RECCOMENDED LOCATIONS TO OPEN A NEW BEER BAR FOR STANLEY PARK BREWING CO

IBM Capstone Project

ABSTRACT

This report aims to recommend locations in Toronto area to open a new Beer Bar for Stanley Park Brewing Co meeting the following conditions: least amount of beer bars and close to popular attractions. From the data gathered from Wikipedia, Foursquare and Open Data Toronto, it was fond the number of beer bars increase towards the shore of the lake and around Toronto University. To find recommended locations, the data was filtered to display neighbourhoods that have popular attractions and beer bars. Popular attractions in neighbourhoods was chosen as a measure of high traffic of people in those areas. K-means was used to group the data based on the similarity of the data points. One of the recommended locations was Central Bay Street, with no beer bars and 4 popular attractions. All analysis was made in Jupyter Notebook.

Fernando Silva

Data collected on Jun 01, 2021

Table of Contents

1. Introduction

The food and beverage industry can be attractive to people due to its potential profitability and relatively lower investment compared to other industries. However, due to the current pandemic, challenges often faced by business owners of this industry have been amplified. Many factors are to be considered when open new businesses and any advantage can be imperative for their success.

Competition is often a factor that influences business owners' decisions. This report aims to recommend locations in Toronto area to open a new Beer Bar for Stanley Park Brewing Co meeting the following conditions: least amount of beer bars and close to popular attractions. Reduced competition combined with proximity to popular locations will contribute to the business' success.

The model in this report can be used for other venues in neighbourhoods in the city of Toronto, so that other business owners may know which neighbourhoods are less competitive in their business area.

1.1 Data Acquisition

Data was acquired in June 01, 2021. This report used datasets from 3 sources: Wikipedia, Foursquare and Open Data Toronto. Below is a summary of the information each source provided:

- Wikipedia:
  - Neighbourhood names and addresses in Toronto
- Foursquare:
  - Geographical location of neighbourhoods in Toronto
  - Venues in each neighbourhood within 1000 m of the neighbourhood geographical coordinate
- Open Data Toronto
  - Popular locations in Toronto. Popularity is measured by the field “Attraction Level”
  - Geospatial location of neighbourhoods

Wikipedia data was acquired using request libraries in python to pull the data directly from the website. Personal credentials were used to gather data from Foursquare by sending requests to its database. Open Data Toronto had data available in several formats such as csv files, which were downloaded and later uploaded to Jupyter Notebook.

2. Methodology

2.1 Relating data from data sources

The use of common parameters such as postal code and neighbourhood name allowed to relate data from the sources used in the report to one another. Because the area of interest is Toronto and the target venue is beer bar, Wikipedia and Foursquare data were primarily filtered based on those conditions. The resulting data showed neighbourhoods in Toronto that have beer bars.

The second condition, popular attractions in Toronto, was later applied. The resulting data showed locations with beer bars and popular attractions.

2.2 Location Criteria

Data about popular locations provided by Open Data Toronto has a field called “Attraction Level”. This field is a very simple measure to indicate how popular the location is. It is a categorical variable that ranges from 1 to 5, in which 1 is most popular.

Assuming that popular locations attract many people, ideal locations will be in neighbourhoods that have popular attractions and the least number of beer bars. Number of popular attractions in neighbourhoods was chosen as a measure of high traffic of people in those areas. A lower number of beer bars will decrease the competition for those potential customers.

2.3 Machine Learning Method

K-means was the chosen machine learning technique for this report. It is an unsupervised clustering method that groups the data into clusters based on how similar the data points are to one another. K-means clustering is a very popular clustering algorithm and provides an idea of the structure of the dataset. The goal of k-means is to group data points into distinct non-overlapping subgroups.

The number of clusters was determined by the elbow method, using the Within-Cluster-Sum-of-Squares (WCSS) as a criterion. WCSS stands for the sum of the squares of distances of the data points in each and every cluster from its centroid.

