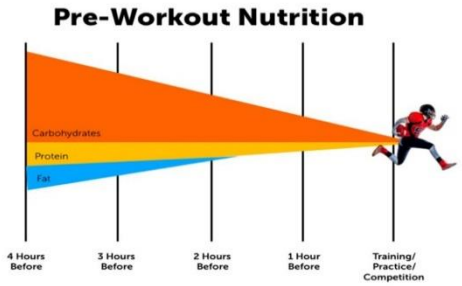
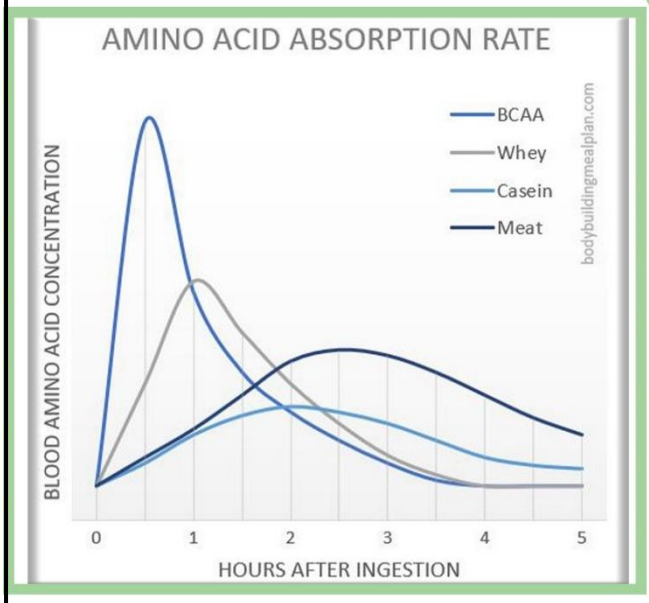
"STATISTICAL ANALYSIS ON EFFECTS OF BODY-BUILDING SUPPLEMENTS"





Gokhale Education Society's

H.P.T ARTS & R.Y.K SCIENCE COLLEGE NASHIK – 422005

DEPARTMENT OF STATISTICS

CERTIFICATE

This is to certify that the project work entitled

"STATISTICAL ANALYSIS ON EFFECTS OF BODY BUILDING SUPPLEMENTS"

is a bonafied work carried out by,

Roll no	Name of the student
05	Bhoi Nilesh Sawan
11	Gurule Mayur Sharad
17	Kotharkar Mahesh Vishweshwar
23	Pandit Sumeet Hareram
29	Sharma Vivek Mahesh
35	Sonawane Prasad Rajendra

With partial fulfilment for the statistics project of the Savitribai Phule Pune University during the year 2022-23. The project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the said degree.

Signature of the guide **Examiner** **Head of Department**

Mr. M. M. Kazi **Prof. Mrs. V. S. Joshi**

ACKNOWLEDGEMENT

A project usually falls short of its expectations unless guided by the right person at the right time. Success of a project is an outcome of sincere efforts, channelled in the right direction, efficient supervision and the most valuable professional guidance. This project would not have been completed without the direct and indirect help and guidance of such luminaries. They provided us with the necessary resources and atmosphere conductive for healthy learning and training. We would like to thank Savitribai Phule Pune University for giving us an opportunity to perform the project because of which could apply the theoretical knowledge in Statistics at an undergraduate level we express our gratitude to our Principal **Dr. Mrs. M.D. Deshpande** Ma'am H.P.T ARTS AND R.Y.K SCIENCE COLLEGE, NASHIK for allowing us to present this project. At the outset we would like to guidance that we have received from **Prof**. **Mrs. V. S. Joshi** Ma'am and our project guide **Mr. M. M. Kazi** sir. We would also like to thank the other teaching staff Mrs. J. D. Vetal, Ms.S.M.Pawar, Ms.V.Aher, Mr.V.Sonar and Non-teaching staff Mr.Patil and Mr.Karanjikar**.** Without their critical evaluation and suggestion at every stage of the project, this project could not have reached its present form. Faculty has critically evaluated each step in developing this project.

We would like to extend the special thanks to our respondents who gave us fruitful information about our topic Statistical analysis on gym using body building supplements and finally, the students of our college, friends and family for their support without which the project could not have been a successful one.

Heartfelt gratitude to all of you.

INDEX

ABSTRACT

The topic of research is "Statistical analysis on effects of bodybuilding supplements". This study helps to identify the aim of people for their fitness goal and their use of supplements for their muscle gain. The research work will propose a true idea that how supplement is useful for human body and when it should be taken in which quantity. The research designed used in this study is Descriptive Research Design. In this context the study also tries to throw light on people's usage of supplements and analysing the diet followed by the people for their muscle gain.

The dosage of supplements and different diet of people has seen a major transformation of their body (muscle gain, weight loss, and bodybuilding). This can be specifically attributed to increase in muscle size, reducing excess fat, maintaining diet, better hormonal changes which leads to healthy life style, increasing focus and maintaining muscle mind connection. Now-a-days supplements play an important role for fast muscle recovery, muscle gain and respective goals.

"Manohar Aich" father of Indian Body building. He was the 1st Indian to become Mr Universe when he won the 1952 National Body-Builders Association Universe championship, Earlier the body-building was done without steroids or even if by limited use of steroids. But now-a-day each and every Body-building beginner wants to take supplements to gain muscles faster. So, our project is on analysing for how much and when to take supplements.

INTRODUCTION

Now-a-days in the ongoing world, fitness is necessary for all of us. People now a days prefer gym more than walking on roads or running in the parks and performing yoga. We all want to join gym for the better muscle growth, to stay fit, healthy and look more muscular. We people also prefer gyms for various reasons like body building, fitness, weight loss, weight gain, cardio, strength training, and many more. In today's world no one is satisfied with their natural physic obtained. So, for the betterment in their physic, people take many of the supplements like whey protein, multivitamins, pre-workout (**BCAA, EAA, GLUTAMINE, COFFEE POWDER, CREATINE** etc.) many of them have better effect of this supplements (good muscularity, better muscle recovery, increase in hormone, etc.) but also some of them have negative effects (digestion problem, kidney issues, pimples or acne, hair fall, etc.) on their body. It also depends upon the genetics of the people. The one hazardous thing is STEROIDS which big body builder's take for excess muscle growth. This project is very interesting for us, for performing and analysing too because it a related to each and every people who prefer going gyms and stay perfectly fit and muscular. We designed a survey to learn about the where people are interested to take their physics. The main purpose is also to study on the diet which people like to consume for their respective fitness goals.

Statistics is an essential applied science. It is the science which has been developed to help in solving problems faced in several fields in our day-to-day life. It is helps us to make decision wise, reliable and logical. Statistics may be defined as the collections, presentation, analysis and interpretations of numerical data. With the help of statistics, we can easily understand and visualize the things. We have chosen this topic for the project because now a day's every one of us wants to go to gym for fitness, bodybuilding, muscle growth, etc. we also are using this as a project to see that age depends on the fitness or not. Or which diet is efficient and better to achieve our fitness goals.

Here, we are checking which time is mostly preferred and is effective for the people. Also, to test that Supplement/Steroids is necessary or not.

Also, we have collected information and observed the correlation on the monthly earning and expenditure of the people on their dietary supplements. And also observe that what are the pros and cons of the supplements taken by the respective people.

OBJECTIVES

Here, we want to check the following objectives:

1) To analyse that how supplement effects in increasing or decreasing of body weight.

2) To analyse which diet, do the people prefer for their muscle building / muscle gaining.

3) To study the side effects of supplements on the body. (Pimple, Hair fall, Digestion issues, Overweight, etc.)

4) To analyse that which type of profession of people are more interested for joining gyms.

5) To analyse the monthly expenditure on supplements.

6) To study the skewness of monthly income.

7)To analyse which type of whey protein people have preferred. (Isolate, Concentrate, Hydrolysate)

8) To analyse that how much time people usually spends in gym.

9) To analyse that if a person buys whey protein is most likely to buy which supplements.

10) To observe that which time is mostly preferred by the people for going to gym and how much time is required for people to do their work out.

11) To analyse which supplements (whey protein, creatine, fish oil, mass gainers, etc.) are taken by the people from their trainer.

DATA COLLECTION

Collection of data is the most important and very first step of statistical methods. Almost care must be exercised while collecting the data, because constitutes the foundations on which the statistical analysis is built.

We have collected the data of several gyms of Nashik city by questionnaire method. In all we have collected the data of 200 gym users from different gyms of different areas.

- ***PRIMARY DATA:***

  Any data that an investigator collects himself are termed as "Primary data". Since the primary data is original in character, they are reliable than any order.

- ***SECONDARY DATA:***

  Data taken from figures collected by others are termed as "Secondary data". The user of secondary data cannot have a thorough understanding of the background as the user of primary data.

In our project we preferred to use only a primary data. Primary data means first-hand information collected by an investigator. It is collected for the first time. It is original and more reliable. For example, the population census conducted by the government of India after every ten years is primary data.

METHOD OF COLLECTION OF PRIMARY DATA

We now consider various methods of collecting primary data, with their merits, demerits and situations in which they are.

- ***DIRECT PERSONAL INTERVIEW:***

  A face-to-face contact is made with the informants (persons from whom the information is to be obtained) under this method of collecting data. The interviewer asks them questions pertaining to the survey and collects the desired information. The information collected in this manner is first hand and also original in character.

VARIABLES AND ATTRIBUTES USED UNDER STUDY

- ***VARIABLES:***
  1) Time gives in gym (in hrs.)
  2) Monthly Income (in Rs.)
  3) Monthly expenditure on supplement (in Rs.)
  4) Age of a gym users.
  5) Weight of a gym users (in kg.)

- ***ATTRIBUTES:***
  1) Gender
  2) Profession of gym users
  3) Types of protein

QUESTIONNAIRE

1) Email:
   ---
2) Age:
   a) 15-30
   b) 30-45
   c) 45-60
   d) Above 60
3) Gender:
   a) Male
   b) Female
   c) Others
4) What was your Weight before joining the gym?
   ---
5) What is your weight after joining the gym after using supplements?
   ---
6) Which type of profession you are belonging?
   a) Sports
   b) Modelling
   c) Business
   d) Official jobs
   e) Students
   f) Other
7) What is your monthly income?
   a) Below 15000
   b) 15000-30000
   c) 30000-45000
   d) 45000-60000
   e) Above-60000
8) Which time you prefer to go to gym?
   a) 5-8 am
   b) 8-11am
   c) 5-8 pm
   d) 8-11pm
9) How much time you give in gym?
   a) 0-1(Hours)
   b) 1-2(Hours)
   c) 2-3(Hours)
   d) 3-4(Hours)
10) Did you had any type of disease before joining the gym?
   a) Yes
   b) No

11) Why you prefer to go gym?
    a) Weight gain
    b) Body building
    c) Power lifting
    d) Fitness
    e) Weight loss
12) Which diet is good for your muscle gain?
    a) Vegetarian
    b) Non vegetarian
    c) Supplement
13) Which type of supplement you use for your body?
    a) Creatine
    b) Protein supplement
    c) Weight gainers
    d) Caffeine
    e) Fish oil
    f) Natural diet
    g) Others
14) What is your monthly expenditure on supplement?
    a) 0-3000
    b) 3000-6000
    c) 6000-9000
    d) 9000-12000
15) Which type of whey protein do you take?
    a) Isolate
    b) Concentrate
    c) Hydrolysate
16) Which type of protein has good number of contents?
    a) Muscle baize
    b) Big muscle
    c) My protein
    d) Morphe nutrition
    e) Other
17) Why do we take whey protein?
    a) Helps to lose weight
    b) Build muscle with proper recovery of tissue
    c) To increase/improve digestion
    d) All the above
    e) Other
18) Which is the important content in whey protein for the body?
    a) Soya product
    b) Caffeine
    c) BCAA
    d) Glutamine
    e) Others

19) Who preferred / suggested you the whey Protein?
    a) Gym trainer
    b) Social media
    c) Friends
    d) Online(google)
    e) Other
20) Do you have kidney problems?
    a) Yes
    b) No
21) What is the right way of taking pre-workout?
    a) During workout
    b) Before workout
    c) After workout
    d) Any time in a day
    e) Other
22) Which type of pre-workout should be taken for Beginners?
    a) BCAA
    b) EAA
    c) Glutamine
    d) Coffee (Black)
23) What is the effect of Glutamine on your muscles?
    a) Strength training recovery
    b) Muscles recovery
    c) Muscle gain
    d) Fat loss
    e) Other
24) What type of effect have you seen after taking supplements?
    a) Hair fall
    b) Pimples
    c) Over weight gain
    d) Digestion issues
    e) Others
25) Which is good /beneficial?
    a) Natural diet
    b) Whey protein
    c) Both
26) Are steroids essential for bodybuilding?
    a) Yes
    b) No
27) Which gives the better effect to the body without any harm?
    a) Steroids
    b) Whey Protein
    c) Daily diet natural + whey protein
    d) Only pre-workout and whey protein
    e) Other

