

# 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.

For more information about PCOS, here is a link to the Mayo Clinic (<https://www.mayoclinic.org/diseases-conditions/pcos/symptoms-causes/syc-20353439>).

## 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	3	
## 2	2	4	0	1	
## 3	1	7	1	3	
## 4	3	5	5	5	
## 5	4	6	2	3	
## 6	5	4	1	2	
##	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	PCOS_Medication
## 1	Less than 6 hours	No	No	Somewhat	No.
## 2	6-8 hours	No	No	Somewhat	No.
## 3	6-8 hours	Yes	No	Somewhat	No.
## 4	6-8 hours	Yes	No	Somewhat	No.
## 5	6-8 hours	Yes	No	Not at All	No.
## 6	6-8 hours	Yes	No	Not Much	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

```
sum(is.na(pcos_data_raw))
```

```
## [1] 0
```

The dataset does not contain any missing values.

```
sum(duplicated(pcos_data_raw))
```

```
## [1] 0
```

The dataset does not contain any duplicated rows.

## 3.2 Adding and Modifying Columns

```
summary(pcos_data_raw$Height_ft)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 124.5   157.5   160.0   160.6   165.1   182.9
```

The column above is called “Height\_ft”, but the values range from 124 to 182. That range is not possible for a human. According to the dataset’s author, this column is actually in cm. The column name will be changed to reflect that.

```
pcos_data_clean <- pcos_data_raw
```

```
pcos_data_clean <- rename(pcos_data_clean, Height_cm = Height_ft)
```

```
head(pcos_data_clean)
```

```
##           Age Weight_kg Height_cm 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	PCOS_Medication
## 1	Less than 6 hours	No	No	Somewhat	No.
## 2	6-8 hours	No	No	Somewhat	No.
## 3	6-8 hours	Yes	No	Somewhat	No.
## 4	6-8 hours	Yes	No	Somewhat	No.
## 5	6-8 hours	Yes	No	Not at All	No.
## 6	6-8 hours	Yes	No	Not Much	No.

Since this study will be used within the USA, new columns will be added to convert the kg and cm columns to lbs and in respectively.

```
pcos_data_clean <- mutate(pcos_data_clean, Height_in = Height_cm / 2.54, Weight_lb = Weight_kg * 2.204)
# 2.54 cm in an inch
# 2.204 kg in a lb
```

```
head(select(pcos_data_clean, Weight_kg, Weight_lb, Height_cm, Height_in))
```

```
## Weight_kg Weight_lb Height_cm Height_in
## 1      66   145.464   157.48      62
```

```
## 2      56    123.424    165.10      65
## 3      89    196.156    167.64      66
## 4      55    121.220    160.02      63
## 5      55    121.220    160.02      63
## 6      45     99.180    160.02      63
```

The Weight\_kg and Height\_cm columns will now be dropped from the dataset.

```
pcos_data_clean <- select(pcos_data_clean, -c(Weight_kg, Height_cm))
```

```
head(pcos_data_clean)
```

```
##      Age Marital_Status PCOS Family_History_PCOS Menstrual_Irregularity
## 1      20-25      Unmarried      No      No      Yes
## 2 Less than 20      Unmarried      No      No      No
## 3 Less than 20      Unmarried      No      Yes      No
## 4      20-25      Unmarried      No      Yes      No
## 5 Less than 20      Unmarried      No      No      No
## 6 Less than 20      Unmarried      No      No      Yes
##      Hormonal_Imbalance Hyperandrogenism Hirsutism Mental_Health
## 1      No      No      No      Yes
## 2      No      No      No      No
## 3      No      No      Yes      Yes
## 4      Yes      No      Yes      Yes
## 5      No      No      No      Yes
## 6      Yes      No      No      Yes
##      Conception_Difficulty Insulin_Resistance Diabetes Childhood_Trauma
## 1      No      No      No      No
## 2      No      No      No      No
## 3      No      No      No      Yes
## 4 Yes, not diagnosed by a doctor      No      No      Yes
## 5      No      No      No      Yes
## 6      No      No      No      Yes
##      Cardiovascular_Disease Diet_Bread_Cereals Diet_Milk_Products Diet_Fruits
## 1      No      7      7      2
## 2      No      4      4      4
## 3      No      6      2      1
## 4      No      2      3      1
## 5      No      4      7      2
## 6      No      7      3      2
##      Diet_Vegetables Diet_Starchy_Vegetables Diet_NonStarchy_Vegetables Diet_Fats
## 1      2      3      1      7
## 2      4      4      2      4
## 3      3      2      1      7
## 4      2      5      3      5
## 5      3      4      4      6
## 6      4      3      5      4
##      Diet_Sweets Diet_Fried_Food Diet_Tea_Coffee Diet_Multivitamin Vegetarian
## 1      3      3      7      0      No
## 2      0      1      0      0      No
## 3      1      3      7      0      No
## 4      5      5      7      1      No
## 5      2      3      5      0      No
```

