Capstone Project

THE BATTLE OF NEIGHBORHOODS – TORONTO CITY

- SHASHI KANT

Introduction

- The main objective of this project is to cluster and segment various localities of Toronto city on the basis of neighborhoods, famous venues, available ameneties.
- It will help different segment of people to take smart and selective decision to choose neighborhood as per their requirements.
- This project aims to select various features and segment areas onthe basis of similarities in the neighborhood.
- Through this project main idea is to help wide variety of audience to help them take better decisions when it comes to choosing the neighborhood which is the basic yet most important need of life.

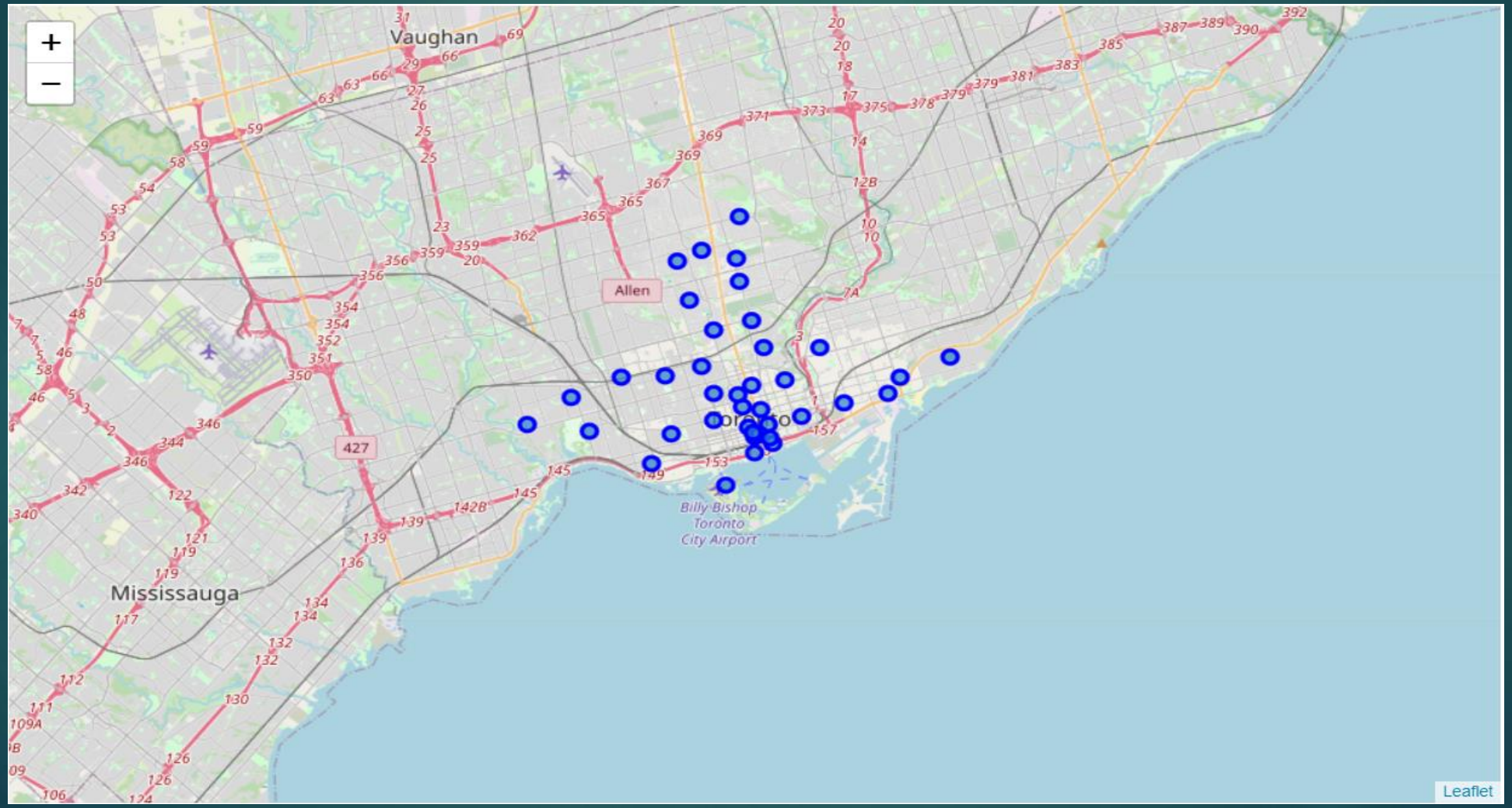
Target Audience

- Immigrants to Canada
- Students
- Business Groups
- Real-Estate
- Startups
- Jobseeker
- Eateries

Data Source

- Wikipedia : To get postalcodes of Canada cities
- Geo Spatial csv : To get Latitude and Longitude of locations
- Four Square : To get Four Square API data

Map Of Toronto City



Methodology

- Approach for Clusterting:

  To compare the similarities of two cities, we decided to explore neighbourhoods, segment them, and group them into clusters to find similar neighbourhoods in a big city like New York and Toronto.