28) How much do you think gym helps you to keep your body fit?
    a) Not at all
    b) Moderate
    c) Good
    d) Very good
29) Do you smoke cigarettes or drink alcohol?
    a) Yes
    b) No

Q-1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
snehashukla2628@gmail.com	0	0	0	1	1	0	1	0	0	3	0	6	0	0	3	1	2	1	0	0	0	0	1	0	0	1	2	0
bhoyarjanvi25@gmail.com	0	0	0	1	1	0	2	1	0	3	1	6	0	0	5	5	0	4	0	1	3	9	0	2	0	2	3	0
ismitayadav0123@gmail.com	0	0	0	0	1	0	3	0	0	1	1	4	0	1	2	8	0	0	1	2	3	0	3	0	1	2	3	0
kamyanishelar@gmail.com	0	0	0	1	1	0	2	1	0	3	6	4	0	2	0	8	2	3	0	2	3	2	1	2	0	2	1	0
poonaprajapati2803@gmail.com	0	0	0	1	1	0	2	1	0	3	6	1	0	0	6	4	5	1	0	1	6	5	3	2	0	2	2	0
shubhangij32@gmail.com	0	0	0	1	2	0	0	1	0	1	0	1	0	0	3	8	3	0	0	1	3	2	1	1	1	1	3	0
chaitralipatil9561@gmail.com	0	0	1	1	2	3	3	1	0	1	1	4	0	0	2	1	0	2	0	1	3	2	4	2	0	2	3	0
madhurakadu533@gmail.com	0	0	1	1	1	0	3	0	0	3	3	6	0	2	5	5	4	0	0	0	6	7	1	0	0	2	2	0
snehashukla2613@gmail.com	0	0	1	1	1	0	0	0	0	1	0	8	0	1	2	8	2	0	1	1	10	5	0	2	0	2	3	0
ketkigorhe2000@gmail.com	0	0	1	2	2	1	0	1	0	1	0	4	0	0	2	8	9	0	0	1	0	4	4	2	0	2	3	0
rishu.v0220@gmail.com	0	0	1	2	4	2	0	1	0	1	0	0	0	0	5	8	11	2	0	1	1	7	12	1	0	2	3	1

sharmapreeti@gmail.com	0	0	1	1	1	0	2	1	0	1	0	4	0	2	1	7	9	3	0	1	3	1	4	0	0	2	2	0
prasadgunjan67@gmail.com	0	0	1	2	0	0	0	1	0	3	1	4	0	0	2	2	0	0	0	2	3	3	2	0	1	2	2	0
nupurabhardwaj@gmail.com	0	0	1	0	0	0	0	0	0	1	1	4	0	2	1	1	0	0	0	1	3	0	1	2	0	2	2	0
nancyjangra20034@gmail.com	0	0	2	2	1	0	0	1	0	1	0	4	0	0	2	1	5	1	0	1	0	3	3	2	0	2	3	0
pritisharma20062000@gmail.com	0	0	2	2	1	0	0	1	0	3	0	1	0	2	2	0	0	0	0	0	3	0	1	0	0	2	2	0
suryawanshiishwari4@gmail.com	0	0	2	2	1	0	0	1	0	1	0	4	0	2	2	1	0	2	0	1	3	2	0	2	0	2	2	0
raviyadav633026@gmail.com	0	0	2	2	1	0	2	0	0	3	2	1	0	2	1	1	3	1	0	1	2	1	1	2	0	2	3	0
jragini438@gmail.com	0	0	2	2	1	0	0	1	0	4	3	4	0	2	2	8	0	2	0	1	3	9	3	0	1	2	2	0
preetipandit@gmail.com	0	0	2	2	1	0	0	2	0	1	2	2	0	2	9	1	2	2	0	1	3	0	4	0	1	4	2	0
pratima4717@gmail.com	0	0	3	2	1	0	2	1	0	4	0	4	0	2	0	4	0	3	0	2	0	3	3	0	0	2	1	0
suchitashukla08@gmail.com	0	0	3	2	3	0	2	1	0	4	0	1	0	0	2	4	0	2	0	1	0	5	9	2	0	2	2	0
shreyaapatel26@gmail.com	0	0	3	3	2	1	0	0	0	4	0	4	0	0	4	3	1	0	0	2	3	1	1	2	1	2	1	0

Notations

Gender	Coding
Male	1
Female	0

Age	Coding
15-30	0
30-45	1

Weight	Coding
30-40	0
40-50	1
50-60	2
60-70	3
70-80	4
80-90	5
90-100	6
100-110	7
110-120	8
120-130	9
130-140	10
140-150	11

Profession	Coding
Sports	0
Student	1
Business	2
Official Job	3
Modelling	4
Others	5

Income	Coding
Below 15000	0
15000-30000	1
30000-45000	2
45000-60000	3
Above 60000	4

Time	Coding
5-8 AM	0
8-11 AM	1
5-8 PM	2
8-11 PM	3

How much time you do workout?	Coding
0 to 1 (hours)	0
1 to 2 (hours)	1
2 to 3 (hours)	2
3 to 4 (hours)	3

Did you had any type of disease before joining the gym?	Coding
Yes	1
No	0

Why you prefer to go to gym?	Coding
Weight gain	0
Body building	1
Power lifting	2
Fitness	3
Weight loss	4

Which diet is good for your muscle gain?	Coding
Vegetarian	0
Non-vegetarian	1
Supplement	2
Vegetarian & Non-vegetarian	3
Vegetarian & Supplement	4
Non-vegetarian & Supplement	5
All	6

Which type of Supplement you use for your body?	Coding
Creatine	0
Protein Supplement	1
Weight gainers	2
Caffeine	3
Natural diet	4
Creatine & Caffeine	5
Others	6

No	7
All of them	8
Fish oil	9

What is your monthly expenditure on supplement?	Coding
0-3000	0
3000-5000	1
6000-9000	2
9000-12000	3

Which type of whey protein do you take?	Coding
Concentrate	0
Hydrolysate	1
Isolate	2

Which type of protein has good number of contents?	Coding
Muscle Blaze	0
Big muscle	1
My Protein	2
Morphe nutrition	3
Muscle Blaze & Big muscle	4
Muscle Blaze & My Protein	5
Muscle Blaze &	6
Muscle Blaze & Morphe nutrition	7
Big muscle & My Protein	8
My Protein & Morphe nutrition	9
All	10
Others	11

Why do we take whey protein?	Coding
Helps to lose weight	0
Build muscle with proper recovery of tissue	1
To increase/improve digestion	2
Helps to lose weight & Build muscle with proper recovery of tissue	3
Helps to lose weight &To increase/improve digestion	4
Build muscle with proper recovery of tissue & to increase/improve digestion	5
All	6
Others	7

Which are the important content in whey protein for the body?	Coding
Soya product	0
Caffeine	1
BCAA	2
Glutamine	3
Soya product & Caffeine	4
Soya product & BCAA	5
Soya product & Glutamine	6
Caffeine & BCAA	7
Caffeine & Glutamine	8
BCAA & Glutamine	9
Others	10
All	11

Who preferred/suggested you Whey Protein?	Coding
Gym trainer	0
Social Media	1
Friends	2
Online (Google)	3
Other	4

Do you have kidney problems?	Coding
Yes	1
No	0

Which type of pre-workout should be taken for Beginners?	Coding
BCAA	0
EAA	1
Glutamine	2
Coffee (Black)	3
BCAA & EAA	4
BCAA & Glutamine	5
BCAA & Coffee (Black)	6
EAA & Glutamine	7
EAA & Coffee (Black)	8
All	9
Others	10

What is the effect of Glutamine on your muscles?	Coding
Strength training recovery	0
Muscles recovery	1
Muscle gain	2
Fat loss	3
Strength training recovery & Muscles recovery	4
Strength training recovery & Muscle gain	5
Strength training recovery & Fat loss	6
Muscles recovery & Muscle gain	7
Muscles recovery & Fat loss	8
All	9
Others	10

What type of effect have you seen after taking supplements?	Coding
Hair fall	0
Pimples	1
Over weight gain	2
Digestion issues	3
Nothing	4
Hair fall & Pimples	5
Hair fall &Over weight gain	6
Hair fall & Digestion issues	7
Pimples & Over weight gain	8
Pimples & Digestion issues	9
Over weight gain & Digestion issues	10
Others	11

Which is good/beneficial?	Coding
Natural diet	0
Whey Protein	1
Both	2

Is steroids essential for bodybuilding?	Coding
Yes	1
No	0

Which gives the better effect to the body without any harm?	Coding
Steroids	0
Whey Protein (only)	1
Daily diet natural + whey protein	2
Only pre-workout and whey protein	3
Others	4

How much do you think gym helps you to keep your body fit?	Coding
Not at all	0
Moderate	1
Good	2
Very good	3

Do you smoke cigarettes or drink alcohol?	Coding
Yes	1
No	0

Do you like our survey?	Coding
Yes	1
No	0

STATISTICAL TOOLS USED

1) Simple Bar diagram.
2) Histogram.
3) Frequency Polygon.
4) Sub divided bar diagram.
5) Pie Chart.
6) Box Plot.
7) Process Control Tools-
    a) Check sheet.
    b) Pareto Diagram.
8) Normality Test.
9) Bartlett Test.
10) Sign Paired Test.
11) Proportion Test.
12) Chi-square- Independence of attributes.
13) Multiple Linear Regression.
14) Data Analytics-
    a) K-Nearest Neighbor.
    b) Naive Bayes.
    c) Decision Tree.
    d) Association Rules.

STATISTICAL SOFTWARE USED

1) MS Excel.
2) Advanced MS Excel.
3) R-Studio.
4) Python.

THEORY OF STATISTICAL TOOLS USED

- ***GRAPHICAL REPRESENTATION:***

Graphical Representation is a visual display of data and statistical results. It is often more effective than presenting the data in tabular form. There are many different types of graphical representations which is used depending upon the nature of data and type of the statistical results. It is very effective way to serve the purpose of comparison at a glance and revealing the patterns in the data. Graphs and diagrams are easy to understand and create an effect. Graphs and charts are often used to ease understanding of large quantities of data and relationships between parts of the data. Graphs can usually be read more quickly than the raw data that they are produced form. They are used in wide variety of fields and can be created by hands often on graph papers or by Computer using a chart application. Therefore, Graphs and Charts are believed to be powerful tools to convey information. The different types of graphical representation used in this project are:

- ***Pie-Chart:***

A pie chart, sometimes called a circle chart, is a way of summarizing a set of nominal data or displaying the different values of a given variable (e.g., percentage distribution). This type of chart is a circle divided into a series of segments. Each segment represents a particular category. The area of each segment is the same proportion of a circle as the category is of the total data set.

- ***Simple Bar Diagram:***

A simple bar chart is used to represent data involving only one variable classified on a spatial, quantitative or temporal basis. In a simple bar chart, we make bars of equal width but variable length, i.e., the magnitude of a quantity is represented by the height or length of the bars.

- ***Subdivided Bar Diagram:***

Sub-divided bar diagrams are those diagrams which simultaneously present, total values as well as part values of a set of data. Different parts of a bar must be shown in the same order for all bars of a diagram.

- ***Histogram:***

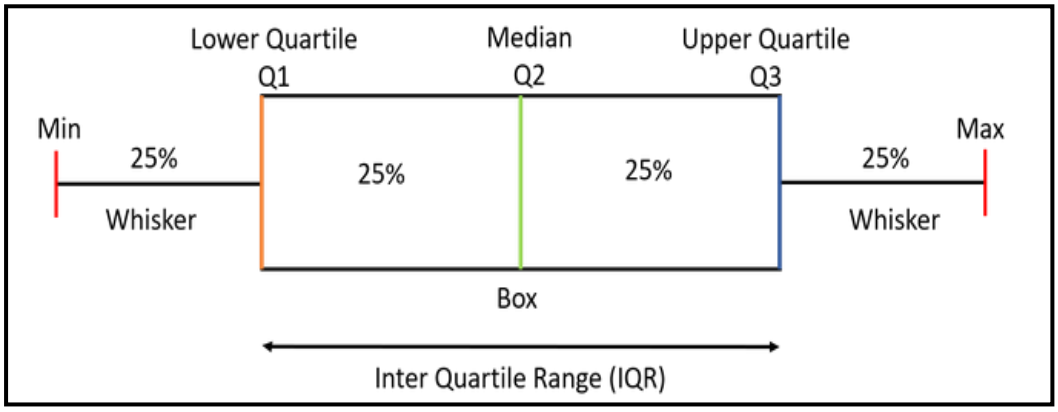
A histogram is a display of statistical information that uses rectangles to show the frequency of data items in successive numerical intervals of equal size.

- ***Frequency polygon:***

Frequency polygons are a graphical device for understanding the shapes of distributions. They serve the same purpose as histograms, but are especially helpful for comparing sets of data. Frequency polygons are also a good choice for displaying cumulative frequency distributions.

- ***Boxplot:***

A box and whisker plot also called a box plot displays the five-number summary of a set of data. The five-number summary is the minimum, first quartile, median, third quartile, and maximum. In a box plot, we draw a box from the first quartile to the third quartile. A vertical line goes through the box at the median. The whiskers go from each quartile to the minimum or maximum.



