Investor Behaviour and Market Segmentation Analysis

Presented by Vaibhav Pawar

2024

Level 1: Beginner

Task 1: Data Overview

Objective: Understand the dataset structure.

Steps:

1. **Load the dataset:** Import the dataset into a data analysis tool such as Python with pandas or spreadsheet software.
2. **Descriptive Statistics:** Use descriptive functions (e.g., info() in pandas) to gather information about the number of entries, columns, and data types.

Dataset

Data Type	Columns
Categorical (Text)	gender, Investment_Avenues, Stock_Marktet, Factor, Objective, Purpose, Duration, Invest_Monitor, Avenue, What are your savings objectives?, Reason_Equity, Reason_Mutual, Reason_Bonds, Reason_FD, Source
Numeric (Integer)	age, Mutual_Funds, Equity_Market, Debentures, Government_Bonds, Fixed_Deposits, PPF, Gold , Expect.

Dataset Summary

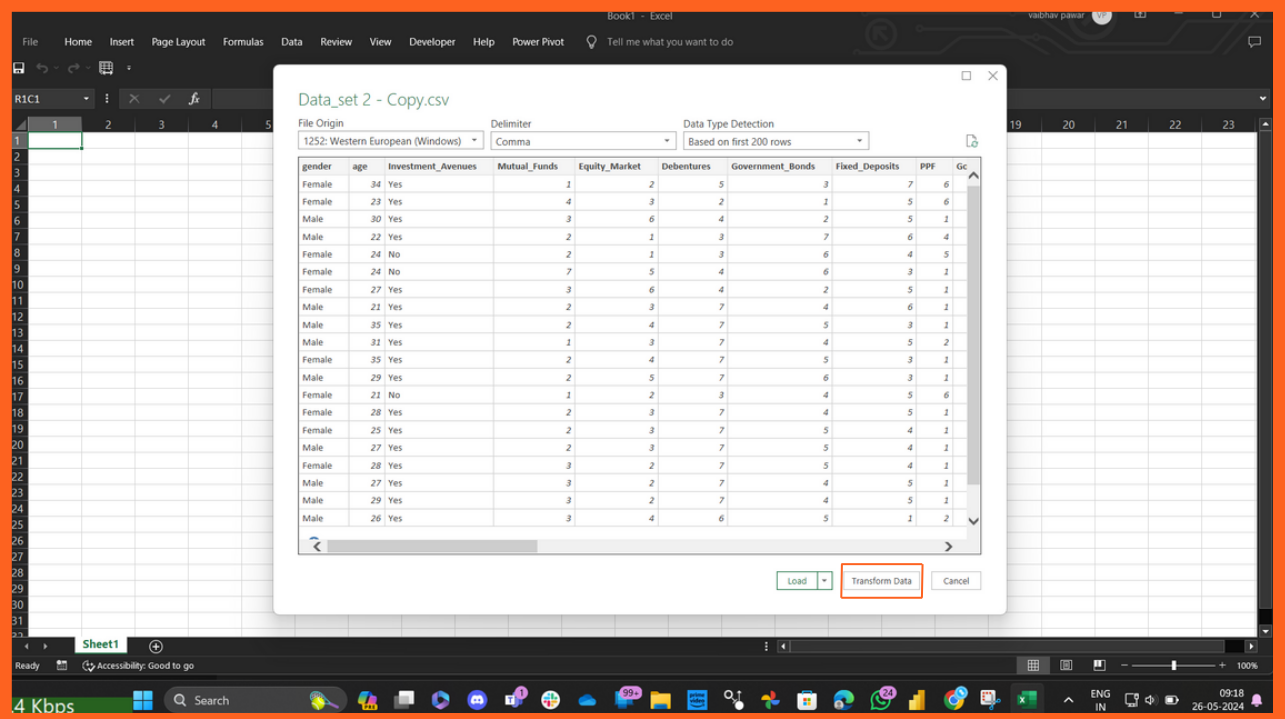
Total Entries : 40 rows
Total Columns : 24 columns

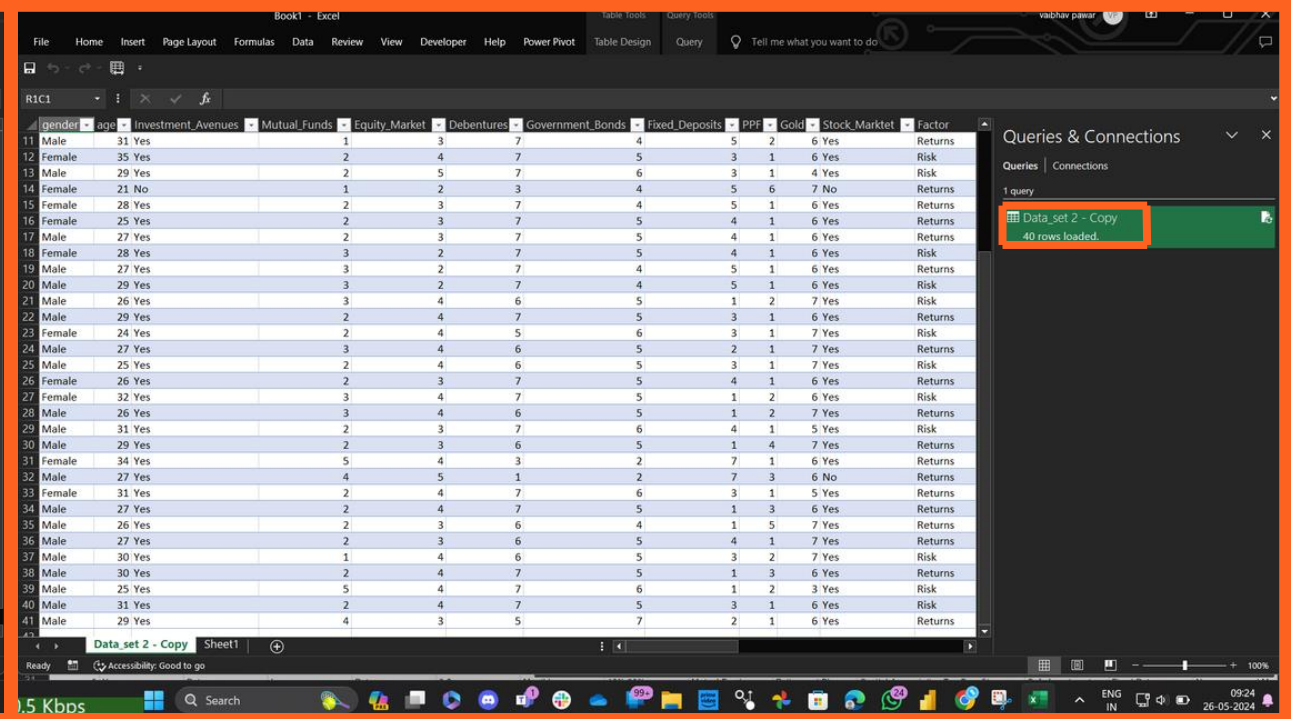
Key Statistical Measures for Selected Columns

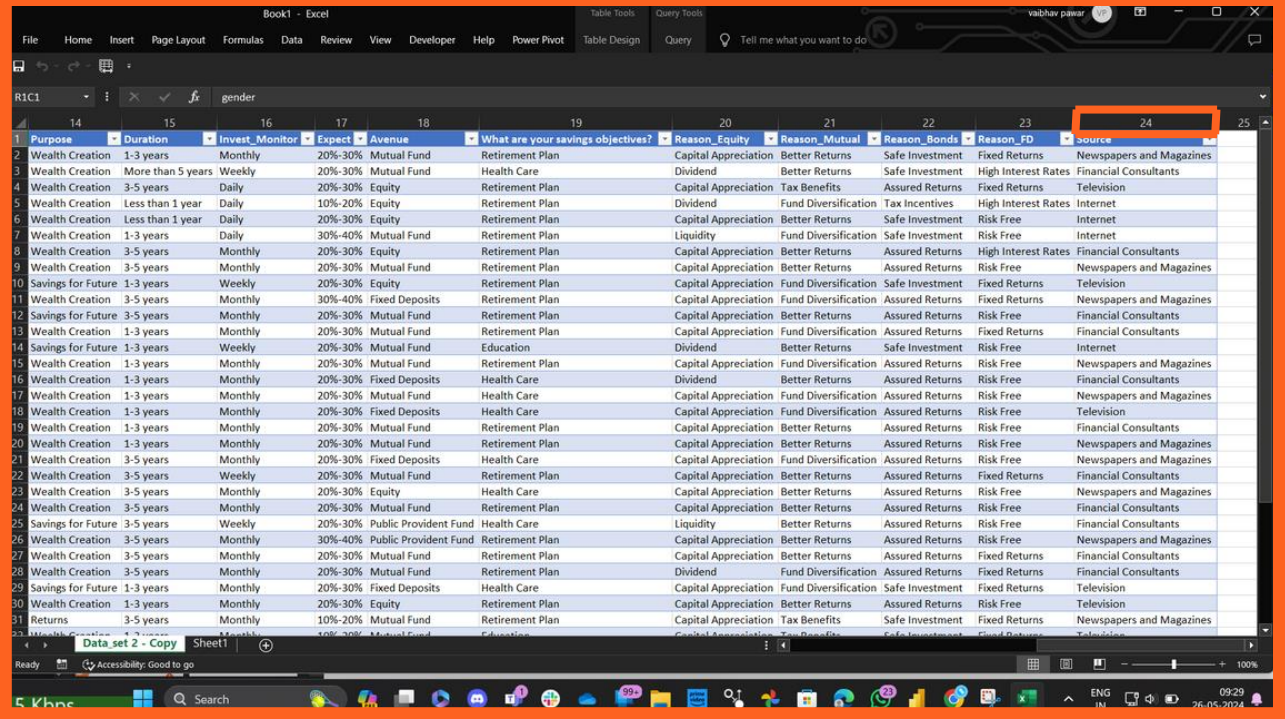
Age
- Mean : 27.8 years
- Median : 27 years
- Mode : 27 years
- Range : 21 to 35 years

