Readme:

The objective of the exercise is to find a tradable signal in the dataset.

The dataset contains sectoral data for 3 separate types of investments made in the US (no international funds are included) and represents 60-70% of activity in the overall asset classes for that week (not all investors report data at the same time). The weekly data spans 10 years from 2006 through end-Jan 2017. Not all sectors have data available for all the dates since new investment vehicles are introduced at various points in time.

1. Institutional Mutual Fund Holdings (investments made/redeemed by institutional investors like Fidelity, Vanguard on behalf of institutions like CALPERS etc)

2. Retail Mutual Fund Holdings (investments made by individuals in their portfolios)

3. ETF (Exchange Traded Funds - institutional investors)

This dataset can be used in isolation - i.e. no additional information is supplied to the models and the models only use the data in these sets to identify a signal. Alternatively, the dataset can be augmented by including information from sources like the stock market, macro indicators from sources like the Fed, Bureau of Labor Statistics, etc.

Modeling Techniques:

Standard time series models can be used or non-parametric models like neural networks can be deployed. Suggested course of actions is as follows:

1. Understand the dataset and what it represents

2. Define the problem: what does 'a tradable signal' mean based on the data available (i.e. develop your modeling equation, if any)

3. Start with simple analysis to identify patterns and outliers

4. Develop models with increasing complexity - i.e. don't start with neural nets etc

5. Define your success criteria - i.e. when do you know you have a tradable signal - positive (buy)/negative (sell), both are signals

Data fields:

ReportDate: Weekly data aggregated and released every Wednesday

AssetClass: Industry/Sector/Asset Class

Flow: Amount of positive (inflow) or negative (outflow) in Millions of USD

FlowPct: Flows as percent of assets at beginning of the week

AssetsEnd: Assets at end of the week in Millions of USD

PortfolioChangePct: Percent change in overall portfolio during the week