

# Raport 1

## Analiza opisowa danych CHURN

Jan Solarz 243889  
Szymon Suszek 237288

13 kwietnia 2020

### Spis treści

<b>1</b>	<b>Krótki opis przeprowadzonej analizy</b>	<b>1</b>
<b>2</b>	<b>Etap 1. Podstawowe informacje i przygotowanie danych</b>	<b>1</b>
<b>3</b>	<b>Etap 2. Analiza opisowa danych</b>	<b>2</b>
3.1	Analiza danych ilościowych . . . . .	2
3.2	Interpretacja wyników zmiennej ilościowych . . . . .	14
3.3	Analiza danych jakościowych . . . . .	15
3.4	Interpretacja wyników zmiennej jakościowych . . . . .	18
<b>4</b>	<b>Etap 3. Analiza skupiająca się na podgrupach Churn.</b>	<b>18</b>
4.1	Analiza zmiennych ilościowych . . . . .	18
4.2	Analiza zmiennych jakościowych Churn. . . . .	21
<b>5</b>	<b>Etap 4. Podsumowanie</b>	<b>24</b>
5.1	Ogólne wnioski . . . . .	24
5.2	Wnioski dotyczące podgrup Churn. . . . .	24
5.3	Ewentualne rozwiązania, działania . . . . .	24

## 1 Krótki opis przeprowadzonej analizy

- Analizować będziemy grupę 3333 klientów sieci telefonii komórkowej charakteryzujących się 21 zmiennymi (cechami)
- Naszym zadaniem jest bliższe przyjrzenie się poszczególnym cechom i na ich podstawie wyciągnięcie wniosków, które pomogą nam zrozumieć ich odejście z sieci oraz znalezienie rozwiązań, które przyniłyby się do polepszeń statystyk.

## 2 Etap 1. Podstawowe informacje i przygotowanie danych

- Na tym etapie interesuje nas wielkość analizowanych danych, ich typ oraz to czy nie występują brakujące obserwacje.

```
#liczba zmiennych
ncol(dane)

## [1] 21

#liczba obserwacji
nrow(dane)

## [1] 3333

# sprawdzenie, które zmienne są typu jakościowego
jakosciowe <- which(sapply(dane, is.factor))
jakosciowe

##      State      Phone Int.l.Plan VMail.Plan      Churn.
##         1         4         5         6         21

# sprawdzenie typu ilościowego zmiennych
ilosciowe <- which(sapply(dane, is.numeric))
ilosciowe

## Account.Length      Area.Code VMail.Message      Day.Mins      Day.Calls
##           2           3           7           8           9
##      Day.Charge      Eve.Mins      Eve.Calls      Eve.Charge      Night.Mins
##          10          11          12          13          14
##      Night.Calls      Night.Charge      Intl.Mins      Intl.Calls      Intl.Charge
##          15          16          17          18          19
## CustServ.Calls
##          20

#usunięcie zmiennej identyfikacyjnej numer telefonu
dane<-dane[, -4]
# brakujące wartości
czy_brak <- is.na(dane)
#liczba wszystkich brakujących wartości
l_brakujacych <- sum(czy_brak)
l_brakujacych

## [1] 0

#indeksy obserwacji brakujących
indeksy <- which(czy_brak, arr.ind = TRUE)
indeksy

##      row col

#brak indeksów
```

- Analizujemy 3333 klientów (liczba obserwacji) i 21 cech. Zmienna "Phone" nie jest istotna w dalszej analizie dlatego możemy ją usunąć. Po sprawdzeniu dane nie posiadają żadnych brakujących obserwacji.
- Mamy do czynienia z 16 zmiennymi ilościowymi i 5 jakościowymi (jedna z nich została usunięta). Istotna dla nas jest cecha "Churn.", na której skupimy się w dalszym etapie.

## 3 Etap 2. Analiza opisowa danych

### 3.1 Analiza danych ilościowych

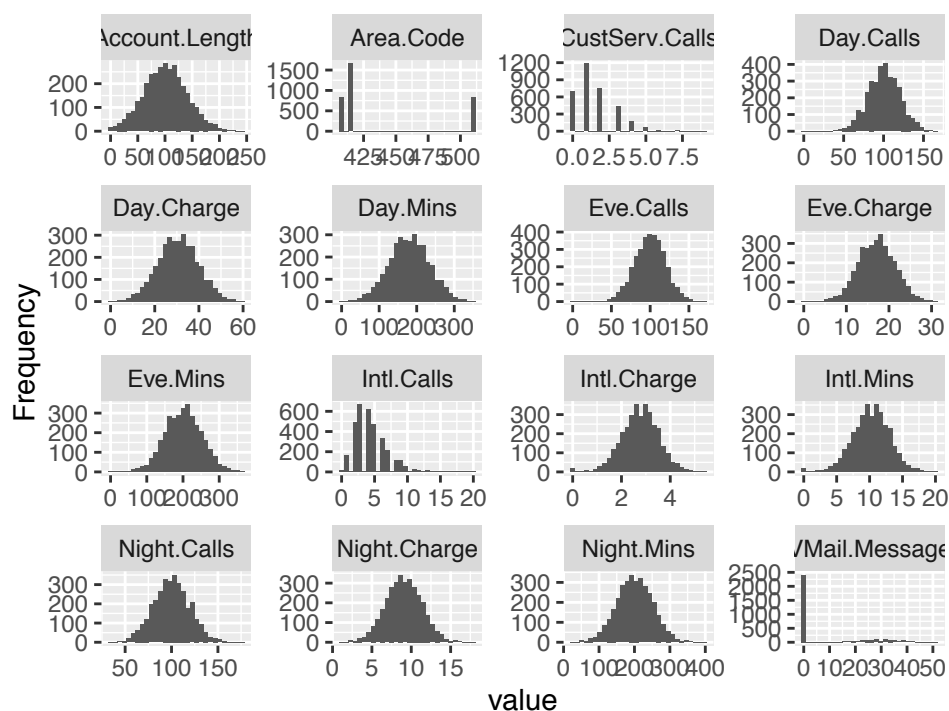
- Wyznamy podstawowe wskaźniki sumaryczne (miary położenia i rozproszenia) w celu skupiania się na tych cechach, z których otrzymamy porządkane przez nas informacje.
- Następnie zajmujemy się dokładniejszą wizualizacją danych

##	Account.Length	VMail.Message	Day.Mins	Day.Calls	Day.Charge
## minimum	1.00000	0.00000	0.00000	0.00000	0.00000
## Q1	74.00000	0.00000	143.70000	87.00000	24.43000
## mediana	101.00000	0.00000	179.40000	101.00000	30.50000
## średnia	101.06481	8.09901	179.77510	100.43564	30.562307
## Q3	127.00000	20.00000	216.40000	114.00000	36.79000
## maksimum	243.00000	51.00000	350.80000	165.00000	59.64000
## rozstęp	242.00000	51.00000	350.80000	165.00000	59.64000
## IQR	53.00000	20.00000	72.70000	27.00000	12.36000
## odch. stand.	39.82211	13.68837	54.46739	20.06908	9.259435
## wariancja	1585.80012	187.37135	2966.69649	402.76814	85.737128
##	Eve.Mins	Eve.Calls	Eve.Charge	Night.Mins	Night.Calls
## minimum	0.00000	0.00000	0.00000	23.20000	33.00000
## Q1	166.60000	87.00000	14.16000	167.00000	87.00000
## mediana	201.40000	100.00000	17.12000	201.20000	100.00000
## średnia	200.98035	100.11431	17.08354	200.87204	100.10771
## Q3	235.30000	114.00000	20.00000	235.30000	113.00000
## maksimum	363.70000	170.00000	30.91000	395.00000	175.00000
## rozstęp	363.70000	170.00000	30.91000	371.80000	142.00000
## IQR	68.70000	27.00000	5.84000	68.30000	26.00000
## odch. stand.	50.71384	19.92263	4.310668	50.57385	19.56861
## wariancja	2571.89402	396.91100	18.581856	2557.71400	382.93047
##	Night.Charge	Intl.Mins	Intl.Calls	Intl.Charge	CustServ.Calls
## minimum	1.04000	0.00000	0.00000	0.00000	0.00000
## Q1	7.52000	8.50000	3.00000	2.30000	1.00000
## mediana	9.05000	10.30000	4.00000	2.78000	1.00000
## średnia	9.039325	10.237294	4.479448	2.7645815	1.562856
## Q3	10.59000	12.10000	6.00000	3.27000	2.00000
## maksimum	17.77000	20.00000	20.00000	5.40000	9.00000
## rozstęp	16.73000	20.00000	20.00000	5.40000	9.00000
## IQR	3.07000	3.60000	3.00000	0.97000	1.00000
## odch. stand.	2.275873	2.791840	2.461214	0.7537726	1.315491

