John Daniel S. Detablan

12001619

Task 1

Complete Data
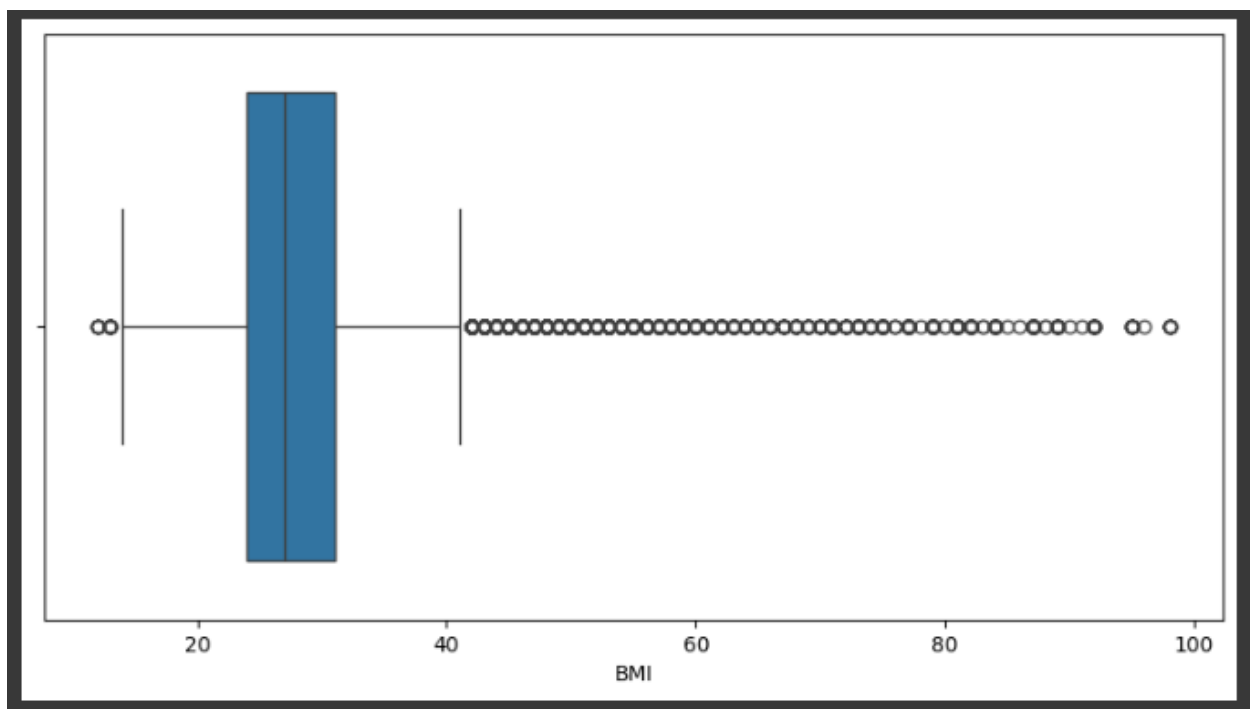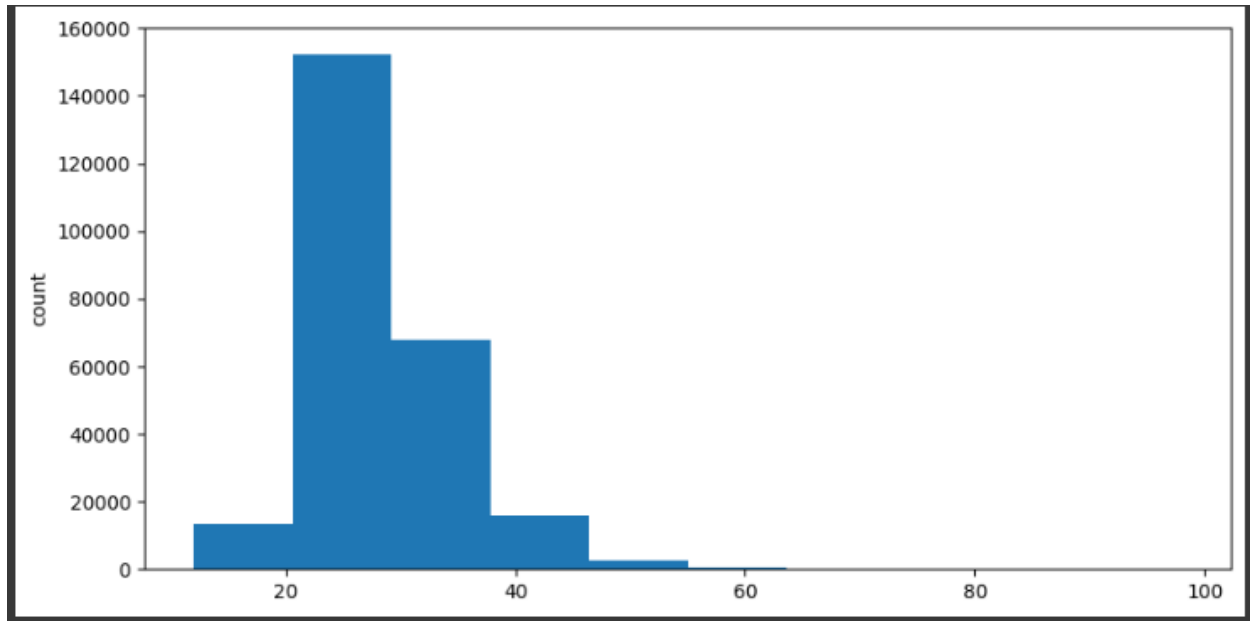
	HighBP	HighChol	CholCheck	BMI	Smoker	Stroke	HeartDiseaseorAttack	PhysActivity	Fruits	Veggies	...	NoDocbcCost	GenHlth	MentHlth	PhysHlth	DiffWalk	Sex	Age	Education	Income	Diabetes_binary
0	1	1	1	40	1	0	0	0	0	1	...	0	5	18	15	1	0	9	4	3	0
1	0	0	0	25	1	0	0	1	0	0	...	1	3	0	0	0	0	7	6	1	0
2	1	1	1	28	0	0	0	0	1	0	...	1	5	30	30	1	0	9	4	8	0
3	1	0	1	27	0	0	0	1	1	1	...	0	2	0	0	0	0	11	3	6	0
4	1	1	1	24	0	0	0	1	1	1	...	0	2	3	0	0	0	11	5	4	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
253675	1	1	1	45	0	0	0	0	1	1	...	0	3	0	5	0	1	5	6	7	0
253676	1	1	1	18	0	0	0	0	0	0	...	0	4	0	0	1	0	11	2	4	1
253677	0	0	1	28	0	0	0	1	1	0	...	0	1	0	0	0	0	2	5	2	0
253678	1	0	1	23	0	0	0	0	1	1	...	0	3	0	0	0	1	7	5	1	0
253679	1	1	1	25	0	0	1	1	1	0	...	0	2	0	0	0	0	9	6	2	1

