

Telco Customer Churn

Stanisław Olek

1 kwietnia 2025

Spis treści

1 Wprowadzenie	1
2 Wczytanie i przygotowanie danych	1
2.1 Rozmiar danych	1
2.2 Typy poszczególnych cech	2
2.3 Przydatność cech	2
2.4 Brakujące obserwacje	4
2.5 Nietypowe wartości	4
3 Analiza zmiennych ilościowych	6
3.1 Analiza dla zbioru bez podziału na grupy	6
3.1.1 Podstawowe wskaźniki sumaryczne	6
3.1.2 Histogramy	7
3.1.3 Wykresy gęstości	8
3.1.4 Wykresy pudełkowe	9
3.1.5 Wykresy rozrzutu	11
3.1.6 Macierz korelacji	11
3.2 Analiza z podziałem na grupy według zmiennej Churn	13
3.2.1 Podstawowe wskaźniki sumaryczne	13
3.2.2 Histogramy	14
3.2.3 Wykresy gęstości	15
3.2.4 Wykresy pudełkowe	16
3.2.5 Wykresy rozrzutu	18
3.2.6 Macierze korelacji	18
4 Analiza zmiennych jakościowych	20
4.1 Analiza dla zbioru bez podziału na grupy	20
4.1.1 Tabele częstości	20
4.1.2 Wykresy słupkowe	22
4.2 Analiza dla zbioru z podziałem na grupy według zmiennej Churn	23
4.2.1 Tabele częstości	23
4.2.2 Wykresy słupkowe	25
4.2.3 Zmienne najlepiej rozróżniające grupy klientów rezygnujących i pozostających	26

5 Podsumowanie - wnioski z przeprowadzonej analizy	27
5.1 Podsumowanie wyników	27
5.2 Charakterystyka klientów firmy	27
5.3 Główne przyczyny odchodzenia klientów i rekomendacje	27

1 Wprowadzenie

Niniejszy raport przedstawia analizę danych dotyczących rezygnacji klientów z usług telekomunikacyjnych (customer churn). Zrozumienie czynników wpływających na decyzję klientów o rezygnacji z usług jest kluczowe dla firm telekomunikacyjnych, ponieważ pozwala im opracować strategie utrzymania klientów.

2 Wczytanie i przygotowanie danych

```
# wczytanie zbioru danych
dane <- read.csv(
  '/Users/cj/Documents/ED/1/WA_Fn-UseC_-Telco-Customer-Churn.csv',
  stringsAsFactors=TRUE)
```

```
dane$SeniorCitizen <- as.factor(dane$SeniorCitizen)
```

Po konwersji zmiennej `SeniorCitizen` jest poprawnie traktowana jako zmienna jakościowa (factor), co odpowiada jej naturze, zmieniona ta wskazuje bowiem, czy klient jest seniorem 1 czy nie 0.

2.1 Rozmiar danych

```
# sprawdzenie rozmiaru danych
dim(dane)

## [1] 7043 21
```

Zbiór danych składa się z 7043 przypadków (wierszy) oraz 21 cech (kolumn)

2.2 Typy poszczególnych cech

Zmienne jakościowe (factor):

- `customerID`,
- `gender`,
- `SeniorCitizen`,
- `Partner`,

- Dependents,
- PhoneService,
- MultipleLines,
- InternetService,
- OnlineSecurity,
- OnlineBackup,
- DeviceProtection,
- TechSupport,
- StreamingTV,
- StreamingMovies,
- Contract,
- PaperlessBilling,
- PaymentMethod,
- Churn

Liczba zmiennych jakościowych: 18

Zmienne ilościowe (numeric):

- tenure,
- MonthlyCharges,
- TotalCharges

Liczba zmiennych ilościowych: 3

2.3 Przydatność cech

W danych występuje cecha `customerID`, która pełni rolę identyfikatora klientów. Cecha ta nie wnosi wartości merytorycznej do analizy, dlatego należy ją usunąć przed dalszą analizą.

```
# usunięcie kolumny customerID
dane <- dane[,-1]

# sprawdzenie struktury danych po usunięciu kolumny customerID
dim(dane)

## [1] 7043    20
```

2.4 Brakujące obserwacje

```
# sprawdzenie czy w danych występują brakujące wartości (NA)
sum(is.na(dane))
```

```
## [1] 11
```

```
# sprawdzenie liczby brakujących wartości dla każdej cechy
na_licznik <- sapply(dane, function(x) sum(is.na(x)))
na_licznik
```

```
##           gender   SeniorCitizen       Partner   Dependents
##             0                 0                 0                 0
##           tenure   PhoneService  MultipleLines InternetService
##             0                 0                 0                 0
##           OnlineSecurity  OnlineBackup DeviceProtection TechSupport
##             0                 0                 0                 0
##           StreamingTV  StreamingMovies      Contract PaperlessBilling
##             0                 0                 0                 0
##           PaymentMethod MonthlyCharges  TotalCharges        Churn
##             0                 0                 11                0
```

W danych występuje 11 brakujących wartości, wszystkie w kolumnie **TotalCharges**. Brakujące wartości są kodowane jako NA w R.

2.5 Nietypowe wartości

```
# sprawdzenie czy zmienne numeryczne zawierają nietypowe wartości
summary(dane[, numeric_zmienne])
```

```
##           tenure   MonthlyCharges  TotalCharges
##   Min.   : 0.00   Min.   : 18.25   Min.   : 18.8
##   1st Qu.: 9.00   1st Qu.: 35.50   1st Qu.: 401.4
##   Median :29.00   Median : 70.35   Median :1397.5
##   Mean   :32.37   Mean   : 64.76   Mean   :2283.3
##   3rd Qu.:55.00   3rd Qu.: 89.85   3rd Qu.:3794.7
##   Max.   :72.00   Max.   :118.75   Max.   :8684.8
##                   NA's   :11
```

```
# sprawdzenie czy wartości TotalCharges = NA mają związek z tenure
dane$tenure[is.na(dane$TotalCharges)]
```

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

```
# sprawdzenie, czy występują nietypowe wartości w zmiennych jakościowych
for (col in names(dane)[sapply(dane, is.factor)]) {
  unikalne_wartosci <- unique(as.character(dane[[col]]))
  nietypowe <- unikalne_wartosci[
```

```
    unikalne_wartosci %in% c("", " ", "unknown", "N/A", "NA", "-", "?")]
}

length(nietypowe)

## [1] 0
```

Wszystkie wiersze z brakującymi wartościami w kolumnie `TotalCharges` mają wartość 0 w kolumnie `tenure`. Sugeruje to, że brakujące wartości `TotalCharges` dotyczą nowych klientów, którzy nie dokonali jeszcze żadnych płatności. Nie zaobserwowano innych nietypowych wartości czy niestandardowego kodowania brakujących danych w danych.

