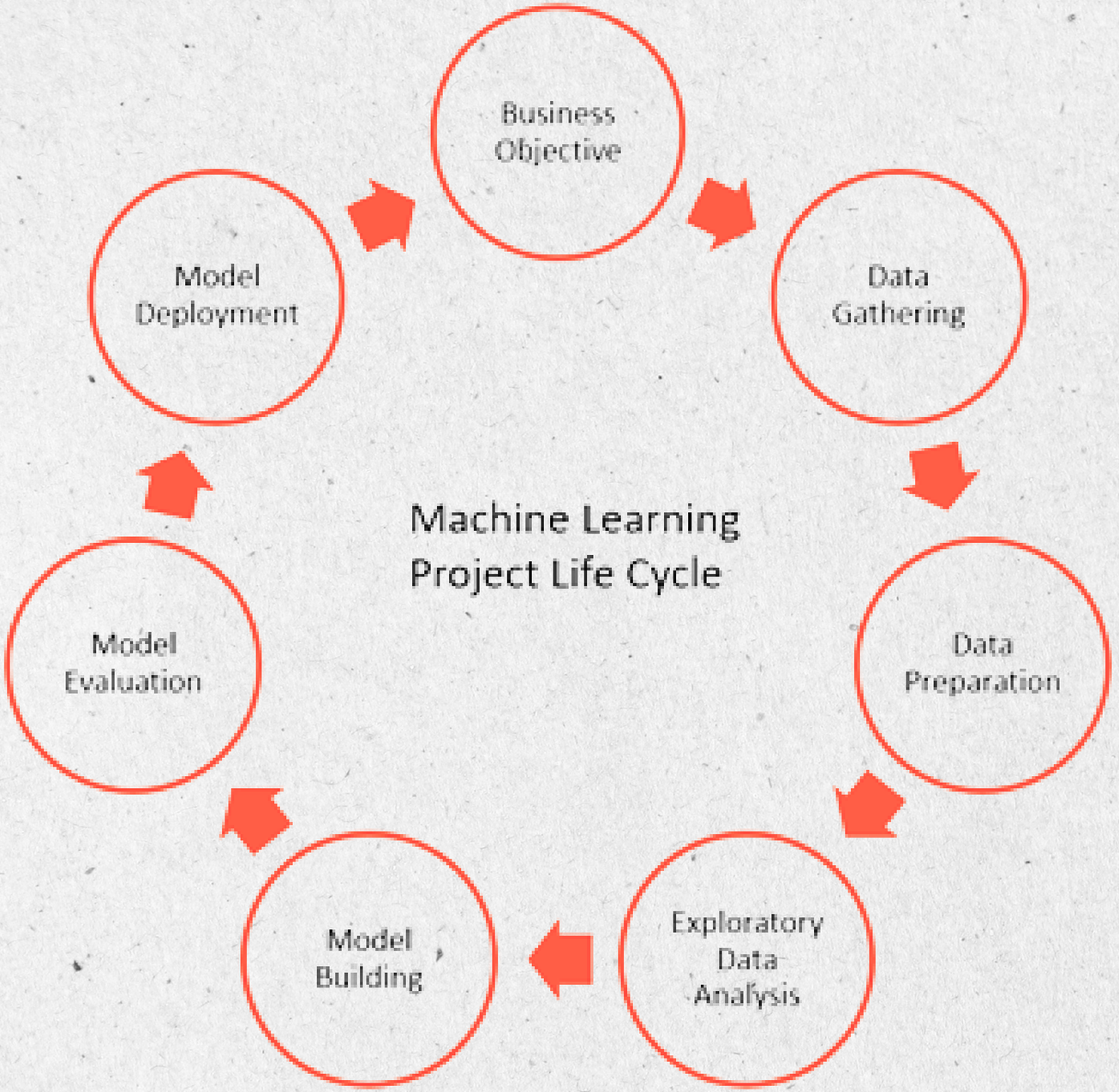
PDS AIML- B2

MULTIPLE DISEASE PREDICTION (PARKINSON'S , DIABETES AND HEART DISEASES) WITH FRONTEND DEPLOYMENT

SNEH PATEL – 22070126113
SHUBH AGRAWAL – 22070126108
SHREYASEE SHINDE – 22070126106

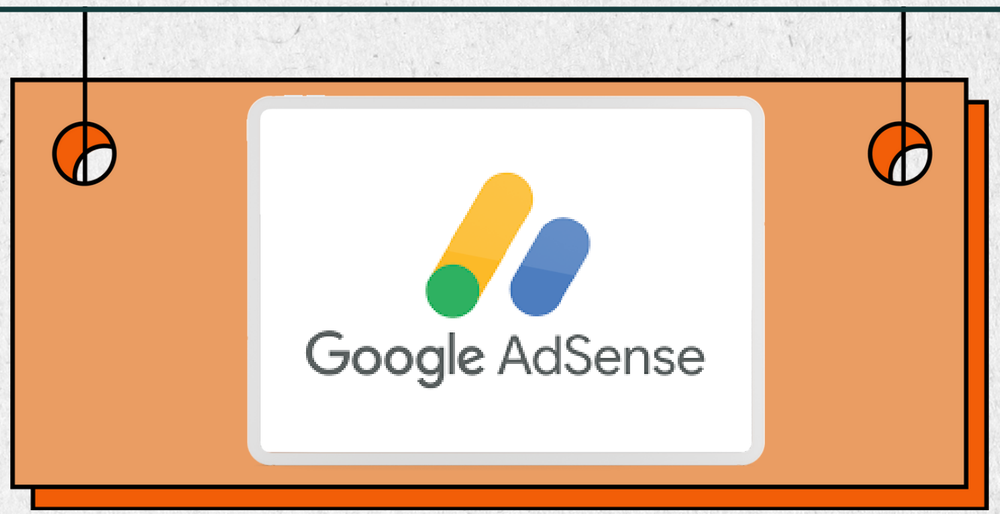
MACHINE LEARNING LIFE CYCLE

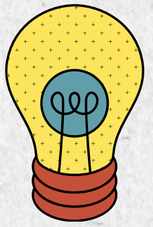


I)BUSINESS OBJECTIVES

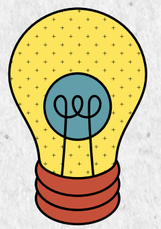
INCOME SOURCE : Google AdSense

Creating a web application for predicting multiple diseases (Parkinson's, diabetes, and heart diseases) and incorporating Google AdSense for monetization including building the backend, frontend, integrating machine learning models, and adding Google AdSense advertisements.





