Application-Tastic

UNIVERSITY OF MUMBAI
DEPARTMENT OF STATISTICS
VIDYANAGARI MUMBAI – 400098

CERTIFICATE

This is to certify that Mr./Ms._____of M.Sc. Statistics has successfully completed the project entitled "Application-Tastic" during the academic year 2018-2019.

The Team Comprises of: -

- **Aditi Gangar**
- **Kiran Shirsath**
- **Nikhil Paliwal**
- **Pallavi Pawar**
- **Sailee Kamble**
- **Shweta Bamane**
- **Rambhaji Kolpe**
- **Sanidhya Chandorkar**

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

Dr. Vaijayanti U. Dixit
(*Associate Professor & Project Guide)*

Dr. Santosh. P. Gite
(Head of the department)

ACKNOWLEDGEMENT

From the conception of topic to the final presentation, we are extremely thankful
to our guide **Dr. Vaijayanti U. Dixit** (*Associate Professor Department of Statistics at University of* Mumbai) for her immense support and guidance throughout the process of making this entire project. It was only with the help of her suggestions & corrections, at every step that we were able to complete the project on time & achieve our objectives to our satisfaction.

We are extremely thankful to Dr. Alok Dabade and Professor Priyesh Tiwari for their support and guidance and providing all the necessary information.

Also, we would like to thank Dr. Santosh Gite (*Head of the Department*) for providing us the opportunity to do the project work and giving us all support and guidance, which made us complete the project duly.

Last but not the least, we would like to give our special thanks to our respondents who diligently and honestly filled our survey purely for academic purposes and taking the time out of their schedule to answer our questions genuinely.

INDEX

Introduction

"There's an app for that." How many times have you heard this phrase? Ever since mobile apps came into the market, they took over in almost every category (utilities, productivity, gaming, entertainment, lifestyle, social networking, and news). It is true that there is an app for everything. In fact, you can plan a whole vacation through mobile apps! From choosing a vacation destination to booking your flight to reserving a restaurant for dinner and reviewing that restaurant afterward, life is simply made easier through apps.

Smart phones have become an integral part of our daily life. As technology has gained ample momentum globally. With a strong connectivity, smart gadgets and disrupted network, apps have made a prominent impact in the Indian lifestyle. Mobile users spend 75% on mobile apps and just 25% of the time on mobile websites.

The number of smartphone users in India is estimated to hit 337 million by the end of 2018. The number of smartphone users in India would reach 490.9 million by 2022

According to Digital Intelligence report, 59% of Indians access internet via mobile phones. Irrespective of the industries, mobile apps have taken a huge leap and proved the future of technology.

As a part of promoting mobile apps and emphasizing on e-wallets we have chosen this as our project topic.

Mobile websites

A **website** or **Web site** is a collection of related network web resources, such as web pages, multimedia content, which are typically identified with a common domain name, and published on at least one web server. Notable examples are wikipedia.org, google.com, and amazon.com.

Websites can be accessed via a public Internet Protocol (IP) network, such as the Internet, or a private local area network (LAN), by a uniform resource locator (URL) that identifies the site.

Websites can have many functions and can be used in various fashions; a website can be a personal website, a corporate website for a company, a government website, an organization website.

Mobile Application

A mobile application, most commonly referred to as an app, is a type of application software designed to run on a mobile device, such as a smartphone or tablet computer. Mobile applications frequently serve to provide users with similar services to those accessed on PCs. Apps are generally small, individual software units with limited function.

Type of application

- Native App.
- Hybrid App.
- Web App.

A native application is a software program that is developed for use on a particular platform or device. Because a native app is built for use on a particular device and its OS, it has the ability to use device-specific hardware and software. (eg. WhatsApp**,** Pinterest, Facebook, LinkedIn, Instagram, Google Maps)

Hybrid applications are web applications (or web pages) in the native browser, such as WebView in iOS and WebView in Android (not Safari or Chrome). Hybrid apps are developed using HTML, CSS and JavaScript, and then wrapped in a native application using platforms like Cordova.eg. (Paytm, UBER). A web application or web app is a client–server computer program which the client (including the user interface and client-side logic) runs in a web browser. Common web applications include webmail, online retail sales, and online auction

Advantage and Disadvantage

Mobile Apps

Advantages

- It utilizes the smartphone's existing features like shopping, taking a photo or using GPS location.
- If you have an app, you can improve user experience. Because smartphone market is growing exponentially. It would be an incentive to let users accomplish a specific action via your app.
- It gives more value to customers, because more users prefer mobile devices to desktops
- The best part is generating revenue. There are many ways to let your app gain you like paying a fee to download.
- Mobile apps make your business easily accessible to clients via app store search. Of course, it needs time and effort to be listed on search.
- Mobile apps increase the brand awareness by letting customers easily post, share, or tweet about your services.

Disadvantages

- The high cost of developing an app
- The complexity of app development process
- Require a long-term investment to ensure compliance
- App stores get a substantial cut of the pie
- Fetching data from the app to track metrics is challenging

Objectives

1) To study whether if there is relationship between

i. Educational qualification v/s Awareness of govt. application.
ii. Education Qualification v/s Use of Lifestyle Application.
iii. Gender v/s Awareness of government application.
iv. Education Qualification v/s Do you use govt. application.

2) To find out the reasons that initiates people to download mobile-applications

3) To identify and analyze socio demographic factors that effects people decision to use E-wallet application.

4) To check if demonetization has an effect on usage of E-wallet application.

5) To find out the most preferred E-Wallet application.

6) To obtain an overall review on

i. Different types of different websites.
ii. Usage of top e-wallet application (Paytm, Phonepe, etc.)
iii. Awareness regarding Government Application and reason for not using government application.

METHODOLOGY

Keeping our objective in mind and the techniques to be used, we have designed our questionnaire accordingly. In our survey, we have collected our data through online applications such as WhatsApp, Facebook and mail as well as from offline such as housing societies, colleges, office vicinities etc. We have used convenient sampling in our study. We have conducted pilot study on 100 samples and after a slight modification in our questionnaire; we conducted the actual survey and have collected 896 sample data. After cleaning data, our data reduced to 882.

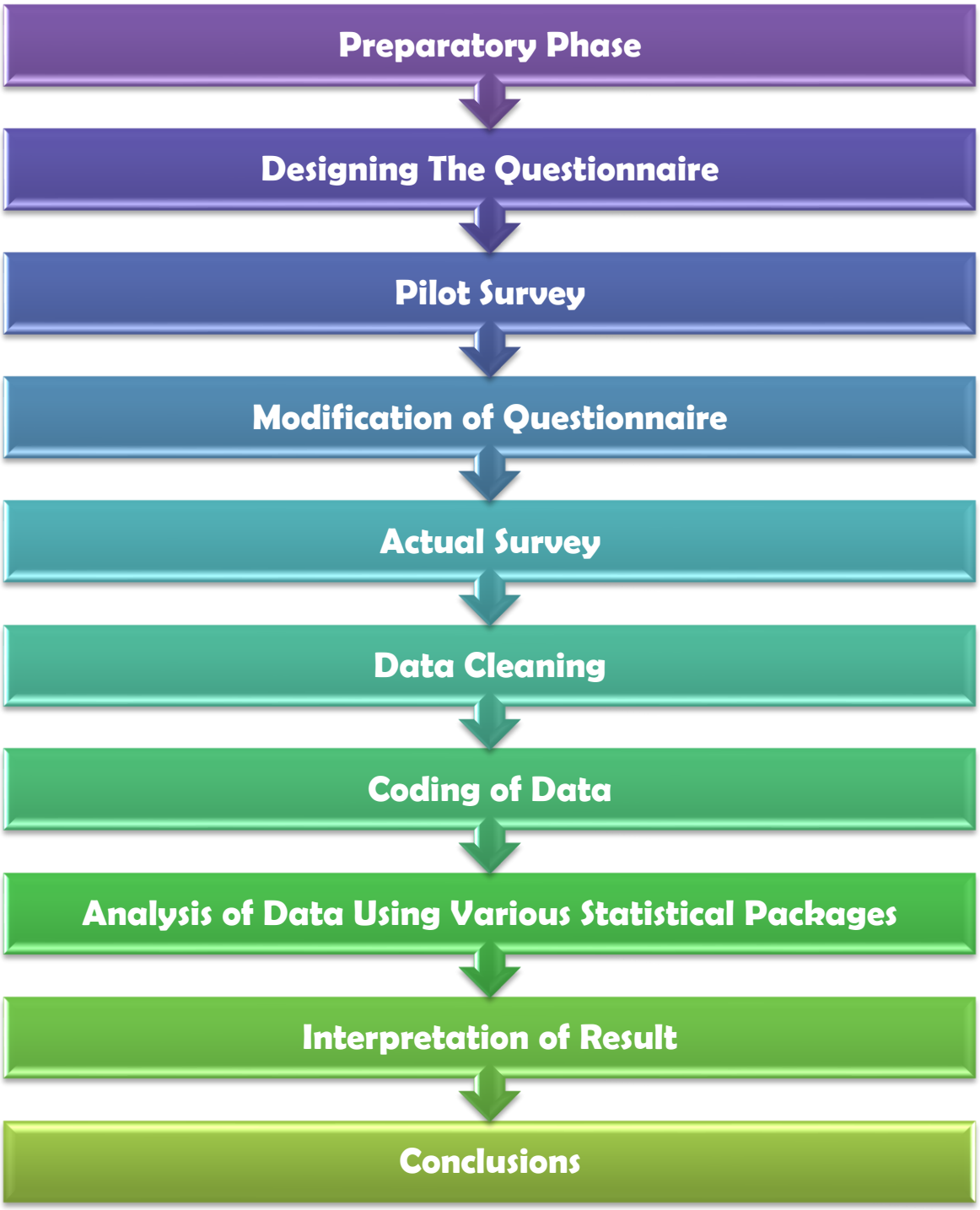
TYPES OF QUESTIONS

- Dichotomous Questions
- Multiple Choice Questions
- Likert Scale Questions
- Categorical Questions
- Open ended question

TECHNIQUES USED

- Graphical Representation
- Chi-square Test
- Binary Logistic Regression
- Wilcoxon Signed Rank Test
- Factor Analysis
- Pareto Analysis
- Word Cloud

STEPS INVOLVED IN CONDUCTING THE SURVEY

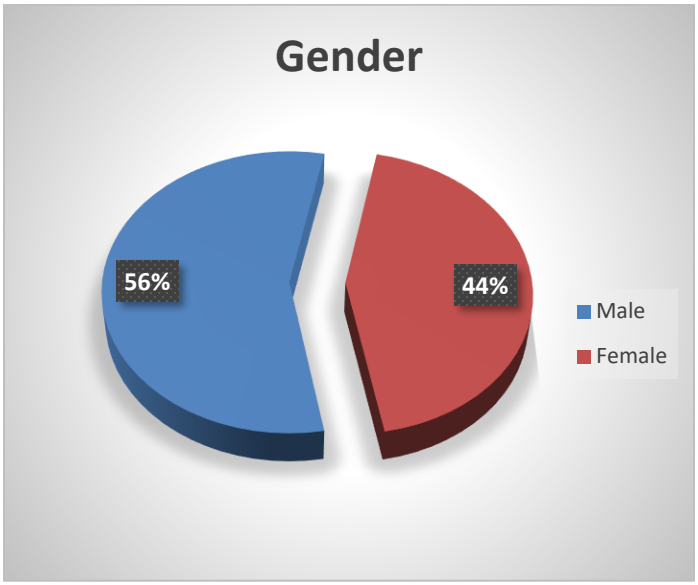


GRAPHICAL REPRESENTATION

The main purpose of graphical representation is to readily give some idea about the entire data and draw instant conclusions.

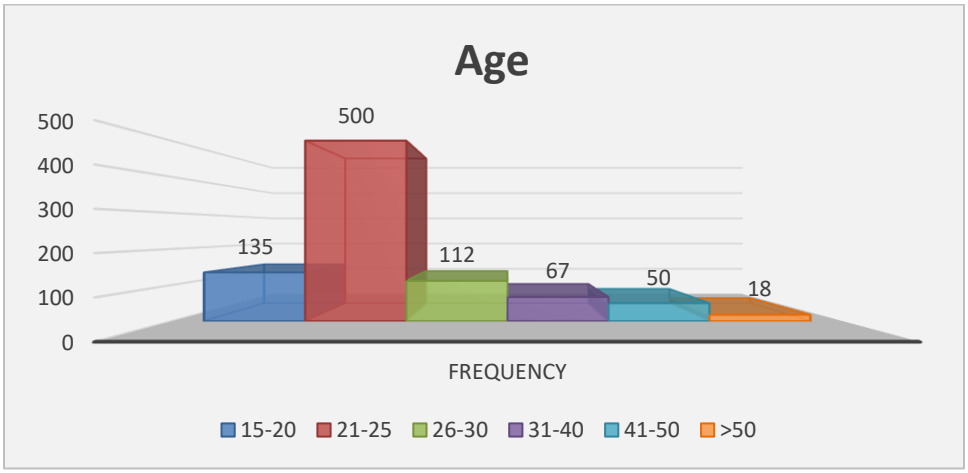
- **Gender**

Gender	Frequency
Male	494
Female	388
Total	882



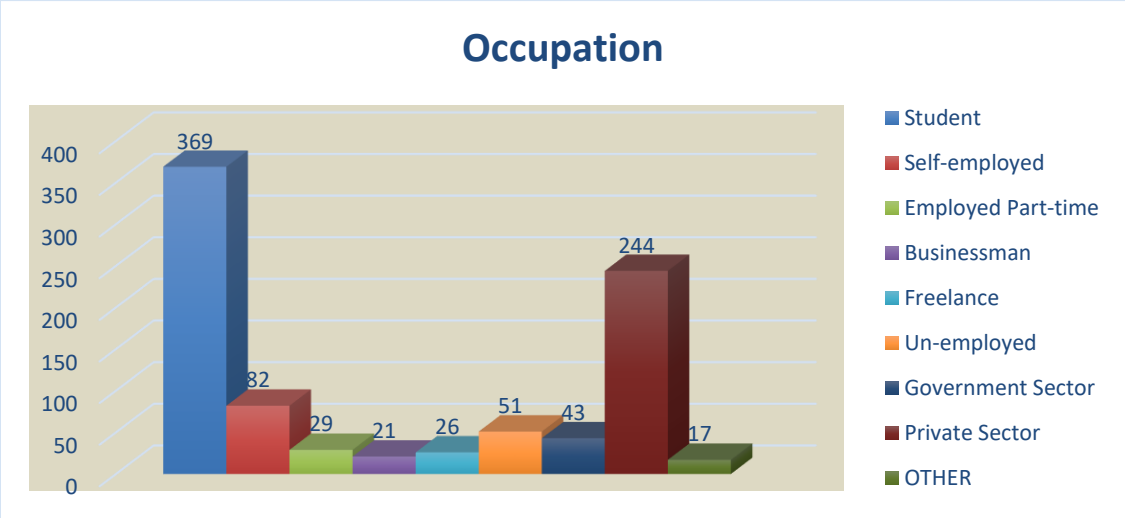
- **Age**

Age group	15-20	21-25	26-30	31-40	41-50	>50
Frequency	135	500	112	67	50	18



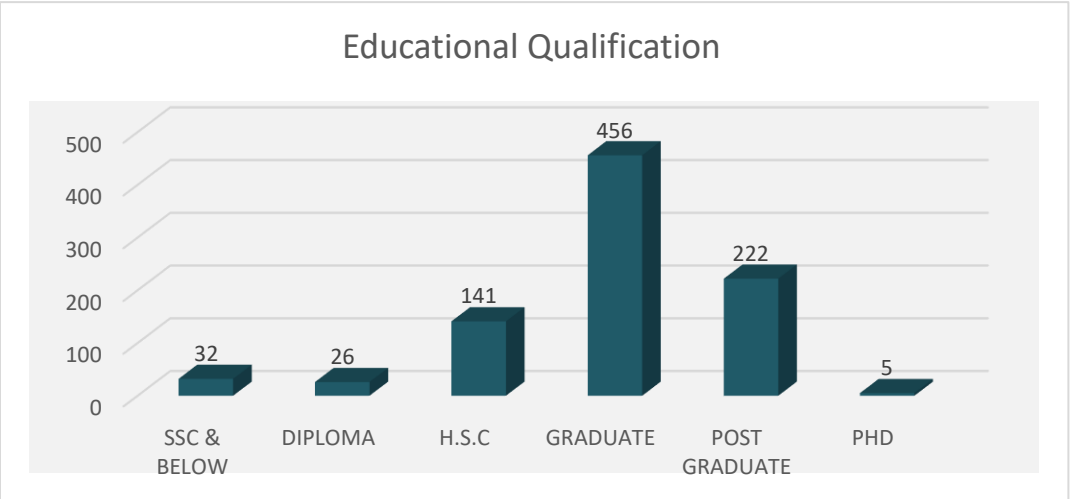
➢ Occupation

Occupation	Student	Self-Employed	Employed Part-time	Businessman	Freelance	Un-employed	Government Sector	Private Factor	Other
Frequency	369	82	29	21	26	51	43	244	17



