



# D-MART, A MARKET REVOLUTION, OR JUST ANOTHER SHOP

Submitted In Partial Fulfilment of Requirements  
For the Degree Of

**Master of Science  
(Statistics)**

Guide  
**Prof. Charulata Avhad**  
Department of Statistics  
S.K.Somaiya College

By  
**Ronak Ashok Kalantre (31031821012)**  
**Harshal Namdev Kothavade(31031821019)**  
**Hasmina Bee Laskar(31031821022)**  
**Soham Vinayak Mali(31031821024)**  
**Nishant Sharad Thanekar(31031821040)**  
**Prathamesh Milind Zinge(31031821043)**



Somaiya Vidyavihar University  
Vidyavihar East, Mumbai 400077

**2022-2023**



**A market revolution,  
or just another shop?`**



# D-MART, A MARKET REVOLUTION, OR JUST ANOTHER SHOP

Submitted In Partial Fulfilment of Requirements  
For the Degree Of

**Master of Science  
(Statistics)**

Guide

**Prof. Charulata Avhad**

Department of Statistics  
S.K.Somaiya College

By

**Ronak Ashok Kalantre (31031821012)**

**Harshal Namdev Kothavade(31031821019)**

**Hasmina Bee Laskar(31031821022)**

**Soham Vinayak Mali(31031821024)**

**Nishant Sharad Thanekar(31031821040)**

**Prathamesh Milind Zinge(31031821043)**



Somaiya Vidyavihar University  
Vidyavihar East, Mumbai 400077

**2022-2023**

## Student details

<b>Name</b>	Ronak Ashok Kalantre Harshal Namdev Kothavade Hasmina Bee Laskar Soham Vinayak Mali Nishant Sharad Thanekar Prathamesh Milind Zinge
<b>Subject</b>	Statistics
<b>Institution</b>	S K Somaiya College
<b>Year</b>	2022-2023
<b>Teacher in Charge</b>	Prof. Charulata Avhad
<b>Title of the Project</b>	D-MART, A MARKET REVOLUTION, OR JUST ANOTHER SHOP ?
<b>Location</b>	Vidyavihar
<b>Duration</b>	Jan 2023-April 2023
<b>Signature of the Teacher</b>	
<b>Signature of the Coordinator of Department (STATISTICS)</b>	

## **CERTIFICATE OF AUTHENTICATION**

This is to certify that the project entitled “*D-MART,A MARKET REVOLUTION,OR JUST ANOTHER SHOP ?*” is a bonafide work of “ *Ronak Ashok Kalantre*”(31031821012) “*Hasmina Bee Laskar*” (31031821022 ) ,”*Nishant Sharad Thanekar*”(31031821040), ”*Harshal Namdev Kothavade*”(31031821019), ”*Soham Vinayak Mali*”(31031821024), ”*Prathamesh Milind Zinge*”(31031821043)submitted to the S K Somaiya College in partial fulfillment of the requirement for the award of the degree of “M.Sc. in the subject of Statistics”.

I considered that the dissertation has reached the standards and fulfilling the requirements of the rules and regulations relating to the nature of the degree. The contents embodied in the dissertation have not been submitted for the award of any other degree or diploma in this or any other university.

**Date:**

**Place:**

(Name and sign)  
sign)  
External Mentor/supervisor  
mentor/supervisor

(Name and  
Internal

(Name and Sign)  
Head of the department

## **Declaration by the student**

we certify that,

- a) The work contained in the thesis is original and has been done by myself under the supervision of my supervisor.
- b) The work has not been submitted to any other Institute for any degree or diploma.
- c) I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.
- d) Whenever I have used materials (data, theoretical analysis, and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references.
- e) Whenever I have quoted written materials from other sources and due credit is given to the sources by citing them.
- f) From the plagiarism test, it is found that the similarity index of whole thesis within 25% and single paper is less than 10 % as per the university guidelines.

**Date:**

**Place:**

-----  
-----  
**Signature**

---

Name of the  
student with  
Exam.  
Seat No.



**SOMAIYA**  
VIDYAVIHAR UNIVERSITY

S K Somaiya College

## **Department of Statistics**

### **CERTIFICATE**

This is to certify that  
Ronak Kalantre  
Harshal Kothavade  
Hasmina Laskar  
Soham Mali  
Nishant Thanekar  
Prathamesh Zinge

of **M.Sc. Statistics**, has satisfactorily completed the Project titled  
**D-MART, A MARKET REVOLUTION, OR JUST ANOTHER SHOP**

for the Partial fulfilment of the Degree by the Somaiya Vidyavihar University, during  
the Academic year **2022-23**.

**Signature of the Teacher In-Charge**

**Signature of the Coordinator**

**Signature of the Examiner/s**

**Date of Examination**

**College**

**Seal**



## **Examiner Approval sheet**

This dissertation/project report “*D-MART,A MARKET REVOLUTION,OR JUST ANOTHER SHOP?*” done by *Ronak Kalantre ,Hasmina Laskar,Soham Mali ,Harshal Kothaavade,Nishant Thanekar,Prathamesh Zinge* is approved for the degree of Master of Science in the subject of Biotechnology.

Examiners

(Name and signature)

1.-----

(Name and signature)

2.-----

Place:

Date:

# Acknowledgement

We would like to take this opportunity to express our heartfelt gratitude toward our Director, Monica Lodha, for her unwavering support and encouragement throughout our project. We would also like to extend our appreciation to Mrs. Jyoti Mantri, Program In-charge and Head of the Department of Statistics at S.K Somaiya College, for her constant guidance and support.

We would like to acknowledge the efforts of our teachers, Mrs. Jyoti Mantri, Miss Manisha Gupta and Mrs. Charulata Avhad, for their dedication and motivation toward keeping us focused and inspired. We are genuinely grateful to our mentor, Mrs. Charulata Avhad, for her continuous guidance, immense knowledge, and unwavering patience throughout the research. Her enthusiasm and motivation have helped us to explore and learn new things, which have been instrumental in the accomplishment of our project.

We owe our sincere gratitude to all the individuals who have contributed and extended their valuable assistance whenever we needed it, without which this research project would not have been possible. Lastly, we would like to thank our family members, friends, and well-wishers who have supported us throughout this journey of study. Their constant encouragement and belief in us have been a source of inspiration and motivation.

Hasmina Laskar  
Nishant Thanekar  
Ronak Kalantre  
Harshal Kothavade  
Soham Mali  
Prathamesh Zinge

<b>SR.NO</b>	<b>CONTENT</b>	<b>PAGE.NO</b>
1	Title Page	1
2	Student Details	3
3	Certificate of authentication	9
4	Declaration by student	10
5	Department Certificate	11
6	Examiner approval sheet	15
7	Acknowledgement	17
8	Abstract	1
Chapter1	Introduction 1.1 Introduction 1.2 Research methodology 1.3 Sampling technique 1.4 Objectives of study 1.5 Variables	2
Chapter2	2.1 Graphical representation	8
Chapter3	SWOT Analysis	22
Chapter 4	Chi- square test Objective 1.1 Objective 1.2 Objective 1.3	25

Chapter 5	Nearest Neighbour Objective 2	31
Chapter 6	Clustering Objective 3	35
Chapter 7	Market Basket Analysis Objective 4	41
Chapter 8	Factor Analysis Objective 5.1 Objective 5.2	45
Chapter 9	Pareto Analysis Objective 6.1 Objective 6.2	56
Chapter 10	Sentiment Analysis Objective 7	60
Chapter 11	Conclusion Codes Bibliography	66
Chapter 12	References	82

## **EXECUTIVE SUMMARY & ABSTRACT**

---

### **EXECUTIVE SUMMARY**

D Mart is a retail chain that was founded by Radhakishan Damani in the year 2002. The first D Mart store was opened in Powai, Mumbai, with the aim of offering high-quality products at competitive prices to customers. Over the years, D Mart has grown rapidly and expanded its operations to various cities in India. Today, it has over 200 stores across the country, making it one of the largest retail chains in India. The success of D Mart can be attributed to its founder, Radhakishan Damani, who is a renowned entrepreneur and investor. He is known for his astute business acumen and is often referred to as the "Retail King" of India. Under his leadership, D Mart has grown from a single store to a chain with a widespread presence across India.

### **ABSTRACT**

The main aim of this paper is to find the consumer perception regarding D-Mart through a survey. 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 D-Mart products. Further, the study brought out the factors that are most important in the decision to purchase D-Mart products, such as strategy, product quality, pricing and brand image, and consumer awareness about the products.

## **INTRODUCTION & NEED FOR STUDY**

---

### **INTRODUCTION**

Shopping has always been an integral part of our lives. From the necessities to luxury goods, consumers are constantly looking for products and services that cater to their needs and wants. As the world has become increasingly globalized, the way consumers shop has also undergone a significant transformation. With the rise of e-commerce, consumers now have access to a wider range of products and services, from anywhere in the world, at any time of the day. The internet has not only transformed the way we shop, but also the way we interact with brands and make purchasing decisions.

In this study, we will explore the relationship between consumers and shopping, and how it has evolved over time. We will examine the factors that influence consumer behavior and decision-making, as well as the impact of technology on the shopping experience. We will also explore the role of marketing and advertising in shaping consumer perceptions and preferences, and the ethical considerations that arise in the process.

The choice of a particular brand by the consumer over 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.

D-Mart's business model is centered around providing its customers with high-quality products at affordable prices. The company operates on a low-cost, high-volume approach, which means that it sells a large quantity of products at low margins. D-Mart's focus on low margins has allowed it to offer its products at lower prices than its competitors, which has helped it attract price-conscious customers.

Another important aspect of D-Mart's business model is its focus on cost-efficiency. The company operates on a lean operating model, which means that it keeps its overheads low by limiting its spending on advertising, marketing, and store decor. D-Mart's stores are designed to be functional rather than fancy, which helps the company save money on store design and maintenance.

D-Mart has employed several strategies to become successful. One of its key strategies is its focus on private label products. D-Mart sells a large number of private label products, which are products that are manufactured specifically for the company. By selling private label products, D-Mart is able to offer its customers high-quality products at lower prices than branded products. Private label products also help D-Mart differentiate itself from its competitors and create a loyal customer base.

Another strategy that has contributed to D-Mart's success is its focus on operational efficiency. The company has invested heavily in technology and automation to streamline its operations and reduce costs. For example, D-Mart has implemented a sophisticated inventory management system that helps it keep track of its stock levels and reduce wastage. The company has also invested in logistics and distribution to ensure that its products are delivered to its stores in a timely and cost-effective manner.

D-Mart has also been successful in expanding its business by opening new stores in strategic locations. The company has focused on opening stores in Tier 2 and Tier 3 cities, where there is a large untapped market for retail. By targeting these markets, D-Mart has been able to establish itself as a leading retailer in these cities and capture a large market share.

D-Mart's success is also attributed to its practices, which are designed to create a positive customer experience. The company has a customer-centric approach, which means that it focuses on providing its customers with a seamless shopping experience. D-Mart's stores are designed to be easy to navigate, and its staff is trained to provide helpful and friendly customer service.

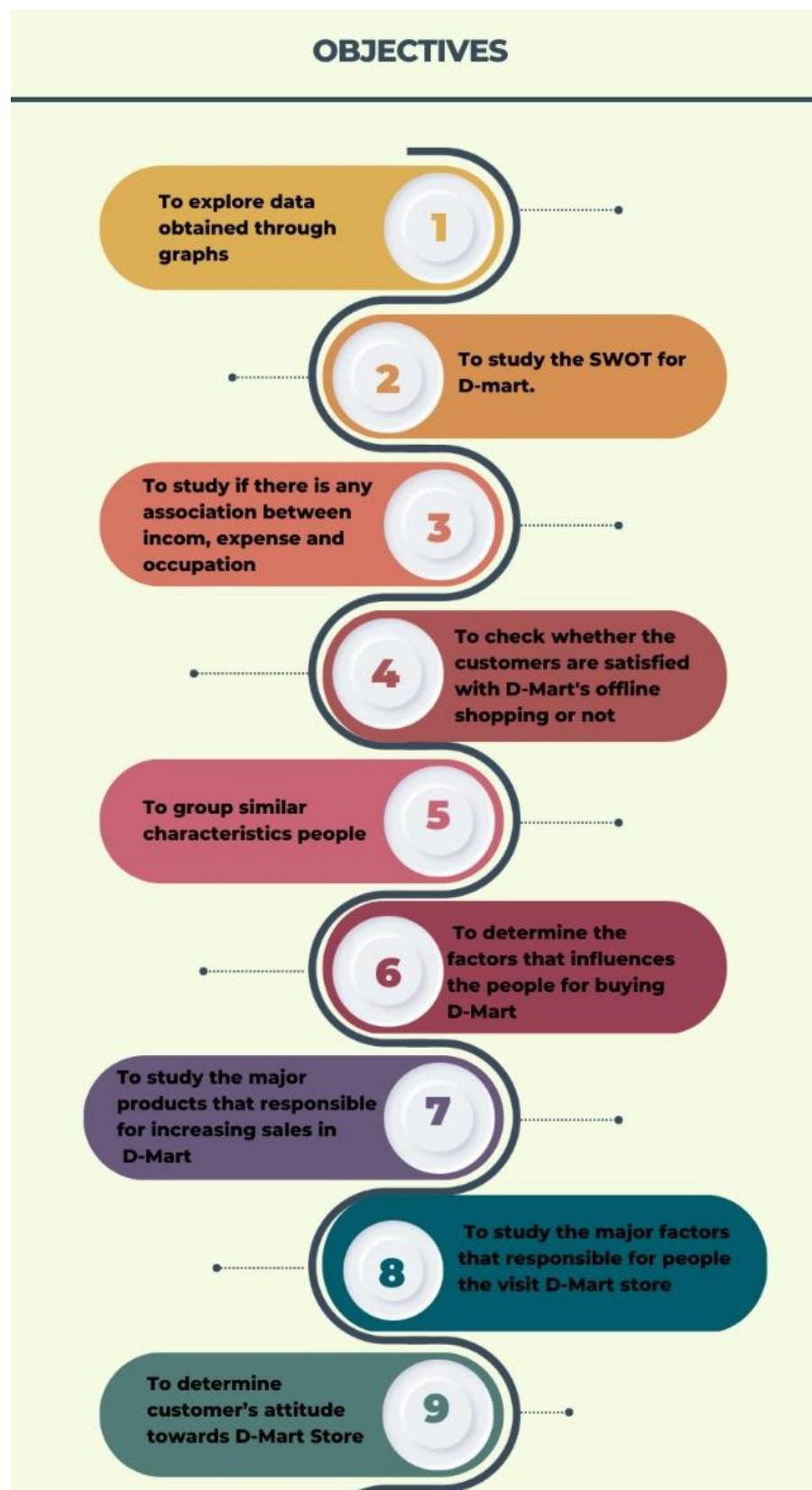
Another practice that has contributed to D-Mart's success is its focus on quality. The company has a rigorous quality control process in place, which ensures that its products meet high-quality standards. D-Mart sources its products directly from manufacturers, which allows it to ensure that its products are of high quality and meet its customers' expectations.

D-Mart also places a strong emphasis on employee training and development. The company invests in its employees by providing them with regular training and development opportunities. This helps to ensure that its staff is knowledgeable about its products and able to provide excellent customer service.