253680 rows × 22 columns

Statistical Summary

|	count	mean	std	min	25%	50%	75%	max
HighBP	253680.0	0.429001	0.494934	0.0	0.0	0.0	1.0	1.0
HighChol	253680.0	0.424121	0.494210	0.0	0.0	0.0	1.0	1.0
CholCheck	253680.0	0.962670	0.189571	0.0	1.0	1.0	1.0	1.0
BMI	253680.0	28.382364	6.608694	12.0	24.0	27.0	31.0	98.0
Smoker	253680.0	0.443169	0.496761	0.0	0.0	0.0	1.0	1.0
Stroke	253680.0	0.040571	0.197294	0.0	0.0	0.0	0.0	1.0
HeartDiseaseorAttack	253680.0	0.094186	0.292087	0.0	0.0	0.0	0.0	1.0
PhysActivity	253680.0	0.756544	0.429169	0.0	1.0	1.0	1.0	1.0
Fruits	253680.0	0.634256	0.481639	0.0	0.0	1.0	1.0	1.0
Veggies	253680.0	0.811420	0.391175	0.0	1.0	1.0	1.0	1.0
HvyAlcoholConsump	253680.0	0.056197	0.230302	0.0	0.0	0.0	0.0	1.0
AnyHealthcare	253680.0	0.951053	0.215759	0.0	1.0	1.0	1.0	1.0
NoDocbcCost	253680.0	0.084177	0.277654	0.0	0.0	0.0	0.0	1.0
GenHlth	253680.0	2.511392	1.068477	1.0	2.0	2.0	3.0	5.0
MentHlth	253680.0	3.184772	7.412847	0.0	0.0	0.0	2.0	30.0
PhysHlth	253680.0	4.242081	8.717951	0.0	0.0	0.0	3.0	30.0
DiffWalk	253680.0	0.168224	0.374066	0.0	0.0	0.0	0.0	1.0
Sex	253680.0	0.440342	0.496429	0.0	0.0	0.0	1.0	1.0
Age	253680.0	8.032119	3.054220	1.0	6.0	8.0	10.0	13.0
Education	253680.0	5.050434	0.985774	1.0	4.0	5.0	6.0	6.0
Income	253680.0	6.053875	2.071148	1.0	5.0	7.0	8.0	8.0
Diabetes_binary	253680.0	0.139333	0.346294	0.0	0.0	0.0	0.0	1.0

