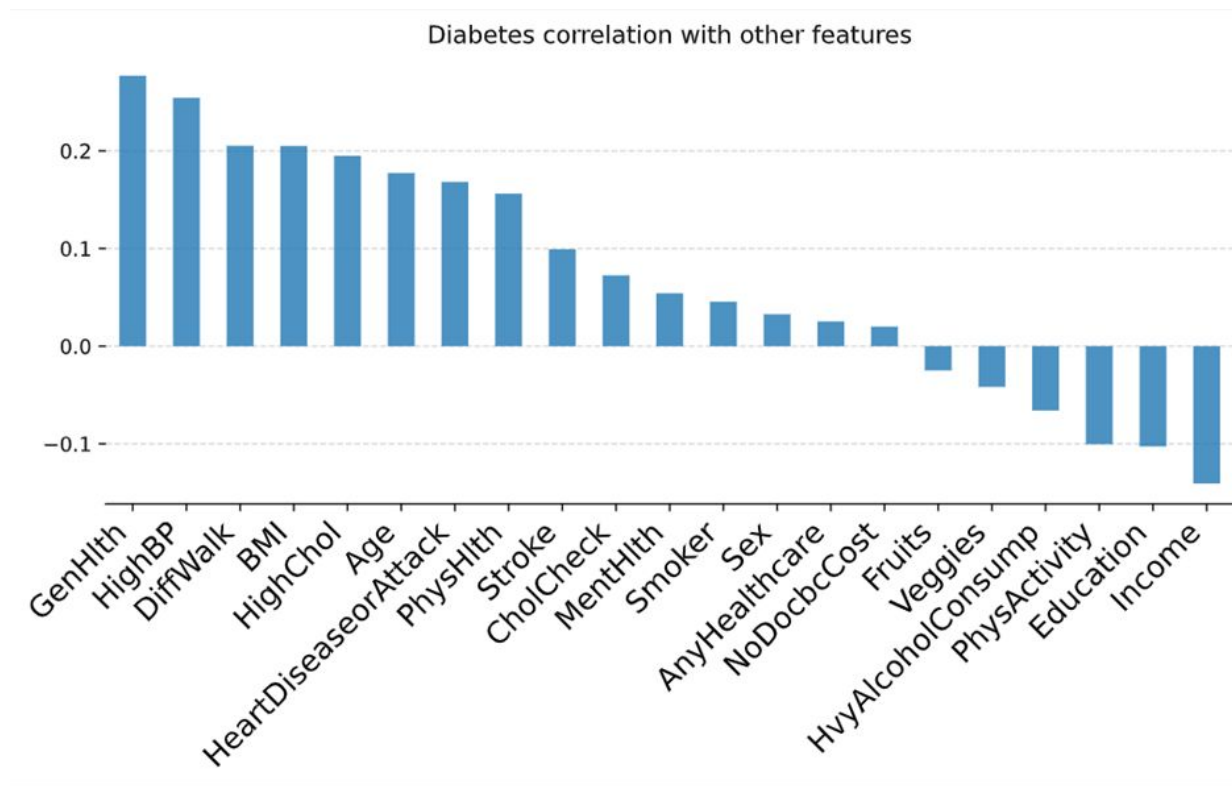


Diabetes Dataset Analysis

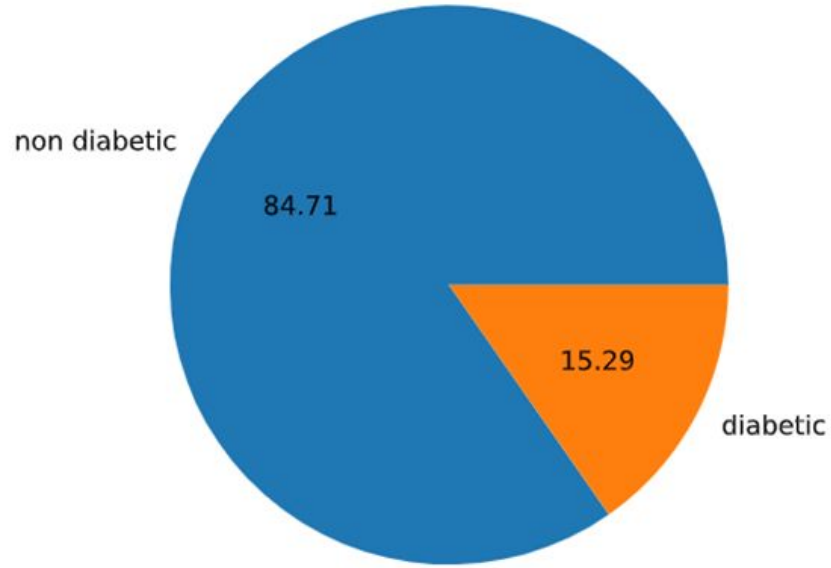
Group 19 - Diabaters

Amna, Haabeel, Kabita, Koket, Mohamed Osman, Mrunal

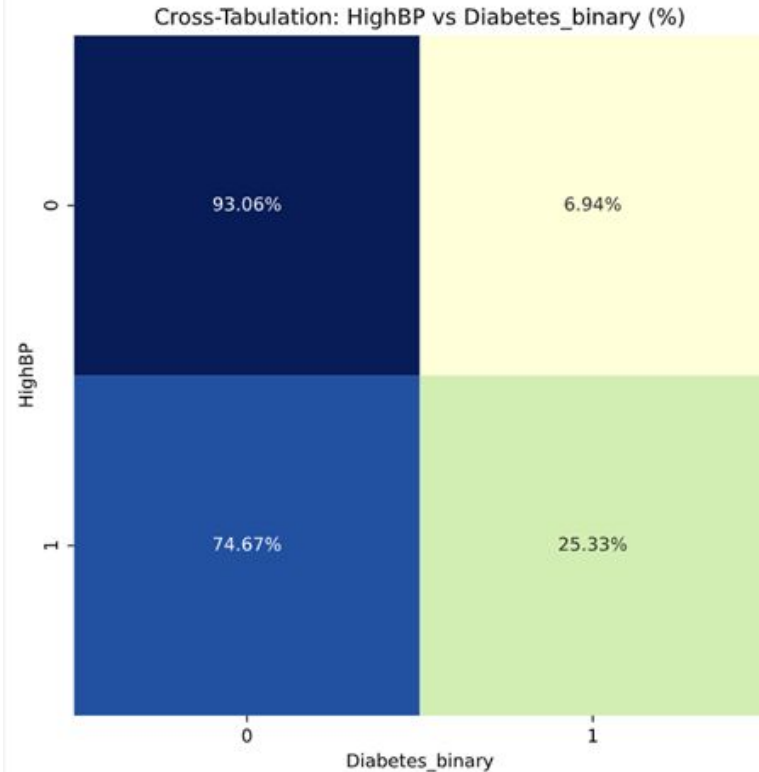
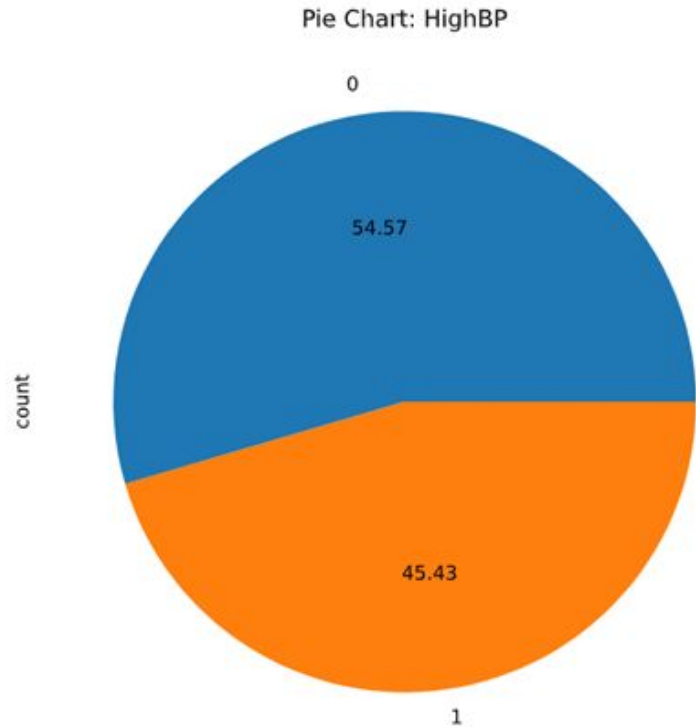
Correlation between Diabetes and chosen features



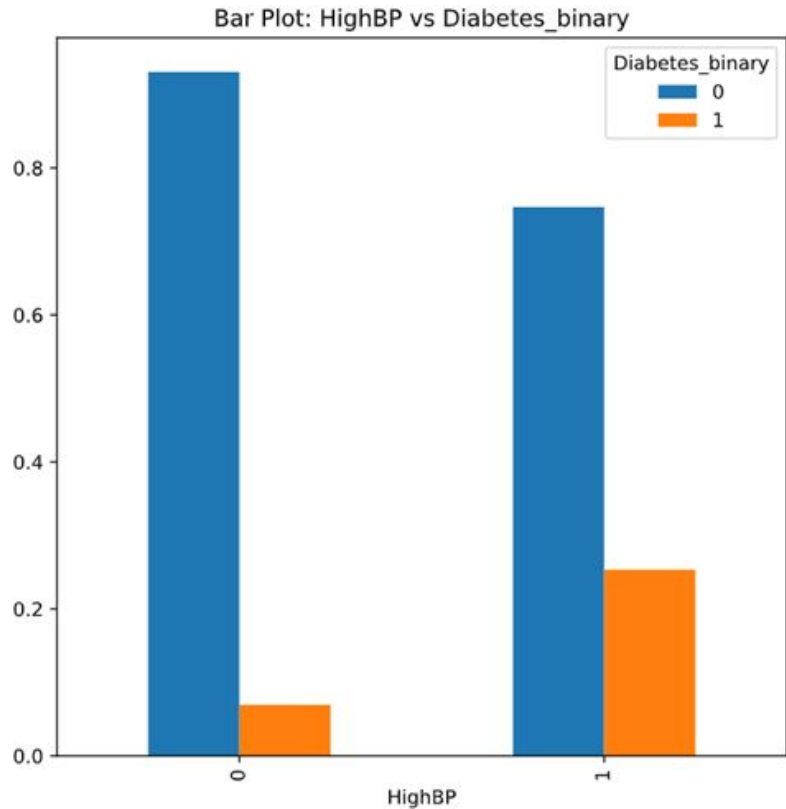
Non Diabetic vs Diabetic Data Distribution



Feature Visualization: High BP

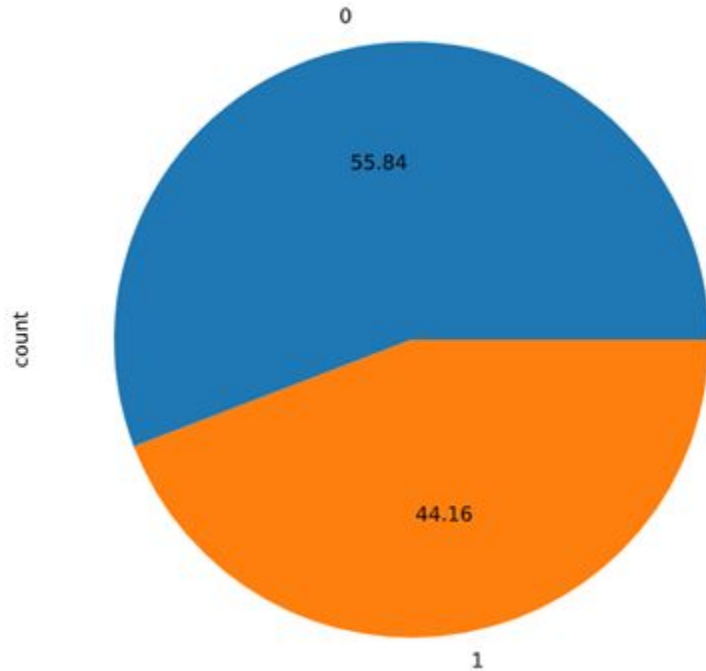


Feature Visualization: High BP

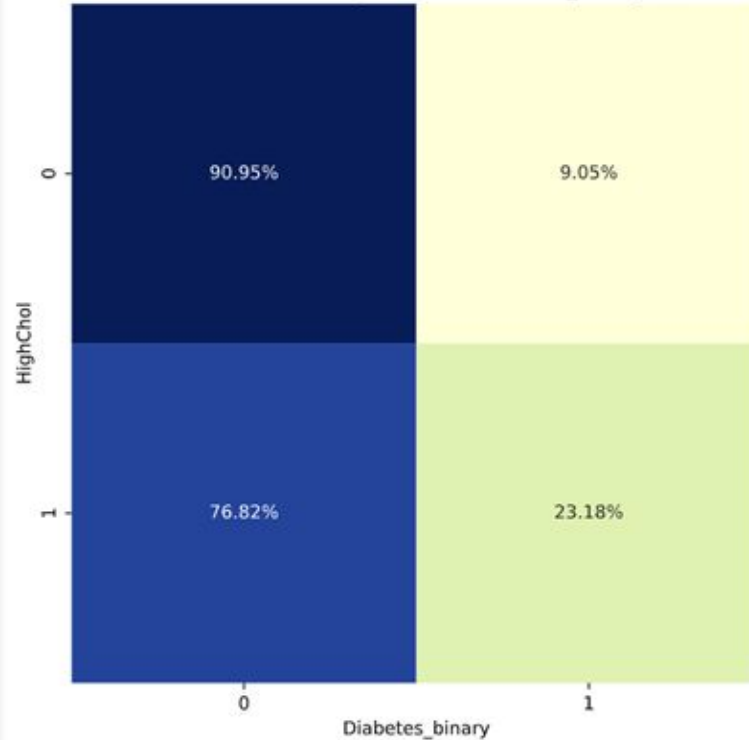


Feature visualization: High Cholesterol

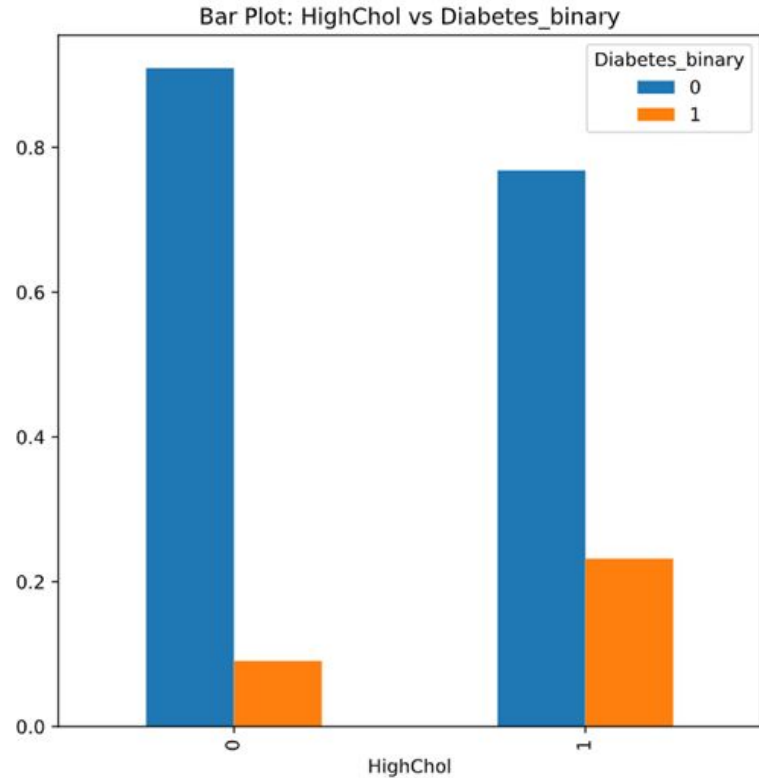
Pie Chart: HighChol



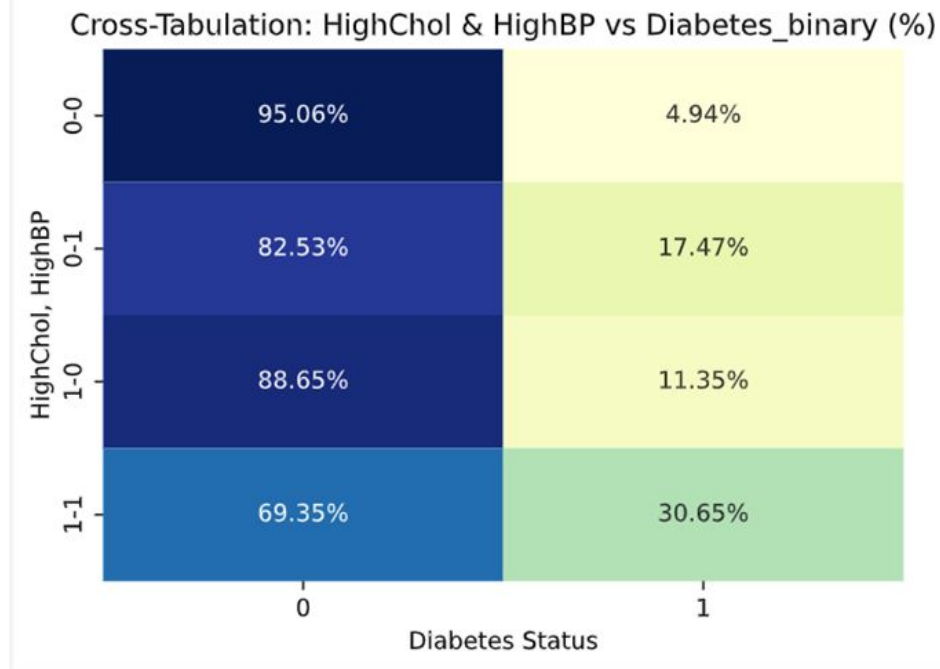
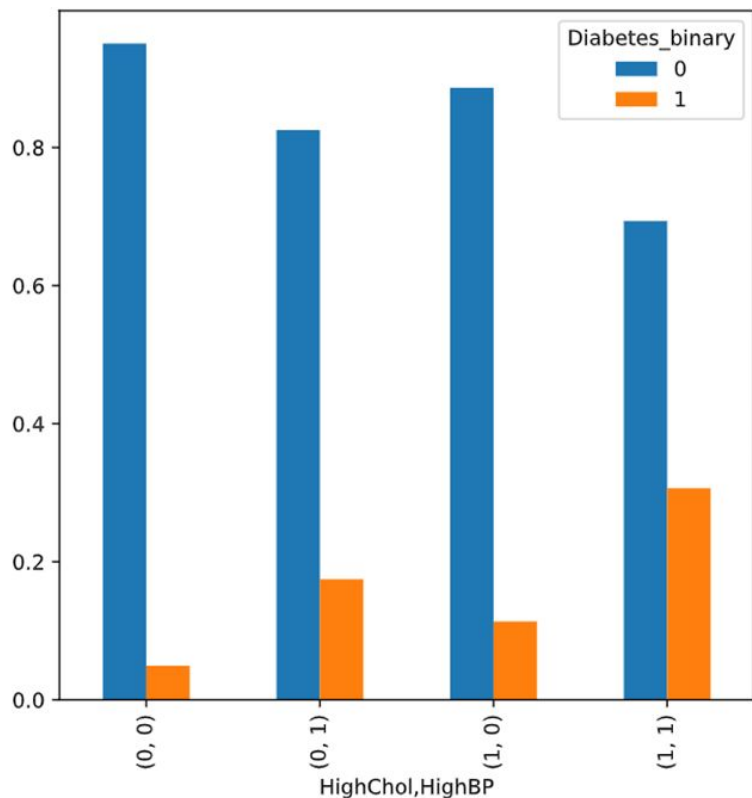
Cross-Tabulation: HighChol vs Diabetes_binary (%)



