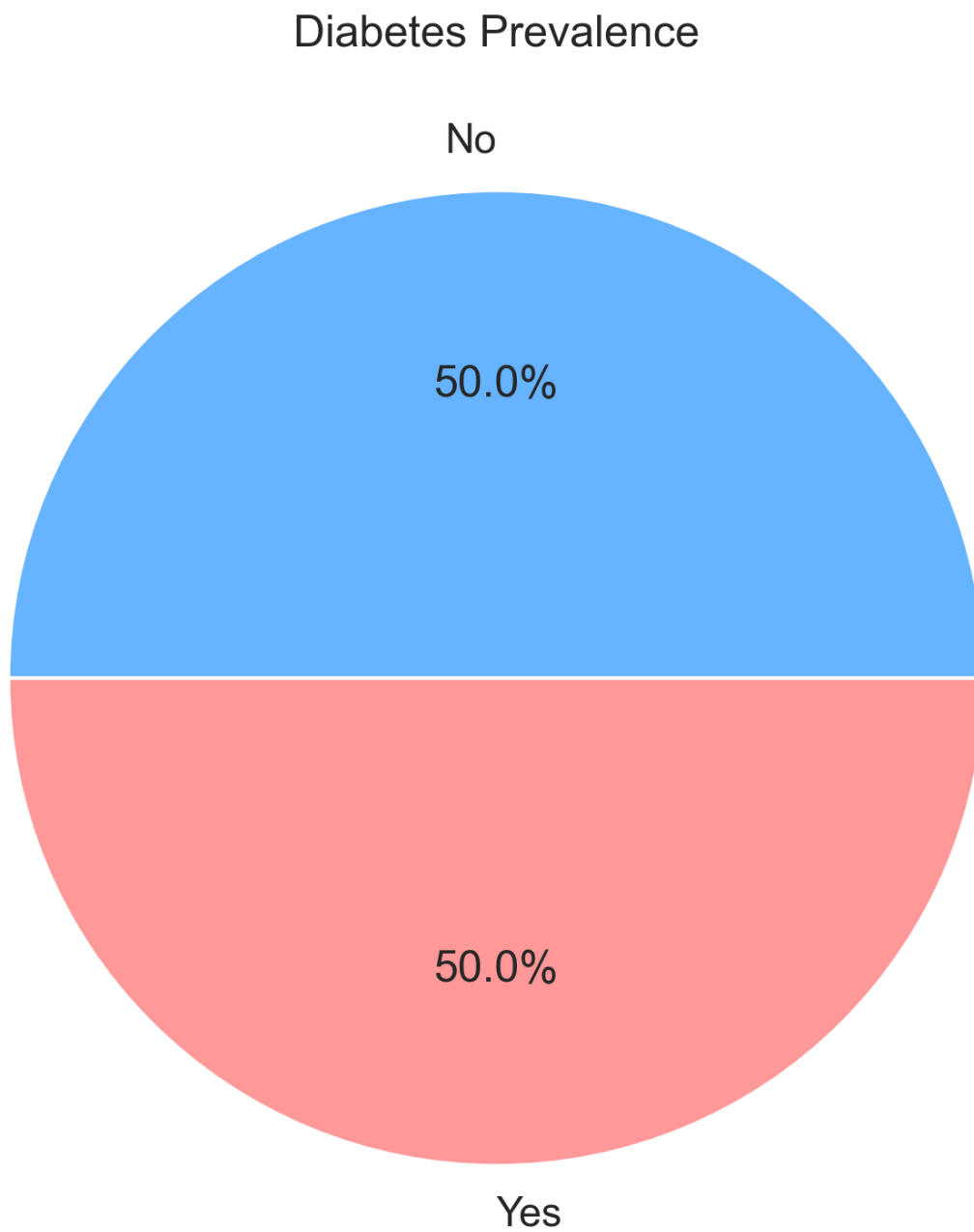# Diabetes Dataset Analysis Report

This report presents an exploratory data analysis of the diabetes dataset, focusing on key risk factors and their relationships with diabetes prevalence. The visualizations below include both static and interactive plots for a comprehensive understanding.

