Analysis of Auckland House Price
Harshikesh Raniga

Summary
The dataset that I used were from koordinates website and otago university website for the year 2018. The dataset contains information on the Capital Value and the properties of the houses such as bedroom,bathroom, land area, and suburbs.
The analysis is based on 1051 properties in auckland. Latitude and longitude represent the latitude and longitude coordinates of the property.

After exploring the data and modelling it. We plotted graphs and found that there was low correlation between all the attributes in the table.
I dropped all the age(peoples) attributes as that would not be involved in making a decision to buy a property. I also dropped the longitude and latitude coordinates of the property as that would be irrelevant data to use for such a model as longitude and latitude are just coordinates where suburbs would be more useful and help identify a correlation.

Below is the table that shows the descriptive statistics of the data.

	Bedrooms	Bathrooms	CV	Latitude	Longitude	SA1	0-19 years	20-29 years	30-39 years	40-49 years	50-59 years	60+ years	SA12018_code	NZDep2018	Population
count	1048.000000	1048.000000	1.048000e+03	1048.000000	1048.000000	1.048000e+03	1048.000000	1048.000000	1048.00000	1048.000000	1048.000000	1048.000000	1.048000e+03	1048.000000	1048.000000
mean	3.779580	2.074427	1.388544e+06	-36.894561	174.799026	7.006332e+06	47.544847	28.915076	27.00000	24.131679	22.597328	29.353053	7.006332e+06	5.065840	179.799618
std	1.167894	0.992904	1.184422e+06	0.128426	0.117991	2.583920e+03	24.713408	20.993232	17.93158	10.956798	10.212455	21.810055	2.583920e+03	2.912027	71.087298
min	1.000000	1.000000	2.700000e+05	-37.265021	174.317078	7.001130e+06	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	7.001130e+06	1.000000	3.000000
25%	3.000000	1.000000	7.800000e+05	-36.950873	174.722226	7.004426e+06	33.000000	15.000000	15.00000	18.000000	15.000000	18.000000	7.004426e+06	2.000000	138.000000
50%	4.000000	2.000000	1.080000e+06	-36.893409	174.798612	7.006334e+06	45.000000	24.000000	24.00000	24.000000	21.000000	27.000000	7.006334e+06	5.000000	174.000000
75%	4.000000	3.000000	1.600000e+06	-36.856280	174.880943	7.008390e+06	57.000000	36.000000	33.00000	30.000000	27.000000	36.000000	7.008390e+06	8.000000	207.750000
max	17.000000	8.000000	1.800000e+07	-36.177655	175.492424	7.011028e+06	201.000000	270.000000	177.00000	114.000000	90.000000	483.000000	7.011028e+06	10.000000	789.000000