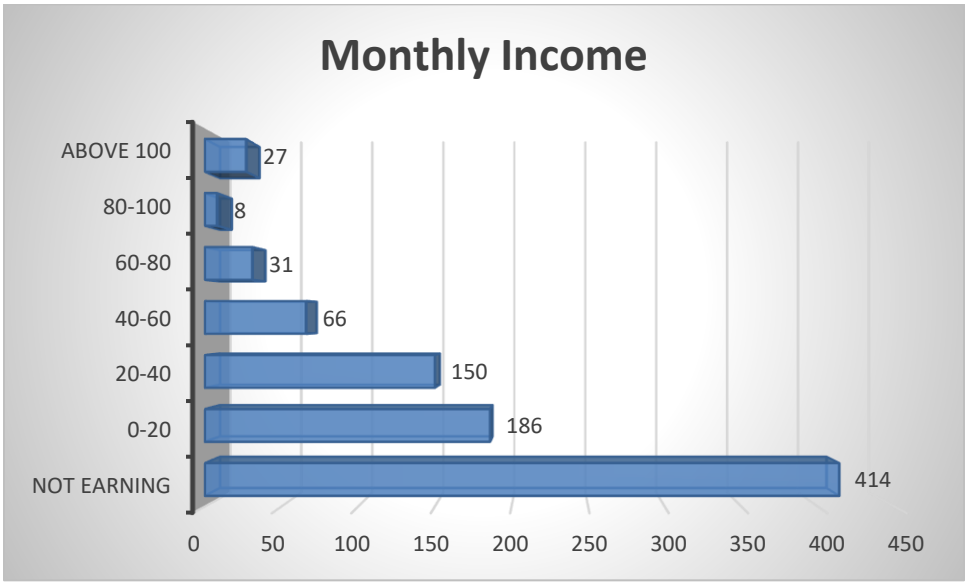
➢ Educational Qualification

Ssc & below	Diploma	H.s.c	Graduate	Post graduate	Phd
32	26	141	456	222	5



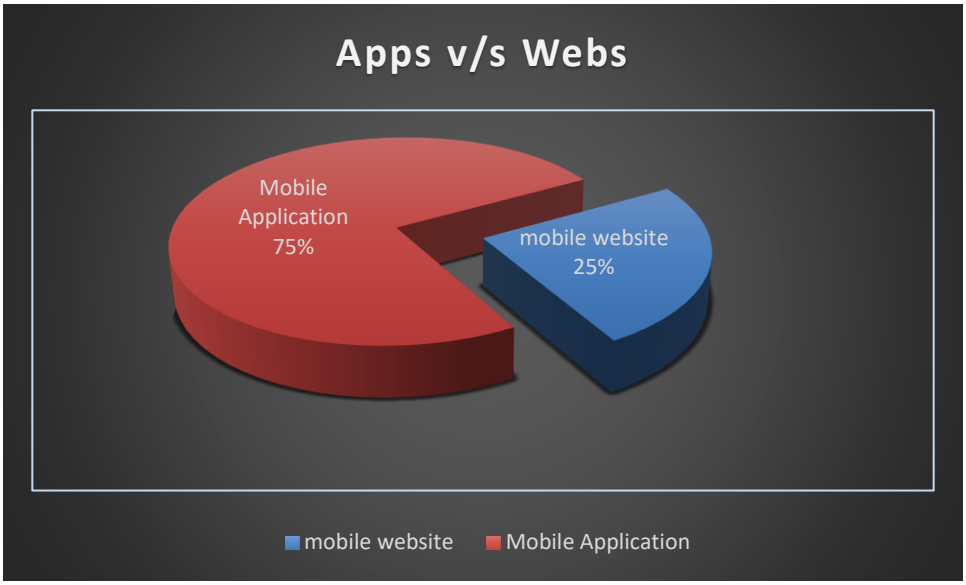
➢ Annual Family Income

Income Range	Frequency
Not Earning	414
0-20,000	186
20,001-40,000	150
40,001-60,000	66
60,001-80,000	31
80,001-100,000	8
Above 100,000	27



➢ Mobile Applications v/s Mobile Websites

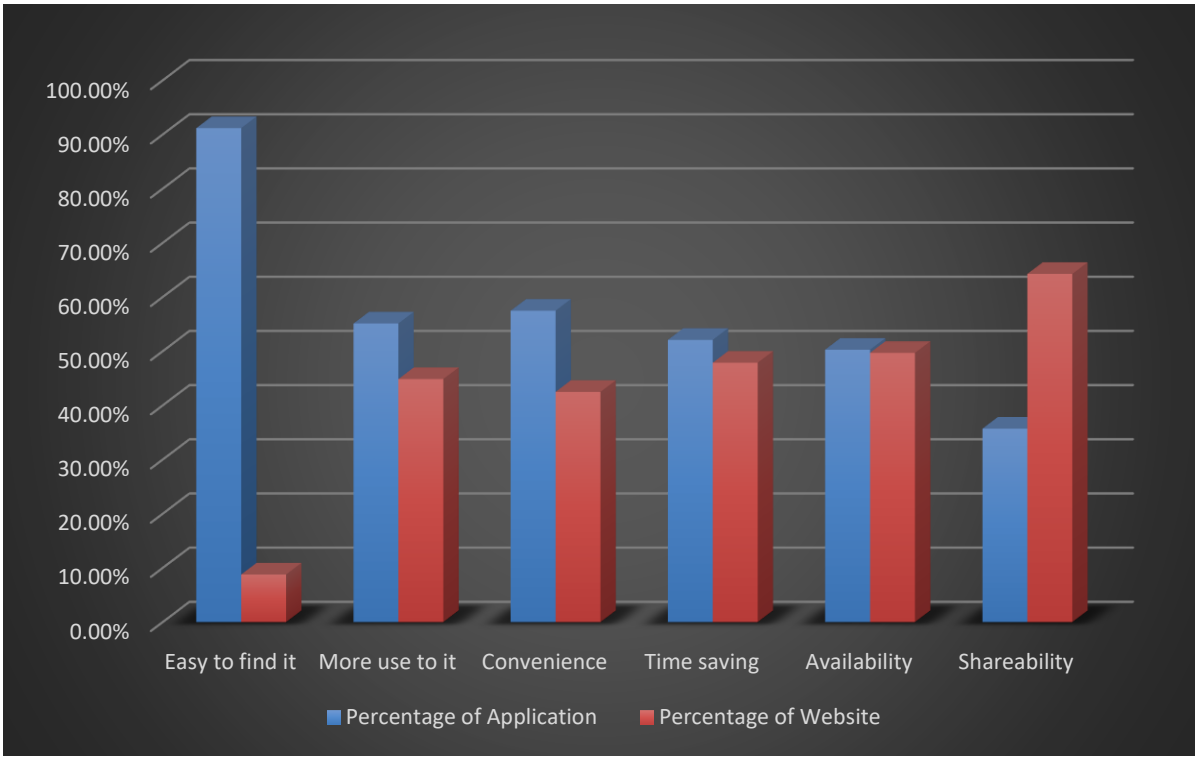
Mobile Applications	663
Mobile Websites	219



Interpretation: According to our survey, 75% from the total sample prefers mobile applications over mobile websites.

Reasons for preferring Mobile Applications & Mobile Websites

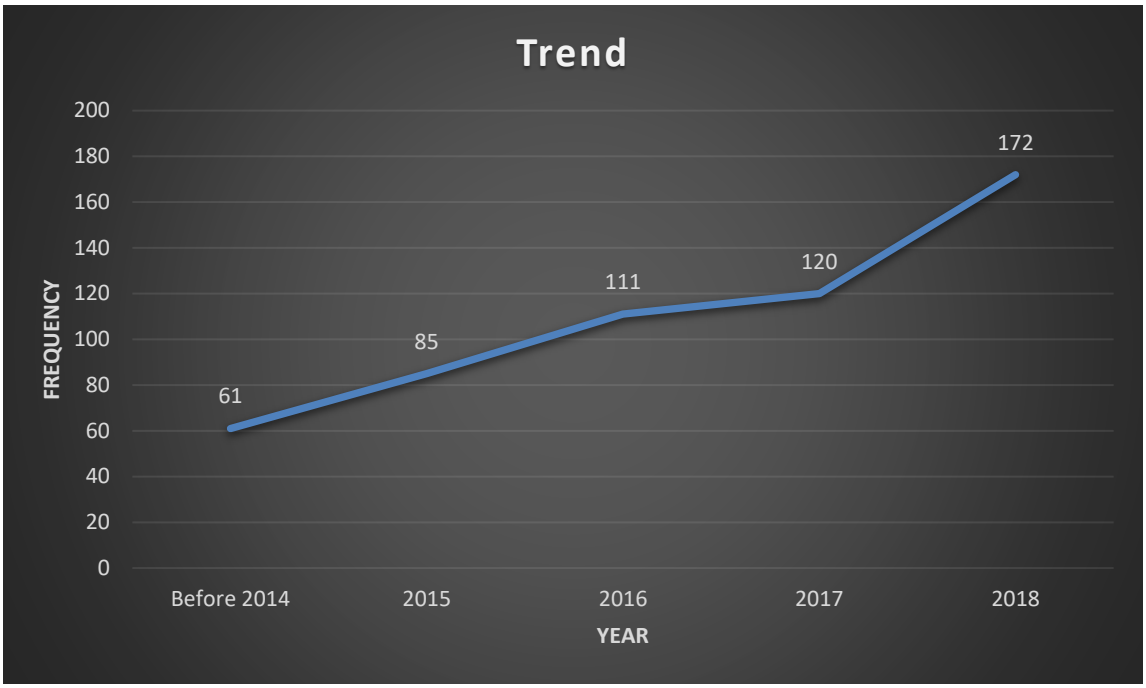
	Mobile Application	Mobile Website	Total	Percentage of Application	Percentage of Website
Easy to find it	404	39	443	91.20%	8.80%
More use to it	258	210	468	55.13%	44.87%
Convenience	422	312	734	57.49%	42.51%
Time saving	383	352	735	52.11%	47.89%
Availability	171	169	340	50.29%	49.71%
Shareability	5	9	14	35.71%	64.29%



Interpretation: We conclude that people usually use mobile application because it is easy to find, more use to it (habit), convenience, time saving factor and availability but when it comes to shareability ability people prefer mobile website.

- **Growth of E-wallets**

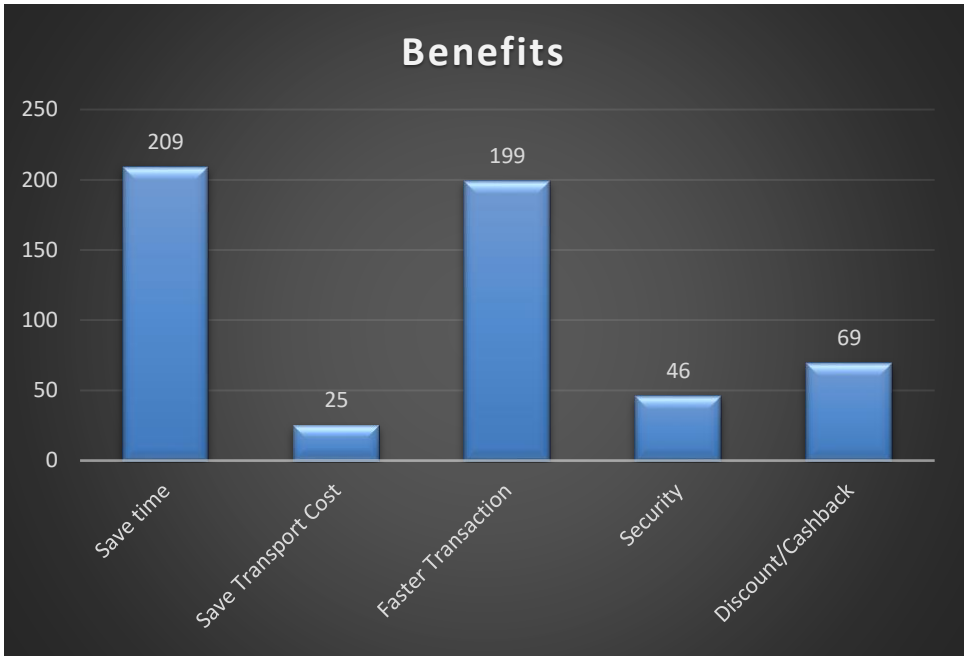
Year	Before 2014	2015	2016	2017	2018
Frequency	61	85	111	120	172



Interpretation: From the above figure result we can conclude that use e-wallets are increasing every year, it also has drastic increase in the year 2018 due to demonetization.

- **Benefits of E-wallet**

Benefits	Save time	Save Transport Cost	Faster Transaction	Security	Discount/Cashback
Frequency	209	25	199	46	29



Interpretation: Our survey shows that time saver and faster transaction is considered to be most important benefits in context of e-wallet application

Objective 1:

To study whether there is relationship between

i. Educational qualification v/s Awareness of govt. application.
ii. Education Qualification v/s Use of Lifestyle Application.
iii. Gender v/s Awareness of government application.
iv. Education Qualification v/s Do you use govt. application.

Cross Tabulation

Cross Tabulation table is the basic technique for examining between two categorical (nominal or ordinal) variables, possibly controlling for additional of variables. Cross tabs procedure offers several measures of and tests association. Additionally, you can obtain estimates of the relative risk of an event given the presence or absence of a characteristic. Number of tests are available to determine if the relationship between two cross tabulated variables is significant.

Pearson chi square tests: Pearson chi-squares used to test the independence of two attributes. A test of independence assesses whether paired observations on two attributes, expressed in a contingency table, independent of each other.

For the test of independence, a chi-square probability of less than or equal to 0.05 (or the chi square statistic being larger than the 0.05 critical point) is commonly interpreted by applied workers as justification for rejecting the null hypothesis that the row attribute is unrelated (that is, only randomly related) to the column attribute.

HYPOTHESIS:

Ho: The two attributes are independent of each other.

