

Dataset

The dataset used for this project is the stock price data of Sony, obtained from the Yahoo Finance library. The dataset contains historical stock price data for the company above, including attributes such as:

- Date: The date for which the stock price data is available.
- Open: The opening price of the stock for the day.
- High: The highest price of the stock for the day.
- Low: The lowest price of the stock for the day.
- Close: The closing price of the stock for the day.
- Adj Close: This is the adjusted closing price of the stock for the day, considering factors such as stock splits and dividends.
- Volume: The number of shares traded for the stock for the day.

The Yahoo Finance website obtained the dataset, which provides historical stock price data for various companies. It was obtained using the pandas-datareader and YahooFinance library in Python, which allows for easy retrieval of data from the Yahoo Finance website.

The dataset covers several years, with daily data points for the company. It is time-series data, which allows for analyzing trends and patterns in stock prices over time. The dataset is also suitable for making predictions about future stock prices using techniques such as machine learning and statistical modeling.

The dataset is clean and well-organized, with each attribute clearly labeled and consistent data types. The dataset is also free of missing values and errors, making it suitable for analysis and modeling.

In summary, the dataset is a comprehensive and reliable source of historical stock price data for Sony, which can be used for various analysis and predictive modeling tasks.

The dataset is well-organized, clean, and error-free, making it an ideal choice for this project.

Literature Review

Data visualization is a graphical representation of information and data. It communicates complex data relationships in a way that is easy to understand. “Data visualization is useful for data cleaning, exploring data structure, detecting outliers and unusual groups, identifying trends and clusters, spotting local patterns, evaluating modeling output, and presenting results(Unwin, 2020).” Python visualization capabilities allow users to create easy-to-understand visual representations of complex data sets that ultimately allow inexperienced users to use these insights to make data-driven decisions. “It is essential for exploratory data analysis and data mining to check data quality and help analysts become familiar with the structure and features of the data before them (Unwin, 2020).” Python provides various libraries and features used for visualizing data. Panda, Matplotlib, NumPy, and Seaborn are popular data analysis, visualization, and statistical computing libraries.

Python’s Panda is used for work involving data sets. It is an object-oriented high-performance programming library that “allows us to analyze big data and make conclusions based on statistical theories. [It] can clean messy data sets and make them readable and relevant,” which is essential in data science(W3Schools). It supports the Matplotlib and NumPy library. Pandas have three fundamental data structures: Series, data frame, and Panel. A panda Series is a one-dimensional array of homogenous data that can be created from a list or an array. You can describe it like a column in a table. A data frame can be described as an Excel spreadsheet for Python; it’s a two-dimensional array with heterogeneous data. Just like we described a Series as a column in a table, you can think of a DataFrame as a whole table. You can load data from many sources, such as CSV, HTML, and JSON, and store it in the DataFrame. A Panel is a three-dimensional data structure with heterogeneous data. It can be described as a container of DataFrames. Pandas use the plot() method to create diagrams.

Pyplot is a submodule of Matplotlib, which allows you to visualize diagrams on screen. “Matplotlib is a comprehensive library for creating static, animated, and

interactive visualizations in Python”(Matplotlib). It was originally written as a Python alternative for MATLAB users, which is reflected in its syntax. Matplotlib has dual interfaces: the pyplot/functional interface and the object-oriented interface. The object-oriented interface is an explicit interface that uses methods on a figure or axes objects to create and build a visualization step-by-step. The implicit pyplot interface is a state-based interface that preserves various function calls and allows users to quickly and efficiently generate plots. It allows users to add/modify plots as needed.

