Calgary Private Cab Services

Table of Contents

1. Introduction

1.1. Calgary

Calgary is a city in the western Canadian province of Alberta. It is located in the south of the province where Bow River and the Elbow River converges, in an area of foothills and prairie, about 80 km (50 mi) east of the front ranges of the Canadian Rockies, roughly 299 km (186 mi) south of the provincial capital of Edmonton and approximately 240 km (150 mi) north of the Canada–United States border. The city anchors the south end of the Statistics Canada-defined urban area, the Calgary–Edmonton Corridor.

The city had a population of 1,285,711 in 2019, making it Alberta's most-populous city, the most-populous in western Canada, fourth-largest census metropolitan area in Canada and second-largest in western Canada.

Calgary's diverse economy includes activity in the energy, financial services, film and television, transportation and logistics, technology, manufacturing, aerospace, health and wellness, retail, and tourism sectors. The Calgary Metropolitan Region is home to Canada's second-highest number of corporate head offices among the country's 800 largest corporations. Calgary is also famous of having highest number of millionaires per capita of any major Canadian city. Moreover, in 1988 it became the first Canadian city to host the Winter Olympic Games.

The Economist Intelligence Unit ranked Calgary the most livable city in North America in both 2018 and 2019. Calgary was also ranked the best city in the world for drivers in 2019.

1.2. Background Information

Calgary is divided into four quadrants, and the quadrants in which a particular neighbourhood of Calgary is reflected as in its street address, which ends in NW, SW, SE or NE. Avenues typically run east-west, while streets run north-south. Once you adjust to this system, navigating around Calgary becomes very easy. However, Calgary is also one of the most geographically dispersed cities in the world. Rather than high-rise multi-story apartments, housing in Calgary generally consists of separate, detached homes and buildings. Unfortunately, the public transport system in Calgary isn’t as comprehensive as those seen in other large Canadian cities and compared to some other medium scaled cities. Therefore, citizens heavily depend on either personal vehicles or select their houses very close to their job reluctantly only because of the fact that proximity and time you spend on road during travelling to your workplace is really important. Moreover, travelling between quadrants (e.g Southwest to Northeast) can be tricky on public transport, and heavy traffic congestion is also common if you have to drive across bridges during rush hour. Last but not least, hefty parking costs and high insurance costs on personnel vehicles should not be disregarded.

1.3. Business Interest

As being the golden rule: Problem for somebody can be an investment opportunity for some others. After a high demand for a cheap, private and high level cab service from local citizens and businessmen, a private well-known investment firm would like to step in to the picture and would like to establish a Private Taxi \ Cab Service Company in Calgary in order to meet citizens demand. This business should provide cheap, efficient, high standard and punctual transportation services for their customers while being profitable.

2. Business

2.1. Challenges

Below are the main challenges Calgary population are facing daily in terms of transportation:

- Calgary is one of the most geographically dispersed cities in the world.
- The public transport system in Calgary isn't as comprehensive as those seen in other large Canadian cities
- Travelling between quadrants can be tricky on public transport
- Heavy traffic congestion is also common if you have to drive across bridges during the rush hours
- Hefty parking costs
- High personal vehicle insurance costs
- High Safety & Health Standards
- Increasing unemployment rate
- Increasing Competition
- Attracting Customer's attention
- Up & Downs in Economy
- Increase in gasoline prices

2.2. Opportunities

- Some neighborhoods are attractive among the young Canadian adults such as universities, cafes, etc.
- Rich economy & customer profile
- Nightlife is famous among other Canadian cities and business life is 24/7.
- Calgary relies heavily on the oil and gas industry.
- Many commercial and industrial enterprises are situated.

2.3. Business Scenario

Calgary shows promising results for establishing a Private Taxi Service Company due to structure of the city and high number of potential customers. Also, Canada is a really safe country to invest due to its top-notch business friendly government regulations. Moreover, customer class varies from mid to high within young to mid age population profile who need this type of service daily for many possible reasons such as schooling, travelling, entertainment, business, education, etc.

Initial investment cost will be high due to quantity of the branches and cabs. Therefore, to be profitable:

- A “Cab Service App” will be commercialized on all platforms. This platform will be the interface between company & customers. Some of the important features of this app are requesting a cab through the app as well as direct calling platform, giving feedback for the drivers and seeing estimated driving hours.
- Branches must be close to all potential customers. Especially to highly populated neighborhoods and business areas.
- Driving hours are very critical both for the customers and drivers.

3. Methodology

Though it requires high investment cost, in order to be profitable and efficient against competition, company needs to establish many branches in different parts of the city to optimize transportation costs and meet customer demand.

