MEDISTIC PANDEMIC ERA

UNIVERSITY OF MUMBAI

DEPARTMENT OF STATISTICS

Vidyanagari, Mumbai-400098

CERTIFICATE

This is to certify that the following students of M.Sc. Part-II have successfully completed the project **(MEDISTIC PANDEMIC ERA)** during the academic year 2020-2021. This work has been done independently to the best of our knowledge and Awareness.

Students involved in this academic project group are:

- Ms. Rasika Ashok Panire
- Ms. Rushika Anil Suryawanshi
- Ms. Ashwini Vijay Sarvade
- Ms. Nikita Ashok Kapade
- Ms. Shreya Suresh Thakur
- Mr. Aaditya Keshav Kamble
- Mr. Kiran Bhagwan Avhad
- Mr. Smit Ravindra More
- Mr. Akshay Digamber Gaikwad

Dr. Mr. Santosh P. Gite

Guide and Mentor

Dr. Mrs. Vaijayanti U. Dixit

Head of the Department

ACKNOWLEDGEMENT

We are extremely thankful to our guide Dr. Santosh Gite for his guidance and support. We are extremely glad to have him as our project guide.

We would also like to thank Dr. Alok Dabade, Dr. Vaijayanti Dixit, Prof.Sheetal Chabukswar, Prof. Priyesh Tiwari and Prof. Ashwini Ghanekar (Co-mentor) for extending their help and support.

We also thank other professors of the department and the entire staff of Statistics Department, Mumbai University for providing us with all the necessary infrastructure and co-operation with us at every stage of our project.

Last but not the least, our special thanks to allthe people for providing us with their vital information, opinions, precious time and their support.

INDEX

INTRODUCTION

Insurance acts as a shield against risks and unforeseen circumstances.

In 1986, Medical Insurance was launched in India. The first health policies in India were Mediclaim policies. The health insurance industry has grown significantly mainly due to liberalization of economy and general awareness. In year 2000, Government of India liberalized insurance and allowed private companies into the insurance sector.

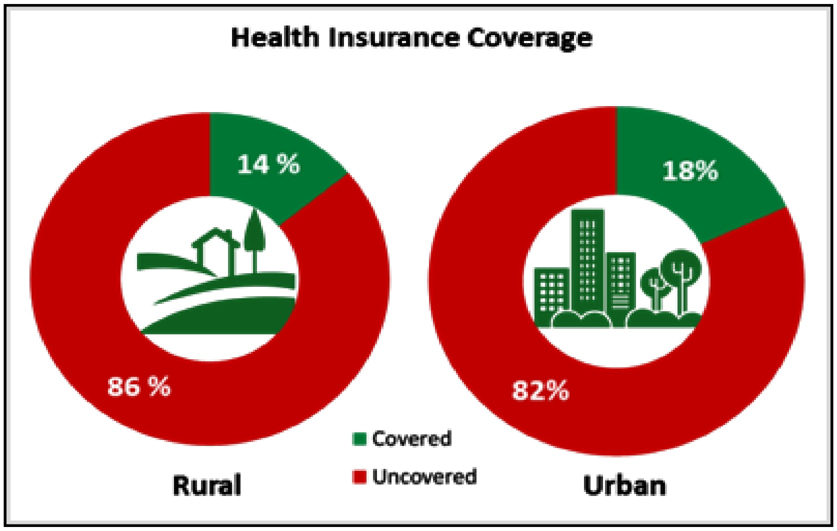
Mediclaim policy is an insurance coverage to claim reimbursement of medical treatment bills generated due to Health-related hospitalization. There are two Mediclaim policies to customers, get your bills claimed either by cashless facility i.e., your bills are directly paid to the hospital or you can pay your bills in the hospital and get a reimbursement after submission of the same to the insurance company. Mediclaim policy is an essential for the peoples because it saves financial loss in case of hospitalization for any sickness, disease or accident.

The Mediclaim policy can be taken on an individual basis or for the entire family if the need be. The insurance premium which is charged by the insurance company will defer from company to company and will depend on whether the Mediclaim policy has been taken for an individual or a group or whether the insurance policy is a cashless or not. The medical age for a Mediclaim policy differs and thus in certain cases the applicant may be required to undergo a medical test or in certain cases the applicant may not be asked to undergo a medical test.

The insurance sector is regulated by the Insurance Regulatory Development Authority (IRDA) of India in our country and the organization has introduced several new rules to standardize this sector aiming to increase the convenience of insurance holders. Some of the guidelines are framed by IRDA insurance include critical illness, pre-existing diseases, hospitalization, etc. It aims to decrease the frauds and to increase satisfaction among insurance policyholders.

As per section 80D, a taxpayer can avail tax deduction on premium paid towards medical insurance for self, spouse, dependent parents and dependent children. Limit of deduction varies with age, for self, spouse and dependent children deduction of Rs 25000 is available. Additional deduction of Rs 25000 is available for insurance paid for parents aged less than 60 years and Rs 50,000 if parents are above 60 years of age.

Below pie-chart shows percentage of Health Insurance Coverage in India



Why this project?

The main objective is to assess the awareness about Mediclaim policy (Health Insurance) in Maharashtra and other states. And if they are willing to buy policy in future or during the Covid-19 pandemic.

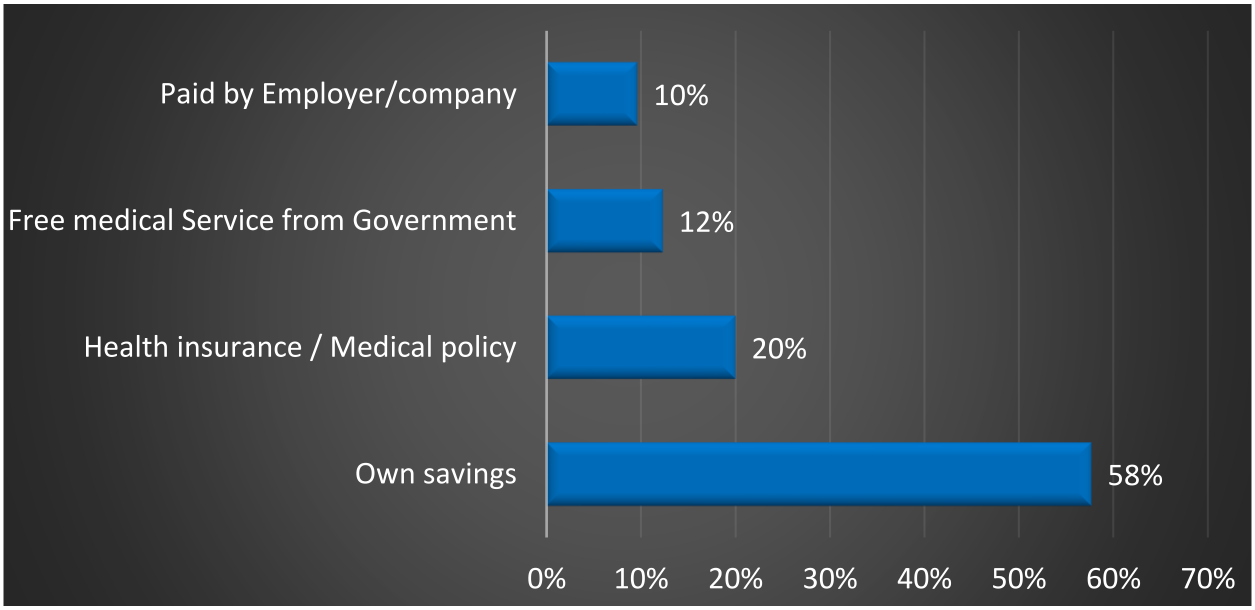
The behavior of the consumer in insurance services is influenced by numerous factors that can be classified as: Situational factors (factors appearing in all the phases of the decision process: before the acquisition of the insurance policy, during the acquisition and after the purchase), characteristics of the insurance products, Factors related to the premiums paid and the payment facilities offered by the insurance companies plays an important role. The external environment (natural, demographical, economical, legislative and technological) also influences the purchase decision. At the same time level of education of customer (not only degree point of view) and awareness are two crucial factors.

The current population of India is 1.3billion, only 65 million Indians have been introduced to insurance thus representing huge untapped potential for insurance companies. Health care has always been a problem area for

India, a nation with a large population and a larger percentage of this population living below the poverty line. This can be explained partly by the fact that India is a low-income developing economy whose domestic saving potential in long term assets is not as high as that of developed economies to spread the habit of insurance.

Uncertainty is a fact of life, and this has been very well proven by the ongoing pandemic that has brought the entire world to its knees. A chaos that could not have been predicted but is no less than bitter truth. Covid-19 is spreading like wildfire and has proven to be a life-threatening health hazard. In a crucial situation like this, having an optimum health insurance plan that would provide some financial relief to the insured and the dependents has been the need of the hour. Medical treatments are difficult and expensive, digging deep into the pockets whilst struggling fatal disease is the last thing anyone would ever want to experience.

As medical care advances and treatments increase health care cost also increase. The purpose of Mediclaim policy/Health Insurance is to protect in and one's family financially in the event of an unexpected serious illness or injury that could be very expensive. One of the small reliefs is that insurance companies do cover this ailment under the health insurance plans, and few of the insurers have also come with dedicated plans to make things better for the new buyers (who don't have an existing plan) during this crucial period.



From above graph we can see that 58% of the population use their own savings for the medical expenses. Even after there are many medical policies available individuals are going for it. Therefore, it is necessary to spread awareness about Mediclaim policy and its different benefits. So that people will give preference to the Mediclaim policy and use benefits of it.

METHODOLOGY

Steps involved in conducting the survey:

- Defining our objectives and scope of the survey.
- Specifying information needs.
- Literature survey.
- Identifying primary data sources.
- Designing questionnaire.
- Pilot Survey.
- Modifying Questionnaire.
- Data Collection.
- Data Coding and Data Entry.
- Data Analysis.
- Preparation of Project Report.

Questionnaire Designing:

The questionnaire prepared consisted of basic information of individuals, like their Age, Gender, Marital Status, where do they live, Qualification, Profession, Family Annual Income etc.

Data Collection:

We conducted a pilot survey of sample size 70. After which we made the necessary changes in the questionnaire. We surveyed 583 participants individually and via Google Forms. After data cleaning, we were left with final sample size of 520.

OBJECTIVES

1. To identify the relationship between socio-demographic factors and purchase of Mediclaim Policy.

2. To study the relation between health factors and Purchase of Mediclaim Policy.

3. To check whether there is significant difference in the Purchase of Mediclaim Policy before and after COVID-19.

4. To find preferable companies to Purchase Mediclaim and preferable Mediclaim Plan.

5. a) To study the reasons for not Purchasing Mediclaim.
   b) To determine the best sources to gain information about Mediclaim.

STATISTICAL TECHNIQUES AND SOFTWARE

- Techniques:

1. Binary Logistic Regression Analysis.
2. Chi-Square test of association.
3. McNemar test.
4. Pareto Analysis.
5. Factor Analysis.

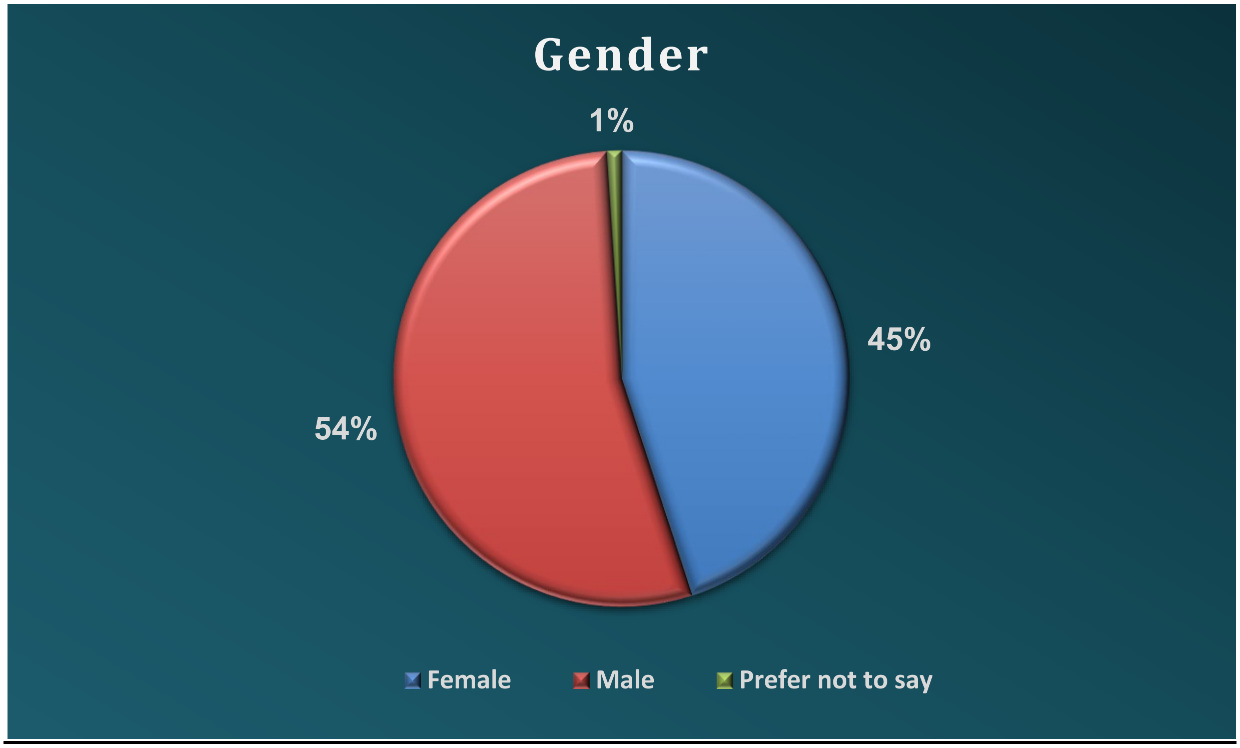
- Software's:

1. Statistical Analysis System (SAS)
2. R Software
3. Python
4. SPSS
5. Minitab
6. MS-Office

GRAPHICAL REPRESENTATION

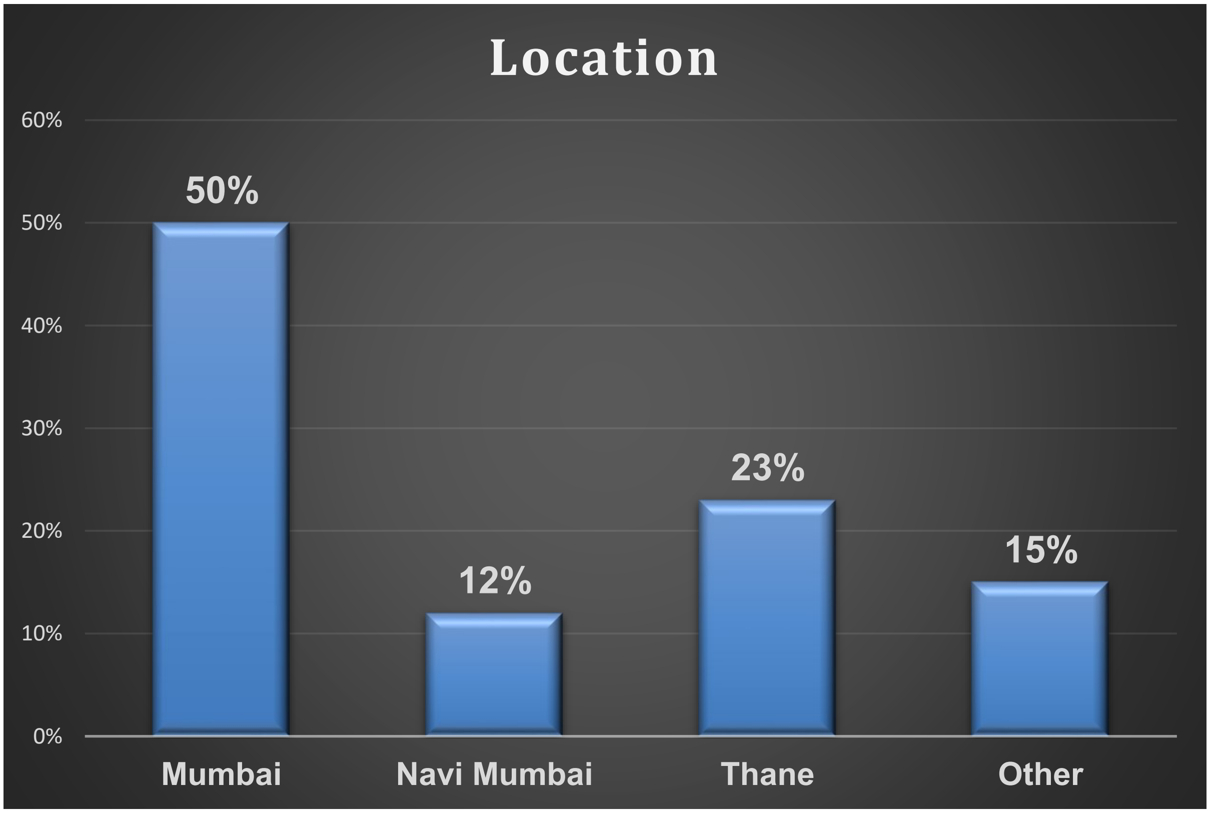
1. <u>Gender</u>

Gender	Percentage
Female	45%
Male	54%
Prefer not to say	1%



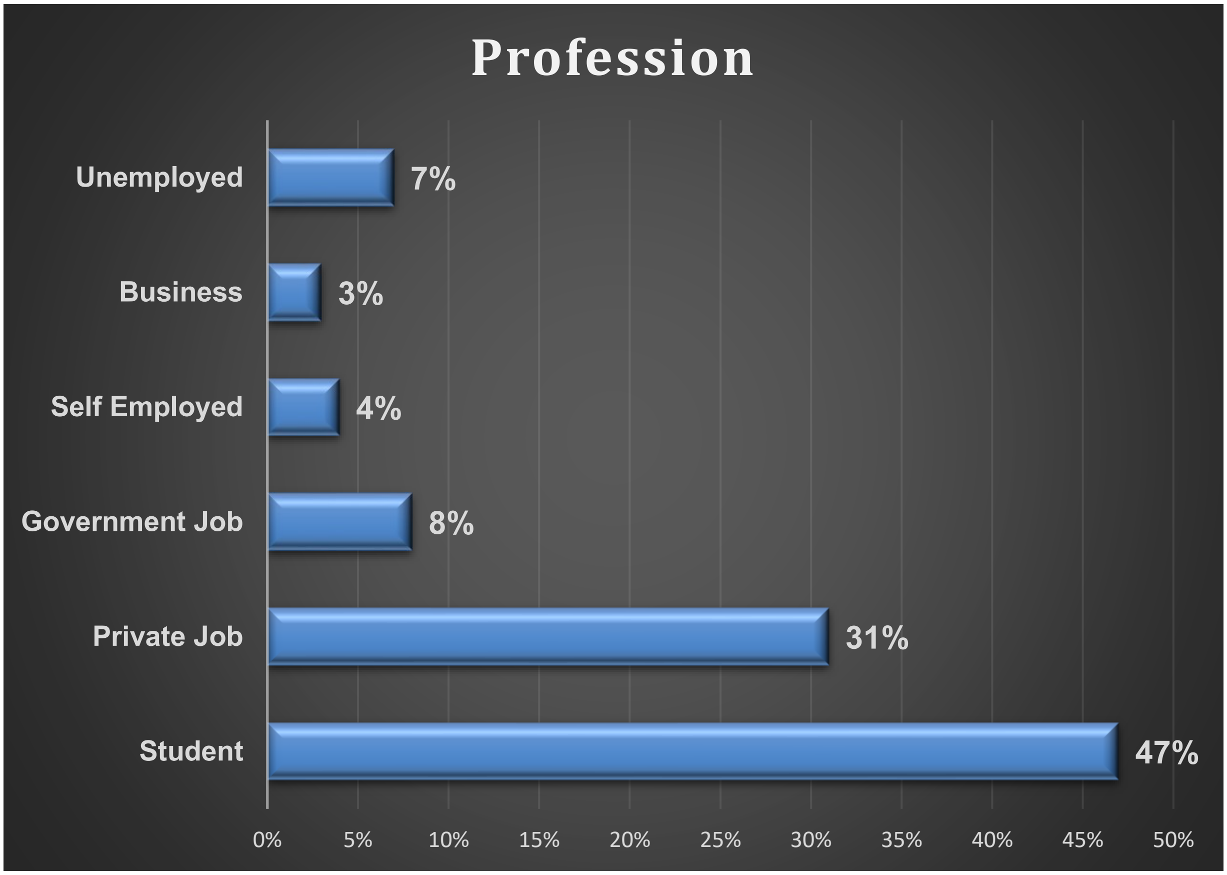
2. Location

Location	Percentage
Mumbai city	50%
Navi Mumbai	12%
Thane	23%
Others	15%



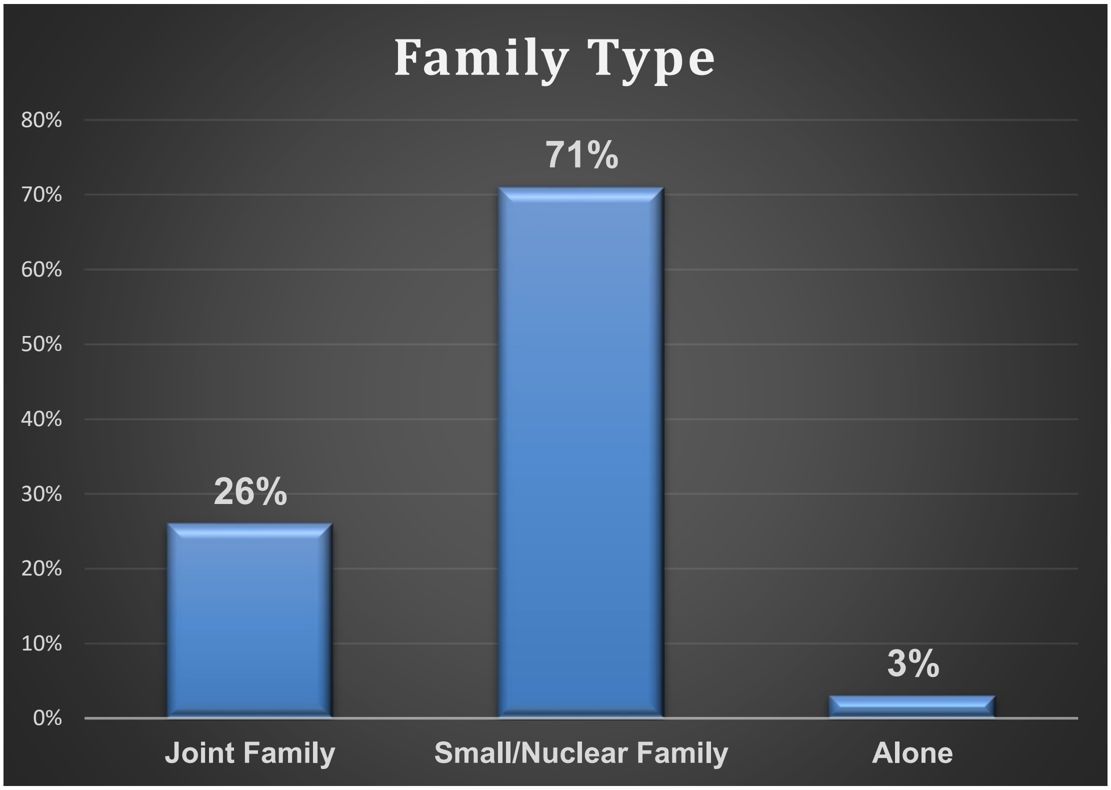
3. Profession

Profession	Percentage
Student	47%
Private job	31%
Government job	8%
Self employed	4%
Business	3%
Unemployed	7%



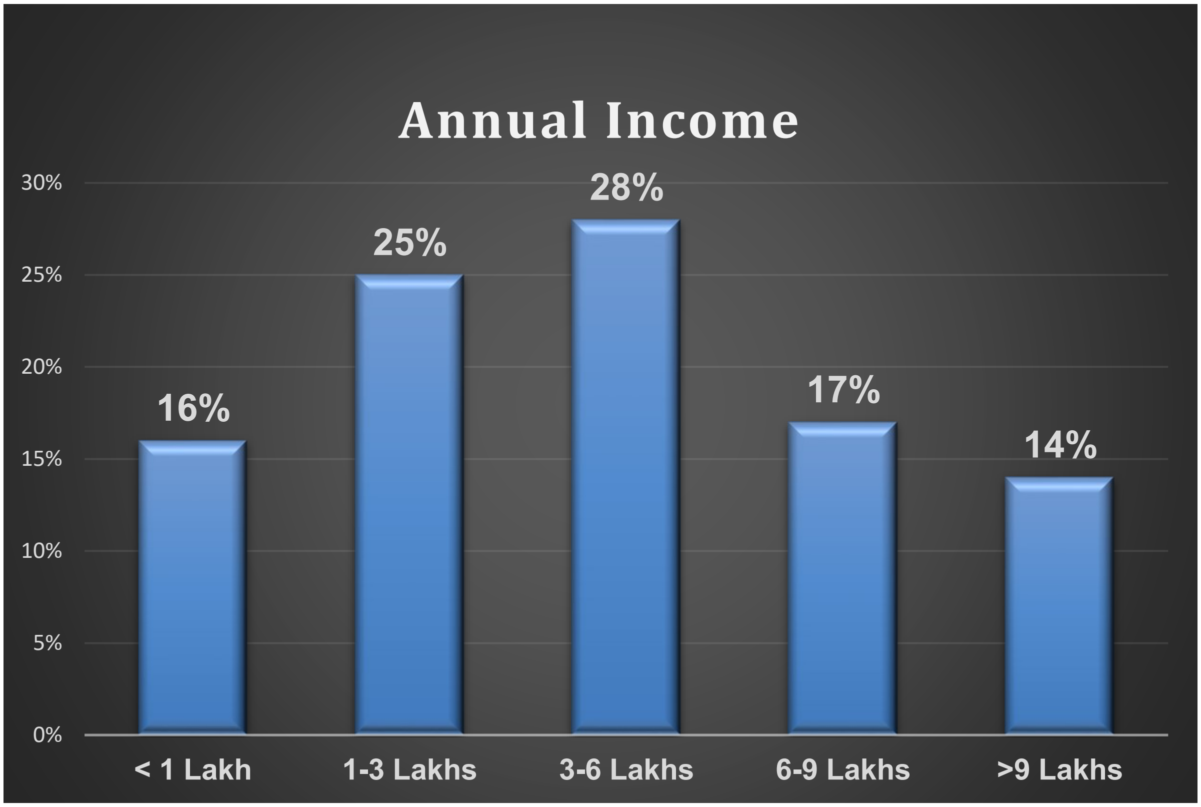
4. Family type

Family type	Percentage
Joint family	26%
Small/Nuclear family	71%
Alone	3%



5. Annual Income

Annual income	Percentage
< 1 Lakh	16%
1 - 3 Lakhs	25%
3 - 6 Lakhs	28%
6 – 9 Lakhs	17%
>9 Lakhs	14%



OBJECTIVE I

To study the relation between socio-demographic factors and purchase of Mediclaim policy.

BINARY LOGISTIC REGRESSION

Binary logistic regression is a form of regression which is used when the dependent variable is binary (= 0 or 1, presence or absence) and the independent variables are of any type. The goal of the analysis using logistic regression method is to find the best fitting and most reasonable model to describe the relationship between the outcome (dependent or response variable) and the set of independent (predictor variable) or understand the impact of explanatory variables and to determine the percent of variation in the dependent variable explained by the independent variable, to rank the relative importance of independents, to assess interaction effect and to covariates the control variables.

