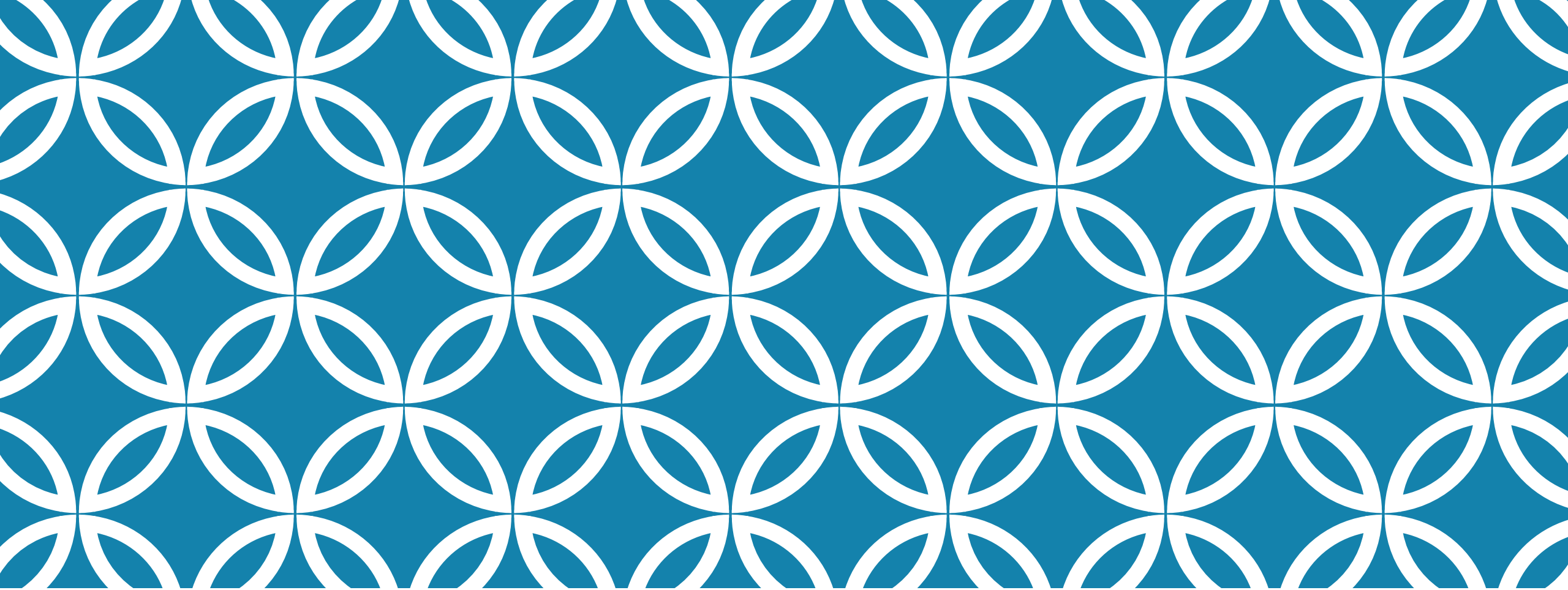


CAPSTONE PROJECT: THE BEST PLACES TO VISIT IN A CITY

Submitted by
Ayush Saxena

Introduction - Business Problem Description

Scenario and Background

Every city has its own different flavors, However, sometimes due to time constraints it gets difficult for the person to identify 'top venues' of different categories that a city is famous for

Problem Statement

Can we use data science to segment city based on its 'top venues' for different categories that can help a person in exploring the city better.?

Target Audience

It will be having two major target groups:

People visiting new places

Travel Guides and tourism companies

Data Description

Data Sources:

1) Wiki page having the details of the postal codes, Borough and Neighborhood of Canada was used to extract the data and read into data frame using pandas.
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
2) Google API was used to gather the coordinates for each of the postal code

Data Cleaning Process

- Initially all the missing column values of the column 'Borough' were removed as they were of no significance.
- Neighbourhoods having same postal codes were combined into one separated by commas.
- For each of the postal code respective Latitude and Longitude coordinates were added using Google API.
- Finally, the postal codes column was dropped as it was irrelevant in the further analysis.