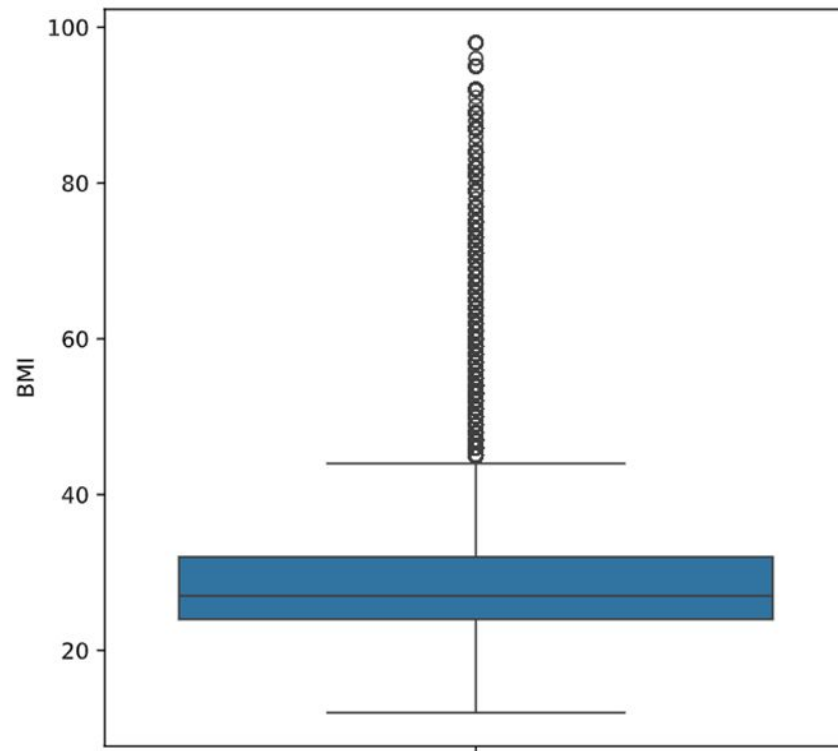
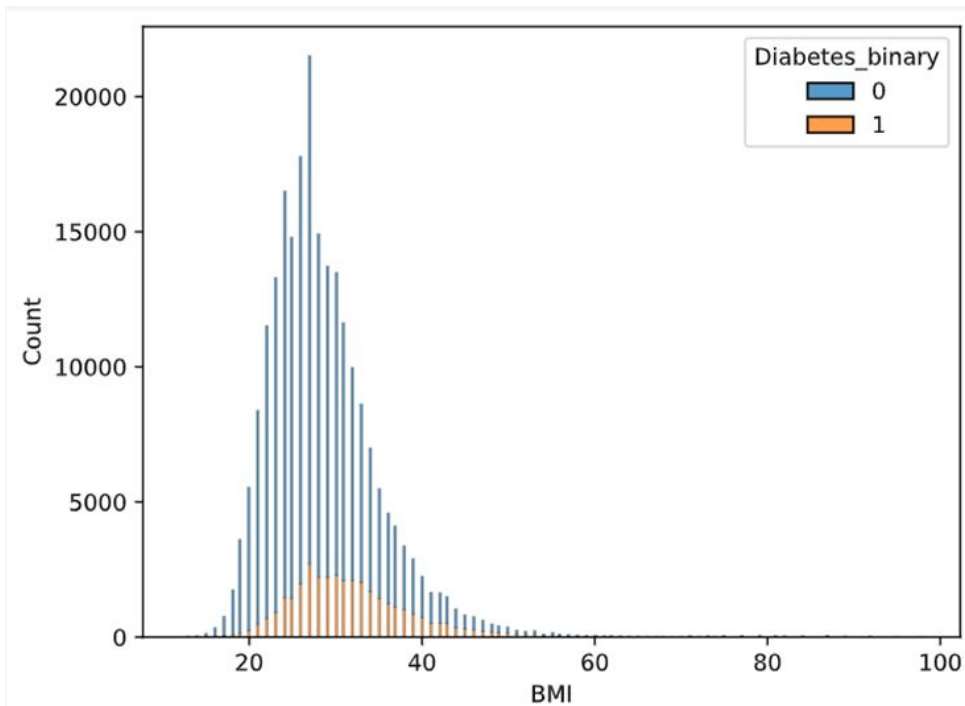
Feature visualization: High Cholesterol



Feature visualization: High BP, High Cholesterol

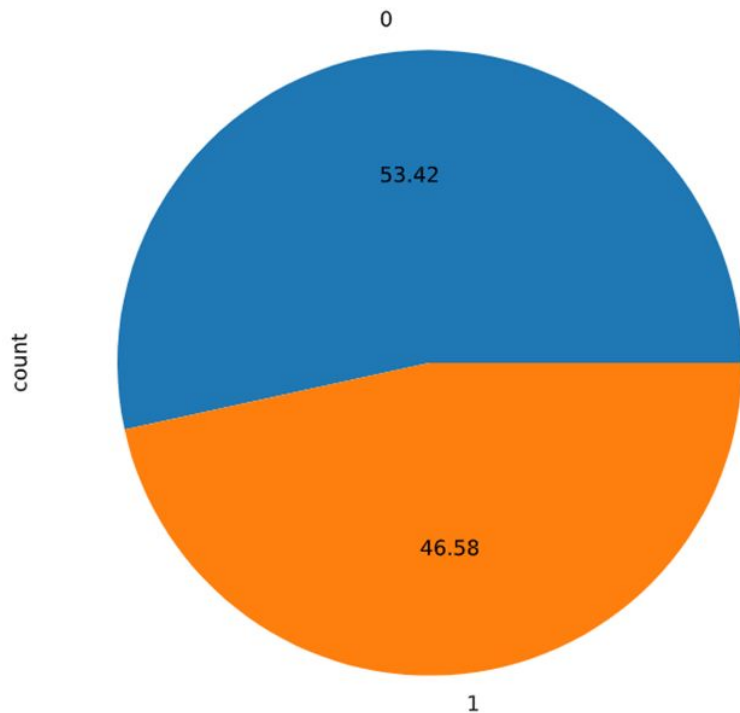


Feature visualization: BMI

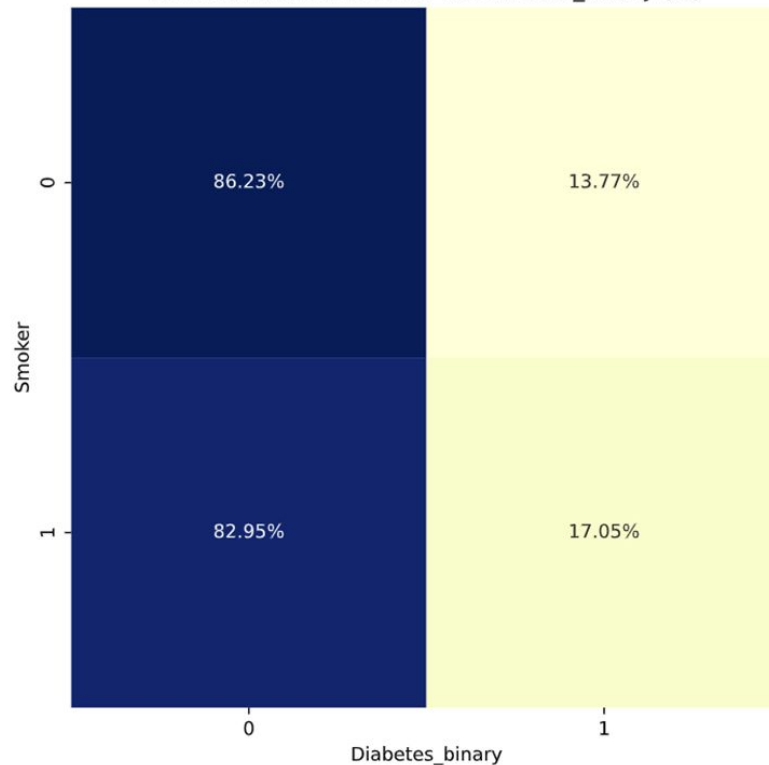


Feature visualization: Smoker

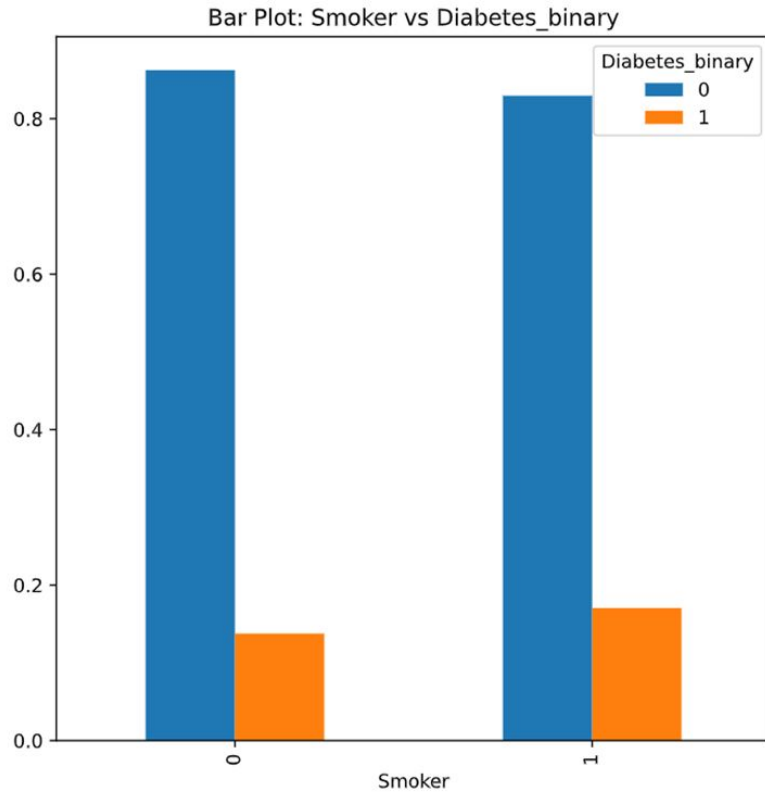
Pie Chart: Smoker



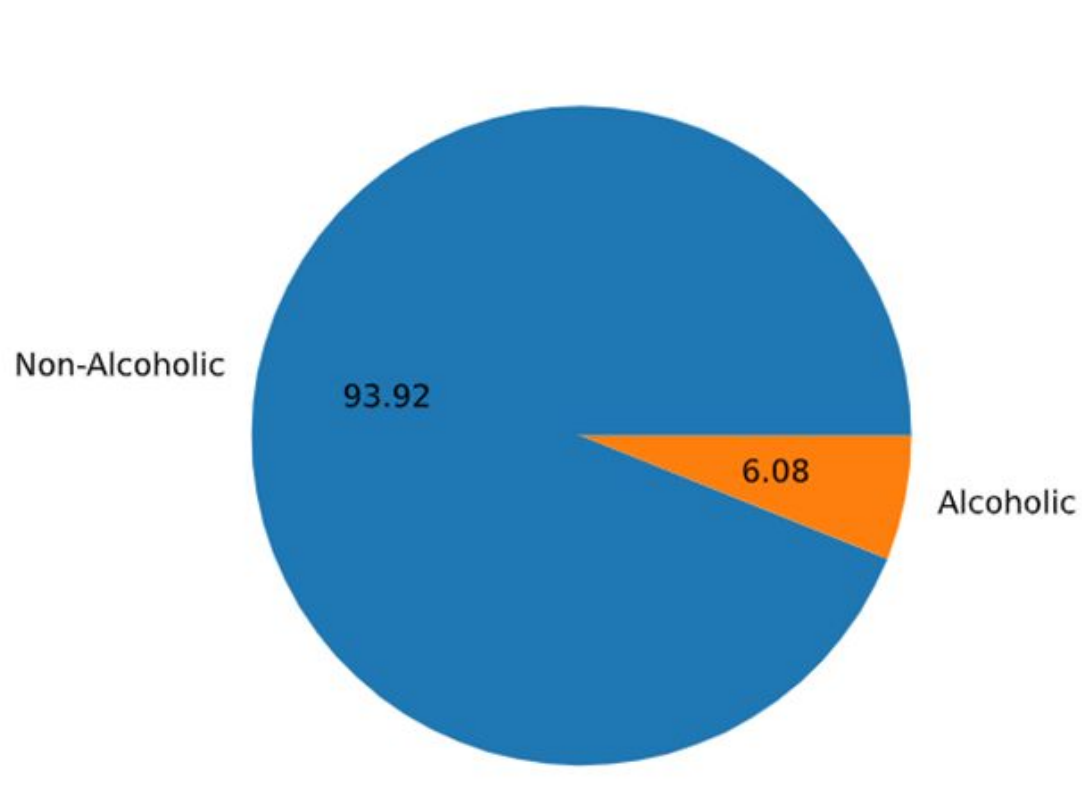
Cross-Tabulation: Smoker vs Diabetes_binary (%)



Feature visualization: Smoker

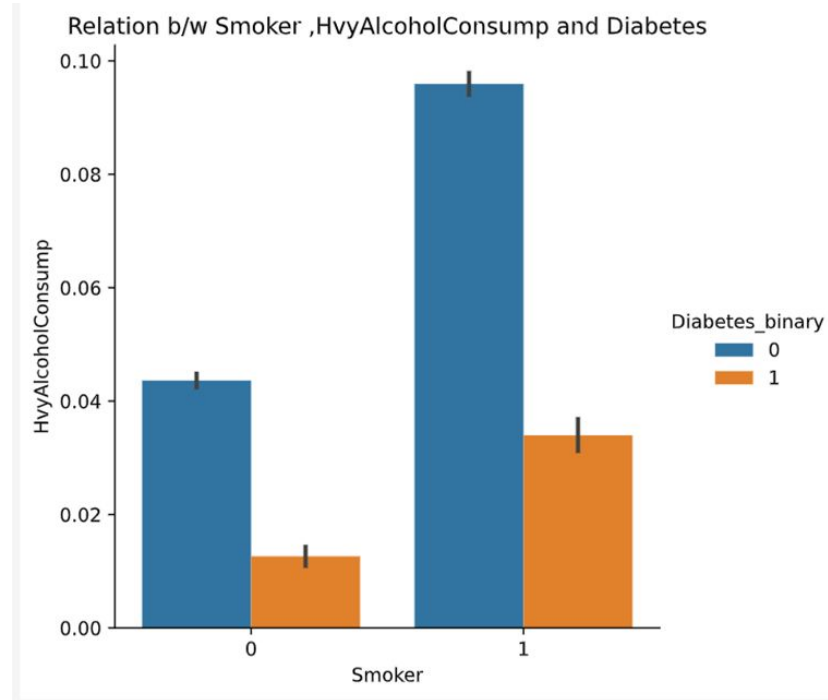
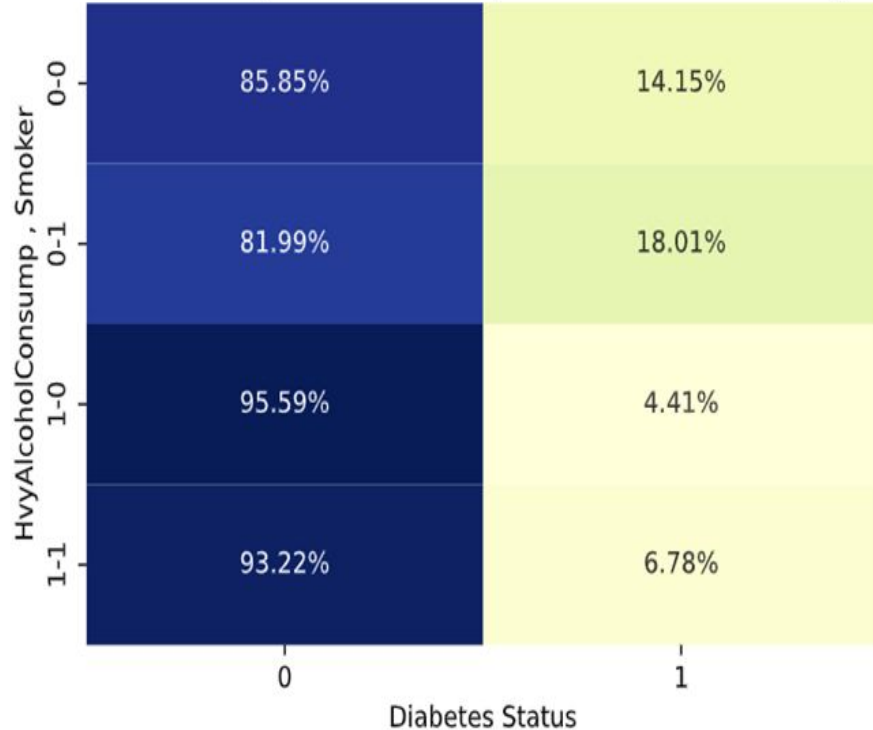


