CAPSTONE PROJECT - PREDICTING NEIGHBORHOOD TO OPEN A RESTAURANT IN TORONTO, CANADA

Applied Data Science Capstone by IBM/Coursera

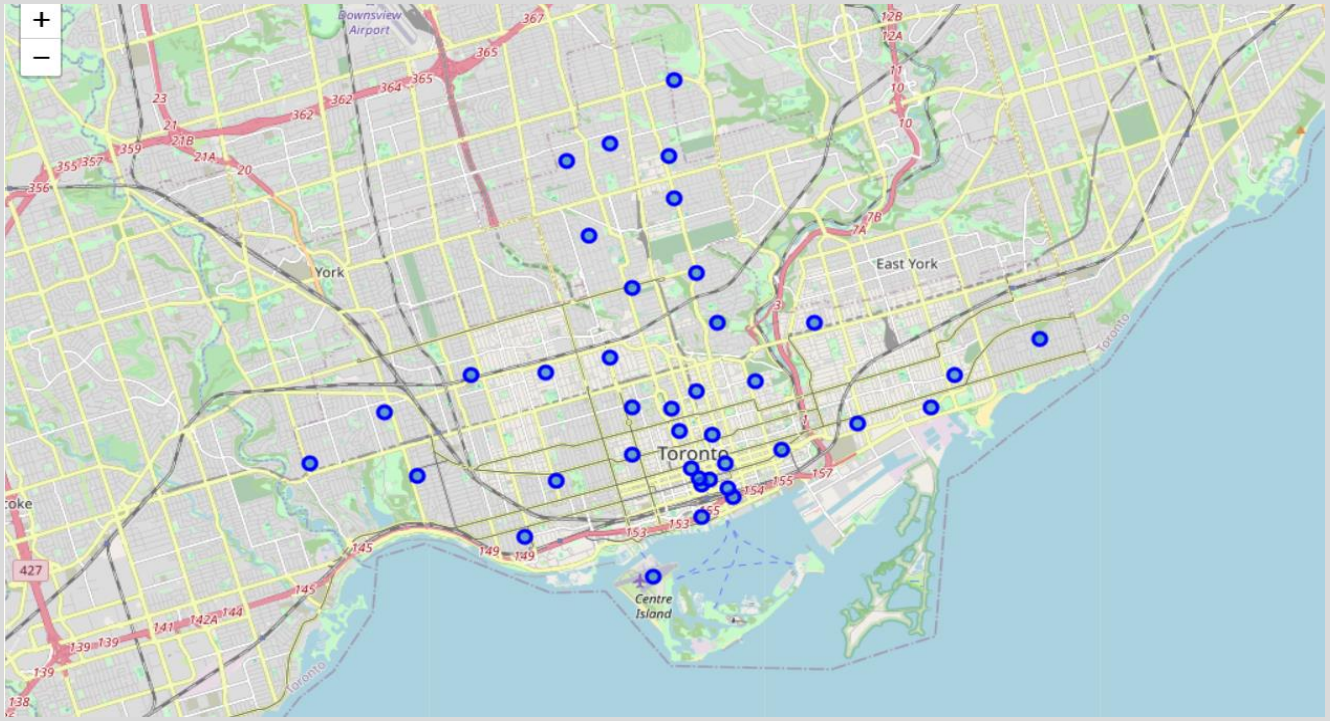
- **In this project we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening an Italian restaurant in Toronto, Canada.**
- Location plays a major role in the success of any restaurant. Therefore, a preliminary market analysis will help in determining a favorable location for the new restaurant. Factors like its neighborhood, or surrounded by the same cuisine restaurant, or have easy access to transportation, will affect sales.
- Since there are lots of restaurants in Toronto, we will try to detect **locations that are not already crowded with Italian restaurants**. This project uses data science find out popular eateries present in the neighborhood and then to predict which neighborhood will be the best to open a restaurant.

DATA ACQUISITION AND CLEANING

- Following data sources will be needed to extract/generate the required information:
- The dataset was present at et: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M,we use Beautifulsoup library to get html data and then algorithmically extract data frame.
- to get coordinates we use https://cocl.us/Geospatial_data
- number of restaurants and their type and location in every neighborhood will be obtained using Foursquare API
- After data cleaning we were left with 180 samples and 5 features.

METHODOLOGY

In first step we have collected the required “data: neighborhood, boroughs, postal codes and location of Toronto, Canada” from Wikipedia using Beautifulsoup library and used geospatial data for location coordinates. From this data we collected Main city locations around Toronto where restaurants will get open. We have shown these locations on map with markers using Folium library.