The Binary logistic regression model is:

$$\Pi(\mathrm{x}) = \frac{\exp\,(\beta_0 + \sum_{i}^{p} \beta_i\, X_i\,)}{1+\exp(\beta_0 + \sum_{i}^{p} \beta_i\, X_i)}$$

Where,

(x): Conditional probability that the outcome is present, i.e., Pr(Y=1|X).

Y: Response variable

X: Vector of independent variables

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

Logit transformation of above model:

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

Where,

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

β_0 : Intercept of the regression variables.

β_i : Slope of ith regression line/ coefficient of ith predictor variable.

PROCEDURE TO CARRY OUT BINARY LOGISTICS REGRESSION ANALYSIS:

GLOBAL TESTING:

Global testing is used to test whether or not at least one of the independent variables influences the dependent variable. i.e., at least one of the variables insignificant.

STEPWISE REGRESSION:

Steps involved in stepwise regression are as follows:

- It begins with no variables in the equation and at each step it enters or removes a predictor based on a partial F-test (i.e., the t-test).
- At each step an appropriate statistical test for each variable currently in the model will be performed to determine whether the variable has significant contribution to the model.
- A negative result may suggest the removal of the variable.
- When no variable in the current model can be removed and no new variables are suggested to be

added to the model, the selection procedure stops.

RESIDUAL CHI-SQUARE:

The residual chi-square test is carried out to test the significance of the remaining independent variables which have not been entered into the model.

WALD STATISTICS:

Wald statistics is used to test the significance of individual coefficients in the model. Each Wald Statistics is compared with chi-square distribution with 1 degree of freedom.

ODDS RATIO:

The odds ratio measures the strength of association between a predictor andthe response variable of interest.

HOSMER AND LEMESHOW GOODNESS OF FIT TEST:

- The Hosmer-Lemeshow statistics 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 the desirable outcome ofnon-significance, indicating that the model prediction does not significantly differ from the observed.
- Classification table: It is used to summarize the result of a fitted logistic regression model. This table is a result of cross-classifying the outcome variable with a dichotomous variable whose values are derived from the estimated logistic probabilities.

ROC CURVE: Receiver Operating Characteristics Curve

ROC Curves are used to evaluate and compare the performance of diagnostic tests. They can also be used to evaluate model fit. ROC Curve is just aplot of proportion of the true positives (events predicted to be events i.e., Sensitivity) v/s the proportion of false positives (non-events predicted to beevents i.e., 1-Specificity).

The accuracy of the test is measured by area under ROC curve. An area of one represents 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 ROC space, the more accurate the test, the true positive rate (sensitivity) is high and 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.

Dependent Variable:

Have Mediclaim or not? (Y)

0 - No

1 - Yes

Independent Variables:

X_1 : Age of an Individual

X_2 : Gender

0- Male
1- Female
2- Prefer not to say

X_3 : Marital status

0- Unmarried

1- Married
2- Widow
3- Divorcee

X_4 : Area
0- Rural
1- Urban
2- Suburban

X_5 : Location
0- Mumbai
1- Navi Mumbai
2- Thane
3- Others

X_6 : Qualification
0- SSC and below
1- HSC
2- Bachelor's degree
3- Master's degree
4- Doctorate degree
5- Diploma
6- Other professional courses

X_7 : Profession
0- Student
1- Private job
2- Government job
3- Self employed
4- Business
5- Unemployed

X_8 : Family type
0- Joint family
1- Nuclear family

2- Alone

X_9 : Annual income

0- Less than 1 lakh
1- 1 to 3 lakhs
2- 3 to 6 lakhs
3- 6 to 9 lakhs
4- More than 9 lakhs

X_{10} : Number of earning members in the family

X_{11} : Number of dependent members in the family

X_{12} : Number of diseases

0- No diseases
1- 1 disease
2- 2 diseases
3- 3 diseases
4- 4 diseases
5- 5 diseases
6- 6 diseases

X_{13} : Do you smoke?

0- No
1- Yes

X_{14} : Do you drink?

0- No
1- Yes

DESIGN VARIABLES:

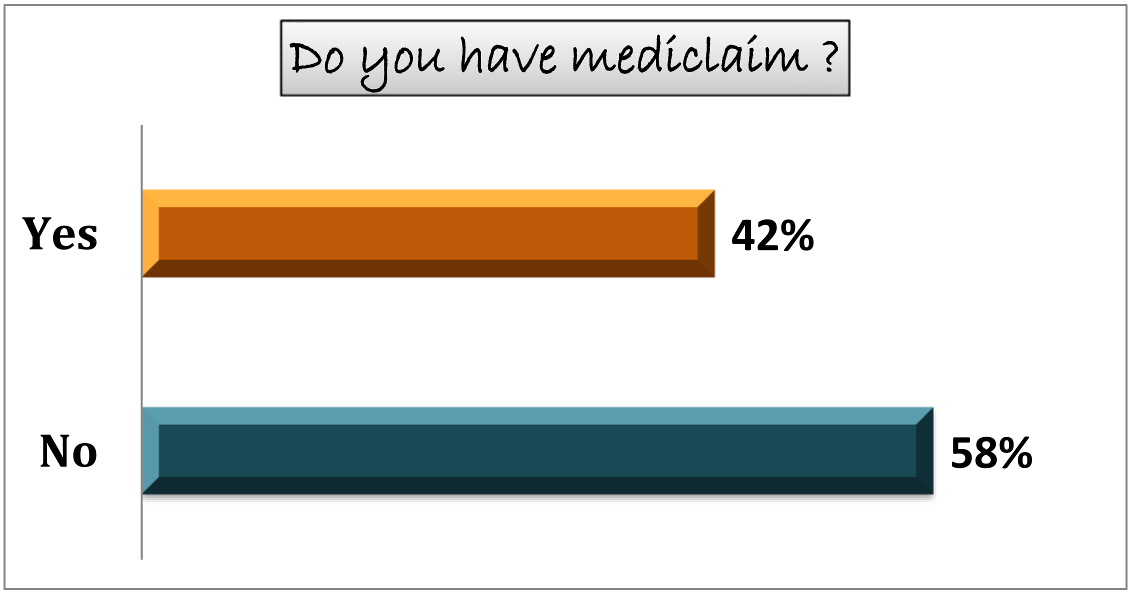
The independent variables are converted to design variables using reference category. For example, variable ‘Qualification’ has 7 categories where the reference category is ‘SSC & Below’ coded as (0,0,0,0,0,0) thus 6 Design variables are defined corresponding to Qualification as

Qualification	**SSC and below**	0	0	0	0	0	0
	HSC	1	0	0	0	0	0
	Bachelor’s degree	0	1	0	0	0	0
	Master’s degree	0	0	1	0	0	0
	Doctorate degree	0	0	0	1	0	0
	Diploma	0	0	0	0	1	0
	Other professional courses	0	0	0	0	0	1

ANALYSIS

a) Elementary Data Analysis:

Have Mediclaim	Coded Variables	Count
Yes	1	216
No	0	304



b) Detection of multicollinearity:

Before going to further analysis, we will check the assumption of multicollinearity. We will proceed further only if there is no serious multicollinearity in our data i.e., variance influence factor should be less than 5.

Variables	VIF
Age	3.241
Gender	1.070
Status	2.939
City	1.152
Area	1.078
Qualification	1.216
Profession	1.466
Family type	1.332
Annual income	1.287
Earning members	1.125
Dependent members	1.293
No of Diseases	1.007
Smoking habit	1.715
Drinking habit	1.575

Conclusion:

As we can see from above table all variance influence factors (VIF) < 5. So, we can conclude that there is no serious collinearity in our dataset. Therefore, we proceed for further analysis.

c) Splitting Data into Train & Test:

Train Data	
Yes	151
No	216
Total	**364**

Test Data	
Yes	65
No	91
Total	**156**

Binary logistic regression procedure

a) Global testing:

Hypothesis:

H_0 : All independent variables are insignificant i.e., $\beta1=...=\beta14=0$.

H_1 : At least one of the independent variables is significant

Test statistics: (L1-L2) ~ χ^2

L1: (-2LogL) for model without the independent variables.

L2: (-2LogL) for model with all the independent variables.

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	183.7789	32	<.0001
Score	134.4690	32	<.0001

Test Criterion**:** Reject H_0, if p-value < 0.05

Conclusion:

The p-value of Likelihood ratio test is less than 0.05, we reject the null hypothesis and conclude that, at least one variable is significant at 5% level of significance.

b) STEPWISE REGRESSION:

After applying Stepwise Selection procedure, at the end of step 5 variable entered were X_9 (Annual income), X_7 (Profession), X_3 (Marital status), X_{11} (Number of dependent family members), X_5 (Location) after which no additional variable entered further at 5% level of significance.

The result and summary of stepwise selection regression procedure is,

Summary of Stepwise Selection

Step	Effect		DF	Number In	Score χ2	Wald χ2	Pr > χ2
	Entered	Removed					
1	Annual income		4	1	58.0209		<.0001
2	Profession		5	2	36.6371		<.0001
3	Marital status		1	3	9.7796		0.0018
4	dependent		1	4	7.8445		0.0051
5	Location		3	5	12.6475		0.0055

Conclusion:

At the end of step 5, we were left with five variables:

1) Annual income
2) Profession
3) Marital status
4) Dependent members in family
5) Location

Thus, there is a relationship between socio-demographic factors such as: Individual’s Annual income, Profession, Marital status, Dependent members in family, Location and their purchase of Mediclaim policy.

C) RESIDUAL CHI-SQUARE:

Hypothesis:

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
28.5229	18	0.0545

Conclusion:

Since the p-value greater than 0.05 (significantly large), we do not reject H_0 hence the reduced model is as good as full model and we proceed with 5 variables given by stepwise selection procedure.

D) WALD'S STATISTIC (INDIVIDUAL TESTING):

Hypothesis:

H_0: $\beta_3=\beta_5=\beta_7= \beta_9= \beta_{11}= 0$

H_1: At least one $\beta i \neq 0$ where i =3,5,7,9,11.

Test Statistics:

Under H_0 the following test statistics follows standard normal distribution.

Test Criterion:

Reject H_0, if p-value < 0.05

Type 3 Analysis of Effects			
Effect	DF	Wald’s χ2	Pr > χ2
Marital status	1	15.2588	<.0001
Location	3	11.9865	0.0074
Profession	5	17.6300	0.0034
annual_income	4	25.3817	<.0001
dependent	1	8.2906	0.0040

Conclusion:

Since p-value of all variables is less than 0.05. Thus, the variables X_3, X_5, X_7, X_9, X_{11} are significant at 5% level of significance.

Parameter estimation and individual’s testing:

To test the hypotheses:

H_0: Individual coefficients of independent variables are zero i.e., $\beta i=0$

H_1: Individual coefficients of independent variables are not zero i.e., $\beta i\neq0$

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald’s χ2	Pr > χ2
Intercept		1	-1.9428	0.5843	11.0555	0.0009
Marital Status	1	1	1.8284	0.4681	15.2588	<.0001
Location	1	1	-0.9812	0.4506	4.7426	0.0294
Location	2	1	0.5096	0.3516	2.1009	0.1472
Location	3	1	0.5830	0.3880	2.2578	0.1329
Profession	1	1	0.9994	0.3218	9.6468	0.0019
Profession	2	1	-0.3575	0.6708	0.2840	0.5941
Profession	3	1	1.8647	0.7927	5.5334	0.0187

Profession	4	1	0.9953	0.7184	1.9196	0.1659
Profession	5	1	-0.2345	0.6057	0.1499	0.6986
annual_income	1	1	1.2268	0.5726	4.5895	0.0322
annual_income	2	1	1.5421	0.5693	7.3376	0.0068
annual_income	3	1	2.2906	0.6047	14.3511	0.0002
annual_income	4	1	2.5917	0.6113	17.9717	<.0001
dependent		1	-0.2881	0.1000	8.2906	0.0040

Conclusion:

Since p-value > 0.05 for Location (Thane, Others), Profession (Government Job, Business, Unemployed). Hence, we do not reject H_0 5% level of significance.

e) ODDS RATIO:

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

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
status Married vs Unmarried	6.224	2.487	15.578
Location Navi Mumbai vs Mumbai	0.375	0.155	0.907
Location Thane vs Mumbai	1.665	0.836	3.316
Location Others vs Mumbai	1.791	0.837	3.832
profession Private job vs Student	2.717	1.446	5.104
profession Gov. job vs Student	0.699	0.188	2.605
Profession Self-employed vs Student	6.454	1.365	30.520
profession Business vs Student	2.706	0.662	11.060
profession Unemployed vs Student	0.791	0.241	2.593
annual_income 1-3 lakhs vs <1lakhs	3.410	1.110	10.476
annual_income 3-6 lakhs vs <1lakhs	4.675	1.532	14.267
annual_income 6-9 lakhs vs <1lakhs	9.881	3.021	32.320
annual_income >9 lakhs vs <1 lakhs	13.352	4.029	44.252
Dependent Members in family	0.750	0.616	0.912

Conclusion:

- Marital status
  - ❖ Married people are 6.224 times more likely to buy Mediclaim than Unmarried people.
- Location
  - ❖ People living in Navi Mumbai are 0.375 times less likely to buy Mediclaim than people living in Mumbai.
  - ❖ People living in Thane are 1.665 times more likely to buy

Mediclaim than people living in Mumbai.

  - ❖ People living in places other than Mumbai, Navi Mumbai, Thane are 1.791 times more likely to buy Mediclaim than people living in Mumbai.

