A PROJECT WORK SUBMITTED TO

# K. J. SOMAIYA COLLEGE OF SCIENCE AND COMMERCE, VIDYANAGAR, VIDYAVIHAR (E)

Re-accredited "A" Grade by NAAC

Autonomous, Affiliated to University of Mumbai

A STUDY ON

**A ₹un for Money**

TOWARDS THE PARTIAL FULFILMENT FOR THE DEGREE

OF

MASTER OF SCIENCE

IN

APPLIED STATISTICS

BY

Sangita Khade

M.Sc. Part II

APPLIED STATISTICS

DEPARTMENT OF STATISTICS

# -"A ₹un for Money"-

# K.J SOMAIYA COLLEGE of SCIENCE and COMMERCE

Affiliated by Mumbai University

"A" Grade by NAAC

## Department of Statistics

Vidyavihar, Mumbai – 400 077

# INVESTMENT- "A ₹un for Money"

**Submitted by:**                                                       **Under Guidance of:**

**Sangita Khade**                                                       **Prof. Asmita Kubal**

# CERTIFICATE OF AUTHENTICATION

This is to certify that the project entitled "*A ₹un for Money*" is a bonafide work of "*Sangita Khade" (1919174)* submitted to the K J Somaiya College of Science and Commerce in partial fulfillment of the requirement for the award of the degree of "M.Sc. in the subject of Applied Statistics.

(Name and sign)

External Mentor/supervisor

(Name and sign)

Internal mentor/supervisor

(Name and Sign)

Head of the department

# Declaration by the student

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

---------------------------------

Signature

Sangita Khade 1919174

Name of the student

Exam. Seat No.

# Examiner Approval sheet


 This dissertation/project report entitled "***A ₹un for Money***" by Sangita Khade is approved for the degree of Mater of Science in the subject of Applied Statistics.




Examiners


(Name and signature)

 1.--------------------------------------------




(Name and signature)

2.-----------------------------




Place:

Date:

# **<u>Acknowledgement</u>**

# **Index**

# [1.1] Abstract:

Investment is the most important thing today. Now-a-days people are earning handsomely. They have all right to invest and spend to some extent. But lack of financial education, put them in much more difficult situation. At present lot of investment avenues are available in the market with investor education. Investors can choose from a variety of instruments and assets. Building on organizational learning theory, we seek to identify the performance effects of security investments that arise from previous failures or external regulatory pressure. This is a project about the study of different types of investment avenues (Stock, Mutual Funds, Real estate, Metals (Gold/Silver etc.), Insurance, Banks, Post offices, PPF (Public Provident Fund), and NPS (National Pension Scheme) and understanding them deeply according to risk return behaviour, demographic factors of investors. Also to know the investment preferences among the investors depending upon many socio-demographic factors and to know the awareness level and problem faced during the investment in different investment avenues and to know about the proactive and reactive investors.

Today's trying economic condition have forced many difficult decisions for people. Most are making conservative decision that reflect a survival mode in investment market. During these difficult times, understanding what investors on an ongoing basis is critical for survival. To explore this study of investment, in this project we have collected data using the convenience sampling. The population of the study constituted all the people of age equal or above 18 years living in the region Mumbai, Navi Mumbai and Thane. Also in this project a total of **594 responses** were collected for the analysis purpose. Out of which **404 were investors** and **190 were non investors** in which 232 responses were from females and 362 from males. As we have collected data from three regions using the convenience sampling method, we got 228 observations from Mumbai, 166 from Mumbai Suburban, 164 were from thane.

1

# [1.2] Objective of the study

1. To find which is the most preferred investment method and factor influencing the Investment.

2. To compare the investment according to socio-demographic factor.

3. To know the relation between "Saving & Investment".

4. To study the pattern of salaried people working in different sectors.

5. To understand the investors behaviour on the basis of demographic factors, awareness, their risk & return in avenues.

6. To know the awareness of investment and problems faced by investors.

7. To understand the proactive & reactive behaviour of investors during uncertainty.

# [1.3] Sampling Design

The population of the study constituted all the people of age equal or above 18 years living in the region Mumbai, Navi Mumbai and Thane. India is a diverse country with a huge cultural, geographical and linguistic diversity. By considering time and some other factors, we know that study of all the population unit is difficult and time consuming task. Also in the absence of list of all units from the population, it was difficult to draw a representative sample from the population. As we know that collecting data from all people living in three regions was not possible so we decided to take sample from it. For probability sampling we should first know the whole population size, which we got from the Census data of 2011, but it was not appropriate for us, as it was 10 years old and also that data considers people of all age but we wanted only 18+.

In this project we have collected data using the convenience sampling. Convenience sampling (also known as availability sampling) is a specific type of non-probability sampling method that relies on data collection from population members who are conveniently available to participate in study. Convenience sampling is a type of sampling where the first available primary data source will be used for the research without additional requirements. In other words, this sampling method involves getting participants wherever you can find them and typically wherever is convenient. In convenience sampling no inclusion criteria identified prior to the selection of subjects. All subjects are invited to participate.

# [1.4] Data Collection Tools

Data collection tools refer to the devices/instruments used to collect data, such as a paper questionnaire or computer-assisted interviewing system. Case Studies, Checklists, Interviews, Observation sometimes, and Surveys or Questionnaires are all tools used to collect data.

Since, we were not getting responses by using only Google form. Therefore we have taken help of paper questionnaire method also

- 150 data from online Google form and
- 444 data from paper questionnaire method.

In this project a total of 594 responses were collected for the analysis purpose. Out of which 232 responses were from females and 362 from males. As we have collected data from three regions using the convenience sampling method, we got 228 observations from Mumbai, 166 from Mumbai Suburban, 164 from thane. Also we received 36 data from other regions using Google form.

**Types of data collected:**

- Demographic information or data e.g. age, gender, educational background, region, family size, income etc.
- Awareness of investment avenues.
- About their investment.
- Satisfaction of respondents.

# [2.1] Review of literature

Literature review is the process of going through the articles related to the topic of the project, which are published in journals, online databases, magazines, newspapers, books or any other source of information including online sources. Literature review helps us to know and understand the findings and views of earlier projects that have carried out in an area similar or related to topic of study. It also helps in understanding the data collection methods and the statistical tools used for analysing the data. The key findings and the conclusion drawn in various projects are of great help for any new project. A large body of literature is available in the area of investments related to institutional investment pattern, portfolio construction methods, portfolio performance evaluation, retirement planning, product preferences and many more associated topics. Studies have been carried out by researchers on the gender differences in allocation of assets, constituents of domestic savings, gender differences in knowledge about financial investments, investors risk tolerance, investors perception of various financial products. Some of the most insightful studies carried out in India and outside are given below.

1.  Umamaheshwari.S, Ashok Kumar (2011) carried out a study to understand the investment pattern and awareness level of individuals belonging to the salaried class from the city of Coimbatore. A structured questionnaire was used to collect the data during December 2010 to July 2011. Responses from 1000 respondents were collected over a period of eight months from the areas of Valparai, Pollachi, Metupalayam and Coimbatore. Statistical tools used for analysis were Chi Square test and ANOVA along with mean value calculation. The awareness level was divided into three types:low, medium and high. Classification was based on the mean score obtained for the respondent. It was found that the most preferred avenue of investment was provident fund followed by insurance gold and Jewellery. The financial product awareness level among men and women was low.

2.  Geetha. N and Ramesh M, (2012), studied the role of demographic factors in investment decisions. Response received from 475 respondents from Nagapattinam district of Tamil Nadu was used for analysis. The sampling method used was convenient sampling. A well-structured questionnaire was used to collect the data from the respondents. Statistical inference was drawn using ANOVA and Chi square tests. The demographic attributes included age, gender, education, occupation, income, savings size 40 and family size. The investment avenue considered for the study were gold, provident fund,life insurance, real estate, bank deposits, postal savings, mutual funds and equities. According to the study, riskprotection, safety of investment, rate of return and liquidity were main factors which influenced investment decisions. This is in line with the findings of Elder & Rudolph (2003).

3.  Suriya Murithi S, Narayanan B and Arivazhagan B (2012) carried out a study to understand the investors preference towards different investment alternatives available in the Indian market. Sample size was 100 respondents. Sampling method used was convenience sampling. Questionnaire was used to collect the data from respondents belonging to the city of Trichy. Simple percentage calculation, correlation andchi square test was used to analyse the data. The two main reasons for saving was for purchasing house and children"s education. Bank deposits was the most preferred investment avenue followed by mutualfunds, gold and post office savings.

4.  Ramanujam.V and Chitra Devi K (2012) conducted a study to analyse the impact of socio economic variables on the attitude of investors towards investments. The sample size for the study was 100 respondents from the city of Coimbatore. The sample consisted of respondents from different age group, educational back ground income level and with varied level of awareness about the financial products. A structured questionnaire was administered for collecting the responses of the respondent. Convenient sampling method was used for picking up the 53 respondents. To analyse the data,ANOVA, mean and Chi square tests were used. It was found that the occupation of the respondent and the frequency of making investment were not significantly associated. The study did not find any relation between annual savings of respondent and purpose of investment. It was also found that, there was no difference in the investment patterns of respondent"s form government, public and private

class of investors.The findings of this study were different from most of the other studies that found association between the nature of jobs and income levels with the investment pattern.

5. P. Parmashivivaiah, Puttaswamy and Ramya (2013) conducted a study in the city of Mysore, to understand he factors influencing investment decisions. The sample size for the study was 120 respondents. They used judgment and snow ball sampling to collect the data. The study was conducted in the first half of 2013 in the city of Mysore. Statistical tools used for analysing the data were Percentage, mean, standard deviation, Chi square test, F test, ANOVA and regression. Data was classified based on the demographic profile of the respondents. It was found that liquidity was the most important factor while choosing an investment portfolio as far as government employees and entrepreneurs were concerned.This is similar to the findings of Geetha. N and Ramesh M, (2012). Private employees and professionals gave equal priority to growth and liquidity. Women did not select investments on the basis of safety of the principal. This is in line with studies conducted by Annika, Sunden and Surrette (2009) and Bernasek (2002).

6. An Empirical Analysis on Investors' perception towards various investment avenues in Vellore city, Tamil Nadu and India (2015) . This article is therefore timely and fills the gap in investment avenues literature. Furthermore, by pooling small amounts of money into a single investment value, individual investors are able to participate in investment strategies that would have otherwise been financially unfeasible. The Indian investment industry is witnessing a rapid growth as a result of infrastructural development, manufacturing and service sector, in personal financial assets and rise in foreign participation. With the growing risk appetite, rising income, and increasing awareness. Mutual funds and shares are becoming preferred investment vehicles like bank fixed deposits and post office savings that are considered safe. The present empirical study is an attempt to examine the investors' perception towards various investment avenues in Vellore city, Tamil Nadu, India. Design/Methodology/Approach: Primary data using convenient sampling through questionnaire and interview method as well as secondary data from wide range of literature from various journal publications had been utilized. Frequency distribution, percentage analysis, Mann Whitney and Kruskal-Wallis Test were used to test the objectives of the study. In order to test the reliability of the scale, the researchers applied Cronbach's Alpha (CFA) to measure the internal consistency of the variables. Findings: The findings of the study on investors' perception towards various investment avenues emphasized that the aged and high income investors prefer to invest only in post office and bank deposits for safety investment reason.

# [2.2] Need for the study

Investment is the most important thing today. Nowadays a people are earning handsomely. They have all right to invest and spend to some extent. But lack of financial education, put them in much more difficult situation. At present lot of investment avenues are available in the market with investor education. Investors can choose from a variety of instruments and assets. While making the choice, they should also consider the rate of return and risk that on their investment.

Each investment alternative has its own strengths and weaknesses. Some options seek to achieve superior returns but with corresponding higher risk. Other provide safety but at the expense of liquidity and growth. Other options such as FDs offer safety and liquidity, but at the cost of return. Mutual funds seek to combine the advantages of investing in arch of these alternatives while dispensing with the shortcomings. Indian stock market is semi-efficient by nature and, is considered as one of the most respected stock markets, where information is quickly and widely disseminated, thereby allowing each security's price to adjust rapidly in an unbiased manner to new information so that, it reflects the nearest investment value. Savings form an important part of the economy of any nation. With the savings invested in various options available to the people, the money acts as the driver for growth of the country. Indian financial scene too presents a plethora of avenues to the investors. Though certainly not the best or deepest of markets in the world, it has reasonable options for an ordinary man to invest his savings. One needs to invest and earn return on their idle resources and generate a specified Sum of money for a specific goal in life and make a provision for an uncertain future. One of the important reasons why one needs to invest wisely is to meet the cost of inflation. Inflation is the rate at which the cost of living increases. The cost of living is simply what it cost to buy the goods and services you need to live. Inflation causes money to lose value because it will not buy the same amount of a good or service in the future as it does now or did in the past. The sooner one starts investing the better.

Comparatively this study reveals investor's mentality on investment and its implications. There are institutions which offer attractive packages to investors. Medias like TV, Newspaper, Magazines, Internet etc., help the investors to access their available avenues for investment. Majority of investor being educated elites in this study, know the available avenues of investment and institutions. Thus, to ascertain people's psychology over investment and financial institutions, an attempt has been made to project the various available avenues for investment and the awareness and satisfactions of that particular investment avenue.

In the investment process, that the investor should have knowledge about the investment alternatives and the markets. The rate of return on investment is highly fluctuation but at the same time, investors have to analyze the rate of return/risk on investment. Financial institutions have been playing a key role to attract investors. Investors are being affected because of the middle agents like broker, jobber etc., everyone is earning profit handsomely but they don't know to access in various available avenues of investment. Lack of financial education, set aside their disposable income in low safety, profitability and marketability of investment as investor, awareness, and they also expect a good rate of return from their investment. For all these, they need adequate flow of information. Hence, the present study entitled INVESTMENT-"A Run for Money" has been taken up.

# [2.3] Introduction

What is Investment? Investment is the employment of funds with the aim of achieving additional income or growth in value. The essential quality of income is that, it involves 'waiting 'for a reward. It involves the commitment of resources which have been saved or put away from current consumption in the hope that some benefits will occur in future. The term 'investment' does not appear to be a simple as it has been defined. Investment has been categorized by financial experts and economists.

Wealth creation is not an art. It is an attribute of one's attitude towards money. How does one know whether have the right kind of attitude towards money? To answer this question, this topic INVESTMENT-"A Run for Money" focus on two basic, money-related aspects and activities viz. avenues and institution available for investment. Now-a-days, investment avenues are widening in the world to create positive sources of income, one can invest disposable income in domestic or offshore market. People in society are investing their savings in a systematic manner and many are in a unsystematic manner. Many do not have financial education. A systematic investment plan always yields a fair return.

Investment is the most important things today. People are earning handsomely, but they do not know where, when and how to invest. Everyone should realize that financial planning is a must today in order to know where one stands financially and also to focus to one's financial efforts in the right direction. A proper understanding of money, its value, the available avenues for investment, various financial institutions, the rate of return/risk etc., are essential to successful manage one's finance for achieving life's goal.

The study has been undertaken to analyse the investment pattern of investment community. The main reasons behind the study are the factors like income, economy condition, risk covering nature of the investors, awareness and satisfaction of investors. This project contains the investor's preferences and as well as the different factors that affect investors decision on the different investment avenues.

There are large numbers of investment available today. Therefore, in this project we have taken nine major types of investments viz. Stock, Mutual Funds, Real estate, Metals (Gold/Silver etc.), Insurance, Banks, Post offices, PPF (Public Provident Fund), and NPS (National Pension Scheme).

Which are explained as follows.

## Stock:

Stock (also capital stock) of a corporation, is all of the shares into which ownership of the corporation is divided. In American English, the shares are collectively known as "stock". A single share of the stock represents fractional ownership of the corporation in proportion to the total number of shares. This typically entitles the stockholder to that fraction of the company's earnings, proceeds from liquidation of assets (after discharge of all senior claims such as secured and unsecured debt),or voting power, often dividing these up in proportion to the amount of money each stockholder has invested. Not all stock is necessarily equal, as certain classes of stock may be issued for example without voting rights, with enhanced voting rights, or with a certain priority to receive profits or liquidation proceeds before or after other classes of shareholders.

## Mutual Funds:

Mutual funds are basically investment vehicles that comprise the capital of different investors who share a mutual financial goal. A fund manager manages the pool of money that is collected from various investors and invests the money into a variety of investment options such as company stocks, bonds, and shares. Mutual funds in India are regulated by the Securities and Exchange Board of India (SEBI), and investing in mutual funds is considered to be the easiest way through which you can increase your wealth.

## Real Estate:

Real estate is "property consisting of land and the buildings on it, along with its natural resources such as crops, minerals or water; immovable property of this nature; an interest vested in this (also) an item of real property, (more generally) buildings or housing in general. Residential real estate may contain either a single family or multifamily structure that is available for occupation or for non-business purposes. Residences can be classified by and how they are connected to neighboring residences and land. Different types of housing tenure can be used for the same physical type. For example, connected residences might be owned by a single entity and leased out, or owned separately with an agreement covering the relationship between units and common areas and concerns.

## Metals (Gold/Silver etc.):

Gold and silver have been recognized as valuable metals, and have been coveted for a long time. Even today, precious metals have their place in a savvy investor's portfolio. But which precious metal is best for investment purposes? And why are they so volatile? There are many ways to buy into precious metals like gold, silver, and platinum, and a host of good reasons why you should give in to the treasure hunt.

## Insurance:

Insurance is a means of protection from financial loss. It is a form of risk management, primarily used to hedge against the risk of a contingent or uncertain loss. An entity which provides insurance is known as an insurer, insurance company, insurance carrier or underwriter. A person or entity who buys insurance is known as an insured or as a policyholder. The insurance transaction involves the insured assuming a guaranteed and known relatively small loss in the form of payment to the insurer in exchange for the insurer's promise to compensate the insured in the event of a covered loss. The loss may or may not be financial, but it must be reducible to financial terms, and usually involves something in which the insured has an insurable interest established by ownership, possession, or pre-existing relationship. The insured receives a contract, called the insurance policy, which details the conditions and circumstances under which the insurer will compensate the insured. The amount of money charged by the insurer to the policyholder for the coverage set forth in the insurance policy is called the premium. If the insured experiences a loss which is potentially covered by the insurance policy, the insured submits a claim to the insurer for processing by a claims adjuster. The insurer may hedge its own risk by taking out reinsurance, whereby another insurance company agrees to carry some of the risks, especially if the primary insurer deems the risk too large for it to carry.

## Banks:

Banks offer numerous "free" services like savings accounts and free checking. In fact, they may even pay you for leaving money in the bank, and you can also boost your earnings by using certificates of deposit (CD) and money market accounts. Unless you work with an online bank, most banks and credit unions also have physical locations staffed by employees. They also run call centers with extended customer service hours. How do they pay for all of those services? Banks earn revenue from investments (or borrowing and lending), account fees, and additional financial services. Whenever you give money to a financial institution, it's essential to understand a firm's business model and exactly how much they charge. But it's not always clear how banks get paid. There are several ways for banks to earn revenue, including investing your money and charging fees to customers.

## Post Offices:

A post office is a public department that provides a customer service to the public and handles their mail needs. Post offices offer mail-related services such as acceptance of letters and parcels; provision of post office boxes; and sale of postage stamps, packaging, and stationery. In addition, many post offices offer additional services: providing and accepting government forms (such as passport applications), processing

government services and fees (such as road tax), and banking services (such as savings accounts and money orders).

## Public Provident Fund (PPF):

The Public Provident Fund is a savings-cum-tax-saving instrument in India, introduced by the National Savings Institute of the Ministry of Finance in 1968. The aim of the scheme is to mobilize small savings by offering an investment with reasonable returns combined with income tax benefits. The scheme is fully guaranteed by the Central Government. Balance in PPF account is not subject to attachment under any order or decree of court. However, Income Tax & other Government authorities can attach the account for recovering tax dues.

## National Pension Scheme (NPS)

The National Pension System (NPS) is a voluntary defined contribution pension system in India. National Pension System, like PPF and EPF is an EEE (Exempt-Exempt-Exempt) instrument in India where entire corpus escapes tax at maturity and entire pension withdrawal amount is tax-free. NPS started with the decision of the Government of India to stop defined benefit pensions for all its employees who joined after 1 January 2004. While the scheme was initially designed for government employees only, it was opened up for all citizens of India between the age of 18 and 60 in 2009.

A citizen of India, whether resident or non-resident can join NPS, subject to the following conditions:

- The subscriber should be between 18 and 65 years old as of the date of submission of his/her application to the Point of Presence (POP) / Point of Presence–Service Provider-Authorized branches of POP for NPS (POP-SP).
- The subscribers should comply with the Know Your Customer (KYC) norms as detailed in the subscriber registration form.
- Should not be Un-discharged insolvent and individuals of unsound mind.

## Others:

Some other types of traditional investments are also present like chit funds, public broker ( *sahukaar*) etc…

# [3.1] Variable:

| Sr. no. | Variables | Levels |
|---|---|---|
| 01 | Gender | 1) Female <br> 2) Male |
| 02 | Age | 1) 18-25 <br> 2) 26-35 <br> 3) 36-45 <br> 4) 46-55 <br> 5) 55 & above |
| 03 | Qualification | 1) Below SSC <br> 2) SSC <br> 3) HSC <br> 4) Diploma <br> 5) Graduate <br> 6) Post Graduate <br> 7) PhD |
| 04 | Occupation | 1) Student and Employed <br> 2) Employed (Government) <br> 3) Employed (Private) <br> 4) Earning Housewife <br> 5) Professionals (Teachers, Doctors, etc) <br> 6) Business <br> 7) Retired |
| 05 | Family Size | 1) 2 <br> 2) 3 <br> 3) 4 <br> 4) 5 <br> 5) 6 <br> 6) 7 <br> 7) More than 7 |
| 06 | Number of Earning Members | 1) 1 <br> 2) 2 <br> 3) 3 <br> 4) 4 <br> 5) 5 <br> 6) 7 |
| 07 | Region | 1) Mumbai <br> 2) Mumbai Suburban <br> 3) Thane <br> 4) Others |
| 08 | Income | 1) Less than 5 lakh <br> 2) 5 lakhs to 10 lakh <br> 3) 10 lakhs to 15 lakh |

| | | |
|---|---|---|
| | | 4) More than 15 lakh |
| 09 | Whether Invest or Not | 1) Yes <br> 2) No |
| 10 | Saving | 1) Less than 10% <br> 2) 10% to 25% <br> 3) 25% to 50% <br> 4) More than 50% |
| 11 | Investment | 1) Less than 5% <br> 2) 5% to 10% <br> 3) 10% to 20% <br> 4) More than 20% |
| 12 | Awareness | 1) Not at all aware <br> 2) Slightly aware <br> 3) Extremely aware |
| 13 | Risk | 1) Low risk <br> 2) Medium risk <br> 3) High risk |
| 14 | Return | 1) Low return <br> 2) Medium return <br> 3) High return |
| 15 | Problems | 1) Lack of knowledge <br> 2) Financial Planning <br> 3) Where and How to invest <br> 4) Return is not fixed <br> 5) No problem |
| 16 | Factor influencing investment | 1) Safety <br> 2) Return <br> 3) Tax saving <br> 4) Future need <br> 5) Liquidity |
| 17 | Precautions | 1) Have invested wisely with proper information <br> 2) Invested in different options <br> 3) Already have Saving <br> 4) No precautions |

11

# [3.2] Graphical Representation:

## [3.2.1] Gender

| Gender | Count of Gender | % of Gender |
|--------|-----------------|-------------|
| Male | 362 | 61% |
| Female | 232 | 39% |
| Total | 594 | 100% |
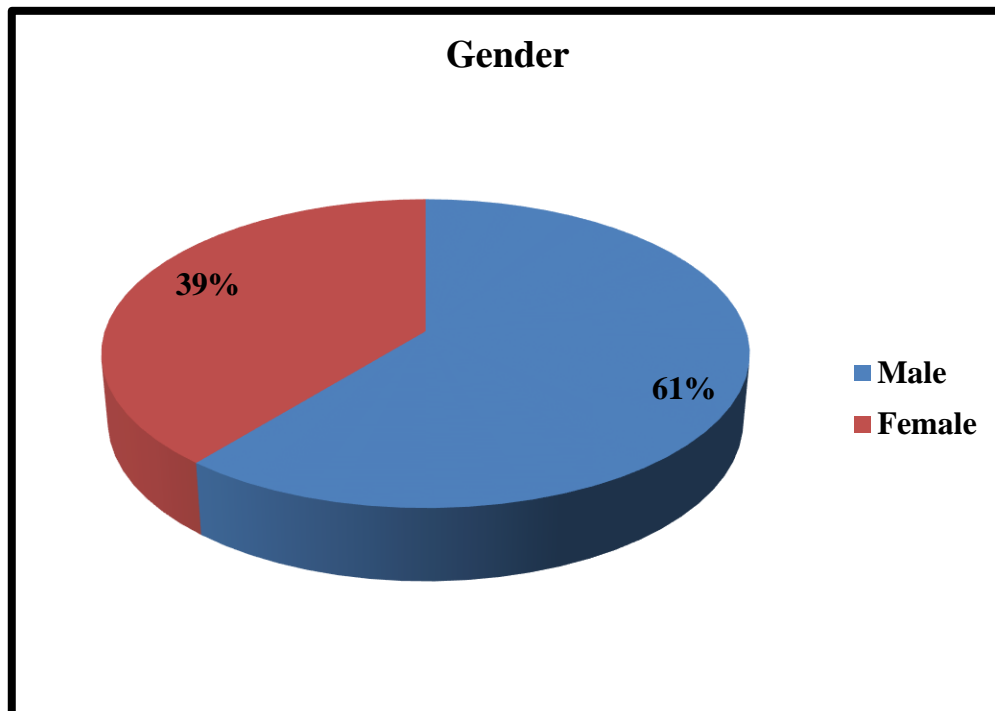


Here we can see that 61% of our respondents are male and the remaining 48% are female.

