*Before doing any practice, make sure to use your Microsoft 365 username for sign-in Power BI and other Microsoft products:*

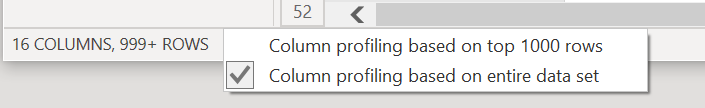
Graphical user interface, text, application, email

Description automatically generated

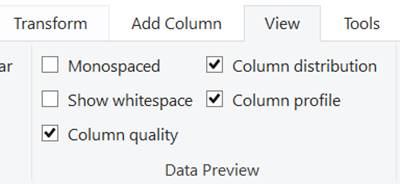
# **Toronto Neighbourhoods**

# **Practice 1 – Create Airbnb Power BI Report**

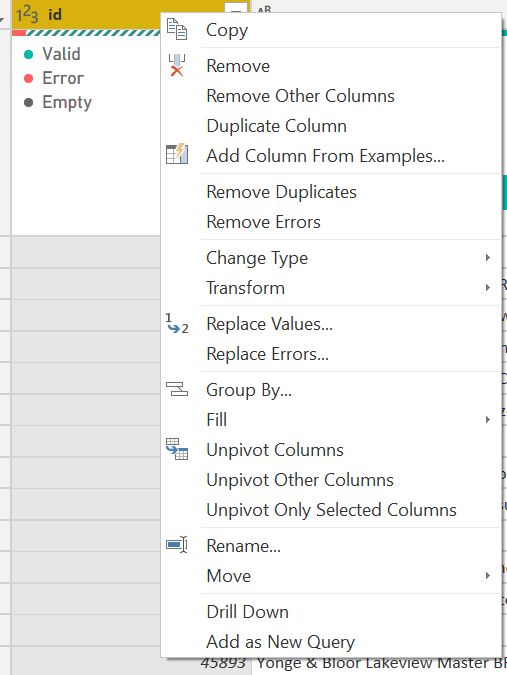
1. Get the data:
   * **listings.csv**: summary information and metrics for listings in Toronto
     + Direct download link: <http://data.insideairbnb.com/canada/on/toronto/2021-02-08/visualisations/listings.csv>
   * **reviews.csv**: Summary Review data and Listing ID
     + Direct download link: <http://data.insideairbnb.com/canada/on/toronto/2021-02-08/visualisations/reviews.csv>
2. Clean the **listings.csv** data (Transform)
   * Before we clean the data, please make sure to select ‘Column profiling based on entire data set’ at the bottom of the Power Query Editor page



* + To help us clean data, select ‘View’ and click ‘Column distribution’, ‘Column profile’, ‘Column quality’, so that it will show columns’ profiles which help us identify data issues:



* + Remove Error for ‘id’, ‘host\_id’;



* + Remove Empty for ‘nieghtbourhood’

Graphical user interface, text, application

Description automatically generated

* + Remove column ‘neightbourhood\_group’

1. Link ‘reviews’ together with ‘listings’ and edit the relationship
   * id from ‘listings’
   * listing\_id from ‘reviews’
   * Right click ‘Properties’ to edit relationship

Graphical user interface, application

Description automatically generated

* + create a Table which has 2 columns – id(or listing\_id) and price (calculate average) to see the differences of Single or Both filters:

**Single filter direction:**

**Table

Description automatically generated**

**Both filter directions:**

Table

Description automatically generated

1. Create calculated column:
   * Annual availability: annual\_avl% = listings[availability\_365] / 365
   * host\_type = if(listings[calculated\_host\_listings\_count]>1, "Multi-listing host", "Single-listing host")
2. Create measure:
   * Avg\_List\_per\_Host = count(listings[id])/DISTINCTCOUNT(listings[host\_id])
3. Create one page dashboard to answer the following questions:
   * What is the average listing price per night? Total number of listings in Toronto?
     + Use Card to show each metric; turn off Category label and turn on Title
     + Average listings per Host? In Data section, you can change the format for each column

