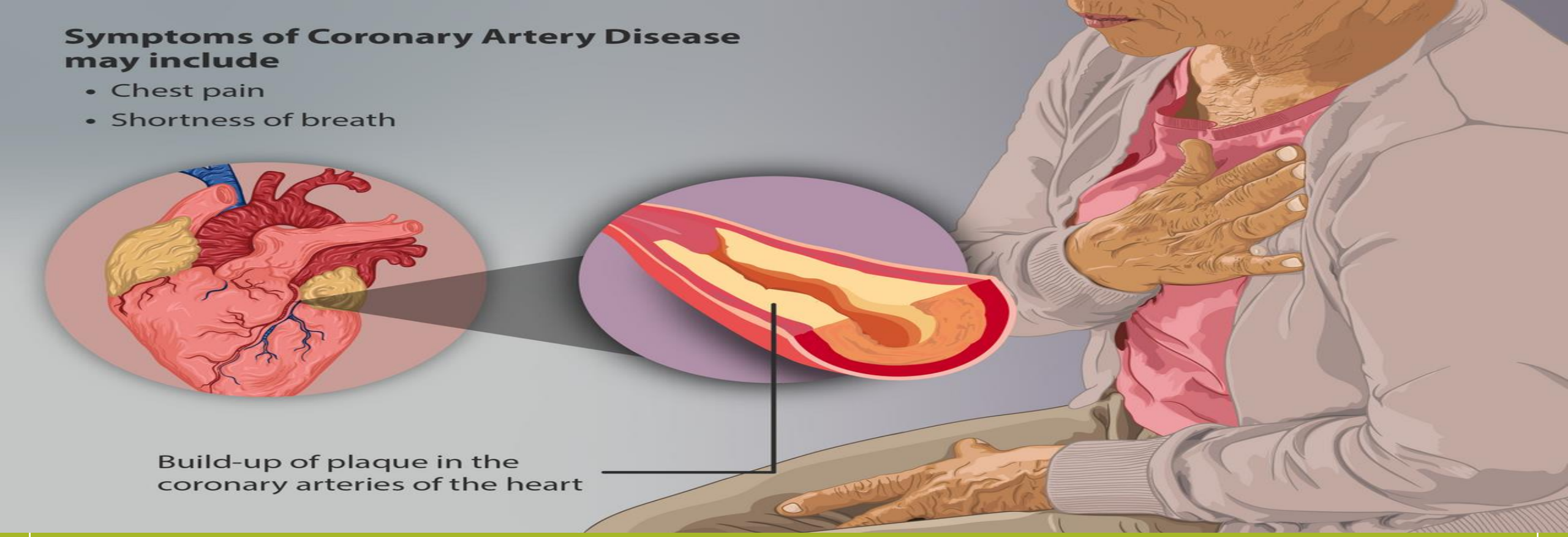


CARDIOVASCULAR HEALTH ASSESSMENT AND RISK PREDICTION

Goal : To predict weather patient has 10 year risk of Coronary Heart Disease(CHD)

01

Introduction

Why this project is and how machine learning can helps in achieving our goal.

02

Executive summary

Summary of the Project that is all explanation in one page.

03

EDA on important features

In this section we will analyse the dataset

04

ML Model Details

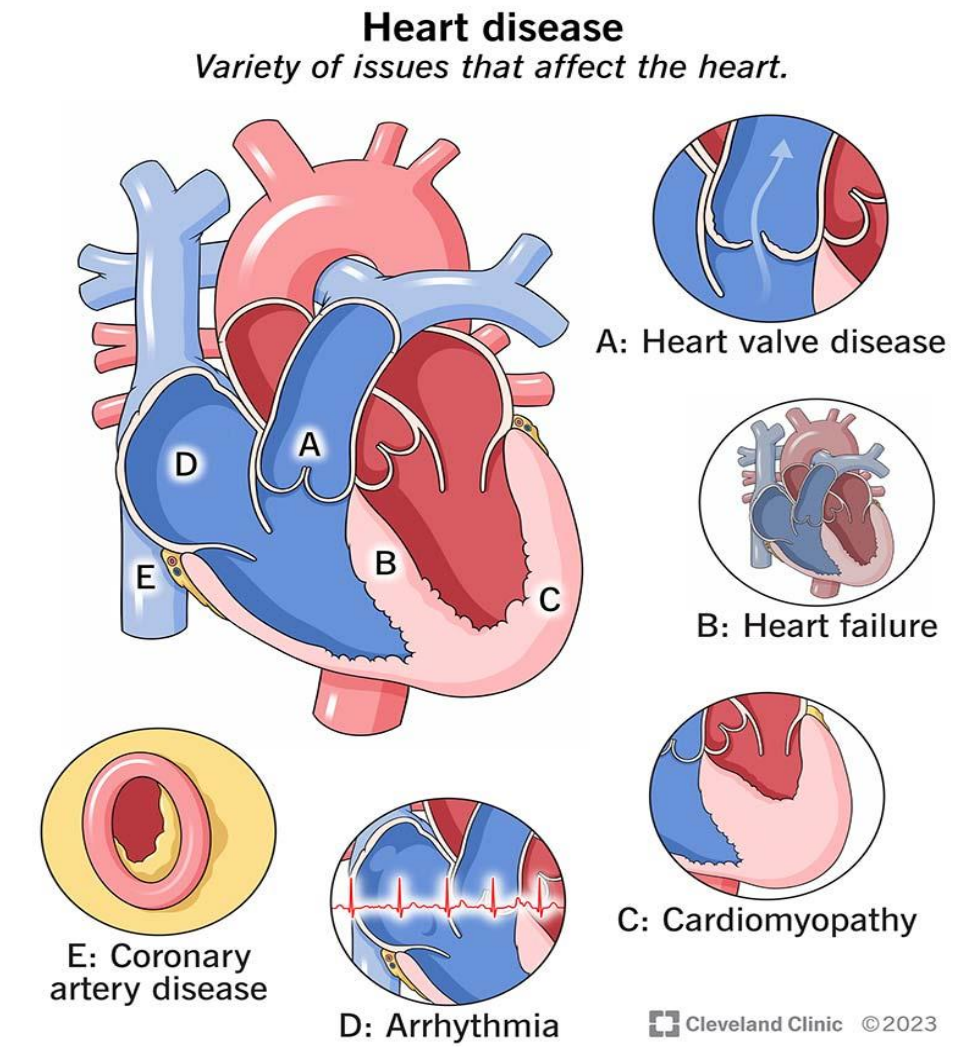
This section, we will know about ml model and its results

05

Conclusion

Recommendation about project

Introduction

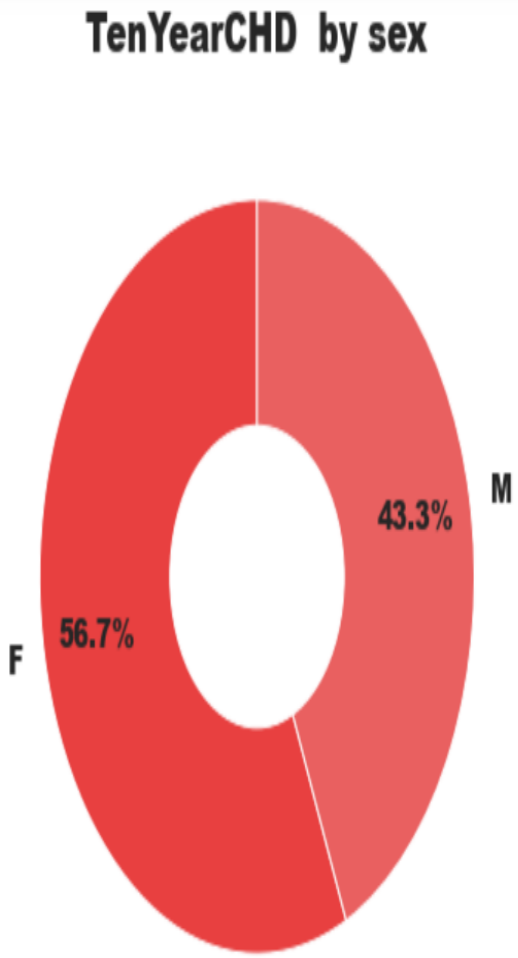


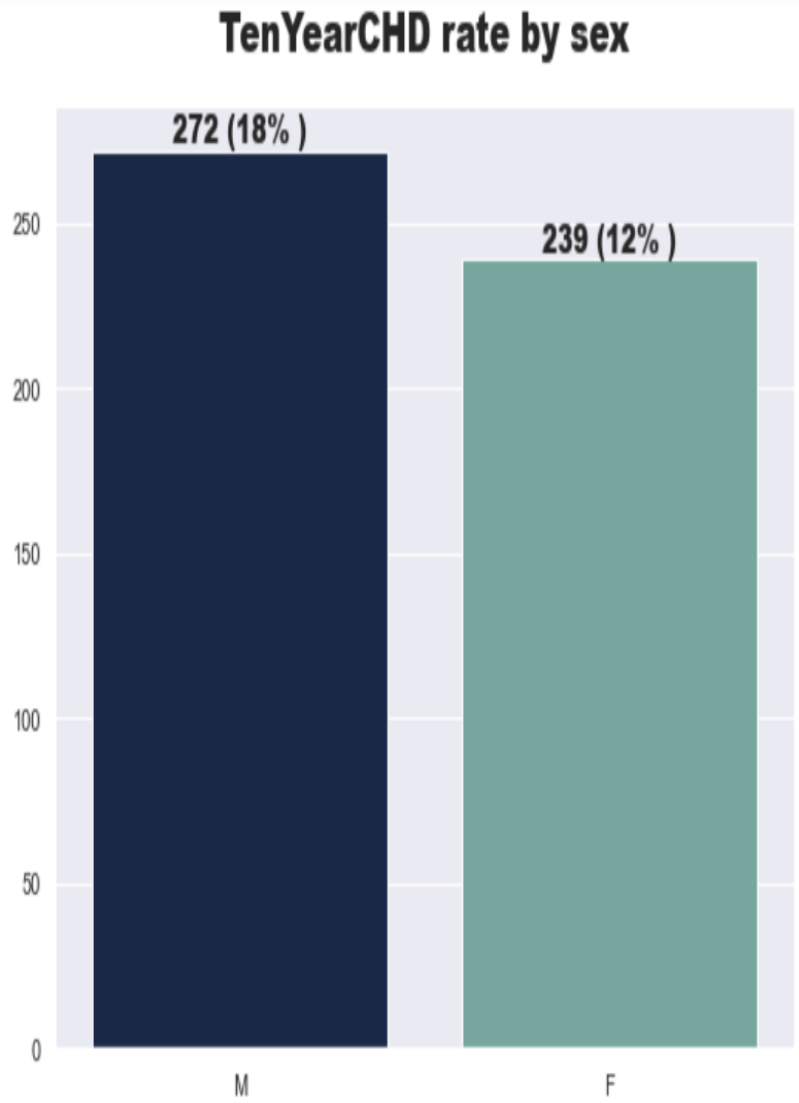
- **Coronary Heart Disease (CHD)** involves the reduction of blood flow to heart muscle due to build-up of plaque in the arteries of the heart.
- The prediction of heart disease is considered one of the most important topics in health domain. With the machine learning algorithms and having large amounts of data, it is possible to use these information that can help doctors make more accurate predictions.
- Prediction of **CHD** is a much complex challenge considering the level of expertise and knowledge required for accurate result. According to **a survey by WHO,** medical professionals can correctly predict heart disease with only **67% accuracy.**
- In this project, a number of independent variables such as **sex, age, cigsPerDay, totChol, sysBP** and **glucose** will be used along with a dependent variable (**TenYearCHD** class) during the training phase to build a classification model. **The classification goal** is to predict whether the patient has 10-year risk of future C*oronary Heart Disease* (CHD) or not.

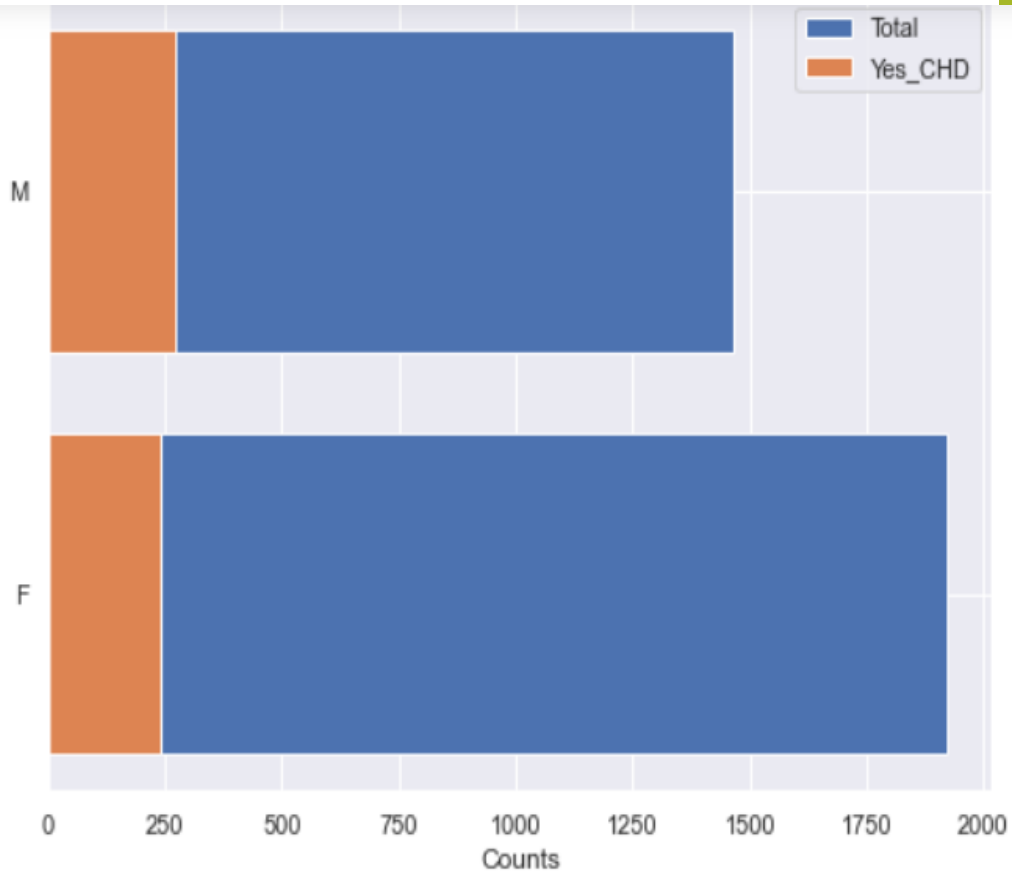
Executive Summary

- There are 16 independent variables and 1 dependent variable in the dataset. Feature id which does not have any contribution in modelling was dropped.
- There are 7 features have missing values, out of these 7, 5 numerical features are imputed with median values of those features as there are outliers in all the 5 numerical features. 2 categorical features are imputed with mode of those features.
- By doing feature engineering 2 new features Hypertension and Glucose_diabetes are created.
- As dependent variable TenYearCHD is highly imbalanced, we use SMOTE for balancing training data.
- Then applying several classifier models, we choose GradientBoostingClassifier as our optimum model as it provides best evaluation metrics for this dataset.

