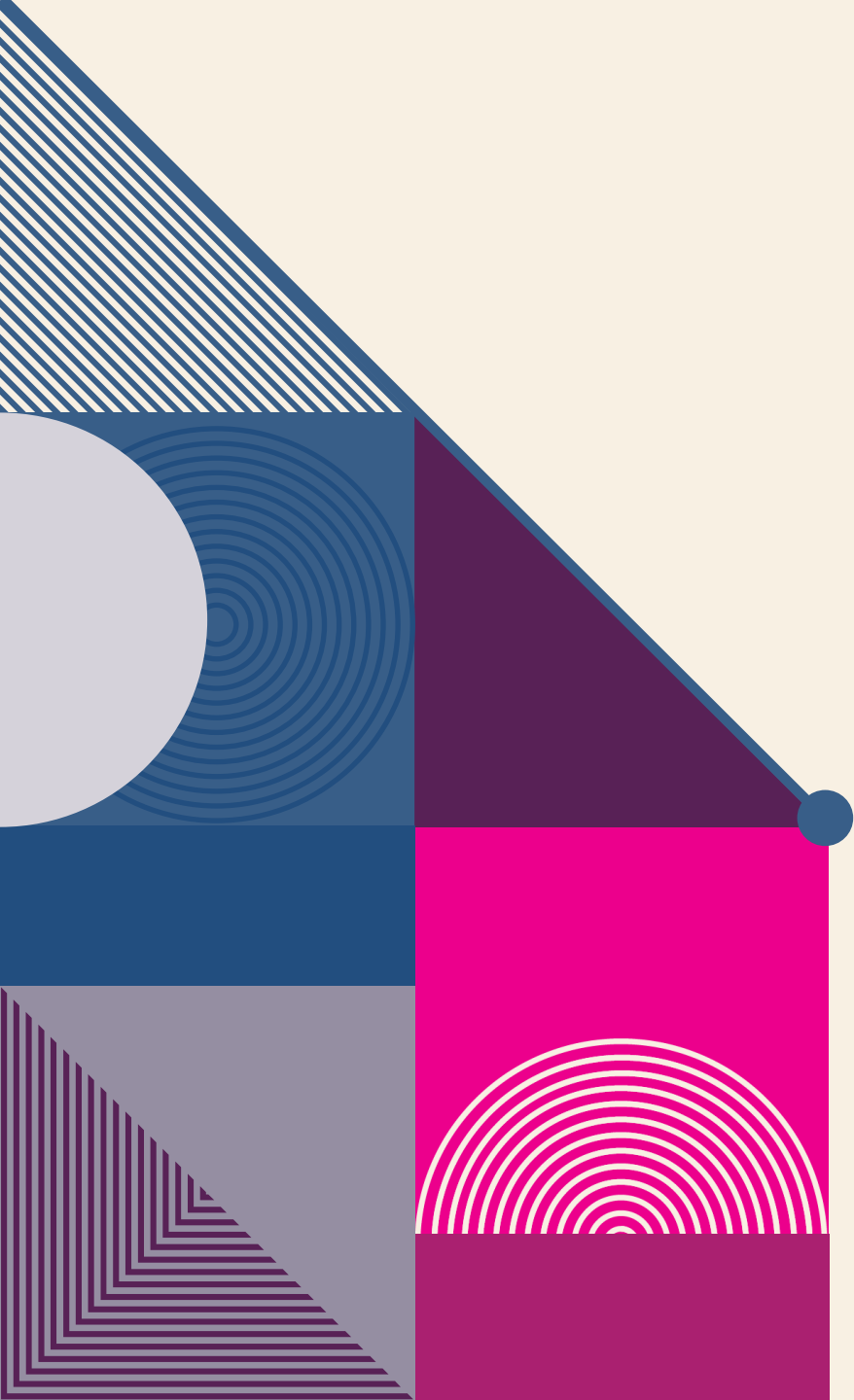




HEART DISEASE PREDICTION

BY NAZIM ATAKAN ERDOGAN,
DATA SCIENCE INTERN &
TRAINEE



AGENDA

Introduction and overview

Exploratory Data Analysis

Machine Learning

Conclusions

OVERVIEW



All the conclusions are embedded to the project as comments in the code.



The interrelating factors were investigated the the relationships among them were demonstrated.



Disease contributing factors such as BMI, blood pressure, exercise level, mental health, vegetariansim diet, income and education were investigated.



Machine learning methods such as linear regression, random forest, decision tree, gradient boost etc. were used to make predictions and verify the correlations between variables and possible disease correlating factors.

DATA OVERVIEW AND STATS

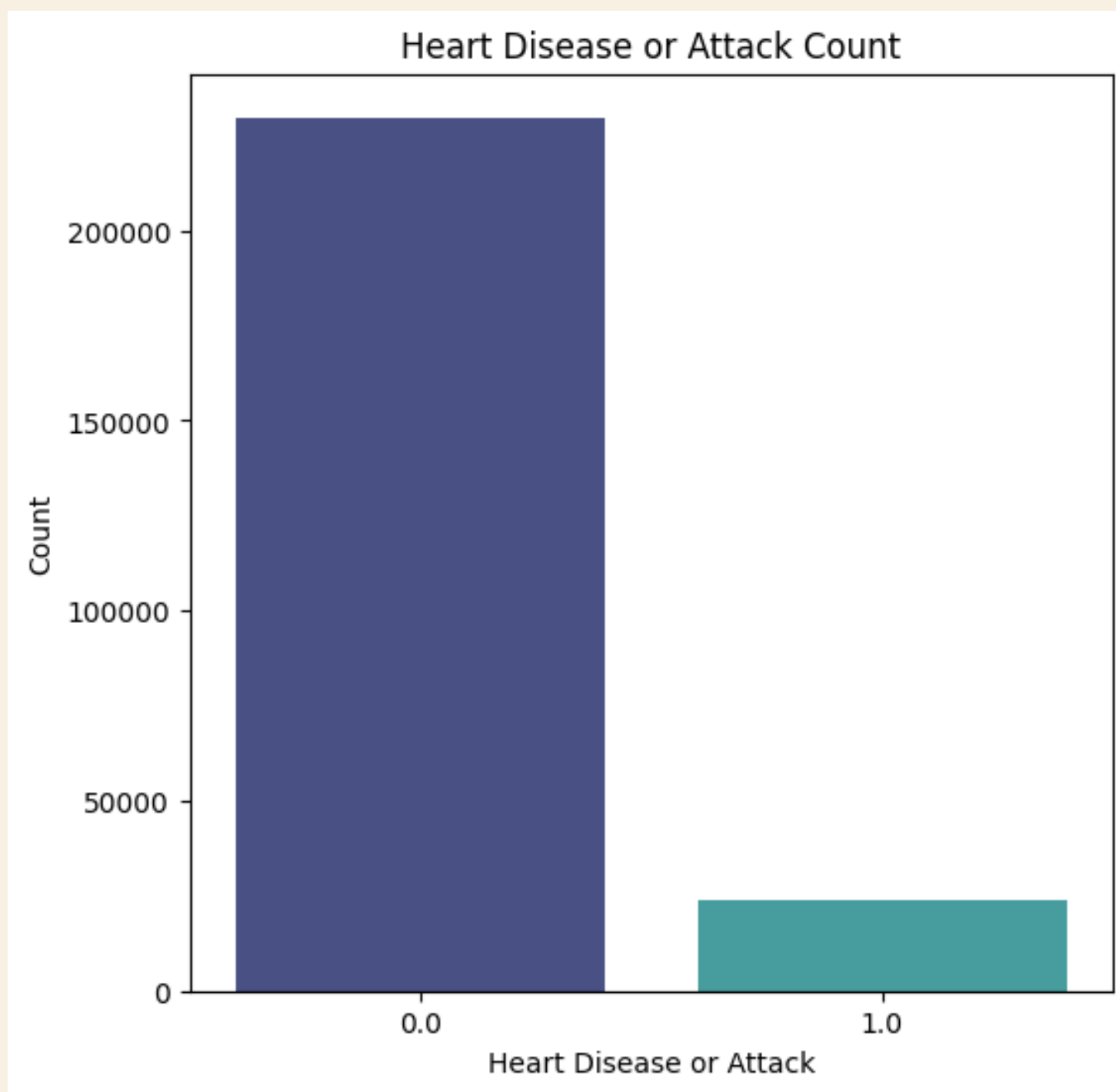
	HeartDiseaseorAttack	HighBP	HighChol	BMI	Smoker	Diabetes	PhysActivity	Fruits	Veggies	HvyAlcoholConsump	MentHlth	PhysHlth	Sex	Age	Education	Income
0	0.0	1.0	1.0	40.0	1.0	0.0	0.0	0.0	1.0	0.0	18.0	15.0	0.0	9.0	4.0	3.0
1	0.0	0.0	0.0	25.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	7.0	6.0	1.0
2	0.0	1.0	1.0	28.0	0.0	0.0	0.0	1.0	0.0	0.0	30.0	30.0	0.0	9.0	4.0	8.0
3	0.0	1.0	0.0	27.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	11.0	3.0	6.0
4	0.0	1.0	1.0	24.0	0.0	0.0	1.0	1.0	1.0	0.0	3.0	0.0	0.0	11.0	5.0	4.0

[140] heart.describe()

