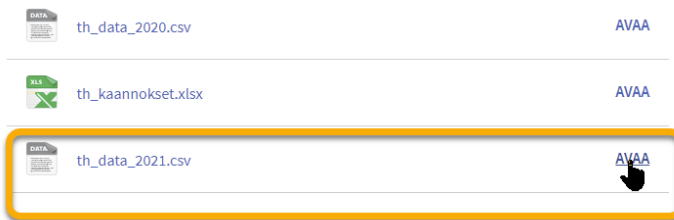


1 Valtion ostolaskut -aineiston lataus sekä kyselyn laatiminen

1.1 Aineiston lataus omaan työasemaan avoindata.fi -palvelusta

Harjoituksessa käytetään Hanselin julkaisemaa csv-muotoista valtion ostolaskudataa. Lataa ensin tarvittava csv-tiedosto työasemaasi avoindata.fi -sivuilta seuraavan ohjeen mukaisesti.

1. Siirry avoindata.fi -sivustolle: <https://www.avoindata.fi/data/fi/dataset/tutkihankintoja-data>
2. Lataa sivuilta löytyvä vuoden **th_data_2021.csv** -tiedosto itsellesi **AVAA**-painiketta napsauttamalla.

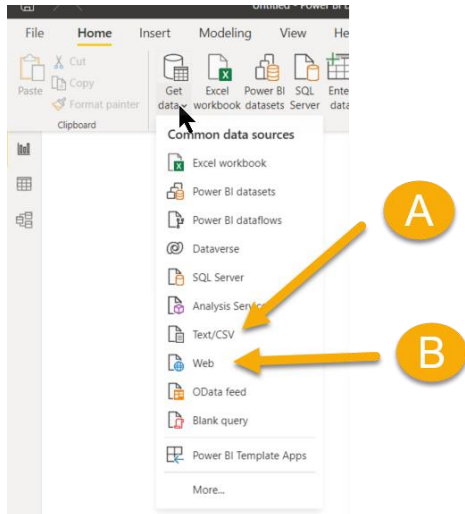


Jos nettiyhteytesi on kovin hidas ja lataus tuntuu kestävän kohtuuttoman pitkään, voit käyttää harjoituksessa myös kouluttajan toimittamaa pienempää muutaman kuukauden vajaata ostolaskuaineistoa (th_data_2021_otos.csv).

Kun tiedosto on latautunut työasemaasi eli **Downloads** (Ladatut tiedostot) -kansioon, kopioi tiedosto harjoitustiedostojen joukkoon.

1.2 Tietoja csv-datan lataamisesta

Datan lataaminen **csv**-tiedostosta tehdään eri tavoin siitä riippuen, onko tiedosto sinulla **levyasemassa** (kuten C: tai D: -asemassa) vai **M365-pilvisijainnissa** (eli https-osoitteessa) kuten SharePoint/Teams. Tässä harjoituksessa oletetaan, että tiedosto on sinulla jossakin levyasemassa.



A) Jos tiedosto on levyasemassa, kyselyn luonti aloitetaan komennolla **Text/CSV**.

B) Jos tiedosto on https-osoitteessa eli pilvisijainnissa, se aloitetaan komennolla **Web** (Verkko). Tässä tapauksessa sinulla pitää olla oikea tiedoston URL-osoite.

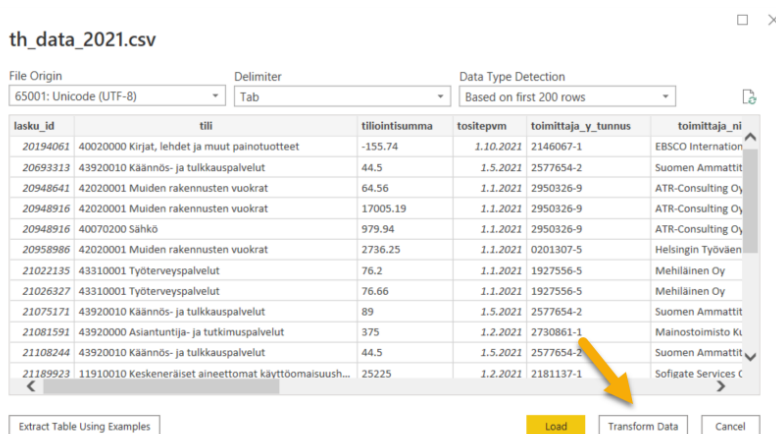
Vastaavasti jos lukisit dataa Excel-tiedostosta, aloittaisit Excel workbook -komennolla, jos tiedosto olisi levyasemassa. Muutoin käyttäisit Web-komentoa.

