

Exploratory Data Analysis Report

Sala2Code

March 9, 2024

1 Dataset Overview

1.1 General Information

- Size: 20758 rows / 18 columns
- Target column: *NObeyesdad*

Table 1: Dataset Columns Overview

Column Name	Data Type ¹	NULL
id	int	0
Gender	string	0
Age	int	0
Height	int	0
Weight	int	0
family_history_with_overweight	string	0
FAVC	string	0
FCVC	int	0
NCP	int	0
CAEC	string	0
SMOKE	string	0
CH2O	int	0
SCC	string	0
FAF	float	0
TUE	float	0
CALC	string	0
MTRANS	string	0
NObeyesdad	string	0

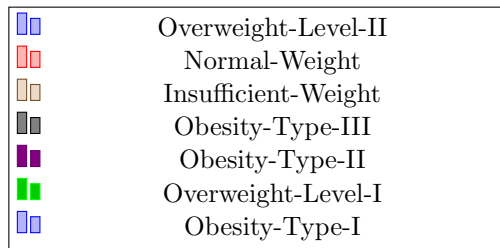
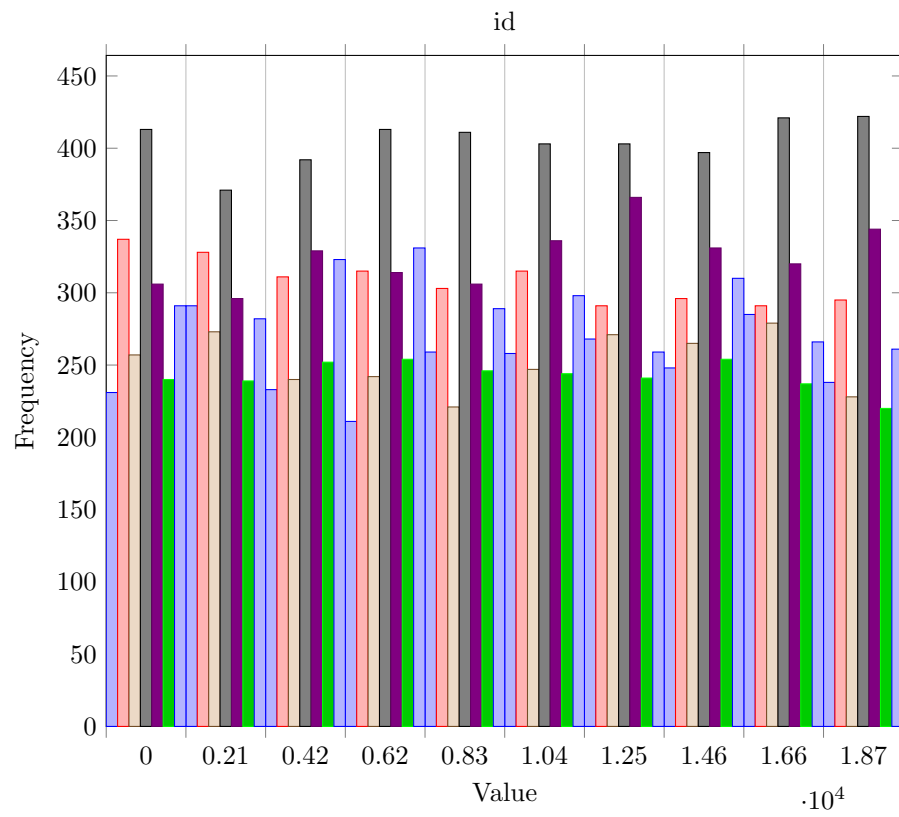
Note: Blue represents continuous variables, Red represents categorical variables.