Increased Consultation Demand



Customized Wellness Plans

2) DATA GATHERING

1) DIABETES DISEASE PREDICTION:
https://www.kaggle.com/datasets/saurabh00007/diabetescsv

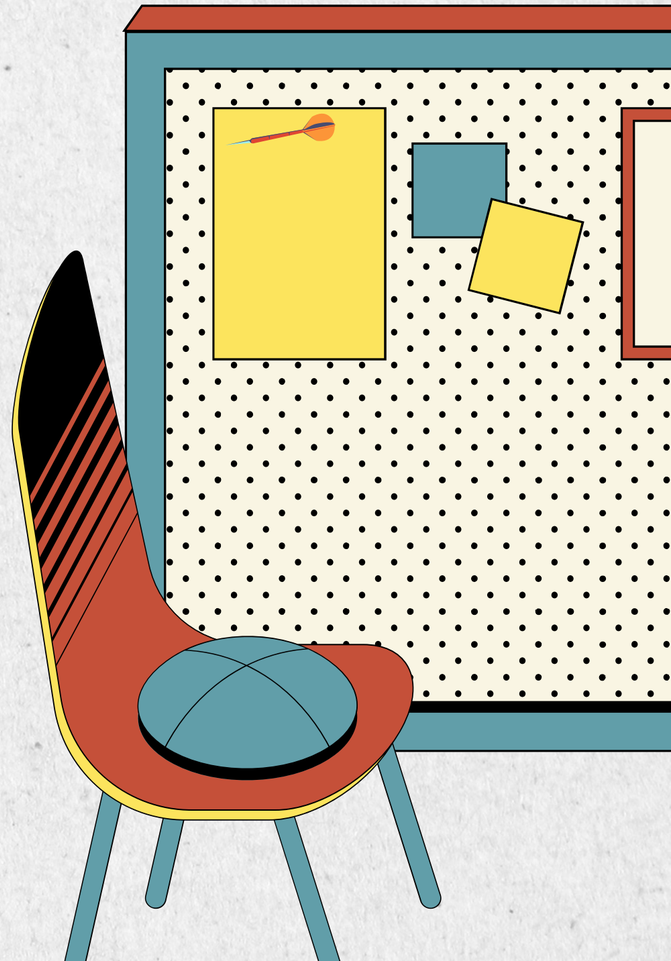
2) PARKINSON's DISEASE PREDICTION:
https://www.kaggle.com/datasets/debasisdotcom/parkinson-disease-detection

3) HEART DISEASE PREDICTION:
https://archive.ics.uci.edu/dataset/45/heart+disease

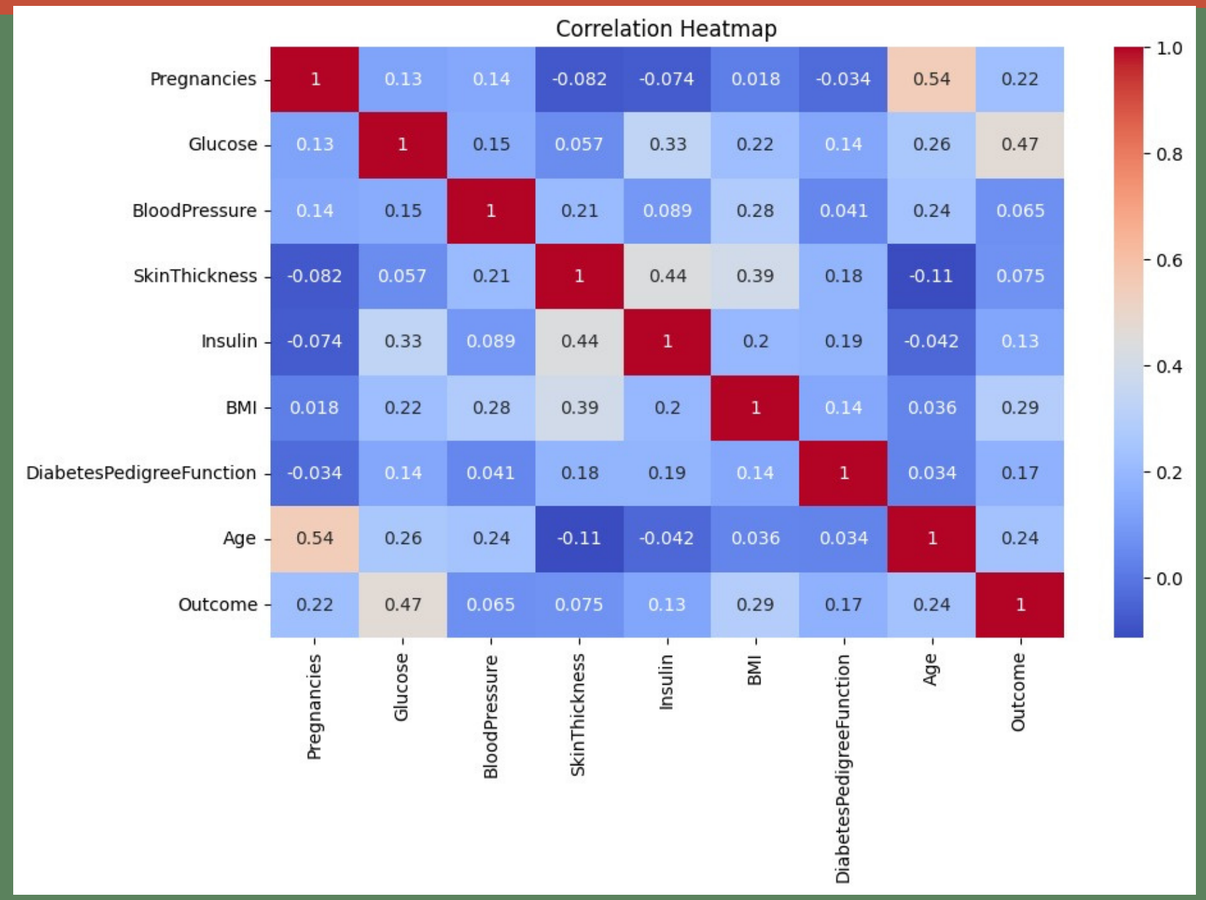
MACHINE LEARNING MODEL OF EACH DISEASE WLL BE CREATED USING THESE DATASETS AND WILL BE COMINBED TO FORM A FINAL WEBSITE FOR MULTIPLE DISEASE PREDICTION.

