

LOCATION RECOMMENDATION FOR THE CONSTRUCTION OF A NEW AMUSEMENT PARK IN CALGARY, ALBERTA, CANADA.

OGUNDIPE OMOTOKÉ

Link to Notebook: [Amusement park recommendation Project](#)

October 27, 2020

1. INTRODUCTION

1.1. Background

When visiting a new environment or travelling to another region, province, or country on vacation, one of the major places we all look out for are places to have fun. A great place to have fun for both children, teenagers and adults is an amusement park. An amusement park is a large outdoor area with fairground rides, shows, refreshments, games of chance or skill, and other entertainments. An amusement park consists of roller coasters, water slides, carousels, Ferris wheel, bumper cars etc. The location of an amusement park greatly determines how much revenue it would generate. According to worldatlas.com and Wikipedia, Calgary is the 4th largest city in Canada by population, with the province of Alberta having the third largest GDP in Canada, making it a profitable place to build a new amusement park. There are a few amusement parks in Calgary with Calaway Park being one of the most popular amusement parks. Potential investors would benefit greatly in investing in building a new amusement park in Calgary, which could potentially become the center of attraction for the residents of Calgary, Canadians and international tourists.

One of the biggest challenges in the construction of a new amusement park is how to make an informed decision in selecting a location for the amusement park. Luckily, with Data science tools, we can observe different parameters and conduct an analysis to help determine the best location for our theme park. The most important thing to consider

when analyzing data is asking the right questions to provide the best insights! So what are the factors that determines the best location for a new amusement park?

1.2. Key Questions

- From available data, is there a relationship between population and entertainment venues in various neighbourhoods
- what is the relationship between current entertainment venues and number of restaurants in various neighbourhoods?
- Are entertainment venues closer to the city center?
- Are entertainment venues located closer to more populated areas or less populated areas

These questions are based on various internet research that implies there is a relationship between amusement park location with restaurant availability, population, available utilities, etc. More information on factors to consider when finding a location to construct a new amusement park.

- Link 1: <http://www.interthemepark.com/where-to-build.html>
- Link 2: : architecturaldesignmanagement.wordpress.com.
- Link 3: maps.unomaha.edu

This project aims to observe this information and analyze data in order to validate these relationships and suggest the best location to build an amusement park from the various neighbourhoods in Calgary

1.3. Description of Data

The data used in this project is obtained from [List of neighbourhoods in Calgary](#). This Wikipedia page contains a table containing the list of all neighbourhoods in Calgary, with useful information such as; population, sector, type, area, etc. The data is scrapped using the beautiful soup library, and using geopy, the latitude and longitude is obtained. Also, Four square API is used in obtaining useful information such as restaurant and other

entertainment venues in each neighbourhood, in order for analysis to determine the best location for a new amusement park. Using packages such as NumPy, matplotlib, folium, pandas, and more, an analysis and visualization of the data would be carried out in order to determine the best neighbourhood to situate a new amusement park.

1.4. Reasons for Data Selection:

- Entertainment venues: The reason for selecting entertainment venues is to analyze and observe each neighbourhood's relationship with other sources of entertainment. A neighbourhood with a high number of entertainment venues would generally attract people. Establishing a new amusement park in a location with a high population and other entertainment sources would generally set the business up for success. This is because people who would visit other entertainment venues would be attracted to an amusement park that is within a close vicinity. Entertainment venues includes; theme parks, zoos, art theatres, movie theatres, stadiums, museums, etc.
- Restaurant: As entertainment venues attract people to a location, so the restaurants. Tourists or residents of a particular neighbourhood are known to be interested in exploring new places to try different kind of cuisines/food. A neighbourhood with a large number of restaurants already attracts people. Situating an amusement park in a location with a high number of people would generally be a plus for an amusement park.

Entertainment, restaurant, and population data are the major factors that can help in identifying the best place to locate a new amusement park.

2. DATA COLLECTION AND PREPARATION

2.1. Data Cleaning

The Data was scrapped from Wikipedia using the Beautiful soup library and urllib library. The initial data scrapped from Wikipedia had 257 rows and 12 columns. The desired columns including data on Neighborhood name, Quadrant, Sector, Type, Population, and Area size (Km^2) was obtained. The NaN rows (empty rows) were then removed leaving the cleaned data with 251 rows and 6 columns. Using Geopy, the coordinates of each

location was obtained, and the Neighbourhood whose coordinates were not retrievable was removed leaving the final data set with 245 rows and 8 columns.

