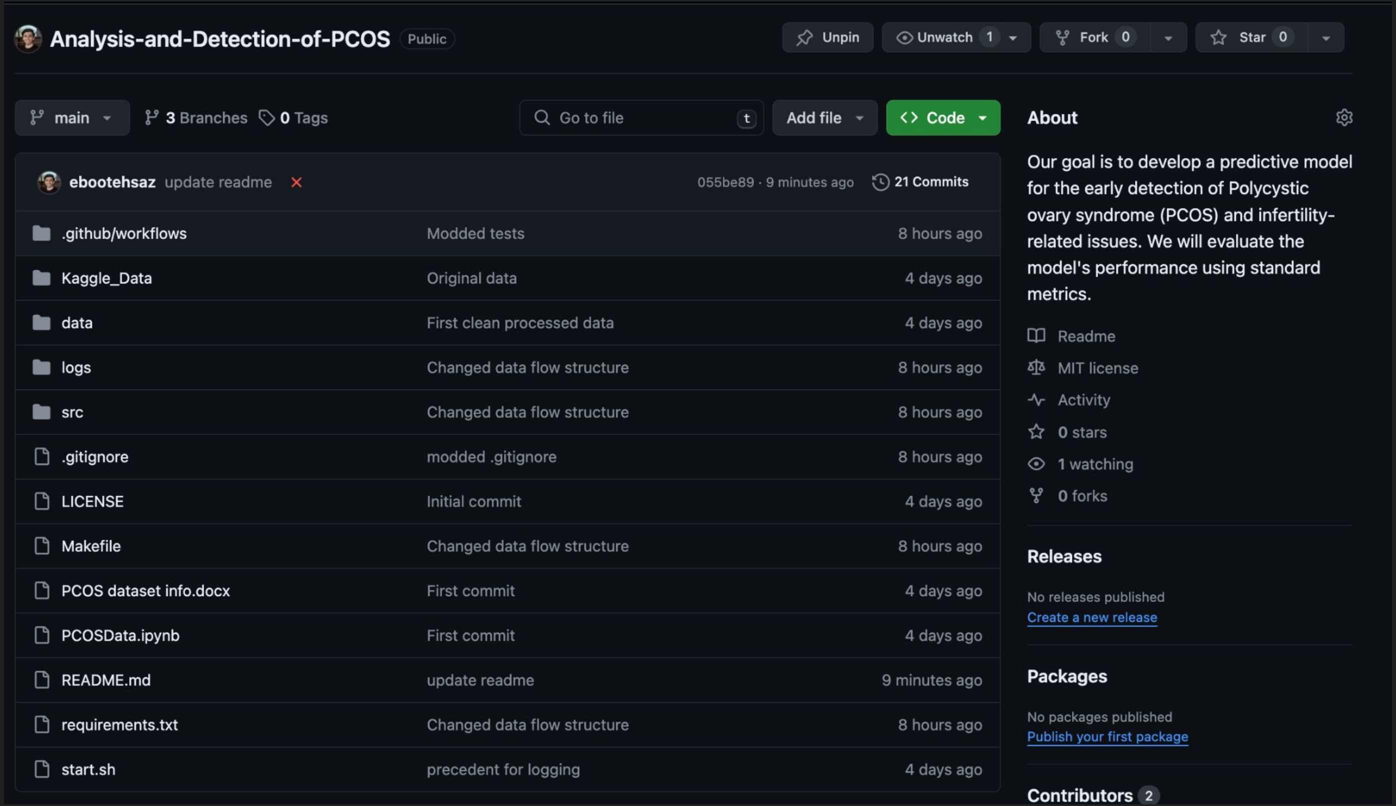
Detection and Analysis of Polycystic Ovary Syndrome

Data Science Boot Camp

Ethan Bootehsaz, Kaiyuan Ma, Joy Wang, Aaron Yang

04/02/2024

Created a Repository



ebootehsaz update readme ✕ 055be89 · 9 minutes ago 21 Commits

.github/workflows	Modded tests	8 hours ago
Kaggle_Data	Original data	4 days ago
data	First clean processed data	4 days ago
logs	Changed data flow structure	8 hours ago
src	Changed data flow structure	8 hours ago
.gitignore	modded .gitignore	8 hours ago
LICENSE	Initial commit	4 days ago
Makefile	Changed data flow structure	8 hours ago
PCOS dataset info.docx	First commit	4 days ago
PCOSData.ipynb	First commit	4 days ago
README.md	update readme	9 minutes ago
requirements.txt	Changed data flow structure	8 hours ago
start.sh	precedent for logging	4 days ago

About

Our goal is to develop a predictive model for the early detection of Polycystic ovary syndrome (PCOS) and infertility-related issues. We will evaluate the model's performance using standard metrics.

- Readme
- MIT license
- Activity
- 0 stars
- 1 watching
- 0 forks

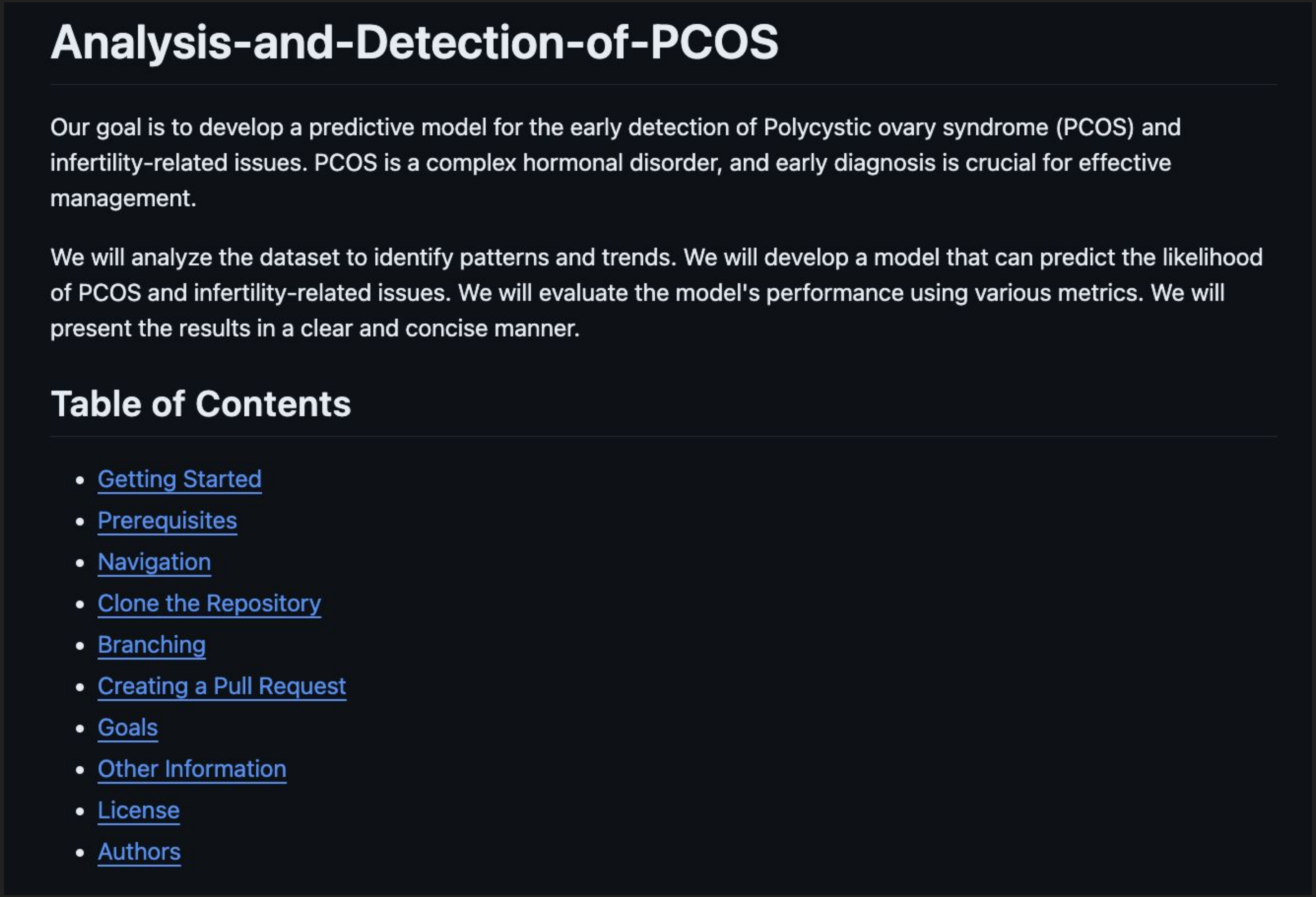
Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Contributors 2

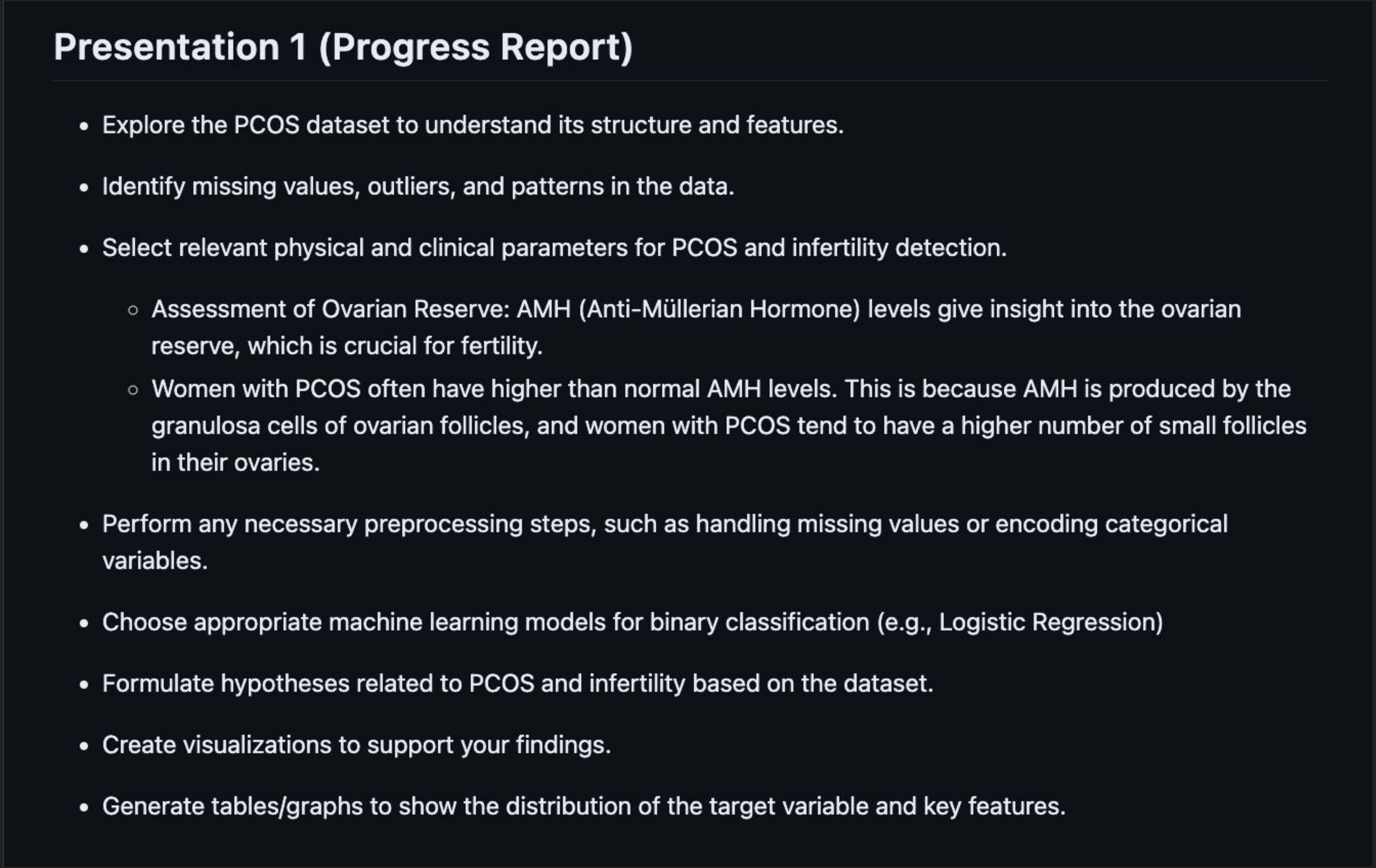


Analysis-and-Detection-of-PCOS

Our goal is to develop a predictive model for the early detection of Polycystic ovary syndrome (PCOS) and infertility-related issues. PCOS is a complex hormonal disorder, and early diagnosis is crucial for effective management.

