

Raport nr 1

Emilia Kowal [249716], Jakub Dworzański [249703]

5 kwietnia 2020

Spis treści

1	Krótki opis zagadnienia	1
2	Opis eksperymentów/analiz	1
3	Wyniki	1
3.1	Przygotowanie danych. Podstawowe informacje o danych	1
3.2	Analiza opisowa — wskaźniki sumaryczne i wykresy	3
3.3	Analiza opisowa z podziałem na grupy	10
4	Podsumowanie	16

1 Krótki opis zagadnienia

Zajmiemy się analizą opisową zbioru danych, zawierającego informacje o klientach operatora telekomunikacyjnego. Na podstawie odkrytych zależności, postaramy się, przede wszystkim, odpowiedzieć na trzy pytania:

- Które elementy obecnej oferty dla klientów można poprawić?
- Z jakich powodów klienci opuszczają naszą firmę?
- Jak można ich zachęcić do pozostania przy naszej ofercie?

2 Opis eksperymentów/analiz

Eksperymenty będą polegały na analizie sumarycznych wskaźników i wykresów. Przeprowadzimy takie analizy dla poszczególnych cech oraz dla par zmiennych. Ponadto, wykonamy je z podziałem klientów na dwie grupy:

- lojalnych, którzy wciąż korzystają z naszych usług,
- nielojalnych, którzy zrezygnowali z naszych usług,

aby móc wyciągnąć prawidłowe wnioski.

Tabela 1: Opis danych

Zmienna	Zmienna jakościowa	Pierwsze wartości
State	TRUE	KS, OH, NJ, OH
Account.Length	FALSE	128, 107, 137, 84
Area.Code	FALSE	415, 415, 415, 408
Phone	TRUE	382-4657, 371-7191, 358-1921, 375-9999
Int.l.Plan	TRUE	no, no, no, yes
VMail.Plan	TRUE	yes, yes, no, no
VMail.Message	FALSE	25, 26, 0, 0
Day.Mins	FALSE	265.1, 161.6, 243.4, 299.4
Day.Calls	FALSE	110, 123, 114, 71
Day.Charge	FALSE	45.07, 27.47, 41.38, 50.9
Eve.Mins	FALSE	197.4, 195.5, 121.2, 61.9
Eve.Calls	FALSE	99, 103, 110, 88
Eve.Charge	FALSE	16.78, 16.62, 10.3, 5.26
Night.Mins	FALSE	244.7, 254.4, 162.6, 196.9
Night.Calls	FALSE	91, 103, 104, 89
Night.Charge	FALSE	11.01, 11.45, 7.32, 8.86
Intl.Mins	FALSE	10, 13.7, 12.2, 6.6
Intl.Calls	FALSE	3, 3, 5, 7
Intl.Charge	FALSE	2.7, 3.7, 3.29, 1.78
CustServ.Calls	FALSE	1, 1, 0, 2
Churn.	TRUE	False., False., False., False.

3 Wyniki

3.1 Przygotowanie danych. Podstawowe informacje o danych

```
dane <- read.csv(file="churn.txt")
```

Opracowywane dane zawierają informacje o 3333 klientach sieci telefonii komórkowej oraz 21 zmiennych charakteryzujących klientów. Po wczytaniu zbioru danych do przestrzeni roboczej, sprawdzamy poprawność poszczególnych typów zmiennych oraz rozmiar danych.

W tabeli 1 widzimy, że zmienna **Area.Code**, zgodnie z opisem, powinna być zmienną jakościową, dlatego zmieniamy jej typ na **factor**.

```
dane$Area.Code <- as.factor(dane$Area.Code)
```

Ponieważ reszta zmiennych została odczytana poprawnie, możemy przeanalizować przydatność zmiennych. Zmienna **Phone** stanowi id klienta i nie będzie kluczowa w dalszej analizie, dlatego ją usuniemy.

```
dane <- dane[, -4]
```

Wyszukujemy brakujących wartości w zbiorze danych:

Tabela 2: Wskaźniki sumaryczne

	średnia	odchylenie standardowe	min	max
Account.Length	101.06	39.82	1.00	243.00
VMail.Message	8.10	13.69	0.00	51.00
Day.Mins	179.78	54.47	0.00	350.80
Day.Calls	100.44	20.07	0.00	165.00
Day.Charge	30.56	9.26	0.00	59.64
Eve.Mins	200.98	50.71	0.00	363.70
Eve.Calls	100.11	19.92	0.00	170.00
Eve.Charge	17.08	4.31	0.00	30.91
Night.Mins	200.87	50.57	23.20	395.00
Night.Calls	100.11	19.57	33.00	175.00
Night.Charge	9.04	2.28	1.04	17.77
Intl.Mins	10.24	2.79	0.00	20.00
Intl.Calls	4.48	2.46	0.00	20.00
Intl.Charge	2.76	0.75	0.00	5.40
CustServ.Calls	1.56	1.32	0.00	9.00

```
sum(is.na(dane))
```

```
## [1] 0
```

Widzimy, że w danych `churn.txt` nie ma wartości brakujących.

3.2 Analiza opisowa — wskaźniki sumaryczne i wykresy

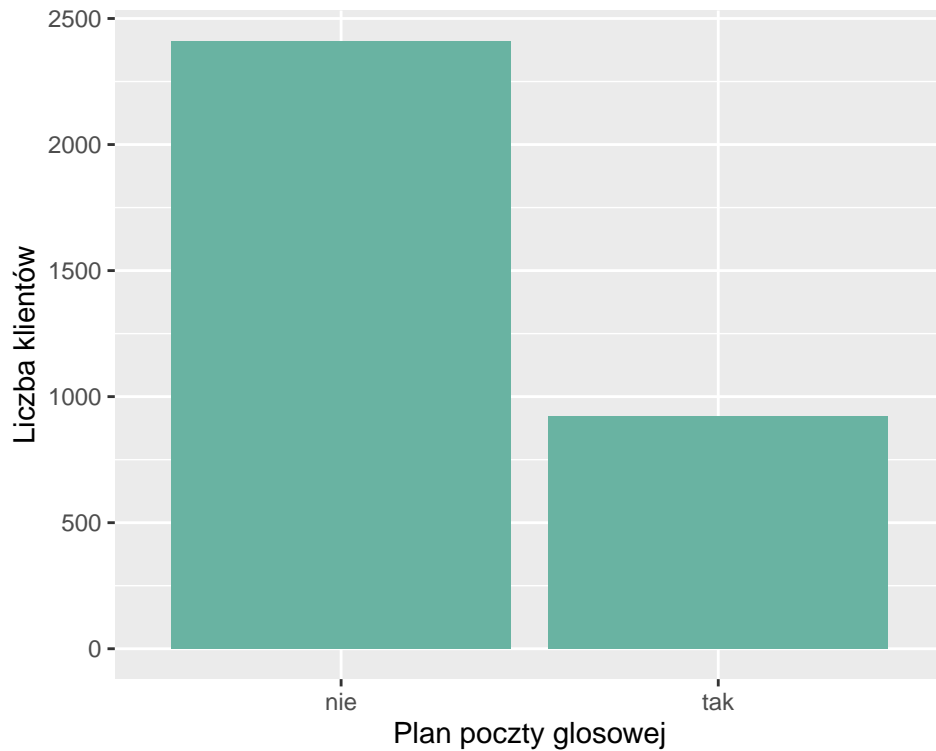
Rozkład poszczególnych zmiennych

Na początek wyznaczamy podstawowe wskaźniki sumaryczne dla wszystkich zmiennych numerycznych, które możemy zobaczyć w tabelach 2 i 3.

