FUND OF FUNDS

Prof. - Gopinath Panda

Group 43

➜ Jenish Vasani (202301057)

➜ Prince Sojitra (202301126)

➜ Siddharth Vala (202301180)

Overview

→ Analyzed 10 years of NAV data for multiple Fund of Funds (FoF) schemes.

→ Performed complete data cleaning, missing value handling, and feature extraction.

→ Conducted univariate, bivariate, and multivariate visualizations to understand fund behavior.

→ Compared Domestic vs Overseas FoFs based on returns, volatility, and risk metrics.

→ Built an LSTM deep learning model to forecast the next 14 days of NAV.

→ Delivered insights, key findings, challenges, and future scope for improving FoF analysis.

What are FoFs?

➡ A Fund of Funds (FoF) is a type of mutual fund that invests indirectly, by allocating money into other mutual funds rather than individual stocks or bonds.

➡ Provides broad diversification across asset classes, fund managers, sectors & geographies.

➡ Types of FoFs included in our project:
  ◦ Domestic Multi-Asset FoFs (Invest in equity, debt, gold, etc.)
  ◦ Gold FoFs (Invest in gold ETF funds)
  ◦ International / Overseas FoFs (Invest in foreign equity funds)

➡ NAV of FoFs depends on the performance of underlying mutual funds → making it complex & interesting for analysis.

Why Analyze FoFs?

➡ FoFs behave differently from normal mutual funds → more diversified.

➡ Useful for:

- Understanding risk–return behaviour
- Identifying which category performs better
- Studying global trends (especially gold & international markets)
- Evaluating stability and volatility

➡ NAV analysis reveals:

- Long-term growth
- Short-term fluctuations
- Market cycle behaviour
- Category-wise differences (Domestic vs Overseas)

Project Purpose

This project aims to:

➡ Explore 10-year NAV history of different FoFs

➡ Conduct univariate, bivariate & multivariate analysis

➡ Understand fund behaviour using rolling & risk metrics

➡ Perform feature engineering to build meaningful indicators

➡ Fit a time-series forecasting model (LSTM)

➡ Predict next 14 days NAV

➡ Provide clear conclusions for real-world FoF analysis

Tools & Technologies Used

- **Python** – Core language for data processing
- **Pandas** – Data manipulation, cleaning & structuring
- **NumPy** – Fast mathematical & array operations
- **Matplotlib / Seaborn** – Data visualization & trend analysis
- **Scikit-learn** – Feature scaling & preprocessing
- **TensorFlow / Keras (LSTM)** – Time-series NAV prediction
- **Jupyter Notebook / Google Colab** – Interactive coding environment
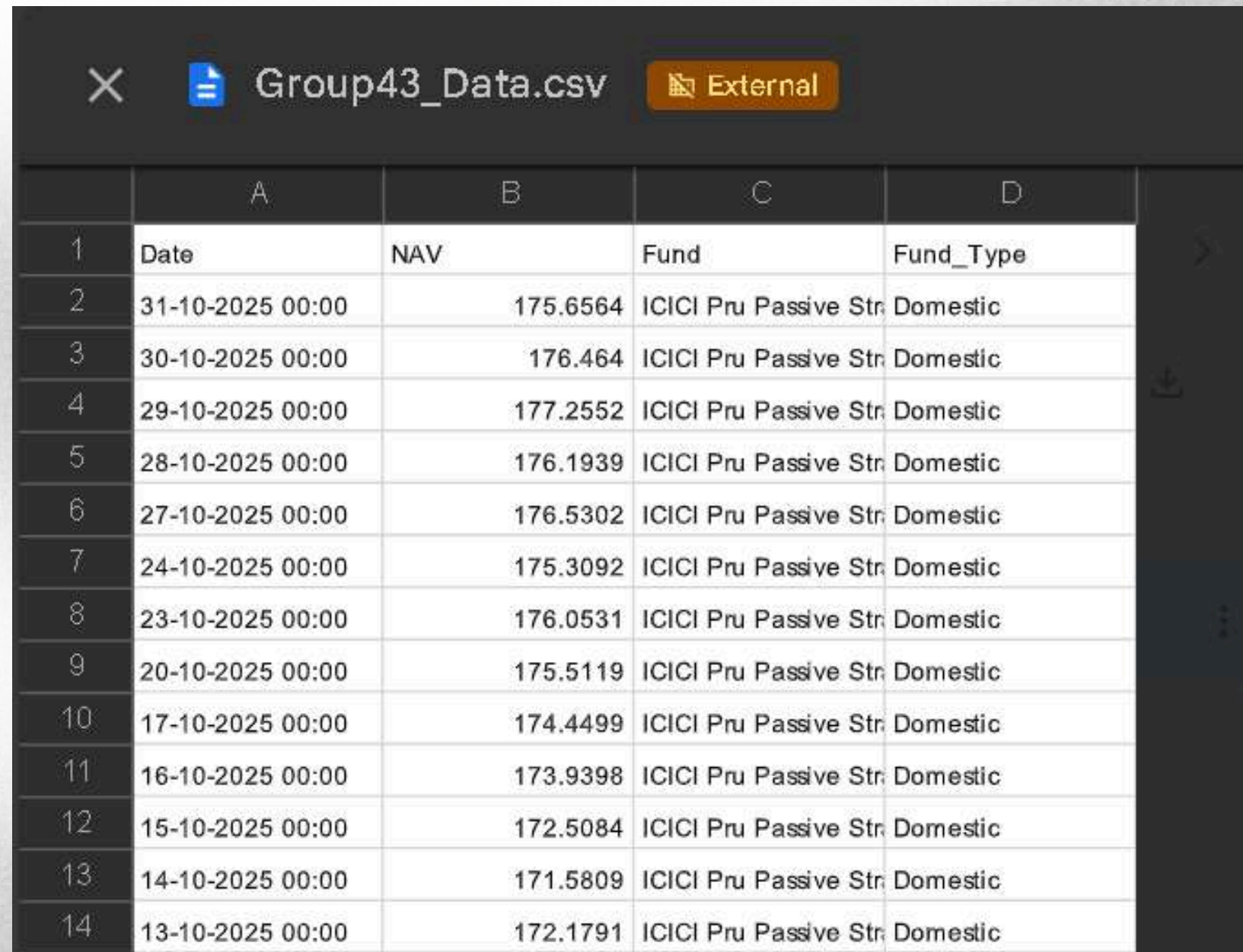
Dataset Overview

Datasets used:

➡ nav_values.csv – 10-year NAV data for all FoFs

➡ returns_correlation.csv – correlation matrix

➡ risk_metrics_classified.csv – Sharpe, Sortino, volatility etc.

