

Raport - Telco

Filip Michewicz 282239
Wiktor Niedźwiedzki 258882

31 marca 2025 Anno Domini

Spis treści

1 Wprowadzenie	2
2 Przygotowanie danych. Podstawowe informacje o danych	3
2.1 Wczytywanie danych	3
2.2 Cechy danych	3
3 Analiza opisowa — wskaźniki sumaryczne	4
3.1 Wskaźniki sumaryczne	4
4 Opis wykonanych analiz	5
5 Analiza opisowa - wykresy	5
5.1 Wskaźniki korelacji	6
5.2 Histogramy i wykresy pudełkowe zmiennych ilościowych	7
5.3 Wykresy słupkowe zmiennych jakościowych	11
5.4 Wykresy rozrzutów	19
6 Analiza opisowa z podziałem na grupy	22
6.1 Histogramy i wykresy pudełkowe	22
6.2 Wykresy słupkowe zmiennych jakościowych	26
6.3 Wykresy rozrzutów	30
7 Dyskusja	32
8 Wnioski	33
9 Rekomendacje	33

Spis wykresów

1 Macierz Korelacji - Klienci Telco	6
2 Histogram dla zmiennej tenure	7
3 Wykres pudełkowy dla zmiennej tenure	8
4 Histogram dla zmiennej MonthlyCharges	8
5 Wykres pudełkowy dla zmiennej MonthlyCharges	9
6 Histogram dla zmiennej TotalCharges	9
7 Wykres pudełkowy dla zmiennej TotalCharges	10
8 Wykres słupkowy zmiennej SeniorCitizen	11
9 Wykres słupkowy zmiennej Partner	11
10 Wykres słupkowy zmiennej Dependents	12

11	Wykres słupkowy zmiennej PhoneService	12
12	Wykres słupkowy zmiennej MultipleLines	13
13	Wykres słupkowy zmiennej InternetService	13
14	Wykres słupkowy zmiennej OnlineSecurity	14
15	Wykres słupkowy zmiennej DeviceProtection	15
16	Wykres słupkowy zmiennej TechSupport	15
17	Wykres słupkowy zmiennej StreamingTV	16
18	Wykres słupkowy zmiennej StreamingMovies	16
19	Wykres słupkowy zmiennej Contract	17
20	Wykres słupkowy zmiennej PaperlessBilling	17
21	Wykres słupkowy zmiennej PaymentMethod	18
22	Wykres słupkowy zmiennej Churn	18
23	Zależność TotalCharges od MonthlyCharges z krzywą wygładzającą	19
24	Zależność TotalCharges od tenure z krzywą wygładzającą	20
25	Zależność MonthlyCharges od tenure z krzywą wygładzającą	21
26	Macierz par zmiennych numerycznych	22
27	Histogram dla zmiennej tenure z podziałem ze względu na churn	22
28	Wykres pudełkowy dla zmiennej tenure z podziałem ze względu na churn	23
29	Histogram dla zmiennej MonthlyCharges z podziałem ze względu na churn	24
30	Wykres pudełkowy dla zmiennej MonthlyCharges z podziałem ze względu na churn	25
31	Wykres pudełkowy dla zmiennej TotalCharges z podziałem ze względu na churn	25
32	Wykres słupkowy zmiennej Dependents z podziałem ze względu na churn	26
33	Wykres słupkowy zmiennej InternetService z podziałem ze względu na churn	26
34	Wykres słupkowy zmiennej OnlineSecurity z podziałem ze względu na churn	27
35	Wykres słupkowy zmiennej OnlineBackup z podziałem ze względu na churn	27
36	Wykres słupkowy zmiennej DeviceProtection z podziałem ze względu na churn	28
37	Wykres słupkowy zmiennej TechSupport z podziałem ze względu na churn	28
38	Wykres słupkowy zmiennej Contract z podziałem ze względu na churn	29
39	Wykres słupkowy zmiennej PaperlessBilling z podziałem ze względu na churn	29
40	Wykres słupkowy zmiennej PaymentMethod z podziałem ze względu na churn	30
41	Wykresy rozrzutów Total Charges ze względu na tenure z podziałem ze względu na zmienną churn	30
42	Wykresy rozrzutów MonthlyCharges ze względu na tenure z podziałem ze względu na zmienną churn	31
43	Macierz par zmiennych numerycznych z podziałem ze względu na churn	32

Spis tabel

1	Opis zmiennych w zbiorze danych	3
2	Tabela z podstawowymi wskaźnikami sumarycznymi dla zmiennych ilościowych	4

1 Wprowadzenie

W ramach tego zadania przeanalizowano dane dotyczące klientów firmy telekomunikacyjnej Telco, ze szczególnym uwzględnieniem problemu odpływu klientów (churn). Cele dokonanej analizy:

- Zbadanie podstawowych cech zbioru danych, zarówno zmiennych ilościowych, jak i jakościowych.
- Identyfikacja zależności między zmiennymi oraz ocena ich rozkładów.
- Porównanie grup lojalnych klientów (Churn = ‘No’) i tych, którzy odeszli (Churn = ‘Yes’).

Chcemy znaleźć odpowiedzi na pytania:

- Jakie są podstawowe właściwości analizowanych zmiennych?

- Które cechy najlepiej różnicują klientów lojalnych od tych, którzy rezygnują z usług?
- Jakie mogą być główne przyczyny odchodzenia klientów i co firma może zrobić, aby temu przeciwdziałać?

2 Przygotowanie danych. Podstawowe informacje o danych

W tej części dokonano wstępniego opracowania danych, identyfikując wszystkie zmienne oraz korygując ich typy w celu zapewnienia poprawności dalszej analizy. Zmieniono również nazwy będące oryginalnie po angielsku na język polski w celu zwiększenia czytelności oraz ułatwienia interpretacji danych.

2.1 Wczytywanie danych

```
klienci <- read.csv(file="WA_Fn-UseC_-Telco-Customer-Churn.csv",
                     stringsAsFactors=TRUE)

# Niepoprawnie wczytane typy dla: customerID oraz SeniorCitizen

klienci$SeniorCitizen <- as.factor(klienci$SeniorCitizen)
klienci$SeniorCitizen <- revalue(klienci$SeniorCitizen,
                                    c("0" = "Nie", "1" = "Tak"))

klienci$customerID <- as.character(klienci$customerID)
```

W Tabeli 1. przedstawiono nazwy zmiennych, ich typy oraz opis, który wyjaśnia, co każda z nich reprezentuje.

Tabela 1: Opis zmiennych w zbiorze danych

Zmienna	Typ	Opis
customerID	character	Unikalny identyfikator klienta
gender	factor	Płeć klienta (Mężczyzna, Kobieta)
SeniorCitizen	factor	Czy klient jest seniorem? (Tak/Nie)
Partner	factor	Czy klient ma partnera/partnerkę? (Tak/Nie)
Dependents	factor	Czy klient ma osoby na utrzymaniu? (Tak/Nie)
tenure	integer	Liczba miesięcy, przez które klient korzysta/korzystał z usług
PhoneService	factor	Czy klient ma usługę telefoniczną? (Tak/Nie)
MultipleLines	factor	Czy klient ma wiele linii telefonicznych? (Tak/Nie/Brak usługi telefonicznej)
InternetService	factor	Typ usługi internetowej (DSL, Światłowód, Brak)
OnlineSecurity	factor	Czy klient ma usługę zabezpieczeń online? (Tak/Nie/Brak internetu)
OnlineBackup	factor	Czy klient ma usługę kopii zapasowej online? (Tak/Nie/Brak internetu)
DeviceProtection	factor	Czy klient ma usługę ochrony urządzenia? (Tak/Nie/Brak internetu)
TechSupport	factor	Czy klient ma usługę wsparcia technicznego? (Tak/Nie/Brak internetu)
StreamingTV	factor	Czy klient korzysta z usługi streamingu TV? (Tak/Nie/Brak internetu)
StreamingMovies	factor	Czy klient korzysta z usługi streamingu filmów? (Tak/Nie/Brak internetu)
Contract	factor	Rodzaj umowy (Miesięczna, Roczna, Dwuletnia)
PaperlessBilling	factor	Czy klient korzysta z faktur elektronicznych? (Tak/Nie)
PaymentMethod	factor	Metoda płatności (np. karta kredytowa, przelew, automatyczne obciążenie)
MonthlyCharges	numeric	Miesięczna opłata za usługi
TotalCharges	numeric	Łączna opłata pobrana od klienta
Churn	factor	Czy klient zrezygnował z usług? (Tak/Nie)

2.2 Cechy danych

- Zbiór danych zawiera 7043 obserwacje (wiersze) oraz 21 zmiennych (kolumn).

- Zmienne *tenure*, *MonthlyCharges* oraz *TotalCharges* mają typ ilościowy (*numeric*).
- Zmienna *customerID* to identyfikator klienta zapisany jako ciąg znaków.
- Pozostałe zmienne mają typ kategoryczny (*factor*).
- Poza unikalnym identyfikatorem klienta, wszystkie pozostałe zmienne są istotne, ponieważ umożliwiają analizę popularności usług, metod płatności oraz średnich kosztów miesięcznych.

`klienci$customerID <- NULL`

- W zbiorze danych występuje 11 brakujących wartości w kolumnie *TotalCharges*, oznaczonych jako *NA*. Braki te można wyjaśnić faktem, że dotyczą nowych klientów, którzy jeszcze nie dokonali płatności.
- Nietypowym rozwiązaniem jest kodowanie zmiennej *SeniorCitizen* za pomocą wartości 0 i 1, zamiast bardziej czytelnych kategorii, takich jak *Tak* i *Nie*.

3 Analiza opisowa — wskaźniki sumaryczne

3.1 Wskaźniki sumaryczne

Dla danych ilościowych obliczono podstawowe wskaźniki sumaryczne, takie jak minimum, pierwszy i trzeci kwartyl, medianę, maksimum, średnią odchylenie standardowe, skośność i kurtozę.

Tabela 2: Tabela z podstawowymi wskaźnikami sumarycznymi dla zmiennych ilościowych

	tenure	MonthlyCharges	TotalCharges
Min	0.000	18.250	18.800
1st Qu.25%	9.000	35.500	401.450
Median	29.000	70.350	1397.475
Średnia	32.371	64.762	2283.300
3rd Qu.75%	55.000	89.850	3794.738
Max	72.000	118.750	8684.800
Odch. Stand.	24.559	30.090	2266.771
Skośność	0.239	-0.220	0.961
Kurtoza	-1.388	-1.258	-0.233

Kurtoza dla wszystkich zmiennych (*tenure*, *MonthlyCharges*, *TotalCharges*) jest ujemna, co oznacza, że ich rozkłady są bardziej płaskie niż rozkład normalny (platykurtyczne). Wskazuje to na większe rozproszenie wartości, czyli brak wyraźnego szczytu i mniej ekstremalnych wartości w porównaniu do typowego rozkładu normalnego.

Skośność informuje o asymetrii rozkładu:

- *tenure* (0.239) – lekko prawostronnie skośny, czyli więcej klientów ma krótszy staż, ale występują też dłuższe okresy.
- *MonthlyCharges* (-0.220) – lekko lewostronnie skośny, co oznacza, że większość klientów płaci stosunkowo niskie kwoty, ale zdarzają się wyższe wartości.
- *TotalCharges* (0.961) – wyraźnie prawostronnie skośny, co sugeruje, że większość klientów zapłaciła stosunkowo niewiele, ale istnieją przypadki z bardzo wysokimi wartościami całkowitych opłat.

4 Opis wykonanych analiz

W analizie przeprowadzono szereg wizualizacji, które miały na celu dokładne zidentyfikowanie kluczowych trendów oraz zależności w analizowanych danych.

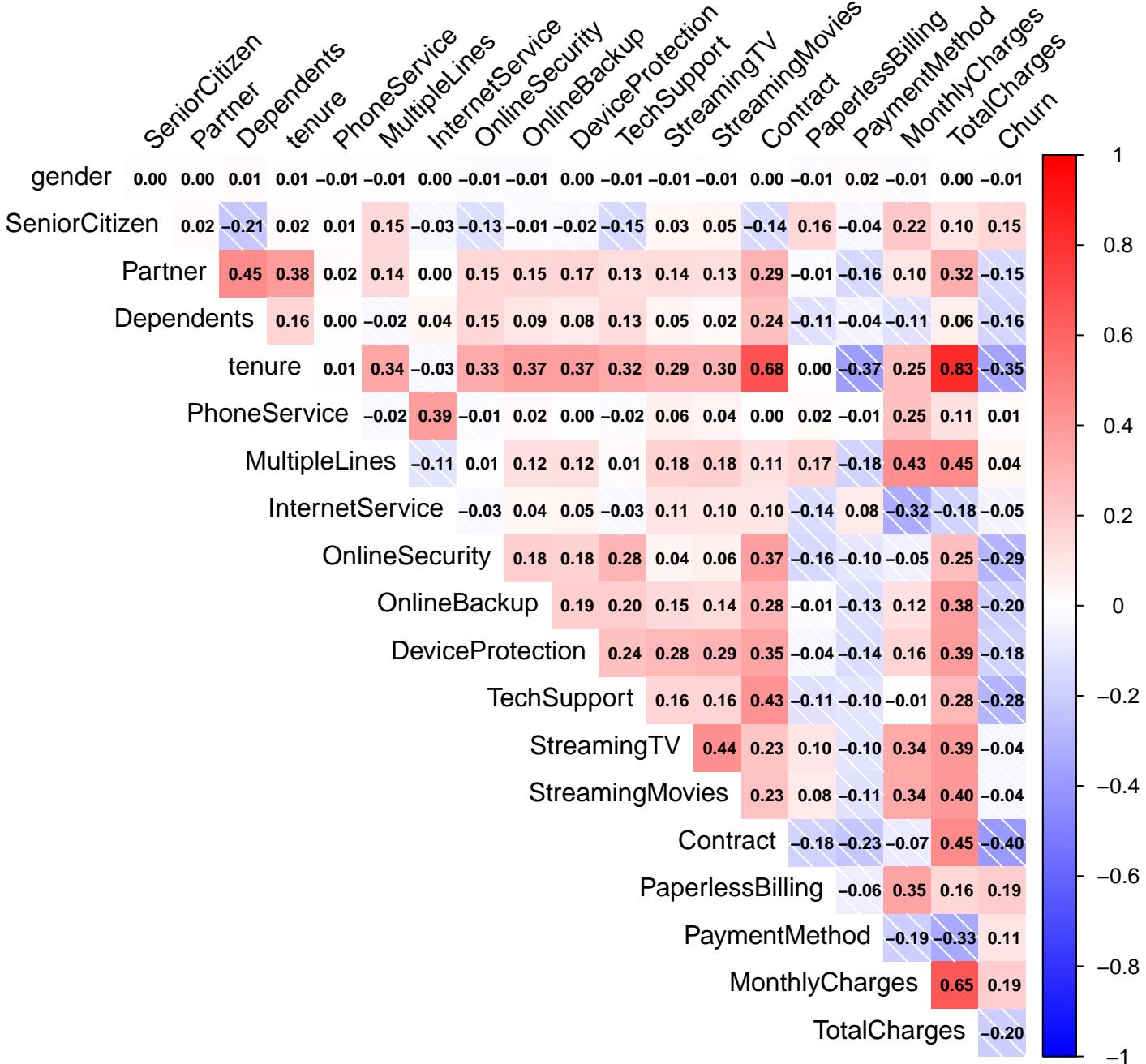
Główne elementy analizy obejmowały:

- Stworzenie macierzy korelacji, mającej na celu uchwycenie zależności między zmiennymi ilościowymi, takimi jak długość współpracy a łączna wysokość opłat.
- Przygotowanie histogramów i wykresów pudełkowych, które umożliwiły szczegółowe zrozumienie rozkładów zmiennych liczbowych oraz ich asymetrii.
- Wykonanie wykresów słupkowych dla zmiennych jakościowych, pozwalających na ocenę struktury klientów, np. pod kątem rodzaju umowy czy usług dodatkowych.
- Zastosowanie wykresów rozrzutów z nałożonymi krzywymi wygładzającymi, które ukazały zależności między zmiennymi liczbowymi, takimi jak opłaty miesięczne a ryzyko rezygnacji.
- Realizację wizualizacji z podziałem na grupy klientów (lojalni vs. rezygnujący), co umożliwiło ocenę różnic w ich zachowaniach.

5 Analiza opisowa - wykresy

W tej części przedstawiono dane w postaci wykresów i dokonano ich podstawowej analizy.