## wariancja	5.179597	7.794368	6.057576	0.5681732	1.730517
--------------	----------	----------	----------	-----------	----------

- W celu lepszego zestawienia cech i dokonania wyboru na które z nich powinniśmy zwrócić szczególną uwagę posłużymy się biblioteką DataExplorer.
- Pierwsze zestawienie ze sobą wszystkich zmiennych

```
plot_histogram(dane)
```



Rysunek 1: Histogramy dla wszystkich zmiennych ilościowych

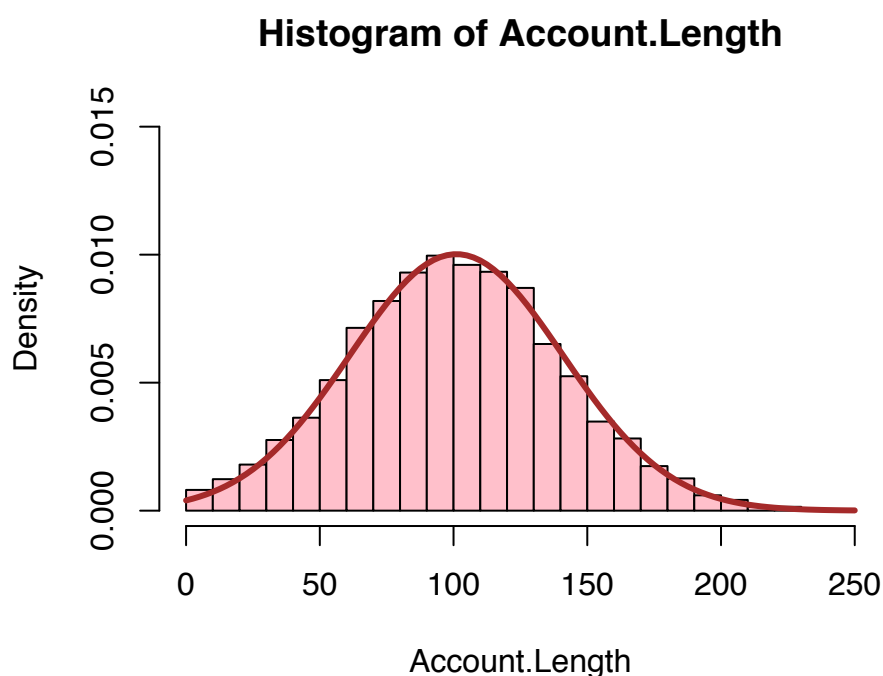
- Przypatrujemy się bliżej pierwszej zmiennej Account.Length

```
library(ggplot2)
library(MASS)
sapply(dane[c(2)] , function(x) my.summary(as.numeric(x)) )

##           Account.Length
## minimum           1.00000
## Q1                74.00000
## mediana           101.00000
## średnia           101.06481
## Q3                127.00000
```

```
## maksimum          243.00000
## rozstęp           242.00000
## IQR               53.00000
## odch. stand.      39.82211
## wariancja         1585.80012

fit2 <- fitdistr(Account.Length, densfun = "normal")
hist(Account.Length, breaks = 30, prob=TRUE, ylim=c(0,0.015), col='pink')
curve(dnorm(x, fit2$estimate[1], fit2$estimate[2]), col = "brown", add=T, lwd=3)
```



Rysunek 2: Wykres zmiennej Account.Length

```
#Jest to rozkład normalny jednomodalny
round(median(Account.Length)) == round(mean(Account.Length))

## [1] TRUE

#Symetryczny
```

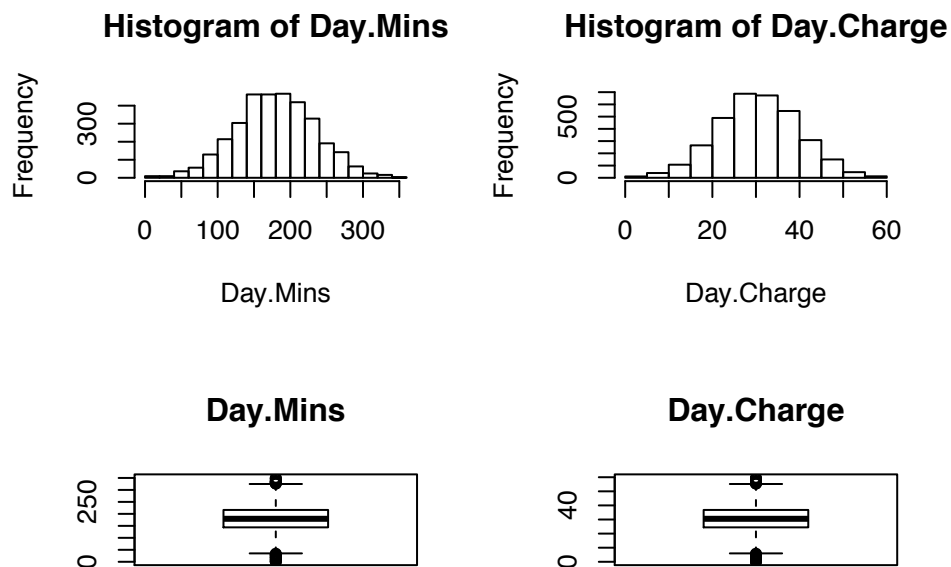
Po zbadaniu zmiennej Account.Length. wiemy, że jest ona rozkładem normalny, jednomodalnym, symetrycznym z dominantą w przedziale 80-120. Min: 1 dzień, Maks: 243 dni

- Przyjrzenie się Day.Mins i Day.Charge
- Następnie zestawienie poszczególnych cech w stosunku do pory dnia

```
sapply(dane[c(7,9)] , function(x) my.summary(as.numeric(x)) )
```

```
##           Day.Mins Day.Charge
## minimum      0.00000  0.000000
## Q1          143.70000  24.430000
## mediana      179.40000  30.500000
## średnia      179.77510  30.562307
## Q3          216.40000  36.790000
## maksimum     350.80000  59.640000
## rozstęp      350.80000  59.640000
## IQR          72.70000  12.360000
## odch. stand.  54.46739   9.259435
## wariancja    2966.69649  85.737128
```

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



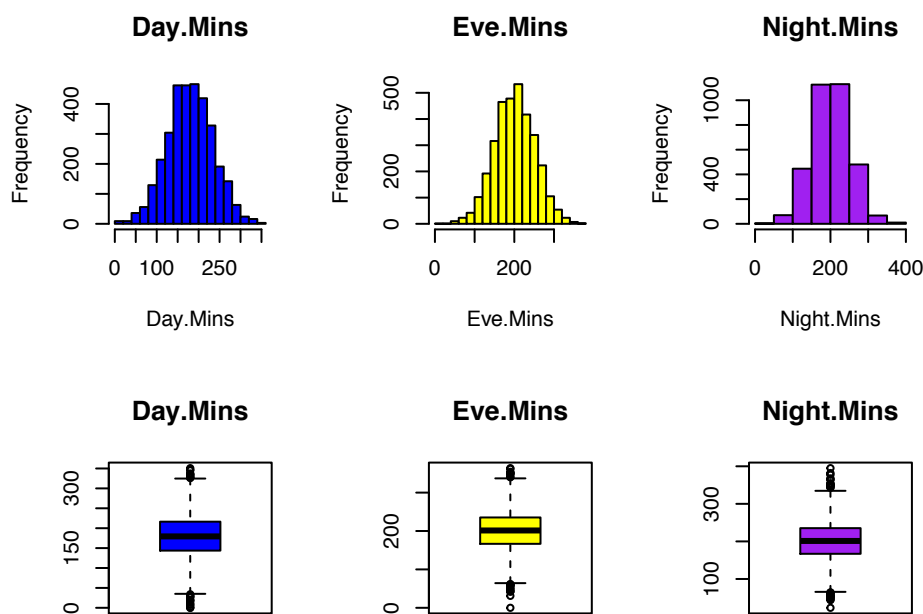
Wykresy te charakteryzują również rozkłady normalne, widzimy że tak samo zachowują się boxploty obu zmiennych.

```
sapply(dane[c(7,10,13)] , function(x) my.summary(as.numeric(x)) )
```

```
##           Day.Mins   Eve.Mins Night.Mins
## minimum      0.00000   0.00000  23.20000
## Q1          143.70000  166.60000  167.00000
```

```
## mediana      179.40000  201.40000  201.20000
## średnia      179.77510  200.98035  200.87204
## Q3           216.40000  235.30000  235.30000
## maksimum     350.80000  363.70000  395.00000
## rozstęp      350.80000  363.70000  371.80000
## IQR          72.70000   68.70000   68.30000
## odch. stand.  54.46739   50.71384   50.57385
## wariancja    2966.69649  2571.89402  2557.71400
```

```
par(mfrow=c(2,3))
hist(Day.Mins, main="Day.Mins", col="blue")
hist(Eve.Mins, main="Eve.Mins", col="yellow")
hist(Night.Mins, main="Night.Mins", col="purple")
boxplot(Day.Mins, main="Day.Mins", col="blue")
boxplot(Eve.Mins, main="Eve.Mins", col="yellow")
boxplot(Night.Mins, main="Night.Mins", col="purple")
```



