**Capstone Project**

**Problem Statement**

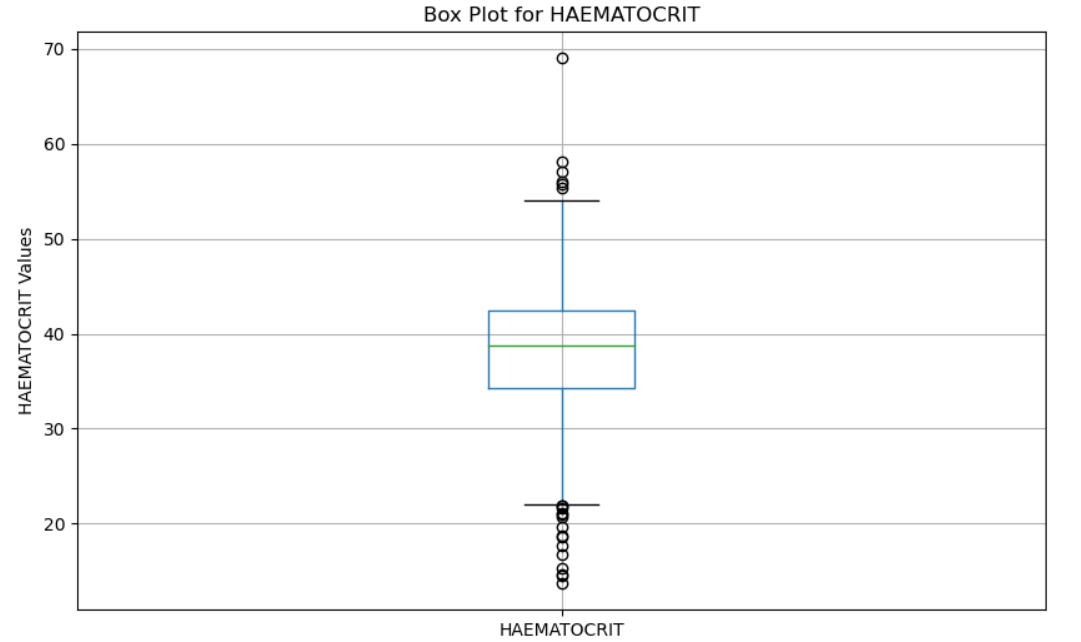
In this project we make one machine learning model that predict patient is well or not after seeing all patient report counts.

* We get this data from kaggle Website
* In this file we got 11 columns
* HAEMATOCRIT
* HAEMOGLOBINS
* ERYTHROCYTE
* LEUCOCYTE
* THROMBOCYTE
* MCH
* MCHC
* MCV
* AGE
* SEX
* SOURCE
* This SOURCE columns is our output column in data set
* Data count of this data set is **3309**

