

# Raport 1

Paweł Matławski  
album 249732

13 marca 2021

## Spis treści

<b>1</b>	<b>Analiza badanego problemu</b>	<b>1</b>
<b>2</b>	<b>Przygotowanie danych. Podstawowe informacje o danych</b>	<b>1</b>
<b>3</b>	<b>Analiza opisowa. Wskaźniki sumaryczne i wykresy.</b>	<b>2</b>
3.1	Wskaźniki sumaryczne . . . . .	2
3.2	Wykresy . . . . .	3
3.3	Interpretacja otrzymanych wyników . . . . .	17
<b>4</b>	<b>Analiza opisowa z podziałem na grupy</b>	<b>17</b>
4.1	Badanie zależności pomiędzy poszczególnymi cechami a przynależnością do jednej z grup . . . . .	17
<b>5</b>	<b>Podsumowanie</b>	<b>27</b>
5.1	Informacje których dowiedzieliśmy się o klientach sieci . . . . .	27
5.2	Główne przyczyny odchodzenia klientów. Proponowane rozwiązania. . . . .	27

## 1 Analiza badanego problemu

- Będziemy analizować migrację klientów w pewnej firmie telefonicznej na podstawie danych.
- Odpowiemy sobie na pytania, jakie cechy zarządzania firmą, jakie oferty i jakie cechy klientów mają największy wpływ na migracje.

## 2 Przygotowanie danych. Podstawowe informacje o danych

- Na początku zbadamy rozmiar danych, typy poszczególnych cech, a więc czy cecha jest typu numerycznego, czy charakterystycznego

```

# ilość cech
ncol(mydata1)

## [1] 20

# ilość przypadków
nrow(mydata1)

## [1] 3333

# rozmiar ramki danych
object.size(mydata1)

## 381680 bytes

factorvar <- which(sapply(mydata1, is.factor))
numericvar <- which(sapply(mydata1, is.numeric))
# cechy typu charakterystycznego
factorvar

##      State Int.l.Plan VMail.Plan      Churn.
##          1           4           5           20

# cechy typu numerycznego
numericvar

## Account.Length      Area.Code  VMail.Message      Day.Mins      Day.Calls
##          2                  3                  6                  7                  8
## Day.Charge      Eve.Mins      Eve.Calls      Eve.Charge      Night.Mins
##          9                 10                 11                 12                 13
## Night.Calls      Night.Charge      Intl.Mins      Intl.Calls      Intl.Charge
##         14                 15                 16                 17                 18
## CustServ.Calls
##          19

# ilość cech brakujących
missingvar <- is.na(mydata1)
sum(missingvar)

## [1] 0

```

- Z widocznych rezultatów możemy zauważać, że nie ma w naszych danych brakujących obserwacji, zatem także po przejrzeniu tabelki upewniliśmy się, że nie ma nietypowych zmiennych. Doszliśmy do wniosku, że cecha Phone jest nieistotna w celach naszego badania, zatem usunęliśmy ją z tabelki

### 3 Analiza opisowa. Wskaźniki sumaryczne i wykresy.

#### 3.1 Wskaźniki sumaryczne

- Przeanalizujemy miary położenia i rozproszenia dla cech numerycznych. Zbadamy średnie, mediany, kwantyle, wartości minimalne i maksymalne, wariancję i odchylenie standardowe, rozstęp międzykwartylowy.

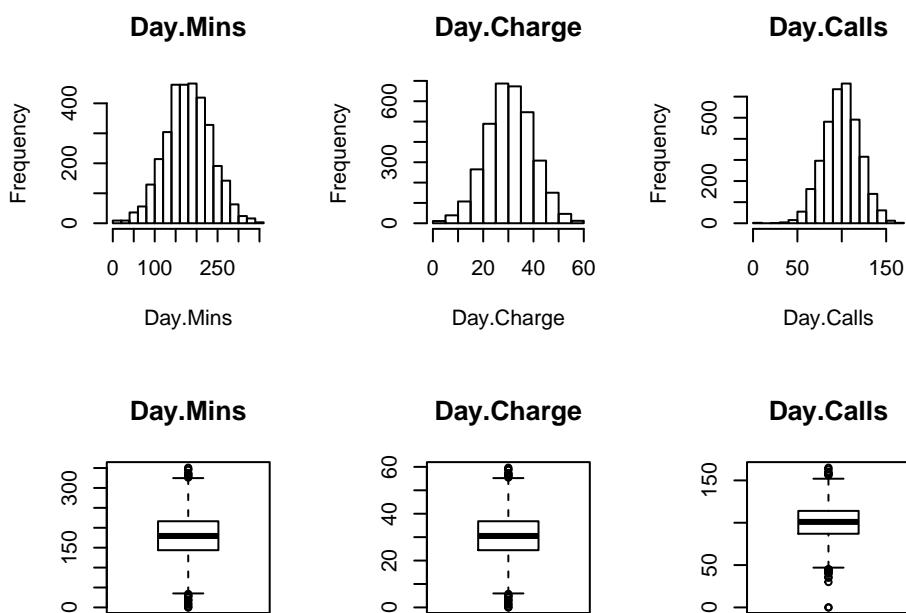
```
my.summary <- function(X)
{
  wynik <- c(min(X), quantile(X, 0.25), median(X), mean(X), quantile(X, 0.75), max(X), var(X))
  names(wynik) <- c("min", "Q1", "median", "mean", "Q3", "max", "var", "sd", "IQR")
  return(wynik)
}
sapply(mydata1[-c(1,3,4,5,20)] , function(x) my.summary(as.numeric(x)) )

##          Account.Length VMail.Message    Day.Mins Day.Calls Day.Charge Eve.Mins
## min           1.000000     0.000000    0.00000   0.00000   0.000000  0.00000
## Q1            74.000000    0.000000  143.70000  87.00000  24.430000 166.60000
## median        101.000000   0.000000 179.40000 101.00000 30.500000 201.40000
## mean          101.06481    8.09901 179.77510 100.43564 30.562307 200.98035
## Q3            127.000000   20.00000 216.40000 114.00000 36.790000 235.30000
## max           243.000000   51.00000 350.80000 165.00000 59.640000 363.70000
## var           1585.80012   187.37135 2966.69649 402.76814 85.737128 2571.89402
## sd             39.82211    13.68837  54.46739  20.06908  9.259435  50.71384
## IQR           53.000000   20.00000  72.70000  27.00000 12.360000  68.70000
##          Eve.Calls Eve.Charge Night.Mins Night.Calls Night.Charge Intl.Mins
## min           0.000000  0.000000  23.20000   33.00000   1.040000  0.000000
## Q1            87.000000 14.160000 167.00000   87.00000   7.520000  8.500000
## median        100.000000 17.120000 201.20000  100.00000   9.050000 10.300000
## mean          100.11431  17.083540 200.87204  100.10771   9.039325 10.237294
## Q3            114.000000 20.000000 235.30000  113.00000  10.590000 12.100000
## max           170.000000 30.910000 395.00000  175.00000  17.770000 20.000000
## var           396.91100 18.581856 2557.71400  382.93047   5.179597  7.794368
## sd             19.92263  4.310668  50.57385   19.56861   2.275873  2.791840
## IQR           27.000000  5.840000  68.30000   26.00000   3.070000  3.600000
##          Intl.Calls Intl.Charge CustServ.Calls
## min           0.000000  0.0000000   0.000000
## Q1            3.000000  2.3000000   1.000000
## median        4.000000  2.7800000   1.000000
## mean          4.479448  2.7645815   1.562856
## Q3            6.000000  3.2700000   2.000000
## max           20.000000  5.4000000   9.000000
## var            6.057576  0.5681732   1.730517
## sd             2.461214  0.7537726   1.315491
## IQR           3.000000  0.9700000   1.000000
```

## 3.2 Wykresy

- Aby nasza analiza była jak najbardziej poprawna, dobierzemy do danych jak najlepsze typy wykresów.

```
par(mfrow=c(2,3))
hist(Day.Mins, main = "Day.Mins")
hist(Day.Charge, main = "Day.Charge")
hist(Day.Calls, main = "Day.Calls")
boxplot(Day.Mins, main="Day.Mins")
boxplot(Day.Charge, main="Day.Charge")
boxplot(Day.Calls, main="Day.Calls")
```



```
skewness(Day.Mins, na.rm=FALSE, type = 3)
## [1] -0.0290509

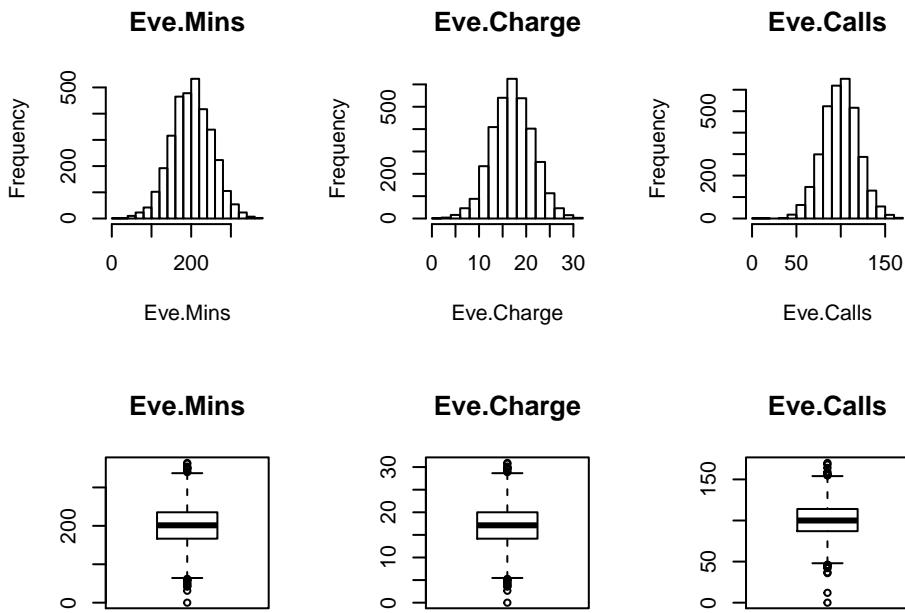
skewness(Day.Charge, na.rm=FALSE, type = 3)
## [1] -0.0290571

skewness(Day.Calls, na.rm=FALSE, type = 3)
## [1] -0.111686
```

```

par(mfrow=c(2,3))
hist(Eve.Mins, main = "Eve.Mins")
hist(Eve.Charge, main = "Eve.Charge")
hist(Eve.Calls, main = "Eve.Calls")
boxplot(Eve.Mins, main="Eve.Mins")
boxplot(Eve.Charge, main="Eve.Charge")
boxplot(Eve.Calls, main="Eve.Calls")