5.1 Wskaźniki korelacji



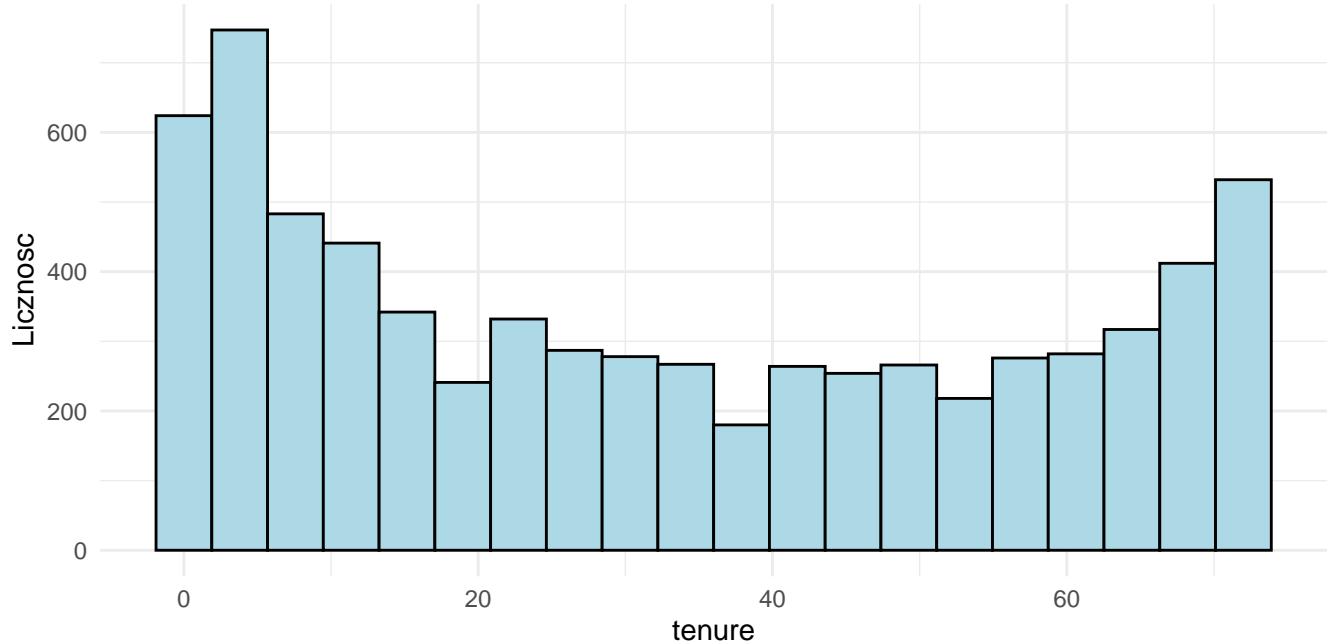
Wykres 1: Macierz Korelacji - Klienci Telco

Na Wykresie 1. przedstawiono macierz korelacji dla wczytanych danych. Widać wyraźne zależności, między innymi:

- Klienci z dłuższym okresem współpracy (*Tenure*) oraz posiadający dłuższe umowy (*Contract*) rzadziej opuszczają firmę, co przekłada się na niższy wskaźnik churn.
- Posiadanie usług, takich jak *OnlineSecurity*, *TechSupport* czy *OnlineBackup*, wiąże się z mniejszym odsetkiem klientów, którzy zdecydowali się odejść.
- Wyższe miesięczne rachunki (*MonthlyCharges*) oraz cechy, takie jak korzystanie z usług internetowych i

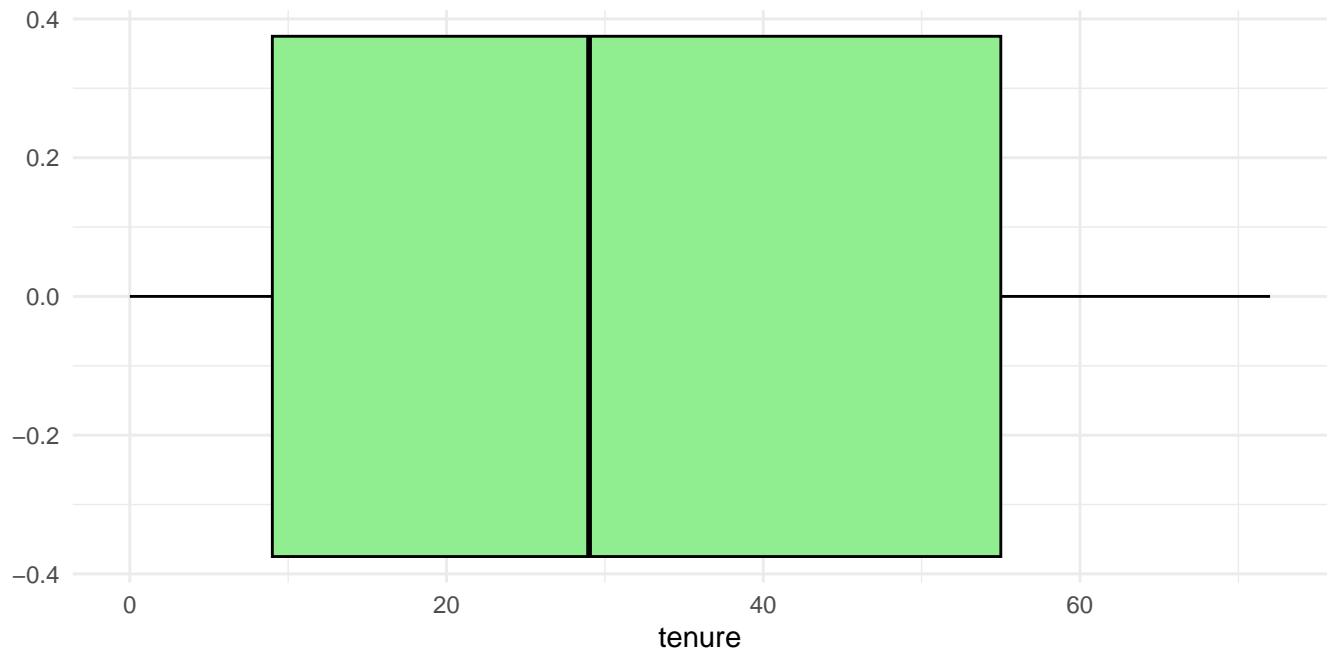
e-fakturowanie (*PaperlessBilling*), korelują z wyższym odsetkiem klientów rezygnujących z usług. Te obserwacje stanowią podstawę do dalszych, bardziej szczegółowych analiz.

5.2 Histogramy i wykresy pudełkowe zmiennych ilościowych



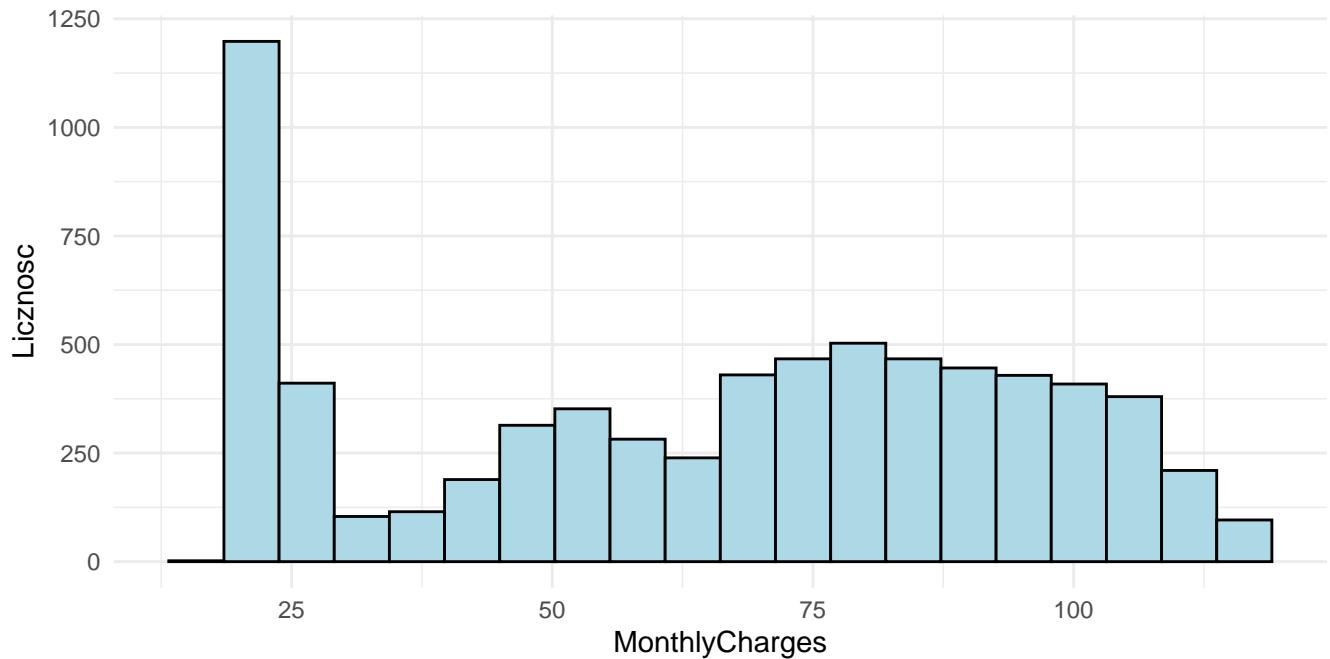
Wykres 2: Histogram dla zmiennej tenure

Wykres 2. przedstawia histogram zmiennej *tenure*, który ilustruje długość korzystania przez klientów z usług firmy. Rozkład ten jest dość równomierny, choć zauważalna jest większa liczba klientów, którzy korzystali z usług przez krótki okres czasu. Wskazuje to na wyraźną grupę klientów, którzy niedawno rozpoczęli korzystanie z naszych usług lub zakończyli współpracę w krótkim czasie.



Wykres 3: Wykres pudełkowy dla zmiennej tenure

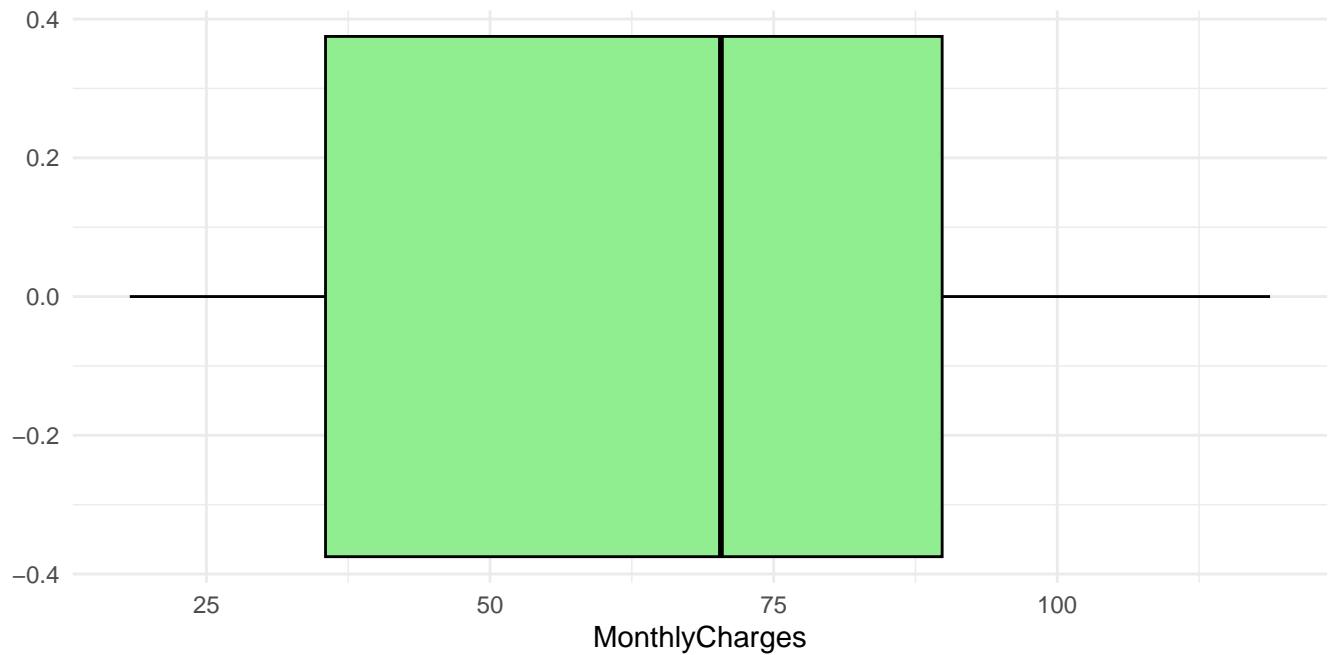
Wykres 3. przedstawia wykres pudełkowy dla zmiennej *tenure*. Z danych wynika, że wśród naszych klientów znajdują się zarówno osoby nowo pozyskane, jak i ci, którzy współpracują z nami od 6 lat. Środkowa część próby wskazuje, że większość klientów jest z nami od 9 do 55 miesięcy.



Wykres 4: Histogram dla zmiennej MonthlyCharges

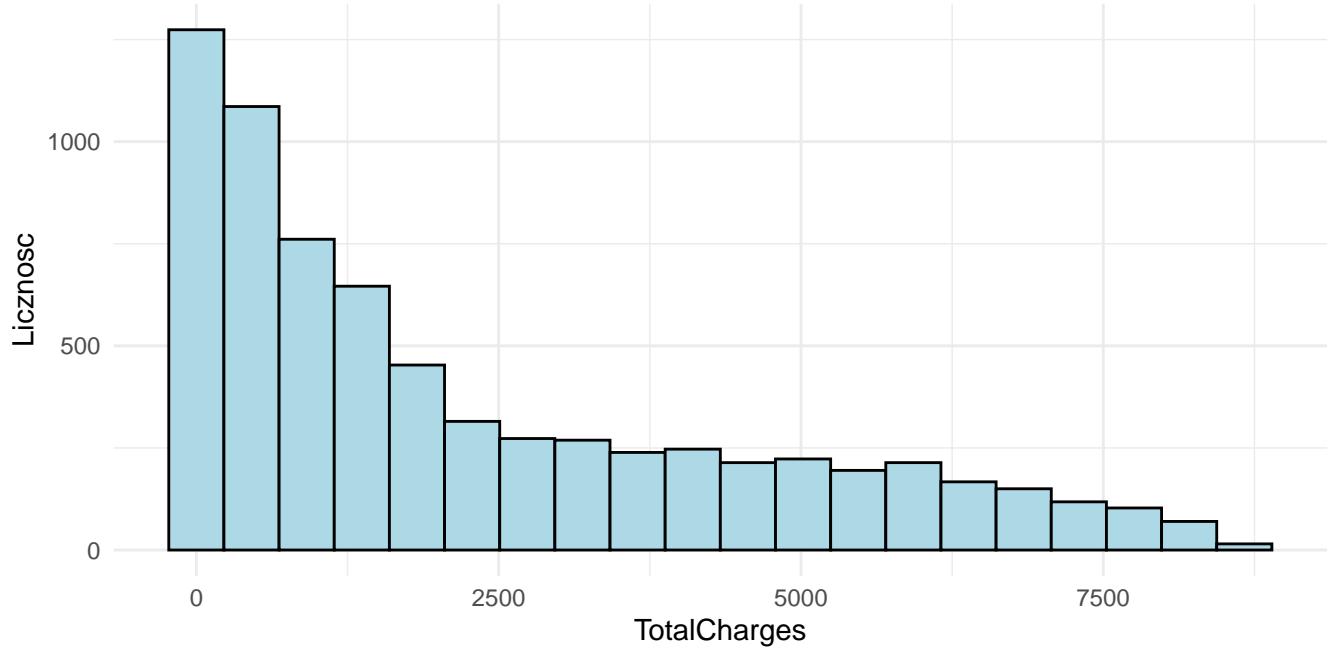
Wykres 4. przedstawia histogram zmiennej *MonthlyCharges*. Z danych wynika, że wielu klientów korzysta z najwyższych usług, ale wyraźnie wyróżnia się grupa klientów, którzy wydają od około 60 do 120 dolarów

miesięcznie.



Wykres 5: Wykres pudelkowy dla zmiennej *MonthlyCharges*

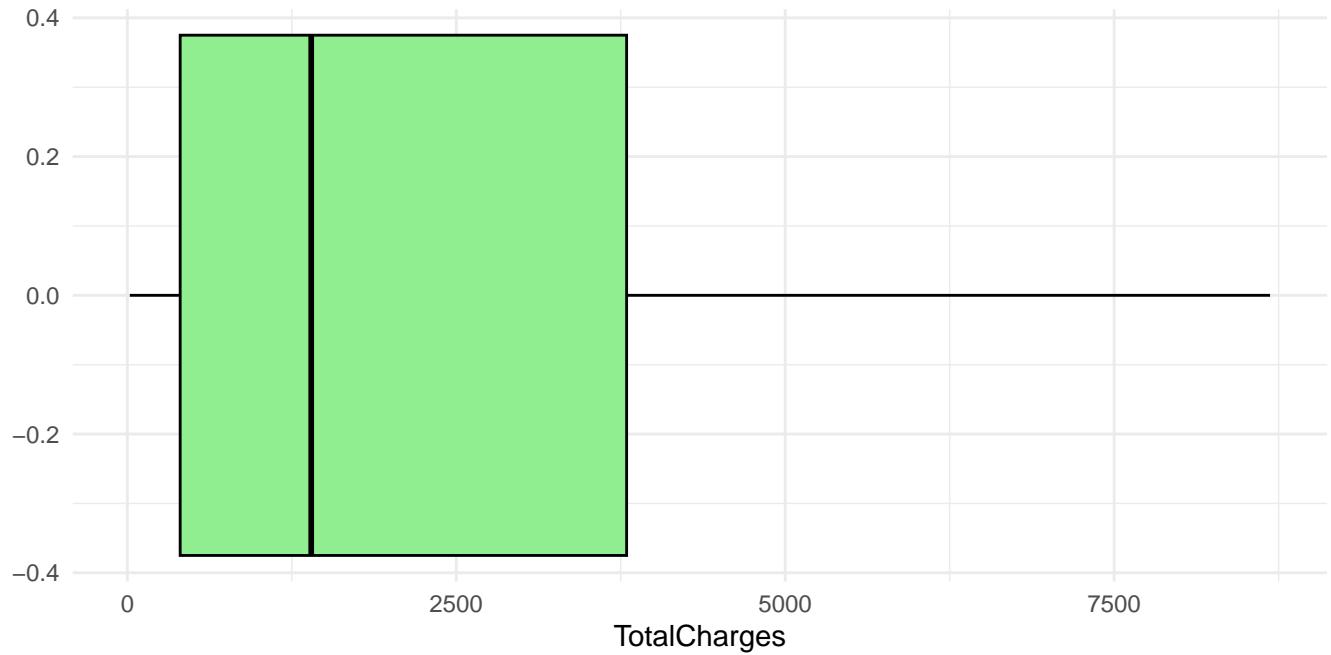
Wykres 5. przedstawia wykres pudelkowy dla zmiennej *MonthlyCharges*. Widać na nim, że większa część klientów ponosi raczej niższe opłaty, jednak mediana jest przesunięta w prawą stronę pudełka, co świadczy o asymetrii prawostronnej (dodatniej) rozkładu.



Wykres 6: Histogram dla zmiennej *TotalCharges*

Wykres 6. przedstawia histogram dla zmiennej *Total Charges*. Zgodnie z oczekiwaniami, rozkład jest

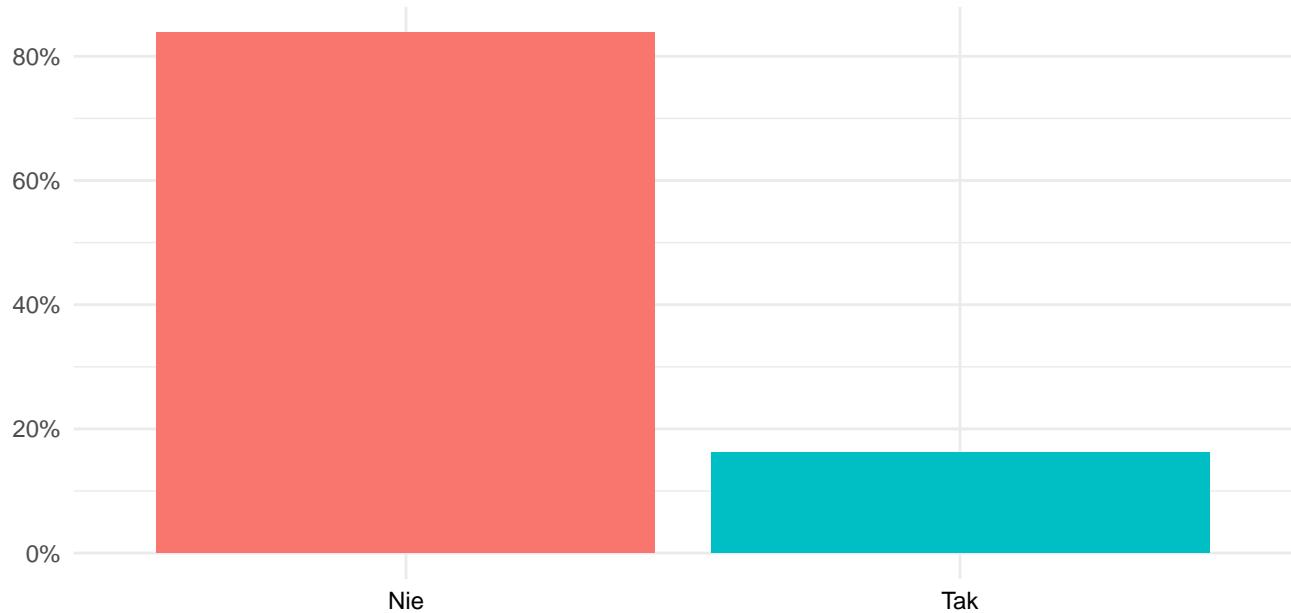
malejący, jednak warto zwrócić uwagę na tempo tego spadku. Dla niższych wartości (poniżej 2500 dolarów) liczność obserwacji zmniejsza się stosunkowo szybko, natomiast w dalszej części rozkładu tempo spadku jest łagodniejsze. Może to sugerować, że większość klientów generuje raczej niższe łączne opłaty, a wyższe wartości występują rzadziej, ale w sposób bardziej rozproszony.