DATA PREPARATION

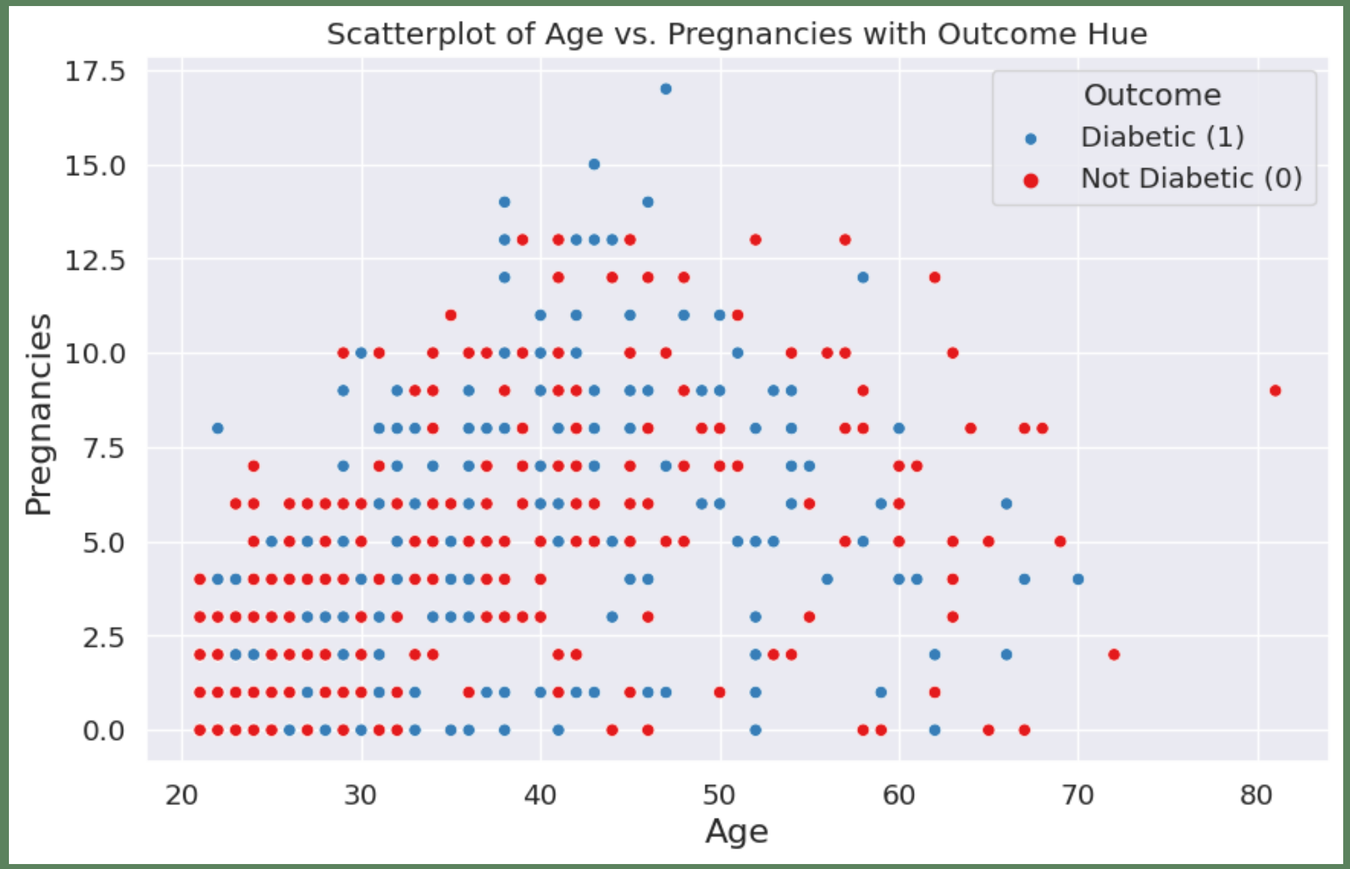
Firstly we checked for NULL values in our 3 datasets using some common function such as isnull() and .sum() , there were no null values found in our datasets , Thus we performed some DPL tasks such as .describe() , .head() , .tail() , .info() to get some particular information like attributes , statistical measures (mean , median , quartiles and standard deviation) .



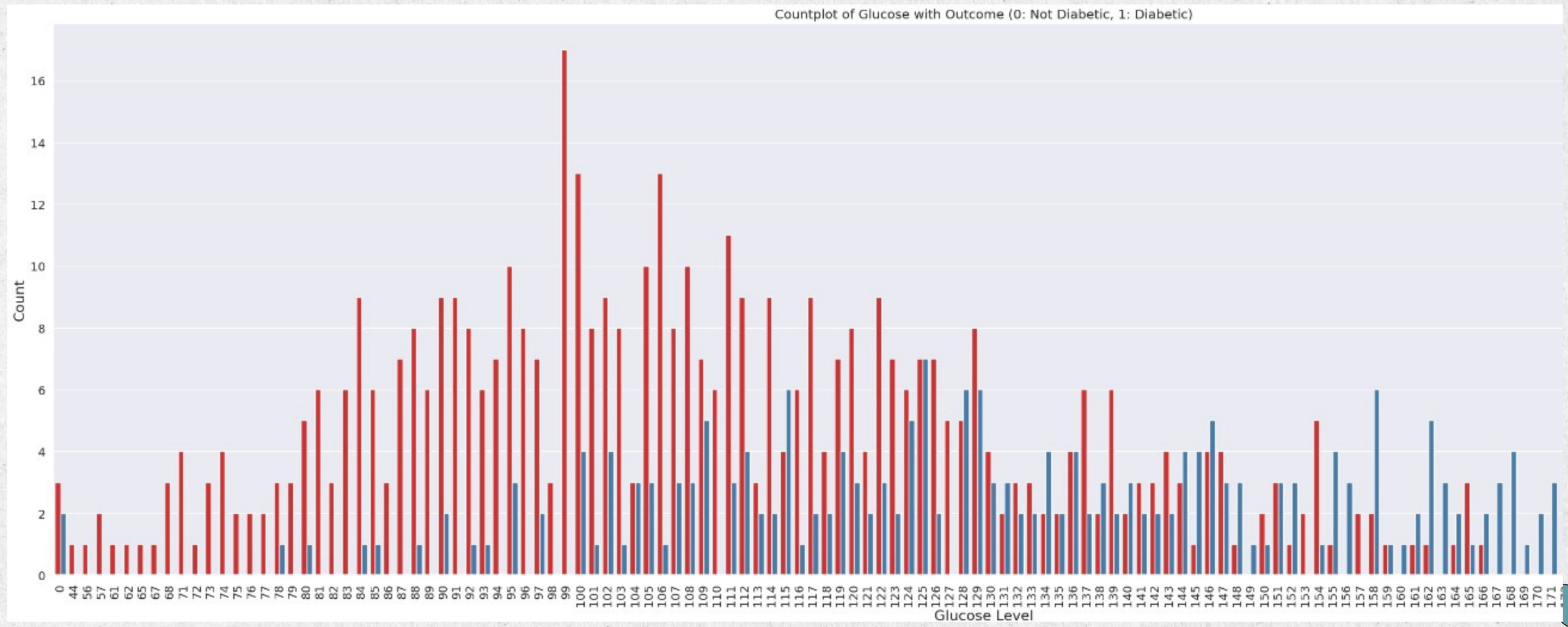
EDA GRAPHS OF DATASET (DIABETES)



