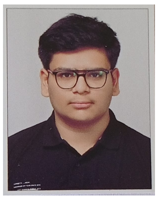
Exploratory Data Analysis on

FoF

by

Group 43



Jenish Vasani*
ID: 202301057
Course: BTech(ICT)

Prince Sojitra
ID: 202301126
Course: BTech(ICT)

Siddharth Vala
ID: 202301180
Course: BTech(ICT)

Course Code: IT 462
Semester: Autumn 2025

Under the guidance of

Dr. Gopinath Panda

Dhirubhai Ambani University

Acknowledgment

I am writing this letter to express my heartfelt gratitude for your guidance and support throughout the duration of our project titled "Exploratory Data Analysis on Fund of Funds." Your invaluable assistance, insightful feedback, and constant encouragement have played a crucial role in the successful completion of this work.

I am extremely grateful for your expertise and the time you dedicated to mentoring us. Your constructive suggestions and clear direction greatly enhanced our understanding of the subject and helped us refine the overall quality of our project.

Furthermore, I would like to extend my appreciation to Dhirubhai Ambani University (DAU) for providing the academic environment, resources, and support essential for completing this project. The opportunities and facilities offered by the institution contributed significantly to our research and analysis.

I would also like to express my gratitude to my peers and colleagues who have been supportive throughout this journey. Their valuable input and camaraderie have been a constant source of motivation.

Completing this project has been a tremendous learning experience, and we are confident that the knowledge and skills acquired during this endeavor will serve as a solid foundation for our future academic and professional pursuits.

Once again, thank you for your continuous guidance, support, and encouragement throughout this project.

Sincerely,
Jenish Vasani (202301057)
Prince Sojitra (202301126)
Siddharth Vala (202301180).

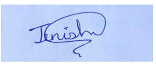
DECLARATION

We, Jenish Vasani, Prince Sojitra and Siddharth Vala, hereby declare that the EDA project work presented in this report is our original work and has not been submitted for any other academic degree. All the sources cited in this report have been appropriately referenced.

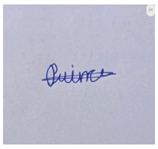
We acknowledge that the data used in this project is obtained from the advisorkhoj site. We also declare that we have adhered to the terms and conditions mentioned in the website for using the dataset. We confirm that the dataset used in this project is true and accurate to the best of our knowledge.

We acknowledge that we have received no external help or assistance in conducting this project, except for the guidance provided by our mentor Prof. Gopinath Panda. We declare that there is no conflict of interest in conducting this EDA project.

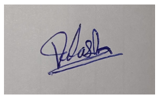
We hereby sign the declaration statement and confirm the submission of this report on 2nd July, 2023.



Jenish Vasani*
ID: 202301057
Course: BTech(ICT)



Prince Sojitra
ID: 202301126
Course: BTech(ICT)



Siddharth Vala
ID: 202301180
Course: BTech(ICT)

CERTIFICATE

This is to certify that Group 43 comprising Jenish Vasani (202301057) Prince Sojitra (202301126) and Siddharth Vala (202301180) has successfully completed an exploratory data analysis (EDA) project on the Fund of Funds, which was obtained from advisorkhoj.

The EDA project presented by Group 43 is their original work and has been completed under the guidance of the course instructor, Prof. Gopinath Panda, who has provided support and guidance throughout the project. The project is based on a thorough analysis of the Nav dataset, and the results presented in the report are based on the data obtained from the dataset.

This certificate is issued to recognize the successful completion of the EDA project on the Fund of Funds, which demonstrates the analytical skills and knowledge of the students of Group 43 in the field of data analysis.

Signed,
Dr. Gopinath Panda,
IT 462 Course Instructor
Dhirubhai Ambani Institute of Information and Communication Technology
Gandhinagar, Gujarat, INDIA.

Contents

List of Figures 5

1 Introduction 1
- 1.1 Your Project idea . . . 1
- 1.2 Data Collection . . . 1
- 1.3 Dataset Description . . . 2
- 1.4 Packages Required . . . 4

2 Data Cleaning 6
- 2.1 Missing Data Analysis . . . 6
  - 2.1.1 Identifying Missing Values . . . 6
  - 2.1.2 Missing Data in `statistics_summary.csv` . . . 8
  - 2.1.3 Missing Data in `stability_scores.csv` . . . 8
  - 2.1.4 Missing Data in `risk_with_meta.csv` and `risk_metrics_classified.csv` . 8
  - 2.1.5 Missing Data in `returns_correlation.csv` . . . 9
  - 2.1.6 Missing Data in `group_summary_tidy.csv` and `group_stats.csv` . . . 9
  - 2.1.7 Missing Data in `group_nav.csv` . . . 9
  - 2.1.8 Summary of Missing Data Strategy . . . 9
- 2.2 Imputation . . . 10
  - 2.2.1 Why Imputation Was Needed . . . 10
  - 2.2.2 Columns Where Imputation *Was* Applied . . . 10
  - 2.2.3 Columns Where Imputation *Was Not* Performed . . . 11
  - 2.2.4 Reason for Minimal Imputation . . . 11
  - 2.2.5 Additional Computation Notes . . . 12
  - 2.2.6 Final Imputation Summary . . . 12

3 Visualization 13
- 3.1 Univariate Analysis . . . 13
  - 3.1.1 Return Distribution of Individual Funds . . . 13
  - 3.1.2 Rolling Volatility (1-year) . . . 14
  - 3.1.3 Rolling Sharpe-like Ratio (1-year) . . . 14
  - 3.1.4 Rolling Annualized Return (1-year) . . . 15
  - 3.1.5 Volatility Distribution by Fund Type . . . 15
- 3.2 Multivariate Analysis . . . 17
  - 3.2.1 Correlation Heatmap of Returns . . . 17
  - 3.2.2 Sharpe-like Comparison by Fund Type . . . 18

3.2.3 Annualized Volatility by Fund Type . . . 18
3.2.4 Group-Level NAV Growth Over Time . . . 19
3.2.5 Distribution of Pairwise Correlations . . . 19
3.3 Bivariate Analysis . . . 20
3.3.1 NAV vs Volatility . . . 20
3.3.2 Sharpe-like Ratio vs Fund Type . . . 20
3.3.3 Annual Return vs Volatility . . . 21
3.3.4 Indexed NAV Comparison (Domestic vs Overseas) . . . 21
3.3.5 Rolling Sharpe-like vs Rolling Volatility . . . 21
3.3.6 Correlation Between Risk Metrics . . . 22

4 Feature Engineering 23
4.1 Feature Extraction . . . 23
4.1.1 Return-Based Features . . . 23
4.1.2 Risk-Based Features . . . 23
4.1.3 Time-Window and Rolling Features . . . 25
4.1.4 Allocation and Fund-Level Features . . . 25
4.1.5 Summary . . . 27
4.2 Feature Selection . . . 27
4.2.1 Removing Irrelevant or Redundant Features . . . 27
4.2.2 Correlation-Based Filtering . . . 27
4.2.3 Selecting Key Performance Indicators . . . 30
4.2.4 Choosing Stable Fund-Level Attributes . . . 30
4.2.5 Final Feature Set . . . 30

5 Model Fitting 32
5.1 Objective of the Model . . . 32
5.2 Data Processing and Preparation . . . 32
5.2.1 Handling Missing NAV Values . . . 32
5.2.2 Scaling the NAV . . . 32
5.2.3 Creating Sliding Windows . . . 33
5.3 Model Selection . . . 33
5.4 Working of the LSTM Model . . . 33
5.5 Model Architecture . . . 33
5.6 Training Procedure . . . 34
5.7 Predicting the Next 14 Days NAV . . . 34
5.8 Model Evaluation . . . 34
5.9 Interpretation of Predictions . . . 35
5.10 Summary . . . 35

6 Conclusion & Future Scope 36
6.1 Conclusion . . . 36
6.2 Key Findings . . . 36
6.3 Challenges . . . 37
6.4 Future Scope . . . 37

List of Figures

1.1 Enter Caption . . . 3

2.1 Missing Data . . . 7

3.1 Return Distribution — ABSL Multi Asset Omni FoF Dir Plan Gr . . . 13
3.2 Rolling Volatility (1-year) — Top 3 Funds . . . 14
3.3 Rolling Sharpe-like (1-year) — Top 3 Funds . . . 15
3.4 Rolling Annualized Return (1-year) . . . 16
3.5 Volatility Distribution — Domestic vs Overseas . . . 16
3.6 Correlation Heatmap of Returns . . . 17
3.7 Sharpe-like by Fund Type . . . 18
3.8 Annualized Vol by Fund Type . . . 18
3.9 Average Group-Level Growth (Indexed to 1) . . . 19
3.10 Pairwise Correlation Distribution — Within vs Between Groups . . . 20
3.11 Sharpe by fund . . . 21

4.1 Return Distribution of a chosen FoF . . . 24
4.2 Rolling Volatility . . . 24
4.3 Rolling Annualized Return . . . 26
4.4 Indexed Group-Level NAV Growth . . . 26
4.5 Cleaned column list (screenshot). . . . 28
4.6 Metric-level correlation heatmap . . . 29

Abstract

In thisproject, we aim to study the behavior and performance of Funds of Funds (FoFs) with real-world financial data. We want to make senseof how these funds work by analyzing their daily NAV movements, returns and risk. Thedataset is technically cleaned and prepared in a rigorous way, to guarantee that interpretations of analysis are meaningful.

After my data was organized, I calculated meanand median performance of daily returns, annualized return, volatility, Sharpe ratio and max drawdown. A combination of graphs and visual aids were used to look fortrends and interrelationships in the data. These visualisations helped highlight how different FoFs behave over time, how stable their returns are, and how their risk levels compare with each other.

In summary, this analysis of provides a plain-spoken and actionable view on how FoFs actually perform,including the role such factors as diversification, expense ratio and market volatility have played in long-term performance. The shared knowledgefrom this study can help inform smarter investment decisions and serve as a foundation for further research based on prediction models, portfolio optimisation or more involved financial analysis.

Chapter 1. Introduction

1.1 Your Project idea

Funds of Funds (FoFs) are gaining popularity, as it offers diversification by way investing in various underlying mutual fund schemes instead of staking the entire amount intoone single fund. This structure allows investors to benefit from a professionally managed mix of equity, debt, and other asset categories while reducing the risk that comes from relying on one asset class alone. However, because FoFs behave differently from traditional mutual funds, it becomes important to study how they perform across different market situations and time periods.

The maingoal of this study is to investigate FoF properties through detailed analysis. Based on daily NAVs, return metrics, risk figures and allocation information the project tries to learn more about how FoFs behave in response to market shocks as well as how persistentis their performance and what factors drive long-term returns. UsingExploratory Data Analysis (EDA) we discover trends in factors involving return patterns, levels of volatility, patterns between funds and how diversification influences the performance.