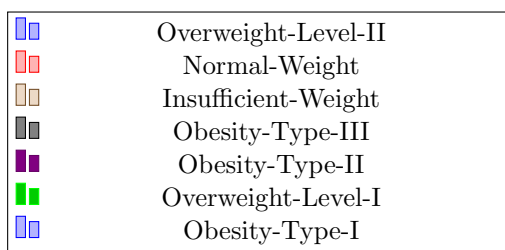
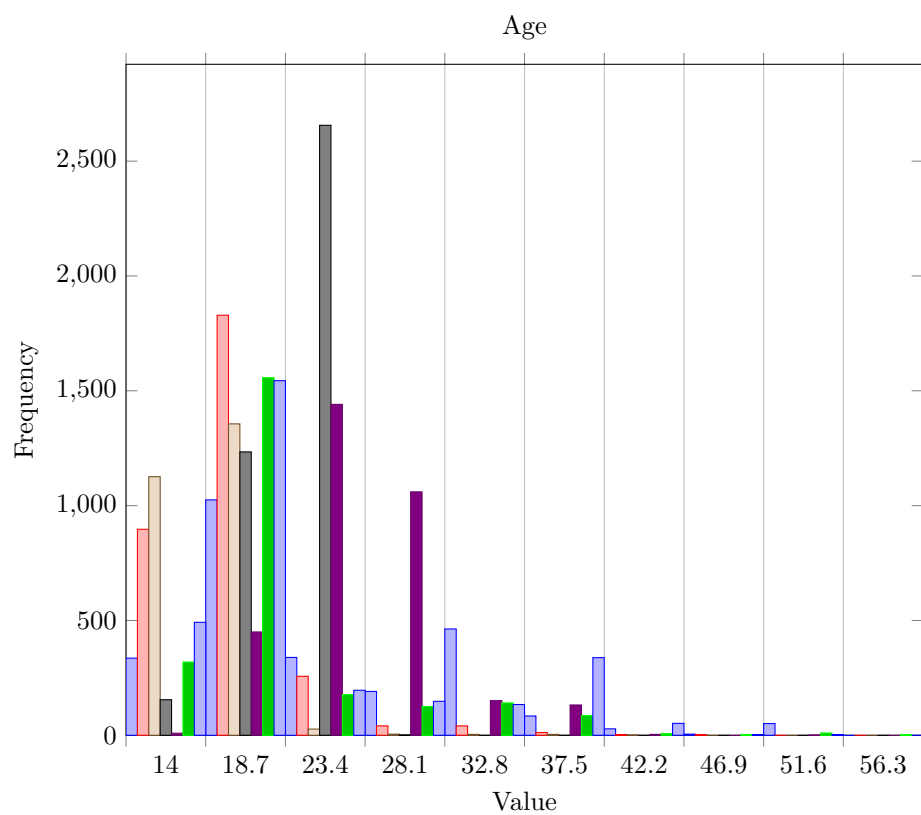
¹Consider the first value in each column. Therefore, it is not precise.

