Best Toronto Neighbourhoods to open a Persian Restaurant

Problem Statement: Prospects of opening a Persian Restaurant in Toronto, Canada.

Introduction

To be awarded a Data Science professional certificate from IBM, the candidate must demonstrate the knowledge on the real-world data to solve a business problem, me as a candidate decided to work on the real dataset related to Toronto area for opening a new business by analysing and combining several data sets.
In this project I will gather the data from the web (Wikipedia) and by using the Foursquare location data will find out which neighborhood is/are best to open the restaurant.

Arash Khosravi 28 Apr 2020

Data acquisition and cleaning

Data Sources

I'm using wiki page to get the information related to each Postal Code and related neighborhood in Toronto.
The data can be retrieved from the link https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

I'm using "https://cocl.us/Geospatial_data" csv file to get all the geographical coordinates of the neighbourhoods.

I'm going to identify the neighbourhoods which are densely populated with Persians, to get information about the distribution of population by their ethnicity I'm using "Demographics of Toronto" (https://en.m.wikipedia.org/wiki/Demographics_of_Toronto#Ethnic_diversity)

To get location and other information about various venues in Toronto I'm using Foursquare's explore API.
Using the Foursquare's explore API (which gives venues recommendations), I'm fetching details about the venues up present in Toronto and collected their names, categories and locations (latitude and longitude).

From Foursquare API (https://developer.foursquare.com/docs), I retrieved the following for each venue:
Name: The name of the venue.
Category: The category type as defined by the API.
Latitude: The latitude value of the venue.
Longitude: The longitude value of the venue.

Data Cleaning

- Wikipedia — package is used to scrape the data from wiki.
- *Scrap the distribution of population from Wikipedia*
- *The final DataFrame with Latitude and Longitude*

	Borough	PostalCode	Neighborhood	Latitude	Longitude
0	Central Toronto	M4N	Lawrence Park	43.728020	-79.388790
1	Central Toronto	M4P	Davisville North	43.712751	-79.390197
2	Central Toronto	M4R	North Toronto West	43.715383	-79.405678
3	Central Toronto	M4S	Davisville	43.704324	-79.388790
4	Central Toronto	M4T	Moore Park / Summerhill East	43.689574	-79.383160
...	...	...	...	...	...
98	York	M6C	Humewood-Cedarvale	43.693781	-79.428191
99	York	M6E	Caledonia-Fairbanks	43.689026	-79.453512
100	York	M6M	Del Ray / Mount Dennis / Keelsdale and Silvert...	43.691116	-79.476013
101	York	M6N	Runnymede / The Junction North	43.673185	-79.487262
102	York	M9N	Weston	43.706876	-79.518188

