

IBM DATASCIENCE
PROFESSIONAL CERTIFICATE

Rent and Price of Houses in Auckland
CAPSTONE PROJECT



Asha Rajagopal
May 2020

1. Introduction

Houses are a dream for every human and one of his basic essentials. When you move in to a new place, the first thing that you look for is where to rent/buy a house within the limited budget. Houses are a big part of real estate business. And the prime reason for the boom in real estate is because houses are often considered as an investment. You could either rent a house, buy a house and rent it out later or sell it on a profit. Several factors are considered when buying/renting a house apart from the condition, budget and layout of the house. For example, one looks for the proximity to the amenities like supermarkets, schools, public transport, parks, shops etc. This project aims to get an understanding on the relationship between rent/house price in Auckland region of New Zealand and the most popular amenities nearby and thus choose an apt location to rent/buy a house.

1.1 Business Problem

This project uses the Data Science methodology and Machine Learning techniques like Clustering and Segmentation to analyse the average rent or price of a house (3-bedroom) in the suburbs of Auckland (The study is confined to the suburbs of Auckland and 3 bedroom houses). The business problem we are trying to solve is: Where would be an ideal location to rent/buy a house in Auckland?

1.2 Target Audience

The target audience of this study would be people looking for renting a house in Auckland as well as those who look forward to buy a house. This project would give them an idea on where would be an apt location for their house according to their budget and interests. The study would be useful for those who would like to know about the relationship between the most popular amenities/venues and the rent of house.

2. Data

For the study, we will need the following data:

- a) The list of suburbs of Auckland region
- b) The latest (for the year 2020) average rent and house price for 3 bedroom houses in Auckland region
- c) The latitude and longitude data for each suburb
- d) The most popular venues of each suburb

2.1 Data Source and Extraction Methodology

- a) List of suburbs of Auckland: The area of study is restricted to Auckland region in New Zealand. So we need to collect the list of suburbs of Auckland either from the Wikipedia page https://en.wikipedia.org/wiki/List_of_suburbs_of_Auckland or from the web page of Barfoot & Thompson, the real estate agency in New Zealand. For example, the Wikipedia lists the suburbs in Auckland by categorizing them to eastern, western, north, south and central suburbs. Data scraping techniques (BeautifulSoup or Pandas package in Python) will be used to acquire the required data.
- b) The average rent or house price: We need to get the average rent per week or price of house in a suburb in 2020, to identify the apt location to rent or buy a house. This data can be obtained from the real estate agency web page, <https://www.barfoot.co.nz/market-reports/2020/january/suburb-report>. Data scraping using the BeautifulSoup package in Python will be used to extract the data from the page. The above link contains both the list of suburbs and the rent or house price for each suburb. For example, each row in the table of the above webpage has the suburb name, average rent per week for a 3- bedroom house, average price for 3 bedroom house and gross yield. We will not need the gross yield field for our study.
- c) The latitude and longitude data: In order to plot the map of Auckland and visualise the suburbs and clusters in Auckland, we need the latitude and longitude data. The geo coordinates are also used for retrieving the information on venues of a suburb. The geocoder package of Python will be used to get the geospatial information. In case if the geocoder package doesn't work, the latitude and longitude data will be scraped from the web page, <https://www.geonames.org/postalcode-search.html?q=Auckland&country=NZ>
- d) The most popular venues: The data regarding the most popular venues will be extracted using the FourSquare API. The FourSquare API returns the popular venues in a suburb or location based on the foot traffic. An example query to FourSquare API to get the venues in a location would like, [https://api.foursquare.com/v2/venues/explore?client_id={}&client_secret={}&ll={},{}&v={}&radius={}&limit={}'.format\(CLIENT_ID, CLIENT_SECRET, neighborhood_latitude, neighborhood_longitude, VERSION, radius, LIMIT\)](https://api.foursquare.com/v2/venues/explore?client_id={}&client_secret={}&ll={},{}&v={}&radius={}&limit={}'.format(CLIENT_ID, CLIENT_SECRET, neighborhood_latitude, neighborhood_longitude, VERSION, radius, LIMIT))
The developer console of FourSquare API lets us get venues or other details by passing the latitude and longitude of the location (suburb in our case). It will then be converted to a dataframe and used in our machine learning algorithm. This data would help us to understand the amenities in a suburb.

