

Case Analysis of SP500

We will investigate data related to the S&P500 stocks.

This is the URL of the data we will use: <https://docs.google.com/spreadsheets/d/11ahk9uWxBkDqrhNm7qYmiTwrlSC53N1zvXYfv7tt0CM>

1) Reading S&P500 data from a Google Sheet into a tibble

The Google Sheet ID is: 11ahk9uWxBkDqrhNm7qYmiTwrlSC53N1zvXYfv7tt0CM. We can use the function `gsheet2tbl` in package `gsheet` to read the Google Sheet into a 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 <- "11ahk9uWxBkDqrhNm7qYmiTwrlSC53N1zvXYfv7tt0CM"
# Form the URL to connect to
url500 <- paste(prefix, sheetID)
# Read the Google Sheet located at the URL into a tibble called sp500
sp500 <- gsheet2tbl(url500)
```

No encoding supplied: defaulting to UTF-8.

1) Reviewing the data

We run the `str()` function to better understand the data.

```
str(sp500)
```

```
Classes 'tbl_df', 'tbl' and 'data.frame':   503 obs. of  36 variables:
 $ Date                : chr  "7/15/2023" "7/15/2023" "7/15/2023" "7/15/2023" ...
 $ Stock               : chr  "A" "AAL" "AAP" "AAPL" ...
 $ Description          : chr  "Agilent Technologies, Inc." "American Airl..."
```

```

$ Sector                : chr  "Health Technology" "Transportation" "Retail"
$ Industry              : chr  "Medical Specialties" "Airlines" "Specialty
$ Market.Capitalization : num  3.54e+10 1.20e+10 4.19e+09 2.99e+12 2.36e+1
$ Price                 : num  120 18.4 70.5 190.2 133.6 ...
$ X52.Week.Low          : num  113.3 11.7 63.6 124.2 131 ...
$ X52.Week.High         : num  160 19.1 212 194 168 194 116 78.3 328 519.
$ Return.on.Equity..TTM. : num  24.8 NA 14.6 146 51.1 389 16 14.8 30.7 33.7
$ Return.on.Assets..TTM. : num  12.7 2.64 3.35 27.6 5.43 2.79 7.82 4.98 14.9
$ Return.on.Invested.Capital..TTM. : num  16.51 5.44 6.17 57.18 9.9 ...
$ Gross.Margin..TTM.    : num  54.1 21.7 43.8 43.2 72.2 ...
$ Operating.Margin..TTM. : num  23.78 7.3 5.63 29.16 41.07 ...
$ Net.Margin..TTM.      : num  19.19 3.39 3.61 24.49 13.3 ...
$ Price.to.Earnings.Ratio..TTM. : num  26.45 7.23 10.48 32.32 31.46 ...
$ Price.to.Book..FY.    : num  6.68 NA 1.56 59.85 13.7 ...
$ Enterprise.Value.EBITDA..TTM. : num  18.8 7.08 8.85 24.5 9.24 12.8 19.1 NA 17.4
$ EBITDA..TTM.          : num  1.97e+09 6.02e+09 9.21e+08 1.24e+11 3.18e+1
$ EPS.Diluted..TTM.     : num  4.54 2.54 6.72 5.89 4.25 ...
$ EBITDA..TTM.YoY.Growth. : num  10.52 NA -16 -5.36 10.6 ...
$ EBITDA..Quarterly.YoY.Growth. : num  8.2 NA -39.01 -4.58 11.68 ...
$ EPS.Diluted..TTM.YoY.Growth. : num  9.17 NA -25.21 -4.33 -39.11 ...
$ EPS.Diluted..Quarterly.YoY.Growth. : num  11.69944 NA -68.3683 -0.00656 -94.8904 ...
$ Price.to.Free.Cash.Flow..TTM. : num  30.13 7.03 NA 30.92 10.09 ...
$ Free.Cash.Flow..TTM.YoY.Growth. : num  11.81 NA -100.23 -7.85 6.68 ...
$ Free.Cash.Flow..Quarterly.YoY.Growth. : num  55.7078 648.1481 -176.135 -0.0312 -15.3392
$ Debt.to.Equity.Ratio..MRQ. : num  0.473 NA 1.582 1.763 4.678 ...
$ Current.Ratio..MRQ.    : num  2.37 0.718 1.244 0.94 0.96 ...
$ Quick.Ratio..MRQ.      : num  1.708 0.624 0.238 0.878 0.821 ...
$ Dividend.Yield.Forward : num  0.75 NA 1.413 0.506 4.386 ...
$ Dividends.per.share..Annual.YoY.Growth. : num  8.25 NA 84.62 5.88 7.53 ...
$ Price.to.Sales..FY.    : num  5.257 0.246 0.383 7.875 4.091 ...
$ Revenue..TTM.YoY.Growth. : num  7.8597 50.2948 1.4153 -0.2544 0.0282 ...
$ Revenue..Quarterly.YoY.Growth. : num  6.85 36.97 1.29 -2.51 -9.7 ...
$ Technical.Rating       : chr  "Neutral" "Sell" "Buy" "Sell" ...

```

The `str(sp500)` output provides valuable insights into the structure and data types of the columns in the `sp500` tibble. Let's delve into the details.

The output reveals that `sp500` is a tibble with dimensions $[503 \times 36]$. This means it consists of 503 rows, each representing a specific S&P500 stock, and 36 columns containing information about each stock.

Here's a breakdown of the information associated with each column:

- 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`, `X52.Week.Low`, `X52.Week.High`, and other numeric columns contain diverse financial metrics and stock prices related to the S&P500 stocks.
- The column labeled `Technical.Rating` is a character column that assigns a technical rating to each stock.

By examining the `str(sp500)` output, you gain a comprehensive understanding of the data types and column names present in the `sp500` tibble, enabling you to grasp the structure of the dataset effectively.

Rename Data Columns