Methodology for this study will be as following:

In the first stage Calgary Population data will be collected and analyzed.

- Highly populated areas will be analyzed and classified by residential, business \ industrial areas.
- Highly populated areas will be visualized on a folium map and categorized in terms of population density.
- Top 10 to 12 locations will be determined and suitable areas for branches will be determined.

During the second stage, hot spots\ Attractive areas will be determined.

- Hot spots such as museums, libraries, hospitals, universities will be determined.
- Any missing data will be parsed from Wikipedia or similar websites and combined into main hotspot data frame

- Location coordinates will be determined via folium coordinate finder
- Finally, all hot spots data will be feed into base map to visualize highly dense attractive areas

3rd stage will focus on finding all close by venues around these Hotspots

- An access to Foursquare API will be established.
- All venues (1000 limit) within 500 m of hotspots will be determined via Foursquare API.
- Afterwards, a json file for all the venues will be generated.
- Finally, all localized venues will be classified

4th stage will mainly focus on clustering venues data

- Clustering will be needed to determine most attractive spots in Calgary.
- Clustered venues data and population data will be combined and help structuring the layers on base maps.
- Finally, folium maps will indicate best areas and help defining sizes of each branch for Taxi Services.

4. DATA

4.1. Sources

Calgary data website is a great source for data scientists. Through the website, one can access and browse many datasets related to Calgary such as crime rate, population data, businesses data, housing data. For more information refer to link below:

https://data.calgary.ca/

https://data.calgary.ca/Demographics/Census-by-Community-2019/rkfr-buzb

https://data.calgary.ca/Services-and-Amenities/Community-Services/x34e-bcjz

4.2. Libraries

Below are the code set for the libraries that have been used in this study.

```
!conda install -c conda-forge geopy --yes
!conda install -c conda-forge folium=0.11.0 –yes
# library to handle data in a vectorized manner
import numpy as np
# library for data analsysis
import pandas as pd
```

```
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)
import json # library to handle JSON files
# convert an address into latitude and longitude values
from geopy.geocoders import Nominatim
from bs4 import BeautifulSoup #For web Scrabing
import requests # library to handle requests
# Tranform JSON file into a pandas dataframe
from pandas.io.json import json_normalize
# Matplotlib and associated plotting modules
import matplotlib as mpl
import matplotlib.pyplot as plt
from matplotlib import cm
import matplotlib.colors as colors
# import k-means from clustering stage
from sklearn.cluster import KMeans
import folium # map rendering library
from folium.plugins import BeautifyIcon
print('Libraries imported.')
```

4.3. Datasets

For this study, there are 2 datasets that meets our criteria: “Calgary Census by Community Data” and “Calgary Community Services”. These are 2 detailed datasets and can be the base for future improvements.

Former dataset has information about neighborhood names, class of neighborhoods (residential or business), Sectors where the neighborhoods belong to, population of each neighborhood, count of dwellings in each neighborhood, etc.

Whereas in latter dataset, we can find information about name of attraction place (Museum, health center, community centers, etc), neighborhood where it belongs to, address, location coordinate data.

4.4. Defining Datasets

After reading these datasets, our initial observation on this dataset is, it is not clean and needs to be modified further. Also, there are many columns that is unnecessary for this study. Below are the fresh datasets that is generated:

- Df_new: only includes some columns from “Calgary Census by Community Data” df_new[["NAME","CLASS","RES_CNT","Clus_Db",'latitude','longitude']]
- Df_business: Industrial areas by grouping df_new dataset

- df_community: Community service locations in Calgary
- University_df: Colleges in Calgary
- Calgary_Hotspot_df: Combined data for colleges and community services
- Allurl_df: Foursquare information about the Venues all around the hotspots.
- Calgary_venues: Venues grouped data for all the hot spots
- Calgary_onehot: Amount and type of all venues in Calgary
- Calgary_grouped : Grouped Venues data for all the hot spots in Calgary_onehot
- Calgary_grouped_clustering: Clustering dataset of Calgary_grouped dataset where 'Neighborhood' data column is dropped so that clustering analysis can be done.
- Calgary_Hotspot_df_final: Final Clustered Data where it combines clustering labels, location coordinates and neighborhood names.

5. Data Analysis – First Step

5.1. Cleaning & Modification & Updates & Selection

Initially dataset for "Calgary Census by Community Data" is filtered and a new dataframe "df_new" is generated. It only includes some column names : NAME, CLASS, RES_CNT, Clus_Db, latitude, longitude. After the needed columns have been filtered, we are passing this data into cleaning process. "0" values is dropped and index reset back to "0". There is only one name which is causing issues during computation process (Douglasdale) while defining location coordinates. Once needed corrections and cleaning are applied, error will be fixed. Finally, df_new dataframe is sorted in terms of descending population quantity to define highly populated neighborhoods.