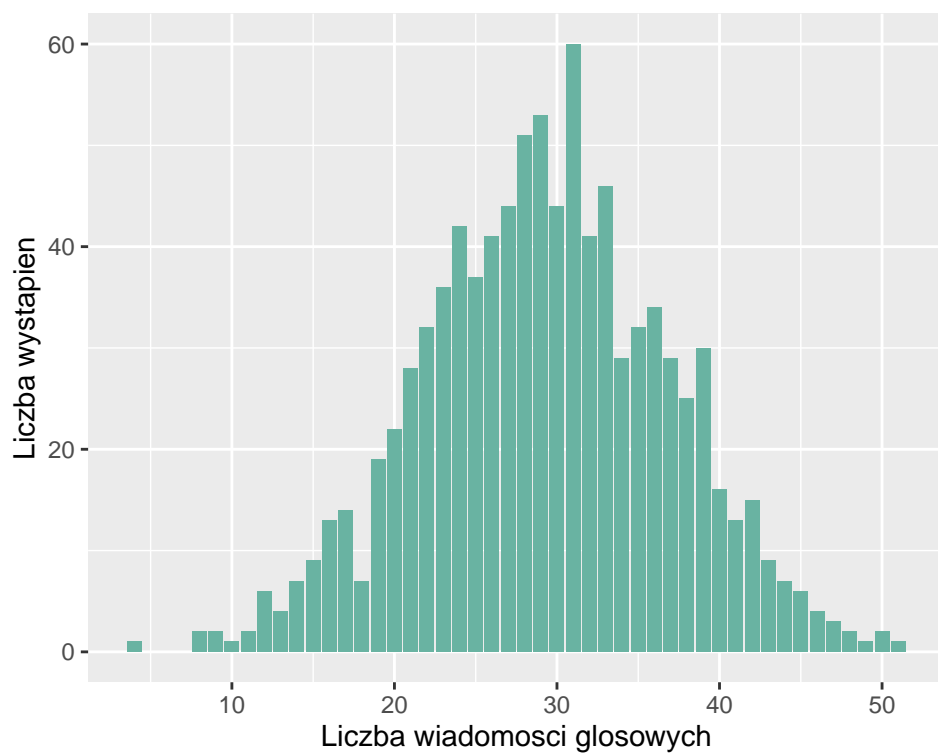
Aby lepiej zbadać poszczególne cechy, sporządzimy wykresy ich rozkładów w badanym zbiorze danych.

Tabela 3: Wskaźniki sumaryczne c.d.

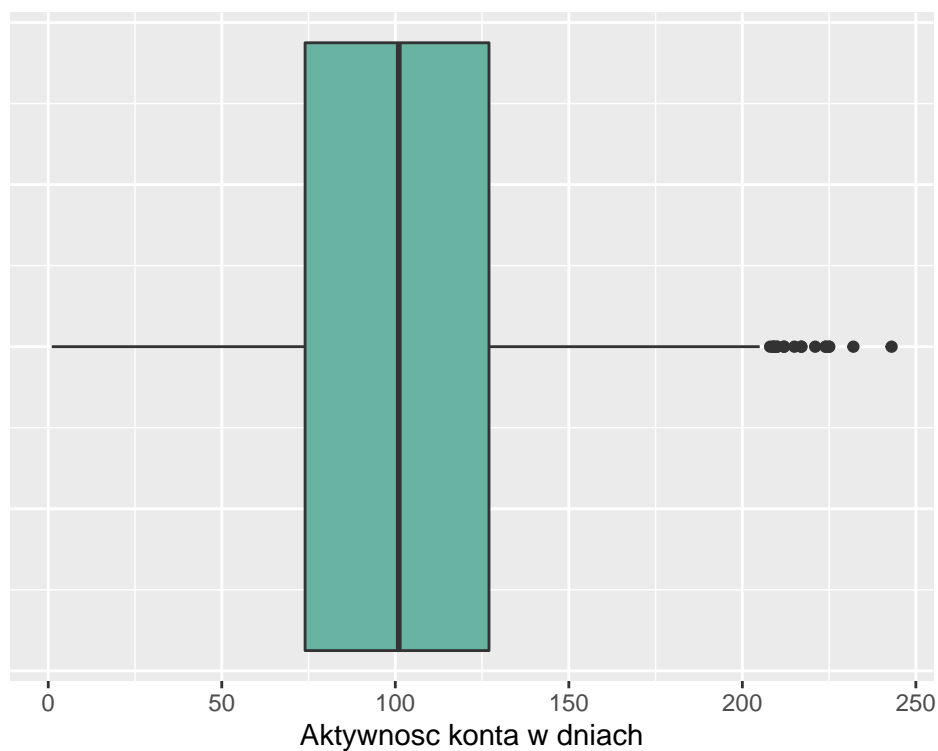
	1. kwartył	mediana	3. kwartył	IQR	skośność	kurtoza
Account.Length	74.00	101.00	127.00	53.00	0.10	-0.11
VMail.Message	0.00	0.00	20.00	20.00	1.26	-0.05
Day.Mins	143.70	179.40	216.40	72.70	-0.03	-0.02
Day.Calls	87.00	101.00	114.00	27.00	-0.11	0.24
Day.Charge	24.43	30.50	36.79	12.36	-0.03	-0.02
Eve.Mins	166.60	201.40	235.30	68.70	-0.02	0.02
Eve.Calls	87.00	100.00	114.00	27.00	-0.06	0.20
Eve.Charge	14.16	17.12	20.00	5.84	-0.02	0.02
Night.Mins	167.00	201.20	235.30	68.30	0.01	0.08
Night.Calls	87.00	100.00	113.00	26.00	0.03	-0.08
Night.Charge	7.52	9.05	10.59	3.07	0.01	0.08
Intl.Mins	8.50	10.30	12.10	3.60	-0.24	0.60
Intl.Calls	3.00	4.00	6.00	3.00	1.32	3.07
Intl.Charge	2.30	2.78	3.27	0.97	-0.25	0.60
CustServ.Calls	1.00	1.00	2.00	1.00	1.09	1.72



