

PX-Web – tilastotietokannan datan muokkaus Power BI:llä

Heidi Enho

Kouluttaja, vanhempi konsultti

Power BI & Excel -blogi: HExcelligent.fi

heidi@onsight.fi

Onsight



Power BI



Power BI Desktop



Julkaisu

Tietolähteet

~ 70 tietolähdettä, uusia julkistetaan säännöllisesti

Datan lataus, siivous, muokkaus

Datan yhdistely, laskenta ja visualisoinnit

- Ilmainen
- Kehittyy voimakkaasti, uusia ominaisuuksia kuukausittain
- Visualisointeja 29 + ~60 kpl
- Power BI Desktop ladattavissa osoitteesta:

<https://www.microsoft.com/en-us/download/details.aspx?id=45331>

Power BI –pilvipalvelu powerbi.com

- Maksuton tai maksullinen Pro*
- Oma työtila
- Ryhmätyötilat yhdessä tekemiseen*
- Raporttien ja visualisointien jakaminen

Hyödyntäminen

- Selain
- Mobiili Appsit (iOS, Android, Windows Phone 10)
- Excel



*) Maksullinen Power BI Pro

Onsight

Avointa dataa

- [Avoindata.fi](https://avoindata.fi)
- Tilastokeskuksen tilastotietokannat, kuten
 - PX-Web StatFi
 - Paavo – postinumeroittainen data
- Trafi
- Luonnonvarakeskuksen tilastotietokanta
- Eläketurvakeskuksen tilastotietokanta
- Verohallinnon tilastotietokanta

PX-Web esimerkki

Lataa ja asenna Power BI Desktop – ohjelma:

<https://www.microsoft.com/en-us/download/details.aspx?id=45331>

Seudullinen tilastorajapinta:

<https://www.avoindata.fi/data/fi/dataset/seudullinen-tilastorajapinta>

ETUSIVU / ORGANISAATIOT / HELSINGIN KAUPUNGIN ... / SEUDULLINEN TILASTORAJAPINTA

Seudullinen tilastorajapinta

Avoim data 08.11.2016 | Helsingin kaupungin tietokeskus

Seudullinen tilastorajapinta PX-Web API mahdollistaa Helsingin seudun aluesarjat -tilastokannassa (<http://aluesarjat.fi/>) olevien tilastotietojen hakemisen koneellisesti eri tiedostomuodoissa mm. xlsx, csv, json ja json-stat.

Aluesarjat-tilastokannassa on tilastollisia aikasarjoja useista alueellisista ilmiöistä. Tilastot ovat pääosin vuositason, pisimmät väestösarjat alkavat vuodesta 1962. Tietokanta muodostuu kolmesta kokonaisuudesta: Helsingin, Espoon ja Vantaan kaupunkien tilastoaluetiedoista sekä kunta- ja seututilastotietoa Helsingin seudusta ja Uudenmaan maakunnasta. Tietokannasta löytyvät uusimmat saatavilla olevat tiedot. Osaa tiedoista päivitetään vuosittain, osaa neljännesvuosittain.

Aihealueet - Asuminen - Asunto- ja rakennuskanta - Asunto- ja rakennustuotanto - Tulotaso - Työmarkkinat ja työssäkäynti - Väestö

PX-Web API:n käytöstä on laadittu ohjeita, joihin kannattaa tutustua ennen rajapinnan käyttöä: - Tilastokeskuksen käyttöohjeet suomeksi ja englanniksi <http://pxnet2.stat.fi/apli1.html> sekä - Ruotsin tilastokeskuksen käyttöohjeet <http://www.scb.se/Grupp/OmSCB/API/API-description.pdf>.

Aiemmin kokeilukäyttöön julkaistu REST-rajapinta löytyy osoitteesta <http://dev.hel.fi/stats/>.

Aineistolinkit (1)



PX-Web -rajapinta
PX-Web -rajapinta

1
Avaa



Helsingin seudun PX-Web-tietokannat

suomi ▼

2

Valitse tietokanta:

Tilastokanta

Helsingin ympäristötilasto

Etsi oikea tietojoukko

- Esim. Uusimaa > Työmarkkinat ja työssäkäynti > Työpaikat > Työpaikat Uudellamaalla (alueella työssäkäyvät) toimialan mukaan

Lähde: [Tilastokeskus](#) ja [Helsingin](#), [Espoon](#) ja [Vantaan](#) tilastotoimet.

 Tulotaso

 **Työmarkkinat ja työssäkäynti**

Valitse vasemmalta valikosta aihealue.

Alue:

Uusimaa kunnittain

Helsingin seutu (14 kuntaa)

Pääkaupunkiseutu (Helsinki, Espoo, Vantaa, Kauniainen)

Kehyskunnat (Hyvinkää, Järvenpää, Kerava, Kirkkonummi, Mäntsälä, Nurmijärvi, Pornainen, Sipoo, Tuusula, Vihti)

Helsingin, Raaseporin, Porvoon ja Loviisan seutukunnat

Lähde: [Tilastokeskus](#)

 **Työmarkkinat**

 **Työpaikat**

 Työpaikat Uudellamaalla (alueella työssäkäyvät) toimialan (TOL 2008) mukaan 31.12.   

 Työpaikat Uudellamaalla (alueella työssäkäyvät) toimialan (TOL 1995) mukaan 31.12.   

Valitse muuttujat

Huomaa, että yleensä jokin muuttujista (esim. vuosi) esitetään yleensä vaakasuuntaisesti eri sarakkeissa, jotta lukuja olisi helppo lukea. Tällainen vaakasuuntaiseksi muodostuva matriisi pitää kääntää pystysuuntaiseksi, jotta se soveltuu raportointiin ja analysointiin.

Mitä enemmän valitset erilaisia muuttujia, sitä vaativammaksi datan muokkaus muuttuu.

1
Valitse
taulukko

2
Valitse
muuttujat

3
Näytä taulukko

Työpaikat Uudellamaalla (alueella työssäkäyvät) toimialan (TOL 2008) mukaan 31.12.

Valitse muuttujat

Tietoa taulukosta

Merkitse valintasi ja valitse esitysmuoto (taulukko ruutuun tai tiedostomuoto). Valintaohje
*-merkityille muuttujille tarvitaan ainakin yksi arvo

Alue *	Toimiala *	Vuosi *
☒	☒	☒
Yhteensä 34 Valittu 34	Yhteensä 23 Valittu 23	Yhteensä 7 Valittu 7
Uusimaa Helsingin seutukunta Raaseporin seutukunta Loviisan seutukunta Porvoon seutukunta Suomi	Q Terveys- ja sosiaalipalvelut R Taiteet, viihde ja virkistys S Muu palvelutoiminta T Kotitalouksien toiminta työnantajina; kotitalouksien eriyttämätön toiminta tavaroiden ja palve U Kansainvälisten organisaatioiden ja toimielinten toiminta X Toimiala tuntematon	2008 2009 2010 2011 2012 2013
Etsi	Etsi	Etsi
☐ Rivin alusta	☐ Rivin alusta	☐ Rivin alusta

