Auckland House Prices Analysis

The following analysis used three sets of data. Housing prices dataset provided with the assignment, Koordinates's geographical API to obtain population data, and Otago University's census 18 datasets to obtain the deprivation index. We merged the data by using longitude, latitude and SA1 information, and analysed a total of 1052 rows of data with 15 columns of variables.

Basic Information

	Bedrooms	Bathrooms	Land area	CV	Latitude	Longitude	SA1
count	1051.00	1051.00	1051.00	1051.00	1051.00	1051.00	1051.00
mean	3.78	2.07	856.99	1387520.55	-36.89	174.80	7006319.18
std	1.17	0.99	1588.16	1182939.36	0.13	0.12	2591.26
min	1.00	1.00	40.00	270000.00	-37.27	174.32	7001130.00
25%	3.00	1.00	321.00	780000.00	-36.95	174.72	7004415.50
50%	4.00	2.00	571.00	1080000.00	-36.89	174.80	7006325.00
75%	4.00	3.00	825.00	1600000.00	-36.86	174.88	7008383.50
max	17.00	8.00	22240.00	18000000.00	-36.18	175.49	7011028.00

0-19 years	20-29 years	30-39 years	40-49 years	50-59 years	60+ years	Population	Deprivation Index
1051.00	1051.00	1051.00	1051.00	1051.00	1051.00	1051.00	1051.00
47.55	28.96	27.04	24.13	22.62	29.36	179.91	5.06
24.69	21.04	17.98	10.94	10.21	21.81	71.06	2.91
0.00	0.00	0.00	0.00	0.00	0.00	3.00	1.00
33.00	15.00	15.00	18.00	15.00	18.00	138.00	2.00
45.00	24.00	24.00	24.00	21.00	27.00	174.00	5.00
57.00	36.00	33.00	30.00	27.00	36.00	210.00	8.00
201.00	270.00	177.00	114.00	90.00	483.00	789.00	10.00

I used Pandas describe() function to generate the count, mean, std, max, 25%, 25% and 75% results for each column based on the 1052 rows of data.

CV Prediction

I split the CV data into two sets(3.5 to 6.5 split) for training and testing purposes, and dropped locational and descriptive data such as addresses, longitude and latitude. These information is unlikely to help during the prediction.

With the remaining variables, I ran the training and prediction and received a R^2 score of 0.317