Cognifyz INTERNSHIP PROGRAM BUSINESS ANALYTICS

Project Objective

This project aims to analyze investment patterns, preferences, and expectations among participants based on demographics, investment choices, savings objectives, information sources, and risk factors. The goal is to extract
insights
that aid financial decision-making and risk management.

Use and Benefits

Provides valuable investor behavior insights to financial advisors and institutions.
Supports better portfolio management by understanding investment horizons and expected returns.
Helps product designers tailor financial products to client needs.
Assists in identifying risk factors and diversification strategies.
Facilitates data-driven decision-making for stakeholders.

How It Helps a Business Analyst

Enables a Business Analyst to gather, clean, and analyze relevant financial data effectively.
Supports requirement gathering by understanding client investment behavior and expectations.
Provides actionable insights to support strategic financial planning and forecasting.
Facilitates visualization and reporting of complex data for stakeholders.
Enhances problem-solving and decision-making skills through risk and correlation analysis.
Bridges communication between technical teams and business stakeholders by translating data insights into business strategies.

Full Summary

The dataset analyzed included 40 participants with detailed demographics and investment information. Key findings
showed a male majority, preference for Equity Market and Mutual Funds, primary motivations of Capital Appreciation,
and retirement planning as a leading savings goal. Information mainly came from financial consultants and newspapers.
Investments typically last around 3 years with expected returns mostly between 20%-30%. Weak correlation between age
and duration indicates independent risk factors, highlighting the importance of diverse strategies.

**This overview contextualizes the analysis and shows the critical role of business analysis skills in extracting and
applying investment insights for business value.**

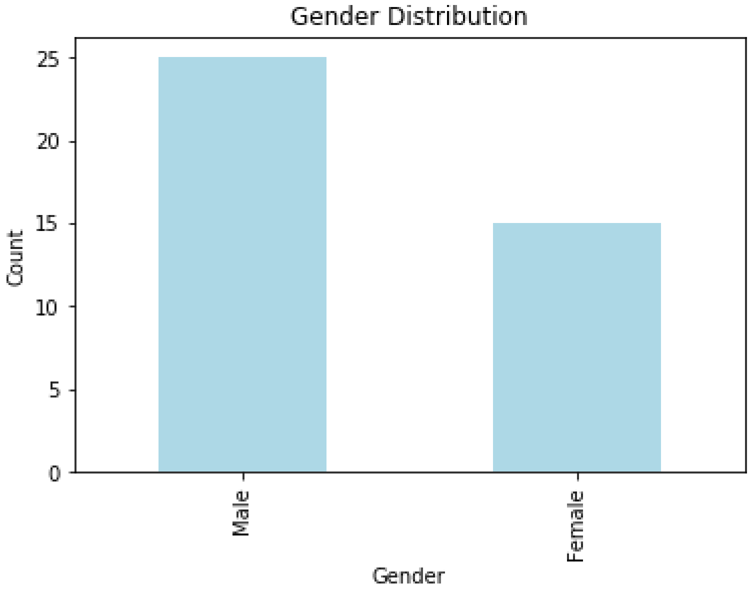
```
# Task 1: Dataset Familiarization
import pandas as pd

# Load the dataset from absolute Windows path
df = pd.read_csv('C:/Users/msec-018-21/Downloads/Data_set 2 - Copy.csv')

# Display data types of each column
print(df.dtypes)

# Show first 5 rows to get overview
print(df.head())

# Check for missing values in each column
print(df.isnull().sum())
```

```
gender                                object
age                                    int64
Investment_Avenues                    object
Mutual_Funds                           int64
Equity_Market                          int64
Debentures                             int64
Government_Bonds                       int64
Fixed_Deposits                         int64
PPF                                    int64
Gold                                   int64
Stock_Marktet                         object
Factor                                object
Objective                             object
Purpose                               object
Duration                              object
Invest_Monitor                        object
Expect                                object
Avenue                                object
What are your savings objectives?    object
Reason_Equity                         object
Reason_Mutual                         object
Reason_Bonds                          object
Reason_FD                             object
Source                                object
dtype: object
   gender  age Investment_Avenues  Mutual_Funds  Equity_Market  Debentures  
\
0  Female   34                Yes             1              2           5
1  Female   23                Yes             4              3           2
2    Male   30                Yes             3              6           4
3    Male   22                Yes             2              1           3
4  Female   24                 No             2              1           3

   Government_Bonds  Fixed_Deposits  PPF  Gold            ...             \
0                 3               7    6     4            ...
1                 1               5    6     7            ...
2                 2               5    1     7            ...
3                 7               6    4     5            ...
4                 6               4    5     7            ...

            Duration Invest_Monitor   Expect       Avenue  \
0          1-3 years        Monthly  20%-30%  Mutual Fund
1  More than 5 years         Weekly  20%-30%  Mutual Fund
2          3-5 years          Daily  20%-30%       Equity
3   Less than 1 year          Daily  10%-20%       Equity
4   Less than 1 year          Daily  20%-30%       Equity

  What are your savings objectives?         Reason_Equity  \
0                   Retirement Plan  Capital Appreciation
1                       Health Care              Dividend
2                   Retirement Plan  Capital Appreciation
3                   Retirement Plan              Dividend
4                   Retirement Plan  Capital Appreciation

          Reason_Mutual     Reason_Bonds            Reason_FD  \
0        Better Returns  Safe Investment        Fixed Returns
1        Better Returns  Safe Investment  High Interest Rates
2          Tax Benefits  Assured Returns        Fixed Returns
3  Fund Diversification   Tax Incentives  High Interest Rates
4        Better Returns  Safe Investment            Risk Free
```

```
                     Source
0  Newspapers and Magazines
1     Financial Consultants
2                Television
3                  Internet
4                  Internet

[5 rows x 24 columns]
gender                               0
age                                  0
Investment_Avenues                   0
Mutual_Funds                         0
Equity_Market                        0
Debentures                           0
Government_Bonds                     0
Fixed_Deposits                       0
PPF                                  0
Gold                                 0
Stock_Marktet                        0
Factor                               0
Objective                            0
Purpose                              0
Duration                             0
Invest_Monitor                       0
Expect                               0
Avenue                               0
What are your savings objectives?    0
Reason_Equity                        0
Reason_Mutual                        0
Reason_Bonds                         0
Reason_FD                            0
Source                               0
dtype: int64
```