Wykres 7: Wykres pudełkowy dla zmiennej *TotalCharges*

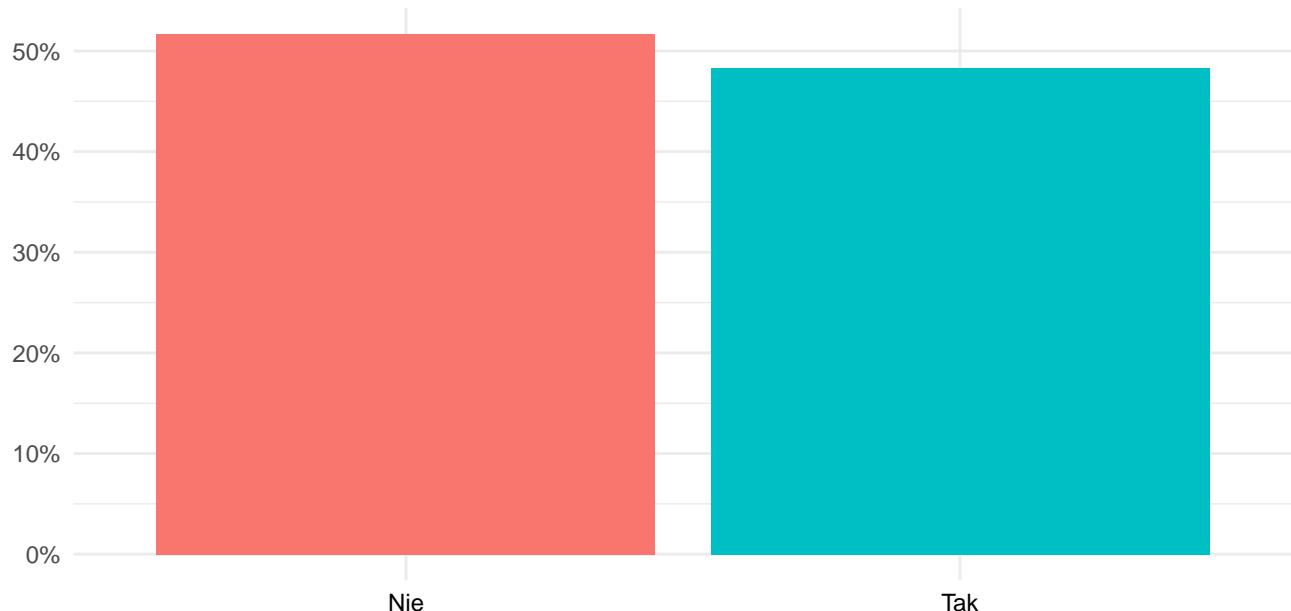
Wykres 7. przedstawia wykres pudełkowy dla zmiennej *Total Charges*. Pudełko oraz mediana są przesunięte w lewą stronę, co wskazuje, że większość klientów ponosi raczej niższe łączne opłaty. Świadczy to o asymetrii prawostronnej rozkładu, gdzie mniejsze wartości są bardziej skupione, a wyższe rozłożone w dłuższym ogonie. Zaskakujący jest brak obserwacji odstających, co może sugerować, że wartości *Total Charges* są stosunkowo jednolite i nie występują znacząco odbiegające przypadki.

5.3 Wykresy słupkowe zmiennych jakościowych



Wykres 8: Wykres słupkowy zmiennej SeniorCitizen

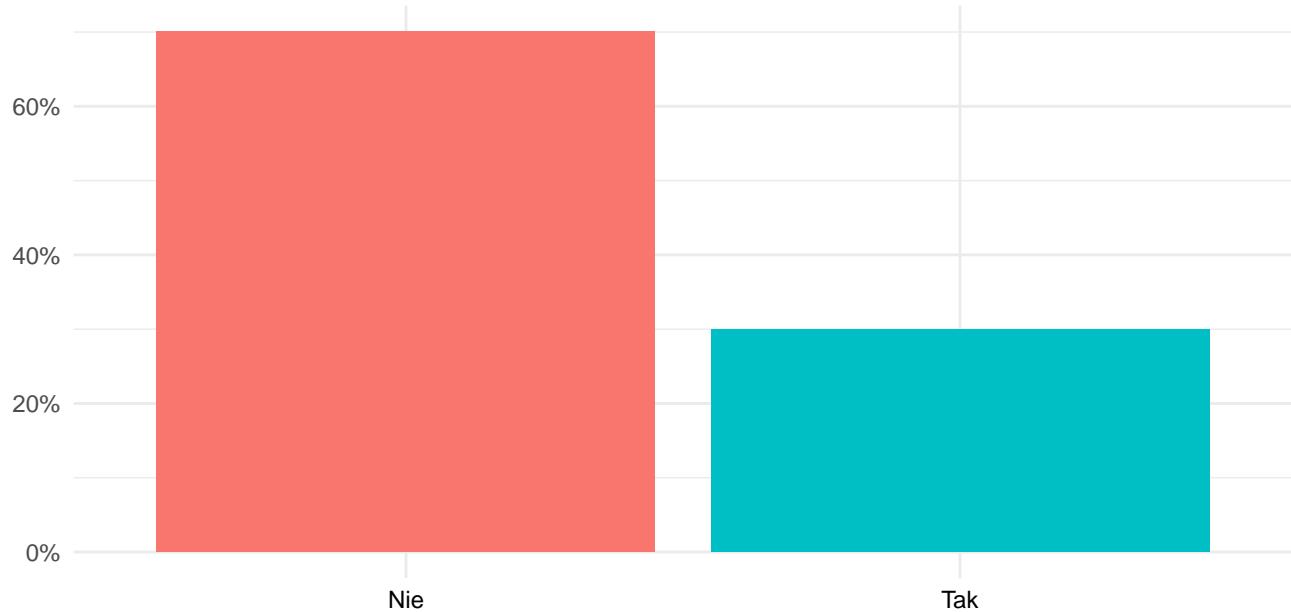
Na wykresie 8. przedstawiono wykres słupkowy dla zmiennej *SeniorCitizen*. Wynika z niego, że większość klientów to osoby, które nie są seniorami. Oznacza to, że w analizowanej grupie przeważają młodsi użytkownicy, co może mieć wpływ na preferencje dotyczące usług oraz wzorce płatności.



Wykres 9: Wykres słupkowy zmiennej Partner

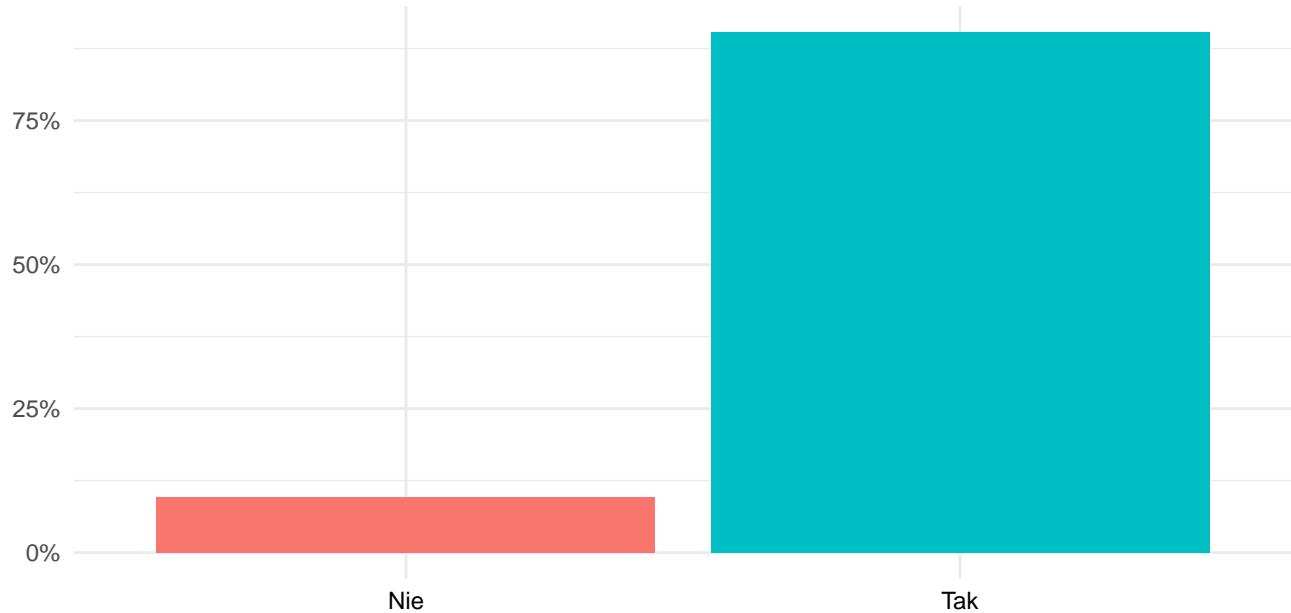
Na wykresie 9. przedstawiono wykres słupkowy dla zmiennej *Partner*. Wynika z niego, że rozkład posiadania partnera jest dość równomierny – liczba klientów będących w związku jest zbliżona do liczby osób, które

nie mają partnera. Taki wynik sugeruje, że status partnerski nie jest czynnikiem silnie dominującym wśród analizowanych klientów.



Wykres 10: Wykres słupkowy zmiennej *Dependents*

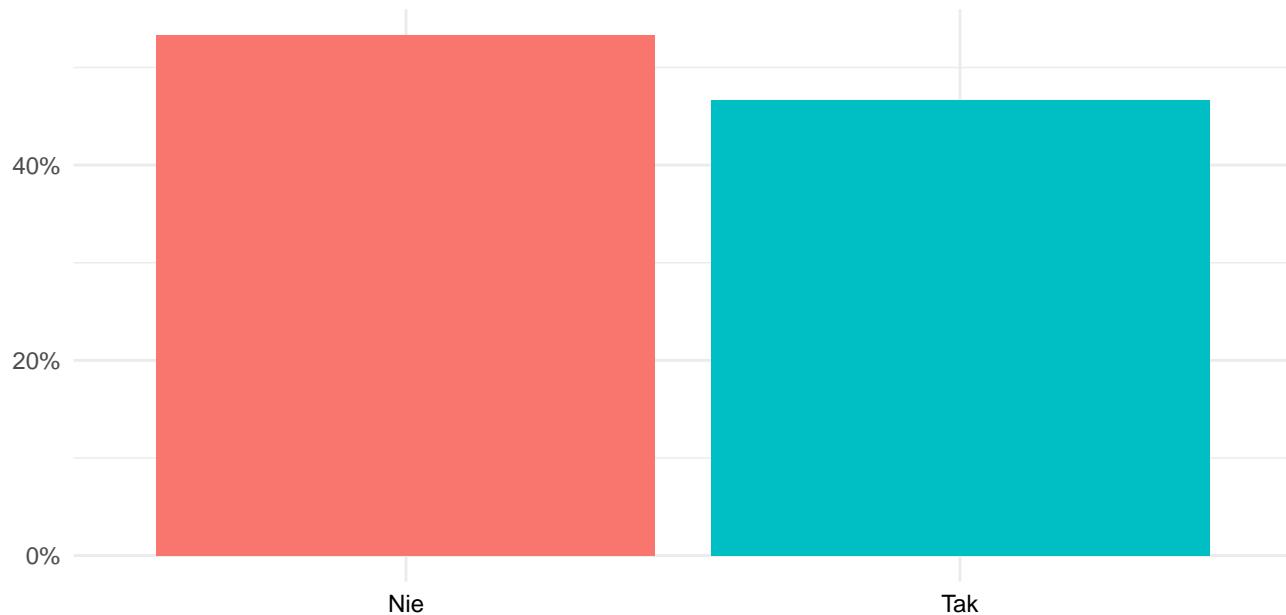
Na wykresie 10. przedstawiono wykres słupkowy dla zmiennej *Dependents*. Wynika z niego, że większość klientów (około 70%) nie ma osób na swoim utrzymaniu. Może to sugerować, że znaczna część użytkowników to osoby samodzielne finansowo, co może wpływać na ich preferencje dotyczące usług – mogą częściej wybierać bardziej elastyczne opcje lub krótsze zobowiązania, niekoniecznie kierując się potrzebami rodziny.



Wykres 11: Wykres słupkowy zmiennej *PhoneService*

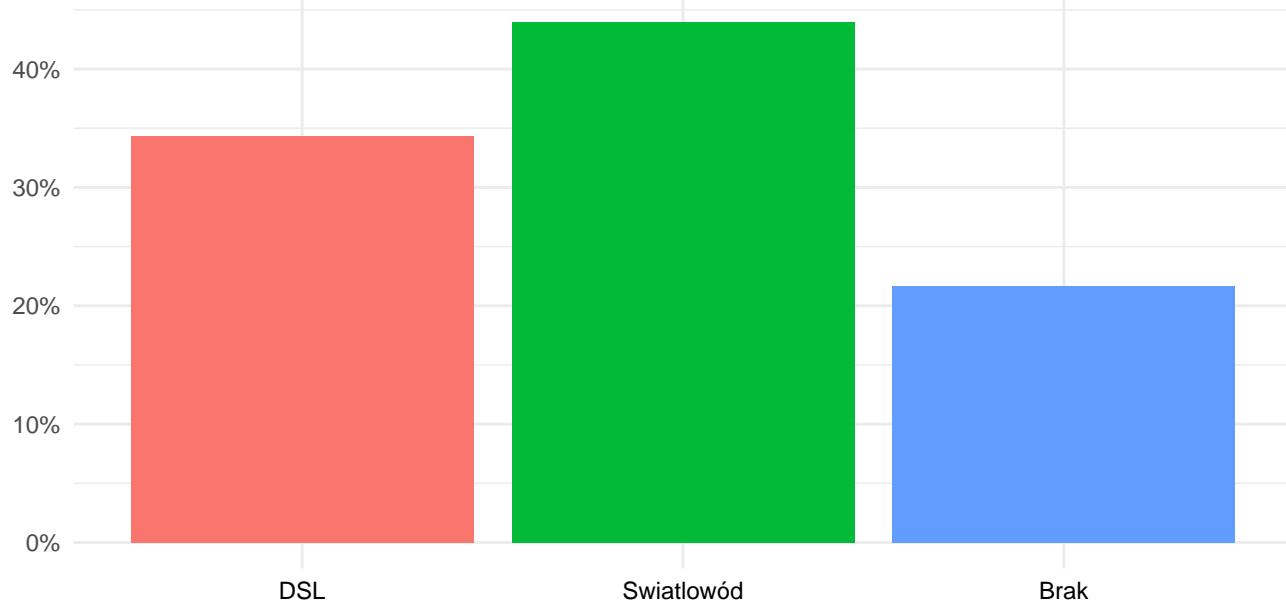
Na wykresie 11. przedstawiono wykres słupkowy dla zmiennej *PhoneService*. Wynika z niego, że zdecydowana

większość klientów korzystała z usług telefonicznych, podczas gdy jedynie niewielka grupa (około 10%) nie miała aktywnej usługi telefonicznej. Może to sugerować, że telefonia jest nadal istotnym elementem oferty, choć pewna część klientów może preferować inne formy komunikacji.



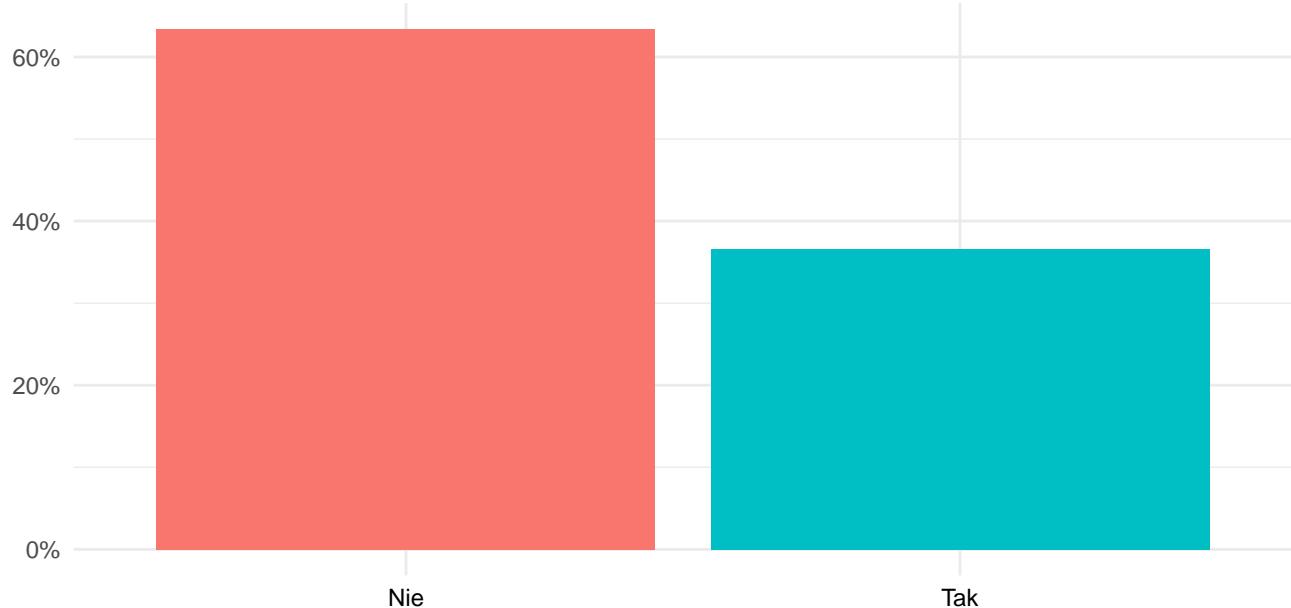
Wykres 12: Wykres słupkowy zmiennej *MultipleLines*

Na wykresie 12. przedstawiono wykres słupkowy dla zmiennej *MultipleLines* wśród klientów, którzy korzystają z usług telefonicznych. Wynika z niego, że wśród tych klientów niewielka większość posiada tylko jedną linię telefoniczną, podczas gdy nieco mniejsza grupa zdecydowała się na korzystanie z wielu linii. Może to sugerować, że dodatkowe linie telefoniczne nie cieszą się dużą popularnością, co może być efektem indywidualnych potrzeb użytkowników lub specyfiki gospodarstw domowych, które nie wymagają więcej niż jednej linii telefonicznej.



