

# Sprawozdanie 1

Jakub Markowiak  
album 255705

31 marca 2021

## Spis treści

1	Krótki opis zagadnienia	1
2	Opis eksperymentów/analiz	1
3	Wyniki	1
4	Podsumowanie	7

## 1 Krótki opis zagadnienia

W tym sprawozdaniu przeprowadzimy analizę opisową danych `churn`, zawierających informacje o klientach pewnej sieci komórkowej. Zajmiemy się również interpretacją otrzymanych wyników oraz spróbujemy scharakteryzować grupę klientów, którzy zrezygnowali z usług tej sieci.

## 2 Opis eksperymentów/analiz

Przeprowadzimy następujące analizy i eksperymenty:

1. analiza opisowa danych `churn` - zastosowanie wskaźników sumarycznych, wykresów oraz zbadanie zależności między zmiennymi,
2. analiza opisowa danych `churn` z podziałem na grupy.

## 3 Wyniki

Rozpoczynamy od wczytania danych z pliku `churn.txt` oraz zliczenia liczby obserwacji, liczby cech jakościowych i ilościowych oraz sprawdzenia, czy w danych występują brakujące wartości. Ponieważ cecha `Phone` służy jako identyfikator, usuwamy ją przed dalszą analizą. Zmienna `Area.Code` została odczytana jako zmienna ilościowa, zatem konwertujemy ją do typu `"factor"`.

Liczba cech	Liczba obserwacji	Cechy ilościowe	Cechy jakościowe	Brakujące wartości
20	3333	15	5	0

Tabela 1: Wstępne spojrzenie na dane

Widzimy, że w danych nie występują brakujące obserwacje. Następnie wyznaczymy podstawowe statystyki opisowe dla cech `Day.Mins`, `Eve.Mins`, `Night.Mins` oraz `CustServ.Calls`. Zdefiniujemy funkcję `my.summary` działającą analogicznie do domyślnego `summary`.

```
#Zdefiniowanie funkcji wyznaczającej podstawowe wskaźniki sumaryczne
my.summary <- function(x)
{
  wskaźniki <- c(
    Srednia = mean(x, na.rm = T),
    Mediana = median(x, na.rm = T),
    IQR = IQR(x, na.rm = T),
    Min = min(x, na.rm = T),
    Maks = max(x, na.rm = T),
    Odch.stand. = sd(x, na.rm = T),
    Rozstep = max(x, na.rm = T) - min(x, na.rm = T)
  )
  return(wskaźniki)
}
```

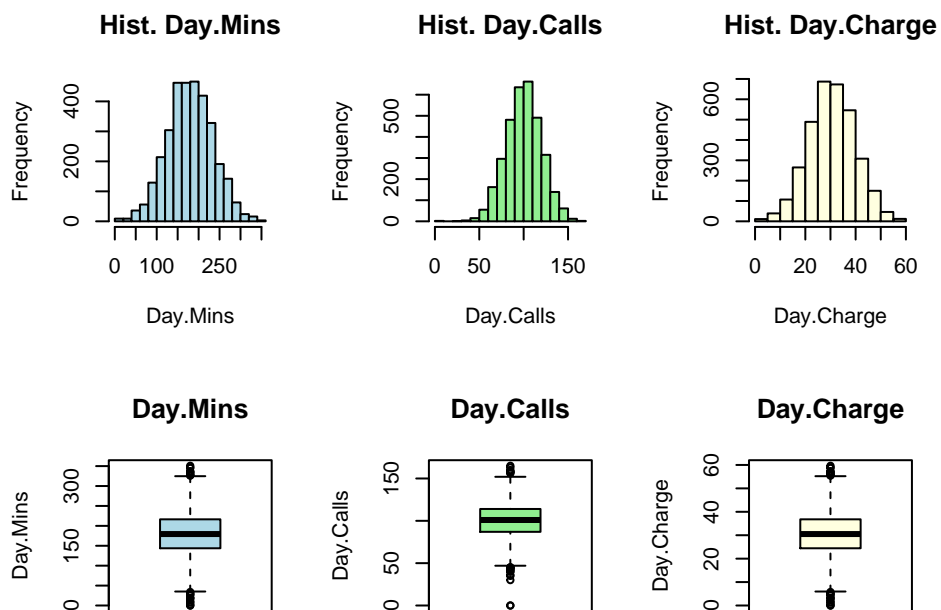
Wykorzystując tę funkcję sporządzamy tabelę zawierającą podstawowe wskaźniki sumaryczne.

