Mid-Program Presentation Present

Presenter's Name: Eleni Yanyan and Alex

04/02/2024

Analysis and Detection of PCOS

Objective: Develop a predictive model for early detection of PCOS and infertility.

Importance: Early diagnosis crucial for effective management.

Dataset: Comprehensive physical and clinical parameters related to PCOS.

Problem Statement

The goal is to develop a predictive model for the early detection of PCOS and infertility-related issues. PCOS is a complex hormonal disorder, and early diagnosis is crucial for effective management.

Tasks

Structure and Features: Understand dataset dimensions and variables.

Missing Values: Identify and handle missing data.

Outliers: Detect and address outliers.

Patterns: Explore patterns in the data.

Data Preprocessing

Encode categorical variables, normalize data.

Data

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 541 entries, 0 to 540
Data columns (total 44 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Sl. No                    541 non-null    int64
 1   Patient File No.          541 non-null    int64
 2   PCOS (Y/N)                541 non-null    int64
 3    Age (yrs)                541 non-null    int64
 4   Weight (Kg)               541 non-null    float64
 5   Height(Cm)                541 non-null    float64
 6   BMI                       241 non-null    float64
 7   Blood Group               541 non-null    int64
 8   Pulse rate(bpm)           541 non-null    int64
 9   RR (breaths/min)          541 non-null    int64
 10  Hb(g/dl)                  541 non-null    float64
 11  Cycle(R/I)                541 non-null    int64
 12  Cycle length(days)        541 non-null    int64
 13  Marraige Status (Yrs)     540 non-null    float64
 14  Pregnant(Y/N)             541 non-null    int64
 15  No. of aborptions         541 non-null    int64
 16    I   beta-HCG(mIU/mL)    541 non-null    float64
