Comparing two cities Windsor and Detroit

Sogra Bilal Memon

April 15th 2020

1. Introduction:

1.1. Background

This project aims to compare the neighborhoods of two cities: The Canadian city Windsor and the American city Detroit. Windsor and Detroit are located right across the Canadian – US boarder separated by the Detroit River that runs in between them. These two cities have shared multiple partnerships and agreements such as the Detroit and Windsor Tunnel Corporation which is jointly owned by both the cities. The two cities are connected to each other through the Michigan railway tunnel, Ambassador bridge and the Detroit – Windsor tunnel. More than 40,000 people travel between these two cities per day.

1.2. Problem

As discussed earlier more than 40,000 people cross the US – Canada border between these two cities on a daily basis. This could be for several reasons including tourism, people living in one city while working in the other city or the availability of certain venues in one city that are absent in the other. In April 2013 the development of a new bridge The Gordie Howe International Bridge was approved between these two cities to reduce the traffic on the existing bridge. The Gordie Howe International Bridge is expected to be constructed by 2024 and is predicted to cost a total of $5.7billion. It has also been predicted that this bridge will earn $70.5 million in toll revenue. Understanding the differences between the neighborhoods and the available venues in each of these cities could help us understand why residence one city would commute to the other city.

1.3. Interest

Many stakeholders can benefit from understanding this data. The government can make important decisions related to providing important services such as public transport or required infrastructure based on its citizen's needs. The analysis of this data can also be helpful to new comers, tourists and business owners. New comers can compare neighborhoods to help them select a suitable location for residence that is close enough to their workplace and meets their other requirements. On the other hand, tourists may be interested in planning a trip that includes different types of venues from both cities. Business owners can understand the missing venues in their city to set up a new business venture that gives users the same facility while giving an added advantage of being closer to the user.

2. Data Acquisition and Cleaning

2.1. Data Source

For this project I will be extracting the data from two Wikipedia pages using the web scraping python library Beautiful Soup. I selected two Wikipedia pages that have the names of all the neighborhoods in each of the two cities. You can find the links to these pages here: Windsor and Detroit. From these two pages all the neighborhoods and the boroughs they belong to can be extracted. The geolocation for each of these neighborhoods can be found using the Nominatim python module that converts addresses into latitude and longitude. Using the latitude and longitude of each neighborhood we can call the Foursquare API to get the nearby venues. We can manipulate the data received from the Foursquare API to find the most popular venues in each neighborhood. The neighborhoods can then be cluster using the top 5 venues for each neighborhood as the features. The k-means clustering algorithm will group similar neighborhoods in the same cluster and the dissimilar neighborhoods in a different cluster. The results can then be plotted on a folium map using the geolocation of each neighborhood to position the marker correctly. The markers can then be given different colors depending on which cluster each neighborhood belongs to.

2.2. Data cleaning

This data has been scraped from two web pages which means neighborhood and borough names need to be separated from the rest of the information on the page such as images and description of places. The required data was extracted using the Beautiful Soup python library. In the Wikipedia page for Windsor I was able to extract the Borough names and the neighborhood names from the contents side bar. This added a few extra unnecessary rows such as see also, references etc. which I removed using simple data frame slicing. For the city of Detroit, I was able to extract the data from a table. Similar to the data extraction for Windsor the data frame for Detroit also had extra rows such as the title of the table. These where removed. The data frame for Windsor also had duplicate entries for some Boroughs. This happened because some Boroughs are divided into sub-areas which are then divided into neighborhoods. The python extraction script first recorded all the neighborhoods in a borough and then repeated the process for each of the sub-areas which means the neighborhood was recorded twice once under the Borough and again under it sub-area. I removed all rows which had sub-areas to remove the duplicate entries.

2.3. Feature Selection

Using the Foursquare API we get at most 100 venues within the radius of 500 meters of each neighborhood. For each of the Venues the Foursquare API also gives us the name of the venue, the latitude and longitude of the venue and the categories the venue belongs to. The goal of this project is to cluster neighborhood based on the type of venues present in the neighborhood. To do this we need to extract and select features that will help the K-means clustering algorithm separate neighborhoods into different clusters. The geolocation of each venue and venue name are not required to cluster the neighborhoods. So we create a new data frame that has each venue in each of the neighborhoods and each available category label as a feature. The value of each category feature will be Zero if the venue does not belong to this category. On the other hand, the value of the category type will be one if the venue fits in that particular category. This is done using one hot encoding. We then take the mean for each category for all the venues that belong to a particular neighborhood. This shows us the frequency of each venue category in a neighborhood. Using these features we can compare the neighborhoods based on the frequency of each venue category that is present in the neighborhoods.

