



BATTLE OF THE NEIGHBORHOODS

DETERMINING THE BEST
LOCATION TO OPEN AN
ITALIAN RESTAURANT IN
TORONTO

BACKGROUND

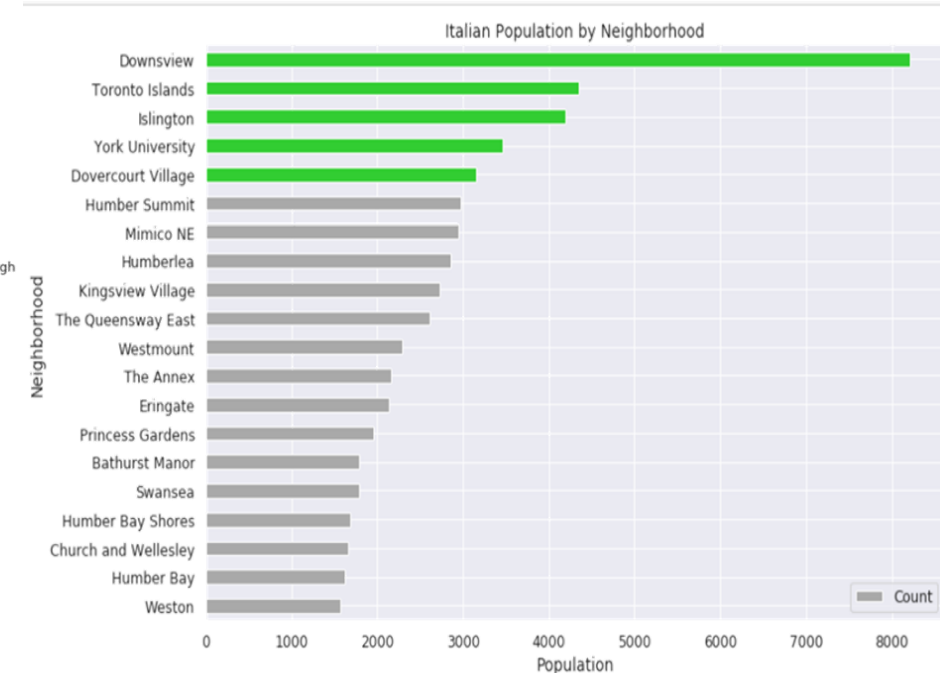
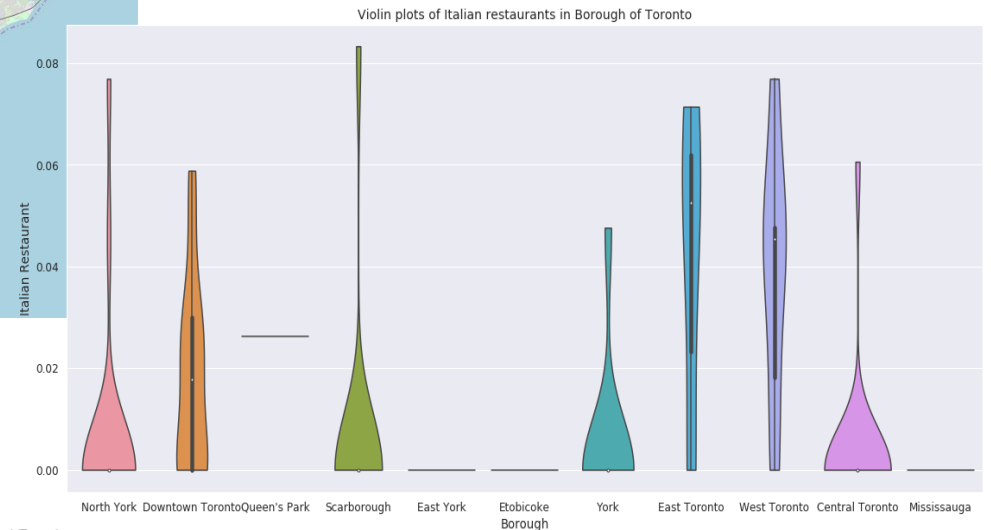
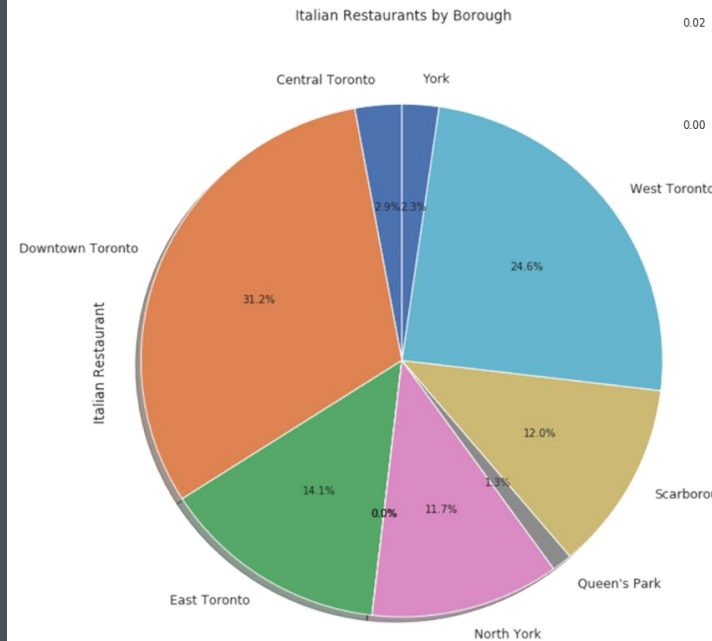
- Toronto is one of the most culturally diverse cities in Canada
- The city is considered one of the largest entertainment hubs in the country
- One of the most popular themed restaurant categories is Italian restaurants
- The Italian community is the 6th largest ethnic group in the country
- Toronto presents a great opportunity of new business owners or franchisors to set up an Italian restaurant