 17  II    beta-HCG(mIU/mL)    541 non-null    object
 18  FSH(mIU/mL)               541 non-null    float64
 19  LH(mIU/mL)                541 non-null    float64
 20  FSH/LH                    9 non-null      float64
 21  Hip(inch)                 541 non-null    int64
 22  Waist(inch)               541 non-null    int64
 23  Waist:Hip Ratio           9 non-null      float64
 24  TSH (mIU/L)               541 non-null    float64
 25  AMH(ng/mL)                541 non-null    object
 26  PRL(ng/mL)                541 non-null    float64
 27  Vit D3 (ng/mL)            541 non-null    float64
 28  PRG(ng/mL)                541 non-null    float64
 29  RBS(mg/dl)                541 non-null    float64
 30  Weight gain(Y/N)          541 non-null    int64
 31  hair growth(Y/N)          541 non-null    int64
 32  Skin darkening (Y/N)      541 non-null    int64
 33  Hair loss(Y/N)            541 non-null    int64
 34  Pimples(Y/N)              541 non-null    int64
 35  Fast food (Y/N)           540 non-null    float64
 36  Reg.Exercise(Y/N)         541 non-null    int64
 37  BP _Systolic (mmHg)       541 non-null    int64
 38  BP _Diastolic (mmHg)      541 non-null    int64
 39  Follicle No. (L)          541 non-null    int64
 40  Follicle No. (R)          541 non-null    int64
 41  Avg. F size (L) (mm)      541 non-null    float64
 42  Avg. F size (R) (mm)      541 non-null    float64
 43  Endometrium (mm)          541 non-null    float64
dtypes: float64(19), int64(23), object(2)
memory usage: 186.1+ KB
```

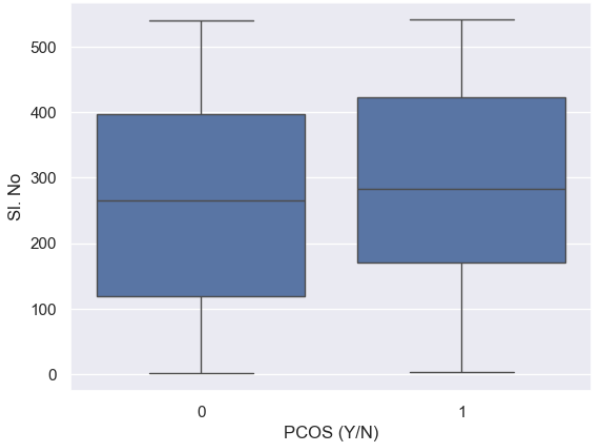
Missing Values

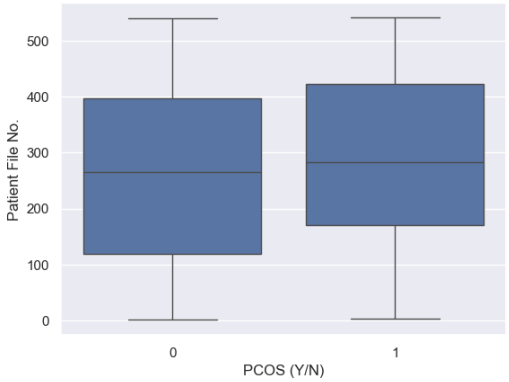
Sl. No	0
Patient File No.	0
PCOS (Y/N)	0
Age (yrs)	0
Weight (Kg)	0
Height(Cm)	0
BMI	300
Blood Group	0
Pulse rate(bpm)	0
RR (breaths/min)	0
Hb(g/dl)	0
Cycle(R/I)	0
Cycle length(days)	0
Marraige Status (Yrs)	1
Pregnant(Y/N)	0
