**Exp No:5 Data Collection and Initial Exploration**

**NAME:ENIYA.B.A**

**ROLL NO:230701085**

**Objective:** To collect, load, and perform initial exploration of the diabetes dataset.

import pandas as pd

db=pd.read\_csv("diabetes.csv")

print(db.head())

**OUTPUT:**

Pregnancies Glucose BloodPressure SkinThickness Insulin BMI \

0 6 148 72 35 0 33.6

1 1 85 66 29 0 26.6

2 8 183 64 0 0 23.3

3 1 89 66 23 94 28.1

4 0 137 40 35 168 43.1

DiabetesPedigreeFunction Age Outcome

0 0.627 50 1

1 0.351 31 0

2 0.672 32 1

3 0.167 21 0

4 2.288 33 1

print(db.info())

print(db.describe())

**OUTPUT:**

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 9 entries, 0 to 8

Data columns (total 9 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Pregnancies 9 non-null int64

1 Glucose 9 non-null int64

2 BloodPressure 9 non-null int64

3 SkinThickness 9 non-null int64

4 Insulin 9 non-null int64

5 BMI 9 non-null float64

6 DiabetesPedigreeFunction 9 non-null float64

7 Age 9 non-null int64

8 Outcome 9 non-null int64

dtypes: float64(2), int64(7)

memory usage: 776.0 bytes

None

Pregnancies Glucose BloodPressure SkinThickness Insulin \

count 9.000000 9.000000 9.000000 9.000000 9.000000

mean 4.000000 127.555556 55.777778 22.111111 99.222222

std 3.464102 42.497386 23.631429 17.567331 177.146675

min 0.000000 78.000000 0.000000 0.000000 0.000000

25% 1.000000 89.000000 50.000000 0.000000 0.000000

50% 3.000000 116.000000 66.000000 29.000000 0.000000

75% 6.000000 148.000000 70.000000 35.000000 94.000000

max 10.000000 197.000000 74.000000 45.000000 543.000000

BMI DiabetesPedigreeFunction Age Outcome

count 9.000000 9.000000 9.000000 9.000000

mean 30.788889 0.538444 33.888889 0.555556

std 5.997175 0.686137 10.635371 0.527046

min 23.300000 0.134000 21.000000 0.000000

25% 26.600000 0.167000 29.000000 0.000000

50% 30.500000 0.248000 31.000000 1.000000

75% 33.600000 0.627000 33.000000 1.000000

max 43.100000 2.288000 53.000000 1.000000

import matplotlib.pyplot as plt

import seaborn as sns

db.hist(bins=50,figsize=(20,15))

plt.show()

sns.pairplot(db)

plt.show()

**OUTPUT:**