Which gender is more prone to CHD

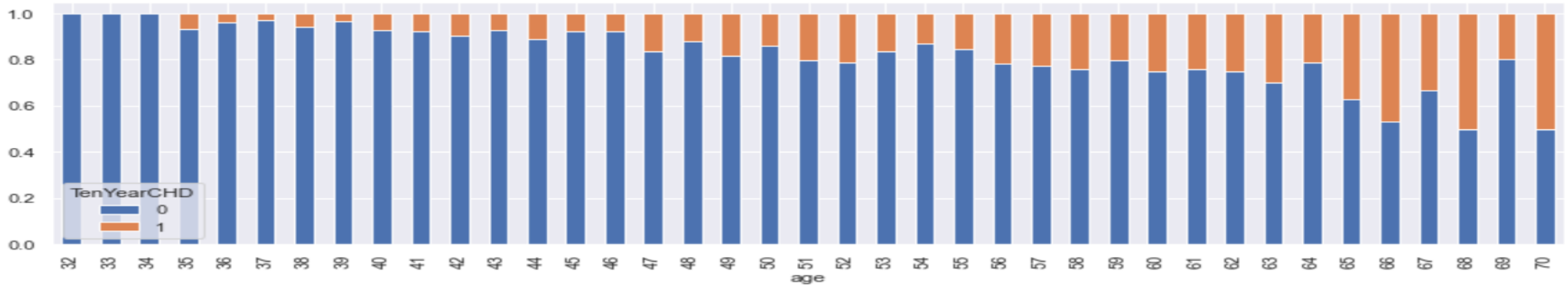


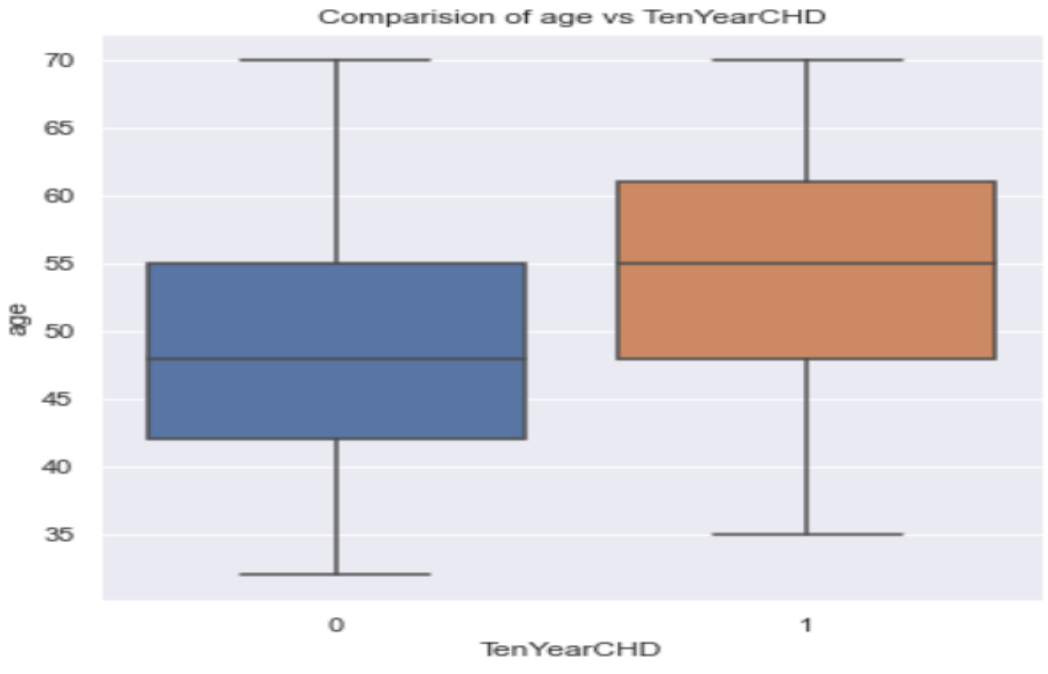


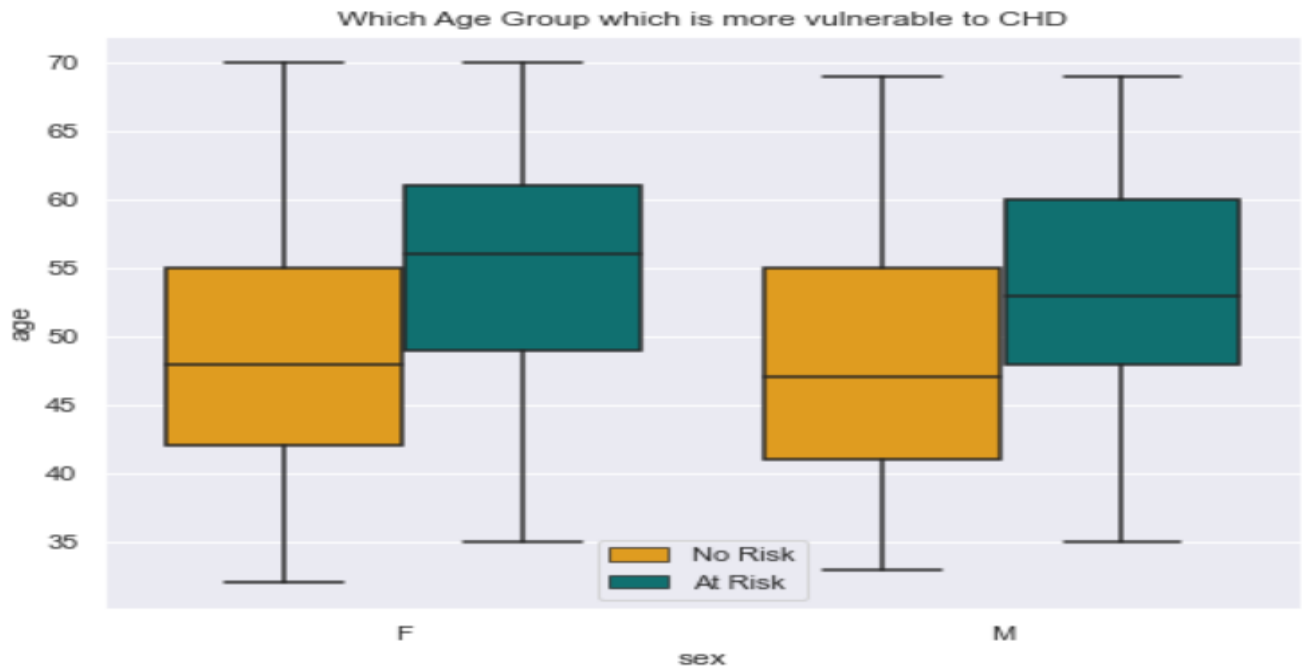


	sex	Total_Count	% of Total	Yes_CHD	No_CHD	%Yes_CHD	%No_CHD
0	F	1923	56.725664	239	1684	12.428	87.572
1	M	1467	43.274336	272	1195	18.541	81.459

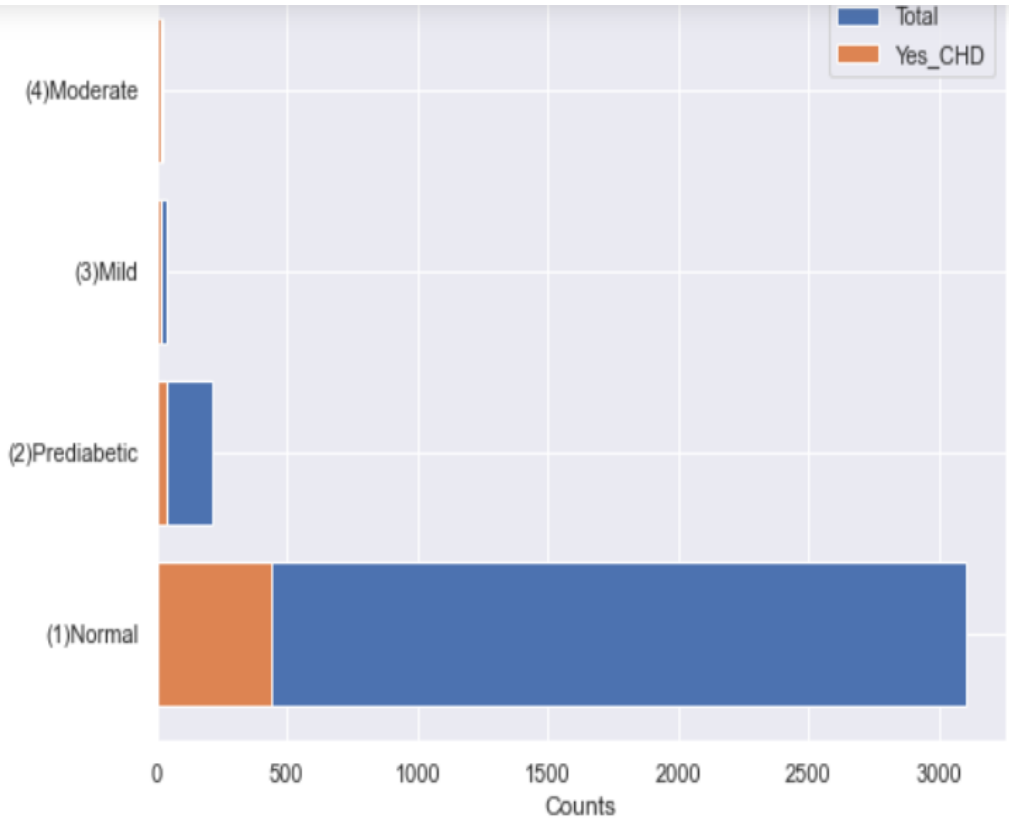
Analysis on age group vulnerable CHD





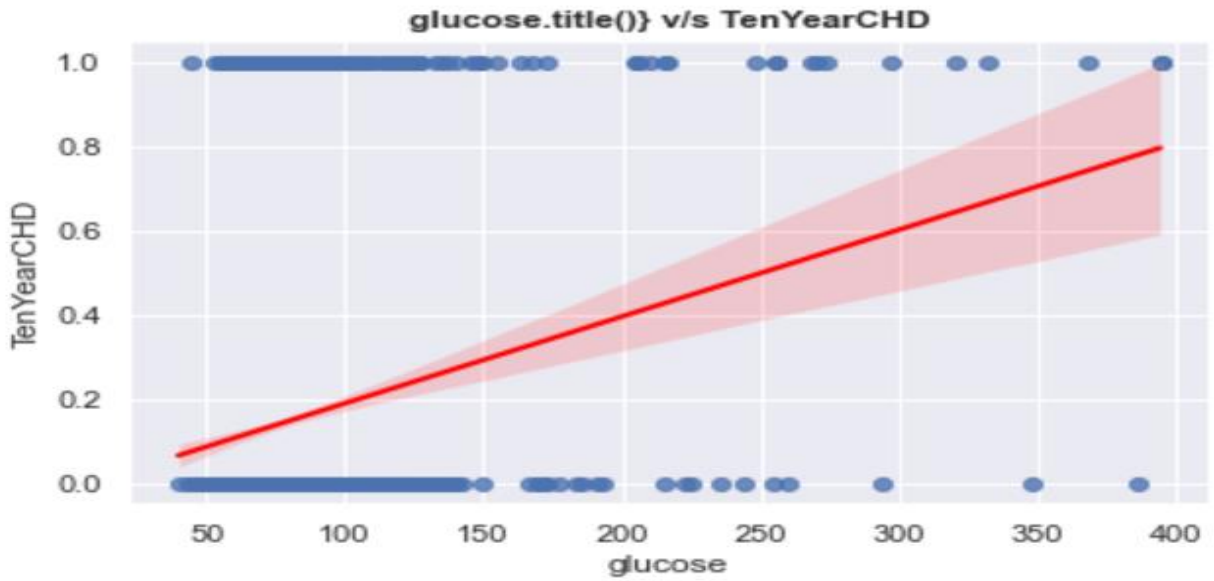


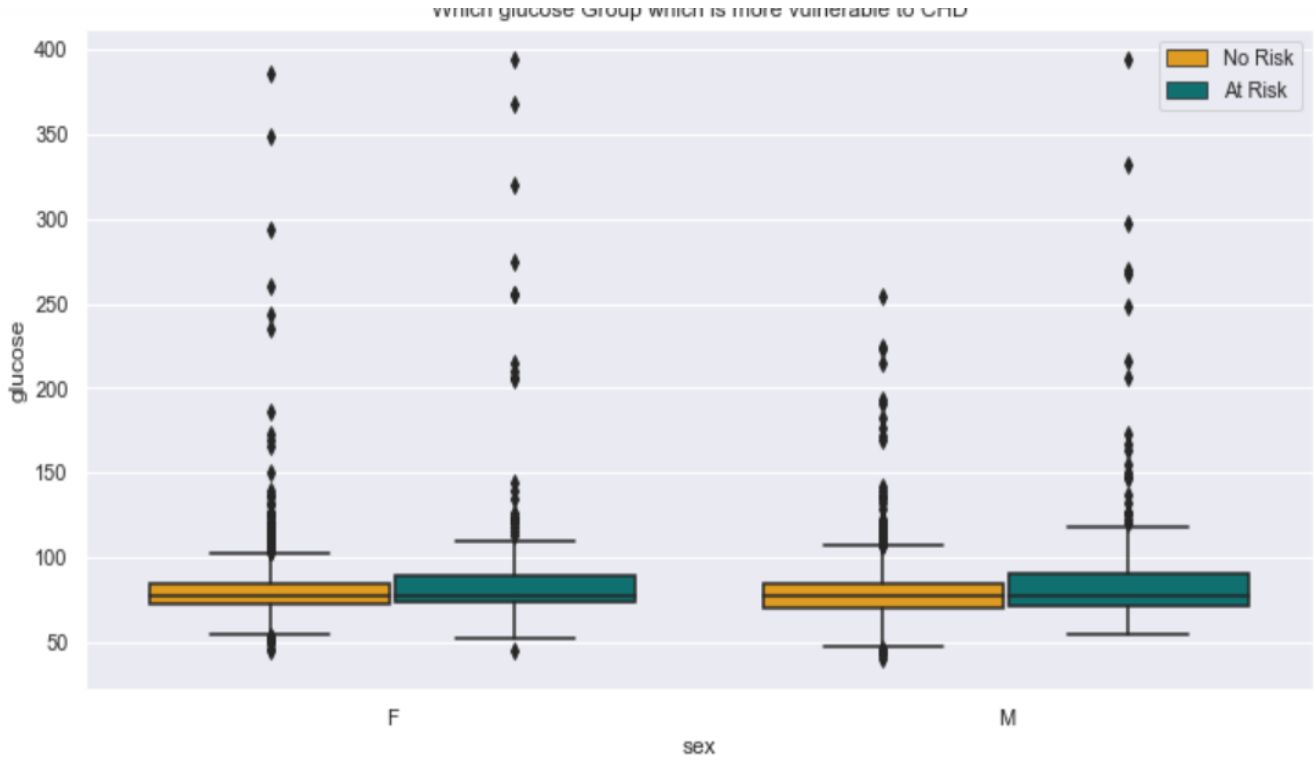
Effect of glucose on CHD



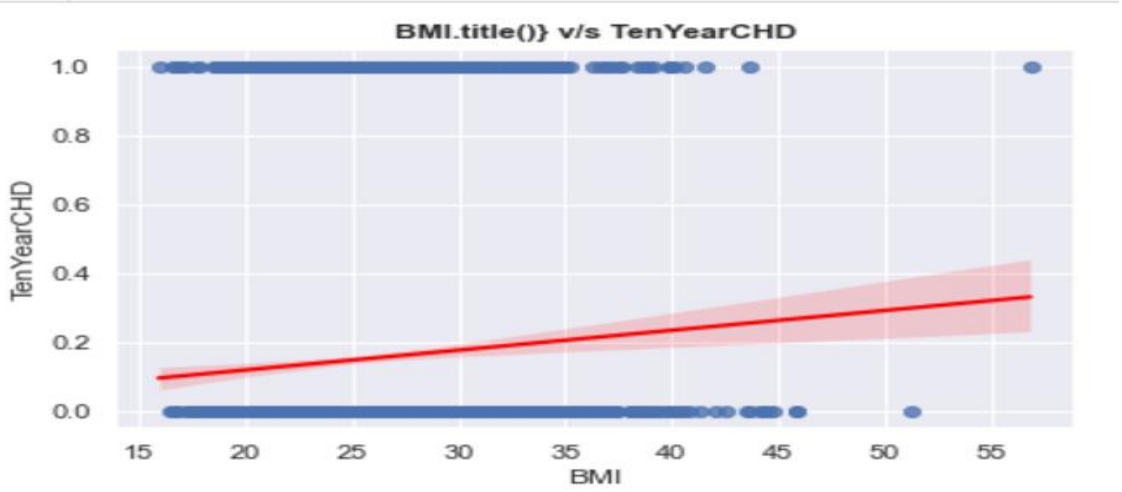
t[241]:

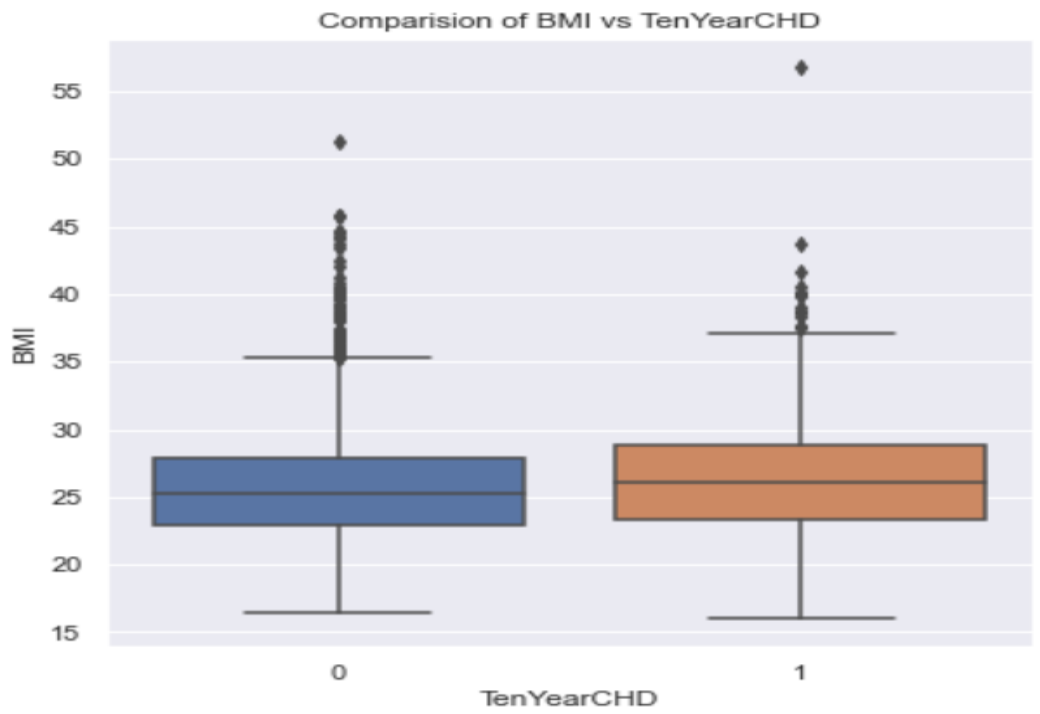
	Glucose_diabetes	Total_Count	% of Total	Yes_CHD	No_CHD	%Yes_CHD	%No_CHD
0	(1)Normal	3103	91.533923	441	2662	14.212	85.788
1	(2)Prediabetic	216	6.371681	36	180	16.667	83.333
2	(3)Mild	43	1.268437	16	27	37.209	62.791
3	(4)Moderate	28	0.825959	18	10	64.286	35.714

