

Opinnäytetyö (AMK)

Liiketalous

2023

Sampsa Virta

Data-analytiikkaan perustuva tiedon käsittely, muotoilu ja visuaalinen esitys Exceliä käyttäen



Opinnäytetyö (AMK) | Tiivistelmä

Turun ammattikorkeakoulu

Liiketalous

2023 | 79 sivua

Sampsa Virta

Data-analytiikkaan perustuva tiedon käsittely, muotoilu ja visuaalinen esitys Exceliä käyttäen

Opinnäytetyön tavoitteena oli tarkastella data-analytiikkaa ja datan visualisointia. Tavoitteena oli myös selvittää selkeät visualisointitekniikat ja kertoa niistä esimerkkien avulla. Lisäksi tavoitteena oli laatia kokoelma hyödyllisistä Excel-funktioista ja toiminnoista esimerkkien kanssa.

Työn teoriaosuudet toteutettiin kirjallisten ja verkkolähteiden avulla. Visualisointien ja funktioiden esimerkit toteutettiin Microsoft Excelillä ja muu kuvamateriaali on verkosta tai kirjoista lainattu.

Työn tulokseksi saatiin teoriaa data-analytiikasta, esimerkki PowerQuery:stä, esimerkkejä monista data-analytiikan toiminnoista Excelillä ja teoriaa sekä neuvoja datan graafisesta esityksestä.

Tuloksia voidaan hyödyntää apuna Excelin käytössä tai visualisointien luomisessa.

Asiasanat:

Microsoft Excel, data, tiedonhallinta, visualisointi, big data, business intelligence, muotoilu, kaaviot

Sisältö

Käytetyt lyhenteet tai sanasto	8
1 Johdanto	9
2 Data-analytiikka	11
2.1 DIKW-pyramidi	11
2.2 Business Intelligence	13
2.3 Traditional Data vs. Big Data	13
3 Datan tuonti ja muokkaus	16
3.1 PowerQuery	16
3.2 Tiedon tuonti verkosta	23
4 Analytiikkaa funktioilla	26
4.1 Tekstin käsittely ja merkkijonot	26
4.2 Aika ja aikafunktiot	30
4.3 Lajittelu ja suodatus	32
4.4 Datan tarkastelu	35
4.5 Hakufunktiot	36
4.6 Loogiset funktiot	39
4.7 Tilastolliset funktiot	44
4.8 Yhdistäminen	47
4.9 Muut Excel-työkalut - Optimointi	47
5 Graafinen esitys	51
5.1 Visualisointitavat	52
5.1.1 Taulukko ja lämpökartta	53
5.1.2 Pisteet ja viivat	54
5.1.3 Pylväät	55
5.1.4 Aluekaaviot (puukaavio ja vohvelikaavio)	61
5.2 Muotoilu	63
5.2.1 Vältettävät visualisoinnit	64

5.2.2 Luettavuuden parantaminen	68
5.2.3 Ylimääräisen tiedon vähentäminen	69
5.2.4 Korostus	70
5.2.5 Tekstit	71
5.2.6 Värit	72
5.2.7 Muut visuaaliset elementit	73
6 Johtopäätökset	75
Lähteet	77

Kuvat

Kuva 1. Datan tuonti.	17
Kuva 2. Datan tuonti, toinen vaihe.	17
Kuva 3. PowerQuery-ikkuna.	18
Kuva 4. Sarakkeiden ja rivien hallinta.	18
Kuva 5. Muokkauksen vaiheet, korostettuna poistetut rivit ja sarakkeet.	19
Kuva 6. Suodatus.	19
Kuva 7. Muokatut tiedot.	20
Kuva 8. Sarakkeen lisääminen.	20
Kuva 9. Sarakkeen lisääminen, toinen vaihe.	21
Kuva 10. Tallennetut vaiheet.	21
Kuva 11. Taulukon tuominen.	22
Kuva 12. Lopullinen taulukko.	23
Kuva 13. Verkkosisällön hakeminen.	24
Kuva 14. Sarjataulukko Excelin verkkosisältönavigaattorissa.	25
Kuva 15. Esimerkki CONCAT-funktiosta.	26
Kuva 16. Esimerkki TEXTJOIN-funktiosta.	27
Kuva 17. Esimerkki LEFT-funktiosta.	27
Kuva 18. Esimerkki MID-funktiosta.	28
Kuva 19. Esimerkki LEN-funktiosta.	28
Kuva 20. Esimerkki FIND-funktiosta.	29

Kuva 21. Esimerkki merkkijonofunktioiden yhdistämisestä.	29
Kuva 22. Esimerkki virheellisestä muotoilusta.	30
Kuva 23. Aikasolun muotoilun muutos.	31
Kuva 24. Esimerkki virheellisestä laskutoimituksesta.	31
Kuva 25. Virheellisen laskutoimituksen korjaus.	31
Kuva 26. Esimerkki SORT-funktiosta.	32
Kuva 27. Esimerkki SORTBY-funktiosta.	33
Kuva 28. Esimerkki FILTER-funktiosta.	34
Kuva 29. Esimerkki SORT- ja FILTER-funktioiden yhdistämisestä.	35
Kuva 30. Esimerkit COUNTIFS ja SUMIFS-funktioista.	35
Kuva 31. Esimerkki XLOOKUP-funktiosta.	37
Kuva 32. XLOOKUP-funktion lopputulos.	37
Kuva 33. Esimerkki XLOOKUP-funktion yhdistämisestä toiseen funktioon sekä useamman tiedon palautus.	37
Kuva 34. Esimerkki kahdensuuntaisesta XLOOKUP-hausta.	38
Kuva 35. Kahdensuuntaisen XLOOKUP-haun lopputulos.	38
Kuva 36. Esimerkki yksinkertaisesta IF-lausekkeesta.	39
Kuva 37. Sovellus IF-lauseketta käyttäen.	39
Kuva 38. IF-lausekkeen lopputulos.	40
Kuva 39. Esimerkki IFS-funktiosta.	40
Kuva 40. IFS-funktion lopputulos.	41
Kuva 41. Esimerkki sisäkkäisestä IF-rakenteesta.	41
Kuva 42. Sisäkkäisen IF-rakenteen lopputulos.	42
Kuva 43. Esimerkki AND-funktiosta.	42
Kuva 44. AND-funktion lopputulos.	43
Kuva 45. Esimerkki OR-funktiosta.	44
Kuva 46. OR-funktion lopputulos.	44
Kuva 47. Esimerkki SUMPRODUCT-funktion toiminnasta.	45
Kuva 48. SUMPRODUCT-funktion lopputulos.	45
Kuva 49. Esimerkki RANK-funktiosta.	46
Kuva 50. Muita hyödyllisiä funktioita.	46
Kuva 51. Esimerkki funktioiden yhdistämisestä XLOOKUP-funktiolla.	47

Kuva 52. Indeksirahastosovellus ennen optimointia.	47
Kuva 53. Esimerkki Solver-toiminnosta.	48
Kuva 54. Solver ratkaisee.	48
Kuva 55. Yksinkertainen katetuottolaskelma.	49
Kuva 56. Goal Seek.	49
Kuva 57. Goal Seek-ratkaisu.	50
Kuva 58. Suomen työeläkkeensaajat-julkaisusta poimittuja korotuksia.	52
Kuva 59. Z-kuvio: luonnollinen lukusuunta.	74

Kuviot

Kuvio 1. DIKW-Pyramidi.	12
Kuvio 2. Sisäkkäisen funktion toiminta.	30
Kuvio 3. Pistekaavio, jossa ajatut mailit Y-akselilla ja auton hinta X-akselilla.	54
Kuvio 4. Viivakaavio, jossa polttonesteiden hinnat Y-akselilla ja aika X-akselilla.	55
Kuvio 5. Muutettu Y-akseli eri perusviiva.	56
Kuvio 6. Pylväskaavioita yhdellä ja useammalla sarjalla.	57
Kuvio 7. Pinottu pystysuora pylväskaavio.	58
Kuvio 8. Pinottu vaakasuuntainen pylväskaavio, jossa X-akseli summattu 100 prosenttiin.	59
Kuvio 9. Vesiputouskaavio kuukausituloista.	60
Kuvio 10. Osakkeiden hintojen jakauma Helsingin pörssissä, histogrammi.	61
Kuvio 11. Puukaavio verkkokaupan tuotosta tuotekategorioittain.	62
Kuvio 12. Vohvelikaavio haastatteluista.	63
Kuvio 13. Esimerkki ympyräkaaviosta	65
Kuvio 14. Esimerkki donitsikaaviosta.	65
Kuvio 15. Saman tiedon kommunikointi pylväskaavion avulla.	66
Kuvio 16. Myytyjen polkupyörien määrä ensisijaisella Y-akselilla ja polkupyörien hintojen keskiarvo toissijaisella Y-akselilla.	67
Kuvio 17. Vaihtoehdot kahden Y-akselin käyttöön.	67
Kuvio 18. Esimerkki 3D-kaaviosta kolmella akselilla.	68

Kuvio 19. Alkuperäinen kaavio.	69
Kuvio 20. Muokattu kaavio.	70
Kuvio 21. Esimerkki korostuksesta kuviota 20 käyttäen.	71
Kuvio 22. Esimerkki annotaatiosta.	72
Kuvio 23. Värin käyttö.	73
Kuvio 24. Pylväsdiagrammi vakaatasossa.	74

Taulukot

Taulukko 1. Esimerkki taulukosta ja lämpökartasta	53
Taulukko 2. Suurin arvo korostettu jokaisesta mitasta.	53

Käytetyt lyhenteet tai sanasto

Big Data	Massadata
Traditional Data	Perinteinen data
Dashboard	Interaktiivinen tiedonhallintapaneeli, käytetään tiedon näyttämiseen, seuraamisen ja analysoimiseen.
DIKW	Data, Information, Knowledge, Wisdom
BI	Business Intelligence
Business Intelligence	Liiketoimintatiedon hallinta
.csv, .txt, xlsx	Yleisiä tiedostopäätteitä tietojoukkotiedostoille
PowerQuery	Microsoft Exceliin sisällytetty tiedon muunnos- ja valmistelutyökalu.
Delimiter	Arvojen välissä oleva erotin
Lookup	Excel-funktio, jolla haetaan tietoa taulukosta
Argumentti	Funktioissa käytetty syötetieto.
Likert-asteikko	Kyselytutkimusasteikko, jossa vastaukset jaetaan viiteen seuraavasti: Ehdottomasti samaa mieltä, Osittain samaa mieltä, Ei samaa eikä eri mieltä, Osittain eri mieltä, Ei lainkaan samaa mieltä.
Logiikkatesti	Funktioissa käytetty testi, jonka tulos TOSI tai EPÄTOSI.

1 Johdanto

Nykyaikana tiedolla johtaminen on suuri osa monen yrityksen liiketoiminnan päätöksentekoa. Tiedolla johtaminen tarkoittaa, että päätösten takana on perusteena jokin tieto. Tieto saadaan yritysten keräämästä materiaalista eli datasta data-analyysin avulla. Tästä syystä data-analyysi on tärkeässä asemassa. On myös todella tärkeää osoittaa visuaalisesti ja kommunikoida tehokkaasti datan ja tiedon esille tuomat tulokset. Tämä nostaa datan visualisoinnin asemaa.

Opinnäytetyön aihe on data-analytiikka Microsoft Excelillä. Data-analytiikan aihepiiri on kokonaisuudessaan suuri, mutta opinnäytetyö keskittyy data-analytiikan selventämiseen, funktioiden toteuttamiseen Excelillä sekä tiedon esittämiseen visuaalisesti. Aihe on merkityksellinen, koska data-analytiikka on tärkeä osa sekä taloushallinnon kuin muiden alojen nykyaikaa ja tulevaisuutta. Data-analytiikkaa tarvitaan, jotta voidaan erottaa tieto datasta ja kommunikoida se selkeästi. Päädyin aiheeseen kiinnostuksesta data-analytiikkaan ja tiedon esittämiseen erilaisilla työkaluilla. Excel on minulle sopiva, sillä olen jo sen edistynyt käyttäjä ja voin opinnäytetyön kautta paneutua syväisemmin erilaisiin toimintoihin data-analytiikan tarkoituksiin. Opinnäytetyössä käytettyjä työkaluja on siis tarkoituksella rajoitettu saavutettavuuden vuoksi. Halusin myös oppia tietoteknisen soveltamisen lisäksi kuvioden suunnittelusta muotoilun kannalta sekä löytää käytännöllisiä tapoja lisätä kuvioden luettavuutta tiedon kommunikoinnin parantamiseksi.

Tavoitteena on tarkastella erilaisia data-analytiikan toimintoja Excelissä, toteuttaa niistä esimerkit ja kirjoittaa niistä. Lisäksi opinnäytetyö voi toimia oppaana näihin Excelin toimintoihin. Tavoitteena on myös soveltaa data-analytiikan alan osia pienellä skaalalla ja tutulla työkalulla eli Excelillä. Työn on tarkoitus antaa lukijalle selkeä kokonaiskuva työssä käsitellyistä aiheista ja tavoitteena on saada lukija miettimään enemmän miten visualisoida tietoa tarkoituksellisesti ja selkeästi. Työn aihe muuttui useasti kirjoittamisen aikana, sillä aihe on kiinnostava ja teoriaa sekä kiinnostavia työkaluja löytyy todella

paljon. Rajoitin aiheen kuitenkin ainoastaan teoriaan data-analytiikasta, visualisointeihin sekä Excel-funktioiden käyttöön. Tästä syystä päätin olla kirjoittamatta esim. Pivot-työkaluista Excelissä, sillä ne painottavat enemmän interaktiivisuutta ja tiedon esittämistä interaktiivisilla elementeillä eikä tiedon saantia tai muokkausta.

Työn alussa tarkastellaan data-analytiikkaa alana ja sen eri osia. Lisäksi kirjoitan trendikkään Big Datan ja perinteisemmän ”Traditional” Datan eroavaisuuksista. Tämän jälkeen tutustutaan työssä käytettyihin työkaluihin, sitten toteutetaan data-analyysiä Excel-funktioilla ja lopuksi tarkastellaan datan visualisointitapoja sekä visualisointien muotoilua. Työssä käytetään useita englanninkielisiä nimityksiä, mutta suomennokset mainitaan tarvittaessa.

2 Data-analytiikka

Data-analytiikan määritelmänä voidaan pitää tietomäärien muodostamista muotoihin, joissa datan kokonaisuus on helpommin ymmärrettävissä ja sen avulla on helpompi tehdä päätöksiä. Tähän käytetään mm. erilaisia visualisointeja, tunnuslukuja ja näiden yhdistelmiä dashboardilla. Data-analytiikan tavoitteena on jalostaa raakadata käytännöllisempään muotoon.

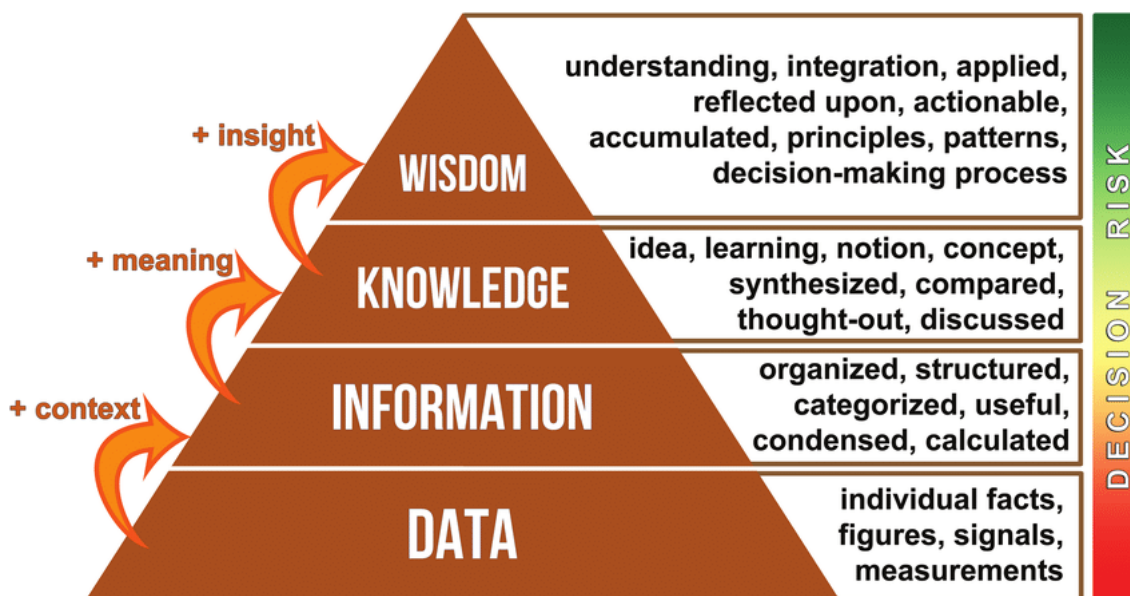
Data on uusi luonnovara ja sitä kerääntyy luonnollisesti erilaisten prosessien yhteydessä. Sen kerääminen mahdollistaa sen käsittelyn, data-analytiikan hyödyntämisen ja johtopäätösten tekemisen. Usein dataa ajatellaan numeerisena tietona, mutta datalla on monta muotoa. Se voi olla numeerista: diskreettiä (esim. kokonaisluvut, arvosanat) tai jatkuvaa (esim. euromäärät tileillä, pituus/paino), tai kategorista: nominaalista (silmien väri esim. "ruskea"/"sininen") tai ordinaalista (esim. "harvoin", "joskus", "usein"). (Donges 2018.)

Kuitenkin data on melkein aina jatkuvaa numeerista tietoa, ellei analyysissa käytetä mukana tietoja asiakastytyväisyyskyselyistä tai vastaavaa dataa. Lisäksi jotkut datatyypit soveltuvat erilaisiin visualisointeihin, hyvänä esimerkkinä geograafinen data. Vaikka koordinaattien data on numeerista, voidaan ne esittää kartalla.

Lähes kaikki yritykset ja muut yhteisöt keräävät dataa. Taloushallinnossa voidaan yksityisen datan lisäksi hyödyntää erilaista julkista avointa dataa analyyseissä. Esimerkiksi Suomen kaupunkien keräämät liikennetiedot voivat olla hyödyllistä autoalalla toimiville yrityksille.

2.1 DIKW-pyramidi

Tietojenkäsittelyssä datasta usein käytetään DIKW (data, information, knowledge, wisdom)-hierarkiaa (Mäkelä & al. 2020).



Kuvio 1. DIKW-Pyramidi (Cannas & al. 2019).

Hierarkiaa voidaan soveltaa eri puitteissa, mutta esimerkkinä taloushallinnossa data on usein lukuja ja numeroita. Datasta voidaan tehdä havaintoja, mutta se ei sisällä tulkintaa. Seuraavalla tasolla luodaan datasta informaatiota esimerkiksi erilaisilla kaavioilla, jolla voidaan vastata kysymykseen "mitä" eli mitä tietoa saadaan irti datasta. Seuraava tietämyksen (knowledge) taso vastaa kysymykseen "miten" yhdistämällä kerätyt faktat, niiden suhteet ja kontekstit (Mäkelä & al. 2020).

Kuvion 1 alimmat kolme tasoa perustuvat historialliseen tietoon ja sen jalostamiseen, ymmärtämiseen ja johtopäätöksiä luomiseen. Pyramidin huipulla viisauden (wisdom) kohdalla saadaan tämän prosessin kautta jonkinlaista ymmärrystä menneestä. Viisauden avulla voidaan arvioida mennyttä eri kannoilta (onko jokin toistuva kuvio, oliko aiemmat päätökset hyviä) ja tehdä parempia päätöksiä tulevaisuudessa. (Mäkelä & al. 2020).

Data-analytiikan voidaan siis ajatella sisältyvän alimmille kolmelle portaille, keskittyen toiselle informaatioportaille, sillä tässä välissä tapahtuu kaikki mekaaninen tiedon käsittely, jonka tavoitteena on luoda tietämystä.

2.2 Business Intelligence

Business Intelligence, joskus suomeksi liiketoimintatiedon hallinta, tarkoittaa yleisesti strategioita ja teknologioita, joita yritykset käyttävät analysoidakseen erilaisia liiketoimintatietoja ja luomaan analyysin kautta strategisia ja taktisia liiketoimintapäätöksiä. BI-työkalut muuntavat ja analysoivat tietojoukkoja ja esittävät analyttisiä tuloksia mm. raporteissa, yhteenvedoissa, dashboardeissa, kuvioissa ja kaavioissa. Usein BI:llä myös viitataan erilaisiin työkaluihin, jotka tarjoavat helposti hahmotettavaa tietoa organisaation nykytilasta saatavilla olevien tietojen perusteella (Olavsrud & al. 2023).

Lyhyesti voidaan todeta, että BI:llä tarkoitetaan siis data-analytiikan prosessien ja käytäntöjen hyödyntämistä organisaation tietojen avulla organisaation historian ja nykytilanteen arvioimiseksi (Olavsrud & al. 2023). BI keskittyy siis historialliseen tietoon, jolla saadaan älykkyyttä (intelligence) liiketoiminnan (business) toiminnasta. Kontrastina predikttiivinen analytiikka, joka keskittyy tulevaisuuden ennakoimiseen. (Olavsrud & al. 2023).

Nykyään lyhenteellä BI usein tarkoitetaan ulkoisen osapuolen tarjoamia ja yritysten käyttämiä järjestelmiä kokonaisuutena, sisältäen käyttöönoton, datan keruun ja varastoinnin sekä analytiikkatyökalut.

2.3 Traditional Data vs. Big Data

Usein puhutessa data-analytiikasta, sen mukana mainitaan big data. Yritysten keräämät tiedot ovat kuitenkin melkein aina "traditional dataa" eli perinteistä dataa. Tunnusomaista tälle datalle on sen rakenteellisuus, mikä tekee siitä helpompaa tallentaa, hallita ja analysoida (Pure Storage 2023). Hyvä ja yleinen esimerkki traditional data:sta on toiminnanohjausjärjestelmien avulla kerätty ja hallinnoitu liiketoimintadata. Kuitenkaan tämä data ei ole määrältään tai arvoltaan kovinkaan pieni, vaan esimerkiksi suurien yritysten taloudelliset tiedot ja kirjanpidon raportit luokitellaan traditional data:ksi.

Big dataa kuvataan usein viiden V:n mallilla, jossa jokainen V-sana kuvaa yhtä big datan ominaisuutta ja näiden viiden yhdistelmällä saadaan siitä hyvä kokonaiskuva. Käymällä läpi nämä viisi V:tä voidaan siis samalla verrata niitä traditional dataan, jonka lopputuloksena saadaan hyvä kuva molemmista ja niiden eroavaisuuksista. (Kacfeh Emani 2015.)

Volume (Määrä)

Suurien tietomäärien käsittelystä saatava hyöty on big data -analytiikan suurin vetovoima. Big datan luodun ja varastoidun datan määrä tuo enemmän arvoa kuin hyvä tiedon mallinnus. Seuraksena on se, että monet yritykset keräävät valtavia määriä erilaista tietoa välittämättä sen hyödyllisyydestä (Kacfeh Emani 2015).