## [3.2.2] Age

| Age | Count of Age | % of Age |
|---|---|---|
| 18 - 25 | 154 | 26% |
| 26 - 35 | 171 | 29% |
| 36 - 45 | 141 | 24% |
| 46 - 55 | 78 | 13% |
| 55 & above | 50 | 8% |
| Total | 594 | 100% |
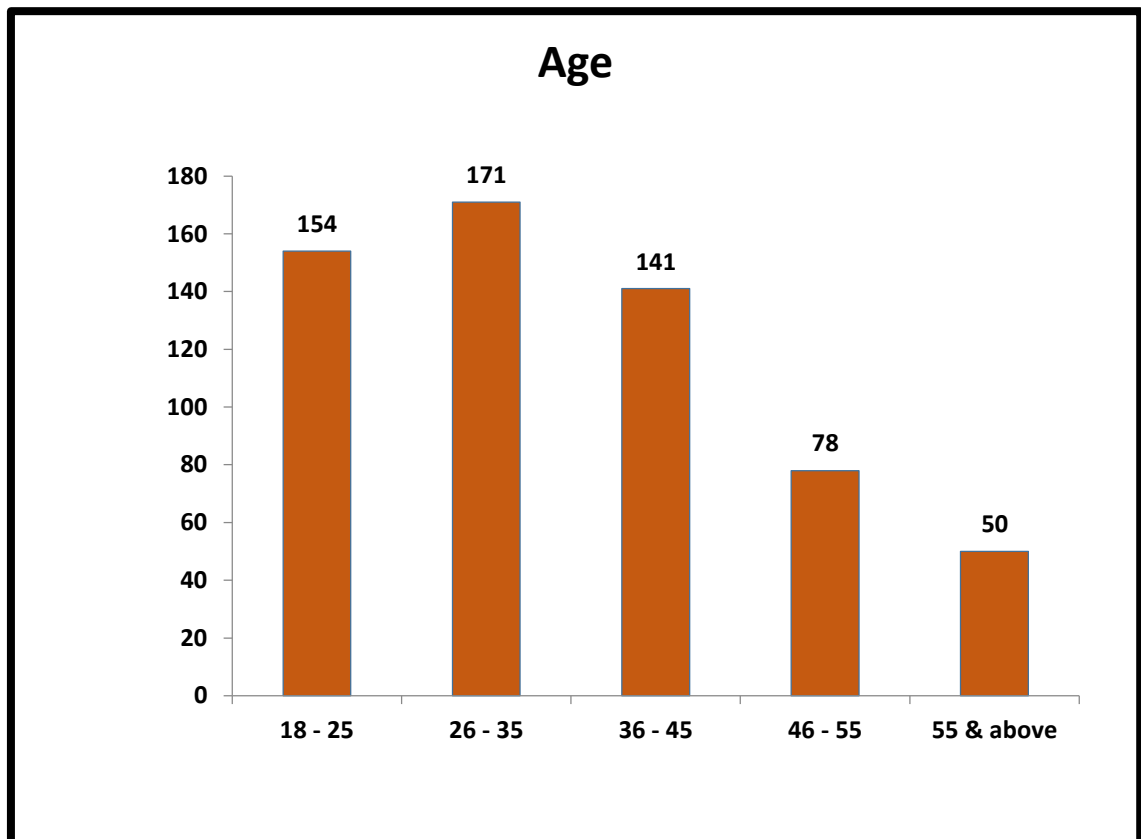


Here we can see that most of our respondents are between age group 18-25, 26-35, and 36-45.

# [3.2.3] Qualification

| Qualification | Count of Qualification | % of Qualification |
|---|---|---|
| Below S.S.C | 9 | 2% |
| S.S.C | 38 | 6% |
| H.S.C | 22 | 4% |
| Diploma | 80 | 13% |
| Graduation | 262 | 44% |
| Post-Graduation | 180 | 30% |
| Doctorate | 3 | 1% |
| Total | 594 | 100% |
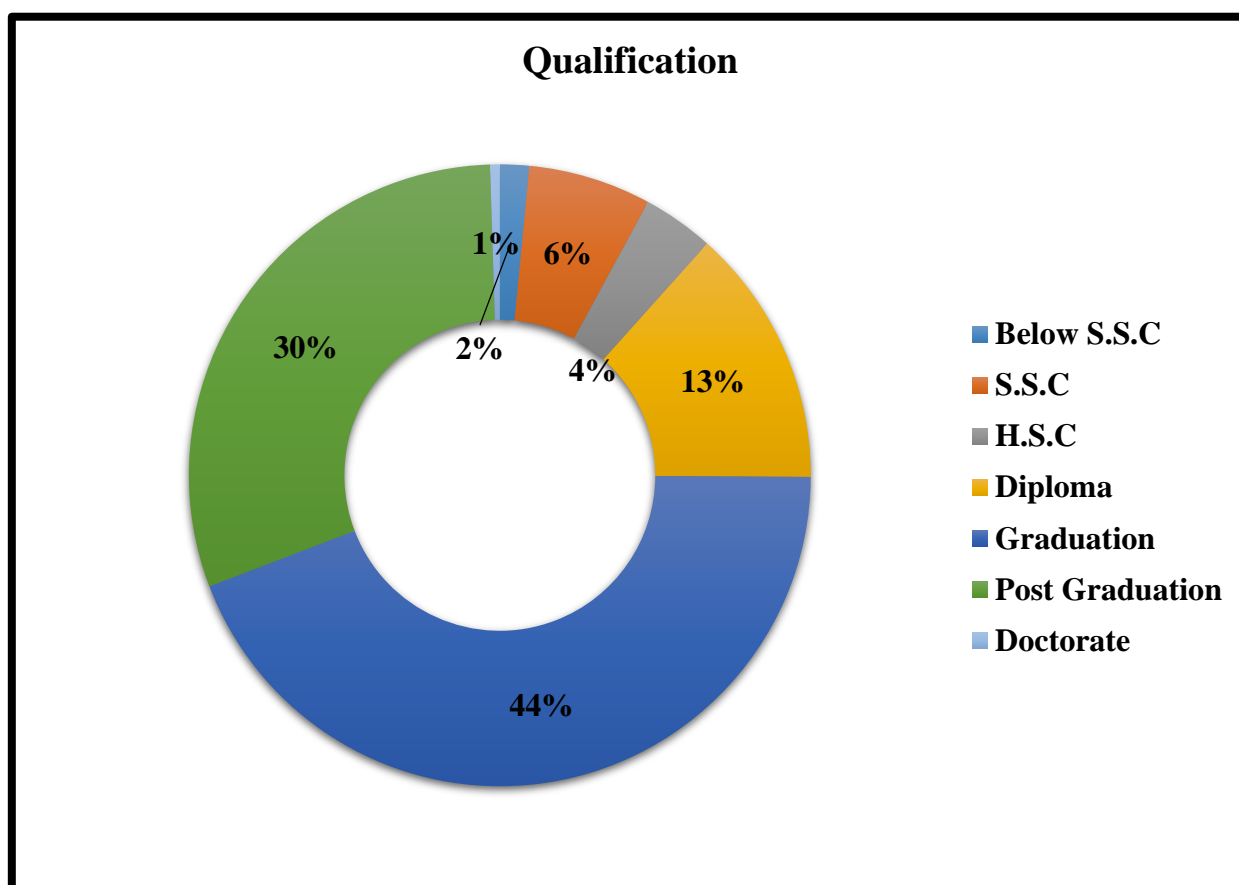
## Qualification



**Here we can see that most of our respondents have qualification Graduation and Post-Graduation.**

# [3.2.4] Occupation

| Occupation | Count of Occupation | % of Occupation |
|---|---|---|
| Student and Employed | 103 | 17% |
| Employed (Government) | 93 | 16% |
| Employed (Private) | 229 | 38% |
| Earning Housewife | 34 | 6% |
| Business | 60 | 10% |
| Professionals ( Teachers, Doctors, etc.) | 47 | 8% |
| Retired | 28 | 5% |
| Total | 594 | 100% |

**Occupation**
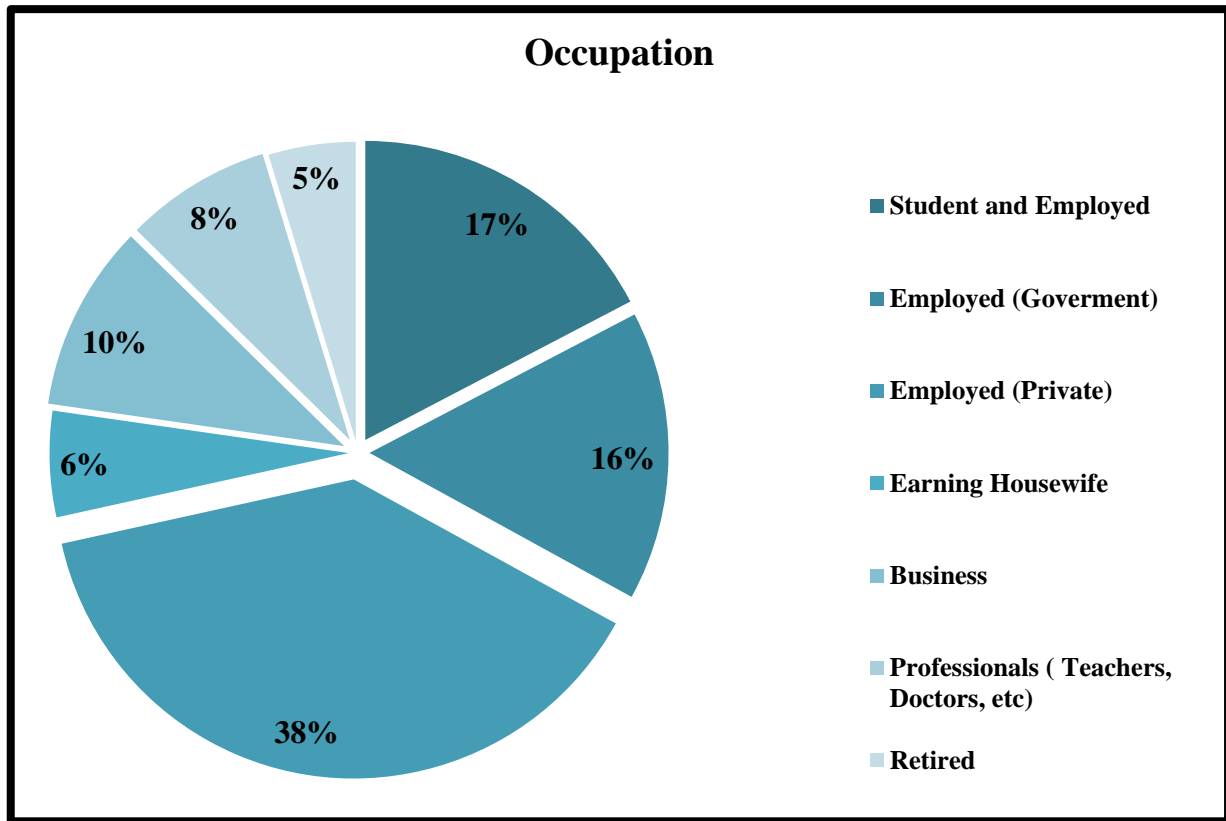


**Here we can see that most of our respondents are Government Employees, Private Employees and Employed Student.**

# [3.2.5] Family Size

| Family Size | Count of Family Size | % of Family Size |
|:---:|:---:|:---:|
| 2 | 36 | 6.06% |
| 3 | 132 | 22.22% |
| 4 | 204 | 34.34% |
| 5 | 129 | 21.72% |
| 6 | 51 | 8.59% |
| 7 | 20 | 3.37% |
| More Than 7 | 22 | 3.70% |
| Total | 594 | 100.00% |
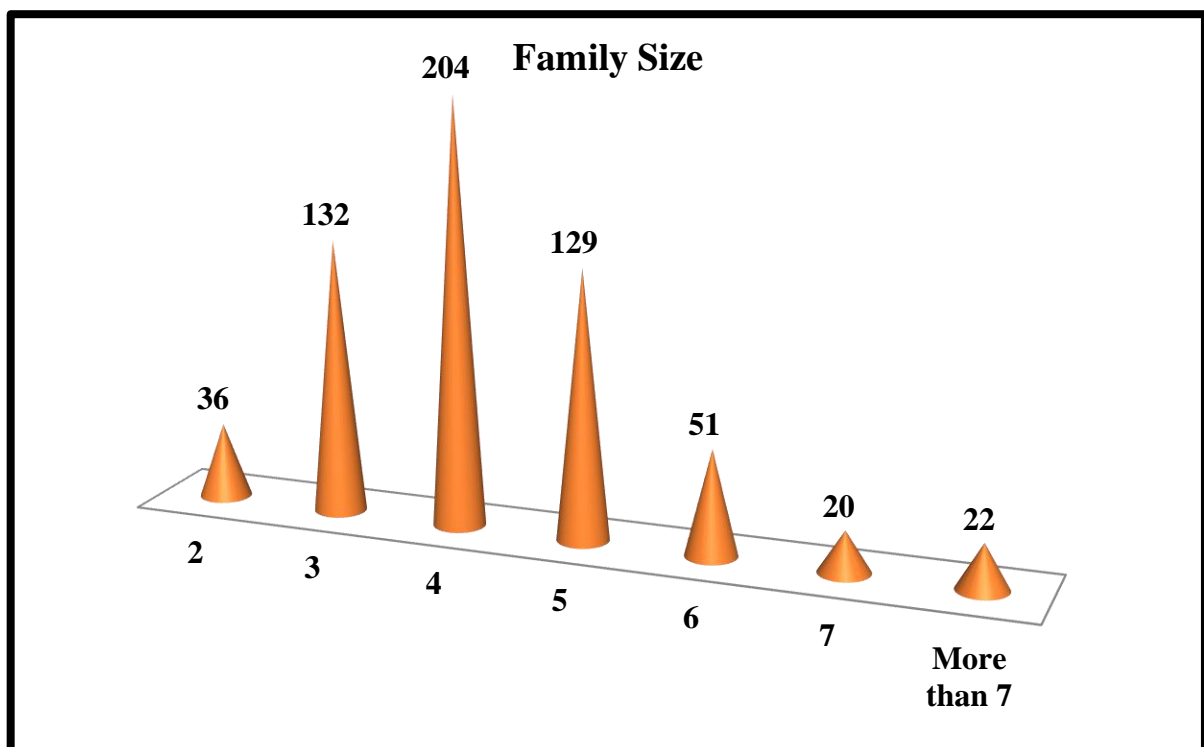


**Here we can see that most of our respondents have Family Size 3, 4 or 5.**

# [3.2.6] Number of Earning Members

| No. of earning members | Count of Earning Members | % of Earning Members |
|:---:|:---:|:---:|
| 1 | 183 | 31% |
| 2 | 279 | 47% |
| 3 | 94 | 16% |
| 4 | 28 | 5% |
| 5 | 8 | 1% |
| 6 | 2 | 0.3367% |
| Total | 594 | 100% |

**Number of Earning Members**
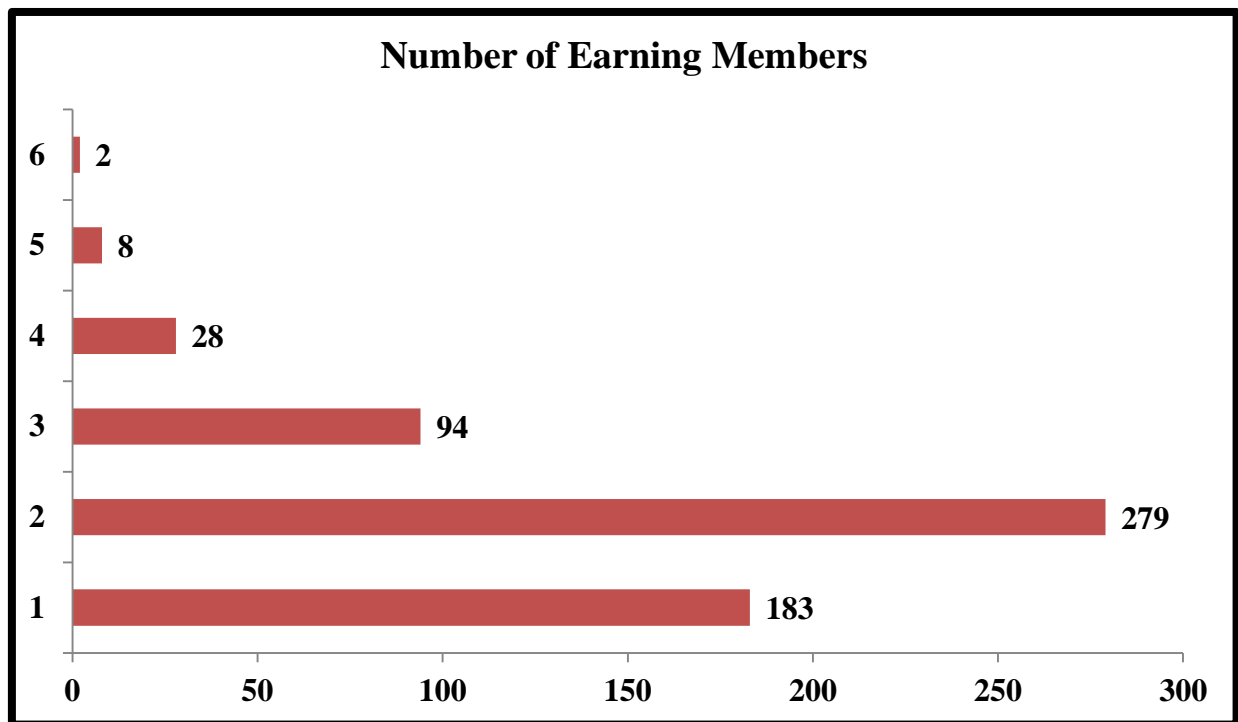


Here we can see that most of our respondents have Earning Members 1, 2 or 3.

# [3.2.7] Region

| Region | Count in Region | % in Region |
|---|---|---|
| Mumbai | 228 | 38% |
| Mumbai Suburban | 166 | 28% |
| Thane | 164 | 28% |
| Others | 36 | 6% |
| Total | 594 | 100% |
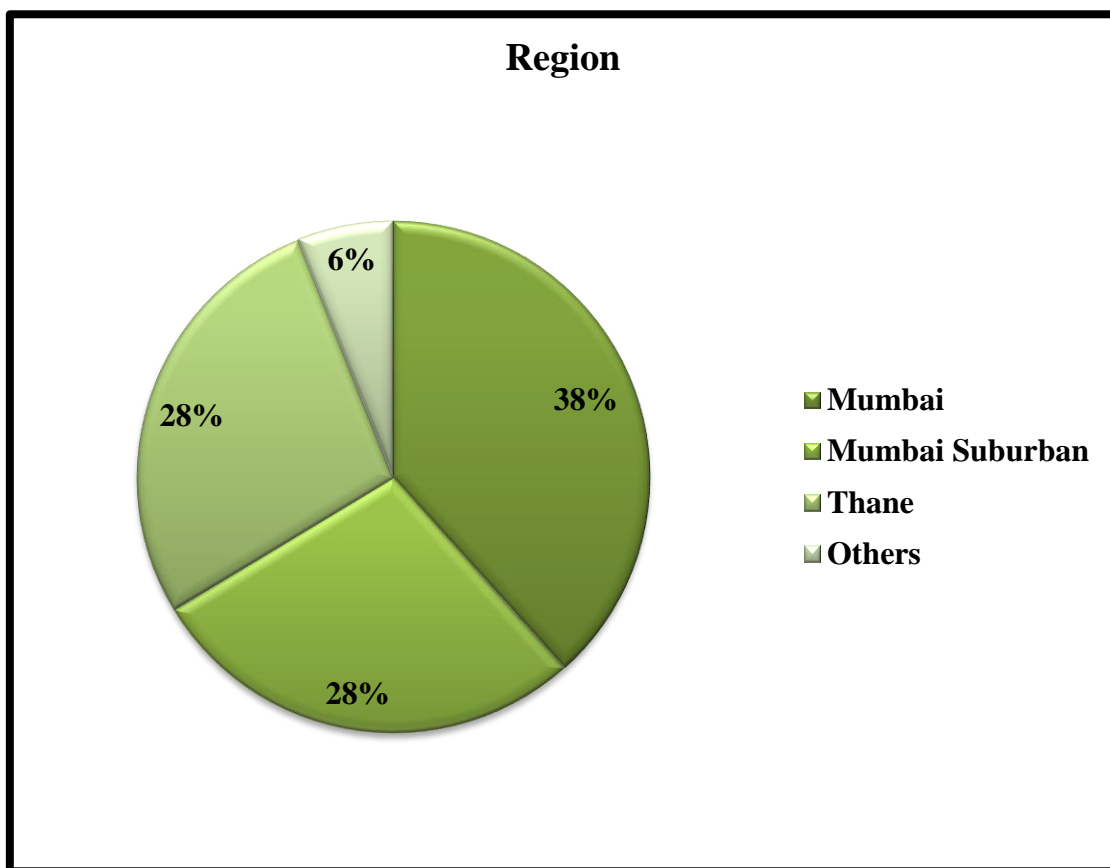
**Region**



**Here we can see that most of our respondents are from region Mumbai.**

# [3.2.8] Annual Income

| Annual Income | Count of Annual Income | % of Annual Income |
|---|---|---|
| Less than 5 lakh | 284 | 48% |
| 5 lakh to 10 lakh | 177 | 30% |
| 10 lakh to 15 lakh | 78 | 13% |
| more than 15 lakh | 55 | 9% |
| Total | 594 | 100% |
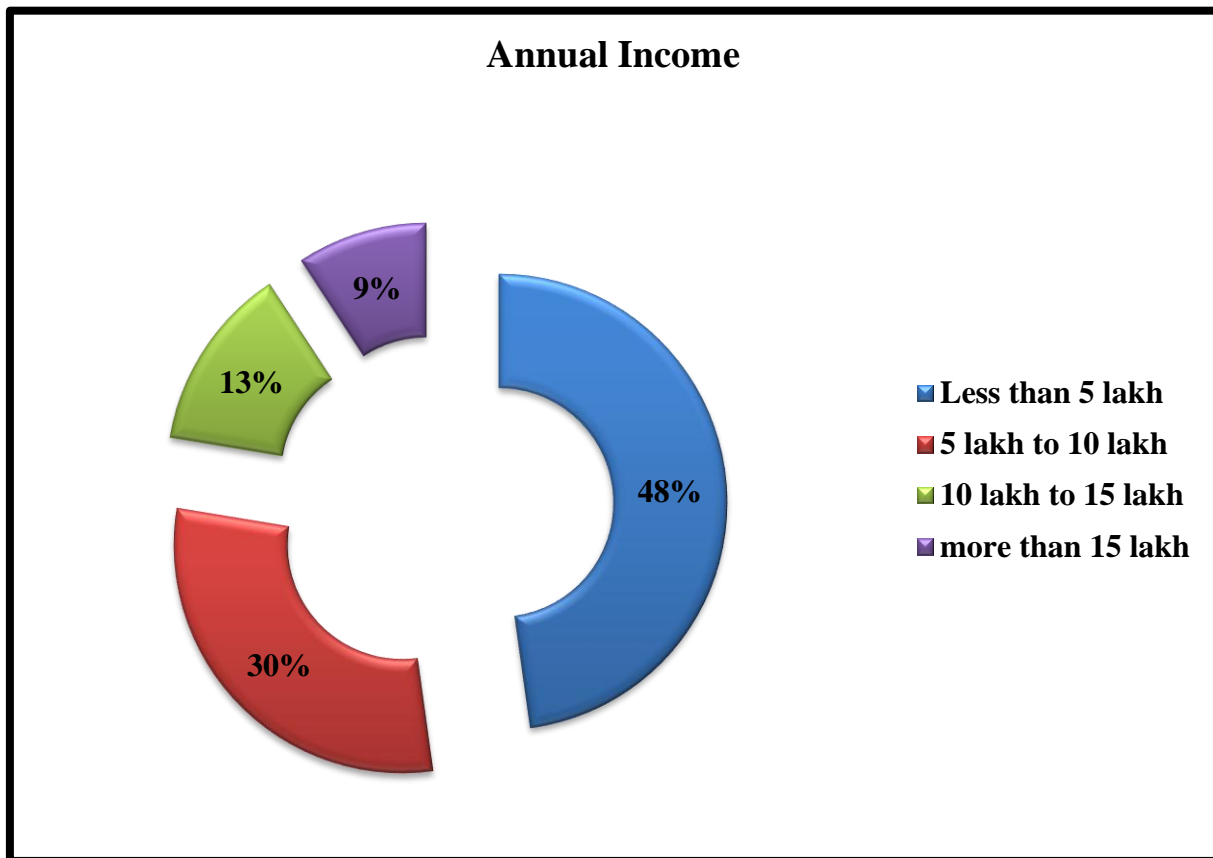
**Annual Income**



Here we can see that most of our respondents have Annual Income Less than 5 lakhs and 5 lakhs to 10 lakhs.

