* Look at the column features. Make a hypothesis.
* Learn more about the dataset. See the numerical data stats. See the categorical data stats. Are you sure the data columns are the right data type?
  + For instance, the year could be a numerical value. Would year as an object be better suited for your project? You have to make these decisions.
  + Is there missing data? How much in a column? Are you going to drop the column? Or can you fill it up with something else? You need to make decisions on this.
    - You can also learn more about the missing data and make guesses why by looking at the missing rows.
* Explore the dependent variable?
  + Not sure how this works.
* After learning more about the data, EDA
  + Ask questions about the dataset. What more do you want to learn?
  + Make visualizations and analysis based on the questions.
* Creating a “pipeline”
  + Process and clean the data based on what you want to do. It is “automatic”
  + Pipeline is just a fancy term for combining all the preprocessing into one condense function.
    - Practice this in your next project
* Initialize models.
* PCA – used when dealing with high-dimensional data
  + Reduce data
  + Visualize in a way original feature space couldn’t
  + Serve as feature engineering by creating new features
* Feature engineering
  + Overlaps with processing data
  + More focused on how to create data that is best suited to the model you are working with.