3. Methodology

3.1 Web scraping and Data wrangling

To begin with, we need to web scrape the web page

<https://www.barfoot.co.nz/market-reports/2020/january/suburb-report>.

This is done using the BeautifulSoup package in Python. The web page contains suburbs, average sale price for a three bedroom house, average rent per week for a three bedroom house and gross yield data categorized into different areas. The data is in tabular form. So at first, the table data is scraped and converted to a dataframe, then a new column 'Area' is added to the dataframes. The multiple dataframes are then combined into a single dataframe and 'gross yield' column is dropped/removed since it is not required in our study. The snapshot of dataframe is shown in the image below.

	Area	Suburb	Average sale price (3 bdrm)	Average rent per week (3 bdrm)
0	Auckland Central	Auckland Central	\$987,500	\$1,055
1	Central	Avondale	\$855,105	\$554
2	Central	Blockhouse Bay	\$930,663	\$563
3	Central	Epsom	\$1,601,250	\$773
4	Central	Freemans Bay	\$1,702,167	\$881

The above dataframe is further cleansed by renaming the columns, changing the data type of sale price and rent column to integer. The final dataframe looks like the one below.

```
Area          object
Suburb        object
Average sale price  int64
Average rent per week  int64
dtype: object
```

	Area	Suburb	Average sale price	Average rent per week
0	Auckland Central	Auckland Central	987500	1055
1	Central	Avondale	855105	554
2	Central	Blockhouse Bay	930663	563
3	Central	Epsom	1601250	773
4	Central	Freemans Bay	1702167	881

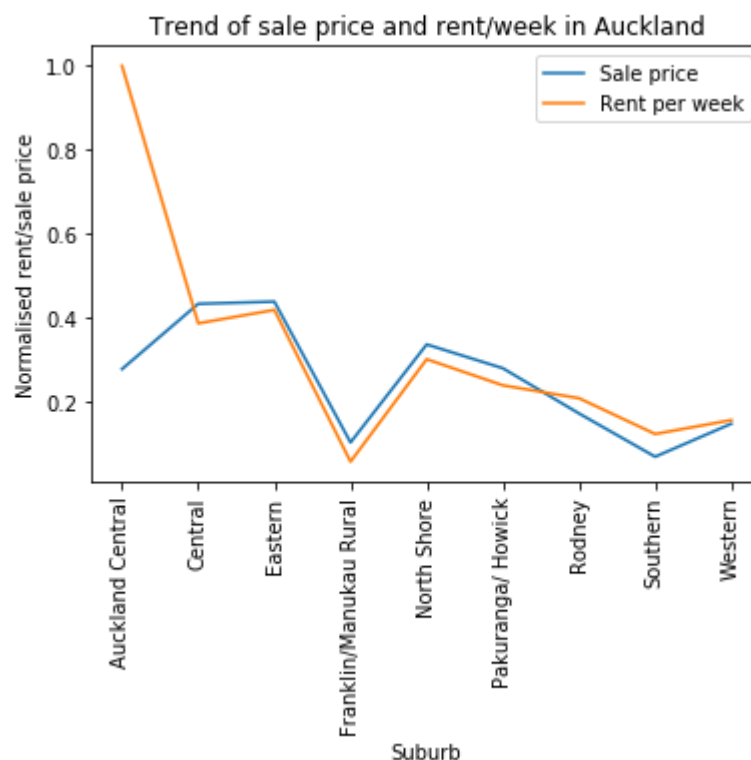
In order to find the nearby venues using Foursquare API, we need the latitude and longitude information for each suburb. The geo coordinates are retrieved using geocoder package and then added as columns to the above clean dataframe. The final dataset to work with is

	Area	Suburb	Average sale price	Average rent per week	Latitude	Longitude
0	Auckland Central	Auckland Central	987500	1055	-36.848399	174.764388
1	Central	Avondale	855105	554	-36.890448	174.687017
2	Central	Blockhouse Bay	930663	563	-36.928200	174.700190
3	Central	Epsom	1601250	773	-36.888454	174.772938
4	Central	Freemans Bay	1702167	881	-36.852879	174.750353