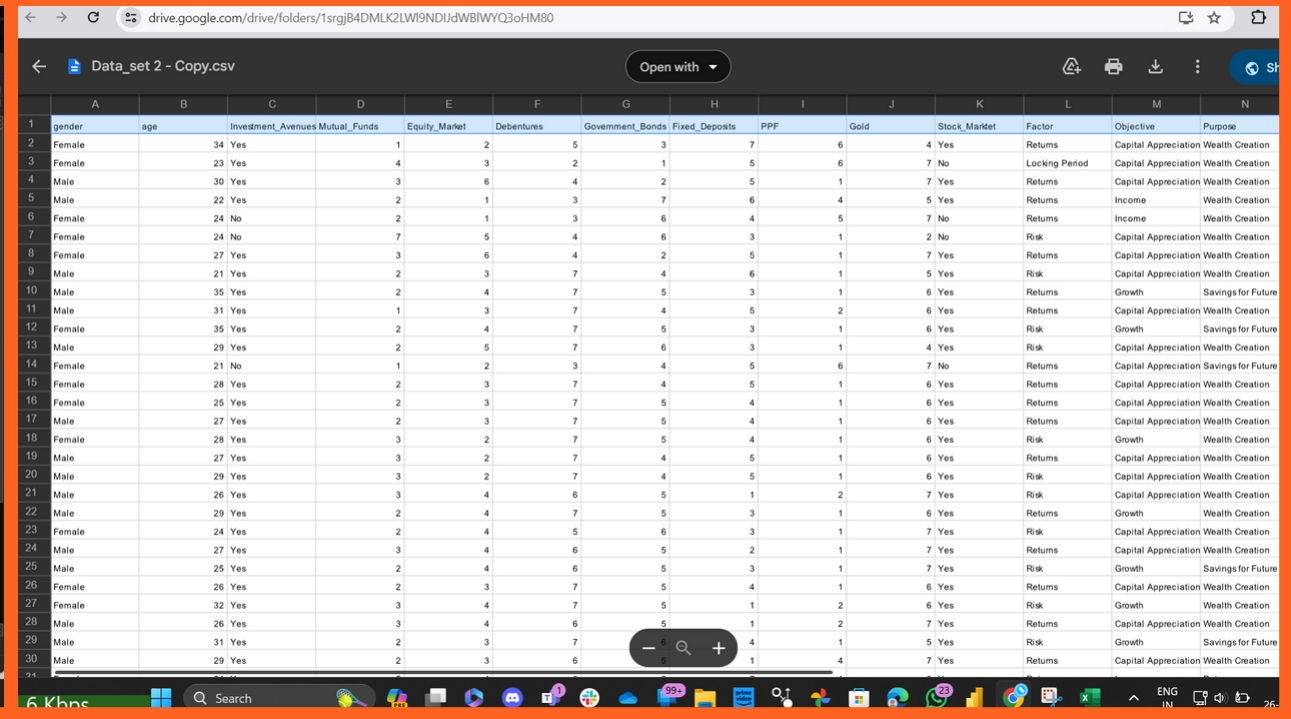
Investment Avenues (Average Investment in Units)
- Mutual Funds : 2.55
- Equity Market : 3.475
- Debentures : 5.75
- Government Bonds : 4.65
- Fixed Deposits : 3.575
- PPF : 2.025
- Gold : 5.975

Insights
- Investors are typically around 28 years old.
- Debentures and gold are the most invested avenues.
- PPF has the lowest average investment.
- Investment patterns show that while some investors prefer larger, more consistent investments in gold and debentures, others have more variability in their investments across mutual funds, equity markets, and fixed deposits.


Data_set 2 - Copy.csv
File Origin
1252: Western European (Windows)
Delimiter
Comma
Data Type Detection
Based on first 200 rows
Load
Transform Data
Cancel


Queries & Connections
Queries | Connections
1 query
Data_set 2 - Copy
40 rows loaded.




Data_set 2 - Copy.csv
Open with

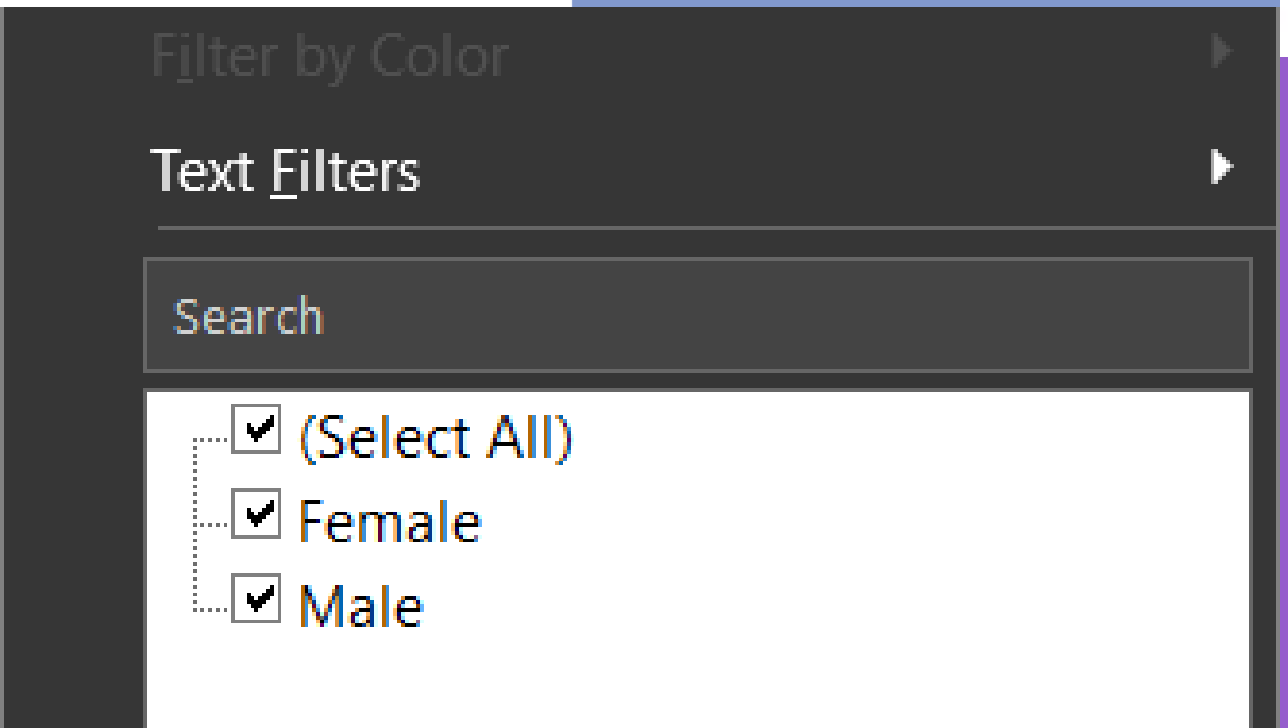
Level 1: Beginner

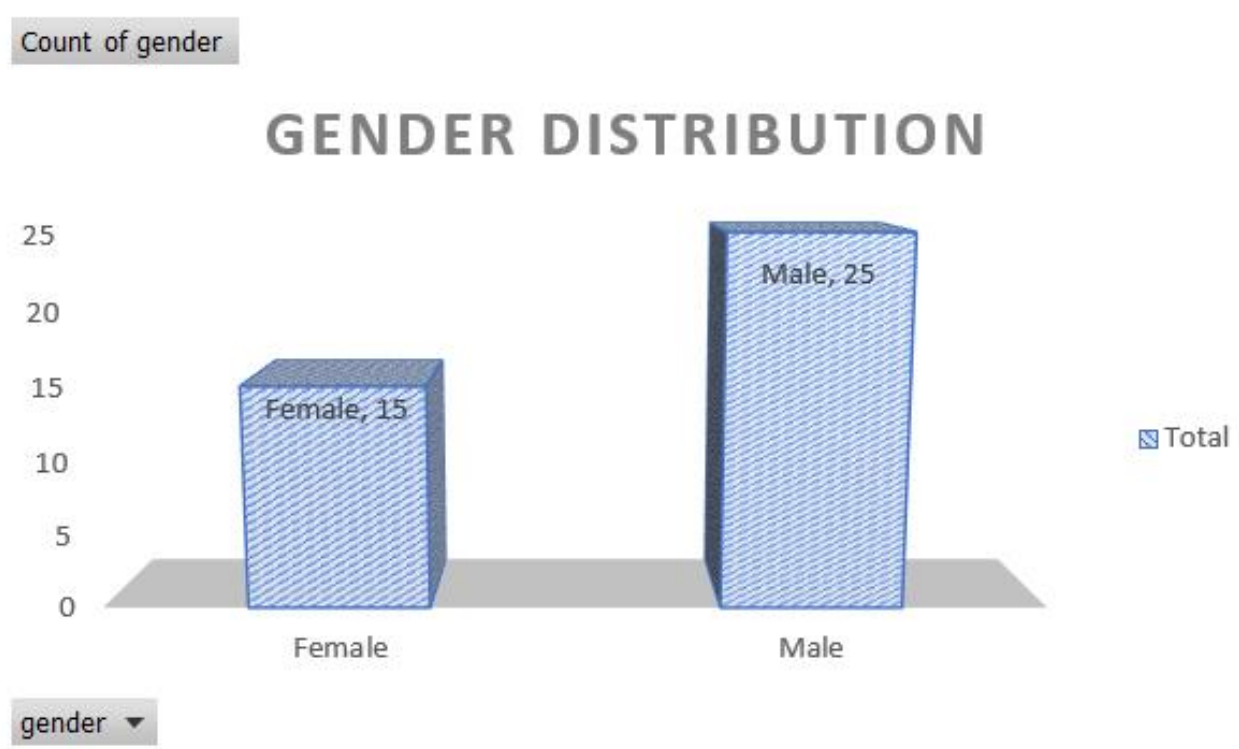
Task 2: Gender Distribution

Objective: Visualize gender distribution in the dataset.

Steps:

1. **Extract Gender Information:** Identify and extract the gender column from the dataset.
2. **Visualization:** Create a simple visualization, such as a bar chart or pie chart, to represent the distribution of genders in the dataset.





Based on the provided gender data:

- Total Entries: 40
- Gender Breakdown:
  - Female: 15 (37.5%)
  - Male: 25 (62.5%)

Visualization

- Bar Chart

Level 2: Intermediate

Cognifyz
Where Data Meets Intelligence

Task 3: Descriptive Statistics

Objective: Present basic statistics for numerical columns.

Steps:

1. **Identify Numerical Columns:** Review the dataset to identify columns containing numerical data (e.g., age, income).
2. **Calculations:** Use statistical functions (e.g., mean(), median(), std()) to calculate the mean, median, and standard deviation for each numerical column.

