HSNC UNIVERSITY, MUMBAI
KISHINCHAND CHELLARAM COLLEGE

DEPARTMENT OF STATISTICS
CHURCHGATE – 400020

Comprehensive Study on Consumer Behaviour in Online Retail Using Different Statistical Tools

Kartavi Naik

Lajari Murkar

Riya Gidwani

Nikita Pakhale

Guide: Dr. S.B. Muley

HSNC UNIVERSITY, MUMBAI

KISHINCHAND CHELLARAM COLLEGE

DEPARTMENT OF STATISTICS
CHURCHGATE – 400020

CERTIFICATE

This is to certify that the following students of M.Sc. Part-II, have successfully completed the project entitled

"Comprehensive Study on Consumer Behaviour in Online Retail Using Different Statistical Tools"

During the Academic year of 2022-2023 Students involved in this research group are

KARTAVI NAIK

LAJARI MURKAR

RIYA GIDWANI

NIKITA PAKHALE

This research project is to the best of our knowledge and belief is original.

Dr. S. B. Muley	Dr. Asha Jindal
Assistant Professor ,	Professor
Project Guide	(UG:Head , PG: Coordinator)
Department of statistics	Department of Statistics

Acknowledgement

A profound sense of gratitude binds us to our reputable institution for a phenomenal environment and steady assistance towards the globe of research. Providing us with all the possible facilities to enhance our learning and providing us necessary amenities and required support in completion of this assignment.

We would like to express our special thanks to our principal, Dr. ________ for a great learning environment making the assimilation smooth and providing a wonderful opportunity to make this project report.

We would also like to thank to Dr. Asha Jindal, Head of the department of statistics, for being the key-person in inculcating our academic enhancement and moulding our learning attitude towards research study.

We would like to express our very great appreciation to our Mentor, Dr.Sakharam B Muley , Assistant Professor, for his valuable and constructive suggestions during the planning and development of this research work and willingness to give his time so generously has been very much grateful for us.

Our sincere thanks to the department of statistics, for their patient guidance and enthusiastic encouragement.

In addition, we would like to thank our parents, family and friends for great support in deliberating over our problems and findings, as well as providing happy distraction to rest our mind outside of the research.

Contents

ABSTRACT

Online shopping is the activity or action of buying products or services over the Internet. It means going online, landing on a seller’s website, selecting something, and arranging for its delivery. This study is to understand consumer behaviour for online shopping and preferred online platforms on the basis of factors influencing and satisfaction level.

INTRODUCTION:

Due to the busy lives and stressful schedules of the majority of people in the 21st century, online shopping has taken on a significant role. Online shopping became the most convenient and practical option for them in this circumstance. The internet has altered how consumers shop and quickly expanded to a worldwide perspective. The act of purchasing goods and services from an online store is physically comparable, and this type of online shopping is known as business to consumer online shopping. The classical model behaviour assumption underlies the current paper.

Online shopping is a type of e-commerce that enables customers to make direct purchases of products or services from a vendor through the Internet. E-store, e-shop, Internet shop, web store, virtual store, and online store are a few other names. The physical similarities between purchasing goods and services from an online store is what gives rise to the term "business-toconsumer" online shopping. Customers who choose to shop online choose to purchase a product from a website.

The internet has evolved into a brand-new method of product distribution. Together with looking up items and learning more about them, online shopping has emerged as one of the main reasons people use the internet. As a result, the internet has created a fiercely competitive market where customers compete.

Growing opportunities for online purchasing are provided by India's Internet usage. Consumer behaviour is treated as an applied discipline because some decisions have a significant impact on customers' behaviour and expected actions. If e-customers are aware of the factors that influence online behaviour and the relationships between these factors, they can build new marketing strategies to convert potential customers into active customers. The social and micro perspectives are the two primary ones that aim to solicit this awareness. The internet has altered how consumers shop and quickly evolved into a global viewpoint.

Several businesses began utilising online shopping with the intention of lowering marketing expenses, which would push them to reduce the price of their products in order to compete in extremely cutthroat marketplaces. Businesses utilise the Internet to connect, deliver, and distribute information and goods. Consumers utilize this data in a variety of ways, including to evaluate product features, costs, warranties, and delivery services in addition to making purchases. Several experts are optimistic about the future of the internet marketing industry. In addition to the E-enormous market's potential, the Internet offers businesses a singular opportunity to more effectively engage both current and new consumers.

History of online shopping

- Electronic shopping was invented by an English inventor called Michael Aldrich. In 1979, he invented the earliest form of e-commerce which allowed online transaction processing between business and customers, as well as between business and business.
- The invention of the first ever web browser, i.e. the World Wide Web in 1990 is the second major proponent of online shopping. Without an interconnected internet, there would be no online marketplace at all. We owe this marvellous invention called the ‘WWW’ to Tim

Berners Lee. That man is the major reason why millions of people have access to the internet which hosts numerous e-commerce platforms.

- 1995 welcomed the biggest transformation of online shopping. The first online marketplaces were established. First came Amazon.com, arguably the biggest online marketplace launched by Jeff Bezos—yup, the richest man in the world as of 2019 with an estimated net worth of $115 billion.

- All these online marketplaces with no reliable payment gateway. It is 1998, PayPal gets a full swing at the online payment system niche; quickly becomes a success.

- The history of online shopping 1979 – It all began when Michael Aldrich 'invented' online shopping. Using videotext, a two way message service, it revolutionized businesses. Online shopping started early in 1995 by the introduction of internet in India. Online shopping became popular during the Internet boom in 1999-2000 with the well know auction site know as bazee.com. As of 2020, customers can shop online using a range of different computers and devices, including desktop computers, laptops, tablet computers, smartphones and smart speakers. India plaza was the first ever online shopping store; founder was K. Vaitheeswaran, founded in June 1999. Earlier it was named as Fabmall and was US based company, but after a period of time it was renamed and launched as Indiaplaza in India. Amazon launches first online shopping site in India. Online retailer Amazon has launched its first shopping website in India. The company took its first steps into the Indian market in February 2012 when it launched Junglee.com, a site which allowed customers to compare prices online but not purchase items directly.

- As of 2020, it's clear to see that the entire internet is now a virtual shopping mall. With people choosing to carry out trades on Instagram, Whatsapp and Facebook, it's safe to say that we are in the online shopping era.

- All well and good but exactly how far can online shopping go? Will it replace the traditional way of shopping at malls? Yes it can. In fact, online shopping has all the precepts to become 'shopping' itself. While in-store retail still enjoys some patronisation, there has been a general shift to ecommerce and all with good reason. The future of ecommerce is now upon us. Eventually, online shopping will replace in-store retail for certain products. In time, shopping will mean 'online'.

If you own a physical store somewhere on earth or outer space, it's about time you considered a transition to online trade. We aren't very far from a global online shopping space. No longer will people have to queue up in stores to buy a product that may not be in stock. People can now make quality purchases from their comfort zones. Nothing beats comfort.

- India's social commerce has the potential to expand to US$16–20 billion in FY25, growing at a CAGR of 55-60%.
- India's e-commerce market is expected to reach US$ 111 billion by 2024 and US$ 200 billion by 2026.
- In 2022, the Indian ecommerce market is predicted to increase by 21.5%, reaching US$ 74.8 billion.
- India's e-commerce market is expected to reach US$ 350 billion by 2030.

The recent rise in digital literacy has led to an influx of investment in E-commerce firms, levelling the market for new players to set up their base, while churning out innovative patterns to disrupt old functioning.

Sample design:

The study is descriptive and analytical. It is descriptive in the sense it exist at present and it includes facts and findings. It is analytical in the sense it involves analysis of the collected data and information. The relevant data were collected through questionnaire.

SOURCE OF DATA ANALYSIS:

Primary Data:

Primary data is the original information or details collected from first-hand research. In other words, it's collected by the researcher himself. For this research we have collected the data through surveys for which the questionnaire was arranged and composed by ourselves.In order to carryout statistical enquires the questionnaire prepared included age, gender, educational qualification and information about the preference of the respondents.

Survey research has been chosen by the researcher as the strategy for examining customer perception and behaviour about internet purchasing. The research begins with the identification of many factors that contribute to how consumers perceive online buying. Primary data was used in the research. 324 people were the sample size for the survey that collected the primary data for this investigation. The questionnaire was utilised to gather first-hand information.

Research methodology:

Investigation has used a variety of techniques, including questionnaires and observations, to study the rise of online shopping among different genders, demographic groups, and age groups. Based on this research, a questionnaire survey that is chosen as the only significant public and professional sample has been designed. The purpose of the study was to look at how 324 participants felt about filling out an online shopping survey with closed-ended questions. Data was acquired from a variety of people in various occupations and from the general public.

Research includes:

1. Learning customer behaviour.

2. Getting a deeper understanding of frequency of online shopping.

3. Primary research.

Methods used:

Using a combination of RFM analysis and multiple statistics and machine learning, we have been able to better understand consumer loyalty.

We have utilised logistic regression, Naïve Bayes classification, SVM algorithm and Linear Discriminant analysis to examine consumer satisfaction levels across a range of variables, including product quality, discounts and offers, pricing, return and exchange policies, payment options, availability of a larger range of brands, varieties of items, and shipping times. The most crucial satisfaction criteria that differentiate between various client segments in terms of their loyalty have been identified through analysis of the model's coefficients.

We gathered information through survey satisfaction levels across a range of variables which forms independent variable in our study. The customer loyalty segment, which is divided into two and four classes using RFM, served as the dependent variable. Logistic regression ,Naïve Bayes classification and SVM algorithm used RFM segment with two classes whereas Linear Discriminant analysis used RFM segment with Four classes.

On applying logistic regression, Naïve Bayes classification, SVM algorithm and Linear Discriminant analysis we get the model's coefficients, and utilise stats models to get the pvalues for those coefficients. On the basis of the results we found out the significant factors that influence and discriminate the loyalty of customers and develop marketing strategies to drive sales.

Purpose of study :

The impact of factors influencing online shopping, satisfaction level on various aspects of online shopping, frequency of shopping, and demographic factors can have a significant impact on customer loyalty in the online shopping industry.

For instance, customer satisfaction can increase the likelihood of customers becoming loyal to a brand, while negative experiences can lead to customers leaving and switching to competitors. Similarly, the frequency of shopping can also have a strong influence on customer loyalty, with more frequent shoppers being more likely to become loyal customers.

Demographic factors such as age, gender, income, and education level can also impact customer loyalty. For example, younger customers may be more likely to switch between brands and be less loyal, while older customers may be more loyal to brands they have used for longer periods.

In summary, understanding the impact of these factors on customer loyalty can help businesses better understand and target their customer base, increase customer satisfaction, and improve overall retention rates.

Literature review:

1. Agarwal, Anuraag, and Sanskrity Joseph. (February,2021). Most companies run their online portals to sell their products / services online. The potential growth of online shopping has given rise to the idea of conducting online shopping research in India. Trust is one of the biggest barriers to success in Internet media. Lack of confidence and the risks involved can prevent online customers from participating in e-commerce. The

study intends to identify key variables and construct which has a significant influence one trust and e risk in India. The researcher through literature review has identified few dimensions of trust and risk which will be explored on the basis of sociodemographic variables to get broad picture and to arrive at conclusions. The data was collected through Questionnaires.

2. Barbosa, Diogo (march,2018). In a world where online shopping and digital interactions are growing constantly, one must be aware of their consequences for retail and how it can benefit from this. Big retail chains are keeping up with the phenomena by reinventing its way to attract more shoppers, either by creating very complete online portals or by bringing technology to its physical spaces.This paper brings some data on how consumers are using digital platforms for buying goods while inside a physical store. It concludes that a growing number of consumers are using mobile as an in-store support device as well as retailers are using it to attract consumers to stores.The paper also shares some data from a small inquiry made to Portuguese users about their use of mobile for in-store shopping purposes.
3. Dubey, Anil Roy, and B. Balaji (April, 2021). In the digital world, it’s somewhat difficult to manage the expectations of the present customers especially E-Commerce Shoppers who are often challenging and pushed to keep up with what is anticipated to them. Numerous E-shoppers are under pressure to locate and approach to convey an offline in-store ride to the online store ride of higher level interaction with their customers.The www is to remodel in the region of populace where communal networked group impact and leads to online shopping.

4. G. Bhongade, Bhagyashree, and Ashwini V.Z(2018). The growth of the internet as a secure online shopping channel has developed since 1994. With the Increasing number of e-commerce portal, we are now heavily inclined to online shopping. One of the benefits of online shopping is the ability to read reviews about the product purchased. This paper presents a semi-supervised approach for opinion mining using online product reviews obtained from Amazon website.
5. Prashar, Sanjeev, T. Sai Vijay, and Chandan Parsad(2015).The increased use of smartphones and tablets, along with advanced security features being offered by the online retailers are adding strength to e-commerce industry. Growing at an astonishing rate at 85%, as against 65% growth of regular shopping over the previous year, internet retailing in India touched US $10.672 billion in 2013, making it one of the most anticipated destinations for national and multinational online retailers. Several web portals are looking to capture a share of this huge market. This study gains importance as Indian arms of multinational online selling companies like Amazon and eBay are fighting various home-grown players like Snapdeal, Flipkart and many more. The objective of this paper is to identify and rank the factors that influence the selection of web portal among online shoppers in India.
6. Prashar, Sanjeev, Sai Vijay Tata, Chandan Parsad, Abhishek Banerjee, Nikhil Sahakari, and Subham Chatterjee(April,2019). This article describes how the exponential growth of e-commerce in India and the presence of many national and multinational e-retailers has set the trend for the major overhaul of the online industry. Most of the e-retailers have failed to differentiate themselves from the competitors. This has resulted in their failure to attract and retain the right set of consumers for their respective businesses. The article identifies four types of online shoppers – ‘Information Seekers,' ‘Utility

Seekers,' ‘Value Seekers' and ‘Core Shoppers.' Each of these four segments display significant differences and this information can be strategically used by web retailers in targeting their markets effectively.

7. Rao, Katta Rama Mohana, and Chandra Sekhar Patro. (July,2017). Over the last decade, the trend of web shopping has been increasing rapidly with the development and ease in accessibility of internet. Web shopping is a recent experience in the field of EBusiness and is definitely the future of shopping in the world. Most of the companies are operating online portals to sell their products or services to the customers. The growth of web shopping is an impetus for further exploration of the feelings towards the web, built between online businesses and shoppers. The study identifies the variables influencing shopper's stance towards web shopping and also offers some insights for improvement of the web stores. Furthermore, the results show that the variables have significant influence on the shopper's stance towards web shopping. Finally, some measures have been proposed for e-retailers to take initiatives for making web shopping experience more effective and trustworthy.
8. Reddy, Anuradha, and G. V. Bhavani Prasad.(April , 2012). With the introduction of Internet and e-commerce many companies have been performing their business transactions through e-portals. Increasing technology has bought tremendous changes in online business transactions (buying and selling). This paper examines consumer perceptions of varying characteristics of e-portals, identifies various factors that influence consumer trust and privacy e-portals, and analyzes how various security and privacy factors affect consumer perceptions toward e-portals. A survey questionnaire consisting 21 questions was developed and mailed to 150 e-commerce (B2B and B2C) consumers in 3 emirates of UAE wherein 108 individuals responded. Questions were developed from a literature review of news, as well as security and privacy issues. Factor analysis that included principal component analysis and varimax rotation was performed on all multiple scale items that determined retention of items. Results indicated that most participants are concerned about security and privacy issues while they are using e-portals, but few participants stated that security is the main issue that creates a barrier for their online shopping. Most participants are not aware of internet privacy and security policies and are not interested in knowing technology used for security of e-portals.
9. Srivastava, Gautam, Neeraj Anand, and Arvind Kumar Jain (2018). The market scenario in present time is not only limited to take a bag and go out for shopping but instead of that the pattern has now shifted towards going digital. Consumers is only away from its requirements by just a click and the goods will be delivered at his door steps. What can be more beneficial than saving time and money without wandering in the market in search of a right product in reasonable price. In todayâ€™s world digital marketing has totally changed the buying behaviour pattern of the consumers. Now-adays consumers don’t want to go physically to purchase the products.
10. Shaheen, Musarrat, Farrah Zeba, Namrata Chatterjee, and Raveesh Krishnankutty (September ,2019). Purpose Electronic commerce (e-commerce) is growing rapidly and the e-retailers are finding it pertinent to enhance customers’ online shopping experiences and engage them with e-commerce portals. The present study contributes to the theories of online marketing in the space of e-shopping, online reviews, customer trust, customer engagement and online shopping behavior. The study aims at understanding the role of different attributes associated with the online reviews’ credibility and information usefulness in driving customer engagement with specific

focus on online shopping through the utility of online devices. The study is one of the pioneering empirical studies that explore the role of online reviews in driving customer engagement.

DATA ANALYSIS

OBJECTIVE:

Use of different classification and prediction methods to analyse multivariate data based on:

- Factor influencing
- Preferred platform
- Satisfaction level
- Preferred products

RELIABILITY ANALYSIS:

In statistics, analysing the consistency and stability of a measurement instrument or research tool, such as a questionnaire, a test, or an observation procedure, is known as reliability testing. The goal of reliability testing is to identify the degree to which the measurement instrument produces reliable data across time and in a variety of contexts or scenarios.

In order to assess the degree of consistency, reliability testing often requires using the same measurement instrument on the same set of people on various occasions or under various circumstances. The following techniques are used to evaluate reliability:

1. *Test-retest reliability*: This entails giving the same measurement tool to the same set of people on two separate occasions, comparing the outcomes, and determining the instrument's consistency across time.

2. *Inter-rate reliability*: includes assigning the same action or event to numerous raters or observers, and then comparing the ratings to check for consistency.

3. *Inter consistency reliability*: assessing the consistency of the items within a single measurement instrument which is Cronbach's alpha coefficient, which evaluates how closely the items on a scale or questionnaire are correlated with one another, is a popular technique for evaluating internal consistency.

Reliability testing is a crucial phase in the research process since an unreliable measurement tool may produce inaccurate or inconsistent results, which may damage the validity of the research findings. To ensure that their findings are reliable and accurate, researchers therefore work to increase the reliability of their measurement tools.

A popular statistical test to evaluate the internal consistency reliability of a measurement instrument, such as a questionnaire or a scale, is called Cronbach's alpha. The degree to which the items in a single assessment instrument are consistent with one another and measure the same concept is referred to as internal consistency reliability.

The range of the Cronbach's alpha coefficient is 0 to 1, with higher numbers indicating more internal consistency reliability. For research purposes, a Cronbach's alpha value of 0.70 or above is typically considered as acceptable; however, the acceptable level may change based on the particular environment and study issue.

Assumptions:

1. *The measurement tool is stable over time*: Reliability assumes that the data collection tool provides consistent results over time. This implies that the instrument shouldn't deteriorate or change after each administration.
2. *The measurement tool is internally consistent*: Reliability presupposes that the components of a measurement tool are measuring the same construct or concept and are consistent with one another. This implies that there should be a strong correlation between the items.
3. *The sample is accurate*: Reliability implies that the population sampled for the study is representative of the larger population at large. This means that the sample should be randomly selected and should be large enough to ensure a representative distribution of the parameters.
4. *Reliability* in definitional terms take into account the normal distribution of the data obtained from the measurement tool. As a result, the distribution of scores should be symmetrical and bell-shaped.
5. *The lack of measurement error in the data*: Reliability assumes that the data obtained from the measurement device is free of bias or random error. This implies that the tool must be authentic and reliable

RELIABILITY TEST OF ONLINE RETAIL CONSUMERS

Execution on SPSS:

- *Using Cronbach's Alpha Test:*

1. On Factor Influencing

Case Processing Summary			
		N	%
Cases	Valid	278	100.0
	Excluded[a]	0	.0
	Total	278	100.0
a. List wise deletion based on all variables in the procedure.			

Table no.01

Item Statistics

	Mean	Std. Deviation	N
Ease of Use	3.47	1.418	278
Product Presentation	3.10	1.292	278
Shipping time and cost	3.45	1.276	278
Secure transactions	3.52	1.299	278
Pricing and Discount	3.66	1.355	278
Advertisement	2.88	1.269	278
Social Media	2.93	1.277	278
Availability of wide range of products	3.42	1.321	278
Value for money	3.39	1.330	278
Brand Consciousness	3.27	1.355	278
Product Marketing	3.15	1.330	278

Table no.02

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	3.294	2.885	3.662	.777	1.269	.062	11
Item Variances	1.745	1.611	2.012	.400	1.248	.014	11
Inter-Item Covariance	.910	.550	1.266	.716	2.300	.031	11
Inter-Item Correlations	.521	.338	.720	.382	2.129	.009	11

Table no.03

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Ease of Use	32.77	97.167	.718	.591	.915
Product Presentation	33.14	100.222	.671	.505	.917
Shipping time and cost	32.79	99.711	.703	.584	.915
Secure transactions	32.72	99.047	.716	.609	.915
Pricing and Discount	32.58	98.310	.712	.599	.915
Advertisement	33.35	102.085	.607	.548	.920
Social Media	33.31	102.567	.582	.499	.921
Availability of wide range of products	32.82	97.823	.753	.597	.913
Value for money	32.85	98.203	.731	.566	.914
Brand Consciousness	32.97	98.118	.719	.596	.915
Product Marketing	33.09	99.256	.688	.600	.916

Table no.04

Scale statistics			
Mean	Variance	Std. Deviation	N of Items
36.24	119.258	10.921	11

Table no.05

Reliability Statistics	
Cronbach's Alpha	N of Items
.923	11

Table no.06

Interpretation:

The value of Cronbach's alpha coefficient should be between 0 and 1, with a higher number indicating better reliability. Cronbach's alpha coefficient should be higher than 0.70; that scale has good internal validity and reliability.

Therefore, the results indicate that scale On what occasions we prefer online shopping has Cronbach's alpha value = 0.923 implying Excellent reliability and internal consistency.

2. On Product that you buy online

Case Processing Summary			
		N	%
Cases	Valid	278	100.0
	Excluded[a]	0	.0
	Total	278	100.0
a. Listwise deletion based on all variables in the procedure.			

Table no.07

Item Statistics			
	Mean	Std. Deviation	N
Fashion	3.06	1.329	278
Electronics	3.65	1.170	278
Footwear	3.46	1.245	278
Baby Products	4.43	1.034	278
Home decor and Furnishing	4.06	1.118	278
Food	3.39	1.349	278
Jewellery	4.06	1.196	278
Beauty Products	3.56	1.397	278
Grocery	3.74	1.348	278
Gift cards and Supplement	4.27	1.126	278
Toys and Video Games	4.31	1.164	278

Books	4.01	1.176	278
Handmade Products	4.36	1.104	278
Stationary Products	3.96	1.219	278
Medical Supplies	3.99	1.212	278

Table no.08

Summary Item Statistics

	Mean	Min.	Max.	Range	Maximum / Minimum	Variance	N of Items
Item Means	3.888	3.058	4.432	1.374	1.449	.159	15
Item Variances	1.480	1.069	1.951	.881	1.824	.065	15
Inter-Item Covariance	.452	.032	1.001	.970	31.425	.035	15
Inter-Item Correlations	.312	.020	.632	.611	30.974	.018	15

Table no.09

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Fashion	55.26	107.104	.302	.371	.871
Electronics	54.67	107.912	.324	.248	.869
Footwear	54.86	103.387	.483	.376	.862
Baby Products	53.89	104.395	.554	.466	.859
Home Decor and Furnishing	54.26	101.976	.618	.485	.856
Food	54.93	102.771	.460	.323	.863

Jewellery	54.26	100.633	.630	.531	.855
Beauty Products	54.76	99.824	.551	.489	.858
Grocery	54.58	102.042	.489	.343	.862
Gift cards and Supplement	54.05	102.059	.608	.497	.856
Toys and Video Games	54.01	102.888	.547	.507	.859
Books	54.31	104.561	.467	.302	.863
Handmade Products	53.96	101.912	.629	.558	.855
Stationary Products	54.36	101.371	.583	.505	.857
Medical Supplies	54.33	102.497	.538	.440	.859

Table no.10

Scale Statistics			
Mean	Variance	Std. Deviation	N of Items
58.32	117.164	10.824	15

Table no.11

⇨

Reliability Statistics	
Cronbach's Alpha	N of Items
.868	15

Table no.12

Interpretation:

The value of Cronbach's alpha coefficient should be between 0 and 1, with a higher number indicating better reliability. Cronbach's alpha coefficient should be higher than 0.70; that scale has good internal validity and reliability.

Therefore, the results indicate that scale On On Product that you online has Cronbach's alpha value = 0.868 implying Good reliability and internal consistency.

3. **On Preferred online shopping Platform**

Case Processing Summary			
		N	%
Cases	Valid	278	100.0
	Excluded[a]	0	.0
	Total	278	100.0

a. Listwise deletion based on all variables in the procedure.

Table no.13

Item Statistics			
	Mean	Std. Deviation	N
Myntra	2.86	1.457	278
Amazon	2.51	1.225	278
Flipcart	2.70	1.358	278
Jiomart	3.51	1.408	278
AJIO	3.81	1.365	278
BigBasket	4.10	1.246	278
Snapdeal	4.17	1.213	278
Netmeds	4.27	1.175	278

TATA1MG	4.28	1.144	278
Firstcry	4.26	1.216	278
Other	3.94	1.337	278

Table no.14

Summary Item Statistics							
	Mean	Min.	Max.	Range	Maximum / Minimum	Variance	N of Items
Item Means	3.673	2.507	4.284	1.777	1.709	.460	11
Item Variances	1.663	1.309	2.124	.815	1.623	.072	11
Inter-Item Covariance	.475	-.150	.960	1.110	-6.419	.082	11
Inter-Item Correlations	.298	-.082	.688	.771	-8.347	.039	11

Table no.15

Item-Total Statistics					
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Myntra	37.55	62.176	.273	.154	.822
Amazon	37.90	63.701	.275	.293	.818
Flipcart	37.71	63.861	.225	.302	.825
Jiomart	36.90	59.824	.403	.210	.808
AJIO	36.59	58.545	.487	.295	.799
BigBasket	36.30	56.948	.642	.549	.784

Snapdeal	36.23	56.923	.666	.572	.783
Netmeds	36.14	57.146	.679	.589	.782
TATA1MG	36.12	57.776	.661	.624	.784
Firstcry	36.14	57.373	.637	.575	.785
Other	36.47	59.254	.464	.370	.802

Table no.16

Scale Statistics			
Mean	Variance	Std. Deviation	N of Items
40.41	70.581	8.401	11

Table no.17

Reliability Statistics	
Cronbach's Alpha	N of Items
.815	11

Table no.18

Interpretation:

The value of Cronbach’s alpha coefficient should be between 0 and 1, with a higher number indicating better reliability. Cronbach’s alpha coefficient should be higher than 0.70; that scale has good internal validity and reliability.

Therefore, the results indicate that scale On On Product that you online has Cronbach’s alpha value = 0.815 implying Good reliability and internal consistency.

4. On Satisfaction Level of online shopping

Case Processing Summary			
		N	%
Cases	Valid	278	100.0
	Excluded[a]	0	.0
	Total	278	100.0
a. Listwise deletion based on all variables in the procedure.			