➡ stability_scores.csv – rolling stability

➡ group_summary.csv – Domestic vs Overseas stats

➡ categorised_risk.csv – Risk level categories

Coverage:

➡ More than 15 FoF schemes

➡ 2015 to 2025

➡ Mixed categories → Multi-Asset, Gold, Overseas equity

Raw Data Problems

Raw FoF NAV data contained multiple issues:

➡ Missing values (NA, blank cells, "–", "N.A", "null")

➡ Unaligned dates due to holidays

➡ New fund launches → missing early years

➡ Invalid columns auto-generated during extraction

➡ Irregular formats

➡ Different naming conventions across sources

➡ Duplicate rows in some datasets

Missing Data Analysis

Types of missing NAV data:

- Market Holidays: NAV not published on weekends/festivals
- MFAPI gaps: API sometimes skips older data
- New schemes: Recently launched funds do not have long history
- ZeroNAV / NullNAV: Some values returned as 0 before correction

For each missing case:

- Holiday gaps → harmless
- Large missing ranges → removed
- No artificial interpolation (kept realistic behaviour)

Cleaning Strategy

Cleaning steps in detail:

- Converted special characters ("–", "N.A") → NaN
- Removed fully empty columns
- Aligned NAV data across all dates
- Removed invalid & unnamed index columns
- Converted % values in risk metrics into numeric
- Verified rolling-window compatibility
- Ensured smooth continuous time-series for model

Final Result After Cleaning

The cleaned dataset:

- Contains only valid NAV data
- Has consistent date index
- Maintains natural missing behaviour
- Is usable for rolling calculations
- Is ready for visual analysis & modelling

Visualization Purpose

➡ Detect trends & cycles

➡ Compare Domestic vs Overseas

➡ Visualize return distributions

➡ Understand risk characteristics

➡ Identify outliers

➡ Support model building

➡ Confirm correctness of cleaned data

Univariate Analysis

Return Distribution — ABSL Multi Asset Omni FoF Dir Plan Gr
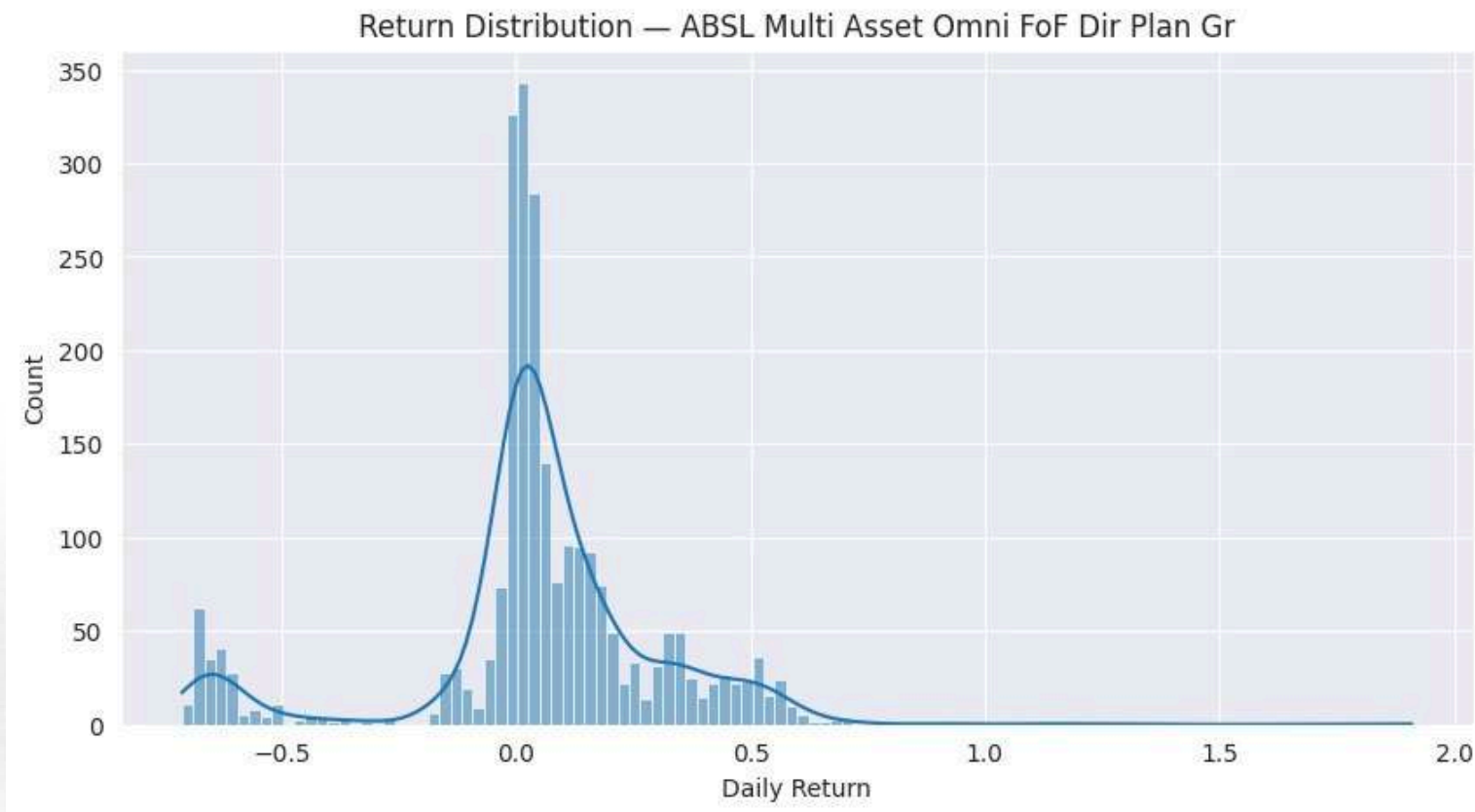
Univariate visuals included:
- NAV Trend Graphs: Long-term upward/downward behaviour
- Return Distribution Histogram: Normality/deviation
- Volatility Distribution: Stability assessment
- Sharpe-like Ratio Distribution: Performance quality
- Stability Score Histogram

Each reveals behaviour of individual funds.