5.2. Library Features Selection Process

Pandas "unique" feature helps on defining categories in this dataset and on defining missing data categories to conduct further analysis.

"Sort ascending" feature helps on defining highly populated neighborhoods.

"Nominatim(user_agent="Calgarypop_explorer")" feature helps on defining location coordinates. After filtering population quantity and neighborhood, all the location coordinates can be found and assigned on needed columns by only feeding the code with the neighborhood names.

For missing data or some critical information, "Beautifulsoup" can help us on scraping data from websites. When values in "df_community" data frame "TYPE" column is examined, college names will be found missing. By scraping data from wikipedia page for "Calgary Universities & college List", full college names list can be found for further analysis.

Matplotlib bar charts helps on visualization of the highly dense neighborhoods.

Finally, folium maps helps for initial and final visualization of location data. Also, clustered data can be fed into the map to categorize the locations and markers can be used the define location borders for the Taxi Service Branches.

5.3. Initial Analysis

Initial analysis shows that Calgary is suitable for private cab services.

Chart below depicts “Age Categories vs Population”.

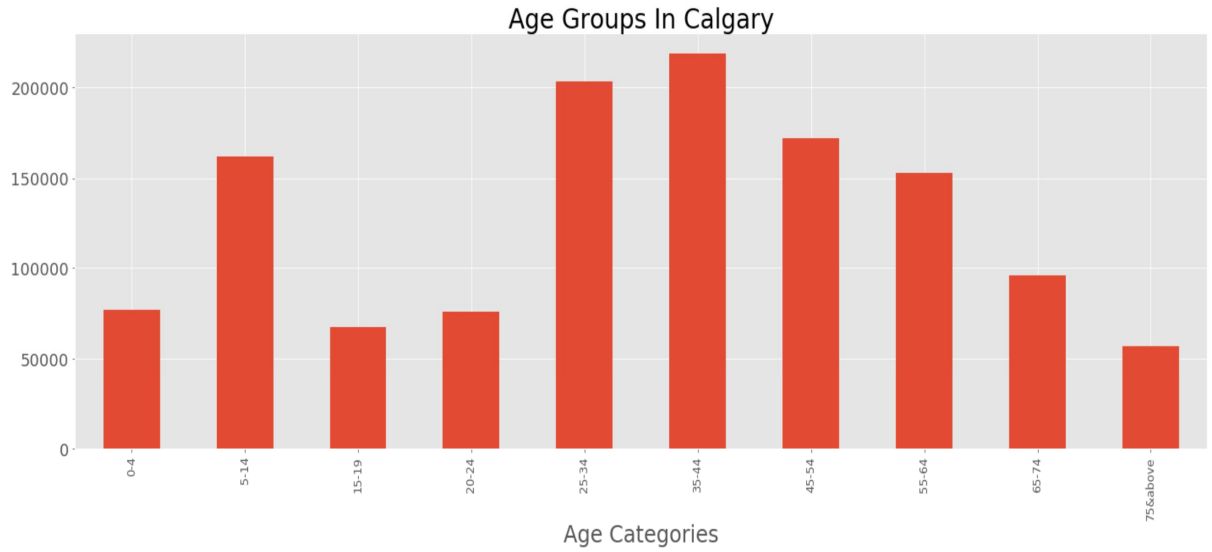


Figure 1. Age Groups in Calgary

It shows City has a diversified age group and majority of the people are fitting in working or studying class who can afford private cab services and use it in their daily life.

5.4. Location Data

Neighborhood locations are defined through a “for loop” code which iterates through “ geolocator = Nominatim(user_agent="Calgarypop_explorer") ” which is a library from “geopy.geocoders”. Same approach can also be adopted for the next “attractive places” dataset and “ Foursquare Venues” dataset.