Valittuja taulukkosoluja: 5 474 (enimmäisraja on 100 000)
Selailunäkymä on rajoitettu 1 000 riviin ja 30 sarakkeeseen

Taulukkonäkymä 1

Jatka

Tallenna lopputulos joko csv- tai xlsx-tiedostoksi

Näytä taulukko Tietoja taulukosta

Muuta ja laske

Tallenna nimellä

Taulukkonäkymä 1







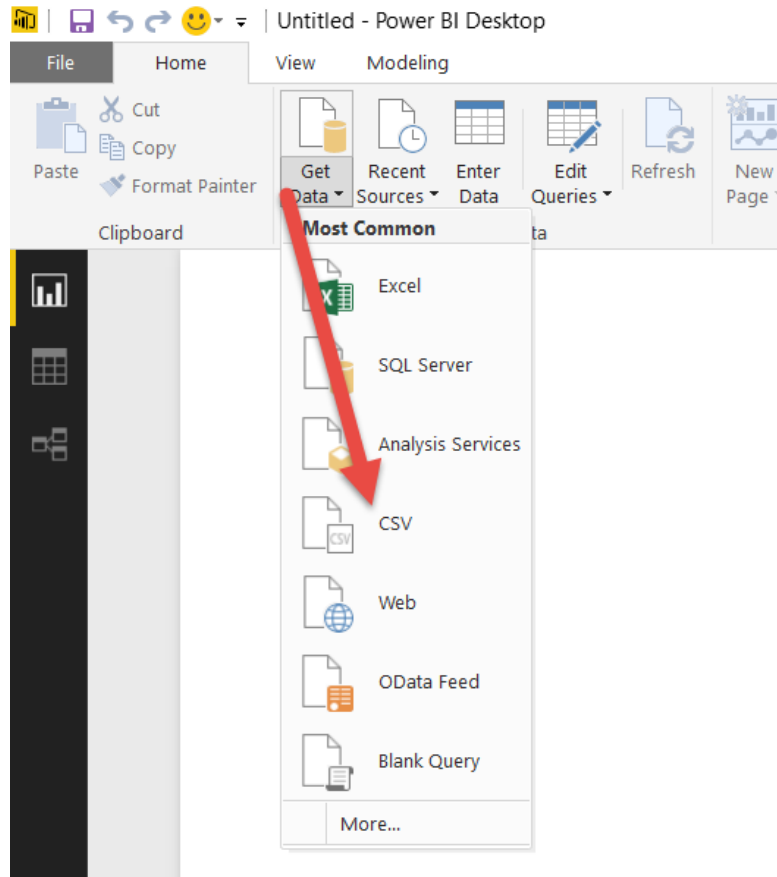
 Taulukon asetukset

 Tallenna poiminta

Työpaikat Uudellamaalla (alueella työssäkäyvät) toimialan (TOL 2008) mukaan 31.12.

	2007	2008	2009	2010	2011	2012	2013
Helsinki							
Kaikki toimialat yhteensä	385 356	388 053	380 579	381 625	386 357	386 634	381 942
A Maatalous, metsätalous ja kalatalous	381	392	490	492	490	484	483
B Kaivostoiminta ja louhinta	178	227	198	97	43	17	20
C Teollisuus	22 949	22 476	20 212	19 438	17 986	17 104	16 199
D Sähkö-, kaasu- ja lämpöhuolto, jäähdytysliiketoiminta	1 812	1 825	1 823	1 792	1 853	1 781	1 771

Muokkaa ja lataa csv-tiedosto Power BI Desktopissa



USUMTP2.csv

File Origin: 1252: Western European (Windows) | Delimiter: Semicolon | Data Type Detection: Based on first 200 rows

Column1	Column2	Column3	Column4	Column5	Column6
Työpaikat Uudellamaalla (alueella työssäkäyvät) toimial...		null	null	null	
		null	null	null	
Alue	Toimiala	2007	2008	2009	
Helsinki	Kaikki toimialat yhteensä	385356	388053	380579	3
Helsinki	A Maatalous, metsätalous ja kalatalous	381	392	490	
Helsinki	B Kaivostoiminta ja louhinta	178	227	198	
Helsinki	C Teollisuus	22949	22476	20212	
Helsinki	D Sähkö-, kaasu- ja lämpöhuolto, jäähdytysliiketoiminta	1812	1825	1823	
Helsinki	E Vesihuolto, viemäri- ja jätevesihuolto, jätehuolto ja m...	894	832	845	
Helsinki	F Rakentaminen	18975	18461	17000	
Helsinki	G Tukku- ja vähittäiskauppa; moottoriajoneuvojen ja m...	47243	46431	44502	
Helsinki	H Kuljetus ja varastointi	21138	21459	20442	
Helsinki	I Majoitus- ja ravitsemistoiminta	15617	16111	15605	
Helsinki	J Informaatio ja viestintä	35857	36063	35585	
Helsinki	K Rahoitus- ja vakuutustoiminta	18195	18328	18677	
Helsinki	L Kiinteistöalan toiminta	5365	5362	5318	
Helsinki	M Ammatillinen, tieteellinen ja tekninen toiminta	35672	36169	34290	
Helsinki	N Hallinto- ja tukipalvelutoiminta	28838	29024	26984	
Helsinki	O Julkinen hallinto ja maanpuolustus; pakollinen sosiaal...	32089	31975	31767	
Helsinki	P Koulutus	23240	23034	23865	

Buttons: Load, Edit, Cancel

Poista datan yläpuolelta turhat rivit

- Datan muokkaus ja käsittely tapahtuu Power BI Desktop –ohjelman ns. kyselyeditorissa (Query Editor)

