POLYCYSTIC OVARIAN SYNDROME SYMPTOMS

Exploratory Data Analysis

Fall 2022 | Sarah Beth Powell

WHAT YOU SHOULD KNOW ABOUT

PCOS

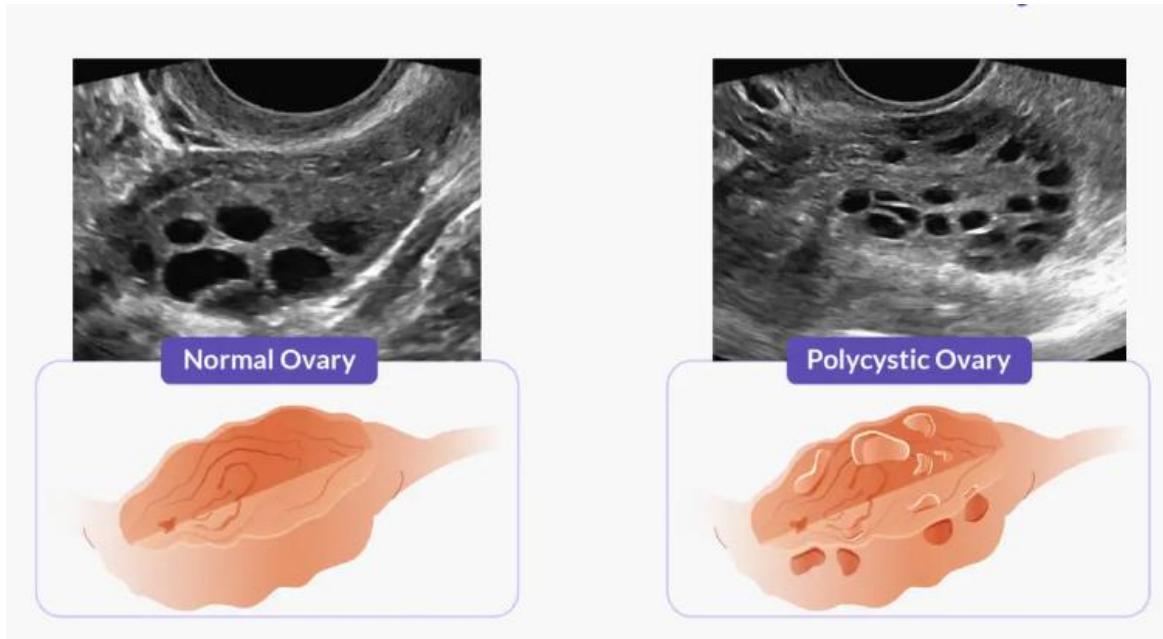### (POLYCYSTIC OVARY SYNDROME)

What is PCOS?

Affecting an estimated 5 to 6 million women in the United States, PCOS is the most common cause for infertility. The exact cause of PCOS is unknown, but research suggests that elevated levels of androgens and insulin maybe related to the condition. PCOS impacts women of all ages, even post-child bearing age.

Common Symptoms & Impacts

Excess hair growth on face & body
Irregular cycles
Infertility
Weight gain
Diabetes
Acne

An estimated
$8 Billion
was spent to diagnose & treat PCOS just in the year 2020.