3. Results

3.1 Beer Bars in Toronto Neighbourhoods

Table 1 below shows the number of beer bars per neighbourhood in Toronto. Only a small set of the table is displayed in this report.

Table 1: Number of beer bars per neighbourhood in Toronto

Neighbourhood	***Number of beer bars***
Berczy Park	3
Brockton, Parkdale Village, Exhibition Place	1
Business reply mail Processing Centre	0
CN Tower, King and Spadina, Railway Lands	0
Central Bay Street	0
Christie	0
Church and Wellesley	0
Commerce Court, Victoria Hotel	3
Davisville	0

Neighbourhood	*Number of beer bars*
Davisville North	0
Dufferin, Dovercourt Village	0
First Canadian Place, Underground city	1

Distribution of beer bars across Toronto seams to be sparse. Most common number of beer bars in neighbourhoods is zero (0). We can observe this pattern with a simple histogram. Figure 1 shows the distribution of beer bars in Toronto. As the count of beer bars increases, the number of neighbourhoods that have beer bars decreases.

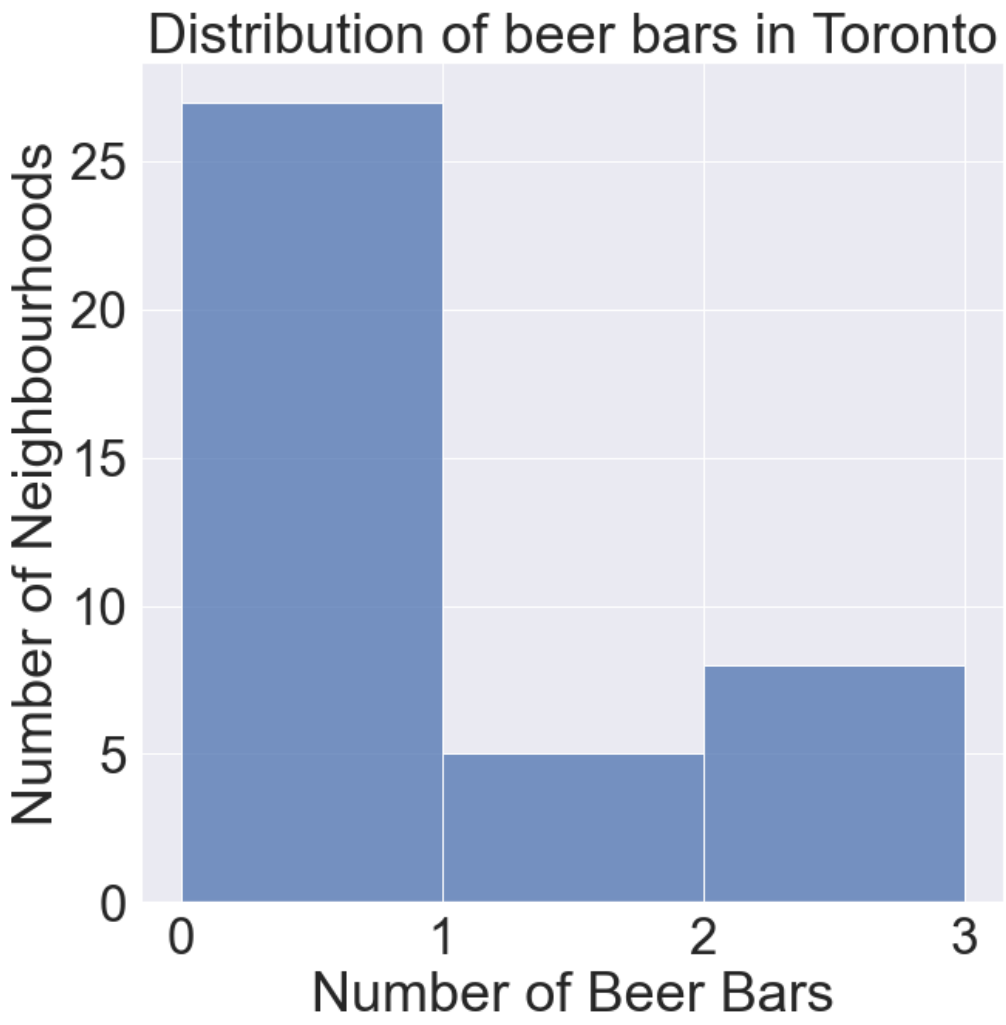

Figure 1: Distribution of beer bars in Toronto neighbourhoods

Most neighbourhoods in Toronto, approximately 26, do not have a beer bar. 5 neighbourhoods have 1 to 2 beer bars and 6 neighbourhoods have 2 to 3 beer bars. Figure 2 shows the number of beer bars only in neighbourhoods that have beer bars.

