

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)	540 non-null	float64
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
44	Unnamed: 44	1 non-null	float64

dtypes: float64(22), int64(23)

memory usage: 190.3 KB

```
data.isnull().sum()
```

Sl. No	0
Patient File No.	0
PCOS (Y/N)	0
Age (yrs)	0
Weight (Kg)	0
Height(Cm)	0
BMI	0
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)	0
Pregnant(Y/N)	0
No. of aborptions	0
I beta-HCG(mIU/mL)	0
II beta-HCG(mIU/mL)	0
FSH(mIU/mL)	0
LH(mIU/mL)	0
FSH/LH	0
Hip(inch)	0
Waist(inch)	0
Waist:Hip Ratio	0
TSH (mIU/L)	0
AMH(ng/mL)	0
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)	0
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
Unnamed: 44	0

dtype: int64

```
# Split data into training and testing sets
x_train, x_test, y_train, y_test = train_test_split(x,y, test_size=0.2, random_state=42)

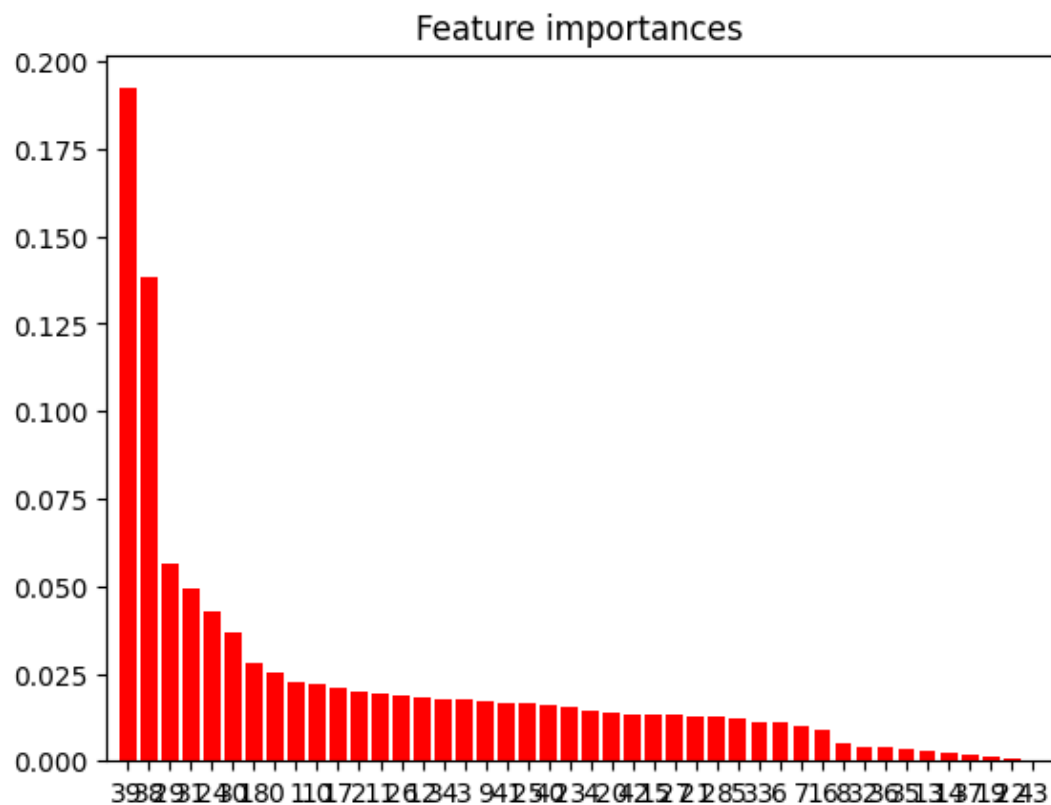
# Build RandomForestClassifier model
rf_classifier = RandomForestClassifier()
rf_classifier.fit(x_train, y_train)
```

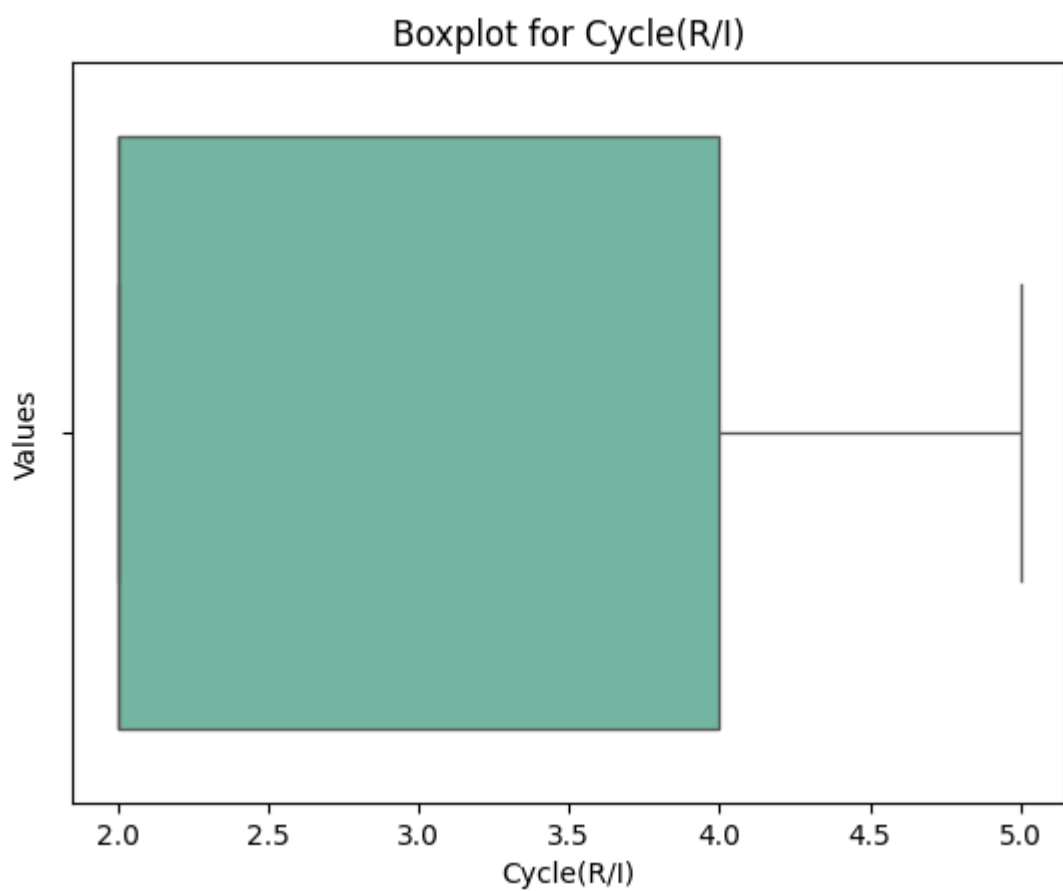
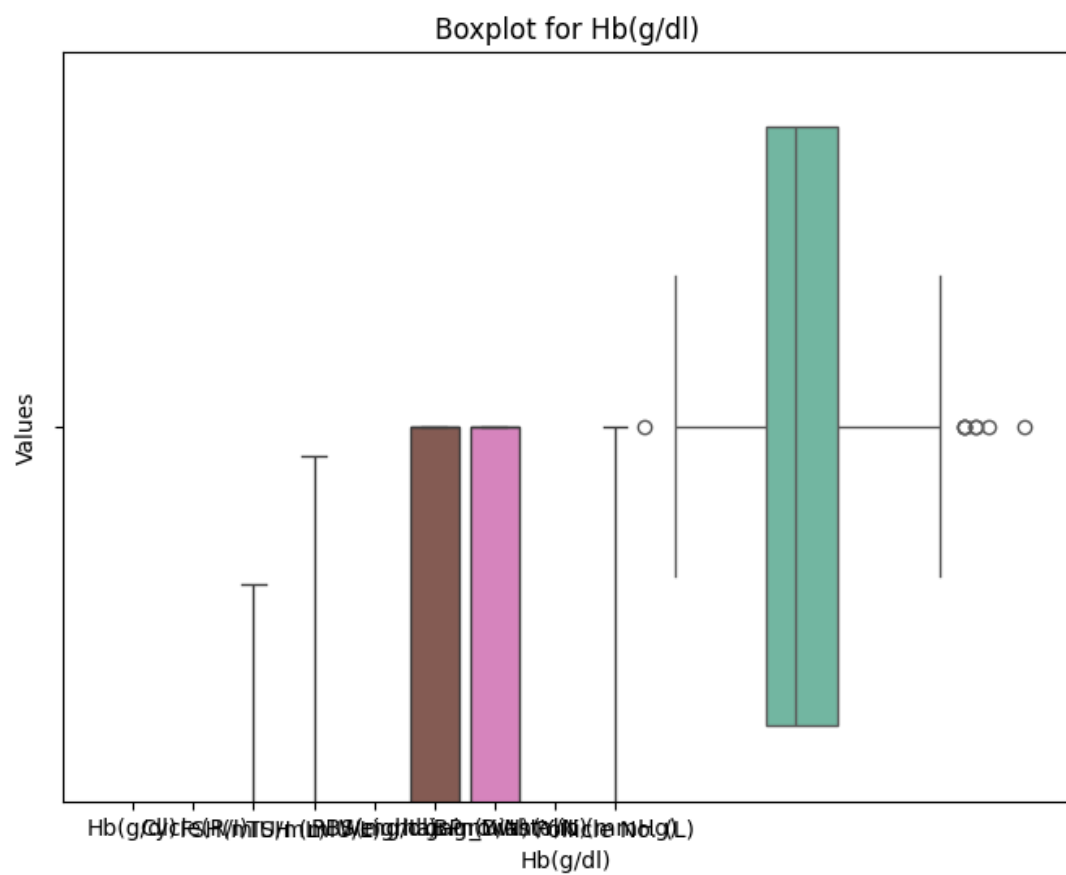
▼ RandomForestClassifier
RandomForestClassifier()

