Analysation of Auckland House Price

Han Lee, July 2020

Executive Summary

The dataset is the capital value of properties based on Auckland, which was downloaded from GitHub. It contains the number of bathrooms, bedrooms each property have followed by the geographical locations, the total population and population per different age groups of the suburbs where the properties are located, and deprivation index according to SA1 units.

The analysis is based on 1052 properties for each of 17 variables. The variables are self-explanatory except CV, which stands form capital value of property.

After exploring the data by data collection using an API call, and by creating visualisations of correlation between each numerical variable, moderate correlated variables are found.

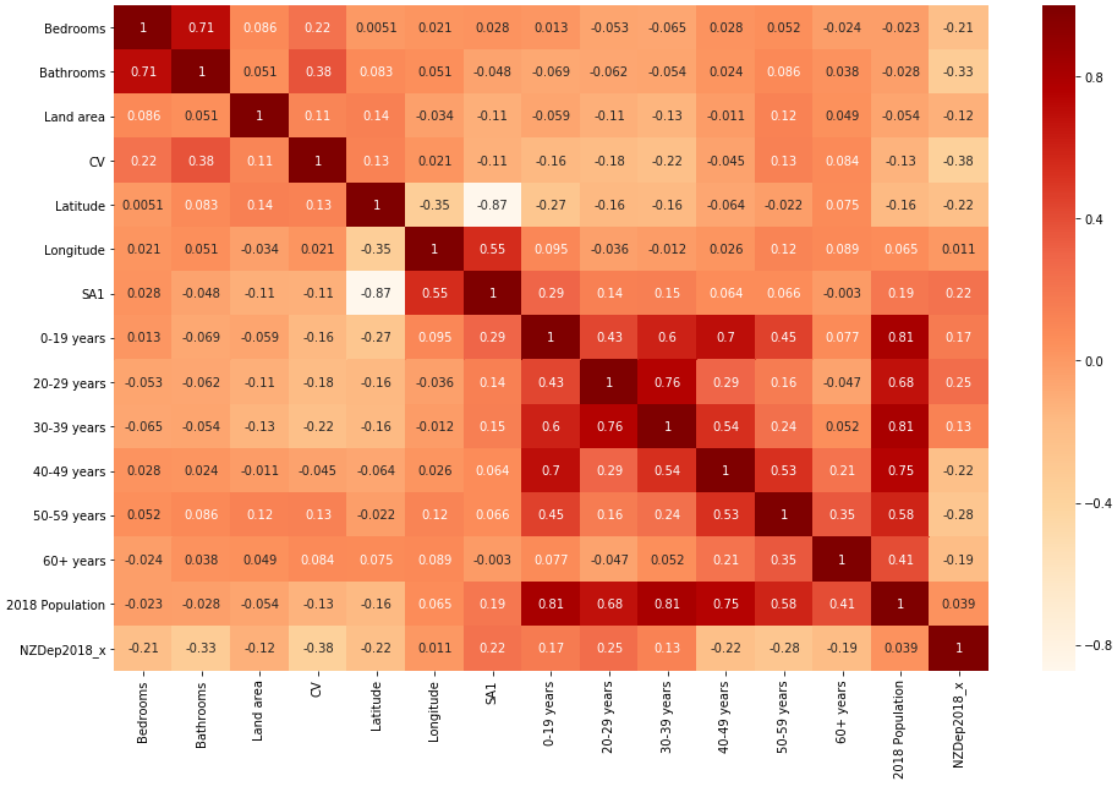
Initial Data Analysis

The dataset is the first five properties in the csv file with all the variables used. NZDep2018 represent the New Zealand Index of Deprivation 2018, where Decile 1.0 represents areas with the least deprived scores and Decile 10.0 represents areas with the most deprived scores.

	Bedrooms	Bathrooms	Address	Land area	CV	Latitude	Longitude	SA1	0-19 years	20-29 years	30-39 years	40-49 years	50-59 years	60+ years	Suburbs	2018 Population	NZDep2018_x
0	5	3.0	106 Lawrence Crescent Hill Park, Auckland	714.0	960000	-37.012920	174.904069	7009770	48	27	24	21	24	21	Manurewa	174	6.0
1	5	3.0	8 Corsica Way Karaka, Auckland	564.0	1250000	-37.063672	174.922912	7009991	42	18	12	21	15	30	Karaka	129	1.0
2	6	4.0	243 Harbourside Drive Karaka, Auckland	626.0	1250000	-37.063580	174.924044	7009991	42	18	12	21	15	30	Karaka	129	1.0
3	2	1.0	2/30 Hardington Street Onehunga, Auckland	65.0	740000	-36.912996	174.787425	7007871	42	6	21	21	12	15	Onehunga	120	2.0
4	3	1.0	59 Israel Avenue Clover Park, Auckland	601.0	630000	-36.979037	174.892612	7008902	93	27	33	30	21	33	Clover Park	231	9.0

Analysis of Correlations and Pattern in the data

The correlation between the numeric columns were calculated and observed in the below correlation plot. (The right colour bar indicated the correlation values. For example, dark red means correlation value is positive 1 and light orange means correlation value is negative 1.)



The graph shows that there are mild or moderate relation between CV and population, number of bathrooms and bedrooms, and deprivation index.

Analysis

The analysis has been done by using linear regression with 80% trained data and 20% test data. The data produced a somewhat correaltion between deprived index and capital value, yet there are no significant correaltionship. Though, since the accuracy of outcome of the model is only 0.34 in R^2 score, the outcome needed to be handled with care.