Traditional data on määrältään paljon pienempi, rakenteeltaan järjestelmällisempi ja usein keskittyy yhteen datalähteeseen.

Variety (Valikoima) viittaa datan tyyppiin ja luonteeseen. Saatu data koostuu useasta eri lähteestä ja on usein jäsentymätöntä. Big data vaatii siis aina käsittelyä jatkotoimenpiteitä varten. Tällainen data voi olla jäsenneiltyä (esim. tiedot relaatiotietokannoista), puolirakenteista (verkkolokit, sosiaalisen median syötteet, raakasyötteet suoraan anturilähteestä, sähköpostista jne.) tai jäsentämättömiä (videot, kuvat, äänitiedostot) (Kacfeh Emani 2015). Traditional datalla on kiinteä rakenne ja on usein täydellisesti järjestetyssä muodossa jatkokäsittelyä varten. Traditional data on siis joko jäsenneiltyä tai puolirakenteista, jota on jollain tavalla jo muunneltu jatkokäsittelyä varten.

Velocity (Nopeus) viittaa siihen, kuinka nopeasti dataa luodaan ja kuinka nopeasti tiedot liikkuvat. Tietovirta on jatkuvaa, sitä luodaan ja lähetetään jatkuvasti. Tieto voi virrata esim. tietokoneista, verkoista, älypuhelimista tai sosiaalisesta mediasta. Nämä tiedot voidaan sitten käsitellä ja analysoida melkein jopa reaaliajassa. (Kacfeh Emani 2015.) Traditional data voi usein olla

myös nopeaa, mutta useimmiten historiallista dataa kerätään suuremmalla tarkkuudella ja rakenteellisuudella, mikä tekee siitä myös hitaampaa. Esimerkki hitaasta datasta ovat erilaiset päivittäin, viikottain tai kuukausittain kerätyt raportit.

Value (Arvo) viittaa siihen, että datan tulee tuottaa arvoa ja sitä tulee pystyä hyödyntämään. Tietenkin tämä tarve on molemmilla datatyypeillä. Big datalla useimmiten yritetään löytää uusia tarpeita tai mahdollistaa uusia liiketoimintamallia, tuotteita tai palveluja (fokus tulevaisuuteen) (Kacfeh Emani 2015).

Traditional datalla useimmiten autetaan esim. yrityksiä ymmärtämään tietyn strategian tai muutoksen vaikutukset rajalliseen mittausalueeseen tietyn ajanjakson aikana (Pure Storage 2023).

Veracity (Todenmukaisuus)

Big datan kerättyjen tietojen laatu voi vaihdella suuresti, mikä taas vaikuttaa analyysin tarkkuuteen. Todenmukaisuuden (tai sen puutteen) vuoksi Big Datasta saatuja tuloksia ei voida todistaa, mutta niille voidaan määrittää todennäköisyys (Kacfeh Emani 2015).

Traditional datan keruu on suunniteltua ja rakenteellista, jotta voidaan luottaa analyysin tarkkuuteen.

Näiden viiden lisäksi myös tärkeä eroavaisuus on se, että usein big datan käsittelyyn tarvitaan todella tehokas supertietokone, kun taas traditional datan käsittelyyn voidaan käyttää perinteisiä työkaluja, esimerkiksi Exceliä tai SQL-kieltä. Traditional data voi kuitenkin esiintyä massiivisena, järjestettynä tietojoukkona.

3 Datan tuonti ja muokkaus

Excel on taulukkolaskentaohjelma. Se ei usein ole paras työkalu, mutta se on helposti opittavissa ja yleisesti helppokäyttöinen. Useilla on jo osaaminen sen perustoiminnoista ja sen käyttöliittymä on tuttu, joten uusien asioiden oppimisessa ja soveltamisessa ei ole suuria esteitä. Lisäksi sovelluksessa on paljon erilaisia ominaisuuksia ja siinä tehdyt laskentataulukot voidaan tarvittaessa siirtää tai viedä muihin ohjelmiin, mikä tekee siitä ideaalin työkalun tiedon nopeaan käsittelyyn ja visualisointien luomiseen.

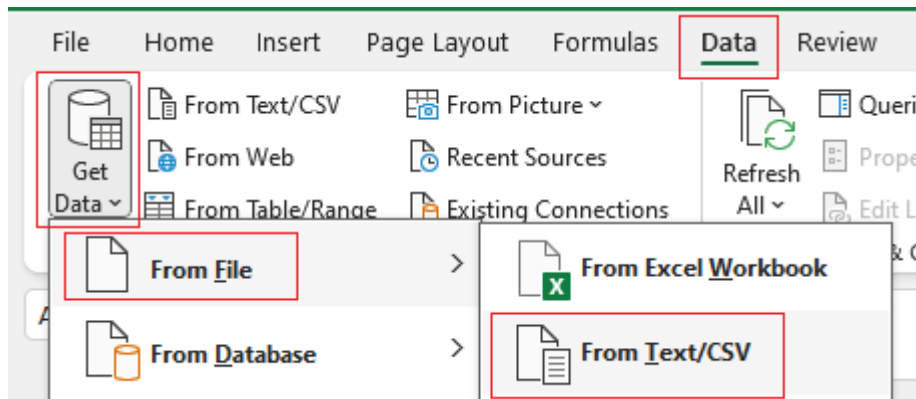
3.1 PowerQuery

PowerQuery:ä käytetään usean datan muunnokseen, valmisteluun, siivoukseen. PowerQuery:llä voi tuoda dataa kerralla useasta datalähteestä ja valita tarvittavat tiedot ja muuntaa se halutulla tavalla.

Datalähteenä käytetään usein jostain sovelluksesta vietyä (export) dataa, esimerkiksi taloushallinnon sovelluksen raportit. Data voi olla erilaisissa muodoissa, useimmiten .csv, .txt tai Excelin .xlsx muodoissa. Dataa voidaan myös tuoda Exceliin ja muihin ohjelmiin suoraan tietokannoista esim. Microsoft Access, SQL Server tai Oracle Database.

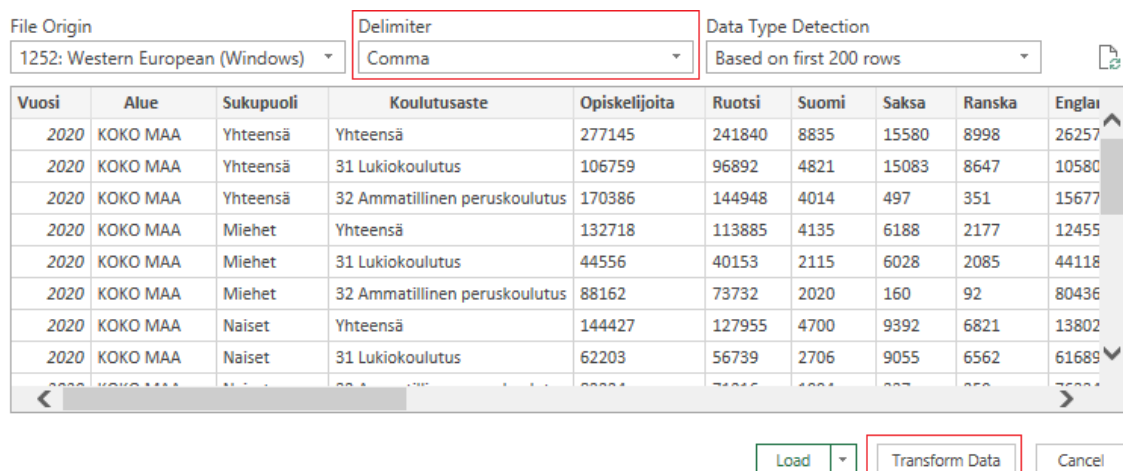
Dataa voidaan myös haravoida verkosta (web scraping), joko suoraan Excelillä, muulla ohjelmalla tai koodilla. Varsinkin haravoidussa datassa voi olla muodollisia virheitä, jotka täytyy siivota pois PowerQuery:n avulla.

Seuraavaksi esimerkki tietojen valinnasta PowerQueryä käyttäen. Esimerkissä käytetään verkosta ladattua tilastokeskuksen CSV-tiedostoa, joka sisältää kaikki toisen asteen opiskelijoiden valitsemat vieraat kielet vuosilta 2020-2021 (Tilastokeskus 2020-2021). CSV-tiedosto on tekstimuotoinen ja tästä syystä ei vie paljon tilaa.



Kuva 1. Datan tuonti.

Kuvassa 1 data tuodaan valitsemalla Data->Get Data->From File-> From Text/CSV ja valitsemalla tiedosto.



Kuva 2. Datan tuonti, toinen vaihe.

Tässä kyseisessä csv-tiedostossa tiedot on erotettu pilkuilla, joten valittuna on kuvassa 2 erottimena (delimiter) on pilkku (comma). Seuraavaksi painetaan Transform Data, joka avaa Power Query editorin. Tästä datakokonaisuudesta voidaan esimerkiksi tiedustella englannin ja ruotsin kielten opiskelun osuutta maakunnittain vuonna 2021. Tarvittaessa voidaan myös tuoda jostain muusta lähteestä lisäyksenä tähän kokonaisuuteen. Tuotu tieto nähdään PowerQuery-ikkunassa kuvassa 3.

005_12ad_2021_20231022-174334 - Power Query Editor

File Home Transform Add Column View

Close & Load Refresh Preview Advanced Editor Manage Columns Reduce Rows Sort Split Column Group By Data Type: Whole Number Use First Row as Headers Replace Values Combine Manage Parameters Data source settings New Source Recent Sources Enter Data New Query

Queries [3] OMXH OMXH (2) 005_12ad_2021_2023...

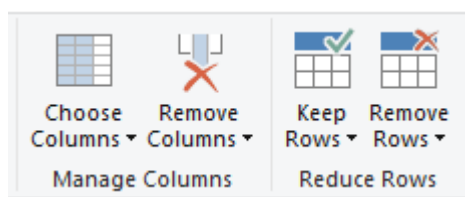
= Table.TransformColumnTypes(#"Promoted Headers",{{"Vuosi", Int64.Type}, {"Alue", type text},

	Vuosi	Alue	Sukupuoli	Koulutusaste	Opiskelijoita
1	2020	KOKO MAA	Yhteensä	Yhteensä	277145
2	2020	KOKO MAA	Yhteensä	31 Lukiokoulutus	106759
3	2020	KOKO MAA	Yhteensä	32 Ammatillinen peruskoulutus	170386
4	2020	KOKO MAA	Miehet	Yhteensä	132718
5	2020	KOKO MAA	Miehet	31 Lukiokoulutus	44556
6	2020	KOKO MAA	Miehet	32 Ammatillinen peruskoulutus	88162
7	2020	KOKO MAA	Naiset	Yhteensä	144427
8	2020	KOKO MAA	Naiset	31 Lukiokoulutus	62203
9	2020	KOKO MAA	Naiset	32 Ammatillinen peruskoulutus	82224
10	2020	KOKO MAA	Tuntematon	Yhteensä	.
11	2020	KOKO MAA	Tuntematon	31 Lukiokoulutus	.
12	2020	KOKO MAA	Tuntematon	32 Ammatillinen peruskoulutus	.
13	2020	MK01 Uusimaa	Yhteensä	Yhteensä	92487
14	2020	MK01 Uusimaa	Yhteensä	31 Lukiokoulutus	37591
15	2020	MK01 Uusimaa	Yhteensä	32 Ammatillinen peruskoulutus	54896
16	2020	MK01 Uusimaa	Miehet	Yhteensä	43708
17	2020	MK01 Uusimaa	Miehet	31 Lukiokoulutus	16428
18	2020	MK01 Uusimaa	Miehet	32 Ammatillinen peruskoulutus	27280
19	2020	MK01 Uusimaa	Naiset	Yhteensä	48779
20	2020	MK01 Uusimaa	Naiset	31 Lukiokoulutus	21163
21	2020	MK01 Uusimaa	Naiset	32 Ammatillinen peruskoulutus	27616
22	2020	MK01 Uusimaa	Tuntematon	Yhteensä	.
23	2020	MK01 Uusimaa	Tuntematon	31 Lukiokoulutus	.
24					

29 COLUMNS, 504 ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 17:56

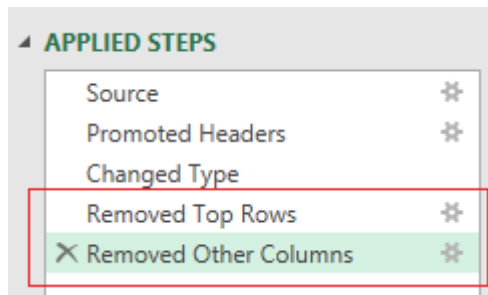
Kuva 3. PowerQuery-ikkuna.

Ladatuissa tiedoissa on paljon ylimääräistä, joten poistetaan tarpeettomat tiedot kuvassa 4 olevien toimintojen avulla.



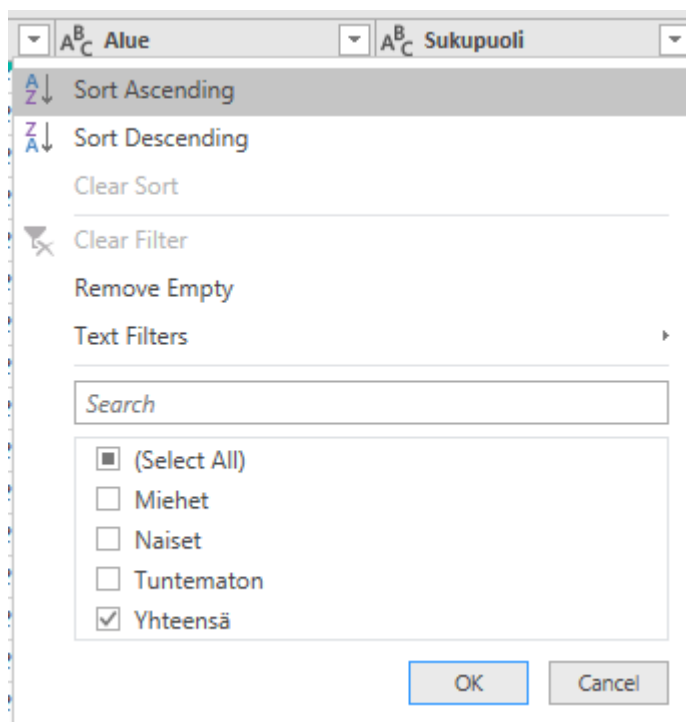
Kuva 4. Sarakkeiden ja rivien hallinta.

Power Query kerää ikkunan sivulle kaikki muokkaamiseen käytetyt toiminnot, joita voi selata takaisinpäin, muokata tai poistaa kuvan 5 mukaisesti:



Kuva 5. Muokkauksen vaiheet, korostettuna poistetut rivit ja sarakkeet.

Excelin tapaan voi tietoja myös suodattaa (filter, kuva 6) ja tämän tekeminen luo myös käytetyn toiminnon listan perälle.



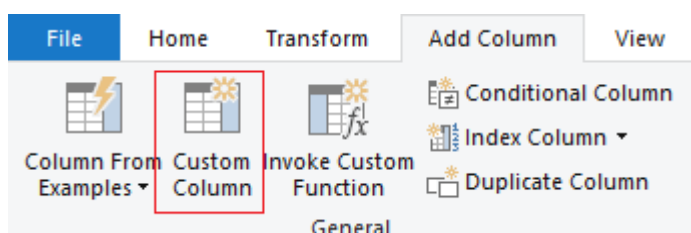
Kuva 6. Suodatus.

Muokkausten jälkeen näyttää kokonaisuus selvemältä (kuva 7).

	A ^B _C Alue	A ^B _C Opiskelijoita	A ^B _C Englanti	A ^B _C Englanti, %	A ^B _C Ruotsi	A ^B _C Ruotsi, %
1	MK01 Uusimaa	93986	92293	98.2	81022	86.2
2	MK02 Varsinais-Suomi	21872	21826	99.8	20254	92.6
3	MK04 Satakunta	10493	10490	100.0	9878	94.1
4	MK05 Kanta-Häme	7684	7432	96.7	7117	92.6
5	MK06 Pirkanmaa	27008	26995	100.0	25419	94.1
6	MK07 Päijät-Häme	10133	10130	100.0	9688	95.6
7	MK08 Kymenlaakso	7460	2861	38.4	6599	88.5
8	MK09 Etelä-Karjala	5056	2107	41.7	4560	90.2
9	MK10 Etelä-Savo	6823	6819	99.9	6391	93.7
10	MK11 Pohjois-Savo	12010	12010	100.0	11384	94.8
11	MK12 Pohjois-Karjala	8288	7824	94.4	7469	90.1
12	MK13 Keski-Suomi	13293	13290	100.0	12817	96.4
13	MK14 Etelä-Pohjanmaa	9830	9824	99.9	9359	95.2
14	MK15 Pohjanmaa	8821	6946	78.7	4409	50.0
15	MK16 Keski-Pohjanmaa	4248	4244	99.9	4035	95.0
16	MK17 Pohjois-Pohjanmaa	23047	23027	99.9	21990	95.4
17	MK18 Kainuu	3587	3582	99.9	3405	94.9
18	MK19 Lappi	8485	8477	99.9	8222	96.9
19	MK21 Ahvenanmaa	1070	556	52.0	2	0.2

Kuva 7. Muokatut tiedot.

Sarakkeita voidaan myös lisätä muokkaamalla saatavilla olevia tietoja (kuva 8).



Kuva 8. Sarakkeen lisääminen.

Luodaan yksinkertainen sarake, joka laskee prosenttimäärien erotuksen kuvassa 9 olevan kaavan mukaisesti.

Custom Column

Add a column that is computed from the other columns.

New column name

Eroavaisuus %

Custom column formula ⓘ

= [#"Englanti, %"]-[#"Ruotsi, %"]

Available columns

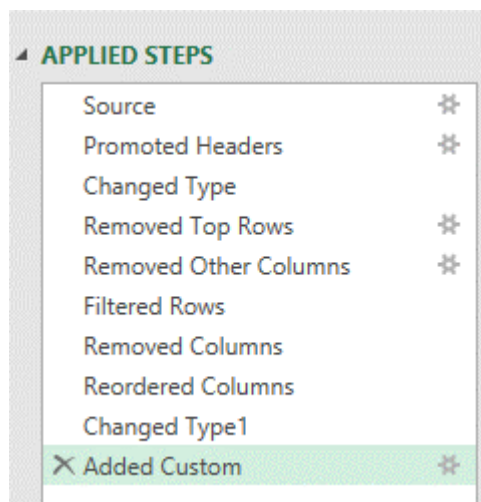
- Alue
- Opiskelijoita
- Englanti
- Englanti, %
- Ruotsi
- Ruotsi, %**

<< Insert

[Learn about Power Query formulas](#)

Kuva 9. Sarakkeen lisääminen, toinen vaihe.

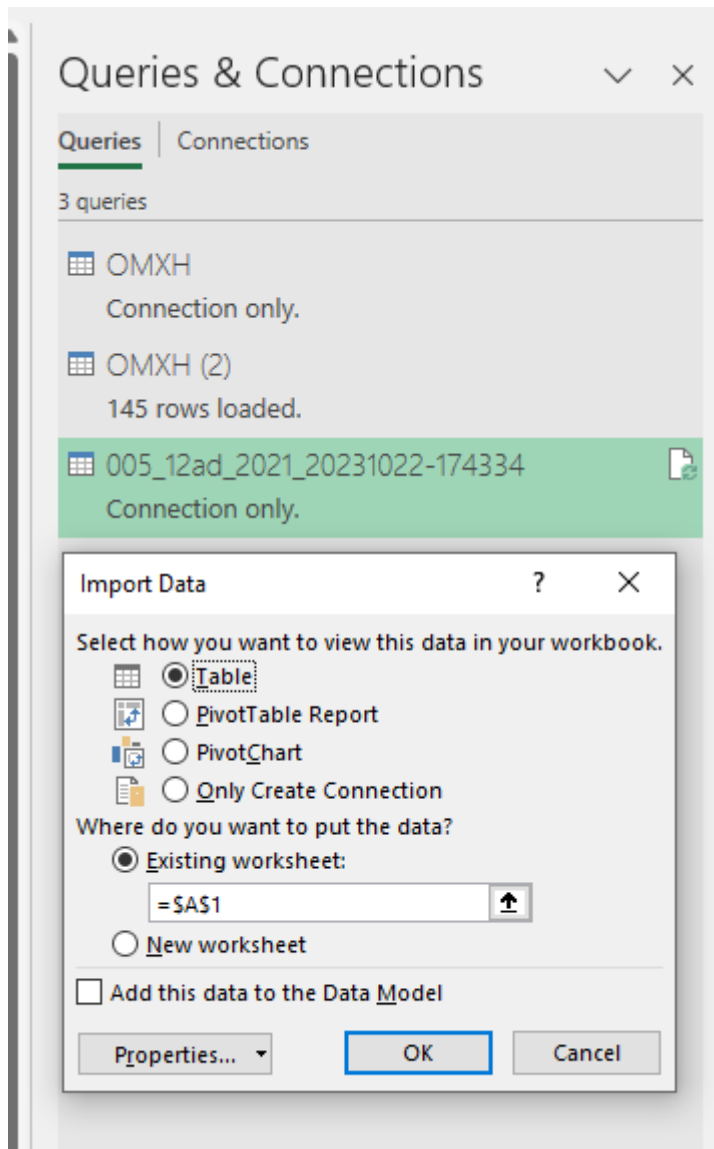
Nyt tietojen valinta on valmis ja Power Query on tallentanut kaikki vaiheet (kuva 10).



Kuva 10. Tallennetut vaiheet.

Vaiheiden tallennus on hyödyllinen siksi, että siihen voidaan myöhemmin tehdä muutoksia lisäämällä tietoa tai mikäli tiedot päivitetään, voidaan tuoda päivitettyt tiedot suoraan säilyttämällä samat vaiheet ja sama muotoilu.

Suljetaan Power Query editor "Close & Load" näppäintä käyttäen.



Kuva 11. Taulukon tuominen.

Tehdyt kyselyt (query) näkyvät Excelissä oikealla kuvan 11 mukaisessa ikkunassa, ja hiiren oikealla näppäimellä voidaan tuoda tiedot asiakirjalle.

