WORKFLOW: MUTUAL FUND RECOMMENDATION SYSTEM

🔷 PHASE 0: USER INPUTS & GOALS

NEW INVESTORS:

- **Budget** (NEW)
- **Risk Tolerance**: Low / Medium / High
- **Fund Type Preference**: Equity / Debt / Hybrid / ELSS etc.
- **Market Cap Preference**: Large / Mid / Small / No Preference etc.

EXISTING INVESTORS:

- Fund Name / Scheme ID
- NAV at time of purchase
- Units held
- Purchase Date

PHASE 1: DATA LOADING & CLEANING

Purpose: Ensure clean and usable NAV data
Actions:

- Load data using pandas
- Convert:
    - `Date` to datetime

    - `NAV` to float
- Check and handle nulls:
    - Drop rows with missing NAV, Scheme, or Fund House
- Sort data by SchemeID and Date
- Remove duplicates if any

Run null checks after each operation to ensure integrity

PHASE 2: FEATURE ENGINEERING

Return Calculations:

- Group by SchemeID and compute:
    - Daily Return: % change in NAV
    - Monthly, Quarterly, Yearly returns using `.resample('M')`, `.resample('Q')`, `.resample('Y')`
    - CAGR for 1Y, 3Y, 5Y

Risk Metrics:

- Standard Deviation (Monthly, Quarterly, Yearly)
- Sharpe Ratio
- Sortino Ratio (only downside std)
- Maximum Drawdown
- Volatility (rolling std)

Trailing Returns:

- Use CAGR formula for 1Y, 3Y, 5Y trailing periods

PHASE 2.1: FUND CLASSIFICATION

Add **Market Cap** & **Asset Class**

- Use Scheme name keywords:
  - MarketCap: "Large Cap", "Mid Cap", "Small Cap" → Column: `MarketCap`
  - AssetClass: "Equity", "Debt", "Hybrid", "ELSS" → Column: `AssetClass`

PHASE 3: DATA ENCODING & TRANSFORMATION

Label Encoding:

- Fund House, Scheme Name, AssetClass, MarketCap

Scaling & Normalization:

- Apply **StandardScaler** or **MinMaxScaler** on:
  - NAV, Returns, Sharpe, Drawdown, etc.
- Should be done **after feature engineering** is complete

PHASE 4: OUTLIER DETECTION

Don't remove — just flag for reference

- Use **Z-score method** for:
  - Return

    - Sharpe
    - Drawdown

Mark rows as outliers using an "is_outlier" flag column

PHASE 5: RISK PROFILING & CLUSTERING

Cluster funds by risk level

- Use **KMeans** / Agglomerative Clustering
- Features: Sharpe, Std Dev, Max Drawdown, Return
- Assign Risk Cluster: Low / Medium / High

PHASE 6: RECOMMENDATION ENGINE

For New Investors:

- Filter funds using:
    - FundType, MarketCap, RiskCluster, InvestmentGoal, Budget (NEW)
- Rank by:
    - Sharpe ↑
    - 3Y/5Y Return ↑
    - Drawdown ↓
- Show latest NAV

For Existing Investors:

- Calculate:

    - Current Value = Latest NAV × Units Held
    - CAGR
- Find similar category funds with better performance
- If better fund found:
    - Suggest switch (with reason)
- Else:
    - Suggest hold

PHASE 7: STREAMLIT APP STRUCTURE

Input Page:

- Dynamic form: New vs Existing investor
- Upload CSV (for portfolio analysis)

Output Page:
For New Investor:

- Fund Name
- Latest NAV
- Units to purchase

For Existing Investor:

- Current FundID
- Scheme Name
- NAV at purchase

- Latest NAV
- Top similar fund name
- Recommendation
- Reason