- Profession
  - ❖ People working in private sector are 2.717 times more likely to buy Mediclaim than students.
  - ❖ People working in government sector are 0.699 times less likely to buy Mediclaim than students.
  - ❖ Self-employed people are 6.454 times more likely to buy Mediclaim than students.
  - ❖ People doing business are 2.706 times more likely to buy Mediclaim than students.
  - ❖ Unemployed people are 0.791 times less likely to buy Mediclaim than students.
- Annual income
  - ❖ People whose annual income is between 1 to 3 lakhs are 3.410 times more likely to buy Mediclaim than people whose annual income is less than 1 lakh.
  - ❖ People having annual income between 3 to 6 lakhs are 4.675 times more likely to buy Mediclaim than people whose annual income is less than 1 lakh.
  - ❖ People whose annual income is between 6 to 9 lakhs are 9.881 times more likely to buy Mediclaim than people whose annual income is less than 1 lakh.
  - ❖ People whose annual income is more than 9 lakhs are 13.352

times more likely to buy Mediclaim than people whose annual income is less than 1 lakh.

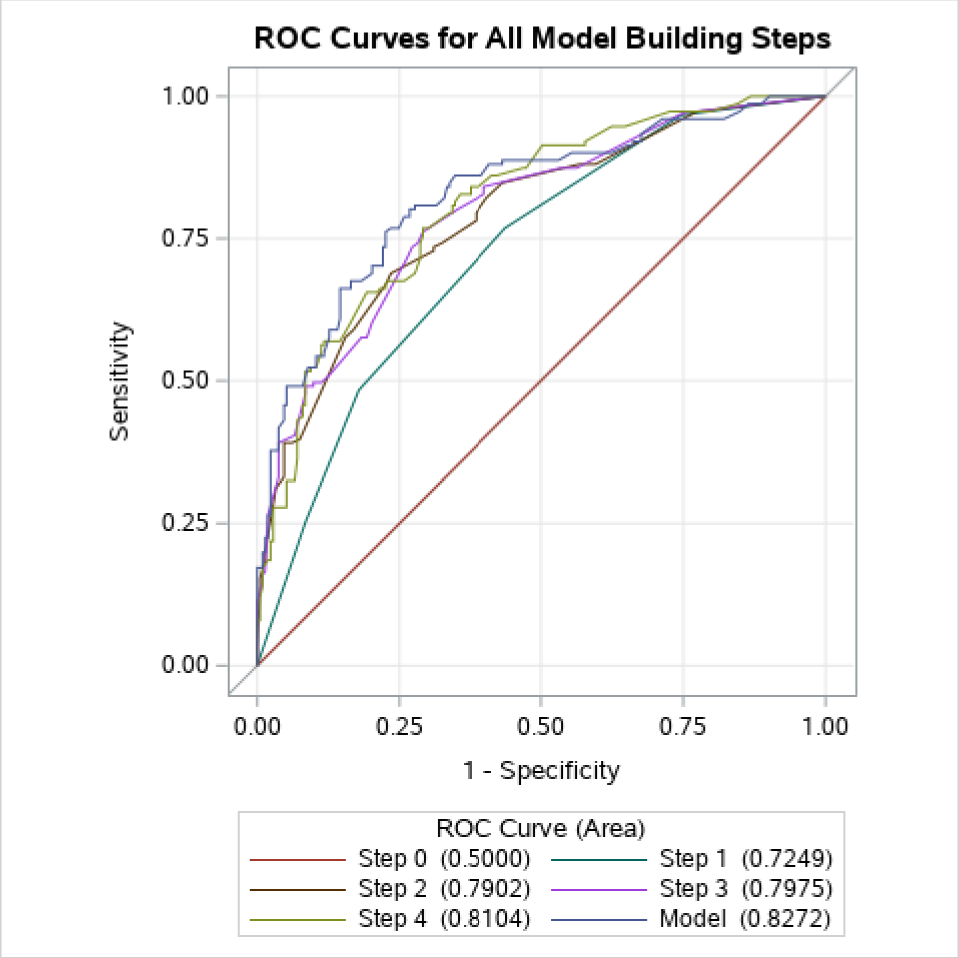
f) ROC CURVE:

The Receiver Operating Characteristic (ROC) curve is a graphical representation of the trade-off between the false positive and true positive rates for various cut-off values.

Y- axis: Sensitivity

X-axis: 1-Specificity

The accuracy of the model can be me measured by area under the ROC curve (C).



Association of Predicted Probabilities and Observed Responses

Percent Concordant	82.5	**Somers' D**	0.654
Percent Discordant	17.1	**Gamma**	0.657
Percent Tied	0.4	**Tau-a**	0.319
Pairs	32163	**c**	0.827

Conclusion:

Area under the ROC curve is estimated by the statistic ‘c’ in the “association of predicted probabilities and observed responses” table. Hence, the area under the ROC curve is 0.827. Since, the ROC curve rises quickly i.e., both sensitivity and specificity are high. Hence the model has high predictive accuracy.

g) HOSMER & LEMESHOW (TEST FOR GOODNESS OF FIT):

Hypothesis:

H_0 : Model is a good fit.

H_1 : Model is not a good fit.

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
12.4262	8	0.1332

Conclusion:

Since p-value is greater than 0.05, we do not reject H_0 and conclude that the fitted model is good.

Model Validation (Using confusion matrix):

Using R software, we calculate Confusing Matrix as shown below:

Confusion Matrix

Actual	Predicted	
	Yes	No
Yes	38 (True Positive)	10 (False Negative)
No	27 (False Positive)	81 (True Negative)

Error Rate:

$$\frac{False\ Positive\ +\ False\ Negative}{Total\ number\ of\ observations}$$

Here, False Positive = 27; False Negative = 10;

Total no. of observations = 156

Error Rate = (27+10) / 156

= 0.2371

Thus, Error rate for Logistic Regression model is 23.71%.

Fitted Model:

$$g(x) = -1.9428 + 1.8284(X_{3.1}) - 0.9812(X_{5.1}) + 0.9994(X_{7.1}) + 1.8647(X_{7.3}) + 1.2268(X_{9.1}) + 1.5421(X_{9.2}) + 2.2906(X_{9.3}) + 2.5917(X_{9.4})$$

OBJECTIVE II

To check the association between health factors and purchase of Mediclaim policy.

CHI-SQUARE TEST FOR ASSOCIATION

This test is also called as Pearson's Chi-Square test or the Chi-Square test of independence. It is used to discover if there is a relationship between two categorical variables. The two variables should be measured at an ordinal or nominal level (i.e., categorical data). The two variables should consist of two or more categorical, independent groups. The procedure involves comparing the observed cells frequencies with expected cell frequencies. Expected frequencies are number of cases that should fall in each cell if there is no relationship between the two categorical variables.

Test statistic:

$$\chi^2_{\text{cal}} = \frac{\sum(\text{Oi}-\text{Ei})^2}{\text{Ei}}$$

Where,

Oi : Observed frequency

Ei : $\frac{(Row\ total * Column\ total)}{N}$

N: Total number of observations

Decision Criteria: Reject H_0, if p-value < 0.05

If the calculated chi-square value is greater than the tabulated chi-square value, then the null hypothesis is rejected and we conclude that, the two variables under consideration are not independent.

A) Blood Pressure and Policy

Hypothesis:

H_0: There is no association between Blood pressure and individual's purchase of Mediclaim policy.

H_1: There is association between Blood pressure and individual's purchase of Mediclaim policy.

Table of Observed and Expected Frequencies:

Health factor	Have Mediclaim or not		
	Yes	No	Total
Blood pressure	11 (8.31)	9 (11.81)	20
No Blood pressure	205 (207.6)	295 (292.31)	500
Total	216	304	520
Chi-Sq = 1.552, DF= 1, p-value=0.213			

Conclusion:

Since the p-value is greater than 0.05, we do not reject H_0 and conclude that There is no association between Blood pressure and individual's purchase of Mediclaim policy.

B) Diabetes and Policy

Hypothesis:

H_0: There is no association between Diabetes and individual's purchase of Mediclaim policy.

H_1: There is association between Diabetes and individual's purchase of Mediclaim policy.

Table of Observed and Expected Frequencies:

Health factor	Have Mediclaim or not		
	Yes	No	Total
Diabetes	11 (9.14)	11 (12.86)	22
No Diabetes	205 (206.86)	293 (291.14)	498
Total	216	304	520
Chi-Sq = 0.677, DF= 1, p-value=0.411			

Conclusion:

Since the p-value is greater than 0.05, we do not reject H_0 and conclude that There is no association between Diabetes and individual's purchase of Mediclaim policy.

C) Cholesterol and Policy

Hypothesis:

H_0: There is no association between Cholesterol and individual's purchase of Mediclaim policy.

H_1: There is association between Cholesterol and individual's purchase of Mediclaim policy.

Table of Observed and Expected Frequencies:

Health factor	Have Mediclaim or not		
	Yes	No	Total
Cholesterol	10 (7.06)	7 (9.94)	17
No Cholesterol	206 (208.94)	297 (294.06)	503
Total	216	304	520
Chi-Sq = 2.164, DF= 1, p-value=0.141			

Conclusion:

Since the p-value is greater than 0.05, we do not reject H_0 and conclude that There is no association between Cholesterol and individual's purchase of Mediclaim policy.

D) Thyroid and Policy

Hypothesis:

H_0: There is no association between Thyroid and individual's purchase of Mediclaim policy.
H_1: There is association between Thyroid and individual's purchase of Mediclaim policy.

Table of Observed and Expected Frequencies:

Health factor	Have Mediclaim or not		
	Yes	No	Total
Thyroid	11 (8.31)	9 (11.69)	20
No Thyroid	205 (207.69)	295 (292.31)	500
Total	216	304	520
Chi-Sq = 1.552, DF= 1, p-value=0.213			

Conclusion:

Since the p-value is greater than 0.05, we do not reject H_0 and conclude that There is no association between Thyroid and individual's purchase of Mediclaim policy.

E) Anemia and Policy

Hypothesis:

H_0: There is no association between Anemia and individual's purchase of Mediclaim policy.
H_1: There is association between Anemia and individual's purchase of Mediclaim policy.

Table of Observed and Expected Frequencies:

Health factor	Have Mediclaim or not		
	Yes	No	Total
Anemia	7 (5.82)	7 (8.18)	14
No Anemia	209 (210.18)	297 (295.82)	506
Total	216	304	520
Chi-Sq = 0.424, DF= 1, p-value=0.515			

Conclusion:

Since the p-value is greater than 0.05, we do not reject H_0 and conclude that there is no association between Anemia and individual's purchase of Mediclaim policy.

F) Asthma and Policy

Hypothesis:

H_0: There is no association between Asthma and individual's purchase of Mediclaim policy.
H_1: There is association between Asthma and individual's purchase of Mediclaim policy.

Table of Observed and Expected Frequencies:

Health factor	Have Mediclaim or not		
	Yes	No	Total
Asthma	8 (5.40)	5 (7.60)	13
No Asthma	208 (210.6)	299 (296.40)	507
Total	216	304	520
Chi-Sq = 2.196, DF= 1, p-value=0.138			

Conclusion:

Since the p-value is greater than 0.05, we do not reject H_0 and conclude that there is no association between Asthma and individual's purchase of Mediclaim policy.

G) Surgery and Policy

Hypothesis:

H_0: There is no association between Surgery and individual's purchase of Mediclaim policy.
H_1: There is association between Surgery and individual's purchase of Mediclaim policy.

Table of Observed and Expected Frequencies:

Health factor	Have Mediclaim or not		
	Yes	No	Total
Surgery	41 (30.74)	33 (43.26)	74
No Surgery	175 (185.26)	271 (260.74)	446
Total	216	304	520
Chi-Sq = 6.832, DF= 1, p-value=0.009			

Conclusion:

Since the p-value is less than 0.05, we reject H_0 and conclude that there is association between Surgery and individual's purchase of Mediclaim policy.

H) Critical illness and Policy

Hypothesis:

H_0: There is no association between Critical illness and individual's purchase of Mediclaim policy.
H_1: There is association between Critical illness and individual's purchase of Mediclaim policy.

Table of Observed and Expected Frequencies:

Health factor	Have Mediclaim or not		
	Yes	No	Total
Critical illness	8 (6.65)	8 (9.35)	16
No Critical illness	208 (209.35)	296 (294.65)	504
Total	216	304	520
Chi-Sq = 0.487, DF= 1, p-value=0.485			

Conclusion:

Since the p-value is greater than 0.05, we do not reject H_0 and conclude that there is no association between Critical illness and individual's purchase of Mediclaim policy.

OBJECTIVE III

To check whether there is significant difference in the purchase of Mediclaim policy before and after covid-19

MCNEMAR TEST

The McNemar test is a non-parametric test for paired nominal data. It's used when you are interested in finding a change in proportion for the paired data. This test is sometimes referred to as **McNemar's Chi-Square test** because the test statistic has a chi-square distribution. The McNemar test is used to determine if there are differences on a dichotomous dependent variable between two related groups. It can be considered to be similar to the paired-samples t-test, but for a dichotomous rather than a continuous dependent variable.

The McNemar's test has three assumptions that must be met. If these assumptions are not met, you cannot use a McNemar's test, but may be able to use another statistical test instead.

Assumptions for McNemar test:

1. You must have one nominal variable with two categories (i.e. dichotomous variables) and one independent variable with two connected groups.
2. The two groups in your dependent variable must be mutually exclusive. In other words, participants cannot appear in more than one group.
3. Your sample must be a random sample.

Hypothesis:

H_0: there is no significant difference in purchase of Mediclaim policy before and after covid-19.

H_1: there is significant difference in purchase of Mediclaim policy before and after covid-19.

Test statistics:

$$\chi^2 = \frac{(b-c)^2}{(b+c)}$$

Decision criterion: Reject H_0, if $p\text{-value} < 0.05$

Purchase of Mediclaim	Have Mediclaim		
	No	Yes	Total
Before Covid-19	247	205	452
After Covid-19	57	11	68
Total	304	216	520

McNemar's Test			
Chi-Square	DF	Pr > ChiSq	Exact Pr >= ChiSq
83.6031	1	<.0001	<.0001

Conclusion:

Since, p-value is less than 0.05 we reject the null hypothesis and conclude that there is significant difference in the purchase of Mediclaim policy before and after covid-19.

