

Sprawozdanie z listy nr 1

Eksploracja Danych

Dawid Skowroński 282241, Mateusz Cieślak 272633

2025-04-30

Spis treści

1	Wstęp	2
2	Przygotowanie danych	2
2.1	Opis danych	2
3	Analiza opisowa	4
3.1	Wskaźniki sumaryczne	4
3.2	Wykresy poszczególnych zmiennych	6
3.2.1	Histogramy	6
3.2.2	Wykresy pudełkowe	7
3.2.3	Wykresy słupkowe	8
3.2.4	Wykresy rozrzutu	10
4	Analiza opisowa z podziałem na grupy	12
4.1	Wskaźniki sumaryczne z podziałem na grupy	12
4.2	Wykresy z podziałem na grupy	13
4.2.1	Zmienne ilościowe	13
4.2.2	Zmienne jakościowe	14
4.3	Zależności liniowe z podziałem na grupy	16
4.4	Które ze zmiennych wykazują największe zróżnicowanie?	17
5	Podsumowanie – wnioski z przeprowadzonej analizy	18
5.1	Krótkie podsumowanie	18
5.2	Co wiemy o klientach firmy?	18
5.3	Analiza etapu trzeciego	18
5.3.1	Przyczyny odchodzenia	18
5.3.2	Propozycje przeciwdziałania	19

Spis wykresów

1	Histogramy zmiennych tenure, MonthlyCharges, TotalCharges wraz z gęstościami empirycznymi	6
2	Wykresy pudełkowe zmiennej tenure z podziałem na typ umowy	7
3	Wykresy zmiennych Dependents, Partner oraz zmiennej Partner pogrupowanej względem zmiennej Dependents	8
4	Wykresy zmiennych SeniorCitizen, PhoneService oraz MultipleLines	8
5	Wykresy zmiennych OnlineBackup, StreamingTV oraz TechSupport	9
6	Wykresy zmiennych Contract, PaperlessBilling oraz Churn	9
7	Wykresy rozrzutu dla zmiennych TotalCharges, tenure oraz opłat całkowitych wyliczonych za pomocą MonthlyCharges i tenure	10

8	Wykresy rozrzutu dla par zmiennych ciągłych (tenure, MonthlyCharges, TotalCharges)	11
9	Porównanie rozkładu zmiennych tenure, MonthlyCharges, TotalCharges dla dwóch grup: lojalnych i odchodzących klientów	13
10	Wpływ typu usługi internetowej i metody płatności na wskaźnik rezygnacji (Churn)	14
11	Wpływ typu umowy i osób na utrzymaniu na wskaźnik rezygnacji (Churn)	14
12	Wpływ bycia seniorem i płatności bezgotówkowej na wskaźnik rezygnacji (Churn)	15
13	Wpływ bycia w związku na wskaźnik rezygnacji (Churn)	15
14	Zależność łącznej sumy opłat od czasu subskrypcji z podziałem na klientów lojalnych i nielojalnych.	16
15	Wykres rozrzutu zmiennej MonthlyCharges od zmiennej z podziałem na Churn	17

Spis tabel

1	Struktura zbioru danych Telco Customer	2
2	Struktura zbioru danych Telco Customer	3
3	Liczba brakujących obserwacji z podziałem na cechy	4
4	Wybrane cechy obserwacji zbioru Telco, zawierające wartości brakujące	4
5	Wskaźniki sumaryczne zmiennych ilościowych	5
6	Macierz korelacji między zmiennymi ilościowymi	11
7	Porównanie wskaźników sumarycznych zmiennej tenure dla grup klientów lojalnych i odchodzących	12
8	Porównanie wskaźników sumarycznych zmiennej MonthlyCharges dla grup klientów lojalnych i odchodzących	12
9	Porównanie wskaźników sumarycznych zmiennej TotalCharges dla grup klientów lojalnych i odchodzących	12

1 Wstęp

Celem pracy jest zastosowanie metod analizy opisowej (wykresów i wskaźników sumarycznych) do analizy wybranych danych oraz szczegółowa interpretacja otrzymanych wyników. Wykorzystany został zbiór danych `WA_Fn-UseC-Telco-Customer-Churn.csv`, zawierający informacje o klientach firmy telekomunikacyjnej (źródło: Kaggle).

Dane tego typu są powszechnie używane w analizie migracji klientów (*churn analysis*). Proces ten polega na zbieraniu informacji o zachowaniach klientów w celu przewidzenia, którzy z nich mogą odejść z usług firmy, a następnie wdrożeniu działań zachęcających ich do pozostania.

2 Przygotowanie danych

2.1 Opis danych

Sprawdzamy poprawność typów zaimportowanych danych.

