



PROJECT REPORT

Indira Lakshmi Gundavarapu



Variable	Description
Age	age of the participant at the time of examination
Male	gender of the participant (male =1, female = 0)
Education	Educational level of the patient (1 = less than high school, 2 = completed high school or equivalent, 3 = some college, 4= completed college or higher)
Income	Income of the patient
Current Smoker	whether the participant is currently a smoker (yes or no)
Cigarettes per Day	the average number of cigarettes smoked per day by current smokers
BP Meds	whether the participant is taking blood pressure medication (yes or no)
Prevalent Stroke	whether the participant has a history of stroke (yes or no)
Prevalent Hyp	whether the participant has a history of hypertension (yes or no)
Diabetes	whether the participant has diabetes (yes or no)
Total Chol	total cholesterol level in milligrams per deciliter
Sys BP	systolic blood pressure in millimeters of mercury
Dia BP	diastolic blood pressure in millimeters of mercury
BMI	body mass index in kilograms per square meter
Heart Rate	resting heart rate in beats per minute
Glucose	Blood glucose level in milligrams per deciliter
A1c	Hemoglobin A1c (%)
Ten Year CHD	whether the participant developed coronary heart disease (CHD) within 10 years of the examination (yes or no)



- › There are a total of 18 attributes, out of which I divided them into 10 measures and 8 categories.

- › **Measures:**

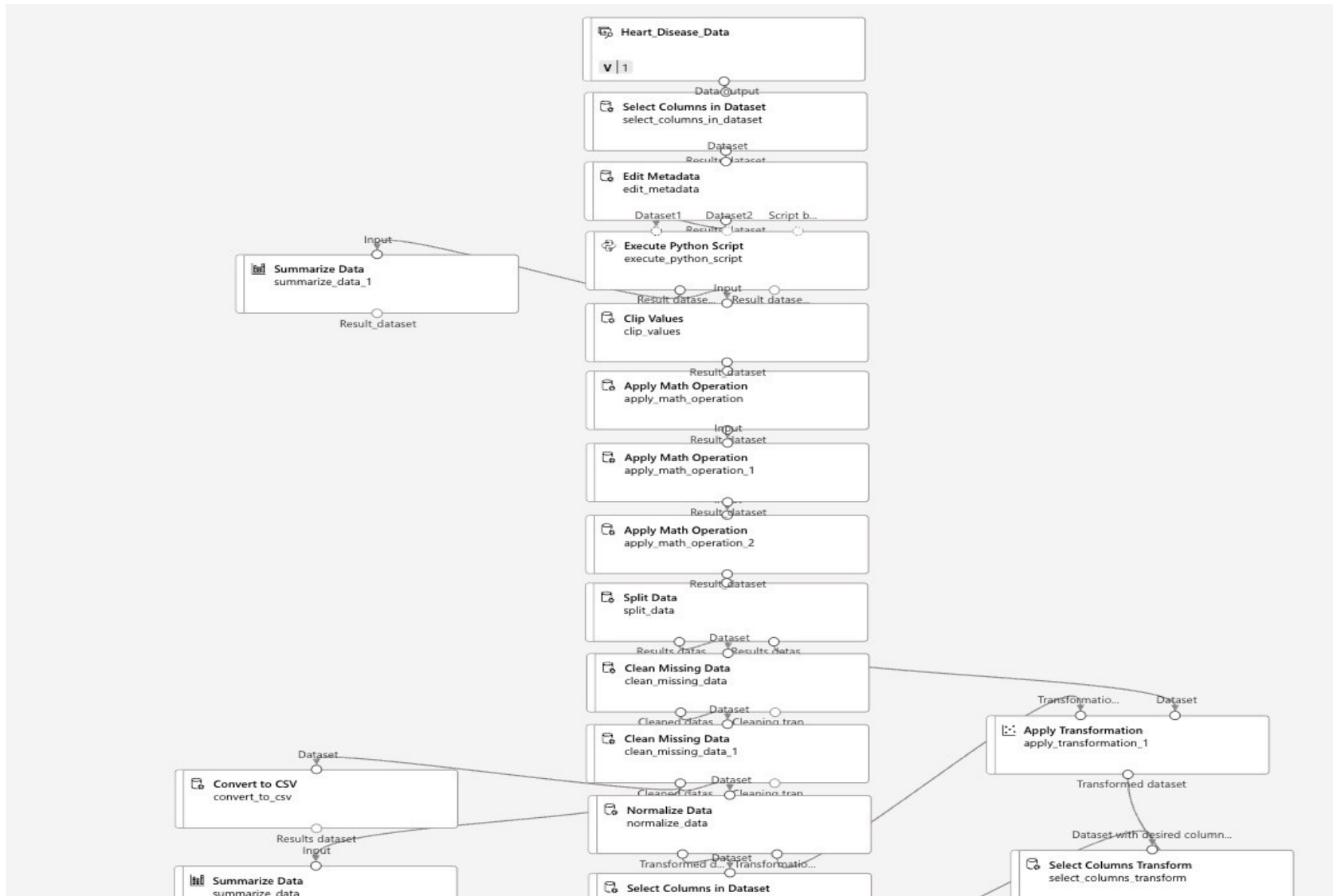
- Income
- BMI
- Age
- DiaBP
- CigsPerDay
- totChol
- sysBP
- heartrate
- Glucose
- a1c

- › **Categories:**

- male
- education
- currentSmoker
- BPMeds
- prevalentStroke
- prevalentHyp
- diabetes
- TenYearCHD

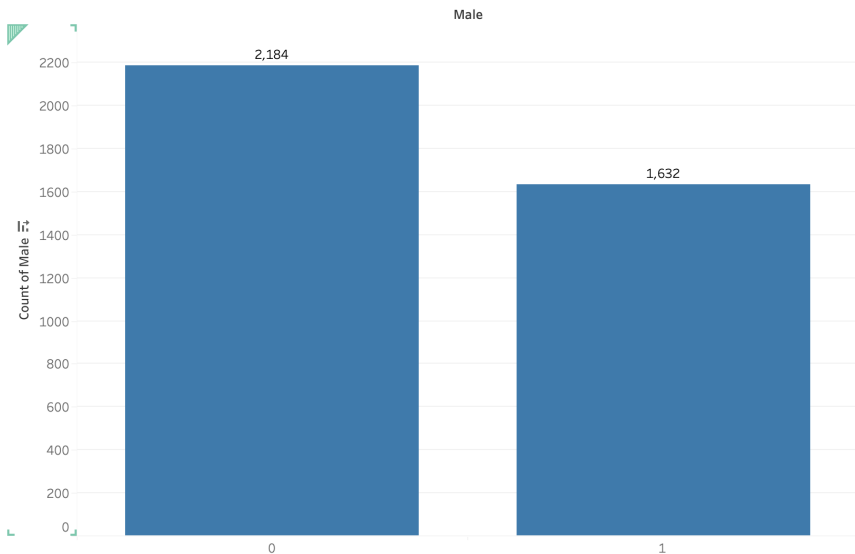


- › Removed patientID
- › Changed cigsPerDay to double
- › Used execute python script to fill 1975 null values in cigsPerDay with the corresponding values of non-current smokers of value '0'. After this, there were only 27 null values left in cigsPerDay
- › Clip values for totChol because the cholesterol values of 6,960 and 9,280 did not make sense
- › Cleaned the missing data for binary variables using fill with mode and for non-binary variables using fill with mean
- › Converted this to CSV to generate an EDA report in Tableau
- › Normalized and removed the variable 'a1c' because of its high correlation with the variable 'glucose.'
- › Applied filter-based feature selection