## **NEED FOR STUDY**

Companies need to understand the behavior of their customers to effectively market their products and services. By studying consumer behavior, companies can identify the needs and preferences of their target audience, which helps them to design their products and marketing strategies. This study will reveal the different aspects of consumers' perceptions regarding price, quality, range, availability, and advertisements of the products. The need for the study is essential as the competition in the 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 D-Mart

# OBJECTIVES



## LITERATURE REVIEW

---

We carried a study on consumer perception regarding D-Mart and its reputation amongst Indian consumers. The objective of the study was to know about the perception, satisfaction level, and attributes of consumers with regards to D-mart products. The data was collected from 403 respondents and conclusions were drawn looking at the data.

The questionnaire had questions for both the ones who preferred shopping at D-Mart and for those who preferred shopping at other places. Out of 403 respondents, 300 preferred D-Mart while 103 preferred other places. We performed analysis on the obtained data through techniques like SWOT analysis, Factor analysis, chi-square, Market Basket Analysis etc. to draw conclusions as accurate as possible. 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 the preference of the consumer's and their needs while choosing the product.

Through Sentimental analysis we got to know that people trust the brand and hence choose D-Mart products. K-NN was used to predict whether customers preferred online or offline shopping, we also used it to check accuracy. Decision Tree was used to predict how often customers would choose D-mart to shop. Market Basket Analysis was used to check the association between the grocery and other staples. Pareto Analysis was done to understand which sector had the major influence on sales of D-Mart products.

## **METHODOLOGY & SAMPLING TECHNIQUE**

---

### **RESEARCH METHODOLOGY**

For collecting primary data, a survey approach was used and for this purpose a research questionnaire consisting of a set of questions was presented to the respondents to know their perception towards D-Mart products. We then conducted a pilot survey for a sample size of 50 personally & made the necessary changes. We added a review question with personal feedback, we also added some textbox questions for more detailed analysis. We surveyed a total of 403 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.

### **SAMPLING TECHNIQUE**

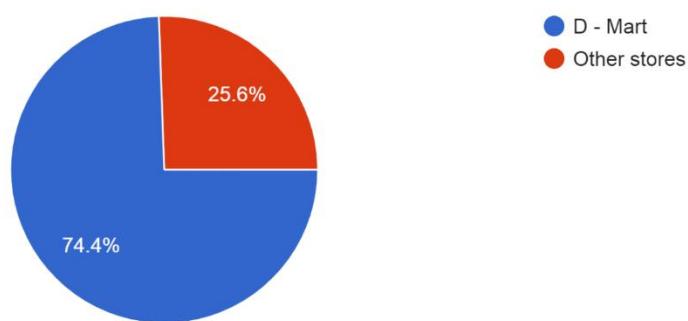
In present study convenience sampling was used to collect the data from the defined universe. Convenience sampling (also known as availability sampling) is a specific type of non-probability sampling method that relies on data collection from population members who are conveniently available to participate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

## GRAPHICAL REPRESENTATIONS

---

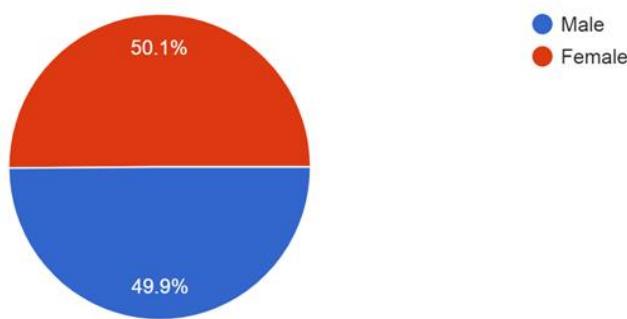
+

Which store comes in your mind when you think of a purchasing a product ?  
403 responses



- Out of 403 participants, 300 participants thought of purchasing the product from D-Mart and 103 participants thought of purchasing the product from other stores .

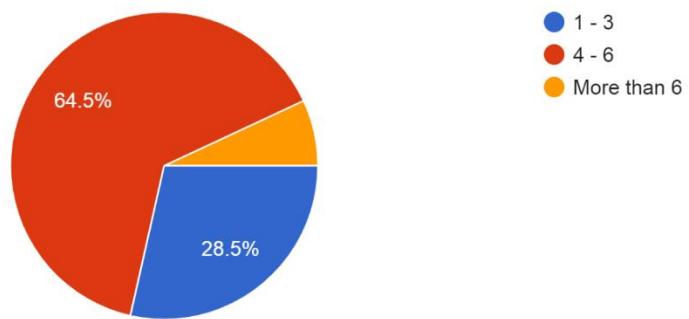
Gender  
403 responses



We have almost equal responses from male and female with 50.1% & 49.9% respectively.

How many members are in your family ?

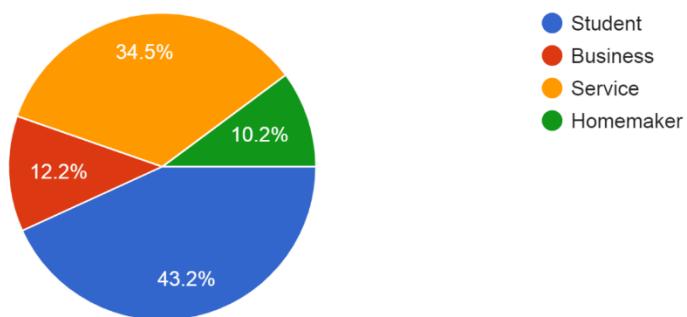
403 responses



Families with 1-3 members are 28.5%, large amount of response are from families with 4-6 members 64.5% and only 7% families are of more than 6 members.

Occupation

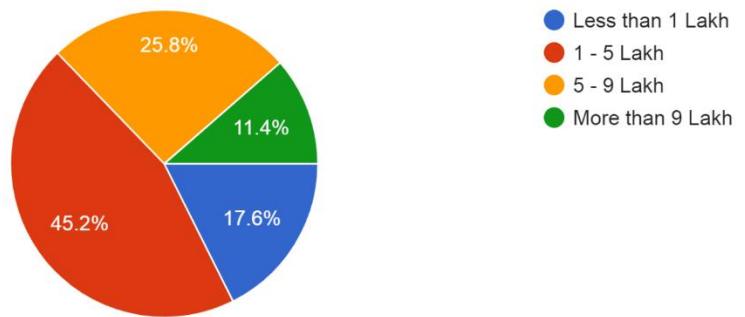
403 responses



We have 43.2% responses from students, 34.5% responses from services, 12.2% responses from businesspersons and 10.2% are homemakers. Majority of our responses are from students are least from homemakers.

### Annual income of the family ?

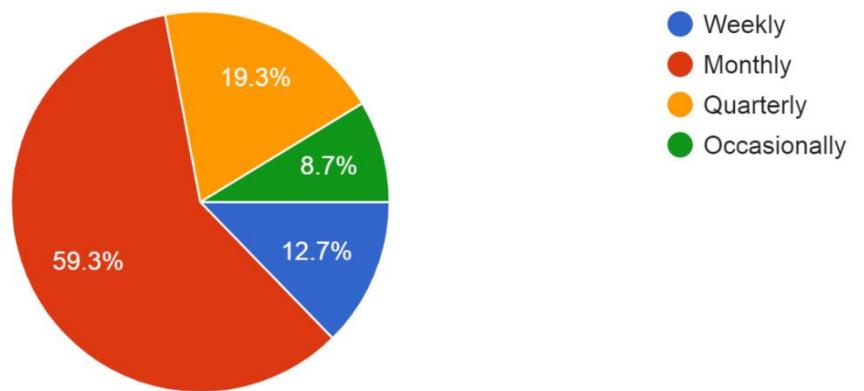
403 responses



Our majority responses have family income within 1-5 lakhs with 45.2% and the least are of annual income of more than 9 lakhs with 11.4%. Annual income of less than 1 lakh and 5-9 lakhs are with the percentage 17.6 and 25.8 respectively.

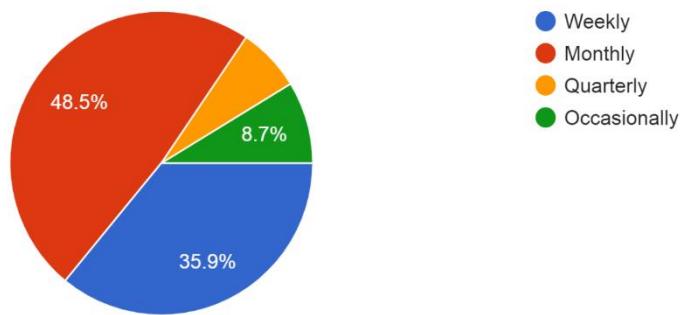
### How frequently do you visit D-Mart store ?

300 responses



### How frequently do you visit Other store ?

103 responses

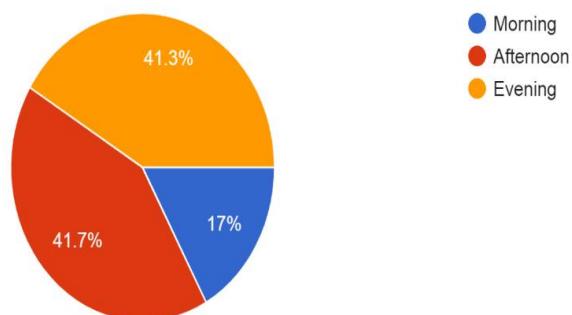


59.3% of the visitors visit D-mart on monthly basis to shop.	48.5% of the visitors visit other stores on monthly basis to shop.
12.7% of the visitors visit D-mart on weekly basis to shop.	35.9% of the visitors visit other stores on weekly basis to shop.
19.3% of the visitors visit D-mart on monthly basis to shop.	6.9% of the visitors visit other stores on monthly basis to shop.
8.7% of the visitors visit D-mart on occasionally basis to shop.	8.7% of the visitors visit other stores on occasionally basis to shop.

0

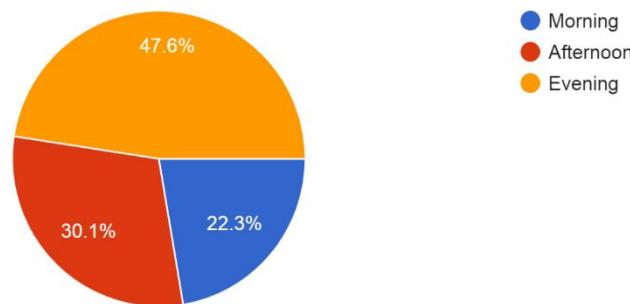
### What time of you prefer purchasing D-Mart ?

300 responses



What time of you prefer purchasing Other Stores ?

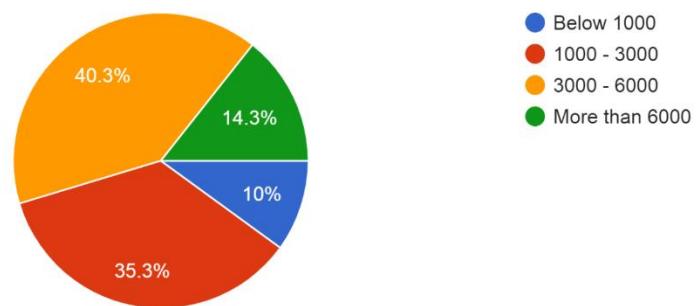
103 responses



17% of the people shop in morning in D-mart.	22.3% of the people shop in morning in other stores.
41.7% of the people shop in afternoon in D-mart.	30.1% of the people shop in afternoon in other stores.
41.3% of the people shop in evening in D-mart.	47.6% of the people shop in evening in other stores.

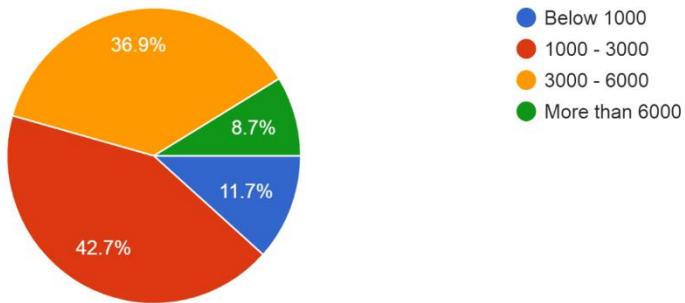
What is your average monthly expenditure while shopping in D-Mart ?

300 responses



What is your average monthly expenditure while shopping in Other Stores ?

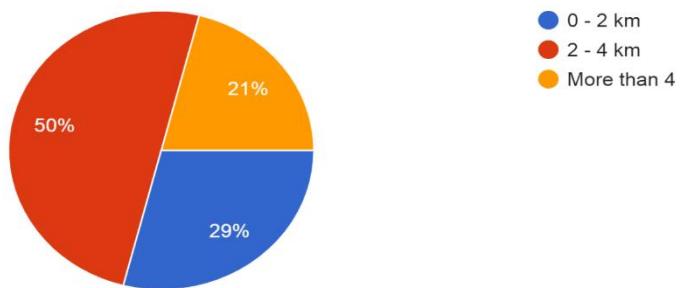
103 responses



10% of the visitors visiting D-mart shops are for price below 1000 Rs.	11.7% of the visitors visiting other stores shops are for price below 1000 Rs.
35.3% of the visitors visiting D-mart shops are for price between 1000-3000 Rs.	42.7% of the visitors visiting other stores shops are for price between 1000-3000 Rs.
40.3% of the visitors visiting D-mart shops are for price between 3000- 6000Rs.	36.9% of the visitors visiting other stores shops are for price below 3000- 6000 Rs.
14.3% of the visitors visiting D-mart shops are for price more than 6000 Rs.	8.7% of the visitors visiting other stores shops are for price more than 6000 Rs.

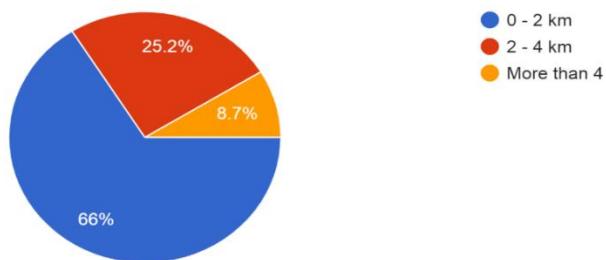
How much distance do you travel to reach the store ?

300 responses



How much distance do you travel to reach the Other Stores ?

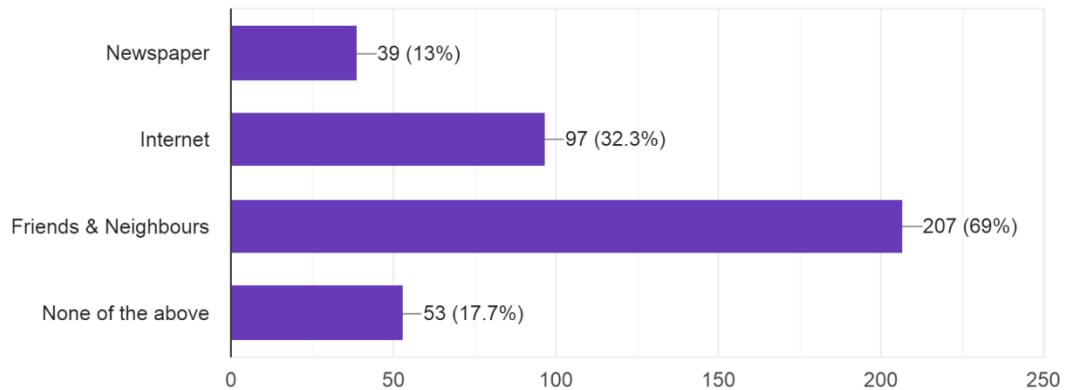
103 responses



29% of the people live at 0-2 km from the D-mart	56 % of the people live at 0-2 km from other stores.
50% of the people live at 2-4 km from the D-mart.	25.2 % of the people live at 2-4 km from other stores.
21% of the people live at more than 4 km from the D-mart.	8.7 % of the people live at more than 4 km from other stores.

