

IBM DATASCIENCE  
PROFESSIONAL CERTIFICATE

**Rent and Price of Houses in**  
**Auckland**  
CAPSTONE PROJECT



Asha Rajagopal  
May 2020

# 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.

# Data

For the study, we will need the following data:

1. The list of suburbs of Auckland region
2. The latest (for the year 2020) average rent and house price for 3 bedroom houses in Auckland region
3. The latitude and longitude data for each suburb
4. The most popular venues of each suburb

## 1.1 Data Source and Extraction Methodology

1. 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](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.
2. 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.
3. 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>

4. 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.