Table no.19

Item Statistics			
	Mean	Std. Deviation	N
Product Quality	3.62	1.121	278
Discounts and Offers	3.61	1.196	278
Products Prices	3.63	1.112	278
Return Policy	3.65	1.305	278
Exchange Policy	3.62	1.318	278
Payment options	3.88	1.162	278

Availability of wide range of brands	3.71	1.239	278
Varieties of items	3.77	1.204	278
Shipping time	3.66	1.150	278

Table no.20

Summary Item Statistics							
	Mean	Min.	Max.	Range	Maximum / Minimum	Variance	N of Items
Item Means	3.682	3.612	3.878	.266	1.074	.008	9
Item Variances	1.447	1.237	1.738	.501	1.405	.033	9
Inter-Item Covariance	.865	.698	1.422	.724	2.037	.019	9
Inter-Item Correlations	.599	.478	.827	.349	1.730	.004	9

Table no.21

Item-Total Statistics					
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Product Quality	29.53	62.257	.668	.501	.927
Discounts and Offers	29.53	59.882	.758	.616	.922
Products Prices	29.51	61.406	.728	.572	.924
Return Policy	29.49	58.287	.770	.733	.921
Exchange Policy	29.52	58.236	.763	.723	.921
Payment options	29.26	60.404	.752	.584	.922

Availability of wide range of brands	29.43	59.401	.754	.657	.922
Varieties of items	29.37	59.621	.767	.679	.921
Shipping time	29.48	60.864	.732	.550	.923

Table no.22

Scale Statistics			
Mean	Variance	Std. Deviation	N of Items
33.14	75.334	8.680	9

Table no.23

Reliability Statistics	
Cronbach's Alpha	N of Items
.931	9

Table no.24

Interpretation:

The value of Cronbach's alpha coefficient should be between 0 and 1, with a higher number indicating better reliability. Cronbach's alpha coefficient should be higher than 0.70; that scale has good internal validity and reliability.

Therefore, the results indicate that scale On On Product that you online has Cronbach's alpha value = 0.931 implying Excellent reliability and internal consistency.

5. On what occasions we prefer online shopping

Case Processing Summary			
		N	%
Cases	Valid	278	100.0
	Excluded[a]	0	.0
	Total	278	100.0

a. Listwise deletion based on all variables in the procedure.

Table no.25

Item Statistics			
	Mean	Std. Deviation	N
During sales	2.09	.990	278
During festive seasons	2.31	.911	278
Occasionally	2.51	.898	278

Table no.26

Summary Item Statistics							
	Mean	Min	Max	Range	Maximum / Minimum	Variance	N of Items
Item Means	2.305	2.094	2.507	.414	1.198	.043	3
Item Variances	.872	.807	.980	.174	1.215	.009	3

Inter-Item Covariance	.235	.079	.321	.242	4.074	.015	3
Inter-Item Correlations	.273	.089	.392	.304	4.429	.021	3

Table no.27

Item-Total Statistics					
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
During sales	4.82	2.278	.258	.118	.563
During festive seasons	4.60	1.945	.494	.248	.162
Occasionally	4.41	2.423	.286	.156	.506

Table no.28

Scale Statistics			
Mean	Variance	Std. Deviation	N of Items
6.91	4.029	2.007	3

Table no.29

Reliability Statistics	
Cronbach's Alpha	N of Items
.526	3

Table no.30

Interpretation:

The value of Cronbach's alpha coefficient should be between 0 and 1, with a higher number indicating better reliability. Cronbach's alpha coefficient should be higher than 0.70; that scale has good internal validity and reliability.

Therefore, the results indicate that scale On On Product that you online has Cronbach's alpha value = 0.526 implying poor reliability and internal consistency.

Demographic Factors

Age distribution of consumer:

AGE					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	under 18	40	12.3	12.3	12.3
	18-24	134	41.4	41.4	53.7
	25-34	53	16.4	16.4	70.1
	35-44	36	11.1	11.1	81.2
	44-54	35	10.8	10.8	92.0
	above 54	26	8.0	8.0	100.0
	Total	324	100.0	100.0	

Table no.31

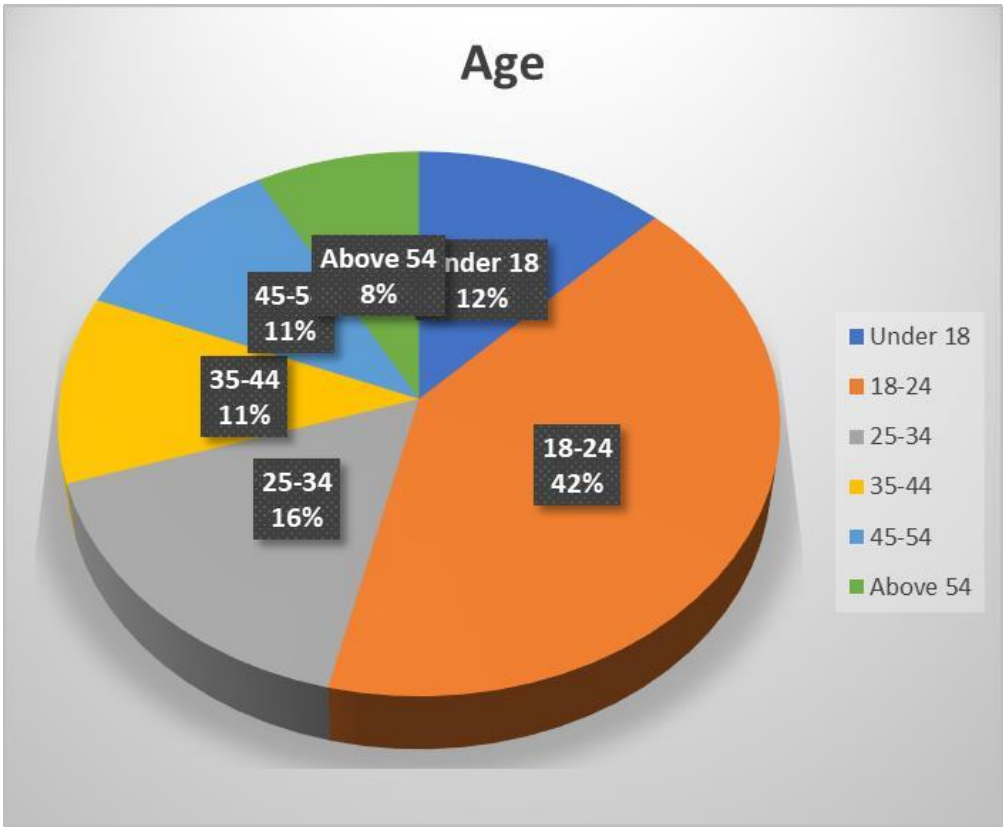


Figure no.1

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	male	126	38.9	38.9	38.9
	female	186	57.4	57.4	96.3
	prefer not to answer	12	3.7	3.7	100.0
	Total	324	100.0	100.0	

Table no.32

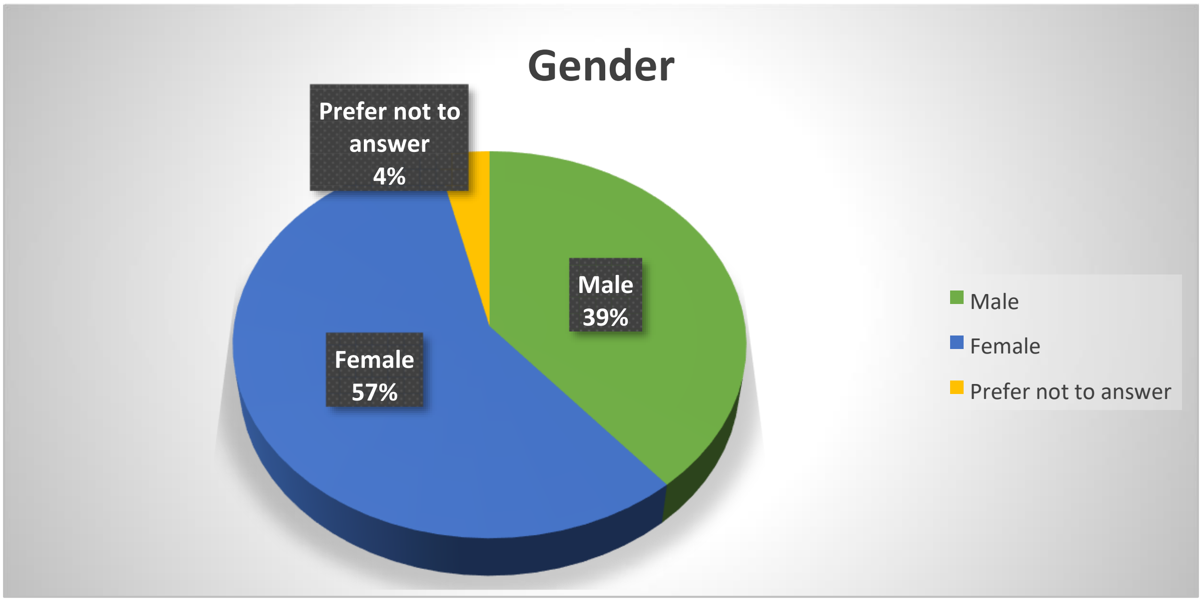


Figure no.2

Educational distribution of consumers:

EDUCATION					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Up to ssc	36	11.1	11.1	11.1
	Up to hsc	58	17.9	17.9	29.0
	bachelors degree	128	39.5	39.5	68.5
	masters degree or diploma	71	21.9	21.9	90.4
	professional	21	6.5	6.5	96.9
	other	10	3.1	3.1	100.0
	Total	324	100.0	100.0	

Table no.33

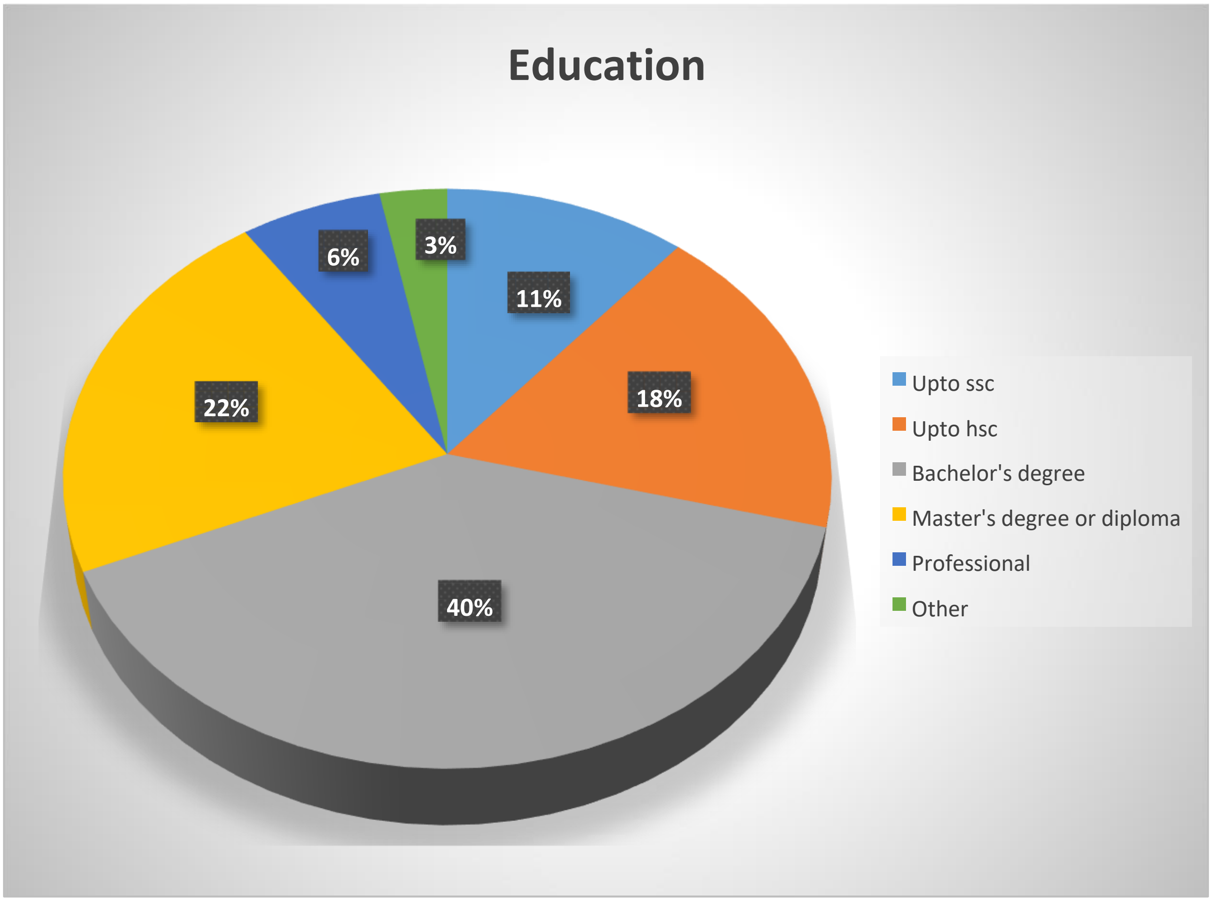


Figure no.3

Profession of consumers:

CURRENT STATUS					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	employed	111	34.3	34.3	34.3
	unemployed	39	12.0	12.0	46.3
	housewife	32	9.9	9.9	56.2
	a student	131	40.4	40.4	96.6
	retired	11	3.4	3.4	100.0
	Total	324	100.0	100.0	

Table no.34

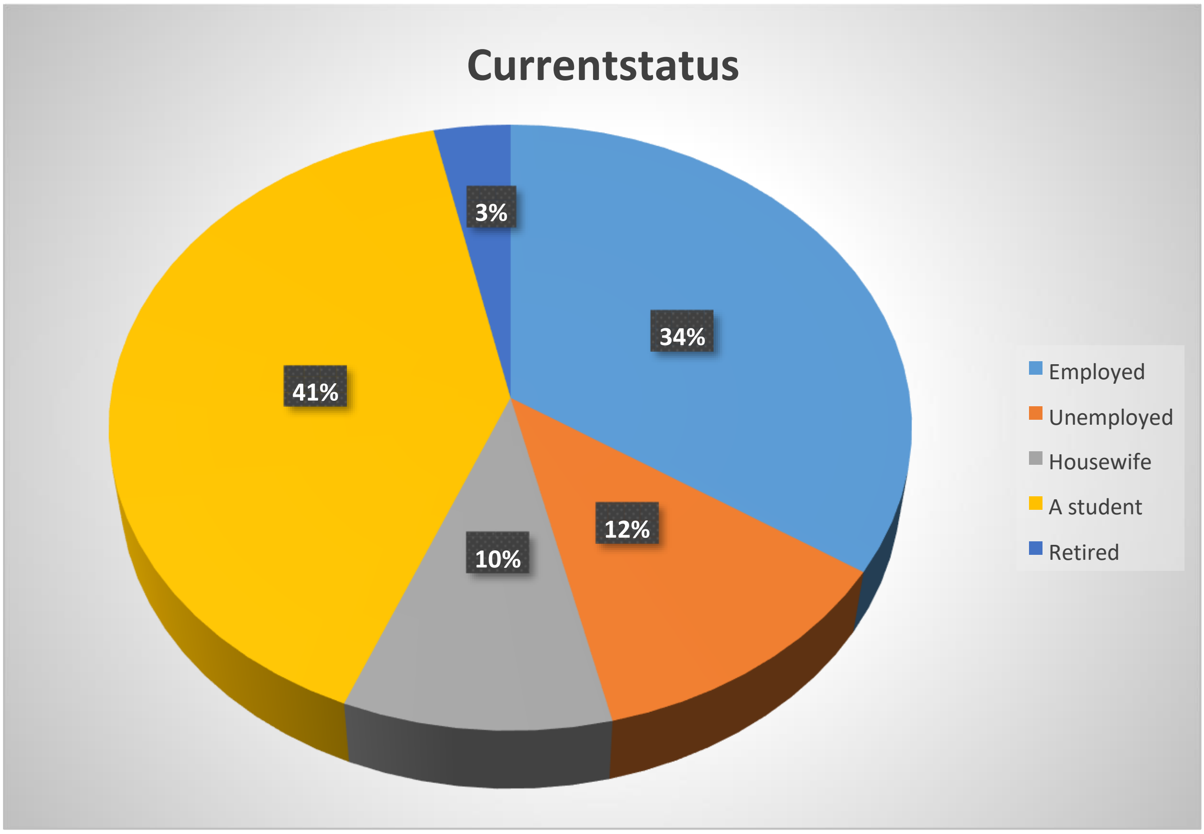


Figure no.4

Distribution on the basis of marital status of consumers:

MARITAL STATUS					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	married	107	33.0	33.0	33.0
	unmarried	199	61.4	61.4	94.4
	divorced	8	2.5	2.5	96.9
	widowed	10	3.1	3.1	100.0
	Total	324	100.0	100.0	

Table no.35

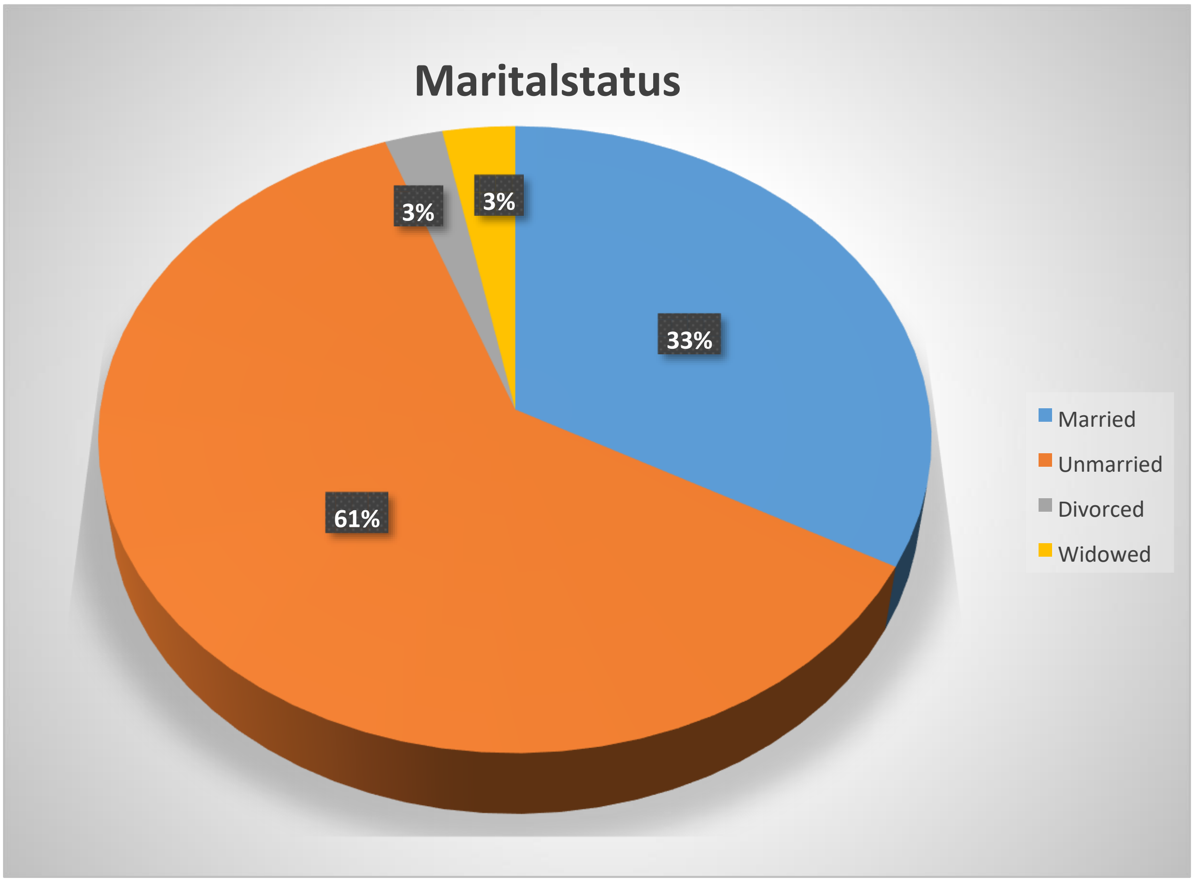


Figure no.5

Distribution of members in the family of consumers:

MEMBERS IN FAMILY					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	6	1.9	1.9	1.9
	2	21	6.5	6.5	8.3
	3	85	26.2	26.2	34.6
	4	136	42.0	42.0	76.5
	more than 4	76	23.5	23.5	100.0
	Total	324	100.0	100.0	

Table no.36

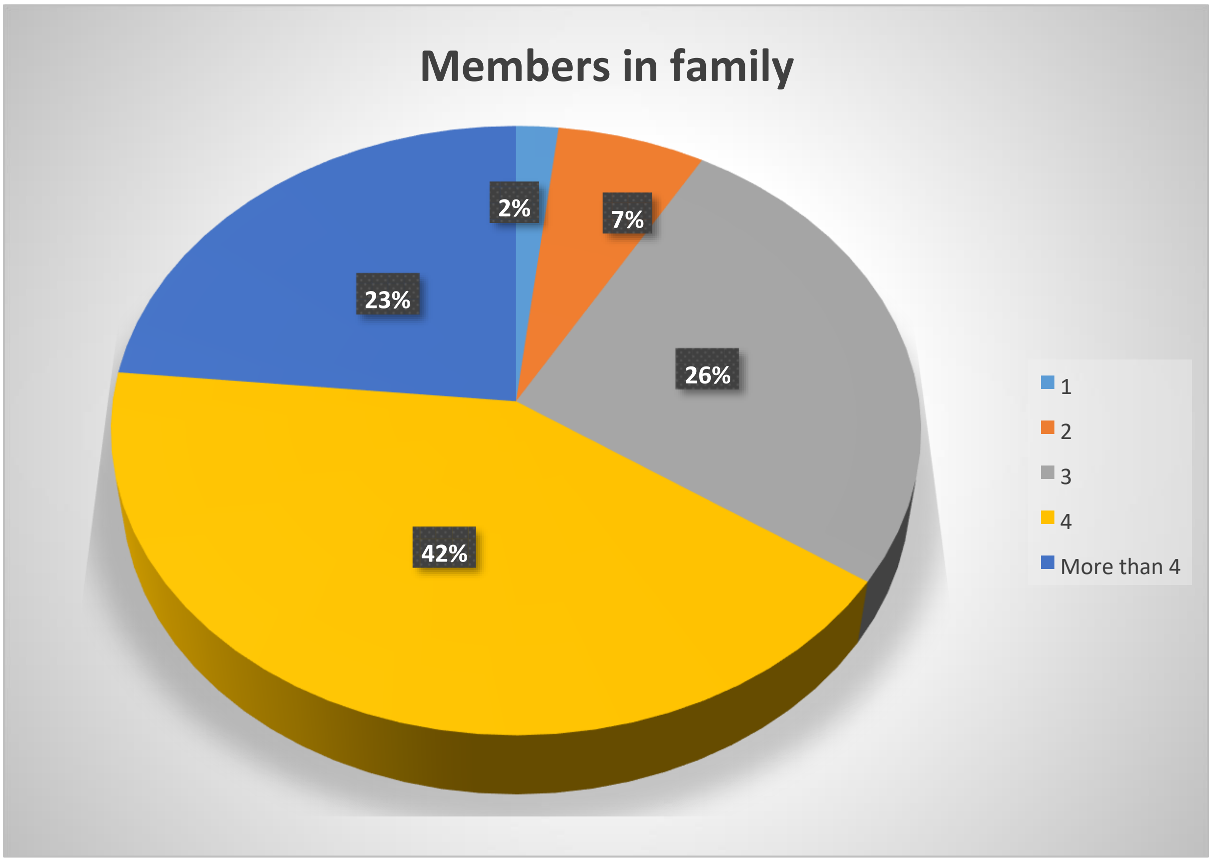


Figure no.6

Distribution on the basis of income of the consumer:

Income					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	less than 1 lakh	116	35.8	35.8	35.8
	1-5 lakh	129	39.8	39.8	75.6
	5-9 lakh	46	14.2	14.2	89.8
	more than 9 lakh	32	9.9	9.9	99.7
	5	1	.3	.3	100.0
	Total	324	100.0	100.0	

Table no.37

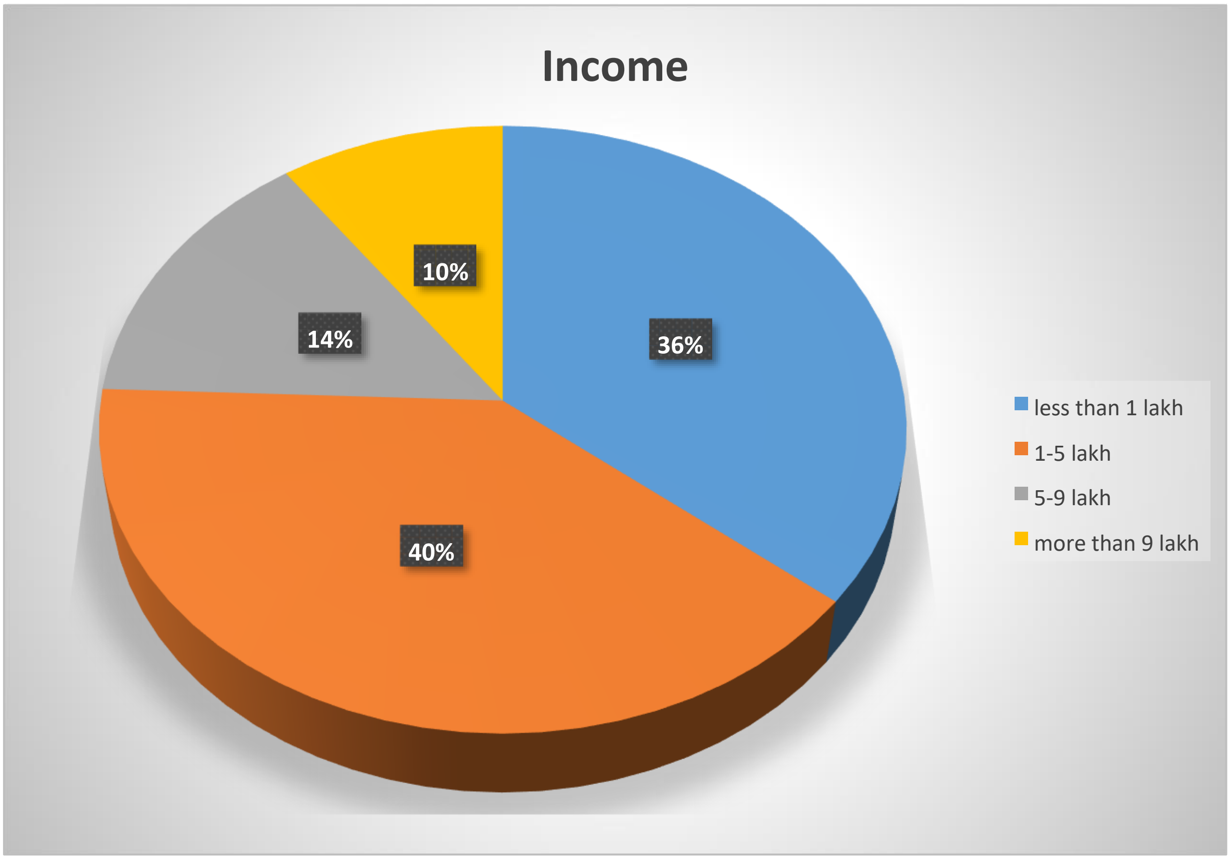


Figure no.7

Frequency of online purchase:

Frequently Shopping						
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	twice a week	13	4.0	4.7	4.7
	weekly	34	10.5	12.2	16.8
	monthly	84	25.9	30.1	47.0
	occasionally	148	45.7	53.0	100.0
	Total	279	86.1	100.0	
Missing	System	45	13.9		
Total		324	100.0		