Feature ranking:

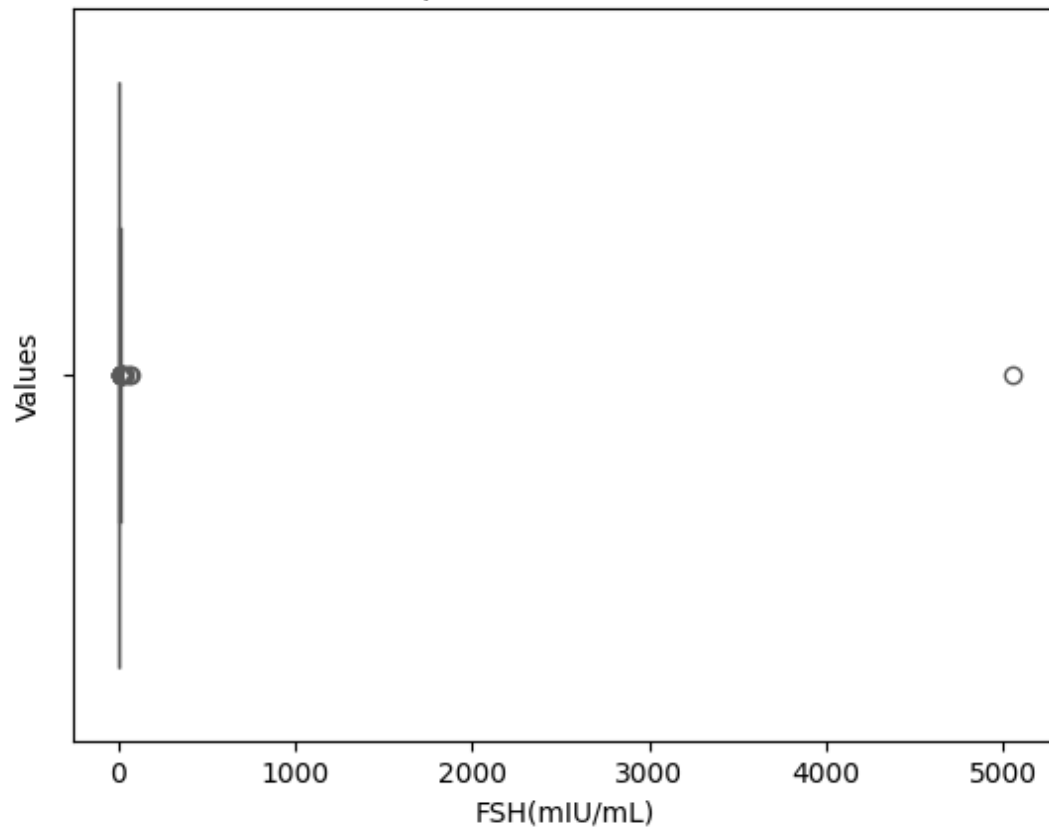
1. feature 39 (0.192279) Follicle No. (L)
2. feature 38 (0.138409) BP_Diastolic (mmHg)
3. feature 29 (0.056602) RBS(mg/dl)
4. feature 31 (0.049465) hair growth(Y/N)
5. feature 24 (0.042875) TSH (mIU/L)
6. feature 30 (0.036787) Weight gain(Y/N)
7. feature 18 (0.028215) FSH(mIU/mL)
8. feature 0 (0.025201) Sl. No
9. feature 1 (0.022400) Patient File No.
10. feature 10 (0.021934) Hb(g/dl)
11. feature 17 (0.021010) II beta-HCG(mIU/mL)
12. feature 2 (0.019750) PCOS (Y/N)
13. feature 11 (0.019522) Cycle(R/I)
14. feature 26 (0.018908) PRL(ng/mL)
15. feature 12 (0.018222) Cycle length(days)
16. feature 34 (0.017822) Pimples(Y/N)
17. feature 3 (0.017385) Age (yrs)
18. feature 9 (0.016892) RR (breaths/min)
19. feature 41 (0.016540) Avg. F size (L) (mm)
20. feature 25 (0.016444) AMH(ng/mL)
21. feature 40 (0.015817) Follicle No. (R)
22. feature 23 (0.015445) Waist:Hip Ratio
23. feature 4 (0.014129) Weight (Kg)
24. feature 20 (0.013861) FSH/LH
25. feature 42 (0.013286) Avg. F size (R) (mm)
26. feature 15 (0.013270) No. of abortions
27. feature 27 (0.013168) Vit D3 (ng/mL)
28. feature 21 (0.012692) Hip(inch)
29. feature 28 (0.012581) PRG(ng/mL)
30. feature 5 (0.012391) Height(Cm)
31. feature 33 (0.011264) Hair loss(Y/N)
32. feature 6 (0.010811) BMI
33. feature 7 (0.010198) Blood Group