We will analyze the dataset to identify patterns and trends. We will develop a model that can predict the likelihood of PCOS and infertility-related issues. We will evaluate the model's performance using various metrics. We will present the results in a clear and concise manner.

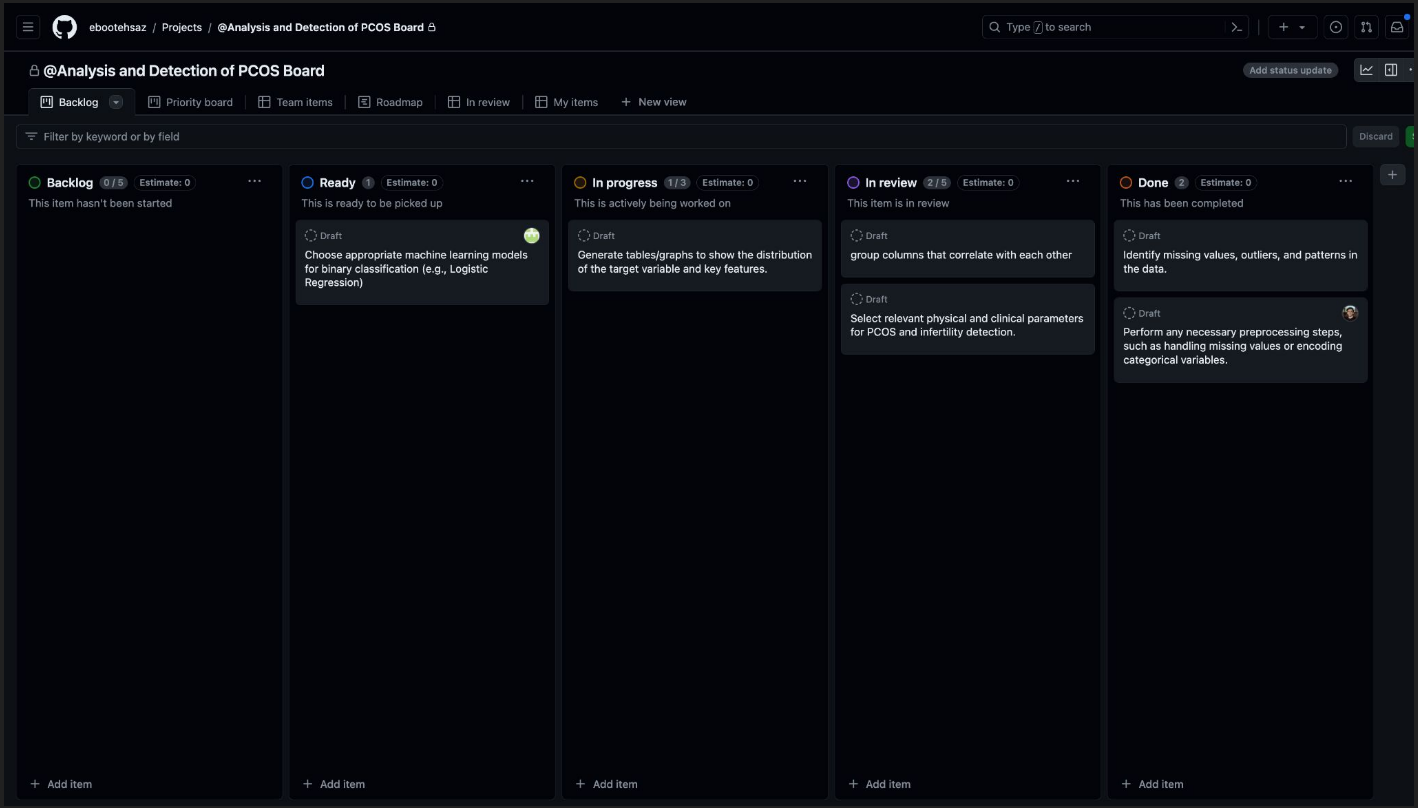
Table of Contents



Presentation 1 (Progress Report)

- Explore the PCOS dataset to understand its structure and features.
- Identify missing values, outliers, and patterns in the data.
- Select relevant physical and clinical parameters for PCOS and infertility detection.
    - Assessment of Ovarian Reserve: AMH (Anti-Müllerian Hormone) levels give insight into the ovarian reserve, which is crucial for fertility.
    - Women with PCOS often have higher than normal AMH levels. This is because AMH is produced by the granulosa cells of ovarian follicles, and women with PCOS tend to have a higher number of small follicles in their ovaries.
- Perform any necessary preprocessing steps, such as handling missing values or encoding categorical variables.
- Choose appropriate machine learning models for binary classification (e.g., Logistic Regression)
- Formulate hypotheses related to PCOS and infertility based on the dataset.
- Create visualizations to support your findings.
- Generate tables/graphs to show the distribution of the target variable and key features.

Created a Project Board



@Analysis and Detection of PCOS Board

Backlog 0 / 5 Estimate: 0
This item hasn't been started

Ready 1 Estimate: 0
This is ready to be picked up

Draft
Choose appropriate machine learning models for binary classification (e.g., Logistic Regression)

In progress 1 / 3 Estimate: 0
This is actively being worked on

Draft
Generate tables/graphs to show the distribution of the target variable and key features.

In review 2 / 5 Estimate: 0
This item is in review

Draft
group columns that correlate with each other

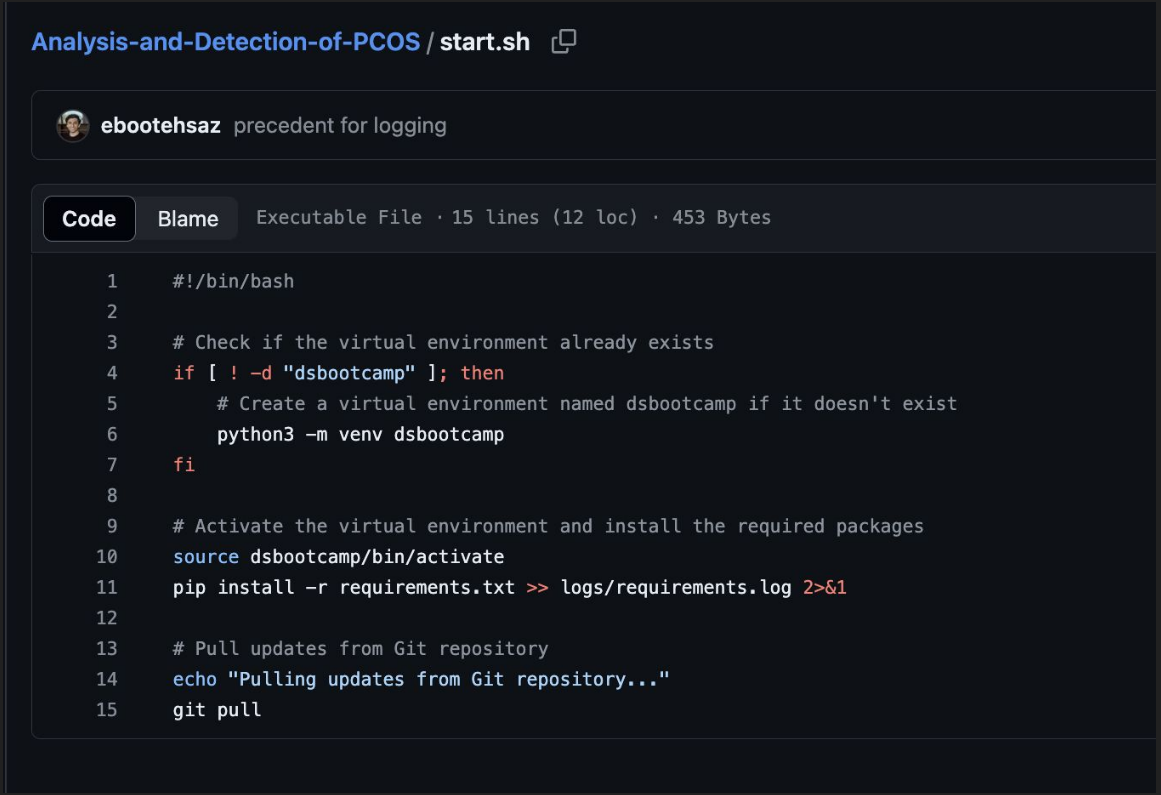
Draft
Select relevant physical and clinical parameters for PCOS and infertility detection.

Done 2 Estimate: 0
This has been completed

Draft
Identify missing values, outliers, and patterns in the data.

Draft
Perform any necessary preprocessing steps, such as handling missing values or encoding categorical variables.

Script to help get started

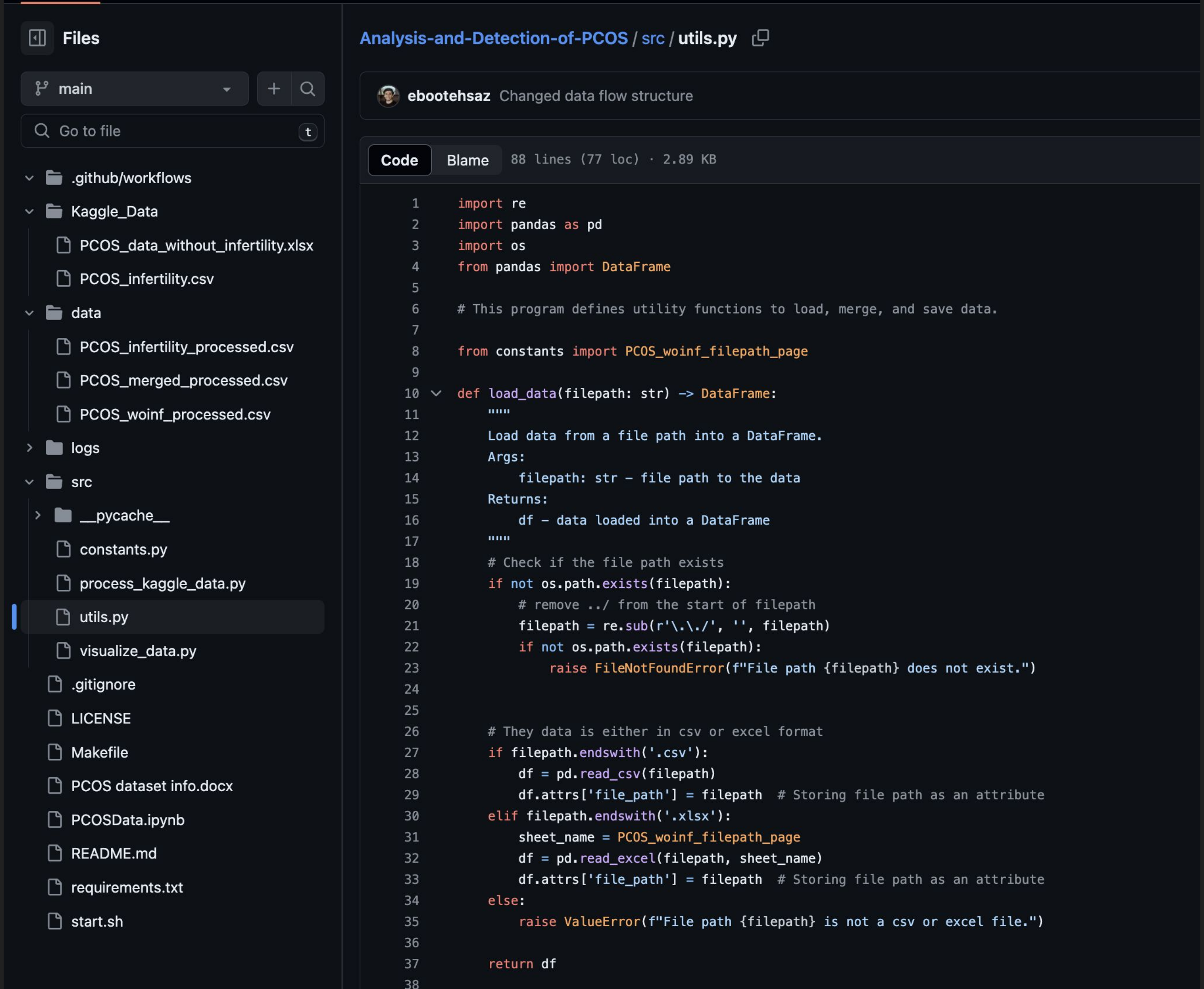


Analysis-and-Detection-of-PCOS / start.sh

ebootehsaz precedent for logging

Executable File · 15 lines (12 loc) · 453 Bytes

```bash
#!/bin/bash

# Check if the virtual environment already exists
if [ ! -d "dsbootcamp" ]; then
    # Create a virtual environment named dsbootcamp if it doesn't exist
    python3 -m venv dsbootcamp
fi

# Activate the virtual environment and install the required packages
source dsbootcamp/bin/activate
pip install -r requirements.txt >> logs/requirements.log 2>&1

# Pull updates from Git repository
echo "Pulling updates from Git repository..."
git pull
```

