CAL STATE EAST BAY

INDE 330: Engineering Statistics and Probability

Analysis of Lifestyle habits on Abdominal Fat Obesity

PREPARED FOR

Dr. Farnaz Ganjeizadeh

Department of Engineering, California State University East Bay

PREPARED BY: Niharika Patil, Bhagyashri Desai

Contents

1 Introduction

As per recent research, abdominal fat is of more concern than subcutaneous fat. Clinical observations and basic research results agree that excessive fat inside the abdomen is a major contributor to cardiometabolic conditions. As a result of the rising epidemic of obesity, understanding body fat distribution and its clinical implications is critical to timely treatment. Visceral adipose tissue is a hormonally active component of total body fat, which possesses unique biochemical characteristics that influence several normal and pathological processes in the human body. In our analysis, we compared the lifestyle habits of two groups. One group consisted of individuals who had smoking and alcohol consumption habits whereas the other group didn't have these lifestyle habits.

2 Raw Data

This analysis is based on a data set obtained from Nutrition and Exercise consultancy firm; 'S Fitness Club' based in India. S Fitness Club provides customized diet and exercise plans to clients along with better lifestyle guidance. They provided us with the data of their 300 clients for our analysis.

3 Sampling Method

In our project, the sampling method we used was stratified sampling. Stratified sampling is a type of sampling method in which the total population is divided into smaller groups or strata to complete the sampling process. The strata are formed based on some common characteristics in the population data. Out of 300 clients 144 had complete information required for our analysis. In our data cleaning process, we eliminated those clients who had null values. The two subgroups we chose were the distribution of individuals having smoking and alcohol consumption habits (commonly used group name as SD in this report) compared to the distribution of individuals who didn't have those habits (commonly used group name as non-SD in this report). We have collectively taken a total of seventy-two people in each group.

4 Statistical Methods

For our project, we decided to use five statistical methods. The statistical methods are: Descriptive Statistics, Histogram, Boxplot, Normality Test and Hypothesis testing.

4.1 Descriptive Statistics

Descriptive Statistics is used to describe the basic features of the data in a study. They provide simple summaries about the sample and the measures. It is used to present quantitative descriptions in a manageable form. We have used Measure of Dispersion, which is a type of descriptive statistics, and it refers to the idea of variability within the data.

Here we have taken the population with smoking and alcohol consumption habits in Descriptive Statistics: SD and the population without smoking and alcohol consumption habits in Descriptive Statistics: Non-SD. We can see that the Mean, Standard Deviation, Variance, and all other attributes are higher for Descriptive Statistics: SD, that is, the population with smoking and alcohol consumption habits.

Descriptive Statistics: SD

Statistics

Variable	Mean	SE Mean	StDev	Variance	Minimum	Median	Maximum	Range	Mode
SD	40.255	0.547	4.642	21.552	32.000	40.000	50.000	18.000	39

Variable	N for Mode
SD	6

Figure 1: Descriptive Statistics SD

Descriptive Statistics: Non-SD

Statistics

Variable	Mean	SE Mean	StDev	Variance	Minimum	Median	Maximum	Range	Mode
Non-SD	37.941	0.369	3.131	9.802	31.500	38.000	45.000	13.500	36, 39, 40

Variable	N for Mode
Non-SD	8

Figure 2: Descriptive Statistics Non-SD

4.2 Good fit method

Statisticians say that a regression model fits the data well if the differences between the observations and the predicted values are small and unbiased. Unbiased in this context means that the fitted values are not systematically too high or too low anywhere in the observation space.

The goodness of fit of a statistical model describes how well it fits a set of observations. Measures of goodness of fit typically summarize the discrepancy between observed values and the values expected under the model in question. Such measures can be used in statistical hypothesis testing, e.g., to test for normality of residuals, to test whether two samples are drawn from identical distributions, or whether outcome frequencies follow a specified distribution.

