Hannah D'Souza

ENSF 611 Project Proposal

30026013

Due: March 24th, 2023

Why: Question being investigated

a. I want to investigate the relationship of how different parameters affect the price of houses in a particular neighborhood like the California housing dataset.

How: Plan of attack

a. I will use a dataset of house sales in the neighborhood, which contains information such as house size, number of bedrooms, total rooms, and median income. First, I will preprocess the data by handling missing values and encoding categorical variables. Then, I will train a linear regression model to predict the price of houses based on the number of bedrooms. I will evaluate the model using metrics such as mean squared error and R^2 score and plot the model's predictions against the actual prices. Additionally, I will use feature scaling to see if it improves the model's performance. Finally, I will interpret the model's coefficients to gain insights into the relationship between number of bedrooms and house prices.

What: Dataset, models, framework, components

a. California Housing Prices (10 features, 20000 samples)

b. Python with scikit-learn and Panda's libraries.

c. Linear regression model with feature scaling and coefficient interpretation.