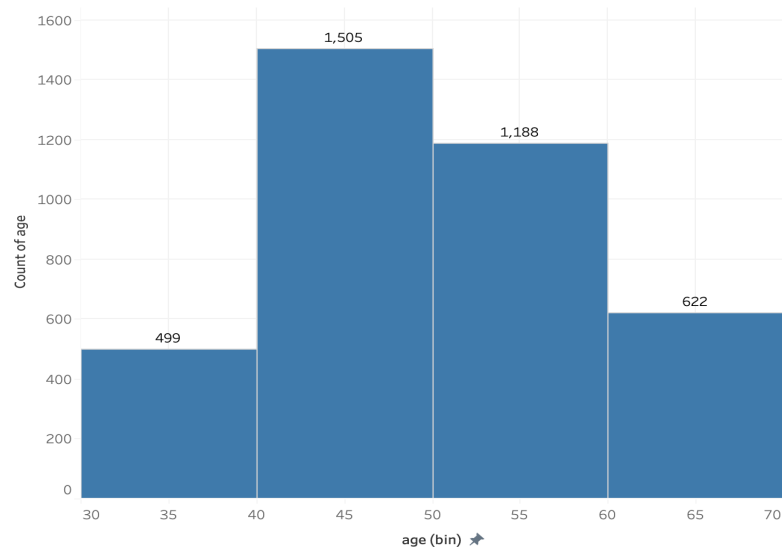




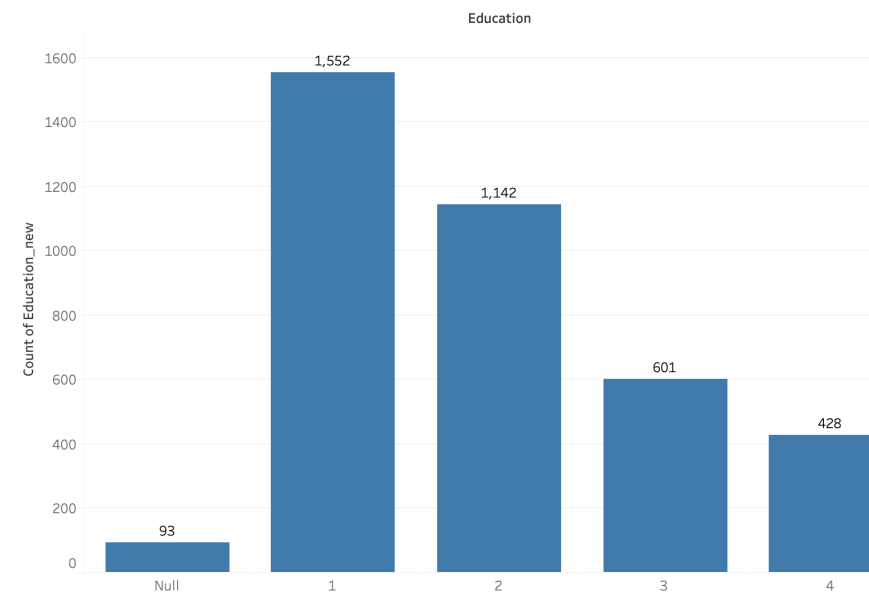
Gender



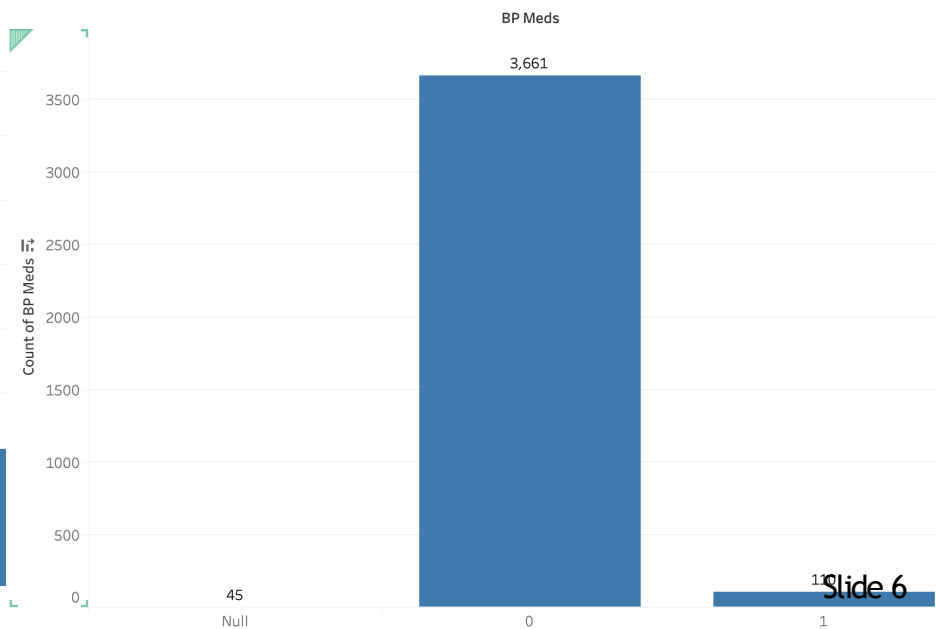
Age



Education

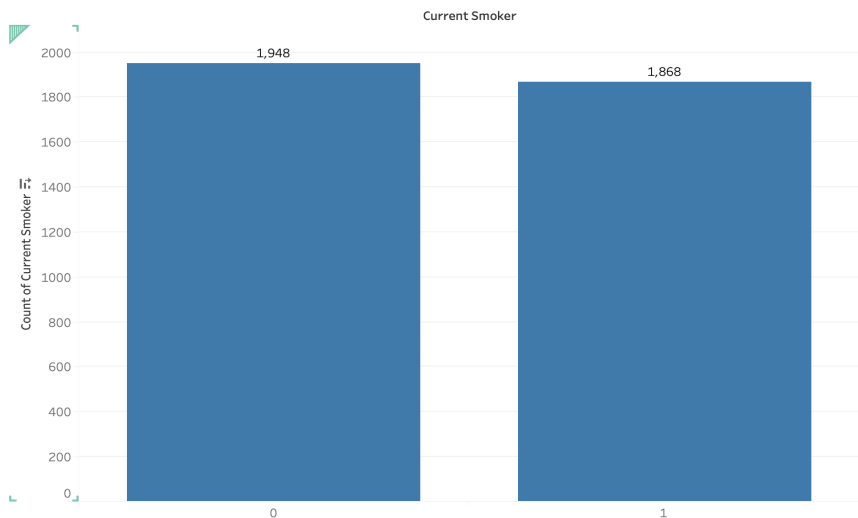


BP Meds

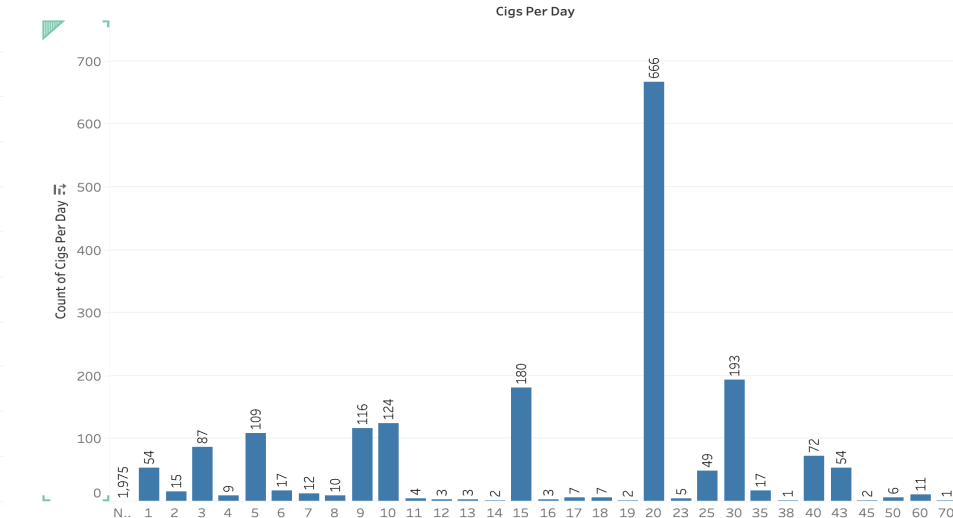




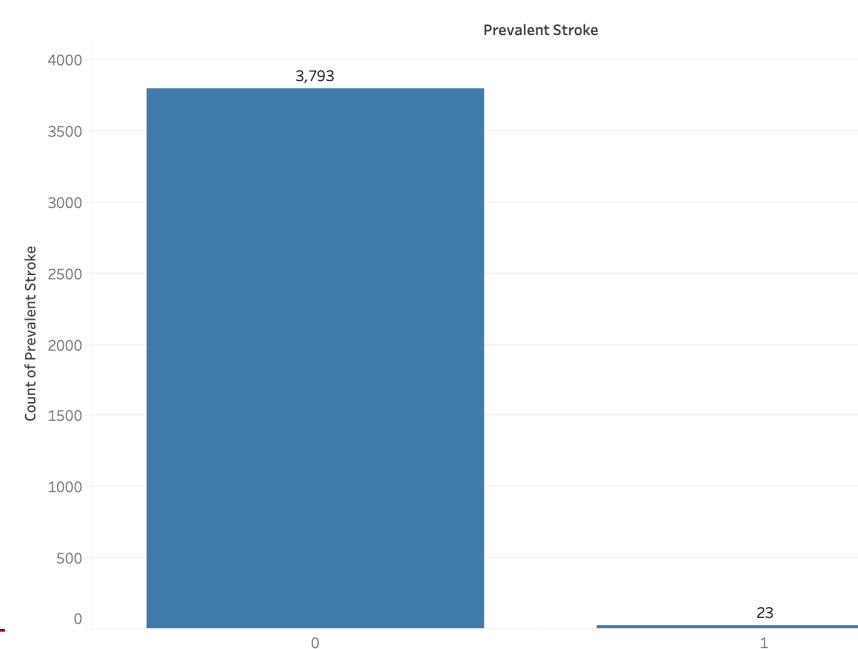
Current Smoker



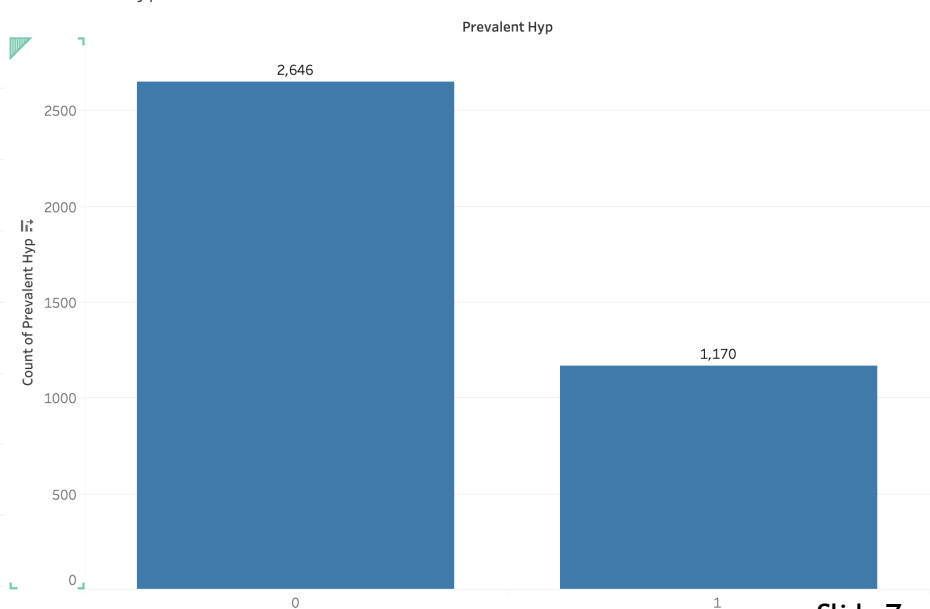
Cigs Per Day



Prevalent Stroke

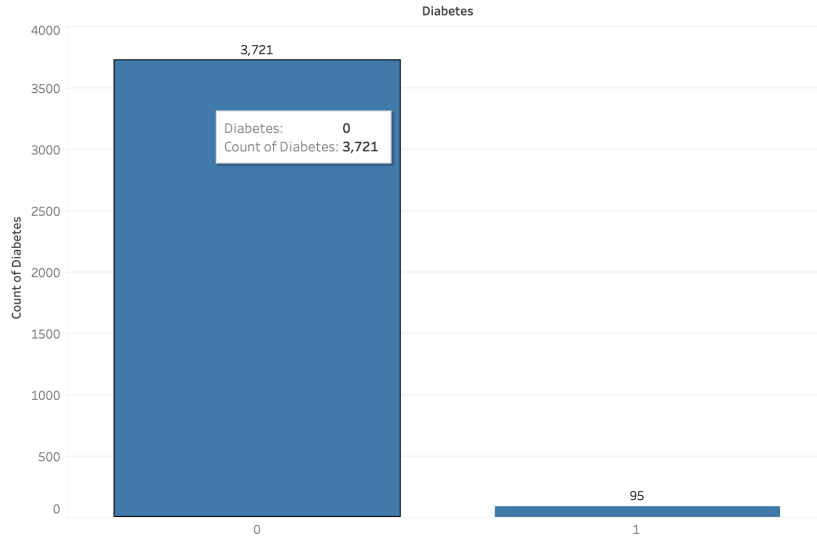