Tabela 2: Podstawowe wskaźniki sumaryczne

	Srednia	Mediana	IQR	Min	Maks	Odch.stand.	Rozstep
Day.Mins	179.78	179.4	72.70	0	350.80	54.47	350.80
Day.Calls	100.44	101.0	27.00	0	165.00	20.07	165.00
Day.Charge	30.56	30.5	12.36	0	59.64	9.26	59.64
CustServ.Calls	1.56	1.0	1.00	0	9.00	1.32	9.00

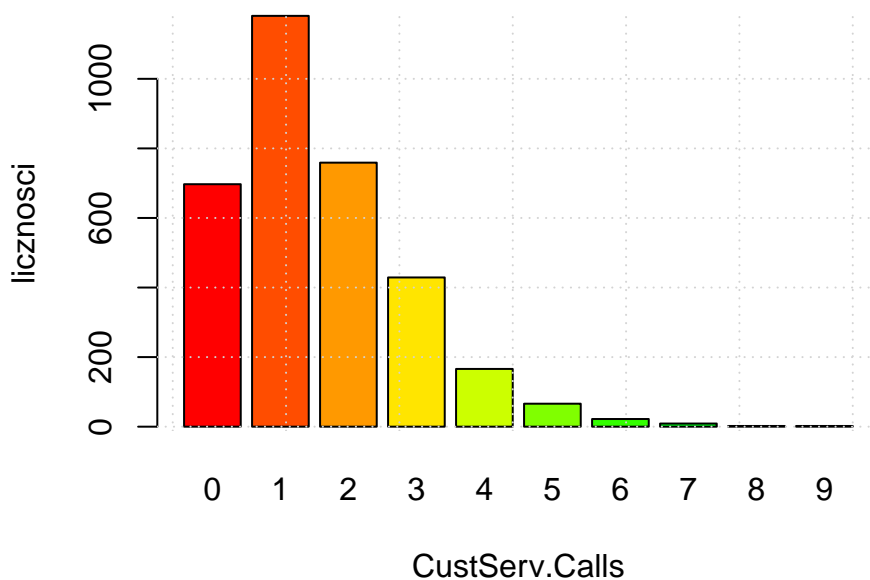
Możemy zauważyć, że zmienna `Day.Mins` przyjmuje wartości z zakresu  $[0; 350, 8]$ , a przedział typowych wartości to  $[125, 31; 234, 25]$ . Podobnie dla `Day.Calls` przedziałem typowych wartości jest  $[80, 37; 120, 51]$ .

Teraz wykonamy histogramy oraz wykresy pudełkowe dla tych zmiennych. Widzimy, że te zmienne mają rozkłady zbliżone do symetrycznych oraz kształtem przypominają gęstość rozkładu normalnego. Możemy też dostrzec, że zmienne te przyjmują wartości z szerokiego zakresu. Największą zmiennością charakteryzują się zmienne `Day.Mins` oraz `Day.Charge`.



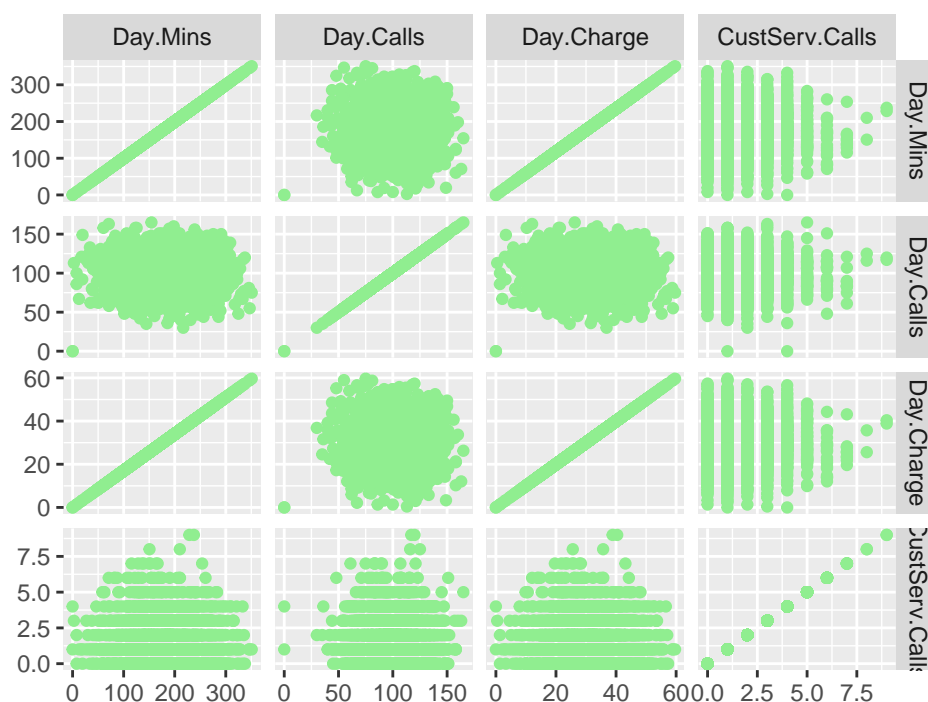
Spoglądając natomiast na diagram słupkowy cechy `CustServ.Calls` obserwujemy, że rozkład tej zmiennej jest prawostronnie skośny.