3 Analiza zmiennych ilościowych

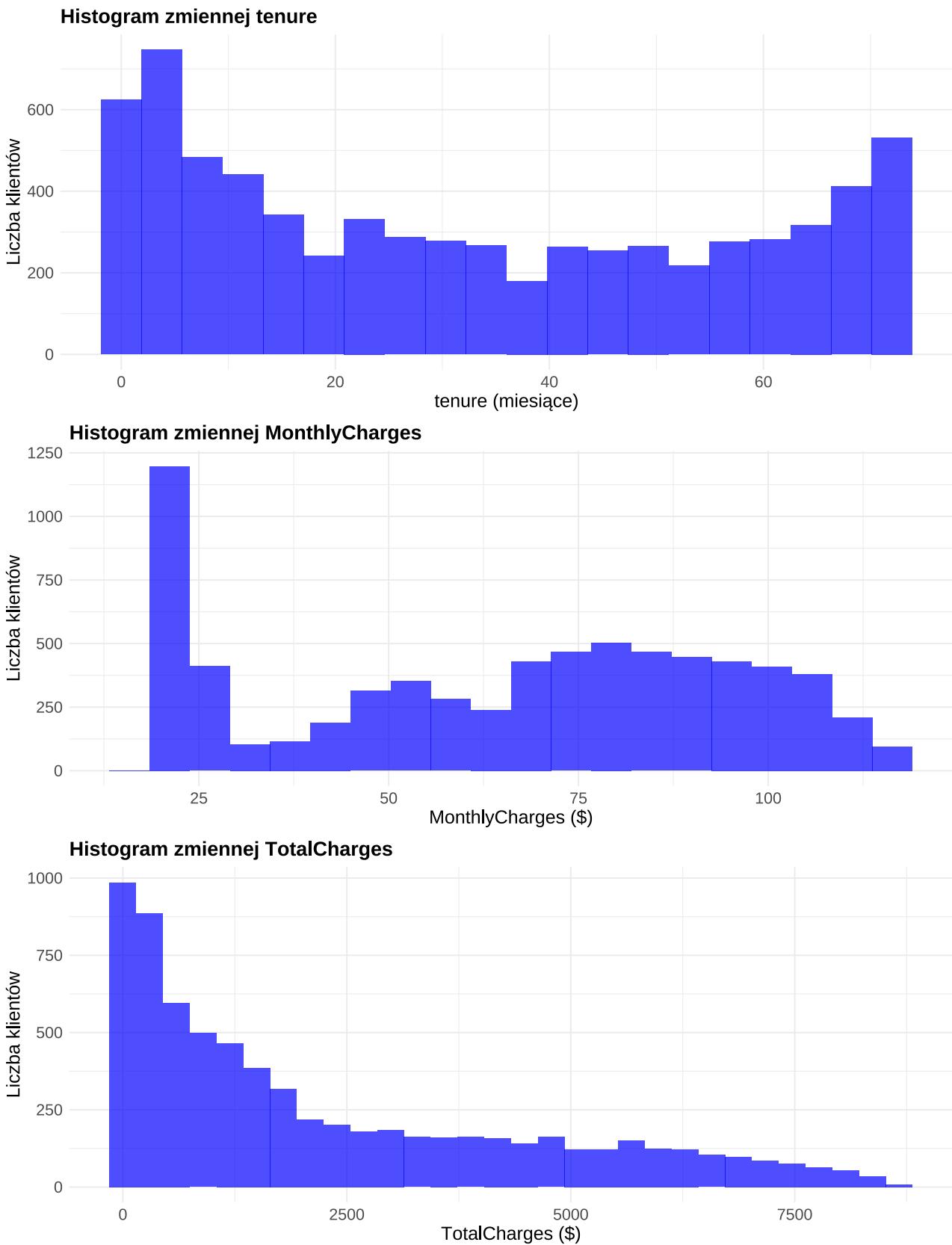
3.1 Analiza dla zbioru bez podziału na grupy

3.1.1 Podstawowe wskaźniki sumaryczne

Tabela 1: Podstawowe statystyki dla zmiennych ilościowych

	tenure	MonthlyCharges	TotalCharges
Średnia	32.37	64.76	2283.30
Mediana	29.00	70.35	1397.47
Min	0.00	18.25	18.80
Max	72.00	118.75	8684.80
Q1.25%	9.00	35.50	401.45
Q3.75%	55.00	89.85	3794.74
SD	24.56	30.09	2266.77
IQR	46.00	54.35	3393.29

3.1.2 Histogramy



Rysunek 1: Histogramy dla zmiennych ilościowych bez podziału na grupy

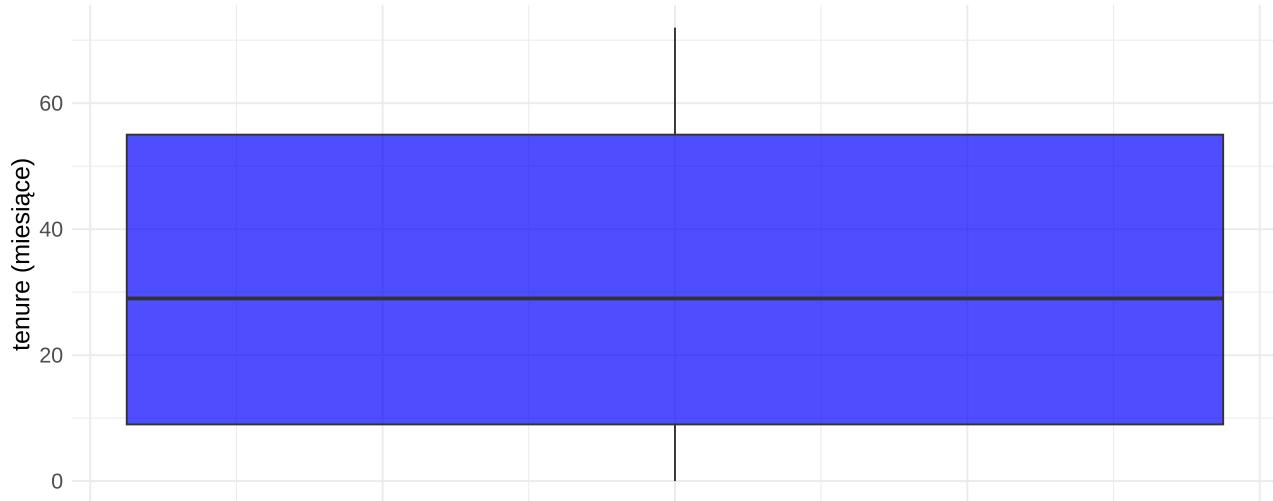
3.1.3 Wykresy gęstości



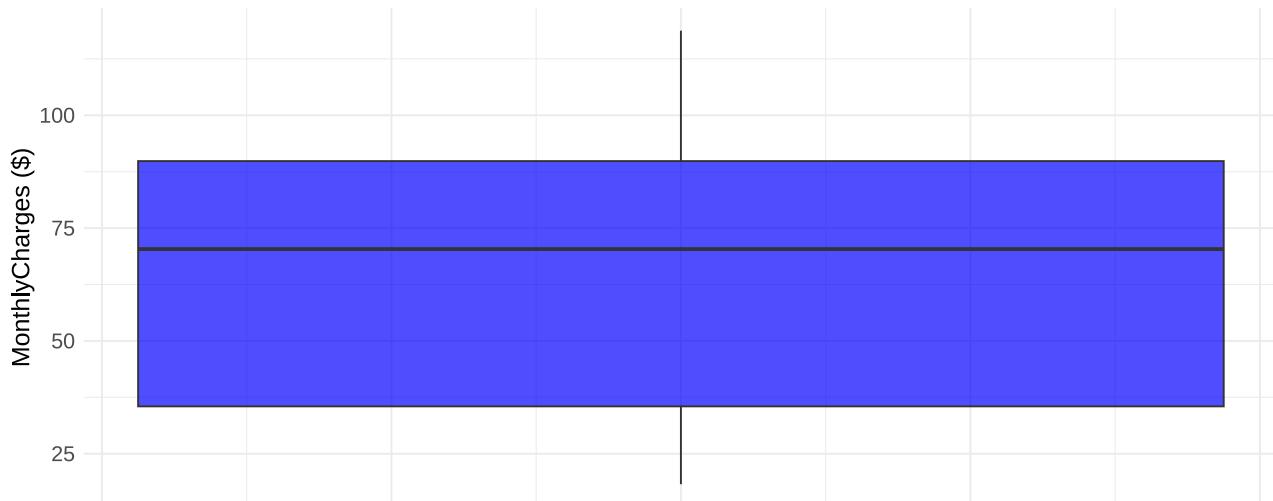
Rysunek 2: Wykresy gęstości dla zmiennych ilościowych bez podziału na grupy