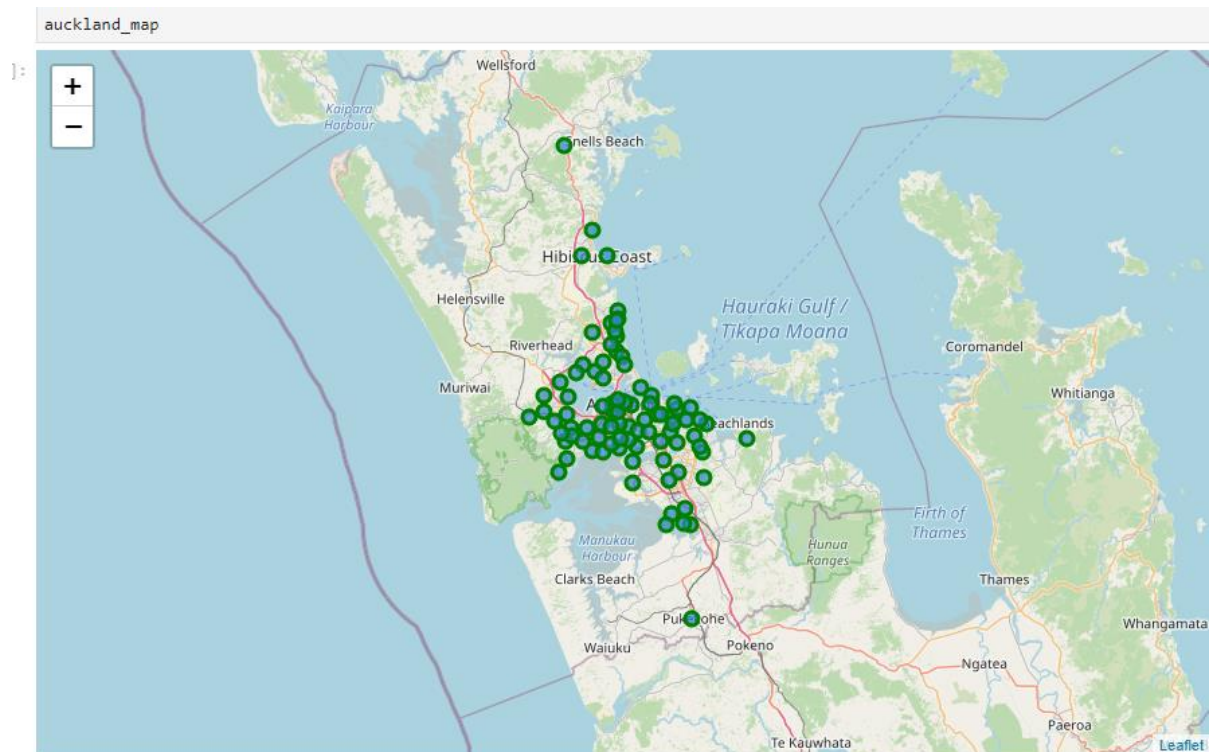
Also, a temporary dataset is also built with just Average sale price, Average rent per week which is in turn grouped based on Area. The sale price and rent columns are normalised as well since there will be a huge difference between rent and sale price otherwise and will be difficult to visualize or analyze.

3.2 Data Visualization

Once the dataset is cleansed and ready to work on, the data is visualized to understand the trends and how they are distributed. At first, the normalised rent and sale price data is visualized. The graph shows a high correlation between rent and sale price except for the Auckland Central area.



Now, we will visualise the Auckland suburbs on a map to get an idea about how they are distributed geographically. The Folium package is used to create maps in python.



The Foursquare API is used to retrieve the venues in each suburb. The retrieved data is then converted to a dataset with list of venues, their category, their location, geo coordinates, suburb and area they are in. This dataset then undergoes one-hot encoding. This helps us to get the frequency of a venue category in each suburb. And we create a new dataset with the top five venues in each suburb within a radius of 1500 meters. The dataset looks as below.

