Data Science Capstone project

Opening a Vietnamese Restaurant in Toronto, Canada
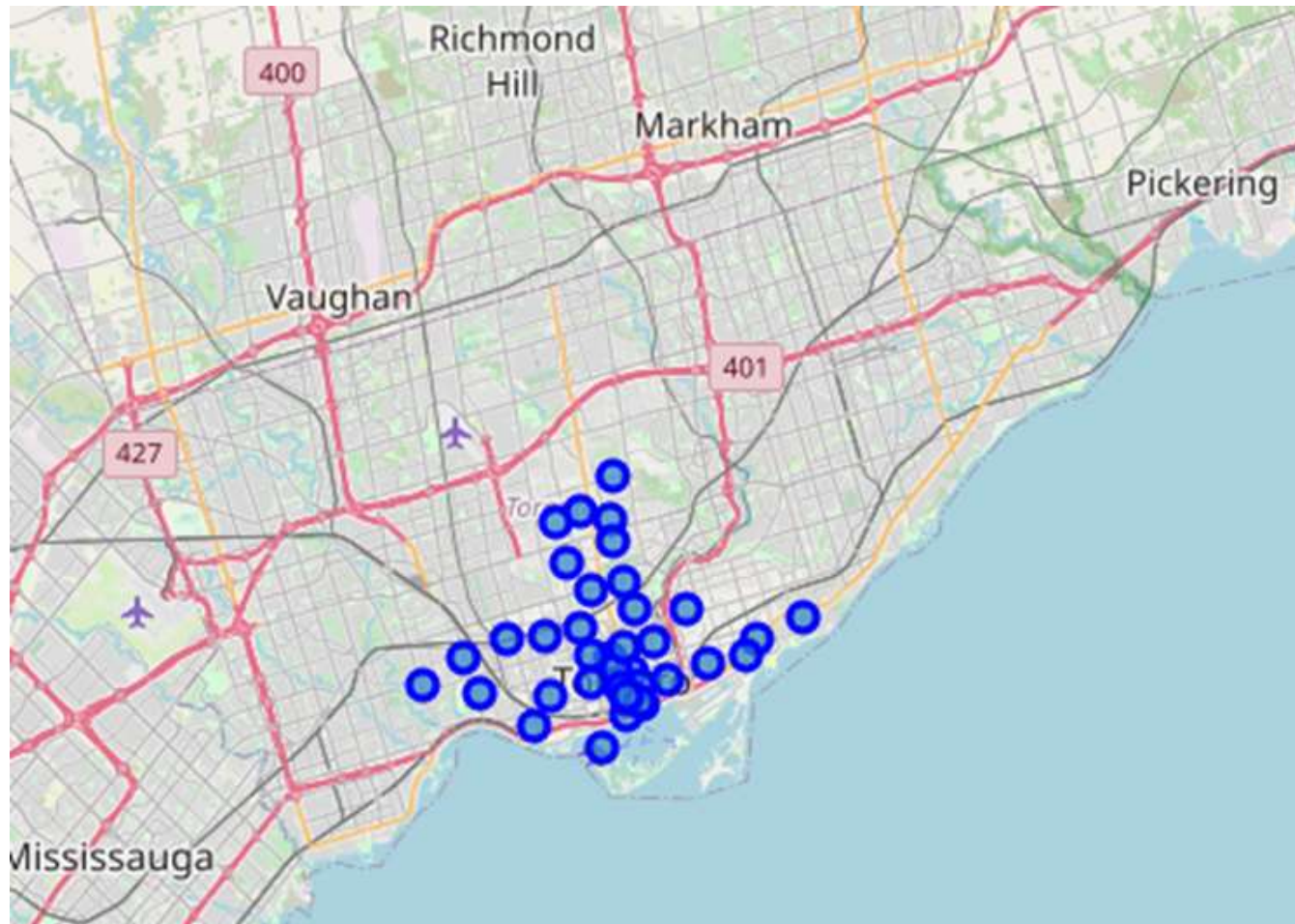
Target: Restaurant business owner

▶ **Background:** Toronto is a diversified city which welcomes different cultures in the world but this advantage could become a waste if one does not have a clear strategy, especially, of where the best location to open a restaurant should be.

▶ **Topic description:** Vietnamese cuisine as topic for the research of the optimal neighborhood(s) to open a restaurant in Toronto

▶ **Target Audience:**

    ▶ Business owner seeking to open a new restaurant

    ▶ Investors targeting to invest in Food and Beverage sector in Toronto

▶ **Why helpful?**

    ▶ Provide the target audience a general picture on how all types of venues are distributed across all the neighborhoods in Toronto

    ▶ Detailed comparison on how the current restaurants are spread through the city, across the neighborhoods and by cuisine, especially Asian and Vietnamese

Data Acquisition and Pre-processing

► **Geographic data on neighborhoods in Toronto:**

  ► Using BeautifulSoup to scrape data from Wikipedia page List of Postal Code of Canada

  ► Latitude and longitude coordinates of each neighborhood by importing the csv file containing the latitudes and longitudes of neighborhoods in Canada from cocl.us/geospatial_data.