Distribution Identification for Non-SD

Goodness of Fit Test

Distribution	AD	P	LRT P
Normal	0.382	0.390	
Box-Cox Transformation	0.382	0.390	
Lognormal	0.426	0.307	
3-Parameter Lognormal	0.396	*	0.492
Exponential	27.915	<0.003	
2-Parameter Exponential	8.061	<0.010	0.000
Weibull	0.627	0.098	
3-Parameter Weibull	0.367	0.389	0.015
Smallest Extreme Value	0.857	0.025	
Largest Extreme Value	0.938	0.017	
Gamma	0.417	>0.250	
3-Parameter Gamma	0.459	*	1.000
Logistic	0.524	0.141	
Loglogistic	0.548	0.114	
3-Parameter Loglogistic	0.582	*	0.679

Figure 3: Good Fit Method Non-SD

Distribution Identification for SD

Goodness of Fit Test

Distribution	AD	P	LRT P
Normal	0.407	0.341	
Box-Cox Transformation	0.407	0.341	
Lognormal	0.517	0.183	
3-Parameter Lognormal	0.422	*	0.142
Exponential	24.754	<0.003	
2-Parameter Exponential	11.919	<0.010	0.000
Weibull	0.478	0.236	
3-Parameter Weibull	0.367	0.368	0.140
Smallest Extreme Value	0.710	0.063	
Largest Extreme Value	0.971	0.014	
Gamma	0.476	0.244	
3-Parameter Gamma	0.441	*	0.437
Logistic	0.556	0.105	
Loglogistic	0.631	0.064	
3-Parameter Loglogistic	0.557	*	0.294

Figure 4: Good Fit method SD

4.3 Histogram:

The histogram is a visual representation of the frequency distribution. For each interval of the histogram, the area of the bar equals the relative frequency of the measurements in the interval. The relative frequency is an estimate of the probability that a measurement falls in the interval.

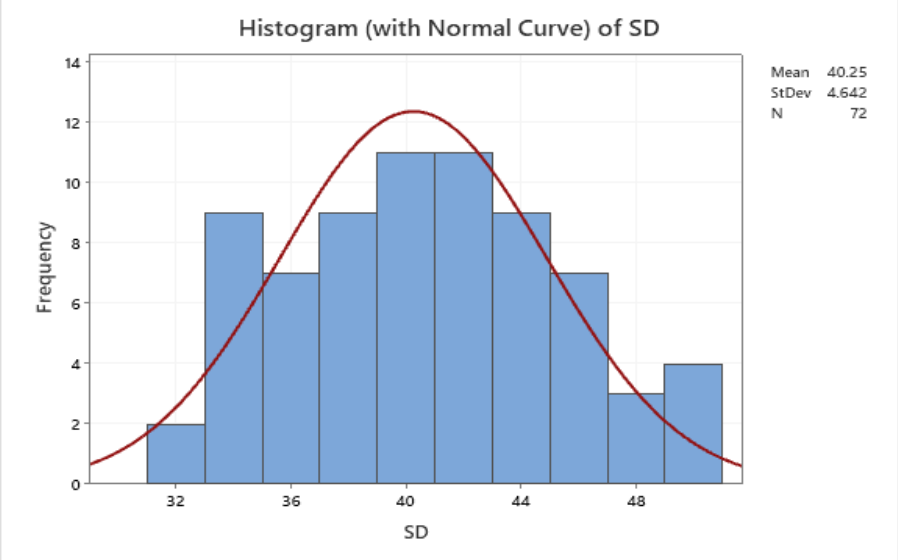

Figure 5: Histogram SD

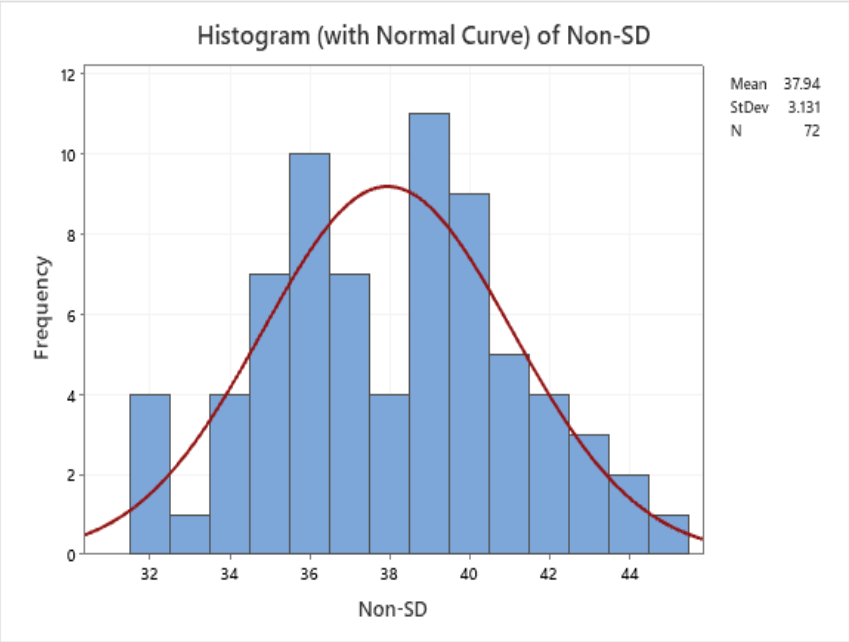

Figure 6: Histogram Non-SD

4.4 Box Plots:

The box plot is a graphical representation that describes several important features of a data set, such as center, spread, departure from symmetry, and identification of unusual observations or outliers. A box plot, sometimes called box-and-whisker plots, displays the three quartiles, the minimum, and the maximum of the data on a rectangular box, aligned either horizontally or vertically.