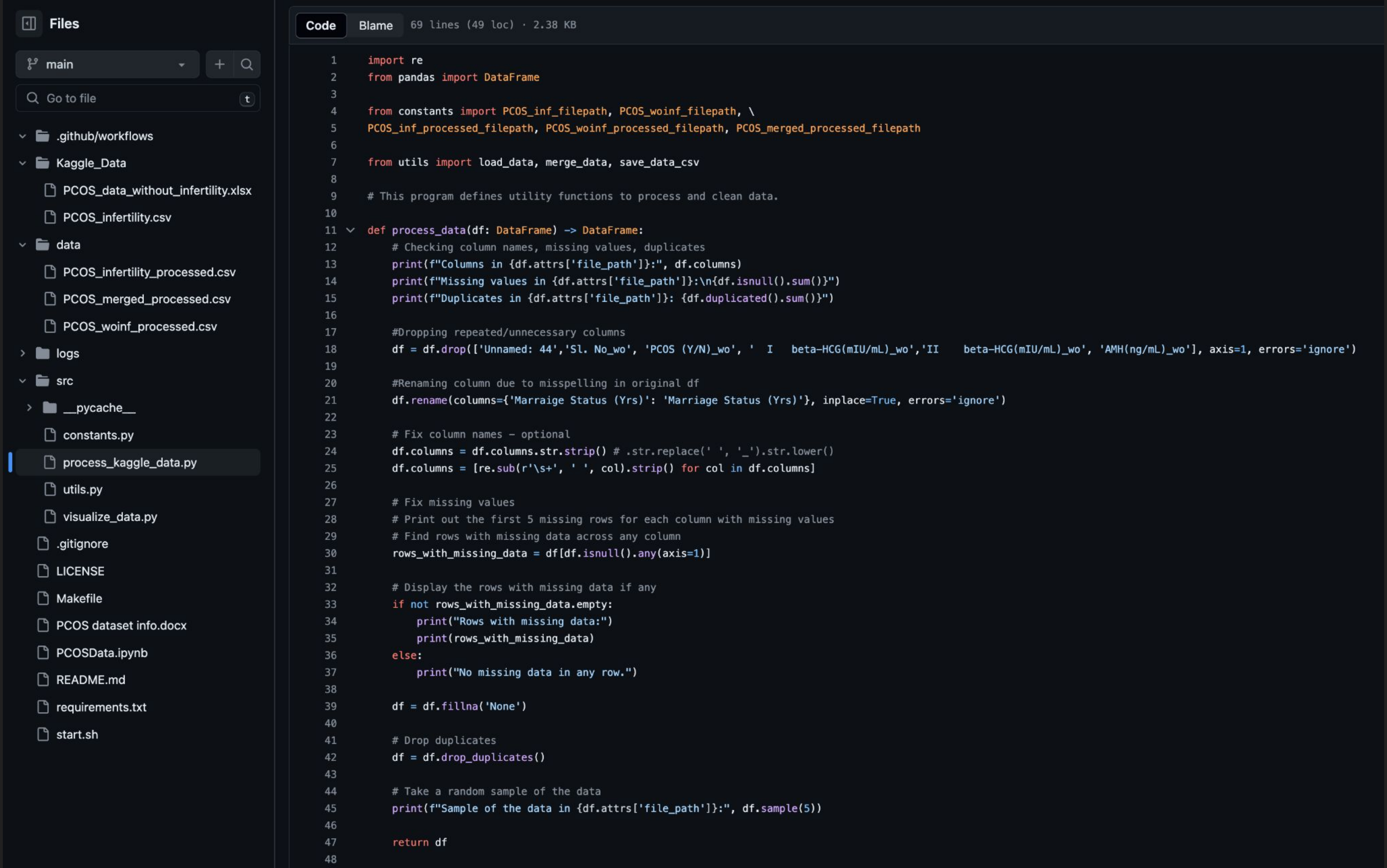


Analysis-and-Detection-of-PCOS / src / utils.py

ebootehsaz Changed data flow structure

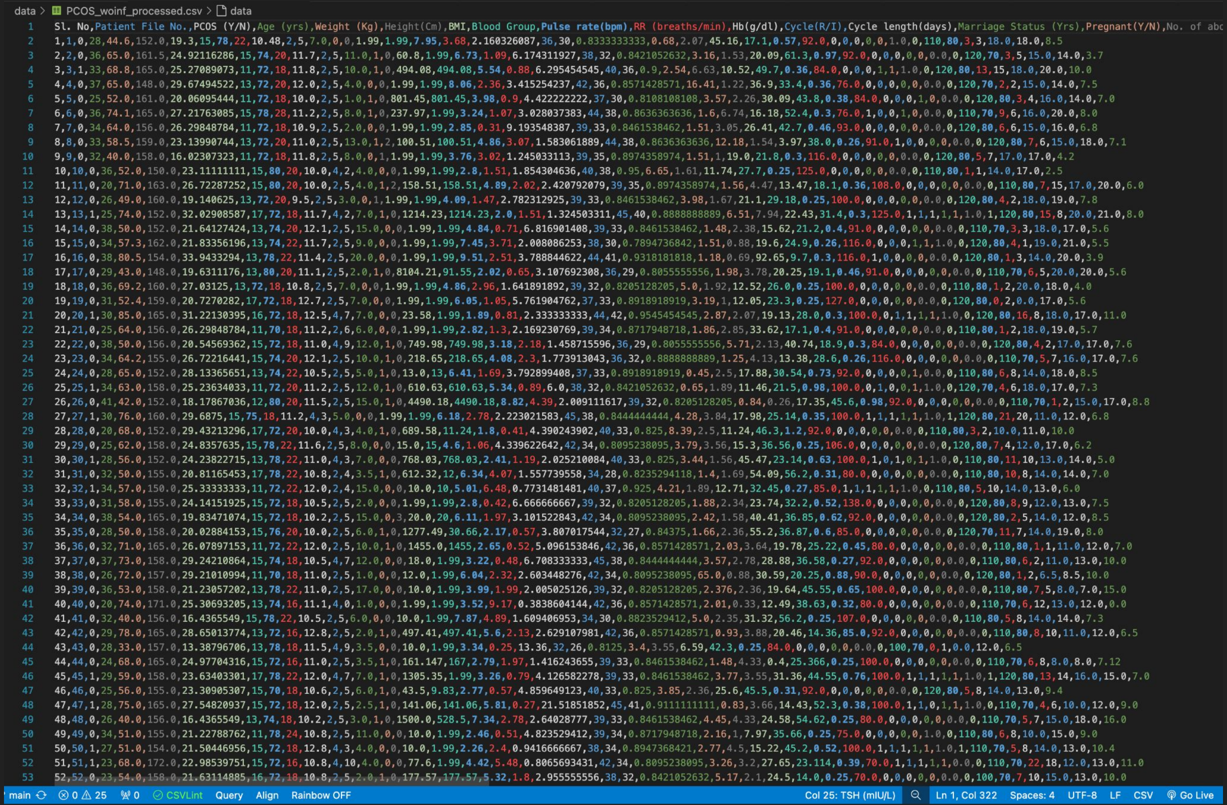
88 lines (77 loc) · 2.89 KB

```python
import re
import pandas as pd
import os
from pandas import DataFrame

# This program defines utility functions to load, merge, and save data.

from constants import PCOS_woinf_filepath_page

def load_data(filepath: str) -> DataFrame:
    """
    Load data from a file path into a DataFrame.
    Args:
        filepath: str - file path to the data
    Returns:
        df - data loaded into a DataFrame
    """
    # Check if the file path exists
    if not os.path.exists(filepath):
        # remove ../ from the start of filepath
        filepath = re.sub(r'\.\./', '', filepath)
        if not os.path.exists(filepath):
            raise FileNotFoundError(f"File path {filepath} does not exist.")


    # They data is either in csv or excel format
    if filepath.endswith('.csv'):
        df = pd.read_csv(filepath)
        df.attrs['file_path'] = filepath  # Storing file path as an attribute
    elif filepath.endswith('.xlsx'):
        sheet_name = PCOS_woinf_filepath_page
        df = pd.read_excel(filepath, sheet_name)
        df.attrs['file_path'] = filepath  # Storing file path as an attribute
    else:
        raise ValueError(f"File path {filepath} is not a csv or excel file.")

    return df

```



```python
import re
from pandas import DataFrame

from constants import PCOS_inf_filepath, PCOS_woinf_filepath, \
PCOS_inf_processed_filepath, PCOS_woinf_processed_filepath, PCOS_merged_processed_filepath

from utils import load_data, merge_data, save_data_csv

# This program defines utility functions to process and clean data.

def process_data(df: DataFrame) -> DataFrame:
    # Checking column names, missing values, duplicates
    print(f"Columns in {df.attrs['file_path']}:", df.columns)
    print(f"Missing values in {df.attrs['file_path']}:\n{df.isnull().sum()}")
    print(f"Duplicates in {df.attrs['file_path']}: {df.duplicated().sum()}")

    #Dropping repeated/unnecessary columns
    df = df.drop(['Unnamed: 44','Sl. No_wo', 'PCOS (Y/N)_wo', '  I   beta-HCG(mIU/mL)_wo','II    beta-HCG(mIU/mL)_wo', 'AMH(ng/mL)_wo'], axis=1, errors='ignore')

    #Renaming column due to misspelling in original df
    df.rename(columns={'Marraige Status (Yrs)': 'Marriage Status (Yrs)'}, inplace=True, errors='ignore')

    # Fix column names - optional
    df.columns = df.columns.str.strip() # .str.replace(' ', '_').str.lower()
    df.columns = [re.sub(r'\s+', ' ', col).strip() for col in df.columns]

    # Fix missing values
    # Print out the first 5 missing rows for each column with missing values
    # Find rows with missing data across any column
    rows_with_missing_data = df[df.isnull().any(axis=1)]

    # Display the rows with missing data if any
    if not rows_with_missing_data.empty:
        print("Rows with missing data:")
        print(rows_with_missing_data)
    else:
        print("No missing data in any row.")

    df = df.fillna('None')

    # Drop duplicates
    df = df.drop_duplicates()

    # Take a random sample of the data
    print(f"Sample of the data in {df.attrs['file_path']}:", df.sample(5))

    return df