  ► The total number of neighborhoods after cleaning up is 100

► **Data on venues using Foursquare API:**

  ► Get the top 100 venues within a radius of 500 meters for all the neighborhoods in Toronto and list down the list of the extracted venues together with their category, latitude and longitude

  ► Number of venues after cleaning up is 1603

Exploratory Data Analysis (1)

▶ **Visualization of the neighborhoods in Toronto:** map out the geographical visualization of the neighborhoods in Toronto for further in-depth analysis

Exploratory Data Analysis (2)

► **Filtering venue data to focus on restaurants, Asian restaurants and Vietnamese restaurants:**

```
# numbers of restaurants by cuisine
viet_restaurants = toronto_restaurants[ toronto_restaurants['Venue Category'].isin(viet_restaurant_list) ]
asian_restaurants = toronto_restaurants[ toronto_restaurants['Venue Category'].isin(asian_restaurant_list) ]

print('Total number of restaurants:', len(toronto_restaurants['Venue'].unique()))
print('Total number of Asian restaurants:', len(asian_restaurants['Venue'].unique()))
print('Total number of Vietnamese restaurants:', len(viet_restaurants['Venue'].unique()))
```
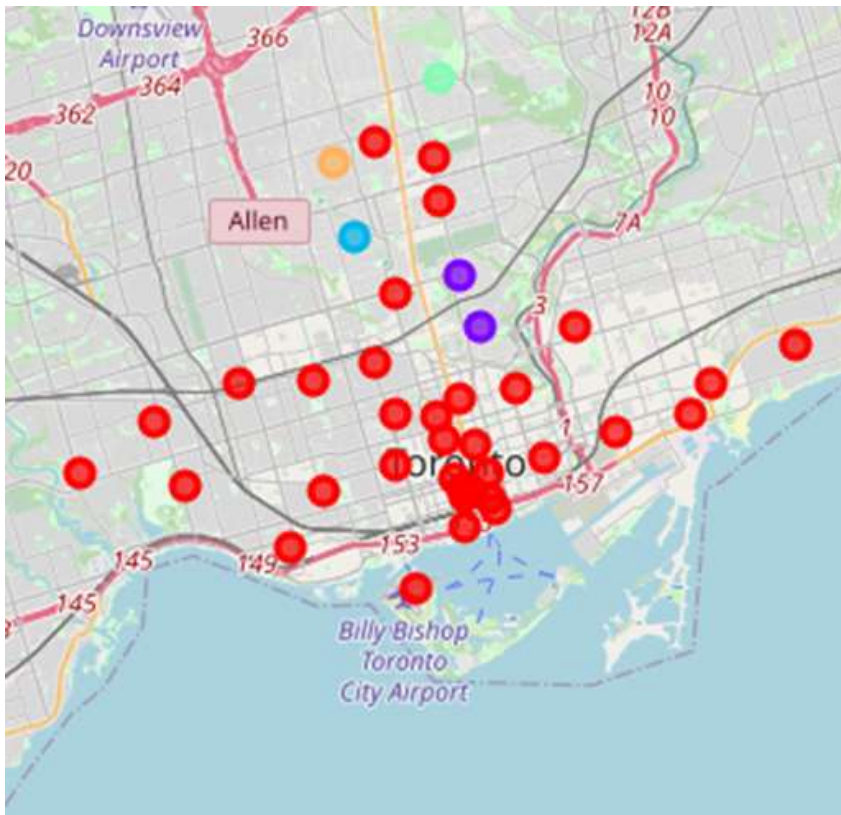
```
Total number of restaurants: 348
Total number of Asian restaurants: 83
Total number of Vietnamese restaurants: 6
```

