Personal Journal:

02/12/2024: Team had a discussion regarding the topic selection and project work distribution was formed.

All of us sat together and started discussing the project deliverables, what is expected of the results. Since the objective is revolving around Datasets and Machine Learning Algorithms, we decided to think of the topic on which we are going to work.

05/12/2024: Topic finalization

In the last meeting itself, we were clear about the objective, and everybody came up with some topics which we can consider. The team had suggested various topics like Sentiment analysis, whether forecasting, and prediction using AI/ML techniques. We all agreed to work on Sentiment Analysis from Twitter data at the very beginning.

10/12/2024: Topic change because we wanted to work on something different.

Sentiment analysis using twitter data was already done by Illiyaz in Data Analytics project. Illiyaz had a grip on some algorithms like LogisticRegression and Classifiers, but Illiyaz wanted to work on some different algorithms which are related to time series forecasting. So, we decided to work on Stock price analysis and prediction using Regression models.

15/12/2024: Dataset collection and further analysis.

After going through many websites and datasets, some good datasets were selected by the team. I gave an idea for using one of the inbuilt libraries like y-finance which gives data for different stock prices. Which seemed to be a good idea at first. And we started to work on it, how to include y-finance and working on the coding part.

20/12/2024: Coding process

As I was working on including y-finance module to fetch the stock data, Illiyaz was thinking on how we can apply the Machine Learning Algorithms to the datasets which gives best results. Initially Linear Regression was looking a better option, as it is simple to understand and implement and it is easier to interpret the output coefficients. Even tough it is vulnerable to over-fitting but we could manage it by dimensionality reduction techniques.

22/12/2024: Taking proper datasets(csv files) instead of y-finance module

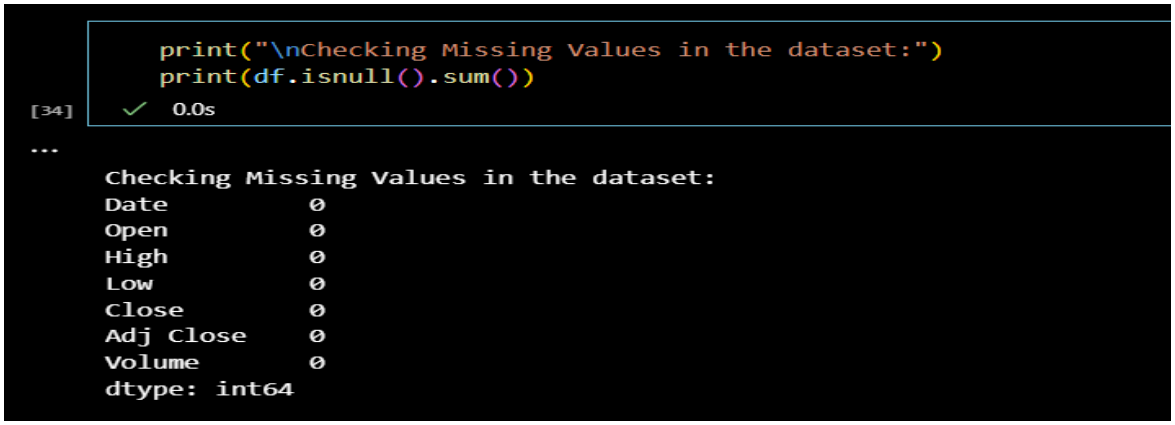
Hitesh suggested that the yfinance module has certain disadvantages and the objective of the project may get missed. Because we have to collect data and then pre-process, we need

to keep checking and comparing the trends, so we can get proper datasets which include stock values, Opening, Closing, and highest stock values for the day.

Shravani got the datasets from Kaggle which suits best for our project. Here is the link for the datasets (https://www.kaggle.com/datasets/zongaobian/netflix-stock-data-and-key-affiliated-companies). We have stock prices of many different companies (multiple datasets) but which are interrelated to each other. So, we decided to do Netflix Stock Prediction along with other stocks (Amazon, Nvidia, SONY) which are highly interrelated with Netflix stocks.

23/12/2024: All of us took one dataset each and started working on the datasets and working on the code part. I chose NVidia Stock data, imported the required dependencies, and loaded the dataset into VS-Code. I started showing its dimensionality and important features.

