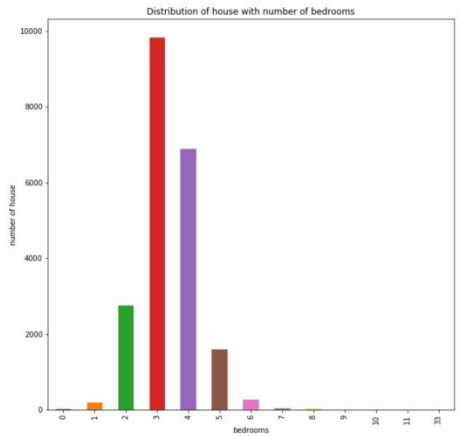
Capstone Project 1: Exploratory Data Analysis
-Mehreteab Kidane

To gain a sense of the relationship of the features with each other and with house prices, the target variable, I tried to use a diverse set of data visualization tools, including the following: boxplots, histogramplots, scatter plots and correlation plots. The first EDA I performed was to examine the distribution of the home sale prices.
The price of a house is dependent on various factors like size of area, how many bedrooms, location and many other factors.

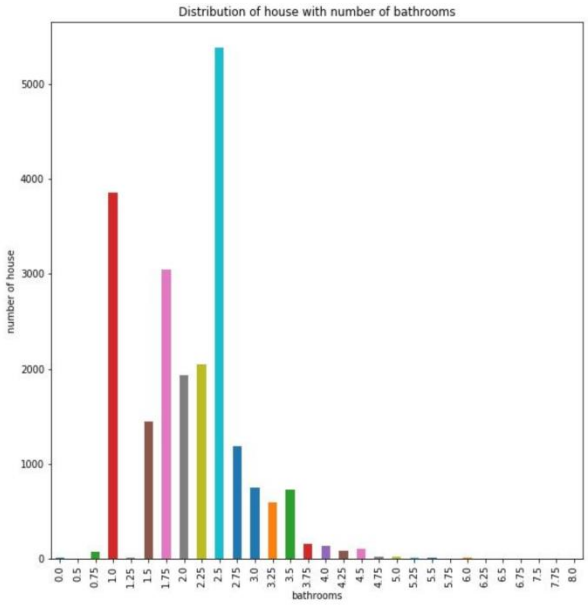
I conducted a hypothesis test to check if there is no significant correlation between a number of bedrooms and price. The p-value for the hypothesis test is less than the level of significance 0.05, so I rejected the null hypothesis. So I support that there is a correlation between a number of bedrooms and price.



The distribution of the house with the number of bedrooms plot tells us that most of the houses have 3 bedrooms.

I also conducted a hypothesis test to check the correlation between a number of bathrooms and price. The p-value for the hypothesis test is less than the level of significance 0.05, so I reject the null hypothesis and suggest that there is a correlation between a number of bathrooms and price.

Similarly, I conducted a hypothesis test to check the correlation between sqft_living and price. The p-value for the hypothesis test is less than the level of significance 0.05, so I reject the null hypothesis and suggest that there is a correlation between sqft_living and price. I also conducted a hypothesis test to check if there is a correlation between grade and price.



The distribution of the house with the number of bathrooms plot tells us that most of the houses have 2.5 bathrooms.

The test suggests that there is correlation between grade and price. I also conducted a hypothesis test to check if there is no statistical importance between mean house price and a number of bedrooms less than 3 and greater than 3. The p-value for the test is greater than the level of significance 0.05, so I fail to reject the null hypothesis. This suggests us that there is no statistical importance between mean house price and a number of bedrooms less than 3 and greater than 3.