What are your sources of awareness for D-Mart ?

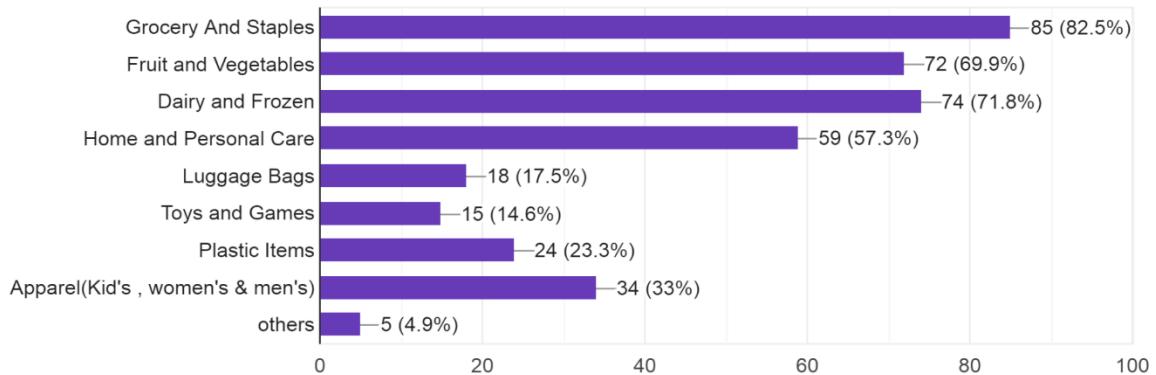
300 responses



The majority of the people, up to 69% of the population heard of D-mart from their friends and neighbors. 32.2% of the people know about it from the internet and 13% of the population got to know about it from the internet. About 17.7% of the population got information about it from other sources. As we see here, most of the people got to know about D-mart from friends and neighbors, its mouth publicity is good and effective.

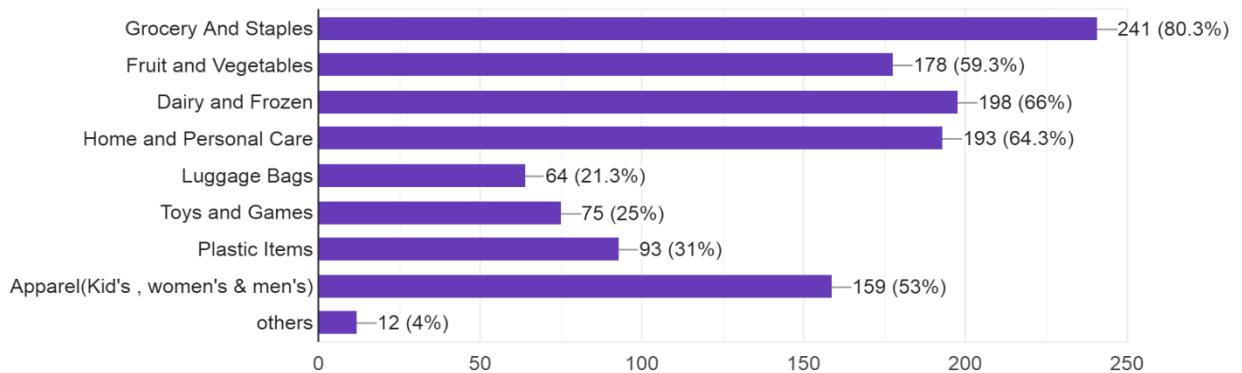
Tick the column according to your buying behavior for the following categories of product ?

103 responses



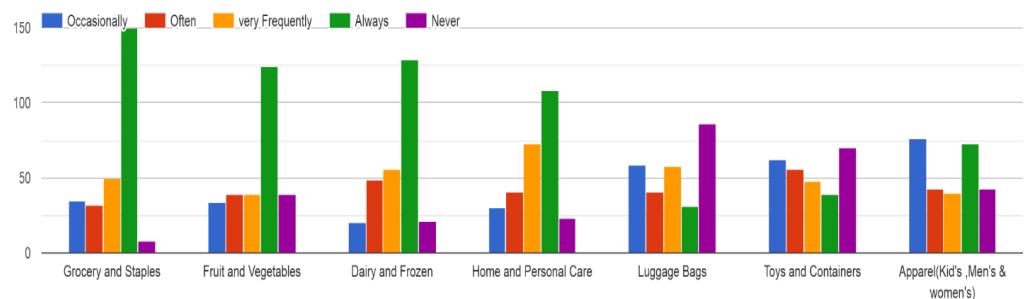
Tick the column according to your buying behavior for the following categories of product ?

300 responses

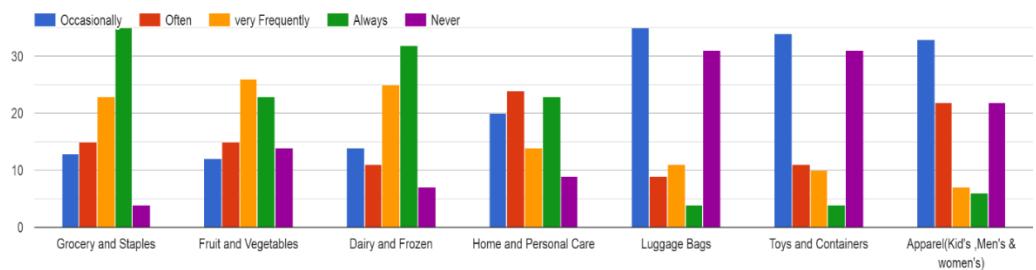


The purchase of grocery and staple products is high in both D-mart and other stores with total 80.3% and 82.5% respectively. Fruits and vegetables are more purchased in other stores than D-mart with 69.9% and 59.3% respectively. Dairy and frozen products are also purchased in good amount with 66% & 71.8% in D-mart & other stores respectively.

Tick the column according to your buying behavior for the following categories of product ?

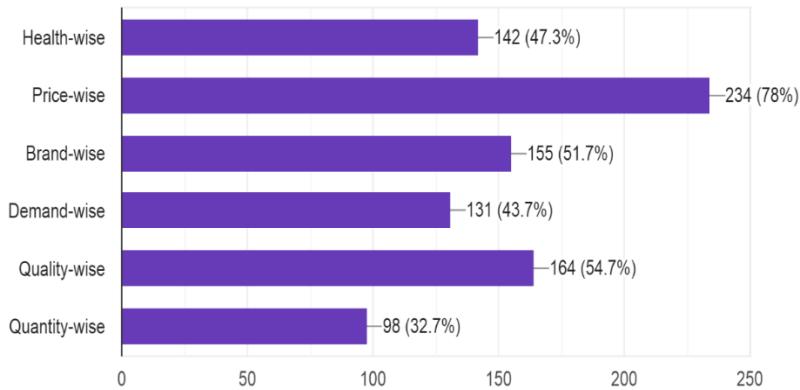


Tick the column according to your buying behavior for the following categories of product ?



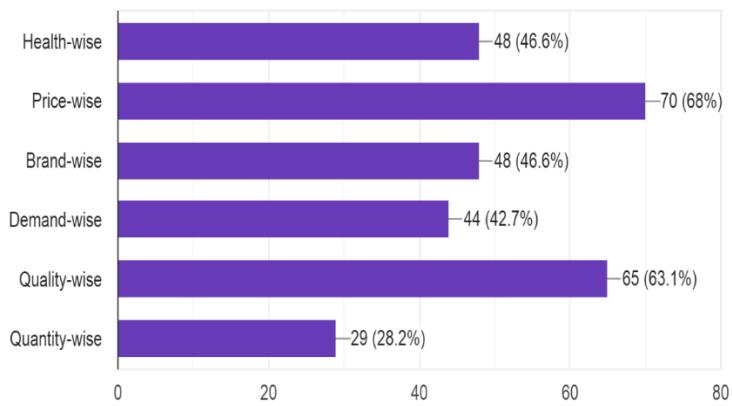
### Reasons for buying D-Mart products ?

300 responses



### Reasons for buying Other Stores products ?

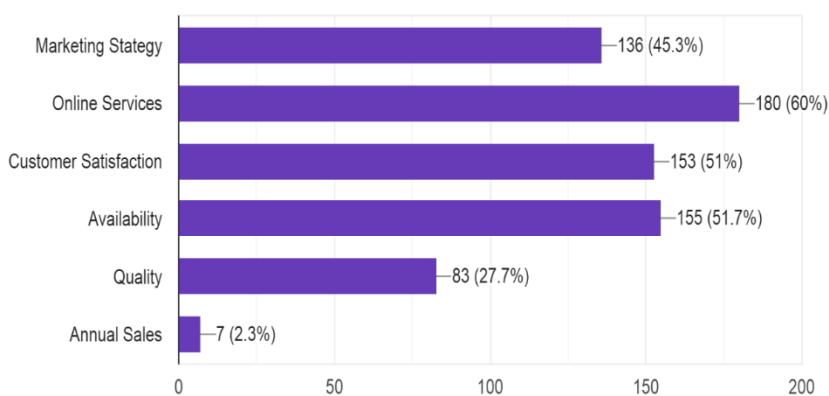
103 responses



We can compare by looking at the graphs of D-mart and other stores that the best reason for purchasing from both the stores is its price with 78% and 68% respectively. the reason for health wise to purchase the product is 47.3% for D-mart and 48.6% for other stores. The reason that brand influences the purchase of the product is 51.7% in D-mart and 46.6% in other stores. The demand wise purchase of the products is 43.7% in D-mart and 42.75 in other stores. The quality wise purchase of the product is 54.7% in D-mart and 63.1% in other stores. Quantity wise 32.7% and 28.2% of products are being purchased in D-mart and other stores respectively.

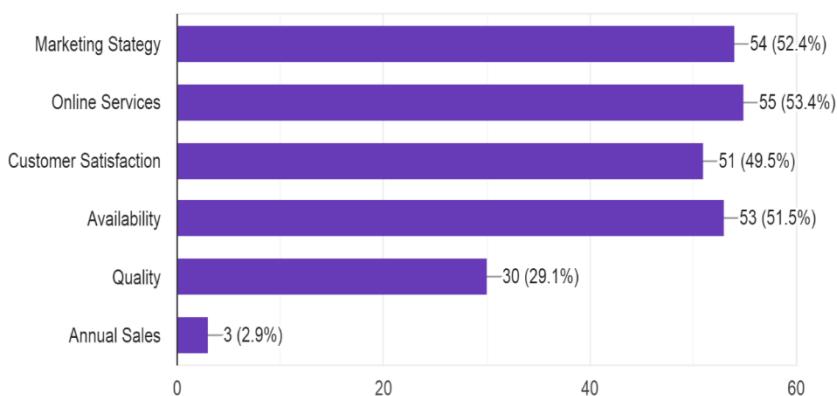
According to you , in which category D-Mart needs to improve to be better brand ?

300 responses



According to you , in which category Other Stores needs to improve to be better brand ?

103 responses



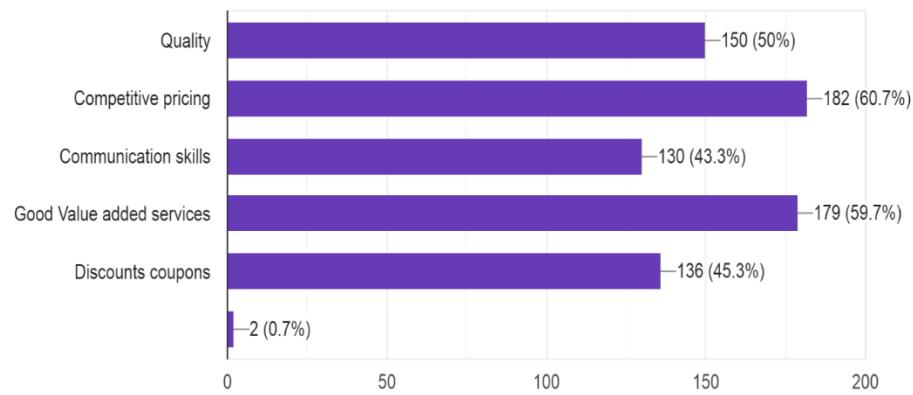
	D-mart	Other stores
<b>Marketing strategy</b>	<b>45.3%</b>	<b>52.4%</b>
<b>Online Services</b>	<b>60%</b>	<b>53.4%</b>

<b>Customer Satisfaction</b>	<b>51%</b>	<b>49.5%</b>
<b>Availability</b>	<b>51.7%</b>	<b>51.5%</b>
<b>Quality</b>	<b>27.7%</b>	<b>29.1%</b>
<b>Annual Sales</b>	<b>2.3%</b>	<b>2.9%</b>

Online services can improve both D-mart and other stores with 60% & 53.4% respectively.

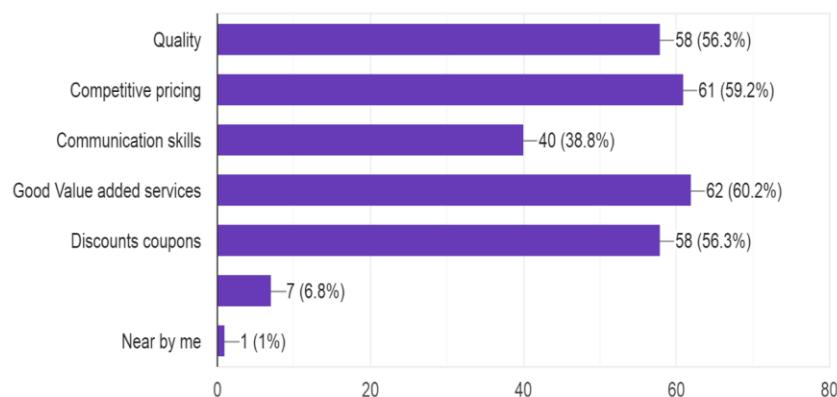
Which of the following according to you, will help build the good image for D-Mart ?

300 responses



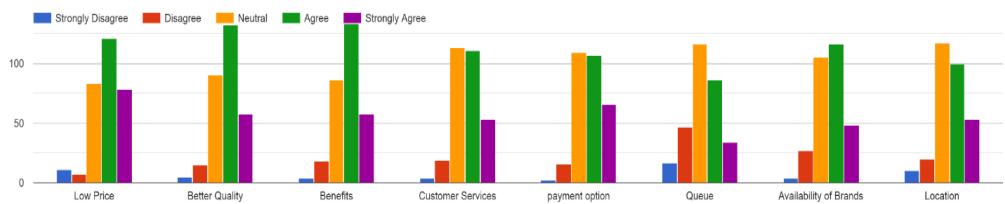
Which of the following according to you, will help build the good image for Other Stores ?

103 responses

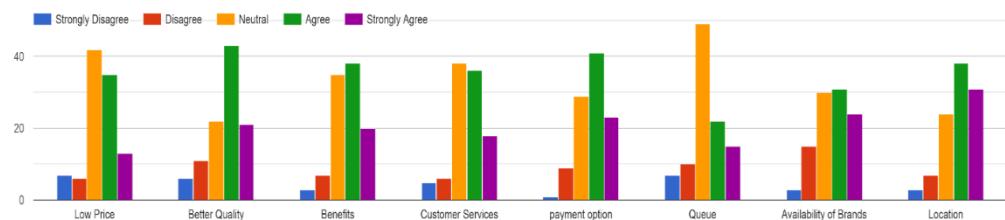


**To build a good image for D-mart competitive pricing will help the most with 60.7% & for other stores it can build its image the most is through good value added services with 60.2%**

Please give your opinion on D-Mart ?

