

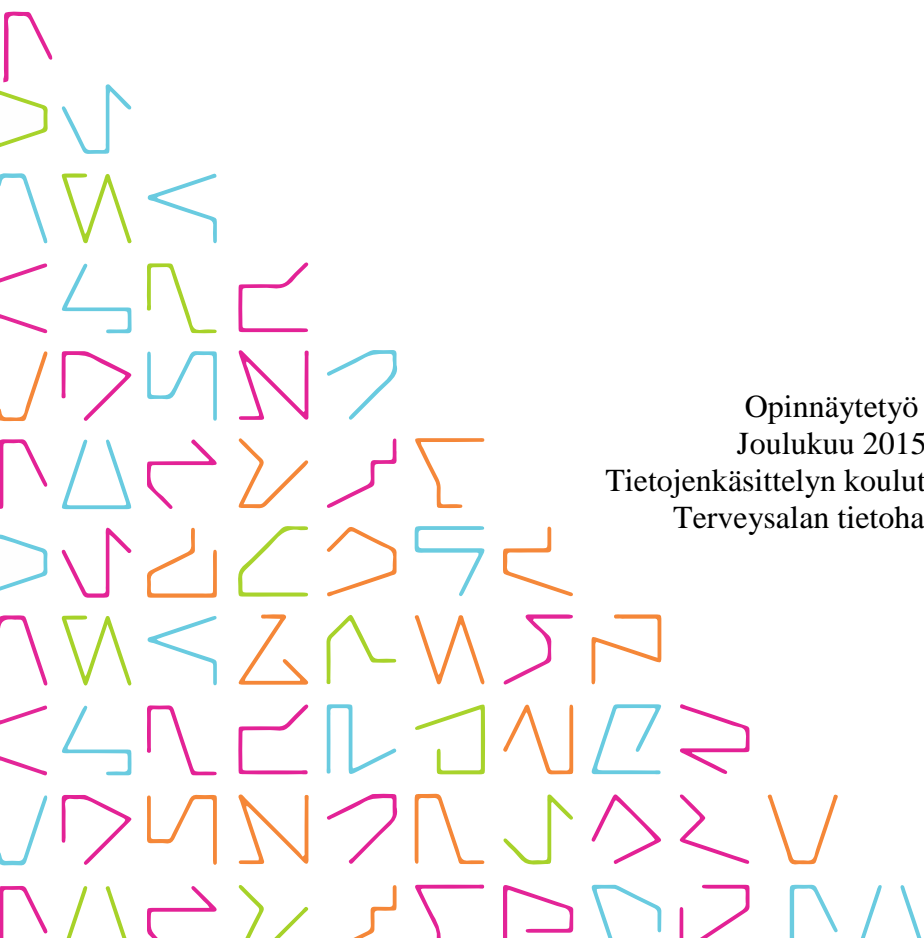


TAMPEREEN  
AMMATTIKORKEAKOULU

# ANALYTIKKAOHJELMISTOT ERI LIIKETOIMINTATARPEISIIN

Joona Hirvasmaa

Opinnäytetyö  
Joulukuu 2015  
Tietojenkäsittelyn koulutusohjelma  
Terveysalan tietohallinta



## TIIVISTELMÄ

Tampereen ammattikorkeakoulu  
Tietojenkäsittelyn koulutusohjelma  
Terveysalan tietohallinta

HIRVASMAA, JOONA:  
Analytiikkaohjelmistot eri liiketoimintatarpeisiin

Opinnäytetyö 52 sivua  
Joulukuu 2015

---

Opinnäytetyön tarkoituksena oli selvittää ja vertailla eri analytiikkaohjelmistoja. Tavoitteena oli saada toimeksiantajan käyttöön selvitys analytiikkaohjelmistoista ja niiden käytöstä erilaisissa liiketoimintatarpeissa. Toimeksiantajana oli ROIgrow Oy, joka on tiedolla johtamisen ja digitaalisen liiketoiminnan palveluyritys. Yritys tarjoaa asiakkailleen analytiikkapohjaisia kehityspalveluita ja halusi tämän vuoksi selvityksen eri ohjelmistoista.

Työhön valittiin viisi analytiikkaohjelmistoa: Alteryx Designer, IBM SPSS Statistics, KNIME, PSPP ja RapidMiner Studio. Ohjelmistoja selvitettiin liiketoiminnan kannalta ja selvitettäviin osa-alueisiin kuuluivat ominaisuudet, toiminnot, käytettävyys, tukipalvelut ja hinnoittelut. Selvityksen tulokset kerättiin testaamalla ohjelmistoja sekä tutkimalla ohjelmistojen kotisivuja. Selvityksen lisäksi työssä käytiin läpi yleisesti liiketoiminnan analytiikkaa.

Selvityksen perusteella selvisi, että ohjelmistoista toiminnoiltaan ja ominaisuuksiltaan kattavimmat olivat KNIME ja RapidMiner Studio. Kummatkin ohjelmistot sisälsivät satoja tehokkaita työkaluja analyysien suorittamiseen, ja niillä onnistui niin kuvaavan kuin ennustavan analytiikankin suorittaminen. Käytettävyydeltään parhain oli IBM SPSS Statistics, jonka käyttö oli varsin helppo oppia. Ohjelmiston käyttöä helpottivat selkeät interaktiiviset ikkunat, vetovalikot ja taulukot. Tukipalveluiltaan Alteryx Designer tarjosi tehokkaimman ohjelmistotuen. Tukipalvelut sisälsivät erinomaisen henkilökohtaisen tuen, mutta myös hyvän yhteisötuen ja laajat opetusmateriaalit. Hinnoittelultaan KNIME, PSPP ja RapidMiner Studio olivat edullisimmat ratkaisut. Loppujen lopuksi kaikki ohjelmistot olivat täysin päteviä erityyppisten analyysien suorittamiseen ja niiden erot olivat enemmän painotuksia kuin vakavia puutteita. Ohjelmistoista selvästi heikoin oli PSPP, mutta muuten jokainen ohjelmisto soveltui hyvin liiketoiminta-analytiikan harjoittamiseen.

---

Asiasanat: analytiikkaohjelmistot, liiketoiminnan analytiikka, analytiikka

## ABSTRACT

Tampereen ammattikorkeakoulu  
Tampere University of Applied Sciences  
Degree Programme in Business Information Systems  
Data Administration for the Health Care Industry

HIRVASMAA, JOONA:  
Analytics Software for Different Business Needs

Bachelor's thesis 52 pages  
December 2015

---

The purpose of this thesis was to research and compare different analytics tools. The objective was to create a report about these tools and about their use cases in connection with various business needs presented by the commissioner of this thesis, ROIgrow Oy. The company focuses on analytical growth services, and it offers analytics software for its clients which is why they needed the report about analytics software.

Five tools were chosen for this thesis: Alteryx Designer, IBM SPSS Statistics, KNIME, PSPP and RapidMiner Studio. The factors under examination included features, functions, usability, support services and prices. Research data were collected by testing the aforementioned programs and examining their homepages. Business analytics in general were also included in the discussion.

According to the research KNIME and RapidMiner Studio had the most extensive functions and features. Both tools included hundreds of effective tools for creating an analysis. IBM SPSS Statistics had the best usability especially because of its ease of use. Alteryx Designer offered the most efficient software support services. As regards pricing, KNIME, PSPP and RapidMiner Studio were the cheapest solutions. In the end all differences between the analytics software examined here were mostly related to emphasis rather than severe shortcomings. PSPP was clearly the weakest piece of software, but each tool was suitable for doing business analyses.

---

Key words: analytics software, business analytics, analytics

## SISÄLLYS

1	JOHDANTO.....	7
2	TAUSTATIETOJA .....	8
	2.1 Tarkoitus ja tavoite .....	8
	2.2 Rajaukset.....	8
	2.3 Työn rakenne .....	9
3	ANALYTIKKA LIIKETOIMINNASSA .....	10
	3.1 Analytiikka.....	10
	3.2 Liiketoiminta tiedonhallinta.....	11
	3.3 Liiketoiminnan analytiikka .....	12
	3.4 Analytiikka, liiketoiminta tiedonhallinta ja liiketoiminnan analytiikka ..	13
	3.5 Analytiikan tyyppejä.....	14
	3.5.1 Kuvaileva analytiikka.....	14
	3.5.2 Ennustava analytiikka .....	14
	3.5.3 Ohjaava analytiikka.....	15
4	ANALYTIKKAOHJELMISTOT .....	16
5	TUTKITTAVAT OHJELMISTOT .....	17
	5.1 Alteryx Designer.....	18
	5.1.1 Ominaisuudet ja toiminnot.....	18
	5.1.2 Käytettävyys.....	20
	5.1.3 Tukipalvelut .....	21
	5.1.4 Hinnoittelu .....	22
	5.1.5 Yhteenveto .....	22
	5.2 IBM SPSS Statistics.....	23
	5.2.1 Ominaisuudet ja toiminnot.....	23
	5.2.2 Käytettävyys.....	25
	5.2.3 Tukipalvelut .....	26
	5.2.4 Hinnoittelu .....	26
	5.2.5 Yhteenveto .....	27
	5.3 KNIME .....	28
	5.3.1 Ominaisuudet ja toiminnot.....	28
	5.3.2 Käytettävyys.....	30
	5.3.3 Tukipalvelut .....	31
	5.3.4 Hinnoittelu .....	31
	5.3.5 Yhteenveto .....	31
	5.4 PSPP.....	32
	5.4.1 Ominaisuudet ja toiminnot.....	32

5.4.2 Käytettävyys.....	35
5.4.3 Tukipalvelut .....	35
5.4.4 Hinnoittelu .....	36
5.4.5 Yhteenveto .....	36
5.5 RapidMiner Studio.....	36
5.5.1 Ominaisuudet ja toiminnot.....	37
5.5.2 Käytettävyys.....	39
5.5.3 Tukipalvelut .....	40
5.5.4 Hinnoittelu .....	40
5.5.5 Yhteenveto .....	41
6 JOHTOPÄÄTÖKSET .....	42
6.1 Ohjelmistojen ominaisuudet ja toiminnot.....	42
6.2 Ohjelmistojen käytettävyys.....	44
6.3 Ohjelmistojen tukipalvelut.....	46
6.4 Ohjelmistojen hinnoittelu .....	47
6.5 Yhteenveto ohjelmistoista.....	48
7 POHDINTA.....	50
LÄHTEET.....	51

**LYHENTEET JA TERMIT**

Analytiikka	tilastotieteen menetelmät datan muuntamiseen hyödylliseksi tiedoksi
BA	liiketoiminnan analytiikka (Business analytics)
BI	liiketoimintatiedon hallinta (Business intelligence)
Big Data	suuret järjestelemättömät tietomassat ja teknologiat, joilla datasta jalostetaan ymmärrystä ja tietoa
CSV	tiedostomuoto, jolla tallennetaan taulukkomuotoista tietoa tekstitiedostoon (Comma seperated values)
Data	tietoa, jolla ei välttämättä ole merkitystä ja järjestystä
R-kieli	avoimeen lähdekoodin perustava ohjelmointikieli, jota käytetään varsinkin tilastollisissa menetelmissä
Raportointi	asian dokumentointia ja välittämistä
Segmentointi	markkinointistrategia, jossa laajat kohdemarkkinat jaetaan useisiin osiin eri tarpeiden tyydyttämiseksi
Syntaksi	ohjelmointikielen lauseoppi
Tekstinlouhinta	menetelmiä, joilla suurista tekstimassoista löydetään oleellinen tieto
Tiedonlouhinta	menetelmiä, joilla suurista tietomassoista löydetään oleellinen tieto

## 1 JOHDANTO

Monet entiset kilpailukeinot ovat kokonaan kadonneet. Maantieteelliset sijainnit eivät enää anna turvaa globaaleilla markkinoilla ja tuotteita sekä palveluita voidaan nopeasti kopioida. Nykyään kilpaillaan siitä, kuinka tehokasta liiketoiminta on, ja kuka tekee parhaimmat päätökset toimintatapojen, tuotteiden ja palveluiden muuttamisessa. Erottavaa tekijää on markkinoilla kyettävä optimoimaan ja näin analytiikan merkitys kasvaa. (Helkiö 2012.) Liiketoiminnan analytiikkaa voidaan soveltaa moninaisesti ja eri tavoin riippuen yrityksen strategioista. Analytiikan toteuttaminen pitää aloittaa oikean ohjelmiston valinnasta. Analytiikkaohjelmistoja löytyy paljon ja jokaisella on omat erottelevat vahvuutensa. On tärkeää löytää oikeat ohjelmistot erilaisiin tarpeisiin.

Oikean analytiikkaohjelmiston valitseminen ja käyttäminen onkin oleellinen osa liiketoiminnan analytiikkaa. Opinnäytetyön aiheena on tutkia ja selvittää analytiikkaohjelmistojen käyttöä eri tarpeisiin. Toimeksiantajana työllä toimii ROIgrow Oy, joka on tiedolla johtamisen ja digitaalisen liiketoiminnan palveluyritys. Yritys tarjoaa asiakkailleen tukea liiketoiminnan kasvun edistämiseen erilaisilla palveluilla, kuten analytiikalla, nettisivuilla ja mainostuotintopalveluilla. Koska yritys tarjoaa asiakkailleen analytiikkapalveluita, on ohjelmistojen kattava läpikäyminen tarpeellista.

Opinnäytetyö toimii osana tietojenkäsittelyn tradenomin tutkintoa Tampereen ammattikorkeakoulussa. Opinnäytetyö toteutetaan kirjallisena työnä ja lähteinä käytetään lähinnä artikkeleita, kirjallisuutta ja erinäisiä luotettavia nettisivuja.

## 2 TAUSTATIETOJA

Tässä luvussa käydään läpi työn taustatietoja. Selvitetään työn tavoite ja tarkoitus, eli mitä aiotaan tehdä ja mitä hyötyä siitä on. Laaja aihe vaatii myös tiettyjä rajoituksia. Luvun lopussa esitetään yleiskatsaus työn rakenteesta.

### 2.1 Tarkoitus ja tavoite

Opinnäytetyön tarkoituksena on tutkia ja vertailla eri analytiikkaohjelmistoja. Työssä tutkitaan ohjelmistojen toimintoja, käytettävyyttä, tukipalveluita ja hinnoittelua.

Tavoitteena on saada toimeksiantajan käyttöön selvitys analytiikkaohjelmistoista ja niiden käytöstä erilaisissa liiketoimintatarpeissa. Opinnäytetyön avulla toimeksiantaja pystyy valitsemaan oikean ohjelmiston eri asiakasprojekteihin. Opinnäytetyö tarjoaa myös oivan dokumentin analytiikkaohjelmistoista. Opinnäytetyön tavoitteena on myös kehittää tekijän kykyjä soveltaa tutkimustietoa ja käyttää valittuja menetelmiä työelämän ongelmien erittelyyn ja ratkaisemiseen sekä lisätä valmiutta itsenäiseen vaativaan asiantuntijatyöskentelyyn.

Yhteistyössä toimeksiantajan ja ammattikorkeakoulun kanssa tekijä oppii myös arvokkaita vuorovaikutustaitoja. Opinnäytetyön tuotokset ovat hyödynnettävissä suoraan työelämään, koska niitä voidaan soveltaa myöhemmin myös käytännön työssä. Työ antaa tärkeää tietoa analytiikkaohjelmistoista ja niiden erilaisista käyttötarpeista. Tuloksista voi päätellä, mitä ohjelmistoja tulisi käyttää mihinkin tarpeeseen. Opinnäytetyöprosessin aikana tekijä oppii arvioimaan myös itseään, mikä kehittää asiantuntijuutta.

Opinnäyte tarjoaa uutta tietoa analytiikkaohjelmistoista ja järjestää vanhaa uuteen muotoon. Tästä koituu hyötyä toimeksiantajalle ja muulle ammattiyhteisölle. Tekijälle hyödyllisyys on henkilökohtaista ja konkreettista.

### 2.2 Rajaukset



Opinnäytetyö keskittyy liiketoiminnan analytiikkaan. Työssä esitetään ensin liiketoiminnan analytiikkaa pintapuolisesti, koska pääpaino on itse ohjelmistoilla. Työstä jää kokonaan pois web-analytiikka. Opinnäytetyö tulee keskittymään työpöytäympäristöissä toimiviin ohjelmistoihin, ja ne toimivat Microsoft Windows-, Mac OS X- ja/tai Linux-käyttöjärjestelmissä. Kaikki tutkittavat ohjelmistot toimivat vähintään Microsoft Windows-käyttöjärjestelmässä. Runsaan ohjelmistotarjonnan vuoksi rajausta tutkittavien ohjelmistojen kohdalla oli välttämätöntä tehdä. Toimeksiantaja suoritti ohjelmistojen rajaamisen, ja tutkittavaksi valittiin kaupallisia ohjelmistoja ja avoimen lähdekoodin ohjelmistoja yhteensä viisi kappaletta. Ohjelmistojen selvittämisessä tullaan tarkastelemaan niitä mahdollisimman laaja-alaisesti. Ohjelmistoista tutkitaan niiden ominaisuuksia, toimintoja, käytettävyyttä, tukipalveluita ja hinnoitteluja.

### **2.3 Työn rakenne**

Opinnäytetyö voidaan jakaa karkeasti kahteen osaan. Aluksi luodaan yleiskatsaus liiketoiminnan analytiikkaan, jonka jälkeen siirrytään ohjelmistojen selvitykseen. Työ alkaa johdannolla ja taustatietojen läpikäymisellä. Seuraavaksi luodaan yleissilmäys liiketoiminnan analytiikkaan ja käydään lävitse yleisimmät käsitteet. Tämän jälkeen selvitetään mitä analytiikkaohjelmistot ovat, jonka jälkeen siirrytään itse ohjelmistojen selvitykseen. Ohjelmistojen selvityksen jälkeen seuraavat johtopäätökset ohjelmistoista ja lopuksi pohdinta.

### 3 ANALYTIikka LIKETOIMINNASSA

Mikä tekee päätöksien teosta hankalaa liiketoiminnassa? Epävarmuus kuuluu varmasti suurimpiin haasteisiin. Jos tietäisimme kuinka paljon tietyllä tuotteella tulee olemaan kysyntää, niin voisimme paremmin suunnitella ja aikatauluttaa tuotantoa. Entä jos tietäisimme kuinka kauan projektin eri vaiheet vievät aikaa, jolloin voisimme paremmin arvioida projektin kustannuksia ja sen valmistumispäivää. Toinen päätöksen tekoa yleensä hankaloittava tekijä on useat vaihtoehdot. (Camm ym. 2015, 5.)