H1: The two attributes are dependent of each other

The first step in the chi-square test is to calculate the chi-square statistic. The chi-square statistic is calculated by finding the difference between each observed and theoretical frequency for each possible outcome, squaring them, dividing each by the theoretical frequency, and taking the sum of the results.

The test statistics is defined as:

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

X^2= Pearson's cumulative test statistic, which asymptotically approaches a X^2 Distribution.

O_i= the number of observations of type *i*.

n= total number of observations.

E_i = the expected (theoretical) frequency of type *i*, asserted by the null hypothesis.

The chi square statistic can then be used to calculate a p-value by comparing the value of the statistic to a chi square distribution. The number of degrees of freedom is equal to (k-1) *(r-1) where, k and r are the levels of two attributes.

1. Checking independence of 'Educational Qualification & Awareness of Government application

According to preliminary analysis, there are some people who are aware of Government Application, so we try to find out whether there is any association between Educational Qualification and Whether they are aware of Government application.

Attribute 1 = Educational Qualification

Attribute 2 =Awareness of Government application

To test:

H_0: There is no association in Educational Qualification & Awareness of Government application

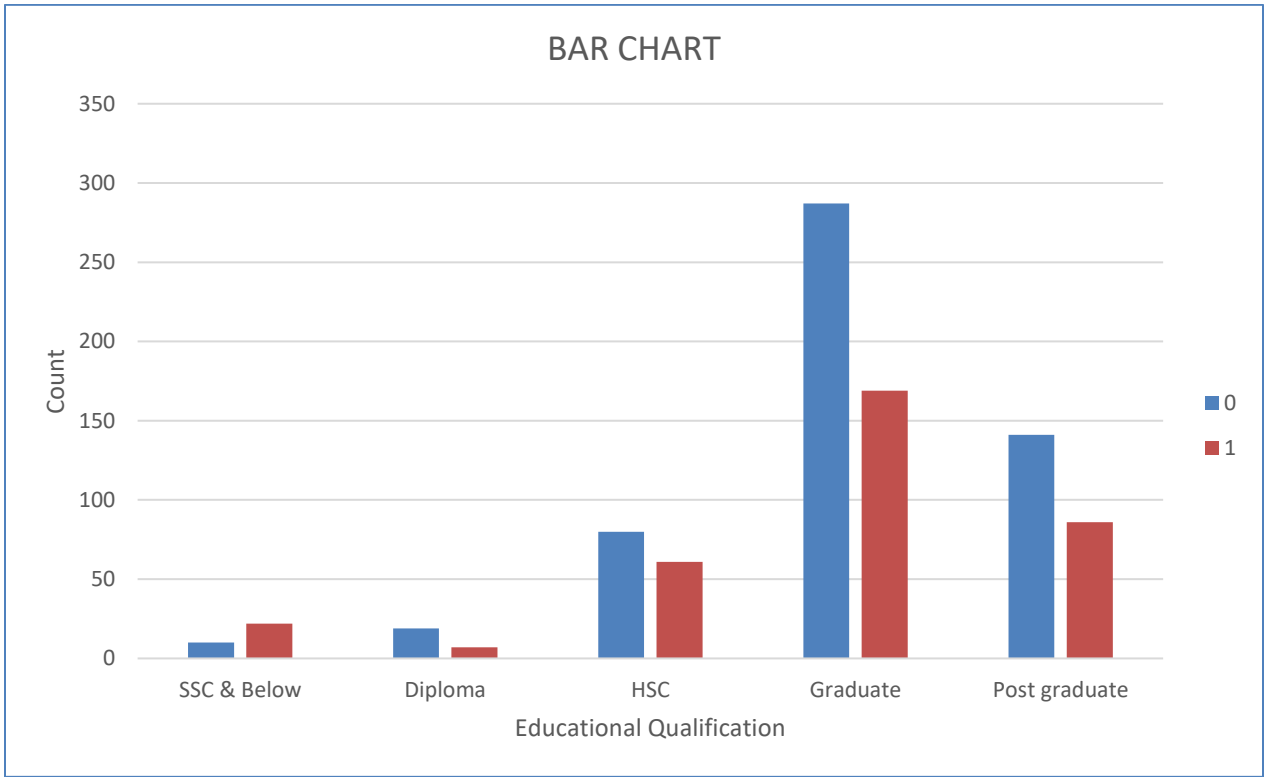
H_1: There is association in Educational Qualification & Awareness of Government application

Education Qualification * Are you aware of Government Applications? Cross tabulation

Count

		Are you aware of Government Applications?		Total
		Yes	No	
Education Qualification?	SSC & Below	10	22	32
	Diploma	19	7	26
	HSC	80	61	141
	Graduate	287	169	456
	Post-Graduate	141	86	227
Total		537	345	882

	Value	Df	Asymp. Sig. (2-sided)
Pearson Chi-Square	15.393[a]	4	.004



Interpretation: Since P-value (0.004) for Pearson Chi-Square test is less than alpha (0.05) hence we Reject H_0.

Conclusion: Educational Qualification and Awareness of Government application are associated with each other. According to the above result we come to know that more educated people are aware of government apps.

2) <u>Checking Independence of 'Gender & Use of Lifestyle application</u>

Attribute 1: Gender

Attribute 2: Use of lifestyle application

To test:

H_0: There is no association in Gender & Use of lifestyle application

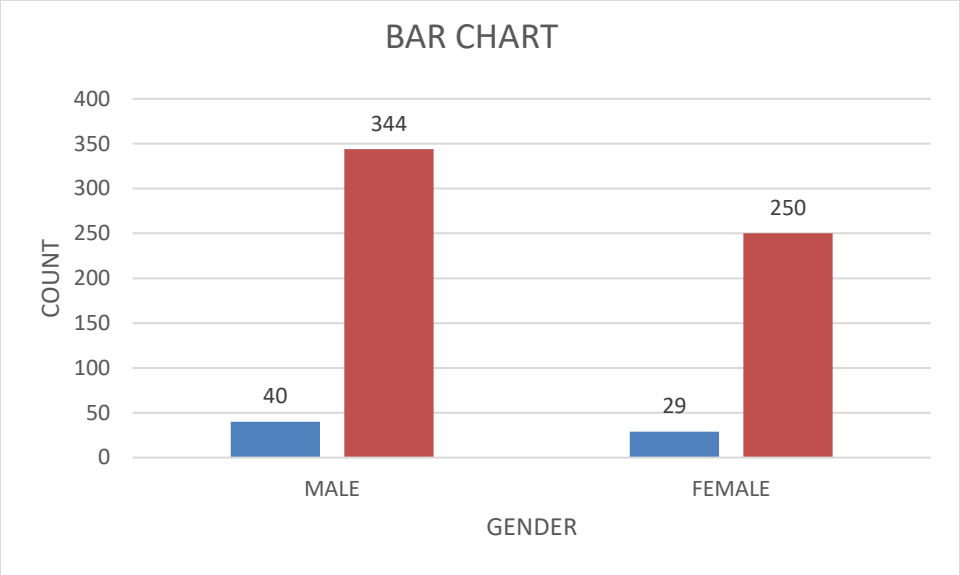
H_1: There is association in Gender & Use of lifestyle application

Gender * Do you use Lifestyles Apps Crosstabulation

Count

		Do you use Life styles Apps		Total
		No	Yes	
Gender	Male	40	344	384
	Female	29	250	279
Total		69	594	663

	Value	Df	Asymp. Sig. (2-sided)
Pearson Chi-Square	.000[a]	1	.993



Interpretation: Since P-value (0.993) for Pearson Chi-Square test is greater than alpha (0.05), hence we do not Reject Ho.

Conclusion: There is no association in Gender & frequent use of lifestyle application.

3) <u>Checking Independence of 'Gender & Awareness of govt. apps.</u>

Attribute 1: Gender.

Attribute 2: Awareness of govt app

To test:

H_0: There is no association in Gender and govt app

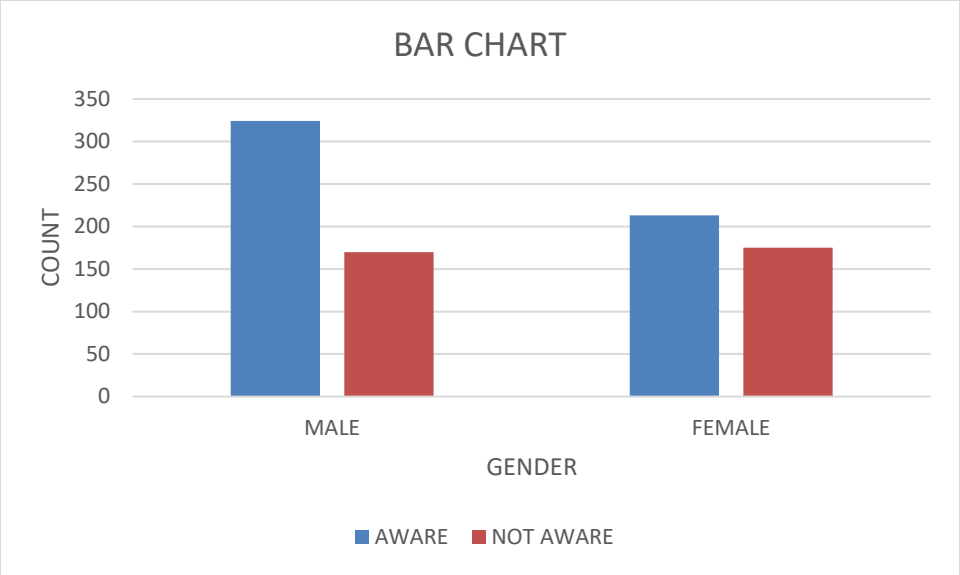
H_1: There is association in Gender and govt app

Gender * Are you aware of Government Applications? Crosstabulation

Count

		Are you aware of Government Applications?		Total
		Yes	No	
Gender	Male	324	170	494
	Female	213	175	388
Total		537	345	882

	Value	Df	Asymp. Sig. (2-sided)
Pearson Chi-Square	10.428[a]	1	.001



Interpretation: Since P-value (0.001) for Pearson Chi-Square test is less than alpha hence we Reject H0.

Conclusion: Gender is associated with awareness of govt app. This is because female respondents do not prefer to indulge more time on application because of which they are unaware of govt application. 66% of males are aware than that of females 52%

4) <u>Checking Independence of 'Educational Qualification and Do you use government apps or not'.</u>

Attribute 1: Educational Qualification

Attribute 2: Do you use government apps?

To test:

H_0: There is no association between educational qualification and usage of government apps.

H_1: There is association between educational qualification and usage of government apps.

Education Qualification? * Do you use Government app or not? Crosstabulation

Count

		Do you use Government App?		
		Yes	No	Total
Education Qualification?	SSC & Below	22	10	32
	Diploma	9	17	26
	HSC	82	59	141
	Graduate	208	248	456
	Post Graduate	117	110	227
Total		438	444	882

	Value	Df	Asymp. Sig. (2-sided)
Pearson Chi-Square	14.398[a]	4	.006

.

<u>Interpretation</u>: Since P-value (0.006) for Pearson Chi-Square test is less than alpha, we Reject H0.

<u>Conclusion:</u> There is an association between educational qualification and usage of government apps.

Objective 2:

To find out the reasons for downloading Mobile-Applications

WORD CLOUD

A word cloud is a graphical representation of frequently used words in
A collection of text files. The height of each word in a word cloud is an indication of frequency of occurrence of the word in the entire text. Such diagrams are very useful when doing text analytics.
Text mining methods allow us to highlight the most frequently used keywords in a paragraph of texts. One can create a word cloud, also referred as *text* cloud or tag cloud, which is a visual representation of text data.
We used word cloud technique to find out what are the factors that influences people choice to download an application.
We took a survey and got 883 responses out of which 664 respondents have responded to this particular question.
This is result we received:

Conclusion:

After applying the word cloud technique on our responses for text mining purposes, we can conclude that there are factors which people look forward while downloading an application:

1. Features that an application provides
2. Entertainment factor that an application provides is consider to be vital
3. The performance of an application is also an important aspect.

We have also noticed that people before downloading an application look at Ratings that application has and also compare with several application before downloading.

OBJECTIVE 3:

To identify and analyze the socio-demographic factors that affect the people's decision on whether to use e-wallet application.

Binary logistic regression

Logistic regression is part of a category of statistical models called generalized linear models.

Logistic regression is a predictive analysis, like linear regression. Logistic regression allows one to predict a discrete outcome from a set of variables that may be continuous, discrete and dichotomous or a mix of any of these. Here the dependent variable is dichotomous such as presence/ absence.

Binary logistic regression is a form of regression which is used when the dependent is a binary and the independents are of any type. Continuous variables are not used as dependents in logistic regression. Unlike logit regression, there can be only one dependent variable.

The goal of an analysis using logistic regression method is find the best fitting and most parsimonious, yet biologically reasonable model to describe the relationship between an outcome (dependent or response variable) and a set of independent (predictor or explanatory) variables and to determine the percent of variance in the dependent variable explained by the independents; to rank the relative importance of independents; to assess interaction effects; and to understand the impact of covariate control variables.

We used binary logistic regression since our dependent variable is categorical taking 2 values viz. using e-wallet application and not using e-wallet application.

Logistic regression is based on Maximum Likelihood (ML) Estimation which says coefficients should be chosen in such a way that it maximizes the Probability of Y given X (likelihood). With ML, the computer uses different "iterations" in which it tries different solutions until it gets the maximum likelihood estimates. **Fisher Scoring** is the most popular iterative method of estimating the regression parameters.

logit(p) = b0 + b1X1 + b2X2 + ------ + bk Xk

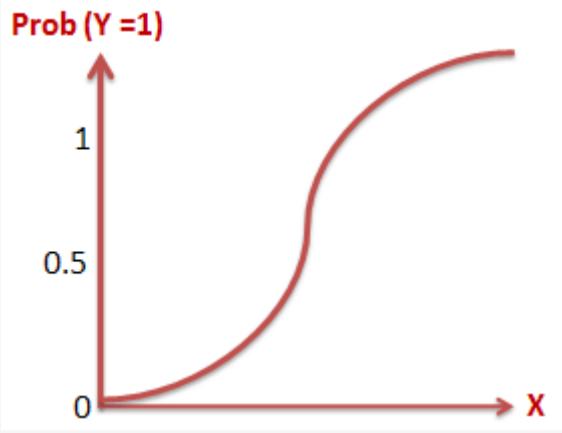
where logit(p) = loge (p / (1-p))

Take exponential both the sides

$$p = \frac{1}{1 + e^{-(b_0 + b_1 X_1 + b_2 X_2 + --- + b_k X_k)}}$$

Logistic Regression Equation

p: *the probability of the dependent variable equaling a "success" or "event".*



Logistic Regression Curve

Note that LHS of the model can lie between $-\infty$ *to* ∞

We have 6 independent variables, out of which 1 is continuous and 5 are categorical. The independent variables are Age, Gender, Education qualification, Professional status, Individual Monthly Income, Marital Status.

Assumption:

1. The logit transformation of the outcome variable has a linear relationship with the predictor variables. The one way to check the assumption is to categorize the independent variables. Transform the numeric variables to 10/20 groups and then check whether they have linear or monotonic relationship.

2. No multicollinearity problem. No high correlation between predictors.

3. No influential observations (Outliers).

4. Large Sample Size - It requires at least 10 events per independent variable.

Testing for multicollinearity

To test for multi-collinearity before we run the logistic regression model, we decided to run correlation of estimates on our chosen variables to make sure we only include those variables are actually relevant to our model analysis and ensure that bias remains minimum.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	.696	.097		7.146	.000		
	age	-.006	.003	-.100	-1.982	.048	.401	2.492
	gender	-.195	.032	-.200	-6.142	.000	.962	1.039
	education	.056	.017	.107	3.233	.001	.936	1.069
	status	.017	.006	.108	2.810	.005	.689	1.451
	income	.051	.015	.145	3.408	.001	.564	1.773
	marriage	-.030	.062	-.024	-.487	.626	.435	2.300

a. Dependent Variable: wallet

Since all the VIF values are less than 5 So, there is very little or no multi-collinearity in our selected variables.