Rysunek 3: Wykresy minut połączeń o różnych porach dnia

Wszystkie zmienne charakteryzują rozkłady normalne, są symetryczne, jednomodalne. Kolejno wartości przedziałów: 350.8, 363.7, 371.8 Kolejno Średnie: 180, 201, 201

```
par(mfrow=c(1,2))

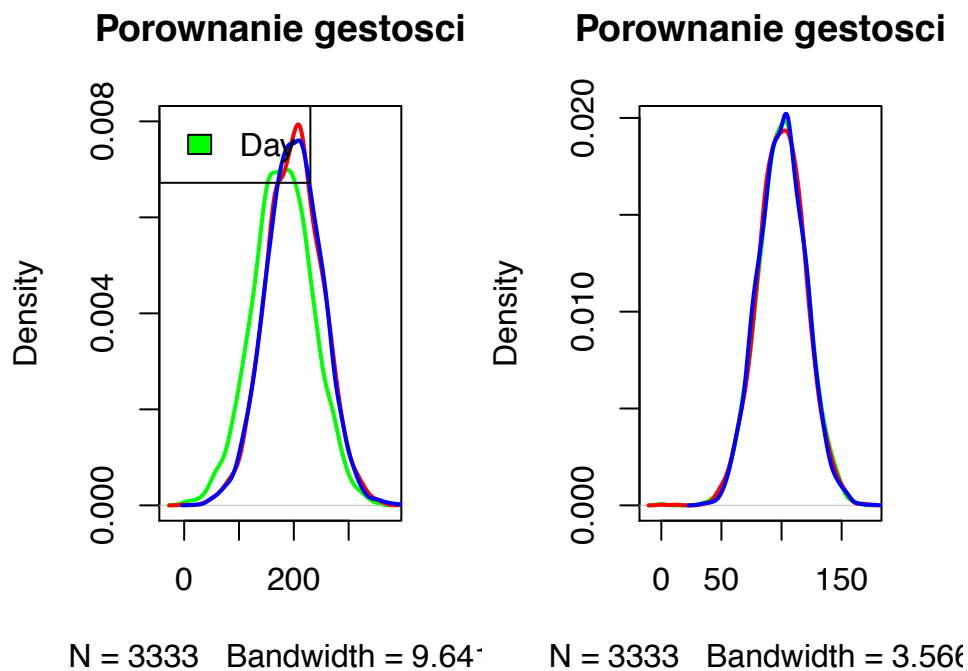
d11 <- density(Day.Mins)
d22 <- density(Eve.Mins)
```

```

d33 <- density(Night.Mins)
plot(d11, col="green", ylim=c(0,0.008), lwd=2, main="Porównanie gestosci")
lines(d22, col="red", lwd=2)
lines(d33, col="blue", lwd=2)
legend(x = "topleft", legend=c('Day'), fill=c("green"))

d1 <- density(Day.Calls)
d2 <- density(Eve.Calls)
d3 <- density(Night.Calls)
plot(d1, col="green", lwd=2, main="Porównanie gestosci")
lines(d2, col="red", lwd=2)
lines(d3, col="blue", lwd=2)

```



Rysunek 4: Porównanie gęstości

```

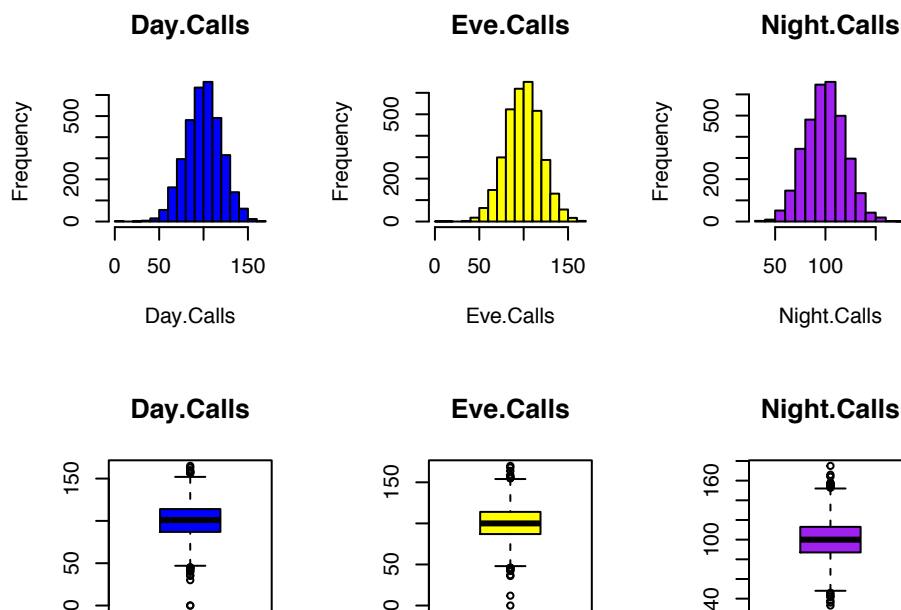
#Badamy zestawione ze sobą ilości wykonanych telefonów o różnych porach dnia
sapply(dane[c(8,11,14)] , function(x) my.summary(as.numeric(x)) )

```

	Day.Calls	Eve.Calls	Night.Calls
## minimum	0.00000	0.00000	33.00000
## Q1	87.00000	87.00000	87.00000
## mediana	101.00000	100.00000	100.00000
## średnia	100.43564	100.11431	100.10771
## Q3	114.00000	114.00000	113.00000
## maksimum	165.00000	170.00000	175.00000



## rozstęp	165.00000	170.00000	142.00000
## IQR	27.00000	27.00000	26.00000
## odch. stand.	20.06908	19.92263	19.56861
## wariancja	402.76814	396.91100	382.93047

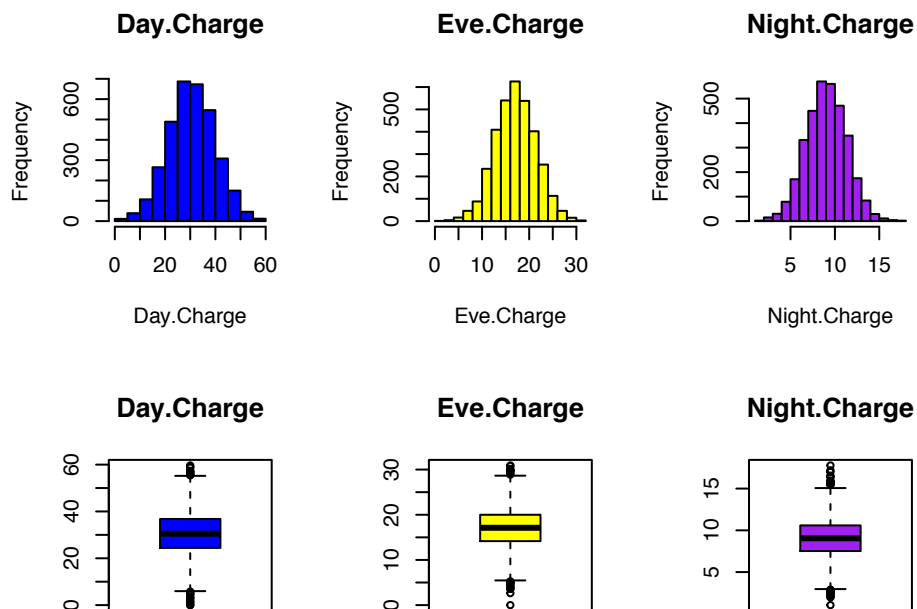


Rysunek 5: Wykresy liczby połączeń o różnych porach dnia

Wszystkie zmienne charakteryzują rozkłady normalne, są symetryczne, jednomodalne. Kolejno wartości przedziałów: 165, 170, 142 Kolejno Średnie: 100.5, 100, 100

```
#Badamy zestawione ze sobą ilości pobranych opłat o różnych porach dnia
sapply(dane[c(9,12,15)] , function(x) my.summary(as.numeric(x)) )
```