Rolling Volatility (1-year) — Top 3 Funds

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

Rolling Annualized Return (1-year) — Top 3 Funds

Univariate Analysis

Volatility Distribution — Domestic vs Overseas
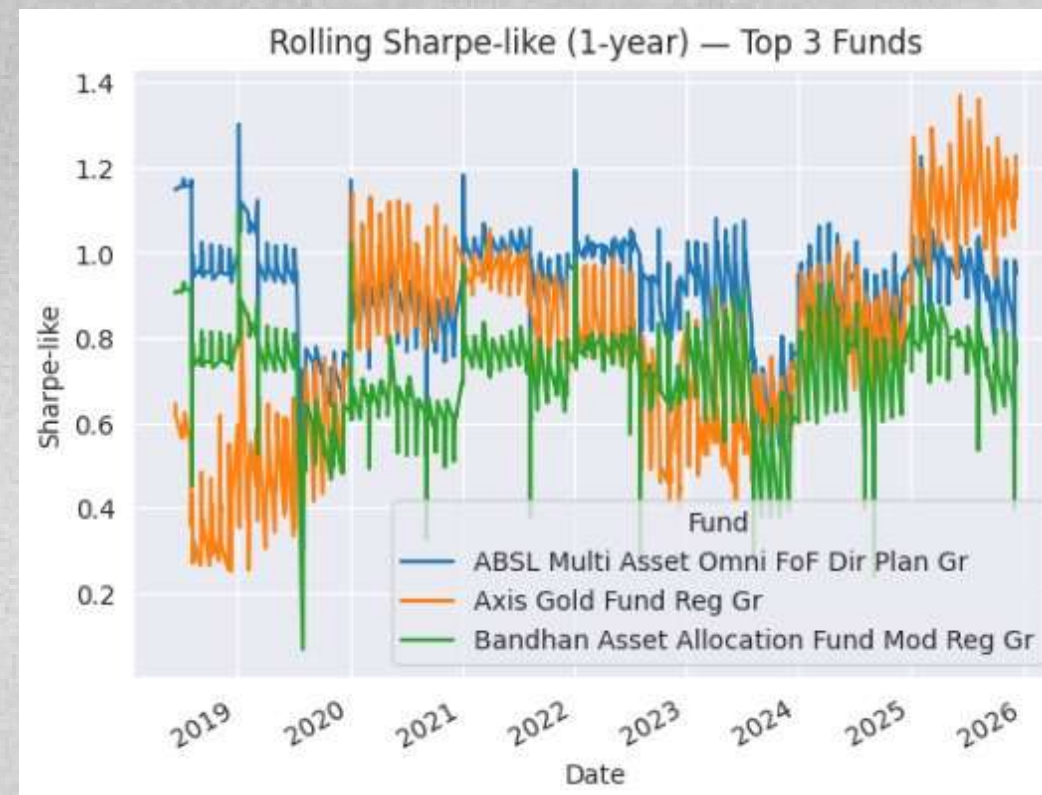
Correlation Heatmap of Returns
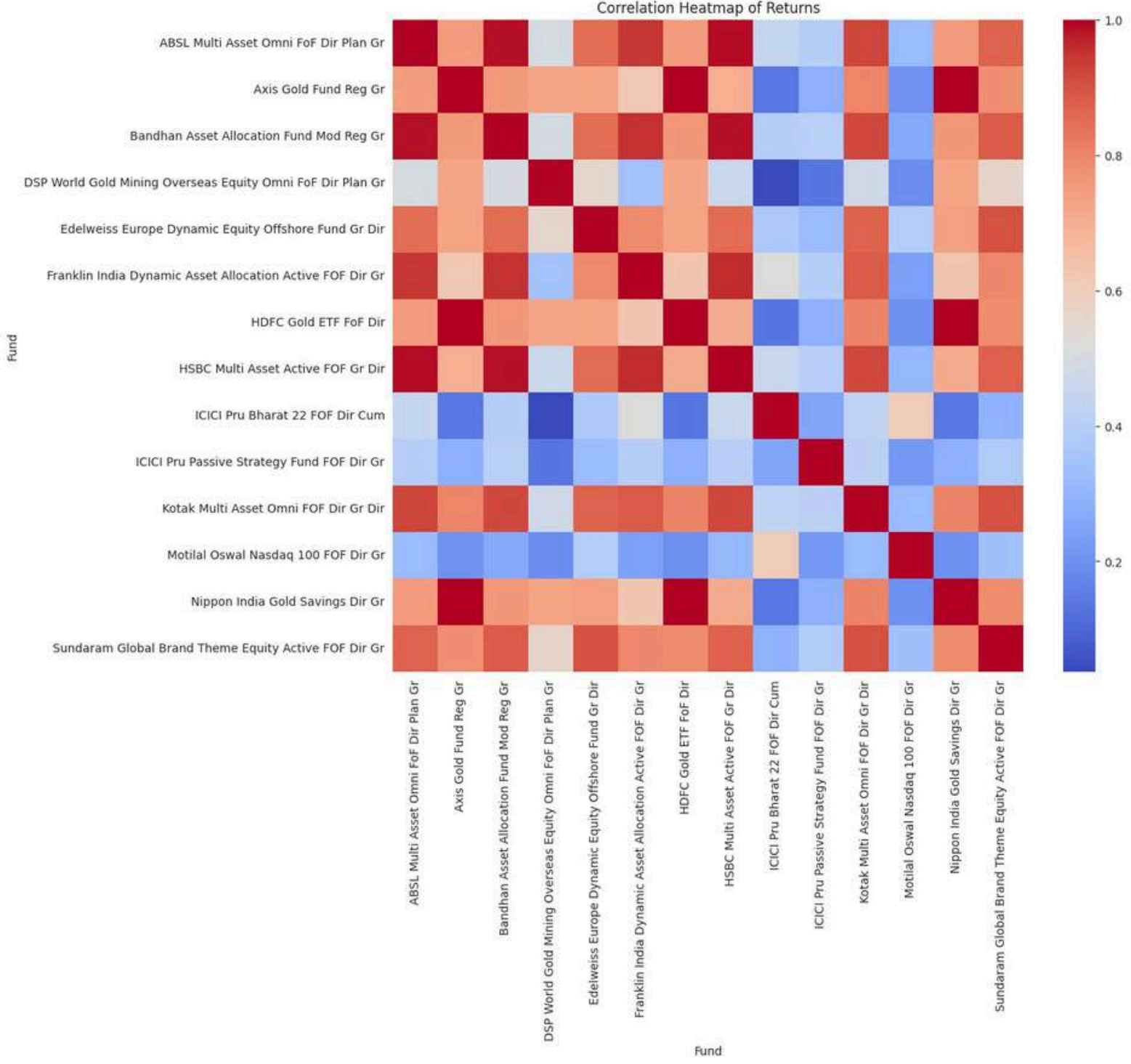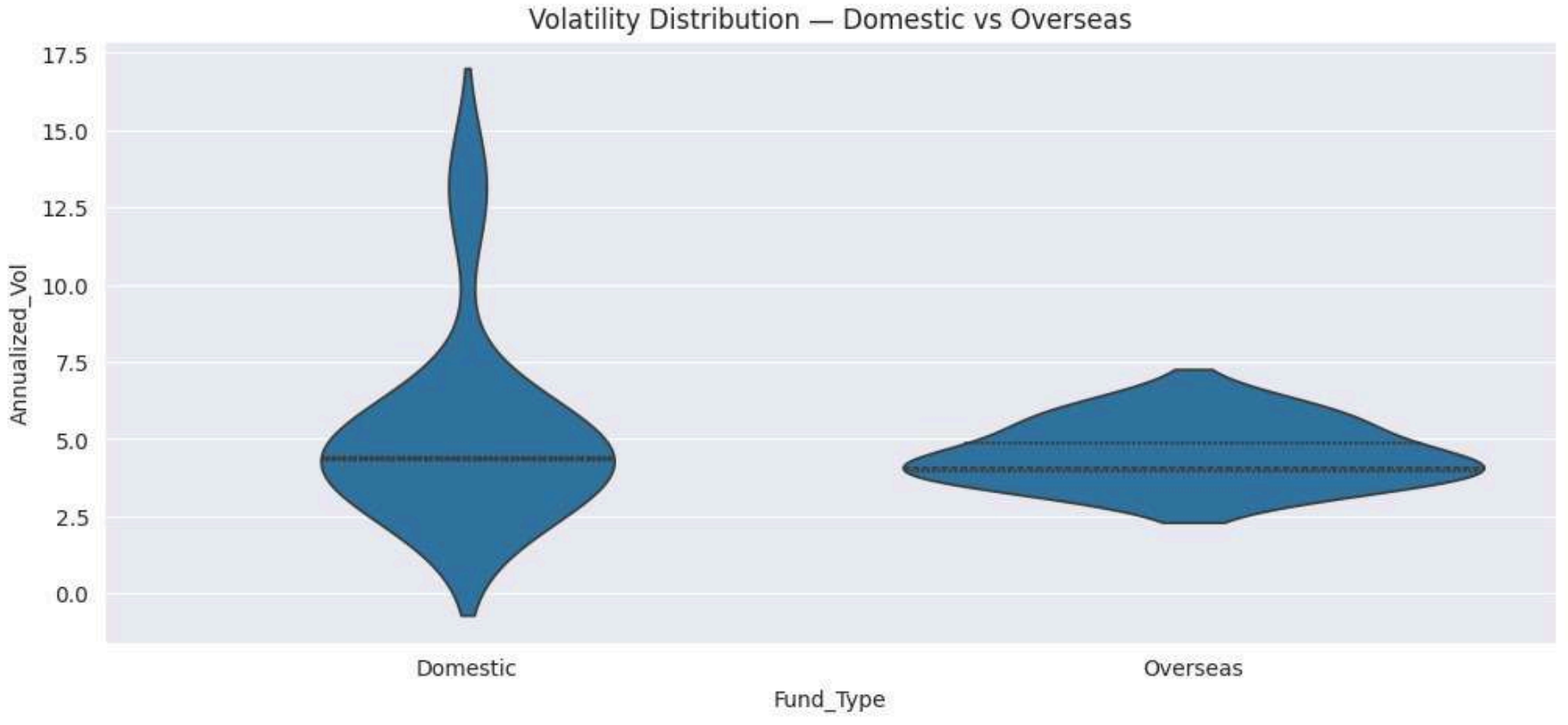
Key Observations