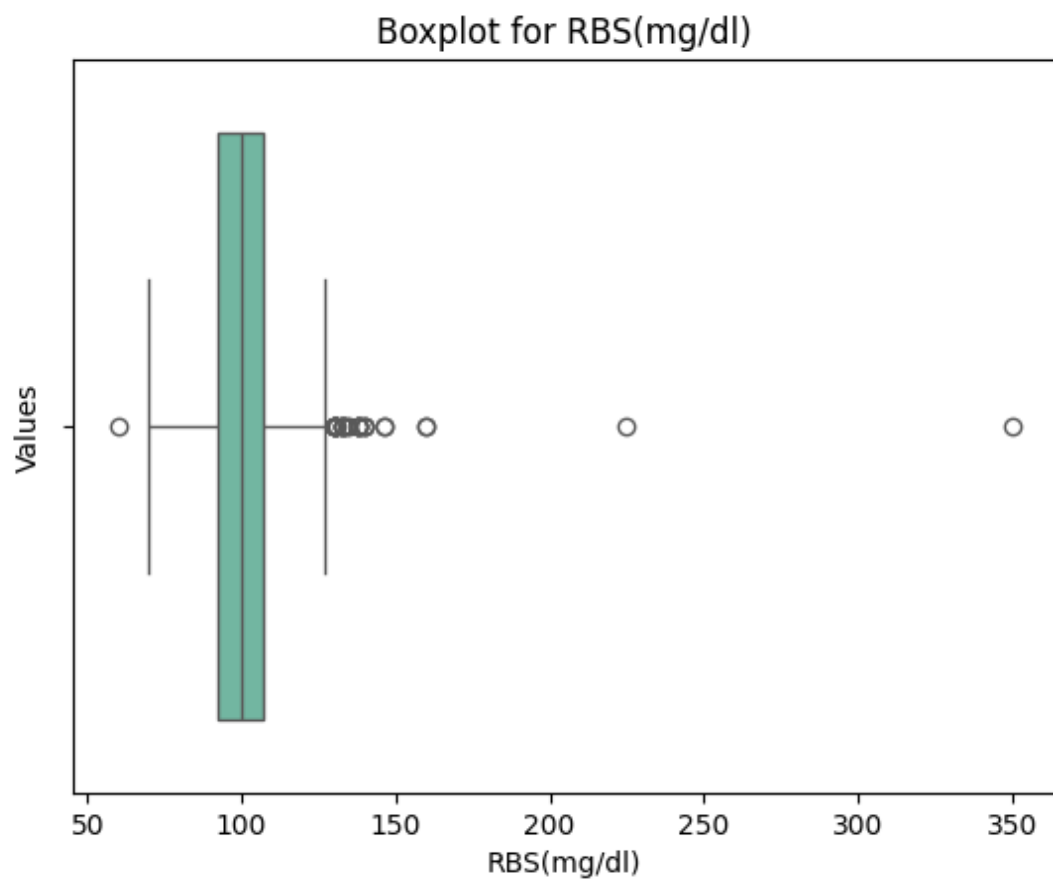
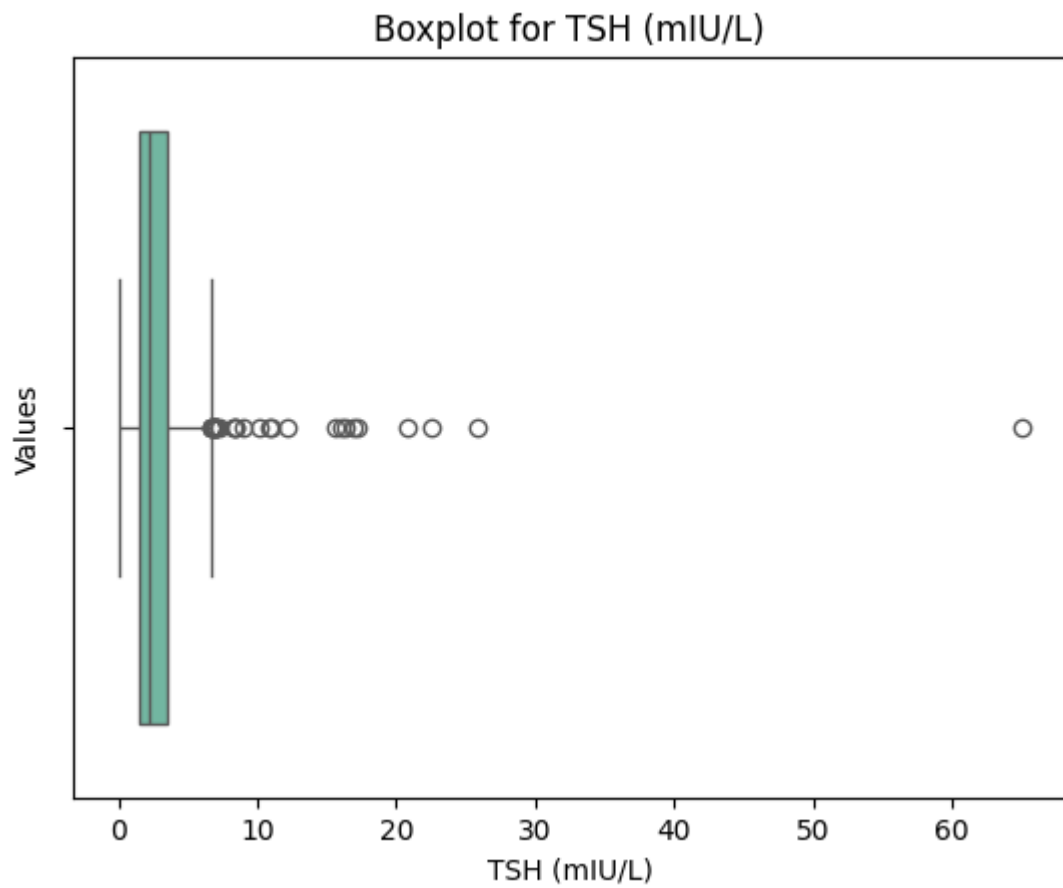
34. feature 16 (0.009147) I beta-HCG(mIU/mL)
 35. feature 8 (0.004801) Pulse rate(bpm)
 36. feature 32 (0.004153) Skin darkening (Y/N)
 37. feature 36 (0.003891) Reg.Exercise(Y/N)
 38. feature 35 (0.003615) Fast food (Y/N)
 39. feature 13 (0.002861) Marraige Status (Yrs)
 40. feature 14 (0.002200) Pregnant(Y/N)
 41. feature 37 (0.001784) BP _Systolic (mmHg)
 42. feature 19 (0.001082) LH(mIU/mL)
 43. feature 22 (0.000895) Waist(inch)
 44. feature 43 (0.000000) Endometrium (mm)



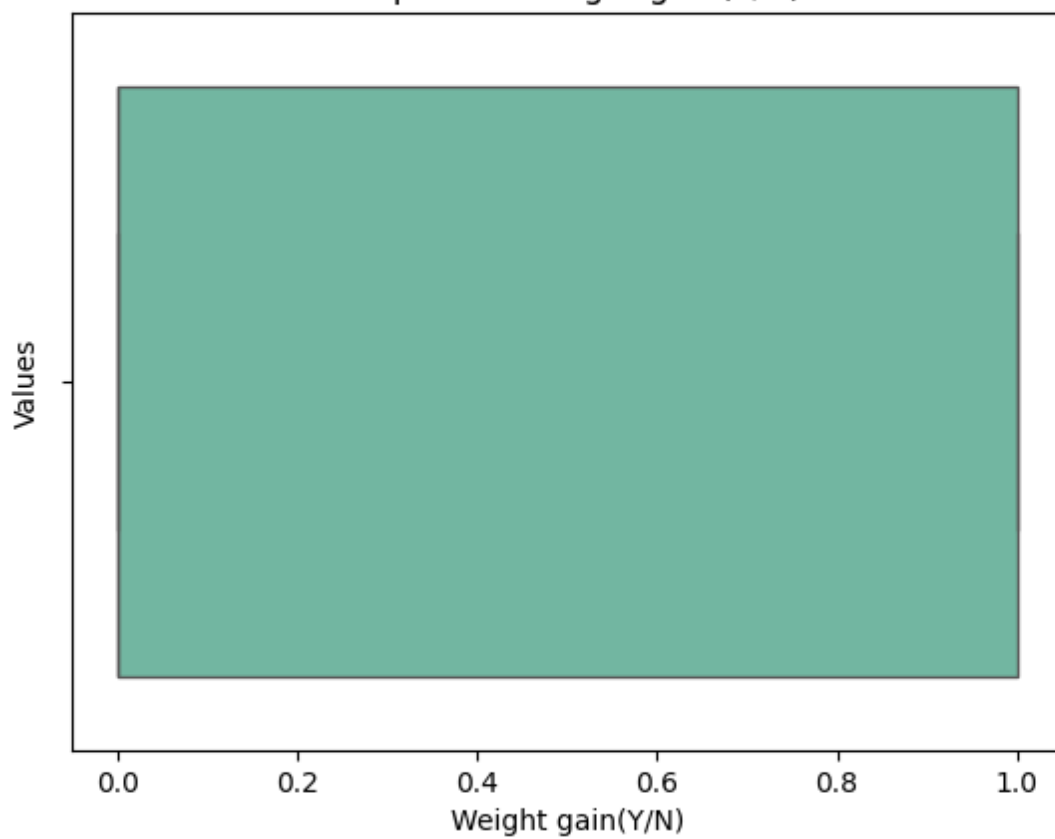


Boxplot for FSH(mIU/mL)