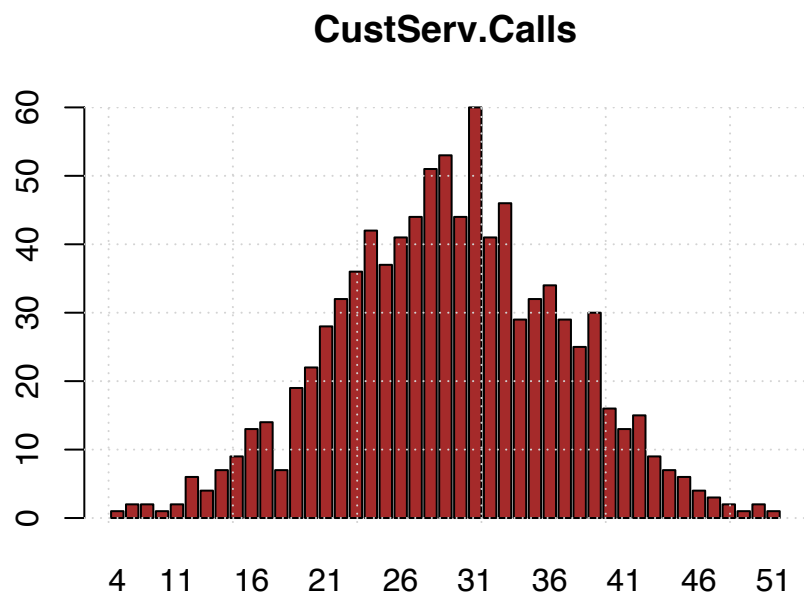
##	Day.Charge	Eve.Charge	Night.Charge
## minimum	0.000000	0.000000	1.040000
## Q1	24.430000	14.160000	7.520000
## mediana	30.500000	17.120000	9.050000
## średnia	30.562307	17.083540	9.039325
## Q3	36.790000	20.000000	10.590000
## maksimum	59.640000	30.910000	17.770000
## rozstęp	59.640000	30.910000	16.730000
## IQR	12.360000	5.840000	3.070000
## odch. stand.	9.259435	4.310668	2.275873
## wariancja	85.737128	18.581856	5.179597



Rysunek 6: Wykresy opłat połączeń o różnych porach dnia

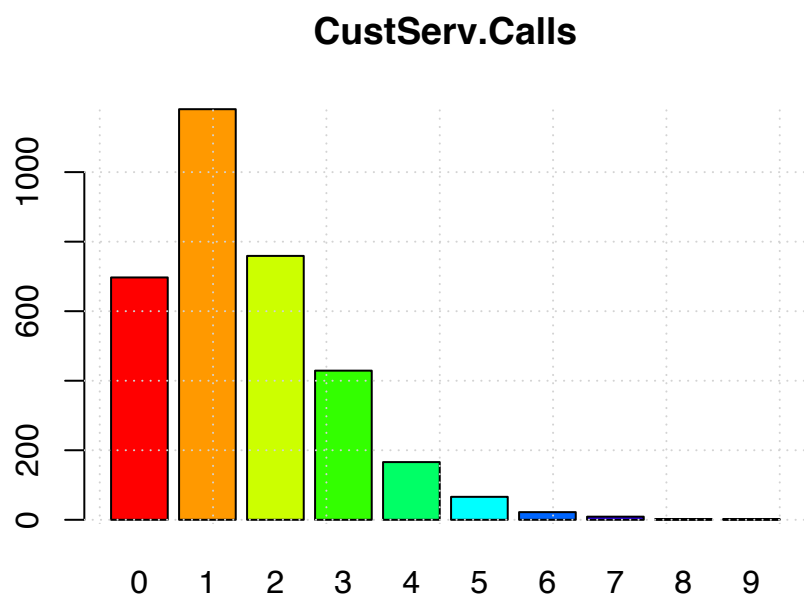
Wszystkie zmienne charakteryzują rozkłady normalne, są symetryczne, jednomodalne. Kolejno wartości przedziałów: 60, 31, 17 Kolejno Średnie: 30.5, 17, 9

```
barplot(table(VMail.Message[VMail.Plan=="yes"]), main="CustServ.Calls", col="brown")
grid()
```



Rysunek 7: Histogram ilości wiadomości głosowych

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

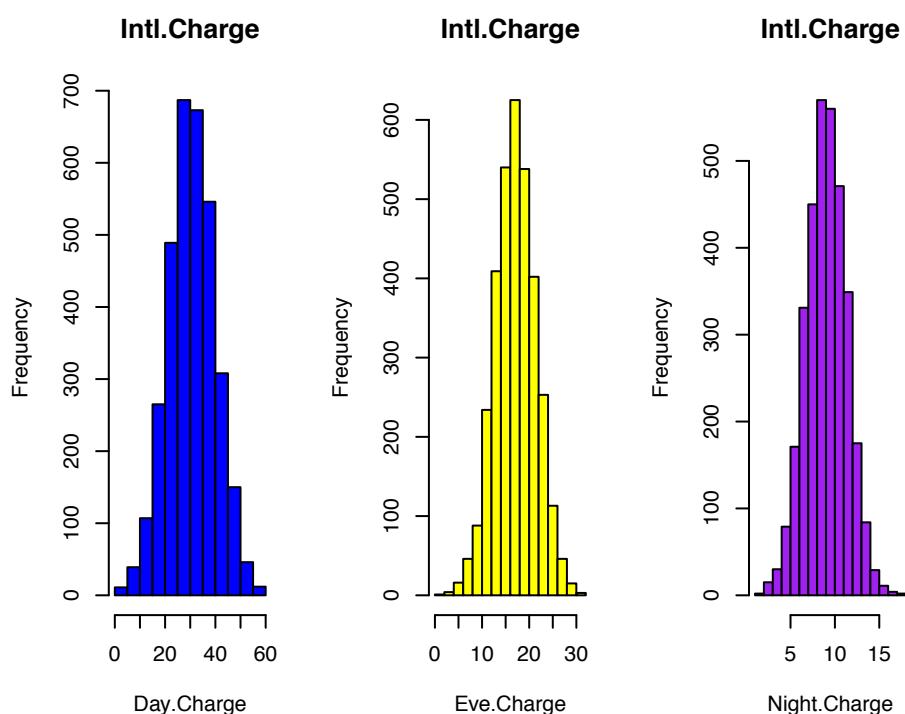


Rysunek 8: Wykres słupkowy ilości połączeń do biura obsługi klienta

```
sapply(dane[c(16,17,18)] , function(x) my.summary(as.numeric(x)) )
```

```
##          Intl.Mins Intl.Calls Intl.Charge
## minimum      0.000000  0.000000  0.000000
## Q1           8.500000  3.000000  2.300000
## mediana     10.300000  4.000000  2.780000
## średnia     10.237294  4.479448  2.7645815
## Q3          12.100000  6.000000  3.270000
## maksimum    20.000000  20.000000  5.400000
## rozstęp     20.000000  20.000000  5.400000
## IQR         3.600000  3.000000  0.970000
## odch. stand. 2.791840  2.461214  0.7537726
## wariancja    7.794368  6.057576  0.5681732
```

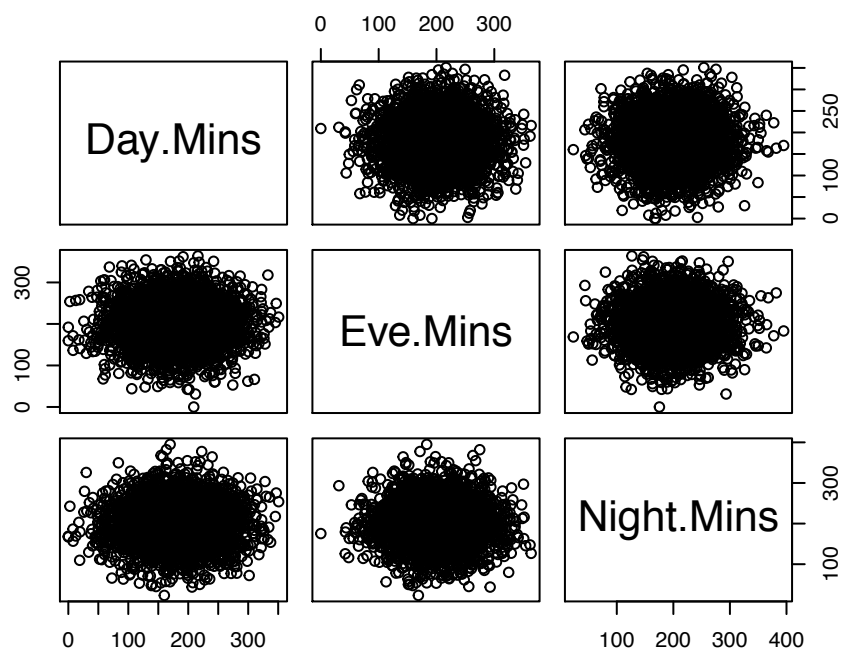
```
par(mfrow=c(1,3))
hist(Day.Charge, main="Intl.Charge", col="blue")
hist(Eve.Charge, main="Intl.Charge", col="yellow")
hist(Night.Charge, main="Intl.Charge", col="purple")
```