Model building

Frequencies of dependent variables are given in following tables

The logistic Procedure

Model Information		
Data Set	WORK.S1	
Response Variable	wallet	wallet
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	882
Number of Observations Used	882

Response Profile		
Ordered Value	**wallet**	**Total Frequency**
1	1	549
2	0	333

Probability modeled is wallet=1.

DESIGN VARIABLES
The independent variables are converted to design variables. Six design variables are defined corresponding to the variable

Class Level Information								
Class	**Value**	**Design Variables**						
Gender	**1(Male)**	0						
	2(Female)	1						
Education qualification	**1(SSC and below)**	1	0	0	0			
	2(Diploma)	0	1	0	0			
	3(HSC)	0	0	1	0			
	4(Graduate)	0	0	0	0			
	5(Post Graduate & PhD)	0	0	0	1			
Professional Status	**1(Student)**	0	0	0	0	0	0	0
	2(Self-employed)	1	0	0	0	0	0	0
	3(Employed part-time)	0	1	0	0	0	0	0
	4(Businessman)	0	0	1	0	0	0	0
	5(Freelancer)	0	0	0	1	0	0	0
	6(Un-employed)	0	0	0	0	1	0	0
	7(Government Sector)	0	0	0	0	0	1	0
	8(Private Sector)	0	0	0	0	0	0	1
Individual Monthly Income	**1(Not earning)**	1	0	0	0	0		
	2(0-20 Thousand)	0	1	0	0	0		
	3(20-40 Thousand)	0	0	1	0	0		
	4(40-60 Thousand)	0	0	0	1	0		
	5(60-80 Thousand)	0	0	0	0	1		
	6(80 and above)	0	0	0	0	0		
Marriage	**1(Single)**	0						
	2(Married)	1						

Model

Here our dependent variable has categories:

Y = 0 not using e-wallet application

= 1 using e-wallet application

The logit function is:

$$y = \beta 0 + \beta 1X1 + \sum_{j=1}^{2} \beta\, 2jX2j + \sum_{j=1}^{4} \beta 3jX3j + \sum_{j=1}^{8} \beta 4jX4j + \sum_{j=1}^{6} \beta 5jX5j + \sum_{j=1}^{2} \beta 6jXj$$

Where,

$\beta 0 = Intercept$

$\beta 1 = Coefficient\ of\ Agevariable$

$\beta 2j = Coefficient\ of\ gender$

$\beta 3j = Coefficient\ of\ Educational\ qualification$

$\beta 4j = Coefficient\ of\ Professional\ status$

$\beta 5j = \ Coefficient\ of\ Individual\ monthly\ income$

$\beta 6j = Coefficient\ of\ marital\ status$

The conditional probabilities of each Outcome Category, given the covariate vectors are as follows:

$P_1 = P(Y=1|X) = e^{\left(\frac{g(x)}{1+e^{g(x)}}\right)}$

$P_2 = P(Y=0|X) = e^{\left(\frac{g(x)}{1+e^{g(x)}}\right)}$

Likelihood function

The likelihood function expresses the probability of the observed data as a function of the unknown parameters. The maximum likelihood estimators of these parameters are chosen to be those values which maximize this function.

The likelihood function: -

$$I(\beta)=\prod_{i=1}^{n} p^{yi}\,(1-p)^{1-yi}$$

Where,

If Y = 1 then $Y_1 = 1$, $Y_2 = 0$

If Y = 0 then $Y_1 = 0$, $Y_2 = 1$

The likelihood equation is found by taking the first partial derivative of

$L(\beta) = \ln l(\beta)$ with respect to each of the unknown parameters. The maximum likelihood estimator is obtained by setting these equations equal to zero and solving for β.

Global testing

Global testing is used to test whether or not, at least one of the independent variables influence the dependent variables i.e. at least one of the independent
variables insignificant.
H_0 = The Design variables entered into the model by stepwise procedure are insignificant
$\beta_1 = \beta_2 = \ldots = \beta_p = 0$
V/s
H_1 = The Design variables entered into the model by stepwise procedure are significant OR At least one coefficient is not zero.

Test Statistic:

$\chi^2 = L_1 - L_2$ which follows chi-square distribution with k df.

$L_1 = -2$ Log L with only constant term and no independent variables.

$L_2 = -2$ Log L with k independent variables and a constant term.

Where, L is the likelihood function.

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	42.4169	1	<.0001
Score	42.3572	1	<.0001
Wald	41.4880	1	<.0001

Conclusion:

Since p-value < 0.05, we reject H0.

Hence, at least one independent variable is significant

Stepwise selection procedure

Stepwise method is a process of building a model by successively adding or removing variables.

Any stepwise procedure for selection or deletion of variables from a model is based on statistical algorithm which check for the importance of variables, and either include or exclude them on the basis of fixed decision rule. The importance of the variable is defined in terms of a measure of the statistical significance is assumed via likelihood ratio chi-squared test. Thus, at any step in the procedure the most important variable, in statistical terms, will be the one that produces the greatest change in the log-likelihood relative to a model not containing the variable (i.e., the one that would result in the largest likelihood ratio statistic). For the stepwise selection procedure, we kept SLS and SLE at **5%**.

Note: No (additional) effects met the 0.05 significance level for entry into the model.

Summary of Stepwise Selection								
Step	Effect		DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Variable Label
	Entered	Removed						
1	Gender		1	1	42.3572		<.0001	gender
2	Income		5	2	33.8230		<.0001	income
3	Education		4	3	20.7022		0.0004	education
4	Age		1	4	9.3773		0.0022	age

Note: No (additional) effects met the 0.05 significance level for entry into the model.
As we can see from the summary of stepwise selection procedure, there were 4 variables extracted from the 6 we originally entered into the model. We can see that these 4 variables have p-value < 0.05 and so have significant impact on our dependent variable.

Residual Chi-square:

H_0: The reduced model is as good as full model.
H_1: The reduced model is not as good as full model.

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
6.9760	8	0.5392

Since the p-value>0.05, we do not reject H_o
Hence the reduced model is as good as full model and we with 4 variables given by step-wise procedure.

Wald Statistics (Individual Testing)

Hypothesis:
H_0: βi=0; i=1,2,3,5
H_1: not H0

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
Age	1	8.8985	0.0029
Gender	1	34.7231	<.0001
Education Qualification	4	17.4775	0.0016
Individual monthly income	5	32.5866	<.0001

From the table of analysis of effects, the p-value of all the variables is less than 0.05.
Thus, the variables Age, Gender, Education qualification and Individual monthly Income are significant at 5% los.

Parameter estimation and individual testing

The parameter (logit) estimates are nothing but the M.L.E. estimates obtained by partial differentiation of the natural log of likelihood function with respect to each of the unknown parameters and equating the resultant equations to zero. Iterative method is used for computing the estimates. The standard interpretation of the multinomial logit is that for a unit change in the predictor change, the logit of the outcome relative to the reference group is expected to change by its respective parameter (which is in log-odds unit) given the other variables in the models are held constant

Intercept: This is the multinomial logit estimates when the predictor variables in the model are evaluated at zero.
To test the hypotheses:
H_0 = Individual coefficients of independent variables are zero OR βi=0
H_1 = Individual coefficients of independent variables are not zero OR βi≠0
To test the above hypotheses, Wald's Statistic is used. It is defined as the ratio of estimated coefficient to its estimated standard error.

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	2.4466	0.5739	18.1739	<.0001
age		1	-0.0327	0.0110	8.8985	0.0029
gender	**2**	1	-0.8793	0.1492	34.7231	<.0001
education_	**1**	1	-1.0655	0.4202	6.4305	0.0112
education_	**2**	1	0.2941	0.4690	0.3932	0.5306
education_	**3**	1	-0.5260	0.2062	6.5062	0.0107
education_	**5**	1	0.2294	0.1832	1.5674	0.2106
income	**1**	1	-1.1082	0.4380	6.4010	0.0114
income	**2**	1	-0.4121	0.4440	0.8614	0.3533
income	**3**	1	-0.2251	0.4425	0.2586	0.6111
income	**4**	1	0.4637	0.5174	0.8034	0.3701
income	**5**	1	0.0455	0.5744	0.0063	0.9368

The variables – Education Qualification (2,5) i.e. diploma and post-graduation/PhD are insignificant. And almost all individual monthly income except not earning category are insignificant.

Goodness of fit

With logistic regression, instead of R2 as the statistics for overall fit of the linear regression model, deviance between observed values from the expected values is used. In linear regression, residuals can be defined as yi−yi(hat). Where i is the observed dependent variable for the ith subject, and yi(hat) the corresponding prediction from the model. The same concept applied to logistic regression, where yi is equal to either 1 or 0.

Hosmer & Lemeshow test:

The Hosmer–Lemeshow test is a statistical test for goodness of fit for logistic regression models. It is used frequently in risk prediction models. The test assesses whether or not the observed event rates match expected event rates in subgroups of the model population. The Hosmer–Lemeshow test specifically identifies subgroups as the deciles of fitted risk values. Models for which expected and observed event rates in subgroups are similar are called well calibrated. The Hosmer-Lemeshow statistic evaluates the goodness of fit by creating ordered groups of subjects and then comparing the number actually in each group (observed) to the number predicted by the logistic regression model (predicted). The statistic used is a chi-square statistic with desirable outcome of no significance, indicating the model prediction does not significantly differ from the observed.

Hypothesis:
H_0 = Model is a good fit for the data

H_1 = Model is not a good fit for the data

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	**DF**	**Pr > ChiSq**
10.5343	8	0.2295

Since p-value > 0.05, we fail to reject H0 and conclude that our model is a good fit

Fitted model:
$g(x) = 2.4466 – 0.0327(x_{11}) – 0.8793(x_{21}) - 1.0655(x_{31}) – 0.526(x_{33}) - 1.1082(x_{41})$
Where
x_{11}= Age
x_{21}= Gender
x_{31}= Education qualification (SSC and below)
x_{33}= Education qualification (HSC)
x_{41}= Income Not earning

Odds Ratio:

Odds ratio is a measure of association. It approximates how much more likely it is for outcome to be present among the different levels of independent variables.

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
Age	0.968	0.947	0.989
Female vs Male	0.415	0.310	0.556
SSC and below vs Graduate	0.345	0.151	0.785
Diploma vs Graduate	1.342	0.535	3.365
HSC vs Graduate	0.591	0.394	0.885
Post graduate and PhD vs Graduate	1.258	0.878	1.801
Not earning vs 80 thousand and above	0.330	0.140	0.779
0-20 thousand vs 80 thousand and above	0.662	0.277	1.581
20-40 thousand vs 80 thousand and above	0.798	0.335	1.901
40-60 thousand vs 80 thousand and above	1.590	0.577	4.383
60-80 thousand vs 80 thousand and above	1.047	0.339	3.226

Conclusion

1. **Gender:** Males are 0.415 times less likely to use e-wallet applications than female.
2. **Education Qualification:**
   a. Graduates are 0.345 times less likely to use e-wallet applications as compared people who have studied up to SSC or below.
   b. Graduates are 1.342 times more likely to use e-wallet applications as compared people who have studied up to diploma.
   c. Graduates are 0.591 times less likely to use e-wallet applications as compared to people who have studied till HSC.
   d. Graduates are 1.258 times more likely to use e-wallet application as compared to people who are have done post-graduation/PhD.
3. **Individual Monthly Income:**
   a. People who earn 80 thousand and above per month are 0.330 times less likely to use e-wallet application as compared to people who are not earning.
   b. People who earn 80 thousand and above per month are 0.662 times less likely to use e-wallet application as compared to people who earn 0-20 thousand per month.
   c. People who earn 80 thousand and above per month are 0.798 times less likely to use e-wallet application as compared to people who earn 20-40 thousand per month.
   d. People who earn 80 thousand and above per month are 1.59 times more likely to use e-wallet application as compared to people who earn 40-60 thousand per month.
   e. People who earn 80 thousand and above per month are 1.047 times more likely to use e-wallet application as compared to people who earn 60-80 thousand per month.

Classification Table:
The classification table is a cross-tabulation of observed and predicted frequencies for the dependent values of Y. These observed and predicted values y values are cross tabulated to get the classification table as follows

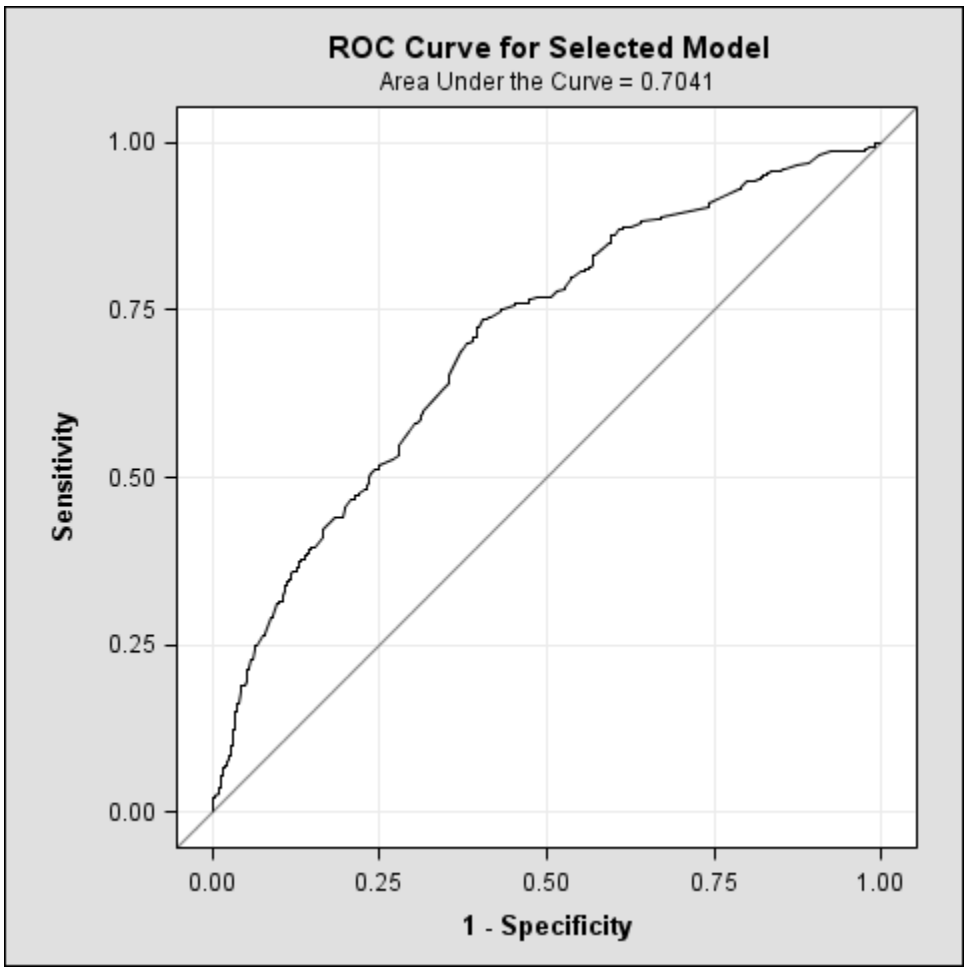
Classification Table									
Prob Level	Correct		Incorrect		Percentages				
	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.500	448	142	191	101	66.9	81.6	42.6	29.9	41.6