Basic Descriptive Statistics

Statistical parameters	age	Mutual_Funds	Equity_Market	Debentures	Government_Bonds	Fixed_Deposits	PPF	Gold
Mean	27.8	2.55	3.475	5.75	4.65	3.575	2.025	5.975
Median	27	2	4	6.5	5	3.5	1	6
Mode	27	2	4	7	5	3	1	6
min	21	1	1	1	1	1	1	2
Max	35	7	6	7	7	7	6	7
Count	40	40	40	40	40	40	40	40
standard deviation	3.52	1.18	1.12	1.65	1.35	1.77	1.59	1.13

Investment Avenues (Average Investment in Units)

Age
- Mean: 27.8
- Median: 27
- Mode: 27
- Range: 21 to 35
- Std Deviation: 3.52

Equity Market
- Mean: 3.475
- Median: 4
- Mode: 4
- Range: 1 to 6
- Std Deviation: 1.12

Government Bonds
- Mean: 4.65
- Median: 5
- Mode: 5
- Range: 1 to 7
- Std Deviation: 1.35

PPF
- Mean: 2.025
- Median: 1
- Mode: 1
- Range: 1 to 6
- Std Deviation: 1.59

Mutual Funds
- Mean: 2.55
- Median: 2
- Mode: 2
- Range: 1 to 7
- Std Deviation: 1.18

Debentures
- Mean: 5.75
- Median: 6.5
- Mode: 7
- Range: 1 to 7
- Std Deviation: 1.65

Fixed Deposits
- Mean: 3.575
- Median: 3.5
- Mode: 3
- Range: 1 to 7
- Std Deviation: 1.77

Gold
- Mean: 5.975
- Median: 6
- Mode: 6
- Range: 2 to 7
- Std Deviation: 1.13

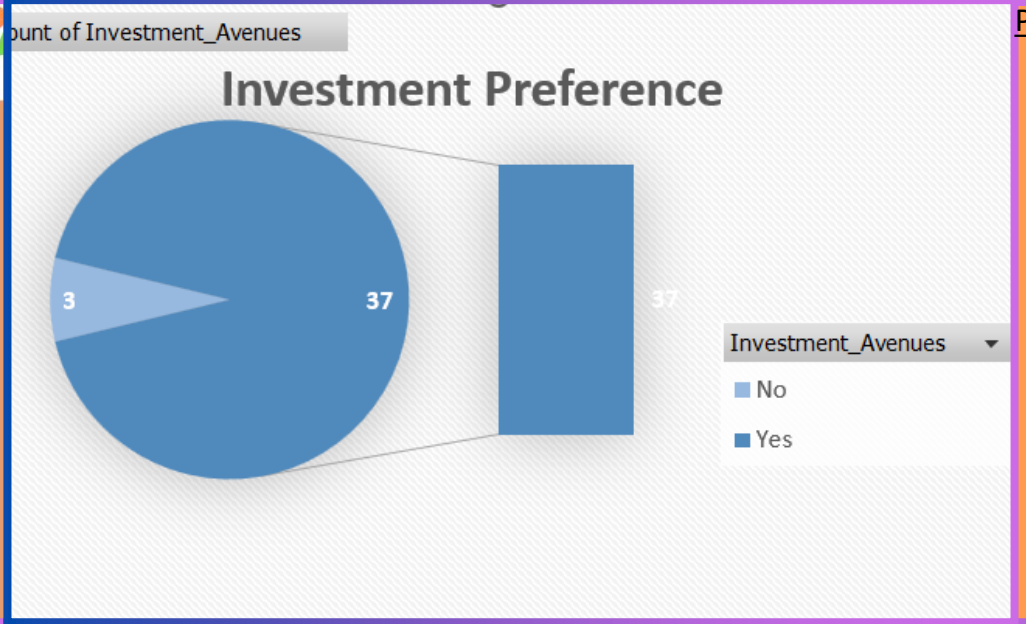
Level 2: Intermediate

Task 4: Most Preferred Investment Avenue

Objective: Identify the most preferred investment avenue.

Steps:

1. **Analyze Investment Avenues:** Examine the column containing information about different investment avenues (e.g., equity, mutual funds).
2. **Frequency Analysis:** Determine the investment avenue with the highest frequency or occurrence.



Peapoles preference Summary:

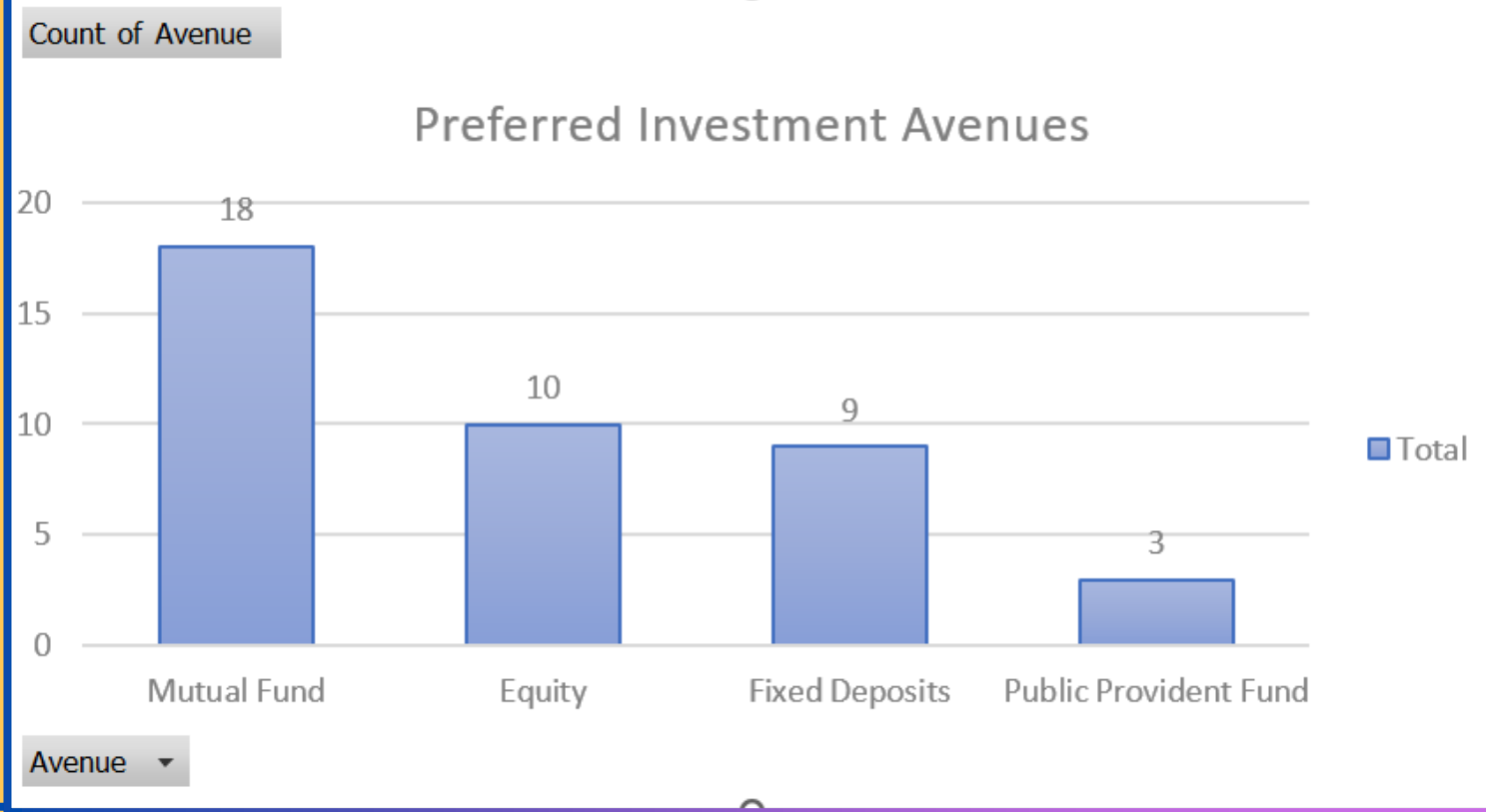
- Total YES: 37 Peapole say Yes i.e they preferred to invest.
- Total NO: 3 people say No i.e They not to Preferred Investment Avenues.

Investment Preferences Summary:

The table below showing the count of people who prefer each investment avenue:

1. Mutual Fund: Preferred by 18 individuals.
2. Equity: Preferred by 10 individuals.
3. Fixed Deposits: Preferred by 9 individuals.
4. Public Provident Fund: Preferred by 3 individuals.

Total 40 individuals are there in the dataset.
Peapole most preferred ‘Mutual Fund’ as a investment Avenue.



Level 3: Advanced

Task 5: Reasons for Investment

Objective: Analyze and summarize reasons for investment choices.

Steps:

1. **Explore Reasons Column:** Investigate the column where participants provided reasons for their investment choices.
2. **Summarize Reasons:** Identify common themes or recurring reasons and provide a summary.

Click Here to see SQl Script

Result Grid | Filter Rows: | Export:

Reason	Count
Better Returns, Assured Returns And Fixed Ret...	68
Safe and Risk Free	32
Capital Appreciation	30
Liquidity and Fund Diversification	15
Dividend	8
Tax Benefit And Tax Incentives	4
High Interest Rates	3

Investment Reasons Summary:

1. Better Returns, Assured Returns, and Fixed Returns
   Count: 68

2. Safe and Risk-Free
   Count: 32

3. Capital Appreciation
   Count: 30

4. Liquidity and Fund Diversification
   Count: 15

5. Dividend
   Count: 8

6. Tax Benefits and Tax Incentives
   Count: 4

7. High Interest Rates
   Count: 3

The majority of investors prioritize better returns, safety, and capital appreciation in their investment choices.