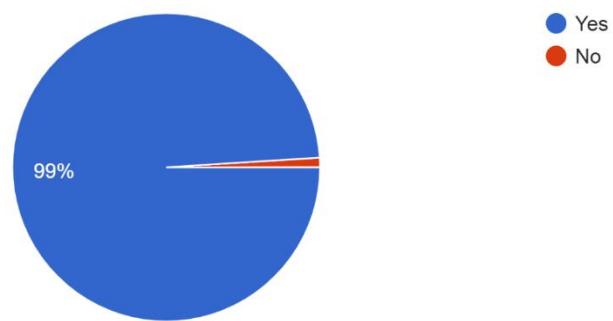


Please give your opinion on Other Stores ?



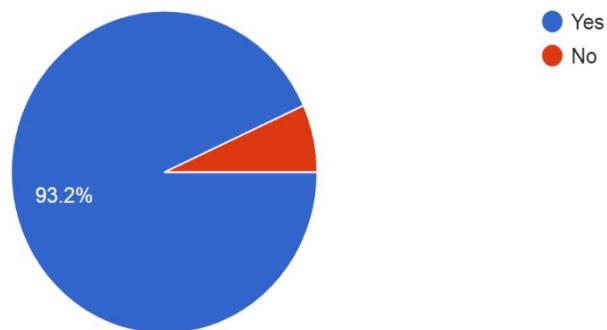
Are you satisfied with D-Mart Products ?

300 responses



Are you satisfied with Other Stores Products ?

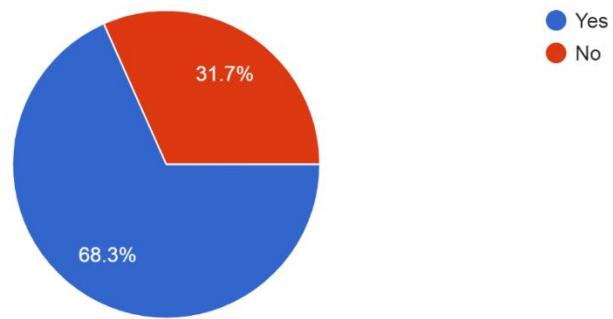
103 responses



From the graphs we can see that there customers are more satisfied from D-mart rather than other stores

Do you prefer D-Mart ready over offline purchase ?

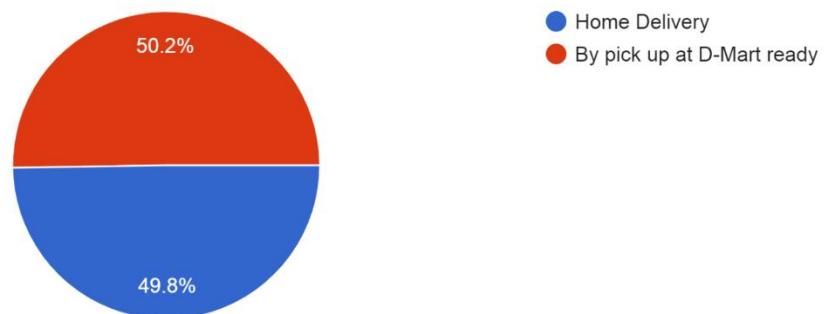
300 responses



**68.3% of participant are ready for offline purchase in D-mart.**

How would you like to get your order ?

203 responses

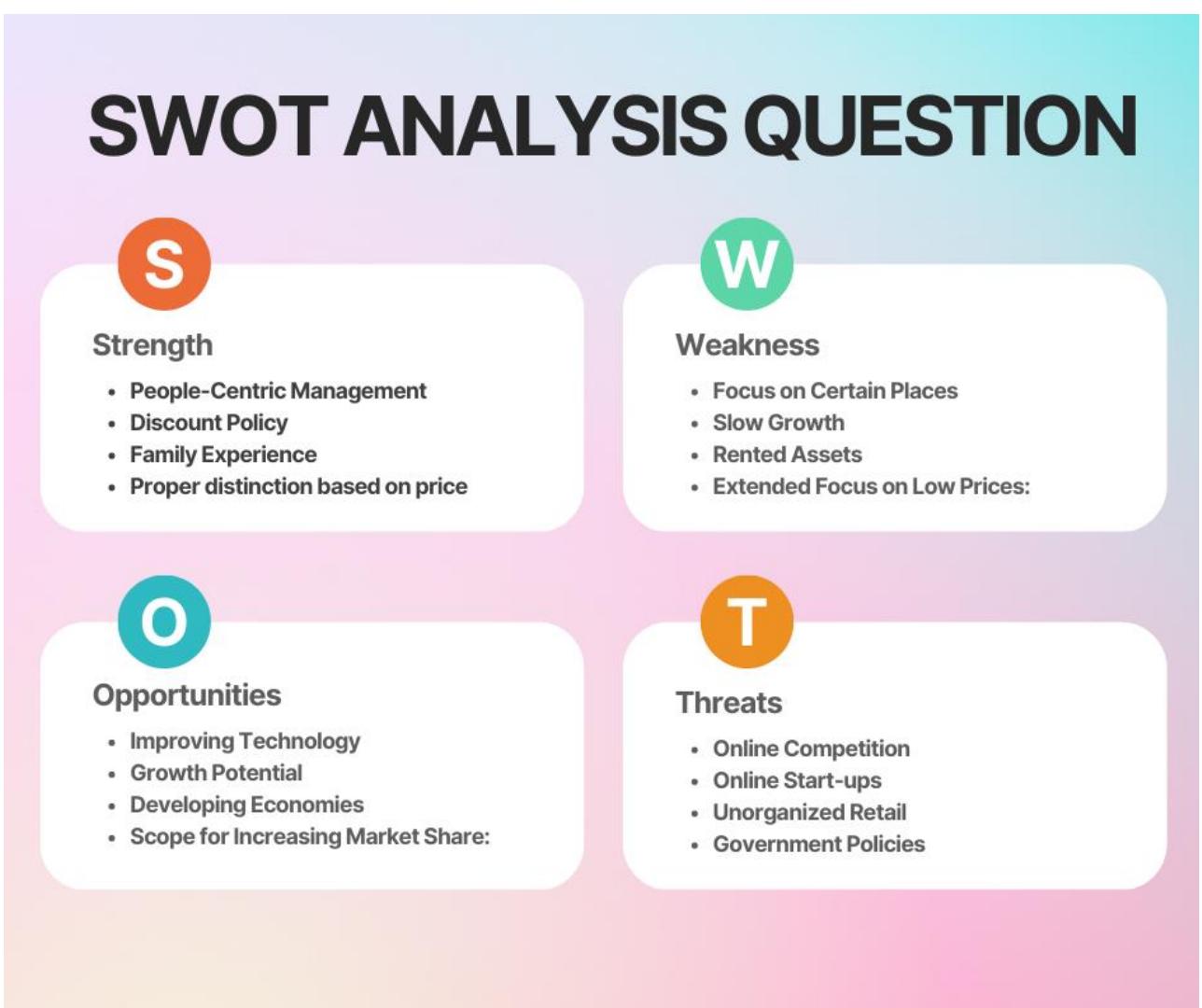


## SWOT ANALYSIS

### 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

- **People-Centric Management:** It maintains good relations with all of its stakeholders including its vendors and suppliers. It has a strong employee policy and is transparent in employee relations.
- **Discount Policy:** It is known for its low price and various offers and discounts. This gives D-Mart an edge over the competition. Through its pricing strategies, it is sufficiently generating value.
- **Family Experience:** D-Mart stores offer a full family shopping experience. It is a one-stop destination for all of the family's needs which makes the customer shopping experience way much easier and people are liking this.
- **Proper distinction based on price:** D-Mart chose not to follow the trends set by other retail competitors and instead, created their own. They achieved dominance by offering their products at significantly lower prices than their rivals, using a simple price-based differentiation strategy.

## WEAKNESSES

- **Focus on Certain Places:** The majority of D-Mart stores are in the Western States and fewer in the southern market which makes D-Mart leave a lot of untapped markets.
- **Rented Assets:** Retail stores like D-Mart operate mostly through rented stores in malls, resulting in higher costs and the majority of sales going to pay rents.
- **Slow Growth:** D-Mart was established 20 years ago, but it still hasn't been able to capture the market as much as it should have. It's mainly because of its long term focus on only one mission.
- **Extended Focus on Low Prices:** In its attempt to provide low prices to its customers they have to continuously haggle with vendors which may aggravate vendor relations.

## OPPORTUNITIES

- **Improving Technology:** Improved technology will enhance the in-store experience of customers and it, therefore, can charge a premium for that.
- **Growth Potential:** D-Mart stores aspire to be India's most valuable retail stores for its customers. They can take this as a motivation to grow more in the market such as global expansion and tie-ups with international brands.
- **Developing Economies:** The entire economy is open, and D-Mart has a rising opportunity to access burgeoning developing economies.
- **Scope for Increasing Market Share:** D-Mart has a great scope of expanding its market share by opening more stores in the southern states

## THREATS

- **Online Competition:** People today are more and more inclined to shop online rather than visit a store and do all the hard work. Online competitors like Amazon Pantry and Local Platforms are becoming tough competitors of supermarkets like D-Mart.
- **Online Start-ups:** There is a massive increase in the number of startups and they generally provide even lower prices and personalized experiences.
- **Unorganized Retail:** A large population of the target market still prefers to buy goods directly from local convenience stores and shops.
- **Government Policies:** There is a continuous change in government policies across different countries. Moreover, political unrest in the country can impede business, resulting in lower performance and higher costs.

## **CHI-SQUARE TEST OF INDEPENDENCE**

---

### **OBJECTIVE**

To test the association between occupation and their income for D-mart 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.

## **STEPS TO IMPLEMENT FOR CHI-SQUARE TEST OF INDEPENDENCE**

•SPSS

2 •Import the table

3 •Weight Cases-Frequency

4 •Analyze

HO2: There is no significant association 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 three sub-objectives for further analysis.

1. To test the association between occupation and their income for D-mart products.

2. To test the association between occupation and their the expense for D-mart products.

3. To test the association between income and their the expense for D-mart products.

## **HYPOTHESIS**

HO1: There is no significant association between the occupation and their income for D-mart products.

H11: Not HO1 between the occupation and their expense for D-mart products.

H12: Not HO2

HO3: There is no significant association between the income and their expense for D-mart products.

H13: Not HO3

## **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.

Occupation

Student 1

Business 2

Service 3

Homemaker 4

Annual income

Less than 1 lakh 1

1 - 5 Lakh 2

5 - 9 Lakh 3

More than 9 Lakh 4

Expenditure

Below 1000 1

1000 - 3000 2

3000 - 6000 3

more 6000 4

### Occupation \* Income Crosstabulation

Occupation			Income				Total
			1	2	3	4	
1	Count		37	47	19	12	115
	Expected Count		19.9	51.4	31.8	11.9	115.0
2	Count		1	14	14	8	37
	Expected Count		6.4	16.5	10.2	3.8	37.0
3	Count		9	58	38	9	114
	Expected Count		19.8	50.9	31.5	11.8	114.0
4	Count		5	15	12	2	34
	Expected Count		5.9	15.2	9.4	3.5	34.0
Total	Count		52	134	83	31	300
	Expected Count		52.0	134.0	83.0	31.0	300.0

### Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)		Exact Sig. (2-sided)
			sided)		
Pearson Chi-Square	41.377 <sup>a</sup>	9	.000		. <sup>b</sup>
Likelihood Ratio	42.117	9	.000		
Linear-by-Linear Association	6.618	1	.010		
N of Valid Cases	300				

p-value(0.000)<0.05

Thus, we reject H01. Hence, we conclude that There is significant association between the occupation and their income for D-mart products., i.e., they are dependent of each other.

**OBJECTIVE** To test the association between occupation and their expense for D-mart products.

### Occupation \* Expense Crosstabulation

Occupation	1		Expense				Total
			1	2	3	4	
Occupation	1	Count	18	45	44	8	115
		Expected Count	11.5	40.6	46.4	16.5	115.0
	2	Count	2	7	19	9	37
		Expected Count	3.7	13.1	14.9	5.3	37.0
	3	Count	7	43	44	20	114
		Expected Count	11.4	40.3	46.0	16.3	114.0
	4	Count	3	11	14	6	34
		Expected Count	3.4	12.0	13.7	4.9	34.0
Total		Count	30	106	121	43	300
		Expected Count	30.0	106.0	121.0	43.0	300.0

### Chi-Square Tests

	Value	df	Asymptotic	Exact Sig. (2-sided)
			Significance (2-sided)	
Pearson Chi-Square	19.111 <sup>a</sup>	9	.024	<sup>b</sup>
Likelihood Ratio	19.987	9	.018	
Linear-by-Linear Association	6.721	1	.010	
N of Valid Cases	300			

p-value(0.024)<0.05

Thus, we reject H02. Hence, we conclude that There is significant association between the occupation and their expense for D-mart products., i.e., they are dependent of each other.

### Income \* Expense Crosstabulation

Income		Count	Expense				Total
			1	2	3	4	
1	Count	13	25	12	2	52	
	Expected Count	5.2	18.4	21.0	7.5	52.0	
2	Count	9	53	58	14	134	
	Expected Count	13.4	47.3	54.0	19.2	134.0	
3	Count	6	21	39	17	83	
	Expected Count	8.3	29.3	33.5	11.9	83.0	
4	Count	2	7	12	10	31	
	Expected Count	3.1	11.0	12.5	4.4	31.0	
Total		30	106	121	43	300	
		30.0	106.0	121.0	43.0	300.0	

### Chi-Square Tests

		Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square		40.628 <sup>a</sup>	9	.000	. <sup>b</sup>
Likelihood Ratio		38.031	9	.000	
Linear-by-Linear Association		28.403	1	.000	
N of Valid Cases		300			

p-value(0.000)<0.05

Thus, we reject H03.

Hence, we conclude that There is significant association between the income and their expense for D-mart products., i.e., they are dependent of each other.

## K-NEAREST NEIGHBORS

---

### OBJECTIVE

To check whether the Customers are satisfied with the D-Mart for offline shopping or not and check the accuracy.

The K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems. It is a non-parametric algorithm, which means it does not make any assumption on underlying data. It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset. K-NN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much like the new data.

### PROS

1. K-NN is intuitive and simple.
2. K-NN has no assumptions.
3. It constantly evolves.
4. Very easy to implement for multi-class problems.
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 a 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.

