

Opening a New Restaurant in Calgary

Introduction

The objective of this capstone project is to analyse and select the best locations in the city of Calgary, Canada to open a new restaurant. This will be a preliminary overview of the data providing the framework for a more detailed analysis later.

Target Audience

The target audience of this project is entrepreneurs looking to get an overview of the current restaurant scene in Calgary with the eventual goal of obtaining financing and/or investors to open a new establishment.

Data

Data will be obtained from the City of Calgary 2019 census. The original data will be cleaned and sorted using SQL queries so only the relevant data is imported.

The census data that is of interest is the type (ie: residential), community, community population, and community location.

Methodology

The census data is already grouped by community with a total of 212 unique values. The original location data was stored as a multipolygon that was split into two columns for the latitude and longitude location information. The multipolygon could be used with the population count to calculate a relative population density later. Although the census data contains a lot of useful data for population density and demographic distributions, it is outside the scope of this initial investigation.

After the census data has been wrangled, Foursquare API will be used to get the venue data in the respective neighborhoods. The venue data will be combined with the census data and then passed through the Scikit-Learn k-means clustering algorithm to reveal trends and patterns in the data distribution. The k-means clustering algorithm will combine the neighborhoods into groups that share similar attributes based on the venue data returned from Foursquare. From there, the individual clusters will be analysed to determine the most suitable location for a new restaurant based on current restaurant locations and distribution density.

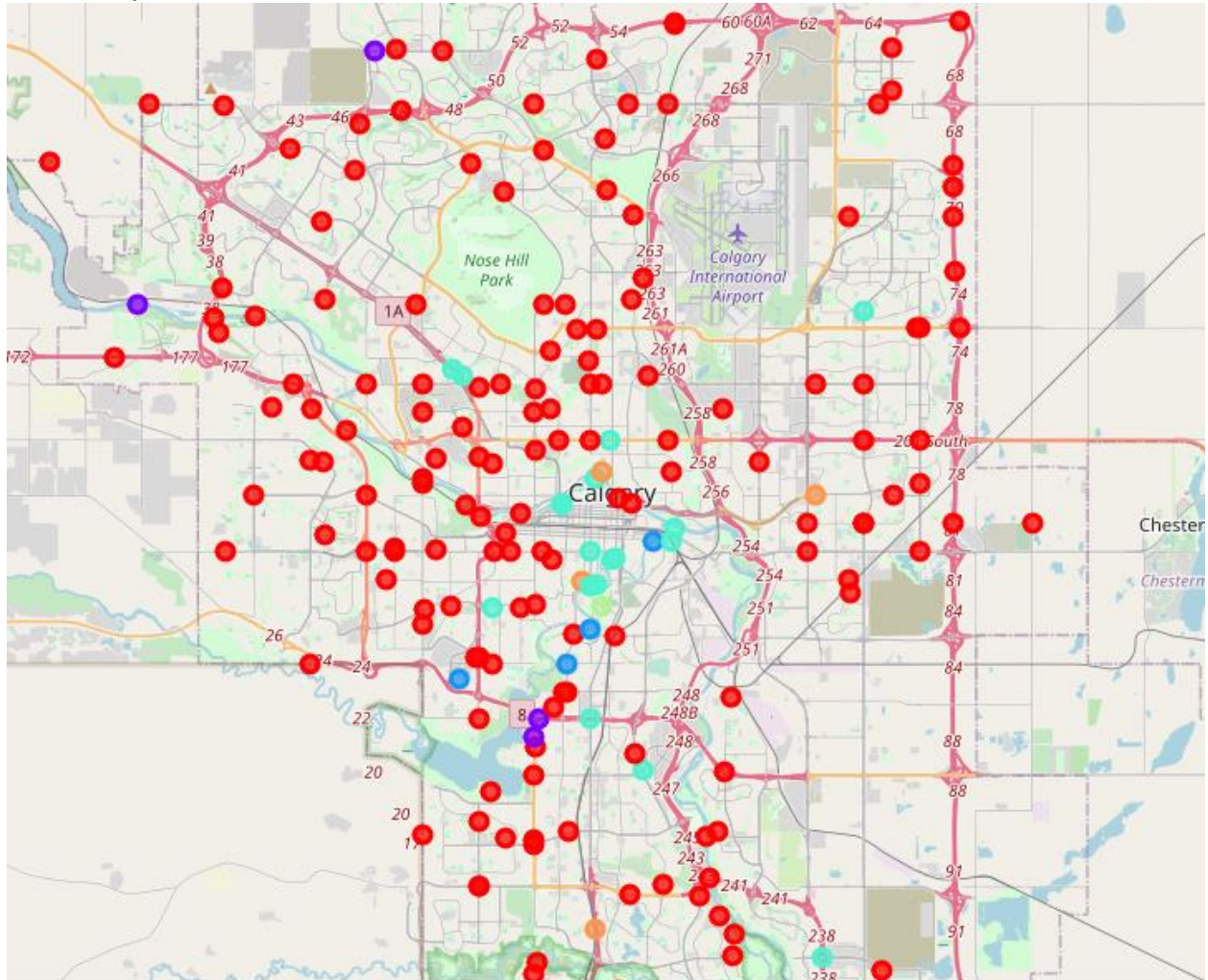
Prepared Census Data:

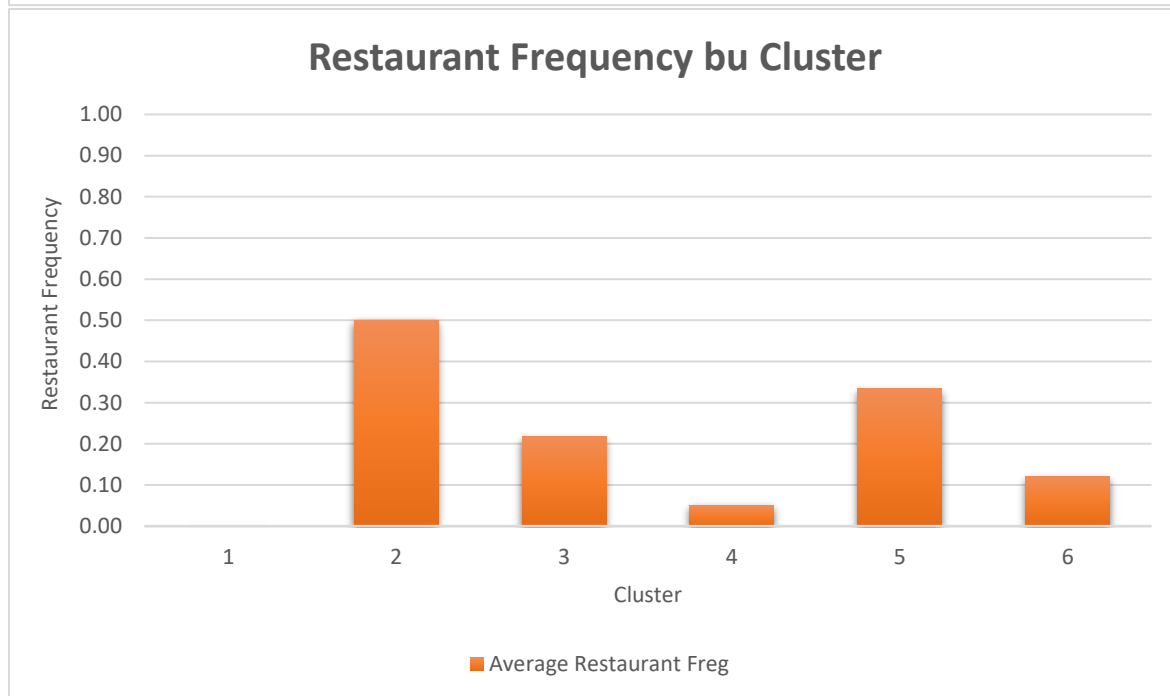
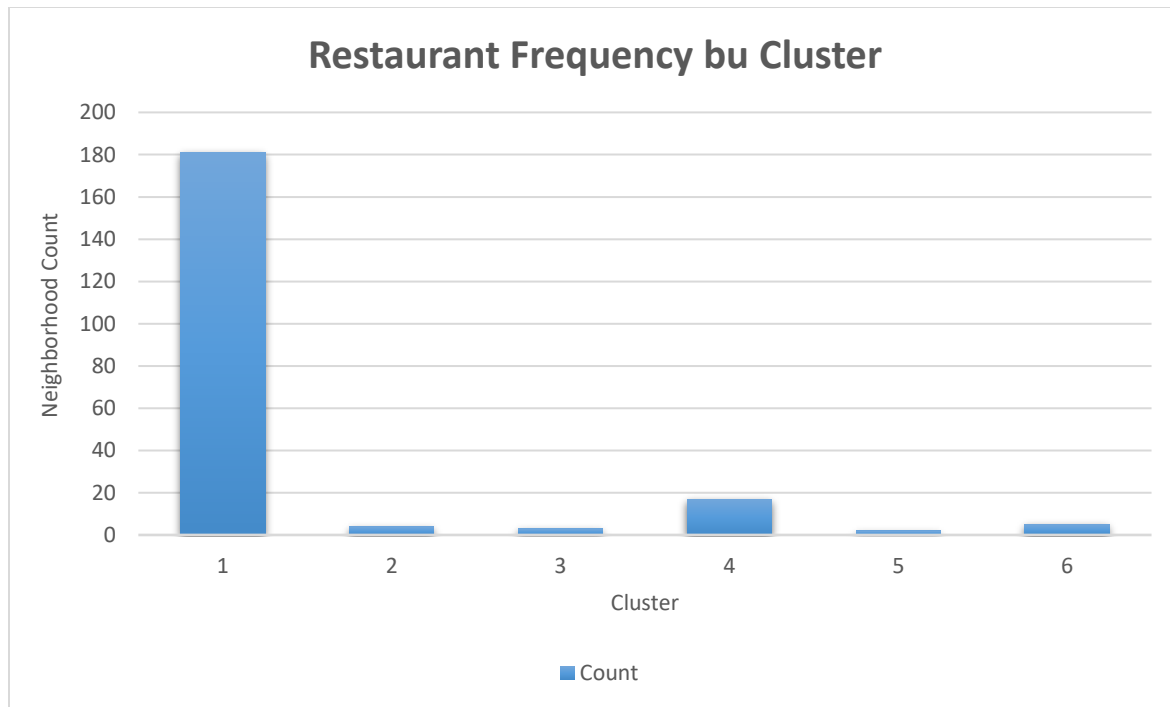
	CLASS	CLASS_CODE	COMM_CODE	NEIGHBORHOOD	RES_CNT	LATITUDE	LONGITUDE
0	Residential	1	ABB	ABBEYDALE	5957	51.066971	-113.935119
1	Residential	1	ACA	ACADIA	10520	50.980779	-114.050012
2	Residential	1	ALB	ALBERT PARK/RADISSON HEIGHTS	6997	51.045153	-113.981603
3	Residential	1	ALT	ALTADORE	6942	51.023194	-114.100546
4	Residential	1	APP	APPLEWOOD PARK	6981	51.045203	-113.921220

Results

The neighborhoods were split into 6 clusters using the Scikit-Learn k-means clustering algorithm based on restaurant frequency. Cluster 1 contained the highest frequency with 0.5 and a total of 4 neighborhoods. Most of the neighborhoods fell into cluster 0 with a frequency of 0 restaurants for 181 of the total 213 neighborhoods (85% of the neighborhoods).

Cluster Map





Discussion

While most of the neighborhoods do not contain any restaurants, this does not make them a suitable candidate for a new restaurant. Further analysis needs to be done on the type and distribution of restaurants based on what type of restaurant the stakeholders are looking to invest in. Further analysis may include calculating the population density from the multipolygon information and superimposing

that on the neighborhood map along with the distribution of the type of restaurant being investigated. Another avenue of investigation could be population demographic information and choosing a suitable restaurant type based on a combination of census and polling data.

Conclusion

The 2019 City of Calgary census data, combined with Foursquare API data, provided enough information for a preliminary analysis of restaurant distribution. Most of the neighborhoods in Calgary do not contain any restaurants. This, however does not make them suitable candidates for a new restaurant. Further investigation needs to be completed to provide a complete picture of all the contributing factors.