DATA ANALYTICS

TEAM 1

CONTENTS ››››

INTRODUCTION

Preface:

A **mutual fund** is a professionally-managed investment vehicle, usually run by an asset management company that brings together a group of investors and permeates their money in stocks, bonds and other securities. The company administering the mutual fund scheme is referred to as a fund house. It invests pooled funds of retail and institutional investors in equity, fixed income or other such securities in accordance with the stated investment objective of the fund.

Mutual funds are classified by their structure as **Open-ended** funds and **Close-ended** funds. In the former, investors can buy and sell the units from the fund at any point of time while in the latter the fund money is raised from investors only once. Subsequently, if the fund is listed on a stock exchange then the units can be traded like stocks.

The characterizing features of mutual funds are **Assets Under Management** (AUM), and **Net Asset Value** (NAV). Assets under management (AUM) is the total market value of assets that an investment company or financial institution manages on behalf of investors. Net asset value (NAV) is the value per share of a mutual fund on a specific date or time.

$$NAV = \frac{\text{Assets - Liabilities}}{\text{Number of Outstanding Shares}}$$

The average Assets Under Management (AUM) of Indian Mutual Fund Industry for the month of December 2017 stood at ₹ 22.60 lakh crore. The Mutual Funds industry tends to be considerably dynamic as represented by the fact that it's associated values tend to vary on an hourly basis. The following report attempts to define a comprehensive methodology of rating the various fund houses in India by encompassing all the aforementioned facets.

Objective:

The objective of this analysis is to compare the performance of the given Mutual Fund Houses on the basis of the following pre-eminent factors:

1. Asset Sizes or AUM (Assets under Management)
2. AUM Growth
3. NAV (Net Asset Value) of Funds or NAV per unit
4. NAV or NAV per unit growth

Furthermore, an attempt has been made to encompass the impact of other influencing factors that pertain to the nature of the funds, such as age of the fund, age of the fund house, minimum investment etc. Macroeconomic parameters such as GDP, CPI, Nifty prices, Exchange Rate have also been associated with fluctuations in the markets, which makes them another entity that should have a role in the eventual comparison.

ABOUT DATA

Data provided:

1. Text files constituting the details of Open Ended Schemes of the given Mutual Fund Houses. The Scheme Code, Scheme Name, Net Asset Value, Repurchase Price, Sale Price and Date for each scheme has been provided.

2. Excel sheet contains data for 34 mutual funds. Details include name of Fund House, Fund Class, Plan, Options, Riskometer, Crisil Mutual Fund Rank, Sector Allocation Size, Sector wise name and percentage.

Data extracted:

The values of the additional variables utilized in the analysis such as AUM, GDP, CPI, Nifty, Exchange rate (rupees vs $) has been crawled using “selenium” library from the hyperlinks to the webpages given in the problem statement. The data crawled from the aforementioned sources has been used to analyse the effects of macroeconomic parameters.

Dividend Funds Annotation

Certain schemes in the dataset yield dividends, for instance, a scheme may have a constant NAV throughout the span of 5 years, which is contrary to the apparent fact that the return of this scheme comes out to be zero, a fact prudently incorporated in the analysis.

APPROACH

The approach adopted in this analysis attempts to accommodate all relevant factors which influence the performance of the Mutual Fund Houses. Naturally, these factors play an important role in the final rating of these houses. The performance of the Mutual Fund Houses have been measured using a linear weighted model. Consequently, each parameter has been assigned a weight using Analytic Hierarchy Process (AHP).

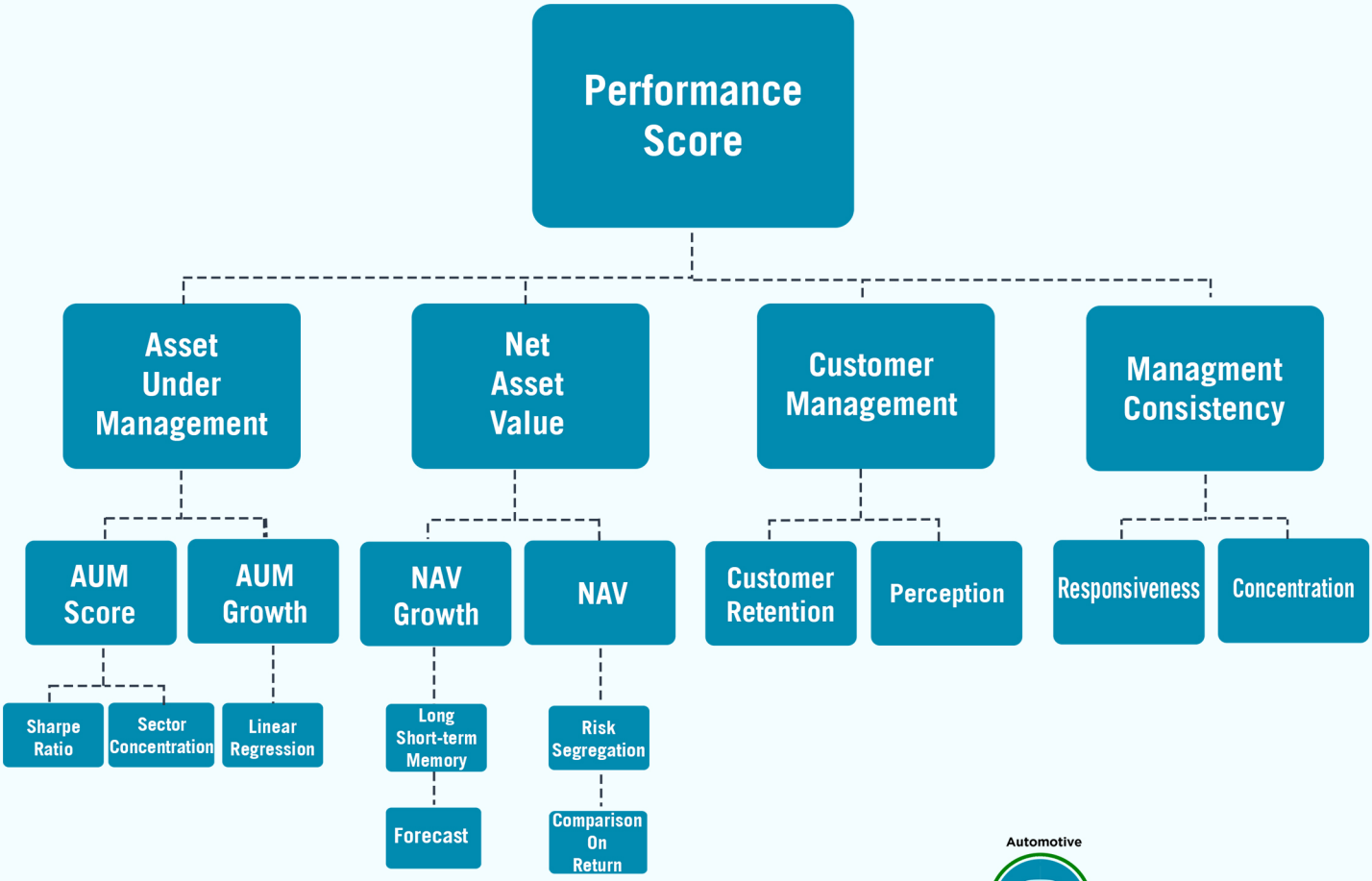
$$\text{Performance score} = \sum (\text{Factors} \times \text{Weights})$$

