

Systematic Investing



Agenda

- I. Business Problem and Objectives
- II. Executive Summary
- III. Data Overview and Key Business Assumptions (if any)
- IV. Data Preparation and Pre-processing
 - Sanity checks, treatment and transformations for analytical dataset preparation
- V. Exploratory Data Analysis
 - Key Business Findings and Insights
- VI. Model Development and Validations
 - Model comparisons on key scoring metrics and model finalization
- VII. Dashboarding (required only when it is in scope of analysis)
- VIII. Businesss Recommendations and Potential Business Impact
- IX. Next Steps
- X. Appendix

Business Problem

Anything, any event in the outside world, can influence the price of the stock.

- ▶ Political events
- ▶ Economic news
- ▶ Rumors
- ▶ Anxiety,
- ▶ Psychological factors.
- ▶ Etc..

So, unlike other data science problems that generally are predicted, the predictions of stock prices are rather complicated

Objectives

Create an investment approach that emphasizes data-driven insights, scientific testing, and disciplined portfolio construction technique to seek varied portfolio outcomes.

Systematic Investing

Executive Summary

- ▶ Stock price prediction model is created by applying forecasting and ML techniques on the historical prices.
- ▶ An investment portfolio is then created by applying the above model on a universe of stocks and rank the stocks based on the expected returns. The top 10 performing stocks are actively managed.
- ▶ The portfolio is adjusted on a regular basis and the performing stocks are retained in the portfolio and the non-performers are replaced with the new stocks.
- ▶ The above model is tested by creating a backtest engine for testing the model on historical data.
- ▶ The performance of the portfolio is then compared with the benchmark index to check the outperformance

Data Overview

- ▶ The daily closing prices for **50 large cap stock** Universe from **2015 to 2021** is used as raw data for predicting the prices and is sourced from financial data vendors.
- ▶ As shown, the raw data provides Date, Open price, High Price, Low Price, Closing Price, Adjusted Closing price and Volume data of the stocks
- ▶ Covid-19 period is chosen to understand the model preparedness against the black swan events.
- ▶ The predictions will be used to rank the stocks and the top 10 stocks will be used for investment and will be tracked for the period of back test.

date	open	high	low	close	volume	symbol
2015-01-01	1283.5	1283.5	1270.5	1272.78	366830	TCS
2015-01-02	1275.5	1295.47	1275.3	1289.72	925740	TCS
2015-01-05	1290.5	1299.95	1262.33	1270.13	1754242	TCS
2015-01-06	1264.55	1264.55	1220	1223.3	2423784	TCS
2015-01-07	1235	1239.58	1203.72	1208.85	2636332	TCS
2015-01-08	1221.2	1224.5	1210.28	1221.9	1565408	TCS
2015-01-09	1227.5	1259.95	1225	1256.15	3197642	TCS
2015-01-12	1258.5	1264	1240.13	1254.85	1596006	TCS
2015-01-13	1260	1265.2	1240.05	1248.95	1468432	TCS
2015-01-14	1258	1265.9	1250.25	1260.97	1787096	TCS
2015-01-15	1266.5	1290.88	1261.63	1269.55	5009920	TCS
2015-01-16	1271.85	1272.5	1242.15	1266.08	5293830	TCS
2015-01-19	1268	1275.5	1251.5	1255.5	1135074	TCS

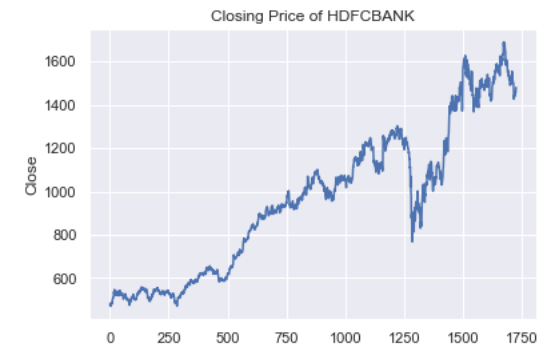
Data Preparation and Preprocessing

- ▶ Data Sanity checks is done on the data to check if the prices are split adjusted
- ▶ The Datetime and Close price is the only data that will be used as an input data and fed to the models for future price forecasting
- ▶ The Data for the complete universe is merged in a single data set on the date.
- ▶ The above dataset becomes the master data for ranking the individual stocks based on the expected returns and used by the back-testing Engine for actively managing the portfolio on the historical dates and create the equity curves.

