



The World Of Herbology



A Project Report on

PATANJALI-The World Of Herbology

Submitted By

Hetvi Gala

Mansi Makwana

Vaibhavi Oza

Akshay Barot

Towards the Partial Fulfillment for the Degree of

MASTER OF SCIENCE IN APPLIED STATISTICS

Under the Guidance of

K. J. SOMAIYA COLLEGE OF SCIENCE AND COMMERCE,

VIDYAVIHAR (E)

Re-accredited “A” Grade by NAAC Autonomous,

Affiliated to University of Mumbai

ACADEMIC YEAR 2020-2021



DECLARATION

We, Hetvi Gala, Mansi Makwana, Vaibhavi Oza and Akshay Barot, studying in the Second year of Master of Science in Applied Statistics course in the academic year 2020-2021 at K.J.Somaiya College Of Science & Commerce, hereby declare that we have completed the project titled, "PATANJALI-The World Of Herbology" as a part of the course requirements of Master in Applied Statistics.

We further declare that the information presented in this project is true and original to the best of our knowledge.

Date:

Place:

Name of the student:

Signature:



CERTIFICATE

I, , hereby certify that **Hetvi Gala, Mansi Makwana, Vaibhavi Oza and Akshay Barot** studying in the Second Year of Master of Science in Applied Statistics course at the K.J.Somaiya College Of Science and Commerce, have completed a project on "**PATANJALI-The World Of Herbology**", under my guidance in the academic year 2020-2021.

I further declare that the information presented in this project is true and original to the best of my knowledge.

Date:

Place:

Signature of-

(Project Incharge):



CERTIFICATE OF AUTHENTICATION

*This is to certify that the project entitled "**PATANJALI-The World Of Herbology**" is a bonafide work of **Hetvi Gala, Mansi Makwana, Vaibhavi Oza & Akshay Barot**, submitted to the K.J.Somaiya College of Science and Commerce in partial fulfillment of the requirement for the award of the degree of M.Sc. in the subject of Applied Statistics.*

Date:

Place:

Signature of-

(Project Incharge):

Mr. Prakash Shah (Head Of the Department):

External Invigilator:



Examiner Approval sheet

*This project report entitled “PATANJALI-The World Of Herbology” done by, **Hetvi Gala, Mansi Makwana, Vaibhavi Oza & Akshay Barot**, is approved for the degree of Master of Science in the subject of Applied Statistics.*

Date:

Place:

Name & Signature of-

External Invigilator:



ACKNOWLEDGEMENT

We would like to express our gratitude and sincere thanks to our Project Guide, for instilling confidence in us to carry out this study and extending valuable guidance and encouragement from time to time, without which it would not have been possible to undertake and complete this project.

*We also wish to extend my appreciation to our Head - Academics, Mr. **Prashant Shah** and the Program Coordinators, for their kind co-ordination and support.*

Also, our special thanks to the people of Mumbai for providing us with their vital information, opinions, valuable time & support.

INDEX

SR.NO	CONTENT	PAGE.NO
1	Executive summary & Abstract	9
2	Introduction & Need for study	10
3	Objectives of the study	12
4	Literature Review	13
5	Research Methodology & Sampling Technique	14
6	Graphical Representations	15
7	SWOT Analysis	26
8	Chi-Square Test of Independence	29
9	Logistic Regression	34
10	Artificial Neural Networks	44
11	K-Nearest Neighbors	50
12	Decision Trees	53
13	Market Basket Analysis using Apriori Algorithm	59
14	Pareto Analysis	64
15	Factor Analysis	68
16	Sentiment Analysis	75
17	Conclusion	81
18	Software's used & Codes	83
19	Questionnaire	90
20	References	96

EXECUTIVE SUMMARY & ABSTRACT

EXECUTIVE SUMMARY

Patanjali Ayurved Kendra Ltd was founded by Swami Ramdev and Acharya Balakrishna on September 27th, 2007. It started its operation in Katmandu, Nepal to provide holistic, natural and effective ayurvedic treatment. Patanjali Ayurved Limited (PAL) has three manufacturing units at Haridwar, Uttarakhand with its registered office in New Delhi. The initial project of Ramdev was his Patanjali Yogpeeth and since its inauguration in 2006, it has been labeled as one of the largest centers for research on Yoga and Ayurveda in the world.

ABSTRACT

The main aim of this paper is to find the consumer perception regarding the Patanjali brand through a survey in the city of Mumbai. The responses to our questionnaire suggested that the reasonable pricing, quality of products and good health benefits were the most important reasons for the popularity of Patanjali products. Further the study brought out the various factors as being the most important in affecting the decision to purchase Patanjali products such as: communication strategy, product quality, pricing and brand image of Baba Ramdev and consumer awareness about the products.

INTRODUCTION & NEED FOR STUDY

INTRODUCTION

The choice of a particular brand by the consumer over the time is mainly affected by the quality and benefits offered by the brand especially when it comes to the brand of eatables and cosmetics products. Consumer satisfaction is derived by comparing the actual performance with the expected performance of the product after usage. Perceptions are highly subjective and thus can easily distort. The confusion of buying and not buying continues into our mind unless our decision is not supported by many. Thus, in order to survive in the country like India, brands need to be positioned in the minds of people. India is known to be a hub of Herbal brands and Ayurveda as the herbal products are deeply associated with the spiritual sentiments of the people. One of the study of WHO (World Health Organization) shows that near about 82% of the world population depends upon the natural products rather than the other medicines due to their high cost and side effects. **Patanjali Ayurved**, is an Indian multinational consumer packaged goods company, based in Haridwar, India, that was started by **Ramdev** and **Balkrishna** in 2007. The company manufactures mineral and herbal products. “Patanjali” is represented by Guru Swami and the company is set up with a motive to provide high quality Indian products at fair price which are organic and natural.

Patanjali Food and Herbal Park at Haridwar is the main production facility operated by Patanjali Ayurved. The company has a production capacity of ₹35,000 crore and is in the process of expanding to a capacity of ₹60,000 crore through its new production units at several places, including Noida, Nagpur, and Indore. Patanjali Ayurved produces products in the categories of personal care and food. The company manufactures more than 2,500 products including 45 types of cosmetic products and 30 types of food products. Patanjali has also launched beauty and baby products. Patanjali Ayurvedic manufacturing division has over 300 medicines for curing a range of ailments and body conditions, from common cold to chronic diseases like paralysis. In 2016, Patanjali had announced to enter the textile manufacturing centre. The company is reported to manufacture not only traditional clothes such as Kurta-Pajama but also popular western clothes such as jeans. On 5th November 2016, Patanjali announced that it will set up a new manufacturing plant Patanjali Herbal and Mega Food Park in Balipara, Assam by investing ₹1,200 crore with the manufacturing capacity of 10 lakh tonnes of goods per year. The plant is the largest facility of Patanjali in India.

Patanjali Ayurved sells through nearly 4,700 retail outlets as of May 2016. Patanjali also sells its products online and is planning to open outlets at railway stations and airports. Patanjali Ayurveda products are also available in modern trade stores including Reliance retail, Hyper city and Star Bazaar apart from online channels. Patanjali Ayurved has also started its FMCG (Fast Moving Consumer Goods) expansion in form of dealership and distributorship channels across the country and expects wider growth in overseas distribution as well. Ram-Dev not only became a television celebrity teaching yoga but, is sitting on the fastest growing FMCG Company clocking a turnover close to Rs. 5000 Crore, more than Colgate, GlaxoSmithKline and Emami. All throughout this Ram-Dev never had his contempt for multinationals and vows to kill their market. He even has been discouraged of starting anti Coke and Maggie drive.

Patanjali as a brand established with the combination of both brand as well as personality Identity. In the beginning Baba Ram Dev has established himself as a Yoga Guru and posted Yoga as a solution to many health challenges which are prevalent because of western influence and food which made our lifestyle sedentary and unhealthy. When Baba Ram Dev established his own identity in customers mind he then used the interconnected Strategies to establish his brand identity in the Indian market. Baba Ram Dev has been successfully able to impact the minds of the people and the brand is doing really well in current times as people are choosing herbal products over any other products.

NEED FOR STUDY

This study will reveal the different aspects of consumer's perception regarding price, quality, range, availability, and advertisements of the products. The need for the study is very essential as the competition in the Ayurvedic and herbal products is ever increasing. Competitors are mainly struggling to shut down the market by capturing its market share. Today scenario is such that the competitors are coming up with sales promotion and incentives to compete with Patanjali brand.

OBJECTIVES

- 1 • To study the SWOT for the brand Patanjali.
- 2 • To test the association between gender and their health-conscious behavior while buying the Patanjali products.
- 3 • To study the influence of various factions on overall satisfaction of the products for Patanjali users.
- 4 • To predict if Non-Patanjali users will be overall satisfied after using Patanjali products.
- 5 • To predict how often customers use Patanjali products based on socio-demographic factors, sources of awareness & their buying and spending patterns.
- 6 • To identify next product purchase that might interest a customer.
- 7 • To study the factors that affect Patanjali from being a better brand.
- 8 • To determine the factors that influences the people for buying patanjali products.
- 9 • To analyze sentimental behaviour of the customers towards the brand Patanjali.

LITERATURE REVIEW

We carried a study on consumer perception regarding Patanjali product. The objective of the study were to know about the perception, satisfaction level, and attributes of consumers with regards to Patanjali products. The data was collected from 501 respondents of Mumbai and conclusions were drawn looking at the data.

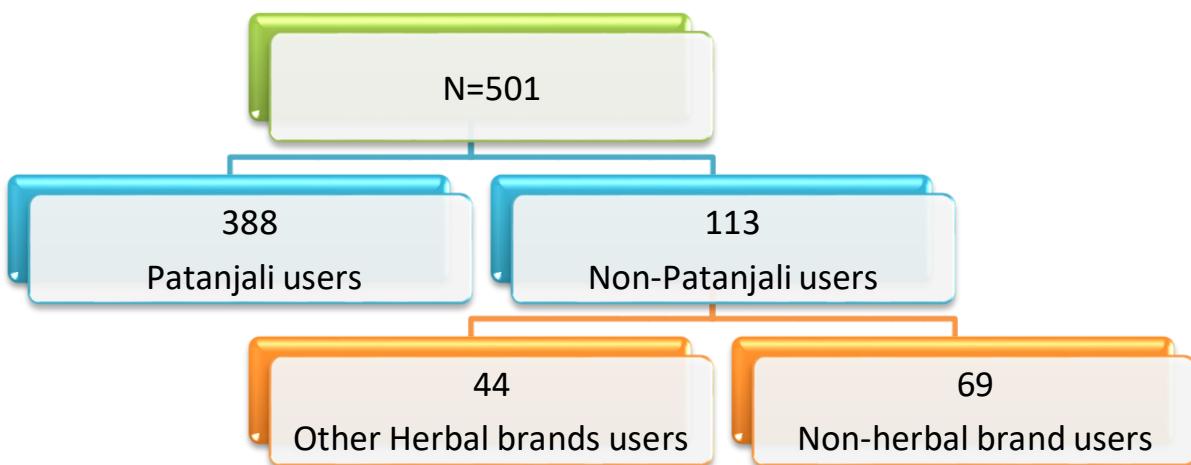
The questionnaire had questions for both, the one who had used Patanjali products and the one who had never used Patanjali products. Out of 501 respondents 388 had used the products while 113 had never used any Patanjali product. We performed analysis on the obtained data through techniques like SWOT analysis, Factor analysis, chi-square, Market Basket Analysis etc. Based on the outputs we made predictions about the brand. We checked for various associations of people with the product to draw relevant results. We checked for the preference of the consumer's and their needs while choosing the product. Through Sentimental analysis we got to know that people trust's the brand and hence chooses Patanjali products. K-NN was used to predict whether non – Patanjali users would be satisfied using Patanjali product. Decision Tree was used to predict how often customers would use Patanjali products. Market Basket Analysis was used to check the assosciation between the products to identify the next purchases of the customer. Pareto Analysis was done so as to understand where the brand needs to improve based on the consumer data.

The objectives of the study were to understand about the mindset of the people while they choose Patanjali products and to help the brand improvise where-ever necessary.

RESEARCH METHODOLOGY & SAMPLING TECHNIQUE

RESEARCH METHODOLOGY

For collecting primary data, a survey approach was used only for Mumbai city and for this purpose research questionnaire consisting of a set of questions was presented to the respondents to know their perception towards Patanjali ayurvedic products. We then conducted a pilot survey for a sample size of 50 personally & made the necessary changes. We surveyed a total of 501 individuals through Google forms and a personal survey for primary data. We entered the data into an excel sheet, cleaned our data & encoded our variables. Out of 501 users, we have considered the data for only 388 Patanjali users and 44 other herbal brand users.

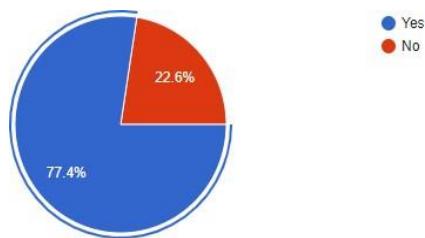


SAMPLING TECHNIQUE

In present study convenience sampling was used to collect the data from the defined universe. Convenience sampling(also known as availability sampling) is a specific type of non-probability sampling method that relies on data collection from population members who are conveniently available to participate in study. In other words, this sampling method involves getting participants from wherever you can find and typically wherever is convenient. In convenience sampling, no inclusion criteria are identified prior to the selection of subjects. Allsubjects are invited to participate.

GRAPHICAL REPRESENTATIONS

People who have used the Patanjali Products:

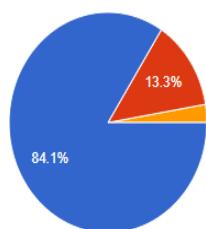


- Out of 501 people there are 388 people who have used Patanjali products and 113 people who have never used Patanjali products.

OTHER HERBAL BRANDS VS PATANJALI

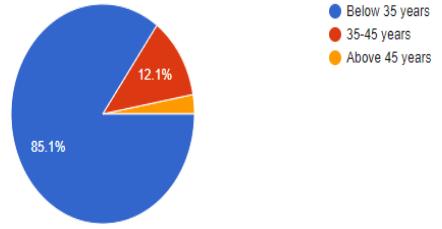
Age responses of the people:

Age
113 responses



Age
388 responses

● Below 35 years
● 35-45 years
● Above 45 years

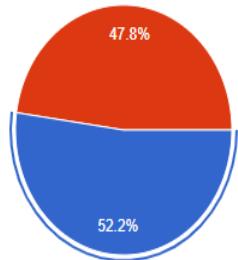


84.1% which means there are 95 people falling in the age group of below 35 years of age who have not used the products.	85.1% which means 330 people falling in the age group of below 35 years of age who have used the products.
13.3% which means there are 15 people falling in the age group of 35-45 years of age who have not used the products.	12.1% which means 47 people falling in the age group of 35-45 years of age who have used the products.
2.7% which means there are 3 people falling in the age group of above 45 years of age who have not used the products.	2.8% which means 11 people falling in the age group of above 45 years of age who have used the products.

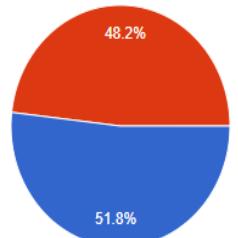
Gender responses of the people:

Gender

113 responses



Male
Female
Other



Male
Female
Other

There are 52.2% which accounts to 59 males who have not used the brand.

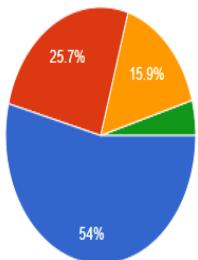
There are 47.8% which accounts to 54 females who have not used the brand.

There are 51.8% which accounts to 201 males who have used the brand.

There are 48.2% which accounts to 187 females who have used the brand.

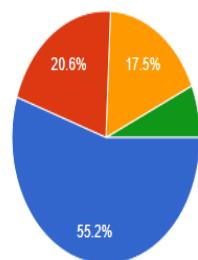
Income:

113 responses



Below 3
3-5
5-10
Above 10

388 responses



Below 3
3-5
5-10
Above 10

54% which means there are 61 people whose income is below 3 lakhs who have not used the brand.

25.7% which means there is 29 people whose income is between 3-5 lakhs who have not used the brand.

15.9% which means there is 18 people whose income is between 5-10 lakhs who have not used the brand.

4.4% which means there is 5 people whose income is above 10 lakhs who have not used the brand.

55% which means there are 214 people whose income is below 3 lakhs who have used the brand.

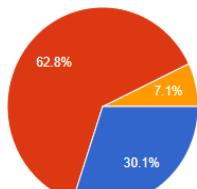
20.6% which means there is 80 people whose income is between 3-5 lakhs who have used the brand.

17.5% which means there is 68 people whose income is between 5-10 lakhs who have used the brand.

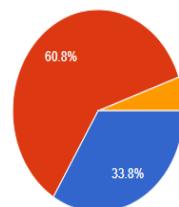
6.7% which means there is 26 people whose income is above 10 lakhs who have used the brand.

Location wise responses of people :

Location
113 responses

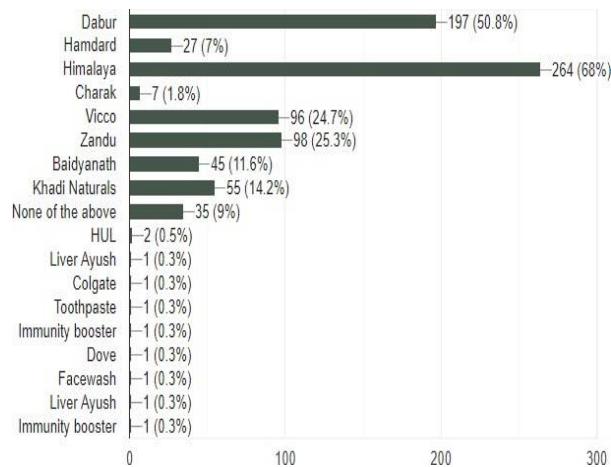
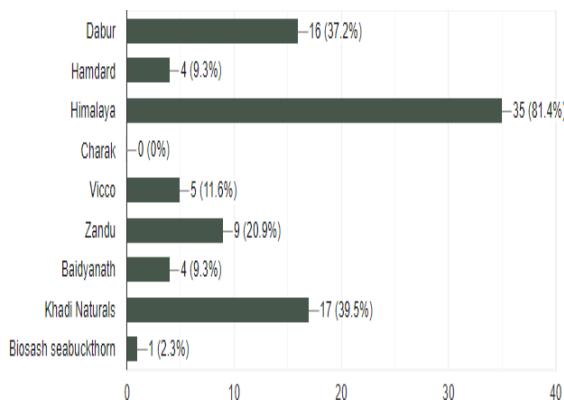


Location
388 responses



62.8% which means there are 71 people residing in western line that does not use Patanjali products.	60.8% which means there are 236 people residing in western line that use Patanjali products.
30.1% which means there are 34 people residing in central line that does not use Patanjali products.	33.8% which means there are 131 people residing in central line that use Patanjali products.
7.1% which means there are 8 people residing in harbour line that does not use Patanjali products.	5.4% which means there are 21 people residing in harbour line that use Patanjali products.

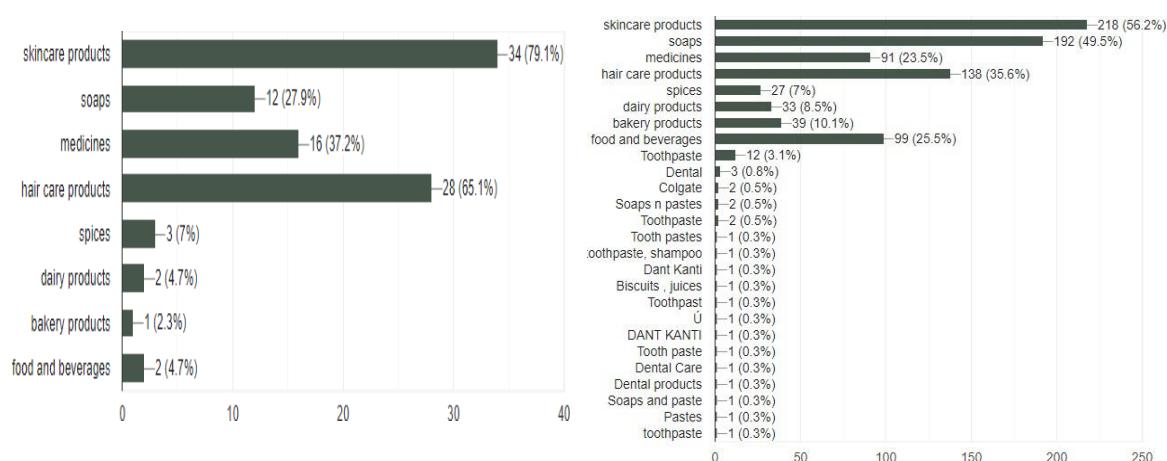
Herbal brands used by people:



Brands people use other than Patanjali	Brands people use along with Patanjali
81.4% people use Himalaya.	68% of the people use Himalaya along with Patanjali.
39.5% people use Khadi Naturals.	50.8% of the people use Dabur along with Patanjali.
37.2% people use Dabur.	25.3% of the people use Zandu along with Patanjali.
20.9% people use Zandu.	24.7% of the people use Vicco along with Patanjali.

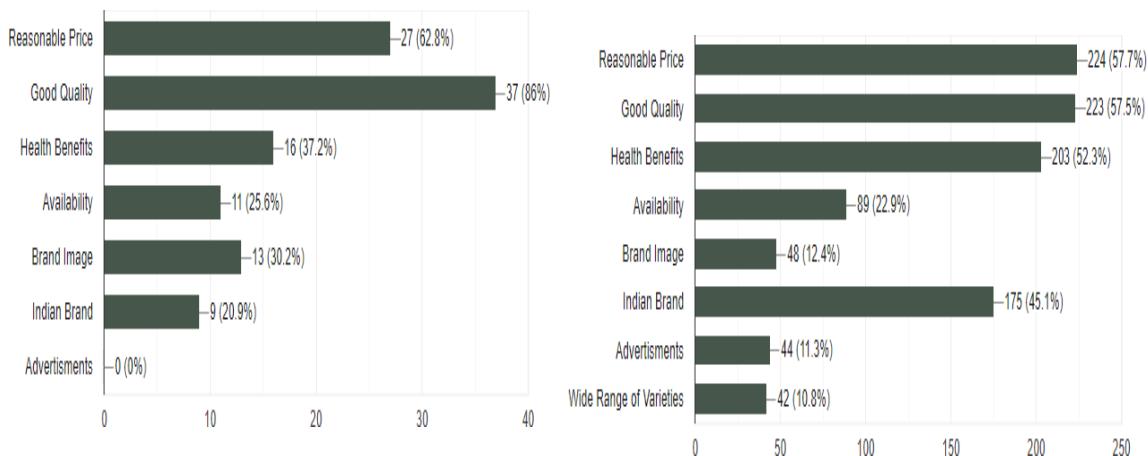
11.6% people use Vicco.	14.2% of the people use Khadi naturals along with Patanjali.
9.3% people use Baidyanath and Hamdard.	11.6% of the people use Baidyanath along with Patanjali.
2.3% people use Biosash seabuckthorn.	9% of the people use no other brand than Patanjali.
0% people use charak.	7% of the people use Hamdard along with Patanjali.
	1.8% of the people use Charak along with Patanjali.