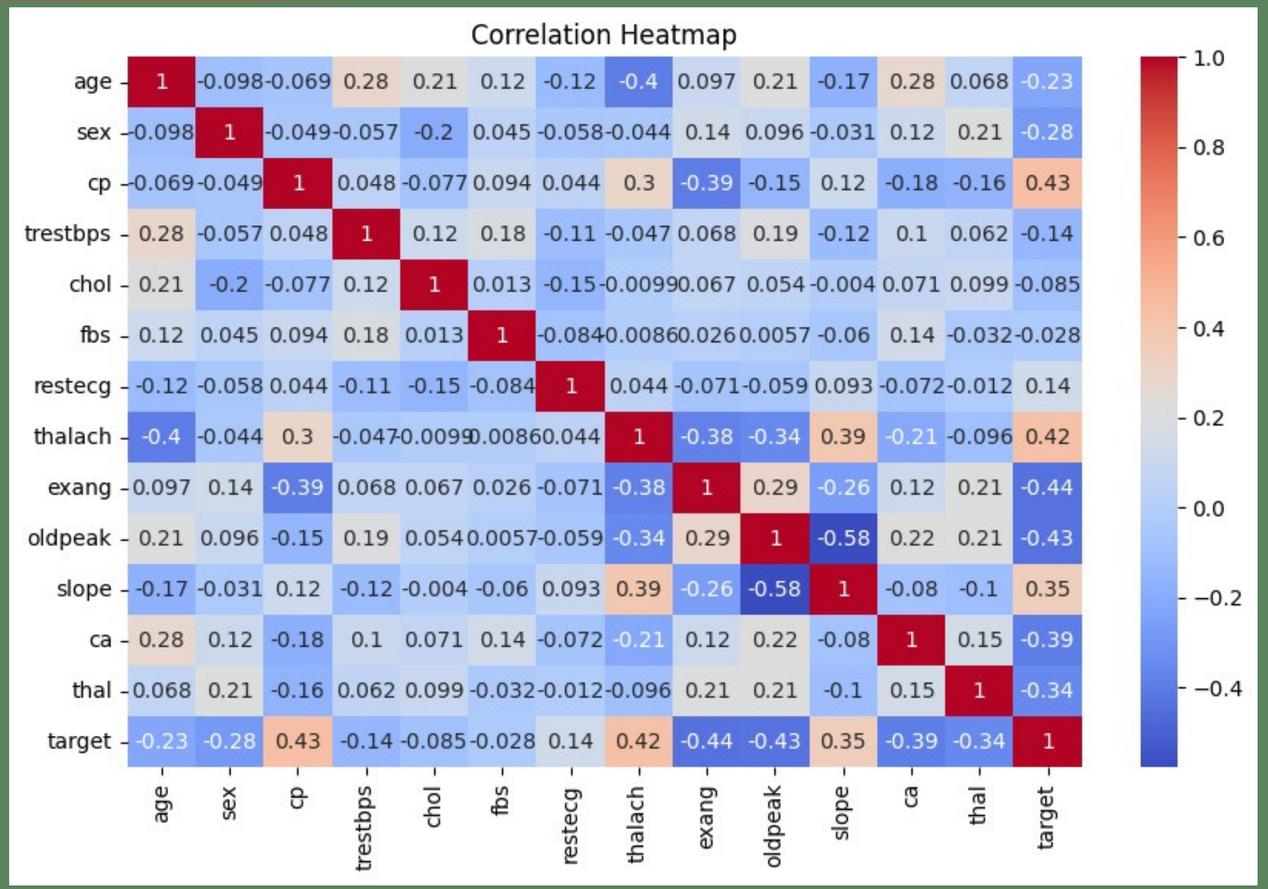
EDA GRAPHS OF DATASET (DIABETES)



