Moneytalks - Report

Adriano Augusto - Grace Achenyo Okolo

STEP3 - Data analysis Vol.1

```
library(data.table)
library(dplyr)
library(tidyr)
library(ggplot2)
library(pdist)
library(caret)
data <- read.csv("dataset.csv")
```

We decided to check whether there are months where most of the times a stock price closes in positive or negative. To do so, we counted the positive days (i.e. daily change > 0.5), and the negative days (i.e. daily change < -0.5), in each month. Then, we evaluated the difference, and we did group the counting by month and year and plot them. The values are normalized according to the number of stocks composing the NASDAQ100 (107).

Simultaneously, we did the same for the stock opening price. To do so, we counted the positive openings (i.e. opening change positive), and the negative openings (i.e. opening change negative), in each month. Then, we evaluated the difference, and we did group the counting by month and year and plot them. The values are normalized according to the number of stocks composing the NASDAQ100 (107).

```
dcount_pos <- function(df, value) {
  summarise(filter(df, d_change>value), n())
}

dcount_neg <- function(df, value) {
  summarise(filter(df, d_change<value), n())
}

ocount_pos <- function(df, value) {
  summarise(filter(df, o_change>value), n())
}

ocount_neg <- function(df, value) {
  summarise(filter(df, o_change<value), n())
}
```

```
years <- c(2013, 2014, 2015, 2016, 2017)
months <- c(1, 2, 3, 4, 5,6, 7, 8, 9, 10, 11, 12)
```

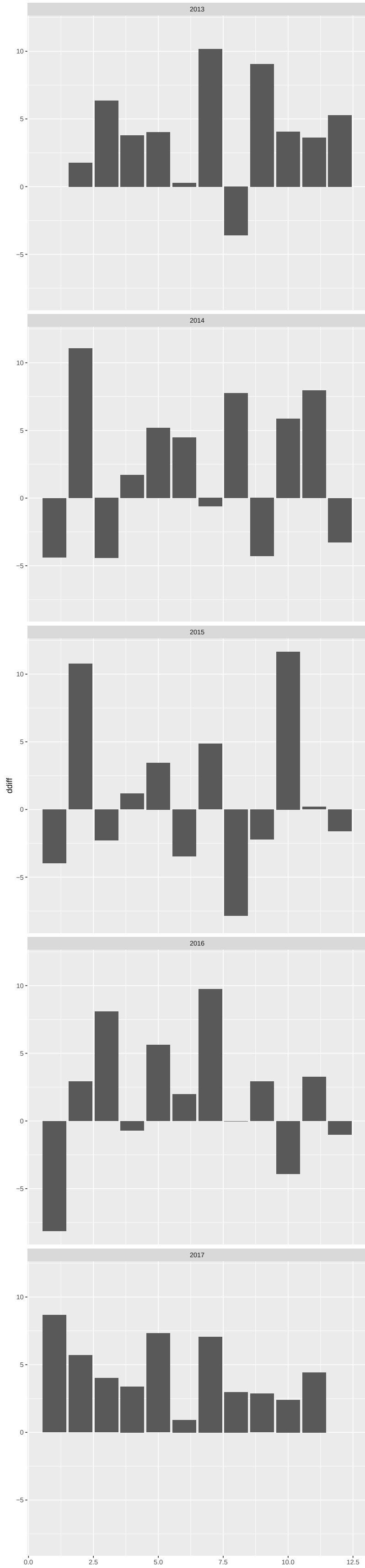
```
iterations = 60
features = 6

changes <- matrix(ncol=features, nrow=iterations)

c <- 0
 for(y in years) {
  for(m in months) {
```

```
   monthly_data  <- filter(data, year == y & month == m)

   dmonthly_tot_pos <- dcount_pos(monthly_data, 0.5)[1,]
   dmonthly_tot_neg <- dcount_neg(monthly_data, -0.5)[1,]

   omonthly_tot_pos <- ocount_pos(monthly_data, 0.5)[1,]
   omonthly_tot_neg <- ocount_neg(monthly_data, -0.5)[1,]

   changes[c,1] <- y
   changes[c,2] <- m
   changes[c,3] <- dmonthly_tot_pos/107
   changes[c,4] <- dmonthly_tot_neg/107
   changes[c,5] <- omonthly_tot_pos/107
   changes[c,6] <- omonthly_tot_neg/107
   c = c+1
  }
 }

changes <- data.frame(changes)
changes <- rename(changes, year = X1, month = X2, dtot_pos = X3,
                  dtot_neg = X4, otot_pos = X5, otot_neg = X6)
changes <- filter(changes, dtot_pos != 0 | dtot_neg != 0)
```

```
changes <- transform(changes, ddiff= dtot_pos-dtot_neg)
changes <- transform(changes, odiff= otot_pos-otot_neg)
```