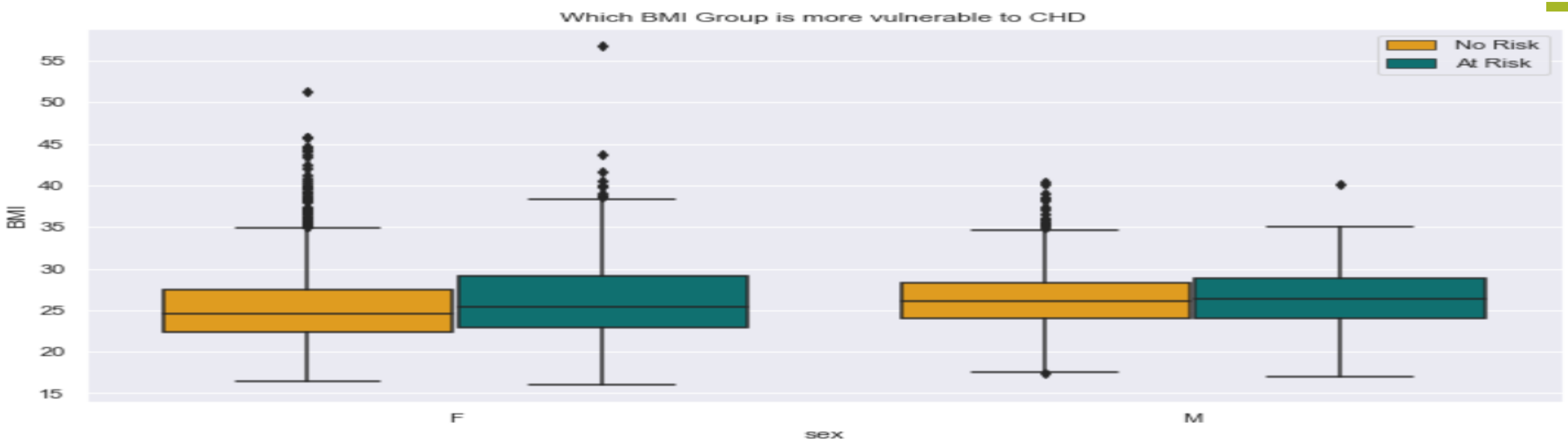




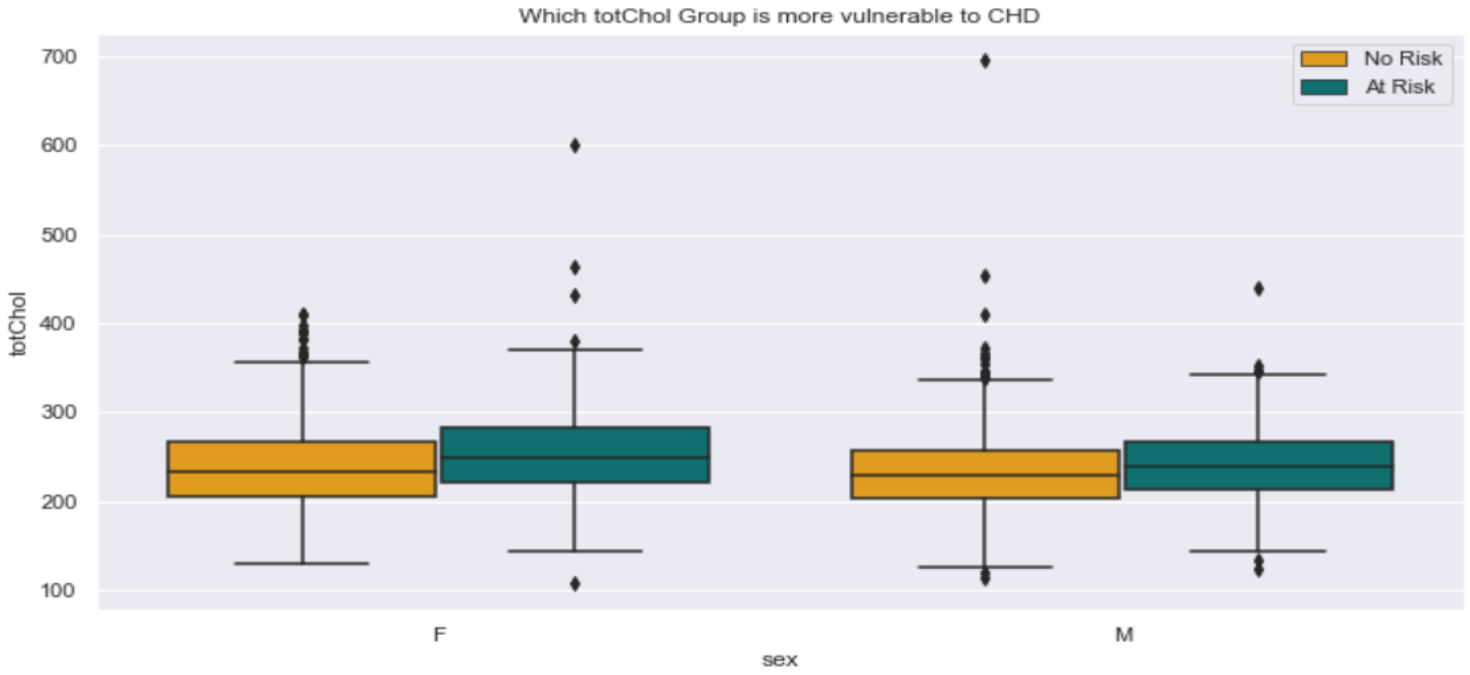
Analysis of BMI related to CHD

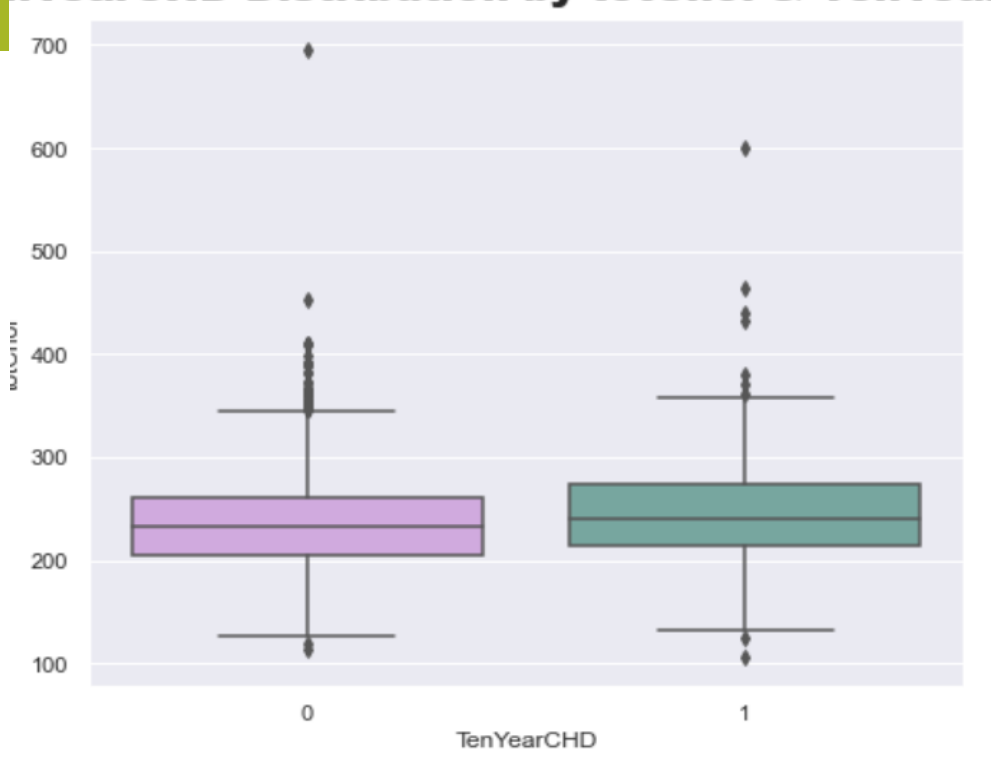


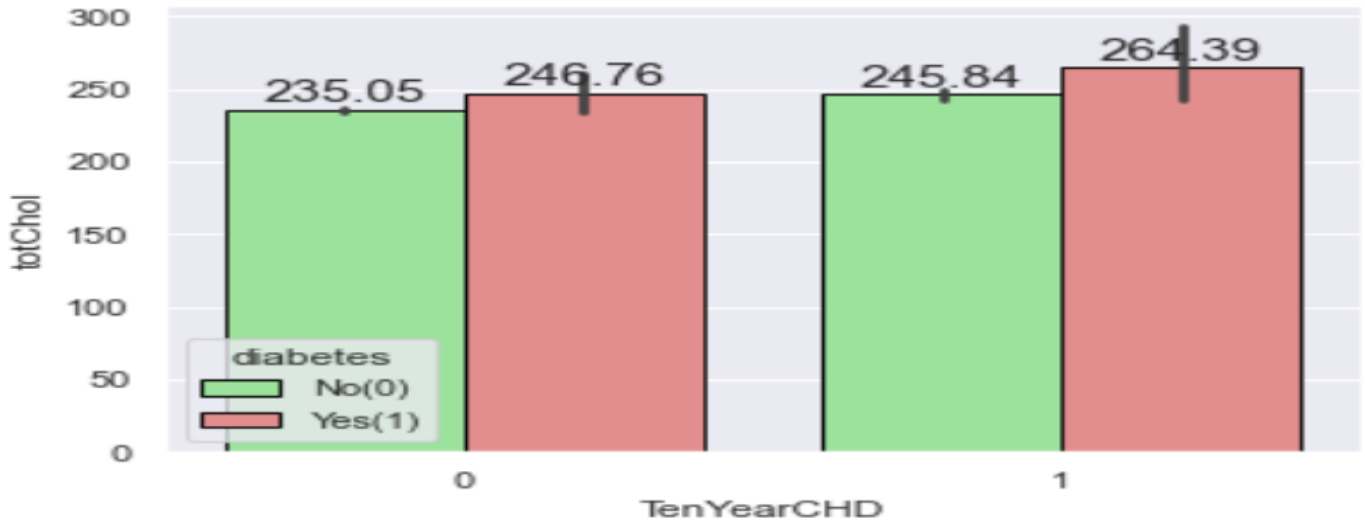


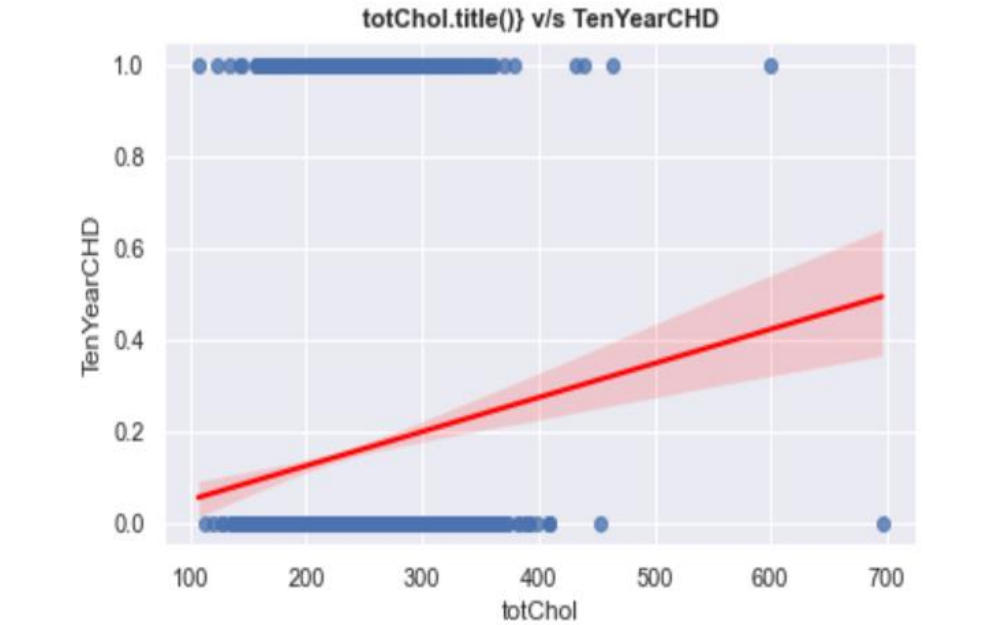


Analysis of totChol effect on CHD

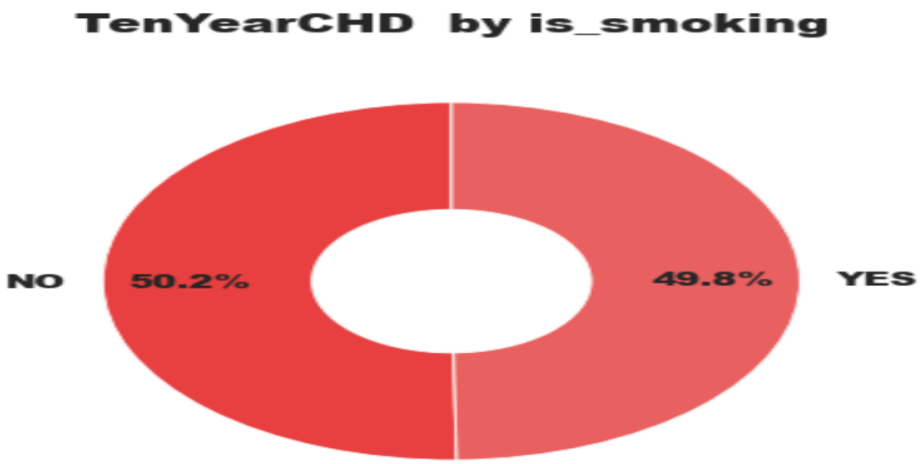


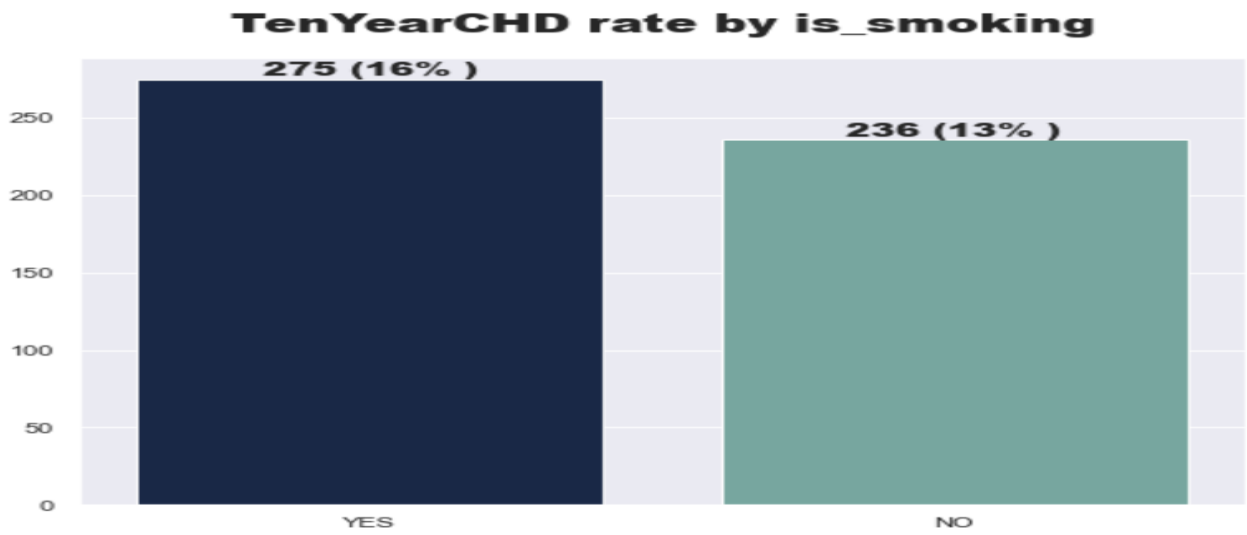


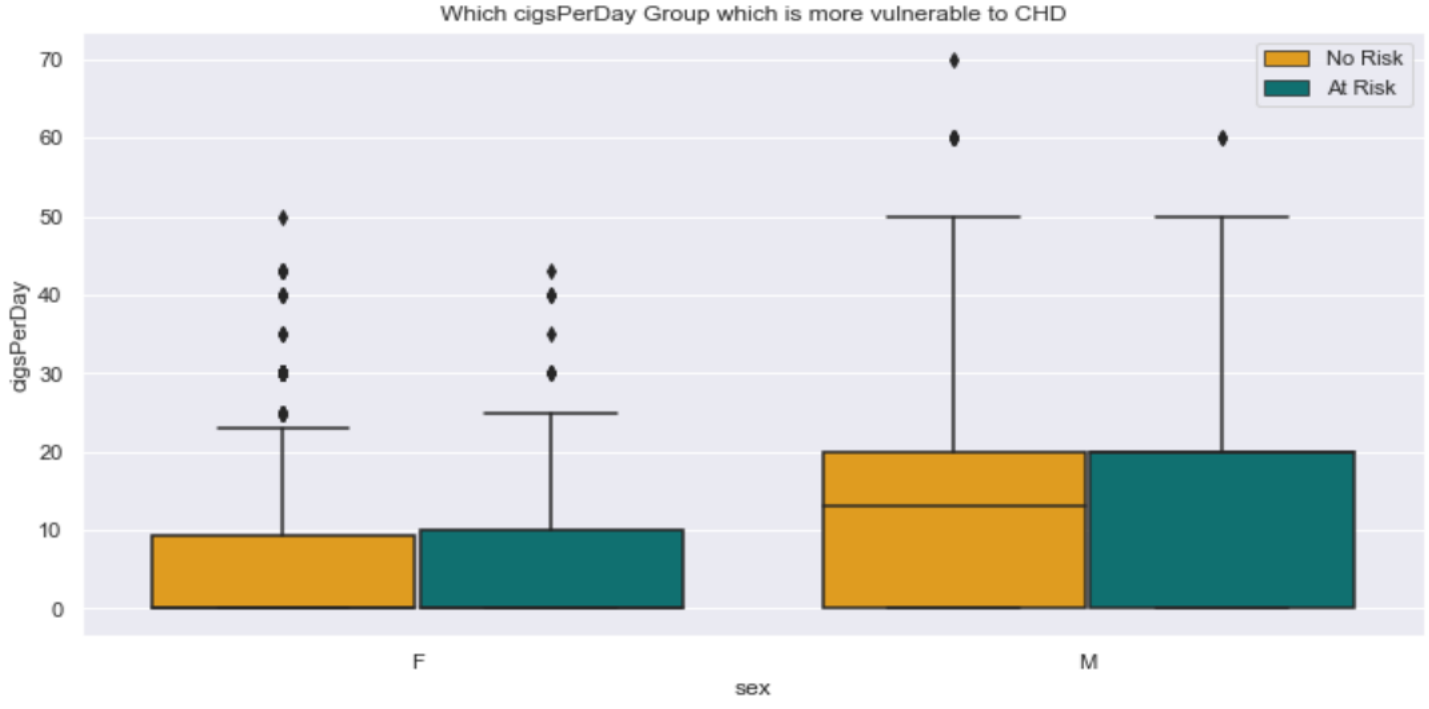


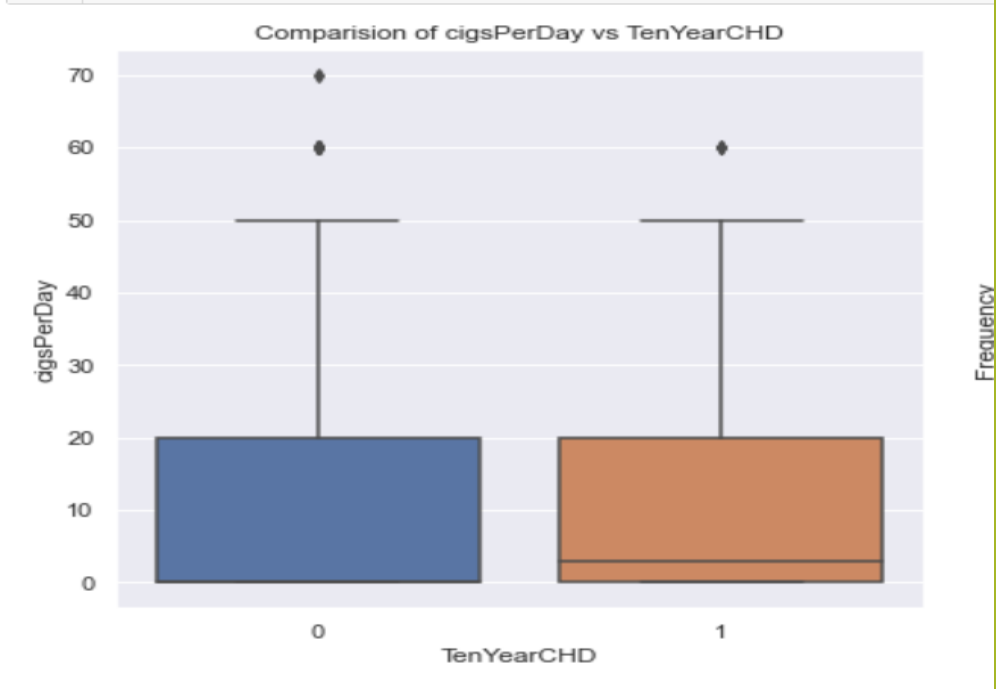


Effect of smoking related to risk of CHD

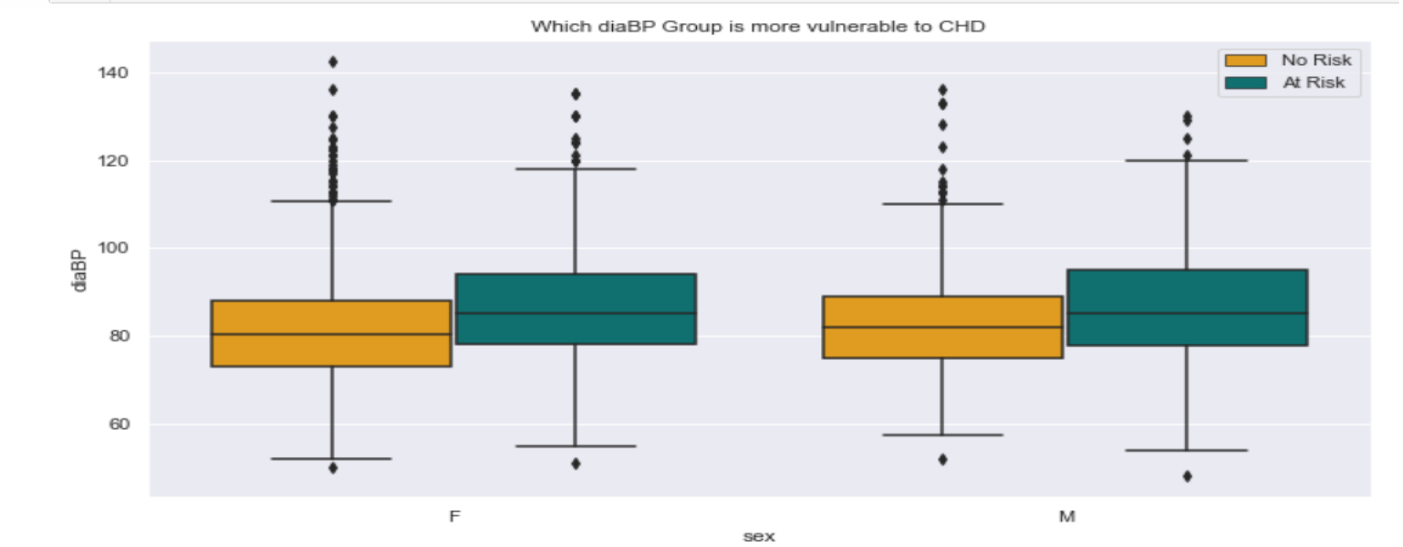


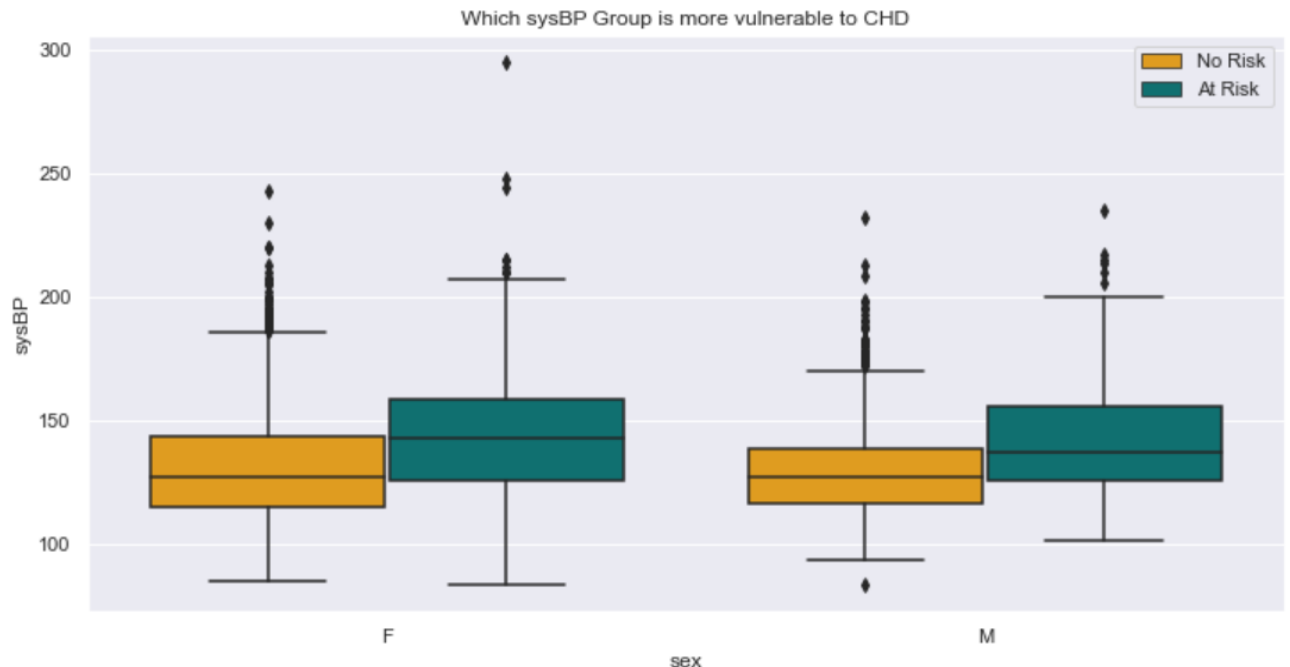


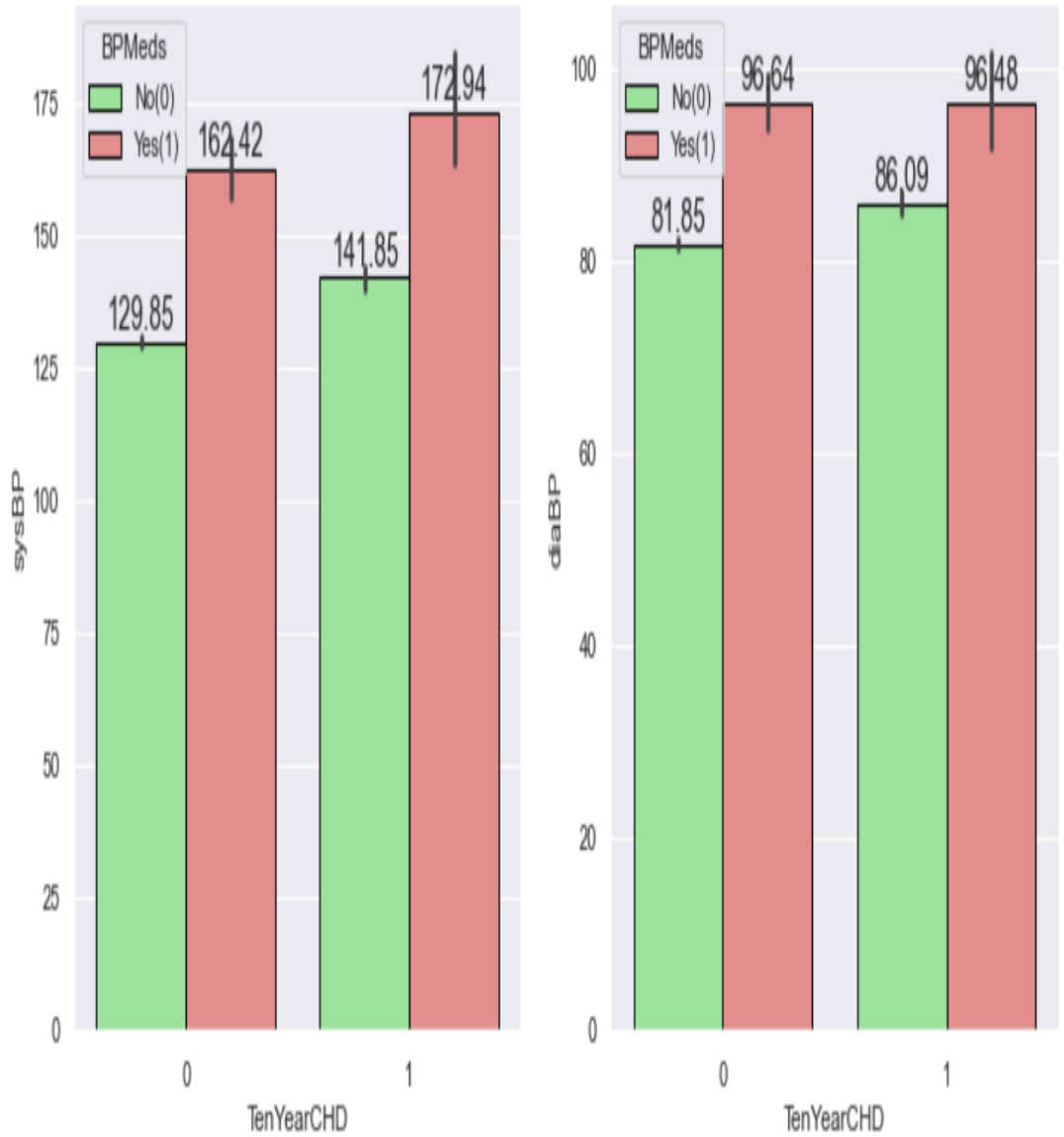




Systolic and diastolic bp effect on risk of CHD







Effect of Hypertension on risk of getting CHD

- Hypertension is combination of SysBP and DiaBP when,

(SysBP < 90) or (DiaBP < 60) = '0(Hypotension)',

(SysBP < 120) and (DiaBP < 80) = '1(Optimal)',