**Diagram słupkowy `CustServ.Calls`**

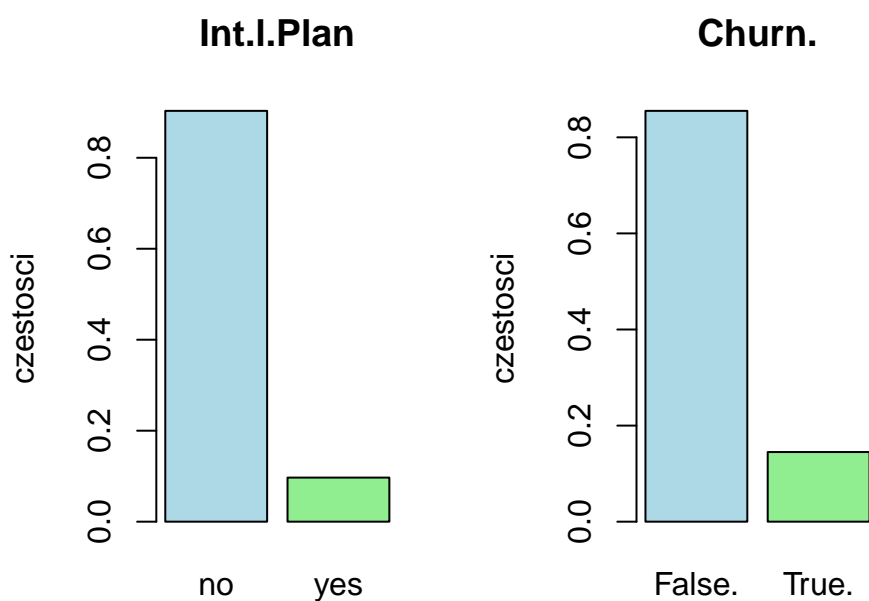


Wykonamy teraz wykresy rozrzutu dla badanych cech. Widzimy, że istnieje liniowa zależność między długością prowadzenia rozmów, a opłatą za korzystanie z usług.

## Wykresy rozrzutu dla wybranych cech



Sprawdzimy również, jak często klienci korzystają z planu międzynarodowego oraz rezygnują z usług firmy.



Około 10% zbadanych klientów korzysta z planu międzynarodowego, a około 17% osób zrezygnowało z usług tej firmy.

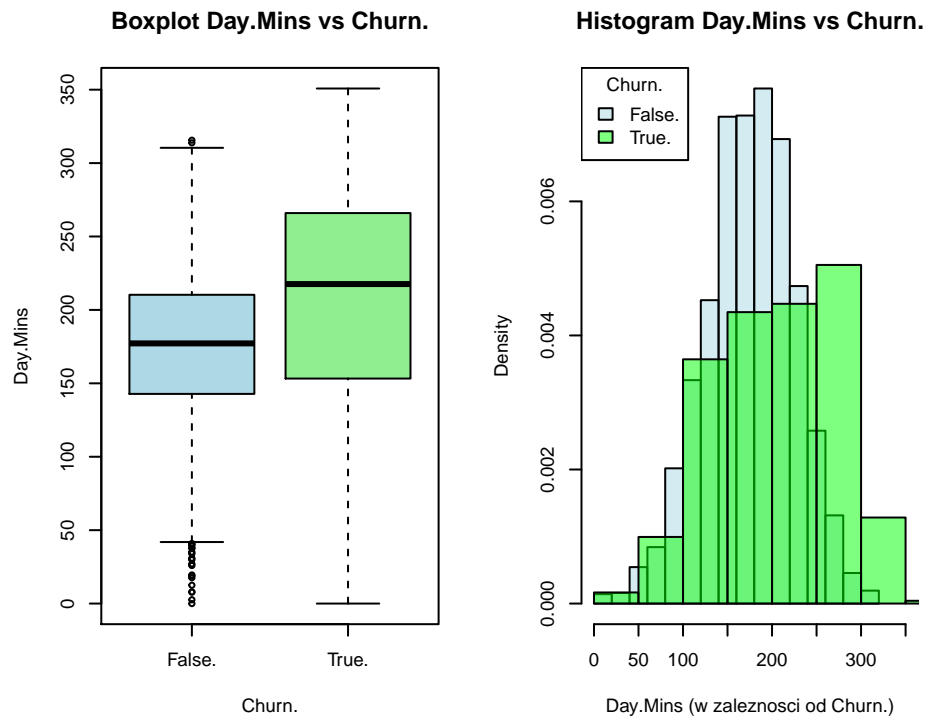
Zajmiemy się teraz analizą opisową z podziałem na dwie grupy - klientów, którzy odeszli oraz tych, którzy wciąż korzystają z usług firmy.