Feature visualization: Alcohol Consumption



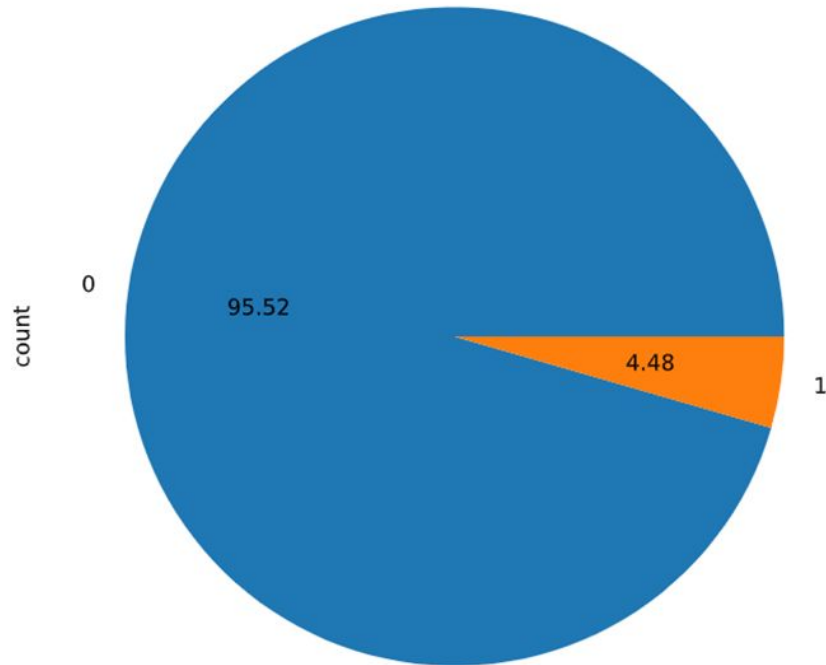
Feature visualization: Smoker and Alcohol Consumption

Cross-Tabulation: HvyAlcoholConsump & Smoker vs Diabetes_binary (%)

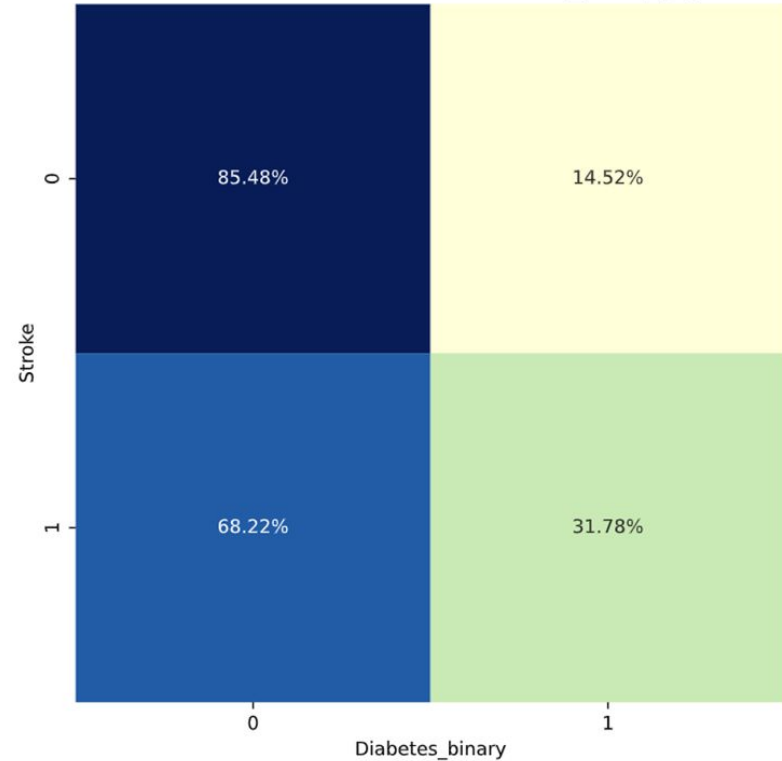


Feature visualization: Stroke

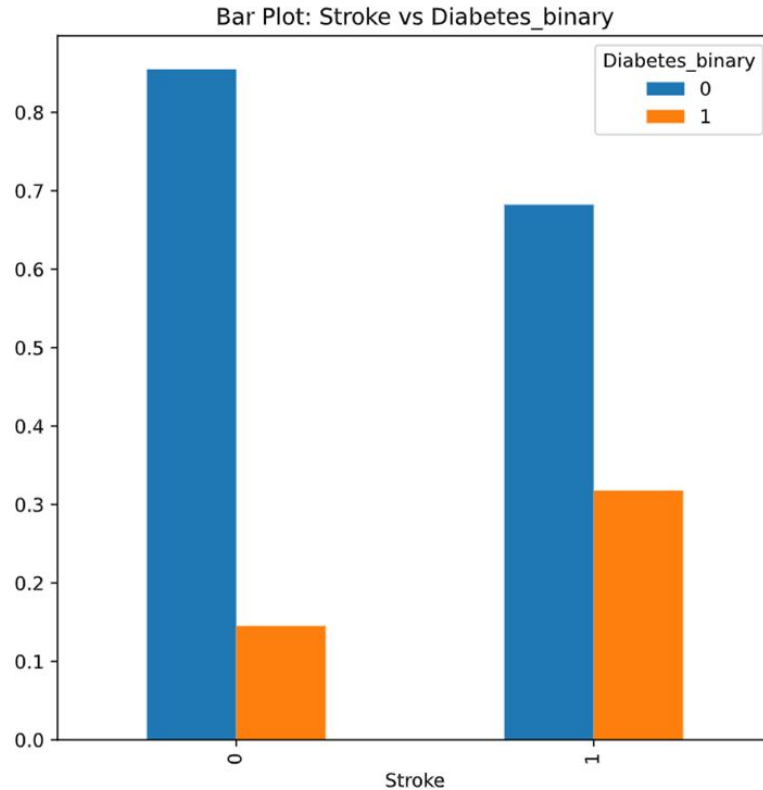
Pie Chart: Stroke



Cross-Tabulation: Stroke vs Diabetes_binary (%)

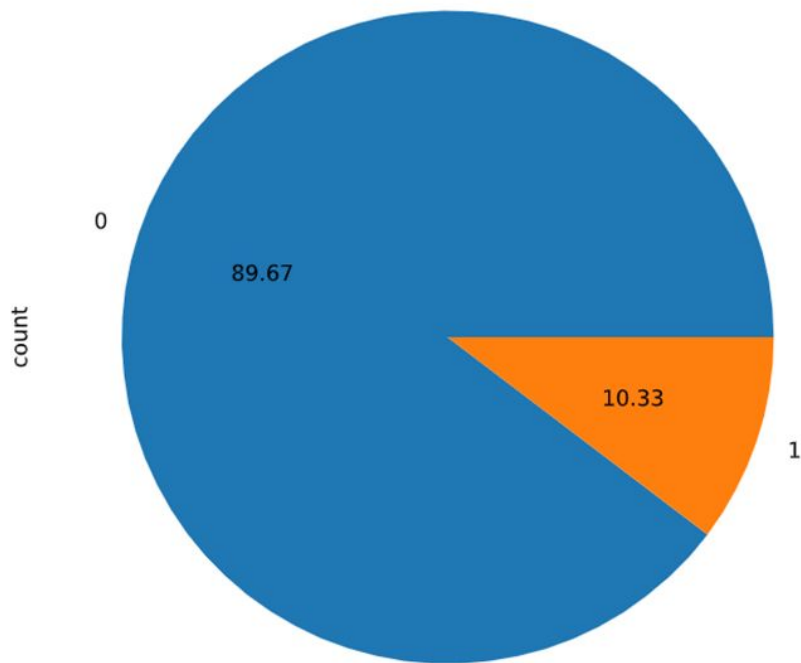


Feature visualization: Stroke

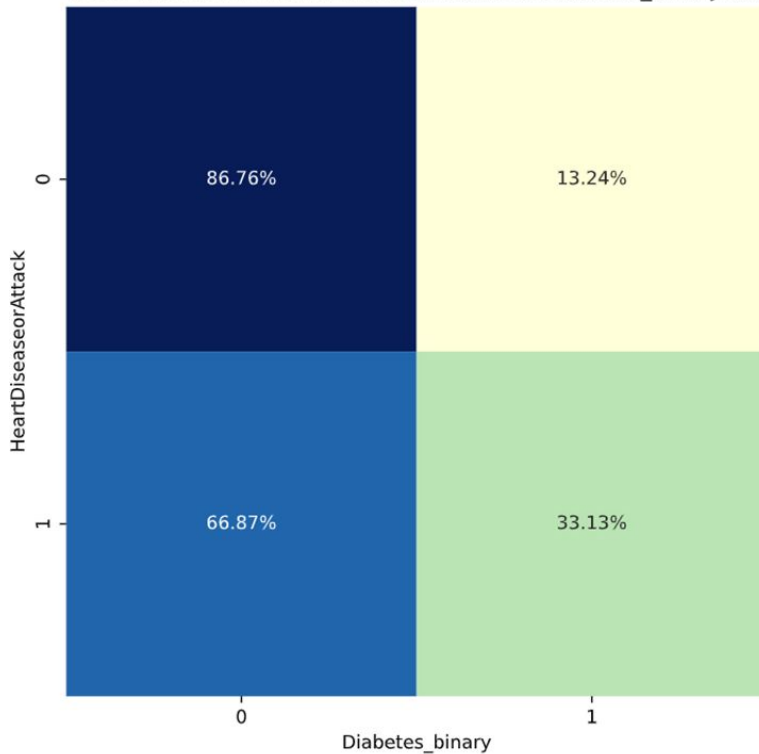


Feature visualization: History of Heart Conditions

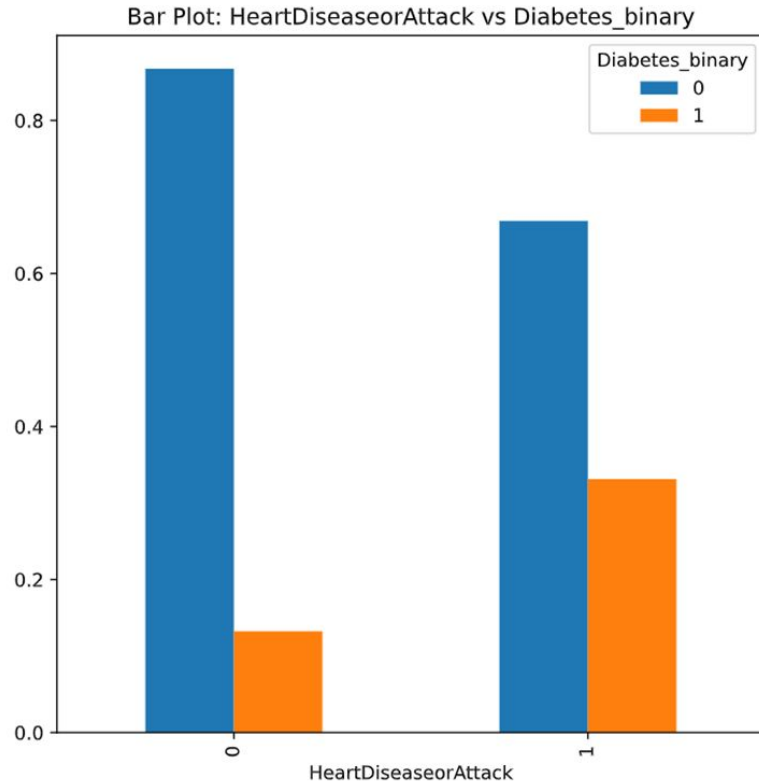
Pie Chart: HeartDiseaseorAttack



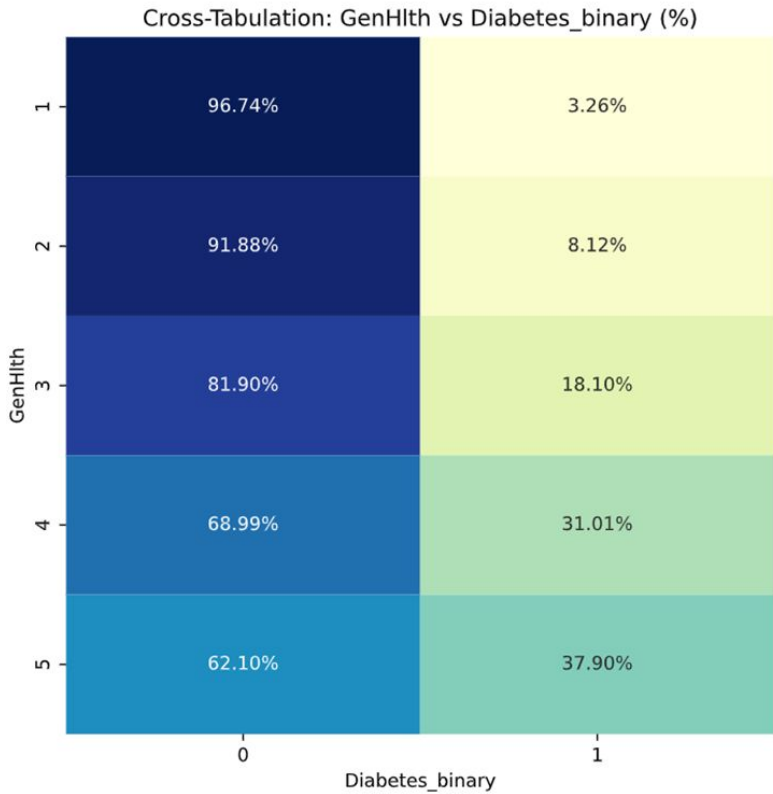
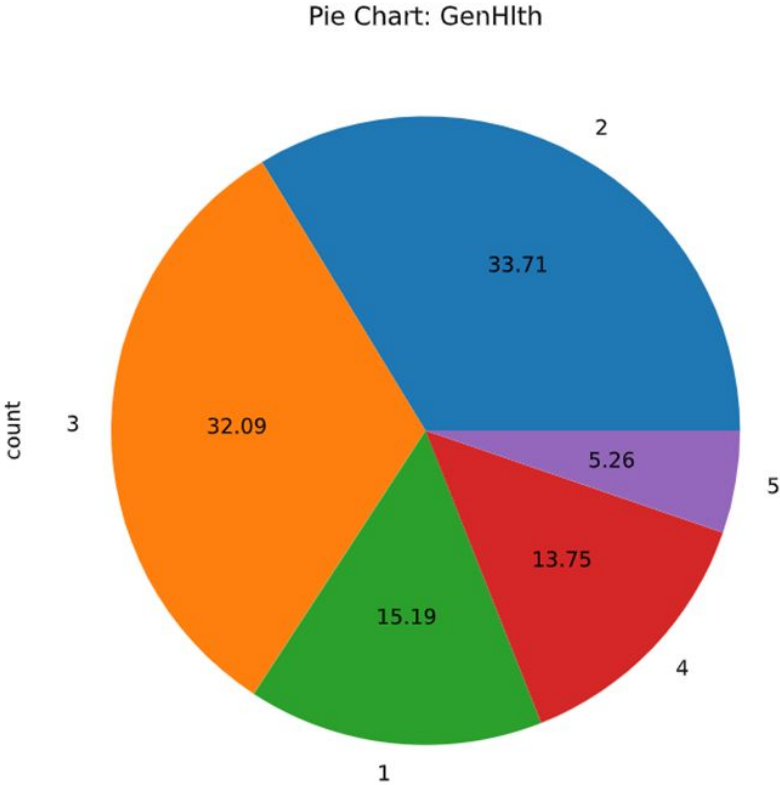
Cross-Tabulation: HeartDiseaseorAttack vs Diabetes_binary (%)