Graphical user interface, text, application

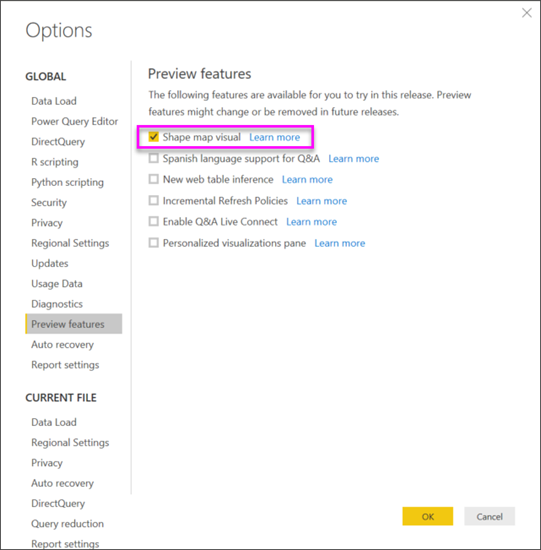
Description automatically generated

* + Line chart showing the count of reviews over time => How was the Airbnb business affected by Covid-19?
    - Use basic filtering to only select year 2015 -2021
  + Average price and number of listings by room type (Donut chart)
  + What is the overall annual availability rate?
  + Most hosts have multiple listings or single listing?
    - Use Pie chart and count distinct host\_id
  + Create a map showing each listing’s location
  + Calculate number of listings and average price for each Neighbourhood
    - Use Table

1. Create shape map:
   * Download neightborhood GeoJSON file:

<http://data.insideairbnb.com/canada/on/toronto/2021-04-09/visualisations/neighbourhoods.geojson>

* + Go to <https://mapshaper.org/> and use this website to convert the geojson file you just download to a TopoJSON file
    - Import and export as TopoJSON file
  + To enable Shape Map, select File > Options and Settings > Options > Preview Features, then select the Shape Map Visual checkbox



* + Select ‘Shape Map’ and drag ‘neighborhood’ to ‘Location’; click Format and select ‘+ Add map’ under ‘Shape’ section

Table

Description automatically generated

* + Upload the TopoJSON file (.json) file you just converted to add map
  + Add columns for ‘Color of saturation’ and ‘Tooltips’ to complete the shape map

Map

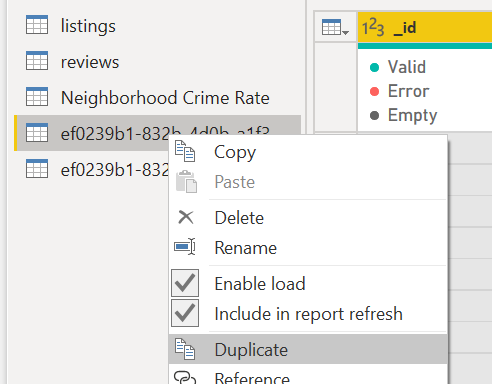
Description automatically generated

# **Practice 2 – Prepare data Neighborhood Crime Rates**

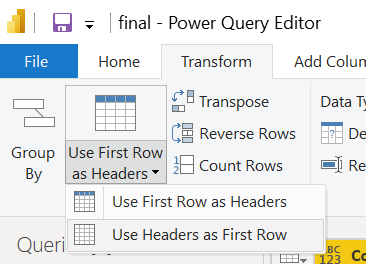
1. Get the data:
   * Neighbourhood Crime Rates: this dataset includes the Crime Data by Neighborhood
     + Direct download link: <https://ckan0.cf.opendata.inter.prod-toronto.ca/download_resource/4754baf5-3715-4166-a347-eda9813521de?format=csv&projection=4326>
2. Transform in Query Editor
   * Before we clean the data, please make sure to select ‘Column profiling based on entire data set’ at the bottom of the Power Query Editor page
   * Only **keep** columns by using ‘Choose Columns’:
     + Neighbourhood
     + Hood\_ID
     + F2020\_Population\_Projection
     + Assault\_2020
     + AutoTheft\_2020
     + BreakAndEnter\_2020
     + Robbery\_2020
     + TheftOver\_2020
     + Homicide\_2020
     + Shootings\_2020
   * Rename column ‘Hood\_ID’ to ‘Neighbourhood\_id’
   * Add Column => Custom Column => create a new column:
     + crime\_rate% = ([Assault\_2020] + [AutoTheft\_2020] + [BreakAndEnter\_2020] + [Robbery\_2020] + [TheftOver\_2020] + [Homicide\_2020] + [Shootings\_2020])/[F2020\_Population\_Projection]
     + Don’t forget to change to percentage format in Transform
   * Remove columns which you do not need:
     + Assault\_2020, Robbery\_2020, TheftOver\_2020, Homicide\_2020, Shootings\_2020, AutoTheft\_2020, BreakAndEnter\_2020
   * Rename data to ‘Neighborhood Crime Rate’