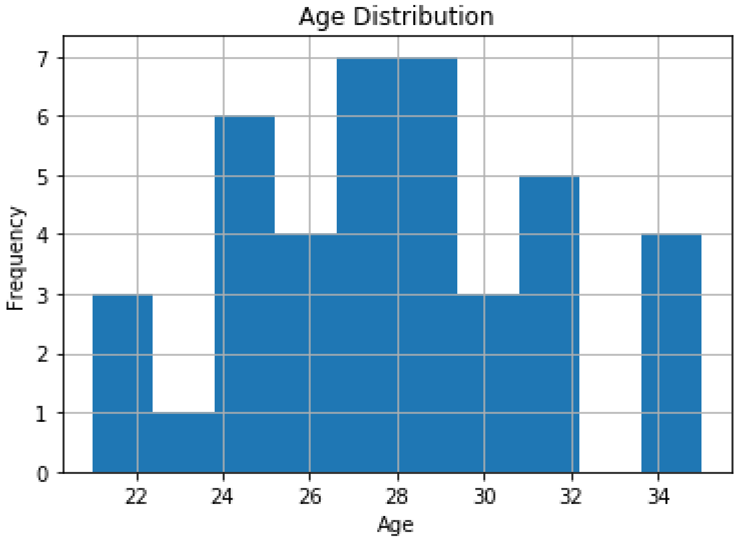
```
# Task 2: Gender Distribution
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('C:/Users/msec-018-21/Downloads/Data_set 2 - Copy.csv')

# Count how many participants belong to each gender category
gender_counts = df['gender'].value_counts()
print(gender_counts)

# Plot gender distribution as bar chart gender_counts.plot(kind='bar',
color='lightblue')
plt.title('Gender Distribution')
plt.xlabel('Gender')
plt.ylabel('Count')
plt.show()
```

```
Male      25
Female    15
Name: gender, dtype: int64
```