	datetime	ADANI PORTS	ASIANPAINT	AXISBANK	BAJAJ-AUTO	BAJFINANCE
0	2019-05-31	416.40	1407.30	808.30	2925.90	3467.10
1	2019-06-30	410.20	1358.15	808.55	2827.05	3681.10
2	2019-07-31	377.55	1521.10	674.10	2516.50	3252.05
3	2019-08-31	365.70	1616.20	663.90	2788.30	3332.50
4	2019-09-30	413.95	1762.15	685.00	2941.80	4046.05
5	2019-10-31	398.10	1821.30	747.40	3222.90	3992.00
6	2019-11-30	382.05	1706.15	739.05	3176.00	4074.20
7	2019-12-31	366.00	1784.95	754.10	3185.00	4234.75
8	2020-01-31	369.40	1795.65	729.30	3180.05	4365.90
9	2020-02-29	342.20	1797.95	697.30	2890.00	4465.85
10	2020-03-31	251.30	1666.50	379.00	2022.35	2215.80
11	2020-04-30	290.10	1758.70	444.90	2623.30	2318.10
12	2020-05-31	323.80	1683.10	384.95	2710.50	1953.65
13	2020-06-30	343.90	1687.45	406.65	2826.05	2831.00
14	2020-07-31	315.20	1715.50	431.65	3004.95	3251.30
15	2020-08-31	355.75	1899.10	496.75	2967.00	3487.80
16	2020-09-30	341.75	1986.40	424.65	2881.10	3278.60
17	2020-10-31	359.85	2211.50	492.50	2886.90	3309.00
18	2020-11-30	411.55	2215.30	601.60	3173.55	4908.80
19	2020-12-31	483.75	2764.50	620.45	3444.05	5295.20
20	2021-01-31	509.70	2407.35	662.90	4005.80	4734.55

Exploratory Data Analysis

- The Closing prices of the different stocks are plotted in a line chart to check the data completeness



Model Development and Validation

- ▶ Various Time series forecasting and ML techniques were explored to forecast and predict future prices and create the forecasting engine.

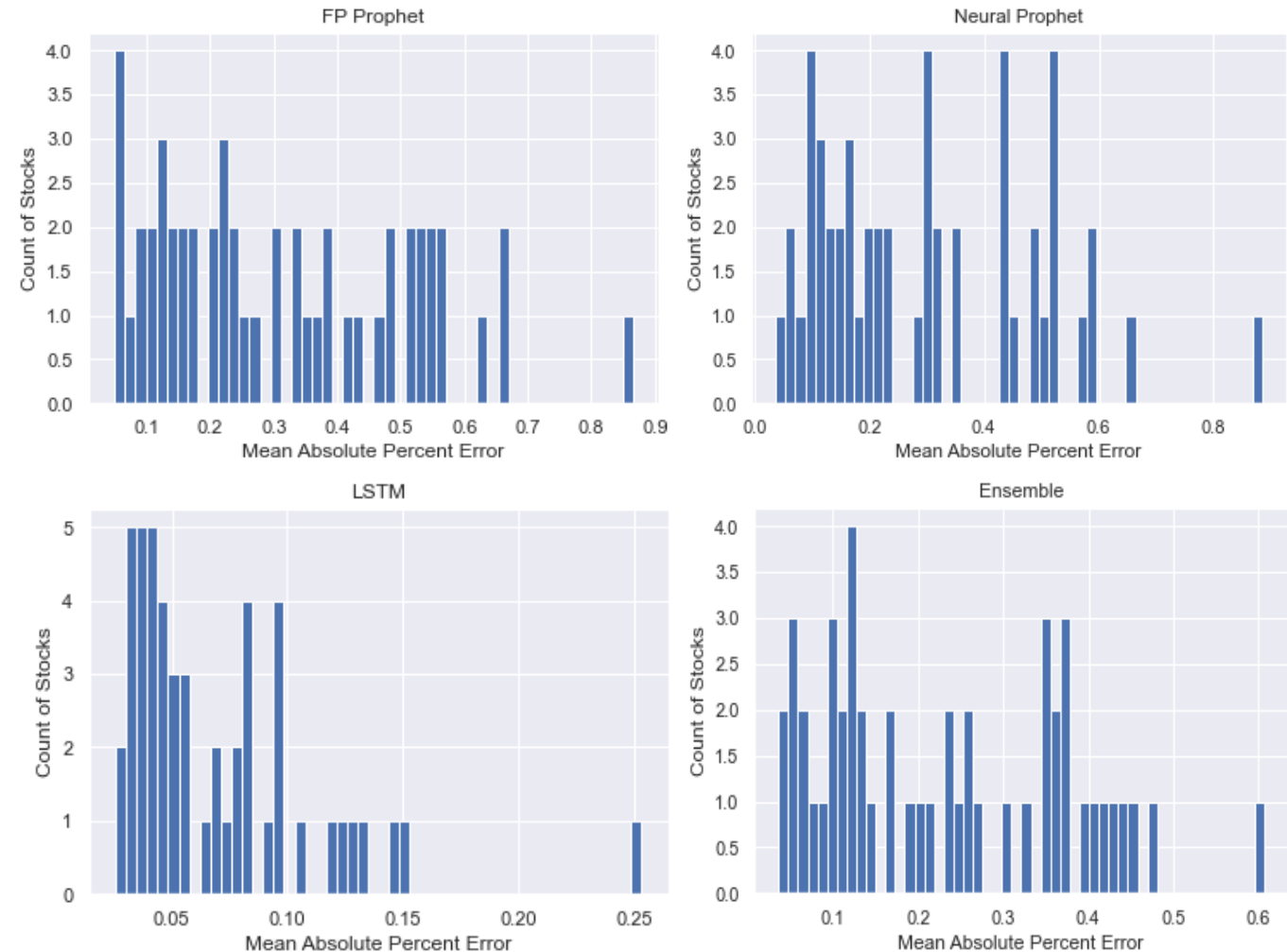
- Linear regression
- Holt winter's method
- ARIMA
- FB prophet
- Neural Prophet
- Random Forest
- LSTM