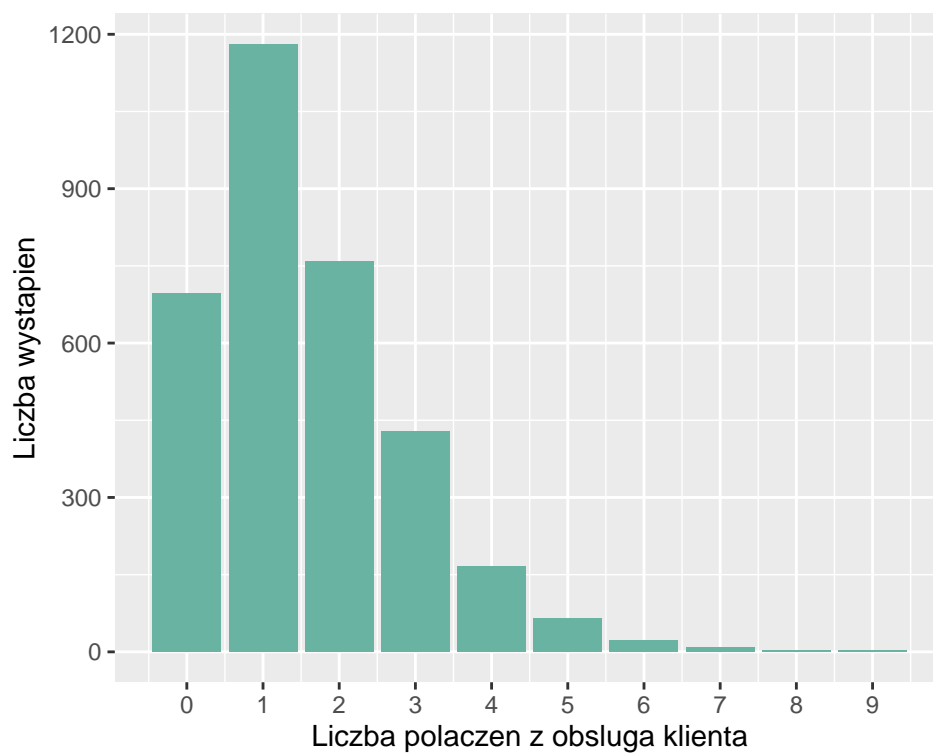
Rysunek 1: Wykres słupkowy dla planu pocztu głosowej



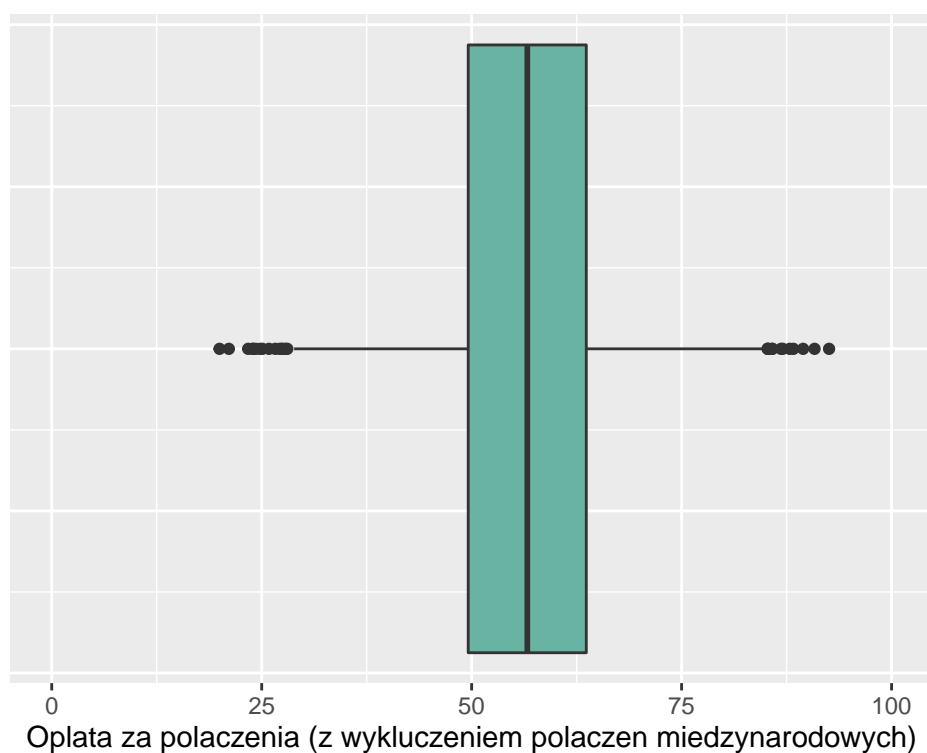
Rysunek 2: Histogram liczby wiadomości głosowych dla klientów korzystających poczty głosowej



Rysunek 3: Wykres pudełkowy długości aktywności konta



Rysunek 4: Histogram liczby połączeń z obsługą klienta



Rysunek 5: Wykres pudełkowy dla opłat za połączenia

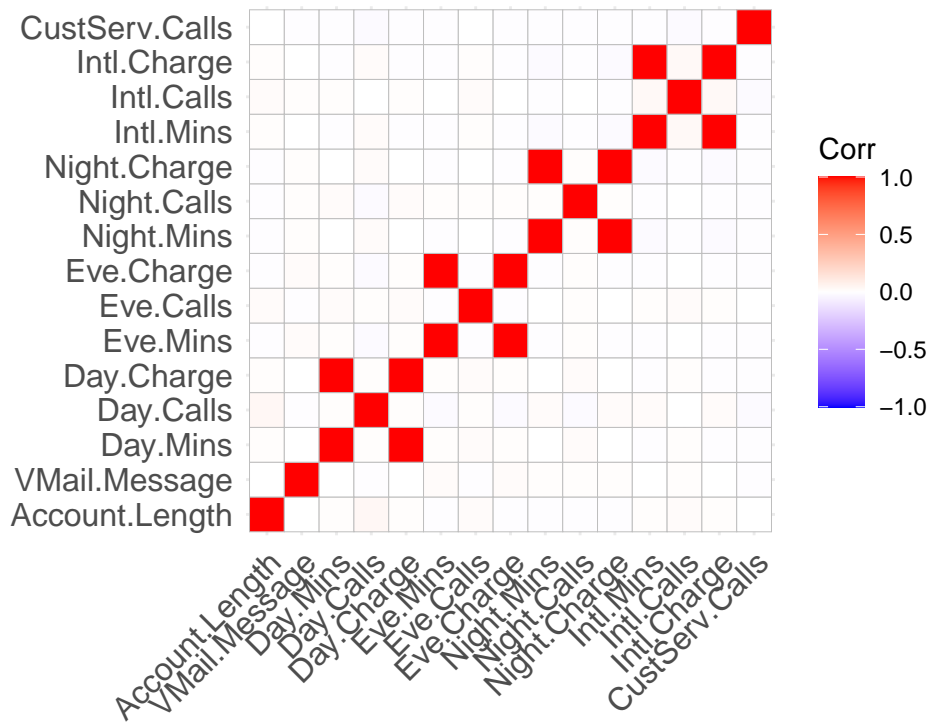
Na wykresie 1 można zauważyć, że rozkład nie jest symetryczny. Ponieważ plan międzynarodowy również charakteryzuje się dość istotną asymetrią, to możemy wnioskować, iż klienci

niechętnie korzystają z dodatkowych usług oferowanych przez firmę.