```



```

skewness(Eve.Mins, na.rm=FALSE, type = 3)
## [1] -0.02385597

skewness(Eve.Charge, na.rm=FALSE, type = 3)
## [1] -0.02383652

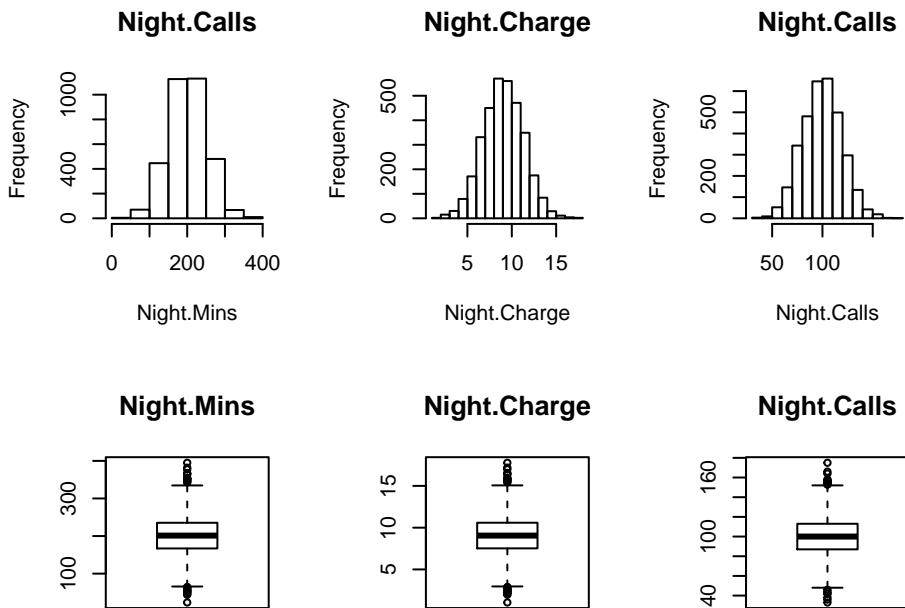
skewness(Eve.Calls, na.rm=FALSE, type = 3)
## [1] -0.05551314

```

```

par(mfrow=c(2,3))
hist(Night.Mins, main = "Night.Calls")
hist(Night.Charge, main = "Night.Charge")
hist(Night.Calls, main = "Night.Calls")
boxplot(Night.Mins, main="Night.Mins")
boxplot(Night.Charge, main="Night.Charge")
boxplot(Night.Calls, main="Night.Calls")

```

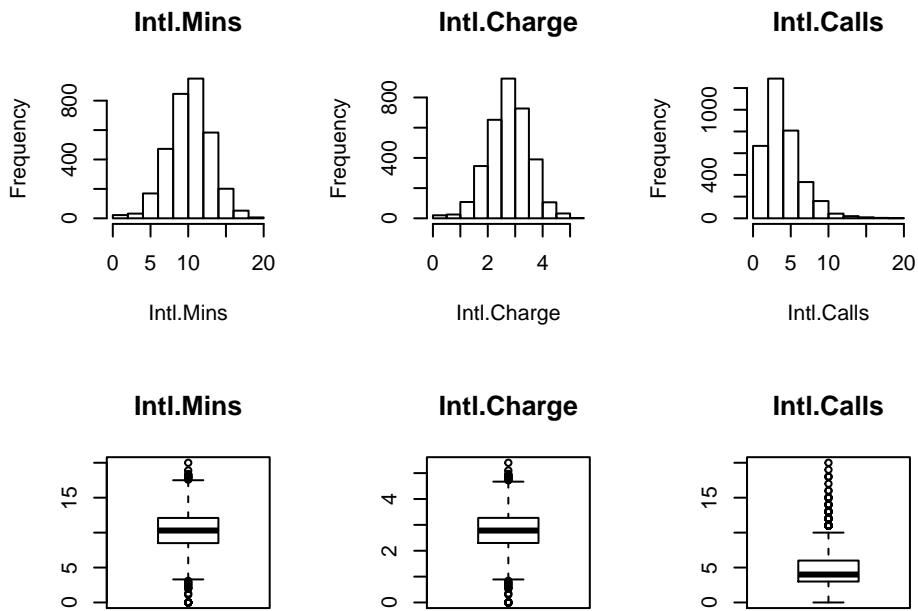


```
skewness(Night.Mins, na.rm=FALSE, type = 3)
## [1] 0.008913263

skewness(Night.Charge, na.rm=FALSE, type = 3)
## [1] 0.00887824

skewness(Night.Calls, na.rm=FALSE, type = 3)
## [1] 0.03247032
```

```
par(mfrow=c(2,3))
hist(Intl.Mins, main = "Intl.Mins")
hist(Intl.Charge, main = "Intl.Charge")
hist(Intl.Calls, main = "Intl.Calls")
boxplot(Intl.Mins, main="Intl.Mins")
boxplot(Intl.Charge, main="Intl.Charge")
boxplot(Intl.Calls, main="Intl.Calls")
```



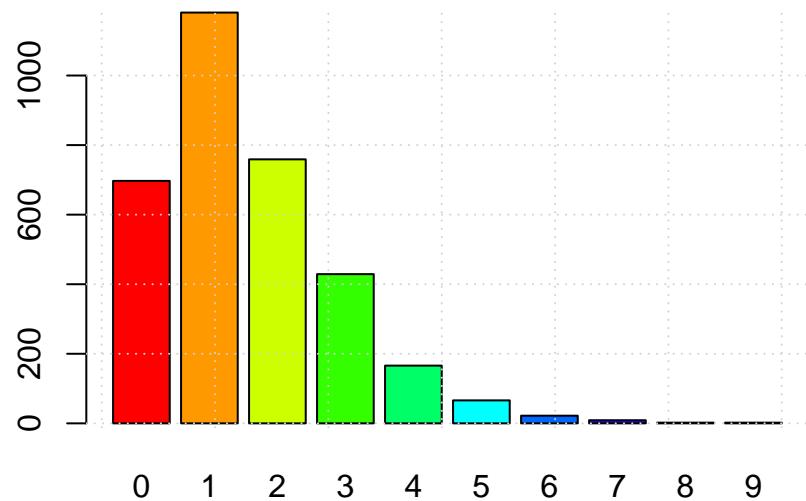
```
skewness(Intl.Mins, na.rm=FALSE, type = 3)
## [1] -0.2449153

skewness(Intl.Charge, na.rm=FALSE, type = 3)
## [1] -0.2450658

skewness(Intl.Calls, na.rm=FALSE, type = 3)
## [1] 1.320289
```

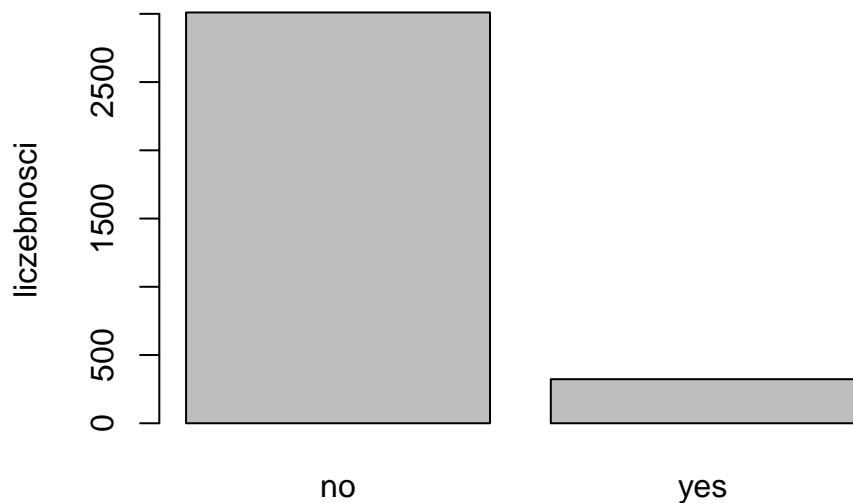
```
barplot(table(CustServ.Calls), main="CustServ.Calls", col=rainbow(10))
grid()
```

