**Student Name:** **Weight:** 17.5%

**Student ID:** **Marks:** /27

# Assignment: Individual Project 2

## Overview

This assessment is for you to demonstrate data visualization techniques by creating an exploratory report/dashboard on a real-life dataset. You will be using a variety of City of Calgary datasets. In this project you will set up your data model including its relationships, then use it to create some exploratory visuals. The model you create for this exercise will also be used for Group Presentation 2, so recommended table nicknames (aliases) have been provided for consistency.

## Instructions

**Scenario**

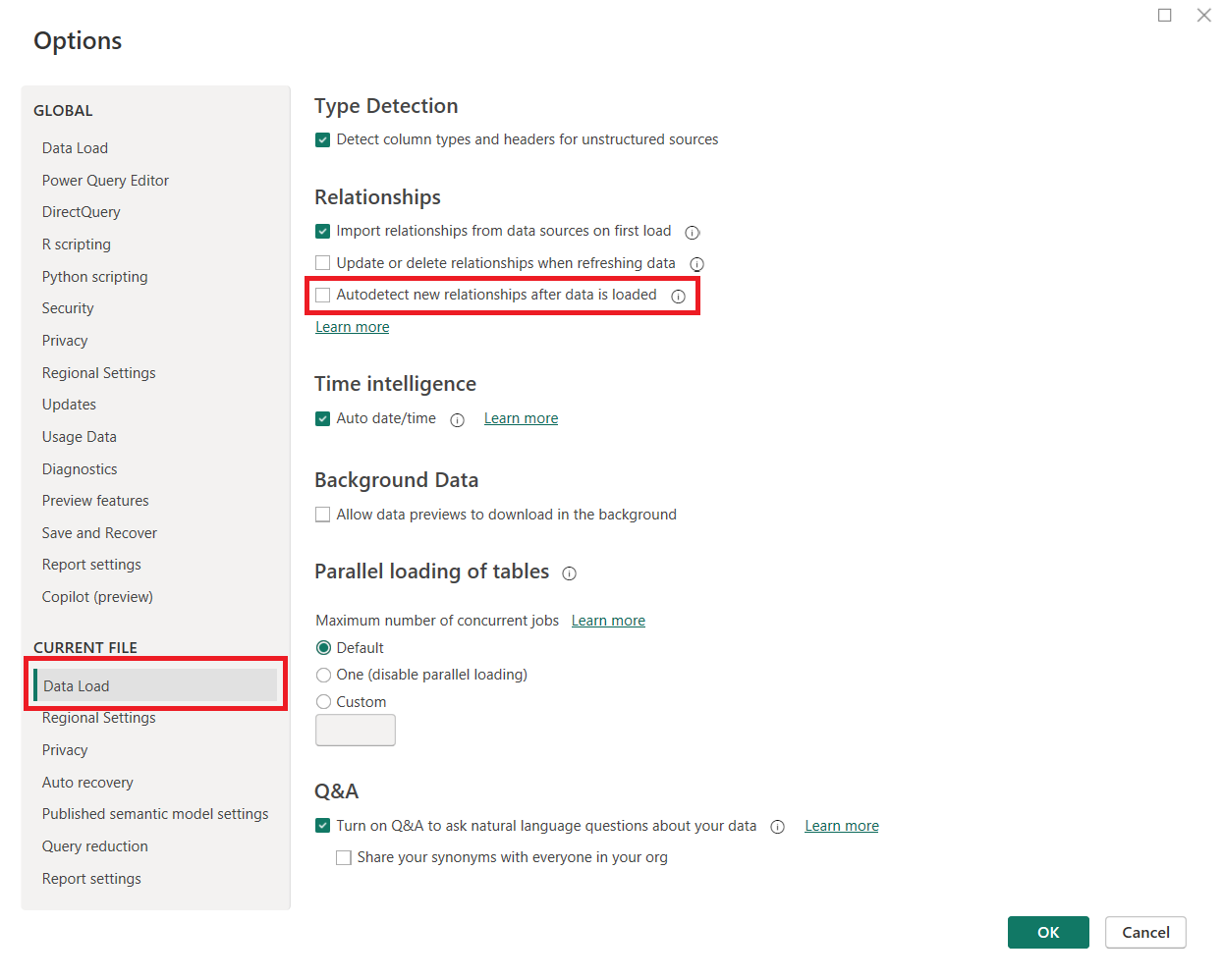
The City of Calgary has contracted you to perform exploratory analysis on their publicly available data. They have asked you to create an exploratory report/dashboard in Power BI to highlight any key insights that you uncover in your investigation. It will be your responsibility to identify meaningful business questions based on the dataset and answer them with effective data visualizations.