```

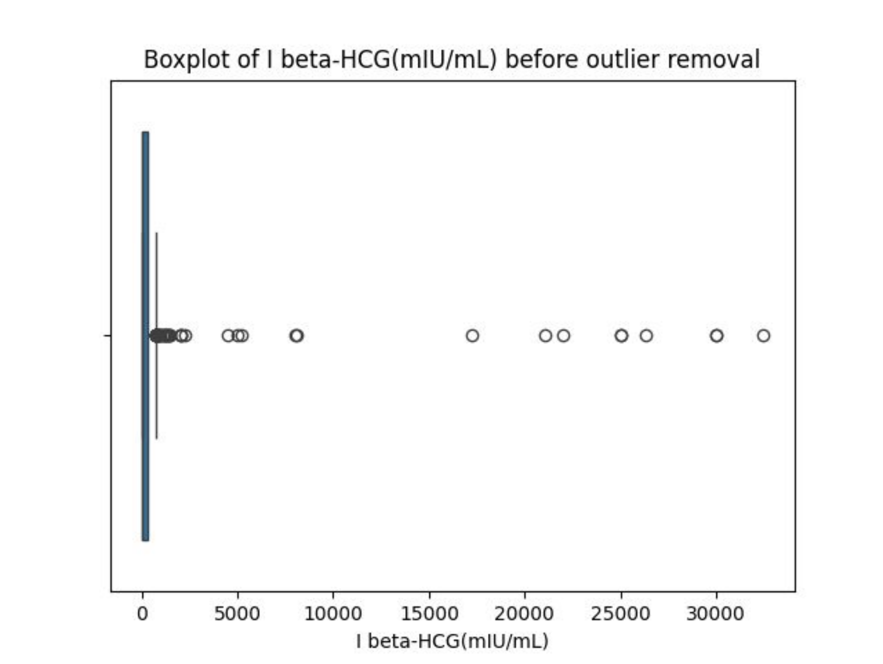
Sample of data

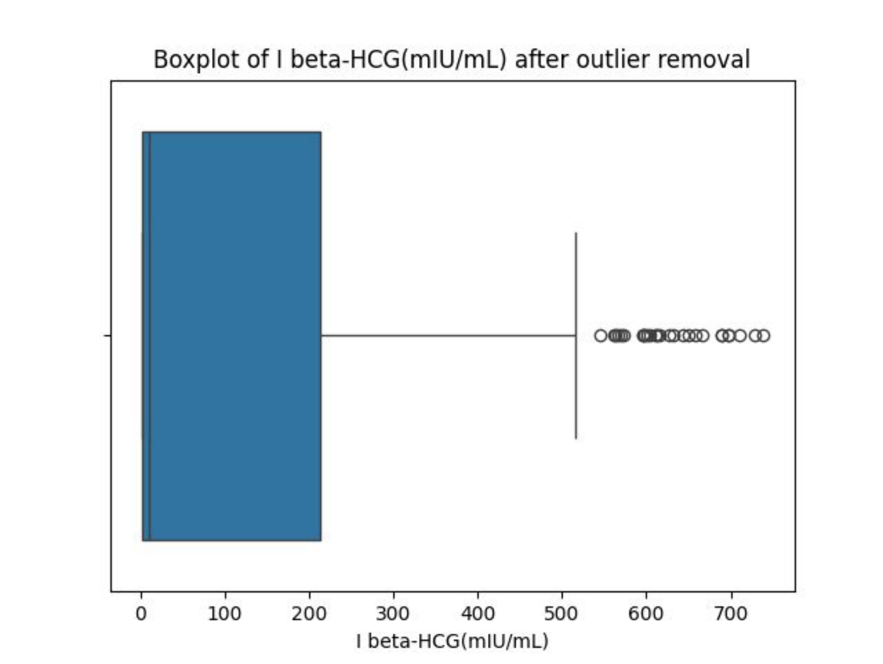


data > PCOS_woinf_processed.csv > data

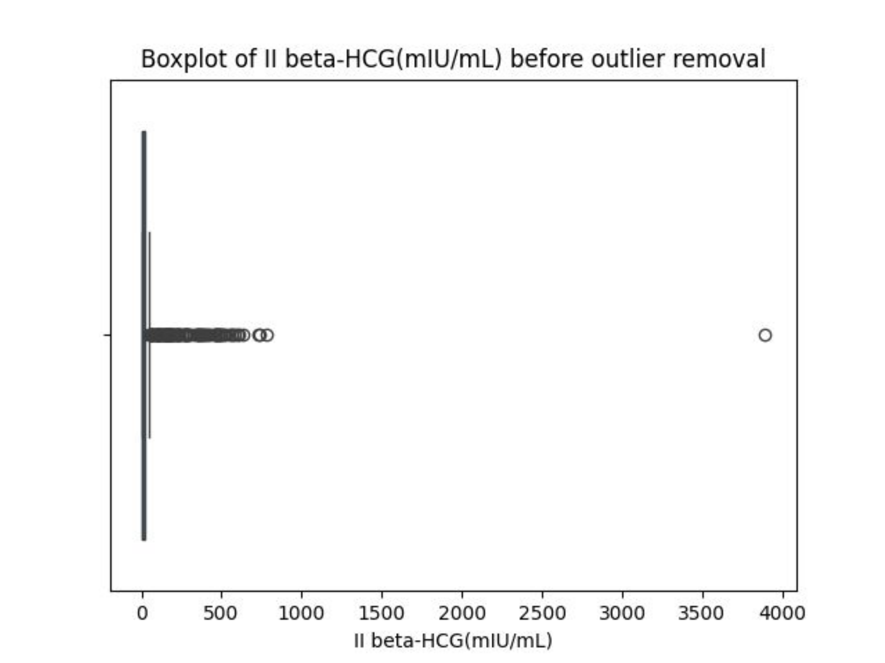
```
Sl. No,Patient File No.,PCOS (Y/N),Age (yrs),Weight (Kg),Height(Cm),BMI,Blood Group,Pulse rate(bpm),RR (breaths/min),Hb(g/dl),Cycle(R/I),Cycle length(days),Marriage Status (Yrs),Pregnant(Y/N),No. of abo
1,1,0,28,44.6,152.0,19.3,15,78,22,10.48,2,5,7.0,0,0,1.99,1.99,7.95,3.68,2.160326087,36,30,0.8333333333,0.68,2.07,45.16,17.1,0.57,92.0,0,0,0,0,0,1.0,0,110,80,3,3,18.0,18.0,8.5
2,2,0,36,65.0,161.5,24.92116286,15,74,20,11.7,2,5,11.0,1,0,60.8,1.99,6.73,1.09,6.174311927,38,32,0.8421052632,3.16,1.53,20.09,61.3,0.97,92.0,0,0,0,0,0,0.0,0,120,70,3,5,15.0,14.0,3.7
3,3,1,33,68.8,165.0,25.27089073,11,72,18,11.8,2,5,10.0,1,0,494.08,494.08,5.54,0.88,6.295454545,40,36,0.9,2.54,6.63,10.52,49.7,0.36,84.0,0,0,0,1,1,1.0,0,120,80,13,15,18.0,20.0,10.0
4,4,0,37,65.0,148.0,29.67494522,13,72,20,12.0,2,5,4.0,0,0,1.99,1.99,8.06,2.36,3.415254237,42,36,0.8571428571,16.41,1.22,36.9,33.4,0.36,76.0,0,0,0,0,0,0.0,0,120,70,2,2,15.0,14.0,7.5
5,5,0,25,52.0,161.0,20.06095444,11,72,18,10.0,2,5,1.0,1,0,801.45,801.45,3.98,0.9,4.422222222,37,30,0.8108108108,3.57,2.26,30.09,43.8,0.38,84.0,0,0,0,1,0,0.0,0,120,80,3,4,16.0,14.0,7.0
6,6,0,36,74.1,165.0,27.21763085,15,78,28,11.2,2,5,8.0,1,0,237.97,1.99,3.24,1.07,3.028037383,44,38,0.8636363636,1.6,6.74,16.18,52.4,0.3,76.0,1,0,0,1,0,0.0,0,110,70,9,6,16.0,20.0,8.0
7,7,0,34,64.0,156.0,26.29848784,11,72,18,10.9,2,5,2.0,0,0,1.99,1.99,2.85,0.31,9.193548387,39,33,0.8461538462,1.51,3.05,26.41,42.7,0.46,93.0,0,0,0,0,0,0.0,0,120,80,6,6,15.0,16.0,6.8
8,8,0,33,58.5,159.0,23.13990744,13,72,20,11.0,2,5,13.0,1,2,100.51,100.51,4.86,3.07,1.583061889,44,38,0.8636363636,12.18,1.54,3.97,38.0,0.26,91.0,1,0,0,0,0,0.0,0,120,80,7,6,15.0,18.0,7.1
9,9,0,32,40.0,158.0,16.02307323,11,72,18,11.8,2,5,8.0,0,1,1.99,1.99,3.76,3.02,1.245033113,39,35,0.8974358974,1.51,1,19.0,21.8,0.3,116.0,0,0,0,0,0,0.0,0,120,80,5,7,17.0,17.0,4.2
10,10,0,36,52.0,150.0,23.11111111,15,80,20,10.0,4,2,4.0,0,0,1.99,1.99,2.8,1.51,1.854304636,40,38,0.95,6.65,1.61,11.74,27.7,0.25,125.0,0,0,0,0,0,0.0,0,110,80,1,1,14.0,17.0,2.5
11,11,0,20,71.0,163.0,26.72287252,15,80,20,10.0,2,5,4.0,1,2,158.51,158.51,4.89,2.02,2.420792079,39,35,0.8974358974,1.56,4.47,13.47,18.1,0.36,108.0,0,0,0,0,0,0.0,0,110,80,7,15,17.0,20.0,6.0
12,12,0,26,49.0,160.0,19.140625,13,72,20,9.5,2,5,3.0,0,1,1.99,1.99,4.09,1.47,2.782312925,39,33,0.8461538462,3.98,1.67,21.1,29.18,0.25,100.0,0,0,0,0,0,0.0,0,120,80,4,2,18.0,19.0,7.8
13,13,1,25,74.0,152.0,32.02908587,17,72,18,11.7,4,2,7.0,1,0,1214.23,1214.23,2.0,1.51,1.324503311,45,40,0.8888888889,6.51,7.94,22.43,31.4,0.3,125.0,1,1,1,1,1,1.0,1,120,80,15,8,20.0,21.0,8.0
14,14,0,38,50.0,152.0,21.64127424,13,74,20,12.1,2,5,15.0,0,0,1.99,1.99,4.84,0.71,6.816901408,39,33,0.8461538462,1.48,2.38,15.62,21.2,0.4,91.0,0,0,0,0,0,0.0,0,110,70,3,3,18.0,17.0,5.6
15,15,0,34,57.3,162.0,21.83356196,13,74,22,11.7,2,5,9.0,0,0,1.99,1.99,7.45,3.71,2.008086253,38,30,0.7894736842,1.51,0.88,19.6,24.9,0.26,116.0,0,0,0,1,1,1.0,0,120,80,4,1,19.0,21.0,5.5
16,16,0,38,80.5,154.0,33.9433294,13,78,22,11.4,2,5,20.0,0,0,1.99,1.99,9.51,2.51,3.788844622,44,41,0.9318181818,1.18,0.69,92.65,9.7,0.3,116.0,1,0,0,0,0,0.0,0,120,80,1,3,14.0,20.0,3.9
17,17,0,29,43.0,148.0,19.6311176,13,80,20,11.1,2,5,2.0,1,0,8104.21,91.55,2.02,0.65,3.107692308,36,29,0.8055555556,1.98,3.78,20.25,19.1,0.46,91.0,0,0,0,0,0,0.0,0,110,70,6,5,20.0,20.0,5.6
18,18,0,36,69.2,160.0,27.03125,13,72,18,10.8,2,5,7.0,0,0,1.99,1.99,4.86,2.96,1.641891892,39,32,0.8205128205,5.0,1.92,12.52,26.0,0.25,100.0,0,0,0,0,0,0.0,0,110,80,1,2,20.0,18.0,4.0
19,19,0,31,52.4,159.0,20.7270282,17,72,18,12.7,2,5,7.0,0,0,1.99,1.99,6.05,1.05,5.761904762,37,33,0.8918918919,3.19,1,12.05,23.3,0.25,127.0,0,0,0,0,0,0.0,0,120,80,0,2,0.0,17.0,5.6
20,20,1,30,85.0,165.0,31.22130395,16,72,18,12.5,4,7,7.0,0,0,23.58,1.99,1.89,0.81,2.333333333,44,42,0.9545454545,2.87,2.07,19.13,28.0,0.3,100.0,0,1,1,1,1,1.0,0,120,80,16,8,18.0,17.0,11.0
21,21,0,25,64.0,156.0,26.29848784,11,70,18,11.2,2,6,6.0,0,0,1.99,1.99,2.82,1.3,2.169230769,39,34,0.8717948718,1.86,2.85,33.62,17.1,0.4,91.0,0,0,0,0,0,0.0,0,110,80,1,2,18.0,19.0,5.7
22,22,0,38,50.0,156.0,20.54569362,15,72,18,11.0,4,9,12.0,1,0,749.98,749.98,3.18,2.18,1.458715596,36,29,0.8055555556,5.71,2.13,40.74,18.9,0.3,84.0,0,0,0,0,0,0.0,0,120,80,4,2,17.0,17.0,7.6
23,23,0,34,64.2,155.0,26.72216441,15,74,20,12.1,2,5,10.0,1,0,218.65,218.65,4.08,2.3,1.773913043,36,32,0.8888888889,1.25,4.13,13.38,28.6,0.26,116.0,0,0,0,0,0,0.0,0,110,70,5,7,16.0,17.0,7.6
24,24,0,28,65.0,152.0,28.13365651,13,74,22,10.5,2,5,5.0,1,0,13.0,13,6.41,1.69,3.792899408,37,33,0.8918918919,0.45,2.5,17.88,30.54,0.73,92.0,0,0,0,0,1,0.0,0,110,80,6,8,14.0,18.0,8.5
25,25,1,34,63.0,158.0,25.23634033,11,72,20,11.2,2,5,12.0,1,0,610.63,610.63,5.34,0.89,6.0,38,32,0.8421052632,0.65,1.89,11.46,21.5,0.98,100.0,0,1,0,0,1,1.0,0,120,70,4,6,18.0,17.0,7.3