### CustServ.Calls



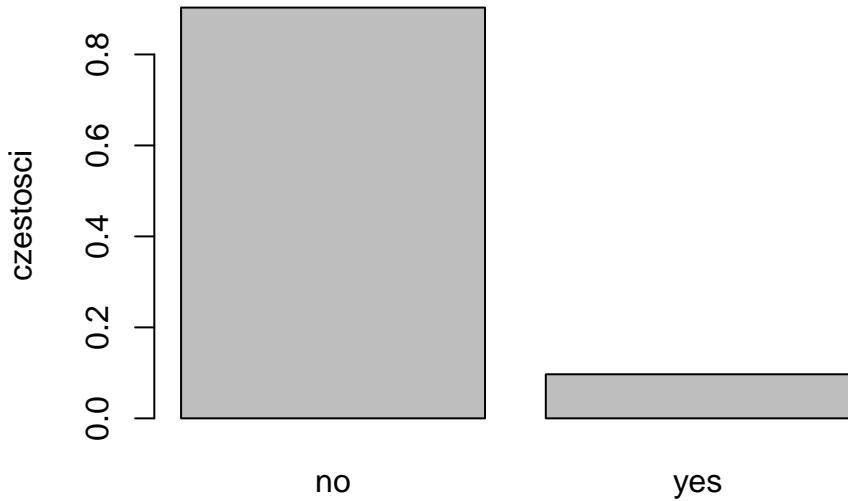
```
barplot(table(Int.l.Plan), main="Int.l.Plan", ylab="liczebnosci")
```

### Int.l.Plan



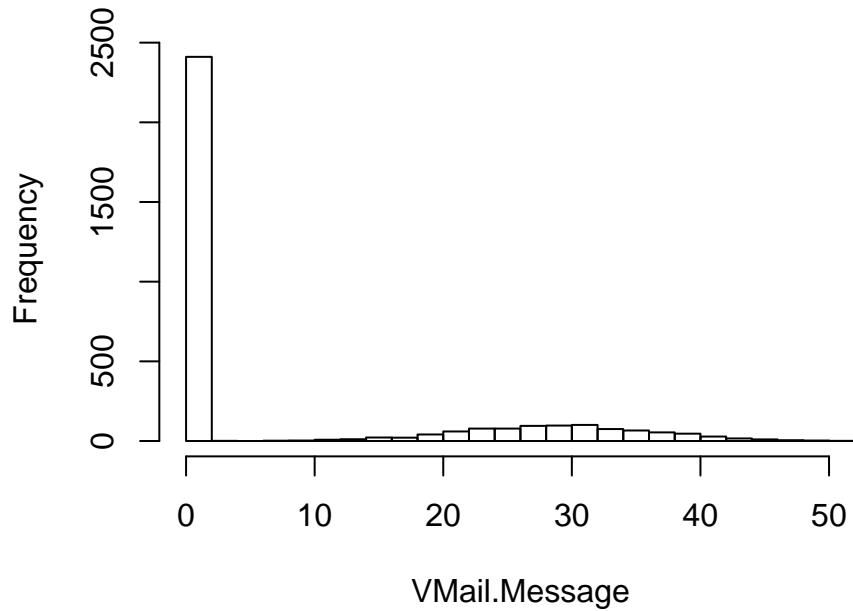
```
barplot(table(Int.l.Plan)/length(Int.l.Plan), main="Int.l.Plan", ylab="czestosc")
```

**Int.I.Plan**



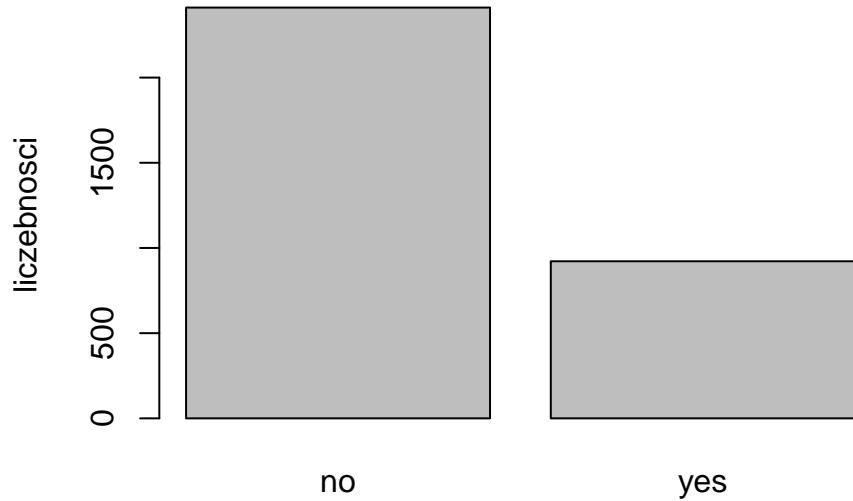
```
hist(VMail.Message, breaks = 20)
```

**Histogram of VMail.Message**



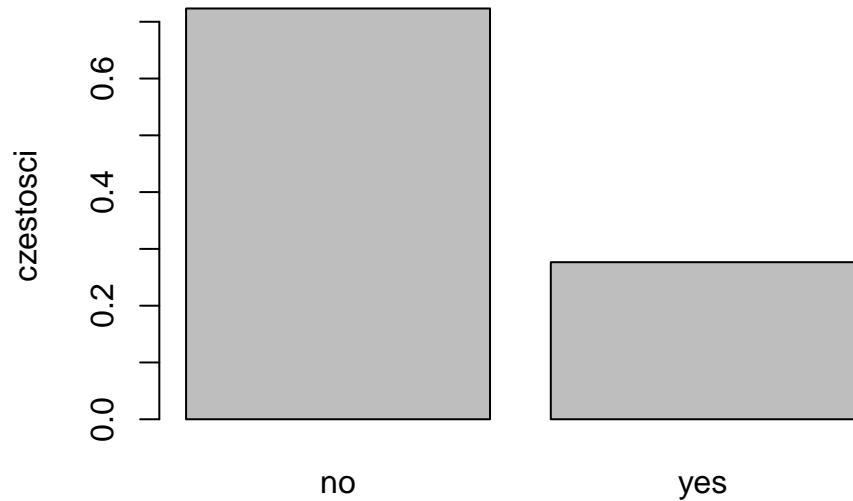
```
barplot(table(VMail.Plan), main="VMail.Plan", ylab="liczebnosci")
```

### VMail.Plan



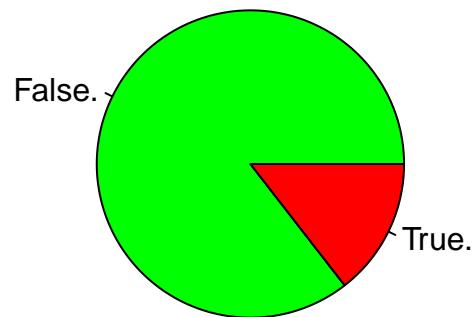
```
barplot(table(VMail.Plan)/length(VMail.Plan), main="VMail.Plan", ylab="czestosci")
```

### VMail.Plan



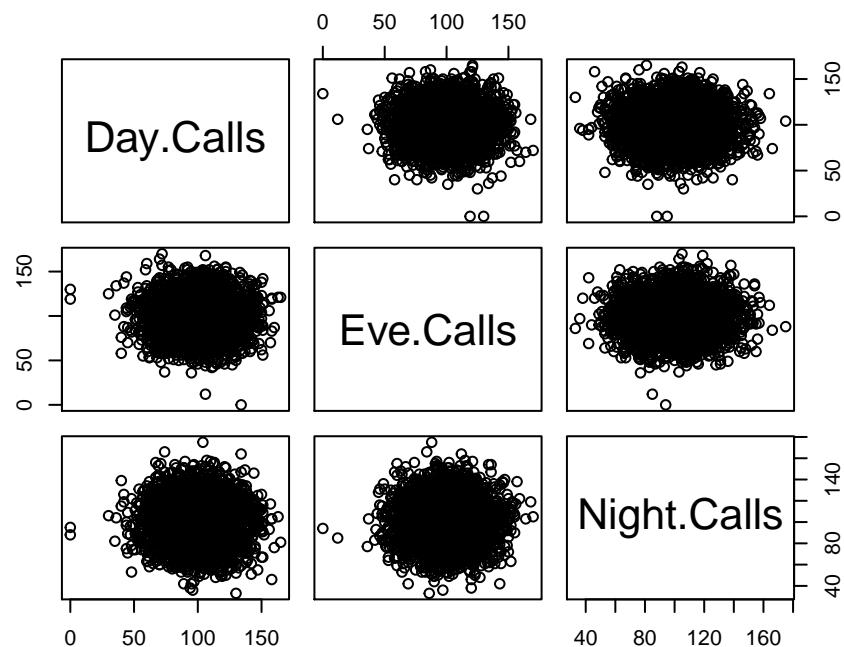
```
pie(table(Churn.), col=c("green","red"), main = "Statystyka klientów, którzy opuścili fi
```

## Statystyka klientów, którzy opuscili firme

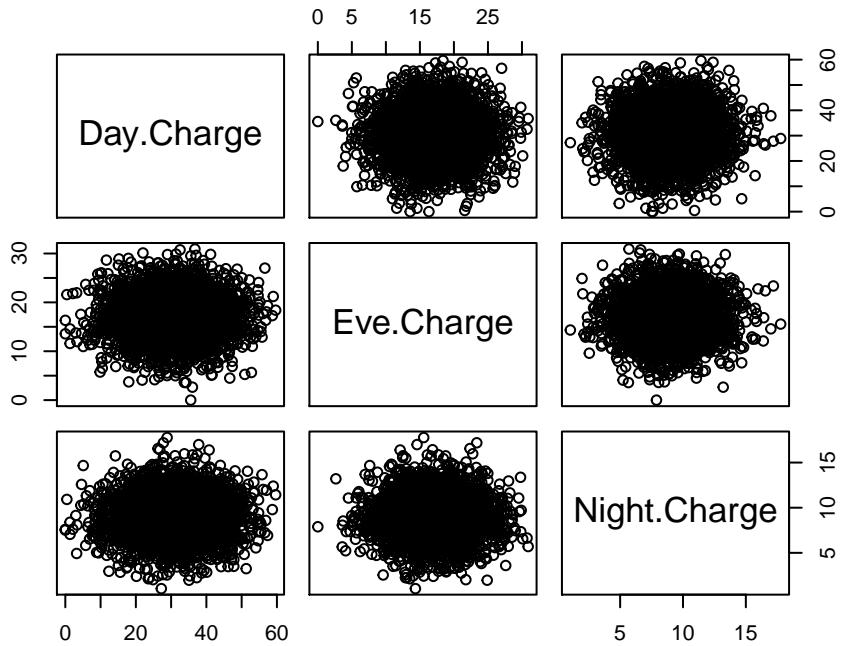


- Wykorzystując wykresy rozrzutu, przeanalizujemy czy występują istotne zależności pomiędzy niektórymi zmiennymi.