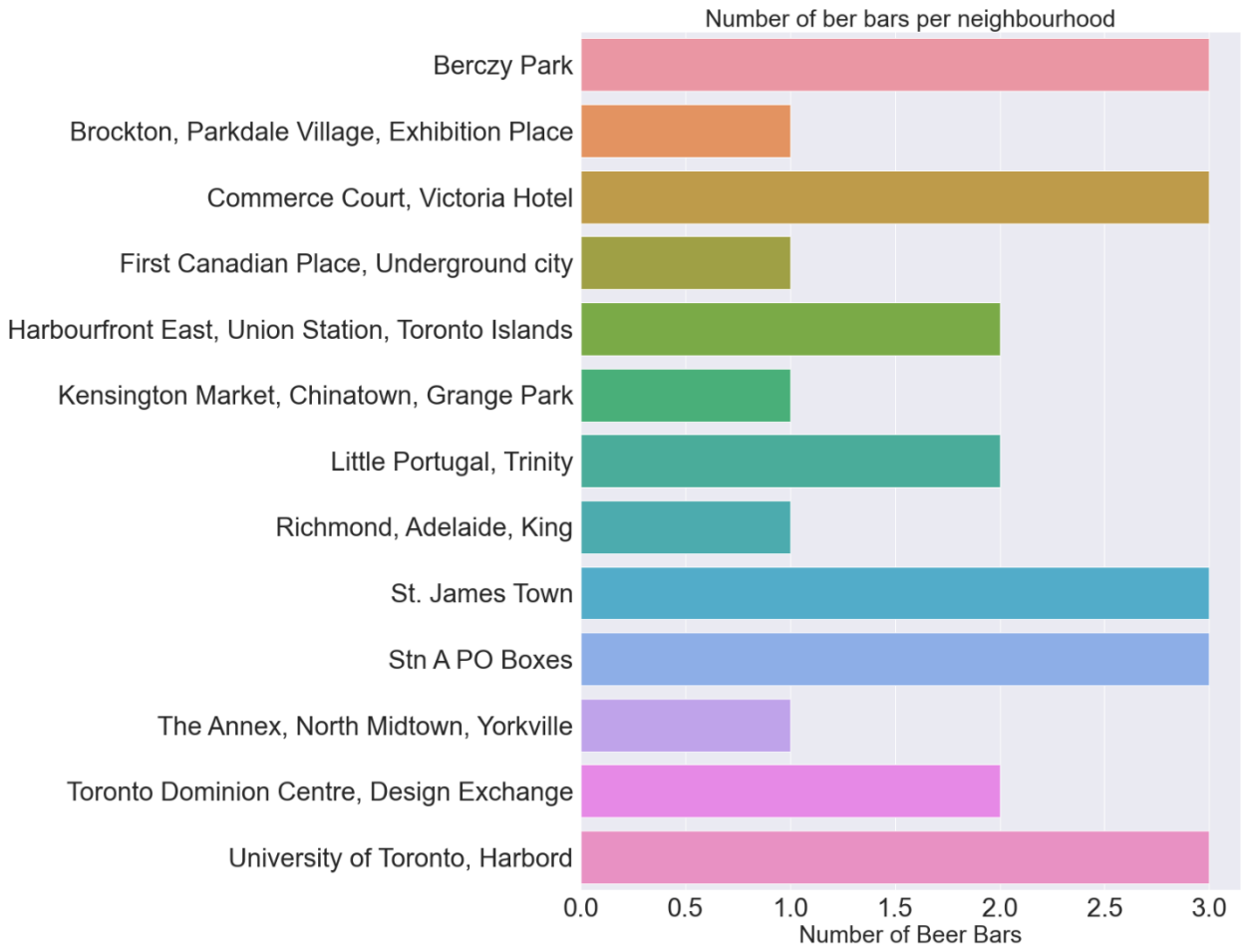

Figure 2: Number of beer bars in neighbourhoods that have beer bars

The largest count of beer bars in the data set was 3 in *University of Toronto, Harbord*, possibly due to higher number of young people in the neighbourhood that study in the university, *Berczy* Park, *Commerce Court, Victoria Hotel*, *St. James Town*, and in *Stn A PO Boxes.* These neighbourhoods are the most competitive for beer bar owners.

Figure 3 shows neighbourhoods with beer bars in a map. According to the color scale, orange represents the are with highest largest number of beer bars, followed by pink, dark green and dark blue.