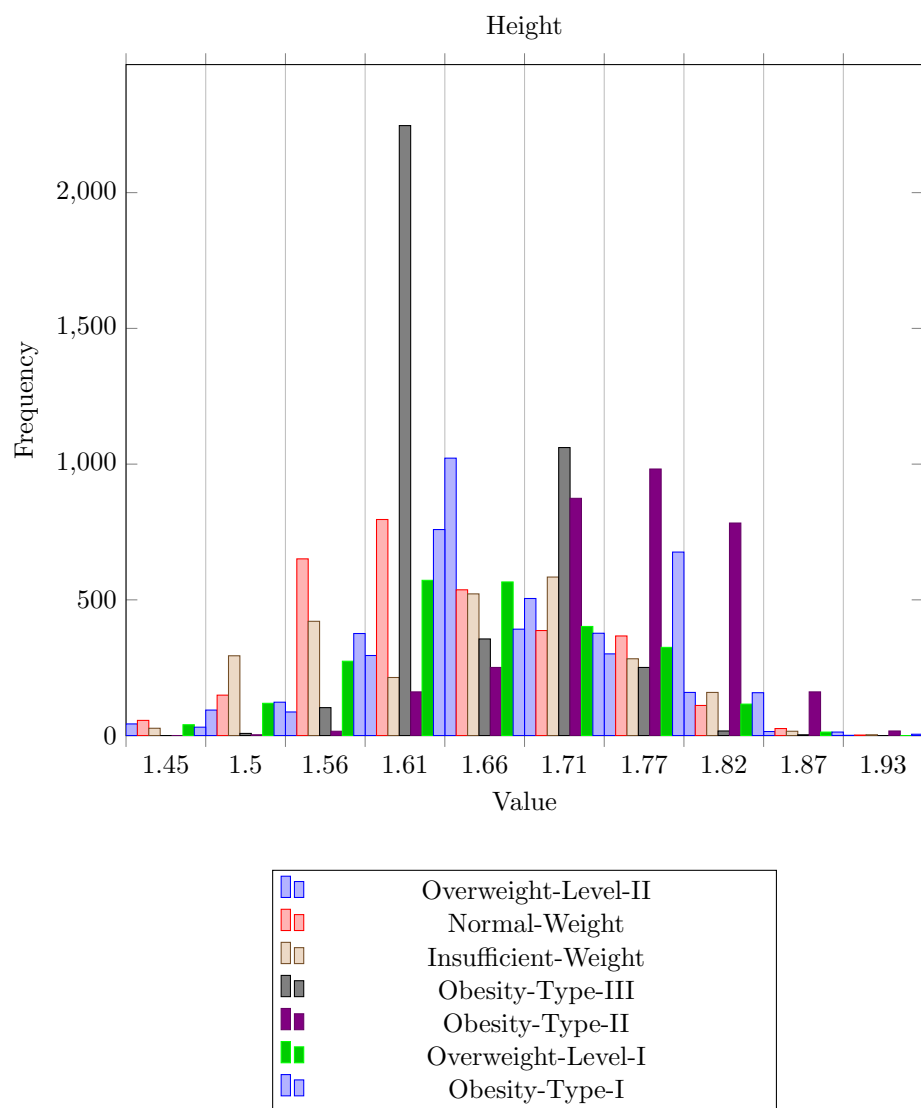
1.2 Continuous Columns

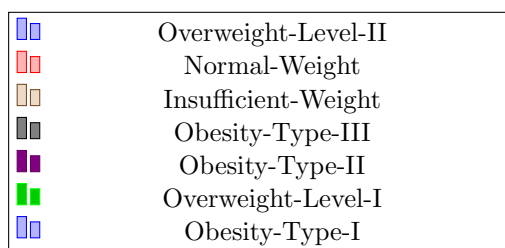
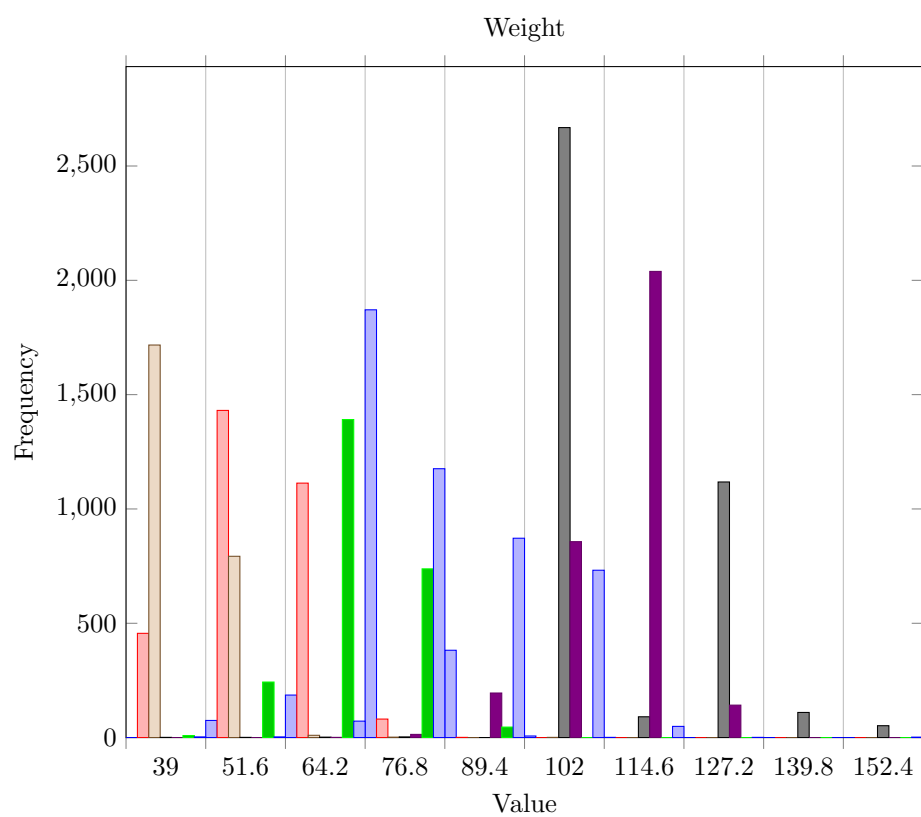
Column Name	Mean	Median	Min	Max	Variance	Std	Outliers
id	1.04e+04	1.04e+04	0	2.08e+04	3.59e+07	5.99e+03	0
Age	23.8	22.8	14	61	32.4	5.69	1074
Height	1.7	1.7	1.45	1.98	0.00762	0.0873	4
Weight	87.9	84.1	39	165	696	26.4	0
FCVC	2.45	2.39	1	3	0.284	0.533	0
NCP	2.76	3	1	4	0.498	0.705	6052
CH2O	2.03	2	1	3	0.37	0.608	0
FAF	0.982	1	0	3	0.703	0.838	0
TUE	0.617	0.574	0	2	0.363	0.602	0

