APPLIED DATA SCIENCE CAPSTONE

Project: Finding the best location in Toronto for setting up an office (HUB) for a food delivery company

INTRODUCTION/ BUSINESS PROBLEM

- Lieferando is a Dutch dot-com company specialized in online food ordering and home delivery

- They are planning to set up an office in Toronto to deliver foods from restaurants to home

- They want to use Data Science to fin the best location for their office set up

DATA COLLECTION

- A list of postal codes in Canada where the first letter is M is taken from wikipedia.

- Basic information of borough , postcode and neighbourhood is gathered from this webpage (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

- The latitude and longitude coordinates of a given postal code is taken from this link (: http://cocl.us/Geospatial_data).

- Information of neighourhoods of each postcode as well as borough in Toronto is taken from at Fourthsquare APi website (https://developer.foursquare.com/).

METHODOLOGY

The whole work is divided into three sections.

- Step 1: Data importing and pre processing
- Step 2: Using the Final DataFrame with FourthSquare API to extract maps
- Step 3: Applying K-Means Clustering algorithm to cluster the restaurants

METHODOLOGY

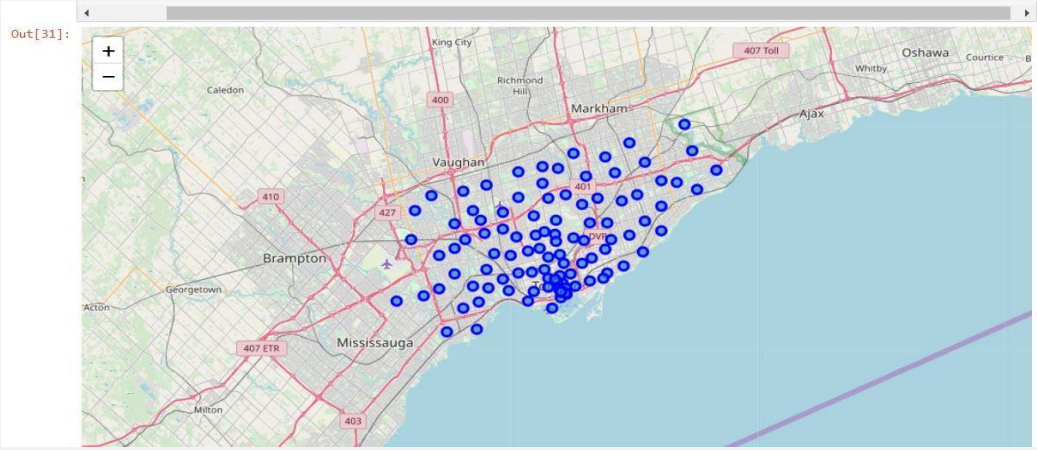
Step 1: Data Importing and Pre Processing

The Latitude and Longitude columns of Geospatial Data is concatenated with our Toronto Postcode data frame. Final dataframe look like this

	Postcode	Borough	Neighbourhood	Postal Code	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	M1B	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	M1C	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	M1E	43.763573	-79.188711
3	M1G	Scarborough	Woburn	M1G	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	M1H	43.773136	-79.239476
5	M1J	Scarborough	Scarborough Village	M1J	43.744734	-79.239476
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	M1K	43.727929	-79.262029
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge	M1L	43.711112	-79.284577
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West	M1M	43.716316	-79.239476
9	M1N	Scarborough	Birch Cliff, Cliffside West	M1N	43.692657	-79.264848

METHODOLOGY

Step 2: Using the Final DataFrame with FourthSquare API to Extract Maps



- A map of Toronto with neighbourhoods on top is created.
- This map shows neighourhoods of all the postcode
- DataFrame is further processes to find out the borough of highest number of neighbourhood's- which is found North York