103 rows × 5 columns

Get location data using Foursquare

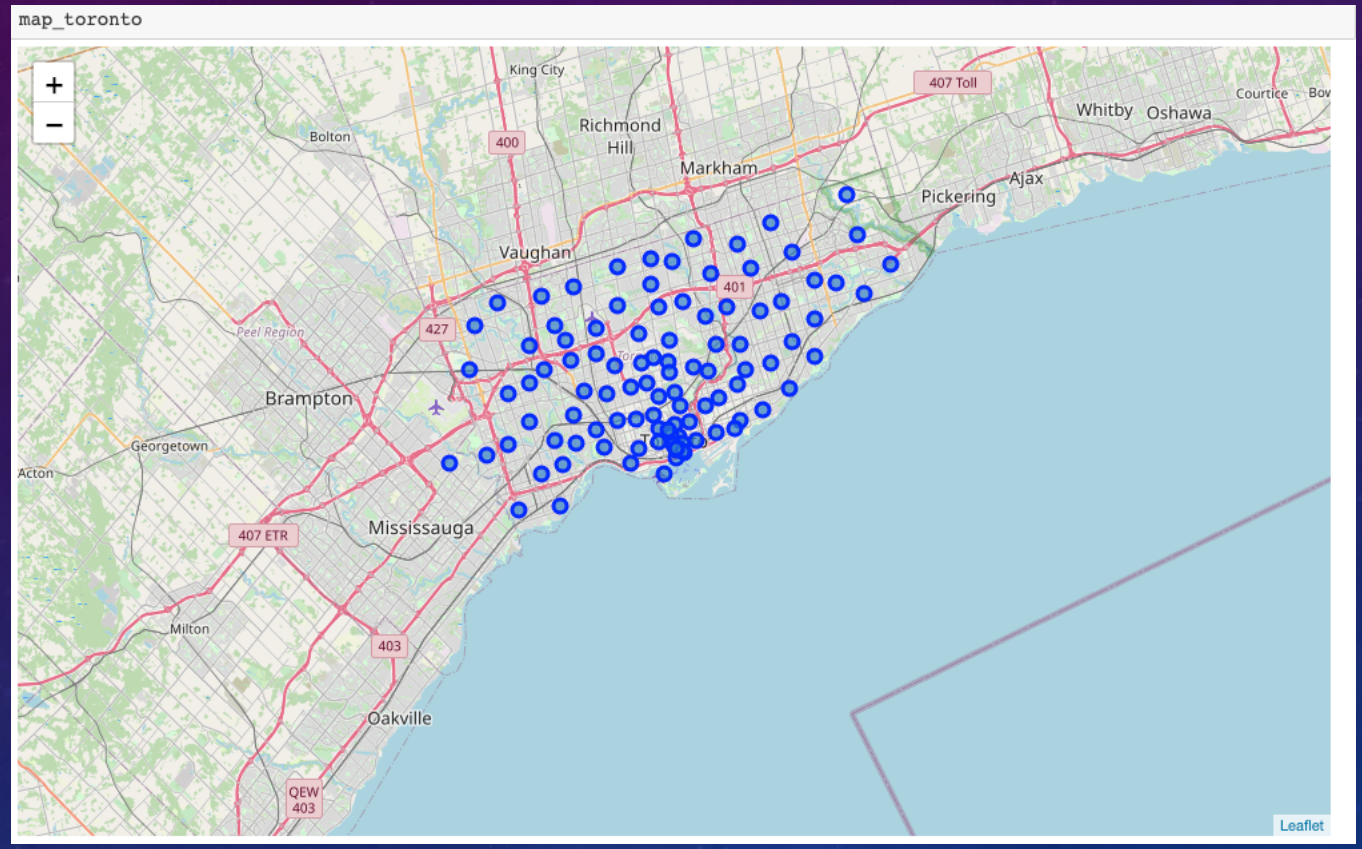
```
In [32]: toronto_venues.head(10)
```

Out[32]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Lawrence Park	43.728020	-79.388790	Lawrence Park Ravine	43.726963	-79.394382	Park
1	Lawrence Park	43.728020	-79.388790	Zodiac Swim School	43.728532	-79.382860	Swim School
2	Lawrence Park	43.728020	-79.388790	TTC Bus #162 - Lawrence-Donway	43.728026	-79.382805	Bus Line
3	Davisville North	43.712751	-79.390197	Homeway Restaurant & Brunch	43.712641	-79.391557	Breakfast Spot
4	Davisville North	43.712751	-79.390197	Summerhill Market North	43.715499	-79.392881	Food & Drink Shop
5	Davisville North	43.712751	-79.390197	Sherwood Park	43.716551	-79.387776	Park
6	Davisville North	43.712751	-79.390197	Winners	43.713236	-79.393873	Clothing Store
7	Davisville North	43.712751	-79.390197	Best Western Roehampton Hotel & Suites	43.708878	-79.390880	Hotel
8	Davisville North	43.712751	-79.390197	Subway	43.708378	-79.390473	Sandwich Place
9	Davisville North	43.712751	-79.390197	Gym	43.713126	-79.393537	Gym