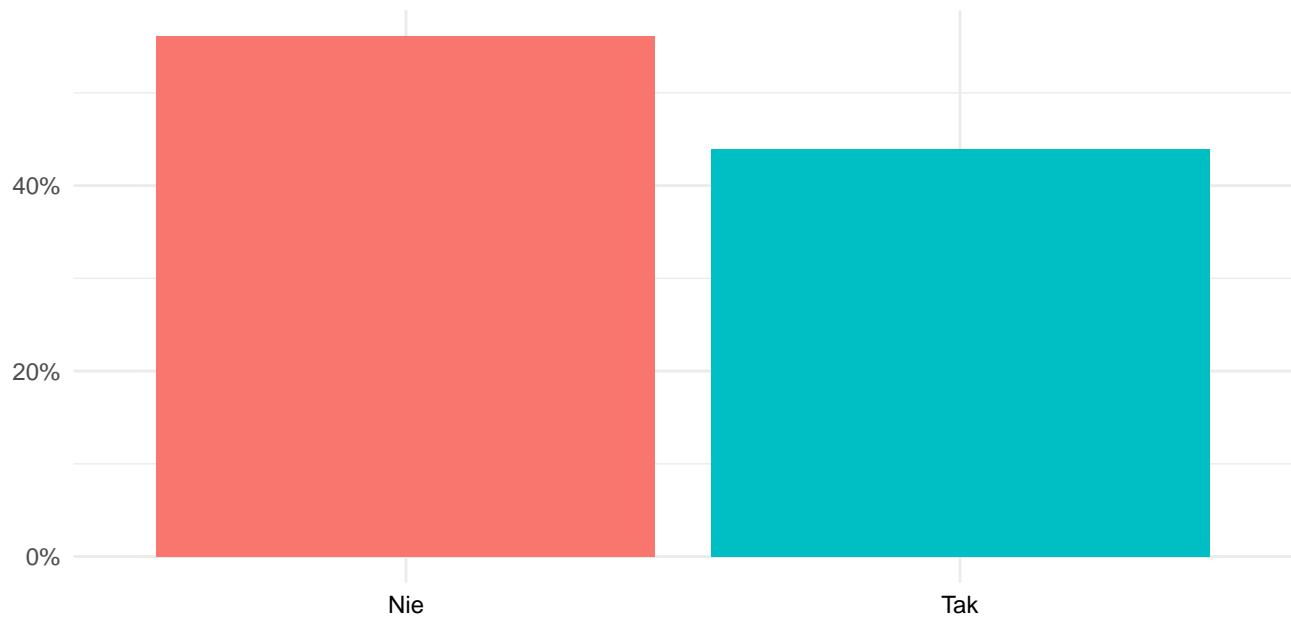
Wykres 13: Wykres słupkowy zmiennej *InternetService*

Na wykresie 13. przedstawiono wykres słupkowy dla zmiennej *InternetService*. Wynika z niego, że około 22% naszych klientów nie korzysta z usług internetowych, natomiast większość (około 43%) korzysta z rozwiązań światłowodowych. Pozostałe 35% korzysta z połączenia DSL. To wskazuje na różnorodność preferencji w zakresie usług internetowych, z dominacją technologii światłowodowych, które mogą oferować szybsze i bardziej stabilne połączenia.



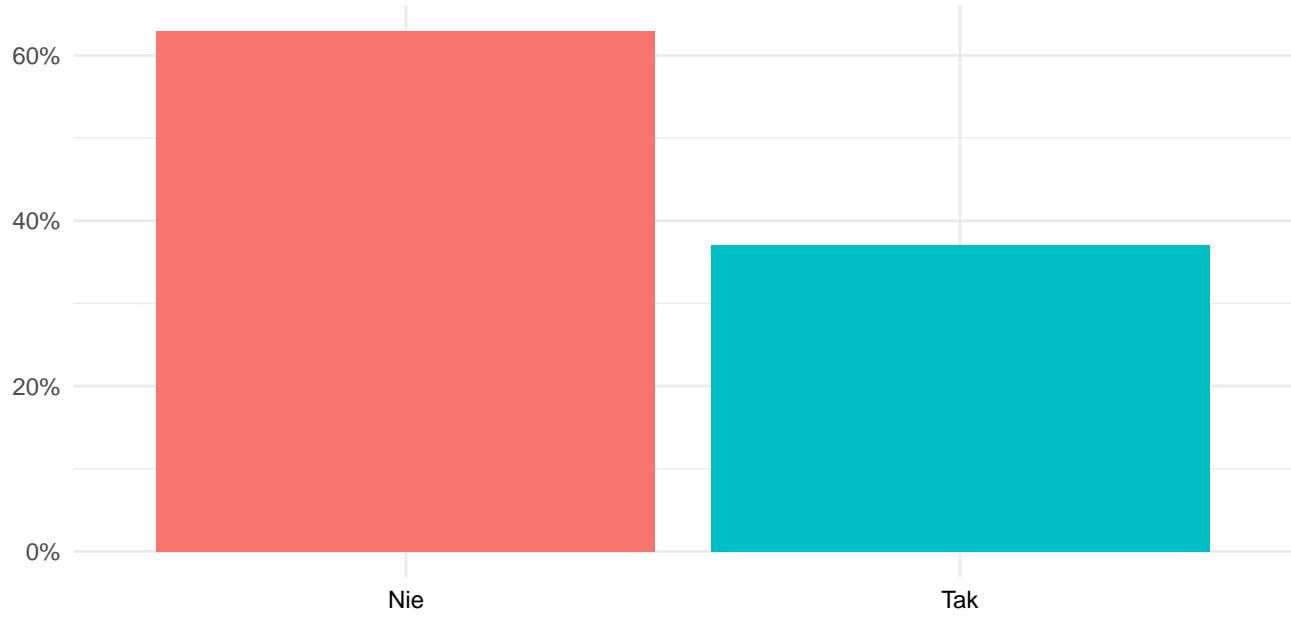
Wykres 14: Wykres słupkowy zmiennej *OnlineSecurity*

Na wykresie 14. przedstawiono wykres słupkowy dla zmiennej *OnlineSecurity* wśród klientów, którzy korzystają z usług internetowych. Widać, że większość z nich nie korzysta z tej usługi, co może sugerować, że klienci nie zwracają wystarczającej uwagi na bezpieczeństwo w internecie. Może to mieć reperkusje w przyszłości, gdyż brak odpowiedniej ochrony online może prowadzić do większego ryzyka związanych z bezpieczeństwem, co w konsekwencji może skutkować rezygnacją z naszej usługi.



Wykres 15: Wykres słupkowy zmiennej *DeviceProtection*

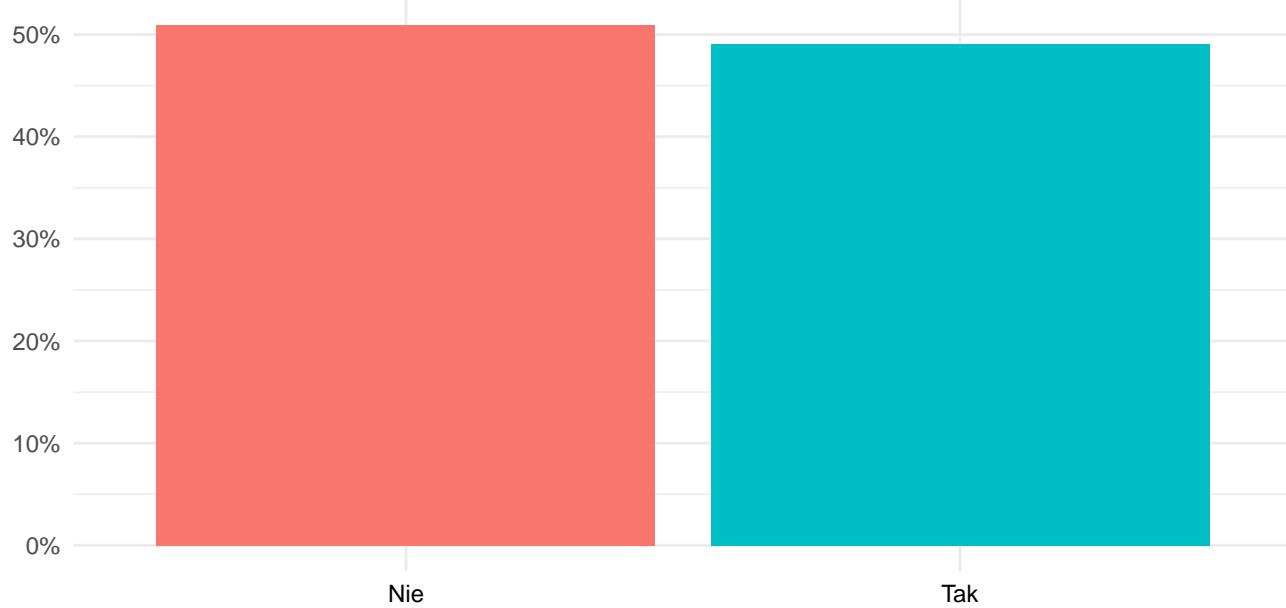
Na wykresie 15. przedstawiono wykres słupkowy dla zmiennej *DeviceProtection* wśród klientów, którzy korzystają z usług internetowych. Wynika z niego, że podobnie jak w przypadku zmiennej *OnlineSecurity*, większość klientów nie korzysta z tej usługi. Taki brak zainteresowania ochroną urządzeń może prowadzić do problemów związanych z bezpieczeństwem, co w przyszłości może przyczynić się do utraty klientów.



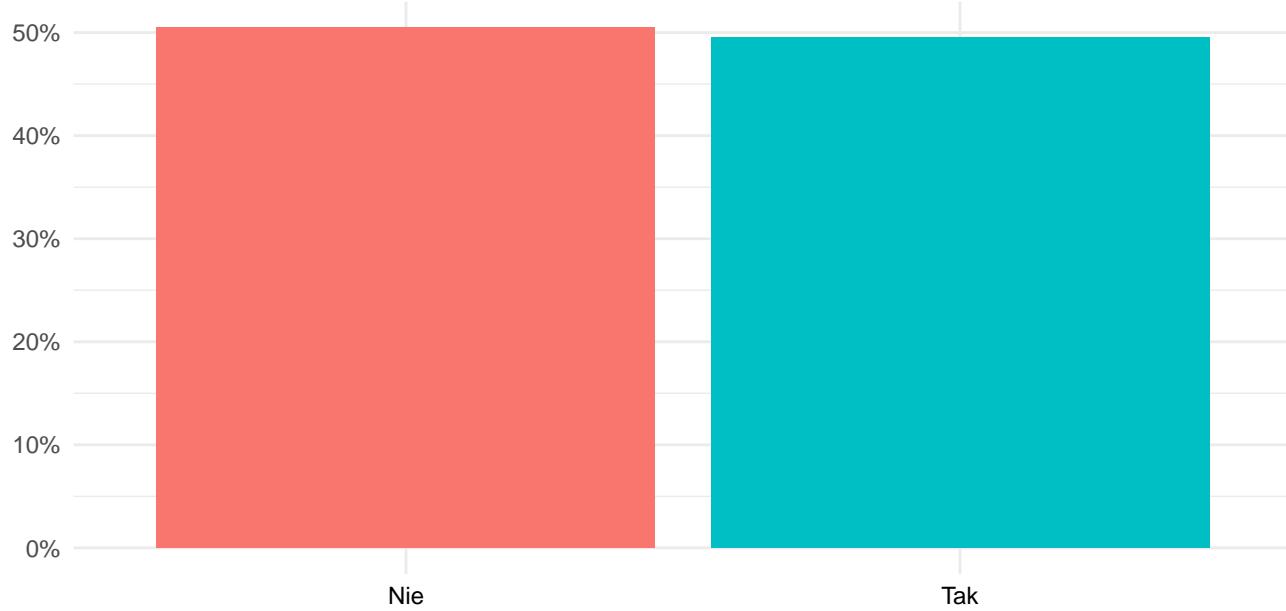
Wykres 16: Wykres słupkowy zmiennej *TechSupport*

Na wykresie 16. przedstawiono wykres słupkowy dla zmiennej *TechSupport* wśród klientów, którzy korzystają z usług internetowych. Wynika z niego, że większość klientów nie korzysta z tej usługi. Może to sugerować, że klienci nie czują potrzeby wsparcia technicznego lub są w stanie radzić sobie z problemami samodzielnie. Niemniej jednak, brak korzystania z tej usługi może również wskazywać na potencjalne ryzyko w przyszłości,

gdyby klienci napotkali trudności, które wymagałyby pomocy technicznej, co mogłoby prowadzić do ich niezadowolenia lub odejścia od usługi.

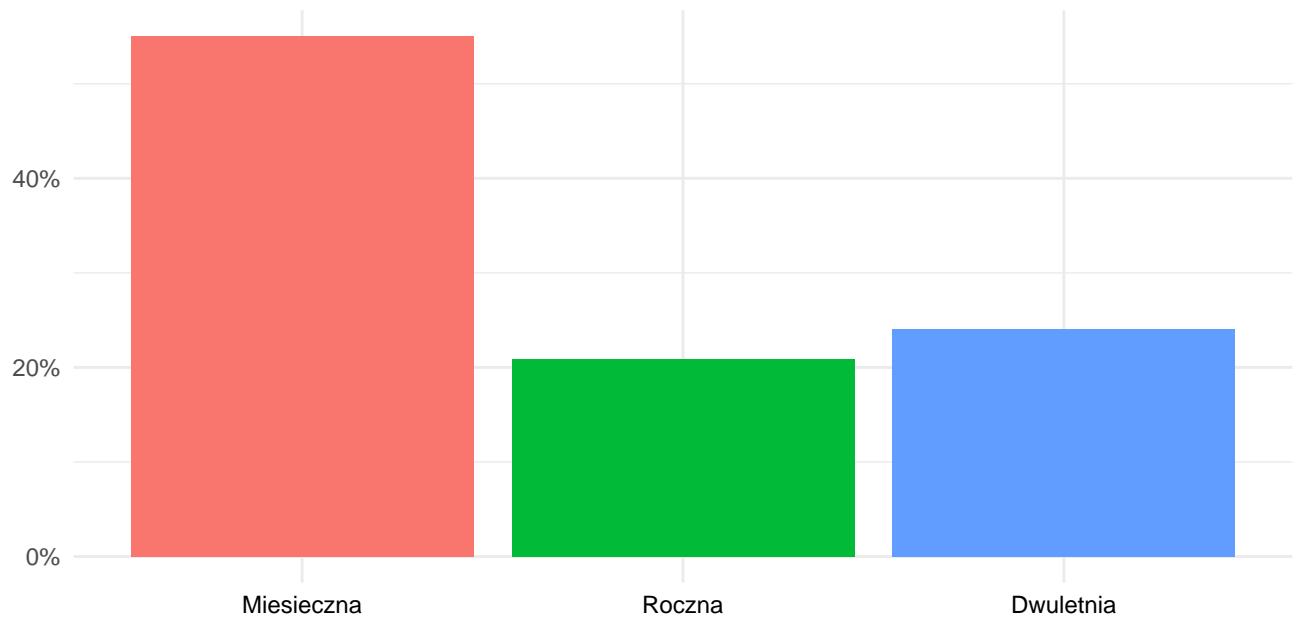


Wykres 17: Wykres słupkowy zmiennej *StreamingTV*



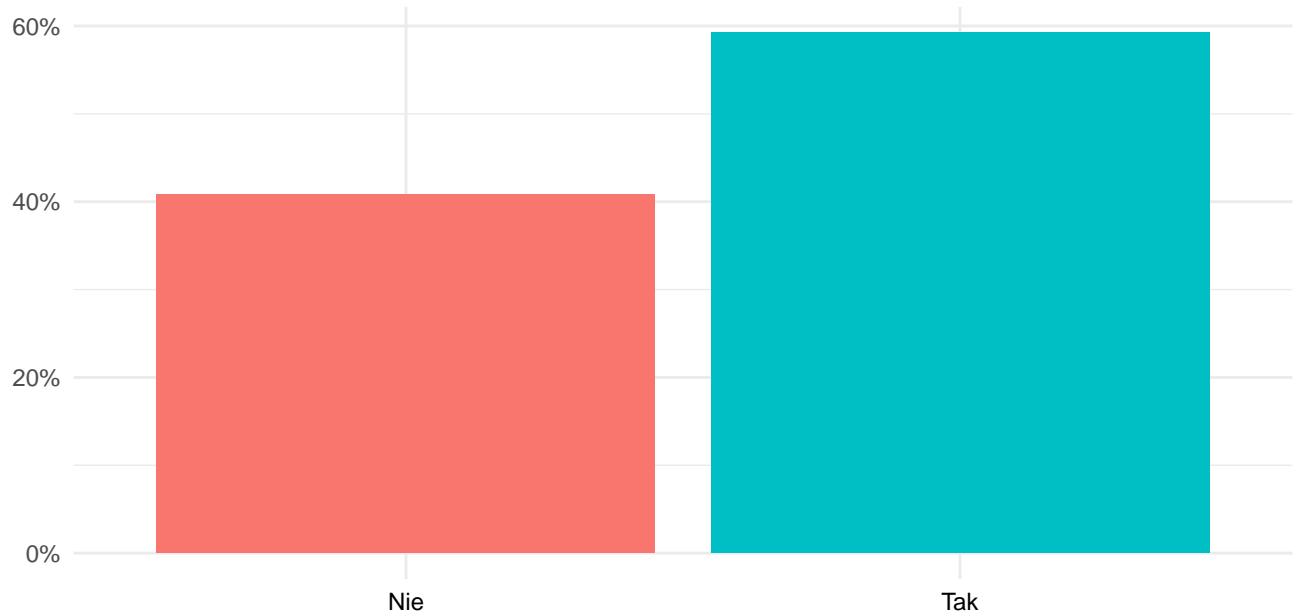
Wykres 18: Wykres słupkowy zmiennej *StreamingMovies*

Na wykresach 17. i 18. przedstawiono wykresy słupkowe dla zmiennych *StreamingTV* oraz *StreamingMovies*. Wynika z nich, że równa liczba klientów korzysta z tych udogodnień. Jednakże, jak wskazuje paradoks Simpsona, może to się zmienić w przypadku dogłębszej analizy, szczególnie gdy uwzględnimy zmienną *churn*. Może się okazać, że w różnych grupach klientów (np. o różnym statusie "churn") korzystanie z usług streamingowych ma różne zależności, co może wpływać na rzeczywisty obraz zachowań użytkowników w kontekście rezygnacji z usług.