26,26,0,41,42.0,152.0,18.17867036,12,80,20,11.5,2,5,15.0,1,0,4490.18,4490.18,8.82,4.39,2.009111617,39,32,0.8205128205,0.84,0.26,17.35,45.6,0.98,92.0,0,0,0,0,0,0.0,0,110,70,1,2,15.0,17.0,8.8
27,27,1,30,76.0,160.0,29.6875,15,75,18,11.2,4,3,5.0,0,0,1.99,1.99,6.18,2.78,2.223021583,45,38,0.8444444444,4.28,3.84,17.98,25.14,0.35,100.0,1,1,1,1,1,1.0,1,120,80,21,20,11.0,12.0,6.8
28,28,0,20,68.0,152.0,29.43213296,17,72,20,10.0,4,3,4.0,1,0,689.58,11.24,1.8,0.41,4.390243902,40,33,0.825,8.39,2.5,11.24,46.3,1.2,92.0,0,0,0,0,0,0.0,0,110,80,3,2,10.0,11.0,10.0
29,29,0,25,62.0,158.0,24.8357635,15,78,22,11.6,2,5,8.0,0,0,15.0,15,4.6,1.06,4.339622642,42,34,0.8095238095,3.79,3.56,15.3,36.56,0.25,106.0,0,0,0,0,0,0.0,0,120,80,7,4,12.0,17.0,6.2
30,30,1,28,56.0,152.0,24.23822715,13,78,22,11.0,4,3,7.0,0,0,768.03,768.03,2.41,1.19,2.025210084,40,33,0.825,3.44,1.56,45.47,23.14,0.63,100.0,1,0,1,0,1,1.0,0,110,80,11,10,13.0,14.0,5.0
31,31,0,32,50.0,155.0,20.81165453,17,78,22,10.8,2,4,3.5,1,0,612.32,12,6.34,4.07,1.557739558,34,28,0.8235294118,1.4,1.69,54.09,56.2,0.31,80.0,0,0,0,0,0,0.0,0,110,80,10,8,14.0,14.0,7.0
32,32,1,34,57.0,150.0,25.33333333,11,72,22,12.0,2,4,15.0,0,0,10.0,10,5.01,6.48,0.7731481481,40,37,0.925,4.21,1.89,12.71,32.45,0.27,85.0,1,1,1,1,1,1.0,0,110,80,5,10,14.0,13.0,6.0
33,33,0,31,58.0,155.0,24.14151925,15,72,18,10.5,2,5,2.0,0,0,1.99,1.99,2.8,0.42,6.666666667,39,32,0.8205128205,1.88,2.34,23.74,32.2,0.52,138.0,0,0,0,0,0,0.0,0,120,80,8,9,12.0,13.0,7.5
34,34,0,38,54.0,165.0,19.83471074,15,72,18,10.2,2,5,15.0,0,3,20.0,20,6.11,1.97,3.101522843,42,34,0.8095238095,2.42,1.58,40.41,36.85,0.62,92.0,0,0,0,0,0,0.0,0,120,80,2,5,14.0,12.0,8.5
35,35,0,28,50.0,158.0,20.02884153,15,76,20,10.0,2,5,6.0,1,0,1277.49,30.66,2.17,0.57,3.807017544,32,27,0.84375,1.66,2.36,55.2,36.87,0.6,85.0,0,0,0,0,0,0.0,0,120,70,11,7,14.0,19.0,8.0
36,36,0,32,71.0,165.0,26.07897153,11,72,22,12.0,2,5,10.0,1,0,1455.0,1455,2.65,0.52,5.096153846,42,36,0.8571428571,2.03,3.64,19.78,25.22,0.45,80.0,0,0,0,0,0,0.0,0,110,80,1,1,11.0,12.0,7.0
37,37,0,37,73.0,158.0,29.24210864,15,74,18,10.5,4,7,12.0,0,0,18.0,1.99,3.22,0.48,6.708333333,45,38,0.8444444444,3.57,2.78,28.88,36.58,0.27,92.0,0,0,0,0,0,0.0,0,110,80,6,2,11.0,13.0,10.0
38,38,0,26,72.0,157.0,29.21010994,11,70,18,11.0,2,5,1.0,0,0,12.0,1.99,6.04,2.32,2.603448276,42,34,0.8095238095,65.0,0.88,30.59,20.25,0.88,90.0,0,0,0,0,0,0.0,0,120,80,1,2,6.5,8.5,10.0
39,39,0,36,53.0,158.0,21.23057202,13,78,22,11.0,2,5,17.0,0,0,10.0,1.99,3.99,1.99,2.005025126,39,32,0.8205128205,2.376,2.36,19.64,45.55,0.65,100.0,0,0,0,0,0,0.0,0,110,80,7,5,8.0,7.0,15.0
40,40,0,20,74.0,171.0,25.30693205,13,74,16,11.1,4,0,1.0,0,0,1.99,1.99,3.52,9.17,0.3838604144,42,36,0.8571428571,2.01,0.33,12.49,38.63,0.32,80.0,0,0,0,0,0,0.0,0,110,70,6,12,13.0,12.0,0.0
41,41,0,32,40.0,156.0,16.4365549,15,78,22,10.5,2,5,6.0,0,0,10.0,1.99,7.87,4.89,1.609406953,34,30,0.8823529412,5.0,2.35,31.32,56.2,0.25,107.0,0,0,0,0,0,0.0,0,110,80,5,8,14.0,14.0,7.3
42,42,0,29,78.0,165.0,28.65013774,13,72,16,12.8,2,5,2.0,1,0,497.41,497.41,5.6,2.13,2.629107981,42,36,0.8571428571,0.93,3.88,20.46,14.36,85.0,92.0,0,0,0,0,0,0.0,0,110,80,8,10,11.0,12.0,6.5
43,43,0,28,33.0,157.0,13.38796706,13,78,18,11.5,4,9,3.5,0,0,10.0,1.99,3.34,0.25,13.36,32,26,0.8125,3.4,3.55,6.59,42.3,0.25,84.0,0,0,0,0,0,0.0,0,100,70,0,1,0.0,12.0,6.5
44,44,0,24,68.0,165.0,24.97704316,15,72,16,11.0,2,5,3.5,1,0,161.147,167,2.79,1.97,1.416243655,39,33,0.8461538462,1.48,4.33,0.4,25.366,0.25,100.0,0,0,0,0,0,0.0,0,110,70,6,8,8.0,8.0,7.12
45,45,1,29,59.0,158.0,23.63403301,17,78,22,12.0,4,7,7.0,1,0,1305.35,1.99,3.26,0.79,4.126582278,39,33,0.8461538462,3.77,3.55,31.36,44.55,0.76,100.0,1,1,1,1,1,1.0,1,120,80,13,14,16.0,15.0,7.0
46,46,0,25,56.0,155.0,23.30905307,15,70,18,10.6,2,5,6.0,1,0,43.5,9.83,2.77,0.57,4.859649123,40,33,0.825,3.85,2.36,25.6,45.5,0.31,92.0,0,0,0,0,0,0.0,0,120,80,5,8,14.0,13.0,9.4
47,47,1,28,75.0,165.0,27.54820937,15,72,18,12.0,2,5,2.5,1,0,141.06,141.06,5.81,0.27,21.51851852,45,41,0.9111111111,0.83,3.66,14.43,52.3,0.38,100.0,1,1,0,1,1,1.0,0,110,70,4,6,10.0,12.0,9.0
48,48,0,26,40.0,156.0,16.4365549,13,74,18,10.2,2,5,3.0,1,0,1500.0,528.5,7.34,2.78,2.64028777,39,33,0.8461538462,4.45,4.33,24.58,54.62,0.25,80.0,0,0,0,0,0,0.0,0,110,70,5,7,15.0,18.0,16.0
49,49,0,34,51.0,155.0,21.22788762,11,78,24,10.8,2,5,11.0,0,0,10.0,1.99,2.46,0.51,4.823529412,39,34,0.8717948718,2.16,1,7.97,35.66,0.25,75.0,0,0,0,0,0,1.0,0,110,80,6,8,10.0,15.0,9.0
50,50,1,27,51.0,154.0,21.50446956,15,72,18,12.8,4,3,4.0,0,0,10.0,1.99,2.26,2.4,0.9416666667,38,34,0.8947368421,2.77,4.5,15.22,45.2,0.52,100.0,1,1,1,1,1,1.0,1,110,70,5,8,14.0,13.0,10.4
51,51,1,23,68.0,172.0,22.98539751,15,72,16,10.8,4,10,4.0,0,0,77.6,1.99,4.42,5.48,0.8065693431,42,34,0.8095238095,3.26,3.2,27.65,23.114,0.39,70.0,1,1,1,1,1,0.0,0,110,70,22,18,12.0,13.0,11.0
52,52,0,23,54.0,158.0,21.63114885,16,72,18,10.8,2,5,2.0,1,0,177.57,177.57,5.32,1.8,2.955555556,38,32,0.8421052632,5.17,2.1,24.5,14.0,0.25,70.0,0,0,0,0,0,0.0,0,100,70,7,10,15.0,13.0,10.0
```

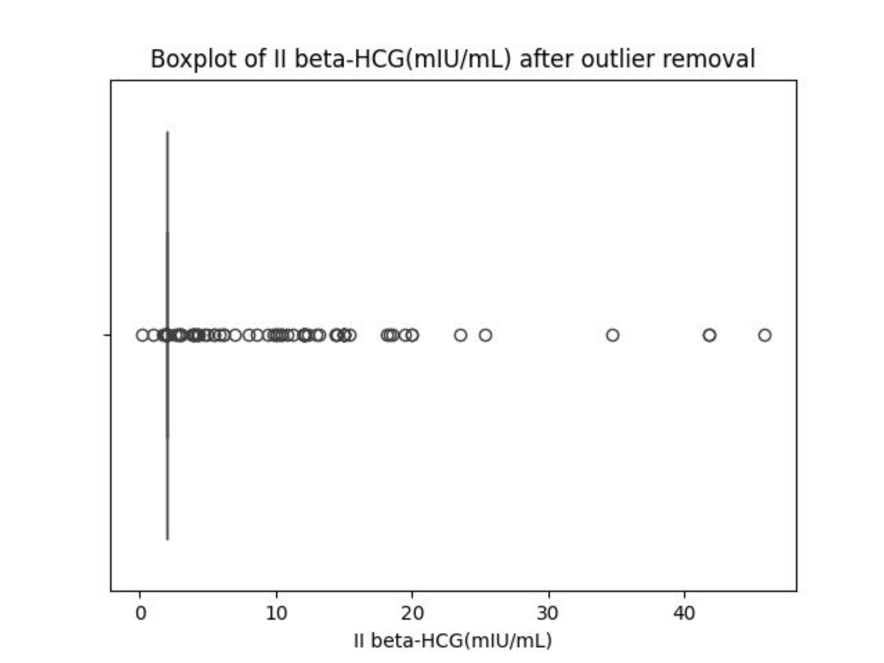
Boxplot of I beta-HCG (mIU/mL)