Analytiikan avulla voidaan mahdollistaa myynnin osuvuutta tarjoamalla asiakkaille juuri oikeita tuotteita ja palveluita juuri oikeaan aikaan. Tuotteiden ja palveluiden hinnoittelua pystytään optimoimaan niin, että niistä saadaan korkein hinta, jolla kapasiteetti saadaan myytyä loppuun. Kysyntää pystytään ennakoimaan paremmin ja tuotteiden saatavuutta optimoimaan. Analytiikka tarjoaakin monia mahdollisuuksia yrityksen liiketoiminnan kasvuun. (Helkiö 2012.)

Analytiikkaa on mahdollista soveltaa moninaisesti ja erilaisesti riippuen toimialasta, strategioista ja liiketoimintaprosesseista. Ennakoiva analytiikka ja tiedonlouhinta soveltuvat hyvin segmentointityyppisiin kysymyksiin ja numeeriseen ennustamiseen. Tiedon visualisoinnilla puolestaan voidaan löytämään esimerkiksi korrelaatioita, poikkeamatietoja ja trendejä. (Helkiö 2012.)

#### 3.1 Analytiikka

Analytiikka ei ole yksiselitteinen termi. Analytiikalla tarkoitetaan datan laajaa hyväksikäyttöä, tilastollista ja kvantitaavista analyysia, ennustavia ja selittäviä malleja sekä päätöksenteon, johtamisen ja toiminnan perustamista tosiasioihin. Analyttisellä toiminnalla voidaan tuottaa aineistoa päätöksenteon tueksi tai ohjata automatisoitua päätöksentekoa. Käytännössä analyysin tekoon ei tarvita muuta kuin kynä, paperi ja laskukone, mutta nykyään on parasta käyttää teknologian tarjoamia ratkaisuja. (Davenport & Harris 2007, 26.)

Yksinkertaisuudessaan analytiikka muuntaa dataa hyödylliseksi tiedoksi. Analytiikka alkaa olla jo vanha termi, ja sitä käytetään kuvaamaan kaikenlaista analytiikkaa, ei pelkästään liiketoiminnallista. Tyypillinen esimerkki analytiikasta on analytiikan hyödyntäminen sää tietojen tutkimisessa ja tilastoimisessa. Analytiikka pelkästään on yksi termeistä, jota käytetään puhuttaessa analytiikasta liiketoiminnassa.

### **3.2 Liiketoiminta tiedonhallinta**

Liiketoiminnasta ja analytiikasta puhuttaessa termi liiketoiminta tiedonhallinta eli Business Intelligence (BI) nousee usein esille. Liiketoimintatiedon hallinnalla tarkoitetaan toimintaa, jonka avulla yritys kerää, analysoi, jakaa ja hyödyntää yrityksen kannalta oleellista liiketoimintatietoa. Sen tehtävänä on hankkia merkitykselliseksi arvioitua tietoa eri lähteistä sekä varastoida ja luokitella sitä jatkokäyttöä varten. Liiketoimintatiedon hallinnan keskeisenä tehtävänä on analysoida ja yhdistellä irrallisia tiedonpaloja eri merkityksien ja asiayhteyksien ymmärtämiseksi ja jakaa näin uutta syntyneitä tietoja kaikille sitä tarvitseville päätöksentekijöille. Parhaimmassa tapauksessa liiketoimintatiedon hallinta toimii proaktiivisesti päätöksenteon tukena tarjoten ennakkovaroituksia liiketoimintaympäristön tapahtumista ja niiden seurauksista yritykselle. Tarkoituksena on mahdollistaa parempien päätöksien tekeminen ja sitä kautta tuloksekkaampi liiketoiminta. (Laihonen ym. 2013, 45-46.)

Tosiasiasa kaikki yritykset harjoittavat liiketoimintatiedon hallintaa ainakin jossain muodossa. Toimintaa ei vain aina tiedosteta. Tietoa saadaan ja hankitaan sitä suuremmin ajattelematta. Sen merkitystä pohditaan ja siihen nojaten tehdään päätöksiä ilman, että sitä huomataan liiketoimintatiedon hallinnaksi. Kun näitä kaikkia eri toimintoja suoritetaan systemaattisesti yhdessä, ne muodostavat liiketoimintatiedon hallintaprosessin. Tehtävät on jaoteltu viiteen eri vaiheeseen: tietotarpeiden määrittely, tiedon hankinta, tiedon prosessointi ja analysointi, tiedon jakaminen sekä tiedon hyödyntäminen ja palaute. Vaikka vaiheet on jaoteltu osiin, ovat ne osittain myös päällekkäisiä. (Laihonen ym. 2013, 46.)



KUVA 1. Liiketoimintatiedon hallinnan prosessi (Laihonen ym. 2013, 46)

### 3.3 Liiketoiminnan analytiikka

Liiketoiminnan analytiikka eli Business Analytics (BA) on myös tärkeä käsite analytiikasta ja liiketoiminnasta puhuttaessa. Liiketoiminnan analytiikassa aivan kuten liiketoimintatiedon hallintakin on prosessi, jossa muunnetaan dataa oivalluksiksi, joilla pyritään tekemään parempia päätöksiä. Liiketoiminnan analytiikassa tuotokset nähdään usein ennemminkin tavoitteina kuin vaihtoehtoina päätöksenteolle. (Camm 2015, 5.) Liiketoiminnan analytiikalla voidaan käsittää osaamista, teknologioita ja toimintatapoja jatkuvassa iteratiivisessa prosessissa, jossa liiketoiminnan dataa hyödynnetään paremman liiketoiminnan toteuttamiseksi ja suunnittelun tueksi. Liiketoiminnan analytiikka pyritään vastaamaan kysymyksiin miksi jotain tapahtuu, mitä tapahtuu seuraavaksi ja mitä voisi tapahtua. (Strickland 2014, 7.)

Liiketoiminnan analytiikka pitää sisällään analytiikan, raportoinnin ja erinäiset kyselyt. Liiketoiminnan analytiikka tuo yhteen edistyneen tiedon hallinnan, analytiikan metodit ja tulosten esittämisen. Nimestä huolimatta liiketoiminnan analytiikka ei ole rajoitettu prosessi ja termistä huolimatta liiketoiminnan analytiikkaa voivat hyödyntää myös voittoa tavoittelemattomat yritykset. Sana viittaa välttämättömään tehtävään tai tavoitteeseen, joita analytiikalla suoritetaan. (Davenport 2010, 4-6.)

### 3.4 Analytiikka, liiketoiminta tiedonhallinta ja liiketoiminnan analytiikka

Kaikki kolme termiä ovat yleisiä puhuttaessa liiketoiminnasta ja analytiikasta. Liiketoiminnan analytiikan ja liiketoimintatiedon hallinnan välille on vaikea tehdä selkeää eroa, koska ne ovat käsitteinä hyvin lähellä toisiaan ja menevät osittain päällekkäin. Alla oleva taulukko selvittää termejä ja niiden eroavaisuuksia. Taulukossa mainitaan myös yleisiä analytiikkatyyppisiä, joita tullaan avaamaan seuraavassa osiossa.

TAULUKKO 1. Analytiikan, liiketoiminnan analytiikan ja liiketoimintatiedon hallinnan luonteet (Schniederjans ym. 2014, 6)

<b>Tyypit</b>	<b>Analytiikka</b>	<b>Liiketoiminnan analytiikka</b>	<b>Liiketoiminta tiedonhallinta</b>
<b>Liiketoiminnan suunnittelu rooli</b>	Mitä tapahtuu ja mitä tulee tapahtumaan?	Mitä tapahtuu nyt, mitä tulee tapahtumaan?	Mitä tapahtuu nyt ja miten olemme käsitelleet asiaa?
<b>Kuvailevan analytiikan käyttö</b>	Kyllä	Kyllä	Kyllä
<b>Ennustavan analytiikan käyttö</b>	Kyllä	Kyllä	Ei (ainoastaan historiallisesti)
<b>Ohjailevan analytiikan käyttö</b>	Kyllä	Kyllä	Ei (ainoastaan historiallisesti)
<b>Kaikkien kolmen käyttö yhdisteltynä</b>	Ei	Kyllä	Ei
<b>Liiketoiminta painotteen</b>	Ehkä	Kyllä	Kyllä
<b>Painottuu tiedon varastointiin ja hallintaan</b>	Ei	Ei	Kyllä
<b>Keskittyminen liiketoiminnan arvon kehittämiseen</b>	Ei	Kyllä	Ei

### 3.5 Analytiikan tyyppejä

Liiketoiminnan analytiikka voi pitää sisällään kaiken aina yksinkertaisista raporteista metodeihin, joilla löydetään parhaat toimintamuodot. On olemassa monia eri analytiikkatyyppejä. Seuraavaksi otetaan tarkasteluun kolme yleisintä, jotka ovat kuvaileva, ennustava ja ohjaileva analytiikka. Nämä analytiikan tyypit voidaan esittää myös kysymyksinä: mitä tapahtui, miksi se tapahtui ja mitä tulee tapahtumaan. Näitä voidaan käyttää yhdistelminä tai erikseen. Esimerkiksi yritys voi käyttää vain kuvailevaa analytiikkaa päätöksen teon tueksi tai vaihtoehtoisesti eri analytiikkatyyppien yhdistelmiä saadakseen tarpeeksi tietoa suunnittelun ja päätöksen teon tueksi. (Schniederjans ym. 2014, 3-4.)

#### 3.5.1 Kuvaileva analytiikka

Kuvailevalla analytiikalla (descriptive analytics) tarkoitetaan tekniikoita, joilla selvitetään, mitä on tapahtunut tähän mennessä ja mitä tällä hetkellä tapahtuu kuvaten sekä menneitä että nykyä. Kuvailevaa analytiikkaa on mahdollista käyttää monella eri tavalla, esimerkiksi tarkkaan työvoimatietojen ymmärtämiseen. Analytiikalla voidaan porautua työvoiman alaryhmiin määriteltyjen piirteiden perusteella ja selvittää ryhmien suhteita ja eroja toisiinsa. Voidaan myös tutkia ryhmien käyttäytymisen ja mielenkiinnon kohteiden muutoksia niin nykyajassa kuin menneessä. (Fitz-Enz 2010, 12.)

Onkin monia tapoja ymmärtää työvoimaa. Melkein minkä tahansa ryhmän ominaisuuksia voidaan tutkia, tarvitaan vain muuttujat, jotka luovat ominaisuudet, kuten ikä, koulutus, taidot ja kiinnostukset. Yksinkertaisuudessaan tämä on työvoimasegmentointia, aivan kuten markkinoinnissa tehdään asiakassegmentointia. Kaiken tämän tarkoituksena on tietenkin parantaa sijoitetun pääoman tuottoa. (Fitz-Enz 2010, 12.)

#### 3.5.2 Ennustava analytiikka

Ennustavalla analytiikalla (predictive analytics) tarkoitetaan tekniikoita, joiden avulla voidaan analysoida historia- ja nykytietoja, jotta voidaan päätellä, mitä mahdollisesti tapahtuu tai ei tapahdu. Kyseinen analytiikka käyttää kuvailevan analytiikan kuvauksia

apunaan tulevaisuuden mallintamiseksi. Ennustava analytiikka on yleensä vain suuntaa antava, eikä pelkästään sen perusteella voida vielä lieventää tai arvioida tulevia riskejä etenkin jatkuvasti arvaamattomilla ja laajentuvilla globaaleilla markkinoilla. Ennustava analytiikka ilmaisee tulevaisuuden mahdollisuuksia ja auttaa tekemään päätöksiä riskien vähentämiseksi. (Fitz-Enz 2010, 13.)

Mallinnukset ovat esimerkkejä ennustavasta analytiikasta. Kun esimerkiksi rakennetaan menestyksekkäs mallinnus, joka perustuu työvoiman taitoihin, ominaisuuksiin ja kokemuksiin, pystytään kasvattamaan todennäköisyyttä oikeiden ihmisten valintaan koulutettaviksi ja palkattaviksi. Eri mallinuksilla voidaan löytää myös eroja yksilöiden väliltä havaitsemalla syy-seuraussuhteita ja hankitun tiedon avulla voidaan luoda tehokkaita algoritmeja henkilöstöresurssien optimointiin. (Fitz-Enz 2010, 13.)

### **3.5.3 Ohjaava analytiikka**

Ohjaava analytiikka (prescriptive analytics) tarkoittaa tekniikoita, joiden avulla pystytään laskennallisilla malleilla kehittämään ja analysoimaan vaihtoehtoja organisaation toiminnalle. Analytiikka mahdollistaa päättäjien tehdä päätöksiä, joilla voidaan minimoida liiketoiminnan eri riskejä sekä lisätä pääoman tuottoa. Hyvin toteutettuna ohjaava analytiikka vähentää epävarmuutta ja parantaa liiketoimintapäätöksiä. Tulee kuitenkin muistaa, että yhdelläkään analytiikan sovelluksella ei ole mahdollista ennustaa tulevaisuutta ehdottomalla varmuudella. (Fitz-Enz 2010, 13.) Ohjaava analytiikka on kuitenkin haasteellista määrittää, koska se on niin lähellä ennustavaa analytiikkaa. (Vorhies 2014.)

## 4 ANALYTIKKAOHJELMISTOT

Ensimmäisten analyttisten menetelmien käyttö oli vähäistä ja satunnaista. Jo 1960-luvulla eräät tutkijat ja yksityisyrittäjät alkoivat kokeilla tietokonejärjestelmien kykyä tukea päätöksentekoa ja analysoida dataa. Sovelluksia käytettiin analyttisiin, toistuviin ja jokseenkin rajattuihin toimintoihin, kuten tuotannon suunnitteluun ja kuljetusten reititykseen. Näitä ohjelmistoja kutsuttiin päätöstukijärjestelmiksi (decision support systems, DSS). DSS:n juurien voidaan katsoa ulottuvan toisen maailmansodan aikaisiin ja sen jälkeisiin sotilaallisiin sovelluksiin. Nykyaikaisia tietokoneita ei tosin silloin vielä ollut olemassa. Tietokoneilla suoritettujen tilastollisten analyysien teko yleistyi 1970-luvulla, kun SPSS:n ja SAS Instituten kaltaiset yritykset alkoivat lanseerata ohjelmistokokonaisuuksiaan. DSS-päätöstukijärjestelmät muovautuivat puolestaan johdon tietojärjestelmiksi. Ylin johto käytti dataa ja tietokonesovelluksia liiketoiminnan valvomiseen ja raportointiin eikä päätöksenteon apuna. Osasyynä voidaan pitää johdon haluttomuutta käyttää ohjelmistoja itse. (Davenport & Harris 2007, 31-32.)

Tiedon analysoinnista alkoi tulla tärkeää, koska muun muassa toiminnanohjaus- ja myymäläjärjestelmät tuottivat paljon dataa. Myöhemmin internet kasvatti tiedon tulvaa vielä enemmän. Nykyisin useimmilla suurilla yrityksillä on käytössään jonkinlaisia analytiikkaohjelmistoja. Analytiikkaohjelmistojen tarjonta on laaja ja kasvaa edelleen. SPSS:n ja SAS:n kaltaiset ohjelmistoyhtiöt ovat tarjonneet jo yli 40 vuoden ajan asiakkaidensa käyttöön jatkuvasti parempia analyysijä ja niiden kehitys vain jatkuu. SPSS ja SAS kuuluvat kaupallisiin ratkaisuihin. Näiden lisäksi on tarjolla avoimen lähdekoodin ohjelmistoja, kuten RapidMiner, R-ohjelma ja Knime. (Davenport & Harris 2007, 32-33.)

Pystymme nykyään ymmärtämään ja hyödyntämään erilaista dataa. Teknologiasta etsitään tukea johtamiseen. Tehokkaat tietokoneet ja analytiikkaohjelmistot mahdollistavat nopeiden analyysien teon ja suurien tiedostojen käsittelyn. Analytiikkaohjelmistoilla pystytään suorittamaan erilaisia tilastollisia analyysijä ja tukemaan näin päätöksien tekoa.

Ohjelmistoja löytyy runsaasti ja niistä löytyy myös eroja. Niillä on omat käyttötarkoituksensa ja ohjelmistokohtaiset painotukset. Ohjelmistot ovat visuaalisesti eri näköisiä ja toimivat eri tavoilla. Myös hinnoittelussa ja tukipalveluissa löytyy eroja.



## 5 TUTKITTAVAT OHJELMISTOT

Ohjelmistoseelvitykseen valittiin täysin kaupallisista ohjelmistoista Alteryx Designer ja IBM SPSS Statistics, sekä avoimeen lähdekoodiin perustuvat KNIME, PSPP ja RapidMiner Studio. Kaupallisista ohjelmistoista oli tarjolla useita eri versioita, koska selvitys sisältää myös täysin ilmaisia ohjelmistoja, työhön valittiin kaupallisista halvimmat ohjelmistoversiot. Ohjelmistot toimivat Microsoft Windows-, Mac OS X- ja/tai Linux-käyttöjärjestelmillä. Kaikki tutkittavat ohjelmistot toimivat ainakin Microsoft Windows-käyttöjärjestelmällä, jossa niitä myös tutkittiin.

Alustava tarkastelu osoitti, että kaikki analytiikkaohjelmistot tarjosivat tarvittavat toiminnot liiketoiminta-analytiikan toteuttamiseen. Ohjelmistojen testauksessa testattiin ja tutkittiin niiden ominaisuuksia, toimintoja, käytettävyyttä ja tukipalveluita. Apuna testauksessa käytettiin myös testiaineiston tiedostoja. Testiaineistot sisälsivät kuvitteellisia liiketoimintatietoja CSV-tiedostomuodoissa, ja niitä hyväksi käyttäen toteutettiin erilaisia analyysejä ohjelmistoissa. Lisäksi selvitettiin ohjelmistojen mahdolliset kustannukset. Tarkoituksena oli muodostaa tiivis ja perusteellinen katsaus tutkittavaksi valittujen ohjelmistojen sopivuudesta eri käyttötarpeisiin. Ohjelmistot käytiin lävitse aakkosjärjestyksessä.

Toiminnoissa ja ominaisuuksissa keskityttiin ohjelmistojen päätoimintoihin ja ominaisuuksiin. Miten ohjelmisto toimii ja millainen sen käyttöliittymä on. Miten tietoa tuodaan ohjelmistoon ja muokataan. Tutkittiin ja selvitettiin myös itse analyyysien toteuttamista. Liiketoiminnan kannalta oleelliset kuvailevan ja ennustavan analytiikan mahdollisuudet käytiin myös läpi. Ohjaavan analytiikan tekniikat jätetään käymättä, koska ne sivuavat ennustavan analytiikan tekniikoita. Tutkittavana oli myös tiedon ulostulo, visualisointi ja raportointi.