METHODOLOGY

- Second step in our analysis will be calculation and exploration of "restaurants" around these major neighborhood locations. This new data helps us to get a clear view of type of restaurants in the locality. We have also identified venue location, venue category, name, latitudes and longitudes of restaurants using "Foursquare API". This gives a new dataframe with 349 samples and 7 features. These 7 features are venue latitudes, venue longitudes, neighborhood latitude, neighborhood longitude, Venue, Venue category and Neighborhood.

```
[23]: city_venues.head(15)
```

[23]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	The Beaches	43.676357	-79.293031	Domino's Pizza	43.679058	-79.297382	Pizza Place
4	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
5	The Beaches	43.676357	-79.293031	Seaspray Restaurant	43.678888	-79.298167	Asian Restaurant
6	The Danforth West, Riverdale	43.679557	-79.352188	MenEssentials	43.677820	-79.351265	Cosmetics Shop
7	The Danforth West, Riverdale	43.679557	-79.352188	Pantheon	43.677621	-79.351434	Greek Restaurant
8	The Danforth West, Riverdale	43.679557	-79.352188	La Diperie	43.677702	-79.352265	Ice Cream Shop
9	The Danforth West, Riverdale	43.679557	-79.352188	Dolce Gelato	43.677773	-79.351187	Ice Cream Shop
10	The Danforth West, Riverdale	43.679557	-79.352188	Cafe Fiorentina	43.677743	-79.350115	Italian Restaurant
11	The Danforth West, Riverdale	43.679557	-79.352188	Mezes	43.677962	-79.350196	Greek Restaurant
12	The Danforth West, Riverdale	43.679557	-79.352188	Louis Cifer Brew Works	43.677663	-79.351313	Brewery
13	The Danforth West,	43.679557	-79.352188	Messini Authentic Gyros	43.677704	-79.350480	Greek Restaurant

METHODOLOGY

- Out of these 349 samples we narrow it down to those who has restaurant string in their venue category. This gives our final data set "df_restaurants" with 81 samples and 7 features.