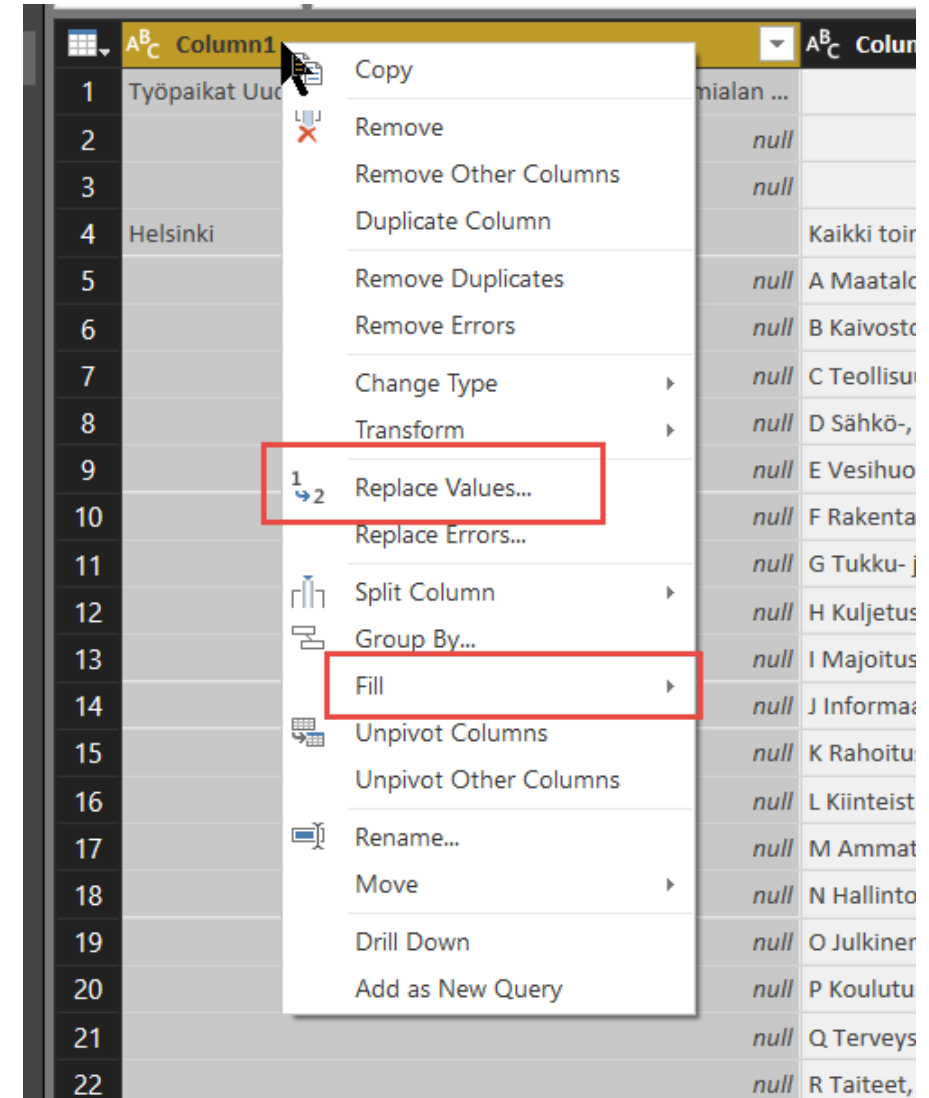
The screenshot shows the Power BI Desktop Query Editor interface. The 'Remove Rows' button in the 'Manage Columns' group is selected, opening a dropdown menu with the following options: Remove Top Rows, Remove Bottom Rows, Remove Alternate Rows, Remove Duplicates, Remove Blank Rows, and Remove Errors. The background shows a data table with columns 'Column1' and 'Column2'.

	Column1	Column2
1	Työpaikat Uudellamaalla (...)	
2		
3	Alue	Toimiala
4	Helsinki	Kaikki toimialat yhteensä
5	Helsinki	A Maatalous, metsätalous ja kalatalous
6	Helsinki	B Kaivostoiminta ja louhinta

Täytä sarakkeet

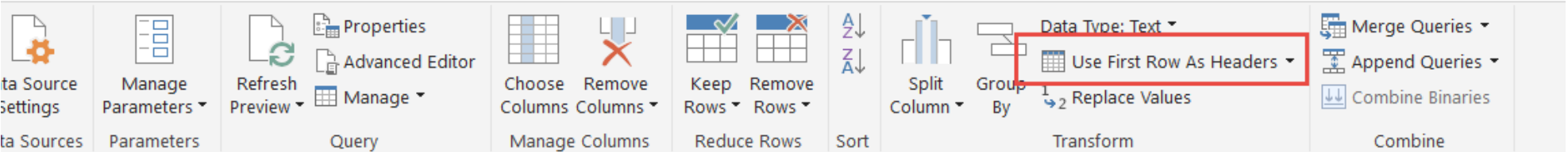
Kun käsitellään PX-Web tilastotietokannan dataa, eri tietojoukkojen aineistot tulevat hieman eri muodoissa riippuen käyttöliittymästä ja valitusta tiedostomuodosta. Tiedot voivat olla joko valmiiksi täytettyinä sarakkeissa tai "otsikoituina".

Jos sarakkeessa on otsikon kaltaisia tietoja eli esimerkiksi kunnan nimi on vain yhdellä rivillä, voit täyttää sarakkeessa tiedot alas pikavalikon **Fill > Down** –komennolla. **Fill** täyttää alla oleviin soluihin. Mikäli soluissa ei ole *null*-arvoja, vaan esimerkiksi tyhjä välilyönti, se pitää korvata ensin **Replace Values** –komennolla *null*-arvoiksi.



Nosta ylin rivi otsikoiksi

- Voit joko nostaa otsikkorivin tai kirjoittaa itse otsikot kaksoisnapsauttamalla otsikkokenttiä



The screenshot shows the Power Query ribbon with the 'Transform' tab selected. The 'Use First Row As Headers' option is highlighted with a red box. Below the ribbon, a data table is shown with the first row highlighted in red, indicating it has been used as the header row.

	Column1	Column2	Column3	Column4	Column5	Column6
1	Alue	Toimiala	2007	2008	2009	
2	Helsinki	Kaikki toimialat yhteensä	385356	388053	380579	
3	Helsinki	A Maatalous, metsätalous ja kalatalous	381	392	490	
4	Helsinki	B Kaivostoiminta ja louhinta	178	227	198	
5	Helsinki	C Teollisuus	22949	22476	20212	
6	Helsinki	D Sähkö-, kaasu- ja lämpöhuolto, jäähdytysliiketoiminta	1812	1825	1823	
7	Helsinki	E Vesihuolto, viemäri- ja jätevesihuolto, jätehuolto ja muu ympäristön ...	894	832	845	

Muuttujat sarakkeissa

- PX-Web kääntää yleensä jonkun muuttujista sarakkeiksi, jotta lopputulos olisi sellaisenaan helposti luettavassa muodossa
- Raportointiin tällainen data ei kuitenkaan kelpaa
- Muuttujat pitää kääntää sarakkeiksi eli datalle pitää suorittaa ns. Unpivot-operaatio
- Unpivotin voi suorittaa vasta kun kyseisen muuttujan arvot on nostettu otsikkoriville eli esimerkiksi vuodet näkyvät otsikoissa
- Mikäli aineistossasi on useita eri otsikkotasoja vaakasuuntaisesti (esim. yhdellä rivillä vuosi ja sen alla kuukaudet), et pysty käsittelemään dataa yhdessä kyselyssä, vaan joudut luomaan esimerkiksi eri kyselyn eri vuosia varten

Käännä (unpivot) data raportointiin sopivaksi

Table.PromoteHeaders("#Removed Top Rows")