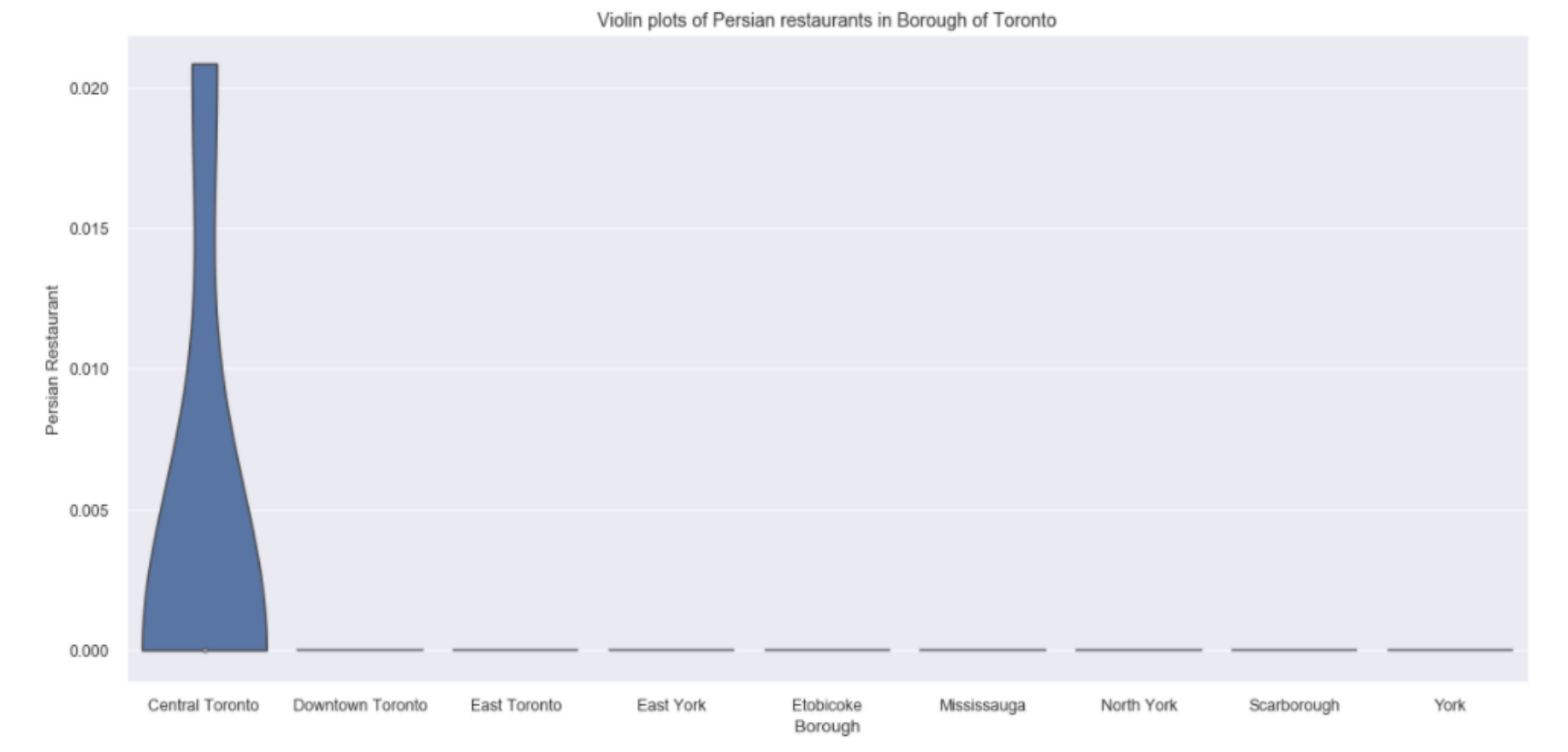
Exploratory Data Analysis

Folium Library and Leaflet Map



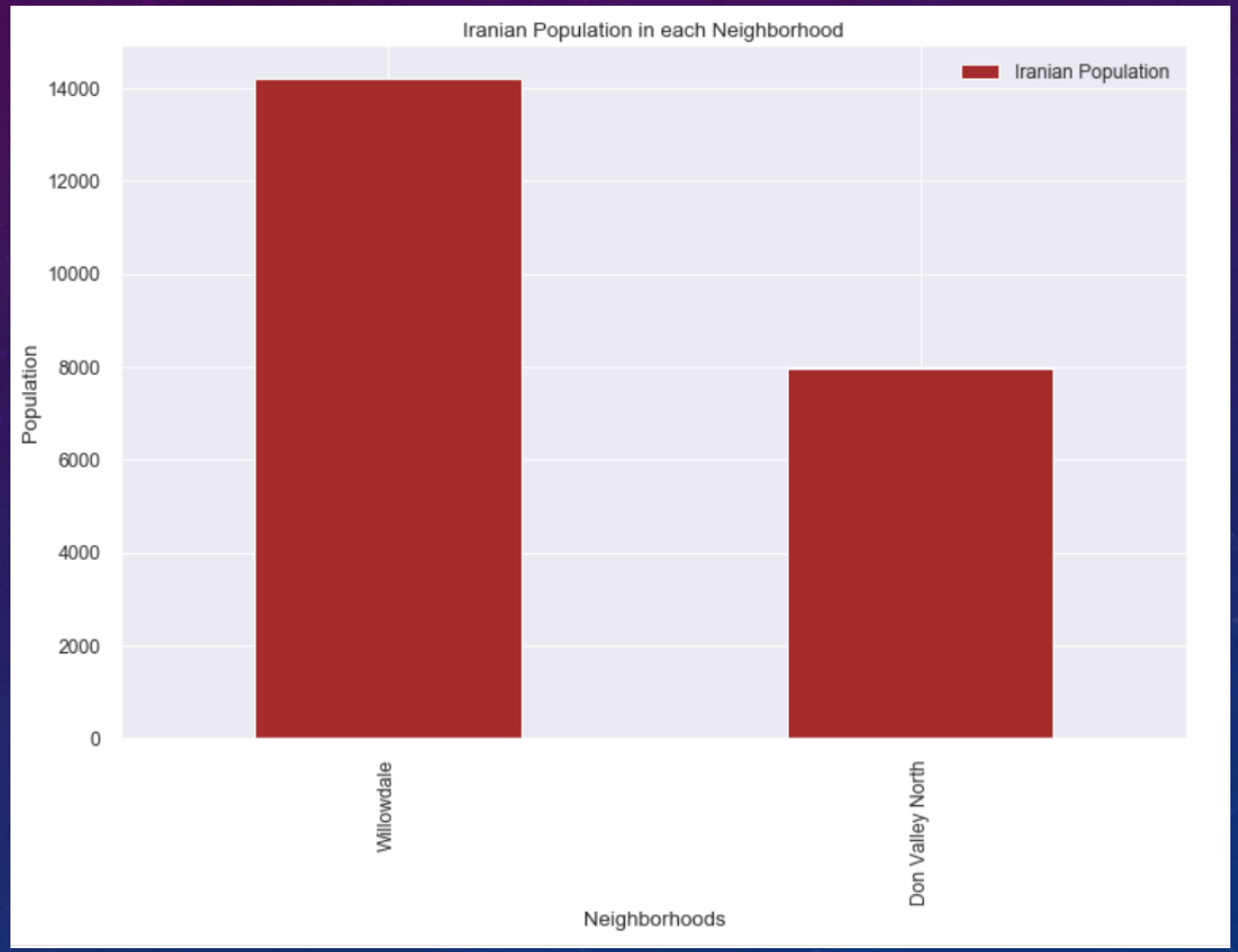
The neighborhoods in Toronto map