Univariate Analysis

Pie chart showing the proportion of diabetic vs. non-diabetic individuals.

BMI Distribution (Histogram with KDE)
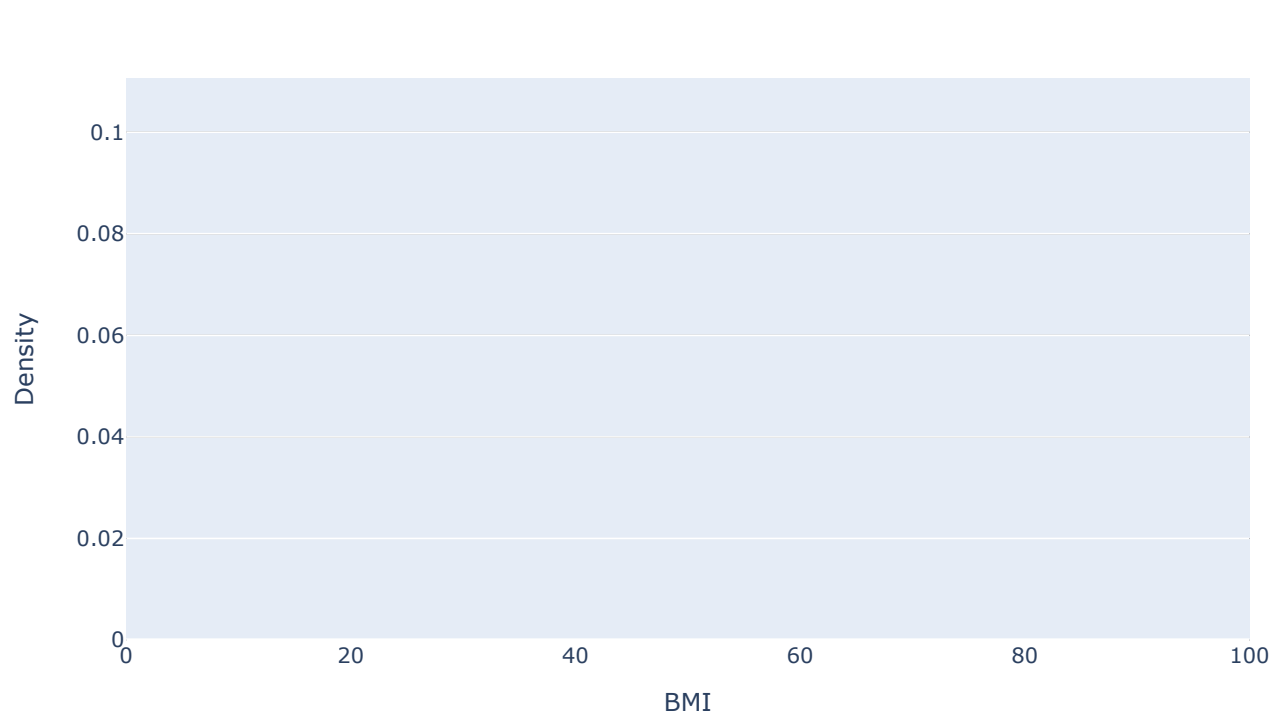
Interactive histogram with KDE overlay for BMI distribution.