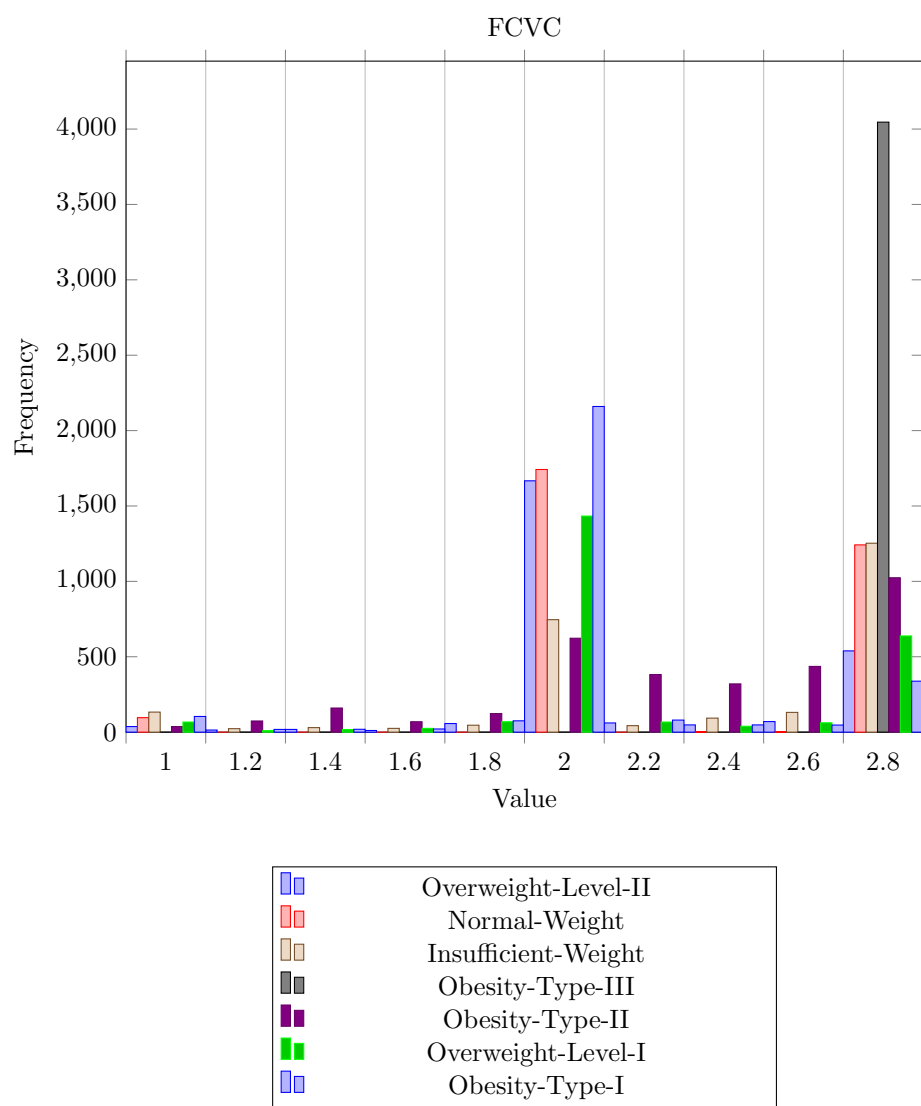
Table 2: Data types and null value counts for each column.