Product Categories used by people:



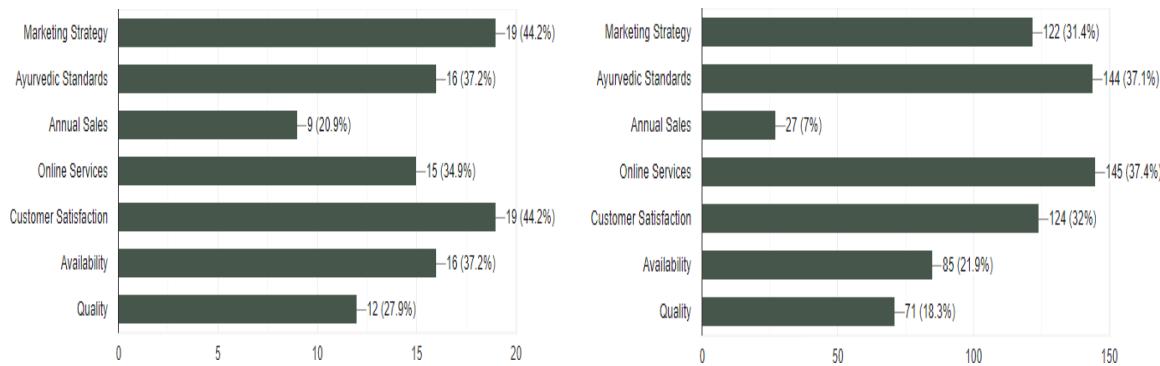
Other brand Categories	Patanjali brand categories
79.1% people use skincare products of other brands.	56.2% people use skincare products of Patanjali brand.
65.1% people use hair care products of other brands.	49.5% people uses soaps of Patanjali brand.
37.2% people uses medicines of other brands.	35.6% people use hair care products of Patanjali brand.
27.9% people uses soap of other brands.	25.5% people uses food and beverages of Patanjali.
7% people use spices of other brands.	23.5% people use medicines of Patanjali brand.
4.7% people use dairy and food and beverages of other brands.	10.1% people use bakery products of Patanjali brand.
2.3% people use bakery products of other brands.	8.5% people use dairy products of Patanjali brand.
	7% people use spices of Patanjali brand.

Reasons for Buying these products:



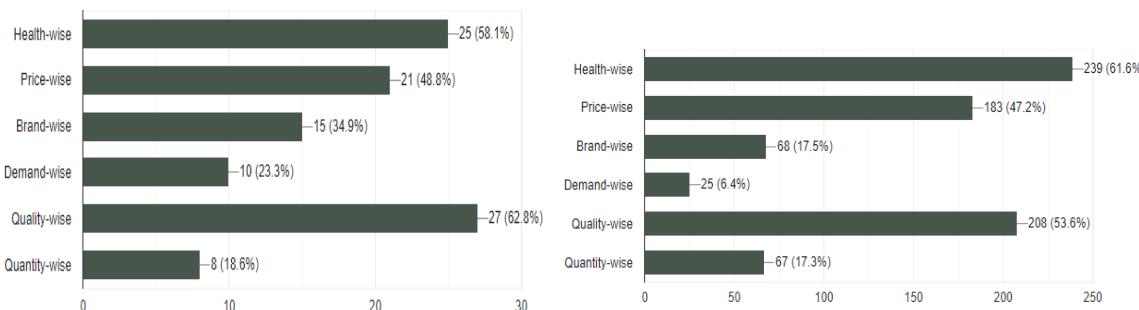
Reasons for buying other products	Reasons for buying Patanjali products
86% people buy because of Good Quality	57.7% people buy because of reasonable price.
62.8% people buy because of reasonable price.	57.5% people buy because of Good Quality.
37.2% people buy because of health benefits.	52.3% people buy because of health benefits.
30.2% people buy because of brand image.	45.1% people buy because of Indian Brand.
25.6% people buy because of Availability of the product.	22.9% people buy because of Availability of the product.
20.9% people buy because it is an Indian brand.	12.4% people buy because of brand image.
	11.3% people buy because of advertisements.
	10.8% people buy because of wide range of varieties.

Improvement of brand:



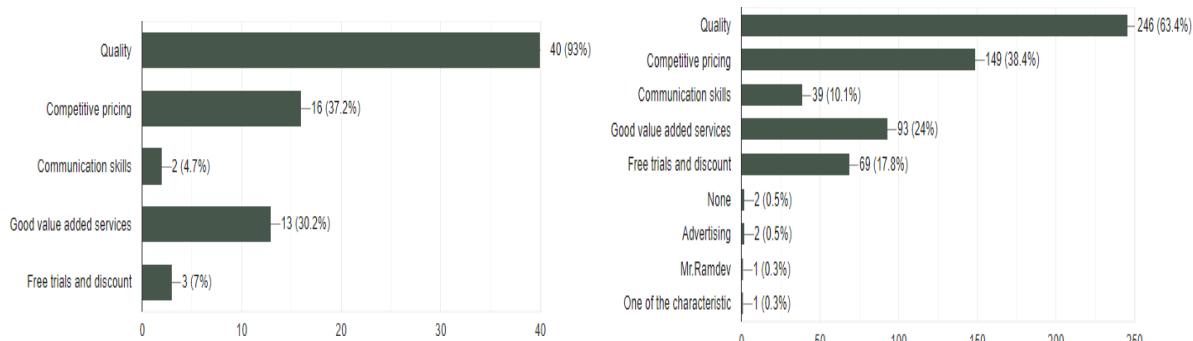
Categories of improvement in other brands	Categories of improvement in Patanjali Brand
44.2% people think the brand needs to improve in marketing strategy and customer satisfaction.	37.4% people think the brand needs to improve in online services.
37.2% people think the brand needs to improve in Ayurvedic standards and availability of product.	37.1% people think the brand needs to improve in Ayurvedic standards.
34.9% people think the brand needs to improve in online services.	32.1% people think the brand needs to improve in customer satisfaction.
27.9% people think the brand needs to improve in Quality.	31.4% people think the brand needs to improve in marketing strategy.
20.9% people think the brand needs to improve in Annual sales.	21.9% people think the brand needs to improve in the availability of the product.
	18.3% people think the brand needs to improve in Quality.
	7% people think the brand needs to improve in Annual sales.

Benefits of the product:



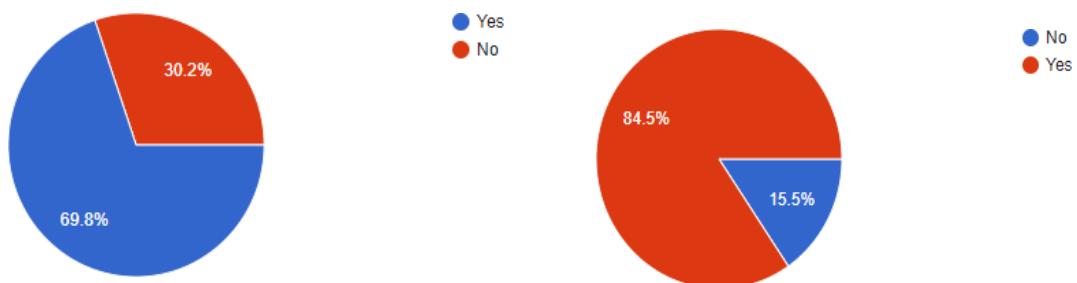
Benefits of using other products	Benefits of using Patanjali products
62.8% people think the product is beneficial quality wise.	61.6% people think the product is beneficial health wise.
58.1% people think the product is beneficial health-wise.	53.6% people think the product is beneficial quality wise.
48.8% people think the product is beneficial price wise.	47.2% people think the product is beneficial price wise.
34.9% people think the product is beneficial brand wise.	17.5% people think the product is beneficial brand wise.
23.3% people think the product is beneficial demand wise.	17.3% people think the product is beneficial quantity wise.
18.6% people think the product is beneficial quantity wise.	6.4% people think the product is beneficial demand wise.

Better brand:



Other brands a better brand	Patanjali a better brand
93% people think quality makes it a better brand.	63.4% people think quality makes it a better brand.
37.2% people think competitive pricing makes it a better brand.	38.4% people think competitive pricing makes it a better brand.
30.2% people think good value added services makes it a better brand.	24% people think good value added services makes it a better brand.
7% people think free trials and discount makes it a better brand.	17.8% people think free trials and discount makes it a better brand.
4.7% people think communication skills makes it a better brand.	10.1% people think communication skills makes it a better brand.

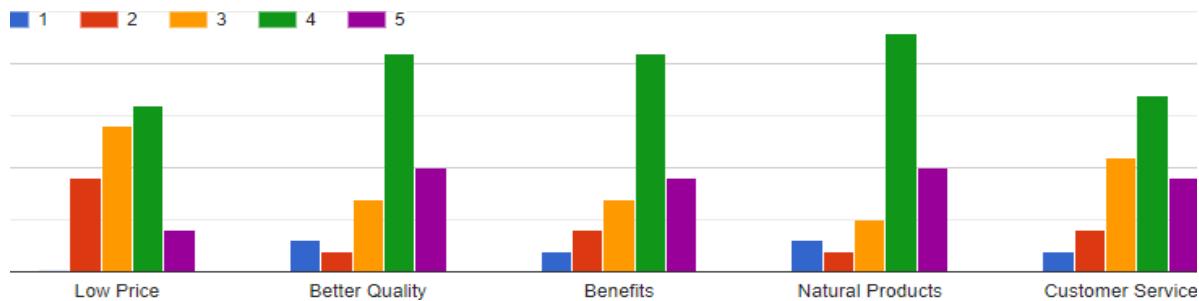
Overall Satisfaction of the brand:



Overall Satisfaction from other brands	Overall Satisfaction from Patanjali brand
69.8% people are satisfied using other brands.	84.5% people are satisfied using Patanjali brand.
30.2% people are not satisfied using other brands.	15.5% people are not satisfied using Patanjali brand.

Satisfaction from following characteristics of other brands:

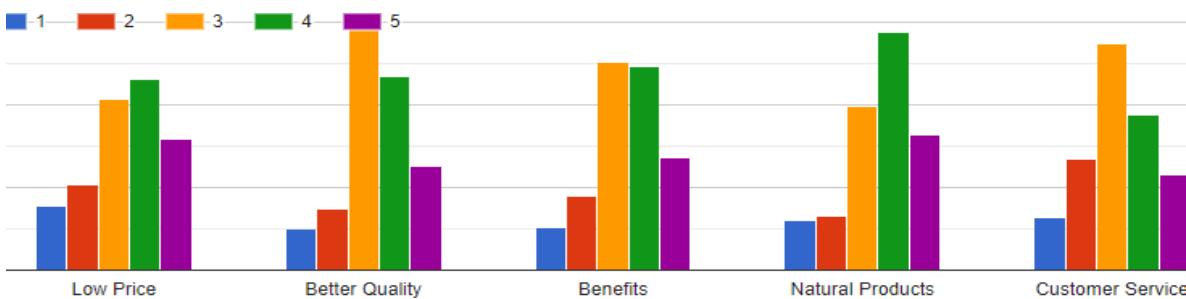
(1-Strongly Disagree,2-Disagree,3-Neutral, 4-Agree,5- Strongly Agree)



- People agrees buying product with low price, better quality, benefits, natural products and customer service.

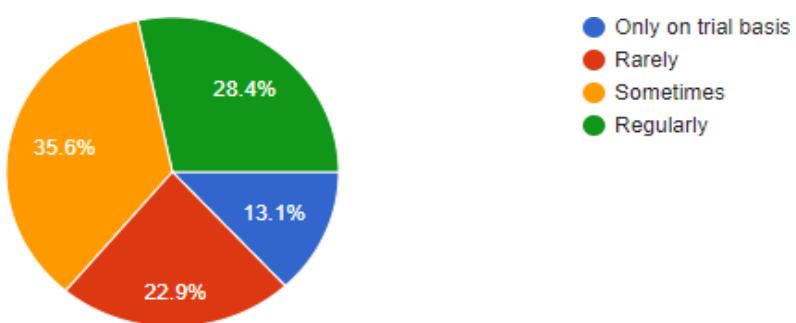
Satisfaction from following characteristics of Patanjali brand:

(1-Strongly Disagree,2-Disagree,3-Neutral, 4-Agree,5- Strongly Agree)



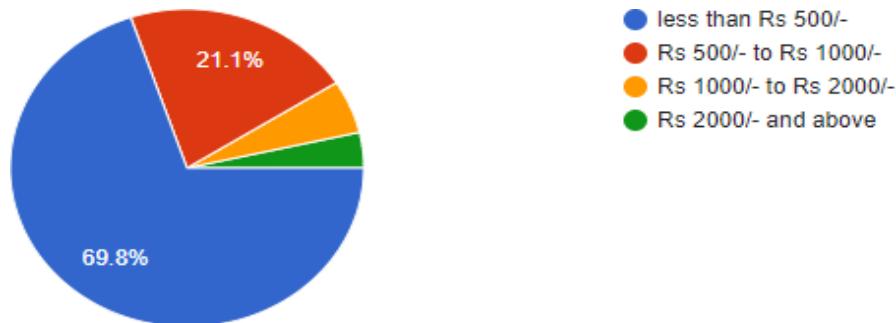
- People agree buying Patanjali product with low price and the product being natural.
- People are neutral towards the quality, benefits and customer service of the product.

Frequency of usage of Patanjali Products:



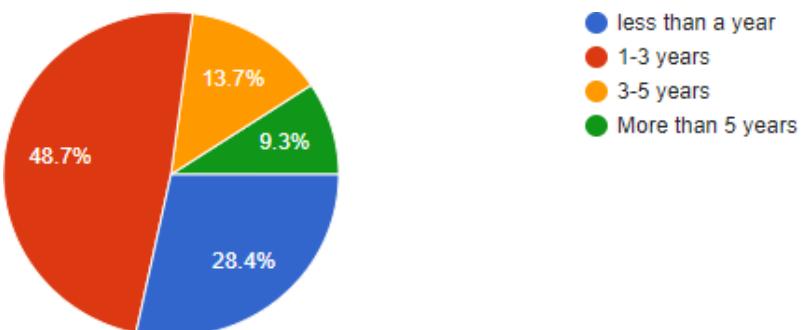
- 35.6% people uses Patanjali products sometimes and 13.1% people uses Patanjali product only on trial basis.

Monthly Spent on Patanjali Product:



- 69.8% people spend less than 500/- monthly on Patanjali products whereas 3.6% people spend more than 2000/- monthly on Patanjali products.

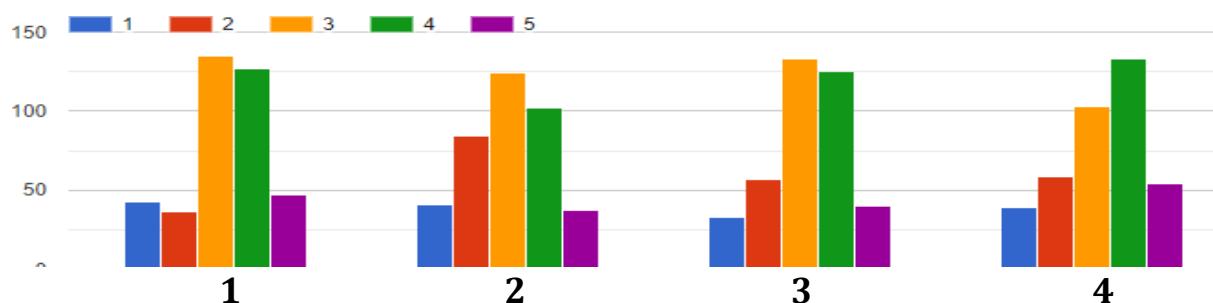
Years of usage of Patanjali Products:



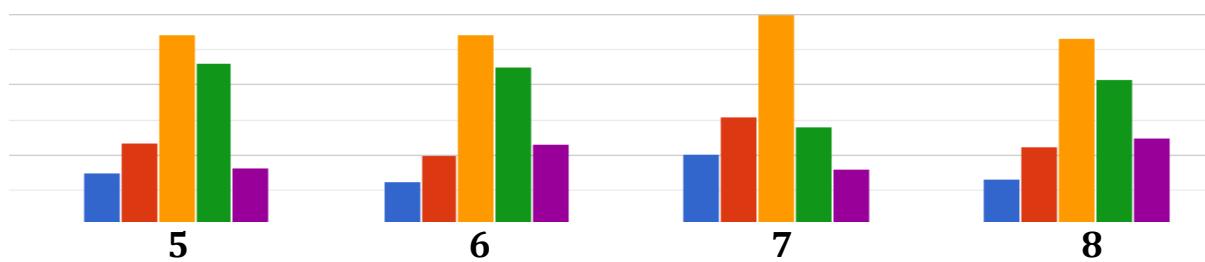
- 48.7% people are using Patanjali from 1-3 years and 9.3% people are using Patanjali for more than 5 years.

Opinion on Patanjali Products:

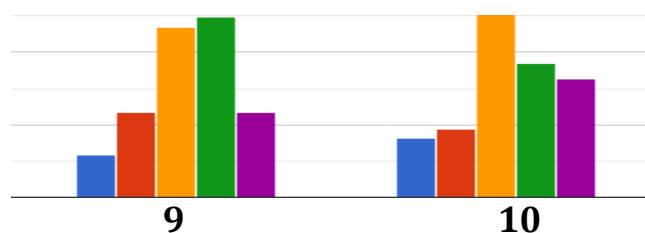
(1-Strongly Disagree,2-Disagree,3-Neutral, 4-Agree,5- Strongly Agree)



- The first graph is of Patanjali products gaining popularity in the market where people have a neutral opinion towards it.
- The second graph is about the Advertisements of Patanjali products that are more visible in recent times in which people have neutral opinion towards it.
- The third graph is on Patanjali gaining reputation due to superior quality products and benefits towards which people have neutral opinion.
- The fourth graph is on whether Patanjali products are easily available at better place to which people have agreed.

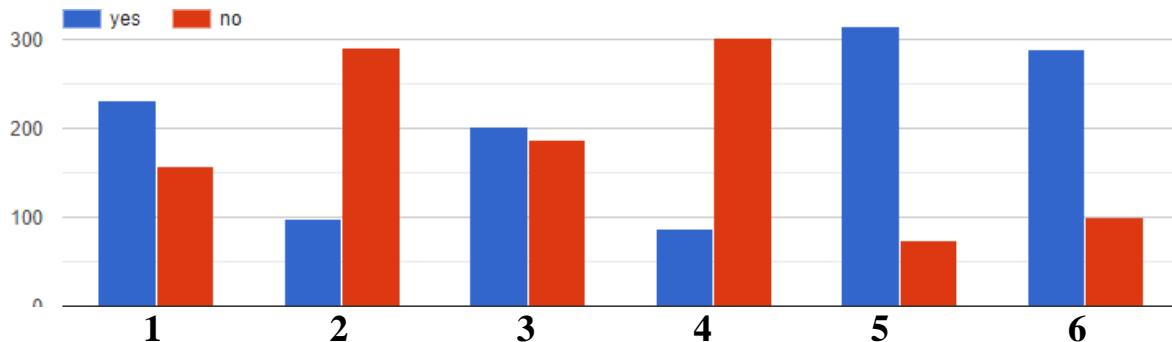


- The fifth graph is about whether Patanjali products symbolize purity to which people have a neutral opinion.
- The sixth graph is about whether Patanjali products are safe as they are made from natural ingredients to which people have neutral opinions.
- The seventh graph is about whether Patanjali products are appealing to youngsters to which people have a neutral opinion.
- The eighth graph is about whether Patanjali should increase the promotion and advertisements of their products to which people have a neutral opinion.



- The Ninth graph is about whether Patanjali has emerged as a powerful brand in terms of purity, prices and quality to which people have agreed.
- The tenth graph is about whether a Brand ambassador affects the popularity of the products to which people have a neutral opinion.

Yes/No Factors:



- The first graph is on would people consider Patanjali over any other competitive brands to which people have majorly answered yes.
- The second graph is about, have they ever complained about the Patanjali product to which people have majorly answered no.
- The third graph is about whether people prefer doctor's recommendation while choosing the product to which people have majorly answered yes.
- The fourth graph is about if they have ever experienced any side effects while using the products to which people have majorly answered no.
- The fifth graph is about whether Patanjali products have gained reputation due to the popularity of Baba Ramdev to which people have majorly answered yes.
- The sixth graph is about whether they would recommend their friends to use the Patanjali product to which people have majorly answered yes.

SWOT ANALYSIS

OBJECTIVE 1

To study the internal and external environments of a company, through the identification and analysis of the strengths and weaknesses of the organization, and the opportunities and threats to which it is exposed.

SWOT Analysis is a simple but useful framework for analyzing your organization's strengths, weaknesses, opportunities, and threats (SWOT)



STRENGTHS

1. **Baba Ramdev**: The fan following and goodwill of Baba Ramdev guaranteed that Patanjali grows quickly and becomes a routine name in the Indian households.
2. **Strong patriotism**: Patanjali has used the Indian card to its advantage and has always marketed that it's a brand made in India, for Indians. Most of the brands in India are international brand. Patanjali actively asks Indians to buy India made products to help the economy of the country.
3. **Comparatively Low Prices**: Patanjali products are generally priced at 20-30 % lower than the competitive brands and thus it becomes impossible for the competitive brands to compete with Patanjali. The company sources the products directly from farmers and hence cuts on middlemen.
4. **Ayurveda and Herbal**: The Products that Patanjali offers are made from Ayurveda and Herbal natural components. The Swadeshi products also have played an important role in the success of Patanjali. India has never been low on plants or vegetation and we get a lot of naturally grown medicines in our dense forests.

WEAKNESSES

1. **Over dependency on Ramdev**: For many of its consumers, Patanjali is still synonymous to Baba Ramdev and hence any actions of Baba Ramdev will have repercussions on the brand itself. Baba Ramdev's political affiliations are also well known and hence if at all he is targeted for any political vendetta, Patanjali will also suffer.
2. **A low number of manufacturing units**: Patanjali would need to set up manufacturing units in different parts of the country which would require heavy investments. It also would have to move from the word-of-mouth strategy to nationwide promotional campaigns.
3. **Penetration pricing is not long term**: Patanjali might have to compromise on its pricing strategies if it wants to expand and thus it's a big challenge for Patanjali. It cannot sell at such low costs for a very long term. Any company needs profits to drive more sales and therefore earn more profits. It's a cycle. But if Patanjali does not earn much, then it cannot spend much and cannot expand.
4. **Product Dependence**: While Patanjali has many products in its kitty but a major part of its revenues is dependent on 5-6 of its main products such as its toothpaste and shampoo. They need to push the other products more to achieve its ambitious target.
5. **Lack of experienced management graduates**: Patanjali does not have a large pool of management graduates and thinks tanks which can be a problem when they look for expansion throughout the country or globally.

