Opening a new shopping center in Auckland, New Zealand

IBM Applied Data Science Capstone

SAMIRA YADEGARI

DECEMBER 2020

1. Introduction

A shopping center is a group of retail stores such as grocery stores, restaurants, fashion outlets, cinemas, cafes, and children's playgrounds. It usually has ample parking facilities and designed to serve a community or neighborhood. Shopping malls provide a great distribution channel for retailers to market their products and services, so it helps them to improve in their business. Also, Property developers are taking advantage of this trend to build more shopping centers to answer to the demands.

Create a new shopping center can increase comfort for people living around and attract people from other areas to build their house near to the new shopping mall. In addition, it can increase property value and has an effect on property management.

Auckland is the biggest city in New Zealand and has special attributes: Auckland is a very vast city, has high traffic jam and because it is the main business city in New Zealand have an annual population increase.

Opening a new shopping center is a complicated decision and depends on various variables such as the neighborhood population, distance to other malls, the future of other industries in the area and etc. Therefore, making a good decision about the place of the new mall has a straight impact on successfulness of that.

2. Business Problem

The objective of this capstone project is to analyse the Neighborhoods' data and select the best locations in Auckland city to open a new shopping center. So the business question is:

- If the city authorities are looking to open a new shopping center, where is the best place for that?

In this project, I will use data science and machine learning techniques such as Clustering to provide a possible answer to the business question.

3. Target Audience:

This project is particularly useful for Auckland authorities to decide on the future of the city and how to develop it. As I said before, Auckland has special features such as its bigness, traffic jam, high rate of immigrants, and etc.

In addition, this project is useful for property developers and investors looking to open or invest in new shopping centers in a vast city such as Auckland.

4. Data resources:

To answer the business question, we will need some special piece of information about Auckland neighborhoods, their geographical coordinates, and venues of them:

1. List of neighborhoods in Auckland, New Zealand is fetched from Wikipedia page (https://en.wikipedia.org/wiki/Demographics_of_Auckland). The extracted table had a few extra columns to remove and change to make a table with neighborhood and population columns only. So, in this section, we used python codes to extract and clean data.

2. For achieving geographical (Latitude and longitude) coordinates of the Auckland neighborhoods we used the Python Geocoder package. This is required in order to plot the map and also to get the venue data (Figure 1).

	Neighborhood	Population	Latitude	Longitude
0	Rodney	73932	42.205546	-95.952511
1	Hibiscus and Bays	112380	-36.646629	174.783579
2	Upper Harbour	67137	-36.764037	174.675784
3	Kaipatiki	97593	-36.794515	174.716734
4	Devonport-Takapuna	62193	-36.791967	174.787492
5	Henderson-Massey	139749	-36.848997	174.629905
6	Waitākere Ranges	60390	-36.982943	174.449241
7	Great Barrier	1104	-1.696481	29.245338
8	Waiheke	10221	-36.793097	175.088401
9	Waitemata	89880	-36.821227	174.669807
10	Whau	89727	-36.907050	174.692776
11	Albert-Eden	108819	-36.870200	174.731366
12	Puketapapa	62532	-36.942041	174.723668
13	Orakei	90960	-36.870288	174.814415
14	Maungakiekie-Tamaki	87396	-36.902262	174.845038
15	Howick	151947	45.184903	-73.849060
16	Mangere-Otahuhu	90387	-36.983504	174.761372
17	Otara-Papatoetoe	97992	-36.992209	174.859371
18	Manurewa	114447	-37.024150	174.888660
19	Papakura	68496	-37.059960	174.944542
20	Franklin	84504	33.160119	-95.216521

Figure 1. Coordinates of the Auckland neighborhoods

3. Venue data, particularly data related to shopping malls. We will use Foursquare API to get the venue data for those neighborhoods and will use this data to perform clustering on the neighborhoods (Figure 2).