Reference to Endocrine Society

ABOUT THE DATA

Normal Ovary

Polycystic Ovary

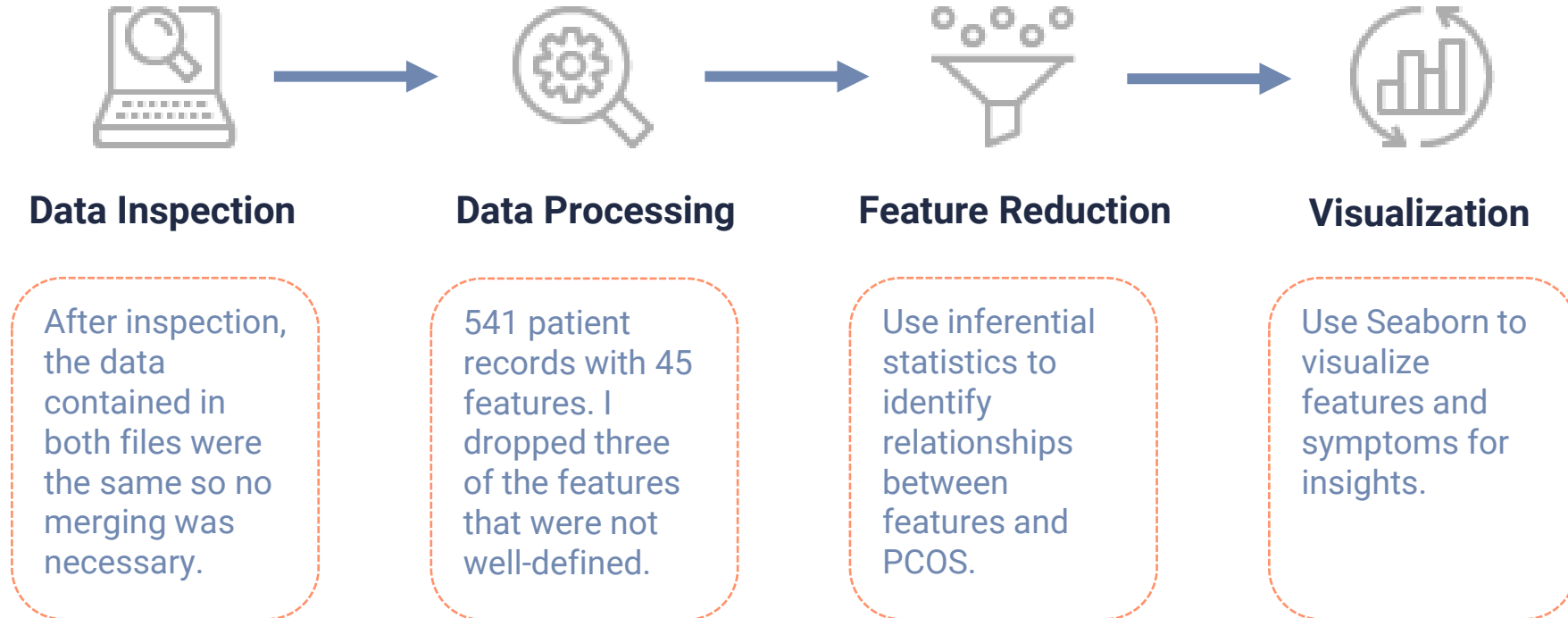**Kaggle PCOS Data** was collected from 10 different hospitals across Kerala, India.

Two different files containing **infertility & fertility patient data** in a Comma Separated Value (CSV) file and an Excel file.

45 features included in the fertility dataset which includes various hormone levels from blood tests and symptoms of the patients.

QUESTIONS OF INTEREST

1 Because symptoms vary, are there any features that are **correlated with PCOS**?

2 What are the **most frequent symptoms** PCOS patients exhibit?

3 Do **non-PCOS patients** exhibit **similar symptoms** to those diagnosed with PCOS?

EXPLORATORY ANALYSIS PROCESS

Data Inspection

Data Processing

Feature Reduction

Visualization

After inspection, the data contained in both files were the same so no merging was necessary.

541 patient records with 45 features. I dropped three of the features that were not well-defined.

Use inferential statistics to identify relationships between features and PCOS.