3.1.4 Wykresy pudełkowe

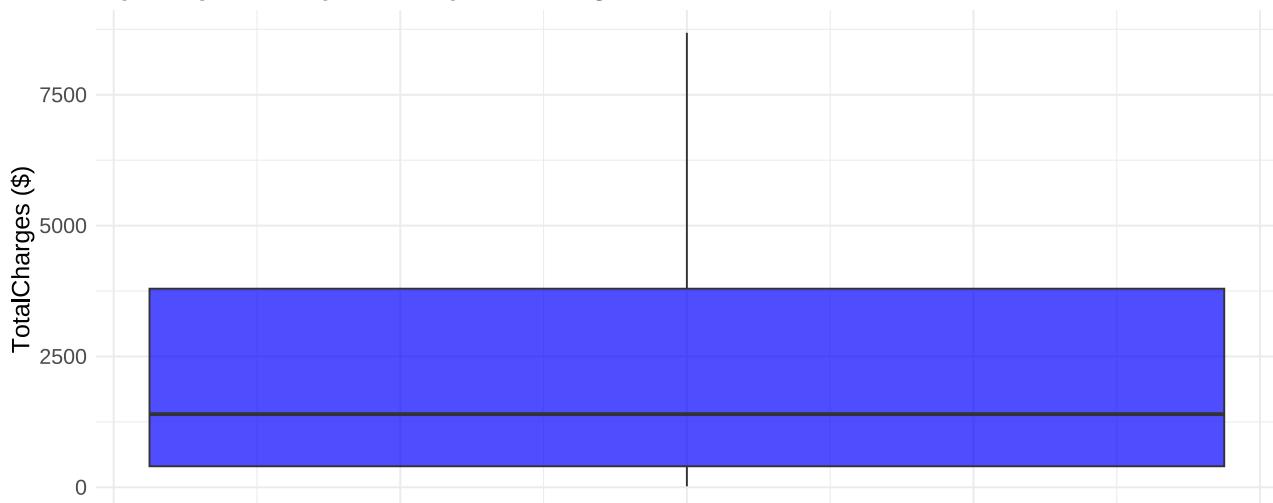
Wykres pudełkowy zmiennej tenure



Wykres pudełkowy zmiennej MonthlyCharges



Wykres pudełkowy zmiennej TotalCharges

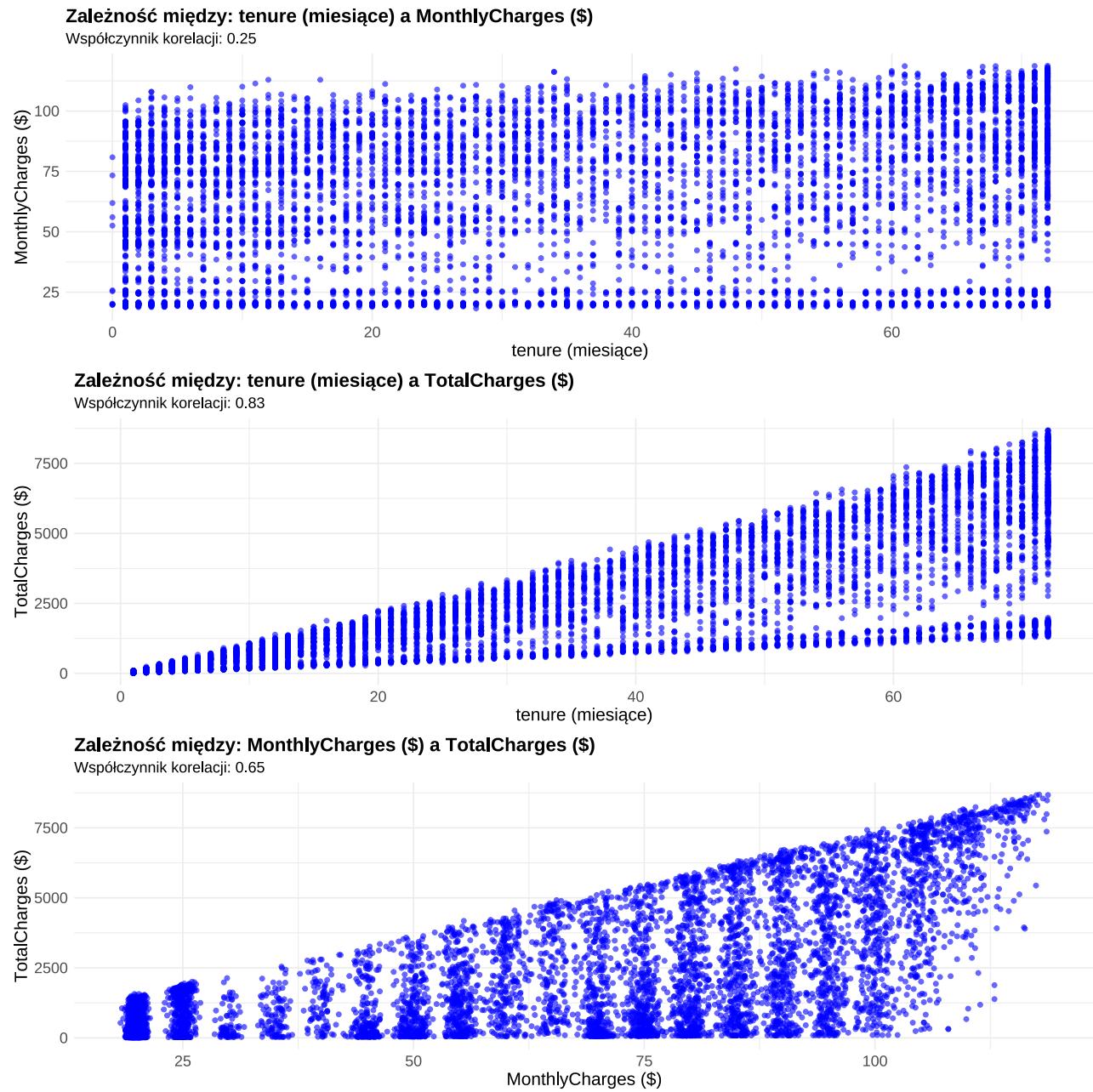


Rysunek 3: Wykresy pudełkowe dla zmiennych ilościowych bez podziału na grupy

Analizując dane z tabeli (Tabela 1), histogramów (Rysunek 1), wykresów gęstości (Rysunek 2) i wykresów pudełkowych (Rysunek 3) można zauważyć, że:

- Zmienna **tenure** ma wartości od 0 do 72 miesięcy, ze średnią 32.4 miesiące i medianą 29 miesięcy. Rozkład tej zmiennej jest dość równomierny, co sugeruje podobną liczbę klientów z różnym stażem
- Zmienna **MonthlyCharges** ma zakres od \$18.25 do \$118.75, ze średnią \$64.76 i medianą \$70.35. Różnica między średnią a medianą sugeruje lekką asymetrię rozkładu. Wykres gęstości pokazuje dwa wyraźne szczyty - pierwszy dla niższych opłat (około 20\$) i drugi dla wyższych (około 80\$)
- Zmienna **TotalCharges** ma największe zróżnicowanie, z wartościami od \$18.8 do \$8684.8. Duża różnica między średnią (\$2283.3) a medianą (\$1397.47) wskazuje na wyraźną asymetrię prawostronną rozkładu. Wykres gęstości potwierdza to, pokazując wysoką koncentrację obserwacji w niższych wartościach i długi ogon rozkładu w kierunku wartości wysokich

3.1.5 Wykresy rozrzutu



Rysunek 4: Wykresy rozrzutu dla par zmiennych ilościowych bez podziału na grupy

3.1.6 Macierz korelacji

Tabela 2: Macierz korelacji dla zmiennych ilościowych

	Okres umowy	MonthlyCharges	TotalCharges
tenure	1.00	0.25	0.83
MonthlyCharges	0.25	1.00	0.65
TotalCharges	0.83	0.65	1.00

Na podstawie analizy wykresów rozrzutu (Rysunek 4) oraz macierzy korelacji (Tabela 2), można sformułować następujące wnioski dotyczące zależności między zmiennymi ilościowymi i zmiennością poszczególnych cech ilościowych:

- Między okresem umowy (**tenure**) a opłatami całkowitymi (**TotalCharges**) występuje najsilniejsza zależność liniowa, co potwierdza wysoki współczynnik korelacji wynoszący 0.83. Jest to logiczne, ponieważ im dłużej klient korzysta z usług, tym więcej płaci w sumie. Na wykresie rozrzutu widać wyraźny trend liniowy
- Między opłatami miesięcznymi (**MonthlyCharges**) a opłatami całkowitymi (**TotalCharges**) występuje umiarkowana korelacja (0.65). Związek ten nie jest idealnie liniowy. Na wykresie rozrzutu można zaobserwować duże rozrzut punktów, co potwierdza, że opłaty całkowite zależą nie tylko od wysokości miesięcznych opłat, ale także od okresu umowy
- Między okresem umowy (**tenure**) a opłatami miesięcznymi (**MonthlyCharges**) widoczny jest prawie brak korelacji (0.25), co sugeruje, że wysokość miesięcznych opłat nie zależy od długości relacji klienta z firmą. Wykres rozrzutu dla tej pary zmiennych pokazuje chmurę punktów bez wyraźnego trendu
- Największą zmienność wykazują zmienne:
 - **TotalCharges**, jej punkty są rozproszone w bardzo szerokim zakresie wartości, od bliskich zeru do ponad \$8000
 - **tenure**, co widoczne jest w równomiernym rozłożeniu punktów wzduż całej osi X (od 0 do ponad 70 miesięcy)

