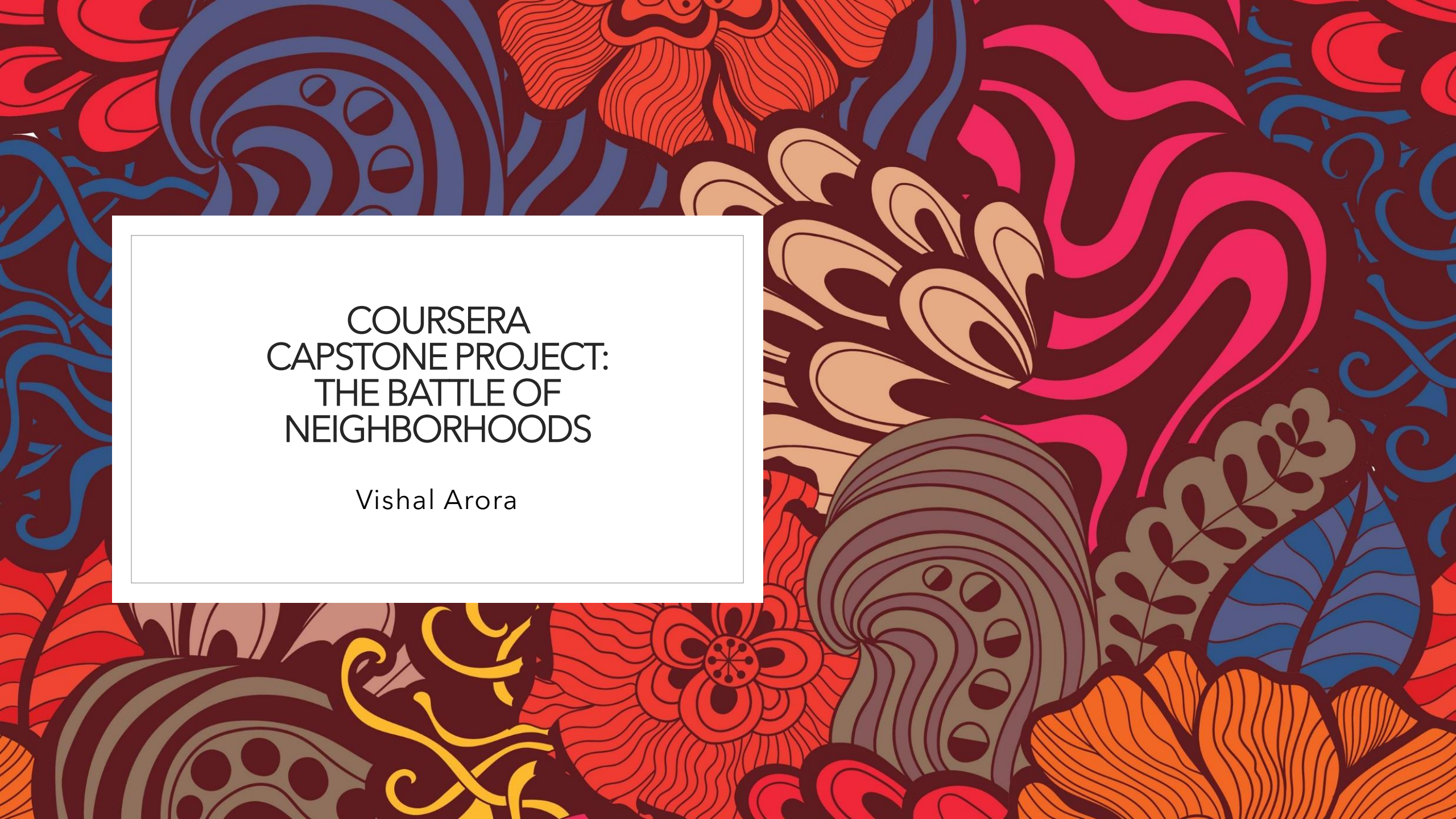


COURSERA CAPSTONE PROJECT: THE BATTLE OF NEIGHBORHOODS

Vishal Arora

Introduction: Business Problem

- **Find a suitable location for *EAT-TO-GET-FIT* first outlet in Toronto**
- The fitness industry in Canada is growing at a rapid pace of more than 5% year on year
- We all know that exercise plays a vital in becoming fit. However, the *DIET* plays an even more important role in the overall process.
- Introducing *EAT-TO-GET-FIT* outlets in Canada which will take care of the complete meal for the person enrolling for the service.
- In this project, we will put our data science mind to work and find out the most promising location in Toronto. We plan to have the outlet near the Gym or Fitness Centers because post the workout, people need food to re-energize themselves.
- The **target audience** for this project will be the entrepreneur who wants to find the location to open an eat to get fit outlet(restaurant)

