Week 1

***A description of the problem and a discussion of the background. (15 marks)***

As I was exploring potential problem to propose for the project, I was approached by a friend who is currently in the process of opening his own Ramen Restaurant in Toronto.

Based on his past experience of running multiple Ramen establishments, they are looking for the following criteria to determine initial potential locations

1. High population
2. High population density area to allow for highest potential number of initial customers when opening
3. High percentage of population with income greater than $30,000
4. High percentage of population who are living alone
5. High percentage of population who are aged between 15 - 54 years old
6. High population percentage of immigrants originating from East Asian countries
7. Neighbourhood with high residential real estate price
8. Neighbourhood with lower number of existing vendors that are similar to Ramen

|  |  |  |
| --- | --- | --- |
| Criteria | Qualitative Weight | Quantitative Weight |
| 1 | Medium | 0.125 |
| 2 | High | 0.2 |
| 3 | Medium | 0.125 |
| 4 | High | 0.2 |
| 5 | Medium | 0.125 |
| 6 | Low | 0.05 |
| 7 | Low | 0.05 |
| 8 | Medium | 0.125 |

After initial exploration of the available data, it also determined to limit the geographical area to be close to where the owner of the ramen place resides to reduce traveling time:

* South of St. Clair Street
* East of Roncesvalles Avenue
* West of Coxwell Avenue
* Excluding the following areas to avoid existing competitions: Bay Corridor Yonge-Church Corridor.

This reduces the total number of neighbourhoods being examined from 140 to 32. By concentrating on the neighbourhoods close to the downtown core we are also targeting those who may work in downtown are looking to find dinner spots on their commute home to the suburbs.

***A description of the data and how it will be used to solve the problem. (15 marks)***

Most of the data to be used to determine the best neighbourhood to open a Ramen Restaurant will be obtained from Toronto Open Data project:

The following particular dataset will be utilized

**Neighbourhood Profiles**

<https://open.toronto.ca/dataset/neighbourhood-profiles/>

This data set was last refreshed in July 2019 and contains the data for each of City of Toronto's 140 social planning neighbourhoods of Toronto collected during the latest Census of Population in 2016. Data collected contains: age and sex, families and households, language, immigration and internal migration, ethnocultural diversity, Aboriginal peoples, housing, education, income, and labour.

These social planning neighbourhoods were developed by the City of Toronto to help government and community organizations with local planning by providing socio-economic data at a meaningful geographic area. The boundaries of these social planning neighbourhoods are consistent over time, allowing for comparison between Census years

The Census data to provide a portrait of the demographic, social and economic characteristics of the people and households in each City of Toronto neighbourhood, where in this case we are particularly interested in characteristics of people in favour over households.

The data was sourced from a number of Census tables released by Statistics Canada.

Particularly, we are interested in the following data points for each neighbourhoods

* Population and dwellings
  + Population, 2016
  + Population density per square kilometre
* Age characteristics
  + Youth (15-24 years)
  + Working Age (25-54 years)
* Family characteristics of adults
  + Persons living alone (per cent)
* Income of individuals in 2015
  + Total - Employment income groups in 2015 for the population aged 15 years and over in private households - 100% data
    - Under $5,000 (including loss)
    - $5,000 to $9,999
    - $10,000 to $19,999
    - $20,000 to $29,999
    - $30,000 to $39,999
    - $40,000 to $49,999
    - $50,000 to $59,999
    - $60,000 to $69,999
    - $70,000 to $79,999
    - $80,000 and over
* Total income: Average amount ($)
* Immigrants by selected place of birth
  + Total - Selected places of birth for the immigrant population in private households - 25% sample data
    - Asia
      * China
      * Hong Kong
      * Japan
      * Korea; South
      * Philippines
      * Taiwan
      * Viet Nam

The percentage is calculated by tabulating the total of each subcategory and dividing by the total respondents.

**Neighbourhoods**

<https://open.toronto.ca/dataset/neighbourhoods/>

This dataset in .geojson format contains boundaries as well as the geographic coordinates (latitude and longitude) of the 140 City of Toronto neighbourhoods for social planning purposes.

The geographic coordinates will be used to call the Foursquare API to determine the upto 120 food vendors within the center of each neighbourhood.

The boundaries information will be used to generate choropleth map with Folium on a Leaflet map.

The file was converted to .json file for the purpose of this project and easy reading into

**Foursquare**

<https://developer.foursquare.com/>

Foursquare API is used to collect data on close by venues within a specific radius of a given geographic coordinate.

For the purpose of this project we are only looking at food related venues, by pass ‘food’ to the parameter ‘Section’ in the Get Venue Recommendations API call.

(<https://developer.foursquare.com/docs/api/venues/explore>)

The data returned will be used to explore existing make up for the food venue landscape of each neighbourhood.

**2018 Toronto Detached Prices by Neighbourhood**

<https://docs.google.com/spreadsheets/d/1GizBkDvGaYm5AAPJ7O0y9eq-n7i81to6WpjejvYFzw0/edit#gid=754169703>

<http://www.century21.ca/scott.ingram/blog/Toronto_s_most_and_least_expensive_neighbourhoods_in_2018>

This data set contain average detached house prices by 144 TREB (Toronto Real Estate Board) neighbourhoods aggregated by Scott Ingram. As Scott mentioned, detached house prices were used to keep things simple and more apples-to-apples between neighbourhood.

In order for the data to be used, the TREB neighbourhood must be converted to the 140 City of Toronto social planning neighbourhoods.

For neighbourhood that does not have any sales, the overall average price for Toronto will be used .

This data can be used to determine the state of the real estate price.

For the second week, the final deliverables of the project will be:

A link to your Notebook on your Github repository, showing your code. (15 marks)

A full report consisting of all of the following components (15 marks):

Introduction where you discuss the business problem and who would be interested in this project.

Data where you describe the data that will be used to solve the problem and the source of the data.

Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, and what machine learnings were used and why.

Results section where you discuss the results.

Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.

Conclusion section where you conclude the report.

3. Your choice of a presentation or blogpost. (10 marks)