Capstone Project Report: Franchise Expansion
By : Kali Benavides

1. Introduction

1.1 Background

Chatime is a bubble tea restaurant with a very profitable location in the Toronto neighborhood called Bay Street Corridor. The owners of Chatime would like to expand their franchise into the United States. The owners have selected Miami, FL as the city they would like to expand to. The profitability of their first location is due in large part to the consumer traffic that occurs in the Bay Street Corridor neighborhood. Therefore, they would like to identify a neighborhood in Miami, FL that is similar to the Bay Street Corridor neighborhood in Toronto.

1.2 Problem Data

Data that would be helpful to finding a similar neighborhood would be data on what type of venue categories occur most frequently, population density of the neighborhood and geographical location with respect to landmarks. This project aims to predict which neighborhood in Miami is most similar to the Bay Street Corridor neighborhood and thereby select a new location for a Chatime franchise.

1.3 Interest

The business owners of Chatime as well as other developers within Toronto or Miami neighborhoods.

2. Data acquisition and cleaning

2.1 Data sources

The Toronto neighborhood data listing the neighborhood names, and corresponding GPS data is sourced from https://open.toronto.ca/dataset/neighbourhoods/. The Toronto neighborhood population density data is sourced from https://open.toronto.ca/dataset/neighbourhood-profiles/.

The Miami neighborhood names, GPS coordinates and population density are sourced from url='https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Miami'

The data on venues located in each neighborhood is sourced from Foursquare API by querying the coordinates for each neighborhood location and yielding a GET request to get a list of venues within 500 meters of each neighborhood center.

This data will be used to determine which venue categories occur most frequently in the Bay Street Corridor neighborhood. The Toronto neighborhood will then be clustered with all the Miami neighborhoods to determine which neighborhood has the most similar combination of venue categories. The clustered neighborhoods will then be sorted using population density. Finally the neighborhoods will be mapped to compare geographical landmarks near the neighborhood to determine if that increases the overall similarity of the neighborhoods. The Miami neighborhood falling into the same cluster as the Bay Street Corridor neighborhood and with a similar population density and geographical landmarks will be selected as the location for the next franchise.

2.2 Data cleaning

Data downloaded or scraped from multiple sources were combined into two tables one listing all of the Miami neighborhood data and one listing the neighborhood data for the Toronto neighborhood of Bay Street Corridor. The neighborhood data contained in these initial tables lists all of the names, GPS coordinates, and population density for each neighborhood. The data set for the Miami neighborhoods was scraped from a Wikipedia

page and so had to be cleaned up. The Latitude and Longitude data were split into two separate columns and converted to data type float so that they could be in the appropriate form for mapping and Foursquare API searches. The population density data was also converted to data type float so that it could be used to sort the data rows at the end of the analysis. The column name for Population density was also changed to match the nomenclature used in the Bay Street Corridor data. Several columns were dropped from the Miami neighborhood data because they were not applicable to this analysis. Finally the country and city data was added as two columns to both the Miami neighborhood data table and the Toronto (Bay Street Corridor) data table.

3. Methodology

Exploratory data analysis was performed on the Miami neighborhood data by first listing out the unique neighborhood names to understand how many distinct neighborhoods are located in the City of Miami. Next, the city of Miami was mapped with markers for each neighborhood location. The Bay Street Corridor neighborhood location was then marked on a map of the CIty of Toronto.

To begin to gather the venue data for each neighborhood I used the GPS coordinates for each neighborhood to do a GET request to get a list of venues within 500 meters of each neighborhood center. The venue category data was then compiled for each neighborhood. And the neighborhoods were grouped to yield the mean of the frequency of occurrence of each category. A table was created with all of the neighborhoods and the top ten venue categories that occur within each.

This venue category data can then be used to cluster the neighborhoods into groups. I used K-means partitioning clustering to divide the data into 4 different clusters based on their top ten most common venues.

Once the clusters are created, I then added the population density data to sort each cluster from greatest population density to least.

This will yield a cluster containing the Bay Street Corridor neighborhood in Toronto. I will then select the Miami neighborhood in the same cluster with the most similar population density. The venue information for the Miami neighborhoods will also be used to identify any existing bubble tea restaurants which would compete with Chatime. I will also compare the mapping of the neighborhood locations to look at similar geographical landmark data. Such as vicinity to the coast line, major highway infrastructure or other advantageous traits that are similar to those found in the Bay Street Corridor neighborhood.

4. Results

The K-Means clustering analysis yielded five different clusters. These clusters are summarized below with observations about their classification.