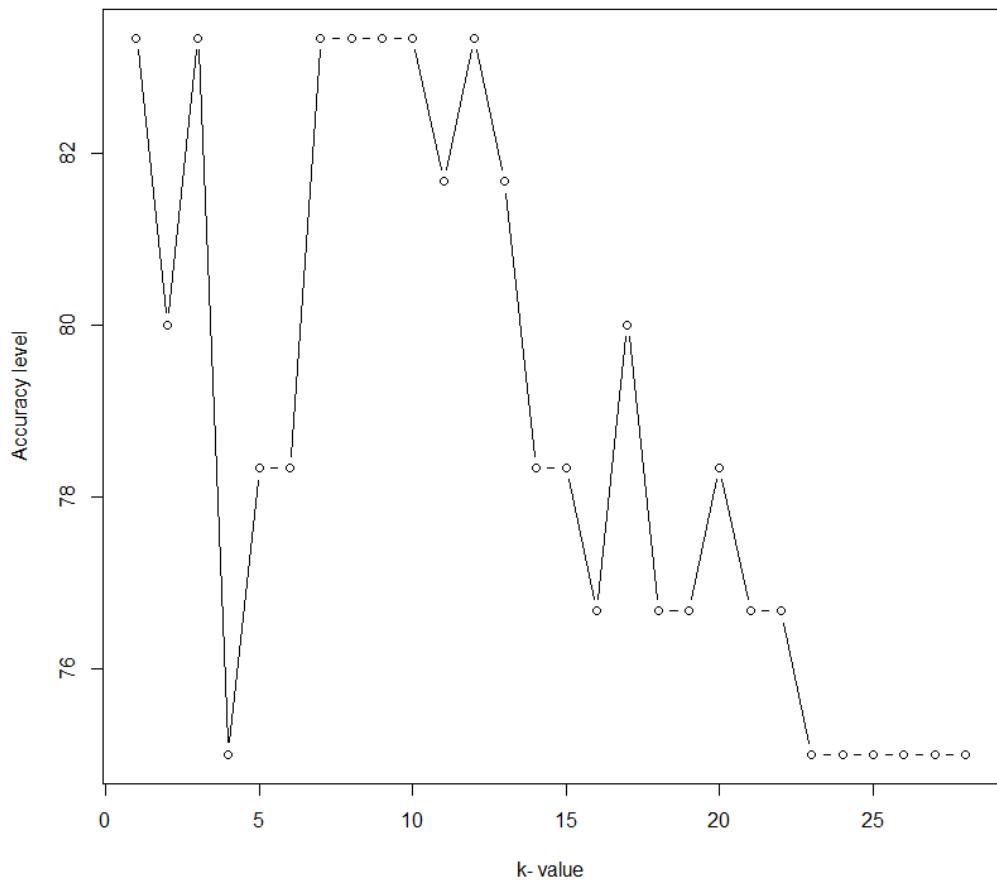
## STEPS TO IMPLEMENT K-NN ALGORITHM

- 1) Import the data.
- 2) Install the packages like “Class”, “caret”, “dplyr”
- 3) Summarize the data.
- 4) Create features and target variables.
- 5) Drop the columns which are not required and make a new data frame.
- 6) Convert the suitable column into factor.
- 7) Split the data into training and testing set.
- 8) Generate the K-NN model using neighbor value.
- 9) Train or fit the data into the model.
- 10) Predict the future.

### Data Analysis and Interpretation

knn.10	1	2
1	45	11
2	0	4

- 1) Out of 56 observations, for 45 observations we are correctly able to predict the offline purchase value as 1, which means that 45 customers are willing to shop offline at D-Mart. Similarly for 4 observations we have correctly predicted that the 4 are willing to use online D-Mart shopping.
- 2) But for 11 customers we have predicted that they prefer online shopping but due to some reasons are still using offline D-Mart stores
- 3) We have 300 observations in our data set, so 240 observations are for our train data set. The square root of 300 is around 17.03, but we have made a model in that we have given all values from 1 to 28.
- 4) So, the value for k is higher for 1,3,7,8,9,10. i.e., **83.33%** as shown in above graph.
- 5) Therefore, we have created a model with the value of k = 10
- 6) The model predicted that the frequency of the class with an accuracy of **83.33%**.



## Kappa Statistics

The Cohen's kappa is a statistical coefficient that represents the degree of accuracy and reliability in a statistical classification. It means the agreement between two raters (judges) who each classify items into manually exclusive categories.

Kappa statistic = 0.3529

The value of the kappa statistic indicates that there is a fair agreement between the raters.

	Class 1
Sensitivity	1.0000
Specificity	0.2667
Pos Pred value	0.8036
Neg Pred value	1.0000
Prevalence	0.7500
Detection Rate	0.7500
Detection Prevalence	0.9333
Balanced Accuracy	0.6333

# CLUSTERING

---

## Objective:

Grouping similar characteristics people

K-means clustering is a machine learning clustering technique used to simplify large datasets into smaller and simple datasets. Distinct patterns are evaluated, and similar data sets are grouped together. The variable K represents the number of groups in the data. This article evaluates the pros and cons of the K-means clustering algorithm to help you weigh the benefits of using this clustering technique.

## Pros:

- **Simple:** It is easy to implement k-means and identify unknown groups of data from complex data sets. The results are presented in an easy and simple manner.
- **Flexible:** K-means algorithm can easily adjust to the changes. If there are any problems, adjusting the cluster segment will allow changes to easily occur on the algorithm
- **Suitable in a large dataset:** K-means is suitable for a large number of datasets, and it's computed much faster than the smaller dataset. It can also produce higher clusters.
- **Efficient:** The algorithm used is good at segmenting the large data set. Its efficiency depends on the shape of the clusters. K-means works well in hyper-spherical clusters.
- **Time complexity:** K-means segmentation is linear in the number of data objects thus increasing execution time. It doesn't take more time to classify similar characteristics in data like hierarchical algorithms.
- **Tight clusters:** Compared to hierarchical algorithms, k-means produce tighter

clusters especially with globular clusters.

- **Easy to interpret:** The results are easy to interpret. It generates cluster descriptions in a form minimized to ease understanding of the data.
- **Computation cost:** Compared to using other clustering methods, a k-means clustering technique is fast and efficient in terms of its computational cost  $O(K^*n^*d)$ .
- **Accuracy:** K-means analysis improves clustering accuracy and ensures information about a particular problem domain is available. Modification of the k-means algorithm based on this information improves the accuracy of the clusters.
- **Spherical clusters:** This mode of clustering works great when dealing with spherical clusters. It operates with an assumption of joint distributions of features since each cluster is spherical. All the clusters features or characters have equal variance and each is independent of the other.

Cons:

- **No-optimal set of clusters:** K-means doesn't allow the development of an optimal set of clusters and for effective results, you should decide on the clusters before.
- **Lacks consistency:** K-means clustering gives varying results on different runs of an algorithm. A random choice of cluster patterns yields different clustering results resulting in inconsistency.
- **Uniform effect:** It produces clusters with uniform sizes even when the input data has different sizes.
- **Order of values:** The way in which data is ordered in building the algorithm affects the final results of the data set.
- **Sensitivity to scale:** Changing or rescaling the dataset either through normalization or standardization will completely change the final results.
- **Crash computer:** When dealing with a large dataset, conducting a dendrogram technique will crash the computer due to a lot of computational load and Ram limits.
- **Handle numerical data:** K-means algorithm can be performed in numerical data only.
- **Operates in assumption:** K-means clustering technique assumes that we deal with spherical clusters and each cluster has equal numbers for observations. The spherical assumptions have to be satisfied. The algorithm can't work with clusters of unusual

size.

- **Specify K-values:** For K-means clustering to be effective, you have to specify the number of clusters (K) at the beginning of the algorithm.
- **Prediction issues:** It is difficult to predict the k-values or the number of clusters. It is also difficult to compare the quality of the produced clusters.

## How does the K-Means Algorithm Work?

### STEPS FOR K-MEANS ALGORITHM

- 01** Select the number K to decide the number of clusters.
- 02** Select random K points or centroids.
- 03** Assign each data point to their closest centroid.
- 04** Calculate the variance and place a new centroid of each cluster.
- 05** Repeat the third step.
- 06** If any reassignment occurs, then go to step-4 else go to FINISH.
- 07** Model is ready.

#### ***Variables considered***

- V1: Age***  
***V2: Occupation***  
***V3: Annual Income***  
***V4: Time of visiting store***  
***V5: Monthly Expense***

## VARIABLES ENCODED

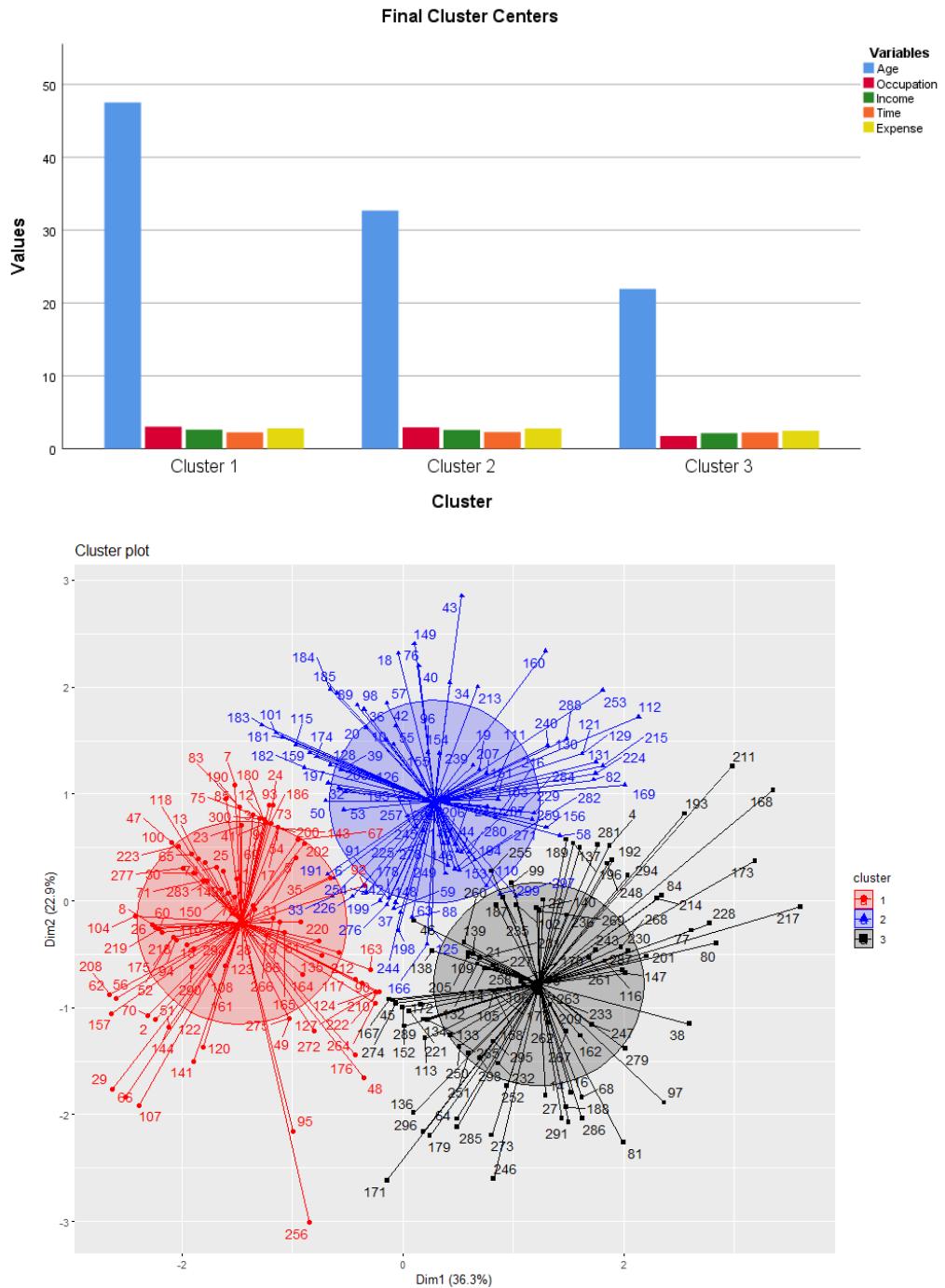
<b><i>Occupation</i></b>	
Student	1
Business	2
Service	3
Homemaker	4
<b><i>Annual Income</i></b>	
Less than 1 lakh	1
1 – 5 lakh	2
5 - 9 lakh	3
More than 9 lakh	4
<b><i>Monthly Expense</i></b>	
Less than 1000	1
1000 – 3000	2
3000 – 6000	3
More than 6000	4
<b><i>Time of Visiting Store</i></b>	
Morning	1
Afternoon	2
Evening	3

## Analysis

### Final Cluster Centers

	Cluster		
	1	2	3
Age	48	33	22
Occupation	3	3	2

Income	3	3	2
Time	2	2	2
Expense	3	3	2



### ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Age	11650.721	2	11.080	297	1051.494	.000
Occupation	52.973	2	.822	297	64.455	.000
Income	7.088	2	.727	297	9.746	.000
Time	.096	2	.529	297	.181	.834
Expense	3.699	2	.711	297	5.202	.006

### Interpretation K means Cluster Analysis

1. From Anova table, the age has greater influence in deciding the cluster, since the F value is 1051.494, next variable is occupation (64.455), income (9.746), expense (5.202) and the least important factor to influence cluster is time (0.181).
2. In cluster 1 on average the income of respondents is 5-9 lakh, age is 48 years, occupation is service and monthly expense is 3000-6000 Rs.
3. In cluster 2 on average the income of respondents is 5-9 lakh, age is 33 years, occupation is service and monthly expense is 3000-6000 Rs.
4. In cluster 3 on average the income of respondents is 1-5 lakh, age is 22 years, occupation business and monthly expense is 1000-3000 Rs.

## MARKET BASKET ANALYSIS USING APRIORI ALGORITHM

---

### OBJECTIVE 6

To identify the 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:

## APRIORI ALGORITHM

**Association Rule:** X 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.

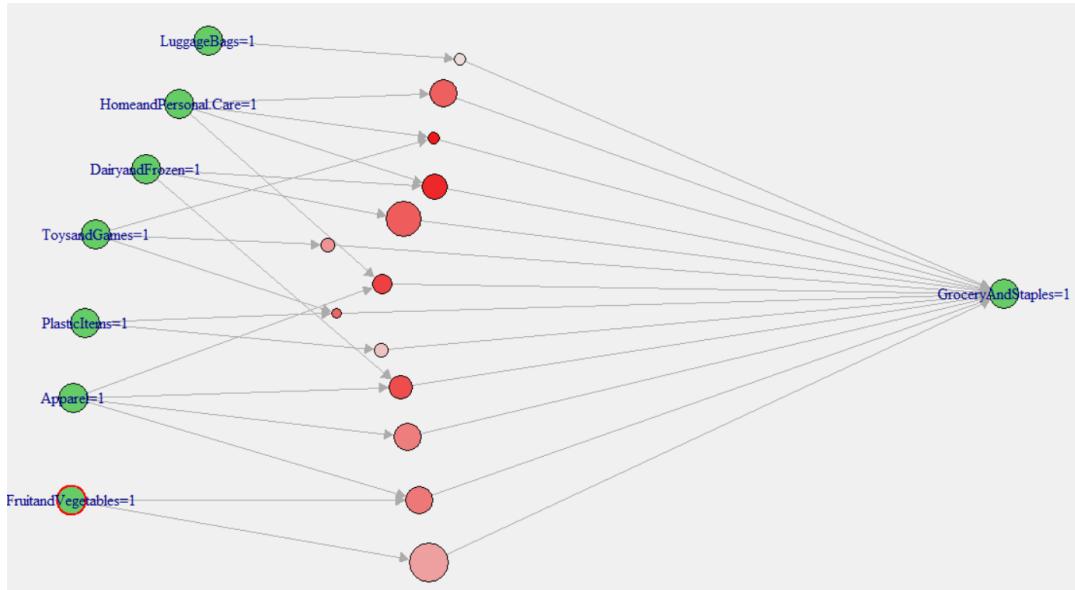
1. **Support:** This says how popular an item set is, it is number of times an item appears in total number of transaction in other word we say frequency of item.
2. **Confidence:** This says how likely item Y is purchased when item X is purchased, expressed as  $\{X \rightarrow Y\}$ . This is measured 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).