OBJECTIVE IV

A. To find out preferable companies to purchase Mediclaim.

PARETO ANALYSIS

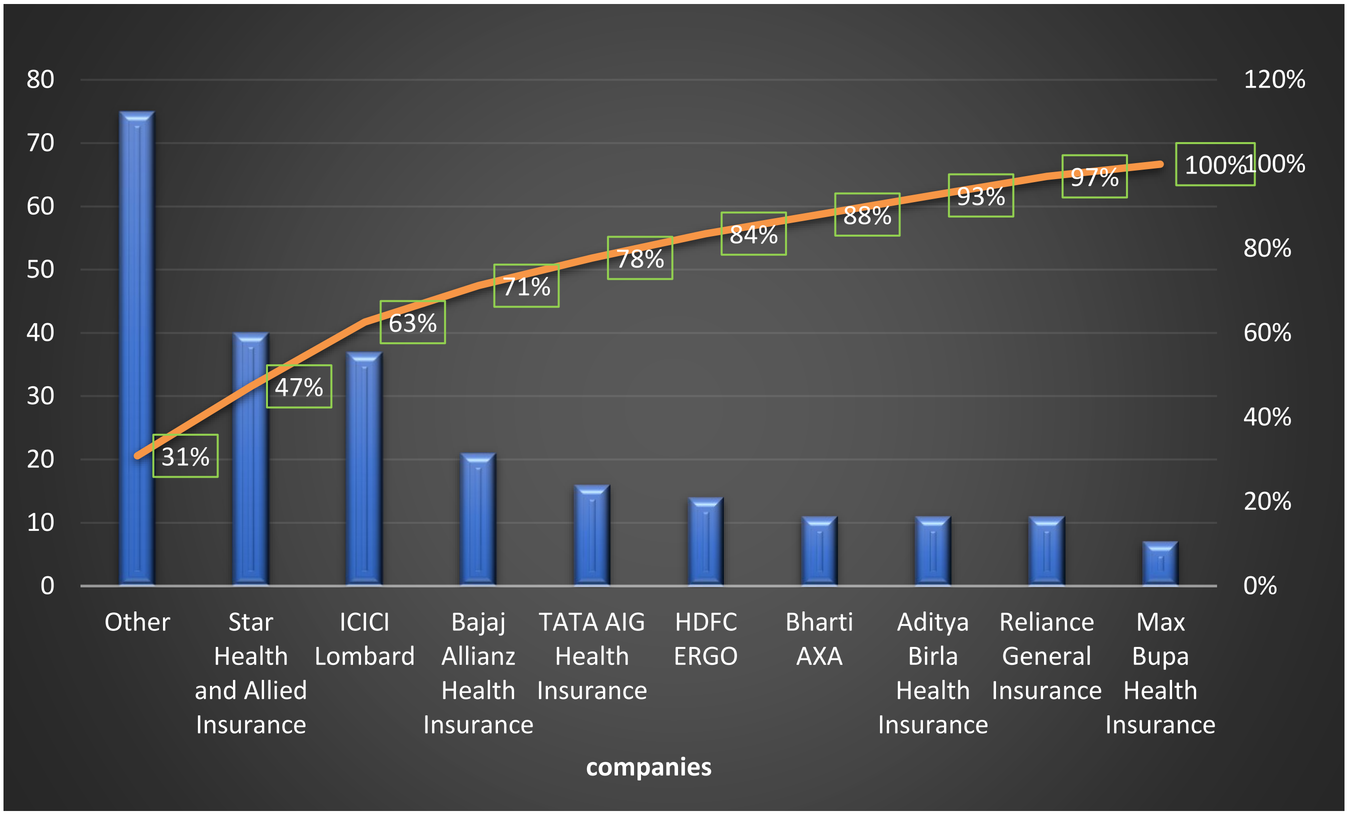
Pareto Analysis is a statistical technique in decision making, used for the selection of a limited number of tasks, that produce significant overall effect. It uses the Pareto Principle (also known as 80/20 rule), the idea that by doing 20% of the work, you can generate 80% of the benefit of doing the entire job. This technique helps to identify the top portion of causes, that need to be addressed to resolve the majority of problems. This is also known as the 'vital few' and the 'trivial many' effects.

A Pareto chart, also called a Pareto distribution diagram, is a vertical bar graph, in which values are plotted in decreasing order of relative frequency from left to right. The problem categories or causes are shown on the 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 bar 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'.

Pareto charts are extremely useful for analyzing what problems need attention first, because the taller bars on the chart, which represent frequency, clearly illustrate which variables have the greatest cumulative effect on a given system. The Pareto chart is one of the seven basic tools of quality control. The purpose of the Pareto chart is to highlight the most important among a (typically large) set of factors.

We know that there many different benefits of having Mediclaim policy. In our project we asked individuals if they have Mediclaim policy then from which company they bought it and what is the type of that policy. We tried to find out most preferred companies to purchase Mediclaim and most preferred type of policy/plan.

Preferred companies:

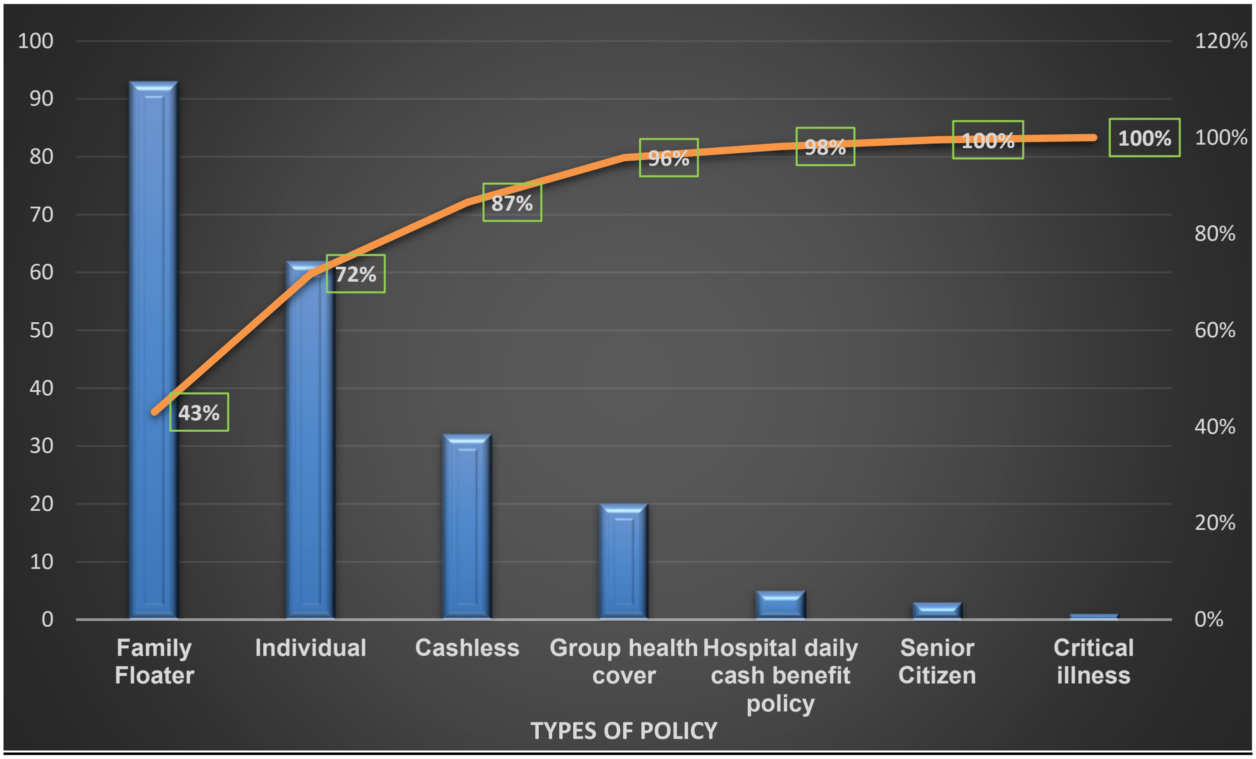


Conclusion:

From pareto analysis, we came to know that, most preferred Mediclaim companies by individuals are Star Health and Allied Insurance, ICICI Lombard, Bajaj Allianz Health Insurance, TATA AIG Health Insurance and Others

B. To find out preferable Mediclaim plan.

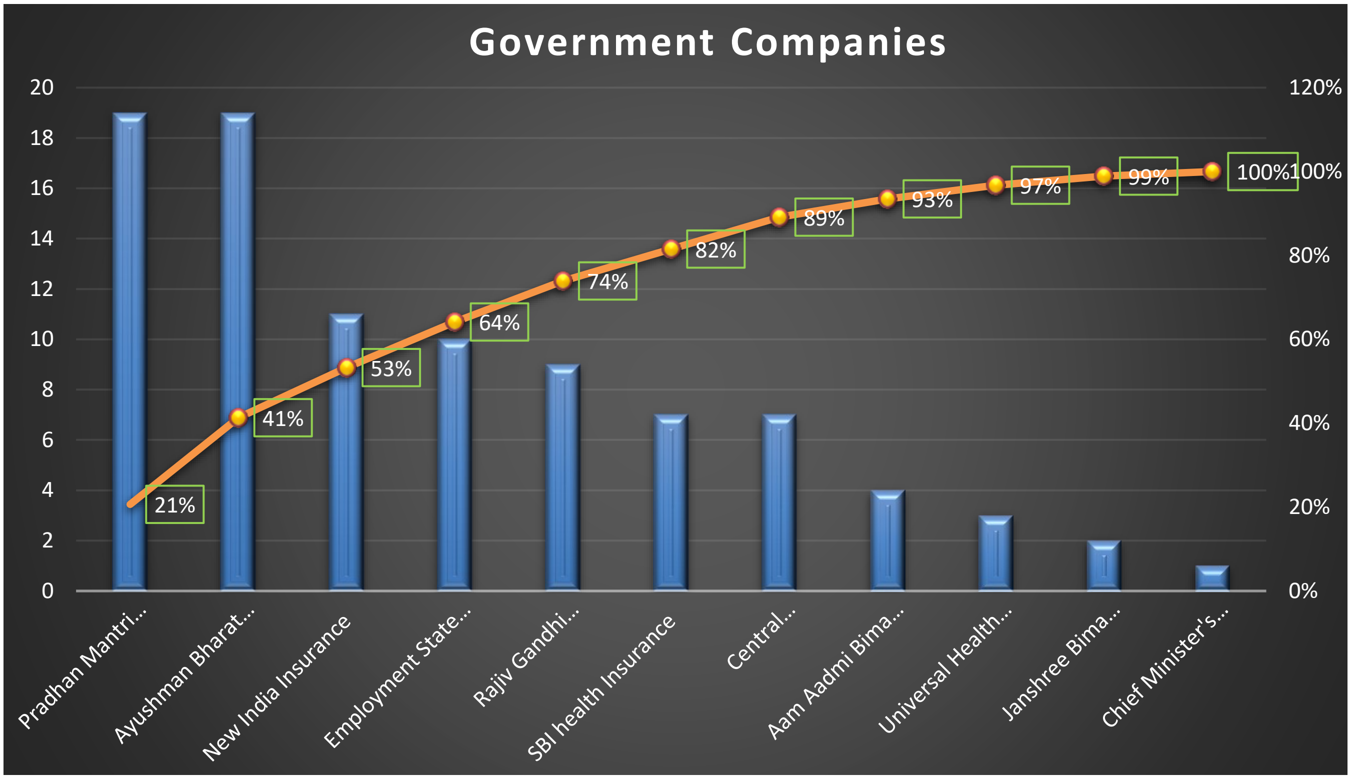
Preferable plans:



Conclusion:

From pareto analysis, we come to know that, most preferred type of policies by individual’s are Family floater and Individual type policy.

C. To find out preferable government companies to purchase Mediclaim.



<u>Conclusion</u>:

From pareto analysis, we came to know that, most preferred government Mediclaim companies by individuals are Pradhanmantri suraksha bima, Ayushman Bharat yojana, New India insurance, Employment state insurance and Rajeev Gandhi yojana.

FINAL CONCLUSION (PARETO ANALYSIS)

The most preferred companies to purchase Mediclaim by individuals are:

- Star Health and Allied Insurance
- ICICI Lombard
- Bajaj Allianz Health Insurance
- TATA AIG Health Insurance

The most preferred Type of policies by people are:

- Family Floater Plan
- Individual Insurance Plan

The most preferred government companies to purchase medical policy are:

- Pradhanmantri Suraksha Bima Yojana
- Aayushman Bharat Yojana
- New India Insurance
- Employment State Insurance
- Rajeev Gandhi Jeevandayee Arogya Yojana

Hence, the companies should try to improve the above plans by giving maximum coverage benefits to the people.

OBJECTIVE V

A. To determine the factors(reasons) affecting individual's decision to don't have policy.

FACTOR ANALYSIS

Factor analysis is a useful method of reducing data complexity by reducing the number of variables under study. In general, factor analysis is a set of techniques which analyses correlations between variables, and reduces their number into fewer uncorrelated and unobservable "factors" which explain much of the original data more economically. Thus, factor analysis is a general name denoting a class of procedures primarily used for data reduction and summarization.

We use the following statistics in factor analysis:

- Correlation (or Covariance) matrix: A Correlation (or Covariance) matrix is symmetric matrix showing the simple correlations (or covariance) between all possible pairs of variables included in the analysis.
- Communality: The portion of variance of ith variable contributed by the 'm' common factors is called ith communality.
- Eigen values: It represents the total variance explained by each factor.

PREREQUISITES:

Statistics for testing the appropriateness of the factor analysis.

a) <u>Kaiser-Meyer-Olkin (KMO) Measure:</u>

- The KMO measure is used for sampling adequacy.
- A small value of KMO statistic indicates the correlation between pairs of variables cannot be explained by the other variables and that factor analysis may not be appropriate.
- Generally, value greater than 0.5 is desirable.

b) <u>Bartlett's Test of Sphericity:</u>

- Bartlett's test of sphericity tests hypothesis that the correlation matrix is identity matrix.
- This test which is often done prior to factor analysis, tests whether the data comes from multivariate normal distribution with 0 co-variances.
- We proceed with factor analysis only if the above null hypothesis is rejected.