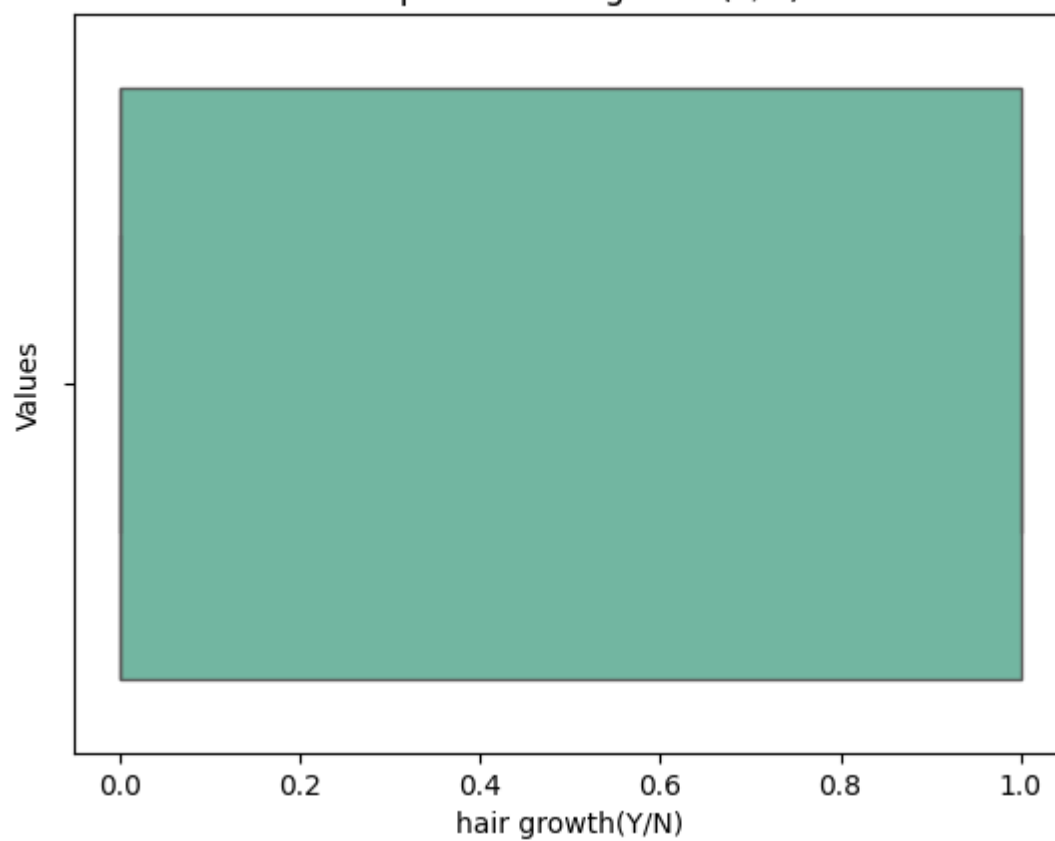




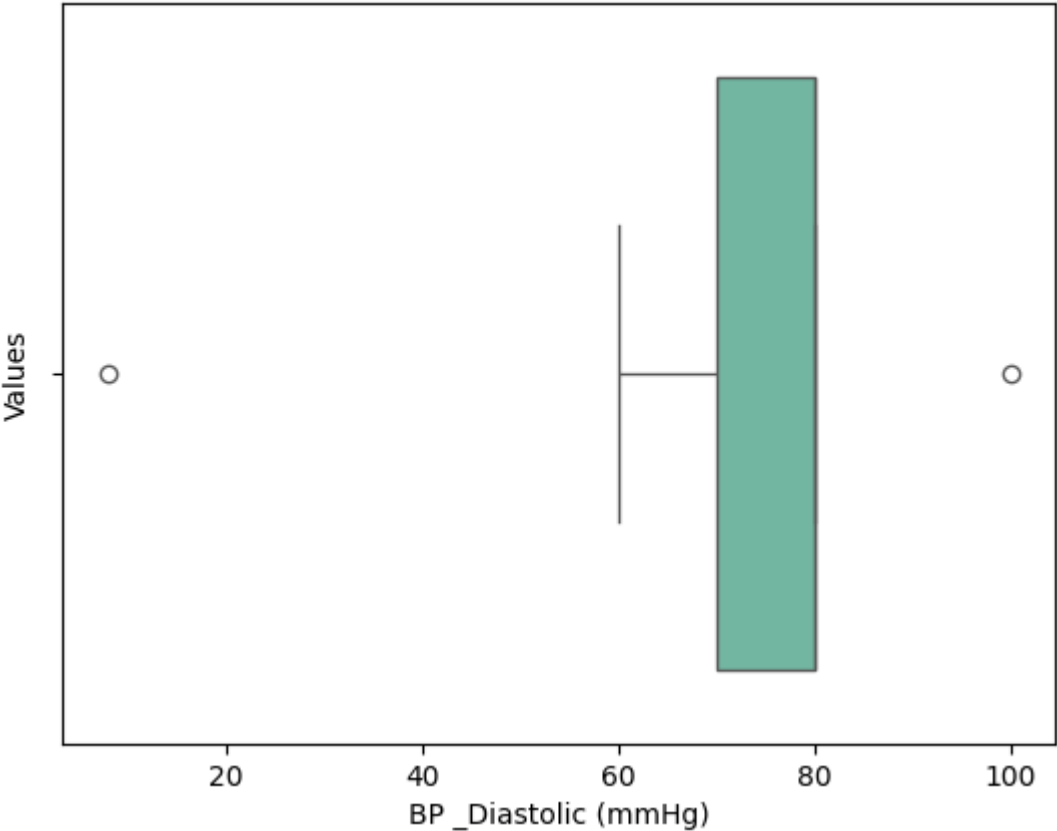
Boxplot for Weight gain(Y/N)



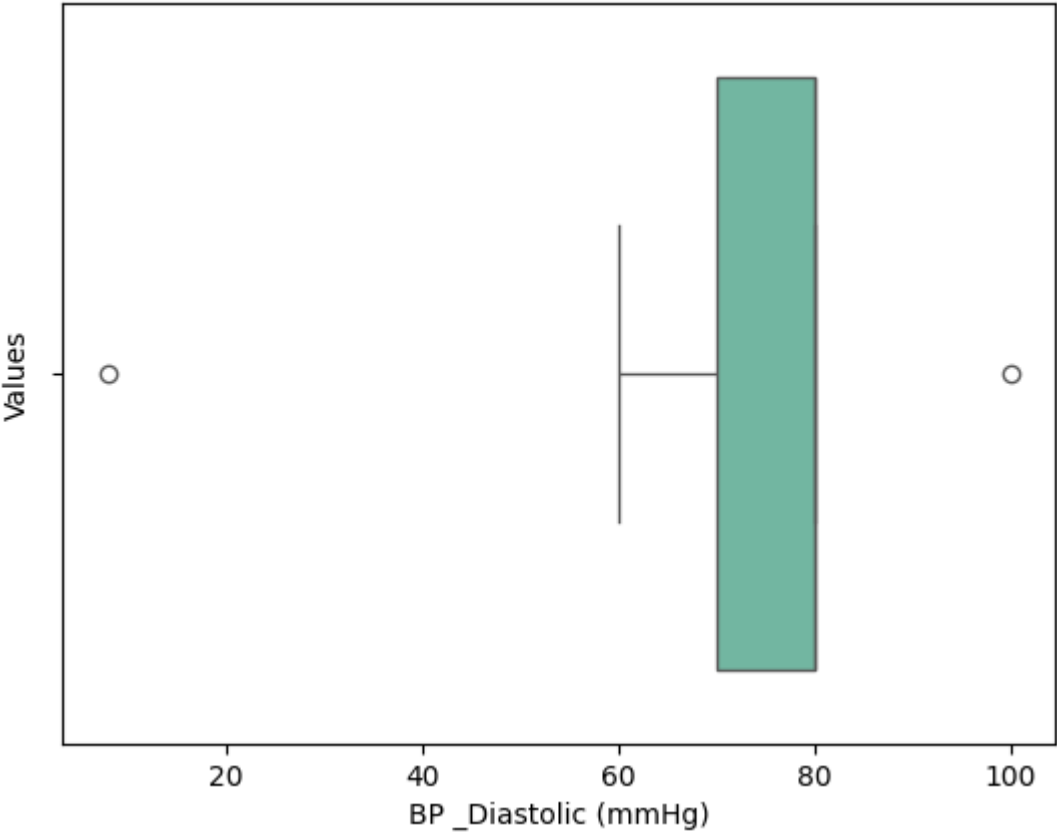
Boxplot for hair growth(Y/N)