Feature visualization: History of Heart Condition

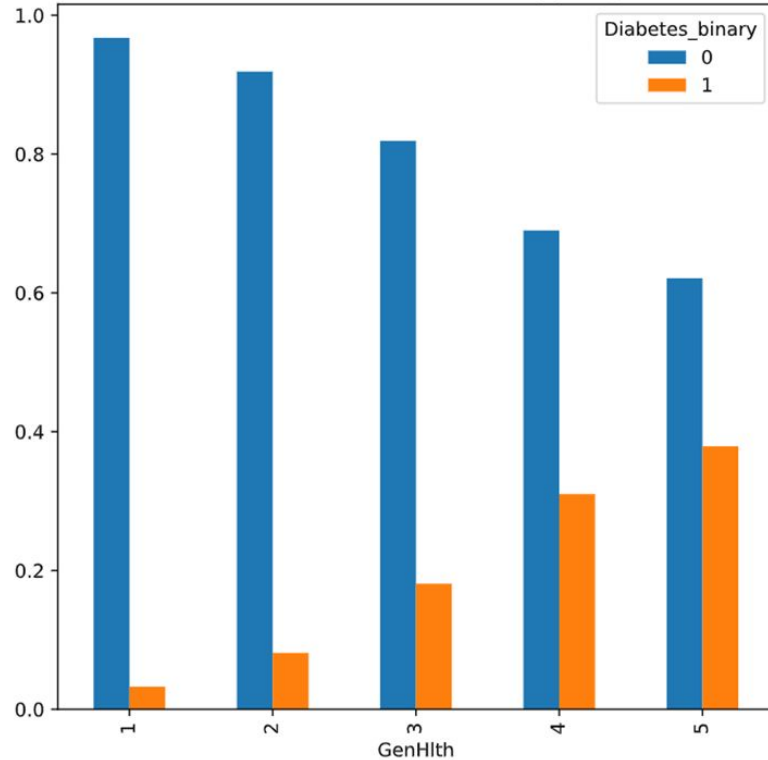


Feature visualization: General Health

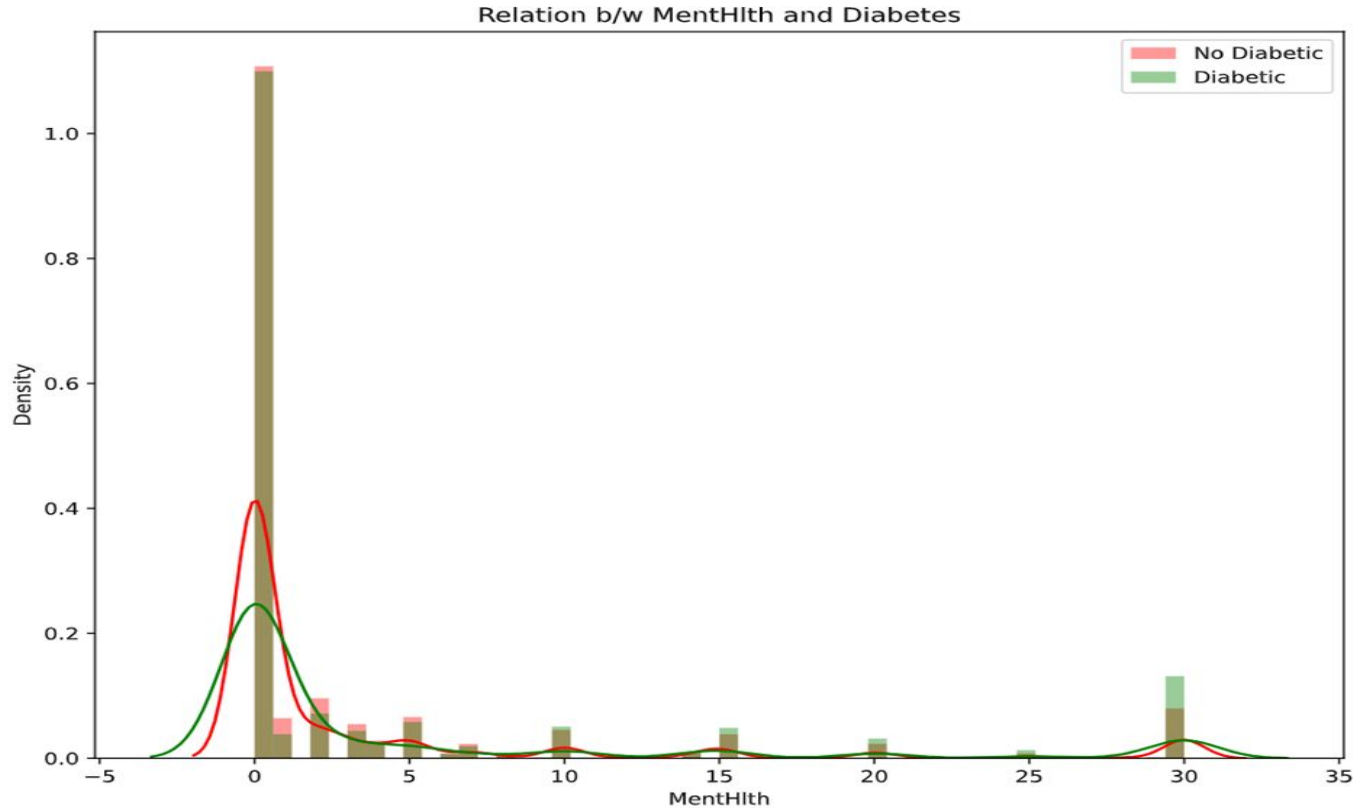


Feature visualization: General Health

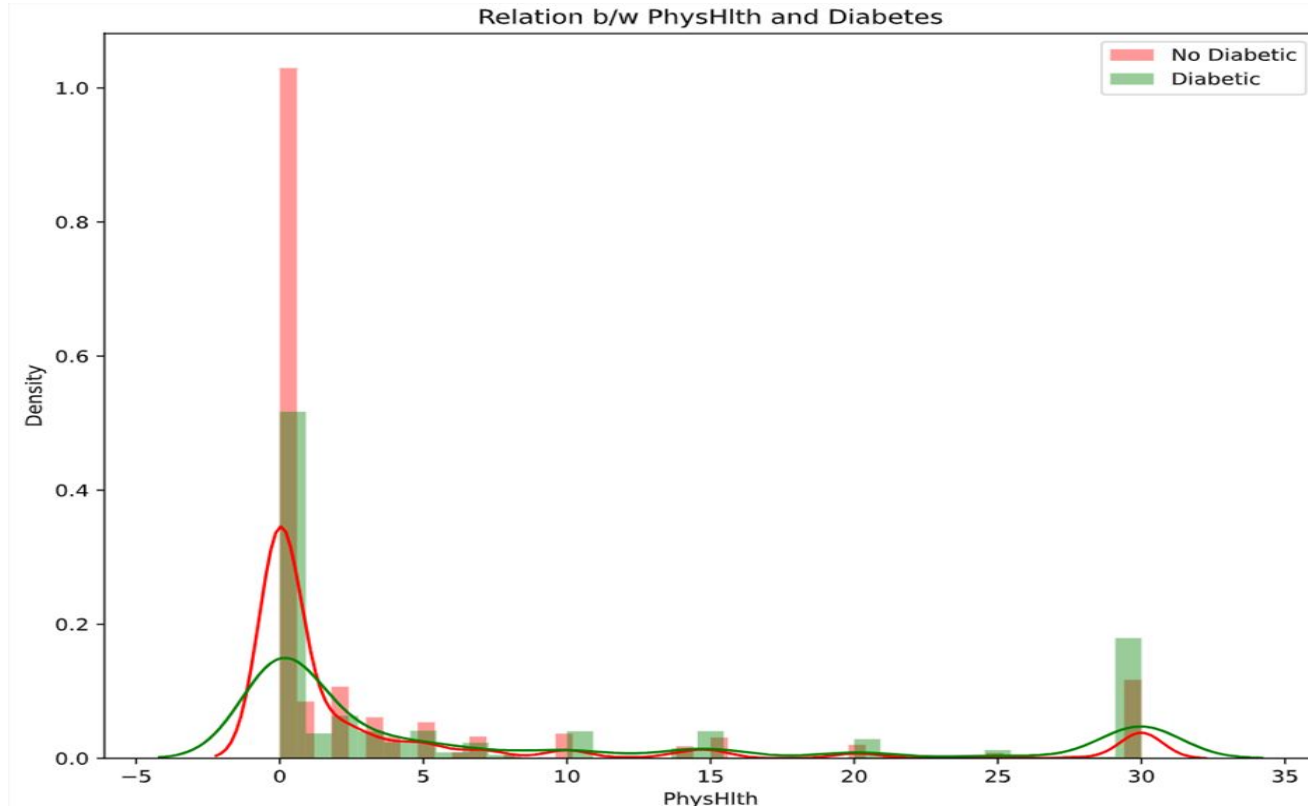
Bar Plot: GenHlth vs Diabetes_binary



Feature visualization: Mental Health

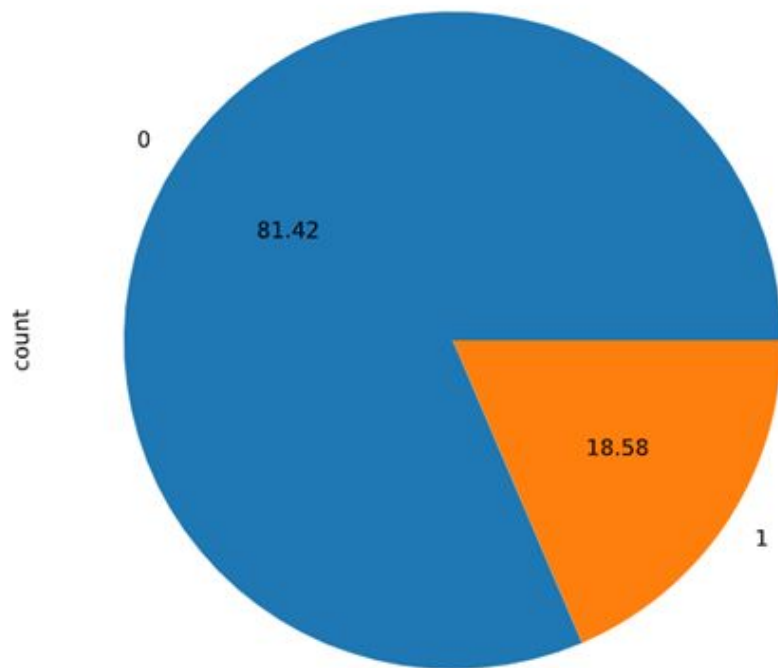


Feature visualization: Physical Health

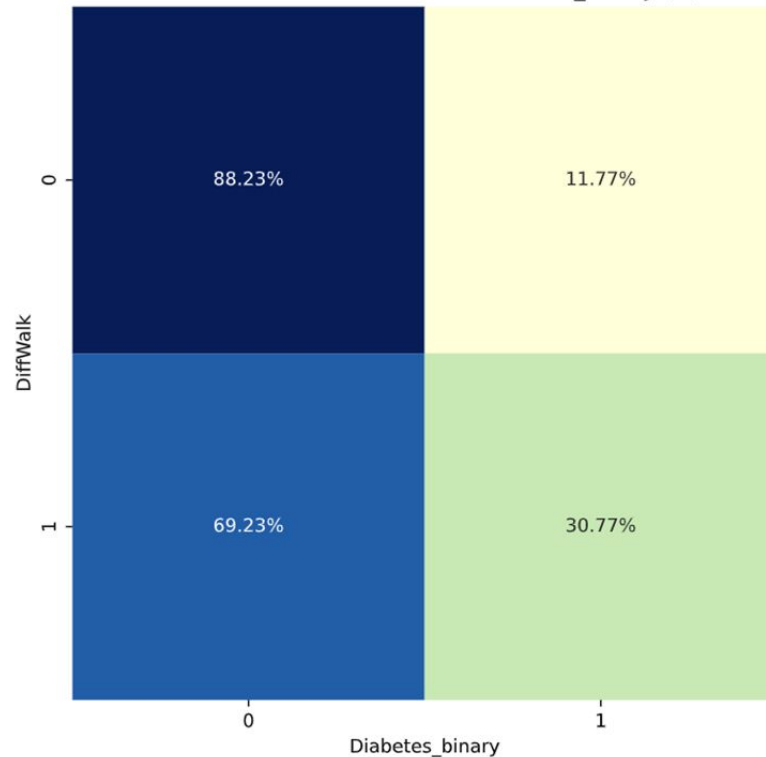


Feature visualization: Difficulty Walking

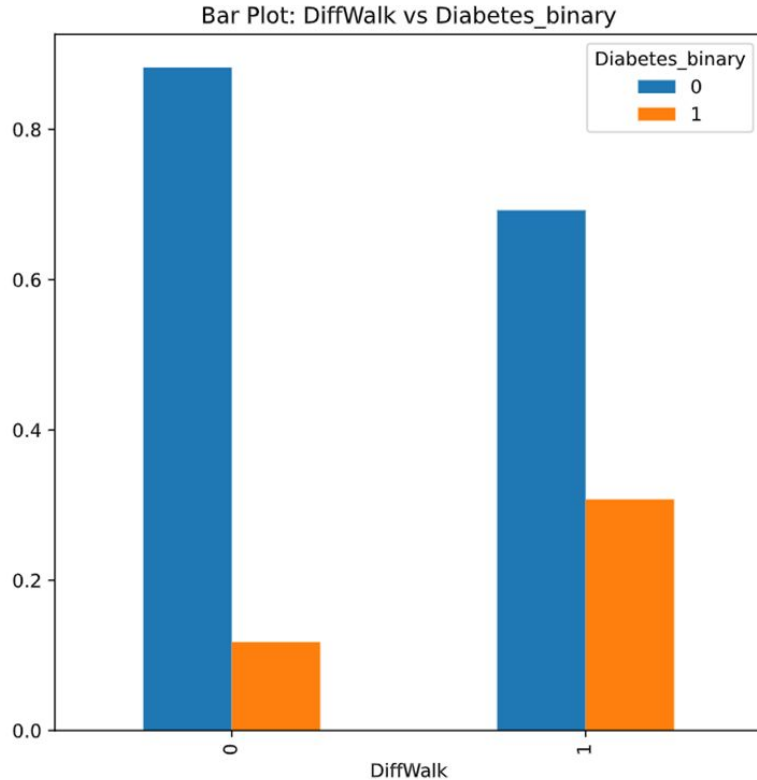
Pie Chart: DiffWalk



Cross-Tabulation: DiffWalk vs Diabetes_binary (%)

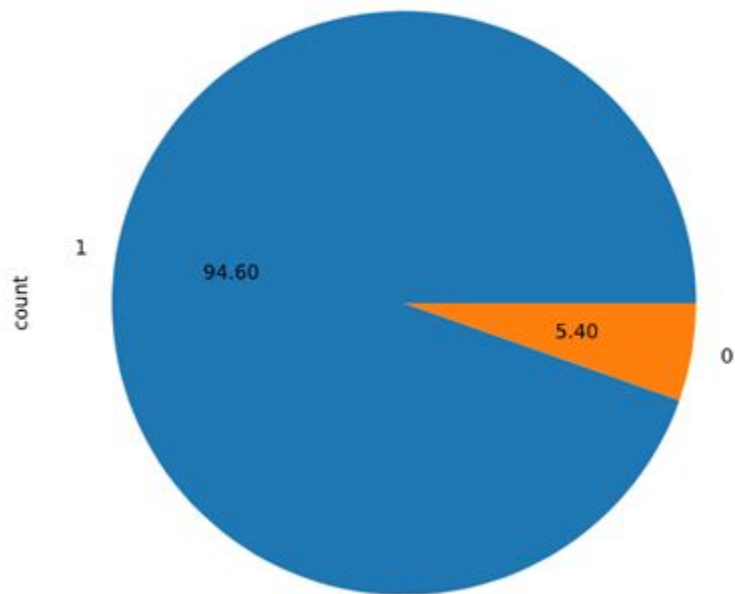


Feature visualization: Difficulty Walking

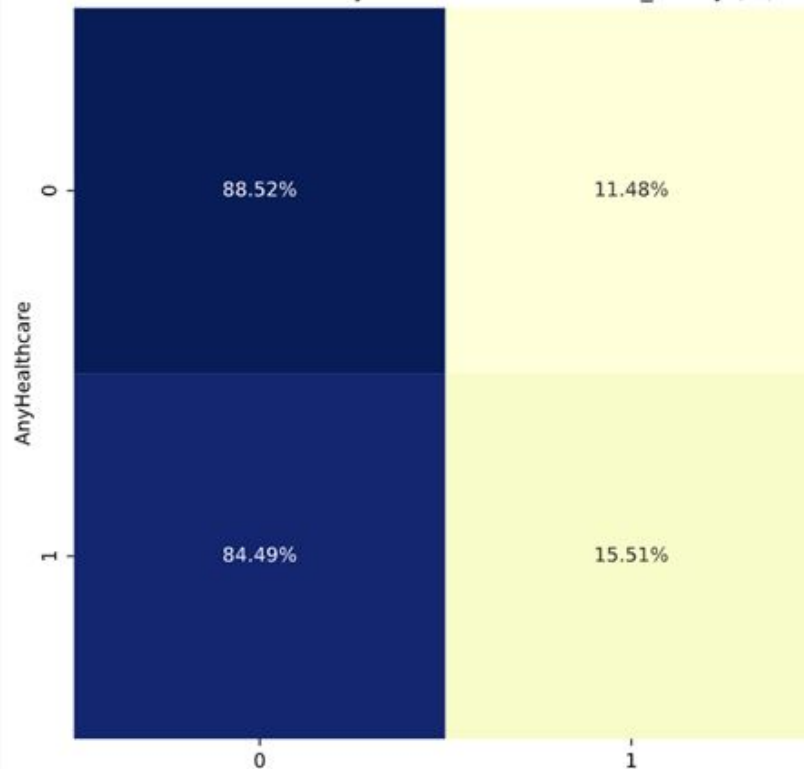


Feature visualization: Access to Healthcare

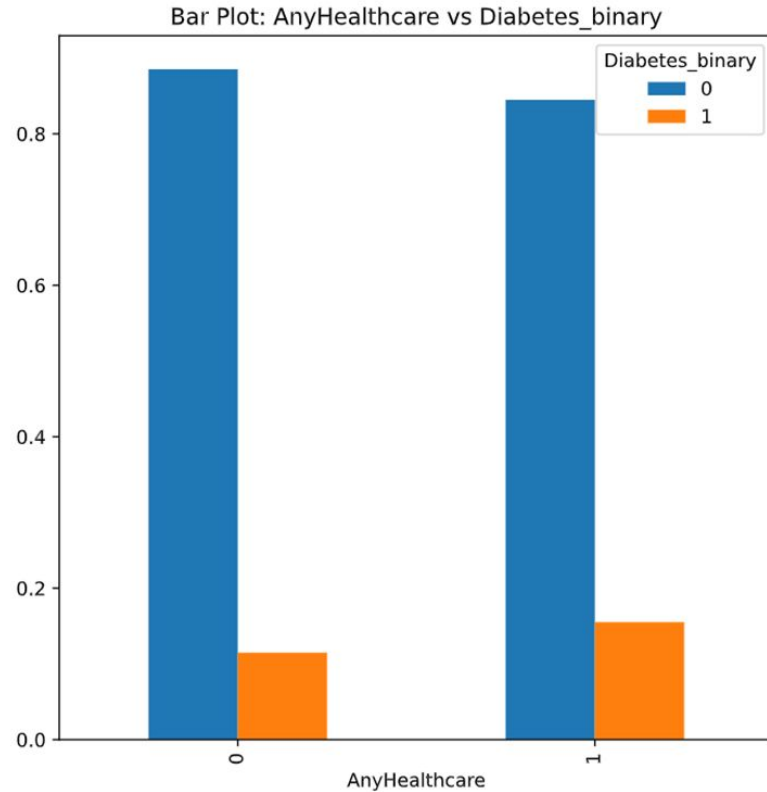
Pie Chart: AnyHealthcare



Cross-Tabulation: AnyHealthcare vs Diabetes_binary (%)