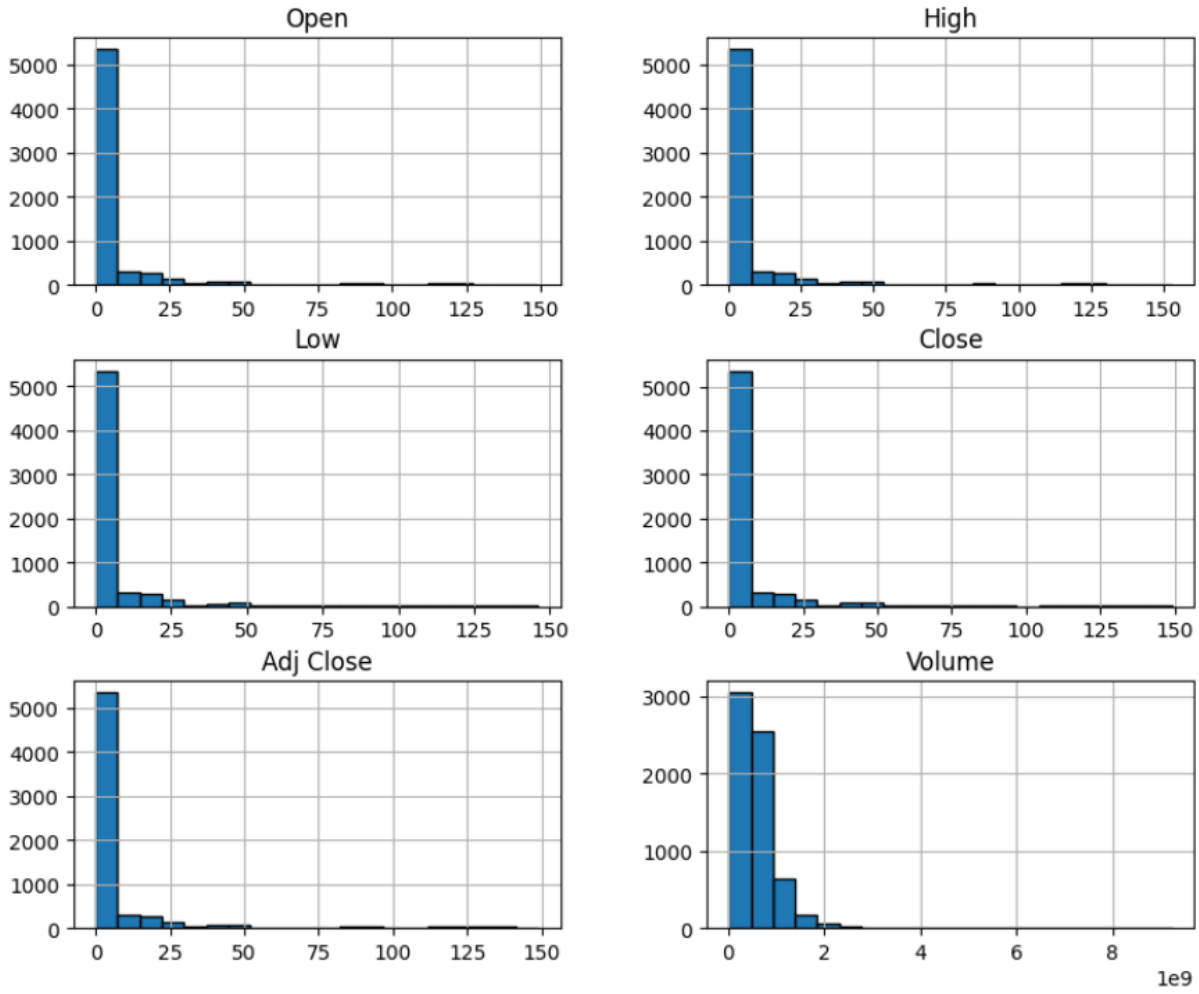
As it is important to understand the important features and data types of the datasets, we decided to visualize my dataset. But before that we needed to check If there are any missing values in the datasets, because missing values might create problems and causes errors in the code while applying ML Algorithms or while plotting graphs.

