LACHOO MEMORIAL COLLEGE OF SCIENCE & TECHNOLOGY, JODHPUR

Affiliated To:-

JAI NARAIN VYAS UNIVERSITY, JODHPUR

2025-2026

Project Title: Investment Behavior Analysis Dashboard

Team Members : Shivrani Kanwar (243501195)
Ranvijay Singh Bhati (243501178)
Rahul Hudda (243501167)

Course Code : CCBCA321

Course Name : BCA-Python Programming

Semester/year : Semester III -2025

Submitted to : Ms. Sanjana Khatri

Abstract

This project analyzes a survey dataset of 1,120 respondents to understand investment objectives, preferred investment avenues, information sources, and demographic patterns that influence investment preferences. We perform end-to-end data processing — acquisition, cleaning, transformation, visualization, and interpretation — to produce actionable insights for financial product teams and retail-investor educators. Key findings indicate that *Capital Appreciation* and *Wealth Creation* are dominant objectives, with equities and mutual funds scoring high on preference when investors value returns. Traditional media (newspapers, financial consultants) remain important information sources, though the internet is emerging. Gender differences are visible: males show higher counts in many investment categories while females exhibit distinct trust patterns (e.g., financial consultants). The analysis uses Python (Pandas, NumPy), and visualization tools (Matplotlib, Seaborn for exploratory figures; Power BI for polished dashboard). We present recommendations to tailor financial products, improve communication channels, and target segments by age and gender.

Introduction

Background & Motivation

India's retail-investor landscape is rapidly evolving: more individuals are exploring formal investment avenues (equity markets, mutual funds, government bonds) for long-term wealth creation. Understanding *why* and *how* people invest — their objectives, information sources, and behavioral patterns — helps financial institutions design better products and communications.

2. Problem Statement

What drives an individual to choose a particular investment avenue? How do demographic variables (gender, age) and information sources influence investment preferences and frequency? Which products score highest on preference scores (average rating) and which objectives dominate across the population?

3. Objectives of the Project

- Profile investor objectives (capital appreciation, growth, income).
- Rank and compare investment avenues by average preference scores and by demographic slices.
- Identify the most trusted information sources and their impact on investment choices.
- Provide recommendations based on observed patterns.

Dataset Description

Dataset & Data Acquisition

The dataset is a structured CSV survey file with 1,120 observations and fields including:

- respondent_id — unique identifier
- gender — Male / Female
- age — age in years
- objective — investment objective (Capital Appreciation, Growth, Income)
- purpose — purpose of saving (Wealth Creation, Returns, Savings for Future)
- investment_type — avenue chosen (Equity, Mutual Fund, Fixed Deposits, Bonds, PPF, etc.)
- preference_score — numeric rating (1–7) of preference for each avenue
- information_source — source of investment information (Financial Consultants, Newspapers, Television, Internet)
- investment_frequency — Daily/Weekly/Monthly
- duration — investment duration category
- other survey metadata (location, education, etc.)

Data was imported into Python using pandas.read_csv().

Methodology (Python Data Cleaning)

Overview

We followed a standard data-science pipeline:

- Data loading and initial inspection
- Data cleaning (missing values, duplicates, inconsistent categories)
- Feature transformation and encoding
- Exploratory Data Analysis (EDA) with plots
- Aggregation and cross-tab analysis
- Dashboard creation in Power BI for stakeholder-ready visuals