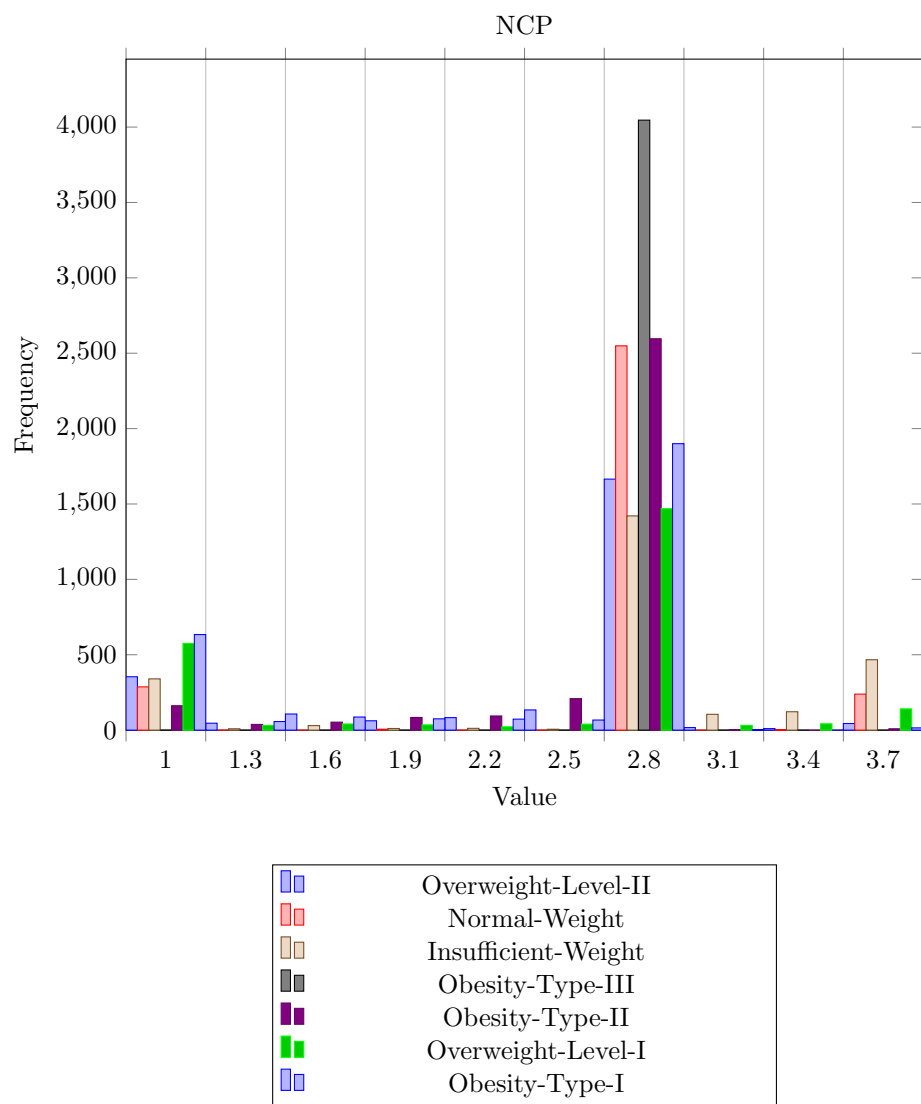


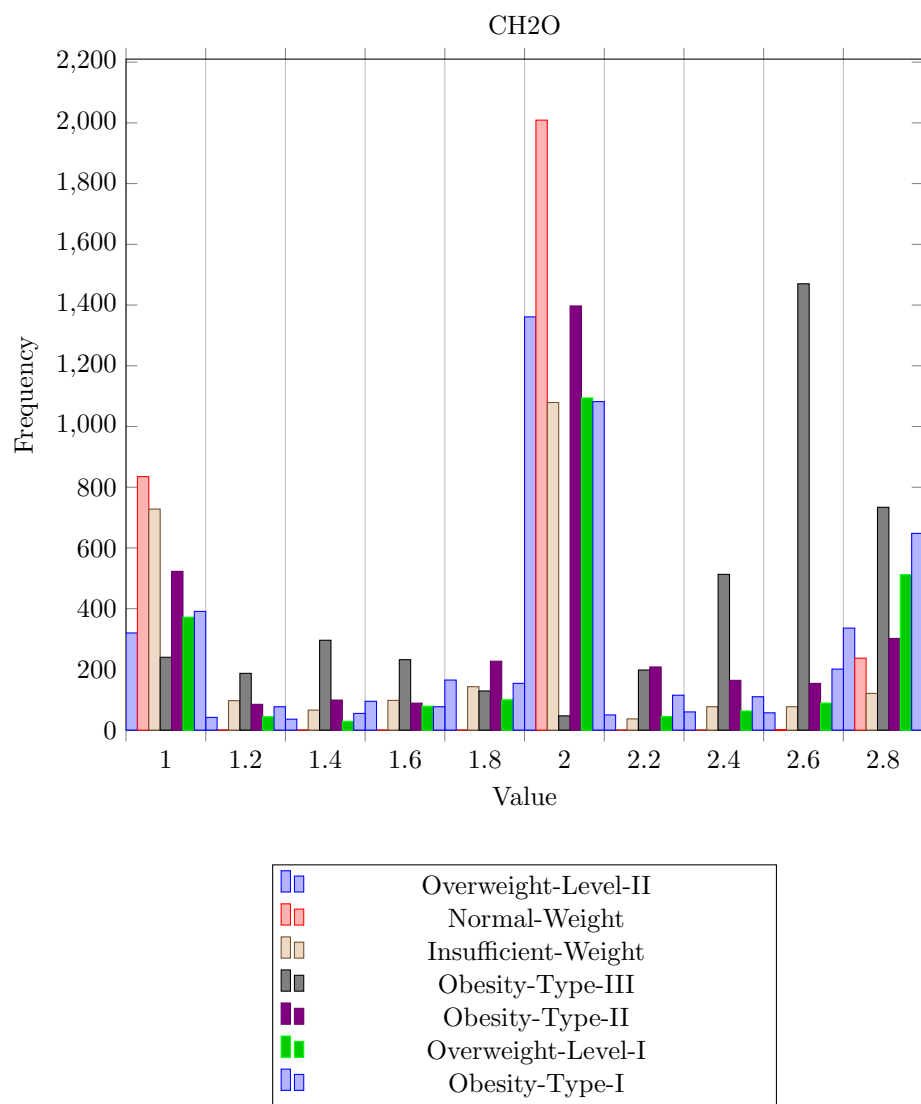