3.2 Analiza z podziałem na grupy według zmiennej Churn

3.2.1 Podstawowe wskaźniki sumaryczne

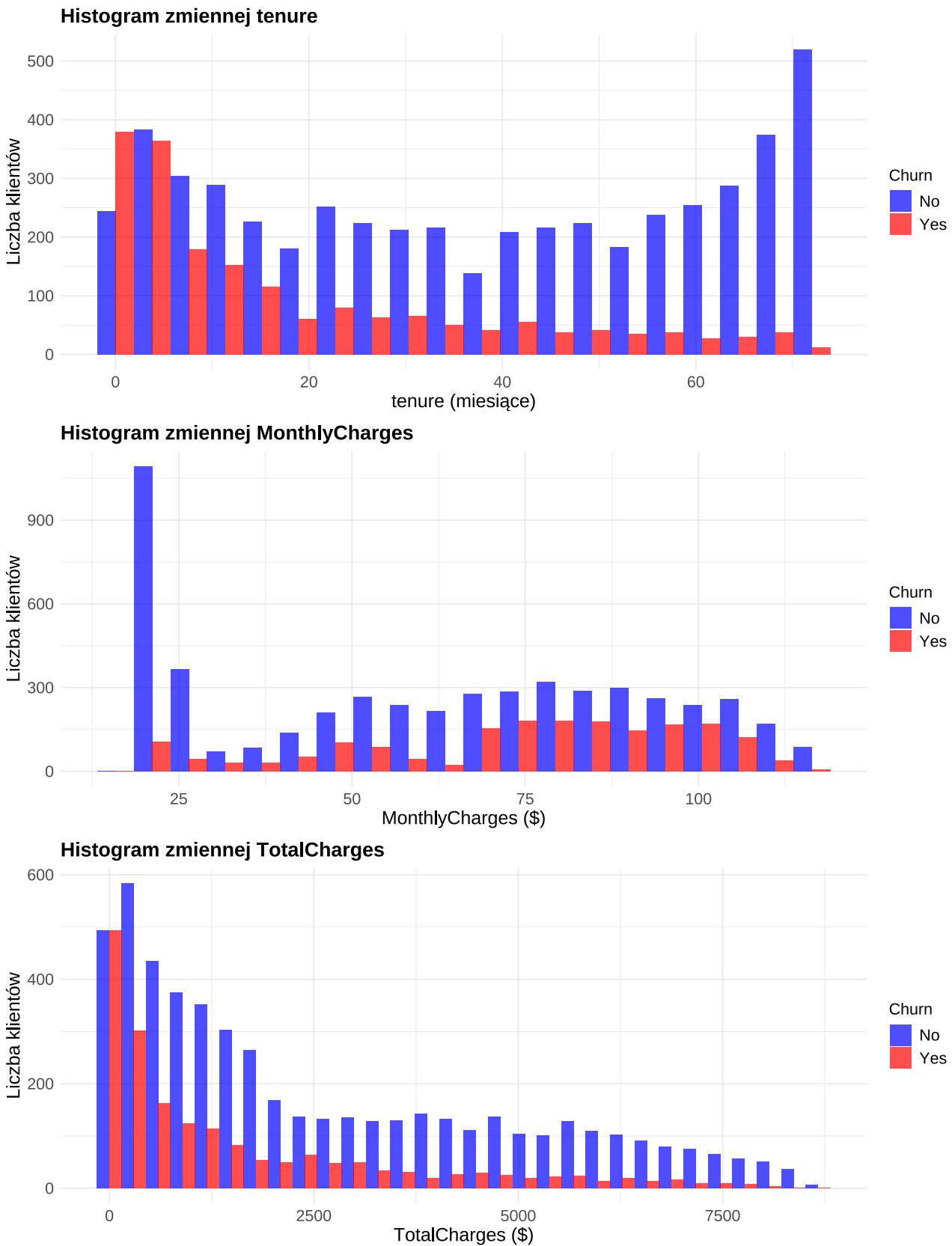
Tabela 3: Podstawowe statystyki dla klientów, którzy odeszli (Churn = Yes)

	tenure	MonthlyCharges	TotalCharges
Średnia	17.98	74.44	1531.80
Medianą	10.00	79.65	703.55
Min	1.00	18.85	18.85
Max	72.00	118.35	8684.80
Q1.25%	2.00	56.15	134.50
Q3.75%	29.00	94.20	2331.30
SD	19.53	24.67	1890.82
IQR	27.00	38.05	2196.80

Tabela 4: Podstawowe statystyki dla klientów lojalnych (Churn = No)

	tenure	MonthlyCharges	TotalCharges
Średnia	37.57	61.27	2555.34
Medianą	38.00	64.43	1683.60
Min	0.00	18.25	18.80
Max	72.00	118.75	8672.45
Q1.25%	15.00	25.10	577.82
Q3.75%	61.00	88.40	4264.12
SD	24.11	31.09	2329.46
IQR	46.00	63.30	3686.30

3.2.2 Histogramy



Rysunek 5: Histogramy dla zmiennych ilościowych z podziałem na grupy według zmiennej Churn