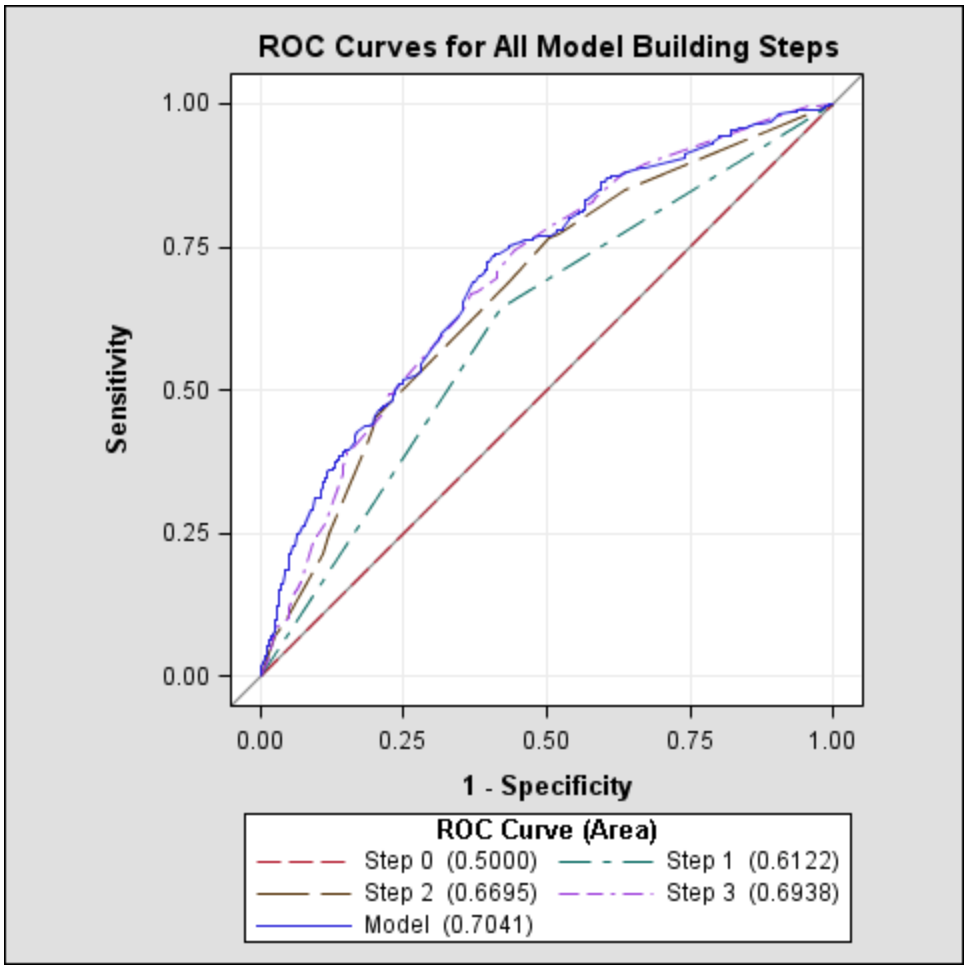
The table shows that 590 out of 882 (i.e. 66.9%) matches are correct. Hence, we can conclude that the model predicts the values correctly 66.9% of the times.

Roc curve
Receiver Operating Characteristic i.e. ROC curve is used to evaluate and compare the performance of diagnostic tests. They can also be used to evaluate model fit. A ROC curve is just a plot of proportion of true positives (events predicted to be events i.e. Sensitivity) versus the proportion of false positives (non-events predicted to be events i.e. Specificity). The accuracy of test is measured by the area under the ROC curve. An area of 1 represent a perfect test, while an area of 0.5 represents a worthless test. The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test, the true positive rate (sensitivity) is high and the false positive rate (1-specificity) is low. Statistically, more area under the curve means that it is identifying more true positives while minimizing the number/ percent of false positives.

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	70.0	Somers' D	0.408
Percent Discordant	29.2	Gamma	0.411
Percent Tied	0.8	Tau-a	0.192
Pairs	182817	c	0.704

Area under the ROC curve is estimated by the statistic c in the "Association of predicted Probabilities and Observed Responses" table. Hence, the areas under the ROC curve is 0.704.



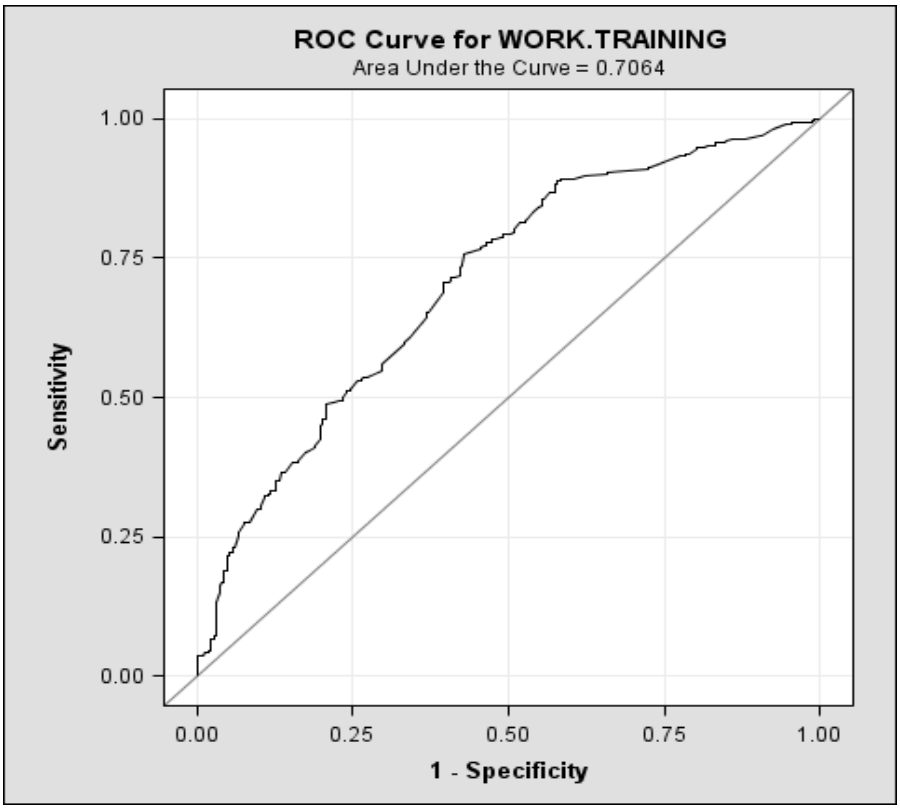

ROC Curves for All Model Building Steps
Sensitivity
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
1 - Specificity
ROC Curve (Area)
Step 0 (0.5000)
Step 1 (0.6122)
Step 2 (0.6695)
Step 3 (0.6938)
Model (0.7041)

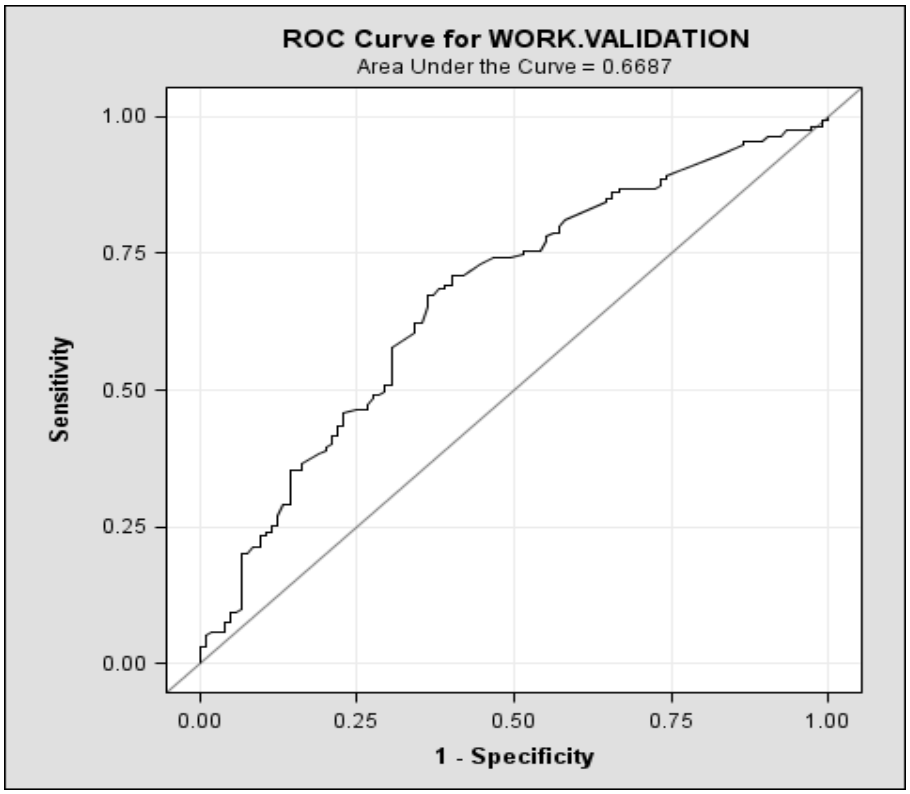
Training and validation:
We have used 70% of responses for training and 30% for validation

Number of Observations Read	618
Number of Observations Used	618

Response Profile		
Ordered Value	**wallet**	**Total Frequency**
1	1	390
2	0	228

Fit Statistics for SCORE Data				
Data Set	**Total Frequency**	**Log Likelihood**	**Misclassification Rate**	**Area Under the ROC Curve**
WORK.TRAINING	618	-365.8	0.2880	0.7064
WORK.VALIDATION	264	-169.9	0.3447	0.6687





Here we conclude that Area Under the ROC for training data is 0.7064 & Area Under the ROC for validation data is 0.6687

Kolmogorov-Smirnov Statistics:
Kolmogorov-Smirnov Statistics is one of the commonly used measures to assess predictive-power. Kolmogorov-Smirnov is the maximum difference between the cumulative true positive and cumulative false positive rate.

Kolmogorov-Smirnov Two-Sample Test (Asymptotic)			
KS	0.152237	**D**	0.311051
KSa	2.473564	**Pr > KSa**	<.0001

KS=0.311051 > 0.25, it indicates model is good.

Interaction Term
Now to check if interaction term were significant, we included all interaction term in our model and then we used Backward elimination method and concluded that Interaction term Age*Income comes out to be significant. so, we include these term in our fitted model.

Objective 4

To check if demonetization has effect on usage of e-wallet application.

Wilcoxon signed rank test

The Wilcoxon signed-rank test is a non-parametric test that can be used to determine whether two dependent samples are selected from populations having the same distribution. This test uses more of the information for the case of two related samples when measurement scale allows us to determine the relative magnitude of the difference of pairs of observations. After determining the magnitude of differences, we can rank them and through this ranking, the test utilizes more or additional information. It can be used as an alternative to the paired Student's t-test (t-test for matched pairs).

Assumptions:

1. Data are paired and come from the same population.

2. Each pair is chosen randomly and independently.

3. The data are measured on at least an interval scale when, as is usual, within-pair difference calculated to perform the test. (though it does suffice that within-pair comparisons are on an ordinal scale).

Hypothesis:

H_0: Difference between the pairs follows a symmetric distribution around zero.

V/S

H_1: Difference between the pairs does not follow a symmetric distribution around zero.

Procedure:

Let N be the sample size, i.e., the number of pairs. Thus, there are a total of 2N data points. For pairs i = 1, ..., N, let X_{1i} and X_{2i} denote the measurements.

{In our case - X_1: Pre-ban and X_2: Post-ban}

1. For i = 1, ..., N, calculate $|X_{2i} - X_{1i}|$ and sgn $(X_{2i} - X_{1i})$, where 'sgn' is the sign function.
2. Exclude pairs with $| X_{2i} - X_{1i}| = 0$. Let N_r be the reduced sample size.
3. Order the remaining N_r pairs from smallest absolute difference to largest absolute difference, $|X_{2i} - X_{1i}|$.
4. Rank the pairs, starting with the smallest as 1. Ties receive a rank equal to the average of the ranks they span. Let R_i denote the rank.
5. Calculate the test statistic W

   $$W = \sum_{i=1}^{N_r} \{sgn(X_{2i} - X_{1i}) * R_i\}$$

   Where, W=the sum of the signed ranks.
6. Under null hypothesis, W follows a specific distribution with no simple expression.
   This distribution has an expected value of 0 and variance of $\frac{N_r(N_r+1)(2N_r+1)}{6}$.
   W can be compared to a critical value from a reference table.

The two-sided test consists in rejecting H_0

if $|W| > W_{critical, Nr}$

Analysis

Case 1: To check if demonetization has effect on mobile recharges done using e-wallet applications.

H_{01}: There is no significance difference between mobile recharges done using e-wallet application before and after demonetization.

V/s

H_{11}: Not H01.

Wilcoxon signed rank test

data: v1 and v2

p-value = 1.084e-09

alternative hypothesis: true location shift is not equal to 0

95 percent confidence interval:

-1.0000217 -0.9999746

sample estimates:

(pseudo)median

-1.000039

Interpretation: Since p-value < 0.05, we reject H_{01} and conclude that there is significant significance difference between mobile recharges done using e-wallet application before and after demonetization.

In other words, there is a positive effect of demonetization on doing mobile recharges using e-wallet applications i.e. after demonetization usage of e-wallet application for mobile recharges purpose has increased.

Case 2: To check if demonetization has effect on DTH recharges done using e-wallet applications.

H_{01}: There is no significance difference between DTH recharges done using e-wallet application before and after demonetization.

V/s

H_{11}: Not H01.

Wilcoxon signed rank test

data: v1 and v2

p-value = 1.03e-05

alternative hypothesis: true location shift is not equal to 0

95 percent confidence interval:

-1.4999850 -0.9999878

sample estimates:

(pseudo)median

-1.00005

Interpretation: Since p-value < 0.05, we reject H_{01} and conclude that there is significant significance difference between DTH recharges done using e-wallet application before and after demonetization.

In other words, there is a positive effect of demonetization on doing DTH recharges using e-wallet applications i.e. after demonetization usage of e-wallet application for mobile recharges purpose has increased.

Case 3: To check if demonetization has effect on payment of electricity bill using e-wallet applications.

H_{01}: There is no significance difference between payment of electricity bill done using e-wallet application before and after demonetization.

V/s

H_{11}: Not H01.

Wilcoxon signed rank test

data: v1 and v2

p-value = 2.402e-08

alternative hypothesis: true location shift is not equal to 0

95 percent confidence interval:

-1.500063 -1.000007

sample estimates:

(pseudo)median

-1.499928

Interpretation: Since p-value < 0.05, we reject H_{01} and conclude that there is significant significance difference between payment of electricity bill done using e-wallet application before and after demonetization.

In other words, there is a positive effect of demonetization on paying electricity bill using e-wallet applications i.e. after demonetization usage of e-wallet application for payment of electricity bill has increased.

Case 4: To check if demonetization has effect on shopping bills payment via e-wallet applications.

H_{01}: There is no significance difference between shopping bills payment done using e-wallet application before and after demonetization.

V/s

H_{11}: Not H01.

Wilcoxon signed rank test

data: v1 and v2

p-value = 1.551e-05

alternative hypothesis: true location shift is not equal to 0

95 percent confidence interval:

-1.0000376 -0.9999862

sample estimates:

(pseudo)median

-0.9999364

Interpretation: Since p-value < 0.05, we reject H_{01} and conclude that there is significant significance difference between shopping bills payment done using e-wallet application before and after demonetization.