Rysunek 9: Histogramy planu międzynarodowego

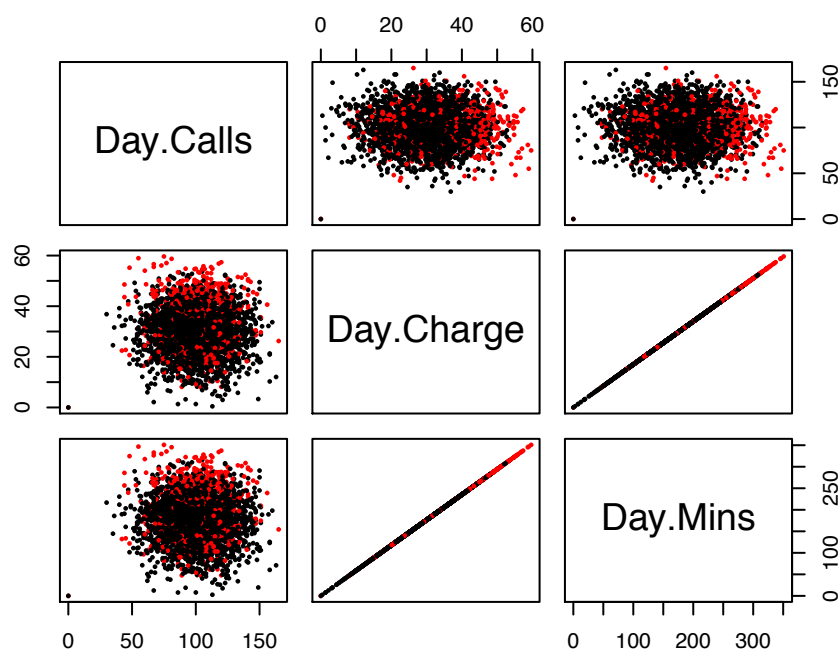
- Wykresy rozrzutu dla par zmiennych ciągłych

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



Rysunek 10: Korelacje między cechami charakteryzującymi rozmowy podczas dnia

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



Rysunek 11: Korelacje między cechami charakteryzującymi rozmowy podczas dnia

```
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
```

Współczynniki korelacji między cechami charakteryzującymi rozmowy podczas dnia

### 3.2 Interpretacja wyników zmiennych ilościowych

- Zakres możliwych wartości
  1. Account.Length: 1-243
  2. Vmail.Message: 0-51
  3. Day.Mins: 0-360
  4. Day.Calls: 0-165
  5. Day.Charge: 0-60
  6. Eve.Mins: 0-364
  7. Eve.Calls: 0-170
  8. Eve.Charge: 0-31
  9. Night.Mins: 23-395
  10. Night.Calls: 33-175

11. Night.Charge: 1-18
12. Intl.Mins: 0-20
13. Intl.Calls: 0-20
14. Intl.Charge: 0-5
15. CustServ.Calls: 0-9

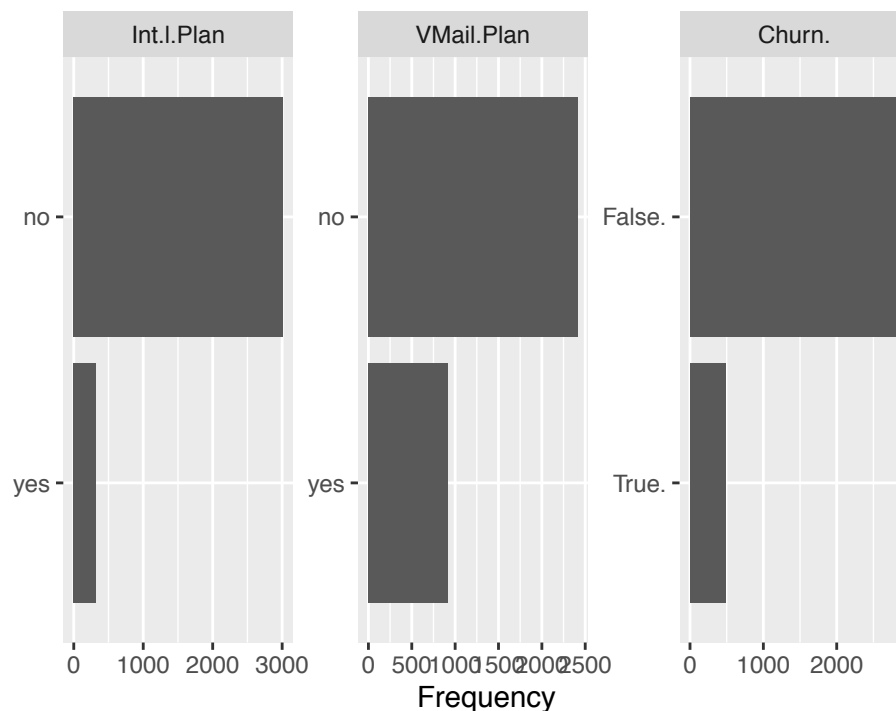
- Rozkłady zmiennych Account.Length, Day.Calls, Day.Charge, Day.Mins, Eve.Calls, Eve.Charge, Eve.Mins, Intl.Charge, Intl.Mins, Night.Calls, Night.Charge oraz Night.Mins charakteryzują się rozkładami normalnymi i są symetryczne (wynika to z przyrównania średniej i mediany)
- Największe zmienności możemy zauważyć przy zmiennych ilości minut wykonanych połączeń o różnych porach dnia oraz długości konta. (wynika to z odchylenia standardowego)

### 3.3 Analiza danych jakościowych

- Wyznaczymy podstawowe wskaźniki sumaryczne (miary położenia i rozproszenia) w celu skupiania się na tych cechach, z których otrzymamy porządkane przez nas informacje.
- Następnie zajmujemy się dokładniejszą wizualizacją danych

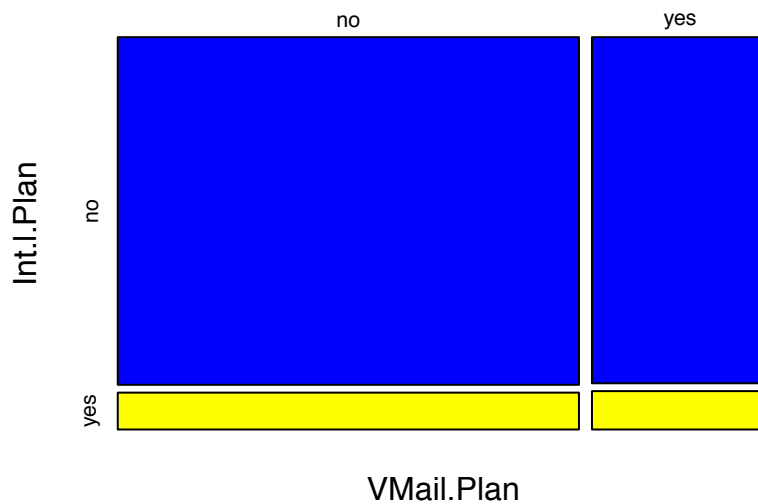
```
# wykresy słupkowe dla wszystkich zmiennych jakościowych
plot_bar(dane)

## 1 columns ignored with more than 50 categories.
## State: 51 categories
```



Rysunek 12: Wykresy cech jakościowych

```
library(descr)
par(mfrow=c(1,1))
crosstab(Int.l.Plan, VMail.Plan, col=c("blue","yellow", prop.t = TRUE))
```



Rysunek 13: Wykres zestawienia klientów z planem międzynarodowym i poczty głosowej