Use Seaborn to visualize features and symptoms for insights.
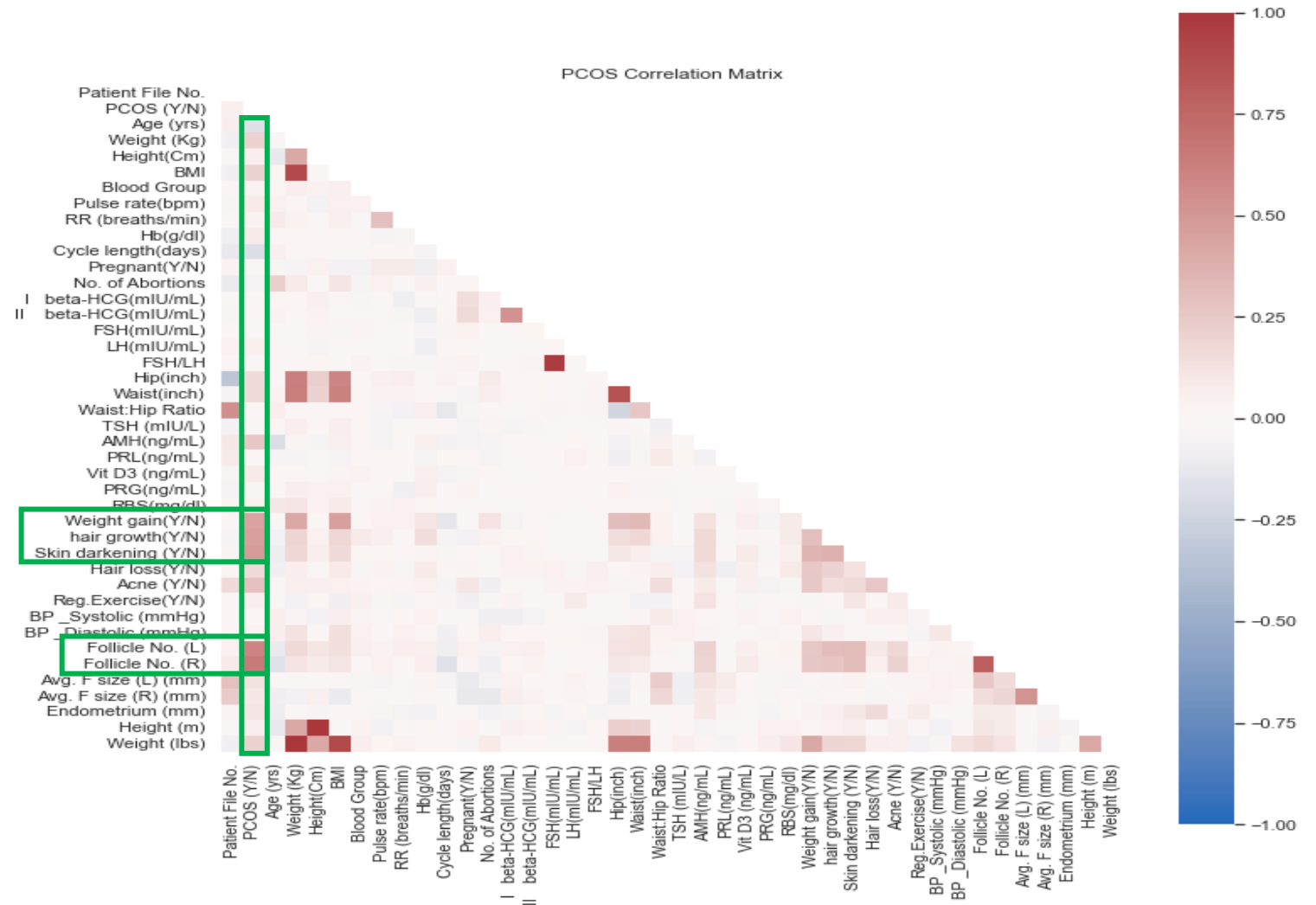
Question 1