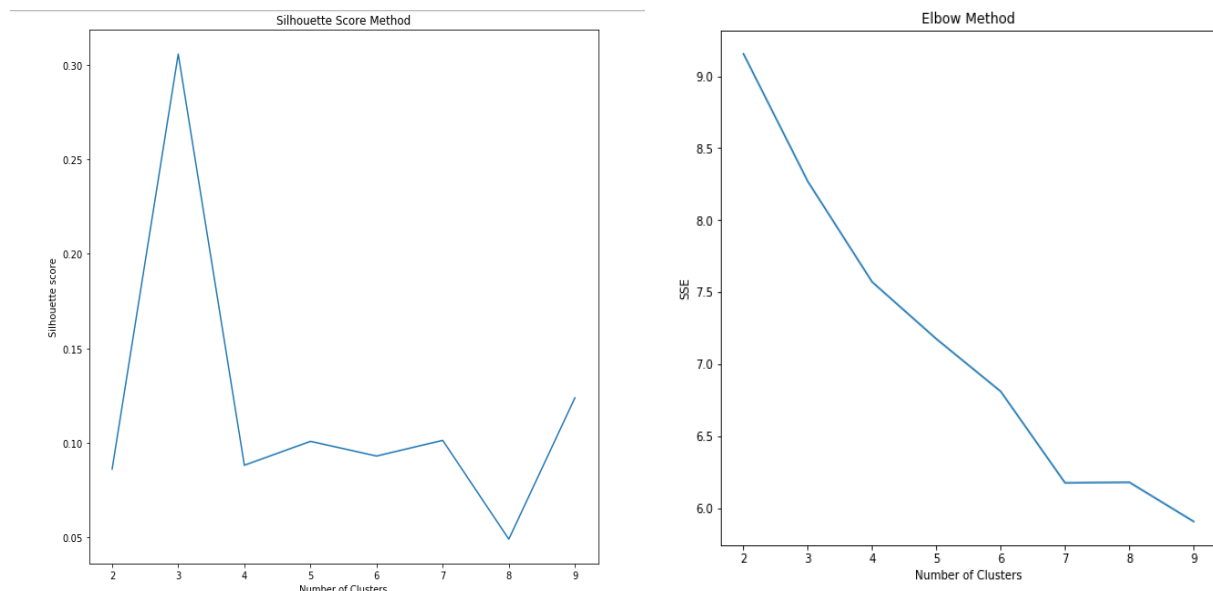
	Suburb	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Albany	Café	Bar	Sporting Goods Shop	Fast Food Restaurant	Department Store
1	Auckland Central	Café	Hotel	Bar	Park	Coffee Shop
2	Avondale	Café	Chinese Restaurant	Market	Mongolian Restaurant	Neighborhood
3	Beach Haven	Park	Pier	Trail	Chinese Restaurant	Farmers Market
4	Beachlands	Trail	Coffee Shop	Resort	Tennis Court	Eastern European Restaurant

3.3 Data Analysis using K-Means Clustering method

The K-Means Clustering Machine Learning Algorithm is used to cluster and segment the Auckland suburbs based on the most common venues nearby. The dataset on which the K-Means model is applied is shown below (which is obtained after performing one-hot encoding technique and grouping data based on suburbs). K-Means algorithm is chosen to create tighter clusters based on the features.

	American Restaurant	Aquarium	Arcade	Argentinian Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Australian Restaurant	BBQ Joint	...	Video Store	Vietnamese Restaurant	Water Park	Wat
0	0.017857	0.0	0.0	0.0	0.00	0.017857	0.035714	0.0	0.0	0.017857	...	0.0	0.0	0.0	
1	0.000000	0.0	0.0	0.0	0.01	0.000000	0.000000	0.0	0.0	0.000000	...	0.0	0.0	0.0	
2	0.000000	0.0	0.0	0.0	0.00	0.000000	0.000000	0.0	0.0	0.000000	...	0.0	0.0	0.0	
3	0.000000	0.0	0.0	0.0	0.00	0.000000	0.000000	0.0	0.0	0.000000	...	0.0	0.0	0.0	
4	0.000000	0.0	0.0	0.0	0.00	0.000000	0.000000	0.0	0.0	0.000000	...	0.0	0.0	0.0	