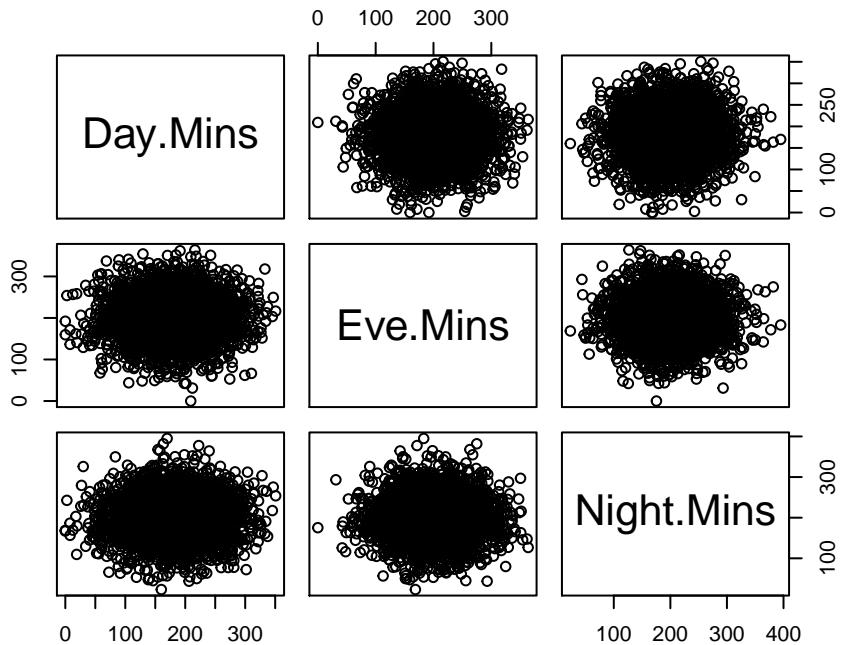
```
pairs(cbind(Day.Calls,Eve.Calls, Night.Calls))
```



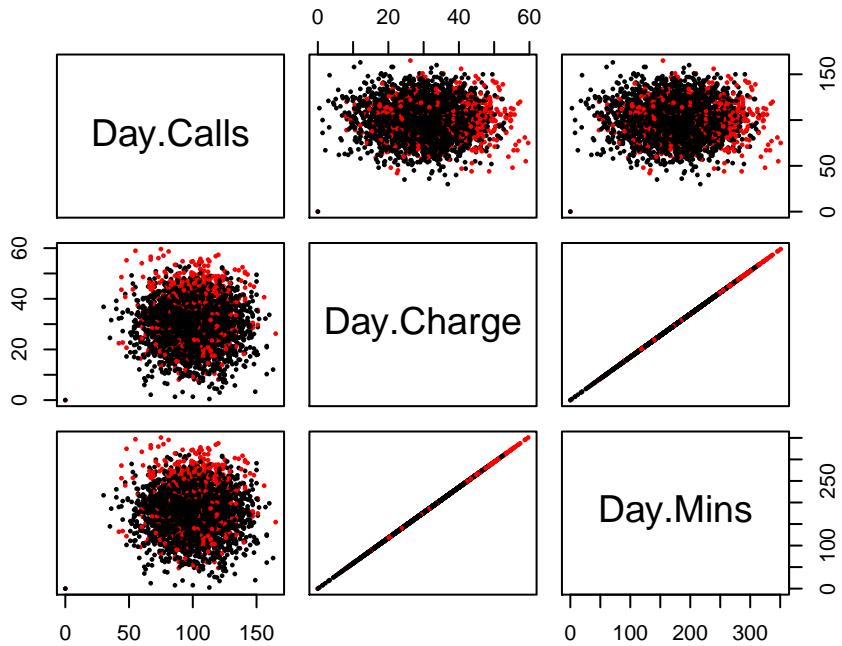
```
pairs(cbind(Day.Charge,Eve.Charge, Night.Charge))
```



```
pairs(cbind(Day.Mins,Eve.Mins, Night.Mins))
```



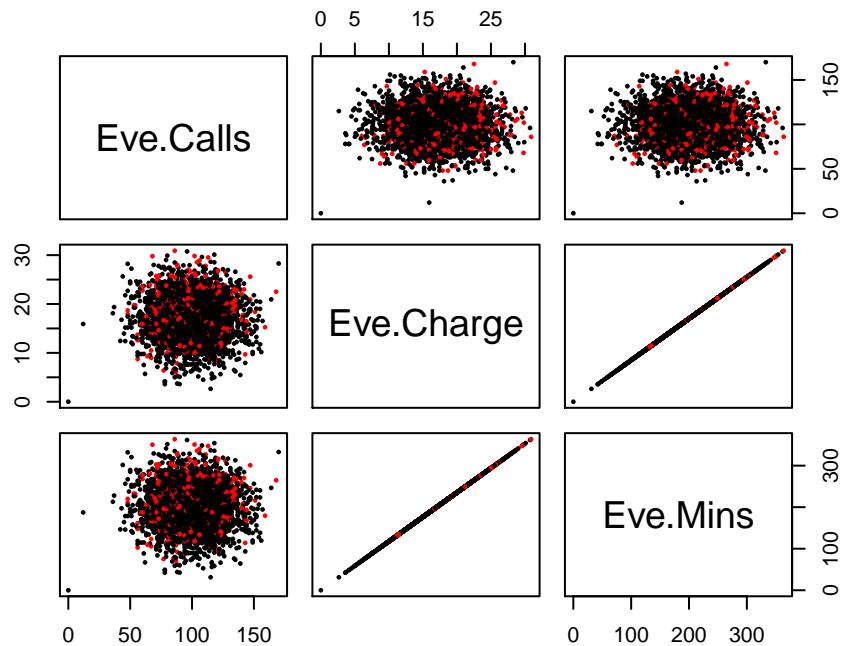
```
pairs(cbind(Day.Calls,Day.Charge, Day.Mins), col=Churn., pch=16, cex=0.5)
```



```
cor(cbind(Day.Calls,Day.Charge, Day.Mins ))
```

```
##           Day.Calls  Day.Charge   Day.Mins
## Day.Calls  1.000000000  0.006752962  0.006750414
## Day.Charge  0.006752962  1.000000000  0.999999952
## Day.Mins    0.006750414  0.999999952  1.000000000
```

```
pairs(cbind(Eve.Calls,Eve.Charge, Eve.Mins), col=Churn., pch=16, cex=0.5)
```



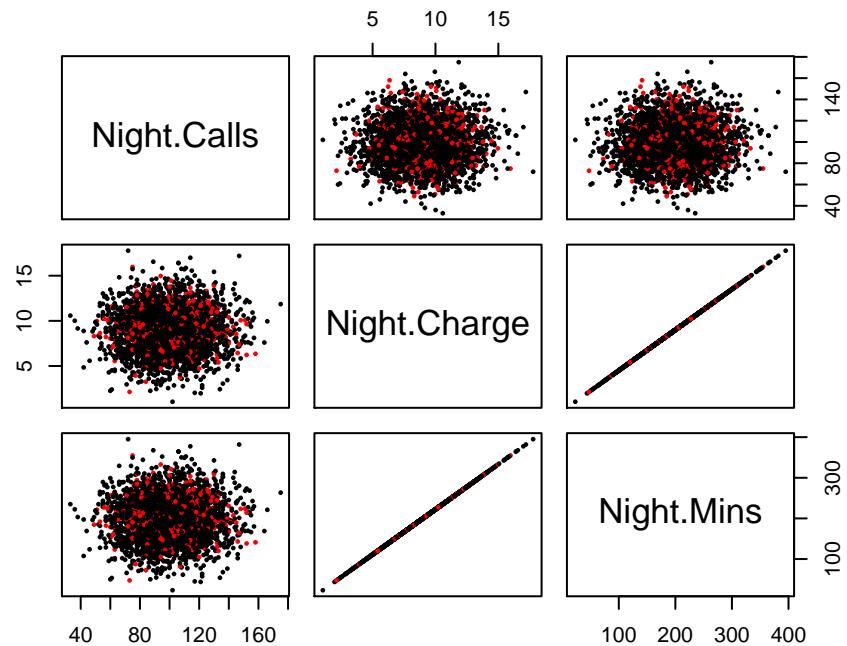
```

cor(cbind(Eve.Calls,Eve.Charge, Eve.Mins))

##           Eve.Calls   Eve.Charge   Eve.Mins
## Eve.Calls  1.00000000 -0.01142289 -0.01143011
## Eve.Charge -0.01142289  1.00000000  0.99999978
## Eve.Mins   -0.01143011  0.99999978  1.00000000

pairs(cbind(Night.Calls,Night.Charge, Night.Mins), col=Churn., pch=16, cex=0.5)

