Finance & Stock Market Analysis – Investor Behavior & Decision Insights

---

🟦 Title Page

Project Title: Finance & Stock Market Analysis – Investor Behavior & Decision Insights

Author: Rahul Mahakal

Domain: Finance & Investment Analytics

Tools & Technologies Used:

- Python
- Pandas & NumPy
- Matplotlib & Seaborn
- Jupyter Notebook
- Modular Python Architecture
- OS-independent file handling using `pathlib`

🟦 Executive Summary

This project analyzes investor demographic data and financial behavior to understand investment preferences, risk appetite, monitoring patterns, and decision-making drivers. By combining descriptive statistics with advanced behavioral visualizations, the study provides actionable insights into how different investor segments choose financial instruments and manage their portfolios.

🟦 Introduction

Business Problem

Individual investors exhibit diverse financial behaviors influenced by age, risk tolerance, investment objectives, and monitoring habits. Understanding these patterns is essential for financial institutions, advisors, and policymakers to design targeted products and improve investor outcomes.

Objectives

- Analyze preferred investment avenues
- Study equity market participation
- Understand investment objectives and duration preferences
- Examine risk appetite and monitoring behavior
- Identify correlations among financial instruments
- Generate data-driven financial recommendations

🟦 Dataset Description

Dataset Overview

The dataset contains survey-based investment behavior data, capturing demographic information, preferred investment avenues, risk factors, objectives, and monitoring habits.

Key Columns

- `gender`, `age` – Demographics
- `Investment_Avenues` – Primary investment choice
- `Equity_Market`, `Mutual_Funds`, `Fixed_Deposits`, `Gold`, `PPF` – Instrument preferences
- `Objective` – Investment objective
- `Duration` – Investment horizon
- `Factor` – Risk factor
- `Invest_Monitor` – Portfolio monitoring frequency
- `What are your savings objectives?` – Savings motivation
- `Reason_Equity`, `Reason_Mutual` – Behavioral motivation

🟦 Data Source

Finance behavior dataset used for analytical and portfolio development purposes.

Dataset Size

- Records: Based on CSV file
- Data Type: Predominantly categorical with demographic attributes
- Missing Values: Validated during preprocessing

🟦 Methodology

Analytics Pipeline

1. Data ingestion and validation
2. Feature engineering (age groups, equity participation)
3. Exploratory data analysis
4. Behavioral pattern analysis
5. Statistical summarization
6. Insight and recommendation generation

Tools & Techniques

- GroupBy and Crosstab analysis
- Categorical distribution analysis
- Behavioral finance visualization techniques
- Correlation analysis between financial instruments

🟦 DATA LOADING & VALIDATION

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import sys
from pathlib import Path

sns.set_theme(style="whitegrid")

BASE_DIR = Path.cwd().parent
DATA_DIR = BASE_DIR / "datasets"

# Detect project root (Multi-Domain-Data-Analysis-Portfolio)
PROJECT_ROOT = Path.cwd().parents[0]

# Add to PYTHONPATH
sys.path.insert(0, str(PROJECT_ROOT))

print("=" * 95)
print("Project root added:", PROJECT_ROOT)
print("=" * 95)
print("src visible:", (PROJECT_ROOT / "src").exists())
print("=" * 95)

df = pd.read_csv(DATA_DIR / "Finance_data.csv")
df.head()
```

```
===============================================================================================
=
Project root added: C:\Users\mahakal.r\PycharmProjects\Multi-Domain-Data-Analysis-Portfolio
===============================================================================================
=
src visible: True
===============================================================================================
=
```

Out[49]:

	gender	age	Investment_Avenues	Mutual_Funds	Equity_Market	Debentures	Government_Bonds
0	Female	34	Yes	1	2	5	3
1	Female	23	Yes	4	3	2	1
2	Male	30	Yes	3	6	4	2
3	Male	22	Yes	2	1	3	7
4	Female	24	No	2	1	3	6

5 rows × 24 columns

◄ ▬▬▬▬▬▬▬▬▬▬▬ ►

In [50]:
```python
df.columns
```

Index(['gender', 'age', 'Investment_Avenues', 'Mutual_Funds', 'Equity_Market',
       'Debentures', 'Government_Bonds', 'Fixed_Deposits', 'PPF', 'Gold',
       'Stock_Marktet', 'Factor', 'Objective', 'Purpose', 'Duration',
       'Invest_Monitor', 'Expect', 'Avenue',
       'What are your savings objectives?', 'Reason_Equity', 'Reason_Mutual',
       'Reason_Bonds', 'Reason_FD', 'Source'],
      dtype='object')

In [51]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 40 entries, 0 to 39
Data columns (total 24 columns):
 #   Column                             Non-Null Count  Dtype
---  ------                             --------------  -----
 0   gender                             40 non-null     object
 1   age                                40 non-null     int64
 2   Investment_Avenues                 40 non-null     object
 3   Mutual_Funds                       40 non-null     int64
 4   Equity_Market                      40 non-null     int64
 5   Debentures                         40 non-null     int64
 6   Government_Bonds                   40 non-null     int64
 7   Fixed_Deposits                     40 non-null     int64
 8   PPF                                40 non-null     int64
 9   Gold                               40 non-null     int64
 10  Stock_Marktet                      40 non-null     object
 11  Factor                             40 non-null     object
 12  Objective                          40 non-null     object
 13  Purpose                            40 non-null     object
 14  Duration                           40 non-null     object
 15  Invest_Monitor                     40 non-null     object
 16  Expect                             40 non-null     object
 17  Avenue                             40 non-null     object
 18  What are your savings objectives?  40 non-null     object
 19  Reason_Equity                      40 non-null     object
 20  Reason_Mutual                      40 non-null     object
 21  Reason_Bonds                       40 non-null     object
 22  Reason_FD                          40 non-null     object
 23  Source                             40 non-null     object
dtypes: int64(8), object(16)
memory usage: 7.6+ KB
```