No. of aborptions	0
I beta-HCG(mIU/mL)	0
II beta-HCG(mIU/mL)	1
FSH(mIU/mL)	0
LH(mIU/mL)	0
FSH/LH	532

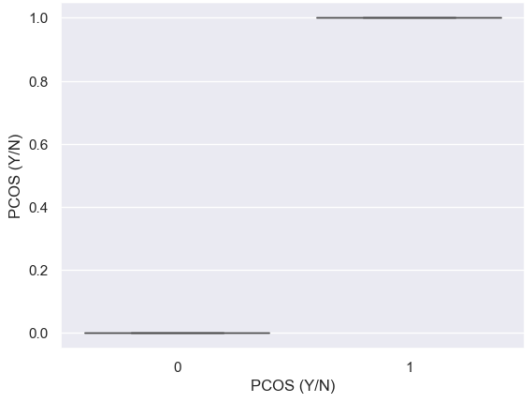
Hip(inch)	0
Waist(inch)	0
Waist:Hip Ratio	532
TSH (mIU/L)	0
AMH(ng/mL)	1
PRL(ng/mL)	0
Vit D3 (ng/mL)	0
PRG(ng/mL)	0
RBS(mg/dl)	0
Weight gain(Y/N)	0
hair growth(Y/N)	0
Skin darkening (Y/N)	0
Hair loss(Y/N)	0
Pimples(Y/N)	0
Fast food (Y/N)	1
Reg.Exercise(Y/N)	0
BP _Systolic (mmHg)	0
BP _Diastolic (mmHg)	0
Follicle No. (L)	0
Follicle No. (R)	0
Avg. F size (L) (mm)	0
Avg. F size (R) (mm)	0
Endometrium (mm)	0

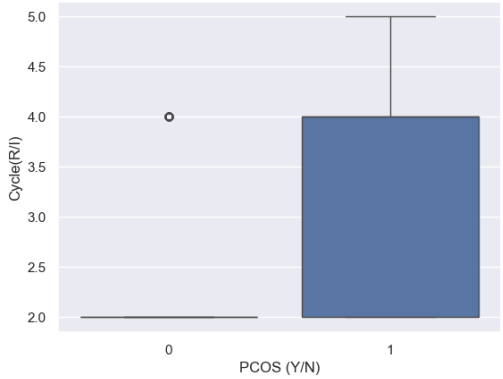
dtype: int64

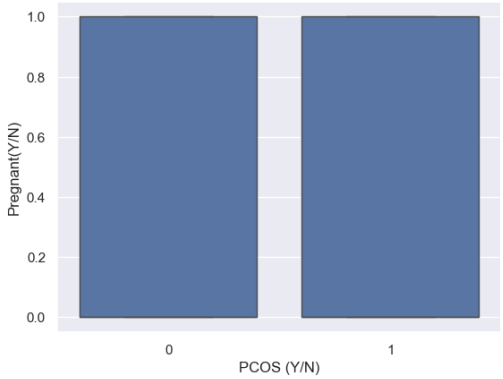
Descriptive Table

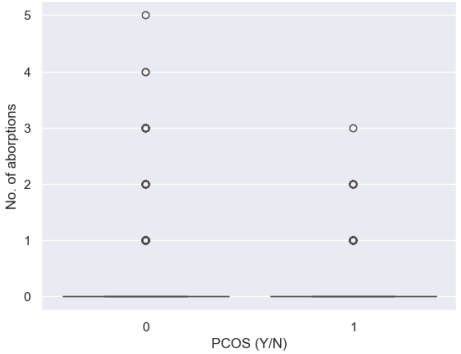




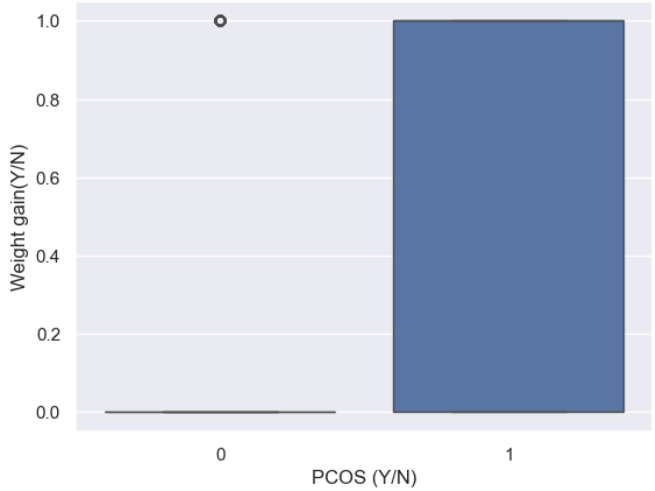


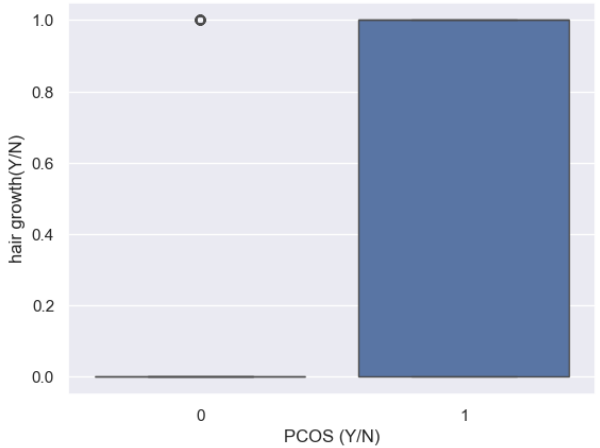


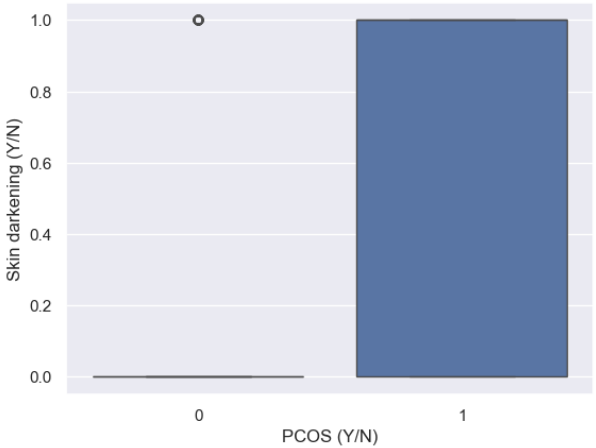


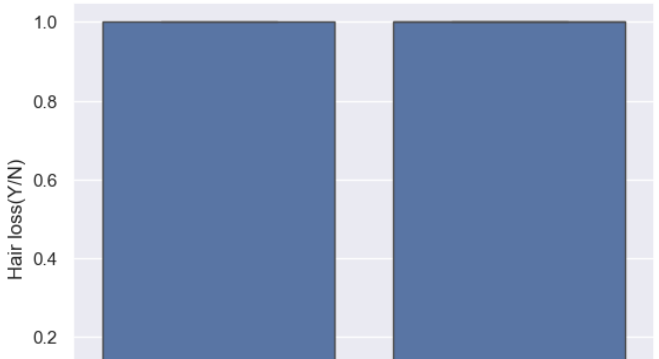


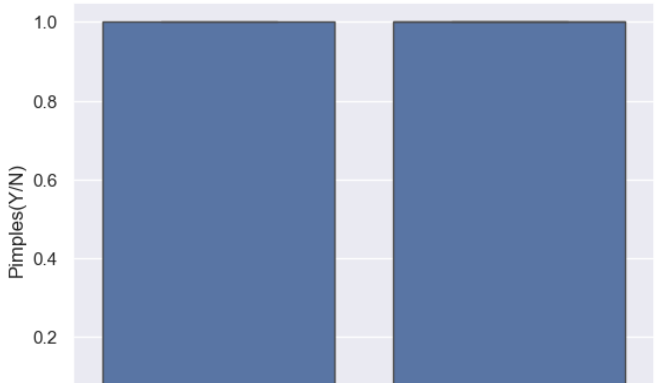
Descriptive Table