Based on our initial analysis and requirement of creating short and long term price forecasting engine we choose the below there models and created an ensemble of these models.

- FB prophet
- Neural Prophet
- LSTM

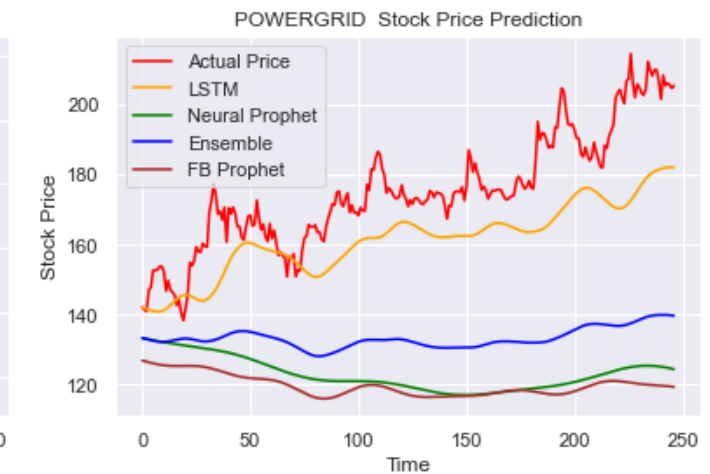