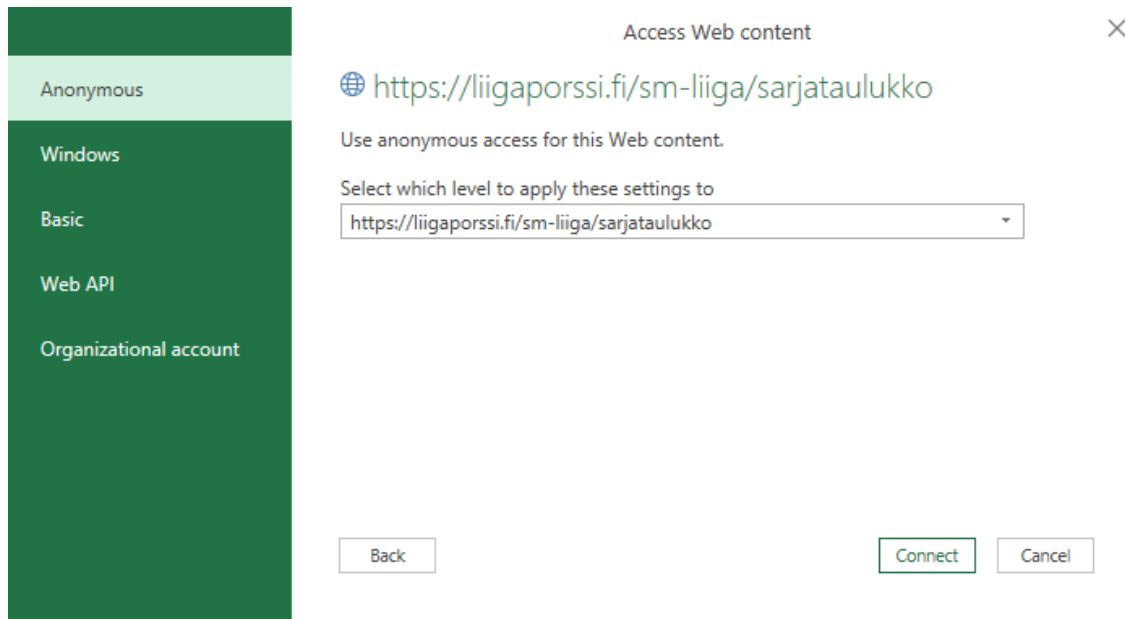
Alue	Opiskelijoita	Englanti	Englanti, %	Ruotsi	Ruotsi, %	Eroavaisuus %
MK01 Uusimaa	93986	92293	98.2	81022	86.2	12
MK02 Varsinais-Suomi	21872	21826	99.8	20254	92.6	7.2
MK04 Satakunta	10493	10490	100	9878	94.1	5.9
MK05 Kanta-Häme	7684	7432	96.7	7117	92.6	4.1
MK06 Pirkanmaa	27008	26995	100	25419	94.1	5.9
MK07 Päijät-Häme	10133	10130	100	9688	95.6	4.4
MK08 Kymenlaakso	7460	2861	38.4	6599	88.5	-50.1
MK09 Etelä-Karjala	5056	2107	41.7	4560	90.2	-48.5
MK10 Etelä-Savo	6823	6819	99.9	6391	93.7	6.2
MK11 Pohjois-Savo	12010	12010	100	11384	94.8	5.2
MK12 Pohjois-Karjala	8288	7824	94.4	7469	90.1	4.3
MK13 Keski-Suomi	13293	13290	100	12817	96.4	3.6
MK14 Etelä-Pohjanmaa	9830	9824	99.9	9359	95.2	4.7
MK15 Pohjanmaa	8821	6946	78.7	4409	50	28.7
MK16 Keski-Pohjanmaa	4248	4244	99.9	4035	95	4.9
MK17 Pohjois-Pohjanmaa	23047	23027	99.9	21990	95.4	4.5
MK18 Kainuu	3587	3582	99.9	3405	94.9	5
MK19 Lappi	8485	8477	99.9	8222	96.9	3
MK21 Ahvenanmaa	1070	556	52	2	0.2	51.8

Kuva 12. Lopullinen taulukko.

PowerQueryn avulla muotoillussa taulukossa on ainoastaan tarvittavat sekä vaiheiden mukaan lisätyt tiedot (kuva 12).

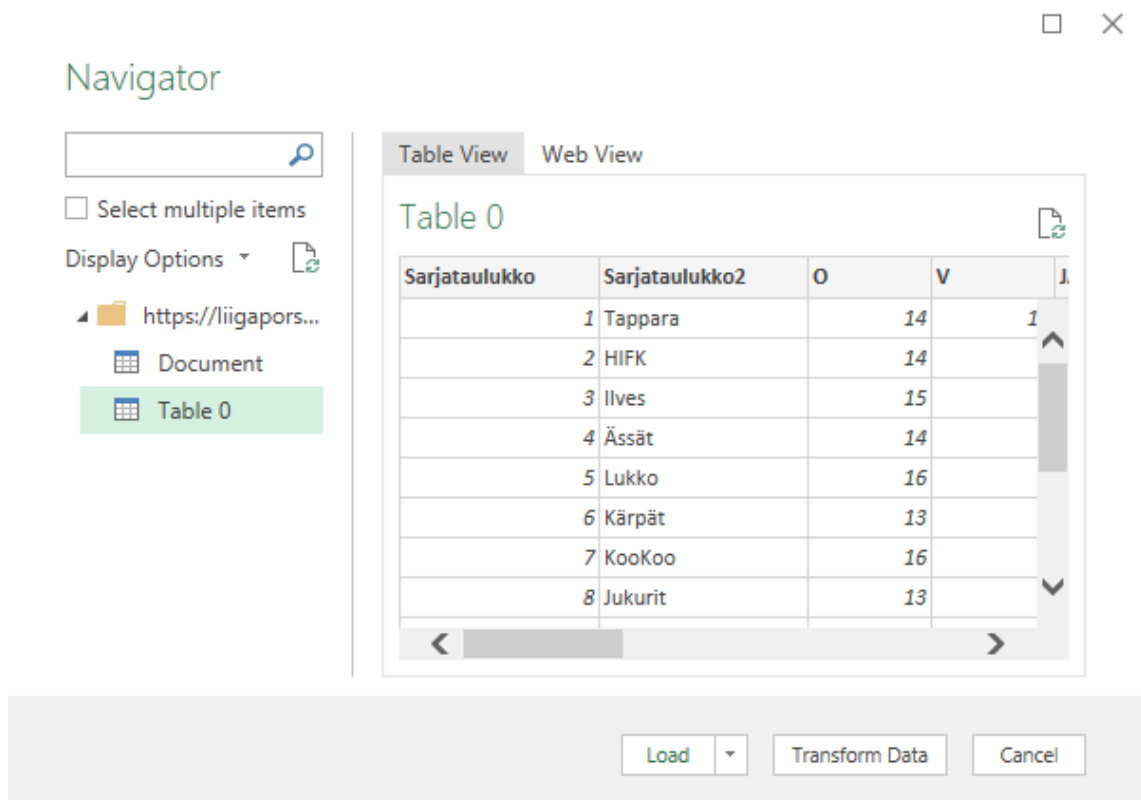
3.2 Tiedon tuonti verkosta

Verkosta tiedon tuominen voi olla myös hyödyllistä ja sitä voi joskus hyödyntää luovasti. Tässä esimerkissä käytetään SM-liigan sarjataulukkoa (Liigapörssi 2023). Data tuodaan samalla tavalla kuin aiemmin, mutta valitaan lähteeksi From Web. Linkki liitetään avautuvaan laatikkoon ja painetaan Connect kuvan 13 mukaisesti.



Kuva 13. Verkkosisällön hakeminen.

Koska tiedot ovat sivulla jo asetettu taulukkoon, löytää Excel sen automaattisesti (kuva 14).



Kuva 14. Sarjataulukko Excelin verkkosisältönavigaattorissa.

Painamalla Transform Data voidaan tuoda tiedot PowerQueryyn ja muokata niitä tai painamalla Load, Excel tuo tiedot suoraan asiakirjalle. Sarjataulukko on tässä hyvä esimerkki siksi, koska tiedot muuttuvat aina kun joukkueet pelaavat pelejä. Excelissä voi päivittää tiedot joko manuaalisesti tai automaattisesti ja kaikki PowerQueryssä tehdyt muutokset säilyvät.

4 Analytiikkaa funktioilla

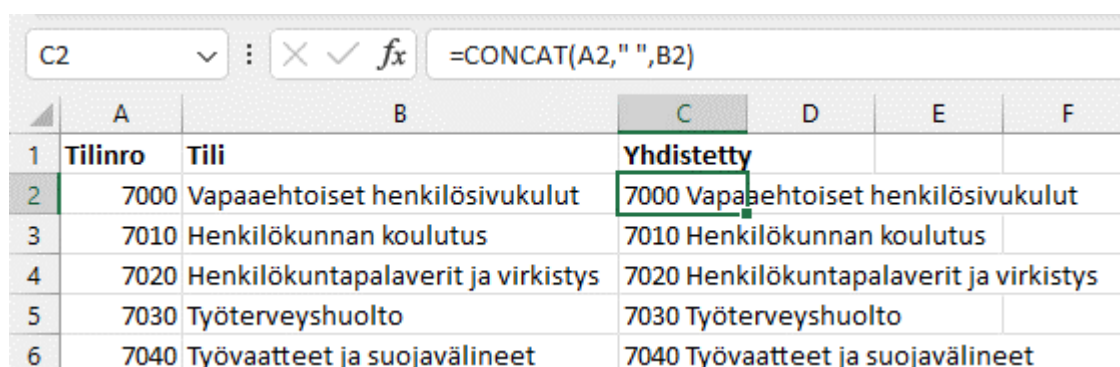
Dataa voidaan käsitellä Excelissä tehokkaasti funktioita ja kaavoja käyttäen. Useimmat funktiot tekevät jonkin matemaattisen, tilastollisen tai loogisen operaation yksinkertaisesti käyttäjän antamista argumenteista: usein valituista soluista tai annetuista arvoista. Niitä on helppo soveltaa, kun on ensin harjoitellut jonkin funktion käyttöä. Funktioita voidaan myös tarvittaessa yhdistää toisiinsa sekä myös matemaattisiin kaavoihin. Olen valinnut tähän kappaleeseen hyödyllisimmät funktiot datan käsittelyyn. Suomenkielisessä Excel-versiossa funktioiden argumentit erotetaan puolipisteellä ; pilkun sijaan.

4.1 Tekstin käsittely ja merkkijonot

Usein data sisältää sanoja, nimityksiä tai muita merkkijonoja. Merkkijonoja voidaan käsitellä usealla eri funktiolla ja soveltaa koko tietosarjaan, jotta välttyään manuaaliselta muokkaukselta.

CONCAT

CONCAT eli YHDISTÄ-funktio yhdistää useamman solun merkkijonot yhteen soluun.



	A	B	C	D	E	F
1	Tilinro	Tili	Yhdistetty			
2	7000	Vapaaehtoiset henkilösivukulut	7000 Vapaaehtoiset henkilösivukulut			
3	7010	Henkilökunnan koulutus	7010 Henkilökunnan koulutus			
4	7020	Henkilökuntapalaverit ja virkistys	7020 Henkilökuntapalaverit ja virkistys			
5	7030	Työterveyshuolto	7030 Työterveyshuolto			
6	7040	Työvaatteet ja suojavälineet	7040 Työvaatteet ja suojavälineet			

Kuva 15. Esimerkki CONCAT-funktiosta.

Esimerkissä (kuva 15) yhdistetään tilinumero ja tilin nimi yhdeksi tekstikentäksi yhteen soluun. Usein kirjanpidon raportit tulevat sovelluksista erotetussa

muodossa, mutta yhdistettyä solua on helpompi käyttää esimerkiksi kaavion luomisessa ja lukemisessa.

Excel 2016 ja uudemmissa versioissa on myös TEXTJOIN-funktio, joka toimii samoin, mutta se hyväksyy argumentiksi myös erottimeen. Kuvassa 16 on erottimeksi asetettu pilkku ja välilyönti, jotta funktio luo nimikkeistä luettelon.

The screenshot shows the Excel formula bar with the formula `=TEXTJOIN(", ",,A2:A5)`. Below the formula bar is a table with columns A through E and rows 1 through 5.

	A	B	C	D	E
1	Lista	Luettelo			
2	Appelsiini	Appelsiin, Banaani, Omena, Päärynä			
3	Banaani				
4	Omena				
5	Päärynä				

Kuva 16. Esimerkki TEXTJOIN-funktiosta.

LEFT, RIGHT ja MID

LEFT ja RIGHT eli VASEN ja OIKEA-funktiot palauttavat vasemmalta tai oikealta luettuna ensimmäiset merkit argumentiksi annetun luvun mukaan.

The screenshot shows the Excel formula bar with the formula `=LEFT(A2,4)`. Below the formula bar is a table with columns A and B and rows 1 through 6.

	A	B
1	Yhdistetty	Tilinumero
2	7000 Vapaaehtoiset henkilösivukulut	7000
3	7010 Henkilökunnan koulutus	7010
4	7020 Henkilökuntapalaverit ja virkistys	7020
5	7030 Työterveyshuolto	7030
6	7040 Työvaatteet ja suojavälineet	7040

Kuva 17. Esimerkki LEFT-funktiosta.

Esimerkissä (kuva 17) funktiolla palautetaan 4 ensimmäistä merkkiä. Funktio RIGHT palauttaisi solusta A2 tiedon "ulut".

MID eli POIMI.TEKSTI-funktio ottaa vastaan kaksi argumenttia, aloituspaikan sekä kirjainten määrän. Funktio palauttaa aloituspaikasta seuraavat kirjaimet annetun luvun mukaan (kuva 18).

	A	B
1	Teksti	MID-erotettu teksti
2	Microsoft Excel 2016	Excel
3		

Kuva 18. Esimerkki MID-funktiosta.

Aloituspaikaksi kuvassa 18 on annettu 11. kirjain eli E ja pituudeksi 5 merkkiä.

LEN

LEN eli PITUUS-funktio palauttaa merkkijonon pituuden lukuna (kuva 19).

	A	B	C	D
1	Teksti	Tekstin pituus		
2	Excel	5		
3				

Kuva 19. Esimerkki LEN-funktiosta.

FIND

FIND eli ETSI-funktio hakee merkkijonosta argumentiksi annetun tekstin ensimmäisen esiintymisen ja palauttaa sen sijainnin lukuna (kuva 20). Funktio hyväksyy myös vapaaehtoisen argumentin aloituspaikan asettamista varten.



	A	B
1	Teksti	
2	AAABAAAA	4
3	ABAAAAAAB	2
4		

Kuva 20. Esimerkki FIND-funktiosta.

SEARCH eli HAKU-funktio toimii täysin samoin, mutta ottaa myös huomioon kirjainkoon.

Funktioiden yhdistäminen

Suurimman hyödyn saa merkkijonojen käsittelyyn yhdistämällä funktioiden toimintoja käyttäen niitä sisäkkäin eli käyttäen yhtä tai useampaa funktiota jonkin funktion argumenttinä.

B2		:			<i>fx</i>	=RIGHT(A2,LEN(A2)-FIND(" ",A2))				
	A	B	C	D	E	F				
1	Nimi	Etunimi								
2	Keinonen, Pertti	Pertti								
3	Näsä, Matti	Matti								
4	Silakka, Timo	Timo								

Kuva 21. Esimerkki merkkijonofunktioiden yhdistämisestä.

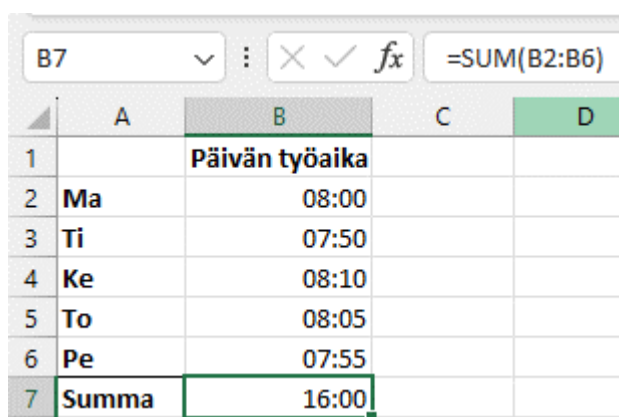
Kuvan 21 funktio palauttaa nimistä etunimet ilman ”kovakoodausta” ja vaikka nimien pituudet vaihtelevat. Funktio toimii kuvion 2 mukaisesti:

=RIGHT(A2,LEN(A2)-FIND(" ",A2))	
↳	=RIGHT(A2,16-FIND(" ",A2))
↳	=RIGHT(A2,16-10)
↳	=RIGHT(A2,6)
↳	Pertti

Kuvio 2. Sisäkkäisen funktion toiminta.

4.2 Aika ja aikafunktiot

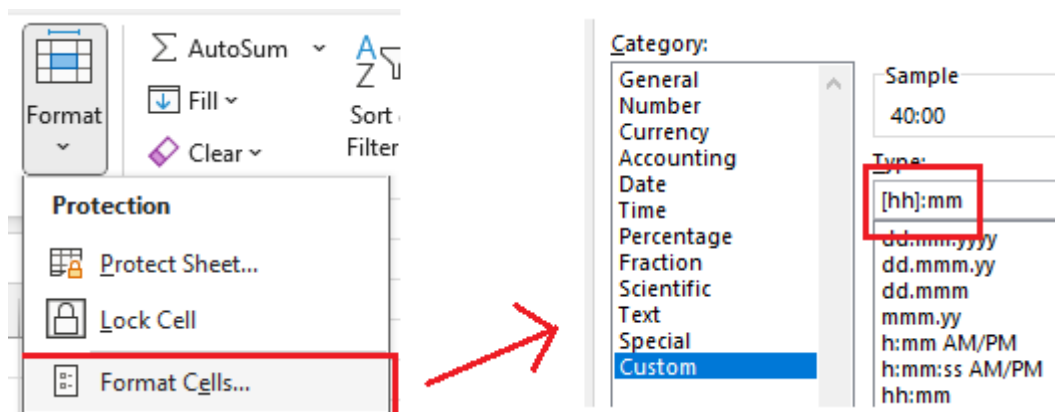
Aika toimii Excelissä hieman omituisesti, ja tästä syystä on hyvä tietää pari asiaa aikatiedon muotoilusta. Vaikka Excel näyttää esim. 1.1.2023, sen oikea arvo on 44927. Tämä johtuu siitä, että Excel laskee päivät 00.01.1900 alkaen kokonaislukuna ja tunnit, minuutit ja sekunnit sen desimaaleina. Mikäli syötetään vain aika, esim. 08:00, ohjelman sisällä se on "00.01.1900 08:00" sekä 0,33. Mikäli syötetään 25:00, se on "01.01.1900 01:00" sekä 1,04. Tämä sekavuus aiheuttaa silloin tällöin ongelmia, mikäli halutaan tehdä laskutoimituksia aikaa käyttäen.



	A	B	C	D
1		Päivän työaika		
2	Ma	08:00		
3	Ti	07:50		
4	Ke	08:10		
5	To	08:05		
6	Pe	07:55		
7	Summa	16:00		

Kuva 22. Esimerkki virheellisestä muotoilusta.

Sisäisesti kaikki toimii hyvin, eli arvot summataan, mutta koska summattu arvo on yli 24 tuntia, ensimmäiset 24 tuntia katoavat ja Excel ilmoittaa seuraavan "päivän" kuluneet tunnit (kuva 22).



Kuva 23. Aikasolun muotoilun muutos.

Muotoilu täytyy muuttaa lisäämällä hakasulkeet tuntimuotoilun hh ympärille kuvan 23 mukaisesti. Tämän jälkeen solu näyttää summatut tunnit oikein, 40:00.

Laskutoimituksissa omituisuus tulee myös esille.

C2

⌵

:

✕

✓

fx

=B2+2

	A	B	C
1		Päivän työaika	Uusi työaika
2	Ma	08:00	56:00

Kuva 24. Esimerkki virheellisestä laskutoimituksesta.

Summaksi tulee 56 (kuva 24), koska Excel tulkitsee kokonaisluvun 2 kahdeksi päiväksi eli 48 tunniksi. Excel ei myöskään hyväksy yhteenlaskua syntaksilla =B2+02:00.

C2

fx

=B2+2/24

	A	B	C
1		Päivän työaika	Uusi työaika
2	Ma	08:00	10:00

Kuva 25. Virheellisen laskutoimituksen korjaus.

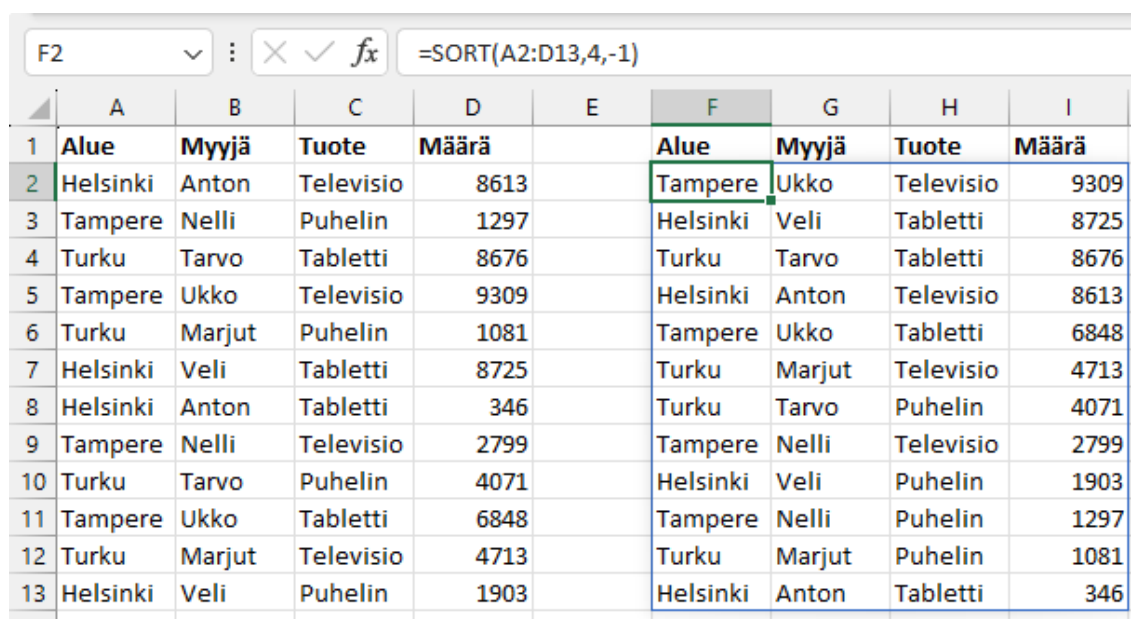
Tästä syystä täytyy laskutoimitus kirjoittaa muodossa $=B2+2/24$. Näin saadaan oikea lopputulos 10:00 (kuva 25). Sama pätee myös toisin päin, eli mikäli halutaan ilmoittaa aika desimaalilukuna, kerrotaan aika luvulla 24. Esim. $08:30 \cdot 24 = 8,5$.

4.3 Lajittelu ja suodatus

Funktioilla voidaan myös lajitella ja suodattaa tietoa. Saman voi toteuttaa myös muuttamalla tiedon taulukoksi ja käyttämällä taulukkotyökaluun rakennettuja toimintoja, mutta funktioiden avulla voidaan pitää alkuperäinen data alkuperäisessä muodossa.

SORT

SORT eli LAJITTELE-funktio lajittelee taulukon valitun sarakkeen tai rivin mukaisesti.