**Variables :**

- **V1 = Grocery and Staples**
- **V2= Fruits and Vegetables**
- **V3= Dairy and Frozen**
- **V4= Home and Personal Care**
- **V5= Luggage Bags**
- **V6= Toys and Containers**
- **V7= Apparel (Kid's, Men's & women's)**

## Data Analysis:

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{HomeandPersonal.Care=1, ToysandGames=1}	=> {GroceryAndStaples=1}	0.10666667	1.000000	0.1066667	1.2448133	32
[2]	{DairyandFrozen=1, HomeandPersonal.Care=1}	=> {GroceryAndStaples=1}	0.29333333	0.9887640	0.2966667	1.2308266	88
[3]	{HomeandPersonal.Care=1, Apparel=1}	=> {GroceryAndStaples=1}	0.23000000	0.9583333	0.2400000	1.1929461	69
[4]	{DairyandFrozen=1, Apparel=1}	=> {GroceryAndStaples=1}	0.28333333	0.9444444	0.3000000	1.1756570	85
[5]	{DairyandFrozen=1}	=> {GroceryAndStaples=1}	0.43333333	0.9285714	0.4666667	1.1558980	130
[6]	{HomeandPersonal.Care=1}	=> {GroceryAndStaples=1}	0.33333333	0.9259259	0.3600000	1.1526049	100
[7]	{PlasticItems=1, ToysandGames=1}	=> {GroceryAndStaples=1}	0.07333333	0.9166667	0.0800000	1.1410788	22
[8]	{FruitandVegetables=1, Apparel=1}	=> {GroceryAndStaples=1}	0.32333333	0.8981481	0.3600000	1.1180267	97
[9]	{Apparel=1}	=> {GroceryAndStaples=1}	0.32333333	0.8899083	0.3633333	1.1077696	97
[10]	{ToysandGames=1}	=> {GroceryAndStaples=1}	0.13000000	0.8666667	0.1500000	1.0788382	39
[11]	{FruitandVegetables=1}	=> {GroceryAndStaples=1}	0.50333333	0.8435754	0.5966667	1.0500939	151
[12]	{PlasticItems=1}	=> {GroceryAndStaples=1}	0.12666667	0.7755102	0.1633333	0.9653654	38
[13]	{LuggageBags=1}	=> {GroceryAndStaples=1}	0.09666667	0.7250000	0.1333333	0.9024896	29



Here, the association is highest between (Home & personal care, Dairy & Frozen to Grocery and Staples ) and (Dairy & Frozen ,Apparel to Grocery & Staples) that is people that are buying Home & personal care ,Dairy & Frozen And Dairy & Frozen ,Apparel products are highly to buy Grocery and Staples as well

## FACTOR ANALYSIS

---

### OBJECTIVE

To determine the factors that influences the people for buying D-Mart 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  $i^{\text{th}}$  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

### STEPS TO IMPLEMENT FACTOR ANALYSIS

- 01 Selecting and Measuring a set of variables in a given domain
- 02 Datascreening in order to prepare the correlation matrix.
- 03 Factor Extraction
- 04 Factor Rotation to increase interpretability.
- 05 Interpretation
- 06 Validation and Reliability of the measures

### **Variables used:**

- V1= Grocery and Staples
- V2= Fruits and Vegetables
- V3= Dairy and Frozen
- V4= Home and Personal Care
- V5= Luggage Bags
- V6= Toys and Containers
- V7= Apparel (Kid's, Men's & women's)

## **ASSUMPTIONS CHECKING**

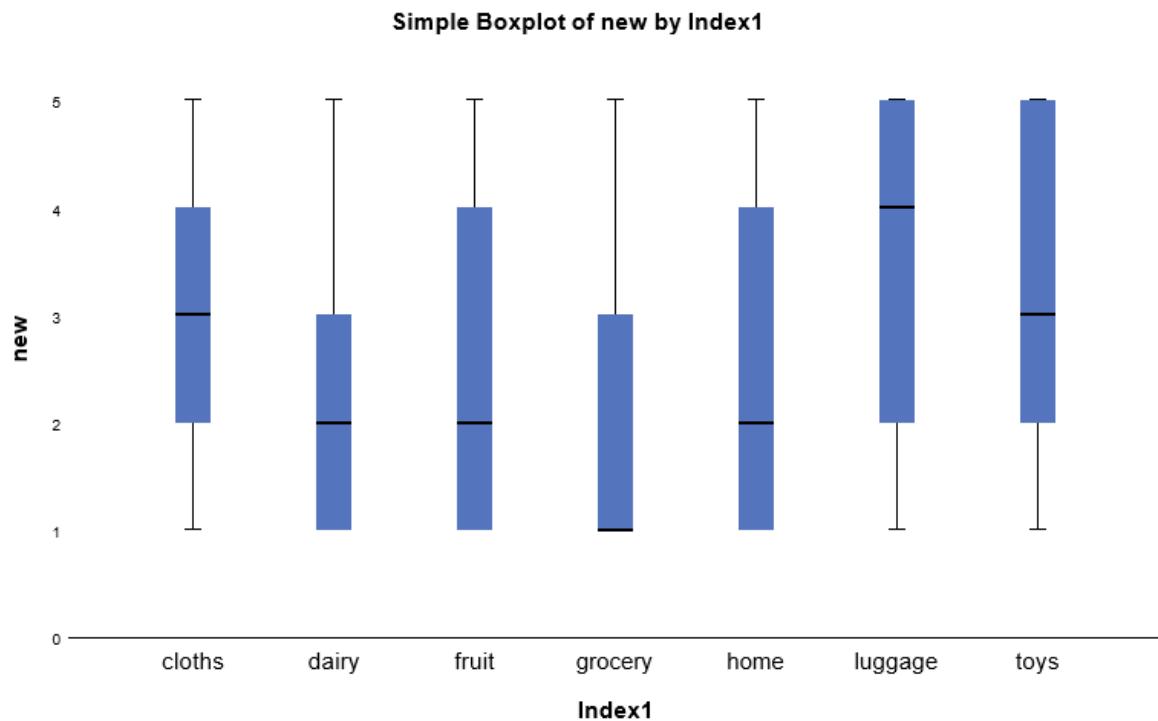
### **MULTICOLLINEARITY**

#### **Correlation Matrix<sup>a</sup>**

grocery		fruit	dairy	home	luggage	toys	cloths	
Correlation	grocery	1.000	.435	.416	.097	-.150	-.060	.016
	fruit	.435	1.000	.491	.066	.080	.094	.121
	dairy	.416	.491	1.000	.105	.061	.088	.056
	home	.097	.066	.105	1.000	.277	.318	.318
	luggage	-.150	.080	.061	.277	1.000	.533	.432
	toys	-.060	.094	.088	.318	.533	1.000	.416
	cloths	.016	.121	.056	.318	.432	.416	1.000

a. Determinant = .244

### **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.		.695
Bartlett's Test of Sphericity	Approx. Chi-Square	417.709
	df	21
	Sig.	.000

### Kaiser-Mayer-Olkin Statistics:

KMO Criterion: **0.695**

The value of KMO statistics is **0.695** 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

417.709

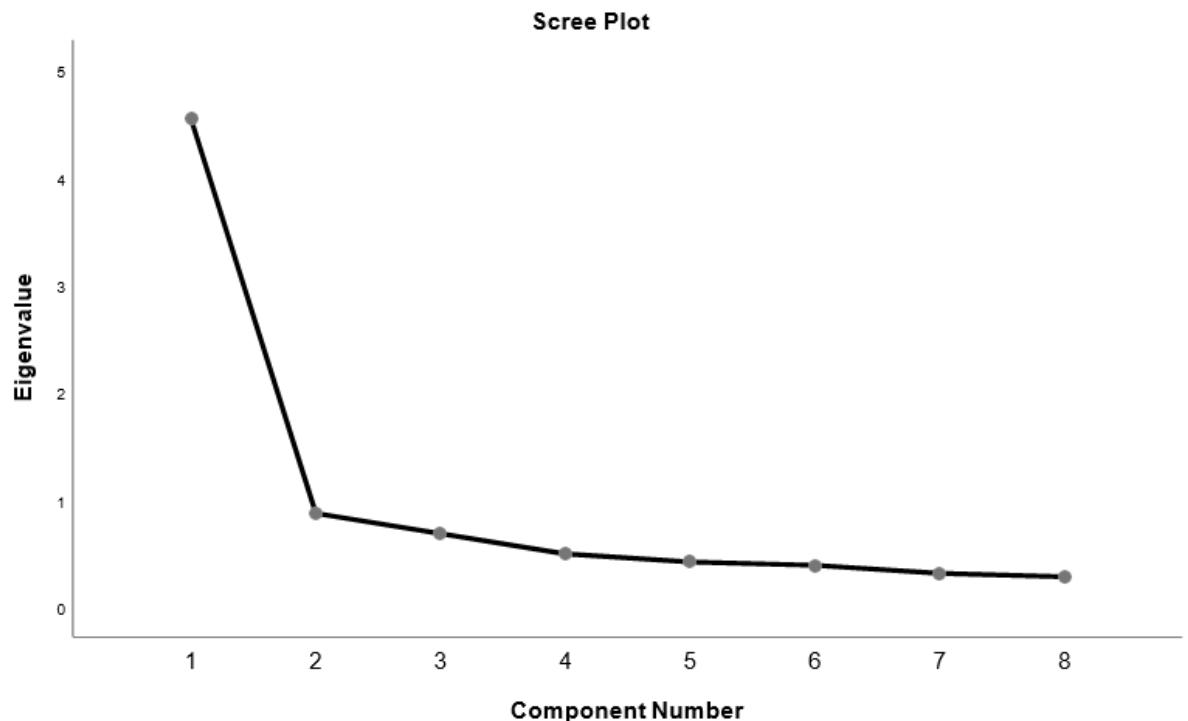
df=21

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 2<sup>nd</sup> component.

### Rotated Component Matrix :

	Component	
	1	2
luggage	.798	
toys	.795	
cloths	.734	
home	.586	
fruit		.796
dairy		.790
grocery		.789

## 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 eigenvalues more than 1.
- The total variance explained by these 2 factors is **58.476%**

Here, we get two factors which divides 7 variable



(ii) To determine the factors that are affecting to visit D-Mart Store

Variables used :

- V1= Low Price
- V2= Quality
- V3= Benefit
- V4= customer service
- V5= Payment option
- V6=Queue
- V7=Availability
- V8=Location

## ASSUMPTIONS CHECKING

### MULTICOLLINEARITY

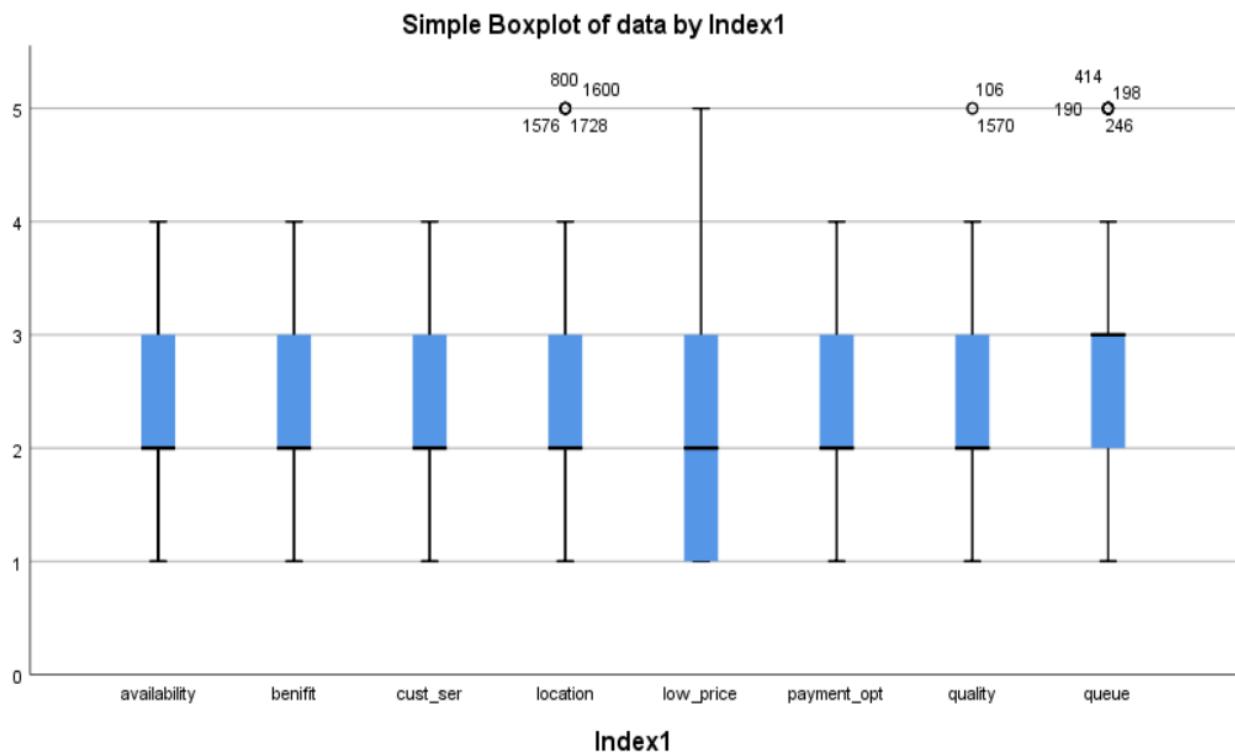
**Correlation Matrix<sup>a</sup>**

	low_price	quality	benifit	cust_ser	payment_opt	queue
Correlation	low_price	1.000	.626	.659	.555	.468
	quality	.626	1.000	.646	.687	.547
	benifit	.659	.646	1.000	.626	.571
	cust_ser	.555	.687	.626	1.000	.598
	payment_opt	.468	.547	.571	.598	1.000
	queue	.257	.409	.292	.446	.404
	availability	.533	.590	.600	.561	.502
	location	.393	.433	.396	.480	.369

**Correlation Matrix<sup>a</sup>**

		availability	location
Correlation	low_price	.533	.393
	quality	.590	.433
	benifit	.600	.396
	cust_ser	.561	.480
	payment_opt	.502	.369
	queue	.388	.383
	availability	1.000	.548
	location	.548	1.000

## Outliers:



There are 10 outliers present in the data but our rating scale is 1 to 5 and all 10 lies under that scale so we need those observations also

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.907
Bartlett's Test of Sphericity	Approx. Chi-Square	1165.170
	df	28
	Sig.	.000

## Kaiser-Meyer-Olkin Statistics:

KMO Criterion: **0.907**

The value of KMO statistics is **0.907** 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

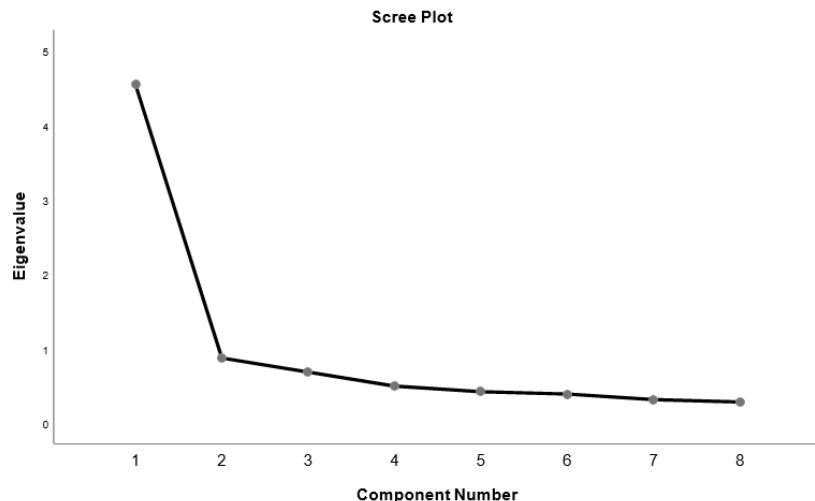
1165.170

df=28

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.