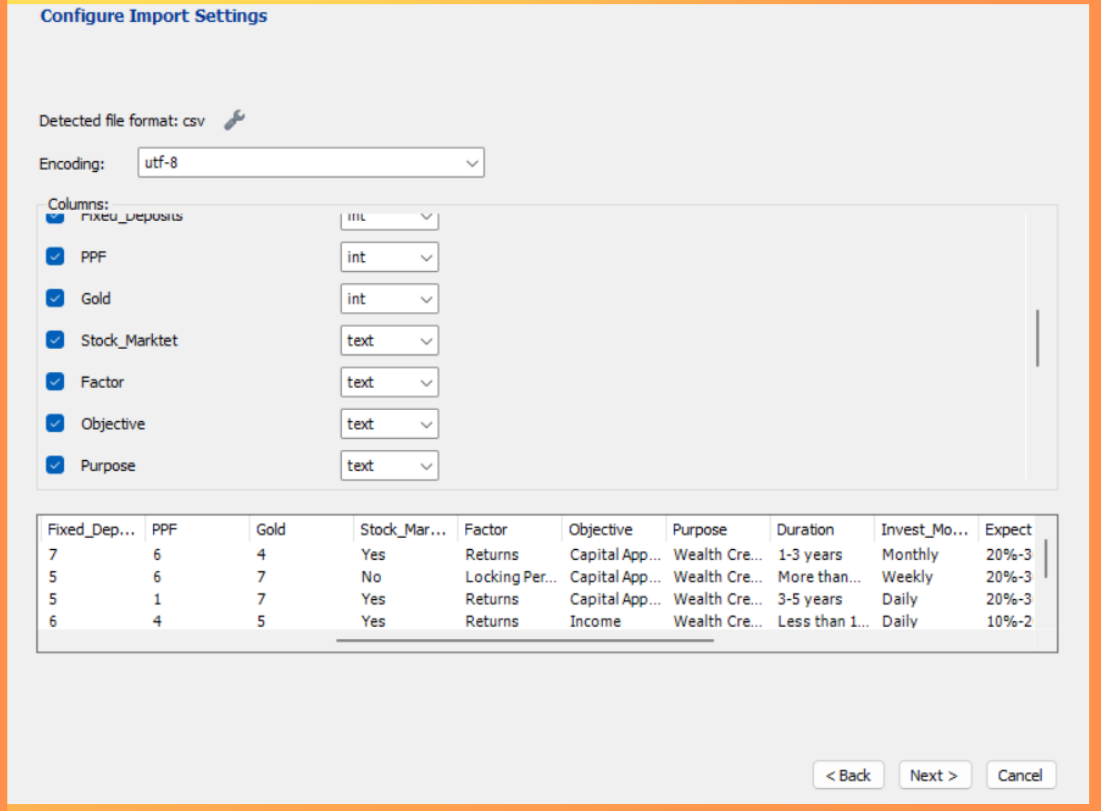
Level 3: Advanced

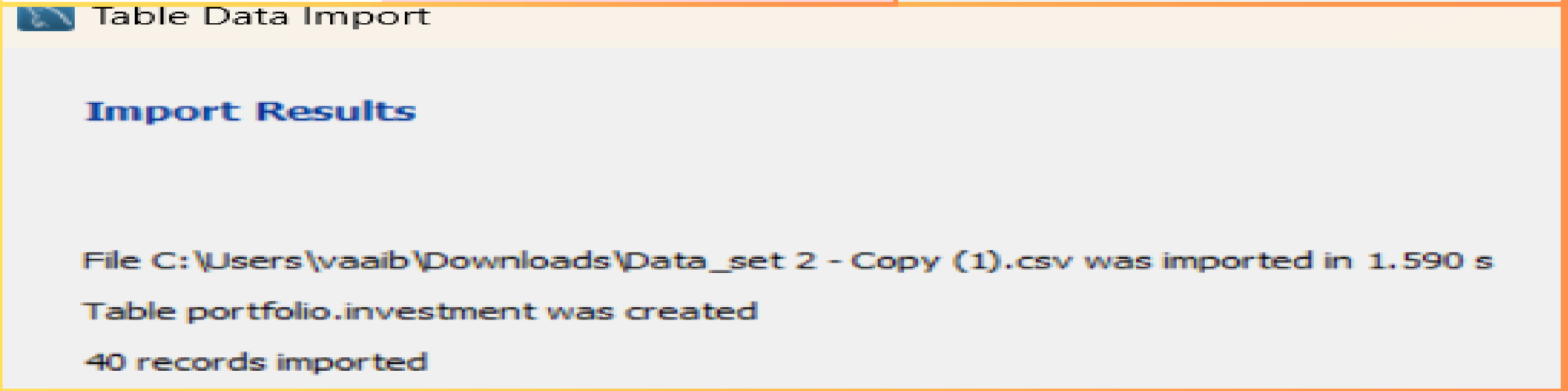
Task 6: Savings Objectives

Objective: Identify and present main savings objectives.

Steps:

1. **Analyze Savings Objectives:** Examine the column containing information about participants' savings objectives.
2. **List and Describe Objectives:** Create a list and describe the main savings objectives mentioned by participants.

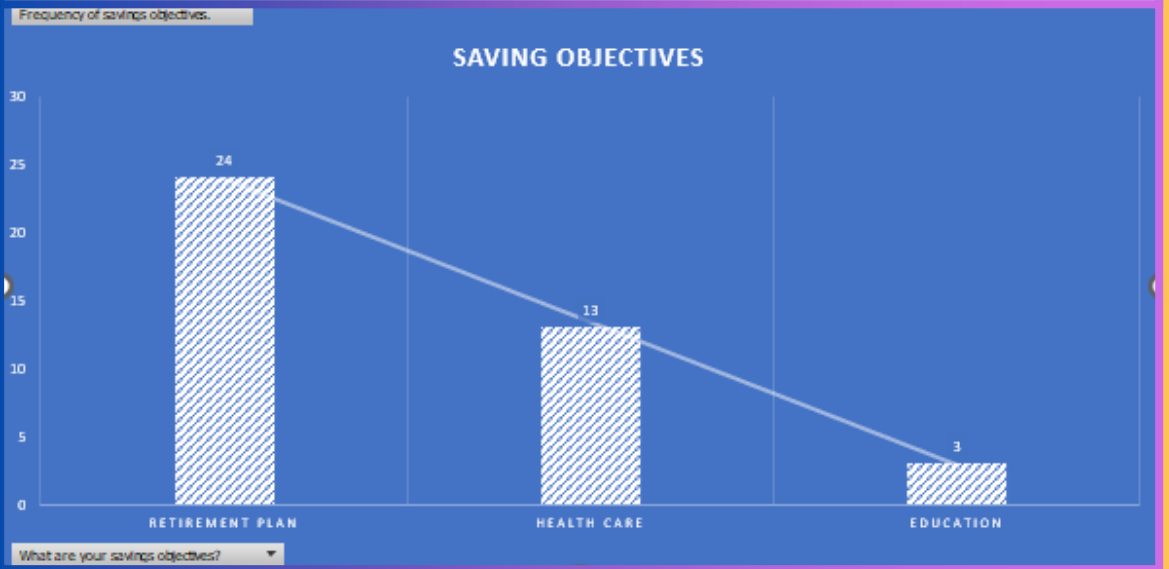




```sql
SELECT
    `What are your savings objectives?`, COUNT(*) AS Frequency
FROM
    investment
GROUP BY `What are your savings objectives?`
ORDER BY Frequency DESC
```

Result Grid | Filter Rows:

What are your savings objectives?	Frequency
Retirement Plan	24
Health Care	13
Education	3



Summary of Savings Objectives:

Based on the frequency analysis of the savings objectives:

1. Retirement Plan
   - Frequency: 24 participants
   - Description: Most common objective, chosen by 60% of participants, indicating a focus on long-term financial security.

2. Health Care
   - Frequency: 13 participants
   - Description: Chosen by 32.5% of participants, showing significant concern for medical expenses.

3. Education
   - Frequency: 3 participants
   - Description: Selected by 7.5% of participants, indicating a minority focus on educational savings.

Insights
- Primary Objective: Retirement planning
- Secondary Objective: Health care savings
- Tertiary Objective: Education savings

Participants prioritize retirement and health care over education for their savings goals.

Level 4: Expert

Cognifyz

Task 7: Common Information Sources

Objective: Analyze common sources participants rely on for investment information.

Steps:

1. **Explore Information Sources Column:** Review the column where participants indicated their sources of investment information.
2. **Identify Common Sources:** Analyze the data to identify and summarize the most common sources participants rely on.

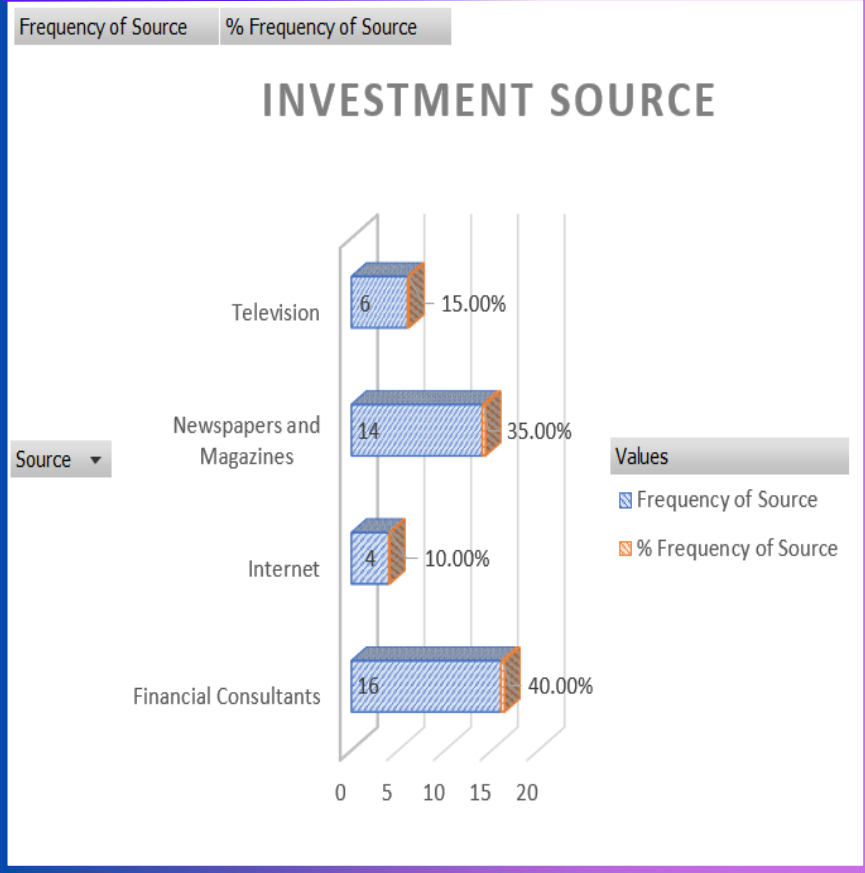
- Summary of Common Information Sources
- Based on the frequency analysis of the sources of investment information:

- Participants mainly rely on financial consultants and print media (newspapers and magazines) for their investment information. Television and the internet are used less frequently. This shows the ongoing importance of professional advice and traditional media sources in making investment decisions in India.

```
SELECT
    Source, COUNT(*) AS Frequency
FROM
    investment
GROUP BY Source
ORDER BY Frequency DESC;
```

Result Grid | Filter Rows:

Source	Frequency
Financial Consultants	16
Newspapers and Magazines	14
Television	6
Internet	4



Financial Consultants:

- Frequency: 16 participants
- Percentage: 40.00%

Newspapers and Magazines:

- Frequency: 14 participants
- Percentage: 35.00%

Television:

- Frequency: 6 participants
- Percentage: 15.00%

Internet:

- Frequency: 4 participants
- Percentage: 10.00%

Level 4: Expert

Task 8: Investment Duration

Objective: Calculate the average investment duration.

