

Raport 1

Eksploracja danych

Tomasz Warzecha, album 282261

2025-03-31

Spis treści

1	Przygotowanie danych	2
1.1	Krótki opis zagadnienia	2
1.2	Analiza struktury danych	3
2	Analiza opisowa	3
2.1	Wskaźniki sumaryczne	3
2.2	Wykresy	5
2.3	Wykresy zmiennych jakościowych	6
3	Wykresy rozrzutu	8
4	Analiza opisowa z podziałem na grupy	10
4.1	Wskaźniki sumaryczne dla grup	10
4.2	Wykresy zmiennych ilościowych dla grup	13
4.3	Wykresy zmiennych jakościowych dla grup	15
4.4	Największe zróżnicowanie wartości/rozkładu zmiennych w grupach	16
5	Podsumowanie	17
5.1	Podsumowanie etapów 1-3	17
5.2	Charakteryzacja klientów firmy	17
5.3	Przyczyny odchodzenia klientów	17

1 Przygotowanie danych

1.1 Krótki opis zagadnienia

Niniejszy raport przedstawia analizę danych dotyczących rezygnacji klientów (churn) firmy telekomunikacyjnej. Głównym celem badania jest identyfikacja czynników wpływających na decyzję klientów o rezygnacji z usług. Dane pochodzą ze zbioru WA_Fn-UseC_-Telco-Customer-Churn.csv dostępnego na platformie Kaggle i zawierają informacje o 7043 klientach, w tym ich charakterystyki demograficzne, subskrybowane usługi oraz status konta.

Tabela 1: Struktura danych

x	
customerID	factor
gender	factor
SeniorCitizen	integer
Partner	factor
Dependents	factor
tenure	integer
PhoneService	factor
MultipleLines	factor
InternetService	factor
OnlineSecurity	factor
OnlineBackup	factor
DeviceProtection	factor
TechSupport	factor
StreamingTV	factor
StreamingMovies	factor
Contract	factor
PaperlessBilling	factor
PaymentMethod	factor
MonthlyCharges	numeric
TotalCharges	numeric
Churn	factor

Zmienna **SeniorCitizen** powinna być jakościowa, a nie ilościowa, gdyż przyjmuje dwie wartości. Pozostałe zmienne zostały poprawnie rozpoznane jako jakościowe (factor) lub ilościowe (numeric).

Tabela 2: Brakujące dane

	x
TotalCharges	11

1.2 Analiza struktury danych

Nasze dane mają 7043 przypadków i 21 cech. W powyższej tabeli możemy zobaczyć wszystkie cechy oraz ich typy. Mamy 3 cechy o wartościach numerycznych oraz 18 cech o wartościach typu factor(po zmianie typu zmiennej **SeniorCitizen**). Cecha **CustomerID** (factor) nie będzie nam potrzebna w dalszej analizie, zatem możemy ją usunąć. W naszych danych mamy 11 brakujących wartości, wszystkie znajdują się w kolumnie **TotalCharges** i są typu NA.

2 Analiza opisowa

2.1 Wskaźniki sumaryczne

Poniżej znajdują się wskaźniki sumaryczne dla zmiennych ilościowych.

Tabela 3: Podstawowe wskaźniki sumaryczne dla zmiennych **tenure**, **MonthlyCharges** i **TotalCharges**

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
tenure	0.00	9.00	29.00	32.37	55.00	72.00
MonthlyCharges	18.25	35.50	70.35	64.76	89.85	118.75
TotalCharges	18.80	401.45	1397.47	2283.30	3794.74	8684.80

Analiza zmiennych ilościowych pokazuje, że:

- **tenure** (czas trwania usługi w miesiącach) waha się od 1 do 72 miesięcy, przy średniej około 32 miesięcy
- **MonthlyCharges** (miesięczne opłaty) wynoszą średnio 64.76, z zakresem od 18.25 do 118.75
- **TotalCharges** (całkowite opłaty) zgodnie z intuicją wykazują największą zmienność, od 18.8 do 8684.8, gdzie średnia wynosi 1397.5

