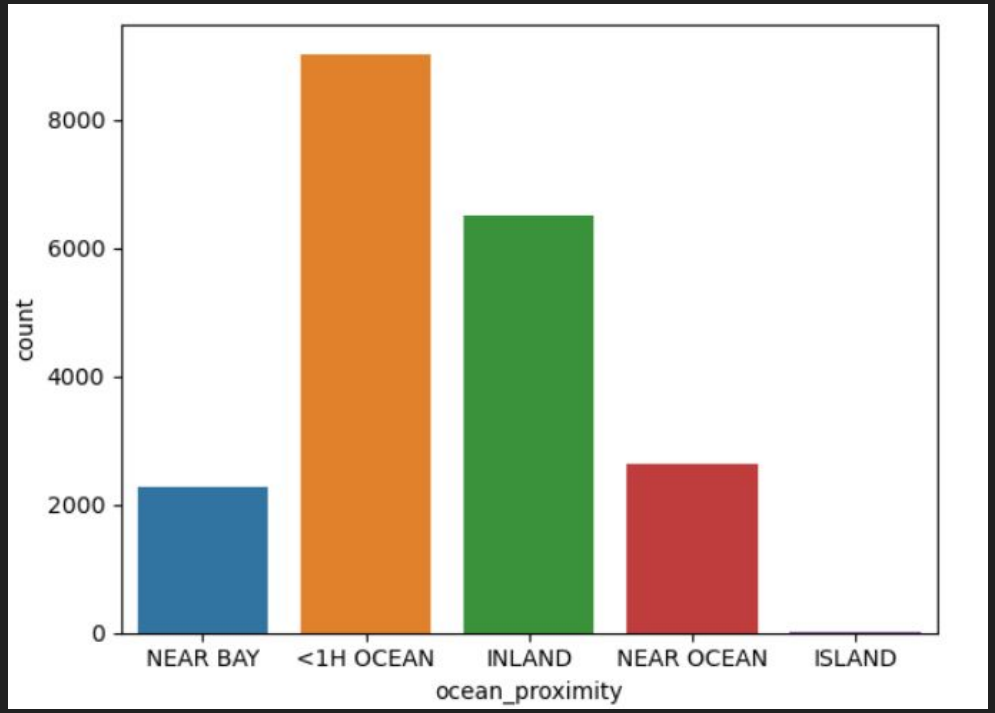
San Francisco 1990's Housing Price Prediction

