Stock Screener Application

Lokesh Gorrela Krishna Reddy
Statistics MS
UC Davis
lgorrelakrishnareddy@ucdavis.edu

March 21, 2024

1 Abstract

This project aims to address the challenges faced by aspiring investors in forming investment strategies or solutions for equity markets. Existing tools do not do justice for an user to create customized charts and dashboards. By scraping data from financial websites allow one to create one's personalized dashboards with personalized metrics, for top stocks or stocks belonging to a certain sector.This offers an cost effective process.Utilizing python libraries and databases hosted on cloud service providers enables us to store and process data.

2 Motivation

The motivation behind developing the application mainly stems from the need to formulate personal strategies for investment in equity markets. As an aspiring investor myself, I have often encountered challenges or difficulties in finding tools, that cater to my very specific requirements and goals. Existing tools in the market show predefined charts on a generic level, that do not align with my specific requirements. This application would allow me to create my specific dashboards, with personalized metrics to make investment decisions. This application lists and allows me to traverse through articles written about a particular stock or a company and analyze the opinions of experts and investment companies. Existing products do not allow customization and are on a subscription basis, therefore a cheap way to analyze market data is by exploring sites that allow web scraping and storing that data for creating a personalized dashboard, which is a cheap and customized way to analyze data.

3 Methods and Challenges

3.1 Data Description

Being primarily interested in Index investing and cherry-picking the top 50 stocks that are the most profitable and advanced companies in various sectors ranging from Banking, Technology, Manufacturing, Automobiles, Power, Healthcare, Services, Chemicals, etc. Being not knowledgeable enough to perform a very high technical analysis, or macro-market analysis, or micro-market analysis, it is reasonable to find articles written by experts on some trustworthy financial news websites in one place to help us analyze the overall market situation or a stock situation for the upcoming day. Here data for India's top 50 companies in the stock market is being scraped from several sources. The names of the companies, stock symbols associated with those companies, their daily closing prices, and calculating the percentage change in value, and articles associated with those stocks from financial news websites.

3.2 Data Access

The top 50 stocks lie in the index named **Nifty 50**, and the list of all those stocks are scraped from Financial Express website. Amongst a lot of websites researched to access data, financial express provides a complete list at once, compared to truncated list provided by some other websites, and there were no API's provided by the websites.Several articles and websites that provide historical data use stock symbols and company names inconsistently. Stock symbols are not provided by the website mentioned above. The main issue while accessing this data and all the data following is the **403 Access Denied** response error. Needed to set up user-agent in the request headers, this agent helps server deliver content that is compatible with the client, considering factors such as browser type, version, and operating system.

Data for stock symbols was not provided anywhere in the financial express website. Lot of data sources were checked for stock symbols, until they were found in Groww website, which is also a stock trading service provider. The company names had to be processed to match the syntax accepted by the service provider. For example, a company with name *HDFC Life Insurance Company Ltd.*, need to be processed to *hdfc-standard-life-insurance-co-ltd.* There were some inconsistencies because of renaming, mergers and acquisitions. Some of those include *Larsen & Toubro Ltd.* as *larsen-toubro-ltd*, *LTIMindtree Ltd.* as *larsen-toubro-infotech-ltd*, and *Tata Consumer Products Ltd.* as *tata-global-beverages-ltd. Naming inconsistency* is a major challenge, since each and every name processing needs to be validated. The ***url*** needs to be formatted accordingly for each company.

Stock symbols were acquired, and multiple resources were checked for historic daily data for each stock, for customized time ranges. Many such service providers were on a subscription model and many free resources doesn't provide data at various customization's like, time ranges and frequency. Yahoo Finance website provides the historic data for free and at various levels of customization. Needed a symbol conversion for making requests from *ADANIENT* to *ADANIENT.NS*. Request payload had to be customized for required data frequency, and time ranges. The request returns an JSON object with all the required data. Only a part of the object data is being accessed for this particular application are *epoch timestamps*, and daily *closing prices*. A data frame with columns *stock symbol*, *timestamp*, and *closing prices* is created. The epoch timestamp column is then converted into a *datetime* object into a 'YYYY-MM-DD' format.

Data is being pushed into *Github* for back up and data is being pushed into a *MySQL* database hosted on *Amazon Web Services*. Figuring out how to host a database on internet is the hard part, but hosting services provides a very easy interface to host database. Pre-built libraries like *PyGithub* allows an easy interface to connect to *Github* API's and upload the data as csv files. One need to create an access key from the Github portal to access Github API's. Hosting a database on AWS (Amazon Web Services) is well documented and an easy interface is provided. Check for Relational Database Service and Select the type of database from SQL, MySQL, or PostgreSQL. Enabling database to be accessed from anywhere is the most challenging part. Security groups had to be configured to allow TCP/IP connections from all the sources, and that security group needs to be assigned to the database configurations. Free tier does not support applying the new security group assignment instantly, this will be applied at a three day scheduled job run.

