Exercise 2- Report

Group-9
Amber Sethi
Joshua Dominic Dmello
Mohamed Bennis
Muhammad Attique
Tanvi Sharma

Identifying Key variables and metrics.
Presenting descriptive statistics on each of them.

Table of Contents:

Variables:

Age....2

BMI....3

Blood Pressure....4

Diabetes Pedigree Function....5

Age

Age is one of the factors that can be used to determine whether a person has diabetic condition or not. For instance, approximately 25% of adults over the age of 60 years have diabetes.

Below are the statistics defined for the Age variable.

	Age
Mean	33.24088542
Std	11.76023154
Median	29
Count	768
Total	25529

-Using Excel

Quantile statistics

Minimum	21
5-th percentile	21
Q1	24
median	29
Q3	41
95-th percentile	58
Maximum	81
Range	60
Interquartile range (IQR)	17

Descriptive statistics

Standard deviation	11.76023154
Coefficient of variation (CV)	0.3537881556
Kurtosis	0.6431588885
Mean	33.24088542
Median Absolute Deviation (MAD)	7
Skewness	1.129596701
Sum	25529
Variance	138.3030459
Monotonicity	Not monotonic

-Using Pandas Profiling in Python

We observe here that the median age is 29 and the average age of a person diagnosed with diabetes is 33. It can also be seen that is positively skewed. The older a human is, more is the chance of getting diagnosed with diabetes.

BMI

Another chosen metric is BMI. For instance, in general, at 18 years of age, for male subjects, the remaining lifetime diabetes risk ranged from 7.6% for those with BMI <18.5 kg/m^2 to 70.3% for those BMI >35 kg/m^2.

Below are the statistics defined for the BMI variable.

	BMI
Mean	31.99257813
Std	7.88416032
Median	32
Count	768
Total	24570.3

-Using Excel

Quantile statistics

Minimum	0
5-th percentile	21.8
Q1	27.3
median	32
Q3	36.6
95-th percentile	44.395
Maximum	67.1
Range	67.1
Interquartile range (IQR)	9.3

Descriptive statistics

Standard deviation	7.88416032
Coefficient of variation (CV)	0.2464371671
Kurtosis	3.290442901
Mean	31.99257812
Median Absolute Deviation (MAD)	4.6
Skewness	-0.4289815885
Sum	24570.3
Variance	62.15998396
Monotonicity	Not monotonic

-Using Pandas Profiling in Python

We see here that the average BMI is very close to 32 which is also the value for median and we can say that is very close to a normal distribution curve. A higher BMI can be a cause of diabetes.

Blood Pressure

Blood pressure has a direct relation with Diabetes. For instance, people with high blood pressure have a higher risk of developing type 2 diabetes.

Below are the statistics defined for the Blood Pressure variable.

	BloodPressure
Mean	69.10546875
Std	19.35580717
Median	72
Count	768
Total	53073

-Using Excel

Quantile statistics

Minimum	0
5-th percentile	38.7
Q1	62
median	72
Q3	80
95-th percentile	90
Maximum	122
Range	122
Interquartile range (IQR)	18

Descriptive statistics

Standard deviation	19.35580717
Coefficient of variation (CV)	0.2800908166
Kurtosis	5.18015656
Mean	69.10546875
Median Absolute Deviation (MAD)	8
Skewness	-1.843607983
Sum	53073
Variance	374.6472712
Monotonicity	Not monotonic

-Using Pandas Profiling in Python

A person on average has a Blood Pressure of around 69 and here the median is 72. The skewness is negative 1.84. Blood pressure relates to diabetes directly and the more the blood pressure, the more is the chance of being diagnosed with diabetes.

Diabetes Pedigree Function

The last key metric is Diabetes Pedigree Function which is a function that assesses the risk of diabetes depending on a person's family history.

Below are the statistics defined for the Diabetes Pedigree Function variable.

	DiabetesPedigreeFunction
Mean	0.471876302
Std	0.331328595
Median	0.3725
Count	768
Total	362.401

-Using Excel

Quantile statistics

Minimum	0.078
5-th percentile	0.14035
Q1	0.24375
median	0.3725
Q3	0.62625
95-th percentile	1.13285
Maximum	2.42
Range	2.342
Interquartile range (IQR)	0.3825

Descriptive statistics

Standard deviation	0.331328595
Coefficient of variation (CV)	0.7021513764
Kurtosis	5.594953528
Mean	0.4718763021
Median Absolute Deviation (MAD)	0.1675
Skewness	1.919911066
Sum	362.401
Variance	0.1097786379
Monotonicity	Not monotonic

-Using Pandas Profiling in Python

Diabetes Pedigree Function determines the risk of diabetes depending mainly on the genetics. We can use this metric to determine whether the diabetic patients have obtained it genetically or not.