Poniżej znajdują się wskaźniki dla zmiennych jakościowych

Tabela 4: wskaźniki sumaryczne dla zmiennych jakościowych

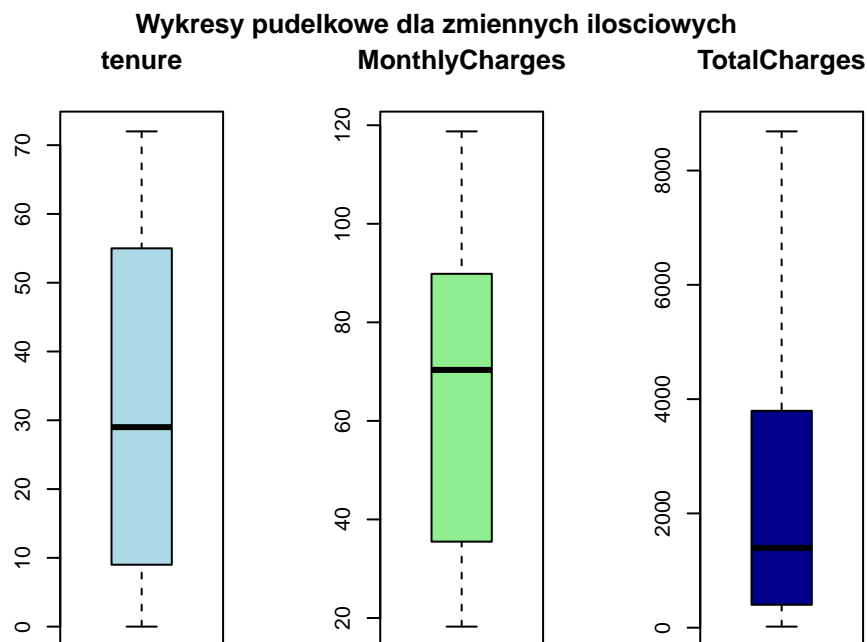
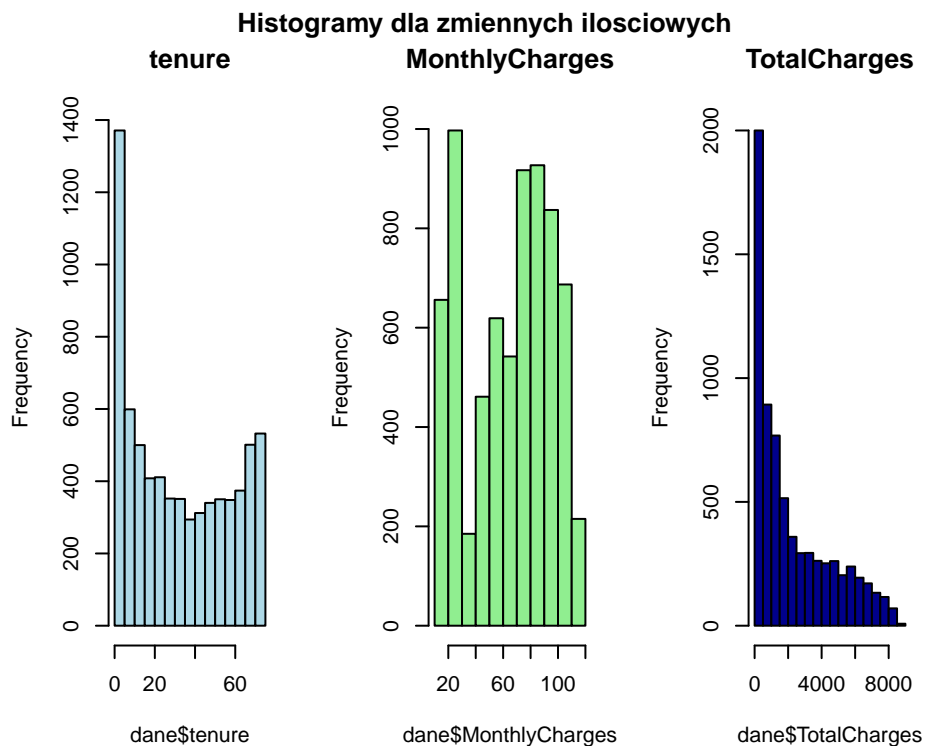
Zmienna	Rozkład
gender	Female : 3488; Male : 3555
SeniorCitizen	No : 5901; Yes : 1142
Partner	No : 3641; Yes : 3402
Dependents	No : 4933; Yes : 2110
PhoneService	No : 682; Yes : 6361
MultipleLines	No : 3390; No phone service : 682; Yes : 2971
InternetService	DSL : 2421; Fiber optic : 3096; No : 1526
OnlineSecurity	No : 3498; No internet service : 1526; Yes : 2019
OnlineBackup	No : 3088; No internet service : 1526; Yes : 2429
DeviceProtection	No : 3095; No internet service : 1526; Yes : 2422
TechSupport	No : 3473; No internet service : 1526; Yes : 2044
StreamingTV	No : 2810; No internet service : 1526; Yes : 2707
StreamingMovies	No : 2785; No internet service : 1526; Yes : 2732
Contract	Month-to-month : 3875; One year : 1473; Two year : 1695
PaperlessBilling	No : 2872; Yes : 4171
PaymentMethod	Bank transfer (automatic) : 1544; Credit card (automatic) : 1522; Electronic check :
Churn	No : 5174; Yes : 1869

Kluczowe obserwacje dotyczące zmiennych jakościowych:

Rozkład płci wśród klientów jest niemal równy – liczba kobiet i mężczyzn jest bardzo zbliżona. Wśród użytkowników stosunkowo niewielki odsetek stanowią seniorzy. Znaczna część klientów zdecydowała się na rezygnację z usług, co wskazuje na dość wysoką rotację. Jeśli chodzi o preferowane typy umów, większość osób wybiera opcję miesięczną, a ponad połowa klientów korzysta z elektronicznego systemu rozliczeń, rezygnując z tradycyjnych papierowych faktur.

2.2 Wykresy

Poniżej znajdują się histogramy oraz wykresy pudełkowe dla zmiennych ilościowych.