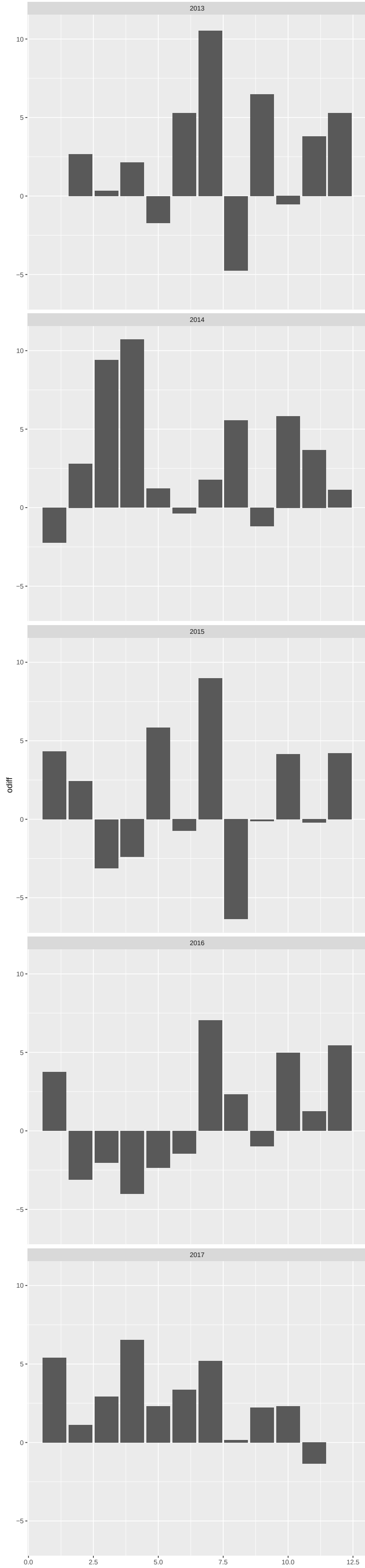
```
#summary(changes$dtot_pos)
#summary(changes$dtot_neg)
#summary(changes$otot_pos)
#summary(changes$otot_neg)
```

```
ggplot(changes, aes(x=month, y=ddiff)) +
  geom_bar(stat = "identity") +
  facet_wrap(~ year, ncol = 1)
```


2013
2014
2015
2016
2017
ddiff
month
10
5
0
−5
0.0
2.5
5.0
7.5
10.0
12.5

This first series of plots show how many days the stock price closed in positive (>0.5) for a specific month, w.r.t. the number of days the stock price closed in negative (<-0.5). A positive value X means that, for that month, there were X-days positive more than the negative days. What we can see from these plots is that, overall, there are more positive days than negative. Also, we can see that months having many positive days, are eventually followed by month with more negative days, and vice-versa. This entails that the stock prices movements follow a cycle that repeats itself. An example is the plot for year 2017, where first half and second half shows the same pattern. Also, the month of February was the only one able to have across the all 5 years an excess of positive days over the negative ones. Although this may be just casual, it may also be related to the fact that at the end of January, companies report the financial results of their first quarter financial year.

```
ggplot(changes, aes(x=month, y=odiff)) +
  geom_bar(stat = "identity") +
  facet_wrap(~ year, ncol = 1)
```


2013
2014
2015
2016
2017
10
5
0
−5
odiff
0.0
2.5
5.0
7.5
10.0
12.5
month

This second series of plots report a similar analysis of the previous, however this time we focused on the opening price change instead of the closing price. The results show a very interesting fact, when compared with the results of the previous plots. Indeed, the number of positive days reduce overall. A clear example is the year 2016, first semester. Considering the closing price, there were more positive days than negative. Instead, when considering the opening price, there were more negative days than positive. This means that most of the times for those months a stock price opening with a negative change, turned the negative change into positive over the day. This is a first step toward the analysis of the relation between the opening price change and the closing price change. From this results, we can see that if there is a correlation it may be negative.