- ***Correlation Coefficient:***

A correlation coefficient is a statistical measure of the degree to which changes to the value of one variable predict change to the value of another. In positively correlated variables, the value increases or decreases in tandem.

Karl Pearson's Correlation coefficient is given by,

$$\mathbf{r}=\frac{\mathbf{cov(x,y)}}{\boldsymbol{\sigma x \sigma y}}$$

Where, r is such that $-1 \leq r \leq 1$.

The positive and negative signs are used for positive correlation and negative correlation respectively.

A prefect correlation of ± 1 occurs only when the data points all lie exactly on a straight line. Correlation coefficient number is greater than 0.8 is generally described as a strong, where as a correlation coefficient number is less than 0.5 is generally described as a weak.

- ***Testing of Hypothesis:***

Testing of hypothesis is a form of statistical inference that uses data from a sample to draw conclusions about a population parameter or a population probability distribution First, a tentative assumption is made about the parameter or distribution. This assumption is called the null hypothesis and is denoted by H_0. An alternative hypothesis (denoted H_1), which is the opposite of what is stated in the null hypothesis, is then defined. The hypothesis-testing procedure involves using sample data to determine whether or not H_0 can be rejected. If H_0 is rejected, the statistical conclusion is that the alternative hypothesis H_1 is true.

- ***Proportion Test-Testing of equality of two proportions ($P_1 = P_2$)***

Suppose we draw two samples. Suppose these samples give proportions of specific items as P_1 and P_2 respectively. One may be interested in knowing that the population proportions from which these samples are chosen are same. In other words, we want to whether difference between two sample proportions is negligible and it has arisen merely due to sampling variations.
Let,
P_1=proportion of specific items in first population.
P_2=proportion of specific items in second population.
n_1=size of sample drawn from first sample.
n_2=size of sample drawn from second sample.
X_1-Number of items of specific type in first sample.
X_2-Number of items of specific type in second sample.
P_1=X_1/n_1=proportion of specific items in first sample.
P_2=X_2/n_2= proportion of specific items in second sample.

The hypothesis for such problems will be:
Null Hypothesis, H_0: $P_1=P_2$
V/s Alternative Hypothesis,
$H_1: P_1 \neq P_2$
$H_1: P_1 < P_2$
$H_1: P_1 > P_2$
R commands for null hypothesis H_0: $P_1=P_2$
(a) Consider the alternative hypothesis: H_1: $P_1=P_2$
prop.test (x,n,conf.level=c)

(b) Consider the alternative hypothesis H_1:$P_1>P_2$.
prop.test (x,n,conf.level=c alternative="greater")

(c) Consider the alternative hypothesis $H_1:P_1<P_2$.
prop.test(x,n,conf.level=c alternative="less")

Criteria: Here level of significance α% is less than p-value then we may accept H_0.

- ***Chi-square test for independence of two attributes:***

Suppose that the given data is classified into r-levels of attributes A denoted as $A_1,\ldots., A_r$ and s levels of attribute B represented by $B_1,\ldots.,B_s$.
Then different class frequencies can be represented in the following tabular form:

B / A	B_1	B_2	…	B_i	…	B_s	Total
A_1	O_{11}	O_{12}	…	O_{1i}	…	O_{1s}	(A_1)
A_2	O_{21}	O_{22}	…	O_{2j}	…	O_{2s}	(A_2)
…	…	…	…	…	…	…	…
A_i	O_{i1}	O_{i2}	…	O_{ij}	…	O_{is}	(A_i)
…	…	…	…	…	…	…	…
A_r	O_{r1}	O_{r2}	…	O_{rj}	…	O_{rj}	(A_r)
Total	(B_2)	(B_2)	…	(B_j)	…	(B_s)	N

This table is known as (r x s) contingency table.
$N=\Sigma\Sigma O_{ij}$= Total observed frequency
$A_i=\Sigma O_{ij}$= Total of observed frequencies in *ith* row; i=1, 2, …, r
$B_j=\Sigma O_{ij}$= Total of observed frequencies in *jth* row; j=1, 2, …, s

Here,
Hypothesis under consideration is,
H_0: Two attributes A and B are independent
v/s
H_1: Two attributes A and B are not independent.

$E_{ij}=(A_i)(B_j)/N$; i=1, 2, …., r; j=1, 2, …., s
The test statistics under Ho is,

$$\chi^2 =\Sigma\Sigma\ (O_{ij}-e_{ij})^2/e_{ij} \quad =\Sigma\Sigma\ (O^2_{ij}/e_{ij})-N$$

Criteria: (1) Reject H_0 at α% l.o.s if $\chi^2_{(r-1)\,(s-1)} \geq \chi^2_{(r-1)\,(s-1),\,\alpha}$ accept otherwise.

(2) Reject H_0 at α% l.o.s if p- value less than l.o.s, otherwise accept it.

- ***Bartlett-Test***

We consider here Bartlett-test for equality of two variances.
In this test we test the null hypothesis,
H_0: $\sigma_{12} = \sigma_{22}$, against one of the alternative hypothesis v/s **H_1: $\sigma_{12} \neq \sigma_{22}$**.

a) Consider, **H_1: $\sigma_{12} \neq \sigma_{22}$** then the R command will be,
bartlett.test(x)

- ***Normality Test:***

In statistics, normality tests are used to determine if a data set is well-modelled by a ***Normal distribution*** and to compute how likely it is for a random variable underlying the data set to be normally distributed. A number of statistical tests, such as the student's t-test and the one-way and two-way ANOVA, require a normally distributed sample population.

To test the hypothesis,

H_0: The given sample is from normal distribution. V/s

H_1: The given sample is not from normal distribution.

We can check the normality of data using R-studio with ***Shapiro test.***

Criteria: If l.o.s < p-value, then we accept H_0 at α% los.

- ***Paired T-Test:***

The paired t-test gives a hypothesis examination of the difference between population means for a set of random samples whose variations are almost normally distributed. Subjects are often tested in a before-after situation or with subjects as alike as possible.

To test the hypothesis**, H_0: $\mu = \mu_0$** against one of the alternative hypothesis

H_1: $\mu \neq \mu_0$.

H_1: $\mu < \mu_0$.

H_1: $\mu > \mu_0$.

R supports the command t.test(mu=μ,(conf.level=c)

- ***Linear Regressions:***

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

There are two types of regressions,

A) ***Simple Linear Regression:*** Simple linear regression is a regression model that estimates the relationship between one independent variable and one dependent variable using a straight line. Both variables should be quantitative.

B) ***Multiple Linear Regressions:*** Multiple linear regression is a regression model that estimates the relationship between a quantitative dependent variable and two or more independent variables using a straight line.

- ***Process Control Tools:***

A) ***Check sheet:*** Check sheet is one of the seven process control tools, A check sheet is a structured, prepared form for collecting and analysing data. This is a generic data collection and analysis tool that can be adapted for a wide variety of purposes and is considered one of the seven basic quality tools.

B) ***Pareto Diagram:*** A Pareto diagram is a simple bar chart that ranks related measures in decreasing order of occurrence. The principle was developed by Vilfredo Pareto, an Italian economist and sociologist who conducted a study in Europe in the early 1900s on wealth and poverty. He found that wealth was concentrated in the hands of the few and poverty in the hands of the many. The principle is based on the unequal distribution of things in the universe. It is the law of the "significant few versus the trivial many." The significant few things will generally make up 80% of the whole, while the trivial many will make up about 20%. The purpose of a Pareto diagram is to separate the significant aspects of a problem from the trivial ones. By graphically separating the aspects of a problem, a team will know where to direct its improvement efforts. Reducing the largest bars identified in the diagram will do more for overall improvement than reducing the smaller ones.

- ***Machine Learning Algorithms:***

A) K-NN (K-Nearest Neighbor):

The k-nearest neighbour algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point. While it can be used for either regression or classification problems, it is typically used as a classification algorithm, working off the assumption that similar points can be found near one another.

B)Naive-Bayes Algorithm:

Naive Bayes is a Supervised Non-linear classification algorithm in R Programming. Naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes theorem with strong (Naive) independence assumptions between the features or variables. The Naive Bayes algorithm is called "Naive" because it makes the assumption that the occurrence of a certain feature is independent of the occurrence of other features. Naive Bayes algorithm is based on Bayes theorem. Bayes theorem gives the conditional probability of an event A given another event B has occurred.

$$P(A|B)= \frac{P(B|A)P(A)}{P(B)}$$

Where,

P (A|B) = Conditional probability of A given B.

P (B|A) = Conditional probability of B given A.

P (A) = Probability of event A.

P (B) = Probability of event B.

C)*Decision Tree:*

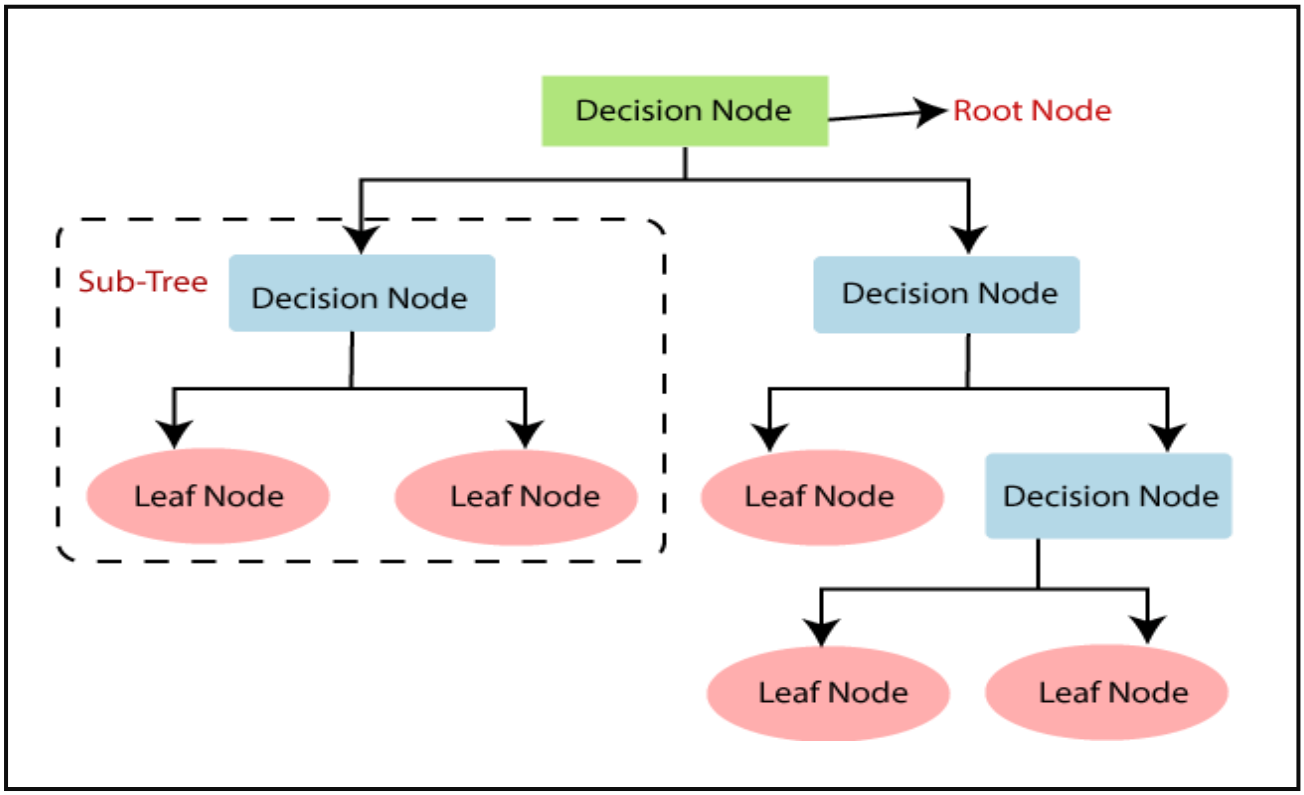
Decision Tree is a supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.

The decisions or the test are performed on the basis of features of the given dataset. It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions. It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches

and constructs a tree-like structure. In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm. A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into subtrees.

Below diagram explains the general structure of a decision tree:



D) *Association rule:*

Association rule mining finds interesting associations and relationships among large sets of data items. This rule shows how frequently an item set occurs in a transaction. A typical example is a Market Based Analysis. Market Based Analysis is one of the key techniques used by large relations to show associations between items. It allows retailers to identify relationships between the items that people buy together frequently. Given a set of transactions, we can find rules that will predict the occurrence of an item based on the occurrences of other items in the transaction.