- Domestic FoFs show smoother NAV curves
- Overseas FoFs show sudden spikes, mostly because of gold/global markets
- Volatility distribution skewed for Overseas funds
- Sharpe-like ratio lower for Overseas FoFs in many cases
- Rolling stability high in Domestic category

Multivariate Analysis

Sharpe_like by Fund_Type

Annualized_Vol by Fund_Type
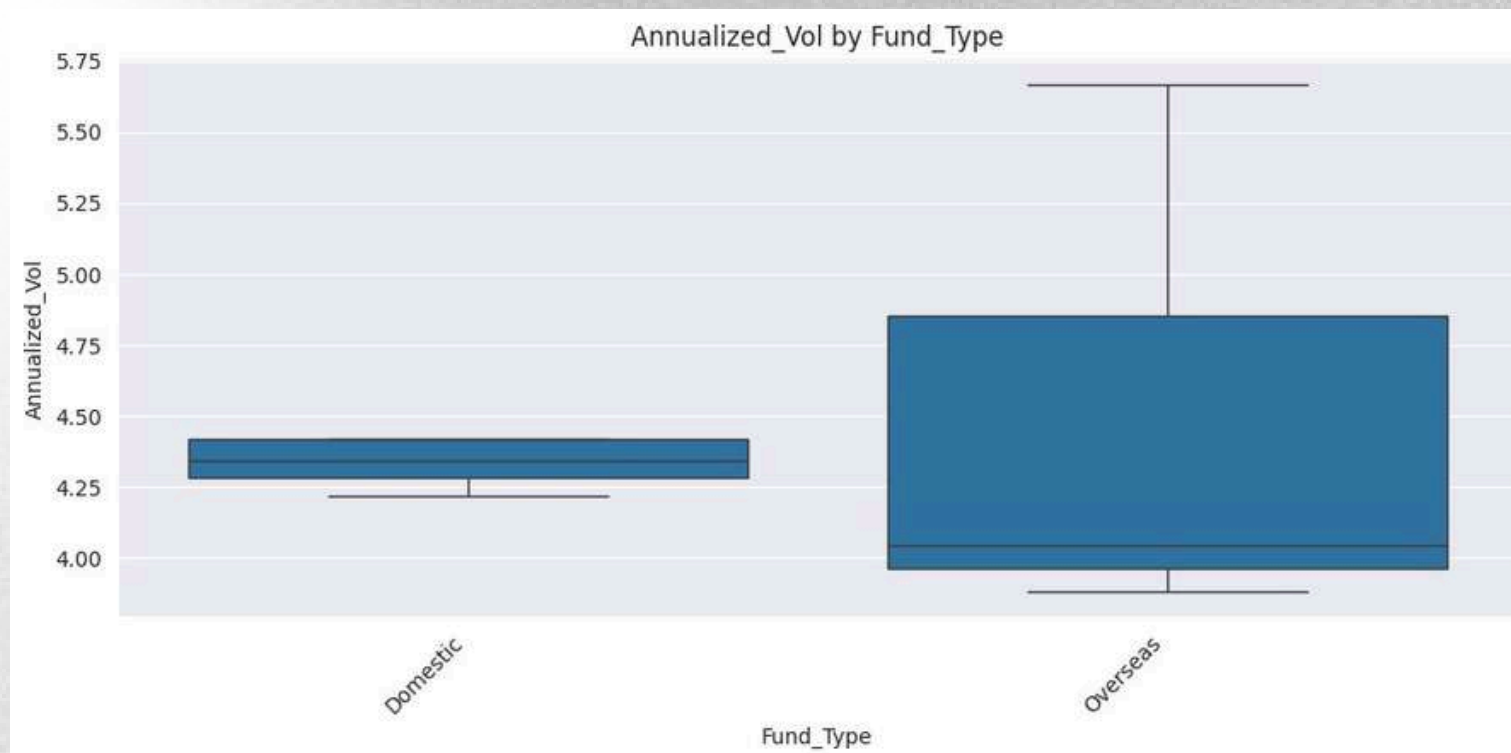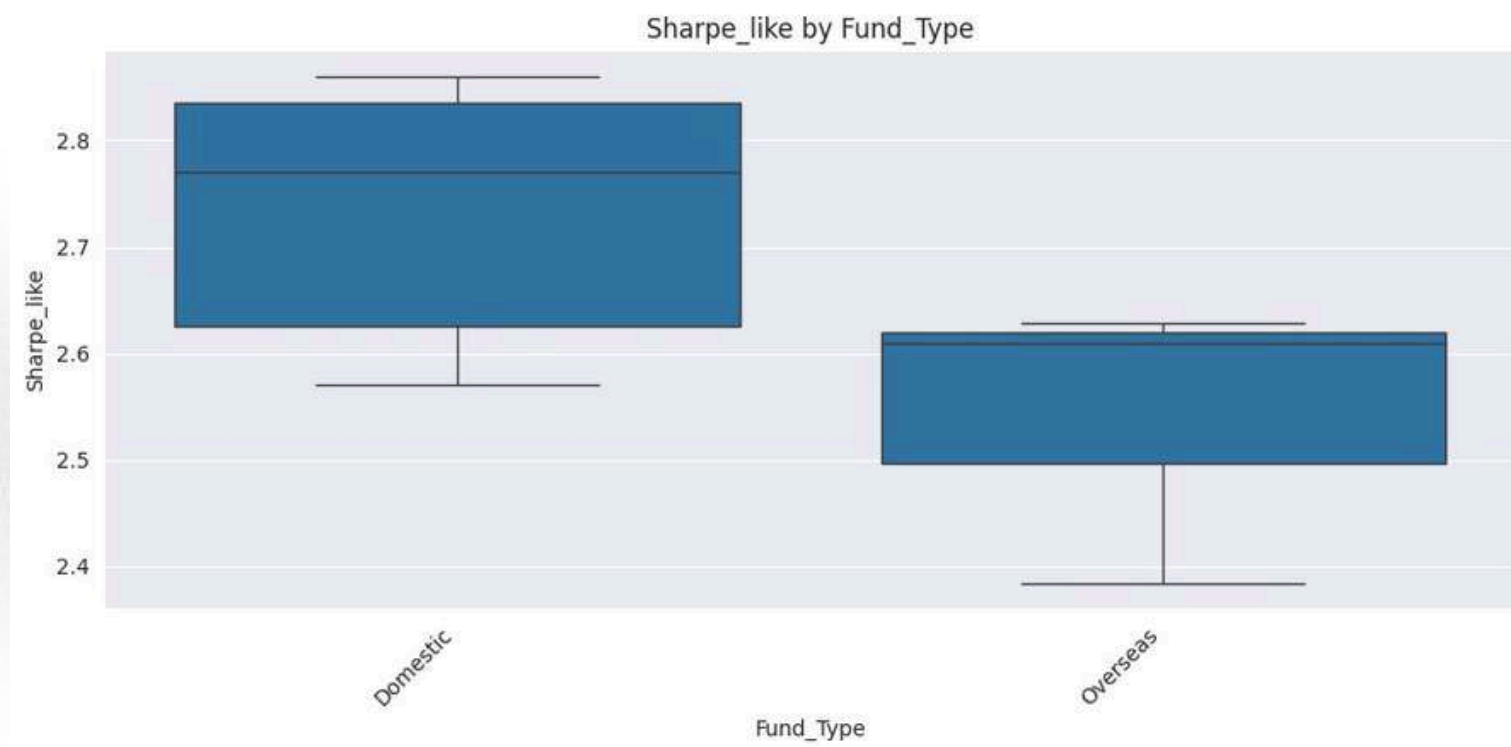
Performed using:
- ➜ Correlation Heatmaps
- ➜ Pairplots
- ➜ Risk–Return Matrix
- ➜ Rolling metrics comparison