STEP4 - Data analysis Vol.2

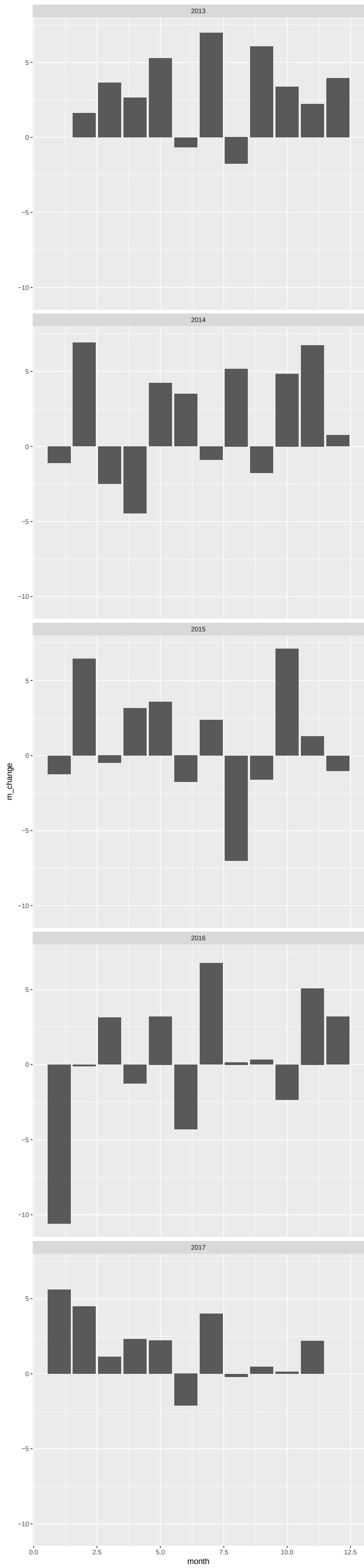
Successively, we decided to analyse what is the average monthly return for the NASDAQ100. First, we considered the monthly return of each month per year per each listed company in the NASDAQ100, and we evaluated the average. Then, we plot such average and we analysed the results, as well the summary of the values.

```
iterations = 60
features = 3

changes <- matrix(ncol=features, nrow=iterations)

c <- 0
 for(y in years) {
  for(m in months) {
   monthly_data  <- filter(data, year == y & month == m)

   i <- which.max(monthly_data$day)
   if(length(monthly_data[i,'m_change']) == 0) break

   d <- monthly_data[i,'day']

   monthly_data <- filter(monthly_data, day == d)

   changes[c,1] <- y
   changes[c,2] <- m
   changes[c,3] <- as.numeric(summarise(monthly_data, avg_mchange = mean(m_change)))
   c <-  c+1
  }
 }

changes <- data.frame(changes)
changes <- rename(changes, year = X1, month = X2, m_change = X3)
changes <- filter(changes, m_change != 0 )
```

```
summary(changes$m_change)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -10.5870  -0.9833   2.2114   1.5380   4.0028   7.1289
```

```
ggplot(changes, aes(x=month, y=m_change)) +
  geom_bar(stat = "identity") +
  facet_wrap(~ year, ncol = 1)
```


2013
2014
2015
2016
2017
m_change
month
5
0
−5
−10
0.0
2.5
5.0
7.5
10.0
12.5

The summary shows that the average return over the last 5 years has been 1.5% monthly. Despite some outliers (e.g. January 2016), we can say that across the last 5 years the NASDAQ100 had positive returns the majority of the month. Also, positive returns are often followed by negative returns.

STEP4 - Advices for investors

From the findings we got so far we designed four questions about stock market and stock prices and we did answer. The answers are backed by the data analysis did so far. The suggestions, despite simple, are fundamental for any investor.

1. Should I invest? Yes, everyone should invest in the stock market, i.e. everyone should be an investor. This statement is backed by the fact that the company listed in the NASDAQ100 showed to be solid and have positive returns across the 5 years, with an average of 1.5% monthly. Such return is way greater than any saving account offered by any bank in the world.
2. Short or long time investments? Buy and hold strategies are the most valuable, since the stocks' prices tend to grow over long time. The data exploration showed that the longer is the time-frame, the greater is the growth of the price stock. Despite there is no hint about how the prices change daily or weekly, it was clear that on a long run, the price tent to go up. This is of course strictly related to the fact that the company selected from the NASDAQ100 index are stable and solid, with strong financial setting.
3. When to buy? We saw in different graphs that despite it is not possible to forecast the magnitude of the price change, we can see patterns of the type "up-down-up-down". In other words, the best moment to buy is within negative periods. This does not mean that there is a warranty of the stock price to go up again quickly, however, the chances are higher that it could happen.
4. When to sell? The later the better. This question and its answer are strictly related to the type of investment strategy. Although the best rational strategy is to buy and hold a stock, it is clear that having knowledge of when a price could go up and when down, the returns would be much more profitable. However, the analysis of the data did so far showed that there is no way to forecast price movements, nor direction nor magnitude. Indeed, no specific patterns were found in the prices' movements.