The box plots here indicate that the distribution of both the population with smoking and alcohol consumption habits, Boxplot of SD and without the habits, Boxplot of Non−SD is symmetrical around the central value because the left and right whiskers and the lengths of the left and right boxes around the median are about the same. For the Boxplot of SD, the upper whisker extends to observation 50 because it is the highest observation. This limit is 44+1.5(44 − 36) = 56. The lower whisker extends to observation 32 because it is the smallest observation. This limit is 36−1.5(44 − 36) = 24.

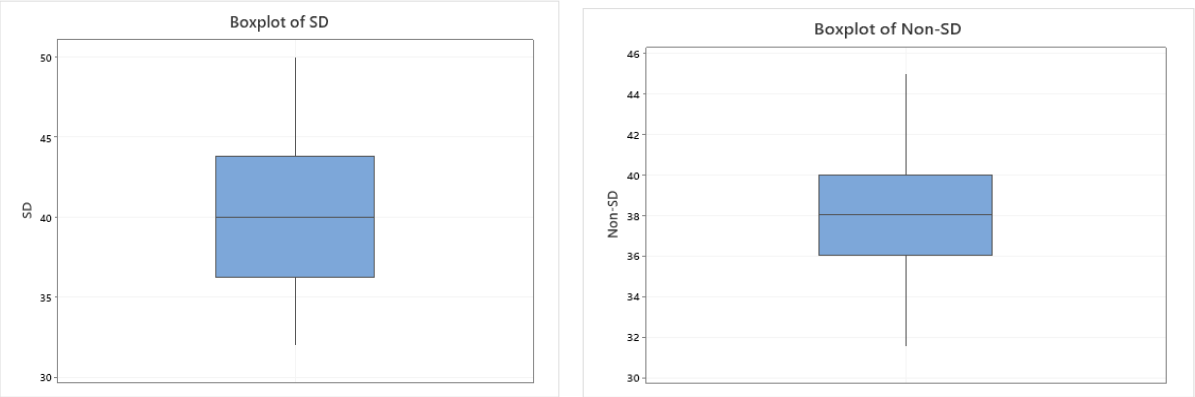

Figure 7: Box Plots

4.5 Normality Test:

Probability plot is a graphical method for determining whether sample data conform to a hypothesized distribution based on a subjective visual examination of the data. Probability plotting typically uses special axes that have been scaled for the hypothesized distribution.

For both the Probability Plot of population with smoking and alcohol consumption habits, Probability Plot of SD and population without the habits, Probability Plot for Non-SD, in assessing the closeness of the points to the straight line, all the points are covered by this line, so a normal distribution adequately describes the data for both the population. Here the Mean, Standard Deviation, AD and P-Value is higher for the Probability Plot for SD group in comparison to the Probability Plot for Non-SD.