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 2<sup>nd</sup> component

## Rotated Component Matrix :

	1	2
benifit	.862	
low_price	.843	
quality	.769	.346
cust_ser	.696	.458
availability	.642	.462
payment_opt	.624	.405
queue		.882
location	.349	.658

## 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 eigenvalues more than 1.
- The total variance explained by these 2 factors is **67.734%**

Here, we get two factors which divide 8 variables



Consumer needs:

Queue

Location

## PARETO ANALYSIS

---

### OBJECTIVE

- (i) To study the major products that responsible for increasing sales in D-Mart

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.

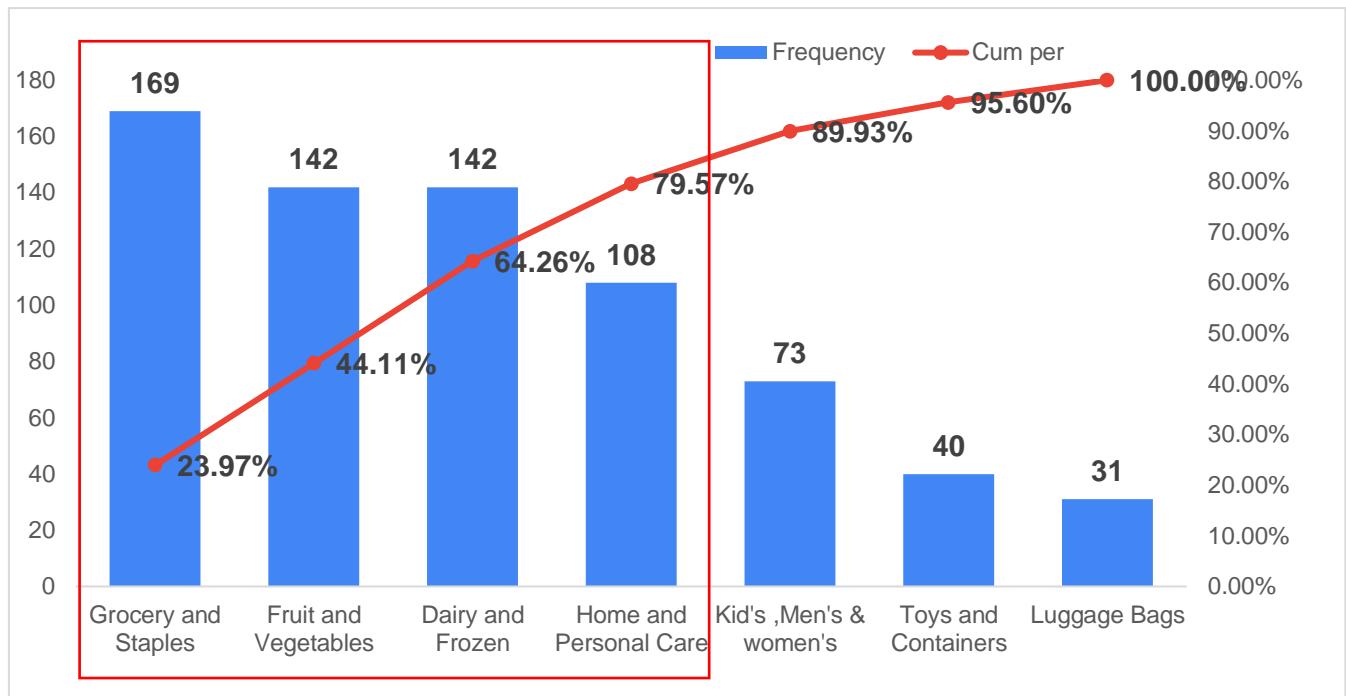
### PROS

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 decisionmaking, 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 damaged has already happened. It should focus on past data as well as present and future data also.

<b>Products</b>	<b>Frequency</b>	<b>Cum Freq</b>	<b>Cum per</b>
<b>Grocery and Staples</b>	169	169	23.97%
<b>Fruit and Vegetables</b>	142	311	44.11%
<b>Dairy and Frozen</b>	142	453	64.26%
<b>Home and Personal Care</b>	108	561	79.57%
<b>Kid's ,Men's &amp; women's</b>	73	634	89.93%
<b>Toys and Containers</b>	40	674	95.60%
<b>Luggage Bags</b>	31	705	100.00%
<b>total</b>	705		



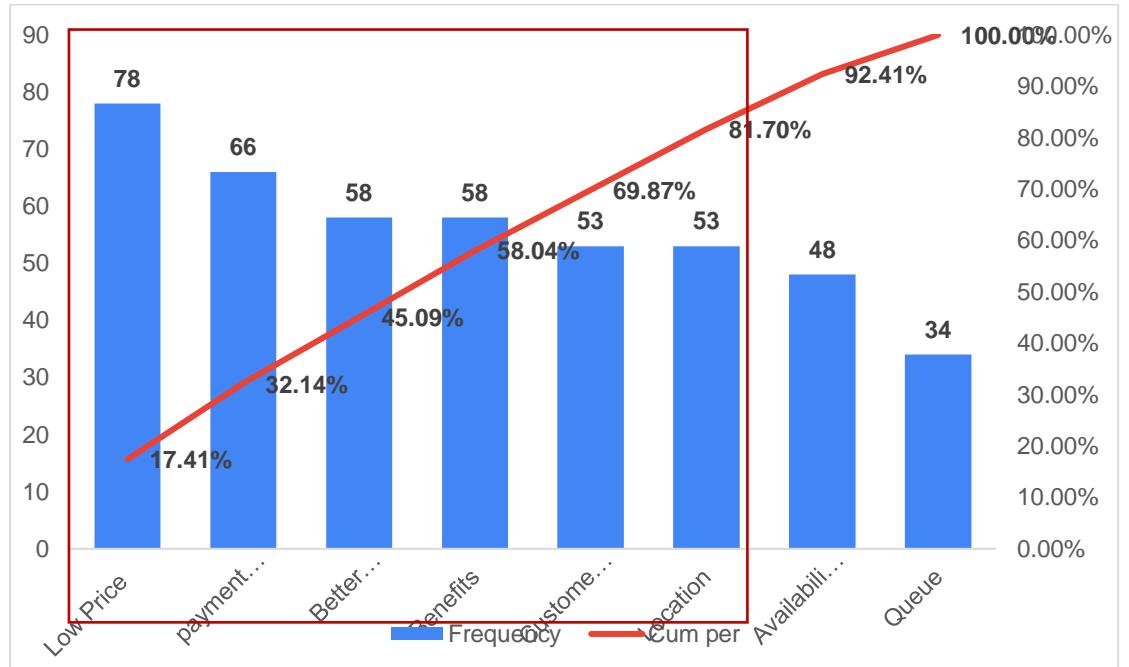
### Interpretation :

Grocery & Staples, Fruit & Vegetables , Dairy & Frozen and home & personal care are the major factors that influence people to shop at D-Mart

(ii)To study the major factors that responsible for people to visit D-Mart Store

Factors	Frequency	Cum Freq	Cum per
<b>Low Price</b>	78	78	17.41%
<b>payment option</b>	66	144	32.14%
<b>Better Quality</b>	58	202	45.09%
<b>Benefits</b>	58	260	58.04%
<b>Customer Services</b>	53	313	69.87%
<b>Location</b>	53	366	81.70%

<b>Availability of Brands</b>	48	414	92.41%
<b>Queue</b>	34	448	100.00%
<b>Total</b>	448		



### Interpretation:

Low Price , payment option ,Better quality ,Benefits, customer service and location are the major factors influence people to visit D-Mart Store.

## **SENTIMENT ANALYSIS**

---

### **OBJECTIVE**

To determine customers' behavior towards the products for D-Mart

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 analyse 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.

## **WORDCLOUD**

- **WHAT IS A WORD CLOUD?**

A word cloud is a popular visualization of words typically associated with text data.

They are most commonly used to highlight popular or trending terms based on the frequency of use to prominent and important. A Word Cloud is beautiful, Informative image that communicates much in a single glance.

- **WHAT IS THE PURPOSE OF THE WORLD CLOUD?**

Word clouds (also known as text clouds or tag clouds) work in a simple way: the more a specific word appears in a source of textual data (such as a speech, blog post, or database), the bigger and bolder it appears in the word cloud.

A word cloud is a collection, or cluster, of words depicted in different sizes. The bigger and bolder the word appears, the more often it's mentioned within a given text and the more important it is.

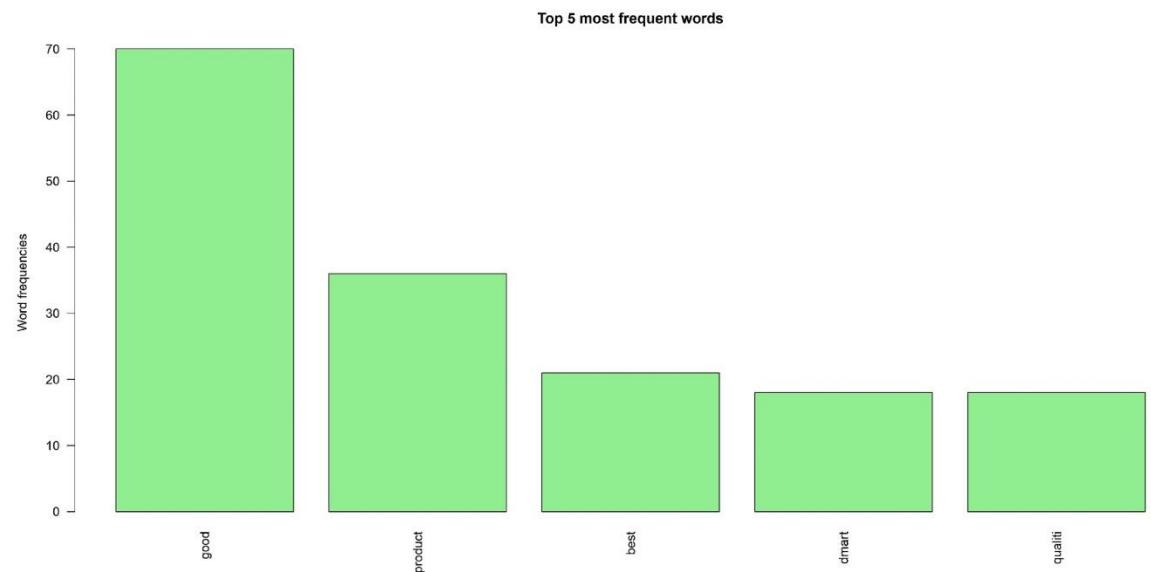
### **STEP TO BE IMPLEMENTED FOR SENTIMENT ANALYSIS**

- 1) Product Review
- 2) Sentiment Identification
- 3) Feature Selection
- 4) Sentimental Classification
- 5) Sentimental Polarity
- 6) Obtaining World Cloud

## Data Analysis & Interpretation

### Case 1: -

	Word	Frequency
Good	Good	70
Product	product	36
Best	best	21
D-Mart	D-Mart	18
Quality	Quality	18



Here,

- The most frequently occurring word is “good”. Also notice that no negative word is used in the data.
- “Good”, “Product”, “Best”, “D-Mart”, and “Quality” are the most frequently occurring words, which indicate that most people feel good about their work and their health.

## Case 2: -

### 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.

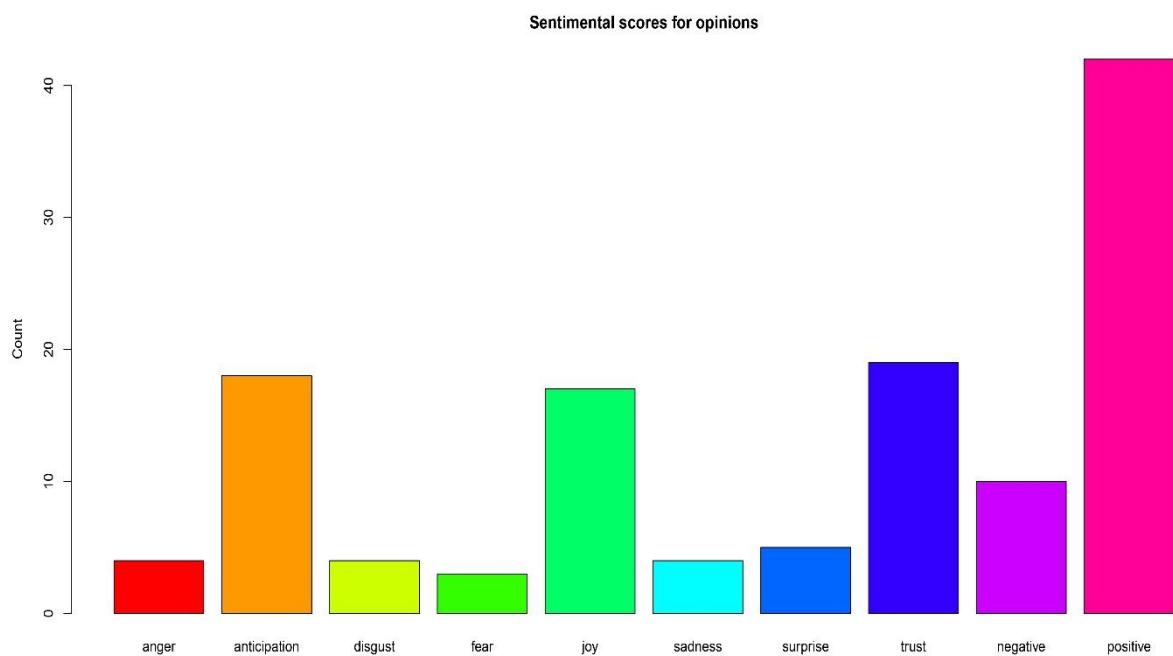
Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust	Negative	Positive
0	1	0	0	1	0	1	1	0	1
0	0	0	0	0	0	0	0	0	0
0	1	0	0	1	0	0	2	0	1
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
1	2	0	0	1	0	1	1	0	1
1	3	0	0	1	0	1	1	0	2
0	1	0	0	1	0	1	0	0	1
1	2	0	0	2	0	0	1	0	2

anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative
13	133	4	3	119	6	91	119	16
positive								
	168							

	count	emotion
anger	13	anger
anticipation	133	anticipation
disgust	4	disgust
fear	3	fear
joy	119	joy
sadness	6	sadness
surprise	91	surprise
trust	119	trust
negative	16	negative
positive	168	positive

We have a total of 300 observations, but the above tables only consist of the first 10 observations.

***Bar plot showing the count of words in the text, associated with each emotion.***



This bar chart demonstrates that words associated with the best emotion of “positive” occurred above forty times in the text, whereas words associated with the worst emotion of “disgust” occurred less than ten times. A deeper understanding of the overall emotions occurring in the survey response can be gained by comparing these numbers as a percentage of the total number of meaningful words.

### **Case 3: -**

For the final step word Cloud was obtained to focus and to shed light on what the people feel.

## Interpretation of Word Cloud

- It has been observed that words such as “good”, “Product” and “quality” have been expressed to a greater extent.
  - Some of the words like “best”, “D-Mart” and “nice” are the above point. Those customers are shopping from D-Mart are satisfied with the D-Mart with all positive feedback.



