

Input Features

Pregnancies

0

17

Glucose

120

199

Blood Pressure

122

122

BMI

31.99

67.10

Diabetes Pedigree Function

0.47

2.42

Age

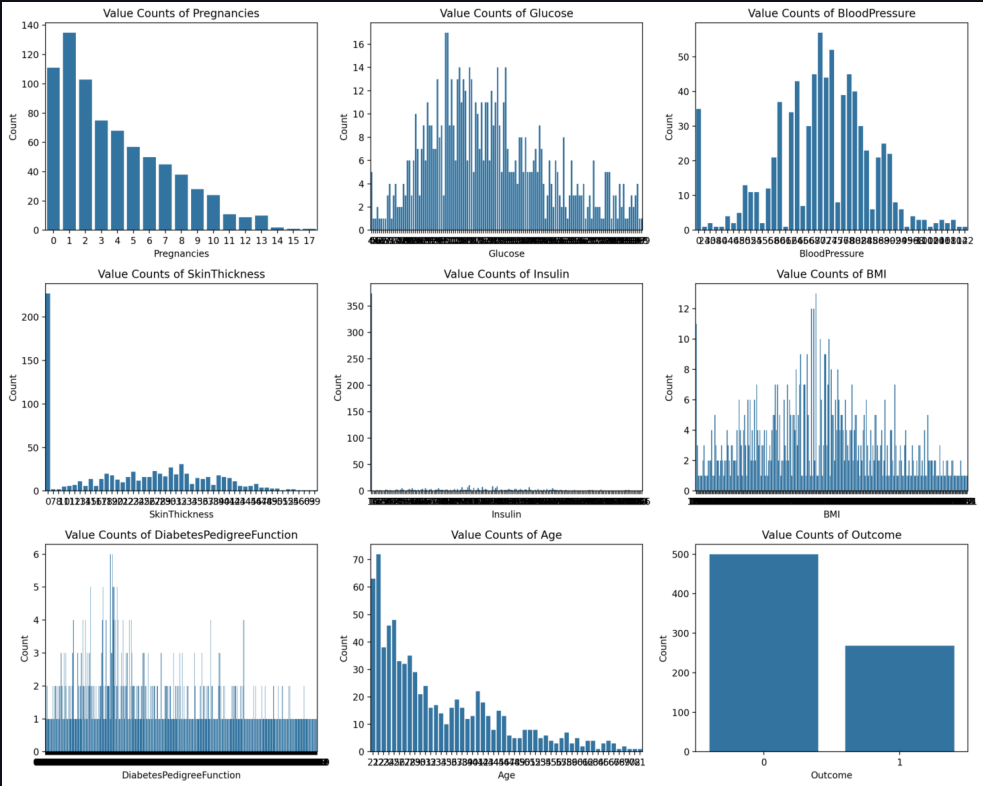
33

81

Predict Diabetes

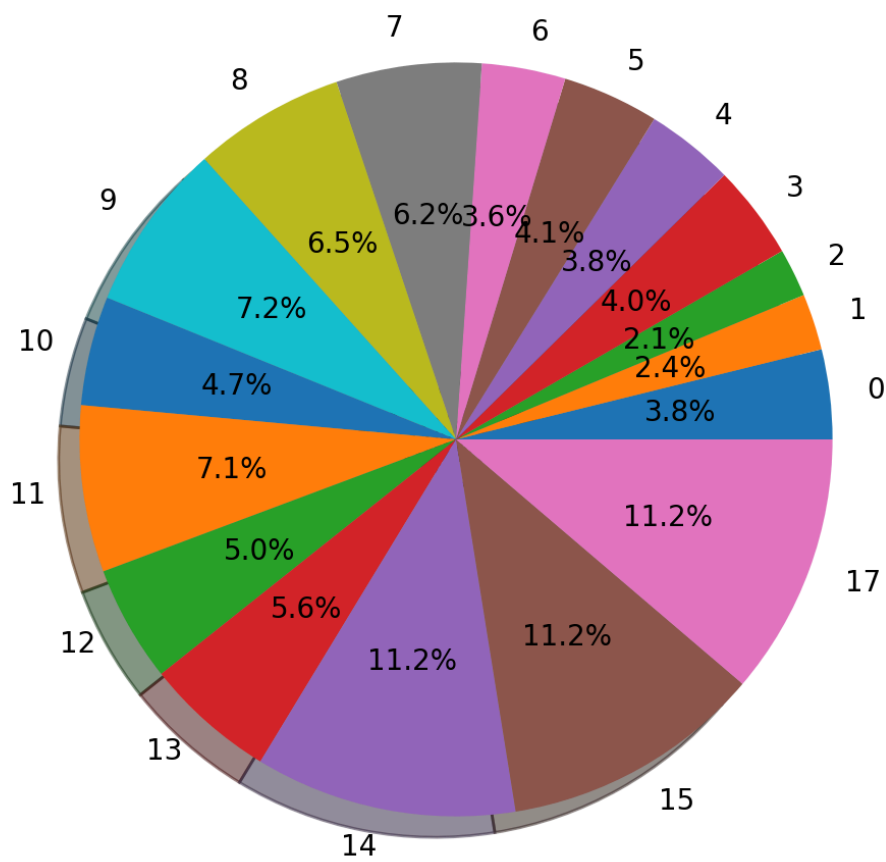
# Diabetes Prediction and Data Analysis

## Data Distribution

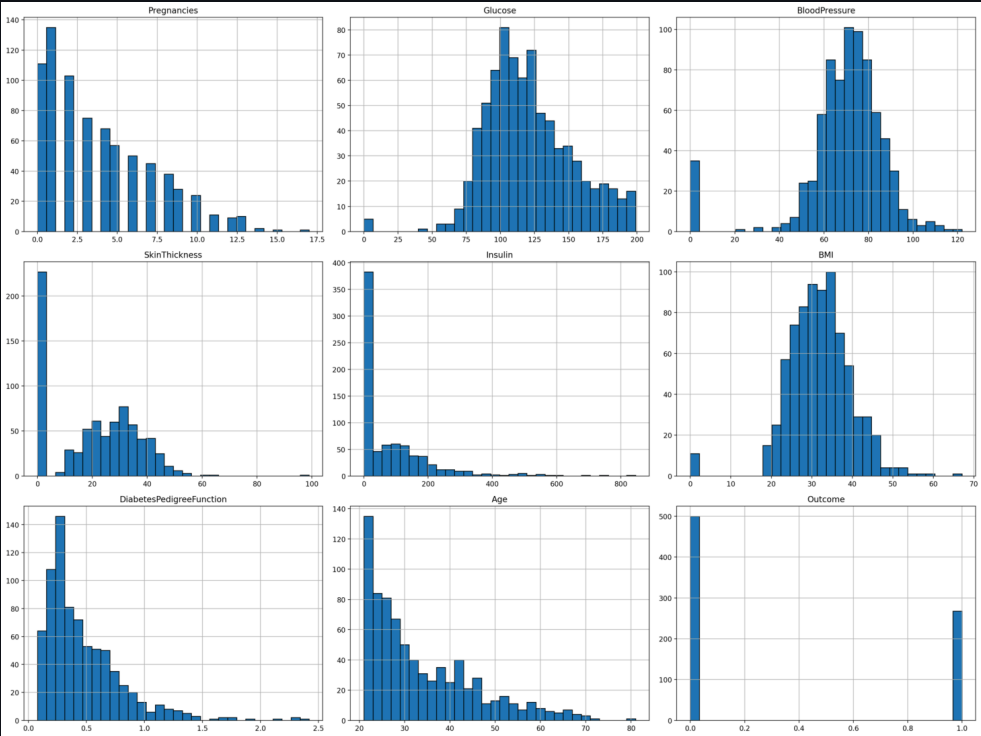


## Diabetes Cases vs. Number of Pregnancies

# Diabetes Cases versus Number of Pregnancies



## Feature Histograms



Missing values in dataset:

	0
Pregnancies	0
Glucose	0
BloodPressure	0
SkinThickness	0
Insulin	0
BMI	0
DiabetesPedigree	0
Age	0
Outcome	0

Data description after removing outliers:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction
count	713	713	713	713	713	713	713
mean	3.7658	118.5905	68.899	19.7658	62.23	31.7927	0.4652
std	3.2941	28.2676	19.8454	15.6989	78.4716	7.9582	0.3216
min	0	44	0	0	0	0	0.078
25%	1	99	62	0	0	27.1	0.243
50%	3	115	72	22	0	32	0.365
75%	6	136	80	32	115	36.2	0.613
max	13	193	122	63	318	67.1	2.42

Model Results:

	model	best_score	best_params
0	svm	0.7785	{"C":1,"kernel":"rbf"}
1	random_forest	0.742	{"n_estimators":10}
2	logistic_regression	0.7686	{"C":1}
3	decision_tree	-0.3474	{"criterion":"friedman_mse","splitter":"random"}