In [52]: `df.isnull().sum()`

```
Out[52]: 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
```

🟦 SAFE FEATURE ENGINEERING

```python
In [53]:  # Age bands (financially meaningful)
          df["AGE_GROUP"] = pd.cut(
              df["age"],
              bins=[18, 25, 35, 45, 60, 100],
              labels=["18-25", "26-35", "36-45", "46-60", "60+"]
          )

          # Binary equity preference
          df["EQUITY_INVESTOR"] = df["Equity_Market"].apply(
              lambda x: 1 if str(x).lower() == "yes" else 0
          )
```

🟦 Exploratory Data Analysis (EDA)

1️⃣ Preferred Investment Avenues

```python
In [54]:  plt.figure(figsize=(6, 5))

          sns.countplot(
              data=df,
              x="Investment_Avenues",
              hue="Investment_Avenues",     # REQUIRED
              palette="Set2",
              legend=False
          )

          plt.title("Preferred Investment Avenues", fontsize=14)
          plt.xlabel("Investment Avenue")
          plt.ylabel("Number of Investors")
          plt.xticks(rotation=0)
          plt.tight_layout()
          plt.show()
```
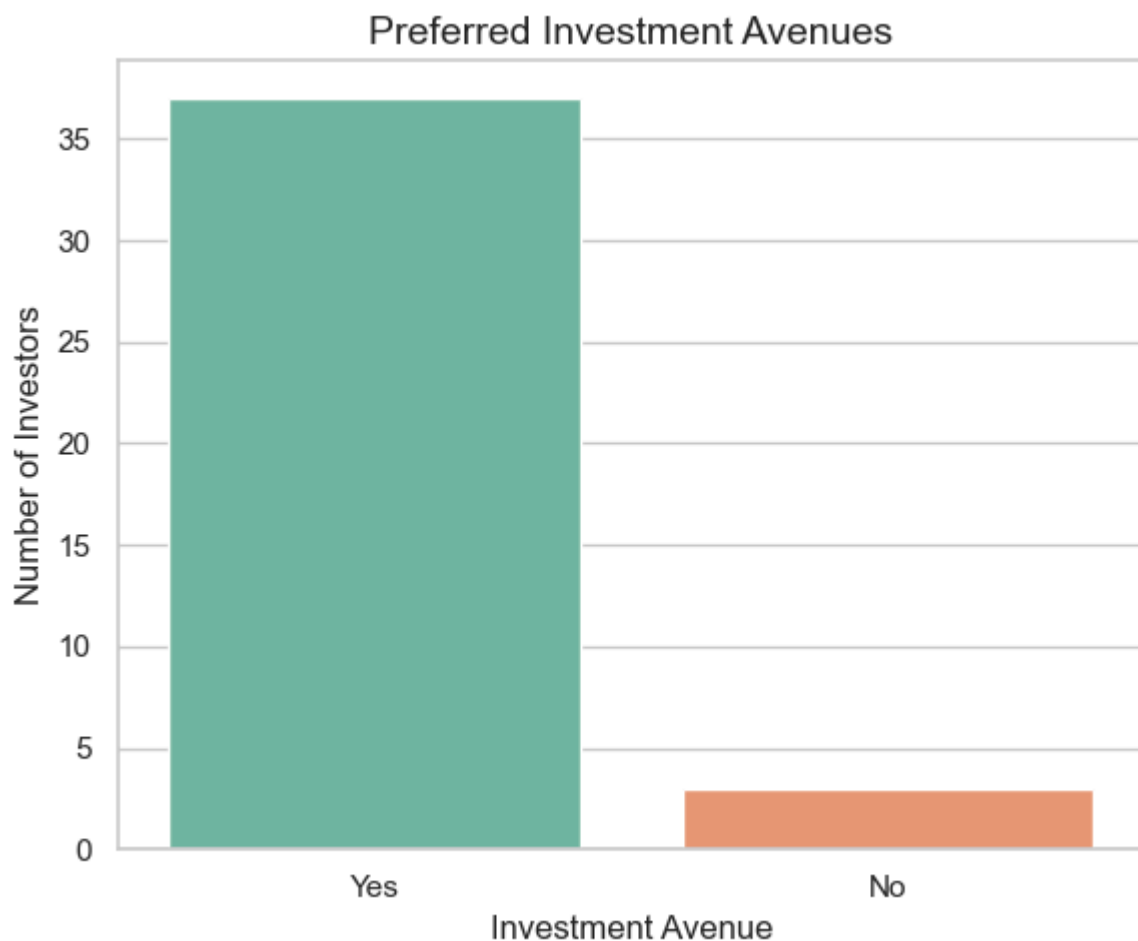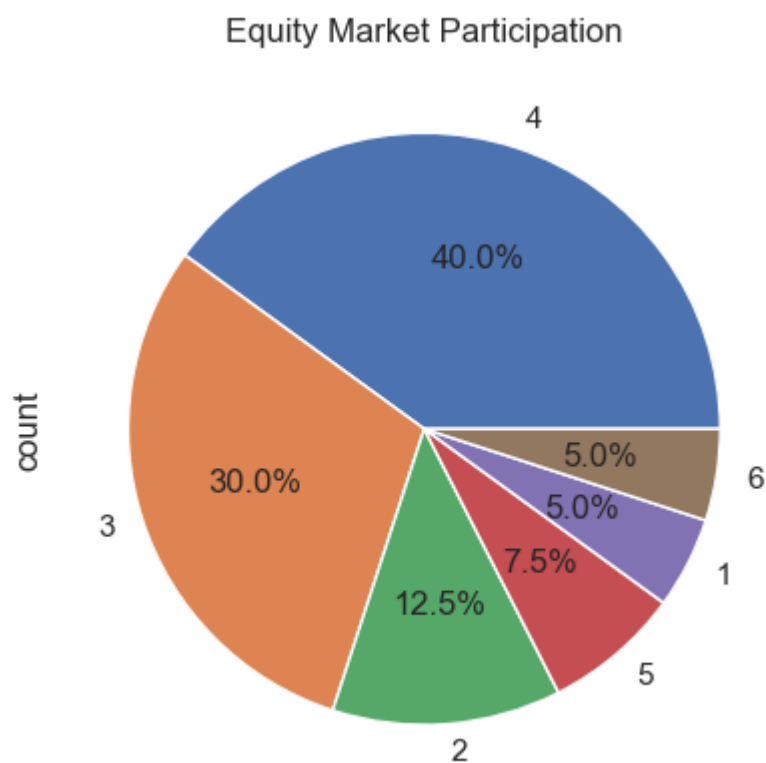
Caption: - Identifies the most commonly chosen investment instruments.