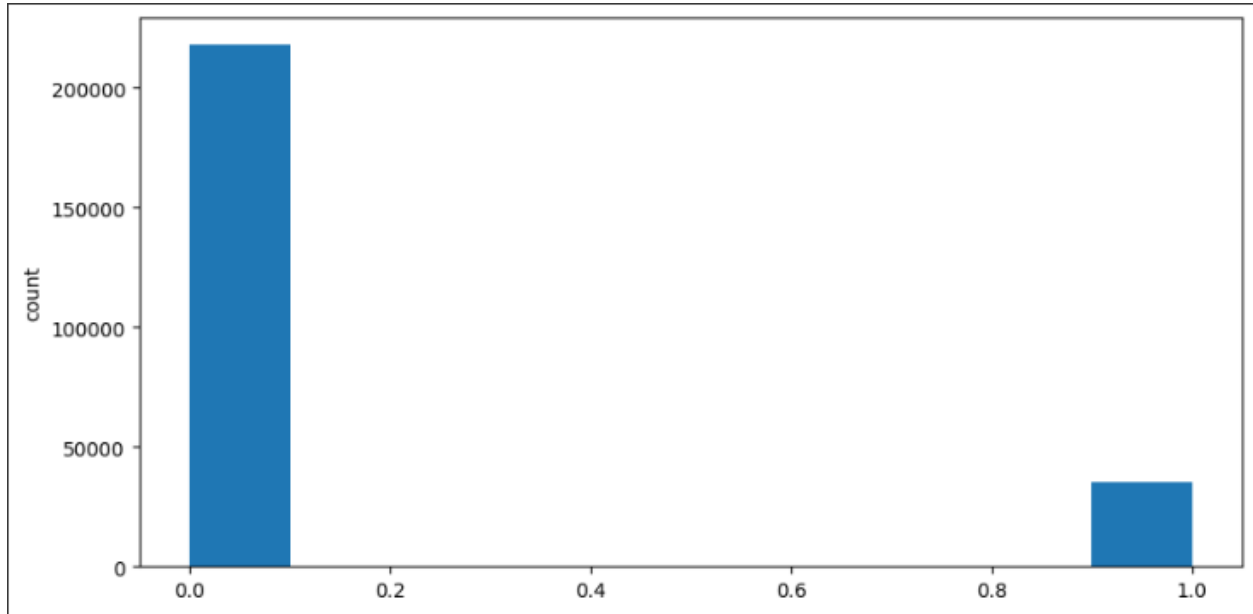Histogram and Box Plot of the BMI Data

Histogram and Box Plot of the Age Data