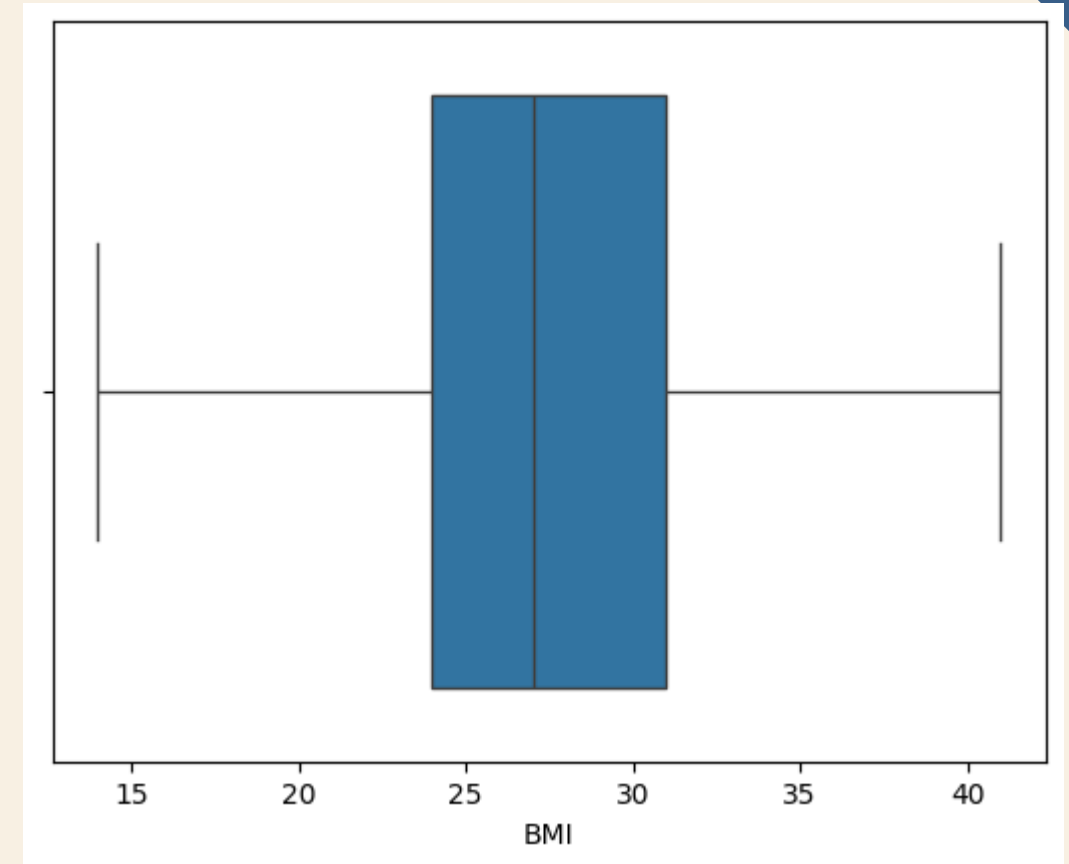
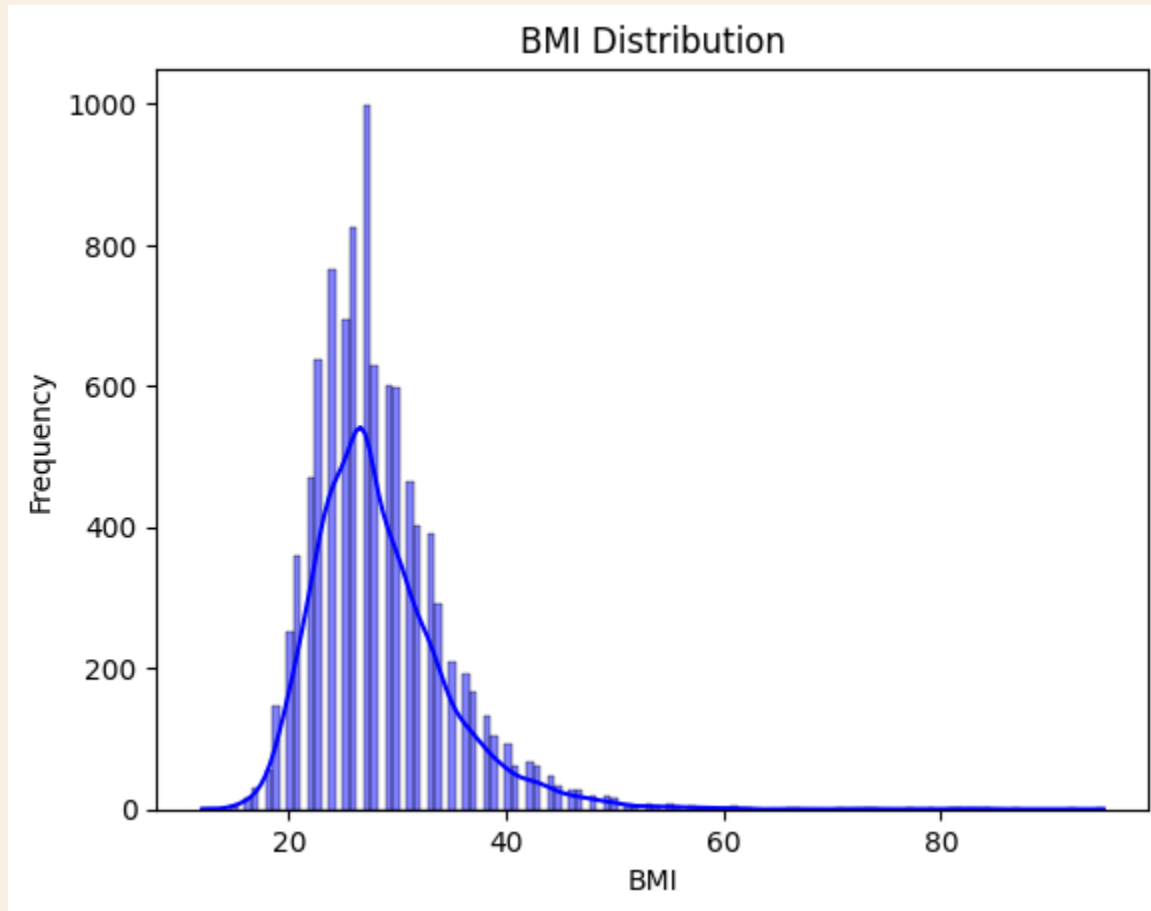
	HeartDiseaseorAttack	HighBP	HighChol	BMI	Smoker	Diabetes	PhysActivity	Fruits	Veggies	HvyAlcoholConsump	MentHlth	PhysHlth
count	253680.000000	253680.000000	253680.000000	253680.000000	253680.000000	253680.000000	253680.000000	253680.000000	253680.000000	253680.000000	253680.000000	253680.000000
mean	0.094186	0.429001	0.424121	28.382364	0.443169	0.296921	0.756544	0.634256	0.811420	0.056197	3.184772	4.242081
std	0.292087	0.494934	0.494210	6.608694	0.496761	0.698160	0.429169	0.481639	0.391175	0.230302	7.412847	8.717951
min	0.000000	0.000000	0.000000	12.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	24.000000	0.000000	0.000000	1.000000	0.000000	1.000000	0.000000	0.000000	0.000000
50%	0.000000	0.000000	0.000000	27.000000	0.000000	0.000000	1.000000	1.000000	1.000000	0.000000	0.000000	0.000000
75%	0.000000	1.000000	1.000000	31.000000	1.000000	0.000000	1.000000	1.000000	1.000000	0.000000	2.000000	3.000000
max	1.000000	1.000000	1.000000	98.000000	1.000000	2.000000	1.000000	1.000000	1.000000	1.000000	30.000000	30.000000

STATS

	HeartDiseaseorAttack	HighBP	HighChol	BMI	Smoker	Diabetes	PhysActivity	Fruits	Veggies	HvyAlcoholConsump	MentHlth	PhysHlth	Sex	Age	Education	Income
mean	1.000000	0.750345	0.701168	29.466622	0.619470	0.687231	0.640355	0.604696	0.763906	0.035492	4.670322	9.154439	0.572887	10.131210	4.745951	5.148161
std	0.000000	0.432822	0.457756	6.740591	0.485527	0.935105	0.479906	0.488926	0.424690	0.185023	9.192712	11.873898	0.494669	2.218853	1.061990	2.198956
min	1.000000	0.000000	0.000000	12.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	1.000000	1.000000
25%	1.000000	1.000000	0.000000	25.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	9.000000	4.000000	3.000000
50%	1.000000	1.000000	1.000000	28.000000	1.000000	0.000000	1.000000	1.000000	1.000000	0.000000	0.000000	2.000000	1.000000	10.000000	5.000000	5.000000
75%	1.000000	1.000000	1.000000	33.000000	1.000000	2.000000	1.000000	1.000000	1.000000	0.000000	4.000000	20.000000	1.000000	12.000000	6.000000	7.000000
max	1.000000	1.000000	1.000000	98.000000	1.000000	2.000000	1.000000	1.000000	1.000000	1.000000	30.000000	30.000000	1.000000	13.000000	6.000000	8.000000

