedolla johtamisen prosesseissa. Sen avulla mittareita voidaan esittää visuaalisemmassa ja vuorovaikutteisemmassa muodossa. Käyttäjäkohtaisten näkymien luominen ja valmiit vertailut esimerkiksi edelliseen vuoteen voi helpottaa ja yksinkertaistaa tiedon hyödyntämistä. Tietoa voidaan noudata Power BI:hin lukuisista tietolähteistä ja sen ansiosta siitä voidaan rakentaa tiedolla johtamisen keskus, joka yhdistää niin sisäisen kuin ulkoisen datan yhteen paikkaan. Power BI:ssä on monipuoliset analytiikkaominaisuudet, joiden kanssa pääsee nopeasti alkuun ja joita koko ajan kehitetään. Monista kaavioista löytyy sisäänrakennettuna analytiikkaominaisuuksia, joiden avulla voidaan selvittää tarkempia taustasyitä.

Opiskelijoiden opintojen keskeyttämiseen johtavien syiden selvittämiseksi Power BI:llä rakennettiin visualisointeja niin määrällisten seikkojen kuin taustatekijöiden todennäköisyyksien esittämiseksi. Tausta-aineiston tuntemus helpottaa datan valmistelua, joka on välttämätön vaihe luotettavien tulosten saamiseksi. Tässä tiedonlouhinnassa haasteen aiheutti se, että järjestelmän rakenne ja muuttuneet merkintätavat oli huomioitava sekä tiedonkeruussa että Power BI:n datan valmistelussa, jotta lopputulemana oli ehjä ja analysoitavissa oleva tausta-aineisto.

CRISP-DM-menetelmän avulla tiedonlouhinta eteni sujuvasti ja jokainen projektin vaihe tuli hoidetuksi huolella. Lopputulokseksi saatujen visualisointien avulla voitiin todeta säännönmukaisuutta keskeyttämisten

määrällisessä toteutumisessa ja saatiin selville todennäköisyyksiä eri taustatekijöille.

Aikajakson valinta vaikutti saatuihin tuloksiin ja kertoo sen, että keskeyttämisen taustalla olevat tekijät muuttuvat jatkuvasti. Lisäksi tästä voidaan päätellä, että tausta-aineistossa tapahtuu muutoksia ja data on sitä luotettavampaa mitä uudempaa käytetään. Vuonna 2018 alkaneet valtakunnalliset tiedonsiirrot opiskelijahallintojärjestelmästä opetushallituksen tietovarantoihin on tuonut merkintätapoihin säännönmukaisuutta ja se vahvistaa osaltaan uudemman tiedon luotettavuuden.

Suosittelien yhdyskäytävän rakentamista opiskelijahallintojärjestelmän tietokannasta Power BI:hin. Tämä jättää pois välivaiheen, jossa data tuotetaan ensin Exceliin ja noudetaan siitä Power BI:hin. Yhdyskäytävän avulla Power BI:ssä on aina käytettävissä opiskelijoista uusin data, joka päivittyy automaattisesti päivittäin. Uusien tietolähteiden tuominen joko osaksi tietokantaa tai suoraan Power BI:hin mahdollistaa jatkuvan seurannan keskittämisen ja sujuvoittaa tiedolla johtamista. Koontinäyttöjä voidaan julkaista esimerkiksi intranetsivuilla, mikä edistää tiedolla johtamisen jalkauttamista ja näkyvyyttä koko henkilöstölle.

Tulevaisuuden datan hyödyntämisen kannalta haastan vielä pohtimaan, millaisista mittareista olisi organisaation arjessa suurin hyöty. Nykyisten määrällisten mittareiden ohessa laadullisilla mittareilla voisi saavuttaa uutta sisältöä joka päiväiseen tiedolla johtamiseen. Opiskelijadatan kannalta on hyvä pohtia, millaista taustatietoa opiskelijoista halutaan tulevaisuudessa analysoida. Opiskelijahallintojärjestelmässä on jo nyt mahdollisuus esimerkiksi annettujen tukitoimien kirjaamiselle, mutta kirjaamiseen liittyvien ohjeistusten selkeyttämisellä tietoa varmemmin myös kirjataan sille tarkoitettuihin kohteisiin. Vielä hyödyntämättömät sisäiset ja ulkoiset tietolähteet kannattaa kartoittaa, jotta niitäkin voidaan tarvittaessa tuoda osaksi vastaamaan huolellisesti määriteltyihin tiedolla johtamisen kysymyksiin.