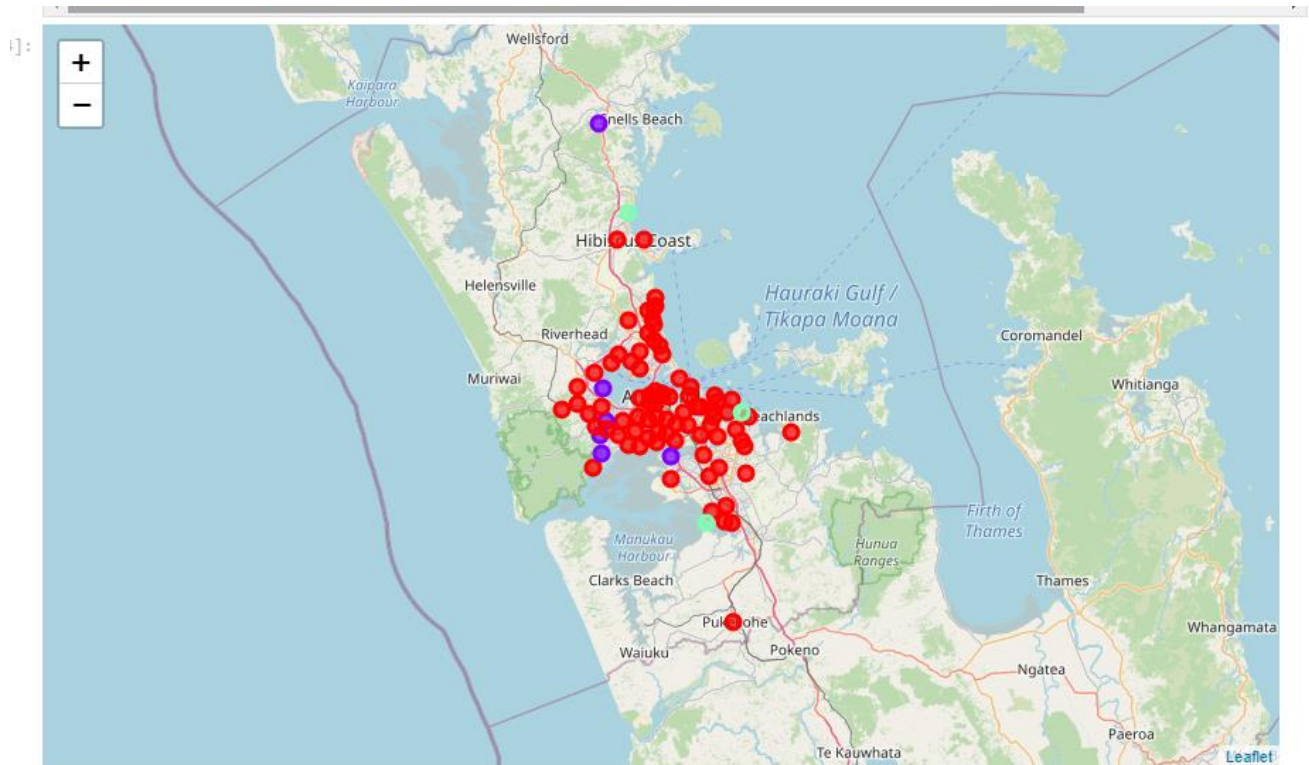
The optimum number of clusters or 'k' in K-Means is found out using Elbow Method and Silhouette score. The K- Means algorithm is run for different values of k and it was obvious from the results of Elbow and Silhouette score method that, the best value of k is 3.



The K-Means clustering algorithm is run again on our dataset which segments and clusters Auckland suburbs into three different clusters. A new dataset is then created with top five venues in each suburb, the cluster labels for each suburb, the area, latitude and longitude for each suburb. The suburbs are again visualised on map according to their clusters.

4. Results and Discussion

The suburbs of Auckland after clustering and segmentation using K-Means machine learning algorithm is as shown below.



There are three clusters, the characteristics of each cluster is as discussed below.