Table no.38

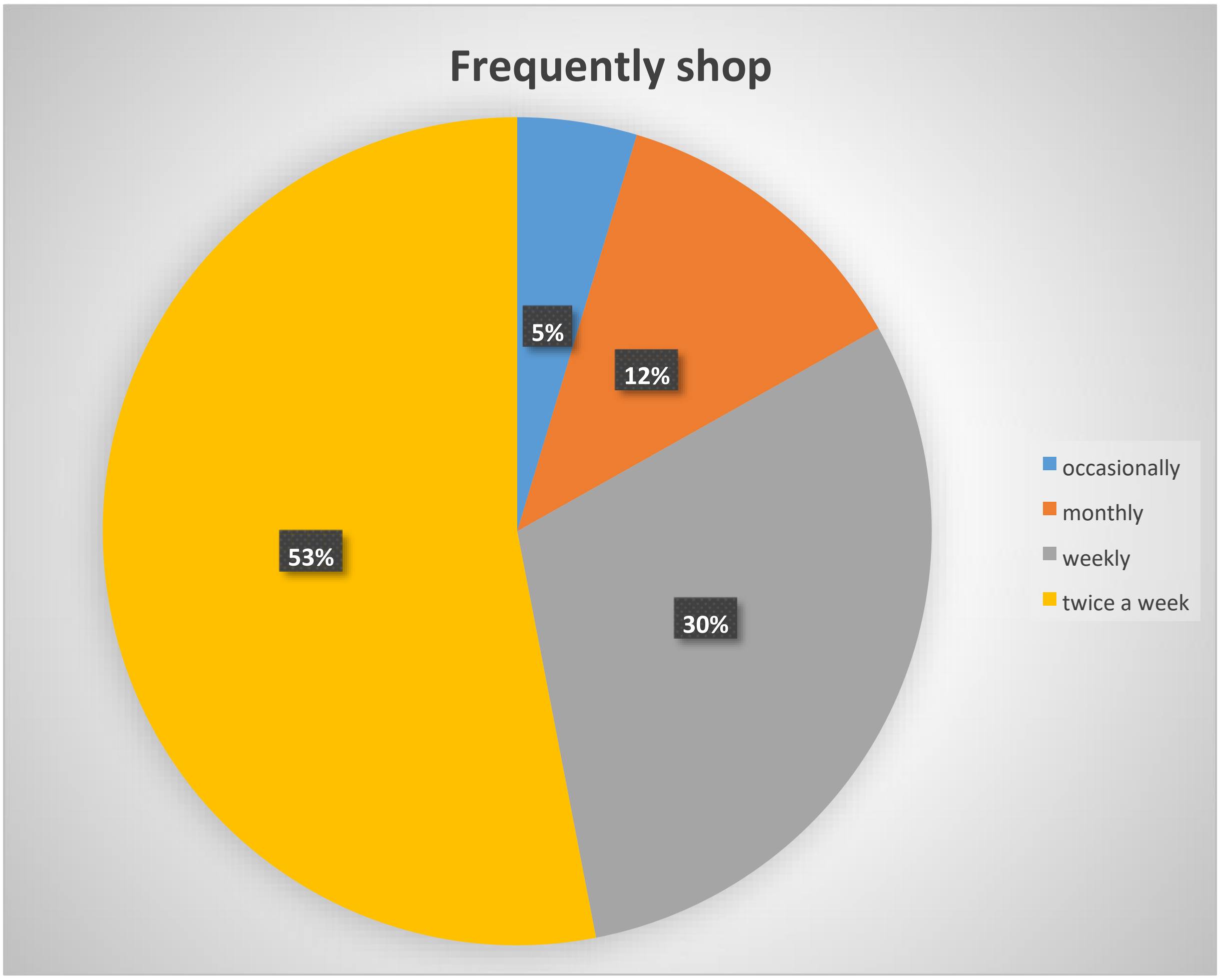


Figure no.8

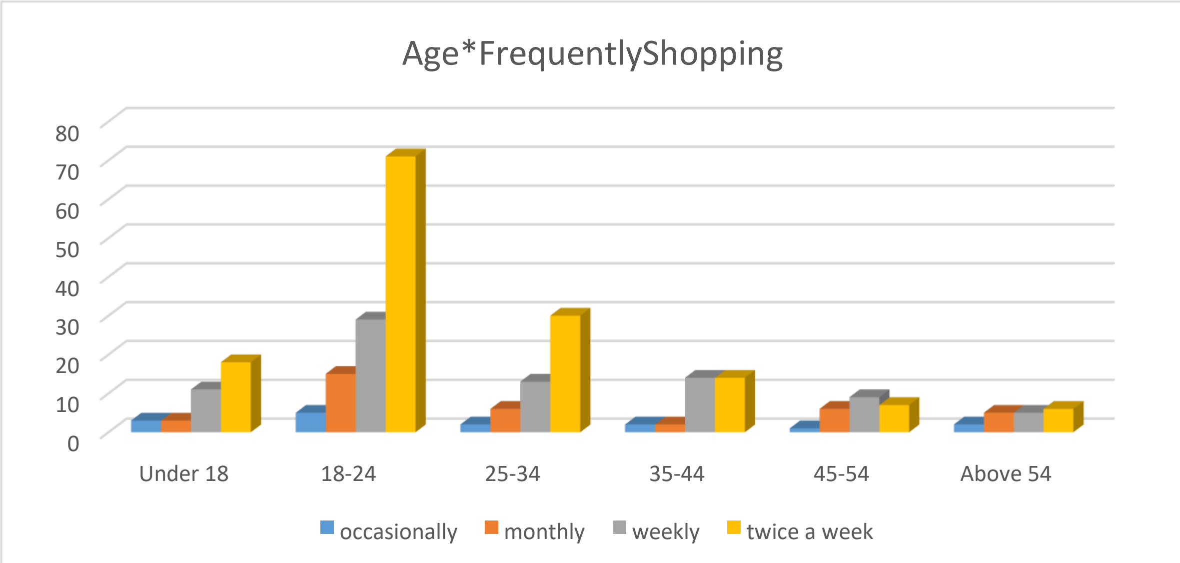
CROSS TABULATION:

1. Frequently of shopping on the basis of age:

Age * Frequently Shopping Cross tabulation						
Count						
		Frequently Shopping				Total
		occasionally	monthly	weekly	twice a week	
Age	Under 18	3	3	11	18	35
	18-24	5	15	29	71	120
	25-34	2	6	13	30	51
	35-44	2	2	14	14	32
	45-54	1	6	9	7	23

	Above 54	2	5	5	6	18
Total		15	37	81	146	279

Table no.39



Graph no.1

- In the age group of under 18 there are 3 people who shop occasionally, 3 shop monthly, 11 people do shopping weekly and 18 people shop twice a week.
- In the age group 18-24 there are 5 people who shop occasionally, 15shop monthly, 29 people do shopping weekly and 71 people shop twice a week.
- In the age group 25-34 there are 2 people who shop occasionally, 6 shop monthly, 13 people do shopping weekly and 30 people shop twice a week.
- In the age group 35-44 there are 2 people who shop occasionally, 2 shop monthly, 14 people do shopping weekly and 14 people shop twice a week.
- In the age group 45-54 there is 1 person who shops occasionally, 6 people shop monthly, 9 people do shopping weekly and 7 people shop twice a week.
- In the age group of above 54 there are 2 people who shop occasionally, 5 people shop monthly, 5 people do shopping weekly and 6 people shop twice a week.

Chi-Square Tests			
	Value	df	Asymptotic Significance (2sided)
Pearson Chi-Square	19.762[a]	15	.181
Likelihood Ratio	18.694	15	.228
Linear-by-Linear Association	6.037	1	.014

N of Valid Cases	279		
a. 9 cells (37.5%) have expected count less than 5. The minimum expected count is .97.			

Table no.40 **Interpretation:**

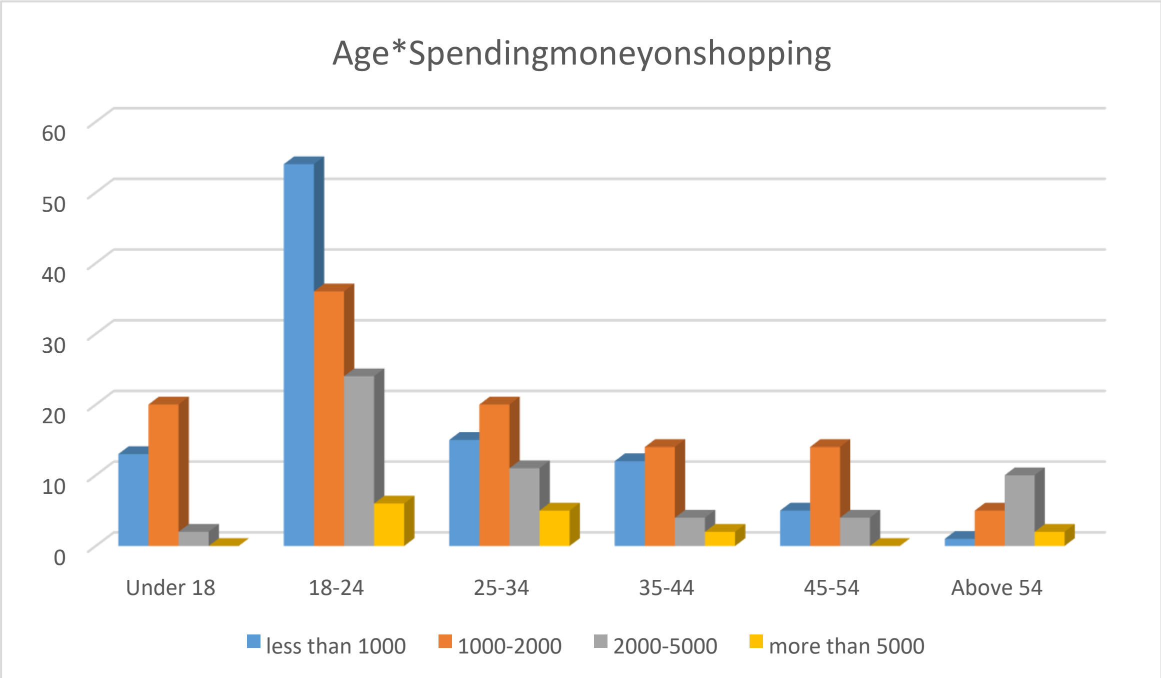
Here, p-value > 0.05; hence we conclude that the variables Frequently Shopping and Age are independent

i.e. there is no correlation and is insignificant.

2. Spending money on shopping on the basis of age:

Age * Spending money on shopping Cross tabulation						
Count						
		Spending money on shopping				Total
		less than 1000	1000-2000	2000-5000	more than 5000	
Age	Under 18	13	20	2	0	35
	18-24	54	36	24	6	120
	25-34	15	20	11	5	51
	35-44	12	14	4	2	32
	45-54	5	14	4	0	23
	Above 54	1	5	10	2	18
Total		100	109	55	15	279

Table no.41



Graph no.2

- In the age group of under 18 there are 13 people who spend less than 1000rs on monthly basis 20 people spend 1000 to 2000rs on monthly basis, 2 people spend 2000 to 5000rs on monthly basis and there is no one in the age group who spend more than 5000 on monthly basis.
- In the age group of 18-24 there are 54 people who spends less than 1000 on monthly basis , 36 people spends 1000 to 2000 on monthly basis, 24 people spends 2000 to 5000 on monthly basis and there is 6 who spend more than 5000 on monthly basis.
- In the age group of 25-34 there are 15 people who spends less than 1000 on monthly basis , 20 people spends 1000 to 2000 on monthly basis, 11 people spends 2000 to 5000 on monthly basis and there is 5 who spend more than 5000 on monthly basis.
- In the age group of 35-44 there are 12 people who spends less than 1000 on monthly basis , 14 people spends 1000 to 2000 on monthly basis, 4 people spends 2000 to 5000 on monthly basis and there is 2 who spend more than 5000 on monthly basis.
- In the age group of 45-54 there are 5 people who spends less than 1000 on monthly basis , 14 people spends 1000 to 2000 on monthly basis, 4 people spends 2000 to 5000 on monthly basis and there is no one in the age group who spend more than 5000 on monthly basis.

- In the age group of above 55 there are 1 people who spends less than 1000 on monthly basis , 5 people spends 1000 to 2000 on monthly basis, 10 people spends 2000 to 5000 on monthly basis and there is 2 who spend more than 5000 on monthly basis.

Chi-Square Tests			
	Value	df	Asymptotic Significance (2sided)
Pearson Chi-Square	40.705[a]	15	.000
Likelihood Ratio	42.880	15	.000
Linear-by-Linear Association	11.871	1	.001
N of Valid Cases	279		
a. 7 cells (29.2%) have expected count less than 5. The minimum expected count is .97.			

Table no.42 **Interpretation:**

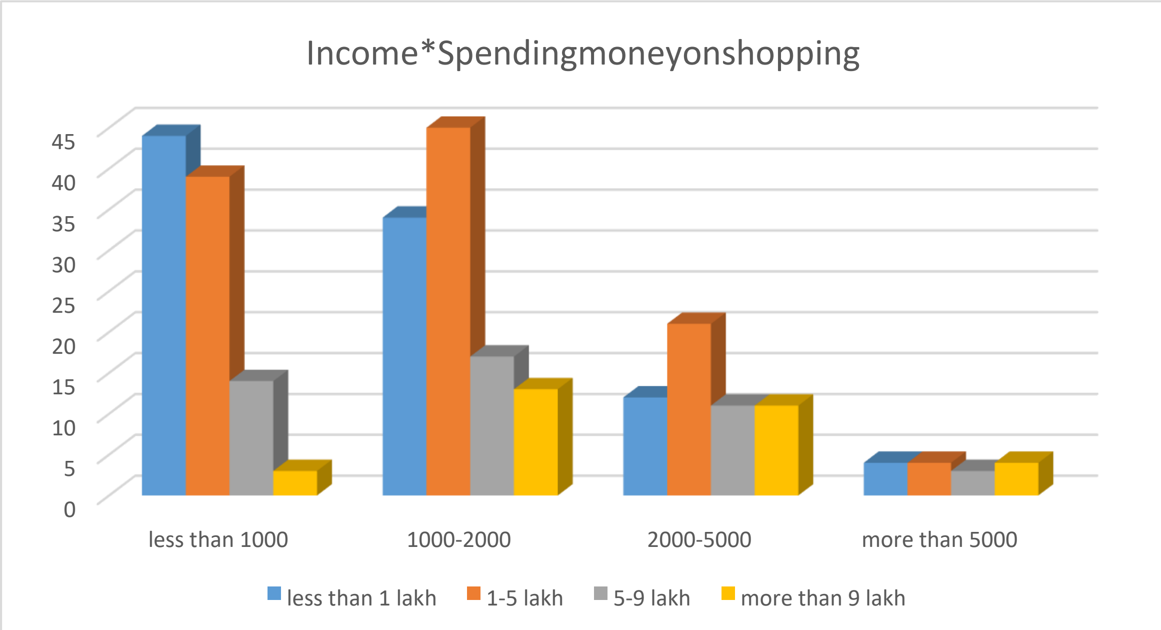
Here, p-value < 0.01; hence we conclude that the variables Spending money on Shopping and Age are dependent

i.e. there is correlation and is significant.

3. Spending money on online shopping on the basis of income:

Income * Spending money on shopping Cross tabulation						
Count						
		Spending money on shopping				Total
		less than 1000	1000-2000	2000-5000	more than 5000	
Income	less than 1 lakh	44	34	12	4	94
	1-5 lakh	39	45	21	4	109
	5-9 lakh	14	17	11	3	45
	more than 9 lakh	3	13	11	4	31
Total		100	109	55	15	279

Table no.43



Graph no.3

From the above chart , we can visualize that

- 44 people in households with less than one lakh rupees per year spend less than 1000rs on shopping, 34 spend between 1000rs to 2000rs per month, 12 spend between 2000rs to 5000rs per month, and four spend more than 5000rs per month.
- 39 individuals with incomes within 1 and 5 lakh rupees spend less than 1000rs on shopping each month, 45 individuals spend between 1000 and 2000rs each month, 21 individuals spend between 2000 and 5000rs each month, and four individuals spend more than 5000rs each month.
- In households earning between 5 and 9 lakh rupees, 14 people spend less than 1000rs a month on shopping, 17 people spend between 1000 and 2000rs , 11 people spend between 2000 and 5000rs a month on shopping, and 3 people spend more than 5000rs a month on shopping.
- 3 people make less than 1000rs per month, 13 people spend between 1000 and 2000rs per month, eleven people spend between 2000 and 5000rs per month, and 4 people make more than 5000rs per month.

Chi-Square Tests

	Value	df	Asymptotic Significance (2sided)
Pearson Chi-Square	20.730[a]	9	.014
Likelihood Ratio	21.600	9	.010
Linear-by-Linear Association	17.629	1	.000
N of Valid Cases	279		
a. 2 cells (12.5%) have expected count less than 5. The minimum expected count is 1.67.			

Table no.44

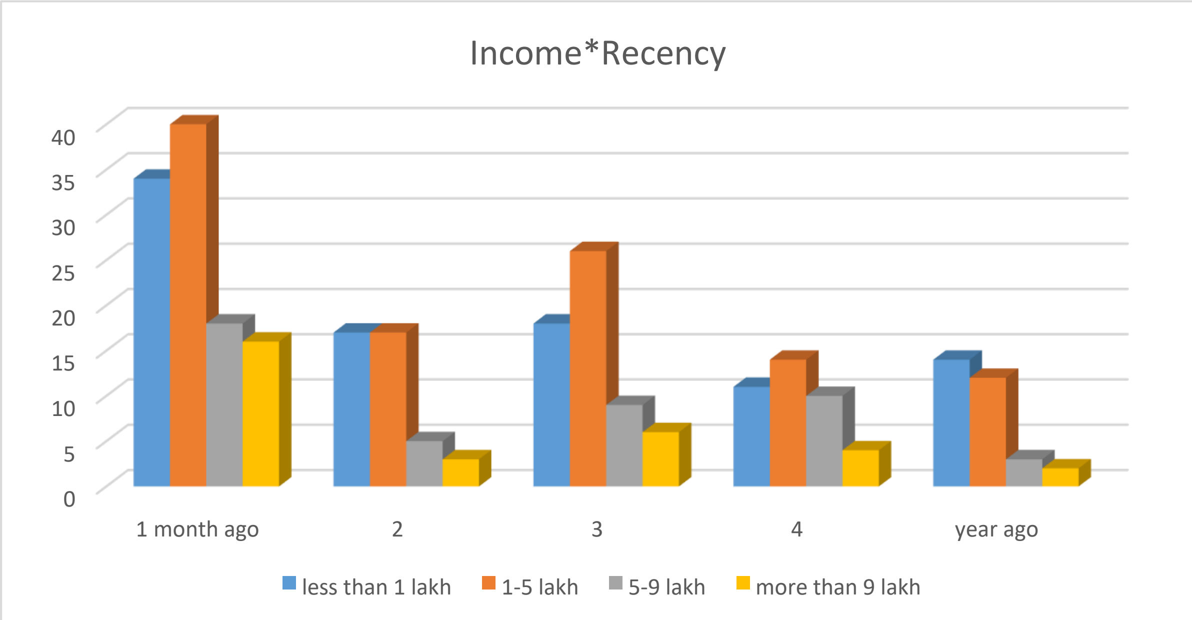
Interpretation:

Here, p-value < 0.05; hence we conclude that the variables Frequently Shopping and Age are dependent i.e. there is correlation and is significant.

4. Recency on the basis of income:

Income * Recency Cross tabulation							
Count							
		Recency					Total
		1 month ago	2	3	4	year ago	
Income	less than 1 lakh	34	17	18	11	14	94
	1-5 lakh	40	17	26	14	12	109
	5-9 lakh	18	5	9	10	3	45
	more than 9 lakh	16	3	6	4	2	31
Total		108	42	59	39	31	279

Table no.45



Graph no.4

- Customers who last shopped less than one lakh rupees were 34, those who last shopped three months ago were 17, those who last shopped six months ago were 18, those who last shopped nine months ago were 11, and those who last shopped a year ago were 14.
- Consumers who made between one and five lakh purchases a month ago are 40, those who made purchases three months ago are 17, those made six months ago are 26, those made nine months ago are 14, and those made a year ago are twelve.
- Customers who last shopped between 5 and 9 lakh about one month ago are 18, those who last shopped around three months ago are five, those who last shopped around nine months ago are ten, and those who last shopped around a year ago are three.
- Customers who last spent more than 9 lakh about a month ago are 16, those who last shopped around 3 months ago are 3, those who last shopped around 6 months ago are 4, and those who last shopped around 9 months ago are 2.

Chi-Square Tests			
	Value	df	Asymptotic Significance (2sided)
Pearson Chi-Square	9.188[a]	12	.687
Likelihood Ratio	8.971	12	.705
Linear-by-Linear Association	1.274	1	.259
N of Valid Cases	279		
a. 3 cells (15.0%) have expected count less than 5. The minimum expected count is 3.44.			

Table no.46 **Interpretation:**

Here, p-value > 0.05; hence we conclude that the variables Frequently Shopping and Age are independent

i.e. there is no correlation and is insignificant.

CUSTOMER SEGMENTATION

- ***What is customer segmentation?***

Customer segmentation is the practise of grouping your consumers according to shared traits, such as behaviours or demographics, so your sales team or marketing team may more effectively sell to those customers. The creation of a marketing persona or product user persona may also be discussed using these customer segmentation categories as a starting point. This is due to the fact that efficient customer segmentation analysis is frequently used to guide a brand's messaging and positioning, aids organisations in determining which new goods or services they might want to invest in, and reveals methods to enhance the sales process. As a result, in order for marketing personas to be successful, they must be intimately related to those groups.

In business to consumer marketing, customers are segmented according to demographic factors that include:

- AGE
- GENDER
- EDUCATION
- CURRENT STATUS
- LIFE STAGE

- ***What is the purpose of segmentation?***

Customer segmentation helps businesses better connect with customers by tailoring their communications, with the added benefits of improving customer experiences, optimising ad spend, and more. The goal of customer segmentation is to gain a deeper understanding of your customers by categorising them into nuanced category groups based on their actions, interests, geographic location, demographic data, and other factors.

Businesses may tailor their marketing strategy by using segmentation to provide customers the brand experiences they want. This may include everything from creating in-store displays to providing personalised digital adverts to conducting targeted email campaigns. Customers experience greater engagement and connection as a consequence. When businesses properly cultivate their businessconsumer relationships, personalization may result in customer loyalty and brand advocacy.

Benefits Of Customer Segmentation:

- Drives conversions and sales
- Reduces risks and costs
- Improves customer service and brand experiences
- Identifies opportunities and challenges
- Creates committed relationships and engaged communities.

To determine which customers are most likely to respond to a marketing effort or develop into loyal customers, RFM analysis serves as a method for customer segmentation. Recency, Frequency, and Monetary, or RFM, are the three main variables that are used to assess consumer behaviour.

Recency refers to how recently the customer made a purchase, Frequency refers to how often a customer makes a purchase, and Monetary refers to how much money a customer spends on purchases.

By examining these criteria, organisations may identify their most valued consumers, target specific customer groups with targeted marketing campaigns, and establish strategies to boost customer loyalty and retention.

RFM Score calculation:

- Recency was measured on the scale of 1 to 5, where 1 being the purchase a month ago and 5 being the purchase a year ago. Customers with high recency are the one with scale 1, 2 and 3 and customers with low frequency are the one with scale 4 and 5
- Frequency was measured on the scale of 1 to 4, high frequency customers falls under the scale of 1 and 2, low frequency customers falls under the scale of 3 and 4.
- Spending was measured on the scale of 1 to 4, where high spending customers are the one who spends 2000- more than 5000 and falls under the scale of 3 and 4 and low spending customers are the one who spends less than 1000-2000 and falls under the scale of 1 and 2
- For eg RFM Score of 111 means, high recency, high frequency and low spending.

2 CLASSES SEGMENTATION:

Based on RFM Score customers were segmented into 2 Classes, loyal online consumers, loyal offline consumers.

- Loyal online consumers are the consumers with high recency, high frequency and high spending.
- Loyal offline consumers are the consumers with low recency, low frequency and low spending.

4 CLASSES SEGMENTATION:

We have further classified the loyal online and loyal offline consumers into loyal and potential loyal segments , this was done to gain a deeper understanding of the behaviour and characteristics of these customers. By creating these sub-segments, you can identify the customers who are most valuable to the business and design targeted marketing strategies to retain their loyalty.

For instance, the loyal online customers are those who have a high recency, high frequency, and high spending score. They are the most valuable customers who have a strong connection with the business. However, within this group, there may be some customers who have the potential to increase their spending or frequency of purchase with the right marketing approach. These customers are identified as the potential loyal online customers.

Similarly, the loyal offline customers are those with low recency, low frequency, and low spending scores. They are the least valuable customers and require a different marketing approach to encourage them to become more loyal. By identifying the potential loyal offline customers, you can design targeted strategies to engage them and convert them into loyal customers.

The purpose of creating these sub-segments is to gain insights into customer behaviour, identify the most valuable customers, and design targeted strategies to increase customer loyalty and retention.

Based on RFM Score customers were further segmented into 4 Classes, loyal online consumers, loyal offline consumers, potential loyal online consumers and potential loyal offline consumers.

- Loyal online consumers are the consumers with high recency, high frequency and high spending.
- Loyal offline consumers are the consumers with low recency, low frequency and low spending.
- Potential loyal online consumers are the consumers with high recency, high frequency and low spending.
- Potential loyal offline consumers are the consumers with low recency, low frequency and high spending.

- **The table below shows the different statistical method employed:**

Method	Purpose	Variables under study
1.Logistic Regression	To develop model to predict customer loyalty segment on the basis of satisfaction level and to identify the factors that have a significant impact on customer loyalty segment.	Loyalty segment (2 classification), Satisfaction level
2.Naïve Bayes	To develop model to predict customer loyalty segment on the basis of satisfaction level and to identify the	Loyalty segment (2 classification), Satisfaction level
	factors that have a significant impact on customer loyalty segment	
3.SVM	To identify the factors that significantly influences customer loyalty to an online purchasing platform.	Loyalty segment (2 classification), Satisfaction level
4.Linear Discriminant Analysis	To identify the factors discriminating customers into loyalty segment based on satisfaction level	Loyalty segment (4 classification), Satisfaction level, Factor influencing online shopping
5.PCA	To factorize preferred online shopping platform into components and to analyse the variation of frequency and spending in each component on the basis of product preference	Preferred online shopping platforms, preferred product, Frequency, spending.
6.SEM	To see the relationship between factors influence online shopping and satisfaction level of customers who shop online	Satisfaction level, Factor influencing online shopping
7.CFA	To identify if there is any relationship between SL and MONTHLY SPEND, FI and MONTHLY SPEND, SL-Fi and MONTHLY SPEND	Satisfaction level, Factor influencing online shopping, loyalty segment (4 classification), monthly spend
8.Mediation	To understand effect of loyalty segment on association of SL and monthly spend	Satisfaction level, loyalty segment (4 classification), monthly spend

Table no.47

Logistic Regression

Objective:

To develop model to predict customer loyalty segment on the basis of satisfaction level and to identify the factors that have a significant impact on customer loyalty segment.

What is logistic Regression?

In regression analysis, logistic regression (or logit regression) is estimating the parameters of a logistic model (the coefficients in the linear combination). Logistic regression is a statistical analysis method to predict a binary outcome, such as yes or no, based on prior observations of a data set. A logistic regression model predicts a dependent data variable by analyzing the relationship between one or more existing independent variables.

Assumptions of Logistic Regression:

- Linearity of the logit: The relationship between the independent variables and the log odds of the dependent variable should be linear.
- Independence of observations: Each observation in the dataset should be independent of all other observations.
- No multicollinearity: There should be no perfect or high correlation between independent variables in the model.
- Large sample size: Logistic regression requires a sufficiently large sample size to ensure stable estimates and avoid overfitting.
- Absence of outliers: Outliers can have a significant effect on the coefficients of logistic regression and can influence the results.

Mathematical Model:

In the logistic regression model, the log odds (or logit) of the probability of the dependent variable is modelled as a linear combination of the independent variables. The equation for the logistic regression model can be represented as:

$$\text{logit}(p) = \beta 0 + \beta 1x1 + \beta 2x2 + ... + \beta nxn$$

where logit(p) is the log odds of the probability of the dependent variable, β0 is the intercept, β1-βn are the coefficients of the independent variables, and x1-xn are the values of the independent variables.

