Introduction/Business Problem

Large cities often have many restaurants due to the density of people. We are looking at the feasibility of starting a new restaurant. We will look at a metropolitan area, and analyze the neighborhoods to determine if a new restaurant will have sufficient clientele. To determine clientele, we will analyze specific areas and determine cuisine preference over an area. We will determine which restaurant to start based on scarcity of specific cuisines in different areas of our city. Our target meal will be lunch, so we will have to analyze businesses in the area as well. The target audience will be the employees at nearby businesses, who will want a diverse option of lunches.

Data

Data needed will be businesses and their coordinates, and restaurant types and their coordinates. Data will be analyzed to determine which new restaurants should be placed in an area based on scarcity of cuisine in that area. I will use a geocoder to find locations. The Foursquare API will be used extensively.

Methodology

In order to obtain the data needed, segmenting and clustering was used. Wikipedia was used as a data source, from where data was scraped. The Pandas data frame was utilized.

Results

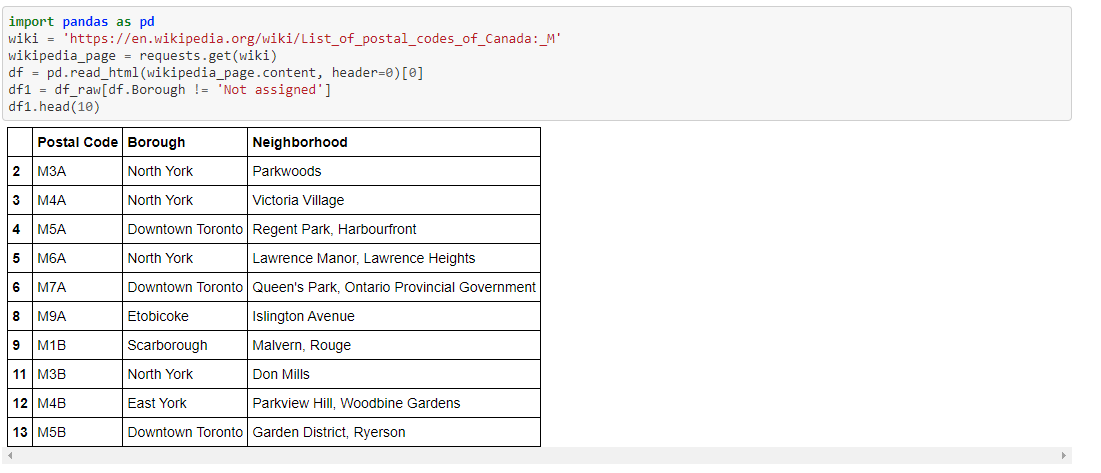


Table 1: Unsorted Data

Figure 1: New York and Surrounding Area

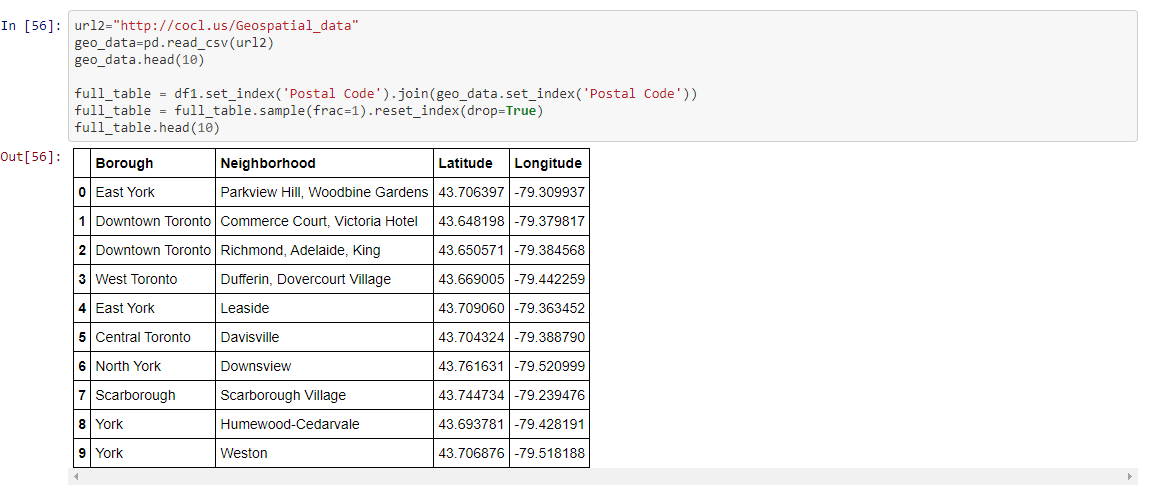


Table 2: Sorted Data

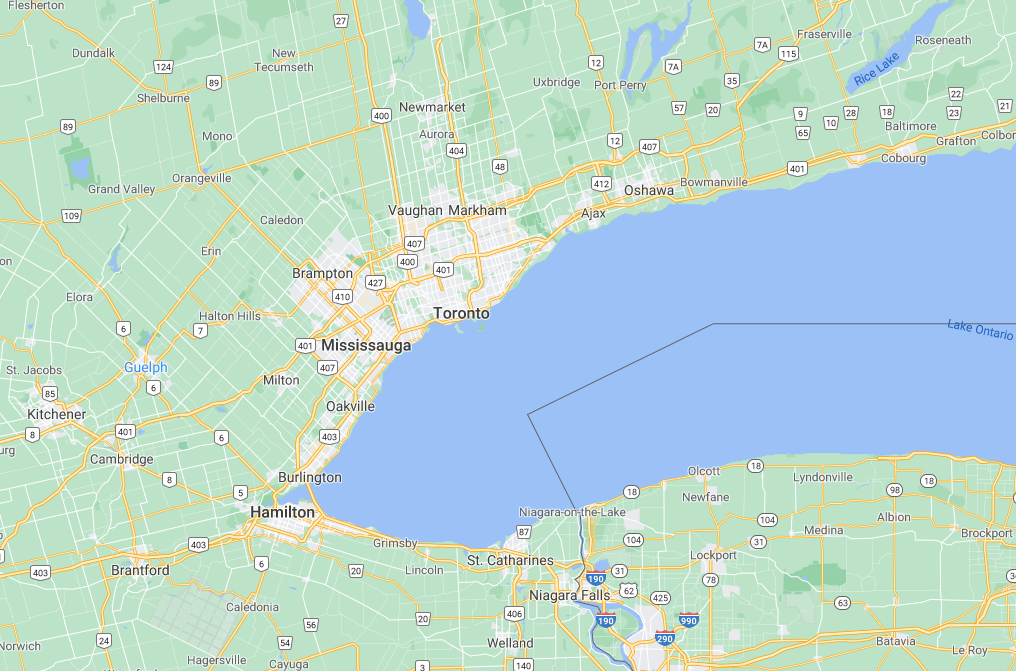


Figure 1: Toronto and Surrounding Area

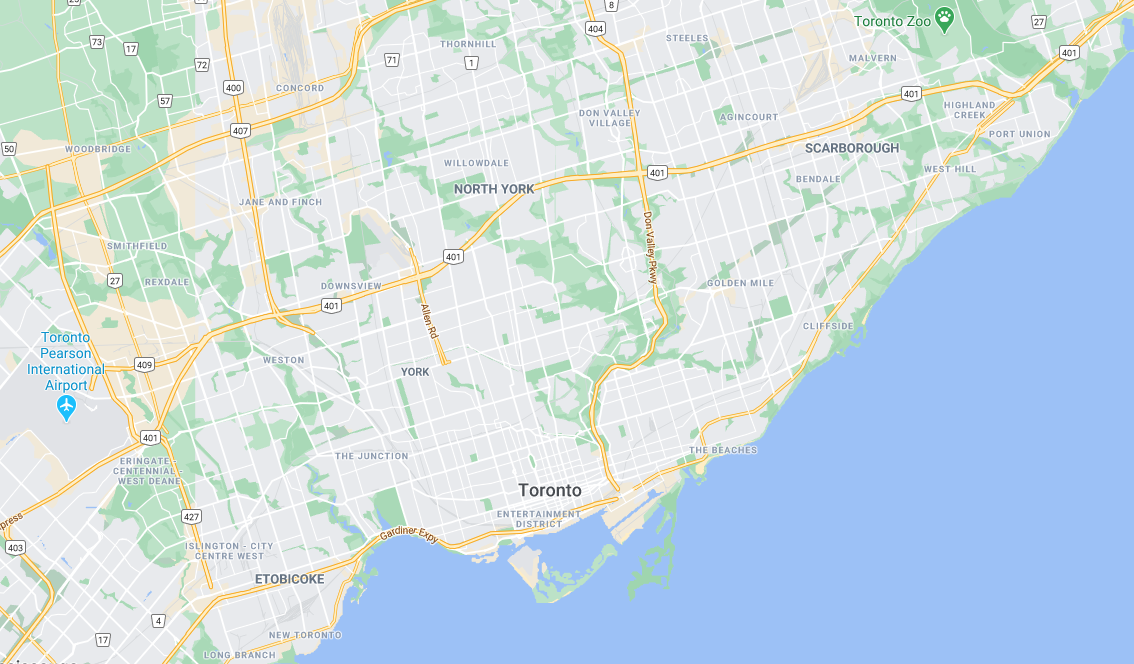


Figure 2: Toronto



Figure 3: Toronto



Figure 4: East York



Figure 5: York



Figure 6: Scarborough

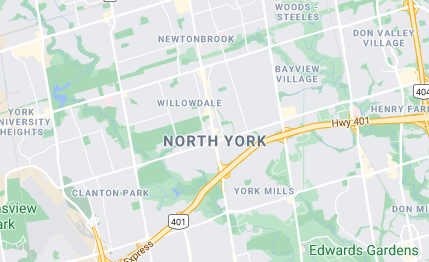


Figure 7: North York

Discussion and Conclusion:

To keep it simple, my data frame is organized by neighborhood seen above. This gives geographical, neighborhood, and location information (See Table 1-2; figure 1-7).