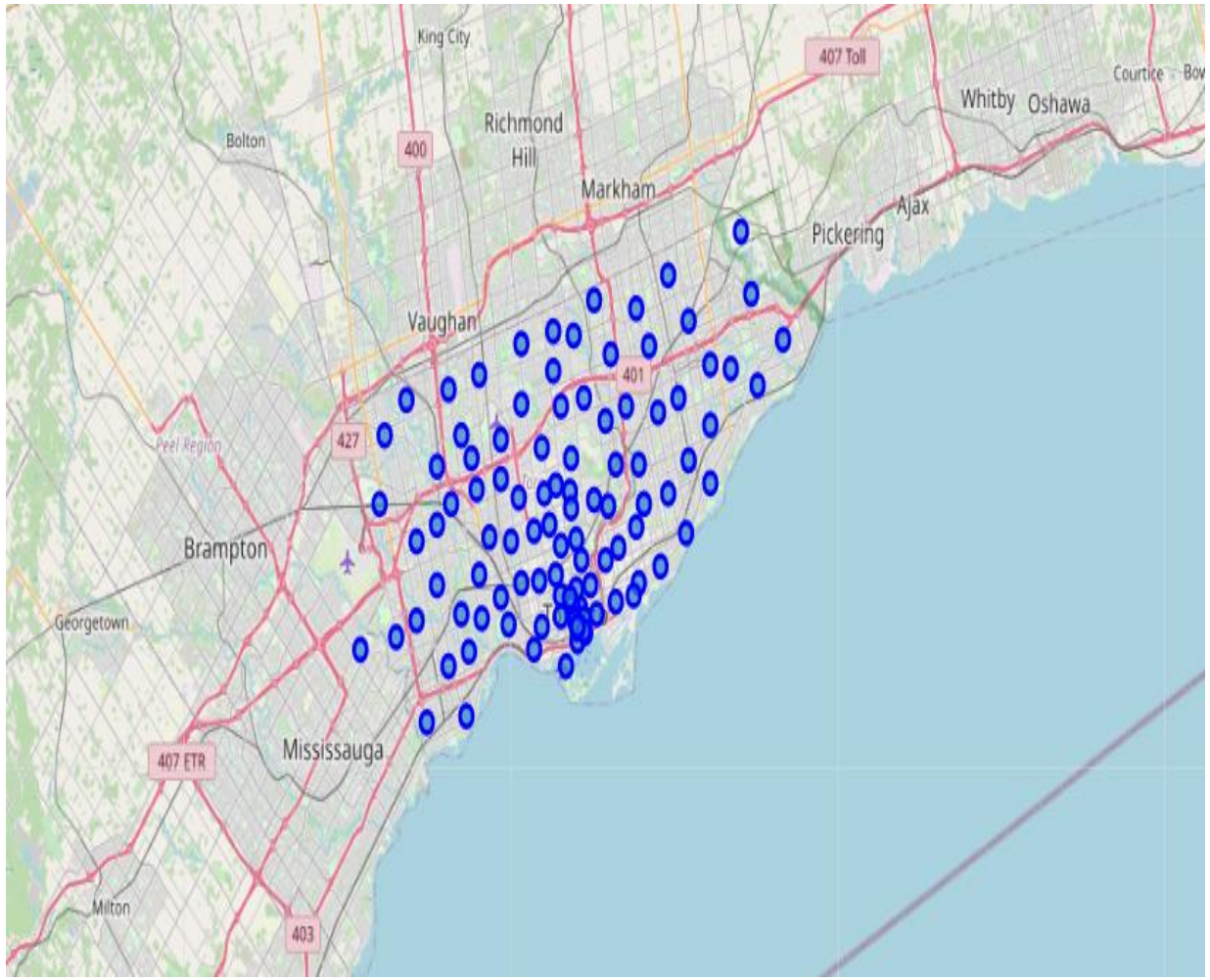
Dataset Ready to use

	Borough	Neighborhood	Latitude	Longitude
0	Scarborough	Malvern , Rouge	43.806686	-79.194353
1	Scarborough	Rouge Hill , Port Union , Highland Creek	43.784535	-79.160497
2	Scarborough	Guildwood , Morningside , West Hill	43.763573	-79.188711
3	Scarborough	Woburn	43.770992	-79.216917
4	Scarborough	Cedarbrae	43.773136	-79.239476

Methodology Used (1/2)