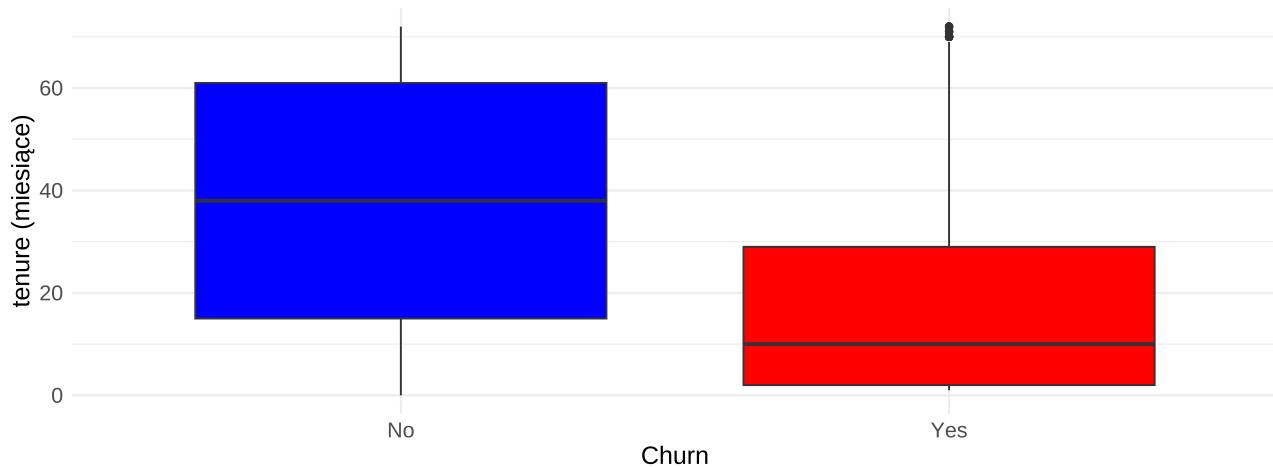
3.2.3 Wykresy gęstości



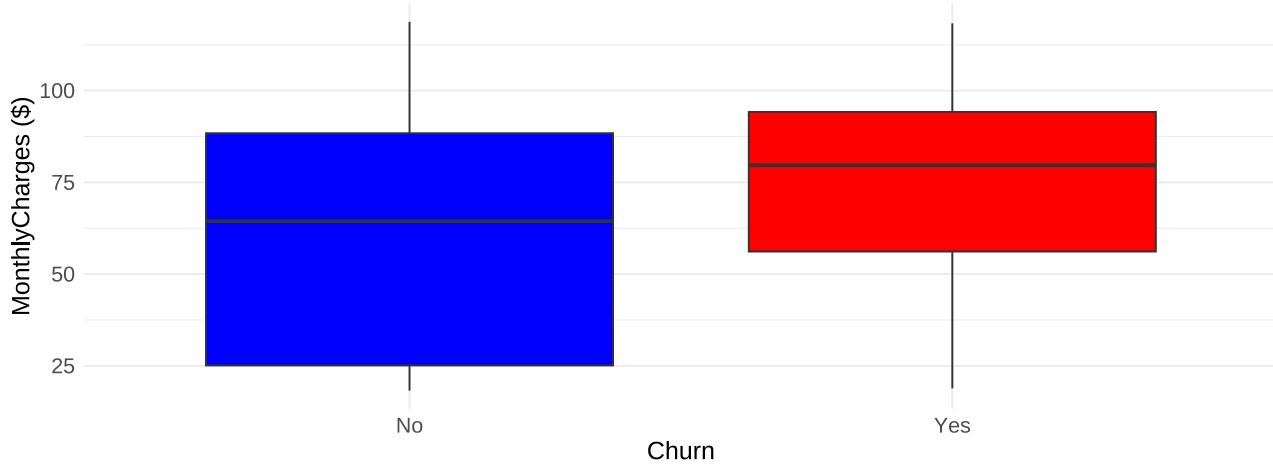
Rysunek 6: Wykresy gęstości dla zmiennych ilościowych z podziałem na grupy według zmiennej Churn

3.2.4 Wykresy pudełkowe

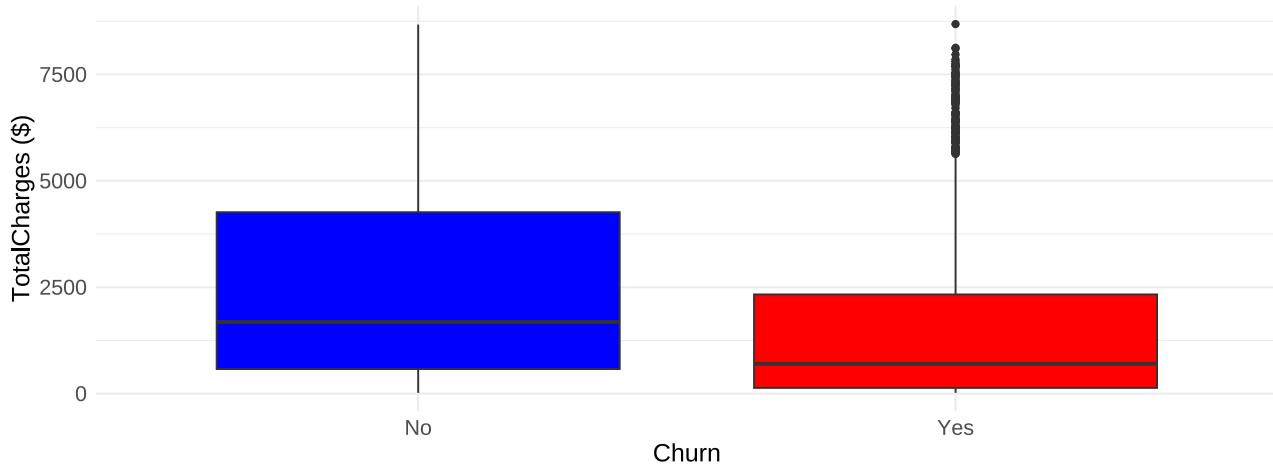
Wykres pudełkowy zmiennej tenure



Wykres pudełkowy zmiennej MonthlyCharges



Wykres pudełkowy zmiennej TotalCharges



Rysunek 7: Wykresy pudełkowe dla zmiennych ilościowych z podziałem na grupy według zmiennej Churn

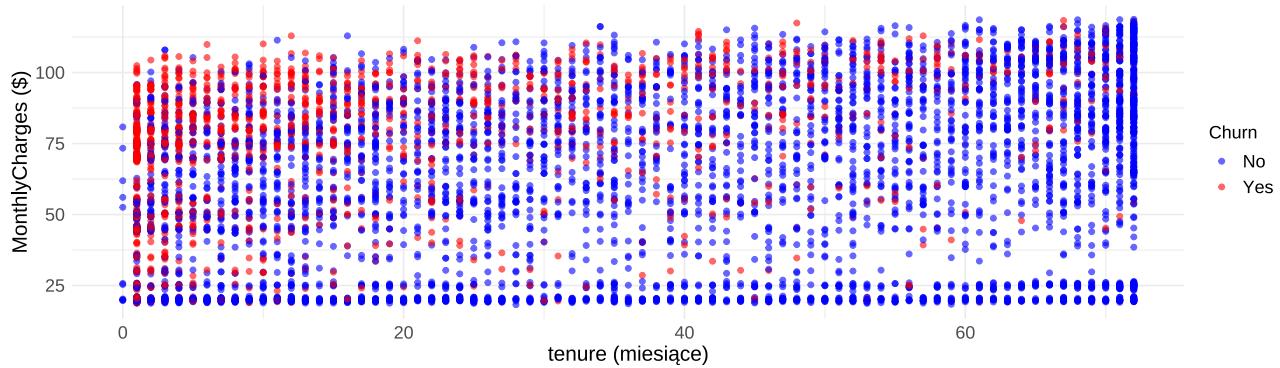
Analizując dane z tabel (Tabela 3 i 4), histogramów (Rysunek 5), wykresów gęstości (Rysunek 6) oraz wykresów pudełkowych (Rysunek 7) dla grup wydzielonych według zmiennej Churn, można zauważyć, że:

- Zmienna **tenure** wykazuje znaczące różnice między grupami:
 - Dla klientów, którzy odeszli, średnia wynosi jedynie 18 miesięcy, podczas gdy dla lojalnych klientów - 37.6 miesięcy
 - Mediana dla klientów odchodzących to 10 miesięcy, w porównaniu do 38 miesięcy dla klientów lojalnych
 - Wykres gęstości pokazuje, że większość rezygnacji następuje w początkowym okresie umowy - rozkład dla grupy Churn=Yes jest silnie prawoskoły
- Zmienna **MonthlyCharges** również wykazuje istotne różnice:
 - Klienci, którzy odeszli, płacili średnio \$74.44 miesięcznie wobec \$61.27 dla klientów lojalnych
 - Wykresy pudełkowe ujawniają, że mediana i kwartyle dla klientów rezygnujących są wyraźnie przesunięte w kierunku wyższych wartości
 - Wykres gęstości dla klientów, którzy odeszli, ma większe zagnieszczenie w obszarze wysokich opłat miesięcznych (70-100\$)
- Zmienna **TotalCharges** pokazuje:
 - Klienci rezygnujący mają niższe średnie opłaty całkowite (\$1531.8) niż klienci lojalni (\$2555.34) (co jest naturalną konsekwencją krótszego okresu umów klientów rezygnujących)
 - Wykresy pokazują, że dla klientów rezygnujących rozkład jest silnie skoncentrowany w obszarze niskich wartości

3.2.5 Wykresy rozrzutu

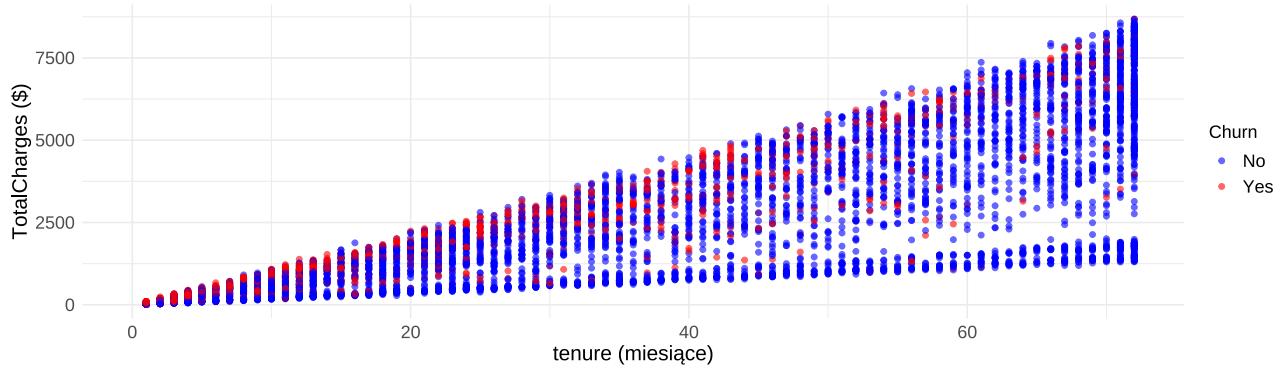
Zależność między: tenure (miesiące) a MonthlyCharges (\$)

