**Coursera Capstone Project**

**Finding a list of popular venues closed to TTC stations in Toronto**

**BY: Gayathri Sithampalam**

**Introduction**

TTC (Toronto Transit Commission) plays a huge role is taking commuters from one end of the city to another in Toronto. In extreme weather they are the only hope for many commuters. Being closely located near a TTC subway station gives a great amount of advantage to a lot of small business. It increases the number of customers and there by the profits. Even businesses that don’t rely on walk-in customers, like offices, might also benefit from being close to transit stations due to shorter commute times for their employees. Buying a house close to a TTC station brings you to a step closer to all the amenities in the neighborhood. Many property developers are taking this into consideration prior to building their condominiums. In Toronto, a TTC station that is within reach is considered as an added luxury by many.

**Business Problem**

The objective of this project is to find the best location, or a TTC subway station, around which it is ideal to buy a house, open or business or find an employment opportunity for that matter. Using some of the techniques that were taught in this course, this project will find the perfect cluster or neighborhood which would serve as an advantageous spot for an individual like me who solely relies on public transportation. My ideal location would consist of a wide of restaurants, banks, and bus station. Let’s see if I could I find a place that fulfills all these criteria.

**Data**

* A list of all the TTC stations in Line 1,2,3, and 4 will be used. Their coordinates will be used to map them and to retrieve all the venues close to them. Data for all the four lines were retrieved from [: https://scruss.com/blog/2005/12/14/toronto-subway-station-gps-locations/](:%20https:/scruss.com/blog/2005/12/14/toronto-subway-station-gps-locations/). The four .csv files were merged and read as one file to get all the station points in one. Since the coordinates were readily available, there was no need of other packages to retrieve them
* All the venues in the vicinity were obtained with the help of Foursquare API. With the help of Foursquare API, and machine learning technique such as K means clustering, the locations were plotted in a folium map. In addition, data cleaning, wrangling and visualization were utilized to achieve the result.

**Methodology**

The process was begun by reading the .csv file that contained the names of all TTC stations and their coordinates. Once the crucial part was over, a folium map with all the subway stations in Toronto was plotted, with the help of Geocoders (Nominatim), and a for loop. This produced a beautiful plot with the full subway extensions within the city.

Map

Description automatically generated

Foursquare API is a social media platform, which people use to explore their neighborhoods and actively keep up with their friends and colleagues. We will be using it for the same purpose. Developer account of Foursquare gives us the opportunity to retrieve location data. To begin with, the credentials were set up, to request for up to 100 venues, within a radius of 500 meters. With the help of a function, we acquire all the details of the venues that are within the range. We quickly group all venues that were returned for each TTC station based on the distances.

Graphical user interface

Description automatically generated with low confidence

One hot coding was used to convert the categorical variables into binary values. In this case, it provides with information as to if a certain venue is present within the vicinity of the point of interest. This would also give us an insight as to which venues dominate around that particular TTC station.

Calendar

Description automatically generated with low confidence

Top 10 venues around each station were then determined and stored in a data frame. Once this has been set up, we utilized K means clustering, with K clusters=5. Each venue was given an appropriate cluster number next. A final folium map with the clusters was created to show the segregation.

Graphical user interface

Description automatically generated

Map

Description automatically generated

**Results**

Table

Description automatically generated

Cluster 0 (red) and 1 (purple) consisted of all the amenities within the mentioned radius. Cluster 1 had a lot more amenities than cluster 0.

Cluster 2 (light blue),3 (cyan blue) and 4 (chrome) did not have many venues around them, with each of them consisting of less than 2 types of amenities around them.

**Discussion**

Based on the results, it was clear that cluster 0 consisted of a huge number of cafes, and other amenities

Cluster 1- was dominated by cafes and coffeeshops equally. Not to mention, it also had all the required range of restaurants, banks, and bus stations

Cluster 2, 3 and 4- were dominated by parks, coffee house and restaurants respectively.

Chart, bar chart

Description automatically generated

Cluster 0 and 1 seem to have a higher advantage as they are located within downtown Toronto or more central locations, there by attracting more businesses. In addition, the presence of many companies, universities and banks has promoted more and more businesses to move there.

**Conclusion**

Based on the analysis, I would like to conclude Cluster 1 as the clear winner. Cluster 0 was not far off from cluster 1, and it also had its additional perks. The other clusters, in particular cluster 2, though located in and around the same area do not have a lot of amenities, this seems to be a bit of stretch. To answer the question in the picture, I would prefer to live in cluster 1, since it is located around Downtown Toronto and it also had a wide range of restaurants, is located close to banks and bus stations.