Bivariate Analysis

High Blood Pressure vs. Diabetes
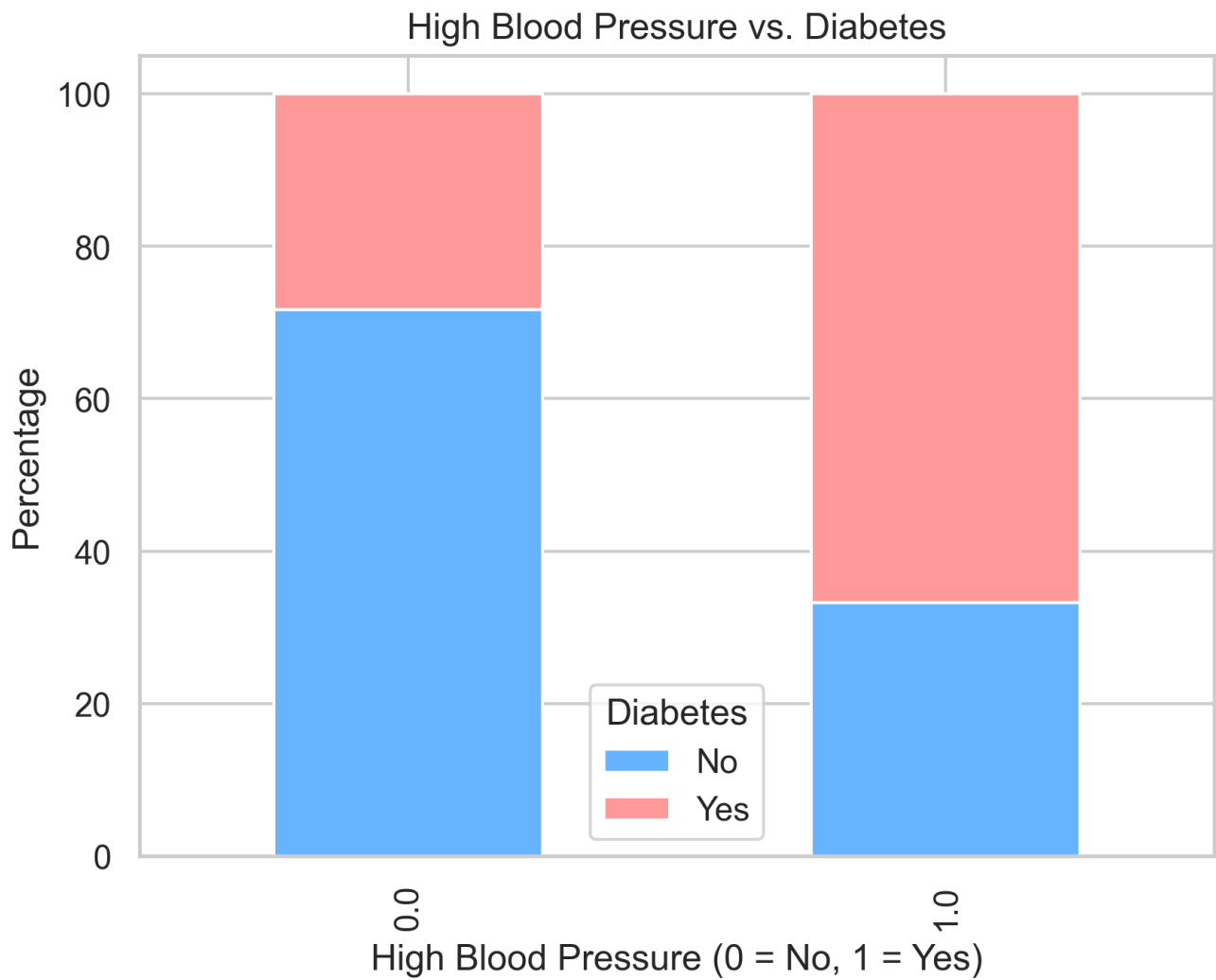
Stacked bar chart showing the percentage of diabetic individuals by high blood pressure status.

