**Calgary & Toronto**

**Neighborhood Comparison**

**Introduction**

**1.1 Problem**

Relocation is common among many jobs, and it is difficult decision to make. This project will explore the downtown location of two major cities in two different provinces, namely Calgary and Toronto. Venues in both locations will be explored and used for decision making.

**1.2 Interest**

We are interested in discovering how those neighborhoods are similar or different, and this information will be an aid when making decisions such as moving, migration, or relocation.

**Data**

**2.1 Data Acquisition**

Calgary, and Toronto postal code are scraped from:

<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_T>

<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

Toronto longitude and latitude data is reused from previous project, and for Calgary, data is scraped from:

<https://www.geonames.org/postal-codes/CA/AB/alberta.html>

Both cities popular venues are collected from foursquare.com, using latitude and longitude to locate the neighborhood and explore all venues within 500 meters of the neighborhood.

**2.2 Data Cleaning**

All data collected are stored into two separate data frames. For Calgary data, there were lots of unassigned postal codes, and they are deleted before combing with locational data. At the same time, there were many neighborhoods with different names, but refer to the same postal code, thus neighborhood names were trimmed down to a single neighborhood name for each postal code. It is then combined with the longitude and latitude data, and ready for further processing.

As of Toronto data, unlike Calgary data, it is much less ambiguous, each postal code only refers to a single neighborhood, it is simply scraped from the website and then combined with locational dat.

Calgary downtown neighborhoods is selected by plotting all the neighborhoods in a map and selected manually. Then foursquare.com is used to explore all the venues within 500 meters of each neighborhood, however, for Calgary, only Connaught neighborhood has enough venues worth comparing, all other neighborhoods have insufficient venues for comparison. Therefore, only Connaught neighborhood is selected for Calgary.

Since only one neighborhood is selected, therefore for Toronto, only the most central location is selected as well, which is the Central Bay Street Area.

For both neighborhoods, one hundred venues are explored, and top 20 is selected for comparison.

Since only top 20 is selected in each neighborhood, the frequency value is adjusted and converted to a percentage value using the following equation:

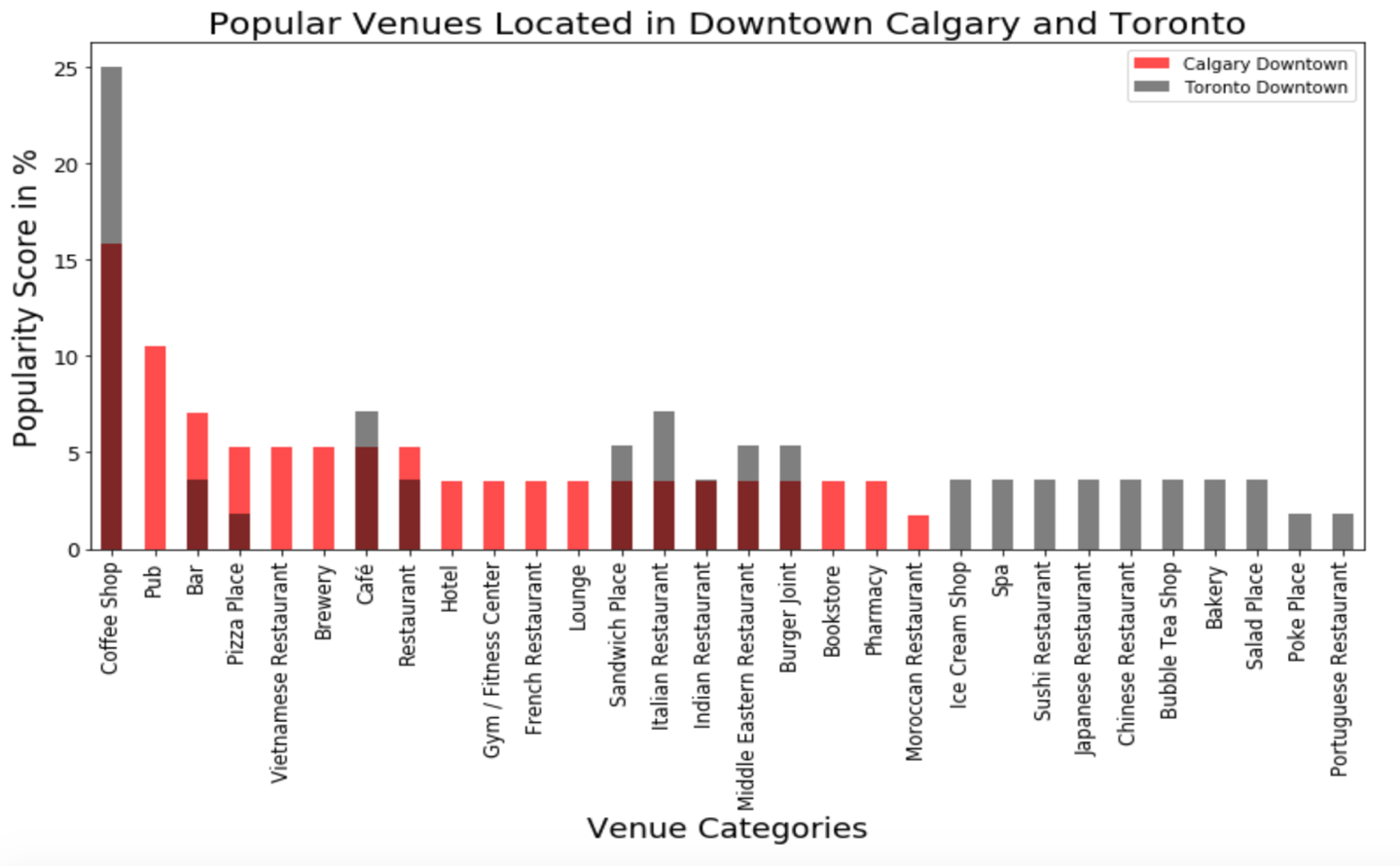
However, both neighborhoods do not share all venues, therefore, all venue information are combined in a single data frame, and if any venue is not in the top 20 list of either neighborhood, then a zero percent of popularity is assumed.