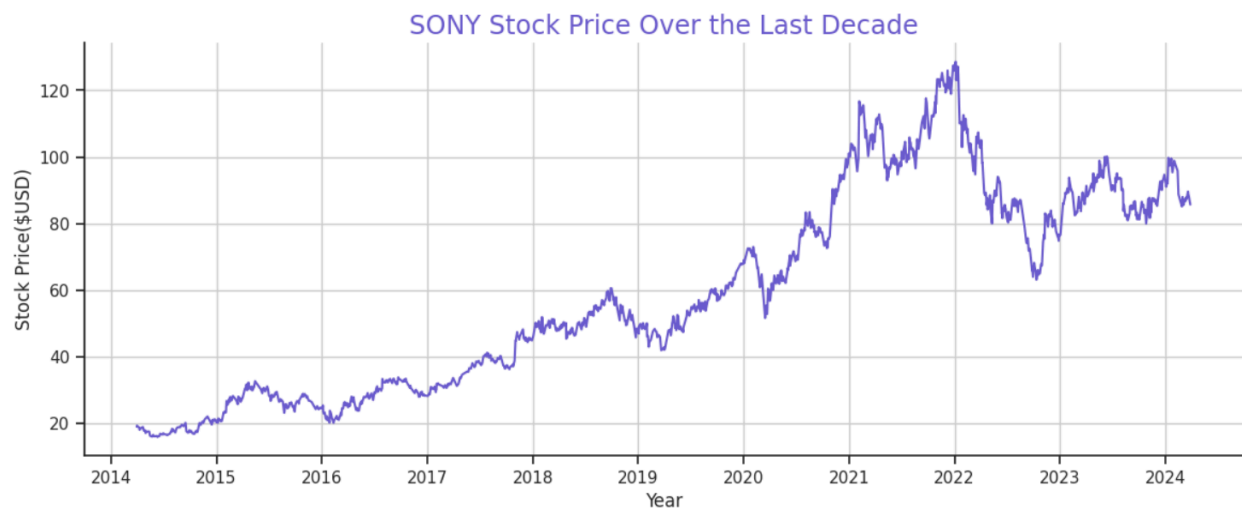
“[Seaborn] provides a high-level interface to MATPLOTLIB and integrates closely with Pandas data structures. Functions in the seaborn library expose a declarative, dataset-oriented API that makes it easy to translate questions about data into graphics that can answer them”(Wascom, 2021). It provides default styles and color palettes, making statistical plots more appealing to users. It is also useful because one can switch between different visual representations for the same variables to better understand a dataset.

Pandas has a built-in library called NumPy, used for scientific computing. NumPy, which stands for Numerical Python, is used to manipulate multidimensional array objects. Why Lists should not be used instead of NumPy? NumPy is significantly faster than Lists because it is stored in memory. Numpy arrays are stored as continuous blocks of memory, while Python lists are stored as small blocks scattered in memory. This causes memory access for Numpy arrays to be faster than lists. Having most of NumPy’s operations implemented in C reduces the cost(the cost of loops) and increases speed (the use of parallelism).

We used these libraries to help us assist with documenting and analyzing the stock data for Sony. We used Panda, Matplotlib, NumPy, Seaborn, and other statistical analysis tools to help us relate a data story. We analyzed and observed our stock data for the past decade, including the Covid-19 pandemic. These libraries made it easier for us to see how the stock prices moved visually and tell a visual story to our audience.