3. Methodology

3.1. Exploratory Data Analysis

Exploratory data analysis can reveal interesting things about the data that we are working with. For this section of the project I have looked at the five top venues in each of the neighborhoods. Figure 1 shows the partial list of neighborhoods with the top five venues listed based on frequency.

```
----Art Center----
          venue  freq
0          Café  0.12
1  Intersection  0.08
2      Boutique  0.08
3    Art Museum  0.08
4   Pizza Place  0.08


----Atkinson Ave.----
               venue  freq
0       Intersection  0.29
1     Hardware Store  0.14
2  Mobile Phone Shop  0.14
3      Garden Center  0.14
4     Rental Service  0.14


----Atkinson Avenue----
```

Figure 1: shows the top 5 venues in a neighborhood

	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
0	Art Center	Café	History Museum	Art Museum	Pizza Place	Intersection	Boutique	Pharmacy	Garden	Art Gallery	Science Museum
1	Atkinson Ave.	Intersection	Mobile Phone Shop	Hardware Store	Furniture / Home Store	Rental Service	Garden Center	Electronics Store	Ethiopian Restaurant	Duty-free Shop	Distillery
2	Atkinson Avenue	Intersection	Mobile Phone Shop	Hardware Store	Furniture / Home Store	Rental Service	Garden Center	Electronics Store	Ethiopian Restaurant	Duty-free Shop	Distillery
3	Aviation	Convenience Store	Pizza Place	IT Services	Lounge	Women's Store	Dive Bar	Falafel Restaurant	Factory	Event Space	Ethiopian Restaurant
4	Bagley	Food	Clothing Store	Grocery Store	Intersection	BBQ Joint	Event Space	Factory	Ethiopian Restaurant	Dive Bar	Duty-free Shop

Figure

3.2. Predictive Modeling

For this project I have chosen to use an unsupervised machine learning model K-means clustering. In this project I create a data frame that all the neighborhoods in both Detroit and Windsor each with the 10 feature. The features are the

4. Discussion

5. Results

6. Figure 1:

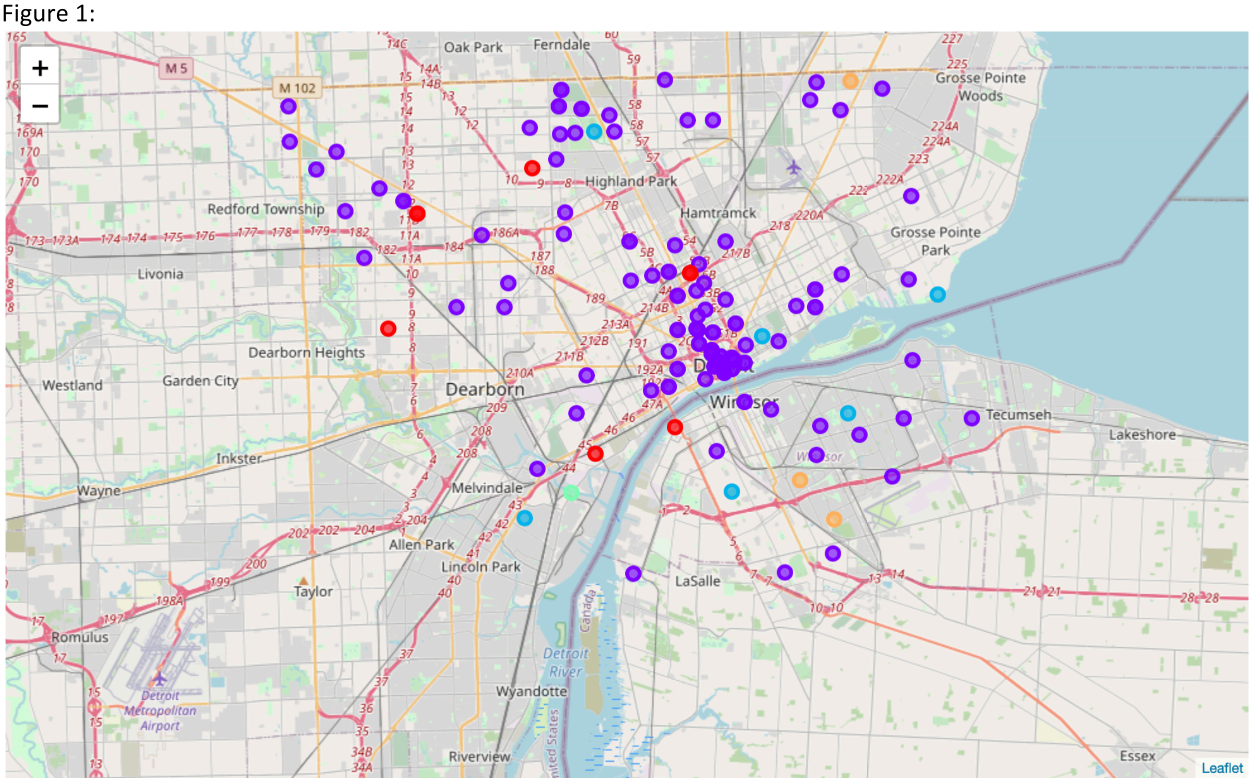

Figure 1:

```
downtown_merged.loc[downtown_merged['Cluster Labels'] == 0, downtown_merged.columns[[1] + list(range(5, downtown_merged
```