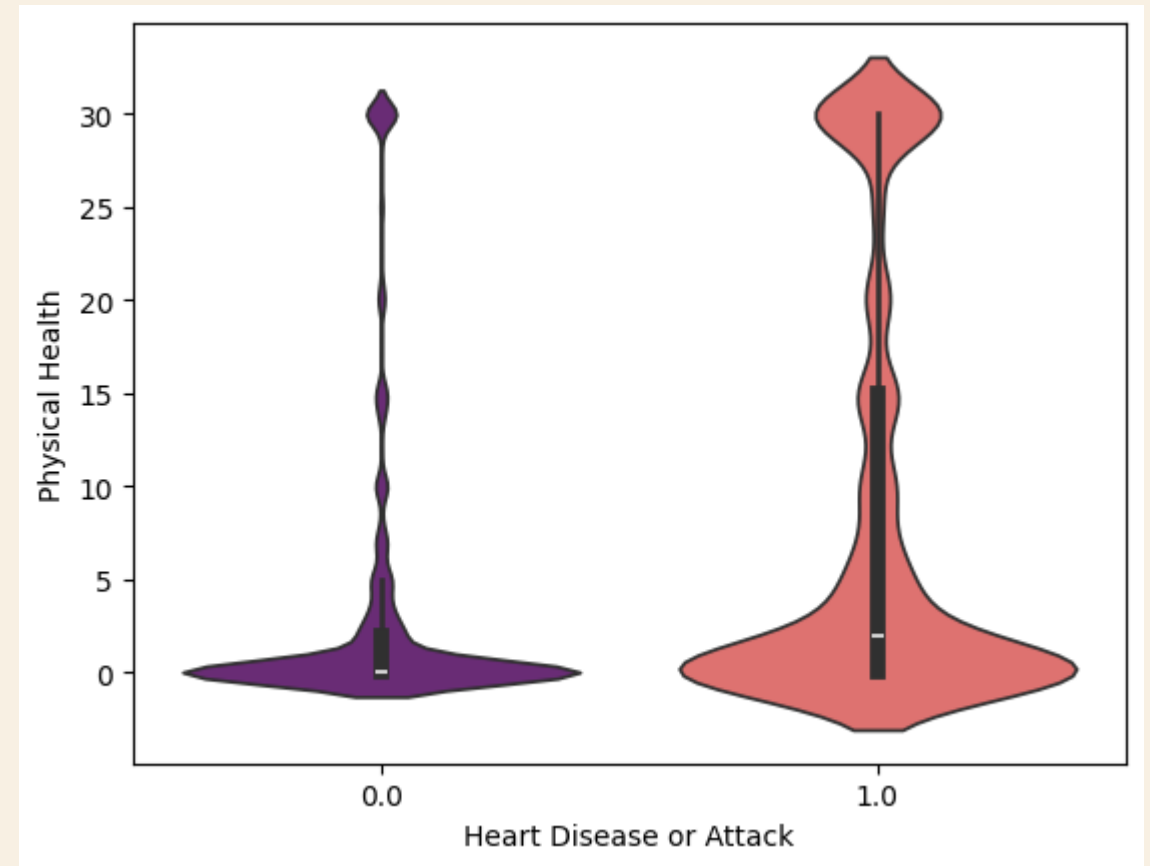
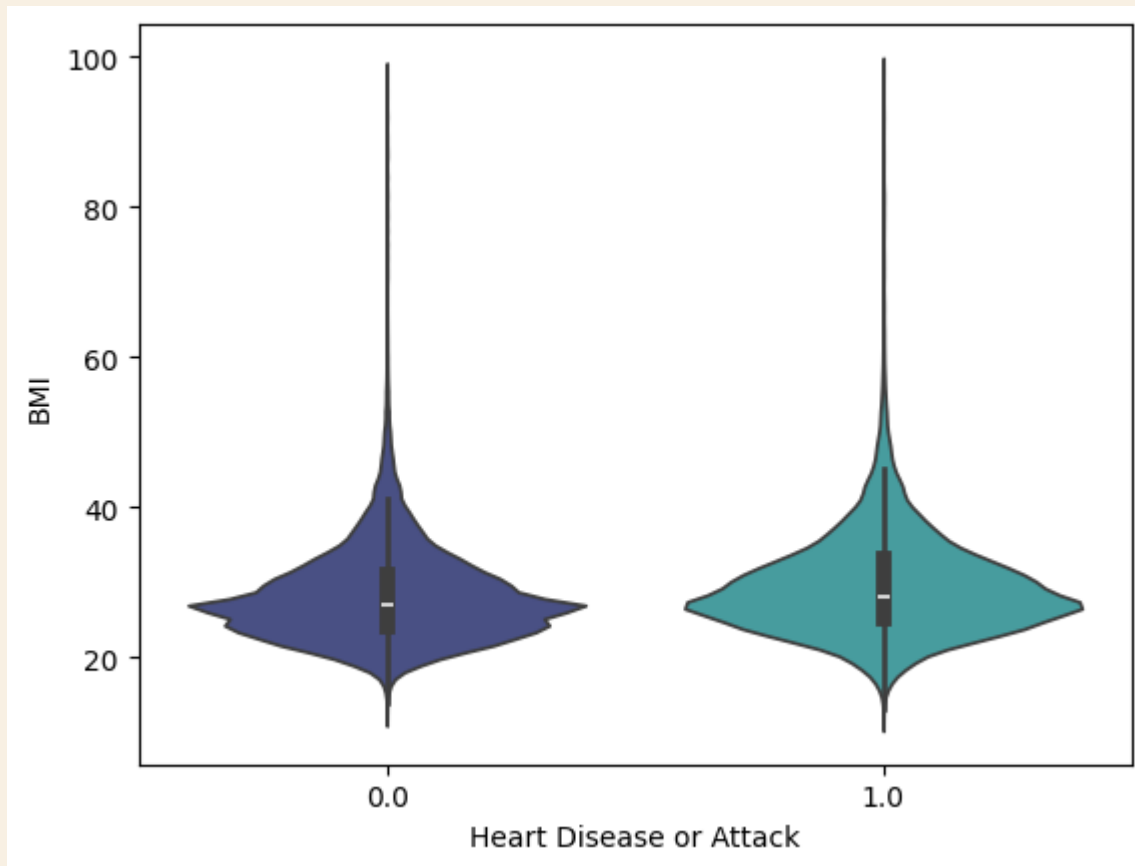


BMI ANALYSIS



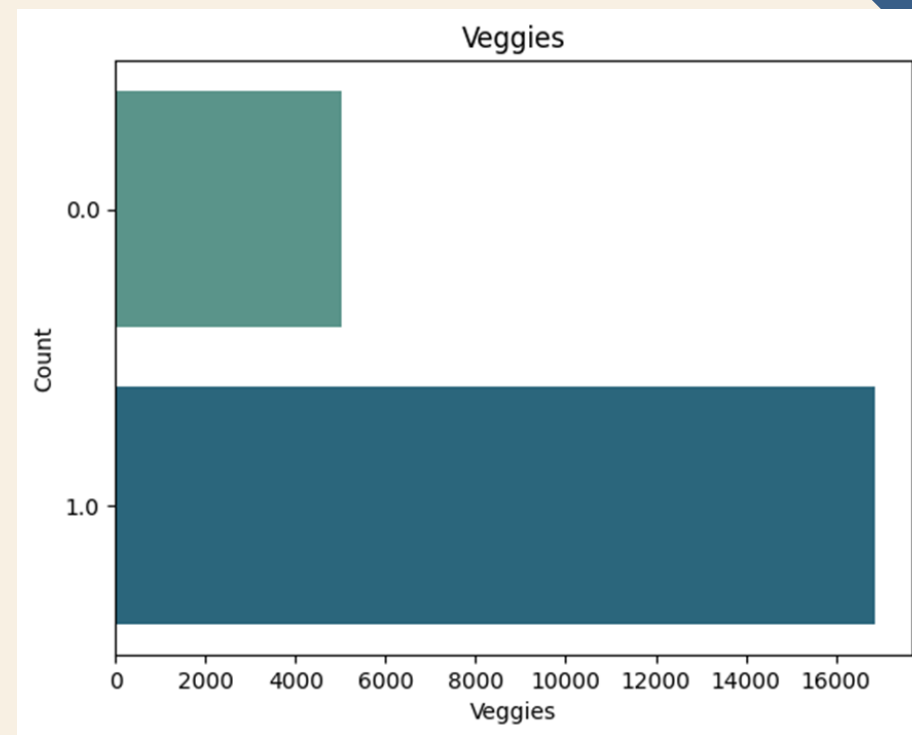
BMI	
HeartDiseaseorAttack	
0.000000	27.474992
1.000000	28.490699

BMI ; PHYSICAL HEALTH



PhysActivity	
HeartDiseaseorAttack	
0.000000	0.768625
1.000000	0.640355

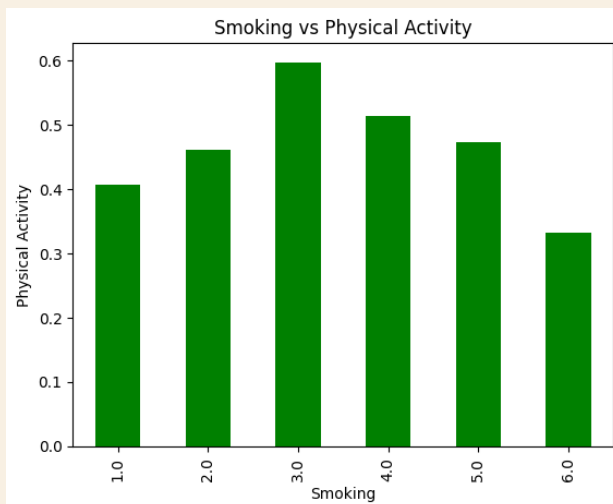
Smoker	
HeartDiseaseorAttack	
0.000000	0.424837
1.000000	0.619470