Factors

>AUM (w=0.12) >AUM Growth (w=0.09) >NAV (w=0.47) >NAV Growth (w=0.35)
>Perception(w=0.09) >Customer Retention(w=0.06) >Management Consistency(w=0.29)

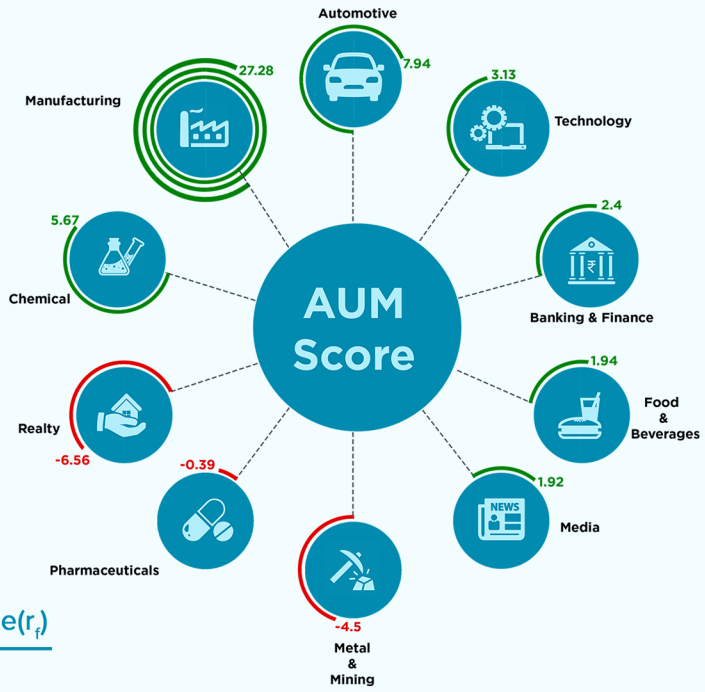
As evidenced by the analysis, the performance metric of the Mutual Fund Houses cannot be solemnly judged on the basis of NAV and AUM (and their growth). Thus, an endeavor has been made to incorporate other key parameters like **Pecrception**, **Customer Retention** and **Management Consistency** to make the prediction more pragmatic.

Overview:



AUM SCORE

The AUM score is a measure of the dexterity with which the company manages its portfolio. It outlines the ability of the fund house to invest in the sectors which give us the higher returns with minimal levels of risk. The industry standard for this is Sharpe Ratio, which is a measure of the same.

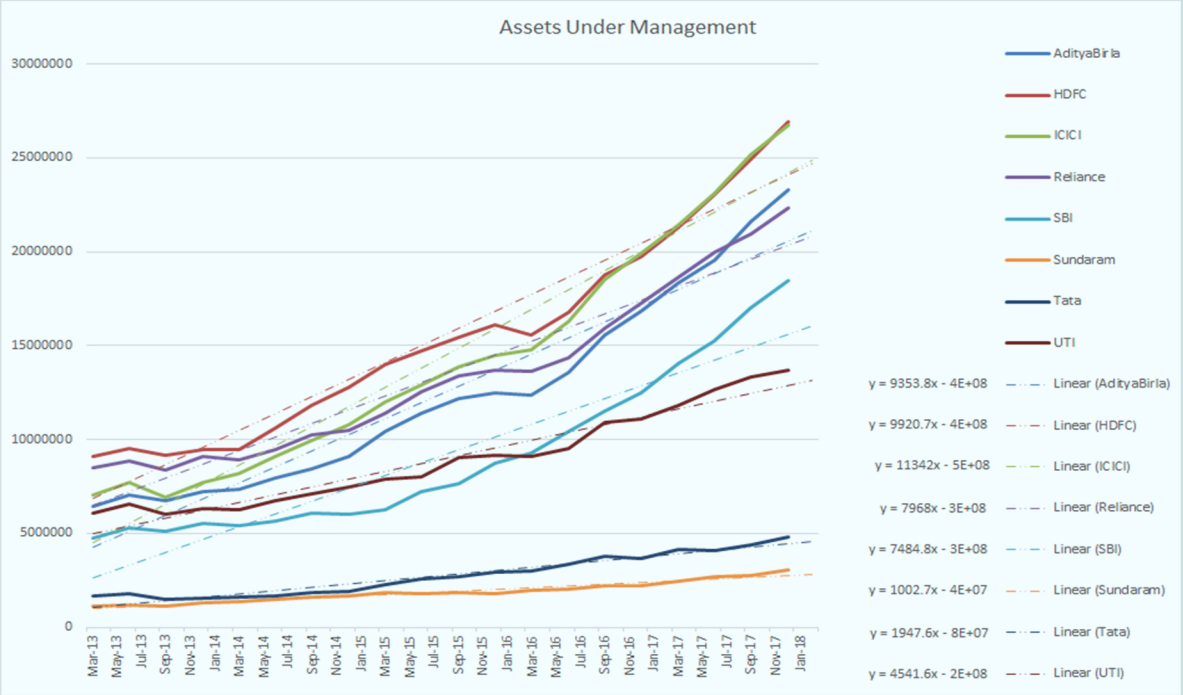


$$\text{Sharpe Ratio} = \frac{\text{Mean Portfolio Return}(r_p) - \text{Riskfree rate}(r_f)}{\text{Standard Deviation of Portfolio}(\sigma_p)}$$

$$\text{AUM}_{\text{Fund House}} = \sum_{\text{All Sectors}} (\text{Sector concentration} * \text{Sharpe Ratio}_{\text{Sector}})$$