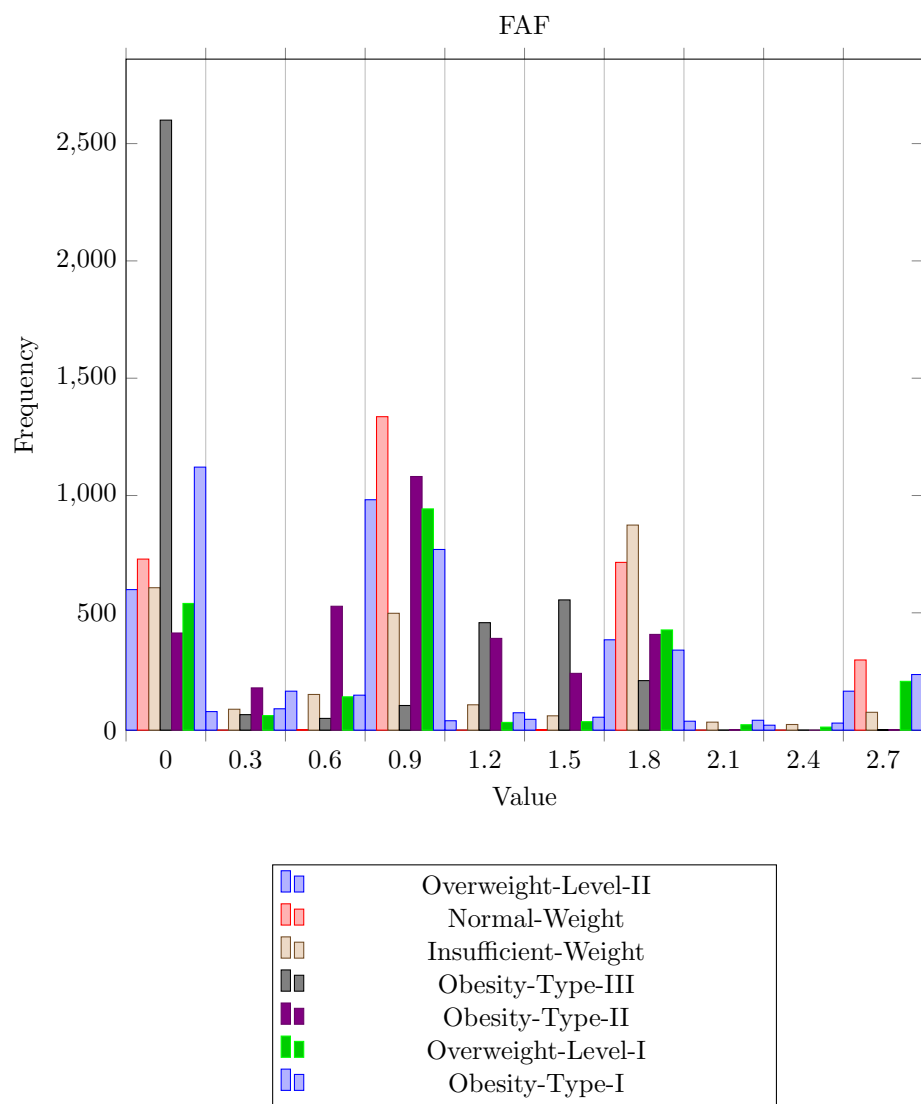


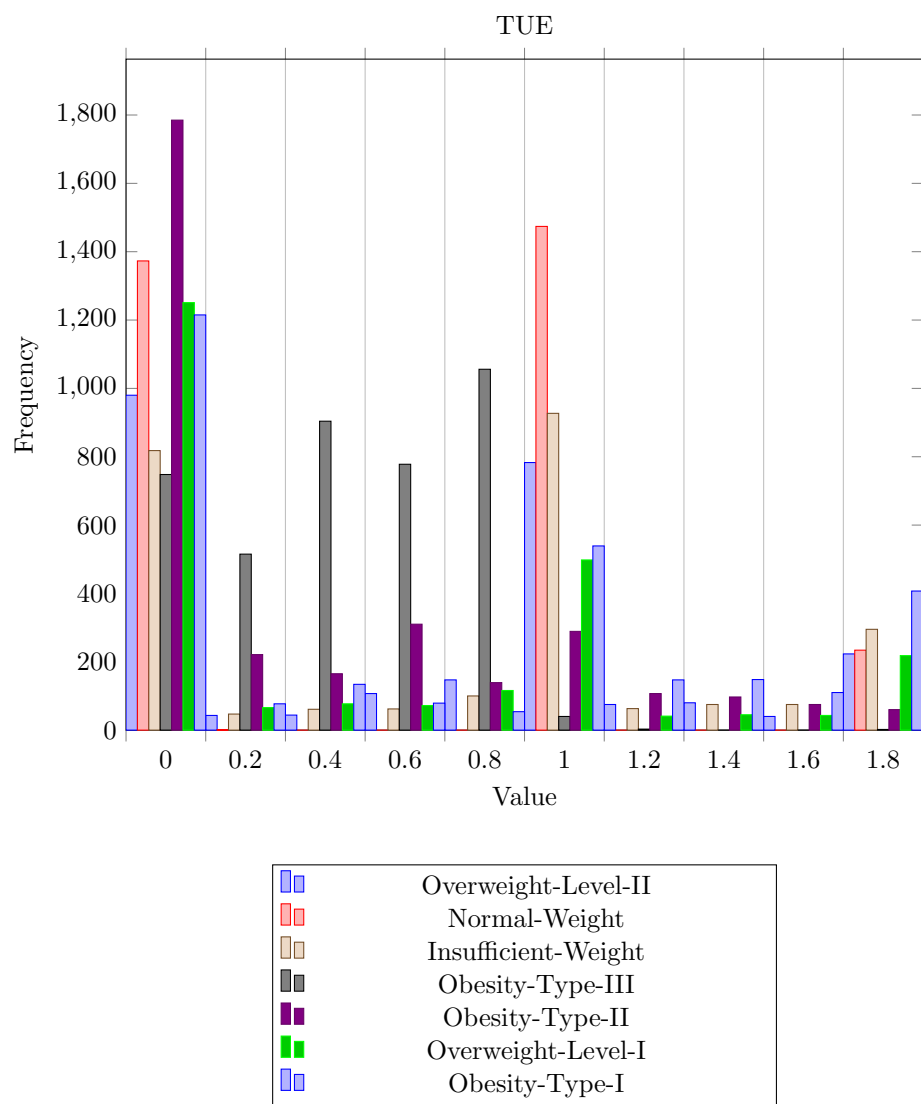