# [3.2.9] Whether invest or Not?

| Invest or Not? | Count | % |
|:---:|:---:|:---:|
| No | 190 | 32% |
| Yes | 404 | 68% |
| Total | 594 | 100% |

**Invest or Not?**
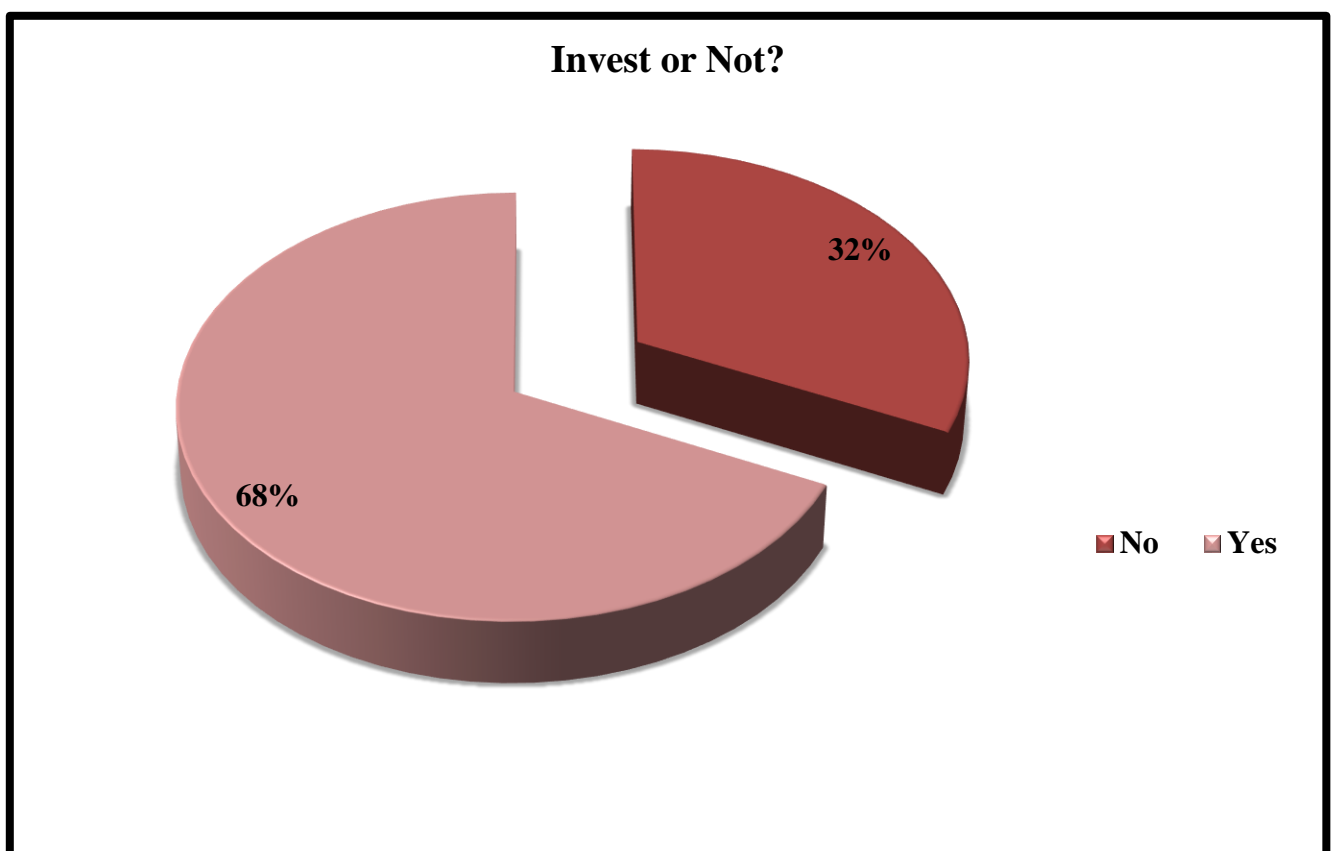


In Above graphical representations we have seen that there are overall 32% people who don't invest anywhere in any investment avenues. Hence to explore more about the reasons of not investing we will also make some graphical representation of people who don't invest by considering the factors Gender, Qualification and Annual Income.

## [3.2.10] Gender of People who don't invest

| Gender | Count | % |
|:------:|:-----:|:-----:|
| Female | 83 | 44 % |
| Male | 107 | 56 % |
| Total | 190 | 100 % |
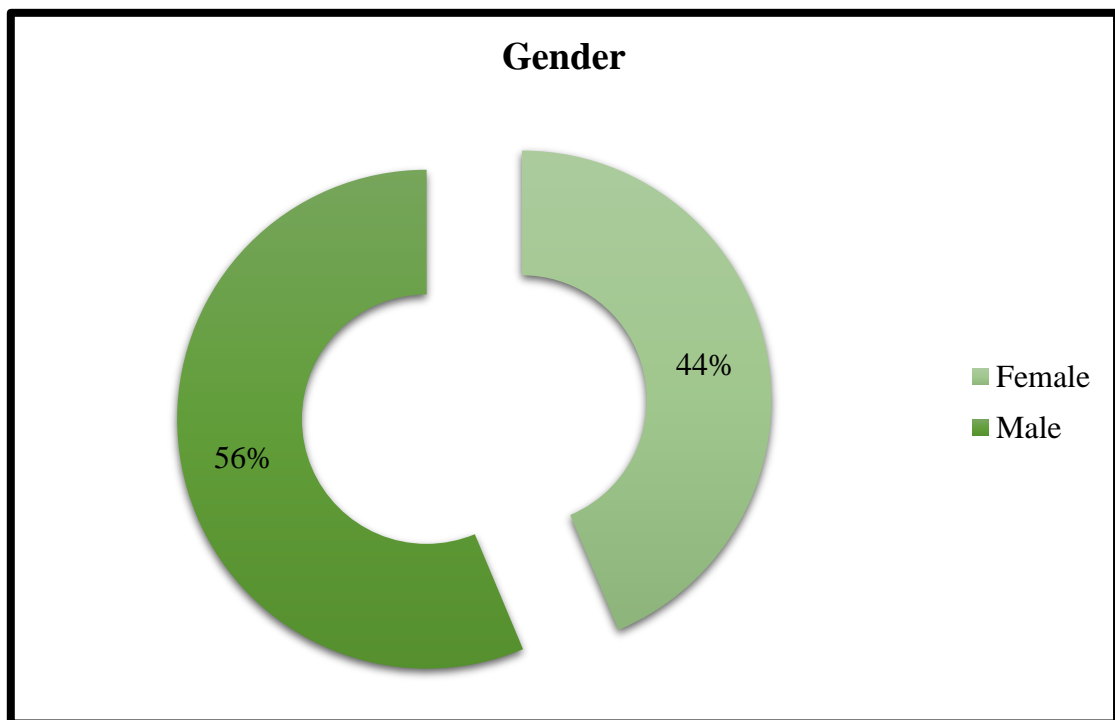
### Gender



As we have already found that out of 594 people 190 of our respondents (32% of respondents) don't invest in any type of investment avenue. By analysing the Gender, we can see that Gender don't have much effect on investment as same number of male and female respondents don't invest. i.e. 56% male and 44% females are not investors.

| Educational Qualification | Count | % |
|---|---|---|
| Below S.S.C | 4 | 2 % |
| S.S.C | 9 | 5 % |
| H.S.C | 23 | 12 % |
| Diploma | 10 | 5 % |
| Graduation | 78 | 41 % |
| Post-Graduation | 66 | 35 % |
| Total | 190 | 100 % |
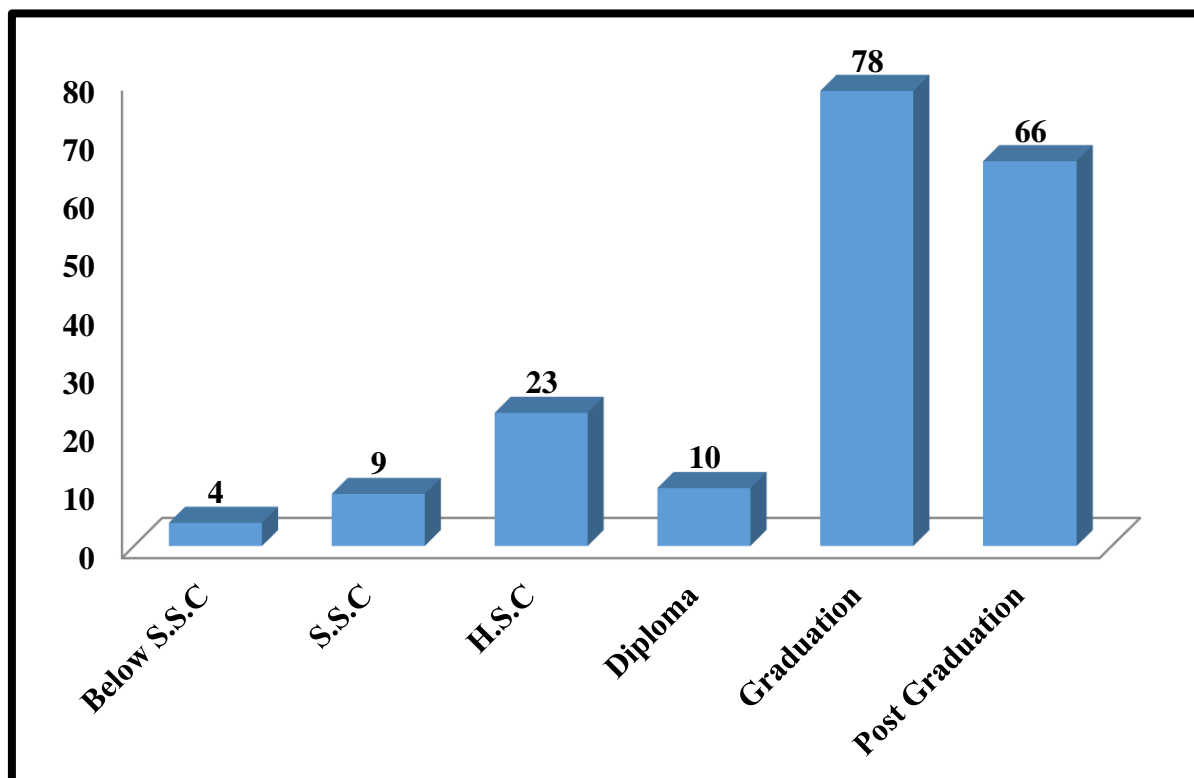


Here we can see that most of people who don't invest have highest qualification among all non investors (78 are Graduate & 66 Post Graduate). It is bit difficult to find out the reason for not investment but we will now analyse reason for not investing.

| Graduation | Count | % | Post-Graduation | Count | % |
|---|---|---|---|---|---|
| Financial Issue | 30 | 38.46% | Financial Issue | 20 | 30.30 |
| Lack of Knowledge | 24 | 30.77% | Lack of Knowledge | 18 | 27.27 |
| More Risk | 3 | 3.85% | More Risk | 9 | 13.64 |
| Not Interested | 21 | 26.92% | Not Interested | 19 | 28.79 |
| Total | 78 | 100% | Total | 66 | 100 |
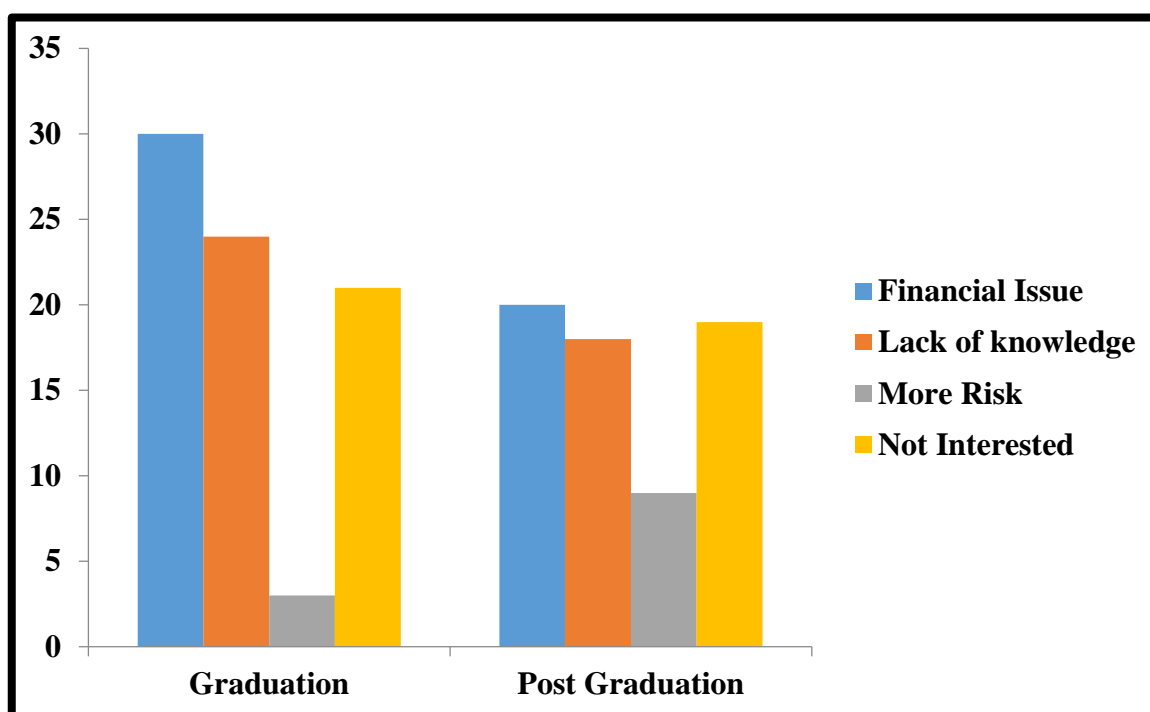


From here we can conclude that among the Graduates non investor the biggest reason for not investing is financial issue followed by Lack of knowledge and some people also don't show interest in investing. Also among the Post Graduate non investor the biggest reason for not investing is Financial issue followed by Lack of knowledge and some people also don't show interest in investing. In both the group very less number of people have risk issue.

## [3.2.13] Annual Income of People who don't invest

| Annual income | Count | % |
|---|---|---|
| Less than 5 lakh | 133 | 70 % |
| 5 lakh to 10 lakh | 44 | 23 % |
| 10 lakh to 15 lakh | 9 | 5 % |
| more than 15 lakh | 4 | 2 % |
| Total | 190 | 100 % |
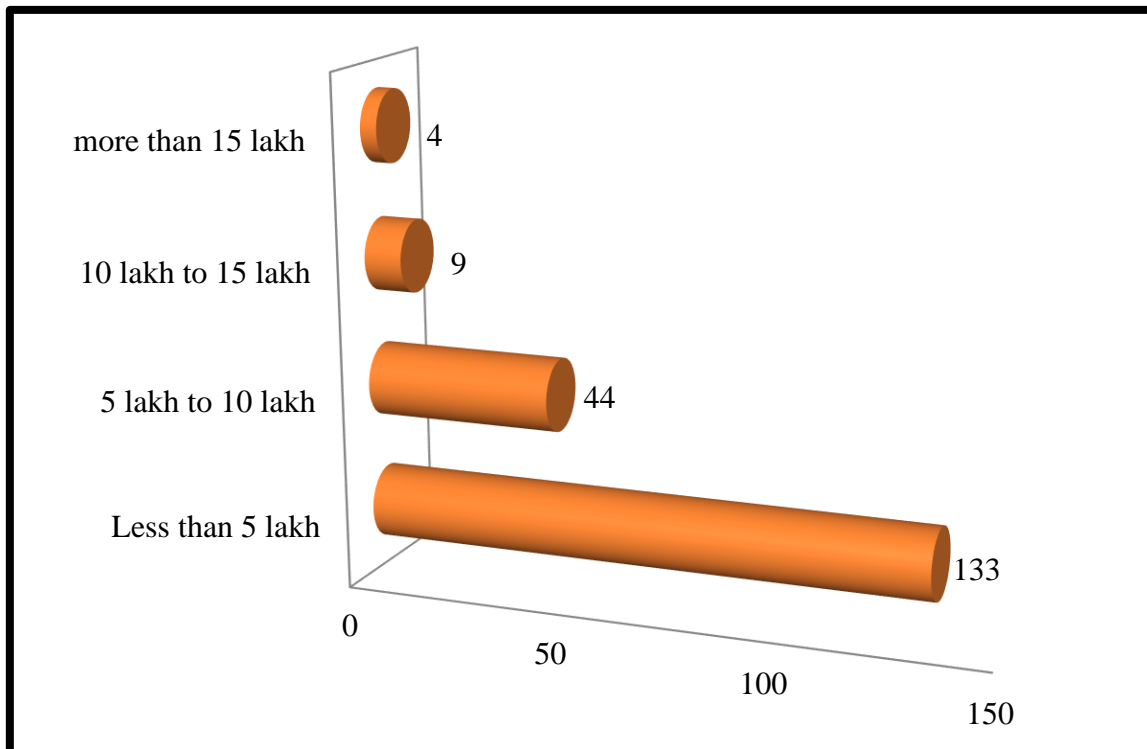


From above graph we can conclude that Annual Income perform a major role in ones investment. In our data of non investor most of people have income less than 5 lakh i.e. less as compared to other people. So, from here we can conclude that income can be a reason for not investing.

# [4.1] Objective (1):

**To find which is the most preferred investment method and factor influencing the Investment.**

# [4.2] Analysis:

| Avenues | Stocks | Mutual Funds | Real Estate | Metals | Insurance | Banks | Post Office | PPF | NPS |
|---------|--------|--------------|-------------|--------|-----------|-------|-------------|-----|-----|
| **Count** | 100 | 155 | 88 | 121 | 241 | 234 | 79 | 94 | 45 |

**From the above table we can see that Insurance is most preferred investment avenue followed by Banks and NPS is least preferred investment avenue.**
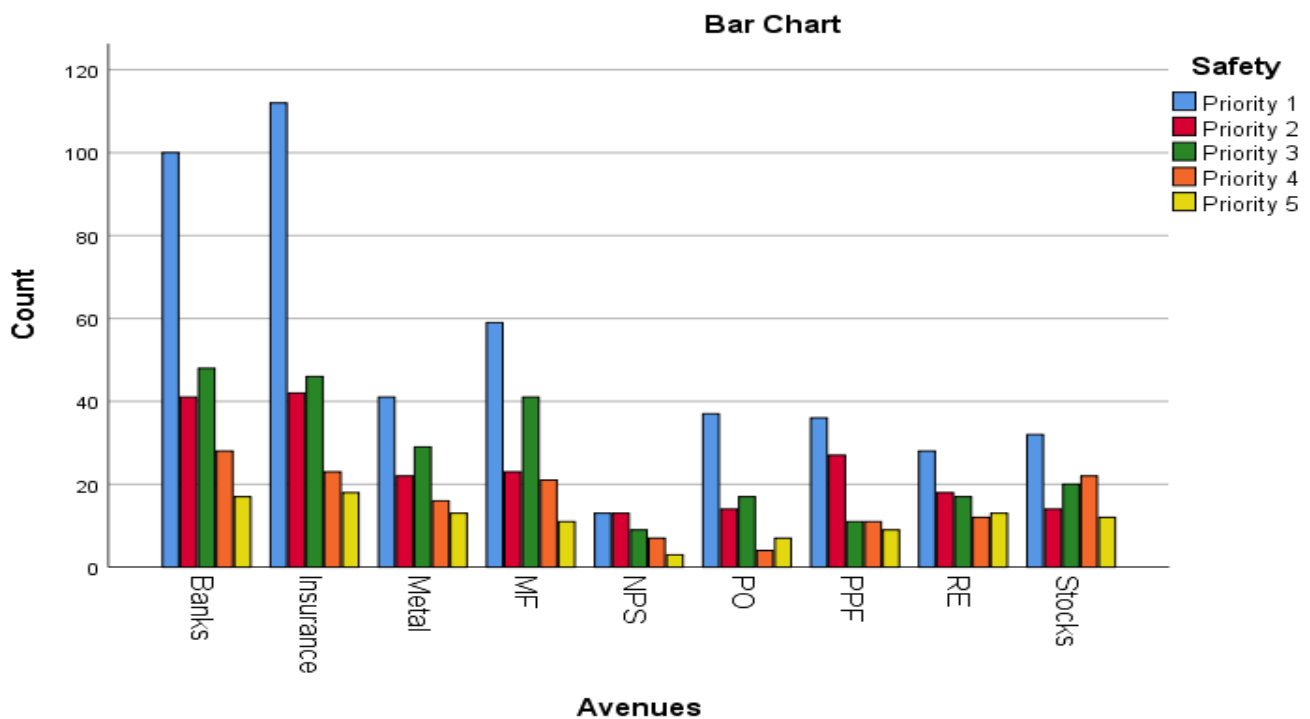


Fig 4.2.1

**From the fig 4.2.1, we see that there are different priorities of "Safety" for different avenues. So, we can conclude that majority investors have safety as their 1ˢᵗ priorities for all the avenues.**

**Bar Chart**



Fig 4.2.2

**From the fig 4.2.2, we see that there are different priorities of "Returns" for different avenues. So, we can conclude that majority investors have returns as their 2nd priorities for all the avenues, except for Insurance as there is a tie between 2nd and 3rd priority.**

**Bar Chart**



Fig 4.2.3

26

From the fig 4.2.3, we see that there is variation in priorities of "Tax Savings" for different avenues. So, we can conclude that majority investors of Stocks have 1st priority, majority investors of PPF and Mutual Funds have 3rd priority, majority investors of Banks and Insurance have 4th priority, whereas investors of Post Office have tie between 4th and 5th priority and majority investors of Metal and Real Estate have 5th priority.
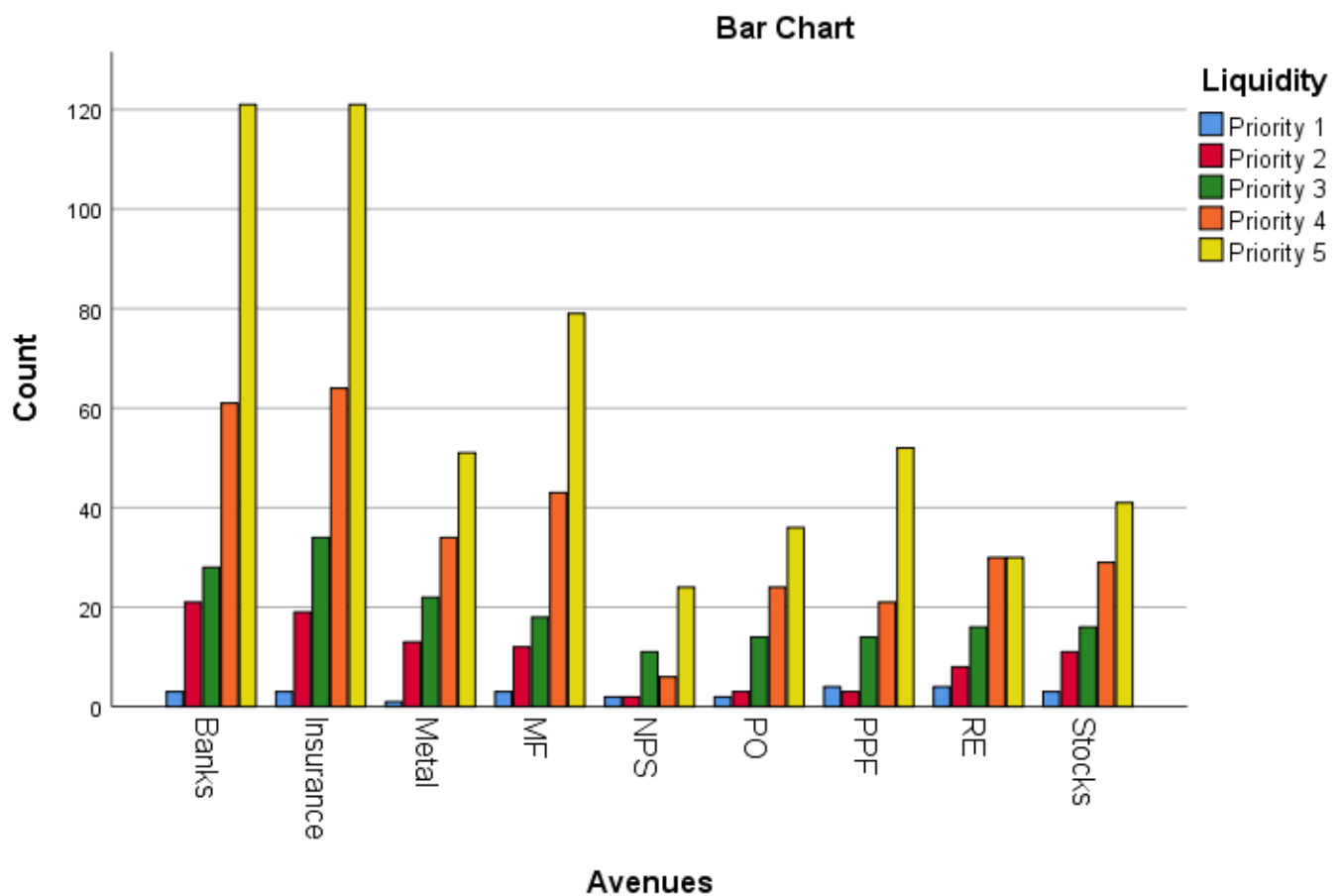


