ADS506 Assignment 1.2 sp500_stocks

Gabriel E. Mancillas Gallardo

2024-10-28

Assignment: Propose a Time Series Dataset for Your Final Project ### Data Source

The data source is the S&P 500 stock data from Kaggle however becasue the data exceed the 50MB compacity we will only include 2022 - 2024 dataset. The dataset contains the stock prices of the S&P 500 companies from 2013 to 2018. The dataset contains the following columns:

```r
#library(<PACKAGE_DEPENDENCIES>)
library(readr)
library(dplyr)
library(tsibble)

# Load data
sp500_data <- read_csv("/Users/gabrielmancillas/Documents/GitHub/S-P-500/data/sp500_stocks_filtered.csv"

# Show data
head(sp500_data)
```

```
## # A tibble: 6 x 8
##   Date       Symbol `Adj Close` Close High   Low  Open  Volume
##   <date>     <chr>        <dbl> <dbl> <dbl> <dbl> <dbl>   <dbl>
## 1 2022-01-03 MMM           131.  149.  150.  147.  149. 2309117
## 2 2022-01-04 MMM           132.  151.  152.  149.  149. 3016551
## 3 2022-01-05 MMM           132.  150.  152.  148.  148. 3531070
## 4 2022-01-06 MMM           131.  149.  152.  148.  151. 2996458
## 5 2022-01-07 MMM           132.  150.  151.  148.  149. 3349039
## 6 2022-01-10 MMM           130.  148.  150.  148.  150. 2772328
```

```r
#total number of rows and columns
dim(sp500_data)
```

```
## [1] 356124      8
```

Data Source

Include public links to data if it is too large to upload (do not upload datasets larger than 50MB). **There is no need to upload the data (its under 40MB)** [github repository[(https://github.com/Gabeleo24/S-P-500/tree/main/data)

Time Series Plot

```r
# Check for duplicates
duplicates <- sp500_data %>%
    group_by(Date) %>%
    filter(n() > 1)
```

```r
# Remove duplicates if any
sp500_data <- sp500_data %>%
    distinct(Date, .keep_all = TRUE)

# Plot the time series
library(tsibble)
library(ggplot2)

# Convert to tsibble for time series structure
sp500_data_ts <- sp500_data %>%
    as_tsibble(index = Date)

# Time series plot for Adjusted Close prices
ggplot(sp500_data_ts, aes(x = Date, y = `Adj Close`)) +
    geom_line(color = "blue") +
    labs(
        title = "S&P 500 Adjusted Close Prices (2022-2024)",
        x = "Date",
        y = "Adjusted Close Price (USD)"
    ) +
    theme_minimal() +
    theme(
        plot.title = element_text(hjust = 0.5, size = 16),
        axis.text.x = element_text(angle = 45, hjust = 1)
    )
```

S&P 500 Adjusted Close Prices (2022–2024)
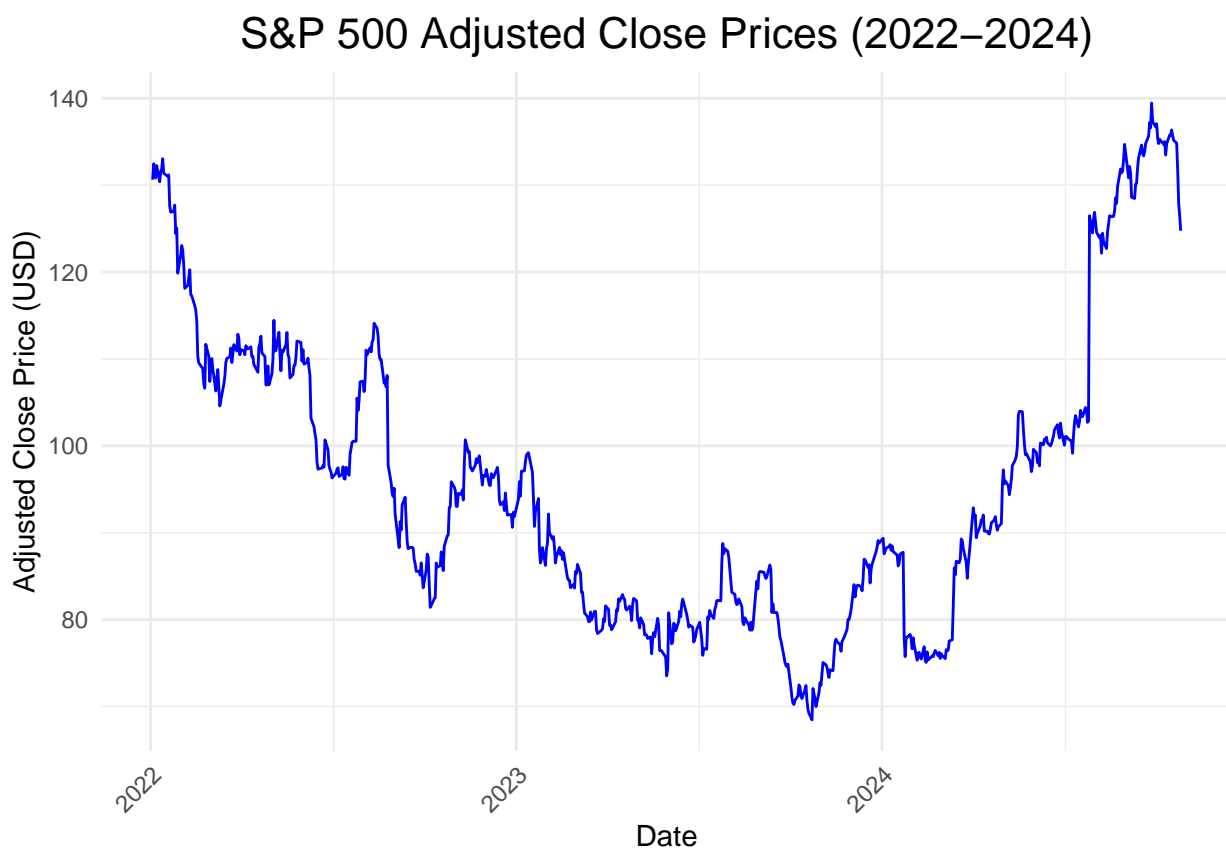
Discussion

The S&P 500's Adjusted Close Prices experienced notable changes from 2022 to 2024, with a prominent decrease in early 2022, a recovery throughout 2023, and a further increase into 2024. This trend probably indicates larger economic issues in 2022, like worries about inflation and slowdowns, followed by a possible bounce back influenced by better investor confidence and stabilization-focused economic strategies.

Market volatility is highlighted by data showing significant drops and rebounds, revealing a market that is greatly influenced by global events. This time frame highlights key patterns in investor actions, demonstrating caution when the market is down and regained trust as it bounces back.

Predicting outcomes using this set of data could be extremely beneficial for investors and financial analysts who want to predict future trends and modify their investment strategies accordingly. Yet, the stock market's innate unpredictability, particularly in a tumultuous period such as 2022-2024, presents obstacles to precise prediction. Adding more economic indicators and market factors would probably improve the accuracy of predictions and offer stronger insights for planning strategic investments.