Future and Option Chain Data Analysis of Indian Stock Market for Deriving Profitable Trades

PYCK Project

Team no 24

Team Members:

A V N S Kalyan (203010024)

Basireddy Vineeth Kumar Reddy (203010025)

Resham Dnyaneshwar Khade(203010034)

Shawnam (203010035)

Program : Masters in Aerospace Propulsion

Abstract

This project is based on Data science, in which majorly python libraries such as NumPy, Pandas, Plotly, Matplotlib, Mplfinance will be used to interpret the option chain data at important strike price along with open interest, change in open interest and spot price of underlying asset on an interactive plot. The major challenge being reading the large number of data files (downloaded from the National Stock Exchange official website) in sequence and digging out the relevant data from the data pool for the particular date, month or year for specific company or domain at certain strike prices. Other challenge being arranging the required data on plots so that it's easy for anyone to interpret the trend by giving an input function to select the required data of interest. The data analysis is carried out for particular domain which will automatically plot the year wise trend of different symbols of interest.

This project is motivated by increasing interest of younger generation towards financial market and plenty of lucrative job opportunities in the area. This gives flexibility to understand the trend of strike prices for different symbols.

Introduction:

Data analysis and interpretation of data in finance sector plays very important role. There are many majors taken into consideration for analyzing the data and finding better and flexible solutions to read and interpret it for such a huge data acquired by many company's. The main purpose of doing this project is to analyze the data for different domains with respect to date, month, year for one case of symbols in the data of strike prices of all dates February 2020. Some of its important applications are to evaluate earning capacity and profitability of the company as compared to industries. This also helps us studying the strengths, weaknesses, position of company with other firms, there trend of performance. The python programing language is been used for studying and sorting out the data analysis with respect to input of our interest.

Following are the Python libraries used for our analysis:

Matplotlib library

- Matplotlib is a low level graph plotting library in python that serves as a visualization utility.
- Matplotlib was created by John D. Hunter.
- Matplotlib is open source and we can use it freely.
- Matplotlib is mostly written in python, a few segments are written in C, Objective-C and Javascript for Platform compatibility.
- This library is been used for plotting the data in box type plotting.

Installing plotly and its overviews:

The **plotly Python library** is an interactive, open-source plotting library that supports over 40 unique chart types covering a wide range of statistical, financial, geographic, scientific, and 3-dimensional use-cases.

Built on top of the Plotly JavaScript library (plotly.js), plotly enables Python users to create beautiful interactive web-based visualizations that can be displayed in Jupyter notebooks, saved to standalone HTML files, or served as part of pure Python-built web applications using Dash. The plotly Python library is sometimes referred to as "plotly.py" to differentiate it from the JavaScript library.

```
pip install plotly
```

```
Requirement already satisfied: plotly in c:\users\dell\anaconda3\lib\site-packages (5.1.0)Note: you may need to restart the ker
nel to use updated packages.
Requirement already satisfied: six in c:\users\dell\anaconda3\lib\site-packages (from plotly) (1.15.0)
Requirement already satisfied: tenacity>=6.2.0 in c:\users\dell\anaconda3\lib\site-packages (from plotly) (7.0.0)
```

String Concatenation

To concatenate, or combine, two strings you can use the + operator this is used for combining all data.

Functions:

A function is a block of code which only runs when it is called. You can pass data, known as parameters, into a function. A function can return data as a result.

User Inputs():

Python allows for user input. That means we are able to ask the user for input.

This is used for analysis of particular data from huge data and call the data from the particular directory.

Python Code:

Steps:

Install plotly:

```
pip install plotly
```

```
Requirement already satisfied: plotly in c:\users\dell\anaconda3\lib\site-packages (5.1.0)Note: you may need to restart the ker
nel to use updated packages.
Requirement already satisfied: six in c:\users\dell\anaconda3\lib\site-packages (from plotly) (1.15.0)
Requirement already satisfied: tenacity>=6.2.0 in c:\users\dell\anaconda3\lib\site-packages (from plotly) (7.0.0)
```