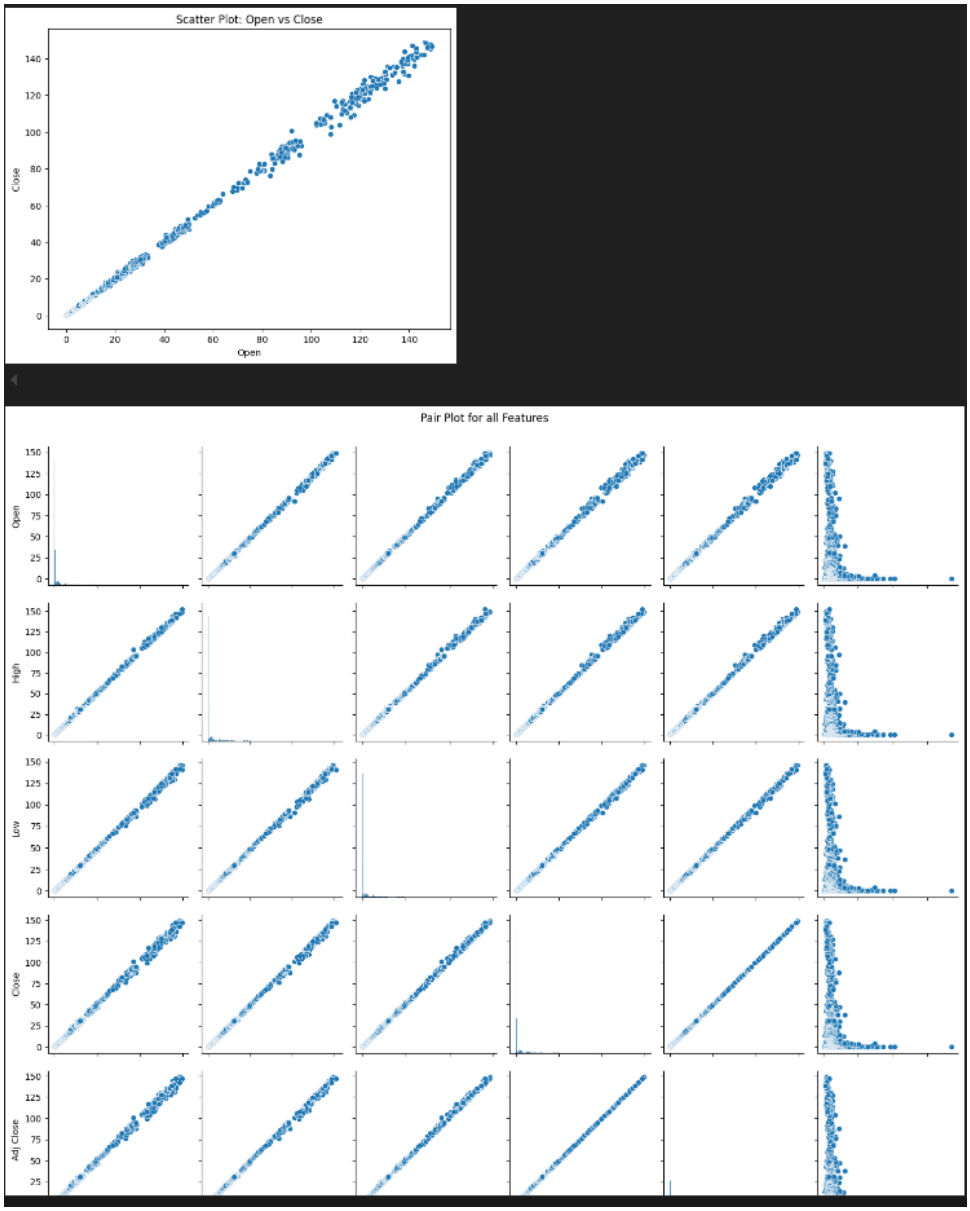


This is a clean dataset, hence there are no missing values, so there wasn't any need to remove any row or column.