AUM Growth:

AUM Growth is a measure of the quarterly increment of Assets Under Management under the Fund House. From the scraped data, the quarterly AUM values of the individual mutual fund schemes have been obtained. The quarterly AUM data of the Mutual Fund Houses is calculated by summing up the AUM values of their respective mutual funds. Subsequently, this data is plotted with respect to time and a plot for each Fund House is created. A best fit linear model is generated for these plots. The slope of this line divided by the mean AUM, is the measure of AUM Growth of a Fund House.



NAV:

Securities of different risks cannot be compared on the basis of returns. For the mitigation of this problem, the individual Mutual Funds of all the houses are segregated on the basis of risk. This is accomplished by clustering the different mutual funds on the basis of their risk which is described by the parameters Alpha, Beta and Volatility.

Alpha - It is a measure of an investment's performance on a risk-adjusted basis.
R_i = the realized return of the portfolio or investment
R_m = the realized return of the appropriate market index
R_f = the risk-free rate of return for the time period
β = the beta of the portfolio of investment with respect to the chosen market index

$$\text{Alpha} = R_i - (R_f + \beta(R_m - R_f))$$

Beta - It is a measure of the volatility of a security or a portfolio in comparison to the market as a whole.

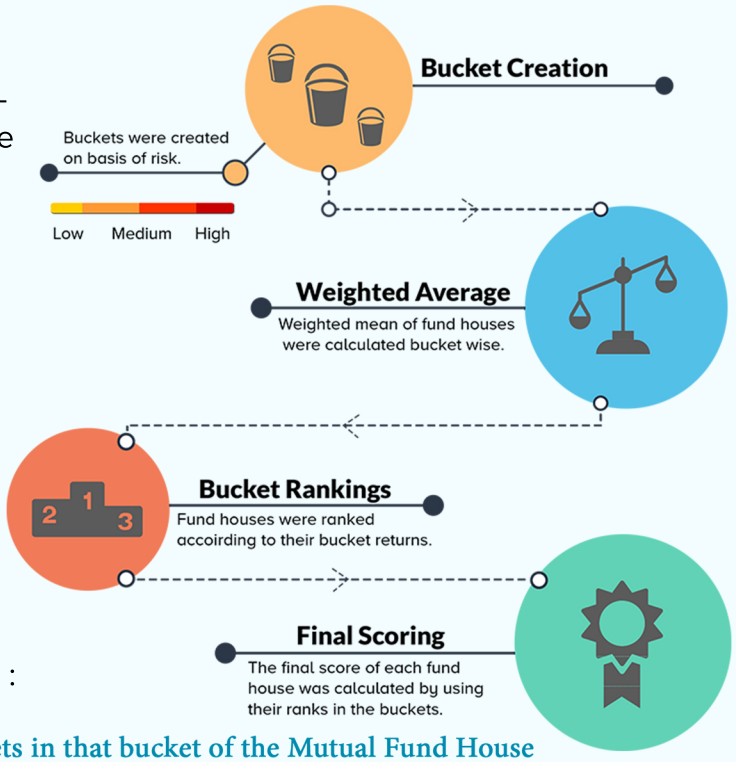
$$\text{Beta} = \frac{\text{Covariance}(r_i, r_m)}{\text{Variance}(r_m)}$$

Standard Deviation - It quantifies the deviation of return on a fund from the expected returns based on its historical performance.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Eventually, the data has been clustered into 3 buckets, after consideration on the basis of their risk which is described by the aforementioned parameters Alpha, Beta and Volatility. Eventually we obtain the buckets:

>High Risk >Medium Risk >Low Risk



It is ensured that the definition of buc kets remains constant for all the mutual fund houses, for instance high risk, is same for all the fund houses. Subsequently, the performance of the individual fund houses are ranked in these buckets (according to their average returns) to ensure that the comparison of the fund houses is on the same standard.

Further, to compare the overall performance of fund houses, the measure defined was as follows :

$$\text{NAV Score of Mutual Fund House} = \frac{\text{Proportion of assets in that bucket of the Mutual Fund House}}{\text{Rank of Mutual Fund House in the bucket}}$$

NAV Growth:

For the purpose of evaluation of the growth of mutual funds in the stipulated time frame (2013 to 2018), the returns of the respective mutual funds were calculated. Firstly, the daily returns are obtained from the NAV prices which are then annualised to yield the annual return (or growth) of the mutual fund schemes.

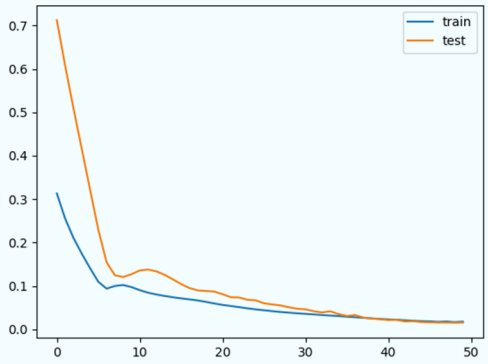
$$\text{Daily Returns}(r_{t+1}) = \ln(NAV_{t+1}/NAV_t)$$

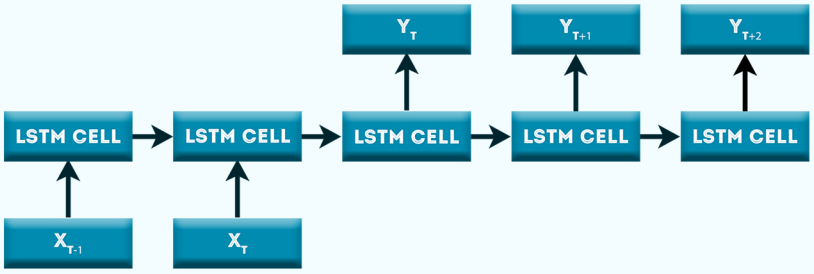
$$\text{Annual Return Rate} = (\Pi_n(1+r_i))^{1/n}-1$$

The returns are also compared to the crawled data of return of Mutual Funds for verification.