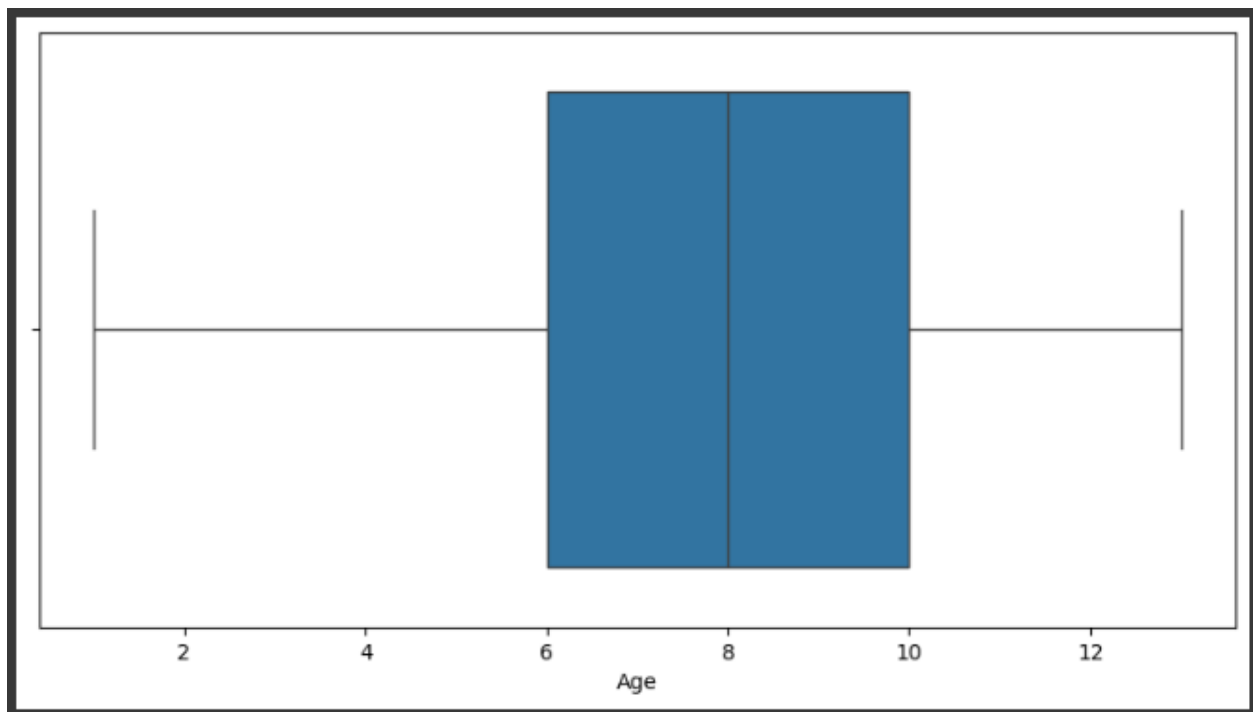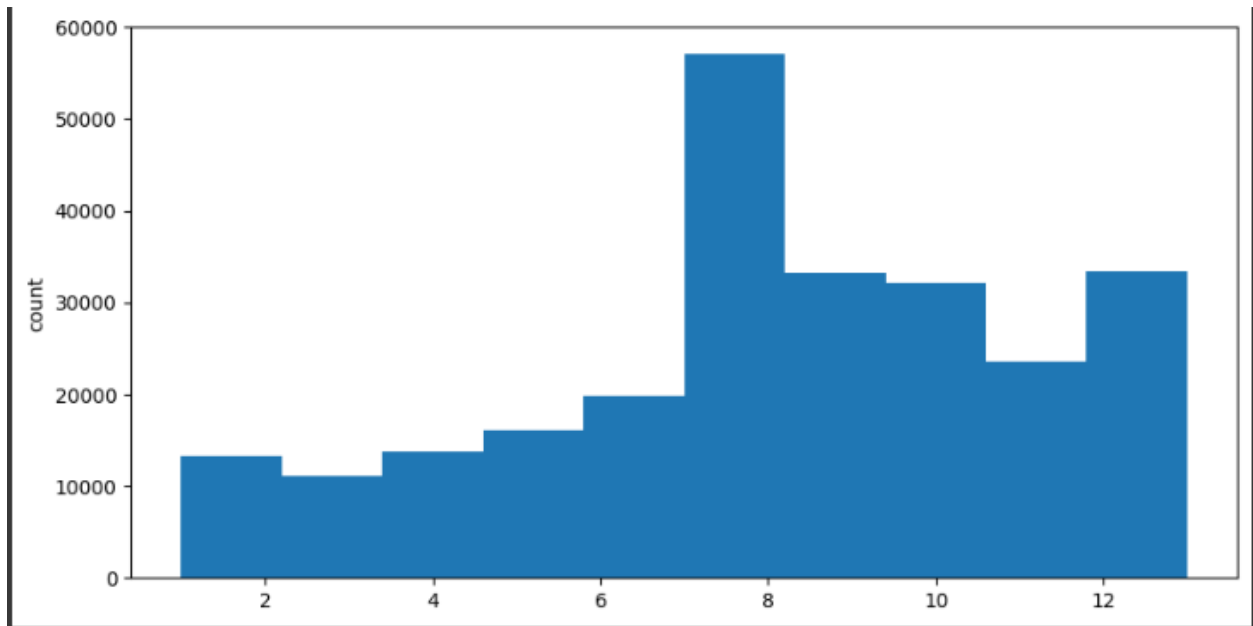
Histogram and Box Plot of the General Health Data