(SysBP < 129) or (DiaBP < 84) = '2(Normal)',

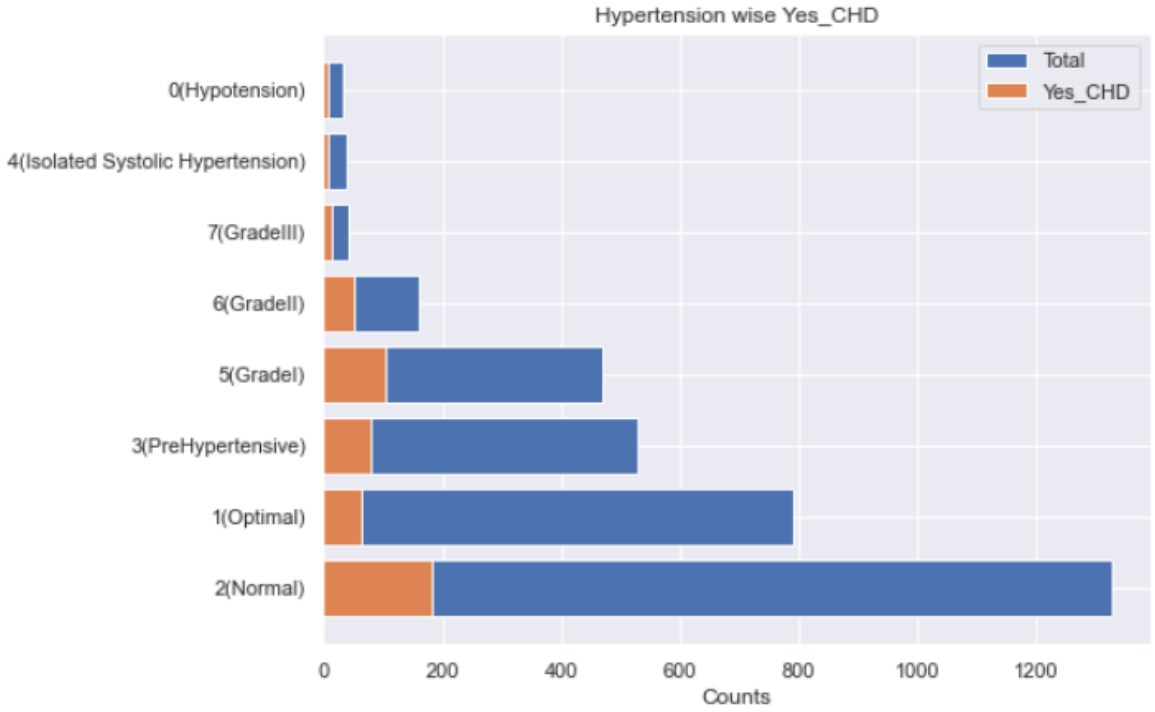
(SysBP < 140) or (DiaBP < 89)=3 (PreHypertensive),

(SysBP > 140) and (DiaBP < 90)= '4(Isolated Systolic Hypertension)',

(SysBP < 160) or (DiaBP < 100)='5(GradeI),

(SysBP < 180) or (DiaBP >=110):='6(GradeII)'

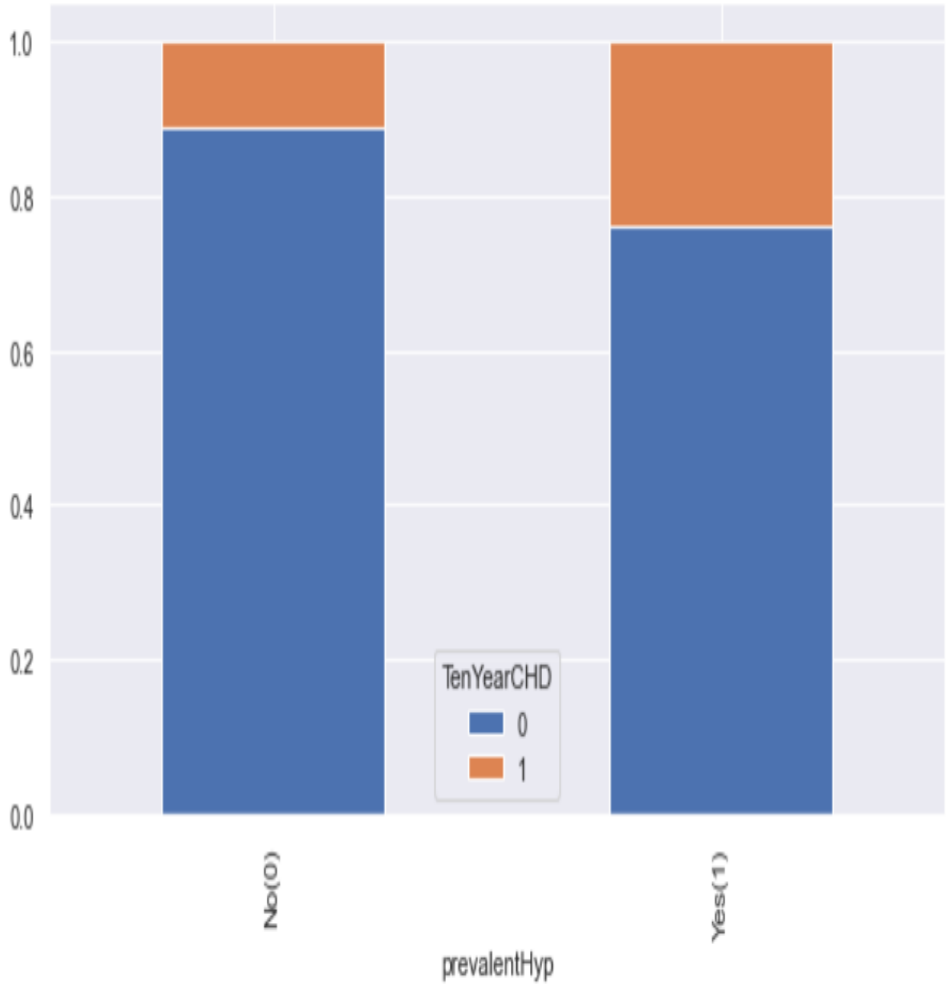
Others = '7(GradeIII)

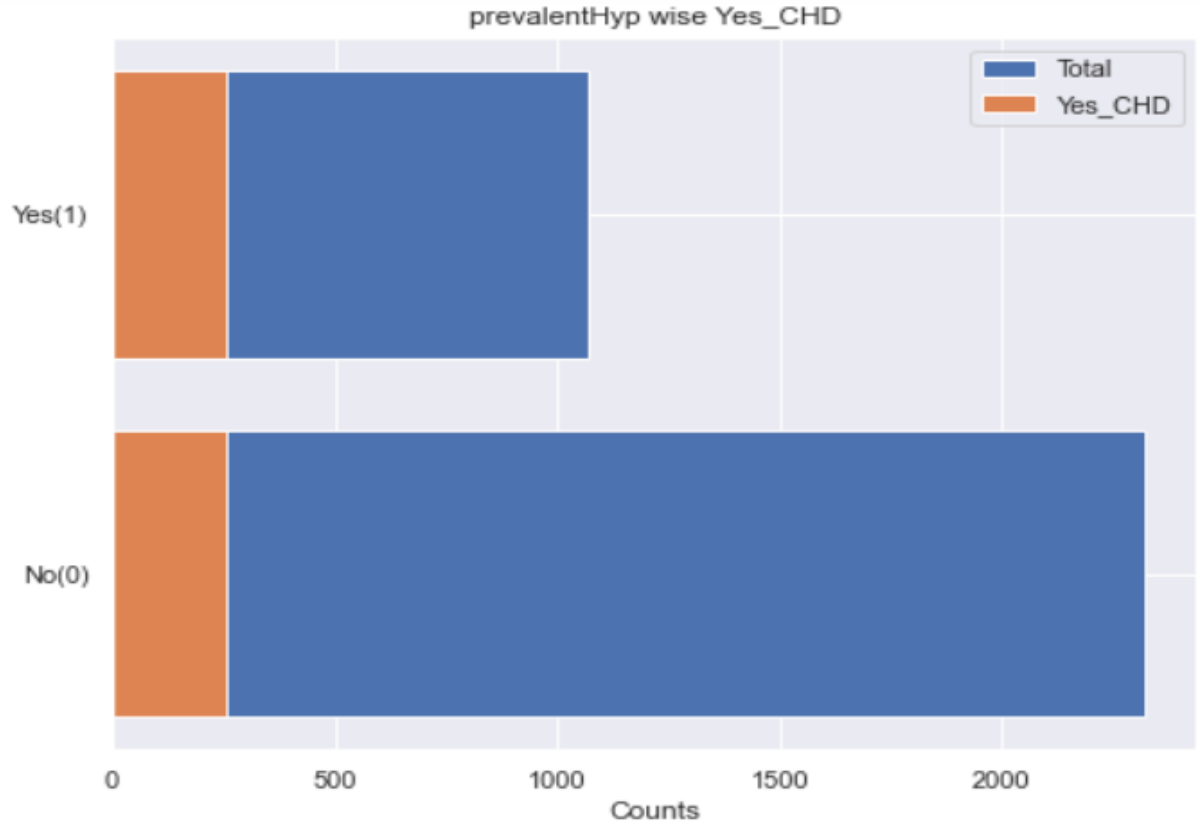


Out[230]:

	Hypertension	Total_Count	% of Total	Yes_CHD	No_CHD	%Yes_CHD	%No_CHD
2	2(Normal)	1326	15.604720	182	1144	13.725	86.275
1	1(Optimal)	792	23.362832	64	728	8.081	91.919
3	3(PreHypertensive)	529	13.834808	79	450	14.934	85.066
5	5(GradeI)	469	1.209440	105	364	22.388	77.612
6	6(GradeII)	161	1.150442	51	110	31.677	68.323
7	7(GradeIII)	41	0.973451	14	27	34.146	65.854
4	4(Isolated Systolic Hypertension)	39	4.749263	9	30	23.077	76.923
0	0(Hypotension)	33	39.115044	7	26	21.212	78.788