Steps:

1. **Analyze Investment Duration Column:** Examine the column containing information about participants' investment durations.
2. **Calculate Average Duration:** Use appropriate statistical methods to calculate the average investment duration.

- Using Excel:
- Splitting the duration column into two different columns naming Duration_start and Duration end.
- Splitting can help us to calculate the average investment duration.
- We can use the following formula to create another column named 'Average year'
- Average year = ((Duration_start + Duration_end) / 2)
- Then Calculate average of Average year

e	Purpose	Duration	Invest_Monitor	Expe
Appreciation	Wealth Creation	1-3 years	Monthly	20%-
Appreciation	Wealth Creation	More than 5 years	Weekly	20%-

Find and Replace

Find | Replace

Find what: More than 5 years

Replace with: 5-5

Options >>

Replace All | Replace | Find All | Find Next | Close

Book	Sheet	Name	Cell	Value	Formula
cognify Internship project.xlsx	Data_set 2 - Copy		R3C15	More than 5 years	

ive	Purpose	Duration	Invest_Monitor	Ex
Appreciation	Wealth Creation	1-3 years	Monthly	20
Appreciation	Wealth Creation	05-May	Weekly	20
Appreciation	Wealth Creation	3-5 years	Daily	20
	Wealth Creation	Less than 1 year	Daily	10

Find and Replace

Find | Replace

Find what: Less than 1 year

Replace with: 1-1

Options >>

Replace All | Replace | Find All | Find Next | Close

Book	Sheet	Name	Cell	Value	Formula
cognify Internship project.xlsx	Data_set 2 - Copy		R5C15	Less than 1 year	
cognify Internship project.xlsx	Data_set 2 - Copy		R6C15	Less than 1 year	

Objective	Purpose	Duration	Invest_Monitor	Expe
Capital Appreciation	Wealth Creation	1-3 years	Monthly	20%

Convert Text to Columns Wizard - Step 2 of 3

This screen lets you set the delimiters your data contains. You can see how your text is affected in the preview below.

Delimiters

Tab

Semicolon

Comma

Space

Other: -

Treat consecutive delimiters as one

Text qualifier: "

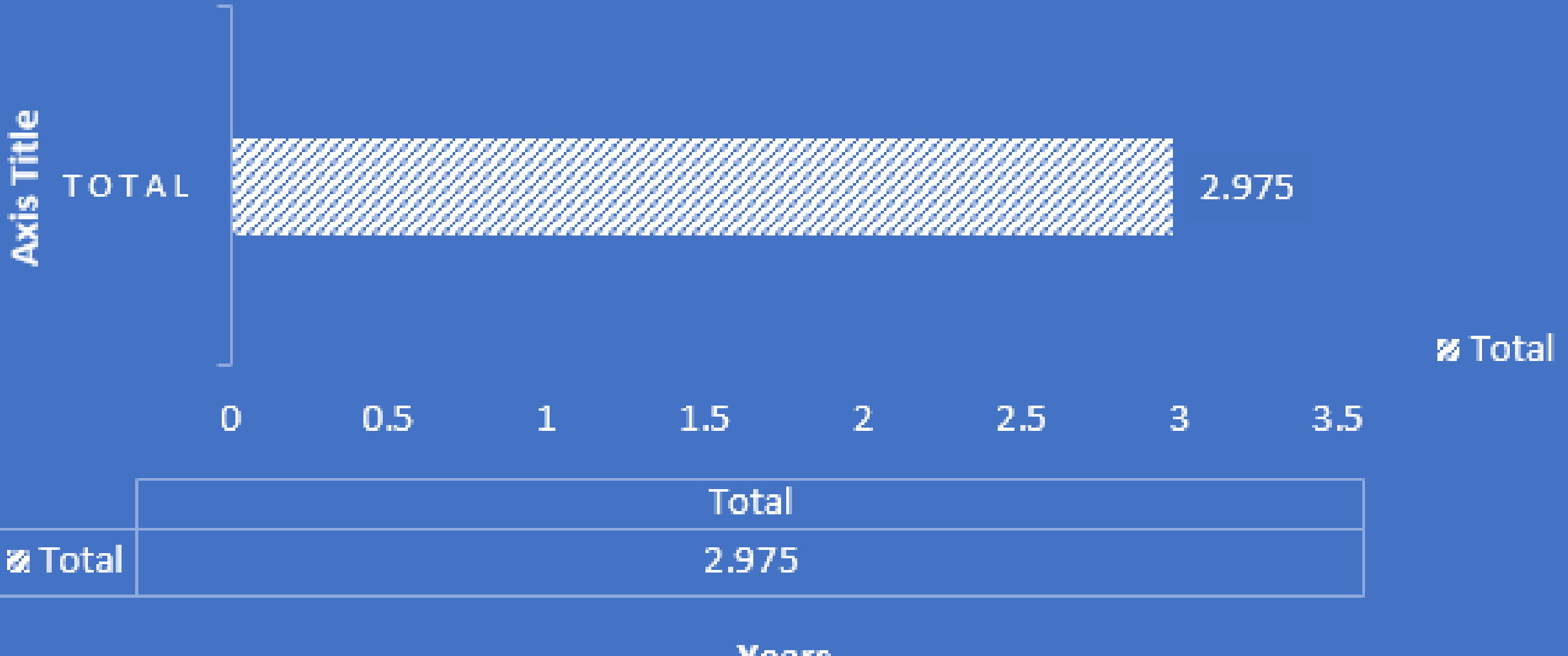
Data preview

Duration
1 3 years
5
3 5 years
1
1

Cancel | < Back | Next > | Finish

Average of Average year

AVERAGE INVESTMENT DURATION



	Total
Total	2.975

- Calculate Average investment duration by using SQL queries.

```sql
with DurationMidPoint AS (
    SELECT Duration,
    CASE
        WHEN Duration LIKE 'More than%' THEN CAST(TRIM(SUBSTRING(Duration, 11,2)) AS UNSIGNED)
        WHEN Duration LIKE 'Less than%' THEN CAST(TRIM(SUBSTRING(Duration, 11,2)) AS UNSIGNED)
        ELSE (
            CAST(SUBSTRING_INDEX(Duration, '-' , 1) AS UNSIGNED)+
            CAST(SUBSTRING_INDEX(SUBSTRING_INDEX(Duration, ' ' , 1), '-' , -1) AS UNSIGNED)) / 2
            END AS Midpoint
     FROM investment
)
SELECT
    AVG(Midpoint) AS average_duration_in_years
FROM
    DurationMidPoint;
```

Result Grid | Filter Rows:

average_duration_in_years
2.97500000

- Explanation of the SQL Query:

- WHEN Duration LIKE 'more than%' THEN CAST(TRIM(SUBSTRING(Duration, 11, 2)) AS UNSIGNED):

- WHEN Duration LIKE 'less than%' THEN CAST(TRIM(SUBSTRING(Duration, 11, 2)) AS UNSIGNED):
- Extracts the substring starting at the 11th character and takes 2 characters (e.g., "5 ").
- TRIM is used to remove any leading or trailing spaces to clean up the extracted substring.
- CAST(... AS UNSIGNED) converts the trimmed substring to an integer.

- CAST(SUBSTRING_INDEX(Duration, '-', 1) AS UNSIGNED)
- extracts the value before (‘-‘) of the range.

- CAST(SUBSTRING_INDEX (SUBSTRING_INDEX(Duration, ' ', 1), '-', -1) AS UNSIGNED)
- extracts the value before (space) and after (‘-’) of the range.
- The midpoint is calculated as the average of These two extracted values.
- The outer SELECT statement calculates the average of all midpoints.

- Summary of Full Code:
- Extract lower and upper bounds from the duration ranges.
- Calculate the midpoint for each duration range.
- Compute the average of these midpoints to find the average investment duration.
- The average investment duration in years = 2.975 years.

Level 4: Expert

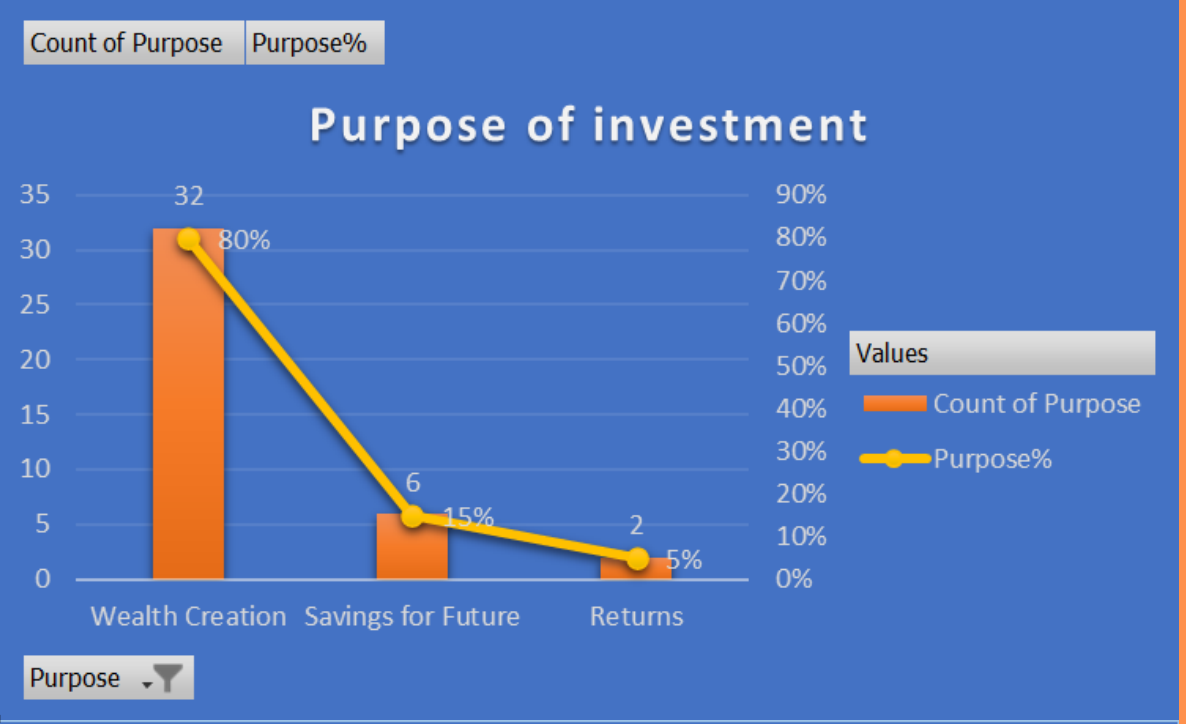
Cognifyz
Where Data Meets Intelligence

Task 9: Expectations from Investments

Objective: Summarize participants' expectations from investments.