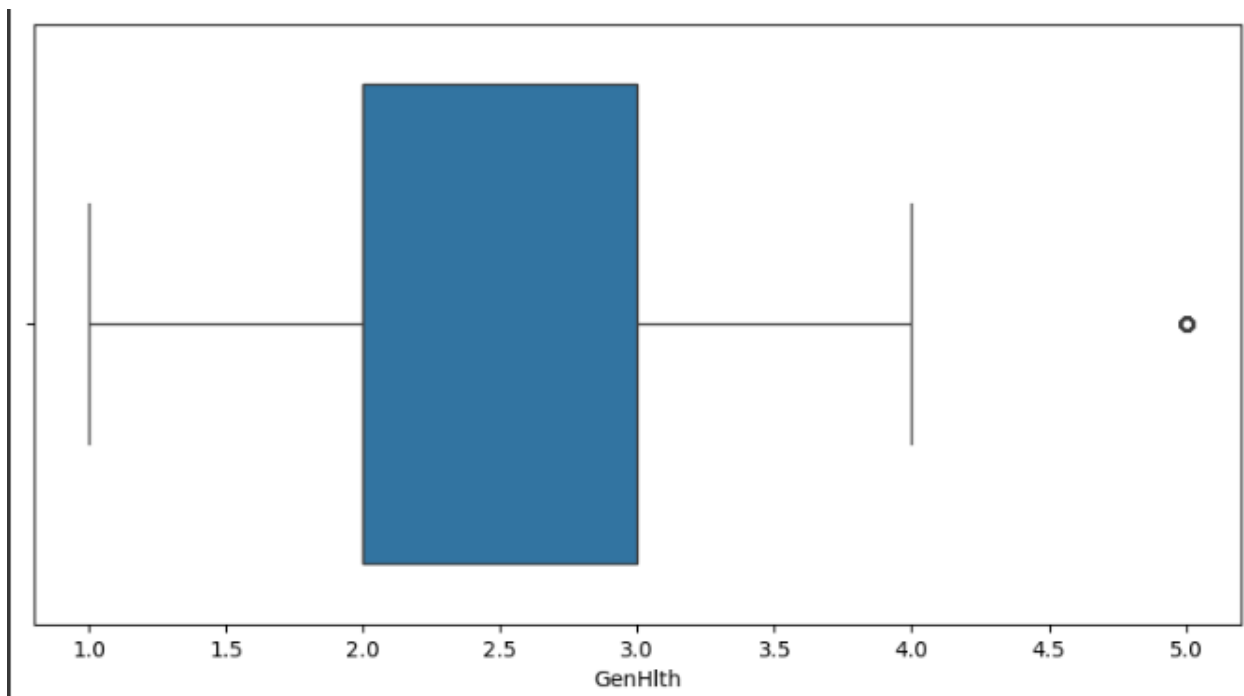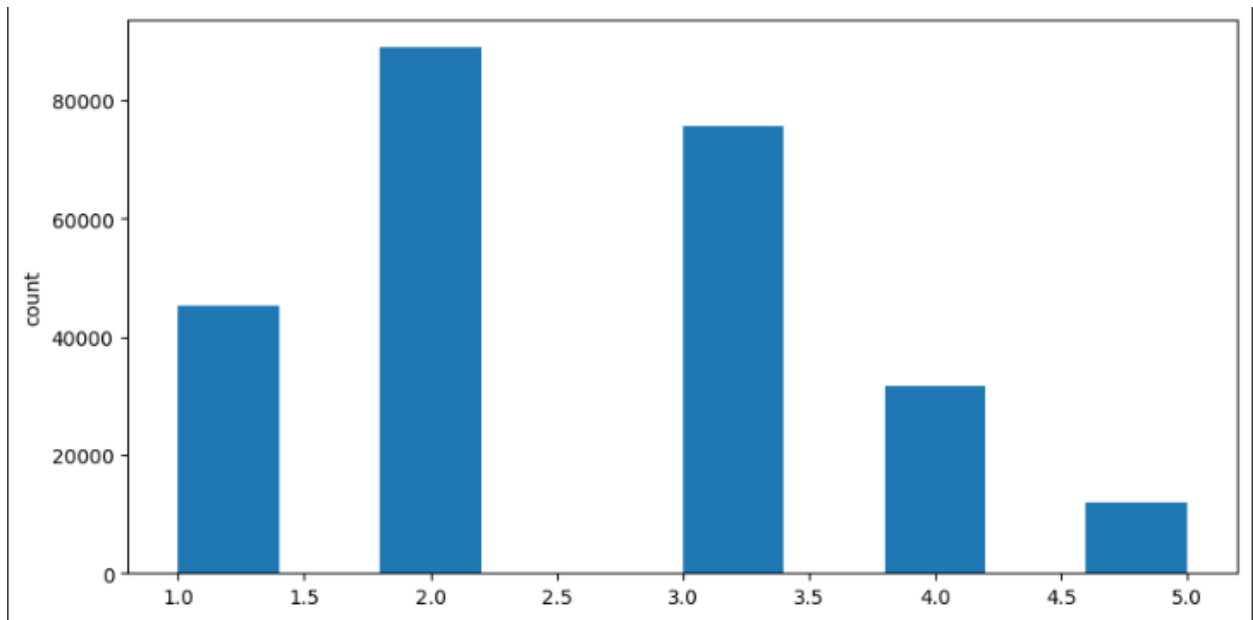
Histogram of the Diabetes_binary Data

