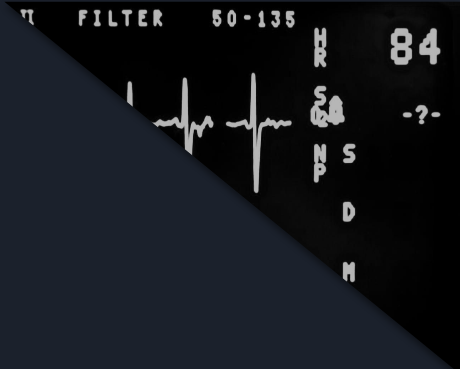


Predicting Cardiovascular Disease

Module 5 Project
By Desmond Webb

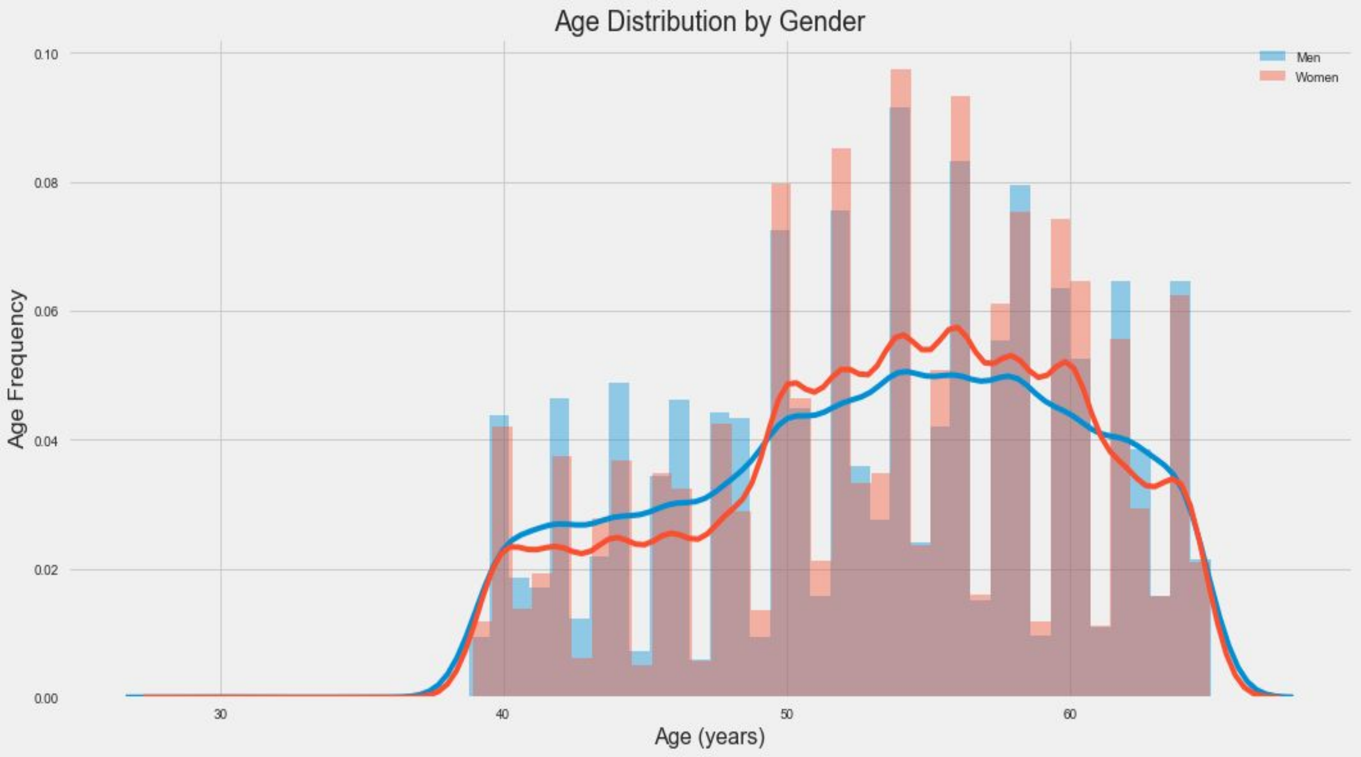
Project Scope

- Present Dataset
- Risk Factor Relationships