Code for data analysis

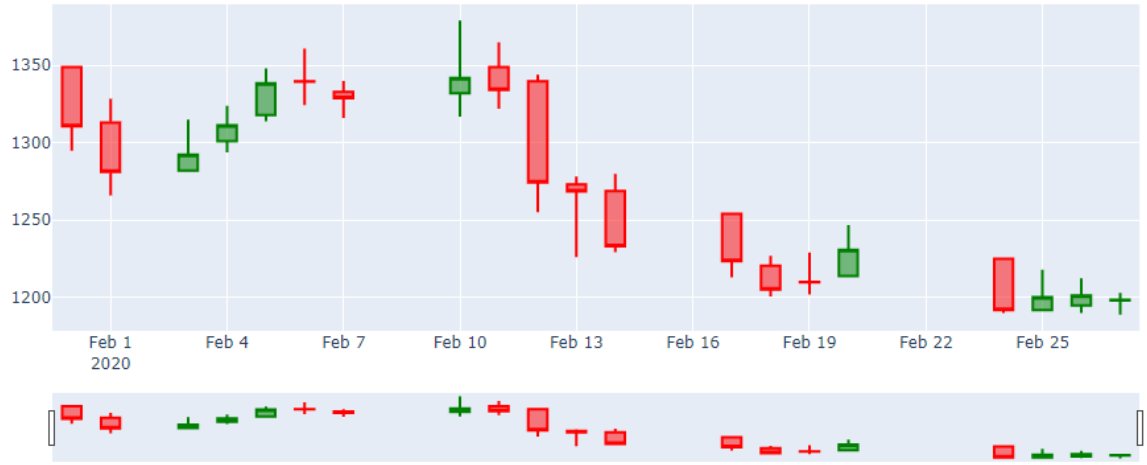
```
%matplotlib inline

import pandas as pd

import plotly.graph_objects as go

import glob

path = r'C:\Users\Dell\.jupyter\Cash_Data'

all_files = glob.glob(path + "/*.csv")

frame = pd.concat([pd.read_csv(f) for f in all_files], ignore_index = True)

#frame.to_csv('output.csv', index = False)

p = input("Enter the keyword:   ")



df = frame[frame.SYMBOL == p]

#k.columns

df.drop(columns=['SYMBOL', 'SERIES', 'LAST', 'PREVCLOSE','TOTTRDQTY', 'TOTTRDVAL', 'TOTALTRADES', 'ISIN','Unnamed: 13'], inplace=True)



df.TIMESTAMP = pd.to_datetime(df.TIMESTAMP)
```

```
df = df.set_index('TIMESTAMP')


figure = go.Figure(
   data = [
         go.Candlestick(
            x = df.index,
           low = df['LOW'],
           high = df['HIGH'],
           close = df['CLOSE'],
           open = df['OPEN'],
           increasing_line_color = 'green',
           decreasing_line_color = 'red'
         )


   ]
)
figure.show()
```

Enter the keyword for the data you want to anlayze:

```
Enter the keyword:   ABB
```

Plotting:



Errors :

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-5-f00e5f4ffcd4> in <module>
     13
     14
---> 15 frame = pd.concat([pd.read_csv(f) for f in all_files], ignore_index = True)
     16 #frame.to_csv('output.csv', index = False)
     17

~\anaconda3\lib\site-packages\pandas\core\reshape\concat.py in concat(objs, axis, join, ignore_index, keys, levels, names, v
fy_integrity, sort, copy)
    269     ValueError: Indexes have overlapping values: ['a']
    270     """
--> 271     op = _Concatenator(
    272         objs,
    273         axis=axis,

~\anaconda3\lib\site-packages\pandas\core\reshape\concat.py in __init__(self, objs, axis, join, keys, levels, names, ignore_
ex, verify_integrity, copy, sort)
    327
    328         if len(objs) == 0:
--> 329             raise ValueError("No objects to concatenate")
    330
    331         if keys is None:
```

This error can be resolved by specifying the path and proper file uploads for .csv files.

Conclusion:

This project have given us an understanding of python programming and its application in different fields. The data analysis of strike price for a month of

February is been studied and have learned about different librarys of mplfinance, plotly and matplotlib with there interpretation in coding.

References :

Youtube link:

https://youtu.be/Ow-tFJThOpA

Google drive link for files(data,ipynb):

https://drive.google.com/file/d/1zYcdzEKIOgejVNvXelGcVPDousT-ja6v/view?usp=sharing