Steps:

1. **Explore Expectations Column:** Review the column where participants provided information about their expectations.
2. **List and Describe Expectations:** Create a list and describe the common expectations mentioned by participants.



- Purpose:
- Wealth Creation

  Count: 32 investors
- Expectations:

- 26 investors expected 30%-40% returns.
- 4 investors expected 20%-30% returns.
- 2 investors.10%-20% returns.

- Purpose:
- Savings for Future

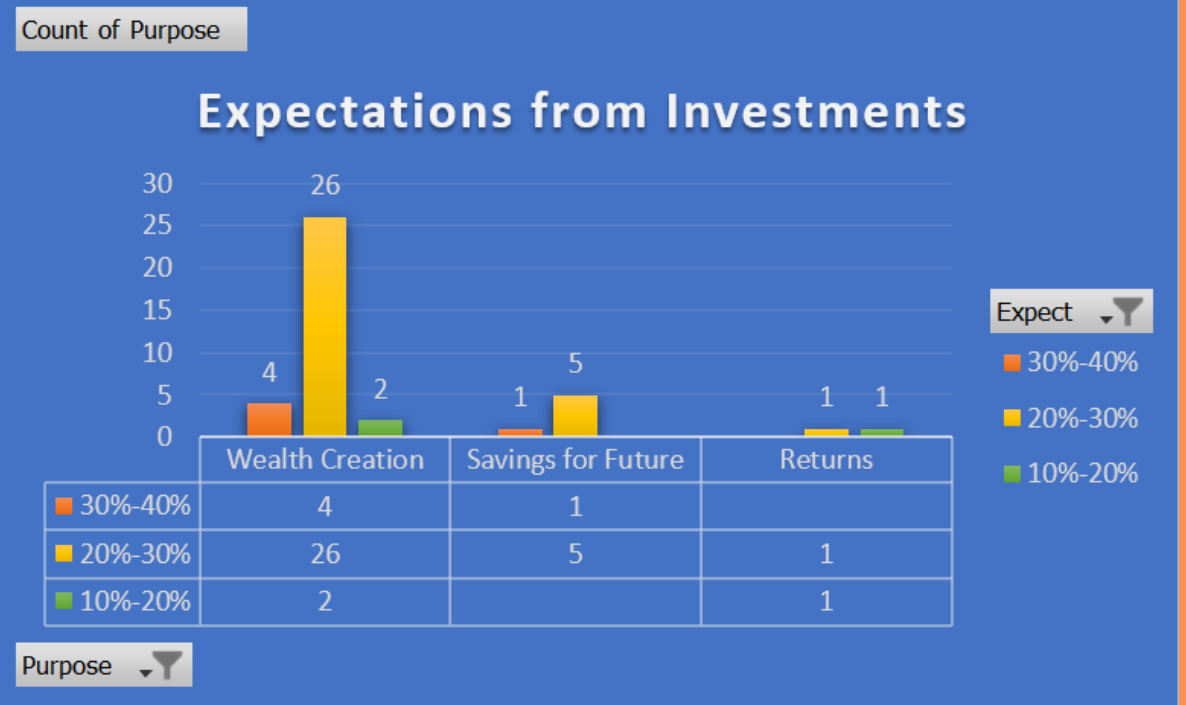
  Count: 6 investors
- Expectations:

- 1 investor expected 30%-40% returns.
- 5 investors. expected 20%-30% returns.
- 0 investors. expected 10%-20% returns:

- Purpose:
- Returns

  Count: 2 investors
- Expectations:

- 0 investors expected 30%-40% returns:
- 1 investor. expected 20%-30% returns.
- 1 investor expected 10%-20% returns.



- Summarize participants' expectations from investments, using SQL

```sql
SELECT
    purpose,
    Expect AS Expect_returns,
    COUNT(*) AS Count_of_investors
FROM
    Investment
GROUP BY Expect_returns , purpose;
```

Result Grid | Filter Rows: | Export:

purpose	Expect_returns	Count_of_investors
Wealth Creation	20%-30%	26
Wealth Creation	10%-20%	2
Wealth Creation	30%-40%	4
Savings for Future	20%-30%	5
Returns	10%-20%	1
Returns	20%-30%	1
Savings for Future	30%-40%	1

Summary:

Wealth Creation:

- The majority of participants 32 (80%) are investing to create wealth. Most of these investors expect a return between 20%-30%, with a few expecting higher or lower returns.

Savings for Future:

- 6 (15%) participants invest with the goal of saving for the future. Most of these investors also expect a return between 20%-30%.

Returns:

- A smaller group of 2 participants (5%) is primarily focused on earning returns. Their expectations are split between 20%-30% and 10%-20% returns.

Level 4: Expert

Task 10: Correlation Analysis

Cognifyz
Where Data Meets Intelligence

Objective: Explore potential correlations between factors.

Steps:

1. **Select Relevant Columns:** Identify columns such as age, investment duration, and expected returns for correlation analysis.
2. **Use Statistical Methods or Visualizations:** Employ statistical methods (e.g., correlation coefficients) or visualizations (e.g., scatter plots) to explore and visualize potential correlations.

Correlation between Age and Investment Duration	0.06
Correlation between Age and Expected Returns	-0.09
Correlation between Investment Duration and Expected Returns	0.26

Used the 'CORREL' function in Excel to calculate the correlations between the columns.

Insights from Correlation Values

1. Age and Investment Duration (0.06):
Almost no relationship. Age doesn't significantly impact investment duration.

2. Age and Expected Returns (-0.09):
Weak inverse relationship. Younger investors slightly expect higher returns.

3. Investment Duration and Expected Returns (0.26) :
Moderate positive relationship. Longer investment durations are associated with higher expected returns.

Summary

Age vs. Duration: Age has minimal effect on investment duration.

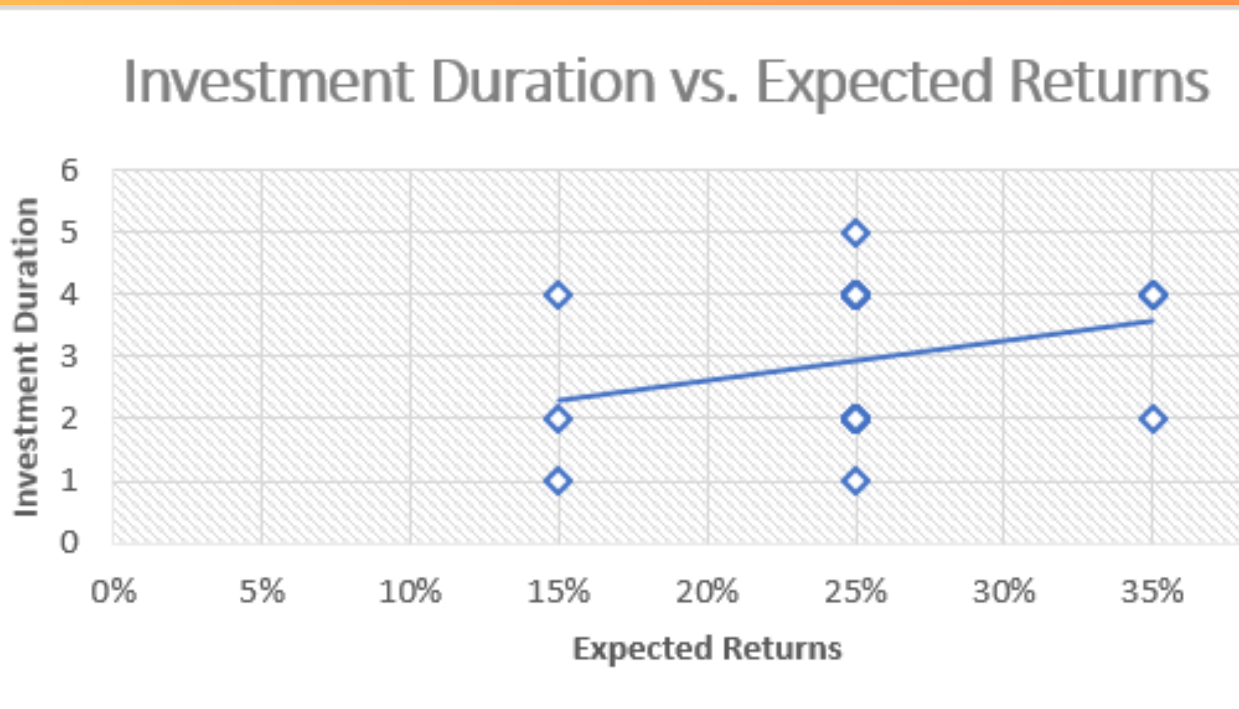
Age vs. Returns: Younger participants have slightly higher return expectations.

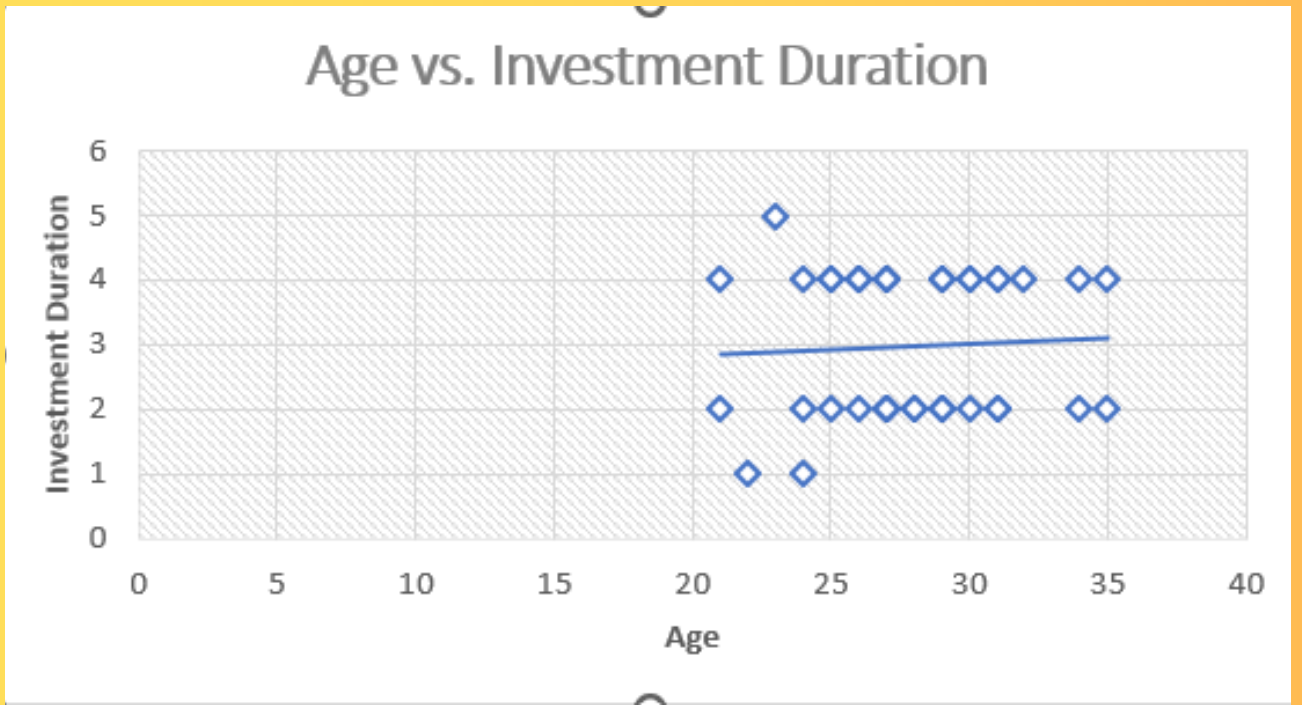
Duration vs. Returns: Longer investments tend to have higher expected returns.