- Model Development

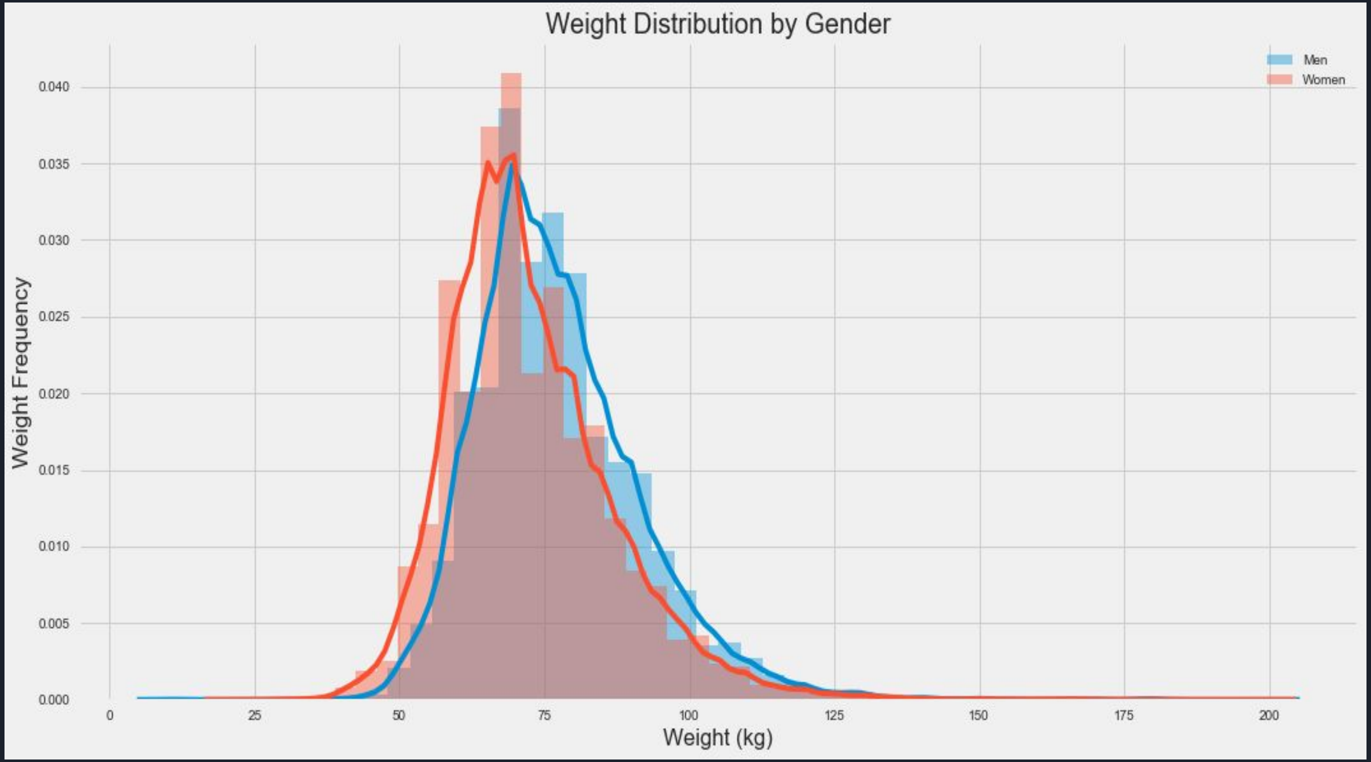
Project Objectives

- Cardiovascular disease classification model (CD)
- Explore risk factor contributions to CD
- Explore risk factor interactions