Prevalent Hypertension

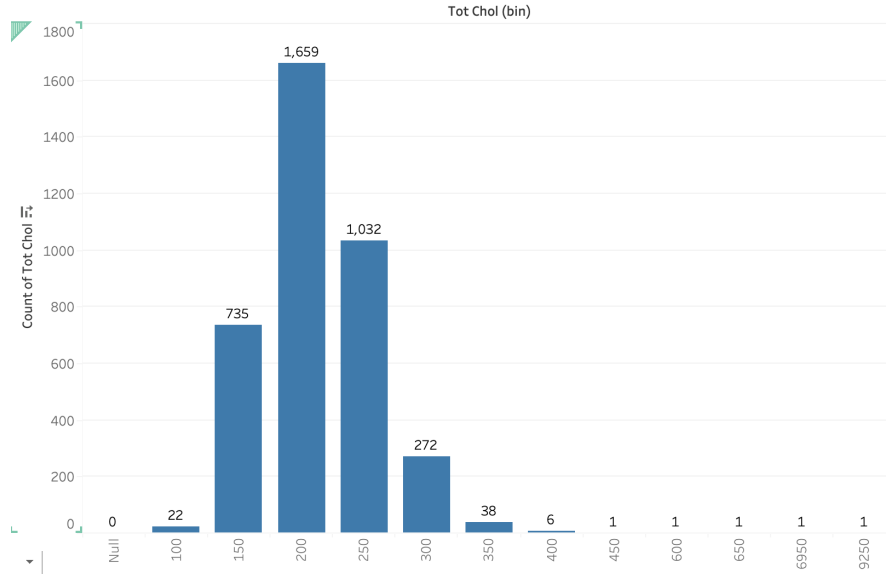




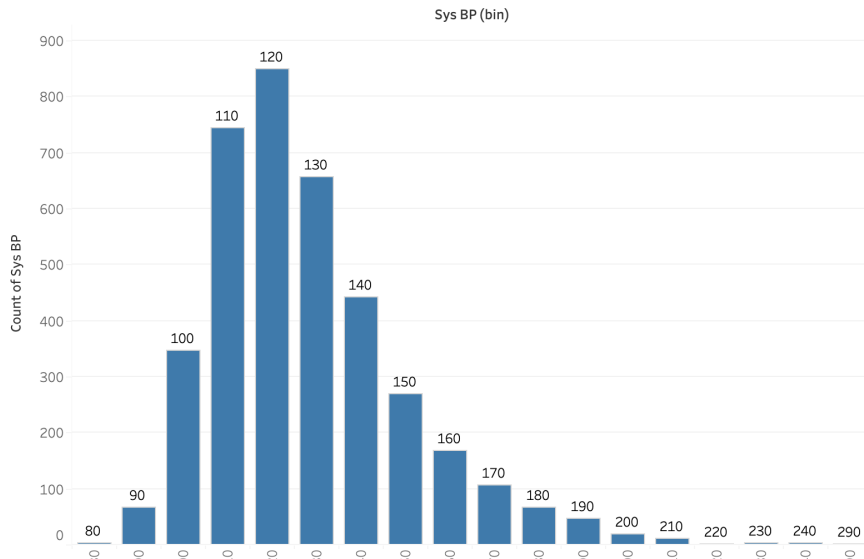
Diabetes



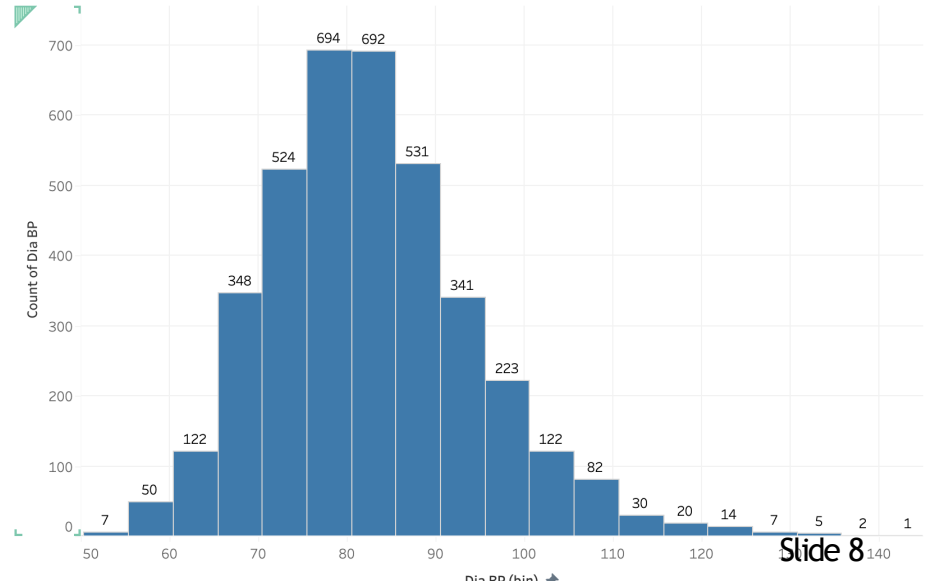
Total Cholesterol



sys BP

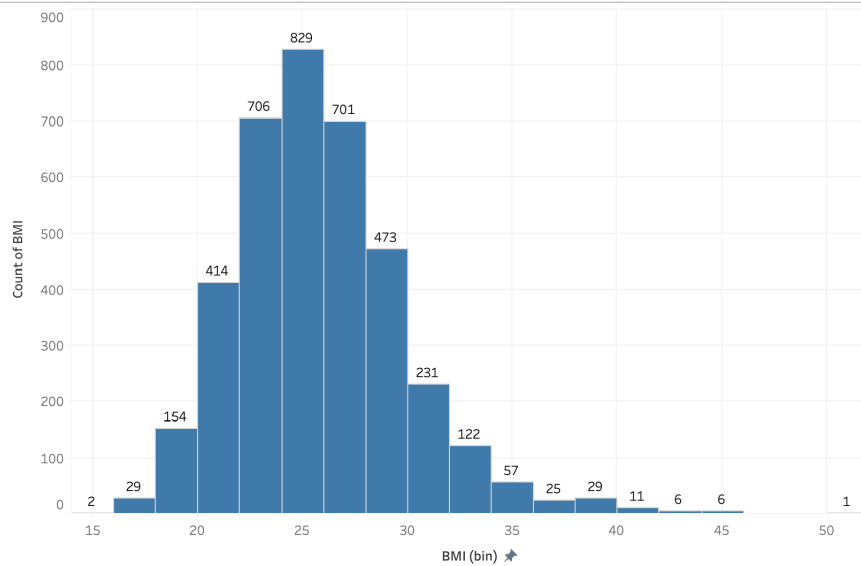


Dia BP

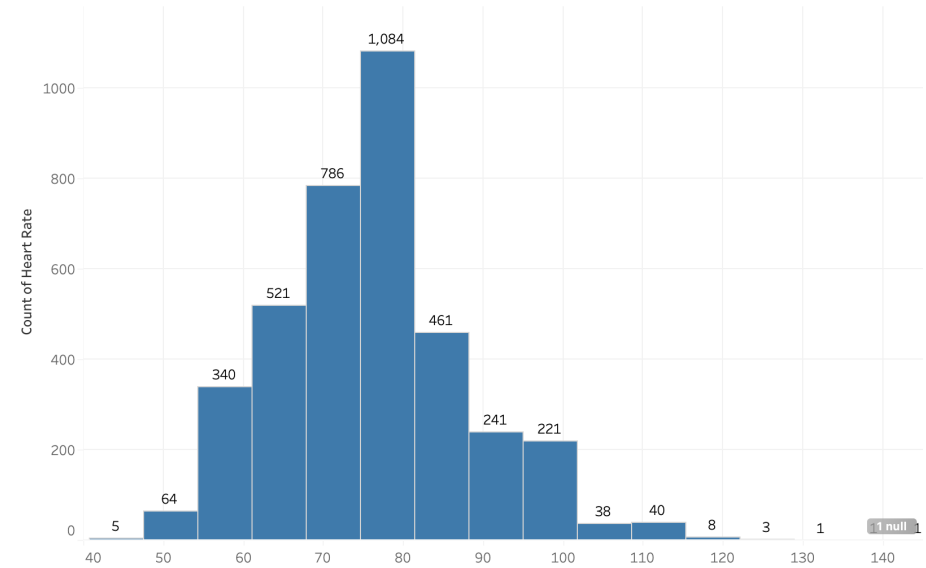




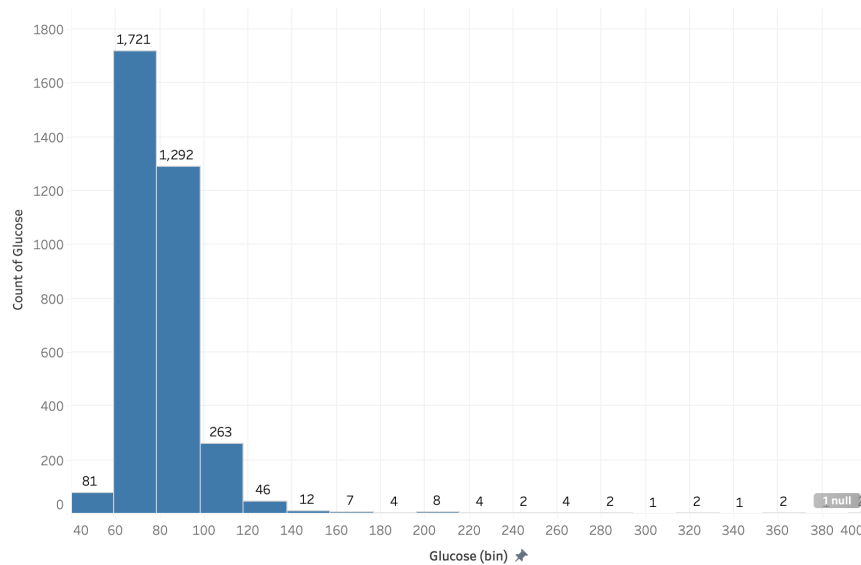
BMI



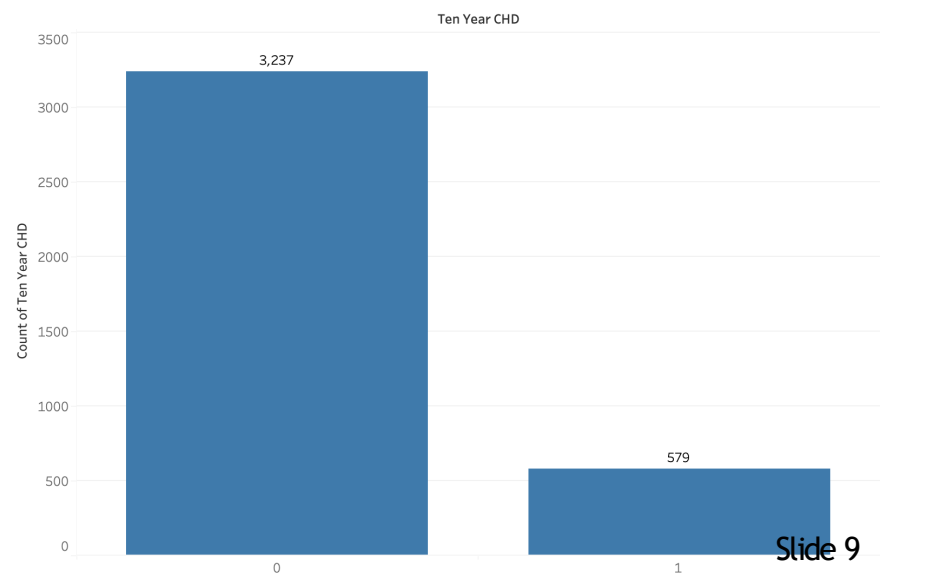
Heart Rate



Glucose