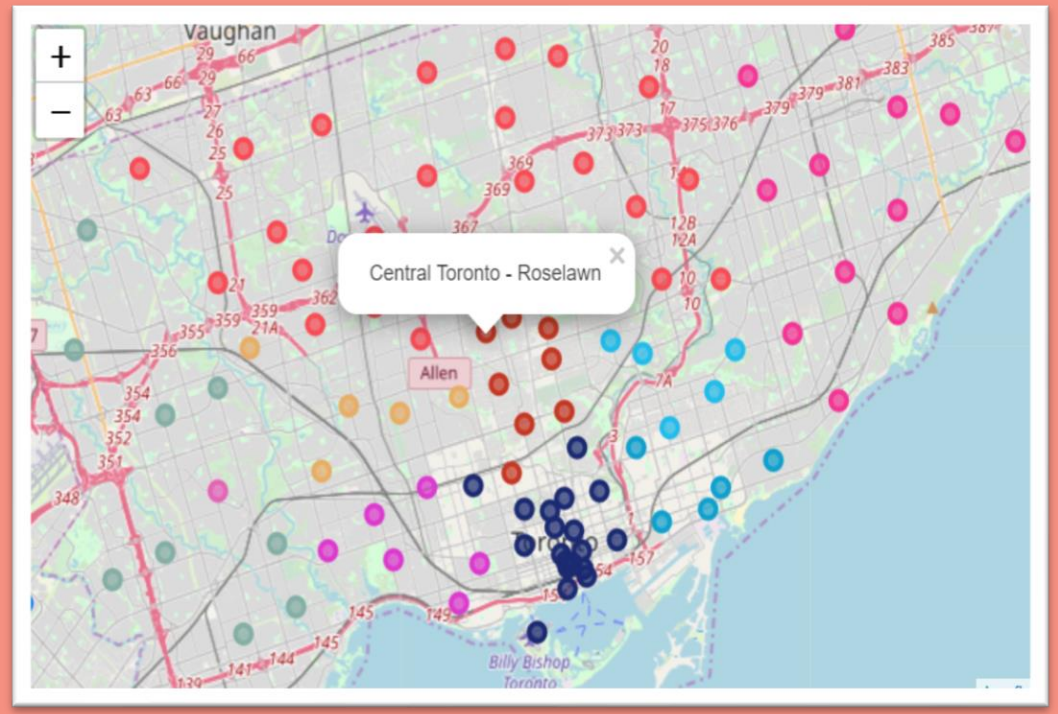
Data Objective

- Data Requirements:
  - List of neighborhoods in Toronto, Canada
  - Latitude and Longitude of these neighborhoods
  - Venue Category data to understand the type of venues in each neighborhood