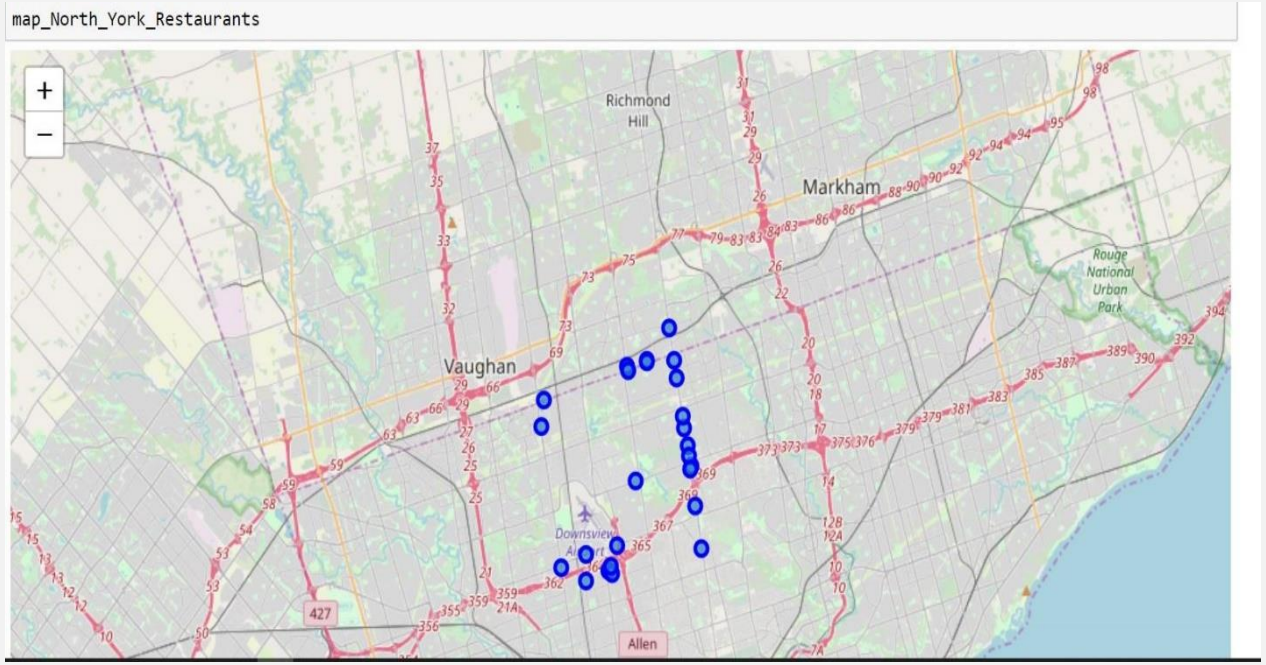
Then the dataframe is filtered by restaurant, since we need only the list of restaurants of North York

	name	categories	lat	lng
6	Kinka Izakaya	Japanese Restaurant	43.760161	-79.409827
8	Bagel Plus	Restaurant	43.755395	-79.440686
13	RH Courtyard Café	Restaurant	43.724874	-79.455536
15	Auberge du Pommier	French Restaurant	43.746962	-79.407879
23	Konjiki Ramen	Ramen Restaurant	43.766998	-79.412222

METHODOLOGY

Step 2: Using the Final DataFrame with FourthSquare API to Extract Maps

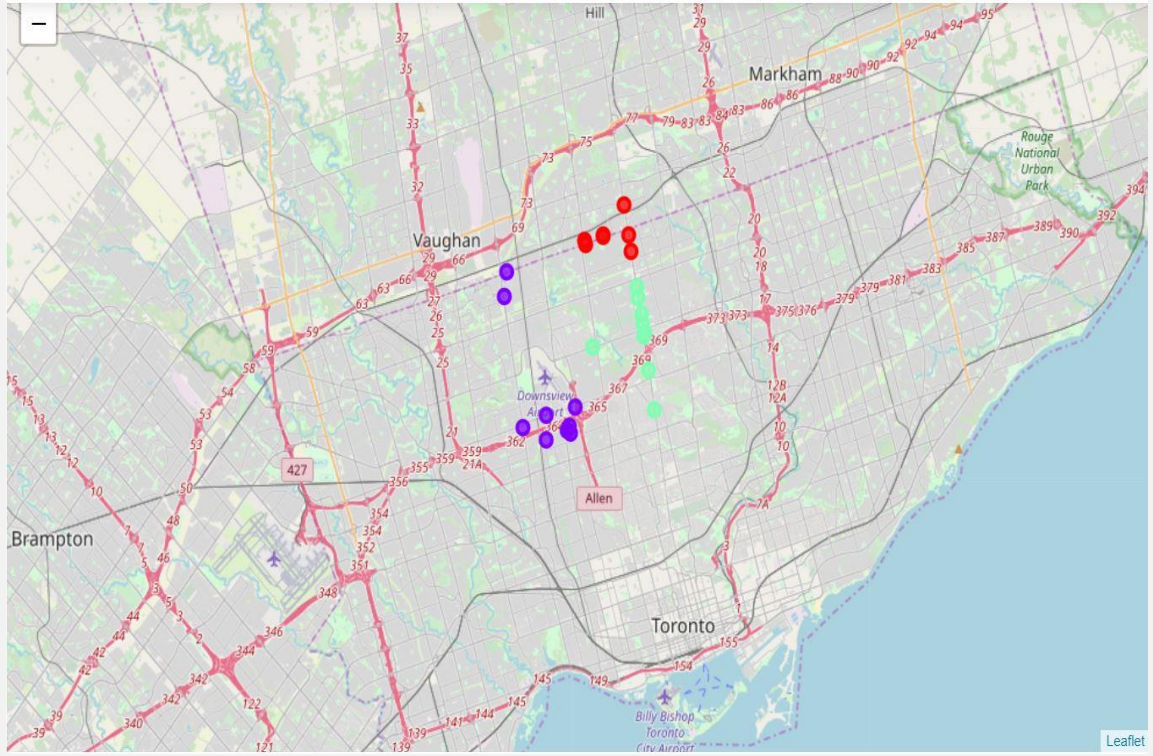
Then only the restautants of borough-NorthYork was created