# CONCLUSIONS

1

## Factor analysis

- Consumer Daily Needs : Grocery and Staples, Fruits and Vegetables, Dairy and Frozen
- Consumer Functional Needs : Home and personal care, luggage, toy, containers and apparel
- Consumer preference : Benefit, Low Price, Customer Service, Availability, Payment Option
- Consumer Needs : Queue, Location

## PARETO ANALYSIS

- Grocery & Staples , Fruit & Vegetables , Dairy & Frozen and Home &are the major factors that influence
- Price , payment option ,Better quality ,Benefits, customer service and location are the major factors influence

2

## MARKET BASKET ANALYSIS

3

- association is highest between Home & personal care ,Dairy & Frozen
- Apparel products are highly to buy   Grocery and Staples as well
- To increase the sales of Fruit and Vegetables , Plastic Items and Luggage bags gives offer that are associated with Grocery and Staples

## KNN ALGORITHM

4

- This technique is conducted to check whether to predict that the D-Mart customers is satisfied with them or not.
- After checking the accuracy for KNN technique the highest accuracy we conclude that the D-Mart customers is satisfied with them or not.

5

## CHI SQUARE TEST OF INDEPENDENCE

- There is a significant association between occupation and income.
- There is a significant association between occupation and expense.
- There is a significant association between income and expense.

## CLUSTERING

- 3 clusters are formed. Age is the most important factor in deciding groups of people.
- People working in service sector most frequently visit D-Mart store.

6

## SENTIMENT ANALYSIS

7

- The analysis shows that D-mart customers have very high emotion of "positive" on their product & very low emotion of "disgust" for the product.
- Hence, this is good sign of customer's satisfaction.

## R Codes: -

```
data= read.csv(file.choose())
View(data)
str(data)
df=na.omit(data)

install.packages('dplyr')
df=data[,-!(names(data) %in% c("Timestamp", "Age", "Total.Mem.Fam",
"Annual.inc_fam", "Visit.store", "Time", "Visit.D_mart", "Dis_travelled",
"Source.D_Mart", "Main.Col.1", "Main.col.2", "Main.col.3", "Main.col.4",
"Result.of.offline.pur", "CP1", "CP2", "CP3", "CP4", "CP5", "CP6", "CP7",
"Com.D_mart"))]
View(df)
df1 = na.omit(df)
View(df1)
nrow(df1)
ncol(df1)

## install.packages("class")
library('class')
## install.packages('caret')
library('caret')

## Normalize
normalize = function(x) {
  return ((x - min(x)) / (max(x) - min(x))) }

D-Mart.subset.n = as.data.frame(lapply(df1[,2:38], normalize))
head(D-Mart.subset.n)

set.seed(123)      # to get the same random sample
```

```

df2 = sample(1:nrow(D-Mart_subset.n),size = nrow(D-Mart_subset.n)*0.8,replace =
FALSE) # Random select of 80% data

train.D-Mart = df1[df2,] ## 80% train data
test.D-Mart = df1[-df2,] ## remaining 20% test data

## Now creating separate data frame for Offline_pur feature is our target variable
train.D-Mart_lables = df1[df2,39]
test.D-Mart_lables = df1[-df2,39]

NROW(train.D-Mart_lables) ## to find the no. of observation

knn.17 = knn(train=train.D-Mart, test=test.D-Mart, cl=train.D-Mart_lables, k=17)
knn.18 = knn(train=train.D-Mart, test=test.D-Mart, cl=train.D-Mart_lables, k=18)

## Lets find the proportion of correct classification for k = 17 & 18
Acc.17 = 100 * sum(test.D-Mart_lables == knn.17)/NROW(test.D-Mart_lables) ## for
knn = 17
Acc.18 = 100 * sum(test.D-Mart_lables == knn.18)/NROW(test.D-Mart_lables) ## for
knn = 18

Acc.17 ## 78.33 % this is accuracy for k = 17
Acc.18 ## 76.66 % this is accuracy for k = 18

## to check the prediction against actual value in tabular form
table(knn.17 , test.D-Mart_lables)
knn.17

confusionMatrix(table(knn.17 , test.D-Mart_lables))

## One more step to added in this is that to improve the performance of the

```

model we can use the elbow method or maximum % accuracy  
graph

```
i=1          ## declaration to initiate for loop
k.optm=1
for (i in 1:28){
  knn.mod = knn(train = train.D-Mart, test = test.D-Mart, cl=train.D-Mart_lables, k=i)
  k.optm[i] = 100 * sum(test.D-Mart_lables == knn.mod)/NROW(test.D-Mart_lables)
  k = i
  cat(k, '=',k.optm[i],'\n')      ## to print % accuracy
}

# to plot % accuracy wrt to k-value
plot(k.optm, type = "b", xlab="k- value", ylab = "Accuracy level")

# By using the above model, we use knn = 10
knn.10 = knn(train=train.D-Mart, test=test.D-Mart, cl=train.D-Mart_lables, k=10)
Acc.10 = 100 * sum(test.D-Mart_lables == knn.10)/NROW(test.D-Mart_lables)
## for knn = 10
Acc.10    # 83.33% this is accuracy for k = 10

# To check the prediction against actual value in tabular form
table(knn.10 , test.D-Mart_lables)
knn.10

confusionMatrix(table(knn.10 , test.D-Mart_lables))
```

## 1) Clustering

# Code

```

install.packages("factoextra")
library(cluster)
library(factoextra)
df=read.csv(file.choose())
df1=df[,-1]
head(df1)
df1=na.omit(df1)
df1=scale(df1)
k=kmeans(df1,3)
k
fviz_cluster(k,data=df1,palette=c("red","blue","black"),
             ellipse.type="euclid",
             star.plot=T,
             repel=T,
             ggtheme=theme())

```

## 2) Market Basket Analysis

```

library(arules)
install.packages("arulesViz")
library(arulesViz)
x<-read.csv("C:/Users/nisha/Downloads/Resume/new.csv",header = T,colClasses =
             "factor")
summary(x)
rules1=apriori(x)
rules1
rules2=apriori(x,
                parameter = list(minlen=2,maxlen=3,confidence=.5,support=0.07),
                appearance =
list(lhs=c("FruitandVegetables=1","DairyandFrozen=1","Apparel=1","HomeandPerso
nal.Care=1","LuggageBags=1","PlasticItems=1","ToysandGames=1"),rhs=c("Grocer
yAndStaples=1")))

```

```

inspect(rules2)
r1=sort(rules2,decreasing = T,by="lift")
inspect(r1)
r2=is.redundant(r1)
r2
summary(r2)
r3=r1[!r2]
inspect(r3)
plot(r3,method="graph",
      engine="interactive")

```

### **3) Factor Analysis**

**Using spss**

### **4) Pareto Analysis**

**Using MS-Excel**

### **5) sentiment Analysis**

```

library('tm')
library('SnowballC')
library('RColorBrewer')
library('syuzhet')
library('ggplot2')
library('wordcloud2')

data<-read.csv("C:/Users/nisha/Downloads/sentiment.csv")
corpus=iconv(data)
corpus
corpus=Corpus(VectorSource(corpus))
inspect(corpus[1:30])
corpus=tm_map(corpus,tolower)
corpus=tm_map(corpus,removePunctuation)
corpus=tm_map(corpus,removeWords,stopwords("english"))

```

```
stopwords("english")
inspect(corpus[1:50])
corpus=tm_map(corpus,stripWhitespace)
final_data=corpus
tdm=TermDocumentMatrix(final_data)
tdm=as.matrix(tdm)
f=rowSums(tdm)
f=subset(f,f>=5)
barplot(f,las=2,col = rainbow(50))
d=data.frame(names(f),f)
colnames(d)=c('word','freq')
wordcloud2(d,size=0.8,shape="circle",rotateRatio=0.1,minSize=1)
review=as.character(data)
v=get_nrc_sentiment(review)
z=barplot(colSums(v),col=rainbow(10),ylab = "Count",main = "Sentimental scores for
opinions")
```

## QUESTIONNAIRE

---

1) Gender:

- MALE
- FEMALE

2) Age: \_\_\_\_\_

3) How many members are in your family ?

- 1-3
- 4-6
- More than 6

4) Occupation

- Student
- Buisness
- Service
- Homemaker

5) Annual income of the family

- Less than 1 lakh
- 1-5 lakhs
- 5-9 lakhs
- More than 9 lakhs

6) Which store comes in your mind when you think of a purchasing a product ?

- D-mart
- Other stores

7) How frequently do you visit D-Mart store ?

- Weekly
- Monthly
- Quarterly
- Occasionally

8) How long have you been visiting D-Mart ?

- 1-3 years
- 3-5 years
- More than 5 years

9) What time of you prefer purchasing D-Mart ?

- Morning
- Afternoon
- Evening

10) What is your average monthly expenditure while shopping in D-Mart ?

- Less than 1000
- 1000-3000
- 3000-6000
- More than 6000

11) How much distance do you travel to reach the store ?

- 0-2 km
- 2-4 km
- More than 4 km

12) What are your sources of awareness for D-Mart ?

- Newspaper
- Internet
- Friends and neighbours
- None of the above

13) Tick the column according to your buying behavior for the following categories of product ?

- Grocery and Staples
- Fruit and Vegetables
- Dairy and Frozen
- Home and Personal Care
- Luggage Bags
- Toys and Containers

- Plastic items
- Apparel(Kid's ,Men's & women's)

14) Tick the column according to your buying behavior for the following categories of product ?

	Occasionally	Often	Very Frequently	Always	Never
Grocery and Staple					
Fruit and Vegetables					
Dairy and Frozen					
Home and Personal Care					
Luggage Bags					
Toys and Containers					
Apparel(Kid's ,Men's & women's)					

15) Reasons for buying D-Mart products ?

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

16) According to you , in which category D-Mart needs to improve to be better brand ?

- Marketing strategy
- Online service
- Customer satisfaction
- Availability
- Quality

17) Which of the following according to you, will help build the good image for D-Mart ?

- Quality
- Competitive pricing
- Communication skills
- Good value added services
- Discount coupons

18) Please give your opinion on D-Mart ?

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Low Price					
Better Quality					
Benefits					
Customer Services					
payment option					
Queue					
Availability of Brands					
Location					

19) Are you satisfied with D-Mart Products?

- Yes

- No

20) What is your comment regarding D-Mart store

21) Do you prefer D-Mart ready over offline purchase ?

- Yes
- No

22) How would you like to get your order ?

- Home delivery
- By pick up at store

23) How frequently do you visit Other store ?

- Weekly
- Monthly
- Quarterly
- Occasionally

24) How long have you been visiting Other stores ?

- 1-3 years
- 3-5 years
- More than 5 years

25) What time of you prefer purchasing Other Stores ?

- Morning
- Afternoon
- Evening

26) What is your average monthly expenditure while shopping in Other Stores ?

- Less than 1000
- 1000-3000
- 3000-6000

- More than 6000

27) How much distance do you travel to reach the Other Stores ?

- 0-2 km
- 2-4 km
- More than 4 km

28) Tick the column according to your buying behavior for the following categories of product ?

- Grocery and Staples
- Fruit and Vegetables
- Dairy and Frozen
- Home and Personal Care
- Luggage Bags
- Toys and Containers
- Plastic items
- Apparel(Kid's ,Men's & women's)

29) Tick the column according to your buying behavior for the following categories of product ?

	Occasionally	Often	Very Frequently	Always	Never
Grocery and Staple					
Fruit and Vegetables					
Dairy and Frozen					
Home and Personal Care					
Luggage Bags					
Toys and Containers					

Apparel(Kid's ,Men's & women's)					
---------------------------------	--	--	--	--	--

30) Reasons for buying Other Stores products ?

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

31) According to you , in which category Other Stores needs to improve to be better brand ?

- Marketing strategy
- Online service
- Customer satisfaction
- Availability
- Quality

32) Which of the following according to you, will help build the good image for Other Stores ?

- Quality
- Competitive pricing
- Communication skills
- Good value added services
- Discount coupons

33) Please give your opinion on Other Stores ?

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Low Price					

Better Quality					
Benefits					
Customer Services					
payment option					
Queue					
Availability of Brands					
Location					

34) Are you satisfied with Other Stores Products ?

- Yes
- No

35) What is your comment regarding Other store ?

---

## Bibliography

---

- ⇒ Arbaz, A. (2019). "A STUDY ON CUSTOMER BUYING BEHAVIOUR TOWARDS D-. K.H.M.W.College, MUMBAI UNIVERSITY. Mumbai: HSC Projects.
- ⇒ Bhat, A. K. (2020, December 3-4). A Study on Customer Satisfaction towards D Mart Retail Outlet in. *International E Conference on Adapting to the New Business Normal –The way ahead*, 21.
- ⇒ Chechani, S. (2019-21). A Study on "Impact of D-Mart on small grocery retailers-consumer. (Shree, Ed.) *Journal of University of Shanghai for Science and Technology*, 13.
- ⇒ Dabre, M. H. (2019). A Study on the Consumer Buying Behaviour of D-Mart Shopping Center. *International Journal of Social Sciences & Humanities*, 5.
- ⇒ Jainy m. Patel, S. A. (2017-18). *Consumer Buying Behaviour of D-Mart*. A. S. B. Patel College, BUSINESS ADMINISTRATION. Patan: slideshare.
- ⇒ Narkhede, R. V. (2016-17). *A Consumer Buying Behavior with Reference to D-Mart in Jalgaon*. Moolji Jaitha College, BBA. Jalgaon: slideshare.
- ⇒ Patel, C. V. (.). *A Study on the effective promotional strategy influencing customer for BigBazaar and D-Mart*. KSV University, MBA. Gandhinagar: Academia Edu.
- ⇒ Swapna, D. (2014, December). A STUDY ON CONSUMER BEHAVIOR TOWARD D-MART's. *JETIR*, 8(12), 27.

## REFERENCES

---

- <http://www.analyticsvidhya.com/>  
<https://stackoverflow.com/>
- <https://youtu.be/pwrMW-fIwH4>
- <https://www.slideshare.net/JeniPatel4/consumer-buying-behaviour-of-D-Mart>
- <https://www.scribd.com/document/404999775/D-mart-black-book-docx>
- [https://www.google.com/url?sa=t&source=web&rct=j&url=https://tradebrains.in/how-big-is-D-Mart/&ved=2ahUKEwir1pXBwsn-AhUSSWwGHcscAt4QFnoECDgQAQ&usg=AOvVaw1HsR-du\\_GWPCqJFBK0M5ZT](https://www.google.com/url?sa=t&source=web&rct=j&url=https://tradebrains.in/how-big-is-D-Mart/&ved=2ahUKEwir1pXBwsn-AhUSSWwGHcscAt4QFnoECDgQAQ&usg=AOvVaw1HsR-du_GWPCqJFBK0M5ZT)
- <https://www.google.com/search?q=youtube+video+for+D-Mart+analysis&oq=youtube+video+for+D-Mart+analysis+&aqs=chrome..69i57j33i160l4.13615j0j7&client=ms-android-oneplus-rvo3&sourceid=chrome-mobile&ie=UTF-8#>



**Progress report for the Month of Jan2023-April 2023**

Date of report:	Guide's signature	Internal mentor signature	Roll Number:
			31031821012 31031821019 31031821022 31031821024 31031821040 31031821043
Name of student	Ronak Kalantre Harshal Kothavade Hasmina Laskar Soham Mali Nishant Thanekar Prathamesh Zinge		
Leave/ early off taken by the student during the reporting period	Number of days	With prior permission	Reason
	0		
Current status of project: Jan 2023	Deciding Topic Finding Research papers Decinding Objectives		
Feb 2023	Forming Questionnaire Performing Pilot Survey Performing pilot survey analysis Doing appropriate changes		
Mar 2023	Finalize the google form Collecting the data		
April 2023	Analysing the data Obtaining appropriate results Making Report and PPT		
Planned work executed in time		Yes	
Remarks by guide			