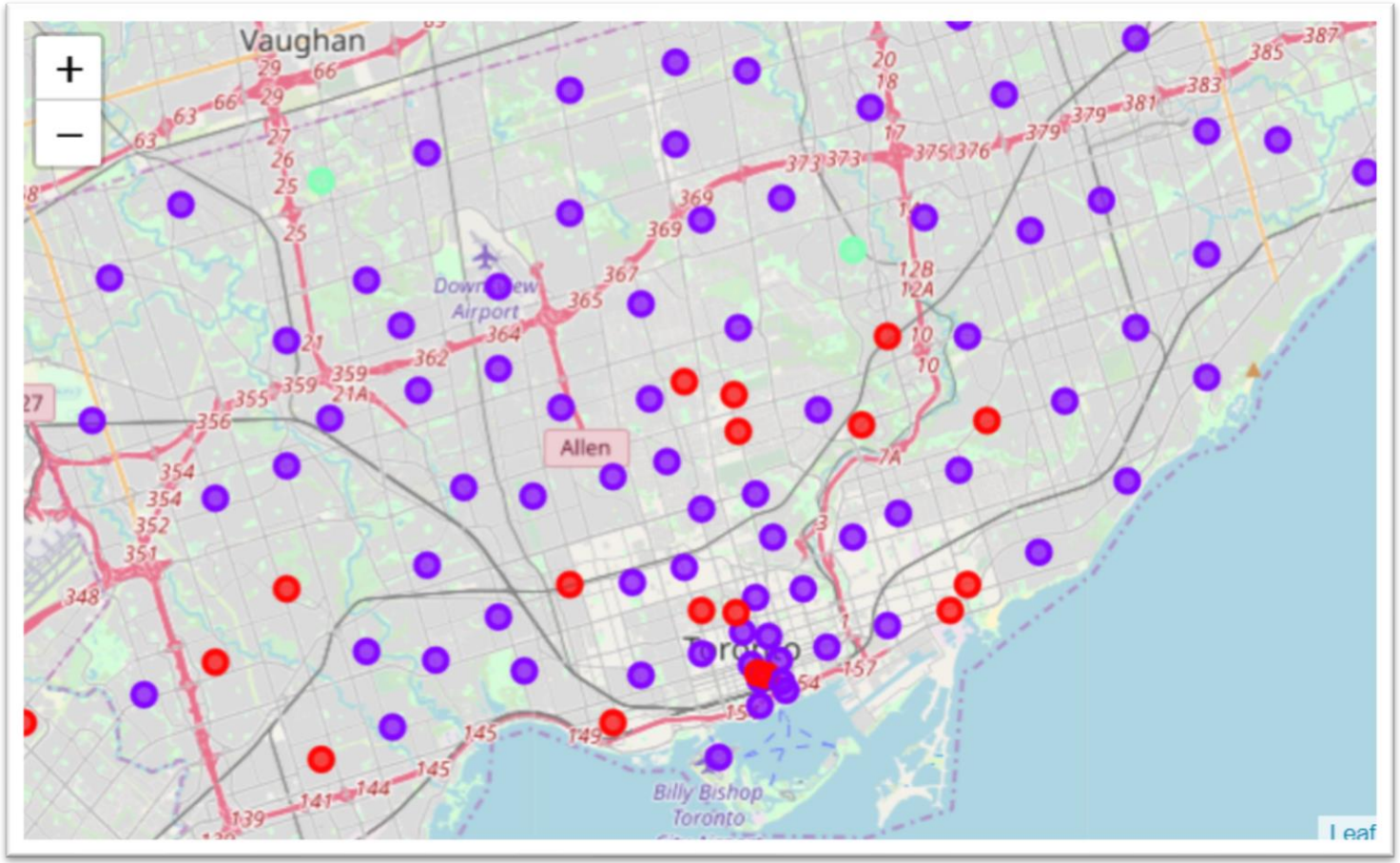
- Data Fetching:
  - Using the technique of web-scrapping to gather the list of neighborhoods in Toronto
  - Installing the Geocoder package to fetch the Latitudes and Longitudes the neighborhoods
  - Calling Foursquare API to get the details of various types of venues in these neighborhoods

Out of all the features considered, we will consider Venue Category equal to Gym to segment the data. This data will be used for data analysis (using techniques like clustering) to come up with the most suitable location for opening the first outlet of *EAT-TO-GET-FIT*.

Methodology



- Web scrap to extract the borough and neighborhood data in Toronto, Canada
- Fetch the co-ordinates of these neighborhoods using the postal code
- Call the Foursquare API to get data on venues in every neighborhood.
- Data wrangling to clean it and make it suitable for analysis
- Visualize the neighborhoods on the Toronto Map
- Use the one-hot encoding technique on Venue Category
- Calculate the mean occurrence of “Gym” in every neighborhood
- Perform Cluster analysis
- Visualize the data frame to understand the clusters formed
- Identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location



Results

We have used K-means clustering to form 3 kinds of clusters:

Cluster 2: Neighborhoods with a lot of Gyms in the area,

Cluster 1: Neighborhoods with no Gym in the area,

Cluster 0: Neighborhood with less or no Gyms in the area

Based on the clusters formed, there are 2 Neighborhoods with high number of Gyms/Fitness Centers. These are **Don Mills North** and **Downsview Northwest**. These locations belong to the **Cluster 2** and have **Turquoise** colored markers in the MapView

Discussion

- Based on the results drawn from the analysis, I would like to recommend ***Don Mills North*** and ***Downsview Northwest*** for opening the first outlet of ***EAT-TO-GET-FIT***.
- Next step will be to incorporate factors like size of the available space, distance from gym/fitness center, price of the available space, demographic factors of the neighborhood, etc. recommend the final specific location is outside the scope of this project due to the timeline attached.
- However, the key take-away would be that these 2 areas are great to start with the on-ground street exploration for the ultimate outlet location.

Conclusion

- In this study, I have analyzed the neighborhoods and venue categories in Toronto, Canada.
- As I wanted to open the first outlet near the fitness centers/gyms, I built clustering model to identify the most potential areas.
- Based on the data, Don Mills North and Downsview are the 2 most suitable areas for opening the outlet.
- This model will be very useful when we will expand the chain of EAT-TO-GET-FIT to other parts of Toronto and later, Canada.

THANK YOU