The logistic regression model estimates the coefficients of the independent variables that maximize the likelihood of the observed data. The observed data includes the values of the independent variables and the corresponding values of the dependent variable (customer loyalty).

The p-values of the coefficients in the logistic regression model indicate the statistical significance of each independent variable in explaining the dependent variable. A p-value less than 0.05 indicates that the independent variable is statistically significant in explaining the dependent variable.

Why use logistic Regression?

By using logistic regression, we can identify which independent variables (such as product quality, discounts and offers, etc.) have the greatest impact on the dependent variable (customer loyalty) while controlling for other factors.

In other words, logistic regression enables you to identify the significant factors that influence customer loyalty and how much they influence it. This is important because understanding these factors can help online purchasing platforms to make targeted improvements to their services to enhance customer loyalty.

Additionally, logistic regression models are easy to interpret, making them a useful tool for communicating results to stakeholders. Logistic regression models can provide insights that can help organizations make better decisions and ultimately improve customer loyalty.

Logistic regression is a powerful statistical tool that can help us achieve our objective of identifying the factors that significantly influence customer loyalty to an online purchasing platform.

Using Python:

```
In [16]: X =df[['ProductQuality',
                'DiscountsandOffers', 'ProductsPrices', 'ReturnPolicy',
                'ExchangePolicy', 'Paymentoptions', 'Availabilityofwiderangeofbrands',
                'Varietiesofitems', 'Shippingtime']]
         y = df['LoyaltySegment']
```

```
In [18]: from sklearn.linear_model import LogisticRegression
         from sklearn.model_selection import train_test_split
         from sklearn.metrics import accuracy_score
         import statsmodels.api as sm

         # Split the data into training and testing sets
         X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

         # Create the logistic regression model
         logreg = LogisticRegression()

         # Fit the model to the training data
         logreg.fit(X_train, y_train)

         # Predict on the testing data
         y_pred = logreg.predict(X_test)

         # Get the coefficients of the logistic regression model
         coefficients = logreg.coef_

         # Use statsmodels to fit the logistic regression model
         logit_model = sm.Logit(y_train, sm.add_constant(X_train))
         result = logit_model.fit()

         # Get the p-values of the coefficients
         p_values = result.pvalues[1:]
```

```
# Print the coefficients and p-values
print("Coefficients:", coefficients)
print("P-values:", p_values)

# Calculate the accuracy of the model
accuracy = accuracy_score(y_test, y_pred)

print(f"The accuracy of the logistic regression model is {accuracy:.2f}")
```

```
Optimization terminated successfully.
         Current function value: 0.446397
         Iterations 6
Coefficients: [[-0.62820684  0.12243819 -0.02793768 -0.17064847  0.0297562  -0.20204838
   0.57463113  0.24575499  0.06437767]]
P-values: ProductQuality                     0.006622
DiscountsandOffers                 0.568365
ProductsPrices                     0.895578
ReturnPolicy                       0.494906
ExchangePolicy                     0.884263
Paymentoptions                     0.392011
Availabilityofwiderangeofbrands    0.011167
Varietiesofitems                   0.341218
Shippingtime                       0.769521
dtype: float64
The accuracy of the logistic regression model is 0.88
```

Prediction

```
# Predict on new data
new_data = [[1, 2, 3, 4, 5, 6, 7, 8, 9], [2, 3, 4, 5, 6, 7, 8, 9, 10]]
new_pred = logreg.predict(new_data)
print("New predictions:", new_pred)
```

```
New predictions: [1 1]
```

Findings:

- 'ProductQuality', 'ProductsPrices', 'ReturnPolicy', 'ExchangePolicy', 'Paymentoptions', and 'Shippingtime' have negative coefficients, indicating that they have a negative impact on customer loyalty.
- On the other hand, 'DiscountsandOffers', 'Availabilityofwiderangeofbrands', and 'Varietiesofitems' have positive coefficients, indicating a positive impact on customer loyalty.
- The p-values of the coefficients indicate the significance of each independent variable in the logistic regression model. A p-value less than 0.05 indicates that the independent variable is significant in the model.
- From the output, we can see that 'ProductQuality', 'Availabilityofwiderangeofbrands', and 'Varietiesofitems' have p-values less than 0.05, indicating that they are significant in the model.
- On passing the unseen data for 2 metrics the model predicted that they belong to loyal online consumer segment.

Naïve Bayes Classification

Objective:

To develop model to predict customer loyalty segment on the basis of satisfaction level and to identify the factors that have a significant impact on customer loyalty segment.

What is Naïve Bayes classification?

Naive Bayes is a kind of classifier which uses the Bayes Theorem. It predicts membership probabilities for each class such as the probability that given record or data point belongs to a particular class. The class with the highest probability is considered as the most likely class.

Mathematical Model

We want to predict the Loyalty Segment of a customer based on their features such as Product Quality, Discounts and Offers, Products Prices, Return Policy, Exchange Policy, Payment options, Availability of wide range of brands, Varieties of items, and Shipping time.

To do this, Naive Bayes calculates the probability of each possible Loyalty Segment given the values of these features. It assumes that the features are conditionally independent given the Loyalty Segment, meaning that each feature has no influence on the other features.

Based on this assumption, the conditional probability of a Loyalty Segment y given the feature vector X can be calculated as follows:

P(y | X) = P(y) * P(Product Quality | y) * P(Discounts and Offers | y) * P(Products Prices | y) * P(Return Policy | y) * P(ExchangePolicy | y) * P(Payment options | y) * P(Availabilityofwiderangeofbrands | y) * P(Varieties of items | y) * P(Shipping time | y)

where P(y) is the prior probability of the LoyaltySegment y, which can be estimated using the training data.

Finally, the Naive Bayes algorithm chooses the LoyaltySegment y that maximizes the posterior probability P(y | X). In other words, it predicts the LoyaltySegment with the highest probability based on the given feature values.

Why used Naïve bayes classifier?

Naive Bayes classification assumes that the independent variables are conditionally independent given the loyalty segment, which means that the probability of observing a particular combination of independent variables for a customer can be calculated as the product of the individual probabilities of each independent variable given the loyalty segment.

Naive Bayes classification can help us identify the factors that have the greatest impact on customer loyalty, by calculating the conditional probabilities of each independent variable given each loyalty segment. These conditional probabilities can then be used to rank the independent variables in terms of their influence on customer loyalty.

Using python:

```
In [20]: from sklearn.naive_bayes import MultinomialNB
         from sklearn.model_selection import train_test_split
         from sklearn.metrics import accuracy_score
         import statsmodels.api as sm

         # Split the data into training and testing sets
         X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

         # Create the Naive Bayes model
         nb = MultinomialNB()

         # Fit the model to the training data
         nb.fit(X_train, y_train)

         # Predict on the testing data
         y_pred = nb.predict(X_test)

         # Calculate the accuracy of the model
         accuracy = accuracy_score(y_test, y_pred)

         print(f"The accuracy of the Naive Bayes model is {accuracy:.2f}")
```

```
# Use statsmodels to fit the model and get the p-values of the coefficients
X_train = sm.add_constant(X_train)
nb_model = sm.MNLogit(y_train, X_train)
nb_results = nb_model.fit()
nb_coef = nb_results.params
nb_pvalues = nb_results.pvalues

print("Coefficients:")
print(nb_coef)
print("P-values:")
print(nb_pvalues)
```

```
The accuracy of the Naive Bayes model is 0.84
Optimization terminated successfully.
         Current function value: 0.446397
         Iterations 6
```

```
Coefficients:
                                     0
const                         1.579117
ProductQuality               -0.671663
DiscountsandOffers            0.144085
ProductsPrices               -0.030929
ReturnPolicy                 -0.189976
ExchangePolicy                0.037524
Paymentoptions               -0.211573
Availabilityofwiderangeofbrands  0.610514
Varietiesofitems              0.244380
Shippingtime                  0.068323
P-values:
                                     0
const                         0.038151
ProductQuality                0.006622
DiscountsandOffers            0.568365
ProductsPrices                0.895578
ReturnPolicy                  0.494906
ExchangePolicy                0.884263
Paymentoptions                0.392011
Availabilityofwiderangeofbrands  0.011167
Varietiesofitems              0.341218
Shippingtime                  0.769521
```

Prediction

```
# Predict on new data
new_data = [[1, 2, 3, 4, 5, 6, 7, 8, 9], [2, 3, 4, 5, 6, 7, 8, 9, 10]]
new_data_pred = nb.predict(new_data)

print("Predictions for new data:")
print(new_data_pred)
```

Output:

```
Predictions for new data:
[1 1]
```

Findings:

- The predicted probability of the target variable being in class 1 (or any specific class, depending on the model setup) is exp(1.579117) / (1 + exp(1.579117)) = 0.829.
- The 'ProductQuality' input feature has a negative coefficient of -0.671663 and a low pvalue of 0.006622, indicating that it is significant and negatively affects the target variable.
- The 'Availabilityofwiderangeofbrands' feature has a positive coefficient of 0.610514 and a low p-value of 0.011167, indicating that it is significant and positively affects the target variable.
- The other input features do not have significant p-values, suggesting that they are not important for predicting the target variable in this model.
- The prediction on unseen data through naïve bayes predicted that the customer will be classifies as loyal online consumer.

Support Vector Machine

Objective:

To identify the factors that significantly influence customer loyalty to an online purchasing platform.

What is SVM ?

Support Vector Machine (SVM) is a Supervised Machine Learning algorithm that is used for regression and/or classification. Although it is occasionally quite helpful for regression, classification is where it is most often used. In short, SVM identifies a hyper-plane that establishes a distinction between the different kinds of information.

SVM works by mapping data to a high-dimensional feature space so that data points can be categorized, even when the data are not otherwise linearly separable. A separator between the categories is found, then the data are transformed in such a way that the separator could be drawn as a hyperplane.

Mathematical Model

The objective is to find the hyper plane that best separates the data into the different classes. The coefficients in our output represent the weights assigned to each feature in the hyper plane equation, while the p-values represent the statistical significance of each feature in predicting the target variable.

Specifically, the hyper plane equation for linear SVM can be written as:

w^T * x + b = 0

where w is the weight vector, x is the input vector, and b is the bias term. The weight vector represents the coefficients in your output, and the bias term is also part of the SVM model but is not explicitly shown in your output.

In terms of our variables, each coefficient represents the weight assigned to the corresponding feature in the hyper plane equation. For example, the coefficient for Product Quality represents the weight assigned to that feature in the hyper plane equation.

Why use SVM ?

SVM is used because it helps us to determine the most crucial attributes that help forecast customer loyalty in terms of our objective of examining the variables that influence customers' loyalty to your online shopping platform.

Based on the values of the input attributes, the SVM algorithm finds the hyper plane that best divides the data into distinct classes (in your example, loyal vs. non-loyal consumers). The weights allocated to each feature in the hyper plane equation are represented by the coefficients you acquire from SVM, which show how significant each feature is in predicting customer loyalty.

Hence, you may identify which factors have the most effects on client loyalty and change your marketing approach accordingly by studying the coefficients and p-values of the SVM model.

Using Python:

```
In [21]: from sklearn.svm import SVC
         from sklearn.model_selection import train_test_split
         from sklearn.metrics import accuracy_score
         import statsmodels.api as sm

         # Split the data into training and testing sets
         X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

         # Create the SVM model
         svm = SVC(kernel='linear')

         # Fit the model to the training data
         svm.fit(X_train, y_train)

         # Predict on the testing data
         y_pred = svm.predict(X_test)

         # Get the coefficients of the SVM model
         coefficients = svm.coef_

         # Use statsmodels to fit the SVM model
         svm_model = sm.OLS(y_train, sm.add_constant(X_train))
         result = svm_model.fit()

         # Get the p-values of the coefficients
         p_values = result.pvalues[1:]
```

```
# Print the coefficients, p-values, and accuracy score
print("Coefficients:", coefficients)
print("P-values:", p_values)
print("The accuracy of the SVM model is", accuracy_score(y_test, y_pred))
```

```
Coefficients: [[-1.76912293e-04 -1.10033171e-05  1.03645356e-04  3.98724877e-05
  -3.22755784e-05  4.26661543e-05  1.73155927e-05  8.83102804e-05
   1.08400340e-06]]
P-values: ProductQuality                     0.005393
DiscountsandOffers                 0.583045
ProductsPrices                     0.990823
ReturnPolicy                       0.580631
ExchangePolicy                     0.888167
Paymentoptions                     0.271933
Availabilityofwiderangeofbrands    0.011011
Varietiesofitems                   0.306320
Shippingtime                       0.886314
dtype: float64
The accuracy of the SVM model is 0.8392857142857143
```

Findings:

- The p-values of the coefficients indicate the statistical significance of the relationship between each feature and the LoyaltySegment.
- A p-value less than 0.05 suggests that the relationship is statistically significant, while a p-value greater than 0.05 suggests that the relationship is not statistically significant.
- In this case, only the features ProductQuality and Availabilityofwiderangeofbrands have p-values less than 0.05, indicating that they are statistically significant.

Conclusion:

The results of the logistic regression, Naive Bayes, and SVM models all suggest that the quality of the products and the availability of a wide range of brands have a significant impact on customer loyalty. While the logistic regression and Naive Bayes models also identified the variety of items as a significant factor, the SVM model did not find it to be statistically significant. Other factors such as product prices, return policy, exchange policy, payment options, and shipping time did not show significant impact on customer loyalty in any of the models.

Marketing Strategies

1. *Improve product quality*: As the logistic regression and SVM models both indicate that product quality has a significant impact on customer loyalty, focusing on improving product quality can be a viable strategy. This could involve investing in better raw materials, improving production processes, and introducing quality control measures to ensure consistent quality.

2. *Offer discounts and promotions*: The logistic regression model indicates that discounts and offers have a positive impact on customer loyalty. Offering discounts and promotions can help attract new customers and retain existing ones. These discounts could be targeted at specific products, time-bound offers, or loyalty-based discounts.

3. *Expand brand and product range*: The logistic regression and Naive Bayes models indicate that the availability of a wider range of brands and product varieties has a positive impact on customer loyalty. Offering more brands and products can attract a wider customer base and increase loyalty among existing customers.

4. *Improve customer service*: Although not explicitly mentioned in the models, good customer service is a key factor in customer loyalty. Improving customer service can involve investing in training for customer service personnel, implementing customer feedback systems, and addressing customer complaints promptly.

5. *Streamline return and exchange policies*: The logistic regression model indicates that return and exchange policies have a negative impact on customer loyalty. Streamlining these policies can help reduce customer frustration and improve loyalty.

6. *Improve shipping time*: The logistic regression model indicates that shipping time has a negative impact on customer loyalty. Improving shipping time can involve investing in faster shipping methods, better logistics management, and improving supply chain efficiency.

Linear Discriminant Analysis

Objective:

1. Discriminating customers into loyalty segment on the basis of customer satisfaction level across number of variables.

What is LDA?

A statistical method known as discriminant analysis uses a set of independent factors, called metric predictors, to identify whether a group belongs to a given individual. According to the distinct properties of the data, this technique's main purpose is to categorise each observation.

It can help you determine which independent variables have the most effects on the dependent variable. It explains how each variable affects the categorization, making it easier for you to comprehend.

Assumptions of LDA:

- The independent variables have a normal distribution.
- It is assumed that the variances between categories are constant across predictor levels.
- It is assumed that the predictor variables are independent. A correlation between them may weaken the analysis's predictive ability.
- In addition to the variables' independence, the samples' independence is also taken into account.

Why did we employ LDA?

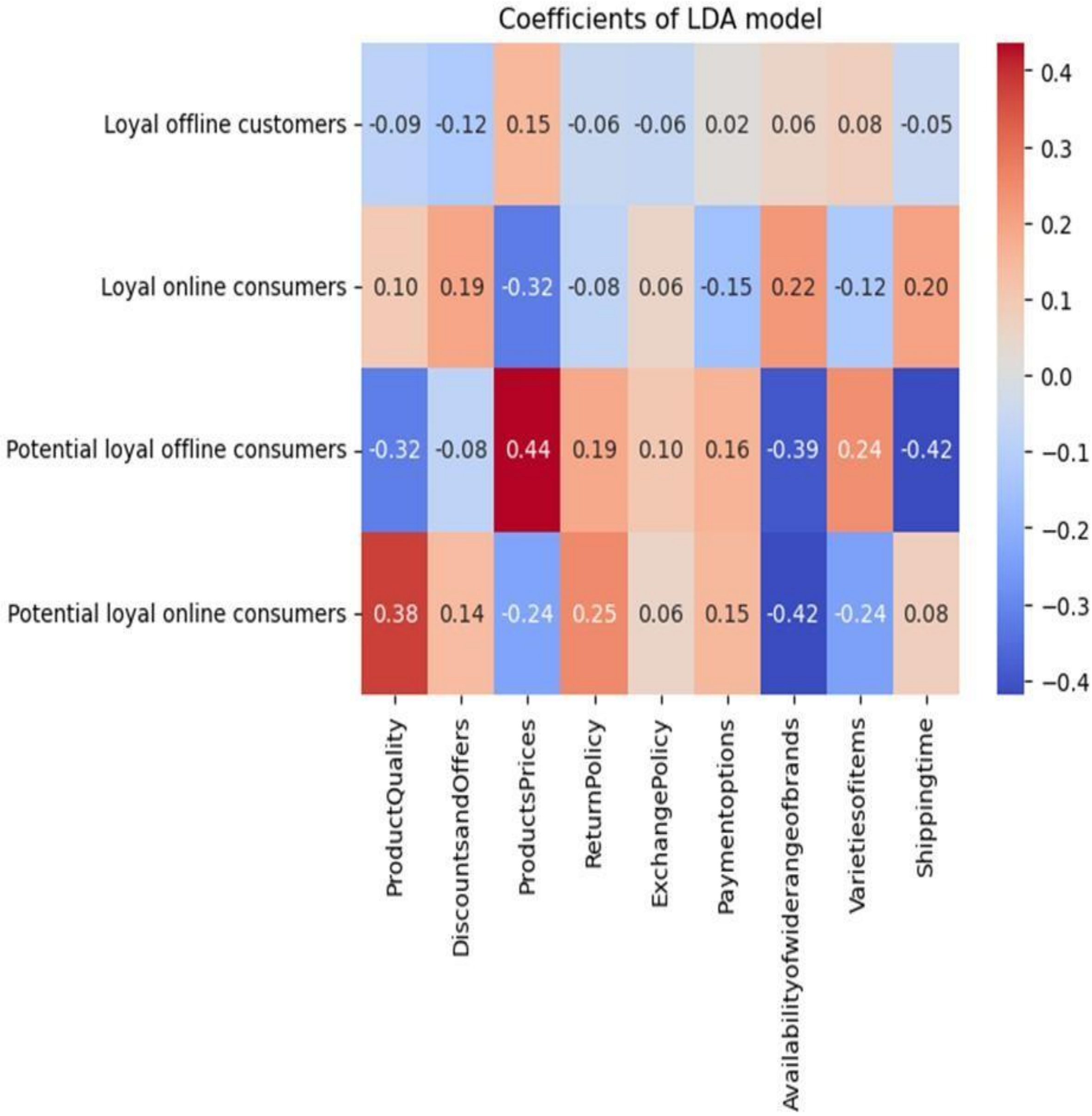
It's crucial to first identify the primary elements that affect customer satisfaction and factor influencing online shopping in order to understand the factors which impact customer loyalty and create marketing strategies that work. We may analyse customer satisfaction levels and Influencing factors across a number of variables utilising LDA, including satisfaction level across product quality, discounts and offers, pricing, return and exchange policies, payment methods, availability of more brands, a greater variety of products, and shipping times and influencing factors across a range of variables. Using LDA, we may determine which elements are most crucial for separating various customer categories based on their degrees of satisfaction and influencing factors. We can discover the relative importance of each factor in identifying the loyalty segment to which a customer belongs by examining at the LDA model's coefficients. With this data, targeted marketing strategies can then be created to meet the specific requirements and preferences of each consumer category. We can develop effective marketing strategies that can help to boost customer loyalty and drive revenue growth by using LDA to gain insights into the key variables that influence customer satisfaction and loyalty.

Using Python:

```
In [5]: import numpy as np
        import pandas as pd
        import seaborn as sns
        import matplotlib.pyplot as plt
        from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
        from sklearn.model_selection import train_test_split
        from sklearn.metrics import f1_score


        # Define the columns for X and y
        satisfaction_cols = ['ProductQuality', 'DiscountsandOffers', 'ProductsPrices', 'ReturnPolicy', 'ExchangePolicy',
                 'Paymentoptions', 'Availabilityofwiderangeofbrands', 'Varietiesofitems', 'Shippingtime']
        y_col = 'LoyaltySegment'

        # Split the data into X and y
        X = df[satisfaction_cols]
        y = df[y_col]

        # Split the data into training and testing sets
        X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

        # Fit the model on the training data
        lda = LinearDiscriminantAnalysis()
        lda.fit(X_train, y_train)

        # Evaluate the model on the test data
        y_pred = lda.predict(X_test)
        y_pred
```

```
# Get the coefficients of the LDA model
coefs = lda.coef_

# Create a dataframe with the coefficients
coef_df = pd.DataFrame(coefs, columns=satisfaction_cols, index=np.unique(y_train))

# Plot the heatmap
sns.heatmap(coef_df, cmap='coolwarm', annot=True, fmt='.2f')
plt.title('Coefficients of LDA model')
plt.show()
```

- **OUTPUT:**



Graph no.5

Findings:

The following interpretations can be drawn from the LDA model's coefficient values:

- The negative coefficient for loyal offline customers and the positive coefficient for loyal online consumers and potential loyal online consumers both point to the importance of product quality for all customer segments, regardless of whether they are loyal or potential loyal customers.
- The positive coefficients for loyal and potential online customers and the negative coefficient for potential offline customers indicate that discounts and offers are more significant to online customers than to offline customers (both loyal and potential).
- According to the positive coefficient, product prices matter more to loyal offline customers and potential loyal offline customers.
- When compared to potential loyal online and offline customers, loyal offline and online customers placed less importance on return and exchange policies.
- The negative coefficient suggests that payment choices are less significant to loyal online consumers.
- According to the positive coefficient for loyal online customers and the negative coefficient for potential loyal offline customers, availability of a wider variety of brands is more significant for loyal online customers than it is for potential loyal offline customers.
- Variety of products is more essential to potential loyal offline and offline customers, as the positive coefficient indicates.
- The positive coefficient shows that shipping time is more significant to loyal online consumers and potential loyal online consumers.

POSSIBLE MARKETING STRATEGIES:

- *<u>Loyal</u> <u>offline</u> <u>customers</u>*: You can concentrate on enhancing the quality of your products and highlighting this in your marketing messaging because this consumer category values product quality the highest. Also, you can provide customised customer service and loyalty programmes to promote customer retention.

- *Loyal online customers*: As discounts and offers are the most crucial elements for this customer group, you can provide them with special discounts and offers only for making purchases online. You can also reward them with reward points for coming back for more. Also, you may strengthen customer loyalty by personalising your online user experience.
- *Potentially loyal offline customers*: You can expand your product offerings and introduce new brands to get their attention since these customers value having access to a greater variety of brands and products. To encourage customers to try new products, you may also provide them with customised recommendations and promotions.
- *Potentially loyal online customers*: You can provide competitive pricing and promotions for online orders since product prices are the most significant element for this customer segment. To programs that offer loyal customers, you can also enhance your online user experience and make it more practical and customised.
- *To increase customer satisfaction and loyalty*, you can also concentrate on enhancing your return and exchange policies for all consumer categories.

Objective:

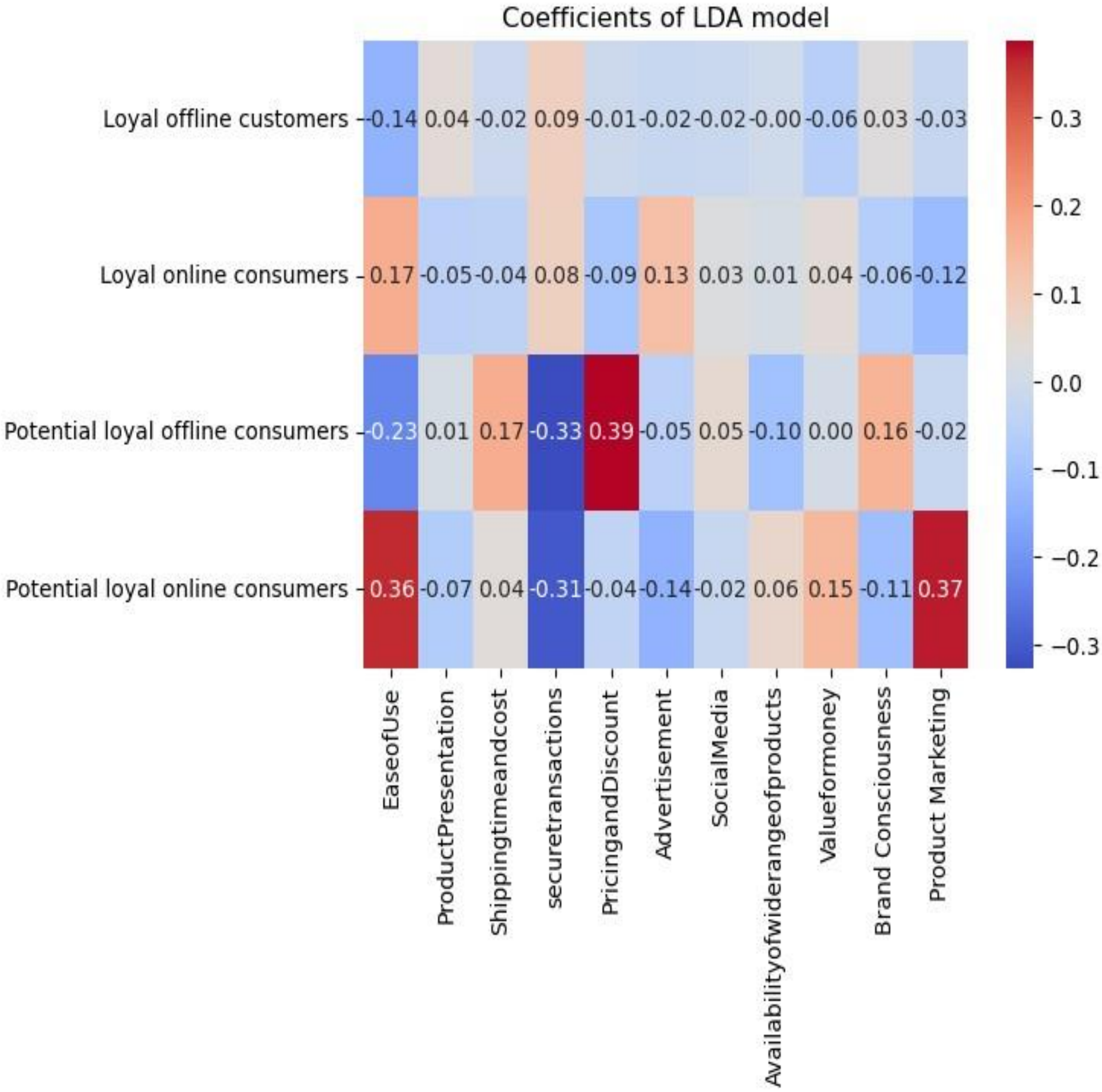
2. Discriminating customers into loyalty segment on the basis of factors influencing online shopping across number of variables.