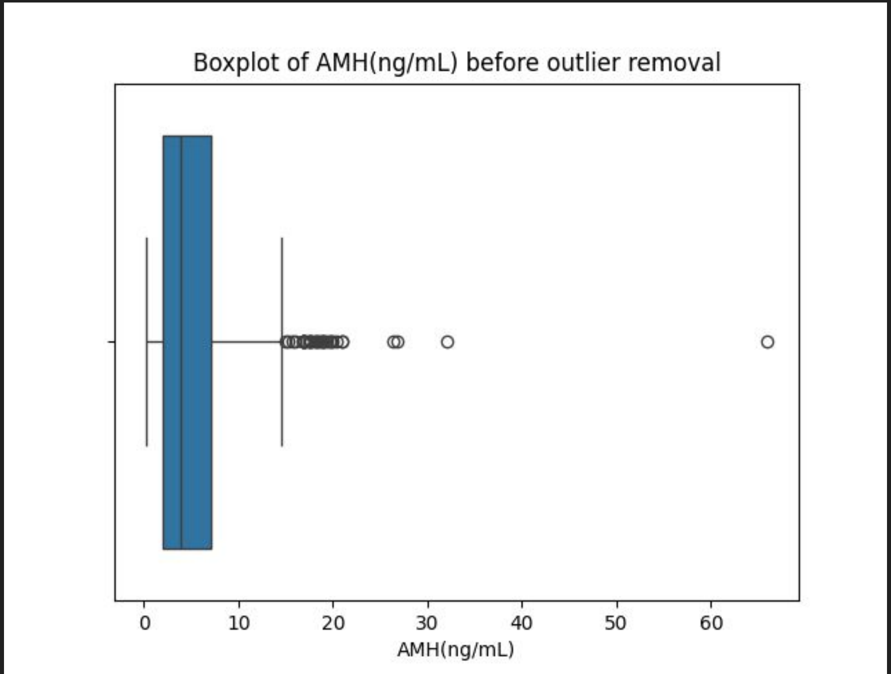


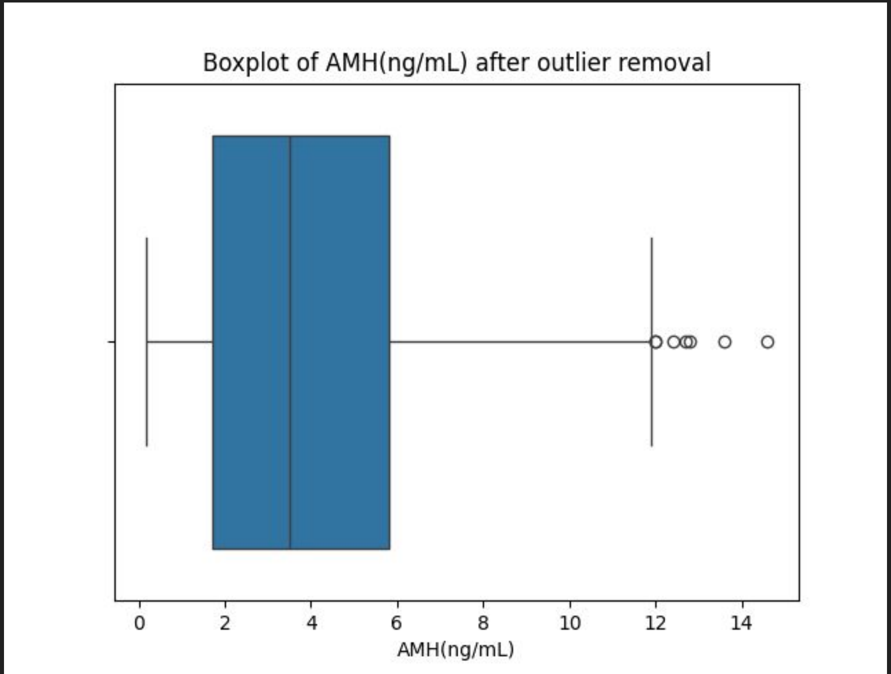
Boxplot II beta-HCG (mIU/mL)