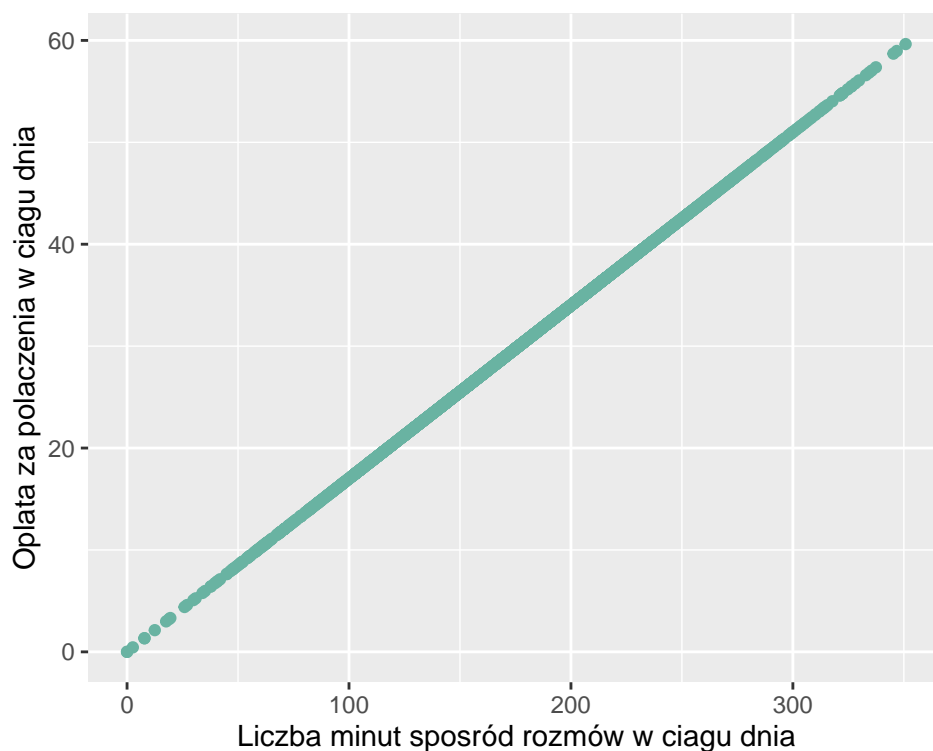
Na podstawie tabel i wykresów (np. 2, 5) możemy zauważyć, że większość zmiennych charakteryzuje się symetrycznym rozkładem. Widać również, że klienci chętniej dzwonią w taryfie wieczornej oraz nocnej.

Zależności pomiędzy parami cech

Wyznamy macierz korelacji, aby sprawdzić, czy pomiędzy poszczególnymi cechami występują liniowe zależności:



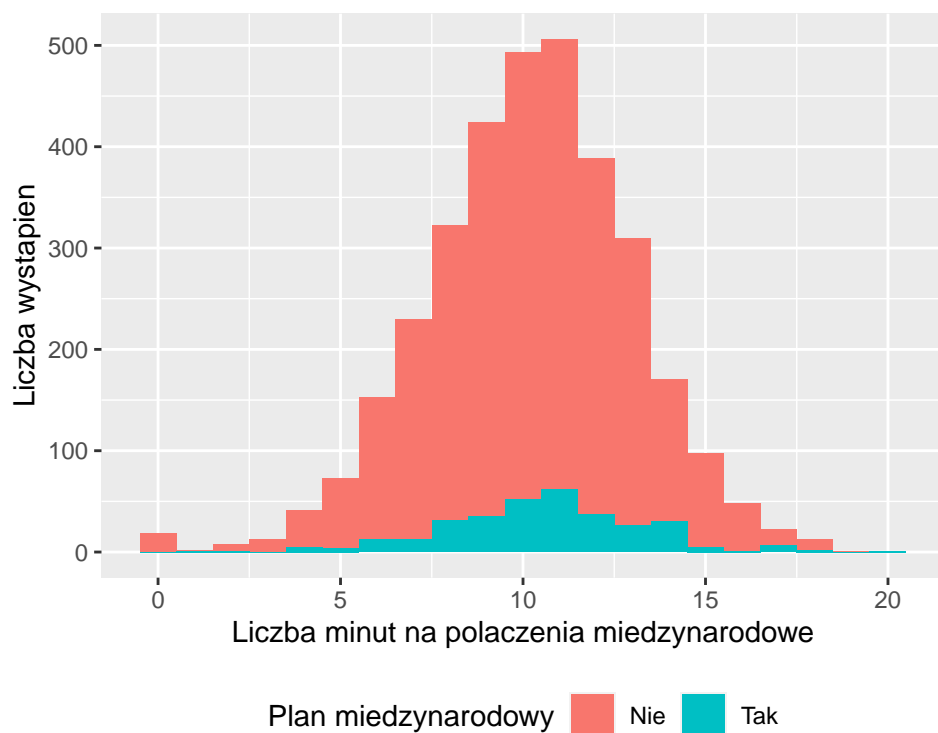
Rysunek 6: Macierz korelacji pomiędzy zmiennymi