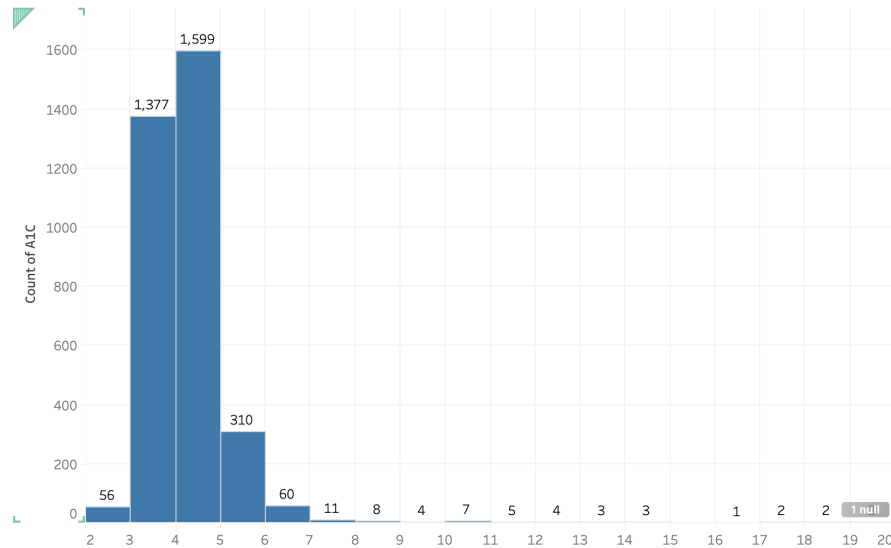


Ten Year CHD

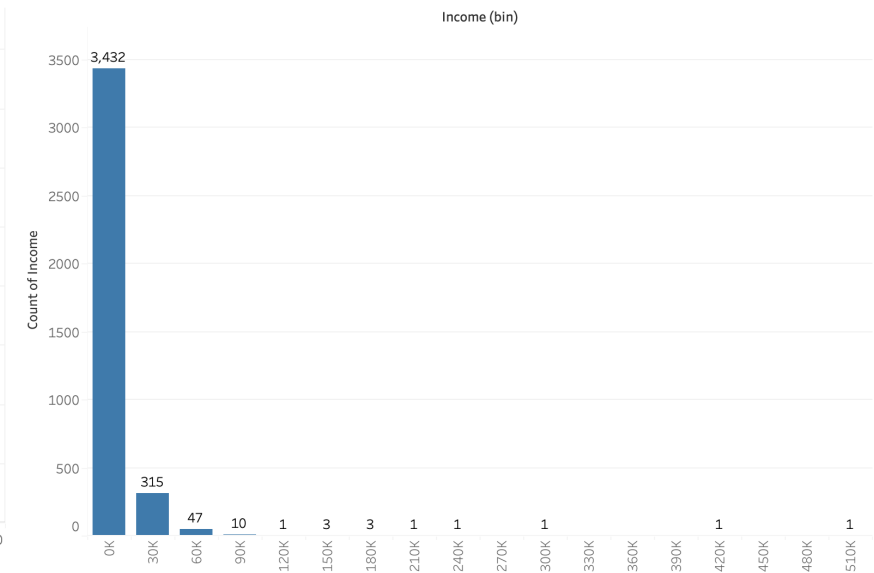




A1C

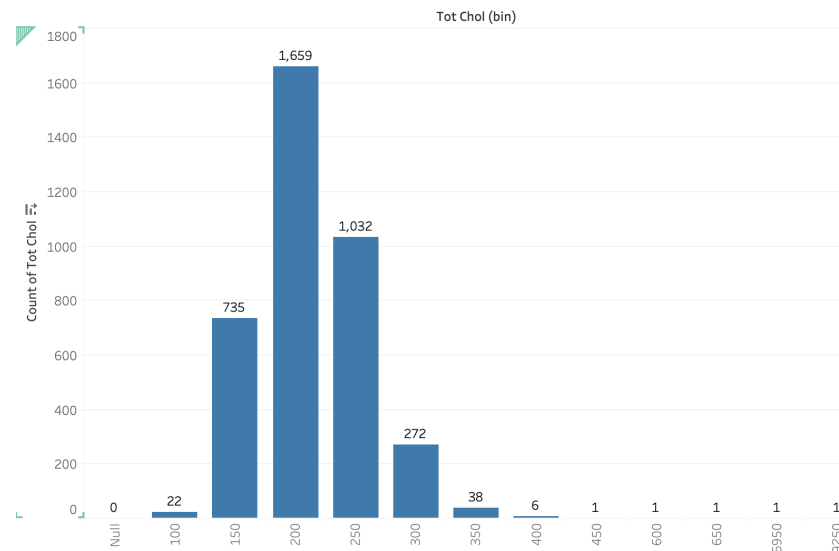


Income



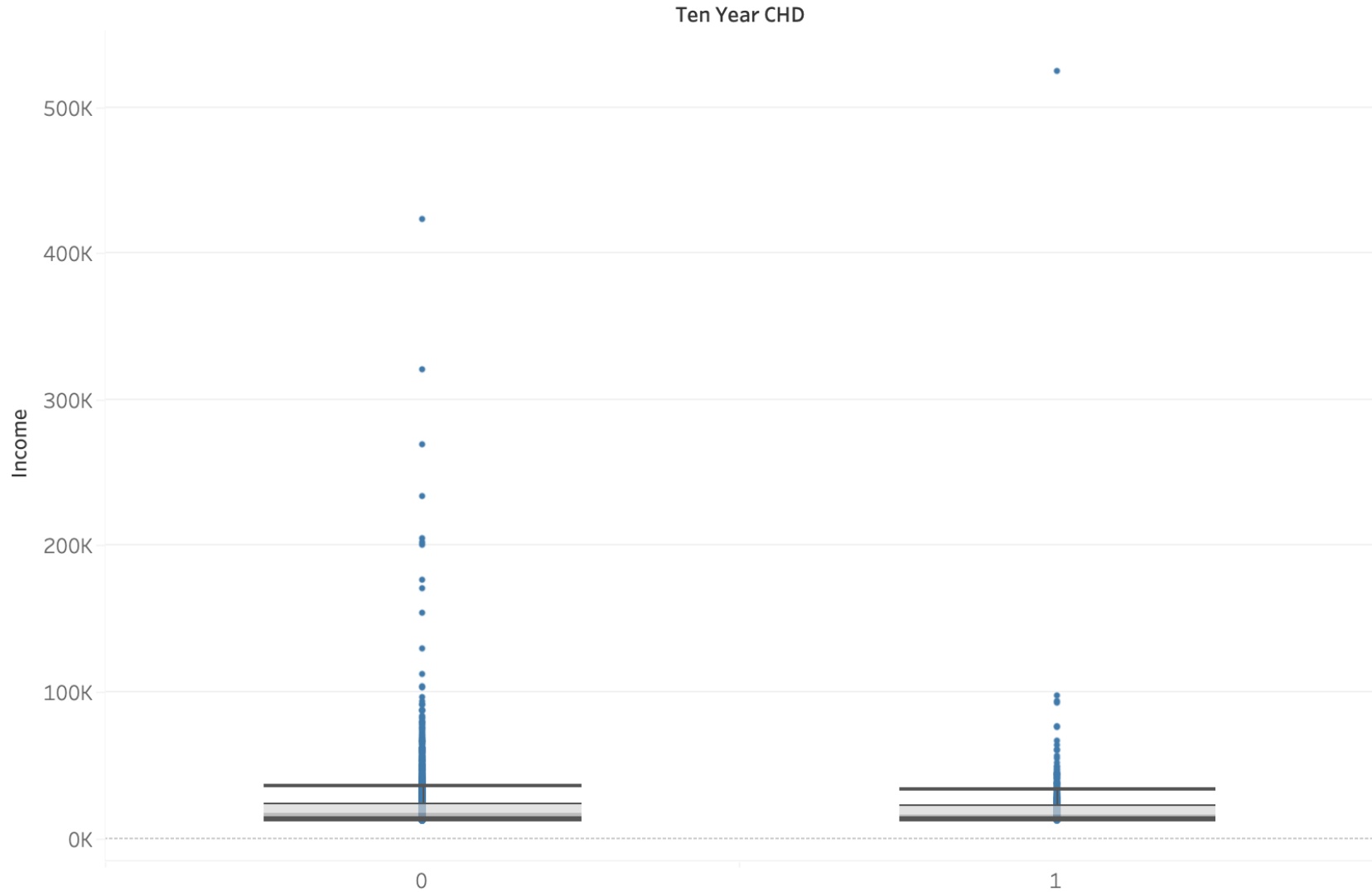
A1C (bin) ★

Total Cholesterol



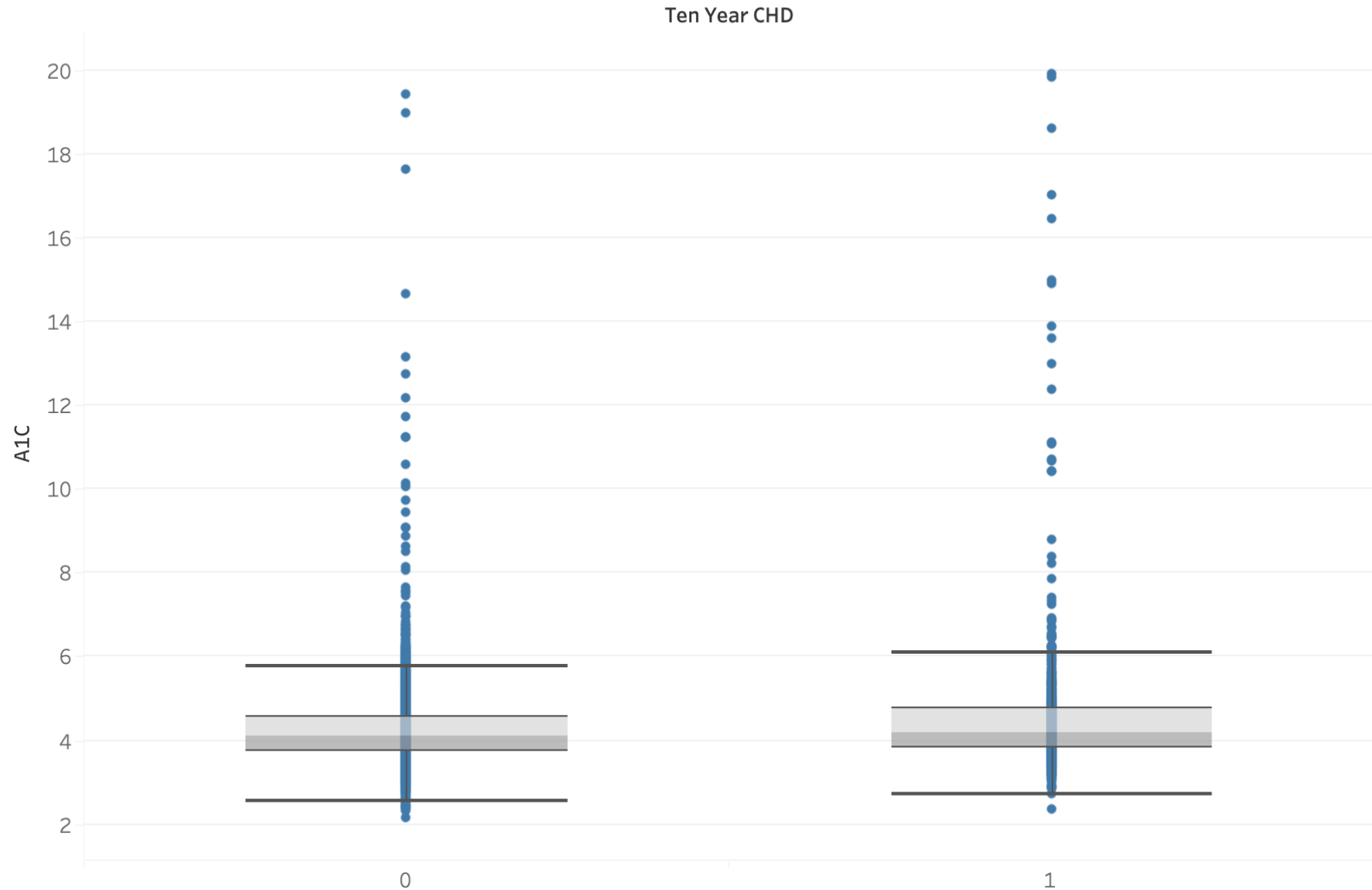


Income vs Ten Year CHD



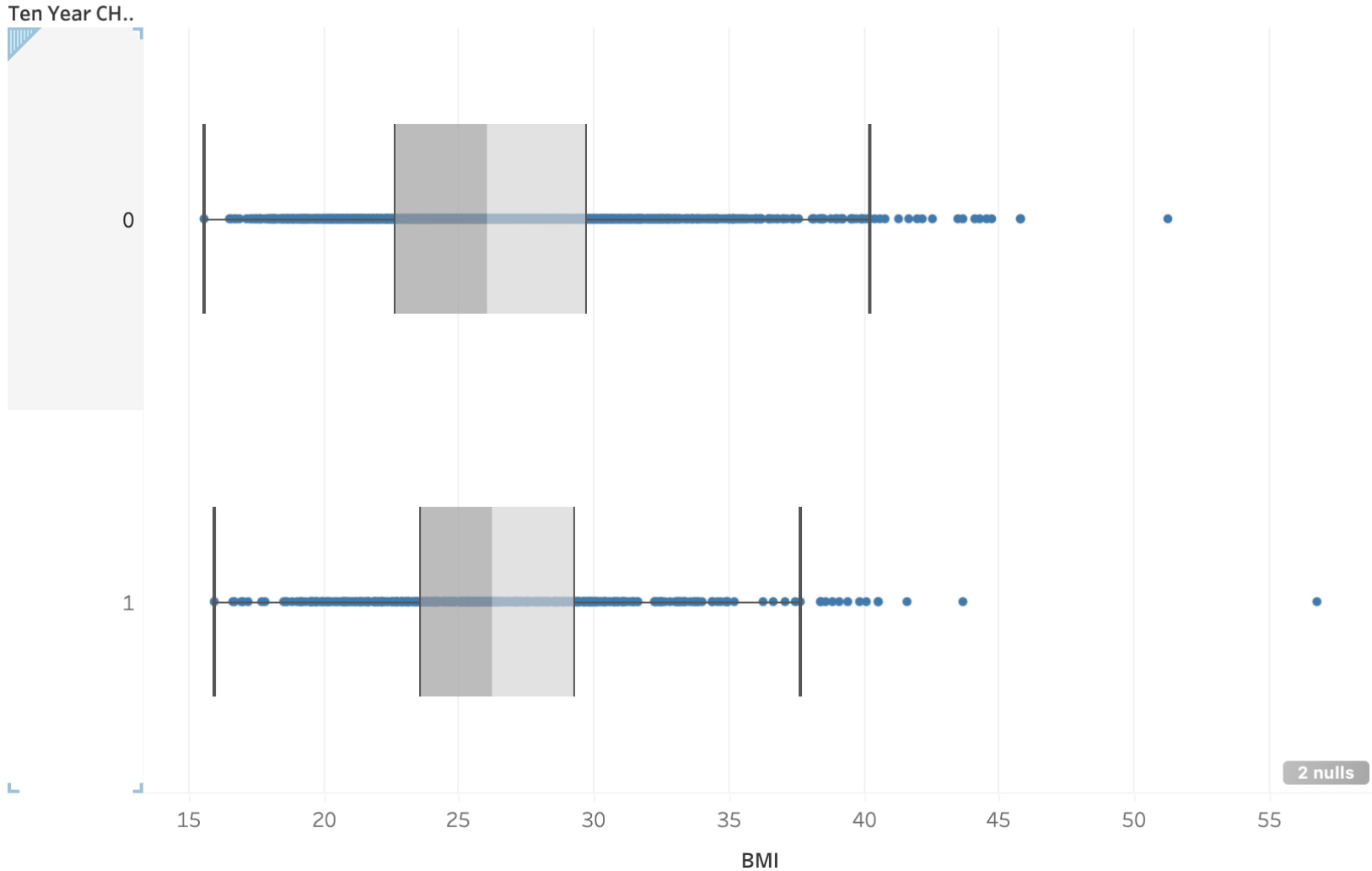


A1C vs Ten Year CHD





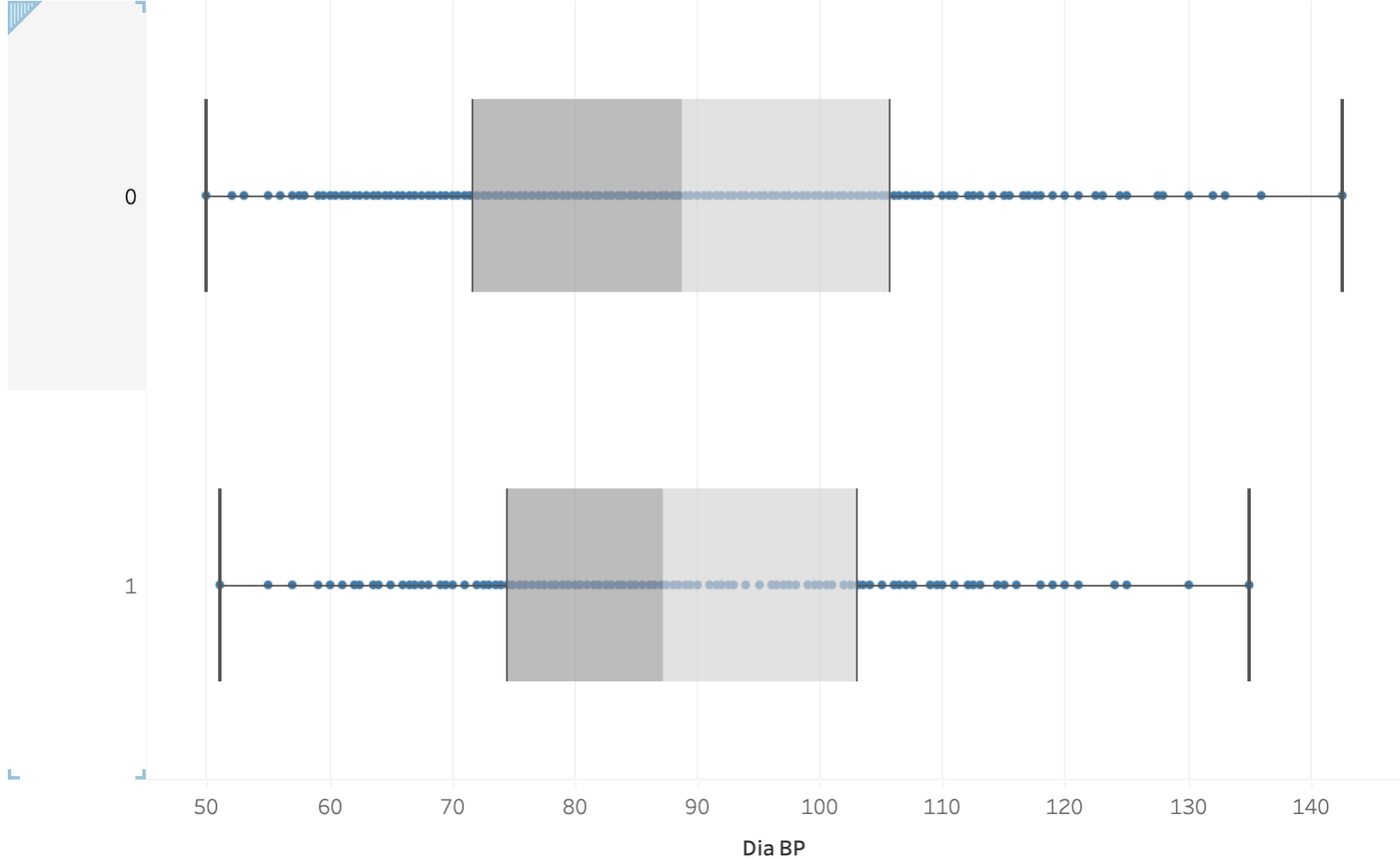