	A	B	C	D	E	F	G	H	I
1	Alue	Myyjä	Tuote	Määrä		Alue	Myyjä	Tuote	Määrä
2	Helsinki	Anton	Televisio	8613		Tampere	Ukko	Televisio	9309
3	Tampere	Nelli	Puhelin	1297		Helsinki	Veli	Tabletti	8725
4	Turku	Tarvo	Tabletti	8676		Turku	Tarvo	Tabletti	8676
5	Tampere	Ukko	Televisio	9309		Helsinki	Anton	Televisio	8613
6	Turku	Marjut	Puhelin	1081		Tampere	Ukko	Tabletti	6848
7	Helsinki	Veli	Tabletti	8725		Turku	Marjut	Televisio	4713
8	Helsinki	Anton	Tabletti	346		Turku	Tarvo	Puhelin	4071
9	Tampere	Nelli	Televisio	2799		Tampere	Nelli	Televisio	2799
10	Turku	Tarvo	Puhelin	4071		Helsinki	Veli	Puhelin	1903
11	Tampere	Ukko	Tabletti	6848		Tampere	Nelli	Puhelin	1297
12	Turku	Marjut	Televisio	4713		Turku	Marjut	Puhelin	1081
13	Helsinki	Veli	Puhelin	1903		Helsinki	Anton	Tabletti	346

Kuva 26. Esimerkki SORT-funktiosta.

Kuvan 26 esimerkissä funktion argumenteiksi ollaan valittu taulukko ilman otsikkoriviä. "4" viittaa neljänteen sarakkeeseen "Määrä" ja "-1" viittaa lajittelun suuntaan eli laskevaan.

SORTBY

SORTBY eli LAJITTELE.ARVOJEN.PERUSTEELLA-funktio toimii samoin kuin SORT, mutta funktio sallii argumenteiksi useamman lajitteluperusteen.

FV : ✖ ✔ fx =SORTBY(A2:D13,A2:A13,-1,D2:D13,-1)									
	A	B	C	D	E	F	G	H	I
1	Alue	Myyjä	Tuote	Määrä		Alue	Myyjä	Tuote	Määrä
2	Helsinki	Anton	Televisio	8613		-1)	Tarvo	Tabletti	8676
3	Tampere	Nelli	Puhelin	1297		Turku	Marjut	Televisio	4713
4	Turku	Tarvo	Tabletti	8676		Turku	Tarvo	Puhelin	4071
5	Tampere	Ukko	Televisio	9309		Turku	Marjut	Puhelin	1081
6	Turku	Marjut	Puhelin	1081		Tampere	Ukko	Televisio	9309
7	Helsinki	Veli	Tabletti	8725		Tampere	Ukko	Tabletti	6848
8	Helsinki	Anton	Tabletti	346		Tampere	Nelli	Televisio	2799
9	Tampere	Nelli	Televisio	2799		Tampere	Nelli	Puhelin	1297
10	Turku	Tarvo	Puhelin	4071		Helsinki	Veli	Tabletti	8725
11	Tampere	Ukko	Tabletti	6848		Helsinki	Anton	Televisio	8613
12	Turku	Marjut	Televisio	4713		Helsinki	Veli	Puhelin	1903
13	Helsinki	Veli	Puhelin	1903		Helsinki	Anton	Tabletti	346

Kuva 27. Esimerkki SORTBY-funktiosta.

Kuvan 27 esimerkissä funktioon on valittu koko taulukko, sitten ensimmäiseksi lajitteluperusteeksi "Alue" (sarake A, punaisella) ja toiseksi lajitteluperusteeksi "Määrä" (sarake D, liilana).

FILTER

FILTER eli SUODATA-funktio suodattaa taulukosta tiedot annetun argumentin mukaisesti.

FV : ✖ ✔ fx =FILTER(A2:D13,D2:D13>5000)									
	A	B	C	D	E	F	G	H	I
1	Alue	Myyjä	Tuote	Määrä		Alue	Myyjä	Tuote	Määrä
2	Helsinki	Anton	Televisio	8613		Turku	Tarvo	Tabletti	8676
3	Tampere	Nelli	Puhelin	1297		Turku	Marjut	Puhelin	1081
4	Turku	Tarvo	Tabletti	8676		Turku	Tarvo	Puhelin	4071
5	Tampere	Ukko	Televisio	9309		Turku	Marjut	Televisio	4713
6	Turku	Marjut	Puhelin	1081					
7	Helsinki	Veli	Tabletti	8725		Alue	Myyjä	Tuote	Määrä
8	Helsinki	Anton	Tabletti	346)	Anton	Televisio	8613
9	Tampere	Nelli	Televisio	2799		Turku	Tarvo	Tabletti	8676
10	Turku	Tarvo	Puhelin	4071		Tampere	Ukko	Televisio	9309
11	Tampere	Ukko	Tabletti	6848		Helsinki	Veli	Tabletti	8725
12	Turku	Marjut	Televisio	4713		Tampere	Ukko	Tabletti	6848
13	Helsinki	Veli	Puhelin	1903					

Kuva 28. Esimerkki FILTER-funktiosta.

Kuvan 28 funktio suodattaa taulukosta rivit, joissa sarakkeen D arvo on yli 5000. Ylemmässä esimerkissä on suodatettu taulukosta rivit, joissa sarakkeen A arvo on "Turku".

Funktioiden yhdistäminen

Näiden funktioiden yhdistäminen on yksinkertaista ja tehokasta tiedon lajitteluun.

FV : ✖ ✔ fx =SORT(FILTER(A2:D13,C2:C13="Tabletti"),4,-1)									
	A	B	C	D	E	F	G	H	I
1	Alue	Myyjä	Tuote	Määrä		Alue	Myyjä	Tuote	Määrä
2	Helsinki	Anton	Televisio	8613)	Veli	Tabletti	8725
3	Tampere	Nelli	Puhelin	1297		Turku	Tarvo	Tabletti	8676
4	Turku	Tarvo	Tabletti	8676		Tampere	Ukko	Tabletti	6848
5	Tampere	Ukko	Televisio	9309		Helsinki	Anton	Tabletti	346
6	Turku	Marjut	Puhelin	1081					
7	Helsinki	Veli	Tabletti	8725					
8	Helsinki	Anton	Tabletti	346					
9	Tampere	Nelli	Televisio	2799					
10	Turku	Tarvo	Puhelin	4071					
11	Tampere	Ukko	Tabletti	6848					
12	Turku	Marjut	Televisio	4713					
13	Helsinki	Veli	Puhelin	1903					

Kuva 29. Esimerkki SORT- ja FILTER-funktioiden yhdistämisestä.

Kuvassa 29 SORT-funktion ensimmäiseksi argumentiksi annetaan FILTER-funktio, joka tuottaa lajiteltavaksi ainoastaan rivit, joissa sarakkeessa C on arvo "Tabletti". SORT-funktio sitten lajittelee ne suuruusjärjestykseen.

4.4 Datan tarkastelu

Taulukkkodataa voidaan tarkastella moneen eri tarkoitukseen seuraavilla funktioilla. Funktiot ovat COUNTIFS, SUMIFS ja AVERAGEIFS eli LASKE.JOS.JOUKKO, SUMMA.JOS.JOUKKO ja KESKIARVO.JOS.JOUKKO. Nämä funktiot laskevat, summaavat tai laskevat keskiarvon datasta jonkin kriteerin tai useamman kriteerin perusteella.

	A	B	C	D	E
1	TUOTEKOODI	TUOTEKUVAUS1	TUOTEKUVAUS2	VARASTOINTI	HINTA ALV 0%
38	8334950	KULMAYHDE AIRCO,PAROC BEND 90 20 V			50.10
39	3110342	VERKKOMATTO PAR,PAROC PRO WMAT V			48.63
40	3110343	VERKKOMATTO PAR,PAROC PRO WMAT E			63.50
41	3110345	VERKKOMATTO PAR,PAROC PRO WMAT E			53.50
42	3110347	VERKKOMATTO PAR,PRO W.MAT 680 ALE			91.25
43	3110348	VERKKOMATTO PAR,PRO W.MAT 680 ALV			55.34
44	3110349	VERKKOMATTO PAR,PRO W.MAT 680 ALE			70.54
45	3110352	VERKKOMATTO PAR,PAROC W.MAT 680 E			60.54
46	3110381	VERKKOMATTO PAR,PAROC PRO WMAT V			73.02
47	3110382	VERKKOMATTO PAR,PRO W.MAT 680 ALV			79.81
48	3110389	PAROC Pro Loose M 50X100X5000 V			26.71
49	3121450	PUTKIKÄYRÄERISTE PAROC BEND ALUC E			21.44
50	3121451	PUTKIKÄYRÄERISTE PAROC BEND ALUC E			31.59
51	3121470	PUTKIKÄYRÄERISTE PAROC BEND ALUC E			25.99
52	3121471	PUTKIKÄYRÄERISTE PAROC BEND ALUC E			31.67
Esimerkki:					
1	Varastossa määrä:			14372	=COUNTIFS(D2:D30679,"V")
2	Varastossa sekä hinta yli 3000 €, määrä:			209	=COUNTIFS(D2:D30679,"V",E2:E30679,">3000")
3	Varastossa olevat Uponor-tuotteet, summa:			138 816.78 €	=SUMIFS(E2:E30679,D2:D30679,"V",C2:C30679,"*uponor*")

Kuva 30. Esimerkit COUNTIFS ja SUMIFS-funktioista.

Kuvan 30 esimerkissä käytetään tietona Ahlsell LV- ja ilmastointituotteiden ohjehinnastoja (Ahlsell 2023). Esimerkissä 1 funktio laskee varastossa olevien tuotenimikkeiden määrän. Esimerkissä 2 funktio laskee lisäksi ainoastaan ne, joiden hinta on sarakkeen E mukaan yli 3000. Tässä funktion "IFS" viittaa useampaan annettuun kriteeriin. Esimerkissä 3 funktio summaa sarakkeen E arvot, jotka ovat sarakkeen D mukaan varastossa sekä tuotekuvauksesta löytyy

teksti "uponor". Tekstin ympärillä olevat asteriskit toimivat tässä esimerkissä jokerimerkkeinä eli funktio palauttaa jokaisen ilmentymän, mikä sisältää tekstin.

Jokaisesta näistä funktioista on myös yksittäiset versiot (COUNTIF, SUMIF, AVERAGEIF – LASKE.JOS, SUMMA.JOS, KESKIVARVO.JOS), mutta näille ei ole käytännöllistä hyötyä, sillä ne ottavat vain yhden kriteerin ja -IFS/JOS.JOUKKO-funktiota voi käyttää useamman kriteerin lisäksi myös vain yhdellä kriteerillä.

4.5 Hakufunktiot

Hakufunktiot eli "lookupit" ovat suosituimpia Excel-funktioita. Hakufunktiot palauttavat arvon taulukosta annettujen argumenttien mukaisesti. VLOOKUP hakee taulukosta tiedon riveittäin eli vertikaalisesti. HLOOKUP taas hakee taulukosta tiedon sarakkeittain eli horisontaalisesti. Suomentokset näille funktioille ovat sekavat, sillä VLOOKUP on PHAKU ja HLOOKUP on VHAKU. Nimitykset tulevat sanoista pysty ja vaaka. Sivuhuomautuksena tästä syystä on hyvä tapa käyttää Excelissä englanninkielistä lokalisointia, mikä samalla varmistaa tiedoston toiminnan yhteensopivuuden muiden kanssa.

Excel for Microsoft 365, Excel 2021 ja Excel for Web ja uudemmissa Excel-versioissa (Cheusheva 2023) on kehittyneempi XLOOKUP eli XHAKU, joka toimii todella tehokkaasti ja jota voi soveltaa useassa eri tilanteessa. XLOOKUP-funktio hakee taulukosta tiedon kontekstin perusteella, riippuen siitä, onko argumentiksi annettu sarakkeita vai rivejä.

Kuvan 31 ja 32 esimerkeissä on käytetty tietojoukkona videopelien myyntitilastoja (Greg Smith 2016).

Formula bar: `=XLOOKUP(K2,A2:A16599,B2:B16599)`

	A	B	C	D	E	F	G	H	I	J	K	L
	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	Global_Sales			
1	1	Wii Sports	Wii	2006	Sports	Nintendo	41.49	29.02	82.74		100	B16599)
2	2	Super Mario Bros.	NES	1985	Platform	Nintendo	29.08	3.58	40.24			
3	3	Mario Kart Wii	Wii	2008	Racing	Nintendo	15.85	12.88	35.82			
4	4	Wii Sports Resort	Wii	2009	Sports	Nintendo	15.75	11.01	33			
5	5	Pokemon Red/Pokemon Blue	GB	1996	Role-Play	Nintendo	11.27	8.89	31.37			
6	6	Tetris	GB	1989	Puzzle	Nintendo	23.2	2.26	30.26			
7	7	New Super Mario Bros.	DS	2006	Platform	Nintendo	11.38	9.23	30.01			
8	8	Wii Play	Wii	2006	Misc	Nintendo	14.03	9.2	29.02			
9	9	New Super Mario Bros. Wii	Wii	2009	Platform	Nintendo	14.59	7.06	28.62			
10	10	Duck Hunt	NES	1984	Shooter	Nintendo	26.93	0.63	28.31			

Kuva 31. Esimerkki XLOOKUP-funktiosta.

K	L	M
100	Battlefield 3	

Kuva 32. XLOOKUP-funktion lopputulos.

Kuvan 31 funktio hakee ensimmäiseksi argumentiksi annetun arvon K2 (100) toiseksi annetusta argumentista sarake A. Funktio palauttaa arvon kolmanneksi annetusta argumentista sarake B. Funktio siis hakee sarakkeesta A arvon 100 ja palauttaa siltä riviltä sarakkeen B arvon, tässä esimerkissä "Battlefield 3" (kuva 32).

Formula bar: `=TRANSPOSE(XLOOKUP(K2,B2:B16599,A2:I16599))`

	A	B	C	D	E	F	G	H	I	J	K	L
	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	Global_Sales			
1	1	Wii Sports	Wii	2006	Sports	Nintendo	41.49	29.02	82.74		Metroid	
2	2	Super Mario Bros.	NES	1985	Platform	Nintendo	29.08	3.58	40.24		Rank	I16599))
3	3	Mario Kart Wii	Wii	2008	Racing	Nintendo	15.85	12.88	35.82		Name	Metroid
4	4	Wii Sports Resort	Wii	2009	Sports	Nintendo	15.75	11.01	33		Platform	NES
5	5	Pokemon Red/Pokemon Blue	GB	1996	Role-Play	Nintendo	11.27	8.89	31.37		Year	1986
6	6	Tetris	GB	1989	Puzzle	Nintendo	23.2	2.26	30.26		Genre	Action
7	7	New Super Mario Bros.	DS	2006	Platform	Nintendo	11.38	9.23	30.01		Publisher	Nintendo
8	8	Wii Play	Wii	2006	Misc	Nintendo	14.03	9.2	29.02		NA_Sales	1.33
9	9	New Super Mario Bros. Wii	Wii	2009	Platform	Nintendo	14.59	7.06	28.62		EU_Sales	0.31
10	10	Duck Hunt	NES	1984	Shooter	Nintendo	26.93	0.63	28.31		Global_Sales	2.73

Kuva 33. Esimerkki XLOOKUP-funktion yhdistämisestä toiseen funktioon sekä useamman tiedon palautus.

TRANSPOSE-funktio transponoi eli kääntää tiedot vastakkaiselle akselille.

Kuvan 33 esimerkissä vaakasuuntainen taulukko saadaan sen avulla muutettua

palautettu rivi selkeäksi sarakkeeksi. Esimerkissä myös aiemman esimerkin mukaisesti haetaan taulukosta tieto "Metroid" sarakkeesta B, mutta palautetaan jokaisen (A-I) sarakkeen tiedot. Kolmanneksi argumentiksi on valittu koko taulukko (Liilalla A2:I16599), joten funktio palauttaa kaiken tiedon riviltä, jolla "Metroid" ilmenee.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Rank	Name	Platform	Year	Genre	Publisher	NA Sales	EU Sales	Global Sales			
2	1	Wii Sports	Wii	2006	Sports	Nintendo	41.49	29.02	82.74			Global_Sales
3	2	Super Mario Bros.	NES	1985	Platform	Nintendo	29.08	3.58	40.24		Halo 2	XLOOKUP(L2,
4	3	Mario Kart Wii	Wii	2008	Racing	Nintendo	15.85	12.88	35.82			
5	4	Wii Sports Resort	Wii	2009	Sports	Nintendo	15.75	11.01	33			
6	5	Pokemon Red/Pokemon Blue	GB	1996	Role-Play	Nintendo	11.27	8.89	31.37			
7	6	Tetris	GB	1989	Puzzle	Nintendo	23.2	2.26	30.26			
8	7	New Super Mario Bros.	DS	2006	Platform	Nintendo	11.38	9.23	30.01			
9	8	Wii Play	Wii	2006	Misc	Nintendo	14.03	9.2	29.02			
10	9	New Super Mario Bros. Wii	Wii	2009	Platform	Nintendo	14.59	7.06	28.62			
11	10	Duck Hunt	NES	1984	Shooter	Nintendo	26.93	0.63	28.31			

Kuva 34. Esimerkki kahdensuuntaisesta XLOOKUP-hausta.

K	L	M
	Global_Sales	
Halo 2	8.49	

Kuva 35. Kahdensuuntaisen XLOOKUP-haun lopputulos.

Kuvan 34 esimerkki voi näyttää hieman sekavalta, mutta sisäkkäisen funktion toiminta on yksinkertainen. Sisimmäinen eli jälkimmäinen XLOOKUP-funktio hakee ensin otsikkoriviltä (Vihreällä, A1:I1) annetun argumentin "Global_Sales". Kolmanneksi argumentiksi on annettu taas koko taulukko, joten funktio palauttaa vastaavalta riviltä kaiken tiedon eli jokaisen rivin Global_Sales arvon. Seuraavaksi toimii ensimmäinen eli uloimmainen XLOOKUP-funktio, joka hakee annettua arvoa "Halo 2" sarakkeesta B ja palauttaa arvon sisimmäisen funktion antamasta taulukosta eli Global_Sales tiedoista. Lopputulos on 8.49 (miljoonaa) eli tämän pelin myynti (kuva 35). Vaihtamalla haetut arvot tai kopioimalla funktio useampaan soluun saadaan haettua nopeasti jokin tieto molempia akseleita käyttäen.

4.6 Loogiset funktiot

Loogiset funktiot tarkistavat, onko funktion argumenteissa annettu ehto tosi vai epätosi. Usein niitä käytetään päätöksenteon apuna tarkistamaan, täyttyykö joku kriteeri, joka johtaisi jatkotoimenpiteisiin. Loogisia funktioita usein yhdistetään toisiinsa ja luodaan ehtorakenne, jossa saman funktion sisällä tehdään useampi tarkistus. Niitä voidaan käyttää esimerkiksi poikkeamien löytämiseen.

IF

IF eli JOS-funktio palauttaa arvon TRUE (TOSI) tai FALSE (EPÄTOSI) logiikkatestin perusteella. Logiikkatestin voi funktion rakentaa Excelin syntaksin mukaisesti niin monimutkaiseksi kun käyttäjä haluaa. Yleisimmin logiikkatestissä käytetään matemaattisia operaattoreita.

	A	B	C
1	Määrä	Kynnysarvo	Ylittääkö?
2	300	200	=IF(A2>B2,"Kyllä","Ei")
3			

Kuva 36. Esimerkki yksinkertaisesta IF-lausekkeesta.

Kuvan 36 esimerkissä logiikkatesti on $300 > 200$, joka on totta, joten funktio palauttaa arvon "Kyllä".

	A	B	C
1	Myyjä	Myynti	Komissio
2	Anton	13 000.00 €	=IF(B2>E3,B2*E4,B2*E5)
3	Nelli	45 000.00 €	
4	Tarvo	35 000.00 €	
5	Ukko	23 000.00 €	

	Kynnysarvo	
20 000.00 €		
Palkkio >20t	25%	
Palkkio <20t	15%	

Kuva 37. Sovellus IF-lauseketta käyttäen.

Kuvan 37 funktion logiikkatesti vertaa myyjän myyntiä kynnysarvoon. Mikäli kynnysarvo ylittyy, ansaitsee myyjä suuremman komission myynnistä. Funktio toimii yksinkertaisella logiikalla.

	A	B	C	D	E
1	Myyjä	Myynti	Komissio		
2	Anton	13 000.00 €	1 950.00 €		
3	Nelli	45 000.00 €	11 250.00 €	Kynnysarvo	20 000.00 €
4	Tarvo	35 000.00 €	8 750.00 €	Palkkio >20t	25%
5	Ukko	23 000.00 €	5 750.00 €	Palkkio <20t	15%
6	Marjut	40 000.00 €	10 000.00 €		
7	Veli	15 000.00 €	2 250.00 €		
8	Anton	22 000.00 €	5 500.00 €		
9	Nelli	31 000.00 €	7 750.00 €		
10	Tarvo	19 000.00 €	2 850.00 €		
11	Ukko	25 000.00 €	6 250.00 €		
12	Marjut	37 000.00 €	9 250.00 €		
13	Veli	11 000.00 €	1 650.00 €		

Kuva 38. IF-lausekkeen lopputulos.

Kuvassa 38 nähdään, että kynnysarvon molemmilla puolella olevilla arvoilla (19 000 € ja 22 000 €) komissioissa on suhteellisen suuri ero.

IFS ja sisäkkäinen IF-rakenne

IFS- eli JOS.JOUKKO-funktio toimii samoin kuin IF, mutta funktio ottaa useamman logiikkatestin ja palauttaa niistä ensimmäisen TOSI-ehdon.

FV \downarrow \times \checkmark f_x $=IFS(B3>E3,E5*E3+(B3-E3)*E4,B3<E3,B3*E5)$						
	A	B	C	D	E	F
1	Myyjä	Myynti	Komissio			
2	Anton	13 000.00 €				
3	Nelli	45 000.00 €	E5)	Kynnysarvo	20 000.00 €	
4	Tarvo	35 000.00 €		Palkkio >20t	25%	
5	Ukko	23 000.00 €		Palkkio <20t	15%	
6	Marjut	40 000.00 €				

Kuva 39. Esimerkki IFS-funktiosta.

Kuvassa 39 ensimmäinen IF-rakenne on alleviivattu oranssilla ja toinen liilalla. Ensimmäinen logiikkatesti tarkistaa, ylittääkö myynti kynnysarvon ja tekee tarvittavat laskutoimenpiteet, mikäli funktio palauttaa arvon TOSI. Mikäli se palauttaa EPÄTOSI, funktio etenee seuraavaan IF-rakenteeseen ja tarkistaa, onko myynti alle kynnysarvon ja laskee komission.