6 YHTEENVETO

Opinnäytetyönä toteutettiin tiedonlouhinta toimeksiantajan opiskelijahallintojärjestelmän dataan ja tavoitteena oli löytää syitä opiskelijoiden opintojen keskeyttämisen taustalla. Samalla saatiin kokemusta Microsoft Power BI -ohjelmistosta ja voitiin pohtia ohjelmiston hyödyntämismahdollisuuksia toimeksiantajan tiedolla johtamisen prosesseissa.

Opinnäytetyön tietopohjana on datan merkitys liiketoiminnalle. Inspiraation lähde teoriaosuudelle oli kuva 4, jossa esitetään raportoinnin ja analytiikan eroavaisuuksia. Helposti jäädään tuijottamaan menneisyyteen, vaikka käden ulottuvilla on välineitä ennakointiin ja optimointiin. Datamäärän kasvu on arkipäivää ja sen tarjoamiin rajattomiin mahdollisuuksiin on helppo eksyä. Teoriaosuuden tärkeimmät ohjenuorat datan käsittelyyn on tavoitteen asettamisen välttämättömyys ja asioihin reagoiminen saatujen tulosten perusteella. Pitää siis määritellä huolellisesti etukäteen, mitä halutaan tietää ja toisaalta, miten saatua tietoa lopulta hyödynnetään.

Tietopohjan kokoamisen perustana oli tuntemus toimeksiantajan toimintaympäristöön ja prosesseihin. Teoria on koottu siitä näkökulmasta, että toimeksiantaja on julkisen puolen toimija, jolloin esimerkiksi puhtaasti liiketoimintaan keskittyvä taustatieto on voitu rajata pois. Myös tuntemus tiedonlouhinnan lähdejärjestelmään oli suuri etu luotettavien tulosten saavuttamisessa.

Olen tottunut Excelin käyttäjä ja sen myötä Power BI:n perustoiminnot oli helppo ottaa haltuun. Ohjelmiston lukuisat mahdollisuudet tarjoavat haastetta ja oppimista myös jatkossa ja aionkin ottaa Power BI:n omaan työhöni jokapäiväiseksi työkaluksi. Power BI-ohjelmisto kehittyy kaiken aikaa ja Microsoftin vision mukaan analytiikkatoimintoihin tuodaan parannuksia ja uusia koneoppimista sekä tekoälyä hyödyntäviä ominaisuuksia lisätään. Näistä on kannattaa pysyä jatkossakin ajan tasalla. Oman oppimisen kannalta antoisin osuus oli DAX-kielen opettelu datan valmistelussa Power BI:ssä. Aion jatkossa syventyä siihen vielä lisää.

Tutkimuskysymyksiin vastaaminen onnistui ja työn tavoitteet saavutettiin. Tiedonlouhinnan löydöksistä tulee keskustella toimeksiantajalla sopivalla monialaisella kokoonpanolla ja pohtia yhdessä seuraavat askelmerkit opintojen keskeyttämisen ehkäisyyn. Jatkoa ajatellen analytiikkaa kannattaa hyödyntää vielä enemmän tiedolla johtamisen prosesseissa ja Power BI:n ominaisuudet tarjoavat lisäarvoa toimeksiantajan raportointijärjestelmälle. Power BI:n avulla uusia tietolähteitä voidaan hyödyntää ja tuoda mittareihin myös laadullisia näkökulmia.