Käytettävyyden arvioinnissa käytettiin hyväksi Nielsenin määrittämää viiden joukkoa. Jakob Nielsenin mukaan käytettävyys koostuu opittavuudesta, tehokkuudesta, muistettavuudesta, virheiden vähäisyydestä ja yleisestä miellyttävyydestä. Opittavuudella tarkoitetaan tuotteen peruskäytön oppimisen nopeutta. Opittavuutta pidetään yleensä yhtenä tärkeimmistä käytettävyyden tekijöistä, koska hyvä opittavuus mahdollistaa tuotteen no-

pean käyttöönoton. Nopea käyttöönotto tarkoittaa tulosten nopeaa saavuttamista ja helpompaa pääsyä sisälle tuotteen syvempiin ominaisuuksiin. Tehokkuudella puolestaan selvitetään, kuinka hyvän tuotettavuuden tasoon käyttäjä saavuttaa. Käyttäjän tulee kuitenkin olla saavuttanut tietty oppimistaso tuotteen käytöstä, jotta tehokkuutta voidaan kunnolla arvioida. Muistettavuus on myös oleellista, koska tällöin kerran opittuja asioita ei tarvitse opetella heti uudestaan käyttötaujan jälkeen. Virheettömyydellä puolestaan selvitetään tuotteen käyttöä virheettömästi. Jos toiminto ei johda haluttuun lopputulokseen, on se virhe. Viimeisenä on miellyttävyys, eli tuotteen käyttäminen tulee olla miellyttävää ja käyttäjän tulee olla tyytyväinen siihen. (Nielsen 1993, 27–31.) Tutkittavien ohjelmistojen käytettävyyttä tullaan testaamaan näitä termejä käyttäen.

Tukipalveluiden selvittämisessä käytiin lävitse kaikki mahdolliset tukipalvelut. Tukipalveluihin laskettiin henkilökohtaiset puhelin- ja sähköpostituet. Yhteisöllisiin tukiin laskettiin muun muassa ylläpidettävät foorumit ja blogit. Tarjolla olevat opetusmateriaalit ja muut ohjeet kuuluivat myös selvitettäviin tukiin.

Monesti ohjelmistoja rajataan suoraan hinnan perusteella. Olikin oleellista selvittää myös mahdollinen hinnoittelu. Useita ohjelmistoja tarjotaan vielä monilla eri paketeilla, joiden tarjoamat ominaisuudet ja hinnat eroavat selvästi toisistaan.

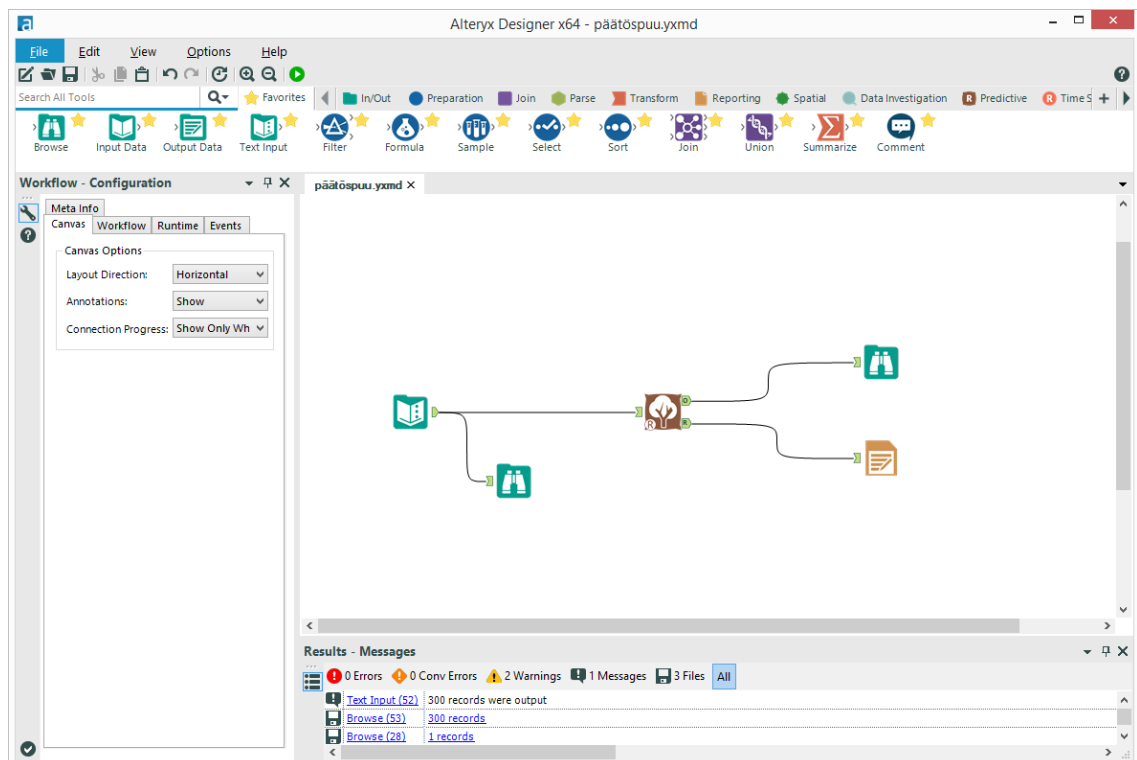
## **5.1 Alteryx Designer**

Alteryx Designer on amerikkalaisen Alteryx ohjelmistoyrityksen kehittämä analytiikka-ohjelmisto. Yritys tunnettiin aiemmin SRC LLC -nimellä ja sen perustivat vuonna 1997 Dean Stoecker, Olivia Duane Adams ja Ned Harding. Vuonna 2010 SRC LLC muutti nimensä Alteryx:ksi. (Alteryx 2015a.) Yrityksen ohjelmisto tarjontaan kuuluu Alteryx Designer –ohjelmisto, joka on suunniteltu varsinkin tiedon yhdistämiseen ja edistyneeseen analytiikkaan. Ohjelmiston vahvuuksia ovat sen tehokkaat ennustavan analytiikan ominaisuudet ja sillä on erityisesti liiketoimintaongelmien ratkomisen. Ohjelmisto on saatavissa Microsoft Windows-käyttöjärjestelmille.

### **5.1.1 Ominaisuudet ja toiminnot**

Alteryx Designer sisältää graafisen käyttöliittymän ja se toimii pääasiassa raahaa ja pudota -menetelmällä eli toimintoja raahataan hiiren avulla. Ohjelmiston päänäkymässä on työkalurivi, josta löytyvät kaikki tarvittavat toiminnot. Työkalurivi muodostuu välilehdistä, joista jokaisesta aukeaa lisää työkaluja. Työkaluvälilehtiin lukeutuu muun muassa syöttö/ulostulo-, tiedonvalmistelu-, yhdistä- ja ennustamistyökalut.

Analyysejä suoritetaan rakentamalla työnkulkuja. Työnkulku tilaa kutsutaan kanvaaksi. Kanvaalle raahataan halutut työkalut, jonka jälkeen ne yhdistetään toisiinsa. Kun työkaluista on rakennettu työnkulku, tulee se vielä lopuksi ajaa. Syöttö-työkalusta pääsee lisäämään aineiston analyysiä varten, kunhan se on ensin raahattu kanvaalle. Alteryx Designer tukee kaikkia yleisimpiä tiedosto- ja tietokantamuotoja. Jos aineistoa haluaa vielä muokata, tulee käyttäjän raahata työnkulkuun haluttu toiminto, kuten suodatus tai rajaaminen. Ohjelmistosta löytyvät kaikki tarvittavat toiminnot aineiston muokkaamiseen. Aineiston valmisteluun löytyy perinteisten muokkaamistyökalujen lisäksi aineistojen yhdistämistyökalu, jolla pystyy yhdistelemään eri aineistoja yhdeksi käsiteltäväksi aineistoksi. Aineistojen yhdistäminen kuuluu ohjelmiston vahvuuksiin. Esimerkiksi xml- ja csv-tiedostot voidaan yhdistää yhdeksi käsiteltäväksi aineistoksi. Tämä onnistuu yhdistämistyökalun käyttämisellä. Toiminto on erittäin kätevä esimerkiksi jos yrityksen omadata on toisessa tiedostomuodossa ja asiakasdata toisessa.



KUVA 2. Alteryx Designer näkymä

Itse analyysien teko ohjelmistossa suoritetaan myös raahattavien työkalujen avulla. Työkaluista on karsittu ylimääräiset ja jätetty vain oleelliset, mutta näidenkin takaa avautuu vielä lisää toimintoja. Ohjelmistolla onnistuu niin kuvaavien kuin ennustavienkin analyysien teko. Kaavatyökalu on esimerkiksi todella tehokas aineiston käsittelyyn. Työkalulla luodaan komentoja, joilla aineistosta voidaan selvittää esimerkiksi yhtäläisyyksiä. Kaikille yleisimmille kuvaavan tilastoinnin toiminnoille, kuten frekvenssille ja ristiintaulukoinnille, on omat työkalunsa. Ohjelmiston omiin erikoisuuksiin voidaan laskea myös sen kyky analysoida paikkatietoja. Alteryx Designer sisältää työkalut, joilla on mahdollista selvittää karttanäkymään esimerkiksi asiakkaiden sijainnit myyntialueeseen katsottuna. Ohjelmiston vahvimpiin toimintoihin kuuluu sen kyvyt suorittaa ennustavaa analytiikkaa. Työkalut perustuvat R-ohjelmointikieleen, mikä selittää niiden tehokkuutta. Ennustavista työkaluista löytyy tuttuja työkaluja, kuten regressiot ja päätöspuut, mutta myös täysin ohjelmiston omia työkaluja. Näihin työkaluihin kuuluvat muun muassa trendienennustamis- ja ostoskorianalyysityökalut. Näillä työkaluilla pystytään luomaan tehokkaita analyyseja tulevaisuuden näkymistä.

Tulokset voidaan tallentaa lisäämällä ulostulotyökalu työnkulun jatkeeksi. Tämän jälkeen analyysin tulos on mahdollista tallentaa esimerkiksi CSV- tai HTML-tiedostomuodoissa. Raportointityökaluvälilehdeltä aukeaa vielä enemmän työkaluja tulosten tallentamiseen. Sieltä löytyy myös työkaluja muun muassa tulosten visualisointiin ja tekstikenttien lisäämiseen. Raportit rakennetaan työkalu kerrallaan. Lopuksi valmiin raportin voi tallentaa kuvantamistyökalun avulla esimerkiksi PDF-tiedostona tai PNG-kuvana. Raportointimahdollisuudet ohjelmistolla ovat rajalliset, koska työkaluja on rajallinen määrä. Kunnollisten raporttien rakentamiseen on parasta käyttää toista ohjelmaa. Visualisointiin löytyvät perinteiset taulukot, ympyräkuvaajat, pylväskuvaajat ja muutama vähän erikoisempi vaihtoehto, kuten tornadokuvaajat ja kuplakuvaajat. Kuvaajia on mahdollista muokata niille varatussa tilassa, mutta visuaalisesti hienompien kuvaajien tekeminen sillä ei onnistu. Visuaalisesti hienompien kuvaajien tekoon ja raportointiin kannattaa käyttää esimerkiksi Tableau-ohjelmistoa, jonka kanssa Alteryx tekee myös yhteistyötä.

### **5.1.2 Käytettävyys**

Ohjelmistossa on panostettu visuaalisuuteen ja yksinkertaisuuteen. Turhaksi katosotut toiminnot on rajattu pois eikä näkymässä näy turhaa tietoa tai toimintoja. Käyttäjän ei tarvitsekaan käyttää liikaa aika ohjelmiston toimintojen opettelemiseen. Työkalut sisältävät tekstin lisäksi myös omat ikonit ja kuvat, jotka helpottavat niiden löytämistä ja käyttämistä. Ne on myös organisoitu järkevästi omiin kategorioihin. Usean työkalun taakse on kumminkin piilotettu enemmän toimintoja, joten käyttäjä voi välillä joutua arvuuttelemaan minkä työkalun takaa haluttu toiminto löytyy. Oikean työkalun löytämisessä auttaa onneksi pikahakuikkuna, joka tunnistaa myös työkalujen takana olevat toiminnot. Analyysien rakentaminen on suoraviivaista ja niiden oppiminen ei vaadi suurempia koulutuksia. Ohjelmistosta pääsee myös käsiksi opastettuihin ohjeisiin, jos käyttäjällä on vaikeuksia toimintojen toteuttamisessa. Ohjelmiston pariin onkin helppo palata käyttötaukojenkin jälkeen.

Kun ohjelmiston käytön on oppinut, sillä pystyy suorittamaan useita erilaisia analyysejä. Työkalujen takaa avautuu tehokkaita toimintoja, joilla analyysien suorittaminen onnistuu ilman suurempia haasteita. Tehokkuuden haasteeksi nousee välillä nopeasti työkaluista täyttyvä kanvas. Jos tarkoitus on tehdä monimutkaisempia analyysejä, työkalujen määrä kasvaa ja niiden sijoittaminen kanvaalle tulee haasteeksi. Pahimmassa tapauksessa työtilasta tulee sekava ja vaikeasti hahmotettava. Isompien analyysien yhteydessä ohjelmisto käy myös huomattavasti hitaammin. Ohjelmistoa vaivaa välillä muutenkin hitaus aineistojen käsittelyssä. Vaikka ohjelmistolla pystyy suorittamaan tehokkaasti analyysejä, niin aineistojen käsittely ennen itse analyysiä on työlästä. Käyttäjän tehdessä virheitä ohjelmisto antaa niistä virheilmoituksia, jotka helpottavat niiden korjaamista. Ohjelmisto ei kuitenkaan aina suostu ajamaan työnkulkuja ja ei anna mitään virheilmoitusta. Tällöin virheen löytämisessä voi kestää pidempään. Ohjelmiston käyttäminen on enimmäkseen miellyttävää. Sen käyttö on helposti ymmärrettävissä ja graafisen käyttöliittymän kanssa työskentely on mielekästä.

### **5.1.3 Tukipalvelut**

Alteryx Designer on kaupallinen ohjelmisto ja tarjoaa sen käyttäjille aktiivisesti tukipalveluita, jotka on lueteltu ohjelmiston kotisivuilla. Nopeata apua tarvitsevat voivat olla suoraa yhteydessä live-chattiin. Soittaminen on myös mahdollista, jos käyttäjä tarvitsee

nopeasti apua. Monimutkaisempien kysymysten osalta on mahdollista lähettää tuelle sähköpostia, jolloin käyttäjä voi jakaa samalla vaikka kuvankaappauksia ja tiedostoja ongelman ratkaisemiseksi.

Alteryx-sivuilta löytyy myös yhteisötukea. Käyttäjät voivat kirjoittaa ja selata keskustelualueella viestejä. Sivuilta löytyy myös artikkeleita ja blogi osio. Yhteisöpalveluissa liikkuu aktiivisia käyttäjiä, mutta vaikeisiin kysymyksiin vastausten saannissa voi kestää kauankin. Sivustoilta löytyy myös laajat määrät opetusmateriaaleja ohjelmiston käytöstä ja toiminnoista. Materiaaleihin lukeutuu tekstien lisäksi myös videoita, joissa käydään askel askeleelta lävitse ohjelmiston käyttöä.

#### 5.1.4 Hinnoittelu

Alteryx Designeria on mahdollista kokeilla 14 päivää ilmaiseksi, jonka jälkeen täytyy ostaa maksullinen versio. Ohjelmistovaihtoehdot on lueteltu alla olevassa taulukossa (Taulukko 2).

TAULUKKO 2. Alteryx Designer hinnoittelu. Hinnat on muunnettu dollareista euroiksi (Alteryx 2015b, muokattu)

Ohjelmisto	Ominaisuudet	Hinta
Alteryx Designer	Datan yhdistäminen, edistynyt analytiikka	4641,44€, per käyttäjä, 1 vuosi
Alteryx Designer w/Spatial	Datan yhdistäminen, edistynyt analytiikka, analytiikkavalmista dataa, analytiikkatyökaluja paikkadatan hyödyntämiseen	15096,73€, per käyttäjä, 1 vuosi
Alteryx Designer w/Data	Datan yhdistäminen, edistynyt analytiikka, analytiikkavalmista dataa, analytiikkatyökaluja pakatun datan hyödyntämiseen	34845,63€, per käyttäjä, 1 vuosi

#### 5.1.5 Yhteenveto

Alteryx Designer tarjoaa nopean tavan käsitellä erilaisia liiketoiminta-aineistoja. Ohjelmisto vakuuttaa etenkin tehokkaiden ennustavan analytiikan työkaluillaan. Ennustavan analytiikan lisäksi sillä onnistuvat muutkin analyysit, joihin lukeutuvat myös paikkatietoanalyysit. Aineistojen yhdistäminen kuuluu myös ohjelmiston hienouksiin. Käyttäjä voi halutessaan analysoida ja yhdistää eri tiedostomuotoja. Sitä ei ole kumminkaan tarkoitettu luomaan näyttäviä kuvaajia ja raportteja.

Analyysin rakentaminen käy vaivattomasti yhdistelemällä eri työkaluja. Ohjelmiston valtteihin kuuluu myös sen miellyttävä graafinen käyttöliittymä. Ohjelmistoa on helppo käyttää ja sen käyttämiseen tarvitsee vain perustiedot data-analyysien suorittamisesta. Ohjelmistoa kuitenkin vaivaa monimutkaisten analyysien kohdalla kanvaan nopea täyttyminen työkaluista. Monimutkaisimpien analyysien kanssa itse ohjelmistokin hidastuu. Ohjelmistoa vaivaa muutenkin välillä hitaus aineistojen käsittelyssä. Ohjelmisto on myös todella hintava verrattuna markkinoiden muihin ratkaisuihin. Toisaalta kallista hintaa selettää ohjelmiston mukana tulevat tehokkaat työkalut ja erinomaiset tukipalvelut.

