* The original dataset had 5000 instances.
* The data types where: 4 ints, 10 floats an 2 strings.
* In the data exploration phase, there were some challenges dealing with missing data like: Finding object type columns, dealing with strings, etc.
* Data deletion and data prediction were needed in this project to clean the dataframe.
* HOA' feature has the most missing values. More than 10% were missing.
* Linear regression model for predicting the 'HOA' feature values has a poor performance so, it was necesary to replace nan values with the median due to outliers.
* Some outliers have to be deleted.
* After the cleaning process, the dataset has 4570 instances.
* 430 instances had to be deleted.
* Correlations between variables tend to be poor.
* Columns and their type once cleaned:
* MLS - Int
* sold\_price - Float
* zipcode - Int
* longitude - Float
* latitude - Float
* lot\_acres - Float
* taxes - Float
* year\_built - Int
* bedrooms - Int
* bathrooms - Float
* sqrt\_ft - Float
* garage - Float
* kitchen\_features - String
* fireplaces - Float
* floor\_covering - String
* HOA - Float