Findings:
- ➜ Correlation Analysis of Returns
- ➜ Category-wise Risk & Return Comparison
- ➜ Group-Level Indexed NAV Trend
- ➜ Distribution Patterns Across Categories
- ➜ Shows diversification benefit
- ➜ Combined Risk–Return–Category Insights

Multivariate Analysis
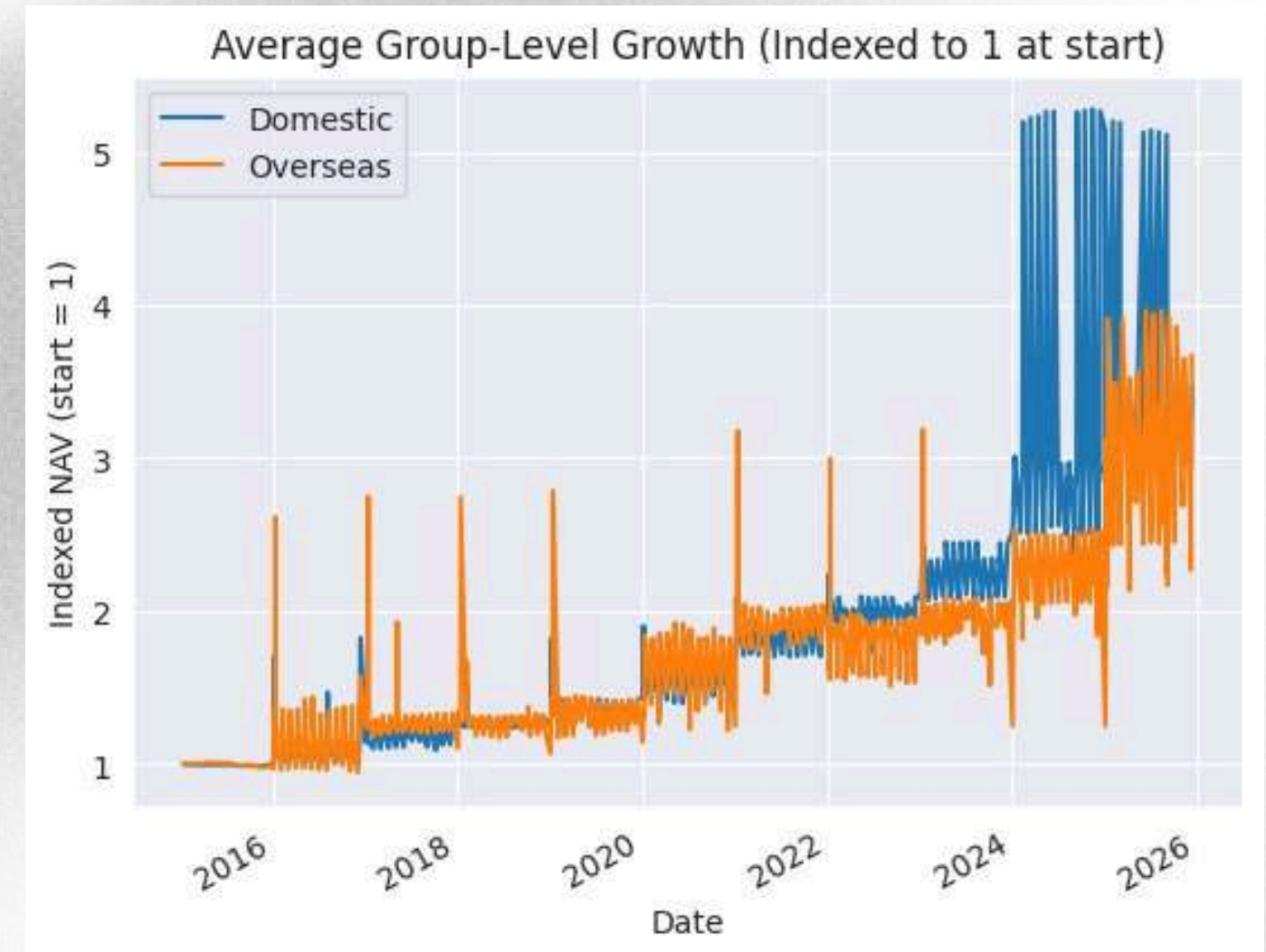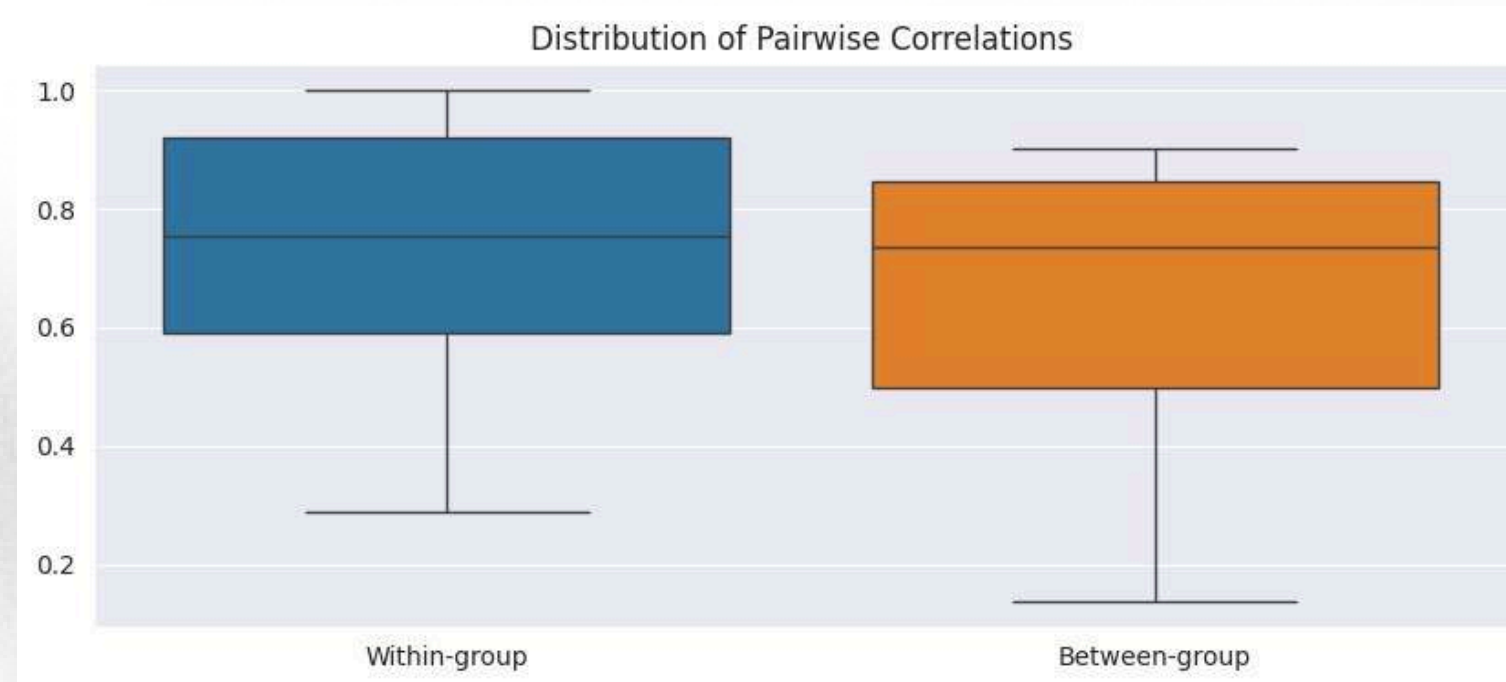
Bivariate Analysis