2 Equity Market Participation

```
In [55]: df["Equity_Market"].value_counts().plot(kind="pie", autopct="%1.1f%%")
         plt.title("Equity Market Participation")
         plt.show()
```


Caption: - Shows the proportion of investors participating in equity markets.

3 Investment Objective Distribution

```python
In [56]: plt.figure(figsize=(10, 5))

sns.countplot(
    data=df,
    y="Objective",
    hue="Objective",
    palette="Set3",
    legend=False
)

plt.title("Investment Objectives Distribution", fontsize=14)
plt.xlabel("Number of Investors")
plt.ylabel("Objective")
plt.tight_layout()
plt.show()
```
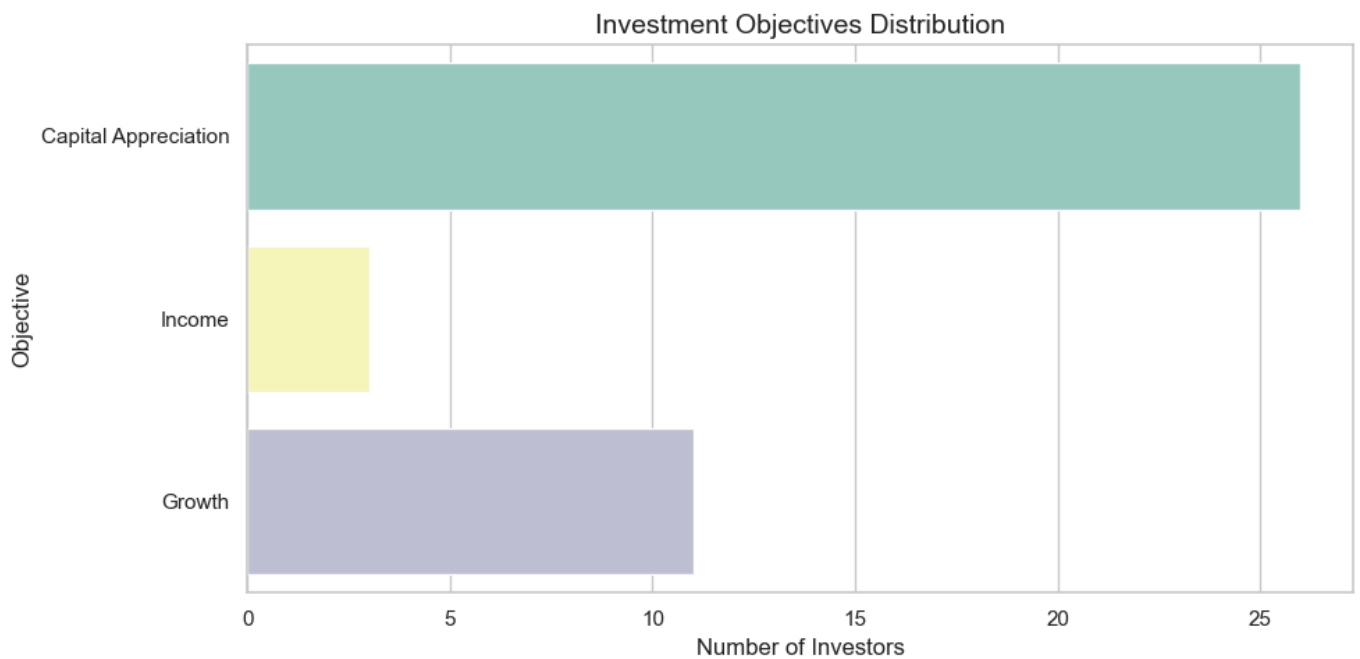

Caption: - Highlights dominant investor objectives such as wealth creation and savings.