1.3 Kyselyn laatiminen

1.3.1 Levyasemassa olevan tiedoston lataamisen aloitus

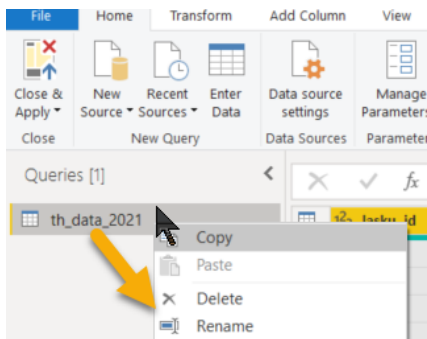
Kun tiedosto on levyasemassa (C:\, D:\, Y:\ tms.), kyselyn aloittaminen on helppoa. Riittää että tiedät missä levyasemassa ja kansiossa tiedosto on. Sinun ei tarvitse tehdä minkäänlaisia ennakkovalmisteluja.

- Käynnistä Power BI Desktop ja valitse **Home > Get Data** (Aloitus > Nouda tiedot) ja valitse **Text/CSV**.
- Etsi ja valitse työasemaasi ladattu .csv-tiedosto ja Open** (Avaa).
- Silmäile tiedoston sisältöä seuraavaksi avautuvassa ikkunassa. Jos esimerkiksi skandimerkit eli ääkköset ovat muuttuneet erikoismerkeiksi, valitse **File Origin** (Tiedoston alkuperä) -luettelosta vaihtoehto **65001: Unicode (UTF-8)**.

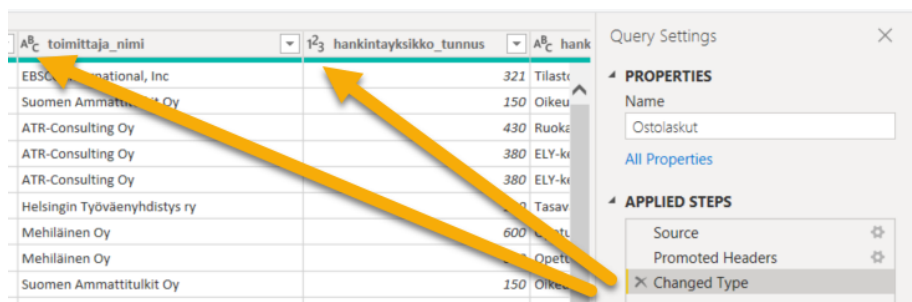


- Valitse **Transform Data** (Muunna tiedot), jotta pääset Power Query -kyselytyökaluun muokkaamaan dataa.

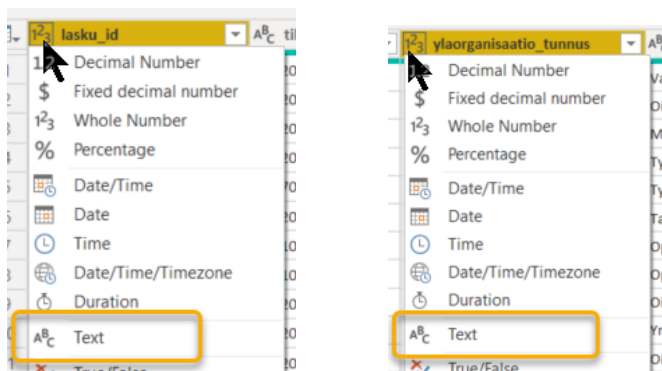
4. Kun olet päässyt Power Query -ikkunaan muokkaamaan dataa, vaihda kyselyn nimeksi **Ostolaskut** esimerkiksi pikavalikosta löytyvällä komennolla.



5. Huomaa, että oikealla näkyy kolme kyselyn askelta, joista kaksi ensimmäistä sinä olet saanut aikaan ja kolmannen Power Query on tehnyt tulkinna-varaisesti aineiston 200 ensimmäisen rivin perusteella:
- a. ensimmäinen **Source**-askel ottaa yhteyden valitsemaasi **csv**-lähdetiedostoon
 - b. toinen **Promoted Headers** -askel nostaa aineiston otsikot ylös otsikkoriviksi
 - c. kolmas askel muuttaa sarakkeiden tietotyyppä, eli yrittää tulkita millaista tietoa sarakkeissa, kuten teksti (ABC), kokonaisluku (123), desimaaliluku (1.2).