Data Cleaning

```
import pandas as pd
import numpy as np

# Load dataset
df = pd.read_csv('investment_survey.csv')

# Quick look
df.info()
df.head()

# 1. Remove exact duplicates
df = df.drop_duplicates()

# 2. Handle missing values
# For categorical fields, replace obvious blanks with 'Unknown', otherwise fill with mode
cat_cols = ['gender', 'objective', 'information_source', 'investment_type']
for c in cat_cols:
df[c] = df[c].fillna('Unknown')
```

```
# For numeric fields, impute with median
num_cols = ['age', 'preference_score']
for c in num_cols:
df[c] = df[c].fillna(df[c].median())

# 3. Fix inconsistent categories (strip whitespace, unify capitalization)
for c in ['gender','objective','information_source','investment_type']:
df[c] = df[c].astype(str).str.strip().str.title()

# 4. Detect and handle outliers in 'age' and 'preference_score'
# Cap ages at 99 and floor at 18 (based on survey scope)
df['age'] = df['age'].clip(lower=18, upper=99)

# 5. Create derived features
# Example: Age bucket
bins = [17,25,35,50,99]
labels = ['18-25','26-35','36-50','50+']
df['age_group'] = pd.cut(df['age'], bins=bins, labels=labels)

# 6. Encode categorical variables where needed (for modeling)
df_encoded = pd.get_dummies(df, columns=['gender','investment_type','information_source'],
drop_first=True)

# Save cleaned version
df.to_csv 'investment_survey_cleaned.csv', index=False)
```

Tools & Libraries

- **Python** (Pandas, NumPy) for cleaning and aggregation
- **Matplotlib / Seaborn** for quick EDA
- **Power BI** for final dashboard and stakeholder-ready visualization

Results & Analysis

The following key results were produced from the cleaned dataset and visualized in the Power BI dashboard shown in the appendix screenshots.

1. Descriptive Statistics

- **Total respondents:** 1,120
- **Average age:** 27.8 years
- **Gender split:** 700 Male (62.5%), 420 Female (37.5%)
- **Investment frequency:** Majority Monthly (812 respondents), followed by Weekly (196) and Daily (112)

2. Investment Objectives

- The most common objective is **Capital Appreciation** (728 respondents, 65%), followed by **Growth** (308, 27.5%) and **Income** (84, 7.5%).
- When segmented by whether respondents already use formal investment avenues, Capital Appreciation remains dominant — a sign of a growth-oriented investor base.

Interpretation: A majority of respondents aim to grow their wealth rather than seek immediate income — this suggests the market for growth-oriented financial products (equity mutual funds, direct equities) is strong.

3. Purpose of Saving

- **Wealth Creation** is the primary reason (896 responses), with **Savings for Future** and **Returns** trailing.

Interpretation: Financial education and long-horizon products (SIPs, retirement funds) are appropriate for targeting this audience.

4. Investment Preferences (Average Scores)

- **Top-scoring avenues** by average preference score are: Gold (~5.98), Debentures (~5.75), Government Bonds (~4.65), Fixed Deposits (~3.58), Equity Market (~3.48), Mutual Funds (~2.55), PPF (~2.03).

Interpretation: Traditional safe-haven assets (gold) and debt-like instruments (debentures, govt bonds) rate highly. Although equities are preferred for capital appreciation, their average preference score is lower — indicating perceived risk or knowledge gap.

5. Information Sources & Trust

- **Most common sources:** Financial Consultants (448, 40%), Newspapers & Magazines (392, 35%), Television (168, 15%), Internet (112, 10%).

- **Trusted sources by gender:** Males place significant trust in financial consultants and newspapers; females show notable trust in television and internet as well (treemap highlights this split).