	A	B	C	D	E
1	Myyjä	Myynti	Komissio		
2	Anton	13 000.00 €	1 950.00 €		
3	Nelli	45 000.00 €	9 250.00 €	Kynnysarvo	20 000.00 €
4	Tarvo	35 000.00 €	6 750.00 €	Palkkio >20t	25%
5	Ukko	23 000.00 €	3 750.00 €	Palkkio <20t	15%
6	Marjut	40 000.00 €	8 000.00 €		
7	Veli	15 000.00 €	2 250.00 €		
8	Anton	22 000.00 €	3 500.00 €		
9	Nelli	31 000.00 €	5 750.00 €		
10	Tarvo	19 000.00 €	2 850.00 €		
11	Ukko	25 000.00 €	4 250.00 €		
12	Marjut	37 000.00 €	7 250.00 €		
13	Veli	11 000.00 €	1 650.00 €		

Kuva 40. IFS-funktion lopputulos.

Tässä esimerkissä myös lasku on suoritettu tarkemmin, nostaen ainoastaan kynnysarvon ylittävän summan korkeammalle komissioprosentille (kuva 40). IFS-funktio on käytännöllinen, koska tarvittaessa voidaan lisätä useampi kriteeri ja useampi logiikkatesti funktion perälle.

FV : ✖ ✔ <i>fx</i> =IF(B3>E3,E3*E5+(B3-E3)*E4,IF(B3<E3,B3*E5))							
	A	B	C	D	E	F	G
1	Myyjä	Myynti	Komissio				
2	Anton	13 000.00 €					
3	Nelli	45 000.00 €	E5))	Kynnysarvo	20 000.00 €		
4	Tarvo	35 000.00 €		Palkkio >20t	25%		
5	Ukko	23 000.00 €		Palkkio <20t	15%		
6	Marjut	40 000.00 €					

Kuva 41. Esimerkki sisäkkäisestä IF-rakenteesta.

Sisäkkäinen IF-rakenne toimii siten, että kirjoitetaan toinen IF-lauseke ensimmäisen EPÄTOSI-ehdoksi (kuva 41). Tämä rakenne on toimiva ja lyhyissa laskuissa selkeä, mutta kolmen tai useamman IF-lausekkeen jälkeen tulisi vaihtaa IFS-funktioon selkeyden vuoksi.

	A	B	C	D	E
1	Myyjä	Myynti	Komissio		
2	Anton	13 000.00 €	1 950.00 €		
3	Nelli	45 000.00 €	9 250.00 €	Kynnysarvo	20 000.00 €
4	Tarvo	35 000.00 €	6 750.00 €	Palkkio >20t	25%
5	Ukko	23 000.00 €	3 750.00 €	Palkkio <20t	15%
6	Marjut	40 000.00 €	8 000.00 €		
7	Veli	15 000.00 €	2 250.00 €		
8	Anton	22 000.00 €	3 500.00 €		
9	Nelli	31 000.00 €	5 750.00 €		
10	Tarvo	19 000.00 €	2 850.00 €		
11	Ukko	25 000.00 €	4 250.00 €		
12	Marjut	37 000.00 €	7 250.00 €		
13	Veli	11 000.00 €	1 650.00 €		

Kuva 42. Sisäkkäisen IF-rakenteen lopputulos.

Lopputulos on sama kuin IFS-rakenteella.

AND

AND- eli JA-funktio toimii samoin kuin IF, mutta logiikkatestejä annetaan kaksi tai useampi ja funktio palauttaa arvon TOSI ainoastaan, jos kaikki kriteerit täyttyvät (kuva 43).

FV ⌵ ✖ ✓ <i>f_x</i> =AND(B2>G2,C2<G3)								
	A	B	C	D	E	F	G	H
1	Henkilö	Pituus	Paino	Kelvollinen?		Kriteeri		
2	Anton	158	78	C2<G3)		Pituus yli	170	cm
3	Nelli	157	78			Paino alle	85	kg

Kuva 43. Esimerkki AND-funktiosta.

	A	B	C	D	E	F	G	H
1	Henkilö	Pituus	Paino	Kelvollinen?		Kriteeri		
2	Anton	158	78	FALSE		Pituus yli	170 cm	
3	Nelli	157	78	FALSE		Paino alle	85 kg	
4	Tarvo	188	101	FALSE				
5	Ukko	185	83	TRUE				
6	Marjut	153	89	FALSE				
7	Veli	177	73	TRUE				
8	Anton	185	66	TRUE				
9	Nelli	151	112	FALSE				
10	Tarvo	196	93	FALSE				
11	Ukko	187	61	TRUE				
12	Marjut	178	97	FALSE				
13	Veli	153	54	FALSE				

Kuva 44. AND-funktion lopputulos.

Funktio palauttaa ainoastaan ne rivit, jossa pituus yli 170 ja paino alle 85.

OR

OR- eli TAI-funktio toimii samoin kuin AND, mutta funktio palauttaa arvon TOSI, jos mikään annetuista logiikkatesteistä on totta.

	A	B	C	D	E	F	G	H
1	Henkilö	Pituus	Paino	Kelvollinen?		Kriteeri		
2	Anton	158	78	C2<G3)		Pituus yli	170 cm	
3	Nelli	157	78			Paino alle	85 kg	

Kuva 45. Esimerkki OR-funktiosta.

	A	B	C	D	E	F	G	H
1	Henkilö	Pituus	Paino	Kelvollinen?		Kriteeri		
2	Anton	158	78	TRUE		Pituus yli	170 cm	
3	Nelli	157	78	TRUE		Paino alle	85 kg	
4	Tarvo	188	101	TRUE				
5	Ukko	185	83	TRUE				
6	Marjut	153	89	FALSE				
7	Veli	177	73	TRUE				
8	Anton	185	66	TRUE				
9	Nelli	151	112	FALSE				
10	Tarvo	196	93	TRUE				
11	Ukko	187	61	TRUE				
12	Marjut	178	97	TRUE				
13	Veli	153	54	TRUE				

Kuva 46. OR-funktion lopputulos.

Funktio palauttaa ne rivit, jossa pituus yli 170 tai paino alle 85.

4.7 Tilastolliset funktiot

Tilastollisia funktioita on monia, mutta olen valinnut tähän pari helposti sovellettavaa funktiota.

SUMPRODUCT

SUMPRODUCT- eli TULOJEN.SUMMA-funktio laskee rivien tai sarakkeiden tulojen summan.

FV ✖ ✓ f_x =SUMPRODUCT(B2:B5,C2:C5)						
	A	B	C	D	E	F
1	Hedelmä	Kilohinta	Määrä kg			
2	Appelsiini	2.50 €	40			
3	Banaani	1.80 €	60			
4	Omena	2.20 €	35			
5	Päärynä	3.00 €	25			
6	Yhteensä		B5,C2:C5)			
7						

Kuva 47. Esimerkki SUMPRODUCT-funktion toiminnasta.

Funktio kertoo valituista sarakkeista ensin vastaavat rivit ja sitten summaa ne yhteen. Kuvan 47 esimerkissä $B2 \cdot C2 + B3 \cdot C3 \dots$ ja summa on yhteensä 360 euroa (kuva 48).

	A	B	C	D
1	Hedelmä	Kilohinta	Määrä kg	Tilauksia
2	Appelsiini	2.50 €	40	3
3	Banaani	1.80 €	60	2
4	Omena	2.20 €	35	10
5	Päärynä	3.00 €	25	6
6	Yhteensä		360.00 €	1 736.00 €

Kuva 48. SUMPRODUCT-funktion lopputulos.

Lisäämällä uuden sarakkeen ja sisällyttämällä sen mukaan funktioon:
 $B2 \cdot C2 \cdot D2 + B3 \cdot C3 \cdot D3 \dots$ ja summa on yhteensä 1736 euroa (kuva 48).

RANK

RANK- eli ARVON.MUKAAN-funktio laskee valitusta matriisista annetun arvon sijoituksen joko nousevana tai laskevana (kuva 49). Suurin hyöty tästä funktiosta on muiden funktioiden sisällä esimerkiksi jos jostain tietosarjasta halutaan laskea jotakin joka kymmenennen arvon perusteella.

FV				=RANK(B2,B2:B13,0)			
	A	B	C	RANK(number, ref, [order])			
1	Myyjä	Myynti	Sijoitus				
2	Anton	13 000.00 €	=RANK(B2,B2:B13,0)				
3	Nelli	45 000.00 €	1				
4	Tarvo	35 000.00 €	4				
5	Ukko	23 000.00 €	7				
6	Marjut	40 000.00 €	2				
7	Veli	15 000.00 €	10				
8	Anton	22 000.00 €	8				
9	Nelli	31 000.00 €	5				
10	Tarvo	19 000.00 €	9				
11	Ukko	25 000.00 €	6				
12	Marjut	37 000.00 €	3				
13	Veli	11 000.00 €	12				

Kuva 49. Esimerkki RANK-funktiosta.

Muut funktiot

M20										
	A	B	C	D	E	F	G	H	I	J
1	Myyjä	Myynti		Funktio	Tulos	Selitys				
2	Anton	13 000.00 €		=COUNTA(B2:B9)	6	Laskee ei-tyhjät solut. =COUNT() laskee vain numerot				
3				=COUNTBLANK(B2:B9)	2	Laskee tyhjät solut				
4	Nelli	45 000.00 €		=MIN(B2:B9)	13 000.00 €	Palauttaa pienimmän arvon				
5	Tarvo	35 000.00 €		=MAX(B2:B9)	45 000.00 €	Palauttaa suurimman arvon				
6	Ukko	23 000.00 €								
7										
8	Marjut	40 000.00 €								
9	Veli	15 000.00 €								
10										

Kuva 50. Muita hyödyllisiä funktioita.

Kuvan 50 funktioiden toiminta on yksinkertainen, mutta ne on hyvä tietää ja niistä saa suuren hyödyn yhdistämällä muihin funktioihin.

4.8 Yhdistäminen

H3 :

<

Kuva 51. Esimerkki funktioiden yhdistämisestä XLOOKUP-funktiolla.

Kuvan 51 esimerkissä XLOOKUP hakee MAX-funktiolla haetun suurimman arvon hintasarakkeesta ja palauttaa tiedot sarakkeista B ja C. TEXTJOIN-funktio erottaa tiedot pilkulla ja välilyönnillä.

4.9 Muut Excel-työkalut - Optimointi

Excelissä on myös pari helppokäyttöistä työkalua optimointiin. Optimoinnilla tarkoitetaan työkirjan arvojen muuttamista annettujen kriteerien perusteella.

Solver

Solver on Excelin mukana tuleva apuohjelma, joka täytyy erikseen aktivoida asetuksista tai Add-Ins-valikosta. Tämän jälkeen se löytyy Data välilehdeltä.

	A	B	C	D	E	F	G	H	I	J	K
1	Sijoitettu raha	5 000.00 €	5 000.00 €	5 000.00 €	5 000.00 €	5 000.00 €	5 000.00 €	5 000.00 €	5 000.00 €	5 000.00 €	5 000.00 €
2	Indeksirahaston numero	1	2	3	4	5	6	7	8	9	10
3	Vuotuinen tuotto	8.00%	7.40%	6.80%	6.20%	5.60%	5.00%	4.40%	3.80%	3.20%	2.60%
4	Vuosittaiset kulut	5.00%	4.50%	4.00%	3.50%	3.00%	2.50%	2.00%	1.50%	1.00%	0.50%
5	Välittäjän palkkio	3.00%	2.00%	1.60%	1.40%	1.20%	1.00%	0.80%	0.60%	0.40%	0.20%
6											
7	Käytettävä raha	50 000.00 €									
8	Sijoituksen pituus vuosina	10									
9	Sijoitettu raha yhteensä:	50 000.00 €									
10											
11	Rahamäärä lopuksi	62 756.65 €									

Kuva 52. Indeksirahastosovellus ennen optimointia.

Kuvitteellinen sijoittaja sijoittaa rahansa kymmeneen eri rahastoon ja kaava laskee tuoton annettujen vuosien jälkeen. Kuvan 52 esimerkissä sijoittajalla on

rahaa 50 000 € ja sijoituksen pituus on 10 vuotta. Indeksirahastoilla on eri tuotto prosentit, kuluprosentit ja palkkioprosentit, joten paras vaihtoehto ei ole heti ilmiselvä. Ainoa kaava on solussa B11. Tämän saa helposti selvitettyä Solver-työkalulla.

	A	B	C	D	E	F	G	H	I	J	K
1	Sijoitettu raha	5 000.00 €	5 000.00 €	5 000.00 €	5 000.00 €	5 000.00 €	5 000.00 €	5 000.00 €	5 000.00 €	5 000.00 €	5 000.00 €
2	Indeksirahaston numero	1	2	3	4	5	6	7	8	9	10
3	Vuotuinen tuotto	8.00%	7.40%	6.80%	6.20%	5.60%	5.00%	4.40%	3.80%	3.20%	2.60%
4	Vuosittaiset kulut	5.00%	4.50%	4.00%	3.50%	3.00%	2.50%	2.00%	1.50%	1.00%	0.50%
5	Välittäjän palkkio	3.00%	2.00%	1.60%	1.40%	1.20%	1.00%	0.80%	0.60%	0.40%	0.20%
6											
7	Käytettävä raha	50 000.00 €									
8	Sijoituksen pituus vuosina	10									
9	Sijoitettu raha yhteensä:	50 000.00 €									
10											
11	Rahamäärä lopuksi	62 756.65 €									
12											
13											
14											
15											
16											
17											

Solver Parameters

Set Objective:

To: ☒ Max ☐ Min ☐ Value Of:

By Changing Variable Cells:

Subject to the Constraints:

Add

Kuva 53. Esimerkki Solver-toiminnosta.

Laittamalla kohteeksi (Set Objective) kaavasolun B1, muuttuviksi soluiksi sijoitetun rahan solut B1-K1 ja rajoitteeksi, että sijoitettu raha ei saa ylittää käytettyä rahaa, laskee Solver parhaan tilanteen maksimaaliseen tuottoon (kuva 53).

	A	B	C	D	E	F	G	H	I	J	K
1	Sijoitettu raha	0.00 €	50 000.00 €	0.00 €	0.00 €	0.00 €	0.00 €	0.00 €	0.00 €	0.00 €	0.00 €
2	Indeksirahaston numero	1	2	3	4	5	6	7	8	9	10
3	Vuotuinen tuotto	8.00%	7.40%	6.80%	6.20%	5.60%	5.00%	4.40%	3.80%	3.20%	2.60%
4	Vuosittaiset kulut	5.00%	4.50%	4.00%	3.50%	3.00%	2.50%	2.00%	1.50%	1.00%	0.50%
5	Välittäjän palkkio	3.00%	2.00%	1.60%	1.40%	1.20%	1.00%	0.80%	0.60%	0.40%	0.20%
6											
7	Käytettävä raha	50 000.00 €									
8	Sijoituksen pituus vuosina	10									
9	Sijoitettu raha yhteensä:	50 000.00 €									
10											
11	Rahamäärä lopuksi	63 911.04 €									
12											

Solver Results

Solver found a solution. All Constraints and optimality conditions are satisfied.

☒ Keep Solver Solution

Reports

Answer

Sensitivity

Limits

Kuva 54. Solver ratkaisee.

Sijoittamalla kaiken käytössä olevan rahan indeksirahastoon 2 saadaan nostettua tuottoa huomattavasti (kuva 54).

What-If analyysi/Goal Seek

Goal Seek toimii samoin kuin Solver, mutta paljon yksinkertaisemmin ja se pystyy ainoastaan muuttamaan yhtä solua kerrallaan. Työkalu on kuitenkin nopea käyttää ja usein käytännöllinen.

	A	B
1	Tuotteita myyty	2000
2	Tuotteen hinta	4.50 €
3	Kate	9 000.00 €
4		
5	Tuotekustannus	1.80 €
6	Kustannukset yht.	3 600.00 €
7		
8	Tulos	5 400.00 €

Kuva 55. Yksinkertainen katetuottolaskelma.

Kuvan 55 esimerkissä on katetuottolaskenta ja halutaan tietää, paljonko täytyisi myydä tuotteita, jotta tulos olisi 8 000 € (kuva 56).

	A	B	C	D	E	F
1	Tuotteita myyty	2000				
2	Tuotteen hinta	4.50 €				
3	Kate	9 000.00 €				
4						
5	Tuotekustannus	1.80 €				
6	Kustannukset yht.	3 600.00 €				
7						
8	Tulos	5 400.00 €				

Goal Seek

Set cell:

To value:

By changing cell:

Kuva 56. Goal Seek.

Goal Seek muuttaa myytyjen tuotteiden määrää siihen asti, että solujen kaavat laskevat tulokseksi 8 000.

	A	B	C	D	E	F
1	Tuotteita myyty	2962.962963				
2	Tuotteen hinta	4.50 €				
3	Kate	13 333.33 €				
4						
5	Tuotekustannus	1.80 €				
6	Kustannukset yht.	5 333.33 €				
7						
8	Tulos	8 000.00 €				

Goal Seek Status ? X

Goal Seeking with Cell B8 found a solution.

Target value: 8000

Current value: 8 000.00 €

Step

Pause

OK Cancel

Kuva 57. Goal Seek-ratkaisu.

Tuotteita täytyisi myydä 2963 kappaletta, jotta tulos olisi 8 000 € (kuva 57).

5 Graafinen esitys

Kuviot ovat visuaalinen esitysmuoto datasta, jossa tiedot esitetään symboleilla ja kuvion tarkoituksena on havainnollistaa ilmiöitä sekä helpottaa oleellisen tiedon hahmottamista. Kuvion avulla yksinkertaistetaan tieto tuloksiin, jotka on helpompi kommunikoida.

Usein mediassa tai työelämässä voi tulla vastaan huonoja tai vaikealukuisia visualisointeja: miksi näin käy, mikä tekee niistä huonoja ja miten niistä voitaisiin tehdä parempia? Visuaalinen esitys kuvioiden avulla ei ole luonnollista, jos verrataan sitä ideoiden kommunikointiin lauseiden ja numeroiden avulla. Kuitenkin kuvioilla voidaan havainnollistaa jokin ilmiö, kertoa tarina ja luoda tällä tavalla tärkeää informaatiota. (Nussbaumer Knaflig 2015). Kuka tahansa voi laittaa numerot Exceliin ja luoda kuvion, mutta tämä on usein kaukana optimaalisesta tai hyvin luettavasta kuviosta.

Pääkohta on siis tarinankerronta datan visualisoinnin avulla, eikä vain datan näyttäminen visuaalisena esityksenä.

Esitystavan valinta

Esitystavan valinnassa tärkeää on konteksti, jossa tieto esitetään. Tärkeimmät kysymykset ovat kenelle tieto esitetään, mistä syystä ja mikä on se informaatio, joka halutaan kommunikoida. On tärkeää miettiä, mitä halutaan yleisön tietävän tai tekevän visualisoinnin tuloksena. Kontekstipohjainen ajattelu on myös relevantti siinä, suunnitellaanko visualisointi esim. sähköpostin liitteeksi tai yleisön edessä esitettäväksi.

On myös tärkeää valita oleellinen esitetty data, sillä kaikki tutkittu data ei ole kommunikoinnin kannalta relevanttia. Joskus voi kuitenkin päätöksenteon apuna olla hyödyllistä visualisoida myös tietoa, joka ei tue tutkimuksen tulosta (Nussbaumer Knaflig 2015).

5.1 Visualisointitavat

Visualisointitapoja on useita ja niillä on omat vahvuutensa ja käyttötapansa. Visualisoinnin valinnassa kannattaa miettiä, mitä tietoa halutaan painottaa. Aina ei myöskään tarvitse visualisoida tietoa kuvioin, vaan voidaan käyttää esitysmateriaalissa visuaalisesti suurennettuja numeroita tai yksinkertaisia lauseita.



Kuva 58. Suomen työeläkkeensaajat-julkaisusta poimittuja korotuksia (Eläketurvakeskus, 2023).

Suomen työeläkkeensaajat -julkaisusta poimitut korotukset esittävät erilaista tietoa havainnollisesti (kuva 58). Eri luvuille on selvästi asetettu eri painotukset ja tekstistä on haluttu nostaa niiden luettavuutta.

Hyvä visualisointi käyttää hyväksi ihmisten visuaalista tulkintaa; pylväsdiagrammin tapauksessa aivot vertailevat automaattisesti pylväiden pituutta ja huomaavat suurimman ja pienimmän. Hyvästä kuviosta on heti selvää mikä on oleellinen tieto, mitä halutaan kommunikoida ja mikäli halutaan vielä painottaa, voidaan korostaa elementtejä suunnittelun avulla tai yksinkertaisemmin väreillä tai huomautuksilla. Melkein kaikkea tietoa voidaan tehokkaasti kommunikoida vain muutamalla eri visualisointitavalla, ja lukija myös hyötyy siitä, että relevantti tieto löytyy tutun näköisestä kuvioista vain vilkaisemalla.

5.1.1 Taulukko ja lämpökartta

	Mitta A	Mitta B	Mitta C			Mitta A	Mitta B	Mitta C
Ryhmä 1	15%	22%	42%		Ryhmä 1	15%	22%	42%
Ryhmä 2	40%	36%	20%		Ryhmä 2	40%	36%	20%
Ryhmä 3	35%	17%	34%		Ryhmä 3	15%	17%	34%
Ryhmä 4	30%	29%	26%		Ryhmä 4	30%	22%	26%
Ryhmä 5	55%	30%	58%		Ryhmä 5	55%	30%	58%
Ryhmä 6	11%	25%	49%		Ryhmä 6	11%	25%	40%

Taulukko 1. Esimerkki taulukosta ja lämpökartasta

Taulukkoa 1 (vasen) lukiessa lukija joutuu siirtämään keskittymisen oleellisesta numeroiden skannaamiseen. Oikealla värikylläisyyden eli käytännössä lämpökartan avulla kaikki tieto pysyy vielä näkyvillä, mutta pienimmät ja suurimmat löytyvät nopeasti. Lisäksi ryhmien verrattavuus korostuu, esimerkiksi nopealla vilkaisulla huomataan, että Ryhmä 5:n luvut ovat kokonaisuudessaan korkeammat kuin Ryhmä 3:n. Lämpökartta ei myöskään luovu taulukon yksityiskohdista. Mikäli taulukko on suuri, lämpökartta voi kuitenkin lisätä sekavuutta. Excelissä tämä on toteutettu ehdollinen muotoilu -toiminnolla. Tällä toiminnolla voidaan sääntöjen avulla korostaa lukuja monella eri tavalla, mitä voidaan käyttää apuna lukujen kommunikoimiseen. Esimerkiksi suurista taulukoista voidaan värjätä vain pari lukua tai värjätä eri väreillä suurimmat 10% ja pienimmät 10% (taulukko 2).