6. Tässä aineistossa osa sarakkeiden tietotyypeistä on tulkittu väärin, sillä 200 ensimmäistä riviä ei edusta riittävän hyvin koko aineiston sisältöä. Esimerkiksi **lasku_id** ja **ylaorganisaatio_tunnus** – sarakkeet on määritetty kokonaisluvuiksi (123), vaikka sarakkeissa on myös tekstiä eli H-alkuisia lasku_id -tunnuksia sekä XX-kirjaimilla merkittyjä yläorganisaatioiden tunnuksia. Jos tekstiä yritetään muuttaa kokonaisluvuiksi, kysely antaa myöhemmin virheilmoituksia. Vaihda siis molemmat sarakkeet tekstiksi sarakeotsikossa olevan painikkeen avulla:

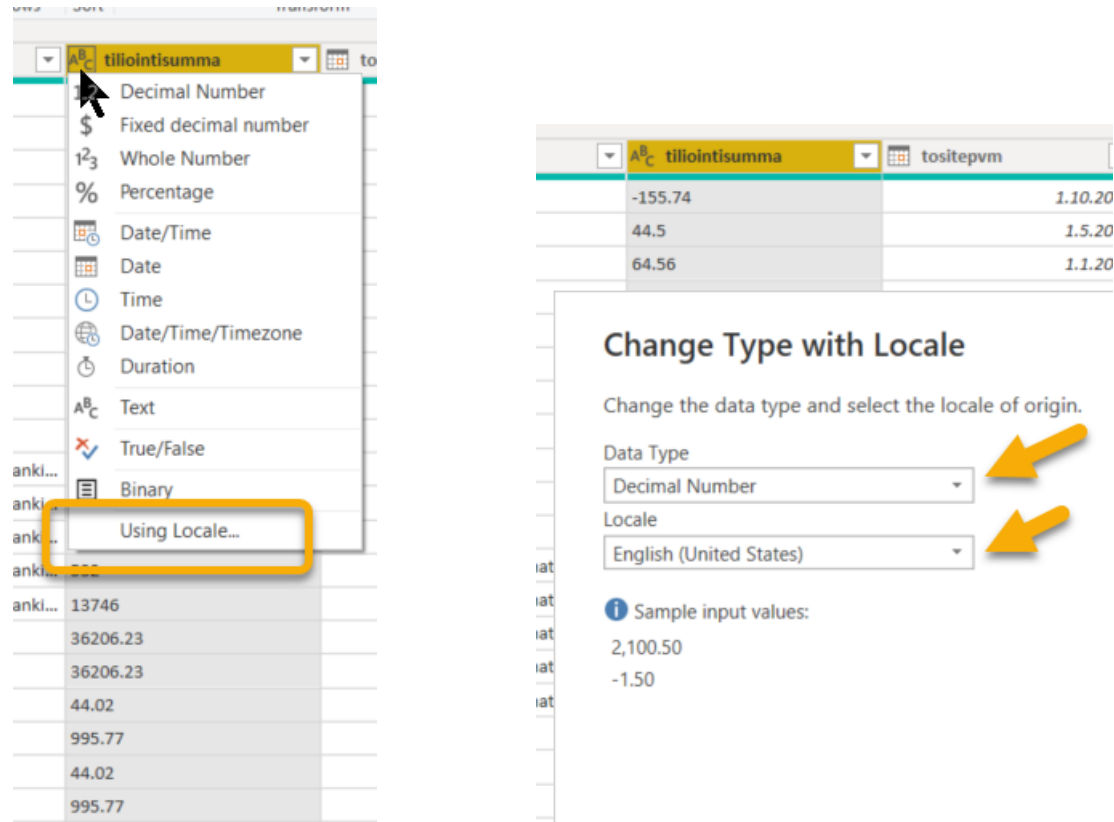


Kun vaihdat tietotyyppä, valitse että korvaat (replace) aiemmat määrittymiset.

Huomaa, että sinun on yleensä hyvä tuntee datasi, jotta tiedät millaista sisältöä eri sarakkeissa on ja osaat määrittellä tietotyytit oikein.