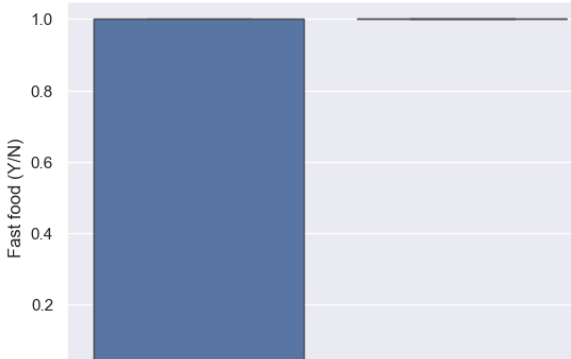




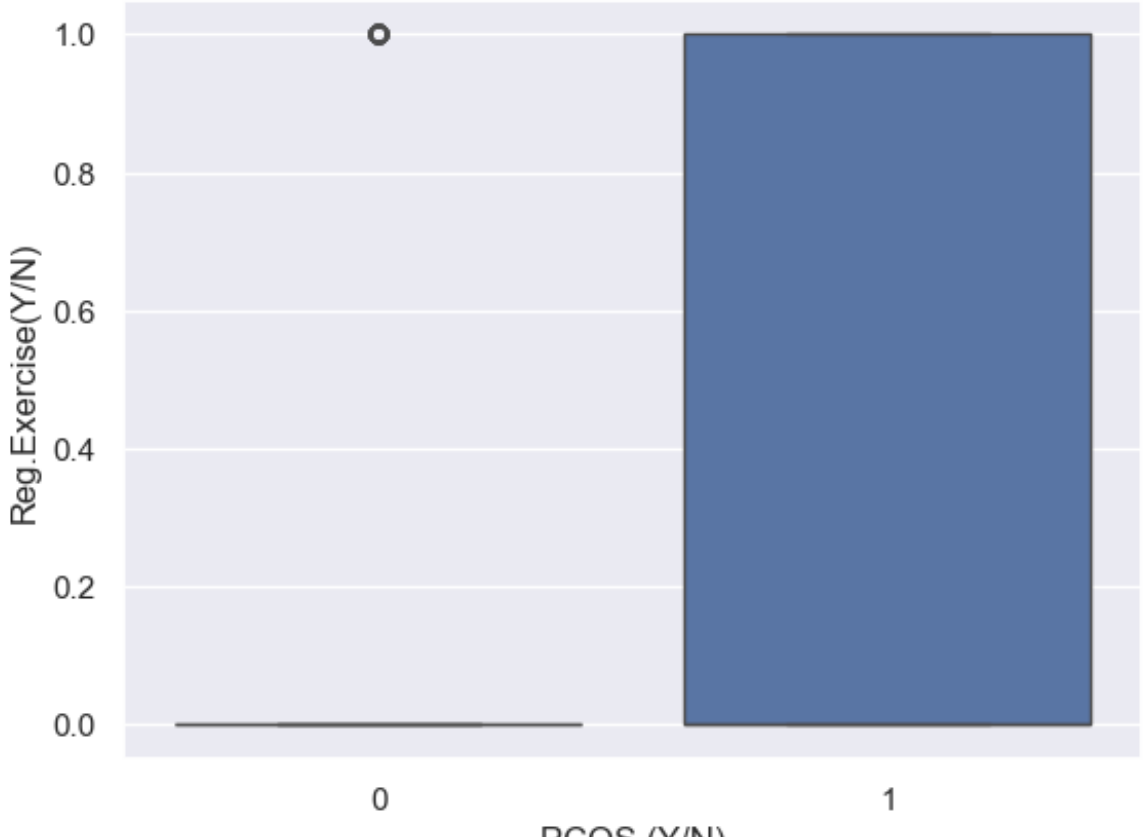




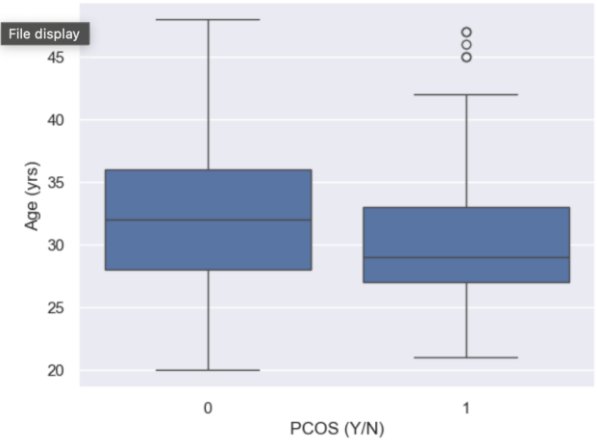


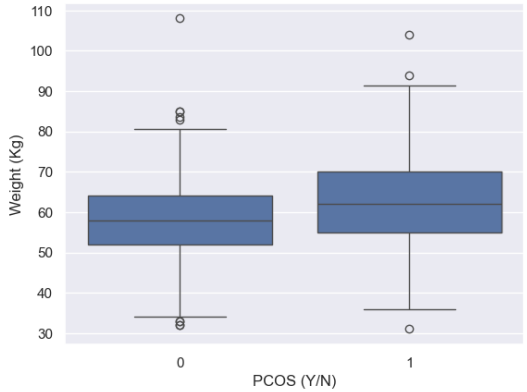


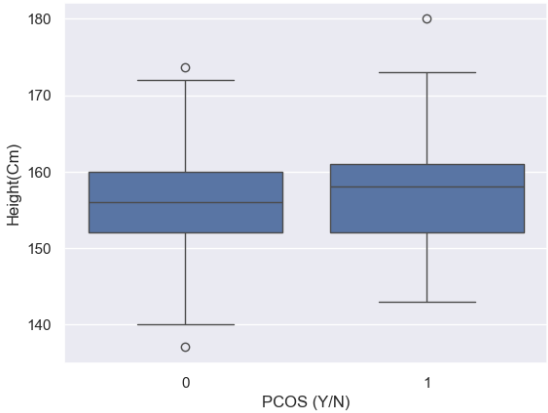
Descriptive Table

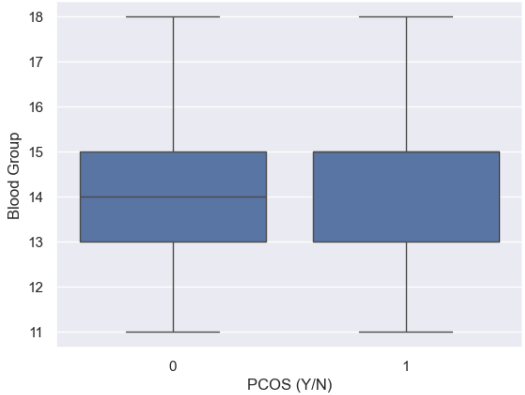


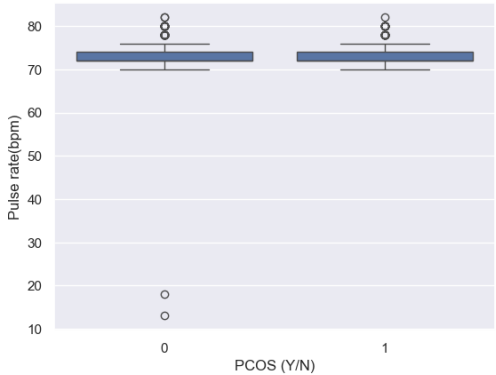
Exploratory Data Analysis

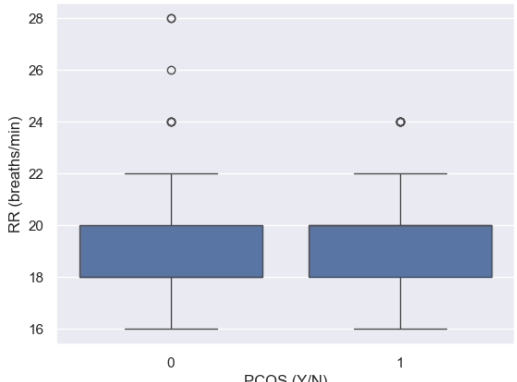


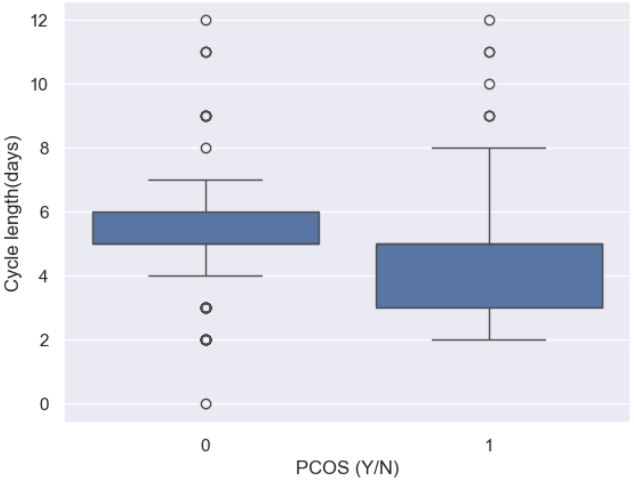


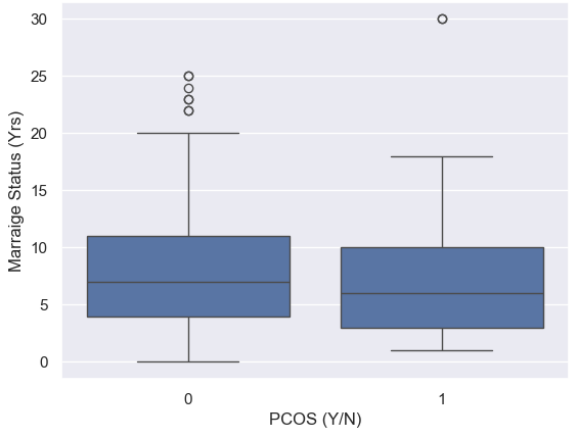


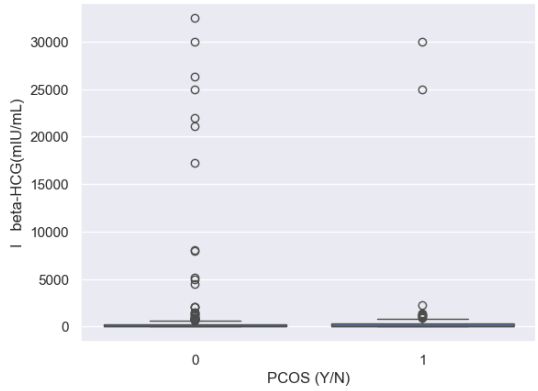


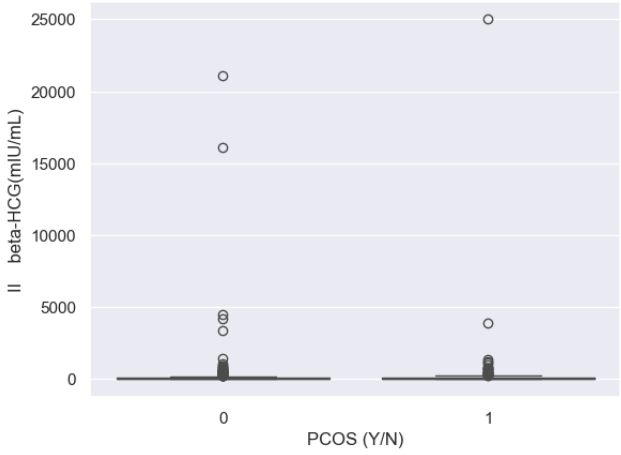


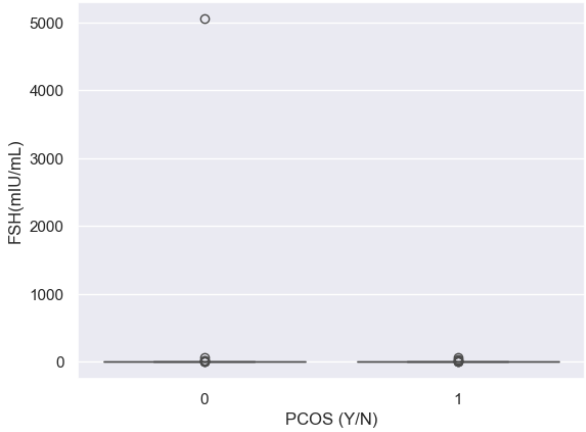


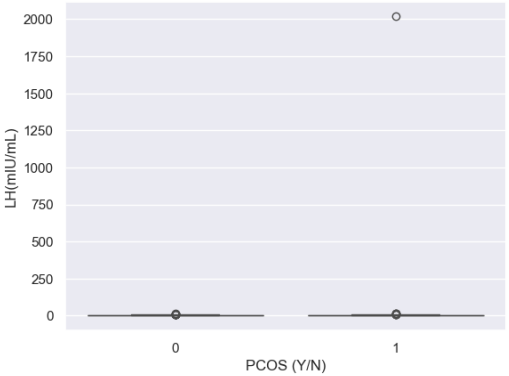

Cycle length(days)
PCOS (Y/N)

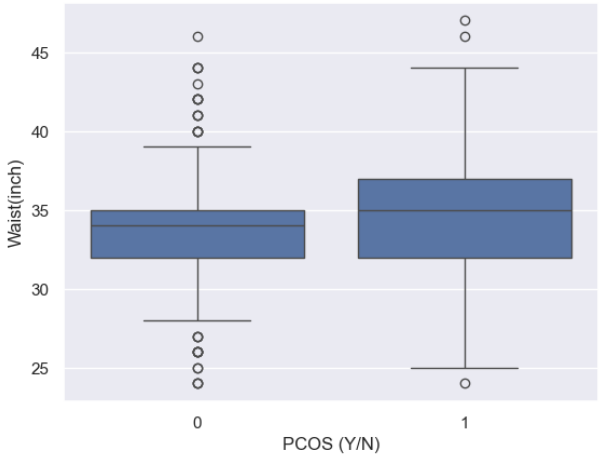

Marraige Status (Yrs)
PCOS (Y/N)

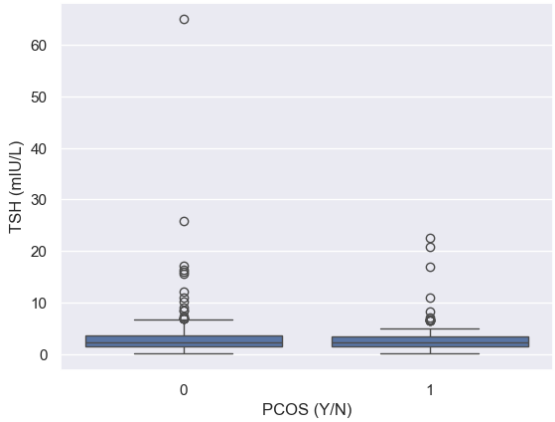

I beta-HCG(mIU/mL)
PCOS (Y/N)

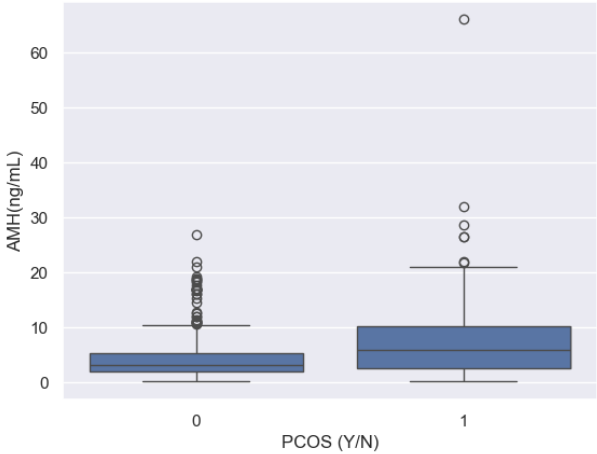

II beta-HCG(mIU/mL)
PCOS (Y/N)

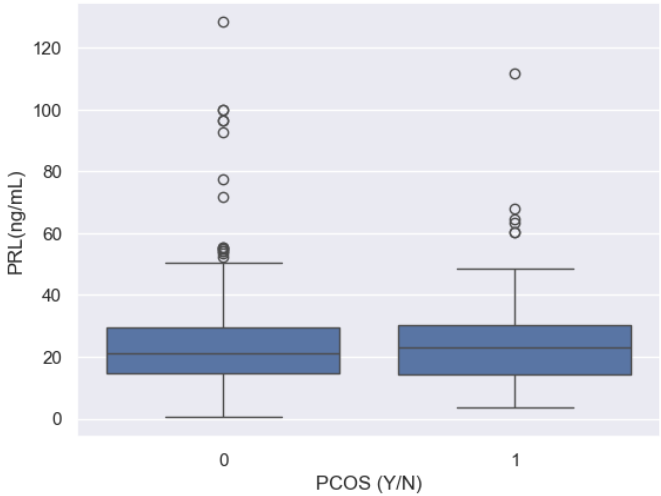