	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Rodney	42.205546	-95.952511	United States Postal Service	42.203652	-95.950499	Post Office
1	Rodney	42.205546	-95.952511	Hamann Trucking	42.200090	-95.966940	Moving Target
2	Upper Harbour	-36.764037	174.675784	Malthouse	-36.775521	174.670324	Pub
3	Upper Harbour	-36.764037	174.675784	Wainoni Park	-36.771903	174.675557	Park
4	Upper Harbour	-36.764037	174.675784	Palette	-36.774340	174.671142	Café

Figure 2. Auckland neighborhood venues list

In this project, we will apply many data science skills such as web scraping (Wikipedia), data wrangling, working with API (Foursquare), machine learning (K-means clustering), and map visualization (Folium).

4. Methodology:

After collecting and cleaning data, we need to do some analysis on it by using the Foursquare API. For example, we can print out the list of categories, and see it clearly contains 'Shopping Mall'.

```
array(['Post Office', 'Moving Target', 'Pub', 'Park', 'Café',
       'Convenience Store', 'Liquor Store', 'Golf Course', 'Burger Joint',
       'Gym', 'Night Market', 'Fast Food Restaurant', 'Gastropub',
       'Supermarket', 'Shopping Mall', 'Pizza Place', 'Department Store',
       'Ramen Restaurant', 'Bus Stop', 'Indian Restaurant', 'Hotel',
       'Pet Store', 'Gymnastics Gym', 'Grocery Store', 'Soccer Field',
       'Portuguese Restaurant', 'Beach', 'Ice Cream Shop', 'Restaurant',
       'Mexican Restaurant', 'Vietnamese Restaurant',
       'Frozen Yogurt Shop', 'Malay Restaurant', 'Japanese Restaurant',
       'Music Venue', 'Mediterranean Restaurant', 'Farmers Market', 'Bar',
       'Thai Restaurant', 'Colombian Restaurant', 'Boutique',
       'Italian Restaurant', 'Coffee Shop', 'Movie Theater',
       'Middle Eastern Restaurant', 'French Restaurant', 'Salad Place',
       'Sandwich Place', 'Chinese Restaurant', 'Bistro'], dtype=object)
```

Figure 3. Auckland neighborhood categories list

Or we can show the most common venues for each Neighborhood:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	0.0	Café	Bar	Bakery	Park	Chinese Restaurant	Pizza Place	Burger Joint	Gym	History Museum	Coffee Shop
1	0.0	Café	Japanese Restaurant	Restaurant	Italian Restaurant	Bar	Malay Restaurant	Burger Joint	Fast Food Restaurant	Ice Cream Shop	Beach
2	0.0	Fast Food Restaurant	American Restaurant	Convenience Store	Taco Place	Mexican Restaurant	Ice Cream Shop	BBQ Joint	Grocery Store	Truck Stop	Fried Chicken Joint
3	0.0	Hotel	Border Crossing	Bistro	Café	Lake	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Electronics Store
4	0.0	Café	Fast Food Restaurant	Asian Restaurant	Convenience Store	Supermarket	Park	Burger Joint	Chinese Restaurant	Coffee Shop	Pet Store

Figure 4. The most common venues for each Neighborhood

And also, because this project is about shopping malls, we can extract the number of shopping malls in different neighborhood:

	Number of malls
Neighborhood	
Albert-Eden	9
Devonport-Takapuna	2
Henderson-Massey	1
Kaipatiki	1
Maungakiekie-Tamaki	3

Figure 5. Number of shopping malls

Finally, we will perform clustering on the data by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project.

We used the Elbow method is used for identifying the best number of clusters and it showed K=4 is the best for our clustering.