24/12/2024: Data Preprocessing and EDA(Exploratory Data Analysis)

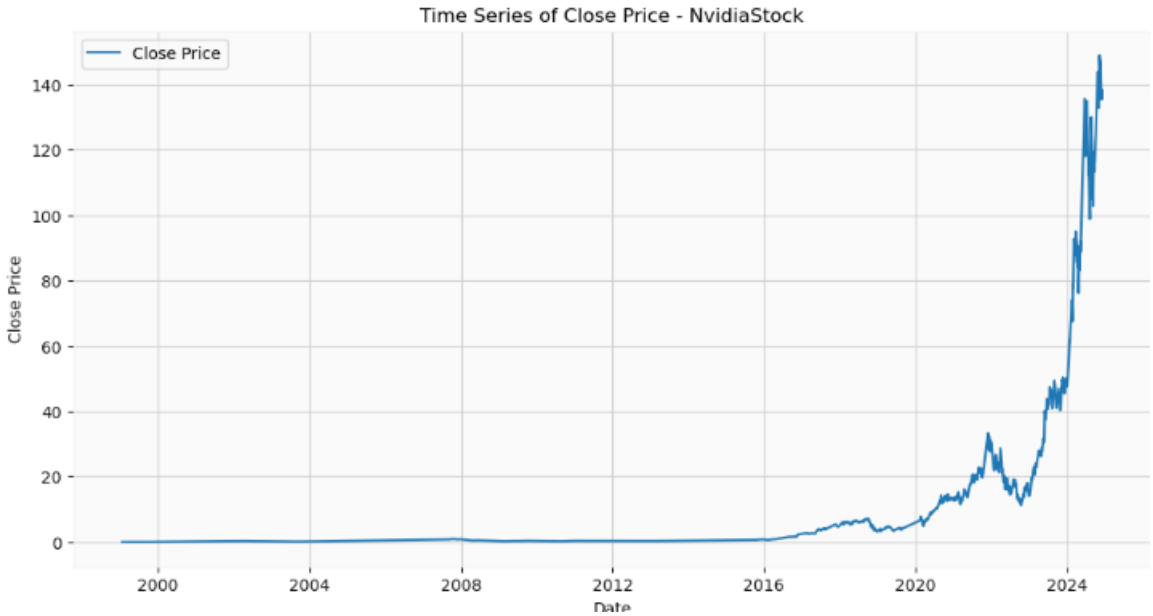
Once I was sure about handling the missing values, I thought of removing outliers; that would be great so that we have less data to work upon. But thinking over, in stock prediction, we cannot remove any entries just because it is an outlier; we need each day's data to check and predict the future stock prices, so we need data for all the days. Hence, I did not remove any outliers. So started plotting graphs for data visualization. First, I plotted a histogram for all the features, then a scatter plot for the Opening and Closing values of the stock. This gives us an idea about the top and bottom constraints of the stock.