5.5. **Initial Results & Visuals**

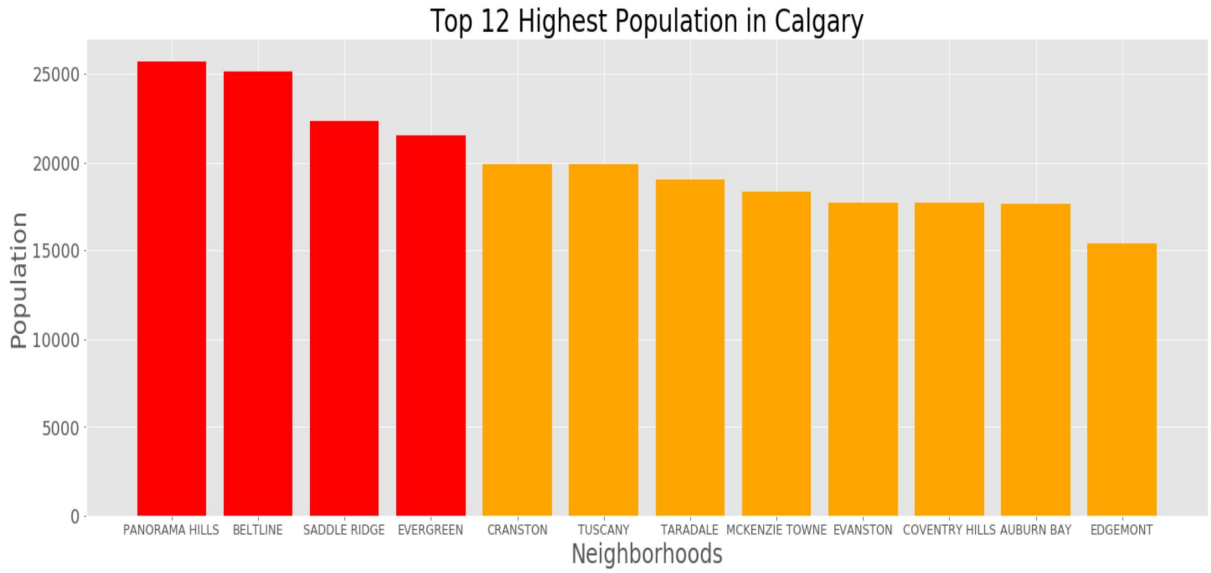
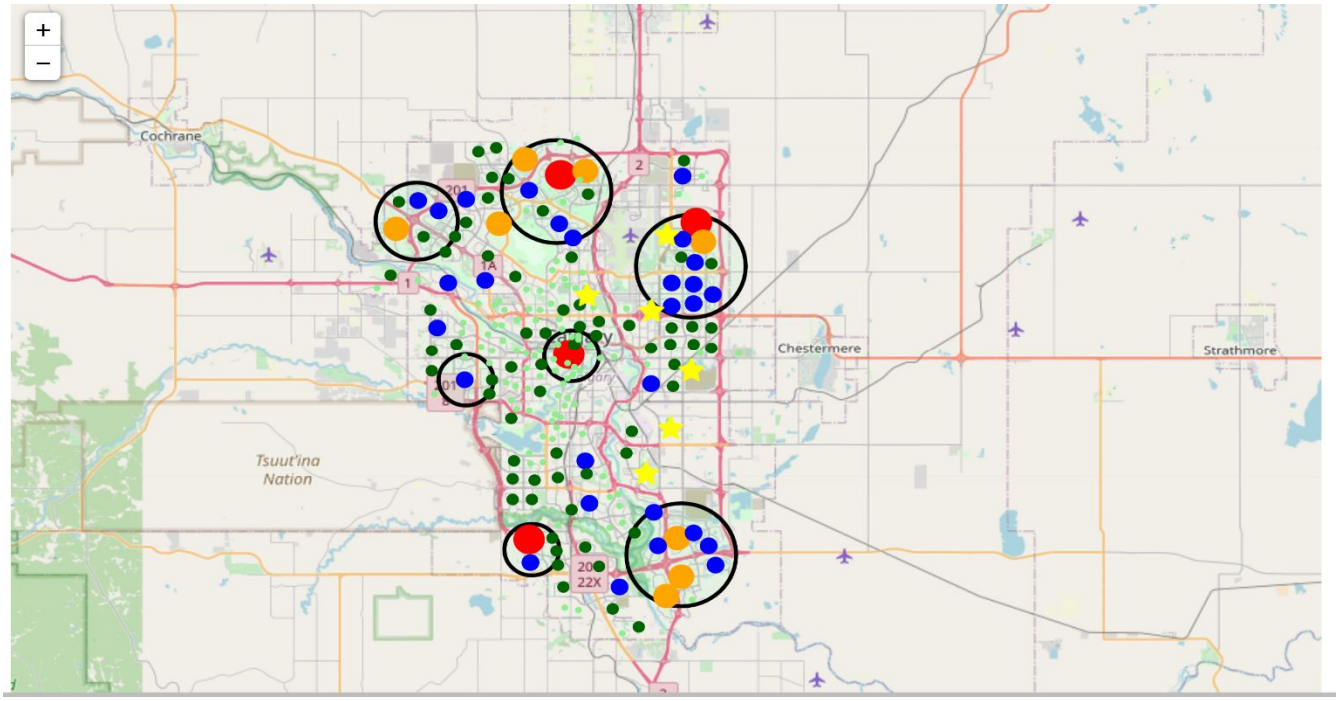


Figure 2. Top 12 Highest Neighborhoods in Calgary

The Matplotlib bar chart shown above illustrates top 12 highest population in Calgary. However, this data is meaningless until a folium map is used to define location coordinates.