FSH(mIU/mL)
PCOS (Y/N)

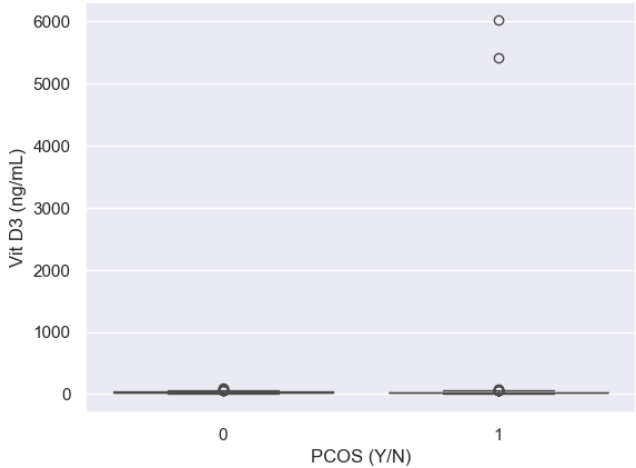

LH(mIU/mL)
PCOS (Y/N)

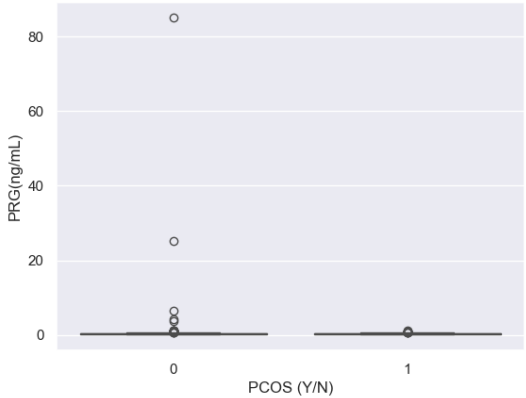

Waist(inch)
45
40
35
30
25
0
1
PCOS (Y/N)

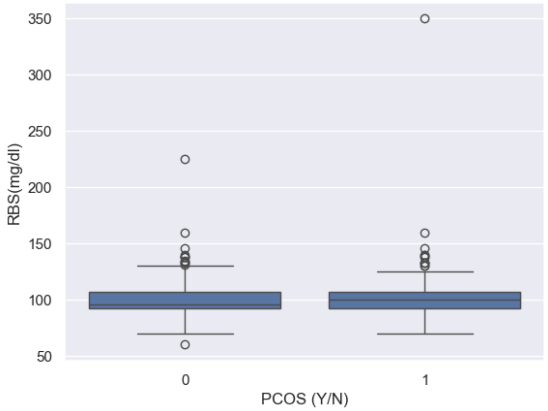

TSH (mIU/L)
60
50
40
30
20
10
0
0
1
PCOS (Y/N)

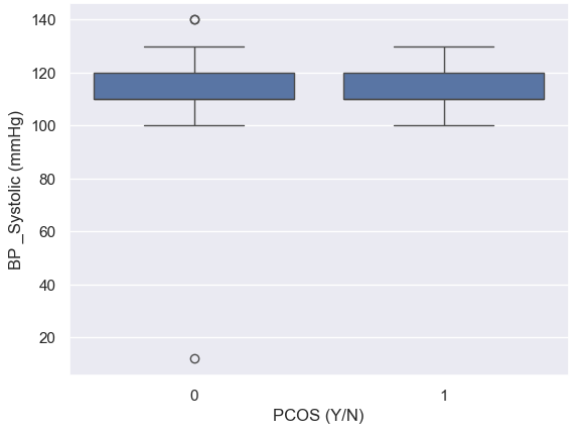

AMH(ng/mL)
60
50
40
30
20
10
0
0
1
PCOS (Y/N)

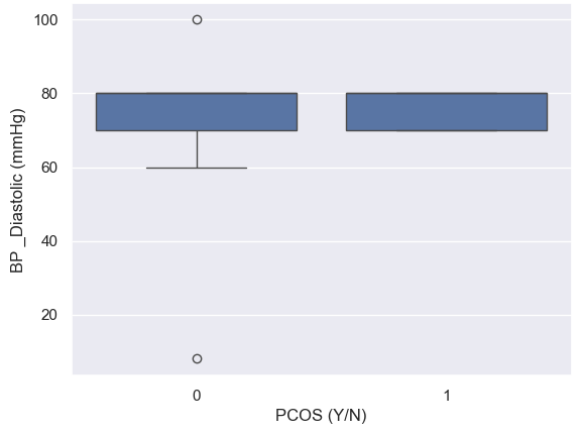

PRL(ng/mL)
120
100
80
60
40
20
0
0
1
PCOS (Y/N)

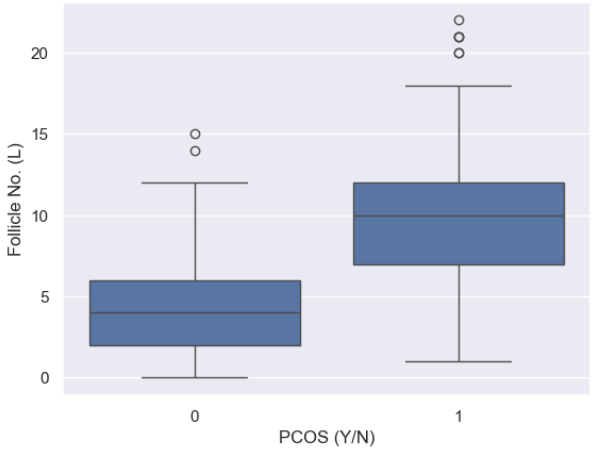

Vit D3 (ng/mL)
6000
5000
4000
3000
2000
1000
0
0
1
PCOS (Y/N)

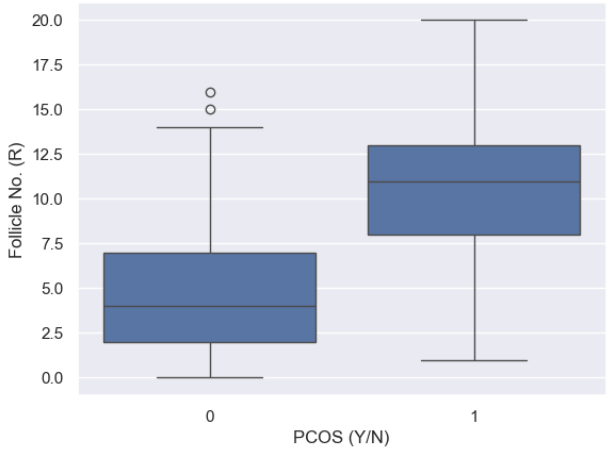

PRG(ng/mL)
80
60
40
20
0
0
1
PCOS (Y/N)

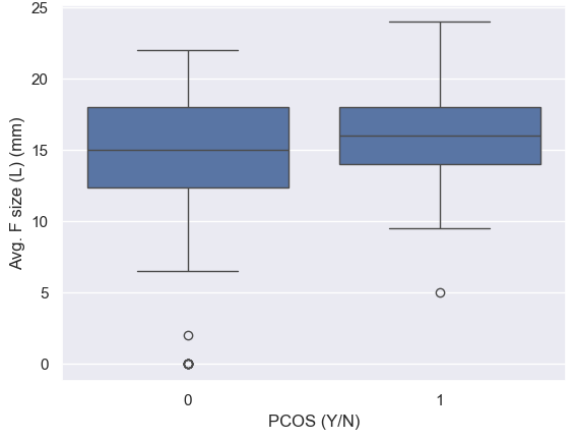

RBS(mg/dl)
350
300
250
200
150
100
50
0
1
PCOS (Y/N)

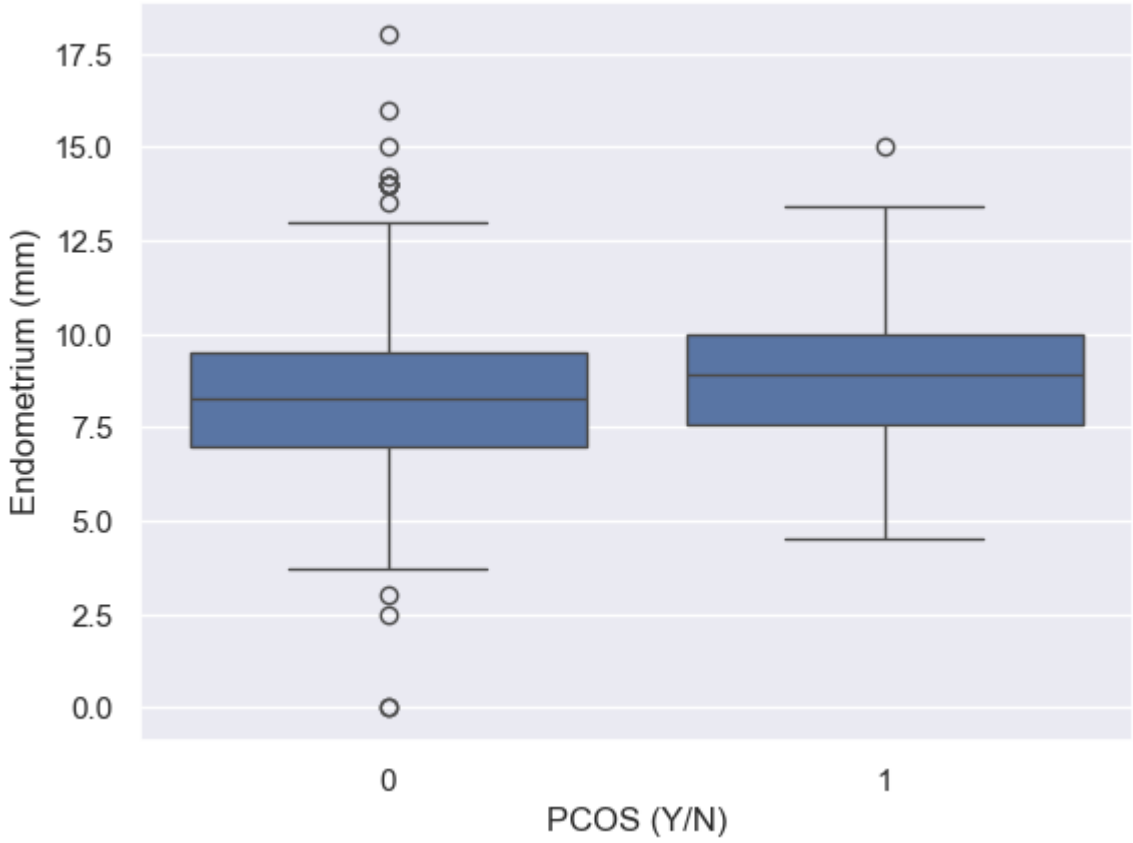

BP _Systolic (mmHg)
140
120
100
80
60
40
20
0
1
PCOS (Y/N)


BP _Diastolic (mmHg)
100
80
60
40
20
0
1
PCOS (Y/N)

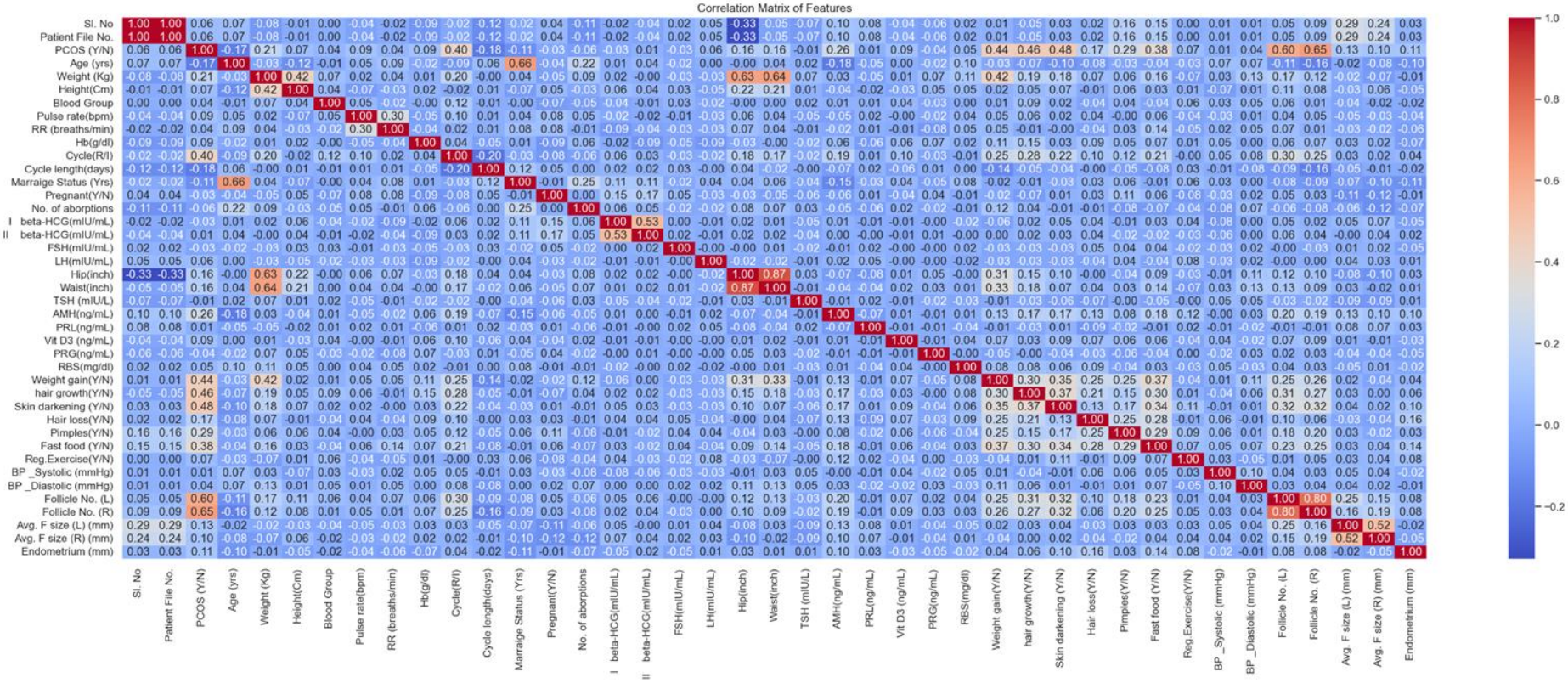

Follicle No. (L)