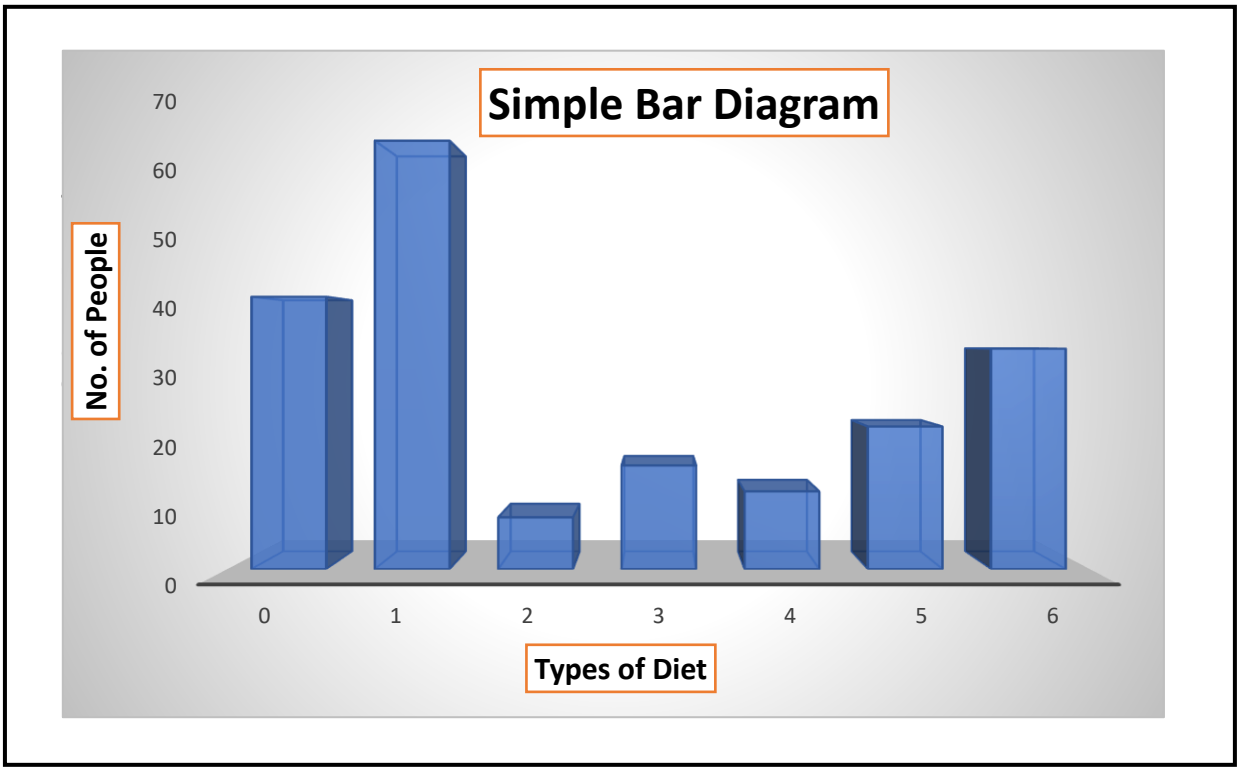
DATA ANALYSIS

- ***Graphical Representation of data-***

a) ***Simple bar diagram-***

Q. Which diet is good for your muscle gain?

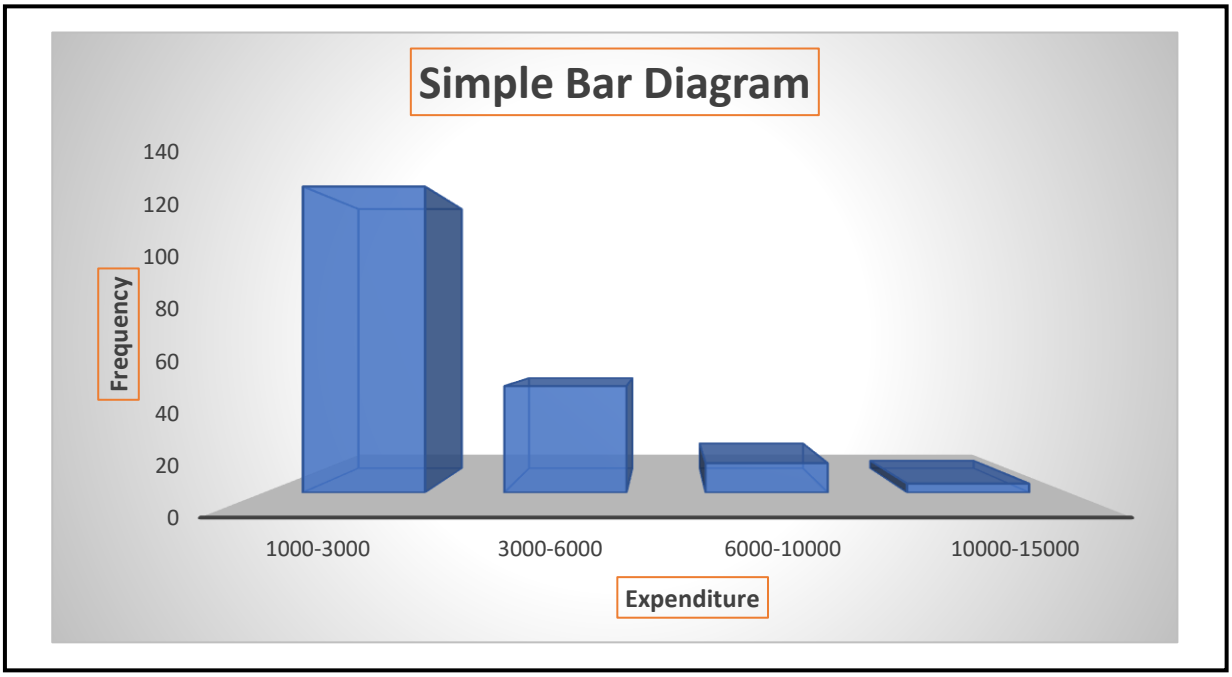
Vegetarian	Non-vegetarian	Supplement	Vegetarian & Non-vegetarian	Vegetarian & Supplement	Non-vegetarian & Supplement	All
0	1	2	3	4	5	6
42	66	8	16	12	22	34



Conclusion: It can be observed from our data that the non- vegetarian diet is good for muscle gain.

Q. Simple bar diagram about income and expenditure on supplements.

Expenditure	Frequency
1000-3000	135
3000-6000	47
6000-10000	13
10000-15000	4



Conclusion: It can be observed from our data that most of the people spend 1000-3000 rupees on the supplements.

b) ***Histogram and frequency polygon-***

Q. How much time you prefer to go gym?

Gym workout time	Notations	Frequency
0-1 (hrs)	0	37
1-2 (hrs)	1	142
2-3 (hrs)	2	20
3-4 (hrs)	4	1

```
>lb=c(0,1,2,3);lb
[1] 0 1 2 3
> ub=c(1,2,3,4);ub
[1] 1 2 3 4
> x=(lb+ub)/2
> x
[1] 0.5 1.5 2.5 3.5
> f=c(37,142,20,1);f
[1]  37 142  20   1
> data.frame(x,f)
   x   f
1 0.5  37
2 1.5 142
3 2.5  20
4 3.5   1
> brks=c(lb[1],ub);brks
[1] 0 1 2 3 4
> y=rep(x,f);y
  [1] 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5
 [21] 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 1.5 1.5 1.5
 [41] 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5
```

[61] 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5

[81] 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5

[101] 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5

[121] 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5

[141] 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5

[161] 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 2.5

[181] 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 3.5

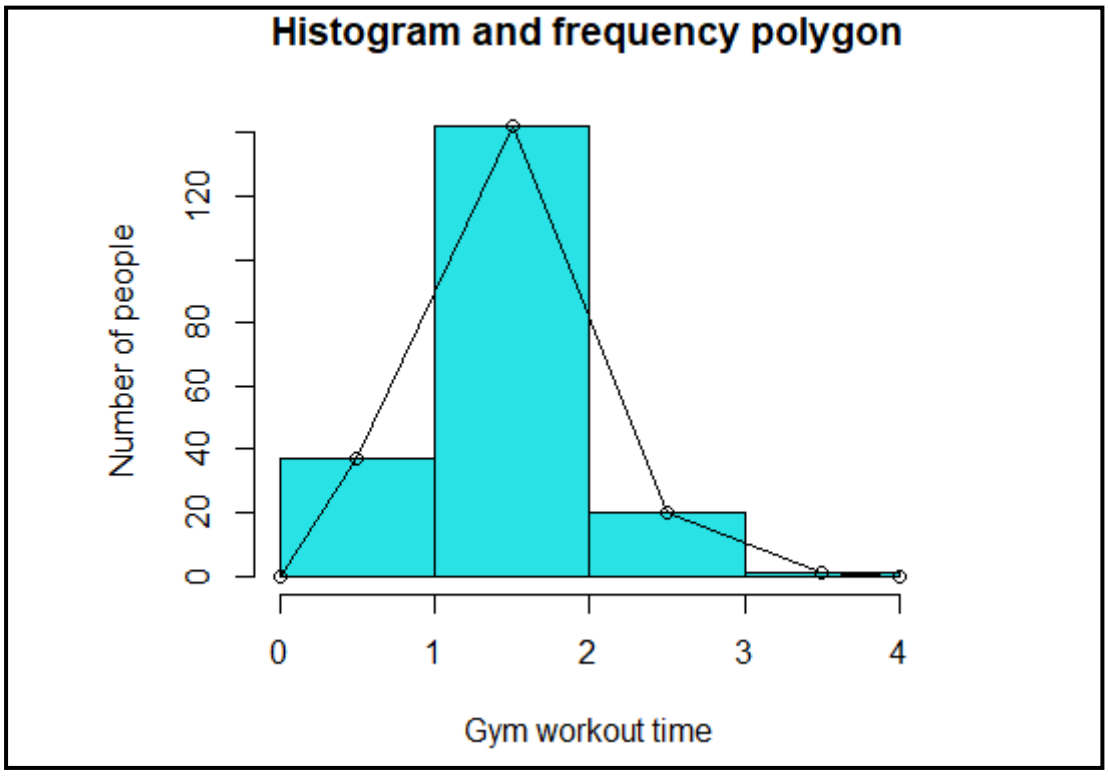
> hist(y,breaks=brks,col=5,xlab="Gym workout time",ylab="Number of people",main="Histogram and frequency polygon")

> x1=c(0,x,4)

> f1=c(0,f,0)

> lines(x1,f1,pch=4)

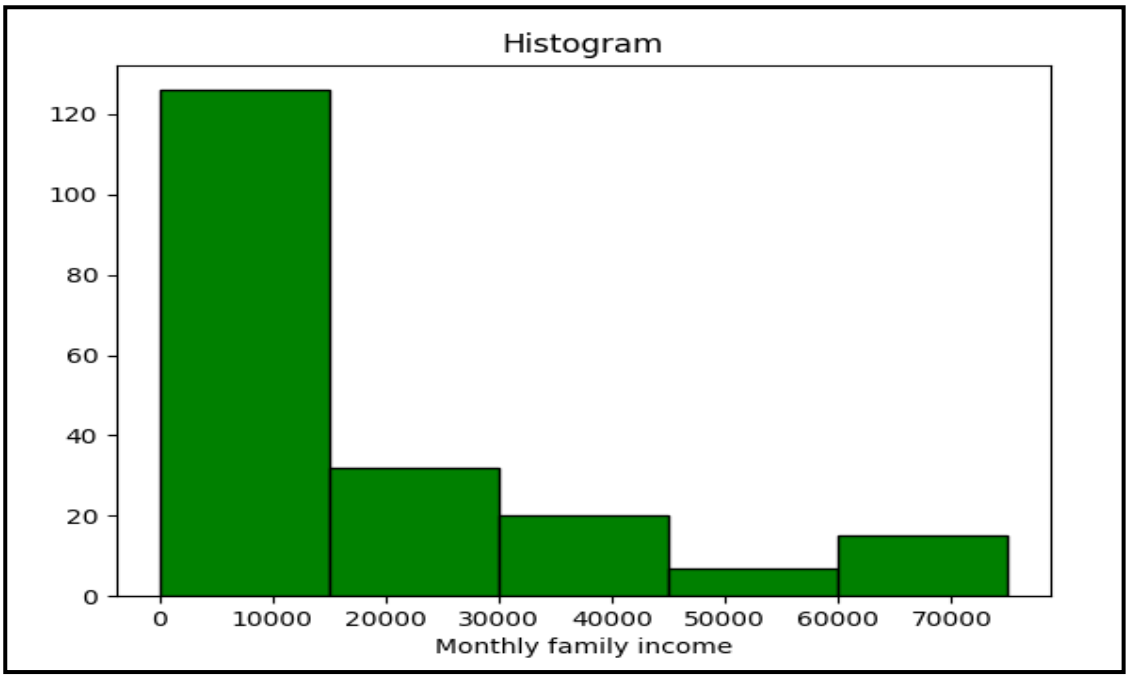
> points(x1,f1)



Conclusion: It can be observed from our data that mostly people prefer 1 to 2 hours daily in the gym.

b) Histogram for Monthly Income-

```
import matplotlib.pyplot as plt
import numpy as np
lb=[0,15000,30000,45000,60000]
ub=[15000,30000,45000,60000,75000]
l=np.array(lb)
u=np.array(ub)
mid=(l+u)/2
f=[126,32,20,7,15]
f1=np.array(f)
y=np.repeat(mid,f1)
interval=[0,15000,30000,45000,60000,75000]
plt.hist(y,interval,edgecolor='black',color='Green')
plt.xlabel("Monthly family income")
plt.title("Histogram")
plt.show()
```