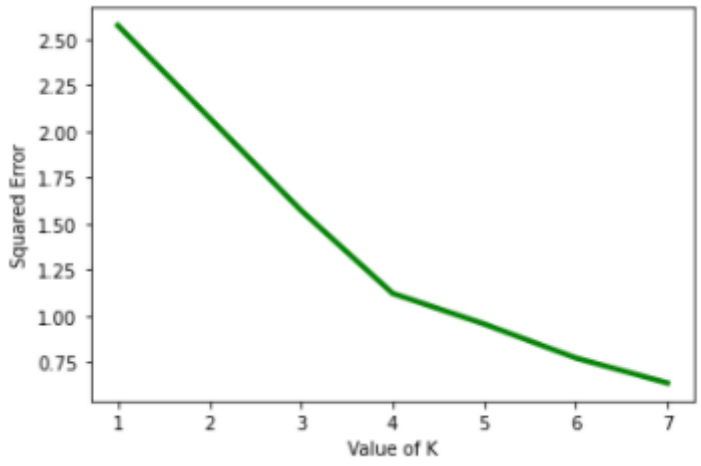

Figure 6. Elbow method is used for identifying the best number of cluster

The results will allow us to identify which neighborhoods have a higher concentration of shopping malls while which ones have less shopping malls. Based on the occurrence of shopping malls in different neighborhoods, it will help us to answer the question as to which neighborhoods are most suitable to open new shopping malls.

	Neighborhood	Shopping Mall	Cluster Labels	Population	Latitude	Longitude
12	Papakura	0.025641	0	68496	-37.059960	174.944542
11	Otara-Papatoetoe	0.031250	0	97992	-36.992209	174.859371
10	Orakei	0.027027	0	90960	-36.870288	174.814415
6	Kaipatiki	0.030303	0	97593	-36.794515	174.716734
13	Puketapapa	0.166667	1	62532	-36.942041	174.723668
8	Manurewa	0.000000	2	114447	-37.024150	174.888660
14	Rodney	0.000000	2	73932	42.205546	-95.952511
17	Waitemata	0.000000	2	89880	-36.821227	174.669807
16	Waiheke	0.000000	2	10221	-36.793097	175.088401
7	Mangere-Otahuhu	0.000000	2	90387	-36.983504	174.761372
5	Howick	0.000000	2	151947	45.184903	-73.849060
4	Henderson-Massey	0.000000	2	139749	-36.848997	174.629905
3	Great Barrier	0.000000	2	1104	-1.696481	29.245338
2	Franklin	0.000000	2	84504	33.160119	-95.216521
1	Devonport-Takapuna	0.000000	2	62193	-36.791967	174.787492
15	Upper Harbour	0.000000	2	67137	-36.764037	174.675784
0	Albert-Eden	0.010000	3	108819	-36.870200	174.731366
9	Maungakiekie-Tamaki	0.017241	3	87396	-36.902262	174.845038
18	Whau	0.017857	3	89727	-36.907050	174.692776

Figure 7. Clustering neighborhoods based on shopping malls numbers

The results from the k-means clustering show that we can categorize the neighborhoods into 4 clusters based on the frequency of occurrence for “Shopping Mall”:

- Cluster 0: Neighborhoods with a moderate number of shopping malls
- Cluster 1: Neighborhoods with a high concentration of shopping malls
- Cluster 2: Neighborhoods without shopping malls
- Cluster 3: Neighborhoods with less than a moderate number of shopping malls

The results of the clustering are visualized in the map below with cluster 0 in red color, cluster 1 in purple color, cluster 2 in mint green color, and cluster 3 in yellow color.