Veggies	
HeartDiseaseorAttack	
0.000000	0.816360
1.000000	0.763906

HvyAlcoholConsump	
HeartDiseaseorAttack	
0.000000	0.058350
1.000000	0.035492

HighChol	
HeartDiseaseorAttack	
0.000000	0.395314
1.000000	0.701168



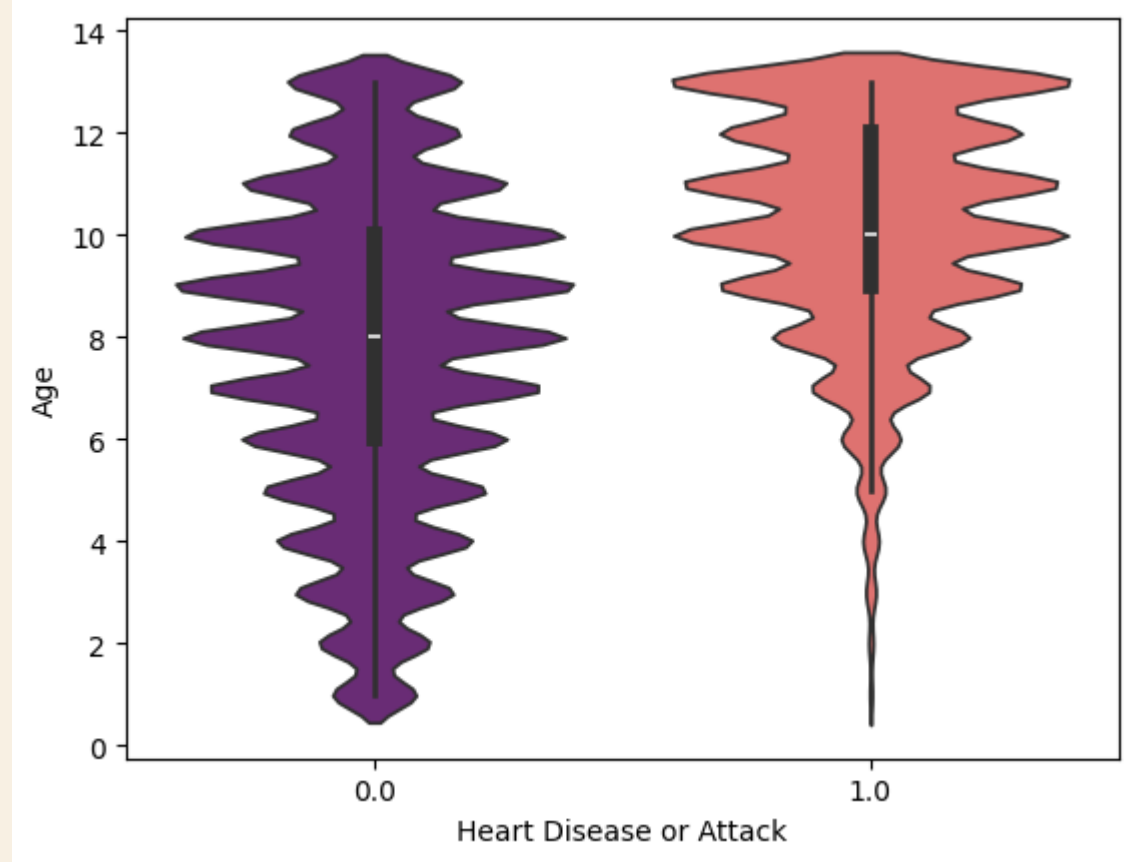
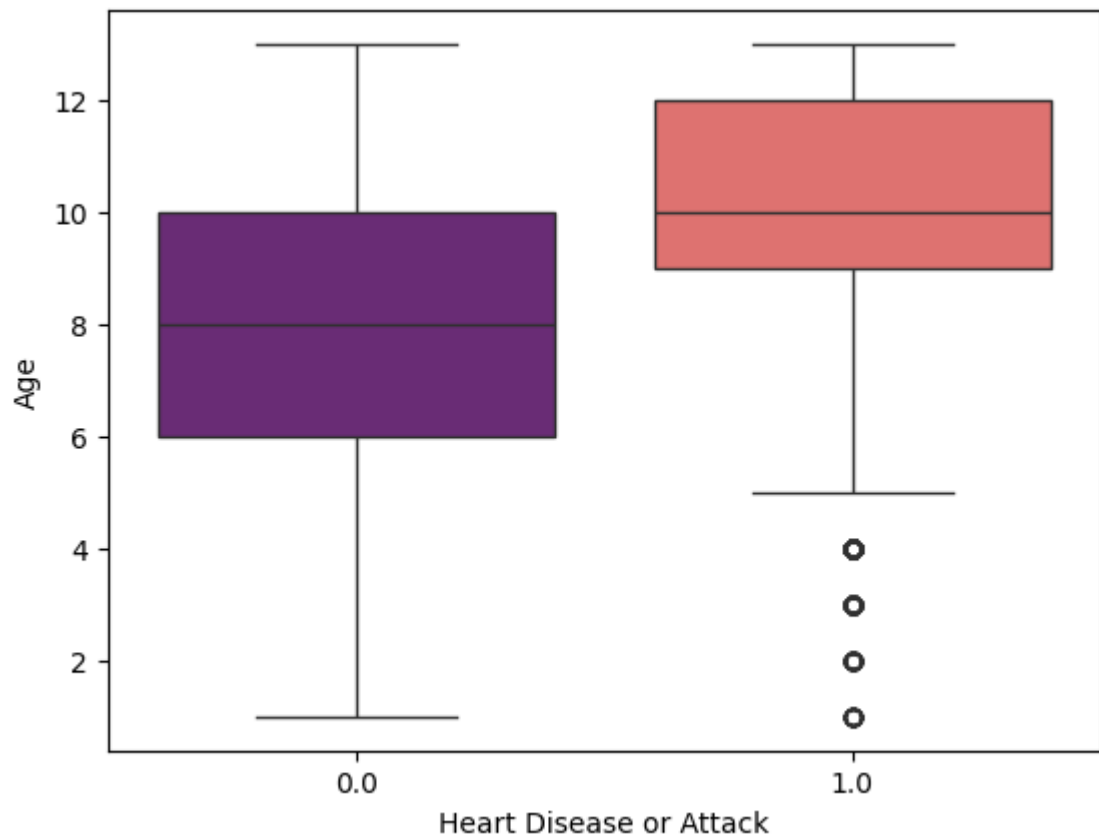
HighBP	
HeartDiseaseorAttack	
0.000000	0.395588
1.000000	0.750345



	HeartDiseaseorAttack	HighBP	HighChol	BMI	Smoker	Diabetes	PhysActivity	Fruits	Veggies	HvyAlcoholConsump	MentHlth	PhysHlth	Sex	Age	Education
HeartDiseaseorAttack	1.000000	0.209361	0.180765	0.052904	0.114441	0.180272	-0.087299	-0.019790	-0.039167	-0.028991	0.064621	0.181698	0.086096	0.221618	-0.099600
HighBP	0.209361	1.000000	0.298199	0.213748	0.096991	0.271596	-0.125267	-0.040555	-0.061266	-0.003972	0.056456	0.161212	0.052207	0.344452	-0.141358
HighChol	0.180765	0.298199	1.000000	0.106722	0.091299	0.209085	-0.078046	-0.040859	-0.039874	-0.011543	0.062069	0.121751	0.031205	0.272318	-0.070802
BMI	0.052904	0.213748	0.106722	1.000000	0.013804	0.224379	-0.147294	-0.087518	-0.062275	-0.048736	0.085310	0.121141	0.042950	-0.036618	-0.103932
Smoker	0.114441	0.096991	0.091299	0.013804	1.000000	0.062914	-0.087401	-0.077666	-0.030678	0.101619	0.092196	0.116460	0.093662	0.120641	-0.161955
Diabetes	0.180272	0.271596	0.209085	0.224379	0.062914	1.000000	-0.121947	-0.042192	-0.058972	-0.057882	0.073507	0.176287	0.031040	0.185026	-0.130517
PhysActivity	-0.087299	-0.125267	-0.078046	-0.147294	-0.087401	-0.121947	1.000000	0.142756	0.153150	0.012392	-0.125587	-0.219230	0.032482	-0.092511	0.199658
Fruits	-0.019790	-0.040555	-0.040859	-0.087518	-0.077666	-0.042192	0.142756	1.000000	0.254342	-0.035288	-0.068217	-0.044633	-0.091175	0.064547	0.110187
Veggies	-0.039167	-0.061266	-0.039874	-0.062275	-0.030678	-0.058972	0.153150	0.254342	1.000000	0.021064	-0.058884	-0.064290	-0.064765	-0.009771	0.154329
HvyAlcoholConsump	-0.028991	-0.003972	-0.011543	-0.048736	0.101619	-0.057882	0.012392	-0.035288	0.021064	1.000000	0.024716	-0.026415	0.005740	-0.034578	0.023997
MentHlth	0.064621	0.056456	0.062069	0.085310	0.092196	0.073507	-0.125587	-0.068217	-0.058884	0.024716	1.000000	0.353619	-0.080705	-0.092068	-0.101830
PhysHlth	0.181698	0.161212	0.121751	0.121141	0.116460	0.176287	-0.219230	-0.044633	-0.064290	-0.026415	0.353619	1.000000	-0.043137	0.099130	-0.155093
Sex	0.086096	0.052207	0.031205	0.042950	0.093662	0.031040	0.032482	-0.091175	-0.064765	0.005740	-0.080705	-0.043137	1.000000	-0.027340	0.019480
Age	0.221618	0.344452	0.272318	-0.036618	0.120641	0.185026	-0.092511	0.064547	-0.009771	-0.034578	-0.092068	0.099130	-0.027340	1.000000	-0.101901
Education	-0.099600	-0.141358	-0.070802	-0.103932	-0.161955	-0.130517	0.199658	0.110187	0.154329	0.023997	-0.101830	-0.155093	0.019480	-0.101901	1.000000
Income	-0.141011	-0.171235	-0.085459	-0.100069	-0.123937	-0.171483	0.198539	0.079929	0.151087	0.053619	-0.209806	-0.266799	0.127141	-0.127775	0.449106