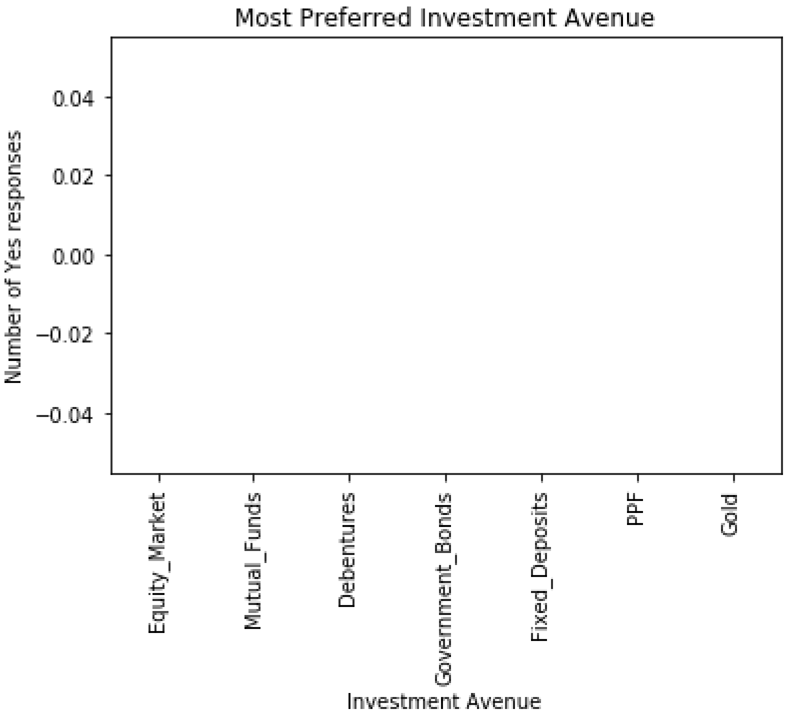
```python
# Task 3: Descriptive Statistics
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('C:/Users/msec-018-21/Downloads/Data_set 2 - Copy.csv')

# Show statistical summary of numerical columns
print(df.describe())

# Print median age
print(df['age'].median())

# Plot histogram of age distribution
df['age'].hist()
plt.title('Age Distribution')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()
```

```
             age  Mutual_Funds  Equity_Market  Debentures  Government_Bonds  
\
count  40.000000     40.000000      40.000000   40.000000         40.000000   
mean   27.800000      2.550000       3.475000    5.750000          4.650000   
std     3.560467      1.197219       1.131994    1.675617          1.369072   
min    21.000000      1.000000       1.000000    1.000000          1.000000   
25%    25.750000      2.000000       3.000000    5.000000          4.000000   
50%    27.000000      2.000000       4.000000    6.500000          5.000000   
75%    30.000000      3.000000       4.000000    7.000000          5.000000   
max    35.000000      7.000000       6.000000    7.000000          7.000000   

       Fixed_Deposits        PPF       Gold  
count       40.000000  40.000000  40.000000  
mean         3.575000   2.025000   5.975000  
std          1.795828   1.609069   1.143263  
min          1.000000   1.000000   2.000000  
25%          2.750000   1.000000   6.000000  
50%          3.500000   1.000000   6.000000  
75%          5.000000   2.250000   7.000000  
max          7.000000   6.000000   7.000000  
27.0
```