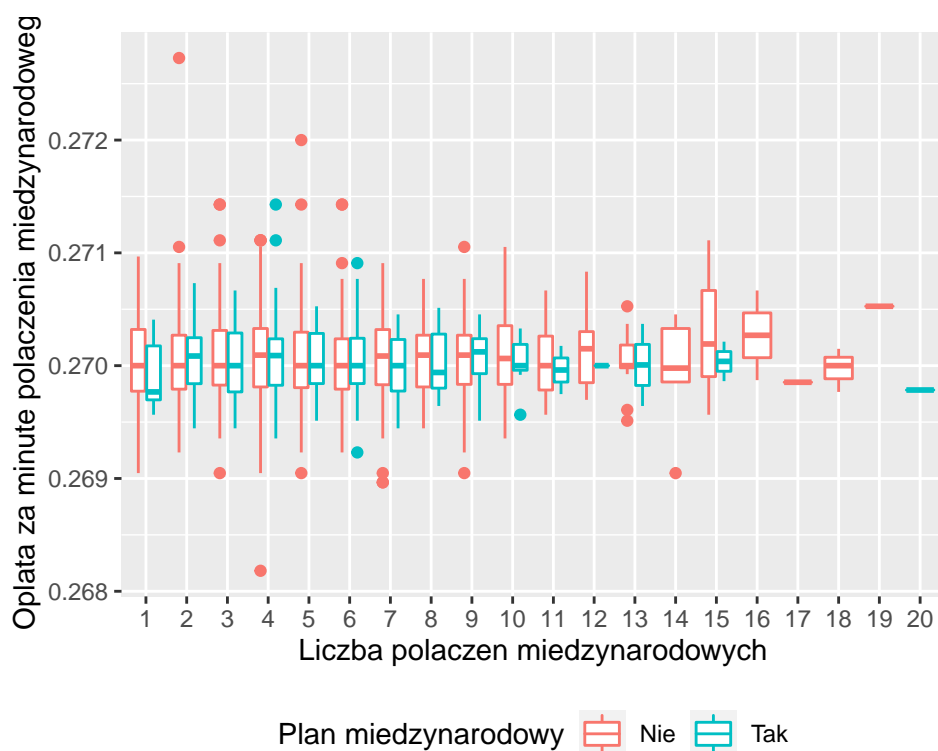
Rysunek 7: Zależność opłaty od liczby minut w ciągu dnia

Istotnie, na Rys. 6 widać, że pomiędzy parami zmiennych `Day.Charge`, `Day.Mins` występuje w pełni liniowa zależność. Liniowa zależność występuje również między analogicznymi zmiennymi w pozostałych częściach dnia, zatem możemy się spodziewać, że opłata za minutę w określonej części dnia jest stała i niezależna od długości rozmowy. Pozostałe zmienne numeryczne nie są ze sobą zbyt dobrze skorelowane.

Przeanalizujemy również ofertę operatora na podstawie usług.



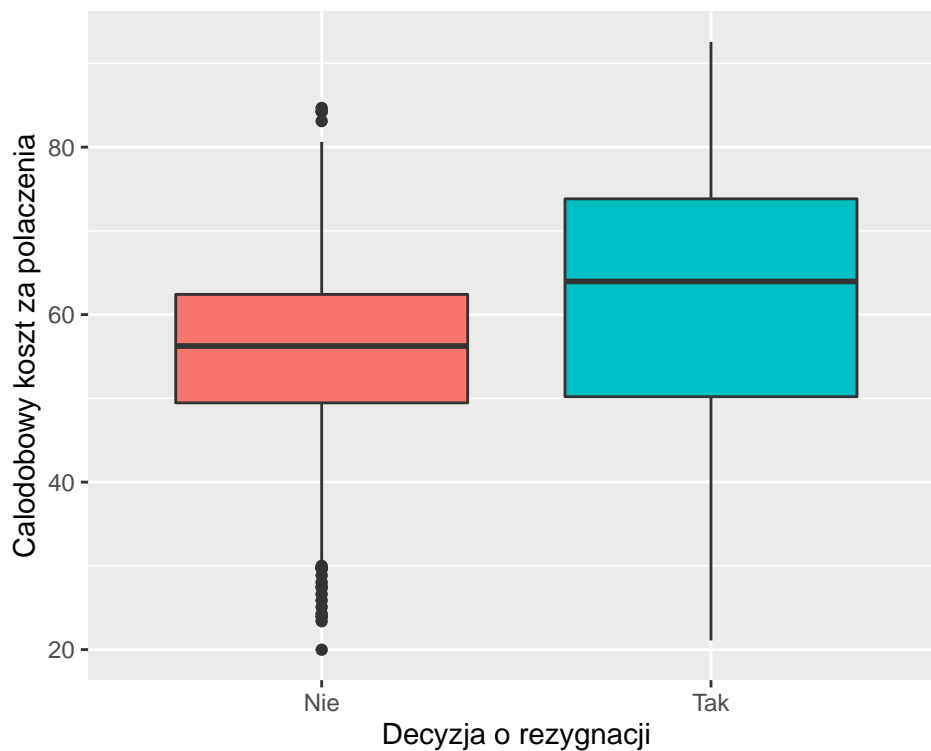
Rysunek 8: Histogram liczby połączeń z podziałem ze względu na plan międzynarodowy



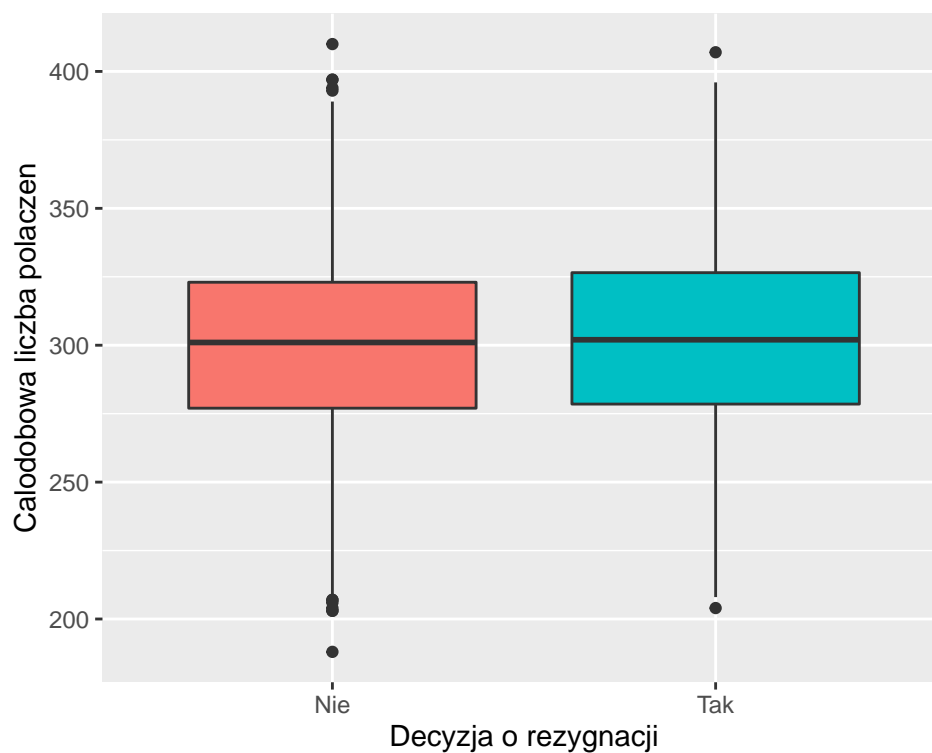
Rysunek 9: Wykresy pudełkowe średniej ceny za minutę połączenia międzynarodowego w zależności od liczby połączeń i z podziałem na usługę planu międzynarodowego

Na podstawie wykresów 8 i 9, możemy wysunąć hipotezę, że plan międzynarodowy jest nieatrakcyjny, ponieważ nie widać, aby przynosił klientom jakiegokolwiek wymierne korzyści, np. opłatę za minutę połączenia. Możemy przypuszczać, że jest to przyczyna małej ilości klientów, wybierających ten produkt.