Korelacja (Yes): 0.4, Korelacja (No): 0.33



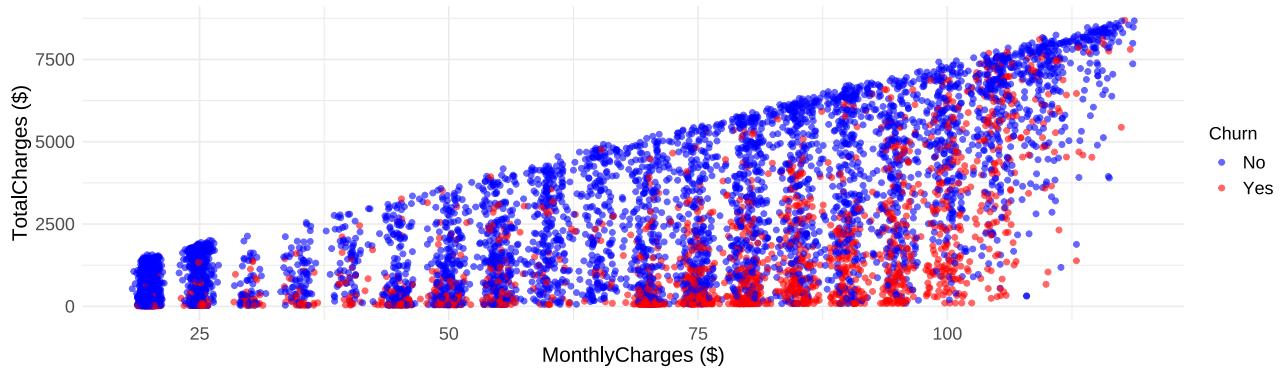
Zależność między: tenure (miesiące) a TotalCharges (\$)

Korelacja (Yes): 0.95, Korelacja (No): 0.79



Zależność między: MonthlyCharges (\$) a TotalCharges (\$)

Korelacja (Yes): 0.55, Korelacja (No): 0.76



Rysunek 8: Wykresy rozrzutu dla par zmiennych ilościowych z podziałem na grupy według zmiennej Churn

3.2.6 Macierze korelacji

Tabela 5: Macierz korelacji dla klientów, którzy odeszli (Churn = Yes)

	tenure	MonthlyCharges	TotalCharges
tenure	1.00	0.40	0.95
MonthlyCharges	0.40	1.00	0.55
TotalCharges	0.95	0.55	1.00

Tabela 6: Macierz korelacji dla klientów lojalnych (Churn = No)

	tenure	MonthlyCharges	TotalCharges
tenure	1.00	0.33	0.79
MonthlyCharges	0.33	1.00	0.76
TotalCharges	0.79	0.76	1.00

Na podstawie analizy wykresów rozrzutu (Rysunek 8) oraz macierzy korelacji (Tabela 5 i 6), można sformułować następujące wnioski dotyczące zależności między zmiennymi ilościowymi w obu grupach klientów:

- Korelacja między **tenure** a **TotalCharges** jest silna w obu grupach:
 - Dla klientów, którzy odeszli, współczynnik korelacji wynosi 0.95
 - Dla klientów lojalnych współczynnik korelacji wynosi 0.79
 - Na wykresie rozrzutu widać, że punkty reprezentujące klientów, którzy odeszli, są skoncentrowane w lewym dolnym rogu (krótki okres, niskie opłaty całkowite)
- Korelacja między **MonthlyCharges** a **TotalCharges** wykazuje różnice między grupami:
 - Dla klientów, którzy odeszli, korelacja jest słabsza (0.55) niż dla klientów lojalnych (0.76)
 - Niższa korelacja w grupie klientów rezygnujących może wynikać z ich krótkiego okresu umowy, który ogranicza wpływ opłat miesięcznych na wartość całkowitą
- Relacja między **tenure** a **MonthlyCharges** ujawnia niewielkie różnice między grupami:
 - Dla klientów, którzy odeszli, występuje umiarkowana dodatnia korelacja (0.4)
 - Dla klientów lojalnych korelacja jest również dodatnia, choć nieco słabsza (0.33)
 - Sugeruje to, że szczególnie w grupie klientów rezygnujących, dłuższy okres umowy jest powiązany z wyższymi opłatami miesięcznymi - być może ci klienci decydują się na rozwiązanie umowy właśnie z powodu rosnących kosztów

Zmiennymi najlepiej różnicującymi obie grupy są:

- **tenure** - bardzo wyraźna separacja grup na wykresach, największa różnica średnich i median
- **MonthlyCharges** - widoczna różnica w rozkładach, klienci rezygnujący płacą więcej miesięcznie

4 Analiza zmiennych jakościowych

W dalszej analizie skupimy się na zmiennych, które wykazują największe zróżnicowanie w danych oraz potencjalnie najsilniejszy związek ze zmienną objaśnianą Churn. Zmienne takie jak SeniorCitizen, Partner, Dependents, Contract, PaperlessBilling oraz TechSupport będą głównym przedmiotem analizy, podczas gdy zmienne o dużej liczbie brakujących danych lub też zmienne nic nie wnoszące do analizy zostaną pominięte, aby uniknąć wprowadzania szumów do analizy.

4.1 Analiza dla zbioru bez podziału na grupy

4.1.1 Tabele częstości

Tabela 7: Tabela częstości dla zmiennej: SeniorCitizen

Kategoria	Liczebność	Procent
0	5901	83.79
1	1142	16.21

Tabela 8: Tabela częstości dla zmiennej: Partner

Kategoria	Liczebność	Procent
No	3641	51.70
Yes	3402	48.30

Tabela 9: Tabela częstości dla zmiennej: Dependents

Kategoria	Liczebność	Procent
No	4933	70.04
Yes	2110	29.96

Tabela 10: Tabela częstości dla zmiennej: Contract

Kategoria	Liczebność	Procent
Month-to-month	3875	55.02
One year	1473	20.91
Two year	1695	24.07

