

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.

Wybrany zbiór danych **Warsaw Stock Exchange - On-Balance Volume**.

```
library('lubridate')

##
## Dołączanie pakietu: 'lubridate'

## Następujące obiekty zostały zakryte z 'package:base':
##
##     date, intersect, setdiff, union

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

Powyższy zbiór został przetworzony. Wykonano następujące kroki:

- Wybranie przedziału lat 2010-2018

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

- Usunięto nieużywane kolumny

```
# 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)

```

Ostatecznie przygotowane dane:

```
head(stock_data)
```

```

##          INSTRUMENT TRADING_DATE CLOSE_VAL_PLN HIGH_VAL_PLN PERCENT_CHANGE
## ALIOR.36761      ALIOR  2012/12/14         60.90        61.80             NA
## ALIOR.36760      ALIOR  2012/12/17         61.25        61.50        0.5747126
## ALIOR.36759      ALIOR  2012/12/18         62.25        62.40        1.6326531
## ALIOR.36758      ALIOR  2012/12/19         62.45        62.55        0.3212851
## ALIOR.36757      ALIOR  2012/12/20         62.25        62.60       -0.3202562
## ALIOR.36756      ALIOR  2012/12/21         62.25        62.40        0.0000000
##          MAX_1_DAY_AHEAD MAX_5_DAYS_AHEAD MAX_10_DAYS_AHEAD
## ALIOR.36761             61.25             62.45             65.55
## ALIOR.36760             62.25             63.25             65.90
## ALIOR.36759             62.45             63.25             65.90
## ALIOR.36758             62.45             64.00             65.90
## ALIOR.36757             62.25             64.90             65.90
## ALIOR.36756             63.25             65.55             65.90
##          PERCENT_CHANGE_1_DAY PERCENT_CHANGE_5_DAYS PERCENT_CHANGE_10_DAYS
## ALIOR.36761             0.5747126             2.545156             7.635468
## ALIOR.36760             1.6326531             3.265306             7.591837
## ALIOR.36759             0.3212851             1.606426             5.863454
## ALIOR.36758             0.0000000             2.481986             5.524420
## ALIOR.36757             0.0000000             4.257028             5.863454
## ALIOR.36756             1.6064257             5.301205             5.863454

```

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 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 tym wykonano **test Wilcoxona**. 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 akcje, których ceny jednego dnia spadły o ponad osiem procent.