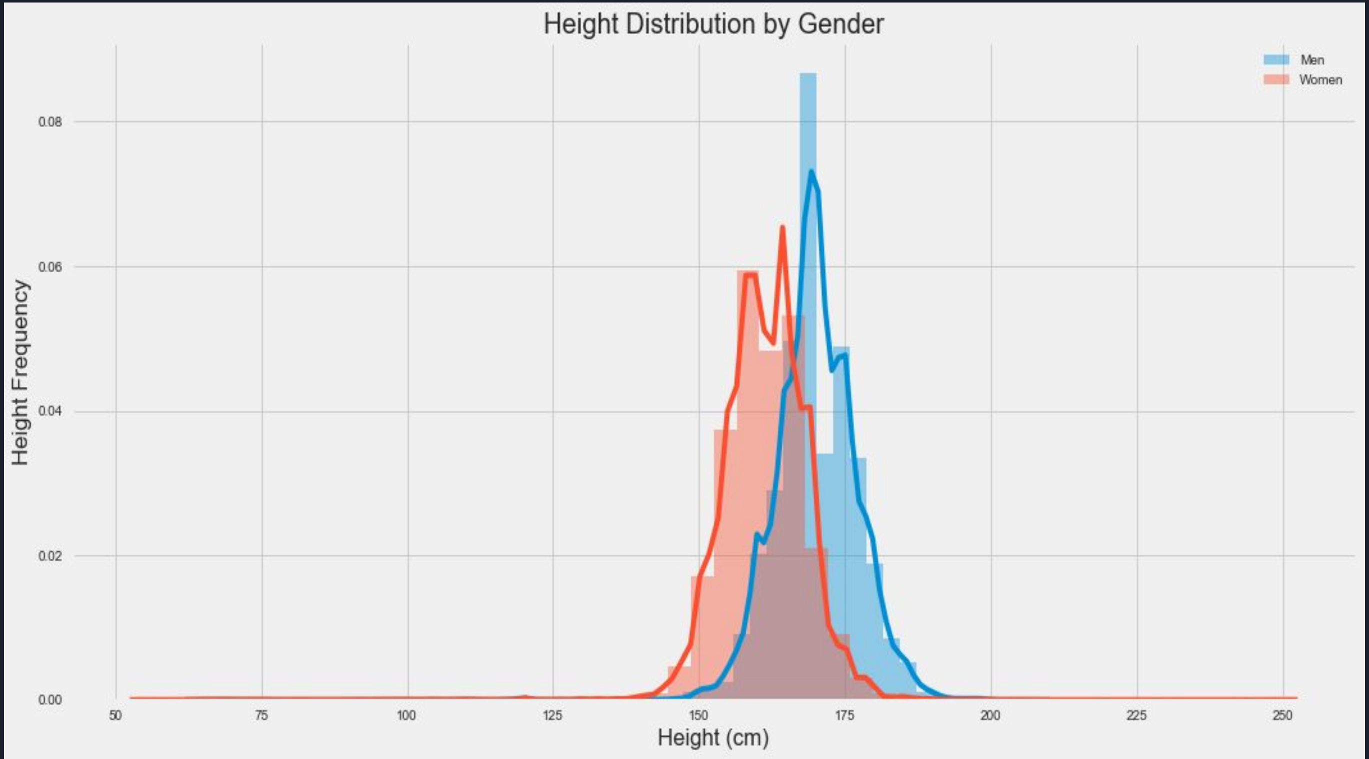
Population Trends



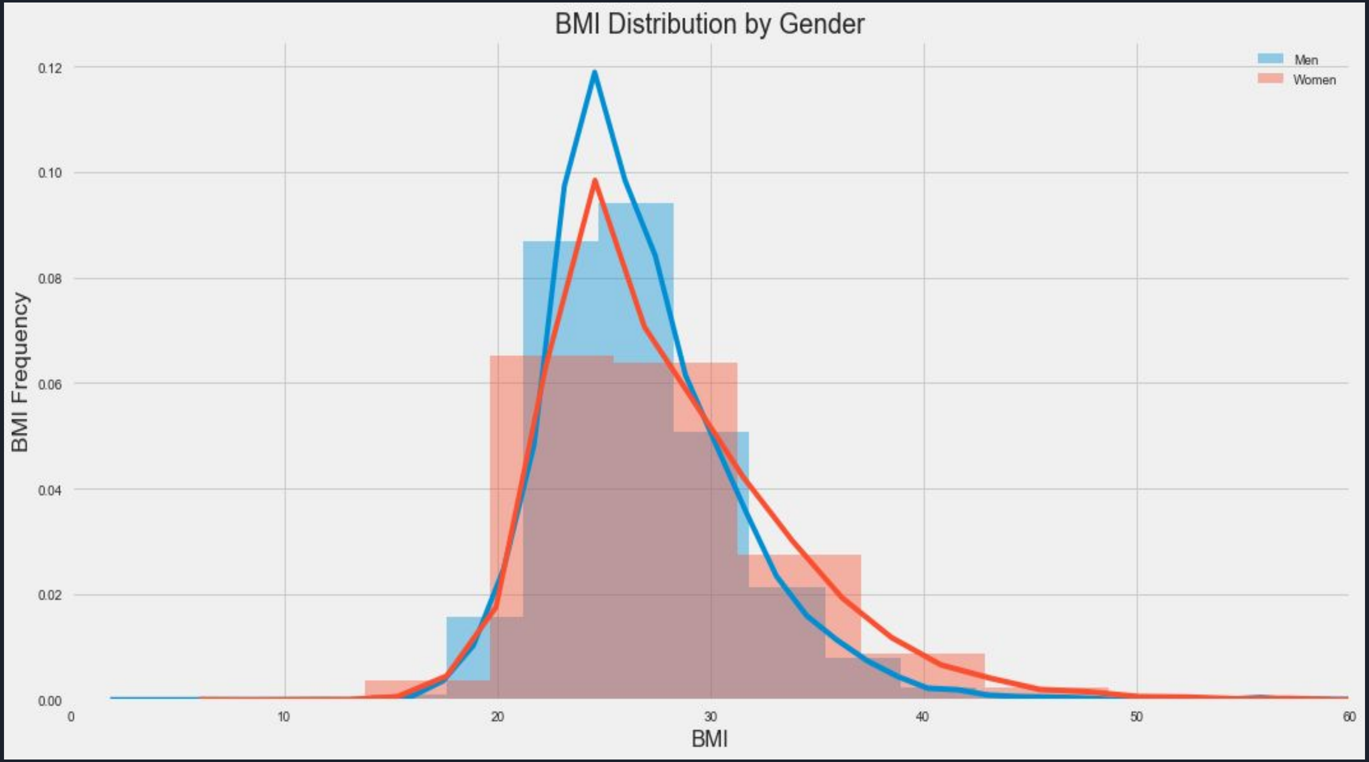
Population Trends



Population Trends



Population Trends



Risk Factor Interactions

✓ cholesterol and physical activity vs CD [1]

- substance use and physical activity [2]

✓ significant relationship

CD Factors

Height/Weight[3]:

age	cholesterol
gender*	glucose
height	smoking
weight	alcohol
ap_hi	physical activity
ap_lo	

BMI[4]:

age	cholesterol
gender	ap_hi
ap_lo	glucose
smoking	alcohol
physical activity	BMI

* non-significant factor

Classification

Height/Weight:

Accuracy: 74.49%[5]

Improvement: 23.78%

BMI:

Accuracy: 73.95%[6]

Improvement: 23.57%

Recommendations:

- BMI model*
- diagnostic training tool
- behavioral therapy

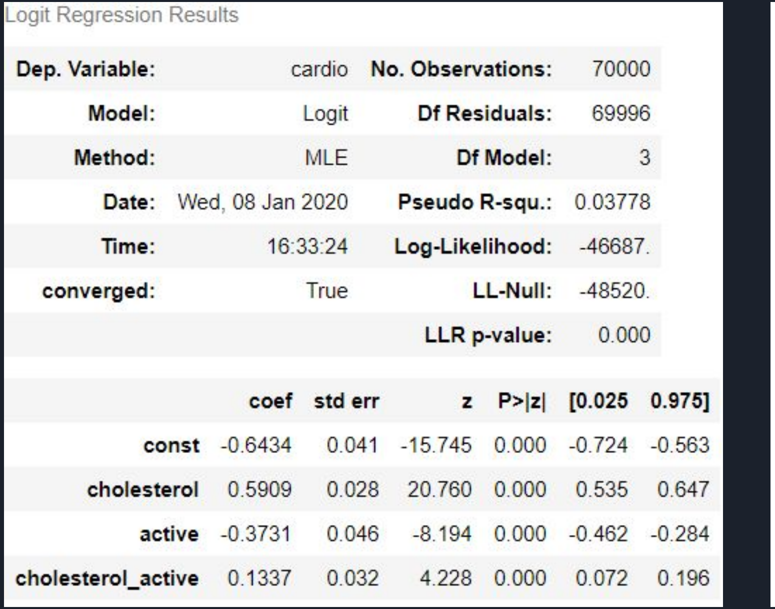
Where To Improve?

- Demographic info
- More specific data collection:
  - glucose (mM/L)
  - cholesterol (mg/dL)
  - alcohol (number of drinks a day)
  - smoking (number of packs per day)
  - physical activity (number of days per week)
  - substance use
- Trunk-fat measurements

Thank you!

Appendix

1. cholesterol * physical activity vs CD

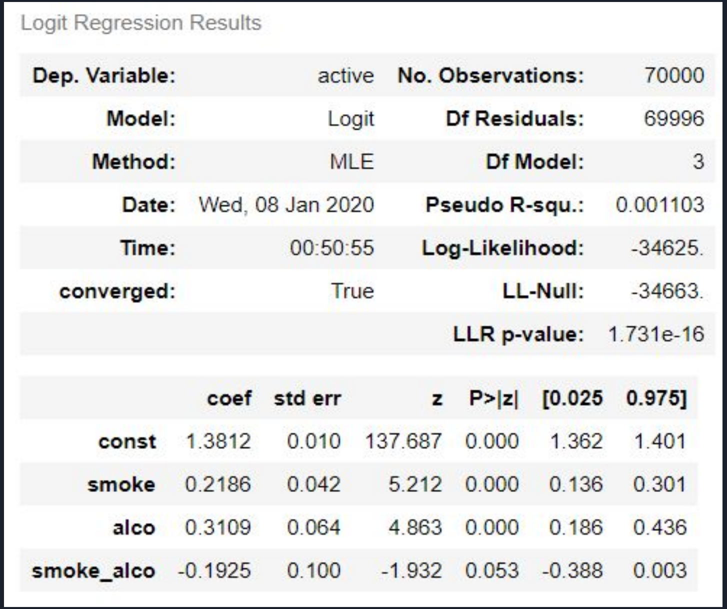


Logit Regression Results

Dep. Variable:	cardio	**No. Observations:**	70000
Model:	Logit	**Df Residuals:**	69996
Method:	MLE	**Df Model:**	3
Date:	Wed, 08 Jan 2020	**Pseudo R-squ.:**	0.03778
Time:	16:33:24	**Log-Likelihood:**	-46687.
converged:	True	**LL-Null:**	-48520.
		LLR p-value:	0.000