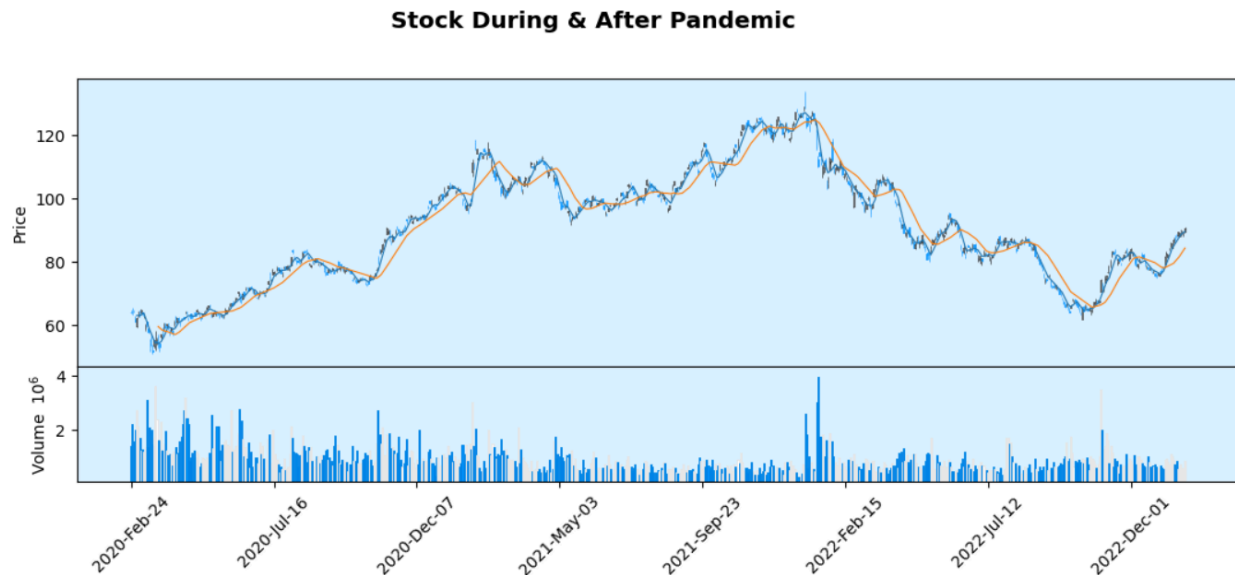
SONY (SONY)

Sony Group Corporation “engages in developing, designing, manufacturing, and selling electronic equipment, instruments, devices, game consoles, and software for consumers, professionals, and industrial markets (Sony, WSJ).” We graphed Sony’s stock closing price for the last decade, as shown in the figure below.