Fig 4.2.4

From the fig 4.2.2, we see that there are different priorities of "Liquidity" for different avenues. So, we can conclude that majority investors have liquidity as their 5thpriorities for all the avenues, except for Real Estate as there is tie between 4th and 5th priority.
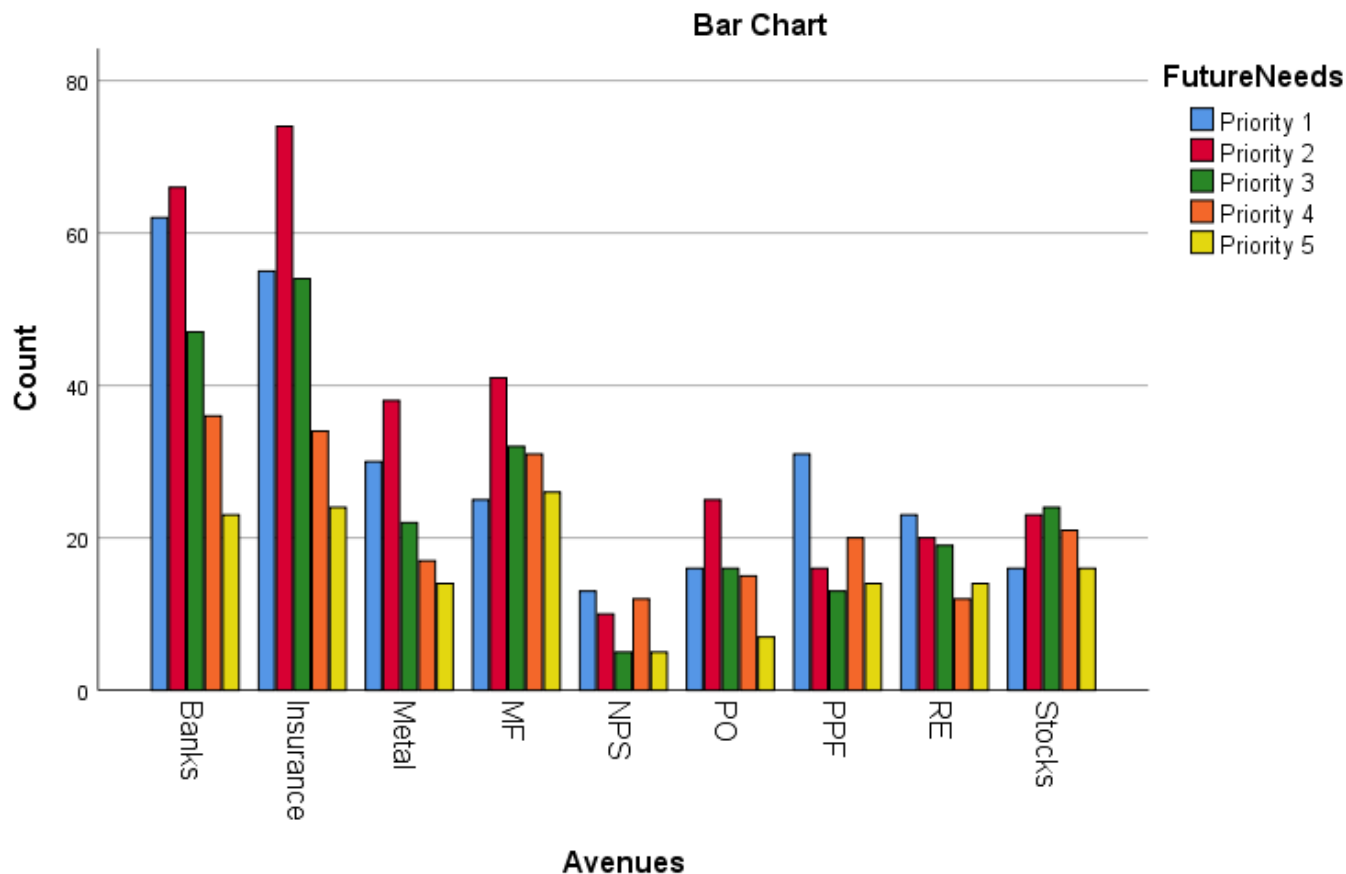
Fig 4.2.5

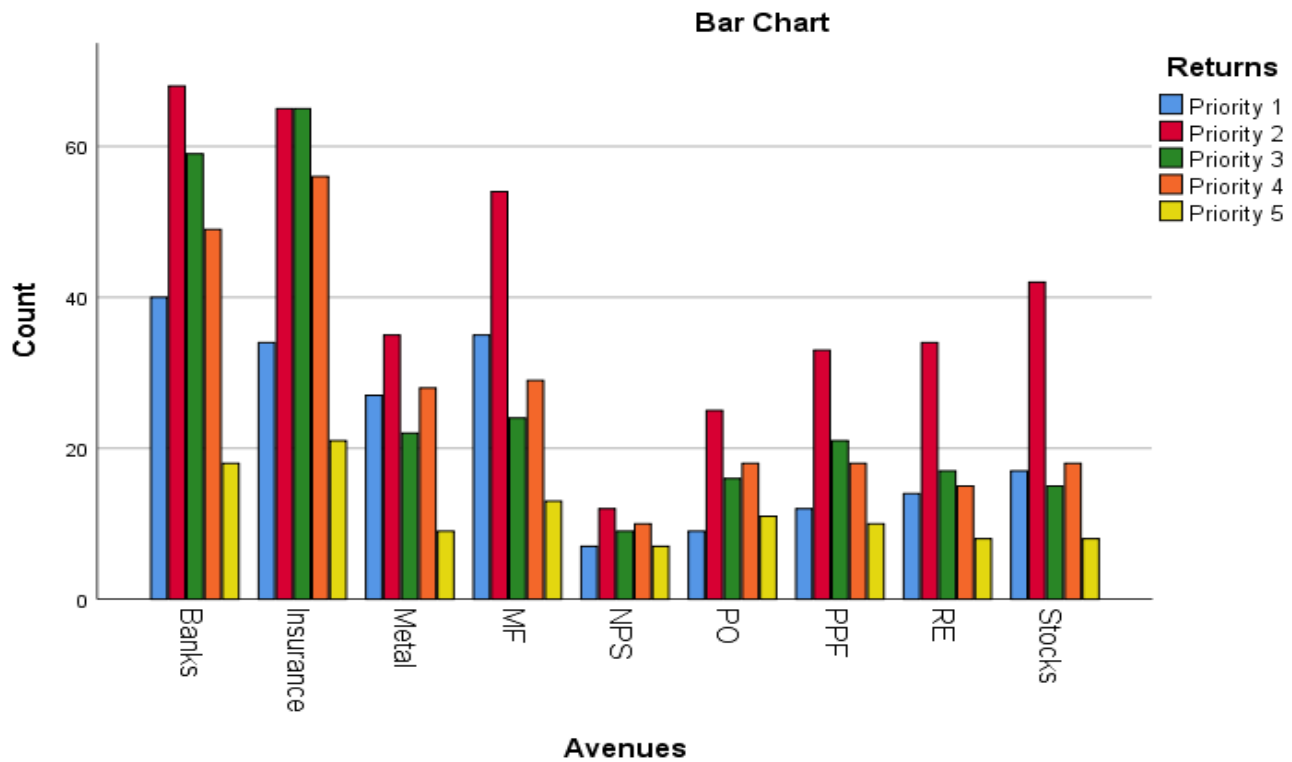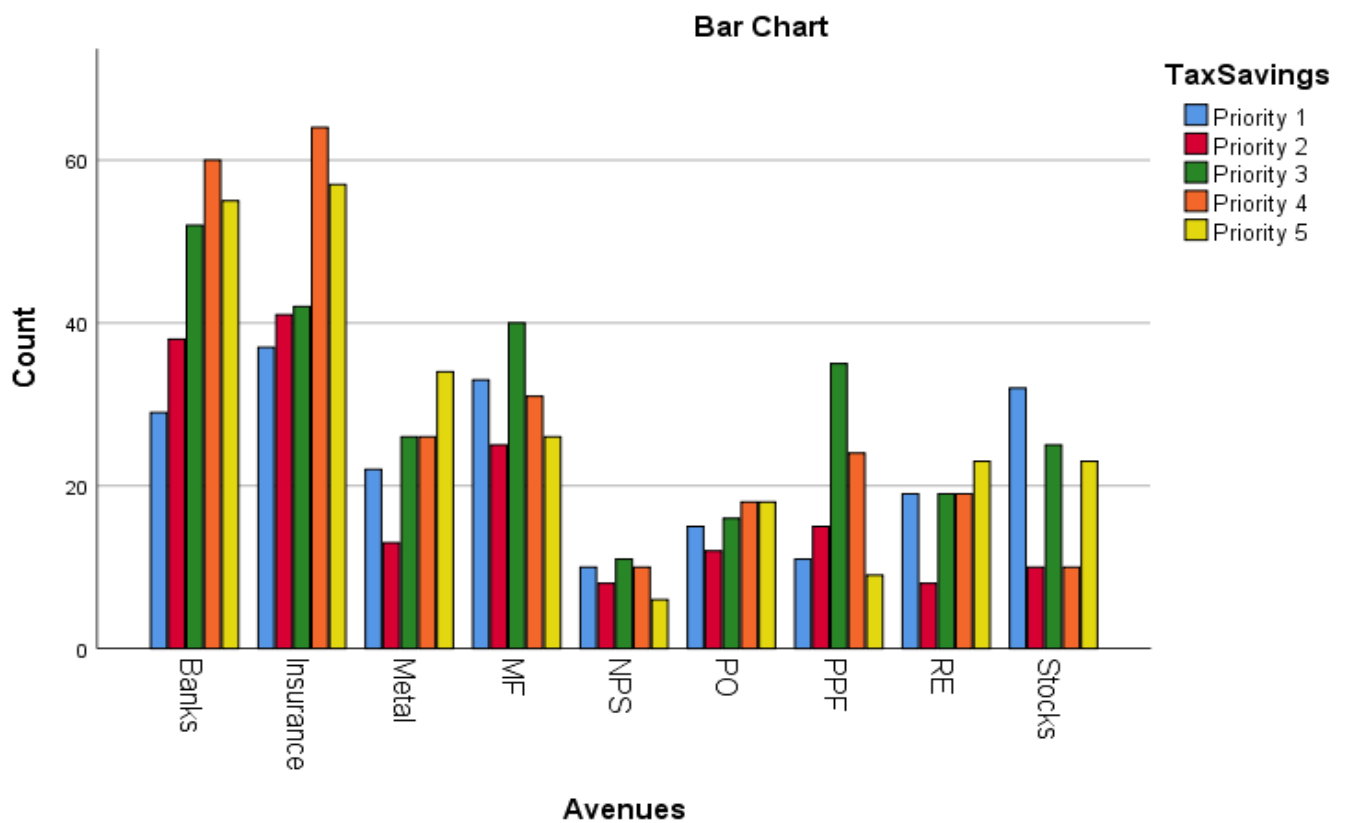**From the fig 4.2.5, we see that there is variation in priorities of "Future Needs" for different avenues. So, we can conclude that majority investors of NPS, PPF and Real estate have 1st priority, majority investors of Banks, Insurance, Metals, Mutual Funds and Post Office have 2nd priority, majority investors of Stocks have 3rd priority.**

# [5.1] Objective (2):

**To compare the investment according to socio-demographic factor.**

# [5.2] Techniques:

## [a] Cross Tabulation:

Cross-tabulation is one of the most useful analytical tools. Cross tabulation is a method to quantitatively analyze the relationship between multiple variables. Also known as contingency tables or cross tabs, cross tabulation groups variables to understand the correlation between different variables. It also shows how correlations change from one variable grouping to another. It is usually used in statistical analysis to find patterns, trends, and probabilities within raw data. When you can use cross tabulation Cross tabulation is usually performed on categorical data — data that can be divided into mutually exclusive groups. Cross-tabulation analysis has its own unique language, using terms such as "banners", "stubs", "Chi-Square Statistic" and "Expected Values."

Bivariate analysis examines the relationship between two variables. Different methods for analyzing bivariate analysis are available, of which cross tabulation is one of the important and frequently used methods. A cross tabulation counts the number of observations in each cross category of two variables. The descriptive result of a cross tabulation is a frequency count for each cell in the analysis. Following section deals with analysis of bivariate data using cross tabulation method.

## [b] Chi-Square Analysis:

The Chi-square statistic is the primary statistic used for testing the statistical significance of the cross-tabulation table. Chi-square tests determine whether or not the two variables are independent. If the variables are independent (have no relationship), then the results of the statistical test will be "non-significant" and we are not able to reject the null hypothesis, meaning that we believe there is no relationship between the variables. If the variables are related, then the results of the statistical test will be "statistically significant" and we are able to reject the null hypothesis, meaning that we can state that there is some relationship between the variables. The chi-square statistic, along with the associated probability of chance observation, may be computed for any table. If the variables are related (i.e., the observed table relationships would occur with very low probability, say only 5%) then we say that the results are "statistically significant" at the .05 or 5% level. This means that the variables have a low chance of being independent.

The chi-square statistic is computed by first computing a chi-square value for each individual cell of the table and then summing them up to form a total chi-square value for the table. The chi-square value for the cell is computed as: (Observed Value – Expected Value)^2 / (Expected Value).

- **Hypothesis:**

  $H_0$=There is no association between both the variables.

  $H_1$= There is a association between both the variables

- **Decision Criteria**

  **If p-value<0.05**

  **Then reject H$_0$**

# [c] Binary logistic regression:

Logistic regression models are used to analyze the relationship between a dependent variable and independent variable(s) when the Dependent Variables is dichotomous. The Dependent Variable is the outcome variable, a.k.a. the predicted variable, and the Independent Variables are the variables that are believed to have an influence on the outcome, a.k.a. predictor variables. If the model contains 1 Independent Variable, then it is a simple logistic regression model, and if the model contains 2+ Independent Variables, then it is a multiple logistic regression model.

Assumptions for logistic regression models:

- The Dependent Variable is categorical (binary)
- If there are more than 2 categories in terms of types of outcome, a multinomial logistic regression should be used
- Independence of observations
- Cannot be a repeated measures design, i.e. collecting outcomes at two different time points.
- Independent variables are linearly related to the log odds
- Absence of multicollinearity

# [5.3] Analysis:

# [5.3.a] Analysis using Cross Tabulation:

**[5.3.1] Cross Tabulation of Age*Gender**

| Age | Gender | | Total |
|---|---|---|---|
| | Female | Male | |
| 18 - 25 | 73 | 81 | 154 |
| 26 - 35 | 59 | 112 | 171 |
| 36 - 45 | 58 | 83 | 141 |
| 46 - 55 | 25 | 53 | 78 |
| 55 & above | 17 | 33 | 50 |
| Total | 232 | 362 | 594 |

**Table 5.3.1 is cross tabulation between Age and Gender of the respondents. It can be seen that majority of the respondents were male and belong to the age group 26-35.**

## [5.3.2] Cross Tabulation of Age*Annual Income

| Age | Annual Income | | | | Total |
|---|---|---|---|---|---|
| | Less than 5 lakh | 5 lakh to 10 lakh | 10 lakh to 15 lakh | More Than 15 lakh | |
| 18 - 25 | 107 | 40 | 4 | 3 | 154 |
| 26 - 35 | 81 | 65 | 18 | 7 | 171 |
| 36 - 45 | 50 | 44 | 31 | 16 | 141 |
| 46 - 55 | 23 | 17 | 19 | 19 | 78 |
| 55 & above | 23 | 11 | 6 | 10 | 50 |
| Total | 284 | 177 | 78 | 55 | 594 |

Table 5.3.2 is cross tabulation between Age and Annual income of the respondents. It can be seen that majority of the respondents have Annual Income and belong to the age group 18-25.

## [5.3.3] Cross Tabulation of Age*Investment Decision

| Age | Do You Invest? | | Total |
|---|---|---|---|
| | No | Yes | |
| 18 - 25 | 102 | 52 | 154 |
| 26 - 35 | 45 | 126 | 171 |
| 36 - 45 | 16 | 125 | 141 |
| 46 - 55 | 13 | 65 | 78 |
| 55 & above | 14 | 36 | 50 |
| Total | 190 | 404 | 594 |

Table 5.3.3 is cross tabulation between Age and investment Decision of the respondents. It can be seen that majority of the respondents invest and belong to the age group 26-35 and 36-45.

## [5.3.4] Cross Tabulation of Gender*Investment Decision

| Gender | Do You Invest? | | Total |
|---|---|---|---|
| | No | Yes | |
| Female | 83 | 149 | 232 |
| Male | 107 | 255 | 362 |
| Total | 190 | 404 | 594 |

Table 5.3.4 is cross tabulation between Gender and investment Decision of the respondents. It can be seen that majority of male respondents invest.

## [5.3.5] Cross Tabulation of Gender*Annual Income

| Gender | Annual Income | | | | Total |
|---|---|---|---|---|---|
| | Less than 5 lakh | 5 lakh to 10 lakh | 10 lakh to 15 lakh | More Than 15 lakh | |
| Female | 125 | 59 | 28 | 20 | 232 |
| Male | 159 | 118 | 50 | 35 | 362 |
| Total | 284 | 177 | 78 | 55 | 594 |

Table 5.3.5 is a cross tabulation between Gender and Annual income of the respondents. It can be seen that in our respondents 159 males and 125 females have annual income less than 5 lakh, 118 males and 59 females have annual income 5 lakh to 10 lakh, 50 males and 28 females have annual income between 10 lakh to 15 lakh and 35 males and 20 females have annual income more than 15 lakh.

Also by looking at the proportion of people in each Annual Income group we can say that, There are total of 284 people have income less than 5 lakh and among them 159 are male respondents, we can find out their proportion as 159/284=0.5599 i.e. almost 55.99% respondents are male.

177 people have income between 5 lakh to 10 lakh and among them 118 are male respondents, we can find out their proportion as 118/177=0.6667 i.e. almost 66.67% respondents are male.

78 people have income less between 10 lakh to 15 lakh and among them 50 are male respondents, we can find out their proportion as 50/78=0.6410 i.e. almost 64.10% respondents are male.

55 people have income more than 15 lakh and among them 35 are male respondents, we can find out their proportion as 35/55=0.6363 i.e. almost 63.63% respondents are male.

## [5.3.6] Cross Tabulation of Investment Decision* Annual Income

| Do you Invest? | Annual Income | | | | Total |
|---|---|---|---|---|---|
| | Less than 5 lakh | 5 lakh to 10 lakh | 10 lakh to 15 lakh | More Than 15 lakh | |
| No | 133 | 44 | 9 | 4 | 190 |
| Yes | 151 | 133 | 69 | 51 | 404 |
| Total | 284 | 177 | 78 | 55 | 594 |

Table 5.3.6 is a cross tabulation between Investment Decision and Annual income of the respondents. It can be seen that majority of respondents having annual income less than 5 lakh invest.

Also by looking at the proportion of people who are investor in each Annual Income group we can say that , There are total of 284 people who have income less than 5 lakh and among them 151 people invest , we can find out their proportion as 151/284=0.5317 i.e. almost 53.17% respondents invest.

177 people have income between 5 lakh to 10 lakh and among them 133 people invest, we can find out their proportion as 133/177=0.7514 i.e. almost 75.14% respondents invest.

78 people have income between 10 lakh to 15 lakh and among them 69 people invest, we can find out their proportion as 69/78=0.8846 i.e. almost 88.46% respondents invest.

55 people have income between more than 15 lakh and among them 51 people invest, we can find out their proportion as 51/55=0.9272 i.e. almost 92.72% respondents invest.

We can say that in our respondents proportion of investors having annual income less than 5 lakh as compare to other respondents is very less.

## [5.3.7] Cross Tabulation of Investment Decision* Qualification

| Do you Invest? | Qualification | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|
| | Below S.S.C | S.S.C | H.S.C | Diploma | Graduation | Post-graduation | Doctorate | |
| No | 4 | 10 | 0 | 78 | 23 | 66 | 9 | 190 |
| Yes | 5 | 12 | 3 | 184 | 57 | 114 | 29 | 404 |
| Total | 9 | 22 | 3 | 262 | 80 | 180 | 38 | 594 |

Table 5.3.7 is cross tabulation between Investment Decision and Qualification of the respondents. It can be seen that majority of respondents having qualification Diploma and Post-Graduation invest.

## [5.3.8] Cross Tabulation of Investment Decision* Region

| Do you Invest? | Region | | | | Total |
|---|---|---|---|---|---|
| | Mumbai | Mumbai Suburban | Thane | Others | |
| No | 57 | 61 | 51 | 21 | 190 |
| Yes | 171 | 105 | 113 | 15 | 404 |
| Total | 228 | 166 | 164 | 36 | 594 |

Table 5.3.8 is cross tabulation between Investment Decision and Region of the respondents. It can be seen that majority of respondents belong to region Mumbai do invest.

# [5.3.b] Analysis using Chi-Square :

### [5.3.9] Chi-Square association between Age*Gender

```
Pearson's Chi-squared test
data:  x1
X-squared = 8.3974, df = 4, p-value = 0.07806
```

We can conclude that there is no association between age and gender of the respondents

### [5.3.10] Chi-Square association between Age*Annual Income

```
Pearson's Chi-squared test
data: x2
X-squared = 102.33, df = 12, p-value < 2.2e-16
```

We can conclude that there is a association between age and Annual income of the respondents

### [5.3.11] Chi-Square association between Age*Investment Decision

```
Pearson's Chi-squared test
data: x3
X-squared = 121.94, df = 4, p-value < 2.2e-16
```

We can conclude that there is association between age and Investment decision of the respondents

33

## [5.3.12] Chi-Square association between Gender*Investment Decision

```
Pearson's Chi-squared test with Yates' continuity correction
data: x4
X-squared = 2.2349, df = 1, p-value = 0.1349
```

We can conclude that there is no association between Gender and Investment decision of the respondents

## [5.3.13] Chi-Square association between Gender*Annual Income

```
Pearson's Chi-squared test
data: x5
X-squared = 5.8628, df = 3, p-value = 0.1185
```

We can conclude that there is no association between Gender and Annual income of the respondents

## [5.3.14] Chi-Square association between Investment Decision* Annual Income

```
Pearson's Chi-squared test
data:  x6
X-squared = 63.332, df = 3, p-value = 1.14e-13
```

We can conclude that there is association between Investment Decision and Annual income of the respondents

## [5.3.15] Chi-Square association between Investment Decision* Qualification

```
Pearson's Chi-squared test
data:  x7
X-squared = 7.8799, df = 6, p-value = 0.247
```

We can conclude that there is no association between Investment decision and Qualification of the respondents

## [5.3.16] Chi-Square association between Investment Decision* Region

```
Pearson's Chi-squared test
data:  x8
X-squared = 18.391, df = 3, p-value = 0.0003653
```

We can conclude that there is association between Investment Decision and Region of the respondents

# [5.3.c] Analysis using Binary Logistic Regression:

**Dependent Variable**

Do you Invest?

**Independent Variable**

Age

Gender

Occupation

Qualification

Family Size

No of Earning Members

Region
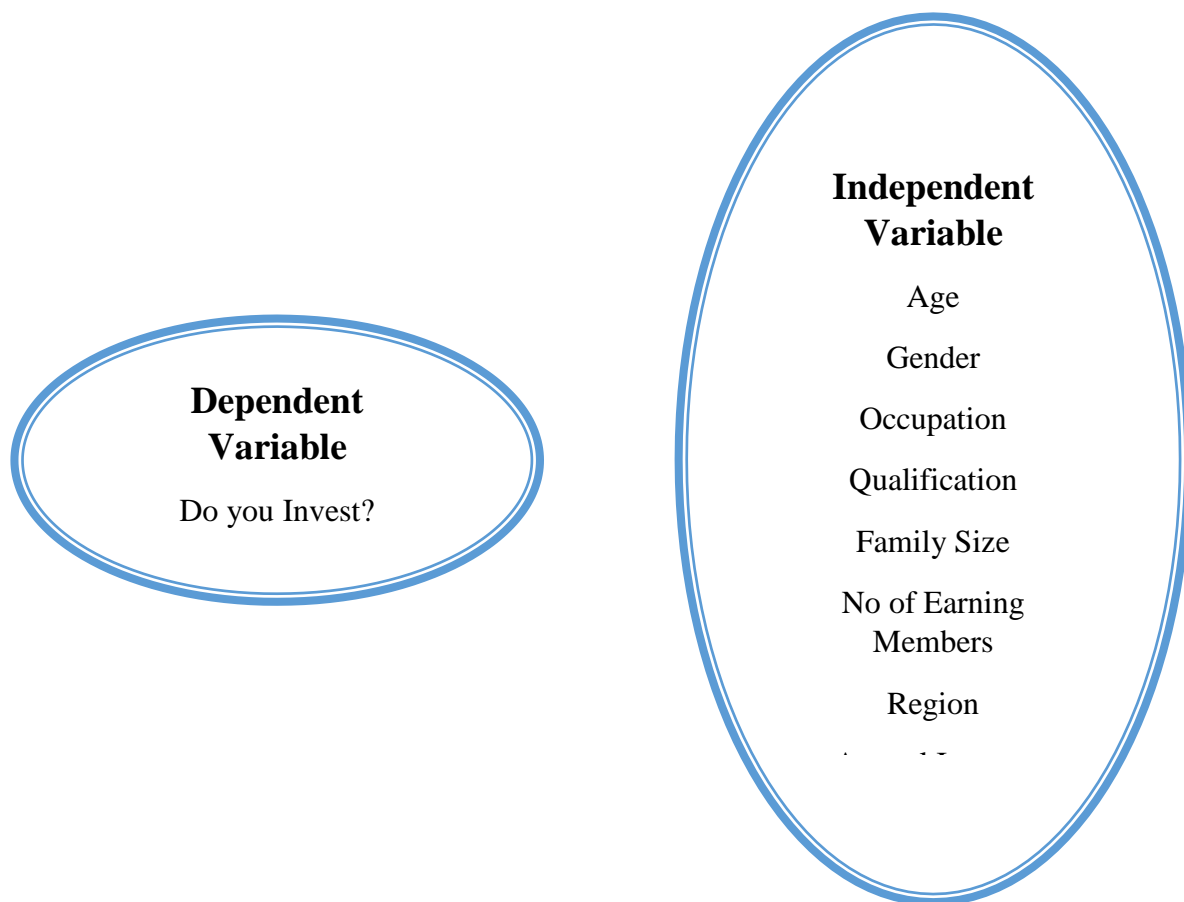
```
Call:
glm(formula = Do.you.Invest. ~ ., family = binomial, data = training_set)

Deviance Residuals:
    Min       1Q    Median       3Q      Max
-2.3485  -0.9251    0.5975    0.8240   1.8619

Coefficients:
                        Estimate Std. Error z value Pr(>|z|)
(Intercept)              2.20160    0.63345   3.476  0.00051 ***
Age                      0.33277    0.10460   3.181  0.00147 **
Gender                   0.23823    0.23731   1.004  0.31544
Qualification            0.02070    0.09457   0.219  0.82673
Occupation              -0.30146    0.07580  -3.977 6.98e-05 ***
Family.Size             -0.22573    0.09916  -2.276  0.02282 *
No.of.earning.members    0.12221    0.13952   0.876  0.38106
Region                  -0.08265    0.09670  -0.855  0.39272
Annual.Income           -0.37740    0.15234  -2.477  0.01323 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 521.37  on 415  degrees of freedom
Residual deviance: 452.38  on 407  degrees of freedom
AIC: 470.38

Number of Fisher Scoring iterations: 4
```