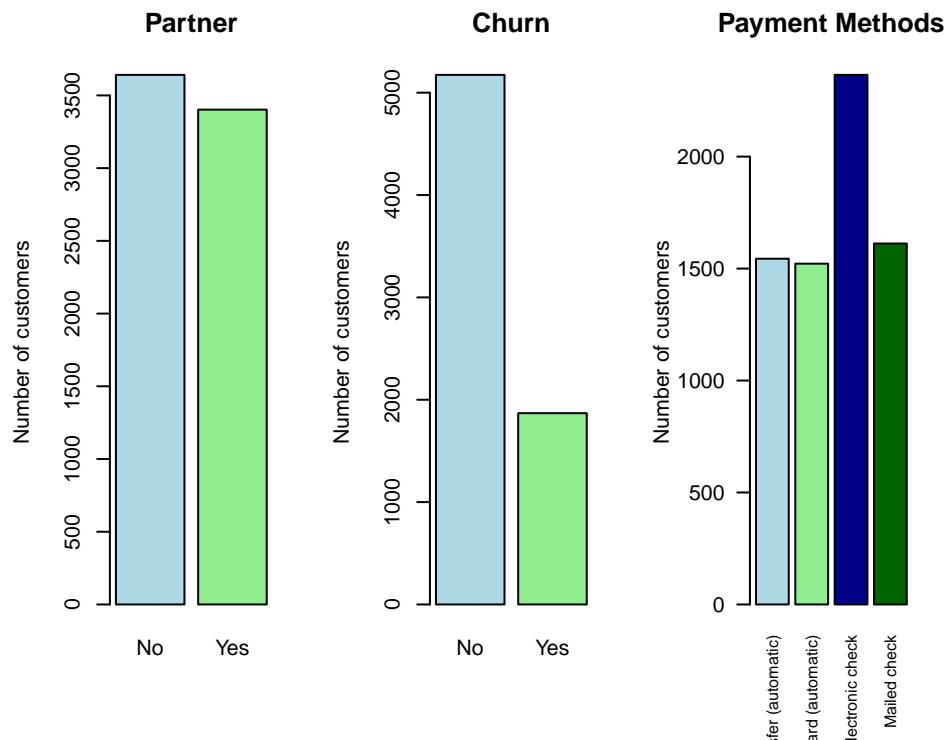


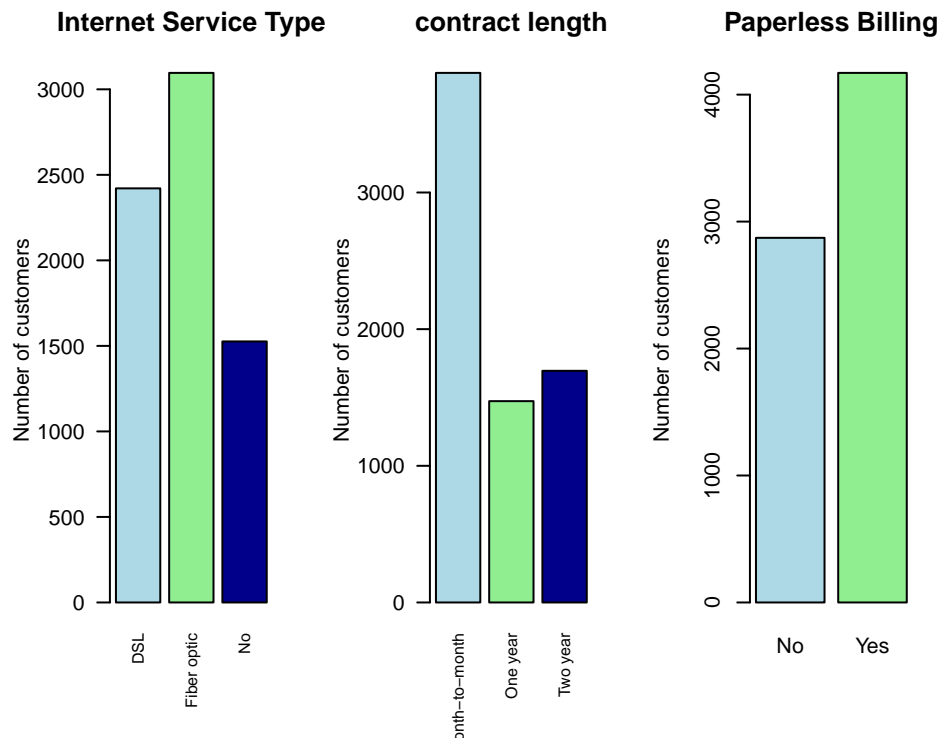
Histogramy i wykresy pudełkowe pokazują, że:

- Rozkład czasu trwania usługi (**tenure**) jest najbardziej zbliżony do symetrycznego, jednak w dalszym ciągu widzimy znaczącą przewagę nowych klientów. Warto jednak zauważyć że jest spora część lojalnych klientów, którzy zostają na dłużej.
- Miesięczne opłaty (**MonthlyCharges**) mają rozkład który wybija się w dwóch miejscach, co może sugerować że mamy dwa najpopularniejsze pakiety usług o różnych cenach.
- Całkowite opłaty (**TotalCharges**) dla większości klientów są niskie, co jest spodziewane przy zmiennej będącej iloczynem czasu i stawek, szczególnie podczas gdy większość naszych klientów jest nowa.

2.3 Wykresy zmiennych jakościowych

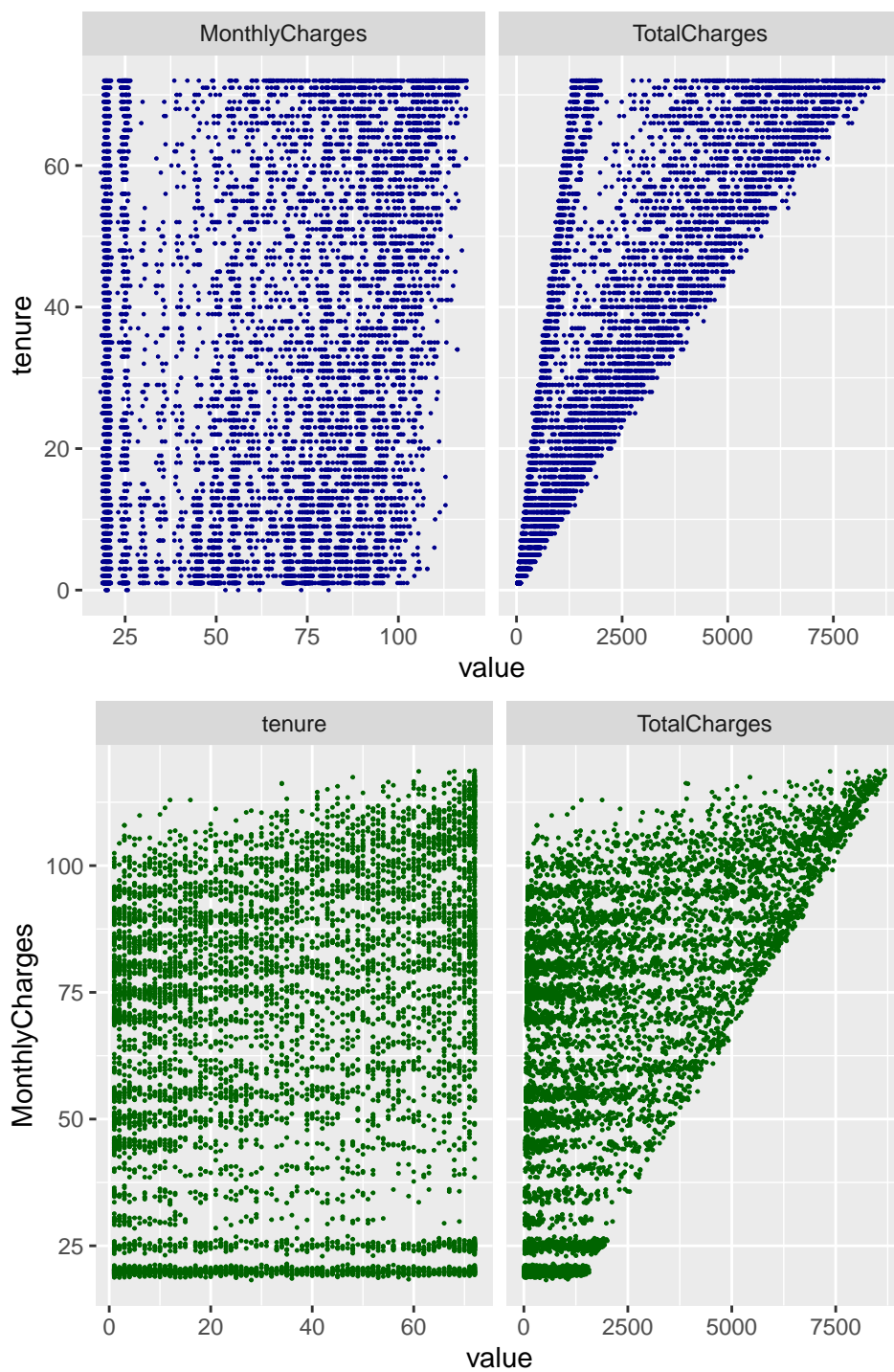
Ponizej wykresy dla zmiennych jakościowych:

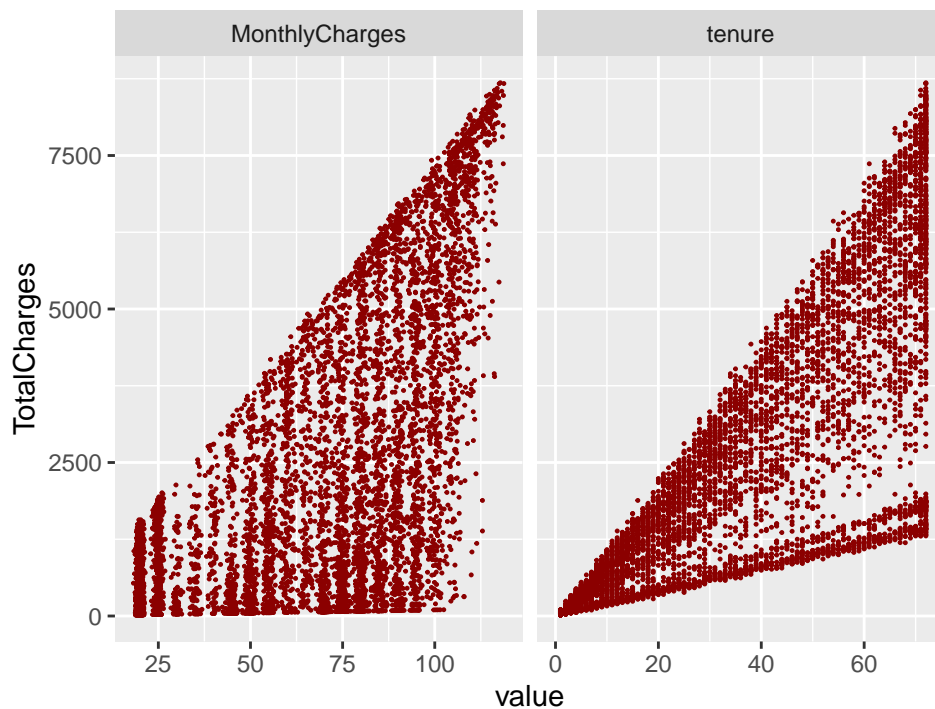




Większość klientów nie posiada partnera, jednak różnice w liczebności obu grup nie są znaczące. Analizując metody płatności, widać, że najczęściej wybraną opcją są czeki elektroniczne, podczas gdy reszta płatności jest mniej popularna. Może to wynikać z wygody oraz przyzwyczajenia klientów. Jeśli chodzi o rodzaj internetu, największa grupa klientów korzysta z internetu światłowodowego, natomiast DSL nadal ma znaczącą liczbę użytkowników. Część klientów nie korzysta z internetu w ramach tej usługi. Najczęściej wybieranym typem umowy jest miesięczna, co może świadczyć o potrzebie elastyczności i unikaniu długoterminowych zobowiązań. Kontrakty dwuletnie są znacznie mniej popularne. Z kolei faktury elektroniczne są wybierane przez większość klientów, co sugeruje, że klienci preferują wygodne i ekologiczne rozwiązania.

3 Wykresy rozrzutu





Wykresy rozrzutu nie wskazują na istnienie wyraźnych zależności liniowych między analizowanymi zmiennymi. Najbardziej zauważalna korelacja pojawia się w przypadku zależności między długością korzystania z usługi (tenure) a całkowitymi opłatami (TotalCharges), jednak nie jest ona idealnie liniowa ze względu na duży rozrzut danych. Jest to zgodne z intuicją – im dłużej klient korzysta z usługi, tym większe łączne koszty ponosi.

4 Analiza opisowa z podziałem na grupy

4.1 Wskazniki sumaryczne dla grup

Wskazniki sumaryczne dla zmiennych ilościowych dla grup:

Tabela 5: Podstawowe wskaźniki sumaryczne dla zmiennych ilościowych (Churn = Yes)

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
tenure	1.00	2.00	10.00	17.98	29.0	72.00
MonthlyCharges	18.85	56.15	79.65	74.44	94.2	118.35
TotalCharges	18.85	134.50	703.55	1531.80	2331.3	8684.80

Tabela 6: Podstawowe wskaźniki sumaryczne dla zmiennych ilościowych (Churn = No)

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
tenure	0.00	15.00	38.00	37.57	61.00	72.00
MonthlyCharges	18.25	25.10	64.43	61.27	88.40	118.75
TotalCharges	18.80	577.82	1683.60	2555.34	4264.12	8672.45

Opis wskaźników sumarycznych dla klientów, którzy odeszli (Churn = Yes):

- **tenure:** Czas trwania usługi waha się od 1 do 72 miesięcy, przy medianie wynoszącej 10 miesięcy i średniej około 17.98 miesięcy. Wskazuje to, że większość klientów, którzy odeszli, korzystała z usług stosunkowo krótko.
- **MonthlyCharges:** Miesięczne opłaty wynoszą średnio 74.44, z medianą 79.65. Zakres wartości jest szeroki (od 18.85 do 118.35), co pokazuje zróżnicowanie wśród klientów.
- **TotalCharges:** Całkowite opłaty wahają się od 18.85 do 8684.80, przy średniej 1531.80 i medianie 703.55. Niższa mediana w porównaniu do średniej sugeruje, że wielu klientów opuściło usługę po stosunkowo niskich całkowitych opłatach.