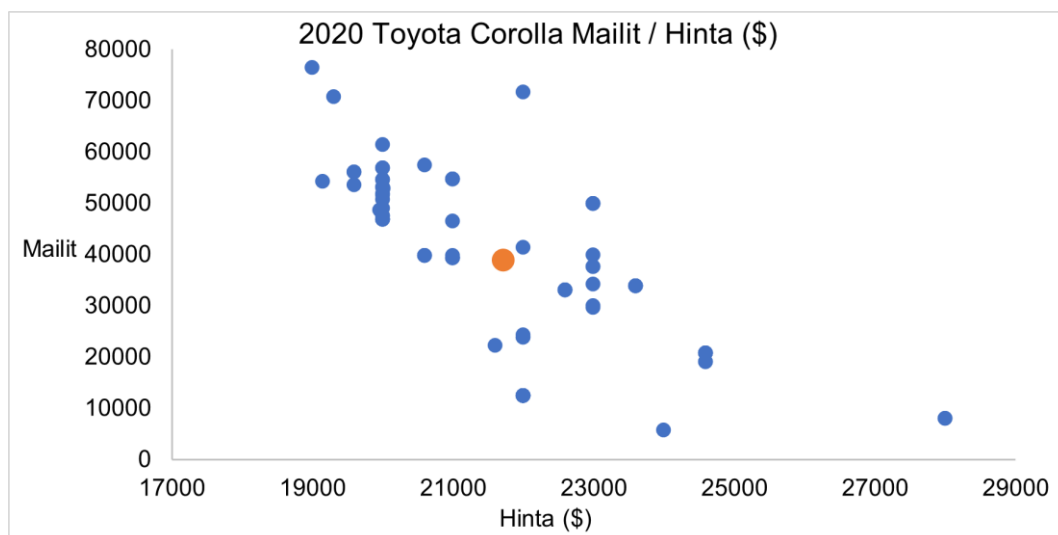
	Mitta A	Mitta B	Mitta C
Ryhmä 1	15%	22%	42%
Ryhmä 2	40%	36%	20%
Ryhmä 3	35%	17%	34%
Ryhmä 4	30%	29%	26%
Ryhmä 5	55%	30%	58%
Ryhmä 6	11%	25%	49%

Taulukko 2. Suurin arvo korostettu jokaisesta mitasta.

5.1.2 Pisteet ja viivat

Usein pidetään piste-, viiva- ja pylväskaavioita keskenään vaihtokelpoisina, sillä jokainen niistä laittaa luvut kahdelle akselille. Eroavaisuus on siinä, mitä halutaan kommunikoida ja visualisoinnilla painottaa. Pisteillä painotetaan korrelaatiota, viivoilla jatkuvuutta ja pylväskaavioilla painotetaan eri kategorioita erottamalla ne eri pylväisiin.

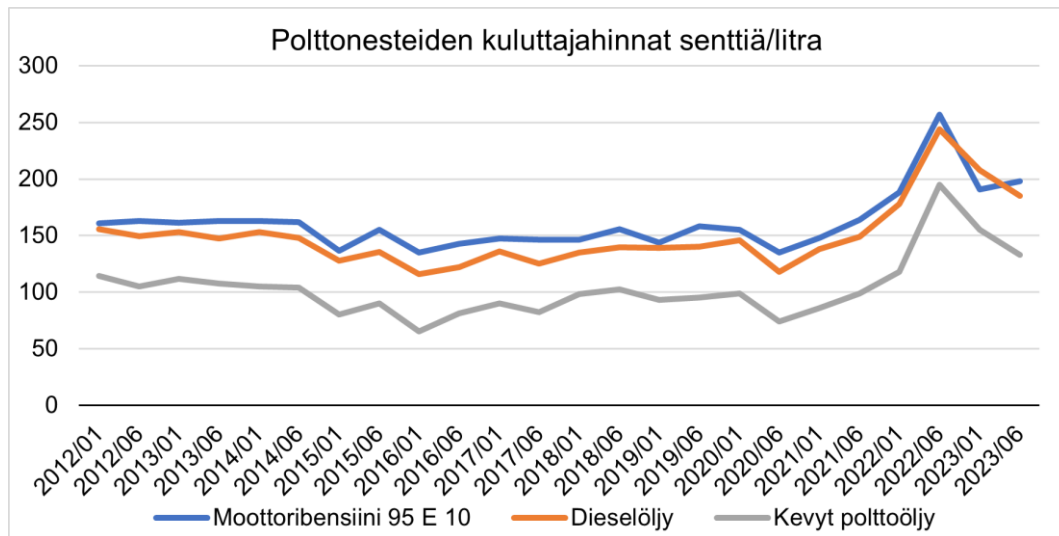
Pisteet



Kuvio 3. Pistekaavio, jossa ajetus mailit Y-akselilla ja auton hinta X-akselilla.

Pistekaaviot ovat hyödyllisiä, kun halutaan osoittaa kahden asian välistä suhdetta. Pisteiden sijoittuminen samanaikaisesti molemmille akseleille näyttää, onko suhde olemassa ja mikä suhde on. Kuvion 3 mukaan arvoilla on selvästi negatiivinen suhde, sillä kun ajetus mailit nousevat, auton arvo markkinoilla laskee. Tietosarjan keskiarvo on myös helppo kommunikoida eri värityksellä, tässä oranssina.

Viivat



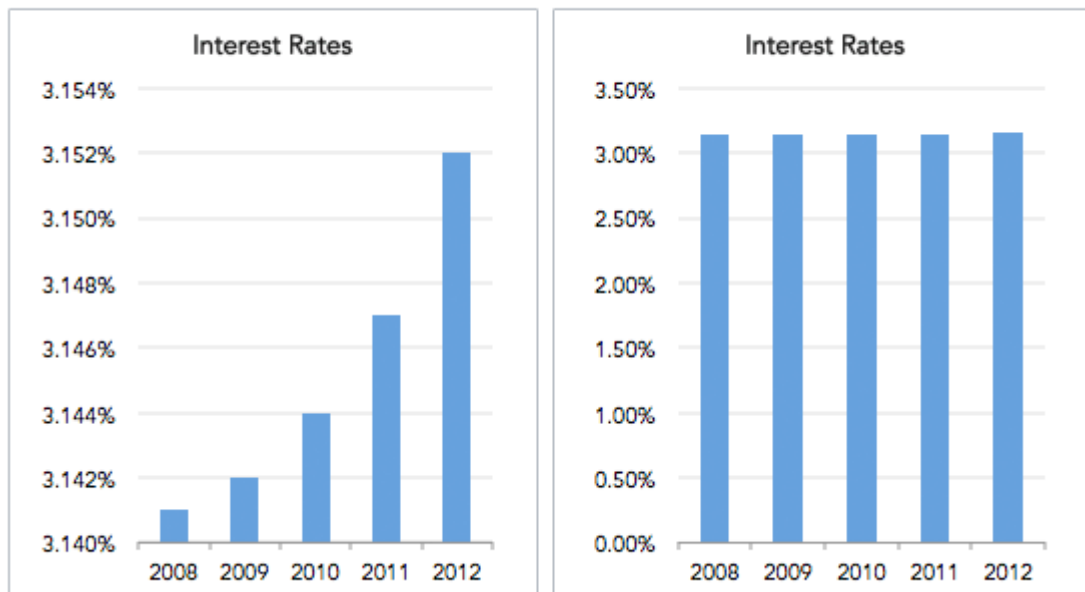
Kuvio 4. Viivakaavio, jossa polttonesteiden hinnat Y-akselilla ja aika X-akselilla.

Viivakaaviota käytetään ajan myötä muuttuvan tiedon näyttämiseen. Kuviossa 4 viiva kommunikoi selkeästi, miten polttonesteiden hinnat kehittyvät n. kuuden kuukauden välein. Lisäksi eriväriset viivat ovat helposti verrattavissa suhteessa toisiinsa.

5.1.3 Pylväät

Usein vältellään pylväskaavioita, koska ne ovat yleisiä. Niiden yleisyys kuitenkin tekee niistä kaikista helposti luettavan ja ymmärrettävän. Silmät huomaavat nopeasti erot pylväiden koossa ja vertaavat pylväitä toisiinsa. Pylväskaaviota rakentaessa kannattaa muistaa painottaa pylväiden pituutta eikä pylväiden käyttämää aluetta.

Same Data, Different Y-Axis

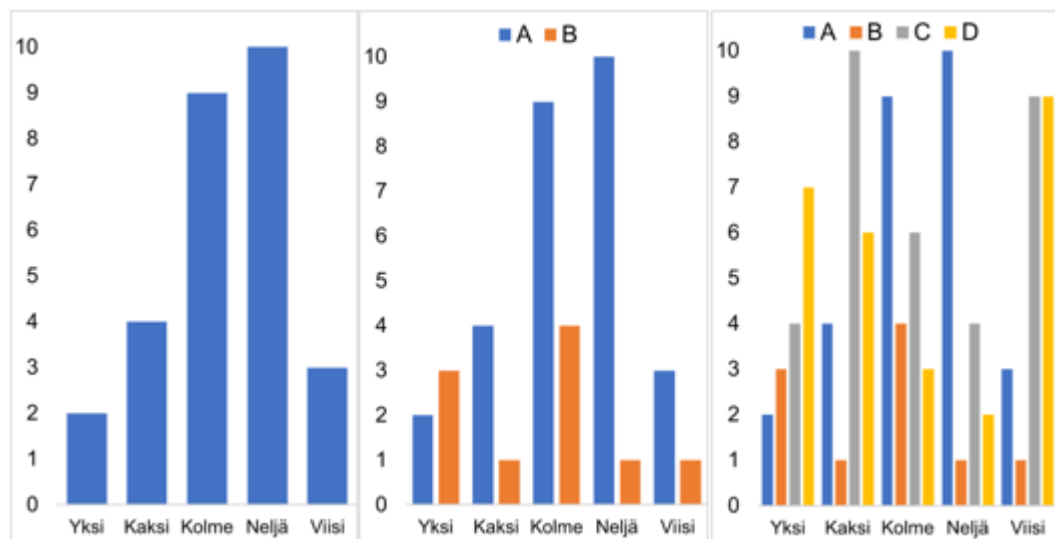


Kuvio 5. Muutettu Y-akseli eri perusviiva (Parikh 2014).

Lisäksi on tärkeää pitää x ja y-akselin risteys eli perusviiva nollassa, jotta pylväiden pituudet ovat verrattavissa (Nussbaumer Knaflic 2015). Perusviivan muuttaminen johtaa harhaanjohtaviin kuvioihin ja se kasvattaa pylväiden pituuden visuaalista eroa (kuvio 5).

Pystysuora pylväskaavio

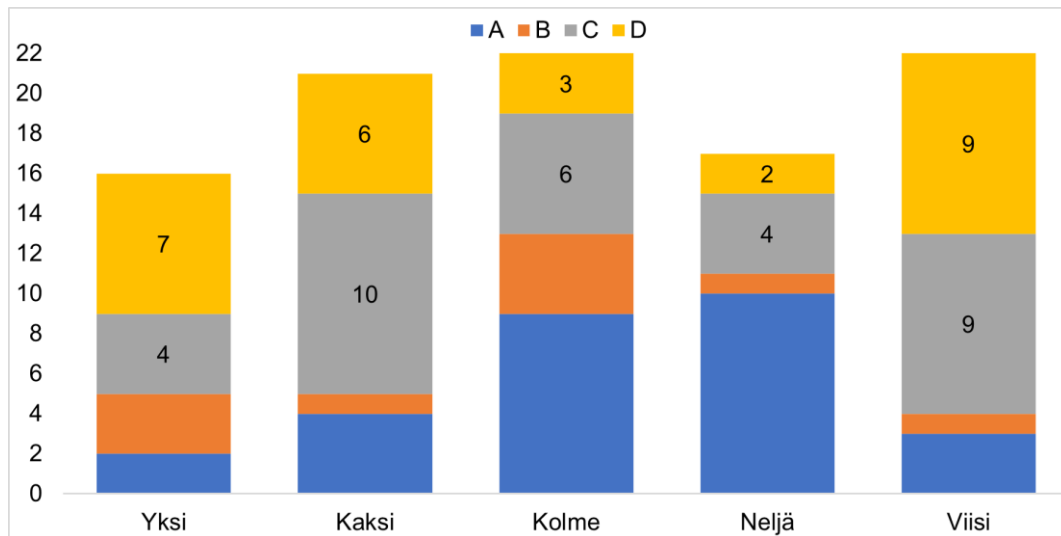
Pystysuora pylväskaavio on yleisin kaaviotyyppi, ja usein ideoita pystytään kommunikoimaan ainoastaan sitä käyttämällä.



Kuvio 6. Pylväskaavioita yhdellä ja useammalla sarjalla.

Sarjoja lisäämällä saadaan moninkertaistettua kuvioon sisältyvän tiedon, mutta luettavuus ja verrattavuus kärsivät (kuvio 6). Tästä syystä kannattaa erottaa luvut useampaan pylvääseen, ellei ole tärkeää verrata eri sarjojen lukuja toisiinsa. Pitämällä mielessä kaavioiden viemän tilan, voi olla parempi idea erottaa tiedot useampaan eri kaavioon, mutta pitää samankokoiset akselit verrattavuuden säilyttämiseksi.

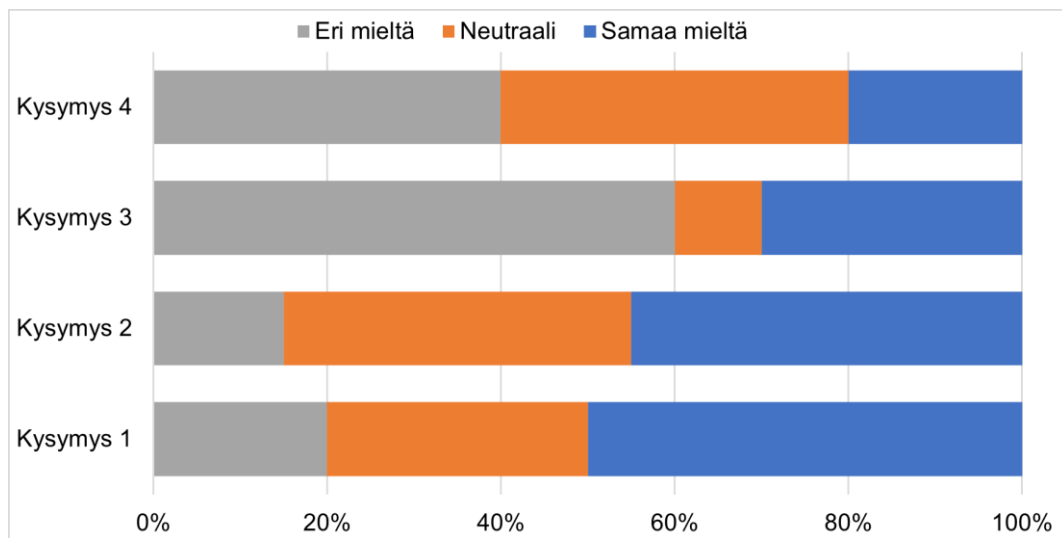
Pinottu pystysuora pylväskaavio



Kuvio 7. Pinottu pystysuora pylväskaavio.

Tälle kuviotyypille on rajoitetusti käyttötapauksia. Sen tarkoitus on visuaalisesti verrata eri luokkien kokonaismääriä ja näyttää, mistä komponenteista kokonaismäärät koostuvat. Luettavuus ja verrattavuus kuitenkin kärsivät. Kuviossa 7 A-sarja (sininen) on paljon helpommin verrattavissa kuin B-sarja (oranssi), sillä A-sarja alkaa samalta akselilta ja B-sarjalla ei ole suoraa alkuviivaa. Verrattavuutta voidaan helpottaa lisäämällä luvut pylväisiin, mutta ne eivät välttämättä mahdu paikoilleen ja kuvio alkaa muistuttamaan normaalia taulukkoristikkoa.

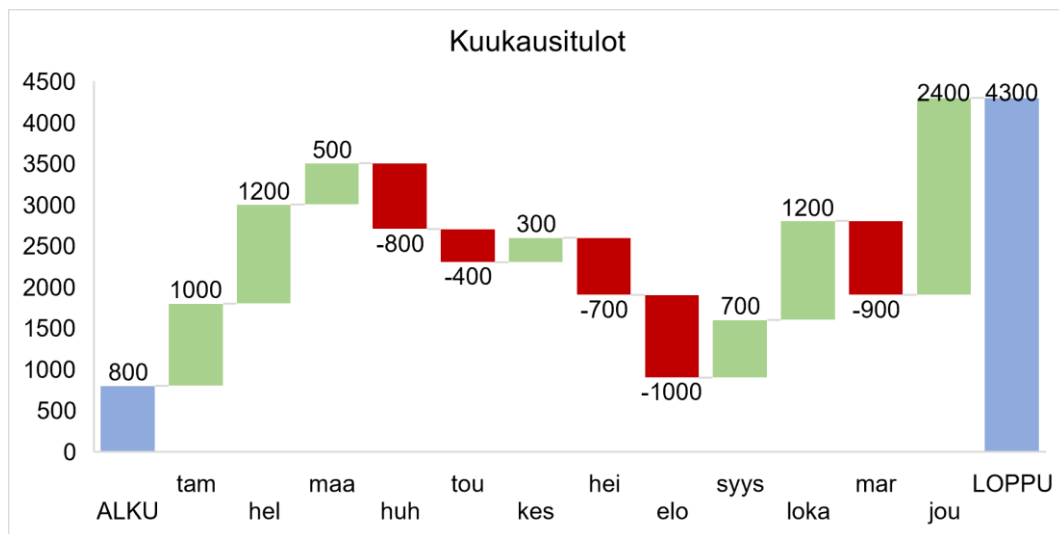
Pinottu vaakasuuntainen pylväskaavio (palkkikaavio)



Kuvio 8. Pinottu vaakasuuntainen pylväskaavio, jossa X-akseli summattu 100 prosenttiin.

Tämä visualisointi toimii todella hyvin tuloksiin, jossa vastaukset vaihtelevat negatiivisesta positiiviseen. 0% ja 100% eli kaavion reunat luovat kaksi eri perusviivaa, jotka pitävät molemmat ääripäät keskenään verrattavissa (kuvio 8). Tämä on toimiva visualisaatio kyselydatalle, sillä se näyttää tuloksen helposti luettavana kokonaisuutena. Usein Likert-asteikolla tuotetut kyselyt visualisoidaan tällä tavalla.

Vesiputouskaavio

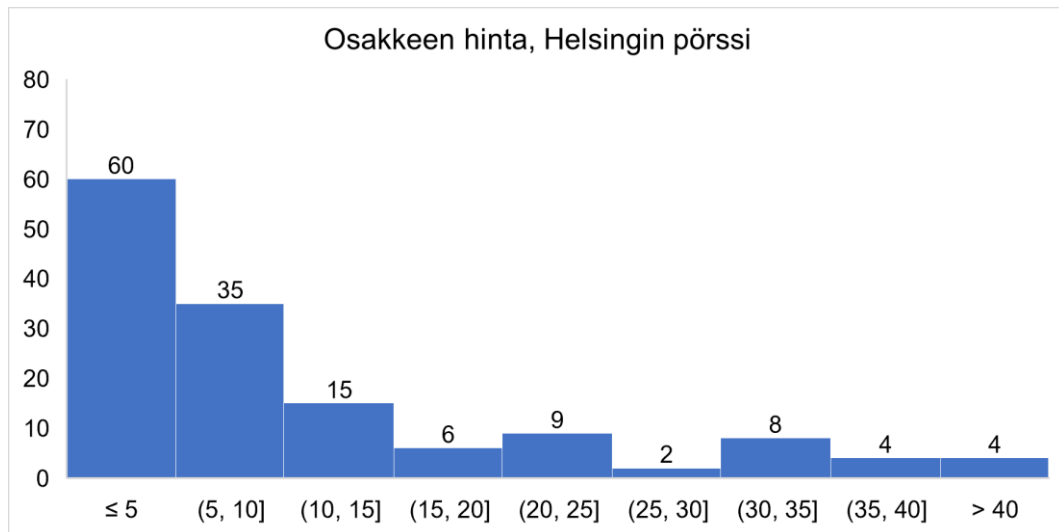


Kuvio 9. Vesiputouskaavio kuukausituloista.

Vesiputouskaaviota voidaan käyttää viivakaavion tilalla, jos halutaan painottaa jatkuvan arvon sijaan jatkuvaa muutosta. Kuvio 9 kommunikoi eri kuukausien tuloa sekä vuoden alku- ja loppusummaa selvästi. Vesiputouskaaviota voidaan myös käyttää pinotun pystykaavion tilalla, erottaen eri sarjat, mutta säilyttäen summaamisen.

Histogrammi

Histogrammia usein käytetään tilastotieteessä kuvaamaan jotain datasta ilmenevää jakaumaa.



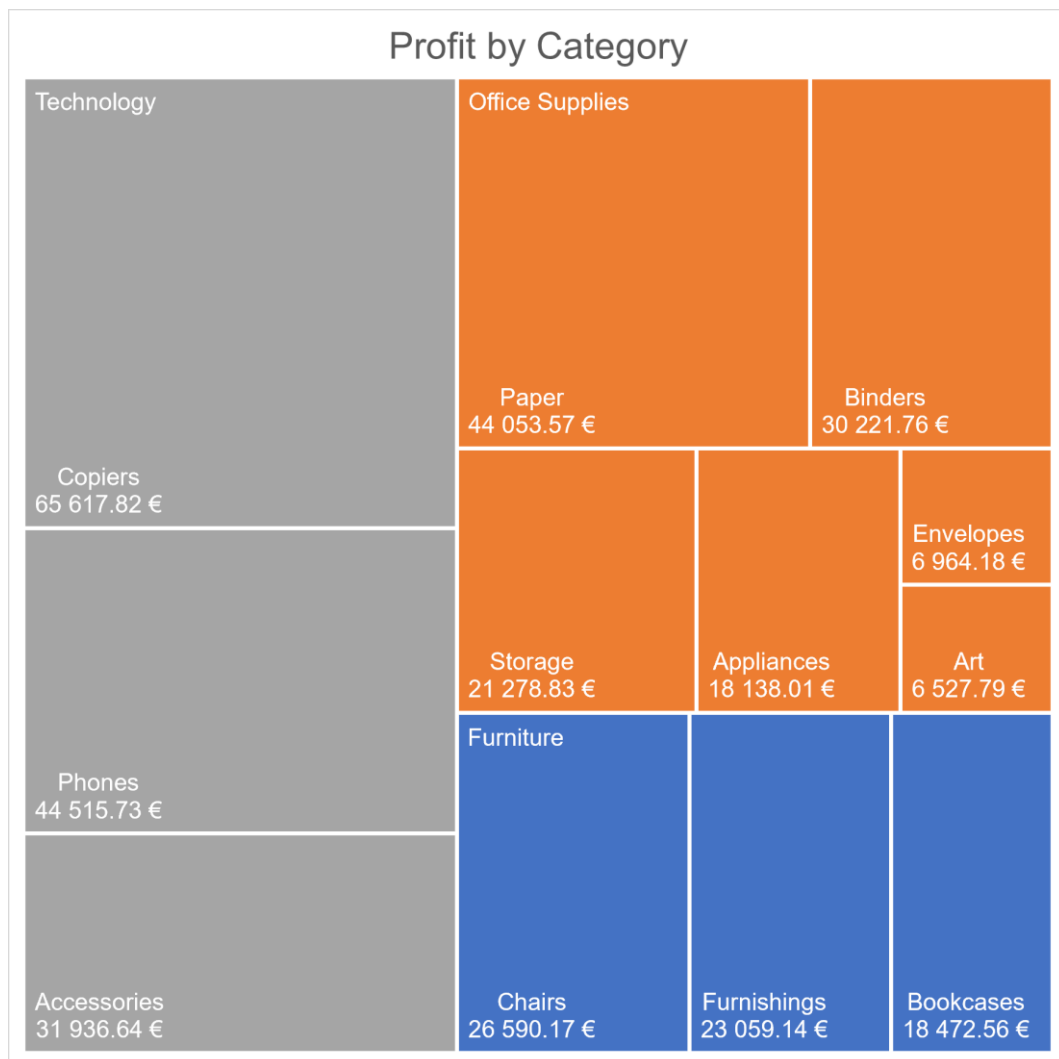
Kuvio 10. Osakkeiden hintojen jakauma Helsingin pörssissä, histogrammi.