```



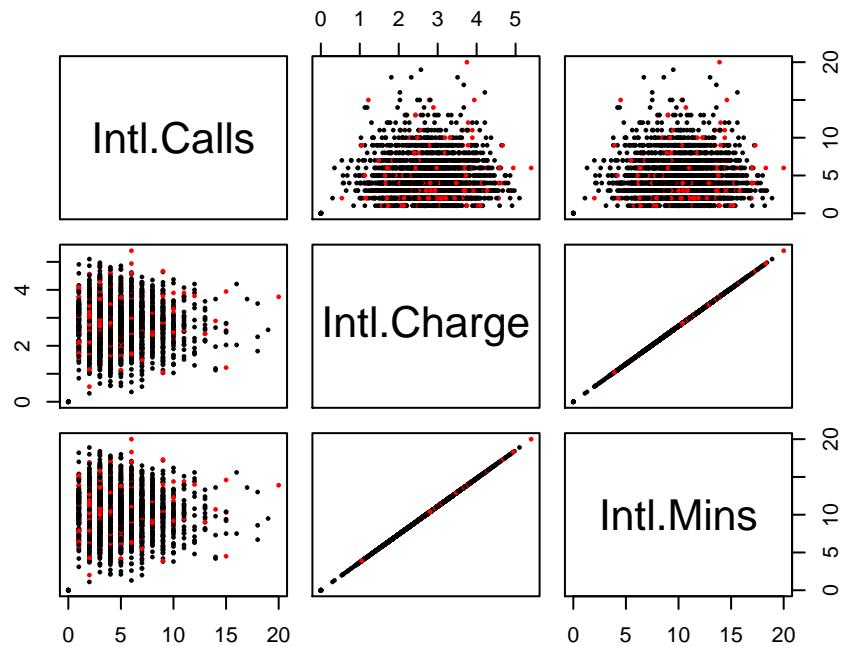
```

cor(cbind(Night.Calls, Night.Charge, Night.Mins))

##           Night.Calls Night.Charge Night.Mins
## Night.Calls    1.00000000  0.01118782  0.01120386
## Night.Charge   0.01118782  1.00000000  0.99999921
## Night.Mins     0.01120386  0.99999921  1.00000000

pairs(cbind(Intl.Calls, Intl.Charge, Intl.Mins), col=Churn., pch=16, cex=0.5)

```



```
cor(cbind(Intl.Calls,Intl.Charge, Intl.Mins))

##                Intl.Calls Intl.Charge  Intl.Mins
## Intl.Calls    1.00000000  0.03237215  0.03230388
## Intl.Charge   0.03237215  1.00000000  0.99999274
## Intl.Mins     0.03230388  0.99999274  1.00000000
```

- Z przedstawionych wykresów rozrzutu i wskaźników korelacji możemy zauważyc, że istotne zależności występują głównie pomiędzy zmiennymi na temat ilości minut połączeń i ich opłat w danych porach doby.

### 3.3 Interpretacja otrzymanych wyników

- Zakres możliwych wartości dla poszczególnych danych numerycznych przedstawia się odpowiednio (podane wartości są przybliżone):
  1. Długość konta: 1-243 miesięcy
  2. Wiadomości poczty głosowej: 0-51
  3. Minuty połączeń w ciągu dnia: 0-360
  4. Ilość połączeń w ciągu dnia: 0-165
  5. Opłata za operacje w ciągu dnia: 0-60 dolarów
  6. Minuty połączeń w godzinach wieczornych: 0-364
  7. Ilość połączeń w godzinach wieczornych: 0-170
  8. Opłata za operacje w godzinach wieczornych: 0-31 dolarów
  9. Minuty połączeń w godzinach nocnych: 23-395
  10. Ilość połączeń w godzinach nocnych: 33-175
  11. Opłata za operacje w godzinach nocnych: 1-18 dolarów
  12. Minuty połączeń międzynarodowych: 0-20
  13. Ilość połączeń międzynarodowych: 0-20
  14. Opłata za operacje międzynarodowe: 0-5 dolarów
  15. Połączenia z obsługą klienta: 0-9
- Zmienne ilościowe takie jak Night.Mins i Night.Charge wykazują rozkład symetryczny. Pomaga nam to określić funkcja skewness. Wartości dla tych zmiennych są bardzo zbliżone do 0, a więc "rozkładają się" symetrycznie.
- Największą zmiennością charakteryzują się zmienne dotyczące minut połączeń w odpowiednich okresach dnia, a także długość konta. Wynika to z największych wartości odchylenia standardowego (a więc także wariancji).
- Ze zmiennych jakościowych można zauważać na przykład, że plan wiadomości poczty głosowej jest częściej przyjmowany niż plan połączeń międzynarodowych.

## 4 Analiza opisowa z podziałem na grupy

### 4.1 Badanie zależności pomiędzy poszczególnymi cechami a przynależnością do jednej z grup

```
dane.lojalni <- subset(mydata1, Churn=="False.")  
dane.odeszli <- subset(mydata1, Churn=="True.")
```

```

my.summary <- function(x)
{
  wynik <- c(min(x), quantile(x, 0.25), median(x), mean(x), quantile(x, 0.75), max(x), var(x))
  names(wynik) <- c("min", "Q1", "median", "mean", "Q3", "max", "var", "sd", "IQR")
  return(wynik)
}
sapply(dane.lojalni[-c(1,3,4,5,20)] , function(x) my.summary(as.numeric(x)) )

##          Account.Length VMail.Message    Day.Mins Day.Calls Day.Charge Eve.Mins
## min           1.000000     0.000000     0.000000  0.000000  0.000000  0.000000
## Q1            73.000000    0.000000   142.82500  87.00000 24.282500 164.50000
## median        100.000000    0.000000   177.20000 100.00000 30.120000 199.60000
## mean          100.79368    8.604561   175.17575 100.28316 29.780421 199.04330
## Q3            127.000000   22.000000   210.30000 114.00000 35.750000 233.20000
## max           243.000000   51.000000   315.60000 163.00000 53.650000 361.80000
## var           1590.60186   193.575058  2518.19852 392.08582 72.775154 2529.30283
## sd            39.88235    13.913125   50.18166 19.80116  8.530835  50.29217
## IQR           54.000000   22.000000   67.47500 27.00000 11.467500  68.70000
##          Eve.Calls Eve.Charge Night.Mins Night.Calls Night.Charge Intl.Mins
## min           0.000000  0.000000  23.20000  33.00000  1.040000  0.000000
## Q1            87.000000 13.980000 165.90000  87.00000  7.470000  8.400000
## median        100.000000 16.970000 200.25000 100.00000  9.010000 10.200000
## mean          100.03860 16.918909 200.13319 100.05825  9.006074 10.158877
## Q3            114.000000 19.820000 234.90000 113.00000 10.570000 12.000000
## max           170.000000 30.750000 395.00000 175.00000 17.770000 18.900000
## var           398.33828 18.274451 2611.72433 380.49362  5.288931  7.753380
## sd            19.95841  4.274863  51.10503  19.50625  2.299768  2.784489
## IQR           27.000000  5.840000  69.00000  26.00000  3.100000  3.600000
##          Intl.Calls Intl.Charge CustServ.Calls
## min           0.000000  0.00000000  0.000000
## Q1            3.000000  2.2700000  1.000000
## median        4.000000  2.7500000  1.000000
## mean          4.532982  2.7434035  1.449825
## Q3            6.000000  3.2400000  2.000000
## max           19.000000  5.1000000  8.000000
## var           5.963285  0.5651797  1.354624
## sd            2.441984  0.7517843  1.163883
## IQR           3.000000  0.9700000  1.000000

sapply(dane.odeszli[-c(1,3,4,5,20)] , function(x) my.summary(as.numeric(x)) )

##          Account.Length VMail.Message    Day.Mins Day.Calls Day.Charge Eve.Mins
## min           1.000000     0.000000     0.000000  0.000000  0.000000  70.90000
## Q1            76.000000    0.000000   153.25000  87.50000 26.05500 177.10000
## median        103.000000    0.000000   217.60000 103.00000 36.99000 211.30000
## mean          102.66460    5.115942  206.91408 101.33540 35.17592 212.41014
## Q3            127.000000    0.000000   265.95000 116.50000 45.21000 249.45000
## max           225.000000   48.000000  350.80000 165.00000 59.64000 363.70000