Figure 3. Population Distribution in Calgary



Above is the folium map including all population data. It has 7 black circles that define optimum areas for the branches and "Yellow Stars" specifies industrial areas. All colored circles have radius proportional to their population range mentioned as below:

- Red marked circle specifies the highest population areas where "neighborhood population" > 20,000 with a radius of 10 unit
- Orange marked circle specifies high population areas where "neighborhood population" is between 15,000 to 20,000 with a radius of 7.5 unit
- Blue marked circle specifies med-high population areas where "neighborhood population" is between 10,000 to 15,000 with a radius of 5 unit
- Dark green marked circle specifies medium population areas where "neighborhood population" is between 5,000 to 10,000 with a radius of 2.5 unit
- Light green marked circle specifies low population areas where "neighborhood population" is between 0 to 5,000 with a radius of 1 unit

When we compare the bar chart to folium map, map gives more information for a decision.

By this analysis, initial branch sizes and locations are determined. However, without venues data it will be really hard to say if citizens are just visiting nearby places or travelling around the city.

6. Data Analysis – Second Step

6.1. Hot Spots in Calgary – Data Cleaning, Modifying and Updating

In previous section, after examining initial population dataset, initial branch locations are planned. But this is only showing static data for the citizens and would not make any sense if none of the citizens move around the city. So that a dynamic locations dataset, where people visit daily such as cafes, libraries, hospitals, universities, is needed. By the help of the latter dataset, we will have a better view on citizens condition in static and dynamic respects so that we can have more chance to take correct business decisions.

Initial check on the "df_community" dataset shows that even the location coordinates data is in place and saved good amount time. However, once we check the unique values in "TYPE" column, all the place names are in place except Universities & Colleges. Therefore, we need to find the names on the web and scrape the data & join it into our "df_community" dataframe list.

With the help of "Beautifulsoup" libraries, we are able to source the data from wikipedia page https://en.wikipedia.org/wiki/Category:Universities_and_colleges_in_Calgary. Afterwards, a "for loop" code is assigned just to receive the college names from the site. Then, location data

for the colleges can be defined. After cleaning and modifying the dataframe list (and the column names), universities dataframe list for the college names can be joined with the main "df_community" data set.

6.2. Results & Visuals

Below is the folium map that indicates all hot spots in Calgary:

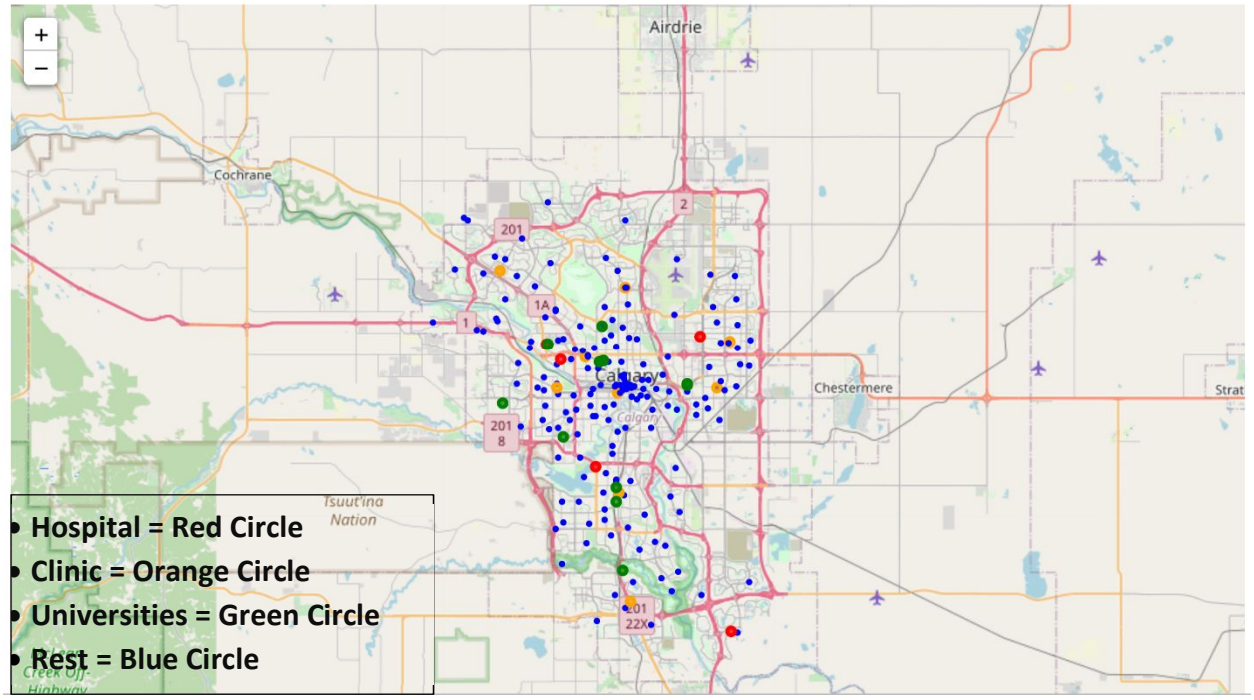


Figure 4. Attractive Hot Spots Areas in Calgary

7. Foursquare API

After defining hot spots in Calgary, Foursquare API can be used to collect venues information around these hotspots. This search is limited to 1000m radius for each hot spot and quantity is limited to 100 venues. Initially, all venue names are defined, and their URL addresses are collected. Then, the collected list of URLs' are feed into foursquare to collect related JSON file for all the venues. Afterwards, all the venue names and their location coordinates can be grouped under same neighborhood names and their presence frequency can be defined for each

neighborhood. Finally, only top 10 venues will be selected and sorted under “neighbourhoods_venues_sorted” dataframe.

8. Clustering Analysis

After collecting and grouping all venues data, clustering helps us on defining locations with most diverse venues options for citizens which is also directly related to highest customer potential areas for the Taxi Services. Even though we tried to setup different clustering numbers, most optimum clustering quantity is 2 which directly separated most and least attractive neighborhoods.

```
[639]: Calgary_grouped_clustering.groupby('Labels').mean()
```

[639]:

Labels	Art Gallery	Asian Restaurant	Athletics & Sports	BBQ Joint	Bakery	Bar	Baseball Field	Bookstore	Brazilian Restaurant	Breakfast Spot	Building	Café	Chinese Restaurant	Clothing Store	Co S
0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000
1	0.008378	0.002336	0.011682	0.011215	0.004206	0.009346	0.028037	0.010125	0.001168	0.004673	0.014019	0.019697	0.018692	0.001869	0.061

```
[639]: Calgary_grouped_clustering.groupby('Labels').mean()
```

[639]:

usic nue	New American Restaurant	Park	Performing Arts Venue	Pet Store	Pharmacy	Photography Studio	Pie Shop	Pizza Place	Playground	Plaza	Pool	Pub	Public Art	Recreation Center	Re (Ap
000	0.000000	0.041667	0.000000	0.000000	0.000000	0.000000	0.000000	0.041667	0.00000	0.000000	0.000000	0.00000	0.000000	0.000000	
673	0.007721	0.035826	0.017053	0.003115	0.004673	0.003115	0.004673	0.006542	0.03271	0.004796	0.004673	0.01729	0.003965	0.023364	