Steps involved in the analysis

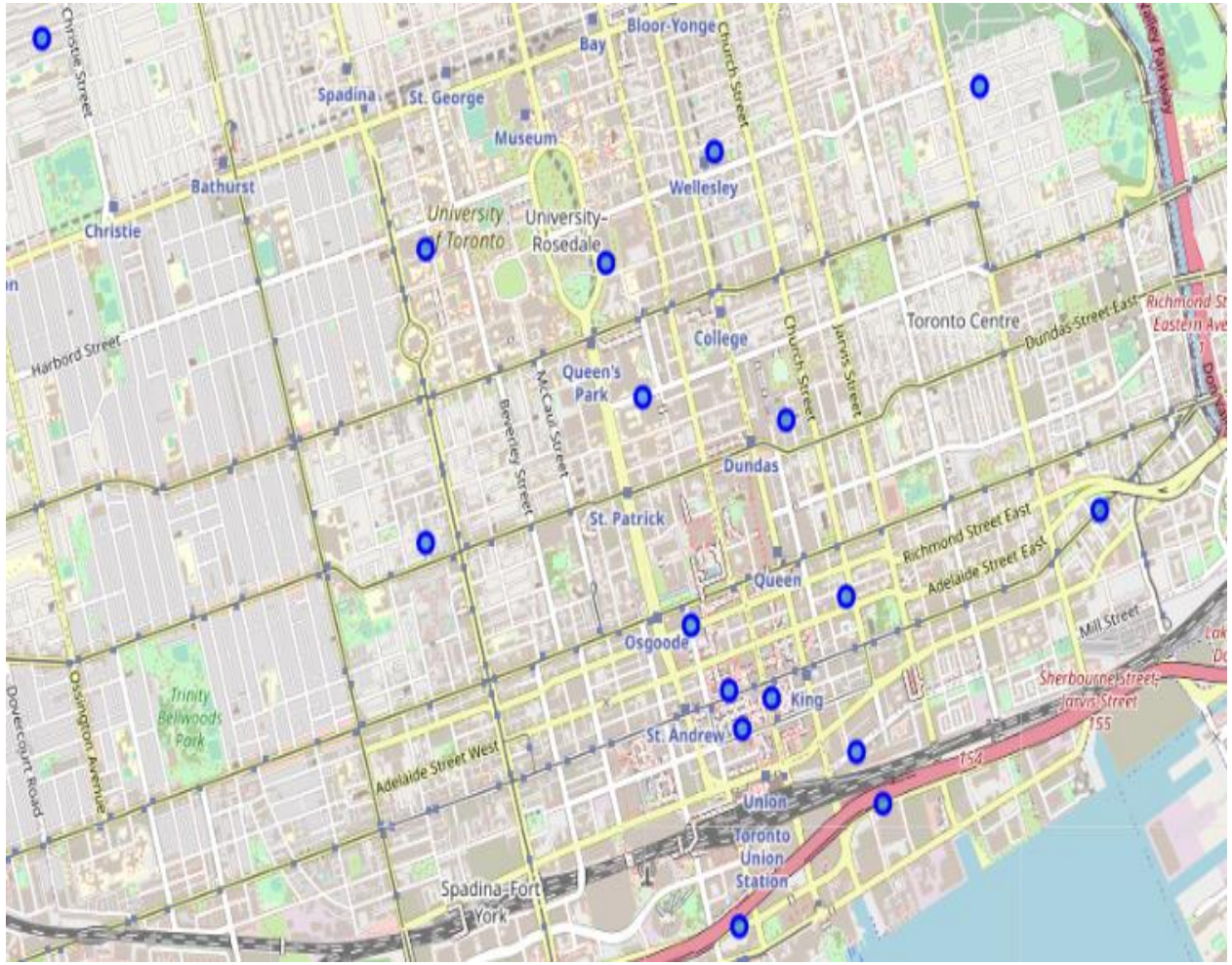
- Initially coordinates of Toronto city were obtained and map was rendered using folium.
- All the neighborhoods of the city were superimposed on the map as shown
- For the project study one of the Borough was chosen for analysis, in our case it is ‘Downtown Toronto
- K means clustering was used to form clusters based on the top venues to visit in the particular Borough.
- Appropriate k was selected using the minimum SSE