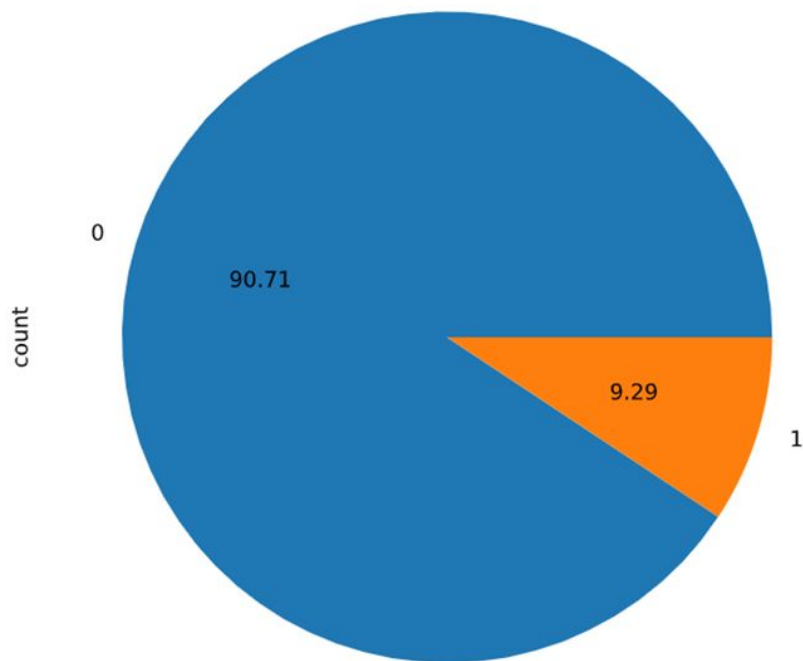


Feature visualization: Access to Healthcare

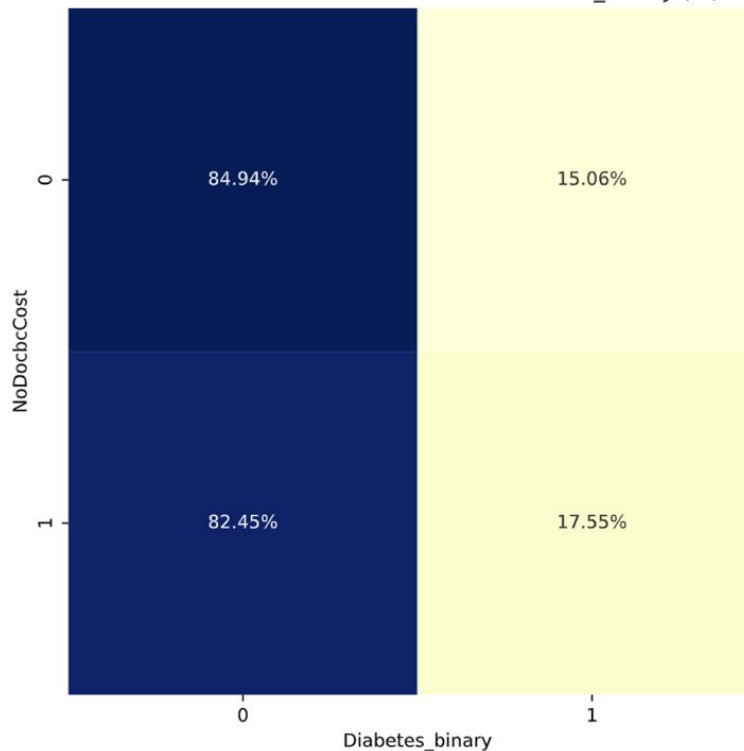


Feature visualization: No access to Doctor due to cost

Pie Chart: NoDocbcCost

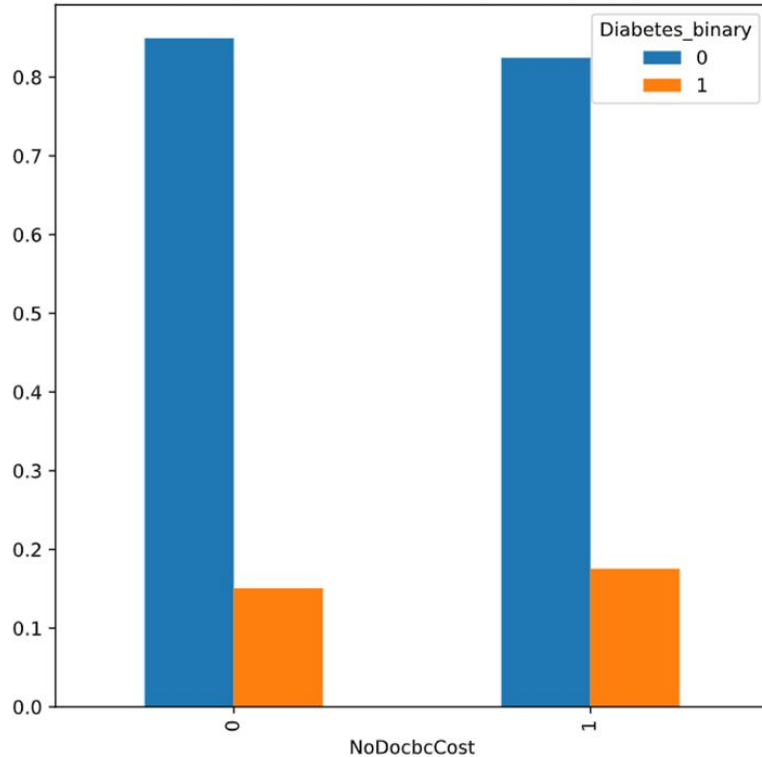


Cross-Tabulation: NoDocbcCost vs Diabetes_binary (%)



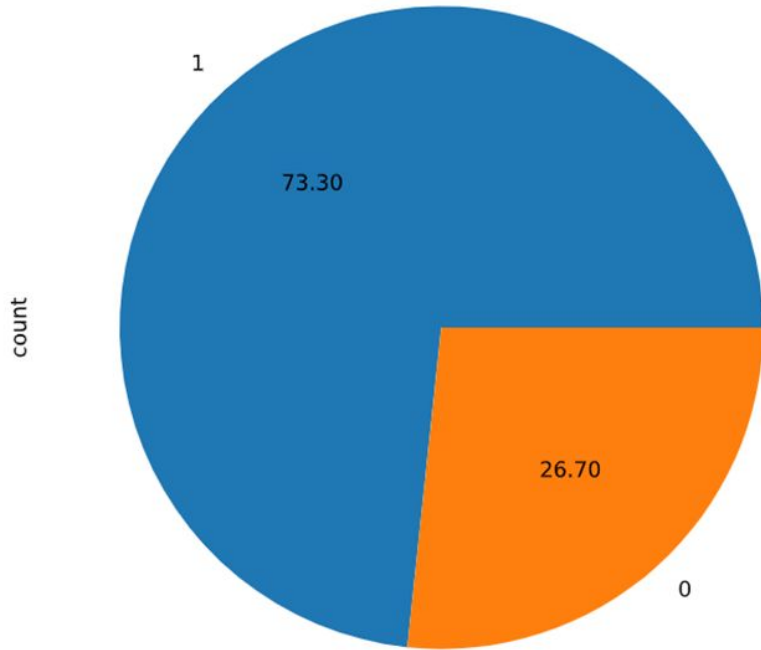
Feature visualization: No access to Doctor due to cost

Bar Plot: NoDocbcCost vs Diabetes_binary

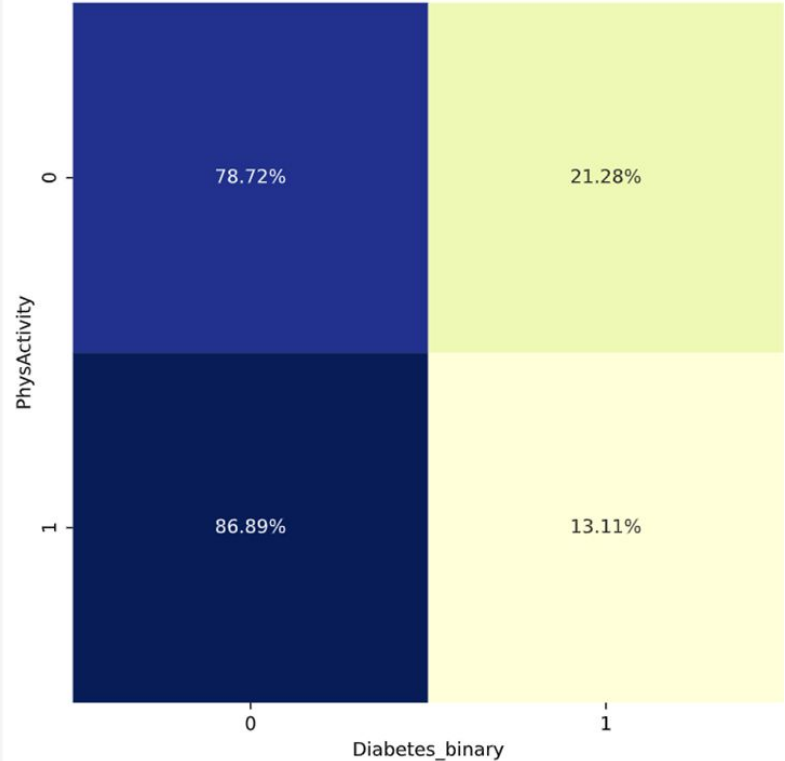


Feature visualization: Physical Exercise

Pie Chart: PhysActivity

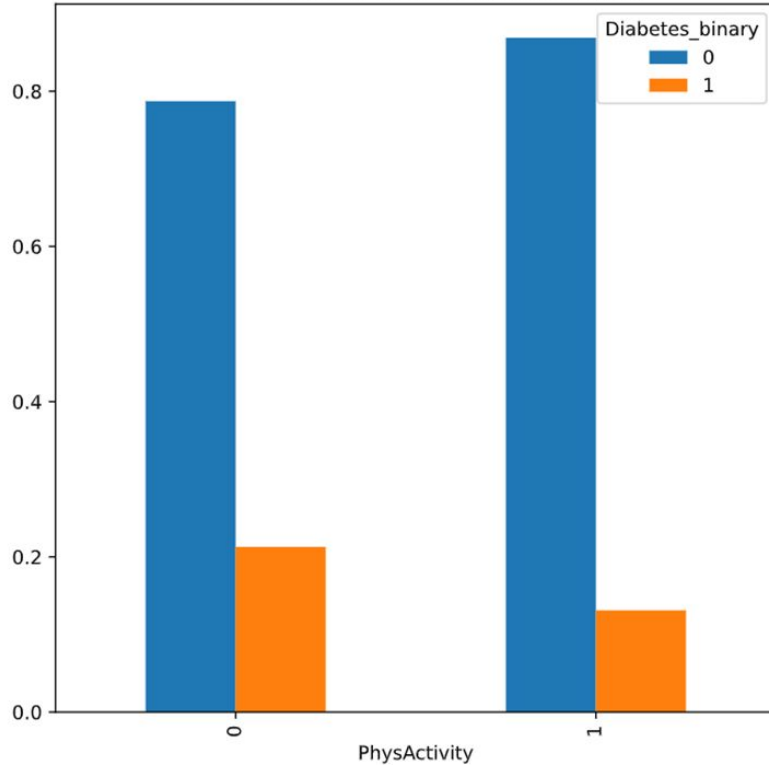


Cross-Tabulation: PhysActivity vs Diabetes_binary (%)



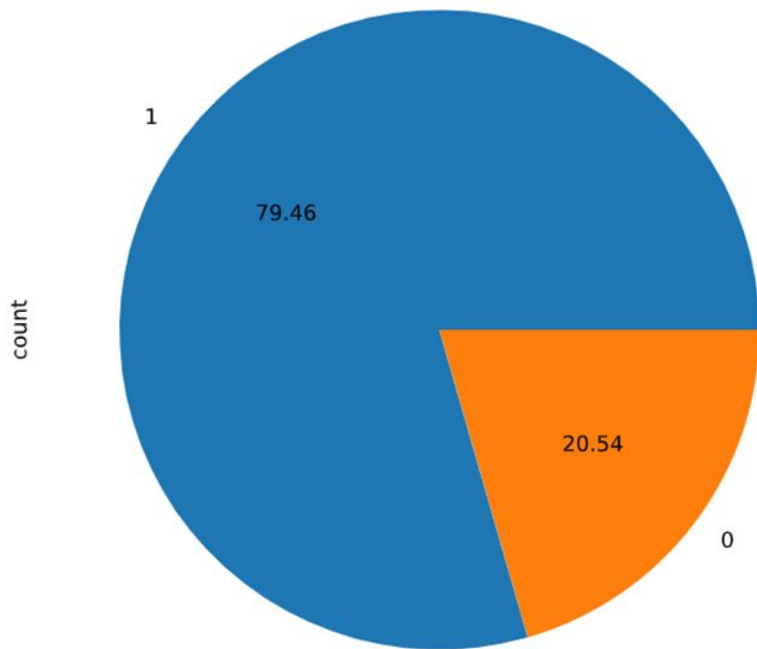
Feature visualization: Physical Exercise

Bar Plot: PhysActivity vs Diabetes_binary

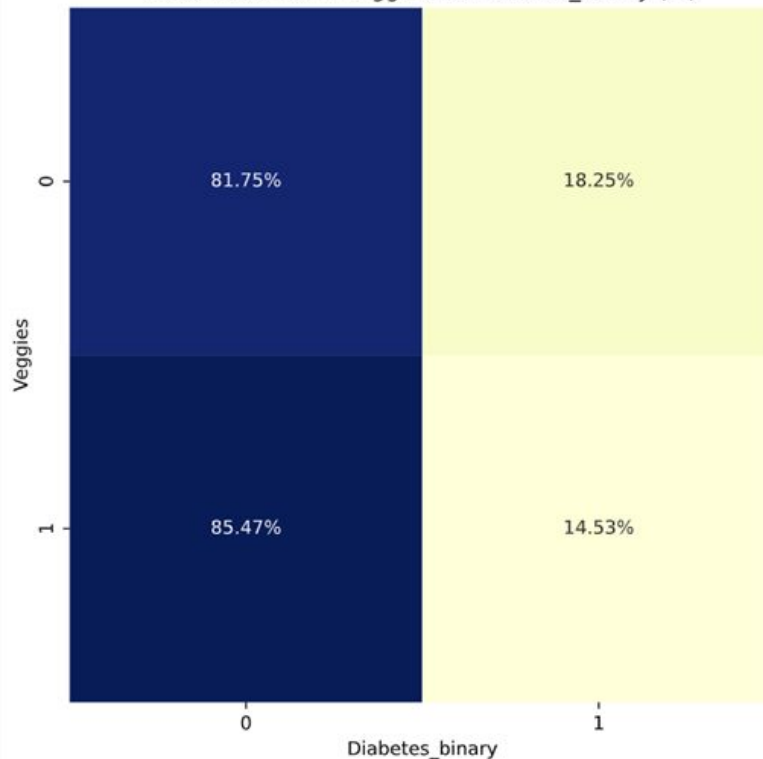


Feature visualization: Veggie Consumption

Pie Chart: Veggies

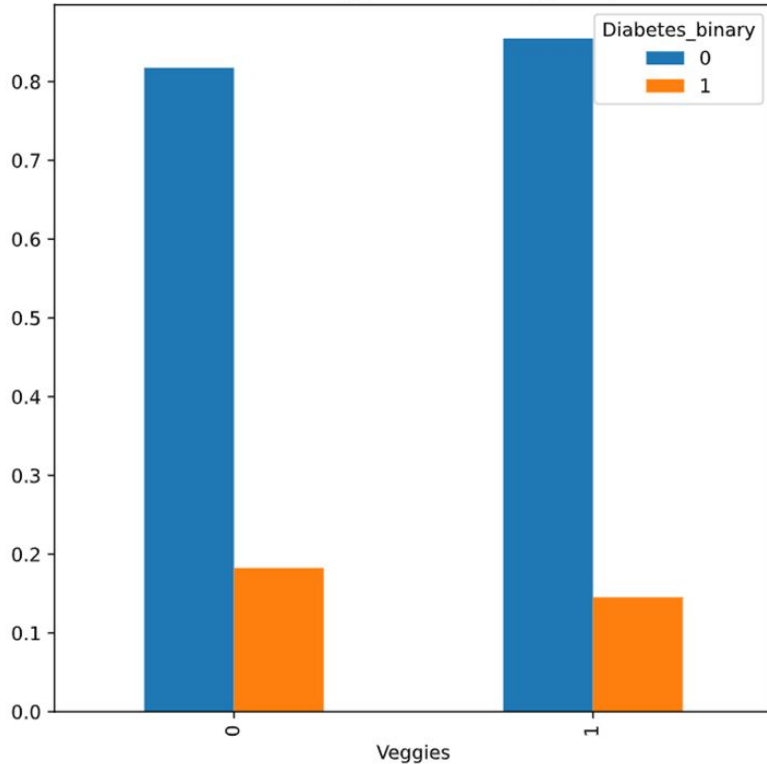


Cross-Tabulation: Veggies vs Diabetes_binary (%)