OPPORTUNITIES

1. **Growing organic sector**: Patanjali has been successful in creating awareness about the benefits of using herbal and natural products which have created a market for itself. The awareness has been spreading and the demand is ever growing.
2. **Expand Rural**: With the portfolio of products that Patanjali has, it has great potential in the rural market and should look to expand its operations in the vast rural market of India.
3. **Going Global**: Patanjali has a great opportunity to expand globally and can look for Middle East and African nation in the beginning. Various other companies such as Dabur have already expanded globally and have been successful.
4. **Tie ups**: Patanjali has successfully tied up with Future group and should continue tie up with modern retail chains and increase its E-commerce sales.

5. **Diversify:** While Patanjali is now present in retail products, it has not entered clothing which is another area where competitors like Reliance and Aditya Birla have expanded successfully. So Patanjali can plan on diversifying its product portfolio even further to Khadi, making it a fashion statement and being true to the roots of Patanjali being an Indian brand.

THREATS

1. **Price war:** A price war is good for consumers but it is detrimental for business. A price war will have drastic effect on Patanjali's profitability, especially because the brand is already selling at very low margins.
2. **Increasing Competition:** FMCG majors such as HUL, Marico, etc. and new entrants such as Sri Ayurveda are also entering the organic market after the awareness created by Patanjali which increases the competition in the market.
3. **Negative Word-of-Mouth:** Any negative word-of-mouth created on social media platforms can affect its position in the market.
4. **Poor reap can affect business:** Patanjali is heavily dependent on natural ingredients and products and hence poor agricultural reap can affect its sales.

CHI SQUARE TEST OF INDEPENDENCE

OBJECTIVE 2

To test the association between gender and their health-conscious behavior towards the Patanjali products.

Chi Square test of independence measures whether there is a relationship between two categorical variables. The Chi Square statistic is a non-parametric tool designed to analyze group differences when the dependent variable is measured at nominal level. It does not require equality of variances among the study groups or Homoscedasticity in the data. Chi Square is robust with respect to the distribution of the data. Unlike most statistics, the Chi Square can provide information not only on the significance of any observed differences, but also provides detailed information on exactly which categories account for any differences found.

PROS

1. It is easier to compute.
2. It identifies the difference between observed and expected values.
3. It does not assume anything about the data distribution.

CONS

1. The number of observations should be more than 20.
2. It can't use percentages.
3. Data must be frequency data.
4. It is sensitive to small frequencies (below 5) which leads to erroneous conclusions.

ASSUMPTIONS

1. A sample with sufficiently large size is assumed.
2. The observations are always assumed to be independent of each other.
3. The categories are mutually exclusive i.e., each subject should fit in only one category.
4. It assumes that the data for the study is randomly picked from the population.



VARIABLES CONSIDERED

GENDER	
Female	0
Male	1
DOCTOR'S RECOMMENDATION	
No	0
Yes	1
SIDE EFFECTS	
No	0
Yes	1

STEPS TO IMPLEMENT FOR CHI-SQUARE TEST OF INDEPENDENCE

- 1 •SPSS
- 2 •Import the table
- 3 •Weight cases-Frequency
- 4 •Analyze
- 5 •Descriptive Statistics-Obtain the crosstabs
- 6 •Enter the variables in rows and variables in columns, then go to Statistics
- 7 •Perform Chi Square

DATA ANALYSIS & INTERPRETATION

Here, we consider two sub-objectives for further analysis.

1. To test the association between gender and their consideration over doctor's recommendation while buying the Patanjali products.
2. To test the association between gender and their experience over any side effects while using the Patanjali products.

HYPOTHESIS

H₀₁: There is no significant association between the gender and their consideration over doctor's recommendation.

H₁₁: Not H₀₁

H₀₂: There is no significant association between gender and their experience over any side effects.

H₁₂: Not H₀₂

TEST STATISTIC

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Where,

O_i = Observed frequency

E_i = Expected frequency

D.F: (r-1) (c-1)

Where, r = Number of rows; c = Number of columns

DECISION CRITERIA(For $\alpha = 0.05$)

If p value is less than 0.05 then we reject null hypothesis & conclude that there is association between two variables, i.e., they are dependent on each other.

CASE 1**Gender * Doctor's recommendation Cross tabulation**

		Doctor's recommendation		Total
		Yes	No	
Gender	Count		90	99
	MALE	Expected Count	97.5	91.5
		% within Gender	47.6%	52.4%
	Count		105	84
	FEMALE	Expected Count	97.5	91.5
		% within Gender	55.6%	44.4%
	Count		195	183
	Total	Expected Count	195.0	183.0
		% within Gender	51.6%	48.4%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	2.383 ^a	1	.123		
Continuity Correction ^b	2.076	1	.150		
Likelihood Ratio	2.386	1	.122		
Fisher's Exact Test				.150	.075
Linear-by-Linear Association	2.377	1	.123		
N of Valid Cases	378				

p-value(0.123)>0.05Thus, we fail to reject H_0 .

Hence, we conclude that there is no association between gender & doctor's recommendation while buying the products, i.e., they are independent of each other.

CASE 2**Gender * side effects Cross tabulation**

		Side effects		Total
		Yes	No	
Gender	Count		41	148
	MALE	Expected Count	42.5	146.5
	% within Gender		21.7%	78.3%
	Count		44	145
	FEMALE	Expected Count	42.5	146.5
	% within Gender		23.3%	76.7%
	Count		85	293
	Total	Expected Count	85.0	293.0
	% within Gender		22.5%	77.5%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.137 ^a	1	.712		
Continuity Correction ^b	.061	1	.805		
Likelihood Ratio	.137	1	.712		
Fisher's Exact Test				.805	.403
Linear-by-Linear Association	.136	1	.712		
N of Valid Cases	378				

p-value(0.712)>0.05Thus, we fail to reject H_0 .

Hence, we conclude that there is no association between gender & their experience over any side effects while using the products, i.e., they are independent of each other.

BINARY LOGISTIC REGRESSION

OBJECTIVE 3

To study the influence of various factors on overall satisfaction for the Patanjali users.

Logistic regression is an extension of simple linear regression. Where the dependent variable is dichotomous or binary in nature, we cannot use simple linear regression. Logistic regression is the statistical technique used to predict the relationship between predictors (our independent variables) and a predicted variable (the dependent variable) where the dependent variable is binary (e.g., sex [male vs. female], response [yes vs. no], score [high vs. low], etc...). There must be two or more independent variables, or predictors, for a logistic regression. The IVs, or predictors, can be continuous (interval/ratio) or categorical (ordinal/nominal). All predictor variables are tested in one block to assess their predictive ability while controlling for the effects of other predictors in the model.

PROS

1. Convenient probability scores for observations.
2. Efficient implementations available across tools.
3. Multi-collinearity is not really an issue and can be countered with L2 regularization to an extent.
4. Wide spread industry comfort for logistic regression solutions.

CONS

1. Doesn't perform well when feature space is too large.
2. Doesn't handle large number of categorical features/variables well.
3. Relies on transformations for non-linear features.
4. Relies on entire data.

ASSUMPTIONS

1. Linearity of continuous independent variables & log odds.
2. Dependent variable should be binary.
3. Observations should be independent of each other.
4. There should be no multicollinearity.
5. Large sample size.

LOGISTIC MODEL

$$\pi(x) = \frac{e^{\beta_0 + \sum_{i=1}^p \beta_i x_i}}{1 + e^{\beta_0 + \sum_{i=1}^p \beta_i x_i}}$$

Where,

$\pi(x)$: Conditional probability that the outcome is Yes,

Y: Response variable.

X: Vector of independent variables.

We use the transformation called logit, which forces the prediction equation to predict values between 0 and 1.

LOGIT TRANSFORMATION OF ABOVE MODEL

$$g(x) = \beta_0 + \sum \beta_{1i} x_{1i} + \sum \beta_{2i} x_{2i} + \dots + \sum \beta_{pi} x_{pi}$$

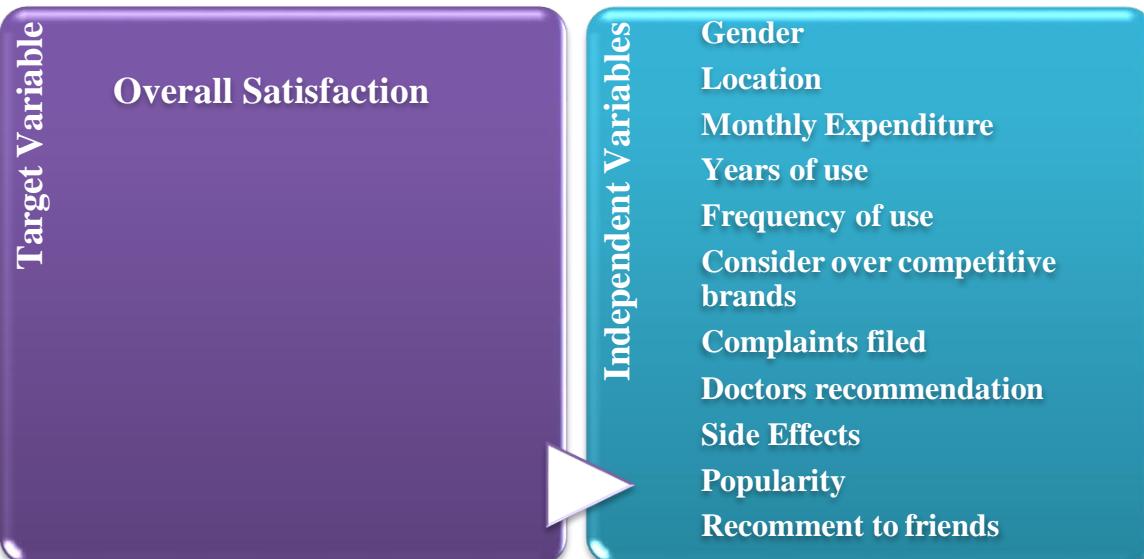
Where,

$g(x)$: Logit transformation of the probability of the event

β_0 : Intercept of the regression variables

β_{ii} : Slope of the i^{th} regression line

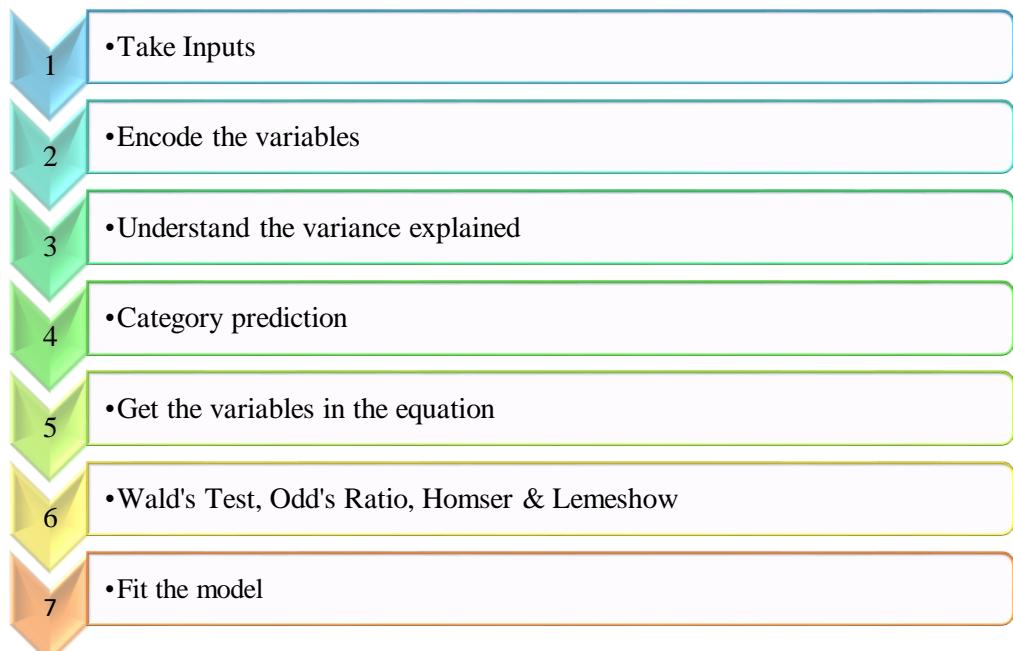
VARIABLE IDENTIFICATION



VARIABLES ENCODED

GENDER		
Female		1
Male		2
LOCATION		
Central Line		1
Harbour Line		2
Western Line		3
MONTHLY EXPENDITURE		
less than Rs 500/-		1
Rs 1000/- to Rs 2000/-		2
Rs 2000/- and above		3
Rs 500/- to Rs 1000/-		4
YEARS OF USE		
1-3 years		1
3-5 years		2
Less than a year		3
More than 5 years		4
FREQUENCY OF USE		
Only on trial basis		1
Rarely		2
Regularly		3
Sometimes		4
CONSIDER OVER COMPETITIVE BRANDS		
No		1
Yes		2
COMPLAINTS FILED		
No		1
Yes		2
DOCTORS RECOMMENDATION		
No		1
Yes		2
SIDE EFFECTS		
No		1
Yes		2
POPULARITY		
No		1
Yes		2
RECOMMEND TO FRIENDS		
No		1
Yes		2
OVERALL SATISFACTION		
No		0
Yes		1

STEPS TO BE IMPLEMENTED IN LOGISTIC REGRESSION



CHECKING THE ASSUMPTIONS

Multicollinearity

Coefficients		Collinearity Statistics	
Model		Tolerance	VIF
		.962	1.039
1	gender	.964	1.038
	location	.927	1.078
	Years of use	.772	1.295
	Frequency of use	.866	1.154
	Monthly Expenditure	.851	1.175
	Consider over competitive brands	.860	1.163
	Complaints filed	.922	1.084
	Doctors Recommendation	.758	1.320
	Side Effects	.937	1.067
	Popularity	.686	1.458

Since, all the VIF's are less than 10, we can conclude that there is no multicollinearity present between the independent variables.

DATA ANALYSIS & INTERPRETATION

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	140.593*	.393	.680

This table contains the Cox & Snell R Square and Nagelkerke R Square values, where in both are the methods of calculating the explained variation. Therfore, the explained variation in the dependent variable based on our model ranges from 39.3% to 68%.

		Observed	Predicted		Percentage Correct
			Overall, are you satisfied with the Patanjali products?	No	
Step 1	Overall, are you satisfied with the Patanjali products?	Yes		33	27
	Overall Percentage			14	314
					55.0
					95.7
					89.4

a. The cut value is .500

"The cut value is 500". This means that if the probability of a case being classified into the "yes" category is greater than 500, then that particular case is classified into the "yes" category. Otherwise, the case is classified in the "no" category. Whilst the classification table appears to be very simple, it actually provides a lot of important information about your binomial logistic regression result, including:

- The **percentage accuracy in classification (PAC)**, which reflects the percentage of cases that can be correctly classified as "no" with the independent variables added (not just the overall model).
- **Sensitivity**, which is the percentage of cases that had the observed characteristic (e.g., "yes" for overall satisfaction) which were correctly predicted by the model (i.e., true positives).
- **Specificity**, which is the percentage of cases that did not have the observed characteristic (e.g., "no" for overall satisfaction) and were also correctly predicted as not having the observed characteristic (i.e., true negatives).

- The **positive predictive value**, which is the percentage of correctly predicted cases "with" the observed characteristic compared to the total number of cases predicted as having the characteristic.
- The **negative predictive value**, which is the percentage of correctly predicted cases "without" the observed characteristic compared to the total number of cases predicted as not having the characteristic.

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	% C.I. for EXP(B)	
							Lower	Upper
	-1.395	.501	7.756	1	.005	.248	.093	.662
			3.004	2	.223			
	new_location							
	new_location(1)	19.565	7057.631	.000	1	.998	314108944.867	.000
	new_location(2)	.856	.494	3.004	1	.083	2.354	.894
	new_term			13.118	3	.004		6.196
	new_term(1)	-1.404	1.139	1.521	1	.217	.246	.026
	new_term(2)	-1.668	.563	8.764	1	.003	.189	.063
	new_term(3)	-3.262	1.028	10.075	1	.002	.038	.005
	new_use			7.771	3	.051		.287
	new_use(1)	1.069	.626	2.914	1	.088	2.913	.853
	new_use(2)	19.803	3465.637	.000	1	.995	398314346.688	.000
Step 1*	new_use(3)	1.821	.653	7.769	1	.005	6.176	1.717
	new_monthlyspend			.294	3	.961		22.219
	new_monthlyspend(1)	-.859	1.636	.276	1	.600	.424	.017
	new_monthlyspend(2)	1.096	9263.596	.000	1	1.000	2.993	.000
	new_monthlyspend(3)	-.161	.769	.044	1	.834	.851	.189
	competitive_brand(1)	1.779	.533	11.133	1	.001	5.922	2.083
	complaints(1)	-.467	.518	.811	1	.368	.627	.227
	doctor_rec(1)	-.201	.503	.160	1	.689	.818	.305
	side_effects(1)	-.213	.522	.167	1	.683	.808	.291
	popularity(1)	-1.676	.794	4.462	1	.035	.187	.039
	recommendation(1)	3.185	.597	28.499	1	.000	24.166	7.505
	Constant	1.258	1.074	1.371	1	.242	3.519	77.809

The statistical significance of the test is found in the "Sig." column. From these results, you can see that variables with $p < 0.05$ added significantly to the model/prediction, but variables with $p > 0.05$, did not add significantly to the model.

WALD'S TEST**Hypothesis:**

$$\mathbf{H_0} : \beta_i = 0 ; i=1,2,3,4,5$$

$$\mathbf{H_1} : \beta_i \neq 0 ; i=1,2,3,4,5$$

Test Statistic: Under H_0 , the following test statistic follows the standard normal distribution.

$$W = \frac{\beta_i}{S.E(\beta ii)} \sim (0,1)$$

DECISION CRITERIA: Reject H_0 when p-value < 0.05

Variable	p-value
Gender	0.005
Location	0.223
Years of use(new_term)	0.004
Frequency of use(new_use)	0.051
Monthly Expenditure	0.951
Consider over Competitive brands	0.001
Complaints filed	0.368
Doctors recommendation	0.689
Side Effects	0.683
Popularity	0.035
Recommend to a friend	0

Hence, our influential variables are Gender, Years of use, Frequency of use, Consider over Competitive brands, Popularity and Recommendation to a friend.

ODD'S RATIO

Variables	ODD'S RATIO
new_gender(1)	0.248
new_term(1)	0.246
new_term(2)	0.189
new_term(3)	0.038
new_use(1)	2.193
new_use(2)	398314346.688
new_use(3)	6.167
Competitive_brand(1)	5.922
Popularity(1)	0.187
Recommendation(1)	24.166

Significant variables:

1. Gender

- Males are 0.248 times less likely to be overall satisfied than females.

2. Years of use

- Customers who use Patanjali products since ‘3-5 years’ are 0.246 times less likely to be overall satisfied than ‘1-3 years’.
- Customers who use Patanjali products since ‘less than a year’ are 0.189 times less likely to be overall satisfied than ‘1-3 years’.
- Customers who use Patanjali products since ‘more than 5 years’ are 0.038 times less likely to be overall satisfied than ‘1-3 years’.

3. Frequency of use

- Customers who use Patanjali products ‘rarely’ are 2.193 times more likely to be overall satisfied than users who use it on trial basis.
- Customers who use Patanjali products ‘regularly’ are 398314346.688 times more likely to be overall satisfied than users who use it ‘on trial basis’.
- Customers who use Patanjali products ‘sometimes’ are 6.167 times more likely to be overall satisfied than users who use it ‘on trial basis’.

4. Consider over Competitive brands

- Customers who prefer considering other competitive brands are 5.922 more likely to be overall satisfied than customers not considering other brands.

5. Popularity

- Customers who consider the popularity of Patanjali products gained due to Ramdev Baba are 0.187 less likely to be overall satisfied than customers not considering it.

6. Recommend to friends

- Customers who will recommend Patanjali products to their friends are 24.166 times more likely to be overall satisfied than customers not recommending it.

HOMSER & LEMESHOW (TEST FOR GOODNESS OF FIT)

H_0 : Fitted model is a good fit

H_1 : Fitted model is not a good fit

DECISION CRITERIA: Reject H_0 when p-value < 0.05

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	9.448	8	.306

Since p-value > 0.05, we fail to reject H_0 and conclude that our model is a good fit.

CONFUSION MATRIX

Actual	Predicted	
	No	Yes
No	33	27
Yes	14	314

$$\text{Accuracy} = (\text{TN} + \text{TP}) / (\text{TN} + \text{TP} + \text{FN} + \text{FP})$$

Accuracy: 89.4%

FITTED MODEL

$$g(x) = 1.258 - 1.668 * \text{new_term}(2) - 3.262 * \text{new_term}(3) + 1.821 * \text{new_use}(3) + 1.779 * (\text{competitive_brand}(1) - 1.676 * \text{popularity}(1) + 3.185 * \text{recommendation}(1)$$

new_term(2)	Less than a year
new_term(3)	More than 5 years
new_use(3)	Sometimes
competitive_brand(1)	Yes
popularity(1)	Yes
Recommendation(1)	Yes

Hence, Years of use(less than a year & more than 5 years),frequency of use(sometimes), consider over competitive brand(yes), popularity(yes) and recommendation to a friend(yes) are the factors that significantly influence the overall satisfaction of Patanjali users.

ARTIFICIAL NEURAL NETWORK & K-NEAREST NEIGHBORS

OBJECTIVE 4

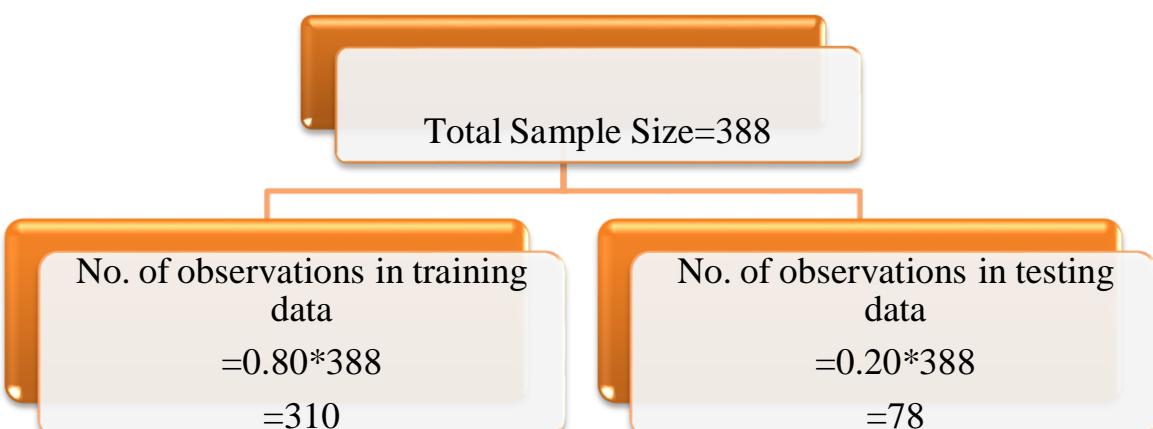
To predict if the Non-Patanjali users will be overall satisfied after using Patanjali products.