<u>Variables:</u>

X_1 : Can't afford it

X_2 : Don't think it's important, no reason to purchase

X_3 : Don't trust company

X_4 : Never thought of buying

X_5 : Employer sponsored cover is enough

X_6 : Bad previous experience

a. Correlation Matrix

		X_1	X_2	X_3	X_4	X_5	X_6
Correlation	X_1	1.000	.285	.399	.045	.088	.228
	X_2	.285	1.000	.419	.151	.228	.267
	X_3	.399	.419	1.000	.265	.222	.379
	X_4	.045	.151	.265	1.000	.194	.110
	X_5	.088	.228	.222	.194	1.000	.405
	X_6	.228	.267	.379	.110	.405	1.000

Determinant = .420

b. KMO and Bartlett Test:

Hypothesis:

H_0: Population correlation matrix is an identity matrix.

H_1: Population correlation matrix is not an identity matrix.

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.703
Bartlett's Test of Sphericity	Approx. Chi-Square	260.046
	Degree of freedom	15
	Significance	.000

Conclusion:

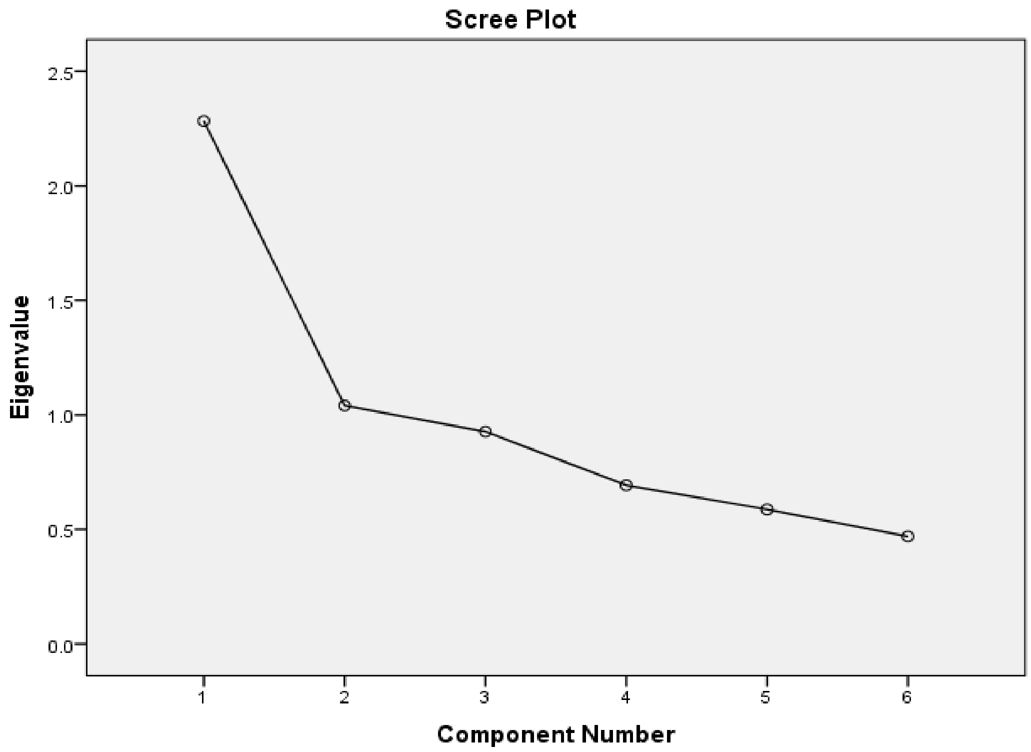
The value of KMO statistic is 0.730 (> 0.5) i.e., significantly large. Hence, we proceed with Factor Analysis as an appropriate technique of data reduction.

c. Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.283	38.053	38.053	2.283	38.053	38.053
2	1.041	17.344	55.397	1.041	17.344	55.397
3	0.927	15.446	70.843			
4	0.693	11.548	82.391			
5	0.587	9.786	92.177			
6	0.4697	0.823	100.000			

From the above table we can see that 2 factors were extracted using minimum eigenvalue criterion which explains 55.397% of the total variation.

d. Scree Plot:



From our data, we can observe an elbow after the 2nd principal component and hence we extract 2 factors. We Use the varimax rotation method of rotation and the following is rotated component matrix.

e. Rotated Component Matrix:

	Component	
	1	2
X_1	0.819	
X_2	0.652	
X_3	0.734	
X_4		0.611
X_5		0.795
X_6		0.572

Factor 1= (X_1, X_2, X_3)

Factor 1= (X_4, X_5, X_6)

Factor 1 **(Lack of budget & trust)**	**Factor 2** **(Lack of need)**
• Can’t afford it • Don’t think it’s important, no reason to purchase • Don’t trust company	• Never thought of buying • Employer sponsored cover is enough. • Bad previous experience

Conclusion:

6 reasons(variables) of not purchasing Mediclaim policy can be classified into two factors lack of budget & trust and lack of need. This account for the 55% of the variation.

B. To determine what are the best source to gain information about Mediclaim policy.

Variables:

X_1 : News paper

X_2 : Insurance agent

X_3 : TV

X_4 : Family/Friends

X_5 : Internet

a. Correlation Matrix

		X_1	X_2	X_3	X_4	X_5
Correlation	X_1	1.000	.483	.647	.517	.600
	X_2	.483	1.000	.495	.661	.627
	X_3	.647	.495	1.000	.488	.533
	X_4	.517	.661	.488	1.000	.686
	X_5	.600	.627	.533	.686	1.000

Determinant = 0.086

b. KMO and Bartlett Test:

Hypothesis:

H_0: Population correlation matrix is an identity matrix.

H_1: Population correlation matrix is not an identity matrix.

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.835
Bartlett's Test of Sphericity	Approx. Chi-Square	1265.372
	Degree of freedom	10
	Significance	.000

Conclusion:

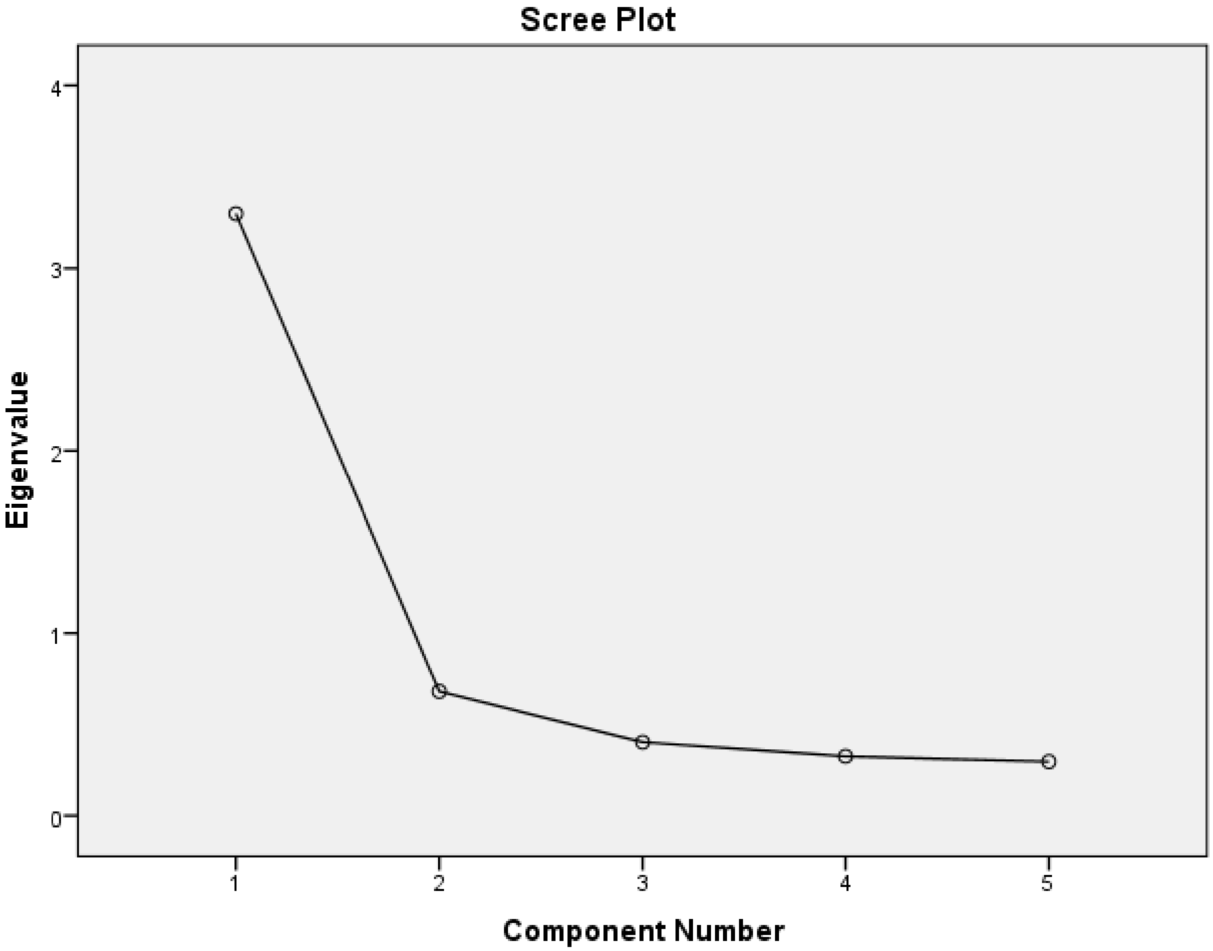
The value of KMO statistic is 0.835 (> 0.5) i.e., significantly large. Hence, we proceed with Factor Analysis as an appropriate technique of data reduction

c. Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.299	65.977	65.977	3.299	65.977	65.977
2	0.680	13.593	79.570	0.680	13.593	79.570
3	0.402	8.034	87.605			
4	0.324	6.485	94.090			
5	0.295	5.910	100.00			

From the above table we can see that 2 factors were extracted by using fix number of factors to extract in SPSS. Because by using minimum eigenvalue criterion only one factor was extracted.

d. Scree Plot:



From our data, we can observe an elbow after the 2nd principal component and hence we extract 2 factors. We use the varimax rotation method of rotation and the following is rotated component matrix.

e. Rotated Component Matrix:

	Component	
	1	2
X_1		0.841
X_2	0.842	
X_3		0.864
X_4	0.857	
X_5	0.753	

Factor 1= (X_2, X_4, X_5)

Factor 1= (X_1, X_3)

Factor 1 (Social sources)	Factor 2 (Old media)
• Insurance agent • Family/Friends • Internet	• News paper • TV

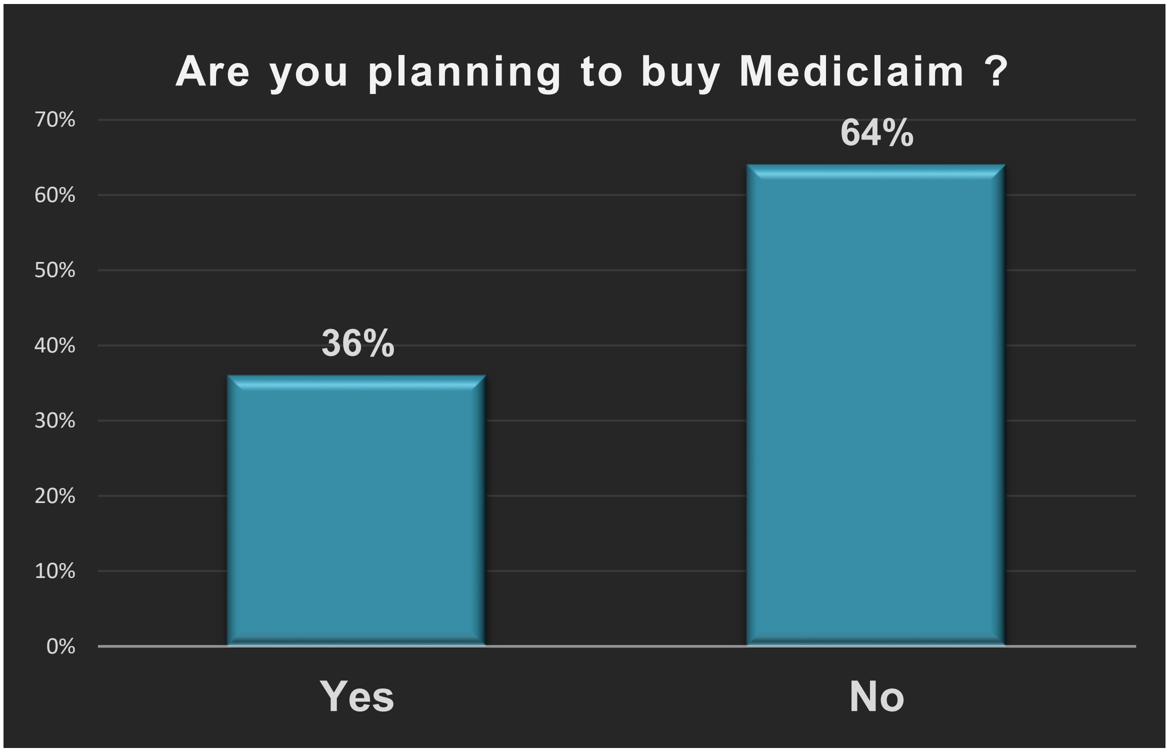
Conclusion:

5 variables can be classified into two factors social sources and old media. These accounts for 80% of the variation.

EXPLORATORY DATA ANALYSIS

In current situation, people are undertaking various health and safety measures to avoid infection including self-isolation and social distancing. But this can't ensure forever health security. It may also create a different situation giving tough time to the individuals and people related to them. it is important to be equipped in advance by ensuring oneself and one's loved ones with all including Mediclaim policy to cover costs incurred on treatments arising in these kinds of situations.