```
## 6          1          2          7          0          No
## Exercise_Frequency          Exercise_Type
## 1          Rarely Cardio (e.g., running, cycling, swimming)
## 2          Daily          No Exercise
## 3          Rarely Cardio (e.g., running, cycling, swimming)
## 4          Never          No Exercise
## 5          Daily Cardio (e.g., running, cycling, swimming)
## 6          Rarely          No Exercise
## Exercise_Duration Sleep_Hours Stress_Level Smoking Exercise_Benefit
## 1          30 minutes Less than 6 hours          No          No          Somewhat
## 2 Less than 30 minutes          6-8 hours          No          No          Somewhat
## 3 Less than 30 minutes          6-8 hours          Yes          No          Somewhat
## 4          Not Applicable          6-8 hours          Yes          No          Somewhat
## 5 30 minutes to 1 hour          6-8 hours          Yes          No          Not at All
## 6          Not Applicable          6-8 hours          Yes          No          Not Much
## PCOS_Medication Height_in Weight_lb
## 1          No.          62 145.464
## 2          No.          65 123.424
## 3          No.          66 196.156
## 4          No.          63 121.220
## 5          No.          63 121.220
## 6          No.          63 99.180
```

A common measure for if someone is at a healthy weight for their height is body mass index (BMI). The formula for this using English units is:

$$BMI = 703 * \frac{weight(lb)}{height^2(in)}$$

This column will be added to the dataset.

```
pcos_data_clean <- mutate(pcos_data_clean, BMI = 703 * (Weight_lb / (Height_in ^ 2)))
```

```
head(select(pcos_data_clean, Weight_lb, Height_in, BMI))
```

```
## Weight_lb Height_in BMI
## 1 145.464 62 26.60281
## 2 123.424 65 20.53659
## 3 196.156 66 31.65695
## 4 121.220 63 21.47081
## 5 121.220 63 21.47081
## 6 99.180 63 17.56703
```

Since the new columns we added to the end of the dataframe, they will be rearranged to the second through fourth columns.

```
pcos_data_clean <- select(pcos_data_clean, c(1, 35, 36, 37, 2:34))
```

```
head(pcos_data_clean)
```

```
##          Age Height_in Weight_lb          BMI Marital_Status PCOS
```



## 1	20-25	62	145.464	26.60281	Unmarried	No
## 2	Less than 20	65	123.424	20.53659	Unmarried	No
## 3	Less than 20	66	196.156	31.65695	Unmarried	No
## 4	20-25	63	121.220	21.47081	Unmarried	No
## 5	Less than 20	63	121.220	21.47081	Unmarried	No
## 6	Less than 20	63	99.180	17.56703	Unmarried	No
##	Family_History_PCOS	Menstrual_Irregularity	Hormonal_Imbalance			
## 1	No		Yes	No		
## 2	No		No	No		
## 3	Yes		No	No		
## 4	Yes		No	Yes		
## 5	No		No	No		
## 6	No		Yes	Yes		
##	Hyperandrogenism	Hirsutism	Mental_Health	Conception_Difficulty		
## 1	No	No	Yes	No		
## 2	No	No	No	No		
## 3	No	Yes	Yes	No		
## 4	No	Yes	Yes	Yes, not diagnosed by a doctor		
## 5	No	No	Yes	No		
## 6	No	No	Yes	No		
##	Insulin_Resistance	Diabetes	Childhood_Trauma	Cardiovascular_Disease		
## 1	No	No	No	No		
## 2	No	No	No	No		
## 3	No	No	Yes	No		
## 4	No	No	Yes	No		
## 5	No	No	Yes	No		
## 6	No	No	Yes	No		
##	Diet_Bread_Cereals	Diet_Milk_Products	Diet_Fruits	Diet_Vegetables		
## 1	7		7	2	2	
## 2	4		4	4	4	
## 3	6		2	1	3	
## 4	2		3	1	2	
## 5	4		7	2	3	
## 6	7		3	2	4	
##	Diet_Starchy_Vegetables	Diet_NonStarchy_Vegetables	Diet_Fats	Diet_Sweets		
## 1	3		1	7	3	
## 2	4		2	4	0	
## 3	2		1	7	1	
## 4	5		3	5	5	
## 5	4		4	6	2	
## 6	3		5	4	1	
##	Diet_Fried_Food	Diet_Tea_Coffee	Diet_Multivitamin	Vegetarian		
## 1	3	7	0	No		
## 2	1	0	0	No		
## 3	3	7	0	No		
## 4	5	7	1	No		
## 5	3	5	0	No		
## 6	2	7	0	No		
##	Exercise_Frequency	Exercise_Type				
## 1	Rarely	Cardio (e.g., running, cycling, swimming)				
## 2	Daily	No Exercise				
## 3	Rarely	Cardio (e.g., running, cycling, swimming)				
## 4	Never	No Exercise				
## 5	Daily	Cardio (e.g., running, cycling, swimming)				