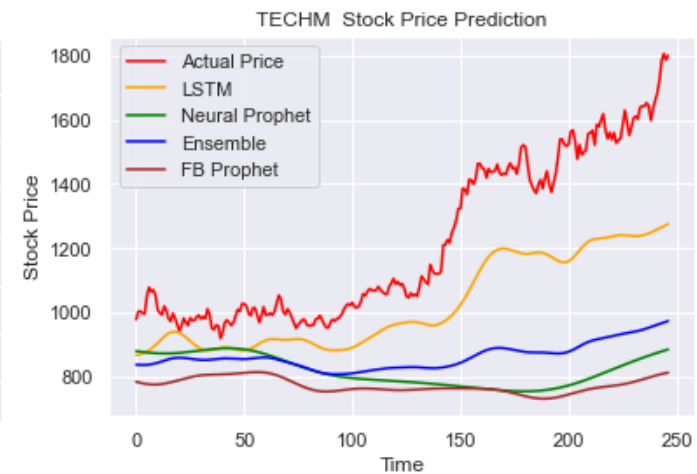
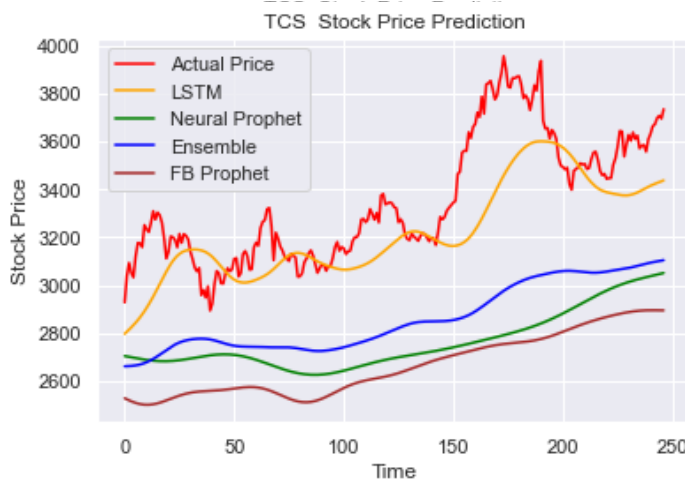
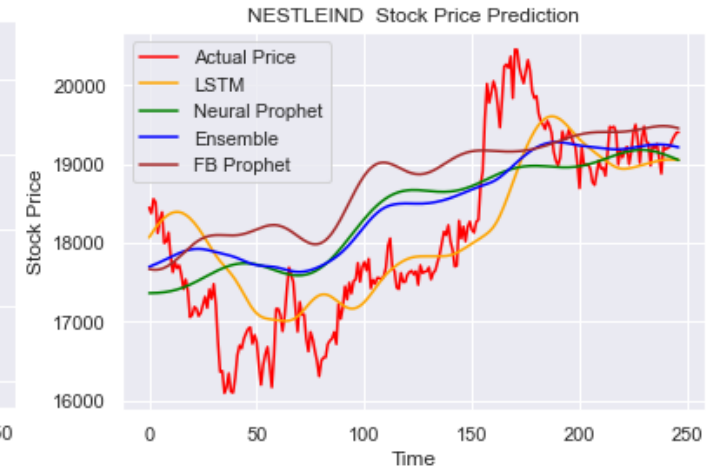
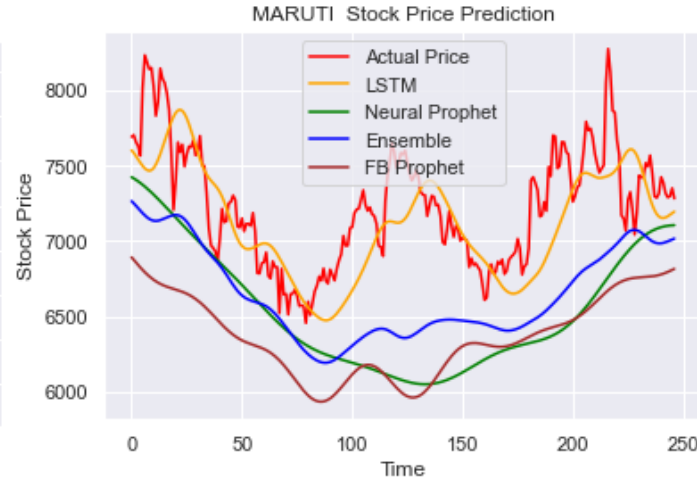
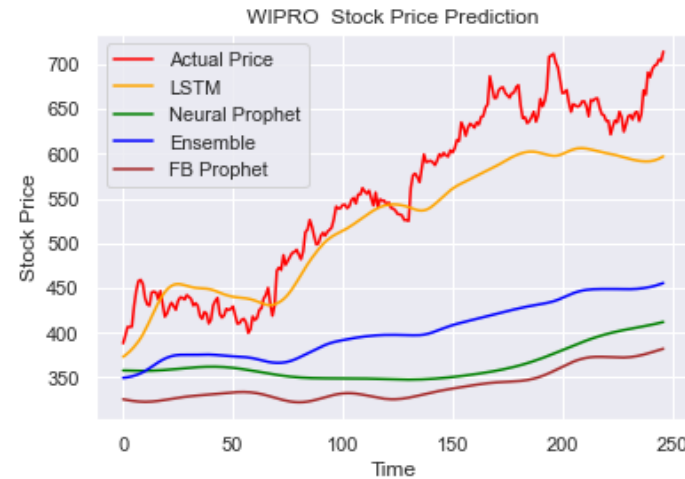
Model Development and Validation

