Business report : SP500 - Company Segmentation

Ralph D. Tasing

Problem Summary

My organization wants to know which companies are similar to each other to help in identifying potential customers of a SAAS software solution (e.g. Salesforce CRM or equivalent) in various segments of the market. The Sales Department is very interested in this analysis, which will help them more easily penetrate various market segments.

Solution Summary

The Analytics Department developed two unsupervised algorithm to classify companies based on how their stocks trade using their daily stock returns (percentage movement from one day to the next). This analysis will deliver value to the stakeholders to determine which companies are related to each other (competitors and have similar attributes).

$$return_{daily} = \frac{price_i - price_{i-1}}{price_{i-1}}$$

Stock prices Analysis

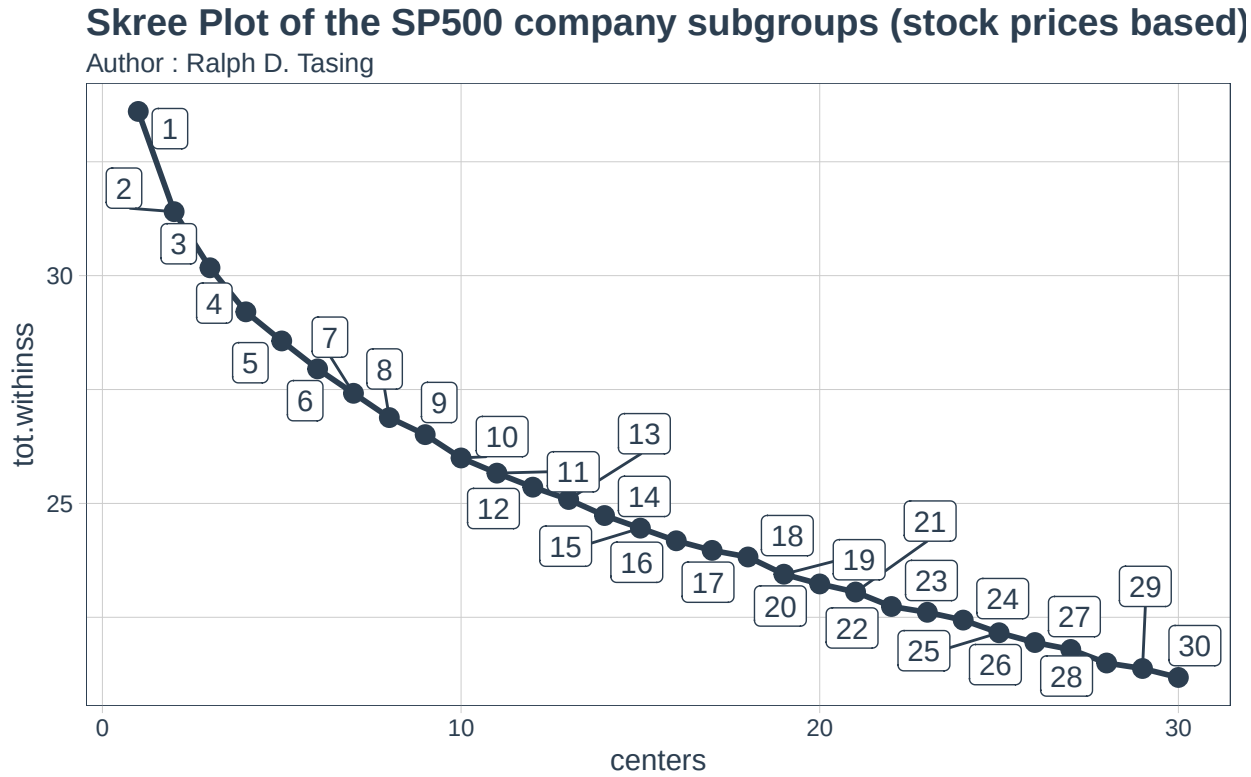
We have stock prices for every stock in the SP 500 Index, which is the daily stock prices for over 500 stocks. The data set is over 1.2M observations.

```
## # A tibble: 1,225,765 x 8
##    symbol date        open  high   low close   volume adjusted
##    <chr>  <date>     <dbl> <dbl> <dbl> <dbl>    <dbl>    <dbl>
##  1 MSFT   2009-01-02  19.5  20.4  19.4  20.3 50084000     15.9
##  2 MSFT   2009-01-05  20.2  20.7  20.1  20.5 61475200     16.0
##  3 MSFT   2009-01-06  20.8  21    20.6  20.8 58083400     16.2
##  4 MSFT   2009-01-07  20.2  20.3  19.5  19.5 72709900     15.2
##  5 MSFT   2009-01-08  19.6  20.2  19.5  20.1 70255400     15.7
##  6 MSFT   2009-01-09  20.2  20.3  19.4  19.5 49815300     15.2
##  7 MSFT   2009-01-12  19.7  19.8  19.3  19.5 52163500     15.2
##  8 MSFT   2009-01-13  19.5  20.0  19.5  19.8 65843500     15.5
##  9 MSFT   2009-01-14  19.5  19.7  19.0  19.1 80257500     14.9
## 10 MSFT   2009-01-15  19.1  19.3  18.5  19.2 96169800     15.0
## # ... with 1,225,755 more rows
```

Daily returns

```
## # A tibble: 141,340 x 3
##    symbol date       pct_return
##    <chr>  <date>          <dbl>
##  1 MSFT   2018-01-03   0.00465
##  2 MSFT   2018-01-04   0.00880
##  3 MSFT   2018-01-05   0.0124
##  4 MSFT   2018-01-08   0.00102
##  5 MSFT   2018-01-09  -0.000680
##  6 MSFT   2018-01-10  -0.00453
##  7 MSFT   2018-01-11   0.00296
##  8 MSFT   2018-01-12   0.0173
##  9 MSFT   2018-01-16  -0.0140
## 10 MSFT   2018-01-17   0.0203
## # ... with 141,330 more rows
```

Skree plot : Optimum number of segments



Conclusion: Based on the skree plot, 10 segments is enough. Behond 10, we don't have too much change

SP500 companies Segmentation : 2D Projection