Kuvio 10:ssä osakkeet on jaettu 5,00 € välisiin joukkoihin. Kuviossa 10 on selvästi positiivisesti vino jakauma eli yhtiöitä ilmaantuu enemmän joukoissa, joissa hinta on pienempi.

5.1.4 Aluekaaviot (puukaavio ja vohvelikaavio)

Aluekaaviot ovat hyviä visualisoimaan suuria kokonaisuuksia tai suuria eroavaisuuksia eri lukujen välillä.

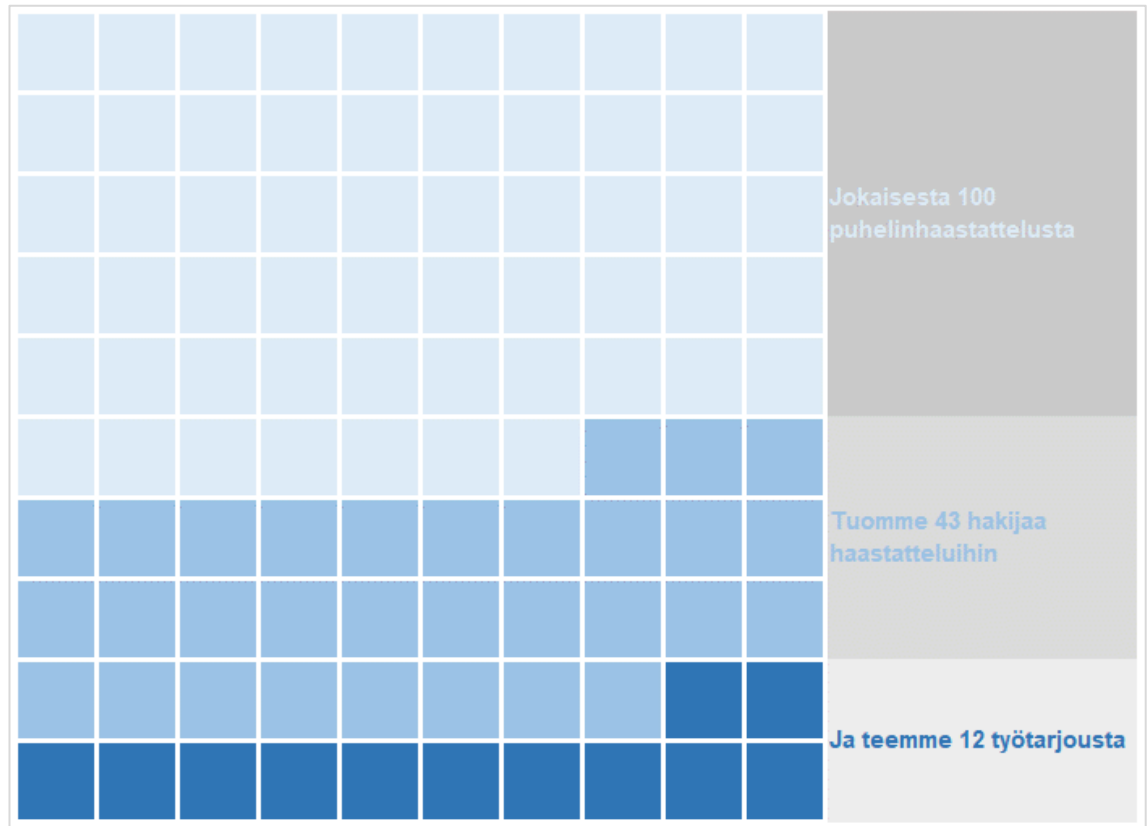
Puukaavio



Kuvio 11. Puukaavio verkkokaupan tuotosta tuotekategorioittain.

Puukaavio esittää usein suuren kokonaisuuden jakautumista osiin (kuvio 11). Tämä visualisointi ei ole kaikista luettavin luokkien lisäntyessä, mutta se antaa hyvän yleiskuvan kaikesta siihen käytetystä datasta. Puukaavio kuvaa kokonaisuuden hierarkkista jakautumista. Puukaavion luettavuutta myös auttaa se, että eri tiedot ovat jaettuna hierarkkisesti ja kirjaimellisesti omiin laatikkoihinsa.

Vohvelikaavio



Kuvio 12. Vohvelikaavio haastatteluista.

Vohvelikaavio on yksinkertainen visualisointi yhden kokonaisuuden prosenttiosuuksista (kuvio 12). Kommunikoinnin kannalta värien valitseminen on myös tärkeää, jotta alueet kommunikoivat selvästi osuutta samasta kokonaisuudesta. Se vastaa ympyräkaaviota, mutta osiot ovat kymmenjärjestelmän sekä kulmikkouden takia helpommin verrattavissa.

5.2 Muotoilu

Kaavioiden muotoilun ajattelemisen on tärkeää tehokkaan ja selkeän tiedonvälityksen saavuttamiseksi. Laajemmin puhutaan informaatiomuotoilusta, jota käytetään monella alalla, varsinkin infografiikassa, esimerkiksi julkisten

palvelujen karttoja suunnitellessa. Informaatiomuotoilun periaatteita voidaan käyttää hyödyksi kaavioita suunnitellessa, parantaen luettavuutta.

Informaatiomuotoilu painottaa yksinkertaisuutta tavoitellen selkeyttä ja ymmärrettävyyttä. Visualisoidun tiedon vähentäminen tai lisääminen voivat molemmat lisätä luettavuutta, joten ratkaisu ei ole aina ilmiselvä.

Yksinkertaisuus on myös subjektiivinen asia, joten tulee pitää mielessä kohderyhmä.

Kirjailija Nancy Duarte ehdottaa viiden vaiheen mallia yksinkertaisuuteen:

1. Kerro totuus
2. Mene asian ytimeen
3. Valitse oikea työkalu (visualisointi/kaavio)
4. Korosta tärkeät asiat
5. "Keep it simple" – lopullinen painotus yksinkertaisuuteen

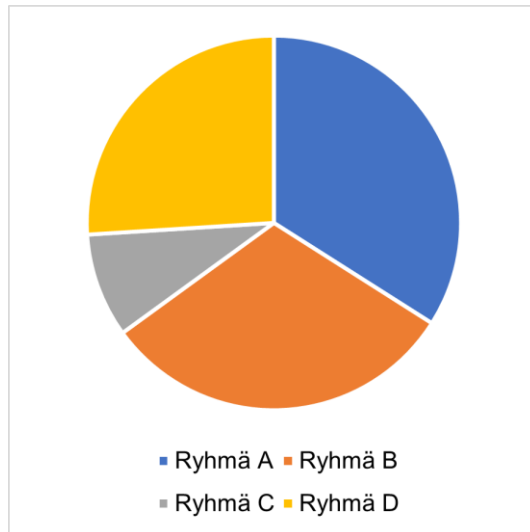
(Duarte 2008.)

Tämä mielestäni kiteyttää tärkeimmät piirteet. Visualisoinnin ei kuuluisi olla harjaanjohtava tai väärinymmärrettävä, sen pitäisi kommunikoida jokin asia nopeasti ja se korostaa tärkeimmät tiedot.

5.2.1 Vältettävät visualisoinnit

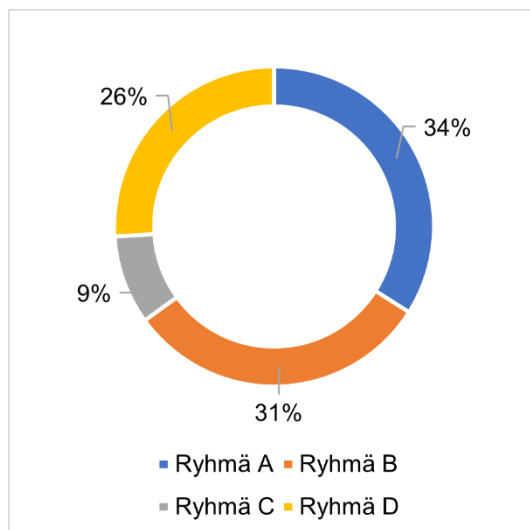
Ympyrä- ja donitsikaavio

Ympyräkaavioita näkee usein ja donitsikaavioita näkyy harvemmin, mutta monet asiantuntijan suosittelevat molempien välttämistä (Few 2007). Usein ne voidaan korvata kokonaan pylväs- tai muulla kaaviolla ja sama idea kommunikoidaan tehokkaammin.



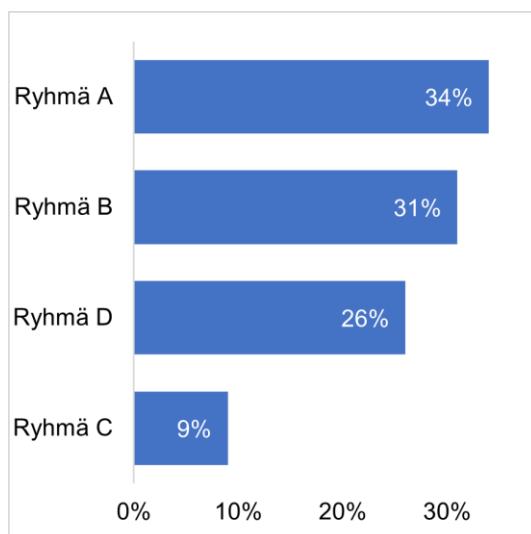
Kuvio 13. Esimerkki ympyräkaaviosta

Ympyräkaavion huonous piilee sen eri lohkojen verrattavuudessa. Kuvion 13 ryhmien A ja B lohkojen koon eroavaisuus ei ole selvästi havaittavissa, vaikka eroavaisuus on kuitenkin merkittävän kokoinen. Ryhmä D on selvästi C:tä suurempi, mutta on todella vaikea hahmottaa, kuinka paljon suurempi se on suhteessa C:hen. Otsikkojen avulla saadaan selville tarkka tieto, mutta kaavion hyöty jää tällöin todella pieneksi.



Kuvio 14. Esimerkki donitsikaaviosta.

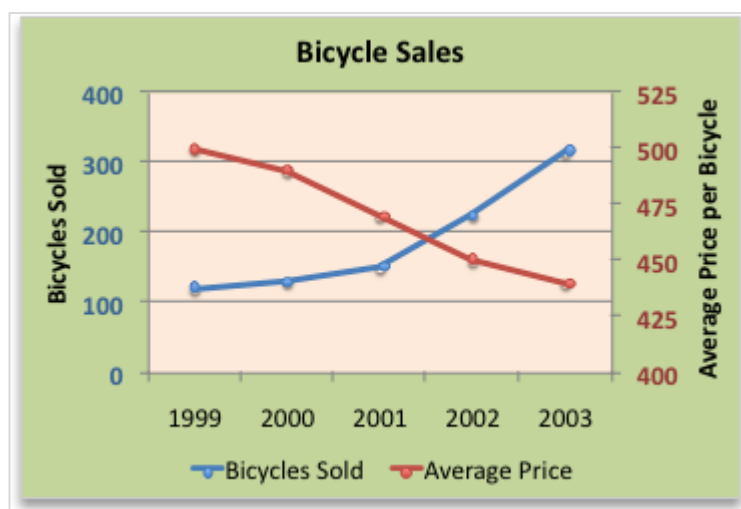
Kaikki samat kritiikit pätevät myös donitsikaavioon (kuvio 14), mutta luettavuus on vielä vaikeampi, sillä lukijan täytyy visuaalisesti verrata eri kaarien pituuksia.



Kuvio 15. Saman tiedon kommunikointi pylväskaavion avulla.

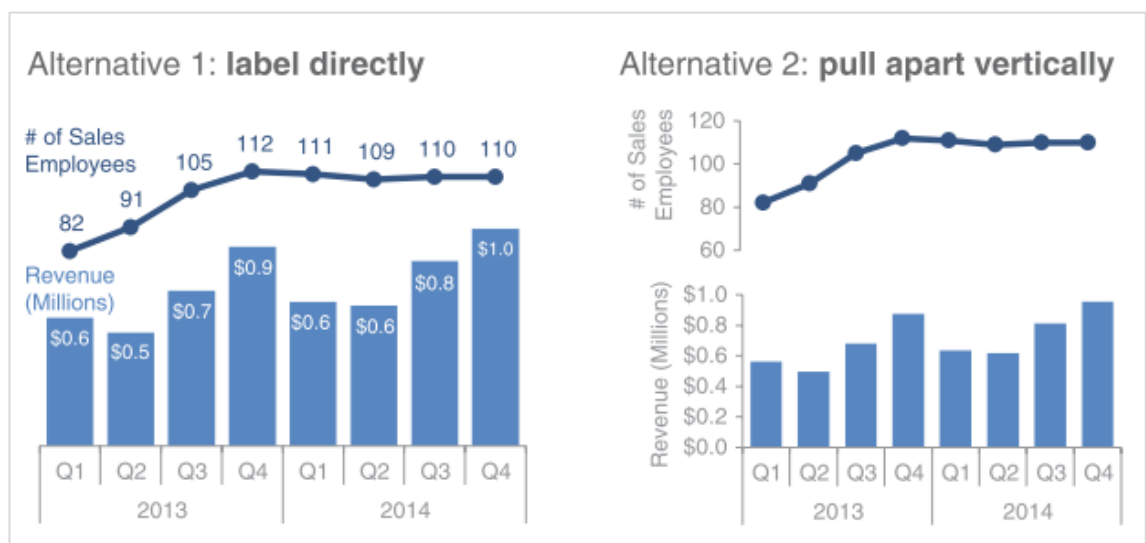
Kuviossa 15 on samat luvut visualisoitu pylväskaaviossa ja luettavuus on parantunut huomattavasti. Pylväät olisivat suoraan verrattavissa toisiinsa ilman otsikkojakin. Lisäksi kaikki sisällytetty tieto löytyy vierekkäin tutuista paikoista ja kaavio on yleisesti tehokkaampi. Suurin eroavaisuus ympyräkaavioon on se, että pylväät eivät osoita, että kokonaisuuden osien summa on 100%. Mutta tästä voidaan mainita erikseen tekstillä tai muulla tavalla.

Toissijainen Y-akseli



Kuvio 16. Myytyjen polkupyörien määrä ensisijaisella Y-akselilla ja polkupyörien hintojen keskiarvo toissijaisella Y-akselilla (Microsoft, 2023).

Toissijainen Y-akseli heikentää luettavuutta merkittävästi, sillä kaavion tulkitsija joutuu aina tarkistamaan, kumpaan akseliin tietty piste on verrattava (kuvio 16). Lisäksi se heikentää kommunikoinnin kokonaisuutta, sillä vilkauksella ei heti ymmärrä, mitä kuviolla halutaan osoittaa. Yhdistämällä kaksi tietoa samalle X-akselille tässä ei lisää kaavion arvoa, vaan lopputulos on heikompi kuin kahden eri tiedon kommunikoiminen kokonaan erikseen.

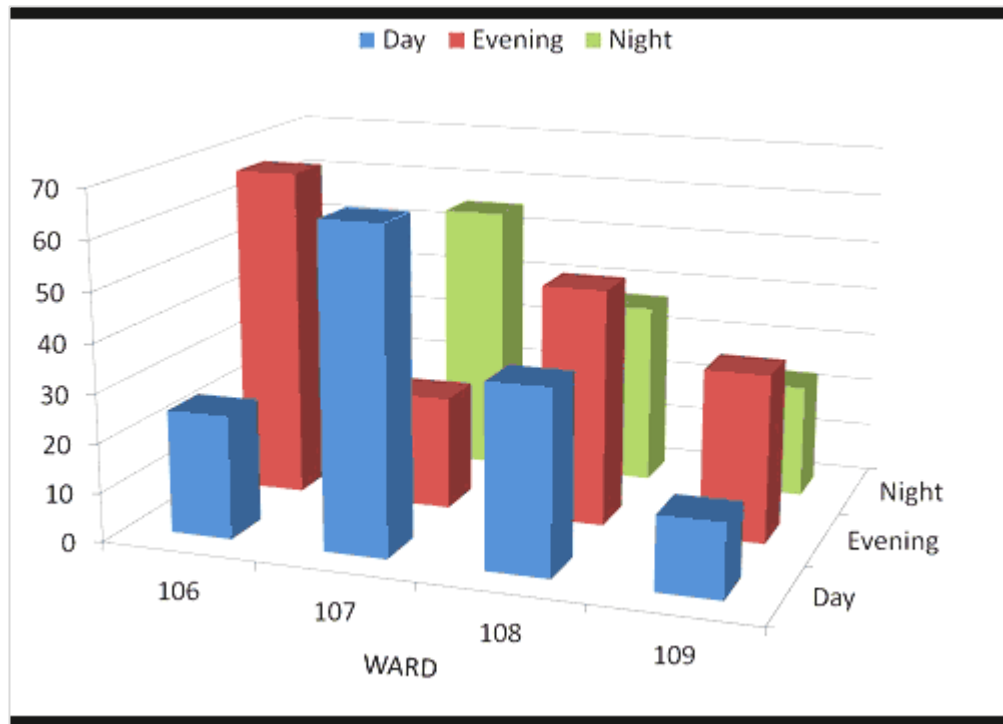


Kuvio 17. Vaihtoehdot kahden Y-akselin käyttöön (Nussbaumer Knaflic 2015).

Kirja antaa kuitenkin kaksi vaihtoehtoa kahden Y-akselin käytölle, mikäli halutaan säästää tilaa visualisoinnilla (kuvio 17). Vaihtoehdossa 1 poistetaan molemmat Y-akselit ja ilmoitetaan luvut kuviossa ja vaihtoehdossa 2 erotetaan toinen akseli ensisijaisen akselin päälle. Vaihtoehdot toimivat hyvin, mutta kahden arvon näyttäminen samalla kaaviolla vihjaa jostain arvojen suhteesta, mitä ei välttämättä ole olemassa (Nussbaumer Knaflic 2015). Tästä syystä mielestäni on paras tapa kommunikoida tarkoituksellisesti yksi idea yhdellä kuviolla ja erottaa tiedot kokonaan.

3D

3D-visualisointeja ei tulisi käyttää koskaan. Kolmas ulottuvuus vain lisää vaikealukuisuutta, usein vie ylimääräistä tilaa ja tuo kaavioon visuaalisia häiriötekijöitä. Visuaalisten elementtien verrattavuus kärsii ja akselien 0-pisteen eli perusviivojen hahmottaminen hankaloittuu (kuvio 18).



Kuvio 18. Esimerkki 3D-kaaviosta kolmella akselilla (Smith 2012).

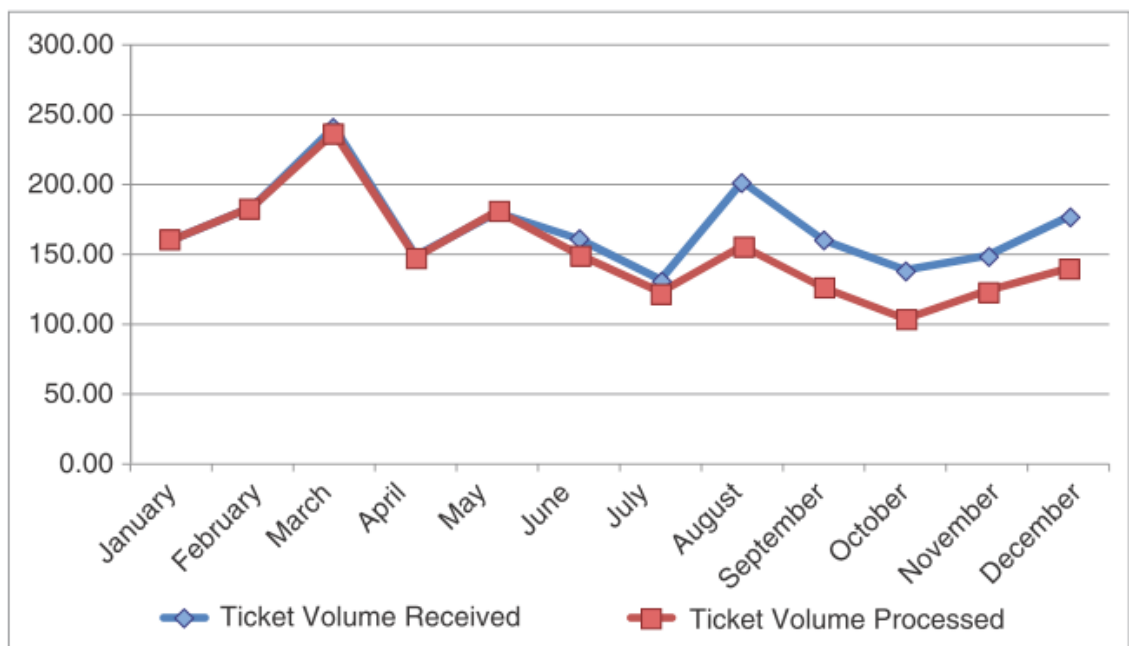
Mikäli 3D-kaaviota on käytetty kolmannen akselin sisällyttämiseen, voidaan kuitenkin erottaa akselit toisistaan ja luoda samasta datasta useampi pylväskaavio. Käytettyä tietoa ei kuitenkaan ole niin paljon, ettei luettavuus säilyisi.

5.2.2 Luettavuuden parantaminen

Luettavuutta voidaan parantaa monella tavalla ja se voi riippua myös kontekstista. Usein paras tapa parantaa luettavuutta on keskittyä siihen, mitä halutaan kuviolla kommunikoida ja poistaa elementit, jotka eivät hyödytä tätä tavoitetta.