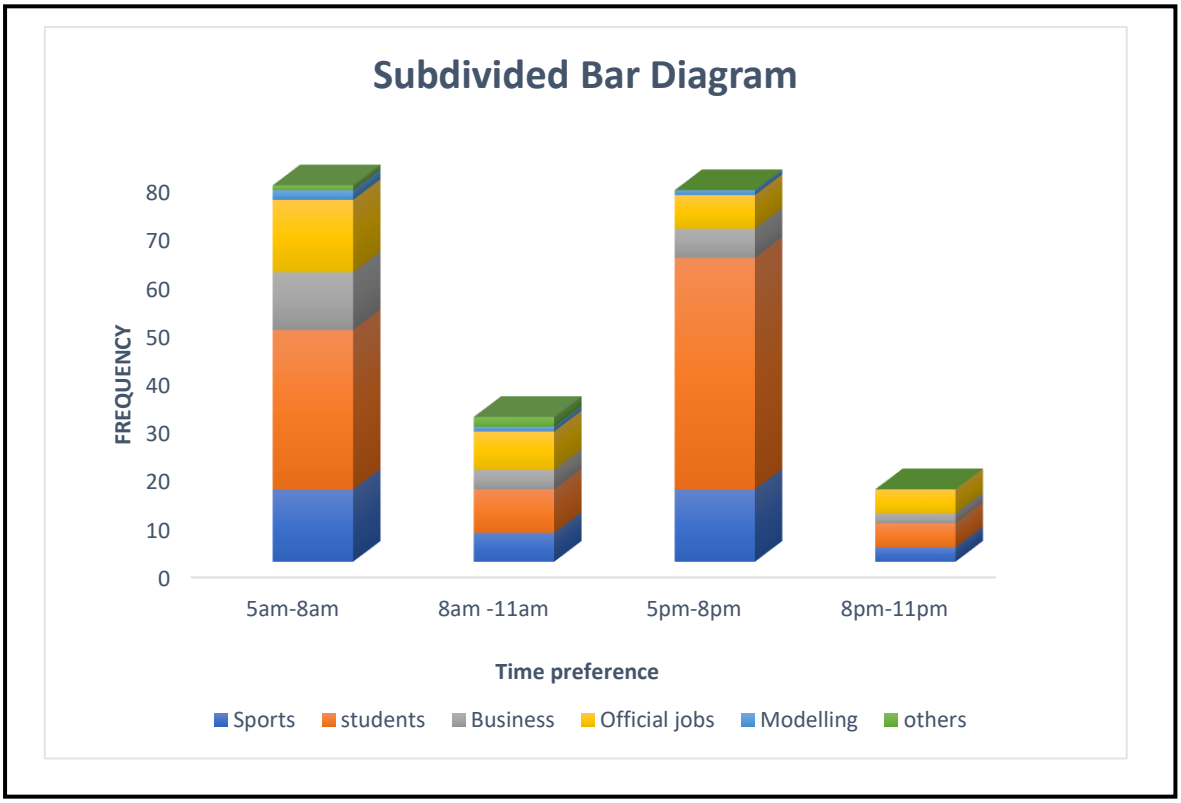


***Conclusion*: It can be observed from our data that monthly income of people are positively skewed.**

c) *Subdivided Bar Diagram-*

Q.Profession wise which time is most preferable?

Profession	5am-8am	8am -11am	5pm-8pm	8pm-11pm
Sports	15	6	15	3
students	33	9	48	5
Business	12	4	6	2
Official jobs	15	8	7	5
Modelling	2	1	1	0
others	1	2	0	0

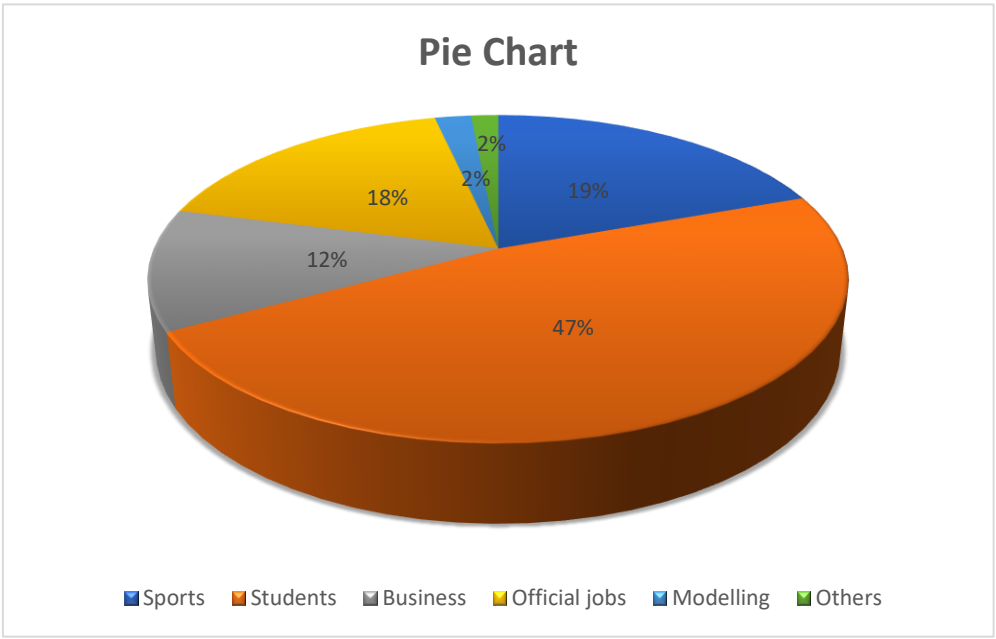


Conclusion: It can be observed from our data that people prefer morning time and evening time 5 to 8 for gym.

d) *Pie Chart-*

Q. Which profession prefers most to go to gyms?

Profession	Sports	Students	Business	Official Jobs	Modelling	Others
Frequency	39	95	24	35	4	3



Conclusion: **It can be observed from our data that students are more likely prefer to go to gym.**

e) *Boxplot-*

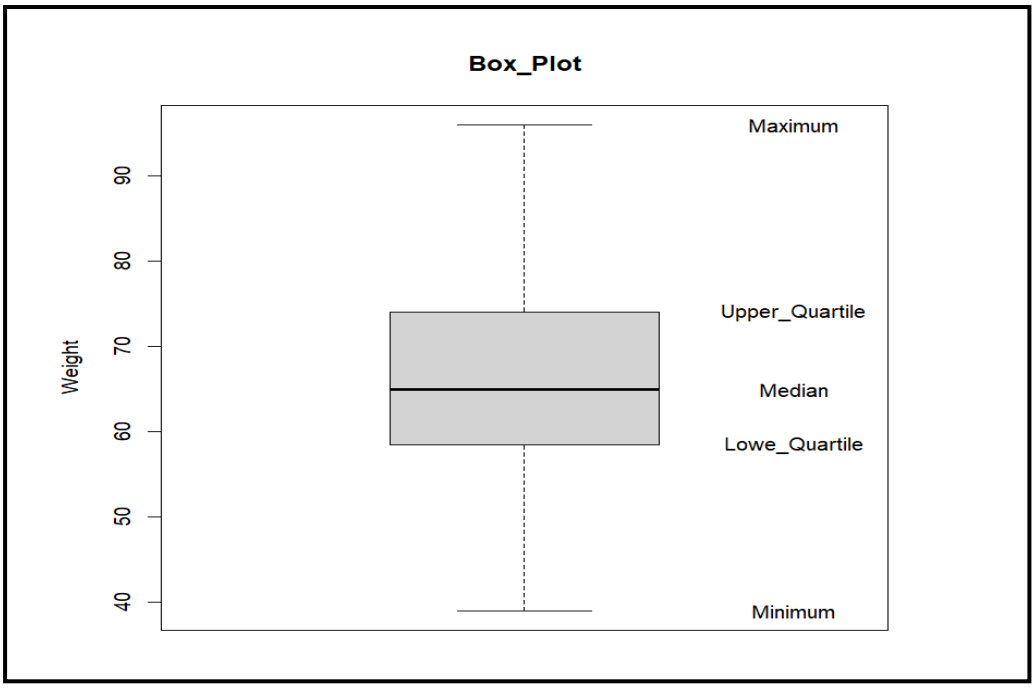
Q. Find the minimum, maximum, and median weight after joining the gym.

```
r=read.csv("C:\\Users\\ABHISHEK\\Desktop\\New folder (2)\\data.csv")
x=Weight after joining the gym by using supplements
x=r$x
length(x)
f=fivenum(x);f
```

```
boxplot(x,main="Box_Plot",ylab="Weight")

lb=c("Minimum","Lowe_Quartile","Median","Upper_Quartile","Maximum")

text(rep(1.4,5),f,labels=lb)
```

Output:

```
r=read.csv("C:\\Users\\ABHISHEK\\Desktop\\New folder (2)\\data.csv")

> x=Weight after joining the gym by using supplements
> x=r$x
> length(x)
[1] 200
> f=fivenum(x)
[1] 39.0 58.5 65.0 74.0 96.0
> boxplot(x,main="Box_Plot",ylab="Weight")
> lb=c("Minimum","Lowe_Quartile","Median","Upper_Quartile","Maximum")
> text(rep(1.4,5),f,labels=lb)
```



Conclusion: It can be observed from our data that the median weight after joining gym is 64 kg, minimum weight after joining the gym is 35 kg and maximum weight after joining the gym is 96 kg.

- ***Statistical Process Control Tools***

a) Check sheet:

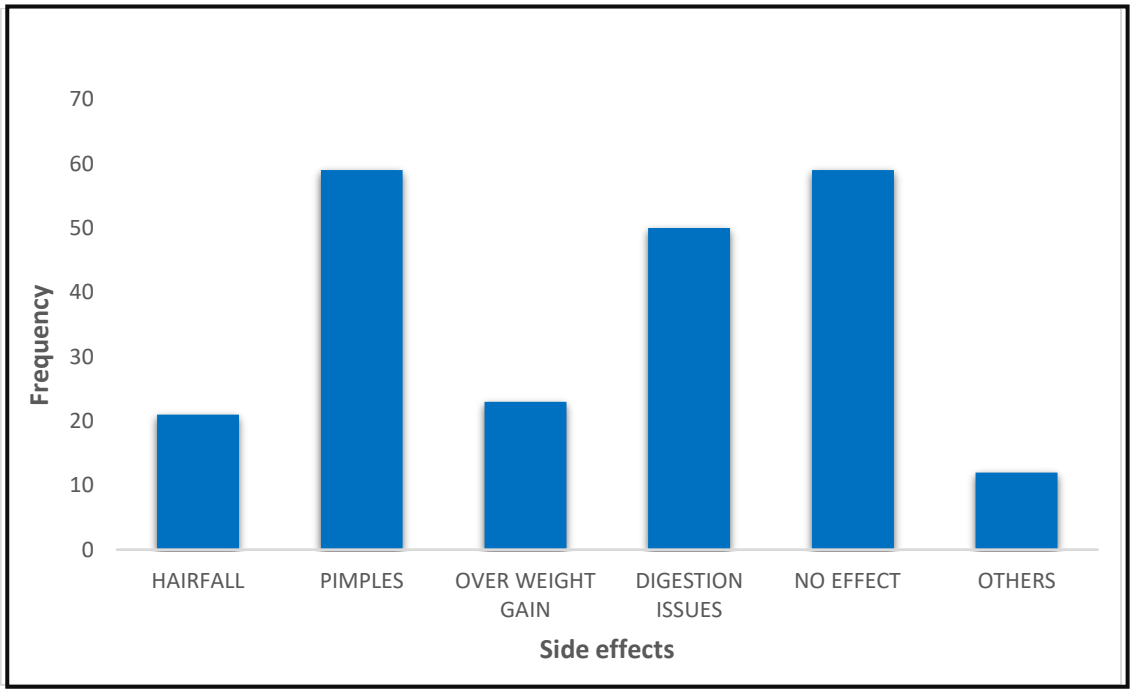
Notations-	0	1	2	3	4	5	
Sample No	***Hair fall***	***Pimples***	***Over Weight Gain***	***Digestion Issues***	***No Effect***	***Others***	***Total***
1		X					1
2	X						1
3				X			1
4		X					1
5				X			1
6		X					1
7					X		1
8		X					1
9	X						1
10					X		1
11						X	1
12					X		1
13			X				1
14		X					1
15				X			1
16		X					1
17	X						1
18		X					1
19				X			1
20					X		1
21				X			1
22		X		X			2
23		X					1
24				X			1
25					X		1
26					X		1
27	X						1
28				X			1
29		X					1
30		X					1
31					X		1
32		X					1
33					X		1
34					X		1
35					X		1
36					X		1

37		X					1
38	X		X				2
39		X					1
40					X		1
41				X			1
42				X			1
43		X					1
44					X		1
45				X			1
46			X				1
47		X					1
48			X				1
49						X	1
50		X					1
51		X					1
52					X		1
53			X				1
54					X		1
55	X			X			2
56					X		1
57					X		1
58					X		1
59					X		1
60						X	1
61		X		X			2
62				X			1
63	X	X					2
64				X			1
65					X		1
66	X						1
67	X						1
68					X		1
69		X					1
70					X		1
71	X	X					2
72				X			1
73			X				1
74		X					1
75		X					1
76						X	1
77				X			1
78			X				1
79	X			X			2
80		X					1
81	X	X					2