# **Practice 3 – Prepare data Neighborhood Income and Neighborhood Diversity**

1. Get the data:
   * **Neighbourhood Profiles**: people and household profile data based on Statistics Canada for the 2016 Census
     + Direct download link: <https://ckan0.cf.opendata.inter.prod-toronto.ca/download_resource/ef0239b1-832b-4d0b-a1f3-4153e53b189e?format=csv>
2. Complete transform in Query Editor to prepare **Neighborhood Income** and **Neighborhood Diversity**
   * Rename and Duplicate the dataset respectively – one for Neighbourhood Income, the other for Neighbourhood Diversity



* + Let’s pick ‘Topic’ – ‘**Income of households in 2015**’ to prepare data **‘Neighbourhood Income’**
    - Use ‘Topic’ column to filters, only keep ‘Income of households in 2015’ and ‘Neighbourhood Information’
    - Remove columns ‘\_id’, ‘Topic’, ‘Data Source’, ‘Category’, and ‘City of Toronto’
    - Demoted Headers by select ‘Use Headers as First Row’ (you must do this before Transpose step):



* + - **Transpose**
    - Promoted Header by select ‘Use First Row as Headers
    - Remove columns we don’t need, and only keep column: ‘**Characteristic’, ‘Neighbourhood Number, *‘Under $5000\_1’, ‘$5,000 to $9,999\_2’, ‘$10,000 to $14,999\_3’……. $100,000 and over\_16’***
    - Change Column types:
      * ‘Characteristic’: type ‘Text’
      * ‘Neighbourhood Number’: type ‘Whole Number’
      * *‘Under $5000\_1’, ‘$5,000 to $9,999\_2’, ‘$10,000 to $14,999\_3’……. $100,000 and over\_16’: type ‘Whole Number’*
    - Rename columns:
      * ‘Characteristic’ to ‘Neighbourhood’
      * ‘Neighbourhood Number’ to ‘**Neighbourhood\_id**’
    - Create Custom Column:

% of $100K+Income = percentage of household with annual income *$100,000 and over (use this as a proximate way to assess income level for each neighborhood):*

* + - * *% of $100K+Income = [#" $100,000 and over\_16"]/([#" Under $5,000\_1"]+[#" $5,000 to $9,999\_2"]+[#" $10,000 to $14,999\_3"]+[#" $15,000 to $19,999\_4"]+[#" $20,000 to $24,999\_5"]+[#" $25,000 to $29,999\_6"]+[#" $30,000 to $34,999\_7"]+[#" $35,000 to $39,999\_8"]+[#" $40,000 to $44,999\_9"]+[#" $45,000 to $49,999\_10"]+[#" $50,000 to $59,999\_11"]+[#" $60,000 to $69,999\_12"]+[#" $70,000 to $79,999\_13"]+[#" $80,000 to $89,999\_14"]+[#" $90,000 to $99,999\_15"]+[#" $100,000 and over\_16"])*

Change the column type to ‘Percentage’

* + - Use ‘Standard’ in ‘Add Column’ to double check calculation
    - Only keep columns: ‘Neighbourhood’, ‘Neighbourhood\_id’, and ‘% of $100K+Income’
    - Rename data name to ‘Neighbourhood Income’
  + Let’s pick another ‘Topic’ – ‘**Mother tongue**’ to prepare data **‘Neighbourhood Diversity’**
    - Keep ‘Mother tongue’ and ‘Neighbourhood Information’
    - Remove column ‘\_id’, ‘Topic’, ‘Category’, ‘Data Source’ and ‘City of Toronto’
    - Keep 4 rows:

1. Neighbourhood\_Number

2. Mother tongue for the total population excluding institutional residents

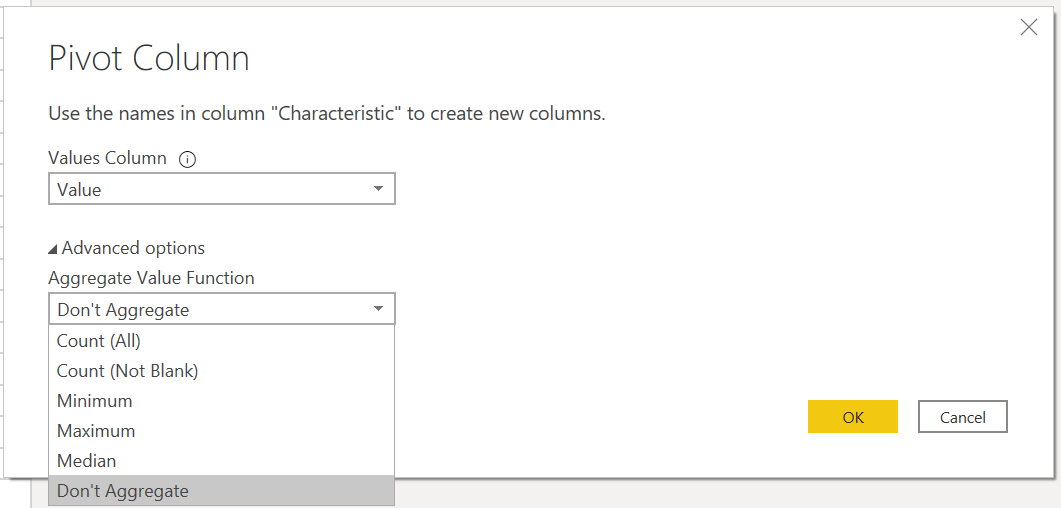
3. Official langurages

4. Non-official langurages

* + - **Unpivoted all Neighbourhood columns**

Select all columns after Characteristic

* + - **Pivot Characteristic column**
      * **Make sure to use value as ‘Values Columns’**
      * Don’t aggregate



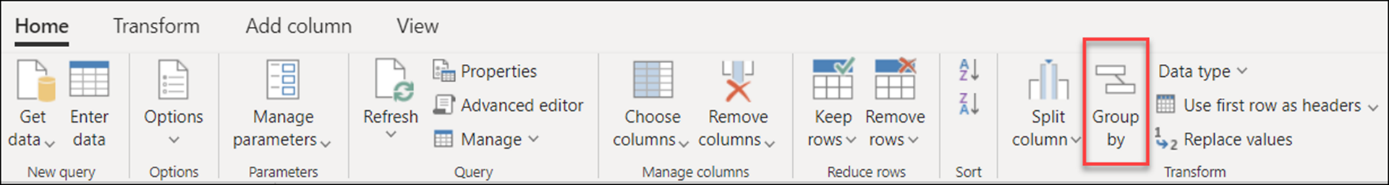
* + - **C**hange all numerical columns’ type to ‘Whole Number’
    - Rename columns:
      * ‘Attribute’ to ‘Neighbourhood’
      * ‘Neighbourhood Number’ to ‘Neighbourhood\_id’
    - Create Custom Colum:
      * *% of nonofficial languages = [#" Non-official languages"]/[Mother tongue for the total population excluding institutional residents]*

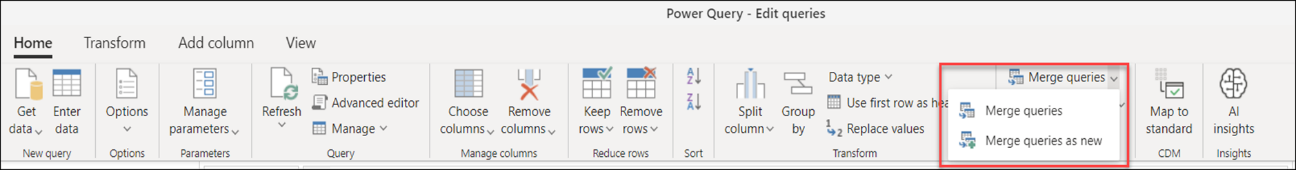
Change the column type to ‘Percentage’

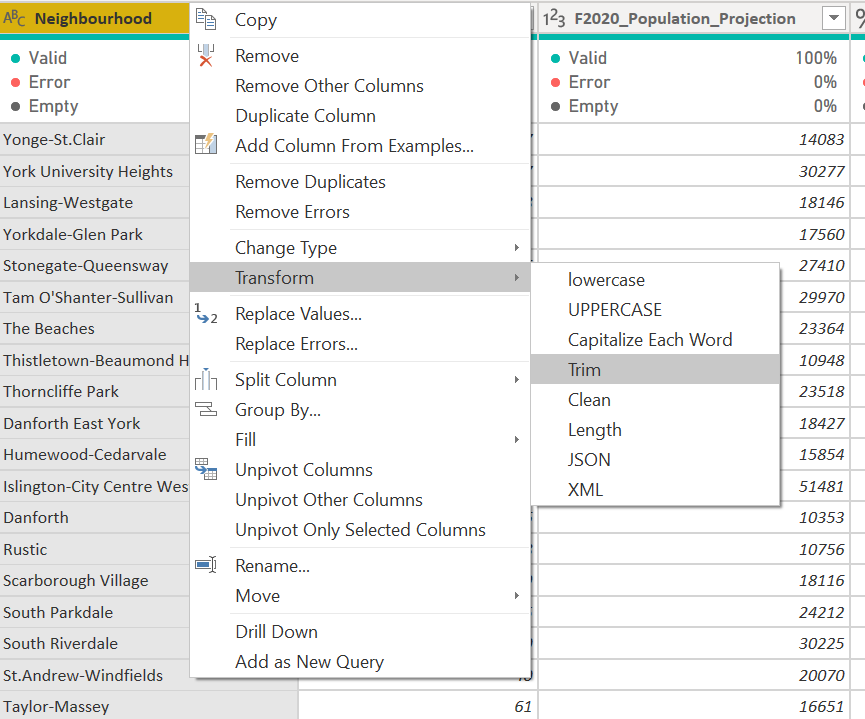
* + - Only keep columns: ‘Neighbourhood’, ‘Neighbourhood\_id’, and ‘% of nonofficial languages ’
    - Rename query name to ‘Neighbourhood Diversity’