- Cluster data using unsupervised machine learning: k-means clustering algorithm

K-Means Clustering Algorithm

- Sample data to show below neighborhood classified into same cluster

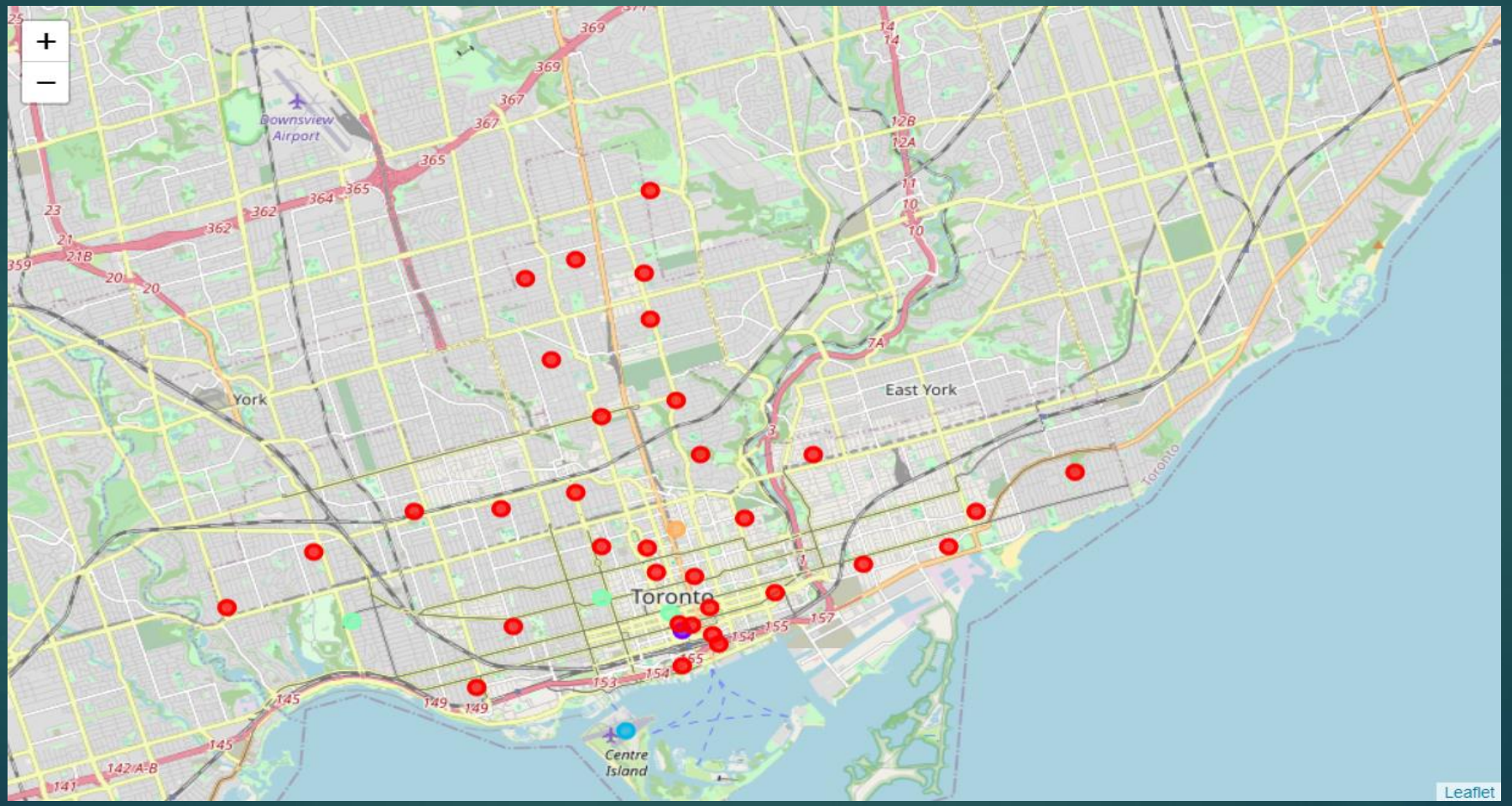
Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
The Beaches	43.676357	-79.293031	0	Park	Trail	Pub	Health Food Store	Wine Bar	Distribution Center	Department Store	Dessert Shop	Diner	Discount Store
The Danforth West, Riverdale	43.679557	-79.352188	0	Greek Restaurant	Coffee Shop	Italian Restaurant	Ice Cream Shop	Furniture / Home Store	Restaurant	Bubble Tea Shop	Bakery	Pub	Pizza Place
India Bazaar, The Beaches West	43.668999	-79.315572	0	Fast Food Restaurant	Sushi Restaurant	Movie Theater	Sandwich Place	Brewery	Restaurant	Italian Restaurant	Ice Cream Shop	Steakhouse	Fish & Chips Shop
Studio District	43.659526	-79.340923	0	Coffee Shop	Gastropub	Bakery	Brewery	Café	American Restaurant	Yoga Studio	Comfort Food Restaurant	Seafood Restaurant	Cheese Shop
Lawrence Park	43.728020	-79.388790	0	Bus Line	Park	Swim School	Deli / Bodega	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Donut Shop	Doner Restaurant	Dog Run