The growth of Mutual Funds schemes in the future have been determined by forecasting the NAV values of different mutual funds using its previous values and macroeconomic parameters. This is a **Multivariate Time Series Forecasting** problem, and **Long Short -Term Memory** (LSTM) has been used for implementing this model. LSTM possesses the capability to learn complex sequential (temporally varying) relationships between the input and output variables. As a result, LSTM gives a high learning capacity and better accuracy over the alternatives namely, RNNs, hidden Markov models and other sequence learning methods.

NAV values are forecasted using their contemporary values, along with GDP, CPI, USD/INR exchange rate, FII/DII, and Nifty index. All these values are transformed to weekly levels, either by averaging over weeks (USD/INR and Nifty index), by extrapolating between quarters (GDP) or duplicating the monthly values (CPI, FII/DII). Owing to the nature of time series data, predictions are made by factoring in the NAV values of the past two weeks. (The duration, two weeks was determined after an iterative process to find the optimum value).





Accordingly, NAV values of upcoming four weeks (Jan 2018) have been predicted using the present NAV values and Macro Economic parameters. Using the forecasted values, the projected growth values of individual mutual funds are calculated. The aforementioned values are subsequently weighed according to:

$$\text{AUM Proportion} = \frac{\text{AUM of Mutual Fund}}{\text{AUM of Mutual Fund House}}$$

to yield average NAV growth of Mutual Fund Houses. The values thus obtained are furthermore normalized and then used as NAV growth score metric in the final Performance score.

PERCEPTION

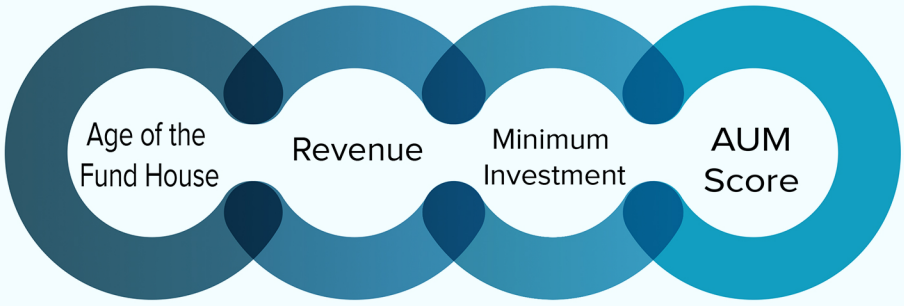
Perception is a measure of the customer's initial perspective of the Mutual Fund House. This metric tries to incorporate the basic parameters that a layman would consider before deciding the Fund House to invest in. It is calculated using a linear weighted model based on Analytic Hierarchy Process (AHP).

The parameters which influence this score are:

>Age of the Fund House (weight = 0.193)
>Annual Revenue generated by parent company (weight = 0.391)
>Minimum Investment (weight = 0.274)
>AUM size (weight = 0.142)

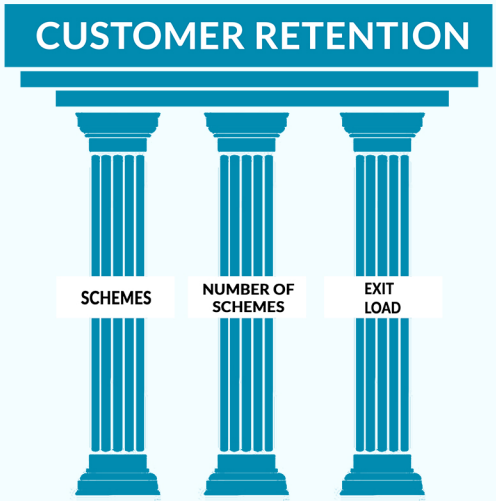
$$\text{Score} = \sum(\text{ Weights x Parameters })$$

Since each Mutual Fund house has a variety of schemes, the minimum investment parameter for each Fund House is taken as the value below which one can invest in 95 percent of the schemes. Concretely, this translates to the usage of 95 percentile value of the minimum investment prices of the fund house for all the schemes.



CUSTOMER RETENTION

Customer retention refers to the actions taken by the companies and organizations to reduce the number of customer defections. It is an estimate of customer contentment after investing in a Mutual Fund House. As per a Harvard Business School report, increasing customer retention rates by 5% can increase profits by 25%-95%. In light of this fact, the importance of customer retention in driving sales and revenue is substantiated.



Customer Retention has been defined on the basis of the following 3 parameters:

New Schemes (X) and **Closed Schemes** (Y): It signifies the cognizance of the mutual fund and its response to changing market demand. New schemes are the number of schemes introduced from 2013 to 2017. Closed Schemes are the number of schemes that have been discontinued from 2013 to 2017.

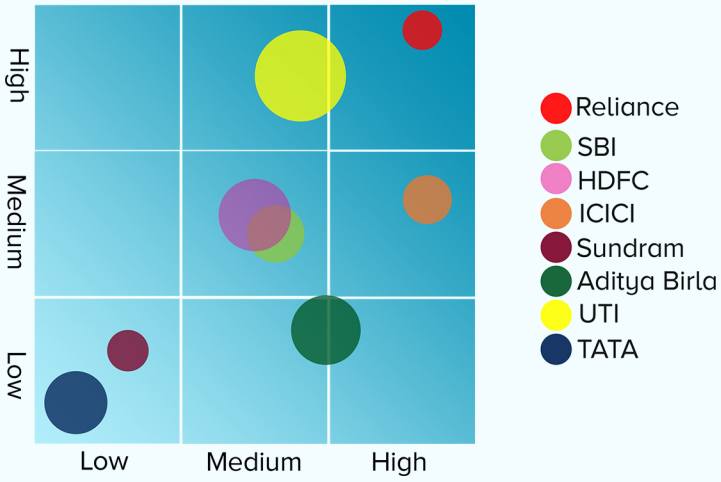
Number of schemes (N): A measure of the options that the potential customers can avail. It signifies the dynamic nature of the mutual fund.