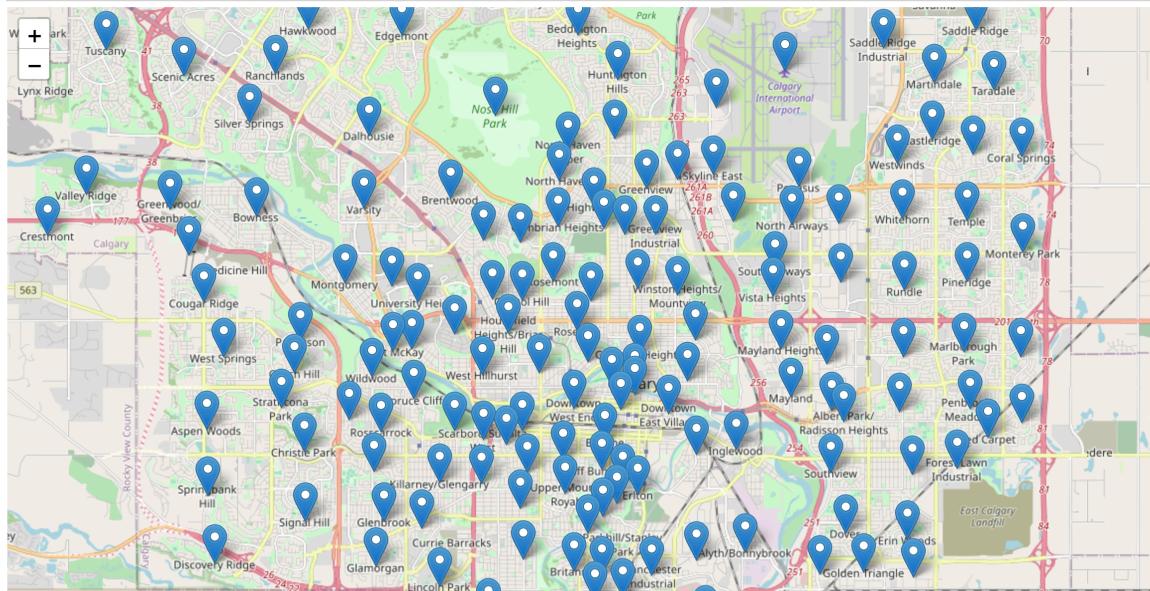


Figure 1: Map of Calgary showing Neighbourhoods

2.2. Data gathering from Four Square API

Using four square API, the restaurant and entertainment locations within a 100 km radius of each neighbourhood was collected with their corresponding category ID. The restaurant and entertainment venues were then indicated on a map using folium library.

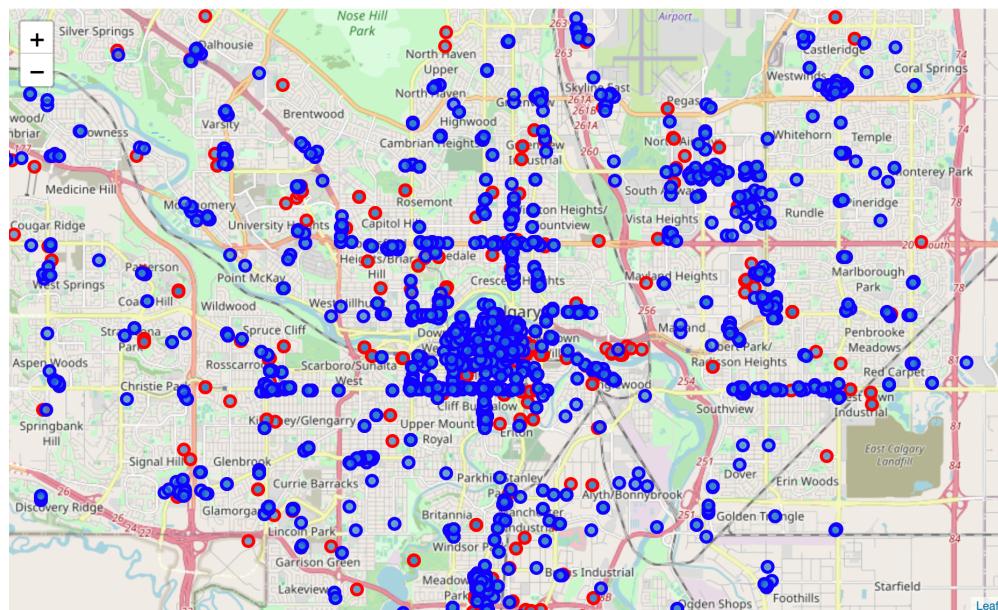


Figure 2: Map of Calgary showing Restaurants and Entertainment venues clusters

With entertainment venues represented in red, and restaurant venues in blue, we can observe that entertainment locations where people go to have fun, are clustered around restaurants, and vice versa. This proves the initial assumptions that there is a relationship between food vendors and entertainment locations. We can also observe from the data that the centre of the city has the most clusters of restaurants and entertainment venues also, which was also part of initial assumptions made. With further analysis, we should be able to narrow down the selection of the best neighbourhood to situate our new amusement park.

2.3. Grouping the data

The restaurant and entertainment location's data were grouped and counted to give the number of restaurant and entertainment venues for each neighbourhood and added to the main data frame. The new data frame consisted of 155 rows and 10 columns indicating that some Neighbourhood did not have restaurant and entertainment locations and were removed from the data frame.