Work Flow

- Fetching the postal codes of Canadian cities from Wikipedia using scrapping method and apply filter to get Toronto city data.
- Fetch Longitude and Latitude details from geo spatial csv file and merge these details with Toronto City data.
- Using credentials of Foursquare API features of near-by places of the neighbourhoods would be fetched along with venue details.
- Target would be to fetch neighborhood and venue data within the radius parameter of 500 metres.
- Prepare single data repository using all above data and apply clustering technique to segment various localities on the basis of neighborhood.

Result

- Map showing different clusters in Toronto city



Discussion

- The main objective of this case study is to suggest various options of localities on the basis of neighborhood, venues, amenities to wide variety of people who are migrating to Toronto city to open a business, immigrants, people on business trip etc.
- Connectivity to the airport, bus stand, city center, markets and other daily needs things nearby.

Conclusion

- In this project, using k-means cluster algorithm the neighborhoods have been segmented into five different clusters for different lattitude and logitude from dataset, which have very-similar neighborhoods around them.
- The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision better with confidence.

Libraries Used

- Pandas : To create and maintain dataframes
- Matplotlib : For visualizations
- Scikit-Learn : Package for machine learning algorithms
- Beautiful Soup : For web scrapping
- Geocoder : For longitude and latitude
- Folium : For visualization of different clusters on map
- Json : To handle json data