# **Practice 4 - Toronto Neighbourhood Rating Tool**

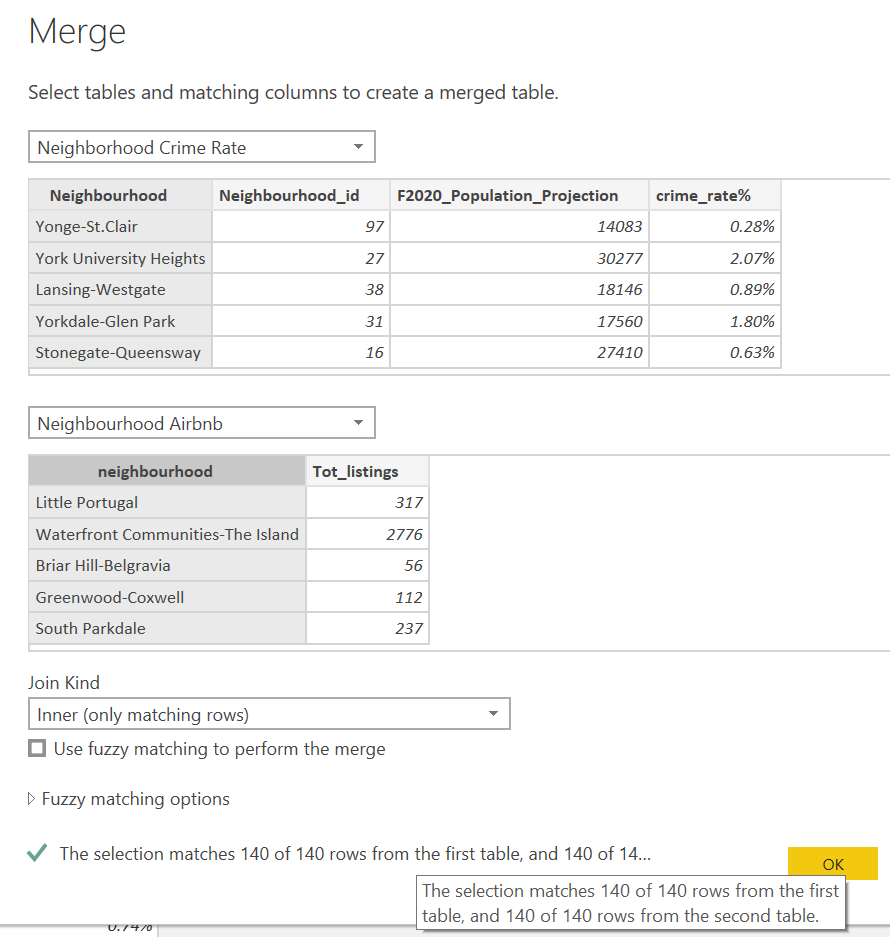
1. Transform in Query Editor
   * **Duplicate** listings to createa new query – Neighbourhood Airbnb
   * **Group by** Neighbourhood and **count** the number of rows as ‘Tot\_listings’



