A Tale of Two Cities: San Francisco and Toronto

By Abeselom Zerit

1. Introduction

1.1 Background

San Francisco and Toronto are both known to be vibrant, multicultural cities with a lot to offer for tourists and locals alike. Not every city can offer everything, but both are best known for their international cuisine and arts & cultural attractions. Food and culture can are often the most exciting aspects of a city, so they can definitely be deciding factors for a city's livability.

1.2 Problem

In this study, we will quantify how similar both cities are in the aforementioned terms. We will be looking at the types and numbers of venues and their location distribution. Please note that due to limited data, the quality and reviews of these venues will not be considered. Regardless, the results of the comparison will be useful to anyone who needs more information on deciding to possibly to move to either of these cities.

2. Data

For this project, we will need to collect the datasets below:
San Francisco neighborhood data with latitude and longitude coordinates
Toronto neighborhood data with latitude and longitude coordinates
Foursquare location data for venues

2.1 San Francisco Data

There was no neighborhood location data available, so zip codes were used instead. The following websites were used to web scrape or download the data. Both datasets were merged to create a neighborhood/location lookup.

For the neighborhood/zip code info:
http://www.healthysf.org/bdi/outcomes/zipmap.htm

For the zip code/location:
https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/

2.2 Toronto Data

Similarly, the Toronto data is merged between neighborhood/zip code and zip code/location lookups.

For the neighborhood/zip code info:
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M”

For the zip code/location:
https://cocl.us/Geospatial_data

2.3 Merged Data
Since we will be clustering neighborhoods from both cities, we will also need a combined dataset.

2.4 Foursquare Location Data
The Foursquare API will be used to get the most common venue categories nearby in each neighborhood in San Francisco and Toronto. This will then be used to cluster similar neighborhoods and compare venue metrics from both cities. The venue categories will also be bucketed by their parent categories to do a high level comparison between the two cities.

3. Methodology
Exploratory data analysis will initially be used to profile our Foursquare neighborhood/ venue breakdown.

Stacked bar charts will then be used to compare the venue parent category distribution in each city.

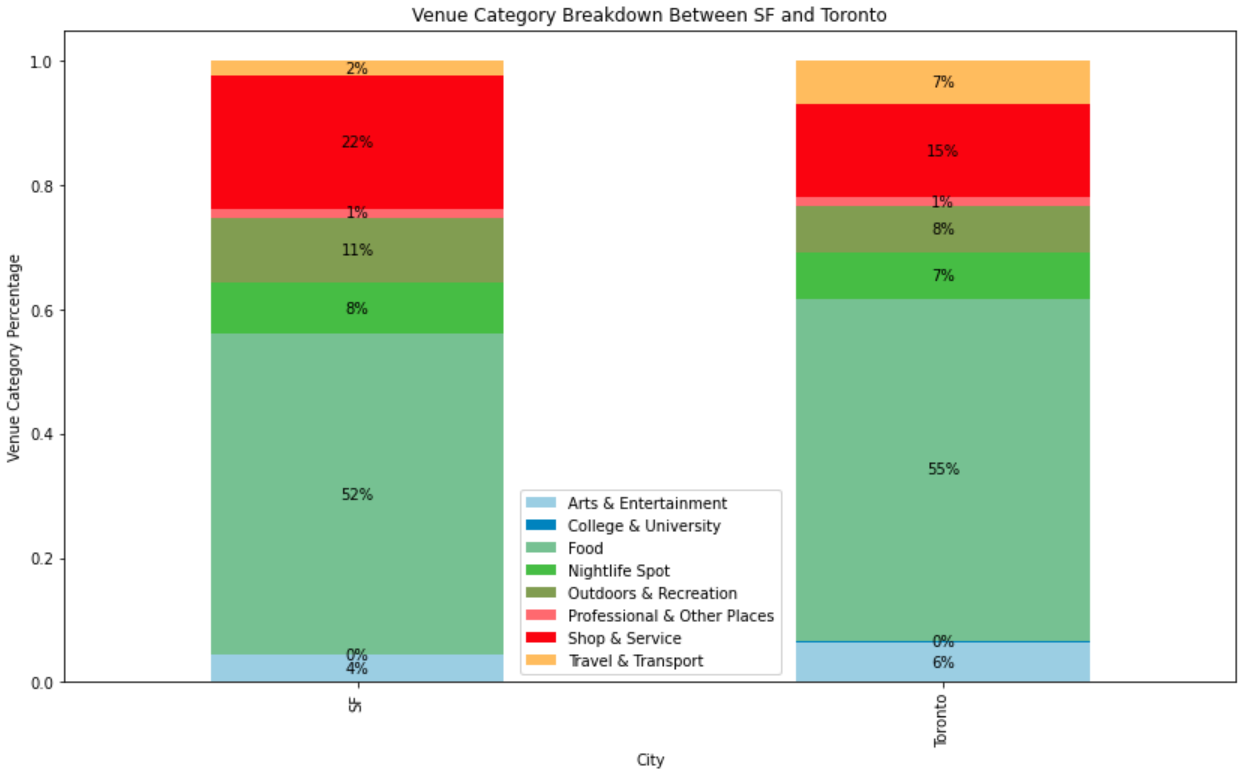
Since there are so many unknown variables, the the K-Means clustering algorithm is ideal for identifying similar neighborhoods from the combined city dataset.