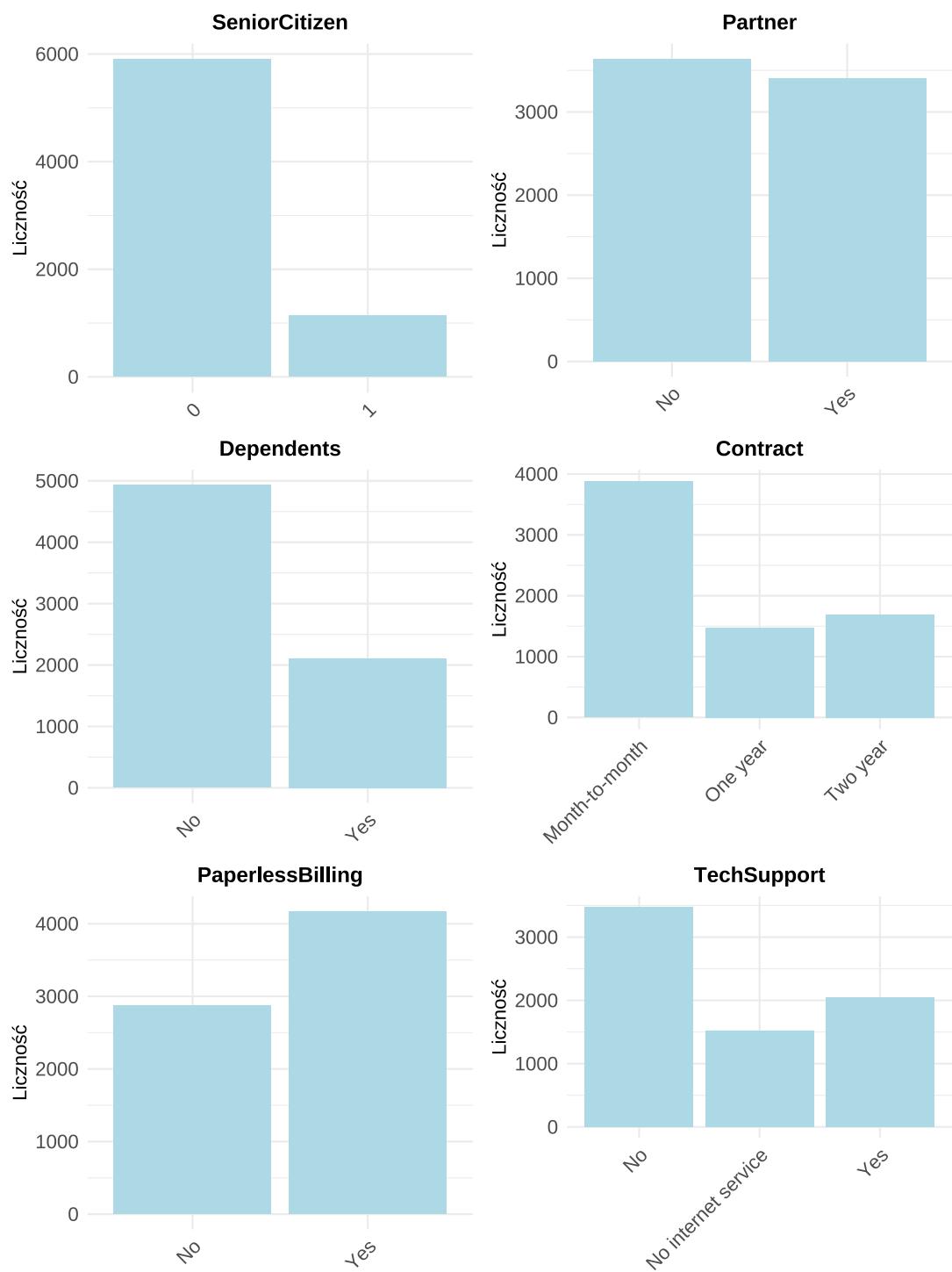
Tabela 11: Tabela częstości dla zmiennej: PaperlessBilling

Kategoria	Liczebność	Procent
No	2872	40.78
Yes	4171	59.22

Tabela 12: Tabela częstości dla zmiennej: TechSupport

Kategoria	Liczebność	Procent
No	3473	49.31
No internet service	1526	21.67
Yes	2044	29.02

4.1.2 Wykresy słupkowe



Rysunek 9: Wykresy słupkowe przedstawiające rozkład częstości najważniejszych zmiennych jakościowych

Jak przedstawiono na wykresach słupkowych (Rysunek 9), zmienne jakościowe w zbiorze danych charakteryzują się różnymi rozkładami. Szczególnie istotna jest zmienna **Contract** (rodzaj umowy), gdzie dominują umowy miesięczne, oraz zmienna **SeniorCitizen**, pokazująca że zdecydowana większość klientów to osoby młodsze. Zmienne **Partner** i **Dependents** wskazują, że większość klientów nie posiada partnerów ani osób zależnych. Zmienna **TechSupport** pokazuje, że znaczna część klientów nie korzysta z usług wsparcia technicznego, co może mieć istotne znaczenie w kontekście ich satysfakcji z usług.

4.2 Analiza dla zbioru z podziałem na grupy według zmiennej Churn

4.2.1 Tabele częstości

Tabela 13: Tabela częstości dla zmiennej **SeniorCitizen** z podziałem na Churn

Kategoria	Liczba_klientow	Liczba_odchodzacych	Procent_odchodzacych
0	5901.00	1393	23.61
1	1142.00	476	41.68

Tabela 14: Tabela częstości dla zmiennej **Partner** z podziałem na Churn

Kategoria	Liczba_klientow	Liczba_odchodzacych	Procent_odchodzacych
No	3641.00	1200	32.96
Yes	3402.00	669	19.66

Tabela 15: Tabela częstości dla zmiennej **Dependents** z podziałem na Churn

Kategoria	Liczba_klientow	Liczba_odchodzacych	Procent_odchodzacych
No	4933.00	1543	31.28
Yes	2110.00	326	15.45

Tabela 16: Tabela częstości dla zmiennej Contract z podziałem na Churn

Kategoria	Liczba_klientow	Liczba_odchodzacych	Procent_odchodzacych
Month-to-month	3875.00	1655	42.71
One year	1473.00	166	11.27
Two year	1695.00	48	2.83

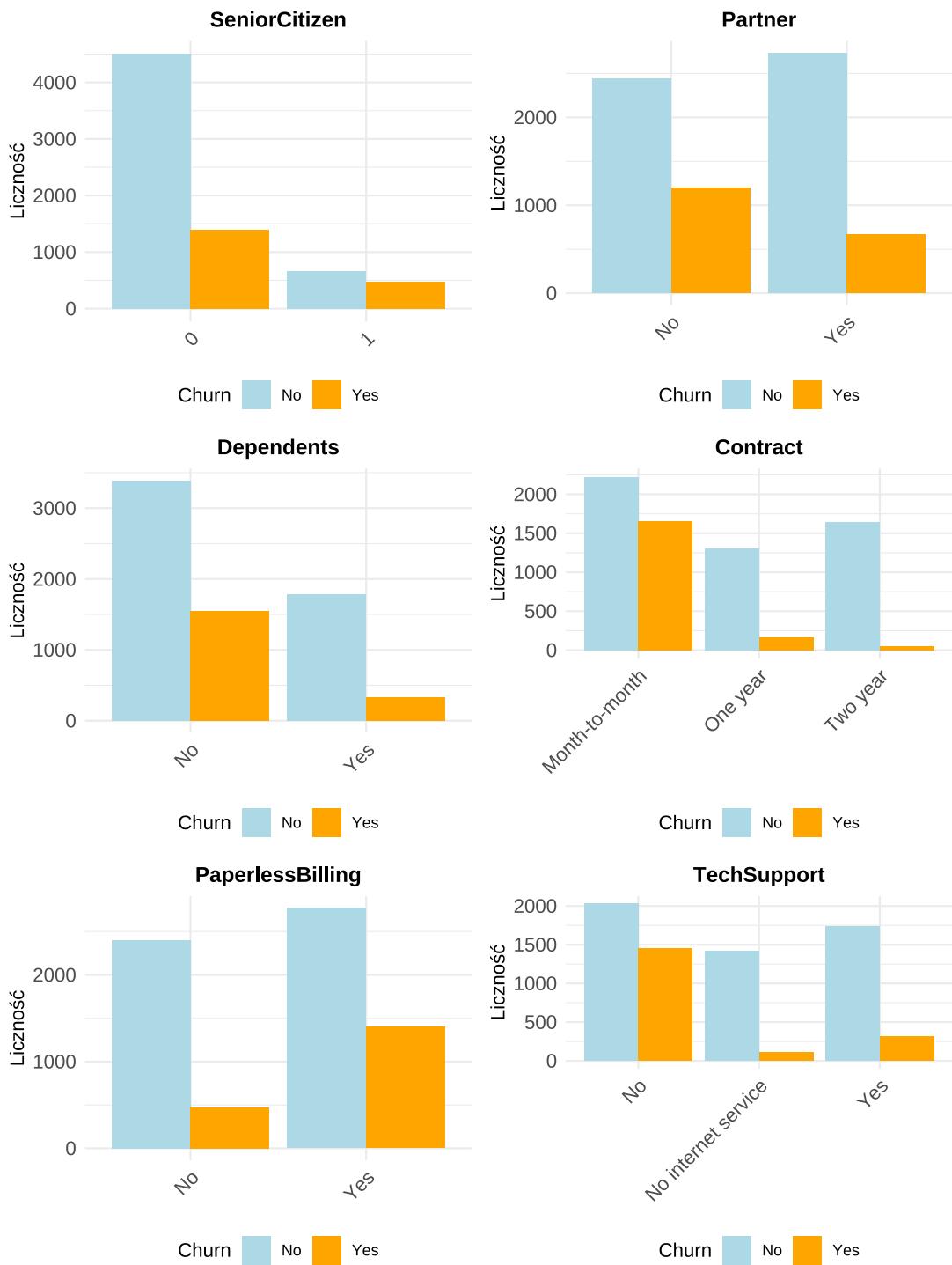
Tabela 17: Tabela częstości dla zmiennej PaperlessBilling z podziałem na Churn