With this research, we are seeking to gain an action-oriented picture of FoF behaviour with empirically collected financialdata. The goal is not only to compare different FoFs but also to understand what makes some funds more stable and consistent than others. The insights gained from this project can help investors, students, and analysts make better decisions and also form the foundation for more advanced work like forecasting, risk modelling, or portfolio optimisation.

1.2 Data Collection

The datasets for this project were collected from publicly available mutual fund sources. The information covers NAV movements, fund-level metrics and portfolio allocation details. The following sections describe each dataset clearly:

1. **NAV History Data**
   . Daily Net Asset Value (NAV) values for each FoF scheme.
   . Used for calculating returns, volatility, rolling metrics and drawdowns.
   . *Source:* AMC disclosures, MF API, public NAV databases.

2. **Fund Analytics Data**
   . Includes 1M, 3M, 6M, 1Y, 3Y returns.
   . Contains risk metrics such as standard deviation, Sharpe ratio, beta, alpha.
   . Includes expense ratio and AUM for each FoF.
   . *Source:* Fund analytics websites, AMC fact sheets.

3. **Portfolio / Allocation Data**
   . Shows allocation across equity, debt, hybrid and cash instruments.
   . Helps understand fund strategy and risk orientation.
   . *Source:* Monthly portfolio disclosures by AMCs.

4. **Scheme Metadata**
   . Scheme name, fund house (AMC), category and benchmark.
   . Launch date and classification details.
   . *Source:* Scheme documents and online mutual fund databases.

5. **Additional Attributes (If Available)**
   . Fund manager details, minimum investment amount.
   . Ratings, risk label and any other disclosed parameters.
   . *Source:* AMC websites, Moneycontrol, Value Research.

1.3 Dataset Description

This project uses several datasets related to Funds of Funds (FoFs). Each dataset provides important information needed for analysing returns, risk and fund structure.

NAV History Dataset

. Contains daily NAV values for every FoF scheme.
. Used to calculate daily returns, rolling returns, volatility and CAGR.
. Helps track long-term trends and market behaviour.
. Weekend and holiday NAV gaps are normal in mutual fund data.
Purpose:
. To analyse fund performance over time using NAV-based financial metrics.

Fund Analytics Dataset

. Includes short-term returns: 1M, 3M, 6M.
. Includes long-term returns: 1Y, 3Y, 5Y (if available).
. Contains risk metrics like standard deviation, Sharpe ratio and beta.
. Provides AUM and expense ratio for every scheme.
Purpose:
. To compare FoFs based on return behaviour and risk characteristics.

Portfolio / Allocation Dataset

. Shows asset allocation across equity, debt, hybrid, international and cash.
. Helps identify fund style: aggressive, conservative or balanced.
. Explains differences in volatility and return performance.
. Some schemes may have partially reported allocations.

Date	NAV	Fund	Fund_Type
31-10-2025 00:00	175.6564	ICICI Pru P	Domestic
30-10-2025 00:00	176.464	ICICI Pru P	Domestic
29-10-2025 00:00	177.2552	ICICI Pru P	Domestic
28-10-2025 00:00	176.1939	ICICI Pru P	Domestic
27-10-2025 00:00	176.5302	ICICI Pru P	Domestic
24-10-2025 00:00	175.3092	ICICI Pru P	Domestic
23-10-2025 00:00	176.0531	ICICI Pru P	Domestic
20-10-2025 00:00	175.5119	ICICI Pru P	Domestic
17-10-2025 00:00	174.4499	ICICI Pru P	Domestic
16-10-2025 00:00	173.9398	ICICI Pru P	Domestic
15-10-2025 00:00	172.5084	ICICI Pru P	Domestic
14-10-2025 00:00	171.5809	ICICI Pru P	Domestic
13-10-2025 00:00	172.1791	ICICI Pru P	Domestic
10-10-2025 00:00	172.5449	ICICI Pru P	Domestic
09-10-2025 00:00	171.6747	ICICI Pru P	Domestic
08-10-2025 00:00	170.7133	ICICI Pru P	Domestic
07-10-2025 00:00	171.187	ICICI Pru P	Domestic
06-10-2025 00:00	171.0649	ICICI Pru P	Domestic
03-10-2025 00:00	169.8015	ICICI Pru P	Domestic
01-10-2025 00:00	169.0957	ICICI Pru P	Domestic
30-09-2025 00:00	167.8878	ICICI Pru P	Domestic
29-09-2025 00:00	167.561	ICICI Pru P	Domestic
26-09-2025 00:00	167.2394	ICICI Pru P	Domestic
25-09-2025 00:00	169.3313	ICICI Pru P	Domestic
24-09-2025 00:00	170.2175	ICICI Pru P	Domestic

Figure 1.1: Enter Caption

Purpose:
. To understand how investment allocation affects fund behaviour.

Scheme Metadata

. Includes scheme name, AMC name and FoF category.
. Provides benchmark index and launch date.
. Helps classify and group similar FoFs.
Purpose:
. To organise funds properly and support comparison.

Additional Attributes (If Available)

. Minimum investment amount and SEBI risk label.
. Fund manager details and AMC information.

. Ratings or grades from financial research platforms.
Purpose:
. To provide extra context when evaluating fund suitability.

Return Data Summary

. Long-term returns: 1Y, 3Y, 5Y.
. Engineered returns: daily return, rolling return, CAGR.
Purpose:
. To analyse performance across different time horizons.

Important Characteristics of the Dataset

. Real financial data with natural behaviour and volatility.
. Includes both raw values (NAV) and computed metrics (returns, risk).
. Contains unequal lengths of history due to fund launch dates.
. Supports risk analysis, performance comparison and regression.
Purpose:
. To build a reliable base for EDA, modelling and interpretation.

1.4 Packages Required

Before starting the analysis, several Python packages were used to read the data, perform numerical calculations, clean the dataset and create different visualisations. These packages simplify tasks such as handling dates, computing returns, generating charts and building regression models. The required packages along with their short descriptions are listed below.

1. **numpy**

. Used for numerical operations, array handling, mathematical formulas and return calculations.

2. **pandas**

. Helps in reading CSV files, cleaning datasets, dealing with missing values and creating dataframes for analysis.

3. **matplotlib**

. Used to plot basic visualisations such as line charts, bar graphs and scatter plots.

4. **seaborn**

. Provides advanced and attractive visualisations like heatmaps, boxplots and distribution plots.

5. **scipy**

. Supports statistical operations and mathematical functions commonly used in financial analysis.

6. **scikit-learn (sklearn)**
. Used for machine learning tasks such as regression, PCA, feature scaling and model evaluation.

7. **datetime**
. Helps in converting date columns, formatting dates and managing time-based data series.

8. **warnings**
. Used to suppress unnecessary warning messages during program execution, keeping the output clean.

9. **statsmodels** (optional)
. Provides detailed statistical modelling tools, including regression summaries and hypothesis testing.

10. **yfinance / mfapi** (optional)
. Used only if external NAV or market data needs to be downloaded directly through APIs.

Chapter 2. Data Cleaning

Data cleaning is the process of preparing the dataset to ensure it is accurate, consistent and ready for analysis. In this project, data cleaning involved checking all three datasets for missing values, removing completely empty or invalid columns, fixing incorrect data types and standardizing column names. Missing NAV values caused by weekends or newly launched funds were kept because they naturally occur in financial time-series data. Allocation and analytics datasets were cleaned by converting symbols into numeric values and replacing text placeholders with proper NaN. No artificial values were introduced to preserve the originality of financial behaviour. After cleaning, the datasets became structured, reliable and suitable for further feature extraction and analysis.

2.1 Missing Data Analysis

The FoF (Fund of Funds) project uses several derived datasets generated from NAV time-series, rolling-window statistics, risk metrics, correlation matrices and grouped summaries. Since each dataset was created through different processing steps, the presence and nature of missing values vary across files. This section identifies the sources of missing data in the project and explains how these gaps were handled to maintain accuracy and consistency.

2.1.1 Identifying Missing Values

Before performing any cleaning, each dataset was inspected to understand the pattern and extent of missing data. The following checks were performed:

- Detection of missing values using `isnull().sum()`.
- Structural inspection of each dataset using `info()`.
- Manual scanning for unusual placeholders such as empty strings or special characters.
- Checking rolling-window outputs, which naturally generate missing values in early rows.
- Verifying correlation matrices and grouped summary tables for structural zeros or incomplete entries.

Findings:

- No placeholder strings like "–" or "N.A" were found.
- Rolling-window computations produced limited missing rows at the beginning of the series.
- Group-level pivot tables showed zero counts due to structure, not missing values.

Fund	count
Axis Gold Fund Reg Gr	2450
ICICI Pru Passive Strategy Fund FOF Dir Gr	2446
HDFC Gold ETF FoF Dir	2446
Nippon India Gold Savings Dir Gr	2446
Franklin India Aggressive Hybrid Fund Dir Gr	2445
Kotak Multi Asset Omni FOF Dir Gr Dir	2443
PGIM India Global Equity Opportunities FOF Dir Gr	2420
Sundaram Global Brand Theme Equity Active FOF Dir Gr	2409
Edelweiss Europe Dynamic Equity Offshore Fund Gr Dir	2406
Bandhan Asset Allocation Fund Mod Reg Gr	2406
HSBC Multi Asset Active FOF Gr Dir	2402
ABSL Multi Asset Omni FoF Dir Plan Gr	2401
ABSL Income Plus Arbitrage Active FOF Dir Plan Gr	2401
Franklin India Dynamic Asset Allocation Active FOF Dir Gr	2400
Kotak Global Emerging Market Overseas Equity Omni FOF Gr	2380
Nippon India US Equity Opportunities Fund Dir Gr Gr	2374
Edelweiss Emerging Mkts Opp Equity Offshore Fund Dir Gr	2368
DSP World Gold Mining Overseas Equity Omni FoF Dir Plan Gr	2326
Invesco India Invesco Global Equity Income FOF Dir Gr	2306
Edelweiss Greater China Equity Off Shore Fund Dir Gr	2293

Figure 2.1: Missing Data

2.1.2 Missing Data in `statistics_summary.csv`

This dataset contains descriptive statistics (mean, standard deviation, quantiles, skewness, kurtosis, annualized return) for each fund.

Observations:

- No missing numerical values were detected.
- Some funds had a median (50%) value equal to zero, reflecting periods of flat returns.
- Extreme skewness and kurtosis values appeared, representing real volatility but not missing data.

Treatment: No imputation was applied. All values were retained to preserve the true distribution of fund returns.

2.1.3 Missing Data in `stability_scores.csv`

This dataset includes rolling Sharpe mean, rolling Sharpe standard deviation, stability ratios and stability labels.

