The Battle of Neighborhoods
(week 4 and week 5)

CAPSTONE PROJECT

Context

- Introduction
- Data
- Methodology
- Results and Discussion
- Conclusion

Introduction

- The aim of the project is to find a perfect location for comfortable Living/Business/Travelling-with a few mouse clicks .

Comfortable Yourself Application

- This application helps to find a perfect area for you keeping in mind your preferences of the day, even help to explore the things around the present location.
- Thus giving you the quality time options for your time.

Data

- Wikipedia help to get the Postal Code, Borough and Neighbourhood in Toronto
- Geospatial data of Toronto contains all the geographic coordinates of each postal code
- Foursquare API allows to obtain the data on what venus are located at each neighbourhood
- Random user data ,with a random number (from 1 to 10) of preferences to check, how our system works

Methodology

- Pandas Library to scrap the table from HTML page

	Postal Code	Borough	Neighbourhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront
5	M6A	North York	Lawrence Manor, Lawrence Heights
6	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

Methodology

- Geospatial Data of Toronto contains the geographical coordinates of each postal code

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

Methodology

- Foursquare API allows to obtain the data on what venues are loated in eact neighbourhood
- Apply one hot encode to get a new table with the neighbourhood as the index and percentage of each category available in the neighbourhood

Methodology

- Random User - To generate the random User to use for a test of the system
  - Select a random number from 1 to 10 to represent the amount of the categories selected by the user
  - Create a table with the categories as the columns and one row, where the value are 1 if the user has the category in his list and 0 for vice versa

Methodology

Recommendation System

- Multiply the user profile with the table that has the neighbourhood and the weight of each category
- Results is a matrix with the score of each neighbourhood
- The higher score ,the better the neighbourhood fit the user interest

Results and Discussions

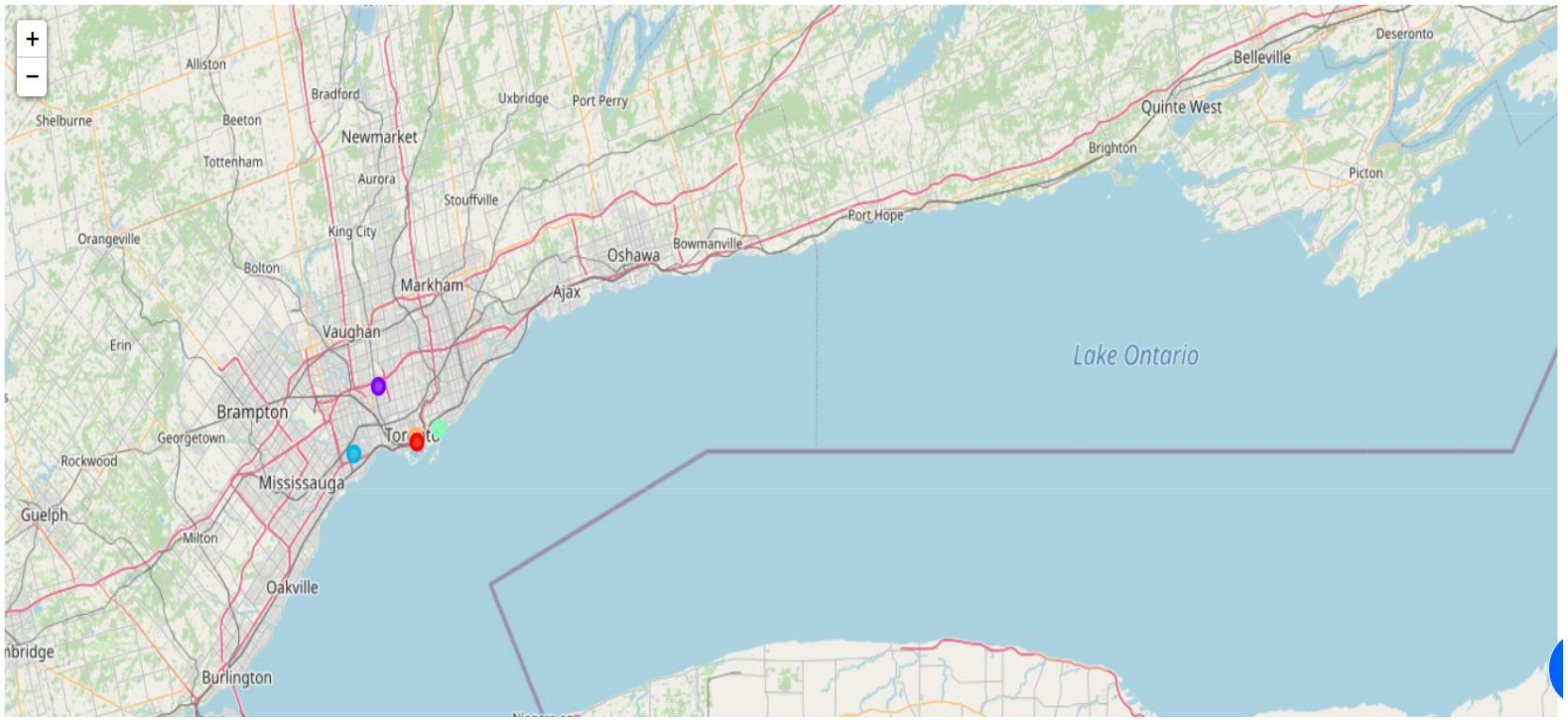
- In this prototype we generated some random categories for the users

```
['Doner Restaurant', 'Taiwanese Restaurant']
```

Results and Discussions

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Score
0	M4X	Downtown Toronto	St. James Town, Cabbagetown	43.667967	-79.367675	0.020833
1	M5T	Downtown Toronto	Kensington Market, Chinatown, Grange Park	43.653206	-79.400049	0.013514
2	M5R	Central Toronto	The Annex, North Midtown, Yorkville	43.672710	-79.405678	0.000000
3	M6G	Downtown Toronto	Christie	43.669542	-79.422564	0.000000
4	M6E	York	Caledonia-Fairbanks	43.689026	-79.453512	0.000000

Results and Discussions



Results and Discussion

- 2 best neighbourhoods for our user are “M4X :Downtown Toronto”,”M5T: Downtown Toronto”,and “M5R :Central Toronto”
- Difference of the score amount the 5 neighbourhoods is not big. A probable reason is that categories, which our user chose are more or less common, they don’t include anything extraordinary as “Airport Food Court”

Conclusion

- This is a sample content-based recommendation system that still need to be improved
- The data and the algorithm need more date and accuracy, especially for some small towns with a few venues
- There are more parameters, which could be use for the search, as distance to work place, bus station and etc