

# Polycystic Ovary Syndrome (PCOS)

## 1. Problem Statement

The goal of this analysis is to use the Polycystic Ovary Syndrome (PCOS) dataset from Kaggle (<https://www.kaggle.com/datasets/hasaanrana/diet-exercise-and-pcos-insights>) to determine if there are any factors that could be linked to having the disease.

## 2. Data Loading and Exploration

```
library('dplyr')
library('ggplot2')
library('tidyr')
library('rmarkdown')
```

```
pcos_data_raw <- read.csv('pcos_data.csv')
```

```
dim(pcos_data_raw)
```

```
## [1] 173 36
```

The dataset contains 173 rows and 36 columns.

```
head(pcos_data_raw)
```

```
##           Age Weight_kg Height_ft Marital_Status PCOS Family_History_PCOS
## 1         20-25      66   157.48   Unmarried    No                No
## 2 Less than 20      56   165.10   Unmarried    No                No
## 3 Less than 20      89   167.64   Unmarried    No                Yes
## 4         20-25      55   160.02   Unmarried    No                Yes
## 5 Less than 20      55   160.02   Unmarried    No                No
## 6 Less than 20      45   160.02   Unmarried    No                No
##  Menstrual_Irregularity Hormonal_Imbalance Hyperandrogenism Hirsutism
## 1                      Yes                No                No        No
## 2                      No                No                No        No
## 3                      No                No                No        Yes
## 4                      No                Yes                No        Yes
## 5                      No                No                No        No
## 6                      Yes                Yes                No        No
##  Mental_Health      Conception_Difficulty Insulin_Resistance Diabetes
## 1              Yes                No                No        No
## 2              No                No                No        No
```

