Monthly Mutual Fund Data Compilation and Analysis

Internship Report

Name: - Pragya Kartik

School: - GEMS Our Own English High School, Dubai

Internship Duration: - 9/07/2025 – 28/07/2025

PROJECT OVERVIEW

This project involved analysing historical Average Assets Under Management (AAUM) in the mutual fund industry, with the goal of identifying patterns and building short-term forecasts. I first used Microsoft Excel extensively to clean, organize, and categorize raw distributor-level investment data from various fund houses across April, May, and June 2025. Once structured, the data was used in R for time-series forecasting and visualization to uncover key industry trends and estimate future AUM movements.

Key Excel Tools and Techniques Used:

1) Multi-level Sorting & Filtering: Applied layered filters to isolate relevant combinations of fund houses, categories, and investment types for cleaner analysis.

2) Custom Formulas with Conditional Logic: Used IF, SUMIFS, and INDEX-MATCH to derive meaningful summaries from unstructured data without needing pivot tables.

3) Data Cleaning Techniques: Standardized inconsistent naming, removed duplicates, handled missing values, and prepared raw CSV sheets for streamlined analysis.

Core R Skills Acquired

I immersed myself in R's capabilities, learning to:

- Work with fundamental data structures like data frames.
- Handle diverse data types, especially Date objects, which are vital for time-series analysis.
- Perform initial data inspections using functions like head() and glimpse().

Key Libraries and Functions Utilized

I extensively used several powerful R libraries to manage and analyze the data:

- **tidyverse (especially dplyr and tidyr):** This suite was instrumental for data manipulation. I learned to:
    - mutate(): Create and modify columns (e.g., converting text to uppercase, cleaning numeric values).
    - filter(): Select specific rows based on conditions (e.g., removing incomplete data or NA values).
    - group_by() and summarise(): Aggregate data by categories like Fund House and Month, calculating sums.
    - The **Pipe Operator (%>%)**: Allowing me to chain complex data operations in a clear, sequential flow.
    - pivot_longer(): Crucial for restructuring distributor AUM data for analysis.
- **lubridate:** Essential for date handling, I used dmy() to parse dates and floor_date() to standardize them to month starts for aggregation.
- **ggplot2:** I learned to create professional-quality line plots to visualize trends, customizing every aspect from axes to legends.

- **scales & ggrepel:** These helped refine plot aesthetics and label readability.