BMI by Diabetes Status
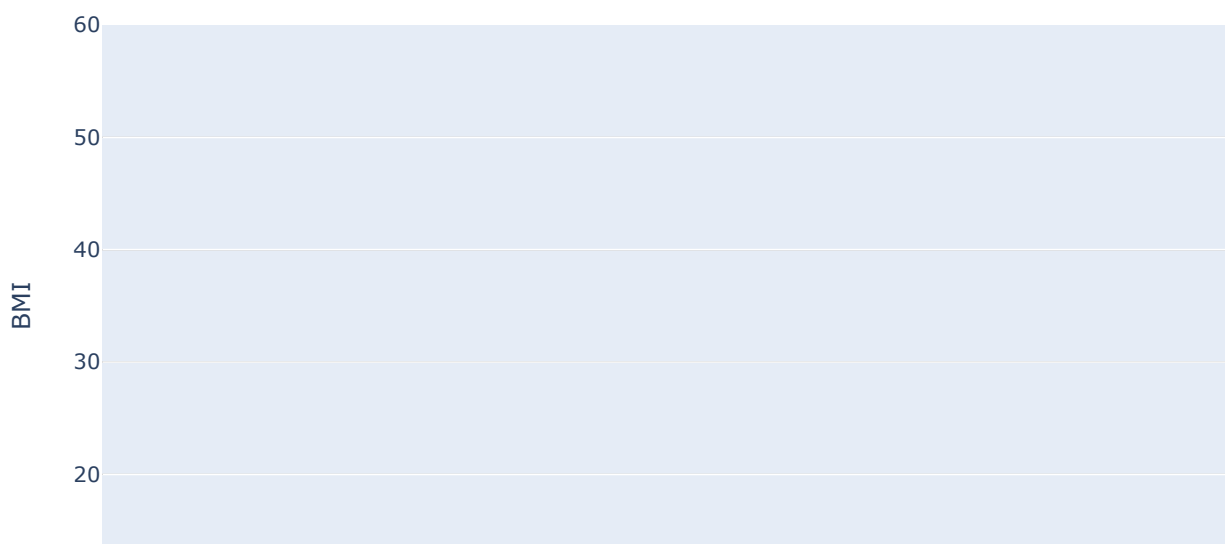
10

Diabetes (0 = No, 1 = Yes)

Interactive violin plot comparing BMI distributions for diabetic and non-diabetic individuals.

BMI vs. Age by Diabetes Status

Interactive scatter plot showing BMI vs. Age, colored by diabetes status.