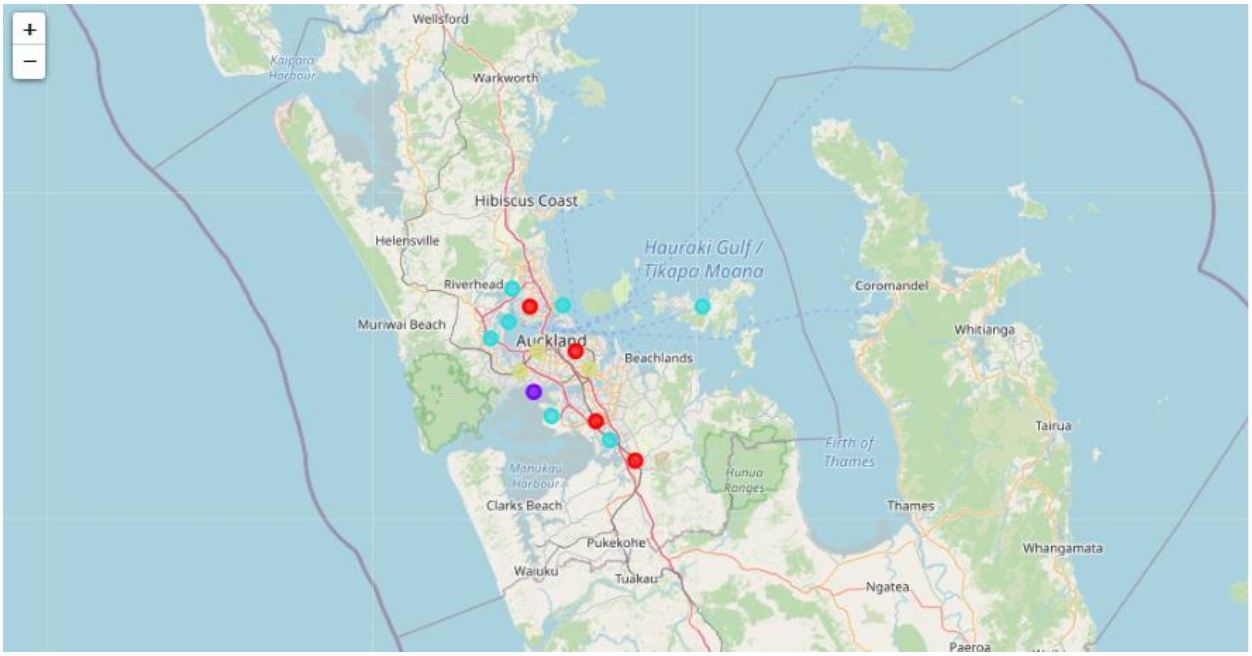

Figure 8. Neighborhoods clustering

5. Discussion

As observations noted from the map in the Results section, most of the shopping malls are concentrated in cluster 1 and a moderate number in cluster 0.

On the other hand, cluster 3 has a very low number of no shopping mall in the neighborhoods. This represents a great opportunity and high potential areas to open new shopping malls as it is very little to no competition from existing malls. Meanwhile, based on provided information by the Foursquare API in cluster 2 is suffering from neighborhoods without shopping malls.

In addition, we have two bar plots that give us a very good perspective of the neighborhood's population and the number of shopping malls in Auckland. It obvious that many neighborhoods need to have shopping malls for the comfort of their people. Property developers can use this information and start investing in different neighborhoods and of course, cluster 2 is the best place to start.