AGE



```
heart_disease.head()
```

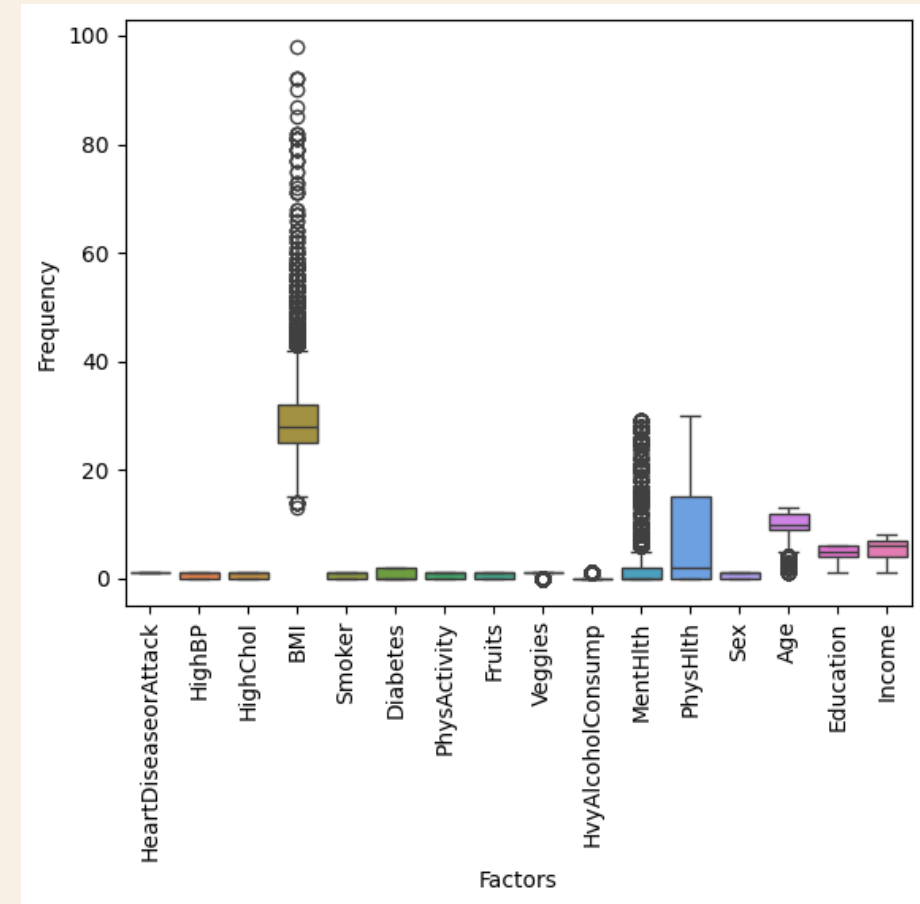
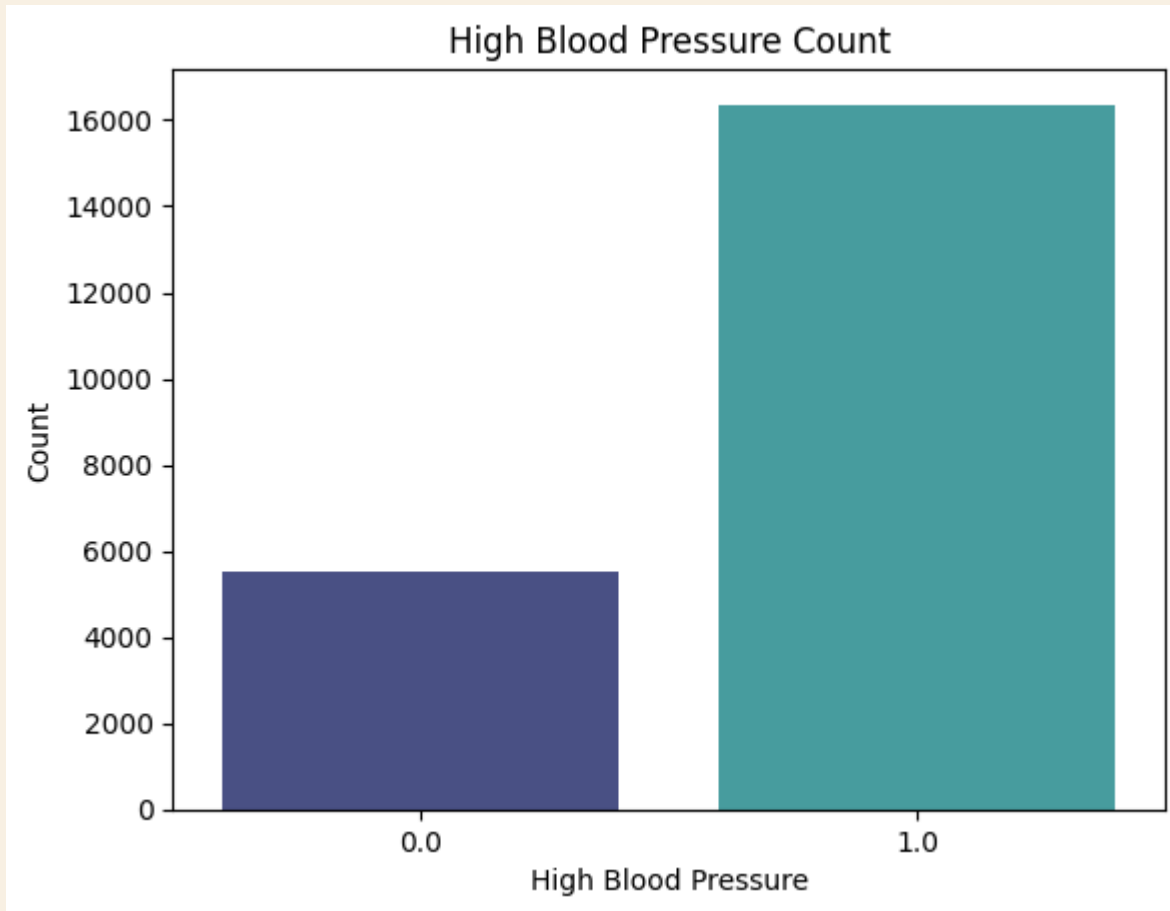


	HeartDiseaseorAttack	HighBP	HighChol	BMI	Smoker	Diabetes	PhysActivity	Fruits	Veggies	HvyAlcoholConsump	MentHlth	PhysHlth	Sex	Age	Education	Income
26	1.0	1.0	1.0	37.0	1.0	2.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	10.0	6.0	5.0
27	1.0	1.0	1.0	28.0	1.0	2.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	12.0	2.0	4.0
47	1.0	1.0	1.0	25.0	1.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0	1.0	10.0	4.0	7.0
50	1.0	1.0	1.0	30.0	1.0	0.0	1.0	0.0	0.0	0.0	10.0	17.0	0.0	9.0	4.0	1.0
58	1.0	1.0	1.0	33.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	11.0	6.0	6.0

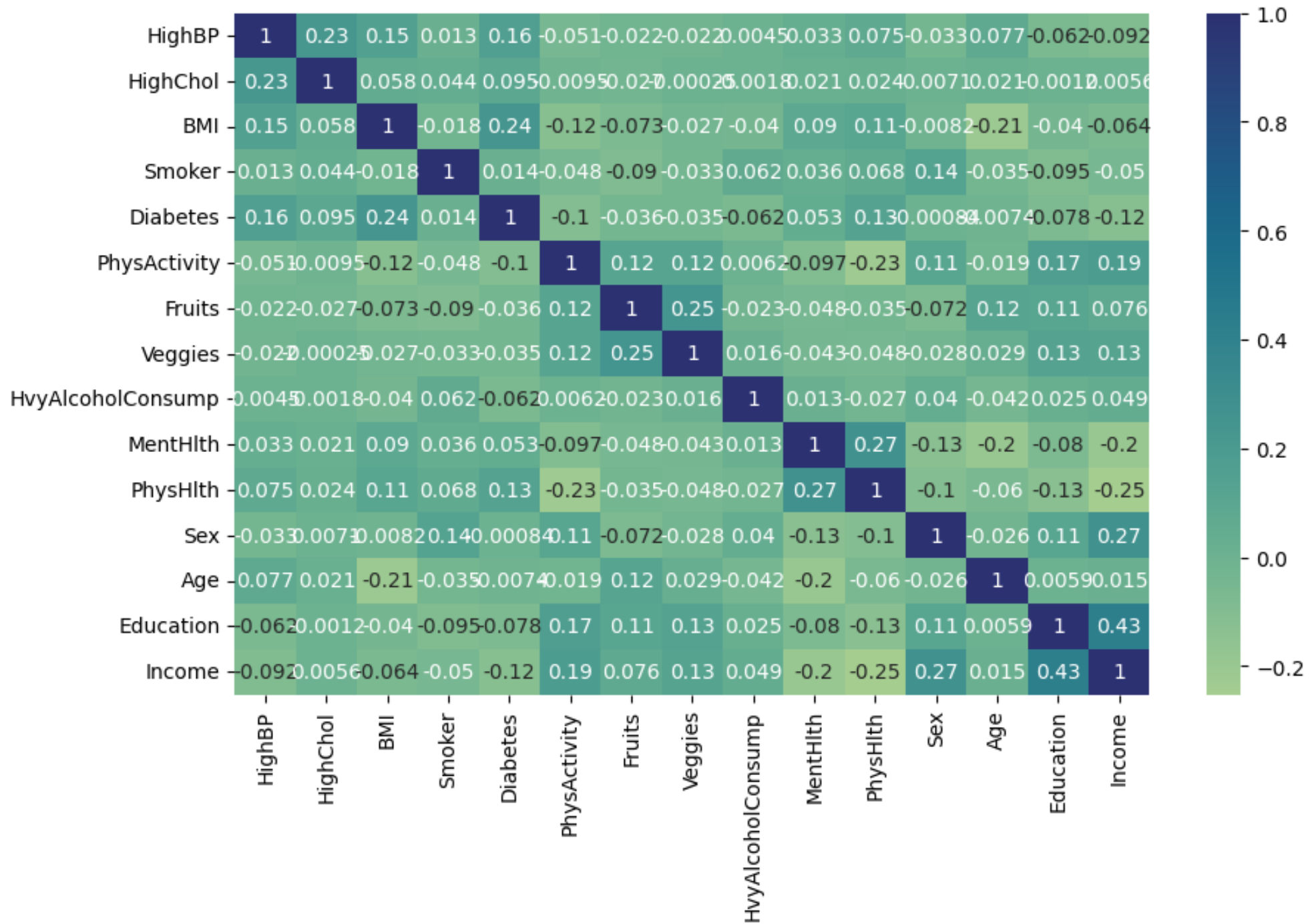
HeartDiseaseorAttack	HighBP	HighChol	BMI	Smoker	Diabetes	PhysActivity	Fruits	Veggies	HvyAlcoholConsump	MentHlth	PhysHlth	Sex	Age	Education	Income			
1.0	1.0	1.0	25.0	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	1.0	13.0	6.0	8.0	6		
			27.0	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	1.0	11.0	6.0	8.0	6		
			26.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	1.0	10.0	6.0	8.0	5		
			29.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	1.0	11.0	6.0	8.0	5		
		0.0	26.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	1.0	10.0	6.0	8.0	5		
				
