

Czy warto kupić akcje spółki, które w ciągu dnia straciły na wartości ponad osiem procent? Przedmiotem analizy są spółki notowane na Warszawskiej Giełdzie Papierów Wartościowych.

Katarzyna Zaleska

Marcin Łojek

Wybrany zbiór danych **Warsaw Stock Exchange - On-Balance Volume**. Domyślnie zbiór danych zawiera informacje o dziennych cenach 1622 różnych instrumentów finansowych notowanych na GPW.

```
library('lubridate')
```

```
##  
## Dołączanie pakietu: 'lubridate'  
  
## Następujące obiekty zostały zakryte z 'package:base':  
##  
## date, intersect, setdiff, union
```

Początkowo odczytano dane pobrane bezpośrednio z platformy Kaggle.

```
# read stock data from csv file  
stock_data <- read.csv('stock_data.csv', header=TRUE, dec='.')
```

Wybrany zbiór składa się z 8 kolumn:

- **INSTRUMENT** - symbol giełdowy akcji/instrumentu
- **TRADING\_\_DATE** - data notowania akcji
- **OPEN\_\_VAL\_\_PLN** - cena akcji na otwarciu rynku
- **CLOSE\_\_VAL\_\_PLN** - cena akcji na zamknięciu rynku
- **HIGH\_\_VAL\_\_PLN** - najwyższa cena akcji w ciągu danego dnia
- **LOW\_\_VAL\_\_PLN** - najniższa cena akcji w ciągu danego dnia
- **VOLUME** - liczba akcji sprzedanych w danym dniu
- **OP\_\_OBV** - wartość wskaźnika On-Balance Volume

```
head(stock_data)
```

```
##   X INSTRUMENT TRADING_DATE OPEN__VAL__PLN CLOSE__VAL__PLN HIGH__VAL__PLN LOW__VAL__PLN  
## 1 0      11BIT   2019/01/18       276.0         275.0         279.5         273.5  
## 2 1      11BIT   2019/01/17       275.5         274.5         282.0         270.0
```

##	3	2	11BIT	2019/01/16	264.0	274.0	274.5	263.5
##	4	3	11BIT	2019/01/15	262.5	263.5	266.0	260.0
##	5	4	11BIT	2019/01/14	265.5	258.0	265.5	258.0
##	6	5	11BIT	2019/01/11	261.5	265.5	269.5	261.5
##			VOLUME	OP_OBV				
##	1		4034	2844835				
##	2		7531	2840801				
##	3		19050	2833270				
##	4		4852	2814220				
##	5		3723	2809368				
##	6		6345	2813091				

Powyższy zbiór został następnie przetworzony. Wykonano następujące kroki:

- **Wybranie przedziału lat 2010-2018** (pierwotny zbiór obejmuje okres od stycznia 1991 roku do stycznia 2019 roku)

```
# get only values from years 2010-2018
stock_data <- stock_data[
  year(stock_data$TRADING_DATE) >= 2010 &
  year(stock_data$TRADING_DATE) <= 2018, ]
```

- Usunięcie nieużywanych dalej kolumn

```
# remove unnecessary columns
stock_data$X <- NULL
stock_data$OPEN_VAL_PLN <- NULL
stock_data$LOW_VAL_PLN <- NULL
stock_data$VOLUME <- NULL
stock_data$OP_OBV <- NULL
```