Effect of prevalentHyp on TenYearCHD

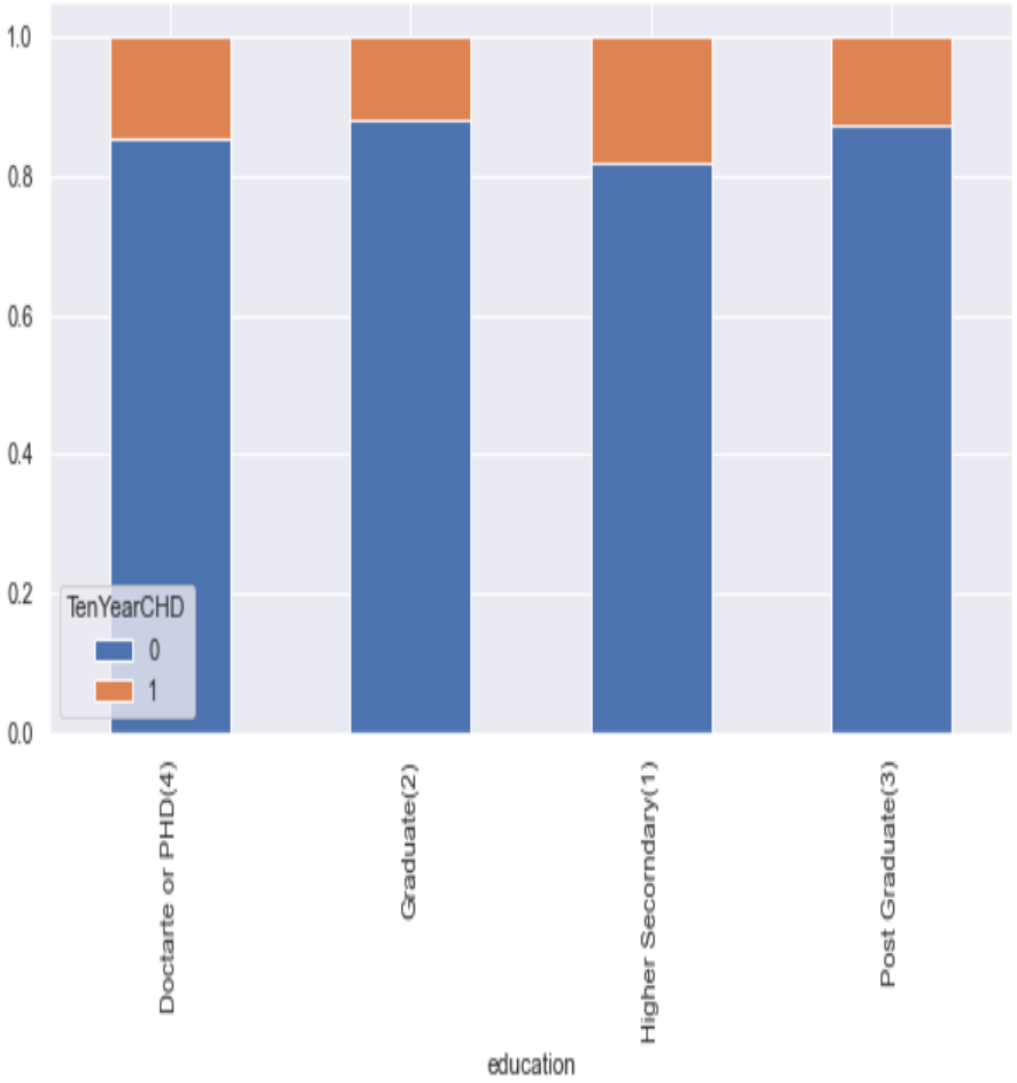


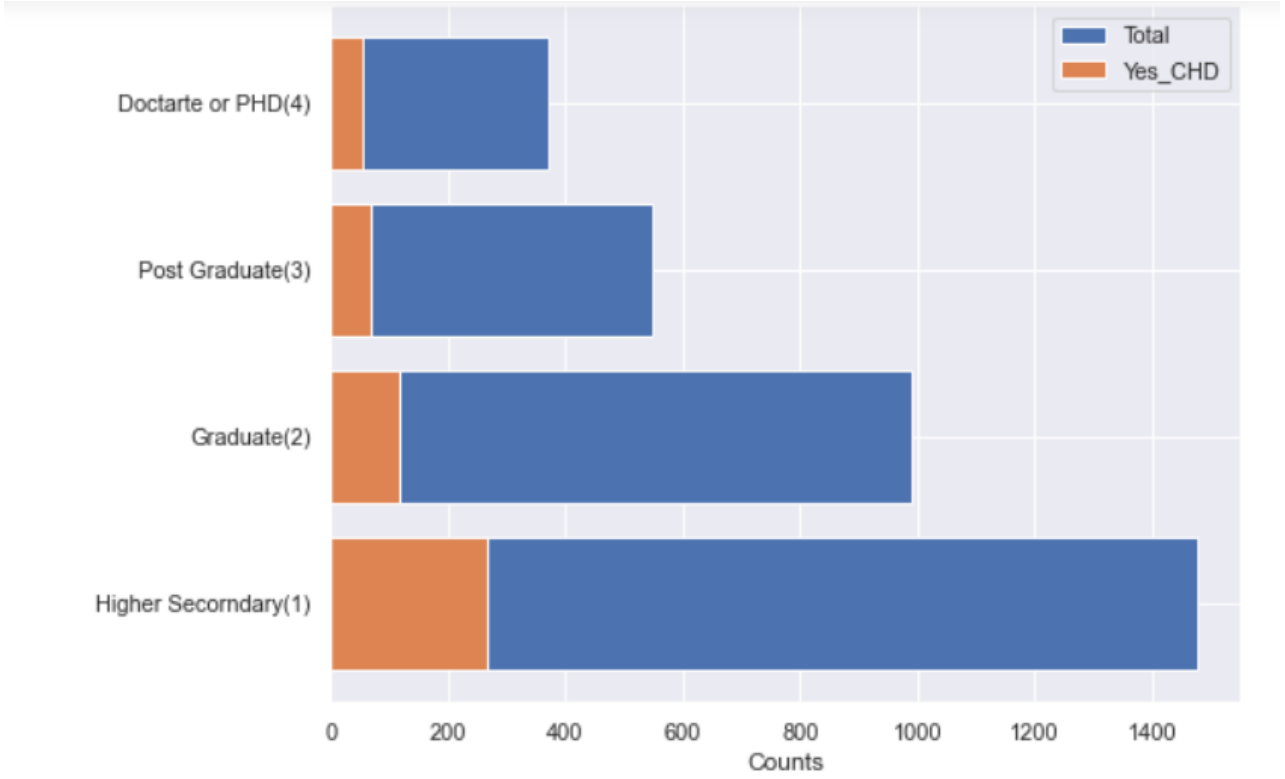


:

	prevalentHyp	Total_Count	% of Total	Yes_CHD	No_CHD	%Yes_CHD	%No_CHD
0	No(0)	2321	68.466077	256	2065	11.030	88.970
1	Yes(1)	1069	31.533923	255	814	23.854	76.146

Education level wrt risk of CHD

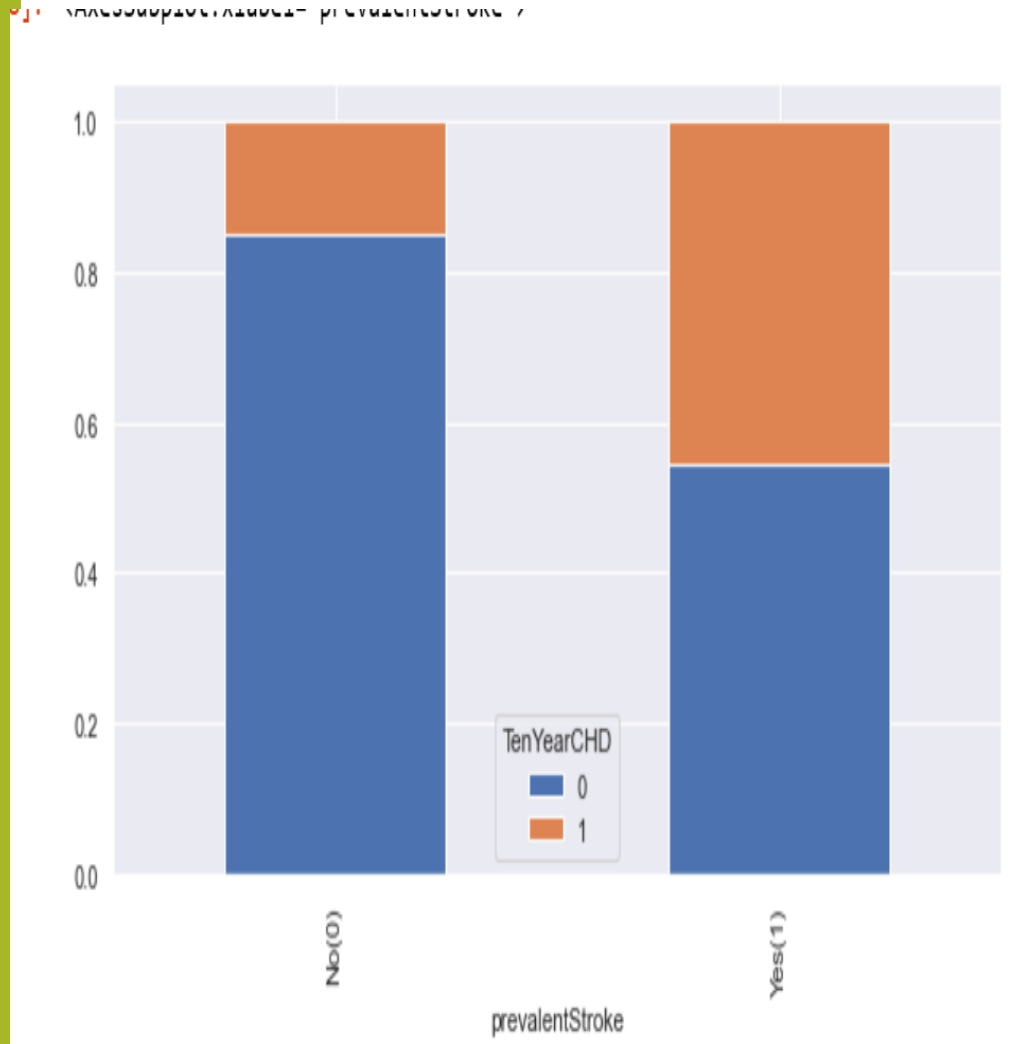


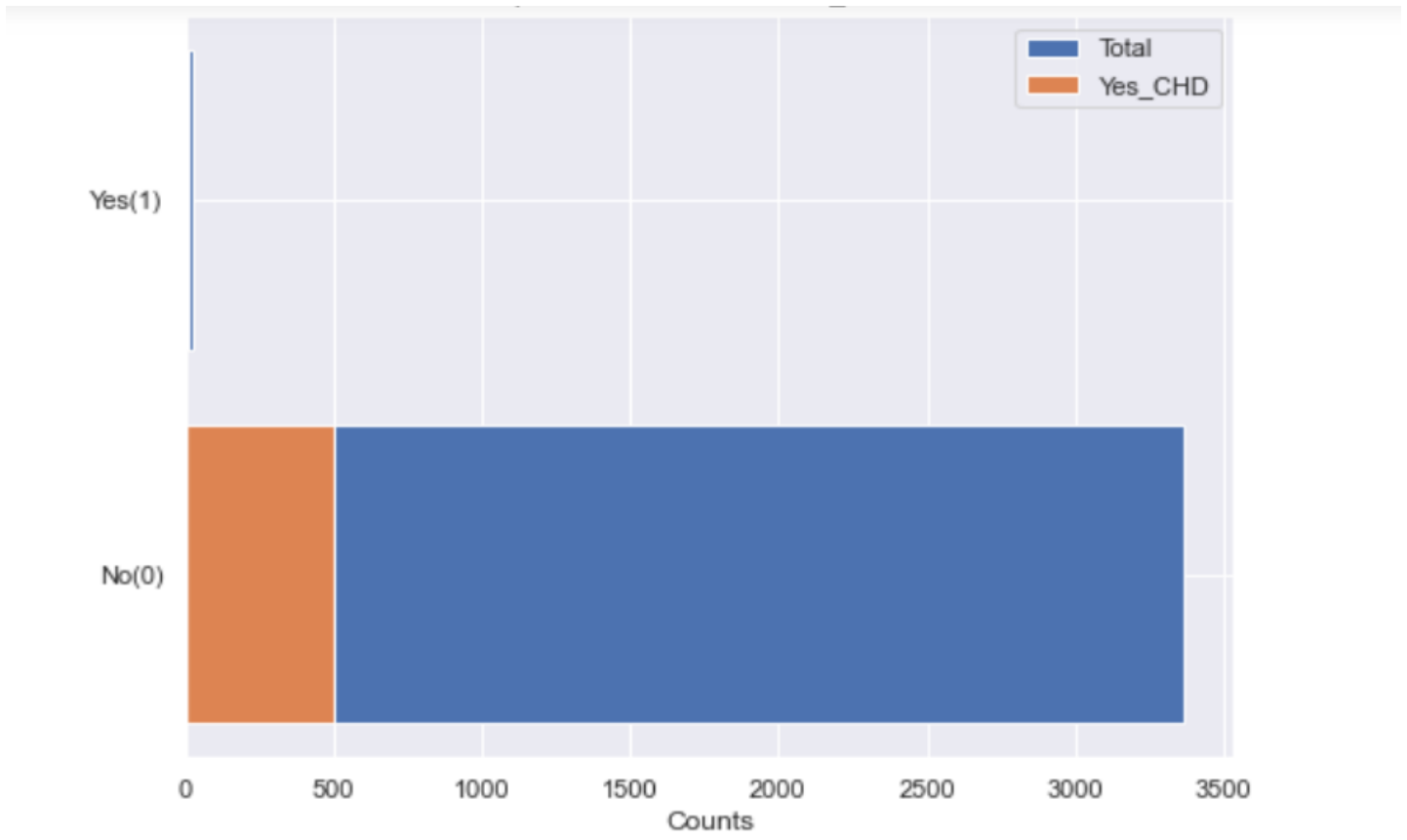


74]:

	education	Total_Count	% of Total	Yes_CHD	No_CHD	%Yes_CHD	%No_CHD
2	Higher Secorndary(1)	1478	16.19469	269	1209	18.200	81.800
1	Graduate(2)	990	29.20354	118	872	11.919	88.081
3	Post Graduate(3)	549	11.00295	70	479	12.750	87.250
0	Doctarte or PHD(4)	373	43.59882	54	319	14.477	85.523