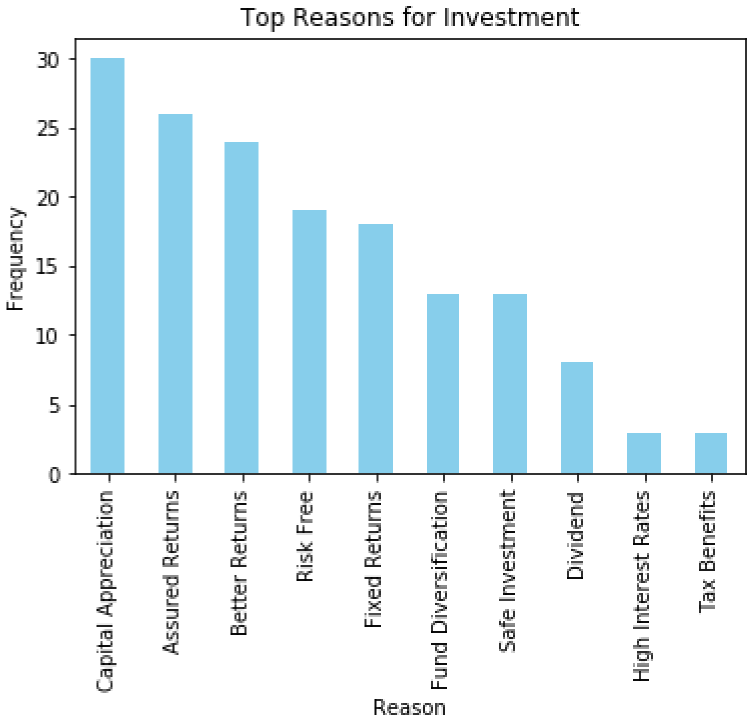
```python
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('C:/Users/msec-018-21/Downloads/Data_set 2 - Copy.csv')

# List of investment avenue columns (edit names to match your file exactly)
avenue_columns = ['Equity_Market', 'Mutual_Funds', 'Debentures', 'Government_Bo
print(df.columns.tolist())  # Check and match column names

# Count 'yes' responses for each avenue
preference_counts = {col: (df[col].astype(str).str.lower() == 'yes').sum() for
preference_series = pd.Series(preference_counts)
print(preference_series)

# Bar chart
preference_series.plot(kind='bar')
plt.title('Most Preferred Investment Avenue')
plt.xlabel('Investment Avenue')
plt.ylabel('Number of Yes responses')
plt.show()
```

```
['gender', 'age', 'Investment_Avenues', 'Mutual_Funds', 'Equity_Market', 'D
ebentures', 'Government_Bonds', 'Fixed_Deposits', 'PPF', 'Gold', 'Stock_Mar
ktet', 'Factor', 'Objective', 'Purpose', 'Duration', 'Invest_Monitor', 'Exp
ect', 'Avenue', 'What are your savings objectives?', 'Reason_Equity', 'Reas
on_Mutual', 'Reason_Bonds', 'Reason_FD', 'Source']
Equity_Market       0
Mutual_Funds        0
Debentures          0
Government_Bonds    0
Fixed_Deposits      0
PPF                 0
Gold                0
dtype: int64
```



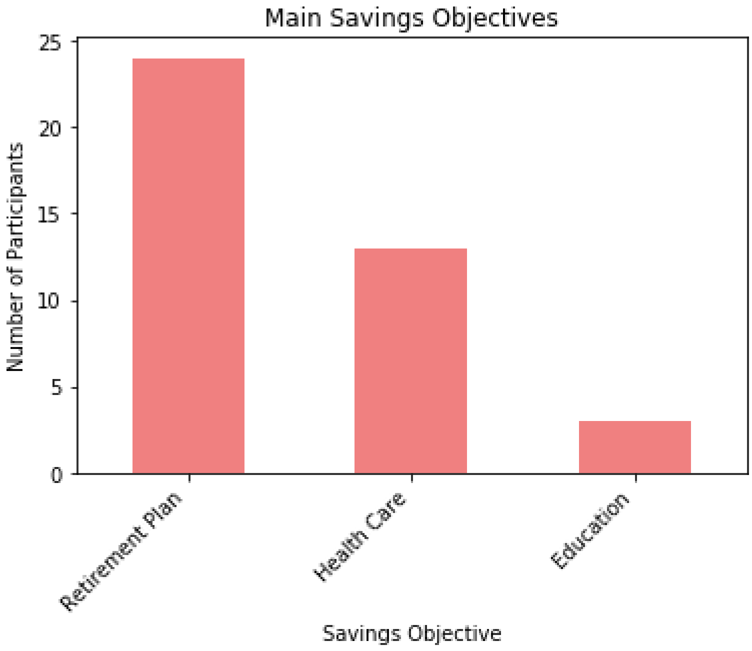
```python
# Task 5: Reasons for Investment
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('C:/Users/msec-018-21/Downloads/Data_set 2 - Copy.csv')

# Columns containing reasons for investment
reason_columns = ['Reason_Equity', 'Reason_Mutual', 'Reason_Bonds', 'Reason_FD'

# Combine all reasons into one series
all_reasons = pd.Series(dtype=str)
for col in reason_columns:
    all_reasons = all_reasons.append(df[col].dropna().astype(str))

# Count frequency of each reason
reason_counts = all_reasons.value_counts()
print(reason_counts)

# Plot top 10 reasons
reason_counts.head(10).plot(kind='bar', color='skyblue')
plt.title('Top Reasons for Investment')
plt.xlabel('Reason')
plt.ylabel('Frequency')
plt.show()
```

```
Capital Appreciation    30
Assured Returns         26
Better Returns          24
Risk Free               19
Fixed Returns           18
Fund Diversification    13
Safe Investment         13
Dividend                 8
High Interest Rates      3
Tax Benefits             3
Liquidity                2
Tax Incentives           1
dtype: int64
```