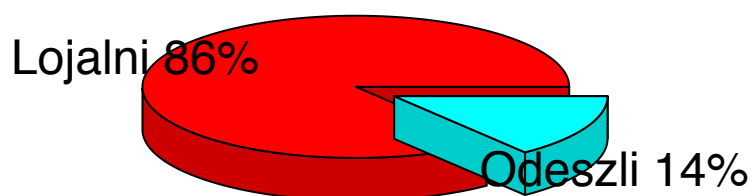
```
##      Cell Contents
## |-----|
## |                      Count |
## |-----|
##
## =====
##              VMail.Plan
## Int.l.Plan   no   yes   Total
## -----
## no           2180  830   3010
## -----
## yes           231   92    323
## -----
## Total        2411  922   3333
## =====
```

```
library(plotrix)
lbls <- c("Lojalni", "Odeszli")
pct <- round(table(Churn.)/sum(table(Churn.))*100)
```



```
lbls <- paste(lbls, pct) # add percents to labels
lbls <- paste(lbls,"%",sep="") # ad % to labels
pie3D(table(Churn.),labels=lbls,explode=0.1,
      main="Wykres lojalnosci klientow")
```

## Wykres lojalnosci klientow



```
table(State)

## State
##  AK  AL  AR  AZ  CA  CO  CT  DC  DE  FL  GA  HI  IA  ID  IL  IN  KS  KY  LA  MA
##  52  80  55  64  34  66  74  54  61  63  54  53  44  73  58  71  70  59  51  65
##  MD  ME  MI  MN  MO  MS  MT  NC  ND  NE  NH  NJ  NM  NV  NY  OH  OK  OR  PA  RI
##  70  62  73  84  63  65  68  68  62  61  56  68  62  66  83  78  61  78  45  65
##  SC  SD  TN  TX  UT  VA  VT  WA  WI  WV  WY
##  60  60  53  72  72  77  73  66  78  106  77

max(table(State))

## [1] 106

min(table(State))

## [1] 34

max(table(State)) - min(table(State))

## [1] 72
```

### 3.4 Interpretacja wyników zmiennych jakościowych

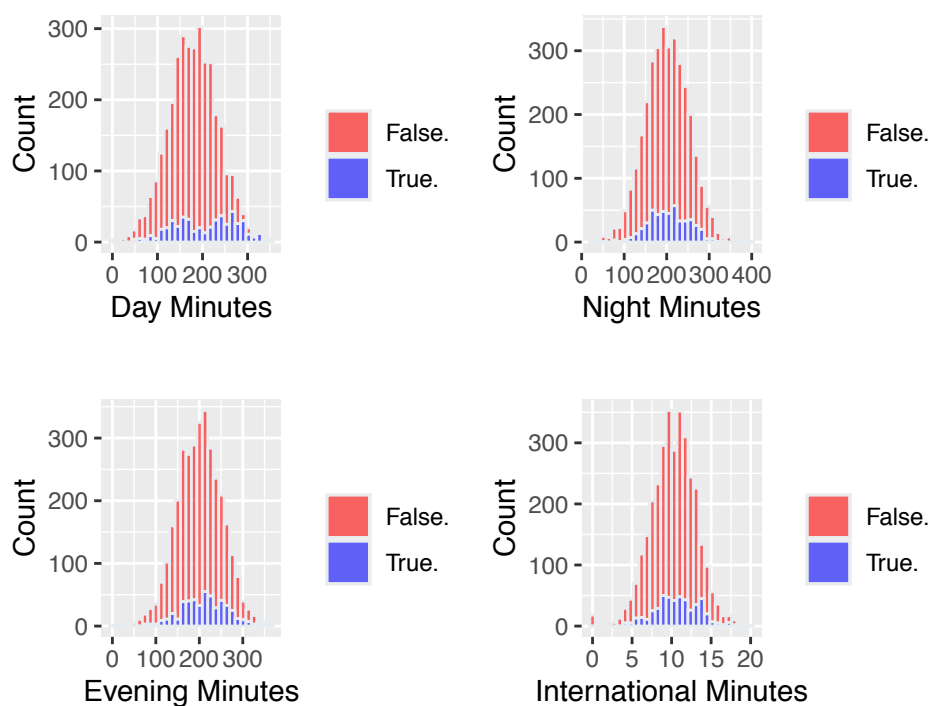
- W przypadku planów jakie klienci mogli wykupić to najwięcej z nich nie zdecydowało się na jakikolwiek z nich – 65.4 procent. Drugą podgrupę stanowili Ci co zasięgneli po plan Vmail.Plan, jednak bez Intl.l.Plan – 24.9 procent.
- 14 na 100 klientów postanowiło odejść z sieci
- Jeśli chodzi o podział na Stany to maksymalna ilość klientów z danego Stanu to 106, minimalna 34. Rozbieżności mogą tu być przypadkowe.

## 4 Etap 3. Analiza skupiająca się na podgrupach Churn.

### 4.1 Analiza zmiennych ilościowych

```
##  
## Attaching package: 'dplyr'  
## The following object is masked from 'package:MASS':  
##  
## select  
## The following objects are masked from 'package:stats':  
##  
## filter, lag  
## The following objects are masked from 'package:base':  
##  
## intersect, setdiff, setequal, union  
##  
## Attaching package: 'gridExtra'  
## The following object is masked from 'package:dplyr':  
##  
## combine
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.  
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.  
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.  
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



Rysunek 14: Histogramy liczby minut o różnych porach dnia oraz połączeń międzynarodowych.

Użytkownicy o dłuższych rozmowach w dzień częściej opuszczają operatora, zaczyna się to powyżej 200 minut. Wśród rozmów podczas wieczorów i w nocy nie widac silnego powiązania odchodnia klientów z zależności od liczby minut. Rozmów międzynarodowych jest około 20 razy mniej, brak znaczących zależności.

```
# Zmienna Calls
p5 <- ggplot( ) +
  geom_histogram(data = dane, aes(x=Day.Calls, fill=Churn.), color="#e9ecef", alpha=0.6,
  scale_fill_manual(values=c("red", "blue"))) +
  labs( x = "Day Calls", y = "Count",title = "") +
  labs(fill="")

p6 <- ggplot( ) +
  geom_histogram(data = dane, aes(x=Night.Calls, fill=Churn.), color="#e9ecef", alpha=0.6,
  scale_fill_manual(values=c("red", "blue"))) +
  labs( x = "Night Calls", y = "Count",title = "") +
  labs(fill="")

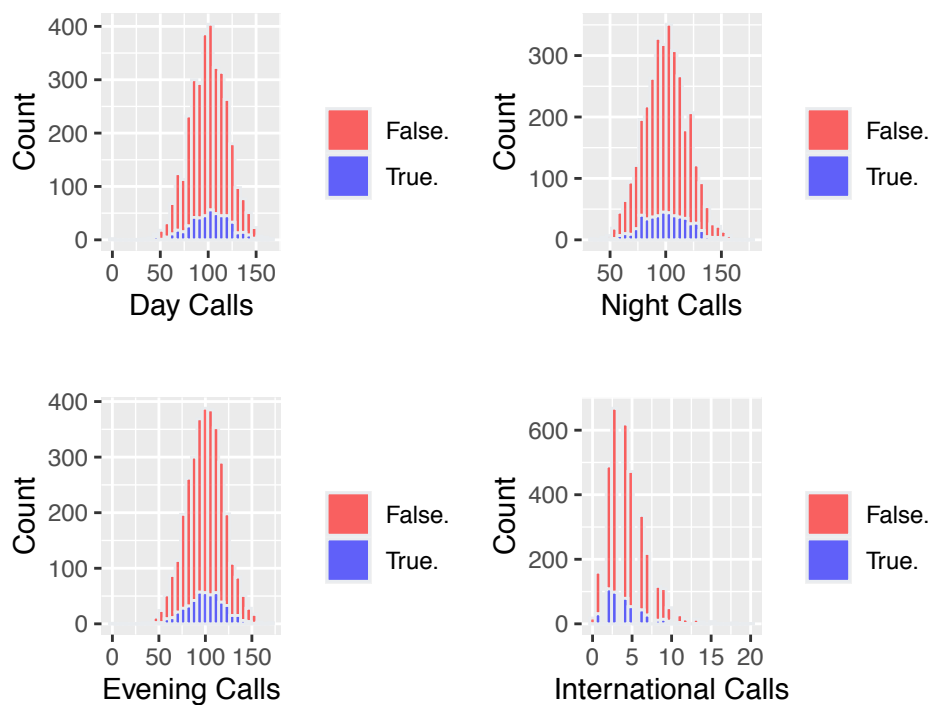
p7 <- ggplot( ) +
  geom_histogram(data = dane, aes(x=Eve.Calls, fill=Churn.), color="#e9ecef", alpha=0.6,
  scale_fill_manual(values=c("red", "blue"))) +
  labs( x = "Evening Calls", y = "Count",title = "") +
  labs(fill="")

p8 <- ggplot( ) +
```

```
geom_histogram(data = dane, aes(x=Intl.Calls, fill=Churn.), color="#e9ecef", alpha=0.6,
scale_fill_manual(values=c("red", "blue"))) +
labs( x = "International Calls", y = "Count",title = "") +
labs(fill="")
```