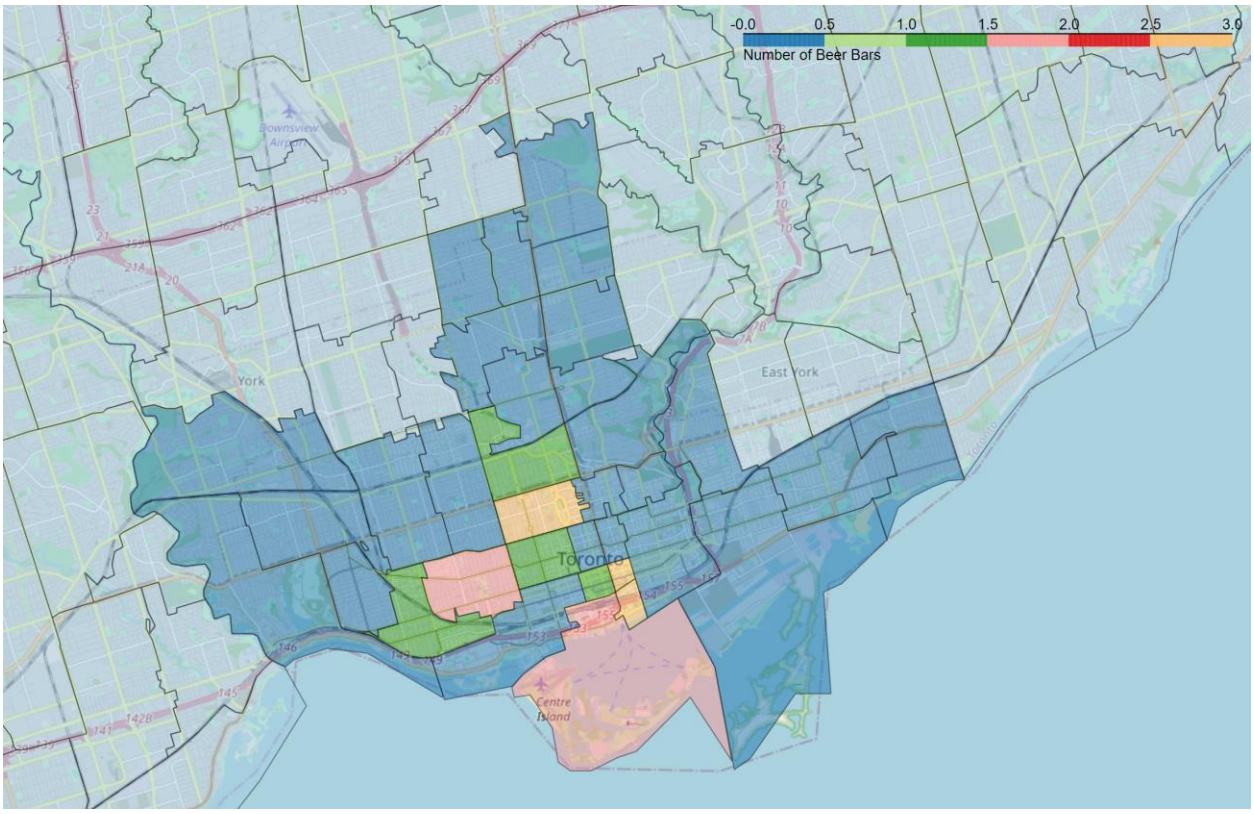

Figure 3: Map of Toronto with neighbourhoods colored according to the number of beer bars

Number of beer bars increase towards the shore of the lake and around Toronto University, with some exceptions. The higher number of beer bars in *Harbourfront East, Union Station, Toronto Islands* may be explained by the number of popular attractions in the region.

3.2 Popular Attractions in Toronto Neighbourhoods