## **5.2 IBM SPSS Statistics**

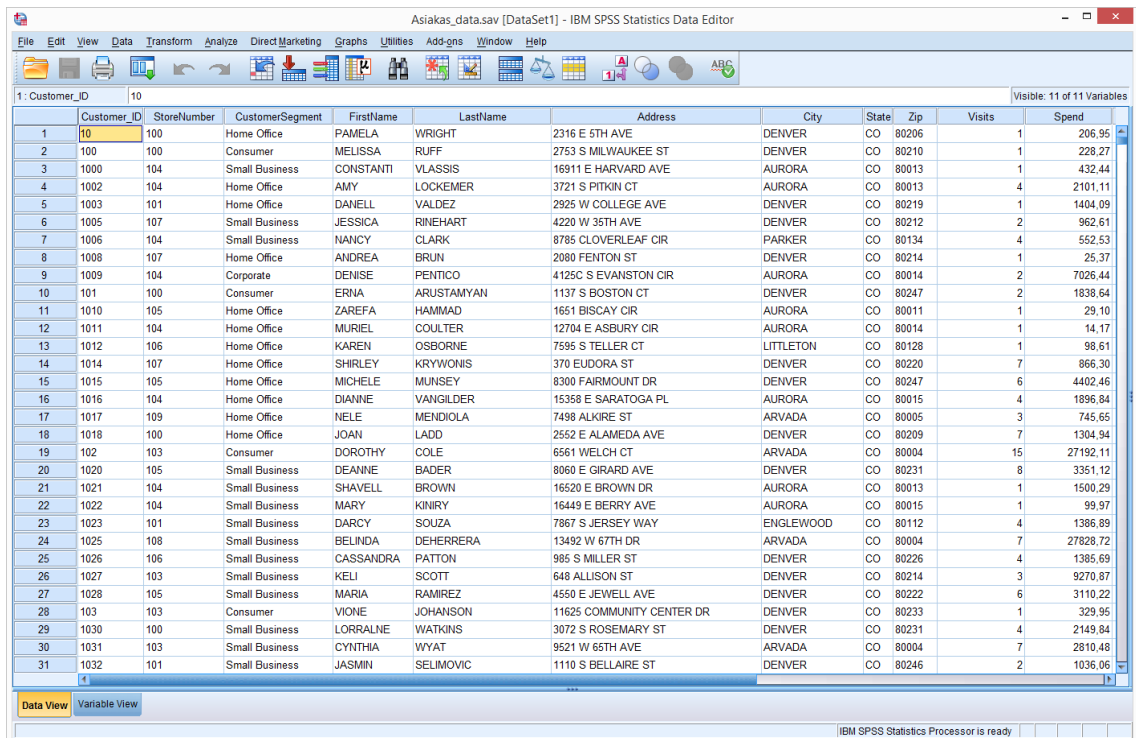
IBM SPSS (Statistical Package for Social Science) Statistics on IBM-yrityksen tuottama tilastotieteelliseen analyysiin suunniteltu ohjelmisto. Se kuuluu tunnetuimpiin analytiikkaohjelmistoihin ja sitä käytetään monilla eri aloilla. Ensimmäinen versio ohjelmistosta julkaistiin vuonna 1968 ja sen kehittämisen aloitti Stanfordin yliopiston valtiotieteen opiskelija Norman Nie. IBM SPSS Statistics on yksi yleisimmin käytetty ohjelmisto tilastotieteellisten analyysien tekoon. Ohjelmistoa käytetään yleensä markkinointi-, tuote-, terveys-, lääketiede-, valtio-, koulutus- ja yritystutkimuksissa sekä monella muulla alalla. (Gupta M. & Gupta D. 2011, 142.) IBM SPSS Statistics tarjoaa tehokkaat taulukot analyysien suorittamiseen ja se on mahdollista saada Microsoft Windows- ja Mac OS X -käyttöjärjestelmille.

### **5.2.1 Ominaisuudet ja toiminnot**

Ohjelmiston käyttö perustuu pääosin vetovalikkoihin ja interaktiivisiin ikkunoihin. Monet toiminnot ovat valmiita käyttöön vetovalikoiden kautta, mutta ohjelmisto tarjoaa myös ohjelmointialustan, joka käyttää 4GL-komentosyntaksikieltä. Syntaksi-ikkunaan

kirjoittamalla voidaan siis myös ohjata ohjelmistoa. Ohjelmointialustalla pystyy suorittamaan muun muassa monimutkaisempia analyysejä.

Ohjelmiston aineistoikkuna sisältää kaksi taulukkoa; tietosivun ja muuttujasivun. Taulukot muistuttavat paljon muita taulukko-ohjelmistoja, kuten Microsoft Exceliä. IBM SPSS Statistics tarjoaa yksinkertaiset ja tehokkaat taulukot tiedon syöttämistä, selaamista ja muokkaamista varten.



Customer_ID	StoreNumber	CustomerSegment	FirstName	LastName	Address	City	State	Zip	Visits	Spend
1	100	Home Office	PAMELA	WRIGHT	2316 E 5TH AVE	DENVER	CO	80206	1	206.95
2	100	Consumer	MELISSA	RUFF	2753 S MILWAUKEE ST	DENVER	CO	80210	1	228.27
3	1000	Small Business	CONSTANTI	VLASSIS	16911 E HARVARD AVE	AURORA	CO	80013	1	432.44
4	1002	Home Office	AMY	LOCKEMER	3721 S PITKIN CT	AURORA	CO	80013	4	2101.11
5	1003	Home Office	DANELL	VALDEZ	2925 W COLLEGE AVE	DENVER	CO	80219	1	1404.09
6	1005	Small Business	JESSICA	RINEHART	4220 W 35TH AVE	DENVER	CO	80212	2	962.61
7	1006	Small Business	NANCY	CLARK	8785 CLOVERLEAF CIR	PARKER	CO	80134	4	552.53
8	1008	Home Office	ANDREA	BRUN	2080 FENTON ST	DENVER	CO	80214	1	25.37
9	1009	Corporate	DENISE	PENTICO	4125C S EVANSTON CIR	AURORA	CO	80014	2	7026.44
10	101	Consumer	ERNA	ARUSTAMYAN	1137 S BOSTON CT	DENVER	CO	80247	2	1838.64
11	1010	Home Office	ZAREFA	HAMMAD	1651 BISCAY CIR	AURORA	CO	80011	1	29.10
12	1011	Home Office	MURIEL	COULTER	12704 E ASBURY CIR	AURORA	CO	80014	1	14.17
13	1012	Home Office	KAREN	OSBORNE	7595 S TELLER CT	LITTLETON	CO	80128	1	98.61
14	1014	Home Office	SHIRLEY	KRYWONIS	370 EUODORA ST	DENVER	CO	80220	7	866.30
15	1015	Home Office	MICHELE	MUNSEY	8300 FAIRMOUNT DR	DENVER	CO	80247	6	4402.46
16	1016	Home Office	DIANNE	VANGILDER	15358 E SARATOGA PL	AURORA	CO	80015	4	1896.84
17	1017	Home Office	NELE	MENDIOLA	7498 ALKIRE ST	ARVADA	CO	80005	3	745.65
18	1018	Home Office	JOAN	LADD	2552 E ALAMEDA AVE	DENVER	CO	80209	7	1304.94
19	102	Consumer	DOROTHY	COLE	6561 WELCH CT	ARVADA	CO	80004	15	27192.11
20	1020	Small Business	DEANNE	BADER	8060 E GIRARD AVE	DENVER	CO	80231	8	3351.12
21	1021	Small Business	SHAVELL	BROWN	16520 E BROWN DR	AURORA	CO	80013	1	1500.29
22	1022	Small Business	MARY	KINIRY	16449 E BERRY AVE	AURORA	CO	80015	1	99.97
23	1023	Small Business	DARCY	SOUZA	7867 S JERSEY WAY	ENGLEWOOD	CO	80112	4	1386.89
24	1025	Small Business	BELINDA	DEHERRERA	13492 W 6TH DR	ARVADA	CO	80004	7	27828.72
25	1026	Small Business	CASSANDRA	PATTON	985 S MILLER ST	DENVER	CO	80226	4	1385.69
26	1027	Small Business	KELI	SCOTT	648 ALLISON ST	DENVER	CO	80214	3	9270.87
27	1028	Small Business	MARIA	RAMIREZ	4560 E JEWELL AVE	DENVER	CO	80222	6	3110.22
28	103	Consumer	VIONE	JOHANSON	11625 COMMUNITY CENTER DR	DENVER	CO	80233	1	329.95
29	1030	Small Business	LORRALNE	WATKINS	3072 S ROSEMARY ST	DENVER	CO	80231	4	2149.84
30	1031	Small Business	CYNTHIA	WYAT	9521 W 65TH AVE	ARVADA	CO	80004	7	2810.48
31	1032	Small Business	JASMIN	SEUMOVIC	1110 S BELLAIRE ST	DENVER	CO	80246	2	1036.06

KUVA 3. IBM SPSS Statistics näkymä

Ohjelmistoon haluttu aineisto tuodaan tiedostovetovalikosta valitsemalla avaa tai avaa tietokanta. Ohjelmisto pystyy lukemaan useita eri tiedosto- ja tietokantamuotoja. Avattua aineistoa pystyy vielä muokkaamaan aineistoikkunassa ennen analyysin aloittamista. Taulukon solujen, rivien ja sarakkeiden muokkaaminen onnistuu hiirellä valitsemalla. Tietoja voidaan nimetä uudelleen tai poistaa kokonaan. Muuttujasivulla pääsee lisäämään muuttujille arvoja ja vaihtamaan niiden luokkia. Haluttu tiedon suodatus ennen analyysin aloittamista voidaan suorittaa eri toiminnoilla. Tietovetovalikosta voi muun muassa valinnoilla tapaus- ja puolita-tiedosto muokata aineistoa. Toiminnot mahdollistavat esimerkiksi tiettyjen muuttujien pois jättämisen ja muun tiedon rajaamiseen. Edistyneet tiedonvalmistelu toiminnot ovat käytettävissä vasta kalliimmista ohjelmisto versioissa.



Kaikkiin analyysin suorittamista tarvittaviin toimintoihin pääsee niille varatun analyysivetovalikon kautta. Kuvailevat tilastointitoiminnot, joita käytetään erinäisten tietojen laskemisessa, kuuluvat ohjelmiston tunnetuimpiin toimintoihin. Ohjelmisto sisältää varsinkin kuvailevaan tilastointiin useita tehokkaita ratkaisuja. Frekvenssin, ristiintaulukoinnin ja monen muun kuvaavan tilastointimenetelmän avulla aineistosta löytää helposti yhteyksiä ja muutoksia. Todennäköisyysslaskennat onnistuvat korrelaatioiden, t-testien ja keskiarvojen avulla. Ennustavaa analytiikkaa on myös mahdollista harjoittaa muun muassa regressioiden avulla. Jos ohjelmistolla on tarkoitus suorittaa enemmänkin ennustavaa analytiikkaa, tulisi tällöin hankkia Professional-versio. Version mukana tulee nimittäin oleellisia ennustavia toimintoja, kuten päätöspuut. Analyysin suorittavia toimintoja ei vetovalikosta aukea satoja, mutta siitä löytyvät kaikki tarvittavat ja ne sisältävät aina itsessään lisää toimintoja.

Pelkkää aineistoa ja itse analyysejä on myös mahdollista visualisoida. Kuvaajien tekoon ohjelmistosta löytyy oma vetovalikkonsa, josta voi valita pylväskuvaajia, ympyräkuvaajia, viivakuvaajia ja paljon muuta. Graafisesti kuvaajista saa yksinkertaisen ja selkeän näköisiä. Kuvaajia pääsee luomaan omassa editorissa ja lisäksi ulostulonäkymässä valmiita kuvaajia voi muokata vielä toisessa editorissa. Ulostulonäkymästä on myös mahdollista ottaa tulokset talteen eri muodoissa, kuten PDF, Microsoft Word, HTML, JPEG ja PNG. Ohjelmiston ulostulonäkymää voikin käyttää raportoinnissa ilman toista ohjelmistoa. Tulee kuitenkin huomioida, että ulostulonäkymässä tulosten muokkaaminen on rajallista. Näkymässä pystyy vaihtamaan tulosten sijaintia ja lisäämään tekstikenttiä, mutta tämäkin on hieman kankeaa ja muokkaustoimintoja puuttuu.

### **5.2.2 Käytettävyys**

Käyttäjien on helppo lähestyä ohjelmistoa sen selkeän ulkoasun ansiosta. IBM SPSS Statistics muistuttaa suurelta osin muita Windows-ohjelmia, joten Windows ympäristöön tottuneet käyttäjät pitkälti tuntevat käyttöön tarvittavat työskentelytavat. Ohjelmisto sisältää myös Excel-tyyppisen taulukon, joka voi olla monelle jo entuudestaan tuttu.

Helposti opittavat vetovalikot ja ikonit helpottavat ohjelmiston käyttöä. IBM SPSS Statistics on erittäin käyttäjäystävällinen ja sen oppiminen on vaivatonta. Laajoja koulutuk-

sia ei tarvitse pitää, jotta uudet käyttäjät oppisivat ohjelmiston käytön. Intuitiivinen ulkoasu mahdollistaa yksinkertaisten analyysien tekemisen ja muuttujien muokkaamisen ilman suurempia selityksiä. Erikoisempien toimintojen käyttö voi aiheuttaa aluksi hankaluuksia, koska niitä tarkentavat selitykset puuttuvat.

Vetovalikoiden takaa suoritettavat analyysit ovat nopeasti ja tehokkaasti toteutettavissa. Valikoista löytää vaivatta haluamansa toiminnot ja opittujen monimutkaisten analyysien toteuttaminen onnistuu suhteellisen vaivattomasti. Aineiston perusmuokkaus onnistuu myös nopeasti taulukoissa ja eri toiminnoilla. Ohjelmiston pariin onkin helppo palata pidemminkin käyttötouon jälkeen, vaikka tarkemmat selitykset toiminnoilta puuttuvatkin. IBM SPSS Statistics voi kyllä kaatuilla suurten data määrien käsittelyn yhteydessä.

Datan ulostuloa on helppo ymmärtää, kun suorittaa muutamia analyyskejä, mutta kun suorittaa kymmeniäkin analyyskejä, ulostulolla on taipumus tulla todella isoksi ja vähän sekavaksikin. Loppujenlopuksi ohjelmiston käyttäminen on melko miellyttävä kokemus.

### **5.2.3 Tukipalvelut**

Kaupallisena ohjelmistoja IBM SPSS Statistics tarjoaa myös monipuolisen tuen ja ohjeistukset. Ohjelmiston omilta nettisivuilta käyttäjä pääsee käsiksi opetusmateriaaleihin, yhteisötukeen ja henkilökohtaiseen tukeen.

Käyttäjä voi ottaa yhteyttä tekniseentukeen IBM SPSS Statistics käyttöä tai asennusta koskien. Henkilökohtaista tukea saa sähköisesti että puhelimitse. Ohjelmiston verkkosivuilta löytyy myös yhteisötukea foorumeiden, blogien ja sosiaalisen median muodoissa, joissa kaikissa on aktiivista toimintaa. Sivustolta löytyy myös useita opetusmateriaaleja. Opetusmateriaalit sisältävät ohjeita aina ohjelmiston asentamisesta sen käyttöön. Materiaaleja on teksti- ja videomuodoissa.

### **5.2.4 Hinnoittelu**

IBM SPSS Statistics on kaupallinen ohjelmisto ja siitä on tarjolla neljä eri versiota: Base, Standard, Premium ja Professional ratkaisut. Jokainen versio jakautuu vielä erilaisiin lisenssivaihtoehtoihin. Jokaiseen ohjelmistoon saa 12 kuukauden lisenssin ja se koskee vain yhtä asennusta. Lisenssin mukana tulee myös 12 kk tuki. Alla taulukossa (Taulukko 3) on lueteltu eri ohjelmistovaihtoehtojen lähtöhinnat

TAULUKKO 3. IBM SPSS Statistics hinnoittelu (IBM 2015, muokattu)

Ohjelmisto	Ominaisuudet	Hinta
Base	Perusanalytiikkaominaisuudet	1100€, per käyttäjä, 1 vuosi
Standard	Perusanalytiikkaominaisuudet, raahaa ja pudota -toiminnot ja lineaarisen mallintamisen toimintoja	2441€, per käyttäjä, 1 vuosi
Professional	Perusanalytiikkaominaisuudet, raahaa ja pudota -toiminnot, lineaarisen mallintamisen toimintoja, edistyneen tiedon valmistelu, päätöspuut ja ennustaminen	4911€, per käyttäjä, 1 vuosi
Premium	Perusanalytiikkaominaisuudet, raahaa ja pudota -toiminnot, lineaarisen mallintamisen toimintoja, edistynyt tiedon valmistelu, päätöspuut ja ennustaminen, strukturoitu yhtälömallintaminen ja harvinaisten tapahtumien analyysit	7323€, per käyttäjä, 1 vuosi

### 5.2.5 Yhteenveto

Yksinkertaisuudessaan IBM SPSS Statistics on kattava ohjelmisto, jolla onnistuu datan analysointi. Ohjelmisto tarjoaa laajan kirjon analytiikan suorittamiseen tarvittavia toimintoja ja varsinkin sen kuvailevan analytiikan ominaisuudet ovat vaikuttavia. Ennustavaan analytiikkaan panostaessa tulisi kumminkin hankkia kalliimpi ohjelmistoversio. Ohjelmistolla onnistuu selkeiden kuvaajien ja raporttien teot, mutta näissäkin on huomattavia rajoituksia.

Ohjelmisto on helposti opittavissa ja erittäin helppokäyttöinen. Tehokkaat ja helposti opittavat vetovalikot kuuluvat ohjelmiston parhaimpiin puoliin. Kaupallisen ohjelmiston etuihin kuuluvat kattavat tukipalvelut. IBM SPSS Statistics -perusversiot eivät ole liian kalliita verrattuna markkinoiden muihin kaupallisiin ratkaisuihin, mutta niistä puuttuu ominaisuuksia, jos haluaa suorittaa monipuolisempaa analytiikkaa.

### **5.3 KNIME**

KNIME eli Konstanz information Miner on avoimen lähdekoodin analytiikkaohjelmisto. Ohjelmisto sisältää graafisen käyttöliittymän, jossa voidaan työkalujen avulla rakentaa helposti analyysejä. KNIME:n kehittäminen alkoi 2004 vuoden alussa, kun Konstanz yliopiston ohjelmoijatiimi aloitti sen kehittämisen. Vuonna 2006 julkaistiin ensimmäinen versio KNIME:sta. (Bulusu 2012, 84-85.) Ohjelmisto toimii Microsoft Windows-, Linux- ja Mac OS X- käyttöjärjestelmillä.