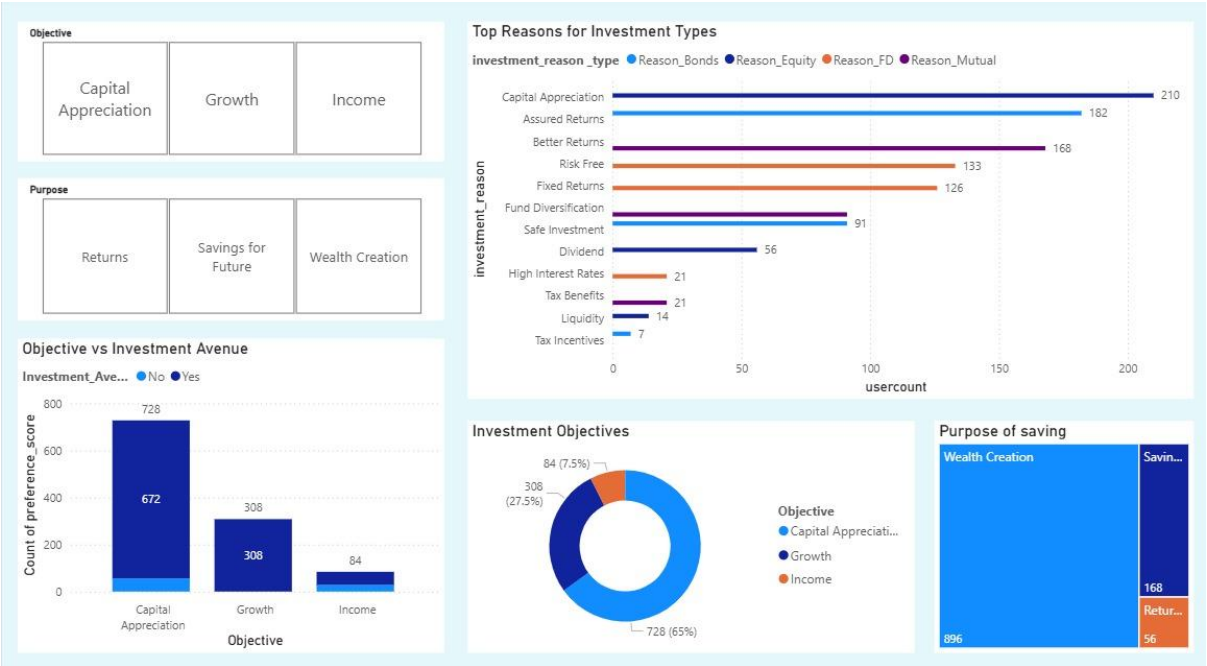
Interpretation: While digital channels are growing, traditional channels (consultants, newspapers) remain central. Outreach strategies should use hybrid channels: seminars and consultant partnerships coupled with targeted digital content.