A similar analysis can be done for popular attractions in Toronto. Figure 4 shows neighbourhoods with beer bars in a map. According to the color scale, orange represents the are with highest largest number of beer bars, followed by pink, dark green and dark blue.

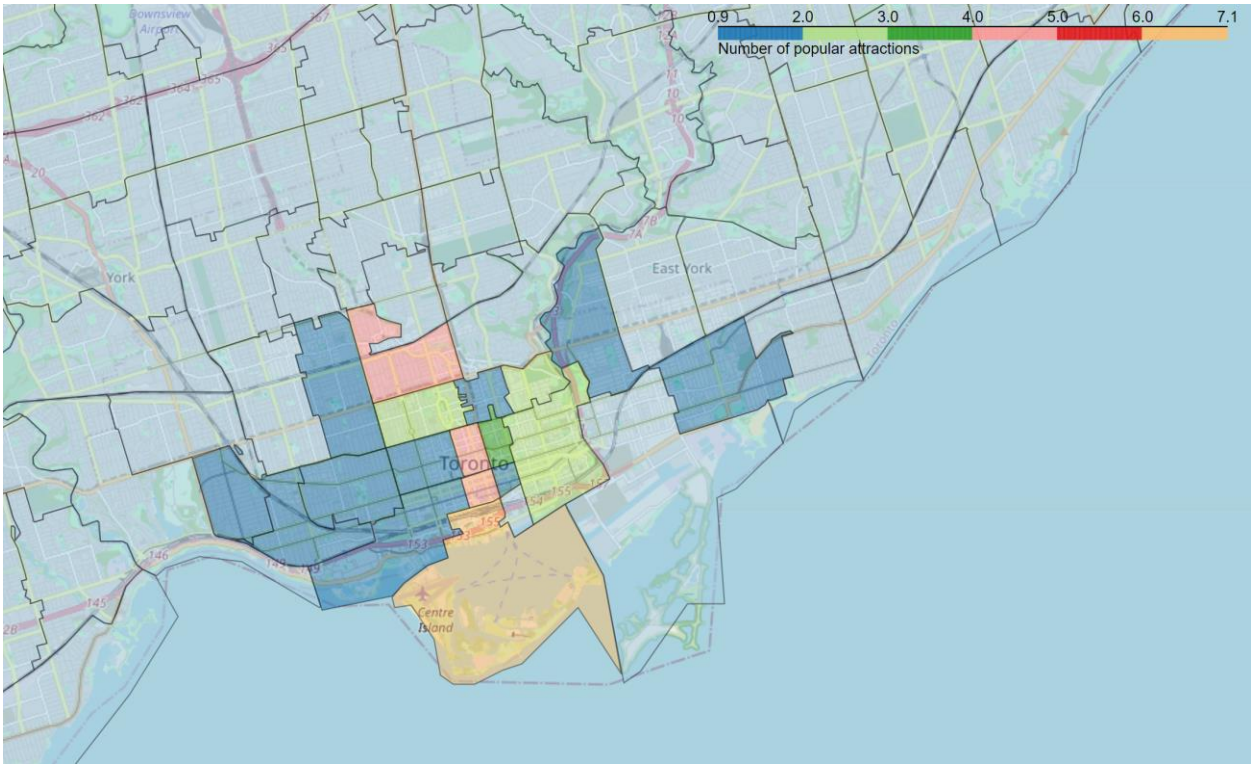

Figure 4: Map of Toronto with neighbourhoods colored according to the number of popular attractions