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
const	-0.6434	0.041	-15.745	0.000	-0.724	-0.563
cholesterol	0.5909	0.028	20.760	0.000	0.535	0.647
active	-0.3731	0.046	-8.194	0.000	-0.462	-0.284
cholesterol_active	0.1337	0.032	4.228	0.000	0.072	0.196

2. smoking * alcohol use vs CD

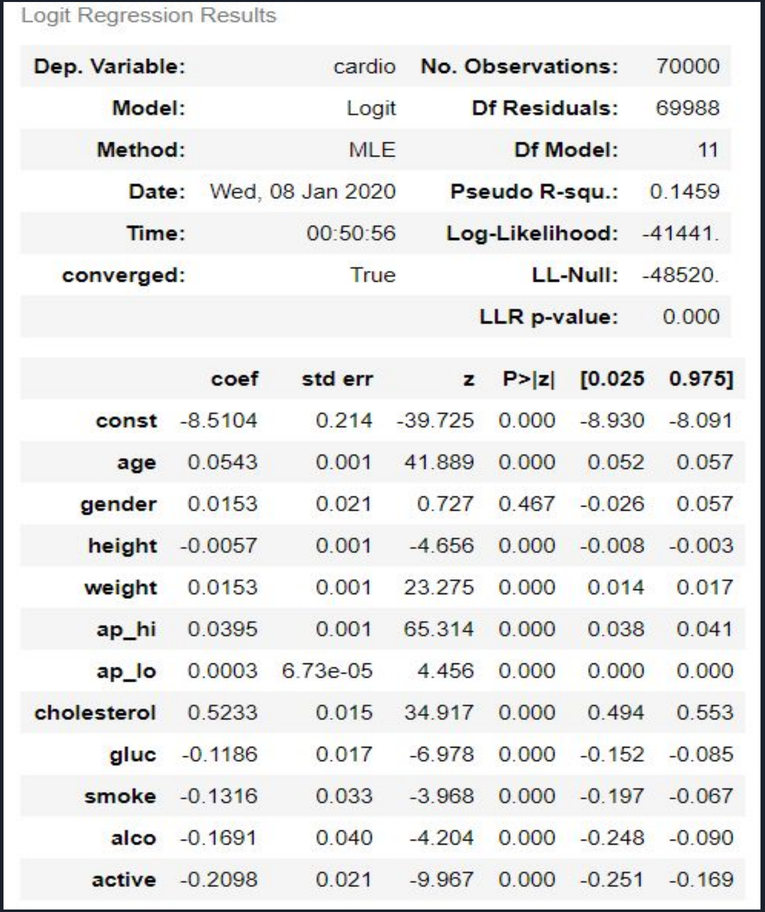


Logit Regression Results

Dep. Variable:	active	**No. Observations:**	70000
Model:	Logit	**Df Residuals:**	69996
Method:	MLE	**Df Model:**	3
Date:	Wed, 08 Jan 2020	**Pseudo R-squ.:**	0.001103
Time:	00:50:55	**Log-Likelihood:**	-34625.
converged:	True	**LL-Null:**	-34663.
		LLR p-value:	1.731e-16

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
const	1.3812	0.010	137.687	0.000	1.362	1.401
smoke	0.2186	0.042	5.212	0.000	0.136	0.301
alco	0.3109	0.064	4.863	0.000	0.186	0.436
smoke_alco	-0.1925	0.100	-1.932	0.053	-0.388	0.003