3.3 Analiza opisowa z podziałem na grupy



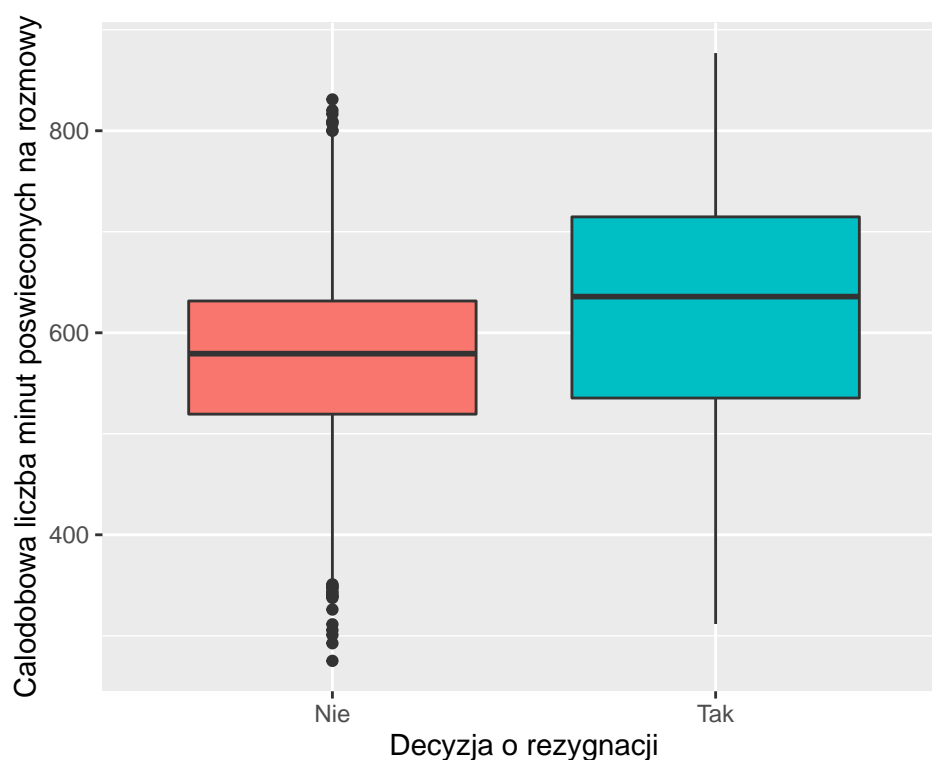
Rysunek 10: Koszty połączeń krajowych na dobę dla obu grup klientów



Rysunek 11: Liczba połączeń krajowych wykonanych przez całą dobę dla obu grup klientów

Jaki widać na wykresach 10 i 11, mimo iż liczba wykonanych połączeń dla obu grup jest podobna, to koszty poniesione za rozmowy są większe dla grupy klientów, która postanowiła zrezygnować z usług sieci telefonii komórkowej.

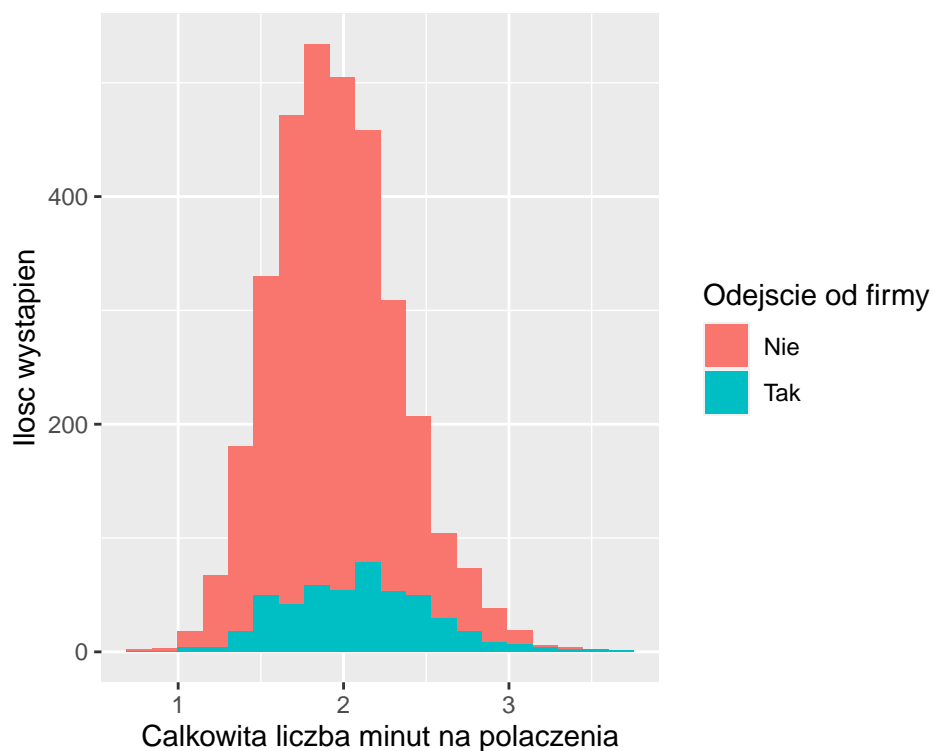
Sprawdzimy, jak długo sumarycznie, trwały rozmowy klientów z podziałem na grupy.



Rysunek 12: Czas trwania połączeń krajowych wykonanych w ciągu doby dla obu grup

Pomimo faktu, że obie grupy wykonują podobną liczbę połączeń, to czas rozmowy jest często dłuższy u grupy, która wykazała brak zainteresowania obecną ofertą sieci komórkowej (Rys. 12), tj. osobom, które spędzają więcej czasu na rozmowach telefonicznych, mogła nie zostać przedstawiona wystarczająco korzystna dla nich oferta.