Number of popular attractions increase towards the shore of the lake. *Harbourfront East, Union Station, Toronto Islands* has the highest number of popular attractions, with 7 attractions in total. *Richmond, Adelaide, King*, located northwest of Toronto University comes in second, with 5 popular attractions.

3.4 K-Means results

Tables containing data about number of beer bars and number of popular attractions were combined to display the number of beer bars in neighbourhoods that have popular attractions. The unsupervised machine learning method grouped the data into clusters. The number of clusters was determined by the elbow method. Figure 5 displays the number of clusters related to Within-Cluster-Sum-of-Squares:

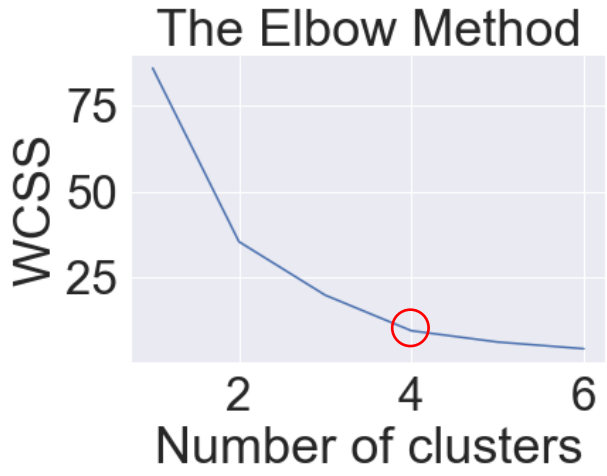

Figure 5: number of clusters related to Within-Cluster-Sum-of-Squares

According to the elbow method, the ideal number of clusters is determined where the rate of decrease of WCSS sharply shifts. Therefore, the ideal number of clusters is 4.

The data was grouped into 4 clusters according to the similarity among data points. Figure 6 displays the clusters generated by the K-means method on a map:

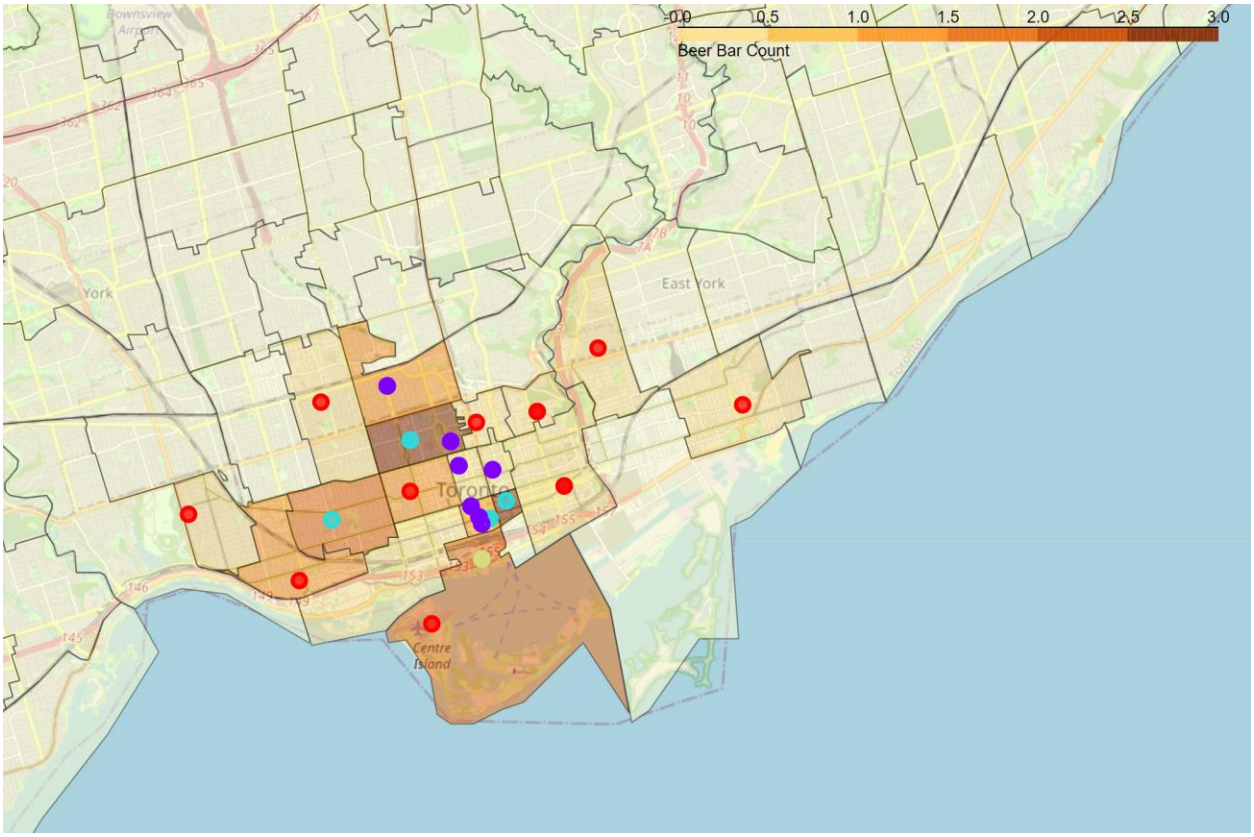

Figure 6: K-means clusters with neighbourhoods colored according to the number of beer bars

According to the color scale, dark brown represents the are with highest largest number of beer bars, followed by light brown, orange and light yellow. Markers coloured in purple represent locations that have relatively high number of popular attractions and relatively low number of beer bars compared to those coloured in red, green, and yellow. Tables 2, 3, 4 and 5 display how the data was grouped by the chosen machine learning method:

Tables 2: K-means grouped data with label 0 (red markers)

Neighborhood	***Cluster***	***Beer Bar Count***	***Popular Attraction count***
Regent Park, Harbourfront	0	0	2
Church and Wellesley	0	0	1

Neighborhood	Cluster	Beer Bar Count	Popular Attraction count
Kensington Market, Chinatown, Grange Park	0	1	1
Christie	0	0	1
The Danforth West, Riverdale	0	0	1
Brockton, Parkdale Village, Exhibition Place	0	1	1
India Bazaar, The Beaches West	0	0	1
Parkdale, Roncesvalles	0	0	1
CN Tower, King and Spadina, Railway Lands, Har...	0	0	1
St. James Town, Cabbagetown	0	0	2

Tables 3: K-means grouped data with label 1 (purple markers)

Neighborhood	Cluster	Beer Bar Count	Popular Attraction count
Queen's Park, Ontario Provincial Government	1	0	4
Garden District, Ryerson	1	0	3
Central Bay Street	1	0	4
Richmond, Adelaide, King	1	1	5
Toronto Dominion Centre, Design Exchange	1	1	4
First Canadian Place, Underground city	1	1	4
The Annex, North Midtown, Yorkville	1	1	4

Tables 4: K-means grouped data with label 2 (light blue markers)

Neighborhood	Cluster	Beer Bar Count	Popular Attraction count
University of Toronto, Harbord	2	3	2
St. James Town	2	3	1
Commerce Court, Victoria Hotel	2	3	2

Neighborhood	Cluster	Beer Bar Count	Popular Attraction count
Little Portugal, Trinity	*2*	*2*	*1*