35

**From above summary we can see that,**

**P – value of Age, Occupation, Family Size & Annual Income are less than 5% level of significance. Therefore, they are significant. Only significant variables are considered for final prediction model.**

```
Call:
glm(formula = Do.you.Invest. ~ ., family = binomial, data = training_set2)

Deviance Residuals:
    Min      1Q   Median      3Q      Max
-2.4264  -0.9335   0.5964   0.8313   1.7278

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.66148    0.43416   3.827  0.00013 ***
Age            0.49492    0.11072   4.470 7.83e-06 ***
Occupation    -0.22391    0.07488  -2.990  0.00279 **
Family.Size   -0.17154    0.08276  -2.073  0.03819 *
Annual.Income -0.28997    0.14525  -1.996  0.04589 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 521.37  on 415  degrees of freedom
Residual deviance: 454.58  on 411  degrees of freedom
AIC: 464.58

Number of Fisher Scoring iterations: 4
```

**From above summary we can see, the final model is**

**Y = 1.66148 + 0.49492(Age) – 0.22391(Occupation) – 0.17154(Family Size) – 0.28997(Annual Income).**

```
Confusion Matrix
```

|   | 0 | 1 |
|---|---|---|
| **0** | 29 | 28 |
| **1** | 14 | 107 |

**From above confusion matrix we can see that 29, 107 times our prediction was accurate and 14, 28 times prediction was not accurate**.

```
[1] "Accuracy of model is: 0.764044943820225"
```
 Accuracy of model was 76%

# [5.4] Conclusion:

## Odds Ratio:

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

| Odds Ratio | | | |
|---|---|---|---|
| **Variable** | Estimate | 95 % Confidence Interval | |
| | | Lower | Upper |
| **Age( 0 vs 1)** | 5.449343 | 3.402920 | 8.862435 |
| **Age( 0 vs 2)** | 15.051087 | 8.292104 | 28.859000 |
| **Age( 0 vs 3)** | 9.624233 | 4.986436 | 19.832782 |
| **Age( 0 vs 4)** | 4.971925 | 2.503838 | 10.357815 |
| **Occupation(0 vs 1)** | 0.3219958 | 0.12381810 | 0.8139527 |
| **Occupation(0 vs 2)** | 2.3106160 | 0.90307131 | 6.1199849 |
| **Occupation(0 vs 3)** | 0.7791295 | 0.37048250 | 1.5346444 |
| **Occupation(0 vs 4)** | 0.6572008 | 0.26181474 | 1.6345245 |
| **Occupation(0 vs 5)** | 0.2221749 | 0.08073247 | 0.5863739 |
| **Occupation(0 vs 6)** | 0.1205792 | 0.0529327 | 0.2507377 |
| **Annual Income(0 vs 1)** | 0.4003609 | 0.1730792 | 0.8367711 |
| **Annual Income(0 vs 2)** | 0.1508439 | 0.0675460 | 0.2999007 |
| **Annual Income(0 vs 3)** | 1.6251550 | 0.4892617 | 6.4971020 |
| **Family Size (0 vs 1)** | 1.3501865 | 0.60501123 | 2.9223960 |
| **Family Size (0 vs 2)** | 1.3930095 | 0.64273818 | 2.9133792 |
| **Family Size (0 vs 3)** | 1.4072868 | 0.62849509 | 3.0586231 |
| **Family Size (0 vs 4)** | 1.1300887 | 0.45308393 | 2.7908033 |
| **Family Size (0 vs 5)** | 0.3138163 | 0.09382392 | 0.9736965 |
| **Family Size (0 vs 6)** | 0.4790508 | 0.15670068 | 1.4197143 |

## Odds Ratio Interpretation

**Significant Variables for Investment Decision are:**

## Age

An investor whose age is 26 - 35 is 5.49 times more likely to invest then investor whose age is 18 – 25.

An investor whose age is 36 – 45 is 15.05 times more likely to invest then investor whose age is 18 – 25.

An investor whose age is 46 - 55 is 9.62 times more likely to invest then investor whose age is 18 – 25.

An investor whose age is 55 & above is 4.97 times more likely to invest then investor whose age is 18 – 25.

## **Occupation**

An investor who is employed(government) is 0.32 times more likely to invest then investor who is student & employed.

An investor who is employed(private) is 2.31 times more likely to invest then investor who is student & employed.

An investor who is earning housewife is 0.77 times more likely to invest then investor who is student & employed.

An investor who has business is 0.65 times more likely to invest then investor who is student & employed.

An investor who are professionals is 0.22 times more likely to invest then investor who is student & employed.

An investor who is retired is 0.12 times more likely to invest then investor who is student & employed.

## **Annual Income**

An investor whose annual income is 5 – 10 Lakhs is 0.40 times more likely to invest then investor whose annual income is less than 5 Lakhs.

An investor whose annual income is 10 - 15 Lakhs is 0.15 times more likely to invest then investor whose annual income is less than 5 Lakhs.

An investor whose annual income is more than 15 Lakhs is 1.62 times more likely to invest then investor whose annual income is less than 5 Lakhs.

## **Family Size**

An investor whose family size is 3 is 1.35 times more likely to invest then investor whose family size is 2.

An investor whose family size is 4 is 1.39 times more likely to invest then investor whose family size is 2.

An investor whose family size is 5 is 1.40 times more likely to invest then investor whose family size is 2.

An investor whose family size is 6 is 1.13 times more likely to invest then investor whose family size is 2.

An investor whose family size is 7 is 0.31 times more likely to invest then investor whose family size is 2.

An investor whose family size is more than 7 is 0.47 times more likely to invest then investor whose family size is 2.

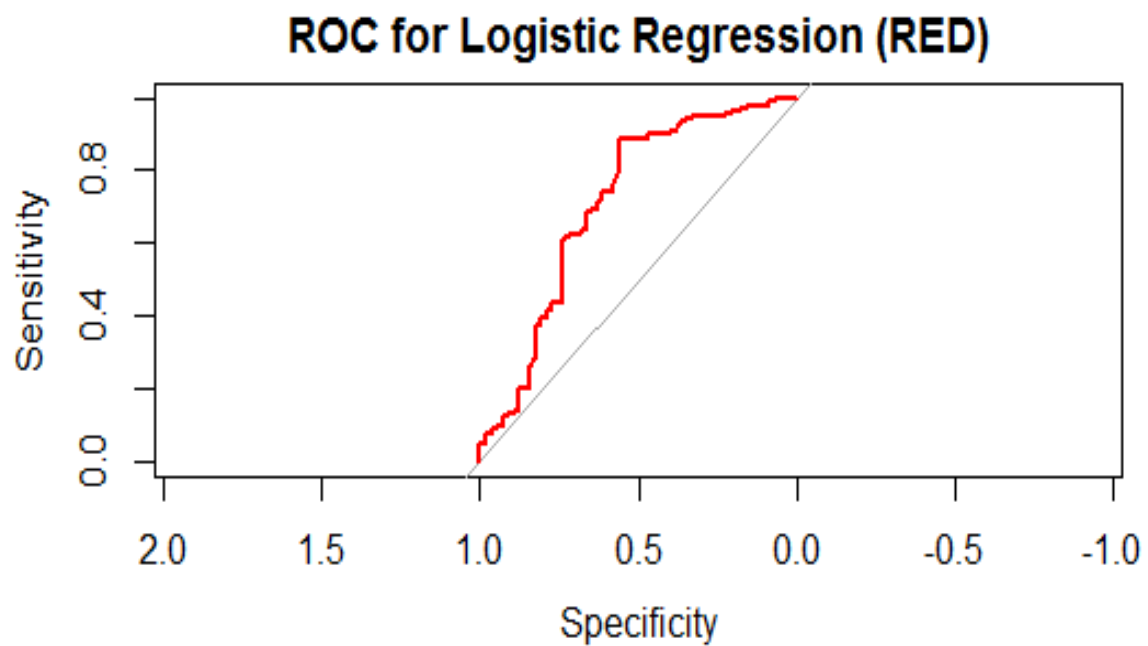# ROC Curve:



**ROC for Logistic Regression (RED)**

**Area under the curve: 0.7107**
**Since, our ROC curve rises quickly i.e. both sensitivity and specificity are high (thus 1-specificity is low). Hence our model has high predictive accuracy.**

# [6.1] Objective (3):

**To know the relation between "Saving & Investment".**

# [6.2] Techniques:

## [6.2.a] Spearman rank correlation:

In statistics, Spearman's rank correlation coefficient or Spearman's ρ, named after Charles Spearman and often denoted by the Greek letter ρ (rho) or as is a nonparametric measure of rank correlation (statistical dependence between the rankings of two variables). It assesses how well the relationship between two variables can be described using a monotonic functional statistics, Spearman's rank correlation coefficient or Spearman's ρ, is a nonparametric measure of rank correlation (statistical dependence between the rankings of two variables). It assesses how well the relationship between two variables can be described using a monotonic function

The Spearman correlation coefficient is defined as the Pearson correlation coefficient between the rank variables. Spearman's correlation coefficient is often denoted rho ρ and measures the monotonic(A monotonic function is a function of a dependent variable that is always increasing or decreasing as its independent variable increases) relationship of the variables rather than the linear association in the Pearson setting. Thus, Spearman's correlation coefficient is more reliable with non-linear data compared to Pearson's r.

When the data contains no ties, ρ can be found by taking the difference of the ranked values using the following equation:

$$r_s = \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Where di is the difference of the ranked variables and n is the number of samples.
Since the dataset contains ties, the above formula will report incorrect values. When the data contain ties, the following equation can be used:

$$r_s = \rho_{rg_x . g_y} = \frac{cov(rg_x, g_y)}{\sigma_{rg_x} \sigma_{rg_y}}$$

Where ρ denotes the usual Pearson coefficient but on the ranked variables. $cov(rg_x, g_y)$ is the covariance and $\sigma_{rg_x} \sigma_{rg_y}$ are the standard deviations of the ranked variables.
To calculate Spearman's ρ in R, first, rank the xx and y variables.

## [6.2.b] Kendall correlation:

The Kendall correlation method measures the correspondence between the ranking of x and y variables. The total number of possible pairings of x with y observations is $n(n - 1)/2$, where n is the size of x and y.
The procedure is as follow:

- Begin by ordering the pairs by the x values. If x and y are correlated, then they would have the same relative rank orders.

- Now, for each $y_i$, count the number of $y_j > y_i$ (concordant pairs (**c**)) and the number of $y_j < y_i$(discordant pairs (**d**)).

Kendall correlation distance is defined as follow:

$$\text{tau} = \frac{n_c - n_d}{\frac{1}{2}n(n-1)}$$

Where,

- $n_c$: total number of concordant pairs
- $n_d$: total number of discordant pairs
- n: size of x and y

# [6.3] Analysis:

## [6.3.a] Analysis using Spearman Rank correlation:

Spearman's rank correlation rho

data:  data1$saving and data1$investment
S = 4182200, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.6166115

# [6.4.a] Conclusion:

Here our rho value is 0.6166115 that is approximately 0.62 , hence it shows the positive monotonic association between "Saving" and "Investment" that is  if our monthly saving increases then the investment amount of peoples are also more . hence we have a positive correlation between "saving" and "investment"

## [6.3.b] Analysis using Kendall correlation:

[1]r= 0.4029974
[2] p-value < 2.2e-16

**Decision Criteria**

Correlation coefficient is comprised between **-1** and **1**:

- **-1** indicates a strong **negative correlation**: this means that every time **x increases**, **y decreases**
- **0** means that there is no **association** between the two variables (x and y)
- **1** indicates a strong **positive correlation**: this means that **y increases** with **x**

# [6.4.b] Conclusion:

Here our rho value is 0.4029974 that is approximately 0.4, hence it shows the positive association between "Saving" and "Investment" that is  if our monthly saving increases then the investment amount of peoples are also more . hence we have a positive correlation between "saving" and "investment"

# [7.1] Objective (4):

**To study the pattern of salaried people working in different sectors.**

# [7.2] Techniques:

[a] **Descriptive Statistics**

[b] **Contingency Table & Chi-Square Association**

# [7.3] Analysis:

**Descriptive of Salaried People investing in different investment avenues**

| Factors | Distribution of factors | Frequency | Percentage |
|---|---|---|---|
| Gender | Female | 149 | 37% |
| | Male | 255 | 63% |
| | **Total** | **404** | **100%** |
| Age | 18 - 25 | 52 | 13% |
| | 26 - 35 | 126 | 31% |
| | 36 - 45 | 125 | 31% |
| | 46 - 55 | 65 | 16% |
| | 55 & above | 36 | 9% |
| | **Total** | **404** | **100%** |
| Qualification | Below S.S.C | 5 | 1% |
| | S.S.C | 29 | 7% |
| | H.S.C | 57 | 14% |
| | Diploma | 12 | 3% |
| | Graduation | 184 | 46% |
| | Post Graduation | 114 | 28% |
| | Doctorate | 3 | 1% |
| | **Total** | **404** | **100%** |
| Occupation | Student and Employed | 33 | 8% |
| | Employed (Government) | 84 | 21% |
| | Employed (Private) | 173 | 43% |
| | Business | 48 | 12% |
| | Professionals ( Teachers, Doctors, etc) | 34 | 8% |
| | Earning Housewife | 19 | 5% |
| | Retired | 13 | 3% |
| | **Total** | **404** | **100%** |
| Annual Income | Less than 5 lakh | 151 | 37% |
| | 5 lakh to 10 lakh | 133 | 33% |
| | 10 lakh to 15 lakh | 69 | 17% |
| | more than 15 lakh | 51 | 13% |
| | **Total** | **404** | **100%** |
| Family Size | 2 | 23 | 6% |
| | 3 | 93 | 23% |
| | 4 | 145 | 36% |

| | | | | |
|---|---|---|---|---|
| | 5 | 92 | 23% | |
| | 6 | 34 | 8% | |
| | 7 | 7 | 1.7% | |
| | More than 7 | 10 | 2.3% | |
| | Total | 404 | 100% | |
| Earning Members | 1 | 119 | 29% | |
| | 2 | 195 | 48% | |
| | 3 | 65 | 16% | |
| | 4 | 21 | 5% | |
| | 5 | 4 | 2% | |
| | Total | 404 | 100% | |
| Region | Mumbai | 171 | 42% | |
| | Mumbai Suburban | 105 | 26% | |
| | Thane | 113 | 28% | |
| | Others | 15 | 4% | |
| | Total | 404 | 100% | |
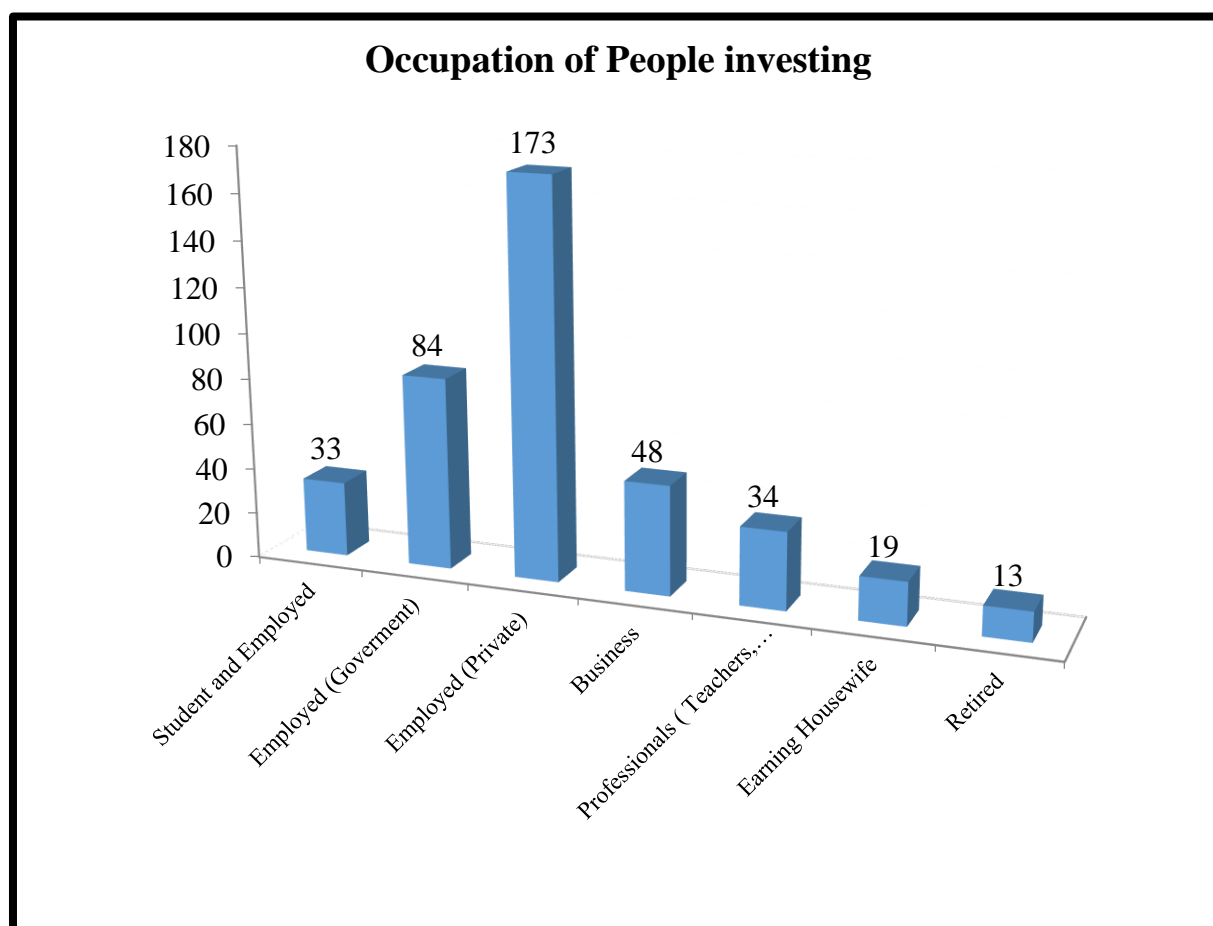


Fig 7.3.1

**From the fig 7.3.1, we can see that the out of 404 people 173 were working as private employee and 84 people works as government employee. We can also see that Retired and housewife invest less as compared to other occupation. Also there is good no. of respondents who are student but they are investing actively.**

**Income distribution of people investing**

| | Less than 5 lakh | 5 lakh to 10 lakh | 10 lakh to 15 lakh | more than 15 lakh |
|---|---|---|---|---|
| **Series1** | 151 | 133 | 69 | 51 |
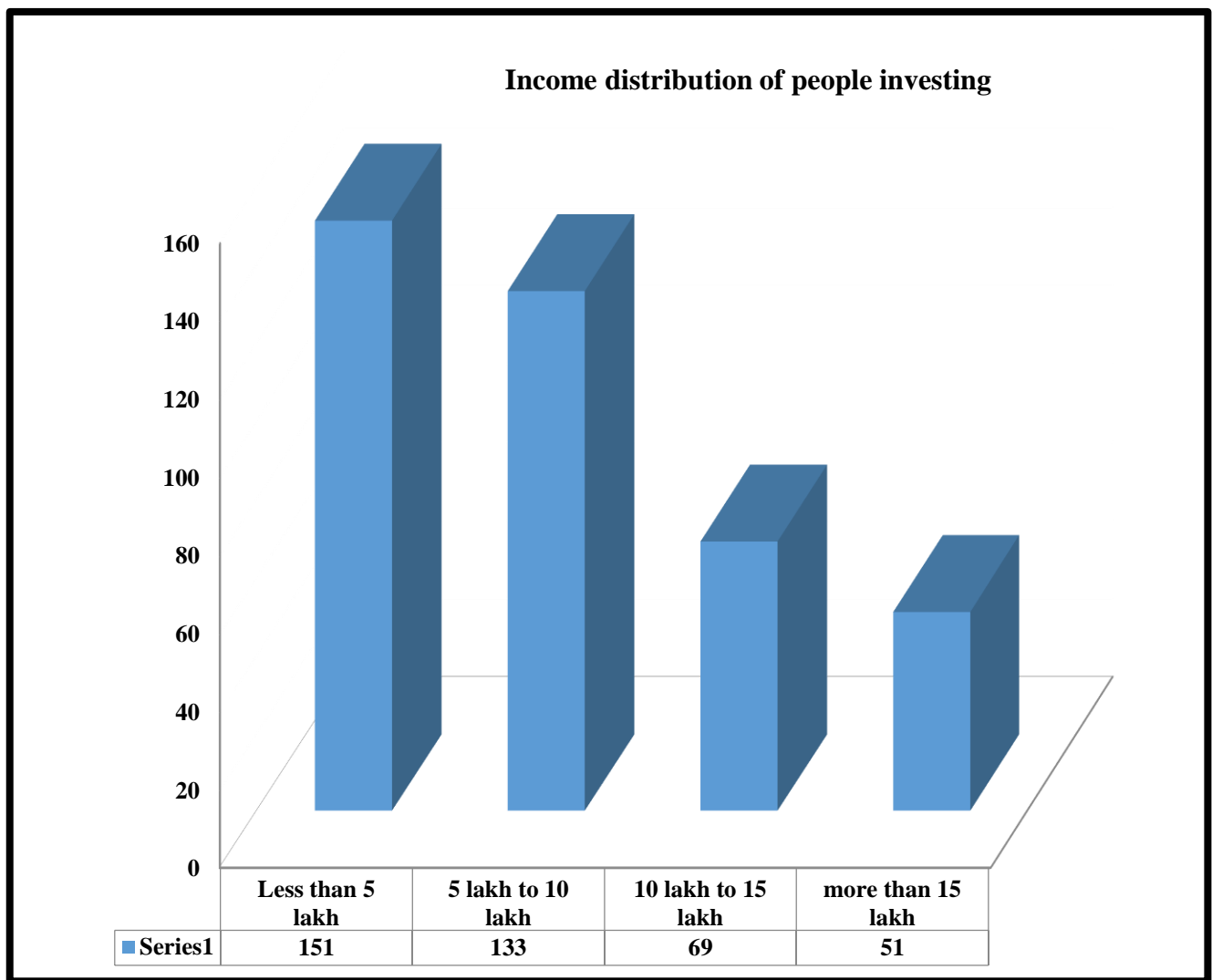
Fig 7.3.2

**From the fig 7.3.2, we can see that the out of 404 people 151 people earn less than 5 lakh Rupees annually. We can also conclude from our data salary and no. of people investing in different avenues is inversely proportional in our data, as salary is more there is less no. of respondents. Also there is good no. of respondents who are earning less as compared to other people but they are investing actively.**

| Occupation | Professionals | | | | | | |
|---|---|---|---|---|---|---|---|
| | Below S.S.C | S.S.C | H.S.C | Diploma | Graduation | Post Graduation | Doctorate |
| Business | 0 | 10 | 9 | 1 | 18 | 10 | 0 |
| Earning Housewife | 5 | 3 | 5 | 0 | 4 | 2 | 0 |
| Employed (Government) | 0 | 8 | 11 | 4 | 38 | 23 | 0 |
| Employed (Private) | 0 | 5 | 23 | 5 | 97 | 43 | 0 |
| Professionals ( Teachers, Doctors, etc) | 0 | 0 | 1 | 0 | 9 | 21 | 3 |
| Retired | 0 | 3 | 2 | 1 | 4 | 3 | 0 |
| Student and Employed | 0 | 0 | 6 | 1 | 14 | 12 | 0 |

## Hypothesis:

$H_0$=There is no association between both the variables.