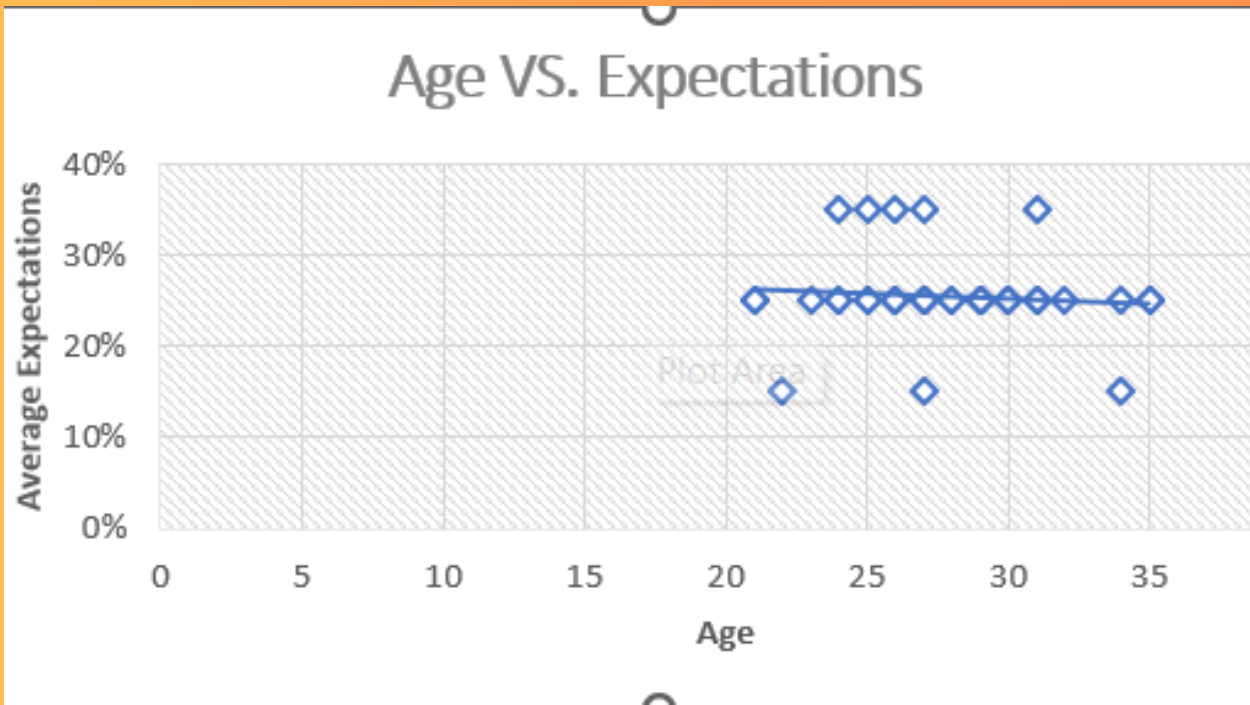
Correlation analysis using ‘EXCEL’

Used the ‘CORREL’ function in Excel to calculate the correlations between the columns.

Correlation between Age and Investment Duration	0.06
Correlation between Age and Expected Returns	-0.09
Correlation between Investment Duration and Expected Returns	0.26