LÄHTEET

- Aspin, A. (2018). *Pro Power BI desktop* (2. painos). Haettu 19.9.2019. Ebook Central - tietokanta.
- Bahrepour, M. (2018). The Forgotten Step in CRISP-DM and ASUM-DM Methodologies Blogijulkaisu 17.8.2018. Haettu 12.10.2019 osoitteesta <https://sharing.luminis.eu/blog/the-forgotten-step-in-crisp-dm-and-asum-dm-methodologies/>
- Biecek, P. (2019). Model Development Process. *Cornell University*. Haettu 12.10.2019 osoitteesta <https://arxiv.org/abs/1907.04461>
- Bošnjak, Z., Grljević, O., & Bošnjak, S. (2019). Transforming Web data into knowledge: Implications for management. *Ekonomski Horizonti*, 21(2), 177–193. Haettu 12.10.2019 osoitteesta <https://doi.org/10.5937/ekonhor1902177b>
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0 Step-by-step data mining guide*. Haettu 16.9.2019 osoitteesta <https://www.the-modeling-agency.com/crisp-dm.pdf>
- Digia. (2018). Kolme visiota: Miltä analytiikan tulevaisuus näyttää? Blogijulkaisu 8.11.2018. Haettu 9.11.2019 osoitteesta <https://blog.digia.com/kolme-visiota-milta-analytiikan-tulevaisuus-nayttaa>
- Espoo. (n.d.). Espoo suunnittelee ja rakentaa kunnallistekniikka 157 miljoonalla vuonna 2018. Haettu 5.11.2019 osoitteesta [https://www.espoo.fi/fi-FI/Espoo_suunnittelee_ja_rakentaa_kunnallis\(132775\)](https://www.espoo.fi/fi-FI/Espoo_suunnittelee_ja_rakentaa_kunnallis(132775))
- Gartner. (n.d.). Magic Quadrant Research Methodology. Haettu 5.11.2019 osoitteesta <https://www.gartner.com/en/research/methodologies/magic-quadrants-research>
- Gartner. (2019). Gartner Reprint. Haettu 5.11.2019 osoitteesta <https://www.gartner.com/doc/reprints?id=1-3TXXSLV&ct=170221&st=sb>
- Jokar, N., Honavar, A. R., AgHamirzadeh, S., & Esfandiari, K. (2016). Web mining and Web usage mining techniques Haettu 2.11.2019 osoitteesta <https://popups.uliege.be/0037-9565/index.php?id=5371>
- Keller, J. (2019). Future of Business Analytics Trends in 2020 & Beyond. Haettu 19.11.2019 osoitteesta <https://selecthub.com/business-analytics/business-analytics-trends/>
- Kielitoimisto. (n.d.). Kielitoimiston sanakirja. Haettu 1.11.2019 osoitteesta <https://www.kielitoimistonsanakirja.fi/>
- Kienzler, R. (2018). Architectural thinking in the Wild West of data science. Haettu 12.10.2019 osoitteesta <https://developer.ibm.com/articles/architectural-thinking-in-the-wild-west-of-data-science/>
- Kirkkonummi. (n.d.). Tietoa taloudesta - Kirkkonummi. Haettu 5.11.2019 osoitteesta <https://www.kirkkonummi.fi/tietoa-taloudesta>
- Knight, D., Knight, B., Pearson, M., Quintana, M., & Powell, B. (2018). *Microsoft Power BI complete reference : bring your data to life with the powerful features of Microsoft Power BI*. Haettu 19.9.2019. Ebook Central -tietokanta.
- Laki ammatillisesta peruskoulutuksesta annetun lain muuttamisesta 246/2015. Haettu 9.11.2019 osoitteesta <https://www.finlex.fi/fi/laki/alkup/2015/20150246>
- Laki ammatillisesta koulutuksesta 531/2017. Haettu 26.10.2019 osoitteesta <https://www.finlex.fi/fi/laki/alkup/2017/20170531>
- Listenmaa, J., Piikkilä, J., & Aho, H. (2018). Parempi tiedolla johtaja. Haettu 14.10.2019 osoitteesta <https://www.innolink.fi/tiedolla-johtajan-opas-julkaistu/>