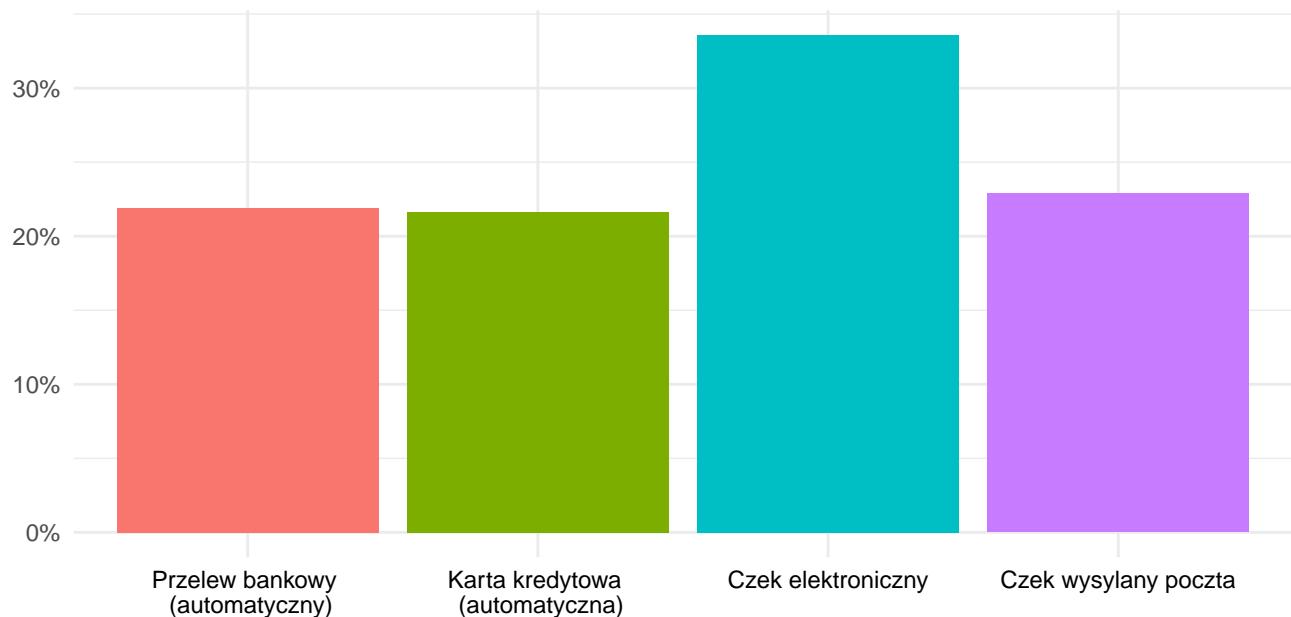
Wykres 19: Wykres słupkowy zmiennej *Contract*

Na wykresie 19. przedstawiono wykres słupkowy dla zmiennej *Contract*. Wynika z niego, że większość klientów preferuje kontrakt miesięczny. Opcje roczne i dwuletnie stanowią około 20% każda, przy czym więcej osób wybiera kontrakt dwuletni niż roczny. Taki rozkład może sugerować, że klienci cenią sobie elastyczność miesięcznych umów, co może wynikać z chęci przetestowania usług firmy lub skorzystania z krótkoterminowych korzyści. Jednocześnie część klientów decyduje się na długoterminowe zobowiązania, oczekując w zamian lepszych warunków, takich jak niższe opłaty miesięczne czy dodatkowe korzyści.



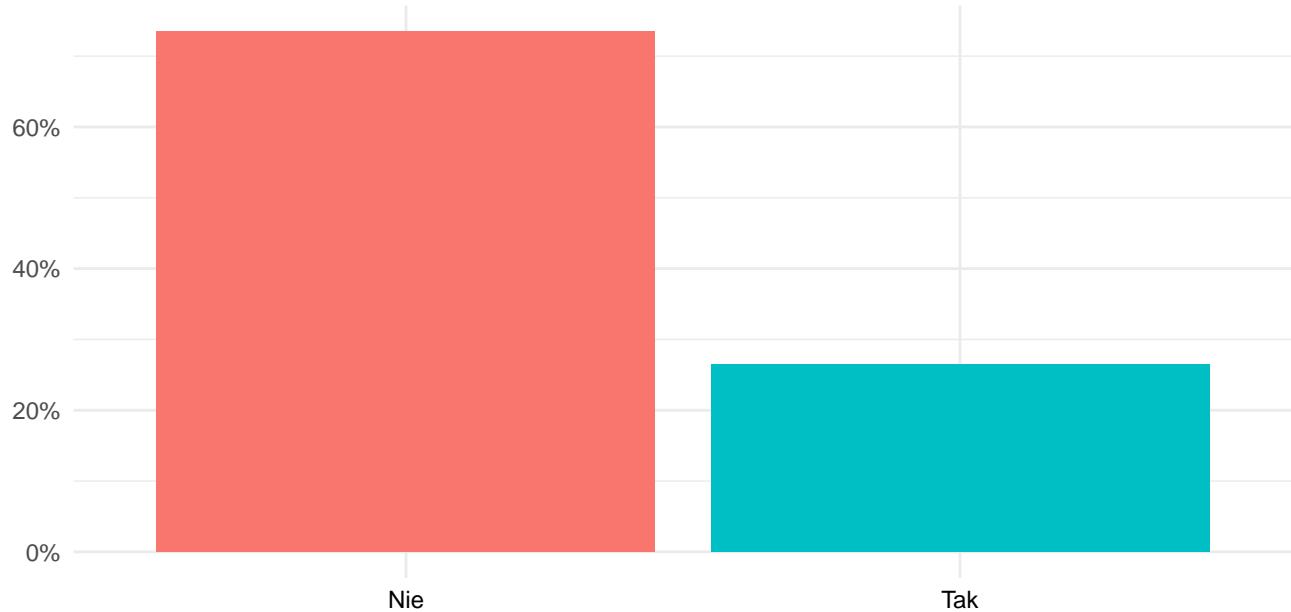
Wykres 20: Wykres słupkowy zmiennej *PaperlessBilling*

Na wykresie 20. przedstawiono wykres słupkowy dla zmiennej *PaperlessBilling*. Wynika z niego, że większość klientów preferuje rozwiązanie bez papierowych faktur.



Wykres 21: Wykres słupkowy zmiennej *PaymentMethod*

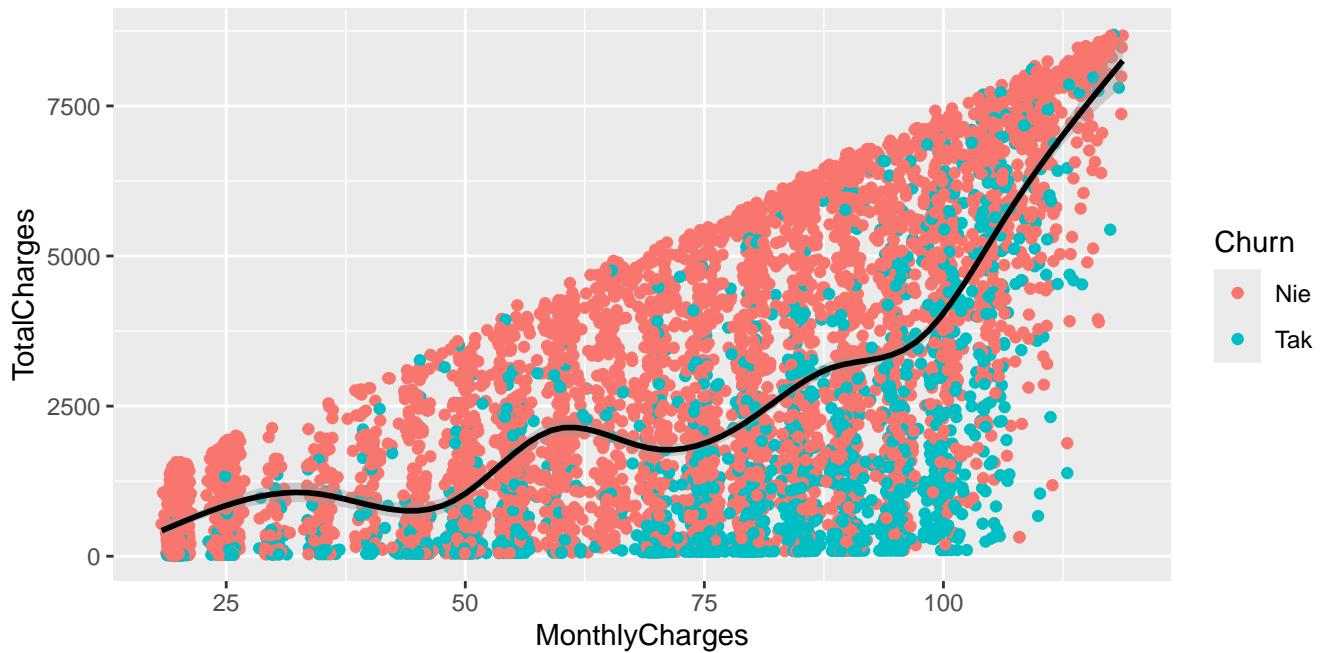
Na wykresie 21. przedstawiono wykres słupkowy dla zmiennej *PaymentMethod*. Widać, że większość osób preferuje płatności za pomocą czeku elektronicznego. Pozostałe metody płatności cieszą się podobną popularnością, co może sugerować, że klienci nie mają wyraźnej preferencji wobec innych opcji płatności, ale cenią sobie różnorodność dostępnych metod.



Wykres 22: Wykres słupkowy zmiennej *Churn*

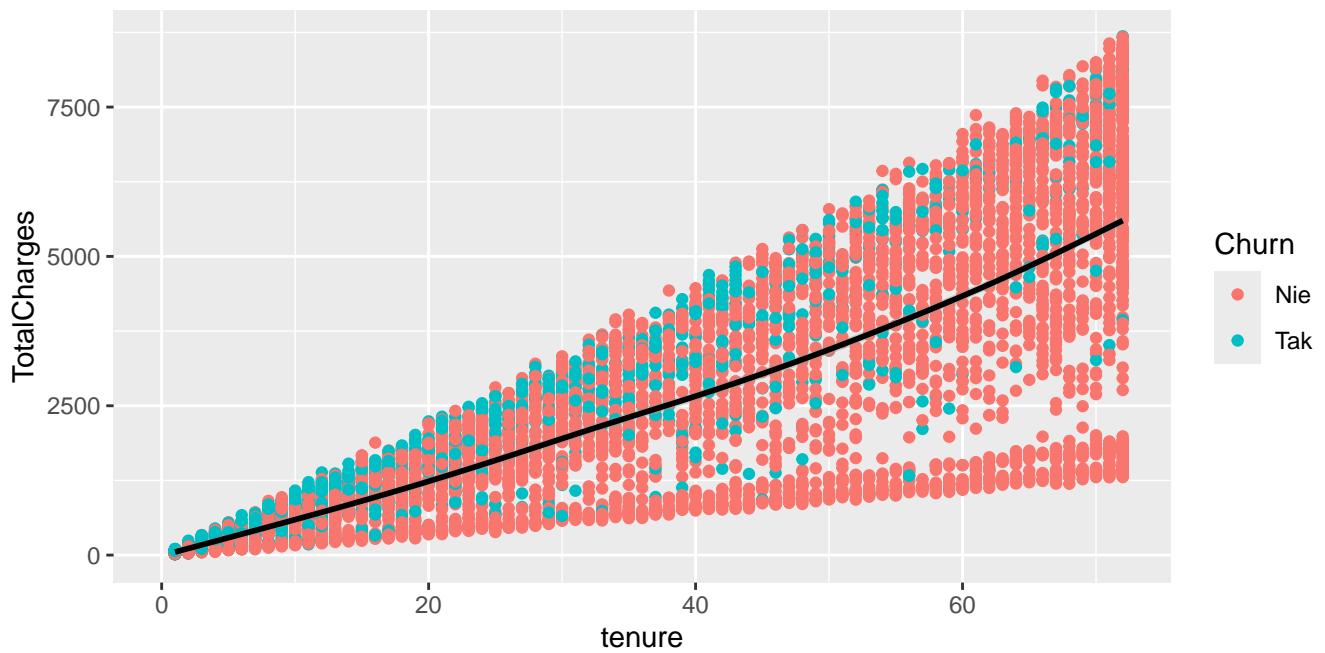
Na wykresie 22. przedstawiono wykres słupkowy dla zmiennej *Churn*. Widać, że większość klientów pozostaje lojalna i nie zrezygnowała z umowy, jednak około 25% zdecydowało się na jej rozwiązanie.

5.4 Wykresy rozrzutów



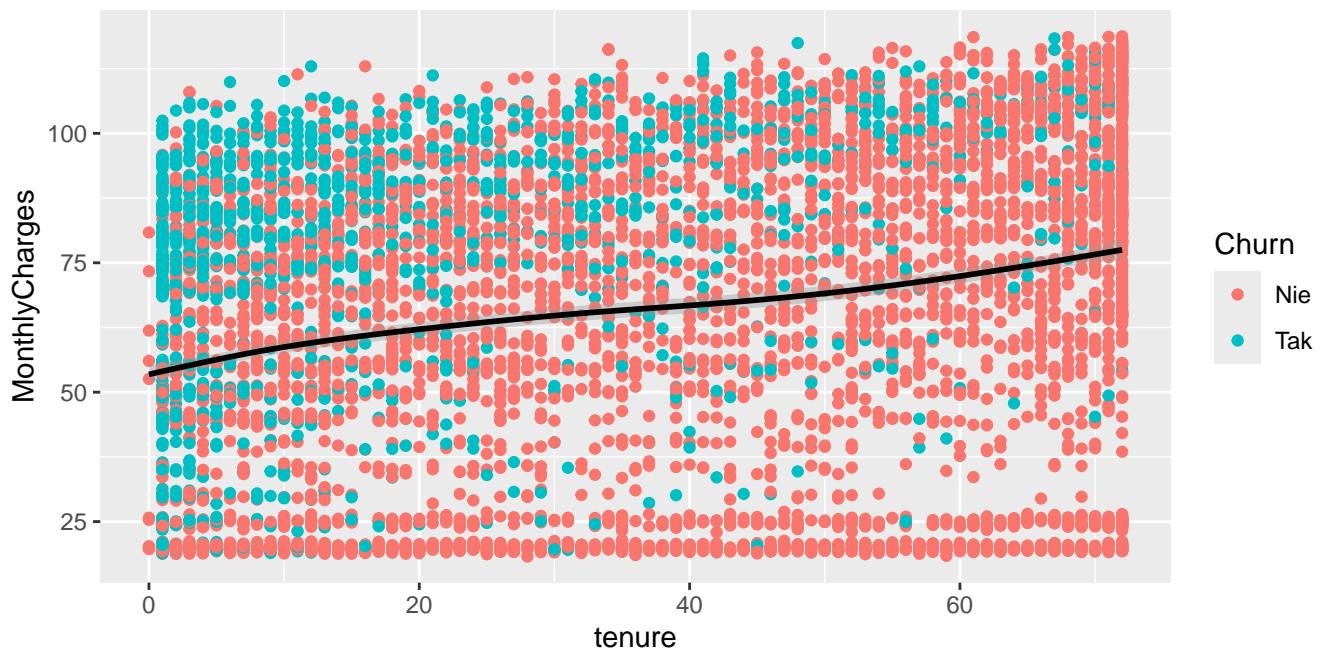
Wykres 23: Zależność TotalCharges od MonthlyCharges z krzywą wygładzającą

Na wykresie 23. przedstawiono zależność *TotalCharges* od *MonthlyCharges* z krzywą wygładzającą, a dane zostały rozdzielone na klientów, którzy odeszli, oraz tych, którzy pozostały. Widać, że istnieje liniowa zależność między maksymalnymi wartościami *TotalCharges* a *MonthlyCharges*, która rośnie proporcjonalnie. Krzywa wygładzająca wskazuje, że dane rozkładają się głównie w średnim zakresie *TotalCharges*, jednak powyżej kwoty 100 dolarów miesięcznie zaczyna się ona zwiększać w sposób liniowy, osiągając wartość maksymalną. Co istotne, nie widać klientów, którzy płacą wysokie miesięczne opłaty, a jednocześnie mają niskie *TotalCharges*. Może to sugerować, że tacy klienci po prostu nie istnieją w zbiorze danych.



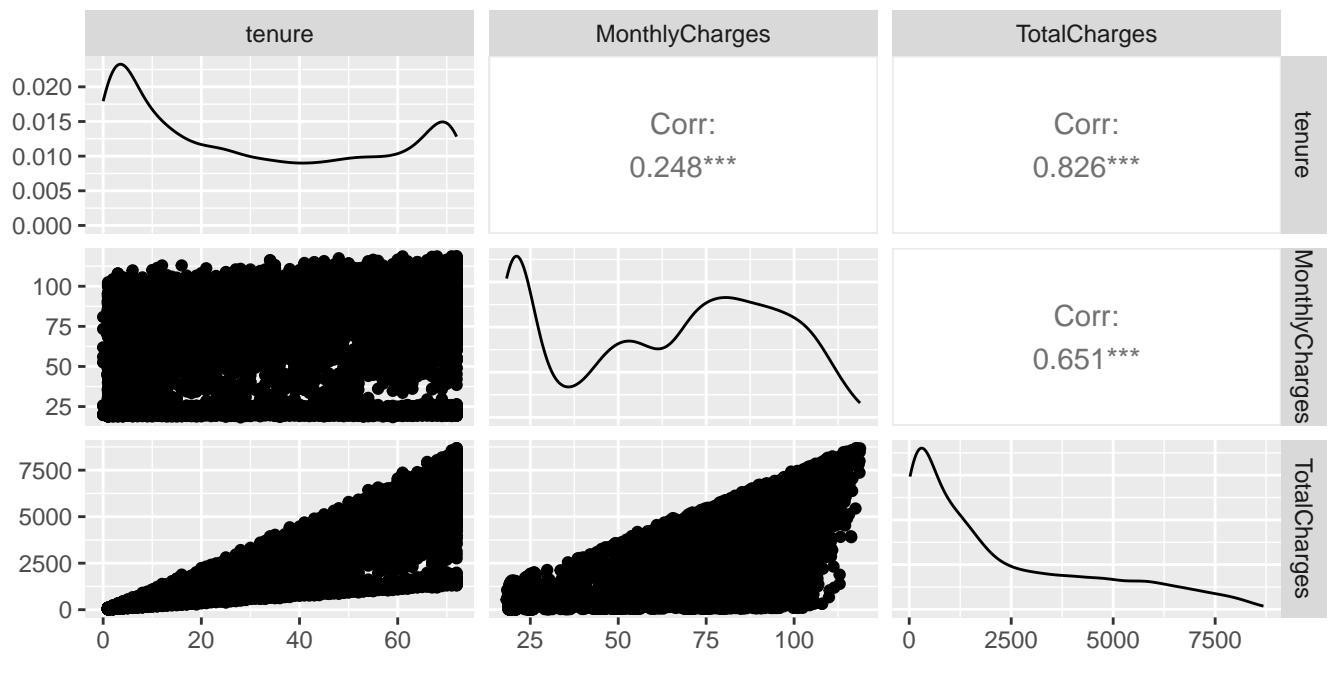
Wykres 24: Zależność TotalCharges od tenure z krzywą wygładzającą

Na wykresie 24. przedstawiono zależność *TotalCharges* od *tenure* z krzywą wygładzającą, a dane zostały rozdzielone na klientów, którzy odeszli, oraz tych, którzy pozostały. Widać liniową zależność między minimalnymi, maksymalnymi i średnimi wartościami *TotalCharges* a *tenure*, co sugeruje, że wraz z długością trwania umowy klienci generują większe całkowite opłaty. Widać również brak klientów o długiej lojalności, którzy jednocześnie mają lekko ponad minimalne wydatki całkowite. Może to wskazywać, że lojalni klienci albo korzystają z minimalnych usług, albo wybierają droższe opcje oferowane przez firmę, co może świadczyć o ich skłonności do długoterminowego zaangażowania w usługi, które są bardziej kosztowne, ale dostosowane do ich potrzeb.