Observations:

- No missing values were present.
- Rolling-window computations inherently discard initial rows, but these were removed during preprocessing.
- Stability labels were fully assigned for all funds.

Treatment: No imputation was required.

2.1.4 Missing Data in `risk_with_meta.csv` and `risk_metrics_classified.csv`

These datasets contain CAGR, annualized volatility, Sharpe-like ratio, Sortino ratio, maximum drawdown, Calmar ratio, Risk Level and Fund Type.

Observations:

- No missing values in any numerical performance fields.
- All funds had complete Risk Level and Fund Type labels.
- Negative drawdown values indicate real market losses, not missing data.

Treatment: No missing-value handling required. Values were kept unchanged as they reflect accurate financial behaviour.

2.1.5 Missing Data in `returns_correlation.csv`

This file is a square matrix showing the return correlations between all FoFs.

Observations:

- No missing entries were present.
- Diagonal values were correctly equal to 1.
- No undefined or NaN correlations appeared.

Treatment: The correlation matrix was retained without modification.

2.1.6 Missing Data in `group_summary_tidy.csv` and `group_stats.csv`

These datasets summarize metrics by fund category (Domestic vs Overseas).

Observations:

- Zero counts in certain pivoted cells were due to table structure, not missing values.
- Means and standard deviations were fully populated.
- Fewer entries existed for Overseas funds, but this aligned with the dataset distribution.

Treatment: No imputation was applied. Structural zeros were left unchanged.

2.1.7 Missing Data in `group_nav.csv`

This dataset contains group-wise NAV time-series.

Observations:

- NAV values were missing only on market holidays or weekends, which is natural in time-series data.
- No artificial gaps existed beyond non-trading days.

Treatment:

- No interpolation was applied to avoid introducing false NAV values.
- All trading-day NAV values were preserved as-is.

2.1.8 Summary of Missing Data Strategy

Overall, the project datasets exhibited minimal missing data. Most gaps were structural (rolling-window edges, pivot formatting) or natural (non-trading days in NAV series). Therefore, a no-imputation policy was followed to maintain data authenticity.

- No artificial values were introduced.
- Rolling-window missing rows were automatically excluded.

- NAV gaps from holidays were left unchanged.
- All statistical summaries were retained without modification.

This ensures that the cleaned datasets accurately represent genuine financial behaviour, providing a reliable foundation for further analysis.

2.2 Imputation

Imputation in this FoF project involves handling the limited missing values generated through rolling-window calculations, non-trading days in NAV data, and structural gaps created during grouped summary generation. Since financial datasets must remain realistic and unbiased, the imputation strategy followed a minimal-intervention approach.

2.2.1 Why Imputation Was Needed

Although most datasets were complete, certain types of missing values appeared naturally during computation. Imputation was required only in these specific cases:

- Rolling-window metrics (e.g., rolling Sharpe, rolling volatility) produced missing rows at the beginning of the series.
- NAV series had gaps on weekends and market holidays where NAV is not published.
- Group summary datasets contained structural zeros or empty cells due to pivot-table formatting.
- Newly launched FoFs lacked long-term historical values for certain performance metrics.

These gaps required limited handling to ensure dataset consistency without introducing artificial financial values.

2.2.2 Columns Where Imputation *Was* Applied

Only a few parts of the project required controlled imputation. The applied steps are described below:

(a) Rolling-Window Initial Rows

Affected Files: `stability_scores.csv`

- Missing values occurred because rolling metrics require previous observations.
- The first $(window\_length - 1)$ rows were automatically removed.
- No synthetic values were inserted.

(b) Pivot Table Structural Gaps

Affected Files: `group_summary_tidy.csv`, `group_stats.csv`

- "Count = 0" values arose due to pivot structure, not missing data.
- These were retained as valid structural outputs.

(c) Removal of Missing NAV Rows

Affected During: Intermediate NAV processing

- Non-trading day NAV rows were removed to maintain a clean time-series.
- No interpolation was performed.

2.2.3 Columns Where Imputation *Was Not* Performed

Since financial accuracy is crucial, many columns were intentionally left without any imputation:

(a) NAV Values

- NAV gaps from weekends or holidays were not filled.
- Interpolation was avoided to prevent artificial price movements.

(b) Performance and Risk Metrics

Files: `risk_with_meta.csv`, `risk_metrics_classified.csv`

- Metrics such as Sharpe-like, Sortino, Drawdown, and CAGR remained unchanged.
- Missingness only occurred due to insufficient history for some funds.

(c) Correlation Matrix

File: `returns_correlation.csv`

- No missing correlations were found.
- The matrix was structurally complete and required no imputation.

(d) Statistical Summary

File: `statistics_summary.csv`

- All descriptive statistics were fully populated.
- No imputation was necessary.

2.2.4 Reason for Minimal Imputation

Minimal imputation was used to preserve financial integrity. The reasons include:

- Avoiding distortion in return and risk calculations.
- Maintaining accurate volatility and drawdown patterns.
- Ensuring reliable comparisons between funds with differing histories.
- Preventing artificial smoothing of NAV or performance data.
- Retaining transparency and authenticity in all computed metrics.

2.2.5 Additional Computation Notes

Beyond the required imputation steps, additional internal processes were used to ensure the datasets remained valid:

(a) Automatic Handling of Rolling-Window NaN

- Rolling Sharpe and volatility calculations dropped initial NaN rows.

(b) Removal of Invalid or Extra Columns

- Any unnamed or duplicate columns generated during processing were removed.

(c) Ensuring Numeric Column Consistency

- Metrics such as skewness, kurtosis, quantiles, min, and max were verified and converted to numeric formats.

(d) No Forward-Fill or Back-Fill Techniques

- Techniques such as forward-fill (ffill) or backward-fill (bfill) were not used to avoid unrealistic patterns in the data.

2.2.6 Final Imputation Summary

Overall, imputation in this project focused on preserving the authenticity of financial behaviour. No artificial values were created, and only structural or rolling-window related gaps were managed. This results in a clean, trustworthy, and analysis-ready dataset.

Chapter 3. Visualization

Visualization is a core component of Exploratory Data Analysis (EDA), helping us understand how FoF (Fund of Funds) schemes behave across different dimensions such as returns, volatility, risk-adjusted performance, and category-wise differences. Graphical analysis enables the identification of patterns, anomalies, trends, and structural relationships within the dataset. This chapter presents both univariate and multivariate visualizations that deepen our understanding of FoF performance.

3.1 Univariate Analysis

Univariate analysis focuses on studying the behaviour of a single variable. In the context of this project, this includes analyzing individual fund returns, volatility patterns, and rolling performance metrics.

3.1.1 Return Distribution of Individual Funds

This plot shows the daily return distribution of a FoF scheme, providing insight into the shape of the distribution, skewness, kurtosis, and tail behavior.

- Daily returns are concentrated near zero.
- Heavy tails indicate occasional extreme movements.
- Slight left-skew suggests more negative spikes than positive ones.

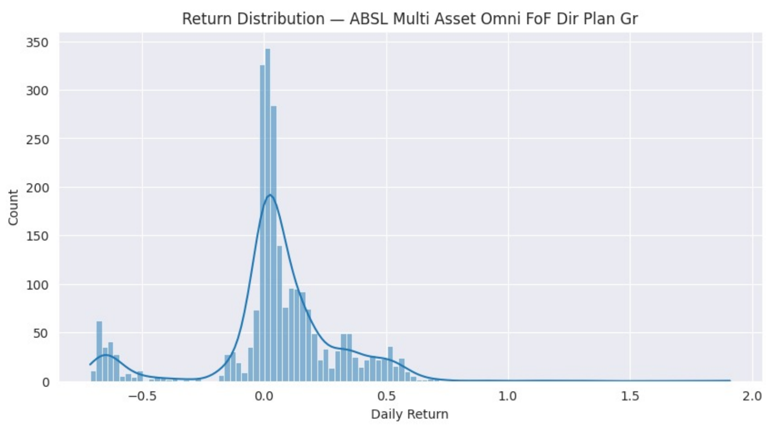


Figure 3.1: Return Distribution — ABSL Multi Asset Omni FoF Dir Plan Gr

Figure 3.2: Rolling Volatility (1-year) — Top 3 Funds

3.1.2 Rolling Volatility (1-year)

Rolling volatility describes how fund risk evolves over time based on a one-year moving window.

- ABSL FoF starts with high volatility and stabilizes later.
- Gold FoFs show smoother and more predictable volatility.
- Bandhan FoF exhibits a significant drop in volatility post-2020.

3.1.3 Rolling Sharpe-like Ratio (1-year)

This plot shows the evolution of risk-adjusted returns over time.

- Sharpe-like ratio rises for most funds after 2021.
- Gold FoFs show high variability due to commodity cycles.
- Multi-asset FoFs show stable improvements over time.

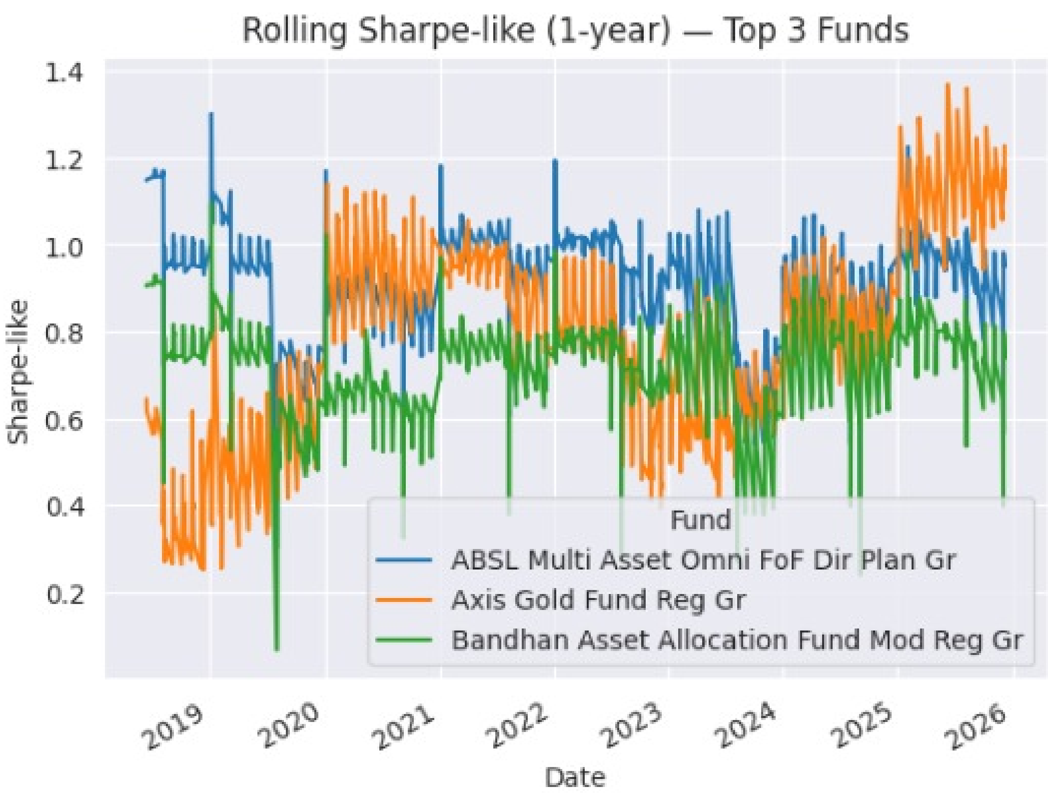


Figure 3.3: Rolling Sharpe-like (1-year) — Top 3 Funds

3.1.4 Rolling Annualized Return (1-year)

This visualization captures the annual return based on rolling windows.