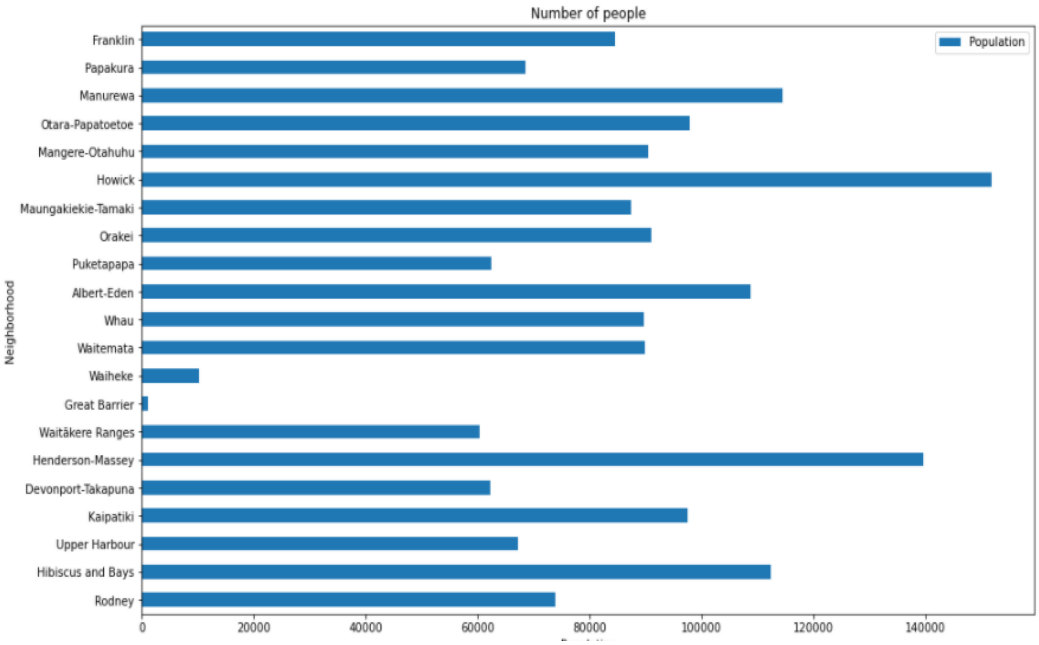

Figure 9. Neighborhoods and population

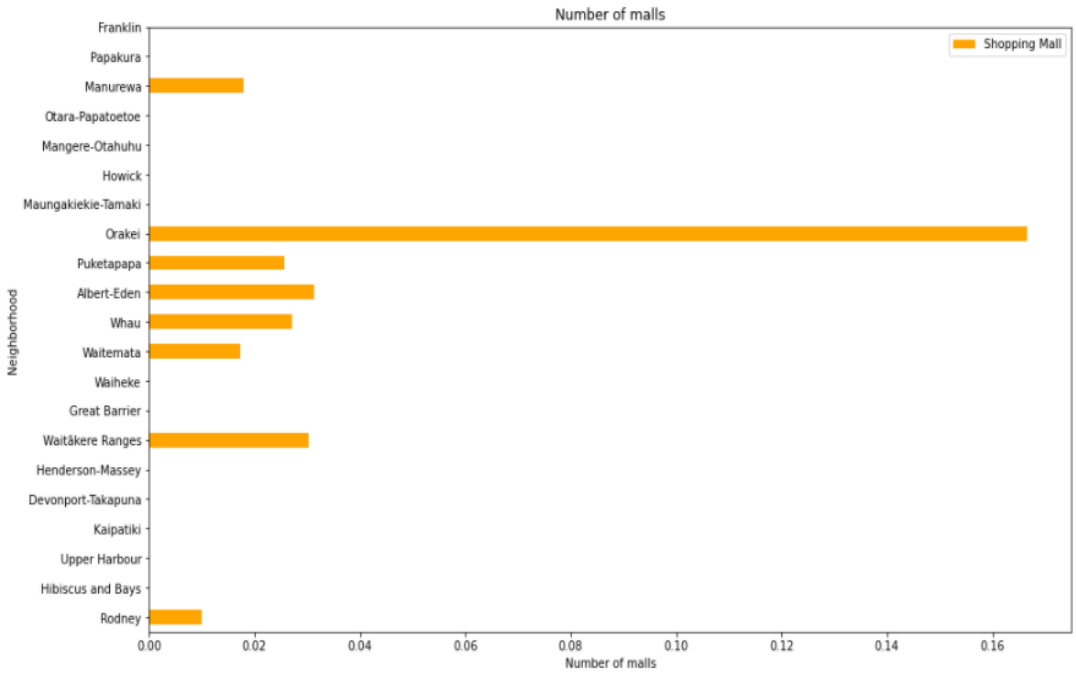

Figure 10. Neighborhoods and shopping mall numbers

6. Limitations and Suggestions for Future Research

In this project, we only used the Foursquare API for extracting data pf shopping malls and maybe it doesn't contain all the related information. So, using other data providers such as google maybe can improve the results. In addition, more work on the relationship between population and number of the shopping mall can give a good idea about the place of the new mall.

7. Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 4 clusters based on their similarities, and lastly providing recommendations to the relevant stakeholders i.e. property developers and

investors regarding the best locations to open a new shopping mall. To answer the business question that was raised in the introduction section, the answer proposed by this project is:

The neighborhoods in cluster 2 are the most preferred locations to open a new shopping mall, although looking at the population of those neighborhoods and their distance are two important factors to consider too.

The findings of this project will help the relevant stakeholders to capitalize on the opportunities in high potential locations while avoiding overcrowded areas in their decisions to open a new shopping mall.

The notebook of this project is on:

https://github.com/Samira000/Coursera_Capstone/blob/7adebc352e1f1041990f9da33e93e40afde6774f/https:/github.com/Samira000/Auckland-New-Shopping-Mall/Auckland_Shopping_Center.ipynb