$H_1$= There is an association between both the variables

## Decision Criteria:

If p-value<0.05

Then reject $H_0$

## [7.3.3] Chi-Square association between Occupation*Qualification Of respondents who are investing in different investment avenues

```
Pearson's Chi-squared test

data:  a
X-squared = 69.563, df = 18, p-value = 5.357e-08
```

## [7.4.a] Conclusion:

We can conclude that there is association between Qualification and Occupation of the respondents

| Occupation | Investment Decision | | Total |
|---|---|---|---|
| | No | Yes | |
| Business | 12 | 48 | 60 |
| Earning Housewife | 15 | 19 | 34 |
| Employed (Government) | 9 | 84 | 93 |
| Employed (Private) | 56 | 173 | 229 |
| Professionals ( Teachers, Doctors, etc) | 13 | 34 | 47 |
| Retired | 15 | 13 | 28 |
| Student and Employed | 70 | 33 | 103 |
| Total | 190 | 404 | 594 |

**[7.3.4] Chi-Square association between Investment Decision\*Occupation of respondents**

```
Pearson's Chi-squared test

data:  b
X-squared = 101.18, df = 6, p-value < 2.2e-16
```

# [7.4.b] Conclusion:

**We can conclude that there is an association between Investment decision and Occupation of the respondents.**
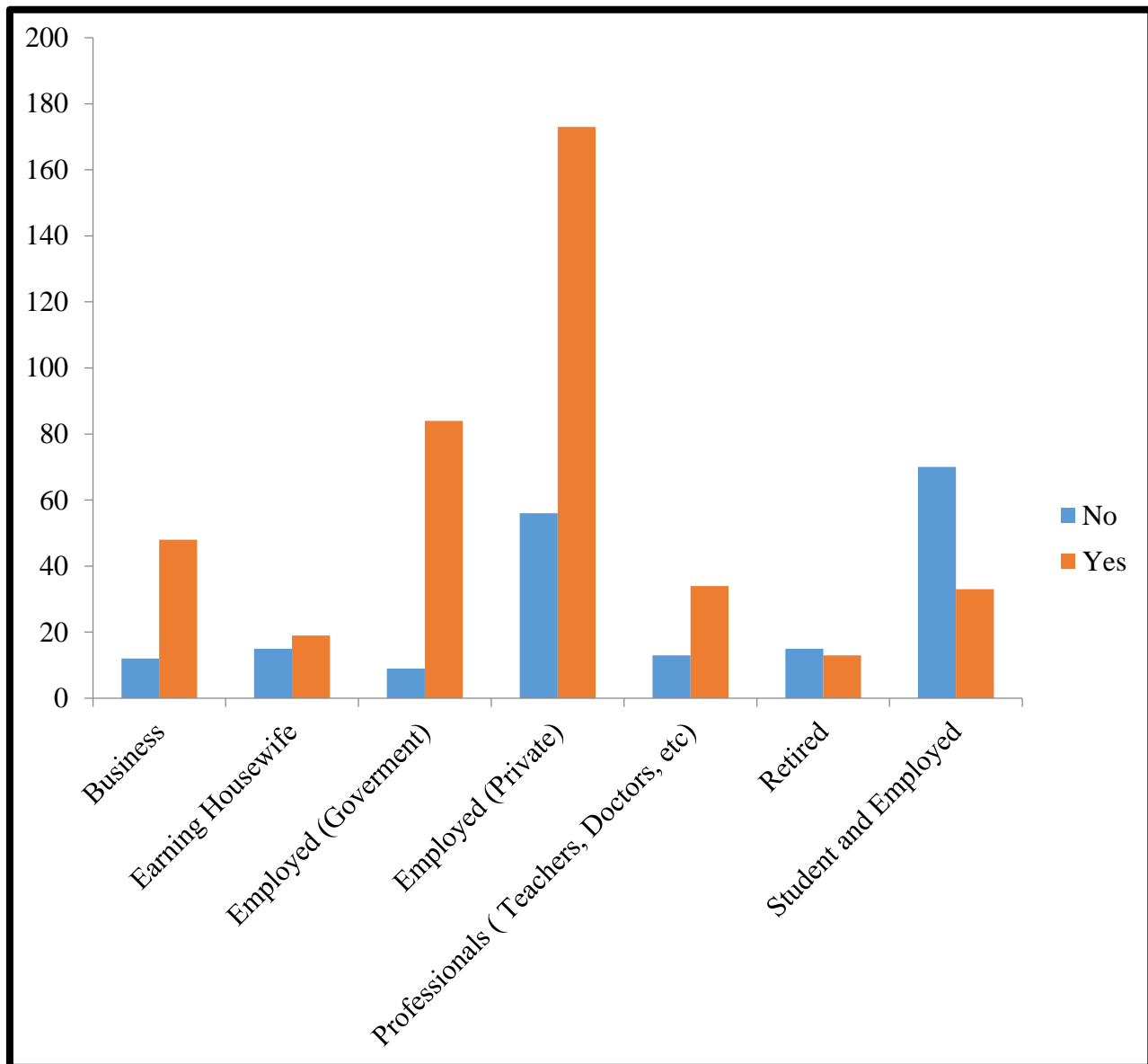


Fig 7.3.5

**From the fig 7.3.5, also we can see that investment decision is depending upon once earning or occupation, as in our data most of the student are not investing and any private and government sector employees are investing.**

# [8.1] Objective (5):

**To understand the investors behaviour on the basis of demographic factors, awareness & risk – return in their avenues**

# [8.2] Techniques:

## [8.2.a] Principal Component Analysis

Principal Component Analysis (PCA) is a statistical technique used for data reduction without losing its properties. Basically, it describes the composition of variances and covariances through several linear combinations of the primary variables, without missing an important part of the original information. In another term, it is about obtaining a unique set of orthogonal axes where the data has the largest variance. Its main aim is to overcome the dimensionality of the problem. The reduction of dimensionality should be such that when dropping higher dimensions, the loss of data is minimum.

Also, the interpretation of principal components can explain associations among variables that are not visible at first glance. It helps analyze the scattering of the observations and recognize the variables responsible for distribution.

**Properties of Principal Components are**:
1. They are a set of primary data variables projected in different directions, similar to the properties of original variables.
2. It is commonly used in Machine Learning and Data Science for dimensionality reduction.
3. They are
orthogonal.
4. If we find PC one by one, then the variance or the variation of the Principal Components reduces as. This means that the 1st PC has the highest variance and the last PC has the least variance.

Before moving on to the computation of Principal Components, you should have the following knowledge:
1. **Variance**: Variance is used to compute the variation of the data points distributed across the dimensionality graph. Mathematically, it is the average squared variation from the mean value. To calculate Var(X) we use the following formula:

$$\text{Var(x)} = \sum \frac{(x_i - \bar{X})^2}{N}$$

2.**Covariance**: With covariance, we can estimate the degree to which analogous components from a couple of sets of grouped data move in an identical direction. In simple words, it is used to identify the dependencies and relationships between the characteristics of datasets. Below is the formula for calculating the Cov (x, y):

$$\text{Cov(x,y)} = \sum \frac{(x_i - \bar{X})(y_i - \bar{Y})}{N}$$

where *xi* and *yi* are the value of x and y in $i^{th}$ dimension. $\bar{x}$ and $\bar{y}$ express the mean.

3. **Eigen Vectors and Eigen Values:** It is used to make alterations in data comprehensible. It can also be understood as expanding/contracting an X-Y graph without altering the directions. An **Eigenvalue** is a value indicating the variance in a particular direction.   4. **Principal Components**: The fresh set of data variables that are collected from the original data set is called Principal Components. The new data variables are extremely meaningful and independent. They possess all the valuable information of the original variables

47

# [8.2.b] Clustering

**Clustering** is one of the most common exploratory data analysis technique used to get an intuition about the structure of the data. It can be defined as the task of identifying subgroups in the data such that data points in the same subgroup (cluster) are very similar while data points in different clusters are very different. In other words, we try to find homogeneous subgroups within the data such that data points in each cluster are as similar as possible according to a similarity measure such as euclidean-based distance or correlation-based distance. The decision of which similarity measure to use is application-specific.

Clustering analysis can be done on the basis of features where we try to find subgroups of samples based on features or on the basis of samples where we try to find subgroups of features based on samples. We'll cover here clustering based on features. Clustering is used in market segmentation; where we try to fined investors that are similar to each other whether in terms of behaviors or attributes, image segmentation/compression; where we try to group similar regions together, document clustering based on topics, etc.

Unlike supervised learning, clustering is considered an unsupervised learning method since we don't have the ground truth to compare the output of the clustering algorithm to the true labels to evaluate its performance. We only want to try to investigate the structure of the data by grouping the data points into distinct subgroups.

# K-Means Clustering

**Kmeans** algorithm is an iterative algorithm that tries to partition the dataset into $K$ pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to **only one group**. It tries to make the inter-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid (arithmetic mean of all the data points that belong to that cluster) is at the minimum. The less variation we have within clusters, the more homogeneous (similar) the data points are within the same cluster.

# [8.3] Analysis:
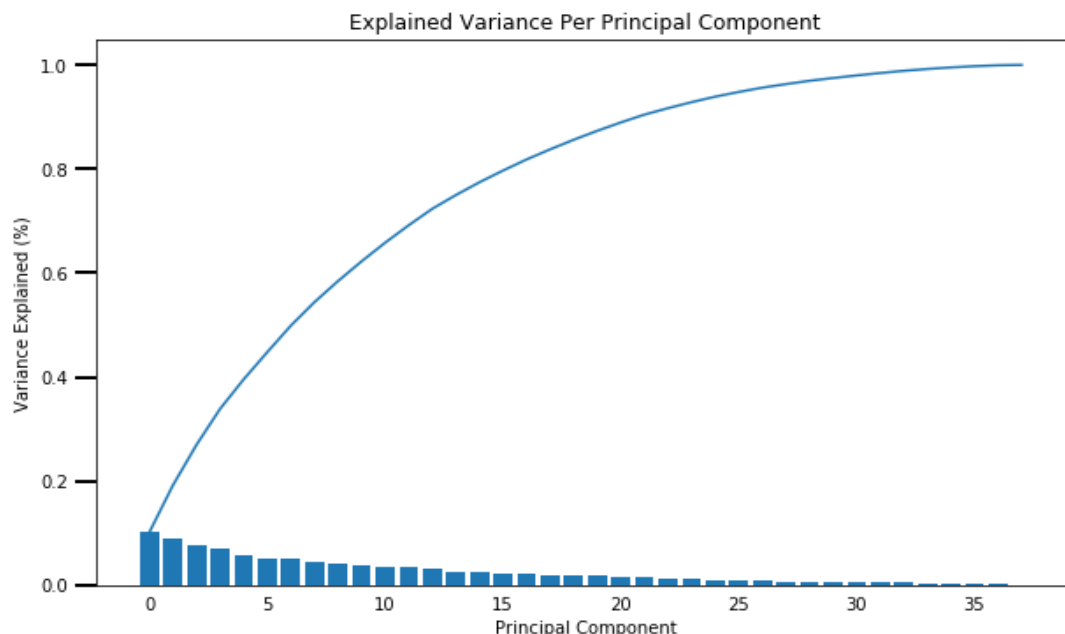


Fig 8.3.1

**From fig 8.3.1, we can see that, 25 components can be selected which explains more than 85% of total variation in the data**
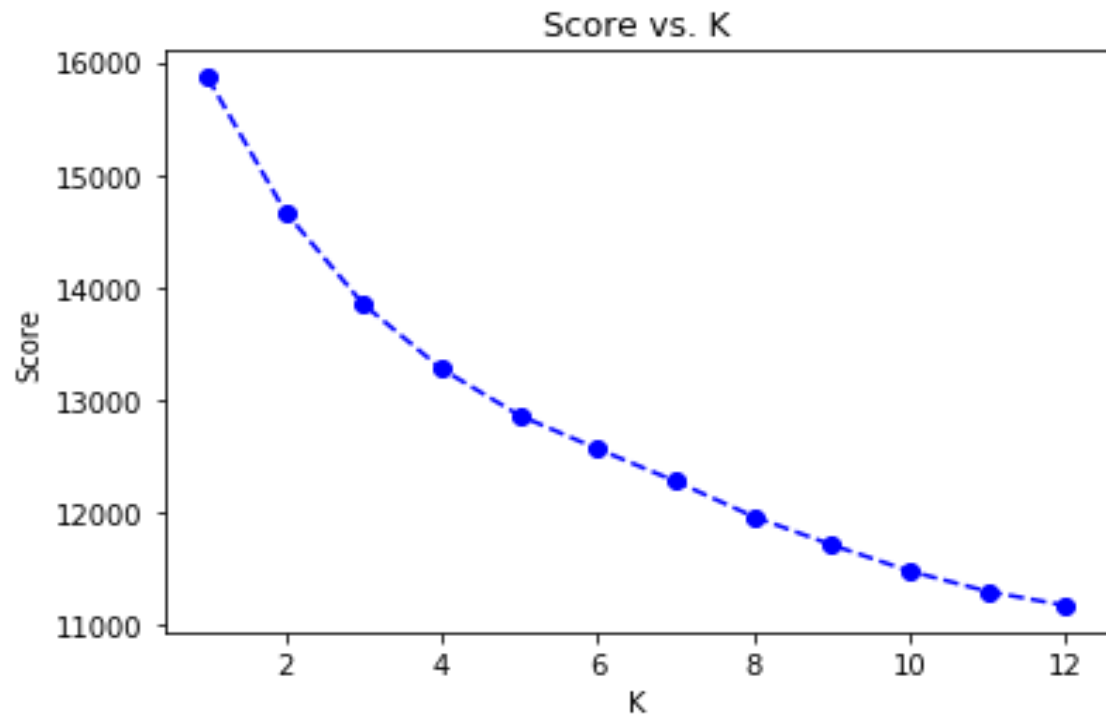
Fig 8.3.2

**From fig 8.3.2 of elbow curve, we see that 8 clusters were formed.**
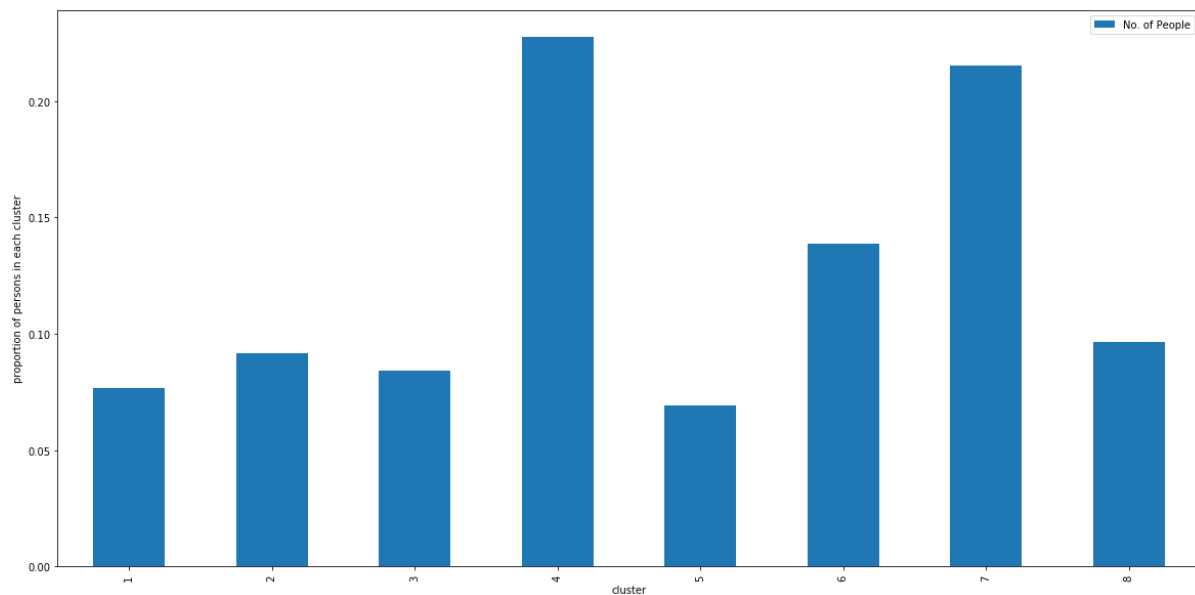


Fig 8.3.3

**From fig 8.3.3, we can see 8 clusters of which we can see 4$^{th}$ cluster contains highest proportion of investors and 5$^{th}$ cluster contains minimum proportion of investors. Now we will analyse the cluster with maximum and minimum proportion to understand the investor behaviour.**

49

**In 4ᵗʰ cluster:**

| Risk of Stocks | Returns of Stocks | | | Grand Total |
|---|---|---|---|---|
| | **High Returns** | **Medium Returns** | **Not Invested** | |
| **High Risk** | 6.52% | 3.26% | 0.00% | 9.78% |
| **Low Risk** | 1.09% | 0.00% | 0.00% | 1.09% |
| **Medium Risk** | 4.35% | 2.17% | 0.00% | 6.52% |
| **Not Invested** | 0.00% | 0.00% | 82.61% | 82.61% |
| **Grand Total** | 11.96% | 5.43% | 82.61% | 100.00% |

Table 8.3.4

| No of earning members | Savings | | | | Grand Total |
|---|---|---|---|---|---|
| | **Less than 10%** | **10% to 25%** | **25% to 50%** | **More than 50%** | |
| **1** | 11.96% | 8.70% | 5.43% | 4.35% | 30.43% |
| **2** | 21.74% | 18.48% | 7.61% | 6.52% | 54.35% |
| **3** | 4.35% | 7.61% | 3.26% | 0.00% | 14.13% |
| **Grand Total** | 38.04% | 34.78% | 16.30% | 10.87% | 100.00% |

Table 8.3.5

| Qualification | Awareness of Mutual Funds | | | Grand Total |
|---|---|---|---|---|
| | **Not at all Aware** | **Slightly Aware** | **Extremely Aware** | |
| **Below S.S.C** | 4.35% | 0.00% | 0.00% | 4.35% |
| **S.S.C** | 9.78% | 3.26% | 0.00% | 13.04% |
| **H.S.C** | 6.52% | 8.70% | 3.26% | 18.48% |
| **Diploma** | 0.00% | 0.00% | 1.09% | 1.09% |
| **Graduation** | 11.96% | 22.83% | 6.52% | 41.30% |
| **Post Graduation** | 3.26% | 13.04% | 5.43% | 21.74% |
| **Grand Total** | 35.87% | 47.83% | 16.30% | 100.00% |

Table 8.3.6

| Region | Percent of investment from savings | | | | Grand Total |
|---|---|---|---|---|---|
| | Less than 5% | 5% to 10% | 10% to 20% | More than 20% | |
| Mumbai | 10.87% | 5.43% | 7.61% | 3.26% | 27.17% |
| Mumbai Suburban | 15.22% | 8.70% | 3.26% | 4.35% | 31.52% |
| Thane | 19.57% | 4.35% | 7.61% | 6.52% | 38.04% |
| Others | 2.17% | 1.09% | 0.00% | 0.00% | 3.26% |
| Grand Total | 47.83% | 19.57% | 18.48% | 14.13% | 100.00% |

Table 8.3.7

| Occupation | Reason for investment | | | Grand Total |
|---|---|---|---|---|
| | Assets | Child Education | Health Care | |
| Business | 3.26% | 18.48% | 3.26% | 25.00% |
| Earning Housewife | 5.43% | 6.52% | 1.09% | 13.04% |
| Employed (Government) | 3.26% | 11.96% | 2.17% | 17.39% |
| Employed (Private) | 9.78% | 29.35% | 0.00% | 39.13% |
| Professionals ( Teachers, Doctors, etc) | 0.00% | 4.35% | 0.00% | 4.35% |
| Retired | 0.00% | 1.09% | 0.00% | 1.09% |
| Grand Total | 21.74% | 71.74% | 6.52% | 100.00% |

Table 8.3.8

| Insurance Risk | Insurance Returns | | | | Grand Total |
|---|---|---|---|---|---|
| | Low Returns | Medium Returns | High Returns | Not Invested | |
| Low Risk | 7.61% | 11.96% | 3.26% | 0.00% | 22.83% |
| Medium Risk | 0.00% | 15.22% | 4.35% | 1.09% | 20.65% |
| Not Invested | 0.00% | 0.00% | 0.00% | 56.52% | 56.52% |
| Grand Total | 7.61% | 27.17% | 7.61% | 57.61% | 100.00% |

Table 8.3.9

# [8.4.a] Conclusion:

**From table no. 8.3.4, we can see that investors who have taken any kind of risks in stocks expects High and Medium Returns. No investors expects low returns. 82.61% don't invest in Stocks**

**From table no. 8.3.5, we can see that if no of earning members is less the percent of savings is less, also if no of earning members increases percent of savings is upto 50% atmost**

**From the table no 8.3.6, we can see that Qualification have more influence on Awareness level of Mutual Funds. The investors with qualification less than H.S.C have less awareness as compared to higher qualification**

**From table no. 8.3.7, we can see that investors from above mentioned regions give more preference to invest less than 5% from their monthly savings. Following most of investors in Mumbai Suburban invest 5% to 10% from their monthly savings. Also few investors from Mumbai and Thane region prefer to invest in more than 10% from their monthly savings**

**From table no. 8.3.8, we can see that most of investors reason for investment was their child's education. And few investors with occupation Business, Earning housewife, Employed (Government as well as Private) had assets purchasing as their reason for investment followed by health care**

**From the table no.8.3.9, we observe that 15.22% of investors takes medium risk in Insurance and expects medium returns, followed by 11.96 % of investors have taken low risk and expects medium returns.  And the majority of 56.52% of investors have not at all invested in Insurance.**

**In 5<sup>th</sup> cluster:**

| Stocks Risk | Stocks Returns | | | Grand Total |
|---|---|---|---|---|
| | **High Returns** | **Medium Returns** | **Not Invested** | |
| **High Risk** | 10.71% | 3.57% | 0.00% | 14.29% |
| **Medium Risk** | 7.14% | 3.57% | 0.00% | 10.71% |
| **Not Invested** | 0.00% | 0.00% | 75.00% | 75.00% |
| **Grand Total** | 17.86% | 7.14% | 75.00% | 100.00% |

Table 8.3.10

| No of earning members | Savings | | | | Grand Total |
|---|---|---|---|---|---|
| | Less than 10% | 10% to 25% | 25% to 50% | More than 50% | |
| 1 | 3.57% | 7.14% | 14.29% | 0.00% | 25.00% |
| 2 | 10.71% | 14.29% | 7.14% | 10.71% | 42.86% |
| 3 | 0.00% | 3.57% | 14.29% | 7.14% | 25.00% |
| 4 | 0.00% | 3.57% | 0.00% | 0.00% | 3.57% |
| 5 | 0.00% | 0.00% | 3.57% | 0.00% | 3.57% |
| Grand Total | 14.29% | 28.57% | 39.29% | 17.86% | 100.00% |

Table 8.3.11

| Qualification | Awareness level of Mutual funds | | | Grand Total |
|---|---|---|---|---|
| | Not at all Aware | Slightly Aware | Extremely Aware | |
| S.S.C | 3.57% | 0.00% | 0.00% | 3.57% |
| H.S.C | 3.57% | 0.00% | 0.00% | 3.57% |
| Diploma | 3.57% | 0.00% | 3.57% | 7.14% |
| Graduation | 3.57% | 28.57% | 3.57% | 35.71% |
| Post Graduation | 3.57% | 28.57% | 17.86% | 50.00% |
| Grand Total | 17.86% | 57.14% | 25.00% | 100.00% |

Table 8.3.12

| Count of Region | Percent of investment from savings | | | | Grand Total |
|---|---|---|---|---|---|
| | Less than 5% | 5% to 10% | 10% to 20% | More than 20% | |
| Mumbai | 0.00% | 3.57% | 21.43% | 14.29% | 39.29% |
| Mumbai Suburban | 3.57% | 10.71% | 7.14% | 7.14% | 28.57% |
| Thane | 0.00% | 7.14% | 0.00% | 17.86% | 25.00% |
| Others | 0.00% | 0.00% | 0.00% | 7.14% | 7.14% |
| Grand Total | 3.57% | 21.43% | 28.57% | 46.43% | 100.00% |

Table 8.3.13

| Occupation | Reason for investment | | | | Grand Total |
|---|---|---|---|---|---|
| | Child Education | Health Care | Marriage | Retirement | |
| Business | 0.00% | 0.00% | 3.57% | 0.00% | 3.57% |
| Employed (Goverment) | 0.00% | 7.14% | 17.86% | 21.43% | 46.43% |
| Employed (Private) | 3.57% | 10.71% | 3.57% | 28.57% | 46.43% |
| Professionals ( Teachers, Doctors, etc) | 0.00% | 0.00% | 0.00% | 3.57% | 3.57% |
| Grand Total | 3.57% | 17.86% | 25.00% | 53.57% | 100.00% |

Table 8.3.14

| Insurance Risk | Insurance Returns | | | | Grand Total |
|---|---|---|---|---|---|
| | Low Returns | Medium Returns | High Returns | Not Invested | |
| Low Risk | 7.14% | 28.57% | 3.57% | 0.00% | 39.29% |
| Medium Risk | 0.00% | 7.14% | 7.14% | 0.00% | 14.29% |
| High Risk | 0.00% | 3.57% | 10.71% | 0.00% | 14.29% |
| Not Invested | 0.00% | 0.00% | 0.00% | 32.14% | 32.14% |
| Grand Total | 7.14% | 39.29% | 21.43% | 32.14% | 100.00% |

Fig 8.3.15

# [8.4.b] Conclusion:

From table no. 8.3.10, we can see that investors who have taken any kind of risks in stocks expects High and Medium Returns. No investor expects low returns. 75% don't invest in Stocks

From table no. 8.3.11, we can see that if no of earning members is less 14.29% of investors save 25% to 50%, If no of earning members increases savings also increases. There are very few investors who have 4 to 5 earning members and their savings is more than 10

From the table no 8.3.12, we can see that Qualification have more influence on Awareness level of Mutual Funds. The investors with qualification less than H.S.C are not at all aware as compared to higher qualification

From table no. 8.3.13, we can see that investors from above mentioned regions give more preference to invest more than 20% from their monthly savings. Following most of investors in Mumbai Suburban invest 5% to 10% from their monthly savings. Also few investors from Mumbai and Mumbai Suburban region prefer to invest in more than 10% from their monthly savings

From table no. 8.3.14, we can see that most of investors reason for investment was their retirement. Also Private job employee invest his money for all reasons whereas Government job employee gets money for his child education from government

From the table no. 8.3.15, we observe that investors taking any risk expects medium and high returns. And 32.14% of investors do not invest at all.

# [9.1] Objective (6):

**To know the awareness of investment and problems faced by investors.**

# [9.2] Techniques:

## [9.2.a] Friedman ANOVA test:

The Friedman test is a non-parametric statistical test developed by Milton Friedman. Similar to the parametric repeated measures ANOVA, it is used to detect differences in treatments across multiple test attempts. The procedure involves ranking each row (or block) together, then considering the values of ranks by columns. Applicable to complete block designs, it is thus a special case of the Durbin test. When you choose to analyze data using a Friedman test, part of the process involves checking to make sure that the data you want to analyze can actually be analyzed using a Friedman test. You need to do this because it is only appropriate to use a Friedman test if your data "passes" the following four assumptions:

**Assumption #1:** One group that is measured on three or more different occasions.
**Assumption #2:** Group is a random sample from the population.
**Assumption #3:** Your dependent variable should be measured at the ordinal or continuous level. Examples of ordinal variables include Likert scales
**Assumption #4:** Samples do NOT need to be normally distributed.