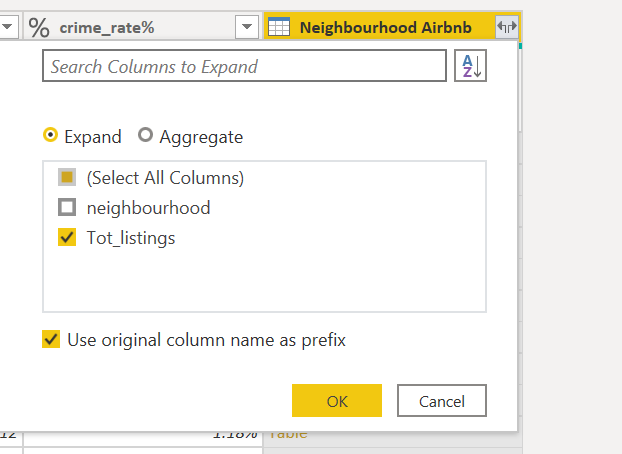
* + **Join** Neighbourhood Income, Neighbourhood Diversity, Neighbourhood Airbnb, and Neighbourhood Crime Rate all together
    - Use **Merge Queries as New** to inner join Neighbourhood Airbnb and Neighbourhood Crime Rate 
      * We need to use neighborhood name to join these 2 datasets
      * In real situation, we should avoid using character columns to join
      * If we have to use character columns to join, we should trim text and Uppercase Neighbourhood columns first



* + - Check whether it is 100% match



* + - After these 2 tables are joined together, we only keep the columns we need (Tot\_listings from Neighbourhood Airbnb:

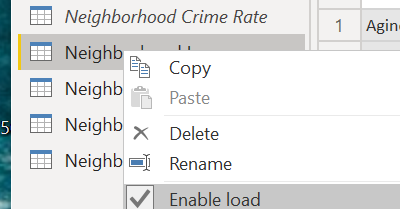


* + - Then, let us continue to Merge Queries:
      * use Neighbourhood\_ID to join other tables together as a new query – **Neighborhood Rating**
    - Adjust column format and rename column names as below:
      * Neighborhood
      * Neighborhood\_id
      * Population
      * Crime\_rate%
      * Tot\_listings
      * %of nonofficial languages
      * %of $100K+Income

Graphical user interface, table

Description automatically generated

* + - Before click ‘Close & Apply’, disable load (unclick Enable load) for Neighborhood Airbnb, Neighborhood Income, Neighborhood Crime Rate and Neighborhood Diversity because we don’t need them anymore



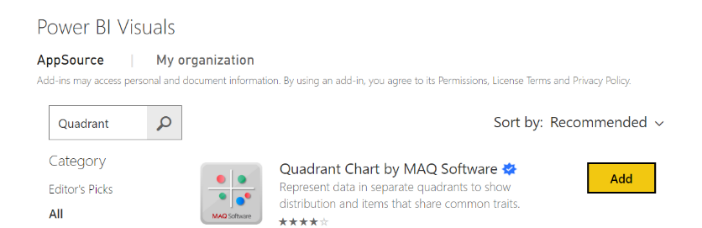
* + - Now we are ready to create dashboard!

1. Create a Shape Map to show neighborhood profile by using the criteria we just created

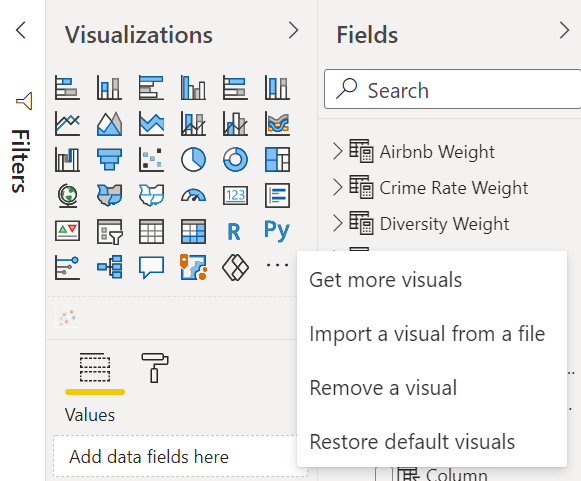
Diagram

Description automatically generated with medium confidenceGraphical user interface, application, PowerPoint

Description automatically generated

1. Optional: create a Quadrant Chart by importing from Custom Visuals

*Note: a quadrant chart is a scatter plot with the background split into four equal sections (quadrants). The purpose of the quadrant chart is to group values into distinct categories based on your criteria. It is widely used to help make well-informed decisions.*



Chart, scatter chart

Description automatically generated

Graphical user interface, application

Description automatically generatedGraphical user interface, application

Description automatically generatedGraphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Now let us start to build the **Toronto Neighborhood Rating Tool!**