82					X		1
83		X		X			2
84				X			1
85			X				1
86		X					1
87				X			1
88				X			1
89						X	1
90		X					1
91			X				1
92					X		1
93	X						1
94					X		1
95					X		1
96				X			1
97				X			1
98		X					1
99					X		1
100		X					1
101			X				1
102					X		1
103		X					1
104		X					1
105			X				1
106					X		1
107		X					1
108	X			X			2
109		X		X			2
110	X	X					2
111					X		1
112					X		1
113						X	1
114		X					1
115					X		1
116					X		1
117					X		1
118				X			1
119	X		X				2
120					X		1
121		X					1
122		X	X				2
123		X					1
124			X				1
125					X		1
126					X		1

127			X				1
128					X		1
129				X			1
130		X	X				2
131				X			1
132					X		1
133				X			1
134		X		X			2
135						X	1
136				X			1
137						X	1
138			X	X			2
139		X					1
140		X					1
141	X			X			2
142	X						1
143					X		1
144					X		1
145					X		1
146		X					1
147					X		1
148		X					1
149					X		1
150		X					1
151		X					1
152				X			1
153				X			1
154		X					1
155					X		1
156		X	X				2
157		X					1
158	X						1
159				X			1
160					X		1
161	X	X					2
162				X			1
163					X		1
164					X		1
165		X		X			2
166						X	1
167				X			1
168			X				1
169					X		1
170		X					1
171					X		1

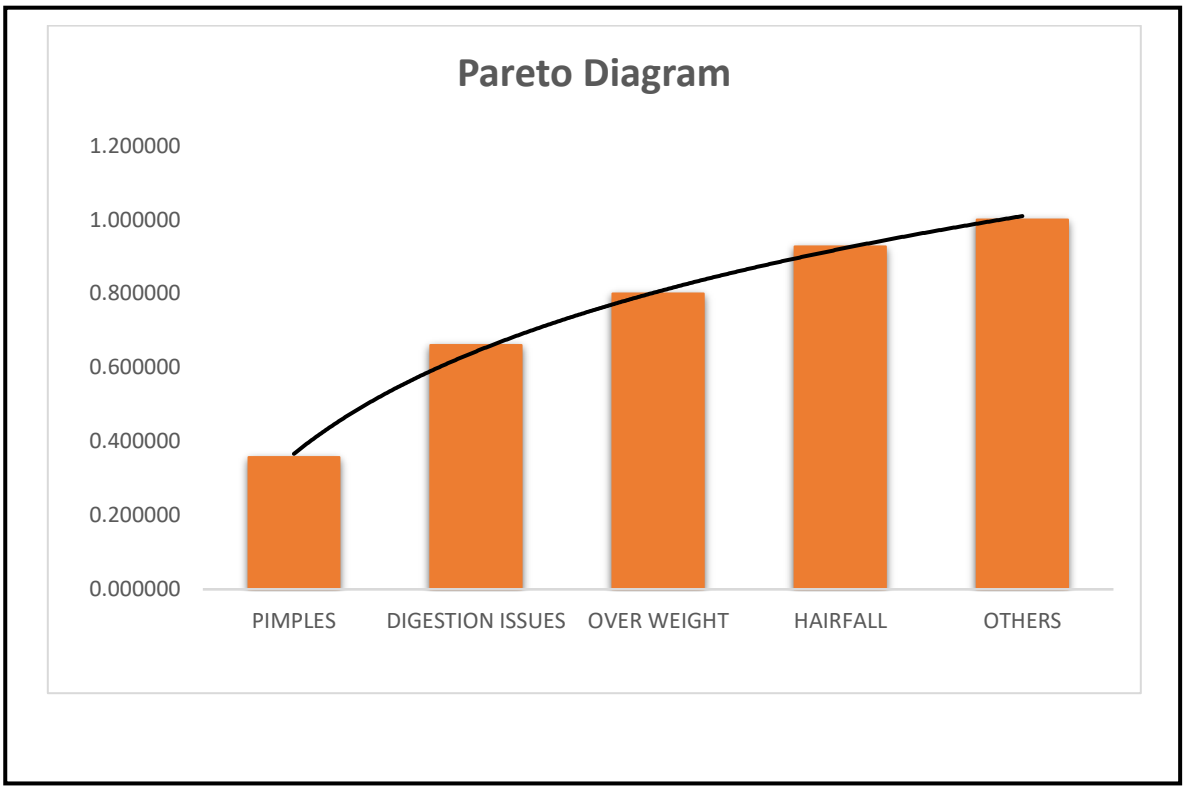
172						X	1
173			X				1
174					X		1
175					X		1
176		X					1
177			X	X			2
178					X		1
179		X					1
180				X			1
181						X	1
182					X		1
183						X	1
184				X			1
185			X	X			2
186					X		1
187	X						1
188		X					1
189				X			1
190					X		1
191			X	X			2
192		X					1
193				X			1
194				X			1
195				X			1
196					X		1
197					X		1
198		X					1
199				X			1
200					X		1
Total	21	59	23	50	59	12	



Conclusion: **It can be observed from our data that pimple is the major side effect of supplements.**

b) Pareto Diagram:

Types Of Causes	No. of Defect	Relative Frequency	Lcf
Pimples	59	0.357576	0.357576
Digestion Issues	50	0.303030	0.660606
Over Weight	23	0.139394	0.800000
Hair fall	21	0.127273	0.927273
Others	12	0.072727	1.000000
Total	**165**		



***Conclusion*: It can be observed from our data that after using the supplements pimples, Digestion issues and Overweight are the most severe causes.**

Testing of Hypothesis

a) Normality Test- To check the normality of population by Shapiro test

i. Here, we want to test,

H_0 =c("Weight before joining the gym is normally distributed")

H_1 =c("Weight before joining the gym is not normally distributed")

```
r=read.csv("D://TYBSC STAT//tybsc project//t-test.csv")
d1=r$bw
shapiro.test(d1)
```

```
        Shapiro-Wilk normality test

data:  d1
W = 0.88509, p-value = 3.116e-11
```

```
 >los=c(0.05)
 >pvalue=c(3.116e-11)
```

Criteria=c("If p-value is less than level of significance alpha then we reject H_0 at alpha% los, otherwise we accept H_0")

```
 >if(pvalue<los)
+{print("reject H0  DECISION: Weight before joining the gym is not normally distributed")}
+else;
+print("otherwise accept H0")
[1] "reject H0
```

Decision: Here, p-value is 3.116e-11< l. o. s (5%)

Hence, we may reject H_0 at 5% l .o. s.

Conclusion: **It can be observed from our data that weight before joining the gym is not normally distributed.**

i. Here, we want to test,

H_0=c("Weight after joining the gym is normally distributed")

H_1=c("Weight after joining the gym is not normally distributed")

```
r=read.csv("D://TYBSC STAT//tybsc project//t-test.csv")
d2=r$aw
shapiro.test(d2)

        Shapiro-Wilk normality test

data: d2
W = 0.99148, p-value = 0.2907
>los=c(0.05)
> pvalue=c(0.2907)
```

Criteria=c("If p-value is less than level of significance alpha then we reject H_0 at alpha% los , otherwise accept H_0")

```
> if(pvalue<los)
+ {print("reject H0 decision: Weight after joining the gym is not normally distributed")}
+else
+ {print("otherwise accept H0 decision: Weight after joining the gym is normally distributed"
)}
[1] "accept H0 “
```

Decision: **Here, p-value is 0.2907> l.o.s(5%)**

Hence, we may accept H_0 at 5% l.o.s.

Conclusion: **It can be observed from our data that weight after joining the gym is normally distributed.**

- ***Testing equality of variance of both population-***

```
# Testing equality of variance using Bartlett test-
  Here, we want to test,
> H0=c("Variances of the both population is same,")
> H1=c("Variances of the both population is not same.")
> x=list(d1,d2)
> bartlett.test(x)

        Bartlett test of homogeneity of variances
```

data: x

Bartlett's K-squared = 0.083689, df = 1, p-value = 0.7724

> los=c(0.05)

> pvalue=c(0.7724)

> if(pvalue<los)

+ {print("reject H_0 “)}else

+ {print(" Accept H_0 ")}

[1] " Accept H_0"

Decision: **Here, p-value is 0.7724> l.o.s(5%)**

Hence, we may accept H_0 at 5% l.o.s.

***Conclusion*:** **Hence, we can conclude that Variances of the both population is same.**

b) Sign paired test:

Here, we want to test,

H_0=c("The weight of person before and after taking a supplement is same.")

H_1 =c(“The weight of a person before and after taking a supplement is increasing.")

where, D= x-y

x= Weight before taking the supplements.

Y=Weight after taking the supplements.

r=read.csv("D://TYBSC STAT//tybsc project//t-test.csv")

x=r$bw

y=r$aw

d=x-y

sp=length(d[d>0])

[1] 66

sn=length(d[d<0])

[1] 131

n=sp+sn

[1] 197

pv=pbinom(sp,n,0.5)

pv=2.117267e-06 #pv:p-value

sp=66 #sp:s plus value

los=0.05

Criteria: c("If p-value is less than level of significance alpha then we reject H_0 at alpha% los, otherwise accept H_0")

if(p v<l o s)

+{print("decision : reject H_0")}

else{

+print("decision : accept H_0")}

Decision: **Here, p-value is 2.117267e-06< l.o.s (5%)**

Hence, we may reject H_0 at 5% l.o.s.

Conclusion: **It can be observed from our data that there is increase in the weight after taking a supplement.**

c) Proportion test:

Test for difference between two population proportions-

In female population out of a random sample of 28 person 22 were found to be (vegetarian and non- vegetarians) while in another male population out of 172 persons 103 were found to be (vegetarian and non- vegetarian) do you find significant difference in food habit of the people of both the profession?

P_1=Diet of the Male populations.

P_2=Diet of the Female populations.

Two tail test H_0: $P_1 = P_2$ V/S H_1: $P_1 \neq P_2$

r=read.csv("D://TYBSC STAT//tybsc project//proportion test.csv")

X_1=r$Which.type.of.profession.you.are.belonging.from.

X_2=r$Which.diet.is.good.for.your.muscle.gain.

>H_0=c(" $P_1 = P_2$")

> H_1=c("$P_1 \neq P_2$")

> x=c(22,103)

> n=c(28,172)

> prop.test(x,n)

2-sample test for equality of proportions with continuity correction

data: x out of n

X-squared = 25.904, df = 1, p-value = 0.08219

alternative hypothesis: two.sided

95 percent confidence interval:

0.3472111 0.7387806

sample estimates:

prop 1 prop 2

0.7857143 0.2427184

```
> alpha=0.05
> pvalue= 0.08219
> if(pvalue<alpha)
+ {
+ print("reject H0")
+ }else{
+ print("accept H0")
+ }
[1] "accept H0"
```

Decision: **Here, p-value is 0.08219 > l. o. s (5%)**
Hence, we may accept H_0 at 5% l. o. s.

Conclusion: **It can be observed from our data that there is no significant difference in the food habits between male and female.**

d) Chi-square test for independence of attributes-

To Check whether the preference of protein intake is independent on effects or not.

Attributes \Effects	Good	Moderate	Very Good
Concentrate	**18**	**1**	**36**
Hydrolysate	**9**	**0**	**10**
Isolate	**38**	**8**	**79**

Chi-square test for independence of attributes

Here, we want to test

H_0=c("The preference of protein intake is independent on effects.")

H_1=c("The preference of protein intake is not independent on effects.")

```
> x=c(18,9,38,1,0,8,36,10,79);x
[1] 18  9 38  1  0  8 36 10 79
> mx=matrix(x,nrow=3,ncol=3)
> chisq.test(mx,correct=F)
```

OUTPUT:

```
     [,1] [,2] [,3]
[1,]   18    1   36
[2,]    9    0   10
[3,]   38    8   79
        Pearson's Chi-squared test
data: mx
X-squared = 4.5545, df = 4, p-value = 0.3361
> alpha=0.05
> pvalue=0.3361
> if(pvalue<alpha)
+print{("reject H0")
```

+ }else{

+print("accept H_0")}

[1] "accept H_0"

Decision: **Here, p-value is 0.3361> l.o.s (5%)**
Hence, we may accept H_0 at 5% l.o.s.