Here, two techniques (KNN & ANN) are used to build our model & check the accuracy score. The model with the best accuracy score is used for further analysis to predict the overall satisfaction of non- Patanjali users.

VARIABLE IDENTIFICATION



DATA SPLITTING



VARIABLES ENCODED

GENDER		
Female		0
Male		1
LOCATION		
Central Line		1
Western Line		2
Harbour Line		3
ALL OTHER HERBAL BRANDS		
No		0
Yes		1
ALL REASONS FOR BUYING		
No		0
Yes		1
ALL BENEFITS		
No		0
Yes		1
ALL IMPROVEMENTS NEEDED		
No		0
Yes		1
ALL CAUSES FOR A BETTER BRAND		
No		0
Yes		1
OVERALL SATISFACTION		
No		0
Yes		1

ARTIFICIAL NEURAL NETWORK

An artificial neural network (**ANN**) is the component of artificial intelligence that **is meant** to simulate the functioning of a human brain. Processing units make up ANNs, which in turn consist of inputs and outputs. The inputs are what the **ANN** learns from to produce the desired output.

PROS

1. Flexible & can be used for both regression problems as well as classification problems.
2. Neural networks are good to model with non-linear data with larger number of inputs.
3. Once trained, the predictions are pretty fast.
4. Can be trained with any number of input and layers.
5. Works best with more data points.

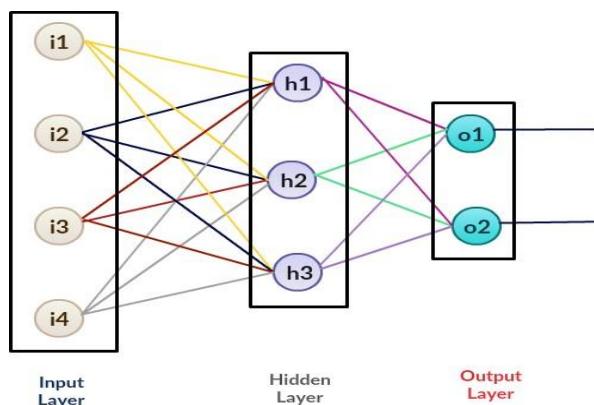
CONS

1. Neural Networks are black boxes, meaning, we cannot know how much each independent variable is influencing the dependent variables.
2. Computationally very expensive & time consuming to train with traditional CPU's.
3. Can lead to problem of over fitting & generalization.

ASSUMPTIONS

1. Artificial Neurons are arranged in layers, which are sequentially arranged.
2. Neurons within the same layer do not interact or communicate to each other.
3. All inputs enter into the network through the input layer and passes through the output layer.
4. All hidden layers at same level should have same activation function.
5. Artificial neurons at consecutive layers are densely connected.
6. Every inter-connected neural network has its own weight and biased associated with it.

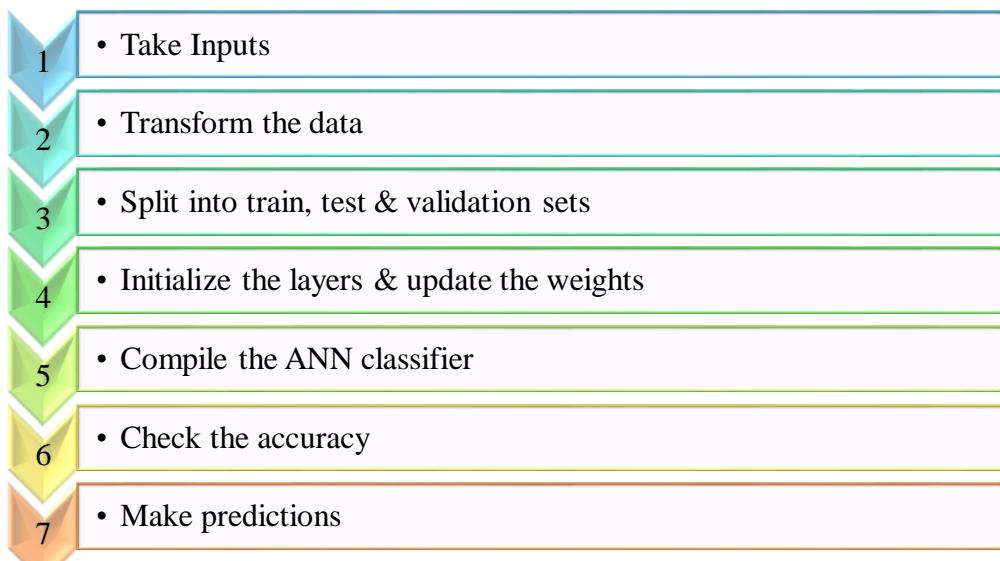
AN ARTIFICIAL NEURAL NETWORK WITH 3 LAYERS



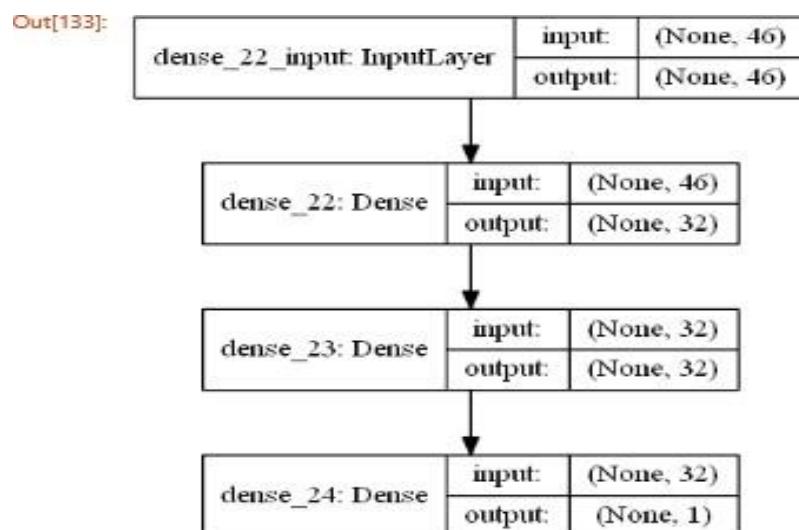
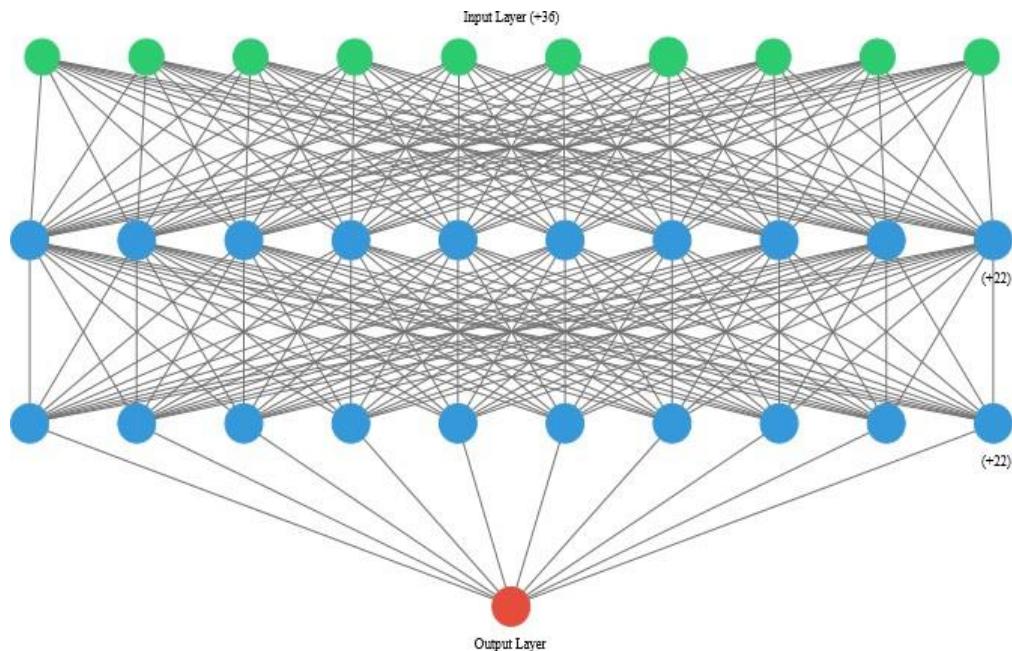
1. **Input Layer:** Initial data for the neural network
2. **Hidden Layer:** Intermediate layer between input and output layer and place where all the communication is done.
3. **Output Layer:** Produce the results for given inputs.

STEPS INVOLVED IN IMPLEMENTATION OF A NEURAL NETWORK

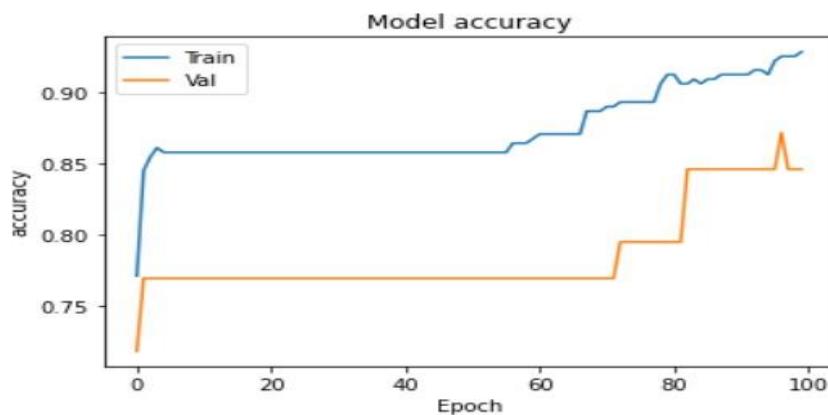
1. **Feed forward:** On a feed forward neural network, we have a set up input features and some random weights.
2. **Back propagation:** During back propagation, we calculate the error between predicted output and target output and then use an algorithm (gradient descent) to update the weight values.

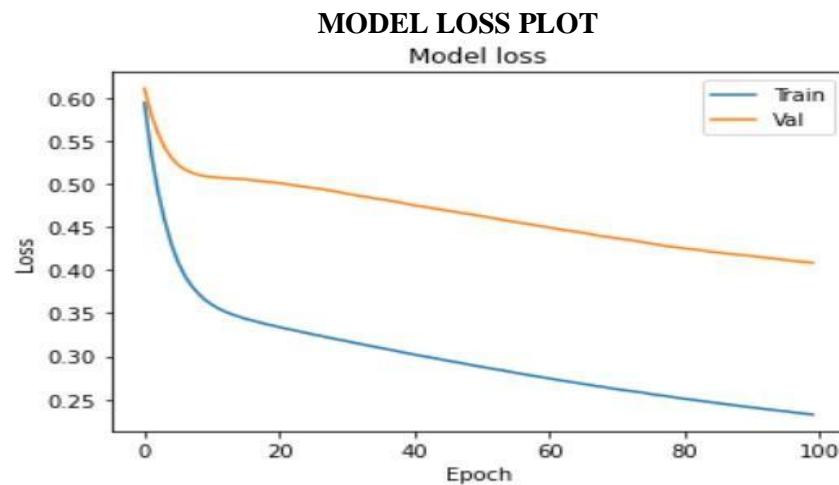


DATA ANALYSIS & INTERPRETATION



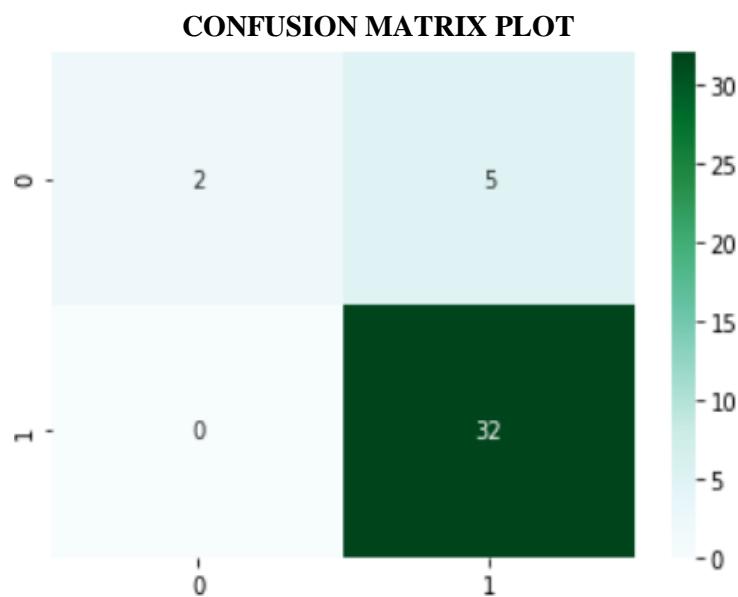
MODEL ACCURACY PLOT





CONFUSION MATRIX

		Predicted	
		Yes	No
Actual	Yes	TP=2	FP=5
	No	FN=0	TN=32



$$\text{Accuracy} = (\text{TN} + \text{TP}) / (\text{TN} + \text{TP} + \text{FN} + \text{FP})$$

Accuracy score = 87.17%

K-NEAREST NEIGHBORS

The k-nearest neighbors (KNN) algorithm is a simple, supervised machine learning algorithm that can be used to solve both classification and regression problems. It's easy to implement and understand, but has a major drawback of becoming significantly slower as the size of that data in use grows. KNN works by finding the distances between a query and all the examples in the data, selecting the specified number of examples (K) closest to the query, then votes for the most frequent label (in the case of classification) or averages the labels (in the case of regression). In the case of classification and regression, choosing the right K for our data is done by trying several Ks and picking the one that works best.

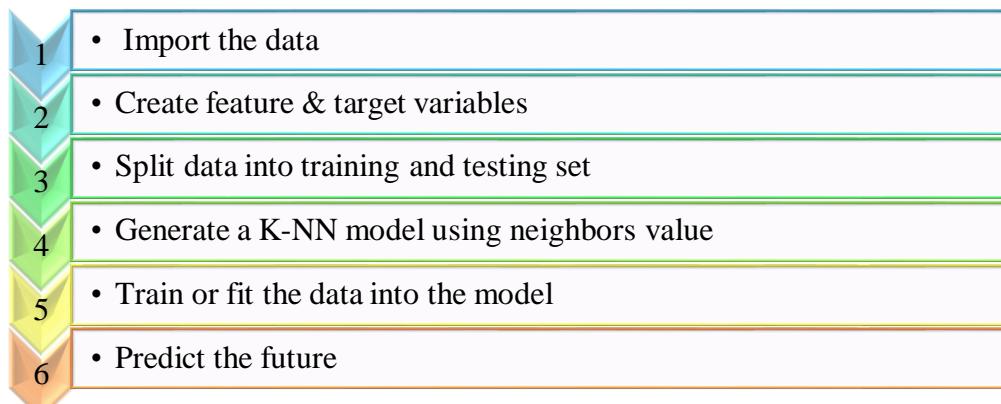
PROS

1. K-NN is pretty intuitive and simple.
2. K-NN has no assumptions.
3. It constantly evolves.
4. Very easy to implement for multi-class problem.
5. Can be used both for Classification and Regression.
6. One Hyper Parameter.
7. Variety of distance criteria to be chosen.

CONS

1. K-NN is slow algorithm.
2. Curse of Dimensionality.
3. K-NN needs homogeneous features.
4. Optimal number of neighbors.
5. Imbalanced data causes problems.
6. Outlier sensitivity.
7. Missing Value treatment.

STEPS TO IMPLEMENT K-NN ALGORITHM



DATA ANALYSIS & INTERPRETATION

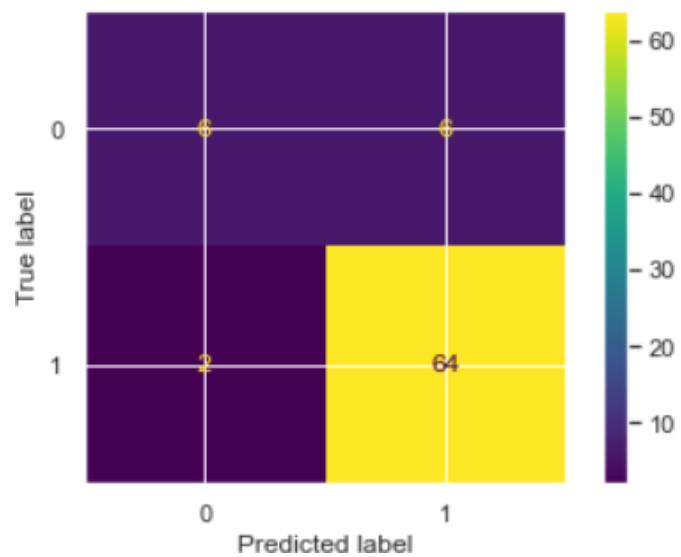
MODEL ACCURACY PLOT



CONFUSION MATRIX

		Predicted	
		Yes	No
Actual	Yes	TP=6	FP=6
	No	FN=2	TN=64

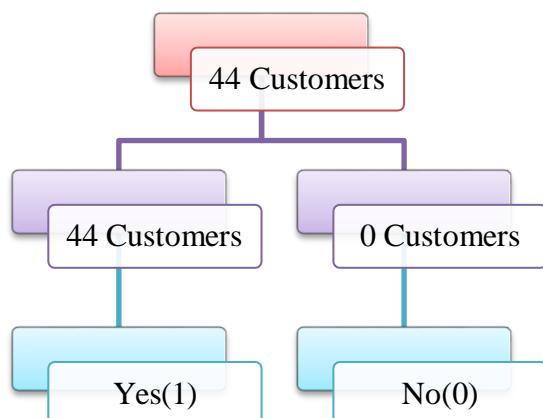
CONFUSION MATRIX PLOT



Using **K=5**,
 $\text{Accuracy} = (\text{TN} + \text{TP}) / (\text{TN} + \text{TP} + \text{FN} + \text{FP})$
Accuracy score = 89.74%

Since, K-NN have higher accuracy score than ANN, we proceed using K-NN for further analysis, i.e. for predicting the overall satisfaction for Non-Patanjali users.

MAKING PREDICTIONS USING K-NN



Thus, out of 44 Non-Patanjali users, i.e. other herbal brand users, all 44 customers will be overall satisfied by the Patanjali products.

DECISION TREES

OBJECTIVE 5

To predict how often Customer use Patanjali products based on Socio-Demographic factor, Sources of Awareness and their Buying and Spending Patterns.

Decision tree analysis is a powerful decision-making tool which initiates a structured nonparametric approach for problem-solving. It facilitates the evaluation and comparison of the various options and their results, as shown in a decision tree. It helps to choose the most competitive alternative.

PROS

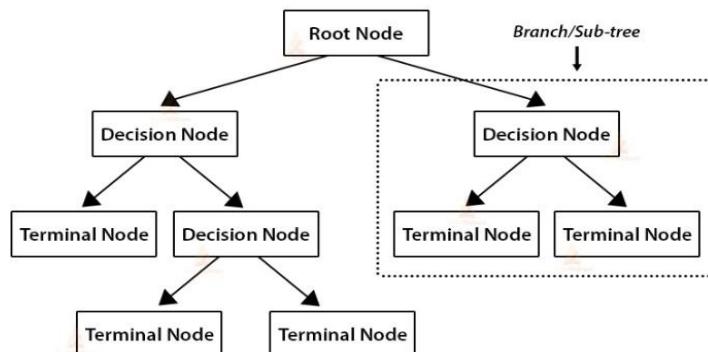
1. They are simple to understand and interpret.
2. Important insights can be generated based on experts describing a situation (its alternatives, probabilities, and costs) and their preferences for outcomes.
3. Decision Trees help to determine worst, best and expected values for different scenarios.
4. Decision Tree can be combined with other decision technique.

CONS

1. They are unstable, meaning that a small change in the data can lead to a large change in the structure of the optimal decision tree.
2. They are often relatively inaccurate. Many other predictors perform better with similar data. This can be remedied by replacing a single decision tree with a random forest of decision trees, but a random forest is not as easy to interpret as a single decision tree.
3. For data including categorical variables with different number of levels, information gain in decision trees is biased in favor of those attributes with more levels.
4. Calculations can get very complex, particularly if many values are uncertain and/or if many outcomes are linked.

TERMINOLOGIES USED

1. **Root Node:** A root node compiles the whole sample, it is then divided into multiple sets which comprise of homogeneous variables.
2. **Decision Node:** That sub-node which diverges into further possibilities, can be denoted as a decision node.
3. **Terminal Node:** The final node showing the outcome which cannot be categorized any further, is termed as a value or terminal node.
4. **Branch:** A branch denotes the various alternatives available with the decision tree maker.
5. **Splitting:** The division of the available option (depicted by a node or sub-node) into multiple sub-nodes is termed as splitting.
6. **Pruning:** It is just the reverse of splitting, where the decision tree maker can eliminate one or more sub-nodes from a particular decision node.