Using Python:

```
In [6]: import numpy as np
        import pandas as pd
        import seaborn as sns
        import matplotlib.pyplot as plt
        from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
        from sklearn.model_selection import train_test_split
        from sklearn.metrics import f1_score


        # Define the columns for X and y
        FI_cols = ['EaseofUse',
               'ProductPresentation', 'Shippingtimeandcost', 'securetransactions',
               'PricingandDiscount', 'Advertisement', 'SocialMedia',
               'Availabilityofwiderangeofproducts', 'Valueformoney',
               'Brand Consciousness', 'Product Marketing']
        y_col = 'LoyaltySegment'

        # Split the data into X and y
        X = df[FI_cols]
        y = df[y_col]

        # Split the data into training and testing sets
        X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

        # Fit the model on the training data
        lda = LinearDiscriminantAnalysis()
        lda.fit(X_train, y_train)
```

```
In [7]: # Get the coefficients of the LDA model
        coefs = lda.coef_

        # Create a dataframe with the coefficients
        coef_df = pd.DataFrame(coefs, columns=FI_cols, index=np.unique(y_train))

        # Plot the heatmap
        sns.heatmap(coef_df, cmap='coolwarm', annot=True, fmt='.2f')
        plt.title('Coefficients of LDA model')
        plt.show()
```

- **OUTPUT:**



Graph no.6

Findings:

- Ease of Use: Loyal online consumers find it easier to use the product than loyal offline customers.

- Product Presentation: There is not a significant difference in the perception of the product presentation between loyal online and offline customers or between potential loyal online and offline customers.
- Shipping time and cost: Potential loyal online customers and potential loyal offline customers perceive that the shipping time and cost are important than loyal offline and online customers.
- Secure transactions: Loyal offline customers and loyal online customers perceive that secure transactions are more important.
- Pricing and Discount: Potential loyal offline customers high utmost importance to pricing and discounts
- Advertisement: Loyal online customers are more likely to respond to advertising than the other groups.
- Social Media: There is not a significant difference in the impact of social media between loyal online and offline customers or between potential loyal online and offline customers.
- Availability of wide range of products: Online consumers appreciate availability of wide range of products than offline consumers.
- Value for money: Value for money has positive and significant impact on online consumers.
- Brand Consciousness: Potential loyal offline customers are more brand conscious than loyal online customers, while potential loyal online customers are less brand conscious than loyal offline customers.
- Product Marketing: Potential loyal online customers are more influenced by product marketing than any other group.

Marketing strategies:

1. *Loyal offline customers:* Provide special discounts or offers to loyal offline customers to incentivize them to purchase more frequently. Create an online platform that allows

loyal offline customers to purchase products online, but also provides additional features such as in-store pick-up or personalized recommendations. Improve the product presentation in offline stores to make the shopping experience more engaging and enjoyable.

2. *Loyal online consumers*: Use targeted advertising to reach loyal online customers and encourage them to make repeat purchases. Enhance the online user experience by improving the ease of use of the website, making it easier for loyal online customers to find products and complete purchases. Use social media platforms to engage with loyal online customers and build a sense of community around the brand.
3. *Potential loyal offline consumers*: Offer promotions or discounts to incentivize potential loyal offline customers to try the brand's products. Use product demonstrations or experiential marketing to showcase the brand's products in stores and provide an engaging and informative experience for potential customers. Develop a loyalty program that rewards offline customers for repeat purchases and incentivizes them to make future purchases.
4. *Potential loyal online consumers*: Offer free shipping or discounted shipping to incentivize potential loyal online customers to make their first purchase. Use social media and influencer marketing to showcase the brand's products to potential customers and build awareness around the brand. Develop a referral program that incentivizes loyal online customers to refer their friends and family to the brand's online store.

PCA

Objective:

To factorize preferred online shopping platform into components and to analyse the variation of frequency and spending in each component on the basis of product preference.

What is PCA ?

PCA stands for Principal Component Analysis. It is a technique used for dimensionality reduction in machine learning and data analysis. PCA is a mathematical technique that transforms a large number of variables into a smaller number of variables, while retaining as much information as possible.

Assumptions:

- Linearity: PCA assumes that the relationship between the variables is linear. It means that the variables have a linear association with each other.
- Normality: PCA assumes that the variables are normally distributed. This assumption is important because PCA involves calculating covariance or correlation matrices, which are based on the assumption of normality.
- Independence: PCA assumes that the variables are independent of each other. This means that the variables do not share any common information or pattern.
- Homoscedasticity: PCA assumes that the variance of the variables is constant across all levels of the other variables. In other words, the spread of the data is the same for all values of the other variables.
- Large sample size: PCA assumes that the sample size is large enough to provide accurate estimates of the population parameters.

Mathematical Model :

Given a dataset X = {x1, x2, ..., xn} consisting of n observations, each of which is a ddimensional vector ($x1 \in R^d$, $x2 \in R^d$, ..., $xn \in R^d$), the aim of PCA is to find a set of $p \leq d$ orthogonal vectors u1, u2, ..., up such that the projection of the data onto these vectors results in a new set of uncorrelated variables z1, z2, ..., zp.

According to the definition of the first principle component, z1, it must have a magnitude of one and be a linear combination of the d original variables to maximise variance in the projected data: u1 is the first main component vector, and z1 = u1T x, where T stands for a matrix's transpose.

Under the condition that it is uncorrelated with the first principle component, the second principal component, or z2, is defined as the linear combination of the initial variables that maximises the variance of the projected data:

u2 is the second principal component vector, and u1T u2 = 0, hence z2 = u2T x.

Up until p principle components are obtained, this method is repeated till the necessary amount of dimensionality reduction is reached. Principal component vectors u1, u2,..., up that are produced serve as an orthogonal basis for the subspace covered by the initial data.

PCA can also be defined in terms of the data's covariance matrix's eigenvalue decomposition. The covariance matrix's eigenvectors are then the principal component vectors, and the accompanying eigenvalues are the proportion of variance that each principal component accounts for.

Why use PCA?

The objective of the analysis was to identify underlying patterns and relationships among the different online shopping platforms and their relationship with customer spending habits and product preferences. Since there were multiple variables and potential correlations among them, we used PCA as a tool to factorize the platform data into components and reduce the dimensionality of the dataset, making it easier to analyze and interpret. By using PCA, we were able to identify the most important factors driving variation in the data and to explore how customer spending habits and product preferences varied across these factors. This allowed us to gain insights into the key drivers of customer behavior and make data-driven decisions about how to optimize our online shopping platform to better meet customer needs and preferences.

Using python:

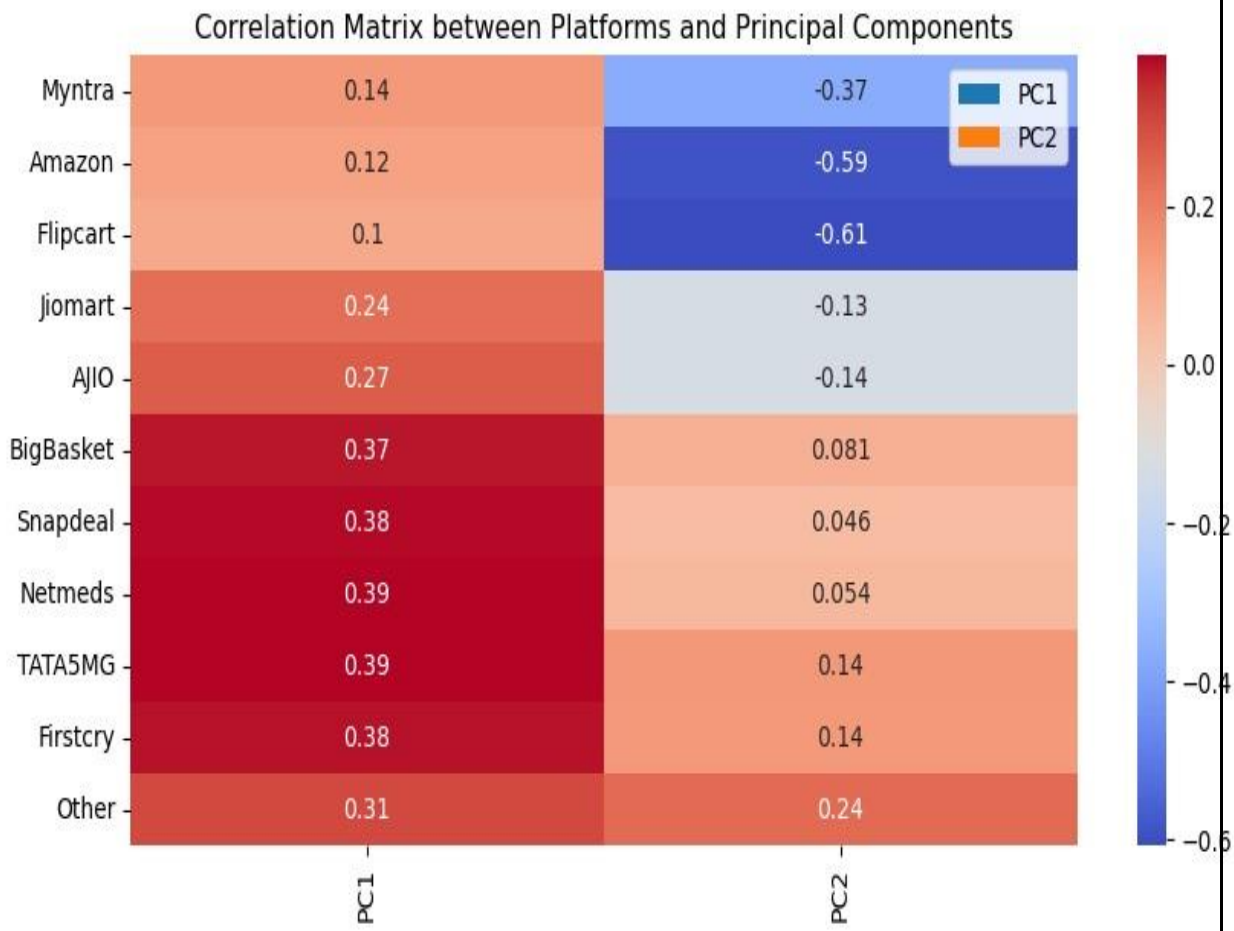
```python
[8]: import pandas as pd
import numpy as np
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
import seaborn as sns
# select the platform columns
platform_cols = ['Myntra', 'Amazon', 'Flipcart', 'Jiomart', 'AJIO', 'BigBasket', 'Snapdeal', 'Netmeds', 'TATA5MG', 'Firstcry', '(

# standardize the platform columns
X = df[platform_cols]
scaler = StandardScaler()
X_std = scaler.fit_transform(X)

# run PCA with all principal components
pca = PCA()
pca.fit(X_std)

# calculate the explained variance ratio for each principal component
var_ratio = pca.explained_variance_ratio_
```

```python
# run PCA with the selected number of principal components (e.g. 2)
n_components = 2
pca = PCA(n_components=n_components)
X_pca = pca.fit_transform(X_std)

# add the PCA scores to the original dataframe
df['PCA1'] = X_pca[:, 0]
df['PCA2'] = X_pca[:, 1]


# extract the loadings of the platforms on the first two principal components
loadings = pd.DataFrame(pca.components_.T, columns=['PC1', 'PC2'], index=platform_cols)

# plot the loadings using a bar plot
fig, ax = plt.subplots(figsize=(10, 5))
loadings.plot(kind='bar', ax=ax)
ax.set_ylabel('Loading')
ax.set_title('Loadings of the Platforms on the First 2 Principal Components')

# display the correlation matrix between the platforms and the principal components
corr_matrix = pd.DataFrame(pca.components_.T, columns=['PC1', 'PC2'], index=platform_cols)
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Matrix between Platforms and Principal Components')
plt.show()
```

```python
In [13]: # select the product and spending columns from the dataframe
product_cols = ['Fashion', 'Electronics', 'Footwear', 'Baby Products', 'Home Décor and Furnishing', 'Food', 'Jewellry', 'BeautyPr
Frequentlyshopping_col = 'FreqentlyShopping'
product_df = df[product_cols + [Frequentlyshopping_col]]

# calculate mean spending for each product in PC1 and PC2
n_products = len(product_cols)
pc1_means = []
pc2_means = []

for i in range(n_products):
    pc1_means.append((product_df.loc[df['PCA1']>0, product_cols[i]] * product_df.loc[df['PCA1']>0, Frequentlyshopping_col]).mean(
    pc2_means.append((product_df.loc[df['PCA2']>0, product_cols[i]] * product_df.loc[df['PCA2']>0, Frequentlyshopping_col]).mean(


# plot the results
fig, axs = plt.subplots(nrows=3, ncols=5, figsize=(15, 10))
for i, ax in enumerate(axs.flat):
    if i < n_products:
        ax.bar(['PCA1', 'PCA2'], [pc1_means[i], pc2_means[i]])
        ax.set_title(product_cols[i])
        ax.set_ylim([0, 10])
plt.tight_layout()
plt.show()
```

```python
In [14]: # select the product and spending columns from the dataframe
product_cols = ['Fashion', 'Electronics', 'Footwear', 'Baby Products', 'Home Décor and Furnishing', 'Food', 'Jewellry', 'BeautyPr
spendings_col = 'MonthlySpending'
product_df = df[product_cols + [spendings_col]]

# calculate mean spending for each product in PC1 and PC2
n_products = len(product_cols)
pc1_means = []
pc2_means = []

for i in range(n_products):
    pc1_means.append((product_df.loc[df['PCA1']>0, product_cols[i]] * product_df.loc[df['PCA1']>0, spendings_col]).mean())
    pc2_means.append((product_df.loc[df['PCA2']>0, product_cols[i]] * product_df.loc[df['PCA2']>0, spendings_col]).mean())

# plot the results
fig, axs = plt.subplots(nrows=3, ncols=5, figsize=(15, 10))
for i, ax in enumerate(axs.flat):
    if i < n_products:
        ax.bar(['PCA1', 'PCA2'], [pc1_means[i], pc2_means[i]])
        ax.set_title(product_cols[i])
        ax.set_ylim([0, 10])
plt.tight_layout()
plt.show()
```

- **Findings:**



Graph no.7

Component	Frequency	Spending
PCA1	Low frequency on the Electronics, footwear, Toys, medical supplement	High spending on Electronics, footwear, Toys, medical supplement
PCA2	High frequency on fashion, beauty product and jewellery.	High spendings on fashion, beauty product and jewellery.

Table no.48

SEM:

WHAT IS SEM?

The multivariate technique known as structural equation modelling (SEM) is incredibly versatile and effective. To represent complex and dynamic relationships within a web of observed and unobserved variables, it includes a conceptual model, path diagram, and system of linked regression-style equations. SEM and regression are fundamentally distinct even though they appear to be identical. There is a distinct difference between dependent and independent variables in a regression model. However, since a dependent variable in one model equation might change into an independent variable in other components of the SEM system, such principles are only applicable relative to SEM. The ability of a variable to play a reciprocal role is precisely what allows SEM to infer causal relationships.

Both endogenous and exogenous variables are included in SEM models. In at least one of the SEM equations, endogenous variables serve as a dependent variable; because they may become independent variables in other SEM equations, they are referred to as endogenous variables rather than response variables. Exogenous factors are dependably free factors in the SEM conditions. SEM conditions model both the causal connections among endogenous and exogenous factors, and the causal connections among endogenous factors.

Path diagrams are the most effective way to show SEM models. The variables are represented by nodes in a path diagram, and arrows show how these variables relate to one another. Latent variables are typically depicted by a circle or ellipse in a path diagram, while observed variables are typically depicted by a rectangle or square. Most of the time, arrows are used to show how the variables are related to one another. A solitary straight bolt shows a causal connection from the foundation of the bolt to the top of the bolt. A reciprocal causal relationship between two variables is indicated by the presence of two straight, single-headed arrows pointing in opposite directions. A bent two-headed bolt shows there might be some relationship between the two factors. Blunder expressions for a variable are embedded into the way chart by drawing a bolt from the worth of the mistake term to the variable with which the term is related.

For example, error terms are typically not connected in path diagrams for cross-sectional data, showing stochastic independence across the error terms. However, the error terms should be connected by curved, two-headed arrows if we detect the correlation between them, which is likely to happen in most research conducted.

Common terms in SEM:

- ***Observed and latent variables*** :

  Observed (manifest) variables and latent variables are two types of theoretical constructs that are frequently studied by researchers, particularly in the behavioural and social sciences. Variables that can be directly observed (such as income, blood sugar, etc.) are observed variables. In any case, explores regularly need to manage idle factors that can't be straightforwardly estimated, for example, character, discernment, purchasing conduct and so on. The latent variables are measured by research using observed variables. Self-report responses to attitudinal scales, coded responses to interview questions, and other similar items may be part of the observation. The latent variables are measured with these measured scores, also known as observed or manifest variables.

- ***Exogenous and endogenous latent variables:***

Endogenous latent variables and exogenous latent variables are similar terms for independent and dependent variables, respectively. Exogenous variables either directly or indirectly affect endogenous variables.

- ***Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA):***

The purpose of factor analysis is to examine the connections between sets of observed and latent variables. Exploratory factor analysis is used when the connections between the observed and latent variables are not known or uncertain. To find out how and to what extent the observed variables are linked to their underline factors, exploratory factor analysis is used. When the researcher has some understanding of the latent variable structure through theory, empirical research, or both, confirmatory factor analysis is appropriate.

- ***The path diagram***:

A path diagram is a visual representation of the relationships that are assumed to exist between the variables in the study. The path diagrams essentially employ four geometric symbols: Single-headed arrows represent the effect of one variable on another variable, double-headed arrows represent covariance or correlation between two variables, and circles or ellipses represent unobserved latent variables. Squares or rectangles represent observed variables.

Two sub models SEM:

Structural and measurement models. The measurement model depicts the connection between latent and observed variables. To put it another way, it represents the CFA model's pattern for how each measure loads on a particular factor. However, the structural model demonstrates the connection between latent variables.

Measurement model:

In a measurement model, the specification includes the number of latent variables, the number of indicators for each latent variable, and the connections between the latent variables and their indicators. A measurement model's objective is to use latent variables to explain the indicators' observed variance and covariance.

CFA commonly utilizes the factor model, which posits that latent variables (factors) are responsible for the observed indicators. The factor model assumes that only one latent variable affects each observed indicator.

The goodness-of-fit of a CFA measurement model to the observed data is evaluated using a variety of fit indices. The chi-square test, the CFI, the Tucker-Lewis Index, and the root mean square error of approximation (RMSEA) are examples of these.

By employing a well-fitting measurement model that assures the dependability and validity of the observed indicators, one can explore the connections between the latent variables and other variables of concern.

Structural model:

The use of CFA's structural model enables the testing of theoretical hypotheses on causal relations between concealed variables. One application is analyzing how intelligence, achievement and motivation interrelate through a structural model.

Within a structural framework, connections between anonymous variables are illustrated through lines or arrows linking these obscure identities. The power attributed to a specific link can range from -1 to 1 and be either positive or negative.

One typically estimates the structural model using maximum likelihood or another suitable estimation technique and evaluates the model's goodness-of-fit through diverse fit indices, including the chi-square test statistic, comparative fit index (CFI), Tucker-Lewis Index (TLI), and root mean square error of approximation (RMSEA).

The CFA structural model is a useful instrument for investigating the connections between latent variables and evaluating theoretical assumptions regarding the fundamental causal mechanisms that result in observable behavioural patterns.

Validity and reliability:

By looking at the validity and reliability scores on an instrument that was used in a specific context, validity, and reliability were checked. Validity and reliability are guaranteed when scores are at an acceptable level. The measurement error of variables is ignored by conventional statistical analysis. However, it has been discovered that measurement error has significant effects. Since the development of software for structural equation modeling, variables' measurement errors can now be taken into account.

Construct validity:

Construct validity can be tried utilizing factor investigation. The goal of factor analysis is to identify which sets of observed variables (latent variables) share common variance-covariance characteristics that define theoretical constructs or factors. The shared variance-covariance between the observed variables is assumed to be caused by some factors, which are smaller in number than the observed variables. Factor analysis is used to either confirm that a set of variables defines those constructs or factors or to investigate which variables relate to factors after one collect data on observed variables.

Confirmatory factor analysis assumes that you have a solid understanding of the number of factors you will encounter and the variables that will most likely load onto each factor when you start the analysis. Typically, the published results of a factor analysis serve as the basis for your expectations. Take the previously validated fatigue scale as an illustration. You want to make sure that the factors are influenced by the variables in your sample in the same way as they were in the original research. To put it another way, you know exactly what you want to find in your sample. This indicates that you are aware of the total number of factors that will be encountered and the variables that will be added to those factors. In contrast to exploratory factor analysis, the criteria for variable inclusion in confirmatory factor analysis are significantly more stringent. Variables with factor loadings below 0.7 are generally eliminated. The purpose of an exploratory factor analysis, which does not have a predetermined number of factors, is to investigate the connections between the variables. Although you may have a general idea of what you anticipate finding, you have not yet chosen a specific hypothesis. Or on the other hand, you might have planned an examination question because of your hypothetical comprehension, and are presently trying it. Obviously, in an exploratory variable examination, the last number of not entirely settled by your information and your translation of the elements. Exploratory factor analyses may have much lower factor loading cut offs. When making scales, you can test a new scale with an exploratory factor analysis before moving on to a confirmatory factor analysis to verify the factor structure in a new sample. The researcher uses exploratory factor model approaches to try to find a model that works with the data. As a result, in practice, researchers develop various alternative models in the hope of discovering a model that is compatible with the data and supported by theory. The primary justification for exploratory factor analysis (EFA) is this. Researchers statistically test the significance of a hypothesized factor model, or whether the sample data support the model, in confirmatory factor model approaches. The hypothesized model's validity is further demonstrated by additional data samples that fit the model. Confirmatory factor analysis (CFA) is conducted primarily for this reason.

TWO TYPES OF CONSTRUCT VALIDITY:

1. **Discriminant validity**:

A subtype of construct validity is discriminant validity. All in all, it shows you how well a test estimates the idea it was intended to quantify.

Discriminant legitimacy explicitly gauges whether develops that hypothetically ought not to be connected are, as a matter of fact, irrelevant.

Discriminant validity is crucial because it reveals whether your test accurately assesses the construct of interest or separate, unrelated constructs. The precision of your operationalization—that is, your capacity to translate abstract ideas into observable variables or data.

You can verify the discriminant validity of your test by proving that there is either no connection or a very weak correlation between measures of unrelated constructs.

A correlation coefficient, like Pearson's r, is used to quantify the degree of correlation and shows the direction and strength of the association between variables. The correlation coefficient's value is always between one and one thousand.

Correlation coefficient values can be understood as follows:

r = 1: there is perfect positive correlation r

= 0: there is no correlation at all r = −1:

there is perfect negative correlation

2. Convergent validity:

A subtype of construct validity is convergent validity. The degree to which a test accurately measures the concept it was designed to measure is known as construct validity.

In that it measures whether constructs that theoretically ought to be related to one another are, in fact, related to one another, convergent validity is a little bit more nuanced.

You must show a positive correlation between measures of related constructs in order to evaluate the convergent validity of your test. In other words, if you have two scales that are connected, folks who perform well on one should perform well on the other.

A correlation coefficient, like Pearson's r, which is a value ranging from 1 to 1, is used to estimate correlation. You can see the strength and direction of the association between the variables due to this coefficient. r = 1: there is perfect positive correlation r = 0: there is no correlation at all r = −1: there is perfect negative correlation

Convergent and discriminant validity work together to support concept validity. In studies, they are both assessed because doing so is necessary to show construct validity. Although neither is sufficient on its own, it's crucial to keep in mind that they are not the same thing.

CFA MODELS:

Stated below the factor model of Factor Influenced (FI) on online shopping and Satisfaction Level (SL) of consumers on online shopping.

We have categorised the Factor Influencing online shopping (FI) in 3 different categories i.e

- ***Finance***
    - Secure Transaction (ST)
    - Pricing And Discount (PAD)
    - Value For Money (VFM)
    - Shipping Time And Cost (STAC)

- ***Marketing***
    - Advertisement (Adv)
    - Social Media (SM)
    - Product Marketing (PM)
    - Product Presentation (PP)

- ***Personal choice***
    - Availiability Of Wide Range of Products (AOWRP)
    - Ease Of Use (EOU)
    - Brand Conciousness (BCon)

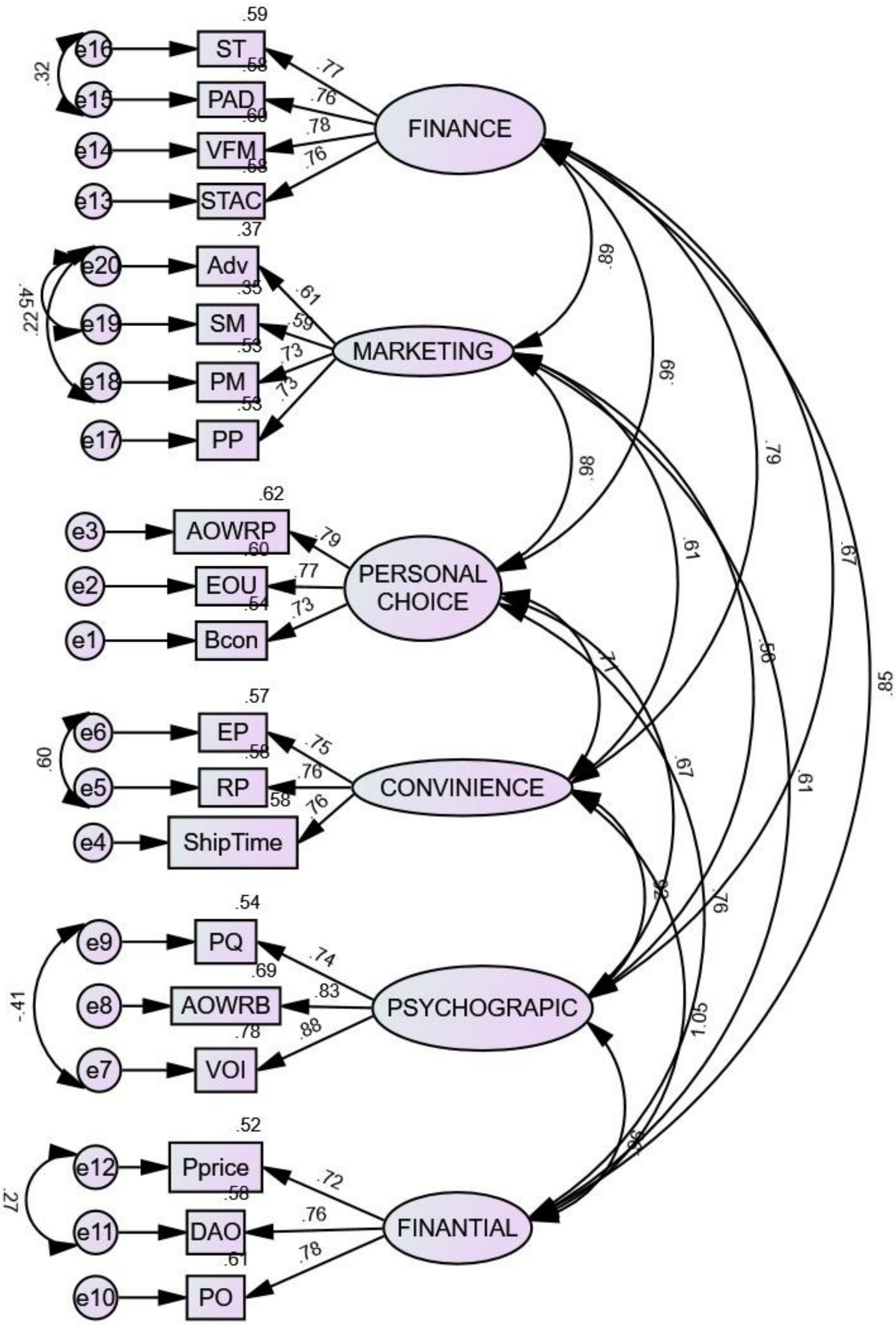
We have categorised the Satisfaction Level of customers on online shopping (SL) in 3 different categories i.e