In other words, there is a positive effect of demonetization on paying shopping bills using e-wallet applications i.e. after demonetization usage of e-wallet application for shopping bills payments has increased.

Case 5: To check if demonetization has effect on buying movie tickets via e-wallet applications.

H_{01}: There is no significance difference between buying movie tickets using e-wallet application before and after demonetization.

V/s

H_{11}: Not H01.

Wilcoxon signed rank test

data: v1 and v2

p-value = 6.548e-06

alternative hypothesis: true location shift is not equal to 0

95 percent confidence interval:

-1.0000740 -0.4999816

sample estimates:

(pseudo)median

-0.9999737

Interpretation: Since p-value < 0.05, we reject H_{01} and conclude that there is significant significance difference between buying movie tickets using e-wallet application before and after demonetization.

In other words, there is a positive effect of demonetization on buying movie tickets using e-wallet applications i.e. after demonetization usage of e-wallet application for buying movie tickets has increased.

Case 6: To check if demonetization has effect on money transfer done via e-wallet applications.

H_{01}: There is no significance difference between money transfer done using e-wallet application before and after demonetization.

V/s

H_{11}: Not H01.

Wilcoxon signed rank test

data: v1 and v2

p-value = 7.686e-07

alternative hypothesis: true location shift is not equal to 0

95 percent confidence interval:

-1.5000612 -0.9999675

sample estimates:

(pseudo)median

-1.499959

Interpretation: Since p-value < 0.05, we reject H_{01} and conclude that there is significant difference between money transfer done using e-wallet application before and after demonetization.

In other words, there is a positive effect of demonetization on doing money transfer using e-wallet applications i.e. after demonetization usage of e-wallet application for money transfer purpose has increased.

Objective 5:

To find out the most preferred E-Wallet/Application.

Factor analysis

Factor analysis is a technique that is used to reduce a large number of variables into fewer numbers of factors. This technique extracts maximum common variance from all variables and puts them into a common score. As an index of all variables, we can use this score for further analysis. Factor analysis is part of general linear model (GLM) and this method also assumes several assumptions: there is linear relationship, there is no perfect multicollinearity, it includes relevant variables into analysis and there is true correlation between variables and factors. Several methods are available, but principle component analysis is used most commonly.

Key concepts and terms:

Principal component analysis: This is the most common method used by researchers. PCA starts extracting the maximum variance and puts them into the first factor. After that, it removes that variance explained by the first factors and then starts extracting maximum variance for the second factor. This process goes to the last factor.

Correlation matrix: A correlation matrix is a lower triangle showing the sample correlations, 'r ' between all possible pairs of variables included in the analysis.

Communality: Communality is the amount of variance a variable share with all the other variables being considered. This is also the proportion of variance explained by common factors.

Factor loading: Factor loading is basically the correlation coefficient for the variable and factor. Factor loading shows the variance explained by the variable on that particular factor. In the SEM approach, as a rule of thumb, 0.7 or higher factor loading represents that the factor extracts sufficient variance from that variable.

Eigen-values: Eigen values are also called characteristic roots. Eigen values shows variance explained by that particular factor out of the total variance. From the commonality column, we can know how much variance is explained by the first factor out of the total variance. For example, if our first factor explains 68% variance out of the total, this means that 32% variance will be explained by the other factor.

Factor score: The factor score is also called the component score. This score is of all row and columns, which can be used as an index of all variables and can be used for further analysis. We can standardize this score by multiplying a common term. With this factor score, whatever analysis we will do, we will assume that all variables will behave as factor scores and will move.

Criteria for determining the number of factors: According to the Kaiser Criterion, Eigen values are good criteria for determining a factor. If Eigen values is greater than one, we should consider that a factor and if Eigen values is less than one, then we should not consider that a factor. According to the variance extraction rule, it should be more than 0.6. If variance is less than 0.6, then we should not consider that a factor.

Rotation method: Rotation method makes it more reliable to understand the output. Eigen values do not affect the rotation method, but the rotation method affects the Eigen values or percentage of variance extracted. There are a number of rotation methods available:

(1) Varimax rotation method (2) Quartimax rotation method (3) Equamax rotation method (4) Direct oblimin rotation method (5) Promax rotation method.

Each of these can be easily selected in SPSS, and we can compare our variance explained by those particular methods.

Varimax rotation Method: Change of coordinates used in principal component analysis (PCA) is known as Varimax rotation. It maximizes the sum of the variances of the squared loadings as all the coefficients will be either large or near zero, with few intermediate values. The goal is to associate each variable to at most one factor.

Assumptions:

1. No outlier: Assume that there are no outliers in data.
2. Adequate sample size: The case must be greater than the factor.
3. Homoscedasticity: Since factor analysis is a linear function of measured variables, it does not require homoscedasticity between the variables.
4. Linearity: Factor analysis is also based on linearity assumption. Non-linear variables can also be used. After transfer, however, it changes into linear variable.
5. Interval Data: Interval data are assumed.

Output:
Variables used in Factor Analysis:
X_1: BHIM
X_2: IRCTC
X_3: MyGov
X_4: Online-seva-Indian digital service
X_5: Swach Bharat
X_6: Voter-Apps
X_7: Online RTI
X_8: GST-Rate finder
X_9: M-Passport Seva
X_{10}: UMANG
X_{11}: Paytm
X_{12}: Freecharge
X13: Mobikwik
X14: PhonePe
X15: Google Pay

Before proceeding with factor analysis on the variables we need to check whether factor analysis is appropriate for our data and are the variables correlated with each other which are the basic assumption for factor analysis.

KMO and Bartlett's test of sphericity:

The Kaiser-Meyer-Olkin measure of Sampling Adequacy is a statistic that indicates the proportion of variance in your variables that might be caused by underlying factors. High values (close to 1 .0) generally indicate that a factor analysis may be useful with your data. If the value is less than 0.50, the results of the factor analysis probably won't be very useful.

Bartlett's test of sphericity tests the hypothesis that your correlation matrix is an identity matrix, which would indicate that your variables are: unrelated and therefore unsuitable for structure detection. Small values (less than 0.05) of the significance level indicate that a factor analysis may be useful with your data.

KMO and Bartlett's Test	
Kaiser-Meyer-Olkin Measure of Sampling Adequacy	0.884
Bartlett's Test of Sphericity (Sig.)	0.000

From the above table we can see that KMO value > 0.5 which shows that factor analysis is useful for our data.
From the Bartlett's test of sphericity, the p-value is less than 0.05 we can conclude that population correlation matrix is not an identity matrix.

Correlation Matrix

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15
	1.000	.600	.638	.601	.490	.538	.565	.611	.551	.616	.031	-.043	-.001	.024	-.088
X1	.600	1.000	.633	.670	.494	.525	.572	.521	.553	.522	.022	.125	.052	.027	.000
X2	.638	.633	1.000	.805	.697	.713	.689	.640	.562	.642	-.041	.042	.033	-.004	-.135
X3	.601	.670	.805	1.000	.717	.709	.782	.716	.688	.730	.055	.106	.101	.111	-.037
X4	.490	.494	.697	.717	1.000	.842	.800	.745	.682	.722	-.041	.079	.072	.049	-.134
X5	.538	.525	.713	.709	.842	1.000	.831	.735	.713	.716	-.091	-.001	-.008	-.004	-.102
X6	.565	.572	.689	.782	.800	.831	1.000	.861	.833	.817	-.074	.027	.007	.016	-.148
X7	.611	.521	.640	.716	.745	.735	.861	1.000	.801	.837	-.059	.010	-.023	-.009	-.141
X7	.551	.553	.562	.688	.682	.713	.833	.801	1.000	.803	.042	.138	.078	.045	-.064
X8	.616	.522	.642	.730	.722	.716	.817	.837	.803	1.000	.033	.063	.045	.093	-.079
X9	.031	.022	-.041	.055	-.041	-.091	-.074	-.059	.042	.033	1.000	.414	.494	.609	.579
X10	-.043	.125	.042	.106	.079	-.001	.027	.010	.138	.063	.414	1.000	.821	.450	.224
X11	-.001	.052	.033	.101	.072	-.008	.007	-.023	.078	.045	.494	.821	1.000	.541	.364
X12	.024	.027	-.004	.111	.049	-.004	.016	-.009	.045	.093	.609	.450	.541	1.000	.633
X13	-.088	.000	-.135	-.037	-.134	-.102	-.148	-.141	-.064	-.079	.579	.224	.364	.633	1.000
X14															
X15															

Total Variance Explained

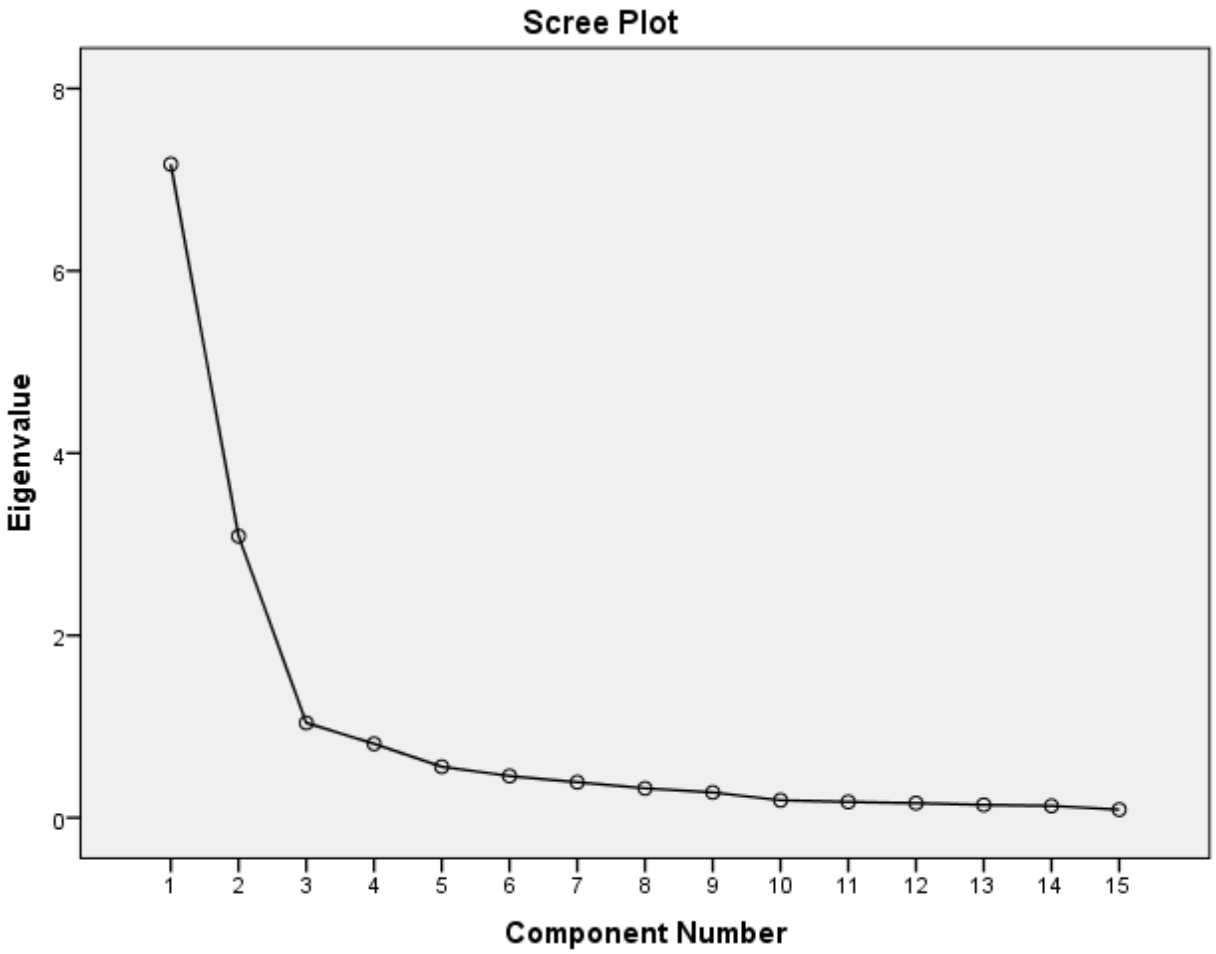
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	7.170	47.801	47.801	7.170	47.801	47.801	7.152	47.682	47.682
2	3.089	20.595	68.396	3.089	20.595	68.396	2.290	15.266	62.948
3	1.040	6.936	75.332	1.040	6.936	75.332	1.857	12.383	75.332
4	.811	5.407	80.738						
5	.559	3.728	84.466						
6	.458	3.051	87.517						
7	.390	2.597	90.114						
8	.322	2.148	92.262						
9	.277	1.847	94.109						
10	.192	1.277	95.386						
11	.173	1.153	96.539						
12	.159	1.063	97.601						
13	.140	.936	98.538						
14	.130	.865	99.403						
15	.090	.597	100.000						

Extraction Method: Principal Component Analysis.

Rotated Component Matrix[a]

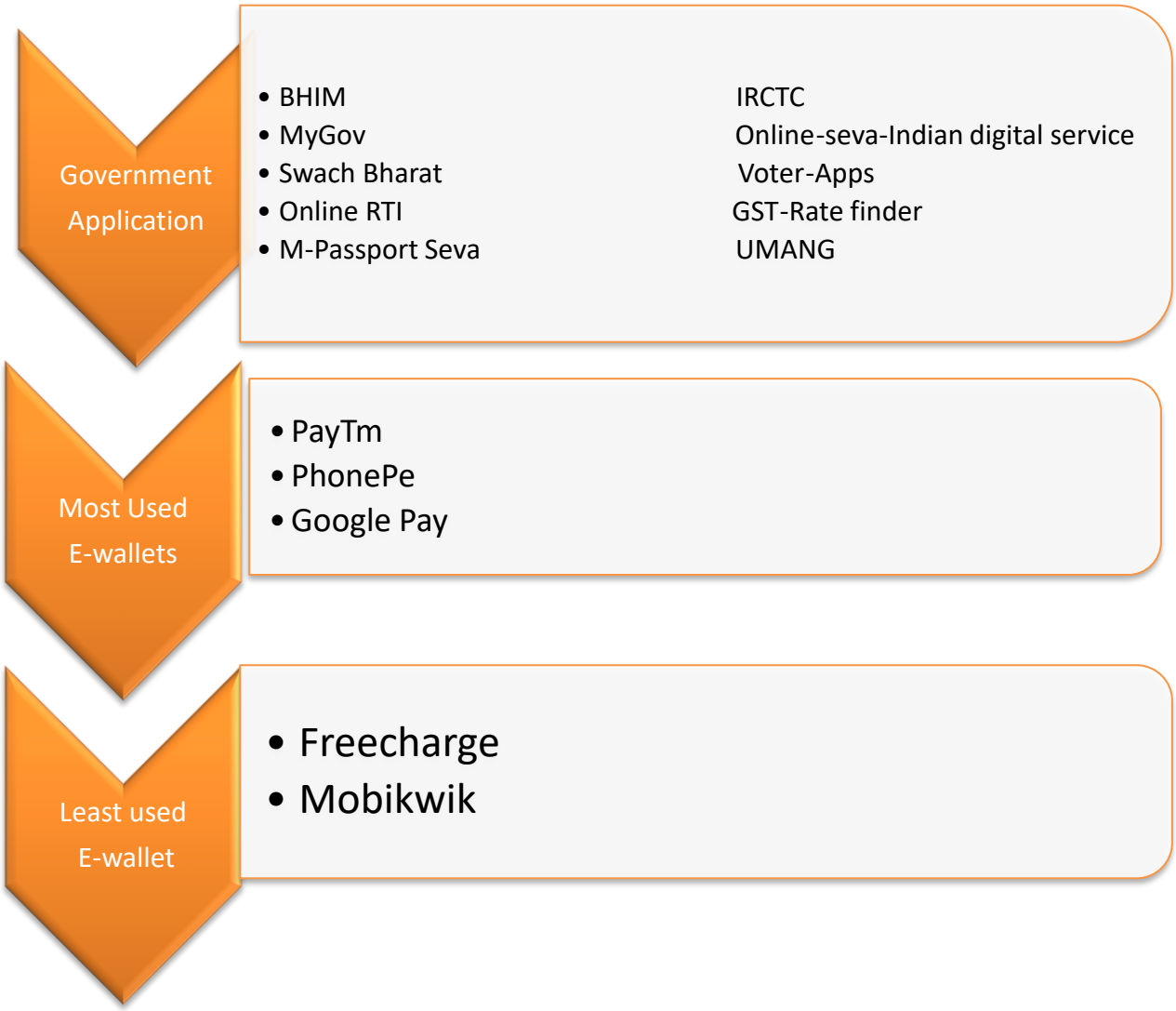
	Component		
	1	2	3
BHIM	.733		
IRCTC	.710		
My Gov	.828		
Online Seva	.880		
swach Bharat	.853		
voter apps	.871		
Online RTI	.923		
GST	.890		
passport	.853		
UMANG	.883		
Paytm		.795	
Mobikwik			.934
Freecharge			.869
PhonePay		.801	
GooglePay		.891	

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 4 iterations.