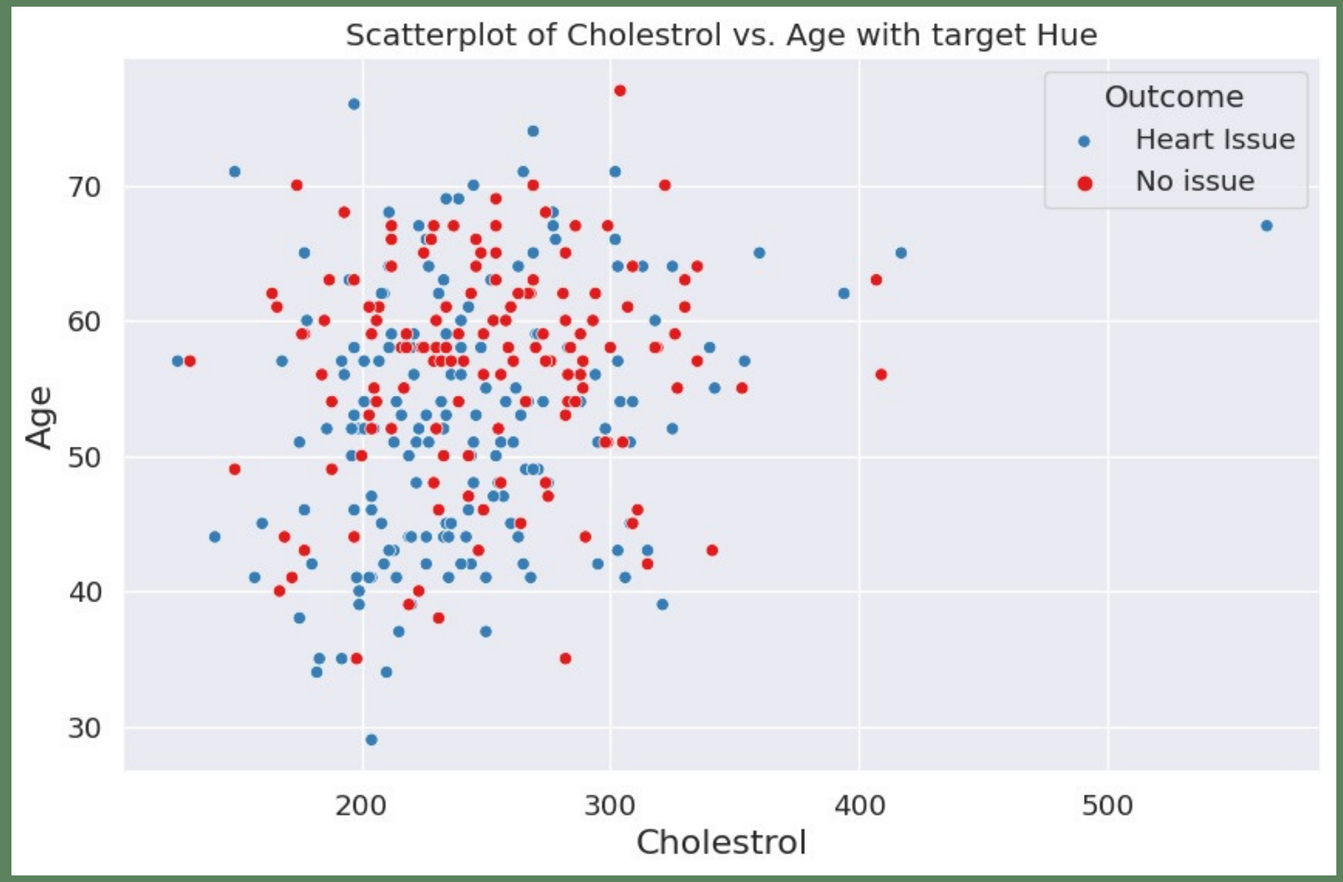
EDA GRAPHS OF DATASET (DIABETES)



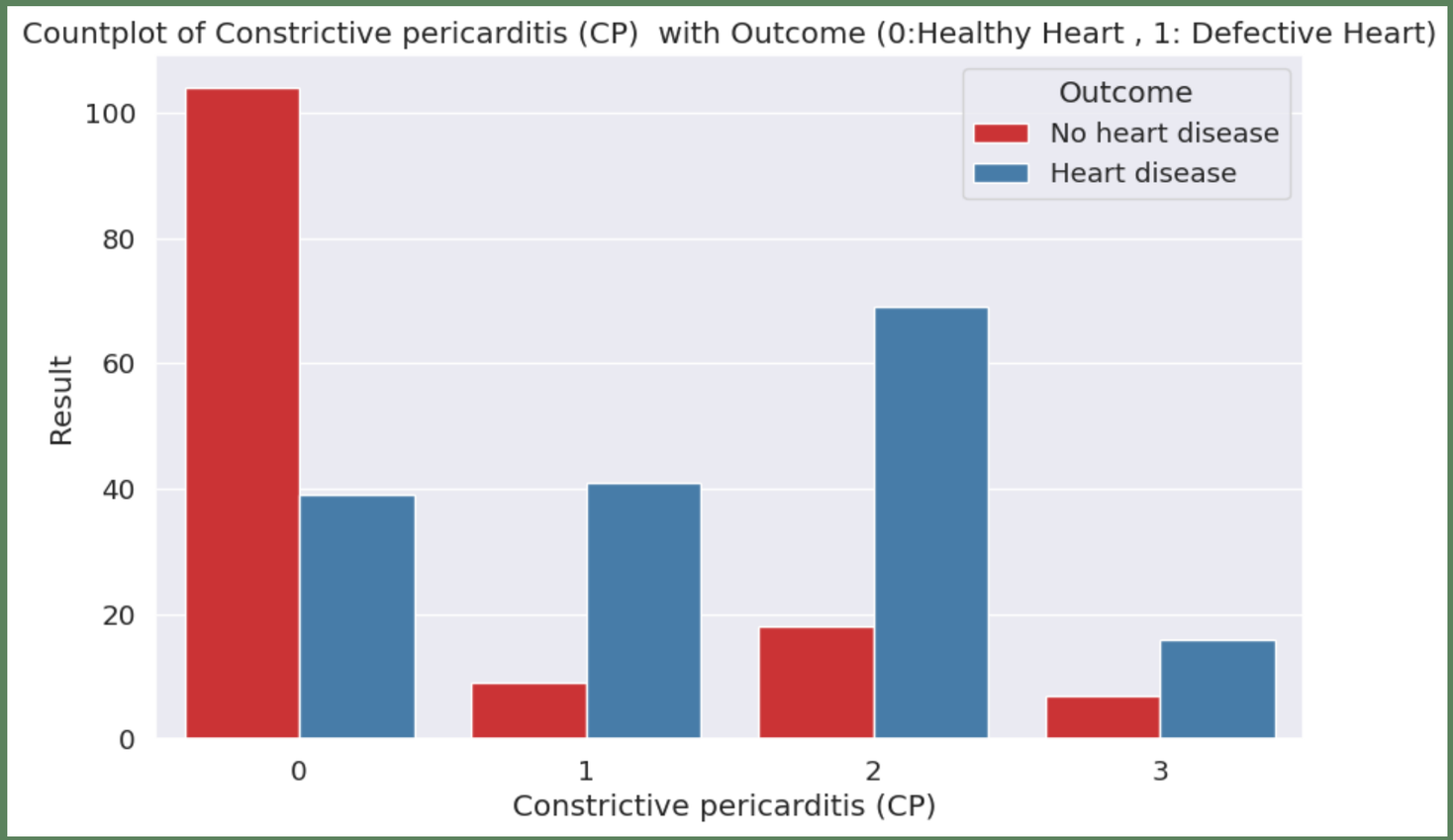
EDA GRAPHS OF DATASET (HEART)