	A ^B Alue	A ^B Toimiala	1 ² 2007	1 ² 2008	1 ² 2009	1 ² 2010	1 ² 2011
1	Helsinki	Kaikki toimialat yhteensä	385356	388053	380579	381625	386357
2	Helsinki	A Maatalous, metsätalous ja kalatalous	381	392	490	492	490
3	Helsinki	B Kaivostoiminta ja louhinta	178	227	198	97	43
4	Helsinki	C Teollisuus	22949	22476	20212	19438	17986
5	Helsinki	D Sähkö-, kaasu- ja lämpöhuolto, jätehuolto	1812	1825	1823	1792	
6	Helsinki	E Vesihuolto, viemäri- ja jätevesihuolto	894	832	845	824	
7	Helsinki	F Rakentaminen	18975	18461	17000	18015	
8	Helsinki	G Tukku- ja vähittäiskauppa; majoitus, ravitsemistoiminta ja liikenne	47243	46431	44502	43940	44384
9	Helsinki	H Kuljetus ja varastointi	21138	21459	20442	19831	19982
10	Helsinki	I Majoitus- ja ravitsemistoiminta	15617	16111	15605	15937	16758
11	Helsinki	J Informaatio- ja viestintä	35857	36063	35585	35262	35980
12	Helsinki	K Rahoitus- ja vakuutustoiminta	18195	18328	18677	18914	19172
13	Helsinki	L Kiinteistöalan toiminta	5365	5362	5318	5383	5344
14	Helsinki	M Ammatillinen, tieteellinen ja tekninen toiminta	35672	36169	34290	35747	36421
15	Helsinki	N Hallinto- ja tukipalvelutoiminta	28838	29024	26984	27331	28831
16	Helsinki	O Julkinen hallinto ja maanpuolustus; pakollinen sosiaalivakuutus	32089	31975	31767	29861	29924
17	Helsinki	P Koulutus	23240	23034	23885	24596	24814
18	Helsinki	Q Terveyden- ja sosiaalipalvelut	49475	51589	52946	53716	53677

Table.UnpivotOtherColumns("#Promoted Headers", {"Alue", "Toimiala"})

	A ^B Alue	A ^B Toimiala	A ^B Attribute	1.2 Val...
1	Helsinki	Kaikki toimialat yhteensä	2007	385356
2	Helsinki	Kaikki toimialat yhteensä	2008	388053
3	Helsinki	Kaikki toimialat yhteensä	2009	380579
4	Helsinki	Kaikki toimialat yhteensä	2010	381625
5	Helsinki	Kaikki toimialat yhteensä	2011	386357
6	Helsinki	Kaikki toimialat yhteensä	2012	386634
7	Helsinki	Kaikki toimialat yhteensä	2013	381942
8	Helsinki	A Maatalous, metsätalous ja kalatalous	2007	381
9	Helsinki	A Maatalous, metsätalous ja kalatalous	2008	392
10	Helsinki	A Maatalous, metsätalous ja kalatalous	2009	490
11	Helsinki	A Maatalous, metsätalous ja kalatalous	2010	492
12	Helsinki	A Maatalous, metsätalous ja kalatalous	2011	490
13	Helsinki	A Maatalous, metsätalous ja kalatalous	2012	484
14	Helsinki	A Maatalous, metsätalous ja kalatalous	2013	483
15	Helsinki	B Kaivostoiminta ja louhinta	2007	178
16	Helsinki	B Kaivostoiminta ja louhinta	2008	227
17	Helsinki	B Kaivostoiminta ja louhinta	2009	198
18	Helsinki	B Kaivostoiminta ja louhinta	2010	97
19	Helsinki	B Kaivostoiminta ja louhinta	2011	43
20	Helsinki	B Kaivostoiminta ja louhinta	2012	17
21	Helsinki	B Kaivostoiminta ja louhinta	2013	20
22	Helsinki	C Teollisuus	2007	22949
23	Helsinki	C Teollisuus	2008	22476
24	Helsinki	C Teollisuus	2009	20212
25	Helsinki	C Teollisuus	2010	19438

- Kaksoisnapsauta sarakeotsikoita ja syötä järkevät otsikot

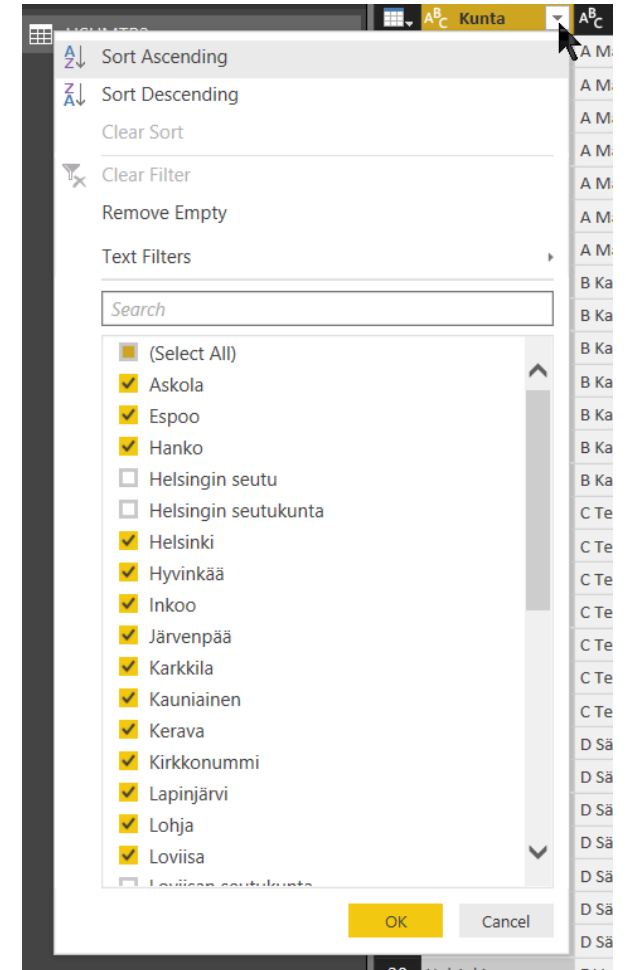
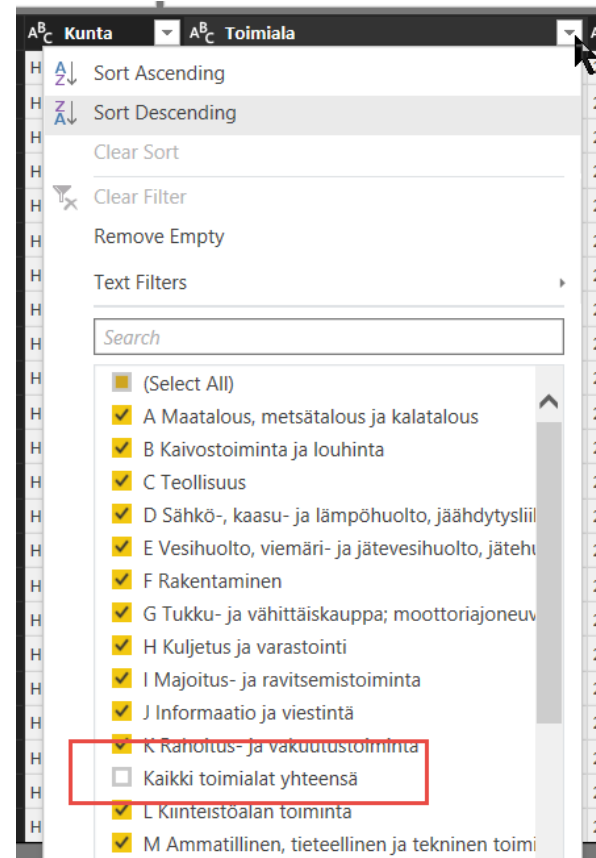
Table.RenameColumns("#Unpivoted Other Columns", {"Attribute", "Vuosi"}, {"Lukumäärä", "Lukumäärä"})

	A ^B Kunta	A ^B Toimiala	A ^B Vuosi	1.2 Lukumäärä
1	Helsinki	Kaikki toimialat yhteensä	2007	385356
2	Helsinki	Kaikki toimialat yhteensä	2008	388053

Suodata pois summatason tiedot, jätä vain data

Huomaa, että PX-Web tarjoaa dataa, johon on yleensä laskettu erilaisia ryhmäsummia. Ne pitää poistaa aineistosta.