From the above scree plot we can see an elbow shape at 3rd component, thus 3 components should be extracted.

On the basis of factor loadings obtained from the matrix we can classify the variables into 3 factors as follows:



Objective 6:

To obtain an overall review on

i. Types of different websites
ii. Types of applications (Paytm, Phonepe, etc.)
iii. Government Application

PARETO ANALYSIS

Pareto Analysis is a statistical technique in decision making that is used for the selection of a limited number of tasks that produce significant overall effect. It uses the Pareto Principle. It is also known as the 80/20 rule. The idea is that by doing 20% of the work, you can generate 80% of the benefit of doing the whole job. This is also known as the "vital few" and the "trivial many" effect.

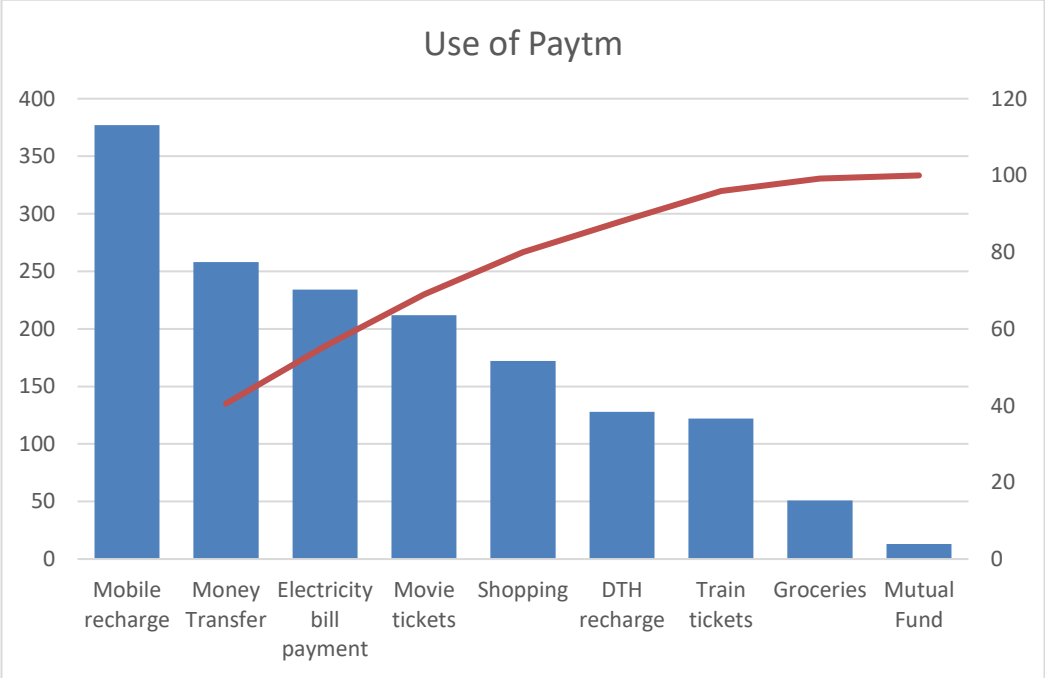
The Pareto principle has many applications in quality control. It is the basis of the Pareto diagram, one of the key tools used in total quality control and Six Sigma.

A Pareto chart is used to graphically summarize and display the relative importance of the differences between groups of data. The Pareto chart is a very simple but effective tool for prioritizing problem causes, which is why it is widely used for problem solving in the manufacturing industry The Pareto Chart is basically a descending bar graph that shows the frequencies of occurrences or relative sizes of various problems or causes of a particular problem. The problem categories or causes are shown on x-axis of the bar graph. Aside from its main bar graph, the Pareto chart may also include a line graph that indicates the cumulative percentage of occurrences at each bar of the graph. This line graph referred to as the "cumulative percentage line", is used to determine which of the bars belong to the 'vital few ' and which ones are relegated to the 'trivial many'.

So, we use pareto analysis for our most used e-wallet and find out for what purpose these e-wallet are used.

Pareto For Use of Paytm

Use	Frequency
Mobile Recharge	377
Money Transfer	258
Electricity bill	234
Movie Tickets	212
Shopping	172
DTH Recharge	128
Train Tickets	122
Groceries	51
Mutual Fund	13

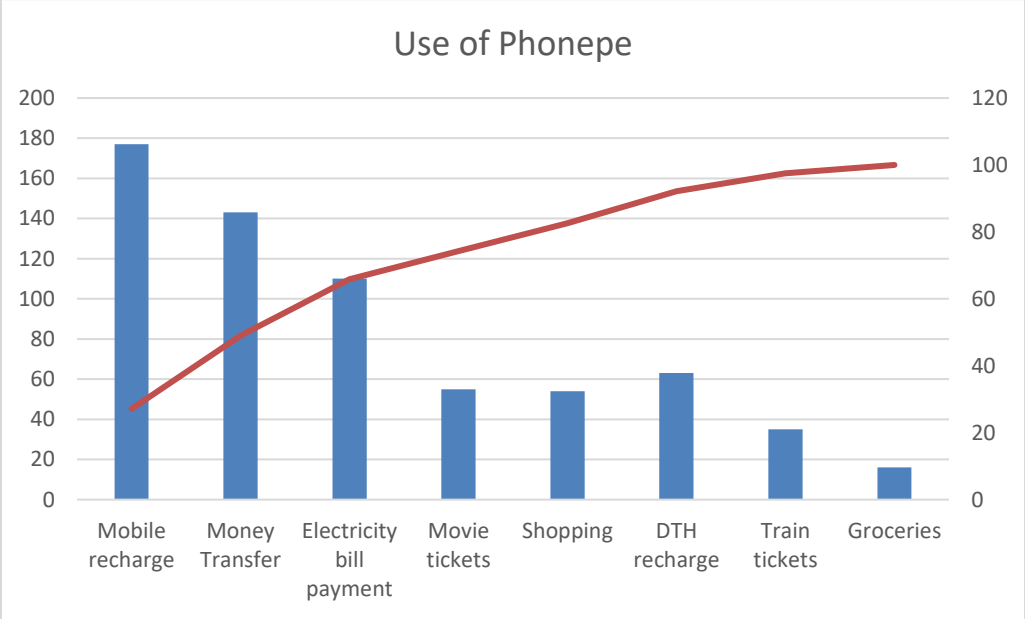


Interpretation:

Our analysis shows that, Paytm is most frequently used for Mobile Recharge, Money transfer, Electricity bill payments, Movie tickets.

Pareto for use of Phonepe

Use	Frequency
Mobile Recharge	177
Money Transfer	143
Electricity bill	110
Movie Tickets	55
Shopping	54
DTH Recharge	63
Train Tickets	35
Groceries	16

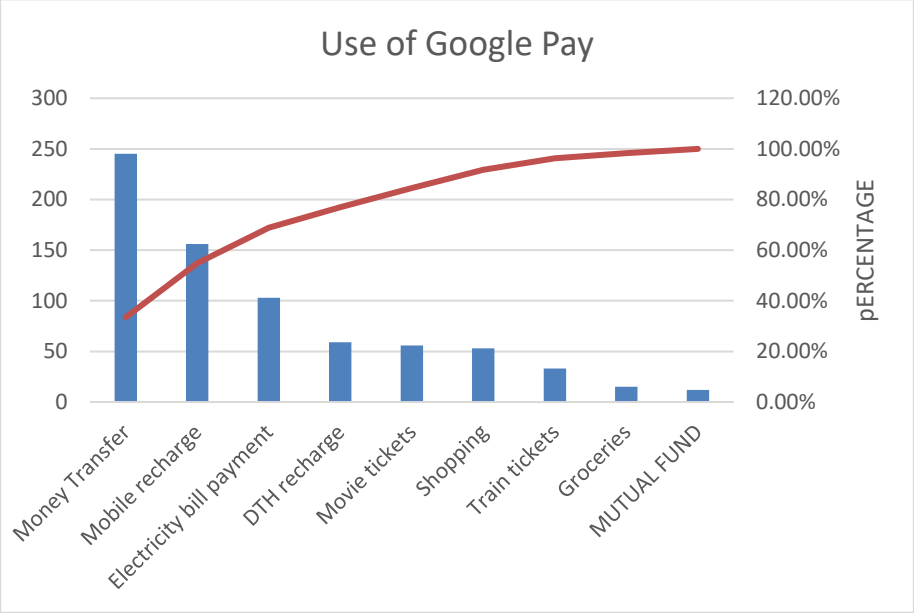


Interpretation:

Our analysis shows that, PhonePe is most frequently used for Mobile Recharge, Money transfer, Electricity bill payments, Movie tickets, shopping.

Pareto for use of Google pay (Tez)

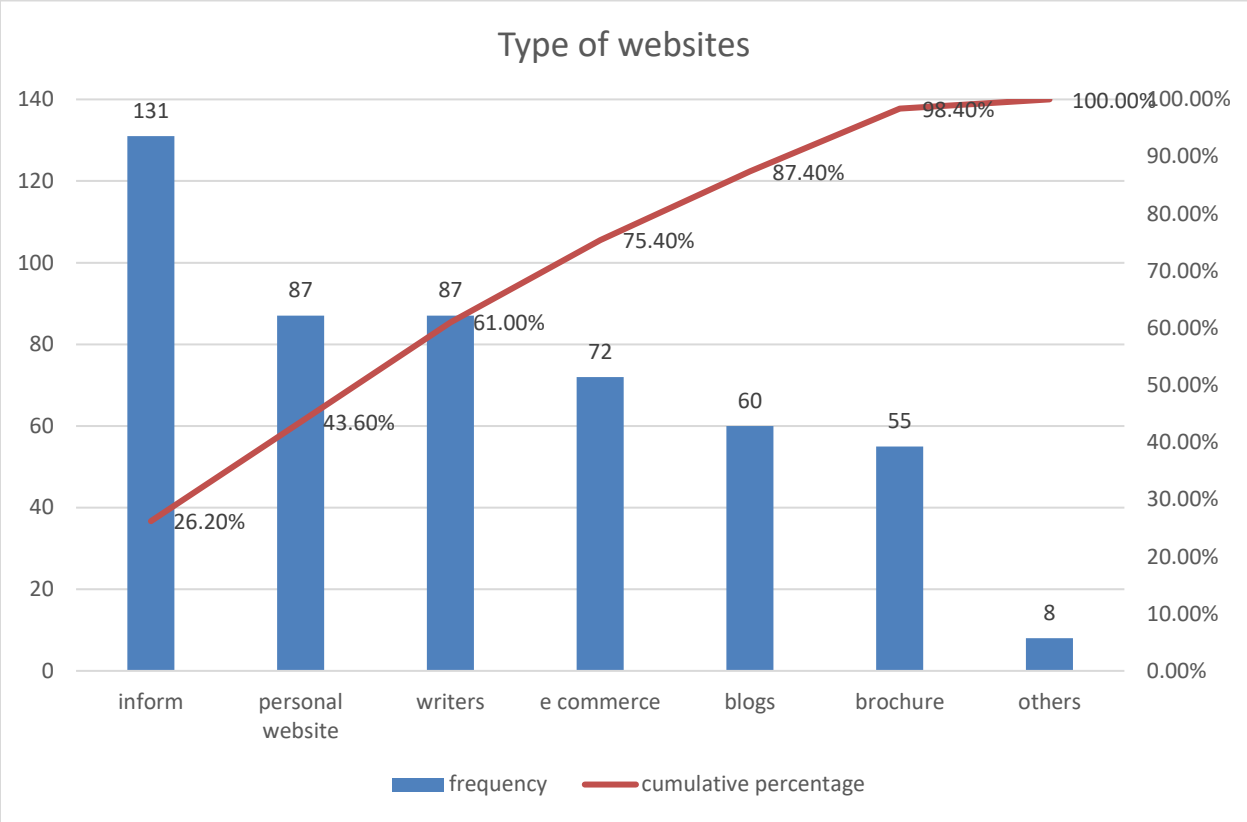
Use	Frequency
Money Transfer	245
Mobile Recharge	156
Electricity bill	106
Movie Tickets	56
Shopping	53
DTH Recharge	59
Train Tickets	33
Groceries	15
Mutual Fund	12



Interpretation: Our analysis shows that, Google Pay (also known as Tez) is most frequently used for Money transfer, Mobile Recharge, Electricity bill payments & DTH recharge.

- **Websites:**

Type of Websites	Information	Personal	Writers	E-commerce	Blogs	Brouchers	Others
Frequency	131	87	87	72	60	55	8



Interpretation: From the above Pareto diagram, we observe that 4 types constitute approximately 80% of the frequencies which are as follows: -

1)Information 2) Personal website 3) Writers
4) E-commerce

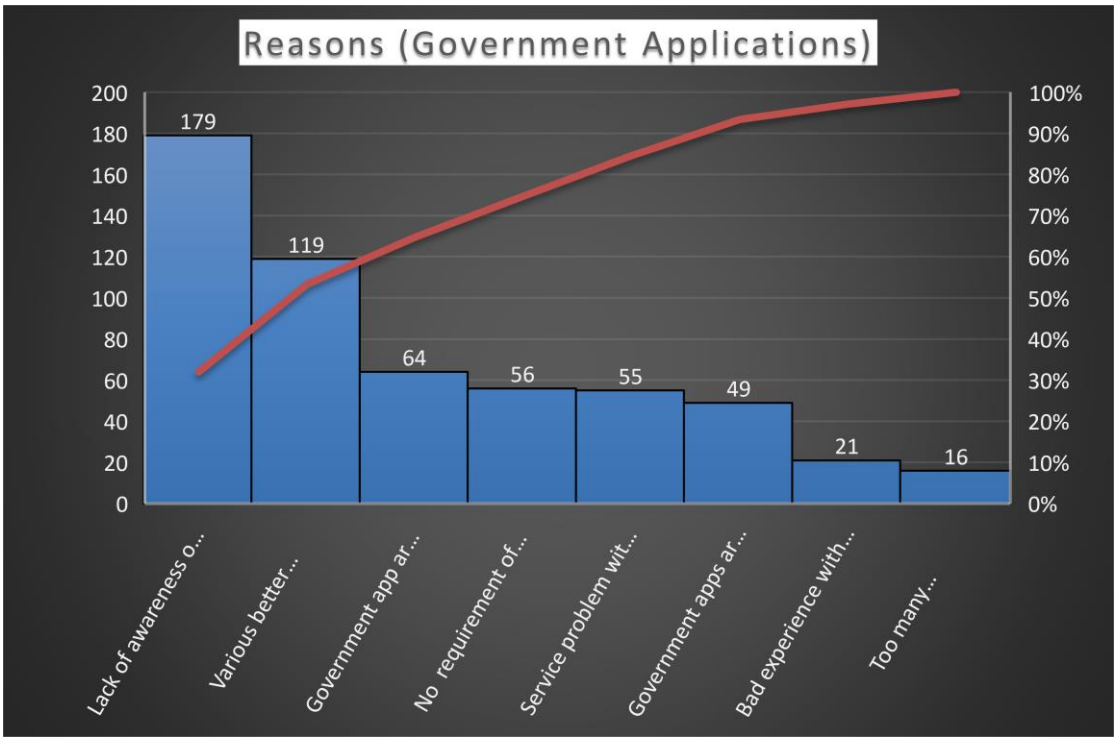
Conclusion: We have concluded that these are the top 4 types of websites which people prefer the most.

