- Main objective of the analysis that specifies whether your model will be focused on prediction or interpretation and the benefits that your analysis provides to the business or stakeholders of this data:
    - The model was focused on prediction. It can be used to help with the identification of heart disease with the given features.
- Brief description of the data set you chose, a summary of its attributes, and an outline of what you are trying to accomplish with this analysis:
    - The dataset that I chose has a set of thirteen features that are correlated with the diagnosis of heart disease. The target is used to identify weather there was a diagnosis of heart disease or not. The target field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4. Experiments with the Cleveland database have concentrated on simply attempting to distinguish presence (values 1,2,3,4) from absence (value 0).
    - I am trying to implement a classification tool used to predict the diagnosis of heart disease in the patients, so it can read feature data points and predict weather or not the patient was diagnosed with heart disease.
- Brief summary of data exploration and actions taken for data cleaning and feature engineering:
    - The data was rounded to the second decimal place and quantile data was examined along with null values. No null values were found in the data, and the data was normalized between values of zero to one across all columns. Pair plots of all the attributes in the data was created, and heatmaps for all attributes was also created. The heatmaps and pair plots show that some of the values are correlated with one another, whereas other values show negative correlations with one another.
- Summary of training at least three different classifier models, preferably of different nature in explainability and predictability. For example, you can start with a simple logistic regression as a baseline, adding other models or ensemble models. Preferably, all your models use the same training and test splits, or the same cross-validation method:
    - A linear regression check with one degree of polynomial features and a training testing split of 20 percent found an R^2 value of 0.7048177057965919.
    - A logistic regression check using a newton-cg solver and a training testing split of 20 percent found an R^2 value of 0.8679653679653679.
    - A logistic regression check using a saga solver and a training testing split of 20 percent found an R^2 value of 0.801948051948052.
    - A lasso regression check using an alpha value of 0.001 and a training testing split of 10 percent found an R^2 value of 0.805797016228342.
    - A SVM check using a C of 0.3 and a training testing split of 20 percent found an R^2 value of 0.9672131147540983.
    - A KNeighborsClassifier check using fifteen neighbors and a training testing split of 20 percent found an R^2 value of 0.9672131147540983.
    - A DecisionTreeClassifier check using an entropy criterion, a maximum depth of 7, and a minimum weight fraction leaf of 0.001 with a training testing split of 10 percent found an R^2 value of 1.0.
    - A RandomForestClassifier check using 24 estimators with a training testing split of 20 percent found an R^2 value of 0.9508196721311475.

- A paragraph explaining which of your classifier models you recommend as a final model that best fits your needs in terms of accuracy and explainability:
    - The best classifier to chose would be the decision tree classifier. The decision tree classifier could be searched through in order to build a stronger picture of what is happening inside. Further, a 100 percent accuracy was found using this classifier.
- Summary Key Findings and Insights, which walks your reader through the main drivers of your model and insights from your data derived from your classifier model.
    - There are many different models that can be used for this sort of problem, with some models appearing much more effective then others in terms of accuracy.
    - SVC, KNeighborsClassifier, DecisionTreeClassifier, and RandomForestClassifier models work the best overall.
    - It is possible to train the model to predict the diagnosis of heart disease.
- Suggestions for next steps in analyzing this data, which may include suggesting revisiting this model after adding specific data features that may help you achieve a better explanation or a better prediction:
    - Using cross validation rather then training and testing splits.
    - Including more features related to heart disease in the dataset.
    - Including more rows or data entries to test and train the data on.
    - Visualizing the results to build a stronger understanding of the data correlations with heart disease.
    - Checking the data for outliers within some acceptable margin or error.