Table 1: Data frame containing Neighbourhoods and the number of restaurants and entertainment venues

	Neighborhood	Quadrant	Sector	Type	Population	Area_size	Longitude	Latitude	food_count	enter_count
0	Calgary, Abbeydale	NE/SE	Northeast	Residential	5,917	1.7	-113.929413	51.058836	5	1
1	Calgary, Acadia	SE	South	Residential	10,705	3.9	-114.055587	50.968655	7	1
2	Calgary, Albert Park/Radisson Heights	SE	East	Residential	6,234	2.5	-113.990195	51.044845	19	1
3	Calgary, Alyth/Bonnybrook	SE	Centre	Industrial	16	3.8	-114.024294	51.016669	7	1
4	Calgary, Arbour Lake	NW	Northwest	Residential	10,836	4.4	-114.202355	51.136786	5	2

3. DATA ANALYSIS

3.1. Data Correlation

In order to analyze our data, we need to observe if there is a correlation between Population, Restaurant location, and Entertainment Location. This is carried out in order to validate the fact that Population, Restaurant location, and Entertainment Location for each Neighborhood is related.

Table 2: Data frame containing correlation information between columns

	Population	Area_size	food_count	enter_count
Population	1.000000	0.716886	-0.083242	-0.058660
Area_size	0.716886	1.000000	-0.291404	-0.250959
food_count	-0.083242	-0.291404	1.000000	0.676363
enter_count	-0.058660	-0.250959	0.676363	1.000000

From the table the correlation of each data can be observed, and our initial assumption that restaurant and entertainment location were correlated is validated. We went on to plot correlation graphs for our data.

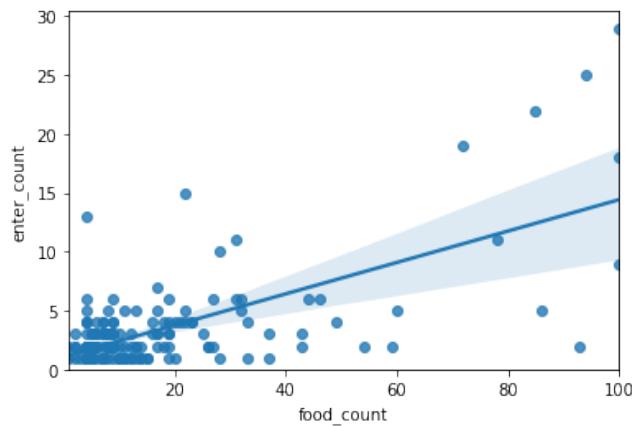


Figure 3: Graph showing correlation between number of entertainment and restaurant venues in a Neighbourhood

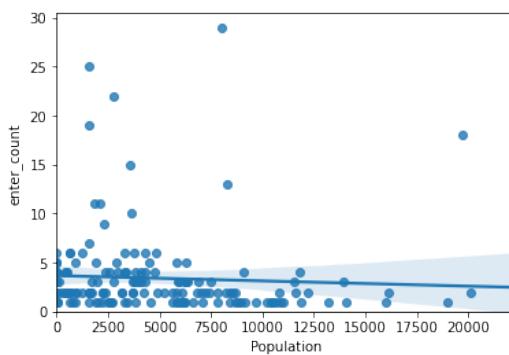


Figure 4: Graph showing correlation between number of entertainment venues and population of a Neighbourhood

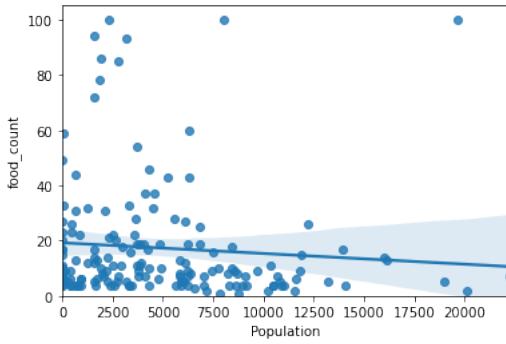


Figure 5: Graph showing correlation between number of restaurants and population of a Neighbourhood

3.2. Grouping the data by Sector

The second step in our analysis is to Narrow down the neighbourhood with the highest number of restaurants and entertainment venues. Luckily our data has information on the sector each Neighbourhood is located. So, what we can do is divide up our Neighbourhood via each sector and locate the sector with the best data qualities we desire.