4 Risk Factor vs Investment Avenue

```python
In [57]: plt.figure(figsize=(7, 5))

sns.countplot(
    data=df,
    x="Factor",
    hue="Investment_Avenues",
    palette="tab10"
)

plt.title("Risk Factor vs Investment Avenue", fontsize=14)
plt.xlabel("Risk Factor")
plt.ylabel("Count")
plt.legend(title="Investment Avenue")
plt.tight_layout()
plt.show()
```
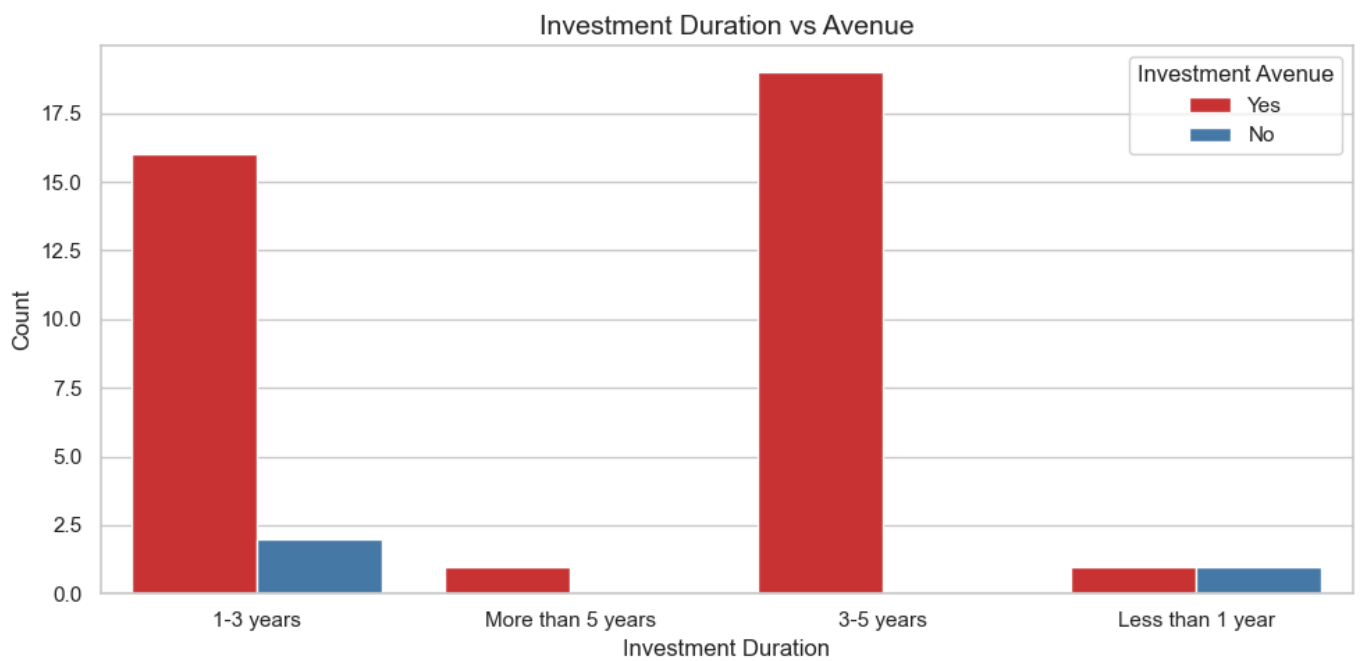
Risk Factor vs Investment Avenue
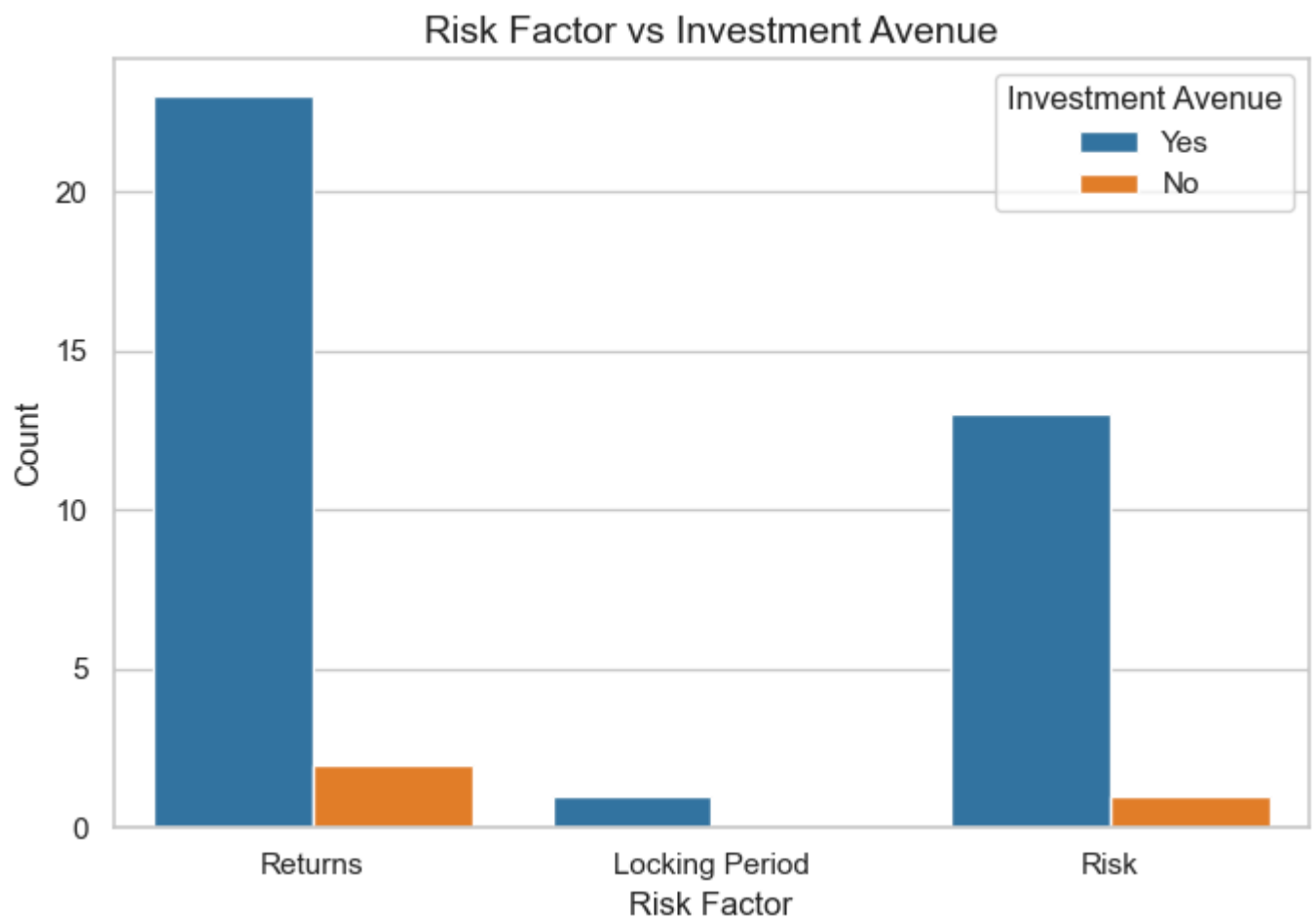
Caption: - Examines how risk tolerance influences investment choices.