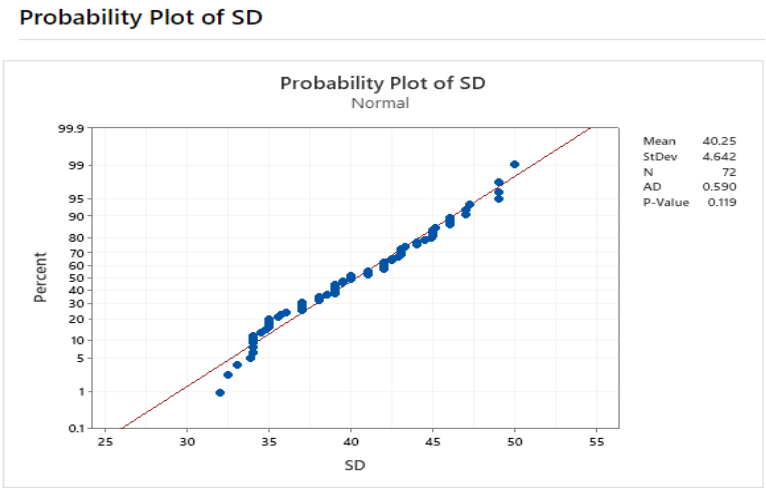

Figure 8: Normality Plot for SD

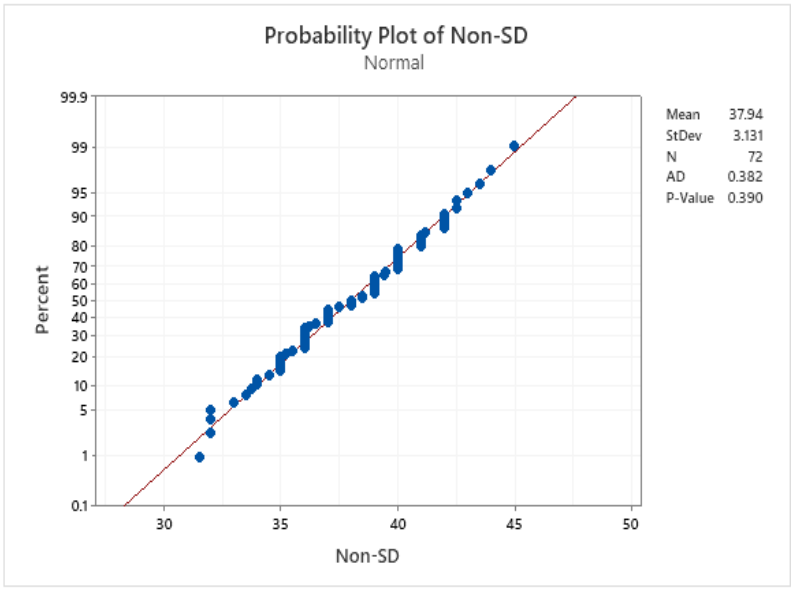

Figure 9: Normality Plot for Non-SD

4.6 Two-Sample T – Hypothesis Testing:

Two sample T test is used to compare the means of two groups. T statistic is used because the variances are unknown. Also in some situations, we cannot reasonably assume unknown variances to be equal. So exact T statistic is not available. However, an approximate result can be applied. In this Two-Sample T - Hypothesis Testing, μ_1 denotes the mean of the population with smoking and alcohol consumption habits, that is the mean of SD and μ_2 denotes the mean of the population without smoking and alcohol consumption habits, that is the mean of non-SD. Here

the Mean of SD is 40.25 and the Mean of Non-SD is 37.94, T-value is 3.51, degrees of freedom, DF is 124 and P-value is 0.001.

$$t = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

The seven-step procedure:

1. Parameters of Interest: mean of waist size for two groups say, μ1 and μ2, and we are Parameters interested in determining whether μ1 − μ2 = 0.
2. Null hypothesis: H0: μ1 − μ2 = 0
3. Alternative hypothesis: H1: μ1 > μ2
4. Test statistic: x1 = 40.25, x2= 37.94, s1 =4.642, s2= 3.131, n1=n2=72

$$t_0^* = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = 3.5$$

5. Reject H0 if: The degrees of freedom on t∗0 are as follows

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_1)^2}{n_2 - 1}} = 124.73$$

6. Therefore, using α=0.05 and a fixed-significance-level test, we would reject H0: μ1 = μ2 if t∗0 > t0.025,124 = 1.96
7. The t*0 = 3.5, which is greater than 1.96. So, we reject the null hypothesis.

Minitab Outputs:

Two-Sample T-Test and CI: SD, Non-SD

Method

μ_1: population mean of SD
μ_2: population mean of Non-SD
Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Estimation for Difference

Difference	95% Lower Bound for Difference
2.314	1.220

Test

Null hypothesis $H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis $H_1: \mu_1 - \mu_2 > 0$