Forecasting Development

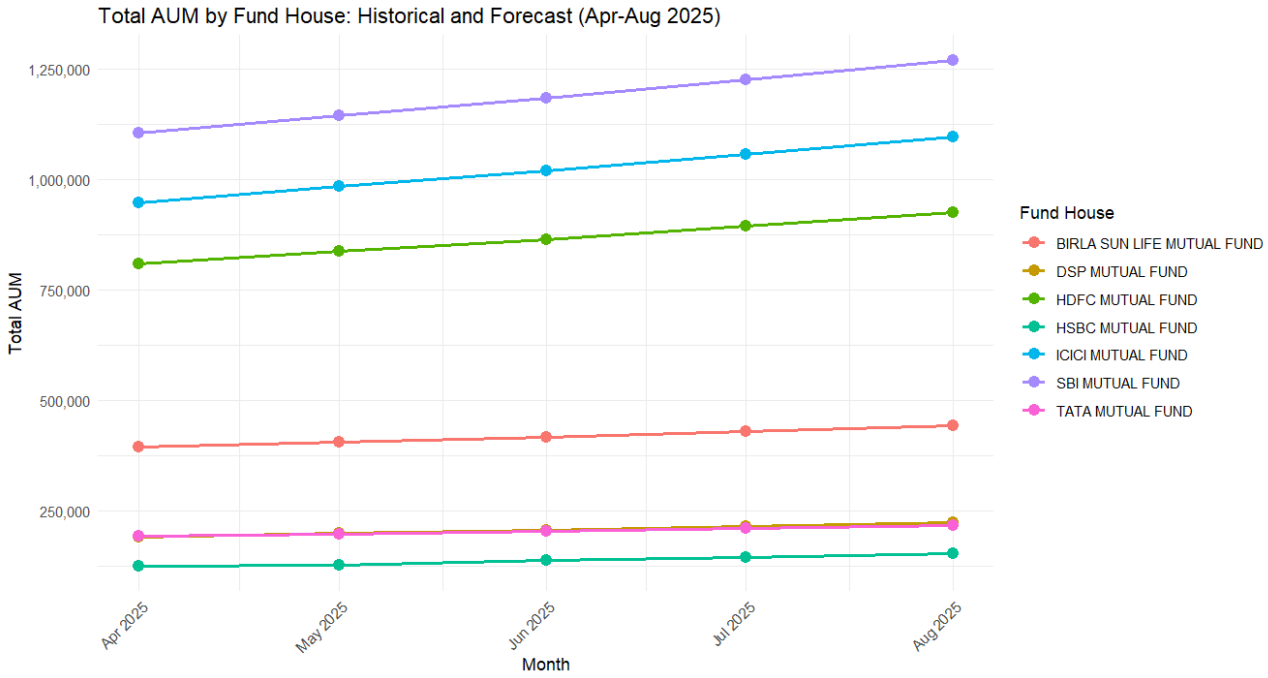
A core achievement was developing the forecasting methodology in R. I calculated the average monthly growth rate for each category (Fund House, Distributor Type, Scheme Category, Scheme Sub-Category) based on the historical April-June 2025 data.

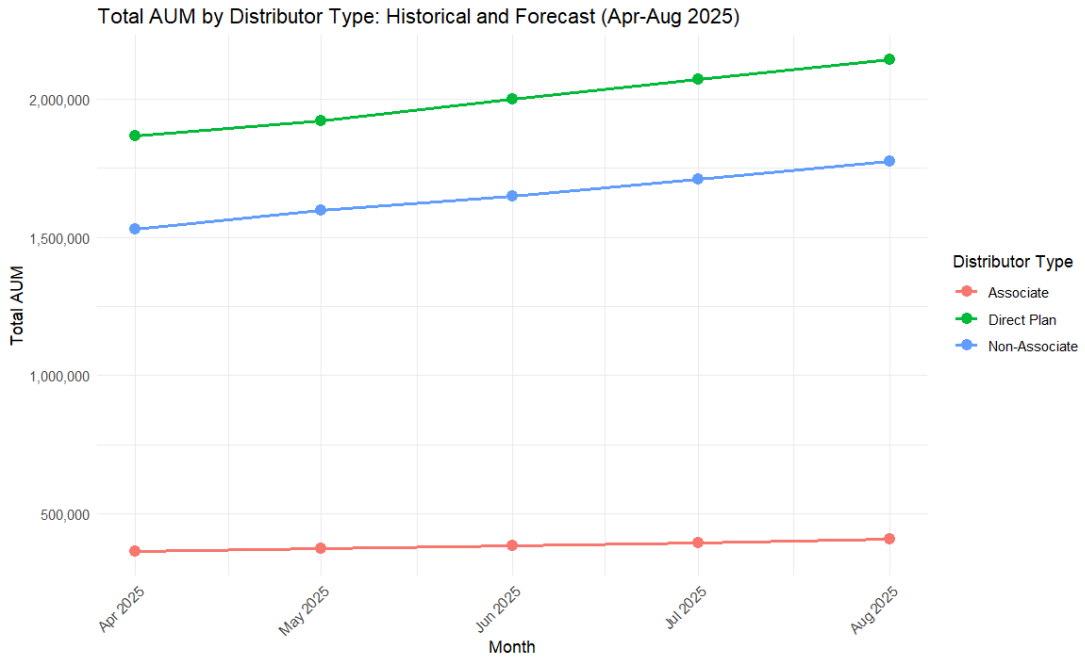
This growth rate was then iteratively applied to project AUM for July and August 2025, assuming recent trends would continue. This allowed me to generate tailored short-term forecasts for each segment.