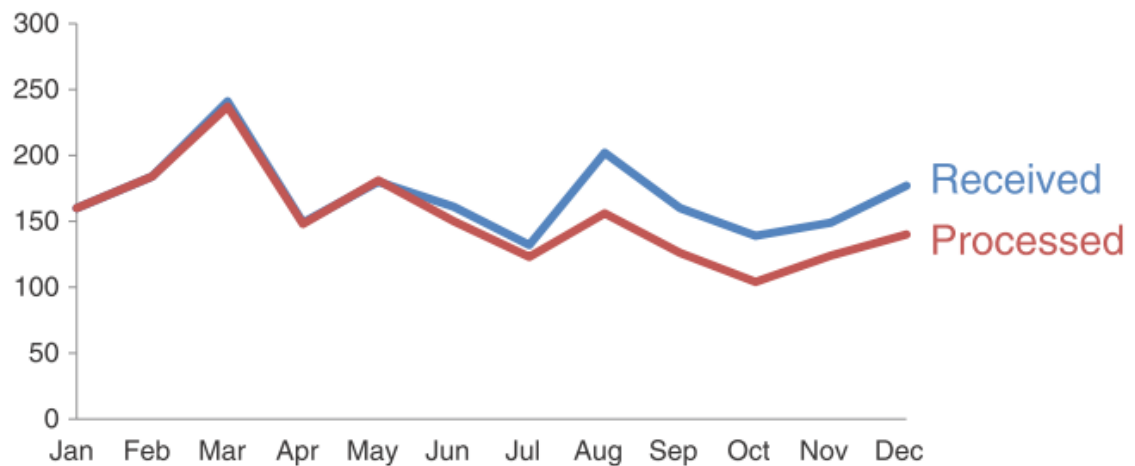
5.2.3 Ylimääräisen tiedon vähentäminen

Ylimääräinen tieto tarkoittaa jokaista näkyvää elementtiä, mikä ei lisää informatiivista arvoa tai ei lisää tarpeeksi arvoa sen läsnäololla. Vähentämällä ylimääräistä tietoa vähennetään myös jokaisen lukijan kognitiivista kuormaa eli aivojen aktivoitumista uuden tiedon oppimiseksi. Hyvää ja yksinkertaista visualisointia ei välttämättä huomaa, mutta huonon tuntee vaikeaksi tai raskaaksi tulkita.



Kuvio 19. Alkuperäinen kaavio (Nussbaumer Knaflic 2015).

Kuvion 19 on visualisointi tehty perusasetuksia käyttäen. Toimiva, mutta ei hyvä.

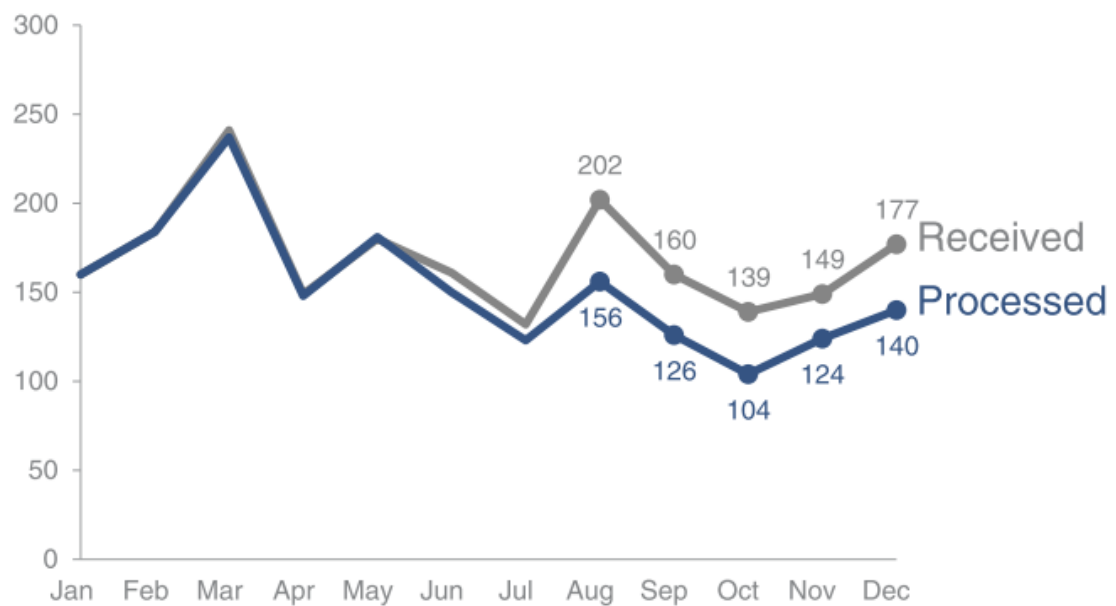


Kuvio 20. Muokattu kaavio (Nussbaumer Knafllic 2015).

Kuviosta 20 huomaa heti, että muokattu kaavio on paljon kevyempi tulkita. Vierekkäin alkuperäinen tuntuu raskaalta. Tämä kaavio kommunikoi, että vastaanotetut liput (ticket) ovat ajan myötä karanneet käsitellyistä lipuista. Kaaviosta on poistettu monta elementtiä, mutta sen sanoma pysyi ehjänä. Myös huomautettavaa on, että akselin apuviivoja ei tarvita, mikä ”avaa” kaavion visuaalisesti.

5.2.4 Korostus

Korostuksella voidaan tuoda esille tärkeimmät elementit. Ihmisen luonnollisen näköaistin takia voidaan strategisesti korostaa tietoa värin, koon, muodon tai sijainnin avulla. Korostettu tieto ohjaa lukijan huomiota ja keskittymistä, sekä jää helpoiten muistiin. Korostuksella voi myös toteuttaa visuaalista hierarkiaa, asettaen tiedon johonkin järjestykseen. Korostuksen voi esimerkiksi toteuttaa niin yksinkertaisesti kuin lihavoidulla tekstillä akseleissa tai pisteissä, tehostaen kommunikointia.



Kuvio 21. Esimerkki korostuksesta kuviota 20 käyttäen (Nussbaumer Knaflic 2015).

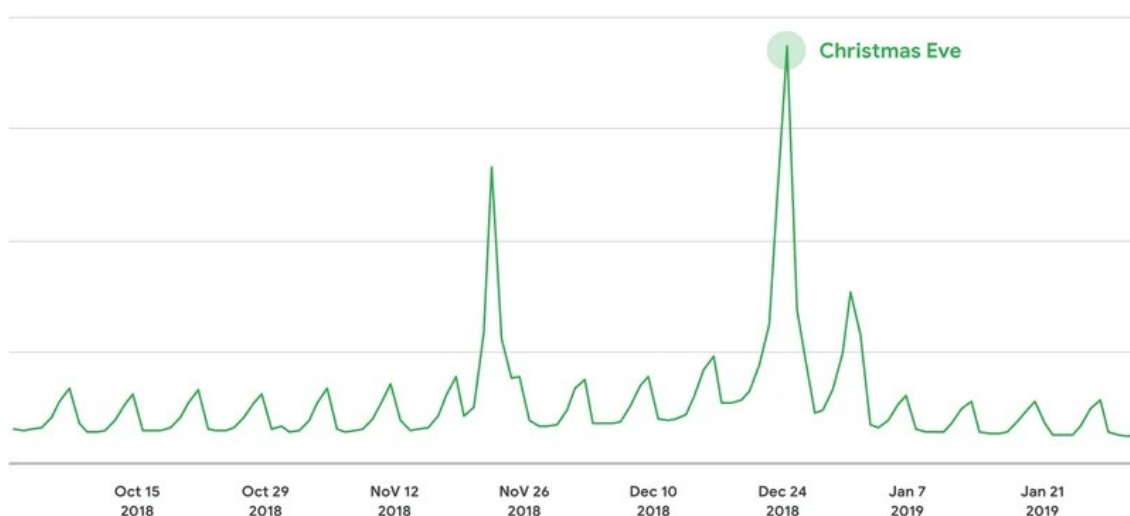
Kuviossa 21 on käytetty väriä, pisteitä ja tekstiä korostuksena. Kaaviota katsoessa huomaa, että huomio automaattisesti siirtyy oikeaan reunaan ja aivot alkavat vertailla pisteiden eroavaisuuksia Y-akselilla. Vasemmalla olevat ”tyhjät” viivat tuovat esille kokonaisuuden kontekstin.

5.2.5 Tekstit

Visualisaation otsikoissa ja muissa osissa käytetyn tekstin luettavuutta voidaan myös parantaa. Joskus akseleilla otsikot ovat joko vinossa tai pystysuorassa, mikä vaikeuttaa luettavuutta, josta syystä tätä olisi vältettävä. Esimerkki kuviossa 20. Tekstielementtejä voidaan korostaa myös *kursivoinnilla*, alaviivauksella tai **lihavoinnilla**. Lihavointi on näistä paras, sillä kursivointi vaikeuttaa luettavuutta varsinkin pienemmällä fontilla ja alaviivaus lisää tarpeetonta visuaalista sekavuutta. Tapauskohtaisesti voidaan myös kirjoittaa korostettuja elementtejä suuremmalla fontilla, kokonaan isoilla kirjaimilla tai eri väreillä.

Jokaisella kaaviolla kuuluisi olla kaavion otsikko sekä molempien akselien otsikot, jotta lukija saa ensisilmäyksellä idean, mitä kaavio esittää ennen numeroiden tutkimista. Tässä opinnäytetyössä käytetyissä kaavioissa ei kuitenkaan aina ole pidetty otsikoita, sillä ne eivät ole olleet ideoiden kommunikoimisen kannalta tärkeitä.

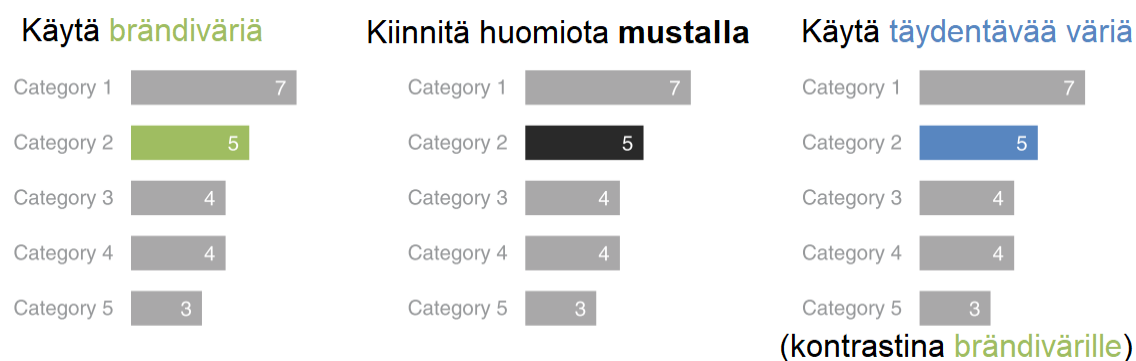
Tekstiä voidaan myös lisätä kaavioihin huomautusten muodossa, kunhan ne eivät tuo liikaa sekavuutta eivätkä mene liikaa itse kaavion päälle.



Kuvio 22. Esimerkki annotaatiosta (Google Data 2019).

Annotaatio tuo heti kontekstin selväksi kun katsotaan korkeaa piikkiä (kuvio 22).

5.2.6 Värit



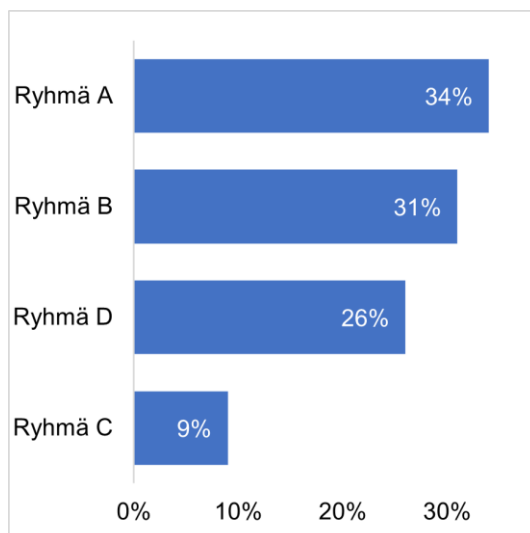
Kuvio 23. Värin käyttö.

Värien käytön tulisi aina olla tarkoituksellista ja värejä tulisi käyttää strategisesti tärkeiden osuuksien korostamiseksi (kuvio 23). Usein värejä käytetään liikaa. Yksikin väri riittää korostukseen, oli se sitten brändiväri tai muu. Voi olla myös hyvä idea käyttää mustan sijaan tummaa harmaasävyä oletuksena ja säästää tumma musta korostukseen. Lisäksi tummempia ja vaaleampia harmaita voi käyttää hierarkian perustamiseen. Visualisoinnissa voi olla myös tärkeää muistaa värisokeus, sillä n. 5 % ihmisistä on värisokeita. Turvallinen korosteväri on sininen, ja sininen on myös rauhallinen sekä ammattimainen. Muillakin väreillä on oma ”äänensävy”, esim. vihreä on hyvä, punainen on huono, oranssi on huomioväri, keltainen on optimistinen. Lisäksi samojen värien käyttäminen jokaisessa kaaviossa tuo yhtenäisyyttä ja lukija tottuu kommunikaation visuaaliseen kieleen.

5.2.7 Muut visuaaliset elementit

Milloin kääntää pylväät kyljelleen?

Klassisen pylväsdiagrammin kääntäminen vaakatasoon lisää luettavuutta, sillä luemme luonnollisesti vasemmalta oikealle. Lisäksi myös pitkät otsikot mahtuvat akselille haittaamatta luettavuutta (kuvio 24).



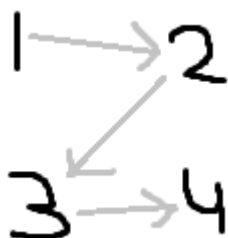
Kuvio 24. Pylväsdiagrammi vakaatasossa.

Silmät alkavat tekemään Z-kuviota: A, 34 %. B, 31 % jne. Luettavuutta parantaa myös se, että elementit luetaan luonnollisesti otsikko ensin ja sitten luku.

Pystykaaviossa vaaka-akselilla ovat pylväiden kärjet, mikä lisää niiden verrattavuutta vaakakaavioon verrattuna.

Lukusuunta

Kuten huomataan, silmät luonnollisesti lukevat vasemmalta oikealle ja ylhäältä alas (kuva 59).



Kuva 59. Z-kuvio: luonnollinen lukusuunta.

Onneksi useimmat ohjelmistot oletuksena laittavat luvut suuruusjärjestykseen ja otsikot järkeviin paikkoihin, mutta varsinkin monimutkaisemmissa kaavioissa kannattaa miettiä, miten asettaa tiedot kaavioon. Hyvä nyrkkisääntö on antaa aina ensin luvuille jokin otsikko ja konteksti akselilla. Muutoin lukija löytää ensimmäisenä luvut kaaviosta ja joutuu miettiä, mitä sillä oikein tarkoitetaan.

Muuta

Hyvä käytännöllinen tapa on kopioida valmis visualisointi ja muokata siitä useamman hieman erilaisen version ja pyytää mielipiteitä muilta. Lopuksi: ei ole oikeita vastauksia tai tapoja luoda visualisaatioita, mutta tärkeää on, että jokaisella valinnalla on jokin tarkoitus.

6 Johtopäätökset

Data-analytiikkaa tarkastellessa tulivat esille datan lajit, datan suhde sen ympärillä toimiviin asioihin, miten dataa voidaan hyödyntää päätöksenteossa, Business Intelligence sekä Big- ja ”Traditional” Datan eroavaisuudet. Loin käytännöllisiä ja ymmärrettäviä esimerkkejä PowerQueryn toiminnasta sekä datan hallinnasta Excel-funktioita ja toimintoja käyttäen. Visualisointeja sekä muotoilua tarkastellessa tärkeimmiksi piirteiksi luettavuuden ja ymmärrettävyyden kannalta ilmentyi kontekstin ymmärtäminen, tarkoituksellisuus, sopivan visualisoinnin valinta, ylimääräisen tiedon välttäminen ja huomion kesittäminen datan sekä visualisoinnin avulla.

Johdannossa asetetut tavoitteet saavutettiin. Analytiikkaa funktioilla -osio keskittyy lähinnä tiedon muotoiluun ja käsittelyyn, mutta käsitellyt funktiot ja niiden toiminta on kuitenkin oleellinen. Opinnäytetyö tuo lisäarvoa kirjoittajalle itseoppimisen kannalta sekä voi tuoda lisäarvoa myös lukijalle. Itsessään työ tuo lisäarvoa sillä, että opinnäytetyöstä saa kokonaiskuvan opinnäytetyössä käsitellyistä aiheista.

Haastavaa työssä oli aiheen rajaaminen sopivan pieneksi ja aiheen sovittaminen opinnäytetyöksi. Työn tuloksena on tietoa joistain data-analytiikan ja visualisointien osista, jotka voivat olla merkityksellisiä myös monelle erilaiselle lukijalle menemättä liian erikoisiin tai teknisiin aiheisiin. Toivottavasti opinnäytetyöstä on hyötyä lukijalle ja se saisi lukijan ajattelemaan dataa ja sen visualisointia syvällisemmin.

Alussa minulla oli kunnianhimoisia ajatuksia opinnäytetyön suhteen, mutta opinnäytetyön rakenne sekä opinnäytetyön kirjoittaminen ilman toimeksiantoa päättyi liian rajoittavaksi. Oli vaikea tehdä yhtenäistä, merkityksellistä ja tiivistä opinnäytetyötä tästä aiheesta, vaikka kiinnostusta olisi. Oli paljon ideoita, jotka olivat liian suuria opinnäytetyön kannalta tai liian vaikeita tuottaa ilman jonkinlaista ulkoista resurssia. Erilaisella ohjauksella opinnäytetyöprosessin alussa olisin mahdollisesti saanut konkreettisen lopputuloksen vision aikaiseksi, jota kohti olisin voinut työskennellä. Olisin halunnut tuottaa jotakin

konkreettisempaa ja hyödyllisempää, mutta toiminnallisen opinnäytetyön tuotos oli hukassa. Oli myös vaikeaa rajata aihe oikein, sillä aihepiiri on todella laaja ja halusin keskittyä asioihin, jotka ovat olennaisia taloushallinnon opiskelijoille eivätkä liian teknisiä, jotta lukijat saisivat opinnäytetyöstä jotakin irti ilman aiempaa tietämystä aiheesta. Lisäksi monen kuvankaappauksen ja tekstin käyttö keskenään ei toimi sujuvasti opinnäytetyöpohjan luomassa rakenteessa luettavuuden kannalta. Valmis opinnäytetyö ei oikeastaan vastaa alkuperäisiä odotuksia tai alkuperäistä potentiaalia, mutta ei opinnäytetyö kokonaan turha ole ja olen silti ylpeä valmiista opinnäytetyöstä. Pidin kuitenkin opinnäytetyön kirjoittamista hyödyllisenä ja se opetti minulle paljon työn tuottamisesta silloin, kun se on vaikeaa ja se ei tunnu luonnolliselta.

Lähteet

Ahlsell. (2023) Excel-ohjehinnastot. LV- ja ilmastointituotteet.

<https://www.ahlsell.fi/palvelut/muut-e-commerce-palvelut/hinnastopalvelut/>.

Hinnastot päivitetty ja viitattu 1.11.2023.

C.Kacfeh Emani, et al.. (2015) Understandable Big Data: A survey. Computer Science Review. s. 70-81. <http://dx.doi.org/10.1016/j.cosrev.2015.05.002>

Cannas, A. Tedeschi, L. Atzori, A. Lunesu, M. (2019) The data-information-knowledge-wisdom (DIKW) hierarchy as a pyramid to manage knowledge.

https://www.researchgate.net/figure/The-data-information-knowledge-wisdom-DIKW-hierarchy-as-a-pyramid-to-manage-knowledge_fig6_332400827. Viitattu

16.8.2023.

Cheusheva, S. (2020) Ablebits.com. XLOOKUP function in Excel – powerful successor of VLOOKUP. <https://www.ablebits.com/office-addins-blog/excel-xlookup-function/>. Viitattu 20.11.2023.

Donges, N. (2018) Data Types in Statistics. Medium.com, Towards Data Science-julkaisu. <https://towardsdatascience.com/data-types-in-statistics-347e152e8bee>. Viitattu 27.8.2023.

Duarte, N. (2008) Slide:ology The Art and Science of Creating Great Presentations. O'Reilly Media.

Eläketurvakeskus 2023. Suomen työeläkkeensaajat 2022. Eläketurvakeskuksen tilastoja-julkaisu. <https://www.julkari.fi/bitstream/handle/10024/146796/suomen-tyoelakkeensaajat-2022.pdf>. Viitattu 25.9.2023.

Few, S. (2007) Perceptual Edge. Visual Business Intelligence Newsletter. Save the Pies for Dessert. <http://www.perceptualedge.com/articles/08-21-07.pdf>. Viitattu 30.8.2023.

Google Data. 2019. Searches that contain “store hours” peak on Christmas Eve Think with Google Source: Google Data, U.S., Mobile searches over the holiday period, Oct. 2018–Jan. 2019. Lähteestä

<https://www.thinkwithgoogle.com/consumer-insights/consumer-trends/last-minute-shopper-behavior/>. Viitattu 1.9.

Greg Smith. (2016) Kaggle Datasets. Video Game Sales. <https://www.kaggle.com/datasets/gregorut/videogamesales/>. Viitattu 20.11.2023.

Liigapörssi. (2023) Iltasanomat. Liigan sarjataulukko. <https://liigaporssi.fi/sm-liiga/sarjataulukko>. Viitattu 22.10.2023.

Mäkelä, L. & Joronen, T. (2020) Dataa ja dadaa – ristiriitainen tiedon visualisointi. Teoksessa Jalonen, H., Helander, N. & Mäkelä, L. (toim.) Arvostustalous — kuinka arvostus rakennetaan ja rakentuu digiyhteiskunnassa. Vastapaino, s. 323 - 346.

Microsoft. (2023) Excel Help & Training. Charts. Add or remove a secondary axis chart in Excel. <https://support.microsoft.com/en-us/office/add-or-remove-a-secondary-axis-in-a-chart-in-excel-91da1e2f-5db1-41e9-8908-e1a2e14dd5a9#OfficeVersion=macOS>. Viitattu 5.10.2023.

Nussbaumer Knaflig C. (2015) Storytelling With Data. John Wiley & Sons.

Olavsrud, T & Fruhlinger, J. (2023) What is business intelligence? Transforming data into business insights. <https://www.cio.com/article/272364/business-intelligence-definition-and-solutions.html>. Viitattu 15.6.2023.

Parikh, R. (2014) Heap.io. Down the Funnel. How to Lie with Data Visualization. <https://www.heap.io/blog/how-to-lie-with-data-visualization>. Viitattu 28.9.2023.

Pure Storage. (2023) The Beginners Guide to Big Data. Big Data vs. Traditional Data. <https://www.purestorage.com/au/knowledge/big-data/big-data-vs-traditional-data.html>. Viitattu 15.6.2023.

Smith, J. (2012). Indezine. Problems with 3D Charts: by James M. Smith. <https://blog.indezine.com/2012/08/problems-with-3d-charts-by-james-m-smith.html>. Viitattu 1.9.

Tilastokeskus. (2021) StatFin. Ainevalinnat. Toisen asteen opiskelijoiden valitsemat vieraat kielet, 2020-2021. Toisen asteen opiskelijoiden valitsemat vieraat kielet, 2020-2021. Viitattu 22.10.2023.