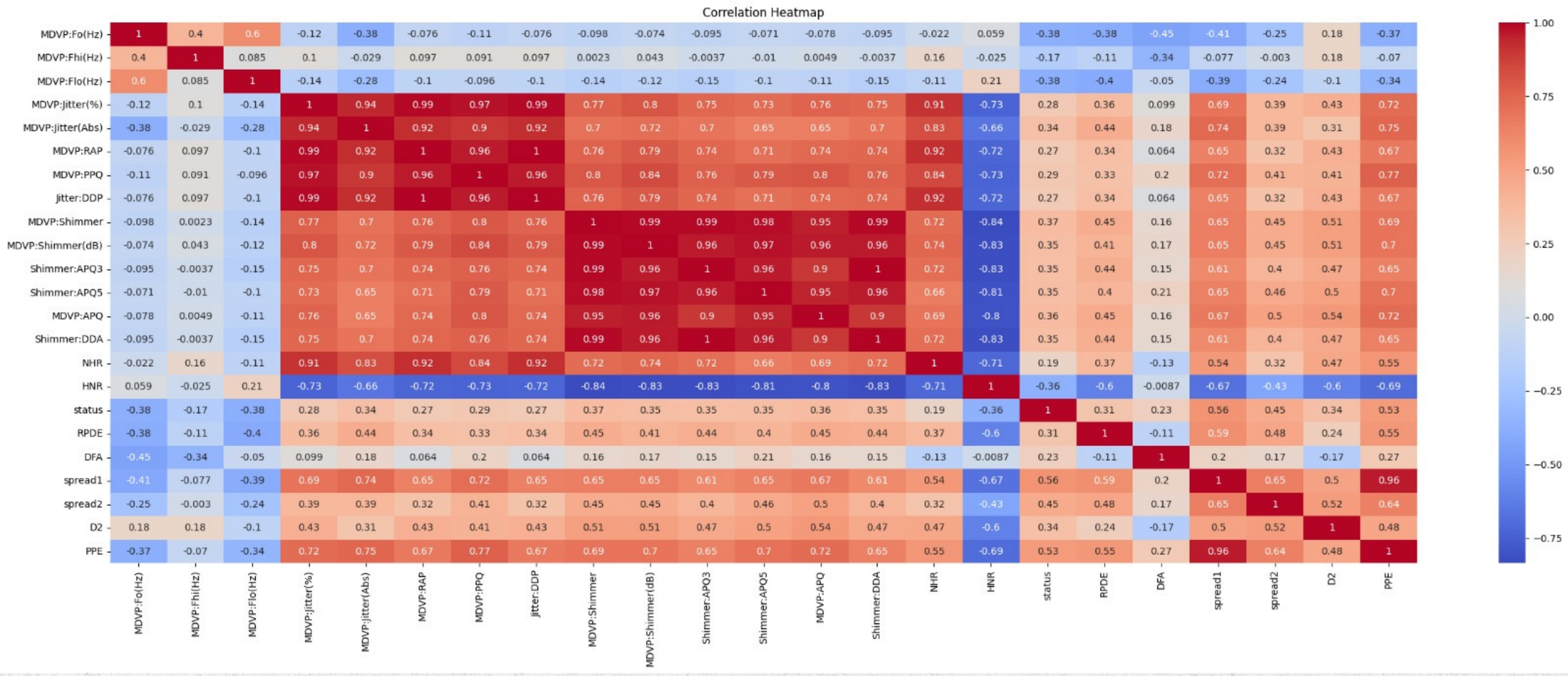
EDA GRAPHS OF DATASET (HEART)



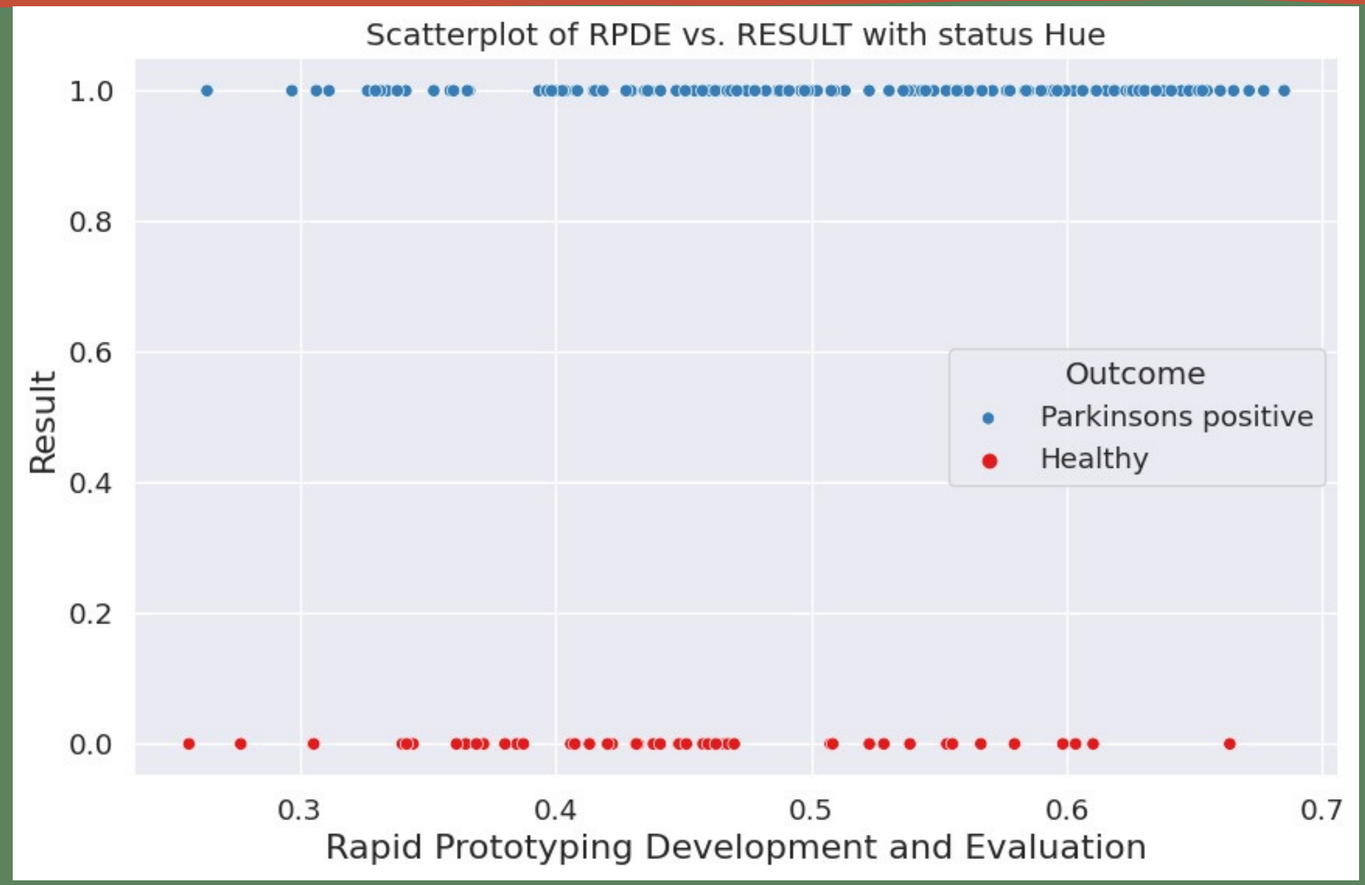
EDA GRAPHS OF DATASET (HEART)



