* Write Milestone Report, share with me and give me edit access at [bsbell21@gmail.com](mailto:bsbell21@gmail.com)
  + Intro
    - Opening sentence that draws in the reader (ie personal story)
    - What is the context for the problem we’re solving
    - Why would this be useful/to whom?
    - In brief (1/2 sentences max), what are we doing
  + Dataset
    - How did you get the data
    - What does the final dataset look like
    - Include any decisions that someone would want to know about (err on the side of being sparse here)
  + EDA:
    - make sure to include titles and axis titles (with units) for each chart
    - you’ll probably want to lead with the univariate chart of the target, general rule is go high level to specific
      * Univariate
        + Univariate chart of the target
      * Bivariate charts
        + Look at you bivariate charts comparing features to the target variable (one example would be job vs % subscribed)
        + Choose 3-5 of these that seem the most interesting, put them in your report, and tell a story about why they are interesting, what the takeaway is, and why we might be seeing what we’re seeing'

A question to answer each time might be: "Is the takeaway of this chart what we expected? If so, why? If not, why might we be seeing this?"

* + - * + Make sure to include your most important features (from the feature importances) x
      * Correlation matrix
        + Comment on what correlations we’re seeing between features that might be causing multicollinearity, or if there isn’t any to be worried about
        + You may also comment briefly on any features that seem highly correlated with the target
      * Feature importances
        + Explain what the most important features are and why that might be. Again, answer the question: "is this what you expected? If so, why? If not, why might we be seeing this?"
  + Intro
    - Opening sentence that draws in the reader (ie personal story)

Housing price prediction is always be high interest to people who would like to buy a home with low price or sale their home with high price. There are many variables associated with house, which can affect the price at different level. It is important to figure out variables with high influence on the sale price.

* + - What is the context for the problem we’re solving

I chose this competition’s dataset (Housing Prices Competition for Kaggle Learn Users) from Kaggle as my CapStone project, because it has as many as 79 explanatory variables describing almost every aspect of residential homes in Ames, Iowa. We would like to take this competition to challenges myself to predict the final price of each home, with skillset I have learned from the data science courses in Springboard.

* + - Why would this be useful/to whom?
    - In brief (1/2 sentences max), what are we doing
  + Dataset
    - How did you get the data

I got the data from Kaggle.

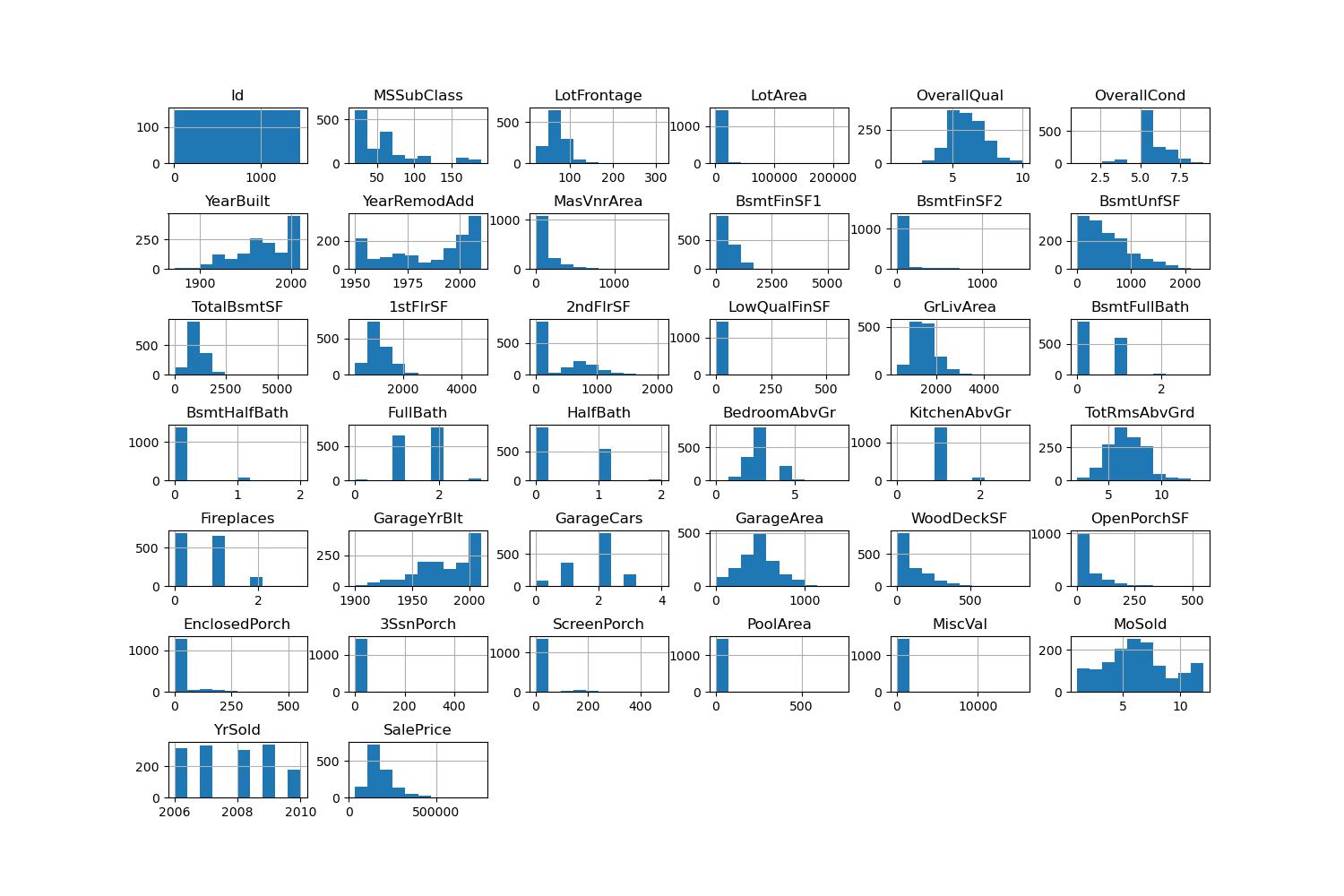
* + - What does the final dataset look like

The dataset have two dataset—train and test. Each dataset has the same size of 1460 sale records. The train dataset include 79 explanatory variables describing almost every aspect of residential homes in Ames, Iowa, while the test dataset has 78 explanatory variables lacking Sale Price variable.

* + - Include any decisions that someone would want to know about (err on the side of being sparse here)
  + EDA:
    - make sure to include titles and axis titles (with units) for each chart

A group of blue and white bars

Description automatically generatedA group of blue and white bars

Description automatically generated

* + - you’ll probably want to lead with the univariate chart of the target, general rule is go high level to specific
      * Univariate
        + Univariate chart of the target
      * Bivariate charts
        + Look at you bivariate charts comparing features to the target variable (one example would be job vs % subscribed)
        + Choose 3-5 of these that seem the most interesting, put them in your report, and tell a story about why they are interesting, what the takeaway is, and why we might be seeing what we’re seeing'

A question to answer each time might be: "Is the takeaway of this chart what we expected? If so, why? If not, why might we be seeing this?"

* + - * + Make sure to include your most important features (from the feature importances) x
      * Correlation matrix
        + Comment on what correlations we’re seeing between features that might be causing multicollinearity, or if there isn’t any to be worried about
        + You may also comment briefly on any features that seem highly correlated with the target
      * Feature importances
        + Explain what the most important features are and why that might be. Again, answer the question: "is this what you expected? If so, why? If not, why might we be seeing this?"