Wstępnie zauważamy, że osoby, które zrezygnowały z usług firmy częściej rozmawiały przez telefon w ciągu dnia.

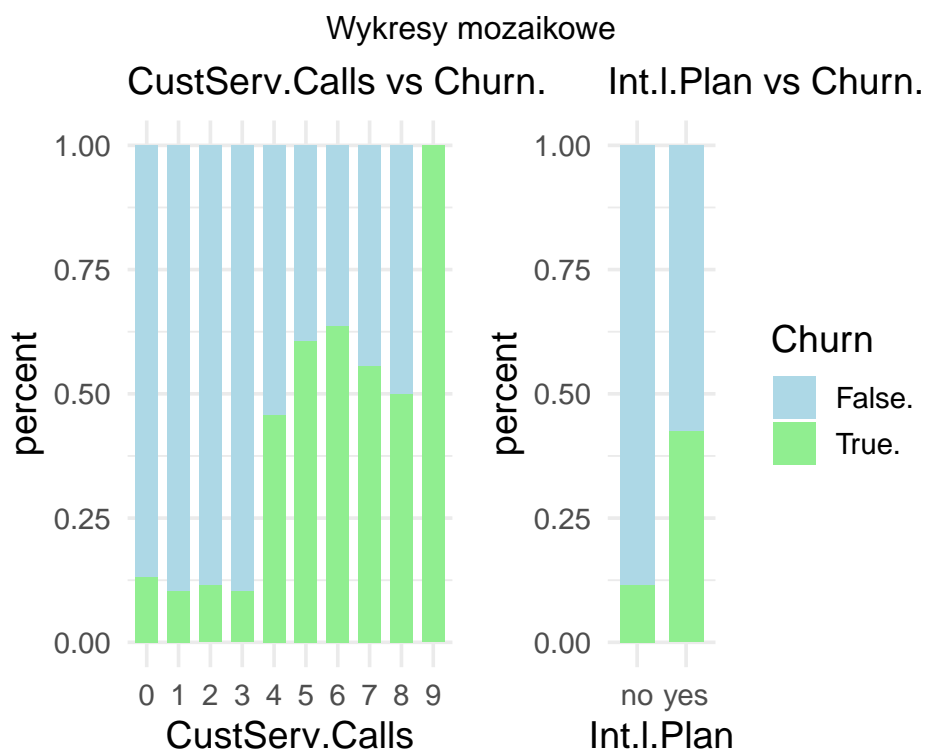
Tabela 3: Porównanie wskaźników summarycznych cechy Day.Mins dla dwóch grup klientów

	Srednia	Mediana	IQR	Min	Maks	Odch.stand.	Rozstep
Zostal	175.18	177.2	67.47	0	315.6	50.18	315.6
Zrezygnowal	206.91	217.6	112.70	0	350.8	69.00	350.8

Wykonamy wykresy pudełkowe dla Day.Mins oraz CustServ.Calls z podziałem na te dwie grupy.