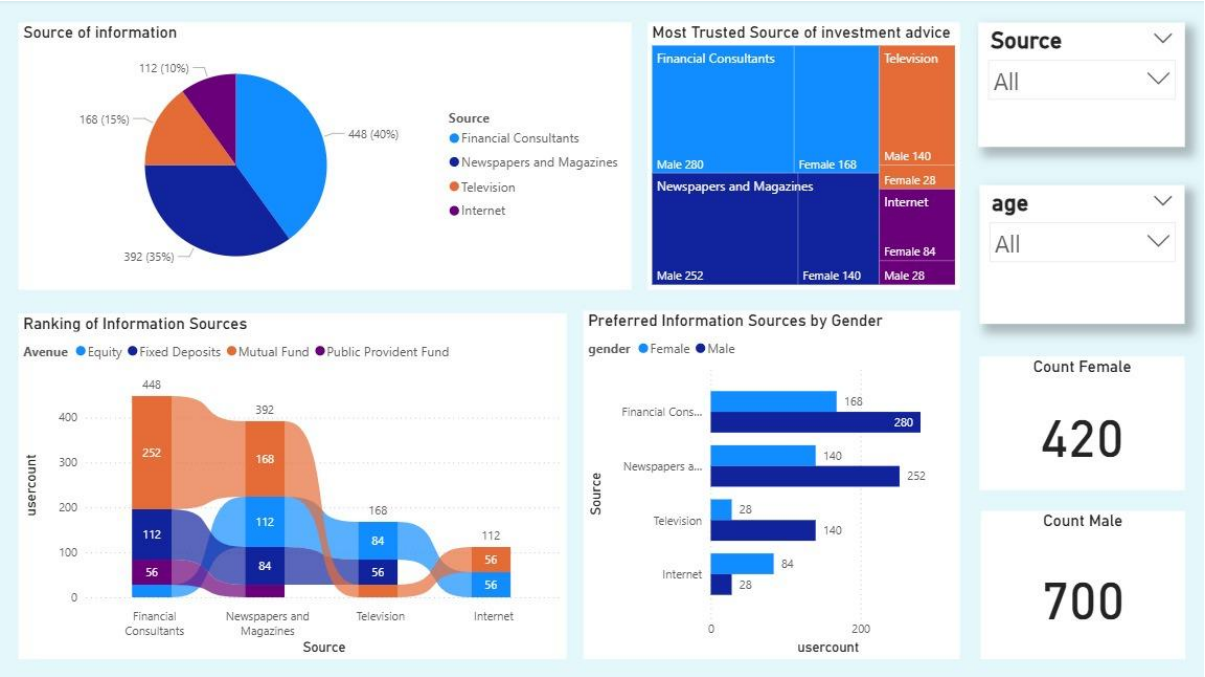
6. Cross-Sectional Findings

- **Gender vs. Investment Preference:** Males report slightly higher average preference scores across several avenues (e.g., Gold, Debentures) compared to females. However, in counts, males participate more in many categories likely due to higher absolute male respondents.
- **Age groups:** Younger respondents (18–25) lean toward equities and mutual funds for capital appreciation, while older groups favor fixed-income and gold.

Dashboard


Total respondents
1120
Average Age
27.80
Percent Investing
0.93
Invest_Monitor Female Male Total
Weekly 56 140 196
Monthly 308 504 812
Daily 56 56 112
Total 420 700 1120
Gender
All
Duration
All
Avenue
All
Investment Duration
56 (5%)
532 (47.5%)
504 (45%)
Duration
3-5 years
1-3 years
Less than 1 year
More than 5 years
Investment Preferences
5.98
5.75
4.65
3.58
3.48
2.55
2.03
Average of preference_score
Gold
Debentures
Government_Bonds
Fixed_Deposits
Equity_Market
Mutual_Funds
PPF
investment_type
Gender vs Investment Preference
gender Female Male
Gold 5.87 6.04 11.91
Debentures 5.20 6.08 11.28
Government_Bonds 4.33 4.84 9.17
Fixed_Deposits 4.20 3.20 7.40
Equity_Market 3.33 3.56 6.89
Mutual_Funds 2.73 2.44 5.17
investment_type
Average of preference_score


Objective
Capital Appreciation
Growth
Income
Purpose
Returns
Savings for Future
Wealth Creation
Top Reasons for Investment Types
investment_reason _type Reason_Bonds Reason_Equity Reason_FD Reason_Mutual
Capital Appreciation 210
Assured Returns 182
Better Returns 168
Risk Free 133
Fixed Returns 126
Fund Diversification 91
Safe Investment
Dividend 56
High Interest Rates 21
Tax Benefits 21
Liquidity 14
Tax Incentives 7
investment_reason
usercount
Objective vs Investment Avenue
Investment_Ave... No Yes
728
672
308
308
84
Count of preference_score
Capital Appreciation
Growth
Income
Objective
Investment Objectives
84 (7.5%)
308 (27.5%)
728 (65%)
Objective
Capital Appreciati...
Growth
Income
Purpose of saving
Wealth Creation
896
Savin...
168
Retur...
56



Key Insights

1. **Target Growth Products to Young Investors:** With capital appreciation and wealth creation top objectives, target SIP campaigns and educational content to the 18–35 age group.
2. **Hybrid Communication Strategy:** Maintain consultant and print presence while boosting trustworthy digital content — FAQs, explainer videos, and calculators.
3. **Risk Education Needed for Equities:** Although many aim for capital appreciation, preference scores show hesitation toward equities — produce risk-mitigation educational programs.
4. **Segmented Product Bundles:** Create bundled products with a mix of gold/debt/equity to match the average investor preference mix.
5. **Gender-Sensitive Messaging:** Tailor messaging channels and content tone (e.g., more visual/digital outreach for female segments that show higher engagement with TV/Internet sources).

Conclusion & Future Scope

This analysis highlights a market leaning toward capital appreciation and wealth creation, with strong roles for consultants and traditional media as trusted information sources. Future work could:

- Build predictive models to recommend the optimal product mix per individual (using classification/regression models on preference scores).
- Run A/B tests on messaging across channels to measure conversion (sign-up, SIP starts).
- Enrich the dataset with income, education, and location data for deeper segmentation.

References

- Survey dataset (internal)
- Relevant articles and financial product descriptions consulted during interpretation (list specific URLs or papers here as needed)

Appendix

GitHub Repository

https://github.com/Shivrani14/Python-Mini-Project