**Codebook for Attrition Dataset**

# Data Overview

## Credentials

The direct link to data is:

1. S&P 500 data – <https://finance.yahoo.com/quote/SPY?p=SPY>

2. TVIX data - <https://finance.yahoo.com/quote/TVIX?p=TVIX>

3. Bitcoin data - <https://www.kaggle.com/johnnyjai/bitcoin/data>

4. S&P 500 companies historical prices- <https://www.kaggle.com/dgawlik/nyse>

## Business goal

This data was collected to answer the following questions:

1. Find some rules- when to open position short/long to make yield in high probability.
2. Check a buying strategy when a stock crosses the peak, on historical data.
3. Predict the price of TVIX and Bitcoin with LSTM (Long Short Term Memory network).
4. Choose the stocks from S&P500 with the highest probability of LSTM, and invest them.

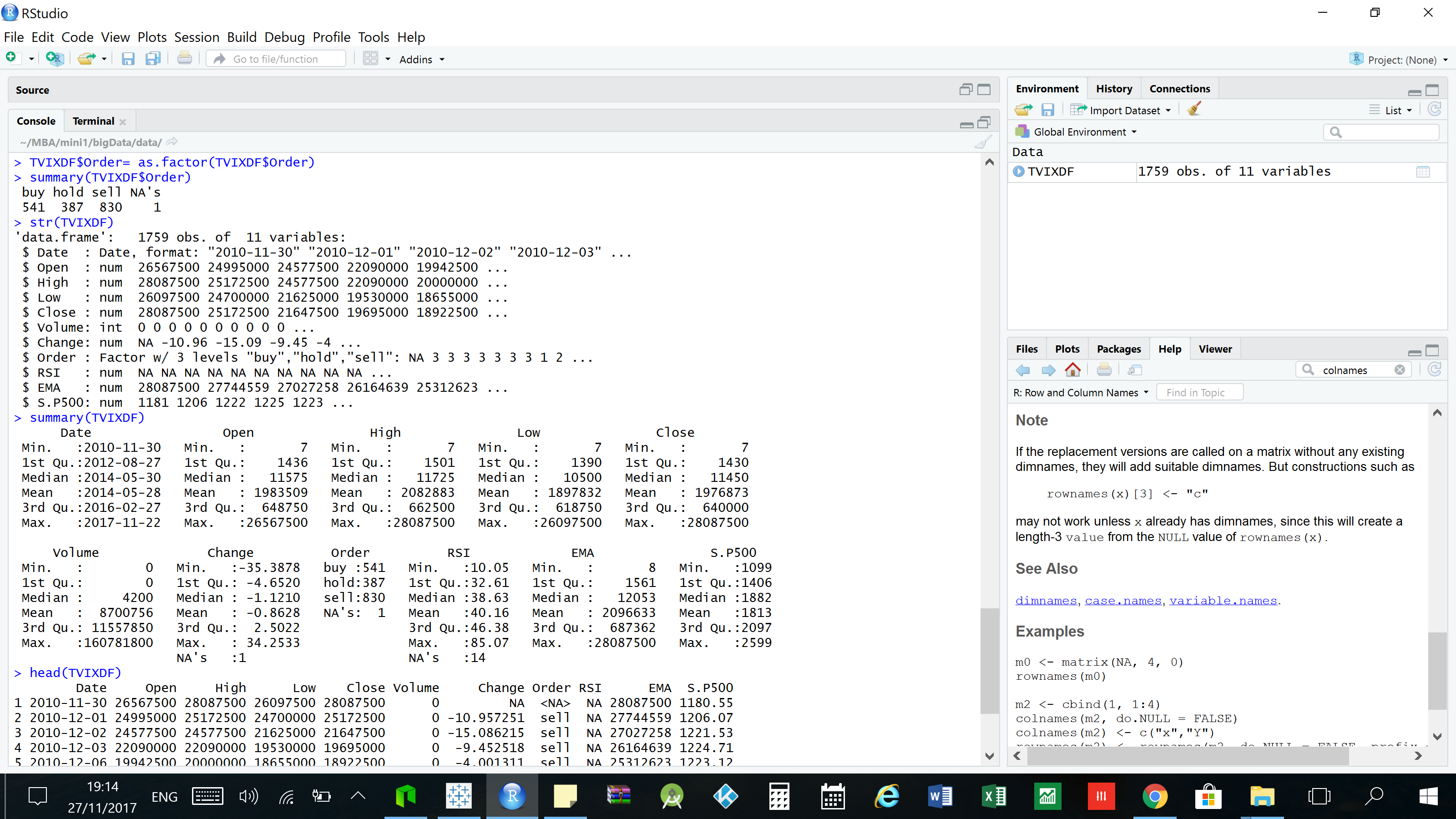
## Data description

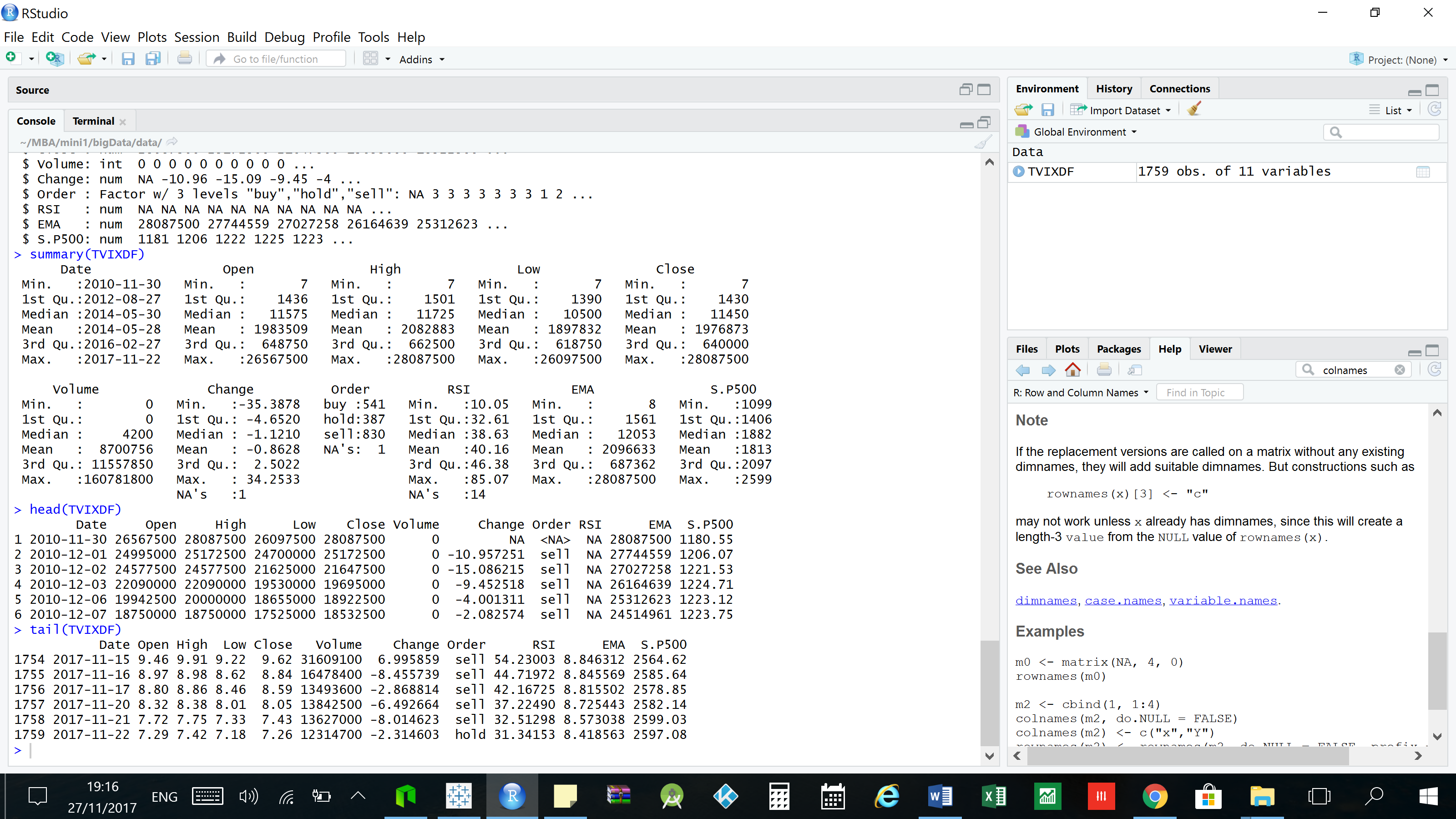
**TVIX** -This data set is a data frame of 10 variables over ~1800 rows. Each row represents an trading date that traded VIX, S&P 500.

**AAPL\_DATA**- This data is a data frame that extract from S&P500 companies, 10 variables over ~1750 rows.

# Variables description

|  |  |  |  |
| --- | --- | --- | --- |
| Variable Name | Description | Type | Possible values |
| Date | Trading Date (format- dd/mm/YY) | Date | 25/11/2015 |
| Symbol | The name of the company | String | "AAPL" |
| Open | The open price at this trading date | Numeric | Positive, USD |
| High | The highest price reached in this trading date | Numeric | Positive, USD |
| Low | The lowest price reached in this trading date | Numeric | Positive, USD |
| Close | The close price at this trading date | Numeric | Positive, USD |
| Volume | The number of total shares that are traded on a given day | Numeric | Positive, USD |
| Change | The percent number of change in the price relative to the previous day price | Numeric | Positive/ Negative percent |