**2.3 Data Reprocessing**

K-means algorithm is intended to be used on the data frame for clustering, however, this is not possible due to insufficient amount of data. To fix this problem, all neighborhoods in Toronto downtown areas is added to the data frame.

**Methodology**

**3.1 Visualization**

Comparison between Central Bay Street, Toronto and Connaught, Calgary

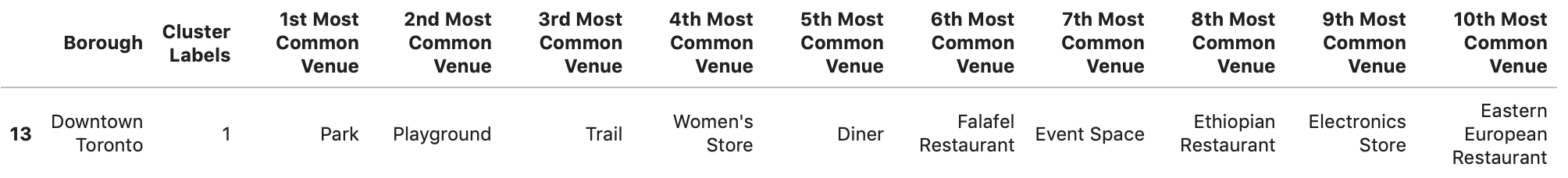
**3.2 K-Means Clustering**

Combined all Toronto downtown neighborhoods and Calgary downtown neighborhood, a k-means clustering algorithm is performed on this new data frame, where k = 5.