Relationship between neighborhood and Persian Restaurant



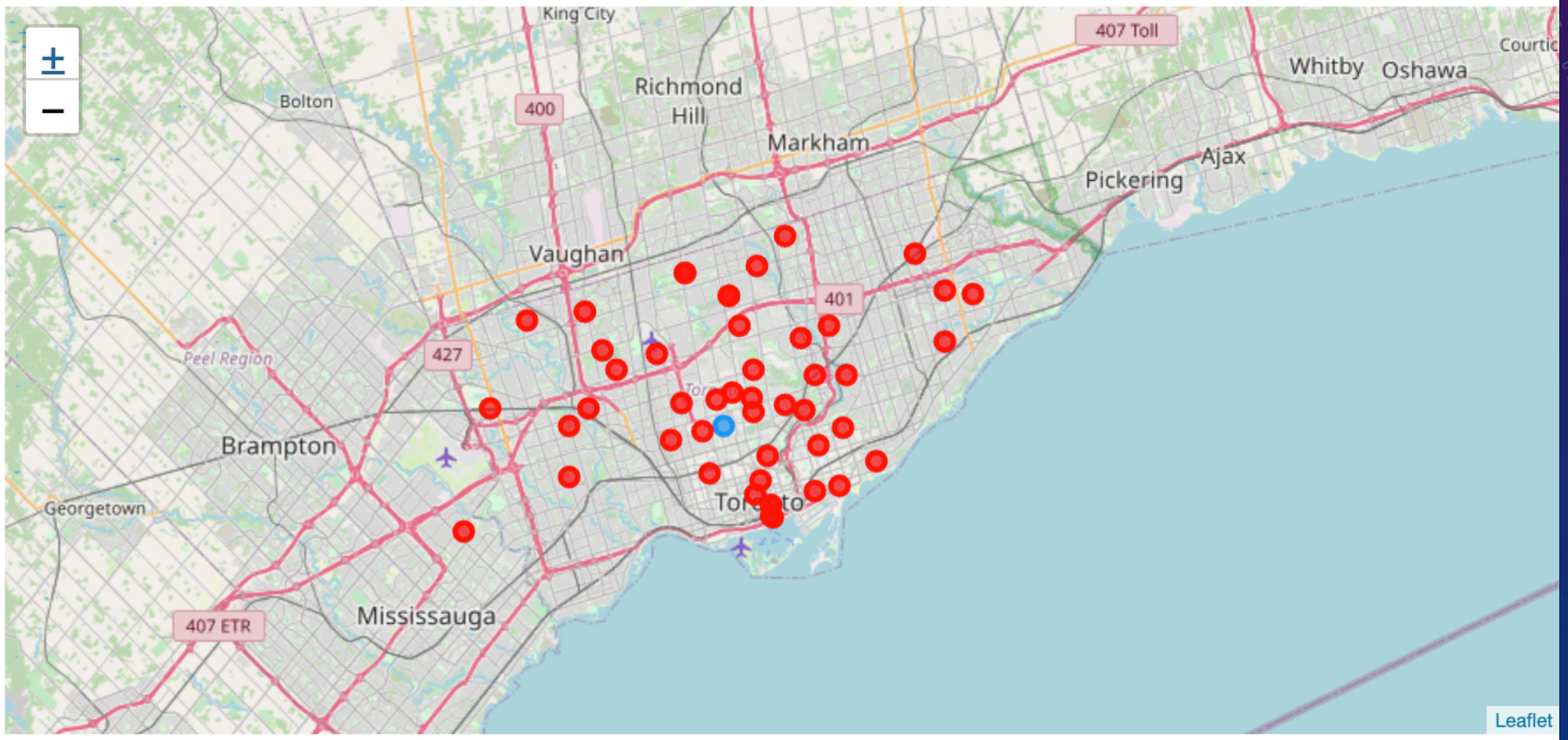
This plot helps in identifying the boroughs with densely populated Persian restaurants.