20
15
10
5
0
0
1
PCOS (Y/N)

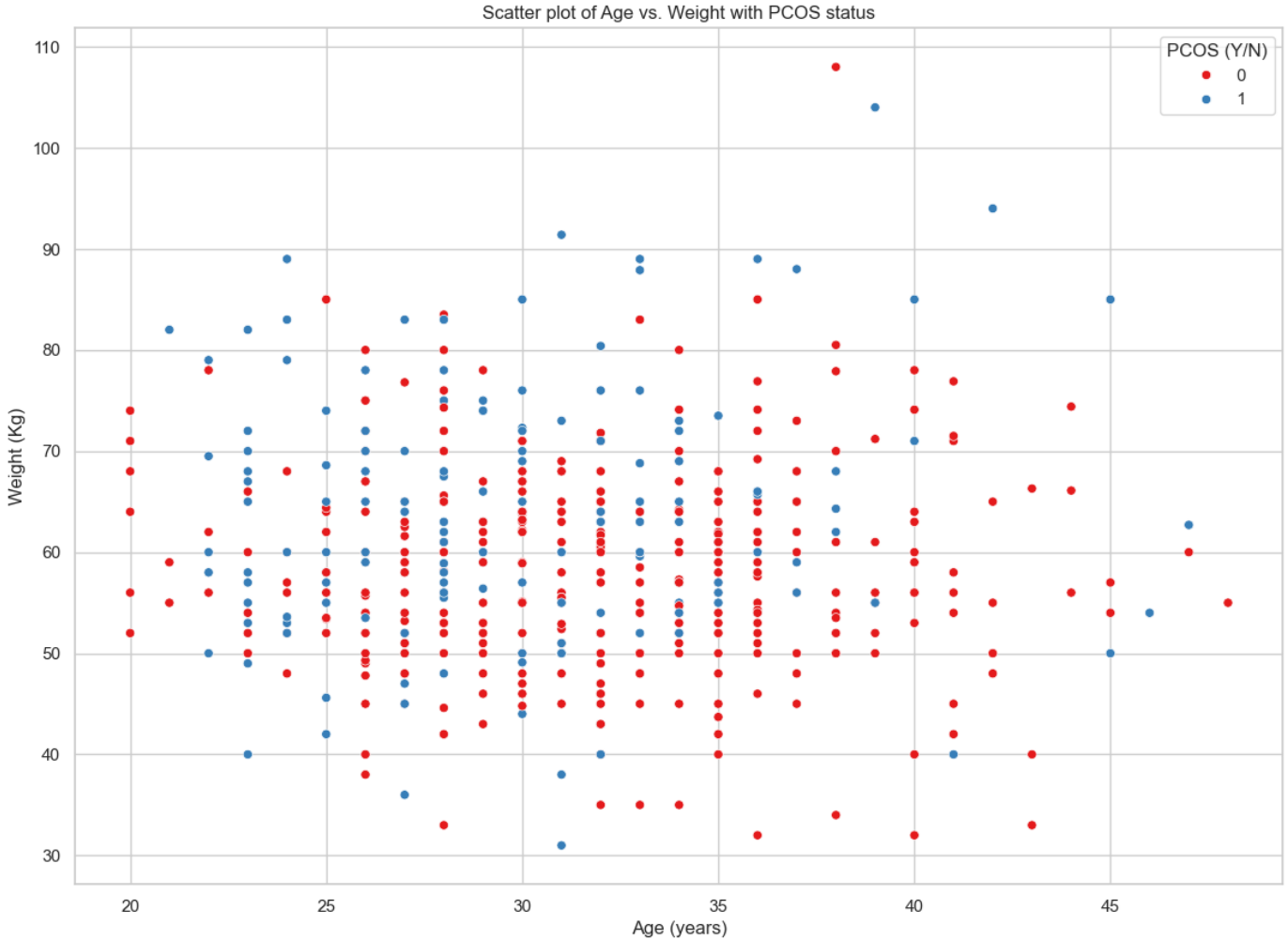

Follicle No. (R)
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0
1
PCOS (Y/N)


Avg. F size (L) (mm)
25
20
15
10
5
0
0
1
PCOS (Y/N)


17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
Endometrium (mm)
0
1
PCOS (Y/N)

Correlation Matrix of all the features




Scatter plot of Age vs. Weight with PCOS status
PCOS (Y/N)
0
1
Weight (Kg)
110
100
90
80
70
60
50
40
30
20
25
30
35
40
45
Age (years)

Next Steps

Model Selection:

- Choose appropriate machine learning models for binary classification, such as Logistic Regression, Decision Trees, Random Forests, or Support Vector Machines.

Model Training and Evaluation:

- Split the data into training and testing sets.
- Train the selected model on the training set.
- Evaluate the model's performance on the testing set using appropriate classification metrics (accuracy, precision, recall, F1 score).