Resources:

1. Data cleaning: <https://www.shanelynn.ie/using-pandas-dataframe-creating-editing-viewing-data-in-python/>
2. seaborn: statistical data visualization: <https://seaborn.pydata.org/>
3. Classification: ROC and AUC: <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc>
4. <https://www.datacamp.com/community/tutorials/understanding-logistic-regression-python>
5. <http://www.data-mania.com/blog/logistic-regression-example-in-python/>
6. <https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/>
7. <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc>
8. <http://logisticregressionanalysis.com/909-understanding-logistic-regression-outputpart-4-making-predictions/>
9. <https://stats.stackexchange.com/questions/203740/logistic-regression-scikit-learn-vs-statsmodels>
10. <https://www.kaggle.com/jasontsmith2718/predicting-heart-disease>
11. <https://datascienceplus.com/building-a-logistic-regression-in-python-step-by-step/>
12. <https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html#sklearn.linear_model.LogisticRegression.decision_function>
13. <https://towardsdatascience.com/building-a-logistic-regression-in-python-step-by-step-becd4d56c9c8>
14. <https://www.drivendata.org/competitions/54/machine-learning-with-a-heart/page/109/#metric>
15. <http://glemaitre.github.io/imbalanced-learn/generated/imblearn.over_sampling.SMOTE.html>
16. <https://www.datacamp.com/community/tutorials/python-data-type-conversion>

Prediction Equation:

heart\_disease\_present =

0.35281083 \* slope\_of\_peak\_exercise\_st\_segment)

+ (0.00134272 \* resting\_blood\_pressure)

+ (0.61461895 \* chest\_pain\_type)

+ (0.97321156 \* num\_major\_vessels)

- (0.21758949 \* fasting\_blood\_sugar\_gt\_120\_mg\_per\_dl)

+ (0.15992521 \* resting\_ekg\_results)

+ (0.00381237 \* serum\_cholesterol\_mg\_per\_dl)

+ (0.54244358 \*oldpeak\_eq\_st\_depression)

+ (1.33501651 \* sex)

- (0.04351537 \* age)

- (0.02687004 \* max\_heart\_rate\_achieved)

+ (1.04901422 \* exercise\_induced\_angina)

Result: 0 = No Heart Disease Present (All Negative Values considered as zero)

1 = Yes Heart Disease Present (All Positive Values considered as one)

Model:

Accuracy: 0.8650793650793651

Precision: 0.8666666666666667