Paina suodatuslistan alaosassa olevaa **Load more** -linkkiä, jos listalla ei näytetä kaikkia arvoja.



Tarkista vielä kunkin sarakkeen tietotyytit

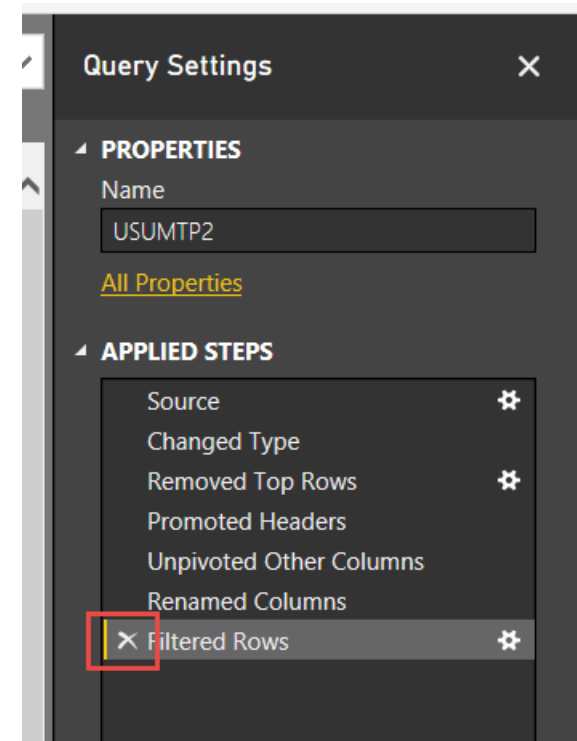
Tietotyytit ovat erittäin tärkeitä, sillä vain luvuilla voi laskea ja näyttää niitä eri visualisoinneissa. Tietotyytit valitaan otsikossa olevan symbolin avulla.

	 Kunta	 Toimiala	 Vuosi	 Lukumäärä
1	Helsinki	A Maatalous, metsätalous ja kalatalous	2007	 Decimal Number
2	Helsinki	A Maatalous, metsätalous ja kalatalous	2008	 Fixed Decimal Number
3	Helsinki	A Maatalous, metsätalous ja kalatalous	2009	 Whole Number
4	Helsinki	A Maatalous, metsätalous ja kalatalous	2010	 Date/Time
5	Helsinki	A Maatalous, metsätalous ja kalatalous	2011	 Date
6	Helsinki	A Maatalous, metsätalous ja kalatalous	2012	 Time
7	Helsinki	A Maatalous, metsätalous ja kalatalous	2013	 Date/Time/Timezone
8	Helsinki	B Kaivostoiminta ja louhinta	2007	 Duration
9	Helsinki	B Kaivostoiminta ja louhinta	2008	
10	Helsinki	B Kaivostoiminta ja louhinta	2009	 Text
11	Helsinki	B Kaivostoiminta ja louhinta	2010	 True/False
12	Helsinki	B Kaivostoiminta ja louhinta	2011	 Binary
13	Helsinki	B Kaivostoiminta ja louhinta	2012	17
14	Helsinki	B Kaivostoiminta ja louhinta	2013	20

Kun olet suorittanut erilaisia muunnosoperaatioita ...

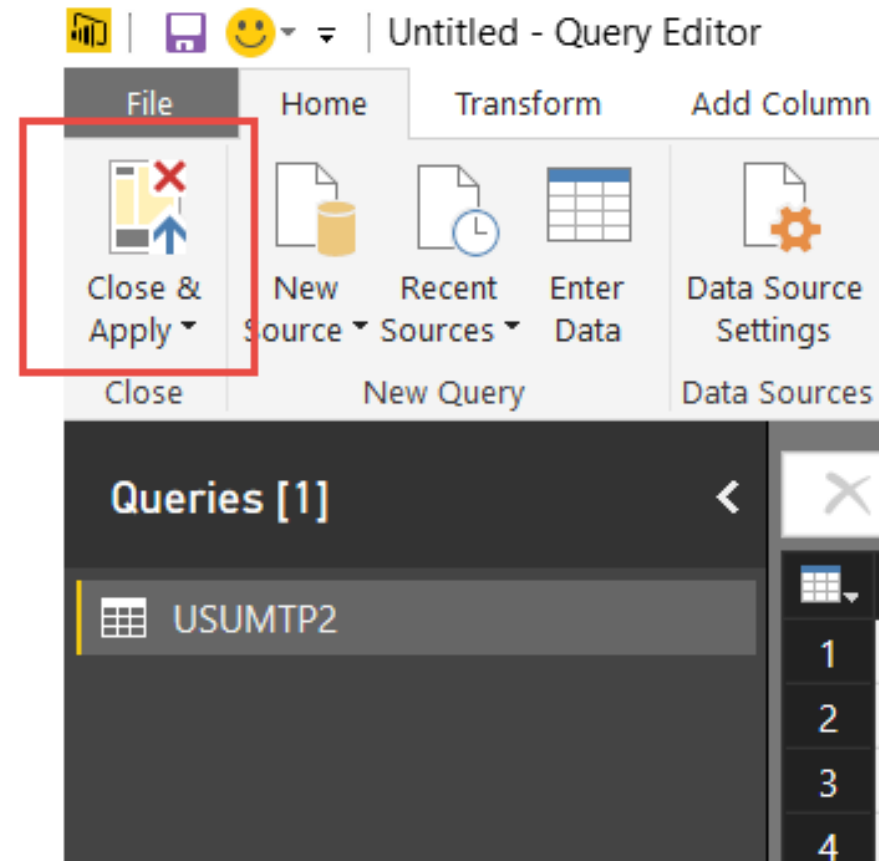
... ne näkyvät oikealla olevassa paneelissa.

Voit napsautella askeleita eli steppejä ja katsoa miltä data on näyttänyt missäkin vaiheessa. Voit myös poistaa tekemiäsi toimenpiteitä askelten edessä olevilla rasteilla.



