Summary Report of PCOD Analysis: Exploratory Data Analysis and Visualizations

Pradeepa Sivakumar

Karpagam Institute of Technology Coimbatore, Tamil Nadu February 2025

Contact Information: Email: pradeepasivakumar23@gmail.com

LinkedIn: linkedin.com/in/pradeepa-s

GitHub: github.com/Pradee23

Introduction:

This report presents the results of an exploratory data analysis (EDA) conducted on the Polycystic Ovary Syndrome (PCOD) dataset. The primary objective of this analysis is to identify patterns and insights related to PCOD through data cleaning, summary statistics, and visualizations. The analysis was conducted using R, leveraging various libraries for data manipulation and visualization.

Data Cleaning:

The initial dataset contained missing values and required preprocessing for accurate analysis. The following steps were taken:

- Replaced missing values in the New$age and Weight$kg columns with their respective means.

- Rounded the values in the New$age and Weight$kg columns for consistency.

- Arranged the data by age for better analysis.

Summary Statistics:

Summary statistics were calculated for the New$age and Weight$kg columns to understand the distribution of these variables.

- **New_age:**

  o Mean: 22.35714

  o Median: 23

  o Minimum: 18

  o Maximum: 45

  o Quantiles: 0% 25% 50% 75% 100%

  18   19   23   23   45

- **Weight_kg:**

  o Mean: 56.20952

  o Median: 55

  o Minimum: 20

  o Maximum: 116

  o Quantiles: 0% 25% 50% 75% 100%

  20   48   55   64   116

Visualizations:

Various visualizations were created to explore the relationships between different variables in the dataset:

1. **Age Distribution by PCOS Status:**

   o   A stacked bar chart that shows the age distribution among individuals with and without PCOS.

2. **Exercise Habits Analysis:**

   o   A heatmap depicting the relationship between exercise frequency, exercise type, and PCOS count, highlighting lifestyle factors associated with PCOS.

3. **Sleep Hours vs. Insulin Resistance:**

   o   A beeswarm plot showing the correlation between sleep hours and insulin resistance status.

4. **Sleep Hours vs. Hormonal Imbalance:**

   o   A bar chart illustrating the relationship between sleep hours and hormonal imbalance, indicating the possible impact of sleep patterns on hormonal health.

5. **Weight Distribution by Exercise Type:**

   o   A boxplot visualizing the weight distribution by different exercise types, providing insights into how physical activity influences weight among individuals with PCOS.

6. **Hormonal Imbalance Across Age Groups:**

   o   A line chart showing hormonal imbalance across different age groups, helping to understand how hormonal health varies with age.

7. **Conception Difficulty Across PCOS Status:**

   o   A bar chart representing the count of individuals with conception difficulties across PCOS status, offering insights into fertility challenges associated with PCOS.

Conclusions:

The EDA and visualizations provide valuable insights into the patterns and relationships between various factors and PCOS. These findings can help healthcare providers and researchers understand and address PCOS more effectively. The analysis highlights the importance of lifestyle factors such as exercise and sleep patterns in managing PCOS and its associated health indicators.

Future Work:

- Further refine visualizations to uncover deeper insights.

- Explore additional features that may contribute to PCOS diagnosis and treatment.