Conclusion: **It can be observed from our data that we conclude that the preference of protein intake is independent on effects.**

e) Multiple linear regression:

Y= Responses for Weight after joining the gym after taking supplements.

x1=Responses for Which diet is good for your muscle gain.

x2= Responses for Which type of Supplement you use for your body.

x3= Responses for Which type of whey protein do you take.

```
r=read.csv("D://TYBSC STAT//tybsc project//multiple regression.csv")
Y=r$What.is.your.weight.after.joining.the.gym.after.using.supplements.
x1=r$Which.diet.is.good.for.your.muscle.gain.
x2=r$Which.type.of.whey.protein.do.you.take.
x3=r$Which.type.of.Supplement.you.use.for.your.body.
reg= lm(Y~x1+x2+x3)
summary(reg)
```

```
Call:
lm(formula = Y~ x1 + x2 + x3)

Residuals:
   Min      1Q  Median      3Q     Max
-45.370  -6.121  -0.430   6.562  29.322

Coefficients:
             Estimate   Std. Error  t value  Pr(>|t|)
(Intercept)  60.8214    1.9911      30.547   2e-16 ***
x1            0.5188     0.3454      1.502    0.13476
x2            0.5700     0.8720      0.654    0.51404
x3            0.8896     0.3171      2.805    0.00554 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Residual standard error: 10.88 on 196 degrees of freedom
Multiple R-squared: 0.04848, Adjusted R-squared: 0.03392
F-statistic: 3.329 on 3 and 196 DF, p-value: 0.02067

> result=cor.test(Y,x1+x2+x3)
> print(result)

Pearson's product-moment correlation

data: Yand x1 + x2 + x3
t = 3.0578, df = 198, p-value = 0.002538
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
0.07584449 0.34104423
sample estimates:
cor
0.212351

Conclusion: **It can be observed from our data that correlation between diet, supplements and type of whey protein is 0.2122351 which is very least. Therefore, diet, supplements and type of whey protein are least correlated.**

Data Analytics

Q. The weight before joining the gym and after joining the gym and type of whey protein is given. We want to predict a person having weight 63 before joining the gym and the weight 66 after joining the gym will use which type of whey protein?

Weight before joining the gym.	63
Weight after joining the gym.	66
k	14

a) *K-Nearest Neighbour:*

X_1= Weight before joining the gym.
X_2= Weight after joining the gym.
Y= Type of whey protein.

What was your Weight before joining the gym?	What is your weight after joining the gym after using supplements?	Which type of whey protein do you take?	Distance		What was your Weight before joining the gym?	What is your weight after joining the gym after using supplements?	Which type of whey protein do you take?	Distance
95	75	Isolate	33.24154		72	80	Isolate	16.64332
74	65	Isolate	11.04536		50	40	Isolate	29.06888
47	55	Isolate	19.41649		145	70	Isolate	82.0975
60	63	Isolate	4.242641		87	79	Concentrate	27.29469
74	82	Isolate	19.41649		86	80	Isolate	26.92582
50	70	Isolate	13.60147		102	71	Concentrate	39.31921
45	60	Isolate	18.97367		95	90	Isolate	40
60	65	Hydrolysate	3.162278		56	65	Concentrate	7.071068
39	50	Hydrolysate	28.84441		64	66	Isolate	1
53	80	Concentrate	17.20465		57	65	Isolate	6.082763
82	75	Isolate	21.0238		88	76	Concentrate	26.92582
45	60	Concentrate	18.97367		55	63	Isolate	8.544004
54	58	Isolate	12.04159		40	50	Concentrate	28.01785
85	80	Isolate	26.07681		65	82	Concentrate	16.12452
67	74	Concentrate	8.944272		54	67	Isolate	9.055385
58	66	Isolate	5		49	53	Concentrate	19.10497
55	59	Hydrolysate	10.63015		85	69	Isolate	22.2036
55	62	Isolate	8.944272		52	78	Isolate	16.27882
65	60	Hydrolysate	6.324555		51	69	Isolate	12.36932

75	78	Isolate	16.97056		57	80	Isolate	15.23155
110	80	Concentrate	49.0408		76	75	Isolate	15.81139
52	65	Isolate	11.04536		70	49	Isolate	18.38478
52	58	Hydrolysate	13.60147		85	70	Isolate	22.36068
100	76	Isolate	38.32754		64	58	Isolate	8.062258
67	62	Isolate	5.656854		60	55	Isolate	11.40175
58	79	Isolate	13.92839		54	57	Isolate	12.72792
60	70	Isolate	5		89	71	Hydrolysate	26.4764
67	89	Isolate	23.34524		58	65	Isolate	5.09902
82	75	Isolate	21.0238		74	86	Isolate	22.82542
50	60	Concentrate	14.31782		45	55	Concentrate	21.09502
85	76	Isolate	24.16609		55	60	Isolate	10
65	72	Isolate	6.324555		50	52	Isolate	19.10497
70	65	Concentrate	7.071068		46	55	Hydrolysate	20.24846
57	65	Concentrate	6.082763		66	78	Isolate	12.36932
55	67	Isolate	8.062258		71	68	Isolate	8.246211
59	60	Isolate	7.211103		78	66	Hydrolysate	15
55	60	Isolate	10		52	56	Isolate	14.86607
39	45	Concentrate	31.89044		62	58	Concentrate	8.062258
67	60	Isolate	7.211103		59	85	Isolate	19.41649
49	57	Hydrolysate	16.64332		59	65	Concentrate	4.123106
45	49	Concentrate	24.75884		95	84	Concentrate	36.71512
48	55	Isolate	18.60108		67	62	Hydrolysate	5.656854
55	60	Concentrate	10		56	58	Isolate	10.63015
56	81	Isolate	16.55295		48	53	Concentrate	19.84943
67	65	Isolate	4.123106		80	67	Isolate	17.02939
60	69	Isolate	4.242641		86	79	Concentrate	26.41969
60	66	Isolate	3		50	59	Isolate	14.76482
60	74.5	Concentrate	9.013878		45	52	Concentrate	22.80351
50	56	Isolate	16.40122		75	70	Concentrate	12.64911
40	70	Concentrate	23.34524		76	66	Isolate	13
45	70	Concentrate	18.43909		49	57	Isolate	16.64332
48	59	Isolate	16.55295		63	74	Concentrate	8
80	70	Isolate	17.46425		55	63	Isolate	8.544004
97	83	Isolate	38.01316		88	87	Concentrate	32.64966
54	67	Isolate	9.055385		54	56	Hydrolysate	13.45362
60	73	Isolate	7.615773		62	65	Concentrate	1.414214
57	60	Isolate	8.485281		62	65	Isolate	1.414214
60	50	Hydrolysate	16.27882		52	68	Isolate	11.18034
65	72	Isolate	6.324555		60	72.4	Isolate	7.068239
98	92	Concentrate	43.60046		75	72	Isolate	13.41641
98	82	Hydrolysate	38.48376		48	46	Isolate	25
60	65	Isolate	3.162278		50	58	Concentrate	15.26434
85	75	Isolate	23.76973		63	68	Isolate	2
70	66	Isolate	7		48	65	Isolate	15.0333

39	39	Hydrolysate	36.12478		90	84	Isolate	32.44996
46	54	Isolate	20.80865		65	64	Concentrate	2.828427
59	75	Concentrate	9.848858		55	60	Hydrolysate	10
48	55	Hydrolysate	18.60108		45	68	Isolate	18.11077
60	65	Hydrolysate	3.162278		40	50	Concentrate	28.01785
59	57	Isolate	9.848858		65	78	Concentrate	12.16553
85	75	Concentrate	23.76973		78	70	Concentrate	15.52417
40	45	Isolate	31.14482		42	56	Concentrate	23.25941
48	56	Concentrate	18.02776		83	82	Isolate	25.6125
55	59	Concentrate	10.63015		45	41	Hydrolysate	30.80584
46	63	Concentrate	17.26268		38	46	Concentrate	32.01562
41	64	Isolate	22.09072		40	55	Isolate	25.4951
79	68	Isolate	16.12452		64	58	Concentrate	8.062258
47	57	Isolate	18.35756		72	75	Isolate	12.72792
45	47	Isolate	26.1725		54	52.8	Isolate	15.97623
54	68	Concentrate	9.219544		70	65	Isolate	7.071068
60	74	Concentrate	8.544004		56	60	Isolate	9.219544
45	55	Isolate	21.09502		84	76	Isolate	23.25941
55	63	Isolate	8.544004		53	60	Isolate	11.6619
65	75	Concentrate	9.219544		98	86	Isolate	40.31129
68	75	Hydrolysate	10.29563		57	65	Isolate	6.082763
60	80	Isolate	14.31782		75	78	Isolate	16.97056
50	56	Isolate	16.40122		58	67	Isolate	5.09902
55	58	Concentrate	11.31371		56	85	Concentrate	20.24846
70	60	Isolate	9.219544		63	70	Isolate	4
62	59	Isolate	7.071068		47	54	Isolate	20
96	79	Isolate	35.4683		57	69	Isolate	6.708204
55	59	Isolate	10.63015		62	70	Isolate	4.123106
70	68	Isolate	7.28011		50	60	Isolate	14.31782
54	58	Concentrate	12.04159		60	70	Isolate	5
51	60	Concentrate	13.41641		60	65	Isolate	3.162278
65	65	Isolate	2.236068		40	65	Concentrate	23.02173
51	60	Isolate	13.41641		132	96	Isolate	75.23962
38	70	Isolate	25.31798		65	70	Concentrate	4.472136
55	60	Isolate	10		77	72	Isolate	15.23155
55	69	Concentrate	8.544004		70	66	Isolate	7

Conclusion: **It can be observed from our data that the person having weight 63 before gym and weight 66 after joining the gym will use isolate type of whey protein.**

b) Naive Bayes Classifier:

X_1= Weight before joining the gym.
X_2= Weight after joining the gym.
Y= Type of whey protein.

```
r=read.csv("C:\\Users\\ABHISHEK\\Desktop\\import.csv")
d=data.frame(r)
sam=sample(2,nrow(d),prob=c(.8,.2),replace = T)
sam
data_train=d[sam==1, ]
View(data_train)
data_test=d[sam==2, ]
View(data_test)
model_nb=naiveBayes(y~.,d=data_train)
model_nb
pred_cl=predict(model_nb,newdata=data_test)
pred_cl
CM=table(pred_cl,data_test[,3])
CM
accuracy=(sum(diag(CM))/sum(CM))*100
accuracy
```

Output:

```
r=read.csv("C:\\Users\\ABHISHEK\\Desktop\\New folder (2)\\import.csv")
> d=data.frame(r)
> model=naiveBayes(y~.,d)
> pred=predict(model,newdata=data.frame(x1=63,x2=66))
> pred
[1] Isolate
Levels: Concentrate Hydrolysate Isolate
> sam=sample(2,nrow(d),prob=c(.8,.2),replace = T)
> sam
  [1] 1 1 1 1 2 2 2 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 2 1 2 1 1 1
 [56] 1 1 1 1 1 1 1 1 1 1 2 1 2 2 1 1 1 2 1 1 1 2 1 1 2 1 2 1 1 1 1 1 1 2 1 1 2 1 1 2 1 1 1 1 1 1
1 1 1 2 1 1 2 1
[111] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1
2 1 1 1 2 1 2 1
[166] 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 2 1 1 1 2 1 1 1 1
> data_train=d[sam==1, ]
> data_test=d[sam==2, ]
> model_nb1=naiveBayes(y~.,d=data_train)
> model_nb1

Naive Bayes Classifier for Discrete Predictors
Call:
naiveBayes.default(x = X, y = Y, laplace = laplace)
```

```
A-priori probabilities:
Y
Concentrate    Hydrolysate      Isolate
 0.29585799    0.09467456   0.60946746

Conditional probabilities:
          x1
Y                   [, 1]             [, 2]
  Concentrate   62.74000         21.02846
  Hydrolysate   61.62500         15.73054
  Isolate           63.07767         15.60586

x2
Y                   [, 1]             [, 2]
  Concentrate   65.96800      11.49171
  Hydrolysate   64.00000      8.13224
  Isolate           66.11165      11.08168

> pred_cl=predict(model_nb1,newdata=data_test)
> pred_cl
 [1] Isolate     Isolate     Isolate     Isolate     Isolate     Isolate     Isolate     Isolate     Isolate
[10] Isolate     Isolate     Isolate     Isolate     Isolate     Isolate     Isolate     Isolate     Isolate
[19] Concentrate Isolate     Isolate     Isolate     Isolate     Isolate     Isolate     Isolate     Isolate
[28] Isolate     Isolate     Isolate     Isolate
Levels: Concentrate Hydrolysate Isolate
> CM=table(pred_cl,data_test[,3])
> CM

pred_cl       Concentrate Hydrolysate Isolate
  Concentrate          0          0       1
  Hydrolysate          0          0       0
  Isolate              5          3      22
> accuracy=(sum(diag(CM))/sum(CM))*100
> accuracy
[1] 70.96774
```

Conclusion: **It can be observed from our data that the predicted value for person having weight before gym 63 and after gym 66 will take isolate type of whey protein. The accuracy of fitted naïve Bayes model is 70%.**

c) *Decision Tree:*

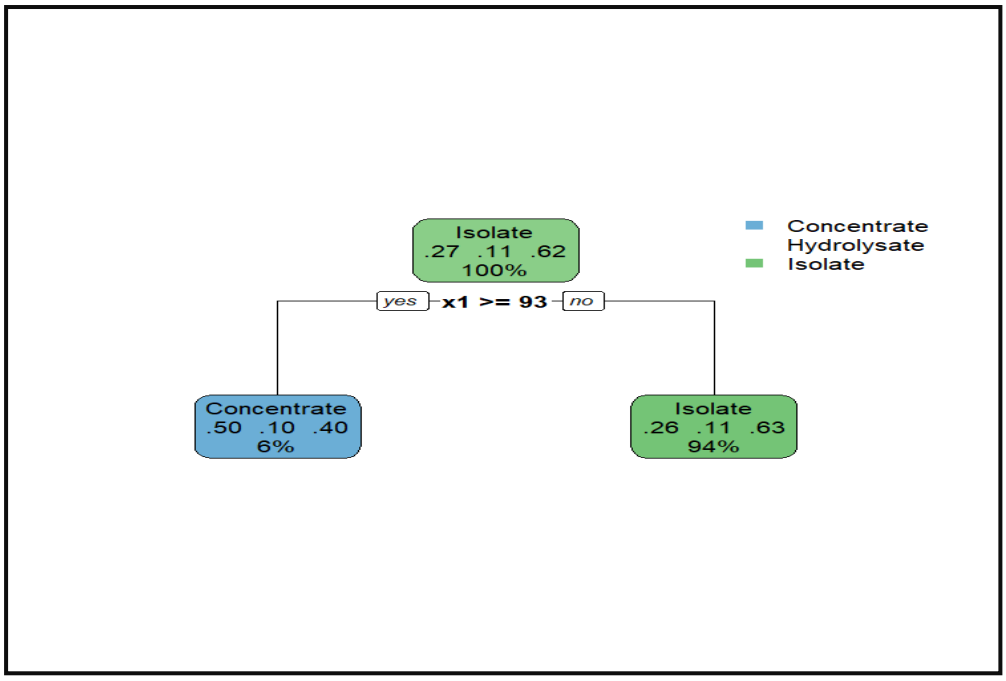
X_1= Weight before joining the gym.
X_2= Weight after joining the gym.
Y= Type of whey protein.