BMI vs Ten Year CHD





Dia BP vs Ten Year CHD

Ten Year CHD

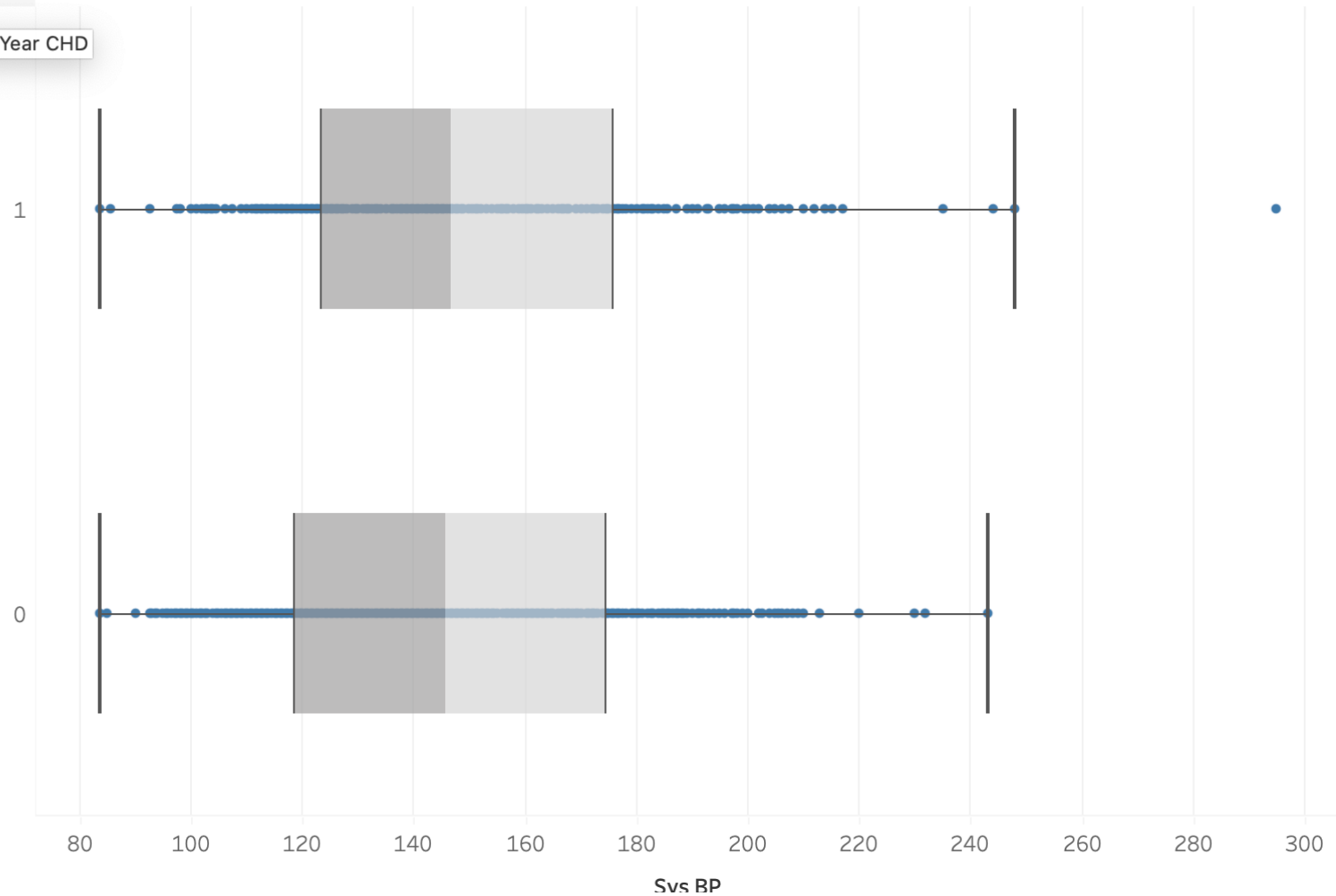




Sys BP vs Ten Year CHD

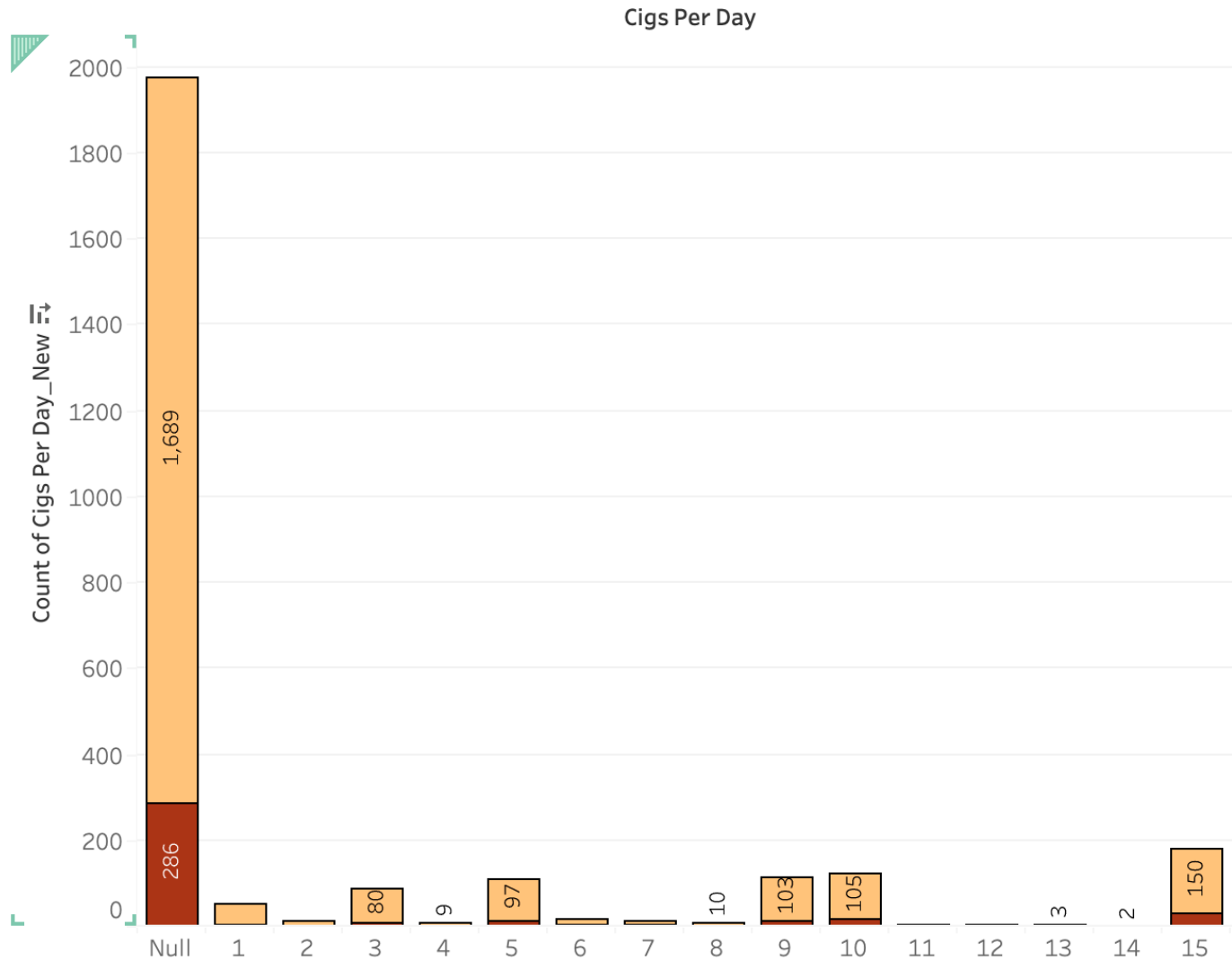
Ten Ye.. ↕

Ten Year CHD



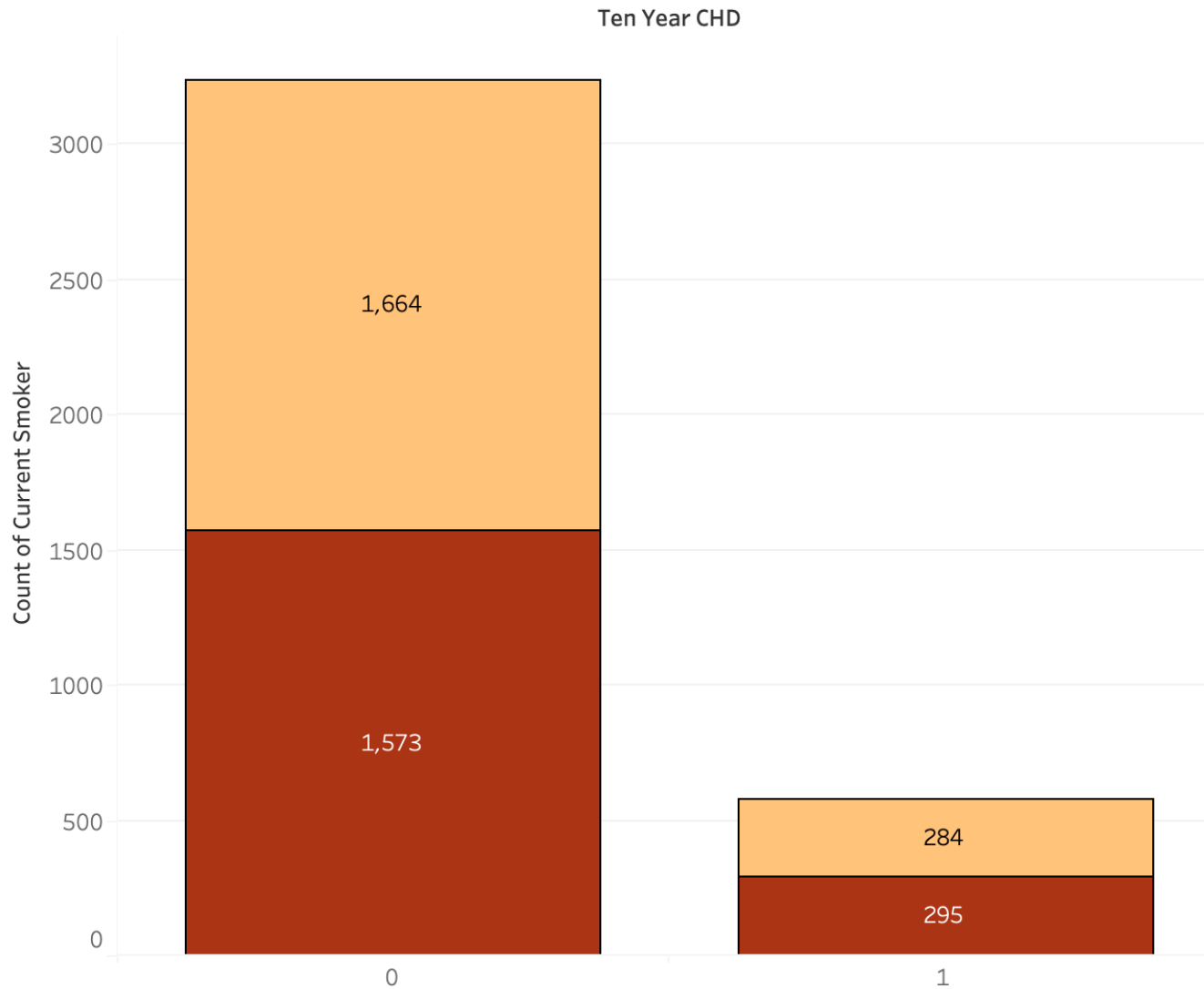


Cigs Per Day VS TenYear CHD



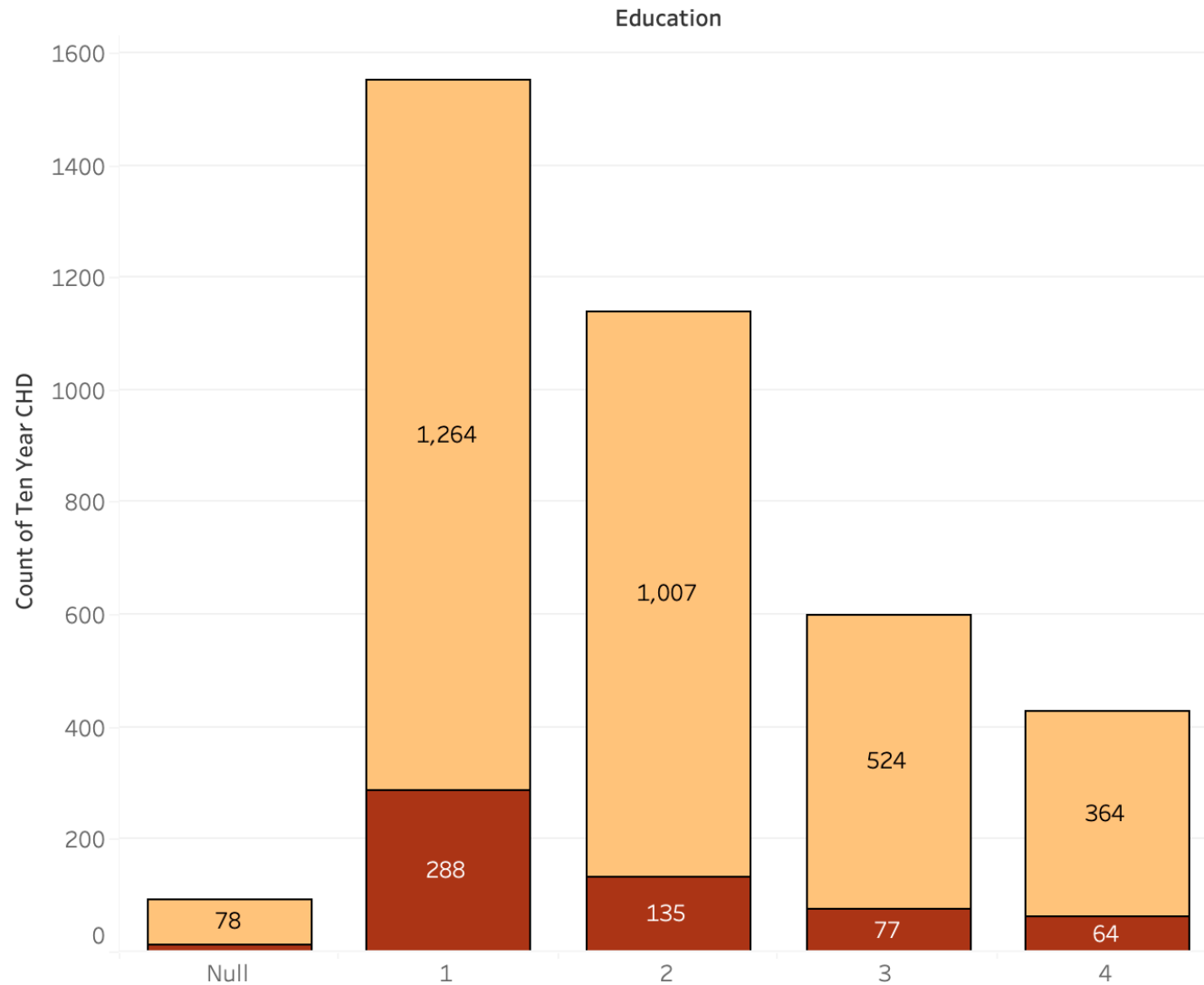


Current Smoker VS TenYearCHD



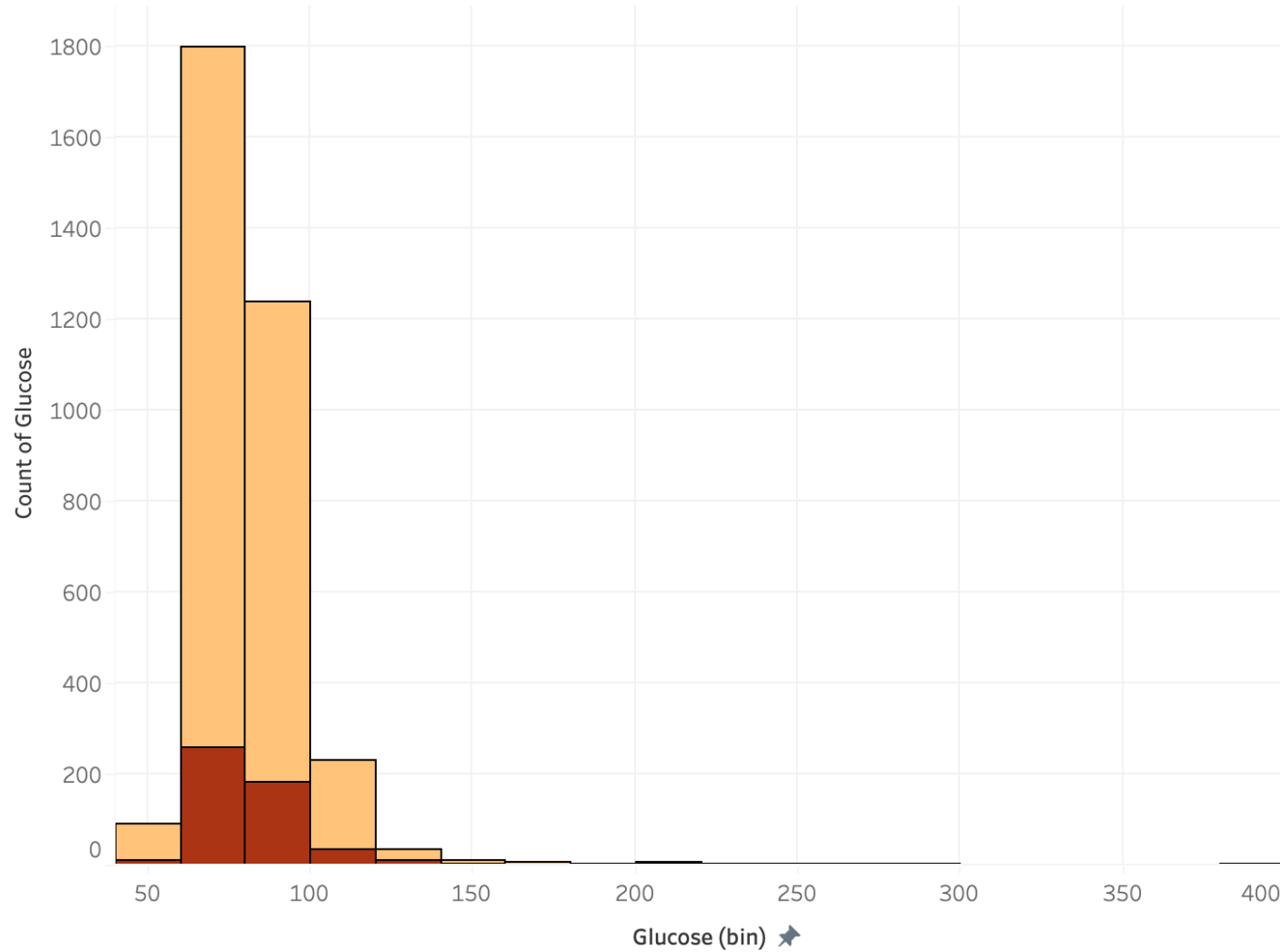


Education VS TenYearCHD





Glucose Vs TenYearCHD





- › Reduced skewness by using apply math operations for different columns with a varied intensity of skewness.
- › **Sqrt transformation attributes:**
 - education
 - cigsPerDay
 - prevalentHyp
 - diaBP
 - BMI
 - heartRate
- › **Ln transformation attributes:**
 - sysBP
- › **Log10 transformation attributes:**
 - BPMeds
 - prevalentStroke
 - diabetes
 - glucose
 - a1c
 - income
 - totChol



