# Documents Included in the Case Study

1. 3 Assignments
2. Answer Sheet for each assignment
3. The csv file
4. Heart Disease Final (combines all three assignments in a cohesive final format, what the three assignments was based off of)

## Assignments

1. Exploratory Data Analysis

* Created with the intention that the student will fill in the blank to create the plots, or do the entire plot by themselves, but using one that was already filled out to aid as an example.

1. Statistical Testing

* From the findings in the first assignment, test the statistical significance of each and determine if they are significant enough to determine for the population.
* Primarily only doing z-test and t-tests
  + Z-test, I did with two methods
    - The first method I used the statsmodel package.
      * I like this method because it was very simple with the package, and it would help the student get more comfortable with working within packages
    - The second method is doing it all “by hand”
      * This one is not that much brain power either, but definitely more than just using the package. However, it drives home the idea that Python can be used in a calculator environment, and can be easy for those who have stats background knowledge to use.
    - I think Its more practical as DASC student to use the statsmodel as they will have access to that library their entire time in the field, but also think doing it by hand is beneficial to practice their stats skills
    - I could do it by asking them to do it by hand 1st then have them do it with the statsmodel package to check it over.
  + T -test
    - The four other tests I have them do are all 2-sample t- tests. It measures the mean of the variable we are testing for those in the dataset that do not have heart disease compared to that of those who do not have heart disease.
    - I did the entirety of the first one by myself for them to use as an example
    - The second one is missing one line of code that they need to include
    - The third and fourth I have them do the entirety of the test
  + Finished with a concluding question

1. Predictive measures

* Because machine learning/predictive learning isn’t a primary goal of this class, I went with a different approach for the final assignment.
* For this case study, I go over two predictive measures for the dataset.
  + Decision Tree Diagram
    - Started with changing the sex variable to a boolean so that it can be used within the analysis.
    - I laid out all the steps and approaches clearly, so that It could be replicated if they wanted to use it later
    - Provided 6 questions for the student to respond to regarding the output of the decision tree as well as the steps to get there.
    - Also did the confusion matrix to determine how the model worked with the data
  + Binary logistic regression
    - Explained why this is a good model for our data
    - Explained the meaning of the output of the regression summary as well as the odds ratio table.
    - Asked 4 questions about the output interpretation and about the importance of using confusion matrix to test the models