Cluster 0

	Area	Suburb	Average sale price	Average rent per week	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Auckland Central	Auckland Central	987500	1055	-36.848399	174.764388	0	Café	Hotel	Bar	Park	Coffee Shop
1	Central	Avondale	855105	554	-36.890448	174.687017	0	Café	Chinese Restaurant	Market	Mongolian Restaurant	Neighborhood
2	Central	Blockhouse Bay	930663	563	-36.928200	174.700190	0	Park	Café	Chinese Restaurant	Neighborhood	Mediterranean Restaurant
3	Central	Epsom	1601250	773	-36.888454	174.772938	0	Café	Japanese Restaurant	Chinese Restaurant	Indian Restaurant	Tennis Court
4	Central	Freemans Bay	1702167	881	-36.852879	174.750353	0	Café	Japanese Restaurant	Restaurant	Steakhouse	Italian Restaurant
...
86	Western	Ranui	654769	490	-36.865026	174.601326	0	Zoo	Pizza Place	Train Station	Convenience Store	Deli / Bodega

The most common venue in this cluster is café and parks. Suburbs in Central and Auckland Central area falls in this cluster. This cluster also has a variety of shops and the average rent per week is NZD 620 and average sale price of house is NZD

1,007,300. This cluster has the highest rent and house price. There are public transport venues like train station among the most common venues.

Cluster 1

	Area	Suburb	Average sale price	Average rent per week	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
65	Rodney	Warkworth	767071	541	-36.428190	174.640190	1	Café	Zoo	Food & Drink Shop	Flower Shop	Flea Market
67	Southern	Conifer Grove	677143	561	-36.946390	174.782401	1	Café	Park	Zoo	Ethiopian Restaurant	Flower Shop
78	Western	Glen Eden	712796	507	-36.913213	174.643600	1	Café	Indian Restaurant	Climbing Gym	Gym / Fitness Center	Park
79	Western	Glendene	746100	517	-36.890409	174.654240	1	Park	Bar	Café	Shopping Mall	Flower Shop
89	Western	Te Atatu Peninsula	850048	543	-36.840714	174.649985	1	Park	Grocery Store	Café	Go Kart Track	Zoo
91	Western	Titirangi	909294	581	-36.940495	174.647251	1	Café	Liquor Store	Restaurant	Trail	Pizza Place

This cluster has lesser suburbs than cluster 0. The most common venues on the cluster are once again café and parks. Western suburbs are the major percentage of this cluster. The average rent per week is NZD 541 and house sale price is NZD 777,075

Cluster 2

	Area	Suburb	Average sale price	Average rent per week	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
56	Pakuranga/Howick	Mellons Bay	1121000	611	-36.87820	174.92019	2	Beach	Playground	Restaurant	Park	Zoo
61	Rodney	Orewa	944679	588	-36.56667	174.70000	2	Beach	Scenic Lookout	Zoo	Electronics Store	Flower Shop
77	Southern	Weymouth	600071	507	-37.04820	174.85019	2	Fast Food Restaurant	Beach	Zoo	Ethiopian Restaurant	Food

This cluster has beaches and restaurants among the most common venues. It has mix of north, south and western areas. But looks like more kids friendly and picnic oriented areas. The average rent per week is NZD 568 and Sale price is NZD 888,583.

Analysing the map and the resultant dataset,

- Cluster 0 has more suburbs that are closer to the Auckland city. And offers a wide range of venues including public transport stations, shopping malls, restaurants, zoo, park etc. The average rent and sale price for houses are relatively high.

- b) Cluster 1 has suburbs that are near to city but seems like away from the chaos and rush of the city. The average rent and sale price is the lowest for this cluster when compared to other clusters. Most of the shops in this cluster are cafes, parks, zoo, grocery stores etc.
- c) Cluster 2 has suburbs tucked into the beauty of beaches and are farther from the city. The rent and sale price is less compared to city but more than cluster 1. And the venues are more suitable for a relaxing peaceful life.

It is hard to predict a specific area for renting/buying a house based on just the nearby venues. But based on the clustering we performed, Cluster 1 would be ideal for the ones who would enjoy the closeness to city and at the same time would love to stay away from the rush of city life.

Cluster 2 would suit the ones looking for a peaceful retired life and Cluster 0 looks like more suitable for the ones who loves city life and are ready to be spoiled for choices.

5. Conclusion

Buying or renting a house in a new place depends on various factors. Amenities available nearby are just one of those criteria. If a person is working, he/she would like to rent/buy a house closer to his/her workplace. Some would prefer a house closer to a school or college if they have children. While some would love a home with scenic views and beach nearby. It is all personal choices.

This study would help those who are looking to buy or rent a house in Auckland region to get an idea on the nearby amenities or venue and a general idea on pricing of houses. And the study can be further extended by including more features like distance from city, schools etc and using more data.