			28.0	0.0	2.0	0.0	1.0	1.0	0.0	10.0	1.0	1.0	6.0	5.0	1.0	1		
											0.0	30.0	0.0	13.0	6.0	2.0	1	
													0.0	0.0	13.0	4.0	4.0	1
									0.0	0.0	15.0	30.0	0.0	9.0	4.0	5.0	1	
						1.0	0.0	1.0	0.0	0.0	20.0	0.0	9.0	4.0	3.0	1		

21337 rows × 1 columns

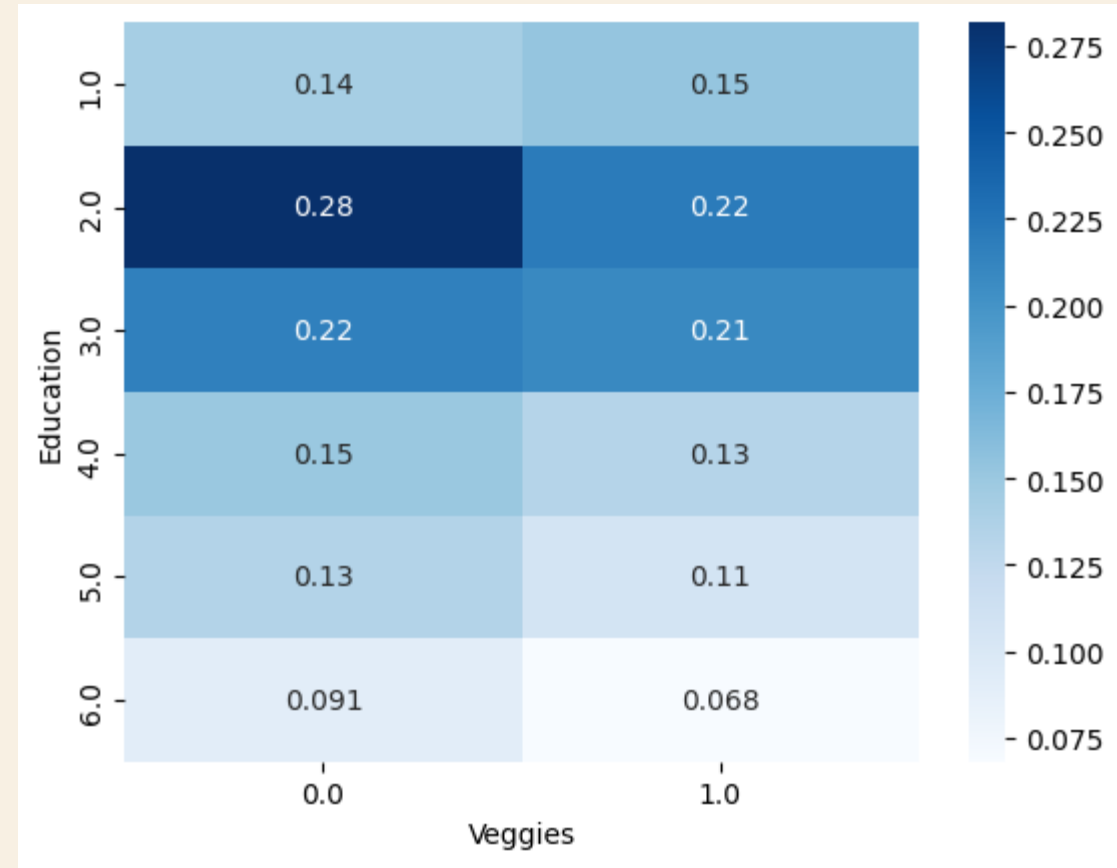
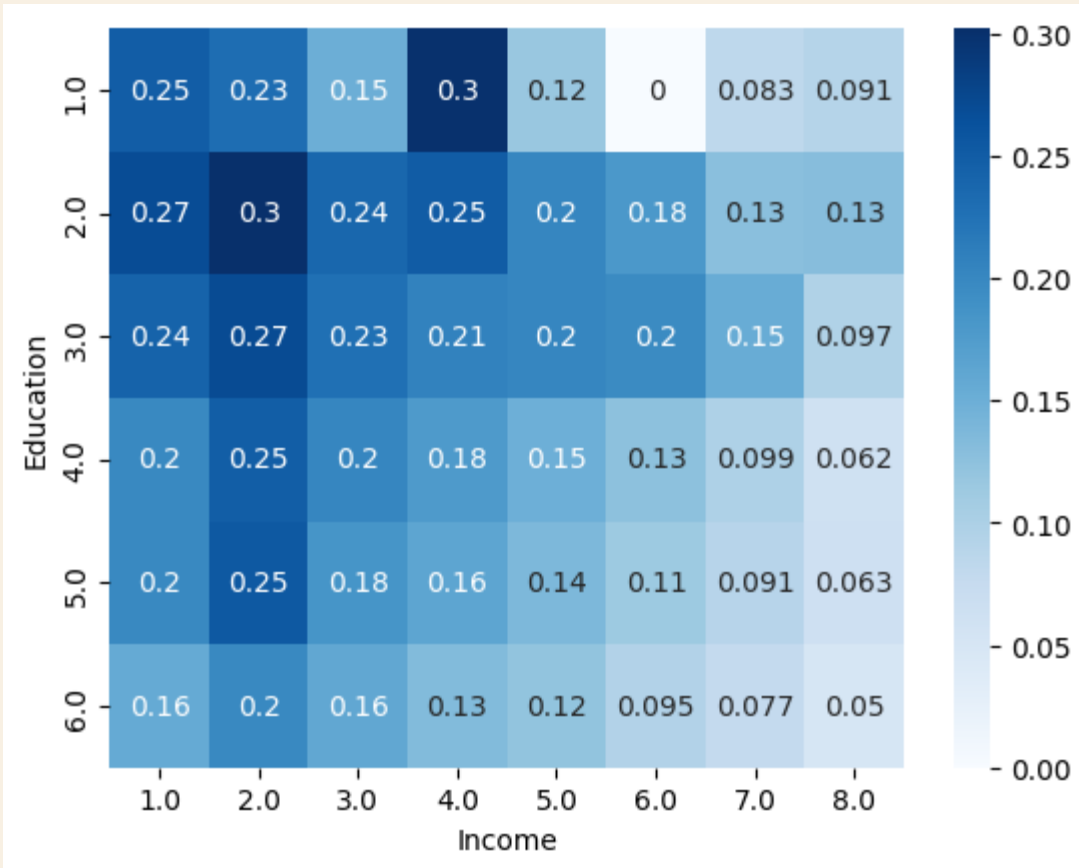
HIGH BLOOD PRESSURE AMONG THE SICK AND FACTORS FOR HEART DISEASE