```
## 6          Rarely          No Exercise
##      Exercise_Duration      Sleep_Hours Stress_Level Smoking Exercise_Benefit
## 1          30 minutes Less than 6 hours          No          No          Somewhat
## 2 Less than 30 minutes          6-8 hours          No          No          Somewhat
## 3 Less than 30 minutes          6-8 hours          Yes          No          Somewhat
## 4          Not Applicable          6-8 hours          Yes          No          Somewhat
## 5 30 minutes to 1 hour          6-8 hours          Yes          No          Not at All
## 6          Not Applicable          6-8 hours          Yes          No          Not Much
##      PCOS_Medication
## 1          No.
## 2          No.
## 3          No.
## 4          No.
## 5          No.
## 6          No.
```

The Marital\_Status column will be checked to make sure the selections are valid.

```
count(pcos_data_clean, Marital_Status)
```

```
##      Marital_Status      n
## 1          Married      16
## 2          Unmarried    156
## 3          Widow       1
```

According to the CDC (<https://www.cdc.gov/bmi/adult-calculator/bmi-categories.html>), there are ranges within BMI that can be used to categorize a person's health as follows:

- Less than 18.5: Underweight
- 18.5 - 24.9: Healthy Weight
- 25 - 29.9: Overweight
- 30 - 34.5: Class 1 Obesity
- 35 - 39.9: Class 2 Obesity
- 40+: Class 3 Obesity

A new column with this category will be added to make it easier to compare people based on BMI.

```
BMI_groups <- cut(pcos_data_clean$BMI, breaks = c(0, 18.4, 24.9, 29.9, 34.5, 39.9, 100),
  labels = c("Underweight",
    "Healthy Weight",
    "Overweight",
    "Class 1 Obesity",
    "Class 2 Obesity",
    "Class 3 Obesity"))
```

```
print(table(BMI_groups))
```

```
## BMI_groups
##      Underweight Healthy Weight      Overweight Class 1 Obesity Class 2 Obesity
##              41              92              30              8              1
## Class 3 Obesity
##              1
```

```
count(pcos_data_clean, BMI < 18.4)
```

```
## BMI < 18.4 n
## 1 FALSE 132
## 2 TRUE 41
```

```
pcos_data_clean$BMI_Cat <- BMI_groups
```

```
head(select(pcos_data_clean, BMI_Cat, BMI))
```

```
## BMI_Cat BMI
## 1 Overweight 26.60281
## 2 Healthy Weight 20.53659
## 3 Class 1 Obesity 31.65695
## 4 Healthy Weight 21.47081
## 5 Healthy Weight 21.47081
## 6 Underweight 17.56703
```

```
pcos_data_clean <- select(pcos_data_clean, 1:4, 38, 5:37)
```

```
head(pcos_data_clean)
```

```
## Age Height_in Weight_lb BMI BMI_Cat Marital_Status PCOS
## 1 20-25 62 145.464 26.60281 Overweight Unmarried No
## 2 Less than 20 65 123.424 20.53659 Healthy Weight Unmarried No
## 3 Less than 20 66 196.156 31.65695 Class 1 Obesity Unmarried No
## 4 20-25 63 121.220 21.47081 Healthy Weight Unmarried No
## 5 Less than 20 63 121.220 21.47081 Healthy Weight Unmarried No
## 6 Less than 20 63 99.180 17.56703 Underweight Unmarried No
## Family_History_PCOS Menstrual_Irregularity Hormonal_Imbalance
## 1 No Yes No
## 2 No No No
## 3 Yes No No
## 4 Yes No Yes
## 5 No No No
## 6 No Yes Yes
## Hyperandrogenism Hirsutism Mental_Health Conception_Difficulty
## 1 No No Yes No
## 2 No No No No
## 3 No Yes Yes No
## 4 No Yes Yes Yes, not diagnosed by a doctor
## 5 No No Yes No
## 6 No No Yes No
## Insulin_Resistance Diabetes Childhood_Trauma Cardiovascular_Disease
## 1 No No No No
## 2 No No No No
## 3 No No Yes No
## 4 No No Yes No
## 5 No No Yes No
## 6 No No Yes No
## Diet_Bread_Cereals Diet_Milk_Products Diet_Fruits Diet_Vegetables
```

```

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

```

```
write.csv(pcos_data_clean,"pcos_data_clean.csv", row.names = FALSE)
```

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

## **5. Analysis**

### **5.1 General Questions**

### **5.2 Inferential Statistics**

## **6. Conclusion**