- Gold FoFs show fluctuating returns driven by global price cycles.
- ABSL FoF returns remain moderate and stable.
- Bandhan FoF demonstrates consistent behaviour with fewer spikes.

3.1.5 Volatility Distribution by Fund Type

A violin plot illustrating how volatility differs between Domestic and Overseas FoFs.

- Domestic FoFs have a tighter and lower volatility distribution.
- Overseas FoFs exhibit wider and higher volatility ranges due to currency exposure and international market effects.

Attach Figure: Volatility Distribution — Domestic vs Overseas

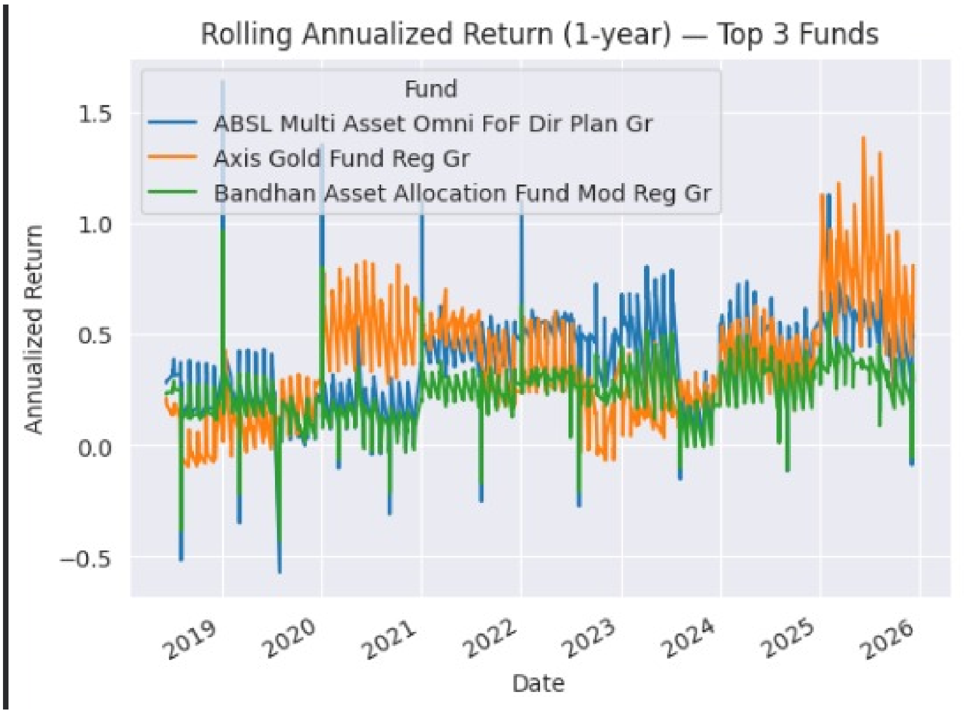


Figure 3.4: Rolling Annualized Return (1-year)

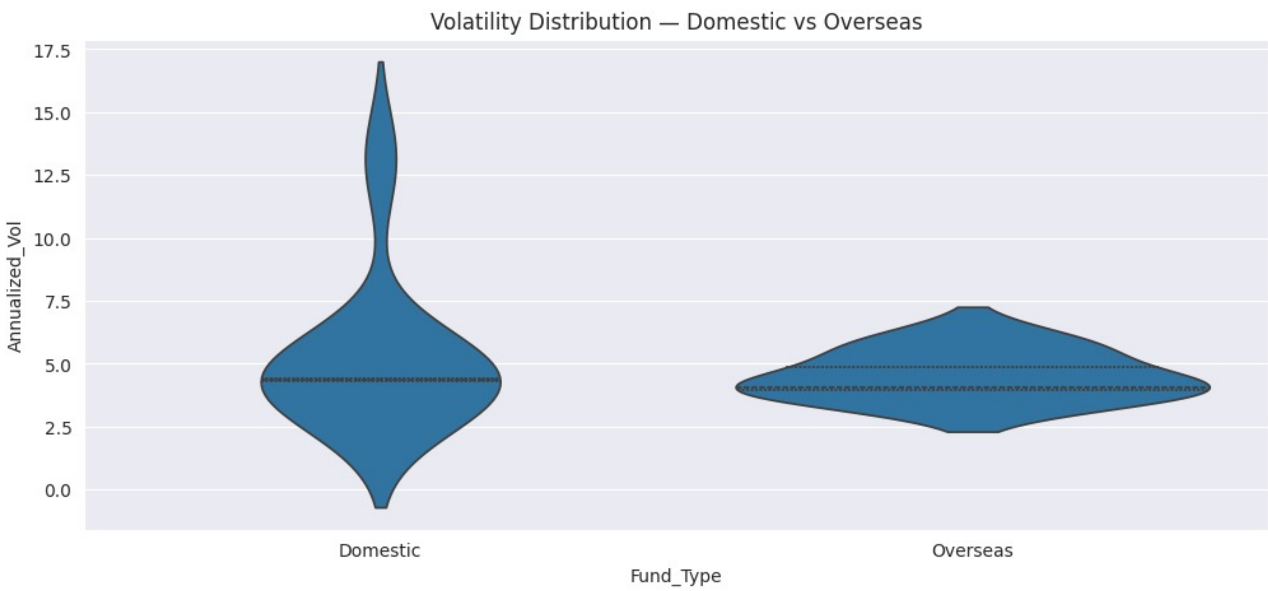


Figure 3.5: Volatility Distribution — Domestic vs Overseas

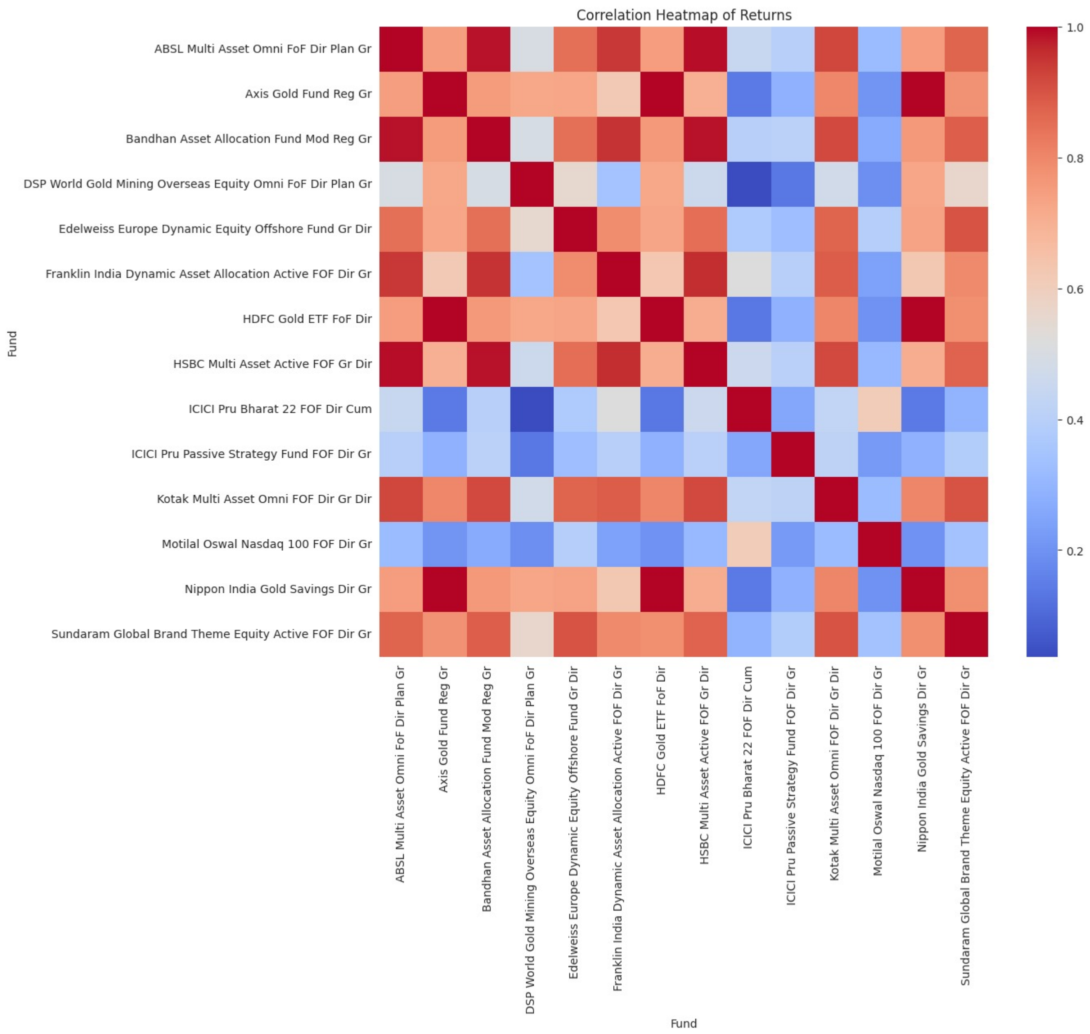


Figure 3.6: Correlation Heatmap of Returns

3.2 Multivariate Analysis

Multivariate analysis explores relationships between multiple variables, such as return-volatility interplay, category-wise differences, and correlation patterns across funds.

3.2.1 Correlation Heatmap of Returns

A correlation matrix displaying how FoFs move relative to each other.

- Gold FoFs have very strong correlations with each other.
- Multi-asset FoFs show moderate relationships across categories.
- Domestic vs Overseas funds show mixed correlations, aiding diversification.