METHODOLOGY

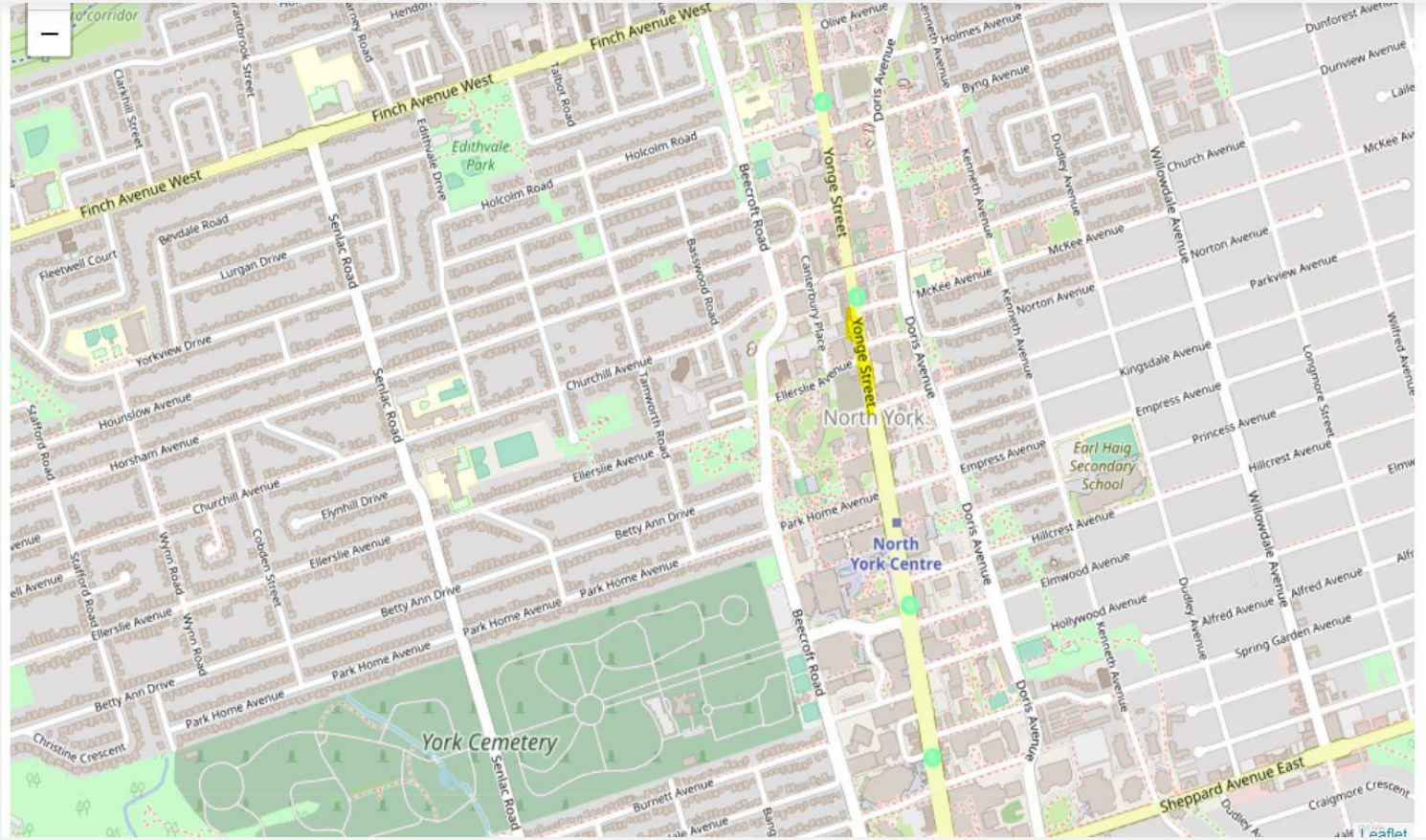
Step 3: Applying K-Means Clustering algorithm to cluster the restaurants

After executing K-Means Clustering algorithm, the output was visualized on Toronto map using folium library



RESULTS

If we zoom in we see that the location of green clusters is **Yonge Street , North York**. Even some of the red marked restaurants also Yonge Street.



DISCUSSION

- some of the data were lost while data pre-processing- which harms the project accuracy
- If there were no limit then I might have got more list of neighbourhoods which finally would have helped to find more accurate results.

THANK YOU!