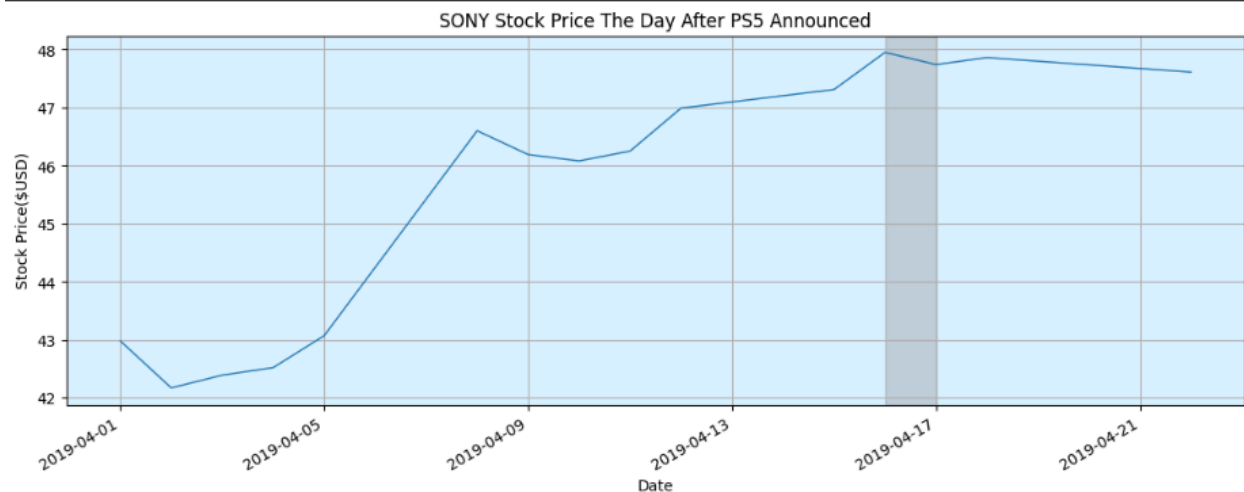


As you can see, Sony’s value has increased over time, reaching its highest stock value during the last decade, precisely on December 28, 2021 (during the COVID-19 pandemic). Many things led to this increase in value, such as Sony acquiring Firesprite, Blueprint Games, and, most importantly, Valkyrie Entertainment during September and December 2021. Valkyrie Entertainment produces high-quality art productions that can be used on consoles and PCs. They have produced the games League of Legends, Valorant, and, most notably, God of War, which has sold over 15 million units since its release in 2018. As you can see below, I have created a candle stick chart documenting Sony's overall stock price and volume during and after the pandemic. The real body of the candle stick represents the ‘Open’ and ‘Close’ of the day’s trading. The blue represents that the ‘Close’ was lower than the ‘Open,’ and the white represents that the ‘Close’ was higher than the ‘Open.’ The volume is the number of stocks traded per period, so every transaction between a buyer or seller contributes to the overall volume count of that stock. The volume pattern of the graph can help indicate the stock’s

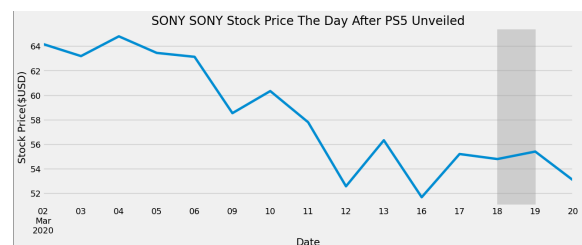
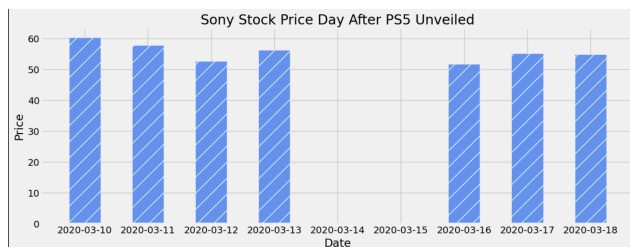
strength- an increasing volume means that the stock has higher liquidity. As you can see, around the December 2021 time period, the volume is high, meaning that the stock, at that time, is strong/healthy.