Table 3: Table showing grouped data by sector with the Number of Neighbourhoods in each sector

Sector	neighborhood_count	
0	Centre	53
1	South	30
2	Northeast	18
3	West	18
4	Northwest	16
5	North	13
6	Southeast	10
7	East	7

Table 4: Table showing Sectors and the number of restaurant and entertainment venues

Sector	food_count	enter_count
0	Centre	1763
1	East	77
2	North	110
3	Northeast	283
4	Northwest	119
5	South	316
6	Southeast	71
7	West	133

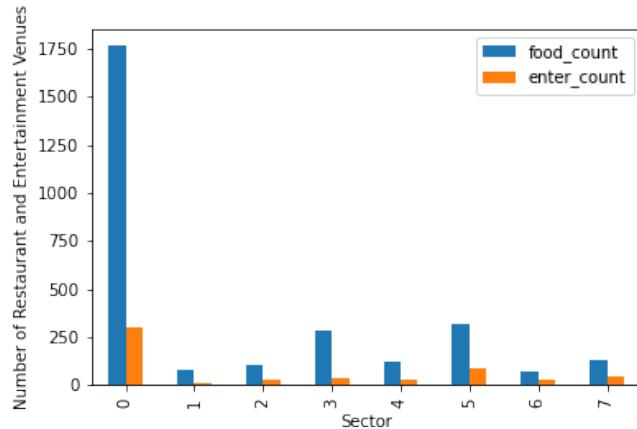


Figure 6: Bar Chart showing the Sectors and the number of restaurant and entertainment venues

From the grouped data, we can observe that the centre sector has the most neighbourhoods, restaurants and entertainment count, proving the initial assumption that entertainment venues are clustered towards the centre of the city.

3.3. Centre Sector Analysis

The neighbourhood in the center sector was then extracted into a new data frame for further analysis as shown in the table below;

Table 5: Table showing the new data frame containing Neighbourhoods in the centre sector

	Neighborhood	Quadrant	Sector	Type	Population	Area_size	Longitude	Latitude	food_count	enter_count
0	Calgary, Alyth/Bonnybrook	SE	Centre	Industrial	16	3	-114.024294	51.016669	7	1
1	Calgary, Banff Trail	NW	Centre	Residential	3837	1	-114.111647	51.071462	19	4
2	Calgary, Bankview	SW	Centre	Residential	5221	0	-114.099518	51.033887	43	2
3	Calgary, Bel-Aire	SW	Centre	Residential	424	0	-114.086811	50.999693	23	4
4	Calgary, Beltline	SW/SE	Centre	Residential	19681	2	-114.072593	51.040498	100	18

A correlation of the data was then carried out again to ensure that the restaurant and entertainment venues were indeed for the data

Table 6: Table showing correlation of data in the center sector

	Population	Area_size	food_count	enter_count
Population	1.000000	0.280217	0.392478	0.296457
Area_size	0.280217	1.000000	-0.217071	-0.101634
food_count	0.392478	-0.217071	1.000000	0.687017
enter_count	0.296457	-0.101634	0.687017	1.000000

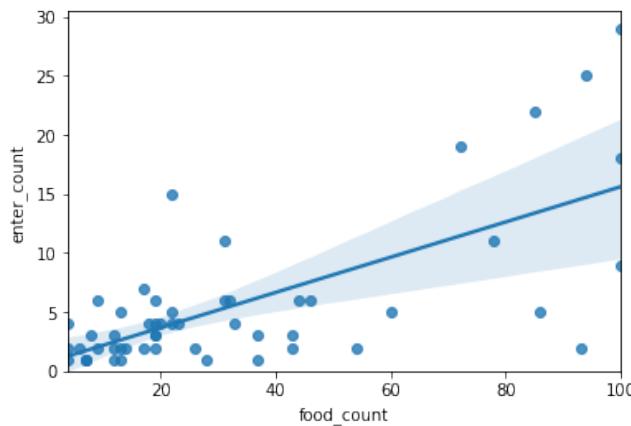


Figure 7: Graph showing correlation between number of entertainment and restaurant venues in a Neighbourhood of the centre sector

3.4. K-means Clustering

Using K-means clustering, the elbow method was used to observe the number of clusters in the Center sector data. From the graph below, the value of K was observed to be 3.

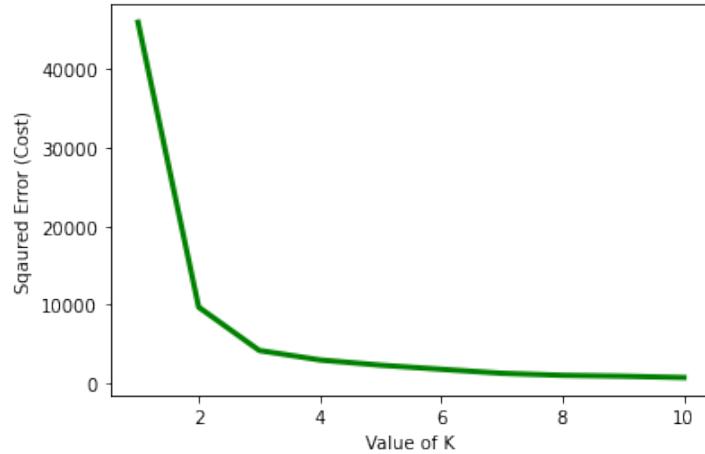


Figure 8: Graph of K-means cluster elbow method

After carrying out the elbow method to find the value of K, the K-means clustering method analysis was carried out, and the cluster of the data was plotted on a graph as shown below.