T-Value	DF	P-Value
3.51	124	0.000

Figure 10: Minitab 2 Sample T test

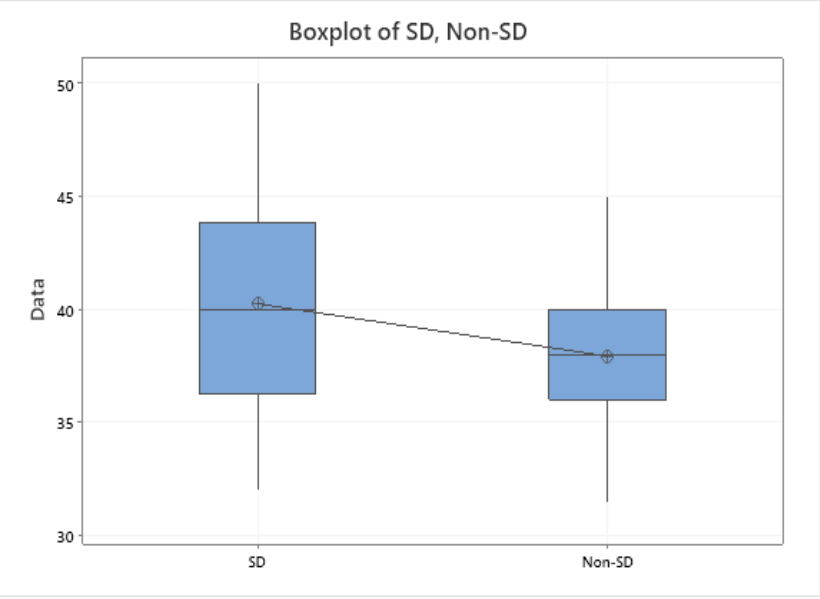

Figure 11: Box Plot Comparison

5 Results and Conclusions

The normality test confirmed that data falls under normal distribution. The 2 Sample T test showed that lifestyle habits smoking, and alcohol consumption has a greater impact on abdominal obesity. The mean value of the SD group was greater by 2.314 than the mean of the non-SD group.

6 Pros & Cons

Pros:

- Able to find the dataset comes under normality.
- With available data able to prove/ support the research paper. As a result of the rising epidemic of obesity, understanding body fat distribution and its clinical implications is critical to timely treatment.

Cons:

- More data is required for accurate estimates.

7 References

1. Montgomery, D.C., and Runger, G. C. (2018). *Applied Statistics and Probability for Engineers*. Wiley. 7th edition.
2. Harvard T H Chan School of Public Health. (n.d.). *Waist Size Matters* https://www.hsph.harvard.edu/obesity-prevention-source/obesity-definition/abdominal-obesity/

8 Appendices

The raw data used for our project in excel form is shown below in snapshot.

Client ID	Gender	Age	Height in m	Height in cm	Weight (in Kg)	BMI	Waist (Largest part of Stomach) in	Client Medical History	Occupation	Lifestyle habits	Physical Activity Levels LOW
1	Male	28	1.76	5.8	80.2	25.89	36	NA	Chartered accountant	Vegetarian, Vegetarian (but eats eggs	Medium
2	Male	21	1.62	5.3	75	28.58	39.5	None	Job	Smoking, Drinking, Vegetarian, Non-ve	High
3	Male	34	1.62	162	66.5	25.34	36	Nothing specific	Chartered Accountant	Non-vegetarian	Medium
4	Female	29	1.63	163	55	20.70	33	None	Service	Vegetarian	Low
5	Male	26	1.7	170	63	21.80	32	N/A	IT Service	Non-vegetarian	Medium
6	Male	26	1.85	185	83	24.25	34	No	BE civil	Non-vegetarian	Medium
7	Male	31	1.6	160	88	34.38	44.88	None	Job	Drinking, Non-vegetarian	Medium
8	Female	27	1.6	160	75	29.30	36	Undergone curretting procedure	Computer Engineer work for Accenture I	Vegetarian (but eats eggs)	Medium
9	Female	25	1.6	160	68	26.56	37	Nill	Architect	Vegetarian, Non-vegetarian	Medium
10	Female	14	1.71	5.6	83	28.38	42	No	Student	Vegetarian	Medium
11	Female	20	1.57	157	56	22.72	36	No medical history	Student	Vegetarian	Medium
12	Female	24	1.49	149	58	26.12	39	Pcos	Student	Vegetarian (but eats eggs)	Low
13	Male	26	1.8	180	105	32.41	41	No	BBA	Vegetarian (but eats eggs)	Medium
14	Male	28	1.64	164	84.3	31.34	43	No	Business	Smoking, Drinking, Vegetarian, Non-ve	Medium
15	Male	25	1.71	171	79	27.02	37	Previous sleep disc problem (lower bac	Student	Non-vegetarian	Low
16	Female	23	1.52	152	42	18.18	28	Acidity problem	Student	Vegetarian (but eats eggs)	Low
18	Male	21	2.1	210	79.5	18.03	37	None	Kirana merchant	None	Medium
19	Female	21	1.65	165	58	21.30	33.5	None	Student	Vegetarian, Vegetarian (but eats eggs	Low
20	Male	23	1.83	6	75	22.40	35.5	No	Business	Drinking, Non-vegetarian	Medium
22	Male	29	1.7	170	105	36.33	48	Nill	Business	Vegetarian, Vegetarian (but eats eggs	Medium
23	Female	21	1.62	162	83.5	31.82	43	Nothing	Student	Vegetarian	Medium
24	Male	26	1.56	156	54	22.19	30	None	Job	Vegetarian (but eats eggs)	Medium
25	Female	23	1.52	152	42	18.18	12	No	Student	Vegetarian (but eats eggs)	Low
26	Male	32	1.7	170	90	31.14	42	No	Busines	Vegetarian	Medium
27	Female	38	1.54	154	58	24.46	35	Slip disk(left shoulder musale tight) Taken Epidural injection for delivery in	Artist	Non-vegetarian	Medium
28	Male	25	1.65	165	82	30.12	41	Nothing	Chartered Accountant	Vegetarian, Non-vegetarian, Vegetaria	Low
29	Female	24	1.65	164.54	67.5	24.93	42	I've a PCOD problem	Work from home	Vegetarian	Medium