Kategoria	Liczba_klientow	Liczba_odchodzacych	Procent_odchodzacych
No	2872.00	469	16.33
Yes	4171.00	1400	33.57

Tabela 18: Tabela częstości dla zmiennej TechSupport z podziałem na Churn

Kategoria	Liczba_klientow	Liczba_odchodzacych	Procent_odchodzacych
No	3473.00	1446	41.64
No internet service	1526.00	113	7.40
Yes	2044.00	310	15.17

4.2.2 Wykresy słupkowe



Rysunek 10: Wykresy słupkowe przedstawiające rozkład częstości wybranych zmiennych jakościowych z podziałem na grupy Churn

Analiza rozkładu zmiennych z podziałem na grupy Churn (Rysunek 10) w połączeniu z danymi z tabel (Tabele: 13, 14, 15, 16, 17, 18) ujawnia zależności:

- **Contract**: Widoczna jest silna zależność między typem umowy a rezygnacją z usług (Tabela 16). Klienci z umowami miesięcznymi znacznie częściej rezygnują z usług (około 42% przypadków) w porównaniu do tych z umowami dwuletnimi (jedynie około 11% rezygnacji)
- **SeniorCitizen**: Jak pokazuje Tabela 13, seniorzy, mimo że stanowią mniejszą część klientów, wykazują proporcjonalnie wyższy wskaźnik rezygnacji - około 41% seniorów rezygnuje z usług, podczas gdy wśród pozostałych klientów odsetek ten wynosi około 24%
- **PaperlessBilling**: Z danych w Tabeli 17 wynika, że klienci korzystający z e-faktur częściej rezygnują z usług (około 33%) w porównaniu do klientów preferujących tradycyjne faktury (około 17%), co może wskazywać na ich większą aktywność w poszukiwaniu lepszych ofert
- **Partner/Dependents**: Tabele 14 i 15 pokazują, że osoby bez partnerów (około 31% rezygnacji) lub osób zależnych (około 30% rezygnacji) wykazują tendencję do częstszego zrywania umowy niż osoby posiadające partnera (około 21% rezygnacji) lub osoby zależne (około 20% rezygnacji), co sugeruje, że zobowiązania rodzinne mogą skłaniać do pozostania przy obecnym dostawcy usług
- **TechSupport**: Tabela 18 ujawnia, że klienci bez wsparcia technicznego mają znacznie wyższy wskaźnik rezygnacji (około 40%) w porównaniu do klientów korzystających z tej usługi (około 15%), co sugeruje, że dobrze funkcjonujące wsparcie techniczne może być kluczowym czynnikiem utrzymania klientów

4.2.3 Zmienne najlepiej rozróżniające grupy klientów rezygnujących i pozostających

Przeprowadzona analiza danych z podziałem na grupy klientów, którzy zrezygnowali z usług (**Churn = Yes**) oraz tych, którzy pozostały (**Churn = No**), pozwala wyciągnąć następujące wnioski dotyczące zmiennych najlepiej rozróżniających te grupy:

1. **Contract** - zmienna ta wykazuje najsilniejsze zróżnicowanie między grupami. Klienci z umowami miesięcznymi rezygnują z usług ponad trzykrotnie częściej (ok. 42%) niż klienci z umowami dwuletnimi (ok. 11%)
2. **TechSupport** - dostęp do wsparcia technicznego istotnie wpływa na utrzymanie klienta. Osoby bez wsparcia technicznego rezygnują znacznie częściej (ok. 40%) niż klienci korzystający z tej usługi (ok. 15%)

5 Podsumowanie - wnioski z przeprowadzonej analizy

5.1 Podsumowanie wyników

Przeprowadzona analiza eksploracyjna danych klientów firmy telekomunikacyjnej wykazała, że zbiór danych zawiera 7043 obserwacje z 21 zmiennymi opisującymi klientów oraz zakres usług, z których korzystają. Analiza jednowymiarowa ujawniła nierównomierne rozkłady w kilku zmiennych, w szczególności przewagę umów miesięcznych oraz niski udział seniorów wśród klientów. Poziom rezygnacji klientów (**Churn**) wynosi około 26,5%, co wskazuje na istotny problem zatrzymania klientów.

W analizie dwuwymiarowej zidentyfikowano kluczowe czynniki różnicujące klientów rezygnujących i pozostających. Zmienne jakościowe **Contract**, **TechSupport**, **SeniorCitizen** oraz **PaperlessBilling** wykazały najsilniejsze zależności ze zmienną **Churn**. Wśród zmiennych ilościowych największe różnice między grupami obserwujemy dla zmiennych **tenure** (długość korzystania z usług) oraz **MonthlyCharges** (miesięczne opłaty).

5.2 Charakterystyka klientów firmy

Klienci analizowanej firmy telekomunikacyjnej charakteryzują się następującymi cechami:

- Przeważającą część stanowią osoby młodsze, nie będące seniorami (około 84%)
- Liczba klientów posiadających i nieposiadających partnerów jest zbliżona (około 50% do 50%)
- Większość klientów nie posiada osób zależnych (około 70%)
- Dominują klienci z umowami miesięcznymi (około 55%), podczas gdy umowy jedno- i dwuletnie stanowią mniejszość
- Większość klientów posiada elektroniczne faktury (około 59%)
- Znacząca część klientów nie korzysta z usług wsparcia technicznego
- Przeciętny czas korzystania z usług (**tenure**) wynosi około 32 miesięcy, przy czym wielu klientów korzysta z usług krócej niż rok
- Średnie miesięczne opłaty (**MonthlyCharges**) wynoszą około 65 dolarów, a ich rozkład jest dość równomierny
- Całkowite opłaty (**TotalCharges**) silnie korelują z czasem korzystania z usług i wykazują znaczne zróżnicowanie wśród klientów

5.3 Główne przyczyny odchodzenia klientów i rekomendacje

Na podstawie przeprowadzonych analiz można wskazać następujące główne przyczyny odchodzenia klientów:

1. **Krótki czas korzystania z usług** - klienci o krótszym stażu (**tenure**) znacznie częściej rezygnują z usług, co wskazuje na kluczowy okres pierwszych 12 miesięcy dla utrzymania klienta

2. **Wyższe miesięczne opłaty** - klienci płacący wyższe miesięczne rachunki (**MonthlyCharges**) mają tendencję do częstszej rezygnacji z usług
3. **Rodzaj umowy** - klienci z umowami miesięcznymi rezygnują trzykrotnie częściej (42%) niż klienci z umowami dwuletnimi (11%)
4. **Brak wsparcia technicznego** - klienci bez tej usługi rezygnują znacznie częściej (40%) niż klienci z dostępem do wsparcia (15%)
5. **Status seniora** - seniorzy rezygnują częściej (41%) niż młodsi klienci (24%)
6. **Elektroniczne faktury** - klienci korzystający z e-faktur częściej rezygnują (33%) niż klienci preferujący tradycyjne faktury (17%)
7. **Brak zobowiązań rodzinnych** - osoby bez partnera oraz bez osób zależnych częściej rezygnują z usług

Rekomendacje dla firmy:

- Wdrożenie specjalnych programów dla nowych klientów
- Wprowadzenie zachęt do podpisywania długoterminowych umów
- Ulepszenie i automatyczne włączenie wsparcia technicznego jako standardowej usługi, szczególnie dla seniorów
- Opracowanie specjalnych pakietów dla seniorów
- Wprowadzenie programu lojalnościowego dla klientów korzystających z e-faktur, aby zrównoważyć ich większą skłonność do zmiany dostawcy
- Stworzenie dedykowanych planów rodzinnych z dodatkowymi korzyściami, aby zwiększyć lojalność klientów z zobowiązaniami rodzinnymi