Correlation Analysis

|	HighBP	HighChol	CholCheck	BMI	Smoker	Stroke	HeartDiseaseorAttack	PhysActivity	Fruits	Veggies	...	NoDocbcCost	GenHlth	MentHlth	PhysHlth	DiffWalk	Sex	Age	Education	Income	Diabetes_binary
HighBP	1.000000	0.298199	0.098508	0.213748	0.096991	0.129575	0.209361	-0.125267	-0.040555	-0.061266	...	0.017358	0.300530	0.056456	0.161212	0.223618	0.052207	0.344452	-0.141358	-0.171235	0.263129
HighChol	0.298199	1.000000	0.085642	0.106722	0.091299	0.092620	0.180765	-0.078046	-0.040859	-0.039874	...	0.013310	0.208426	0.062069	0.121751	0.144672	0.031205	0.272318	-0.070802	-0.085459	0.200276
CholCheck	0.098508	0.085642	1.000000	0.034495	-0.009929	0.024158	0.044206	0.004190	0.023849	0.006121	...	-0.058255	0.046589	-0.008366	0.031775	0.040585	-0.022115	0.090321	0.001510	0.014259	0.064761
BMI	0.213748	0.106722	0.034495	1.000000	0.013804	0.020153	0.052904	-0.147294	-0.087518	-0.062275	...	0.058206	0.239185	0.085310	0.121141	0.197078	0.042950	-0.036618	-0.103932	-0.100069	0.216843
Smoker	0.096991	0.091299	-0.009929	0.013804	1.000000	0.061173	0.114441	-0.087401	-0.077666	-0.030678	...	0.048946	0.163143	0.092196	0.116460	0.122463	0.093662	0.120641	-0.161955	-0.123937	0.060789
Stroke	0.129575	0.092620	0.024158	0.020153	0.061173	1.000000	0.203002	-0.069151	-0.013389	-0.041124	...	0.034804	0.177942	0.070172	0.148944	0.176567	0.002978	0.126974	-0.076009	-0.128599	0.105816
HeartDiseaseorAttack	0.209361	0.180765	0.044206	0.052904	0.114441	0.203002	1.000000	-0.087299	-0.019790	-0.039167	...	0.031000	0.258383	0.064621	0.181698	0.212709	0.086096	0.221618	-0.099600	-0.141011	0.177282
PhysActivity	-0.125267	-0.078046	0.004190	-0.147294	-0.087401	-0.069151	-0.087299	1.000000	0.142756	0.153150	...	-0.061638	-0.266186	-0.125587	-0.219230	-0.253174	0.032482	-0.092511	0.199658	0.198539	-0.118133
Fruits	-0.040555	-0.040859	0.023849	-0.087518	-0.077666	-0.013389	-0.019790	0.142756	1.000000	0.254342	...	-0.044243	-0.103854	-0.068217	-0.044633	-0.048352	-0.091175	0.064547	0.110187	0.079929	-0.040779
Veggies	-0.061266	-0.039874	0.006121	-0.062275	-0.030678	-0.041124	-0.039167	0.153150	0.254342	1.000000	...	-0.032232	-0.123066	-0.058884	-0.064290	-0.080506	-0.064765	-0.009771	0.154329	0.151087	-0.056584
HvyAlcoholConsump	-0.003972	-0.011543	-0.023730	-0.048736	0.101619	-0.016950	-0.028991	0.012392	-0.035288	0.021064	...	0.004684	-0.036724	0.024716	-0.026415	-0.037668	0.005740	-0.034578	0.023997	0.053619	-0.057056
AnyHealthcare	0.038425	0.042230	0.117626	-0.018471	-0.023251	0.008776	0.018734	0.035505	0.031544	0.029584	...	-0.232532	-0.040817	-0.052707	-0.008276	0.007074	-0.019405	0.138046	0.122514	0.157999	0.016255
NoDocbcCost	0.017358	0.013310	-0.058255	0.058206	0.048946	0.034804	0.031000	-0.061638	-0.044243	-0.032232	...	1.000000	0.166397	0.192107	0.148998	0.118447	-0.044931	-0.119777	-0.100701	-0.203182	0.031433
GenHlth	0.300530	0.208426	0.046589	0.239185	0.163143	0.177942	0.258383	-0.266186	-0.103854	-0.123066	...	0.166397	1.000000	0.301674	0.524364	0.456920	-0.006091	0.152450	-0.284912	-0.370014	0.293569
MentHlth	0.056456	0.062069	-0.008366	0.085310	0.092196	0.070172	0.064621	-0.125587	-0.068217	-0.058884	...	0.192107	0.301674	1.000000	0.353619	0.233688	-0.080705	-0.092068	-0.101830	-0.209806	0.069315
PhysHlth	0.161212	0.121751	0.031775	0.121141	0.116460	0.148944	0.181698	-0.219230	-0.044633	-0.064290	...	0.148998	0.524364	0.353619	1.000000	0.478417	-0.043137	0.099130	-0.155093	-0.266799	0.171337
DiffWalk	0.223618	0.144672	0.040585	0.197078	0.122463	0.176567	0.212709	-0.253174	-0.048352	-0.080506	...	0.118447	0.456920	0.233688	0.478417	1.000000	-0.070299	0.204450	-0.192642	-0.320124	0.218344
Sex	0.052207	0.031205	-0.022115	0.042950	0.093662	0.002978	0.086096	0.032482	-0.091175	-0.064765	...	-0.044931	-0.006091	-0.080705	-0.043137	-0.070299	1.000000	-0.027340	0.019480	0.127141	0.031430
Age	0.344452	0.272318	0.090321	-0.036618	0.120641	0.126974	0.221618	-0.092511	0.064547	-0.009771	...	-0.119777	0.152450	-0.092068	0.099130	0.204450	-0.027340	1.000000	-0.101901	-0.127775	0.177442
Education	-0.141358	-0.070802	0.001510	-0.103932	-0.161955	-0.076009	-0.099600	0.199658	0.110187	0.154329	...	-0.100701	-0.284912	-0.101830	-0.155093	-0.192642	0.019480	-0.101901	1.000000	0.449106	-0.124456
Income	-0.171235	-0.085459	0.014259	-0.100069	-0.123937	-0.128599	-0.141011	0.198539	0.079929	0.151087	...	-0.203182	-0.370014	-0.209806	-0.266799	-0.320124	0.127141	-0.127775	0.449106	1.000000	-0.163919
Diabetes_binary	0.263129	0.200276	0.064761	0.216843	0.060789	0.105816	0.177282	-0.118133	-0.040779	-0.056584	...	0.031433	0.293569	0.069315	0.171337	0.218344	0.031430	0.177442	-0.124456	-0.163919	1.000000

