Pima Indians Diabetes Database

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1
5	5	116	74	0	0	25.6	0.201	30	0
6	3	78	50	32	88	31.0	0.248	26	1
7	10	115	0	0	0	35.3	0.134	29	0
8	2	197	70	45	543	30.5	0.158	53	1
9	8	125	96	0	0	0.0	0.232	54	1
10	4	110	92	0	0	37.6	0.191	30	0
11	10	168	74	0	0	38.0	0.537	34	1
12	10	139	80	0	0	27.1	1.441	57	0
13	1	189	60	23	846	30.1	0.398	59	1
14	5	166	72	19	175	25.8	0.587	51	1
15	7	100	0	0	0	30.0	0.484	32	1
16	0	118	84	47	230	45.8	0.551	31	1
17	7	107	74	0	0	29.6	0.254	31	1
18	1	103	30	38	83	43.3	0.183	33	0
19	1	115	70	30	96	34.6	0.529	32	1

Context:

This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective of the dataset is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset. Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of Pima Indian heritage.

Objective:

Extract insights using visual and statistical exploration.

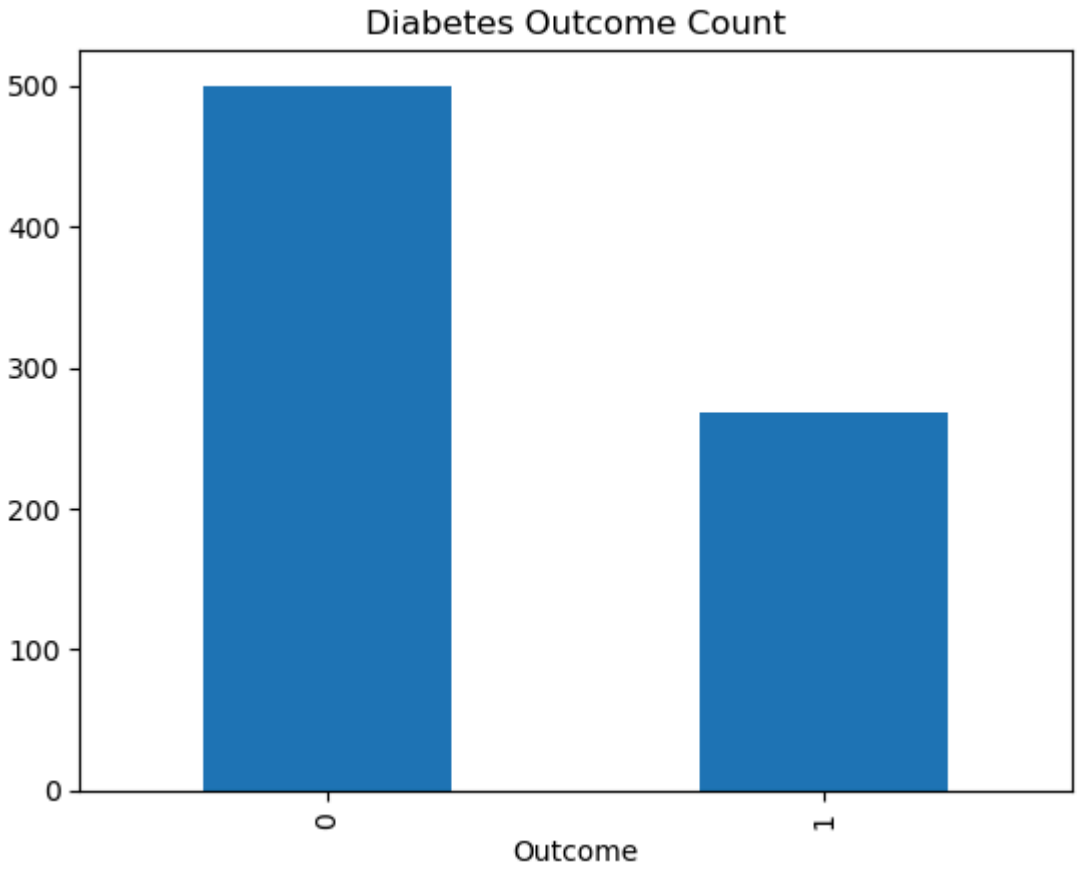
What I have used:

- Pandas
- Seaborn
- Matplotlib

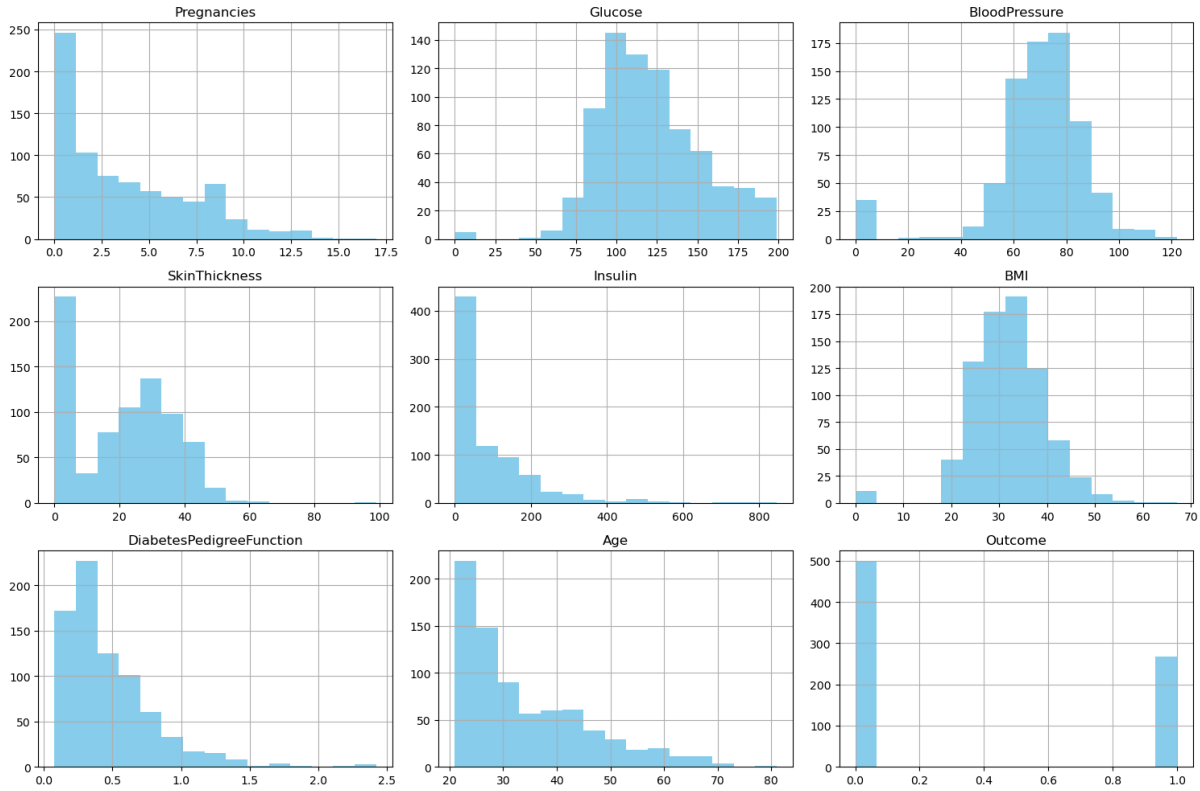
Using these technologies to find the insights in the dataset of a population with diabetes and not diabetes, which tells the factors that affect the possibility of person diagnosing from diabetes.

Univariate Analysis:

Checking class distribution (outcome column), how the affected people are,The bar graph tells that there is a chance of having an imbalance or having a healthy population that are taken on record.



- High glucose levels strongly indicate potential diabetes cases
- Elevated BMI suggests obesity, a key contributor to insulin resistance.
- Insulin levels, particularly when abnormally high or missing, point to metabolic irregularities.
- Age distribution may highlight how diabetes prevalence increases in older individuals.
- The Diabetes Pedigree Function shows whether family history plays a major role.