**Exercise 1: Creating a Data Model (3 points)**

Load all of the tables listed below (excluding those labeled as “optional”) into Power BI using the method listed under “Data Source”. If you choose to, you may also load the optional tables but these are not required. The column labelled “Table Alias” provides recommended names to use for your tables when they are loaded into Power BI. You will be reusing your data model for Group Presentation 2, so it is highly recommended that you follow this naming convention to be consistent with your group members. Data transformation isn’t required but you should ensure that the column names and data types are detected properly wherever possible.

|  |  |  |
| --- | --- | --- |
| Table Alias | Calgary Dataset | Data Source |
| census-comm | [Historical Community Populations](https://data.calgary.ca/Demographics/Historical-Community-Populations/jtpc-xgsh) | Use CSV |
| census-comm-and-gender | [Civic Census by Community, Age and Gender](https://data.calgary.ca/Demographics/Civic-Census-by-Community-Age-and-Gender/vsk6-ghca) | Use CSV |
| census-ward | [Civic Census by Ward](https://data.calgary.ca/Demographics/Civic-Census-by-Ward/dvig-zn6d) | Use API |
| crime | Community Crime and Disorder Statistics (2012-2019) | Use CSV File from Brightspace |
| dim-community | [Communities by Ward](https://data.calgary.ca/Government/Communities-by-Ward/jd78-wxjp) | Use API |
| dim-quadrant | [City Quadrants](https://data.calgary.ca/Base-Maps/City-Quadrants/g2n2-qnvh) | Use API |
| dim-sector | [City Planning Sectors](https://data.calgary.ca/Base-Maps/City-Planning-Sectors/qwju-mqym) | Use API |
| dim-ward | [Ward Boundaries](https://data.calgary.ca/Government/Ward-Boundaries/tz8z-hyaz) | Use API |
| licensed-pets | [Licensed Pets](https://data.calgary.ca/Services-and-Amenities/Licensed-Pets/5dgy-88cq) | Use CSV |
| structures | [Census by Community 2019](https://data.calgary.ca/Demographics/Census-by-Community-2019/rkfr-buzb) | Use API |
| traffic-incidents | [Traffic Incidents](https://data.calgary.ca/Transportation-Transit/Traffic-Incidents/35ra-9556) | Use CSV |
| trees | [Public Trees](https://data.calgary.ca/Environment/Public-Trees/tfs4-3wwa) | Use CSV |
| businesses | [Downtown Calgary Businesses](https://data.calgary.ca/Business-and-Economic-Activity/Downtown-Calgary-Businesses/cibr-88yk) | Optional, use CSV |
| property | [Total Property Assessed Value](https://data.calgary.ca/Government/Total-Property-Assessed-Value/dmd8-bmxh) | Optional, huge table |

Note: it is recommended that you disable the option “Autodetect new relationships after data is loaded” in Power BI which can be found by going to File > Options and settings > Options > Data Load (Under “CURRENT FILE”).