```

```

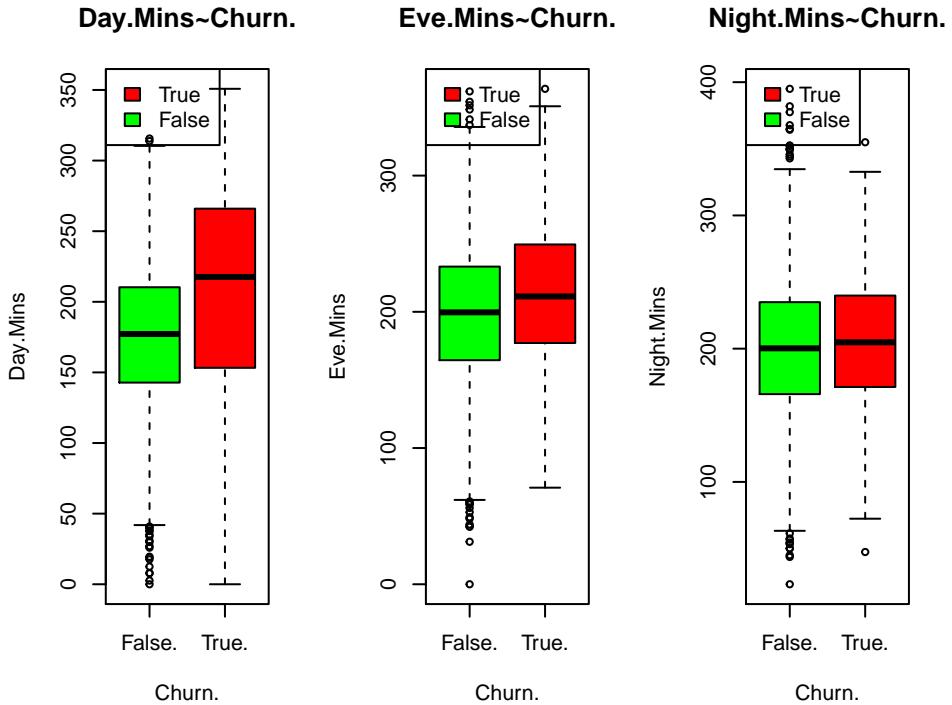
## var      1557.70885   140.662878 4760.69524 465.79598 137.58609 2675.88008
## sd       39.46782    11.860138 68.99779 21.58231 11.72971 51.72891
## IQR     51.00000    0.000000 112.70000 29.00000 19.15500 72.35000
##   Eve.Calls Eve.Charge Night.Mins Night.Calls Night.Charge Intl.Mins
## min     48.00000   6.030000 47.40000 49.00000 2.130000 2.000000
## Q1      87.00000  15.055000 171.25000 85.00000 7.705000 8.800000
## median 101.00000 17.960000 204.80000 100.00000 9.220000 10.600000
## mean   100.56108 18.054969 205.23168 100.39959 9.235528 10.700000
## Q3      114.00000 21.205000 239.85000 115.00000 10.795000 12.800000
## max     168.00000 30.910000 354.90000 158.00000 15.970000 20.000000
## var     389.06421 19.331518 2221.50316 398.02880 4.498983 7.801909
## sd      19.72471  4.396762 47.13282 19.95066 2.121081 2.793190
## IQR     27.00000  6.150000 68.60000 30.00000 3.090000 4.000000
##   Intl.Calls Intl.Charge CustServ.Calls
## min     1.000000 0.5400000 0.000000
## Q1      2.000000 2.3800000 1.000000
## median 4.000000 2.8600000 2.000000
## mean   4.163561 2.8895445 2.229814
## Q3      5.000000 3.4600000 4.000000
## max     20.000000 5.4000000 9.000000
## var     6.510537 0.5687454 3.434628
## sd      2.551575 0.7541521 1.853275
## IQR     3.000000 1.0800000 3.000000

```

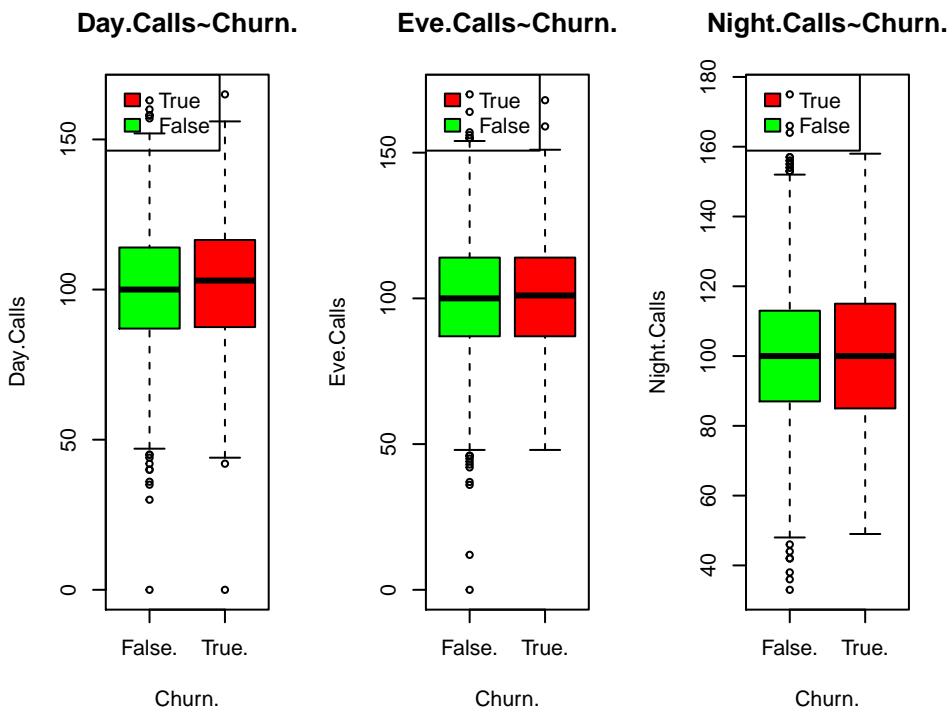
```

par(mfrow=c(1,3))
boxplot(Day.Mins~Churn., main="Day.Mins~Churn.", col=c("green","red"), xlab="Churn.", yl
legend(x = "topleft", legend=c('True','False'), fill=c("red","green"))
boxplot(Eve.Mins~Churn., main="Eve.Mins~Churn.", col=c("green","red"), xlab="Churn.", yl
legend(x = "topleft", legend=c('True','False'), fill=c("red","green"))
boxplot(Night.Mins~Churn., main="Night.Mins~Churn.", col=c("green","red"), xlab="Churn."
legend(x = "topleft", legend=c('True','False'), fill=c("red","green"))

```



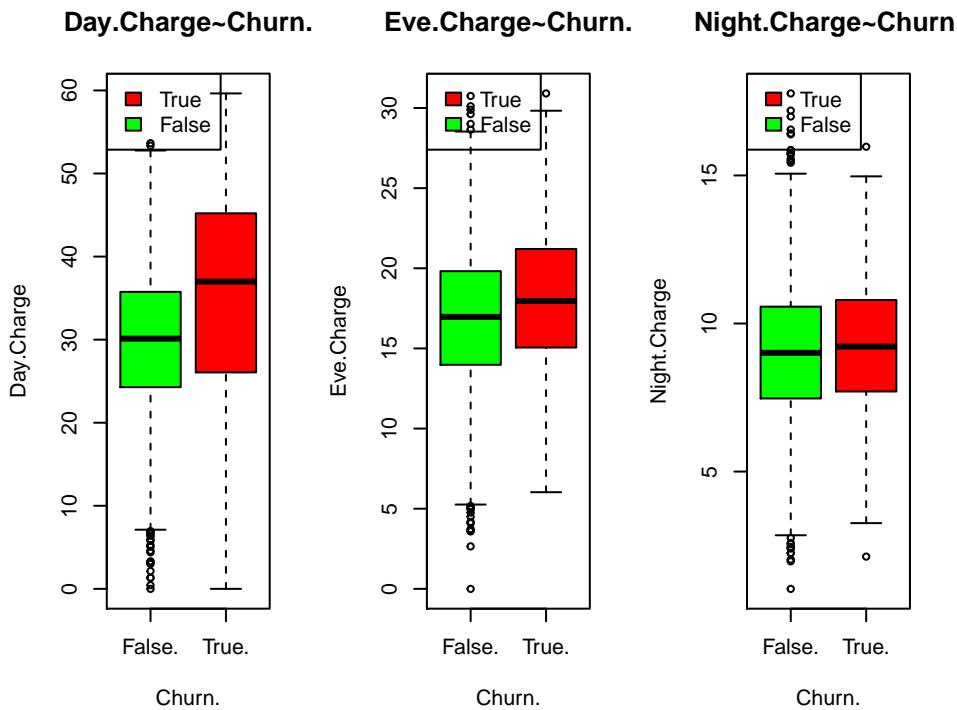
```
par(mfrow=c(1,3))
boxplot(Day.Calls~Churn., main="Day.Calls~Churn.", col=c("green","red"), xlab="Churn.",
legend(x = "topleft", legend=c('True','False'), fill=c("red","green")))
boxplot(Eve.Calls~Churn., main="Eve.Calls~Churn.", col=c("green","red"), xlab="Churn.",
legend(x = "topleft", legend=c('True','False'), fill=c("red","green")))
boxplot(Night.Calls~Churn., main="Night.Calls~Churn.", col=c("green","red"), xlab="Churn.",
legend(x = "topleft", legend=c('True','False'), fill=c("red","green")))
```



```

par(mfrow=c(1,3))
boxplot(Day.Charge~Churn., main="Day.Charge~Churn.", col=c("green","red"), xlab="Churn.")
legend(x = "topleft", legend=c('True','False'), fill=c("red","green"))
boxplot(Eve.Charge~Churn., main="Eve.Charge~Churn.", col=c("green","red"), xlab="Churn.")
legend(x = "topleft", legend=c('True','False'), fill=c("red","green"))
boxplot(Night.Charge~Churn., main="Night.Charge~Churn.", col=c("green","red"), xlab="Churn.")
legend(x = "topleft", legend=c('True','False'), fill=c("red","green"))

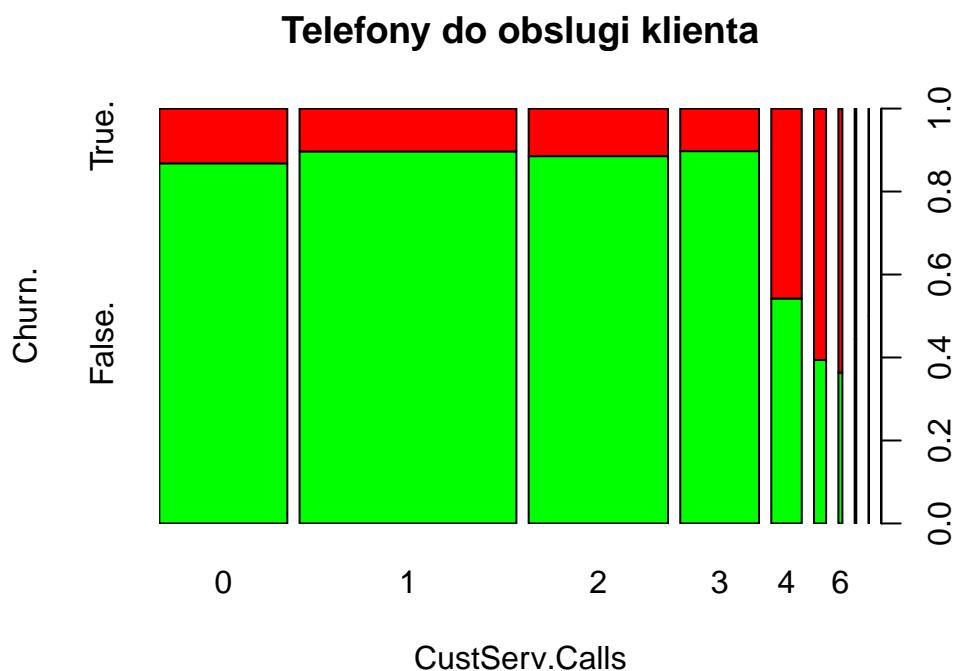
```