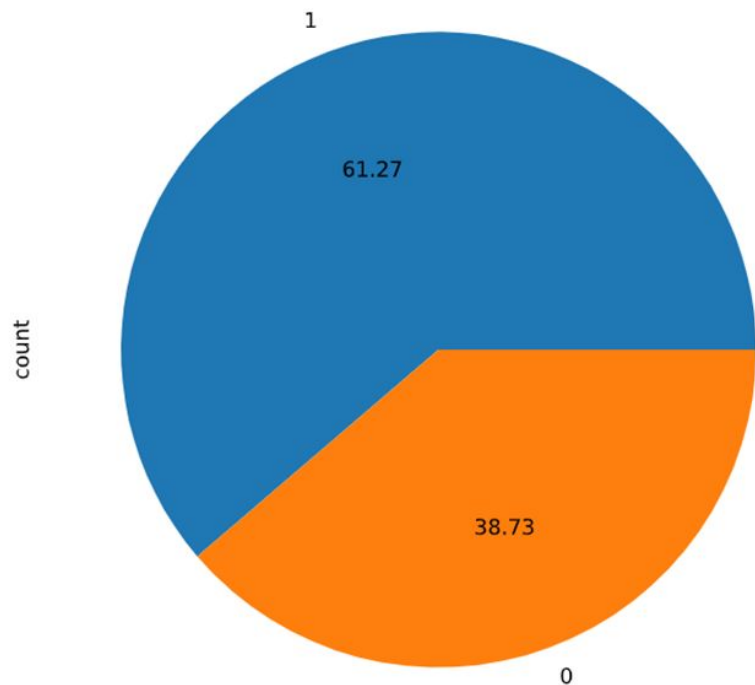
Feature visualization: Veggie Consumption

Bar Plot: Veggies vs Diabetes_binary

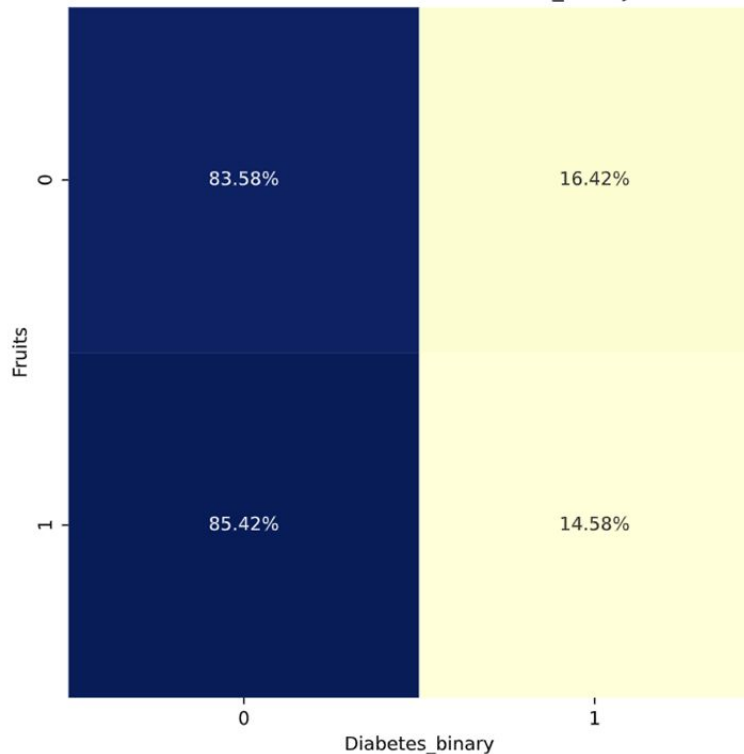


Feature visualization: Fruits Consumption

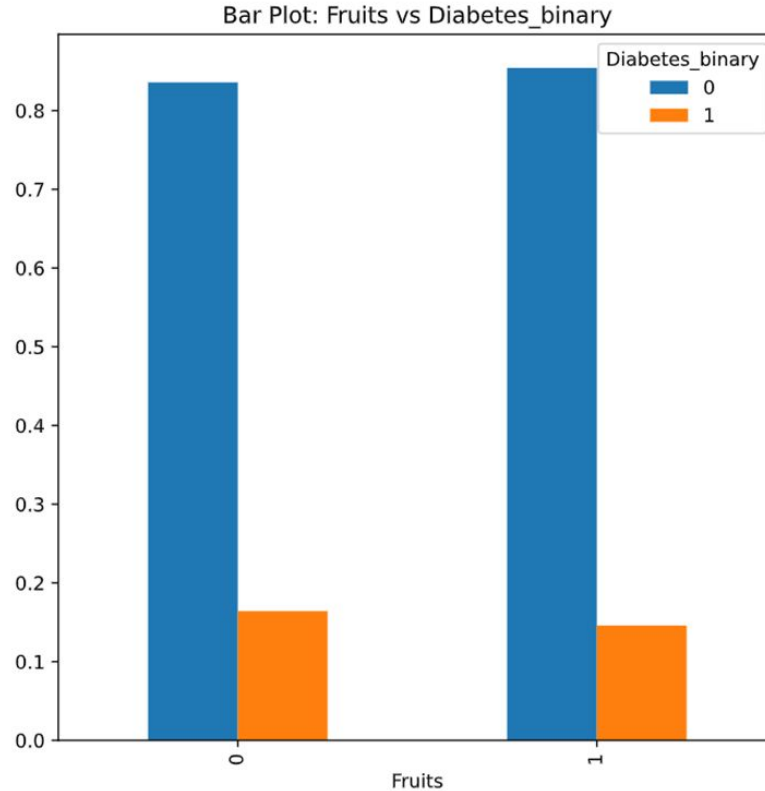
Pie Chart: Fruits



Cross-Tabulation: Fruits vs Diabetes_binary (%)

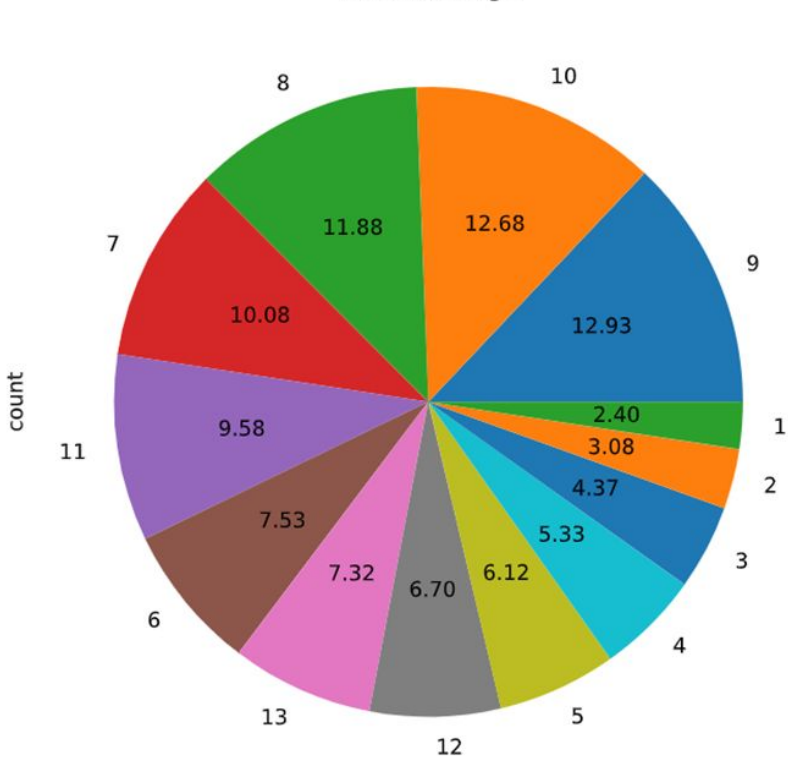


Feature visualization: Fruits Consumption

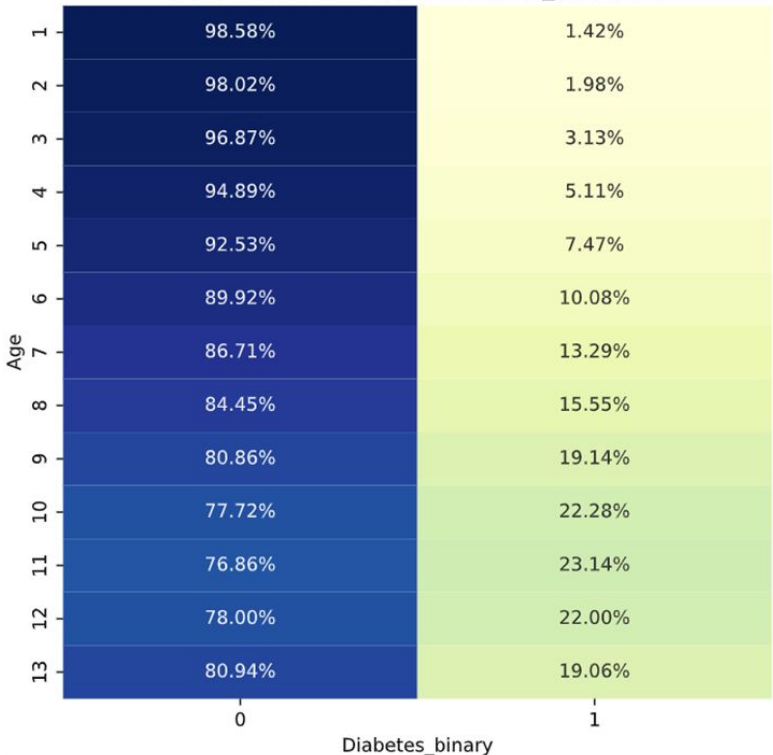


Feature visualization: Age

Pie Chart: Age

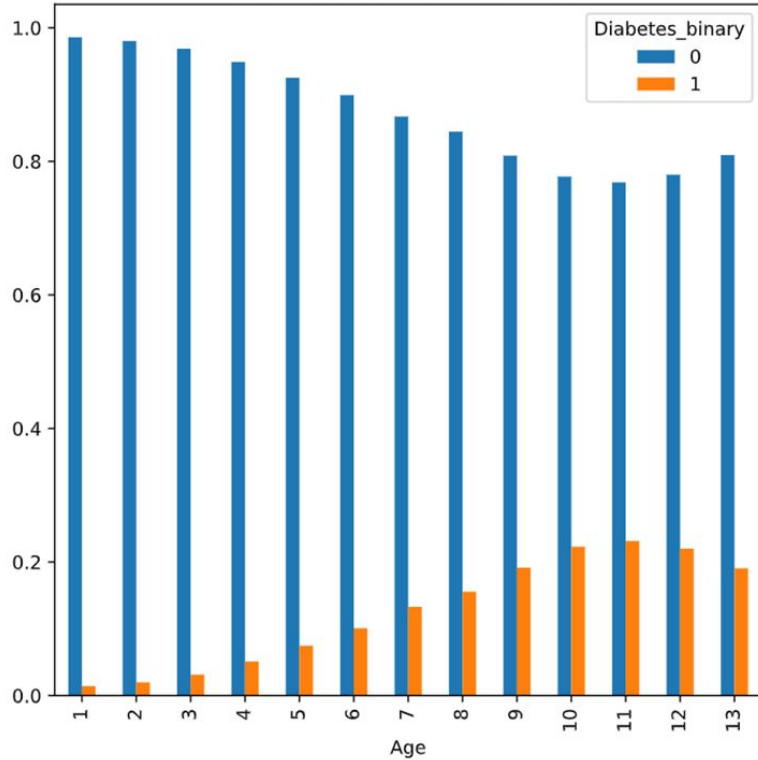


Cross-Tabulation: Age vs Diabetes_binary (%)

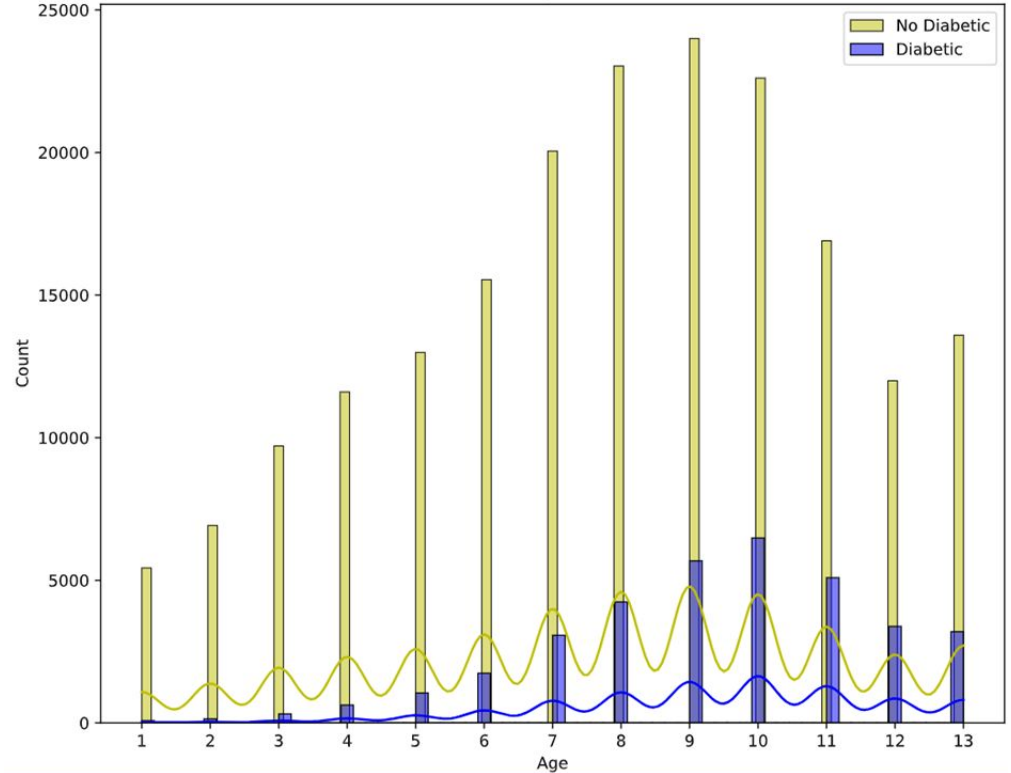


Feature visualization: Age

Bar Plot: Age vs Diabetes_binary

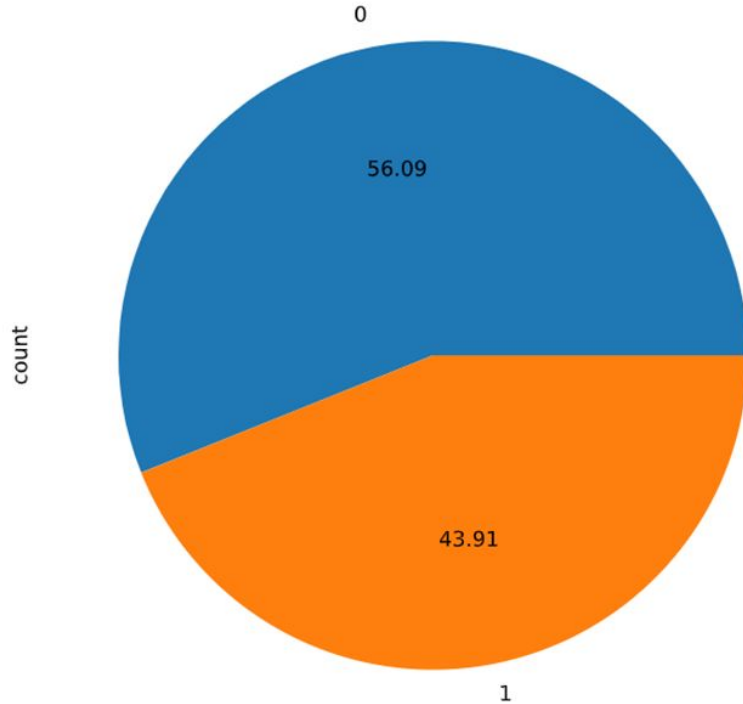


Relation b/w Age and Diabetes

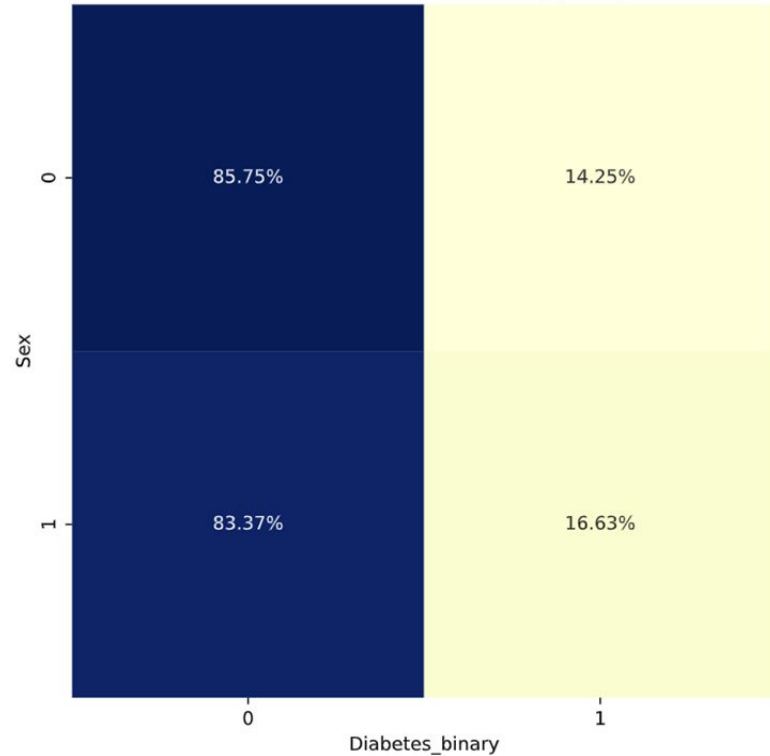


Feature visualization: Sex

Pie Chart: Sex

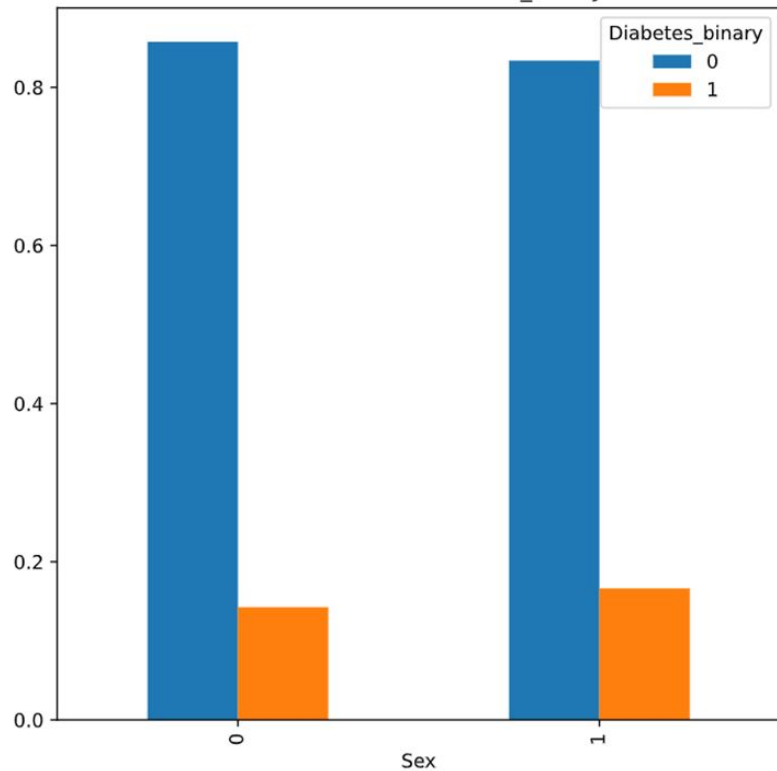


Cross-Tabulation: Sex vs Diabetes_binary (%)