**Exercise 2: Creating Model Relationships (3 points)**

You will need to create relationships between the tables in your data model. Whenever a key can be used (e.g. 3-letter community code), it should be chosen over a longer text field (e.g. full community name). Once all relationship lines have been created, carefully reorganize the tables, grouping and reordering for ease of review. Below are hints for how to create your model relationships:

* The following tables each contain a column containing the community code, so can be connected to **dim-community**: **businesses, census-comm, census-comm-and-gender, trees, structures, licensed-pets, property.**
* **crime** contains community name so can also be connected to **dim-community.**
* **census-ward** and **dim-community** contain ward # so can be connected to **dim-ward.**
* **dim-community** contains sector so can be connected to **dim-sector**.
* **traffic-incidents** contains quadrant so can be connected to **dim-quadrant**.

You may wish to create additional hierarchies: for example, a hierarchy of ward & community, and/or a hierarchy of sector & community. You are welcome to hide or delete any unusable columns, and can create any new columns or measures you like, or even bring in new tables if you can connect them to the Calgary dataset.

**Exercise 3: Creating an Exploratory Report/Dashboard (21 points)**

Finally, you’ll explore the data, looking for trends, outliers, and/or descriptive stats, and create a report/dashboard of the same.

Your report/dashboard should contain the following:

* A clear title (1 point)
* At least 3 visuals (1 point each) with a consistent colour scheme (2 points)

Visuals will be scored on the following:

* The visual title is a clear question (3 points)
* Appropriate chart types are used (3 points)
* Data-to-ink ratio is maximized (3 points)
* Colour is used effectively within visuals (3 points)
* Annotations are clear and accurate (3 points) including tooltips, units, and any applied filters/slicers

**Submission**

To submit, either:

* Submit your .pbix file in the Brightspace Dropbox, or
* Publish your report to Power BI service, then use the “share” feature to share your dashboard URL with your instructor.

## Marking Criteria

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Criteria** | **Missing**   (0 marks) | **Needs Improvement**  (1 mark) | **Good**  (2 marks) | **Excellent**  (3 marks) | **Marks** |
| **Data model** | No tables loaded | Few tables loaded | Most tables loaded | All tables loaded | **/3** |
| **Model relationships** | No model relationships | Most model relationships missing or inappropriate | Most model relationships created and appropriate | All model relationships are created and appropriate | **/3** |
| **Clear title and consistent colour scheme** | No title and inconsistent colour scheme | No title or inconsistent colour scheme | Clear title, some inconsistencies in colour scheme | Clear title and colour scheme is consistent | **/3** |
| **At least 3 visuals** | No visuals created | 1 visual created | 2 visuals created | At least 3 visuals created | **/3** |
| **Visual titles are clear questions** | No questions asked in visual titles or all question unclear | 1 clear question asked in visual title | 2 clear questions asked in visual title | 3 clear questions asked in visual title | **/3** |
| **Appropriate chart types used** | Chart types are inappropriate. | Few chart types are appropriate for the data used. | Most chart types are appropriate for the data used. | All chart types are appropriate for the data used. | **/3** |
| **Data-to-ink ratio** | Data-to-ink ratio is low. Difficult to interpret visualizations | Data-to-ink ratio is ok. Many elements should be removed. | Data-to-ink ratio is high. Some elements could be removed. | Data-to-ink ratio is maximized. Visualizations are clean and well-formatted. | **/3** |
| **Effective use of colour** | Colour is used poorly. Colour is distracting and makes it difficult to understand visualizations. | Colour choices could be better. Colour is distracting in most cases. | Colour is mostly effective for conveying meaning. Some colour may be distracting. | Colour is used effectively to convey meaning. Colour is not distracting. | **/3** |
| **Clear and accurate annotation (e.g. no typos) including tooltips, units, and any applied filters/slicers** | Significant typos and inconsistencies. Annotations are inaccurate or not meaningful. | Many typos and inconsistencies. Some chart annotations are inaccurate or not meaningful. | Minor typos or inconsistencies. Chart annotations are mostly accurate and meaningful. | No typos or inconsistencies. All chart annotations are accurate and meaningful. | **/3** |
| **Total** |  | | | | **/27** |