Exit Load (E.L.) : Mutual funds typically levy exit load to discourage investors from hopping out of schemes before maturity. Customers tend to prefer Fund Houses with lower exit load.

$$\text{Schemes Collation} = \frac{\text{New schemes introduced}}{\text{Closed schemes}}$$

$$\text{Exit Load} = 1 - \left(\frac{\text{Repurchase Price}}{\text{Sale Price}}\right)$$

$$\text{Score} = \frac{(N*X)}{(E.L.*Y)}$$



MANAGEMENT CONSISTENCY

Management Consistency is a measure quantifying the adroitness by which the fund house manages it's mutual fund schemes. The term itself might appear to be abstract at the first, but it has been modelled in a way that adds to the ingenuity of the overall result. We take into consideration two factors while defining this score namely:

> **Concentration** of a mutual fund in various securities, a higher value in this criteria hampers diversification.

> **Responsiveness** is the ability to counter abrupt changes in macroeconomic factors.

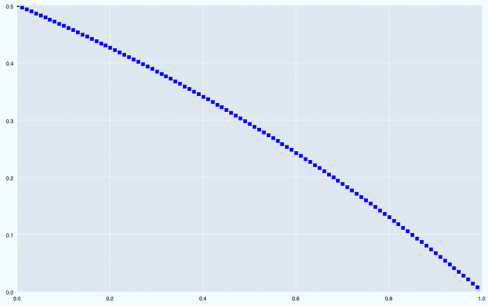
This score is initially calculated for the individual mutual funds.

Consistency Score = Responsiveness + Concentration

Concentration for a mutual fund is modelled using the top five Security Investment % available in the crawled data.

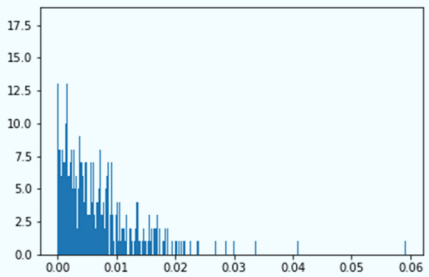
Concentration score of Mutual Fund = $(1-2^{(\text{Top 5 Holding\%-1})})$

.

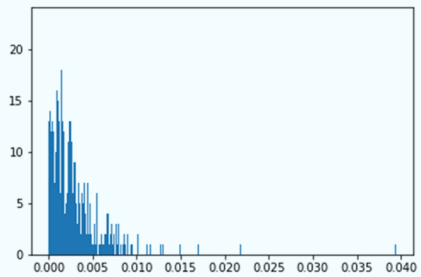


$f(x) = 1 - 2^{1-x}$

Responsiveness is a measure of the tolerance level exhibited, across all the anomalies which occured in different macroeconomics factors. Anomalies for a macroeconomic factor are those instances (dates) for which the percent change in the corresponding factor exceeds a certain threshold. Each Macroeconomic factor has a different threshold which was decided by analysing the knee point from histogram of various Macroeconomic factors.



% ∧ Nifty



% ∧ USD/INR

Subsequently, percent change in NAV values is calculated, across all the anomalies (for all macroeconomic factors) for the anomaly date and two days after it. Consequently, the responsive score of individual mutual fund is calculated.

$$\text{\% Change in Macro Economic Factor} = k1 * \%\triangle \text{ in GVA} + k2 * \% \text{ Inflation} + k3 * \% \triangle \text{in Exchange Rate} + k4 * \%\triangle \text{in Daily Index of NIfty}$$

$$\text{Responsiveness Score (for one Mutual Fund)} = \text{Average over all anomalies} \left[\text{sigmoid} \left(\frac{\% \triangle \text{NAV}}{\% \triangle \text{NAV macro economic factors}} \right) \right]$$

$$\text{Responsiveness (Fund House)} = \text{sigmoid} \left(\frac{\% \triangle NAV_T}{\% \triangle MEF} \right)$$

$$= \text{sigmoid} \left(\frac{\% \triangle AUM_T}{\% \triangle MEF} \right)$$

$$= \text{sigmoid} \left(\frac{\frac{AUM_{t+1} - AUM_t}{\% \, AUM_t}}{\% \triangle MEF} \right)$$

$$= \text{sigmoid} \left(\frac{\sum (NAV_{t+1} - NAV_t) \text{ No. of shares}}{\% \triangle MEF \times AUM_t} \right)$$

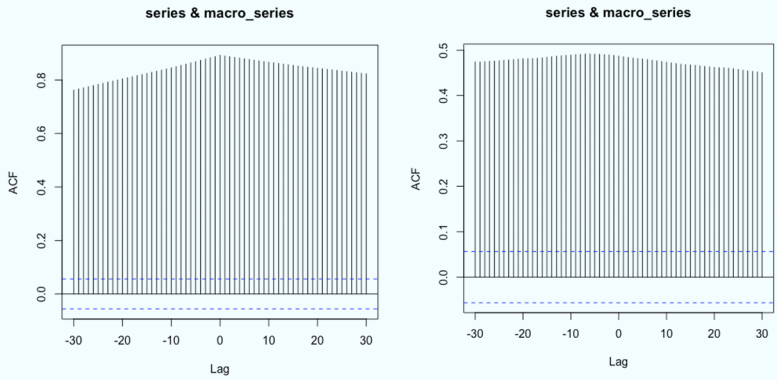
$$= \text{sigmoid} \left(\frac{\sum \left(\frac{NAV_{t+1} - NAV_t}{} \right) NAV_t \times \text{No. of shares}}{\% \triangle MEF \times AUM_t} \right)$$

$$= \text{sigmoid} \left(\frac{\sum \% \triangle (NAV) \, AUM_i}{\% \triangle MEF} \right)$$

$$\Rightarrow \text{Responsiveness(MFH)} = \text{sigmoid} \left(\frac{\sum \text{sigmoid}^{-1}(r_i) \times AUM_i}{AUM_{TOTAL}} \right)$$

Weights of Macro Economic Parameters:

The coefficients k1(0.03121319), k2(-0.03699247), k3(0.09246126), k4(1) are obtained by applying Vector Autoregression Model on the weekly NAV prices (y) and the different Macroeconomic parameters (x's). This address the second part of the problem statement that is, effect of macroeconomic parameters on the NAV values of the mutual fund. The Vector Autoregression demands lag of y with respect to x, as input. To find the lag between NAV values and macroeconomic factors the cross-correlation function was used. However, the analysis reveals that for a majority of mutual fund schemes the value of lag is between zero and one. Hence, one week is taken as the value of lag for the Vector Autoregression.