7. Jos **tiliointisumma**-sarakkeen luvut olisivat suomalaisessa muodossa eli niissä olisi käytetty desimaalipilkkuja, ne olisivat automaattisesti muuttuneet desimaaliluvuiksi. Hansel on julkaissut datan yhdysvaltalaisessa muodossa, joten sinun pitää käyttää tietotyyppivalikossa alinna olevaa **Using Locale** (Käytä aluekohtaisia asetuksia) -komentoa tietotyypin muuttamiseen.

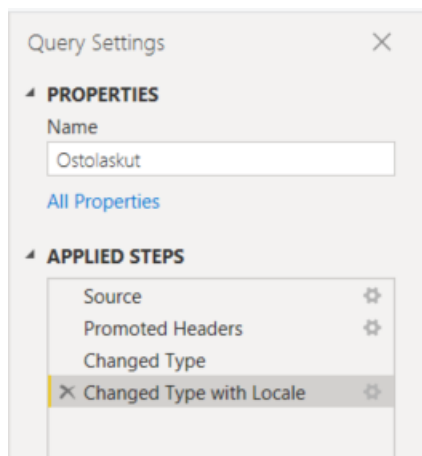
Odota rauhassa, ikkunan avautuminen kestää hetken. Määritä, että luvut ovat desimaalilukuja ja että ne ovat yhdysvaltalaisessa muodossa:



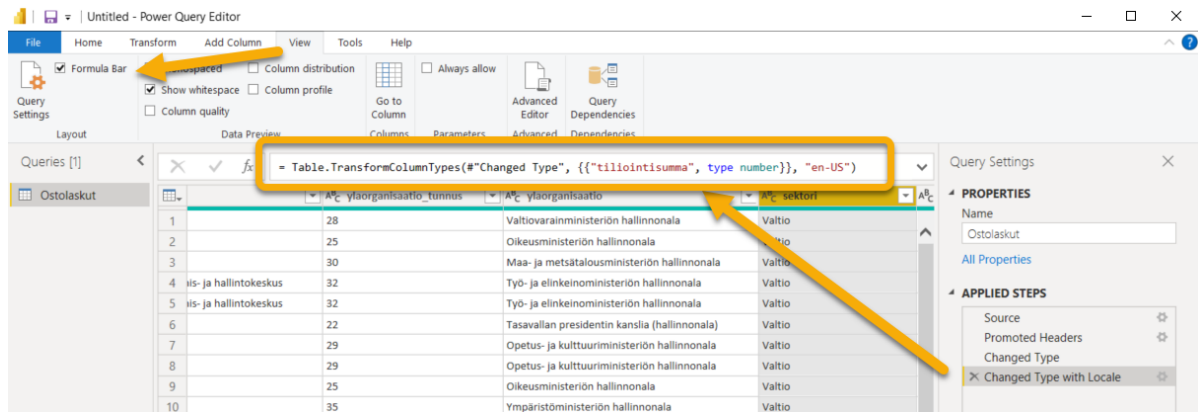
The image shows two screenshots from a data management application. The left screenshot shows a dropdown menu for the 'tiliointisumma' column. The menu options include Decimal Number, Fixed decimal number, Whole Number, Percentage, Date/Time, Date, Time, Date/Time/Timezone, Duration, Text, True/False, and Binary. The 'Using Locale...' option is highlighted with a yellow box. The right screenshot shows a table with two columns: 'tiliointisumma' and 'tositepvm'. The 'tiliointisumma' column contains values: -155.74, 44.5, and 64.56. The 'tositepvm' column contains dates: 1.10.20, 1.5.20, and 1.1.20. Below the table is a dialog box titled 'Change Type with Locale'. The dialog box has a title bar and a close button. It contains the text 'Change the data type and select the locale of origin.' Below this text are two dropdown menus: 'Data Type' (set to 'Decimal Number') and 'Locale' (set to 'English (United States)'). There are two yellow arrows pointing to these dropdown menus. Below the dropdown menus is a section titled 'Sample input values:' with the following values: 2,100.50 and -1.50.

Hyväksy OK:lla.