- ***Convinience***
    - Exchange Policy (EP)
    - Return Policy (RP)
    - Shipping Time (ShipTime)

- ***Psychographic***
    - Product Quality (PQ)
    - Availiability Of Wide Range of Brands (AOWRB)
    - Variety Of Items (VOI)

- ***Finantial***
    - Product Prising (PPrice)
    - Discount And Offers (DAO)
    - Payment Options (PO)

FACTORS INFLUENCING(FI) AND SATISFACTION LEVEL(SL):



MODEL FIT:

Model	Full Form	Calculated Value	Expected Value For Good Fit

CMIN/DF	Discrepancy divided by degree of freedom	2.313	[1-3] or[1-5]or more than 5 weakly fitted
CFI	comparative fit index	0.948	A:0=poor fit close to 1= very good fit B:CFI>.95 : good fit, CFI>0.60 : moderate fit
PCFI	Parsimony comparative fix index	0.744	0= poor fit 1= Exact fit
RMSEA	Root mean square error of approximation	0.069	=0: good fit <0.08: close fit >0.08 : weak fit
GFI	Goodness of fit index	0.889	GFI>.90: good fit GFI>0.60: moderate fit
RMR	Root mean square residual	0.069	RMR=0: good fit RMR<0.08: close fit

Table no.49

Regression Weights: (Group number 1 - Default model)								
			Estimate	S.E.	C.R.	P	Label
Bcon	<---	PERSONAL_CHOICE	1				
EOU	<---	PERSONAL_CHOICE	1.103	0.085	13.029	***	par_1
AOWRP	<---	PERSONAL_CHOICE	1.045	0.079	13.261	***	par_2
ShipTime	<---	CONVINIENCE	1				
RP	<---	CONVINIENCE	1.132	0.086	13.177	***	par_3
EP	<---	CONVINIENCE	1.137	0.087	13.082	***	par_4
VOI	<---	PSYCHOGRAPIC	1				
AOWRB	<---	PSYCHOGRAPIC	0.97	0.056	17.316	***	par_5
PQ	<---	PSYCHOGRAPIC	0.774	0.063	12.366	***	par_6
PO	<---	FINANTIAL	1				
DAO	<---	FINANTIAL	1.006	0.071	14.199	***	par_7
Pprice	<---	FINANTIAL	0.882	0.067	13.195	***	par_8
STAC	<---	FINANCE	1				

VFM	<---	FINANCE	1.062	0.078	13.601	***	par_9
PAD	<---	FINANCE	1.058	0.08	13.235	***	par_10
ST	<---	FINANCE	1.027	0.076	13.431	***	par_11
PP	<---	MARKETING	1				
PM	<---	MARKETING	1.03	0.09	11.438	***	par_12
SM	<---	MARKETING	0.801	0.087	9.258	***	par_13
Adv	<---	MARKETING	0.814	0.086	9.512	***	par_14

Table no.50

Standardized Regression Weights: (Group number 1 - Default model)

			Estimate
Bcon	<---	PERSONAL_CHOICE	.735
EOU	<---	PERSONAL_CHOICE	.775
AOWRP	<---	PERSONAL_CHOICE	.788
ShipTime	<---	CONVINIENCE	.761
RP	<---	CONVINIENCE	.759
EP	<---	CONVINIENCE	.755
VOI	<---	PSYCHOGRAPIC	.884
AOWRB	<---	PSYCHOGRAPIC	.833
PQ	<---	PSYCHOGRAPIC	.735
PO	<---	FINANTIAL	.781
DAO	<---	FINANTIAL	.763
Pprice	<---	FINANTIAL	.720
STAC	<---	FINANCE	.764
VFM	<---	FINANCE	.778
PAD	<---	FINANCE	.761
ST	<---	FINANCE	.770
PP	<---	MARKETING	.730
PM	<---	MARKETING	.730
SM	<---	MARKETING	.591
Adv	<---	MARKETING	.611

Table no.51

Composite reliability:

	CR	MSV	Convergent validity: AVE
PSYCHOGRAPIC	0.859	0.912	0.672
FINANCE	0.852	0.988	0.590
MARKETING	0.762	0.956	0.447
CONVINIENCE	0.802	1.100	0.575
FINANTIAL	0.799	1.100	0.570
PERSONAL_CHOICE	0.810	0.988	0.587

Table no.52

Discriminant validity:

	PSYCHO GRAPIC	FINANCE	MARKETING	CONVINIENCE	FINANTIAL	PERSONAL _CHOICE
PSYCHOGRAPIC	**0.820**					
FINANCE	0.668	**0.768**				
MARKETING	0.560	0.886	**0.669**			
CONVINIENCE	0.917	0.787	0.613	**0.758**		
FINANTIAL	0.955	0.850	0.606	1.049	**0.755**	
PERSONAL_CHOICE	0.665	0.994	0.978	0.711	0.758	**0.766**

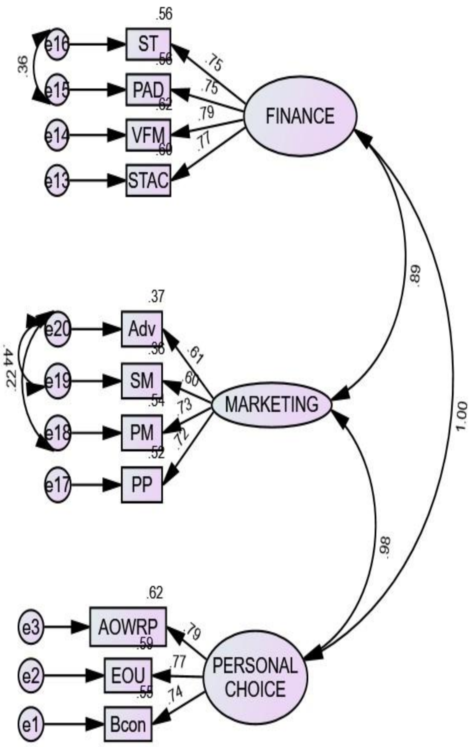
Table no.53

Discriminant Validity:
The square root of the AVE for PSYCHOGRAPIC is less than the absolute value of the correlations with another factor.
The square root of the AVE for FINANCE is less than the absolute value of the correlations with another factor.
The square root of the AVE for MARKETING is less than the absolute value of the correlations with another factor.
The square root of the AVE for CONVINIENCE is less than the absolute value of the correlations with another factor.
The square root of the AVE for FINANTIAL is less than the absolute value of the correlations with another factor.
The square root of the AVE for PERSONAL_CHOICE is less than the absolute value of the correlations with another factor.
The AVE for PSYCHOGRAPIC is less than the MSV.
The AVE for FINANCE is less than the MSV.
The AVE for MARKETING is less than the MSV.
The AVE for CONVINIENCE is less than the MSV.
The AVE for FINANTIAL is less than the MSV.
The AVE for PERSONAL_CHOICE is less than the MSV.

Convergent Validity:

The AVE for MARKETING is less than 0.50. **FACTORS**

INFLUENCING (FI) MODEL:



MODEL FIT:

Model	Full Form	Calculated Value	Expected Value For Good Fit
CMIN/DF	Discrepancy divided by degree of freedom	3.408	[1-3] or[1-5]or more than 5 weakly fitted
CFI	comparative fit index	0.949	A:0=poor fit close to 1= very good fit B:CFI>.95 : good fit, CFI>0.60 : moderate fit
PCFI	Parsimony comparative fix index	0.656	0= poor fit 1= Exact fit
RMSEA	Root mean square error of approximation	0.093	=0 : good fit <0.08 : close fit >0.08 : weak fit
GFI	Goodness of fit index	0.913	GFI>.90: good fit GFI>0.60: moderate fit
RMR	Root mean square residual	0.077	RMR=0: good fit RMR<0.08: close fit

Table no.54

Composite reliability:

	CR	MSV	Convergent validity: AVE
PERSONAL_CHOICE	0.810	0.994	0.587
FINANCE	0.849	0.994	0.584
MARKETING	0.763	0.956	0.449

Table no.55

Discriminant validity:

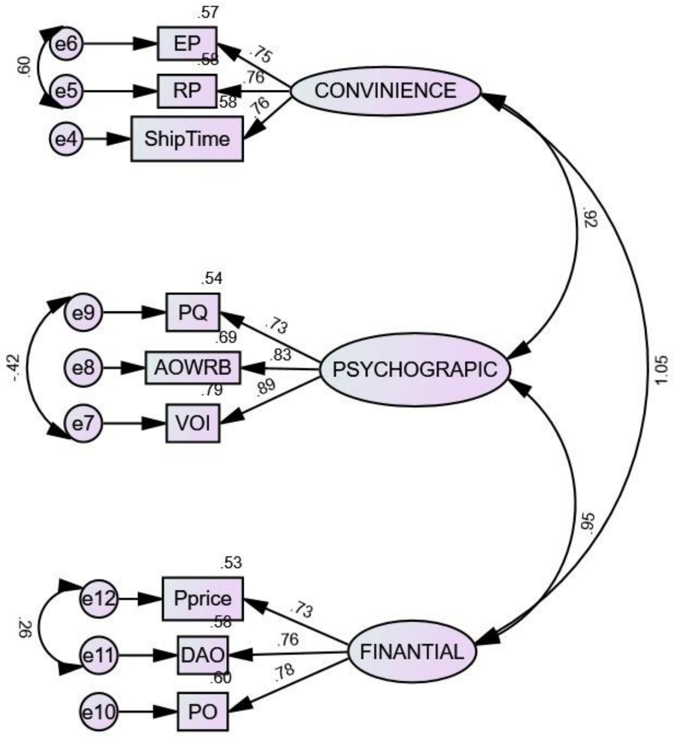
	PERSONAL_CHOICE	FINANCE	MARKETING
PERSONAL_CHOICE	**0.766**		
FINANCE	0.997	**0.764**	
MARKETING	0.978	0.887	**0.670**

Table no.56

Discriminant Validity:
The square root of the AVE for PERSONAL_CHOICE is less than the absolute value of the correlations with another factor.
The square root of the AVE for FINANCE is less than the absolute value of the correlations with another factor.
The square root of the AVE for MARKETING is less than the absolute value of the correlations with another factor.
The AVE for PERSONAL_CHOICE is less than the MSV.
The AVE for FINANCE is less than the MSV.
The AVE for MARKETING is less than the MSV.

Convergent Validity:
The AVE for MARKETING is less than 0.50.

SATISFACTION LEVEL(SL) MODEL:



MODEL FIT:

Model	Full Form	Calculated Value	Expected Value For Good Fit
CMIN/DF	Discrepancy divided by degree of freedom	2.941	[1-3] or[1-5]or more than 5 weakly fitted
CFI	comparative fit index	0.977	A:0=poor fit close to 1= very good fit B:CFI>.95 : good fit, CFI>0.60 : moderate fit
PCFI	Parsimony comparative fix index	0.570	0= poor fit 1= Exact fit
RMSEA	Root mean square error of approximation	0.084	=0 : good fit <0.08 : close fit >0.08 : weak fit
GFI	Goodness of fit index	0.951	GFI>.90: good fit GFI>0.60: moderate fit
RMR	Root mean square residual	0.044	RMR=0: good fit RMR<0.08: close fit

Table no.57

Composite reliability:

	CR	MSV	Convergent validity: AVE
PSYCHOGRAPIC	0.860	0.906	0.673
CONVINIENCE	0.802	1.103	0.575
FINANTIAL	0.800	1.103	0.572

Table no.58

Discriminant validity:

	PSYCHOGRAPIC	**CONVINIENCE**	**FINANTIAL**
PSYCHOGRAPIC	**0.821**		
CONVINIENCE	0.915	**0.758**	
FINANTIAL	0.952	1.050	**0.756**

Table no.59

Discriminant Validity:

The square root of the AVE for PSYCHOGRAPIC is less than the absolute value of the correlations with another factor.

The square root of the AVE for CONVINIENCE is less than the absolute value of the correlations with another factor.

The square root of the AVE for FINANTIAL is less than the absolute value of the correlations with another factor.

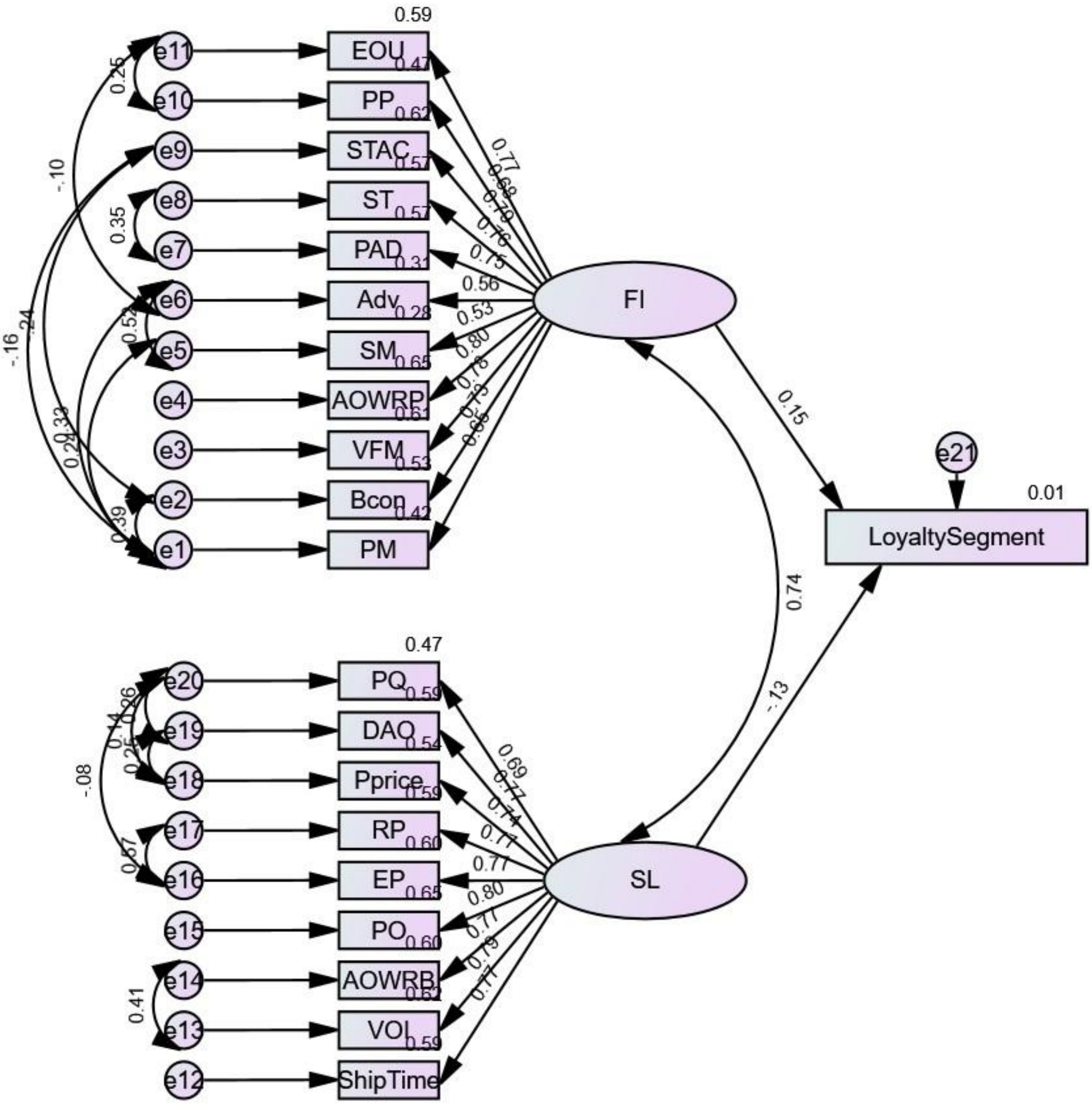
The AVE for PSYCHOGRAPIC is less than the MSV.

The AVE for CONVINIENCE is less than the MSV.

The AVE for FINANTIAL is less than the MSV.

SEM MODELS:

FI-SL-LOYALTY:



MODEL FIT:

Model	Full Form	Calculated Value	Expected Value For Good Fit
CMIN/DF	Discrepancy divided by degree of freedom	1.606	[1-3] or[1-5]or more than 5 weakly fitted
CFI	comparative fit index	0.973	A:0=poor fit close to 1= very good fit B:CFI>.95 : good fit, CFI>0.60 : moderate fit
PCFI	Parsimony comparative fix index	0.797	0= poor fit 1= Exact fit
RMSEA	Root mean square error of approximation	0.047	=0 : good fit <0.08 : close fit >0.08 : weak fit
GFI	Goodness of fit index	0.916	GFI>.90: good fit GFI>0.60: moderate fit
RMR	Root mean square residual	0.065	RMR=0: good fit RMR<0.08: close fit

Table no.60

Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
PM	<---	FI	1.000				
Bcon	<---	FI	1.152	.087	13.304	***	par_1
VFM	<---	FI	1.218	.109	11.158	***	par_2
AOWRP	<---	FI	1.242	.109	11.389	***	par_3
SM	<---	FI	.784	.087	9.039	***	par_4
Adv	<---	FI	.823	.081	10.121	***	par_5
PAD	<---	FI	1.192	.110	10.800	***	par_6
ST	<---	FI	1.148	.106	10.836	***	par_7
STAC	<---	FI	1.173	.112	10.470	***	par_8
PP	<---	FI	1.033	.104	9.980	***	par_9
EOU	<---	FI	1.278	.116	11.001	***	par_10
			Estimate	**S.E.**	**C.R.**	**P**	**Label**
VOI	<---	SL	1.072	.079	13.590	***	par_11

AOWRB	<---	SL	1.084	.081	13.315	***	par_12
PO	<---	SL	1.060	.076	14.024	***	par_13
EP	<---	SL	1.155	.086	13.354	***	par_14
RP	<---	SL	1.137	.086	13.263	***	par_15
Pprice	<---	SL	.931	.074	12.603	***	par_16
DAO	<---	SL	1.037	.079	13.152	***	par_17
PQ	<---	SL	.871	.075	11.553	***	par_18
ShipTime	<---	SL	1.000				
LoyaltySegment	<---	FI	.225	.158	1.422	.155	par_35
LoyaltySegment	<---	SL	-.199	.154	-1.295	.195	par_36

Table no.61

Standardized Regression Weights: (Group number 1 - Default model)

			Estimate
PM	<---	FI	.651
Bcon	<---	FI	.727
VFM	<---	FI	.783
AOWRP	<---	FI	.804
SM	<---	FI	.525
Adv	<---	FI	.557
PAD	<---	FI	.752
ST	<---	FI	.756
STAC	<---	FI	.787
PP	<---	FI	.684
EOU	<---	FI	.770
VOI	<---	SL	.785
			Estimate

AOWRB	<---	SL	.772
PO	<---	SL	.805
EP	<---	SL	.774
RP	<---	SL	.769
Pprice	<---	SL	.738
DAO	<---	SL	.765
PQ	<---	SL	.686
ShipTime	<---	SL	.767
LoyaltySegment	<---	FI	.146
LoyaltySegment	<---	SL	-.133

Table no.62

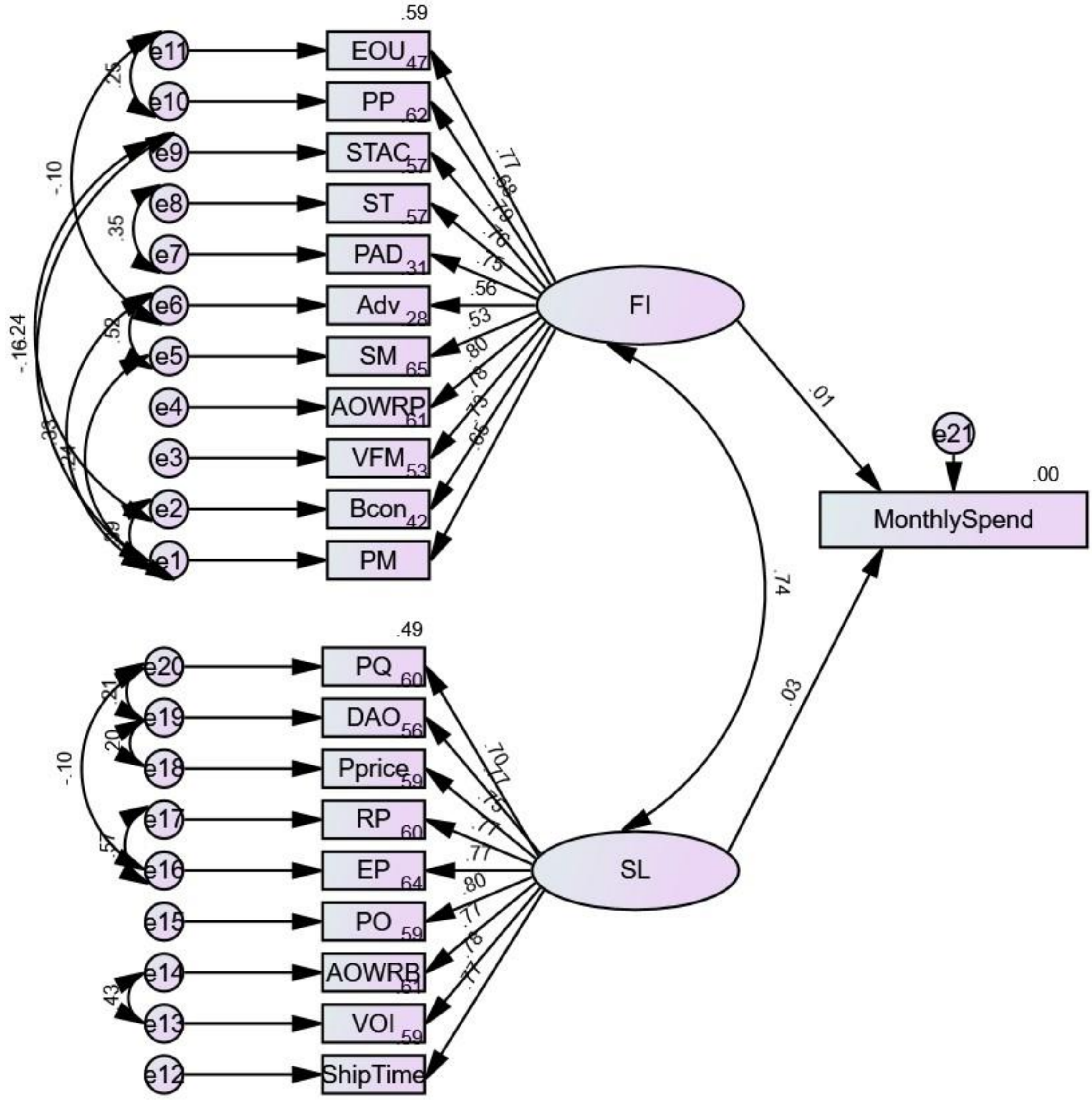
Covariances: (Group number 1 - Default model)

	M.I. Par Change

e17	<--> e18	8.647	.091
e13	<--> e15	6.711	.081
e11	<--> e12	4.632	-.092
e10	<--> e17	5.131	-.087
e9	<--> e13	10.622	-.115
e8	<--> e15	6.319	.092
e8	<--> e9	4.081	.083
e7	<--> e19	5.592	.090
e7	<--> e17	10.817	.117
e7	<--> e16	7.991	-.101
e6	<--> e12	4.414	.085
e4	<--> e16	4.647	.076
e2	<--> e19	4.975	-.087
e2	<--> e13	4.714	-.080
e1	<--> e17	7.701	-.098
e1	<--> e16	11.725	.122

Table no.63

FI-SL-MONTHLY SPEND:



MODEL FIT:

Model	Full Form	Calculated Value	Expected Value For Good Fit
CMIN/DF	Discrepancy divided by degree of freedom	1.664	[1-3] or[1-5]or more than 5 weakly fitted
CFI	comparative fit index	0.970	A:0=poor fit close to 1= very good fit B:CFI>.95 : good fit, CFI>0.60 : moderate fit
PCFI	Parsimony comparative fix index	0.799	0= poor fit 1= Exact fit
RMSEA	Root mean square error of approximation	0.049	=0 : good fit <0.08 : close fit >0.08 : weak fit
GFI	Goodness of fit index	0.913	GFI>.90: good fit GFI>0.60: moderate fit
RMR	Root mean square residual	0.065	RMR=0: good fit RMR<0.08: close fit

Table no.64

Regression Weights: (Group number 1 - Default model)							
			Estimate	S.E.	C.R.	P	Label
PM	<---	FI	1				
Bcon	<---	FI	1.153	0.087	13.315	***	par_1
VFM	<---	FI	1.217	0.109	11.159	***	par_2
AOWRP	<---	FI	1.24	0.109	11.39	***	par_3
SM	<---	FI	0.784	0.087	9.04	***	par_4
Adv	<---	FI	0.823	0.081	10.128	***	par_5
PAD	<---	FI	1.191	0.11	10.802	***	par_6
ST	<---	FI	1.147	0.106	10.843	***	par_7
STAC	<---	FI	1.173	0.112	10.472	***	par_8

PP	<---	FI	1.033	0.103	9.987	***	par_9
EOU	<---	FI	1.277	0.116	10.999	***	par_10
VOI	<---	SL	1.064	0.079	13.535	***	par_11
AOWRB	<---	SL	1.078	0.081	13.295	***	par_12
PO	<---	SL	1.052	0.075	13.974	***	par_13
EP	<---	SL	1.153	0.086	13.382	***	par_14
RP	<---	SL	1.136	0.085	13.313	***	par_15
Pprice	<---	SL	0.945	0.073	12.929	***	par_16
DAO	<---	SL	1.046	0.078	13.369	***	par_17
PQ	<---	SL	0.89	0.075	11.929	***	par_18
ShipTime	<---	SL	1				
MonthlySpend	<---	FI	0.006	0.105	0.054	0.957	par_19
MonthlySpend	<---	SL	0.032	0.102	0.31	0.756	par_20