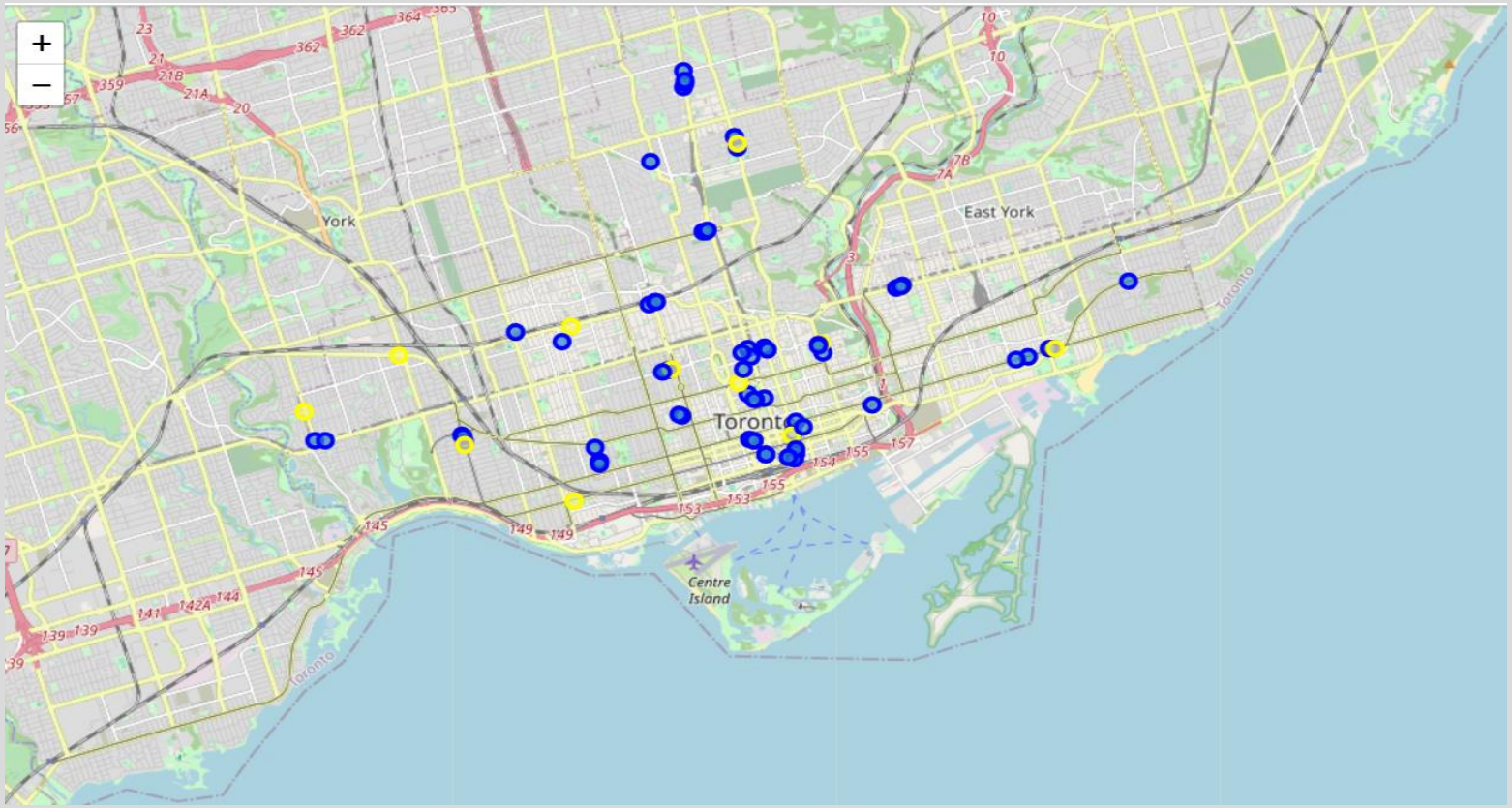
```
df_restaurants.head(15)
```

[33]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
5	The Beaches	43.676357	-79.293031	Seaspray Restaurant	43.678888	-79.298167	Asian Restaurant
7	The Danforth West, Riverdale	43.679557	-79.352188	Pantheon	43.677621	-79.351434	Greek Restaurant
10	The Danforth West, Riverdale	43.679557	-79.352188	Cafe Fiorentina	43.677743	-79.350115	Italian Restaurant
11	The Danforth West, Riverdale	43.679557	-79.352188	Mezes	43.677962	-79.350196	Greek Restaurant
13	The Danforth West, Riverdale	43.679557	-79.352188	Messini Authentic Gyros	43.677704	-79.350480	Greek Restaurant
18	India Bazaar, The Beaches West	43.668999	-79.315572	The Burger's Priest	43.666731	-79.315556	Fast Food Restaurant
20	India Bazaar, The Beaches West	43.668999	-79.315572	O Sushi	43.666684	-79.316614	Sushi Restaurant
24	India Bazaar, The Beaches West	43.668999	-79.315572	Casa di Giorgio	43.666645	-79.315204	Italian Restaurant
49	North Toronto West, Lawrence Park	43.715383	-79.405678	C'est Bon	43.716785	-79.400406	Chinese Restaurant
54	North Toronto West, Lawrence Park	43.715383	-79.405678	Sushi Shop	43.713861	-79.400093	Restaurant
56	North Toronto West, Lawrence Park	43.715383	-79.405678	Tio's Urban Mexican	43.714630	-79.400000	Mexican Restaurant
57	North Toronto West, Lawrence Park	43.715383	-79.405678	A&W	43.715149	-79.399944	Fast Food Restaurant
60	Davisville	43.704324	-79.388790	Marigold Indian Bistro	43.702881	-79.388008	Indian Restaurant
61	Davisville	43.704324	-79.388790	Zee Grill	43.704985	-79.388476	Seafood Restaurant
66	Davisville	43.704324	-79.388790	Sakae Sushi	43.704944	-79.388704	Sushi Restaurant