Driving by the fact that having the change to forecast the near price movements would highly increase the profit of investments, we will try to analyse if any correlation holds between the price change at the opening, an the price change at the closure in terms of percetage.

STEP5 - Relation between Opening and Closing Price Change (%)

We first select the two column for opening change and closing change in terms of percentage, and we evaluate the correlation.

```r
o <- data$o_change
d <- data$d_change

od_corr <- cor(o, d, use = "everything", method = c("pearson", "kendall", "spearman"))
od_corr
```

```
## [1] 0.6645693
```

The correlation between the two features is 0.66. This indicates there is crearly a positive correlation, that makes sense to investigate. Therefore, we evaluate the average distance between the opening and the closing change in %.

```
sum <- 0

for(i in 1:length(o)) {
  sum <- sum + sqrt((o[i] - d[i])^2)
}

avgd <- sum/length(o)
avgd
```

```
## [1] 1.056221
```

The average distance between the two value is 1.06. This is meaningful, since it shows that on average, the opening price gives and hint of the magnitude of the change at closing.

To further analyse this data, we check how many times the opening change was greater than the closing and viceversa.

```
growing <- 0
declining <- 0

for(i in 1:length(o)) {
  if( d[i] > o[i] ) growing <- growing + 1
  else declining <- declining + 1
}

growing
```

```
## [1] 200793
```

```
declining
```

```
## [1] 188718
```

These results are not interesting, since the change of having the closing price greater than the opening (and vice-versa) is 50%. We refine the concept of greater and lower, assuming that as long as the closing price is within +- 0.25 we do not consider it greater nor lower.

```
growing <- 0
declining <- 0
inline <- 0

for(i in 1:length(o)) {
  if( d[i] > (o[i]+0.25) ) growing <- growing + 1
  else if( d[i] < (o[i]-0.25) ) declining <- declining + 1
  else inline <- inline + 1
}

inline
```

```
## [1] 72063
```

```
growing
```

```
## [1] 164562
```

```
declining
```

```
## [1] 152886
```

Also for this case, the results are not interesting, therefore we can say that in a **generic** case, the change the price of a stock closes greater, lower or in-line is roughly 33%, 33% and 33%.