- Maheshwari, A. (2014). *Business intelligence and data mining*. Haettu 12.11.2019. Ebook Central -tietokanta.
- Markkula, T., & Syväniemi, A. (2015). *Analytiikkamatka, datasta tietoon ja tiedolla johtamiseen*. Helsinki: Suomen liikekirjat.
- Marr, B. (2015). *Big data : using SMART big data, analytics and metrics to make better decisions and improve performance*. Haettu 13.9.2019. Ebook Central -tietokanta.
- Marr, B. (2016). *Big Data in Practice: How 45 Successful Companies Used Big Data Analytics*. Haettu 23.9.2019. Ebook Central -tietokanta.
- Microsoft. (n.d.). Decision Tree Chart. Haettu 12.11.2019 osoitteesta <https://appssource.microsoft.com/en-us/product/power-bi-visuals/WA104380817?tab=Overview>
- Microsoft. (2019a). 2019 Gartner Magic Quadrant. Haettu 5.11.2019 osoitteesta <https://info.microsoft.com/ww-landing-gartner-mq-bi-analytics-2019>
- Microsoft. (2019b). Power BI:n suorituskykyä parantavat parhaat käytännöt. Haettu 4.11.2019 osoitteesta <https://docs.microsoft.com/fi-fi/power-bi/power-bi-reports-performance>
- Microsoft. (2019c). Power BI Desktopin tietolähteet. Haettu 4.11.2019 osoitteesta <https://docs.microsoft.com/fi-fi/power-bi/desktop-data-sources>
- Microsoft. (2019d). Tärkeimpien vaikuttajien visualisoinnin opetusohjelma. Haettu 12.11.2019 osoitteesta <https://docs.microsoft.com/fi-fi/power-bi/visuals/power-bi-visualization-influencers>
- Pellinen, J. (2017). *Talousjohtaminen* (2. painos). Haettu 4.10.2019. Alma Talent Pro -tietokanta.
- Powell, B. (2018). *Mastering Microsoft Power BI : expert techniques for effective data analytics and business intelligence*. Haettu 4.11.2019 Ebook Central -tietokanta.
- R Foundation. (n.d.). R: What is R? Haettu 5.11.2019 osoitteesta <https://www.r-project.org/about.html>
- Rogers, D. L. (2016). *The digital transformation playbook : rethink your business for the digital age*. Haettu 13.9.2019. Ebook Central -tietokanta.
- Roiger, R. J. (2017). *Data mining : a tutorial-based primer*. Haettu 18.10.2019. Ebook Central -tietokanta.
- Salo, I. (2014). *Big Data & pilvipalvelut*. Jyväskylä: Docendo Oy.
- SAS. (2014). Data Mining and SEMMA: Data Mining Using SAS(R) Enterprise Miner(TM): A Case Study Approach, Third Edition. Haettu 2.11.2019 osoitteesta <http://support.sas.com/documentation/cdl/en/emcs/66392/HTML/default/viewer.htm#n0pejm83csbja4n1xueveo2uoujy.htm>
- SFS. (n.d.). Laadunhallinta ISO 9001:2015 - Suomen Standardisoimisliitto SFS ry. Haettu 15.10.2019 osoitteesta https://www.sfs.fi/julkaisut_ja_palvelut/tuotteet_valokeilassa/iso_9000_laadunhallinta/iso_9001_2015
- SFS 9001 (2015). Laadunhallintajärjestelmät. Vaatimukset. SFS Online. Haettu 15.10.2019 osoitteesta <https://online.sfs.fi/>
- Smart Vision Europe. (2018). What is the CRISP-DM methodology? Haettu 16.9.2019 osoitteesta <https://www.sv-europe.com/crisp-dm-methodology>
- Succi, S., & Coveney, P. V. (2019, April 8). Big data: The end of the scientific method? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 377. Haettu 25.9.2019 osoitteesta <https://doi.org/10.1098/rsta.2018.0145>

- Suomen Pankki. (n.d.). Dashboard: Suomen Pankin interaktiivisia tilastovisualisointeja. Haettu 5.11.2019 osoitteesta <https://www.suomenpankki.fi/fi/Tilastot/rahalaitosten-tase-lainat-ja-talletukset-ja-korot/dashboard-suomalaisten-lainat-ja-talletukset/>
- Syväniemi, A. (2015). Mutun aika on ohi - tietoon pohjautuva strategia on helpompi perustella. Blogijulkaisu 4.5.2015. Haettu 14.10.2019 osoitteesta <https://blog.kauppalehti.fi/vieraskyna/mutun-aika-on-ohi-tietoon-pohjautuva-strategia-on-helpompi-perustella>
- Syväniemi, A. (2016). Tiedolla vai tunteella – oikoteitä tiedolla johtamiseen. Haettu 14.10.2019 osoitteesta <http://www.sytyke.org/tietojohdaminen/tiedolla-vai-tunteella-oikoteita-tiedolla-johtamiseen/>
- Tampereen kaupunki. (n.d.). Tampere 2030. Haettu 5.10.2019 osoitteesta <http://www.tampereenpulssi.fi/>
- Tietokiri. (2019). Mitä on tiedolla johtaminen? Haettu 14.10.2019 osoitteesta <https://tietokiri.fi/tiedolla-johtaminen/tiedolla-johtaminen-nain-se-tapahtuu/>
- Tilastokeskus. (2019). Suomen virallinen tilasto (SVT): Koulutuksen keskeyttäminen. Haettu 26.10.2019 osoitteesta <https://www.tilastokeskus.fi/til/kkesk/index.html>
- VAAO intranet. (2018). Haettu 15.10.2019.
- Viitala, R., & Jylhä, E. (2019). *Johtaminen. Keskeiset käsitteet, teorial ja trendit* Haettu 16.9.2019 Ellibs-tietokanta.
- Wilson, T. (2013). What is Compiled Data? - Aged Lead Store. Haettu 5.10.2019 osoitteesta <https://agedleadstore.com/what-is-compiled-data/>