**Applied Data Science Capstone Final Project**

**By Marc-Andre St-Laurent**

**Business Problem**

*Determine the best location in Toronto to open a new restaurant.*

Toronto is the largest city in Canada with many existing restaurants. To decide which would be the best location to open a new restaurant, we need to consider many different factors including. This analysis will focus on the location of the restaurant.

A solution to this problem would be of interest to anyone who is interested in opening a new restaurant. While this analysis will focus on Toronto, it can easily be adapted to any other city.

**Data**

Data sources to be used:

* The Foursquare location data of the venues in Toronto
* The 2016 Canada Census Data of the population of the postal code areas of Toronto
* The latitude and longitude coordinates of each postal code area within Toronto.

**Methodology**

To solve our business problem of determining the best location to open a new restaurant, we will segment the data, then determine which area has the highest population, and the lowest number of competing restaurants.

First, we will define the areas to consider. To do this, we will segment the city into sections defined by their postal code area. This is the first three characters of an address’ postal code. This method of segmentation was chosen every location can be clearly defined belonging to only one area. We will use the latitude and longitude of each postal code area to define our geographic areas that we are going to consider.

Another advantage of using the postal code area to segment the data is that the population data is available from the last Canadian Census, which was in 2016. This will factor into our analysis as the more populous a segment, the more potential customers there are. Also, we can remove any area that has a very small population. For example, business district with a housing population of less than 100.

The main source of our data will be the Foursquare API. From this, we can extract all venues within a given area. However, since we are only interested in the competition to our new restaurant, we remove all venues not related to the restaurant industry.

After we have the number of restaurants and population for all segmented areas, we determine the location with the highest population to restaurant ratio.

**Results**

Due to data capacity issues, instead of applying the methodology to all of Toronto, it was applied to one borough: East Toronto.

East Toronto has five postal code areas: M4E, M4K, M4L, M4M, and M7Y. However, M7Y is a business district with a population of 10. This can be removed as an outlier. This leaves four segments to consider.

After extracting all information as described in the Methodology section, here are the results:

|  |  |  |  |
| --- | --- | --- | --- |
| **Postal Code Area** | **Population** | **Restaurant Count** | **Population per Restaurant Ratio** |
| M4L | 32640 | 24 | 1360 |
| M4K | 31583 | 30 | 1053 |
| M4E | 25044 | 24 | 1044 |
| M4M | 24689 | 29 | 851 |

We have our result! With the highest ratio of population to restaurants of 1360:1, the best area to put in a new restaurant is M4L.

**Discussion**

While this analysis focused on the borough of East Toronto, it can be applied to a different borough, or to all of Toronto. It could also be applied to another city if the population and geographic information is available.

Other information that could be used in this analysis is the areas that where people spend their work day. This is another factor in choosing a location for a restaurant. However, this data was not available.

**Conclusion**

Using population data and the Foursquare API we can determine the best location for a new restaurant. While this analysis is limited to where people live, and not where they work, it is still an excellent starting point for a restaurant location.