1. Calculate the score for each criterion
   * Use DAX functions - VAR, RETURN, IF to create Income score, Crime score, Diversity score, Airbnb score
   * To make it simple, let us calculate quartiles and assign scores based on a certain order:
     + Quartiles mark each 25% of a set of data:
       - The first quartile Q1 is the 25th percentile
       - The second quartile Q2 is the 50th percentile
       - The third quartile Q3 is the 75th percentile
     + **Be careful about the order when assign score values** (for example, for crime score, the higher score means lower crime rate; for income score, higher score means higher percentage of $100K income.
     + Do not forget to create **calculated columns** for scores; we need to create:
       - Income score
       - Crime score
       - Airbnb score
       - Diversity score

*Income 25th percentile = PERCENTILE.INC('Neighborhood Rating'[% of $100K+Income], 0.25)*

*Income 50th percentile = PERCENTILE.INC('Neighborhood Rating'[% of $100K+Income], 0.50)*

*Income 75th percentile = PERCENTILE.INC('Neighborhood Rating'[% of $100K+Income], 0.75)*

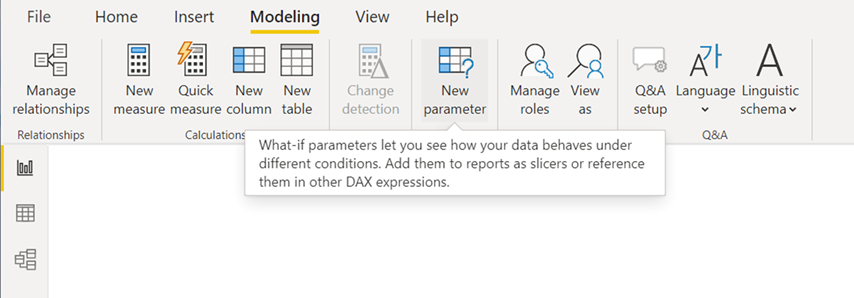
***Income score*** *= if('Neighborhood Rating'[% of $100K+Income] <= 'Neighborhood Rating'[Income 25th percentile], 1, if('Neighborhood Rating'[% of $100K+Income] <= 'Neighborhood Rating'[Income 50th percentile], 2, if('Neighborhood Rating'[% of $100K+Income] <= 'Neighborhood Rating'[Income 75th percentile], 3, 4)))*

***Crime Score*** *= var high=PERCENTILE.INC('Neighborhood Rating'[Crime\_rate%], 0.75) var mid=PERCENTILE.INC('Neighborhood Rating'[Crime\_rate%], 0.50) var low=PERCENTILE.INC('Neighborhood Rating'[Crime\_rate%], 0.25) return if('Neighborhood Rating'[Crime\_rate%] <= low, 4, if('Neighborhood Rating'[Crime\_rate%] <= mid, 3, if('Neighborhood Rating'[Crime\_rate%] <= high, 2, 1)))*

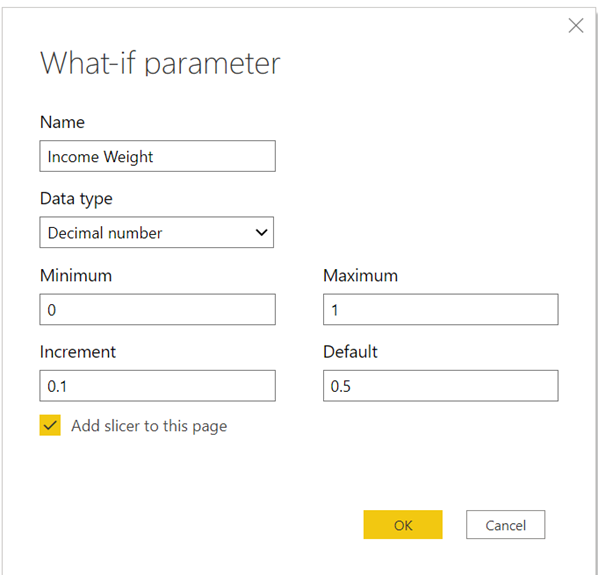
***Airbnb Score*** *= var high=PERCENTILE.INC('Neighborhood Rating'[Tot\_listings], 0.75) var mid=PERCENTILE.INC('Neighborhood Rating'[Tot\_listings], 0.50) var low=PERCENTILE.INC('Neighborhood Rating'[Tot\_listings], 0.25) return if('Neighborhood Rating'[Tot\_listings] <= low, 1, if('Neighborhood Rating'[Tot\_listings] <= mid, 2, if('Neighborhood Rating'[Tot\_listings] <= high, 3, 4)))*