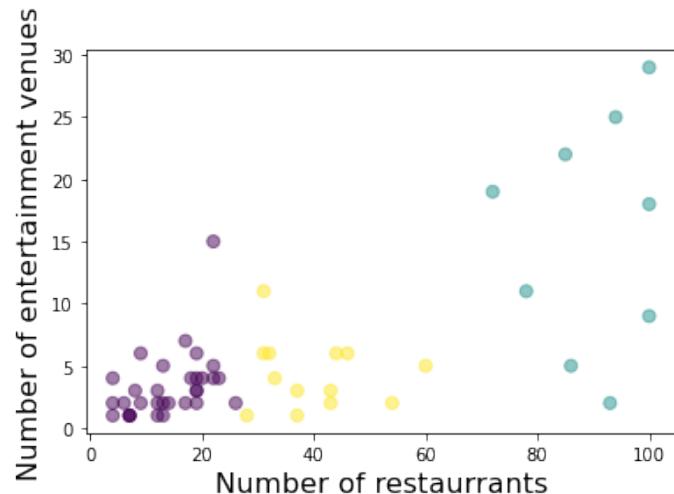


Figure 9: K-means Clusters

Each cluster was then divided up into data frames and plotted on a bar chart to observe the Neighbourhood with the most restaurant and entertainment venues.