Top Reasons for Investment
Frequency
0
5
10
15
20
25
30
Capital Appreciation
Assured Returns
Better Returns
Risk Free
Fixed Returns
Fund Diversification
Safe Investment
Dividend
High Interest Rates
Tax Benefits
Reason

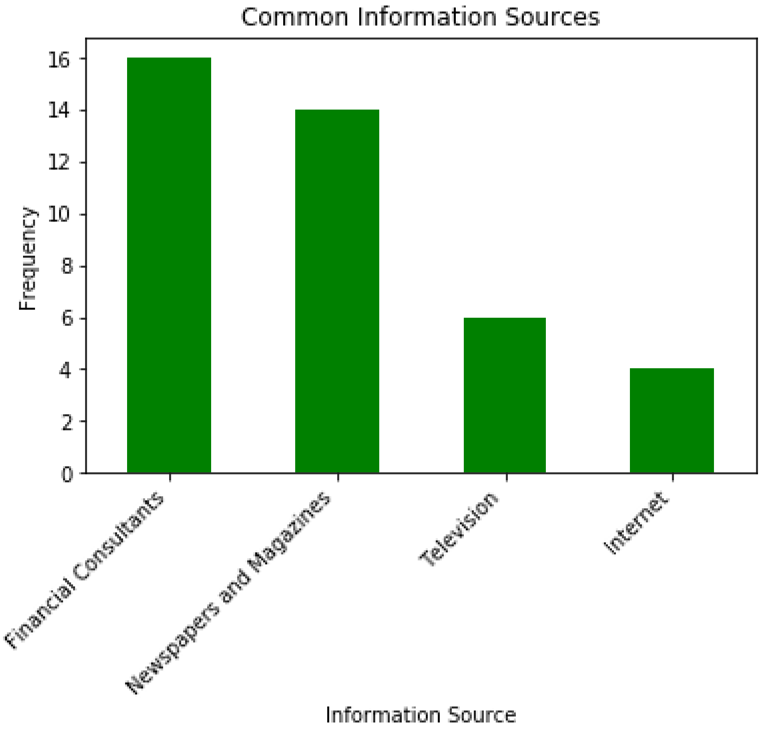
```
# Task 6: Savings Objectives
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('C:/Users/msec-018-21/Downloads/Data_set 2 - Copy.csv')

# Column with savings objectives
savings_obj_col = 'What are your savings objectives?'

# Count how many times each objective appears
savings_counts = df[savings_obj_col].value_counts()
print(savings_counts)

# Bar chart of savings objectives
savings_counts.plot(kind='bar', color='lightcoral')
plt.title('Main Savings Objectives')
plt.xlabel('Savings Objective')
plt.ylabel('Number of Participants')
plt.xticks(rotation=45, ha='right')
plt.show()
```

```
Retirement Plan    24
Health Care        13
Education           3
Name: What are your savings objectives?, dtype: int64
```