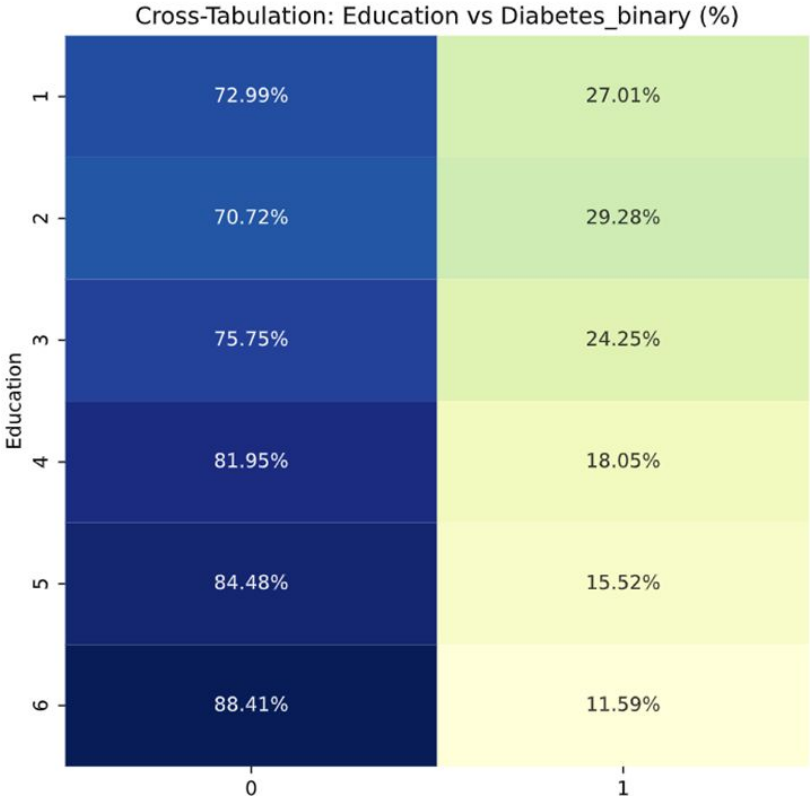
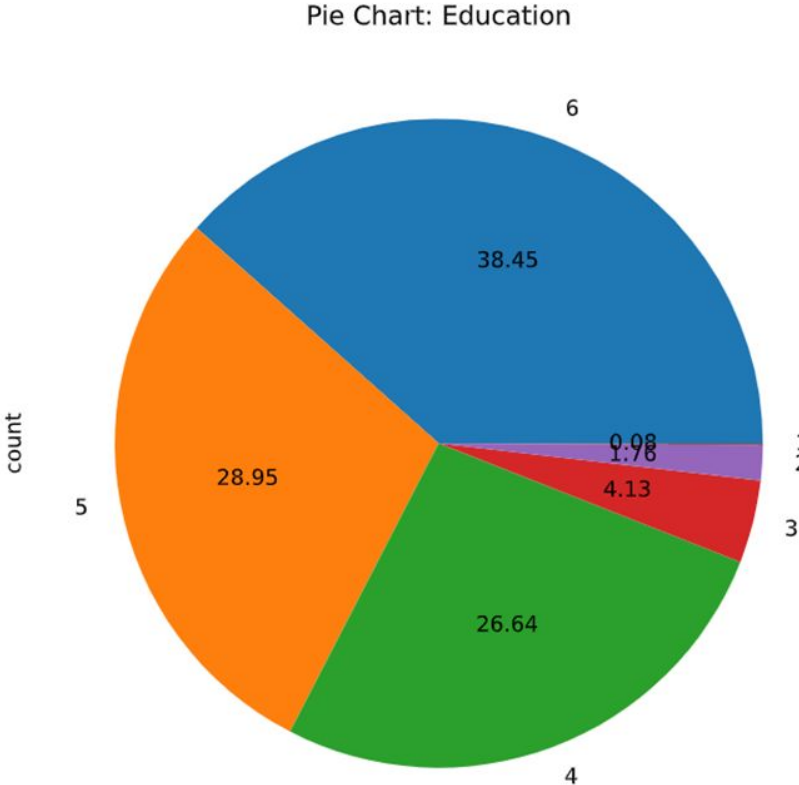


Feature visualization: Sex

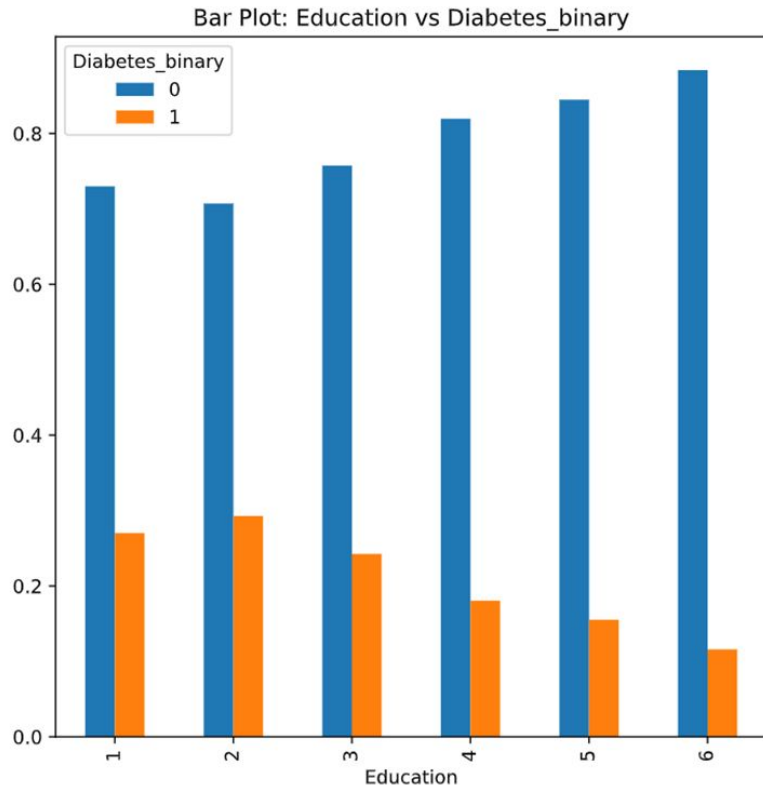
Bar Plot: Sex vs Diabetes_binary



Feature visualization: Education

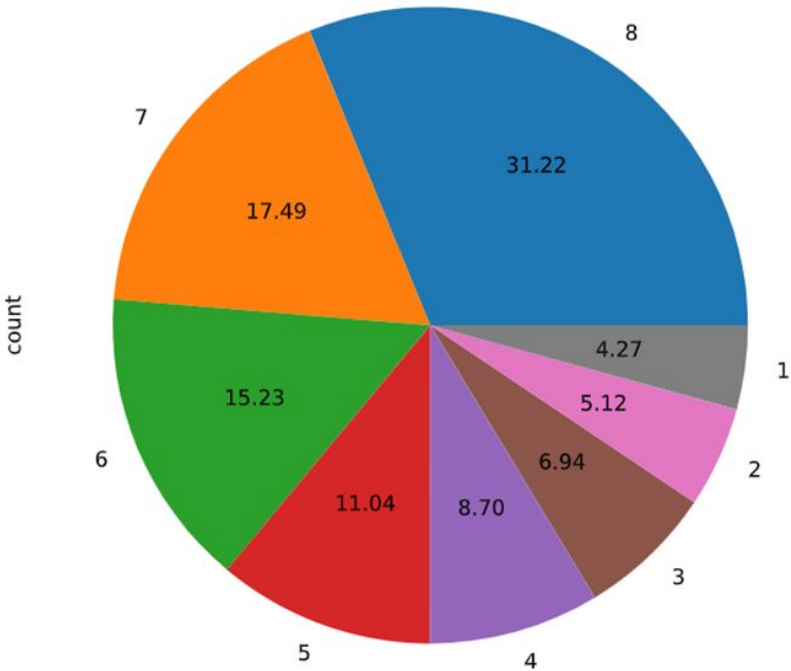


Feature visualization: Education

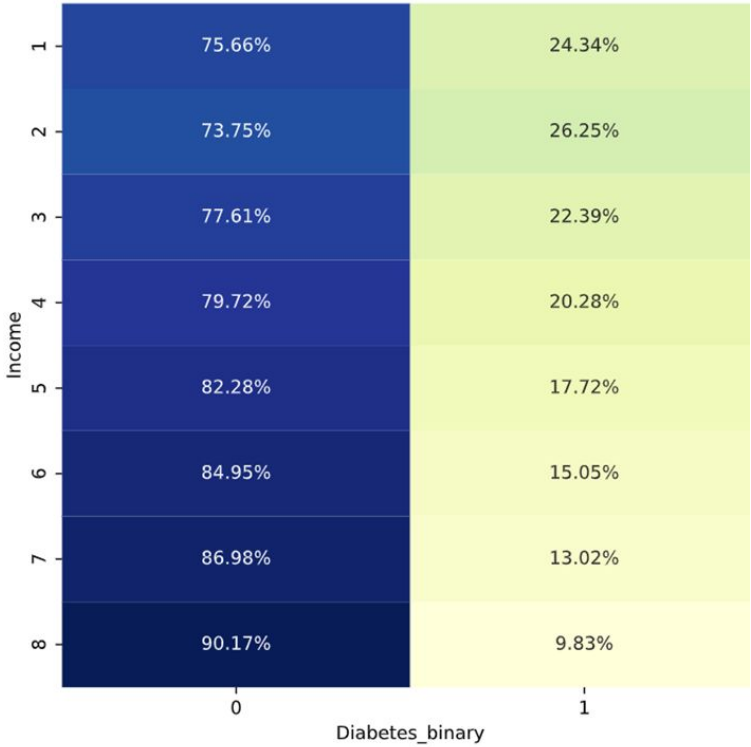


Feature visualization: Income

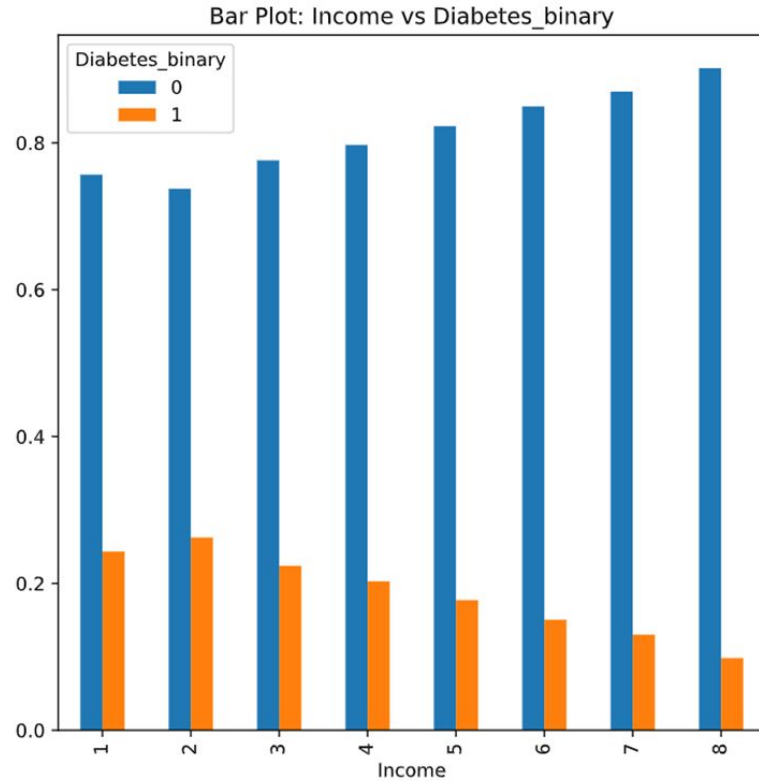
Pie Chart: Income



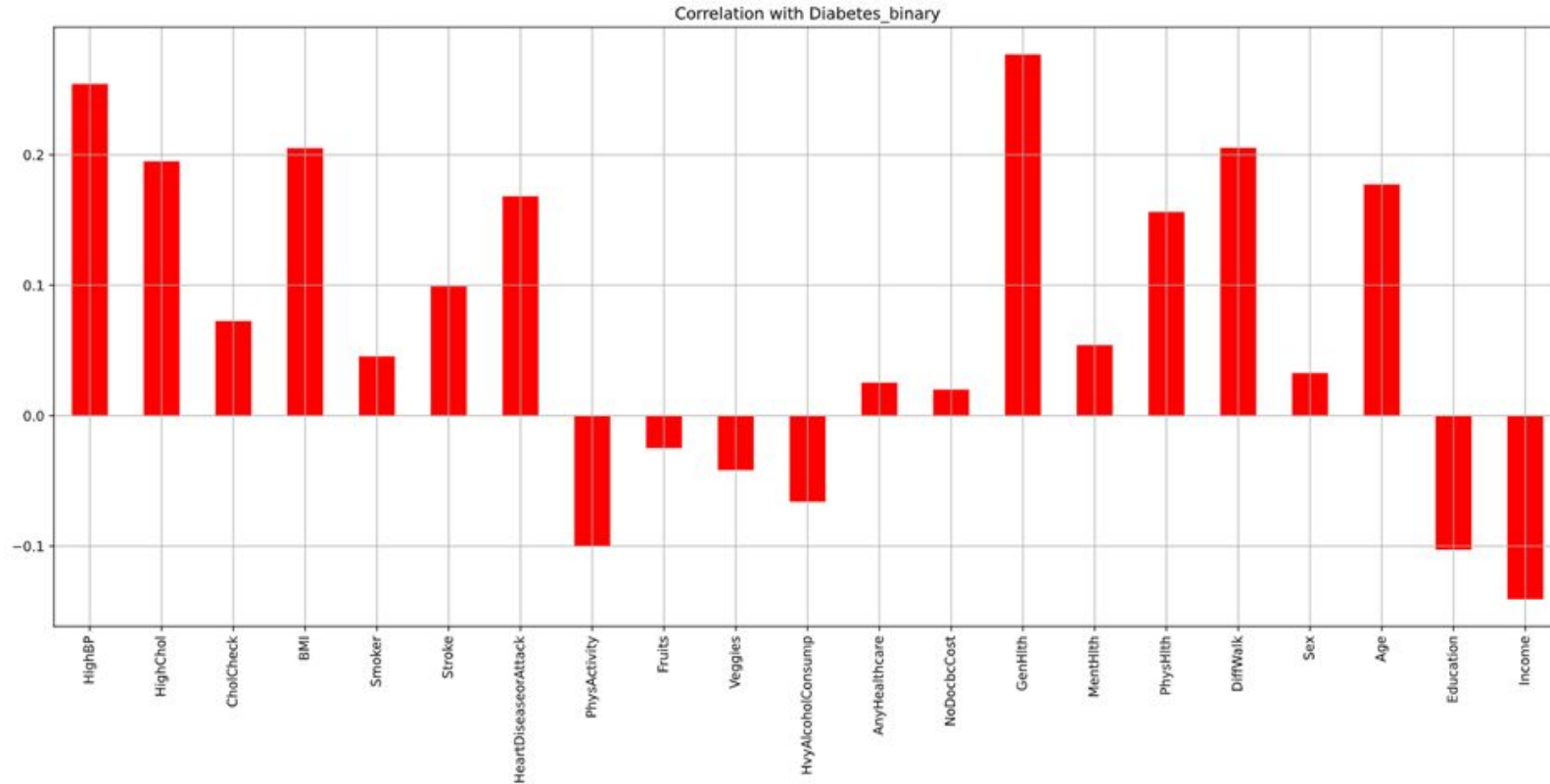
Cross-Tabulation: Income vs Diabetes_binary (%)