5 Duration vs Investment Type

```python
In [58]:  plt.figure(figsize=(10, 5))

          sns.countplot(
              data=df,
              x="Duration",
              hue="Investment_Avenues",
              palette="Set1"
          )

          plt.title("Investment Duration vs Avenue", fontsize=14)
          plt.xlabel("Investment Duration")
          plt.ylabel("Count")
          plt.legend(title="Investment Avenue")
          plt.tight_layout()
          plt.show()
```

Caption: - Analyzes the relationship between investment horizon and instrument selection.

6 Age vs Investment Avenue (Behavioral Pattern)

```
In [59]:  plt.figure(figsize=(10, 6))
          sns.countplot(
              data=df,
              x="AGE_GROUP",
              hue="Investment_Avenues",
              palette="Set2"
          )
          plt.title("Investment Preference by Age Group")
          plt.xlabel("Age Group")
          plt.ylabel("Number of Investors")
          plt.legend(title="Investment Avenue")
          plt.show()
```
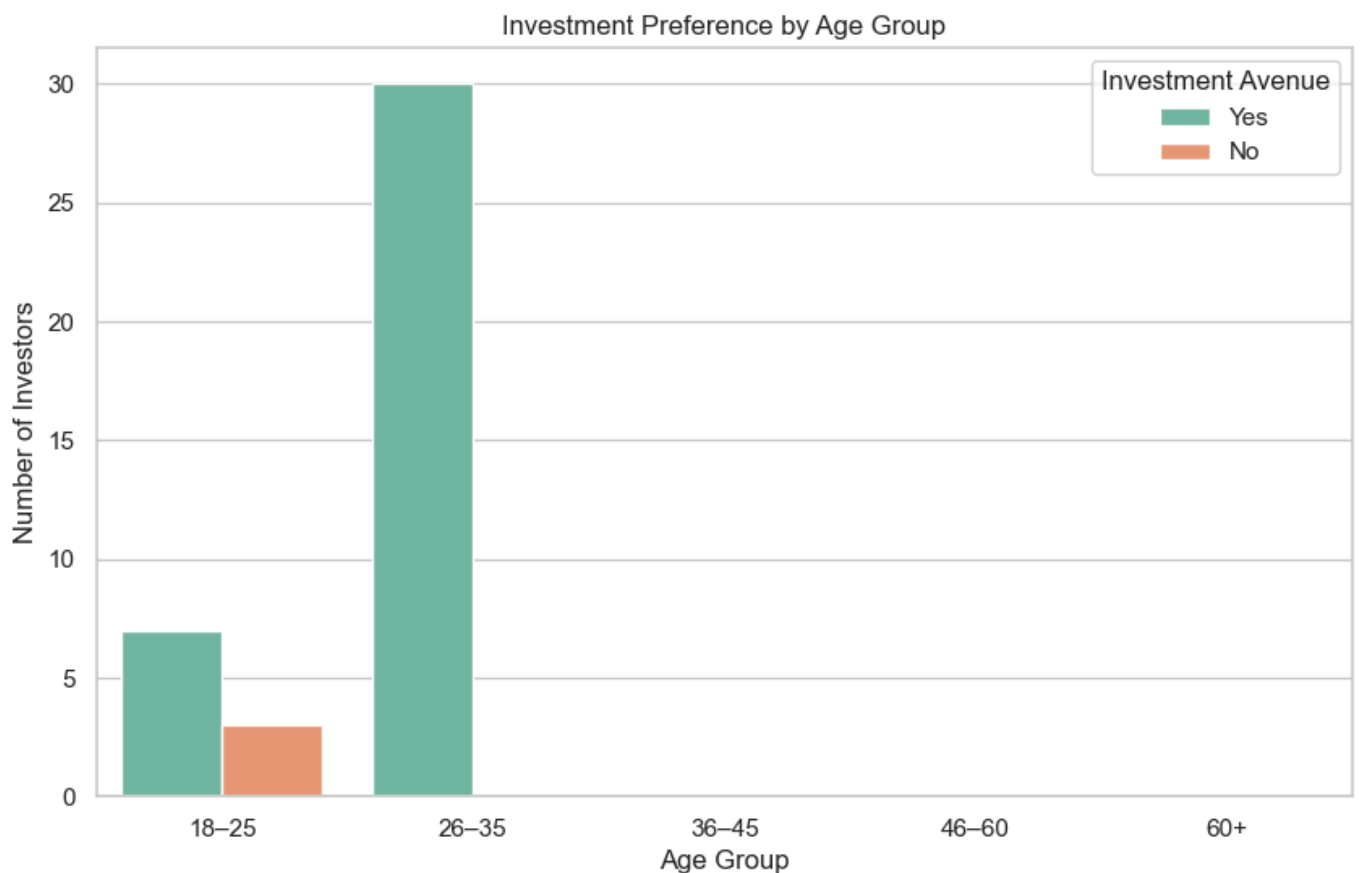
Caption: - Reveals how investment preferences vary across age groups.