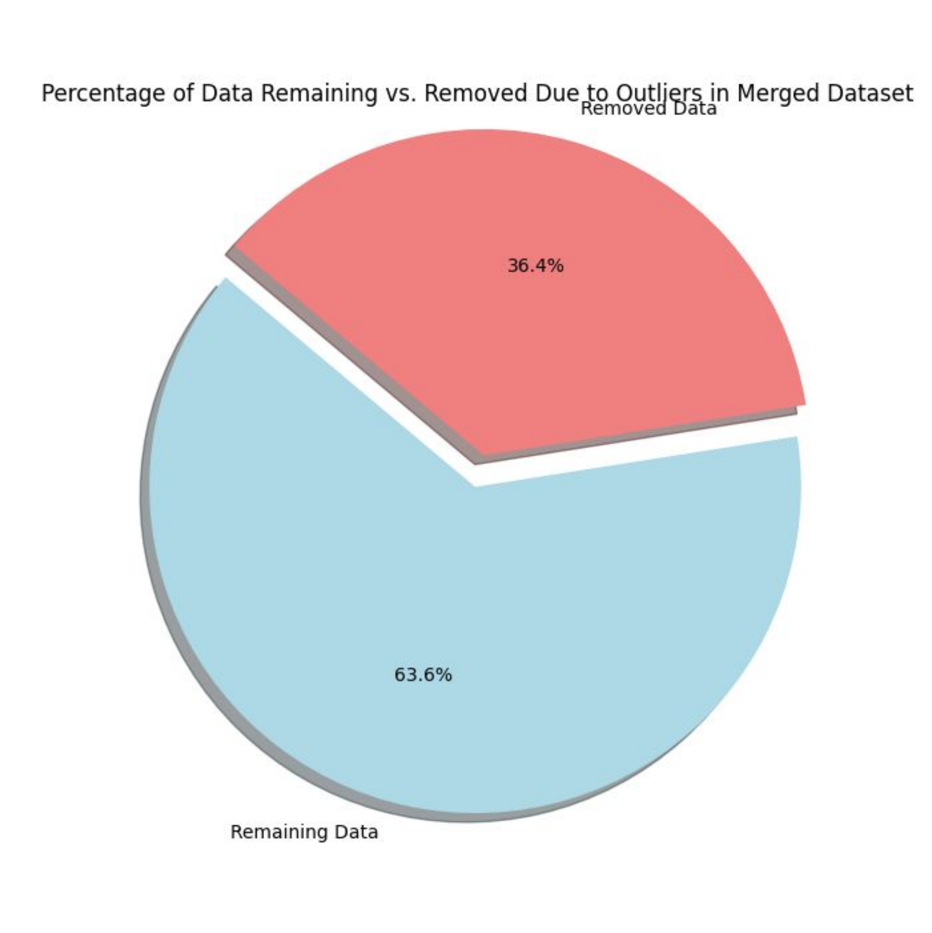


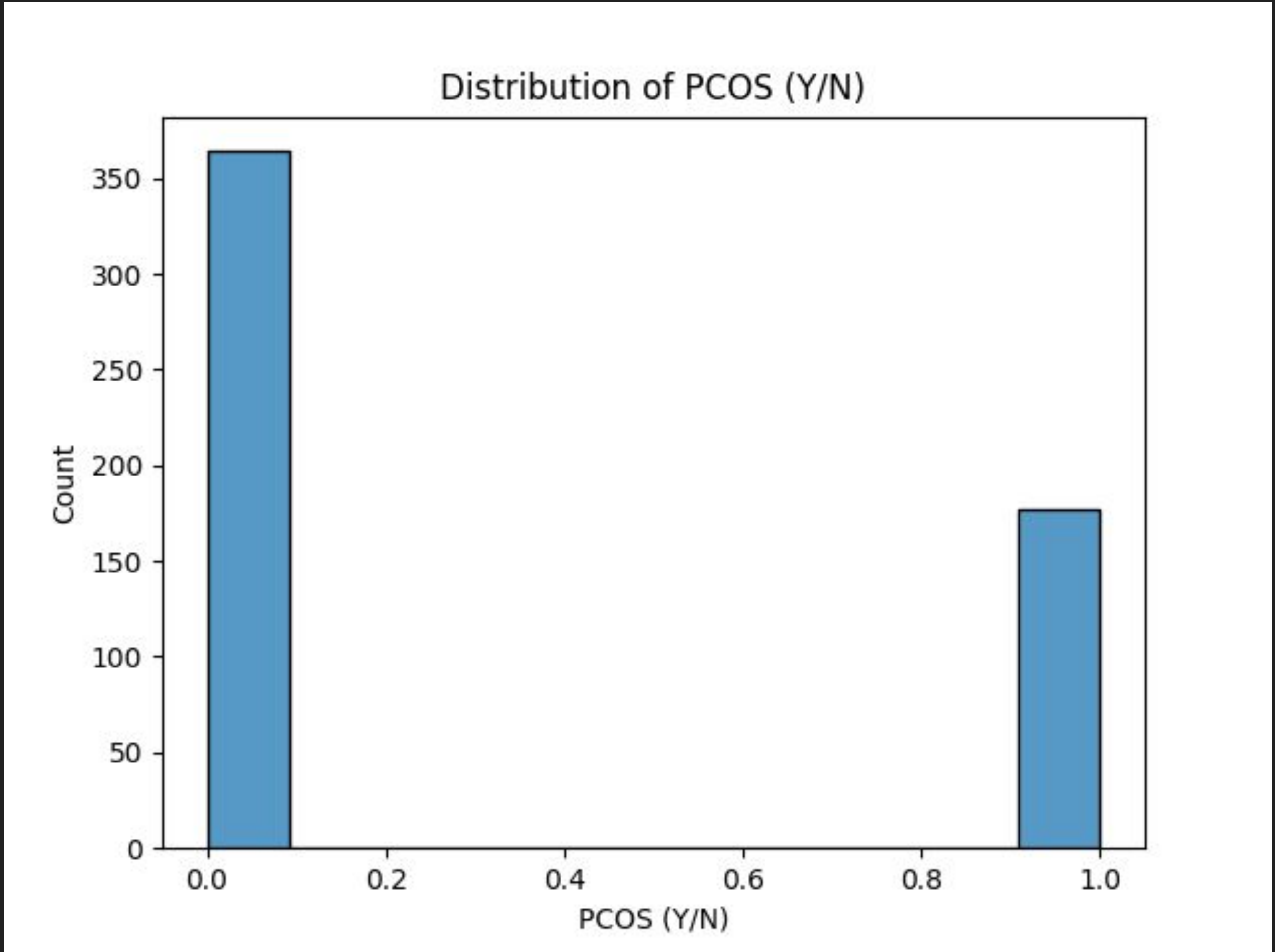


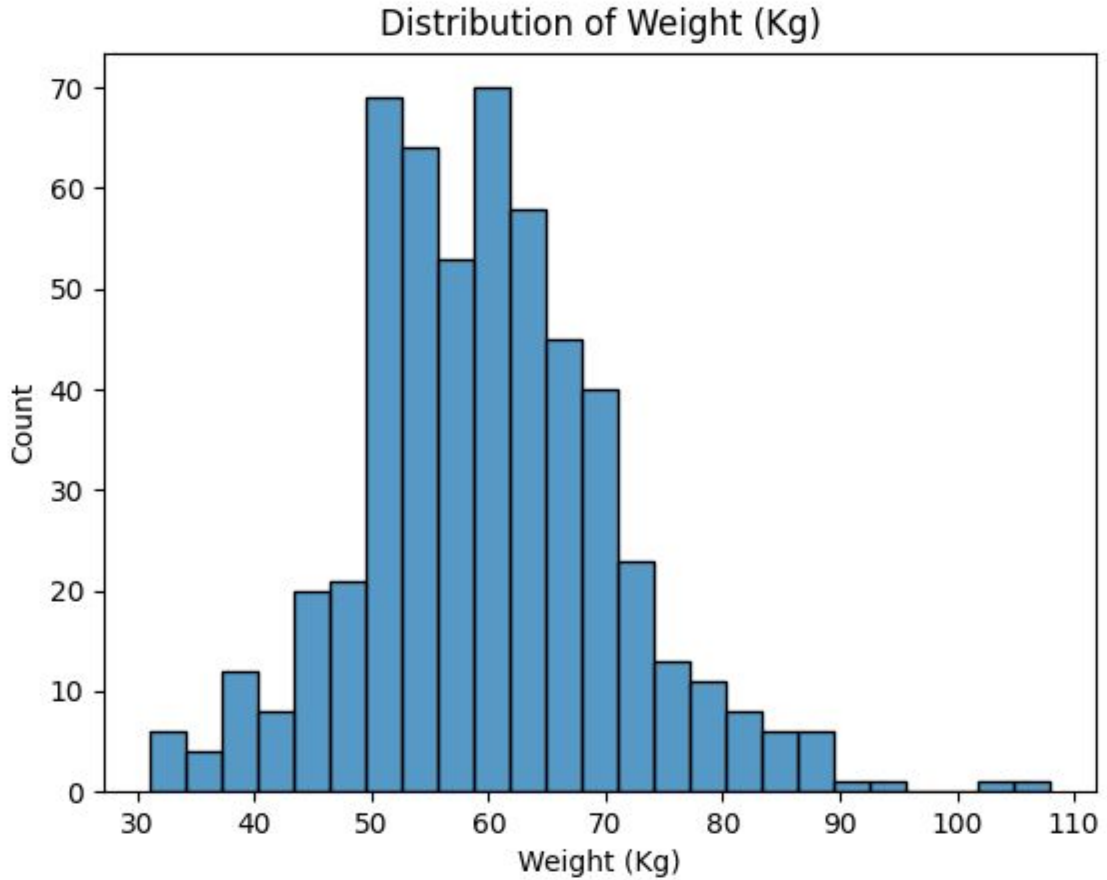
Boxplot of AMH (ng/mL)

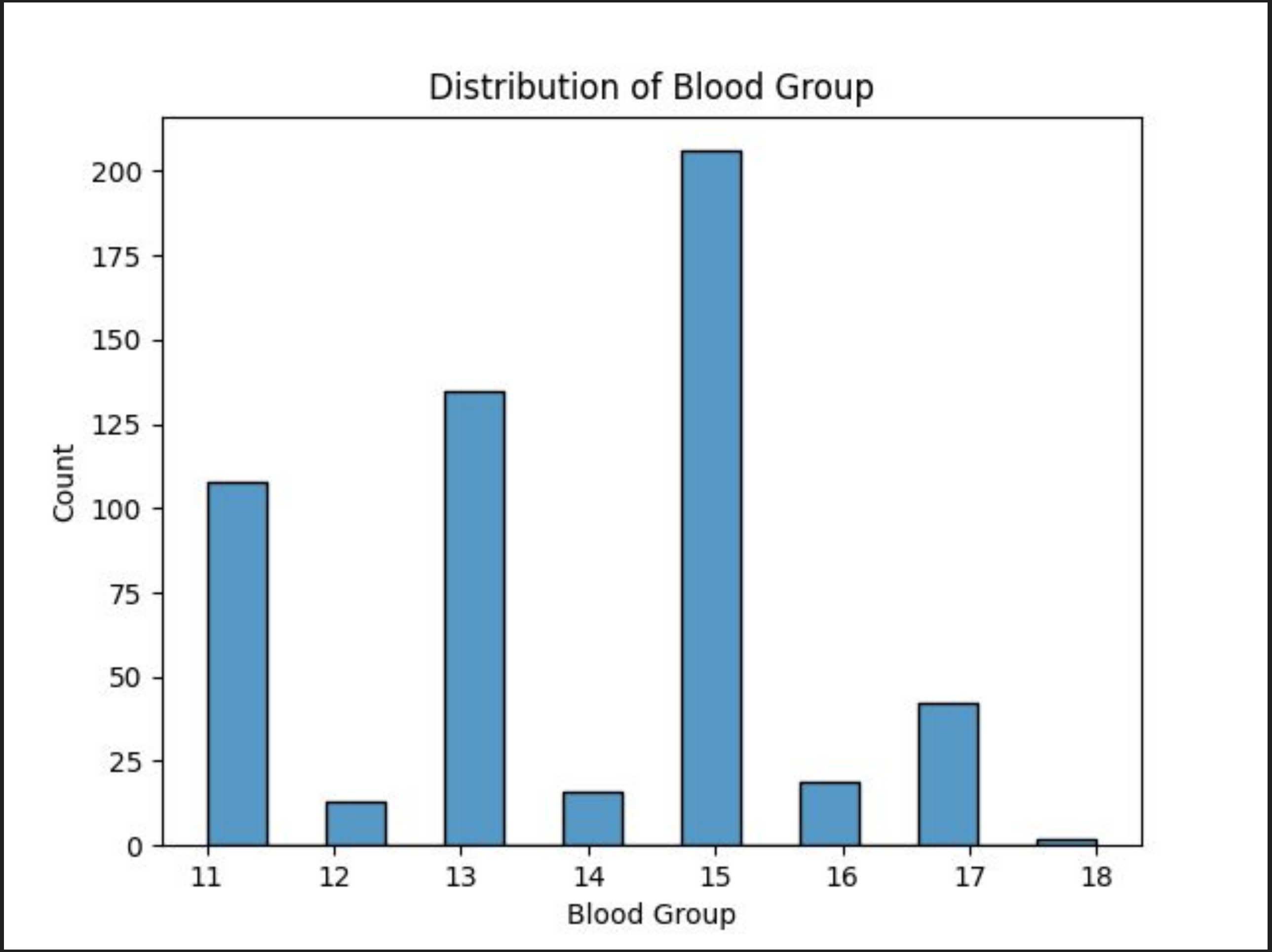


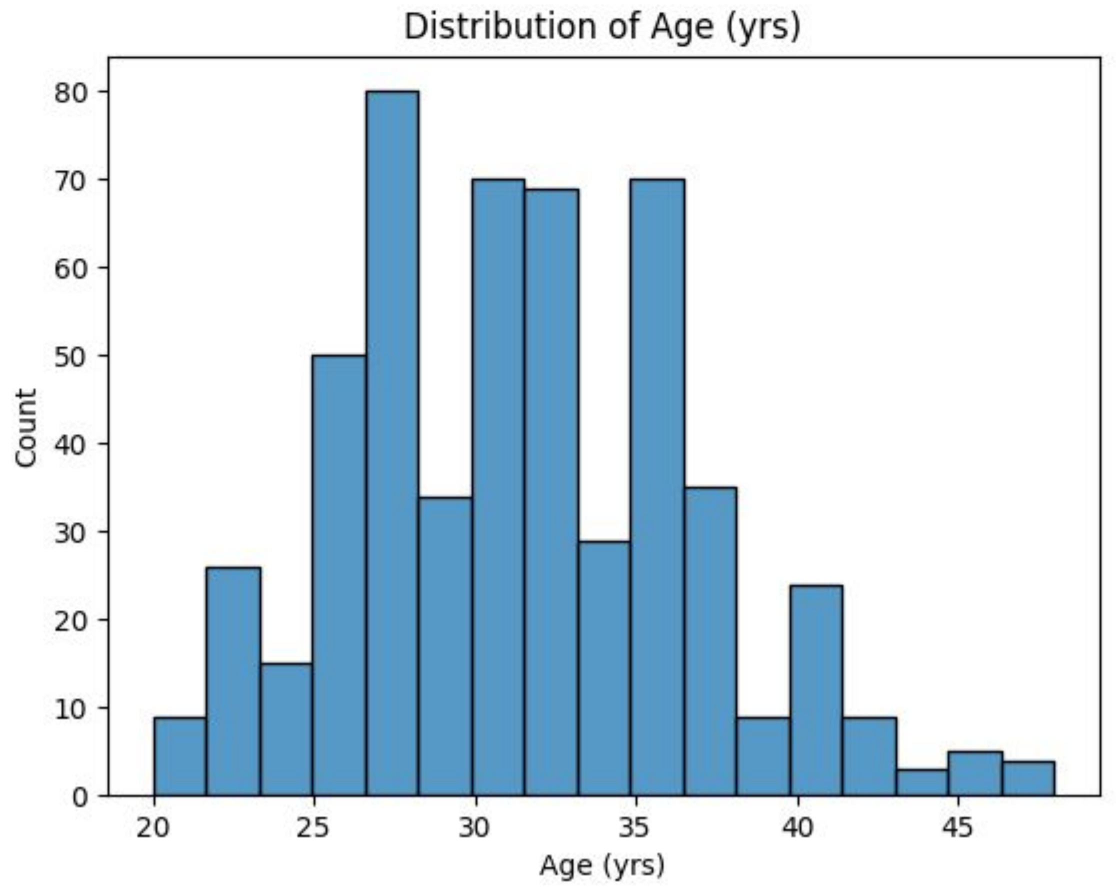


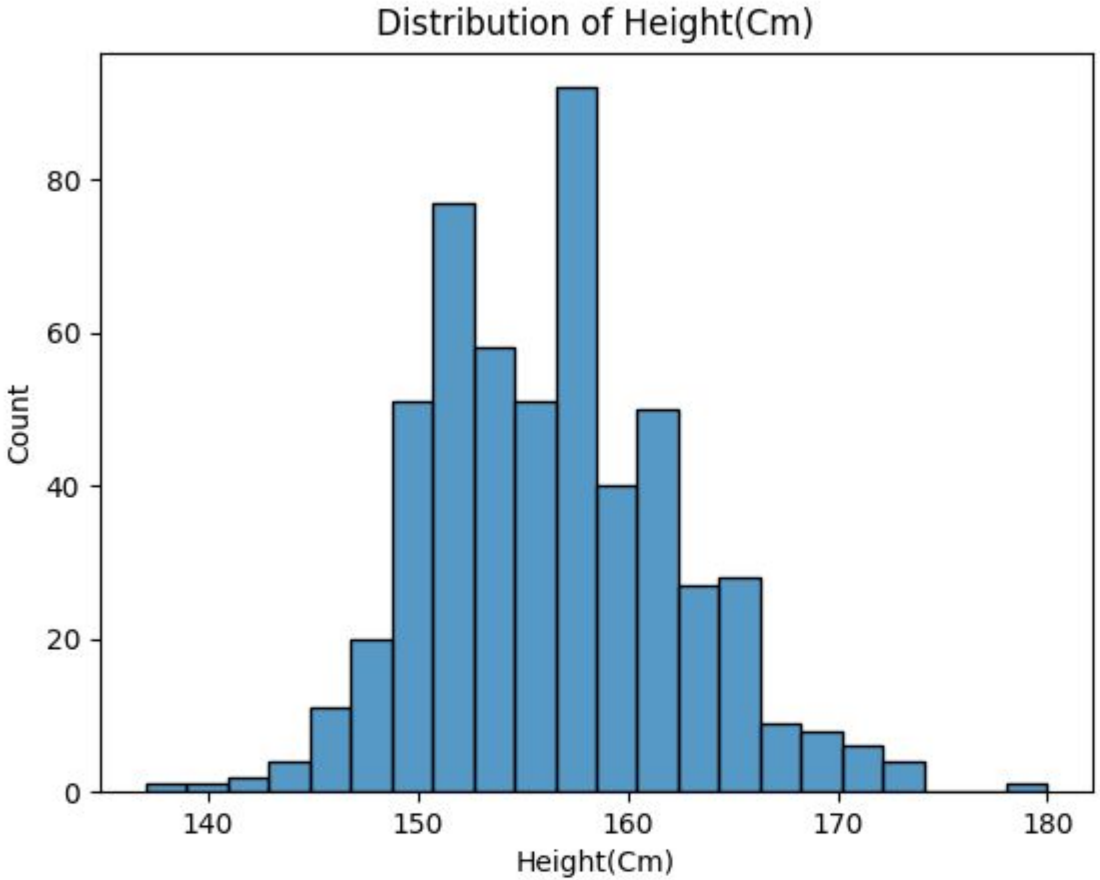

Percentage of Data Remaining vs. Removed Due to Outliers in Merged Dataset
Removed Data
36.4%
63.6%
Remaining Data

