Linear Regression Analysis of Auckland House

Minhan Cho, July 2020

Executive Summary

The dataset is about Auckland house prices. The analysis is based on 1051 observations for each of the 18 different variables.

Initial Data Exploration

The dataset had 3 nulls values: 2 null values in Bathrooms and 1 null value in Suburb. The variable bathrooms are deemed significant correlating value of 0.22 therefore, its value was set to '0'. On the contrary for Suburb, because Suburb was not a numerically value suitable for analysis and NZDep2018 acted like Suburb therefore deletion of the single wrong does not significantly change the study.

CV are skewed so it makes sense to transform it with logarithm when running a linear regression.

Correlation and Relationships

The initial exploration of data began with a heat maps to look for variables with correlations to the CV.

The map shows that Bedrooms, Bathrooms, Land Area and NZDep2018 show the strongest correlations CV.
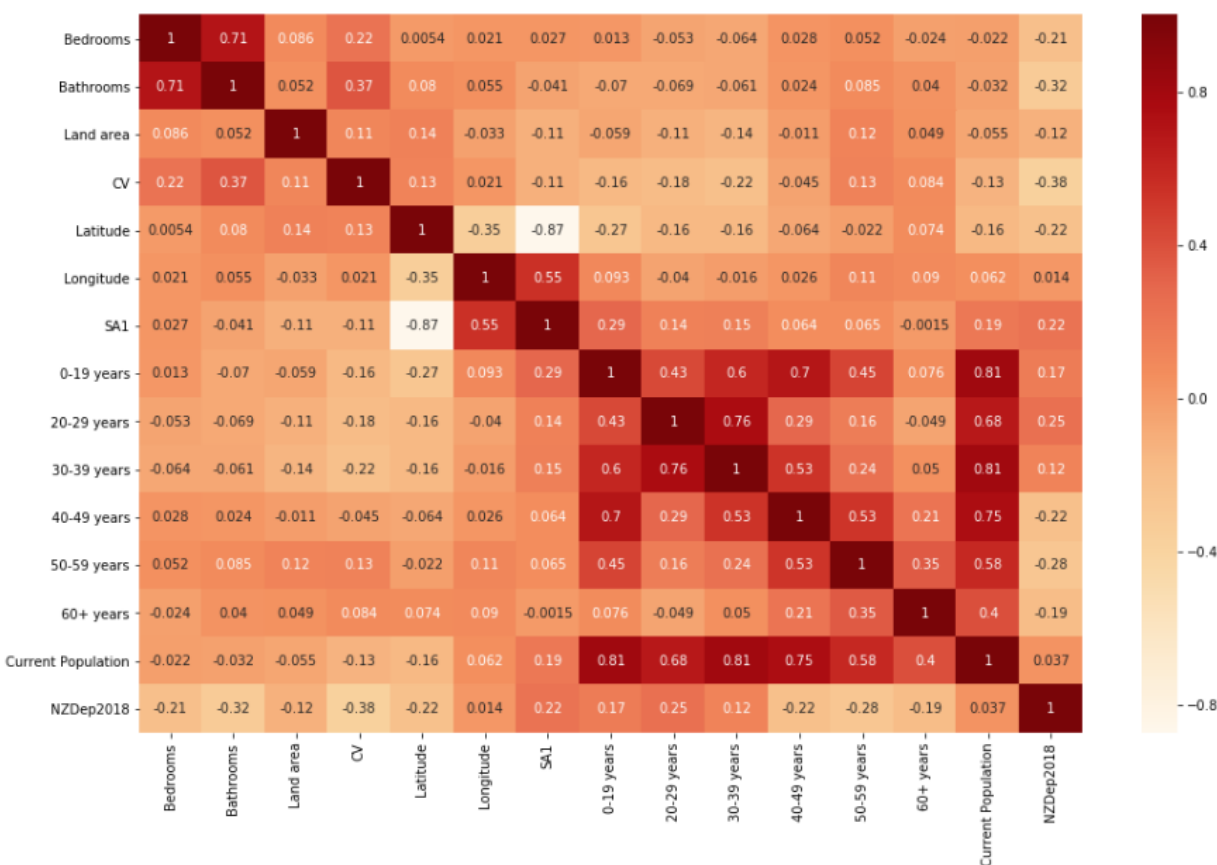
Analysis

In this analysis, we used linear algorithm, and this was trained with 60% of the data and the remaining 40% of the data was used to testing.

Using this method, yielded a model score of 0.346.

Conclusion

Based on model evaluation, I cannot be confident with the score and overall, I think is hard to predict with the given data. However, after going through other models I would still recommend this over other models.