The Friedman test procedure in SPSS Statistics will not test any of the assumptions that are required for this test. In most cases, this is because the assumptions are a methodological or study design issue, and not what SPSS Statistics is designed for. In the case of assessing the types of variable you are using, SPSS Statistics will not provide you with any errors if you incorrectly label your variables as nominal.

# [9.3] Analysis:

## Hypothesis:

**$H_0$: There is no statistically significant difference between awareness of different types of avenues.**
**$H_1$: there is statistically significant difference between awareness of different types of avenues.**

1. **NPar Test**

| Descriptive Statistics | | | | |
|---|---|---|---|---|
| | **N** | **Percentiles** | | |
| | | **25th** | **50th** **(Median)** | **75th** |
| **Stocks** | 594 | 1.00 | 2.00 | 2.00 |
| **Mutual Funds** | 594 | 1.00 | 2.00 | 3.00 |
| **Real Estate** | 594 | 1.00 | 2.00 | 2.25 |

| | | | | |
|---|---|---|---|---|
| **Metal (Gold/Silver/Others)** | **594** | **2.00** | **2.00** | **3.00** |
| **Insurance** | **594** | **2.00** | **3.00** | **3.00** |
| **Banks(FD, RD, Others)** | **594** | **2.00** | **3.00** | **3.00** |
| **Post Office** | **594** | **2.00** | **2.00** | **3.00** |
| **PPF** | **594** | **1.00** | **2.00** | **3.00** |
| **NPS** | **594** | **1.00** | **2.00** | **2.00** |

## 2. Friedman Test

**Ranks**

| | Mean Rank |
|---|---|
| Stocks | 3.95 |
| Mutual Funds | 4.66 |
| Real Estate | 4.45 |
| Metal (Gold/Silver/Others) | 5.23 |
| Insurance | 6.19 |
| Banks(FD, RD, Others) | 6.50 |
| Post Office | 5.19 |
| Public Provident Funds (PPF) | 4.78 |
| National Pension Scheme (NPS) | 4.06 |

**Test Statistics[a]**

| | |
|---|---|
| N | 594 |
| Chi-Square | 760.284 |
| df | 8 |
| Asymp. Sig. | .000 |

a. Friedman Test

# [9.4.a] Conclusion:

**Here p value is 0.000 which is less than 0.05(5 % level of significance) thus we reject null hypothesis that is we conclude that there is statistically significant difference between awareness of different types of avenues**

# [9.2.b] Post hoc test:

Post hoc tests are an integral part of ANOVA. When you use ANOVA to test the equality of at least three group means, statistically significant results indicate that not all of the group means are equal. However, ANOVA results do not identify which particular differences between pairs of means are significant. Use post hoc tests to explore differences between multiple group means while controlling the experiment-wise error rate.

To examine where the differences actually occur, you need to run separate Wilcoxon signed-rank tests on the different combinations of related groups

# [9.3.b] Analysis:

## Hypothesis:

**H$_0$: There is no statistical significant difference between all pairs of investment avenues**

**H$_1$: There is statistical significant difference between all pairs of investment avenues**

1. **NPar Test**

| Descriptive Statistics | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **N** | **Mean** | **Std. Deviation** | **Minimum** | **Maximum** | **Percentiles** | | |
| | | | | | | **25th** | **50th (Median)** | **75th** |
| **Stocks** | 594 | 1.76 | .720 | 1 | 3 | 1.00 | 2.00 | 2.00 |
| **Mutual Funds** | 594 | 1.98 | .746 | 1 | 3 | 1.00 | 2.00 | 3.00 |
| **Real Estate** | 594 | 1.91 | .760 | 1 | 3 | 1.00 | 2.00 | 2.25 |
| **Metal (Gold/Silver/Others)** | 594 | 2.14 | .731 | 1 | 3 | 2.00 | 2.00 | 3.00 |
| **Insurance** | 594 | 2.43 | .667 | 1 | 3 | 2.00 | 3.00 | 3.00 |
| **Banks(FD, RD, Others)** | 594 | 2.51 | .634 | 1 | 3 | 2.00 | 3.00 | 3.00 |
| **Post Office** | 594 | 2.15 | .758 | 1 | 3 | 2.00 | 2.00 | 3.00 |
| **Public Provident Funds (PPF)** | 594 | 2.01 | .806 | 1 | 3 | 1.00 | 2.00 | 3.00 |
| **National Pension Scheme (NPS)** | 594 | 1.80 | .792 | 1 | 3 | 1.00 | 2.00 | 2.00 |

## Wilcoxon Signed Ranks Test

| Ranks | | N | Mean Rank | Sum of Ranks |
|---|---|---|---|---|
| Mutual Funds - Stocks | Negative Ranks | 46[a] | 93.21 | 4287.50 |
| | Positive Ranks | 157[b] | 104.58 | 16418.50 |
| | Ties | 391[c] | | |
| | Total | 594 | | |
| Real Estate - Stocks | Negative Ranks | 120[d] | 147.18 | 17662.00 |
| | Positive Ranks | 188[e] | 159.17 | 29924.00 |
| | Ties | 286[f] | | |
| | Total | 594 | | |
| Metal (Gold/Silver/Others) - Stocks | Negative Ranks | 85[g] | 148.46 | 12619.00 |
| | Positive Ranks | 254[h] | 177.21 | 45011.00 |
| | Ties | 255[i] | | |
| | Total | 594 | | |
| Insurance - Stocks | Negative Ranks | 31[j] | 125.00 | 3875.00 |
| | Positive Ranks | 323[k] | 182.54 | 58960.00 |
| | Ties | 240[l] | | |
| | Total | 594 | | |
| Banks(FD, RD, Others) - Stocks | Negative Ranks | 23[m] | 147.39 | 3390.00 |
| | Positive Ranks | 354[n] | 191.70 | 67863.00 |
| | Ties | 217[o] | | |
| | Total | 594 | | |
| Post Office - Stocks | Negative Ranks | 85[p] | 146.76 | 12475.00 |
| | Positive Ranks | 251[q] | 175.86 | 44141.00 |
| | Ties | 258[r] | | |
| | Total | 594 | | |
| Public Provident Funds (PPF) - Stocks | Negative Ranks | 97[s] | 138.74 | 13457.50 |
| | Positive Ranks | 205[t] | 157.54 | 32295.50 |

| | | | | |
|---|---|---|---|---|
| | Ties | 292[u] | | |
| | Total | 594 | | |
| National Pension Scheme (NPS) - Stocks | Negative Ranks | 128[v] | 123.88 | 15857.00 |
| | Positive Ranks | 135[w] | 139.70 | 18859.00 |
| | Ties | 331[x] | | |
| | Total | 594 | | |
| Real Estate - Mutual Funds | Negative Ranks | 166[y] | 158.26 | 26271.00 |
| | Positive Ranks | 140[z] | 147.86 | 20700.00 |
| | Ties | 288[aa] | | |
| | Total | 594 | | |
| Metal (Gold/Silver/Others) - Mutual Funds | Negative Ranks | 135[ab] | 173.51 | 23424.00 |
| | Positive Ranks | 213[ac] | 175.13 | 37302.00 |
| | Ties | 246[ad] | | |
| | Total | 594 | | |
| Insurance - Mutual Funds | Negative Ranks | 53[ae] | 143.61 | 7611.50 |
| | Positive Ranks | 268[af] | 164.44 | 44069.50 |
| | Ties | 273[ag] | | |
| | Total | 594 | | |
| Banks(FD, RD, Others) - Mutual Funds | Negative Ranks | 39[ah] | 127.50 | 4972.50 |
| | Positive Ranks | 283[ai] | 166.19 | 47030.50 |
| | Ties | 272[aj] | | |
| | Total | 594 | | |
| Post Office - Mutual Funds | Negative Ranks | 117[ak] | 149.14 | 17449.50 |
| | Positive Ranks | 191[al] | 157.78 | 30136.50 |
| | Ties | 286[am] | | |
| | Total | 594 | | |
| Public Provident Funds (PPF) - Mutual Funds | Negative Ranks | 138[an] | 145.42 | 20068.00 |
| | Positive Ranks | 152[ao] | 145.57 | 22127.00 |
| | Ties | 304[ap] | | |
| | Total | 594 | | |
| | Negative Ranks | 188[aq] | 149.73 | 28149.00 |

| | | | | |
|---|---|---|---|---|
| National Pension Scheme (NPS) - Mutual Funds | Positive Ranks | 105[ar] | 142.11 | 14922.00 |
| | Ties | 301[as] | | |
| | Total | 594 | | |
| Metal (Gold/Silver/Others) - Real Estate | Negative Ranks | 71[at] | 120.18 | 8532.50 |
| | Positive Ranks | 180[au] | 128.30 | 23093.50 |
| | Ties | 343[av] | | |
| | Total | 594 | | |
| Insurance - Real Estate | Negative Ranks | 59[aw] | 133.46 | 7874.00 |
| | Positive Ranks | 279[ax] | 177.12 | 49417.00 |
| | Ties | 256[ay] | | |
| | Total | 594 | | |
| Banks(FD, RD, Others) - Real Estate | Negative Ranks | 42[az] | 152.00 | 6384.00 |
| | Positive Ranks | 315[ba] | 182.60 | 57519.00 |
| | Ties | 237[bb] | | |
| | Total | 594 | | |
| Post Office - Real Estate | Negative Ranks | 117[bc] | 152.46 | 17838.00 |
| | Positive Ranks | 214[bd] | 173.40 | 37108.00 |
| | Ties | 263[be] | | |
| | Total | 594 | | |
| Public Provident Funds (PPF) - Real Estate | Negative Ranks | 135[bf] | 161.67 | 21825.50 |
| | Positive Ranks | 184[bg] | 158.77 | 29214.50 |
| | Ties | 275[bh] | | |
| | Total | 594 | | |
| National Pension Scheme (NPS) - Real Estate | Negative Ranks | 169[bi] | 144.96 | 24498.50 |
| | Positive Ranks | 117[bj] | 141.39 | 16542.50 |
| | Ties | 308[bk] | | |
| | Total | 594 | | |
| Insurance - Metal (Gold/Silver/Others) | Negative Ranks | 76[bl] | 129.95 | 9876.00 |
| | Positive Ranks | 208[bm] | 147.09 | 30594.00 |
| | Ties | 310[bn] | | |
| | Total | 594 | | |

| | | | | |
|---|---|---|---|---|
| Banks(FD, RD, Others) - Metal (Gold/Silver/Others) | Negative Ranks | 61[bo] | 131.01 | 7991.50 |
| | Positive Ranks | 232[bp] | 151.20 | 35079.50 |
| | Ties | 301[bq] | | |
| | Total | 594 | | |
| Post Office - Metal (Gold/Silver/Others) | Negative Ranks | 160[br] | 159.00 | 25440.00 |
| | Positive Ranks | 160[bs] | 162.00 | 25920.00 |
| | Ties | 274[bt] | | |
| | Total | 594 | | |
| Public Provident Funds (PPF) - Metal (Gold/Silver/Others) | Negative Ranks | 202[bu] | 170.55 | 34450.50 |
| | Positive Ranks | 139[bv] | 171.66 | 23860.50 |
| | Ties | 253[bw] | | |
| | Total | 594 | | |
| National Pension Scheme (NPS) - Metal (Gold/Silver/Others) | Negative Ranks | 242[bx] | 171.09 | 41403.00 |
| | Positive Ranks | 90[by] | 154.17 | 13875.00 |
| | Ties | 262[bz] | | |
| | Total | 594 | | |
| Banks(FD, RD, Others) - Insurance | Negative Ranks | 68[ca] | 90.28 | 6139.00 |
| | Positive Ranks | 110[cb] | 89.02 | 9792.00 |
| | Ties | 416[cc] | | |
| | Total | 594 | | |
| Post Office - Insurance | Negative Ranks | 203[cd] | 135.56 | 27519.50 |
| | Positive Ranks | 65[ce] | 131.18 | 8526.50 |
| | Ties | 326[cf] | | |
| | Total | 594 | | |
| Public Provident Funds (PPF) - Insurance | Negative Ranks | 243[cg] | 157.88 | 38364.00 |
| | Positive Ranks | 59[ch] | 125.24 | 7389.00 |
| | Ties | 292[ci] | | |
| | Total | 594 | | |
| National Pension Scheme (NPS) - Insurance | Negative Ranks | 310[cj] | 179.53 | 55653.50 |
| | Positive Ranks | 37[ck] | 127.69 | 4724.50 |

61

| | | | | |
|---|---|---|---|---|
| | Ties | 247[cl] | | |
| | Total | 594 | | |
| Post Office - Banks(FD, RD, Others) | Negative Ranks | 218[cm] | 130.22 | 28389.00 |
| | Positive Ranks | 38[cn] | 118.61 | 4507.00 |
| | Ties | 338[co] | | |
| | Total | 594 | | |
| Public Provident Funds (PPF) - Banks(FD, RD, Others) | Negative Ranks | 271[cp] | 165.79 | 44928.50 |
| | Positive Ranks | 50[cq] | 135.05 | 6752.50 |
| | Ties | 273[cr] | | |
| | Total | 594 | | |
| National Pension Scheme (NPS) - Banks(FD, RD, Others) | Negative Ranks | 340[cs] | 184.28 | 62654.00 |
| | Positive Ranks | 24[ct] | 157.33 | 3776.00 |
| | Ties | 230[cu] | | |
| | Total | 594 | | |
| Public Provident Funds (PPF) - Post Office | Negative Ranks | 148[cv] | 125.07 | 18510.00 |
| | Positive Ranks | 91[cw] | 111.76 | 10170.00 |
| | Ties | 355[cx] | | |
| | Total | 594 | | |
| National Pension Scheme (NPS) - Post Office | Negative Ranks | 217[cy] | 143.29 | 31094.00 |
| | Positive Ranks | 63[cz] | 130.89 | 8246.00 |
| | Ties | 314[da] | | |
| | Total | 594 | | |
| National Pension Scheme (NPS) - Public Provident Funds (PPF) | Negative Ranks | 127[db] | 81.44 | 10342.50 |
| | Positive Ranks | 33[dc] | 76.89 | 2537.50 |
| | Ties | 434[dd] | | |
| | Total | 594 | | |

| Test Statistics[a] | | |
|---|---|---|
| | **Z** | **Asymp. Sig. (2-tailed)** |
| **Mutual Funds - Stocks** | -7.975[b] | .000 |
| **Real Estate - Stocks** | -4.253[b] | .000 |
| **Metal (Gold/Silver/Others) - Stocks** | -9.590[b] | .000 |
| **Insurance - Stocks** | -15.011[b] | .000 |
| **Banks(FD, RD, Others) - Stocks** | -15.959[b] | .000 |
| **Post Office - Stocks** | -9.419[b] | .000 |
| **Public Provident Funds (PPF) - Stocks** | -6.616[b] | .000 |
| **National Pension Scheme (NPS) - Stocks** | -1.313[b] | .189 |
| **Real Estate - Mutual Funds** | -1.938[c] | .053 |
| **Metal (Gold/Silver/Others) - Mutual Funds** | -3.962[b] | .000 |
| **Insurance - Mutual Funds** | -11.791[b] | .000 |
| **Banks(FD, RD, Others) - Mutual Funds** | -13.443[b] | .000 |
| **Post Office - Mutual Funds** | -4.331[b] | .000 |
| **Public Provident Funds (PPF) - Mutual Funds** | -.771[b] | .441 |
| **National Pension Scheme (NPS) - Mutual Funds** | -4.877[c] | .000 |
| **Metal (Gold/Silver/Others) - Real Estate** | -6.799[b] | .000 |
| **Insurance - Real Estate** | -12.165[b] | .000 |
| **Banks(FD, RD, Others) - Real Estate** | -13.891[b] | .000 |
| **Post Office - Real Estate** | -5.876[b] | .000 |
| **Public Provident Funds (PPF) - Real Estate** | -2.378[b] | .017 |
| **National Pension Scheme (NPS) - Real Estate** | -3.043[c] | .002 |
| **Insurance - Metal (Gold/Silver/Others)** | -8.027[b] | .000 |

| | | |
|---|---|---|
| Banks(FD, RD, Others) - Metal (Gold/Silver/Others) | -10.023[b] | .000 |
| Post Office - Metal (Gold/Silver/Others) | -.155[b] | .877 |
| Public Provident Funds (PPF) - Metal (Gold/Silver/Others) | -3.076[c] | .002 |
| National Pension Scheme (NPS) - Metal (Gold/Silver/Others) | -8.347[c] | .000 |
| Banks(FD, RD, Others) - Insurance | -2.908[b] | .004 |
| Post Office - Insurance | -8.025[c] | .000 |
| Public Provident Funds (PPF) - Insurance | -10.821[c] | .000 |
| National Pension Scheme (NPS) - Insurance | -14.306[c] | .000 |
| Post Office - Banks(FD, RD, Others) | -10.923[c] | .000 |
| Public Provident Funds (PPF) - Banks(FD, RD, Others) | -12.122[c] | .000 |
| National Pension Scheme (NPS) - Banks(FD, RD, Others) | -15.368[c] | .000 |
| Public Provident Funds (PPF) - Post Office | -4.176[c] | .000 |
| National Pension Scheme (NPS) - Post Office | -8.879[c] | .000 |
| National Pension Scheme (NPS) - Public Provident Funds (PPF) | -6.991[c] | .000 |
| a. Wilcoxon Signed Ranks Test | | |
| b. Based on negative ranks. | | |
| c. Based on positive ranks. | | |

# [9.4.b] Conclusion:

**Here we conclude that pairwise awareness between different type of avenues , thus the following results are as follows**

1. **the p values of pairs [Mutual Funds – Stocks, Real Estate – Stocks, Metal (Gold/Silver/Others) – Stocks, Insurance – Stocks, Banks(FD, RD, Others) – Stocks, Post Office – Stocks, Public Provident Funds (PPF) – Stocks, Metal (Gold/Silver/Others) - Mutual Funds, Insurance - Mutual Funds, Banks(FD, RD, Others) - Mutual Funds, Post Office - Mutual Funds, National Pension Scheme (NPS) - Mutual Funds, Metal (Gold/Silver/Others) - Real Estate, Insurance - Real Estate, Banks(FD, RD, Others) - Real Estate, Post Office - Real Estate, Public Provident Funds (PPF) - Real Estate, National Pension Scheme (NPS) - Real Estate, Insurance - Metal (Gold/Silver/Others, Banks(FD, RD, Others) - Metal (Gold/Silver/Others), Public Provident Funds (PPF) - Metal (Gold/Silver/Others), National Pension Scheme (NPS) - Metal (Gold/Silver/Others, Banks(FD, RD, Others) – Insurance, Post Office – Insurance, Public Provident Funds (PPF) – Insurance, National Pension Scheme (NPS) – Insurance, Post Office - Banks(FD, RD, Others), Public Provident Funds (PPF) - Banks(FD, RD, Others), National Pension Scheme (NPS) - Banks(FD, RD, Others), Public Provident Funds (PPF) - Post Office, National Pension Scheme (NPS) - Post Office, National Pension Scheme (NPS) - Public Provident Funds (PPF) ] is 0.000 which is less than 0.05(5% level of significance) so we conclude that there is significant difference between awareness of these pair of avenues**

2. **The p values for (National Pension Scheme (NPS) – stocks) pair is 0.189 greater than 0.05 so we accept null hypothesis hence we conclude that there is no significant difference between awareness of this pair**

3. **The p values for (Real Estate - Mutual Funds) pair is 0.053 greater than 0.05 so we accept null hypothesis hence we conclude that there is no significant difference between awareness of this pair**

4. **The p values for (Public Provident Funds (PPF) - Mutual Funds) pair is 0.441 greater than 0.05 so we accept null hypothesis hence we conclude that there is no significant difference between awareness of this pair**

5. **The p values for (Post Office - Metal (Gold/Silver/Others)) pair is 0.877 greater than 0.05 so we accept null hypothesis hence we conclude that there is no significant difference between awareness of this pair**
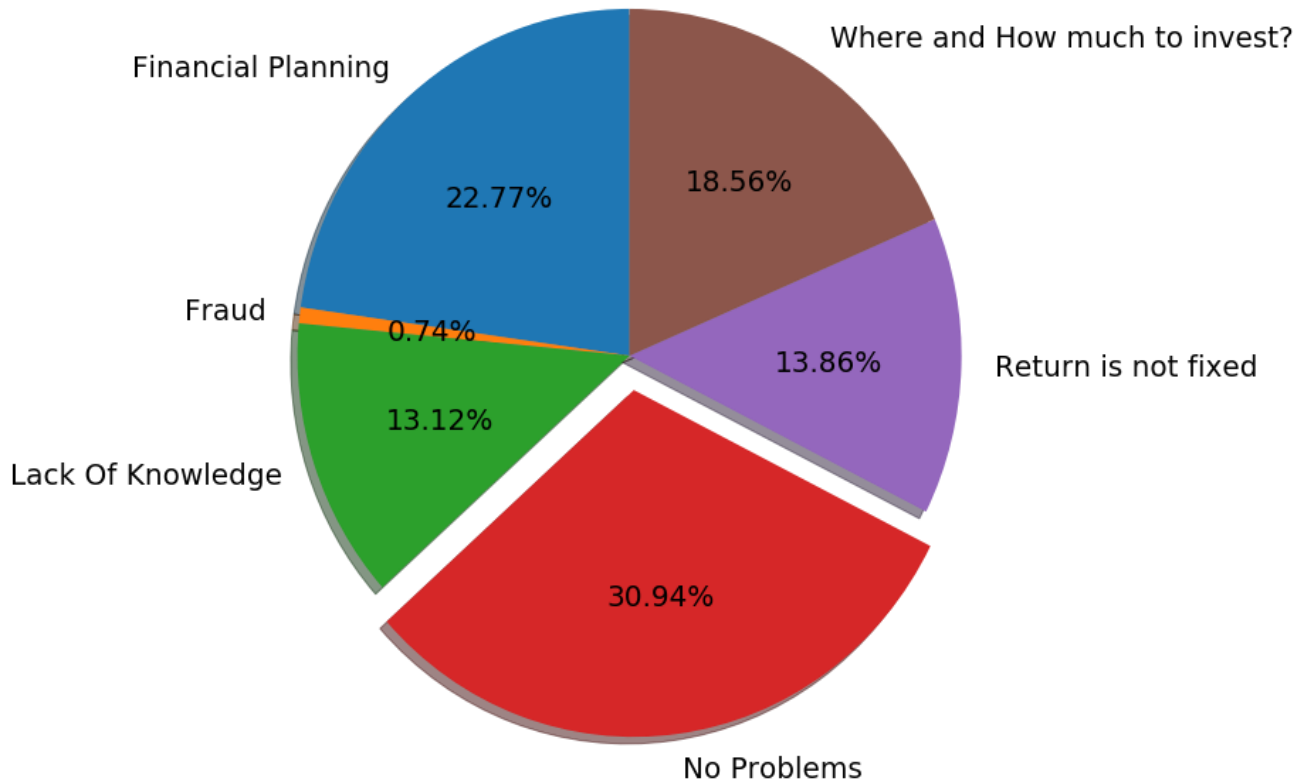
# [9.3.c] Graphical Representation:



Fig 9.3.1

**From the above pie chart, we can see that majority investors of 30.94% does not have any problem as compared to other investors whereas only 0.74% have faced a problem of Fraud which is least among all the investor.**

# [10.1] Objective (7):

To understand the reactive behaviour of investors during uncertainty.