7 Savings Objective vs Investment Avenue (Decision Driver)

```
In [60]: plt.figure(figsize=(12, 6))

         sns.countplot(
             data=df,
             x="What are your savings objectives?",
             hue="Investment_Avenues",
             palette="tab10"
         )

         plt.title("Savings Objectives vs Investment Avenue", fontsize=14)
         plt.xlabel("Savings Objective")
         plt.ylabel("Count")
         plt.xticks(rotation=40, ha="right")
         plt.legend(title="Investment Avenue")
         plt.tight_layout()
         plt.show()
```


Caption: - Explores how savings goals drive financial decisions.

8 Reasons for Equity vs Mutual Funds (Motivation Analysis)

```
In [61]: fig, axes = plt.subplots(1, 2, figsize=(14, 5))

         # Equity Reasons
         sns.countplot(
             data=df,
             y="Reason_Equity",
             hue="Reason_Equity",      # REQUIRED for palette
             palette="Blues",
             legend=False,
             ax=axes[0]
         )
         axes[0].set_title("Reasons for Investing in Equity", fontsize=13)
         axes[0].set_xlabel("Count")
         axes[0].set_ylabel("Reason")

         # Mutual Fund Reasons
         sns.countplot(
```
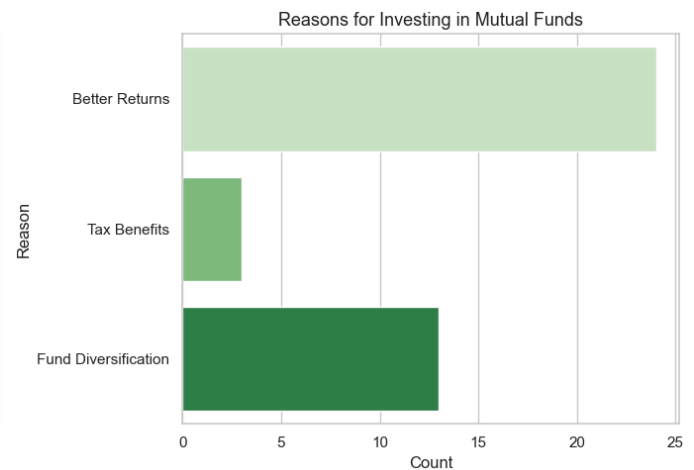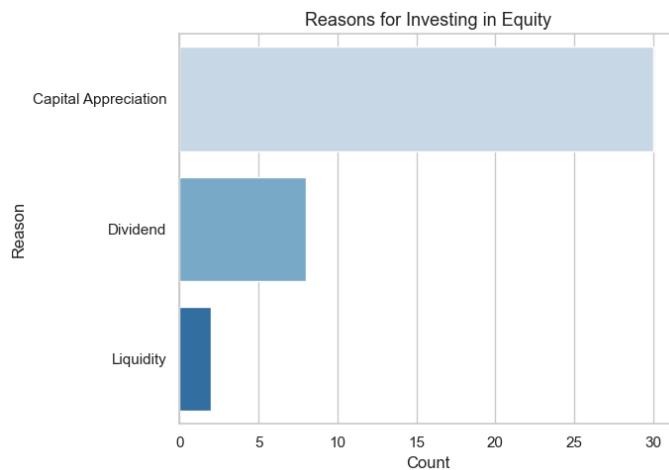
```
    data=df,
    y="Reason_Mutual",
    hue="Reason_Mutual",        # REQUIRED for palette
    palette="Greens",
    legend=False,
    ax=axes[1]
)
axes[1].set_title("Reasons for Investing in Mutual Funds", fontsize=13)
axes[1].set_xlabel("Count")
axes[1].set_ylabel("Reason")

plt.tight_layout()
plt.show()
```


Caption: - Provides insight into motivational factors behind investment choices.

