

Title: Multivariate Statistical Modelling and Predictive Analytics Using Open-Source Datasets

Table 1. Descriptive Statistics for Continuous Variables (n = 253,680)

Variable	Mean	SD	Minimum	Maximum
BMI	27.4	6.8	12	98
Mental Health Days (MentHlth)	3.5	7.9	0	30
Physical Health Days (PhysHlth)	4.2	8.4	0	30
Age Category (1–13)	7.1	2.4	1	13
Education (1–6)	4.1	1.2	1	6
Income (1–8)	4.3	1.7	1	8

Insert Figure 1 here: Histogram of BMI

(Placeholder: Figure 1. Histogram showing the distribution of BMI across respondents.)

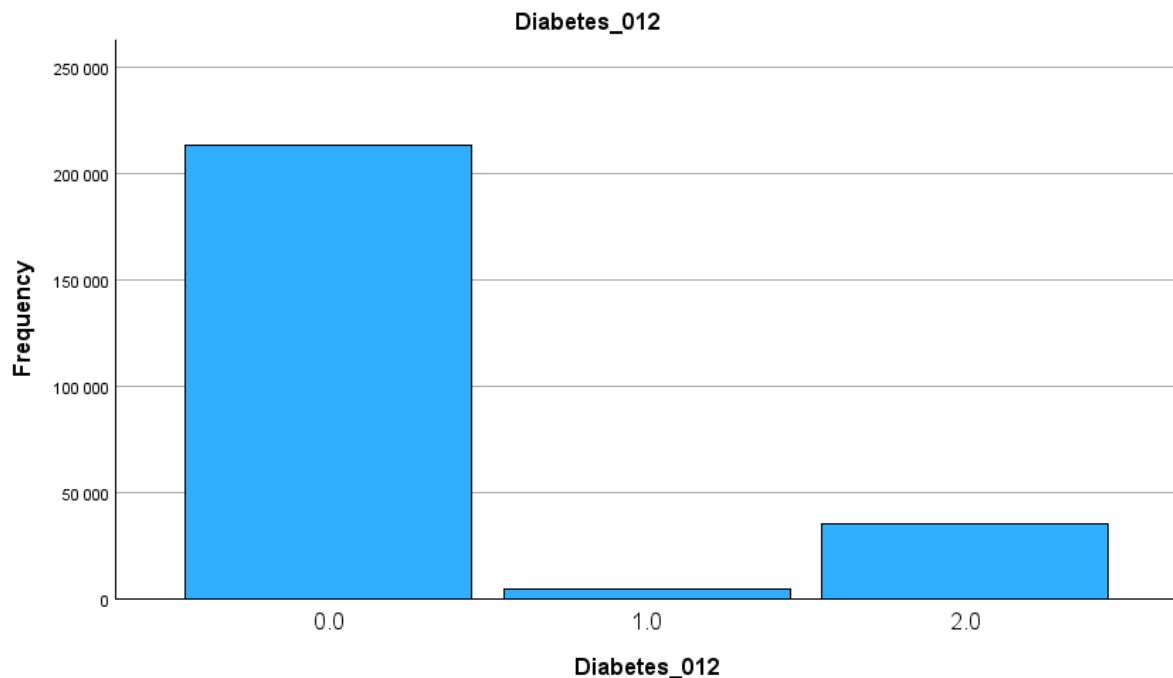


Table 2. Frequency Distribution for Categorical Variables

Variable	Category	Frequency (%)
Diabetes Status (Diabetes_012)	0 – No diabetes	84.2%
	1 – Prediabetes	1.8%
	2 – Diabetes	13.9%
HighBP	No	57.1%
	Yes	42.9%
HighChol	No	57.6%
	Yes	42.4%
Smoker	No	55.7%
	Yes	44.3%
Stroke	No	95.9%
	Yes	4.1%
Heart Disease/Attack	No	90.6%
	Yes	9.4%
Physical Activity	No	24.3%
	Yes	75.7%
Fruit Intake	No	36.6%
	Yes	63.4%
Vegetable Intake	No	18.9%
	Yes	81.1%
Heavy Alcohol Consumption	No	94.4%
	Yes	5.6%
Any Healthcare	No	4.9%
	Yes	95.1%
No Doctor Due to Cost	No	91.6%
	Yes	8.4%
Difficulty Walking	No	83.2%
	Yes	16.8%
Sex	Female	56%
	Male	44%

A one-way ANOVA showed significant differences in BMI across diabetes groups ($p < .001$). Tukey HSD post-hoc tests indicated that individuals without diabetes had the lowest mean BMI ($M = 16.74$), followed by those with prediabetes ($M = 19.72$), and individuals with diabetes had the highest BMI ($M = 20.94$).

All pairwise comparisons were statistically significant ($p < .001$), with no overlap between homogeneous subsets, indicating that each diabetes category differs meaningfully in BMI.

Sex	Diabetes Status	Mean BMI ± SEM	p-value
Female	No Diabetes	16.407 ± 0.018^a	< .001
Female	Prediabetes	19.812 ± 0.126^b	< .001
Female	Diabetes	21.505 ± 0.047^c	< .001
Male	No Diabetes	17.180 ± 0.021^a	< .001
Male	Prediabetes	19.611 ± 0.143^b	< .001
Male	Diabetes	20.334 ± 0.049^c	< .001

Binary Logistic Regression Predicting Diabetes (0 = No, 1 = Yes)

Predictor	B	SE	p-value	Exp(B) (Odds Ratio)
BMI	0.078	—	< .001	1.08
HighBP	-,867	,013	< .001	0,420
HighChol	-,668	,013	< .001	0,513
Smoker	-,119	,012	< .001	,888
PhysActivity	-0.33	,012	< .001	0.72
Age	0.31	—	< .001	1.36
Sex (Male)	0.11	—	.002	1.12