The regression project draws data on house sales in King County Washington (the county within which Seattle is located). This data is available through the statistical competitions website, kaggle.com (direct link; <https://www.kaggle.com/harlfoxem/housesalesprediction>). Our goal is to study the sales price using all or a subset of the following regression variables: date of sale, number of bedrooms, number of bathrooms, square footage of the home, square footage of the whole lot, number of floors, a qualitative response of whether a house has a waterfront view, whether the house has been viewed, an overall rating on condition, the grade given to the house based on a county-wide grading system, square footage of the house not counting the basement, square footage of the basement, the year it was built, the year it was renovated (if renovated), the zipcode (a variable that must be converted to qualitative for proper study), latitude and longitude, and living room area in 2015.

The price is the response variable. The price has a mean 540,008.1 with variance 134782378397 and standard deviation 367,127.3. The descriptive statistics for the explanatory variables are as follows:

* Numbers of bedrooms has a mean of 3.37 with a variance of 0.8649 and a standard deviation of 0.93.
* The number of bathrooms has a mean of 2.1 with a variance of 0.59 and standard deviation of 0.77.
* The square foot of the home has a mean of 2,079.9 with a variance of 843,533.7 and standard deviation of 918.44.
* The number of floors has a mean of 1.49 with a variance of 0.29 and standard deviation of 0.53.
* The square footage of the lot has a mean of 15,106.97, with a variance of 1,715,658,774 and a standard deviation of 41,420.51.
* The condition of the houses had a mean of 3.41 with a variance of 0.42 and standard deviation of 0.65.
* The grade of the houses had a mean of 7.65 with a variance of 1.38 and a standard deviation of 1.17.
* The square footage above the ground, which doesn’t include basements, has a mean of 1788.39 with a variance of 685734.7 and a standard deviation of 828.09.
* The square footage of the basements of the houses has a mean of 291.509 with a variance of 195872.7 and a standard deviation of 442.575.
* The lot square footage has a mean of 12768.46, variance of 745,483,731.4 and standard deviation of 27,304.18. This high variance is likely caused by strong outliers that far exceed the mean.
* The average square footage of the living rooms was 1986.55 with a standard deviation of 685.39. Variance measured at 469739.505. This was also a result of significant outliers for large homes in the county.

We will create a viable predictive model for predicting housing price using at least a reasonable subset of these variables which will lead to noteworthy prediction accuracy. Our hypothesis is that higher square footage leads to a higher selling price.