| Open Change | The percent number of change in the previous day close price relative to the open day price | Numeric | Positive/ Negative percent |
| Change within | The percent number of change in the open day price relative to the close day price | Numeric | Positive/ Negative percent |
| nextday\_open\_change | The percent number of change in the close price relative to the next open day price | Numeric | Positive/ Negative percent |
| nextday\_close\_change | The percent number of change in the close price relative to the close next day price | Numeric | Positive/ Negative percent |
| Market Cup | Total equity market value | Numeric | Positive, USD |
| RSI | The relative strength index (RSI) is a momentum indicator that compares the magnitude of recent gains and losses over a specified trading date to measure speed and change of price movements of a security | Numeric | Positive/ Negative percent |
| EMA | Indicator in [technical analysis](https://www.investopedia.com/terms/t/technicalanalysis.asp) that helps smooth out [price action](https://www.investopedia.com/terms/p/price-action.asp) by filtering out the “noise” from random price fluctuations. The two basic and commonly used MAs are the [simple moving average (SMA),](https://www.investopedia.com/terms/s/sma.asp) which is the simple average of a security over a defined number of time periods, and the [exponential moving average (EMA),](https://www.investopedia.com/terms/e/ema.asp) which gives bigger weight to more recent prices | Numeric | Positive, USD |
| S&P 500 close | The close price of S&P 500 index at this trading date | Numeric | Positive, USD |
| order | The right order in every day depending on the price of tomorrow | factor | "buy","sell","hold" |
| High\_close | True if high price equal to close price | logical | "0", "1" |





# Who needs to review the business question?

Investment fund and stock traders

# Related links:

<http://ieeexplore.ieee.org/abstract/document/7364089/>