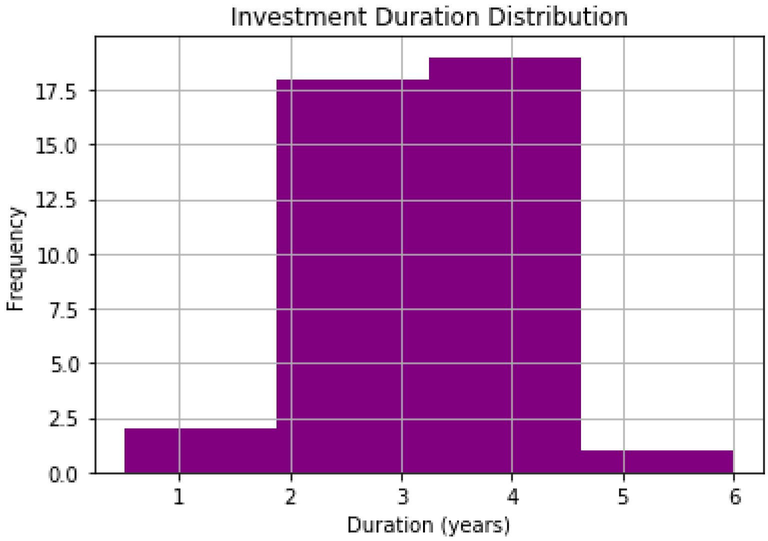
```python
# Task 7: Common Information Sources
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('C:/Users/msec-018-21/Downloads/Data_set 2 - Copy.csv')

# Column indicating sources of investment information
source_col = 'Source'

# Count frequency of each source
source_counts = df[source_col].value_counts()
print(source_counts)

# Bar chart for information sources
source_counts.plot(kind='bar', color='green')
plt.title('Common Information Sources')
plt.xlabel('Information Source')
plt.ylabel('Frequency')
plt.xticks(rotation=45, ha='right')
plt.show()
```

```
Financial Consultants      16
Newspapers and Magazines   14
Television                  6
Internet                    4
Name: Source, dtype: int64
```



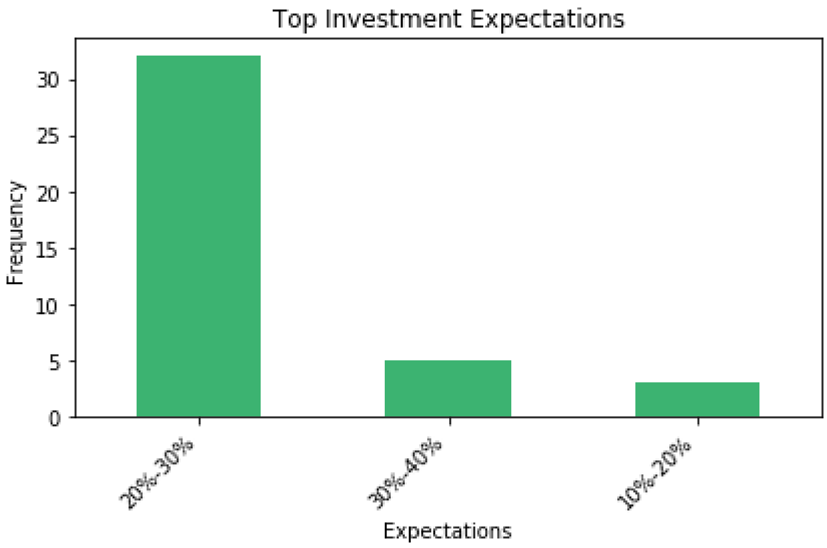
```python
# Task 8: Investment Duration Analysis
import pandas as pd
import matplotlib.pyplot as plt

# Load dataset
df = pd.read_csv('C:/Users/msec-018-21/Downloads/Data_set 2 - Copy.csv')

# Check exact column name for duration
print(df.columns.tolist())

# Map textual durations to numeric years (choose midpoints for range bins)
duration_map = {
    'Less than 1 year': 0.5,
    '1-3 years': 2,
    '3-5 years': 4,
    'More than 5 years': 6  # Or any value >5, for plotting
}

# Apply the mapping
df['Duration_num'] = df['Duration'].map(duration_map)

# Print average investment duration in years
average_duration = df['Duration_num'].mean()
print(f"Average Investment Duration (years): {average_duration:.2f}")

# Plot histogram of investment durations (by years)
df['Duration_num'].hist(bins=4, color='purple')
plt.title('Investment Duration Distribution')
plt.xlabel('Duration (years)')
plt.ylabel('Frequency')
plt.show()
```

```
['gender', 'age', 'Investment_Avenues', 'Mutual_Funds', 'Equity_Market', 'D
ebentures', 'Government_Bonds', 'Fixed_Deposits', 'PPF', 'Gold', 'Stock_Mar
ktet', 'Factor', 'Objective', 'Purpose', 'Duration', 'Invest_Monitor', 'Exp
ect', 'Avenue', 'What are your savings objectives?', 'Reason_Equity', 'Reas
on_Mutual', 'Reason_Bonds', 'Reason_FD', 'Source']
Average Investment Duration (years): 2.98
```