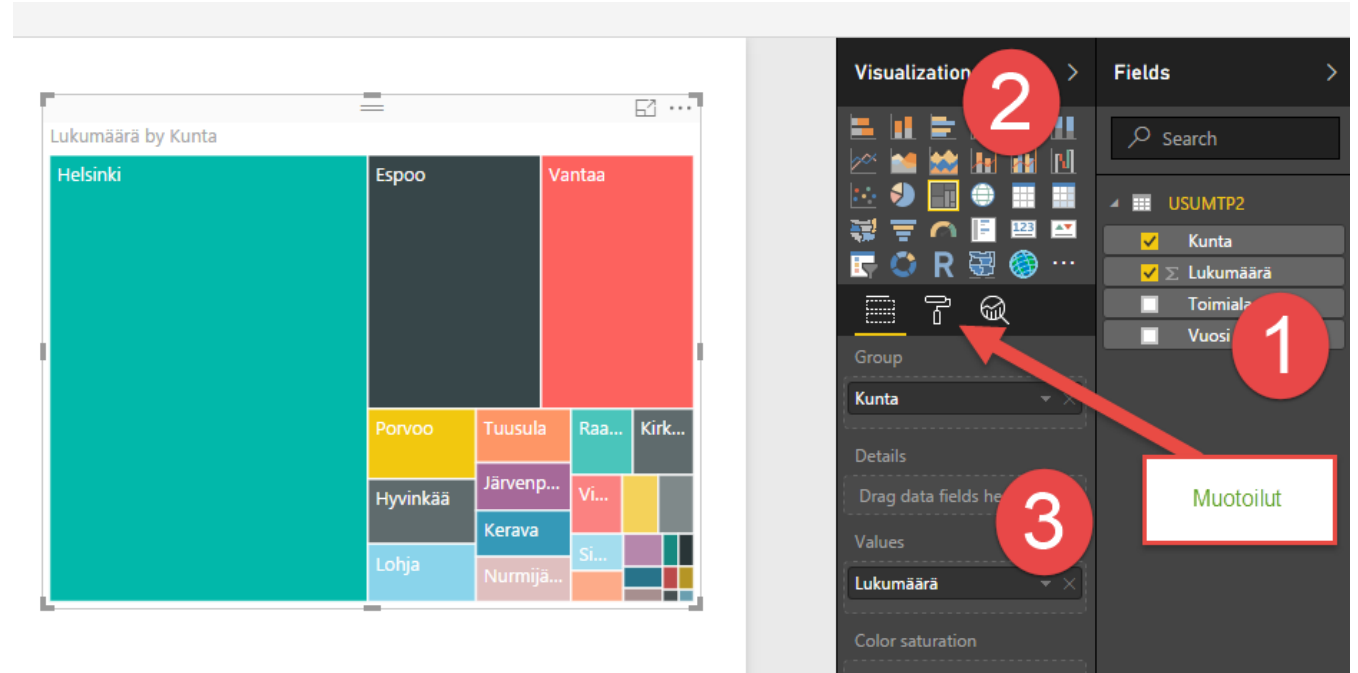
Suorita lopulta varsinainen datan lataus

Kun data on muokattu, se pitää ladata Power BI Desktopiin. Kun valitset **Home > Close & Apply**, tiedot ladataan, kyselyeditori sulkeutuu ja Power BI Desktop –ohjelman pääikkuna tulee jälleen esiin ja voit ryhtyä lisäämään malliin erilaista laskentaa sekä visualisointeja.



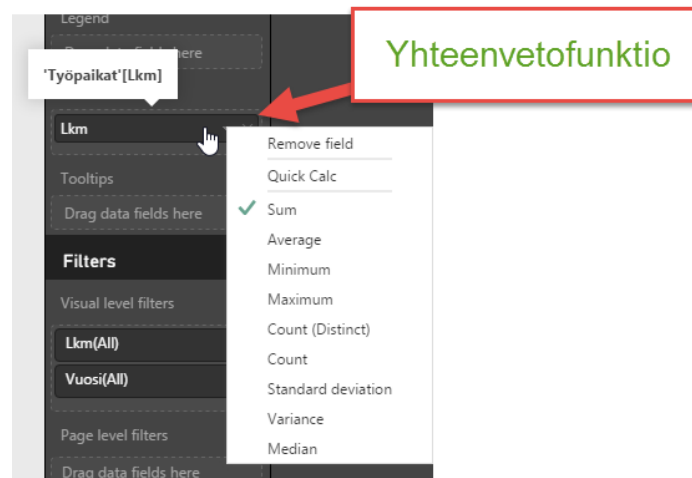
Visuaalien muodostaminen on helppoa

1. Vedät kentät raporttiin ja muodosta erilaisia visualisointeja.
2. Valitse haluamasi visualisointi.
3. Hienosäädä kentät oikeisiin paikkoihin.



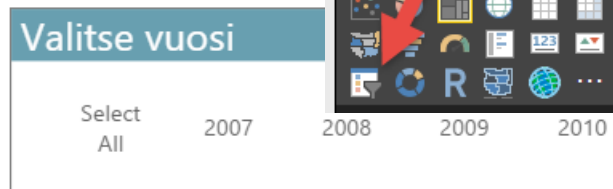
Maalitela-painikkeen avulla löydät runsaasti erilaisia muotoiluita.

Visualisoinnit ovat oletusarvoisesti keskenään vuorovaikuttavia.



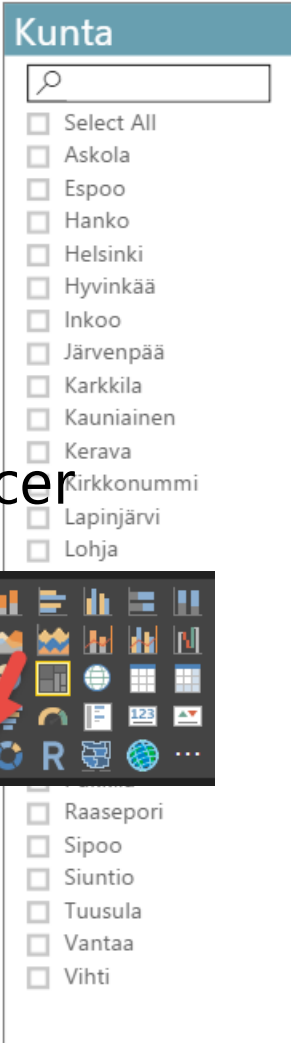
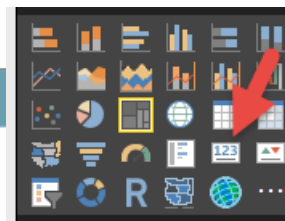
Monipuolisempia laskentasääntöjä voi luoda ns. DAX-kaavoina

