Financial Analysis Using COVID Data

Saurabh Singh, Viddhi Lakhwara, Aditya Panchal, Hemang Bhanushali

♦ Problem Statement:

With the advancement of a global pandemic, the past few months have been nothing short of chaos. This unforeseen hit has not only impacted our everyday lives but also gravely affected various parts of the economy including financial markets.

Our aim is to obtain a deeper understanding on how certain indicators like stock market indexes, forex exchange rates, commodity market, etc. were influenced by covid and up to what extent it had an impact on these economic sectors.

Since these indicators differ from one another, they tend to have a contrasting relation with the impact of COVID-19. For instance, while stock markets might be more vulnerable to the pandemic, foreign exchange rates could have been comparatively more stagnant. Hence, understanding these differences would allow us to gain a better insight into the financial market.

♦ Methodology:

In order to find the impact of COVID-19 on the financial market, we source a comprehensive range of certain economic indicators, explore their features and attributes and depict how each indicator correlates to COVID-19.

Some of the indicators we have taken into account are:

- Stock Market Indexes (with each stocks pricing history)
- Foreign Exchange Market Rate
- Commodity Market (which includes goods like energy, grains, metals, etc)
- Cryptocurrency Market

For data collection we scanned through various resources and found appropriate data sets. This data includes recent changes in the indicators which enables us to find trends associated with each. Most of the data is from NASDAQ, which yields data from a wide range of markets.

We performed ETL operations to clean and organize the data in a uniform pattern. The COVID-19 data had missing columns needed for analysis. Hence, we established relations with other columns and created new fields.

For data analysis, we mainly focused on finding trends between each indicator and COVID.

After examining a relationship between the two, we draw a comparison in an effort to find indicators that were highly susceptible to COVID-19.

Technologies Used:
- Apache Spark
- Cassandra
- Tableau
- Machine Learning models

♦ **Problems and Solutions:**

- Lack of COVID Data: We surfed through multiple sources in order to get current covid data for the duration of 2 years. Most available data either ranged from Jan 2020 to Dec 2020 or only contained information about the second phase. After researching, we found an adequate dataset from: COVID dataset link

- Cleaning COVID Data: In an effort to efficiently utilize this dataset, it had to be heavily cleaned and organized. Since there was no data for the number of new cases on each day, we calculated it by exploiting other columns (new_cases = total_cases – cases_on_previous_day)

- Different Data Sources: Since we made use of multiple indicators, the data for each was obtained from a diverse set of sources. Due to this, variation we had to modify them into a homogeneous format.

- Date Format: To maintain this uniformity, we transformed all "date" columns from 'mm/dd/yyyy' to 'yyyy-mm-dd'

- ML Models: We played with various machine learning algorithms and hyperparameters with the aim of arriving at an optimal model. After experimenting with various scenarios, we decided to go with GBT Regressor.

- Bash Script: Seeing as we had a large array of economic indicators to assess, we deployed bash scripts for every python file. This enabled us to run the entire project using a single file.

- **Results:**

  - Stock Market Indexes

    As the pandemic began, so did a series of extreme turmoil and uncertainty. Panic triggered by economic consequences led to a crash in the stock market. We compared this with both US and global COVID data.

    Visualization: Dashboard for US Stock Market affected by US and Global COVID Cases

    - In US, there was a significant decrease in the percentage change of close values.
    - The stock that took the largest hit at the beginning was RUT; which represents small scale companies. Once the cases started dying down, RUT also rose more than the rest, making it the most influenced by covid.
    - While these stocks experienced various fluctuations throughout, they never stooped as low as they did at the onset of the pandemic.

  - Commodity Market

    We considered four different types of commodities: metals, energies, grains, and meats, each of which were impacted differently by COVID.

    Visualization: Dashboard for US Commodities affected by US Covid Cases

    Following are our observations:
    - Metals like copper and silver dropped with the onset of the pandemic and also took a hit during both waves. China is the largest consumer of copper and due to the lockdown, it's demand was hampered. On the other hand, gold recovered quickly and stayed comparatively stable in contrast to the other metals.
    - Crude oil took a deep plunge in the beginning while other energies remained slightly uniform. This could be due to the fact that during quarantine, there was less transportation and low energy consumption.
    - Most meat products experienced a dramatic drop while continuing to fluctuate later on. Since the pandemic resulted in the closure of many factories and restaurants, the consumption of meat was also influenced.
    - Compared to others, the grain market had slightly varying trends within itself. Soybean and soybean oil dropped in the start whereas rice significantly increased. The increase of rice synchronizes with the fact that consumers hoarded non-perishable food items.