Client ID	Gender	Age	Height in m	Height in cm	Weight (in Kg)	BMI	Waist (Largest part of Stomach) in	Client Medical History	Occupation	Lifestyle habits	Physical Activity Levels LOW
31	Male	20	1.73	172.72	94.75	31.76	45.7	No probs	Student	Vegetarian	Medium
32	Male	25	1.74	174	73.7	24.34	38.2	NA	Job	Vegetarian	Low
33	Female	33	1.58	5.2	63	25.24	35.5	Hypothyroidism	House wife	Vegetarian	Medium
34	Male	26	1.76	176	86.5	27.92	39	No	Doctor	Non-vegetarian	Medium
35	Female	21	1.63	163	86	32.37	45	no	Bsc 3	Non-vegetarian	Medium
36	Male	24	1.66	166.37	85.7	30.96	40	None	Business	Vegetarian, Non-vegetarian, Monday f	Low
37	Male	38	1.8	180	70	21.60	35.8	Hamiya	Self employ	Non-vegetarian	Medium
38	Female	24	1.5	150	42	18.67	28	None	Dancer	Vegetarian	Medium
39	Female	37	1.47	147	59	27.30	34	No medical	Actor	Non-vegetarian	High
40	Female	24	1.58	158	76	30.44	44	No	No	Non-vegetarian	Medium
43	Female	30	1.59	159	63.5	25.12	39	Severe hair loss	Service	Non-vegetarian	Low
44	Female	29	1.55	155	60	24.97	38	Cesarean in April 2019	Housewife	Non-vegetarian	Low
45	Female	27	1.56	156	79.55	32.69	44	Had minor fracture in right leg and ligar	Service	Vegetarian (but eats eggs)	Low
46	Female	26	1.65	5.4	83.65	30.73	40	No	Service	Non-vegetarian	Low
48	Male	25	1.82	182	124	37.44	47	NA	Job	Smoking, Vegetarian (but eats eggs)	Medium
49	Female	29	1.63	163	68	25.59	37	NA	Job	Vegetarian (but eats eggs)	Medium
50	Female	22	1.53	153	58	24.78	36	PCOD from last 6 years	Job	Vegetarian, Non-vegetarian	Low
51	Female	27	1.68	167.64	72	25.62	32	167.64cm	journalist	Drinking, Vegetarian, Non-vegetarian	Medium
52	Female	27	1.5	150	64	28.44	33	Migraine and acidity	Freelancer	Vegetarian	Medium
53	Female	25	1.51	151	37	16.23	28	Migraine/sinus and minimal erosion in r	Engineer	Non-vegetarian	Low
54	Female	18	1.65	165	76	27.92	40	N/A	Student	Vegetarian	Medium
55	Male	25	1.83	183	107.4	32.07	48	NA	Student	Vegetarian	Medium
56	Male	23	1.75	175	87.35	28.52	40	Clear	Civil Engineer	Non-vegetarian	Medium
57	Female	26	1.55	155	79	32.88	44	Pcod	Job	Drinking, Vegetarian, Non-vegetarian,	Low
58	Male	30	1.76	176	80	25.83	39	No	Software Engineer	Smoking, Drinking, Non-vegetarian	Medium
59	Male	56	1.66	166	106	38.47	47	Blood pressure	Business	Nothing	Medium
60	Female	25	1.54	154	64.9	27.37	39	Allergic bronchitis	Doctor(Physiotherapist)	Drinking, Vegetarian (but eats eggs), S	Medium