Because symptoms vary, are there any features that are **correlated with PCOS**?
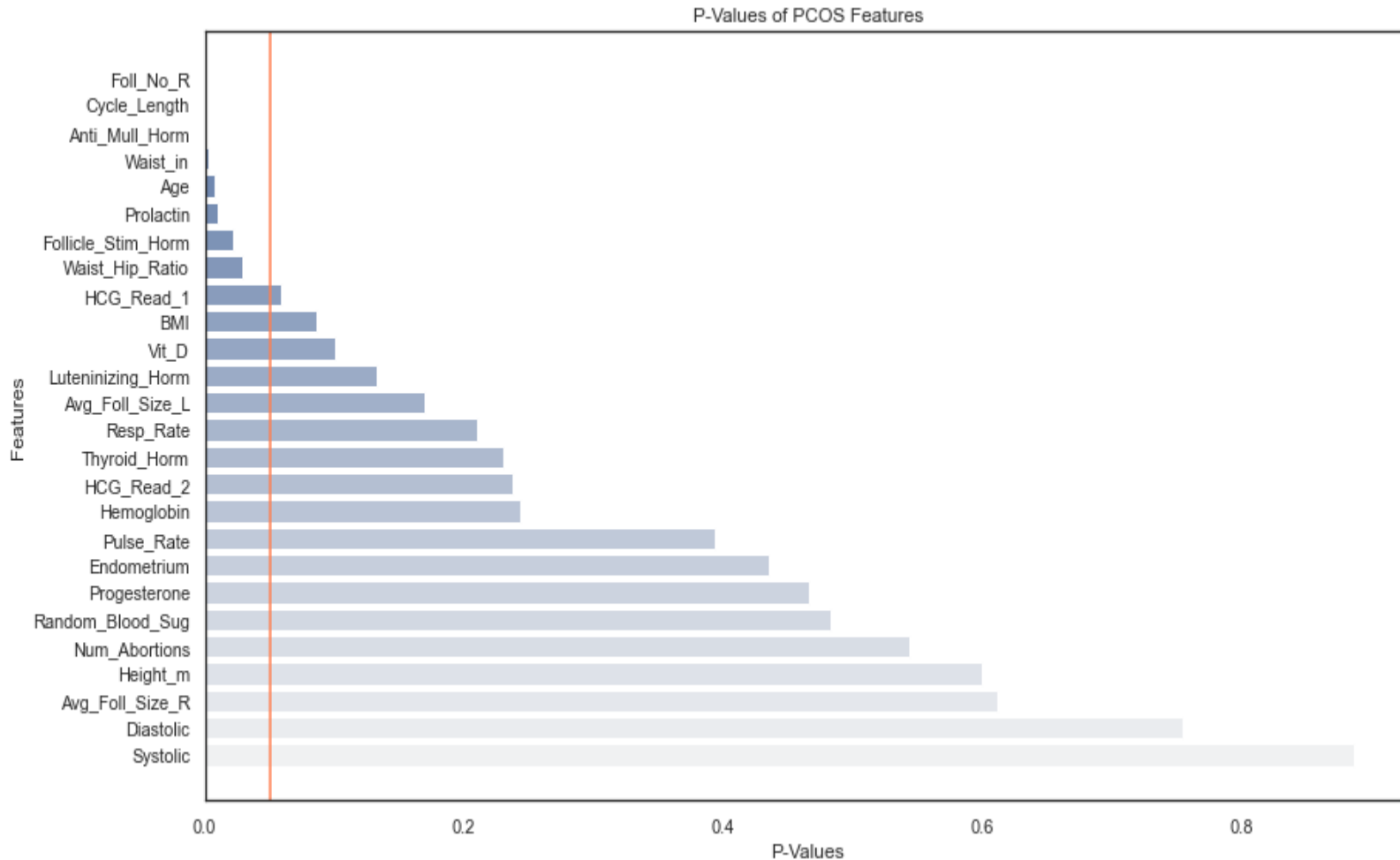
FEATURE REDUCTION

After data cleaning and pre-processing, I needed to analyze which features were correlated with PCOS.

Follicle counts, weight gain, hair growth and skin darkening are correlated with PCOS.

Other features were highly correlated with each other, due to either being the same measurement in different units or being used in the same formula to calculate, like BMI and Weight for instance.