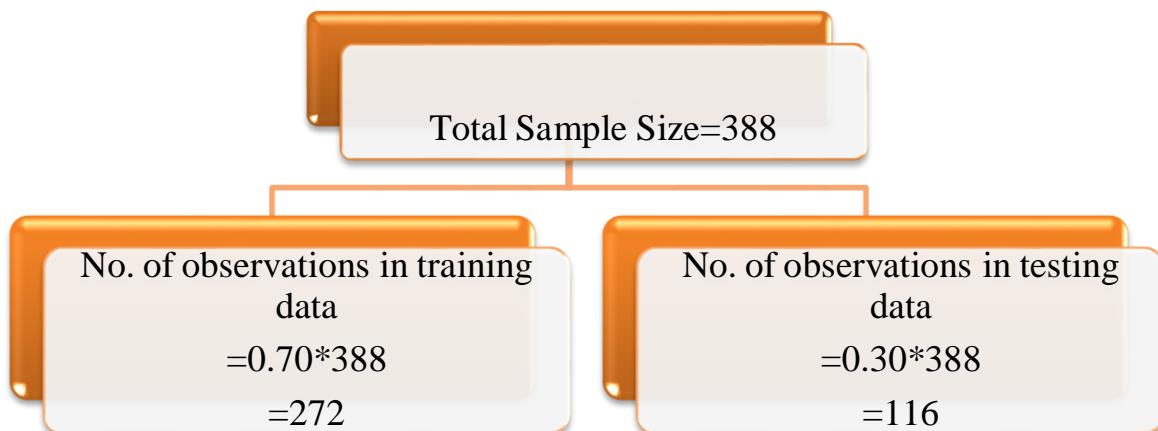
STEPS TO IMPLEMENT DECISION TREES ALGORITHM

- 1 • Load required packages.
- 2 • Load the dataset.
- 3 • Define the feature and the target variable.
- 4 • Split the dataset into train and test sets.
- 5 • Build the model with the help of decision tree classifier.
- 6 • Get the accuracy score
- 7 • Predict the values

VARIABLE IDENTIFICATION



DATA SPLITTING



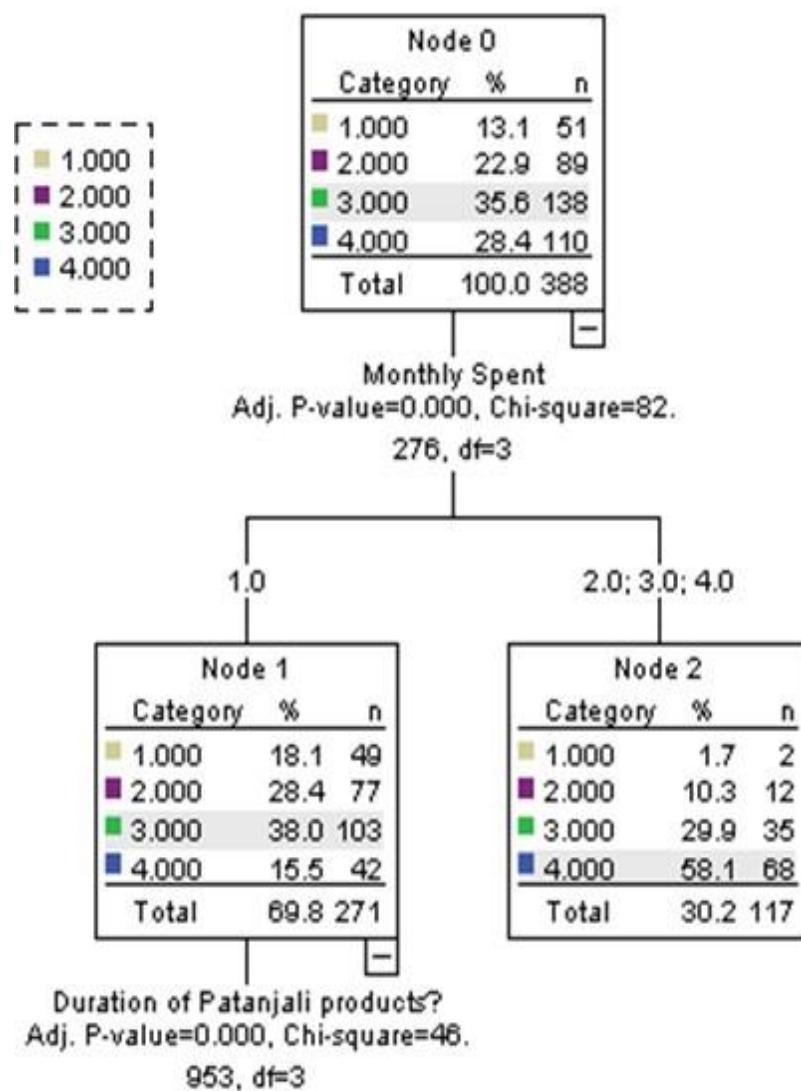
VARIABLES ENCODED

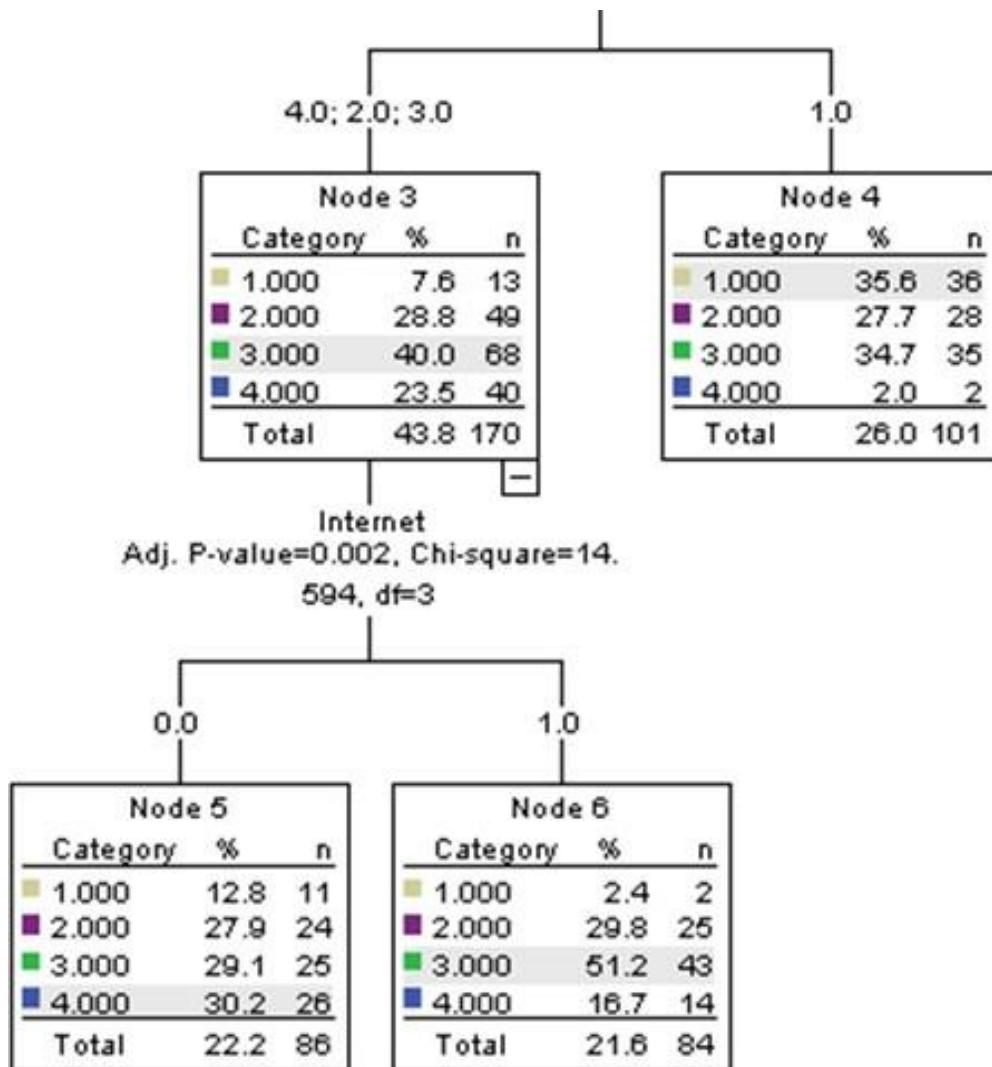
GENDER	
Female	0
Male	1
LOCATION	
Central Line	1
Western Line	2
Harbour Line	3
MONTHLY EXPENDITURE	
less than Rs 500/-	1
Rs 500/- to Rs 1000/-	2
Rs 1000/- to Rs 2000/-	3
More than Rs 2000/-	4
YEARS OF USE	
Less than a year	1
1-3 years	2
3-5 years	3
More than 5 years	4
FREQUENCY OF USE	
Only on trial basis	1
Rarely	2
Sometimes	3

Regularly	4
ALL SOURCES OF AWARENESS	
No	0
Yes	1
OVERALL SATISFACTION	
No	0
Yes	1

DATA ANALYSIS & INTERPRETATION

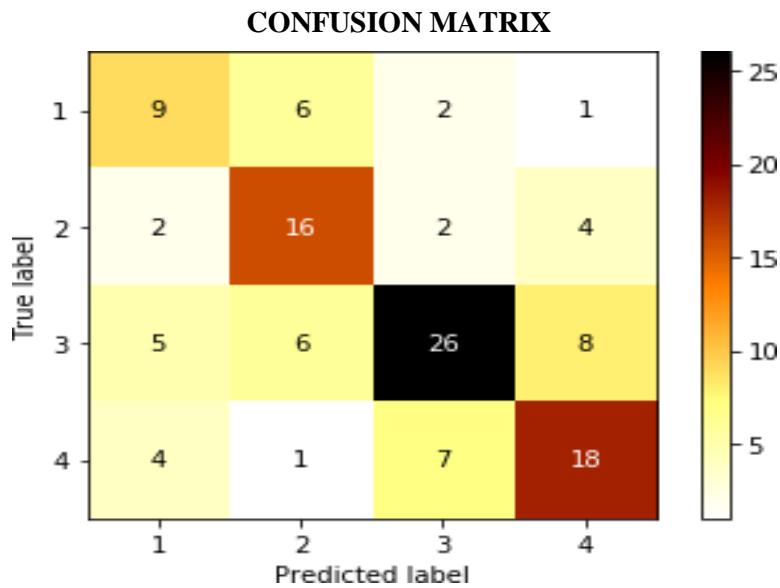
How often do you use Patanjali products?





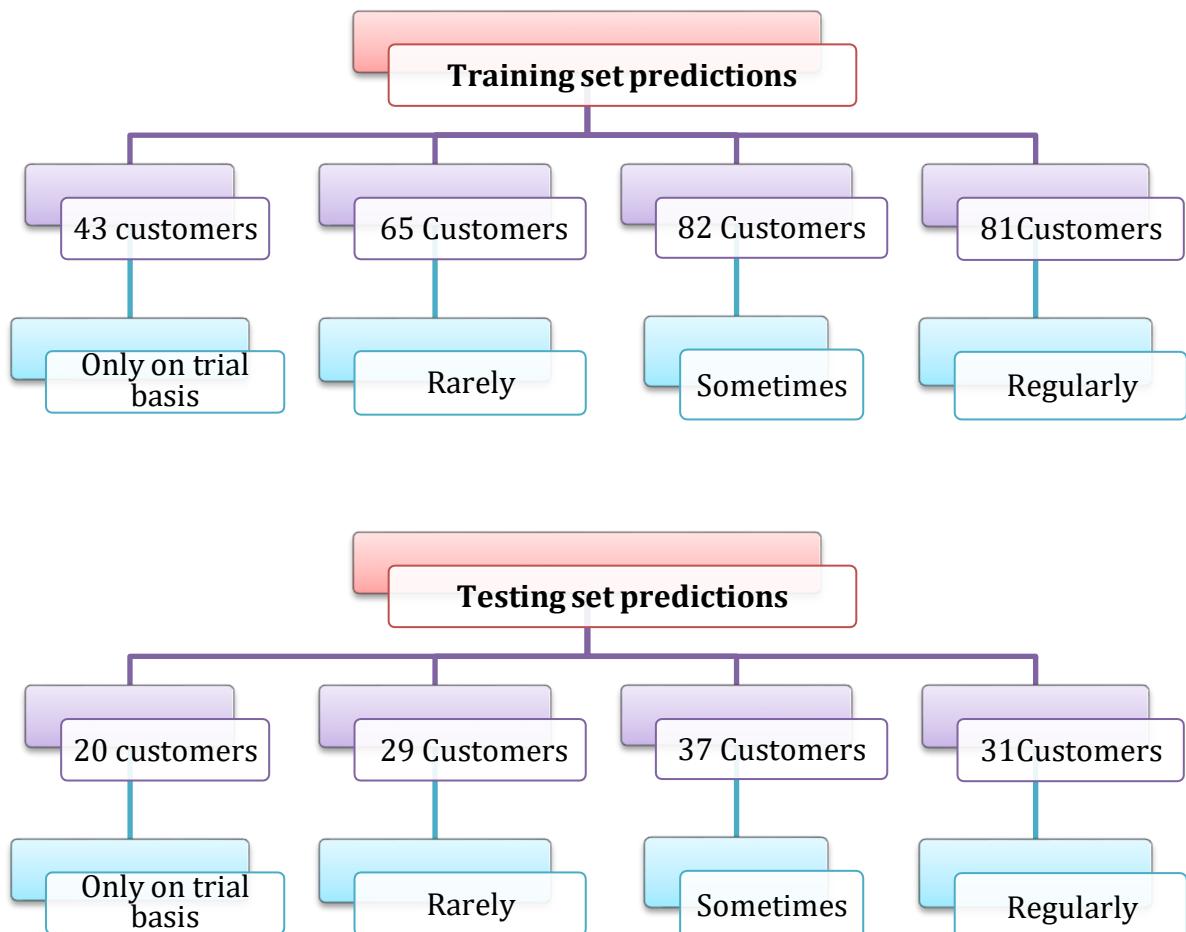
INTERPRETATION

- The highest frequency is 138 which accounts to 35.6% of the population which means majorly people uses Patanjali products sometimes.
- The monthly spent on the products is less than Rs 500/- for maximum people which is for people using it sometimes and for (2,3,4),the highest count is for people using the products regularly.
- For duration of using Patanjali products (2,3,4) the maximum falls under people using the products sometimes where as for (1) it falls under the category of only on trial basis.
- The source of awareness is internet in which (0) accounts for people who are not aware about the product through internet for which maximum count is for the people using the product regularly. (1) accounts for people who are aware about the products through internet for which maximum frequency is for people using the product sometimes.



Accuracy score = 58.97%

MAKING PREDICTIONS



Thus, both the sets predicts that majority of the people will use Patanjali products sometimes, followed by people using them regularly and rarely. Less number of people will use them only on trial basis.

MARKET BASKET ANALYSIS USING APRIORI ALGORITHM

OBJECTIVE 6

To identify next product purchase that might interest a customer.

We have used Market Basket Analysis to identify what product the customer would purchase next looking at the current purchase of the product.

The process of discovering frequent item sets in large transactional database is called Market Basket Analysis. Frequent item set mining leads to the discovery of associations and correlations among items.

PROS

1. It is easy to implement and interpret.
2. It can be used on large datasets and can easily be parallelized.

CONS

1. Calculating support is expensive as it has to go through the entire dataset.
2. It is computationally expensive.

STEPS TO PERFORM MARKET BASKET ANALYSIS

To perform a Market Basket Analysis and identify potential rules, a data mining algorithm called the 'Apriori algorithm' is commonly used, which works in two steps:

Systematically identify itemset's that occur frequently in the data set with a support greater than a pre-specified threshold.

Calculate the confidence of all possible rules given the frequent itemset's and keep only those with a confidence greater than a pre-specified threshold.

APRIORI ALGORITHM

Association Rule: $X \rightarrow Y$

Interpretation: Customer who buy X is likely to buy Y

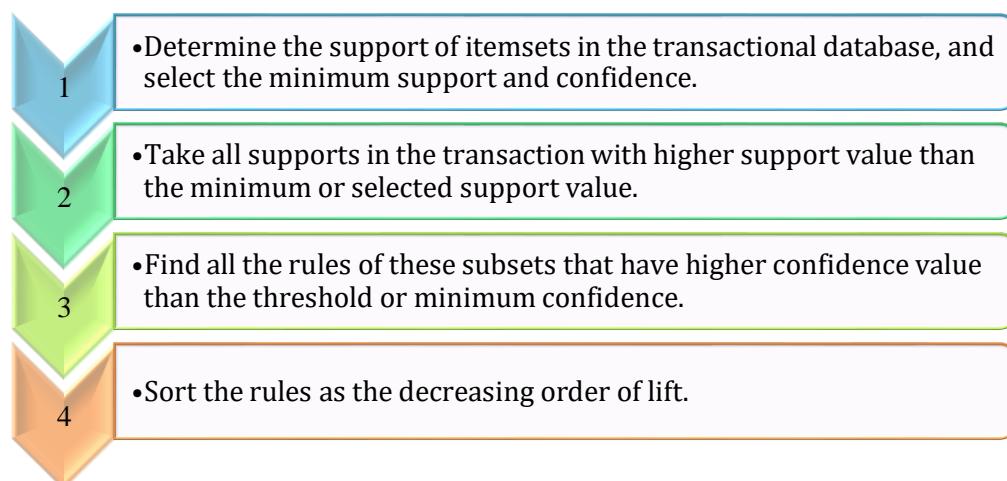
X: Rule antecedent

Y: Rule consequent

Association rules analysis is a technique to uncover how items are associated to each other. There are three common ways to measure association.

1. **Support:** This says how popular an item set is, it is number of times an item appears in total number of transaction in other word we say frequency of item.
2. **Confidence:** This says how likely item Y is purchased when item X is purchased, expressed as $\{X \rightarrow Y\}$. This is measured by the proportion of transactions with item X, in which item Y also appears.
3. **Lift:** It is ratio of expected confidence to observed confidence. it is described as confidence of Y when item X was already known(x/y) to the confidence of Y when X item is unknown. In other words confidence of Y w.r.t. X and confidence of Y without X (means both are independent to each other).

STEPS TO PERFORM APRIORI ALGORITHM

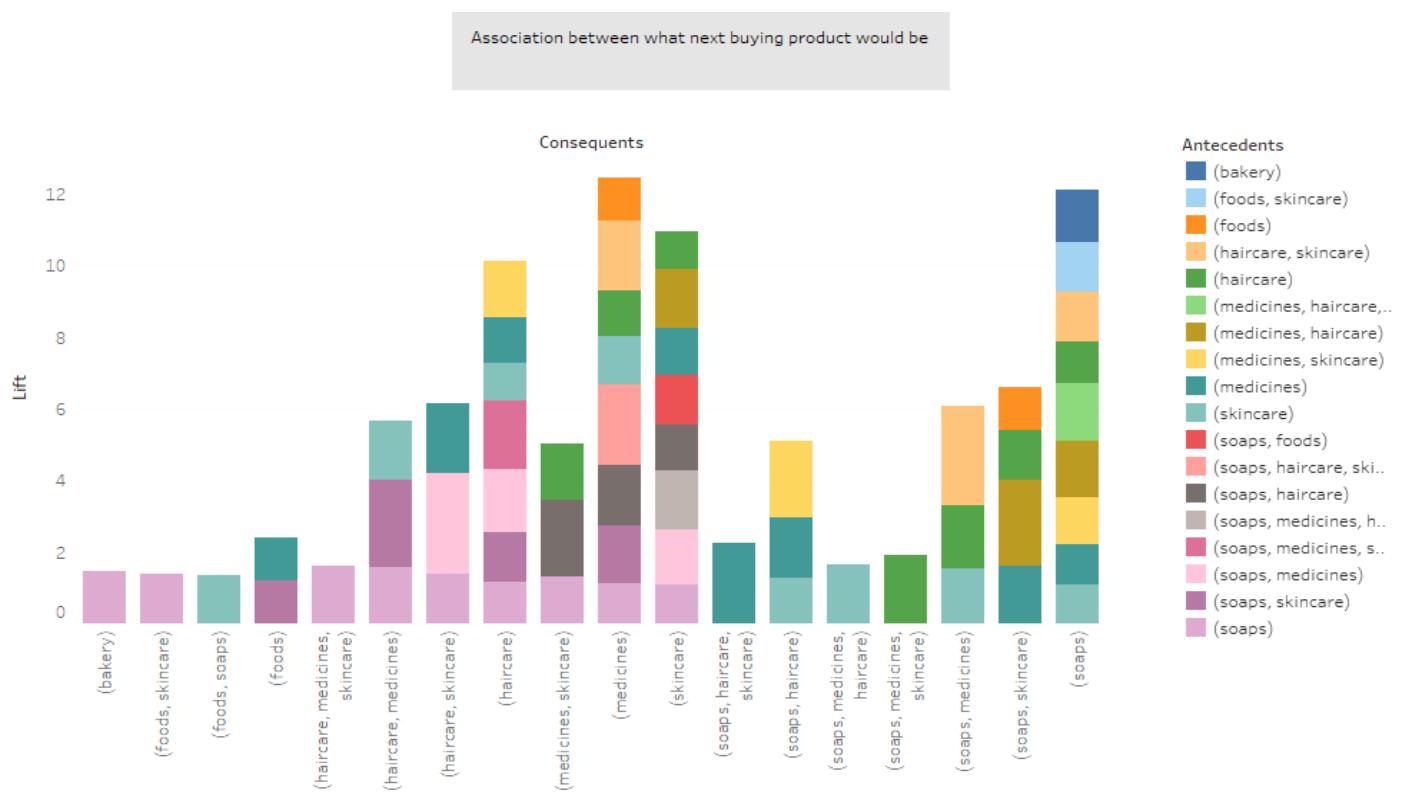


DATA INTERPRETATION & ANALYSIS

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
46	(soaps, medicines, haircare)	(skincare)	0.082474	0.561856	0.077320	0.937500	1.668578	0.030981	7.010309
34	(haircare, medicines)	(skincare)	0.105670	0.561856	0.097938	0.926829	1.649586	0.038567	5.987973
16	(soaps, medicines)	(skincare)	0.131443	0.561856	0.113402	0.862745	1.535528	0.039550	3.192194
49	(haircare, medicines, skincare)	(soaps)	0.097938	0.494845	0.077320	0.789474	1.595395	0.028855	2.399485
42	(haircare, medicines)	(soaps)	0.105670	0.494845	0.082474	0.780488	1.577236	0.030184	2.301260
28	(foods, soaps)	(skincare)	0.121134	0.561856	0.092784	0.765957	1.363264	0.024724	1.872071
2	(medicines)	(skincare)	0.234536	0.561856	0.175258	0.747253	1.329973	0.043482	1.733528
53	(haircare, medicines)	(soaps, skincare)	0.105670	0.301546	0.077320	0.731707	2.426517	0.045455	2.603327
11	(bakery)	(soaps)	0.100515	0.494845	0.072165	0.717949	1.450855	0.022425	1.791003
22	(soaps, haircare)	(skincare)	0.206186	0.561856	0.146907	0.712500	1.268119	0.031061	1.523980
24	(haircare, skincare)	(soaps)	0.211340	0.494845	0.146907	0.695122	1.404726	0.042326	1.656907
47	(soaps, medicines, skincare)	(haircare)	0.113402	0.355670	0.077320	0.681818	1.916996	0.036986	2.025037
29	(foods, skincare)	(soaps)	0.136598	0.494845	0.092784	0.679245	1.372642	0.025189	1.574894
18	(medicines, skincare)	(soaps)	0.175258	0.494845	0.113402	0.647059	1.307598	0.026677	1.431271
40	(soaps, medicines)	(haircare)	0.131443	0.355670	0.082474	0.627451	1.764138	0.035724	1.729517
0	(soaps)	(skincare)	0.494845	0.561856	0.301546	0.609375	1.084576	0.023515	1.121649
4	(haircare)	(skincare)	0.355670	0.561856	0.211340	0.594203	1.057572	0.011505	1.079713
50	(soaps, medicines)	(haircare, skincare)	0.131443	0.211340	0.077320	0.588235	2.783357	0.049540	1.915317
9	(haircare)	(soaps)	0.355670	0.494845	0.206186	0.579710	1.171498	0.030184	1.201920
7	(medicines)	(soaps)	0.234536	0.494845	0.131443	0.560440	1.132555	0.015384	1.149227
36	(medicines, skincare)	(haircare)	0.175258	0.355670	0.097938	0.558824	1.571185	0.035604	1.460481
1	(skincare)	(soaps)	0.561856	0.494845	0.301546	0.536697	1.084576	0.023515	1.090334
48	(soaps, haircare, skincare)	(medicines)	0.146907	0.234536	0.077320	0.526316	2.244072	0.042865	1.615979
23	(soaps, skincare)	(haircare)	0.301546	0.355670	0.146907	0.487179	1.369751	0.039656	1.256443
20	(medicines)	(soaps, skincare)	0.234536	0.301546	0.113402	0.483516	1.603456	0.042679	1.352325
35	(haircare, skincare)	(medicines)	0.211340	0.234536	0.097938	0.463415	1.975878	0.048371	1.426546
13	(medicines)	(haircare)	0.234536	0.355670	0.105670	0.450549	1.266762	0.022253	1.172680
54	(medicines, skincare)	(soaps, haircare)	0.175258	0.206186	0.077320	0.441176	2.139706	0.041184	1.420510
38	(medicines)	(haircare, skincare)	0.234536	0.211340	0.097938	0.417582	1.975878	0.048371	1.354114
8	(soaps)	(haircare)	0.494845	0.355670	0.206186	0.416667	1.171498	0.030184	1.104566
26	(haircare)	(soaps, skincare)	0.355670	0.301546	0.146907	0.413043	1.369751	0.039656	1.189958
41	(soaps, haircare)	(medicines)	0.206186	0.234536	0.082474	0.400000	1.705495	0.034116	1.275773
5	(skincare)	(haircare)	0.561856	0.355670	0.211340	0.376147	1.057572	0.011505	1.032823
17	(soaps, skincare)	(medicines)	0.301546	0.234536	0.113402	0.376068	1.603456	0.042679	1.226839
51	(soaps, haircare)	(medicines, skincare)	0.206186	0.175258	0.077320	0.375000	2.139706	0.041184	1.319588