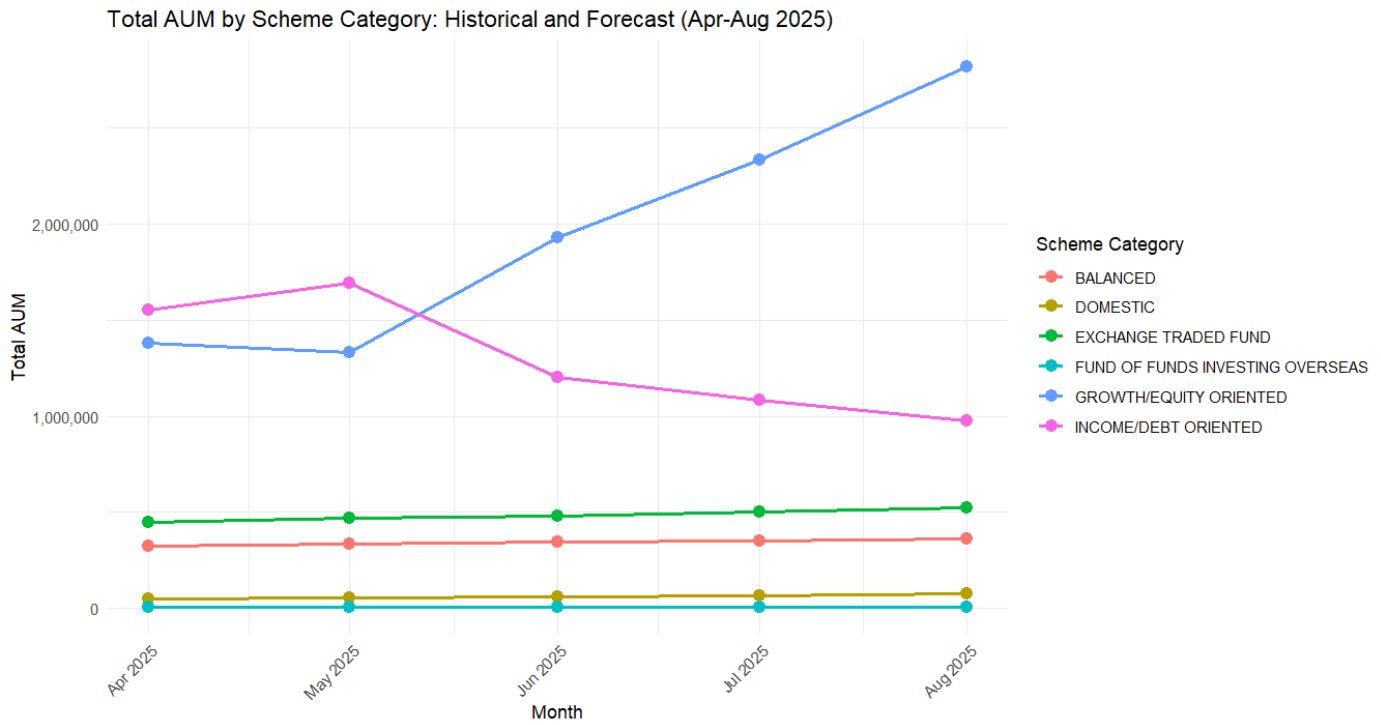
Conclusion

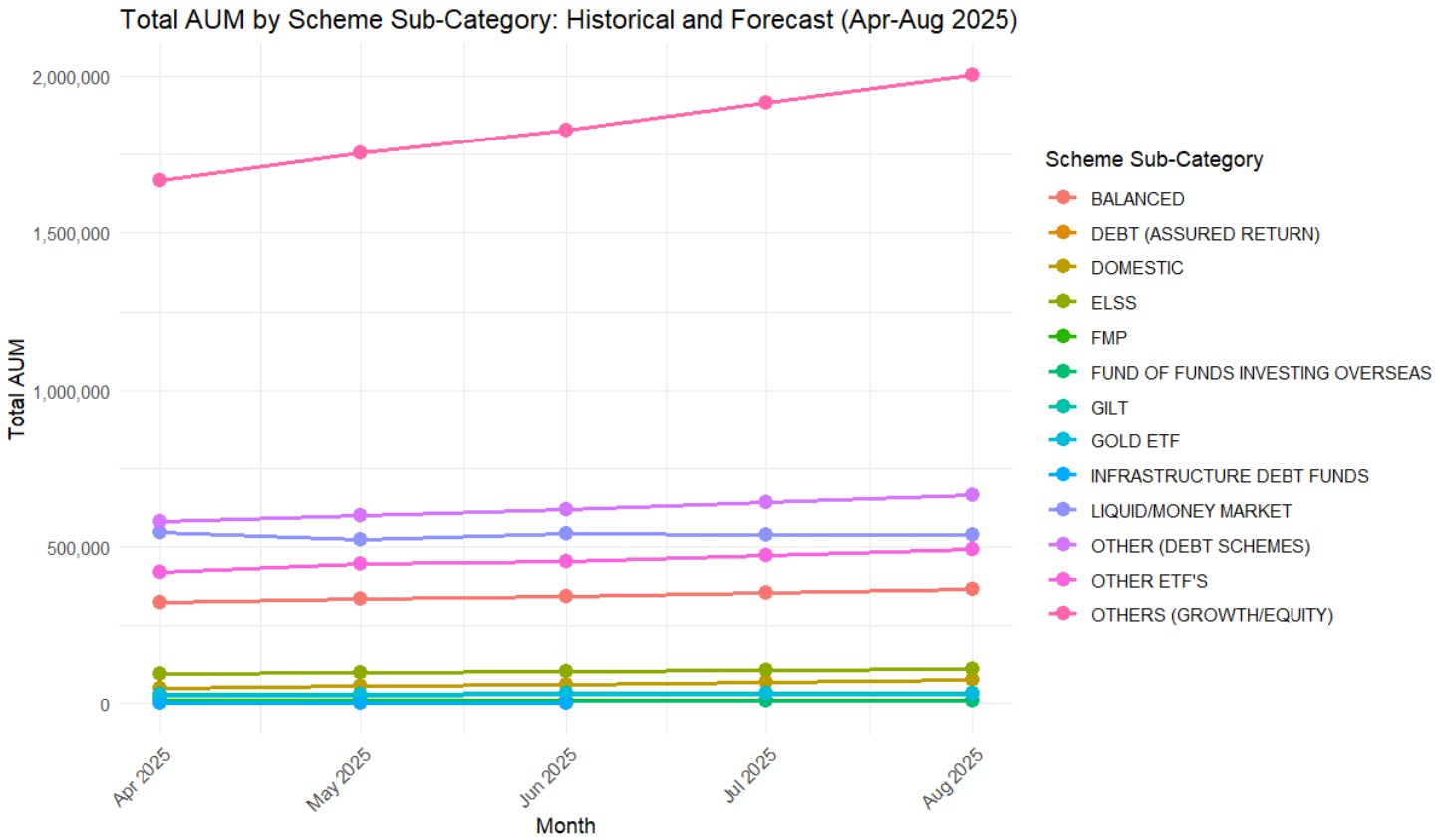
This internship project significantly enhanced my practical skills in data analysis and forecasting using R. I gained hands-on expertise in critical data manipulation, effective visualization, and developing a practical forecasting model. This experience has deepened my understanding of data-driven insights within the financial sector.

KEY VISUALISATIONS: AUM TRENDS AND FORECASTS









GitHub Repository Link: -

https://github.com/PragyaKartik237/Mutual-Fund-AUM-Analysis