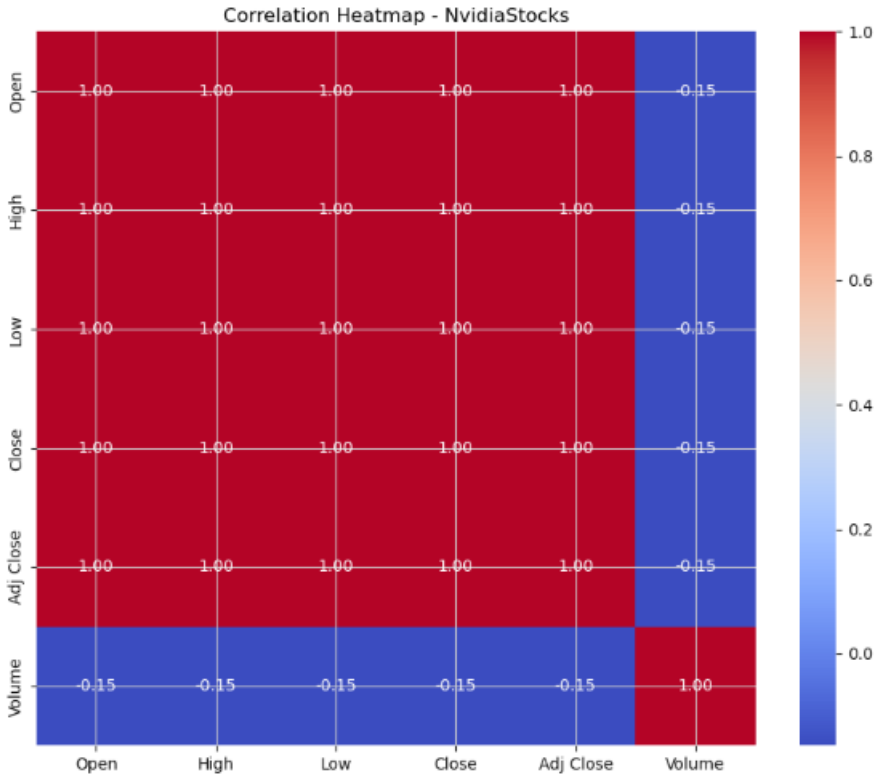

Histograms for all Features
Open
High
Low
Close
Adj Close
Volume
1e9



Then decided to plot the pair plots in order to understand the relations between all the features. But the most important relations pair was of (Open, Close) and (High, Low). This gives the overall picture of the stock price over the years.

Also Line chart for the "Closing Price" over the years were plotted. So we can compare how much a stock has been increased over the years.


Time Series of Close Price - NvidiaStock
Close Price
Close Price
140
120
100
80
60
40
20
0
2000
2004
2008
2012
2016
2020
2024
Date


Correlation Heatmap - NvidiaStocks
Open
High
Low
Close
Adj Close
Volume
1.00
-0.15
1.0
0.8
0.6
0.4
0.2
0.0

28/12/2024: ML Algorithms application

This was the main and the difficult part of our project, where our team needs to decide what ML algorithms we should use and what is the best option for stock predictions. So we decided that each member will work on different ML Algorithms, will get the output results and graphs and then the best algorithm will be picked.

So Shravani decided to work with Linear Regression model, Hitesh decided to work on RandomForestClassifier, I decided to work on LSTM model. And Illiyaz decided to work on prophet model.

So First Illiyaz ransformed the data using MinMaxScaler, so that our ML algorithm works best. Once the scaling is done. ILLIYAZ did splitting of the dataset into training and testing. As the stock prediction is highly complex topic, we need to train the model with more data. Hence ILLIYAZ decided to split the data into 80% training and 20% testing.