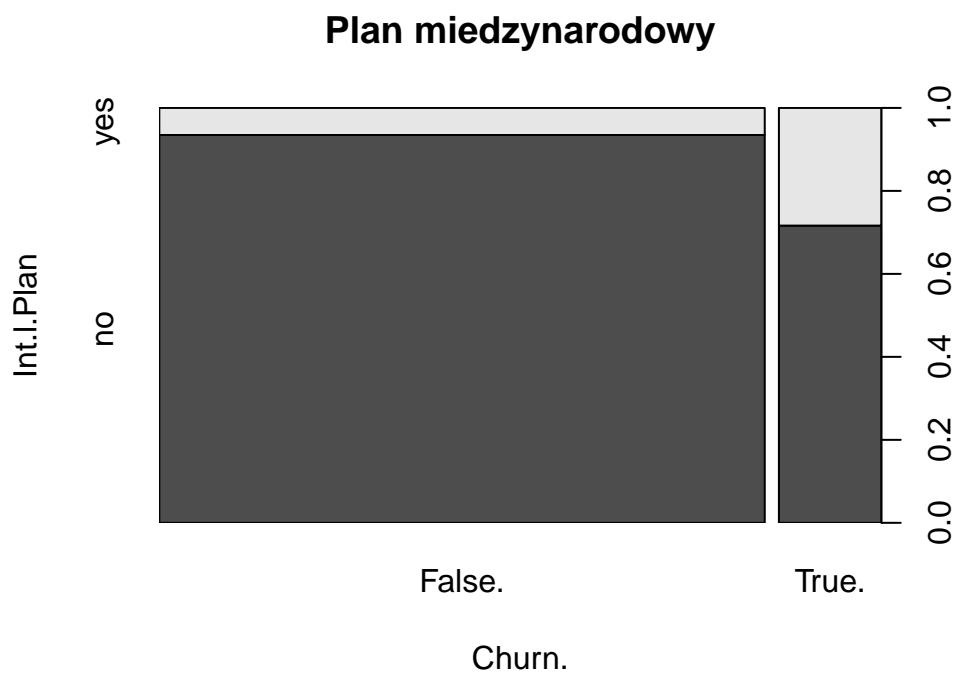
```

plot(Churn.~as.factor(CustServ.Calls), col=c("green","red"), xlab="CustServ.Calls", main="CustServ.Calls~Churn")

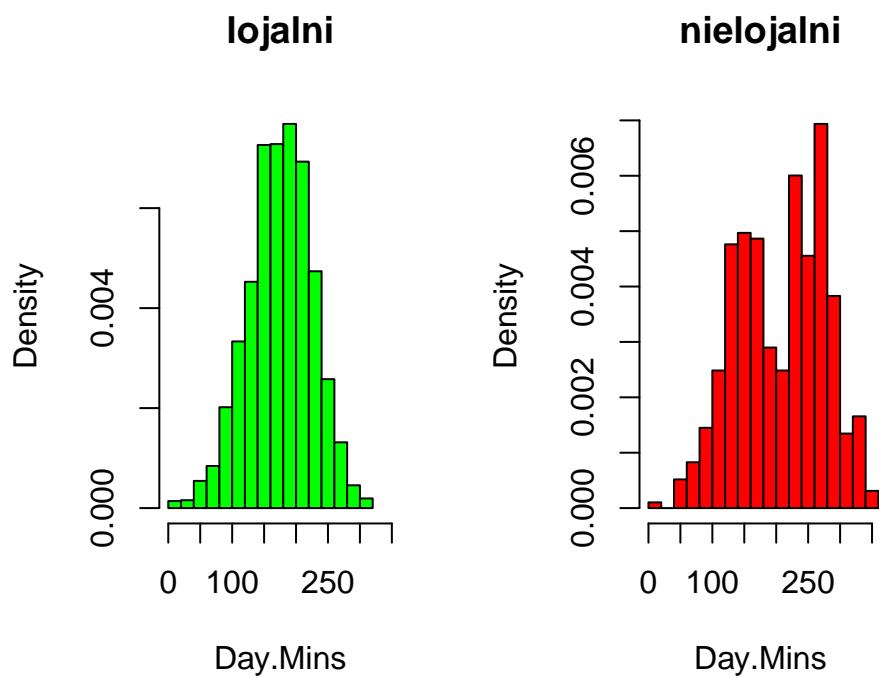
```



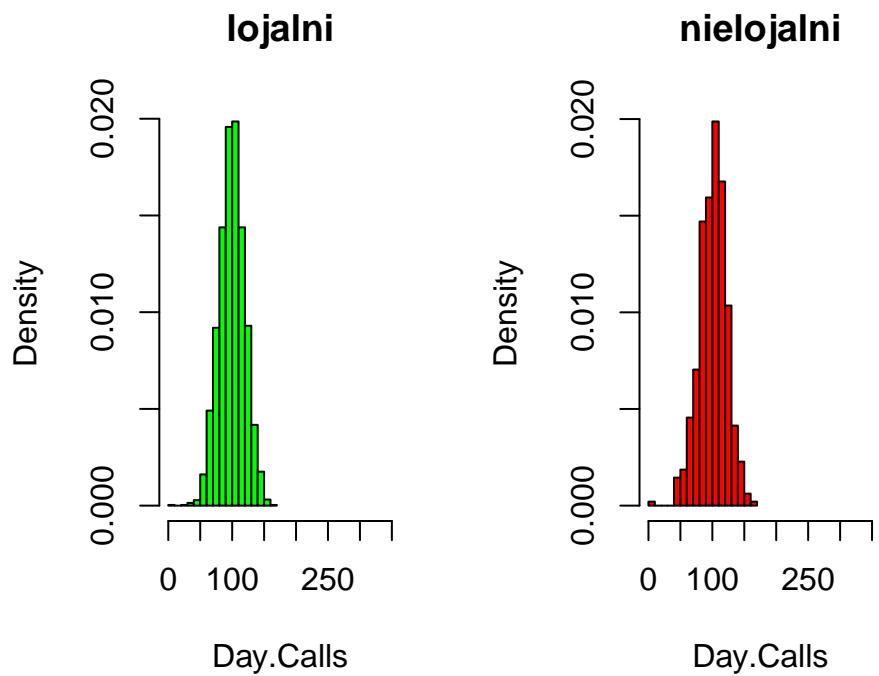
```
plot(Int.l.Plan~Churn., main="Plan międzynarodowy")
```



```
par(mfrow=c(1,2))
zakres <- range(Day.Mins)
hist(dane.lojalni$Day.Mins, xlim=zakres, prob=T, col="green", breaks = 20, main = "lojalni")
hist(dane.odeszli$Day.Mins, xlim=zakres, prob=T, col="red", breaks = 20, main = "nie lojalni")
```



```
par(mfrow=c(1,2))
zakres <- range(Day.Mins)
hist(dane.lojalni$Day.Calls, xlim=zakres, prob=T, col="green", breaks = 20, main = "lojalni")
hist(dane.odeszli$Day.Calls, xlim=zakres, prob=T, col="red", breaks = 20, main = "nielojalni")
```