3. cholesterol * physical activity vs CD

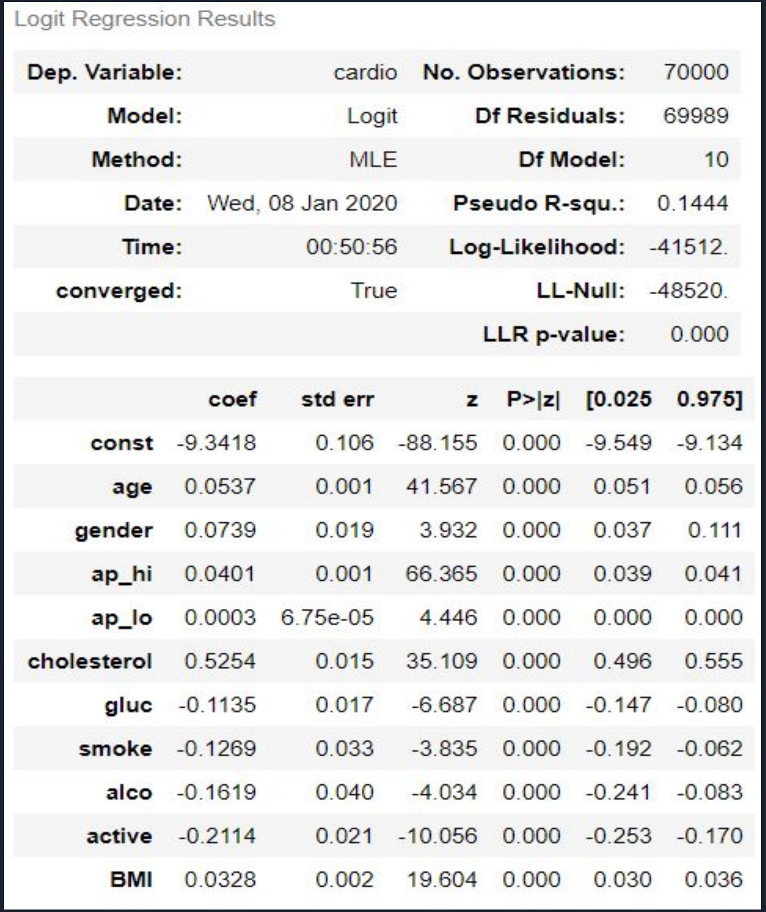


Logit Regression Results

Dep. Variable:	cardio	No. Observations:	70000
Model:	Logit	Df Residuals:	69988
Method:	MLE	Df Model:	11
Date:	Wed, 08 Jan 2020	Pseudo R-squ.:	0.1459
Time:	00:50:56	Log-Likelihood:	-41441.
converged:	True	LL-Null:	-48520.
		LLR p-value:	0.000

	coef	std err	z	P>|z|	[0.025	0.975]
const	-8.5104	0.214	-39.725	0.000	-8.930	-8.091
age	0.0543	0.001	41.889	0.000	0.052	0.057
gender	0.0153	0.021	0.727	0.467	-0.026	0.057
height	-0.0057	0.001	-4.656	0.000	-0.008	-0.003
weight	0.0153	0.001	23.275	0.000	0.014	0.017
ap_hi	0.0395	0.001	65.314	0.000	0.038	0.041
ap_lo	0.0003	6.73e-05	4.456	0.000	0.000	0.000
cholesterol	0.5233	0.015	34.917	0.000	0.494	0.553
gluc	-0.1186	0.017	-6.978	0.000	-0.152	-0.085
smoke	-0.1316	0.033	-3.968	0.000	-0.197	-0.067
alco	-0.1691	0.040	-4.204	0.000	-0.248	-0.090
active	-0.2098	0.021	-9.967	0.000	-0.251	-0.169