- Cluster 1= Lower Economic neighborhood (Characterized by the frequent occurrence of fast food type restaurants and flea markets
- Cluster 2= Restaurant/Social Activities neighborhoods (Characterized by the frequent occurrence of restaurants and other gathering locations) ** Of note is that the Toronto neighborhood of Bay Street Corridor falls into this cluster. So we will analyze this cluster further.*
- Cluster 3=Medium Economic Neighborhood (Characterized by the frequent occurrence of shopping centers and restaurants, but still a high occurrence of fast food/flea market type establishments)

- Cluster 4= Outdoor/Waterside (Characterized by the frequent occurrence of boat/ferry locations, parks, food trucks)
- Cluster 5=Latin American neighborhoods (Characterized by the frequent occurrence of latin american restaurants)

The Bay Street corridor neighborhood was placed in Cluster #2 along with sixteen other Miami neighborhoods. The neighborhoods in this cluster were sorted by descending population density. The Bay Street corridor has a population density of 14,097 people per square kilometer. The Miami neighborhood within this cluster with the closest population density is Brickell with a density of 14,541people per square kilometer. The next closest population density of a Miami neighborhood is Downtown Miami with a density of only 10,613 people per square kilometer. The venue data for the Brickell neighborhood was also reviewed for any existing bubble tea restaurants. There are currently no bubble tea restaurants in the Brickell neighborhood.

*Table of Top Three Neighborhoods in terms of Population Density in Cluster 2

Population density per square kilometre	Neighborhood	City	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
14541.0	Brickell	Miami	1	Hotel	Italian Restaurant	Japanese Restaurant	Argentinian Restaurant	Lounge	Seafood Restaurant	Scenic Lookout	Café	Spanish Restaurant	Pharmacy
14097.0	Bay Street Corridor	Toronto	1	Coffee Shop	Italian Restaurant	Tea Room	Sushi Restaurant	Japanese Restaurant	Ramen Restaurant	Gastropub	New American Restaurant	Dessert Shop	Department Store
10613.0	Downtown	Miami	1	Italian Restaurant	Brazilian Restaurant	Cocktail Bar	Coffee Shop	Breakfast Spot	Café	Seafood Restaurant	Scenic Lookout	Sandwich Place	Salon / Barbershop