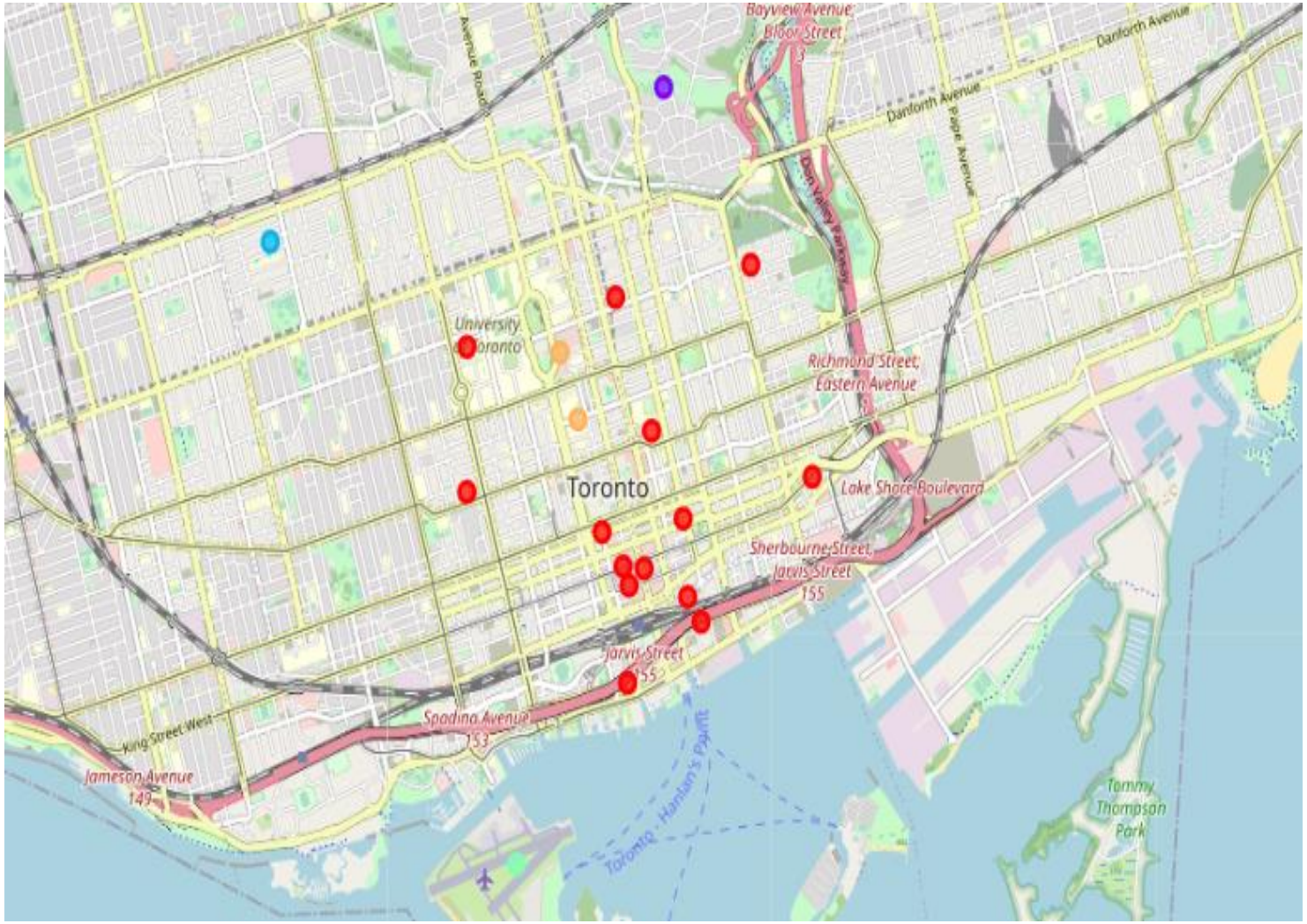
Methodology Used (2/2)

Steps involved in the analysis

- Foursquare API was used to obtain the list of top venues around the neighborhoods.
- Limit of top 100 and radius of within 500 meters was chosen for the study
- The same process was repeated for all the neighborhoods in the Borough, 'Downtown Toronto'
- Once data was collected from Foursquare API, it was grouped based on mean of the frequency of occurrence.
- Top 10 locations per neighborhood were clubbed and then clusters were formed



Results of the analysis



Five clusters were formed based on the top venues to visit:

Cluster 1: **Coffee shops**

.

Cluster 2: **Parks** and **playgrounds**

Cluster 3: **Grocery stores**

Cluster 4: **Airport services**

Cluster 5: **Restaurants**

Hence, if a person wishes to have explore the best coffee in the town then it can directly visit the cluster 1 of the Downtown Toronto as it is the aggregation of the best available in the city and likewise for any other service as required by the person, respective clusters will serve as a much better map.

Discussion and Conclusion

Below are the assumptions made while working on this project

- The clusters/groups are created for the neighbourhoods of Toronto with available data. These can be improved or enhanced using refining the data.
- The value of k chosen for the study is using the best possible guess estimate and can be improved further.
- This project can be replicated for analysis of any locations having Foursquare data.
- The project was done for one of the Borough, Downtown Toronto the same can be extended to other Boroughs and their respective neighbourhoods.
- Regular updating or addition of more data to Foursquare database could refine the results more.

Conclusion

With the help of data science, this project has been successfully been able to segment the city into various clusters that can help a person to explore the city in a much better way with least amount of time. Clusters would guide the tourists to appropriate areas where they can explore the venues of their choices.


Thank you!