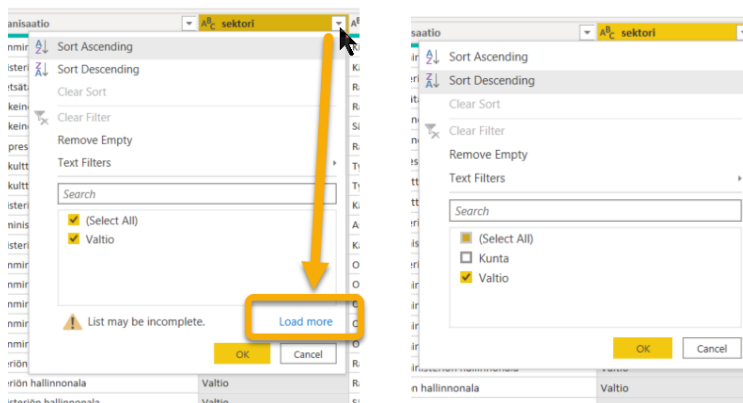
8. Kiinnitä huomiota, että kyselyysi oikealle ilmestyi uusi askel.



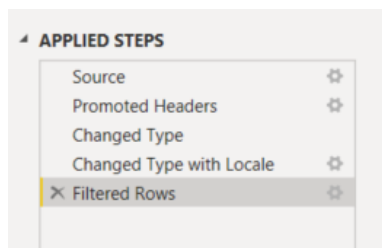
9. Tuo Power Queryyn esiin kaavarivi komennolla **View > Formula Bar** (Näytä > Kaavarivi). Jokainen askel muodostuu joko yhdestä tai useammasta sisäkkäin olevasta M-kielisestä funktiosta. Alla on esimerkkinä jälkimmäisen tietotyyppimuunnos-askelen koodi. Muodostuvaa koodia ryhtyä silmäilemään alusta lähtien, sillä tarvitset vähintään koodin lukutaitoa erilaisten pulmatilanteiden ratkomisessa.



10. Aineistossa on sekä valtion ostolaskuja että joidenkin kuntien ostolaskuja. Tässä raportissa halutaan näyttää vain valtion ostolaskuja, joten suodata pois muut. Lataa ensin suodatusluetteloon näkyviin enemmän arvoja ja jätä dataan vain **Valtion** ostolaskut.

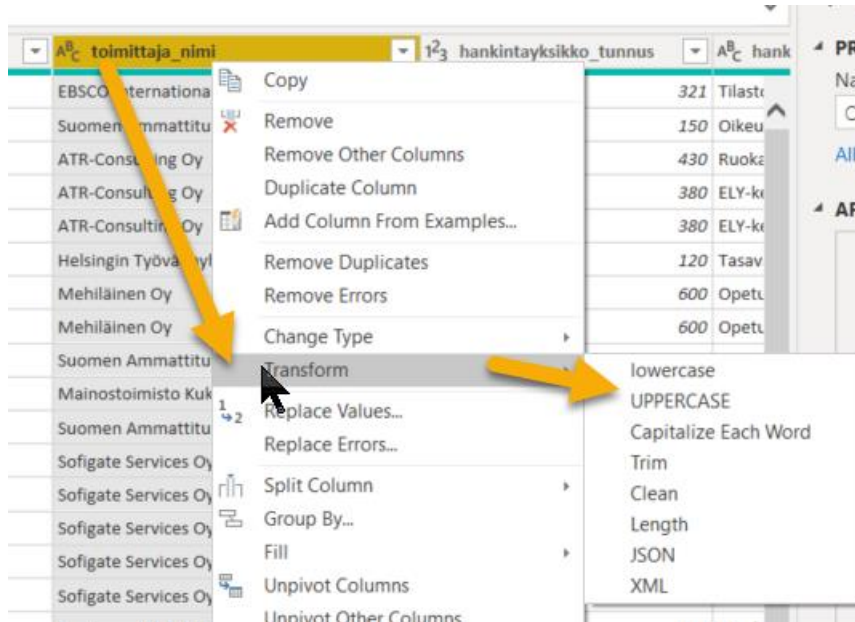


11. Huomaa, että kyselyysi ilmestyi uusi askel. Silmäile kaavariviltä millainen M-koodi taustalla on.
12. Tässä vaiheessa kyselyssä pitäisi olla viisi vaihetta:



13. Napsauta kutakin vaihetta **Source**-vaiheesta alkaen ja katso mitä kaavarivillä lukee ja miltä data näyttää vaiheen seurauksena. Voit kasvattaa kaavarivin korkeutta sen oikeassa reunassa olevan nuolen avulla.

