

# **Analysis of Auckland Properties**

Hang Zhao (Daniel) 9th August 2020

## **Executive Summary**

The dataset is about the information about property data in Auckland which retrieved from [koordinates](#) using Web API and Keys. It contains necessary information about each property. Along with statistical area information acquired from [Stats.NZ](#) and Socioeconomic Deprivation Indexes from Otago University, a full dataset has been built accordingly.

The analysis is based on 1050 observations for each of the 15 variables. The name of the variables is mostly self-explanatory. Few needs to explain more are CV, the Capital value of property; it is an APPROXIMATION of house value. SA1, an area unit classification. Age Groups, the number of people whose age is within 0 - 19 years old living in the SA1 unit area based on the 2018 census.

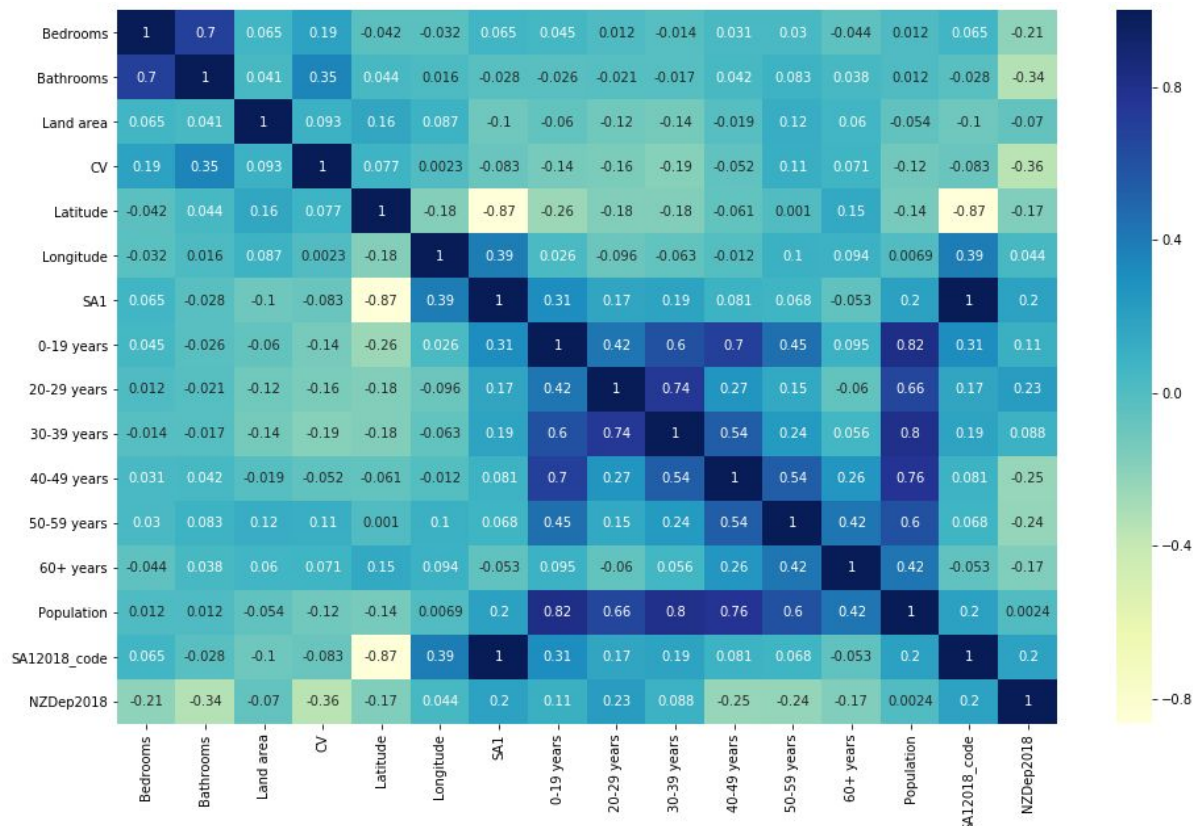
After exploring the data by calculating summary and descriptive statistics, and by creating visualizations of the correlation between each numerical variable, several highly correlated variables are found. After exploring the data, Linear Regression has been tested for this dataset training and the best model has been chosen based on the model accuracy.

## **Initial Data Exploration**

The initial exploration of the data began with some summary and descriptive statistics. Individual Feature Statistics Summary statistics for minimum, maximum, mean, median, standard deviation, and distinct count were calculated for numeric columns. Data from certain column also have been processed to make later analysis convenient. For example, the data type of land area, I turn it from string to numerical data. As the fact that machine learning model can only work on numeric data.

## **Correlation and Relationships**

The correlation between the numeric columns was calculated and observed in the below correlation plot.



The graph shows that the Population & 0~19 years group, Bedrooms & Bathrooms numbers and NZDep2018 & CV have a strong positive correlation with each other.

## Analysis

In this analysis, Linear Regression has been trained with 70% of the dataset. Testing the model with the remaining 30% data shows:

Algorithm	Accuracy	Relation
Linear Regression	0.9641184263062311	0-19 years & Population
Linear Regression	0.4974400825233866	Bedrooms & CV
Linear Regression	0.1337558901074376	NZDep2018 & CV

As shown above, there is a strong connection between 0-19 years ages group and the total population in certain SA1. And the relationship between capital value and bedrooms property has is not very strong.

## **Conclusion**

This analysis has shown that 0-19 years age group numbers can be confidently predicted from its measurements of the Local SA Population. In particular, the Linear Regression has the best performance as the accuracy rate is 96.41%.