Microsoft Azure Machine Learning Studio

University of Southern California > Final_exam_work > Data > Heart_Disease_Data

Heart_Disease_Data

Version: 1 (latest)



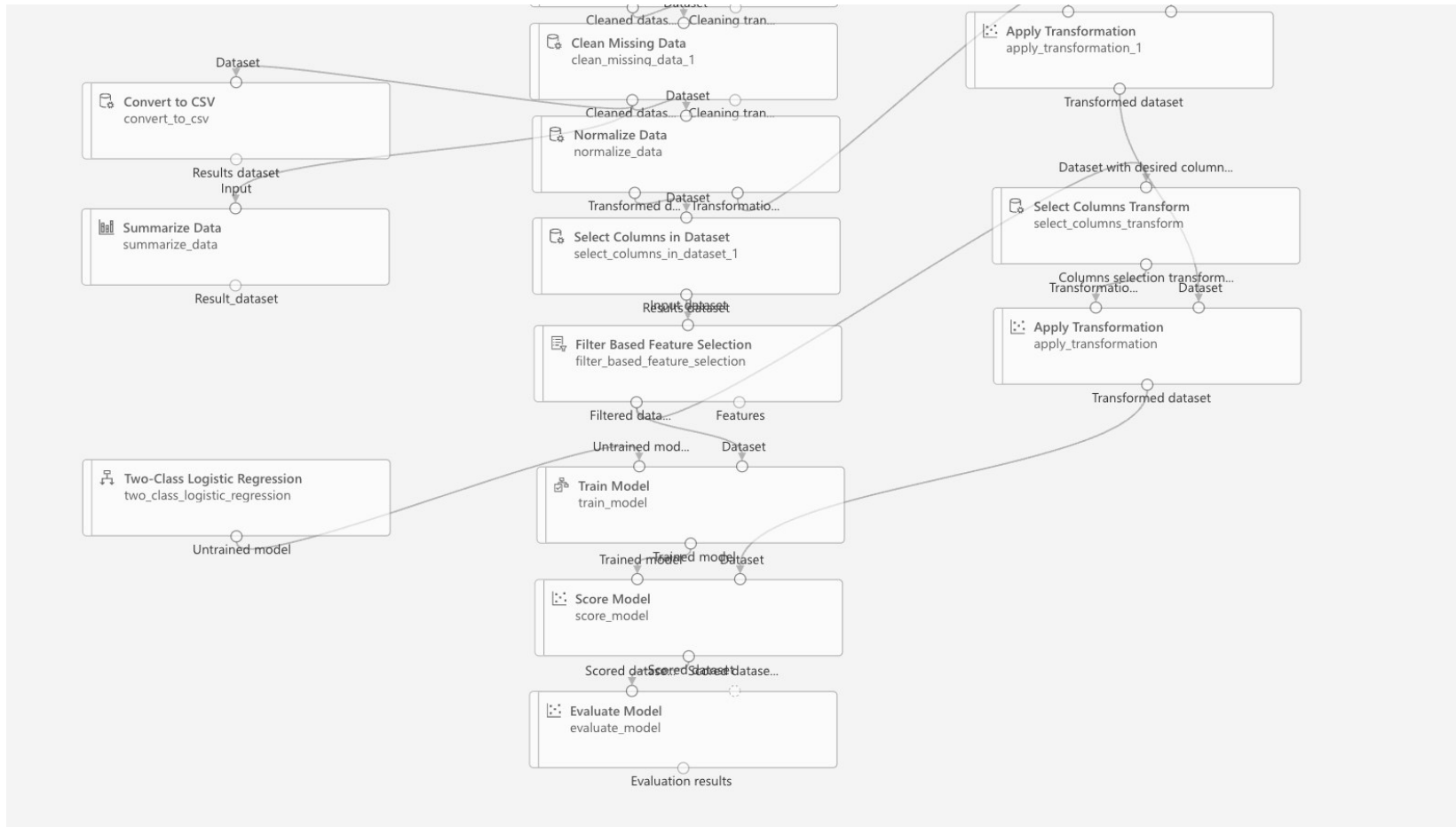
Details Consume **Explore** Models Jobs

Refresh Generate profile

Preview **Profile**

Number of columns: 19 Number of rows: 3816

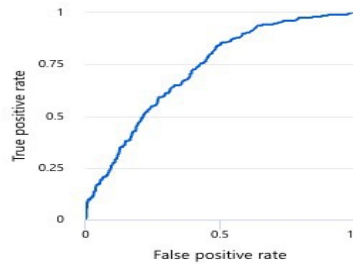
Column	Profile	Type	Min	Max	Count	Missing count	Empty count	Error count	Mean	Std dev...	Variance	Skewness	Kurtosis
prevalentHyp		Integer	0	1	3816	0	0	0	0.31	0.46	0.21	0.84	-1.3
diabetes		Integer	0	1	3816	0	0	0	0.02	0.16	0.02	6.1	35.17
totChol		Decimal	107	9,280	3816	47	0	0	240.85	188.93	35,695.2	41	1,811.61
sysBP		Decimal	83.5	295	3816	0	0	0	132.26	22.12	489.28	1.18	2.29
diaBP		Decimal	50	142.5	3816	0	0	0	82.87	11.91	141.79	0.72	1.23



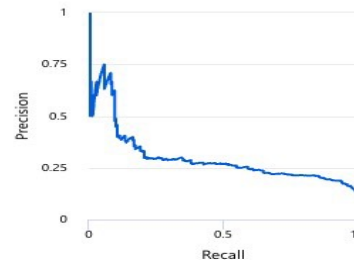


● Scored dataset (left port)

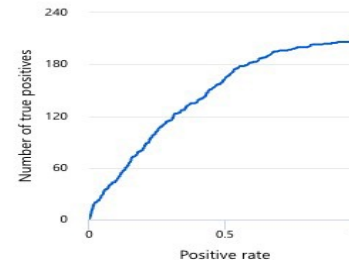
ROC curve



Precision-recall curve



Lift curve

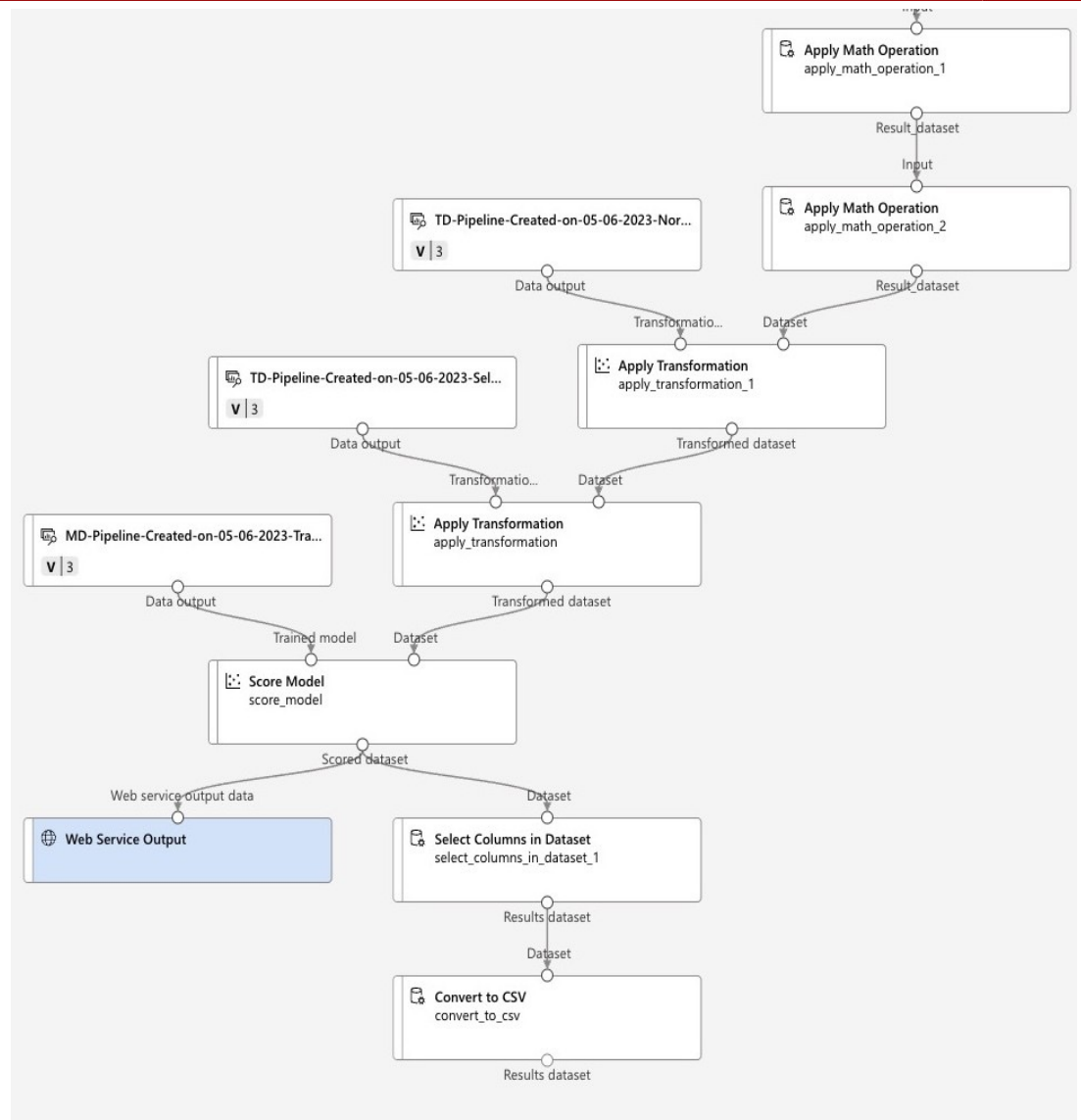
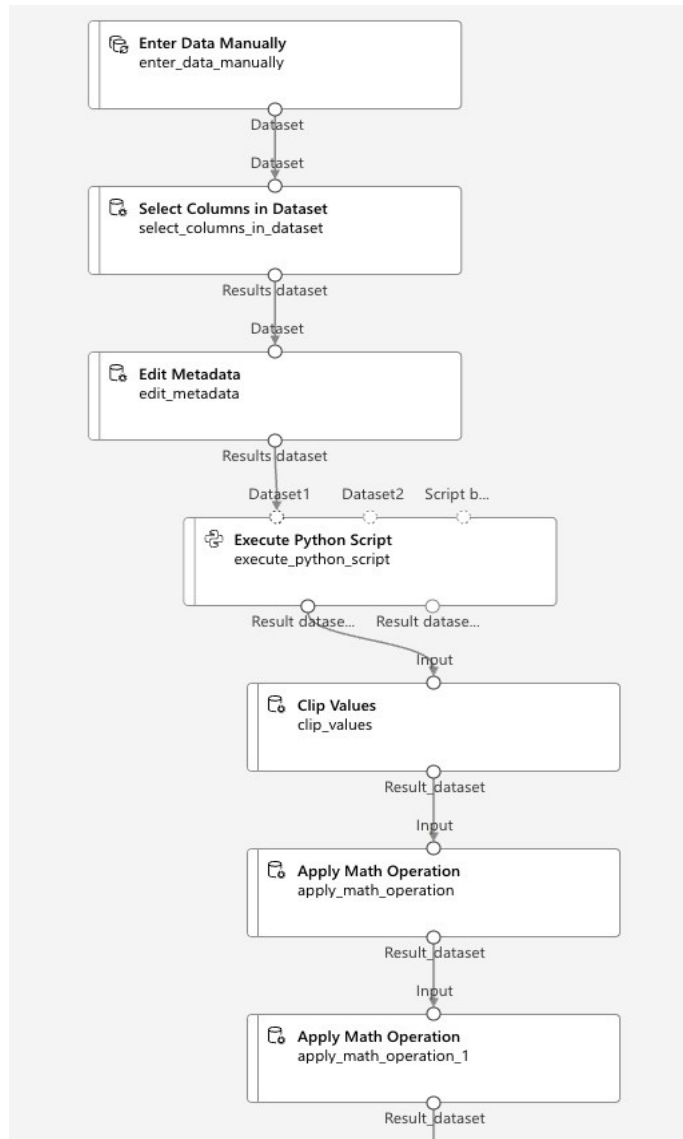


