

Parts to do:

- 1) Explanatory data analysis to see the relationship between the prices and categories: like “prices” vs “area of housing” - can be scatter plots, heatmaps, etc.
- 2) Data preprocessing:
 - a) Cleaning the NAN, null, missing values
 - b) Handling Text and Categorical Attributes (like ocean proximity)
 - c) Feature scaling - normalization
 - d) Creating new parameters out of existing? (total_bedrooms/total_rooms)

DATA SPLITS HERE?

- 3) Modelling:
 - a) Linear Regression - baseline
 - b) Random Forest
 - c) Lasso
 - d) Elastic Net
 - e) Support Vector Machine?
 - f) Gaussian Processes (https://scikit-learn.org/stable/modules/gaussian_process.html)
- 4) Models optimization (hyperparameters tuning)
- 5) Basic error analysis using cross validation (mean squared error)
- 6) Feature extraction/importance (Can experiment with different categories (features, columns) combinations for the models)