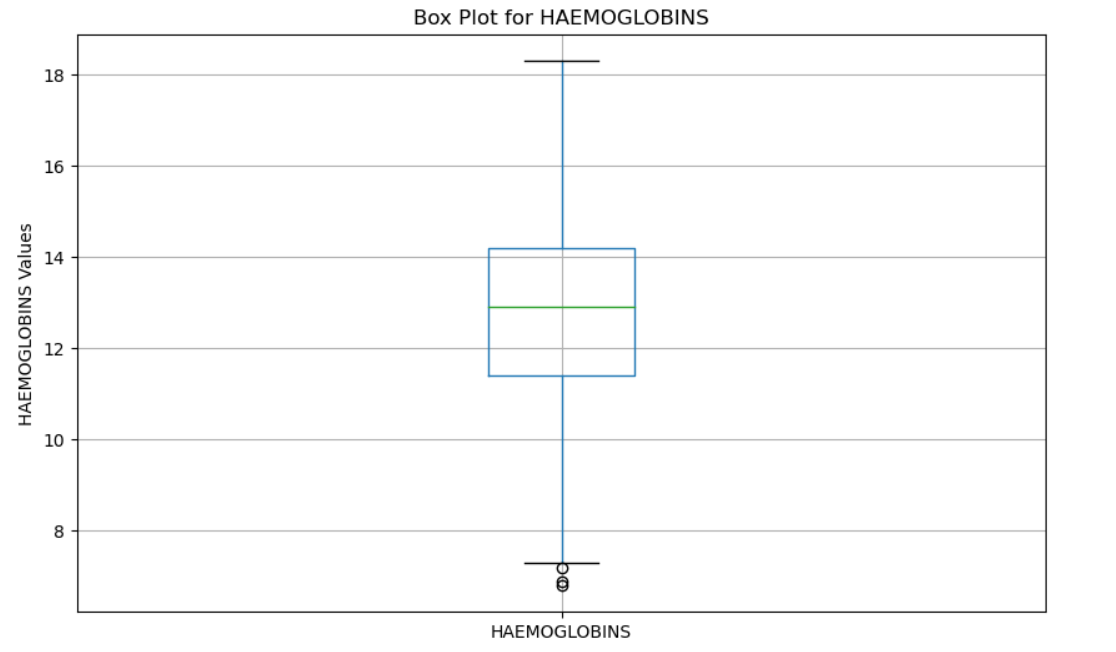
**EDA**

For starting the EDA process we get knowledge of all patient report count from chatgpt.

* + Haematocrit (HCT):
    - Men: 38.3% to 48.6%
    - Women: 35.5% to 44.9%
  + Haemoglobin (Hb or HGB):
    - Men: 13.8 to 17.2 grams per deciliter (g/dL)
    - Women: 12.1 to 15.1 g/dL
  + Erythrocyte (Red Blood Cell) Count:
    - Men: 4.5 to 5.5 million cells/mcL
    - Women: 4.0 to 5.0 million cells/mcL
  + Leucocyte (White Blood Cell) Count:
    - 4,000 to 11,000 cells/mcL
  + Thrombocyte (Platelet) Count:
    - 150,000 to 450,000 cells/mcL
  + MCH (Mean Corpuscular Hemoglobin):
    - 27 to 33 picograms
  + MCHC (Mean Corpuscular Hemoglobin Concentration):
    - 32% to 36%
  + MCV (Mean Corpuscular Volume):
    - 80 to 100 femtoliters
  + Frist we check for null values for there is not any null values.
  + Then we go for remove outliner from dataset.

****

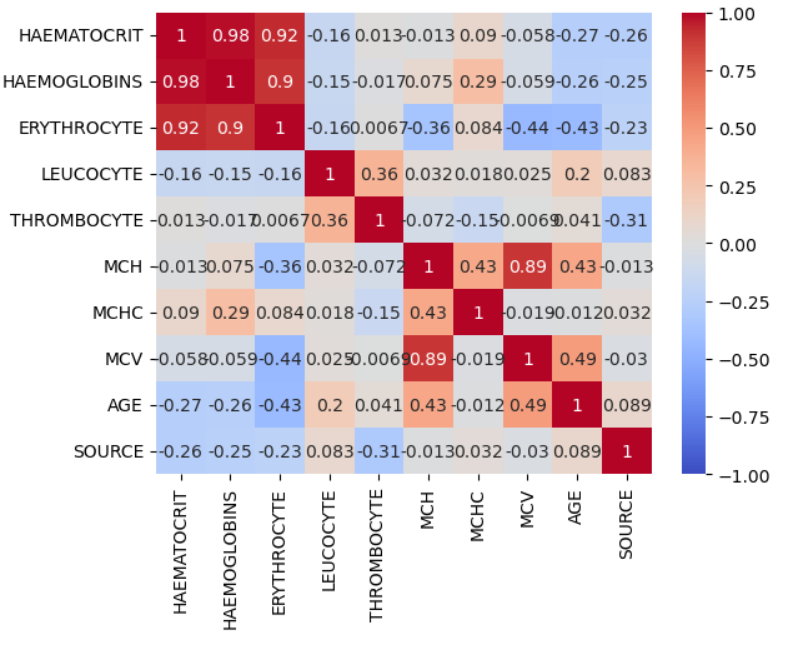
* + In this HAEMATOCIT columns we find outliner
  + so here you can see that there are outliers in that column so we remove the data which has value less than 23 in this columns or greater than 54.
  + check for next column



# In this HENOGLOBINS columns we find outeliners so we remove the data which has value less than 9 in this columns or greter than 18

* + Same why we check all the columns and remove outliners.
  + After removeing outliners our data count is **2643**

**HEATMAP**



**Model Selection**

* + After seeing our dataset and output column we know that this is classification problem.
  + For that classification problem there is some machine learning algorithm is there.
  + Like SVC , Decision Tree, Random Forest,Logistic Regression so we try all the algorithm and final logistic refression.