- ▶ The Historical prices were divided in to Train and test data with Training period from Year 2015 to 2020 and Test Period for Year 2021.
- ▶ The Models were trained on randomly sampled stocks.
- ▶ MAPE: Mean Absolute Percentage error is used for checking forecast accuracy as It comes under percentage errors and are scale independent.
- ▶ The Histogram plot for the three models and its ensemble is plotted to check the accuracy for all the 50 stocks



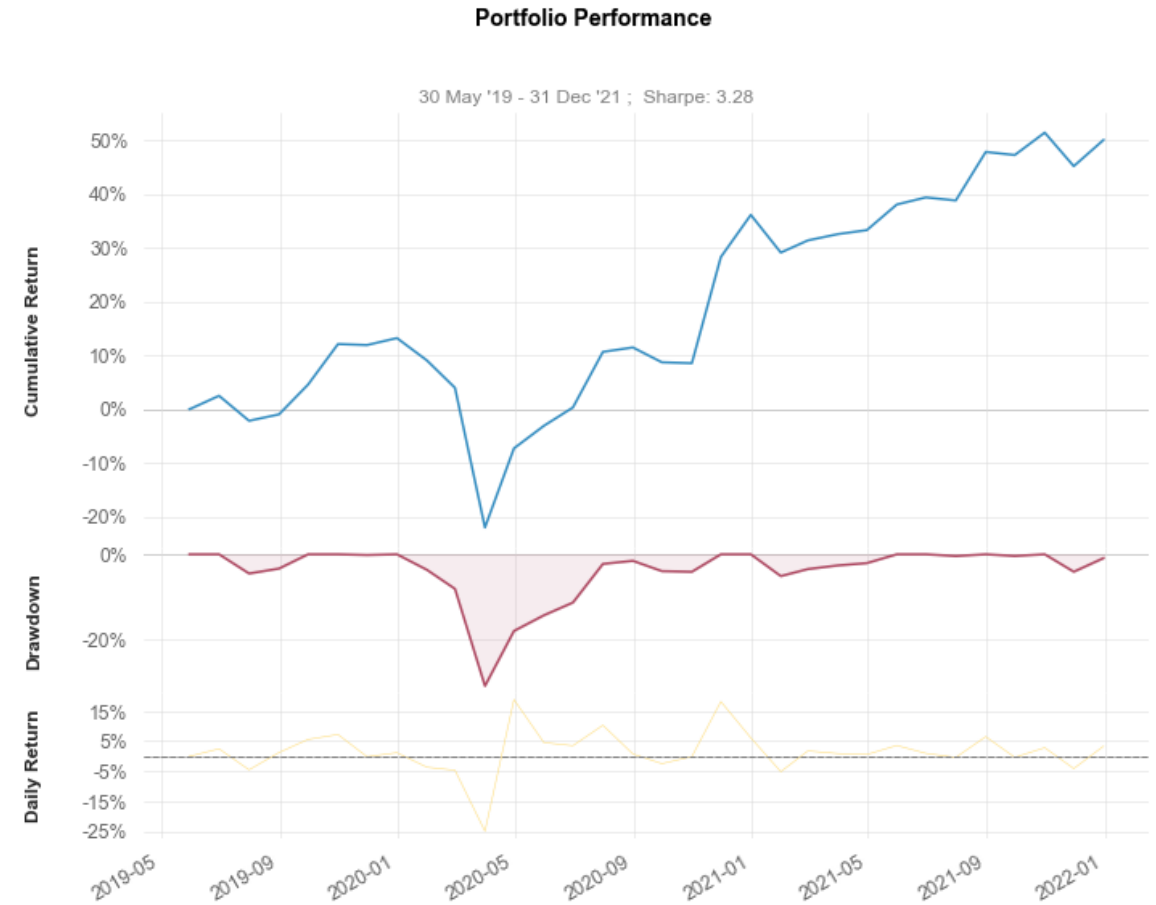
Model Development and Validation

► Below are some of the plots for model comparison



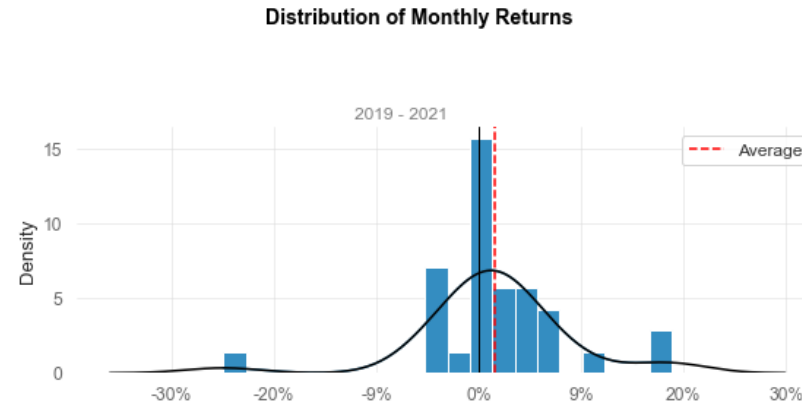
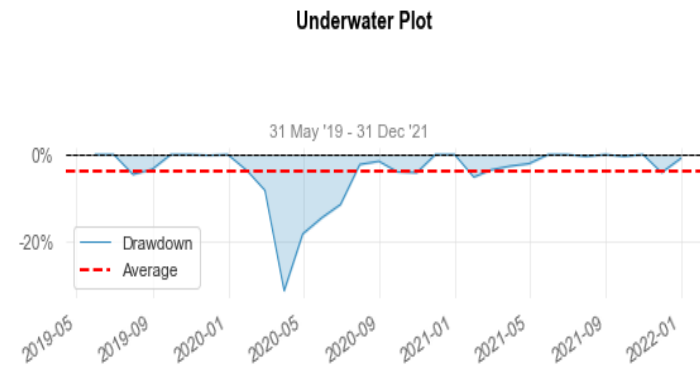