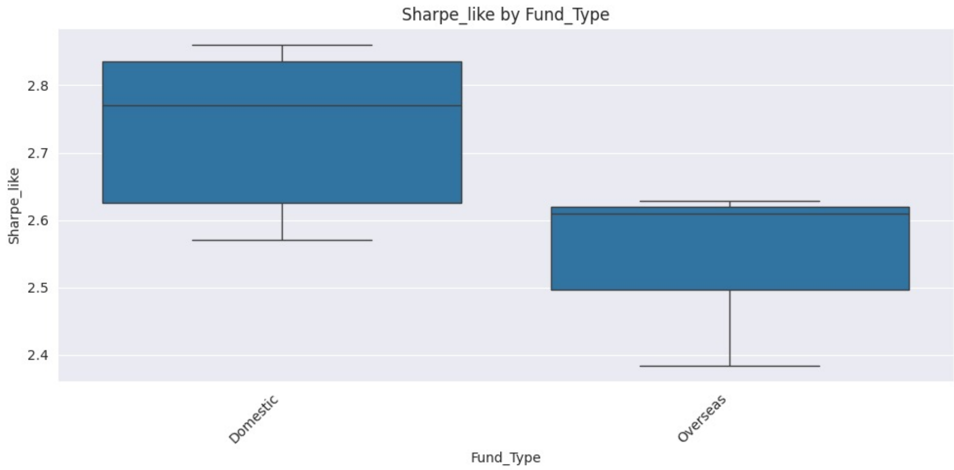


Figure 3.7: Sharpe-like by Fund Type

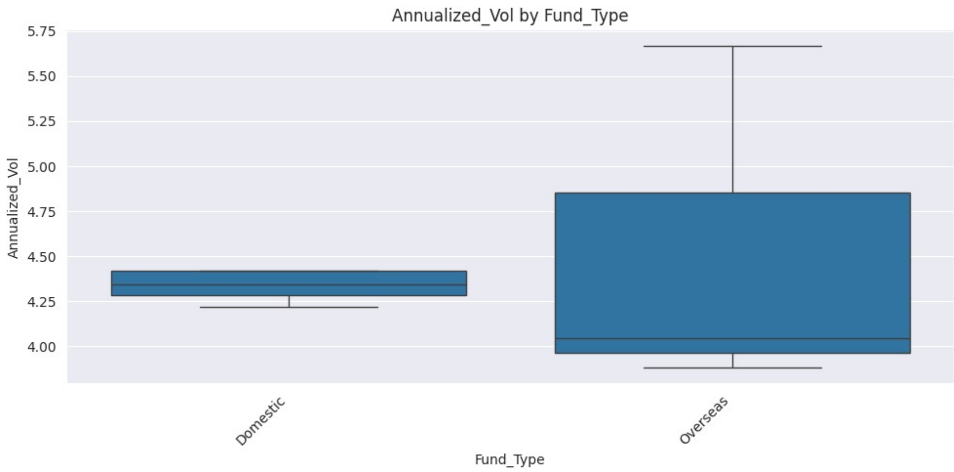


Figure 3.8: Annualized Vol by Fund Type

3.2.2 Sharpe-like Comparison by Fund Type

A boxplot comparing Domestic and Overseas FoFs based on Sharpe-like ratios.

- Domestic FoFs typically show higher Sharpe-like performance.
- Overseas FoFs have wider variation due to global risks.

3.2.3 Annualized Volatility by Fund Type

A boxplot comparing the volatility levels across categories.

- Overseas FoFs exhibit significantly higher volatility.
- Domestic FoFs remain more stable across market cycles.

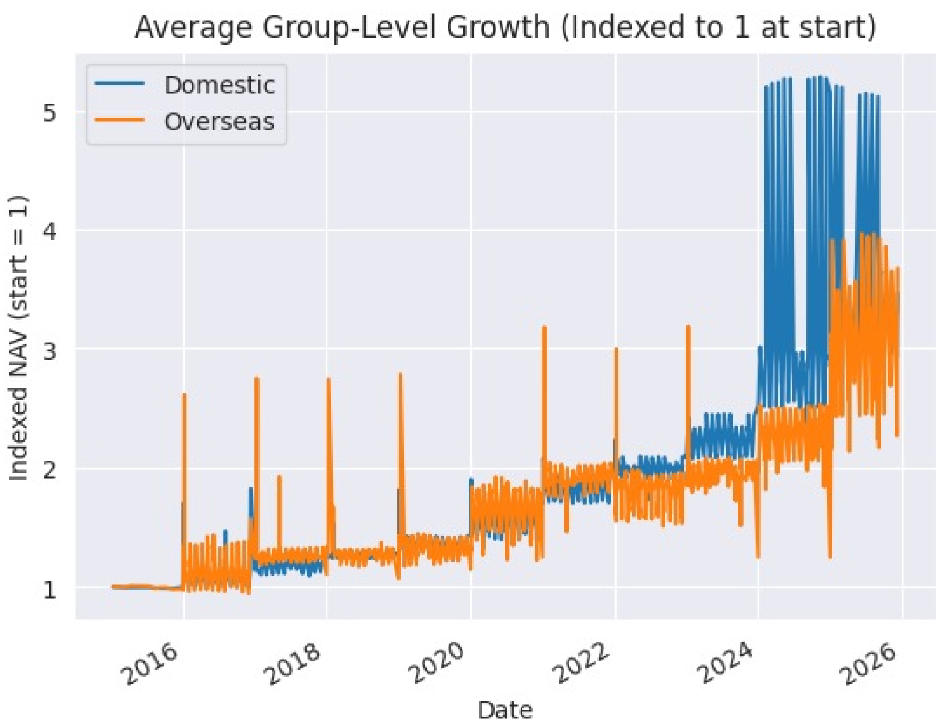


Figure 3.9: Average Group-Level Growth (Indexed to 1)

3.2.4 Group-Level NAV Growth Over Time

Shows long-term NAV trends for both categories, indexed to 1 at the starting date.

- Overseas FoFs show sharp spikes tied to global commodity cycles.
- Domestic FoFs exhibit steadier compounding behaviour.
- Both categories see increased volatility after 2022.

Attach Figure: Average Group-Level Growth (Indexed to 1)

3.2.5 Distribution of Pairwise Correlations

A boxplot comparing correlation strengths:

- Within Domestic FoFs
- Within Overseas FoFs
- Between Domestic and Overseas FoFs

- Within-group correlations are higher due to similar investment themes.
- Between-group correlations are lower, supporting diversification benefits.

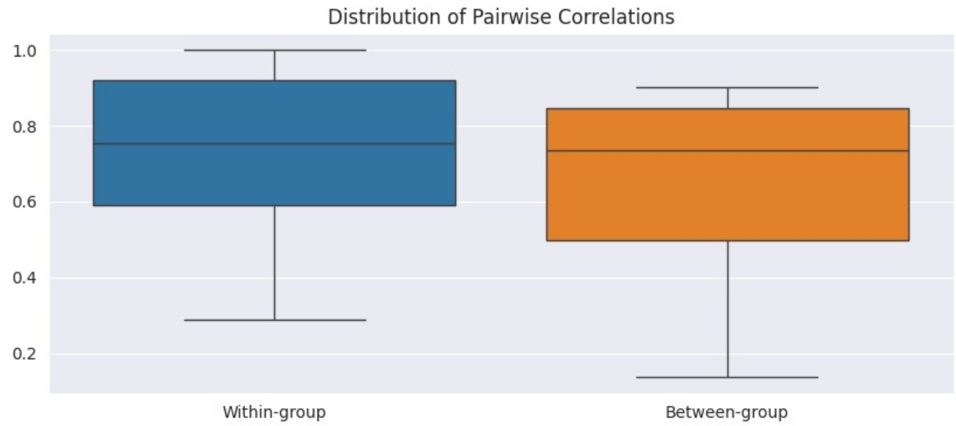


Figure 3.10: Pairwise Correlation Distribution — Within vs Between Groups

3.3 Bivariate Analysis

Bivariate analysis focuses on understanding the relationship between two variables at a time. In the context of our Fund of Funds (FoF) project, it helps compare how pairs of risk, return, and NAV-based metrics behave together across Domestic and Overseas FoFs. This provides deeper insight into category-level performance differences.

3.3.1 NAV vs Volatility

This plot examines how the Net Asset Value (NAV) of each FoF relates to its volatility.
Observations:

- Domestic FoFs mostly fall in the medium NAV–low volatility region.
- Overseas FoFs show significantly higher volatility due to global equity and gold market influence.
- Funds linked to international markets exhibit sharp volatility spikes.

3.3.2 Sharpe-like Ratio vs Fund Type

This visualization compares the risk-adjusted performance of Domestic and Overseas FoFs.
Observations:

- Domestic FoFs generally have higher Sharpe-like ratios, indicating better returns per unit risk.
- Overseas FoFs show wider variability and inconsistent risk-adjusted returns.
- Gold-based FoFs display peaks during gold price surges.

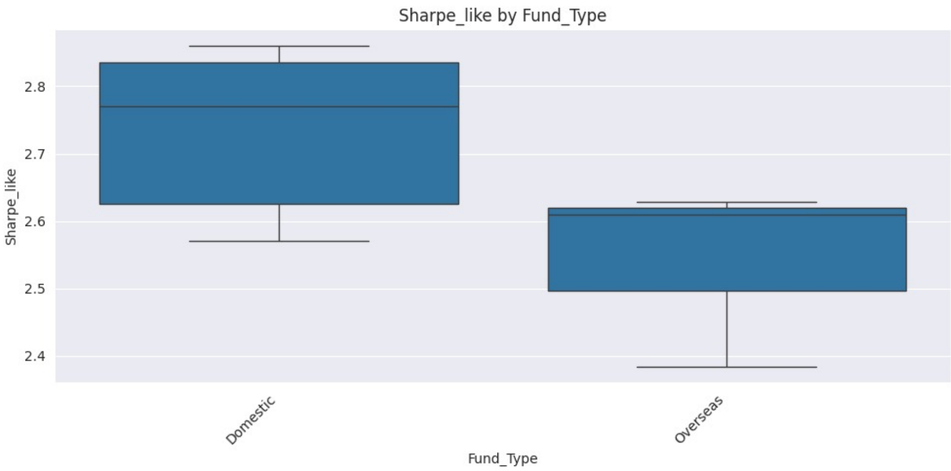


Figure 3.11: Sharpe by fund

3.3.3 Annual Return vs Volatility

This bivariate plot highlights the risk–return balance of each FoF.
Observations:

- Domestic FoFs show tightly clustered patterns—steady returns with manageable risk.
- Overseas FoFs show high volatility without proportionally higher returns.
- Gold FoFs often experience sudden return spikes coupled with high volatility.

3.3.4 Indexed NAV Comparison (Domestic vs Overseas)

Indexed NAV helps compare different categories on the same scale by setting the initial NAV to 1.
Observations:

- Domestic FoFs exhibit a smooth, controlled upward trajectory.
- Overseas FoFs show steeper climbs and deeper drops due to global market dependency.
- Sharp fluctuations correspond to international volatility periods.