Correlation and relationship
Below the graph represents the correlation between the numeric columns. The darker the shade of red it is the stronger the correlation value and the lighter or more orange it is the weaker the correlation value.
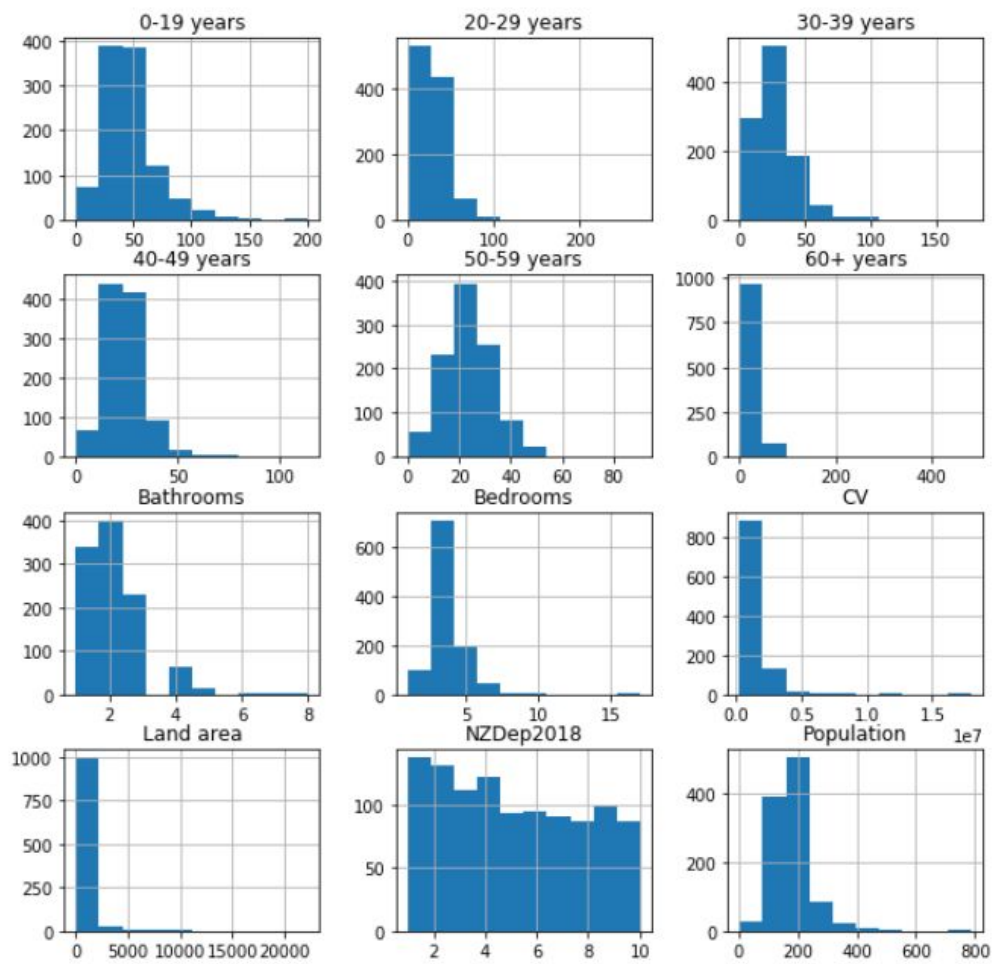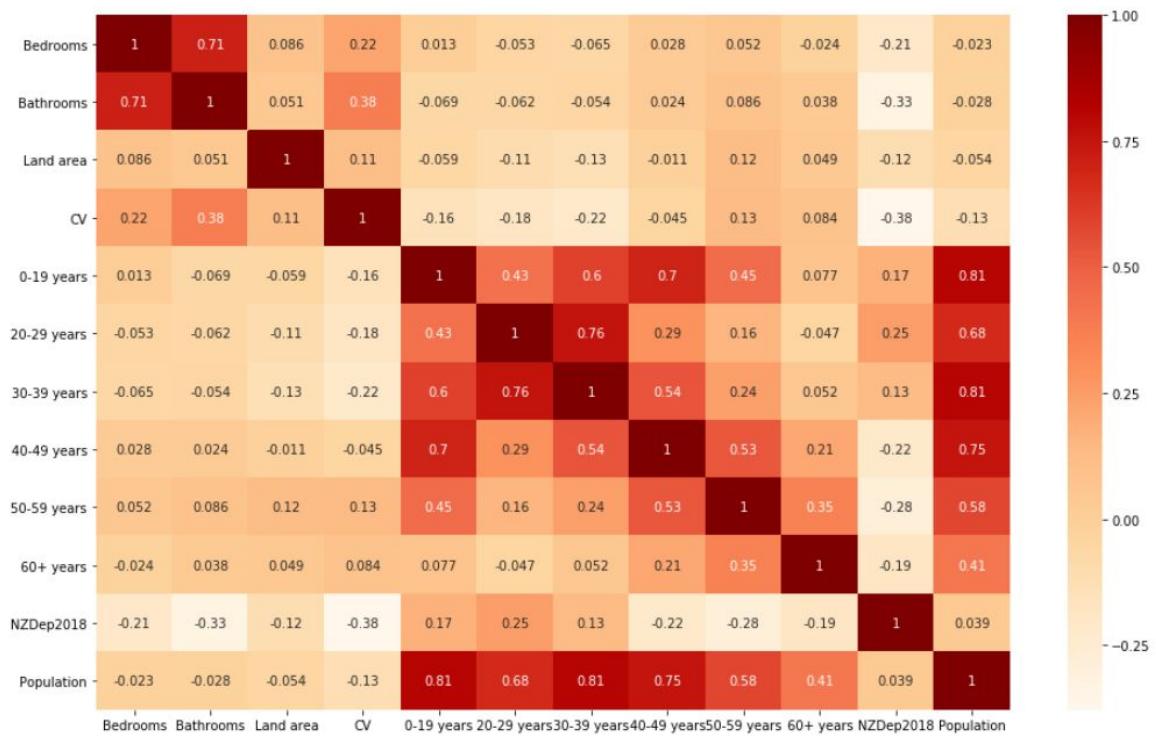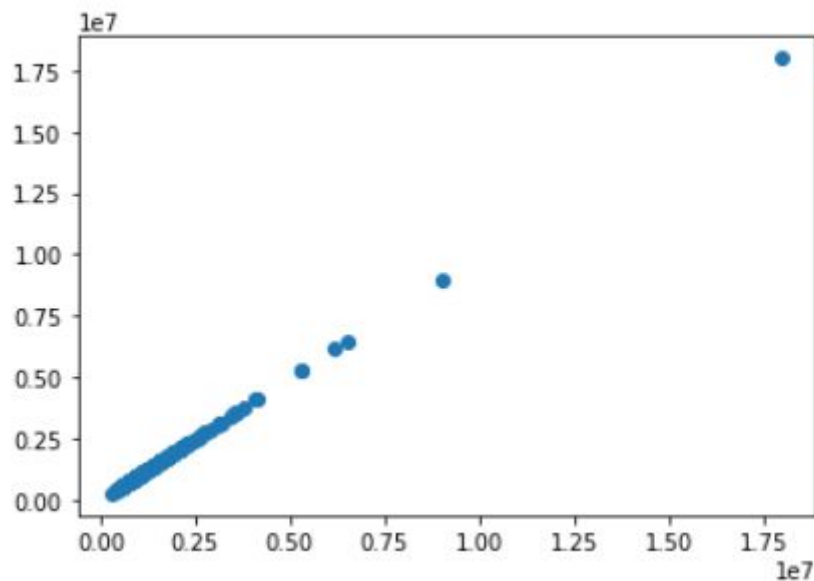
There is an outlier that indicates one of the houses has 17 bedrooms. Which may skew the relationship

Analysis
By looking at the data I can see the correlation between the suburbs and the capital value. Popular areas such as Royal oak, mission bay, epsom tend to have a higher CV than those of suburbs that aren't popular or closer to the city. There is also a relationship between land area and CV the higher the land area the more valued the property is which is also dependent on the suburb.

Conclusion
From the analysis, we saw that the house prices can confidently be predicted by looking at its land area and the suburb the property is located in. Properties closer to the cities tend to have a higher CV and properties further away from the cities or popular locations have a lower CV thus potentially low price.