***Diversity Score*** *= var high=PERCENTILE.INC('Neighborhood Rating'[% of nonofficial languages], 0.75) var mid=PERCENTILE.INC('Neighborhood Rating'[% of nonofficial languages], 0.50) var low=PERCENTILE.INC('Neighborhood Rating'[% of nonofficial languages], 0.25) return if('Neighborhood Rating'[% of nonofficial languages] <= low, 1, if('Neighborhood Rating'[% of nonofficial languages] <= mid, 2, if('Neighborhood Rating'[% of nonofficial languages] <= high, 3, 4)))*

1. Use What-if Parameters to create Weight for each criterion; take **Income Weight** as an example:
   * Open New parameter in Modeling section



* + Fill the information as below:
    - ‘Add slicer to this page’ checkbox automatically puts a slicer with what-if parameter onto the current report page
    - ‘Increment’ is how much the parameter will adjust when interacted with in a report

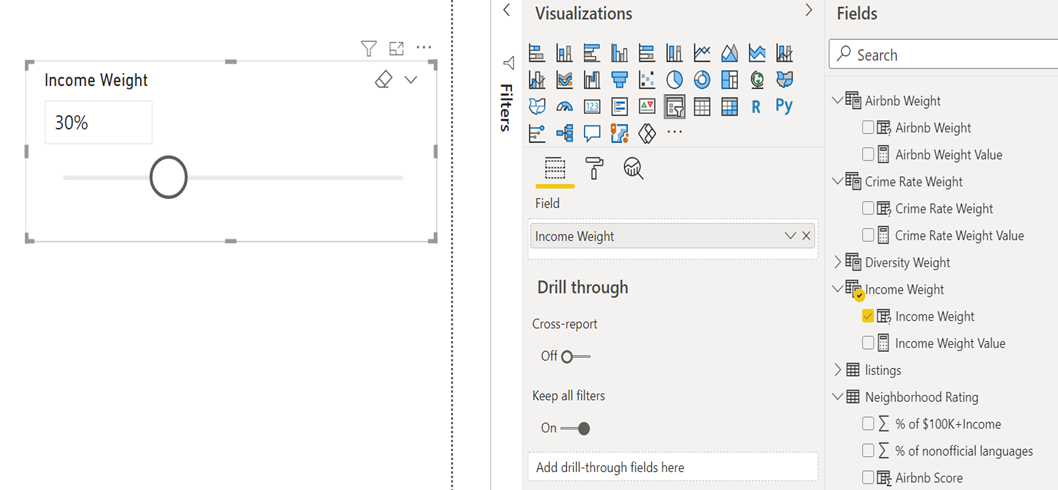


* + Go to ‘Data’ to change the format of ‘Income Weight’ to percentage

A picture containing chart

Description automatically generated

* + Following the steps above to create **Airbnb Weight, Crime Rate Weight and Diversity Weight**
  + In addition to creating the parameter, creating a what-if parameter also creates a measure. Both the parameter and the measure become part of your model



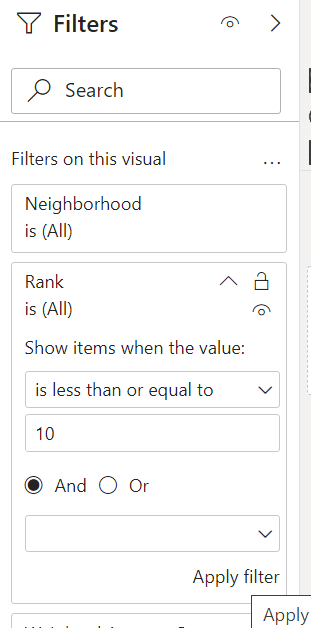
1. **Pass the What-if parameters into a DAX expression to create a Measure, achieving dynamic calculation**
   * Create measure to calculate the dynamic weighted average score for each Neighborhood

*Weighted Average Score = ([Airbnb Weight Value]\*sum('Neighborhood Rating'[Airbnb Score]) + [Crime Rate Weight Value]\*sum('Neighborhood Rating'[Crime Score]) + [Diversity Weight Value]\*sum('Neighborhood Rating'[Diversity Score]) + [Income Weight Value]\*sum('Neighborhood Rating'[Income score]))/([Income Weight Value] + [Airbnb Weight Value] + [Crime Rate Weight Value] + [Diversity Weight Value])*

1. Create measure to rank each Neighbourhood based on dynamic weighted average score

*Rank = rankx(all('Neighborhood Rating'[Neighborhood]), [Weighted Average Score],,0, Dense)*

1. Add a filter by using Rank is less than or equal to 10



1. Complete the tool so that you can rank top 10 Neighbourhoods based on your own preference!

Graphical user interface, website

Description automatically generated