- **Ograniczenie analizy do spółek:** BUDIMEX, CDPROJEKT, PKOBP, CYFRPLSAT, KGHM, PKNORLEN, PGE, PEKAO, ORANGEPL, MBANK, SANPL, KETY, LPP, PZU, KRUK, JSW, ALIOR, DINOPL.

```
# list of tickers we're gonna analyze
wig_tickers <- c(
  'BUDIMEX', 'CDPROJEKT', 'PKOBP', 'CYFRPLSAT', 'KGHM',
  'PKNORLEN', 'PGE', 'PEKAO', 'ORANGEPL', 'MBANK',
  'SANPL', 'KETY', 'LPP', 'PZU', 'KRUK', 'JSW',
  'ALIOR', 'DINOPL'
)

# filter by tickers we're gonna analyze
stock_data <- stock_data[stock_data$INSTRUMENT %in% wig_tickers, ]

# sort the data frame by instrument and then by date
stock_data <- stock_data[order(stock_data$INSTRUMENT, stock_data$TRADING_DATE), ]
```

- Dodanie kolumny **PERCENT\_CHANGE** oznaczającej procentowy spadek wartości akcji względem poprzedniej sesji

```

# add column with percentage change from day to day
stock_data$PERCENT_CHANGE <- ave(
  stock_data$CLOSE_VAL_PLN,
  stock_data$INSTRUMENT,
  FUN = function(x) c(NA, diff(x) / head(x, -1) * 100)
)
# save to CSV for debug purposes
write.csv(stock_data, "preprocessed_data.csv", row.names = FALSE)

```

- Dodanie kolumn **MAX\_1\_DAY\_AHEAD**, **MAX\_5\_DAYS\_AHEAD** oraz **MAX\_10\_DAYS\_AHEAD** oznaczających maksymalną cenę w ciągu 1/5/10 dni po dniu spadku. Odpowiednio dodano również **PERCENT\_CHANGE\_1\_DAY**, **PERCENT\_CHANGE\_5\_DAYS** oraz **PERCENT\_CHANGE\_10\_DAYS** oznaczające procentową zmianę względem ceny na zamknięciu w dniu spadku

```

# add percent rise of value for 1, 5 and 10 days ahead
stock_data <- do.call(rbind, lapply(
  split(stock_data, stock_data$INSTRUMENT),
  function(group) {
    n <- nrow(group)

    group$MAX_1_DAY_AHEAD <- sapply(
      1:n,
      function(i) max(group$CLOSE_VAL_PLN[i:min(i + 1, n)])
    )

    group$MAX_5_DAYS_AHEAD <- sapply(
      1:n,
      function(i) max(group$CLOSE_VAL_PLN[i:min(i + 5, n)])
    )

    group$MAX_10_DAYS_AHEAD <- sapply(
      1:n,
      function(i) max(group$CLOSE_VAL_PLN[i:min(i + 10, n)])
    )

    group$PERCENT_CHANGE_1_DAY <-
      (group$MAX_1_DAY_AHEAD - group$CLOSE_VAL_PLN) / group$CLOSE_VAL_PLN * 100

    group$PERCENT_CHANGE_5_DAYS <-
      (group$MAX_5_DAYS_AHEAD - group$CLOSE_VAL_PLN) / group$CLOSE_VAL_PLN * 100

    group$PERCENT_CHANGE_10_DAYS <-
      (group$MAX_10_DAYS_AHEAD - group$CLOSE_VAL_PLN) / group$CLOSE_VAL_PLN * 100

    return(group)
  }
))

# save to CSV for debug purposes
write.csv(stock_data, "preprocessed_data_rise.csv", row.names = FALSE)

```

Następnie sprawdzono liczbę wierszy dla każdej z wybranych spółek.

```

instrument_counts <- table(stock_data$INSTRUMENT)
instrument_counts_df <- as.data.frame(instrument_counts)
colnames(instrument_counts_df) <- c("INSTRUMENT", "Count")

print(instrument_counts_df)

```

```

##      INSTRUMENT Count
## 1      ALIOR    1503
## 2     BUDIMEX    2248
## 3    CDPROJEKT    2248
## 4    CYFRPLSAT    2248
## 5      DINOPL     423
## 6       JSW    1868
## 7      KETY    2243
## 8      KGHM    2248
## 9      KRUK    1908
## 10     LPP    2240
## 11     MBANK    2248
## 12   ORANGEPL    2248
## 13     PEKAO    2248
## 14      PGE    2248
## 15   PKNORLEN    2248
## 16     PKOBP    2248
## 17      PZU    2159
## 18     SANPL    2245