Threshold 0.5

Accuracy 0.868
Precision 0.706
Recall 0.058
F1 Score 0.107
AUC 0.725

	Actual	
	1	0
Predicted 1	12	5
Predicted 0	196	1313

Score bin ↓	Positive exam...	Negative exam...	Fraction above thresh...	Accuracy	F1 Score	Precisi...	Recall	Negative precisi...	Negative recall	Cumulative AUC
(0.900,1.000]	0	0	0.000	0.864	0.000	1.000	0.000	0.864	1.000	0.000
(0.800,0.900]	1	0	0.001	0.864	0.010	1.000	0.005	0.864	1.000	0.000
(0.700,0.800]	0	0	0.001	0.864	0.010	1.000	0.005	0.864	1.000	0.000
(0.600,0.700]	1	1	0.002	0.864	0.019	0.667	0.010	0.865	0.999	0.000
(0.500,0.600]	10	4	0.011	0.868	0.107	0.706	0.058	0.870	0.996	0.000
(0.400,0.500]	10	23	0.033	0.860	0.171	0.440	0.106	0.874	0.979	0.002
(0.300,0.400]	23	81	0.101	0.822	0.249	0.292	0.216	0.881	0.917	0.012
(0.200,0.300]	63	194	0.269	0.736	0.349	0.263	0.519	0.910	0.770	0.068
(0.100,0.200]	79	480	0.636	0.473	0.317	0.193	0.899	0.962	0.406	0.334
(0.000,0.100]	21	535	1.000	0.136	0.240	0.136	1.000	1.000	0.000	0.725




 Rows ①
178

 Columns ②
17

male	age	BPMeds	prevalentHyp	diabetes	totChol	sysBP	diaBP	glucose	Sqrt(cigsPerDay)	Sqrt(diaBP)	Ln(sysBP)	Log10(totChol)	Log10(glucose)	Log10(a1c)	Scored Labels	Scored Probabilities
1.138811	-0.181031	-0.177571	1.49024	-0.16403	-0.109578	0.246035	0.583968	-0.428919	-0.711428	0.6186	0.339713	-0.018152	-0.510372	-0.17291	0	0.193152
1.138811	-0.998014	-0.177571	-0.671033	-0.16403	-0.956835	0.290551	0.418544	-0.732244	NaN	0.455907	0.385144	-0.957352	-1.014198	-0.506078	0	0.079576
1.138811	0.519241	-0.177571	1.49024	-0.16403	-0.521757	0.646681	-0.739423	0.221063	0.281369	-0.738962	0.737244	-0.454735	0.421835	0.431552	0	0.324693
1.138811	-1.464862	-0.177571	1.49024	-0.16403	-0.26987	-0.354934	0.790747	0.047735	NaN	0.819446	-0.308052	-0.18377	0.18998	0.058835	0	0.071909
-0.878109	1.569648	-0.177571	1.49024	-0.16403	0.691882	1.069585	0.790747	-0.125594	-1.795458	0.819446	1.131281	0.741431	-0.053329	0.20725	0	0.289361
1.138811	0.869376	-0.177571	-0.671033	-0.16403	-0.384364	0.157002	0.087696	-0.775576	NaN	0.124876	0.247855	-0.305255	-1.090571	-1.596855	0	0.18561
-0.878109	0.752664	-0.177571	1.49024	-0.16403	1.355948	2.115716	0.583968	1.651023	NaN	0.6186	2.011558	1.298832	2.005956	1.971907	0	0.260069
-0.878109	-0.181031	-0.177571	1.49024	-0.16403	0.646084	0.557648	-0.491287	-0.775576	NaN	-0.473992	0.651049	0.700785	-1.090571	-0.343654	0	0.110234
1.138811	1.919783	-0.177571	-0.671033	-0.16403	1.172757	-1.044936	-0.987558	-0.082262	NaN	-1.00936	-1.144643	1.150717	0.008634	0.474519	0	0.243055
1.138811	-1.34815	-0.177571	-0.671033	-0.16403	-0.040881	-0.64429	-0.491287	0.004402	NaN	-0.473992	-0.645444	0.051303	0.13027	0.868397	0	0.057717
1.138811	0.402529	-0.177571	-0.671033	-0.16403	0.165208	-1.178485	-1.27705	-0.732244	1.04317	-1.332151	-1.32025	0.25447	-1.014198	-1.134777	0	0.195964
-0.878109	1.1028	-0.177571	1.49024	-0.16403	0.508691	0.379583	0.25312	0.177731	-2.35031	0.291354	0.475033	0.576982	0.364886	0.43191	0	0.196747
-0.878109	-1.34815	5.631544	1.49024	-0.16403	NaN	-0.332676	-0.243151	NaN	-1.795458	-0.214135	-0.282833	NaN	NaN	NaN	0	0.063647
-0.878109	-1.581574	-0.177571	-0.671033	-0.16403	-1.781194	-0.488483	-0.822135	-0.298922	0.281369	-0.828476	-0.461527	-2.06066	-0.309283	-0.41525	0	0.052291
1.138811	-0.998014	5.631544	1.49024	-0.16403	0.165208	0.290551	0.25312	-0.645579	0.281369	0.291354	0.385144	0.25447	-0.864914	-0.833544	0	0.209406
-0.878109	-1.464862	-0.177571	-0.671033	-0.16403	1.447544	-0.911387	-1.897389	-0.298922	NaN	-2.053321	-0.973805	1.371395	-0.309283	-0.530942	0	0.03479
-0.878109	-0.531167	-0.177571	-0.671033	-0.16403	-0.292769	-0.688806	-0.698067	0.221063	-0.172754	-0.694433	-0.698995	-0.20785	0.421835	0.49779	0	0.083414
-0.878109	0.635952	-0.177571	-0.671033	-0.16403	0.188107	0.53539	0.04634	-0.732244	NaN	0.082948	0.629314	0.27658	-1.014198	-1.07641	0	0.127807
1.138811	-1.34815	-0.177571	1.49024	-0.16403	0.371298	0.023454	0.749392	-0.818908	0.281369	0.779496	0.107504	0.450276	-1.168148	-1.012333	0	0.127532
-0.878109	0.635952	-0.177571	-0.671033	-0.16403	0.256804	1.091843	-0.036372	0.047735	-0.834425	-0.001287	1.151351	0.342372	0.18998	0.154885	0	0.191723
-0.878109	-1.34815	-0.177571	-0.671033	-0.16403	-1.506407	-0.955903	-1.235694	0.047735	NaN	-1.285534	-1.030237	-1.667334	0.18998	0.330109	0	0.033189
1.138811	0.285817	-0.177571	-0.671033	-0.16403	0.073613	-1.712679	-1.359762	0.264395	-0.172754	-1.425906	-2.076216	0.165115	0.478132	0.702215	0	0.141031
-0.878109	1.686359	-0.177571	1.49024	-0.16403	0.829275	1.002811	0.25312	0.394392	NaN	0.291354	1.070682	0.861566	0.643265	0.725136	0	0.285803
1.138811	0.402529	-0.177571	-0.671033	-0.16403	-1.002633	0.4241	0.501256	-0.168926	NaN	0.537482	0.519501	-1.013082	-0.116081	0.085027	0	0.162363
1.138811	1.803071	-0.177571	1.49024	-0.16403	-0.407263	0.691197	0.583968	0.654384	-0.172754	0.6186	0.779903	-0.329883	0.957798	0.746124	0	0.472938
1.138811	-0.414455	NaN	1.49024	-0.16403	0.371298	0.646681	0.170408	-0.298922	1.04317	0.20836	0.737244	0.450276	-0.309283	-0.336624	0	0.25223
1.138811	1.336224	-0.177571	-0.671033	-0.16403	-0.292769	-0.377192	-0.739423	-0.298922	NaN	-0.738962	-0.333372	-0.20785	-0.309283	-0.307444	0	0.208916
-0.878109	-0.414455	-0.177571	-0.671033	-0.16403	-0.407263	0.78023	-0.160439	0.481056	NaN	-0.128602	0.864363	-0.329883	0.750357	0.637655	0	0.086605
-0.878109	-0.531167	-0.177571	-0.671033	-0.16403	-0.773644	-0.065579	0.170408	-0.515583	1.04317	0.20836	0.012166	-0.740002	-0.649142	-0.717637	0	0.112188
-0.878109	-0.181031	-0.177571	-0.671033	-0.16403	-0.155376	-0.154611	0.129052	NaN	NaN	0.16668	-0.084638	-0.064956	NaN	NaN	0	0.068579
-0.878109	0.519241	-0.177571	-0.671033	-0.16403	-0.407263	0.490874	0.501256	0.611052	NaN	0.537482	0.585619	-0.329883	0.906751	0.791078	0	0.126541
-0.878109	1.336224	-0.177571	1.49024	-0.16403	1.218555	1.670554	1.245663	0.091067	NaN	1.251963	1.651909	1.188128	0.248975	-0.110625	0	0.278056
-0.878109	0.519241	-0.177571	-0.671033	-0.16403	0.188107	-0.688806	-0.573999	-0.25559	NaN	-0.561732	-0.698995	0.27658	-0.244037	-0.273229	0	0.093492
-0.878109	-0.297743	-0.177571	-0.671033	-0.16403	-1.872789	-0.443967	-0.449931	-0.168926	-1.795458	-0.430335	-0.409951	-2.198396	-1.16081	-0.320166	0	0.069041
1.138811	1.919783	-0.177571	1.49024	-0.16403	0.897971	1.180876	-0.077728	-0.732244	NaN	-0.043596	1.231	0.920647	-1.014198	-0.724132	0	0.407327
1.138811	0.752664	-0.177571	-0.671033	-0.16403	0.417096	0.246035	-0.243151	-0.125594	0.68146	-0.214135	0.339713	0.492842	-0.053329	-0.384738	0	0.297056
-0.878109	-0.998014	-0.177571	-0.671033	-0.16403	-1.323217	-0.733322	-1.318406	-0.602247	NaN	-1.378941	-0.753005	-1.420166	-0.791934	-0.911751	0	0.040728
-0.878109	-0.414455	-0.177571	-0.671033	-0.16403	-1.071329	-1.11171	-0.243151	-0.298922	NaN	-0.214135	-1.231834	-1.097771	-0.309283	-0.198485	0	0.050195
-0.878109	-0.181031	-0.177571	-0.671033	-0.16403	-0.956835	-1.044936	-0.987558	-0.732244	-0.834425	-1.00936	-1.144643	-0.957352	-1.014198	-1.154322	0	0.073998
-0.878109	0.869376	-0.177571	-0.671033	-0.16403	0.302601	-0.332676	-0.739423	-0.428919	-0.172754	-0.738962	-0.282833	0.385792	-0.510372	-1.235086	0	0.170751
-0.878109	-0.297743	-0.177571	1.49024	-0.16403	-0.338566	1.336682	0.790747	-0.082262	NaN	0.819446	1.368004	-0.256333	0.008634	0.020822	0	0.12295
1.138811	-0.414455	-0.177571	-0.671033	-0.16403	-0.040881	0.78023	1.121595	-0.168926	0.281369	1.135234	0.864363	0.051303	-0.116081	-0.161967	0	0.190158