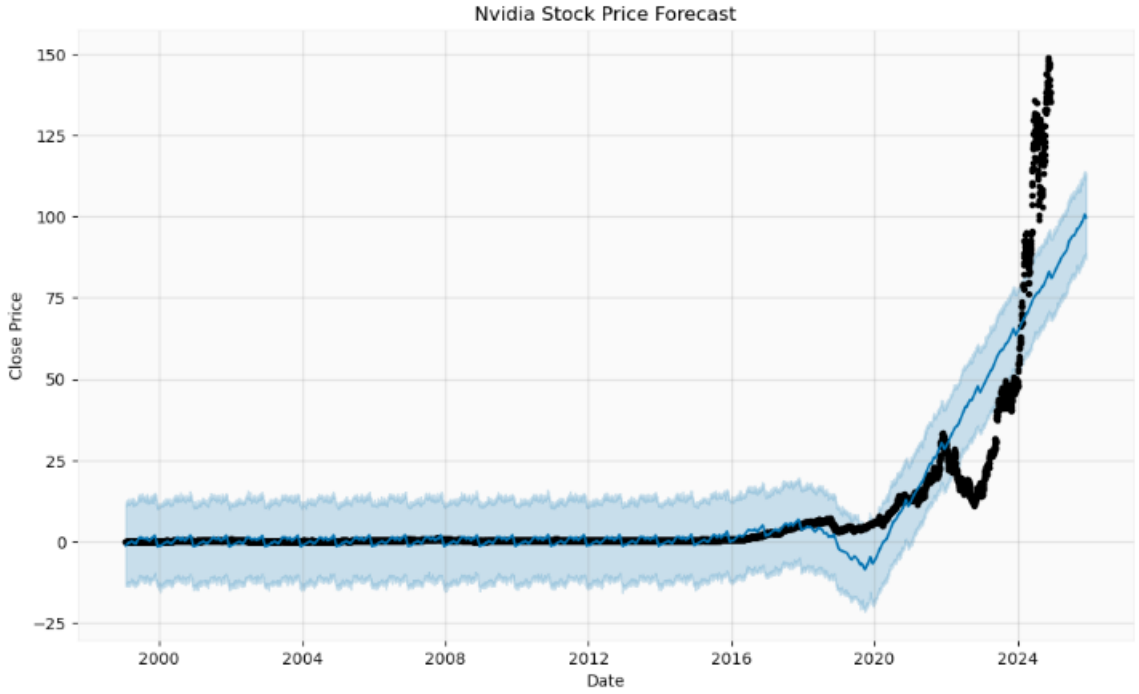
29/12/2024: Multiple error in code

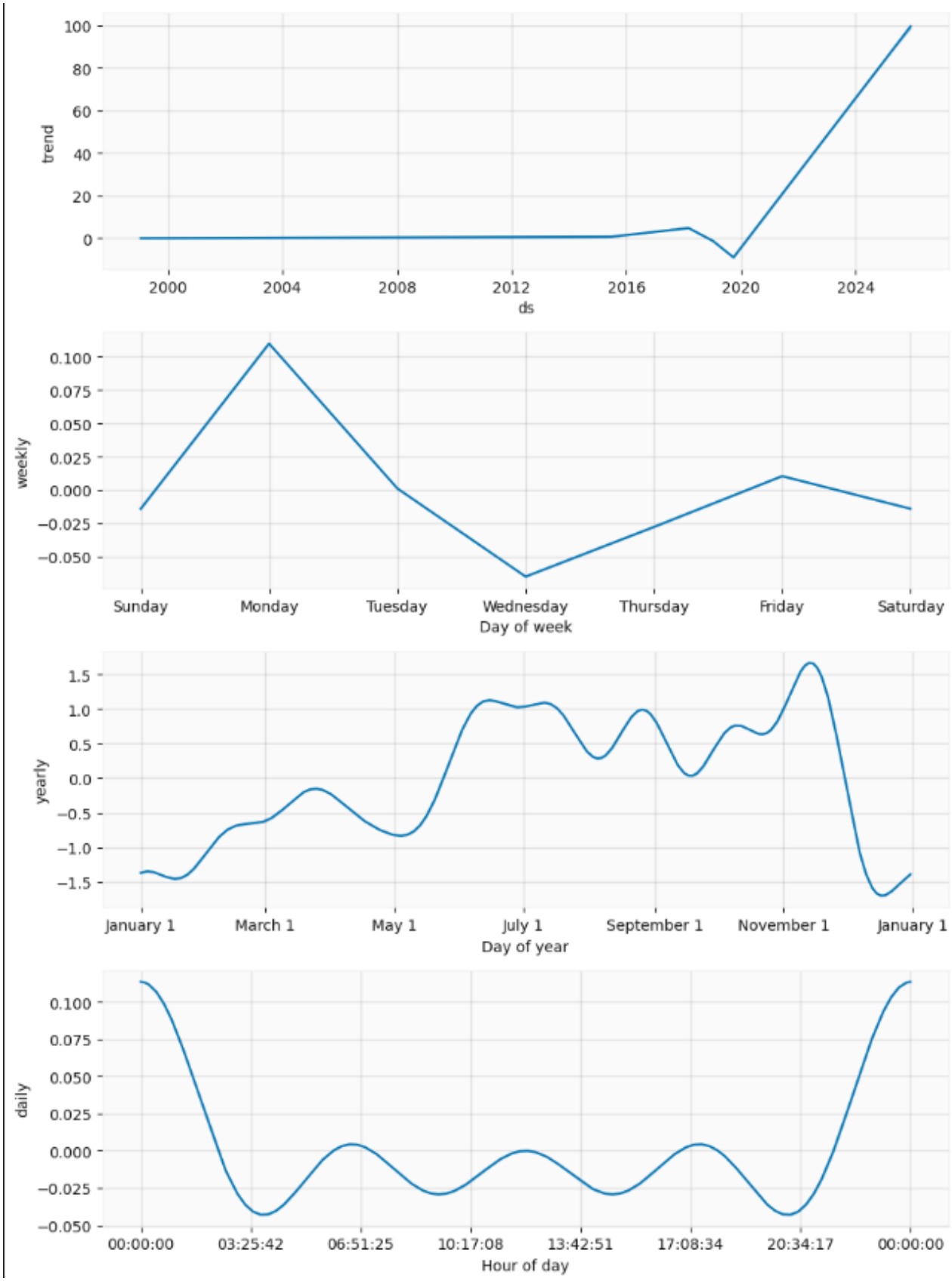
The idea of using the prophet model or how it works was not really known to Illiyaz. So, Illiyaz decided to go through the module on Prophet and understand the concept. It is additive model-based methods for time series data forecasting, which fits non-linear trends with daily, weekly and annual seasonality, as well as holiday impacts. This works the best when applied to time series that contain a number of seasons of historical data with a pronounced seasonal influence. Generally, Prophet is good at outliers, it's robust in case of missing data, and nonconstant trend. First, he learned the prophet model and took the syntax from the official python website for this ML model. However, upon implementing it in my dataset, he received several errors. Later on, Illiyaz came to know that some datetime conversions are to be performed before proceeding further. So Illiyaz referred to some videos on YouTube in which they had used the same prophet model on some dataset, and then he tried to replicate the same thing for his dataset, and it worked.