Wykres 25: Zależność MonthlyCharges od tenure z krzywą wygładzającą

Na wykresie 25. przedstawiono zależność między wartością *MonthlyCharges* a *tenure*, z nałożoną krzywą wygładzającą, a dane zostały podzielone na klientów, którzy zrezygnowali z usług, oraz tych, którzy pozostali. Widać pełen przekrój klientów, korzystających z naszych usług, obejmujący różne kwoty miesięcznych opłat oraz różnych czas lojalności. Zauważalna jest dodatnia zależność liniowa między średnią a maksymalną wartością *MonthlyCharges*. Klienci korzystający z minimalnych usług są obecni w każdym segmencie. Podobnie jak na wykresie 23., widać rozrzedzenie liczby klientów z długim stażem oraz wydatkami na usługi powyżej minimum. Sugeruje to, że klienci, którzy pozostali z nami na dłużej, albo korzystają z podstawowych usług, albo zdecydowali się na zakup dodatkowych usług.

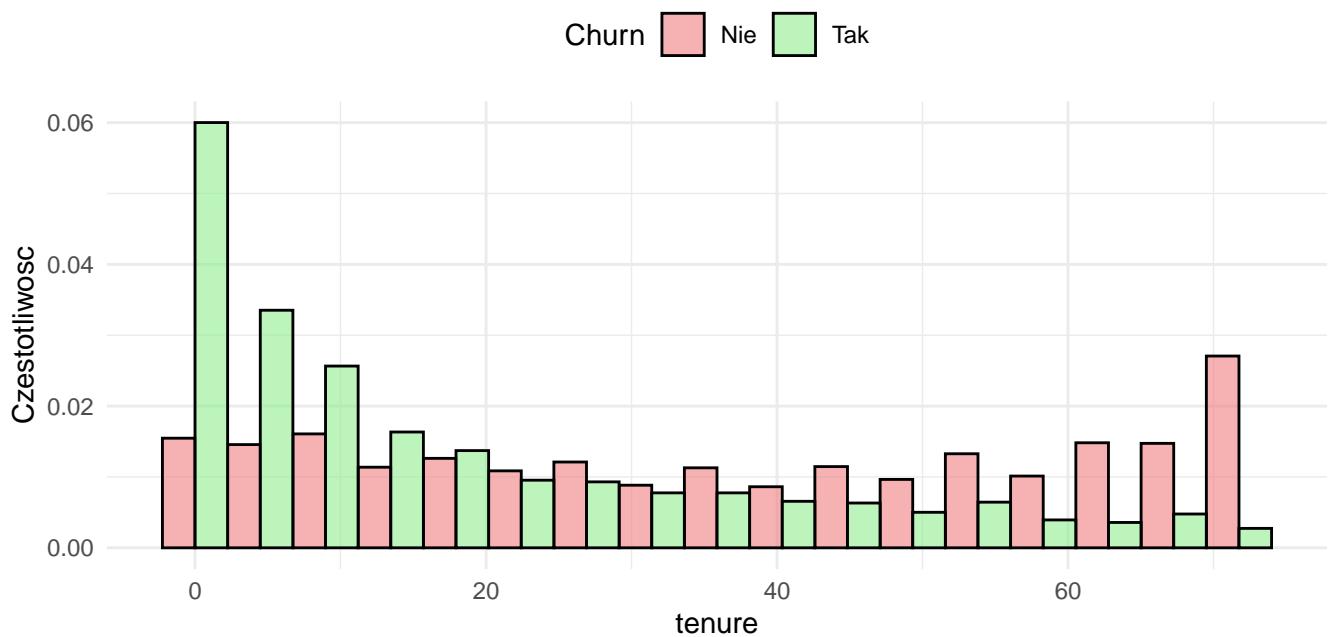


Wykres 26: Macierz par zmiennych numerycznych

Na wykresie 26. zaprezentowano macierz par zmiennych numerycznych, w której zidentyfikowano silną korelację liniową (0,826) między zmiennymi *TotalCharges* a *tenure*.

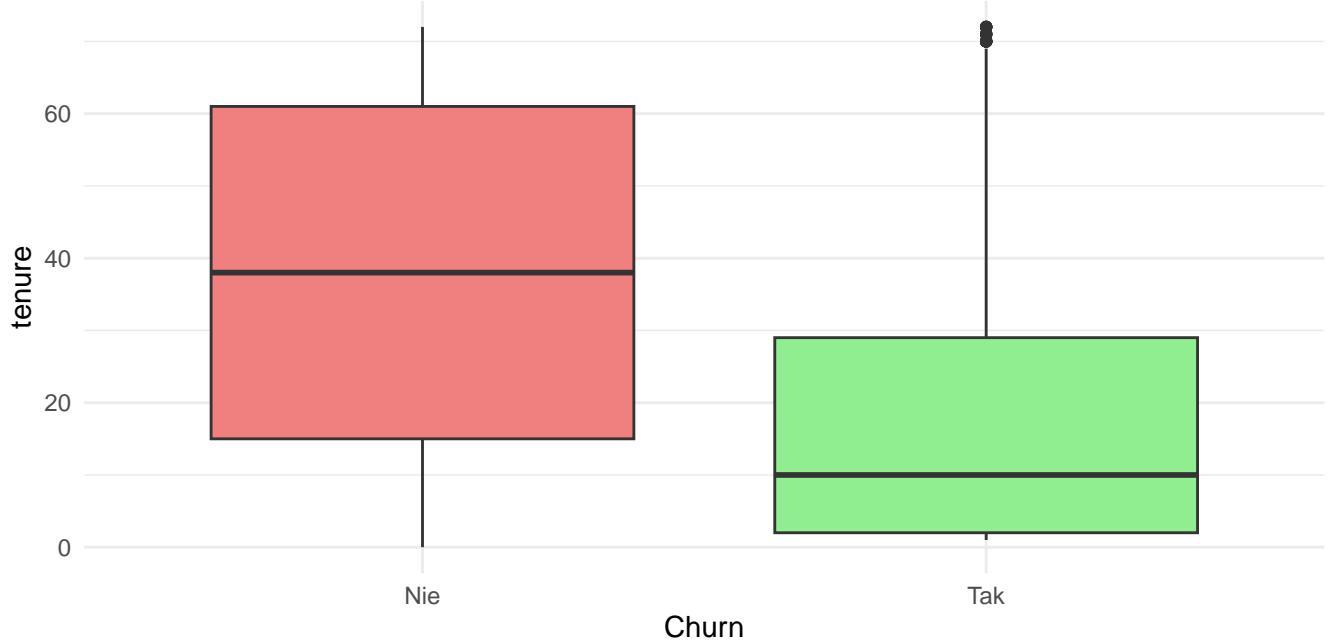
6 Analiza opisowa z podziałem na grupy

6.1 Histogramy i wykresy pudełkowe



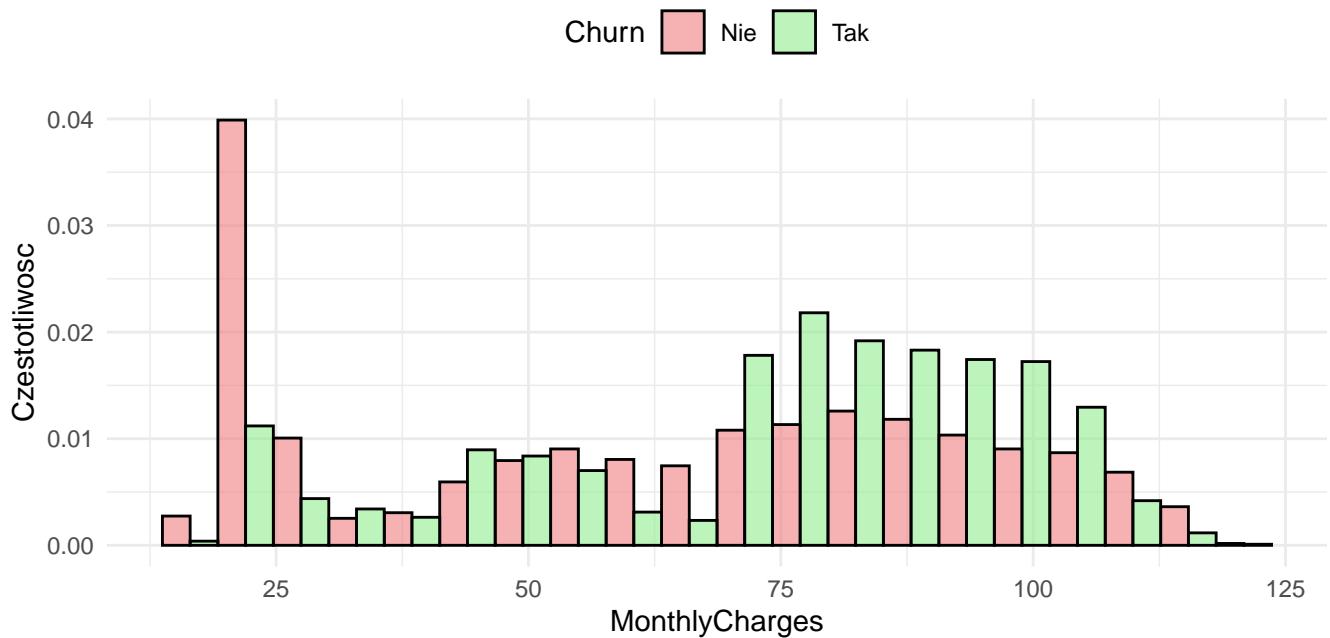
Wykres 27: Histogram dla zmiennej tenure z podziałem ze względu na churn

Na wykresie 27. przedstawiono histogram dla zmiennej *tenure*, z podziałem na klientów, którzy odeszli (churn) i tych, którzy pozostali. Podobnie jak w przypadku ogólnego rozkładu, widać równomierny rozkład dla osób, które zostały, z wyraźnym szczytem dla klientów o dłuższym stażu. Wypłaszczenie się górnego dla niższych wartości jest wynikiem uwzględnienia klientów, którzy odeszli. Ich rozkład przypomina rozkład wykładniczy, co może sugerować, że ci klienci korzystali z naszych usług przez pierwsze miesiące, jednak nie zostali przekonani, co prowadziło do ich stosunkowo szybkiego odejścia.



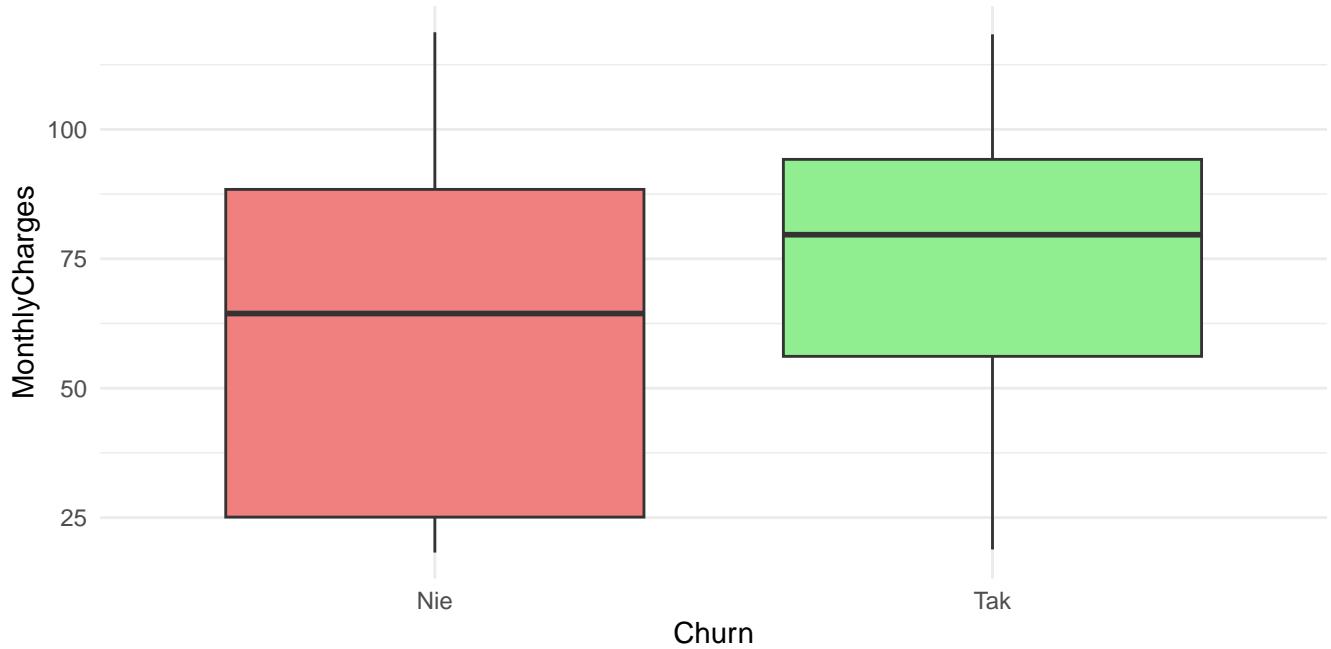
Wykres 28: Wykres pudełkowy dla zmiennej *tenure* z podziałem ze względu na churn

Na wykresie 28. zaprezentowano wykres pudełkowy dla zmiennej *tenure*, z podziałem na klientów, którzy odeszli (churn) i tych, którzy pozostali. Widać, że klienci, którzy zrezygnowali z usług, korzystali z nich przez krótki okres, podczas gdy ci, którzy pozostali, są lojalni przez dłuższy czas. Zauważalne są również odstające dane w grupie klientów, którzy odeszli, co może wynikać z losowego zakończenia współpracy.



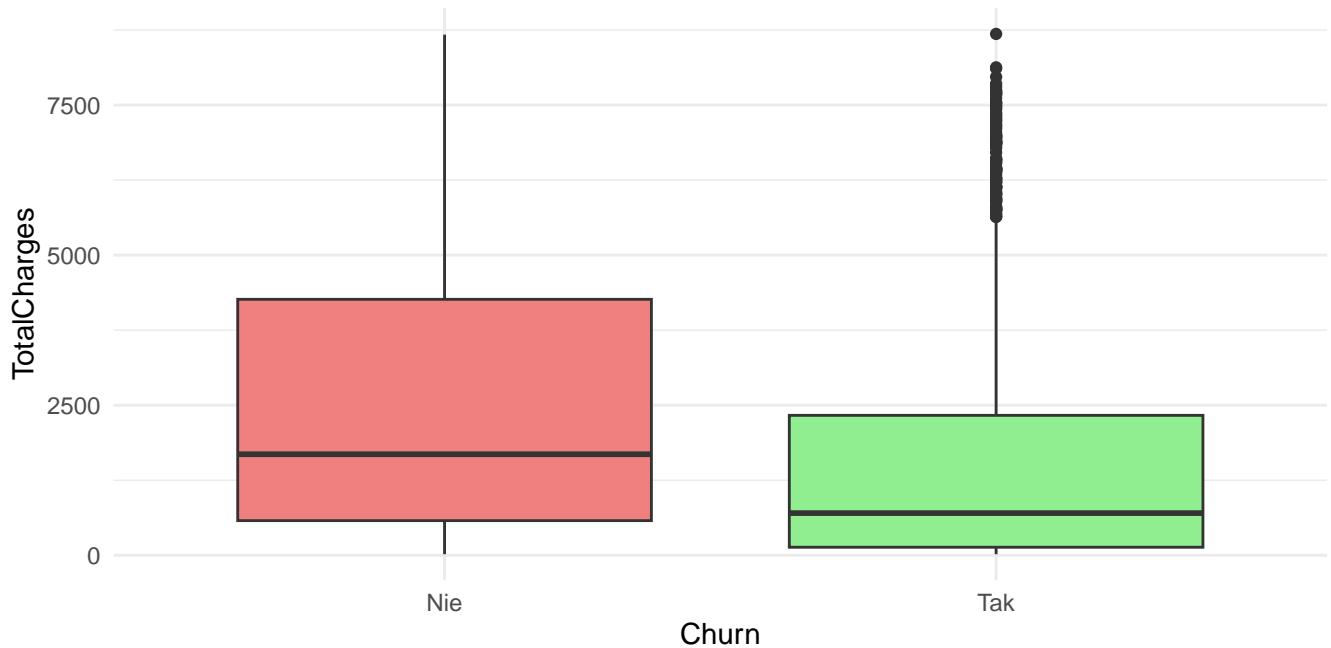
Wykres 29: Histogram dla zmiennej *MonthlyCharges* z podziałem ze względu na churn

Na wykresie 29. przedstawiono histogram dla zmiennej *MonthlyCharges*, z podziałem na klientów, którzy odeszli (churn) i tych, którzy pozostali. Widać, że klienci dzielą się na trzy grupy: tych, którzy płacą minimalnie lub nieco ponad, tych, którzy płacą około 50 dolarów, oraz tych, którzy płacą powyżej 60 dolarów. Zauważalny jest również pik na niewielkich kwotach w przypadku klientów, którzy pozostali, co sugeruje, że mamy wielu stałych klientów korzystających z podstawowych usług. Natomiast dla kwot powyżej 60 dolarów widać, że liczba klientów, którzy odeszli, jest większa niż tych, którzy pozostali. Może to sugerować, że droższe usługi oferowane tym klientom nie przypadły im do gustu, lub że potrzebowali ich tylko przez określony czas.



Wykres 30: Wykres pudełkowy dla zmiennej *MonthlyCharges* z podziałem ze względu na churn

Na wykresie 30. zaprezentowano wykres pudełkowy dla zmiennej *MonthlyCharges*, z podziałem na klientów, którzy odeszli (churn) i tych, którzy pozostali. Widać, że średnia wartość miesięcznych opłat dla klientów, którzy pozostali, mieści się w przedziale od 25 do 88 dolarów. Natomiast dla klientów, którzy odeszli, środkowa część rozkładu pokazuje wyższe kwoty opłat. Sugeruje to, że klienci, którzy pozostali, płacili średnio mniej, podczas gdy ci, którzy odeszli, zazwyczaj korzystali z droższych usług.