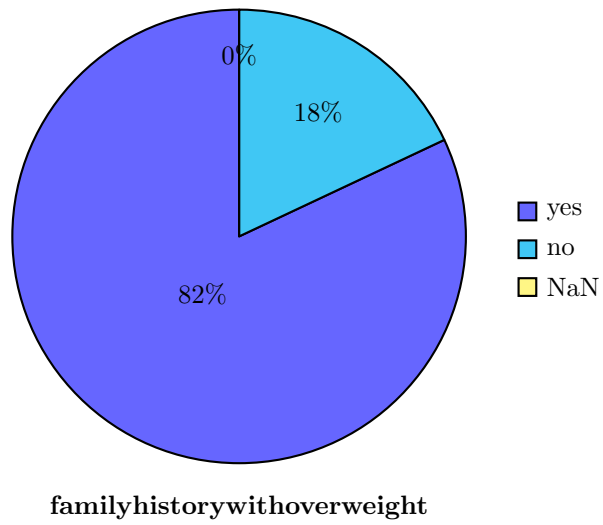
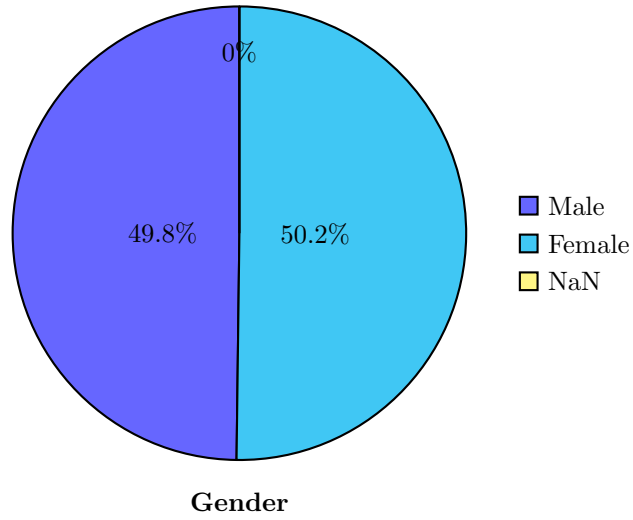