4. smoking * alcohol use vs CD



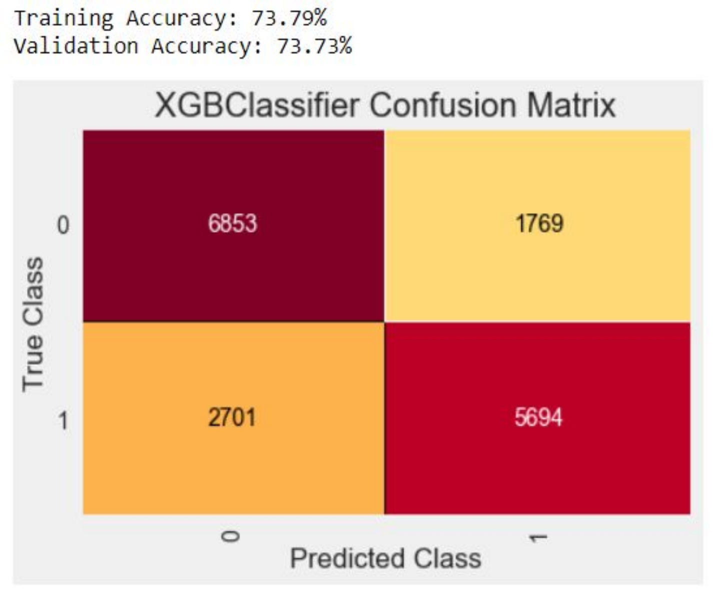
Logit Regression Results

Dep. Variable:	cardio	No. Observations:	70000
Model:	Logit	Df Residuals:	69989
Method:	MLE	Df Model:	10
Date:	Wed, 08 Jan 2020	Pseudo R-squ.:	0.1444
Time:	00:50:56	Log-Likelihood:	-41512.
converged:	True	LL-Null:	-48520.
		LLR p-value:	0.000

	coef	std err	z	P>|z|	[0.025	0.975]
const	-9.3418	0.106	-88.155	0.000	-9.549	-9.134
age	0.0537	0.001	41.567	0.000	0.051	0.056
gender	0.0739	0.019	3.932	0.000	0.037	0.111
ap_hi	0.0401	0.001	66.365	0.000	0.039	0.041
ap_lo	0.0003	6.75e-05	4.446	0.000	0.000	0.000
cholesterol	0.5254	0.015	35.109	0.000	0.496	0.555
gluc	-0.1135	0.017	-6.687	0.000	-0.147	-0.080
smoke	-0.1269	0.033	-3.835	0.000	-0.192	-0.062
alco	-0.1619	0.040	-4.034	0.000	-0.241	-0.083
active	-0.2114	0.021	-10.056	0.000	-0.253	-0.170
BMI	0.0328	0.002	19.604	0.000	0.030	0.036

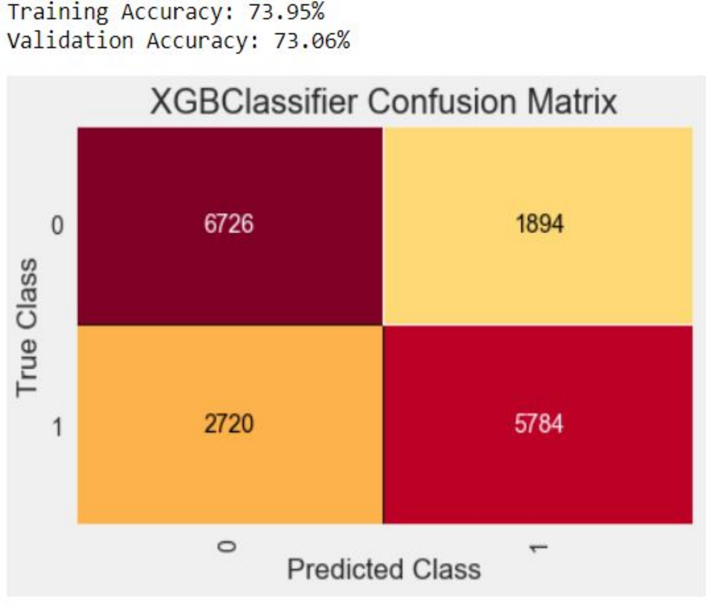
Appendix

5. height/weight classifier



```
Training Accuracy: 73.79%
Validation Accuracy: 73.73%
```

6. BMI classifier



```
Training Accuracy: 73.95%
Validation Accuracy: 73.06%
```