So, we asked the respondent if they don't have Mediclaim policy, are they planning to buy one in the future?



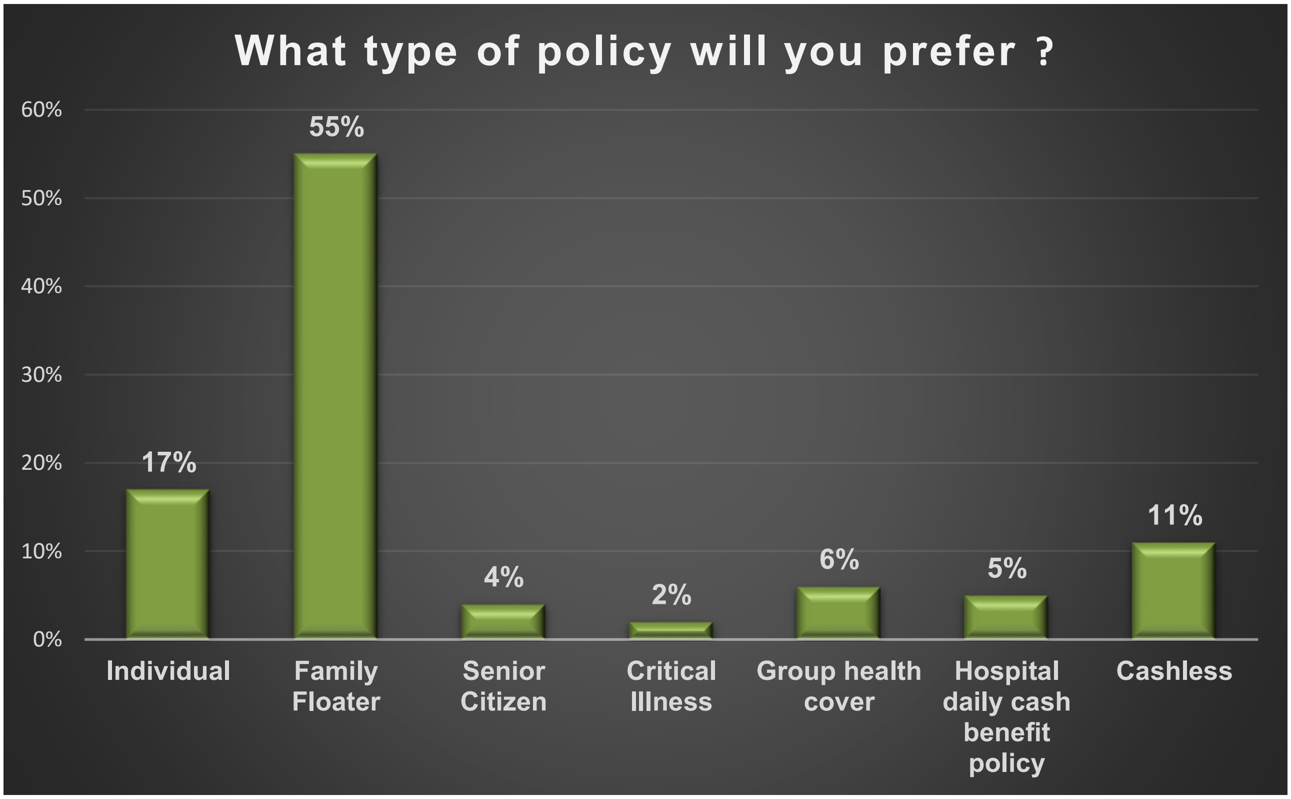
Conclusion:

As we can see that 36% of the people want to buy Mediclaim policy.

Next, we asked them what type of policy will they prefer to buy from following policies:

- Individual
- Family Floater
- Senior Citizen
- Critical Illness
- Group Health Cover
- Hospital daily cash benefit policy
- Cashless

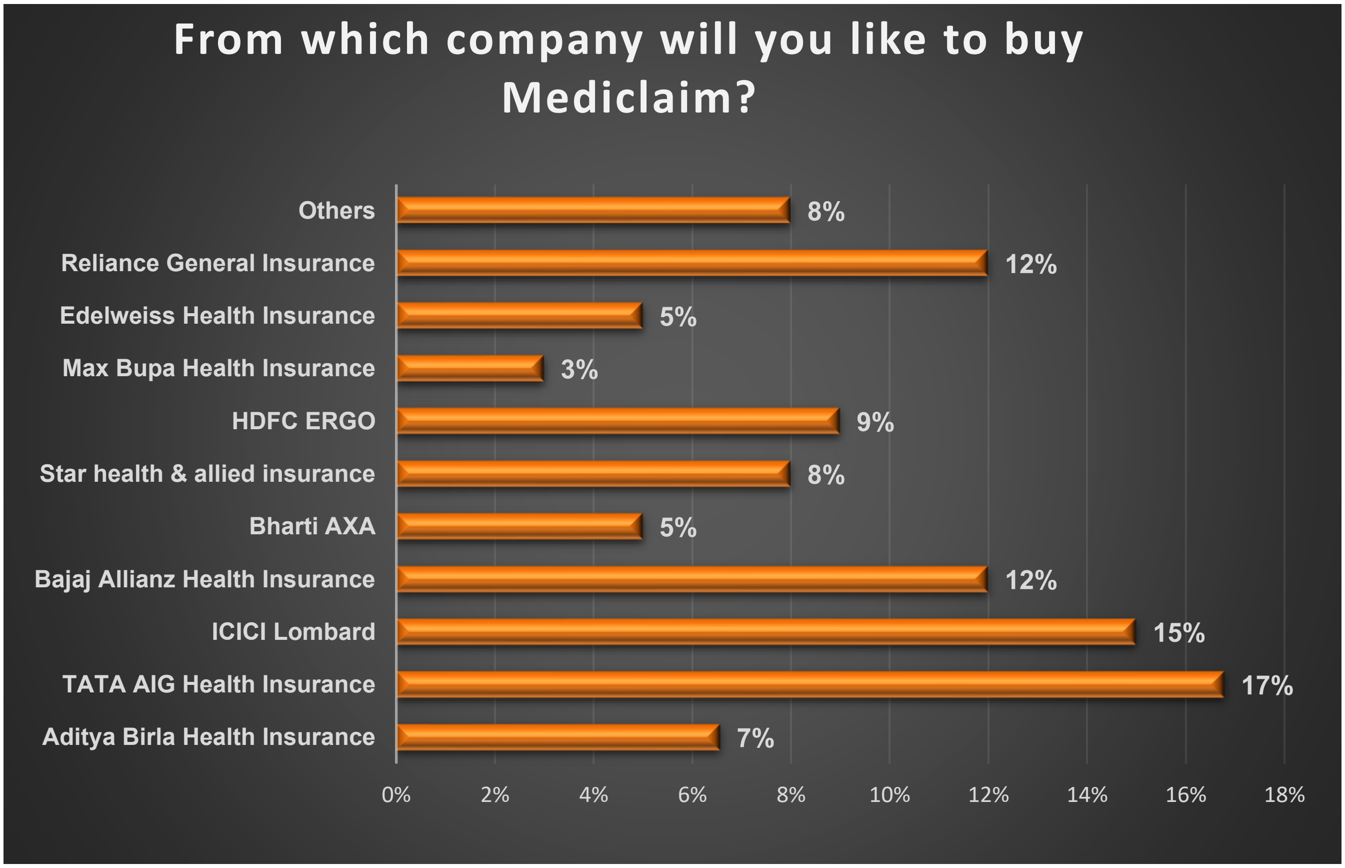
Thus, preferences are shown in the graph below:



<u>Conclusion:</u>

55% of the people prefer to buy Family Floater type of policy followed by 17% who wish to buy Individual type of Mediclaim policy.

Also, we asked them from which company they will like to buy policy. And the results are given below:



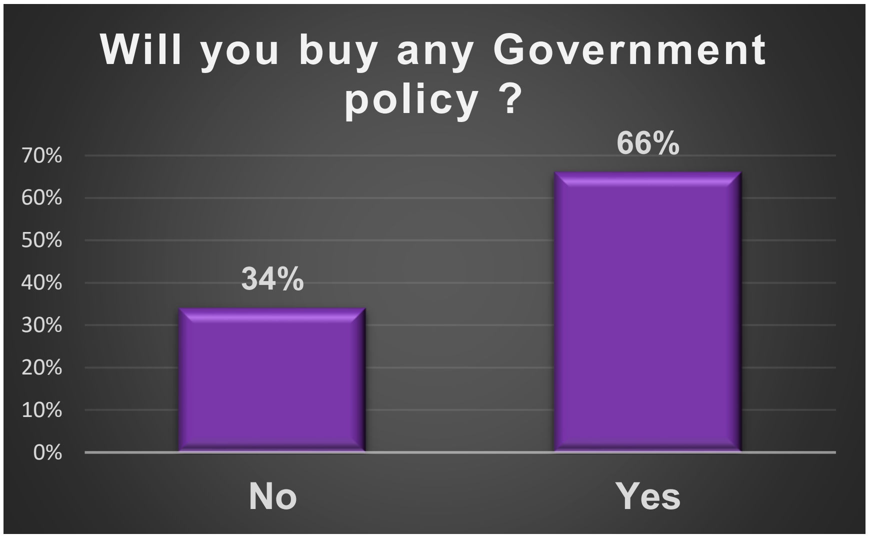
<u>Conclusion:</u>

The top 5 companies that people will like to purchase Mediclaim are:

- TATA AIG Health Insurance
- ICICI Lombard
- Bajaj Allianz Health Insurance
- Reliance General Insurance
- HDFC ERGO

Government of India offers many central or state government powered schemes that are designed to provide adequate health cover at a low-priced insurance cover.

So, we asked them if they want to buy any Government policy?



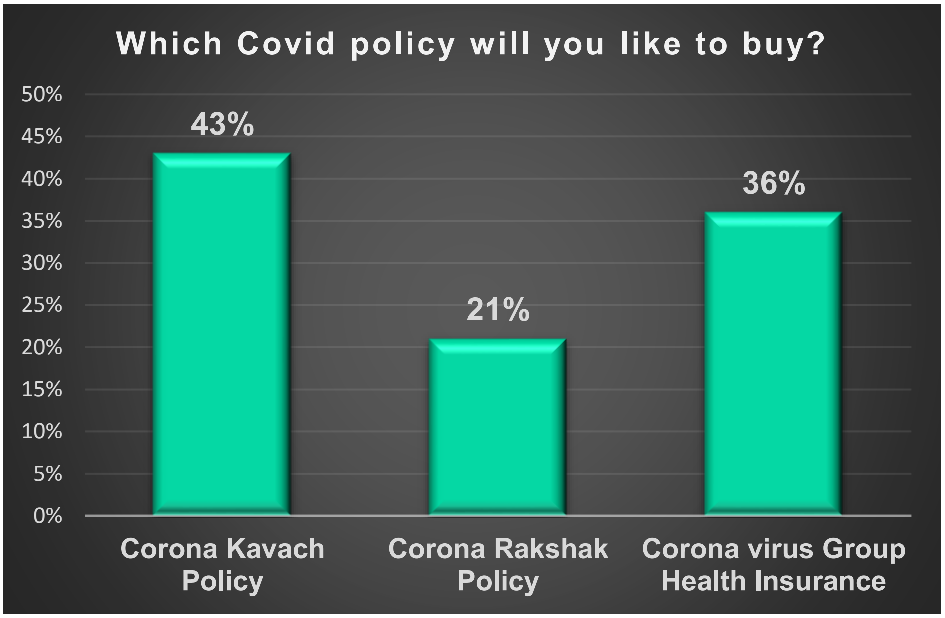
Conclusion:

66% of the people would like to buy Government policy.

Since, the outbreak of coronavirus, many insurers have been offering specific covid-19 insurance plans, corona Kavach and corona Rakshak are two such indemnity and benefit-based plans that have been popular with people in India.

So, we asked the respondent if they like to buy any Covid policy, which policy will they prefer from following policies:

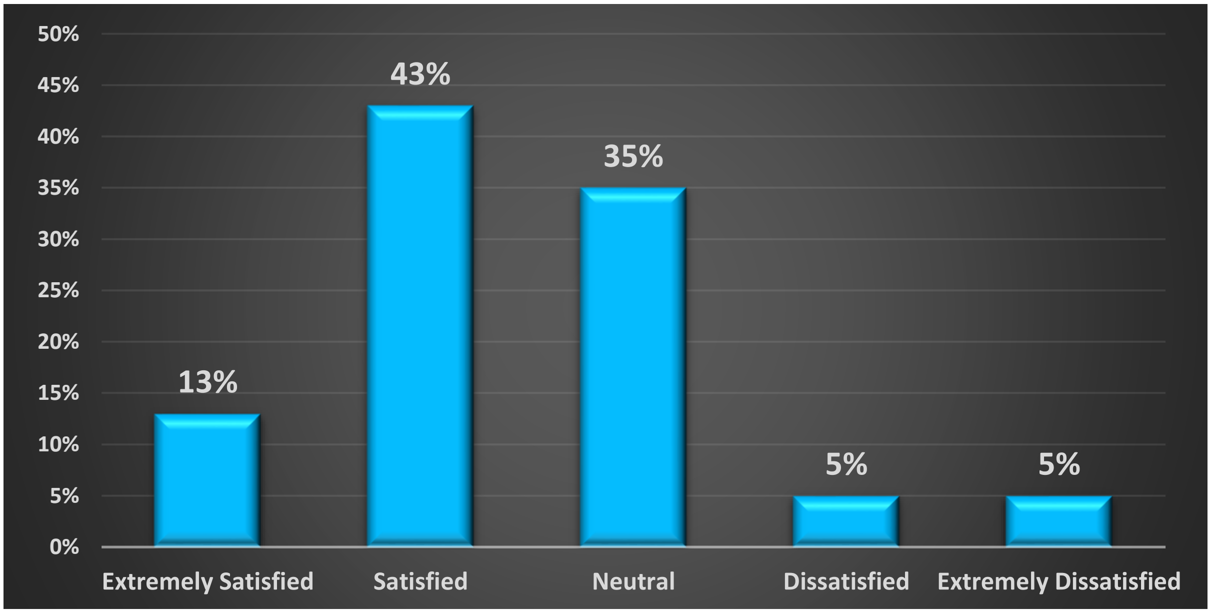
- Corona Kavach Policy
- Corona Rakshak Policy
- Corona virus Group Health Insurance



Conclusion:

43% of the people would like to buy corona Kavach policy followed by 36% of the people would prefer to buy Corona Virus Group Health Insurance.