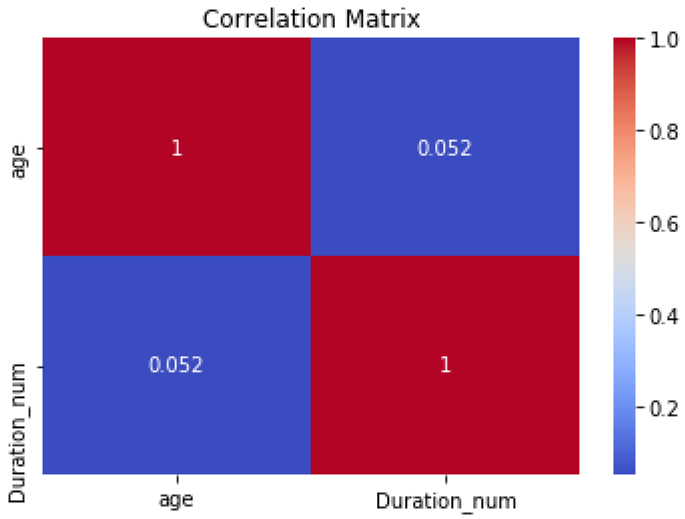
```
import pandas as pd
import matplotlib.pyplot as plt

# Load the dataset
df = pd.read_csv('C:/Users/msec-018-21/Downloads/Data_set 2 - Copy.csv')

# Use the exact expectations column name from your file
expectations_col = 'Expect'  # Change if this is shown differently in your CSV

# Drop missing values and convert to string
expectations = df[expectations_col].dropna().astype(str)

# Count frequency of each unique expectation
expectations_counts = expectations.value_counts()
print(expectations_counts)

# Bar chart for the top 10 expectations
expectations_counts.head(10).plot(kind='bar', color='mediumseagreen')
plt.title('Top Investment Expectations')
plt.xlabel('Expectations')
plt.ylabel('Frequency')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```

```
20%-30%    32
30%-40%     5
10%-20%     3
Name: Expect, dtype: int64
```



```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load data
df = pd.read_csv('C:/Users/msec-018-21/Downloads/Data_set 2 - Copy.csv')

# Map durations as in Task 8 (create numeric duration column if not already cre
duration_map = {
    'Less than 1 year': 0.5,
    '1-3 years': 2,
    '3-5 years': 4,
    'More than 5 years': 6
}
df['Duration_num'] = df['Duration'].map(duration_map)

# Select numeric columns available for correlation
corr_cols = ['age', 'Duration_num']  # Add more if you have other numeric colum

# Build DataFrame with only valid, numeric columns
corr_df = df[corr_cols]
print(corr_df.corr())

# Plot the correlation heatmap
sns.heatmap(corr_df.corr(), annot=True, cmap='coolwarm')
plt.title('Correlation Matrix')
plt.show()
```

```
                   age  Duration_num
age           1.000000      0.051756
Duration_num  0.051756      1.000000
```


Correlation Matrix
age
Duration_num
age
Duration_num