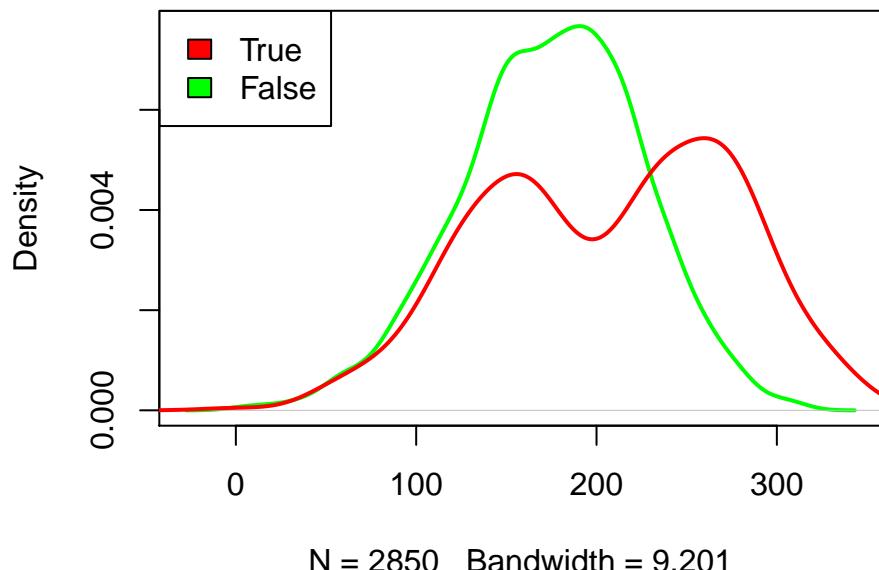


```

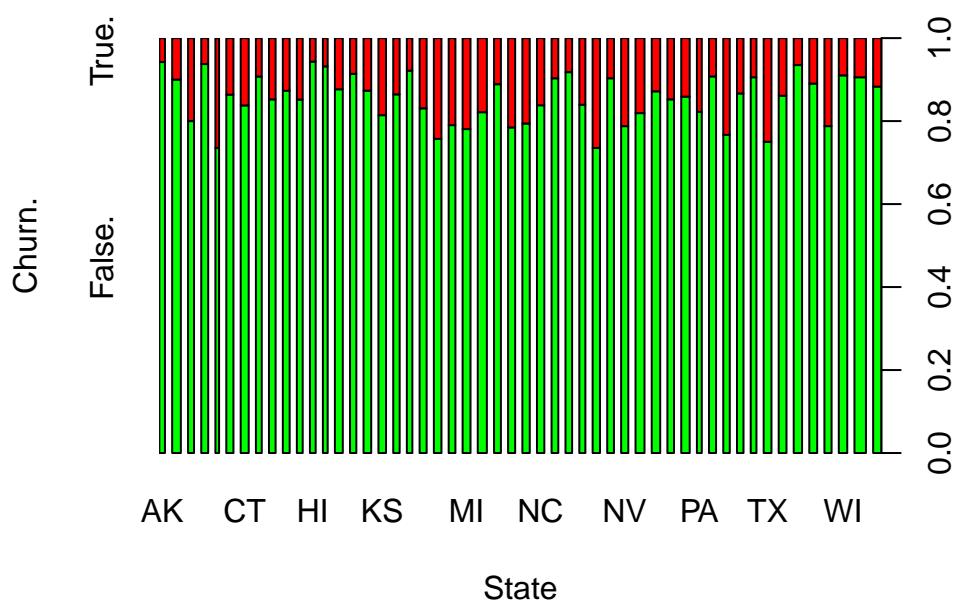
d1 <- density(Day.Mins[Churn=="False"])
d2 <- density(Day.Mins[Churn=="True"])
plot(d1, col="green", lwd=2, main="Porównanie gęstości")
lines(d2, col="red", lwd=2)
legend(x = "topleft", legend=c('True','False'), fill=c("red","green"))

```

## Porównanie gestosci

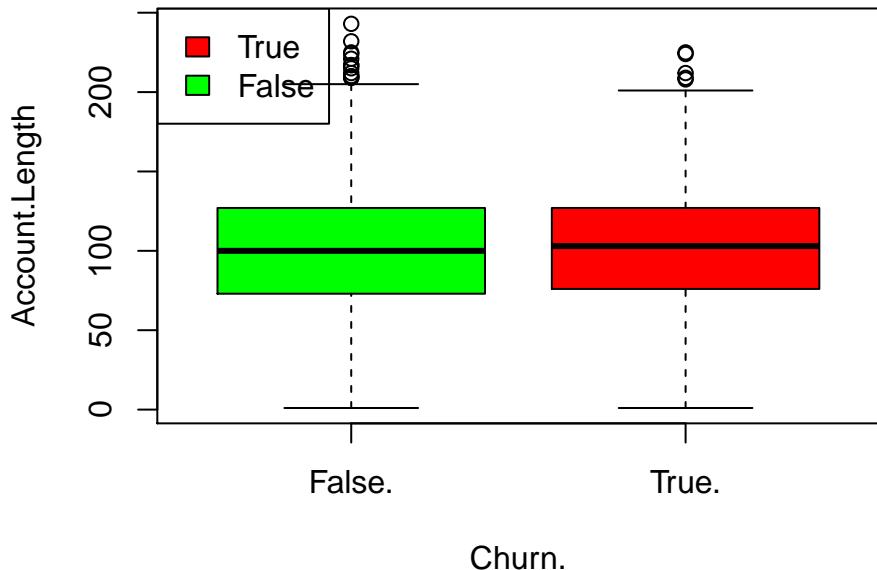


```
plot(Churn ~ as.factor(State), col=c("green","red"), xlab = "State")
```



```
boxplot(Account.Length~Churn., col=c("green","red"), xlab = "Churn.", main="Account.Length~Churn.")
legend(x="topleft", legend=c("True","False"), fill=c("red","green"))
```

**Account.Length~Churn.**



```
by(data=Day.Mins, INDICES=Churn., my.summary)

## Churn.: False.
##      min       Q1     median      mean       Q3      max      var
## 0.00000 142.82500 177.20000 175.17575 210.30000 315.60000 2518.19852
##      sd       IQR
## 50.18166 67.47500
## -----
## Churn.: True.
##      min       Q1     median      mean       Q3      max      var
## 0.00000 153.25000 217.60000 206.91408 265.95000 350.80000 4760.69524
##      sd       IQR
## 68.99779 112.70000
```

```
by(data=Eve.Mins, INDICES=Churn., my.summary)

## Churn.: False.
##      min       Q1     median      mean       Q3      max      var
## 0.00000 164.50000 199.60000 199.04330 233.20000 361.80000 2529.30283
##      sd       IQR
## 50.29217 68.70000
## -----
## Churn.: True.
```

```

##          min        Q1       median        mean        Q3        max     var
## 70.90000 177.10000 211.30000 212.41014 249.45000 363.70000 2675.88008
##          sd        IQR
## 51.72891 72.35000

```

```

by(data=Night.Mins, INDICES=Churn., my.summary)

## Churn.: False.
##          min        Q1       median        mean        Q3        max     var
## 23.20000 165.90000 200.25000 200.13319 234.90000 395.00000 2611.72433
##          sd        IQR
## 51.10503 69.00000
## -----
## Churn.: True.
##          min        Q1       median        mean        Q3        max     var
## 47.40000 171.25000 204.80000 205.23168 239.85000 354.90000 2221.50316
##          sd        IQR
## 47.13282 68.60000

```

```

by(data=Intl.Mins, INDICES=Churn., my.summary)

## Churn.: False.
##          min        Q1       median        mean        Q3        max     var      sd
## 0.000000 8.400000 10.200000 10.158877 12.000000 18.900000 7.753380 2.784489
##          IQR
## 3.600000
## -----
## Churn.: True.
##          min        Q1       median        mean        Q3        max     var      sd
## 2.000000 8.800000 10.600000 10.700000 12.800000 20.000000 7.801909 2.793190
##          IQR
## 4.000000

```

- Odczytując powyższe wykresy zauważamy, że znaczny wpływ na migracje klietów ma ilość minut połączeń w porze dziennej. Widać, że ludzie którzy wykonują więcej operacji dziennie, są bardziej skłonni do rezygnacji z naszej firmy. Wpływ mają także opłaty za operacje w tych porach, co potwierdza liniową zależność, o której mówiliśmy w zadaniu 2.
- Widoczny jest również, że ludzie którzy częściej korzystali z obsługi klienta rezygnowało z ofert firmy, co może sugerować, że obsługa klienta naszej firmy nie funkcjonuje całkowicie prawidłowo. Znaczne różnicowanie jest także w grupie, która używała planu międzynarodowego. Dokładniej, ludzie którzy są użytkownikami tego planu statycznie częściej rezygnowali z usług firmy.
- Histogramy dwóch głównych grup wskazują, że grupa klientów lojalnych reprezentowana jest przez rozkład symetryczny, podczas gdy grupa druga już nie wykazuje takiej cechy.

## **5 Podsumowanie**

### **5.1 Informacje których dowiedzieliśmy się o klientach sieci**

- Plan poczty głosowej jest bardziej popularny wśród naszych klientów niż plan międzynarodowy.
- Pora dnia nie wpływa na ilość, ani czas połączeń.
- Większość klientów odbyła niewiele (od 0 do 3) rozmów z biurem obsługi klienta.

### **5.2 Główne przyczyny odchodzenia klientów. Proponowane rozwiązania.**

- Warto zauważyć, że średni czas trwania jednego połączenia w trakcie dnia u przeciętnego klienta, który zrezygnował z usług naszej firmy, był wyższy niż u przeciętnego klienta, który pozostał jej wierny. Powód takiego stanu rzeczy może być wiele, jednak najprawdopodobniej konkurencja oferuje lepszą ofertę osobom, które prowadzą długie rozmowy telefoniczne. Być może warto byłoby więc rozważyć wprowadzenie specjalnego pakietu stworzonego z myślą o tej grupie klientów.
- Spośród osób, które ponad czterokrotnie rozmawiały z biurem obsługi klienta, ponad połowa zrezygnowała z usług naszej sieci telefonii komórkowej. Warto więc przeanalizować w jaki sposób możemy poprawić działanie biura, aby z lepszą skutecznością rozwiązać problemy klientów, a tym samym zapobiegać procesowi rezygnowania przez nich z naszych usług.
- Niepokojąco duży odsetek osób korzystających z planu międzynarodowego zdecydował się na rezygnację z naszych usług. Warto zatem rozważyć wprowadzenie w nim drobnych korekt i porównanie go z ofertą konkurencji.
- Procent osób, które zdecydowały się na przejście do konkurencji nie rozkłada się identycznie w każdym ze stanów. Przyczyn takiego stanu rzeczy może być jednak bardzo wiele i na podstawie posiadanych danych nie da się jednoznacznie wskazać żadnej z nich.