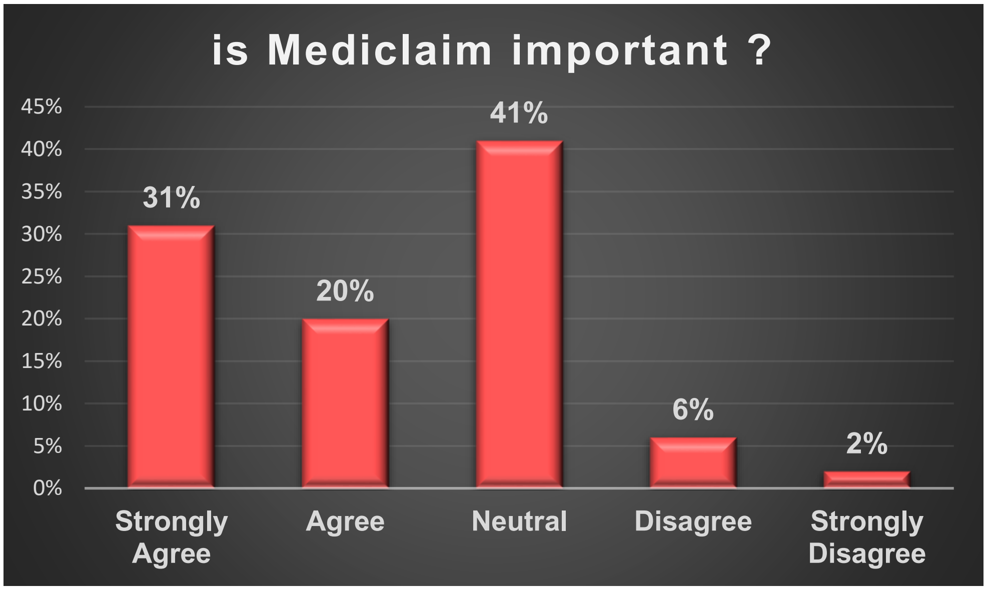
We asked the respondents to give ratings for how satisfied they are with the policy they have



Conclusion:

43% of the people are satisfied with their policy followed by 35% of the people who have neutral thought about satisfaction of their policy.

We asked the respondents to give ratings for importance of Mediclaim



Conclusion:

As we can see that total 51% people agrees that Mediclaim is very important.

While covid-19 has taught us many things, like basic hygiene factors or social distancing, the one very important thing is that we should not ignore health insurance any further.

OVERALL CONCLUSION

1) 51% of the individuals thinks that Mediclaim is very important.
2) From Logistic Regression, we get to know that, individual having Mediclaim or not mainly depend on their Marital status, Location, Profession, Annual Income and Dependent members in their family.
3) Chi-square Test shows that there is no association between various diseases such as Blood Pressure, Cholesterol, Anemia, Thyroid, Asthma, Diabetes, Critical Illness and individual having Mediclaim or not.

But there is association between individual's undergoing surgery and having Mediclaim or not.

4) From McNemar Test, we get to know that there is significant difference in the purchase of Mediclaim policies before and after Covid-19.
5) Pareto Analysis shows that most preferred Mediclaim companies by people in India are:
    - Star Health and Allied Insurance
    - ICICI Lombard
    - Bajaj Allianz Health Insurance
    - TATA AIG Health Insurance

From Pareto Analysis, we also get to know that, people mostly prefer 'Family Floater' and 'Individual' type of policies.

Also, pareto analysis shows that most preferred government Mediclaim companies by people in India are:

- Pradhanmantri Suraksha Bima Yojana
- Aayushman Bharat Yojana
- New India Insurance
- Employment State Insurance
- Rajeev Gandhi Jeevandayee Arogya Yojana

6) From Factor Analysis Results,

❖ 6 Reasons for not purchasing Mediclaim can be classified into 2 factors: ‘Lack of budget and trust’ & ‘Lack of need’.

These accounts for 55% of the variation.

❖ 5 Sources to gain information about Mediclaim can be classified into 2 factors: ‘Social sources’ and ‘Old media’

These accounts for 80% of the variation.

APPENDIX

Binary Logistic Regression Codes:

Splitting data into Training data and Testing Data in R:

```
> medi=read.csv("logisticDATA2.csv")

> library("caTools")

> split = sample.split(medi$policy, SplitRatio = 0.7)

> train = subset(medi, split==TRUE)

> test = subset(medi, split==FALSE)
```

Exporting train and test data as Excel File:

```
> library(writexl)

package ‘writexl’ was built under R version 4.0.5

> write_xlsx(train,"C:\\Desktop\\train.xlsx")

> write_xlsx(test,"C:\\Desktop\\test.xlsx")
```

Logistic Regression (Yes=1, No=0) [Full Model]

```
proc logistic data=train desc;

class gender status city area qualification profession
family_type

annual_income diseases smoke drink ;

model model policy(event='1')=age gender status city area
qualification profession
```

```
     family_type annual_income diseases smoke drink earningM
dependentM;

run;
```

Stepwise Regression with ROC Curve [Reduced Model]

```
ods noproctitle;

ods graphics / imagemap=on;

proc logistic data=WORK.IMPORT plots=(roc);

     class gender(ref='0') status(ref='0') city(ref='0')
area(ref='0') qualification(ref='0') profession(ref='0')
family_type(ref='0') annual_income(ref='0') diseases(ref='0')
smoke(ref='0') drink(ref='0') / param=ref;

model policy(event='1')=gender status city area qualification
profession family_type annual_income diseases smoke drink Age
earningM dependentM / link=logit lackfit selection=stepwise
slentry=0.05 slstay=0.05 hierarchy=single technique=fisher;

run;
```

Predicting On Test Dataset:

```
> model = glm(policy ~
Age+gender+status+area+city+qualification+profession+family_type+
annual_income+earningM+dependentM+diseases+smoke+drink,
family=binomial(link="logit"), data=test)

> summary(model)

> pmodel=predict(model,test)

> tab= table(pmodel>0.5, test$policy)

> tab
```

BIBLIOGRAPHY

a) Referred books:

- Applied Multivariate Statistical Analysis,6th edition, Jhonson R. A. and Wichern D. W. (2015).
- Applied Logistic Regression (Second Edition). -Hosmer, D. and Lemeshow, S. (2000).
- Factor Analysis as a Statistical Method. Second Edition. -Lawley, D. N., Maxwell, A. E. (1971).

b) Sites Used:

- https://www.wikipedia.org/
- https://stats.idre.ucla.edu/r/dae/
- https://www.policyx.com/health-insurance/

c) Software Used:

- SAS
- R Software
- Minitab
- SPSS
- MS – Office
- Python

QUESTIONNAIRE

Q1. Age □

Q2. Gender

- ○ Female
- ○ Male
- ○ Prefer not to say
- ○ Other: ________

Q3. Status

- ○ Unmarried
- ○ Married
- ○ Widow
- ○ Divorcee

Q4. In which city do you live?

- ○ Mumbai
- ○ Navi Mumbai
- ○ Thane
- ○ Other: ________

Q5. Which of the following best describes the area you live in?

- ○ Rural
- ○ Urban
- ○ Suburban

Q6. Qualification

- ○ SSC and below
- ○ HSC
- ○ Bachelor' Degree
- ○ Master' Degree
- ○ Doctorate Degree
- ○ Diploma
- ○ Other Professional Course

Q7.Profession

○ Student
○ Private Job
○ Government Job
○ Self Employed
○ Business
○ Unemployed

Q8. Do you live in?

○ Joint Family
○ Small/Nuclear Family
○ Alone

Q9. Family Annual Income

○ < 1 Lakh
○ 1-3 Lakhs
○ 3-6 Lakhs
○ 6-9 Lakhs
○ >9 Lakhs

Q10. Number of earning member of your family

□

Q11. Number of dependent members in the family

□

Q12. Do you have these habits?

	Yes	No
○ Smoking		
○ Drinking		

Q13. Do you have?

	Yes	No
○ Blood Pressure		
○ Diabetes		
○ Cholesterol		
○ Thyroid		
○ Anemia		
○ Asthma		
○ Critical Illness		

Q14. Have you ever undergone surgery?

○ Yes
○ No

Q15. Which is the most used source to pay your medical expenses?

○ Free medical Service from Government
○ Own savings
○ Paid by Employer/company
○ Health insurance / Medical policy

Q16. Is Health Insurance/ Mediclaim important?

○ Strongly agree
○ Agree
○ Neutral
○ Disagree
○ Strongly Disagree

Q17. Which of the following would you suggest someone to gain information regarding health insurance/Mediclaim policy?

	Strongly agree	agree	neutral	disagree	Strongly disagree
○ Newspaper					
○ Insurance Agent					
○ TV					
○ Family/ Friends					
○ Internet					

Q18. When do you get to know about Mediclaim?

○ Before Covid-19
○ After Covid-19

Q19. Do you have Health Insurance/ Mediclaim policy?

○ Yes
○ No

<u>If said Yes to Q.19,</u>

<u>Have Mediclaim policy</u>

Q1. Which type of policy/Insurance you have?

○ Individual
○ Family Floater
○ Senior Citizen
○ Critical Illness
○ Group Health Cover
○ Hospital Daily Cash Benefit Policy
○ Cashless

Q2. From which company you have purchase your policy?

- ○ ICICI Lombard
- ○ Bajaj Allianz Health Insurance
- ○ Bharti AXA
- ○ Star health & allied insurance
- ○ Aditya Birla Health Insurance
- ○ HDFC ERGO
- ○ Max Bupa Health Insurance
- ○ Edelweiss Health Insurance
- ○ Reliance General Insurance
- ○ TATA AIG Health Insurance
- ○ Other:__________

Q3. When did you purchase Mediclaim?

- ○ Before Covid-19
- ○ After Covid-19

Q4. Do you have any Government Mediclaim policy/Health Insurance? *

- ○ Yes
- ○ No

Q5. Which Government policy do you have? *

- ○ Ayushman Bharat Yojana
- ○ Pradhan Mantri Suraksha Bima Yojana
- ○ Aam Aadmi Bima Yojana (AABY)
- ○ Central Government Health Scheme (CGHS)
- ○ Employment State Insurance Scheme
- ○ Janshree Bima Yojana
- ○ Mahatma Jyotiba Phule Jan Arogya Yojana
- ○ New India Insurance
- ○ Chief Minister's Comprehensive Insurance Scheme
- ○ Rajiv Gandhi Jeevandayee Arogya Yojana (RGJAY)
- ○ SBI health Insurance
- ○ Universal Health Insurance Scheme {UHIS)
- ○ West Bengal Health Scheme
- ○ Yeshasivini Health Scheme
- ○ Mukhyamantri Amrutam Yojana
- ○ Karunya Health Scheme

- ○ Telangana State Government Employees and Journalists Health Scheme
- ○ Dr.YSR Aarogyasri Health Care Trust
- ○ None

Q6. How satisfied are you with your Mediclaim policy/health insurance? *

- ○ Extremely dissatisfied
- ○ Dissatisfied
- ○ Neutral
- ○ Satisfied
- ○ Extremely satisfied

<u>If said No to Q.19</u>

<u>Don't have Mediclaim policy</u>

Q1. What is the reason that you don't have medical policy?

	Strongly agree	agree	neutral	disagree	Strongly disagree
○ Can't afford					
○ Not Important					
○ Don't trust					
○ Never thought of buying					
○ Employer sponsored covered					
○ Dissatisfied with previous coverage					

Q2. Are you planning to buy a medical policy? *

- ○ Yes

○ No

<u>If said Yes to above question Q.2,</u>
<u>If planning to buy Mediclaim policy</u>

Q1. Which type of policy/Insurance will you buy/prefer?

- Individual
- Family Floater
- Senior Citizen
- Critical Illness
- Group Health Cover
- Hospital Daily Cash Benefit Policy
- Cashless

Q2. From which of the following company you will prefer to buy your policy?

○ ICICI Lombard
○ Bajaj Allianz Health Insurance
○ Bharti AXA
○ Star health & allied insurance
○ Aditya Birla Health Insurance
○ HDFC ERGO
○ Max Bupa Health Insurance
○ Edelweiss Health Insurance
○ Reliance General Insurance
○ TATA AIG Health Insurance
○ Other

Q3. Will you buy any Government Mediclaim policy/Health Insurance?

○ Yes
○ No

Q4. If yes, then which Government policy will you buy?

○ Ayushman Bharat Yojana
○ Pradhan Mantri Suraksha Bima Yojana
○ Aam Aadmi Bima Yojana (AABY)
○ Central Government Health Scheme (CGHS)
○ Employment State Insurance Scheme

- SBI health Insurance
- Janshree Bima Yojana
- Mahatma Jyotiba Phule Jan Arogya Yojana
- New India Insurance
- Chief Minister's Comprehensive Insurance Scheme
- Rajiv Gandhi Jeevandayee Arogya Yojana (RGJAY)
- Universal Health Insurance Scheme {UHIS)
- West Bengal Health Scheme
- Yeshasivini Health Scheme
- Mukhyamantri Amrutam Yojana
- Karunya Health Scheme
- Telangana State Government Employees and Journalists Health Scheme
- Dr.YSR Aarogyasri Health Care Trust

Q5. Which COVID health insurance/ Mediclaim policy will you buy? *

- Corona Kavach Policy
- Corona Rakshak Policy
- Corona virus Group Health Insurance
- None