```
> r=read.csv("C:\\Users\\ABHISHEK\\Desktop\\New folder (2)\\import.csv")
> d=data.frame(r)
> model=rpart(y~.,d,method='class')
> pred=predict(model,newdata=data.frame(x1=63,x2=66),type = 'class');pred
     1
Isolate
Levels: Concentrate Hydrolysate Isolate
> sam=sample(2,nrow(d),prob=c(.8,.2),replace = TRUE)
> sam
[1] 1 1 2 1 2 1 1 2 2 2 1 2 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1
1 1 1 1 1 1 2 1
[56] 1 1 2 1 2 1 1 2 1 1 1 1 1 2 2 2 1 1 1 1 2 1 1 1 1 2 1 1 1 1 2 2 1 1 1 2 1 1 1 1 1 1 1 1 2
1 1 1 1 1 1 1 1
[111] 1 2 2 2 1 1 1 1 2 1 1 1 1 1 2 1 2 2 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 2 2 1 1
2 1 1 2 1 1 1 1
[166] 1 2 1 2 1 1 1 1 1 2 1 2 1 1 1 1 1 2 1 1 1 2 1 2 1 1 2 1 1 1 1 1 1 1 2
> data_train=d[sam==1, ]
> data_test=d[sam==2, ]
> model_dt=rpart(y~.,data_train,method='class')
> model_dt
```

Output:

```
n= 154
node, split, n, loss, yval, (yprob)
      * denotes terminal node

1) root 154 59 Isolate (0.2727273 0.1103896 0.6168831)
  2) x1>=92.5 10  5 Concentrate (0.5000000 0.1000000 0.4000000) *
  3) x1< 92.5 144 53 Isolate (0.2569444 0.1111111 0.6319444) *

> rpart.plot(model_dt)
> pred_dt=predict(model_dt,data_test,type="class")
> pred_dt

        3         5         8         9        10        12        15        16        40
  Isolate   Isolate   Isolate   Isolate   Isolate   Isolate   Isolate   Isolate   Isolate
       44        54        58        60        63        69        70        71        76
  Isolate Concentrate   Isolate Concentrate   Isolate   Isolate   Isolate   Isolate   Isolate
       81        87        88        92       102       112       113       114       119
  Isolate   Isolate   Isolate   Isolate   Isolate   Isolate   Isolate   Isolate   Isolate
      126       128       129       138       141       152       154       155       158
  Isolate   Isolate   Isolate Concentrate   Isolate   Isolate   Isolate   Isolate
```

```
   161      167      169      175      177      183      187      189      192
Isolate  Isolate  Isolate  Isolate  Isolate  Isolate  Isolate  Isolate  Isolate
   200
Isolate

Levels: Concentrate Hydrolysate Isolate

> cm=table(pred_dt,data_test[,3]);cm

  pred_dt      Concentrate  Hydrolysate  Isolate
Concentrate        1            0          2
Hydrolysate        0            0          0
Isolate           12            2         29

> accuracy=(sum(diag(cm))/sum(cm))*100
> accuracy
[1] 65.21739
```



Conclusion: **It can be observed from our data that the predicted value for person having weight before gym 63 and after gym 66 will take isolate type of whey protein. The accuracy of the fitted model is 65%.**

d*) Association rules:*

Q. For our data, Transactions of some people taking supplements suggested by their respective trainer is given below, minimum support is set to 0.4 and minimum confidence is 70%. List all the items with their support and confidence, satisfying the requirements of supplements with their confidence.

Transaction	Product 1	Product 2	Product 3	Product 4
1	creatine	caffeine	NA	NA
2	creatine	weight gainers	caffeine	NA
3	creatine	whey protein	caffeine	fish oil
4	creatine	whey protein	caffeine	fish oil
5	whey protein	caffeine	creatine	fish oil
6	whey protein	weight gainers	NA	NA
7	caffeine	whey protein	fish oil	NA
8	whey protein	fish oil	NA	NA
9	weight gainer	whey protein	NA	NA
10	whey protein	weight gainers	creatine	NA
11	creatine	whey protein	fish oil	caffeine
12	creatine	fish oil	whey protein	NA
13	creatine	fish oil	whey protein	NA
14	whey protein	weight gainers	NA	NA
15	creatine	whey protein	fish oil	caffeine
16	whey protein	weight gainers	NA	NA
17	creatine	whey protein	fish oil	caffeine
18	caffeine	whey protein	fish oil	NA
19	whey protein	weight gainers	NA	NA
20	creatine	whey protein	caffeine	NA
21	whey protein	fish oil	NA	NA
22	creatine	NA	fish oil	NA
23	whey protein	caffeine	fish oil	NA
24	creatine	weight gainers	whey protein	caffeine

25	creatine	whey protein	weight gainers	NA
26	creatine	whey protein	fish oil	caffeine
27	creatine	whey protein	NA	NA
28	whey protein	caffeine	NA	NA
29	weight gainer	fish oil	caffeine	NA
30	whey protein	weight gainers	caffeine	creatine

Output:

Support	0.4
Confidence	0.7

Transaction	Creatine	Caffeine	Whey Protein	Fish oil	Weight Gainers
1	1	1	0	0	0
2	1	1	0	0	1
3	1	1	1	1	0
4	1	1	1	1	0
5	1	1	1	1	0
6	0	0	1	0	1
7	0	1	1	1	0
8	0	0	1	1	0
9	1	0	1	0	1
10	1	0	1	0	1
11	1	1	1	1	0
12	1	0	1	1	0
13	1	0	1	1	0
14	0	0	1	0	1
15	1	1	1	1	0
16	0	0	1	0	1
17	1	1	1	1	0
18	1	0	1	1	0
19	0	0	1	0	1
20	1	1	1	0	0
21	0	0	1	1	0
22	1	0	0	1	0
23	0	1	1	1	0
24	1	1	1	0	1
25	1	0	1	0	1

26	1	1	1	1	0
27	1	0	1	0	0
28	0	1	1	0	0
29	0	1	0	1	1
30	1	1	1	0	1
Total	20	16	26	16	11

1 item data set		
Product	Frequency	Support
Creatine	20	0.6667
Caffeine	16	0.5333
Whey Protein	26	0.8667
Fish oil	16	0.5333

1 item frequent data set		
Product	Frequency	Support
Creatine	20	0.666667
Caffeine	16	0.533333
Whey Protein	26	0.866667
Fish oil	16	0.533333

2 item data set		
Product	Frequency	Support
Creatine, Caffeine	12	0.4
Creatine, Whey protein	16	0.533333333
Creatine, Fish oil	9	0.3
Caffeine, Whey Protein	13	0.433333333
Caffeine, Fish oil	10	0.333333333
Whey protein, Fish oil	14	0.466666667

2 item frequent data set		
Product	Frequency	Support
Creatine, Caffeine	12	0.4
Creatine, Whey protein	16	0.533333
Caffeine, Whey Protein	13	0.433333
Whey protein, Fish oil	14	0.466667

Rules	Confidence
Creatine, Caffeine	0.6
Creatine, Whey protein	0.8
Caffeine, Whey Protein	0.8125
Whey protein, Fish oil	0.538461538

3 item data set	Frequency	Support
Creatine, Caffeine, Whey Protein	10	0.333333333
Creatine, Caffeine, Fish oil	7	0.233333333
Creatine, Whey Protein, Fish oil	10	0.333333333
Caffeine, Whey Protein, Fish oil	9	0.3

Conclusion: **It can be observed from our data that the person buying whey protein may most likely buy caffeine (coffee) and creatine.**

INTERPRETATIONS

- ***For Graphs and Charts-***

a) **Simple Bar Diagram:**

1) It can be observed from our data that maximum number of persons are preferring the non-vegetarian diet. So, we may conclude that non-vegetarian diet is mostly good for muscle gain.
2) It can be observed from our data that most of the people spend 1000-3000 rupees on the supplements**.**

b) **Histogram:**

1) It can be observed from our data that mostly people prefer 1-2 hours daily in the gym.
2) It can be observed from our data that monthly income of people are positively skewed.

c) **Sub-Divided Bar Graph:**

It can be observed from our data that people prefer morning time and evening time 5 to 8 for gym.

d) **Pie-Chart:** It can be observed from our data that students are more likely prefer to go to gym.

e) **Boxplot**: It can be observed from our data that the median weight after joining gym is 64 kg, minimum weight after joining the gym is 35 kg and maximum weight after joining the gym is 96 kg.

- ***Statistical Process Control Tools-***

a) **Check Sheet:**

It can be observed from our data that pimple is the major side effect of supplements.

b) **Pareto Diagram:**

It can be observed from our data that 20% of all the causes are responsible for 80% of the side effects. In our data the pimples are responsible for all the other causes.

- ***Testing of Hypothesis-***

a) **Normality:**

1) It can be observed from our data that the weight before joining the gym is not normally distributed.

2) It can be observed from our data that the weight after joining the gym is normally distributed.

b) **Sign paired test:**

It can be observed from our data that the weight after joining the gym is increasing after taking supplements.

c) **Proportion test:**

It can be observed from our data that there is no significant difference between the food habits of the people
of two professions.

d) **Chi-square test for independence of attributes:**

It can be observed from our data that the preference of protein take is independent on effects.

e) **Multiple regression:**

It can be observed from our data that correlation between diet, supplement and type of whey protein is 0.2122351. Therefore, we conclude that they are least correlated with each other.

- ***Data Analytics-***

a) **K-Nearest Neighbour:**

It can be observed from our data that the person having weight before joining the gym 63 and after joining the gym is 66 will prefer Isolate type of whey protein.

b) **Naïve- Bayes and Decision tree:**

It can be observed from our data that by both the classification algorithms the predicted protein type is Isolate. The accuracy with naïve bayes classifier is 70% and by decision tree is 65%. Therefore, for our data the Naïve Bayes classifier is superior to Decision Tree algorithm.

c) **Association Rules:**

It can be observed from our data that person buying whey protein may most likely to buy caffeine (coffee) and creatine.

LIMITATIONS

The limitations of our data are as follows:

1) As our data is not too large, as for such surveys, sample size should be very large.

2) The result would have been better if data from different cities are taken into consideration and compared.

3) The results may vary sample to sample.

4) Our data is heterogeneous so it may be unbiased in nature.

SUGGESTIONS

Some suggestions regarding our project are as follows:

1) According to analysis, people should work out maximum 1.5 to 2 hours a day.

2) People should sleep at least 7-8 hours for better health and proper muscle recovery.

3) People should intake limited amount of supplement (according to their body requirements) for their muscle building otherwise they would suffer from several deceases or effects like hair-fall, pimples, acnes, kidney issues and etc.

4) People should eat less junk foods to keep themselves fit and healthy.

5) People should spend money by managing their income for their supplements.

6) They should avoid dosage off steroids if they do not aim for bodybuilding or any high levels competitions.

BIBLIOGRAPHY

- **Books:**

  1) Statistical computing using R-Software.
  2) Sampling Distribution and exact tests (Nirali publication).
  3) Machine learning, McGraw-Hill (Mitchell T.M.)
  4) Morgan Gaufmann3rd Edition.
  5) Statistical process control (M. Mahajan)

- **Websites:**

  1) https://piktochart.com/graph-maker
  2) https://www.javatpoint.com/classification-algorithm-in-machine-learning
  3) https://www.bodybuilding.com