After obtaining the consistency score for individual mutual funds, these scores are translated to Mutual Fund Houses, using the formula:

$$\text{Responsiveness(MFH)} = \text{sigmoid}\left(\frac{\sum \text{sigmoid}^{-1}(r_i) \times \text{AUM}_i}{\text{AUM}_{\text{TOTAL}}}\right)$$

VERDICT

Evidenced by the sheer number of parameters considered, this analysis aimed to exhaustively compute the performance score. The analysis reveals that Aditya Birla and State Bank of India have consistently performed well in almost all the factors and hence have been ranked at the top. On the other hand, ICICI and HDFC have performed poorly in most of the parameters and hence, are at the bottom. ICICI, has a silver-lining in the form of its AUM growth. Similarly, HDFC performs well in perception and Management Consistency in an otherwise bleak performance.

An overarching insight gained from the analysis is the discrepancy that exists between the perceived and the actual performance of highly reputed entities like ICICI, HDFC and Reliance. These well established entities are unable to do justice to their huge pool of resources unlike Sundaram which does fairly well.

01 Aditya Birla

Favorable : Management Consistency

Unfavorable : Customer Retention

02 SBI

Favorable : NAV Growth

Unfavorable : AUM Growth

03 TATA

Favorable : Management Consistency

Unfavorable : NAV & AUM Score

04 Sundram

Favorable : Management Consistency

Unfavorable : AUM Growth

05 ICICI

Favorable : Management Consistency

Unfavorable : NAV & AUM Score

06 UTI

Favorable : AUM Score & NAV Growth

Unfavorable : AUM Growth & NAV

07 Reliance

Favorable : Customer Retention

Unfavorable : NAV & AUM Growth

08 HDFC

Favorable : Customer Retention

Unfavorable : NAV & AUM Growth

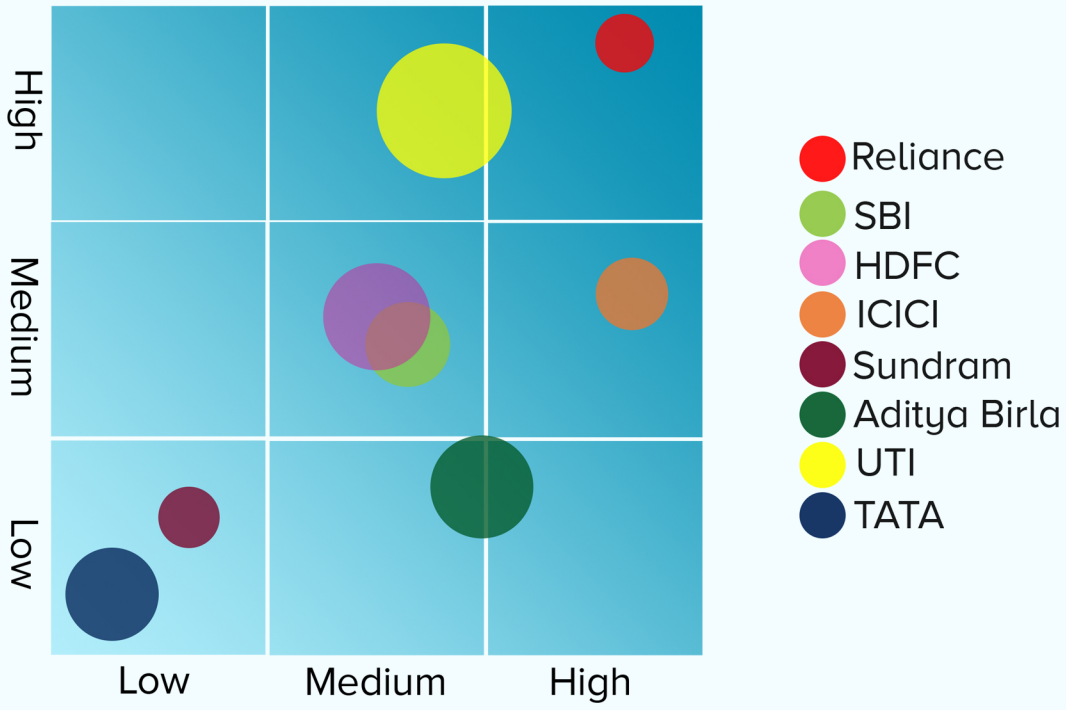
ANNEXURE

CUSTOMER RETENTION SCORE

	A	B	C	D	E	F	G	H	I	J
1	RelianceTi	276	141	367	50	0.004873	0.095512	10835.76	1	
2	SBITimeSe	180	55	220	15	0.006193	0.121384	6645.589	0.613302	
3	HDFCTime	178	54	218	14	0.006503	0.12746	6597.037	0.608821	
4	ICICITimeS	382	214	428	168	0.005256	0.103018	5292.165	0.488398	
5	Sundaram	170	72	178	64	0.002128	0.041709	4801.107	0.44308	
6	AdityaBirla	263	99	287	75	0.00649	0.127205	2978.184	0.274848	
7	UTITimeSe	186	118	270	34	0.05102	1	937.0588	0.086478	
8	TataTimeS	195	77	146	126	0.00617	0.120933	737.7825	0.068088	
9	FH_Name	OLD	NEW	OPEN	Closed	Average E	normalise	Score	Normalised score	
10					Average:	0.01108				
11					STd deviat	0.01516				