Tables 5: K-means grouped data with label 3 (yellow marker)

Neighborhood	Cluster	Beer Bar Count	Popular Attraction count
Harbourfront East, Union Station, Toronto Islands	*3*	*2*	*7*

5. Discussion

5.1 Location of beer bars and popular attractions in Toronto

Requirements for recommending locations were defined by number of beer bars and presence of popular attractions in a location. Number of popular attractions was the measure chosen to infer high traffic of people in the area. According to Figure 3, which shows a map of Toronto with neighbourhoods colored according to the number of beer bars, the distribution of beer bars over the neighbourhoods is located around University of Toronto and towards the shore of the lake. This pattern can be explained by the distribution of popular attractions in the area.

Comparing Figures 3 and 4, we can see all neighbourhoods that popular attractions or neighbourhoods that are close to other neighbourhoods that have popular attractions, also have beer bars. *Harbourfront East, Union Station, Toronto Islands* stands out with the highest number of popular attractions (7 in total) and a relatively low number of beer bars (2 in total) compared to other locations in Toronto.

Although *University of Toronto* neighbourhood has 2 popular attractions only, the popularity of the university and the age group (young people) of people who study and visit the university explains the higher relatively competitivity for beer bars. Neighbourhoods around *University of Toronto* neighbourhood such as Commerce Court, *Victoria Hotel and Berczy Park* also have 3 beer bars. This can be explained by the proximity to University of Toronto, where many young people may live.

5.2 K-means clustering method

K-means method grouped the data based on similarity among data points. The method was able to find neighbourhoods with relatively low number of beer bars and high number of popular attractions. Locations represented by purple markers in Figure 6, labeled by the machine learning model as cluster #1, met the requirements of reduced competition combined with proximity to popular locations, and,

therefore, are the recommended locations to open a new Beer Bar for Stanley Park Brewing Co. Table 3 shows the locations labelled as cluster #1. Recommend locations are as follows:

- Queen's Park, Ontario Provincial Government
- Garden District, Ryerson
- Central Bay Street
- Richmond, Adelaide, King
- Toronto Dominion Centre, Design Exchange
- First Canadian Place, Underground city
- The Annex, North Midtown, Yorkville

Harbourfront East, Union Station, Toronto Islands, represented by yellow marker (Cluster #3) may be recommended due to the high number of popular attractions, 7 in total, despite having the second highest number of beer bars in the region, 2 in total.

4. Conclusion

Distribution of beer bars across Toronto is sparse. Beer bars are concentrated towards the shore of the lake and around University of Toronto, where the number of younger people is higher. *University of Toronto, Harbord* has 3 beer bars in total according to the Foursquare data, the neighbourhood with the highest number of beer bars in Toronto.

Number of popular attractions in neighbourhoods was chosen as a measure of high traffic of people in those areas. The data was arranged in a way to provide the number of beer bars in neighbourhoods that contain at least 1 popular attraction.

K-means was chosen as the machine learning model to group the data into clusters. The model was able to identify neighbourhoods with relatively low number of beer bars and relatively high number of popular attractions. Clusters labeled as #1, met the requirements of reduced competition combined with proximity to popular locations, and, therefore, are the recommended locations to open a new Beer Bar for Stanley Park Brewing Co. Recommended locations are:

- Queen's Park, Ontario Provincial Government
- Garden District, Ryerson
- Central Bay Street
- Richmond, Adelaide, King
- Toronto Dominion Centre, Design Exchange
- First Canadian Place, Underground city
- The Annex, North Midtown, Yorkville

Cluster #3, which represents neighbourhood *Harbourfront East, Union Station, Toronto Islands*, is also recommended due to the high number of popular attractions, 7 in total, despite having the second highest number of beer bars in the region, 2 in total.

This analysis can be repeated for other individual venues or multiple.

5. Future Directions

This report took into consideration 2 conditions: competition and proximity to popular attractions. Real Estate data such as price and location availability may bring significant improvements to this analysis.