Opis wskaźników sumarycznych dla klientów, którzy zostali (Churn = No):

- **tenure:** Czas trwania usługi wynosi od 0 do 72 miesięcy, przy medianie 38 miesięcy i średniej 37.57 miesięcy. Wskazuje to, że klienci, którzy zostali, korzystali z usług znacznie dłużej niż ci, którzy odeszli.
- **MonthlyCharges:** Miesięczne opłaty są średnio niższe (61.27) w porównaniu do klientów, którzy odeszli, z medianą 64.43. Zakres wartości jest podobny (od 18.25 do 118.75).
- **TotalCharges:** Całkowite opłaty są znacznie wyższe (średnia 2555.34, mediana 1683.60) niż w przypadku klientów, którzy odeszli, co jest spójne z dłuższym czasem trwania usługi.

Podsumowanie: Klienci, którzy odeszli, charakteryzują się krótszym czasem korzystania z usług (niższe wartości tenure) i niższymi całkowitymi opłatami (TotalCharges), pomimo wyższych średnich miesięcznych opłat (MonthlyCharges). Natomiast klienci, którzy zostali,

mają dłuższy czas trwania usługi i wyższe całkowite opłaty, co może odzwierciedlać ich większą lojalność i satysfakcję z usług.

Wskaźniki sumaryczne dla zmiennych jakościowych dla grup

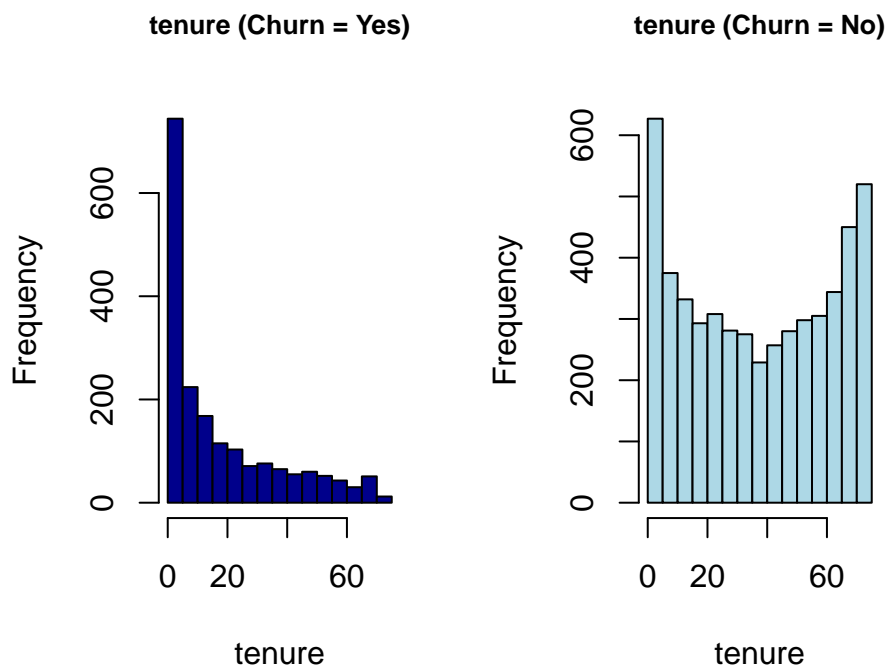
Tabela 7: Wskaźniki sumaryczne dla zmiennych jakościowych (Churn = Yes)

Zmienna	Rozkład
gender	Female : 939; Male : 930
SeniorCitizen	No : 1393; Yes : 476
Partner	No : 1200; Yes : 669
Dependents	No : 1543; Yes : 326
PhoneService	No : 170; Yes : 1699
MultipleLines	No : 849; No phone service : 170; Yes : 850
InternetService	DSL : 459; Fiber optic : 1297; No : 113
OnlineSecurity	No : 1461; No internet service : 113; Yes : 295
OnlineBackup	No : 1233; No internet service : 113; Yes : 523
DeviceProtection	No : 1211; No internet service : 113; Yes : 545
TechSupport	No : 1446; No internet service : 113; Yes : 310
StreamingTV	No : 942; No internet service : 113; Yes : 814
StreamingMovies	No : 938; No internet service : 113; Yes : 818
Contract	Month-to-month : 1655; One year : 166; Two year : 48
PaperlessBilling	No : 469; Yes : 1400
PaymentMethod	Bank transfer (automatic) : 258; Credit card (automatic) : 232; Electronic check : 10