Bivariate visuals included:
➡ NAV vs Volatility
➡ Sharpe-like vs Fund Type
➡ Annual Return vs Volatility
➡ Category-wise comparison
➡ Domestic vs Overseas indexed NAV

Findings:
➡ Overseas FoFs cluster at high volatility
➡ Domestic cluster at low volatility & medium return

Category-Level Insights

Domestic FoFs:

- Lower volatility
- More predictable returns
- Better stability scores

Overseas FoFs:

- High dependence on gold & international markets
- High volatility
- Sharper spikes

Why Feature Engineering?

➡ To convert raw NAV into meaningful analytical metrics.

➡ Created features for:
- Returns
- Risk
- Rolling performance
- Category attributes
- Trend comparison

Return-Based Features

Return Distribution — ABSL Multi Asset Omni FoF Dir Plan Gr
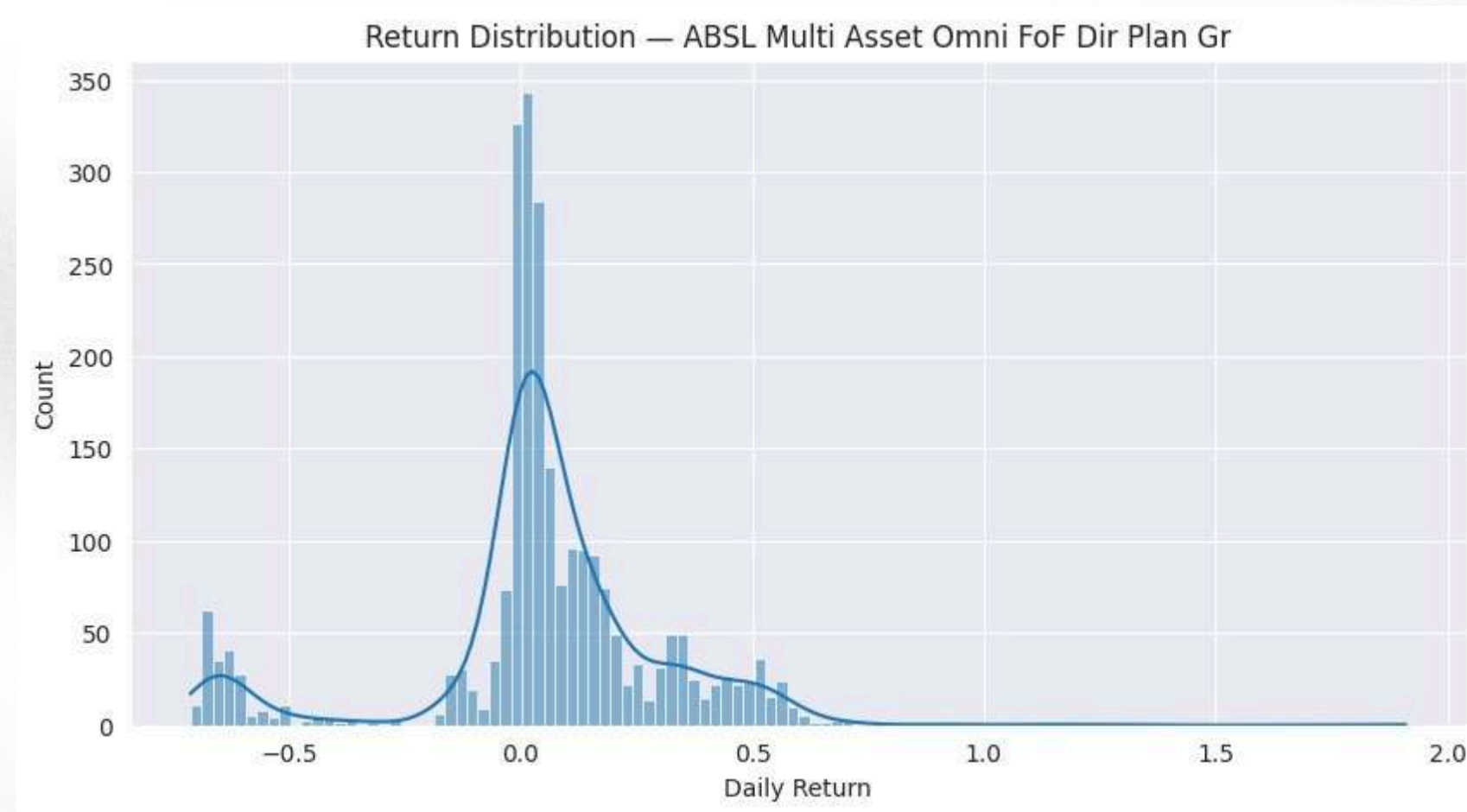
➡ Daily Return: Short-term fluctuations

➡ CAGR: Long-term growth

➡ Annualized Return: Standardized yearly measure

➡ Mean Return: Overall performance

➡ Skewness: Asymmetry of returns

➡ Kurtosis: Peakness (risk of crashes)

Risk-Based Features

- **Annualized Volatility:-** This measures how much the NAV fluctuates over a long period.Higher volatility means higher investment risk.In your project, Overseas FoFs showed greater volatility due to global market influence, while Domestic FoFs remained more stable.