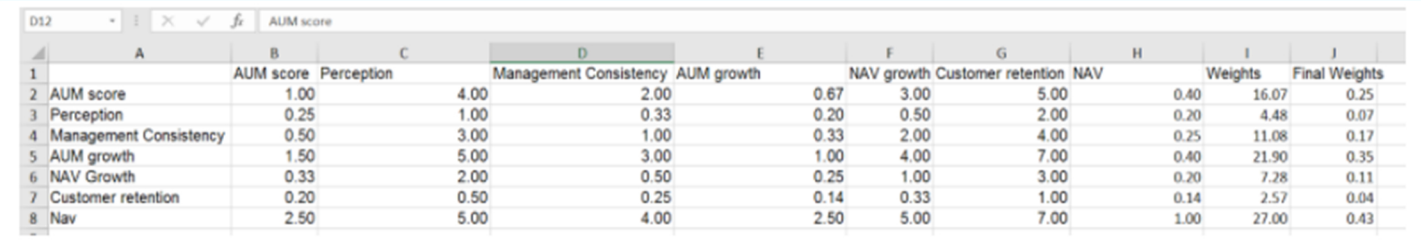
McKinsey matrix

The McKinsey matrix (MKM) is a model to perform a business portfolio analysis on the Strategic Business Units of a corporation. A business portfolio is the collection of Strategic Business Units that make up a corporation.
This matrix considers three factors for ranking.
One factor is plotted on x-axis ,second on the y-axis and third signifies the radius of a circle which is plotted on the matrix depending on other two factors.



ANALYTIC HIERARCHY PROCESS

The Analytic Hierarchy Process (AHP), is an effective tool for dealing with complex decision making. By reducing complex decisions to a series of pairwise comparisons, and then synthesizing the results, the AHP helps to capture both subjective and objective aspects of a decision. In addition, the AHP incorporates a useful technique for checking the consistency of the decision maker's evaluations, thus reducing the bias in the decision making process.
The AHP generates a weight for each evaluation criterion according to the decision maker's pairwise comparisons of the criteria. The higher the weight, the more important the corresponding criterion. Next, for a fixed criterion, the AHP assigns a score to each option according to the decision maker's pairwise comparisons of the options based on that criterion. The higher the score, the better the performance of the option with respect to the considered criterion. Finally, the AHP combines the criteria weights and the options scores, thus determining a global score for each option, and a consequent ranking. The global score for a given option is a weighted sum of the scores it obtained with respect to all the criteria.



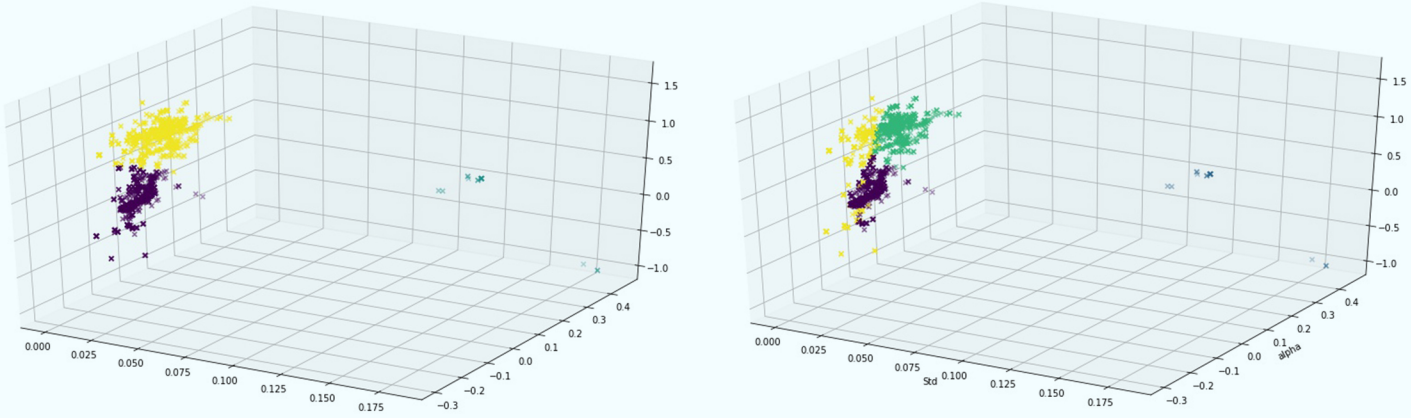
D12 | AUM score

	A	B	C	D	E	F	G	H	I	J
1		AUM score	Perception	Management Consistency	AUM growth	NAV growth	Customer retention	NAV	Weights	Final Weights
2	AUM score	1.00	4.00	2.00	0.67	3.00	5.00	0.40	16.07	0.25
3	Perception	0.25	1.00	0.33	0.20	0.50	2.00	0.20	4.48	0.07
4	Management Consistency	0.50	3.00	1.00	0.33	2.00	4.00	0.25	11.08	0.17
5	AUM growth	1.50	5.00	3.00	1.00	4.00	7.00	0.40	21.90	0.35
6	NAV Growth	0.33	2.00	0.50	0.25	1.00	3.00	0.20	7.28	0.11
7	Customer retention	0.20	0.50	0.25	0.14	0.33	1.00	0.14	2.57	0.04
8	Nav	2.50	5.00	4.00	2.50	5.00	7.00	1.00	27.00	0.43