#### **5.3.1 Ominaisuudet ja toiminnot**

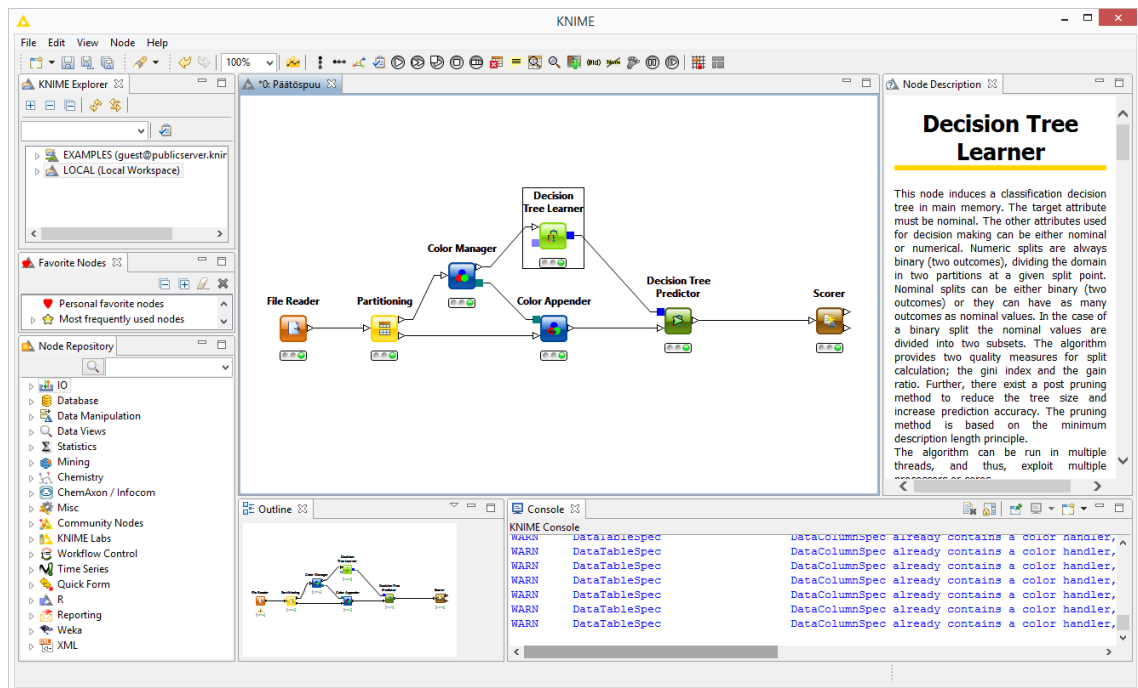
KNIME sisältää graafisen käyttöliittymän ja se perustuu solmujen käyttöön, joilla jokaisella on oma toimintonsa. Solmuja on käytettävissä heti yli 1000. Solmut ovat ohjelmiston työkaluja, joilla suoritetaan kaikki tarvittavat aineiston muokkaamiset ja analyysit. Solmut löytyvät käyttöliittymän päänäkymästä ja ne on jaoteltu omiin kategorioihinsa. Solmujen käyttö toimii raahaa ja pudota -tekniikalla. Halutut toiminnot vain raahataan työtilaan ja niistä muodostetaan työkulkuja. Solmun käsittelyn jälkeen se pitää vielä ajaa. Kun työtilaan raahataan uusi solmu, tulee se myös liittää edelliseen vetämällä vanhasta solmusta nuoli uuteen solmuun. Vaikka ohjelmisto on kirjoitettu Javalla, se sisältää myös solmuja muilla kielillä, kuten Perl ja Python. KNIME:lla on mahdollista kirjoittaa myös R-skriptejä.

Tiedostonlukusolmujen avulla pystytään lisäämään haluttu aineisto. Tiedostojen ja tietokantojen lukusolmuja ohjelmistosta löytyy yli 30 ja niihin kuuluvat muun muassa CSV-, XLS- ja MySQL-solmut. KNIME tukeekin useaa eri tiedosto ja tietokantamuotoa. Kun haluttu aineisto on lisätty työtilaan, pystytään sitä vielä muokkaamaan ennen analyysin

suorittamista. Tiedon manipulaatiolle on varattu omat kansionsa, joista aukeaa kymmennittäin solmuja aineiston valmisteluun. Solmuja löytyy niin suodattamiseen, yhdistämiseen kuin muuntamiseenkin.

Ohjelmistolla onnistuu monimutkaisempikin tilastointi ja tiedonlouhinta, jolla saadaan selville trendejä ja ennustetaan potentiaalisia tuloksia. Solmuista löytyvät kaikki tarvittavat toiminnot niin kuvailevaan kuin ennustavaankin analytiikkaan. Ohjelmiston kansiot pitävät sisällään solmuja t-testien, korrelaatioiden, eri regressioiden, ristiintaulukoiden, eri arvojen ja monen muun toteuttamiseen. Ohjelmistoon pystyy lataamaan myös laajennuksia, jotka mahdollistavat esimerkiksi tekstinlouhinnan, kuvalouhinnan ja sosiaalisen median analyysit. Laajennukset tekevät KNIME:sta todella tehokkaan ja monipuolisen analytiikkaohjelmiston. Avoimen lähdekoodin ohjelmistona KNIME:en on mahdollista kehittää myös omia solmuja ja muokata ohjelmistoa omiin tarpeisiin. Ohjelmisto pystyy lisäksi käsittelemään jopa miljoonia rivejä aineistoa juurikaan jumiutumatta.

Tiedon visualisointiin löytyy yli 20 solmua, joihin lukeutuvat muun muassa ympyräkuvaajat, pylväskuvaajat, viivakuvaajat ja interaktiiviset taulukot. Luodut kuvaajat voi tallentaa suoraan niiden omista näkymistä kuvatiedostoina. Muuten analyysien tallentaminen tapahtuu ulostulosolmujen, kuten CSV- ja XLS-kirjoita -solmujen, avulla. Tulokset voidaan ottaa myös talteen KNIME:n raportointityökalun avulla. Työkalu tarjoaa työtilan, johon pääsee lisäämään tekstikenttiä, kuvia, taulukoita ja erilaisia kuvaajia. Sillä saa tehtyä varsin näyttäviä raportteja, jotka voidaan tallentaa muun muassa PDF-, HTML- tai XLS-muodoissa. Raportointityökalulla on mahdollista tehdä myös edistyneempiä kuvaajia.



KUVA 4. KNIME näkymä

### 5.3.2 Käytettävyys

Yksinkertainen visuaalinen käyttöliittymä mahdollistaa ohjelmiston käyttämisen kaiken tasoilla käyttäjillä. Solmujen lisääminen näkymään on selkeää ja helposti opittavissa. Solmujen määrä on valtaisa, joten niistä oikean löytäminen voi viedä aluksi reilusti aikaa. Toisaalta jokainen solmu on jaettu järkevästi omiin kategorioihinsa ja ne sisältävät omat ikonit, jotka helpottavat oikean solmun löytämistä. Solmut voidaan kätevästi myös ajaa tai ottaa pois käytöstä vain muuttamalla niiden asetuksia. Jokainen solmu sisältää myös oman selitteen sen ominaisuuksista. Solmut eivät ole myöskään turhan isoja, joten monimatkaisemmatkaan analyysit eivät täytä heti tilaa. Työnkulun ymmärrettävyyttä auttaa myös hahmottelunäkymä, joka näyttää koko työnkulun pienoiskoossa. Analyysien ymmärtämistä helpottaa visuaalinen työnkulku, koska KNIME esittää kaikki solmut ja niiden suhteet toisiinsa visuaalisesti. KNIME:ssa on selvästi keskitytty helppokäyttöisyyteen.

KNIME on tehokas ohjelmisto analyysien suorittamiseen, mutta aineiston perusmuokaus on suhteellisen työlästä. Raaha ja pudota -toiminnoilla yksinertaisten muutoksien suorittaminen voi viedä runsaasti aikaa. Virheitä tehtäessä työnkulkua ei pysty ajamaan

ja lokissa ilmoitetaan ongelmat. Ongelma tilanteissa käyttäjä voi myös tutkia useita esimerkkejä KNIME serveriltä, johon pääsee käsiksi ohjelmistosta. Ohjelmisto ei johda muutenkaan käyttäjää harhaan. Graafista käyttöliittymää on miellyttävä käyttää ja sen oppimiseen ei mene suurta aikaa pidemmänkään käyttötaun jälkeen.

### 5.3.3 Tukipalvelut

KNIME-sivusto ylläpitää foorumeita, joilla käyttäjät voivat esittää kysymyksiä ja jakaa mielipiteitään ohjelmistosta. Foorumeille ilmestyy viestejä lähes päivittäin ja niihin vastataan suhteellisen nopeasti. Sivustolta löytyy myös sopivasti opetusmateriaalia ohjelmiston käytöstä. Suurinosa näistä on tekstimuodossa. Ammattitukea haluavat joutuvat hankkimaan ensin kaupallisen version ohjelmistosta.

### 5.3.4 Hinnoittelu

KNIME on avoimen lähdekoodin ohjelmisto ja se on ilmainen, mutta siihen on myös tarjolla kaupallisia ratkaisuja. Alla olevassa taulukossa (Taulukko 4) on käyty lävitse KNIME ratkaisut.

TAULUKKO 4. KNIME hinnoittelu (Knime 2015, muokattu)

Ohjelmisto	Ominaisuudet	Hinta
KNIME Analytics Platform	Kaikki analytiikkaominaisuudet	Ilmainen
KNIME Personal Productivity	Kaikki analytiikkaominaisuudet, metasolmuja, mukautetut solmut	300€, per käyttäjä, 1 vuosi
KNIME Partner Productivity	Kaikki analytiikkaominaisuudet, enemmän metasolmuja, mukautetut solmut	5000€, per yhteistyökumppani, 1 vuosi

### 5.3.5 Yhteenveto

KNIME kuuluu tehokkaimpiin avoimen lähdekoodin analytiikkaohjelmistoihin. Se tuo laajan tarjonnan työkaluja tiedonlouhintaan yli tuhannen solmun avulla. Ohjelmistoa pystyy muokkaamaan omiin tarpeisiin lataamalla lisää solmuja tai luomalla omia. Siitä saa tehtyä todella tehokkaan ja monipuolisen lisäämällä lisäosia siihen. Analytiikan lisäksi ohjelmisto tarjoaa hyvän työkalun raportointiin. Graafinen käyttöliittymä on helposti opittavissa ja käytettävissä. Useiden solmujen tunnistaminen ja käyttäminen vaatii aluksi vähän totuttelua, mutta niiden tultua tutuiksi ohjelmiston käytöstä tulee todella tehokasta. Aineiston muokkaaminen ennen analyysiä on hieman työläämpää.

KNIME on täysin ilmainen ohjelmisto, mutta siitä voi ostaa myös kaupallisen version, jos haluaa päästä käsiksi lisäominaisuuksiin. Ilmainen versio antaa kuitenkin kaikki tarvittavat työkalut analytiikan harjoittamiseen. Ostamalla kaupallisen version tukee lähinnä KNIME:n jatkokehitystä. Tukipalvelut KNIME:ssa keskittyvät lähinnä yhteisötukeen ja verkkosivuilla oleviin opetusmateriaaleihin.

## 5.4 PSPP

PSPP on tilastolliseen analyysiin suunniteltu avoimen lähdekoodin ohjelmisto. Ohjelmisto alkoi GNU-projektina ja ensimmäinen versio julkaistiin elokuussa vuonna 1998. Ohjelmisto rakennettiin ilmaiseksi vaihtoehdoksi kaupalliselle IBM SPSS Statistics:lle. (Kovačić 2009.)

PSPP sisältää graafisen käyttöliittymän ja muistuttaa hyvin paljon IBM SPSS Statistics -ohjelmistoa, mutta se ei silti ole IBM SPSS Statistics-klooni. PSPP on rakennettu täysin erillisenä ohjelmistona, mutta sen suunnittelussa on ollut tavoitteena helpottaa ohjelmiston käyttöä, jos IBM SPSS Statistics oli jo aikaisestaan tuttu. (Gnu 2015.) PSPP on täysin ilmainen ja tarjoaa runsaasti toimintoja analytiikan suorittamiseen. PSPP:tä on mahdollista käyttää graafisella käyttöliittymällä tai komentorivillä. Tässä selvityksessä tullaan keskittymään graafisen käyttöliittymän tutkimiseen. PSPP toimii Microsoft Windows-, Mac OS X-, Linux- ja muutamalla muulla käyttöjärjestelmällä.

### 5.4.1 Ominaisuudet ja toiminnot



Ohjelmiston graafinen käyttöliittymä kulkee nimellä PSPPIRE, ja muistuttaa suurilta osin IBM SPSS Statistics:n käyttöliittymää. Käyttöliittymä ei kuitenkaan sisällä läheskään kaikkia IBM SPSS Statistics:n valikoita ja toimintoja. Graafinen käyttöliittymä on myös rajallisempi verrattuna komentorivin mahdollistamiin toimintoihin.

Aineisto ladataan ohjelmistoon tiedostovalikosta valitsemalla syötä tietoa. Ohjelmisto tukee yleisimpiä tiedosto- ja tietokantamuotoja. Käsiteltävän aineiston PSPP tallentaa savmuodossa aivan kuten IBM SPSS Statistics, joten niiden välillä voidaan helposti vaihtaa käsiteltävää aineistoa. Ladattu aineisto ilmestyy ohjelmiston kahteen taulukkoon; tietosivuun ja muuttujasivuun. Aineistoa on mahdollista muokata vielä näissä taulukoissa muun muassa poistamalla rivejä ja nimeämällä kenttiä uudestaan. Muunna- ja tietovetovalikoista aukeaa vielä lisää toimintoja aineiston muokkaamiseen, kuten aineiston leikkaaminen ja tietyn tiedon valinta. Toiminnot on rajoitettu ainoastaan perusaineiston muokkaamiseen.

PSPP:stä löytyy perusanalytiikan työkalut ja ne ovat lähes identtiset IBM SPSS Statistics:n valikkoihin verrattuna. Kuvailevan tilastoinnin vaihtoehdot ovat ohjelmistossa suhteellisen kattavat. Ohjelmistossa on mahdollista suorittaa frekvenssejä, ristiintaulukointia ja eri arvojen laskemista. Ennustavan analytiikan toiminnoista löytyy muun muassa perinteiset regressiot, mutta muuten ohjelmistosta puuttuu paljon oleellisia toimintoja, kuten päätöspuut. Toimintojen määrä on hyvin vaatimaton.

Case	Customer	Store_Nu	Customer_Segm	Res	First_Name	Last_Name	Address	City	State	Zip
1	2	100	Corporate	No	JEAN	SMITH	376 S JASMINE ST	DENVER	CO	80224
2	3	100	Corporate	No	JULIA	CARRERA	846 S UMATILLA WAY	DENVER	CO	80223
3	5	100	Home Office	No	LINDA	TREVINO	5360 ZUNI ST	DENVER	CO	80221
4	6	106	Home Office	No	H	MACK	1599 WILLIAMS ST	DENVER	CO	80218
5	7	105	Home Office	No	MARISSA	LATTA	12066 E LAKE CIR	ENGLEWOOD	CO	80111
6	8	101	Home Office	No	PHYLLIS	WALKER	7225 S GAYLORD ST	LITTLETON	CO	80122
7	9	105	Home Office	No	VIVIAN	GAULDEN	4497 CORNISH WAY	DENVER	CO	80239
8	10	100	Home Office	No	PAMELA	WRIGHT	2316 E 5TH AVE	DENVER	CO	80206
9	11	106	Home Office	No	MARIA	GONZALEZ	3883 QUITMAN ST	DENVER	CO	80212
10	12	108	Home Office	No	WANDA	MAYBERRY	1965 YUKON ST	DENVER	CO	80214
11	14	105	Small Business	No	KATHLEEN	KIRBACH	2343 S VAUGHN WAY	AURORA	CO	80014
12	15	104	Small Business	Yes	BERNADETTI	MCKALE	4288 S RICHFIELD ST	AURORA	CO	80013
13	16	103	Small Business	No	JEANNINE	QUINLAN	5400 SHERIDAN BLVD	ARVADA	CO	80002
14	17	107	Small Business	No	CYNTHIA	LOUY	195 JADE ST	BROOMFIELD	CO	80020
15	18	104	Small Business	No	MICHAELA	BOLLWITT	4034 S CARSON ST	AURORA	CO	80014
16	19	102	Small Business	No	JENNIFER	CAMPBELL	22873 E CLIFTON PL	AURORA	CO	80016
17	20	100	Small Business	No	KRISTA	FLOREZ	2879 S MEMPHIS ST	AURORA	CO	80013
18	21	100	Small Business	Yes	AVIVA	HEIFETS	3490 S BELLAIRE ST	DENVER	CO	80222

KUVA 5. PSPP näkymä

Tiedon visualisoinnissa on myös rajoituksia. PSPP ei sisällä IBM SPSS Statistics:stä tuttua kuvaaja-valikkoa. Eriaiset kuvaajat piirretäänkin syntaksi-ikkunaan kirjoitettavilla komennoilla, kuten /piechart ja /histogram. Kuvaajia on mahdollista toteuttaa myös joistakin kuvaavien analyysien asetuksista. Frekvenssivalikosta on esimerkiksi mahdollista tehdä ympyräkuvaajia ja histogrammeja. Luotuja kuvaajia ei pysty muokkaamaan jälkikäteen. Niiden visuaalinen ulkoasu on myös hyvin yksinkertainen. Jos tulokset haluaa visualisoida paremman näköisiksi, tulisi valita toinen ohjelmisto visualisointia varten. Ohjelmisto ei myöskään tarjoa kummempia raportointimahdollisuuksia. Tulokset on mahdollista tallentaa ulostulonäkymästä esimerkiksi PDF-, CSV- ja HTML-muodoissa.

PSPP jää toiminnoiltaan kauas IBM SPSS Statistics:stä, mutta perusanalytiikan suorittamiseen se sopii erittäin hyvin. Haasteeksi ohjelmiston käytössä nousee toimintojen puutos, sekä se, että monet toiminnot ovat käytössä vain syntaksi-ikkunan takaa. Toisaalta ohjelmisto ei vaadi mitään lisäosia, vaan kaikki on ohjelmistossa valmiina. Koska kyseessä on avoimen lähdekoodin ohjelmisto, toimintoja on myös mahdollista lisätä vielä itse lataamalla vaikka toisten lähdekoodeja ja lisäämälle ne ohjelmistoon. Ohjelmisto pysyy myös käsittelemään helposti isompaakin aineistoa.

## 5.4.2 Käytettävyys

Ohjelmisto on kaikin puolin yhdenmukainen ja sen käyttö on suhteellisen helposti opittavissa. Ohjelmiston käyttöä helpottaa huomattavasti, jos käyttäjällä on aikaisempaa kokemusta muista taulukko-ohjelmista, kuten IBM SPSS Statistics:stä ja Excel:istä. Toiminnot on piilotettu vetovalikoiden taakse, joilla on niiden toimintoja kuvaavat nimet. Vetovalikoiden käyttöä hieman hankaloittaa merkkien ja ikonien puuttuminen. Muutamasta vetovalikosta aukeaa vielä lisää valikoita, mutta näitä ei mitenkään merkitä valikoihin. Vetovalikoiden toimintojen tunnistavuutta helpottaisi visuaalisuus esimerkiksi ikonien avulla. Toiminnot eivät myöskään sisällä minkäänlaisia selostuksia niiden käyttämisestä.