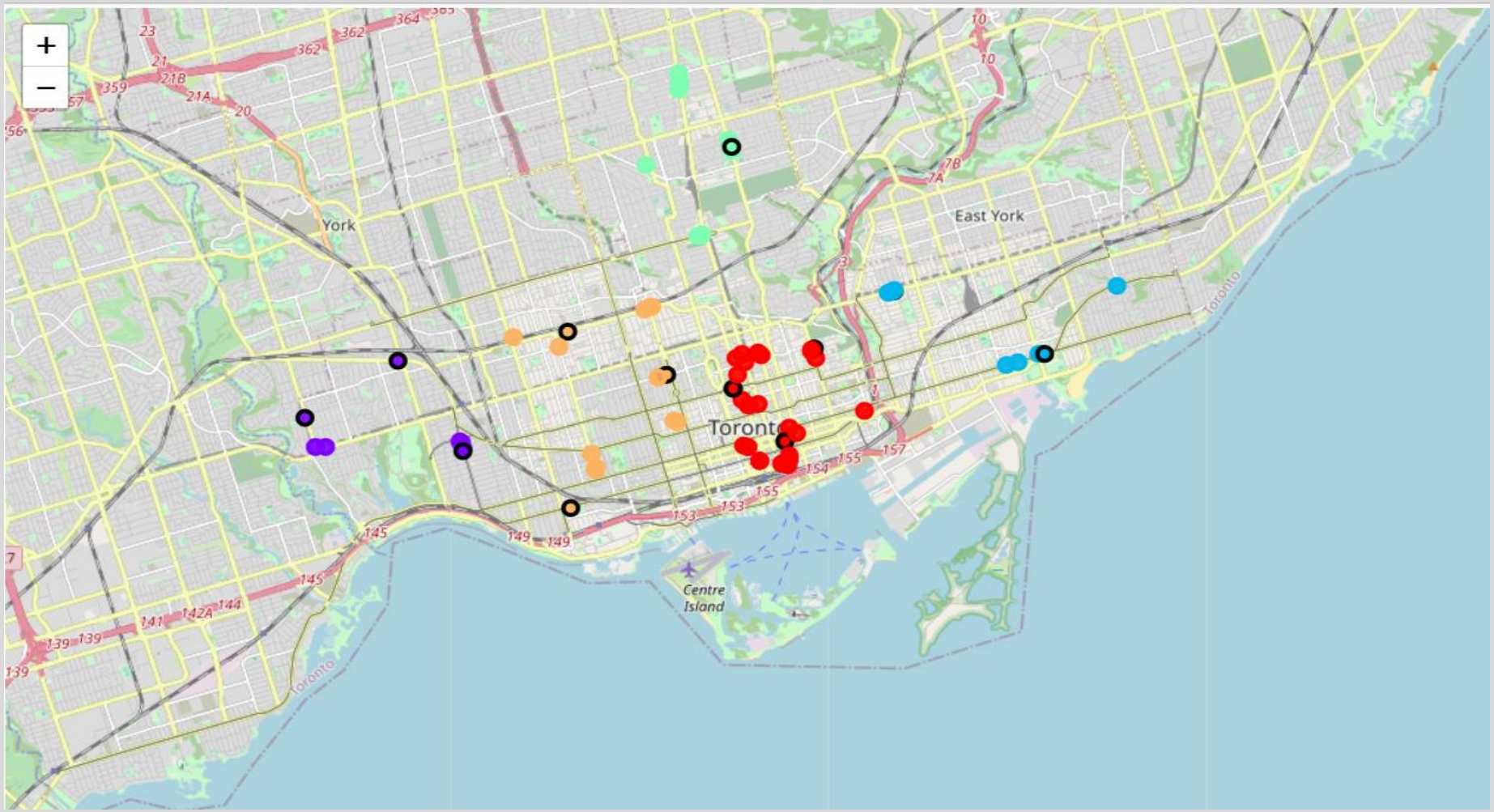
METHODOLOGY

- Then we visualize these restaurant locations on Map of Toronto with "Italian Restaurants" markers different from others, for this we will use Folium library of python. To distinguish between Italian and other type of restaurants, we marked Italian restaurants with yellow marker and others with blue.



METHODOLOGY

- In third and final step we will focus on finding out promising areas by creating "clusters of locations that has any kind of restaurant present." We will only take into consideration the locations with no more than one restaurant in its cluster. We will present map of all clusters (using k-means clustering) of these restaurant locations with clearly distinguish between currently present Italian restaurants from others and search for optimal venue location by finding best fit cluster.
- We will show Toronto Map divided into 5 clusters varying by colors. The Italian Restaurant location in the cluster will be highlighted by black outer ring. We will get a clear view of the Location which can be further considered as a winner.


York
East York
Toronto
Centre Island
Toronto
Toronto

RESULT AND DISCUSSION

- Our result shows that although there are many restaurants in Toronto, with some locations having a greater number of Italian Restaurants, there are still some localities with lesser competition for our stakeholders.
- We came across two clusters which have a smaller number of Italian restaurants than others. **Cluster 2 & 3**. These clusters have location of neighborhood and Venue name etc. By examining both clusters we find out that Cluster 2 has two Italian Restaurants and Cluster 3 has one. Therefore, we choose cluster 3 locations for our new restaurant as there was no other depending feature left.
- By this result, North Toronto West, Lawrence Park are the most suitable localities as they do not have any Italian Restaurant there. If we wish to open a restaurant near Toronto central, Summerhill West will be the best choice.

CONCLUSION

- Purpose of this project was to identify Toronto areas with low number of restaurants (particularly Italian restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Italian restaurant. We have collected the required data: neighborhood, boroughs, postal codes and location of Toronto, Canada from Wikipedia and used geospatial data for location coordinates. Further, by retrieving venues from Foursquare data around major neighborhood locations, we visualize these restaurants on Map of Toronto. Then finding out promising areas by creating clusters of locations that has any kind of restaurant present (with clearly distinguish between currently present Italian restaurants from others) in order to create major zones of interest (containing greatest number of potential locations) which can further to be used as starting points for final exploration by stakeholders.
- Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.

FUTURE DIRECTION

- This project works on the scenario of ‘finding best location to open a restaurant in a particular neighborhood’, but it is not limited to this business problem. We can use this to find out best location for various business problems like constructing apartment, societies, workplace, amusement parks etc.
- Interested user can use postal code data with its corresponding location to predict this analysis for their own use with having access to Foursquare API.