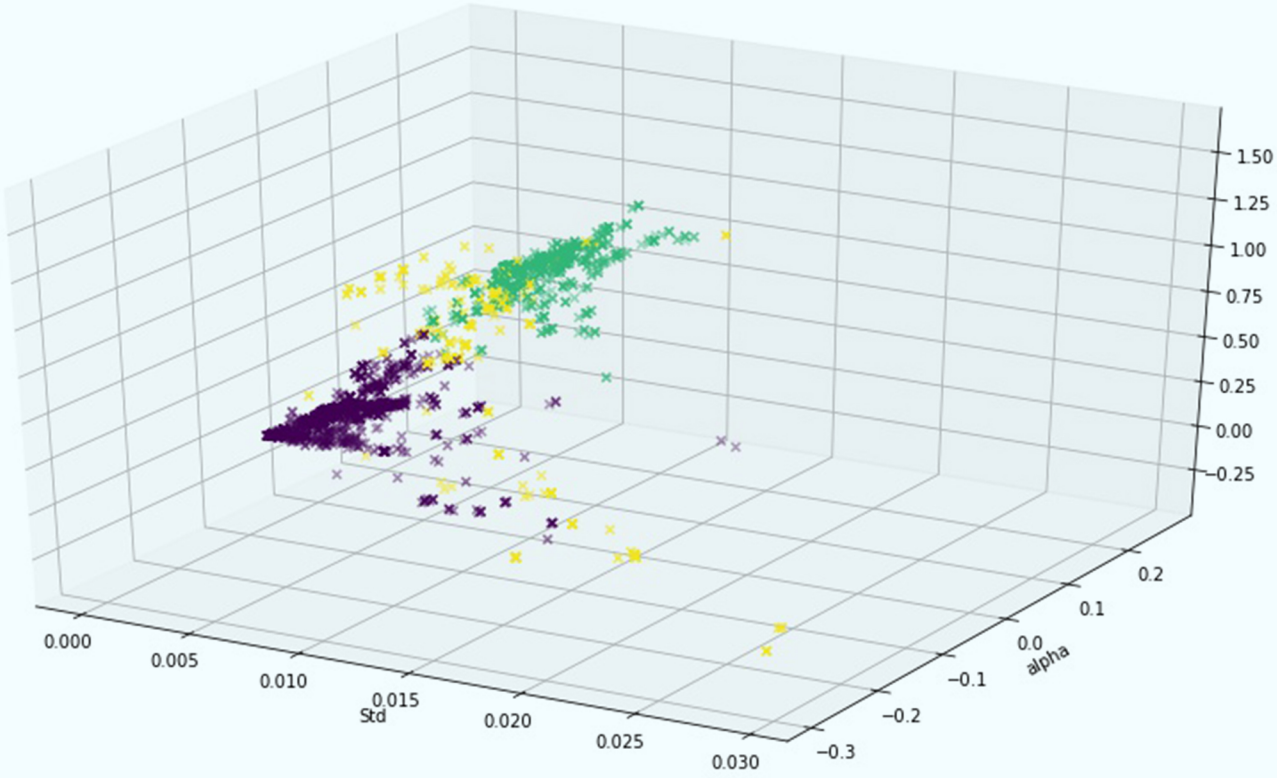
CLUSTERING

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups.
K-means clustering is a type of unsupervised learning, which is used when you have unlabeled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided. Data points are clustered based on feature similarity. The results of the K-means clustering algorithm are:
The centroids of the K clusters, which can be used to label new data and Labels for the training data (each data point is assigned to a single cluster).

Initially, k=3 and k=4 were used.
Although, it was seen that, there were a few outliers in the data.



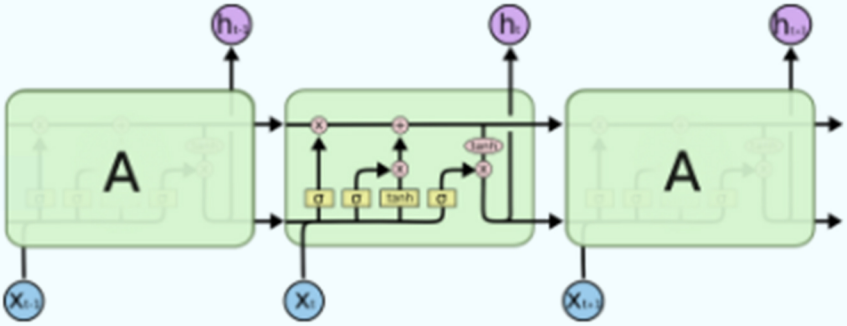
The outliers, were then removed and final clustering was done with k=3.

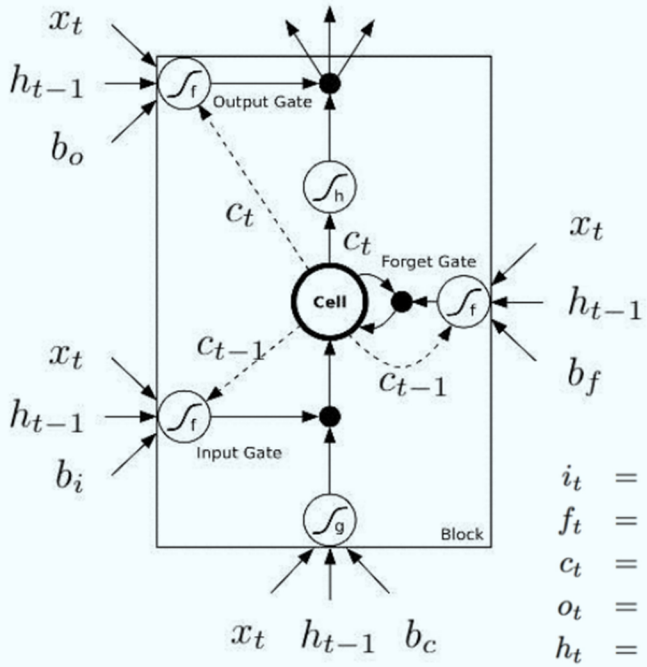


LONG SHORT TERM MEMORY

Long Short Term Memory networks – usually just called “LSTMs” – is a special kind of neural network, capable of learning long-term dependencies. LSTM are able to learn temporal relationship in data, hence able to forecast long sequences.

A benefit of LSTMs in addition to learning long sequences is that they can learn to make a one-shot multi-step forecast which may be useful for time series forecasting.





$$
\begin{aligned}
i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \\
f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \\
c_t &= f_t c_{t-1} + i_t tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \\
o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \\
h_t &= o_t tanh(c_t)
\end{aligned}
$$

We contrived a multi-step forecast.
For a given two weeks in the final 261 weeks of the dataset,
we will be required to make a next4 weeks forecast of NAV values.
We trained the LSTM on all 261 weeks of dataset.
Specifically, from 1st week of January,2013 to last week of December, 2017.
For forecasting, we predicted NAV values of 4 weeks of January, 2018 given last two weeks of December,2017.