Task 2

From the Statistical Summary shown above, the participants have an average age of 8 years old and has an average BMI of 28.38. Now, according to an online article of Food and Nutrition Technical Assistance, given that a child is around 8 years old, and their BMI is greater than or equal to 20.7, they are under the Obese class. However, despite of this, the histogram for General Health and Diabetes_binary suggests that these majority of the children are in good health.
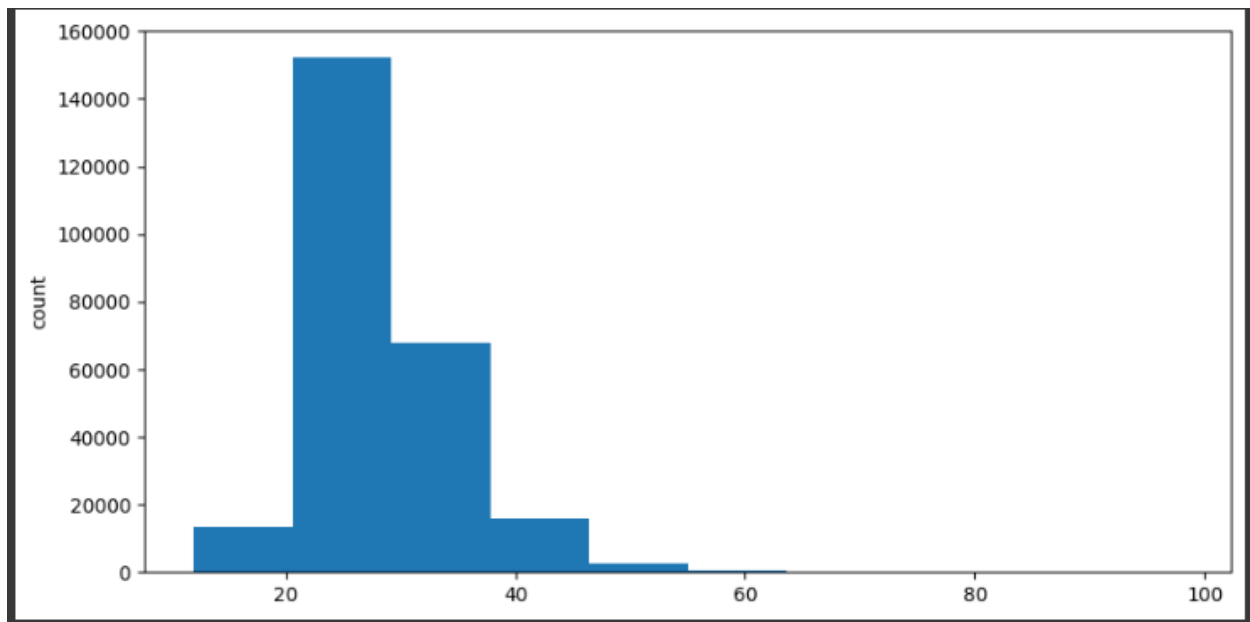
Moreover, the correlation analysis also suggests that there is a positive correlation between Diabetes_binary and the following variables: BMI, GenHlth, and Age.
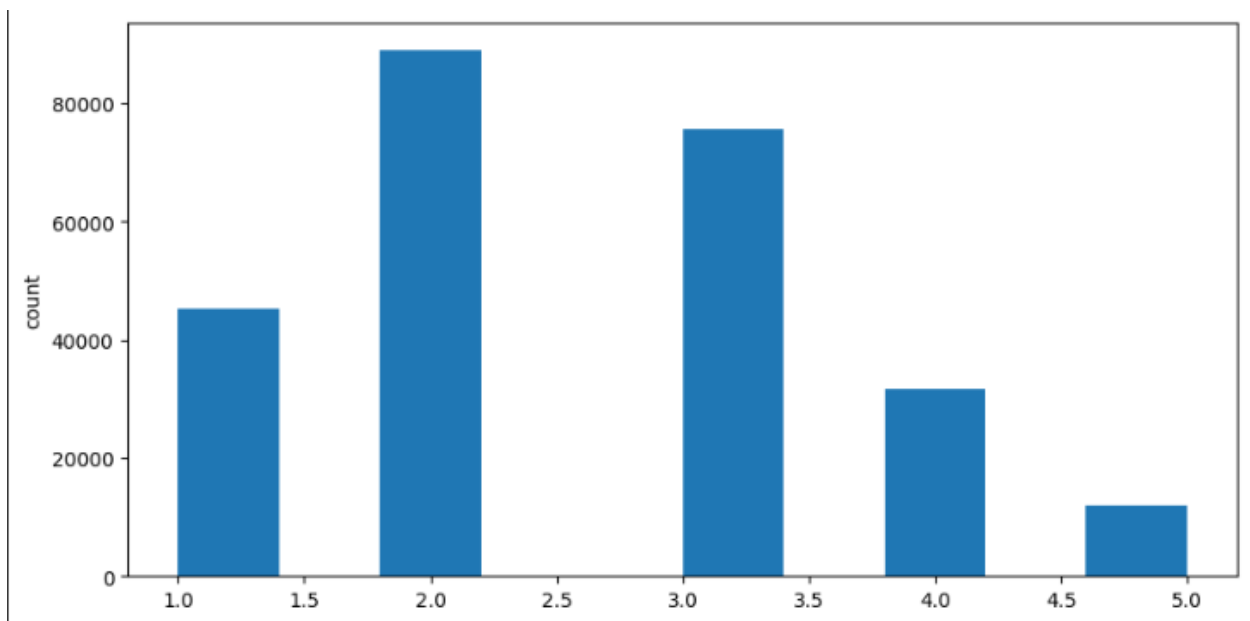
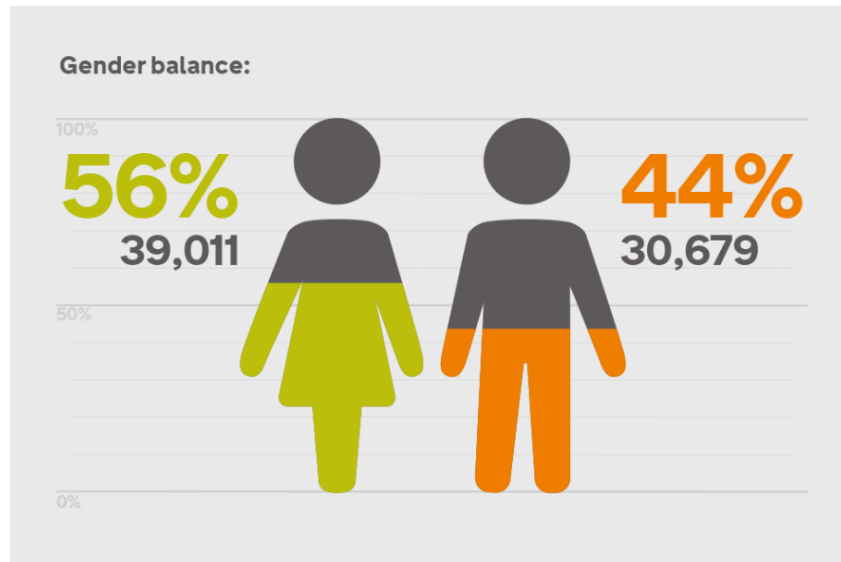# Task 3