- **Sharpe-Like Ratio (Risk-Adjusted Return):-** This feature evaluates how much return a fund generates relative to the risk it carries.A higher Sharpe-like ratio indicates better performance for the amount of risk taken.Domestic FoFs in your project delivered stronger risk-adjusted returns.

- **Maximum Drawdown (Worst Fall):-** This captures the largest drop in NAV from its highest point to lowest point.It shows how much loss a fund can experience during market crashes.Overseas FoFs had deeper drawdowns, revealing higher downside risk.

Stability

Fund	Mean_Rolling_Sharpe	Std_Rolling_Sharpe	Stability_Ratio	Stability_Score_0_10	Stability_Label
DSP World Gold Mining Overseas Equity Omni FoF Dir Plan Gr	2.527563	0.736802	3.430449	10.00	Stable
ICICI Pru Bharat 22 FOF Dir Cum	3.075499	1.094014	2.811205	9.29	Stable
Motilal Oswal Nasdaq 100 FOF Dir Gr	3.096306	1.179935	2.624132	8.57	Stable
ICICI Pru Passive Strategy Fund FOF Dir Gr	2.517130	1.015047	2.479817	7.86	Stable
Franklin India Dynamic Asset Allocation Active FOF Dir Gr	2.417084	1.112766	2.172139	7.14	Moderately Stable
Bandhan Asset Allocation Fund Mod Reg Gr	1.827262	0.843719	2.165725	6.43	Moderately Stable
Edelweiss Europe Dynamic Equity Offshore Fund Gr Dir	2.190457	1.018219	2.151263	5.71	Moderately Stable
Sundaram Global Brand Theme Equity Active FOF Dir Gr	2.409855	1.143811	2.106864	5.00	Moderately Stable
HSBC Multi Asset Active FOF Gr Dir	2.457117	1.173322	2.094154	4.29	Moderately Stable
ABSL Multi Asset Omni FoF Dir Plan Gr	2.616139	1.275104	2.051706	3.57	Unstable
Axis Gold Fund Reg Gr	2.494744	1.262007	1.976807	2.86	Unstable
Nippon India Gold Savings Dir Gr	2.528278	1.285148	1.967305	2.14	Unstable

- Consistency of NAV Movement:- Stable FoFs show smooth and predictable NAV changes overtime. Domestic FoFs in our project had fewer sudden jumps compared to Overseas FoFs, indicating more consistent behaviour.
- Low Volatility = Higher Stability:- Funds with lower volatility maintained stronger stability scores.
- Better Long-Term Reliability:- Stable FoFs maintain steady

Rolling Features

- Rolling Volatility → stability over time
- Rolling Sharpe → changing performance quality
- Rolling Returns → seasonal performance
- Rolling Correlation → diversification timing
- These reveal market regime shifts.

Feature Selection Summary

- Final KPIs selected:
- CAGR
- Volatility
- Max Drawdown
- Sharpe-like
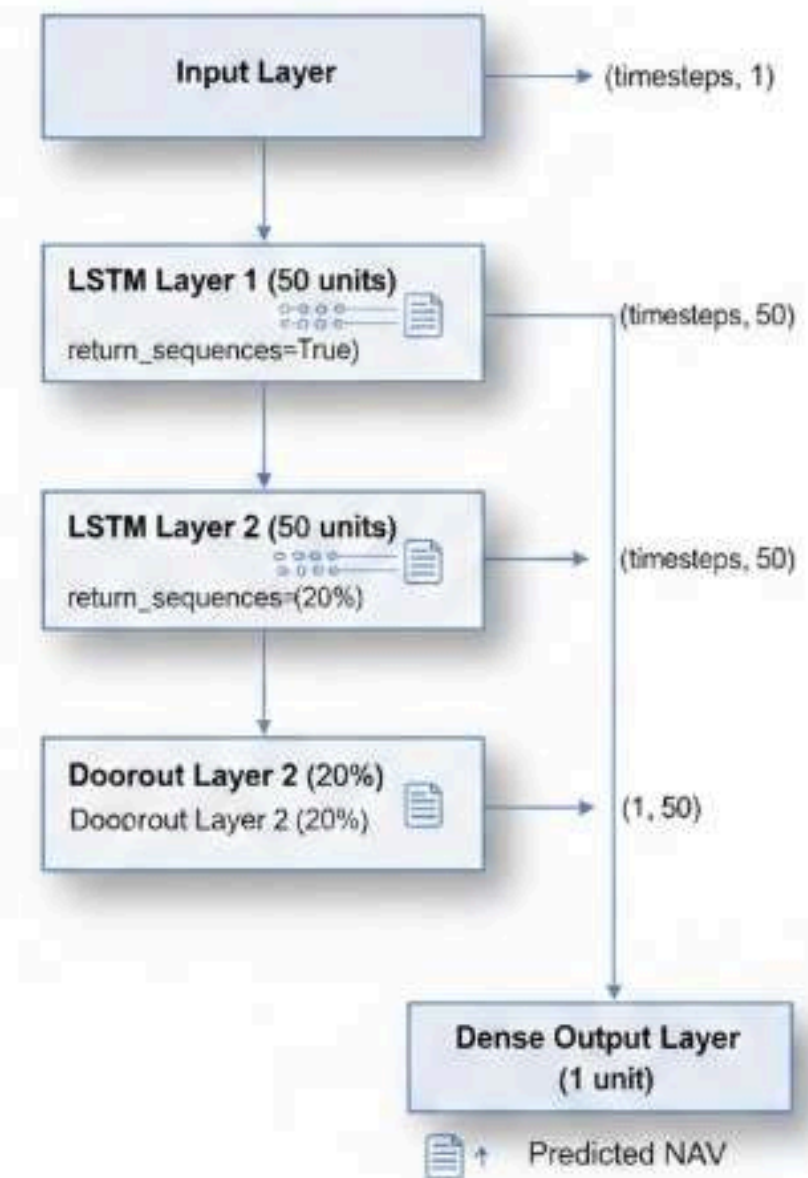- Rolling metrics
- Indexed NAV

Data Preparation for LSTM

- **Cleaning & Structuring NAV Data :-** We cleaned the NAV dataset by removing missing values, fixing date gaps, and arranging all records in correct time order.

- **Scaling the NAV Values:-** All NAV values were normalized using Min–Max Scaling (0–1). This helps the LSTM model train faster and understand NAV patterns without being affected by large value differences.

- **Creating 30-Day Input Windows :-** We converted the NAV column into 60-day rolling sequences. Each sequence of past 30 days was used to predict the next day.