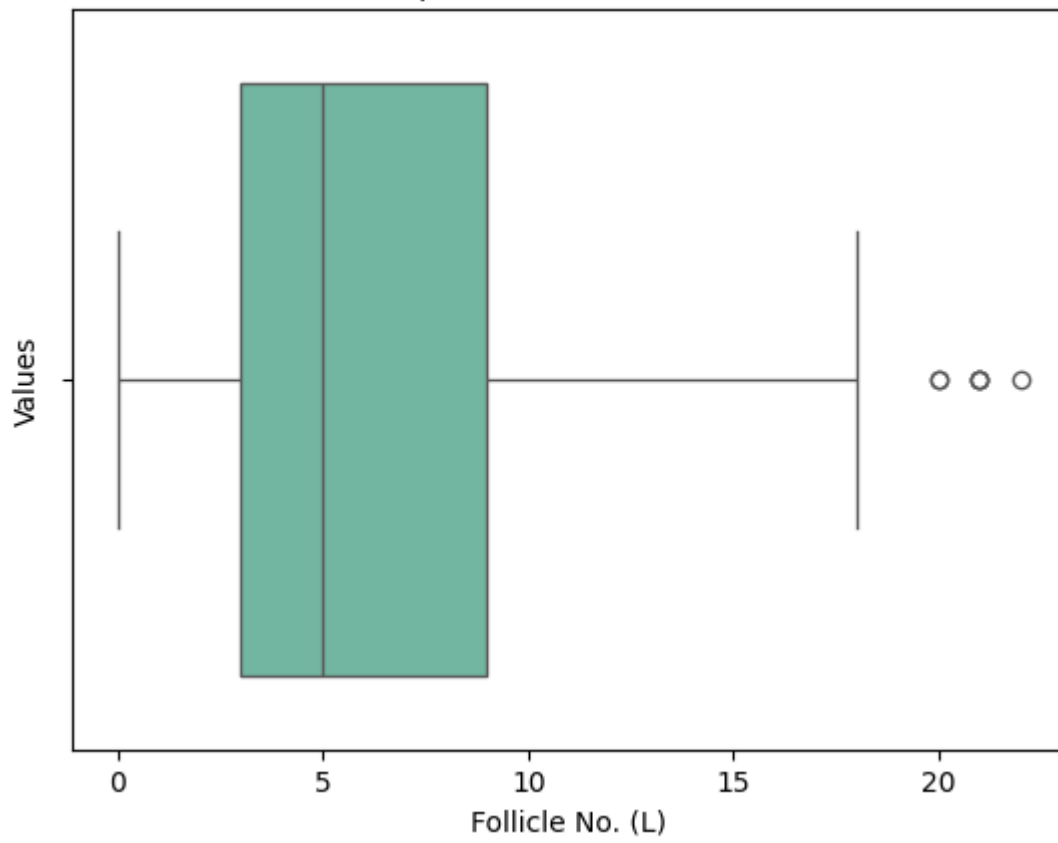
Boxplot for BP_Diastolic (mmHg)



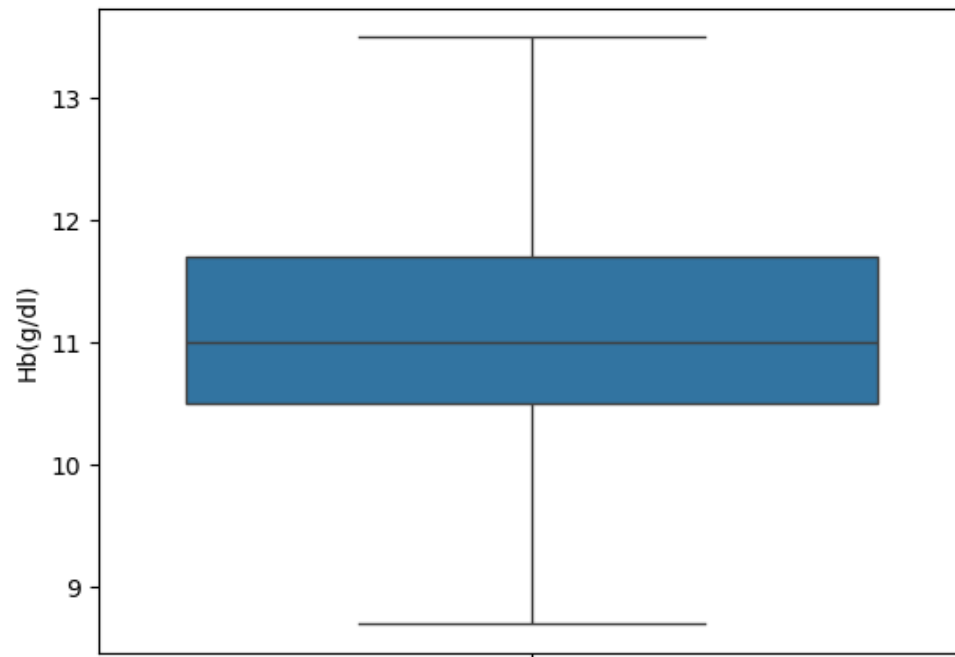
Boxplot for BP_Diastolic (mmHg)



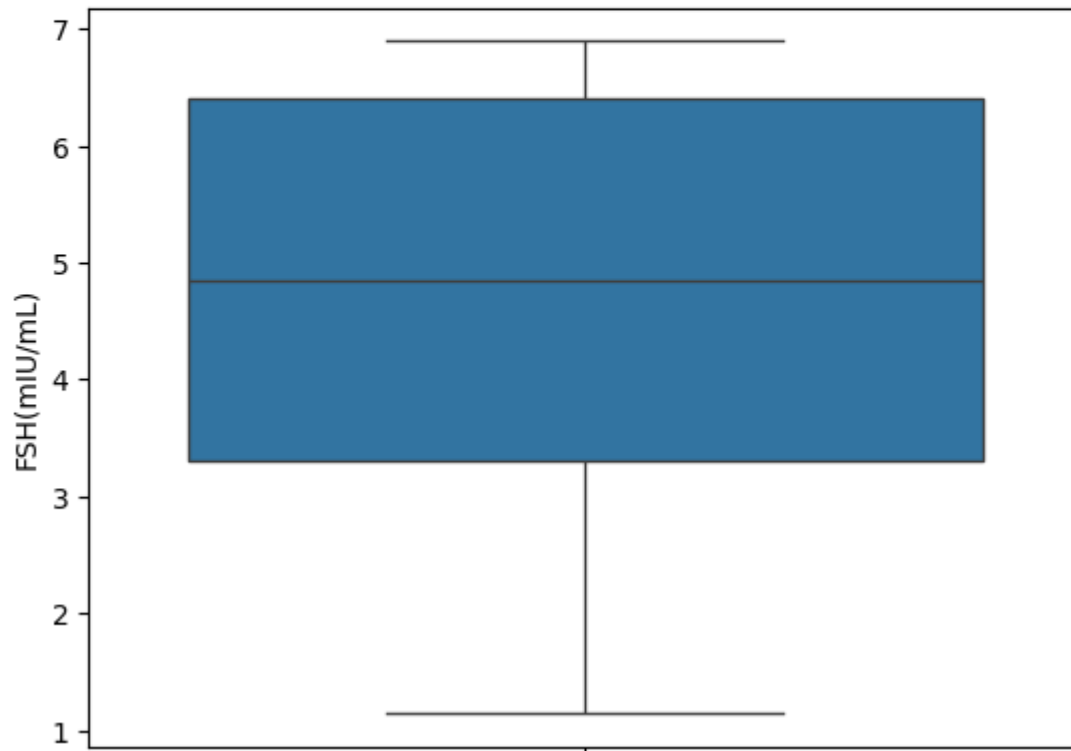
Boxplot for Follicle No. (L)