Summary and Correlations

Correlation Matrix

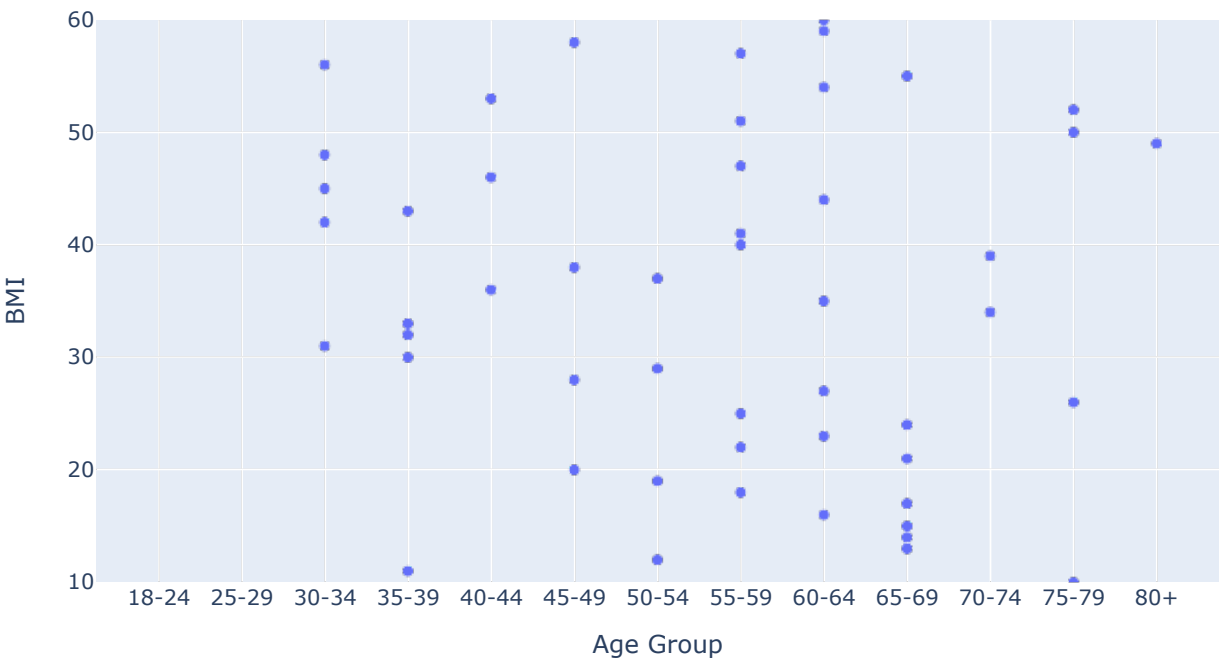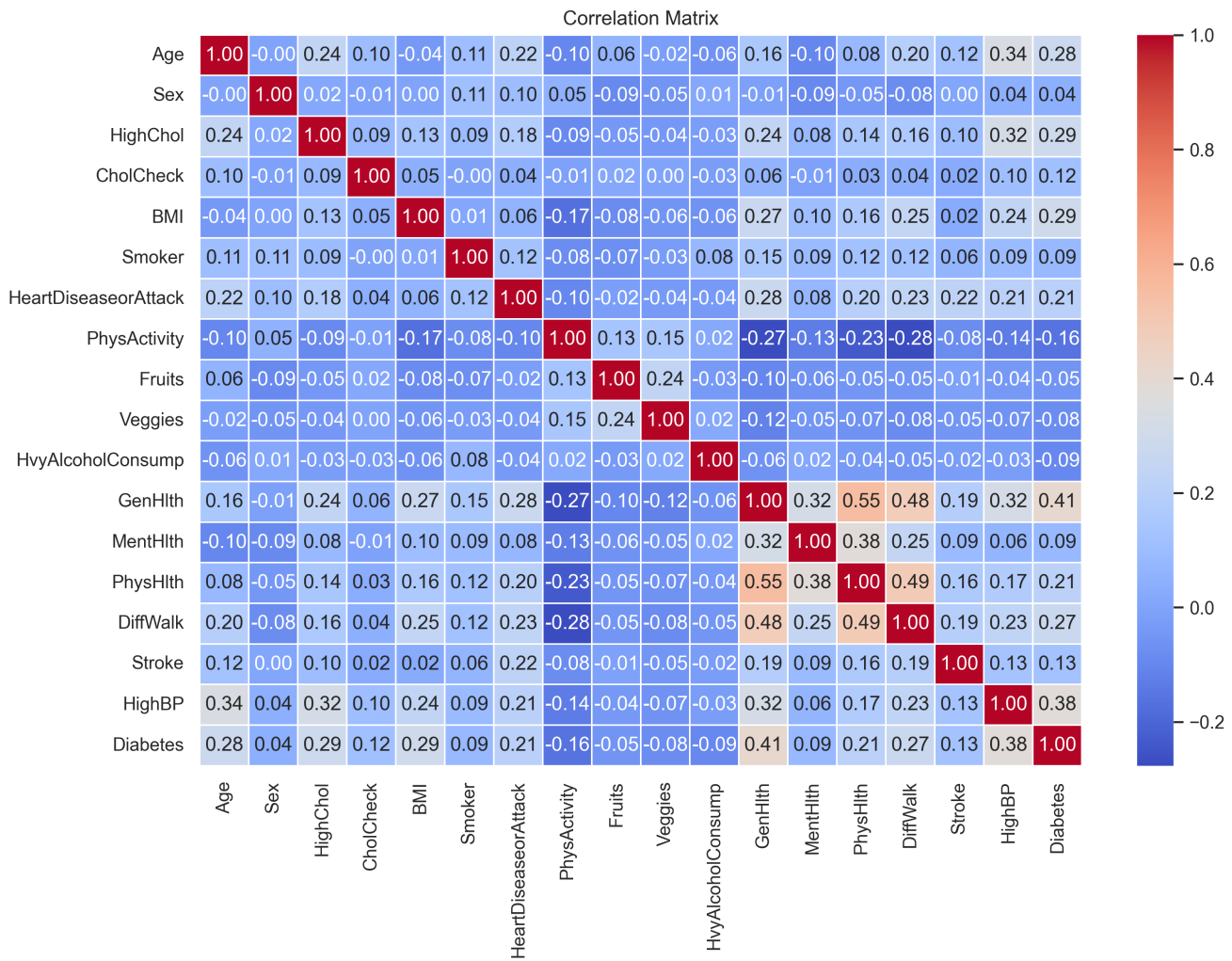|	Age	Sex	HighChol	CholCheck	BMI	Smoker	HeartDiseaseorAttack	PhysActivity	Fruits	Veggies	HvyAlcoholConsump	GenHlth	MentHlth	PhysHlth	DiffWalk	Stroke	HighBP	Diabetes
Age	1.00	-0.00	0.24	0.10	-0.04	0.11	0.22	-0.10	0.06	-0.02	-0.06	0.16	-0.10	0.08	0.20	0.12	0.34	0.28
Sex	-0.00	1.00	0.02	-0.01	0.00	0.11	0.10	0.05	-0.09	-0.05	0.01	-0.01	-0.09	-0.05	-0.08	0.00	0.04	0.04
HighChol	0.24	0.02	1.00	0.09	0.13	0.09	0.18	-0.09	-0.05	-0.04	-0.03	0.24	0.08	0.14	0.16	0.10	0.32	0.29
CholCheck	0.10	-0.01	0.09	1.00	0.05	-0.00	0.04	-0.01	0.02	0.00	-0.03	0.06	-0.01	0.03	0.04	0.02	0.10	0.12
BMI	-0.04	0.00	0.13	0.05	1.00	0.01	0.06	-0.17	-0.08	-0.06	-0.06	0.27	0.10	0.16	0.25	0.02	0.24	0.29
Smoker	0.11	0.11	0.09	-0.00	0.01	1.00	0.12	-0.08	-0.07	-0.03	0.08	0.15	0.09	0.12	0.12	0.06	0.09	0.09
HeartDiseaseorAttack	0.22	0.10	0.18	0.04	0.06	0.12	1.00	-0.10	-0.02	-0.04	-0.04	0.28	0.08	0.20	0.23	0.22	0.21	0.21
PhysActivity	-0.10	0.05	-0.09	-0.01	-0.17	-0.08	-0.10	1.00	0.13	0.15	0.02	-0.27	-0.13	-0.23	-0.28	-0.08	-0.14	-0.16