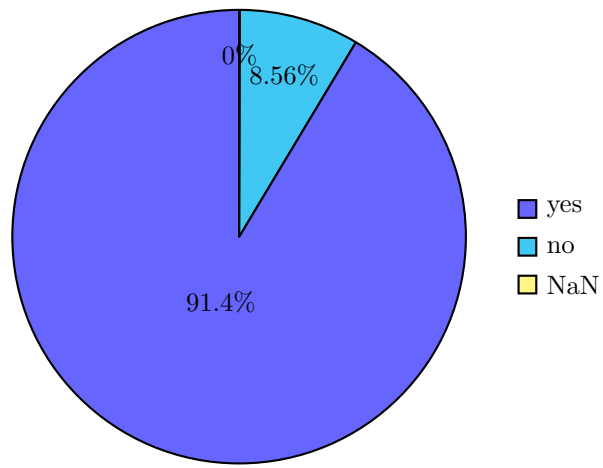




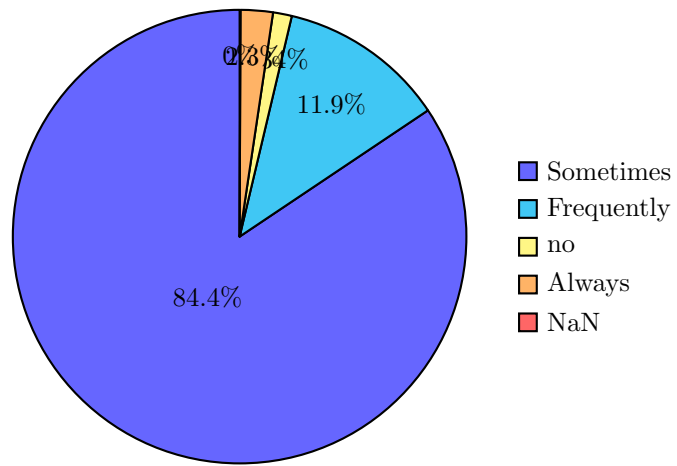


1.3 Categorical Columns

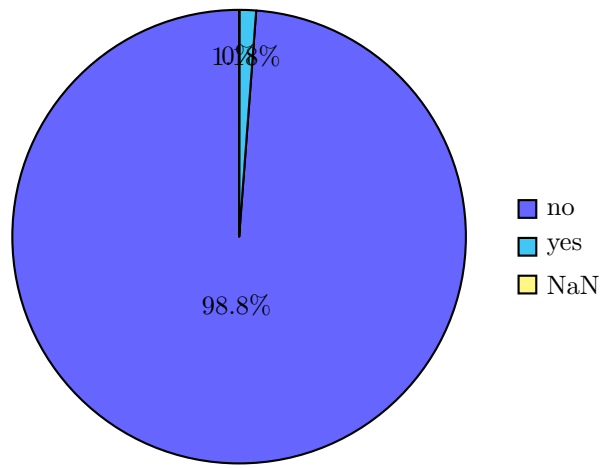




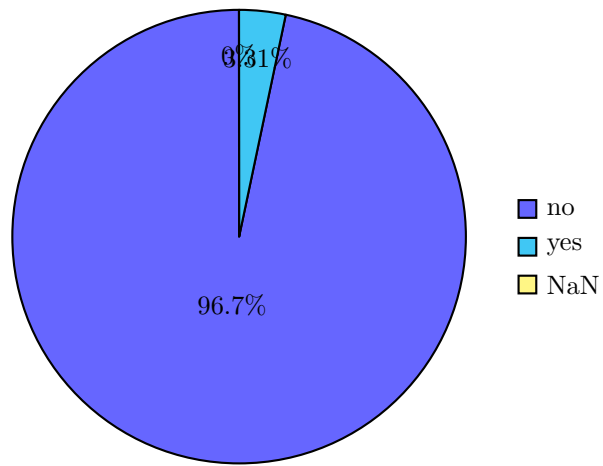
FAVC



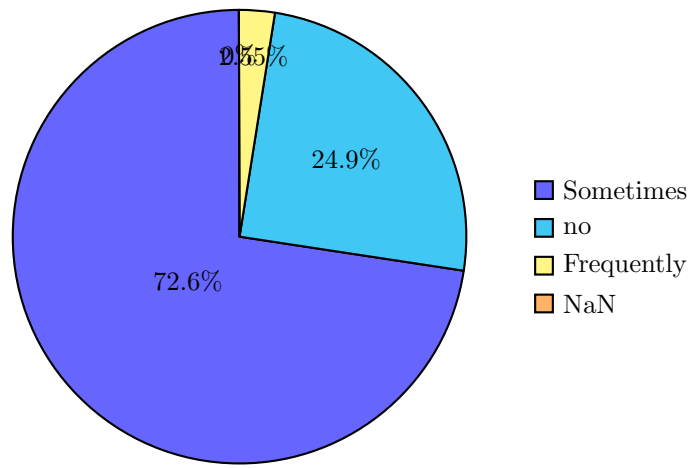
CAEC



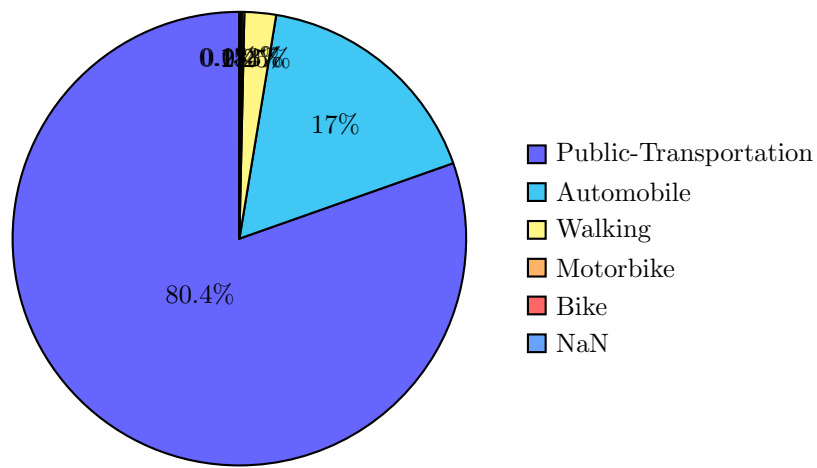
SMOKE



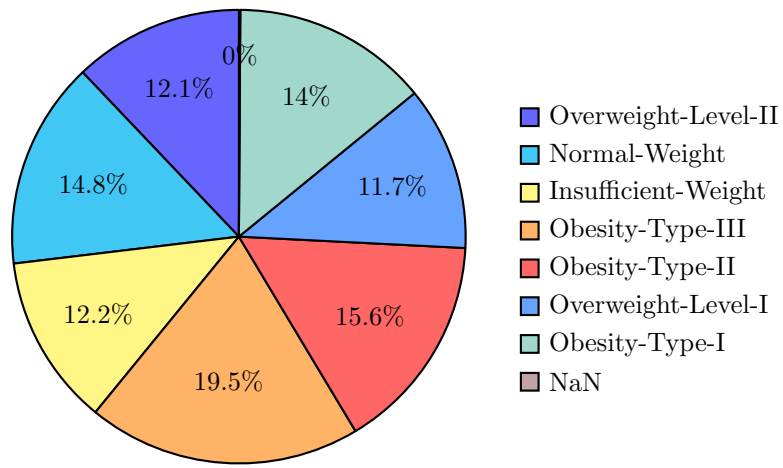
SCC



CALC



MTRANS



NObeyesdad