- Cryptocurrency Market

We took into account cryptocurrencies like Bitcoin, Ethereum and Doge and compared it with both Global and US COVID data

Visualization: Dashboard for Cryptocurrencies affected by Global Covid Cases, Dashboard for Cryptocurrencies affected by US Covid Cases

- The price of most cryptocurrencies plunged by over 50% within a short span of time.
- With an increase in the number of cases this market also rose, while lower cases resulted in lesser bitcoin values.
- All proved to be volatile to the number of covid cases.
- It was affected by cases in the US but did not necessarily have much of an impact on a global scale during the first wave.
- From the visualization, we infer that cryptocurrency was not affected by covid cases as when covid cases increased cryptocurrencies increased and vice versa.

Hence, change in the number of cases did not significantly impact this market in the later stages of the pandemic.

- Foreign Exchange Market

We collected forex rates of currencies like: Brazilian Real (BRL), Canadian Dollar (CAD), Mexican Peso (MXN), Swiss Franc (CHF), Euro (EUR), British Pound Sterling (GBP), Russian Ruble (RUB), Australian Dollar (AUS), Indian Rupee (INR) and Japanese Yen (JPY). In order to better analyze the data, we divided these into three regions (Eastern, Central, Western).

Visualization: Dashboard for Western Region Currencies affected by US Covid cases and Global Covid Cases,
Dashboard for Central Region Currencies affected by US Covid cases and Global Covid Cases,
Dashboard for Eastern Region Currencies affected by US Covid cases and Global Covid Cases

This is how they were impacted by COVID:

- In Western parts of the world, all currencies shot up initially and then gradually decreased. CAD increased slightly during the initial phases but recovered well overall. BRL rose the most and was inversely related to covid cases throughout.
- Central regions had more uniformity during the duration. EUR and GBP dropped down had slight changes there that mostly remained consistent. On the other hand, CHF showed dramatic ups and downs.

- In Eastern regions, INR and JPY dropped after the initial impact and consistently stayed low while rising after the first wave. AUS had a sharp decline in the beginning but recovered and soared later on.

Currencies that did not fluctuate significantly, depict that they have a more robust monetary system and hence weren't largely impacted. The difference in rise and fall of different currencies might be due to a difference in this system and gaps within the country's economy.

- **Results after Training ML Models:**

Stock Type	Accuracy
DOWJONES	93.62
NDX COMPOSITE	93.09
NDX	93.21
NYSE	92.89
RUT	92.17
SPX	93.43

Crypto Type	Accuracy
Binance	93.15
Bitcoin	92.91
Dash	92.87
Dogecoin	92.73
Ethereum	93.02
Litecoin	92.96
Ripple	92.88
Monero	92.79

Forex Type	Accuracy
Australian Dollar (AUD)	93.12
Brazilian Real (BRL)	92.36
Canadian Dollar (CAD)	92.41
Swiss Franc (CHF)	92.22
Euro (EUR)	92.48
Pound Sterling (GBP)	92.57
Indian Rupee (INR)	92.29
Japanese Yen (JPY)	93.18
Mexican Peso (MXN)	92.20
Russian Rouble (RUB)	92.46

Crypto Type	Accuracy
Crude Oil	93.18
Heating Oil	92.92
Natural Gas	93.27
Milk	92.86
Feeder Cattle	92.90
Live Cattle	92.74
Rough Rice	93.32
Soybean Oil	92.11
Soybeans	92.89

- **Project Summary:**

A summary of the emphasis/priorities in our project in the following categories:

Getting the data	Acquiring/gathering/downloading	3
ETL	Extract-Transform-Load work and cleaning the data set.	3
Problem	Work on defining problem itself and motivation for the analysis.	2
Algorithmic work	Work on the algorithms needed to work with the data, including integrating data mining and machine learning techniques.	3
Bigness/parallelization	Efficiency of the analysis on a cluster, and scalability to larger data sets.	3
UI	User interface to the results, possibly including web or data exploration frontends.	1
Visualization	Visualization of analysis results.	3
Technologies	New technologies learned as part of doing the project	2

Total: 20