```
PC <- grid.arrange(p5, p6,p7, p8, ncol=2)
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



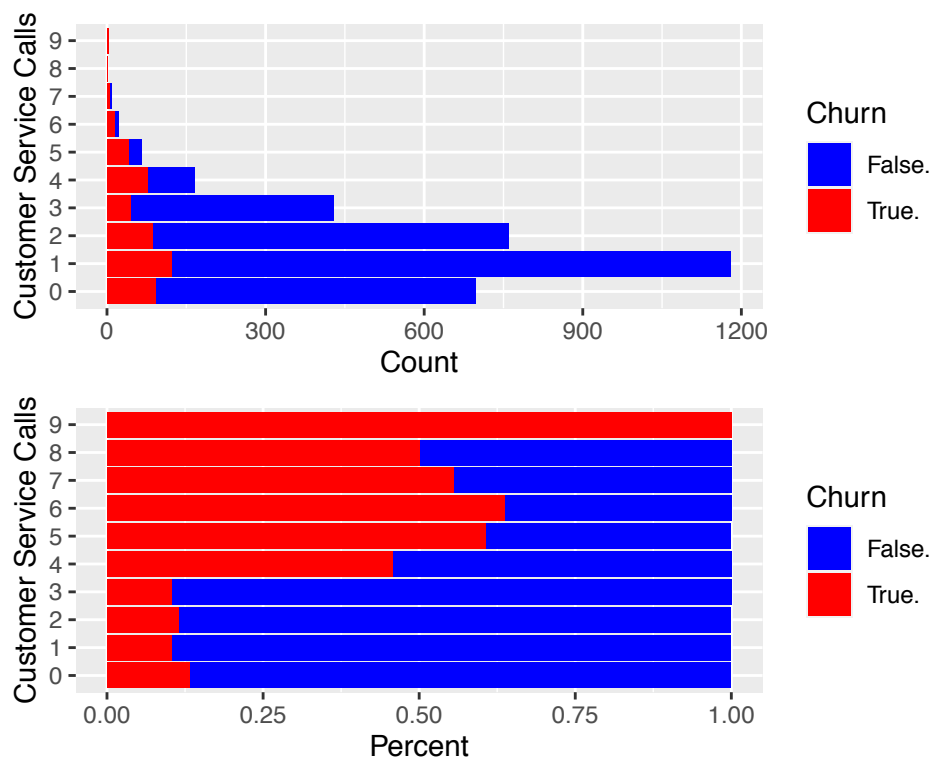
Rysunek 15: Histogramy liczby połączeń o różnych porach dnia oraz międzynarodowych.

```
# Zmienna liczby połączeń z biurem obsługi klienta CustServ.Calls
# wykres słupkowy o częstości liczby połączeń z biurem obsługi klienta
CSC1 <- ggplot() +
  geom_bar(data = dane,
    aes(x = factor(CustServ.Calls),
      fill = factor(Churn.)),
    position = "stack") +
  scale_x_discrete("Customer Service Calls") +
  scale_y_continuous("Count") +
  guides(fill=guide_legend(title="Churn")) +
  scale_fill_manual(values=c("blue", "red"))+
```

```
coord_flip()

# wykres słupkowy w skali procentowej dla liczby połączeń z biurem obsługi klienta
CSC2 <- ggplot() +
  geom_bar(data=dane,
    aes(x = factor(CustServ.Calls),
        fill = factor(Churn.)),
    position = "fill") +
  scale_x_discrete("Customer Service Calls") +
  scale_y_continuous("Percent") +
  guides(fill=guide_legend(title="Churn")) +
  scale_fill_manual(values=c("blue", "red"))+
  coord_flip()

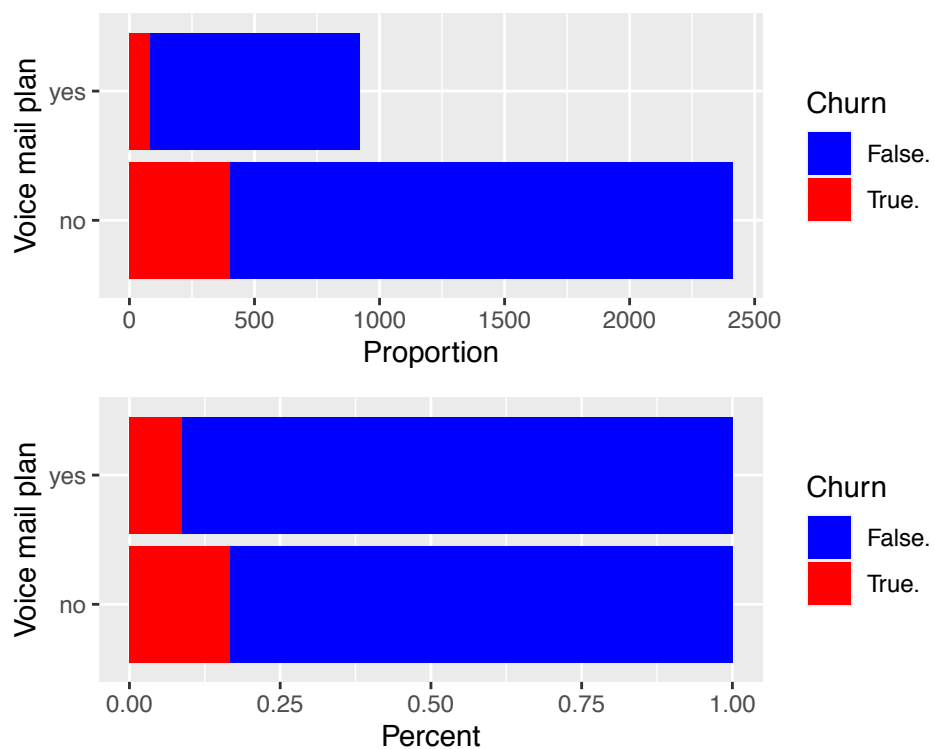
CSC <- grid.arrange(CSC1, CSC2, ncol=1)
```



Rysunek 16: Liczba połączeń serwisu klienta.

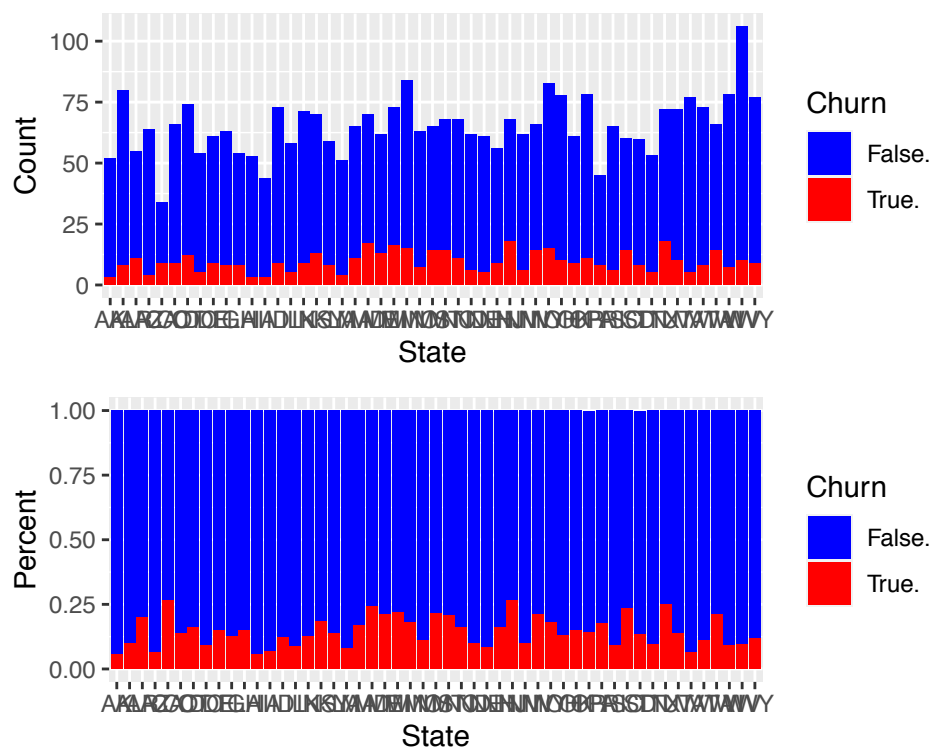
Z wykresu widac ze kategoria osób odchodzących jest duża dla powyżej trzech połączeń. Zmienna wykazuje duże powiązanie z odejściem klienta.