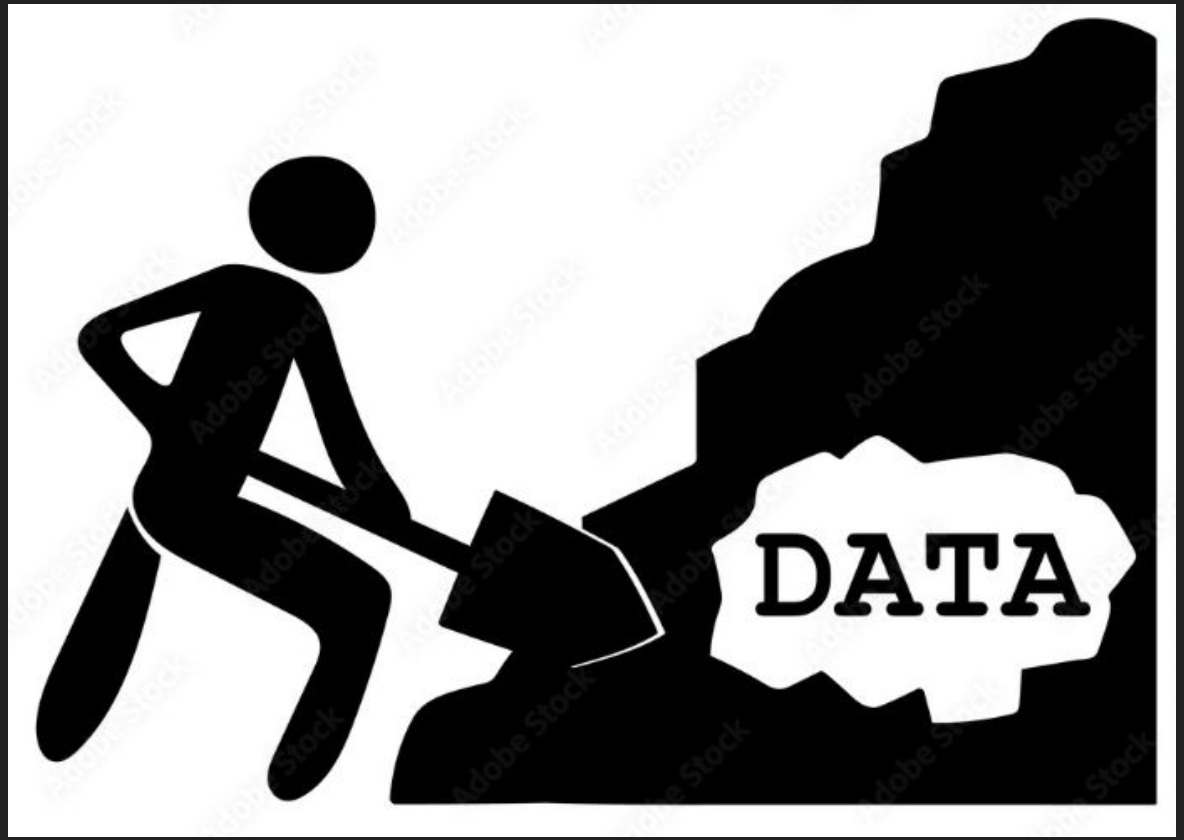
By Anthony Medina

How much does it cost to live in San Francisco?

This project explores the 1990 census data to see what type of housing a family could afford.

How much will it cost to live in these places?





First we dig through the data…

What do we find?

- Missing Values
- Skewed Distributions
- Outliers
- Correlations

Who about missing data?

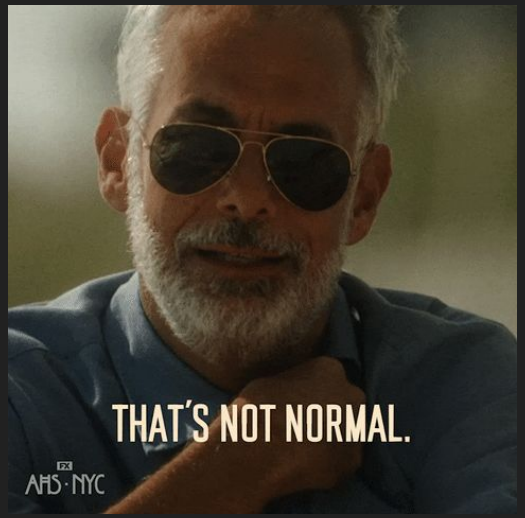
Mean

Remove

Median

Zero

Pre-Processing Data Steps



Model Metrics

	R-Squared	MSE	RMSE
Linear Regression	0.635	490786835.53	70056.18
Polynomial Regression	-1.617	2.17e+30	14749069107868894.8
Ridge Regression	0.634	4912746344.80	70090.99
Lasso Regression	0.635	4907950136.07	70056.76

Model Choice: Linear Regression

- Tied for best R-Squared
- Almost the best MSE (But really close to tied for best)
- Almost the Best RMSE (But really close to tied for best)
- Easily the least complicated of the models.

What did the model have to say?

Predicted 1990's Prices	Adjusted 2023 Prices
Inland: $417,603.01	Inland: $967,883.48
Island: $426,431.03	Island: $989,773.50
Near Bay: $385,457.96	Near Bay: $894,672.40
Near Ocean: $326,171.32	Near Ocean: $757,064.35
Ocean Front: $257,276.63	Ocean Front: $597,157.40

Further questions

1. How does the scaled model compare to the actual values of today's housing in San Francisco?
2. What other models would we have used for to predict a price?
3. Would there be a significant difference if we dealt with missing values differently?