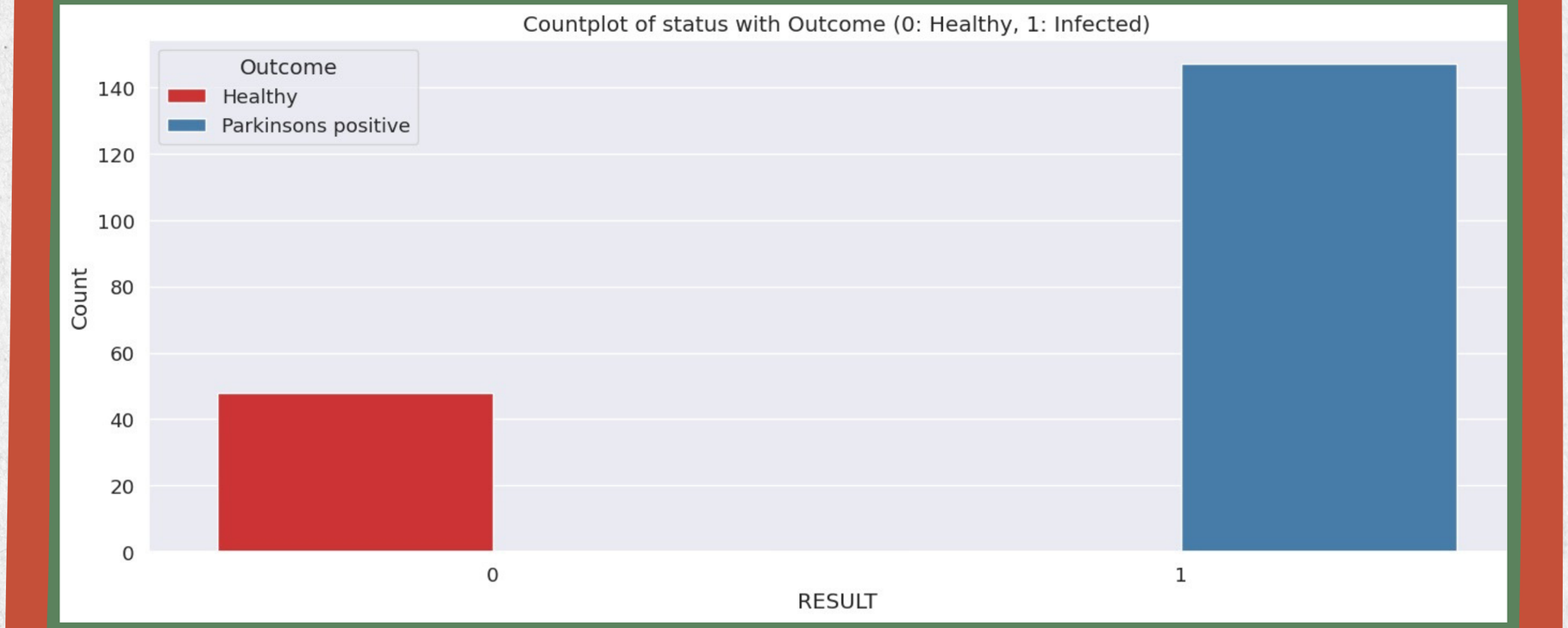
EDA GRAPHS OF DATASET (PARKINSON)



EDA GRAPHS OF DATASET (PARKINSON)



EDA GRAPHS OF DATASET (PARKINSON)



MODEL BUILDING

DIABETES DISEASE PREDICTION

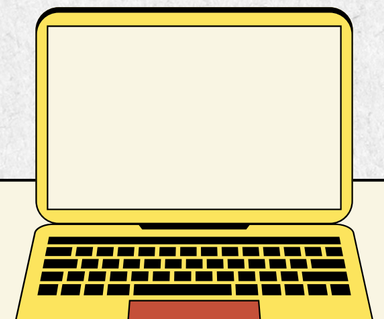
- Splitting the Data into Training data & Test Data (80:20).
- Naive Bayes classifier is used.
- Naive Bayes is preferred for diabetes prediction for its simplicity, efficiency, probabilistic outputs, and suitability for mixed data.

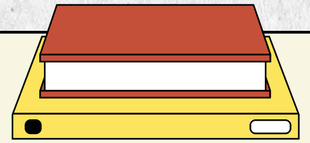
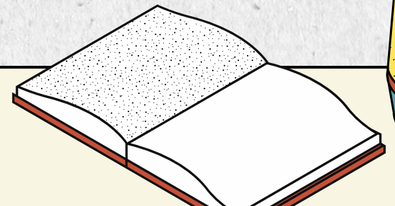
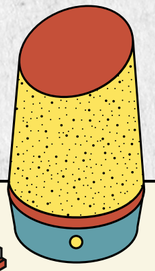
HEART DISEASE PREDICTION

- Splitting the Data into Training data & Test data (80:20).
- LogisticRegression is used.
- LogisticRegression suits heart disease prediction due to its ability to model binary outcomes, interpretability, and minimal overfitting risk.

PARKINSON DISEASE PREDICTION

- Splitting the Data into Training data & Test data (80:20).
- SVM classifier is used.
- SVM is suitable for Parkinson's disease prediction due to its capacity to find complex decision boundaries and handle non-linear relationships in data.




MODEL EVALUTAION

This is the accuracy score of the trained diabetes model.

```
Accuracy score of the training data: 75.57%

Accuracy score of the test data: 77.27%
```

This is the prediction obtained by feeding input in our trained diabetes model.

```
The person is diabetic
Probability of being diabetic for the input data: 67.42%
```

MODEL EVALUTAION

This is the accuracy score of the trained heart model.

```
Accuracy score of the training data: 85.12%
```

```
Accuracy score of the test data: 81.97%
```

This is the prediction obtained by feeding input in our trained heart model.

```
[0]
The Person does not have a Heart Disease
```

MODEL EVALUTAION

This is the accuracy score of the trained Parkinson model.