Currently data is being pushed into three tables in the database. Tables description is provided in Table 1. Python has the ability to access databases using libraries like SQL Alchemy, which is the Python SQL toolkit and Object Relational Mapper that gives application developers the full power and flexibility of SQL. Making sure that database is accessible is the only challenge, SQL Alchemy provides a well documented resource to access the SQL database that was created on AWS and write data into the tables using data frames.

Articles text in the Financial express, in the first 100 pages is scraped and matched against the company names and stock symbols. The matching pattern can be further fine tuned to shorter form of company names that could be written in the articles. The available text can be assigned for each company/stock symbol and date published, can

symbols	**stock_data**	**change_stock_data**
name	symbol	symbol
symbol	date	date
	closing_price	closing_price
		change_value

Table 1: Tables and Columns

symbol	date	closing_price
ADANIENT	2023-08-14	3212
ADANIENT	2023-08-13	3156
ICICI	2023-08-14	1560
ICICI	2023-08-13	1540

Table 2: Original Table

date	ADANIENT	ICICI
2023-08-13	3156	1540
2023-08-14	3212	1560

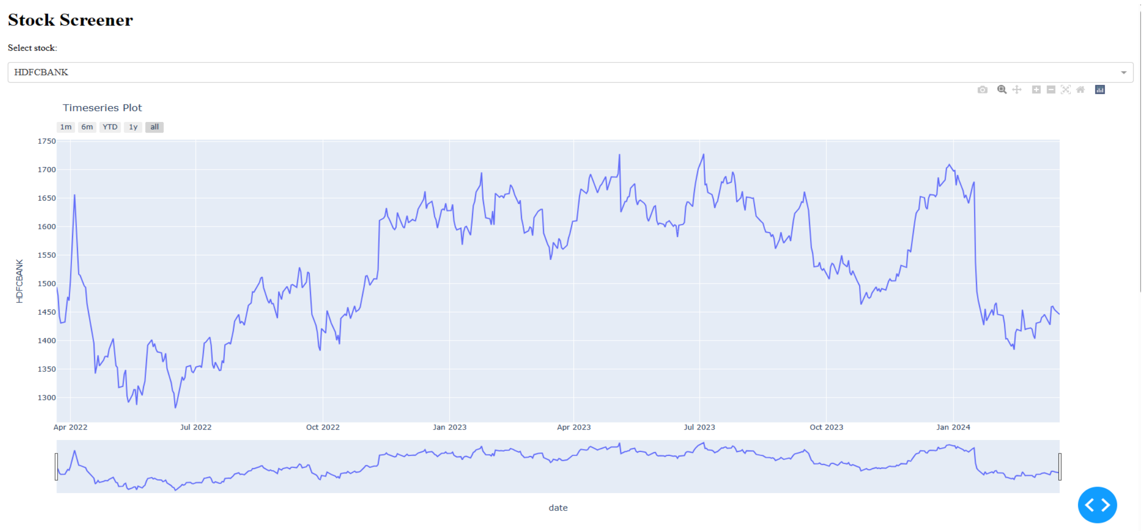
Table 3: Transformed Table

be stored for sentiment analysis. Future plans involve manual encoding each articles text as positive and negative and perform sentiment analysis using that data. Free tier hosting does not allow to hold lot data, hence this data is not pushed into database.

3.3 Data visualization

The current visualization dashboard currently shows the time series chart, and bar chart showing the percentage change from previous trading day. Dash and plotly are used to create the visualization with time range sliders for closer and small time range inspections. Future plans include hosting it on a server and create a domain to access this site with all the dashboards and widgets with more data and metrics to formulate personal investment solutions. The most challenging part of visualization is to transform an existing data frame that matches the graph data structure. Transformation structure is shown in Table 2 and 3. In the future, we strive to display articles that belong to the selected stock from the drop down. A bar chart is displayed showing the count of articles that specifies a certain company or stock symbol in the first 20 pages of Financial express.

Figure 1: Timeseries chart



Stock Screener

Select stock:

HDFCBANK

4 Conclusion

In conclusion, this application developed aims to address the challenges faced by investors in formulating personalized investment strategies. Motivated by need for personalized tools and metrics, this application is developed and will be developed continuously to cater personal needs. This project's methodology involved scraping data from various sources, including financial news websites and stock trading service providers. Challenges such as inconsistent data formats, naming inconsistencies, and access restrictions were overcome through meticulous data processing and configuration adjustments. Data visualization capabilities were implemented using Dash and Plotly libraries.Overall, the application provides users with valuable insights into recent market trends and facilitates informed decision-making regarding investments. Future improvements, such as integrating NLP models for sentiment analysis, aim to further enhance the application's functionality and utility for users in navigating the complexities of equity markets.