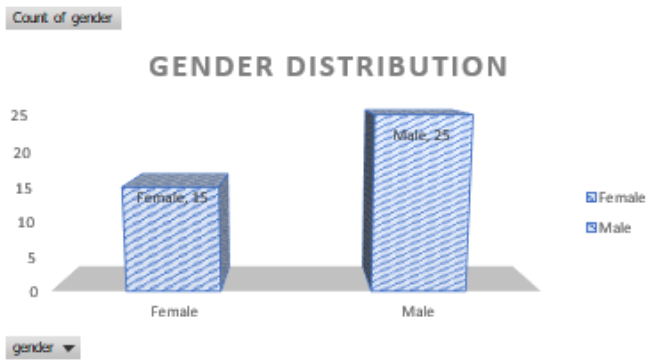


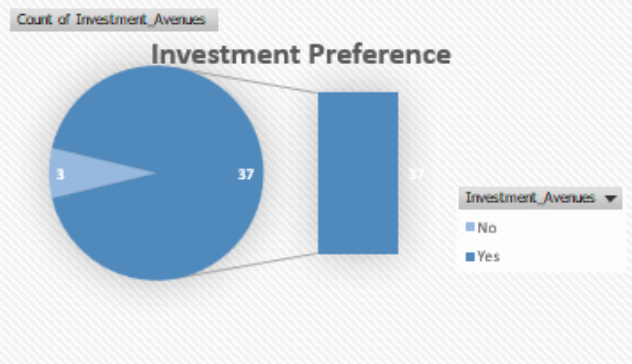


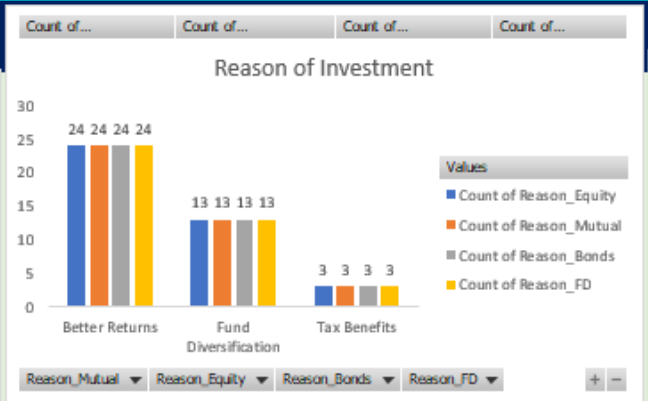


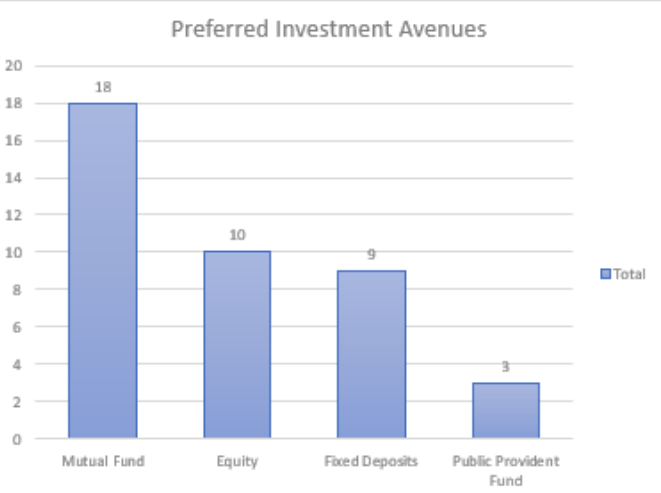
Excel Dashboard

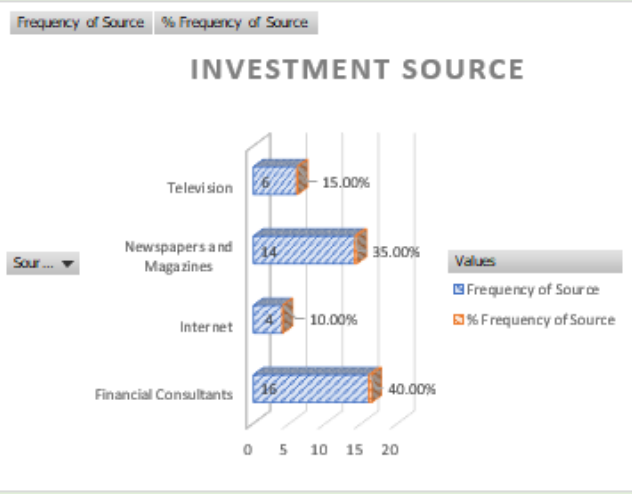
Investment Analysis Dashboard

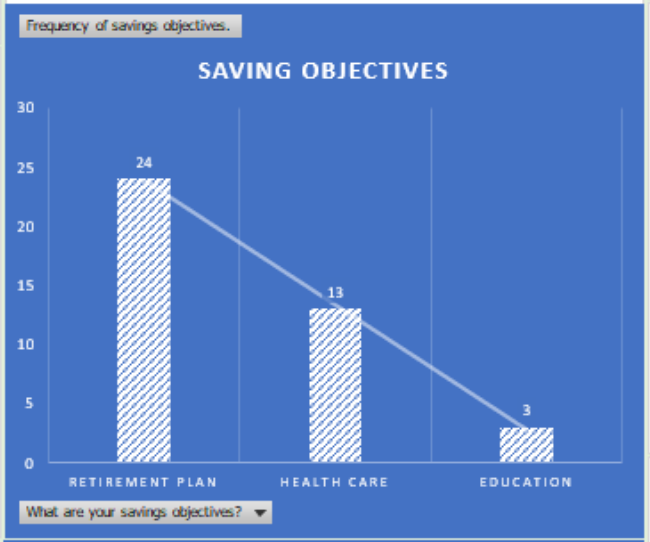


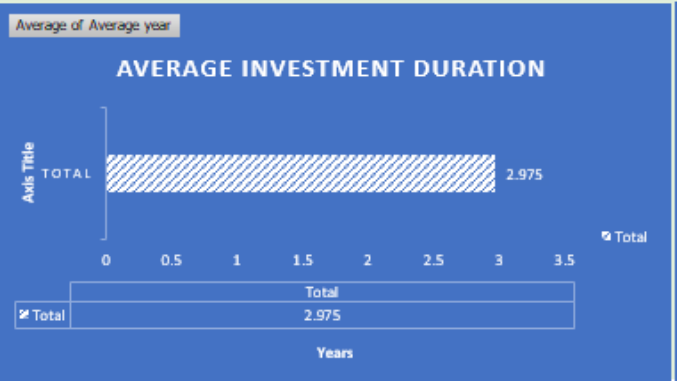


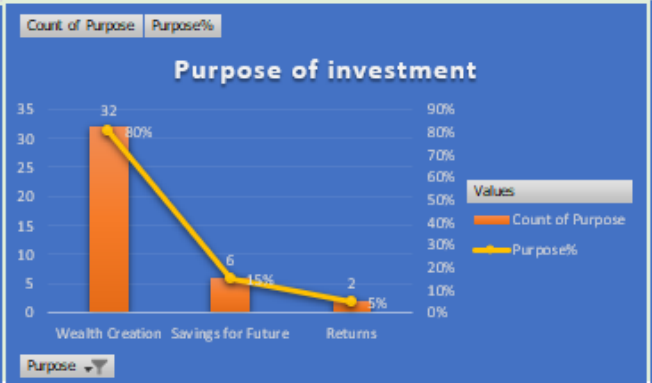


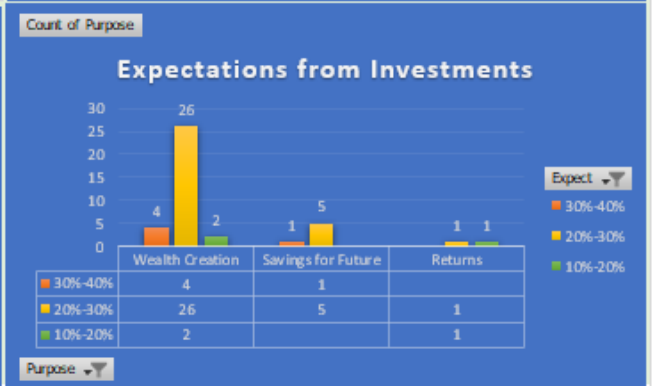




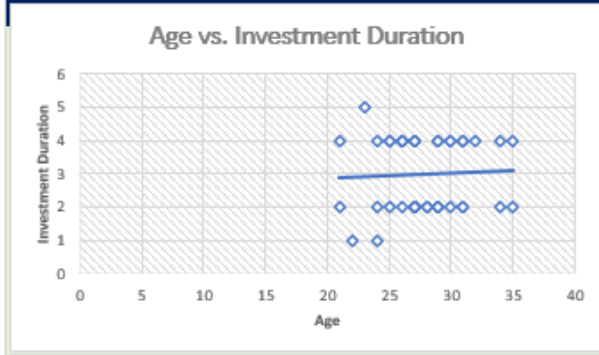


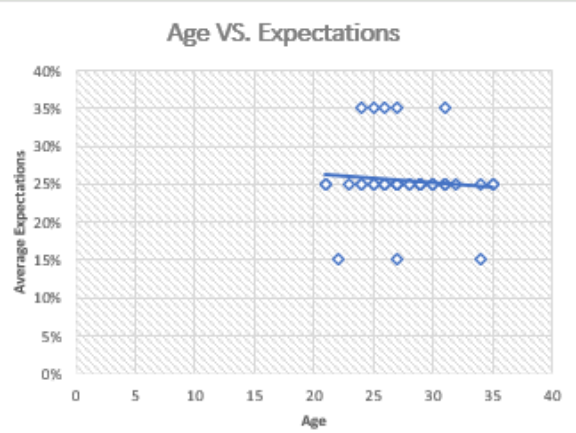


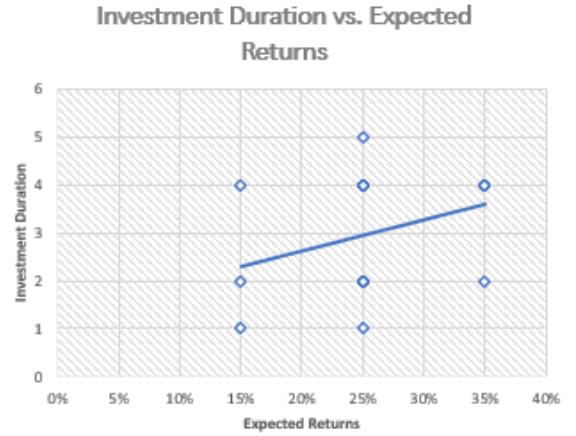




Potential Correlation between factors







Click to open

Microsoft Power BI - An Interactive Power Bi Dashboard

INVESTMENT ANALYSIS

Average investment in governmentt bonds
4.7
Govt_Bond iinvestment n years

Average investment in Fixed Deposits
3.6
FD investment in years

Average investment in debentures
5.8
Debentures investment in ye...

Average investment in Mutual Funds
2.6
Mutual Fund investment in y...

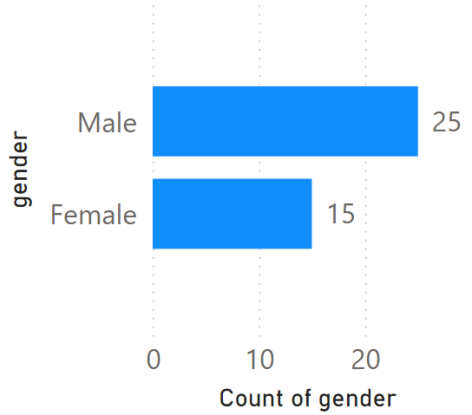
Average investment in Gold
6.0
Gold investment in years

Average investment in PPF
2.0
PPF investment In Years

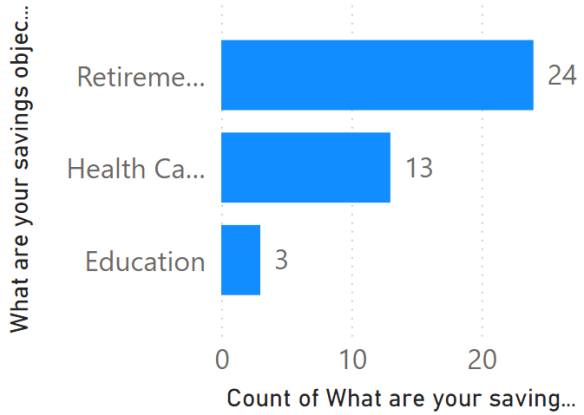
Average investment in equity
3.5
equity investment in years

Average Age
28
age in years

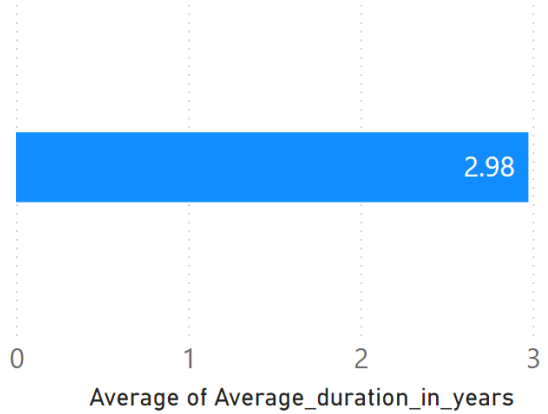
Gender Distribution



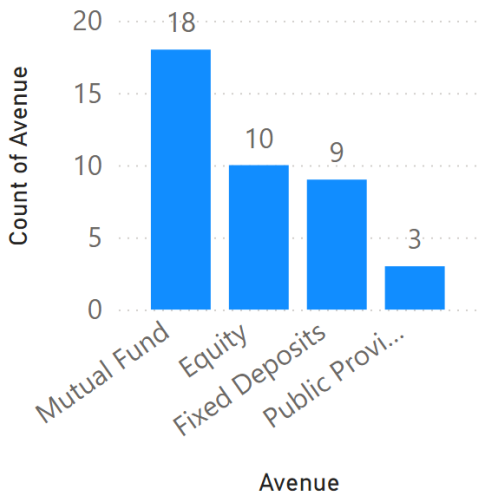
Savings Objectives



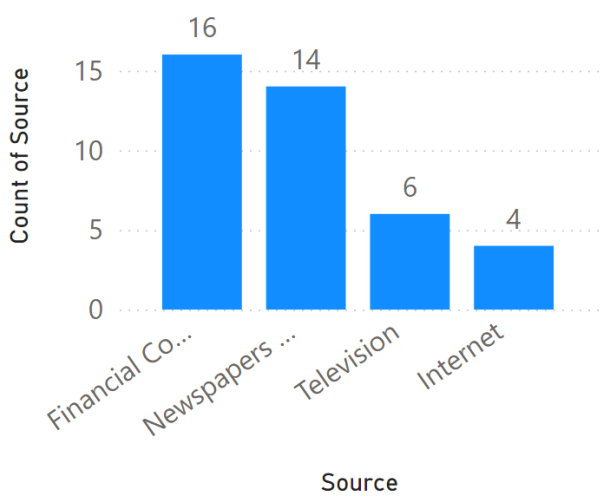
Average Investment duration in years



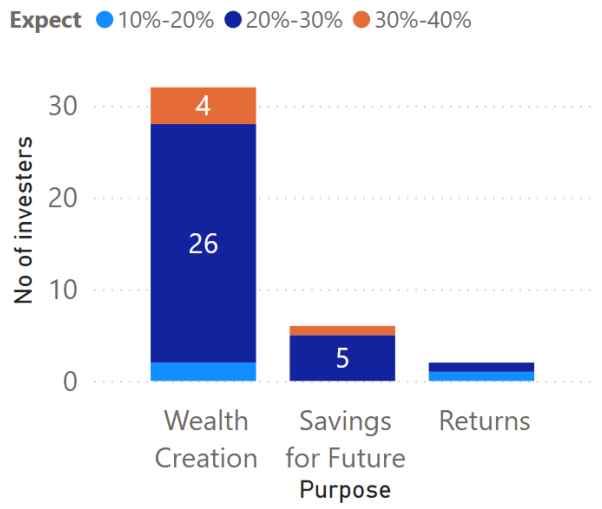
Investment Preference



Investment Source



No of investers by Purpose and Expect



Gender
- Female
- Male

Invest_Moniter
- Daily
- Monthly
- Weekly

Factor
- Locking Pe...
- Returns
- Risk

Duration
- 1-1 years
- 1-3 years
- 3-5 years
- 5-5 years

Avenue
- Equity
- Fixed Depo...
- Mutual Fund
- Public Prov...

Thank you

Contact Details

Linkdin www.linkedin.com/in/vaibhav000

GitHub https://github.com/Vaibhav3046

Email : vaaibhav000@gmail.com