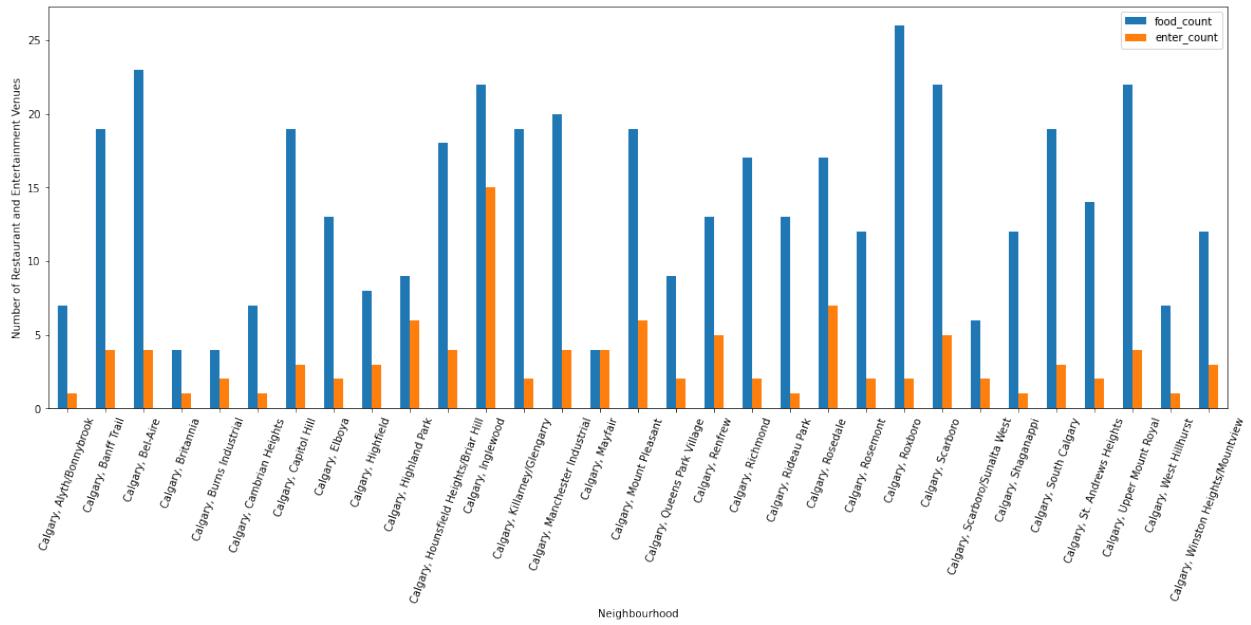


Figure 10: Cluster 0 Bar Chart

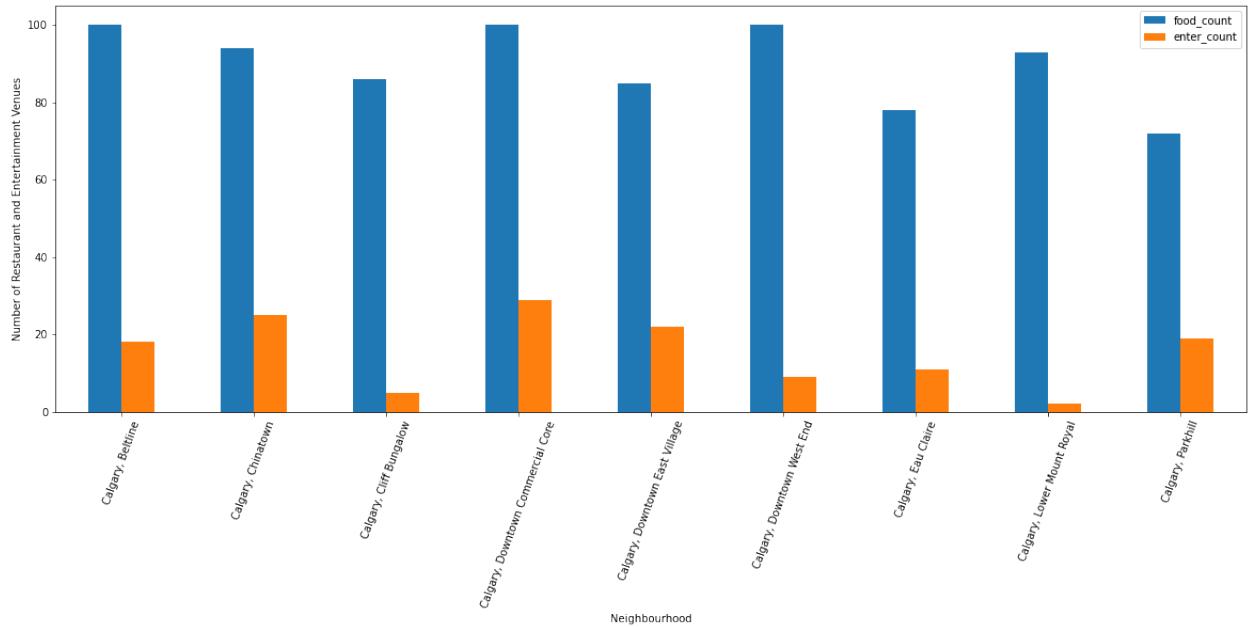


Figure 11: Cluster 1 Bar Chart

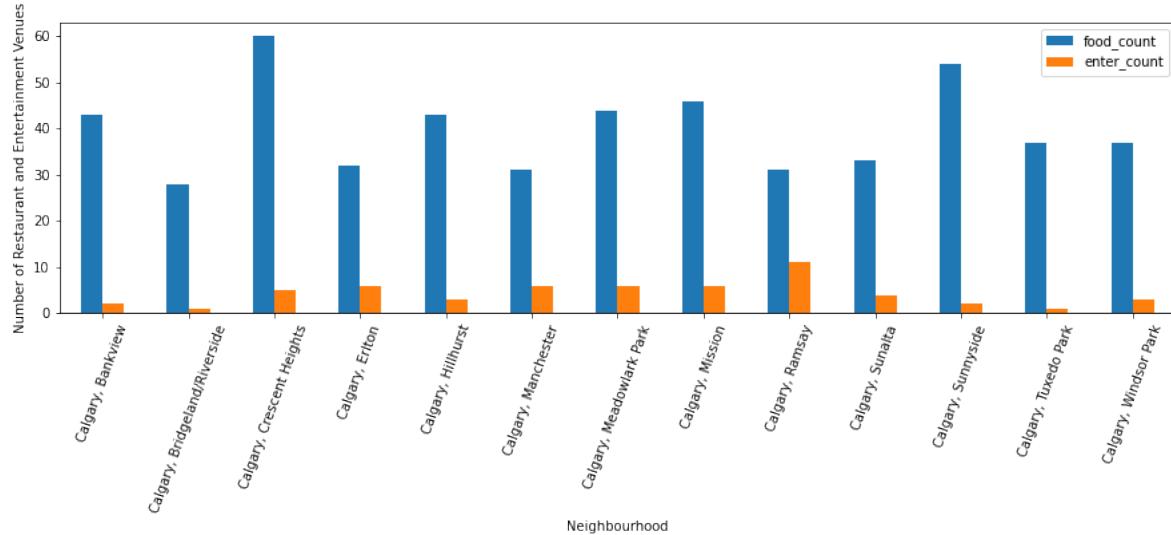


Figure 12: Cluster 2 Bar Chart

3.5. Result of analysis

As we can see, Although Cluster 1 and 2 look similar, Cluster 1 has a higher count of restaurant and entertainment location. From Cluster 1, we see that Calgary Downtown commercial core has the overall highest number of Restaurants and entertainment location.

3.6. Issue with analysis: Area size limitation

Table 7: Calgary Downtown commercial core Data row

	Neighborhood	Quadrant	Sector	Type	Population	Area_size	Longitude	Latitude	food_count	enter_count	Clus_km
13	Calgary, Downtown Commercial Core	SW/SE	Centre	Residential	8015	1	-114.067199	51.047378	100	29	1

As we can see, the area size of Calgary, Downtown Commercial Core is 1.8 km squared. From research carried out, size of Calgary, Downtown Commercial Core is just about 444 acres (1.8 km squared) which restricts the available amount of land to build an amusement park. The size of Calaway park is about 160 acres (0.647497 km squared), so we get a rough idea of how much land size we should look out for. Downtown areas of city are also known to contain high rise buildings used for office spaces etc. and is limited to the available space. Since Center Sector has the highest amount of Restaurant and

Entertainment venues, we expand our search for a location to build our amusement park back to the center sector. We can find the best location to build the amusement park by ranking the Centre Sector based on Area Size, Restaurant, and entertainment venues.