14. **Kokeile:** napsauta **toimittaja_nimi** -saraketta ja konvertoi se isoiksi kirjaimiksi. Huomaa, että askeleiden joukkoon ilmestyy uusi. Lue kaavariviltä koodi.



Poista juuri muodostunut vaihe vaiheen nimen edessä olevalla rastilla, jolloin tekstit palautuvat alkuperäisiksi. Voit siis helposti opetella Power Queryn käyttöä kokeilemalla erilaisia komentoja, sillä pystyt helposti poistamaan vaiheita.

15. **Lisätehtävä:** Vaihda halutessasi seuraavien sarakkeiden otsikot kaksoisnapsauttamalla niitä.

Alkuperäinen otsikko	Uusi otsikko
tiliointisumma	tiliöintisumma
toimittaja_y_tunnus	toimittajan y-tunnus
toimittaja_nimi	toimittajan nimi
ylaorganisaatio	yläorganisaatio
tuote_palveluryhma	palveluryhmä

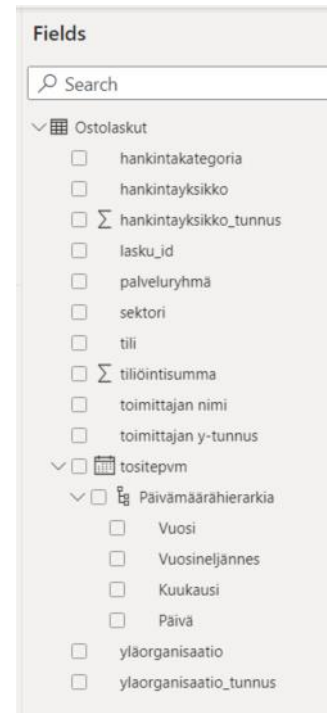
16. Tarkista vielä, että data näyttää hyvältä ja tietotyytit ovat oikein. Huom! lasku_id ja ylaorganisaatio_tunnus -kenttien pitää olla tekstiä. Onhan kyselyssäsi ainakin ensimmäiset viisi askelta ja mahdollisesti myös edellä mainittu sarakkeiden nimeämisaskel.
17. Suorita kysely eli lataa data csv-tiedostosta Power BI Desktop -tiedostoon komennolla **Home > Close & Apply** (Aloitus > Sulje ja ota käyttöön). Power Query sulkeutuu ja data latautuu tiedostoon.

Odota rauhassa, kyselyn suoritus kestää hetken.

18. Jos lataus onnistuu, siitä ei tule erillistä ilmoitusta. Vain virheistä ilmoitetaan. Kerro kouluttajalle, jos saat virheilmoituksen.
19. Tallenna PBI-tiedosto haluamaasi kansioon nimellä **Valtion ostolaskut.pbix**.

Kyselyn suorituksen jälkeen **Ostolaskut**-taulu sekä taulun sarakkeet eli kentät näkyvät Power BI Desktop -pääikkunan oikeassa reunassa **Fields** (Kentät) -paneelissa. →

Huomaa, että numeeristen kokonais- ja desimaalilukukenttien edessä on Σ -merkki ja että **tositepvm**-kentästä on muodostunut automaattisesti päivämäärähierarkia (Vuosi – Vuosineljännes – Kuukausi – Päivä).



Käy silmäilemässä tiedostoon lataamasi aineiston sisältöä **Data** (Tiedot) -näkyvässä. ↓

lasku_id	tili	tiliöintisumma	tositepvm	toimittaja
23657712	42020001 Muiden rakennusten vuokrat	3 017	1.10.2021	1503388-
23657713	42020001 Muiden rakennusten vuokrat	6 342	1.10.2021	1503388-
23657714	42020001 Muiden rakennusten vuokrat	1 927	1.10.2021	1503388-
23657715	42020001 Muiden rakennusten vuokrat	4 731	1.10.2021	1503388-
23657716	42020001 Muiden rakennusten vuokrat	1 233	1.10.2021	1503388-
23657719	42020001 Muiden rakennusten vuokrat	1 233	1.10.2021	1503388-
23657720	42020001 Muiden rakennusten vuokrat	11 663	1.10.2021	1503388-
23657721	42020001 Muiden rakennusten vuokrat	118	1.10.2021	1503388-

