California Housing Prices Prediction Proposal

Problem Statement: What information can we gather from our California Housing Prices data set that we can use to predict the price of houses with varying features.

Context: The housing market is somewhat predictable. It ebbs and flows depending on current interest rates, and the demand on the market. We've had an influx of population that will affect the rate at which people can purchase houses. The intention of this project is to explore prices based on varying features.

Criteria for success: This project is finished when we have a model that will predict a housing price given a location and feature set.

Scope of solution space: The focus on this initiative is to learn regression models using real data and on getting a model that makes predictions at the same time to maximize learning.

Constraints: We are bound by the accuracy of the data collected from Kaggle. We are also going to be limited by my exposure to data science ideas and concepts.

Stakeholders: This project will be presented to my mentor Silvia Seceleanu for accuracy and to Springboard for review.

Data Sources: There is a Kaggle data set of California Housing Prices with 20.6K data points. The features of the data set include longitude, latitude, housing median age, total rooms, total bedrooms, population, households, median income, median house value, and ocean proximity.

Deliverables: This contents of this project will be contained in a Github Repo that includes a PDF of this problem statement, all code written including notebooks and scripts, a slide deck, and a project report.