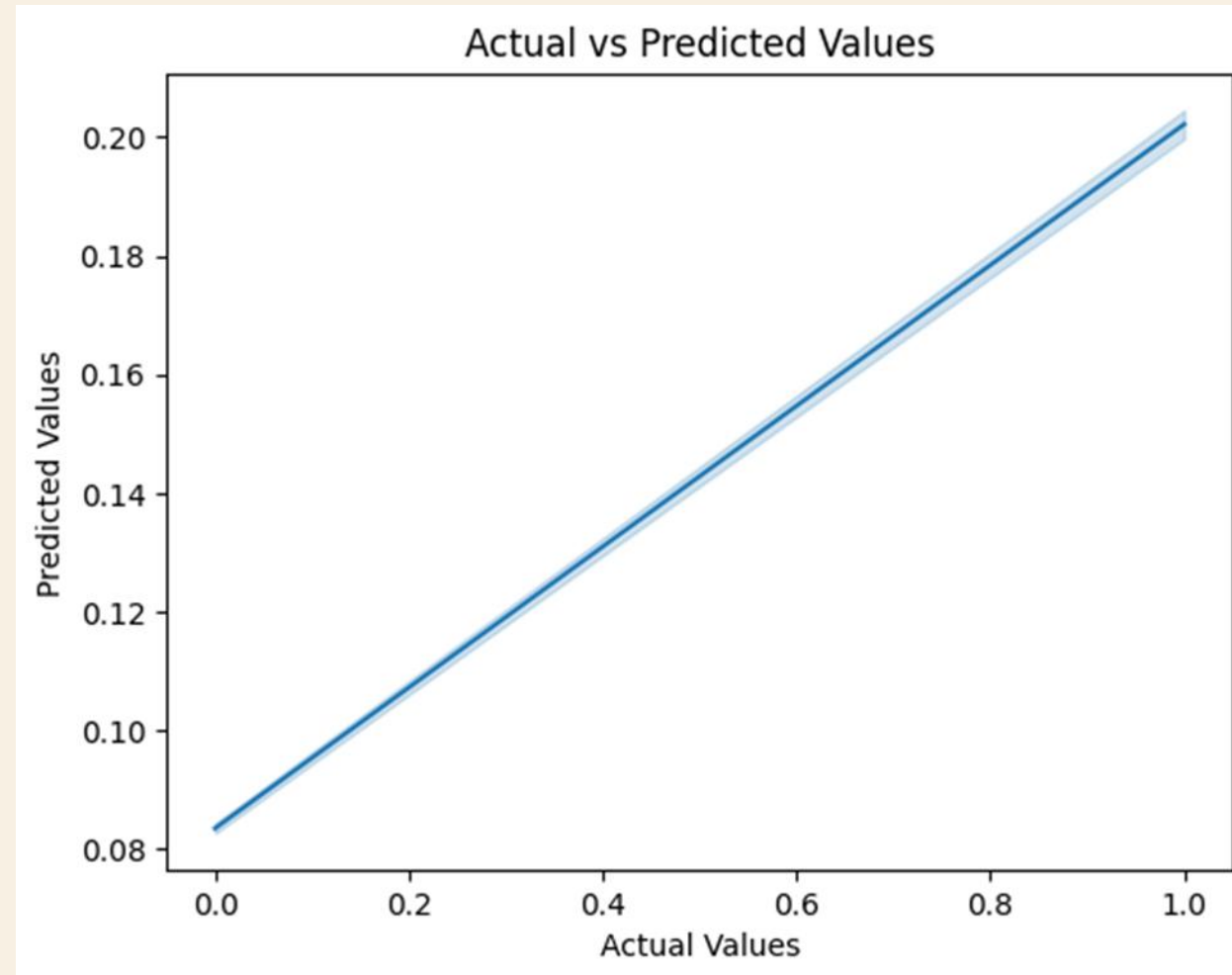
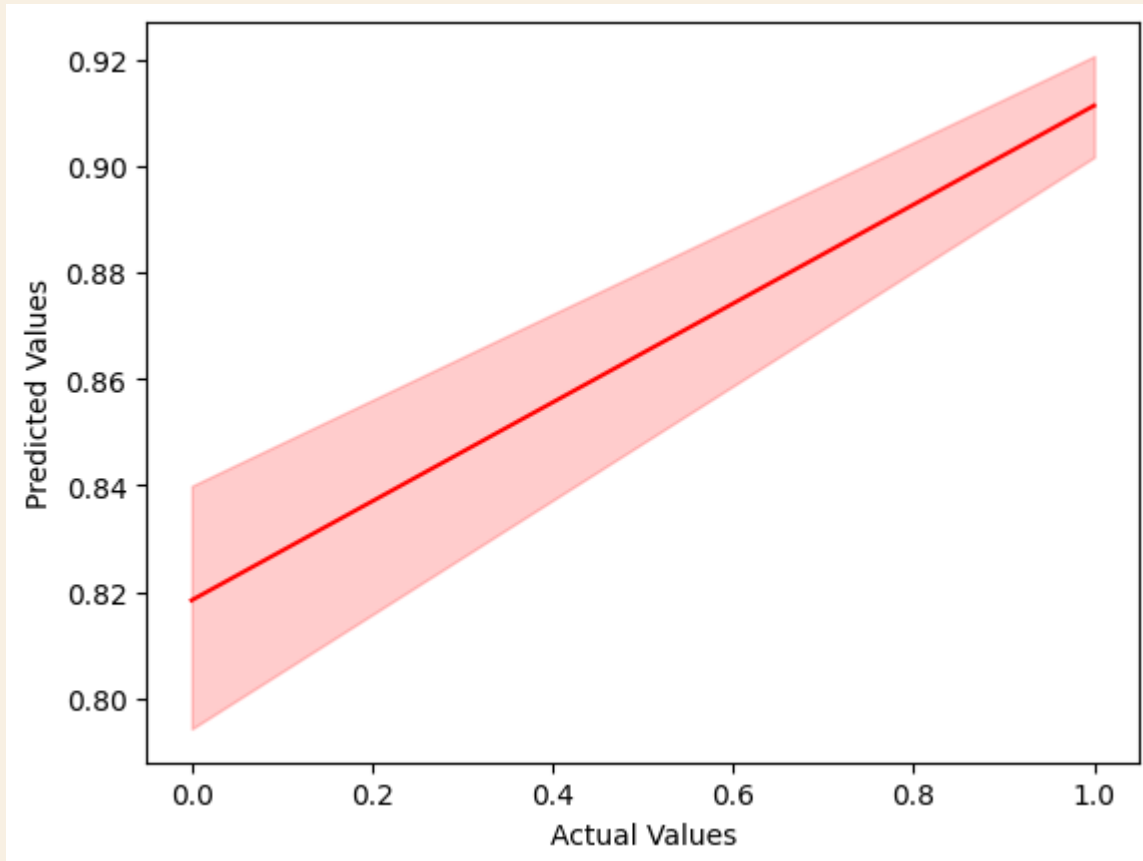
BMI and MentHealth have lots of outliers.



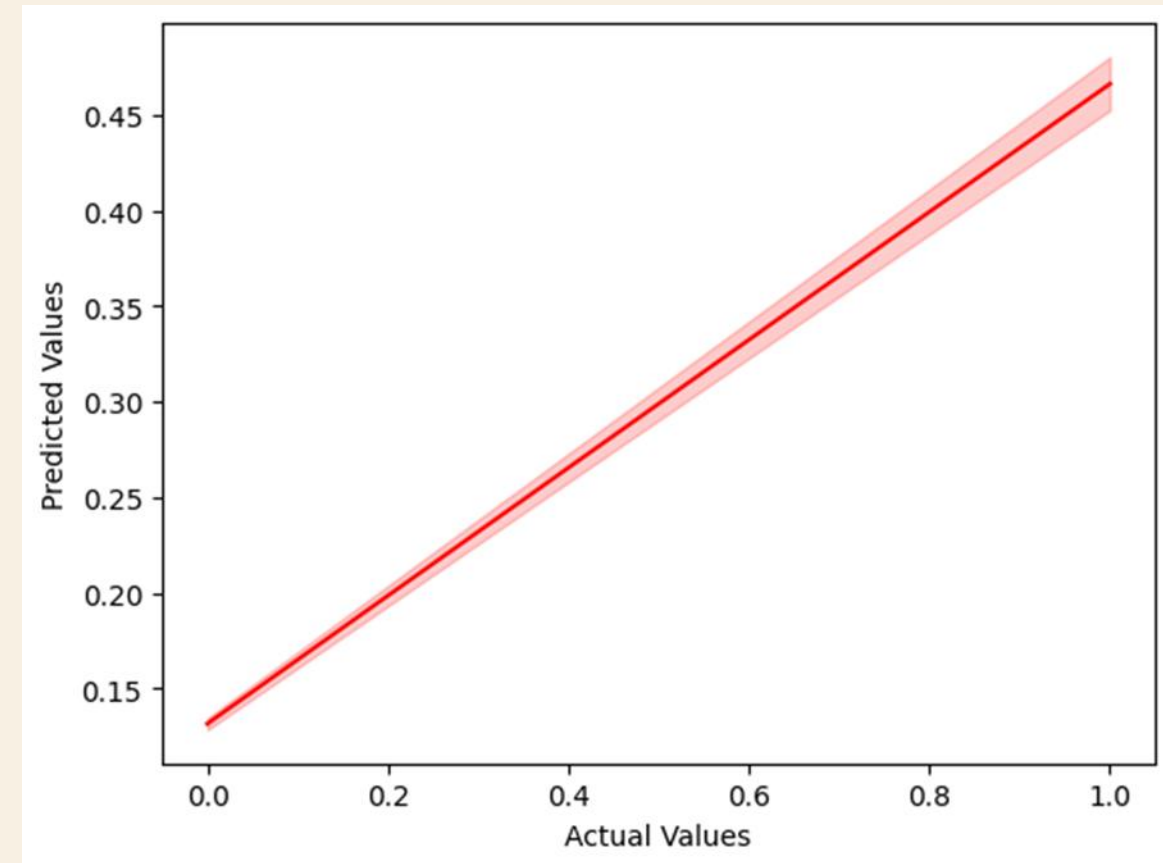
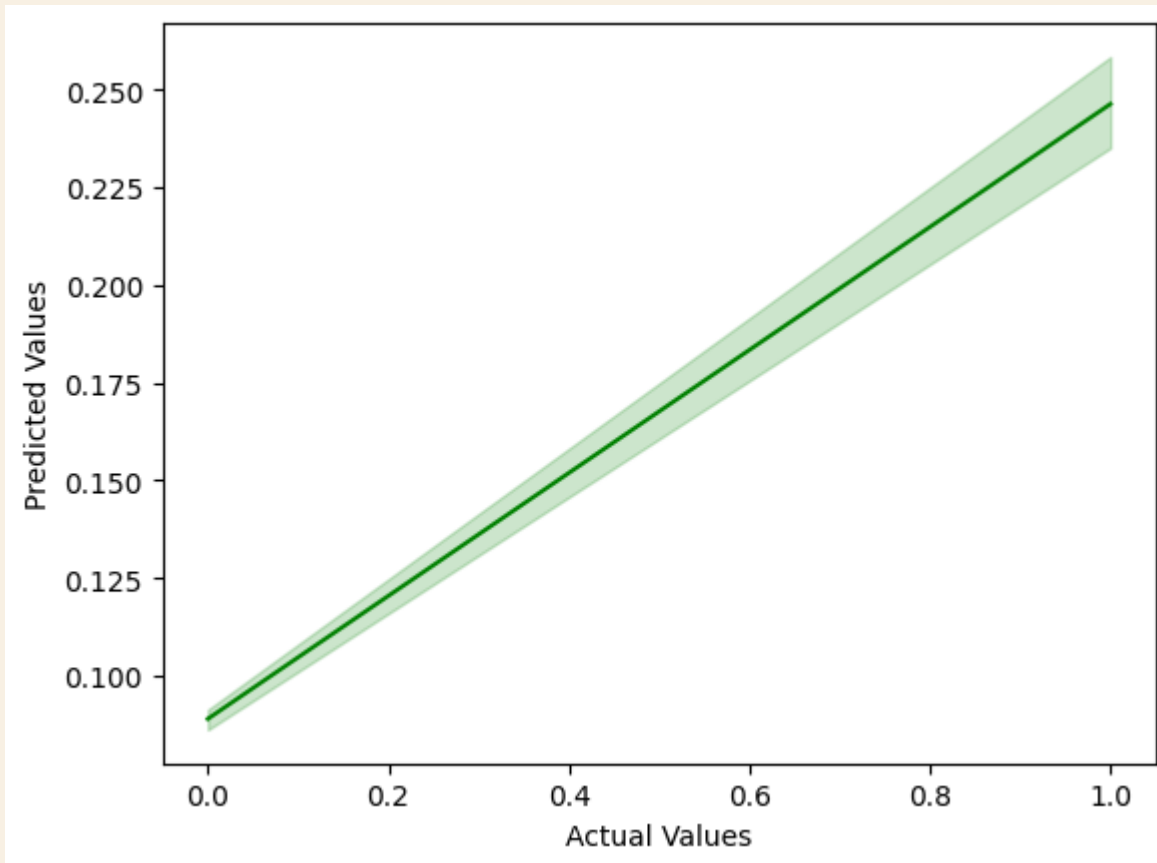
EDUCATION; INCOME; VEGETARIANISM