Folium maps will then be used to display the cross-city venue clusters.

4. Analysis

4.1 Exploratory Data Analysis
There are 58 neighborhoods, 37 Toronto and 21 San Francisco neighborhoods, with a max of 100 nearby venues for each.

We will first examine venue category distributions.



4.2 Top Ten Neighborhood Venue Categories

Since there are so many venue categories, we will only examine the top ten for each neighborhood for the purposes of clustering

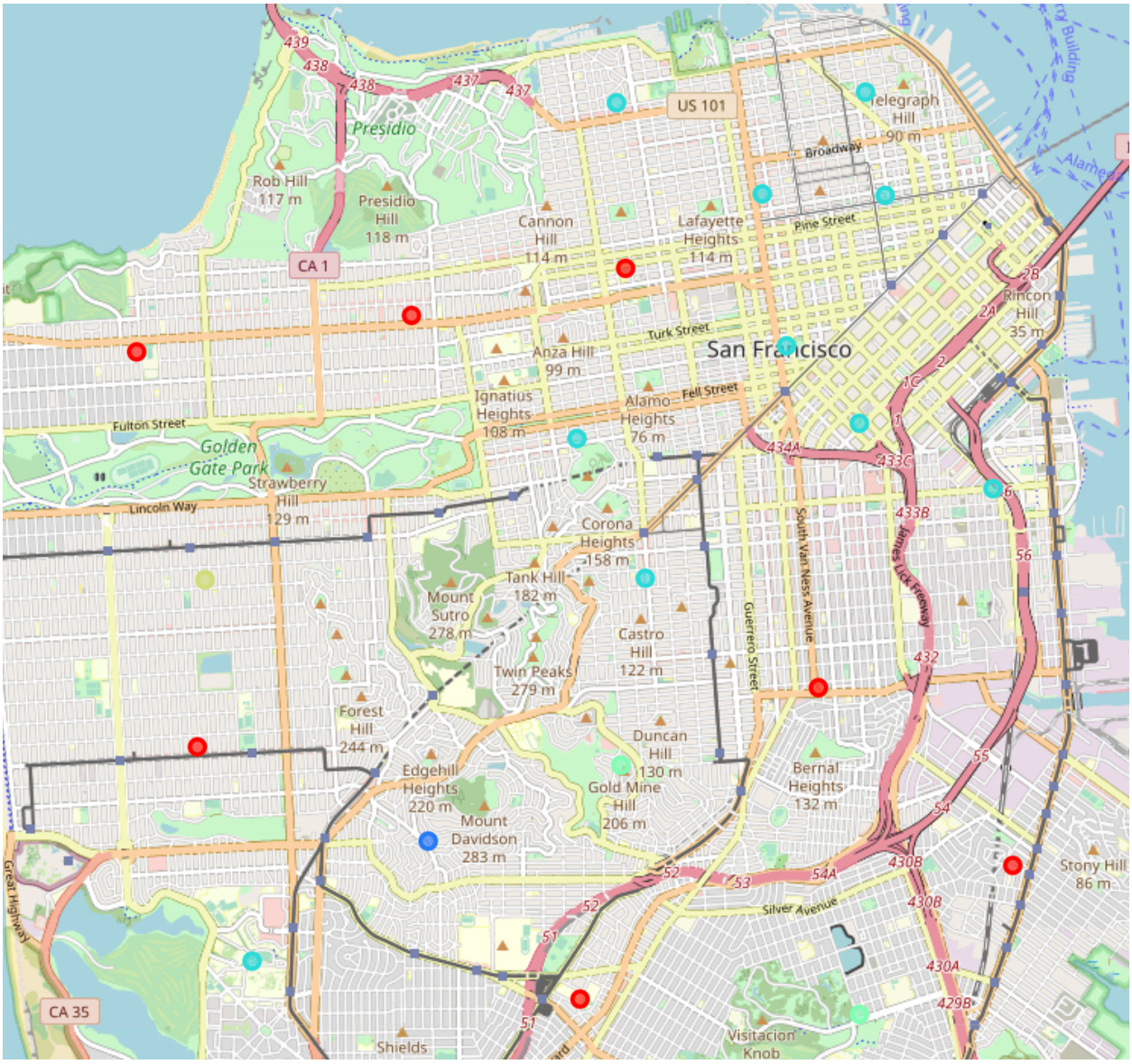
In this section, we will first use the one hot encoding to each categories into a separate column and calculate the frequency of occurrence of each category. Then we will slice the top 10 categories for clustering. Below is the top 10 categories of the first 5 neighborhoods.