Wykres 31: Wykres pudełkowy dla zmiennej *TotalCharges* z podziałem ze względu na churn

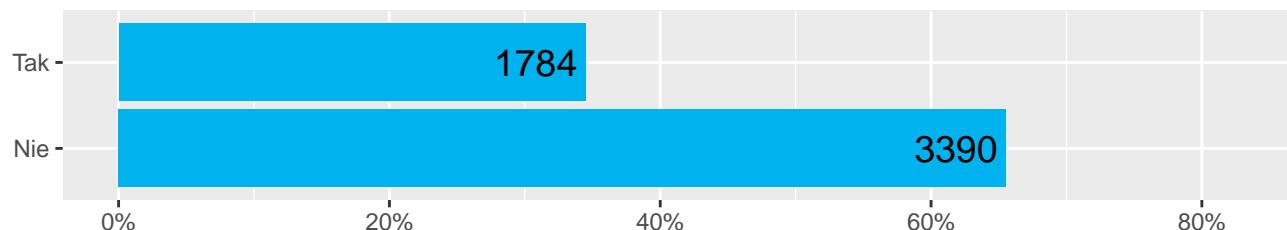
Na wykresie 31. zaprezentowano wykres pudełkowy dla zmiennej *TotalCharges*, z podziałem na klientów,

którzy odeszli (churn) i tych, którzy pozostali. Widać, że część średkowa klientów, którzy odeszli, zostawiła u nas mniej pieniędzy niż ci, którzy pozostali.

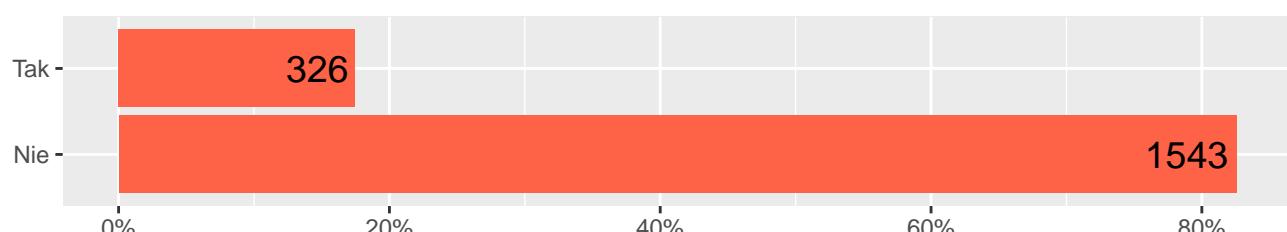
6.2 Wykresy słupkowe zmiennych jakościowych

Ponownie za pomocą for tworzymy wykresy słupkowe zmiennych jakościowych.

Obecni klienci



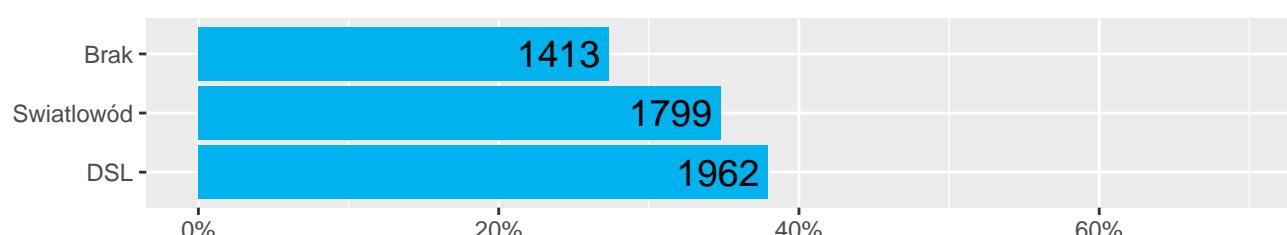
Dawni klienci



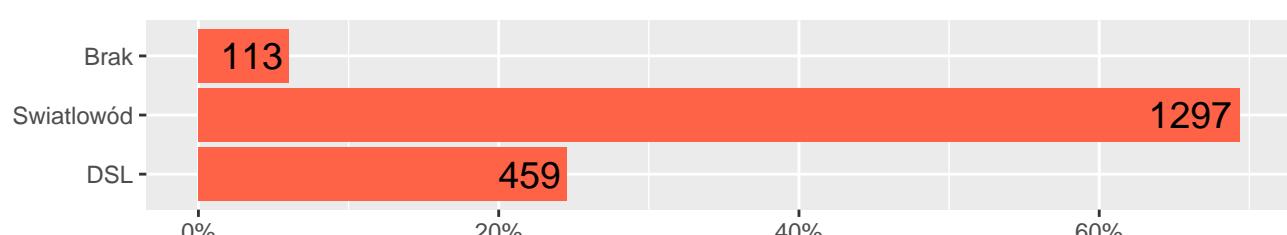
Wykres 32: Wykres słupkowy zmiennej Dependents z podziałem ze względu na churn

Na wykresie 32. przedstawiono wykres słupkowy dla zmiennej *dependants*, z podziałem na klientów, którzy odeszli (churn) i tych, którzy pozostali. Widać, że klienci, którzy odeszli, nie mają nikogo na utrzymaniu, co sugeruje, że osoby bez osób na utrzymaniu częściej decydują się na rezygnację z usług.

Obecni klienci



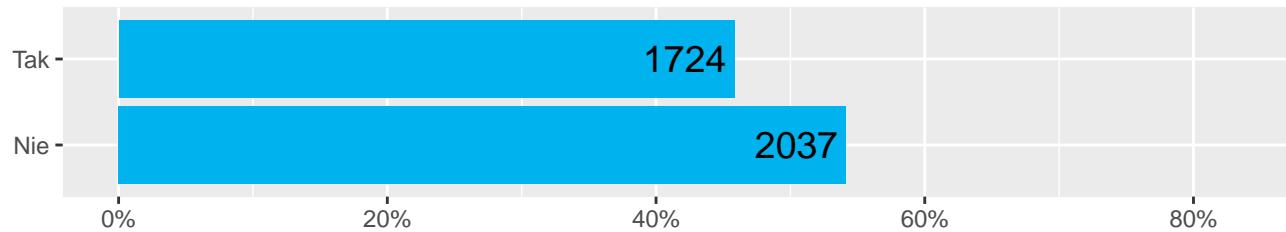
Dawni klienci



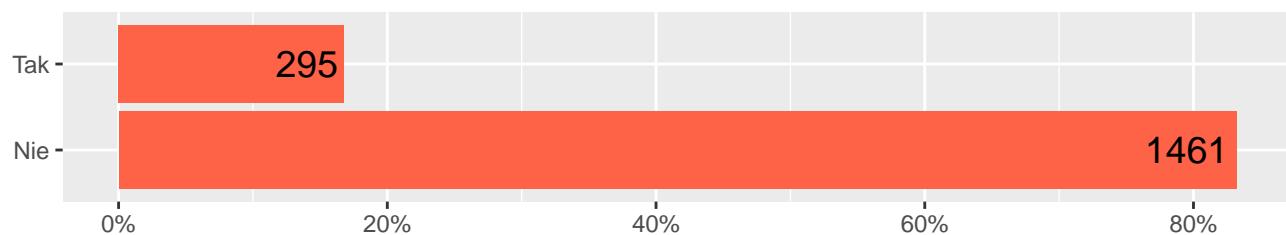
Wykres 33: Wykres słupkowy zmiennej InternetService z podziałem ze względu na churn

Na wykresie 33. przedstawiono wykres słupkowy dla zmiennej *dependants*, z podziałem na klientów, którzy odeszli (churn) i tych, którzy pozostali. Widać, że klienci, którzy odeszli, w większym stopniu korzystali z internetu i mieli dostęp do lepszych rozwiązań światłowodowych w porównaniu do tych, którzy pozostali.

Obecni klienci

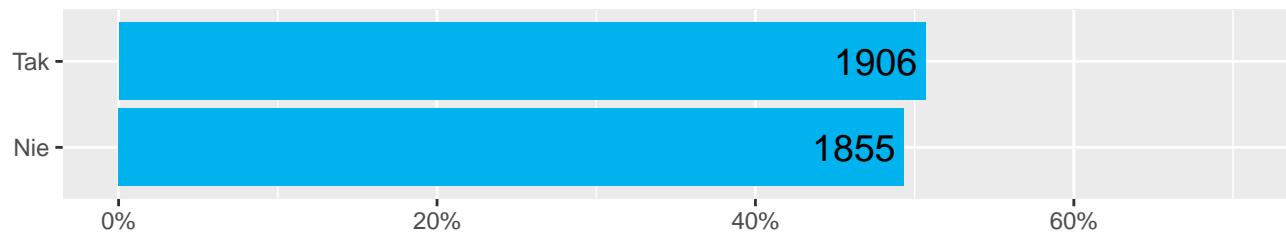


Dawni klienci

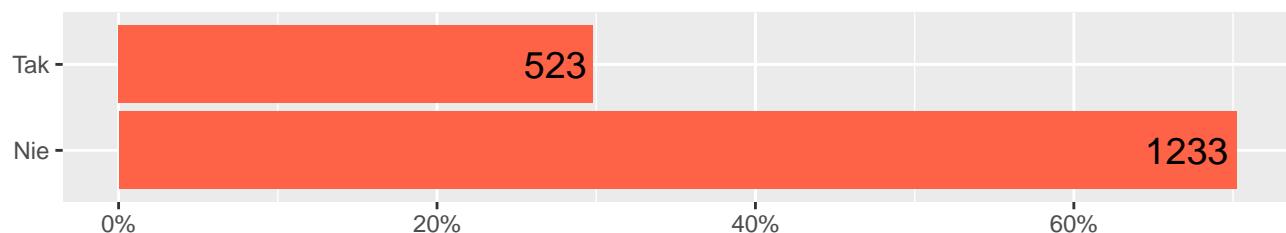


Wykres 34: Wykres słupkowy zmiennej OnlineSecurity z podziałem ze względu na churn

Obecni klienci

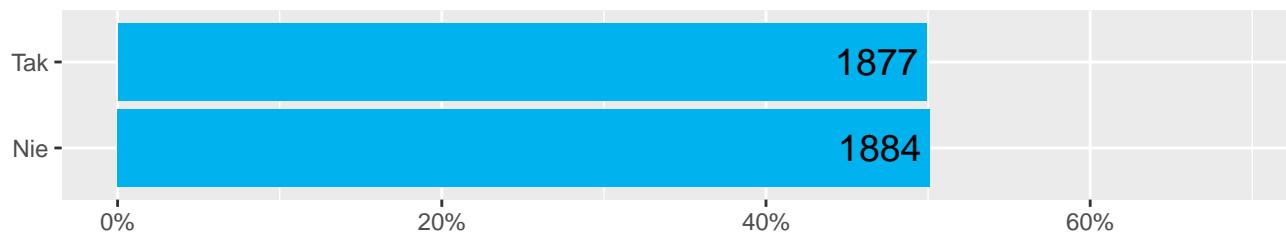


Dawni klienci

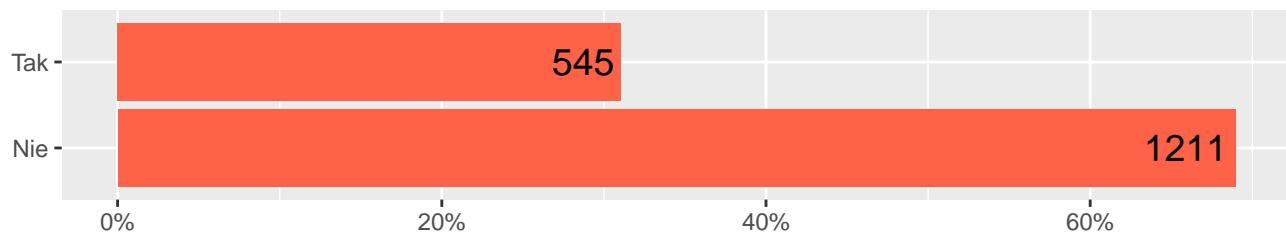


Wykres 35: Wykres słupkowy zmiennej OnlineBackup z podziałem ze względu na churn

Obecni klienci

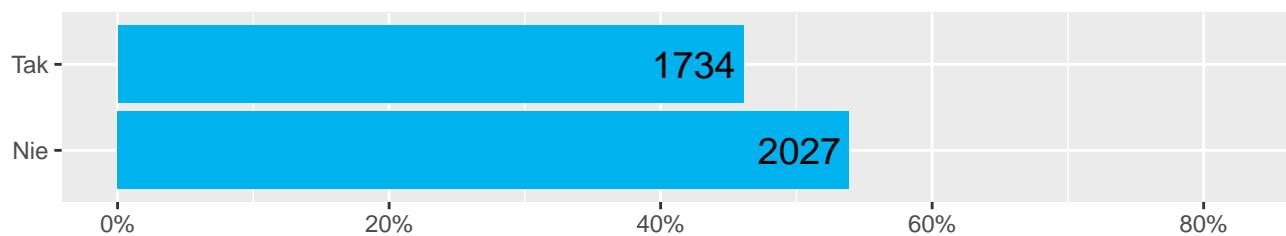


Dawni klienci

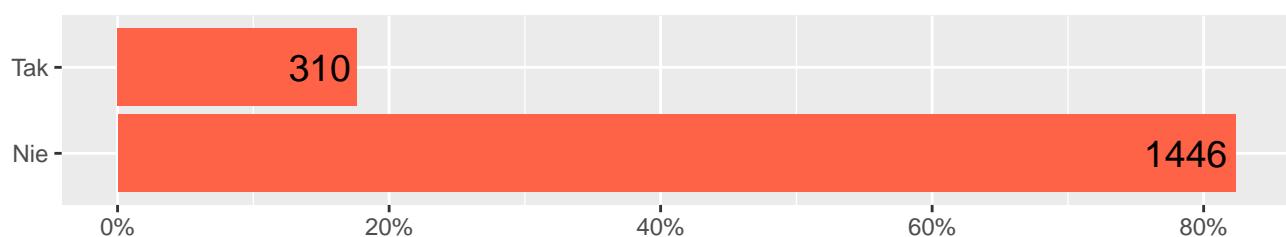


Wykres 36: Wykres słupkowy zmiennej DeviceProtection z podziałem ze względu na churn

Obecni klienci



Dawni klienci



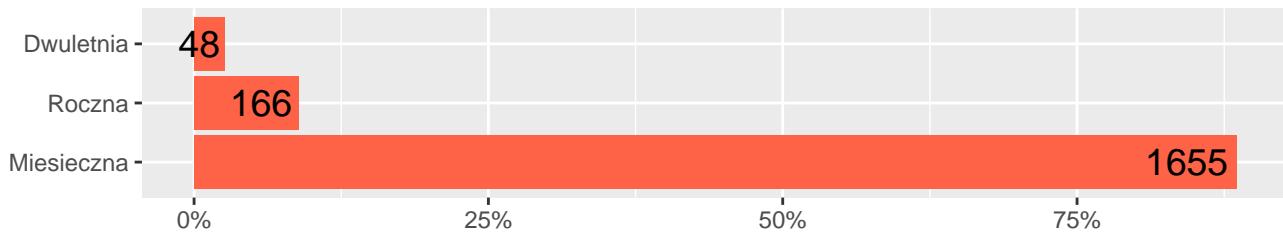
Wykres 37: Wykres słupkowy zmiennej TechSupport z podziałem ze względu na churn

Na wykresach 34, 35, 36 i 37 przedstawiono wykresy słupkowe dla zmiennych *OnlineSecurity*, *OnlineBackup*, *Device Protection* oraz *TechSupport*, z podziałem na klientów, którzy odeszli (churn) i tych, którzy pozostali. Widać, że w większości klientów, którzy odeszli, nie korzystali z tych rozwiązań, w przeciwieństwie do tych, którzy pozostały. Może to sugerować, że brak skorzystania z tych usług mógł mieć wpływ na decyzję o rezygnacji, wskazując na ich niezadowolenie lub brak postrzeganego zapotrzebowania na dodatkowe usługi zabezpieczające.

Obecni klienci



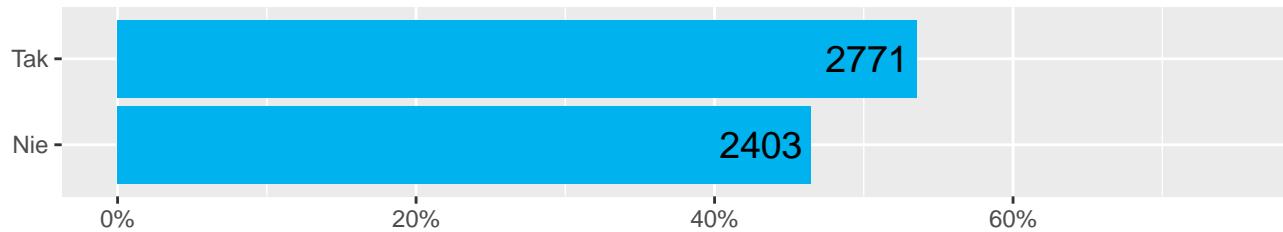
Dawni klienci



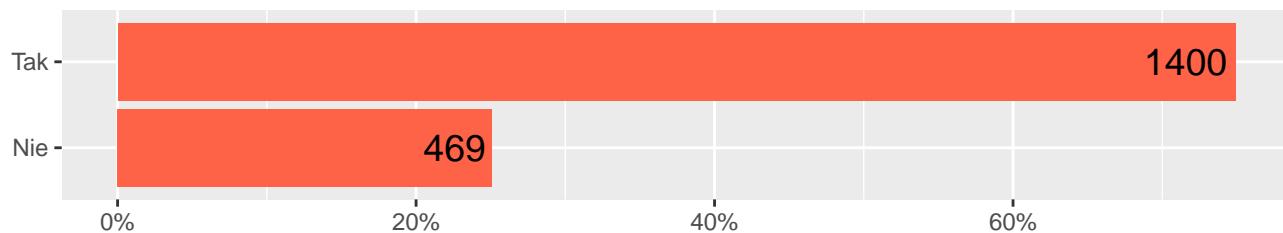
Wykres 38: Wykres słupkowy zmiennej *Contract* z podziałem ze względu na churn

Na wykresie 38. przedstawiono wykres słupkowy dla zmiennej *Contract*, z podziałem na klientów, którzy odeszli (churn) i tych, którzy pozostali. Z analizy wykresu wynika, że klienci, którzy odeszli, w zdecydowanej większości preferowali umowy miesięczne. Natomiast klienci, którzy pozostali, również wybierali umowy miesięczne, ale w znacznie mniejszej skali.