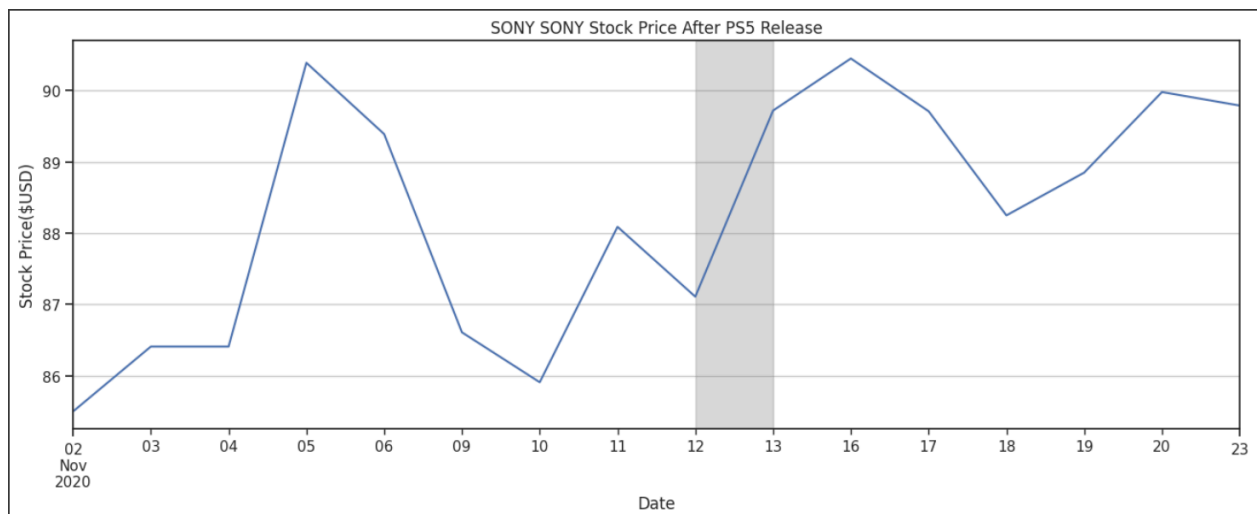
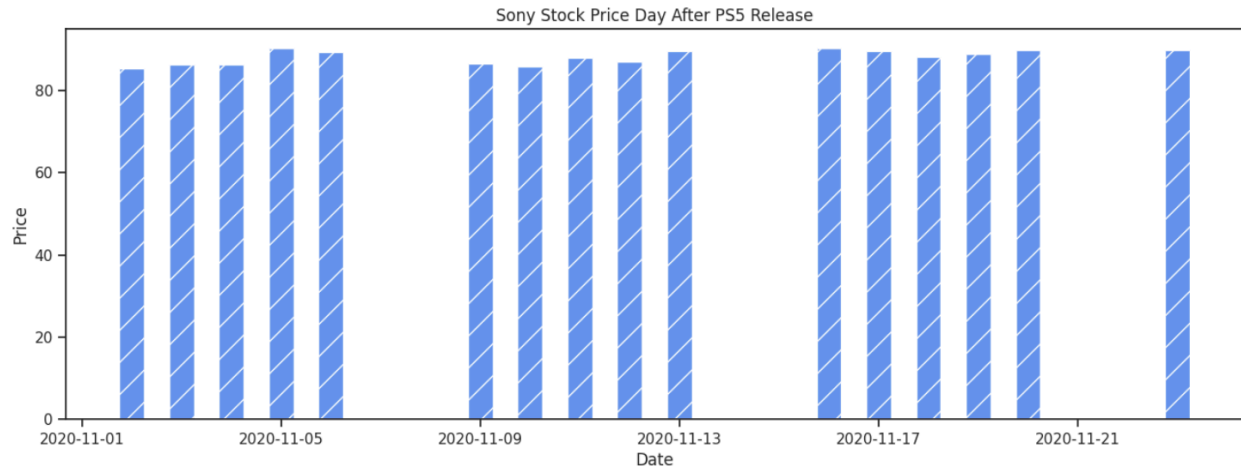
Sony also announced and released the PlayStation 5 console during the pandemic. On April 16th, 2019, Sony's system architect, Mark Cerny, announced that the company was working on the PlayStation 5. Aside from the PlayStation 5 having more software and hardware upgrades than its predecessors, the PS5 features streaming platforms such as Netflix, Hulu, etc. It also lets you download and use apps such as Discord and Spotify. It's the perfect entertainment system where it can potentially replace the use of cellular and cable services.



As you can see from the graphs, although the PlayStation 5 was announced on April 16th, 2019, there wasn't a significant hike in the stocks. I believe this can be potentially because Cerny mentioned the idea of the Play Station 5 being developed but did not necessarily announce that the gaming console was in production and going out for sale soon. On March 18th, 2020, Sony unveiled the PlayStation 5 at the CES press conference and to the public. You can see from the graphs below that Sony's stock slightly increased.



Sony released the PlayStation 5 to the public on November 12th, 2020. The graphs below show an increase in stock due to the demand for the consoles. Compared to the stocks when the PS5 was unveiled to the PS5 being released, it increased from the 50's to the 80s range, which you can notice from the bar graphs presented. Sony has sold over 50 million units of the PS5 gaming consoles worldwide to this date.



Sony was able to stay afloat during the pandemic due to acquiring different popular companies such as Crunchyroll anime business, and other companies. Overall, we can say Sony does a good job of noticing current trends and adapting its products and technologies to the consumers' appetite.