So, I updated my dataframe accordingly, trained my model on this dataframe, and then used "make_future_dataframe", which is an inbuilt function of this prophet model. It gives us the prediction of future stock values and creates a new dataframe that predicts the stock prices for the next year.

Some graphs for the predicted trends were plotted, and a new csv file will be saved that has the future trends and predictions.

```python
# Convert the 'Date' column to datetime format
df['Date'] = pd.to_datetime(df['Date'])

# Prepare data for Prophet (rename columns to 'ds' and 'y')
prophet_df = df[['Date', 'Close']].rename(columns={'Date': 'ds', 'Close': 'y'})

# Initialize the Prophet model
model = Prophet(daily_seasonality=True)

# Fit the model to the data
model.fit(prophet_df)

# Make a dataframe for future predictions
# Let's predict the next 365 days
future = model.make_future_dataframe(periods=365)

# Generate predictions
forecast = model.predict(future)
```




trend
100
80
60
40
20
0
2000
2004
2008
2012
2016
2020
2024
ds
weekly
0.100
0.075
0.050
0.025
0.000
−0.025
−0.050
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Day of week
yearly
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
January 1
March 1
May 1
July 1
September 1
November 1
January 1
Day of year
daily
0.100
0.075
0.050
0.025
0.000
−0.025
−0.050
00:00:00
03:25:42
06:51:25
10:17:08
13:42:51
17:08:34
20:34:17
00:00:00
Hour of day

02/01/2025: Working on PPT and report:

Everyone was ready with the code and outputs, but we decided to select the best models which had better accuracy and predictions. Now, all of us knew what models were applied and which was the better working model. Hence, we decided to work on the report and PPT.

Everyone shared their inputs and research papers which they used over this period. And Shravani and Hitesh acted as a bridge between all 4 of us and decided to work on the report, they collected information from all of us and started working on the report parts like Introduction abstract and literature review. Meanwhile, I and me were working on methodology and results part. Where we shared all the necessary details of the methodologies so that they can append in the report. 03/01/2025: PPT

Presentation slides which involves the important points and highlights from the project was made by me and Hitesh. All the slides were added and necessary key points were highlighted.

04/01/2025: Report and ppt finalization:

Once the report and PPT were ready, we sat together to review all the work. Checked grammar mistakes and alignments. Once it was done, we went with cleaning the report wherever it was required.

05/01/2025: All code is aligned so that every dataset selected has been operated with similar techniques.

Once the report and PPT were ready, we decided that all the datasets used should be aligned, and the same set of operations must be applied so that there will be clarity of all the operations done and algorithms applied. Also, the main reason for this is that the person who will check the code should understand and co-relate all the datasets and results.

Also, all of us uploaded the necessary files to github and added the link. Hitesh and I helped with the instructions file and README file.

06/01/2025: Video presentation

All of us did a teams call and I shared my screen, and we started explaining the code and concept of our project. We also gave a demo by running the code.