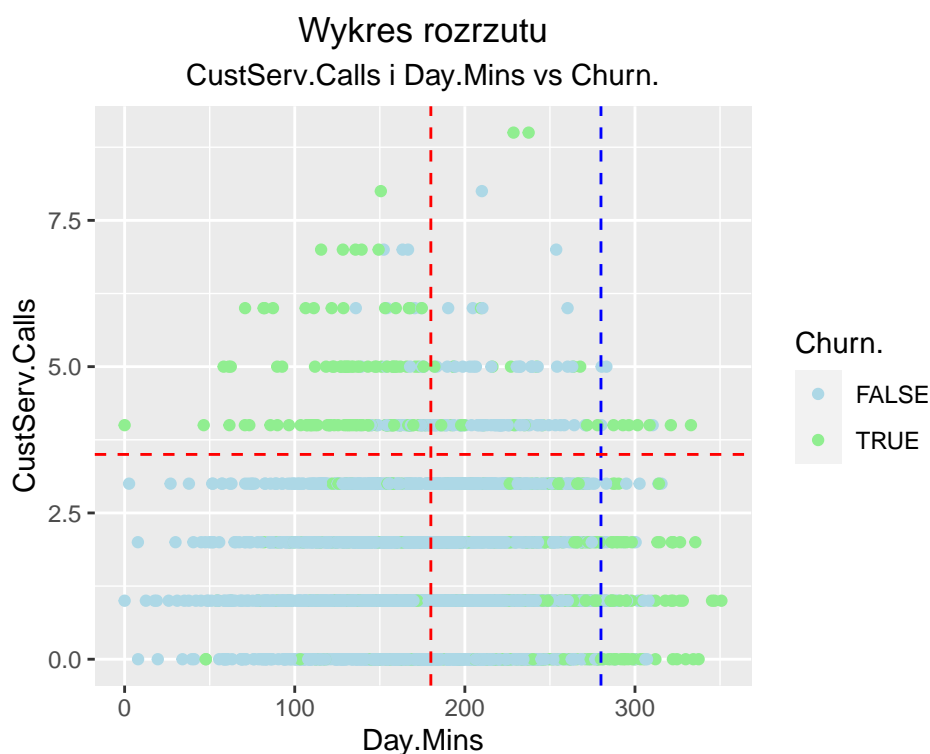


Narysujemy również wykres mozaikowy dla cechy CustServ.Calls oraz wykres częstości dla Int.1.Plan.



Osoby, które kontaktowały się z biurem obsługi klienta cztery lub więcej razy rezygnowały z usług w około 50% przypadków. Spośród osób, które korzystały z planu międzynarodowego około 45% zrezygnowało.

Widzimy zatem, że cechami, które dobrze mogą scharakteryzować osoby rezygnujące z usług, są Day.Mins, Int.l.Plan oraz CustServ.Calls. Wykonamy zatem wykres rozrzutu dla Day.Mins i CustServ.Calls z zaznaczeniem klientów rezygnujących.



Osoby, które rozmawiają w ciągu dnia krócej, ale często kontaktują się z biurem obsługi klienta oraz osoby, które rozmawiają dłużej, najczęściej rezygnują z usług firmy. Przeanalizujemy

zatem te dwie grupy. Pierwsza grupa to klienci, którzy rozmawiają dzień przez mniej niż 180 minut i skontaktowali się z biurem obsługi klienta co najmniej 4 razy. Druga grupa to klienci, którzy rozmawiali przez ponad 260 minut. Zauważamy, że w obu grupach znajduje się około 52% wszystkich przypadków rezygnacji.

Tabela 4: Podział na dwie grupy

	Day.Mins $\leq$ 180, CustServ.Calls $\geq$ 4	Day.Mins $\geq$ 260
zrezygnowało (%)	74.65753	55.02008

Widzimy, że w prawdopodobieństwo rezygnacji klientów z grupy 1 to aż 75%, a w grupie 2 wynosi ono 55%.

## 4 Podsumowanie

Poniżej wypunktujemy najważniejsze wnioski, jakie można wyciągnąć z przeprowadzanych analiz:

- wyznaczenie wskaźników sumarycznych oraz narysowanie podstawowych wykresów pozwala nam wstępnie zaobserwować zależności lub ich brak, aby następnie dogłębnie przeanalizować ewentualnie związki między cechami,
- klienci tej firmy najczęściej korzystają z rozmów w ciągu dnia, rzadko są zainteresowani planem międzynarodowym oraz rzadko kontaktują się z biurem obsługi klienta,
- klienci tej firmy nie są zadowoleni z kontaktów z biurem obsługi klienta, ale niezadowoleni są również najczęściej rozmawiający w ciągu dnia klienci oraz klienci korzystający z planu międzynarodowego,
- aby zmniejszyć prawdopodobieństwo odejścia klienta z firmy można poprawić jakość usług świadczonych przez biuro obsługi klienta, zaproponować zniżkę osobom korzystającym z usług przez ponad 260 minut w ciągu dnia oraz poprawić działanie planu międzynarodowego.