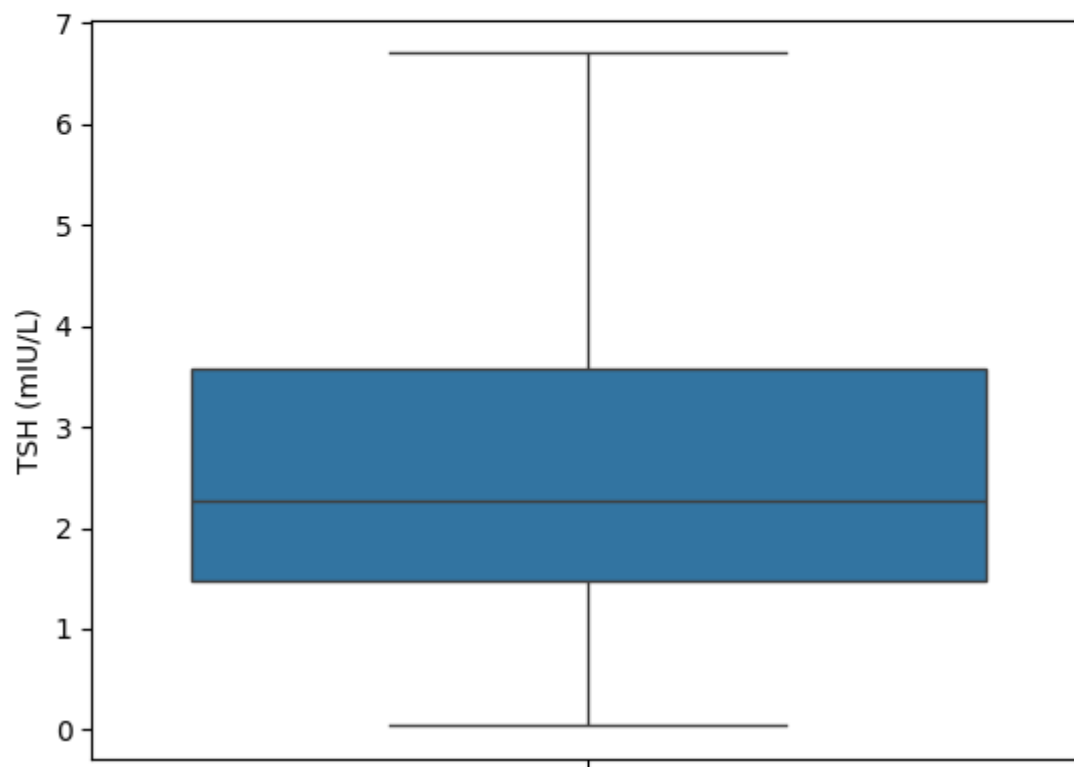
```
0.75    11.7
0.25    10.5
Name: Hb(g/dl), dtype: float64
11.7
10.5
1.1999999999999993
13.499999999999998
8.700000000000001
<Axes: ylabel='Hb(g/dl)'\>
```



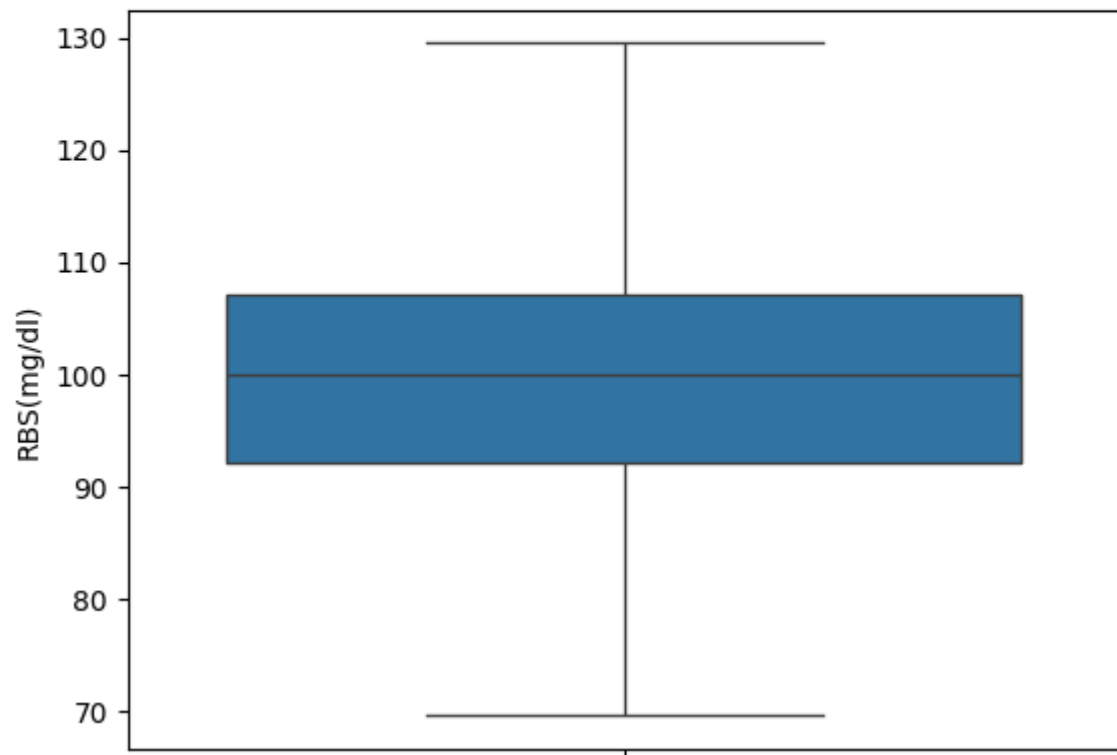
```
0.75    6.41
0.25    3.30
Name: FSH(mIU/mL), dtype: float64
6.41
3.3
1.1999999999999999
8.209999999999999
1.5000000000000001
<Axes: ylabel='FSH(mIU/mL)'\>
```



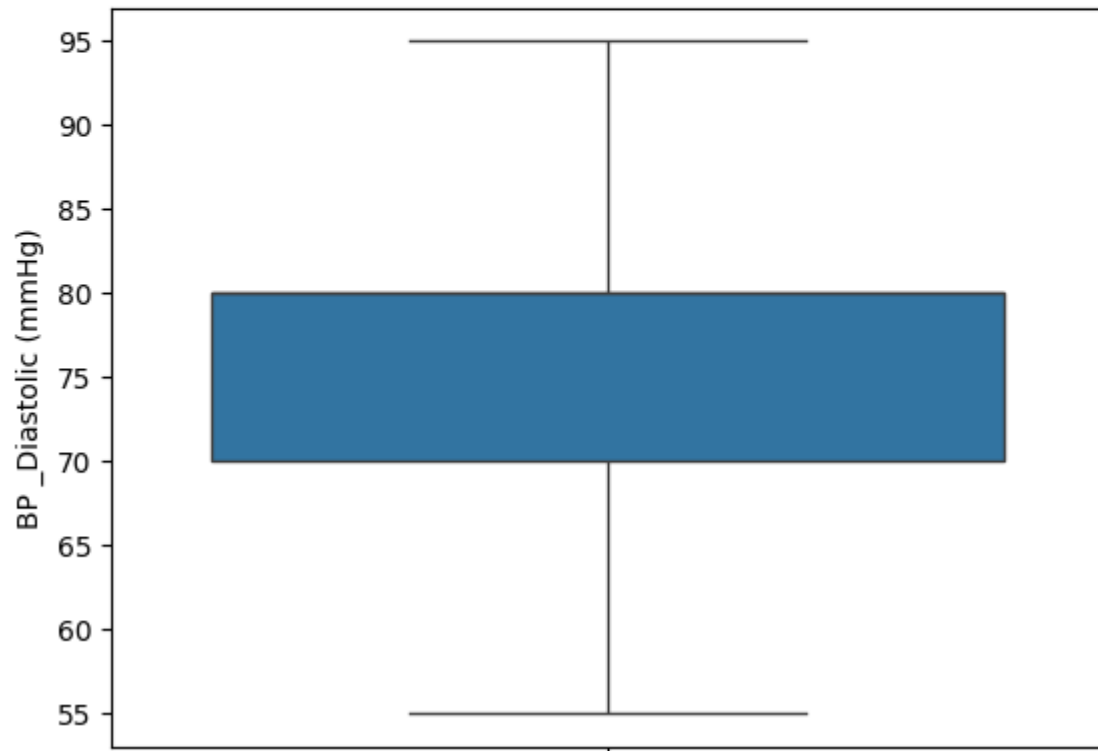
```
0.75    3.57
0.25    1.48
Name: TSH (mIU/L), dtype: float64
3.57
1.48
1.1999999999999993
5.369999999999999
-0.31999999999999895
<Axes: ylabel='TSH (mIU/L)'>
```



```
0.75    107.0
0.25     92.0
Name: RBS(mg/dl), dtype: float64
107.0
92.0
15.0
129.5
69.5
<Axes: ylabel='RBS(mg/dl)'\>
```