```
library(dplyr)

# Define a mapping of new column names
new_names <- c(
  "Date", "Stock", "Desc", "Sector", "Industry",
  "MarketCap", "Price", "Low52Wk", "High52Wk",
  "ROE", "ROA", "ROIC", "GrossMargin",
  "OperatingMargin", "NetMargin", "PE",
  "PB", "EVEBITDA", "EBITDA", "EPS",
  "EBITDA_YOY", "EBITDA_QYOY", "EPS_YOY",
  "EPS_QYOY", "PFCF", "FCF",
  "FCF_QYOY", "DebtToEquity", "CurrentRatio",
  "QuickRatio", "DividendYield",
  "DividendsPerShare_YOY", "PS",
  "Revenue_YOY", "Revenue_QYOY", "TechRating"
)

# Rename the columns using the new_names vector
sp500 <- sp500 %>%
  rename_with(~ new_names, everything())
```

This code is designed to rename the columns of the `sp500` tibble using a predefined mapping of new column names. Let's go through the code step by step:

1. A vector named `new_names` is created, which contains the desired new names for each column in the `sp500` tibble. Each element in the `new_names` vector corresponds to a specific column in the `sp500` tibble and represents the desired new name for that column.
2. The `%>%` operator, often referred to as the pipe operator, is used to pass the `sp500` tibble to the subsequent operation in a more readable and concise manner.
3. The `rename_with()` function from the `dplyr` package is applied to the `sp500` tibble. This function allows you to rename columns based on a specified function or formula.
4. In this case, a formula `~ new_names` is used as the first argument of `rename_with()`. This formula indicates that the new names for the columns should be sourced from the `new_names` vector.
5. The second argument, `everything()`, specifies that the renaming should be applied to all columns in the `sp500` tibble.
6. Finally, the resulting tibble with the renamed columns is assigned back to the `sp500` variable, effectively updating the tibble with the new column names.

In essence, the code uses the `new_names` vector as a mapping to assign new column names to the `sp500` tibble, ensuring that each column is given the desired new name specified in `new_names`.

Review Data Columns

Data Description

```
# A tibble: 7 x 2
  Name      Description
  <chr>     <chr>
1 Date      Date of the observation
2 Stock     Stock ticker symbol
3 Desc      Description of the stock
4 Sector     Sector the stock belongs to
5 Industry  Industry the stock belongs to
6 MarketCap Market capitalization of the company
7 Price     Stock price
```



```
# A tibble: 3 x 2
  Name      Description
  <chr>     <chr>
1 Low52Wk   52-week low price
2 High52Wk  52-week high price
3 TechRating Technical Rating
```

A tibble: 6 x 2

| | Name | Description |
|---|-----------------|----------------------------|
| | <chr> | <chr> |
| 1 | ROE | Return on Equity |
| 2 | ROA | Return on Assets |
| 3 | ROIC | Return on Invested Capital |
| 4 | GrossMargin | Gross Margin |
| 5 | OperatingMargin | Operating Margin |
| 6 | NetMargin | Net Margin |

A tibble: 9 x 2

| | Name | Description |
|---|-------------|--|
| | <chr> | <chr> |
| 1 | PE | Price-to-Earnings Ratio |
| 2 | PB | Price-to-Book Ratio |
| 3 | EVEBITDA | Enterprise Value to EBITDA |
| 4 | EBITDA | EBITDA |
| 5 | EPS | Earnings per Share |
| 6 | EBITDA_YOY | EBITDA Year-over-Year Growth |
| 7 | EBITDA_QYOY | EBITDA Quarterly Year-over-Year Growth |
| 8 | EPS_YOY | EPS Year-over-Year Growth |
| 9 | EPS_QYOY | EPS Quarterly Year-over-Year Growth |

A tibble: 3 x 2

| | Name | Description |
|---|----------|--|
| | <chr> | <chr> |
| 1 | PFCF | Price-to-Free Cash Flow |
| 2 | FCF | Free Cash Flow |
| 3 | FCF_QYOY | Free Cash Flow Quarterly Year-over-Year Growth |

A tibble: 3 x 2

| | Name | Description |
|---|--------------|----------------------|
| | <chr> | <chr> |
| 1 | DebtToEquity | Debt-to-Equity Ratio |
| 2 | CurrentRatio | Current Ratio |
| 3 | QuickRatio | Quick Ratio |

A tibble: 2 x 2

| | Name | Description |
|---|-----------------------|--|
| | <chr> | <chr> |
| 1 | DividendYield | Dividend Yield |
| 2 | DividendsPerShare_YOY | Annual Dividends per Share Year-over-Year Growth |

```
# A tibble: 3 x 2
  Name      Description
  <chr>      <chr>
1 PS        Price-to-Sales Ratio
2 Revenue_YOY Revenue Year-over-Year Growth
3 Revenue_QYOY Revenue Quarterly Year-over-Year Growth
```