## 3	Yes	No	No	No
## 4	Yes	Yes, not diagnosed by a doctor	No	No
## 5	Yes	No	No	No
## 6	Yes	No	No	No
##	Childhood_Trauma	Cardiovascular_Disease	Diet_Bread_Cereals	Diet_Milk_Products
## 1	No	No	7	7
## 2	No	No	4	4
## 3	Yes	No	6	2
## 4	Yes	No	2	3
## 5	Yes	No	4	7
## 6	Yes	No	7	3
##	Diet_Fruits	Diet_Vegetables	Diet_Starchy_Vegetables	
## 1	2	2	3	
## 2	4	4	4	
## 3	1	3	2	
## 4	1	2	5	
## 5	2	3	4	
## 6	2	4	3	
##	Diet_NonStarchy_Vegetables	Diet_Fats	Diet_Sweets	Diet_Fried_Food
## 1		1	7	3
## 2		2	4	0
## 3		1	7	1
## 4		3	5	5
## 5		4	6	2
## 6		5	4	1
##	Diet_Tea_Coffee	Diet_Multivitamin	Vegetarian	Exercise_Frequency
## 1	7	0	No	Rarely
## 2	0	0	No	Daily
## 3	7	0	No	Rarely
## 4	7	1	No	Never
## 5	5	0	No	Daily
## 6	7	0	No	Rarely
##		Exercise_Type	Exercise_Duration	
## 1	Cardio (e.g., running, cycling, swimming)		30 minutes	
## 2		No Exercise	Less than 30 minutes	
## 3	Cardio (e.g., running, cycling, swimming)		Less than 30 minutes	
## 4		No Exercise	Not Applicable	
## 5	Cardio (e.g., running, cycling, swimming)		30 minutes to 1 hour	
## 6		No Exercise	Not Applicable	
##	Sleep_Hours	Stress_Level	Smoking	Exercise_Benefit
## 1	Less than 6 hours	No	No	Somewhat
## 2	6-8 hours	No	No	Somewhat
## 3	6-8 hours	Yes	No	Somewhat
## 4	6-8 hours	Yes	No	Somewhat
## 5	6-8 hours	Yes	No	Not at All
## 6	6-8 hours	Yes	No	Not Much
				PCOS_Medication
## 1				No.
## 2				No.
## 3				No.
## 4				No.
## 5				No.
## 6				No.

```
colnames(pcos_data_raw)
```

```
## [1] "Age" "Weight_kg"
## [3] "Height_ft" "Marital_Status"
## [5] "PCOS" "Family_History_PCOS"
## [7] "Menstrual_Irregularity" "Hormonal_Imbalance"
## [9] "Hyperandrogenism" "Hirsutism"
```

```
## [11] "Mental_Health"          "Conception_Difficulty"
## [13] "Insulin_Resistance"     "Diabetes"
## [15] "Childhood_Trauma"       "Cardiovascular_Disease"
## [17] "Diet_Bread_Cereals"     "Diet_Milk_Products"
## [19] "Diet_Fruits"            "Diet_Vegetables"
## [21] "Diet_Starchy_Vegetables" "Diet_NonStarchy_Vegetables"
## [23] "Diet_Fats"              "Diet_Sweets"
## [25] "Diet_Fried_Food"        "Diet_Tea_Coffee"
## [27] "Diet_Multivitamin"      "Vegetarian"
## [29] "Exercise_Frequency"     "Exercise_Type"
## [31] "Exercise_Duration"      "Sleep_Hours"
## [33] "Stress_Level"           "Smoking"
## [35] "Exercise_Benefit"       "PCOS_Medication"
```

```
summary(pcos_data_raw)
```

```
##      Age      Weight_kg      Height_ft      Marital_Status
## Length:173      Min.   : 36.00      Min.   :124.5      Length:173
## Class :character 1st Qu.: 49.00      1st Qu.:157.5      Class :character
## Mode  :character Median : 55.00      Median :160.0      Mode  :character
##              Mean  : 56.55      Mean  :160.6
##              3rd Qu.: 64.00      3rd Qu.:165.1
##              Max.   :115.00      Max.   :182.9
##      PCOS      Family_History_PCOS      Menstrual_Irregularity
## Length:173      Length:173      Length:173
## Class :character Class :character      Class :character
## Mode  :character Mode  :character      Mode  :character
##
##
##
##      Hormonal_Imbalance      Hyperandrogenism      Hirsutism      Mental_Health
## Length:173      Length:173      Length:173      Length:173
## Class :character      Class :character      Class :character      Class :character
## Mode  :character      Mode  :character      Mode  :character      Mode  :character
##
##
##
##      Conception_Difficulty      Insulin_Resistance      Diabetes      Childhood_Trauma
## Length:173      Length:173      Length:173      Length:173
## Class :character      Class :character      Class :character      Class :character
## Mode  :character      Mode  :character      Mode  :character      Mode  :character
##
##
##
##      Cardiovascular_Disease      Diet_Bread_Cereals      Diet_Milk_Products      Diet_Fruits
## Length:173      Min.   :0.000      Min.   :0.000      Min.   :0.000
## Class :character      1st Qu.:4.000      1st Qu.:1.000      1st Qu.:1.000
## Mode  :character      Median :7.000      Median :3.000      Median :3.000
##              Mean  :5.445      Mean  :3.647      Mean  :3.035
##              3rd Qu.:7.000      3rd Qu.:7.000      3rd Qu.:4.000
##              Max.   :7.000      Max.   :7.000      Max.   :7.000
##      Diet_Vegetables      Diet_Starchy_Vegetables      Diet_NonStarchy_Vegetables
## Min.   :0.000      Min.   :0.000      Min.   :0.000
## 1st Qu.:2.000      1st Qu.:2.000      1st Qu.:1.000
```

```

## Median :3.000 Median :3.000 Median :2.000
## Mean :3.439 Mean :3.069 Mean :2.451
## 3rd Qu.:5.000 3rd Qu.:4.000 3rd Qu.:3.000
## Max. :7.000 Max. :7.000 Max. :7.000
## Diet_Fats Diet_Sweets Diet_Fried_Food Diet_Tea_Coffee
## Min. :0.000 Min. :0.000 Min. :0.000 Min. :0.000
## 1st Qu.:2.000 1st Qu.:2.000 1st Qu.:1.000 1st Qu.:2.000
## Median :5.000 Median :3.000 Median :3.000 Median :6.000
## Mean :4.532 Mean :3.618 Mean :3.179 Mean :4.549
## 3rd Qu.:7.000 3rd Qu.:6.000 3rd Qu.:5.000 3rd Qu.:7.000
## Max. :7.000 Max. :7.000 Max. :7.000 Max. :7.000
## Diet_Multivitamin Vegetarian Exercise_Frequency Exercise_Type
## Min. :0.000 Length:173 Length:173 Length:173
## 1st Qu.:0.000 Class :character Class :character Class :character
## Median :0.000 Mode :character Mode :character Mode :character
## Mean :1.393
## 3rd Qu.:2.000
## Max. :7.000
## Exercise_Duration Sleep_Hours Stress_Level Smoking
## Length:173 Length:173 Length:173 Length:173
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
## Exercise_Benefit PCOS_Medication
## Length:173 Length:173
## Class :character Class :character
## Mode :character Mode :character
##
##
##

```

```
count(pcos_data_raw, PCOS, sort = TRUE)
```

```

##                PCOS    n
## 1                No  134
## 2                Yes   38
## 3 No, Yes, not diagnosed by a doctor    1

```

The dataset contains 134 women without PCOS, 38 with PCOS, and one entry with both values.

### **3. Data Cleaning and Preparation**

#### **3.1 Dealing with Missing and Duplicated Values**

#### **3.2 Adding and Modifying Columns**

### **4. Data Visualization in Tableau**

### **5. Analysis**

#### **5.1 General Questions**

#### **5.2 Inferential Statistics**

### **6. Conclusion**