Fruits	0.06	-0.09	-0.05	0.02	-0.08	-0.07	-0.02	0.13	1.00	0.24	-0.03	-0.10	-0.06	-0.05	-0.05	-0.01	-0.04	-0.05
Veggies	-0.02	-0.05	-0.04	0.00	-0.06	-0.03	-0.04	0.15	0.24	1.00	0.02	-0.12	-0.05	-0.07	-0.08	-0.05	-0.07	-0.08
HvyAlcoholConsump	-0.06	0.01	-0.03	-0.03	-0.06	0.08	-0.04	0.02	-0.03	0.02	1.00	-0.06	0.02	-0.04	-0.05	-0.02	-0.03	-0.09
GenHlth	0.16	-0.01	0.24	0.06	0.27	0.15	0.28	-0.27	-0.10	-0.12	-0.06	1.00	0.32	0.55	0.48	0.19	0.32	0.41
MentHlth	-0.10	-0.09	0.08	-0.01	0.10	0.09	0.08	-0.13	-0.06	-0.05	0.02	0.32	1.00	0.38	0.25	0.09	0.06	0.09
PhysHlth	0.08	-0.05	0.14	0.03	0.16	0.12	0.20	-0.23	-0.05	-0.07	-0.04	0.55	0.38	1.00	0.49	0.16	0.17	0.21
DiffWalk	0.20	-0.08	0.16	0.04	0.25	0.12	0.23	-0.28	-0.05	-0.08	-0.05	0.48	0.25	0.49	1.00	0.19	0.23	0.27
Stroke	0.12	0.00	0.10	0.02	0.02	0.06	0.22	-0.08	-0.01	-0.05	-0.02	0.19	0.09	0.16	0.19	1.00	0.13	0.13
HighBP	0.34	0.04	0.32	0.10	0.24	0.09	0.21	-0.14	-0.04	-0.07	-0.03	0.32	0.06	0.17	0.23	0.13	1.00	0.38
Diabetes	0.28	0.04	0.29	0.12	0.29	0.09	0.21	-0.16	-0.05	-0.08	-0.09	0.41	0.09	0.21	0.27	0.13	0.38	1.00