Now, we analyse a more specific setting, that is the case when the change in percentage at the opening is extremely positive or negative. And for this case we analyse how many times the change at the closing is greater, lower, or in line. We then consider the probability of the three cases, and we compute the difference between the probability the closing change in percentage is greater instead of lower.

```r
probs <- matrix(ncol=5, nrow=16)
it <- 0
for(cp in c(1.0, 2.0, 3.0, 4.0)) {

  ohc <- filter(data, o_change > cp)
  o <- ohc$o_change
  d <- ohc$d_change

  growing <- 0
  declining <- 0
  inline <- 0

  for(r in c(0.5, 1.0, 1.5, 2.0) ){

    for(i in 1:length(o)) {
      if( d[i] > (o[i]+r) ) growing <- growing + 1
      else if( d[i] < (o[i]-r) ) declining <- declining + 1
      else inline <- inline + 1
    }

    it <- it+1
    tot <- inline + growing + declining
    probs[it,1] <- cp
    probs[it,2] <- r
    probs[it,3] <- inline/tot
    probs[it,4] <- growing/tot
    probs[it,5] <- declining/tot
  }

}

probs <- data.frame(probs)
probs <- rename(probs, cp = X1, r = X2, inline = X3,
                growing = X4, declining = X5)
probs <- mutate(probs, gdiff=growing-declining)
probs
```

```
##    cp   r    inline   growing declining       gdiff
## 1   1 0.5 0.2266222 0.3812547 0.3921231 -0.01086840
## 2   1 1.0 0.3316475 0.3286883 0.3396643 -0.01097600
## 3   1 1.5 0.4217153 0.2836544 0.2946304 -0.01097600
## 4   1 2.0 0.4968525 0.2462875 0.2568600 -0.01057247
## 5   2 0.5 0.1578535 0.3963964 0.4457501 -0.04935370
## 6   2 1.0 0.2316882 0.3615354 0.4067763 -0.04524089
## 7   2 1.5 0.3030422 0.3273273 0.3696305 -0.04230317
## 8   2 2.0 0.3702507 0.2945554 0.3351939 -0.04063846
## 9   3 0.5 0.1331195 0.4001604 0.4667201 -0.06655974
```

```
## 10   3 1.0 0.1912590 0.3748998 0.4338412 -0.05894146
## 11   3 1.5 0.2542101 0.3448276 0.4009623 -0.05613472
## 12   3 2.0 0.3165597 0.3143545 0.3690858 -0.05473136
## 13   4 0.5 0.1279683 0.3918206 0.4802111 -0.08839050
## 14   4 1.0 0.1807388 0.3713720 0.4478892 -0.07651715
## 15   4 1.5 0.2387863 0.3469657 0.4142480 -0.06728232
## 16   4 2.0 0.2965040 0.3202507 0.3832454 -0.06299472
```

```r
probs <- matrix(ncol=5, nrow=16)
it <- 0
for(cp in c(1.0, 2.0, 3.0, 4.0)) {

  ohc <- filter(data, o_change < -cp)
  o <- ohc$o_change
  d <- ohc$d_change

  growing <- 0
  declining <- 0
  inline <- 0

  for(r in c(0.5, 1.0, 1.5, 2.0) ){

    for(i in 1:length(o)) {
      if( d[i] > (o[i]+r) ) growing <- growing + 1
      else if( d[i] < (o[i]-r) ) declining <- declining + 1
      else inline <- inline + 1
    }

    it <- it+1
    tot <- inline + growing + declining
    probs[it,1] <- -cp
    probs[it,2] <- r
    probs[it,3] <- inline/tot
    probs[it,4] <- growing/tot
    probs[it,5] <- declining/tot
  }

}

probs <- data.frame(probs)
probs <- rename(probs, cp = X1, r = X2, inline = X3,
                growing = X4, declining = X5)
probs <- mutate(probs, gdiff=growing-declining)
probs
```

```
##    cp   r    inline   growing declining      gdiff
## 1  -1 0.5 0.2333211 0.4510306 0.3156482 0.13538236
## 2  -1 1.0 0.3353458 0.3923039 0.2723503 0.11995365
## 3  -1 1.5 0.4235069 0.3402854 0.2362077 0.10407773
## 4  -1 2.0 0.4978351 0.2974448 0.2047201 0.09272472
## 5  -2 0.5 0.1829876 0.5045643 0.3124481 0.19211618
## 6  -2 1.0 0.2618257 0.4607884 0.2773859 0.18340249
## 7  -2 1.5 0.3365145 0.4175657 0.2459198 0.17164592
## 8  -2 2.0 0.4045643 0.3775934 0.2178423 0.15975104
## 9  -3 0.5 0.1589922 0.5247611 0.3162467 0.20851434
```

```
## 10 -3 1.0 0.2311034 0.4847958 0.2841008 0.20069505
## 11 -3 1.5 0.2959745 0.4471474 0.2568781 0.19026933
## 12 -3 2.0 0.3568636 0.4120330 0.2311034 0.18092963
## 13 -4 0.5 0.1411936 0.5312955 0.3275109 0.20378457
## 14 -4 1.0 0.2059680 0.4927220 0.3013100 0.19141194
## 15 -4 1.5 0.2644347 0.4594857 0.2760796 0.18340611
## 16 -4 2.0 0.3220524 0.4268559 0.2510917 0.17576419
```

At this point, we can summarize the results of the analysis between opening change and closing change. The results are very interesting, and can be summarized with the following statements.

1. although there is a relevant correlation between opening change and closing change, forecasting the direction of the price movement at the closing, taking into account only the change at the opening is almost impossible. Indeed, we showed the chances the price move positively, or negatively or stays in line are very similar.
2. however, when the change at the opening is very positive (+1, +2, +3, +4), the change at the closing can either be greater or lower (with same magnitude, e.g. +-0.5, +-1.0, etc.) with very similar chance. Though, it is more probable that it will be lower than greater. This means, investor should think twice before buying on a day with an opening change in price over 1%.
3. on the other hand, when the change at the opening is very negative (-1, -2, -3, -4), the chances the price will grow by the end of the day are relevant, ranging from a minimum of 9% to a maximum of 21%!

Conclusion:

In light of the new findings, we would draw another suggestion for the investors.

1. Buy at the opening of bad days. You have a chance up to 20% to see you price growing by the end of the day.
2. Stock price movements seem to be impossible to predict. Although this can be a disappointing result, it turns out that the price movements seem to stick around an equilibrium that can be summarized with se simple sentence: "what goes up, will eventually go down and viceversa".

Overall, the analysis of the relation between opening change and closing change of stock prices was useful, although it was limited to the only analysis of the two values. In the next step, we will try to train a classifier that will take into account also weekly and monthly change of a stock price, to draw advices of the type BUY and SELL between opening and closing price.