Sprawdźmy również, jak przedstawia się rozkład połączeń w ciągu dnia:

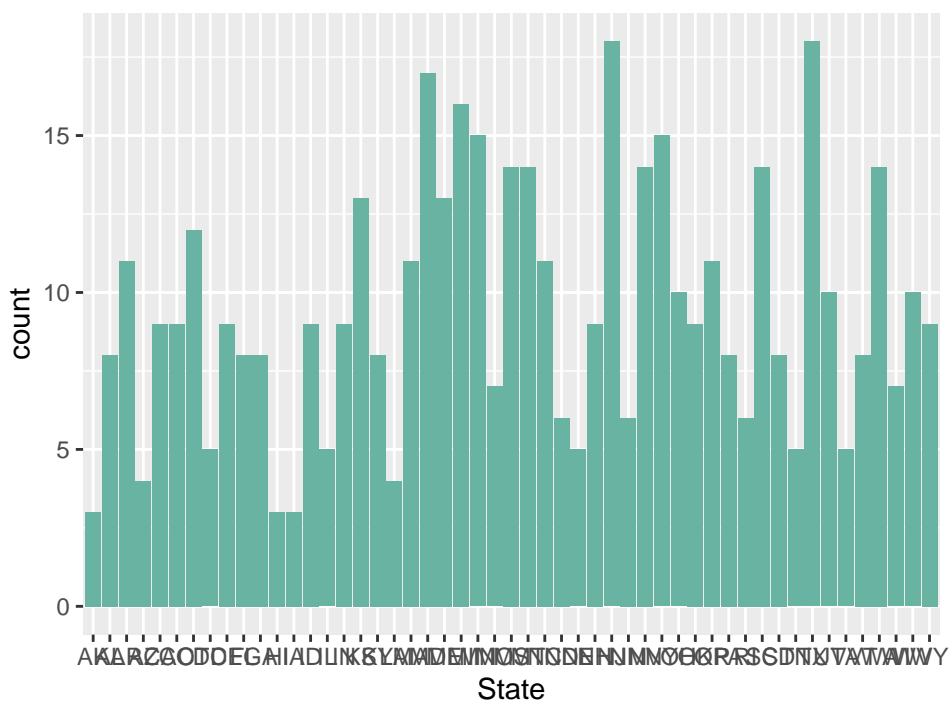


Rysunek 13: Histogram całkowitej liczby minut na połączenia krajowe

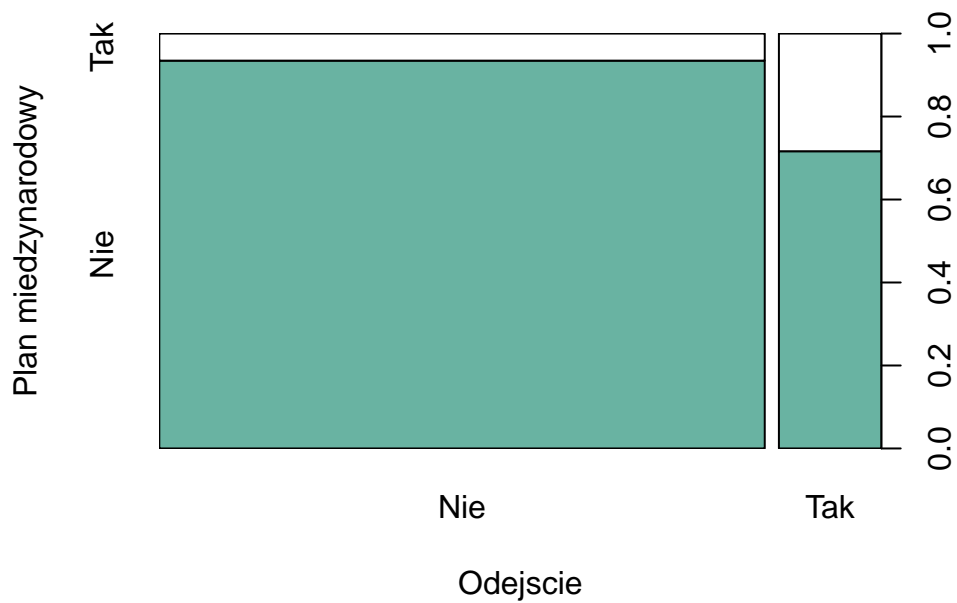
Na podstawie wykresu 13, możemy zauważyć, że wśród osób, które odbywają wiele minut rozmów, potw.

Analizujemy, w których stanach najwięcej klientów zrezygnowało z usług:

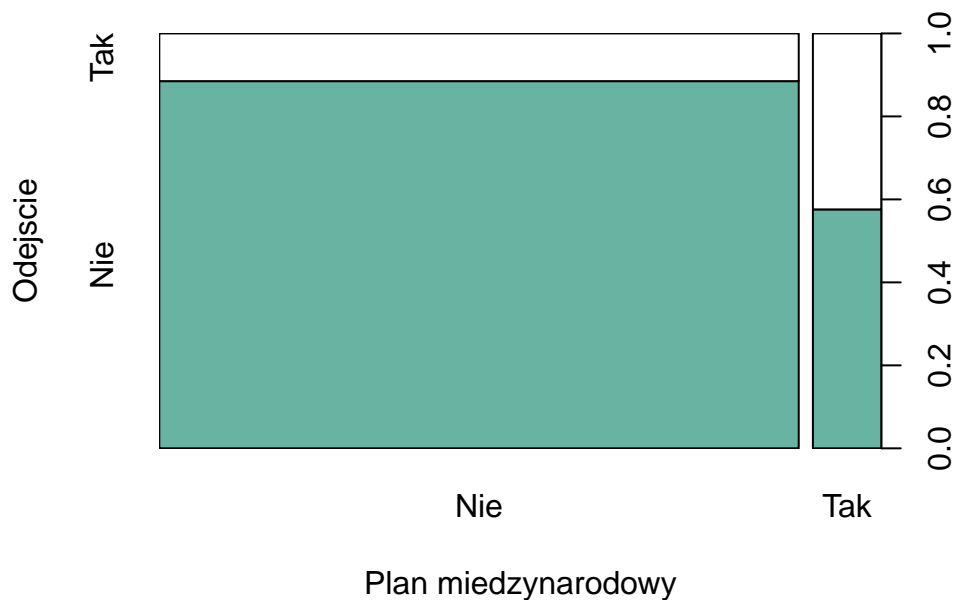
Liczba osób, które zrezygnowały z usług, dla poszczególnych



Spośród klientów rezygnujących z usług, najwięcej zamieszkiwało odpowiednio stany: Teksas, New Jersey oraz Maryland.



Rysunek 14: Osoby rezygnujące, a udział w planie międzynarodowym.