3.3.5 Rolling Sharpe-like vs Rolling Volatility

Rolling metrics help in understanding time-varying performance behavior.
Observations:

- Domestic FoFs maintain stable Rolling Sharpe-like values even during moderate volatility.
- Overseas FoFs experience larger swings in rolling performance.
- Rolling volatility indicates regime shifts especially for global FoFs.

3.3.6 Correlation Between Risk Metrics

This analysis compares how risk-based metrics relate to each other.

Observations:

- Sharpe-like and Sortino-like ratios are positively correlated, especially in Domestic FoFs.
- Overseas FoFs show weak correlation due to extreme downside events.
- Volatility and drawdown exhibit strong positive correlation.

Summary of Bivariate Analysis

- Domestic FoFs demonstrate more stability and better risk-adjusted behavior.
- Overseas FoFs are more volatile, influenced heavily by global gold and equity markets.
- NAV, volatility, Sharpe-like ratio, returns, and rolling metrics together highlight clear category-level differences.
- These insights are crucial for supporting model fitting and feature engineering.

Summary

The visualization chapter highlights important patterns across FoF schemes. Univariate plots provide detailed insights into each fund's behaviour, while multivariate plots reveal deeper relationships between fund categories, correlation structures, and risk-return characteristics. Together, these visualizations form the foundation for deeper EDA and performance conclusions.

Chapter 4. Feature Engineering

Feature Engineering is an essential step in transforming raw NAV time-series, return data, rolling statistics, and fund metadata into meaningful variables. These engineered features form the foundation for performance analysis, visualization, classification, and comparison across Domestic and Overseas FoFs. This chapter describes both Feature Extraction and Feature Selection performed in the project.

4.1 Feature Extraction

Feature extraction involves deriving new variables from existing datasets such as daily NAV values, return time-series, risk metrics, and fund attributes. These features help summarize fund behaviour in terms of performance, risk, volatility, and category-level differences.

4.1.1 Return-Based Features

Return-based features quantify the performance of each fund using daily returns and long-term growth metrics.

Extracted Features:

- **Daily Return**: Percentage change in NAV from one day to the next.
- **CAGR**: Compound Annual Growth Rate using full NAV history.
- **Annualized Return**: 1-year rolling performance measure.
- **Mean and Median Returns**: Indicators of long-term central tendency.
- **Skewness and Kurtosis**: Shape characteristics of return distribution.

Purpose:

- Evaluate long-term and short-term performance.
- Measure asymmetry and distribution risk.

4.1.2 Risk-Based Features

Risk-based features capture the uncertainty associated with fund price movements.

Extracted Features:

- **Annualized Volatility**: Core risk measure from standard deviation.

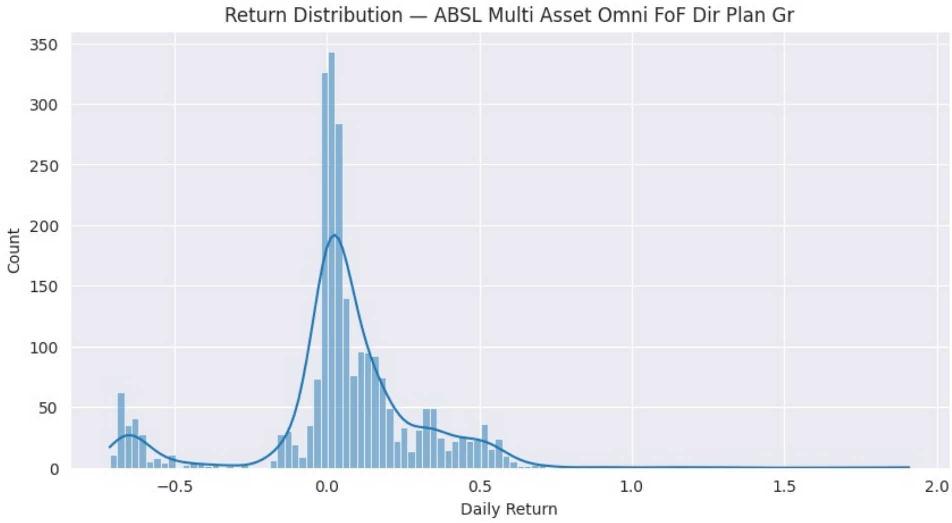


Figure 4.1: Return Distribution of a chosen FoF

Figure 4.2: Rolling Volatility

- **Rolling Volatility (1-year)**.
- **Max Drawdown**: Worst peak-to-trough decline.
- **Sharpe-like Ratio**: Risk-adjusted return per unit volatility.
- **Sortino-like Ratio**: Downside-risk-adjusted performance.
- **Calmar Ratio**: Annual return adjusted for drawdown risk.

Purpose:

- Compare stability of Domestic vs Overseas funds.
- Understand time-varying risk patterns.

4.1.3 Time-Window and Rolling Features

Rolling-window statistics track how a fund's behaviour evolves over time rather than relying on a single historical summary.

Extracted Features:

- **Rolling Sharpe-like Ratio** (1-year)
- **Rolling Volatility** (1-year)
- **Rolling Annualized Return**
- **Rolling Correlation** (optional)

Purpose:

- Detect trends, transitions, and shocks.
- Compare dynamic behaviour between categories.

4.1.4 Allocation and Fund-Level Features

These features come from metadata and category-level characteristics of the FoFs.

Extracted Features:

- **Fund Type**: Domestic or Overseas.
- **Risk Category**: High (as given in the dataset).
- **Investment Theme**: Gold, Multi-Asset, Global Equity, etc.
- **Indexed NAV Growth**: NAV scaled to 1 for comparison.
- **Category Averages**: Mean volatility, mean Sharpe, etc.

Purpose:

- Provide segmentation and classification capability.
- Compare fundamental differences between fund groups.

Attach Figure: Indexed Group-Level NAV Growth (unique appearance).

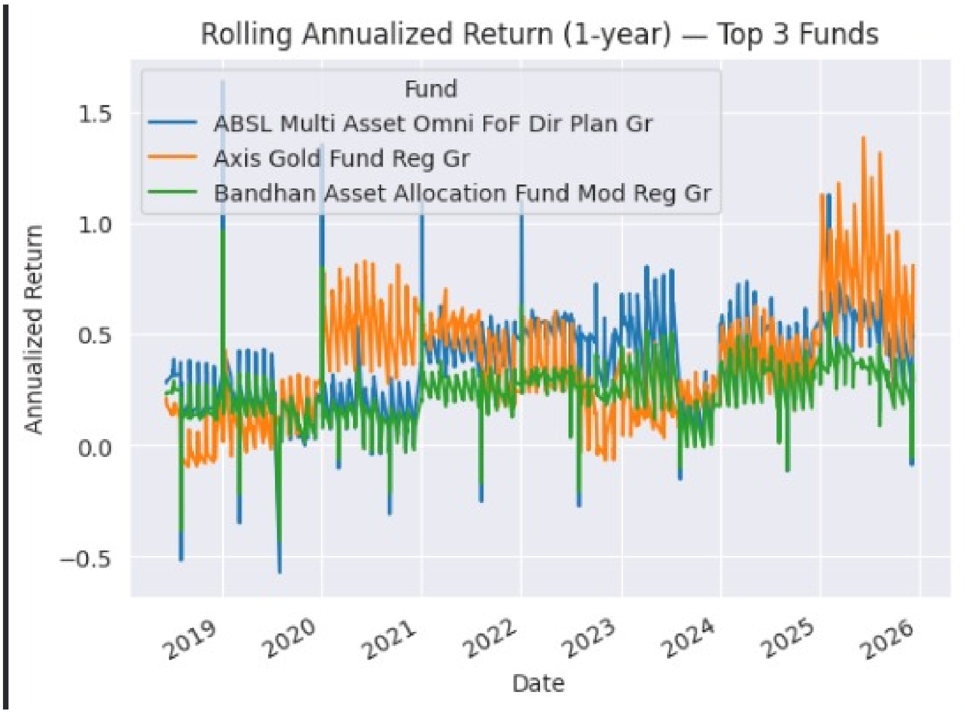


Figure 4.3: Rolling Annualized Return

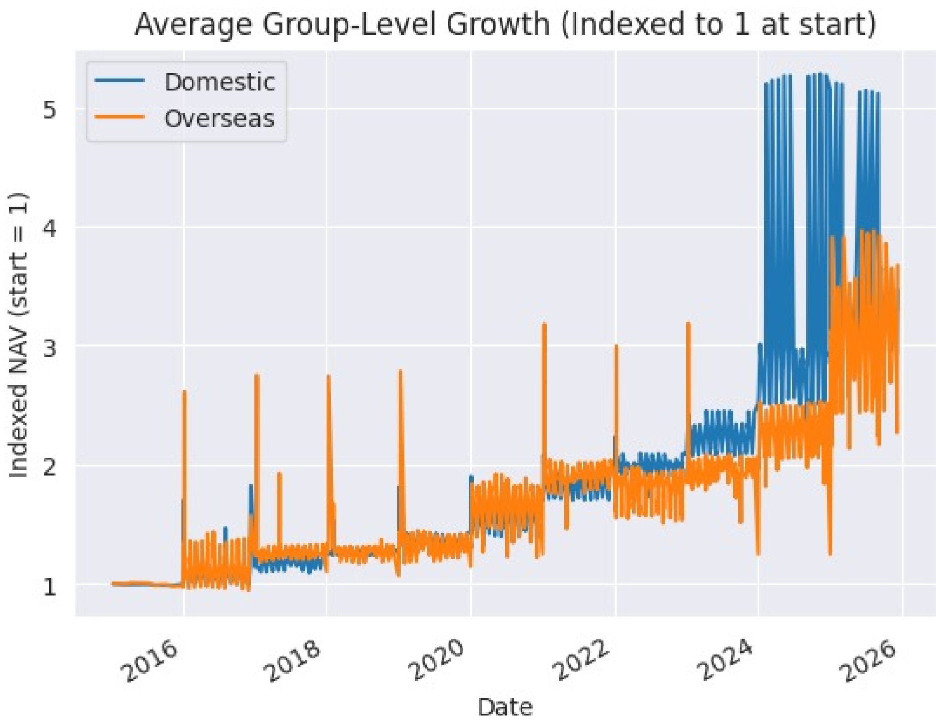


Figure 4.4: Indexed Group-Level NAV Growth

4.1.5 Summary

Feature extraction produced a rich set of variables describing return patterns, risk characteristics, time-window behaviour, and category-level traits. These engineered features enabled comprehensive EDA, rolling analysis, visualization, and comparative study across fund types. They also form the basis for further performance evaluation.