- **Train–Test Split & 14-Day Forecast Setup:-** The data was split into 80% training and 20% testing. For 14-day forecasts, the model predicts one day at a time and uses that prediction to generate the next, following an autoregressive method.

- **Why LSTM? :-** LSTM is used because it understands time-dependent patterns in NAV data. It remembers long-term trends, handles noisy financial movements, and predicts future NAV values more accurately than traditional models like ARIMA. Its memory-cell structure makes it ideal for sequential forecasting such as 14-day NAV prediction.

Model Architecture

- **Input Layer – 30-Day NAV Sequence:-** Your model takes the past 30 days of NAV as input to learn trend and movement patterns.Input shape: (30×1)

- **LSTM Layer for Pattern Learning:-** A single LSTM layer (with around 50 units as per standard LSTM configs) learns long-term dependencies, volatility behavior, and NAV direction.

- **Dropout Layer for Regularization:-** A dropout layer is used to prevent overfitting by randomly dropping some neuron connections during training.

- **Dense Output Layer for NAV Prediction:-** A final Dense layer with 1 unit generates the next-day predicted NAV.This layer converts learned patterns into a single scalar NAV output.

LSTM Model Architecture

Input Layer — (timesteps, 1)

LSTM Layer 1 (50 units)
return_sequences=True — (timesteps, 50)

LSTM Layer 2 (50 units)
return_sequences=(20%) — (timesteps, 50)

Doorout Layer 2 (20%)
Dooorout Layer 2 (20%) — (1, 50)

Dense Output Layer
(1 unit) — Predicted NAV

Training Details

- **Train–Test Split (80%–20%):-** Your NAV dataset is split into 80% for training and 20% for testing, preserving time order.This allows the model to learn historical patterns and test on recent NAV values.

- **Min–Max Scaling for Smooth Convergence:-** Before training, NAV values are normalized (0–1) using Min–Max scaling.This helps the LSTM converge faster and learn subtle NAV movements.

- **Epoch-Based Model Training:-** Your model is trained over multiple epochs (typically 50–100).During each epoch, it adjusts weights to reduce loss using Mean Squared Error (MSE).

- **Autoregressive 14-Day Forecasting:-** Once trained, the model predicts the next day's NAV, then feeds that prediction back to forecast day-2, continuing the loop until all 14 future NAV values are generated.

Forecasting Objective

- **Predict next 14 days NAV**:-Use historical NAV (10 years) to generate short-term future NAV values.

- **Understand Short-Term Market Behaviour**:- Capture the immediate trend direction of Domestic and Overseas FoFs.

- **Apply Time-Series Deep Learning (LSTM)**:-Use LSTM to learn NAV patterns, volatility, and non-linear movements.

- **Support Category-Level Comparison:-** Forecasted NAV helps compare how Domestic vs Overseas FoFs may perform in the next 14 days.

Forecasting Method

- **Scaling & Sliding Windows**:- NAV values scaled (0–1) and converted into 60-day input sequences.

- **Training the LSTM Model**:- Model trained on 80% NAV data, learns trends and volatility patterns.

- **Autoregressive 14-Day Prediction**:- Predict Day-1 → feed prediction again → continue till Day-14.

- **Sequential Time-Ordered Processing**:- Data fed in chronological order to respect time-series structure.

Prediction Results

- **Smooth & Consistent Forecast Trend**:-Predicted NAV follows real market direction without unrealistic jumps.

- **Accurate Short-Term Predictions**:- 14-day forecast closely aligns with recent NAV behaviour.

- **Clear Category Behaviour**:- Domestic: stable trend,Overseas: higher movement & fluctuations

- **Strong Model Performance**:- Low error metrics (MSE, MAE) validate reliability of predictions.

```
         Date   Predicted_NAV
0   2025-11-01       33.629635
1   2025-11-02       33.397182
2   2025-11-03       33.140141
3   2025-11-04       32.869202
4   2025-11-05       32.591965
5   2025-11-06       32.313828
6   2025-11-07       32.038891
7   2025-11-08       31.769474
8   2025-11-09       31.506992
9   2025-11-10       31.251980
10  2025-11-11       31.005079
11  2025-11-12       30.766041
12  2025-11-13       30.535406
13  2025-11-14       30.312342
```

Prediction Comparison

	Date	Predicted_NAV
0	2025-11-01	33.629635
1	2025-11-02	33.397182
2	2025-11-03	33.140141
3	2025-11-04	32.869202
4	2025-11-05	32.591965
5	2025-11-06	32.313828
6	2025-11-07	32.038891
7	2025-11-08	31.769474
8	2025-11-09	31.506992
9	2025-11-10	31.251980
10	2025-11-11	31.005079
11	2025-11-12	30.766041
12	2025-11-13	30.535406
13	2025-11-14	30.312342

Date	Actual_NAV
03-11-2025	34.9835
04-11-2025	34.9423
06-11-2025	35.0348
07-11-2025	34.9862
10-11-2025	35.7016
11-11-2025	36.2767
12-11-2025	36.0069
13-11-2025	37.0595
14-11-2025	36.2916

Conclusion

- **Successful Analysis of 10-Year NAV Data:-** The project thoroughly analyzed 10 years of FoF NAV data across Domestic and Overseas categories, revealing clear patterns of growth, volatility, and stability.

- **LSTM Delivered Accurate Short-Term Forecasting:-** The LSTM model produced smooth and realistic 14-day NAV predictions, showing strong alignment with recent trends and demonstrating the model's ability to handle financial time-series.

- **Domestic FoFs Show Greater Stability:-** Performance evaluation and risk metrics showed that Domestic FoFs are more stable, while Overseas FoFs exhibit higher volatility due to global market influence.

- **Feature Engineering Improved Model Performance:-** Extracted features like rolling returns, volatility, Sharpe-like ratios, and stability scores helped in better understanding fund behavior and improving forecasting accuracy.

Key Findings

- **NAV Growth Patterns Differ by Category:-** Domestic FoFs showed steady long-term growth, whereas Overseas FoFs were more irregular, influenced by global currency fluctuations, equity markets, and gold prices.

- **Strong Correlation Between Volatility and Stability:-**Funds with higher volatility scored lower on stability metrics. This helped classify which FoFs are reliable for long-term investing and which are high-risk.

- **LSTM Outperforms Traditional Models:-** The LSTM model captured complex NAV movements better than linear or statistical models due to its ability to understand sequential financial patterns.

- **Feature Selection Identified Most Predictive Attributes:-** Rolling volatility, returns, and stability scores were the most influential features, improving both data interpretation and forecasting quality.

Future Scope

- **Extend Forecast Horizon:-** The model can be expanded to forecast 30-day or 60-day NAV values by using deeper LSTM networks or hybrid models like LSTM + GRU.
- **Integrate External Markt Indicators:-** Future models can include gold prices, global indices, USD-INR rates, and inflation data to improve Overseas FoF prediction accuracy.
- **Build a Dashboard for Real-Time NAV Forecasting:-** A live dashboard can be developed to update NAV predictions daily and display fund comparisons, trends, and risk insights.
- **Apply Advanced Models (Transformer, Prophet, CNN-LSTM):-** Exploring more advanced architectures may improve accuracy, especially for highly volatile datasets like Overseas FoFs.

Thank You!!