Figure 2: Percentage change chart

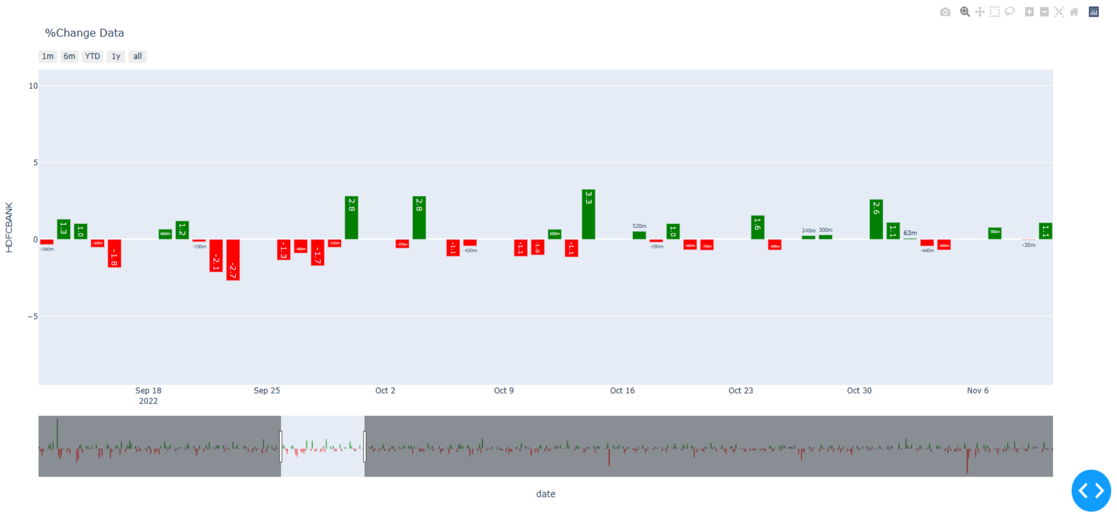
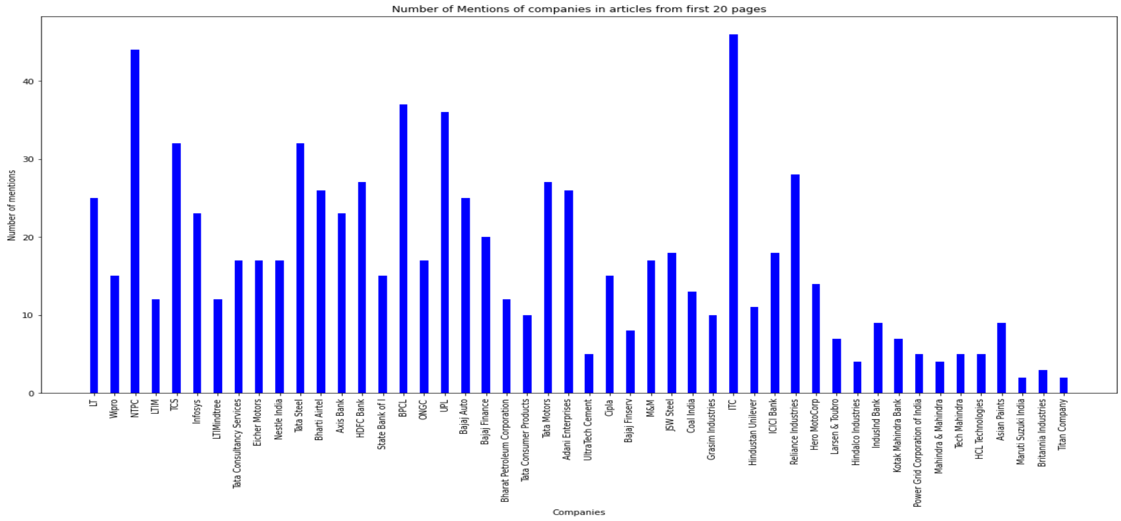


Figure 3: Mentions Chart


Number of Mentions of companies in articles from first 20 pages
Number of mentions
LT
Wipro
NTPC
LTIM
TCS
Infosys
LTIMindtree
Tata Consultancy Services
Eicher Motors
Nestle India
Tata Steel
Bharti Airtel
Axis Bank
HDFC Bank
State Bank of I
BPCL
ONGC
UPL
Bajaj Auto
Bajaj Finance
Bharat Petroleum Corporation
Tata Consumer Products
Tata Motors
Adani Enterprises
UltraTech Cement
Cipla
Bajaj Finserv
M&M
JSW Steel
Coal India
Grasim Industries
ITC
Hindustan Unilever
ICICI Bank
Reliance Industries
Hero MotoCorp
Larsen & Toubro
Hindalco Industries
IndusInd Bank
Kotak Mahindra Bank
Power Grid Corporation of India
Mahindra & Mahindra
Tech Mahindra
HCL Technologies
Asian Paints
Maruti Suzuki India
Britannia Industries
Titan Company
Companies