PCOS Correlation Matrix

HYPOTHESIS TESTING

P-Values of PCOS Features

$$H_0: \mu_{PCOS_\beta} = \mu_{non-PCOS_\beta}$$

$$H_a: \mu_{PCOS_\beta} \neq \mu_{non-PCOS_\beta}$$

Where β is equal to the each independent variable or feature in the dataset and $\alpha = 0.05$.

If p-value > 0.05, retain the null.
If p-value < 0.05, reject null in favor of alternative.

The features that were below the significance level were all features that might be correlated with PCOS:
- Follicle Count
- Cycle Length
- Anti-Mullerian Hormone
- Waist Measurement
- Prolactin
- Follicle Stimulation Hormone
- Waist Hip Ratio

Question 2

Do **non-PCOS patients** exhibit **similar symptoms** to those diagnosed with PCOS?

Minimum Values

Symptom	PCOS (Y/N) 0.000000			PCOS (Y/N) 1.000000	
Exhibits_Symptom	False	True	False	False	True
Acne (Y/N)	0.410351	0.262477	0.099815		0.227357
Hair loss(Y/N)	0.408503	0.264325	0.138632		0.188540
Pregnant(Y/N)	0.410351	0.262477	0.208872		0.118299
Reg.Exercise(Y/N)	0.519409	0.153420	0.232902		0.094270
Skin darkening (Y/N)	0.569316	0.103512	0.123845		0.203327
Weight gain(Y/N)	0.519409	0.153420	0.103512		0.223660
hair growth(Y/N)	0.585952	0.086876	0.140481		0.186691

The most frequent symptoms that **PCOS** patients exhibit:
- Acne
- Weight Gain
- Skin Darkening

Question 2

What are the **most frequent symptoms** PCOS patients exhibit?

CATEGORICAL AGGREGATION

Maximum Values

Symptom	PCOS (Y/N) 0.000000		1.000000	
Exhibits_Symptom	False	True	False	True
Acne (Y/N)	0.410351	0.262477	0.099815	0.227357
Hair loss(Y/N)	0.408503	0.264325	0.138632	0.188540
Pregnant(Y/N)	0.410351	0.262477	0.208872	0.118299
Reg.Exercise(Y/N)	0.519409	0.153420	0.232902	0.094270
Skin darkening (Y/N)	0.569316	0.103512	0.123845	0.203327
Weight gain(Y/N)	0.519409	0.153420	0.103512	0.223660
hair growth(Y/N)	0.585952	0.086876	0.140481	0.186691

The most frequent symptoms that **non-PCOS** patients exhibit:
- Acne
- Hair Loss

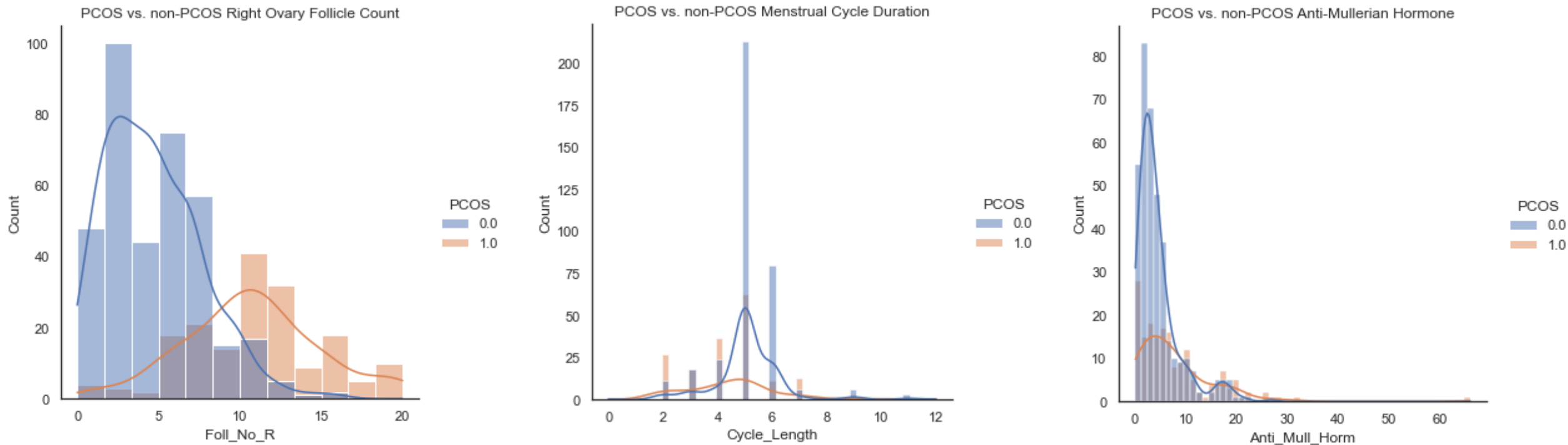Most **non-PCOS** patients do not exhibit any symptoms in this sample.