- **Awareness of Government Apps:**

Are You Aware of Government Application	Frequency	Percentage
Yes	537	60.88%
No	345	39.12%

From our Survey we came to know that around 61% of the People Are not Aware of Government Application and 39% Of the people are Not Aware of Government Application

Reason	Frequency
Lack of awareness of government apps	179
Various better options in private apps	119
Government app are very slow	64
Service problem with government app	55
Government apps are less versatile	49
Bad experience with government app	21
Too many advertisements in government app	16
No requirement of government apps	56



Interpretation: From the above Pareto diagram, we observe that 4 types constitute approximately 80% of the frequencies which are as follows: -

1)Lack of awareness

2)Various better options

3) Government app are very slow

4) Service problem with government app

Conclusion: We have concluded that these are the major reasons for Not preferring the Government App

CONCLUSION

The focus of our project was not only confined to one or two aspects of Mobile Application but also to see increasing trend in using E-wallet Application.

Our project was divided into 3 major objective and we ran statistical technique to find out our concerning result.

1) Objective 2 was to find out the reasons that attracts the person to download mobile application and also to know what are those factors that a person heeds before downloading Application. We use word cloud technique and found out that people look for features, entertainment factor and performance that an application provides to them. We also noticed that people before downloading Application look at the ratings that a particular Application possess and also whether the Application stands tall while being compared to other similar Application.

2) Objective 3 had a major concerned regarding E-wallet Application. This objective was focused on Identifying and Analyzing Socio-demographic factor influencing to download E-wallet Application. We use Binary logistic regression to assess the significance of the different factor in survey responses. We extracted 4 significant factors

3) Objective 4 was majorly concerned regarding different usage of E-wallet Application before and after demonetization and we use Wilcoxon sign rank test to carry out following analysis and we have noticed that there is positive impact i.e. there is increase in usage of E-wallet Application for different purposes i.e. recharges, DTH recharge, etc.

4) In Objective 5 we use factor analysis followed by pareto analysis to find out most preferred Mobile Wallet and for what purpose was it used. We found out Paytm, Phonepe, Google pay were the most preferred ones.

5) Objective 6 was concerned government Application where we wanted to note the reason why people prefer private Application over government Application and we found out the following reasons: Lack of awareness, Various better options, Government app are very slow, Service problem with government app.

Coding

R code for word cloud

Here is how we went about with our analysis:

```
#Create a text file and load the text file

>text=readLines(file.choose())

#Install and load required packages
>install.packages("tm")
>library("tm")
>install.packages("wordcloud")
>library("wordcloud")

#create corpus

>docs=Corpus(VectorSource(text))

#DATA CLEANING
>docs=tm_map(docs,tolower)

>docs=tm_map(docs,stripWhitespace)

#ANALYSIS

>dtm<-TermDocumentMatrix(docs)

>m <- as.matrix(dtm)

>v<- sort(rowSums(m),decreasing = T)

>d<- data.frame(word=names(v),freq=v)

>head(d,30)

>wordcloud(words=d$word,freq=d$freq,min.freq=10,max.word=1000,random.order=F,
rot.per=0.35,colors=brewer.pal(8,"Dark2"))
```

SAS code for binary logistics

```
#after importing the data then view it

proc print data=s1;

run;

#summary of continous variables

proc means data=s1;

var age ;

run;

#summary of categorical variables

proc freq data=s1;

table wallet gender education_ status income marriage;

run;

ods graphics on;

ods html;

#running binary logistic regression

proc logistic data=s1 plots=roc desc;

class gender(ref='1') education_(ref='4') status(ref='1') income(ref='6') marriage(ref='1')/param=ref;

model wallet= age gender education_ status income marriage;

run;

#using stepwise procedure

proc logistic data=s1 plots=roc desc;

class gender(ref='1') education_(ref='4') status(ref='1') income(ref='6') marriage(ref='1')/param=ref;
```

```
model wallet= age gender education_ status income marriage/
selection=stepwise slentry=0.05 slstay=0.05 details ctable lackfit pprob=0.5;
run;
#training and validation
#split data into two datasets:70%-training 30%-validation
proc Surveyselect data=s1 out=split seed=1234 samprate=0.7 outall;
run;
Data training validation;
Set split;
if selected = 1 then output training;
else output validation;
Run;
proc logistic data=training plots=roc desc;
class gender(ref='1') education_(ref='4') status(ref='1') income(ref='6') marriage(ref='1')/param=ref;
model wallet= age gender education_ status income marriage/
selection=stepwise slentry=0.05 slstay=0.05 details ctable lackfit pprob=0.5;
score data = training out=logit_training fitstat outroc=troc;
score data= validation out=logit_validation fitstat outroc=vroc;
run;
#for KS statistics
Proc npar1way data=Logit_Validation edf;
class wallet;
```

```
var p_1;

run;

#to check with interaction terms

proc logistic data=training plots=roc desc;

class gender(ref='1') education_(ref='4') status(ref='1') income(ref='6') marriage(ref='1')/param=ref;

model wallet= age gender education_ status income marriage age*gender age*enducation_ age*status age*income age*marriage gender*education_ gender*status gender*income gender*marriage education_*status education_*income education_*marriage status*income status*marriage income*marriage/

selection=stepwise slentry=0.05 slstay=0.05 details ctable lackfit pprob=0.5;

score data = training out=logit_training fitstat outroc=troc;

score data= validation out=logit_validation fitstat outroc=vroc;

run;

ods html close;

ods graphics off;
```

R code for Wilcoxon sign rank test

```
#import file

mydata=read.csv(file.choose())

attach(mydata)

#to view variables in your data

names(mydata)

#Code for wilcoxon sign rank test

wilcox.test(v1,v2,mu=0,alt="two.sided",paired = T,conf.int = T,conf.level = 0.95,exact = F,correct = F)
```

QUESTIONNAIRE

Applications-Tastic

We are students of Msc Statistics from Mumbai University conducting a survey on Mobile Applications.We request for your honest response. Your participation in the survey is completely voluntary. Thank you for your participation.

*Required

1. What is your age? *

_____

2. Gender *
Mark only one oval.

- ◯ Male
- ◯ Female

3. Education Qualification? * Mark only one oval.

- ◯ S.S.C. & below it.
- ◯ Diploma
- ◯ H.S.C.
- ◯ Graduate
- ◯ Post-Graduate
- ◯ PhD

4. What is your professional status? *
Mark only one oval.

- ◯ Student
- ◯ Self-employed
- ◯ Employed Part-time
- ◯ Businessman
- ◯ Freelance
- ◯ Un-employed
- ◯ Government Sector
- ◯ Private Sector
- ◯ Other: _____

5. Individual Monthly Income (In Thousand) *

Mark only one oval.

- Not earning
- 1 - 20000
- 20001 - 40000
- 40001 - 60000
- 60001 - 80000
- 80001 - 100000
- Above 100000
- Other: _____

6. Marital Status?

Mark only one oval.

- Single
- Married

7. What do you prefer the most? * Mark only one oval.

- Mobile Website Skip to question 8.
- Mobile Application Skip to question 10.

Website

8. If Mobile Website then why? (Multiple Selection) * Tick all that apply.

- [ ] Easy to find it
- [ ] More use to it
- [ ] Convenience
- [ ] Time saving
- [ ] Availability
- [ ] Shareability
- [ ] Other: _____

9. What type of websites do you use? (Multiple Selection) * Tick all that apply.

- [ ] Personal Website
- [ ] Writers/Authors Website (Authors Facebook/twitter page etc.)
- [ ] Blogs
- [ ] Informational/Directory Websites (Eg; Wikipedia)
- [ ] Online Business catalog/brochure
- [ ] E Commerce
- [ ] Other: _____

Application

10. If Mobile Application then why? (Multiple Selection) * Tick all that apply.

- [ ] Easy to find it
- [ ] More use to it
- [ ] Convenience
- [ ] Time saving
- [ ] Availability
- [ ] Shareability
- [ ] Other: _____

11. Do you use following application ? * Mark only one oval per row.

	Yes	No
Life style	◯	◯
Education/Dictionary	◯	◯
Travelling	◯	◯
Social Networking	◯	◯
Entertainment (Hotstar, voot, etc.)	◯	◯
Games	◯	◯
Food Delivery app	◯	◯
Mobile banking/e-wallets	◯	◯
Health & Fitness	◯	◯
Security & Privacy app	◯	◯

12. What are the factors that influence you to download an apps? (Multiple Selection) * Tick all that apply.

- [ ] Icon
- [ ] App size
- [ ] Rating
- [ ] Reviews
- [ ] Features
- [ ] Description
- [ ] Name
- [ ] Performance
- [ ] Comparison
- [ ] Top downloads
- [ ] Price

13. Why do you download apps? (Multiple Selection) *
Tick all that apply.

- [ ] entertainment
- [ ] interaction
- [ ] recommendation
- [ ] reward
- [ ] randomly search
- [ ] curiosity
- [ ] Other: _____

Application (Paid/Free)

14. Do you use Paid Applications? * Mark only one oval.

- ◯ Yes Skip to question 15.
- ◯ No Skip to question 17.

Paid Apps

15.How much money are you willing to spend on paid Apps? * Mark only one oval.

- ◯ 1-100
- ◯ 101-200
- ◯ 201-400
- ◯ 401-600
- ◯ 601-800
- ◯ Above 800

16. Why do you pay for an Apps? (Multiple Selection) * Tick all that apply.

- [ ] To remove advertisements
- [ ] To subscribe to required content
- [ ] To get additional features
- [ ] Cannot find free apps with similar features
- [ ] Paid app is on sale
- [ ] Paid apps have better quality in general
- [ ] Other: _____

Mobile Wallet

17.Do you use mobile wallet/e-wallet (e.g. Paytm, PhonePe, Tez, etc.)? * Mark only one oval.

- ◯ Yes Skip to question 18.
- ◯ No Skip to question 27.

If Yes, for Mobile Wallet then

18. In which year did you start using Mobile Wallet/Banking apps? * Mark only one oval.

- ◯ before 2014
- ◯ 2015
- ◯ 2016
- ◯ 2017
- ◯ 2018

19. How often do you use e-wallet applications for following purpose? (Before Demonetization) * Mark only one oval per row.

	Always	Sometimes	Never
DTH Recharge	◯	◯	◯
Shopping	◯	◯	◯
Movie Tickets	◯	◯	◯
Mobile Recharge	◯	◯	◯
Mutual Funds	◯	◯	◯
Money Transfer	◯	◯	◯
Electricity Bill	◯	◯	◯

20. How often do you use e-wallet for following purpose? (After Demonetization) * Mark only one oval per row.

	Always	Sometimes	Never
Mobile Recharge	◯	◯	◯
DTH recharge	◯	◯	◯
Electricity Bill	◯	◯	◯
shopping	◯	◯	◯
Movie Tickets	◯	◯	◯
Money Transfer	◯	◯	◯
Mutual funds	◯	◯	◯

21. Which are the most important benefits that you see while using the mobile phones for financial transaction? *
Mark only one oval.

- Save time
- Save Transport Cost
- Faster Transaction
- Security
- Discount/Cashback
- Other: _____

PAYTM

22. For what purpose do you use Paytm? (Multiple Selection) * Tick all that apply
.

- Do not use
- Mobile Recharge
- DTH Recharge
- Electricity Bill Payments
- Shopping
- Movie Tickets
- Train Tickets
- Money Transfer
- Groceries
- Mutual Funds
- OTHER

Mobikwik

23.For what purpose do you use Mobikwik? (Multiple Selection) * Tick all that apply.

- Do not use
- Mobile Recharge
- DTH Recharge
- Electricity Bill Payments
- Shopping
- Movie Tickets
- Train Tickets
- Money Transfer
- Groceries
- Mutual Funds
- OTHER

Freecharge

24.For what purpose do you use Freecharge? (Multiple Selection) * Tick all that apply.

- [ ] Do not use
- [ ] Mobile Recharge
- [ ] DTH Recharge
- [ ] Electricity Bill Payments
- [ ] Shopping
- [ ] Movie Tickets
- [ ] Train Tickets
- [ ] Money Transfer
- [ ] Groceries
- [ ] OTHER

PhonePe

25.For what purpose do you use PhonePe? (Multiple Selection) * Tick all that apply.

- [ ] Do not use
- [ ] Mobile Recharge
- [ ] DTH Recharge
- [ ] Electricity Bill Payments
- [ ] Shopping
- [ ] Movie Tickets
- [ ] Train Tickets
- [ ] Money Transfer
- [ ] Groceries
- [ ] OTHER

Google Pay (Tez)

26.For what purpose do you use Google pay (Tez)? (Multiple Selection) * Tick all that apply.

- [ ] Do not use
- [ ] Mobile Recharge
- [ ] DTH Recharge
- [ ] Electricity Bill Payments
- [ ] Shopping
- [ ] Movie Tickets
- [ ] Train Tickets
- [ ] Money Transfer
- [ ] Groceries
- [ ] Mutual Funds
- [ ] OTHER

If No, For Mobile Wallet then?

27. Why you don't use E-wallet? (Multiple Selection) * Tick all that apply
.

- [ ] Lack of awareness
- [ ] Fear of losing money
- [ ] Because It is dependent on gadget (eg. Less storage in mobile, Low Battery, etc...)
- [ ] Not applicable everywhere
- [ ] No need
- [ ] Other: _____

Government App

28.Are you aware of Government Applications? * Mark only one oval.

- ◯ Yes
- ◯ No

29.Do you use government Applications? * Mark only one oval
.

- ◯ Yes Skip to question 30.
- ◯ No Skip to question 31.

If Not Government apps

30. What are the reasons for not preferring government apps? (Multiple Selection) * Tick all that apply.

- [ ] Lack of awareness of government apps
- [ ] Various better options in private apps
- [ ] Government app are very slow
- [ ] Service problem with government app
- [ ] Government apps are less versatile
- [ ] Bad experience with government app
- [ ] Too many advertisements in government app
- [ ] No requirement of government apps
- [ ] Other: _____

Government or both

31. Rate the following apps that you have used. * Mark only one oval per row.

	Very bad	Bad	Good	Very good	Excellent
BHIM					
IRCTC					
MyGov					
Online-seva-Indian digital service					
Swach Bharat					
Voter-Apps					
Online RTI					
GST-Rate finder					
M-Passport Seva					
UMANG					
PayTm					
Freecharge					
Mobikwik					
PhonePe					
Google Pay					