Ohjelmistolla pystyy tekemään perusanalyyskejä jopa todella nopeasti, mutta työskentelyä hidastaa joidenkin toimintojen puuttuminen. Yksi kiusallinen puuttuva ominaisuus on muun muassa kumota ominaisuuden puuttuminen. Ohjelmiston taulukkonäkymät ja toiminnot ovat helposti opittavissa pidemmänkin käyttötaun jälkeen. Ohjelmisto ei kumminkaan tarjoa paljon oikoreittejä, joten käyttäjän tulee tietää mitä tekee.

Ohjelmisto on ulkoasultaan yhdenmukainen, mutta visuaalisesti se saisi olla monipuolisempi. Ohjelmiston toiminnot ja taulukot ovat kaikki sekoitus harmaita sävyjä ja erottavat ominaisuudet puuttuvat. Ohjelmisto kaipaisikin uudistettua, selkeämpää ja nätimpää ulkoasua, jotka edistisivät miellyttävämpää käyttöä.

## 5.4.3 Tukipalvelut

Tukipalvelut ohjelmistolla ovat hyvin vaatimattomat ja ne perustuvat lähinnä dokumentteihin. Ohjelmiston sivuilta löytyy kattavat määrät opetusmateriaaleja ohjelmiston toiminnoista ja käytöstä. Yhteisötukea on mahdollista saada yhteisellä foorumi alueella. Foorumeille saapuu päivittäin viestejä, mutta vastausta voi joutua odottamaan pidempiäkin aikoja.

#### 5.4.4 Hinnoittelu

PSPP on täysin ilmainen ohjelmisto, eikä se sisällä mitään maksullista ominaisuutta. Ohjelmistoon ei ole edes tarjolla mitään maksullista lisäosaa.

#### 5.4.5 Yhteenveto

PSPP on hyvä ohjelmisto, vaikka se jääkin selvästi IBM SPSS Statistics:n taakse. Ohjelmisto tarjoaa kaikki tarvittavat toiminnot perus analytiikan toteuttamiseen. Aineiston visualisointikeinot ohjelmistossa ovat erittäin rajalliset, eikä sillä onnistu kunnan raporttien luominen.

Ohjelmiston käyttöä helpottaa, jos on aikaisemmin käyttänyt muita taulukkolaskentaohjelmia. Varsinkin IBM SPSS Statistics-käyttäjät tuntevat olonsa kotoisaksi PSPP:n kanssa. Silti ohjelmiston käytettävyydessä on ongelmia sen puuttuvien toimintojen ja rajallisen visuaalisuuden vuoksi. Täysin ilmaisena ohjelmistona tukipalvelut eivät ole ihmeelliset, ja ne rajoittuvat lähinnä erilaisiin dokumentteihin.

### 5.5 RapidMiner Studio

RapidMiner Studio kuuluu myös tunnetuimpiin avoimen lähdekoodin analytiikkaohjelmistoihin. Ohjelmisto keskittyy erityisesti tiedonlouhintaan, tekstinlouhintaan ja ennustavaan analytiikkaan. Ohjelmiston kehittäminen aloitettiin vuonna 2001 Dortmundin teknillisen yliopiston tekoälyosastolla. Ingo Mierswa, Simon Fischer ja Ralf Klinkenberg alkoivat kehittää joustavaa ja tehokasta tiedonlouhintaohjelmistoa, jota he kutsuivat YALE:ksi (Yet Another Learning Environment). Vuonna 2006 ohjelmiston kysyntä alkoi olla tarpeeksi suurta, joten kehittäjät päättivät perustaa yrityksen ohjelmiston tueksi. Tällöin ohjelmisto koottiin ja julkaistiin uudestaan, mutta nyt nimeksi otettiin RapidMiner Studio. (Rapidminer 2015a.) RapidMiner Studio on ladattavissa Microsoft Windows-, Mac OS X- ja Linux- käyttöjärjestelmille.

### 5.5.1 Ominaisuudet ja toiminnot

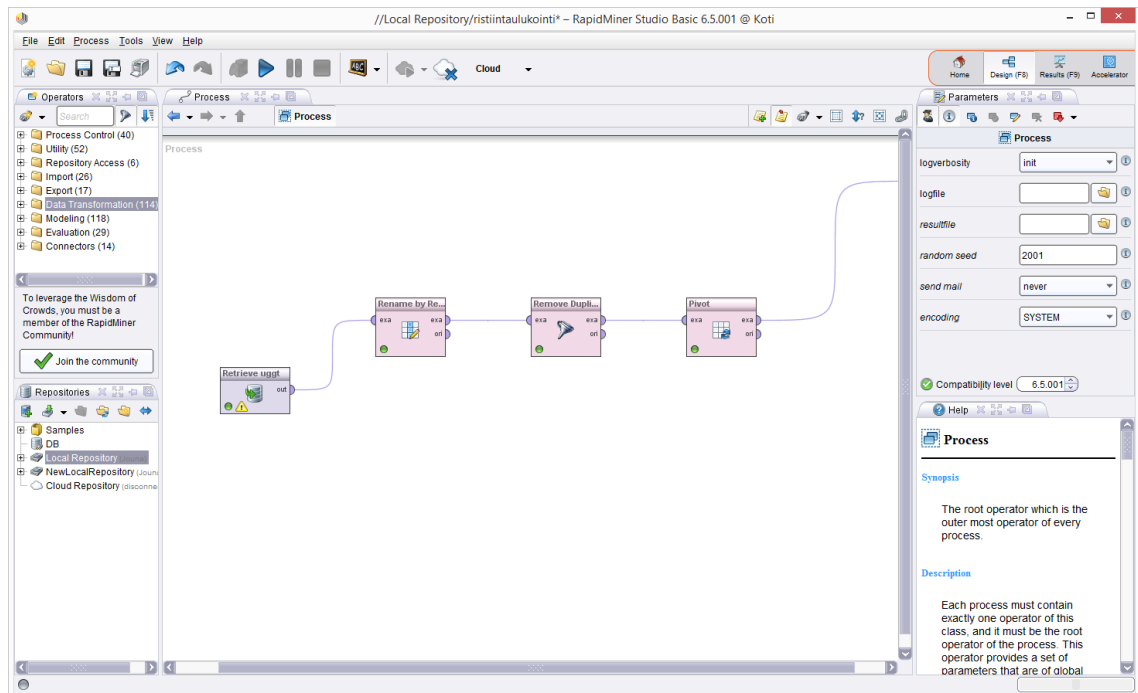
Ohjelmiston käyttö perustuu graafiseen käyttöliittymään ja se toimii raahaa ja pudota - tekniikalla. Ohjelmisto koostuu perspektiiveiksi kutsutuista näkymistä. Perspektiiveihin kuuluvat koti, suunnittelu, tulokset ja kiihdytin. Koti-perspektiivistä pääsee avaamaan aikaisempia analyysyjä, aloittamaan uusia prosesseja ja selaamaan ohjeita. Uuden analyysin pääsee aloittamaan myös sille varatussa suunnittelu- perspektiivissä. Suunnittelu-perspektiivissä itse työtilaa kutsutaan prosessiksi ja se sisältää useita operaattoreita. Jokaisella operaattorilla on oma tehtävänsä ja niitä lisätään osaksi prosessia. Operaattoreita on käytössä yli 1500. Tulokset-perspektiivi näyttää nimensä mukaisesti analyysin tulokset ja kiihdytin-perspektiivissä käyttäjä voi ratkaista tyypillisiä tiedonlouhintaongelmia nopeasti.

Aineiston tuominen analyysiä varten tapahtuu suunnittelu-perspektiivissä. Täältä löytyy oma tietolähdenäkymä, josta avataan halutut aineistot. RapidMiner Studio asennuksen yhteydessä luodaan aina tietolähde, mutta sellaisen pystyy luomaan myös jälkikäteen. Tietolähteelle nimetään jokin kansio, jossa aineistot sijaitsevat. Luodusta tietolähteestä voidaan seuraavaksi hakea haluttu aineisto analyysiä varten. Ohjelmisto pystyy lukemaan lukuisia eri tiedostoformaatteja ja kaikkia yleisimpiä tietokantoja. Kun aineisto on ladattu ohjelmistoon, tulee se tämän jälkeen raahata prosessinäkymään. Valittua aineistoa voidaan vielä muokata eri operaattoreilla, kuten suodatinoperaattorilla. Haluttu operaattori vain raahataan prosessitilaan, jonka jälkeen se pitää tämän vielä yhdistää aineistoon, jollei linkittämistä ole automatisoitu. Aineiston valmisteluun löytyy kaikki tarvittavat operaattorit. Kun halutut operaattorit on lisätty, tulee niistä viimeisin yhdistää tulospäätöseen. Tämän jälkeen koko prosessi ajetaan työkalurivin aja-ikonista. Ohjelmisto tuo nyt tulosperspektiiviin tulokset.

RapidMiner Studio:ta mainostetaan johtavana avoimen lähdekoodin ennustavan analytiikan ohjelmistona. Ohjelmiston ennustaviin malleihin kuuluukin laaja kirjo operaattoreita, kuten korrelaatiot, lineaariset regressiot, loogiset regressiot, markkinakorianalyysit, selviytymisanalyysit, aikasarjojen tiedonlouhinta ja päätöspuut. Jokainen operaattori toimii moitteettomasti ja käsittelee helposti isoakin aineistoa. Kiihdytin-perspektiivissä pääsee käsiksi myös tehokkaihin ennustavan analytiikan toimintoihin. Näihin toimintoihin luokituvat asiakasdatasta ennustaminen, suora markkinointi, ennustava ylläpito ja ajatus-

analyysit. Jokainen toiminto on vielä nopeasti toteutettavissa kolmen yksinkertaisen vaiheen kautta. Toiminnoilla saa nopeasti tehokkaita ennustavia analyysijä. Ennustavaa analytiikkaa on mahdollista toteuttaa myös RapidMinerin Radoop-lisäosalla, joka mahdollistaa Hadoop Big Datan käytön analyysissä. Ohjelmistolla onnistuu myös erilaisten tekstianalyysien teko trendianalyysillä ja luokittelulla.

Vaikka ohjelmisto keskittyy ennustavaan analytiikkaa, tarjoaa se myös hyvät edellytykset kuvailevan analytiikan ja tilastoinnin toteuttamiselle. Ohjelmistosta löytyy operaattorit eri arvojen laskemiseen ja yhteyksien löytämiseen. Ohjelmiston tulostuloksissa pääsee seuraamaan automaattisesti luotua tilastointia, kuten keski-, suuria- ja pieniarvoja. Ohjelmiston tarjoamalla perustoiminnoilla pärjää pitkälle ja niitä on mahdollista laajentaa vielä WEKA ja R-kirjastoilla.



KUVA 6. RapidMiner Studio näkymä

Analyysijä on mahdollista visualisoida tulos-perspektiivissä, josta löytyvät omat välilehdet tulosten näyttämiseen, yleisen tilastoinnin näyttämiseen ja kuvaajien luontiin. Kaikki oleelliset toiminnot tiedon visualisointiin on mahdutettu hyvin samaan työtilaan. Erilaisien kuvaajien luontiin löytyy monipuolisesti erilaisia pylväs- ja ympyräkuvaajia ja paljon muuta. Visualisointia on mahdollista tehdä automatisoidulla tai edistyneellä editorilla. Raportointiin ohjelmisto vaatii kuitenkin erillisen raportointilisäosan tai RapidMiner Ser-

verin käyttöön ottamisen. Ilmaisella raportointilisäosalla saa luotua raportteja rakentamalla operaattori kerrallaan. Raporttien kokoamisessa on silti rajoituksia muun muassa raportin kenttien sijoittelussa.

### 5.5.2 Käytettävyys

Ohjelmiston graafisessa käyttöliittymässä analyttisten prosessien muodostaminen on helppoa ja intuitiivista myös kokemattomille käyttäjille. RapidMiner Studio:lla onnistuvat tekemään tiedonlouhintaa muutkin kuin tilastotieteilijät. Ohjelmisto tarjoaa käyttäjälle myös graafisia askel askeleelta -ohjeita.

Eri operaattoreiden tunnistaminen ja löytäminen voi kuitenkin olla aikaa vievää. Operaattoreita on yli 1500 ja niistä jokainen on piilotettu monen kansion taakse. Operaattoreiden hakemista helpottaisi ikonien käyttö kansioden sijasta. Onneksi operaattoreita on mahdollista myös hakea pikahaussa. Jos käyttäjä tietää nimeltä, mitä operaattoria tarvitsee, sen löytämisessä ei mene kauaa. Operaattoreiden hyväksi puoliksi voidaan laskea niistä avautuvat tietolaatikat, jotka kertovat käyttäjälle operaattorin ominaisuudet. Operaattoreiden muokkaamisnäkyvä aukeaa yleensä myös samaan tilaan, joka nopeuttaa työskentelyä. Muilla ohjelmiston toiminnoilla on yleensä tekstin lisäksi omat ikoninsa, jotka helpottavat niiden ymmärtämistä. Ohjelmiston pariin onkin helppo palata käyttökatkojen jälkeen. Ongelmaksi voi lähinnä muodostua oikeiden operaattoreiden löytäminen. Käytön opittua ohjelmisto tarjoaa tehokkaan alustan analytiikan toteuttamiseen.

Ohjelmiston automatisoidut toiminnot helpottavat ja nopeuttavat työskentelyä. Operaattoreiden välisiä linkityksiä voi automatisoida ja ohjelmisto ohjaa käyttäjää haluttuihin näkymiin. Analyysien suorittamisessa analyysitilan ja tulostilan vaihtaminen onnistuu helposti näkymän oikeasta ylälaidasta. Tilojen edestakainen vaivaton vaihtaminen helpottaa analyysin ymmärtämistä.

Aineiston perusmuokkaaminen on toisaalta ohjelmistossa työlästä ja hidasta ja se voikin olla nopeampaa toisella ohjelmistolla. Ohjelmistoa vaivaavat välillä myös kiusalliset virheet. Tehty prosessikaavio voi näyttää oikein rakennetulta ja loogiselta eikä virheilmoituksia ilmene, mutta silti ohjelmisto ei pysty suorittamaan analyysiä. Tämän kaltaisissa

tapauksissa syyn löytämisessä voi kestää pitkäänkin. Työtila myös täyttyy helposti operaattoreista ja voi pahimmassa tapauksessa muuttua erittäin sekavaksi. Yleisesti ohjelmiston käyttäminen onkin miellyttävää, jollei luoda monimutkaisempia analyysejä.

### 5.5.3 Tukipalvelut

Ohjelmiston tukipalvelut ovat lueteltu ohjelmiston kotisivuilla. Näihin tukipalveluihin kuuluvat yhteisötuki, foorumit, dokumentit ja yritystuki. Yhteisötuen avulla voit selata toisten käyttäjien esittämiä kysymyksiä ja vastauksia sekä esittää omia kysymyksiä. Yhteisötuen käyttöönotto edellyttää vain ilmaista rekisteröitymistä sivustolle. Ohjelmiston sivuilta löytyvät myös aktiiviset foorumit, joissa kuka tahansa voi käydä selaamassa aiempia keskusteluja tai aloittaa omansa. Foorumeilla on jo yli 6000 käyttäjää ja sinne ilmestyy päivittäin uusia viestejä. Foorumin käyttäminen voikin olla nopein tapa saada vastaus kysymyksiin.

Ohjelmiston sivustoilta löytyy myös kattava kirjo erilaisia opetusmateriaaleja, jotka sisältävät ohjeita ohjelmiston asentamiseen ja käyttämiseen. Opetusmateriaalien määrä on yllättävän laaja, niissä on selitetty kaikki ohjelmiston päätoiminnot ja ohjelmiston käyttö askel askeleelta. Perinteisten tekstidokumenttien lisäksi löytyy myös videoita.

Täyden tuotetuen saaminen edellyttää käyttäjältä kaupallisen version ostamista. Esimerkiksi yritystuen saa vasta tässä tapauksessa. Yritystuki mahdollistaa yksityistuen saamisen ohjelmiston asiantuntijoilta.

### 5.5.4 Hinnoittelu

RapidMiner Studio:sta on kolmea ratkaisua, joista yksi on maksullinen. Nämä vaihtoehdot on käyty läpi alla olevassa taulukossa (Taulukko 5). Ohjelmistosta ja muista palveluista on myös tarjolla erikseen sovittavia ratkaisuja, joiden hintoja tulee tiedustella suoraan ohjelmiston tarjoajalta. Ohjelmisto on erittäin kohtuuhintainen verrattuna muihin kaupallisiin ratkaisuihin. Pelkästään ilmaisista paketeistakin löytyy paljon tehokkaita ominaisuuksia, joita myös kaupalliset ratkaisut tarjoavat. Ohjelmisto loistaakin sen edullisella hinta-laatusuhteella.



TAULUKKO 5. RapidMiner Studio ratkaisut. Hinnat muunnettu dollareista euroiksi (RapidMiner 2015b, muokattu)

Ohjelmisto	Ominaisuudet	Vuosittainen hinta
RapidMiner Studio Basic	Perusanalytiikkaominaisuudet	Ilmainen
RapidMiner Studio Community	Perus analytiikkaominaisuudet, lisää analytiikkaominaisuuksia, yhteisöominaisuudet ja edut	Ilmainen, mutta vaatii kirjautumisen yhteisöön
RapidMiner Studio Professional	Perusanalytiikkaominaisuudet, lisää analytiikkaominaisuuksia, yhteisöominaisuudet ja edut, kaupalliset ominaisuudet ja edut	Alkaen 1786,34€, per käyttäjä, 1 vuosi

### 5.5.5 Yhteenveto

RapidMiner Studio on vahvimmillaan perus- ja keskiraskaassa analytiikassa, koska se tarjoaa edullisesti laajan valikoiman valmiita työkaluja analytiikan työstämiseen. Ohjelmiston graafisen käyttöliittymän käyttö on lisäksi sujuvaa ja intuitiivista. Toimintojen automatisointi kuuluu myös ohjelmiston maininnan arvoisiin ominaisuuksiin. Analytiikassa ohjelmisto kuuluu ennustavan analytiikan parhaimmiston. Ohjelmistoon on sisällytetty useita tehokkaita ennustavaa analytiikkaa mahdollistavia toimintoja. Toisaalta ohjelmistoa vaivaa toisinaan aineiston perusmuokkauksen hitaus ja välillä tapahtuvat ongelmat. Ohjelmisto on kuitenkin suhteellisen helppo oppia ja tehokas käytössä.

RapidMiner Studio tarjoaa runsaasti yhteisöllisiä tukipalveluita ja dokumentteja helpottamaan ohjelmiston käyttöä. Henkilökohtaista tukea tarvitsevat joutuvat ostamaan kaupallisen version. Jos liiketoiminta-analytiikassa ilmenee arkoja ja haastavia analyysiongelmiä, olisi parasta ostaa kaupallinen versio ja saada näin henkilökohtaista tukea.