## 4.2 Analiza zmiennych jakościowych Churn.



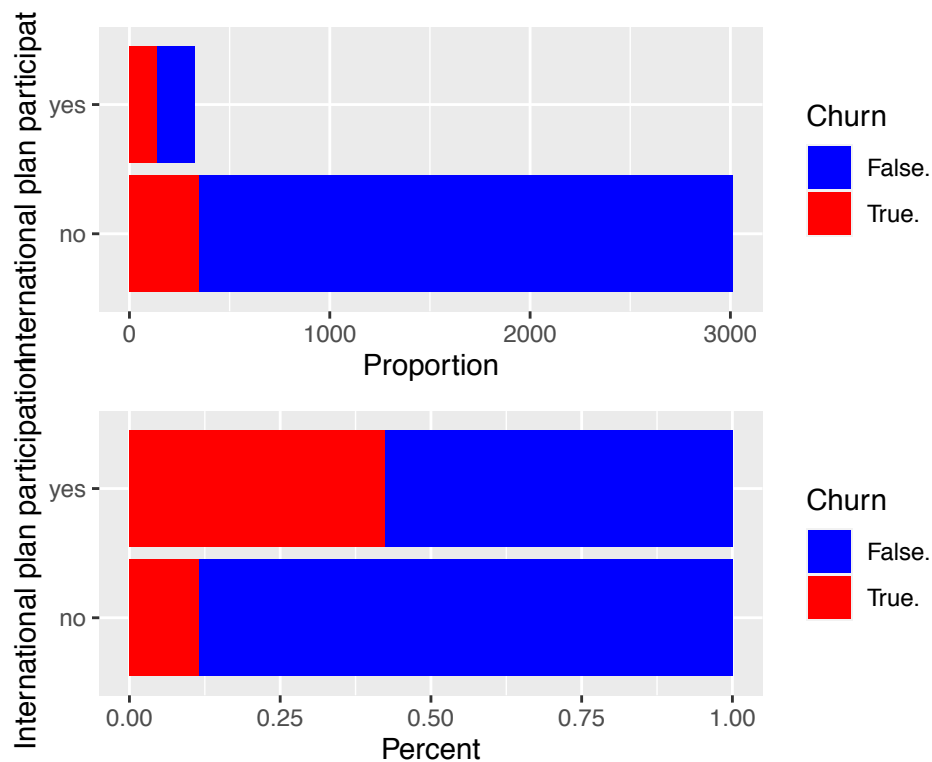
Rysunek 17: Uczestnictwo klientów w planie poczty głosowej

Częściej rezygnują osoby, które nie uczestniczą w kontakcie mailowym.



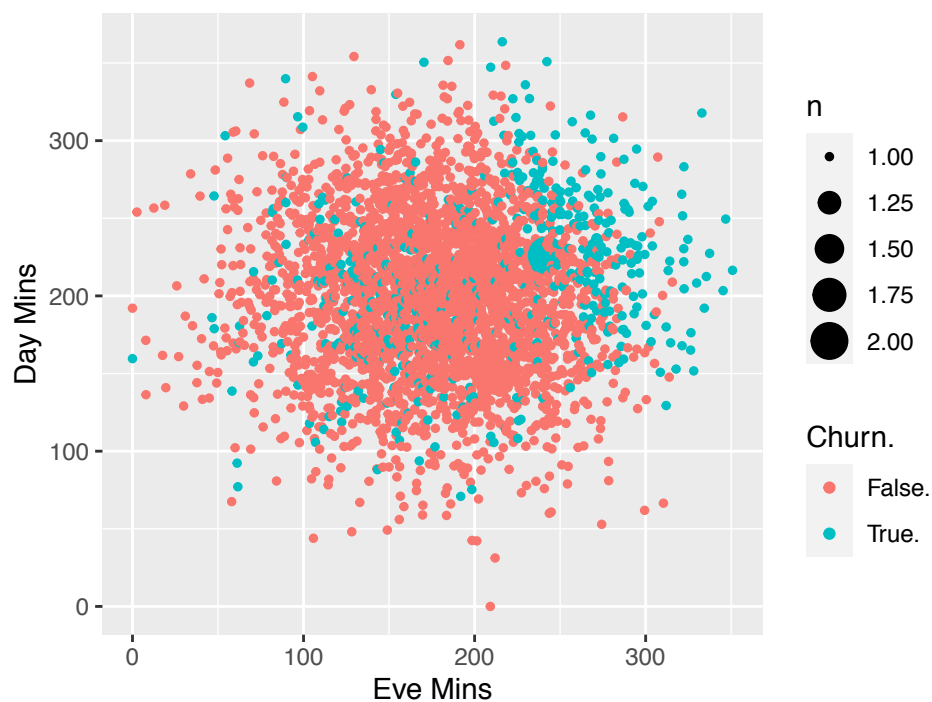
Rysunek 18: Rezygancje klientów w zależności od stanu.

Odejścia w poszczególnych stanach nie są mocno zróżnicowane, lecz charakteryzują się zmiennością. Na podstawie wizualizacji można wybrać stany o największej liczbie rezygnacji.



Rysunek 19: Uczestnictwo klientów w planie międzynarodowym

Osoby biorą udział w planie międzynarodowym częściej opuszczają operatora względem klientów bez tego planu.



Klienci z dużą liczbą minut w dzień jak i wieczorem charakteryzują się większym ryzykiem odejścia.

## **5 Etap 4. Podsumowanie**

### **5.1 Ogólne wnioski**

- Klienci posiadają większą ilość (wydzwonionych) minut o porach wieczornych i nocnych niż w porze dziennej. Zakres wartości też ma tendencje zwykłe. Pokazują nam to parametry położenia oraz wykres gęstości. Natomiast ilość połączeń jest o każdej porze bardzo bliska sobie
- Bardzo wielu klientów nie zdecydowało się na wykupienie żadnego z planów/pakietów, jedynie 9.7 procent z nich skorzystało z planu międzynarodowego
- Istnieje silna korelacja między opłatami a wydzwonionymi minutami odpowiednio o każdej porze dnia
- Jeśli chodzi o zmienną CustServ.Calls dowiadujemy się że najczęstszą liczbą wykonanych połączeń do biura obsługi klienta jest 1. Aż 75 procent klientów wykonuje 2 lub mniej połączeń, przy średniej długości konta 101 dni?
- Występują jedynie 3 numery kierunkowe wśród klientów

### **5.2 Wnioski dotyczące podgrup Churn.**

- Klienci z planem międzynarodowym i planem poczty głosowej częściej opuszczają operatora
- Klienci z powyżej czterech połączeniami z serwisem klienta rezygnują na podobnym poziomie
- Klienci z dużą liczbą minut w dzień częściej rezygnują niż pozostali, zjawisko takie nie występuje jednak wśród minut o innych porach dnia
- Osoby z planem poczty głosowej rzadziej opuszczają operatora

### **5.3 Ewentualne rozwiązania, działania**

- Można obserwować klientów którzy przekraczają liczbę 200 minut i oferować im specjalne bonusy, promocje, pakiety
- Powinno się przeanalizować dlaczego w tym momencie klienci rezygnują, może na rynku jest operator oferujący promocje
- Powinno się monitorować liczbę połączeń wykonanych pomiędzy serwisem a każdym klientem
- Można rozważyć wprowadzenie ofert specjalnych podczas połączeń z klientami jeśli miało to miejsce więcej niż trzy razy ponieważ prawdopodobieństwo odejścia wzrasta znacznie
- Powinno się zbadać dlaczego klienci z planem międzynarodowym częściej odchodzą



- Powinno się zwiększyć uwagę i rozwinąć program planu mailowego lub ułatwić dostęp do dołączenia, może to zwiększyć lojalność klientów
- Łatwiejszy dostęp dołączenia do planu Poczty głosowej- może to zwiększyć lojalność klientów

## Literatura

- [1] dr inż. Adam Zagdański, [http://prac.im.pwr.wroc.pl/~zagdan/polish\\_ver/ED2020/index.html](http://prac.im.pwr.wroc.pl/~zagdan/polish_ver/ED2020/index.html), 2020.