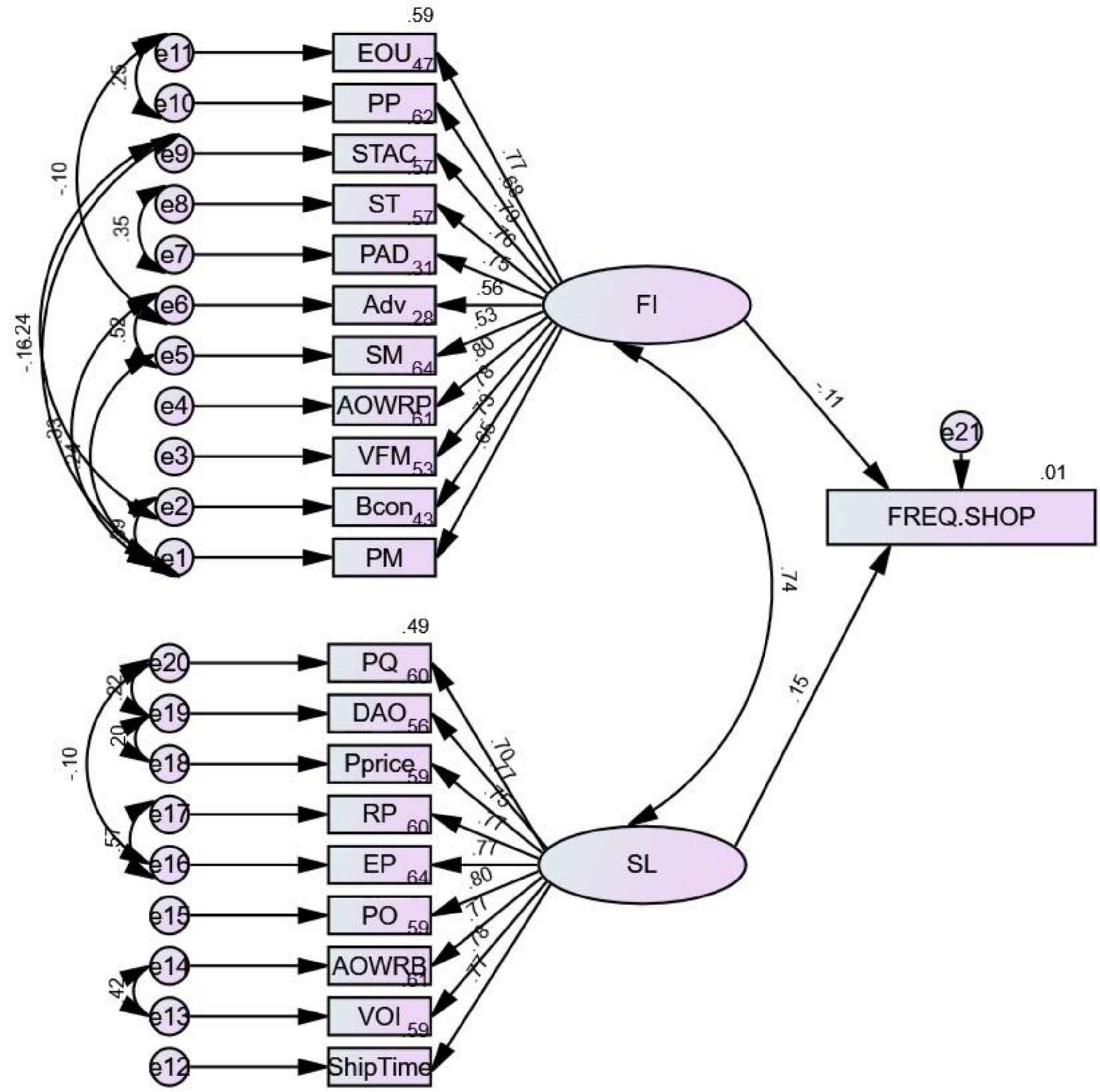
Table no.65

Standardized Regression Weights: (Group number 1 - Default model)			
			Estimate
PM	<---	FI	0.652
Bcon	<---	FI	0.728
VFM	<---	FI	0.782
AOWRP	<---	FI	0.803
SM	<---	FI	0.525
Adv	<---	FI	0.557
PAD	<---	FI	0.752
ST	<---	FI	0.756
STAC	<---	FI	0.787
PP	<---	FI	0.684
EOU	<---	FI	0.77
VOI	<---	SL	0.78
AOWRB	<---	SL	0.768
PO	<---	SL	0.799
EP	<---	SL	0.773
RP	<---	SL	0.769

Pprice	<---	SL	0.75
DAO	<---	SL	0.774
PQ	<---	SL	0.701
ShipTime	<---	SL	0.767
MonthlySpend	<---	FI	0.005
MonthlySpend	<---	SL	0.032

Table no.66

SL-FI-FREQ.SHOP:


EOU
PP
STAC
ST
PAD
Adv
SM
AOWRP
VFM
Bcon
PM
FI
e11
e10
e9
e8
e7
e6
e5
e4
e3
e2
e1
.59
.49
PQ
DAO
Pprice
RP
EP
PO
AOWRB
VOI
ShipTime
SL
e20
e19
e18
e17
e16
e15
e14
e13
e12
FREQ.SHOP
e21
.01
-.11
.74
.15

MODEL FIT:

Model	Full Form	Calculated Value	Expected Value For Good Fit
CMIN/DF	Discrepancy divided by degree of freedom	1.643	[1-3] or[1-5]or more than 5 weakly fitted
CFI	comparative fit index	0.971	A:0=poor fit close to 1= very good fit B:CFI>.95 : good fit, CFI>0.60 : moderate fit
PCFI	Parsimony comparative fix index	0.800	0= poor fit 1= Exact fit
RMSEA	Root mean square error of approximation	0.048	=0 : good fit <0.08 : close fit >0.08 : weak fit
GFI	Goodness of fit index	0.913	GFI>.90: good fit GFI>0.60: moderate fit
RMR	Root mean square residual	0.064	RMR=0: good fit RMR<0.08: close fit

Table no.67

Regression Weights: (Group number 1 - Default model)							
			Estimate	S.E.	C.R.	P	Label
PM	<---	FI	1				
Bcon	<---	FI	1.152	0.087	13.32	***	par_1
VFM	<---	FI	1.216	0.109	11.172	***	par_2
AOWRP	<---	FI	1.24	0.109	11.398	***	par_3
SM	<---	FI	0.783	0.087	9.041	***	par_4
Adv	<---	FI	0.823	0.081	10.129	***	par_5
PAD	<---	FI	1.19	0.11	10.808	***	par_6
ST	<---	FI	1.147	0.106	10.852	***	par_7
STAC	<---	FI	1.172	0.112	10.476	***	par_8
PP	<---	FI	1.032	0.103	9.988	***	par_9

EOU	<---	FI	1.276	0.116	11.009	***	par_10
VOI	<---	SL	1.065	0.078	13.571	***	par_11
AOWRB	<---	SL	1.079	0.081	13.325	***	par_12
PO	<---	SL	1.053	0.075	14.003	***	par_13
EP	<---	SL	1.152	0.086	13.39	***	par_14
RP	<---	SL	1.135	0.085	13.319	***	par_15
Pprice	<---	SL	0.942	0.073	12.908	***	par_16
DAO	<---	SL	1.043	0.078	13.351	***	par_17
PQ	<---	SL	0.887	0.075	11.907	***	par_18
ShipTime	<---	SL	1				
FREQ.SHOP	<---	FI	-0.111	0.106	-1.042	0.297	par_19
FREQ.SHOP	<---	SL	0.146	0.103	1.413	0.158	par_20

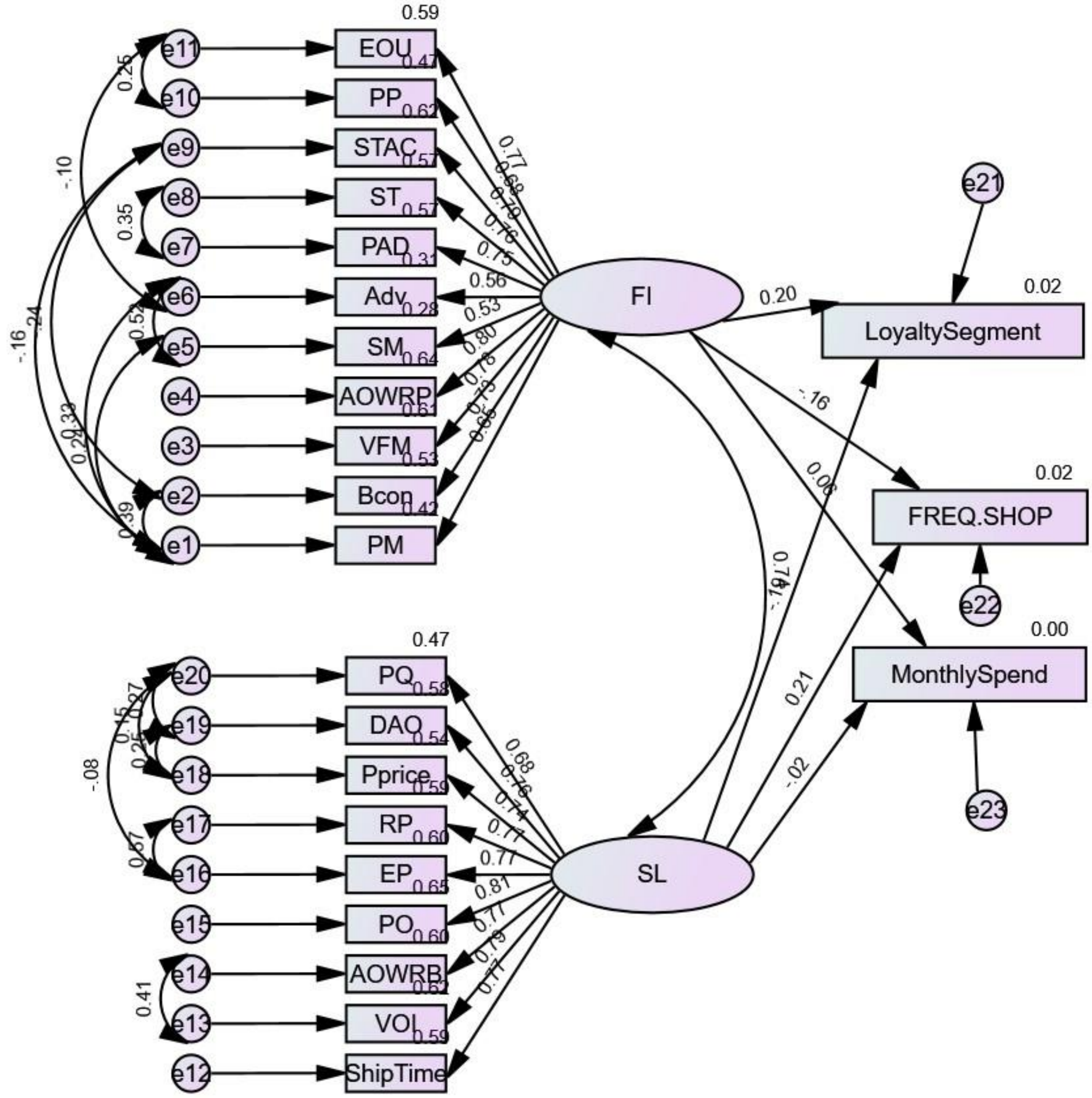
Table no.68

Standardized Regression Weights: (Group number 1 - Default model)			
			Estimate
PM	<---	FI	0.652
Bcon	<---	FI	0.728
VFM	<---	FI	0.783
AOWRP	<---	FI	0.803
SM	<---	FI	0.525
Adv	<---	FI	0.557
PAD	<---	FI	0.752
ST	<---	FI	0.756
STAC	<---	FI	0.787
PP	<---	FI	0.684
EOU	<---	FI	0.77
VOI	<---	SL	0.781
AOWRB	<---	SL	0.769
PO	<---	SL	0.8
EP	<---	SL	0.773
RP	<---	SL	0.769
Pprice	<---	SL	0.748

DAO	<---	SL	0.772
PQ	<---	SL	0.699
ShipTime	<---	SL	0.768
FREQ.SHOP	<---	FI	-0.107
FREQ.SHOP	<---	SL	0.145

Table no.69

FI-SL-LOYALTY-FREQSHOP-MONTHLY SPEND:



MODEL FIT:

Model	Full Form	Calculated Value	Expected Value For Good Fit
CMIN/DF	Discrepancy divided by degree of freedom	2.598	[1-3] or[1-5]or more than 5 weakly fitted
CFI	comparative fit index	0.916	A:0=poor fit close to 1= very good fit B:CFI>.95 : good fit, CFI>0.60 : moderate fit
PCFI	Parsimony comparative fix index	0.764	0= poor fit 1= Exact fit
RMSEA	Root mean square error of approximation	0.076	=0 : good fit <0.08 : close fit >0.08 : weak fit
GFI	Goodness of fit index	0.867	GFI>.90: good fit GFI>0.60: moderate fit
RMR	Root mean square residual	0.084	RMR=0: good fit RMR<0.08: close fit

Table no.70

Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
PM	<---	FI	1.000				
Bcon	<---	FI	1.150	.087	13.293	***	par_1
VFM	<---	FI	1.218	.109	11.162	***	par_2
AOWRP	<---	FI	1.241	.109	11.383	***	par_3
SM	<---	FI	.785	.087	9.038	***	par_4
Adv	<---	FI	.823	.081	10.117	***	par_5
PAD	<---	FI	1.191	.110	10.794	***	par_6
ST	<---	FI	1.148	.106	10.839	***	par_7
STAC	<---	FI	1.174	.112	10.473	***	par_8
PP	<---	FI	1.033	.104	9.973	***	par_9

			Estimate	S.E.	C.R.	P	Label
EOU	<---	FI	1.280	.116	11.011	***	par_10
			Estimate	**S.E.**	**C.R.**	**P**	**Label**
VOI	<---	SL	1.074	.079	13.621	***	par_11
AOWRB	<---	SL	1.086	.081	13.337	***	par_12
PO	<---	SL	1.062	.076	14.051	***	par_13
EP	<---	SL	1.154	.087	13.342	***	par_14
RP	<---	SL	1.136	.086	13.244	***	par_15
Pprice	<---	SL	.927	.074	12.551	***	par_16
DAO	<---	SL	1.034	.079	13.098	***	par_17
PQ	<---	SL	.868	.075	11.503	***	par_18
ShipTime	<---	SL	1.000				
LoyaltySegment	<---	FI	.302	.159	1.903	.057	par_34
LoyaltySegment	<---	SL	-.277	.154	-1.797	.072	par_35
FREQ.SHOP	<---	FI	-.170	.107	-1.585	.113	par_37
MonthlySpend	<---	FI	.058	.106	.546	.585	par_38
FREQ.SHOP	<---	SL	.207	.104	1.985	.047	par_39
MonthlySpend	<---	SL	-.021	.103	-.204	.838	par_40

Table no.71

Covariances: (Group number 1 - Default model)

			M.I.	**Par Change**
e22	<-->	e23	9.300	-.142
e21	<-->	e23	70.111	.575
e21	<-->	e22	111.864	-.730
e17	<-->	e18	8.817	.092
e13	<-->	e15	6.351	.078
e11	<-->	e14	4.007	.080
e11	<-->	e12	4.535	-.091
e10	<-->	e17	5.211	-.088

e9 <--> e13	10.541	-.115
	M.I.	**Par Change**
e8 <--> e15	6.437	.093
e8 <--> e9	4.035	.082
e7 <--> e23	4.035	-.090
e7 <--> e19	5.449	.089
e7 <--> e17	10.701	.116
e7 <--> e16	8.028	-.101
e6 <--> e12	4.391	.085
e5 <--> e9	4.010	-.093
e4 <--> e16	4.563	.075
e2 <--> e19	5.129	-.089
e2 <--> e13	4.735	-.080
e1 <--> e17	7.666	-.098
e1 <--> e16	11.794	.122

Table no.72

Mediation:

Mediation occurs when a structure called a mediator structure intervenes between two other related structures. More specifically ,changes in exogenous constructs lead to changes in mediator constructs , which in turn lead to changes in endogenous constructs in the PLS path way model . In the presence of such effects , mediation can be a useful statistical analysis when supported by theory and properly implemented.

A direct effect **represents a relationship that connects** two **components** with a single arrow. **An indirect effect is a** structural model **pathway** that **includes** a **set** of relationships with at least on intervening **component.** Thus an indirect effect is a **series** of two or more direct **effects,** represented visually by multiple arrows. **The total effect is** the sum of the direct and indirect **effects.**

Three types of mediation:

1. Complementary mediation:
Indirect and direct **effects** are **important** and point in the same direction and both are significant.

2. Competition mediation:
Indirect and direct **effects** are **important** but point in opposite directions and both are significant

3. Indirect mediation only:
Indirect effects are important, direct **effects are not.**

Additionally, they distinguish between two forms of non-mediation:

- **Direct-only non-mediation**: the direct effect is significant, but not the indirect effect.
- **No-effect non-mediation**: neither the direct nor the indirect effect is significant.

As a result, a mediation analysis may reveal that there is no mediation at all (i.e., direct-only non-mediation and no-effect non-mediation) or, in the case of a mediation effect, the mediator construct accounts for all or part of the observed relationship between two latent variables (i.e., complementary and competitive mediation) (i.e., indirect-only mediation).

The explanation of the mechanism by which an intervention affects a result through mediation takes into account both causal and temporal relationships. Mediation analysis aids in providing a focus for future intervention research, enabling the development of more effective and reasonably priced alternative therapies. Mediation analysis is most effective when carried out using solid prior theory and in the proper context. Mediation analysis can be carried out using a fairly open-ended, flexible framework that structural equation modelling offers

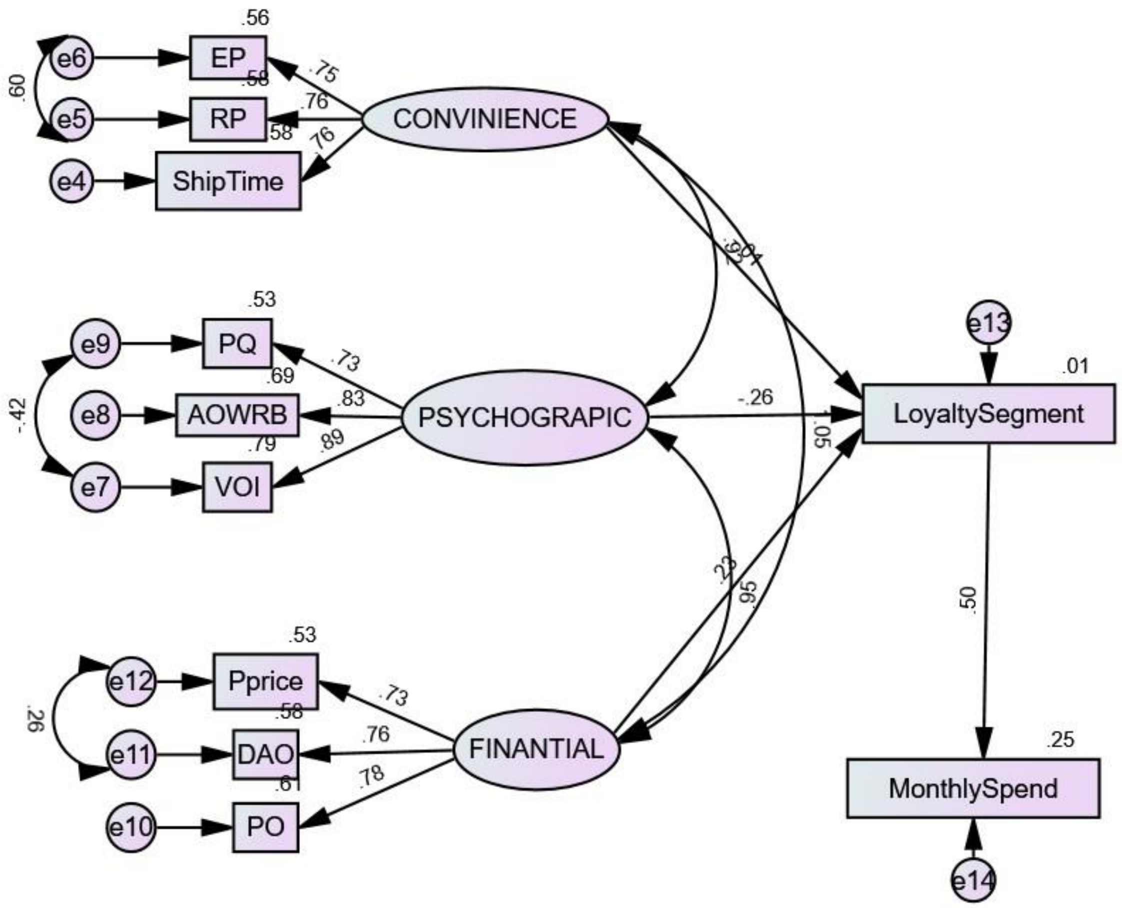
1. OBJECTIVE:

To understand impact of satisfaction level (SL) of customer on their monthly spend mediate through loyalty segment of customer.

MODEL DESCRIPTION:

In this model, Satisfaction Level is proposed to flow through Loyalty segment of customers to their Monthly Spend on shopping.

The satisfaction level is categorised into 3 different factors i.e., convenience, psychographic and financial.



MODEL FIT:

Model	Full Form	Calculated Value	Expected Value For Good Fit
CMIN/DF	Discrepancy divided by degree of freedom	1.944	[1-3] or[1-5]or more than 5 weakly fitted

CFI	comparative fit index	0.981	A:0=poor fit close to 1= very good fit B:CFI>.95 : good fit, CFI>0.60 : moderate fit
PCFI	Parsimony comparative fix index	0.642	0= poor fit 1= Exact fit
RMSEA	Root mean square error of approximation	0.058	=0 : good fit <0.08 : close fit >0.08 : weak fit
GFI	Goodness of fit index	0.954	GFI>.90: good fit GFI>0.60: moderate fit
RMR	Root mean square residual	0.043	RMR=0: good fit RMR<0.08: close fit

Table no.73

Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
LoyaltySegment	<---	CONVINIENCE	-.016	.586	-.028	.978	par_13
LoyaltySegment	<---	PSYCHOGRAPIC	-.317	.321	-.985	.324	par_14
LoyaltySegment	<---	FINANTIAL	.341	.659	.518	.605	par_15
ShipTime	<---	CONVINIENCE	1.000				
RP	<---	CONVINIENCE	1.135	.086	13.152	***	par_1
EP	<---	CONVINIENCE	1.133	.087	12.956	***	par_2
VOI	<---	PSYCHOGRAPIC	1.000				
AOWRB	<---	PSYCHOGRAPIC	.964	.056	17.285	***	par_3
PQ	<---	PSYCHOGRAPIC	.765	.062	12.284	***	par_4
PO	<---	FINANTIAL	1.000				
DAO	<---	FINANTIAL	1.008	.073	13.775	***	par_5
Pprice	<---	FINANTIAL	.895	.069	13.018	***	par_6
MonthlySpend	<---	LoyaltySegment	.335	.035	9.626	***	par_16

Table no.74 **Interpretation:**

The relationship between loyalty of customers who shop online and their monthly spend on online shopping is significant.

Standardized Regression Weights: (Group number 1 - Default model)

			Estimate
LoyaltySegment	<---	CONVINIENCE	-.011
LoyaltySegment	<---	PSYCHOGRAPIC	-.258
LoyaltySegment	<---	FINANTIAL	.234
ShipTime	<---	CONVINIENCE	.761
RP	<---	CONVINIENCE	.761
EP	<---	CONVINIENCE	.752
VOI	<---	PSYCHOGRAPIC	.889
AOWRB	<---	PSYCHOGRAPIC	.833
PQ	<---	PSYCHOGRAPIC	.731
PO	<---	FINANTIAL	.778
DAO	<---	FINANTIAL	.762
Pprice	<---	FINANTIAL	.728
MonthlySpend	<---	LoyaltySegment	.501

Table no.75

Standardized Indirect Effects (Group number 1 - Default model)

	FINANTIAL	PSYCHOGRAPIC	CONVINIENCE	Loyalty Segment
Loyalty Segment	.000	.000	.000	.000
Monthly Spend	.117	-.129	-.005	.000
P price	.000	.000	.000	.000
DAO	.000	.000	.000	.000
PO	.000	.000	.000	.000
PQ	.000	.000	.000	.000
AOWRB	.000	.000	.000	.000
VOI	.000	.000	.000	.000
EP	.000	.000	.000	.000
RP	.000	.000	.000	.000
Ship Time	.000	.000	.000	.000

Table no.76

Modification indices

Covariances: (Group number 1 - Default model)

			M.I.	Par Change
e9	<-->	FINANTIAL	11.156	.085
e9	<-->	e11	15.026	.147
e8	<-->	FINANTIAL	4.070	-.047
e8	<-->	CONVINIENCE	5.055	.057
e8	<-->	e11	4.932	-.076
e8	<-->	e9	5.098	-.081
e7	<-->	e12	4.404	-.065
e5	<-->	FINANTIAL	5.144	.048
e5	<-->	e12	7.637	.085
e5	<-->	e7	4.173	-.059
e4	<-->	e8	5.153	.080

Table no.77

LOWER BOUND:

Indirect Effects - Lower Bounds (BC) (Group number 1 - Default model)

	FINANTIAL	PSYCHOGRAPIC	CONVINIENCE	LoyaltySegment
Loyalty Segment	.000	.000	.000	.000
Monthly Spend	-.276	-.418	-.779	.000
P price	.000	.000	.000	.000
DAO	.000	.000	.000	.000
PO	.000	.000	.000	.000
PQ	.000	.000	.000	.000
AOWRB	.000	.000	.000	.000
VOI	.000	.000	.000	.000
EP	.000	.000	.000	.000
RP	.000	.000	.000	.000
Ship Time	.000	.000	.000	.000

Table no.78

UPPER BOUND:

Indirect Effects - Upper Bounds (BC) (Group number 1 - Default model)

	FINANTIAL	PSYCHOGRAPIC	CONVINIENCE	LoyaltySegment
Loyalty Segment	.000	.000	.000	.000
Monthly Spend	1.034	.093	.362	.000
P price	.000	.000	.000	.000
DAO	.000	.000	.000	.000
PO	.000	.000	.000	.000
PQ	.000	.000	.000	.000
AOWRB	.000	.000	.000	.000
VOI	.000	.000	.000	.000
EP	.000	.000	.000	.000
RP	.000	.000	.000	.000
Ship Time	.000	.000	.000	.000

Table no.79

TWO TAILED:

Indirect Effects - Two Tailed Significance (BC) (Group number 1 - Default model)

	FINANTIAL	PSYCHOGRAPIC	CONVINIENCE	LoyaltySegment
Loyalty Segment	...	...	...	...
Monthly Spend	.481	.321	.913	...
P price	...	...	...	...
DAO	...	...	...	...
PO	...	...	...	...
PQ	...	...	...	...
AOWRB	...	...	...	...
VOI	...	...	...	...
EP	...	...	...	...
RP	...	...	...	...
Ship Time	...	...	...	...

Table no.80

INDIRECT EFFECT:

Indirect Effects (Group number 1 - Default model)

	FINANTIAL	PSYCHOGRAPIC	CONVINIENCE	LoyaltySegment
Loyalty Segment	.000	.000	.000	.000
Monthly Spend	.114	-.106	-.005	.000
P price	.000	.000	.000	.000
DAO	.000	.000	.000	.000
PO	.000	.000	.000	.000
PQ	.000	.000	.000	.000
AOWRB	.000	.000	.000	.000
VOI	.000	.000	.000	.000
EP	.000	.000	.000	.000
RP	.000	.000	.000	.000
Ship Time	.000	.000	.000	.000

Table no.81

Composite reliability:

	CR	**MSV**	**Convergent validity: AVE**
PSYCHOGRAPIC	0.716	0.906	0.521
CONVINIENCE	0.692	1.103	0.431
FINANTIAL	0.737	1.103	0.443

Table no.82

Discriminant validity:

	PSYCHOGRAPIC	**CONVINIENCE**	**FINANTIAL**

PSYCHOGRAPIC	**0.722**		
CONVINIENCE	0.916	**0.656**	
FINANTIAL	0.952	1.050	**0.665**

Table no.83

Discriminant Validity:

The square root of the AVE for PSYCHOGRAPIC is less than the absolute value of the correlations with another factor.

The square root of the AVE for CONVINIENCE is less than the absolute value of the correlations with another factor.

The square root of the AVE for FINANTIAL is less than the absolute value of the correlations with another factor.

The AVE for PSYCHOGRAPIC is less than the MSV.

The AVE for CONVINIENCE is less than the MSV.

T he AVE for FINANTIAL is less than the MSV.

Convergent Validity:

The AVE for CONVINIENCE is less than 0.50.

The AVE for FINANTIAL is less than 0.50.

Reliability:

The CR for CONVINIENCE is less than 0.70.

2. OBJECTIVE:

Impact of factors influencing (FI) and satisfaction level (SL) of customer shopping online on their monthly spend mediate through loyalty segment of customer.