DATA ACQUISITION & CLEANSING

- List of neighborhoods scraped from Wikipedia page 'List of postal codes of Canada' - https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
 - Boroughs with NA values dropped and neighborhoods with NA values matched to Borough
- Geographical coordinates obtained from csv file - https://cocl.us/Geospatial_data
- Demographics data retrieved from Toronto's Open Data portal - <https://open.toronto.ca/dataset/neighbourhood-profiles/>
 - Income columns dropped and only Italian population kept
- Venue Location data recovered from Foursquare API

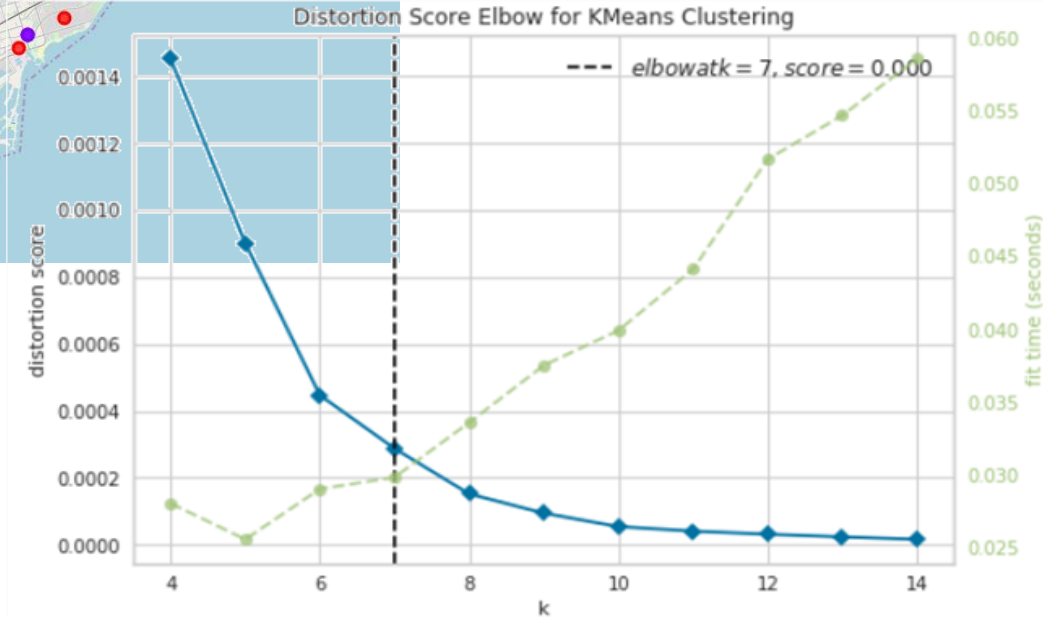
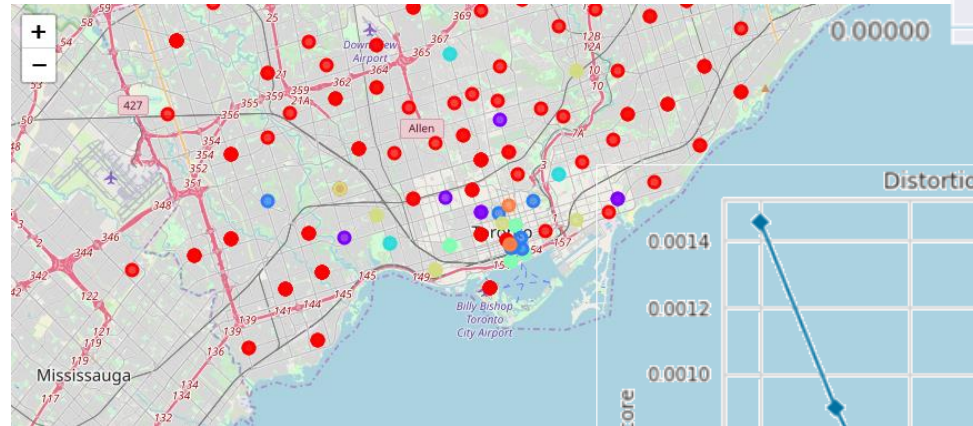
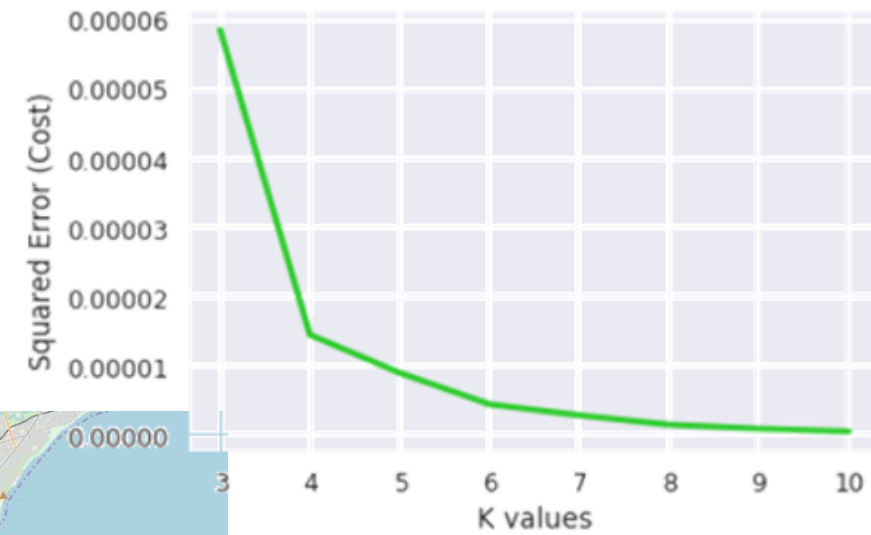
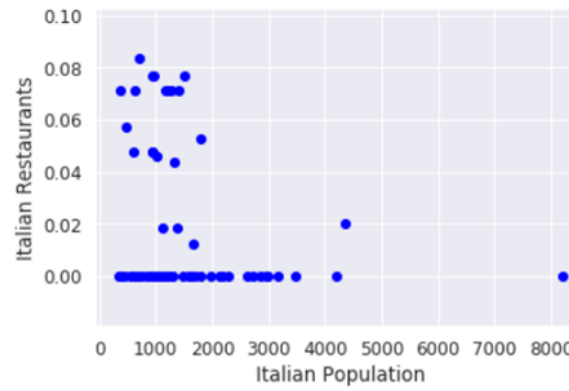
EXPLORATORY ANALYSIS

- Neighborhoods explored – 210, boroughs explored 11
- Boroughs with the most Italian Restaurants – West Toronto, East Toronto, York, North York, Scarborough, Downtown Toronto and Central Toronto
- Most populated neighborhoods – Downsview, Toronto Islands, Islington, York University, Dovercourt Village
- Most populated boroughs – Etobicoke, North York



MODELING

- No distinct linear relationship between Italian population and Italian restaurants
- K-means clustering was used for the analysis
- Distortion score elbow determined 7 as the optimum k value
- 7 Clusters were obtained with cluster 0 having the least number of restaurants and cluster 3 with the largest number of restaurants.



DISCUSSION

- From the K-mean clustering analysis, the cluster with the most restaurants is Cluster 3, followed by Cluster 1, then Cluster 5 followed by Cluster 2, then Cluster 4, Cluster 6 and finally Cluster 0
- Most ideal location is one where there is an established market for Italian restaurants without too much competition but with a significant Italian demographic
- Cluster 3, 1 and 5 respectively have the highest concentration of Italian restaurants, therefore highly competitive clusters
- Cluster 0 and Cluster 6 have the least number of Italian restaurants, therefore may indicate a poor market for Italian food
- Cluster 2 and Cluster 4 are median range for Italian restaurants and are dominated by Downtown neighborhoods, which also represent 5th largest population of Italians

CONCLUSION

- Machine learning techniques were able to help predict the output given the data and used Folium to visualize it on a map.
- Analysis can be improved further with more data and different machine learning techniques to examine relationship through polynomial regression techniques.
- This project is useful in analysing any scenario for opening any type of business eg a spa.
- This project can act as an initial guide to make complex decisions about location using data-science.