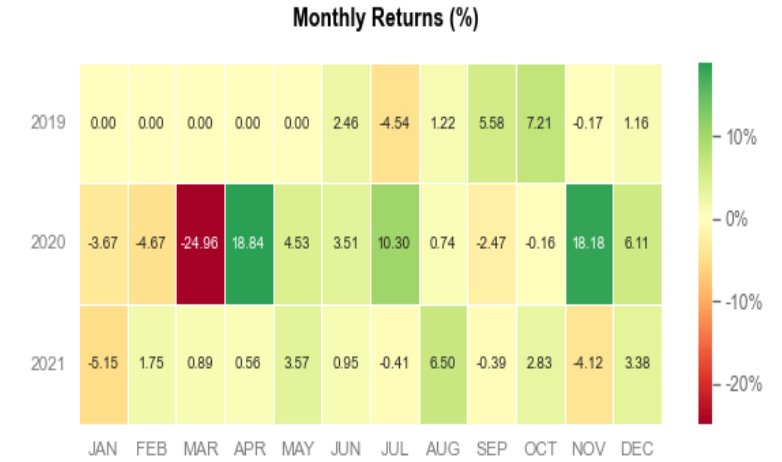
Back-Test the Models

- ▶ Once the models were validated, We tried to run the Model on Historical prices.
- ▶ Running the complete ensemble model meant training all the 4 models each day on each stocks and generating signals for that day and actively managing the same stocks for the rest of the period of backtest while also rebalancing it on a regular intervals incase there is change in the stock holdings
- ▶ Since the above process took a huge amount of time and computing power to run we had to reduce the scale of the project and create a backtest for last 3 years of data and with only 1 Model i.e. FP Prophet.