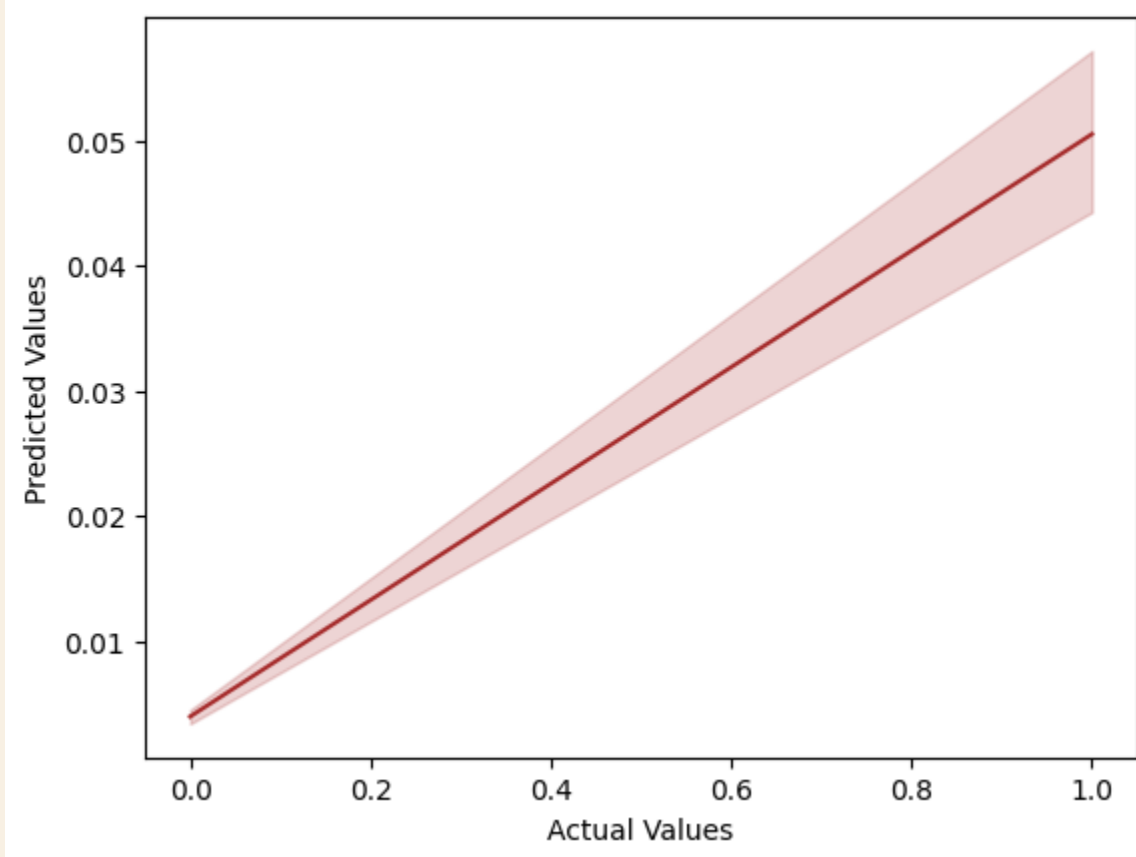
RANDOM FOREST AND LINEAR REGRESSION FOR HEARTH DISEASE OCCURENCE



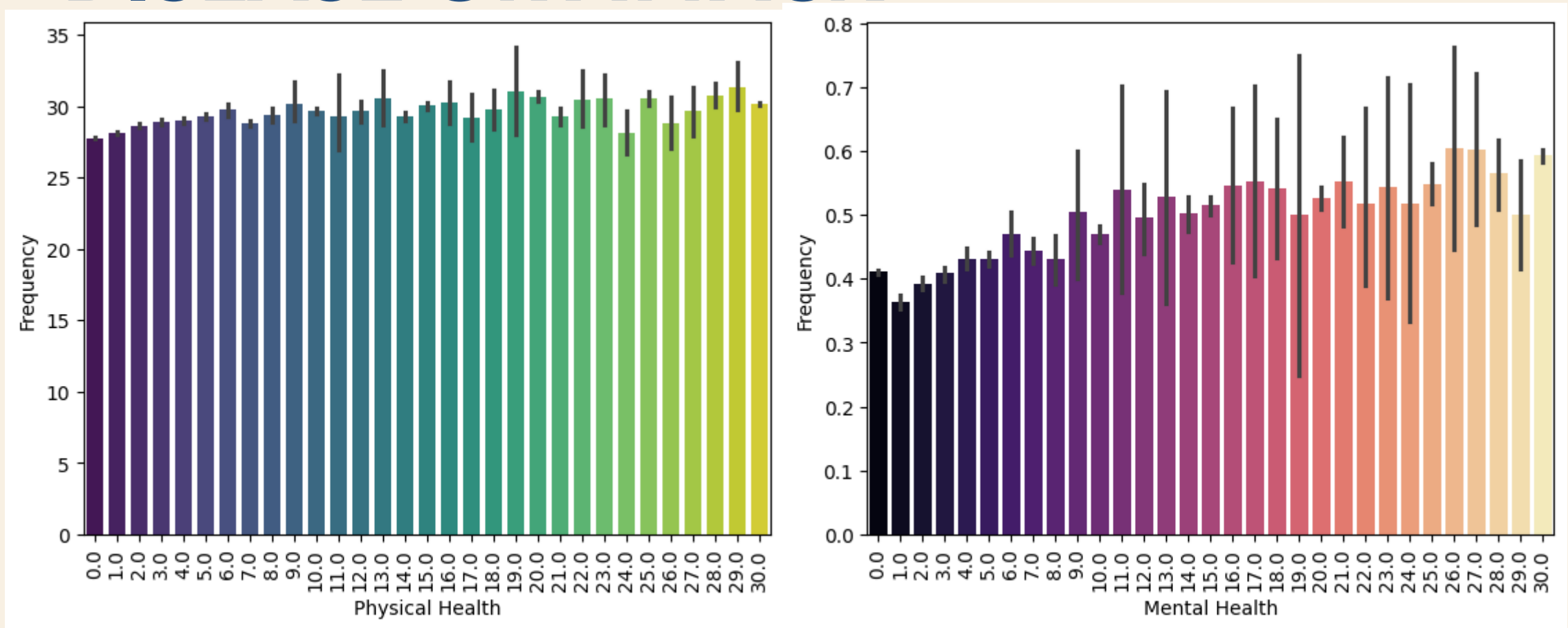
DECISION TREE AND GAUSSIAN NB



GRADIENT BOOST



BY INDIVIDUALS WITHOUT HEART DISEASE OR ATTACK





THANK YOU

NAZIM ATAKAN ERDOGAN