Ohjelmiston ilmaisversiolla pärjää pitkälle, jos ei tarvitse lisäpalveluita, kuten henkilökohtaista tukea, servereitä ja pilvipalveluita. Ohjelmisto on hinta-laatusuhteeltaan silti edullinen verrattuna muihin kaupallisiin ratkaisuihin.

## 6 JOHTOPÄÄTÖKSET

Tässä luvussa tullaan käymään läpi edellä esiteltyjen ohjelmistoselvityksien tulokset. Samalla verrataan ohjelmistoja keskenään kuvaajien avulla ja selvitetään niiden pätevyyttä erilaisiin tarpeisiin. Ohjelmistojen jokaista selvitettävää ominaisuutta tarkastellaan niille varatuissa alaluvuissa. Jokaisessa alaluvussa on omat pylväskuvaajat, joissa arvioidaan ohjelmistoja asteikolla yhdestä viiteen. Paras tulos on viisi ja yksi on heikoin. Pylväskuvaajat eivät perustu mihinkään erilliseen tutkimukseen, vaan ne ovat tämän työn tekijän toteuttamia.

### 6.1 Ohjelmistojen ominaisuudet ja toiminnot

Erot ohjelmistojen toimintojen välillä olivat pieniä, koska ne olivat keskenään pääosin hyvin samankaltaisia. Poikkeuksena voidaan mainita PSPP- ohjelmisto, jossa oli selvästi vähiten toimintoja muihin verrattuna. Alteryx Designer, KNIME ja RapidMiner Studio toimivat kaikki saman tyyppisillä graafisilla käyttöliittymillä, jotka rakentuivat raahaa ja pudota –toiminnon varaan. IBM SPSS Statistics ja PSPP puolestaan sisälsivät taulukot, joissa aineisto käsiteltiin.

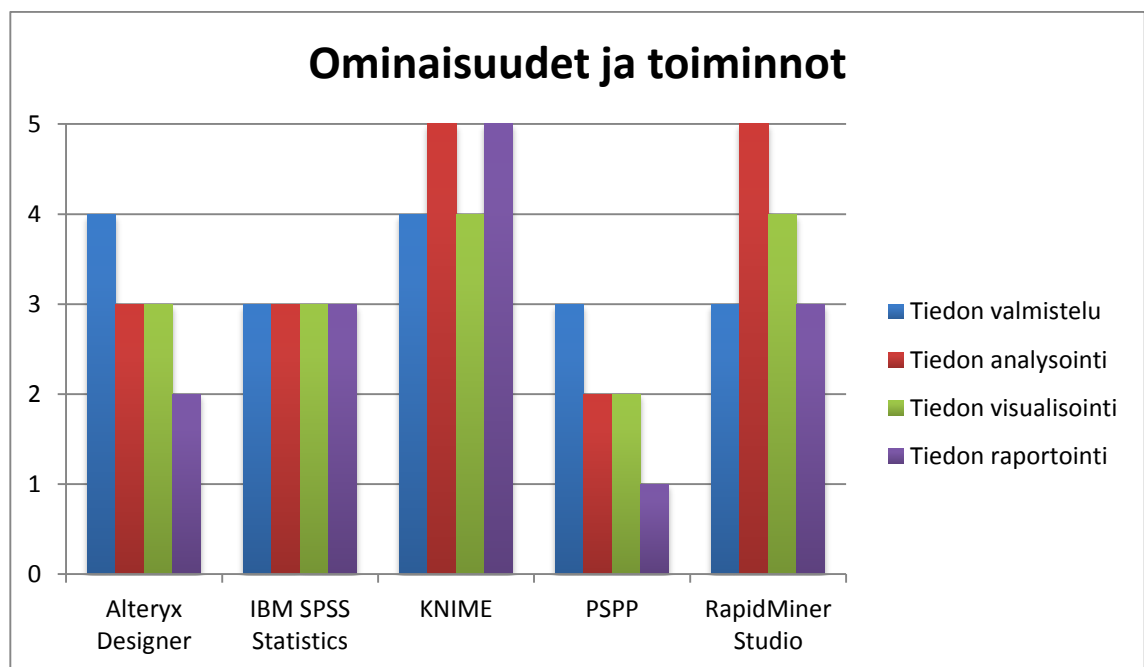
Aineiston valmistelussa Alteryx Designer ja KNIME kuuluivat ohjelmistojen parhaimmiston, sillä ne sisälsivät kattavimmat ja tehokkaimmat toiminnot. Varsinkin Alteryx Designerin tarjoama eri tiedostojen yhdistämistoiminto oli erittäin vaikuttava. KNIME puolestaan sisälsi suuren määrän toimintoja aineiston valmisteluun. Työkaluja tiedon valmisteluun löytyi jokaisesta ohjelmistosta useita, mutta ei yhtä tehokkaita ja monipuolisia.

Tiedon analysointitoimintoja oli myös haastava tutkia, koska kaikista ohjelmista niitä löytyi suuri määrä. Tehokkaimmat ja monipuolisimmat toiminnot löytyi kuitenkin KNIME:sta ja RapidMiner Studio:sta.. KNIME tarjosi analyysien toteuttamiseen satoja työkaluja, ja siihen oli vielä mahdollista ladata ilmaiseksi lisää tehokkaita työkaluja. RapidMiner Studio puolestaan sisälsi myös kattavan työkalulaatikon, josta löytyivät kaikki tarvittavat toiminnot analyysien suorittamiseksi. Alteryx Designer ei sisältänyt sadoittain työkaluja, mutta työkalujen sisälle oli kätkeyty vielä enemmän toimintoja ja ne olivat tehokkaita analyysien suorittamisessa. IBM SPSS Statistics:n edullisin versio sisälsi kaikki

tarvittavat perusanalyysien tekoon. PSPP:llä onnistui myös mainiosti perusanalyysien suorittaminen, vaikka siitä oli karsittu paljon toimintoja.

Tiedon visualisoinnissa parhaimmiston kuuluivat KNIME ja RapidMiner Studio. RapidMiner Studio sisälsi kuvaajien luontiin kaksi omaa editoria, joissa aineistosta sai luotua visuaalisesti hyvännäköisiä kuvaajia. KNIME sisälsi useita työkaluja kuvaajien luontiin, sekä raportointityökalun, jolla pystyi luomaan näyttäviä kuvaajia. Alteryx Designer ja IBM SPSS Statistics sisälsivät työkalut tiedon visualisointiin, mutta toiminnoiltaan ne olivat suppeita. PSPP:n visualisointityökalut olivat todella vaatimattomat. Kuvaajia onnistui lisäämään, mutta visuaalisesti ne olivat yksinkertaisia ja niitä ei pystynyt muokkaamaan. Myös kuvaajatarjonta oli rajallinen.

Raportointiin paras ohjelmisto oli selvästi KNIME, koska se sisälsi oman erillisen raportointityökalun. Työkalulla onnistui näyttävienkin raporttien tekeminen. IBM SPSS Statistics mahdollisti myös raporttien tekemisen ulostulonäkymässä, vaikka erillistä raportointityökalua se ei sisältänyt. Alteryx Designer ja RapidMiner Studio sisälsivät työkaluja raporttien rakentamiseen, mutta ne olivat rajallisia toiminnoiltaan. Näiden ohjelmistojen kanssa tulisikin käyttää erillistä ohjelmistoa aineiston raportointiin. PSPP ei sisältänyt mitään raportointityökalua. Ulostulonäkymästä pystyi tallentamaan analyyskejä, mutta niitä ei juurikaan pystynyt muokkaamaan.

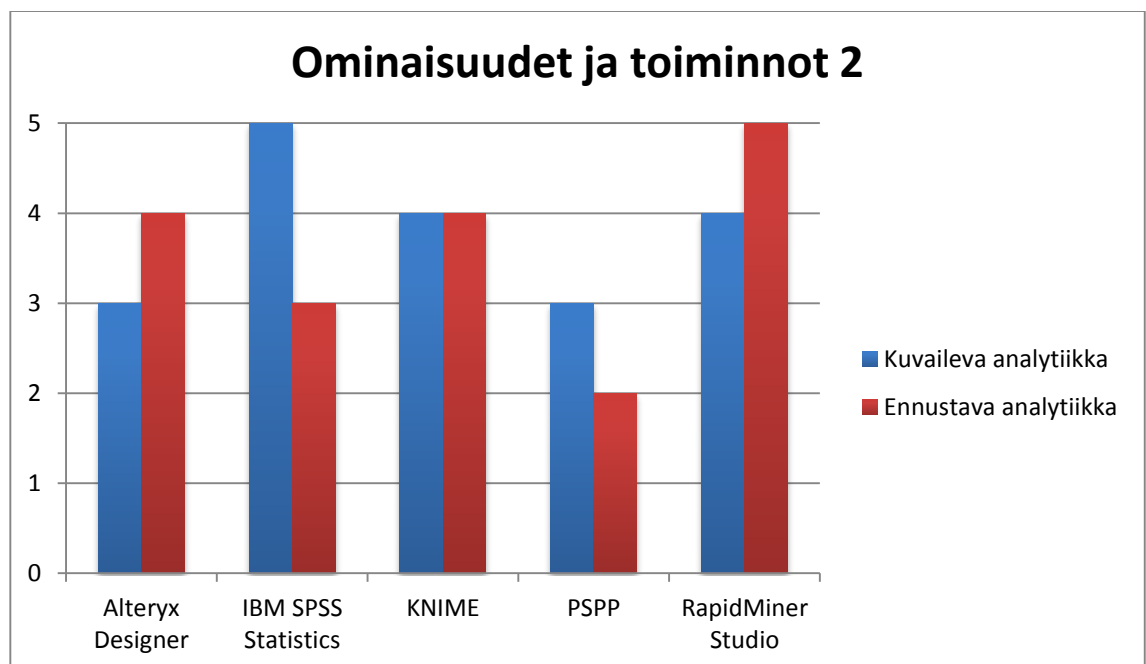


KUVIO 1. Ohjelmistojen ominaisuudet ja toiminnot

Koska analytiikassa itse analyysien toteuttaminen on keskeisessä osassa, tarkastellaan seuraavaksi, kuinka ohjelmistot pärjäävät kuvaavassa ja ennustavassa analytiikassa.

IBM SPSS Statistics selvisi parhaana kuvaavan analytiikan toteuttajana. Ohjelmisto sisälsi omat valikot perinteisten kuvaavien analyysien toteuttamiseen. Ristiintaulukoinnin, frekvenssien ja monen muun tehokkaan kuvaavan tilastointimenetelmän avulla löysi aineistosta helposti yhteyksiä ja muutoksia. Muut ohjelmistot sisälsivät myös kaikki tarvittavat ominaisuudet kuvaavan analytiikan toteuttamiseen, vaikka niissä oli panostettu enemmän liiketoiminnan kannalta oleelliseen ennustavaan analytiikkaan.

RapidMiner Studio selvisi parhaimpana ennustavan analytiikan ohjelmistona monipuolisten toimintojensa ansiosta. Ohjelmisto sisälsi kaikki tarvittavat ominaisuudet päätöspuista eri regressioihin, mutta se sisälsi myös täysin omia toimintoja. Varsin vakuuttava oli ohjelmiston toiminto, jolla pystyi suorittamaan ennustavaa analytiikkaa kolmen yksinkertaisen vaiheen kautta. Kaikista ohjelmistoista heikoin oli jälleen PSPP, vaikka se sisälsi perinteiset kuvaavan analytiikan toiminnot.



KUVIO 2. Ohjelmistojen kuvaavan ja ennustavan analytiikan mahdollisuudet

## 6.2 Ohjelmistojen käytettävyys

Analytiikkaohjelmistojen käytettävyyksistä löytyi paljon samankaltaisuuksia. Minkään ohjelmiston peruskäyttäminen ei muodostunut ongelmaksi, etenkin kun yksikään ohjelmisto ei vaatinut minkäänlaista ohjelmointitaitoa. Kaikilla ohjelmistoilla onnistui liike-toiminta-analyysien suorittaminen ja jokaista ohjelmistoa pystyi käyttämään graafisen käyttöliittymän kautta.

Opittavuudeltaan IBM SPSS Statistics oli selvästi helpoin oppia. Ohjelmisto muistutti suurelta osin muita Windows-ohjelmistoja, joten Windows-ympäristöön totuneet oppivat nopeasti ohjelmiston käytön. Ohjelmisto sisälsi myös helposti opittavat vetovalikot ja taulukot. Alteryx Designer ja KNIME olivat myös helposti opittavissa, koska niissä oli panostettu visuaalisuuteen ja selkeyteen. Muun muassa ohjelmistoissa oli jokainen työkalu nimetty järkevästi ja ne sisälsivät selventävät ikonit. Kummatkin ohjelmistot sisälsivät myös askel askeleelta -ohjeita ohjelmiston peruskäyttämisestä ja muita ohjeita eri toiminnoista. PSPP käyttö oli suhteellisen helposti opittavissa, jos oli aikaisempaa kokemusta muista Windows-ohjelmista tai taulukko-ohjelmista.

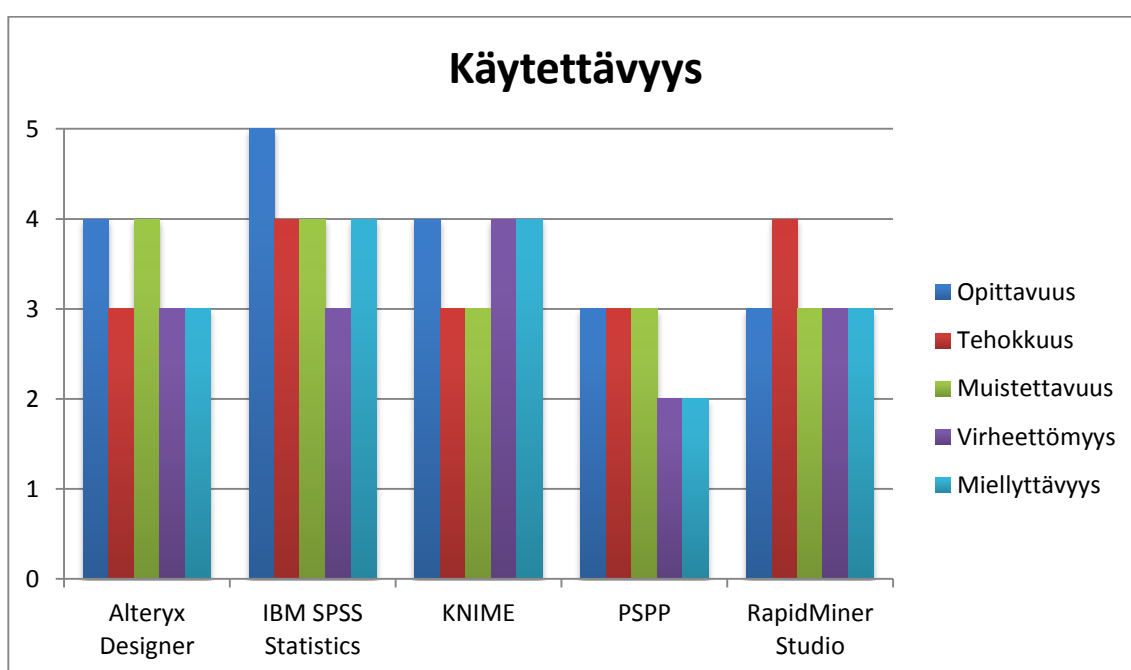
Tehokkaimmat ohjelmistot olivat IBM SPSS Statistics ja RapidMiner Studio. Varsinkin IBM SPSS Statistics:ssa aineiston muokkaaminen ja analyysien tekeminen oli suoraviivaista ja tehokasta. Toiminnot eivät olleet monivaiheisia ja niitä pystyi toteuttamaan nopeasti. RapidMiner Studio:lla puolestaan työskentely tehostui automatisoiduilla toiminnoilla. Alteryx Designer:in ja KNIME:in tehokkuus hiukan kärsi, koska aineiston perusmuokkaaminen oli niillä hitaampaa. Alteryx Designer toimi myös hitaimmin kaikista tutkitavista ohjelmistoista varsinkin jos käsitteli isompia aineistoja. PSPP:ssä perusanalyysien toteuttaminen ohjelmistolla onnistui suhteellisen nopeasti, vaikka toimintoja ohjelmistosta puuttuikin.

Muistettavuudeltaan Alteryx Designer ja IBM SPSS Statistics olivat parhaimmat. Kummatkin ohjelmistot sisälsivät selvät termit ja tunnistettavan visuaalisen ilmeen. Toimintojen määrä oli myös rajattu, joten niiden kaikkien muistaminen ei vienyt kauan aikaa. KNIME ja RapidMiner Studio puolestaan sisälsivät reilu tuhat työkalua, joista oikean löytäminen lyhemmänkin käyttötaun jälkeen oli haastavampaa. PSPP muistettavuutta haittasi lähinnä visuaalisten elementtien puuttuminen.

Virheettömyydeltään KNIME oli hieman parempi muita. Ohjelmisto ei johtanut käyttäjää harhaan ja virheitä tehtäessä ohjelmisto ei suostunut suorittamaan toimintoja loppuun ja

antoi virheilmoituksen, jolla virheen paikantaminen onnistui helposti. Selvästi heikoin virheettömyydeltään oli PSPP, jonka käyttöliittymästä puuttui toimintoja erottavat tekijät. Ohjelmista puuttui myös oleellinen kumota -toiminto.

Ohjelmistoista IBM SPSS Statistics:ia ja KNIMEA oli miellyttävintä käyttää. Kummattakin ohjelmistot sisälsi miellyttävän yhtenäisen graafisen käyttöliittymän ja niillä onnistui analyysien tekeminen. Alteryx Designer sisälsi myös hyvän graafisen käyttöliittymän, mutta ohjelmiston hitaus muodostui sen yhdeksi huomattavaksi ongelmaksi. RapidMiner Studio:ta puolestaan oli miellyttävä käyttää, jos ei lähtenyt tekemään monimutkaisempia analyysejä. PSPP:n kankea ulkoasu ja käyttäminen olivat haasteellisia.