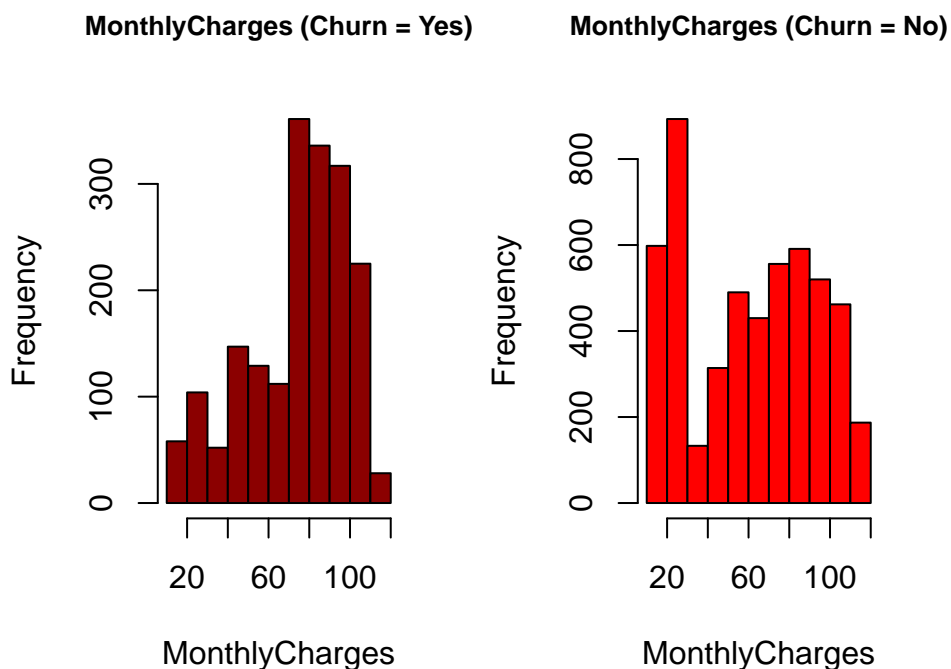
Tabela 8: Wskaźniki sumaryczne dla zmiennych jakościowych (Churn = No)

Zmienna	Rozkład
gender	Female : 2549; Male : 2625
SeniorCitizen	No : 4508; Yes : 666
Partner	No : 2441; Yes : 2733
Dependents	No : 3390; Yes : 1784
PhoneService	No : 512; Yes : 4662
MultipleLines	No : 2541; No phone service : 512; Yes : 2121
InternetService	DSL : 1962; Fiber optic : 1799; No : 1413
OnlineSecurity	No : 2037; No internet service : 1413; Yes : 1724
OnlineBackup	No : 1855; No internet service : 1413; Yes : 1906
DeviceProtection	No : 1884; No internet service : 1413; Yes : 1877
TechSupport	No : 2027; No internet service : 1413; Yes : 1734
StreamingTV	No : 1868; No internet service : 1413; Yes : 1893
StreamingMovies	No : 1847; No internet service : 1413; Yes : 1914
Contract	Month-to-month : 2220; One year : 1307; Two year : 1647
PaperlessBilling	No : 2403; Yes : 2771
PaymentMethod	Bank transfer (automatic) : 1286; Credit card (automatic) : 1290; Electronic check :

4.2 Wykresy zmiennych ilościowych dla grup

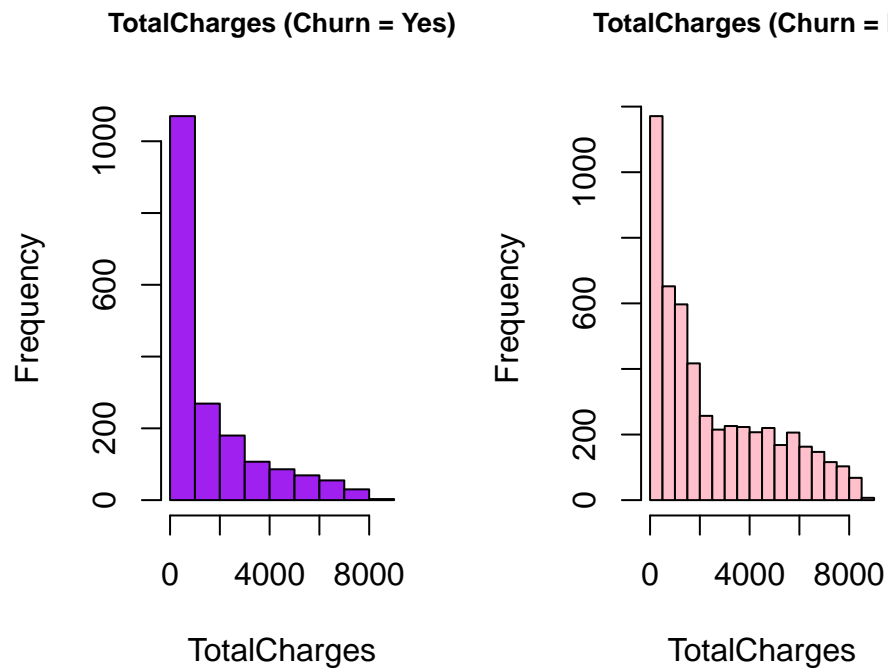


Dla klientów, którzy zrezygnowali, wartości tenure są często niskie – oznacza to, że wielu klientów rezygnuje w pierwszych miesiącach użytkowania. Dla klientów, którzy nie zrezygnowali, rozkład jest bardziej równomierny, a liczba klientów z długim stażem jest znacznie większa. Możemy stwierdzić, że klienci o krótkim stażu są bardziej skłonni do rezygnacji, co może wskazywać na problem z satysfakcją w początkowej fazie korzystania z usługi.