Uudenmaan työpaikat



Slicer, joka
muotoiltu
maalitelan avulla
horisontaaliseksi

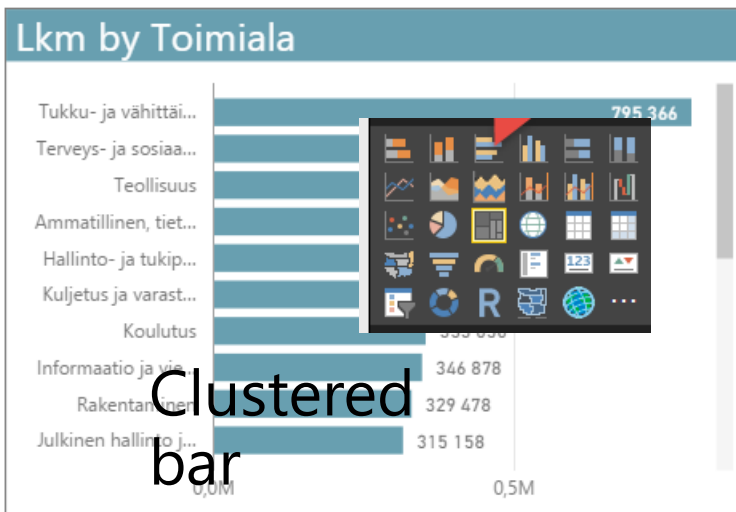
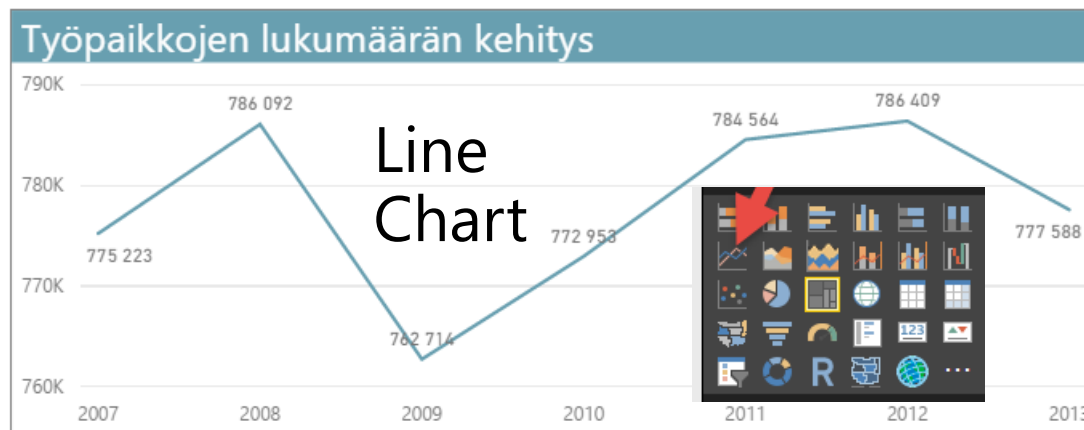
Card



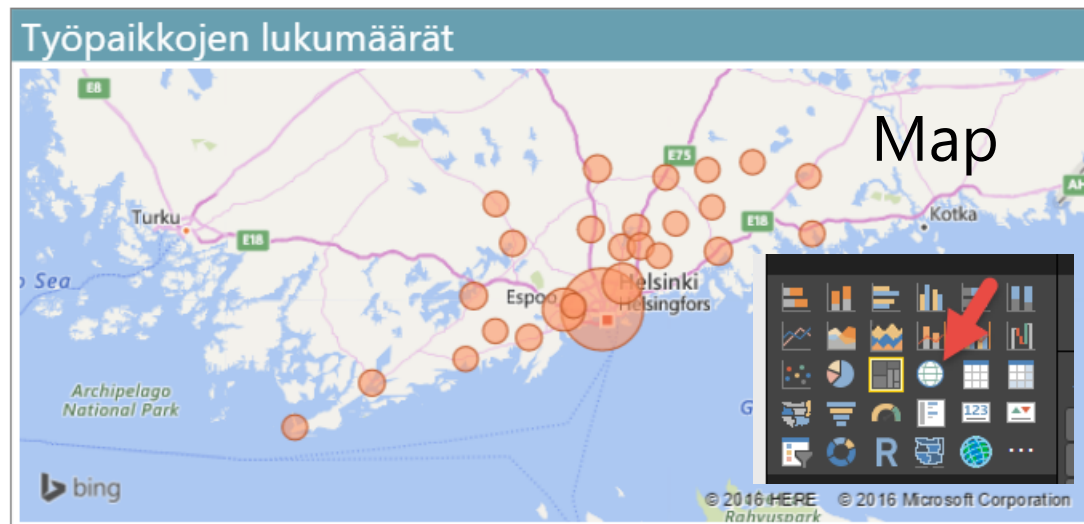
Slicer



Slicer



Clustered
bar

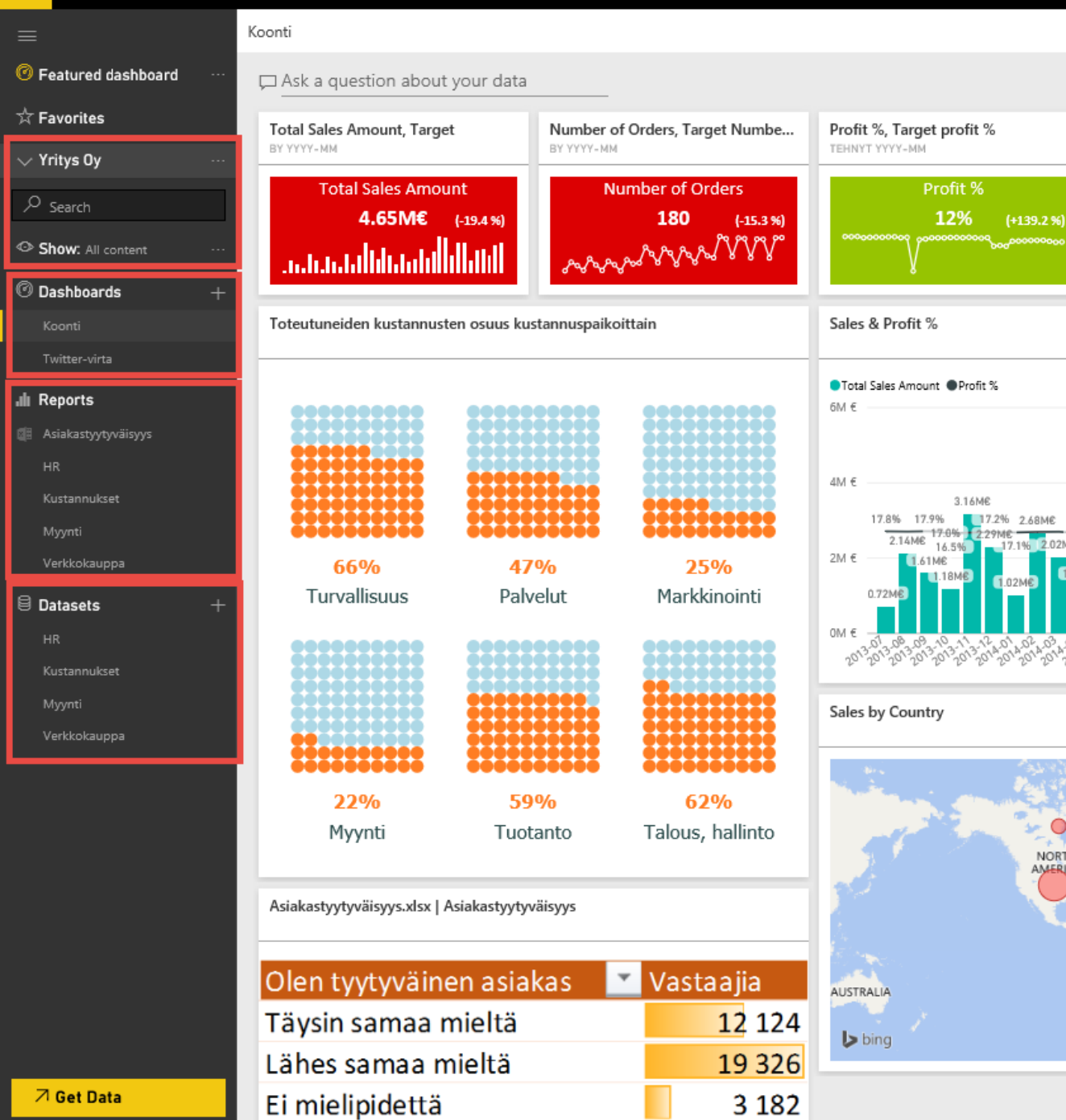


Esimerkkejä

- Vesikulkuneuvot
- Uudenmaan rakennukset ja työpaikat
- Suomalaiset postimerkit
- Miten suosittu etu- tai sukunimesi on
- Euroviisuvoittajat Power BI:ssä
- Hexcelligent-blogin kävijäkysely