Heatmap showing correlations between all variables, with stronger colors indicating higher correlations.
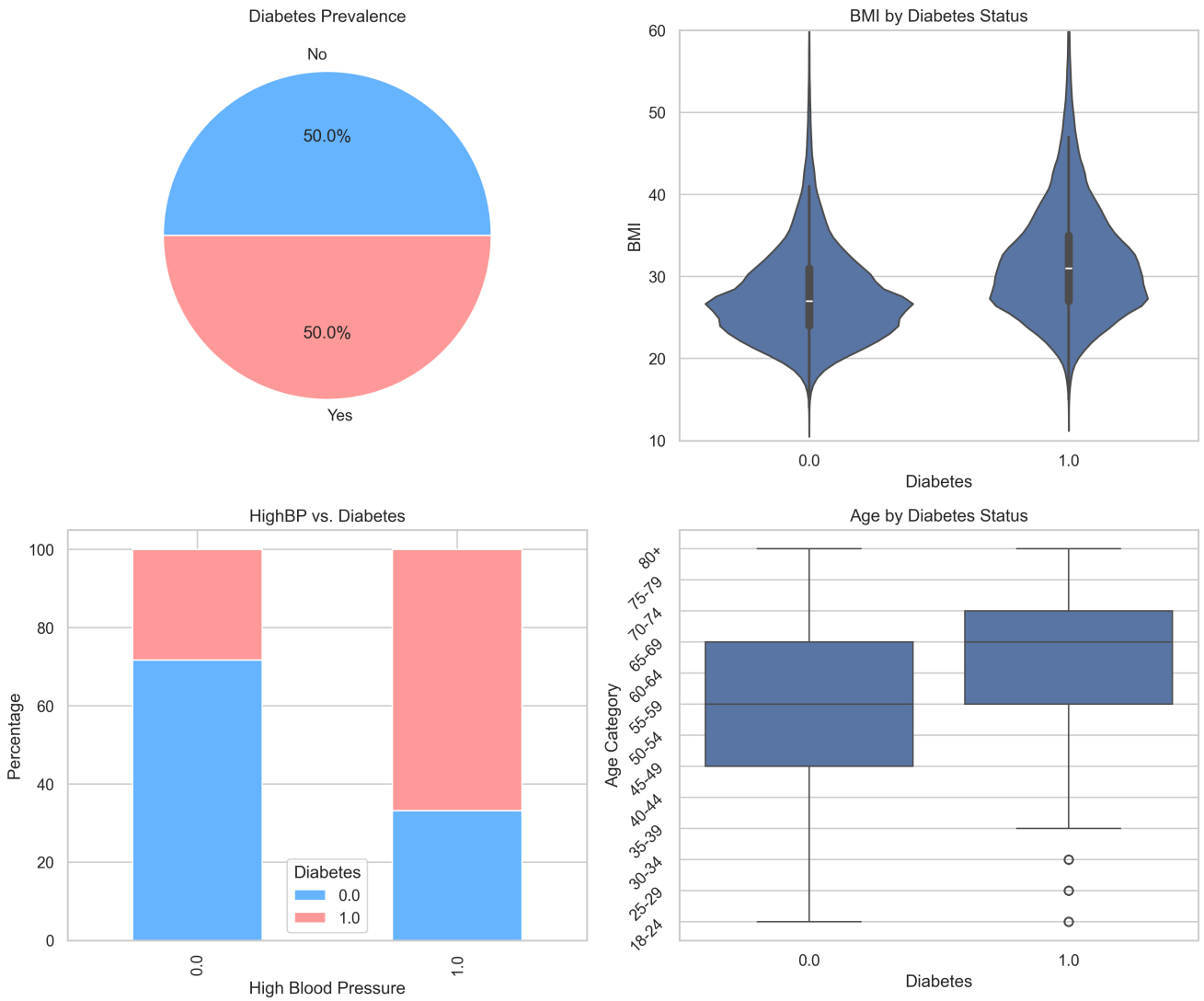
Composite plot summarizing key risk factors for diabetes (prevalence, BMI, HighBP, Age).

Key Insights

- **Prevalence**: Diabetes affects a small but significant portion of the population (~2-5% based on the sample).
- **Risk Factors**: High BMI, high blood pressure, and older age are strongly associated with diabetes.

Recommendations: Focus on screening programs for individuals with high BMI and hypertension, and promote lifestyle interventions like physical activity to reduce diabetes risk.