PATANJALI-The World Of Herbology

55	(haircare, skincare)	(soaps, medicines)	0.211340	0.131443	0.077320	0.365854	2.783357	0.049540	1.369647
31	(foods)	(soaps, skincare)	0.255155	0.301546	0.092784	0.363636	1.205905	0.015843	1.097570
44	(medicines)	(soaps, haircare)	0.234536	0.206186	0.082474	0.351648	1.705495	0.034116	1.224358
57	(medicines)	(soaps, haircare, skincare)	0.234536	0.146907	0.077320	0.329670	2.244072	0.042865	1.272647
3	(skincare)	(medicines)	0.561856	0.234536	0.175258	0.311927	1.329973	0.043482	1.112474
30	(soaps, skincare)	(foods)	0.301546	0.255155	0.092784	0.307692	1.205905	0.015843	1.075888
15	(medicines)	(foods)	0.234536	0.255155	0.072165	0.307692	1.205905	0.012322	1.075888
12	(haircare)	(medicines)	0.355670	0.234536	0.105670	0.297101	1.266762	0.022253	1.089011
25	(soaps)	(haircare, skincare)	0.494845	0.211340	0.146907	0.296875	1.404726	0.042326	1.121649
14	(foods)	(medicines)	0.255155	0.234536	0.072165	0.282828	1.205905	0.012322	1.067337
37	(haircare)	(medicines, skincare)	0.355670	0.175258	0.097938	0.275362	1.571185	0.035604	1.138144
6	(soaps)	(medicines)	0.494845	0.234536	0.131443	0.265625	1.132555	0.015384	1.042334
27	(skincare)	(soaps, haircare)	0.561856	0.206186	0.146907	0.261468	1.268119	0.031061	1.074854
52	(soaps, skincare)	(haircare, medicines)	0.301546	0.105670	0.077320	0.256410	2.426517	0.045455	1.202720
45	(haircare)	(soaps, medicines)	0.355670	0.131443	0.082474	0.231884	1.764138	0.035724	1.130762
19	(soaps)	(medicines, skincare)	0.494845	0.175258	0.113402	0.229167	1.307598	0.026677	1.069936
58	(haircare)	(soaps, medicines, skincare)	0.355670	0.113402	0.077320	0.217391	1.916996	0.036986	1.132875
21	(skincare)	(soaps, medicines)	0.561856	0.131443	0.113402	0.201835	1.535528	0.039550	1.088192
32	(soaps)	(foods, skincare)	0.494845	0.136598	0.092784	0.187500	1.372642	0.025189	1.062649
39	(skincare)	(haircare, medicines)	0.561856	0.105670	0.097938	0.174312	1.649586	0.038567	1.083133
43	(soaps)	(haircare, medicines)	0.494845	0.105670	0.082474	0.166667	1.577236	0.030184	1.073196
33	(skincare)	(foods, soaps)	0.561856	0.121134	0.092784	0.165138	1.363264	0.024724	1.052708
56	(soaps)	(haircare, medicines, skincare)	0.494845	0.097938	0.077320	0.156250	1.595395	0.028855	1.069110
10	(soaps)	(bakery)	0.494845	0.100515	0.072165	0.145833	1.450855	0.022425	1.053055
59	(skincare)	(soaps, medicines, haircare)	0.561856	0.082474	0.077320	0.137615	1.668578	0.030981	1.063939



Here, the association is highest between (hair-care, skincare) and (soaps, medicines) that is people buying hair-care & skincare products are highly likely to buy soaps and medicines as well.

PARETO ANALYSIS

OBJECTIVE 7

To study the factors that affects Patanjali from being a better brand.

Pareto Analysis is a statistical technique in decision-making used for the selection of a limited number of tasks that produce significant overall effect. It uses the Pareto Principle (also known as the 80/20 rule) the idea that by doing 20% of the work you can generate 80% of the benefit of doing the entire job. A Pareto chart is a basic quality tool that helps you identify the most frequent defects, complaints, or any other factor you can count and categorize. The chart takes its name from Vilfredo Pareto, originator of the "80/20 rule," which postulates that, roughly speaking, 20 percent of the people own 80 percent of the wealth. Or, in quality terms, 80 percent of the losses come from 20 percent of the causes. You can use a Pareto chart any time you have data that are broken down into categories, and you can count how often each category occurs.

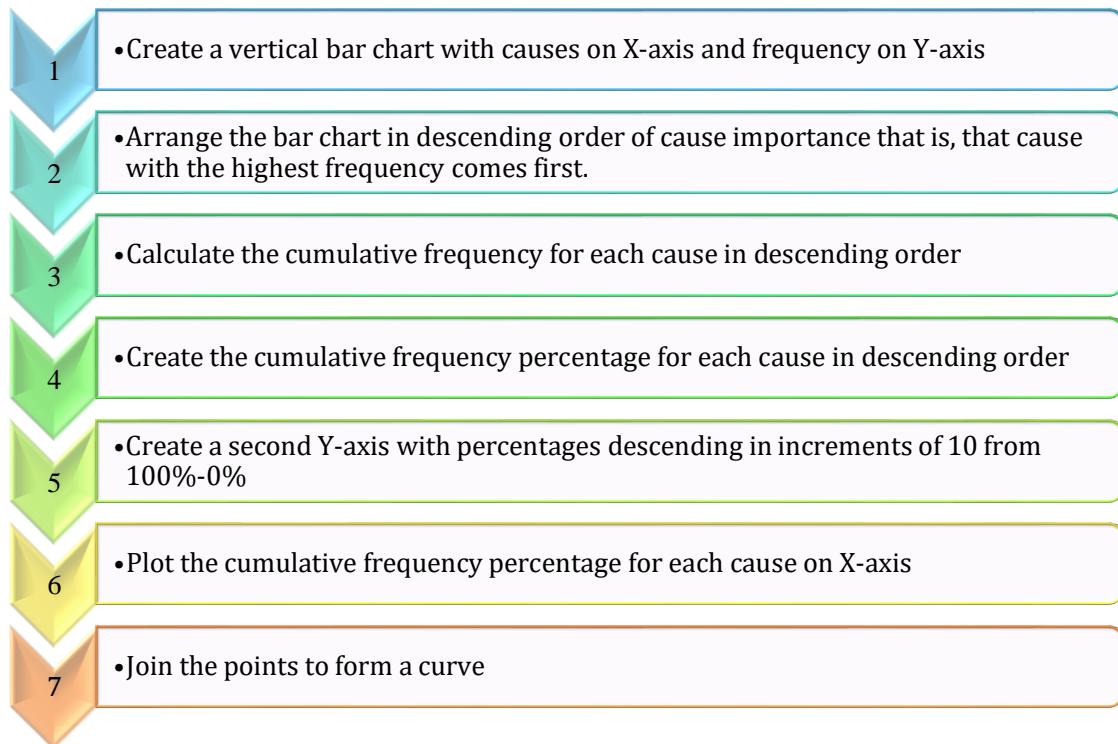
PROS

1. It simply helps to identify and determine main cause i.e. root causes of defects or problems.
2. Defects are organized from highest to lowest priority with help of Pareto Chart. Defects that are on top with highest priority are eliminated first or resolved first.
3. With help of Pareto chart, one can also determine cumulative impact of defect.
4. It also helps in solving issues regarding problem-solving and decision making, time management i.e. be at work on time or personal, change management, etc.
5. One can plan, analyze, and resolve problems or defects with help of Pareto Chart.

CONS

1. One of main disadvantages is that root cause analysis cannot be done by itself in Pareto analysis. There is a requirement of tool i.e. root cause analysis tool for determining or identify root causes or major causes of defect.
2. It does not represent severity of defect or any problem. It only shows qualitative data.
3. Pareto analysis only focuses on past data where damage has already happened. It should focus on past data as well as present and future data also.

STEPS IN IMPLEMENTATION OF PARETO CHARTS

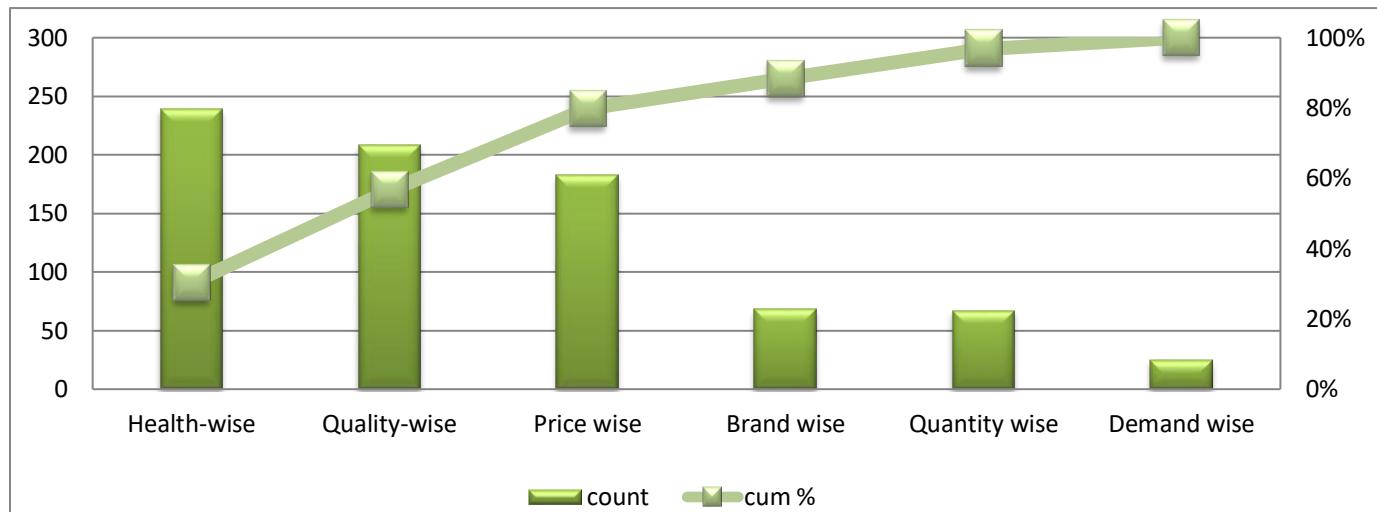


Here, we will consider 3 sub objectives to find the factors that affects the brand.

- 1) To identify factor that affects preferences of people while buying Patanjali products.

Pareto for factors that matters while buying the Patanjali product

Variables	Frequency	Cumulative %
Health-wise	239	0.302532
Quality-wise	208	0.565823
Price wise	183	0.797468
Brand wise	68	0.883544
Quantity wise	67	0.968354
Demand wise	25	1

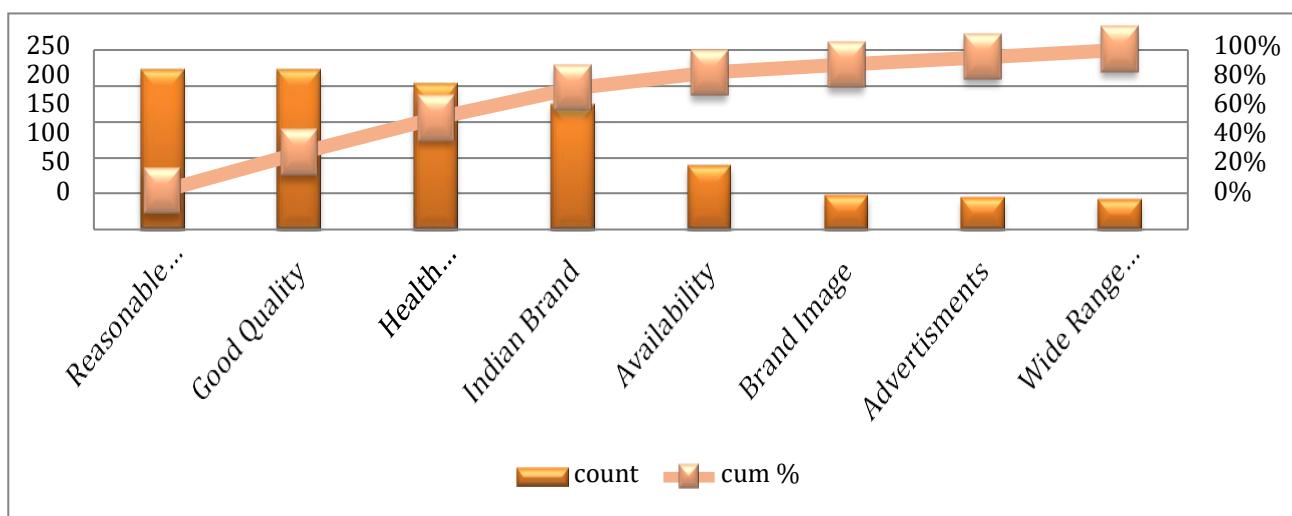


Health-wise, Quality-wise & Price-wise are the major factors that affects the preferences of people while buying the products.

2) To identify reasons for not buying Patanjali Products.

Pareto for not buying Patanjali Product

Variables	Frequency	Cumulative %
Reasonable Price	224	0.21374
Good Quality	223	0.426527
Health Benefits	203	0.620229
Indian Brand	175	0.787214
Availability	89	0.872137
Brand Image	48	0.917939
Advertisements	44	0.959924
Wide Range of Varieties	42	1

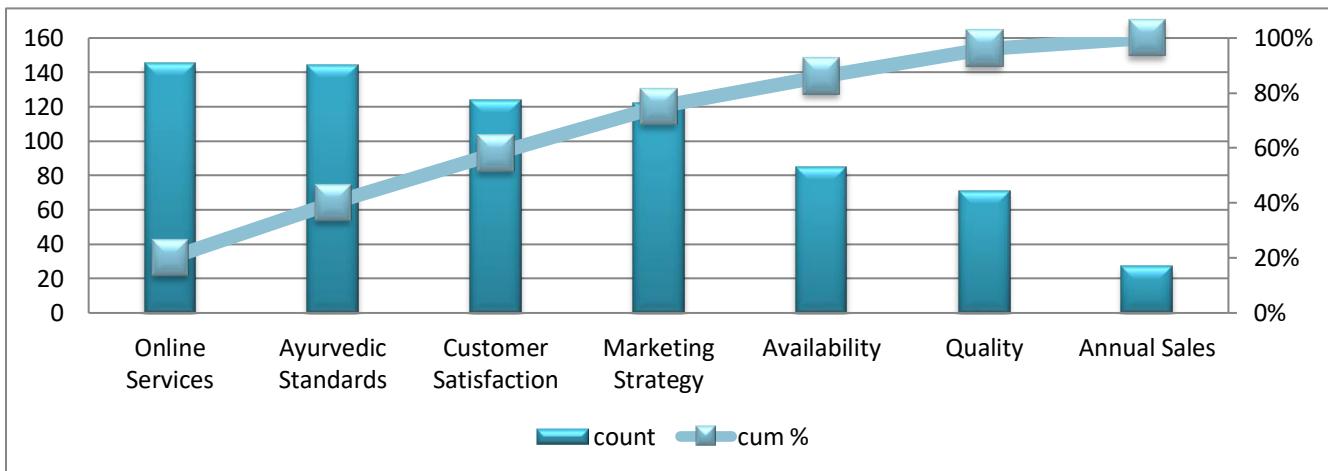


Lack of reasonable prices, lack of good quality, lack of health benefits & lack of Indian brand image are the reasons for not buying the Patanjali products.

- 3) To identify in which category Patanjali needs improvement so as to be a better brand.

Pareto for improvement of brand

Variables	Frequency	Cumulative %
Online Services	145	0.20195
Ayurvedic Standards	144	0.402507
Customer Satisfaction	124	0.575209
Marketing Strategy	122	0.745125
Availability	85	0.86351
Quality	71	0.962396
Annual Sales	27	1



So as to improvise as a brand, Patanjali needs to improve the online services, Ayurvedic standards, customer satisfaction and Marketing strategy.

FACTOR ANALYSIS

OBJECTIVE 8

To determine the factors that influences the people for buying Patanjali products.

Factor analysis is a statistical method used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called factors. In an experiment, the factor (also called an independent variable) is an explanatory variable manipulated by the experimenter. Each factor has two or more levels (i.e., different values of the factor). Combinations of factor levels are called treatments.

PROS

1. Identification of groups of inter-related variables, to see how they are related to each other.
2. Factor analysis can be used to identify the hidden dimensions or constructs which may or may not be apparent from direct analysis.
3. Both objective and subjective attributes can be used.
4. Reduction of number of variables, by combining two or more variables into a single factor.
5. There is flexibility in naming using dimensions.
6. It is not extremely difficult to do, inexpensive, and accurate.

CONS

1. Factor analysis can be good only as the data allows.
2. Naming of the factors can be difficult – multiple attributes can be highly correlated with no apparent reason.

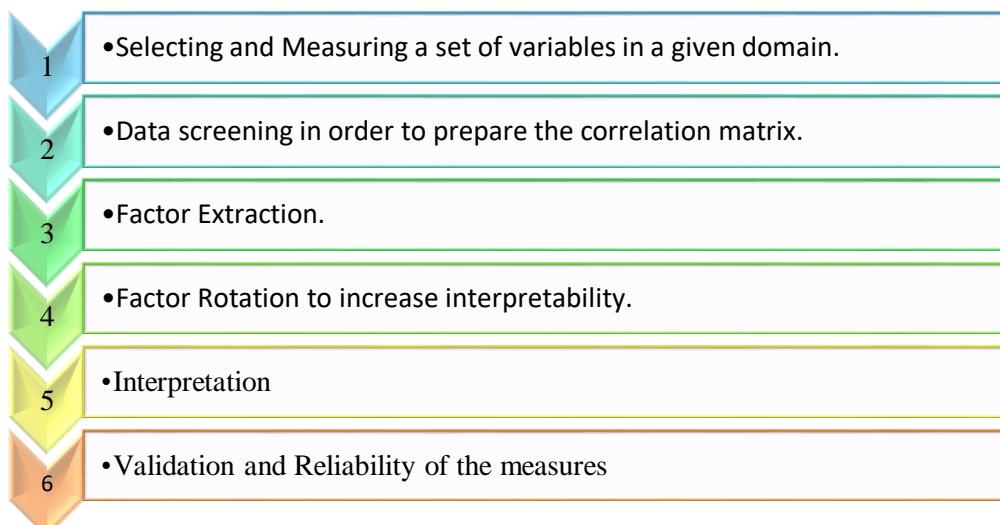
ASSUMPTIONS

1. Adequate sample size
2. No outliers
3. No multi-collinearity
4. Interval data

TERMINOLOGIES USED

1. **Correlation Matrix:** A correlation (covariance) matrix is symmetric matrix showing the simple correlations, (covariance) between all possible pairs of variables included in the analysis.
2. **Communality:** The portion of the variables of the i^{th} variable contributed by the m common factors is called the Communality.
3. **Eigen Values:** The Eigen values represent the total variance explained by each factor.

STEPS TO IMPLEMENT FACTOR ANALYSIS



Variables considered

- V1: Popularity
- V2: Advertisements
- V3: Quality
- V4: Availability
- V5: Purity
- V6: Organic
- V7: Appealing to youngsters
- V8: Increase the promotion
- V9: Brand Ambassadors
- V10: Low price
- V11: Beneficial
- V12: Customer service

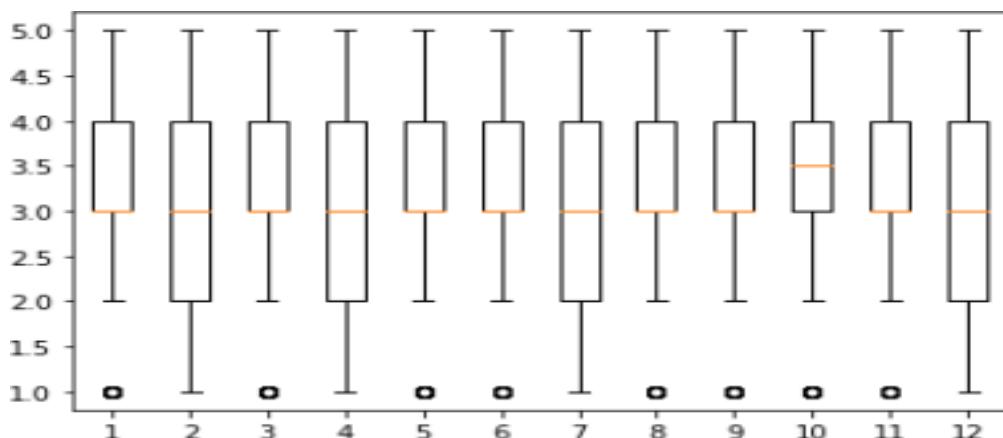
ASSUMPTIONS CHECKING
MULTICOLLINEARITY

Correlation Matrix

	Popularity	Advertisements	Quality	Availability	Purity	Organic	Appealing to Youngsters	Increase the promotion	Brand ambassadors
Correlation	1.000	.698	.646	.611	.612	.618	.480	.511	.498
Popularity	.698	1.000	.547	.490	.564	.607	.476	.355	.389
Advertisements	.646	.547	1.000	.608	.623	.679	.610	.513	.446
Quality	.611	.490	.608	1.000	.588	.624	.480	.521	.484
Availability	.612	.564	.623	.588	1.000	.802	.620	.620	.478
Purity	.618	.607	.679	.624	.802	1.000	.622	.656	.528
Organic	.480	.476	.610	.480	.620	.622	1.000	.478	.361
Appealing to Youngsters	.511	.355	.513	.521	.620	.656	.478	1.000	.589
Increase the promotion	.498	.389	.446	.484	.478	.528	.361	.589	1.000
Brand ambassadors	.362	.282	.260	.292	.270	.319	.139	.394	.363
Low Price	.433	.353	.433	.405	.515	.567	.415	.445	.398
Beneficial	.401	.343	.455	.432	.422	.492	.420	.440	.438
Customer Service									

	Low Price	Beneficial	Customer Service
	.362	.433	.401
	.282	.353	.343
	.260	.433	.455
	.292	.405	.432
	.270	.515	.422
	.319	.567	.492
	.139	.415	.420
	.394	.445	.440
	.363	.398	.438
	1.000	.612	.522
	.612	1.000	.734
	.522	.734	1.000

OUTLIERS



There are no outliers in the data.