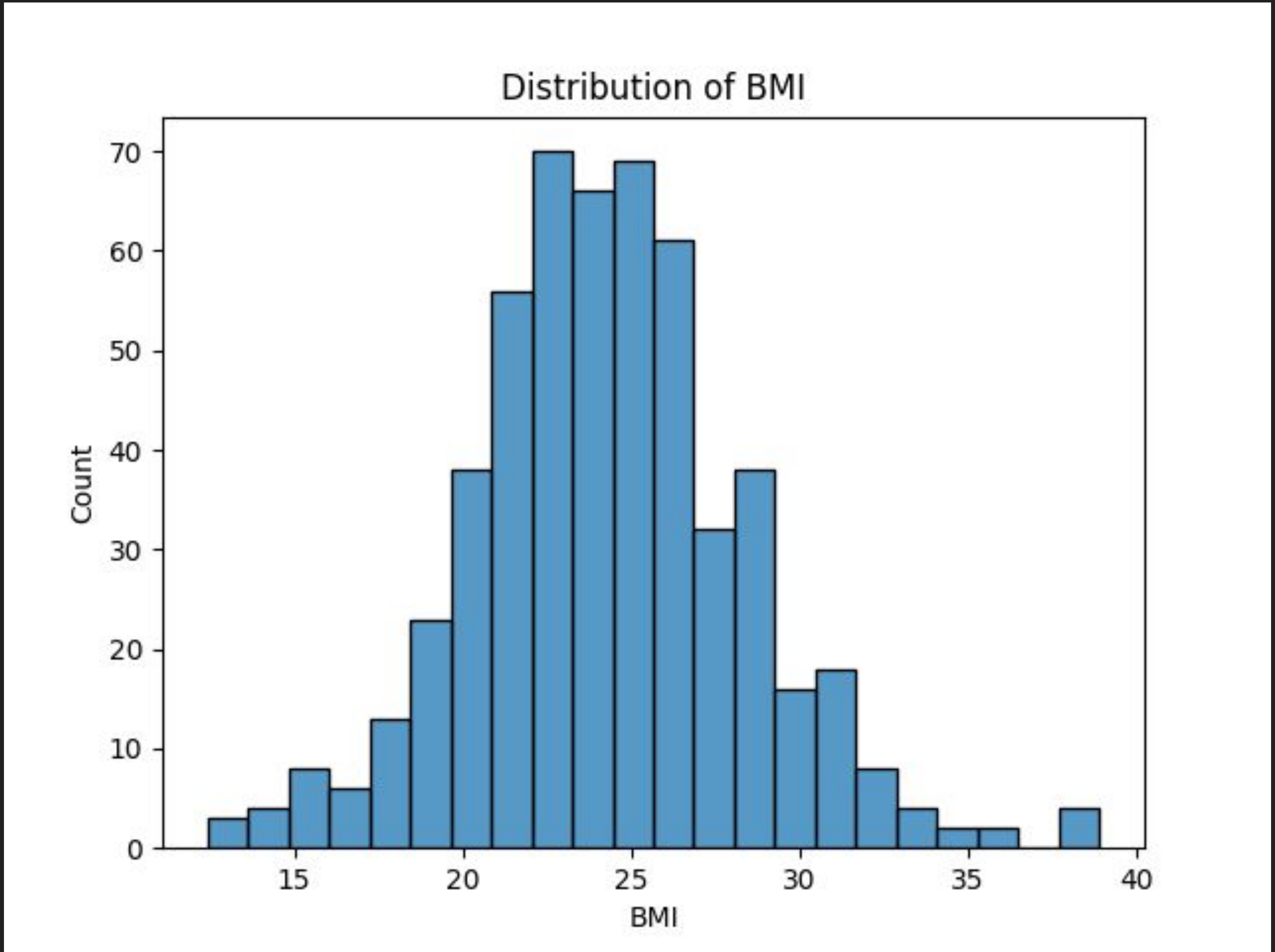

Distribution of PCOS (Y/N)
Count
0
50
100
150
200
250
300
350
0.0
0.2
0.4
0.6
0.8
1.0
PCOS (Y/N)

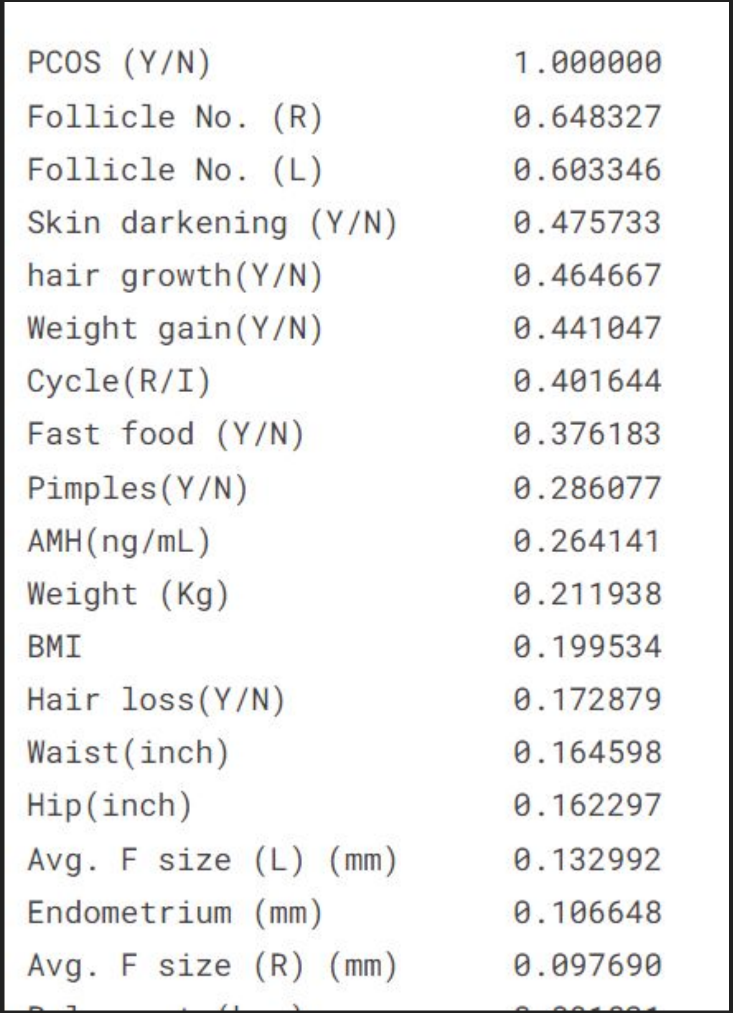

Distribution of Weight (Kg)
Count
0
10
20
30
40
50
60
70
30
40
50
60
70
80
90
100
110
Weight (Kg)


Distribution of Blood Group
Count
0
25
50
75
100
125
150
175
200
11
12
13
14
15
16
17
18
Blood Group

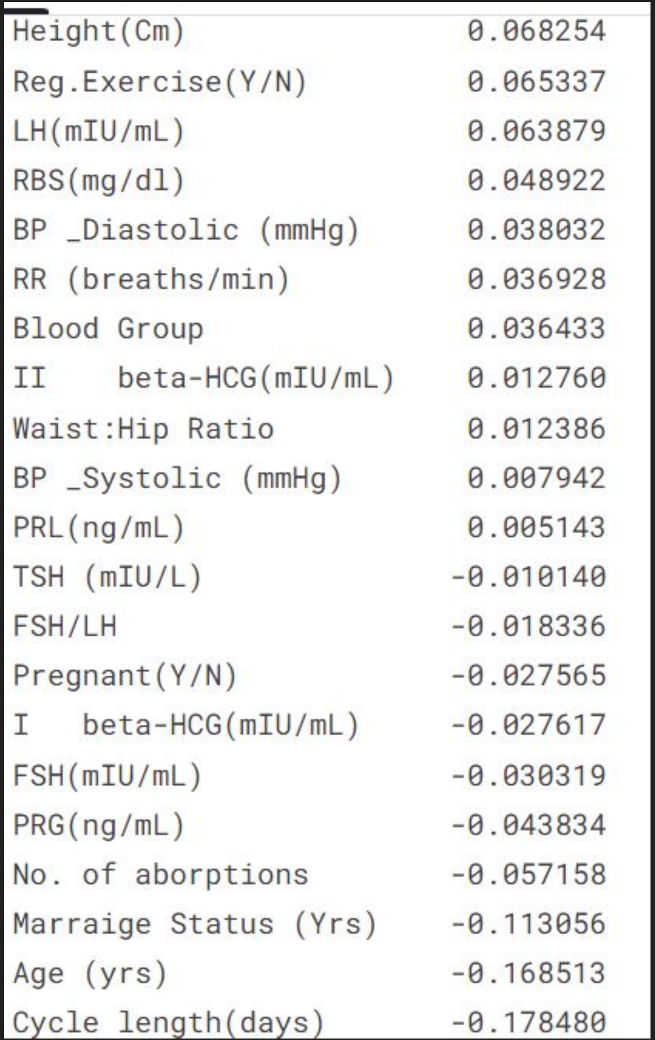

Distribution of Age (yrs)
Count
0
10
20
30
40
50
60
70
80
20
25
30
35
40
45
Age (yrs)


Distribution of Height(Cm)
Count
0
20
40
60
80
140
150
160
170
180
Height(Cm)


Distribution of BMI
Count
0
10
20
30
40
50
60
70
15
20
25
30
35
40
BMI



```
PCOS (Y/N)                 1.000000
Follicle No. (R)           0.648327
Follicle No. (L)           0.603346
Skin darkening (Y/N)       0.475733
hair growth(Y/N)           0.464667
Weight gain(Y/N)           0.441047
Cycle(R/I)                 0.401644
Fast food (Y/N)            0.376183
Pimples(Y/N)               0.286077
AMH(ng/mL)                 0.264141
Weight (Kg)                0.211938
BMI                        0.199534
Hair loss(Y/N)             0.172879
Waist(inch)                0.164598
Hip(inch)                  0.162297
Avg. F size (L) (mm)       0.132992
Endometrium (mm)           0.106648
Avg. F size (R) (mm)       0.097690
```



```
Height(Cm)                 0.068254
Reg.Exercise(Y/N)          0.065337
LH(mIU/mL)                 0.063879
RBS(mg/dl)                 0.048922
BP _Diastolic (mmHg)       0.038032
RR (breaths/min)           0.036928
Blood Group                0.036433
II    beta-HCG(mIU/mL)     0.012760
Waist:Hip Ratio            0.012386
BP _Systolic (mmHg)        0.007942
PRL(ng/mL)                 0.005143
TSH (mIU/L)               -0.010140
FSH/LH                    -0.018336
Pregnant(Y/N)             -0.027565
I   beta-HCG(mIU/mL)      -0.027617
FSH(mIU/mL)               -0.030319
PRG(ng/mL)                -0.043834
No. of aborptions         -0.057158
Marraige Status (Yrs)     -0.113056
Age (yrs)                 -0.168513
Cycle length(days)        -0.178480
```

Assumptions

We first want to focus on the first three feature that is present on both data set which were 'I beta-HCG(mIU/mL)', 'II beta-HCG(mIU/mL)', 'AMH(ng/mL)' as features to train our dataset.

And our target variable will be PCOS positive/negative rate

Future Goals

- Develop a predictive model for the early detection of PCOS and infertility-related issues.
- Analyze the dataset to identify patterns and trends.
- Develop a model that can predict the likelihood of PCOS and infertility-related issues.
- Evaluate the model's performance using various metrics.
- Present the results in a clear and concise manner.

Feedback 4/2 12pm

- Take good traceability of every decision point where the experiment changed / was done / for what reason – Dani
- Oh that’s all

- Need more stats stuff