3.7. Ranking our data based on Area Size

Table 8: Ranked data frame based on Area Size

	Neighborhood	Quadrant	Sector	Type	Population	Area_size	Longitude	Latitude	food_count	enter_count	Clus_km
0	Calgary, Manchester Industrial	SE/SW	Centre	Industrial	0	4	-114.056644	51.011507	20	4	0
1	Calgary, Alyth/Bonnybrook	SE	Centre	Industrial	16	3	-114.024294	51.016669	7	1	0
2	Calgary, Bridgeland/Riverside	NE/SE	Centre	Residential	5594	3	-114.044310	51.053801	28	1	2
3	Calgary, Winston Heights/Mountview	NE	Centre	Residential	3891	3	-114.047588	51.072303	12	3	0
4	Calgary, Beltline	SW/SE	Centre	Residential	19681	2	-114.072593	51.040498	100	18	1
5	Calgary, Renfrew	NE	Centre	Residential	5846	2	-114.041687	51.062572	13	5	0
6	Calgary, Inglewood	SE	Centre	Residential	3564	2	-114.027489	51.038800	22	15	0
7	Calgary, Hillhurst	NW	Centre	Residential	6291	2	-114.095564	51.055414	43	3	2

Based on this data frame, we see the Largest Neighbourhood by area size is Manchester Industrial, but it is an industrial area with zero population. But looking at the data frame, we see Beltline which has the 5th Largest Neighbourhood by area size of 2km squared and a high number of restaurant and entertainment location. This can be a good Neighbourhood to build our new Amusement park. So, we rank our data based on area size, restaurant location and entertainment location.

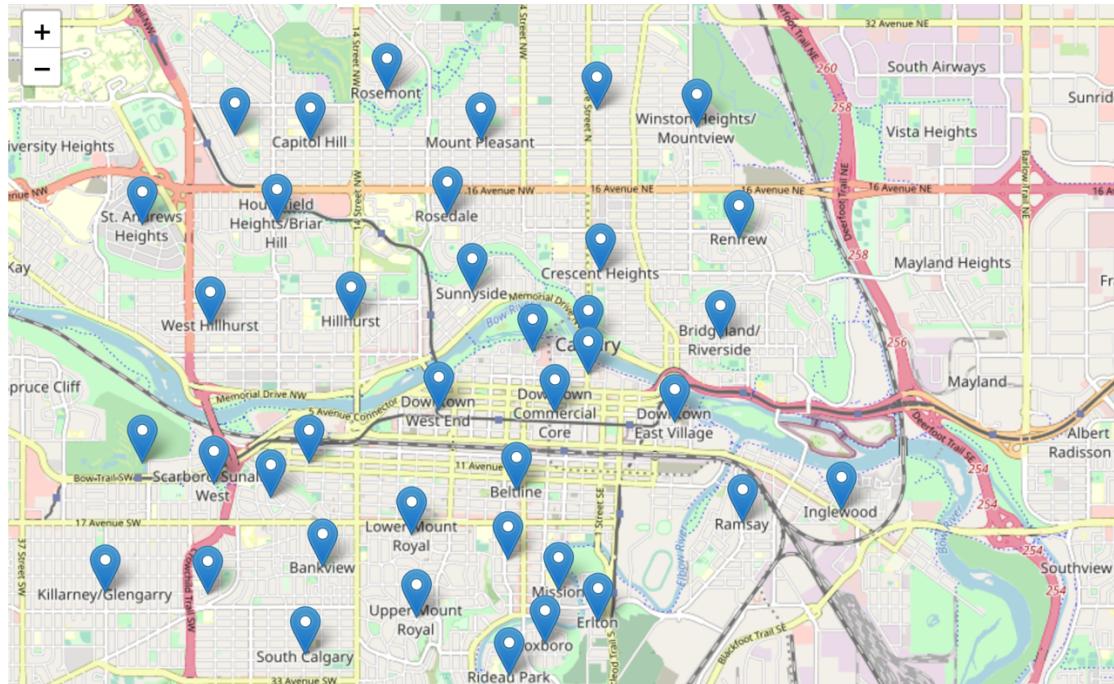
3.8. Ranking our data based on Area Size, number of restaurants, and number of entertainment venues

Table 9: Ranked data frame based on Area Size, number of restaurants, and number of entertainment venues