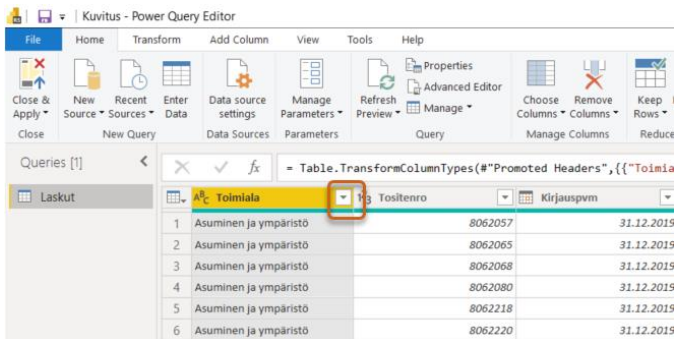
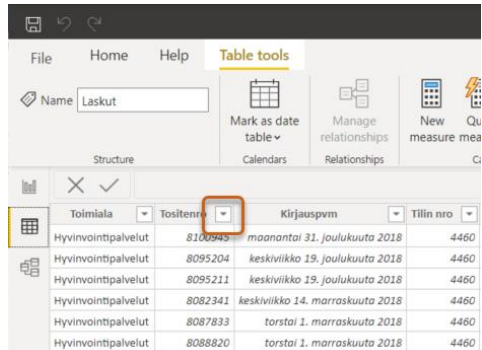
Valitse **Data**-näkyvässä esimerkiksi **tositepvm**-sarake ja huomaa, että näet mm. rivimäärän tilarivillä näytön vasemmassa alakulmassa.

Tilarivillä näytetään myös valitussa sarakeessa esiintyvien erilaisten arvojen lukumäärä (distinct count).

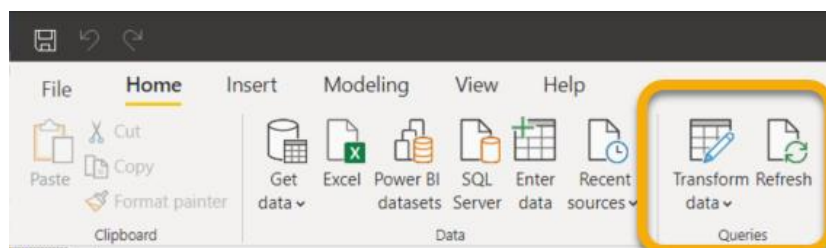
Kun valitsit sarakkeen, valintanauhaan ilmestyi **Saraketyökalut** (Column tools) -välilehti, jonka avulla voit esimerkiksi muotoilla sarakkeiden sisältöä. Muotoile **tositepvm**-sarakkeen päiväys lyhyempään muotoon.

Muotoile myös tiliöintisumma-sarake siten, että siinä näytetään tuhaterotin ilman desimaaleja.

Power Query -näkömön ja yllä olevan Data (Tiedot) -näkömön erot

Power Query	Power BI Desktop -pääikkunan Data (Tiedot) -näkömön
<p>Power Query -näkömön tarkastellaan ja muokataan dataa ennen kuin se ladataan tiedostoon.</p> <p>Power Queryssä määritetään siis, millaista dataa tullaan lataamaan Power BI Desktop -tiedostoon.</p> <p>Jos suodatat dataa Power Queryssä, tiedostoon latautuu fyysisesti vähemmän rivejä.</p> 	<p>Data (Tiedot) -näkömön tarkastellaan tiedostoon ladattua dataa.</p> <p>Data -näkömön ladattua dataa voi mm. muotoilla (desimaalien määrä, päiväysten ulkoasu).</p> <p>Jos suodatat dataa Data-näkömön, rivit eivät fyysisesti poistu. Suodatusnuolia käytetään vain datan tutkimiseen.</p> 

Jos sinun pitää palata muuttamaan Power Query -kyselyä, pääset Power Queryyn komennolla **Home > Transform data** (Aloitus > Muunna tiedot). Voit päivittää datan **Refresh** (Päivitä) -toiminnolla, jolloin Power Query lukee tiedostosta tiedot uudelleen.



Jos sinun pitäisi vaihtaa tietolähdettä eli csv-tiedoston nimi tai sijainti muuttuisi, löydät **Transform data** (Muunna tiedot) -painikkeen valikosta **Data Source Settings** (Tietolähdeasetukset) -komennon.

Muista tallentaa tiedosto.

Jos koulutus järjestetään etänä, kuittaa kouluttajalle Teamsin chatiin esim. sanalla "Valmis", kun olet tehnyt harjoituksen.