Back-Test the Models

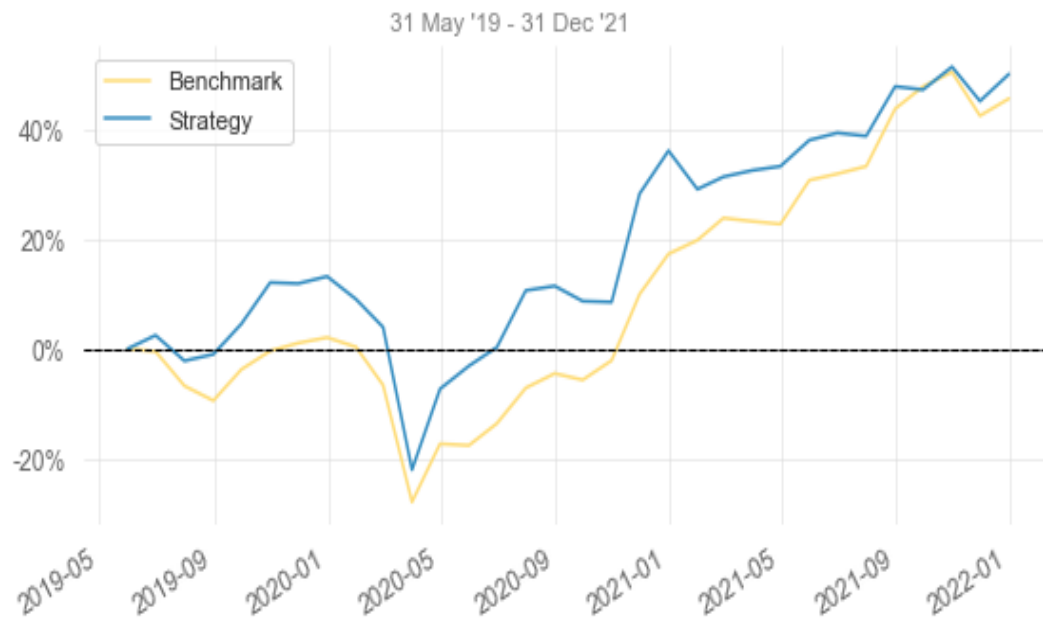
► Portfolio Performance Metrics



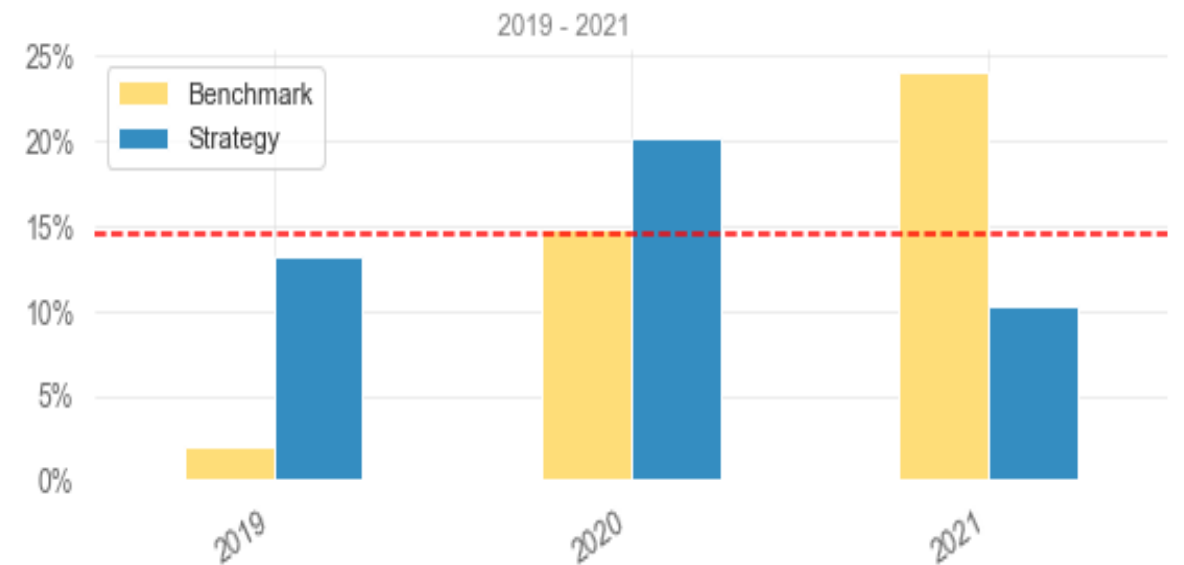
Back-Test the Models

- We compared the model with the benchmark index

Cumulative Returns vs Benchmark



EOY Returns vs Benchmark



Business recommendations

- ▶ The Systematic Investment model was able to provide at 55% of returns for the last 3 years
- ▶ The Model Had a Drawdown of 25% for a black swan event like covid which is considered a good metrics.
- ▶ Though the model was good enough to survive the covid crash it was not able to outperform the benchmark Index as clearly seen in the benchmark comparison equity curve.

Challenges

- ▶ One of the major challenge was to test the complete Model on the historical data to get a clear picture of the Model performance over different market cycles.

Thank You