	Borough	Longitude	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
14	West Side	-83.073334	0.0	Bar	Café	Coffee Shop	Performing Arts Venue	Distillery	Falafel Restaurant	Factory	Event Space	Ethiopian Restaurant	Electronics Store
51	New Center	-83.064966	0.0	Bar	Art Gallery	History Museum	Middle Eastern Restaurant	Women's Store	Doctor's Office	Farmers Market	Falafel Restaurant	Factory	Event Space
94	West	-83.150472	0.0	Bar	Women's Store	Flower Shop	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Factory	Event Space	Ethiopian Restaurant	Electronics Store
96	West	-83.213247	0.0	Art Gallery	Women's Store	Dive Bar	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Factory	Event Space	Ethiopian Restaurant	Electronics Store
97	West	-83.213247	0.0	Art Gallery	Women's Store	Dive Bar	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Factory	Event Space	Ethiopian Restaurant	Electronics Store
114	West	-83.229016	0.0	Bar	Theater	Pizza Place	Discount Store	Women's Store	Farmers Market	Falafel Restaurant	Factory	Event Space	Ethiopian Restaurant
118	Southwest /Near West	-83.116590	0.0	Bar	Home Service	Women's Store	Dive Bar	Farmers Market	Falafel Restaurant	Factory	Event Space	Ethiopian Restaurant	Electronics Store
149	Historic Districts	-83.064966	0.0	Bar	Art Gallery	History Museum	Middle Eastern Restaurant	Women's Store	Doctor's Office	Farmers Market	Falafel Restaurant	Factory	Event Space

Figure 3:

```
downtown_merged.loc[downtown_merged['Cluster Labels'] == 3, downtown_merged.columns[[1] + list(range(5, downtown_merged
```

	Borough	Longitude	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
29	Downtown	-83.129093	3.0	Bridge	Women's Store	Dive Bar	Farmers Market	Falafel Restaurant	Factory	Event Space	Ethiopian Restaurant	Electronics Store	Duty-free Shop

Figure 4:

```
downtown_merged.loc[downtown_merged['Cluster Labels'] == 4, downtown_merged.columns[[1] + list(range(5, downtown_merged
```

	Borough	Longitude	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
17	South Side	-82.986301	4.0	Furniture / Home Store	Women's Store	Dive Bar	Farmers Market	Falafel Restaurant	Factory	Event Space	Ethiopian Restaurant	Electronics Store	Duty-free Shop
18	South Side	-83.004876	4.0	Furniture / Home Store	Baseball Field	Women's Store	Flea Market	Farmers Market	Falafel Restaurant	Factory	Event Space	Ethiopian Restaurant	Electronics Store
69	East	-82.977877	4.0	Furniture / Home Store	Women's Store	Dive Bar	Farmers Market	Falafel Restaurant	Factory	Event Space	Ethiopian Restaurant	Electronics Store	Duty-free Shop

Figure

```
downtown_merged.loc[downtown_merged['Cluster Labels'] == 2, downtown_merged.columns[[1] + list(range(5, downtown_merge
```

	Borough	Longitude	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
4	East End	-82.978921	2.0	Home Service	Basketball Court	Park	Dive Bar	Farmers Market	Falafel Restaurant	Factory	Event Space	Ethiopian Restaurant	Electronics Store
15	West Side	-83.042389	2.0	Park	Trail	Women's Store	Distillery	Farmers Market	Falafel Restaurant	Factory	Event Space	Ethiopian Restaurant	Electronics Store
56	North	-83.117251	2.0	Pool	Park	Discount Store	Falafel Restaurant	Factory	Event Space	Ethiopian Restaurant	Electronics Store	Duty-free Shop	Dry Cleaner
63	East	-82.930019	2.0	Park	Pool	Scenic Lookout	Movie Theater	Harbor / Marina	Lighthouse	Event Space	Ethiopian Restaurant	Electronics Store	Duty-free Shop
64	East	-83.447946	2.0	Soccer Field	Park	Trail	Women's Store	Discount Store	Factory	Event Space	Ethiopian Restaurant	Electronics Store	Duty-free Shop
79	East	-83.025464	2.0	Park	Women's Store	Distillery	Farmers Market	Falafel Restaurant	Factory	Event Space	Ethiopian Restaurant	Electronics Store	Duty-free Shop
116	Southwest /Near West	-83.154776	2.0	Park	Seafood Restaurant	Women's Store	Distillery	Falafel Restaurant	Factory	Event Space	Ethiopian Restaurant	Electronics Store	Duty-free Shop

Figure

7. Conclusion