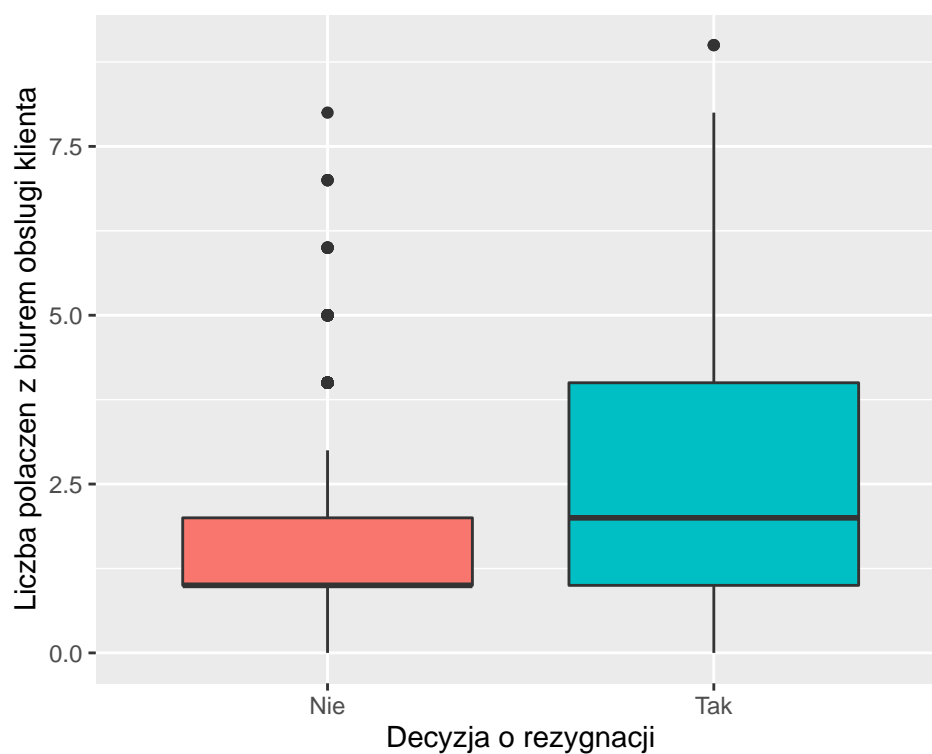


Rysunek 15: Plan międzynarodowy, a rezygnacja usług

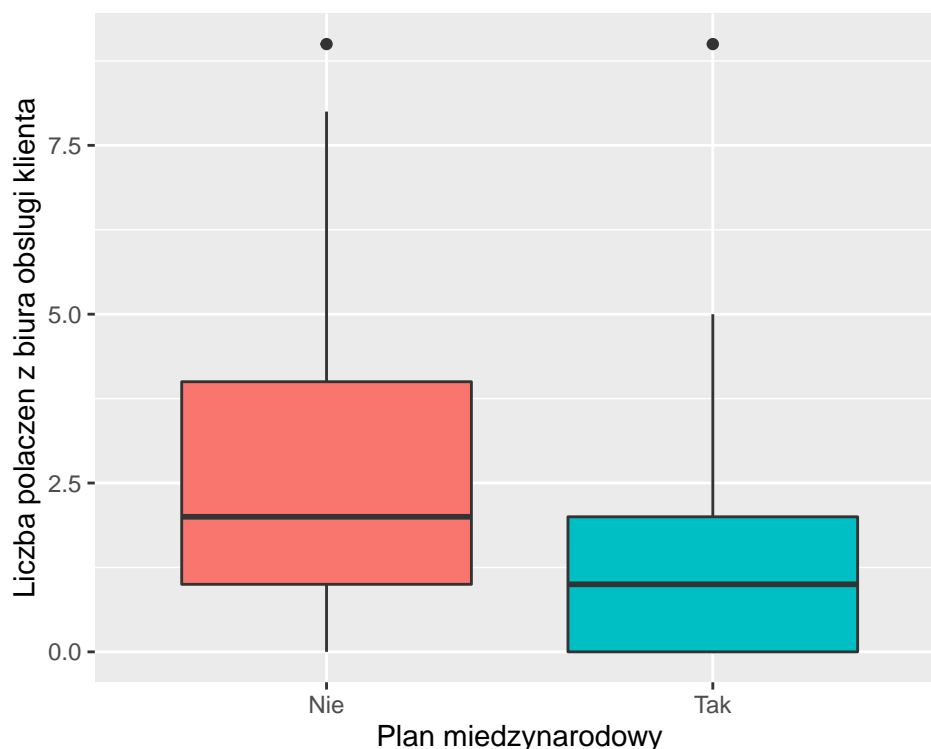
Wśród osób rezygnujących z usług operatora, aż 137 na 483 brało udział w planie międzynarodowym (Rys.14). Ponadto widać, że aż 42% klientów, którzy korzystali z planu międzyna-

rodowego, zrezygnowało z usług (Rys.15).

Przeanalizujemy również liczbę telefonów do biura obsługi klienta, ponieważ przyczyną kontaktu z operatorem zazwyczaj są problemy różnego typu.



Rysunek 16: Liczba telefonów do biura obsługi klienta, a liczba klientów, którzy zrezygnowali



Rysunek 17: liczba telefonów do biura obsługi klienta, a plan międzynarodowy

Widzimy na wykresie 16, że osoby, które podjęły decyzję o rezygnacji, wykonały średnio więcej telefonów do biura obsługi klienta. Jednocześnie, wzmożona liczba telefonów do biura obsługi klienta w przypadku osób, które postanowiły zrezygnować z usług, niekoniecznie musiała wynikać z niezbyt korzystnych warunków planu międzynarodowego (Rys.17), czyli przyczyny można też dopatrywać się w niewykwalifikowanych pracownikach biura obsługi.

4 Podsumowanie

Z uzyskanych wyników, dowiedzieliśmy się, że większość klientów firmy, to klienci niekorzystający z dodatkowych usług, co wskazuje na nieadekwatną lub nieatrakcyjną ofertę.

Na podstawie przeprowadzonych analiz, możemy zauważyć, że odejście klientów jest spowodowane ofertą, która jest dostosowana wyłącznie do podstawowych potrzeb. Widzimy, że dodatkowe usługi, takie jak plan międzynarodowy, nie są korzystne dla klientów i może to być przyczyną odchodzenia klientów z firmy. Ponadto, wygląda na to, że z usług rezygnują osoby, które podczas trwania umowy wykorzystywały więcej minut. Przyczyny można również dopatrywać się w nieodpowiednio przeszkolonych i przygotowanych konsultantach z biura obsługi klienta.

Stąd, możemy wyszczególnić główne 3 problemy:

- Oferta nie jest dostosowana do osób, które podczas trwania umowy wykorzystują wiele minut.
- Plan międzynarodowy skłania ludzi do odejścia od firmy.
- Występują problemy z obsługą klienta, które nie są spowodowane powyższymi problemami.

Aby zwiększyć liczbę klientów, można dostosować ofertę międzynarodową do potrzeb klientów oraz przygotować dodatkową ofertę lojalnościową. Ponadto, można skontrolować biuro obsługi klienta, aby sprawdzić, czy przyczyną odejść są zaniedbania pracowników, czy też inne problemy.