Effect of prevalentStroke on TenYearCHD

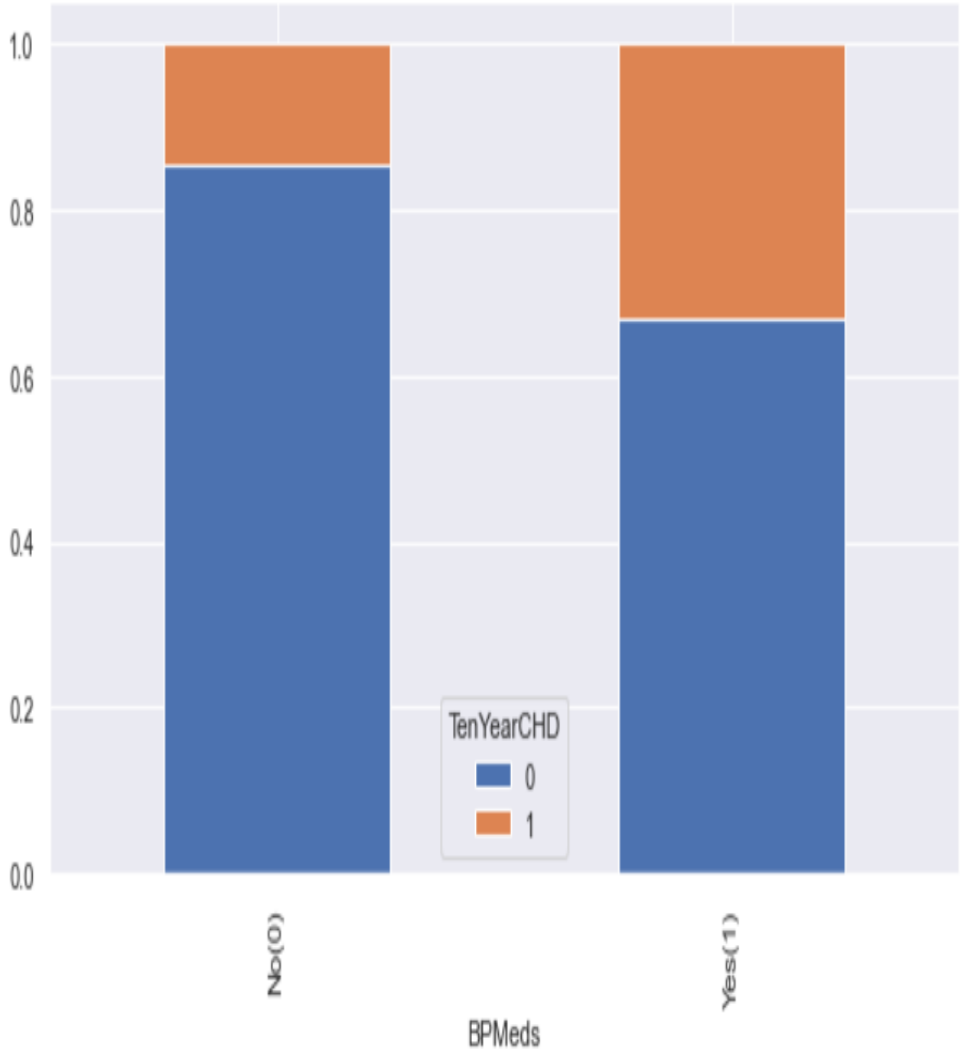


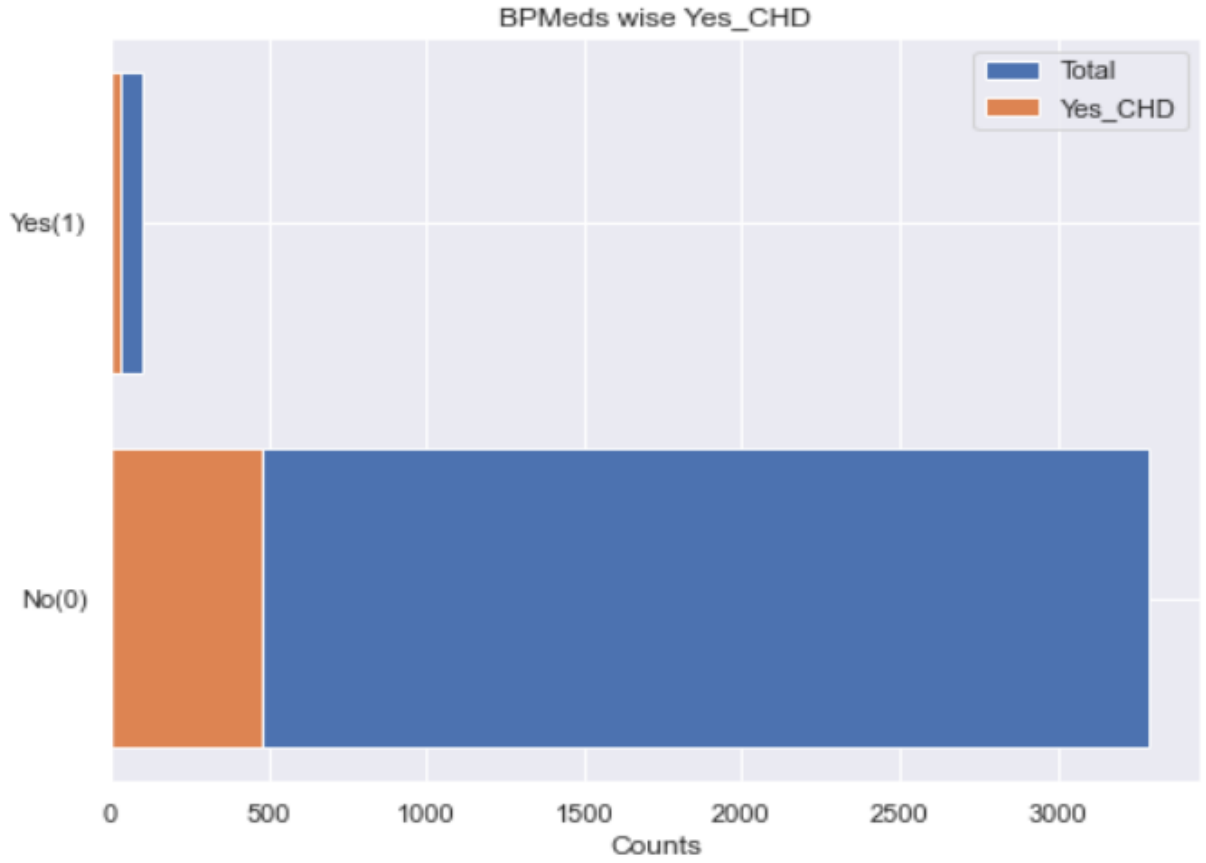


]:

	prevalentStroke	Total_Count	% of Total	Yes_CHD	No_CHD	%Yes_CHD	%No_CHD
0	No(0)	3368	99.351032	501	2867	14.875	85.125
1	Yes(1)	22	0.648968	10	12	45.455	54.545

Effect of BPMeds on risk of getting TenYearCHD

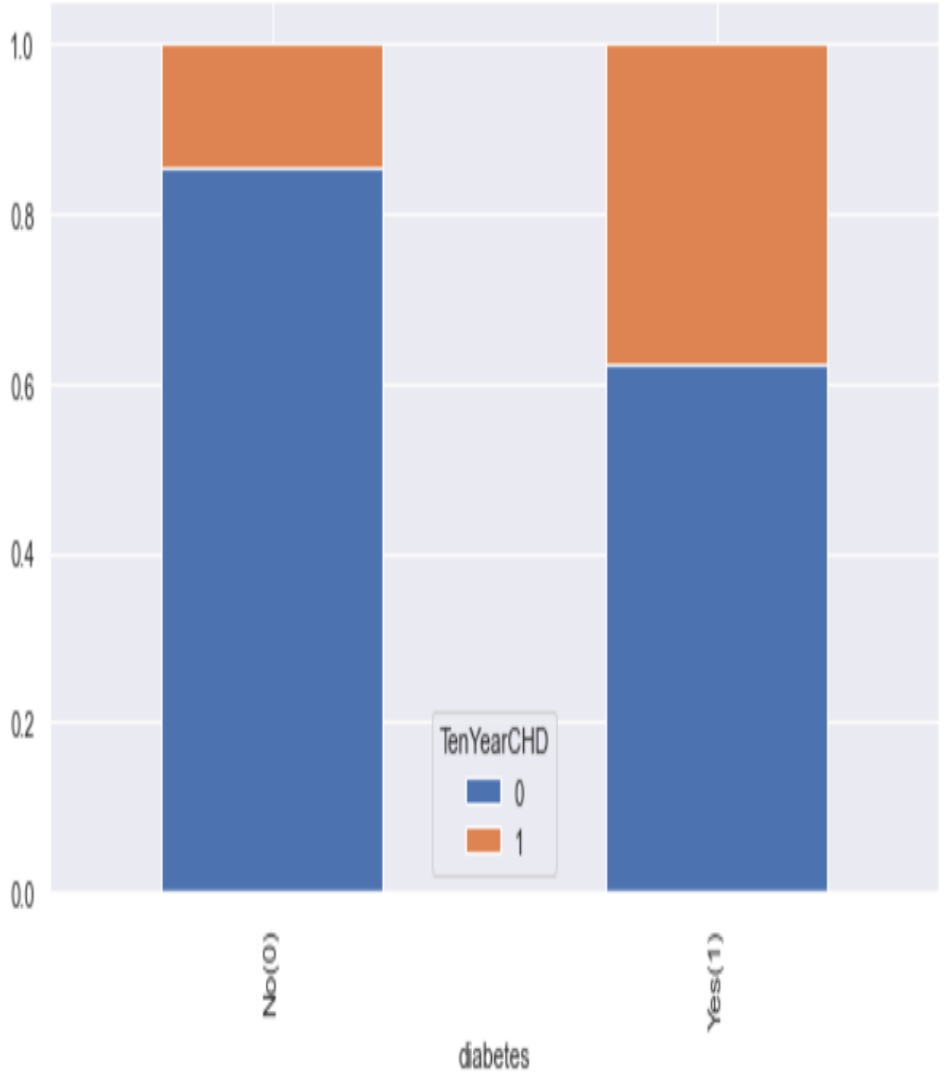


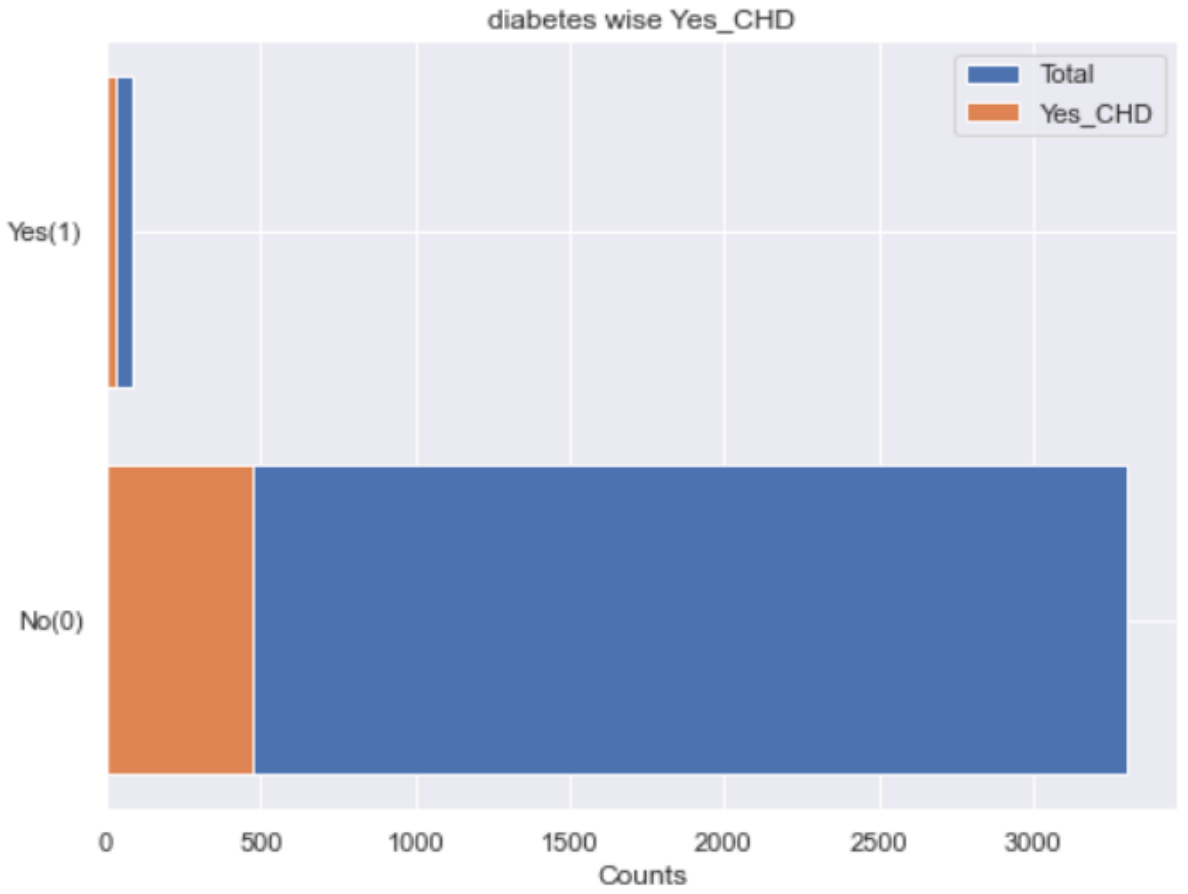


]:

	BPMeds	Total_Count	% of Total	Yes_CHD	No_CHD	%Yes_CHD	%No_CHD
0	No(0)	3290	97.050147	478	2812	14.529	85.471
1	Yes(1)	100	2.949853	33	67	33.000	67.000

Effect of diabetes on risk of getting TenYearCHD



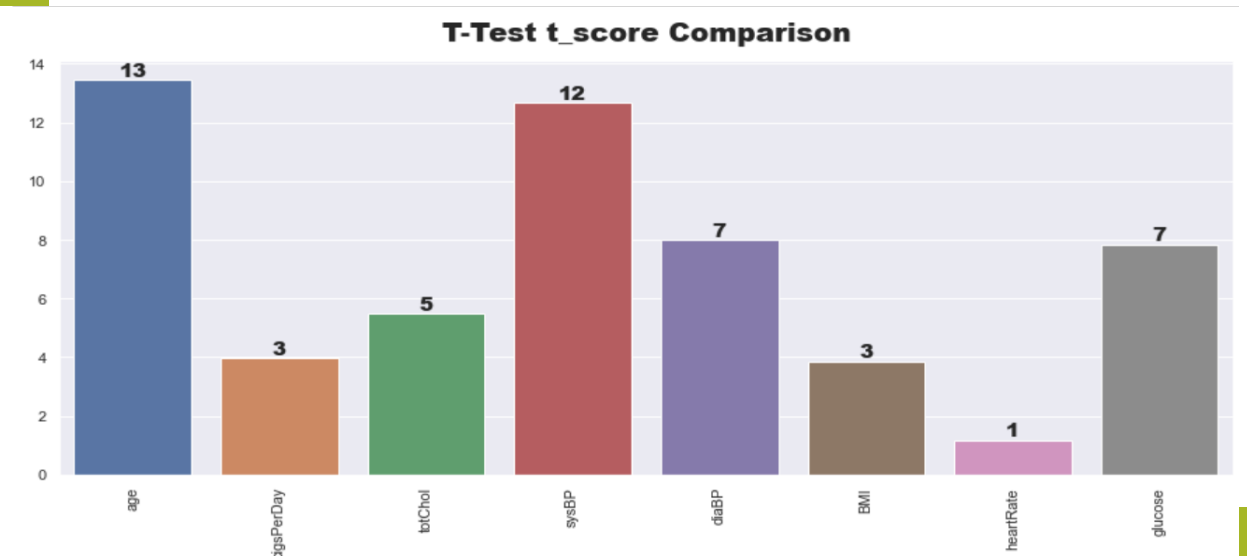


45]:

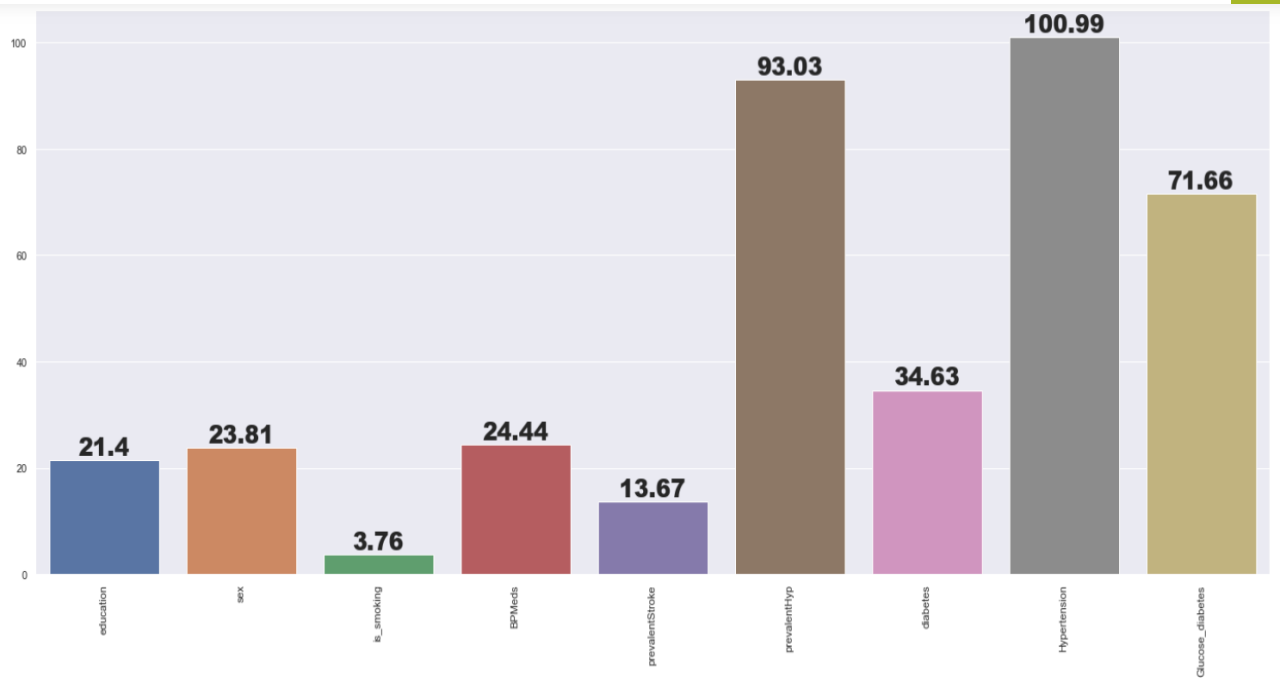
	diabetes	Total_Count	% of Total	Yes_CHD	No_CHD	%Yes_CHD	%No_CHD
0	No(0)	3303	97.433628	478	2825	14.472	85.528
1	Yes(1)	87	2.566372	33	54	37.931	62.069

Hypothesis Test Result

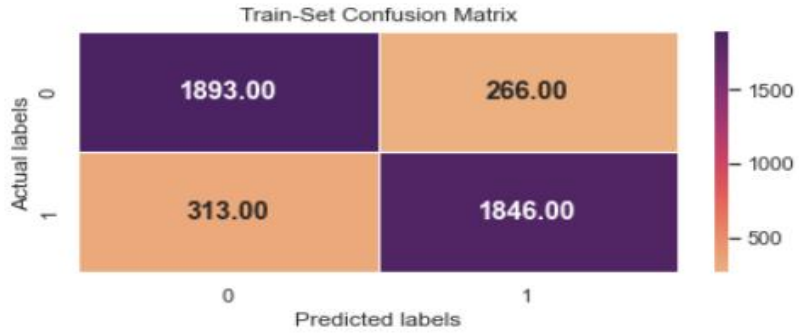
	Features	T_Score	P_value	Yes_mean	No_mean	Yes_std	No_std
0	age	13.436518	0.00000000000000000000	54.1292	48.7280	8.1257	8.4172
1	cigsPerDay	3.968408	0.00007385406644383295	10.9256	8.6707	13.0753	11.6039
2	totChol	5.472428	0.00000004761617144774	247.0352	235.2657	49.1546	43.9873
3	sysBP	12.670642	0.00000000000000000000	143.8542	130.6039	27.0612	20.7105
4	diaBP	7.989049	0.00000000000000184688	86.7632	82.1943	14.0430	11.4953
5	BMI	3.857900	0.00011649398533832090	26.4379	25.6788	4.5607	4.0112
6	heartRate	1.174093	0.24044019024614310398	76.5499	75.8753	12.1593	11.9355
7	glucose	7.839110	0.00000000000000602929	89.0568	80.4179	40.9127	17.9950

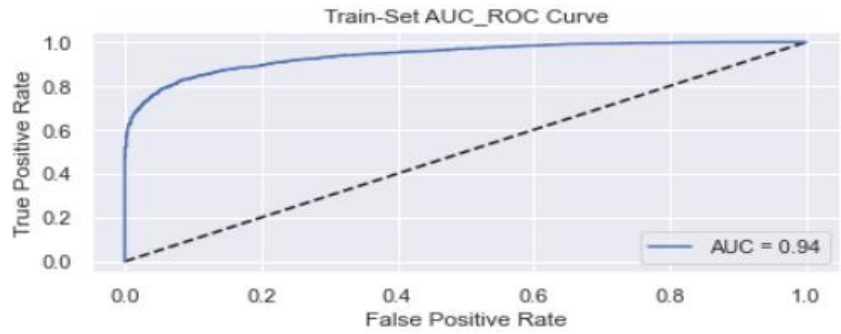


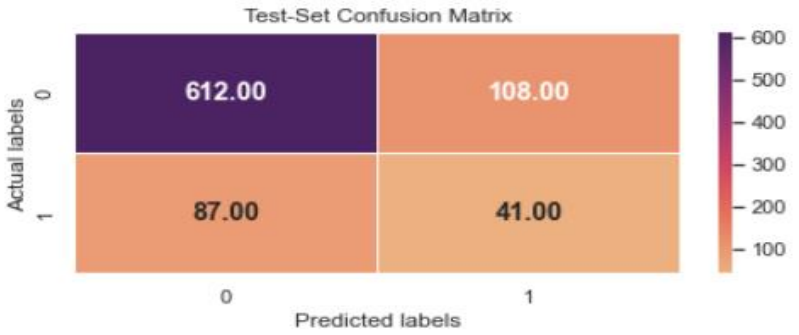
	Features	Chi_2 Statistic	P_value
0	education	21.400066	0.00008693890717147760
1	sex	23.814364	0.00000106087829356180
2	is_smoking	3.763251	0.05239062274853224094
3	BPMeds	24.442673	0.00000076554102539106
4	prevalentStroke	13.666246	0.00021834399343007940
5	prevalentHyp	93.029511	0.00000000000000000000
6	diabetes	34.632033	0.00000000398297868738
7	Hypertension	100.987619	0.00000000000000000067
8	Glucose_diabetes	71.657053	0.00000000000000188528

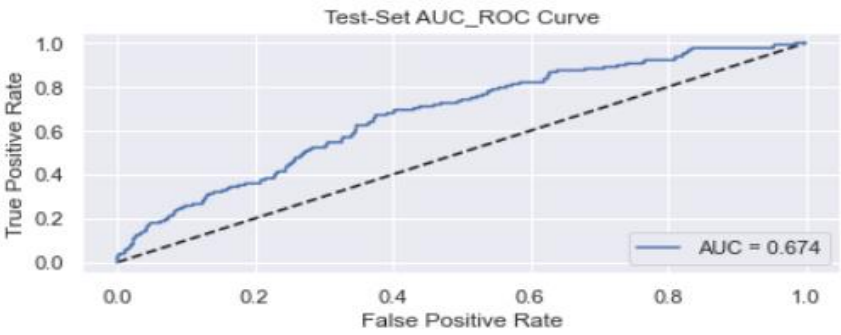


GradientBoostingClassifier Model Evaluation metrics









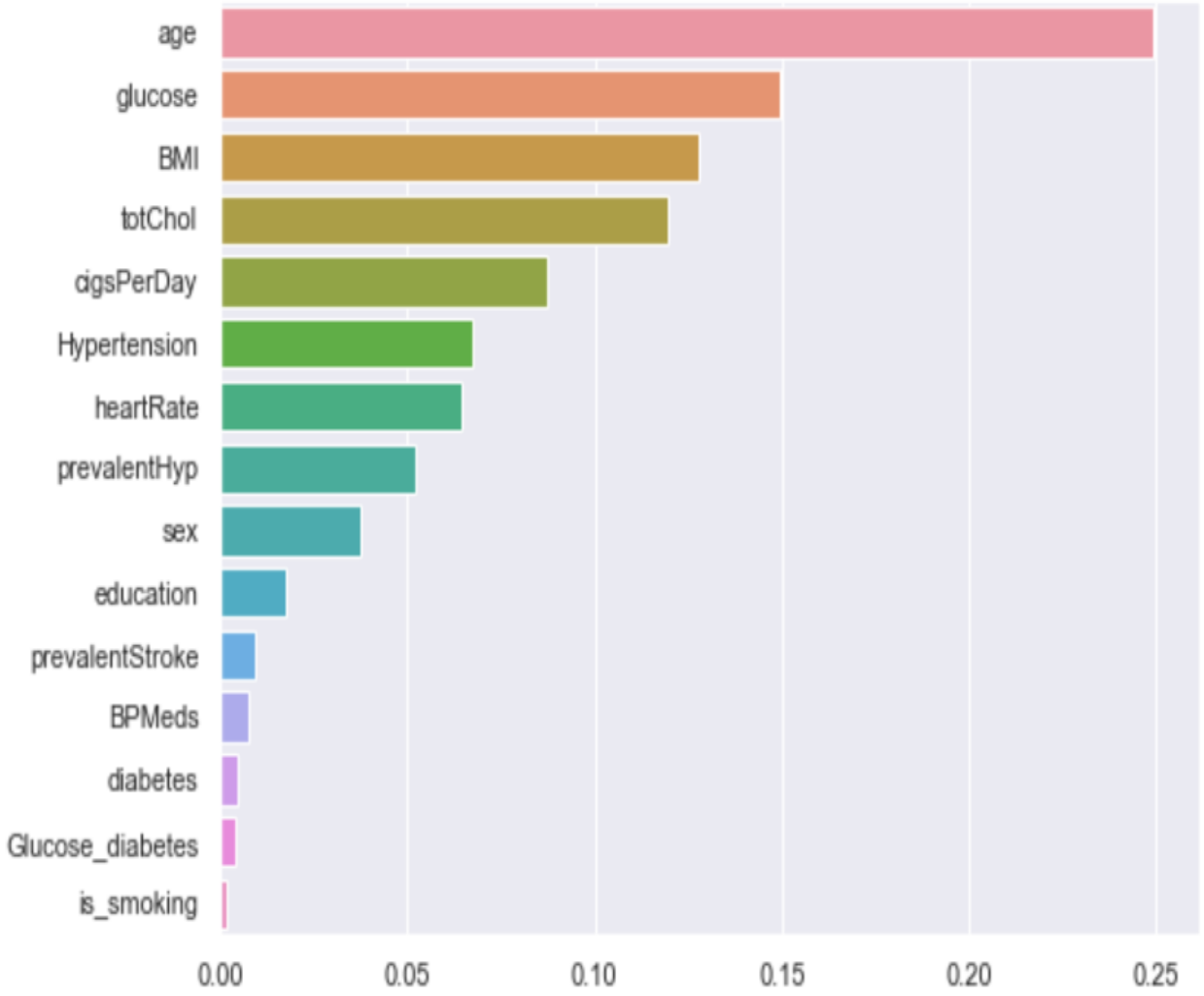
Classification Report(test)

	precision	recall	f1-score	support
0	0.88	0.85	0.86	720
1	0.28	0.32	0.30	128
accuracy			0.77	848
macro avg	0.58	0.59	0.58	848
weighted avg	0.78	0.77	0.78	848

Feature Importance and all model results

[3]:

	Models	Precision	Recall	F1_Score	Accuracy	AUC
0	Logistic_regression	0.79	0.64	0.69	0.64	0.693
1	Decision_tree	0.82	0.64	0.57	0.51	0.640
2	Random_forest	0.79	0.68	0.72	0.68	0.675
3	AdaBoost	0.80	0.74	0.77	0.74	0.681
4	GradientBoosting	0.78	0.77	0.78	0.77	0.674
5	XGboost	0.77	0.81	0.79	0.81	0.624
6	SVM	0.77	0.70	0.73	0.70	0.593
7	knn	0.77	0.70	0.73	0.70	0.560



Recommendation

- As we saw earlier with increase in age risk of getting CHD in future increases and when compare with in genders, median age of risk of CHD is 55, but in case of males it is less than 55 i.e 53 and for women it is more than 55 i.e 56. So, all older people should be regular checked, particularly male as they have higher risk.
- When it comes purely gender basis males are more prone to getting CHD than females, so, males of middle to old age people should regularly do medical check up.
- With increase in glucose level, risk of CHD increases. When glucose level 200 to 400 risk of CHD is 64%, similarly for 120 to 200 is 37%, for 100 to 120 is 16 and for less than 100 is 14%.
- With increase in BMI level risk of CHD increases, particularly among females as in males on median BMI level risk is there. So, for both genders controlling BMI is important, but in females it is important as their risk of getting CHD base level lower compare to males.
- Similarly, with increase in totChol risk of getting CHD increases in both genders. Median totChol level for CHD in males and females are 240 and 248 respectively. Diabetic patients with high totChol level are more prone to getting CHD.
- Smokers have high risk CHD, as males smoke more, there risk is higher compare to females.
- People with GradeI(sysBP < 160 or diaBP <100), GradeII(sysBP<180 or diaBP>=100), GradeIII Hypertension and Hypotension(sysBP < 90 or diaBP < 60) are more prone to getting CHD in future. As we saw earlier people with prevalentHyp are more prone to CHD i.e 23.8% compared to 11.03%.