	Neighborhood	Quadrant	Sector	Type	Population	Area_size	Longitude	Latitude	food_count	enter_count	Clus_km	Rank
0	Calgary, Beltline	SW/SE	Centre	Residential	19681	2	-114.072593	51.040498	100	18	1	1
1	Calgary, Downtown Commercial Core	SW/SE	Centre	Residential	8015	1	-114.067199	51.047378	100	29	1	2
2	Calgary, Inglewood	SE	Centre	Residential	3564	2	-114.027489	51.038800	22	15	0	3
3	Calgary, Ramsay	SE	Centre	Residential	2110	1	-114.041219	51.037727	31	11	2	4
4	Calgary, Chinatown	SW/SE	Centre	Residential	1577	0	-114.062611	51.050654	94	25	1	5
5	Calgary, Downtown East Village	SE	Centre	Residential	2782	0	-114.050643	51.046496	85	22	1	6
6	Calgary, Eau Claire	SW	Centre	Residential	1851	0	-114.070312	51.052579	78	11	1	7
7	Calgary, Parkhill	SW	Centre	Residential	1592	0	-114.062589	51.053423	72	19	1	8

From our ranked data set, according to area size, restaurant and entertainment location, The Beltline Neighbourhood is the best location to build our new amusement park. The best Neighbourhood is therefore **Beltline Neighbourhood**.

Figure 13: Map of Calgary showing the centre sector indicating the Beltline Neighbourhood position



4. METHODOLOGY

To complete the analysis, there were several steps of data gathering, cleaning, and analysis. Data was pulled from Wikipedia and the Foursquare API, cleaned using the padas and Beautiful Soup libraries, and stored into panda's data frames. Data was assembled to align latitude and longitude coordinates with neighborhood names in Calgary, and that was mapped using the folium library.

The Foursquare API was used to find the (up to) 100 closest restaurants and entertainment locations to each neighborhood center and their locations. The data pertaining to the number of restaurant and entertainment locations for each Neighbourhood was stored in a data frame, providing us with useful information about the number of restaurant and entertainment locations for each Neighbourhood. The new data frame was grouped by sector

(North, South, East, West, Northeast, Northwest, Southwest, and center), and the sector with the highest number of restaurant and entertainment location was selected. The Neighbourhoods in Center Sector was selected for further analysis. In order to find out the factors that are correlated to the location of entertainment venues (because an amusement park is a source of entertainment and fun), correlation of the data was carried out for population, area size, restaurant count, and entertainment venues count. A positive correlation was found between the number of restaurant and entertainment venues in a neighbourhood.

The Neighbourhood in the center sector was then clustered using K-means Clustering. The 'elbow method' was used to determine the number of clusters used. The clustered data was then divided and analyzed to identify the neighbourhood with the highest number of restaurant and entertainment locations. Downtown commercial core was selected as the best Neighbourhood. But unfortunately, Downtown is a small Neighbourhood and would not be the best candidate for the construction of a large amusement park. Also, downtown areas are known to be the commercial hub of every city and is usually filled with high rise buildings and has limited land availability. So, the Neighbourhood in the Center Sector was then ranked based on area size, restaurant count, and entertainment venues count. Based on the ranking of the data, Beltline having an area size of 2km squared was selected as the best Neighbourhood.

5. RESULTS

Comparing our analysis to the questions initially asked;

- From available data, is there a relationship between population and entertainment venues in various neighbourhoods

Answer: No correlation exists between population and entertainment venues in various neighbourhoods

- What is the relationship between current entertainment venues and number of restaurants in various neighbourhoods?

Answer: There is a positive correlation between entertainment venues and number of restaurants in a Neighbourhood

- Are entertainment venues closer to the city center?

Answer: Yes, entertainment venues are more clustered towards the center of the city

- Are entertainment venues located closer to more populated areas or less populated areas

Answer: There is no correlation between entertainment venues and population of a Neighbourhood.

Based on initial analysis, downtown commercial core was selected as the best Neighbourhood. But unfortunately, Downtown is a small Neighbourhood and would not be the best candidate for the construction of a large amusement park based on its area size. Based on the ranking of the data (area size, restaurant count, and entertainment venues count), Beltline having an area size of 2km squared was selected as the best Neighbourhood.

6. DISCUSSION

Based on the analysis, the Beltline Neighbourhood was selected as the best Neighbourhood. Although this analysis indicates it as being the best Neighbourhood, further research would need to be carried out along with the City of Calgary, in order to determine if there is a city plan for that Neighbourhood, and if sufficient land exists for the construction of an amusement park. Several enquiries and permits would also be needed in order to determine if potential investor could move forward on determining the best Neighbourhood to build the new amusement park.

7. CONCLUSION

Calgary is the 4th largest city in Canada by population, with the province of Alberta having the third largest GDP in Canada. It is a beautiful city and building an amusement park can be a very profitable business to invest in. A big amusement park with a lot of rides, and fun activities like Disney World, could be a source of tourist attraction, which could potentially increase the economy of Calgary.