Live Case: S&P500 (1 of 3)

Aug 18, 2023.

Overview of the S&P 500

1. The S&P 500, also called the Standard & Poor's 500, is a stock market index that tracks the performance of 500 major publicly traded companies listed on U.S. stock exchanges. It serves as a widely accepted benchmark for assessing the overall health and performance of the U.S. stock market.
2. S&P Dow Jones Indices, a division of S&P Global, is responsible for maintaining the index. The selection of companies included in the S&P 500 is determined by a committee, considering factors such as market capitalization, liquidity, and industry representation.
3. The S&P is a float-weighted index, meaning the market capitalization of the companies in the index are adjusted by the number of shares available for public trading. [1]
4. The performance of the S&P 500 is frequently used to gauge the broader stock market and is commonly referenced by investors, analysts, and financial media. It provides a snapshot of how large-cap U.S. stocks are faring and is considered a reliable indicator of overall market sentiment.
5. Companies that relatively have the highest Market Capitalization include Apple, Microsoft, Amazon, Google's parent company, Alphabet. [1]

S&P 500 Data - Preliminary Setup

1. We will analyze a real-world, recent dataset containing information about the S&P500 stocks. The dataset is located in a Google Sheet [2]
2. Load necessary libraries, suppressing annoying start up messages.

```
library(knitr)
suppressPackageStartupMessages(library(dplyr))
suppressPackageStartupMessages(library(kableExtra))
```

```
suppressPackageStartupMessages(library(lubridate))
library(gsheet)
```

Read the S&P500 data from a Google Sheet into a tibble

1. The complete URL is https://docs.google.com/spreadsheets/d/11ahk9uWxBkDqrhNm7qYmiTwrlSC53N1zvXYfv7ttOCM/
2. The Google Sheet ID is: `11ahk9uWxBkDqrhNm7qYmiTwrlSC53N1zvXYfv7ttOCM`. We can use the function `gsheet2tbl` in package `gsheet` to read the Google Sheet into a tibble or dataframe, as demonstrated in the following code.

```
# Read S&P500 stock data present in a Google Sheet.
library(gsheet)
prefix <- "https://docs.google.com/spreadsheets/d/"
sheetID <- "11ahk9uWxBkDqrhNm7qYmiTwrlSC53N1zvXYfv7ttOCM"
url500 <- paste(prefix,sheetID) # Form the URL to connect to
sp500 <- gsheet2tbl(url500) # Read it into a tibble called sp500
```

3. **Discussion**:

- We load the `gsheet` library to fetch Google Sheets as data frames.
- We create the URL for the specific Google Sheet containing S&P500 stock data.
- Using the `gsheet2tbl` function, we read the data from this Google Sheet into a tibble named `sp500`.

Date

4. We identify the date of the data.

```
d1 <- unique(sp500$Date)
d2 <- mdy(d1)
date <- format(d2, "%d %B %Y") #Save the date
```

- We have 503 stocks of the S&P500 in our dataset, as of 13 September 2023.

Reviewing the data types

1. We need to understand the different data columns and their data types.

```
dataTypes <- tibble(
  ColumnName = names(sp500),
  DataType = sapply(sp500, class)
)
kable(dataTypes, caption = "Data Columns and their data types")
```

3. **Discussion**:

- We create a tibble named `dataTypes` that lists the column names from the `sp500` tibble and the corresponding data types for each column.
- We use `kable` to display this information as a neatly formatted table.
- Note that we use the `sapply()` function to apply the `class()` function to every column in the `sp500` dataset. This will retrieve the data type of each column.

4. **Data Columns**:

- The columns labeled `Date`, `Stock`, `Description`, `Sector`, and `Industry` are character columns. They respectively represent the date, stock ticker symbol, description, sector, and industry of each S&P500 stock.
- Columns such as `Market Capitalization`, `Price`, `52 Week Low`, `52 Week High`, and other numeric columns contain diverse financial metrics and stock prices related to the S&P500 stocks.
- The column labeled `Technical Rating` gives a Buy / Sell technical rating for each stock.

Remove Rows containing no data or Null values

- We check if the "Stock" column in the sp500 dataframe contains any null or blank values and removes them.

```
hasNull <- any(sp500$Stock == "" | is.null(sp500$Stock)) # Check for blank or null values
if (hasNull) {
    sp500 <- sp500[!(is.null(sp500$Stock) | sp500$Stock == ""), ] # Remove any null or bla
}
```

Table 0.1: Data Columns and their data types

ColumnName	DataType
Date	character
Stock	character
Description	character
Sector	character
Industry	character
Market Capitalization	numeric
Price	numeric
52 Week Low	numeric
52 Week High	numeric
Return on Equity (TTM)	numeric
Return on Assets (TTM)	numeric
Return on Invested Capital (TTM)	numeric
Gross Margin (TTM)	numeric
Operating Margin (TTM)	numeric
Net Margin (TTM)	numeric
Price to Earnings Ratio (TTM)	numeric
Price to Book (FY)	numeric
Enterprise Value/EBITDA (TTM)	numeric
EBITDA (TTM)	numeric
EPS Diluted (TTM)	numeric
EBITDA (TTM YoY Growth)	numeric
EBITDA (Quarterly YoY Growth)	numeric
EPS Diluted (TTM YoY Growth)	numeric
EPS Diluted (Quarterly YoY Growth)	numeric
Price to Free Cash Flow (TTM)	numeric
Free Cash Flow (TTM YoY Growth)	numeric
Free Cash Flow (Quarterly YoY Growth)	numeric
Debt to Equity Ratio (MRQ)	numeric
Current Ratio (MRQ)	numeric
Quick Ratio (MRQ)	numeric
Dividend Yield Forward	numeric
Dividends per share (Annual YoY Growth)	numeric
Price to Sales (FY)	numeric
Revenue (TTM YoY Growth)	numeric
Revenue (Quarterly YoY Growth)	numeric
Technical Rating	character