Power BI -pilvipalvelu

- Kun visualisointiraportti on tehty, sen voi julkaista Power BI – pilvipalveluun komennolla **Home > Publish**
- Pilvipalveluun voi luoda ilmaisen tunnuksen powerbi.com –sivulta (Sign up free)
- Pilvipalvelussa olevia raporteista voi tehdä ns. koontisivuja (dashboard) ja niitä voi jakaa kollegoille tai muille Power BI – käyttäjille
- Raportista voi myös luoda julkisesti jaettavan raportin **File > Publish to web** –komennolla (kuten HExcelligentin julkiset raportit)
- Pilveen julkaistuja raportteja voi katsoa mobiiliappseilla



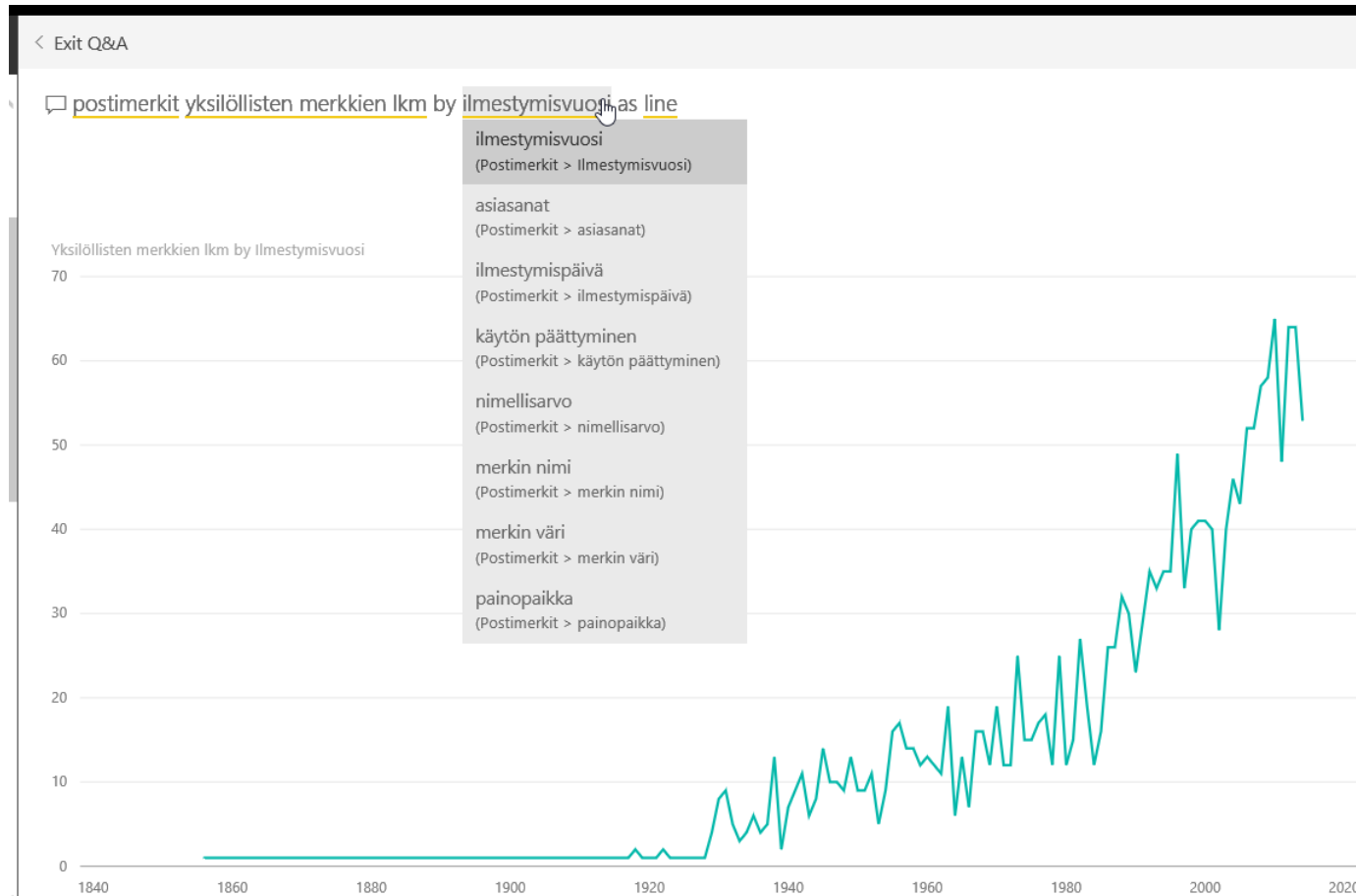
Power BI pilvipalvelu

- Oma henkilökohtainen työtila (My Workspace)
- Ryhmätyötilat (Group Workspaces)
- Koontinäytöt (Dashboards) ja Q & A kyselyt luonnollisella kielellä
- Raporttien ja koontinäytöjen jakaminen
- Raporttien ajastettu päivitys
- Mobiiliappsit

* Power BI Pro

Q & A

- Power BI –pilvipalvelussa voi tehdä kysymyksiä ja saada vastauksia datasta



Lisätietoja

- Powerbi.com > Learn > Guided Learning, Documentation
- Suomeksi "Heidin Power BI blogi" tai "Heidin Excel blogi"
 - <https://hexcelligent.fi/power-bi/>
- Esimerkkivisualisointeja Data Stories Gallery
 - <http://community.powerbi.com/t5/Data-Stories-Gallery/bd-p/DataStoriesGallery>
- Raportteihin voi ladata mukautettuja visuaaleja, joita on jo noin 70, kuten Infographic Designer
 - <https://app.powerbi.com/visuals/>

Kiitos!

- www.onsight.fi/kurssit/
 - Power BI Pro katsaus (1 pv)
 - Power BI Self-Service Workshop (2 pv)
 - Power Query – tiedon lataaminen Exceliin tai Power Query Desktopiin (1 pv)
 - Tulossa teemakurssit sekä visualisoinneista että kohderyhmälle BI-arkkitehdit

Microsoft Partner

Gold Learning
Gold Data Analytics
Gold Cloud Productivity
Gold Devices and Deployment
Gold Collaboration and Content
Gold Small and Midmarket Cloud Solutions
Silver Project and Portfolio Management