Kaiser-Mayer-Olkin (KMO) Measure:

- The KMO measure is used for sampling adequacy.
- The statistic is a measure of the proportion of variance among variables that might be common variance.
- Generally, a value between 0.5 & 1 is desirable.

Bartlett's Test of Sphericity:

- The Bartlett's Test of Sphericity is the test for null hypothesis that the correlation matrix has an identity matrix. Taking this into consideration, these tests provide the minimum standard to proceed for Factor Analysis.
- This test which is often done prior to factor analysis, tests whether the data comes from multivariate normal distribution with zero co-variance.
- We proceed with factor analysis only if the above null hypothesis will be rejected.

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.904
Bartlett's Test of Sphericity	Approx. Chi-Square	2939.04
	df	2
	Sig.	.000

Kaiser-Mayer-Olkin Statistics:**KMO Criterion: 0.904**

The value of KMO statistics is **0.904** which is greater than 0.5 i.e. sample adequacy is met and we proceed with Factor Analysis as an appropriate technique of data reduction.

Bartlett's test of Sphericity:

H_0 : Population correlation matrix is Identity matrix.

H_1 : Population correlation matrix is not an Identity matrix.

Approx. Chi-square value 2939.042

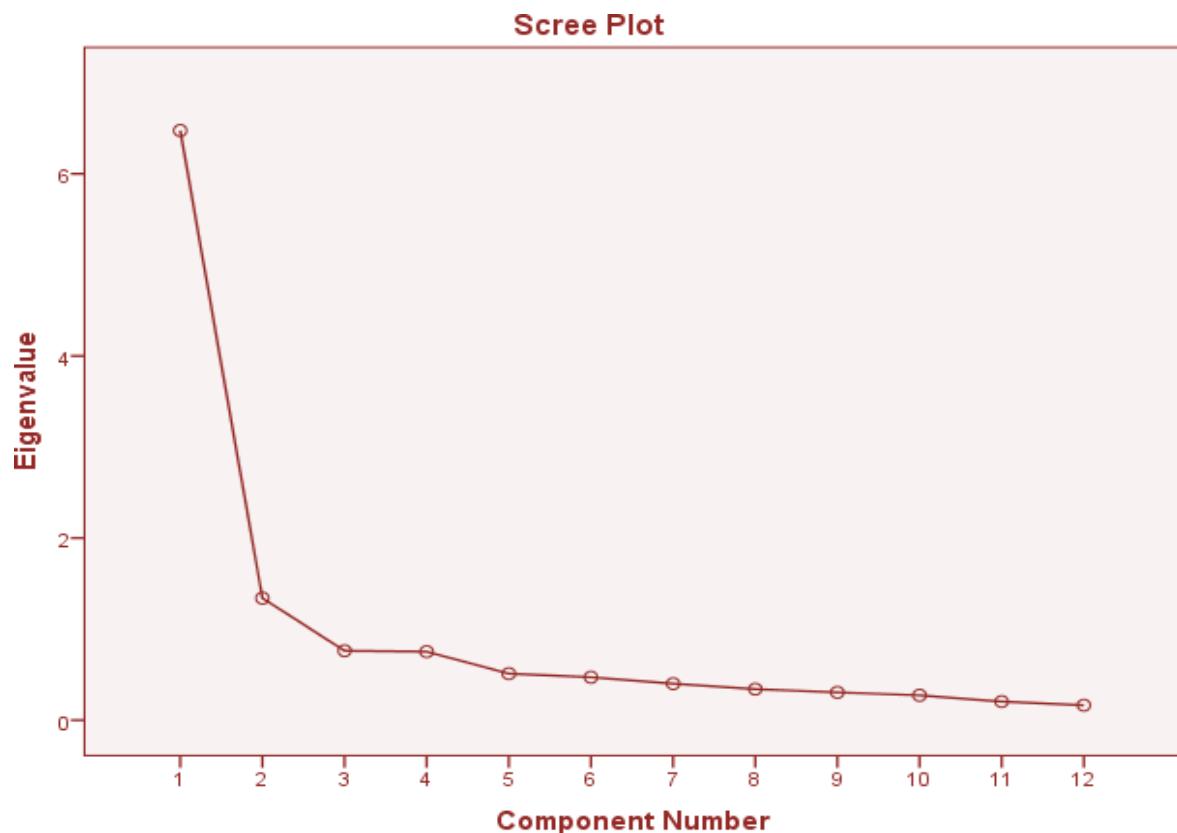
df = 66

p-value(0.000) < 0.05

Since p-value is significant, we reject H_0 and conclude that assumption of multicollinearity is satisfied.

Hence we proceed with factor analysis using principal component analysis as an appropriate technique of data reduction.

DATA ANALYSIS & INTERPRETATION



The point where the slope of the curve is clearly leveling off (the “elbow”) indicates the number of factors that should be generated by the analysis which is after the 2nd component.

Rotated Component Matrix ^a		
	Component	
	1	2
Popularity	.768	.259
Advertisements	.736	.143
Quality	.797	.211
Availability	.725	.258
Purity	.815	.252
Organic	.821	.327
Appealing to Youngsters	.747	.132
Increase the promotion	.615	.423
Brand ambassadors	.525	.433
Low Price	.080	.853
Beneficial	.322	.819
Customer Service	.315	.783

INTERPRETATION

On the basis of factor analysis and graphical representation of the study.

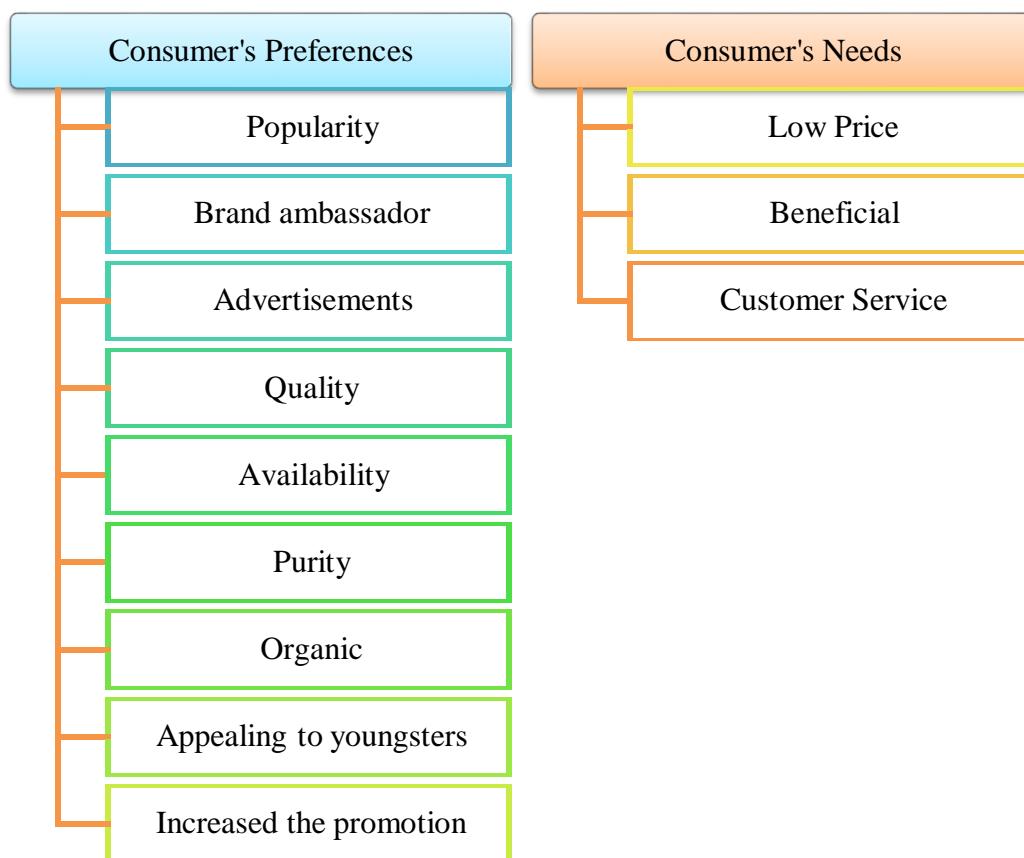
The following are the conclusions:

- Two factors are extracted using those components which are having eigen values more than 1.
- The total variance explained by these 2 factors is **65.127 %**

FACTOR ANALYSIS BY USING VARIMAX METHOD

	R1	R2
V1	0.768	
V2	0.736	
V3	0.797	
V4	0.725	
V5	0.815	
V6	0.821	
V7	0.747	
V8	0.615	
V9	0.525	
V10		0.853
V11		0.819
V12		0.783

Here, we get two factors which divides 12 variables



SENTIMENT ANALYSIS

OBJECTIVE 9

To analyze the sentimental behavior of the customers towards the brand Patanjali.

Sentiment analysis (or opinion mining) is a natural language processing technique used to determine whether data is positive, negative or neutral. Sentiment analysis is often performed on textual data to help businesses monitor brand and product sentiment in customer feedback, and understand customer needs.

PROS

1. Up selling opportunities.
2. Agent monitoring.
3. Tracking overall satisfaction of customer.
4. Handling multiple customers.

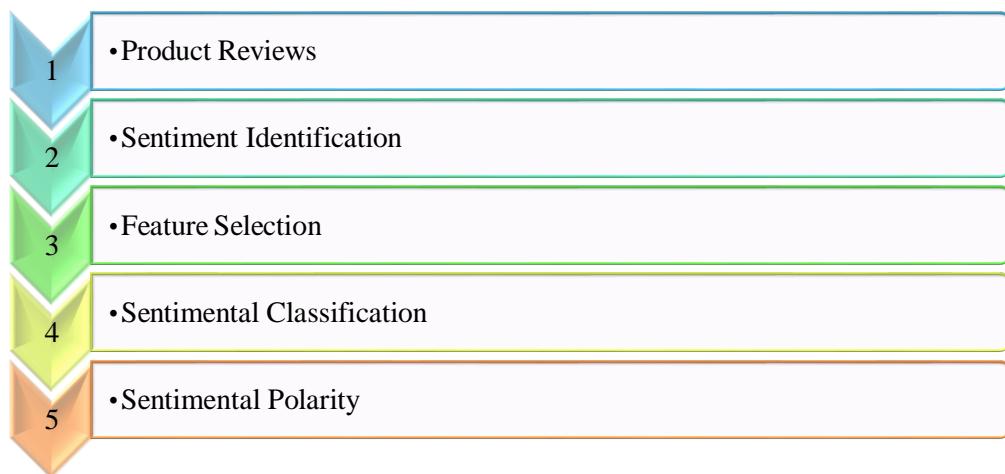
CONS

1. Sentiment analysis tools can identify and analyze many pieces of text automatically and quickly.
2. Computer programs have problems recognizing things like sarcasm and irony, negations, jokes, and exaggerations.
3. 'Disappointed' may be classified as a negative word for the purposes of sentiment analysis, but within the phrase "I wasn't disappointed", it should be classified as positive.

ASSUMPTIONS

1. The data should contain opinionated answers.
2. Large sample size.
3. The data should have polarity. i.e. positive, negative and neutral feedback.

STEPS TO BE IMPLEMENTED FOR SENTIMENT ANALYSIS

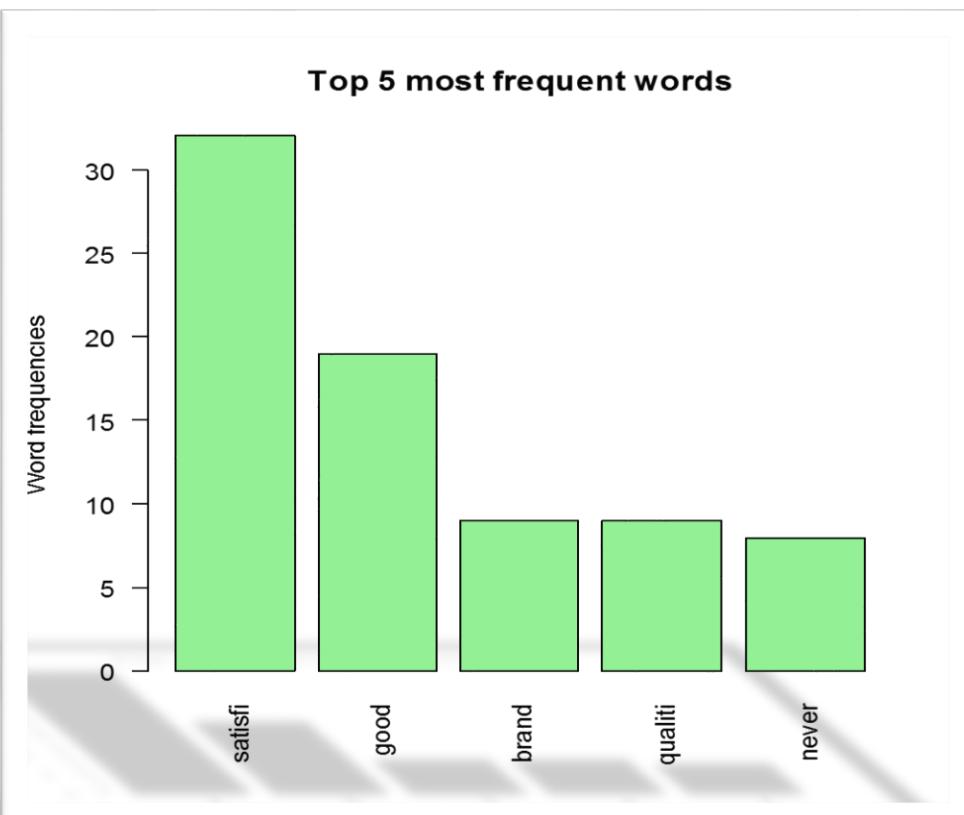


DATA ANALYSIS & INTERPRETATION

5 MOST FREQUENT WORDS

	Word	Frequency
satisfied	Satisfy	32
good	good	19
brand	brand	9
quality	quality	9
never	never	8

Top 5 most frequent words



Here,

- The most frequently occurring word is “satisfied”. Also notice that negative words like “never” don’t feature in the bar chart, which indicates there are no negative prefixes to change the context or meaning of the word “satisfied” (In short, this indicates most responses don’t mention negative phrases like “not good”, “bad” etc.).
- “Good”, “brand” and “quality” are the next three most frequently occurring words, which indicate that most people feel good about their work and their team’s health.
- Finally, the root “improv” for words like “improve”, “improvement”, “improving”, etc. is also on the chart, and you need further analysis to infer if its context is positive or negative.

WORD ASSOCIATION

Correlation is a statistical technique that can demonstrate whether, and how strongly, pairs of variables are related. This technique can be used effectively to analyze which words occur most often in association with the most frequently occurring words in the survey responses, which helps to see the context around these words.

\$work							
togeth							
0.4							
\$good							
integr	synergi						
0.28	0.28						
\$health							
declin	happen	noth	real	sentiment	suppli	wors	
0.29	0.29	0.29	0.29	0.29	0.29	0.29	
\$overal							
bad							
0.26							
\$great							
journey	satisfact	march	goal	pursu	toward	hard	
0.52	0.52	0.36	0.35	0.28	0.26	0.26	
\$feel							
across	board	harsh	system	somewhat			
0.33	0.32	0.32	0.32	0.29			
\$improv							
room	perfect	prop1	thik	attitud			
0.41	0.35	0.35	0.35	0.32			

SENTIMENT SCORES

Sentiments can be classified as positive, neutral or negative. They can also be represented on a numeric scale, to better express the degree of positive or negative strength of the sentiment contained in a body of text.

SYUZHET METHOD

Min	1 st Qu.	Median	Mean	3 rd Qu.	Max
-1.45	0.9	1.6	1.88	2.9	9

BING METHOD

Min	1 st Qu.	Median	Mean	3 rd Qu.	Max
-3	1	-3	2.07	3	9

AFFIN METHOD

Min	1 st Qu.	Median	Mean	3 rd Qu.	Max
-6	2	4	4.36	7	18

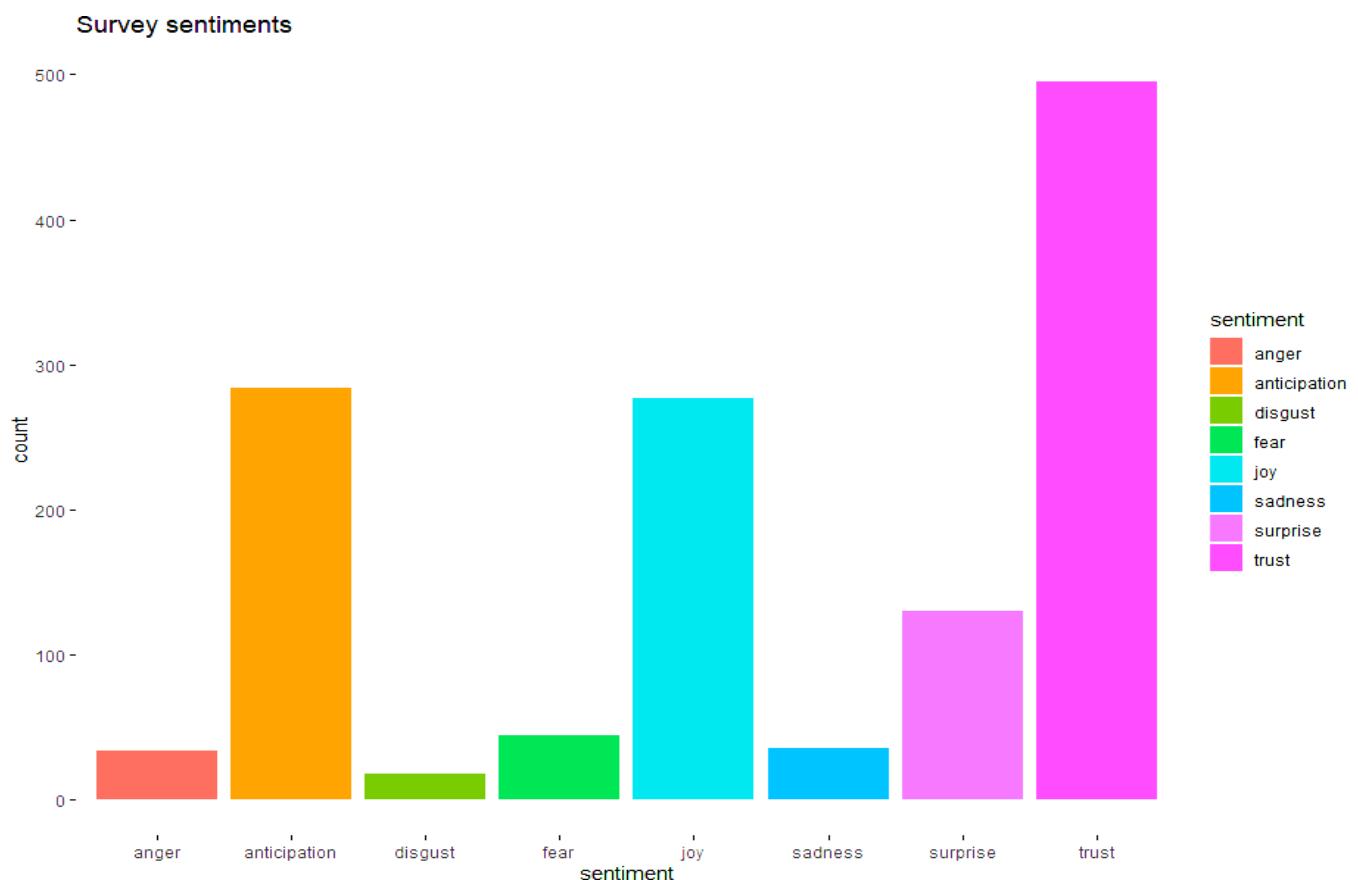
Since we have minimum positive and negative score from SYNZHET method, so this method is best for our case to proceed.

EMOTION CLASSIFICATION

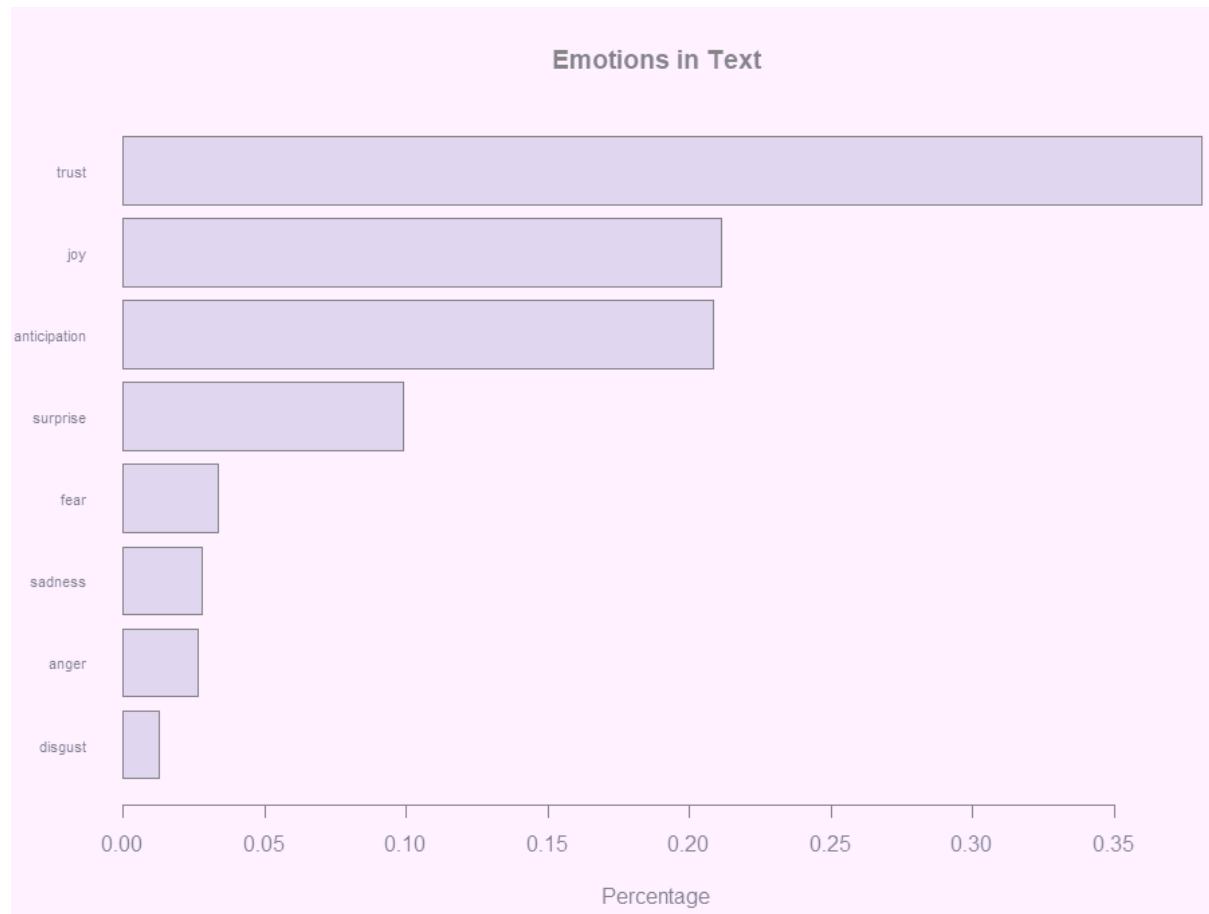
The data frame has ten columns (one column for each of the eight emotions, one column for positive sentiment valence and one for negative sentiment valence). The data in the columns (anger, anticipation, disgust, fear, joy, sadness, surprise, trust, negative, positive) can be accessed individually or in sets.

Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust	Negative	Positive
0	1	0	0	1	0	0	2	1	2
0	3	0	1	0	0	0	1	1	5
0	1	0	0	1	1	0	0	0	4
0	3	0	0	2	0	1	1	2	2
0	2	0	0	2	0	0	3	1	3
0	0	0	0	0	0	0	4	0	1
0	2	0	0	2	0	0	4	0	6
0	4	0	0	4	0	0	4	1	5
0	3	0	0	3	0	0	3	0	5
1	1	0	1	0	0	0	1	0	3

Bar Plot showing the count of words in the text, associated with each emotion



This bar chart demonstrates that words associated with the positive emotion of “trust” occurred about five hundred times in the text, whereas words associated with the negative emotion of “disgust” occurred less than 25 times. A deeper understanding of the overall emotions occurring in the survey response can be gained by comparing these number as a percentage of the total number of meaningful words.

Bar Plot showing the count of words associated with each sentiment expressed as a percentage.

This bar plot allows for a quick and easy comparison of the proportion of words associated with each emotion in the text. The emotion “trust” has the longest bar and shows that words associated with this positive emotion constitute just over 35% of all the meaningful words in this text. On the other hand, the emotion of “disgust” has the shortest bar and shows that words associated with this negative emotion constitute less than 2% of all the meaningful words in this text. Overall, words associated with the positive emotions of “trust” and “joy” account for almost 60% of the meaningful words in the text, which can be interpreted as a good sign of team health.

CONCLUSIONS

SWOT ANALYSIS

- ✓ Patanjali Ayurveda Ltd. have a strong impact on popularity due to the reputation of Baba Ramdev, patriotism, low prices of products & purity.
- ✓ The company needs to improve in the fields of manufacturing units, over dependency on Baba Ramdev, no longer penetration pricing, product dependence & lack of experienced graduates.
- ✓ The company have great opportunities in growth of organic sector, can be expanded in rural markets, it can be expanded globally, huge tie-ups and to be diversified.
- ✓ Price war, increasing competition, negative word- of mouth & poor reap can affect the company's growth & reputation.

CHI-SQUARE TEST OF INDEPENDENCE

- ✓ Patanjali product users (male or female) have no association with the health consciousness.
- ✓ That is, there is no association between gender & doctor's recommendation while buying the products.
- ✓ Also, there is no association between gender & side effects while using the Patanjali products.

FACTOR ANALYSIS

- ✓ **Consumer's Preferences:** Popularity, Purity, Brand ambassadors, Quality, Organic components, Increased promotion, Appealing to youngsters, Advertisements, Availability.
- ✓ **Consumer's Needs:** Low price, Beneficial, Customer Service.

SENTIMENTAL ANALYSIS

- ✓ The analysis shows that Patanjali users have a very high emotion of "Trust" on their products & a very low emotion of "Disgust" for the products.
- ✓ Hence, this is a good sign of team health.

LOGISTIC REGRESSION

- ✓ Patanjali have the highest influence on years of use (less than a year & more than 5 years), frequency of use(sometimes), consideration over other brands, popularity & recommendation to their friends for the overall satisfaction of their customers.

ARTIFICIAL NEURAL NETWORKS & K-NEAREST NEIGHBORS

- ✓ These techniques were conducted to predict if non-Patanjali users will be overall satisfied after using the Patanjali products.
- ✓ After checking the accuracy for both the techniques, K-NN having the highest accuracy, we conclude that all the non-Patanjali users will be overall satisfied after using the Patanjali products.

DECISION TREES

- ✓ After making predictions for training & testing sets, majority of the people will use Patanjali products sometimes, followed by people using them regularly and rarely.
- ✓ Less number of people will use them only on trial basis.

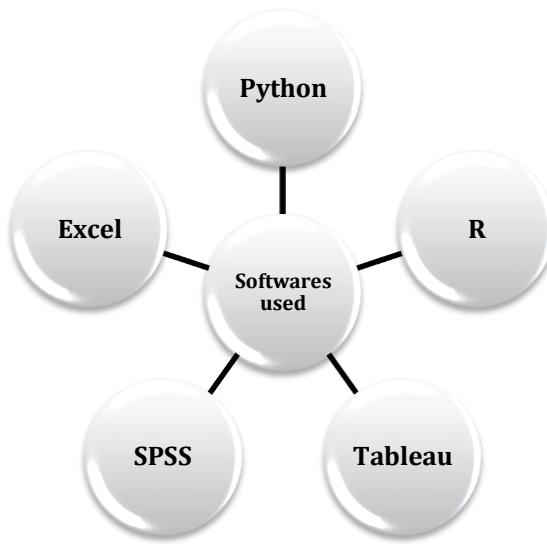
MARKET BASKET ANALYSIS USING APRIORI ALGORITHM

- ✓ The customers would prefer buying soaps & medicines after buying hair care & skincare products on their next purchase, as this technique shows the maximum association between these products.

PARETO ANALYSIS

- ✓ Health-wise, Quality-wise & Price-wise are the major factors that affects the preferences of people while buying the products.
- ✓ Lack of reasonable prices, lack of good quality, lack of health benefits & lack of Indian brand image are the reasons for not buying the Patanjali products.
- ✓ So as to improvise as a brand, Patanjali needs to improve the online services, Ayurvedic standards, customer satisfaction and Marketing strategy.

SOFTWARES USED & CODES



SENTIMENTAL ANALYSIS

Importing Libraries

```

library("tm")
library("SnowballC")
library("RColorBrewer")
library("syuzhet")
library("ggplot2")
  
```

#Reading the Data

```
text <- readLines(file.choose())
```

#Replacing "/", "@" and "|" with space

```

toSpace <- content_transformer(function (x , pattern ) gsub(pattern, " ", x))
TextDoc <- tm_map(TextDoc, toSpace, "/")
TextDoc <- tm_map(TextDoc, toSpace, "@")
TextDoc <- tm_map(TextDoc, toSpace, "\\|")
  
```

Convert the text to lower case

```
TextDoc <- tm_map(TextDoc, content_transformer(tolower))
```

Remove numbers

```
TextDoc <- tm_map(TextDoc, removeNumbers)
```

Remove english common stopwords

```
TextDoc <- tm_map(TextDoc, removeWords, stopwords("english"))
```

Remove punctuations

```
TextDoc <- tm_map(TextDoc, removePunctuation)
```

```

# Eliminate extra white spaces
TextDoc <- tm_map(TextDoc, stripWhitespace)

# Text stemming - which reduces words to their root form
TextDoc <- tm_map(TextDoc, stemDocument)

# Build a term-document matrix
TextDoc_dtm <- TermDocumentMatrix(TextDoc)
dtm_m <- as.matrix(TextDoc_dtm)

# Sort by decreasing value of frequency
dtm_v <- sort(rowSums(dtm_m),decreasing=TRUE)
dtm_d <- data.frame(word = names(dtm_v),freq=dtm_v)

# Display the top 5 most frequent words
head(dtm_d, 5)

#Plot the most frequent words
barplot(dtm_d[1:5]$freq, las = 2, names.arg = dtm_d[1:5]$word, col ="lightgreen", main = "Top 5 most frequent words", ylab = "Word frequencies")
findAssocs(TextDoc_dtm,terms=c("good","image","brand"),corlimit=0.25)
ssocs(textdoc_dtm, terms = c("good","image","brand"), corlimit =
0.25)
syuzhet_vector <- get_sentiment(text, method="syuzhet")

# see the first row of the vector
head(syuzhet_vector)

# see summary statistics of the vector
summary(syuzhet_vector)
bing_vector <- get_sentiment(text, method="bing")
head(bing_vector)
summary(bing_vector)

#affin
afinn_vector <- get_sentiment(text, method="afinn")
head(afinn_vector)
summary(afinn_vector)

#compare the first row of each vector using sign function
rbind( sign(head(syuzhet_vector)), sign(head(bing_vector)), sign(head(afinn_vector)))

# emotions, rather than a score:
# anger, anticipation, disgust, fear, joy, sadness, surprise, trust
# It also counts the number of positive and negative emotions found in each row
d<-get_nrc_sentiment(text)
# head(d,10) - to see top 10 lines of the get_nrc_sentiment dataframe
head (d,10)

```

```
#transpose
td<-data.frame(t(d))

#The function rowSums computes column sums across rows for each level of a grouping
#variable.
td_new <- data.frame(rowSums(td[2:253]))

#Transformation and cleaning
names(td_new)[1] <- "count"
td_new <- cbind("sentiment" = rownames(td_new), td_new)
rownames(td_new) <- NULL
td_new2<-td_new[1:8,]

#Plot One - count of words associated with each sentiment
quickplot(sentiment, data=td_new2, weight=count, geom="bar", fill=sentiment,
ylab="count")+ggtitle("Survey sentiments")

#Plot two - count of words associated with each sentiment, expressed as a percentage
barplot( sort(colSums(prop.table(d[]))), horiz = TRUE, cex.names = 0.7, las = 1, main =
"Emotions in Text", xlab="Percentage")
```

ARTIFICIAL NEURAL NETWORK

```
# Importing Packages
pip install tensorflow
pip install keras
pip install graphviz

#Importing Libraries
import pandas as pd
from sklearn import preprocessing
from sklearn.model_selection import train_test_split
from keras.models import Sequential
from keras.layers import Dense
from ann_visualizer.visualize import ann_viz
from keras.utils.vis_utils import plot_model
from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt
import seaborn as sns
```

```
#Reading the Data
data=pd.read_excel(r"Path\filename.xlsx")
data.head()
```

```
#Converting data into Dataset
dataset=data.values
dataset
```

#Pre-processing the Data

```
X=dataset[:,0:46]
y=dataset[:,46]
min_max_scaler=preprocessing.MinMaxScaler()
X_scale=min_max_scaler.fit_transform(X)
X_scale
```

#Splitting the data into training, testing & validation set

```
X_train,X_val_and_test,y_train,y_val_and_test=train_test_split(X_scale,y,test_size=0.2)
X_val, X_test, y_val, y_test=train_test_split(X_val_and_test,y_val_and_test,test_size=0.5)
print(X_train.shape,X_val.shape,X_test.shape,y_train.shape,y_val.shape,y_test.shape)
```

#Initializing the Artificial Neural Network & Creating the Input Layer, Hidden Layers & Output Layers

```
model=Sequential([
    Dense(32, activation='relu', input_shape=(46,)),
    Dense(32, activation='relu'),
    Dense(1, activation='sigmoid'),])
```

#Compiling the ANN Classifier

```
model.compile(optimizer='sgd',loss='binary_crossentropy',metrics=["accuracy"])
```

#Fitting the Model

```
hist=model.fit(X_train, y_train,
                batch_size=32, epochs=100,
                validation_data=(X_val,y_val))
```

#Checking the Accuracy

```
model.evaluate(X_test,y_test)[1]
```

#Plotting the model

```
plt.plot(hist.history['loss'])
plt.plot(hist.history['val_loss'])
plt.title('Model loss')
plt.xlabel('Epoch')
plt.ylabel('Loss')
plt.legend(['Train','Val'])
plt.show()
```

```
plt.plot(hist.history['accuracy'])
plt.plot(hist.history['val_accuracy'])
plt.title('Model accuracy')
plt.xlabel('Epoch')
plt.ylabel('accuracy')
plt.legend(['Train','Val'])
plt.show()
```

```
ann_viz(model,view=True,filename="network.gv")
```

```
plot_model(model,to_file="model_plot.png",show_shapes=True,show_layer_names=True)
```

```
#Predicting the Values
y_pred=model.predict_classes(X_test)
y_pred
# Creating Confusion Matrix
cm=confusion_matrix(y_test,y_pred)
cm
```

```
#Plotting Confusion Matrix
df=pd.DataFrame(cm)
print(df)
sns.heatmap(df,annot=True,cmap='BuGn')
plt.show()
```

K-NEAREST NEIGHBORS

#Importing Libraries & Packages

```
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
from sklearn import metrics
from sklearn.metrics import confusion_matrix
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import train_test_split
import seaborn as sns
sns.set()
```

#Importing the dataset

```
data=pd.read_excel(r'Path\filename.xlsx')
data.head()
```

#Preprocessing and Train Test Split

```
X = data.iloc[:, :-1].values
y = data.iloc[:, -1].values
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2,random_state=0)
```

#Accuracy check

```
neighbors=np.arange(1,45)
train_accuracy=np.empty(len(neighbors))
test_accuracy=np.empty(len(neighbors))
```

#Fitting the model

```
for i,k in enumerate(neighbors):
    knn=KNeighborsClassifier(n_neighbors=5)
    knn.fit(X_train,y_train)
    train_accuracy[i]=knn.score(X_train,y_train)
    test_accuracy[i]=knn.score(X_test,y_test)
```

#Plotting the graph

```
plt.plot(neighbors, test_accuracy,label='Testing dataset accuracy')
plt.plot(neighbors, train_accuracy,label='Training dataset accuracy')
```

```

plt.legend()
plt.xlabel('n_neighbors')
plt.ylabel('Accuracy')
plt.show()

#Making predictions
y_pred=knn.predict(X_test)

#Accuracy Score
print("Accuracy:",metrics.accuracy_score(y_test,y_pred))

#Accuracy Score using Confusion Matrix
confusion_matrix(y_test,y_pred)

#Plotting of Confusion Matrix
plot_confusion_matrix(knn,X_test,y_test)
plt.show()

#Making further predictions
data2=pd.read_excel(r'Path\filename.xlsx')
yhat=knn.predict(data2)
print(yhat)

```

DECISION TREES

#Importing Libraries & Packages

```

Import numpy as np
import pandas as pd
from sklearn.metrics import confusion_matrix
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score
from sklearn import metrics
from sklearn.metrics import classification_report

```

#Importing the dataset

```

data=pd.read_excel(r'Path\filename..xlsx')
data.head()

```

#Preprocessing and Train Test Split

```

feature_names=['Gender','Location','Income','TV','Radio','Friends and Neighbours',
'Newspaper', 'Internet', 'None of the above', 'Monthly Spent', 'Duration of Patanjali
products?']
x=data[feature_names]
y=data.iloc[:, -1]
X_train, X_test, y_train, y_test = train_test_split(x, y, test_size=0.3, random_state=1)

```

#Fitting the model

```

clf = DecisionTreeClassifier()
clf = clf.fit(X_train,y_train)

```

#Making predictions

```
y_pred = clf.predict(X_test)
train = pd.concat([X_train,y_train],axis=1)
train.head()
train["Predict"]= clf.predict(X_train)
train
train['Predict'].value_counts()
test=pd.concat([X_test,y_test],axis=1)
test.head()
test['Predict'].value_counts()
```

#Plotting Confusion Matrix

```
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(clf,X_test,y_test,cmap='afmhot_r')
import matplotlib.pyplot as plt
plt.show()
```

#Accuracy Score

```
print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
```

MARKET BASKET ANALYSIS USING APRIORI ALGORITHM

#Importing Libraries

```
pip install mlxtend
import pandas as pd
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules
```

#Importing Dataset

```
data=pd.read_excel('Market Basket .xlsx')
data.head()
```

#Checking Association

```
frequent_itemsets = apriori(data, min_support=0.07, use_colnames=True)
frequent_itemsets
rules = association_rules(frequent_itemsets, metric="lift", min_threshold=1)
rules
```

LOGISTIC REGRESSION- SPSS

CHI-SQUARE ANALYSIS- SPSS

FACTOR ANALYSIS-SPSS

PARETO ANALYSIS- MS EXCEL

QUESTIONNAIRE

1) Age

- Below 35 years
- 35-45years
- Above 45 years

1) Gender

- Male
- Female
- Other

2) What is your annual income (lacs)?

- Below 3
- 3-5
- 5-10
- Above 10

3) Location

- Central Line
- Western Line
- Harbour Line

4) What are your sources of awareness for the Patanjali products?

- TV
- Radio
- Friends and Neighbours
- Newspaper
- Internet
- None of the above

5) Have you ever used the brand Patanjali?

- Yes
- No

(IF YES)

1) Which of the following herbel brands do you use other than Patanjali?

- Dabur
- Hamdard
- Himalaya
- Charak
- Vicco
- Zandu
- Baidyanath
- Khadi Naturals
- None of the above
- Other (specify)

2) How often do you use Patanjali products?

- Only on trial basis
- Never
- Rarely
- Sometimes
- Regularly

3) How much do you spend monthly on Patanjali products?

- Less than Rs 500/-
- Rs 500/- to Rs 1000/-
- Rs 1000/- to Rs 2000/-
- Rs 2000/- and above

4) Since how long are you using the Patanjali products?

- Less than a year
- 1-3 years
- 3-5 years
- More than 5 years

5) Which all categories do you use in Patanjali?

- Skin care products
- Soaps
- Medicines
- Hair care products
- Dental Care products
- Spices
- Dairy products
- Bakery products
- Food and beverages
- Others

- 6) What are your reasons for buying Patanjali Products?
- Reasonable Price
 - Good Quality
 - Health Benefits
 - Availability
 - Brand Image
 - Indian Brand
 - Advertisements
 - Wide range of varieties
- 7) According to you, how are using Patanjali products beneficial?
- Health-wise
 - Price-wise
 - Brand-wise
 - Demand-wise
 - Quality-wise
 - Quantity-wise
- 8) According to you, in which category Patanjali needs improvement to be a better brand?
- Marketing Strategy
 - Ayurvedic Standards
 - Annual Sales
 - Online Services
 - Customer Satisfaction
 - Availability
 - Quality
- 9) Which of the following according to you, help build a good brand image for Patanjali?
- Quality
 - Competitive pricing
 - Communication skills
 - Good value added services
 - Free trials and discount
 - Others(specify)
- 10) On a scale of 1-5, please give your opinion on Patanjali products.
 (1-Strongly Disagree, 2-Disagree, 3-Neutral, 4-Agree, 5-Strongly Agree)

Sr No	Statement	Ratings
1	Patanjali products are gaining popularity in the market.	
2	Advertisements of Patanjali products are more visible in recent times.	

3	Patanjali gained reputation due to superior quality products and benefits.	
4	Patanjali products are easily available at better prices.	
5	Patanjali products symbolize purity.	
6	Patanjali products are safe because they are made from natural ingredients.	
7	Patanjali products are appealing to youngsters.	
8	Patanjali should increase the promotion and advertisements of their products.	
9	Patanjali has emerged as a powerful brand in terms of purity, prices and quality.	
10	Brand ambassadors affect the popularity of the products.	

11) According to you, answer in Yes or No.

Sr no	Statement	Yes	No
1	Would you consider Patanjali over any other competitive brands?		
2	Have you ever complained about a Patanjali product?		
3	Would you recommend your friends to use the Patanjali products?		
4	Do you prefer doctor's recommendation while choosing a product?		
5	Have you ever experienced any side effects while using the product?		
6	Patanjali Products gained reputation due to the popularity of Baba Ramdev?		

12) How much satisfied are you with the following characteristics. Rate your opinions.

Sr no	Characteristics	1	2	3	4	5
1	Low price					
2	Better quality					
3	Benefits					
4	Natural Products					
5	Customer Service					

13) Overall, are you satisfied with the Patanjali products?

- Yes
- No

(IF NO)

1) Do you use any herbal brand products?

- Yes
- No

2) Which are the herbel brand products that you use ?

- Dabur
- Hamdard
- Himalaya
- Charak
- Vicco
- Zandu
- Baidyanath
- Khadi Naturals
- None of the above
- Other(Specify)

3) Which all categories do you use in this brand?

- Skin care products
- Soaps
- Medicines
- Hair care products
- Dental Care products
- Spices
- Dairy products
- Bakery products
- Food and beverages
- Other(Specify)

4) What are your reasons for buying these Products?

- Reasonable Price
- Good Quality
- Health Benefits
- Availability
- Brand Image
- Indian Brand
- Advertisements
- Wide range of varieties

5) According to you, in which category this brand needs improvement?

- Marketing Strategy
- Ayurvedic Standards
- Annual Sales
- Online Services
- Customer Satisfaction

- Availability
- Quality

6) According to you, how are using these products beneficial?

- Health-wise
- Price-wise
- Brand-wise
- Demand-wise
- Quality-wise
- Quantity-wise

7) Which of the following according to you, help build a good brand image for this brand?

- Quality
- Competitive pricing
- Communication skills
- Good value added services
- Free trials and discount
- Other (specify)

8) How much satisfied are you with the following characteristics. Rate your opinions.

Sr no	Characteristics	1	2	3	4	5
1	Low price					
2	Better quality					
3	Benefits					
4	Natural Products					
5	Customer Service					

9) What are your reasons for not buying the Patanjali products? (Specify)

10) Overall, are you satisfied using the product this brand?

- Yes
- No

REFERENCES

- <http://www.analyticsvidhya.com/>
 - <https://stackoverflow.com/>
 - <https://www.patanjaliayurved.net/>
 - <https://bstrategyhub.com/patanjali-swot-analysis-2019-swot-analysis-of-patanjali/>
 - <https://github.com/>
 - <https://libguides.library.kent.edu/spss/chisquare#:~:text=The%20Chi%2DSquare%20Test%20of%20Independence%20determines%20whether%20there%20is,Chi%2DSquare%20Test%20of%20Association.>
 - <https://www.kdnuggets.com/2015/05/top-10-data-mining-algorithms-explained.html>
 - <https://pythonprogramminglanguage.com/top-data-mining-algorithms/>
 - https://www.google.com/search?q=logistic+regression+using+spss&rlz=1C1CHBD_enIN710IN711&ei=uRZnYIrzCMmI4-EPy_uo-A4&oq=logistic+regression+using+sp&gs_lcp=Cgdnd3Mtd2l6EALEYADICCAAyAggAMgIIADIGCAAQFhAeMgYIABAWE4yBggAEBYQHjIGCAAQFhAeMgYIABAWE4yBggAEBYQHjIGCAAQFhAeOggIABCRAhCLAz0ECAAQQzoICAAQsQMQgwE6BQgAELEDOgUILhCxAz0HCAAQQxCLAz0FCAAQkQI6CwgAELEDEIMBEIsD0gcIABAKEJECOGoIABCxAxBDEIsD0ggIABCxAxCLAz0QCC4QsQMQxwEOrwEOQxCLAz0HCC4QQxCLAz0HCAAQsQM0QzoECC4QQzoCCC5QIBIYsF9gh2xoAXACeACAAfBiAGBJJIBBjAuMjYuM5gBAKABAaoBB2d3cy13aXqwAQC4AQOLAAQE&sclient=gws-wiz
 - <https://www.geeksforgeeks.org/machine-learning/>