City	Neighborhood	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
SF	Bayview-Hunters Point	Southern / Soul Food Restaurant	Coffee Shop	Mexican Restaurant	Bakery	Café	Taco Place	African Restaurant	Park	Gym	Pharmacy
SF	Castro/Noe Valley	Gay Bar	Thai Restaurant	Coffee Shop	Scenic Lookout	Yoga Studio	Indian Restaurant	Cosmetics Shop	Pizza Place	Playground	Plaza
SF	Chinatown	Coffee Shop	Bakery	Chinese Restaurant	Hotel	Men's Store	Cocktail Bar	Bubble Tea Shop	Dim Sum Restaurant	Church	Szechuan Restaurant
SF	Haight-Ashbury	Coffee Shop	Pizza Place	Park	Boutique	Thrift / Vintage Store	Bakery	Gastropub	Jazz Club	Bookstore	Gym
SF	Hayes Valley/ Tenderloin/North ...	Café	Hotel	Coffee Shop	Mexican Restaurant	Wine Bar	French Restaurant	Theater	Cocktail Bar	Performing Arts Venue	Sushi Restaurant

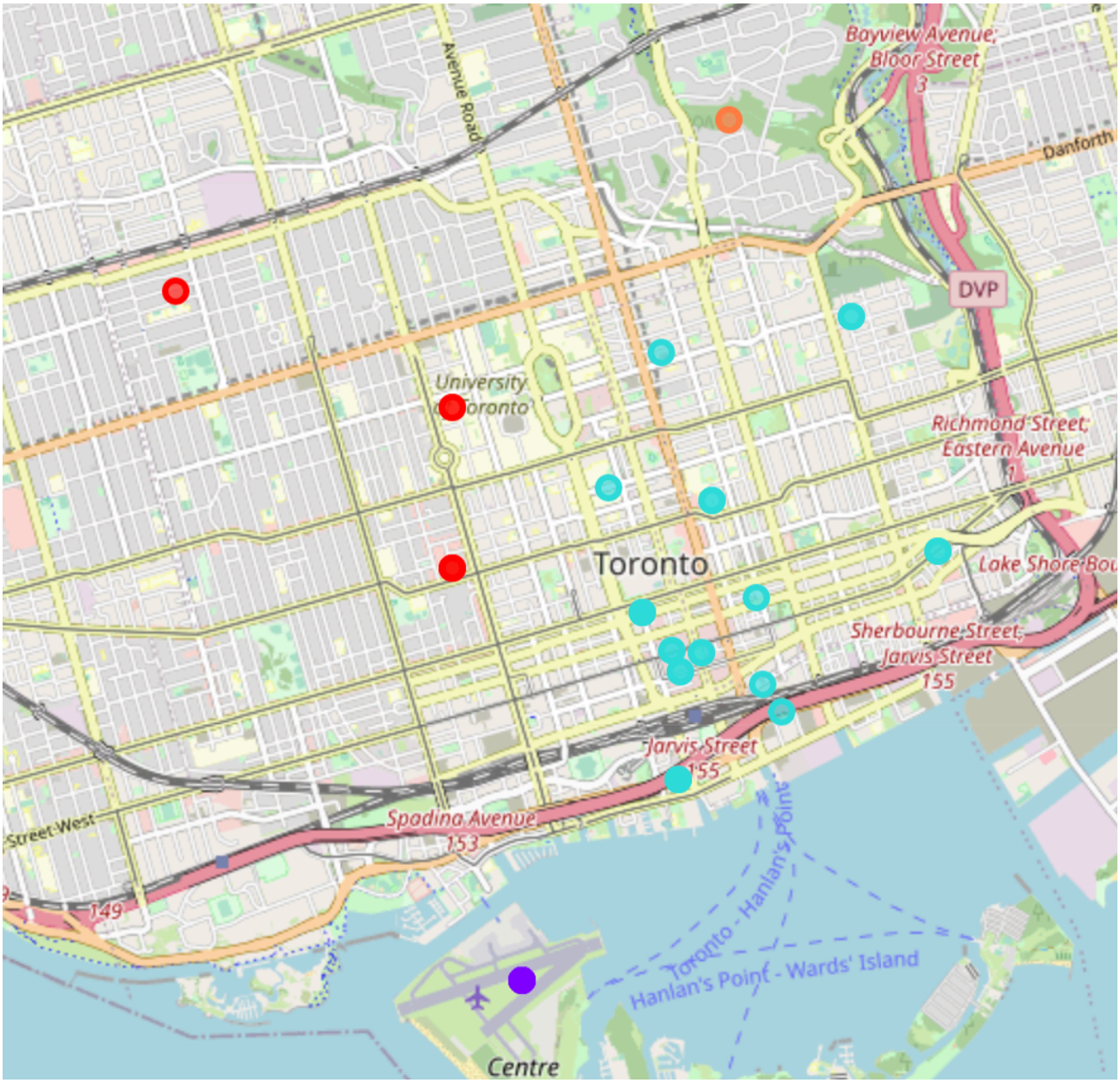
4.3 K-Means Clustering

Next we will use the K-Means clustering machine learning algorithm to cluster all the neighborhoods into 7 clusters and examine the clusters. Here are the clustering results mapped out, with the clusters across both cities share the same color.

San Francisco



Toronto


Bayview Avenue;
Bloor Street
Avenue Road
DVP
University
of Toronto
Richmond Street;
Eastern Avenue
Toronto
Lake Shore Bou
Sherbourne Street;
Jarvis Street
155
Jarvis Street
155
Spadina Avenue
153
149
Toronto - Hanlan's Point - Wards' Island
Centre

5. Results and Discussion

From the venue category breakdown, both cities are very similar is their distribution, with Toronto slightly edging out SF in food (55% vs 52%) and arts & entertainment (6% vs 4%).

From the maps above, the red colored clusters are associated with a higher density of international restaurants. In SF, they are mainly distributed well outside of downtown, where in Toronto, they are clustered near the University of Toronto.

Since the Foursquare data is heavily skewed toward restaurants, the cultural attractions data in the top ten breakdown was sparse. However, there was a noticeable theater presence in the light blue clusters, which were located near the downtown areas of both cities.

6. Conclusion

The purpose of this study was to compare San Francisco and Toronto on the basis on their international cuisine and arts & cultural attractions. Based on the findings, once can conclude that both cities equally have lots to offer, with Toronto offering slightly more in a comparatively smaller area of the city.