Feature visualization: Income



Correlation between features for binary dataset



Results from Lazy Predict Classifier Algorithm

Model	Accuracy	Balanced Accuracy	ROC AUC	F1 Score	\
LGBMClassifier	0.87	0.87	0.87	0.87	
SVC	0.86	0.86	0.86	0.86	
AdaBoostClassifier	0.85	0.85	0.85	0.85	
LogisticRegression	0.85	0.85	0.85	0.85	
CalibratedClassifierCV	0.84	0.84	0.84	0.84	
SGDClassifier	0.84	0.84	0.84	0.84	
LinearSVC	0.84	0.84	0.84	0.84	
RandomForestClassifier	0.83	0.83	0.83	0.83	
ExtraTreesClassifier	0.83	0.83	0.83	0.83	
BaggingClassifier	0.82	0.82	0.82	0.82	
NuSVC	0.82	0.82	0.82	0.82	
GaussianNB	0.81	0.81	0.81	0.81	
QuadraticDiscriminantAnalysis	0.81	0.81	0.81	0.81	
BernoulliNB	0.81	0.81	0.81	0.81	
ExtraTreeClassifier	0.81	0.81	0.81	0.81	
KNeighborsClassifier	0.81	0.81	0.81	0.81	
DecisionTreeClassifier	0.81	0.81	0.81	0.81	
Perceptron	0.81	0.81	0.81	0.81	
LinearDiscriminantAnalysis	0.80	0.80	0.80	0.80	
RidgeClassifier	0.80	0.80	0.80	0.80	
RidgeClassifierCV	0.80	0.80	0.80	0.80	
NearestCentroid	0.79	0.80	0.80	0.79	
XGBClassifier	0.74	0.73	0.73	0.73	
PassiveAggressiveClassifier	0.73	0.73	0.73	0.73	
DummyClassifier	0.50	0.50	0.50	0.33	

Model	Time Taken
LGBMClassifier	1.15
SVC	140.87
AdaBoostClassifier	2.55
LogisticRegression	0.25
CalibratedClassifierCV	2.40
SGDClassifier	0.41
LinearSVC	0.59
RandomForestClassifier	9.24
ExtraTreesClassifier	11.17
BaggingClassifier	2.47
NuSVC	292.70
GaussianNB	0.13
QuadraticDiscriminantAnalysis	0.17
BernoulliNB	0.23
ExtraTreeClassifier	0.18
KNeighborsClassifier	4.72
DecisionTreeClassifier	0.42
Perceptron	0.15
LinearDiscriminantAnalysis	0.28
RidgeClassifier	0.17
RidgeClassifierCV	0.25
NearestCentroid	0.10
XGBClassifier	0.86
PassiveAggressiveClassifier	0.22
DummyClassifier	0.08

LightGBM Model Hyper Parameter Tuning

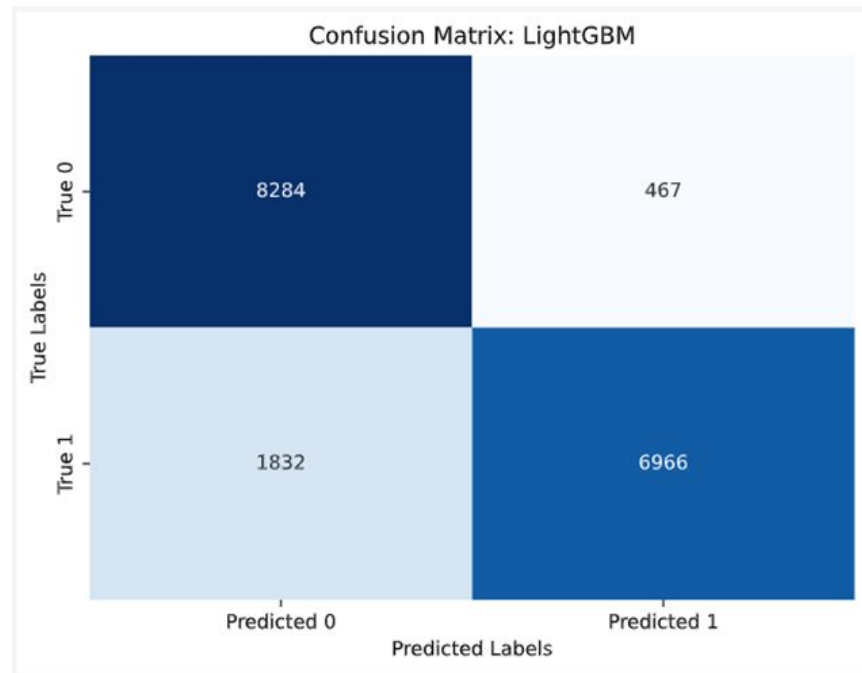
LightGBM Model Results

Accuracy: 0.8689953843523848

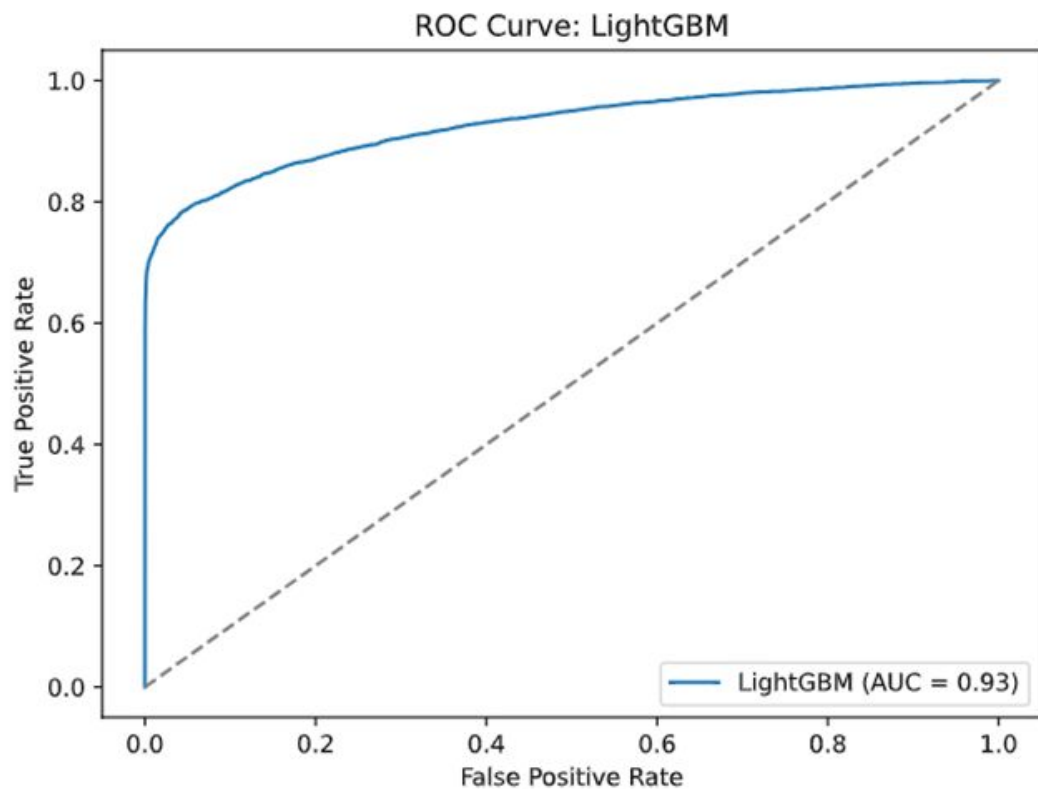
Time Taken: 1.25 seconds

Classification Report:

	precision	recall	f1-score	support
0	0.82	0.95	0.88	8751
1	0.94	0.79	0.86	8798
accuracy			0.87	17549
macro avg	0.88	0.87	0.87	17549
weighted avg	0.88	0.87	0.87	17549



LightGBM Model Hyper Parameter Tuning



AdaBoost Model Hyper Parameter Tuning

AdaBoost Model Results

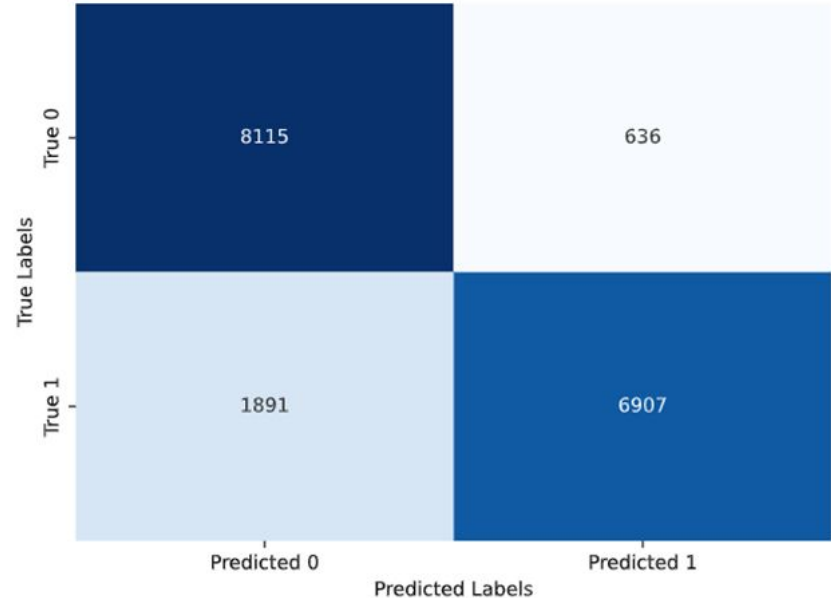
Accuracy: 0.856003191065018

Time Taken: 7.77 seconds

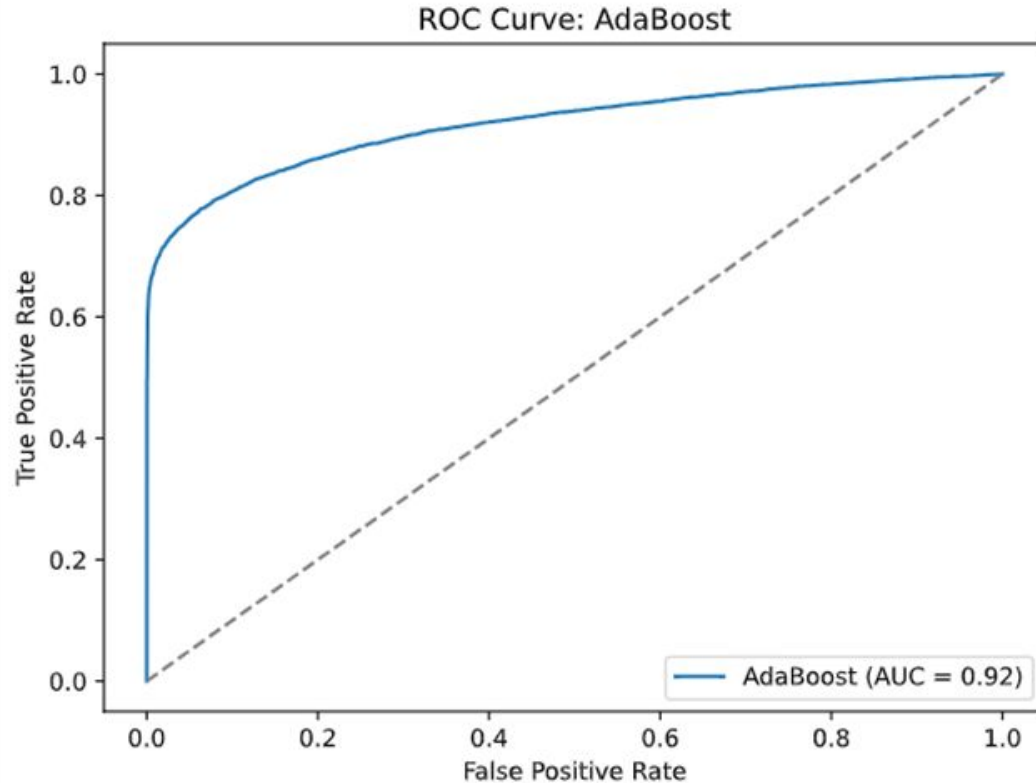
Classification Report:

	precision	recall	f1-score	support
0	0.81	0.93	0.87	8751
1	0.92	0.79	0.85	8798
accuracy			0.86	17549
macro avg	0.86	0.86	0.86	17549
weighted avg	0.86	0.86	0.86	17549

Confusion Matrix: AdaBoost



AdaBoost Model Hyper Parameter Tuning



Logistic Regression Hyper Parameter Tuning

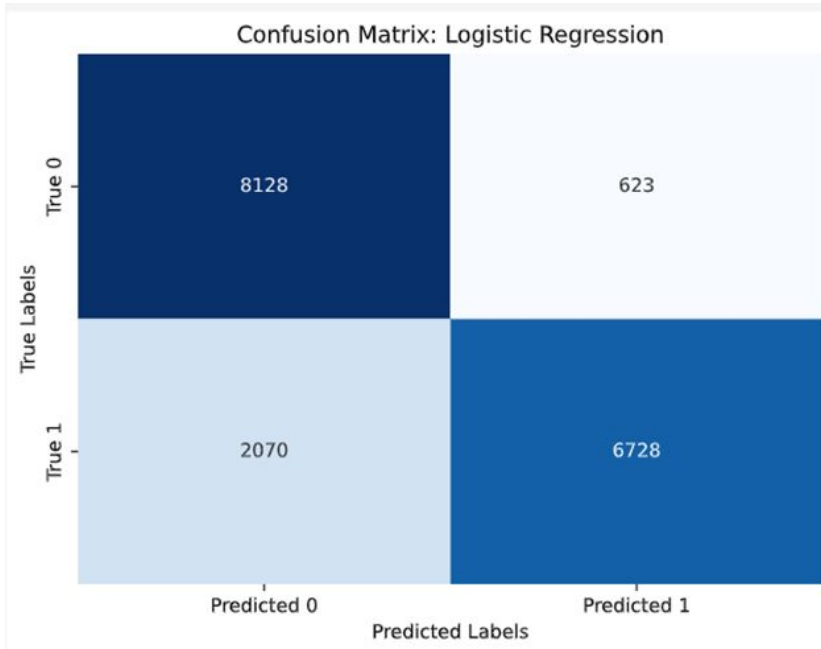
Logistic Regression Model Results

Accuracy: 0.8465439626189526

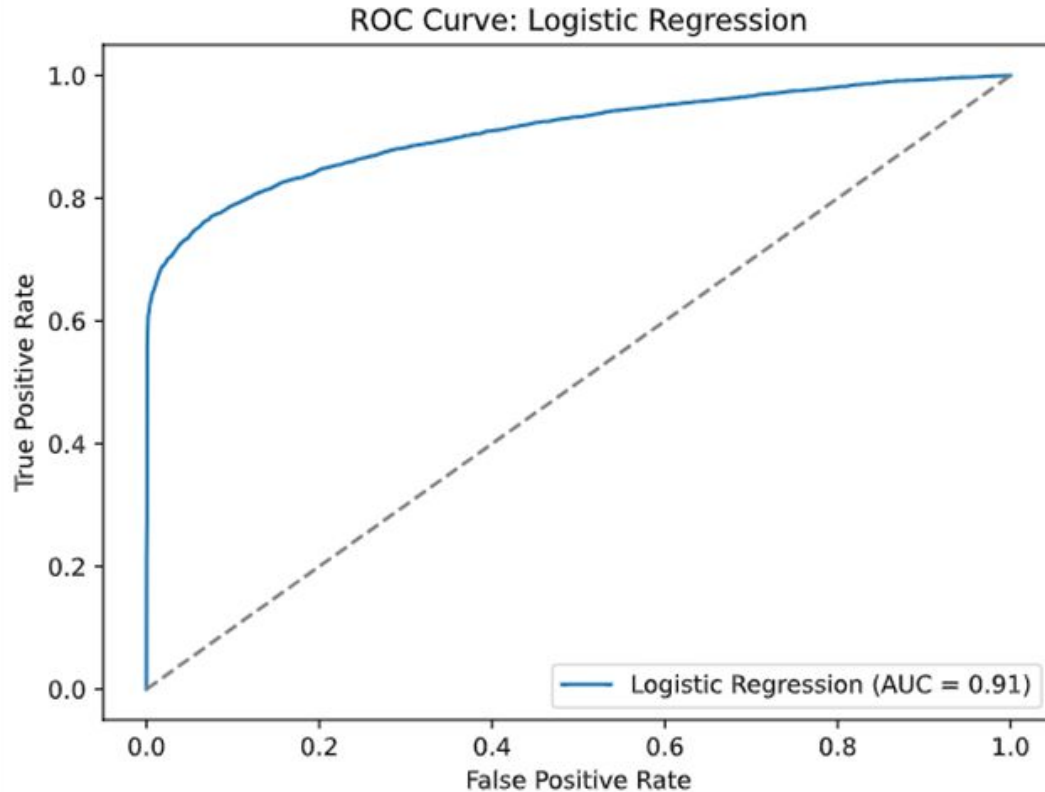
Time Taken: 0.71 seconds

Classification Report:

	precision	recall	f1-score	support
0	0.80	0.93	0.86	8751
1	0.92	0.76	0.83	8798
accuracy			0.85	17549
macro avg	0.86	0.85	0.85	17549
weighted avg	0.86	0.85	0.85	17549



Logistic Regression Hyper Parameter Tuning



Final Model Evaluation

LGBMClassifier evaluation:

	precision	recall	f1-score	support
0	0.82	0.94	0.88	8751
1	0.93	0.79	0.86	8798
accuracy			0.87	17549
macro avg	0.88	0.87	0.87	17549
weighted avg	0.88	0.87	0.87	17549

Extra: Logistic Regression Model Evaluation

Accuracy: 0.8464869793150607

Confusion Matrix:

[[8127 624]

[2070 6728]]

Classification Report:

	precision	recall	f1-score	support
0	0.80	0.93	0.86	8751
1	0.92	0.76	0.83	8798
accuracy			0.85	17549
macro avg	0.86	0.85	0.85	17549
weighted avg	0.86	0.85	0.85	17549