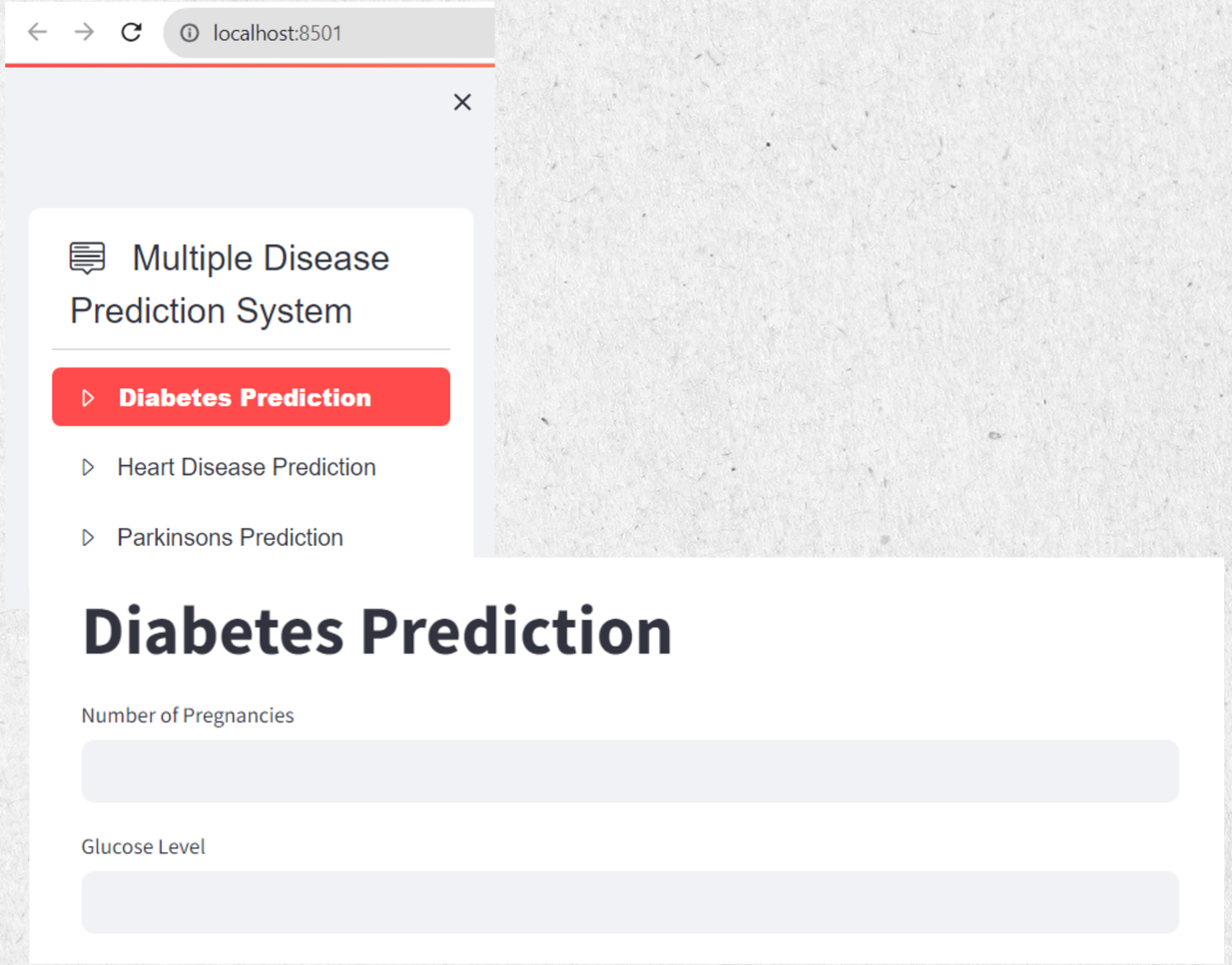
```
Accuracy score of the training data: 88.46%
```

```
Accuracy score of test data : 87.18%
```

This is the prediction obtained by feeding input in our trained Parkinson model.

```
[0]
The Person does not have Parkinsons Disease
```

MODEL DEPLOYMENT

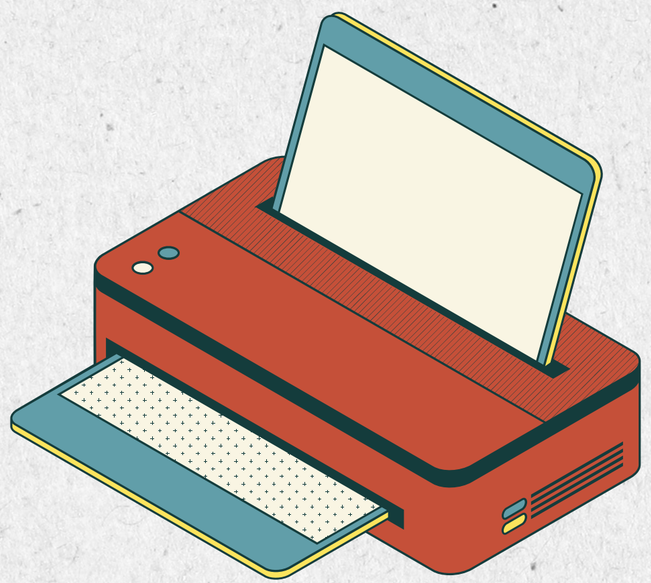


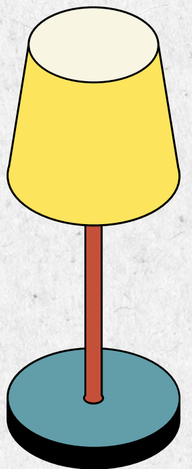
Model Deployment was done using Streamlit which is an open-source Python library for building web applications with simple code, primarily used for data visualization.

In this deployment we have used libraries such as :

- Pickle:- It has been used for loading models of three diseases in spyder .
- streamlit:- This package has been used for frontend deployment whih directly links to the localhost .
- streamlit options menu:- This package has been used for navigating to a particular disease to be predicted .

REFRENCES:





- **[1]** For frontend and to gain knowledge for the deployment using streamlit we refered to a particular video. https://www.youtube.com/watch?v=8Q_QQVQ1HZA&list=PLfFghEzKVmjvuSA67LszN1dZ-Dd_pkus6&index=26

- **[2]** Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical image analysis, 42, 60-88. https://www.sciencedirect.com/science/article/pii/S1361841517301127

- **[3]** Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. https://www.deeplearningbook.org/

- **[4]** Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European conference on computer vision (ECCV). https://openaccess.thecvf.com/content_ECCV_2018/html/Liang-Chieh_Chen_EncoderDecoder_with_Atrous_ECCV_2018_paper.html

THANK YOU!