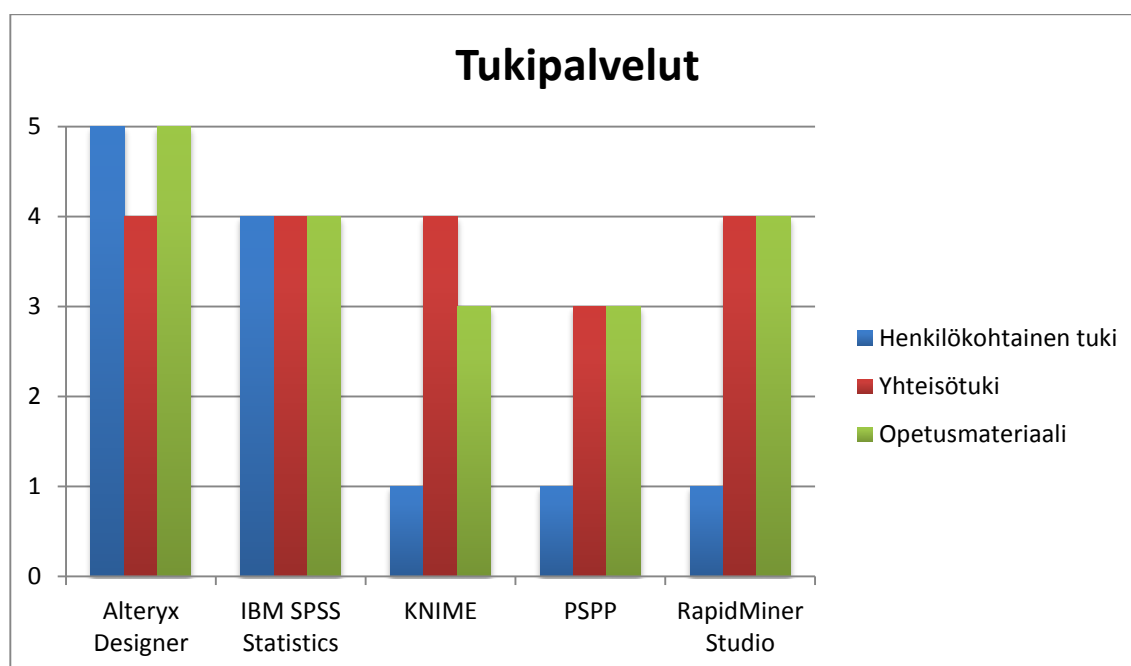
KUVIO 3. Ohjelmistojen käytettävyys

### 6.3 Ohjelmistojen tukipalvelut

Alteryx Designer tarjosi ohjelmistoista parhaan henkilökohtaisen tuen. Nopeaa tukea tarvitsevat voivat olla yhteydessä live-chatissa, puhelimitse ja sähköpostitse. IBM SPSS Statistics tarjosi myös kattavan henkilökohtaisen tuen. Loput ohjelmisto eivät sisältäneet minkäänlaista henkilökohtaista tukea. RapidMiner Studio:n ja KNIME:n on mahdollista saada henkilökohtaista tukea, jos ostaa kaupallisen version, mutta niitä ei tutkittu tässä työssä. Yhteisötuki oli kaikissa ohjelmistoissa hyvää. Jokaisella ohjelmiston sivuilla oli aktiivista yhteisötukea foorumeiden, blogien ja sosiaalisen median muodoissa. PSPP:n

tarjoama yhteisötuki oli silti kaikista vaatimattomin ja se perustui lähinnä yhteen foorumiin.

Opetusmateriaali oli laajinta Alteryx Designer:illa. Ohjelmiston kotisivuilta löytyi laaja kirjo niin tekstejä kuin videoita ohjelmiston käytöstä ja toiminnoista. Ne olivat myös todella havainnollistavia. Myös IBM SPSS Statistics ja RapidMiner Studio tarjosivat samankaltaisia materiaaleja. KNIME:n ja PSPP:n kotisivut sisälsivät myös opetusmateriaalia, mutta rajallisesti.

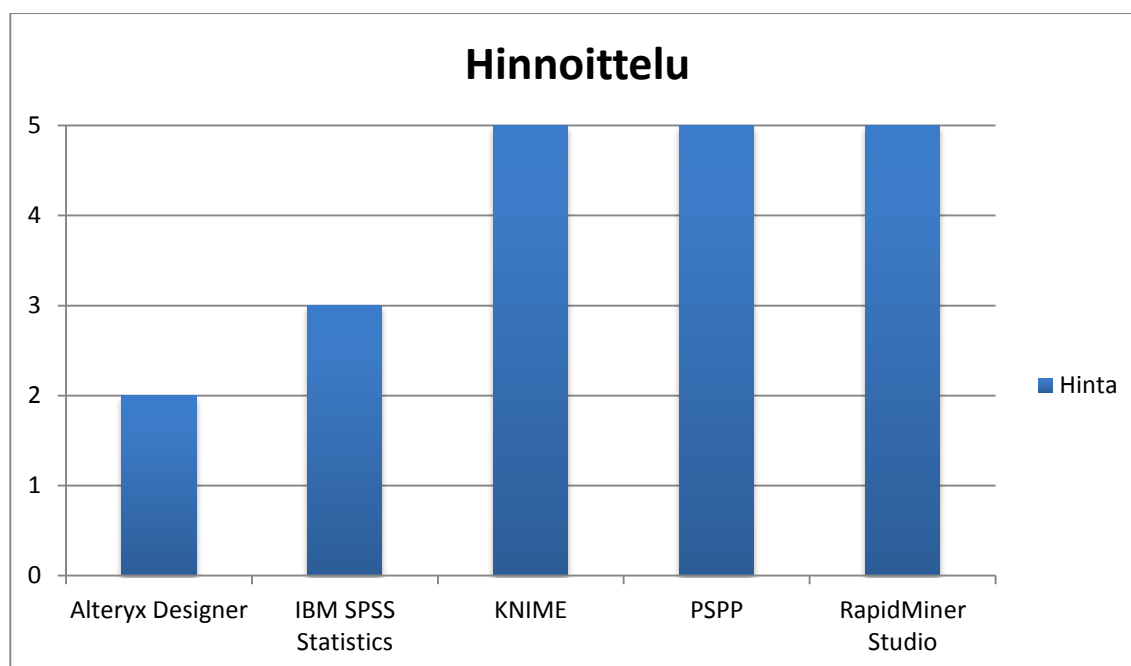


KUVIO 4. Ohjelmistojen tukipalvelut

#### 6.4 Ohjelmistojen hinnoittelu

Ohjelmistojen suurimmat erot löytyivät niiden hinnoitteluista. Alteryx Designer oli kaikista ohjelmista selkeästi kallein, jonka jälkeen tuli IBM SPSS Statistics. Alteryx Designer hieman hämmensi kalliilla hinnalla, varsinkin kun moni muu ohjelmisto tarjoaa samoja ominaisuuksia halvemmalla tai täysin ilmaiseksi. Ohjelmiston hintaa kuitenkin hieman selittää hyvä käytettävyys ja tukipalvelut.

KNIME, PSPP ja RapidMiner Studio olivat täysin ilmaisia ohjelmistoja, mutta RapidMiner Studio:sta ja KNIME:sta oli myös mahdollista hankkia kaupalliset versiot. PSPP ei sisältänyt mitään kaupallista ominaisuutta, ei edes lisäosissa.



KUVIO 5. Ohjelmistojen hinnoittelu

## 6.5 Yhteenveto ohjelmistoista

Loppujen lopuksi kaikki ohjelmistot olivat päteviä erityyppisten analyysien suorittamiseen ja niiden erot olivat enemmän painotuksia kuin vakavia puutteita. Ainoa selvästi heikompi ohjelmisto oli PSPP. Muuten ohjelmistoista jokainen soveltui hyvin liiketoiminta-analytiikan harjoittamiseen. Suurin osa niistä onkin kehitetty liiketoiminta-analytiikan harjoittamista ajatellen. Ohjelmistot tarjosivat laajan kirjon analytiikkaa mahdollistavia toimintoja ja niillä kaikilla pystyi toteuttamaan liiketoiminnan kannalta oleellista kuvaavaa ja ennustavaa analytiikkaa.

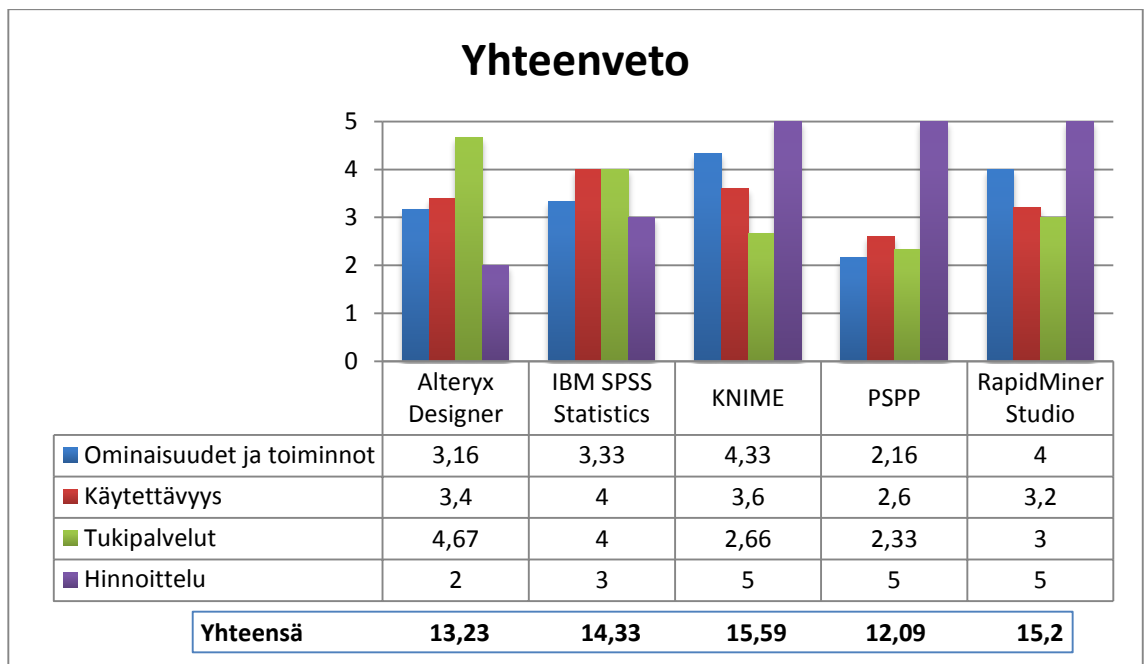
Jokaisella ohjelmistolla oli silti omat vahvuutensa ja heikkoutensa. Tästä syystä niitä oli myös haastava laittaa paremmuusjärjestykseen. Ohjelmistoa valittaessa tulee miettiä, millaiseen käyttöön se tulee. Ominaisuuksiltaan ja toiminnoiltaan RapidMiner Studio ja KNIME tarjosivat eniten mahdollisuuksia ja niiden työkalulistat koostuivat yli tuhannesta työkalusta. Alteryx Designer ja IBM SPSS Statistics ovat silti myös hyviä vaihtoehtoja. Jos harjoitetaan vain perusanalytiikkaa, on myös PSPP hyvä vaihtoehto.

Käytettävyydeltään ohjelmistoissa oli jonkin verran eroja. Käyttäjäystävällisimmät ohjelmistoista olivat IBM SPSS Statistics ja KNIME. Tukipalveluiltaan voittajaksi selvisi Al-



teryx Designer, joka tarjosi varsinkin erinomaisen ja nopean henkilökohtaisen tuen. Kaupallisista ratkaisusista IBM SPSS Statistics sisälsi myös hyvät tukipalvelut. Hinnoittelultaan KNIME, PSPP ja RapidMiner Studio olivat täysin ilmaiset. Jos ohjelmiston kriteerinä on täysin ilmainen ohjelmisto ja halutaan suorittaa edistynyttä analytiikkaa, niin KNIME tai RapidMiner Studio ovat käypiä vaihtoehtoja. Ilmaisista ohjelmistoista PSPP sopii parhaiten perusanalytiikan tehtäviin. Täysin kaupallisista ratkaisuksista IBM SPSS Statistics selvisi hinnaltaan voittajana. Alteryx Designer:in hinta oli yllättävän kallis verrattuna muihin ohjelmistoihin.

Tässä yhteenvedossa tulee kuitenkin muistaa, että jokainen verrattava osa-alue sisälsi vielä lisää painotuksia, jotka käytiin aiemmin lävitse. Yhteenvetokuvioista (Kuvio 6) ei esimerkiksi käy ilmi edellä käytyjä tarkempia osa-alueita. Siinä on laskettu kaikkien aikaisempien osa-alueiden keskiarvot ja näiden yhteispisteet.



KUVIO 6. Ohjelmistojen yhteenveto laskettuna edellisten kuvaajien osa-alueista

## 7 POHDINTA

Opinnäytetyön tekeminen oli pitkä prosessi, joka alkoi teoriapohjan luomisesta, jonka jälkeen siirryttiin selvitystyöhön. Lähteitä teoriapohjaan löytyi runsaasti, joten aluksi suurimmaksi ongelmaksi muodostui oikean tiedon löytäminen. Valitsin lähteiksi luotettavimmat aineistot kirjallisuudesta artikkeleihin. Selvitys toteutettiin ohjelmistoja testaamalla ja käyttämällä hyväksi lähinnä ohjelmistojen omia sivustoja

Opinnäytetyöprosessin aikana itselleni muodostui kattava näkemys analytiikkaohjelmistoista ja siitä, mitkä asiat voivat vaikuttaa niiden valintaan. Vaikka ohjelmistoissa oli paljon samankaltaisuuksia, oli silti mielenkiintoista huomata eroavaisuuksia painotuksissa. Suurimpina ongelmina opinnäytetyön toteutuksessa olikin kunnan eroavaisuuksien löytäminen ohjelmistojen välillä. Varsinkin toiminnoiltaan ja ominaisuuksiltaan ohjelmistoissa oli paljon samaa. Lisäksi tutkittavien ohjelmistojen määrä oli kattava, joten laajempaan syventymiseen ei ollut aikaa. Ohjelmistoista saatiin silti selvitettyä oleellimmat tiedot ja ne ohjaavat näin lukijaa tarkemman tiedon pariin. Tarkoituksena oli kuitenkin muodostaa lyhyt ja ytimekäs katsaus ohjelmistoihin. Mielestäni tavoitteeseen päästiin.

Analytiikkaohjelmistoista löytyy runsaasti erilaista tietoa ja tutkimuksia, mutta näiden löytäminen ja tutkiminen voi osoittautua työlääksi. Ohjelmiston valintaa tehtäessä ei välttämättä ole edes aikaa tutustua yksityiskohtaisesti kaikkeen mahdolliseen tietoon, varsinkin, jos on kiire päästä hyödyntämään analytiikan mahdollisuuksia ja säästää kustannuksissa. Huonoimmassa tapauksessa ohjelmistovalinnat tehdään pelkästään esimerkiksi alhaisen hinnan perusteella. Tämän takia on oleellista tietää, mihin tarpeisiin ohjelmisto valitaan ja mitä se pystyy tarjoamaan. Tämän opinnäytetyön tarkoitus onkin toimia tukena ohjelmiston valinnassa. Olen kerännyt tähän opinnäytetyöhän oleellisia asioita, joita tulisi ottaa huomioon analytiikkaohjelmiston valinnassa nimenomaan liiketoiminta-analytiikan kannalta.

Vaikka ohjelmistoissa riitti tutkittavaa, oli opinnäytetyön tekemiselle varattu tarpeeksi aikaa ja se valmistui ajallaan. Muutenkin opinnäytetyön tekeminen oli antoisaa ja mielekästä. Sekä tutkiessani että lukiessani aiheeseen liittyvää teoriaa opin paljon analytiikkaohjelmistojen toiminnoista, käytettävyydestä, tukipalveluista ja hinnoittelusta sekä näiden vaikutuksesta itse ohjelmiston valintaan.

## LÄHTEET

Alteryx. 2015a. Luettu 20.8.2015.

<http://www.alteryx.com/press-releases/src-llc-is-now-alteryx>

Alteryx. 2015b. Luettu 20.8.2015.

<http://www.alteryx.com/products/pricing>

Azevedo, A. & Santos, M. 2014. Integration of Data Mining in Business Intelligence Systems. Hershey: IGI Global.

Bulusu, L. 2012. Open Source Data Warehousing and Business Intelligence. Boca Raton: CRC Press.

Camm, J., Cochran, J., Fry, M., Ohlmann, J. & Anderson, D. 2014. Essentials of Business Analytics. Boston: Cengage Learning.

Davenport, T. 2010. The New World of "Business Analytics". Luettu 4.7.2015.

[http://www.sas.com/resources/asset/IIA\\_NewWorldofBusinessAnalytics\\_March2010.pdf](http://www.sas.com/resources/asset/IIA_NewWorldofBusinessAnalytics_March2010.pdf)

Davenport, T. & Harris, J. 2007. Analysoi ja voita: Kilpailun uusi tiede. Helsinki: Talentum.

Fitz-Enz, J. 2010. The New HR Analytics: Predicting the Economic Value of Your Company's Human Capital Investments. New York: AMACOM Books.

Gupta, M. & Gupta, D. 2011. Research Methodology. Delhi: PHI Learning Pvt. Ltd.

Gnu. 2015. Luettu 25.8.2015.

<https://www.gnu.org/software/pspp/faq.html>

Helkiö, T. 2012. Analytiikka tekee fiksuksi. Tivi. Luettu 2.7.2015.

<http://www.tivi.fi/Kumppaniblogit/affecto/2012-10-17/Analytiikka-tekee-fiksuksi-3195548.html>

IBM. 2015. Luettu 10.8.2015.

<https://www.ibm.com/marketplace/cloud/spss-statistics/fi/en-fi?step=Plan>

KNIME. 2015. Luettu 18.8.2015.

<https://www.knime.org/products/product-matrix>

Kovačič, M. 2009. PSPP. slo-tech. Luettu 25.8.2015.

<https://slo-tech.com/clanki/09006/>

Laihonen, H., Hannula, M., Helander, N., Ilvonen, I., Jussila, J., Kukko, M., Kärkkäinen, H., Lönnqvist, A., Myllärniemi, J., Pekkola, S., Virtanen, P., Vuori, V. & Yliniemi, T. 2013. Tietojohdaminen. Tamere: Juvenes Print.

Nielsen, J. 1993. Usability Engineering. San Diego: AP Professional.

Schniederjans, M., Schniederjans, D. & Starkey, C. 2014. Business Analytics Principles, Concepts, and Applications: What, Why, and How. Upper Saddle River: Pearson FT Press.

Strickland, J. 2014. Data Science and Analytics for Ordinary People. Raleigh: Lulu.

RapidMiner. 2015a. Luettu 26.8.2015.

<https://rapidminer.com/press-kit/>

RapidMiner. 2015b. Luettu 26.8.2015.

<https://rapidminer.com/products/comparison/>

Vorhies, B. 2014. Prescriptive versus Predictive Analytics – A Distinction without a Difference?. Predictiveanalyticsworld. Luettu 20.9.2015.

<http://www.predictiveanalyticsworld.com/patimes/prescriptive-versus-predictive-analytics-distinction-without-difference/>