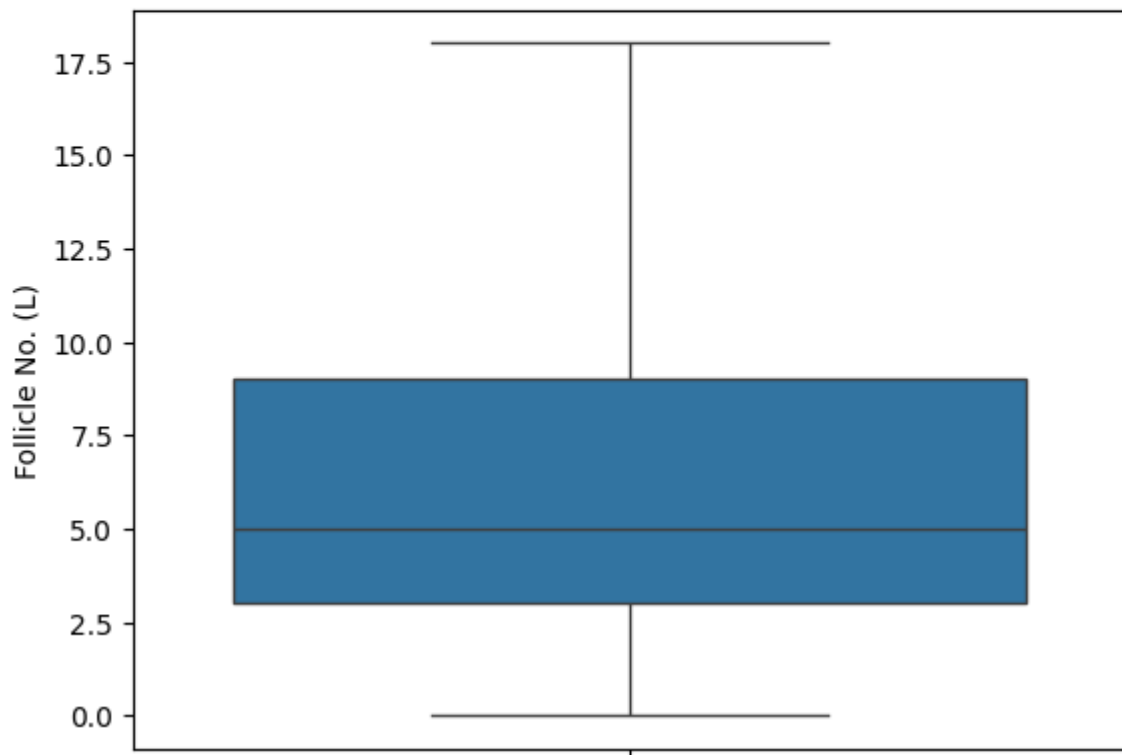
```
0.75    80.0
0.25    70.0
Name: BP_Diastolic (mmHg), dtype: float64
80.0
70.0
15.0
102.5
47.5
<Axes: ylabel='BP_Diastolic (mmHg)'>
```




```

0.75    9.0
0.25    3.0
Name: Follicle No. (L), dtype: float64
9.0
3.0
6.0
18.0
-6.0
<Axes: ylabel='Follicle No. (L)'>