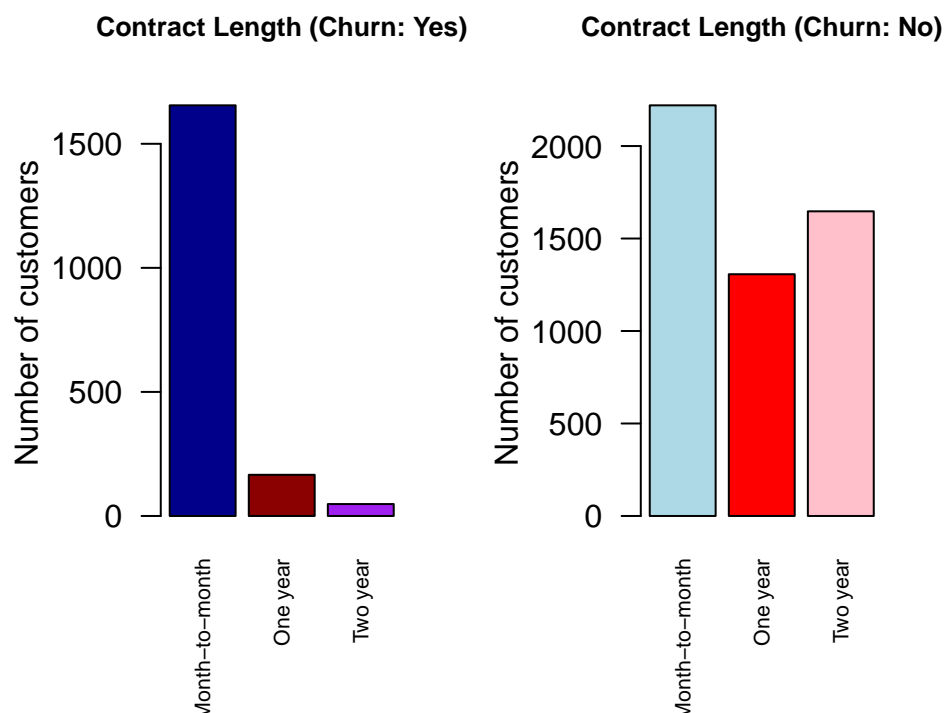
Na wykresach możemy zobaczyć że klienci, którzy zrezygnowali, częściej płacili wyższe miesięczne opłaty. Osoby, które pozostają, mają często niższe opłaty. Może to sugerować, że

droższe plany są mniej atrakcyjne dla klientów lub powodują większe niezadowolenie.

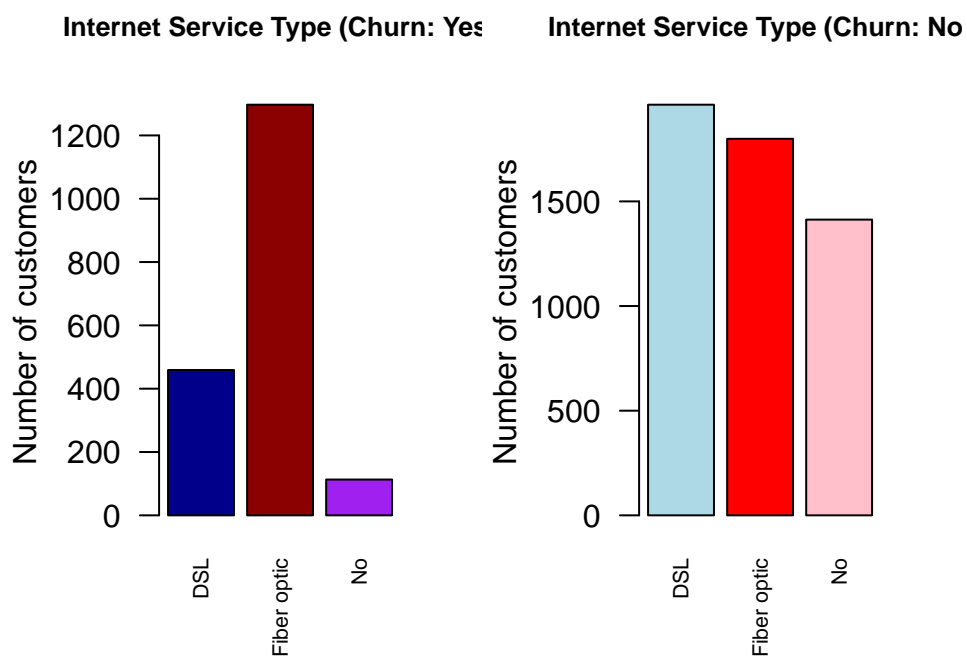


Wartości TotalCharges są niskie dla Churn = Yes, co są zgodne z intuicją, ponieważ ci klienci rezygnują szybko i nie generują dużych sum płatności. Dla Churn = No widzimy szerszy zakres wartości, co jest zgodne z dłuższym czasem trwania umowy.

4.3 Wykresy zmiennych jakościowych dla grup

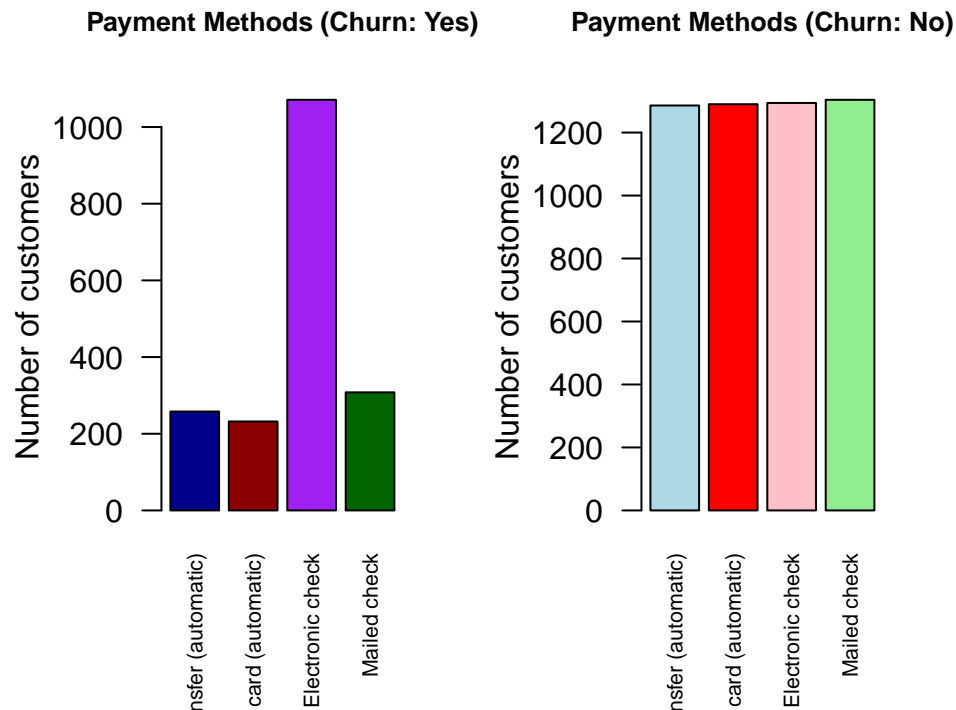


Klienci posiadający umowy miesięczne znacznie częściej decydują się na rezygnację w porównaniu do osób, które zawarły długoterminowe umowy. Może to wynikać z faktu, że nie czują się związani z firmą oraz nie mają żadnych zobowiązań, co ułatwia zmianę dostawcy. Z kolei osoby posiadające długoterminowe zobowiązania rzadziej rezygnują, ponieważ wcześniejsze zerwanie takiej umowy często wiąże się z dodatkowymi kosztami lub karami finansowymi.