Pregnancies
Glucose
BloodPressure
SkinThickness
Insulin
BMI
DiabetesPedigreeFunction
Age
Outcome

Bivariate Analysis :

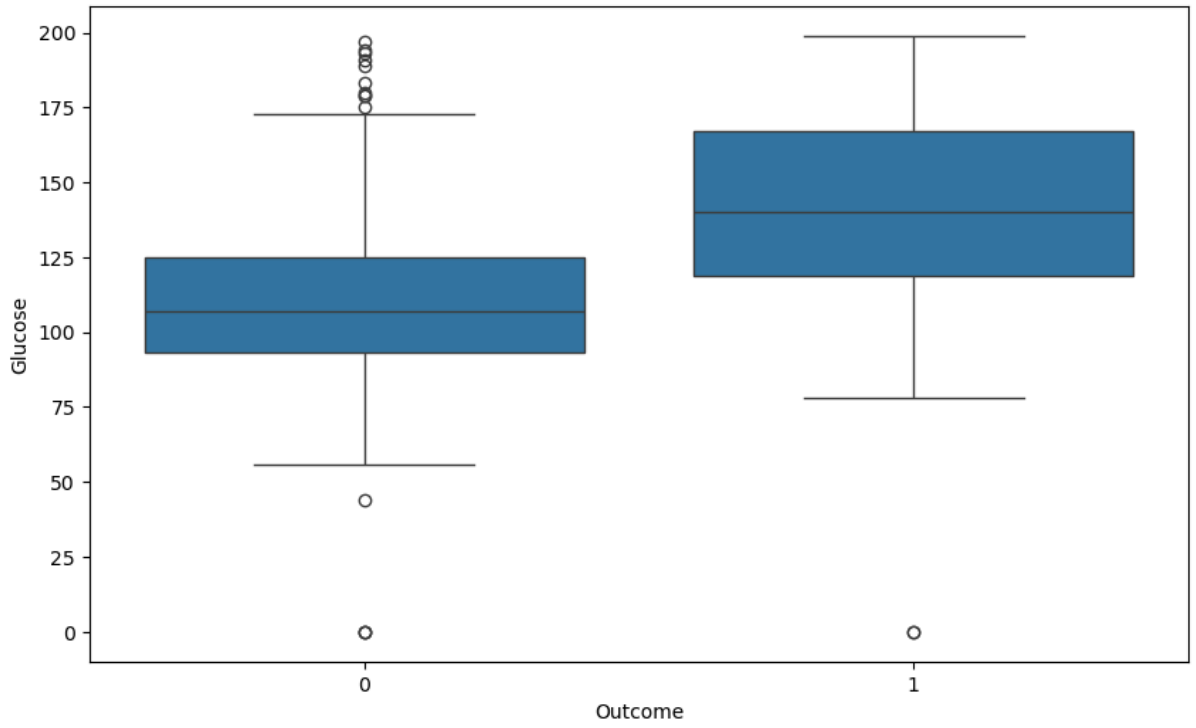
Check how features differ from outcome

1.Median glucose levels for diabetic individuals (Outcome = 1) will likely be higher than non-diabetic individuals (Outcome = 0).

2.Outliers appear as points outside the box range, showing extreme glucose levels.

3.Spread of glucose values in the diabetic group may indicate variability, potentially highlighting cases of uncontrolled diabetes.

4.If the boxes don't overlap much, it suggests glucose is a strong differentiating factor for diabetes classification(glucose is a strong predictor for distinguishing between the two groups).



In case of Glucose and BMI :

If diabetic cases mostly have higher glucose and BMI values, it suggests obesity and high sugar intake are strong diabetes predictors.

In case of Insulin and Glucose:

Clear separation of points can indicate insulin resistance is strongly linked to diabetes.

In case of Age and Glucose:

Older individuals with higher glucose levels might show how age increases diabetes risk.

In case of Blood Pressure and BMI :

If higher BMI values align with increased blood pressure, it strengthens the link between obesity and hypertension.

Correlation Heatmap :

Tells are related the factors are

Positive Correlation, means Strong correlativity (Closer to +1):

1.If glucose and diabetes outcome have a high correlation, it suggests glucose is a key predictor for diabetes.

2.If BMI and blood pressure show strong correlation, it reinforces obesity's impact on hypertension.

Negative Correlation (Closer to -1):

1.If age and insulin levels show negative correlation, it might indicate insulin production decreases with age.