Iranian Population in each Neighborhood



Map of Persian Restaurants in Toronto

```
map_clusters
```



Results

We have reached the end of the analysis; in this section we will document all the findings from above clustering & visualization of the dataset. In this project, we started off with the business problem of identifying a good neighbourhood to open a new Persian restaurant. To achieve that we looked into all the neighbourhoods in Toronto, analysed the Persian population in each neighbourhood & number of Persian restaurants in those neighbourhoods to come to conclusion about which neighbourhood would be a better spot. We have used variety of data sources to set up a very realistic data-analysis scenario. We have found out that:

In those 11 boroughs we identified that only Central Toronto, Downtown Toronto, North York have high number of Persian restaurants with the help of Violin plots between Number of Persian restaurants in Borough of Toronto. In all the ridings, Willowdale, Don Valley North are the densely populated with Persian crowd ridings.

With the help of clusters examining & violin plots looks like Downtown Toronto, Central Toronto, East York are already densely populated with Persian restaurants. So, it is better idea to leave those boroughs out and consider only, East Toronto & North York for the new restaurant’s location.

After careful consideration it is a good idea to open a new Persian restaurant in Willowdale since it has high number of Persian populations which gives a higher number of customers possibility and lower competition since very less Persian restaurants in the neighbourhoods.

Discussion

According to this analysis, Willowdale borough will provide the least competition for the new upcoming Persian restaurant as there is very little Persian restaurants spread or no Persian restaurants in few neighbourhoods. Also looking at the population distribution looks like it is densely populated with Persian crowd which helps the new restaurant by providing high customer visit possibility. So, definitely this region could potentially be a perfect place for starting a quality Persian restaurant. Some of the drawbacks of this analysis are; the clustering is completely based only on data obtained from Foursquare API and the data about the Persian population distribution in each neighbourhood is also based on the 2016 census which is not up-to date. Thus, there is huge gap of 3 years in the population distribution data. Even Though there are lots of areas where it can be improved yet this analysis has certainly provided us with some good insights, preliminary information on possibilities & a head start into this business problem by setting the step stones properly

Conclusion

Finally, to conclude this project, we have got a chance to on a business problem like how a real like data scientists would do. We have used many python libraries to fetch the data, to manipulate the contents & to analyse and visualize those datasets.
We have made use of Foursquare API to explore the venues in neighbourhoods of Toronto, then get good amount of data from Wikipedia which we scraped with help of Wikipedia python library and visualized using various plots present in seaborn & matplotlib.
We also applied machine learning technique to predict the output given the data and used Folium to visualize it on a map. Some of the drawbacks or areas of improvements shows us that this analysis can be further improved with the help of more data and different machine learning technique.
Similarly, we can use this project to analysis any scenario such as opening a different cuisine restaurant or opening of a new gym and etc. Hopefully, this project helps acts as initial guidance to take more complex real-life challenges using data-science