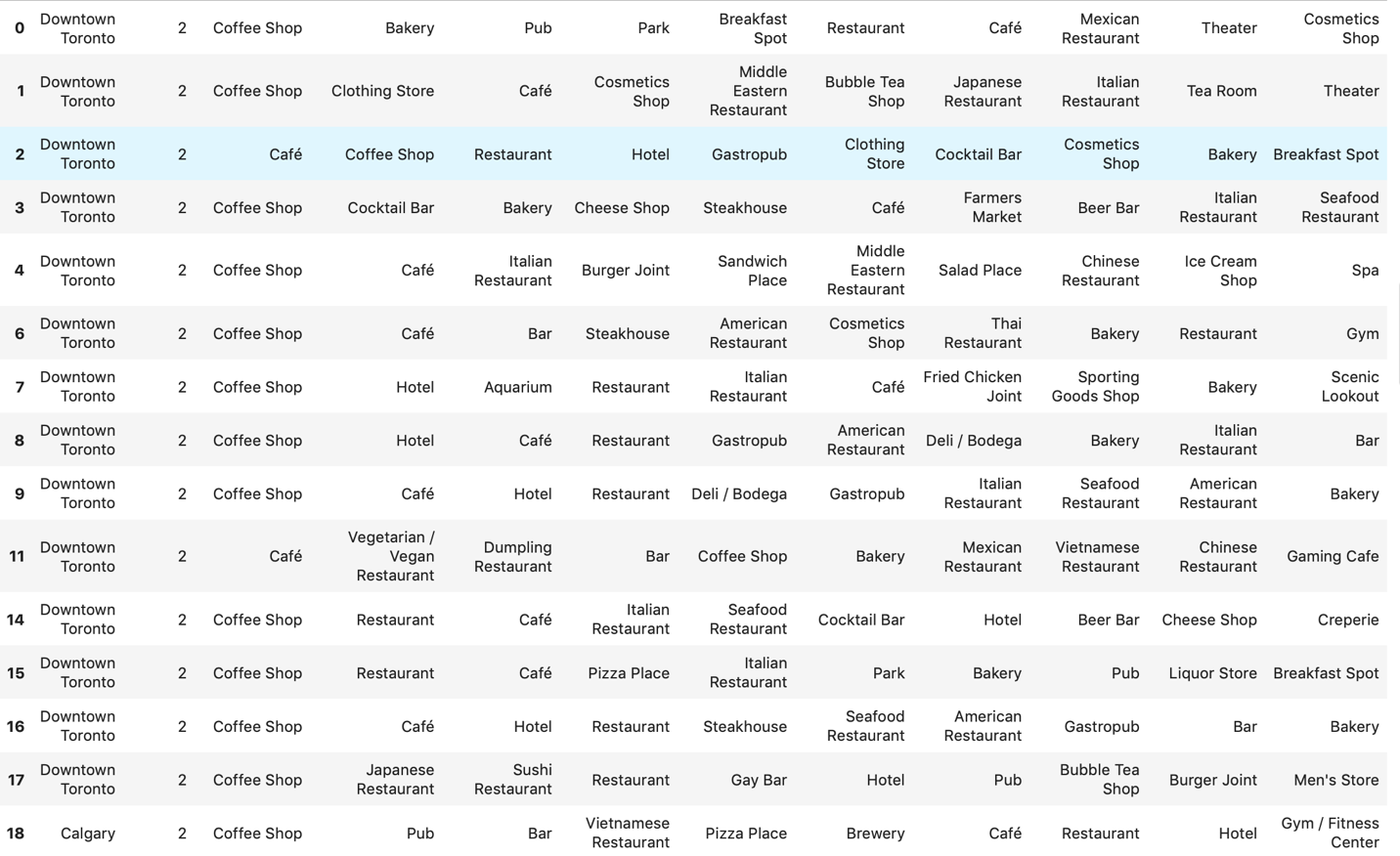
**Cluster #1**

****

**Cluster #2**

****

**Cluster #3**

****

**Cluster #4**

****

**Cluster #5**

****

**Conclusion**

Just compare two most centralized downtown neighborhoods in Toronto and Calgary, it is obvious they are quite distinct, and both cities do not have popular venues from each other. From this observation we can conclude both neighborhoods are quite distinct from each other. Must exert caution before relocating between these two neighborhoods.

However, from the K-means, neighborhoods are divided into five clusters, and it is apparent that the third cluster has most neighborhoods, and it also contains the Calgary downtown neighborhood. Even though some neighborhoods are quite different from others, but majority of them including Calgary downtown neighborhood share a very high degree of similarity. This indicate that these neighborhoods are very similar, and it is safe to relocate between these neighborhoods.