*Figure showing the population density of the neighborhoods in Cluster 2

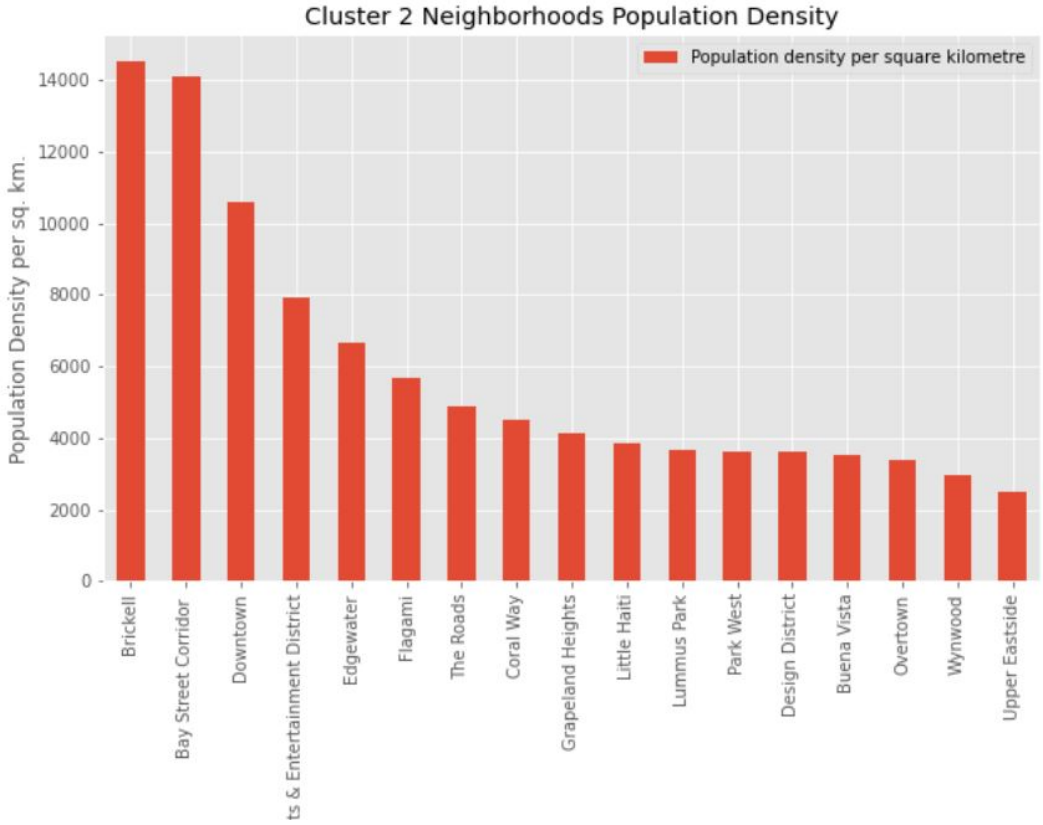
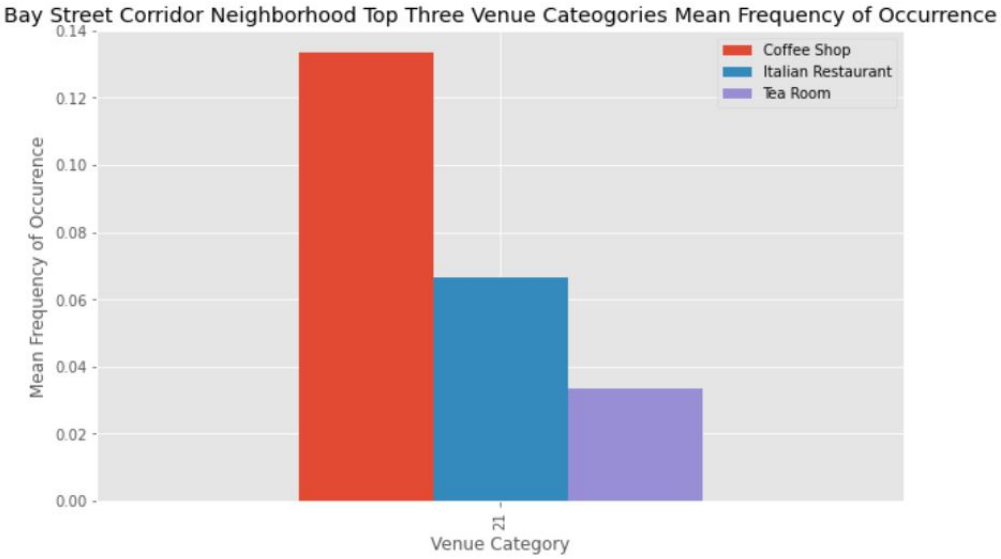
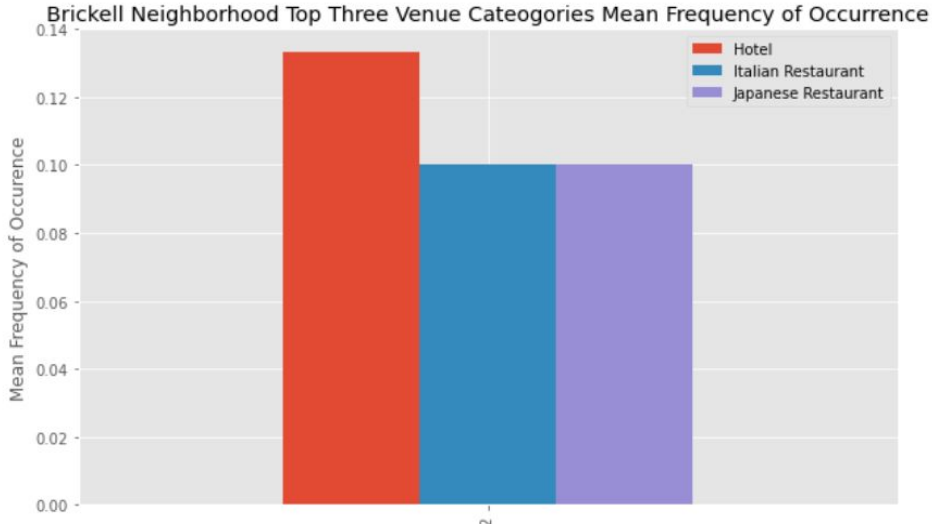
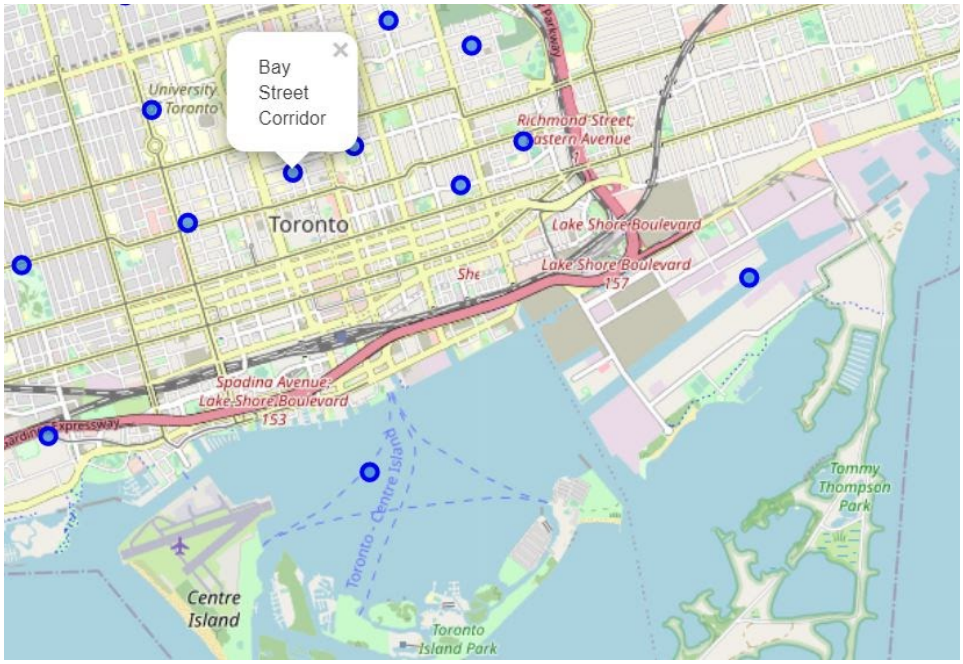