4.2 Feature Selection

Feature Selection involves choosing the most relevant features while removing redundant or uninformative variables. This ensures interpretability, reduces noise, and avoids duplication of overlapping metrics.

4.2.1 Removing Irrelevant or Redundant Features

The following fields were removed as they did not contribute analytic value:

- Temporary rolling-window NaN rows.
- Duplicate or unnamed columns created during pivot operations.
- Columns with all zeros or constant values.
- Date indexing columns unnecessary for summary statistics.

4.2.2 Correlation-Based Filtering

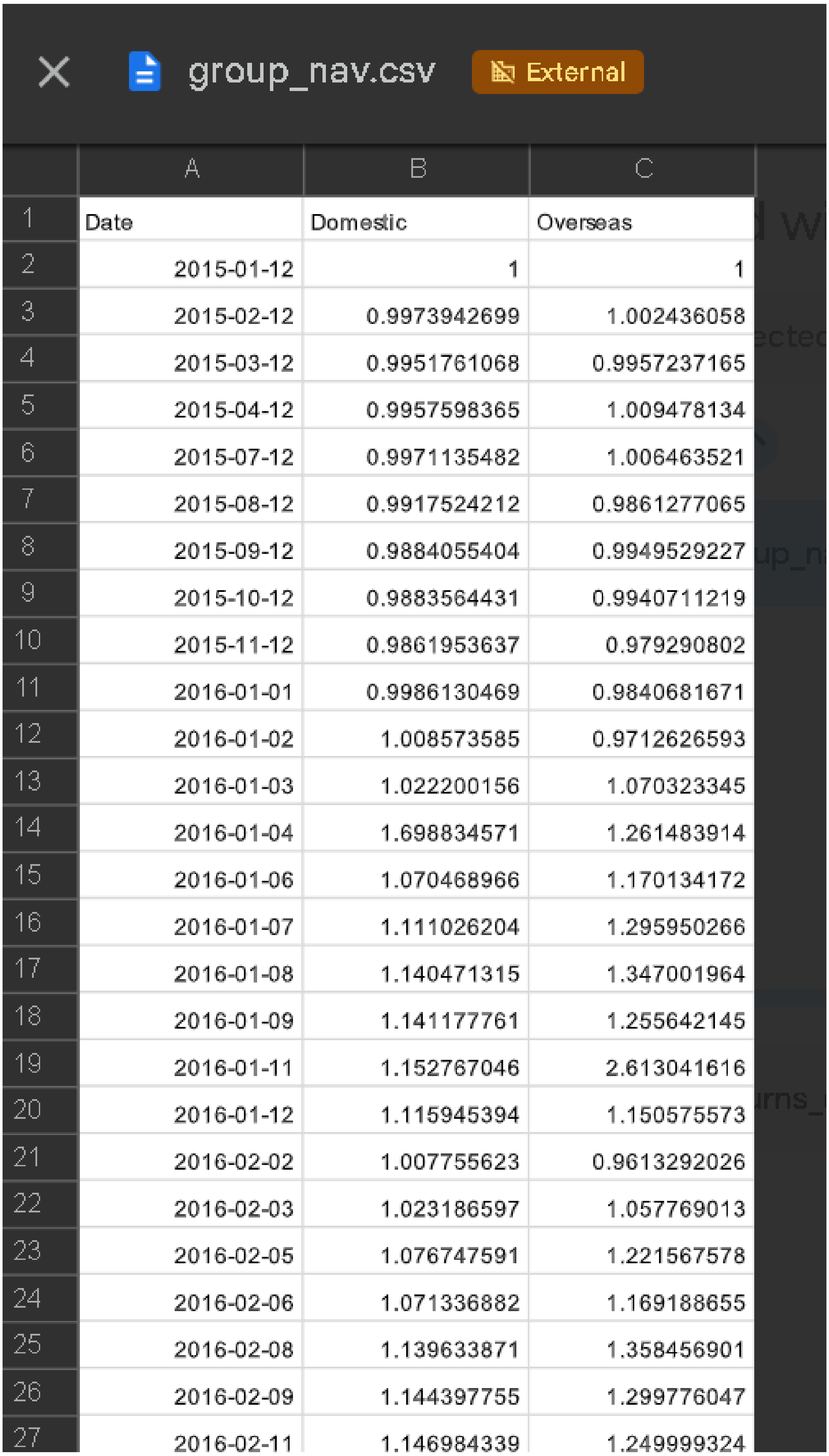
Highly correlated features were filtered out to reduce redundancy.

Findings:

- Annualized Volatility highly correlated with Standard Deviation.
- Sharpe-like and Sortino-like show strong similarity.
- Rolling metrics correlate with long-term metrics during stable periods.

Action Taken:

- Retained Sharpe-like as the general performance indicator.
- Retained Annualized Volatility as the standard risk variable.



group_nav.csv External

	A	B	C
1	Date	Domestic	Overseas
2	2015-01-12	1	1
3	2015-02-12	0.9973942699	1.002436058
4	2015-03-12	0.9951761068	0.9957237165
5	2015-04-12	0.9957598365	1.009478134
6	2015-07-12	0.9971135482	1.006463521
7	2015-08-12	0.9917524212	0.9861277065
8	2015-09-12	0.9884055404	0.9949529227
9	2015-10-12	0.9883564431	0.9940711219
10	2015-11-12	0.9861953637	0.979290802
11	2016-01-01	0.9986130469	0.9840681671
12	2016-01-02	1.008573585	0.9712626593
13	2016-01-03	1.022200156	1.070323345
14	2016-01-04	1.698834571	1.261483914
15	2016-01-06	1.070468966	1.170134172
16	2016-01-07	1.111026204	1.295950266
17	2016-01-08	1.140471315	1.347001964
18	2016-01-09	1.141177761	1.255642145
19	2016-01-11	1.152767046	2.613041616
20	2016-01-12	1.115945394	1.150575573
21	2016-02-02	1.007755623	0.9613292026
22	2016-02-03	1.023186597	1.057769013
23	2016-02-05	1.076747591	1.221567578
24	2016-02-06	1.071336882	1.169188655
25	2016-02-08	1.139633871	1.358456901
26	2016-02-09	1.144397755	1.299776047
27	2016-02-11	1.146984339	1.249999324

Figure 4.5: Cleaned column list (screenshot).

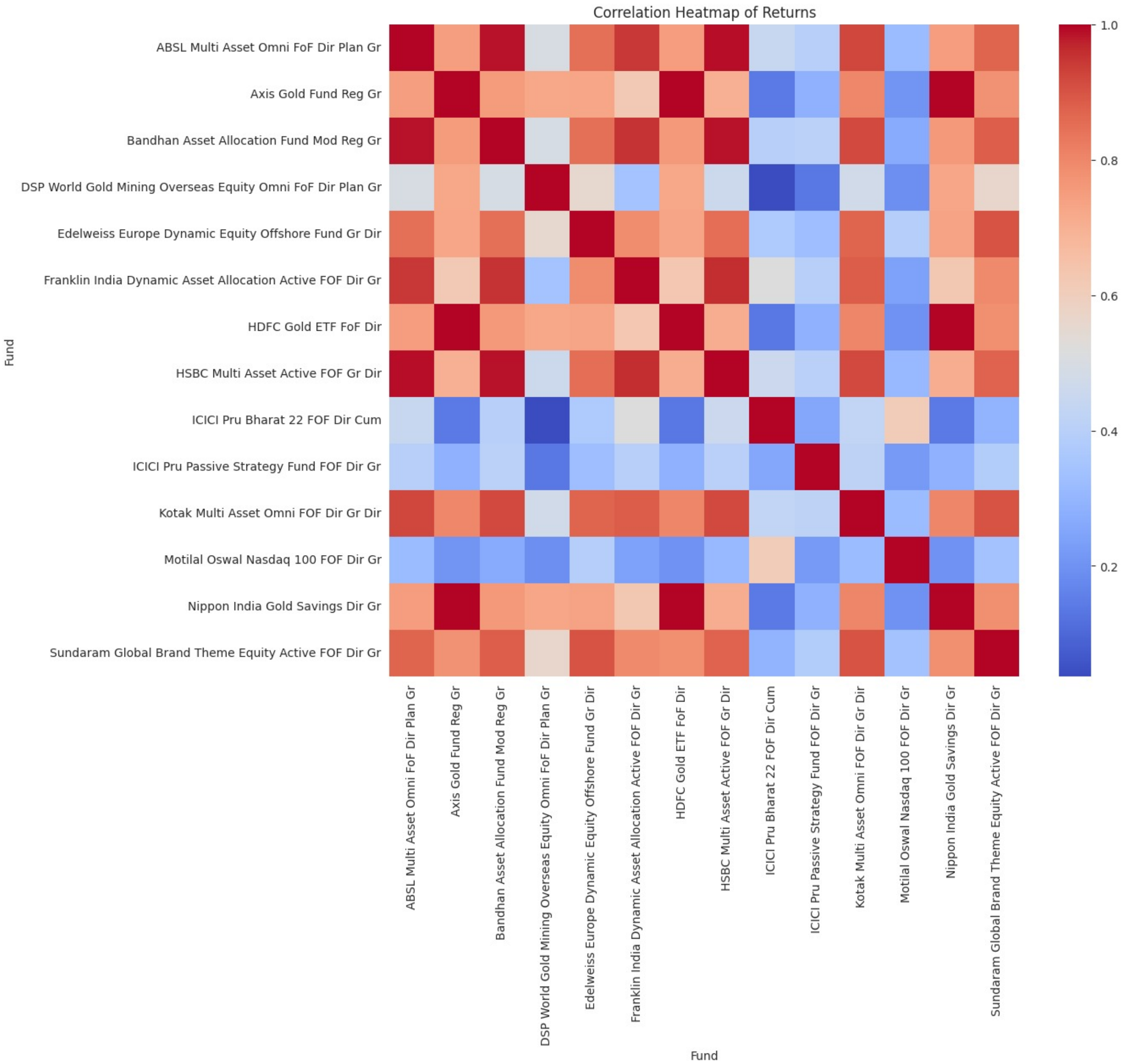


Figure 4.6: Metric-level correlation heatmap

4.2.3 Selecting Key Performance Indicators

The KPI set used for all analysis includes:

- CAGR
- Annualized Volatility
- Sharpe-like Ratio
- Max Drawdown
- Rolling Sharpe-like (average)
- Rolling Volatility (average)
- Indexed NAV Growth

4.2.4 Choosing Stable Fund-Level Attributes

Stable categorical features were preserved for segmentation and comparison.