Heat Map for Correlation Analysis
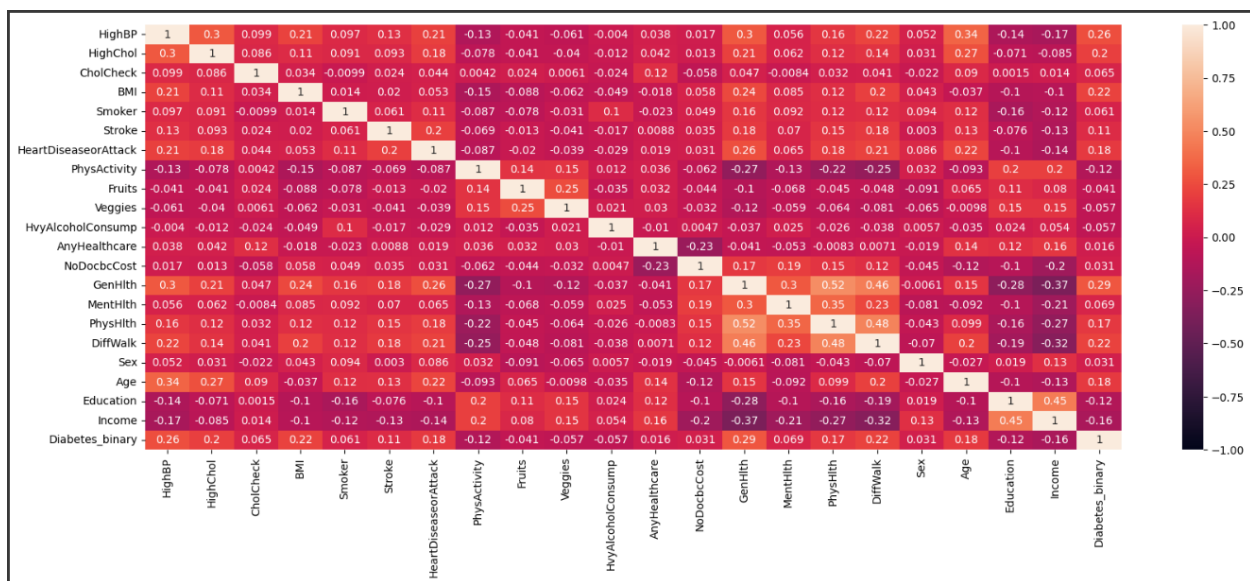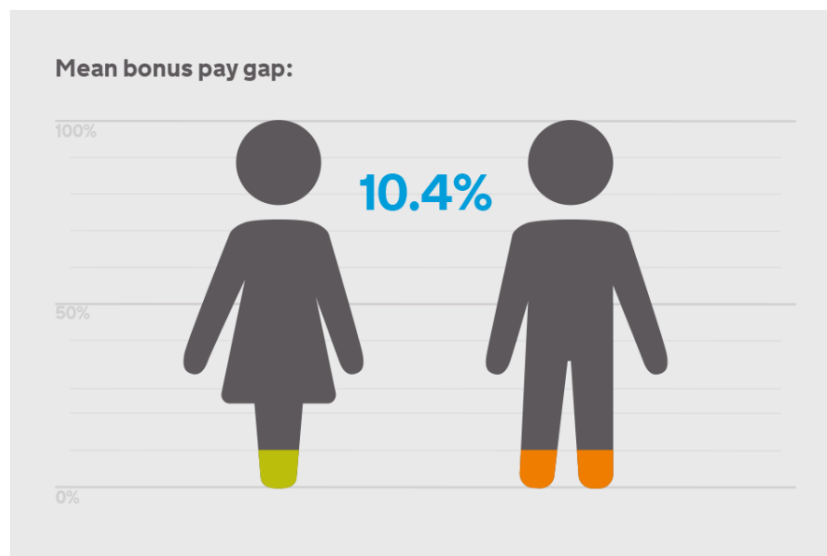
Histogram for BMI Data

Histogram for Diabetes

Task 4

This figure isn't that clear with what it is trying to say or portray. What does "Gender balance" mean? Also, there is no clear indication of what the numbers are or what they represent. Also, the design is a bit much I would say. This kind of design also ruined this graph below

It just says mean bonus pay gap of 10.4% without having a clear understanding whether men do have the higher bonus pay or is it the women, because the figure isn't clear about that as well.

So, if I were to redesign this, a simple bar graph would be a good one with pink indicating the bar for women and blue for men. For me, in terms of design and aesthetic, less is more. In addition, if I were to make a similar graph like this one, I would add few more texts in it to indicate what the numbers are representing and to be clear with what I am trying to compare.

Google Colab Link

Here is the Google Colab link for the python code I made:
https://colab.research.google.com/drive/1kAhS4slI0W7Ttyyj3VlzHTIKDoqY3_y8?usp=sharing

References

Food and Nutrition Technical Assistance. (2013). *BMI and BMI-for-Age Look-Up tables for children and adolescents 5–18 years of age and BMI Look-Up tables for Non-Pregnant, Non-Lactating Adults ≥ 19 years of age*.

https://www.fantaproject.org/sites/default/files/resources/FANTA-BMI-charts-Jan2013-ENG_0.pdf