Klienci korzystający z technologii Fiber optic często rezygnują z usług. Może to sugerować, że

technologia światłowodowa, wiąże się z wyższymi kosztami lub usługa nie spełnia oczekiwań, co skłania klientów do poszukiwania alternatyw. Z kolei osoby korzystające z DSL są bardziej lojalne, możliwe że ze względu na niższe opłaty lub stabilność usługi.



Klienci korzystający z płatności elektronicznym czekiem są znacznie bardziej skłonni do rezygnacji w porównaniu reszty osób. Może to sugerować, że osoby wybierające manualne formy płatności są mniej zaangażowane w korzystanie z usługi. Z kolei automatyczne metody płatności mogą sprzyjać dłuższemu pozostawaniu w usłudze, ponieważ użytkownicy rzadziej podejmują decyzję o zmianie dostawcy.

4.4 Największe zróżnicowanie wartości/rozkładu zmiennych w grupach

Na podstawie przeprowadzonych analiz, największe zróżnicowanie wartości oraz rozkłady między grupami klientów, wykazują następujące zmienne:

-**Contract** (Długość umowy): Ta zmienna pozwala na bardzo dobre rozróżnienie grup klientów. Osoby posiadające umowy miesięczne znacznie częściej rezygnują, podczas gdy klienci z umowami rocznymi i dwuletnimi pozostają znacznie bardziej lojalni. Różnice w rozkładzie są bardzo wyraźne, co czyni tę zmienną jednym z najlepszych predyktorów rezygnacji.

-**PaymentMethod** (Metoda płatności): Klienci korzystający z płatności elektronicznym czekiem znacznie częściej rezygnują z usług, natomiast osoby płacące automatycznie rzadziej odchodzą. To sugeruje, że wybór mniej wygodnej formy płatności może być wskaźnikiem większego prawdopodobieństwa rezygnacji.

-**InternetService** (Rodzaj usługi internetowej): Klienci korzystający z Fiber optic częściej

rezygnują z usług w porównaniu do użytkowników DSL lub osób, które w ogóle nie mają internetu w ramach swojej umowy.

-**MonthlyCharges** (Miesięczne opłaty): Tutaj nie mamy aż tak wyraźnych różnic, natomiast klienci z wyższymi miesięcznymi rachunkami trochę częściej rezygnują względem tych z niższymi opłatami.

5 Podsumowanie

5.1 Podsumowanie etapów 1-3

Przeprowadzona analiza danych pozwoliła na identyfikację kluczowych czynników wpływających na rezygnację klientów z usług firmy telekomunikacyjnej. W etapie przygotowania danych skorygowano typ zmiennych, usunięto zbędne kolumny i przeanalizowano braki w danych. W etapie analizy opisowej zauważono, że klienci są mało wierni firmie, oraz że najpopularniejsze są dwa pakiety w firmie. Między zmiennymi ciągłymi nie zauważono wyraźnych zależności liniowych, jedynie zależności między **tenure** a **TotalCharges** przypominają liniowe. W etapie analizy z podziałem na grupy zauważono, że klienci rezygnujący z usług charakteryzują się krótszym czasem korzystania z nich, wyższymi opłatami miesięcznymi oraz niższymi opłatami całkowitymi. W przypadku zmiennych kategoriowych wykazano, że rezygnujący częściej wybierali umowy miesięczne, płatności elektronicznym czekiem oraz usługi światłowodowe.

5.2 Charakterystyka klientów firmy

Klienci firmy to w większości osoby w średnim wieku, z niemal równym podziałem płci, rzadko korzystające z dodatkowych usług, takich jak ochrona online czy backup. Dominują użytkownicy internetu światłowodowego, choć znacząca część nie ma w ogóle usługi internetowej w ramach oferty. Większość klientów wybiera elastyczne umowy miesięczne i faktury elektroniczne, co może wynikać z preferencji wygody i braku chęci do długoterminowych zobowiązań. Sporo klientów wybiera także чеки elektroniczne jako sposób płatności. Niski odsetek seniorów wśród klientów sugeruje, że firma może nie być wystarczająco atrakcyjna dla starszych użytkowników.

5.3 Przyczyny odchodzenia klientów

Rezygnacja klientów jest silnie powiązana z krótkim czasem korzystania z usług, wysokimi opłatami miesięcznymi oraz wyborem umów miesięcznych. Klienci korzystający z internetu światłowodowego częściej odchodzą, co może wskazywać na problem z ceną lub jakością tej usługi. Dodatkowo, metoda płatności elektronicznym czekiem, wymagająca aktywnego działania klienta, może zachęcać do zmiany operatora. Aby ograniczyć odchodzenie klientów, firma powinna wprowadzić programy zniżkowe skierowane do nowych klientów. Można również wprowadzić zniżki na usługi z kontraktem długoterminowym, w krótkiej perspektywie może się to wydawać nieopłacalne, jednak pomoże to zachować klientów dłużej przy sobie. Należy także zwrócić uwagę na usługi światłowodowe, gdyż może występować jakiś problem z nimi lub ich cena może być zbyt wysoka w porównaniu do jakości, a klientów decydujących się

na to rozwiązanie z pewnością jest najwięcej. Można również zachęcać klientów do płatności automatycznych, ponieważ ułatwia to klientom zapłatę za usługi i sprawia że większa ilość klientów zostaje. Wdrożenie programu lojalnościowego mogłoby również zwiększyć satysfakcję długoletnich klientów i zmniejszyć wskaźnik rezygnacji.