Cluster Label “0” = Least attractive

Cluster Label “1”= Most Attractive

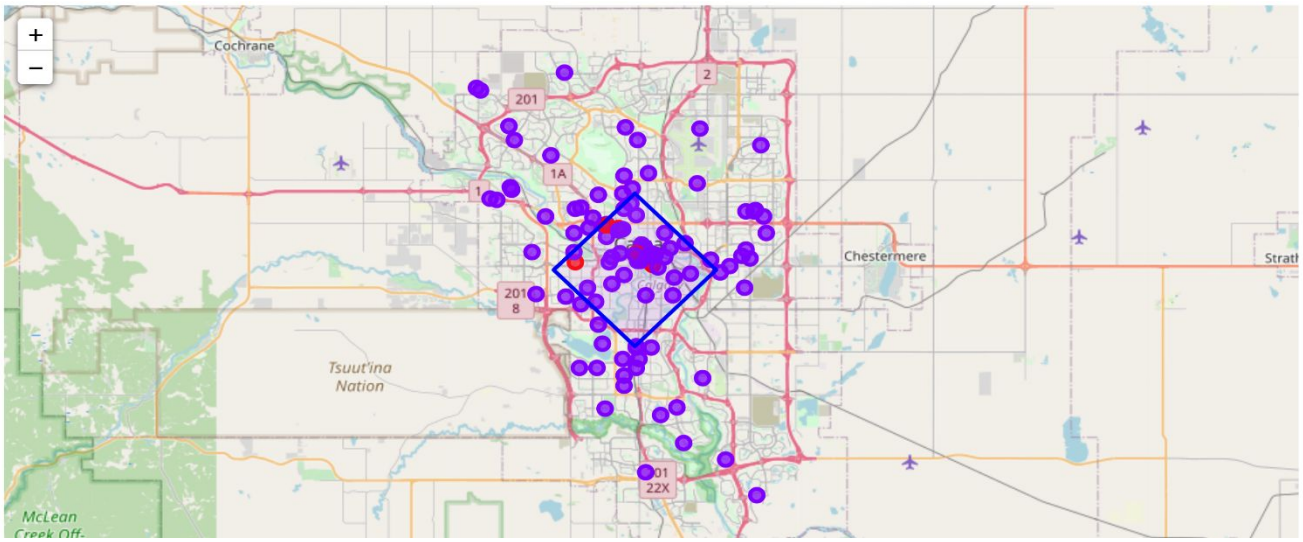
Figure 5. Clustering Neighborhood Data

Finally, all these cluster information will be re-assigned on venues data frame.

9. Final Results & Visuals

Below you may see the location cluster map that indicates location coordinates of most and least attractive zones. One of the findings is, Center of Calgary (Blue Rhombus) is the most attractive neighborhood where all different type of venues are located and this area is the most promising place in Calgary for finding customers which also makes this area the best place for the main office (headquarters).

[651]:



On the map, venues around the hotspots are clustered in to 2 categories.

* Purple indicates the neighborhoods with most attractive venues.

* Red indicates the neighborhoods with least attractive venues.

* Blue rhombus depicts most diverse venues are located in the center of Calgary.

Figure 6. Clustered Hotspots in Calgary Neighborhood Map

Finally, combining both cluster map and population map provides us all the promising locations for Taxi Service Branches.

[653]:

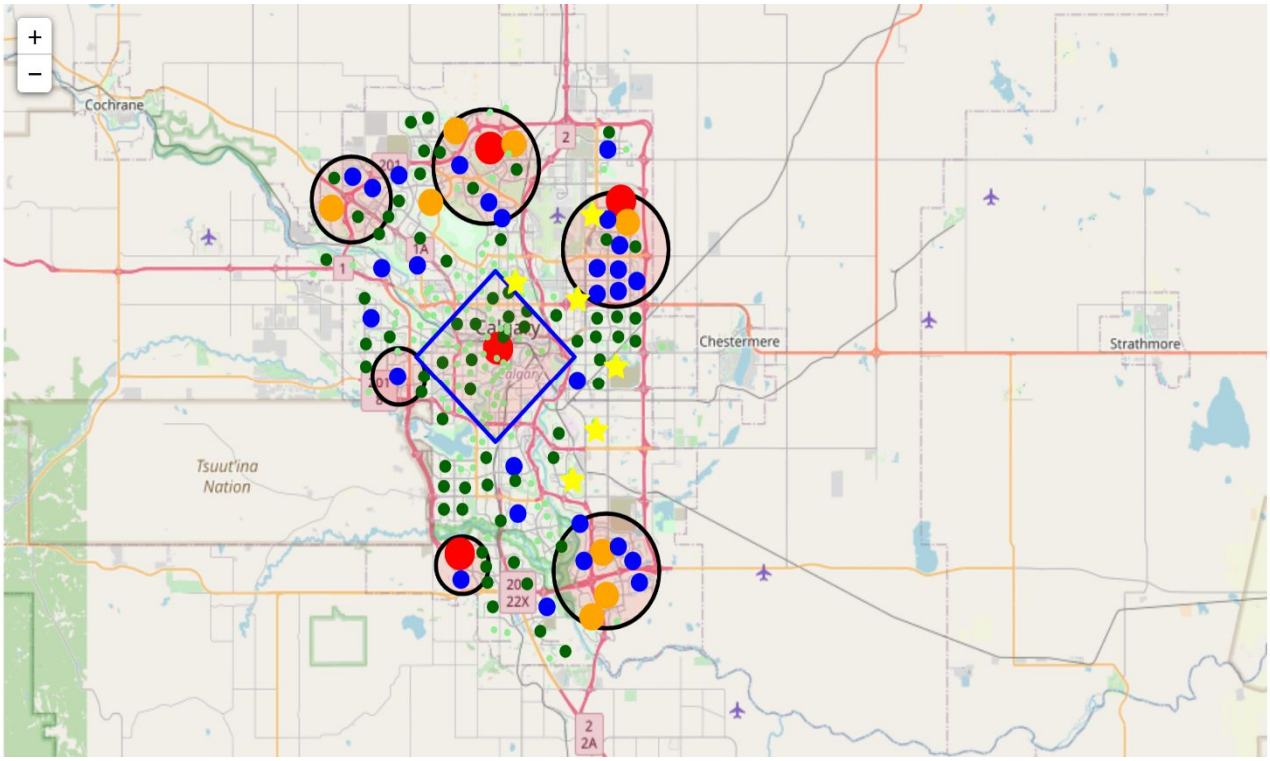


Figure 7. Final Branch Locations in Calgary Population Map

10. Conclusion

This analysis demonstrates that "Taxi Private Cab Company" can be established with 3 main & 1 middle & 3 small branches. By this way, all the important areas in Calgary can be covered and communication between branches can be established efficiently which in return makes the business profitable.

7 branches will be located as following:
* Main Branch - Rhombus @ the Center
* Big Branches - Big Circle North, Big Circle North East, Big Circle South East
* Small Branches – Small Circle West, Small Circle North West, Small Circle South West

Also, branches align with city highway map as show in Figure 8.

Except than the best places for the branches, this study also provided some other valuable information such as: Industrial areas are on the eastside , highly populated areas are on North, East, South East, Center and South West areas, and most attractive neighborhoods are located in the central part of Calgary.

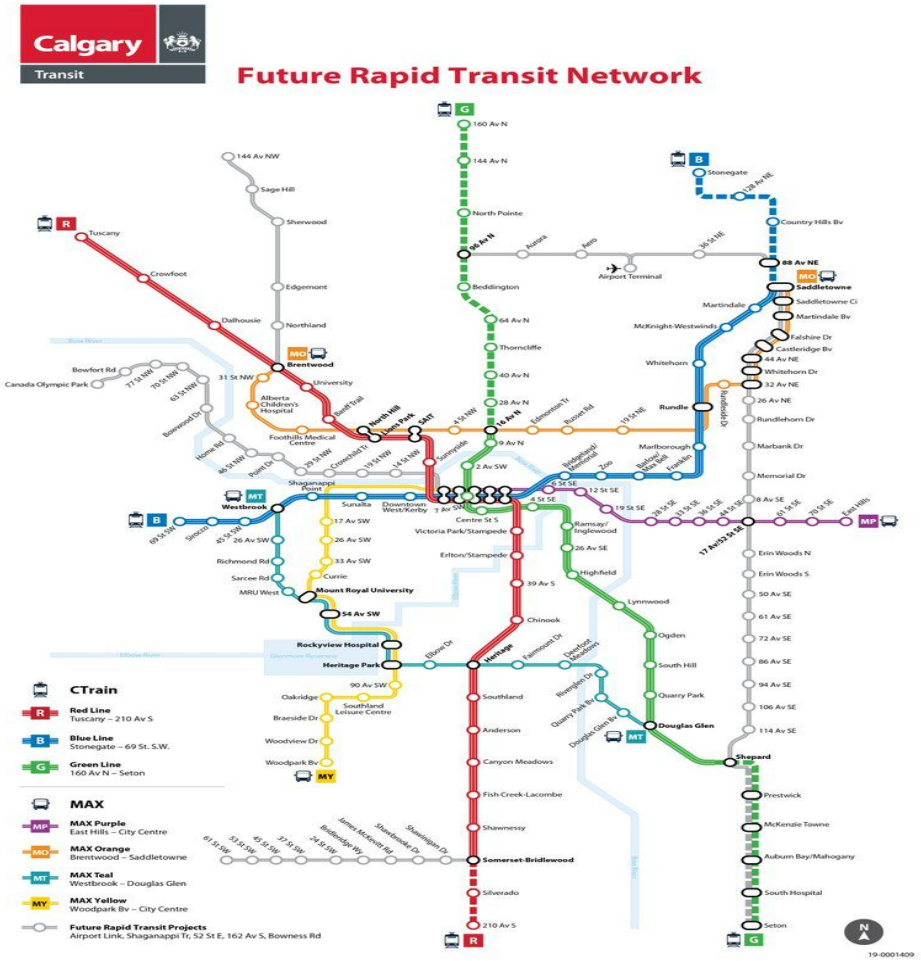


Figure 8. Calgary Transportation Network

11. Discussion

This study should be further extended by examining crime rate, population-age demographics, road conditions, competitor analysis, fuel consumption rate per km, city shop rental prices, etc.

Any extra information will make this study more compact and help on taking better business decisions.