Predicting housing prices is always a challenge to some degree due to the various variables that come play. For the sake of this project, and to demonstrate a working knowledge of the materials covered in class, we have decided to work on the Ames, Iowa dataset.

Our goals is to apply the different regression methods (or one at least) taught in class and find out if we can successfully predicting the housing prices within an acceptable range of error using the features that our analysis deems sufficient to be used in the model.

We want to examine the relationship between the features and their correlations. Investigate weather those features play a major role in deciding a price for specific house. For example, does the neighborhood where the house is located matter? Does having more than two floors matter?...etc

Using libraries such matplotlib and seaborn, we will better understand ‘visually’ how our features are related and explore some basic statistics such as mean, mode and standard deviation. Also, explore the outliers in the dataset and decide how would they affect our analysis and model results.

Of course, our Initial EDA will pave the way on deciding which features will be fitted in the model that we choose and what features to drop. Since prices are continues and individual house prices cannot be classified, we will follow a linear regression model.