```

Po przygotowaniu danych, zidentyfikowaliśmy dni, w których spadek akcji osiąga ponad osiem procent.

```

filtered_by_decrease <- stock_data[stock_data$PERCENT_CHANGE < -8, ]

```

Wykonaliśmy test normalności Shapiro-Wilka odpowiednio dla kolumn MAX\_1\_DAY\_AHEAD, MAX\_5\_DAYS\_AHEAD oraz MAX\_10\_DAYS\_AHEAD.

```

shapiro_1_day <- shapiro.test(filtered_by_decrease$MAX_1_DAY_AHEAD)
print(shapiro_1_day)

```

```

##
##  Shapiro-Wilk normality test
##
## data:  filtered_by_decrease$MAX_1_DAY_AHEAD
## W = 0.22174, p-value < 2.2e-16

```

```

shapiro_5_day <- shapiro.test(filtered_by_decrease$MAX_5_DAYS_AHEAD)
print(shapiro_5_day)

```

```

##
##  Shapiro-Wilk normality test
##
## data:  filtered_by_decrease$MAX_5_DAYS_AHEAD
## W = 0.22296, p-value < 2.2e-16

```

```
shapiro_10_day <- shapiro.test(filtered_by_decrease$MAX_10_DAYS_AHEAD)
print(shapiro_10_day)
```

```
##
## Shapiro-Wilk normality test
##
## data:  filtered_by_decrease$MAX_10_DAYS_AHEAD
## W = 0.22384, p-value < 2.2e-16
```

W każdym przypadku rozkład jest inny niż normalny.

W związku z powyższym wykonano **test Wilcozona**. Dla każdego z wybranych okresów możemy sformułować hipotezę **H0** - **średnia procentowych zmian cen akcji po dniu spadku w wybranym okresie jest mniejsza lub równa zero**. Hipoteza alternatywna zakłada, że średnia cena jest dodatnia.

```
wilcox_result_1_day <- wilcox.test(filtered_by_decrease$PERCENT_CHANGE_1_DAY,
                                   alternative = "greater", mu = 0)
```

```
## Warning in wilcox.test.default(filtered_by_decrease$PERCENT_CHANGE_1_DAY, : nie
## można obliczyć dokładnej wartości prawdopodobieństwa z zerami
```

```
print(wilcox_result_1_day)
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data:  filtered_by_decrease$PERCENT_CHANGE_1_DAY
## V = 741, p-value = 4.029e-08
## alternative hypothesis: true location is greater than 0
```

```
wilcox_result_5_days <- wilcox.test(filtered_by_decrease$PERCENT_CHANGE_5_DAYS,
                                   alternative = "greater", mu = 0)
print(wilcox_result_5_days)
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data:  filtered_by_decrease$PERCENT_CHANGE_5_DAYS
## V = 1953, p-value = 3.883e-12
## alternative hypothesis: true location is greater than 0
```

```
wilcox_result_10_days <- wilcox.test(filtered_by_decrease$PERCENT_CHANGE_10_DAYS,
                                     alternative = "greater", mu = 0)
print(wilcox_result_10_days)
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data:  filtered_by_decrease$PERCENT_CHANGE_10_DAYS
## V = 2211, p-value = 8.396e-13
## alternative hypothesis: true location is greater than 0
```

Na podstawie **testu Wilcoxona** odrzucona została hipoteza zerowa. W związku z tym przyjęto hipotezę alternatywną - w przeciągu wybranego okresu 1/5/10 dni zazwyczaj następuje wzrost cen akcji. Warto kupić akcję, których ceny jednego dnia spadły o ponad osiem procent ponieważ można potem zaobserwować wzrost.