```

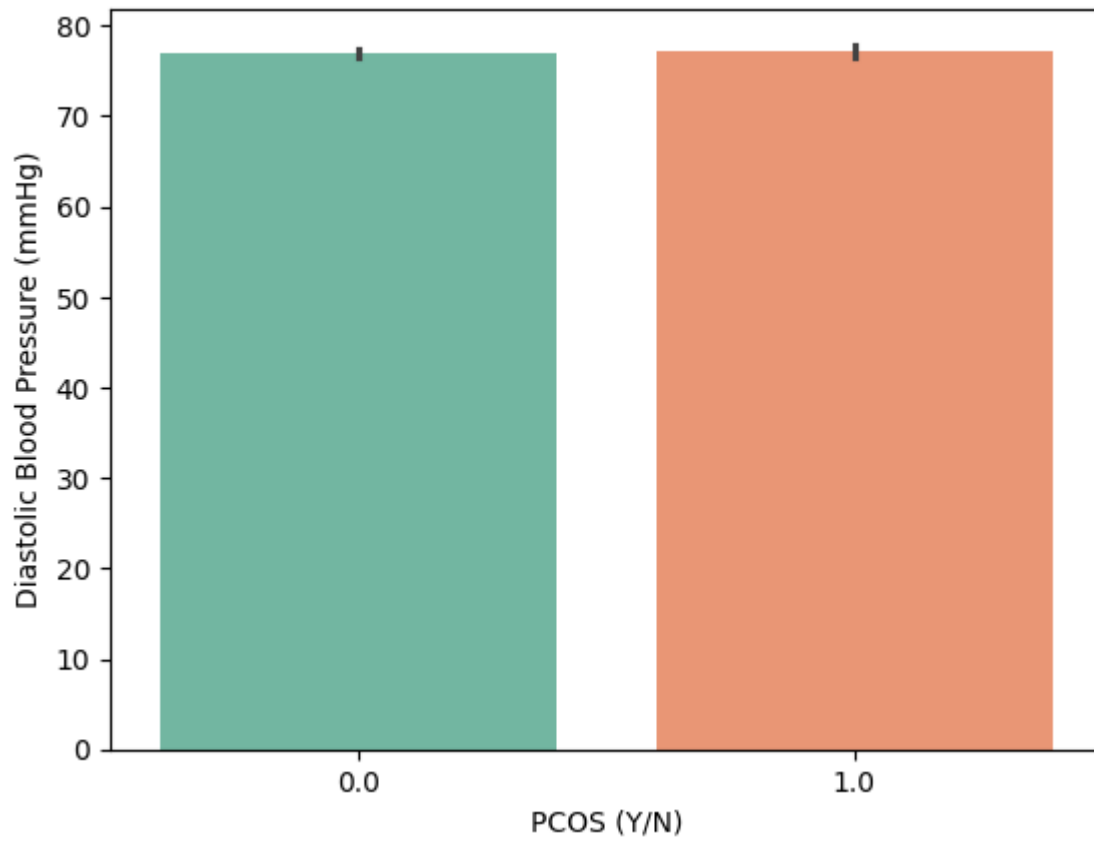


```
data.describe()
```

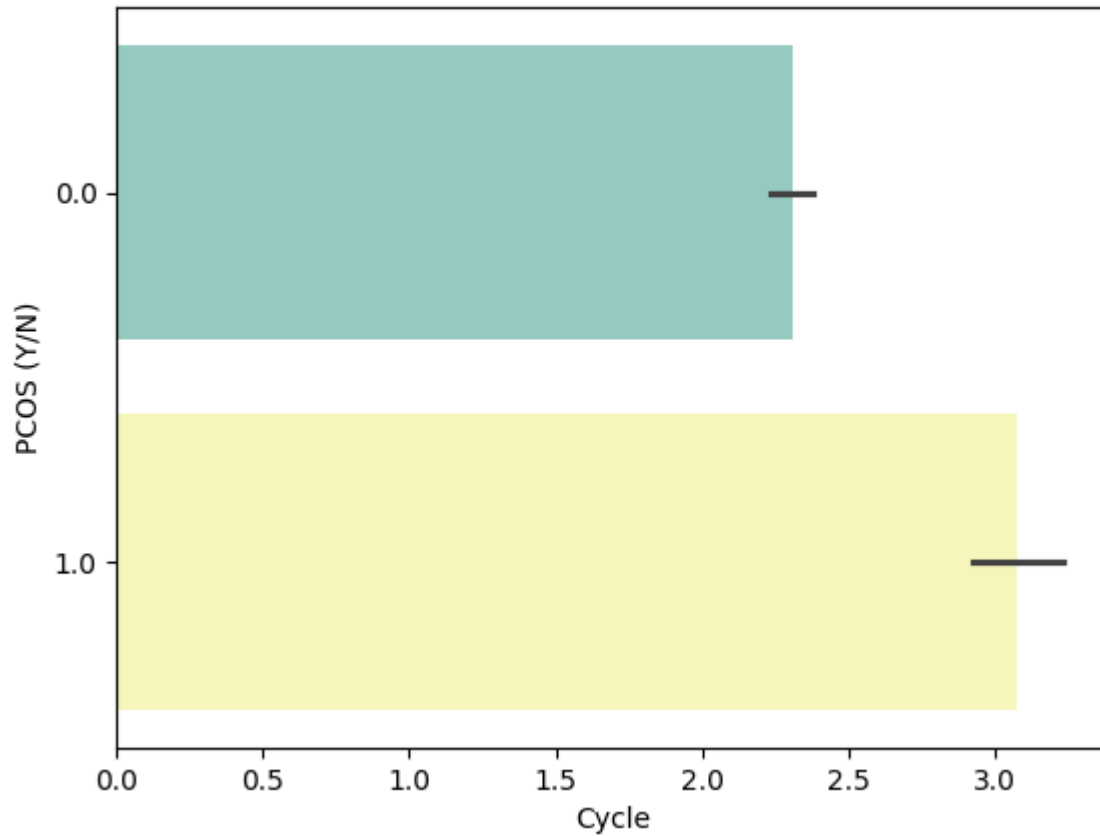
	Sl. No	Patient File No.	PCOS (Y/N)	Age (yrs)	Weight (Kg)	Height(Cm)	BMI	Blood Group	Pulse rate(bpm)	RR (breaths/min)	...	Fast food (Y/N)	Reg.Exercise(Y/N)
count	541.000000	541.000000	541.000000	541.000000	541.000000	541.000000	541.000000	541.000000	541.000000	541.000000	...	541.000000	541.000000
mean	271.000000	271.000000	0.327172	31.430684	59.637153	156.484835	23.929752	13.802218	73.247689	19.243993	...	0.514815	0.247689
std	156.317519	156.317519	0.469615	5.411006	11.028287	6.033545	2.447200	1.840812	4.430285	1.688629	...	0.499780	0.432070
min	1.000000	1.000000	0.000000	20.000000	31.000000	137.000000	15.100000	11.000000	13.000000	16.000000	...	0.000000	0.000000
25%	136.000000	136.000000	0.000000	28.000000	52.000000	152.000000	23.929752	13.000000	72.000000	18.000000	...	0.000000	0.000000
50%	271.000000	271.000000	0.000000	31.000000	59.000000	156.000000	23.929752	14.000000	72.000000	18.000000	...	1.000000	0.000000
75%	406.000000	406.000000	1.000000	35.000000	65.000000	160.000000	23.929752	15.000000	74.000000	20.000000	...	1.000000	0.000000
max	541.000000	541.000000	1.000000	48.000000	108.000000	180.000000	38.900000	18.000000	82.000000	28.000000	...	1.000000	1.000000

8 rows x 14 columns

Diastolic Blood Pressure vs. PCOS



Cycle (R/I) vs. PCOS



```
y_train.value_counts()
```

```
PCOS (Y/N)
0.0      287
1.0      145
Name: count, dtype: int64
```

```
ybal.value_counts()
```

```
PCOS (Y/N)
1.0      287
0.0      287
Name: count, dtype: int64
```

```
model_ = DecisionTreeClassifier()
model_.fit(X_train, y_train)
dt_pred = model_.predict(X_test)
acc_score = accuracy_score(y_test, dt_pred)
print("acc_score of model %.2f"%accuracy_score(y_test, dt_pred))
```

```
acc_score of model 0.80
```

```
from sklearn.ensemble import RandomForestClassifier
rfc = RandomForestClassifier(n_estimators=100)
my_model = rfc.fit(X_train, y_train)
#Making prediction and checking the test set
from sklearn.metrics import accuracy_score
pred_rfc = rfc.predict(X_test)
accuracy = accuracy_score(y_test, pred_rfc)
print(accuracy)
```

```
0.8715596330275229
```

```
from sklearn.metrics import accuracy_score
reg_model = LogisticRegression()
reg_model.fit(X_train, y_train)
reg_pred = reg_model.predict(X_test)
reg_acc_score = accuracy_score(y_test, reg_pred)
print("acc_score of model %.2f"%accuracy_score(y_test, reg_pred))
```

```
acc_score of model 0.88
```

x

	Hb(g/dl)	Cycle(R/I)	FSH(mIU/mL)	TSH (mIU/L)	RBS(mg/dl)	Weight gain(Y/N)	hair growth(Y/N)	BP _Diastolic (mmHg)	Follicle No. (L)
0	10.48	2.0	6.897407	0.680	92.0	0.0	0.0	80.0	3.0
1	11.70	2.0	6.730000	3.160	92.0	0.0	0.0	70.0	3.0
2	11.80	2.0	5.540000	2.540	84.0	0.0	0.0	80.0	13.0
3	12.00	2.0	6.897407	6.705	76.0	0.0	0.0	70.0	2.0
4	10.00	2.0	3.980000	3.570	84.0	0.0	0.0	80.0	3.0
...
536	11.00	2.0	6.897407	1.110	92.0	0.0	0.0	70.0	1.0
537	10.80	2.0	5.070000	2.050	108.0	1.0	0.0	70.0	9.0
538	10.80	2.0	6.897407	2.870	92.0	0.0	0.0	80.0	1.0
539	12.00	4.0	4.400000	2.500	115.0	0.0	0.0	70.0	7.0
540	10.20	4.0	3.990000	1.660	108.0	1.0	1.0	70.0	9.0

541 rows x 9 columns

y

```
0      0.0
1      0.0
2      1.0
3      0.0
4      0.0
...
536    0.0
537    0.0
538    0.0
539    0.0
540    1.0
```

Name: PCOS (Y/N), Length: 541, dtype: float64

```
print(model_.predict([[10.20,4.0,3.990000,1.660,108.0,1.0,1.0,70.0,9.0]]))
```

```
[1.]
```

```
from sklearn.metrics import classification_report
print(classification_report (reg_pred, y_test))
```

	precision	recall	f1-score	support
0.0	0.96	0.88	0.92	84
1.0	0.69	0.88	0.77	25
accuracy			0.88	109
macro avg	0.82	0.88	0.85	109
weighted avg	0.90	0.88	0.89	109

```
print(classification_report(dt_pred, y_test))
```

	precision	recall	f1-score	support
0.0	0.87	0.85	0.86	79
1.0	0.62	0.67	0.65	30
accuracy			0.80	109
macro avg	0.75	0.76	0.75	109
weighted avg	0.80	0.80	0.80	109

```
from sklearn.metrics import classification_report
classi_report = classification_report (y_test, pred_rfc)
print(classi_report)
```

	precision	recall	f1-score	support
0.0	0.87	0.96	0.91	77
1.0	0.88	0.66	0.75	32
accuracy			0.87	109
macro avg	0.87	0.81	0.83	109
weighted avg	0.87	0.87	0.87	109

```
print(classification_report (reg_pred, y_test))
```

	precision	recall	f1-score	support
0.0	0.96	0.88	0.92	84
1.0	0.69	0.88	0.77	25
accuracy			0.88	109
macro avg	0.82	0.88	0.85	109
weighted avg	0.90	0.88	0.89	109