There are in total 348 venues classified as restaurants, in which, 83 (24%) are Asian restaurants, and 6 (1.7%) are Vietnamese restaurants. There are 35 neighborhoods that do not currently have any Vietnamese restaurants
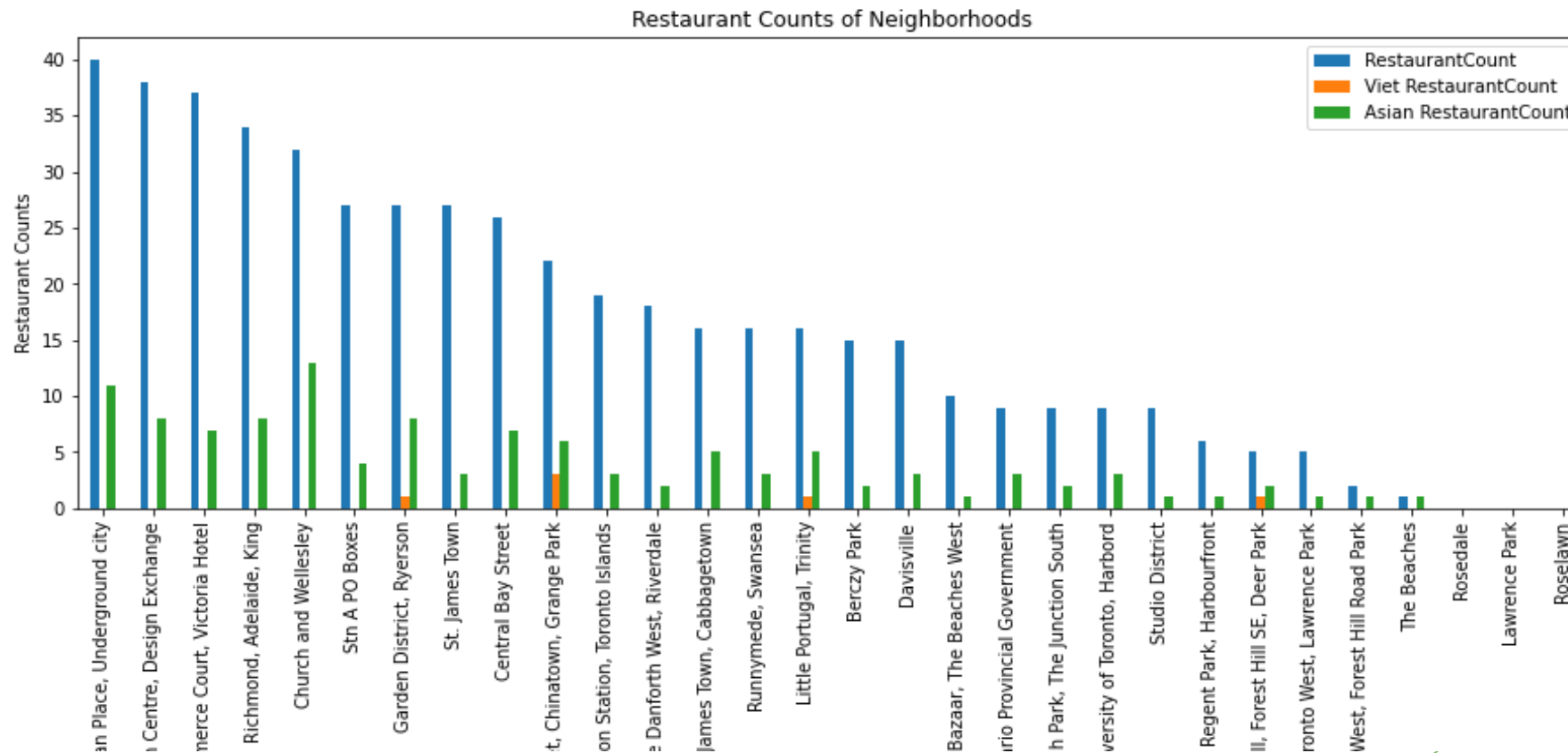
In-depth Analysis – Results (1)

▶ **Clustering the neighborhoods in Toronto by venues using k-means clustering**

▶ Most of the venues are clustered in one single big cluster (Red - called Cluster 1)

▶ Cluster 1-Red: contains restaurants, cafes and eateries.

▶ The cluster shows that restaurants, cafes and bar are distributed evenly and widespread across all the neighborhoods of Toronto

▶ Too vague to make a decision, we need a more zoomed in result and more specific factor for the best decision

In-depth Analysis – Results (2)

▶ **Segmentation and visualization of the restaurants distribution in each neighborhood by cuisines**

Discussion

▶ The clustering of venues by neighborhood shows that restaurants and cafes are evenly spread throughout all the neighborhoods of Toronto, except The Beaches, Roselawn and the Parks.

▶ Not enough to just use this clustering to conclude which neighborhood is the most potential to open a new Vietnamese restaurant.

▶ It is necessary to segregate and visualize the distribution of all restaurants in each neighborhood by cuisine

▶ Create a clearer picture on where the indirect competitors (other Asian restaurants) and the direct competitors (other Vietnamese restaurants) are situated

Conclusion

▶ Choosing neighborhoods with a high number of restaurants are potential to have high diner traffic .

▶ If the said neighborhoods do not have many existing Asian/Vietnamese restaurants, the possibility for a Vietnamese restaurant to be chosen would be higher

Restaurant Counts of Neighborhoods

▶ Stn A PO Boxes; St. James Town; Commerce Court, Victoria Hotel