# [10.2] Analysis:

# [10.2.a] Graphical Representation:

| Options | Count | Percentage |
|---|---|---|
| Already have saving. | 62 | 15.35% |
| Have a backup plan. | 73 | 18.07% |
| Have Invested wisely with proper information. | 75 | 18.56% |
| Invested in different option. | 69 | 17.08% |
| No Precautions. | 125 | 30.94% |
| Grand Total | 404 | 100.00% |
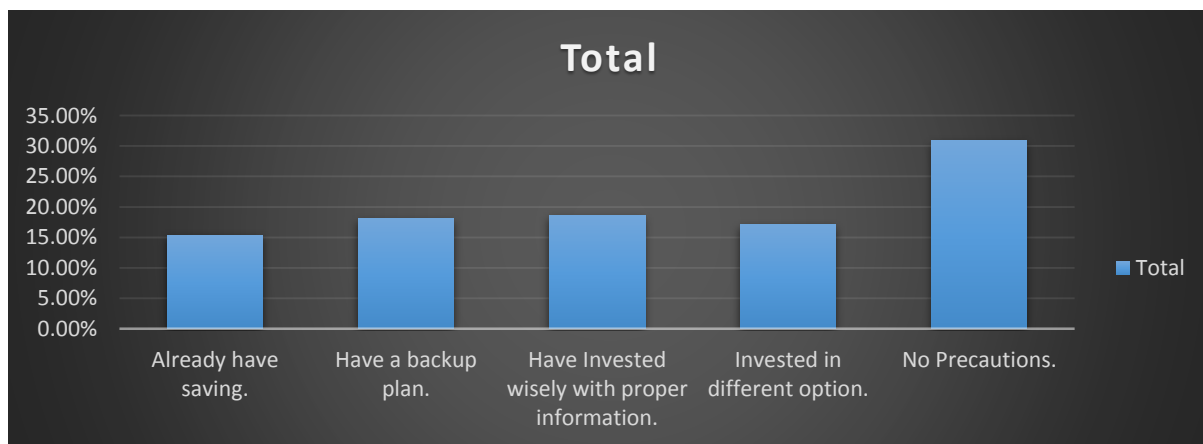


Fig 10.2.1

From fig 10.2.1 we can see that about **69.06%** of investors have a solution towards any kind of uncertainty that arises and they can be said as proactive investors since they have taken precautions

| Options | Count | Percentage |
| --- | --- | --- |
| Accept the situation. | 78 | 19.31% |
| I Will Stop my investment. | 49 | 12.13% |
| Legal Help. | 108 | 26.73% |
| Switch the investment option. | 73 | 18.07% |
| Wait till situation get normal. | 96 | 23.76% |
| Grand Total | 404 | 100 % |
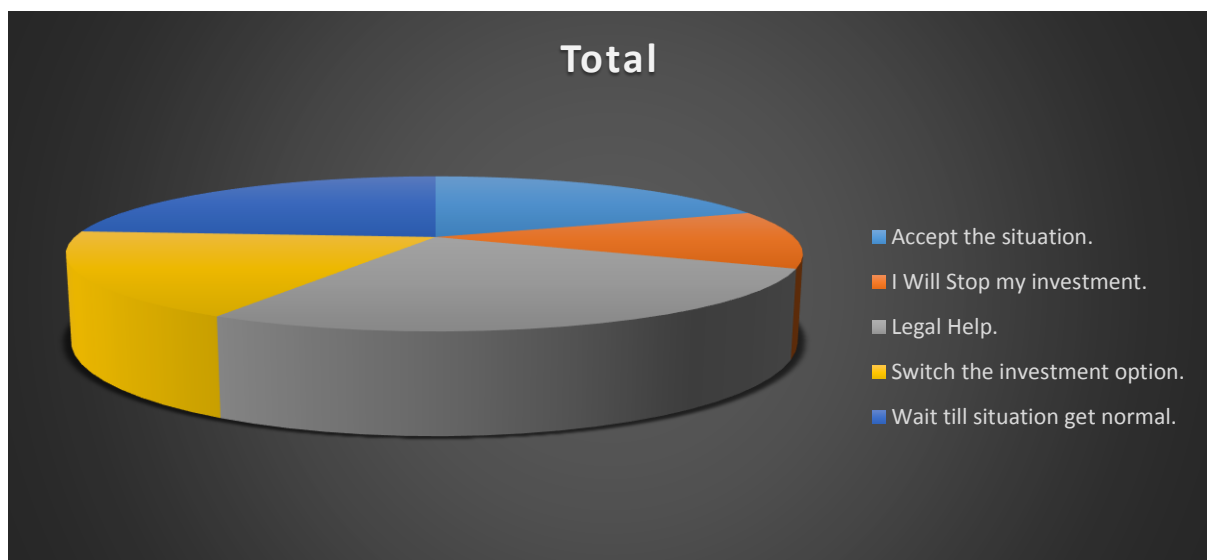


Fig 10.3.2

**From fig 10.3.2 we can say that about 56.93% of investors have reactive behavior towards any uncertainty situation.**

# [11] Findings and Conclusions:

In this project we have did many analyses on the data and result obtained was very good and satisfying. First we have performed operations on the demographic data to see the effect of demographics in our respondents. And we have found that majority of our respondents were male and out of the total respondents 68% people invest in some avenues and 32% don't invest anywhere. i.e. 1/4th people don't invest actively because of reasons like Lack of Knowledge, Fraud etc. problems. Also we did chi square analysis to check association between two demographics variable in investment and we found that the variables in pairs (Age, Investment Decision), (Gender, Investment Decision), (Annual Income, Investment Decision), (Qualification, Investment Decision) and (Region, Investment Decision) are associated with each other in investment.

Then we performed Logistic regression on our data taking Investment Decision as Dependent Variable then we got that only variables Age, Occupation, Family Size & Annual Income are significant variable for the model. Also we checked the pattern of salaried people from different sectors in investment and we found that most of the having working as private employee do most of investment.

As we all know that saving is an important factor for one's investment. So, from our data also we found that saving and investment are positively correlated with each other. i.e. if saving will increase then investment will also increase. Also we have found that nowadays as Investment became an important part of life as many says that only earning is not important so many people of different age groups invest their money, even in our data we have found that many students also invest a small amount of money for their need.

Then to understand the investor behavior on the basis of demographic factors, awareness & risk – return in their avenues we used principle component analysis and got that there were only 25 components which explain 85% of variation in data also did clustering and found 8 clusters from which 4th cluster contains highest proportion of investors and 5th cluster contains minimum proportion of investors then we analyze only those cluster to understand investors behavior. And we got that,1) Investors who have taken any kind of risks in stocks expects High and Medium Returns. No investors expect low returns. 82.61% don't invest in Stocks. 2)If no of earning members is less the percent of savings is less, also if no of earning members increases percent of savings is up to 50% utmost Investors from above mentioned regions give more preference to invest less than 5% from their monthly savings. Following most of investors in Mumbai Suburban invest 5% to 10% from their monthly savings. Also few investors from Mumbai and Thane region prefer to invest in more than 10% from their monthly savings. 3) Most of investors reason for investment was their child's education. And few investors with occupation Business, earning housewife, employed (Government as well as Private) had assets purchasing as their reason for investment followed by health care. 4) 15.22% of investors takes medium risk in Insurance and expects medium returns, followed by 11.96 % of investors have taken low risk and expects medium returns. And the majority of 56.52% of investors have not at all invested in Insurance.

Also we found that Awareness level of different investment avenues differ significantly. Awareness level is totally depending upon the Qualification. From our data we have found that people having more qualification have much knowledge about the investment avenues. And we also studied the reactive behaviour of investors i.e. if some types of uncertainty arise how they will react? And we got result as about 69.06% of investors have a solution towards any kind of uncertainty that arises and they can be said as a proactive investor since they have taken precautions.

# [12] Coding:

## R code for Chi Square Analysis:

```
data <- read.csv("C:/Users/pc/Desktop/data.csv")
View(data)
summary(data)
x1=table(data$Age..completed.,data$Gender)
x1
chisq.test(x1)
x2=table(data$Age..completed.,data$Annual.Income..Personal.)
x2
chisq.test(x2)
x3=table(data$Age..completed.,data$Do.you.Invest.)
x3
chisq.test(x3)
x4=table(data$Gender,data$Do.you.Invest.)
x4
chisq.test(x4)
x5=table(data$Gender,data$Annual.Income..Personal.)
x5
chisq.test(x5)
x6=table(data$Annual.Income..Personal.,data$Do.you.Invest.)
x6
chisq.test(x6) x7=table(data$Qualification.also.currently.pursuing.,data
$Do.you.Invest.)
x7
chisq.test(x7)
x8=table(data$Do.you.Invest.,data$Region)
x8
chisq.test(x8)
```

## R code for Correlation Analysis:

```
#saving & investment objective
#import data
data=read.csv("D:/Analysis_12.csv",header=TRUE)
data
#extract rows and colunms
data1=data[2:404,20:21]
data1
#spearman rank correlation test (ties problem)
rank_cor=cor.test(data1$saving,data1$investment,method="spearman")
rank_cor
rank_cor$p.value
rank_cor$estimate
#spearman rank correlation test with ties
data1.ranked <- data.frame(cbind(rank(data1$saving, ties.method = 'avera
ge'),rank(data1$investment, ties.method = 'average')))
colnames(data1.ranked) <- c('saving', 'investment')
rho <- cov(data1.ranked) / (sd(data1.ranked$saving) * sd(data1.ranked$in
vestment))
rho[[2]]
#kendall rank correlation test(ties problem)
rank_cor=cor.test(data1$saving,data1$investment,method="kendall")
rank_cor
rank_cor$p.value
rank_cor$estimate
#kendall rank correlation test with ties
```

```
#We use the formula to compute tau but first we have to compute Vi
#Kendall's Tau for Tie
KenTauTie <- function(X, Y){
+    #U and Vi function for tie
+    #Here we use a small trick to achieve our goal  #We do not use separ
ate if else
+ #because here we have to do vector computation.
+    U <- function(x, y, i, j){
+      U1 <- ifelse((x[i] - x[j])*(y[i] - y[j]) == 0, 0.5, 0)
+      U2 <- ifelse((x[i] - x[j])*(y[i] - y[j]) > 0, 1, 0)
+      U1 + U2
+    }
+    Vi <- function(x, y){
+      n <- length(y) - 1
+      V <- rep(0, n)
+      for(i in 1:n){
+        V[i] <- sum(U(x, y, i, (i + 1):(n + 1)))
+      }
+      V
+    }
+    (tau <- 2*(sum(Vi(X, Y)))/choose(length(X), 2) - 1)
+ }
(tau <- KenTauTie(data1$saving,data1$investment))
```

## R code for Logistic Regression, Odds Ratio & ROC curve

```
data = read.csv("C://Users//sidh//Desktop//LR.csv")
data = data[,-1]
data2 = data
library(caTools)
set.seed(123)
split = sample.split(data2$Do.you.Invest., SplitRatio = 0.70)
training_set = subset(data2, split == TRUE)
test_set = subset(data2, split == FALSE)

classifier1 = glm(formula = Do.you.Invest. ~ .,
          family = binomial,
          data = training_set)
summary(classifier1)
prob_pred1 = predict(classifier1, type = 'response', newdata = test_set[-9])
y_pred1 = ifelse(prob_pred1 > 0.5, 1, 0)
cm1 = table(test_set[,9], y_pred1 > 0.5)
cm1

AIC(classifier1)
BIC(classifier1)

data2$Gender = NULL
data2$Qualification = NULL
data2$No.of.earning.members = NULL
data2$Region = NULL
split2 = sample.split(data2$Do.you.Invest., SplitRatio = 0.70)
training_set2 = subset(data2, split2 == TRUE)
test_set2 = subset(data2, split2 == FALSE)

classifier2 = glm(formula = Do.you.Invest. ~ .,
```

71

```
            family = binomial,
            data = training_set2)
summary(classifier2)
prob_pred2 = predict(classifier2, type = 'response', newdata = test_set2[-5])
y_pred2 = ifelse(prob_pred2 > 0.5, 1, 0)
cm2 = table(test_set2[,5], y_pred2 > 0.5)
cm2
AIC(classifier2)
BIC(classifier2)
summary(cm2)
paste("Accuracy of model is :" , 136/178)
library(epitools)
table1 = table(data$Age,data$Do.you.Invest.)
table2 = table(data$Occupation,data$Do.you.Invest.)
table3 = table(data$Annual.Income,data$Do.you.Invest.)
table4 = table(data$Family.Size,data$Do.you.Invest.)
or_age = oddsratio(table1)
or_occ = oddsratio(table2)
or_annual = oddsratio(table3)
or_fam = oddsratio(table4)
or_age
or_occ
or_annual
or_fam

ROC_lr <- roc(test_set2$Do.you.Invest., prob_pred2)
ROC_lr_auc <- auc(ROC_lr)
plot(ROC_lr, col = "red", main = "ROC for Logistic Regression (RED)")
ROC_lr_auc
```

## Python code for Principal Component Analysis & Clustering

```python
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.cluster import KMeans
from sklearn.preprocessing import LabelEncoder
import operator

data = pd.read_csv("C://Users//sidh//Desktop//RiskReturn.csv") #Data worked on
data2 = pd.read_csv("C://Users//sidh//Desktop//RiskReturn.csv") # Data assigned for clusters

#LabelEncoding

le=LabelEncoder()
data['Stocks Risk'] = le.fit_transform(data['Stocks Risk'])
data['Bank Risk'] = le.fit_transform(data['Bank Risk'])
data['MF Risk'] = le.fit_transform(data['MF Risk'])
data['RE Risk'] = le.fit_transform(data['RE Risk'])
data['Metal Risk'] = le.fit_transform(data['Metal Risk'])
data['Insu Risk'] = le.fit_transform(data['Insu Risk'])
data['PO Risk'] = le.fit_transform(data['PO Risk'])
data['PPF Risk'] = le.fit_transform(data['PPF Risk'])
data['NPS Risk'] = le.fit_transform(data['NPS Risk'])
```

```python
data['Stocks Returns'] = le.fit_transform(data['Stocks Returns'])
data['MF Returns'] = le.fit_transform(data['MF Returns'])
data['RE Returns'] = le.fit_transform(data['RE Returns'])
data['Metal Returns'] = le.fit_transform(data['Metal Returns'])
data['Insu Returns'] = le.fit_transform(data['Insu Returns'])
data['Bank Returns'] = le.fit_transform(data['Bank Returns'])
data['PO Returns'] = le.fit_transform(data['PO Returns'])
data['PPF Returns'] = le.fit_transform(data['PPF Returns'])
data['NPS Returns'] = le.fit_transform(data['NPS Returns'])
data['Age (completed)'] = le.fit_transform(data['Age (completed)'])
data['Gender'] = le.fit_transform(data['Gender'])
data['Qualification'] = le.fit_transform(data['Qualification'])
data['Occupation'] = le.fit_transform(data['Occupation'])
data['Region'] = le.fit_transform(data['Region'])
data['Annual Income'] = le.fit_transform(data['Annual Income'])
data['Awareness Banks'] = le.fit_transform(data['Awareness Banks'])
data['Awareness Insurance'] = le.fit_transform(data['Awareness Insurance'])
data['Awareness Metal'] = le.fit_transform(data['Awareness Metal'])
data['Awareness Mutual Funds'] = le.fit_transform(data['Awareness Mutual Funds'])
data['Awareness NPS'] = le.fit_transform(data['Awareness NPS'])
data['Awareness PO'] = le.fit_transform(data['Awareness PO'])
data['Awareness PPF'] = le.fit_transform(data['Awareness PPF'])
data['Awareness Real Estate'] = le.fit_transform(data['Awareness Real Estate'])
data['Awareness Stocks'] = le.fit_transform(data['Awareness Stocks'])
data['Reason for investment'] = le.fit_transform(data['Reason for investment'])
data['Savings'] = le.fit_transform(data['Savings'])
data['investing'] = le.fit_transform(data['investing'])

# Applying PCA
from sklearn.decomposition import PCA
Pca = PCA()
data_pca = Pca.fit_transform(data)

# Scree plot function
def scree_plot(pca):
num_components = len(pca.explained_variance_ratio_)
ind = np.arange(num_components)
vals = pca.explained_variance_ratio_
plt.figure(figsize=(10, 6))
ax = plt.subplot(111)
cumvals = np.cumsum(vals)
ax.bar(ind, vals)
ax.plot(ind, cumvals)
ax.xaxis.set_tick_params(width=0)
ax.yaxis.set_tick_params(width=2, length=12)
ax.set_xlabel("Principal Component")
ax.set_ylabel("Variance Explained (%)")
plt.title('Explained Variance Per Principal Component')

# Plotting Scree Plot
scree_plot(Pca)

# Re-apply PCA to the data while selecting for number of components to retain
Pca = PCA(n_components = 25)
data_pca = Pca.fit_transform(data)
data_pca
```

```python
# Alloting weights to PC
def weights(pca, i):
weight = {}
for counter, feature in enumerate(data.columns):
weight[feature] = pca.components_[i][counter]
sorted_weights = sorted(weight.items(), key=operator.itemgetter(1), reverse=True)
return sorted_weights

# Interpreting and investigating feature association from Principal Component
weights(Pca,1)

# Function for finding optimal number of centroids
def Kmeans_Fun(x,centroid):
kmeans=KMeans(n_clusters=centroid)
kmeans_model=kmeans.fit(x)
score=kmeans_model.score(x)
return score

# Finding number of Clusters
score=[]
centroid=[1,2,3,4,5,6,7,8,9,10,11,12]
for i in centroid:
score.append(Kmeans_Fun(data, i))

# Elbow Graph
plt.plot(centroid, np.abs(score), linestyle='--', marker='o', color='b');
plt.xlabel('K');
plt.ylabel('Score');
plt.title('Score vs. K');

# Applying KMeans
kmeans = KMeans(n_clusters=8)
new_kmeans_model = kmeans.fit(data)

assigned_clusters = new_kmeans_model.predict(data)

data2['clusters'] = assigned_clusters

general_proportion = []
x = [k+1 for k in range(8)]
for i in range(8):
general_proportion.append((assigned_clusters == i).sum()/len(assigned_clusters))
df_general = pd.DataFrame({'cluster' : x, 'No. of People' : general_proportion})
df_general.plot(x='cluster', y = ['No. of People'], kind='bar', figsize=(20,10))
plt.ylabel('proportion of persons in each cluster')
plt.show()

cluster4 = data2[data2['clusters'] == 3]

cluster5 = data2[data2['clusters'] == 4]

data2.to_csv("C://Users//sidh//Desktop//clusters1.csv")
```

74

# [13] Questionnaire:

**Q. Age**

- 18-25
- 26-35
- 36-45
- 46-55
- 56 and above

**Q. Gender**

- Male
- Female

**Q. Educational Qualification**

- Below S.S.C
- S.S.C
- H.S.C
- Diploma
- Graduation
- Post Graduation
- Doctorate
- Others...............

**Q. Occupation**

- Student & Employed
- Employed (Government)
- Employed (Private)
- Earning Housewife
- Business
- Professional (Teachers,Dr. etc)
- Retired

**Q. Family Size**................ **Q. Number of Earning Members**..........

**Q. Region**

- Mumbai
- Mumbai Suburban
- Thane
- Others

**Q. Annual Income (Personal)**

- Less than 5 lakh
- 5 lakh to 10 lakh
- 10 lakh to 15 lakh
- More than 15 lakh

**Q. Awareness level of Investment Avenues.**

| Avenue | Not at all aware | Slightly Aware | Extremely Aware |
|---|---|---|---|
| Stocks | | | |
| Mutual Funds | | | |
| Real Estate | | | |
| Metal (Gold/Silver/others) | | | |
| Insurance | | | |
| Banks (FD, RD, others) | | | |
| Post Office | | | |
| National Pension Scheme(NPS) | | | |
| Public Provident Fund (PPF) | | | |

**Q. Do you invest?**

- Yes
- No

**Q. . What percent of your monthly income do you save?**

- Less than 10%
- 10% to 25%
- 25% to 50%
- More than 50%

**Q. What percent of your savings do you invest?**

- Less than 5%
- 5% to 10%
- 10% to 20%
- More than 20%

**Q. Where do you invest? (You can select more than 1 option.)**

- Stocks
- Mutual Funds
- Real Estate

75

- **Metal (Gold/Silver etc..)**
- **Insurance**

- **Banks**
- **Post Office**

- **Public Provident Fund (PPF)**
- **National Pension Scheme (NPS)**

**Q. Factor Influencing on Investment Decision ((Rank according to your priority: 1 for Highest and 5 for lowest)**

| | |
|---|---|
| **Safety** | |
| **Returns** | |
| **Tax Saving** | |
| **Liquidity(quickly conversion ofasset to cash)** | |
| **Future Needs** | |

**Q. Reasons for Investment. (You can select more than 1 option.)**

- **Child Education**
- **Marriage**

- **Asset (Home,Car etc..)**
- **Health Care**

- **Retirement**
- **Others**................

**Q. What is the time period you prefer to invest?**

| Avenue | Short Term (less than 1 year) | Medium Term (1 to 3 years) | Long Term (More than 3 years) |
|---|---|---|---|
| **Stocks** | | | |
| **Mutual Funds** | | | |
| **Real Estate** | | | |
| **Metal (Gold/Silver/others)** | | | |
| **Insurance** | | | |
| **Banks (FD, RD, others)** | | | |
| **Post Office** | | | |
| **National Pension Scheme (NPS)** | | | |
| **Public Provident Fund (PPF)** | | | |

**Q. Whose Advice do you seek? (You can select more than 1 option.)**

- **Self**
- **Internet**
- **Family/Friend**

- **News channel/Papers**
- **Financial Advisor**
- **Others**................

**Q. . How frequently do you invest (your invested avenue only)?**

| Avenue | Daily | Weekly | Monthly | Quarterly | Half yearly | Yearly |
|---|---|---|---|---|---|---|
| **Stocks** | | | | | | |
| **Mutual Funds** | | | | | | |
| **Real Estate** | | | | | | |
| **Metal (Gold/Silver/others)** | | | | | | |
| **Insurance** | | | | | | |
| **Banks (FD, RD, others)** | | | | | | |

75

| Avenue | | | | | |
|---|---|---|---|---|---|
| **Post Office** | | | | | |
| **National Pension Scheme (NPS)** | | | | | |
| **Public Provident Fund (PPF)** | | | | | |

**Q. What amount of risk do you take in your invested avenue?**

| Avenue | Low Risk | Medium Risk | High Risk |
|---|---|---|---|
| **Stocks** | | | |
| **Mutual Funds** | | | |
| **Real Estate** | | | |
| **Metal (Gold/Silver/others)** | | | |
| **Insurance** | | | |
| **Banks (FD, RD, others)** | | | |
| **Post Office** | | | |
| **National Pension Scheme(NPS)** | | | |
| **Public Provident Fund (PPF)** | | | |

**Q. What amount of returns do you expect in your invested avenue?**

| Avenue | Low Returns | Medium Returns | High Returns |
|---|---|---|---|
| **Stocks** | | | |
| **Mutual Funds** | | | |
| **Real Estate** | | | |
| **Metal (Gold/Silver/others)** | | | |
| **Insurance** | | | |
| **Banks (FD, RD, others)** | | | |
| **Post Office** | | | |
| **National Pension Scheme (NPS)** | | | |
| **Public Provident Fund (PPF)** | | | |

**Q. Where had you invested previously? (You can select more than 1 option.)**

- **Stocks**
- **Mutual Funds**
- **Real Estate**
- **Metals (Gold/Silver etc..)**
- **Insurance**
- **Banks**
- **Post Offices**
- **Public Provident Fund (PPF)**
- **National Pension Scheme (NPS)**
- **None of the above**

**Q. If you want to change your investment, where will you invest? (You can select more than 1 option.)**

- **Stocks**
- **Mutual Funds**
- **Real Estate**
- **Metals (Gold/Silver etc..)**
- **Insurance**
- **Banks**
- **Post Offices**
- **Public Provident Fund (PPF)**
- **National Pension Scheme (NPS)**
- **Don't want to change**

**Q. How much are you satisfied with your investment?**

| Avenue | Not at all Satisfied | Slightly Satisfied | Moderately Satisfied | Very Satisfied | Completely Satisfied |
|---|---|---|---|---|---|
| **Stocks** | | | | | |
| **Mutual Funds** | | | | | |
| **Real Estate** | | | | | |
| **Metal (Gold/Silver/others)** | | | | | |
| **Insurance** | | | | | |
| **Banks (FD,RD,others)** | | | | | |

| | | | | | |
|---|---|---|---|---|---|
| **Post Office** | | | | | |
| **National Pension Scheme (NPS)** | | | | | |
| **Public Provident Fund (PPF)** | | | | | |

**Q. What are the problems faced by you during investment?**

- **Lack of Knowledge**
- **Where And How to invest**
- **Financial Planning**
- **Return is not fix**
- **No problem**
- **Others.............**

**Q. What precautions have you taken if uncertainty arises?**

- **Have invested wisely with proper information**
- **Invested in different options**
- **Already have savings**
- **No precautions**
- **Have a Back up plan**
- **Others.............**

**Q. If uncertainty arises; how will you react?**

- **Accept the situation**
- **Wait till situation get normal**
- **Legal Help**
- **Switch the investment**
- **I will stop investing**

# If not investing

**Q. From where do you get investment information?**

- **Family/Friends**
- **Internet**
- **News channel/Papers**
- **Financial Advisor**
- **None of the above**

**Q. Reason for not investing**

- **Lack of Knowledge**
- **Financial Issue**
- **Not Interested**
- **More risk**

-------------

# [14] Statistical Analysis Tools/Software used:

# [15] Problems and Challenges:

- In this type of personal survey, chances of respondent's bias are there. Chances of researcher's bias might have crept in during collection of data and while handling incomplete questionnaires.

- Many respondent refuses to give the real data about their income and investment option, therefore some error might have occurred due to incorrect information.

- The study was restricted to the employed people from the Mumbai, Mumbai Suburban and Thane regions. The findings of the study may not be generalized to the entire population of India.

- Also the obtained sample size was 594, from which we may say that it may not be a good representative of the whole population under study.

- The study is restricted to only the employed people. Enough care has been taken while processing, cleaning, editing and analyzing the data, to minimize the impact of these limitations on the findings of the study.

- The result presented in this report to solely making a projection that may or may not be accurate.

# [16] References:

[Websites]

- http://scholar.google.com/

- https://www.wikipedia.org/

- https://www.youtube.com/

- https://www.kaggle.com/

- https://in.linkedin.com/

[Books]

- STATISTICAL MODELS FOR CASUAL ANALYSIS, (Robert D. Rutherford and Minjs Kim Choe).

- DATA MINING , concepts and Techniques, (Jiawei Han , Micheline Kamber and Jian Pei).

- RESEARCH METHOLOGY, Methods and Techniques, (C.R. Kothari).

- BASICS ECONOMETRICS, (Damodar N. Gujarati).

[Research Papers]

- Bajtelsmit, V.L., & Bernasek, A. (1996). Why do women invest differently than men? Financial Counselling and planning, 7, 1-10.

- Parthiban, U. (2011). A Study on Investment Behaviour of Women Investors.
- Singh, J., & Yadav, P. (2016). A Study on the Factors Influencing Investors Decision in Investing in Equity Shares in Jaipur and Moradabad with Special Reference to Gender. Amity Journal of Finance, 1(1), 117-130.

- Dr. C. M. Shinde, and Priyanka Zanvar, "A STUDY OF INVESTMENT PATTERN ON THE BASIS OF DEMOGRAPHIC TRAITS" International Journal of Research – Granthaalayah, Vol. 3, No. 11(2015): 175-205.

- Giridhari Mohanta, Sathya Swaroop Debasish (2011) A Study on Investment Preferences among Urban Investors in Orissa. Prerna Journal of Management Thought and Practice, 3(1), 1-9

- Rajarajan (2002), "Determinants of Portfolio Choice of Individual Investors", Indian Economic Journal, 81-84.

- Suman and D. P. Warne (2012) Investment Behavior of Individual Investor in Stock Market., IJRFM , 2(2)

- Virani,V. 2012.Saving and Investment pattern of school teachers- A study With special reference to Rajkot City,Gujrat. Abhinav National Refereed journal of research in Commerce and Management. 2(4) :2277-1166.

[Newspaper]

- The Economic Times

80