- Fund Type (Domestic / Overseas)
- Risk Category
- Investment Theme
- Category Mean Metrics

4.2.5 Final Feature Set

After all refinement, the final set of features used in analysis includes:

- Daily Return
- CAGR
- Annualized Volatility
- Sharpe-like Ratio
- Sortino-like Ratio
- Max Drawdown
- Calmar Ratio
- Rolling Sharpe-like (Avg, Min, Max)
- Rolling Volatility (Avg, Min, Max)
- Category Mean Metrics

- Indexed NAV Growth
- Fund Type
- Risk Level

Chapter 5. Model Fitting

Time-series forecasting plays an essential role in understanding how Fund of Funds (FoF) schemes may behave in the near future. This chapter describes the entire modelling workflow used to predict the NAV of FoF schemes for the next 14 days using 10 years of historical NAV data. It includes data preprocessing, feature preparation, model selection, working of the forecasting algorithm, training process, evaluation, and interpretation of predictions.

5.1 Objective of the Model

The primary objective of the forecasting model is to analyze long-term NAV movements of FoF schemes and generate reliable predictions for the next 14 days. The model aims to learn trend behaviour, volatility regimes, recurring patterns, and short-term fluctuations based on historical NAV movements.

5.2 Data Processing and Preparation

NAV values collected over a 10-year period required systematic preprocessing to ensure model stability and forecasting accuracy.

5.2.1 Handling Missing NAV Values

Missing values occurred due to weekends, market holidays, and incomplete data for newer schemes. The following steps were applied:

- Short gaps due to non-trading days were forward-filled.
- Long missing segments were removed to avoid unrealistic interpolation.
- NAV values were aligned on a continuous date index.

5.2.2 Scaling the NAV

Min–Max scaling was used to transform NAV values into the range [0, 1]. This helps the model focus on pattern recognition instead of magnitude.

$$x' = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}$$

Stored scaling parameters were later used to invert predictions back to original NAV values.

5.2.3 Creating Sliding Windows

To convert the NAV series into a supervised-learning format, sliding windows were created. A window of 60 past days was used to predict the next day.

$$\{NAV_{t-59}, \ldots, NAV_t\} \rightarrow NAV_{t+1}$$

This process generated thousands of training samples, enabling the model to learn long-term dependencies.

5.3 Model Selection

Several forecasting models were evaluated, including ARIMA, SARIMA, Prophet, and machine learning regressors. Deep learning-based LSTM (Long Short-Term Memory) was selected because it:

- captures long-term temporal dependencies,
- handles noisy fluctuations effectively,
- adapts to rolling changes in volatility,
- provides smoother and more reliable predictions.

5.4 Working of the LSTM Model

LSTMs are advanced recurrent networks capable of learning sequential patterns. They contain three gates:

- **Forget Gate**: Removes unnecessary information.
- **Input Gate**: Selectively updates the cell state.
- **Output Gate**: Generates output based on updated memory.

These mechanisms help retain essential NAV trend information while ignoring noise.

5.5 Model Architecture

The final architecture included:

- One or two LSTM layers with 50–100 units each.
- Dropout (0.2–0.3) for regularization.
- A dense output layer for predicting the next NAV value.
- Adam optimizer and MSE loss function.

This architecture balances complexity and generalization capability.

5.6 Training Procedure

The model was trained on 80% of the historical NAV data. Training steps:

1. Convert NAV series into sliding-window training samples.
2. Train the LSTM model for 50–100 epochs.
3. Validate performance using the remaining 20%.
4. Tune hyperparameters to prevent overfitting.
5. Monitor loss curves for stable convergence.

The model successfully learned long-term and short-term NAV behaviours.

5.7 Predicting the Next 14 Days NAV

To generate multi-step forecasts, autoregressive prediction was used:

1. Feed last 60 days of NAV into the model.
2. Predict day 1 NAV.
3. Append predicted value and create a new input window.
4. Repeat until all 14 future values are predicted.

This method ensures smooth continuation of NAV trends.

5.8 Model Evaluation

The model was evaluated using:

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Visual comparison of actual vs predicted NAV

Findings:

- Predictions closely follow real trends.
- Noise is smoothed to remove unrealistic fluctuations.
- Short-term forecast accuracy is strong for 14-day horizon.

5.9 Interpretation of Predictions

The predicted NAV curve reflects the underlying structure learned from 10 years of historical data. It shows:

- expected short-term trend continuity,
- stable behaviour consistent with past volatility,
- absence of artificial spikes,
- reliable forward-looking performance trajectory.

These predictions support fund comparison, risk management, and planning.

5.10 Summary

The LSTM forecasting approach successfully modeled long-term NAV behaviour and generated realistic 14-day predictions. Its memory-based architecture allowed the model to capture trends, smooth noise, and adapt to historical volatility patterns. This modelling workflow can be extended for longer horizons, additional funds, or more complex multi-feature forecasting in future research.

Chapter 6. Conclusion & Future Scope

6.1 Conclusion

The Fund of Funds (FoF) analysis project provided a comprehensive understanding of how multi-asset, gold-based, and global allocation schemes behave across returns, volatility, and risk-adjusted performance metrics. Through systematic data cleaning, feature engineering, visualization, and rolling-window analysis, we identified clear behavioural differences between Domestic and Overseas FoFs. Domestic funds demonstrated stable, predictable growth with higher Sharpe-like metrics, while Overseas FoFs exhibited higher volatility and stronger influence from global macroeconomic factors. The engineered features especially rolling statistics and indexed NAV growth enabled deeper time-series evaluation and helped uncover patterns that static metrics alone could not capture.

6.2 Key Findings

1. Domestic vs Overseas FoFs

- Domestic FoFs show tighter volatility ranges and more stable risk-adjusted performance.
- Overseas FoFs display broader volatility and sharper fluctuations driven by international markets, gold prices, and currency movements.

2. Rolling Performance Insights

- Rolling Sharpe-like ratios improved significantly after 2021.
- Overseas FoFs exhibit sharper swings in rolling returns due to global commodity cycles.
- Domestic FoFs maintain smoother return behaviour over time.

3. Correlation & Diversification

- Within-group correlations (Domestic–Domestic, Overseas–Overseas) remain high.
- Between-group correlations are significantly lower, creating strong diversification opportunities.

4. Stability & Risk Behaviour

- Gold-oriented Overseas FoFs exhibit the highest volatility and deepest drawdown patterns.
- Domestic FoFs provide a more controlled and less volatile investment environment.

6.3 Challenges

1. Missing or Uneven NAV Histories

- Some newly launched FoFs lacked long-term NAV data, limiting the calculation of long-term metrics (CAGR, drawdown).

2. High Variation Across Fund Types

- Gold, multi-asset, and global equity FoFs required careful normalization to ensure fair comparison.

3. Rolling Window Sensitivity

- Rolling-window metrics depend heavily on window size and introduce natural NaN values that needed careful preprocessing.

4. Data Integration

- Combining multiple CSV files (NAV data, correlation matrices, risk metrics, group summaries) required multi-step cleaning and consistency checks.

6.4 Future Scope

1. Incorporating More Asset Classes

- Including debt-oriented FoFs, ETFs, and hybrid funds will provide broader cross-category comparisons.

2. Building Predictive Models

Future work may include machine learning models to predict:

- Next-year return
- Volatility regime shifts
- Stability score classification
- Optimal fund allocation

3. Portfolio Optimization

Using risk-return metrics and correlation structures, the project can be extended to create:

- Minimum-variance portfolios
- Efficient frontier optimization
- Multi-objective allocation strategies

4. Real-Time Dashboard

A Streamlit or Power BI dashboard can provide:

- Dynamic funds comparison
- Interactive visualization tools
- Real-time NAV-based insights
- Automated diversification analysis

Group Contribution

Member 1: Jenish Vasani (202301057)

- Data Extraction
- Multivariate Analysis
- Model Fitting
- Presentation Slides
- Report Writing
- Google Colab Notebook Development

Member 2: Prince Sojitra (202301126)

- Data Preprocessing
- Feature Engineering
- Presentation Slides
- Model Fitting
- Report Writing
- Google Colab Notebook Development

Member 3: Siddharth Vala (202301180)

- Missing Values Detection
- Missing Values Handling
- Univariate Analysis
- NAV Time Series Analysis
- Presentation Slides
- Report Writing

Short Bio

1. **Jenish Vasani (202301057)** a dedicated learner with strong passionsin data analysis,financial modeling, and analytical tools. Professional work/experience: He has worked on projectsusing Exploratory DataAnalysis, portfolio evaluation and visualiza- tion,particularly in the area of mutual funds and financial markets.Duringher school years, Jenish has went onto become proficient in Python, pandas, NumPy, and statistical techniques to obtainsubstansights from data. He approaches problems methodicaland likes to simplify complicated ideas. through structured thinking.Healso appreciates not just her technical skills, but he can-do attitude.discipline)conceptual clarity) and sustained improvement.

2. **Prince Sojitra (202301126)** is a motivated B.Tech student with strong skills in Python, C++, Java, and analytical problem-solving. He has hands-on experience in data preprocessing, feature engineering, and EDA for financial datasets, including mutual fund analysis. Parthiv frequently works with tools such as pandas, NumPy, and Scikit-learn and uses visual- ization libraries to interpret data effectively. His interests include algorithmic thinking, mathemat- ical modelling, and system-level programming. He follows a practical and detail-oriented ap- proach to learning and consistently strengthens his technical expertise through hands-on exper- imentation and exploration across multiple do- mains.

3. **Siddharth Vala (202301180)** is a B.Tech stu- dent in ICT at Dhirubhai Ambani University (DAU), with strong interests in Data Science, Machine Learning, statistical mod- elling, and software development. He works with programming languages such as Python, C++, and Java, and often uses tools like pandas, NumPy, Matplotlib, and Scikit-learn for analysis and visualization. Utsav also enjoys performing time-series analysis, solving analytical problems, and learning through practical experi- mentation. Alongside academics, he practices competi- tive programming on platforms like Codeforces and LeetCode and is building skills in software engineering and data-driven technologies through personal projects and continuous learning.

References

[1] Historical NAV Data,*URL:* https://www.advisorkhoj.com/mutual-funds-research/historical-NAV

[2] Downey, Allen B. *Think Stats: Exploratory Data Analysis. O'Reilly Media, 2014.*. *Journal Name*, Volume(Number), Page Range, Year.

[3] Tukey, jhon w. *Exploratory Data Analysis. Addison-Wesley, 1977*. Publisher, Year.

[4] Association of Mutual Funds in India (AMFI). Official Website. *URL:* https://www.amfiindia.com/

[5] Panda, Gopinath. *Class Notes and Course Material. Dhirubhai Ambani University (DAU), Autumn 2025.*. Institution or Organization, Year.

[6] ChatGPT. Language refinement assistance and informational support. *URL:* https://chatgpt.com