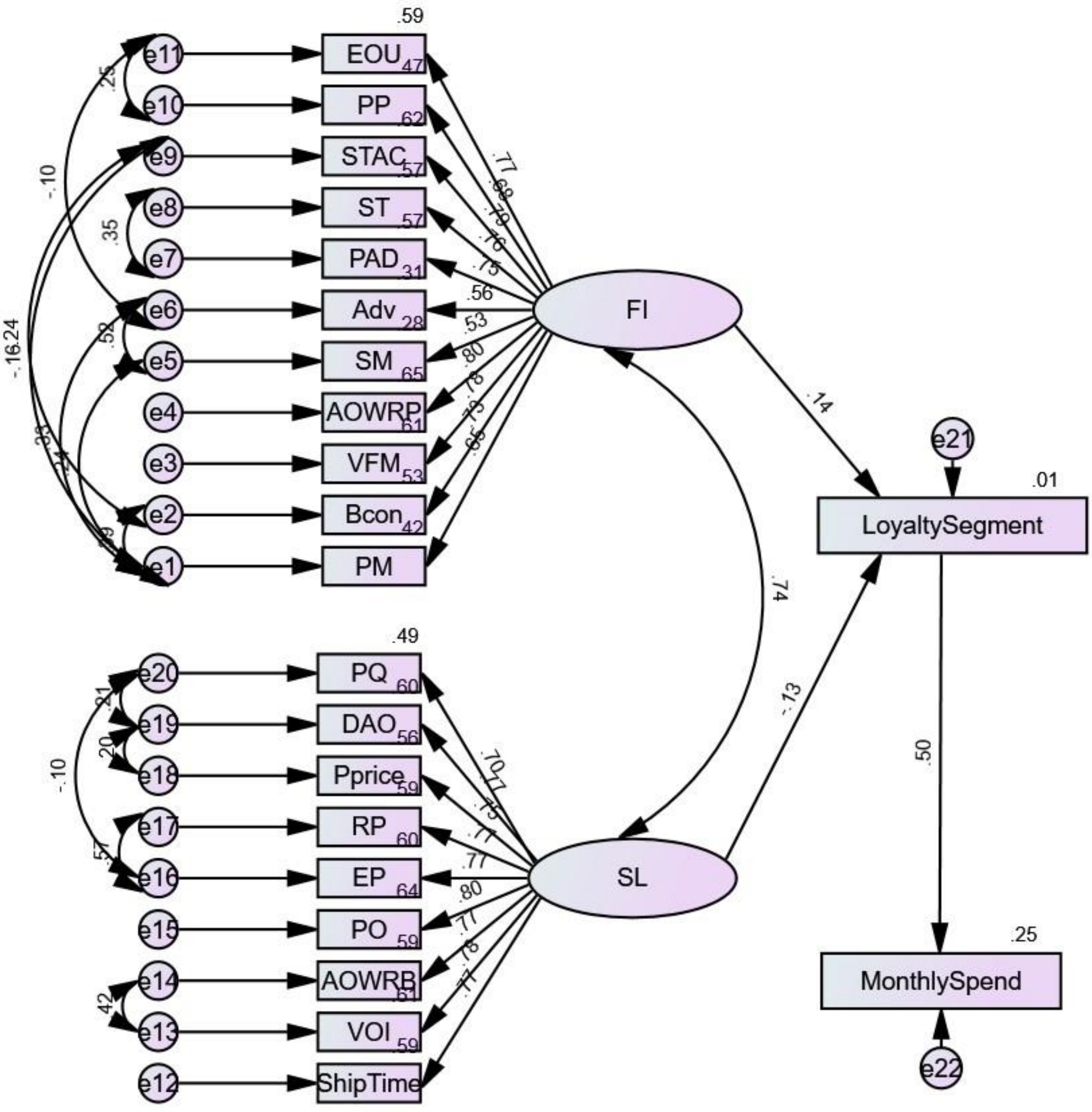
MODEL DESCRIPTION:

In this model, the influence of Factor Influenced online shopping and their Satisfaction Level is proposed to flow through Loyalty segment of customers to their Monthly Spend on shopping.

The concept is that a customer's satisfaction level and factor-influenced online buying have an impact on their loyalty section, which in turn has an impact on their monthly spending on shopping.

It is obvious that customers who show a strong sense of satisfaction during their online shopping activities and further believe that this method is convenient, affordable, and easy to use are more likely to develop a strong sense of loyalty toward online shopping.

Once a customer has become loyal, the model suggests that this will lead to an increase in their Monthly Spend on shopping. This is because loyal consumers are more likely to make repeat purchases and to gradually increase their spending.


.59
EOU
.47
PP
.62
STAC
.57
ST
.57
PAD
.31
Adv
.28
SM
.65
AOWRP
.61
VFM
.53
Bcon
.42
PM
FI
SL
.49
PQ
.60
DAO
.56
Pprice
.59
RP
.60
EP
.64
PO
.59
AOWRB
.61
VOI
.59
ShipTime
LoyaltySegment
.01
MonthlySpend
.25
.14
.74
-.13
.50
e1
e2
e3
e4
e5
e6
e7
e8
e9
e10
e11
e12
e13
e14
e15
e16
e17
e18
e19
e20
e21
e22

MODEL FIT :

Model	Full Form	Calculated Value	Expected Value For Good Fit
CMIN/DF	Discrepancy divided by degree of freedom	1.554	[1-3] or[1-5]or more than 5 weakly fitted
CFI	comparative fit index	0.972	A:0=poor fit close to 1= very good fit B:CFI>.95 : good fit, CFI>0.60 : moderate fit
PCFI	Parsimony comparative fix index	0.812	0= poor fit 1= Exact fit
RMSEA	Root mean square error of approximation	0.064	=0: good fit <0.08: close fit >0.08 : weak fit
GFI	Goodness of fit index	0.914	GFI>.90: good fit GFI>0.60: moderate fit
RMR	Root mean square residual	0.045	RMR=0: good fit RMR<0.08: close fit

Table no.84

Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
LoyaltySegment	<---	FI	.221	.158	1.403	.161	par_19
LoyaltySegment	<---	SL	-.194	.153	-1.270	.204	par_20
PM	<---	FI	1.000				
Bcon	<---	FI	1.152	.087	13.311	***	par_1
VFM	<---	FI	1.217	.109	11.162	***	par_2
AOWRP	<---	FI	1.240	.109	11.390	***	par_3
SM	<---	FI	.784	.087	9.042	***	par_4
Adv	<---	FI	.823	.081	10.125	***	par_5
PAD	<---	FI	1.191	.110	10.802	***	par_6
ST	<---	FI	1.147	.106	10.841	***	par_7
STAC	<---	FI	1.173	.112	10.475	***	par_8
PP	<---	FI	1.033	.103	9.984	***	par_9
EOU	<---	FI	1.278	.116	11.007	***	par_10
VOI	<---	SL	1.065	.079	13.547	***	par_11
AOWRB	<---	SL	1.080	.081	13.306	***	par_12
PO	<---	SL	1.053	.075	13.979	***	par_13

EP	<---	SL	1.153	.086	13.376	***	par_14
RP	<---	SL	1.136	.085	13.302	***	par_15
Pprice	<---	SL	.945	.073	12.925	***	par_16
DAO	<---	SL	1.045	.078	13.347	***	par_17
PQ	<---	SL	.889	.075	11.912	***	par_18
ShipTime	<---	SL	1.000				
MonthlySpend	<---	LoyaltySegment	.335	.035	9.626	***	par_21

Table no.85

Interpretation:

The relationship between loyalty of customers who shop online and their monthly spend on online shopping is significant. To put it more precisely, the conclusion that can be reached from scientific research is that people who show significant loyalty to online shopping are likely to spend more money on online shopping each time. This supports the idea that increasing customer loyalty is an effective strategy for increasing sales, which will eventually increase overall revenue.

Standardized Indirect Effects (Group number 1 - Default model)

	SL	FI	Loyalty Segment
Loyalty Segment	.000	.000	.000
Monthly Spend	-.065	.072	.000
PQ	.000	.000	.000
DAO	.000	.000	.000
Pprice	.000	.000	.000
RP	.000	.000	.000
EP	.000	.000	.000
PO	.000	.000	.000
AOWRB	.000	.000	.000
VOI	.000	.000	.000
Ship Time	.000	.000	.000
EOU	.000	.000	.000
PP	.000	.000	.000
STAC	.000	.000	.000
ST	.000	.000	.000
PAD	.000	.000	.000
Adv	.000	.000	.000
SM	.000	.000	.000
AOWRP	.000	.000	.000
VFM	.000	.000	.000
Bcon	.000	.000	.000
PM	.000	.000	.000

Table no.86

MODIFICATION INDICES:

Covariances: (Group number 1 - Default model)

			M.I.	Par Change
e17	<-->	e18	6.791	.081
e13	<-->	e15	7.782	.088
e11	<-->	e12	4.666	-.092
e10	<-->	e17	5.126	-.087
e9	<-->	e13	10.489	-.115
e8	<-->	e15	6.145	.091
e8	<-->	e9	4.036	.082
e7	<-->	e22	4.411	-.082
e7	<-->	e19	5.688	.091
e7	<-->	e17	11.082	.118
e7	<-->	e16	8.109	-.102
e7	<-->	e13	4.133	.073
e6	<-->	e12	4.406	.085
e5	<-->	e9	4.016	-.093
e4	<-->	e16	4.539	.075
e2	<-->	e19	4.948	-.087
e2	<-->	e13	4.686	-.080
e1	<-->	e17	7.770	-.099
e1	<-->	e16	11.778	.122

Table no.87

LOWER BOUND:

Indirect Effects - Lower Bounds (BC) (Group number 1 - Default model)

	SL	FI	LoyaltySegment
LoyaltySegment	.000	.000	.000
MonthlySpend	-.138	-.004	.000
PQ	.000	.000	.000
DAO	.000	.000	.000
Pprice	.000	.000	.000
RP	.000	.000	.000
EP	.000	.000	.000
PO	.000	.000	.000
AOWRB	.000	.000	.000
VOI	.000	.000	.000
ShipTime	.000	.000	.000
EOU	.000	.000	.000

PP	.000	.000	.000
STAC	.000	.000	.000
ST	.000	.000	.000
PAD	.000	.000	.000
Adv	.000	.000	.000
SM	.000	.000	.000
AOWRP	.000	.000	.000
VFM	.000	.000	.000
Bcon	.000	.000	.000
PM	.000	.000	.000

Table no.88

UPPER BOUND:

Indirect Effects - Upper Bounds (BC) (Group number 1 - Default model)

	SL	FI	LoyaltySegment
LoyaltySegment	.000	.000	.000
MonthlySpend	.019	.159	.000
PQ	.000	.000	.000
DAO	.000	.000	.000
Pprice	.000	.000	.000
RP	.000	.000	.000
EP	.000	.000	.000
PO	.000	.000	.000
AOWRB	.000	.000	.000
VOI	.000	.000	.000
ShipTime	.000	.000	.000
EOU	.000	.000	.000
PP	.000	.000	.000
STAC	.000	.000	.000
ST	.000	.000	.000
PAD	.000	.000	.000
Adv	.000	.000	.000
SM	.000	.000	.000
AOWRP	.000	.000	.000
VFM	.000	.000	.000
Bcon	.000	.000	.000
PM	.000	.000	.000

Table no.89

TWO TAILED:

Indirect Effects - Two Tailed Significance (BC) (Group number 1 - Default model)

	SL	FI	LoyaltySegment
LoyaltySegment	...	...	...
MonthlySpend	.215	.120	...
PQ	...	...	...
DAO	...	...	...
Pprice	...	...	...
RP	...	...	...
EP	...	...	...
PO	...	...	...
AOWRB	...	...	...
VOI	...	...	...
ShipTime	...	...	...
EOU	...	...	...
PP	...	...	...
STAC	...	...	...
ST	...	...	...
PAD	...	...	...
Adv	...	...	...
SM	...	...	...
AOWRP	...	...	...
VFM	...	...	...
Bcon	...	...	...
PM	...	...	...

Table no.90

INDIRECT EFFECT:

Indirect Effects (Group number 1 - Default model)

	SL	FI	LoyaltySegment
LoyaltySegment	.000	.000	.000
MonthlySpend	-.065	.074	.000
PQ	.000	.000	.000
DAO	.000	.000	.000
Pprice	.000	.000	.000
RP	.000	.000	.000
EP	.000	.000	.000
PO	.000	.000	.000
AOWRB	.000	.000	.000
VOI	.000	.000	.000

ShipTime	.000	.000	.000
EOU	.000	.000	.000
PP	.000	.000	.000
STAC	.000	.000	.000
ST	.000	.000	.000
PAD	.000	.000	.000
Adv	.000	.000	.000
SM	.000	.000	.000
AOWRP	.000	.000	.000
VFM	.000	.000	.000
Bcon	.000	.000	.000
PM	.000	.000	.000

Table no.91

INTERPRETATION:

INDIRECT EFFECTS - LOWER BOUND AND UPPER BOUND:

The Following results are from the Bias-corrected percentile method link. It will first present the lower bound indirect effects and then the upper bound indirect effects. With the indirect test of SI to LOYALTY SEGMENT through FI, the lower bound confidence interval is -0.138 and the upper bound is 0.019.Since there is no zero between the Lower Bound and Upper bound Confidence Interval, this shows the significance of indirect effects of the relationship between loyalty of customers who shop online and frequency of online shopping.

INDIRECT EFFECTS - TWO TAILED SIGNIFICANCE:

The values of 0.215 at the intersection of SL and LOYALTY SEGMENT shows the Two Tailed Significance. In this case it is 0.215 which is more than 0.05 .Hence, we can conclude that FI mediates the relationship between SL and LOYALTY SEGMENT.

INDIRECT EFFECTS:

Based on these results, we can conclude that SL has a significant indirect effect on LOYALTY SEGMENT through the FI. The indirect effect of SL to LOYALTY SEGMENT is .065 (a*b).The lower bound confidence interval is .138 and the upper bound is .019. Since this confidence interval did not cross zero, we know that the indirect effect is significant.

CONCLUSION:

Logistic regression model is the best model with highest accuracy of 0.88

Which indicates that 'Discounts and Offers', 'Availability of wide range of brands', and 'Varieties of items' have positive coefficients, that are indicating a positive impact on consumer loyalty and the p-values of the coefficients indicate the significance of each independent variable in the logistic regression model. A p-value less than 0.05 indicates that the independent variable is significant in the model. From the output, we can see that 'Product Quality', 'Availability of wide range of brands', and 'Varieties of items' have p-values less than 0.05, indicating that they are significant in the model.

CODES USED FOR VARIABLES IN SPSS:

DO YOU SHOP ONLINE:

Yes	1
No	2

AGE:

Under 18	1
18 - 24	2
25 – 34	3
35 - 44	4
45 – 54	5
Above 54	6

GENDER:

Female	1
Male	2
Prefer not to answer	3

EDUCATION:

Upto ssc	1
Upto hsc	2
Bachelor's degree	3
Master's degree or diploma	4
Professional	5
Other	6

CURRENT STATUS:

Employed	1
Unemployed	2
Housewife	3

A student	4
Retired	5

MARITAL STATUS:

Married	1
Unmarried	2
Divorced	3
Widowed	4

MEMBERS IN THE FAMILY:

1	1
1	2
3	3
4	4
More than 4	5

INCOME:

less than 1 lakh	1
1-5 lakh	2
5-9 lakh	3
more than 9 lakh	4

FREQUENTLY SHOPPING:

Occasionally	1
Monthly	2
Weekly	3
twice a week	4

DURING SALES:

Always	1
Frequently	2
Sometimes	3
Never	4

DURING FESTIVE SEASON:

Always	1
Frequently	2
Sometimes	3
Never	4

OCCASIONALLY:

Always	1
Frequently	2
Sometimes	3
Never	4

SPENDING MONEY ON SHOPPING:

Less than 1000	1
1000-2000	2
2000-5000	3
More than 5000	4

SHOPPING SINCE:

Less than 1 year	1
1-3 year	2
3-5 year	3
More than 5 year	4

RECENCY:

1 month ago	1
Year ago	2

EASE OF USE:

Least	1
Likely	2
Neutral	3
Extremely likely	4
Most	5

PRODUCT PRESENTATION:

Least	1
Likely	2
Neutral	3
Extremely likely	4
Most	5

SHIPPING TIME AND COST:

Least	1
Likely	2
Neutral	3
Extremely likely	4
Most	5

SECURE TRANSACTION:

Least	1
Likely	2
Neutral	3
Extremely likely	4
Most	5

PRICING AND DISCOUNTS:

Least	1
Likely	2
Neutral	3
Extremely likely	4
Most	5

ADVERTISEMENTS:

Least	1
Likely	2
Neutral	3
Extremely likely	4
Most	5

SOCIAL MEDIA:

Least	1
Likely	2
Neutral	3
Extremely likely	4
Most	5

AVAILABILITY OF WIDE RANGE OF PRODUCTS:

Least	1
Likely	2
Neutral	3
Extremely likely	4
Most	5

VALUE FOR MONEY:

Least	1
Likely	2
Neutral	3
Extremely likely	4
Most	5

BRAND CONSIOUSNESS:

Least	1
Likely	2
Neutral	3
Extremely likely	4
Most	5

PRODUCT MARKETING:

Least	1
Likely	2
Neutral	3
Extremely likely	4
Most	5

FASHION:

Always	1
Very frequently	2
Often	3
Occasionally	4
Never	5

ELECTRONICS

Always	1
Very frequently	2
Often	3
Occasionally	4
Never	5

FOOTWEAR:

Always	1
Very frequently	2
Often	3
Occasionally	4
Never	5

BABY PRODUCTS:

Always	1
Very frequently	2
Often	3
Occasionally	4
Never	5

HOME DÉCOR AND FURNISHING:

Always	1
Very frequently	2
Often	3
Occasionally	4
Never	5

FOOD:

Always	1
Very frequently	2
Often	3
Occasionally	4
Never	5

Always	1
Very frequently	2
Often	3
Occasionally	4

Never	5

JEWELLERY:

Always	1
Very frequently	2
Often	3
Occasionally	4
Never	5

BEAUTY PRODUCTS:

Always	1
Very frequently	2
Often	3
Occasionally	4
Never	5

GROCERY:

Always	1
Very frequently	2
Often	3
Occasionally	4
Never	5

GIFT CARDS AND SUPPLIMENTS:

TOYS AND VIDEO GAMES:

Always	1
Very frequently	2
Often	3
Occasionally	4
Never	5

BOOKS:

Always	1
Very frequently	2
Often	3
Occasionally	4
Never	5

Always	1
Very frequently	2
Often	3
Occasionally	4

Never	5

HANDMADE PRODUCTS:

Always	1
Very frequently	2
Often	3
Occasionally	4
Never	5

STATIONERY PRODUCTS:

Always	1
Very frequently	2
Often	3
Occasionally	4
Never	5

MEDICAL SUPPLIES:

Always	1
Very frequently	2
Often	3
Occasionally	4
Never	5

MYNTRA:

AMAZON

Always	1
Very frequently	2
Often	3
Occasionally	4
Never	5

FLIPKART:

Always	1
Very frequently	2
Often	3
Occasionally	4
Never	5

Always	1
Very frequently	2
Often	3
Occasionally	4

Never	5

JIO MART:

Always	1
Very frequently	2
Often	3
Occasionally	4
Never	5

AJIO:

Always	1
Very frequently	2
Often	3
Occasionally	4
Never	5

BIG BASKET:

Always	1
Very frequently	2
Often	3
Occasionally	4
Never	5

SNAPDEAL:

NETMEDS:

Always	1
Very frequently	2
Often	3
Occasionally	4
Never	5

TATA 1MG:

Always	1
Very frequently	2
Often	3
Occasionally	4
Never	5

Least	1
Likely	2
Neutral	3
Extremely likely	4

Most	5

FIRSTCRY:

Always	1
Very frequently	2
Often	3
Occasionally	4
Never	5

OTHER:

Always	1
Very frequently	2
Often	3
Occasionally	4
Never	5

PRODUCT QUALITY:

Least	1
Likely	2
Neutral	3
Extremely likely	4
Most	5

DISCOUNTS AND OFFERS:

PRODUCT PRICES:

Least	1
Likely	2
Neutral	3
Extremely likely	4
Most	5

RETURN POLICY:

Least	1
Likely	2
Neutral	3
Extremely likely	4
Most	5

EXCHANGE POLICY:

Least	1
Likely	2
Neutral	3
Extremely likely	4
Most	5

PAYMENT OPTIONS:

Least	1
Likely	2
Neutral	3
Extremely likely	4
Most	5

AVAILABILITY OF WIDE RANGE OF BRANDS:

Least	1
Likely	2
Neutral	3
Extremely likely	4
Most	5

VARIETY OF ITEMS:

Least	1
Likely	2
Neutral	3
Extremely likely	4
Most	5

SHIPPING TIME:

Least	1
Likely	2
Neutral	3
Extremely likely	4
Most	5

QUESTIONNAIRE:

Hello, we are a group of students pursuing MSc STATISTICS from KC College, Mumbai conducting a survey on online shopping.

The questionnaire aims to understand online shopping behaviour and pattern of customers. Your views are extremely important for the success of this research. Please be as honest as possible as the result are anonymous. We assure you that the information provided by you will be confidential and strictly used for educational purposes only.
We highly appreciate your response.

THANK YOU.

Gender:

- ○ Male
- ○ Female
- ○ Prefer not to answer

Age:

- ○ Under18 ○ 18-24 ○ 25-34 ○ 35-44 ○ 44-54 ○ Above 54

Marital status:

- ○ Widowed ○ Married ○ Unmarried ○ Divorced

You are currently…:

- ○ Employed ○ Unemployed ○ Housewife ○ A student ○ Retired

Qualification:

- ○ Upto SSC
- ○ Upto HSC ○ Bachelor's degree ○ Master's degree or diploma ○ Professional ○ Other

Annual income:

- ○ Less than 1 lakh ○ 1– 5 lakh ○ 5– 9 lakh ○ More than 9 lakh

Including yourself, how many people currently live in your household?

○ 1 ○ 2 ○ 3 ○ 4 ○ more than 4

Do you shop online?

○ Yes ○ No

IF YES:

Where do you get information about online shopping from? :

○ Friends ○
Relatives ○
Advertisement ○
Other:

How Frequently do you shop online?

○ Twice a week
○ Weekly ○
Monthly ○
Occasionally

On scale of 1-5 how recently have you purchase online?

(1 month ago)

○ 1 ○ 2 ○
3 ○ 4 ○ 5

(1 year ago)

How long have you been shopping online?

○ Less than a year
○ 1 – 3 year ○ 3 – 5
year ○ More than 5
year

On what occasions do you prefer online shopping?

	Always	frequently	Sometimes	Never
During sales				
During festive seasons				
Occasionally				

Tick the Product that you buy online?

	Always	Very Frequently	Often	Occasionally	Never
Fashion					
Electronics					
Footwear					
Baby care products					
Home Décor and furnishing					
Food					
Jewellery					
Beauty product					
Online Grocery					
Gift cards and online supplement					
Toys and video Games					
Books					
Handmade Products					
Stationary products					
Medical Supplies					

Tick on the Preferred online shopping platform? :

	Always	Very Frequently	Often	Occasionally	Never
Myntra					
Amazon					

Flipkart					
Jio Mart					
AJIO					
Big Basket					
Snapdeal					
Netmeds					
TATA 1MG					
Firstcry					
other					

How much do you spend on online shopping every month? :

○ Less than 1000
○ 1000 – 2000 ○
2000 – 5000 ○
More than 5000

What Factors Influences you to shop online? (1 being the least and 5 being the most):

	1	2	3	4	5
Ease Of Use					
Product presentation					
Shipping time and cost					
Secure transactions					
Pricing and Discount					
Advertisements					
Social media					
Availabilty of wide range of products					
Value for money					
Brand consciousness					
Product Marketing					

Rate your satisfaction level on the scale of 1-5 (1 = least satisfied to 5 = most satisfied):

	1	2	3	4	5
Product quality					
Discounts and offers					
Products prices					
Return policy					
Exchange policy					
Payment options					
Availability of wide range of brands					
Varieties Of items					
Shipping time					

IF NO:

Scale the problems you face while online purchase (1= Least problem 5=Very problematic):

	1	2	3	4	5
I don't know about online shopping					
Risk of online transactions					
Risk of identity theft					
Internet availability					
Quality of product					
Shipping issue					
Difficulty in returning product					

Prefer offline shopping					

VARIABLE DESCRIPTION

Variable Name	Label
Frequently shopping	FREQ.SHOP
During Sales	DS
During festive season	DFS
Occasionally	O
Monthly spending	MonthlySpend
Shopping since	ShopSince
Ease of use	EOU
Product presentation	PP

Shipping time and cost	STAC
Secure transaction	ST
Pricing and Discount	PAD
Advertisment	Adv
Social Media	SM
Availability of wide range of product	AOWRP
Value for money	VFM
Brand consciousness	BCon
Product Marketing	PM
Home décor and furnishing	Home Décor
Toys and video games	ToysandGames
Product quality	PQ
Discount and offers	DAO
Product pricing	PPrice
Return policy	RP
Exchange policy	EP
Payment option	PO
Availiability of wide range of brand	AOWRB
Variety of items	VOI
Shipping time	Ship Time

BIBLIOGRAPHY:

SR.NO	REFERENCES	PAGE NO.
1.	Agarwal, Anuraag, and Sanskrity Joseph. "A review on Effect of e trust and e risk on Consumers of retail e markets in India: A Comparative Study Based on Sociodemographic Variables." *Journal of Ravishankar University(PART-A)* 26,no. 1(February 12,2021):49–62. http://dx.doi.org/10.52228/jrua.2020-26-1-8.	6
2.	Barbosa, Diogo. "Digital In-Store Shopping." U.Porto Journal of Engineering 2,no. 1(March 19,2018):44–48. http://dx.doi.org/10.24840/2183-6493_002.001_0005	6

3.	Dubey, Anil Roy, and B. Balaji. "A Study on E-Commerce Aspects Impacting Consumer's Online Shopping Behavior in United Arab Emirates." Journal of Computational and Theoretical Nanoscience 18, no. 4 (April 1,2021):1331–35. http://dx.doi.org/10.1166/jctn.2021.9393	6
4.	G. Bhongade, Bhagyashree, and Ashwini V.Z. "A Semi-supervised Approach for Opinion Mining using Online Product Review." International Journal of Engineering & Technology 7, no. 4.5 (September 22,2018):143. http://dx.doi.org/10.14419/ijet.v7i4.5.20031	7
5.	Prashar, Sanjeev, T. Sai Vijay, and Chandan Parsad. "Antecedents to Online Shopping." International Journal of E-Business Research 11, no. 1(January2015):35–55. http://dx.doi.org/10.4018/ijebr.2015010103.	7
6.	Prashar, Sanjeev, Sai Vijay Tata, Chandan Parsad, Abhishek Banerjee, Nikhil Sahakari, and Subham Chatterjee. "Clustering E-Shoppers on the Basis of Shopping Values and Web Characteristics." Journal of Global Information Management 27, no. 2 (April 2019): 24–38. http://dx.doi.org/10.4018/jgim.2019040102.	7
7.	Rao, Katta Rama Mohana, and Chandra Sekhar Patro. "Shopper's Stance Towards Web Shopping." International Journal of Online Marketing 7no. 3(July2017):42–54. http://dx.doi.org/10.4018/ijom.2017070104	7
8.	Reddy, Anuradha, and G. V. Bhavani Prasad. "Consumer Perceptions on Security, Privacy, and Trust on E-Portals." International Journal of OnlineMarketing 2,no. 2(April2012):10–24. http://dx.doi.org/10.4018/ijom.2012040102.	8
9.	Srivastava, Gautam, Neeraj Anand, and Arvind Kumar Jain. "Impact of supply chain management on the digital marketing of perishable goods (fruits and vegetables)." Journal of Global Economy 14, no. 4 (November 8, 2018): 51–60. http://dx.doi.org/10.1956/jge.v14i4.497.	8
10.	Shaheen, Musarrat, Farrah Zeba, Namrata Chatterjee, and Raveesh Krishnankutty. "Engaging customers through credible and useful reviews: the role of online trust." Young Consumers 21, no. 2 (September 19, 2019): 137–53. http://dx.doi.org/10.1108/yc-01-2019-0943.	8