CONCLUSIONS

1 **Follicle Count**, **Cycle Length**, **AMH**, **PRL**, **FSH**, **Age**, and **Waist Measurement** are highest possible indicators of PCOS.

2 The most frequent symptoms of **PCOS patients** are **Acne**, **Weight Gain** and **Skin Darkening**.

3 The most frequent symptoms exhibited by **non-PCOS patients** were **Acne** and **Skin Darkening**.

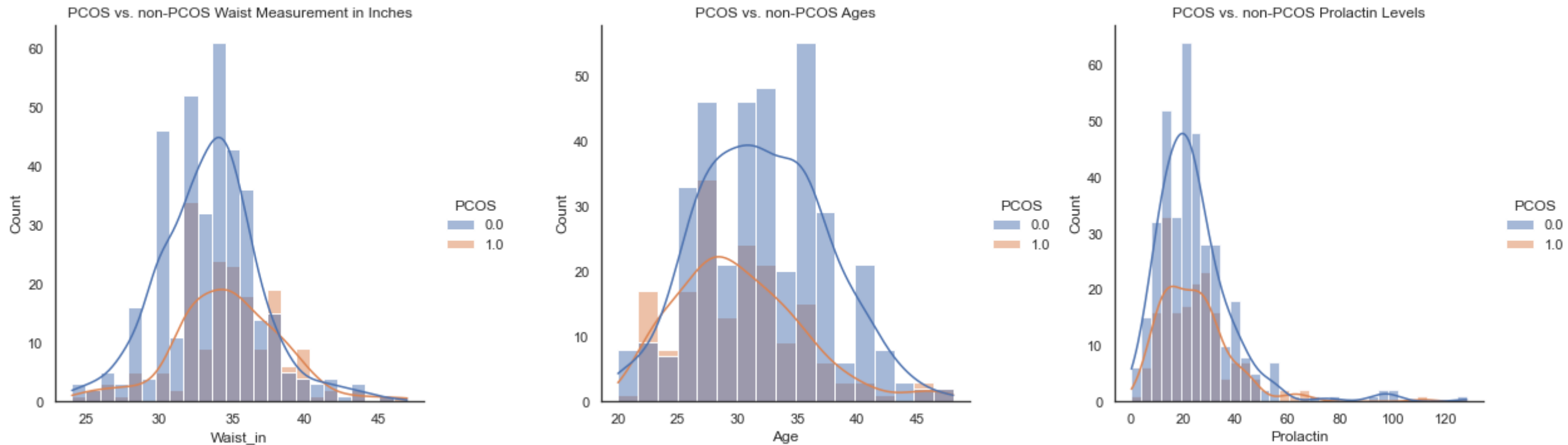# APPENDIX

VISUALIZING DISTRIBUTIONS OF FEATURES

Detailed observations are provided in PCOS_Analysis Jupyter Notebook.

VISUALIZING DISTRIBUTIONS OF FEATURES

Detailed observations are provided in PCOS_Analysis Jupyter Notebook.