9 **Investment Monitoring vs Avenue (Behavior Maturity)**

```
In [62]:  plt.figure(figsize=(10, 5))
          sns.countplot(
              data=df,
              x="Invest_Monitor",
              hue="Investment_Avenues",
              palette="Set3"
          )
          plt.title("Investment Monitoring Frequency vs Avenue")
          plt.xlabel("Investment Monitoring Frequency")
          plt.ylabel("Count")
          plt.show()
```
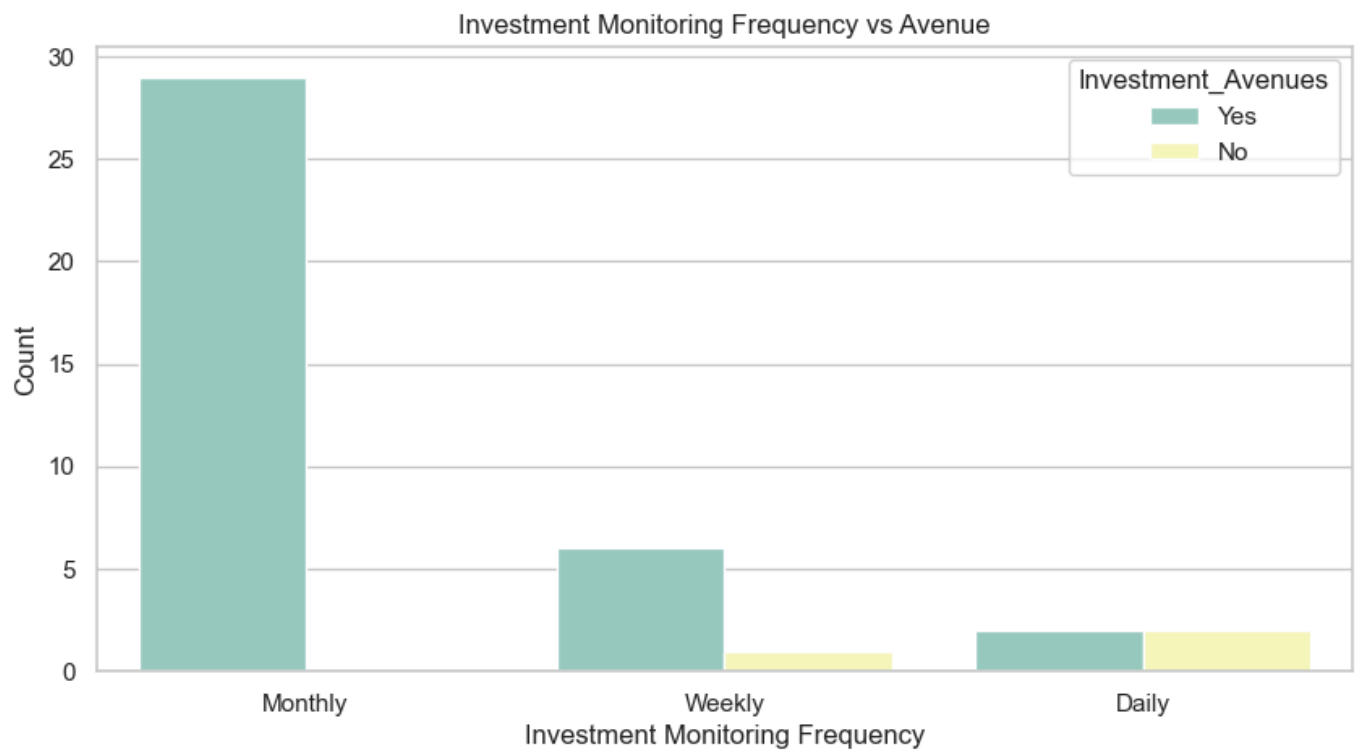
Caption: - Assesses investor maturity through portfolio monitoring behavior.

🔟 Correlation Heatmap (Binary Finance Behavior)

```
In [63]: binary_cols = [
             "Mutual_Funds", "Equity_Market", "Debentures",
             "Government_Bonds", "Fixed_Deposits", "PPF", "Gold"
         ]

         corr = df[binary_cols].corr()

         sns.clustermap(
             corr,
             cmap="vlag",
             center=0,
             annot=True,
             fmt=".2f",
             linewidths=0.5,
             figsize=(8, 7),
             dendrogram_ratio=(.15, .15),
             cbar_kws={"label": "Correlation Strength"}
         )

         plt.suptitle(
             "Clustered Correlation of Investment Instruments",
             fontsize=14,
             y=1.02
         )
         plt.tight_layout()
         plt.show()
```