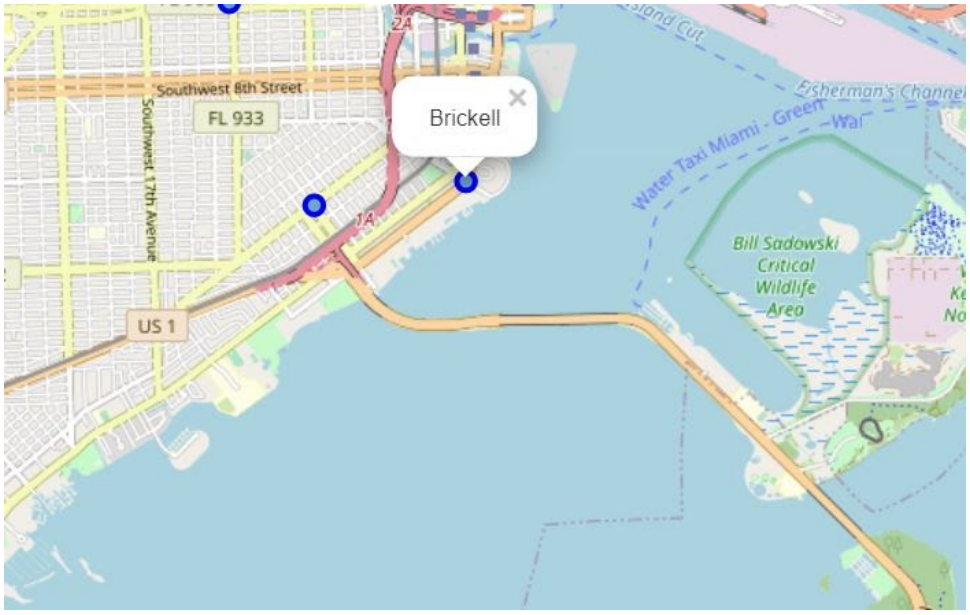
*Figure showing the comparison of the Top three most common venue categories in Brickell and Bay Street Corridor.





Last was a holistic comparison of the location looking at geographical landmark data. The similarities between the Bay Street Corridor neighborhood and the Brickell neighborhood include the following: Vicinity to the bay, close proximity to highway infrastructure, and both contain small areas of green space. In Brickell, Simpson park and in the Bay Street Corridor, Queen's Park.





5.0 Discussion

Based on these findings, the Brickell neighborhood in Miami is most similar to the Bay Street Corridor neighborhood. Based on a combination of the most common venue categories, population density, and geographical landmarks. This indicates that the consumer foot traffic will be similar and predicts similar profitability to the original Chatime location. The Brickell neighborhood does not currently contain any other bubble tea restaurants so there will not be competition for Chatime.Therefore the business owners of the Chatime restaurant should open their next Miami location in the Brickell neighborhood.

6.0 Conclusion

In this study, I analyzed venue and population density data for neighborhoods in Miami to determine which neighborhood was most similar to the Toronto neighborhood of Bay Street Corridor. I built K-Means cluster models to group neighborhoods based on their most common venue categories. These models can be very useful in helping the Chatime business owners to identify which Miami neighborhood they should select for their next bubble tea restaurant location or other business endeavors.