Obecni klienci

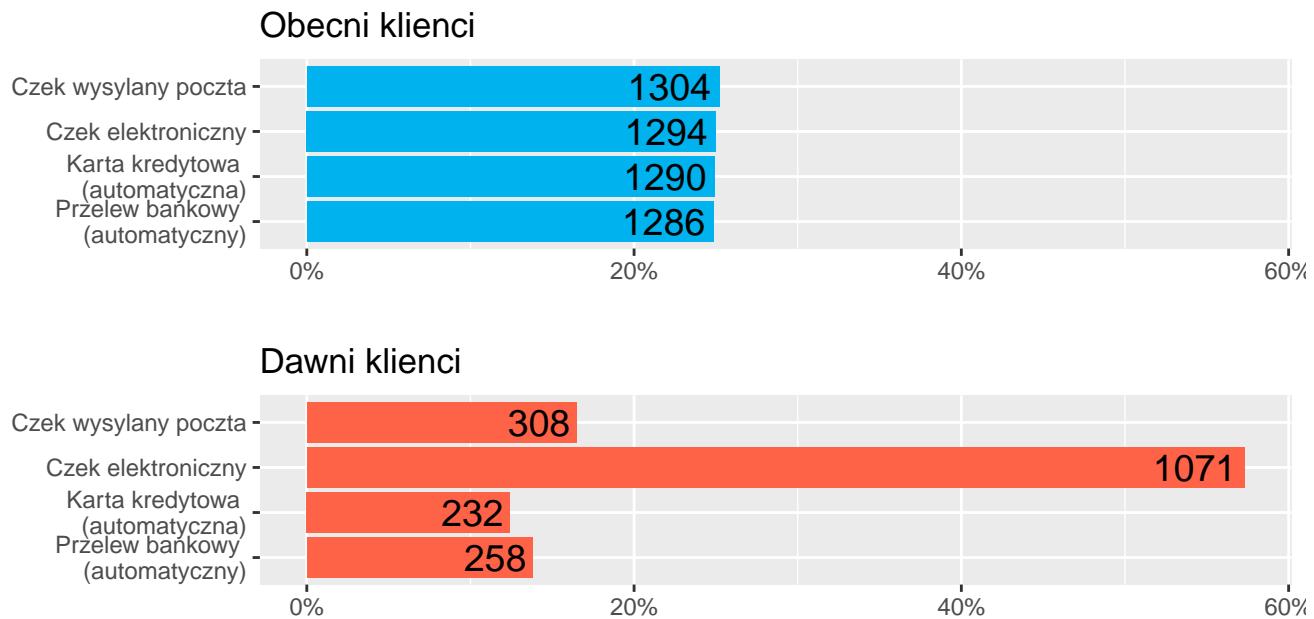


Dawni klienci



Wykres 39: Wykres słupkowy zmiennej *PaperlessBilling* z podziałem ze względu na churn

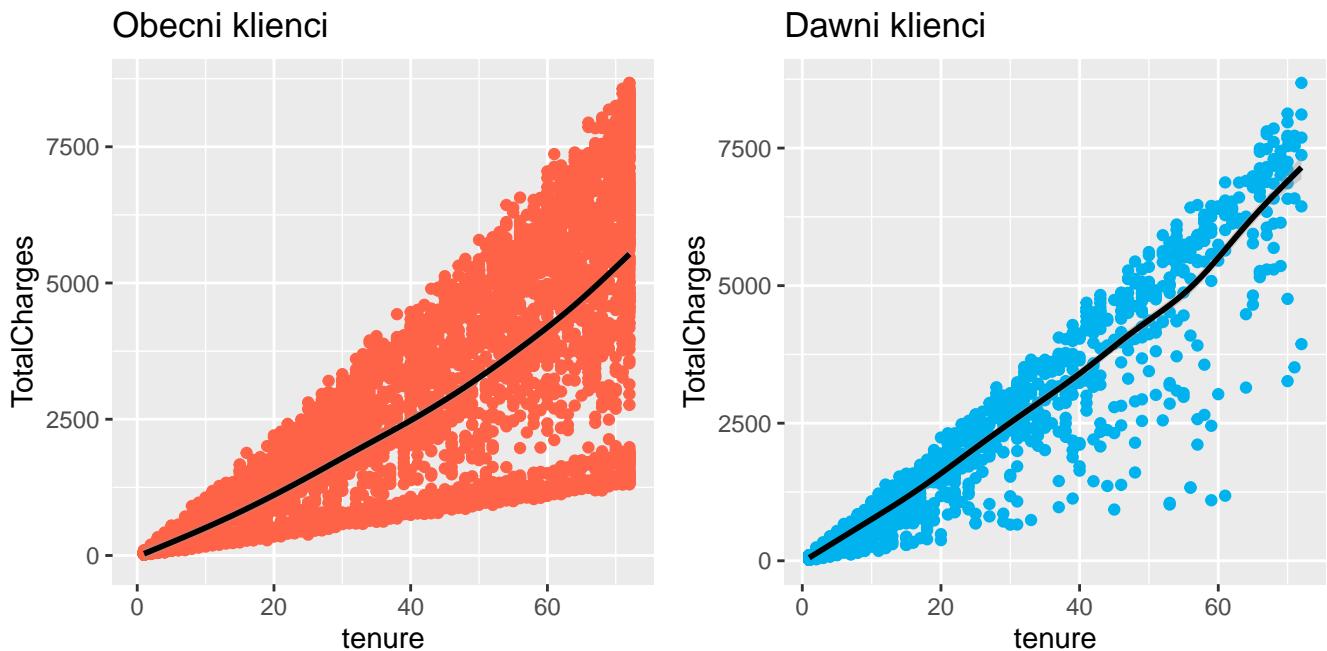
Na wykresie 39. przedstawiono wykres słupkowy dla zmiennej *PaperlessBilling*, z podziałem na klientów, którzy odeszli (churn) i tych, którzy pozostali. Widać, że klienci, którzy odeszli, preferowali biling bezpapierowy. Może to sugerować, że klienci, którzy preferowali tę formę rozliczeń, częściej decydowali się na rezygnację z usług.



Wykres 40: Wykres słupkowy zmiennej *PaymentMethod* z podziałem ze względu na churn

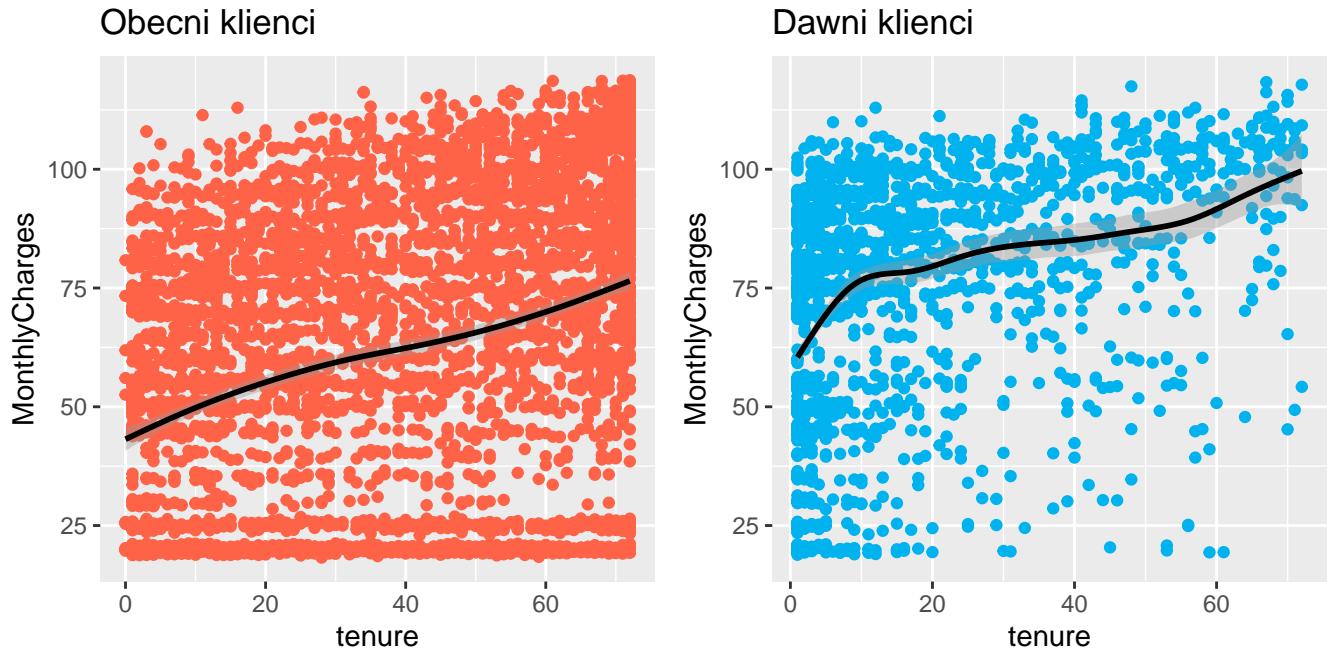
Na wykresie 40. przedstawiono wykres słupkowy dla zmiennej *PaymentMethod*, z podziałem na klientów, którzy odeszli (churn) i tych, którzy pozostały. Widać, że klienci lojalni korzystają w miarę równo ze wszystkich metod płatności, natomiast w przypadku klientów, którzy odeszli, dominującą metodą płatności jest czek elektroniczny.

6.3 Wykresy rozrzutów



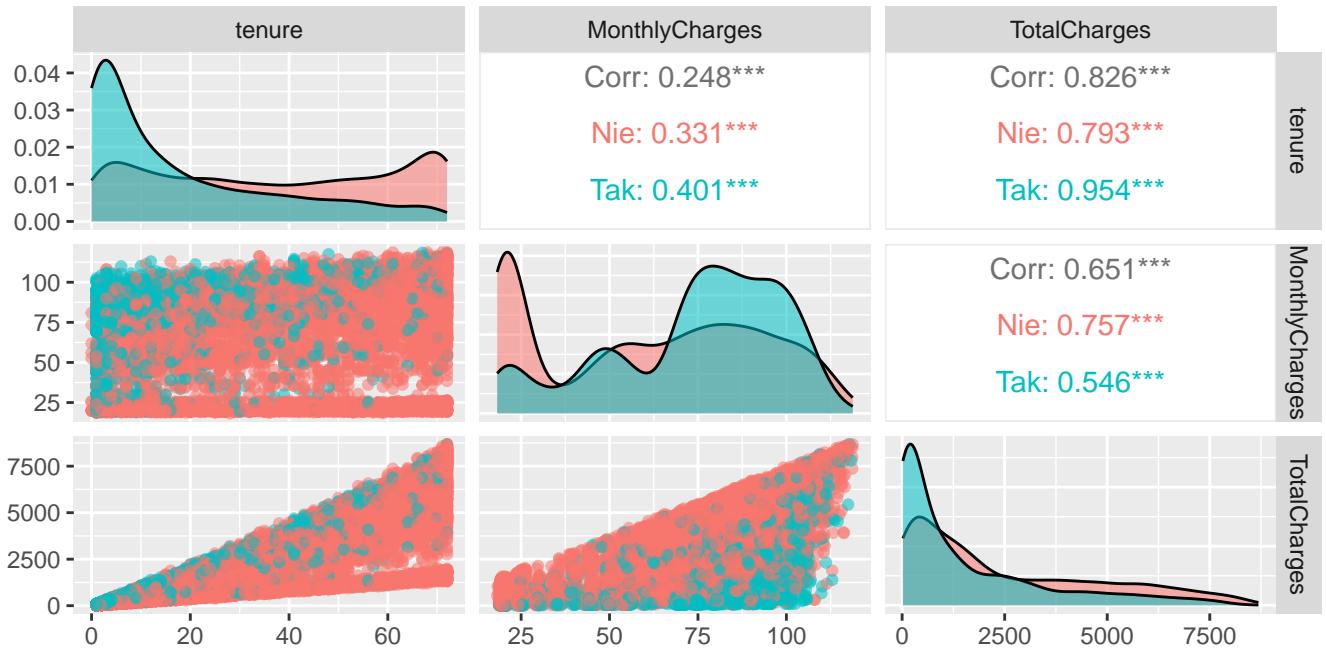
Wykres 41: Wykresy rozrzutów Total Charges ze względu na tenure z podziałem ze względu na zmienną churn

Wykres 41. przedstawia rozkład wartości zmiennej *TotalCharges* w zależności od *tenure*, z podziałem na zmienną *churn*. Widać, że w przypadku byłych klientów, dane koncentrują się wokół wyższych wartości, przy czym ich rozkład wykazuje charakter liniowy. Z kolei dla obecnych klientów, krzywa wygładzająca przechodzi przez środek rozkładu, wskazując na bardziej zróżnicowane zachowania w zakresie *TotalCharges*.



Wykres 42: Wykresy rozrzutów *MonthlyCharges* ze względu na tenure z podziałem ze względu na zmienną *churn*

Wykres 42. przedstawia wykresy rozrzutu *MonthlyCharges* w zależności od *tenure* z podziałem na zmienną *churn*. Widać, że obecni klienci stanowią pełny przekrój naszych klientów, obejmując różne poziomy opłat miesięcznych i długości współpracy. Klienci, którzy pozostali, utrzymują się bliżej maksymalnych wartości *MonthlyCharges*, co może sugerować, że korzystali z usług bardziej premium. Widać także, że jest ich zauważalnie więcej w przypadku wysokich *MonthlyCharges* i niskich *tenure*, co sugeruje, że ci klienci początkowo wykupili droższe usługi, ale później z nich zrezygnowali. Z kolei widać również niewielką liczbę klientów, którzy korzystali z tańszych, bardziej podstawowych usług.



Wykres 43: Macierz par zmiennych numerycznych z podziałem ze względu na churn

Na wykresie 43. przedstawiono macierz par zmiennych numerycznych z podziałem ze względu na *churn*. Widać w niej bardzo wysoką (0,954) korelację między *TotalCharges* a *tenure*, co wskazuje, że długość trwania współpracy z klientem (tenure) ma silny wpływ na całkowite opłaty (*TotalCharges*). Klienci, którzy pozostały z firmą przez dłuższy czas, zazwyczaj generują wyższe całkowite opłaty, co sugeruje, że dłuższa lojalność wiąże się z większymi wydatkami na usługi.

7 Dyskusja

1. Dłuższy okres współpracy (tenure) i zawieranie długoterminowych umów (roczne, dwuletnie) sprzyjają lojalności klientów.
2. Wyższe miesięczne opłaty (MonthlyCharges) korelują z wyższym ryzykiem rezygnacji.
3. Klienci korzystający z dodatkowych usług (OnlineSecurity, TechSupport, OnlineBackup) wykazują większą lojalność.
4. Usługi telefoniczne są kluczowym elementem oferty – około 90% klientów z nich korzysta.
5. Internet – najczęściej technologia światłowodowa (43%) oraz DSL (35%); około 22% klientów nie korzysta z internetu.
6. Klienci preferują biling elektroniczny (PaperlessBilling) oraz umowy miesięczne ze względu na elastyczność.
7. Klienci odchodzący rezygnują z usług wsparcia IT (wykresy 34–37) przy wysokich opłatach miesięcznych (wykres 29), co może wskazywać na postrzeganie tych usług jako zbędnych lub brak atrakcyjnych pakietów łączących wsparcie z innymi usługami.
8. Około 10% dawnych klientów korzystało z usług przez mniej niż rok, a około 15% przez dwa lata (wykres 27) – wskazuje to na podejście „testowe” w korzystaniu z oferty.
9. Wśród obecnych klientów około 21% korzysta z usług krócej niż rok, a 35% poniżej dwóch lat – warto monitorować tę grupę, by zapobiegać ich odejściu.
10. Ponad 85% dawnych klientów miało umowy miesięczne, co ułatwia rezygnację bez długoterminowych zobowiązań.
11. Duża liczba dawnych klientów korzystała ze światłowodu (wykres 33), co może wskazywać na niezadowolenie z tej usługi.
12. Klienci mają tendencję do dokupowania dodatkowych usług – obecni klienci robią to równomiernie,

- natomaiast dawni intensywnie w pierwszym roku, co może sugerować, że intensywne „testowanie” usług nie przekłada się na długoterminową satysfakcję (wykres 42).
13. Umiarkowana zmienność opłat miesięcznych (wykres 23) oznacza, że wyższe łączne wydatki nie muszą iść w parze z wysokimi miesięcznymi opłatami – mogą występować zarówno wieloletni klienci z niskimi opłatami, jak i nowi wybierający droższe pakiety.
 14. Około 20% klientów nie korzysta z internetu, natomiast tylko 10% rezygnuje z usług telefonicznych (wykresy 11 i 13), co sugeruje, że oferta telefoniczna jest bardziej atrakcyjna.

8 Wnioski

- Klienci, którzy odchodzą, zazwyczaj ponoszą wyższe koszty miesięczne, co sugeruje, że dla nich opłaty nie są adekwatne do otrzymywanych usług.
- Dominacja umów miesięcznych wśród klientów odchodzących wskazuje, że wielu z nich korzysta z oferty na zasadzie testowej, nie zobowiązując się długoterminowo.
- Rezygnacja z usług wsparcia IT przy wysokich opłatach sugeruje, że klienci nie widzą wystarczającej wartości w dodatkowych pakietach lub nie są one odpowiednio dopasowane cenowo, przez co są dodatkowo narażeni na większe zagrożenia w internecie, co może ich zniechęcać.
- Brakuje podstawowych danych demograficznych, takich jak wiek klientów, co utrudnia dokładniejszą klasyfikację i identyfikację segmentów, dla których oferta mogłaby być bardziej dopasowana.

Te czynniki łącznie wskazują, że główną przyczyną odchodzenia klientów jest niezadowalająca relacja między kosztem a wartością oferowanych usług, szczególnie w kontekście elastycznych, miesięcznych umów.

9 Rekomendacje

Stworzenie atrakcyjnych pakietów łączonych, które integrują podstawowe usługi ze wsparciem IT w konkurencyjnej cenie.

Wdrożenie programów lojalnościowych premiujących długoterminową współpracę, np. poprzez rabaty czy bonusy.

Personalizacja oferty, w tym przygotowanie specjalnych pakietów startowych dla nowych klientów testujących usługę oraz ekskluzywnych pakietów premium dla klientów lojalnych.

Proaktywna komunikacja z klientami, zwłaszcza tymi, którzy korzystają z usług krócej niż rok lub dwa lata, aby wcześnie zidentyfikować potencjalne niezadowolenie. **Analiza satysfakcji klientów korzystających ze światłowodu**, ponieważ możliwe, że niezadowolenie z tej technologii było przyczyną rezygnacji wielu klientów.

Kontakt z klientami, którzy odeszli przed upływem 24 miesięcy, oraz monitorowanie obecnych klientów w tym przedziale (obecnie 31%), aby zwiększyć ich retencję.

Uatrakcyjnienie usług IT, np. poprzez wprowadzenie okresów próbnych lub oferowanie ich w korzystnych pakietach za niewielką dopłatą.

Identyfikacja klientów o najniższych miesięcznych opłatach i zaproponowanie im dodatkowych usług, które mogłyby zwiększyć ich zaangażowanie i wartość dla firmy.

Nieznajomość szczegółowych informacji o firmie i jej ofercie znacząco utrudnia trafne dopasowanie rekomendacji.