Clustered Correlation of Investment Instruments
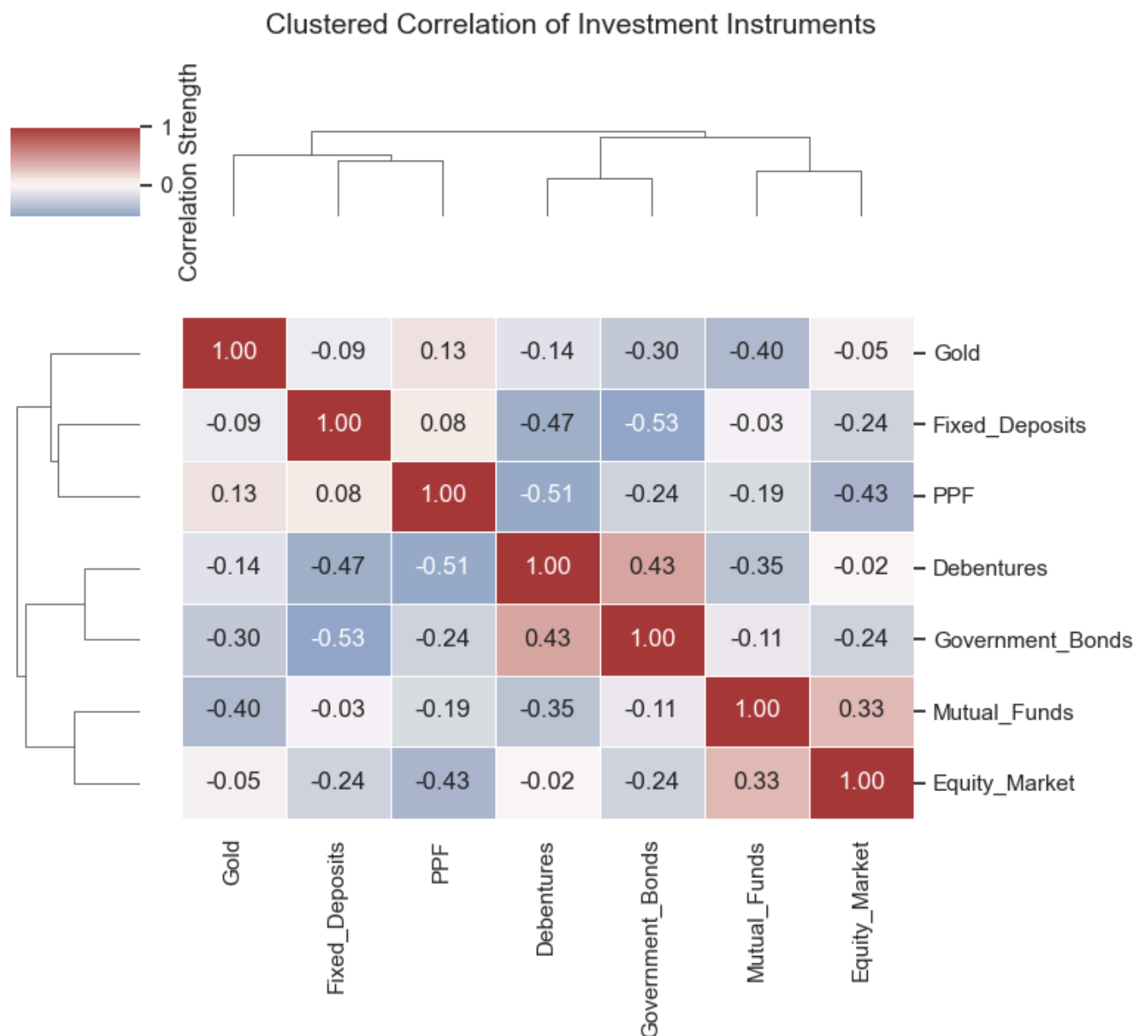
Caption: - Identifies relationships and diversification patterns among financial instruments.

🟦 Statistical Analysis

Key Metrics

- Total number of investors
- Equity participation rate
- Average investor age
- Equity participation by age group and gender
- Risk factor distribution

Interpretation

Statistical analysis indicates that equity participation is higher among younger and middle-aged investors. Long-term investment horizons and frequent monitoring are strongly associated with higher-risk instruments such as equities and mutual funds.

```
In [64]:  df.groupby("AGE_GROUP", observed=True)["EQUITY_INVESTOR"].mean()
```

```
Out[64]:    AGE_GROUP
            18-25    0.0
            26-35    0.0
            Name: EQUITY_INVESTOR, dtype: float64
```

```
In [65]:    df.groupby("gender")["EQUITY_INVESTOR"].mean()
```

```
Out[65]:    gender
            Female    0.0
            Male      0.0
            Name: EQUITY_INVESTOR, dtype: float64
```

```
In [66]:    pd.crosstab(df["Factor"], df["Investment_Avenues"])
```

Out[66]:

Investment_Avenues	No	Yes
Factor		
Locking Period	0	1
Returns	2	23
Risk	1	13

```
In [67]:    from src.finance_stock_market_analysis.visualization import *

            output_dir = Path("../visualizations/finance")

            plot_preferred_investment_avenues(df, output_dir)
            plot_equity_market_participation(df, output_dir)
            plot_investment_objective_distribution(df, output_dir)
            plot_risk_factor_vs_avenue(df, output_dir)
            plot_duration_vs_avenue(df, output_dir)
            plot_age_vs_avenue(df, output_dir)
            plot_savings_objective_vs_avenue(df, output_dir)
            plot_reasons_equity_vs_mutual(df, output_dir)
            plot_investment_monitoring_vs_avenue(df, output_dir)
            plot_clustered_correlation_heatmap(df, output_dir)

            print("✅ All the visualizations exported successfully")
```

✅ All the visualizations exported successfully

🟦 Key Findings

- Mutual funds and equities are the most preferred investment avenues
- Equity participation is dominant among middle-aged investors
- Long-term investment objectives align with higher-risk instruments
- Risk tolerance significantly influences avenue selection
- Active monitoring is linked with equity-oriented portfolios

---

🟦 Financial Insights

1. Younger and middle-aged investors demonstrate higher equity exposure
2. Long-term goals favor equity and mutual fund investments
3. Risk-aware investors diversify across multiple instruments
4. Monitoring frequency reflects financial maturity

🟦 Recommendations

1. Promote diversified portfolios across age groups
2. Encourage long-term, goal-based investment planning
3. Educate conservative investors on risk-adjusted returns
4. Support regular portfolio monitoring for better outcomes
5. Align financial products with investor risk profiles

🟦 Conclusion & Future Scope

Conclusion

This project demonstrates how investor behavior data can be transformed into meaningful financial intelligence. Through structured analysis and advanced visualizations, the study highlights key behavioral patterns that can support better financial decision-making and product design.

Future Scope

- Predictive modeling of investor behavior
- Risk-return optimization analysis
- Time-series analysis of investment evolution
- Integration with real-time market data

End of Report