Tabela 1: Struktura zbioru danych Telco Customer

Zmienna	Typ	Przykładowe.wartości
customerID	factor	7590-VHVEG, 5575-GNVDE, 3668-QPYBK
gender	factor	Female, Male, Male
SeniorCitizen	integer	0, 0, 0
Partner	factor	Yes, No, No
Dependents	factor	No, No, No
tenure	integer	1, 34, 2
PhoneService	factor	No, Yes, Yes
MultipleLines	factor	No phone service, No, No
InternetService	factor	DSL, DSL, DSL
OnlineSecurity	factor	No, Yes, Yes

OnlineBackup	factor	Yes, No, Yes
DeviceProtection	factor	No, Yes, No
TechSupport	factor	No, No, No
StreamingTV	factor	No, No, No
StreamingMovies	factor	No, No, No
Contract	factor	Month-to-month, One year, Month-to-month
PaperlessBilling	factor	Yes, No, Yes
PaymentMethod	factor	Electronic check, Mailed check, Mailed check
MonthlyCharges	numeric	29.85, 56.95, 53.85
TotalCharges	numeric	29.85, 1889.5, 108.15
Churn	factor	No, No, Yes

Zbiór danych składa się z 7043 obserwacji oraz z 21 cech. Możemy zauważyć, że zmienna **customerID** została rozpoznana jako zmienna jakościowa, z racji występowania napisów w numerze identyfikacyjnym klienta. Powinna ona zostać zmieniona na typ **character**. Natomiast zmienna **SeniorCitizen** powinna być zmienną binarną, lecz została rozpoznana jako typ **numeric**.

Po zamianie typów niektórych cech, ponownie sprawdzamy strukturę ramki danych.

Tabela 2: Struktura zbioru danych Telco Customer

Zmienna	Opis	Typ	Przykładowe_wartości
customerID	Identyfikator klienta	character	7590-VHVEG, 5575-GNVDE
gender	Płeć klienta	factor	Female, Male
SeniorCitizen	Czy klient jest seniorem	factor	No, No
Partner	Czy klient ma partnera	factor	Yes, No
Dependents	Czy klient ma osoby na utrzymaniu	factor	No, No
tenure	Staż klienta w miesiącach	integer	1, 34
PhoneService	Czy klient korzysta z usług telefonicznych	factor	No, Yes
MultipleLines	Czy klient korzysta z wielu linii telefonicznych	factor	No phone service, No
InternetService	Typ usługi internetowej	factor	DSL, DSL
OnlineSecurity	Czy klient wykupił usługę bezpieczeństwa online	factor	No, Yes
OnlineBackup	Czy klient korzysta z kopii zapasowej online	factor	Yes, No
DeviceProtection	Czy klient korzysta z ochrony urządzeń	factor	No, Yes
TechSupport	Czy klient korzysta z wsparcia technicznego	factor	No, No
StreamingTV	Czy klient korzysta z telewizji streamingowej	factor	No, No
StreamingMovies	Czy klient korzysta z filmów streamingowej	factor	No, No
Contract	Typ umowy	factor	Month-to-month, One year
PaperlessBilling	Czy klient korzysta z faktur elektronicznych	factor	Yes, No
PaymentMethod	Metoda płatności	factor	Electronic check, Mailed check
MonthlyCharges	Miesięczne opłaty	numeric	29.85, 56.95
TotalCharges	Łączne opłaty	numeric	29.85, 1889.5
Churn	Czy klient zrezygnował z usług	factor	No, No

Zbiór danych zawiera 3 zmienne ilościowe (**tenure**, **MonthlyCharges**, **TotalCharges**). Pozostałe cechy są typu jakościowego. Na tym etapie możemy odrzucić cechy, które nie będą przydatne w analizie owego zbioru. Zmienna **customerID** opisuje numer identyfikacyjny klienta. Nie zawiera ona zatem żadnych wartościowych informacji dla dalszej analizy.

	Liczba wartości NA w kolumnach
gender	0
SeniorCitizen	0
Partner	0
Dependents	0
tenure	0
PhoneService	0
MultipleLines	0
InternetService	0
OnlineSecurity	0
OnlineBackup	0
DeviceProtection	0
TechSupport	0
StreamingTV	0
StreamingMovies	0
Contract	0
PaperlessBilling	0
PaymentMethod	0
MonthlyCharges	0
TotalCharges	11
Churn	0

Tabela 3: Liczba brakujących obserwacji z podziałem na cechy

Analizując wartości tabeli 3 możemy zauważyć, że braki zmiennych występują jedynie w zmiennej **TotalCharges**. Wszystkie wartości brakujące są zakodowane jako NA.

	tenure	MonthlyCharges	Churn
489	0	52.55	No
754	0	20.25	No
937	0	80.85	No
1083	0	25.75	No
1341	0	56.05	No
3332	0	19.85	No
3827	0	25.35	No
4381	0	20.00	No
5219	0	19.70	No
6671	0	73.35	No
6755	0	61.90	No

Tabela 4: Wybrane cechy obserwacji zbioru Telco, zawierające wartości brakujące

Występowanie wartości brakujących dla zmiennej **TotalCharges** może być spowodowane tym, że nowi klienci korzystają z usług firmy krócej niż miesiąc, przez co opłaty nie zostały jeszcze naliczone. Potwierdzają to dane z tabeli 4. Dla wszystkich tych klientów czas stażu wynosi niecały miesiąc.

3 Analiza opisowa

3.1 Wskaźniki sumaryczne

Wyznaczamy podstawowe wskaźniki sumaryczne dla zmiennych ilościowych: **tenure**, **MonthlyCharges**, **TotalCharges**.

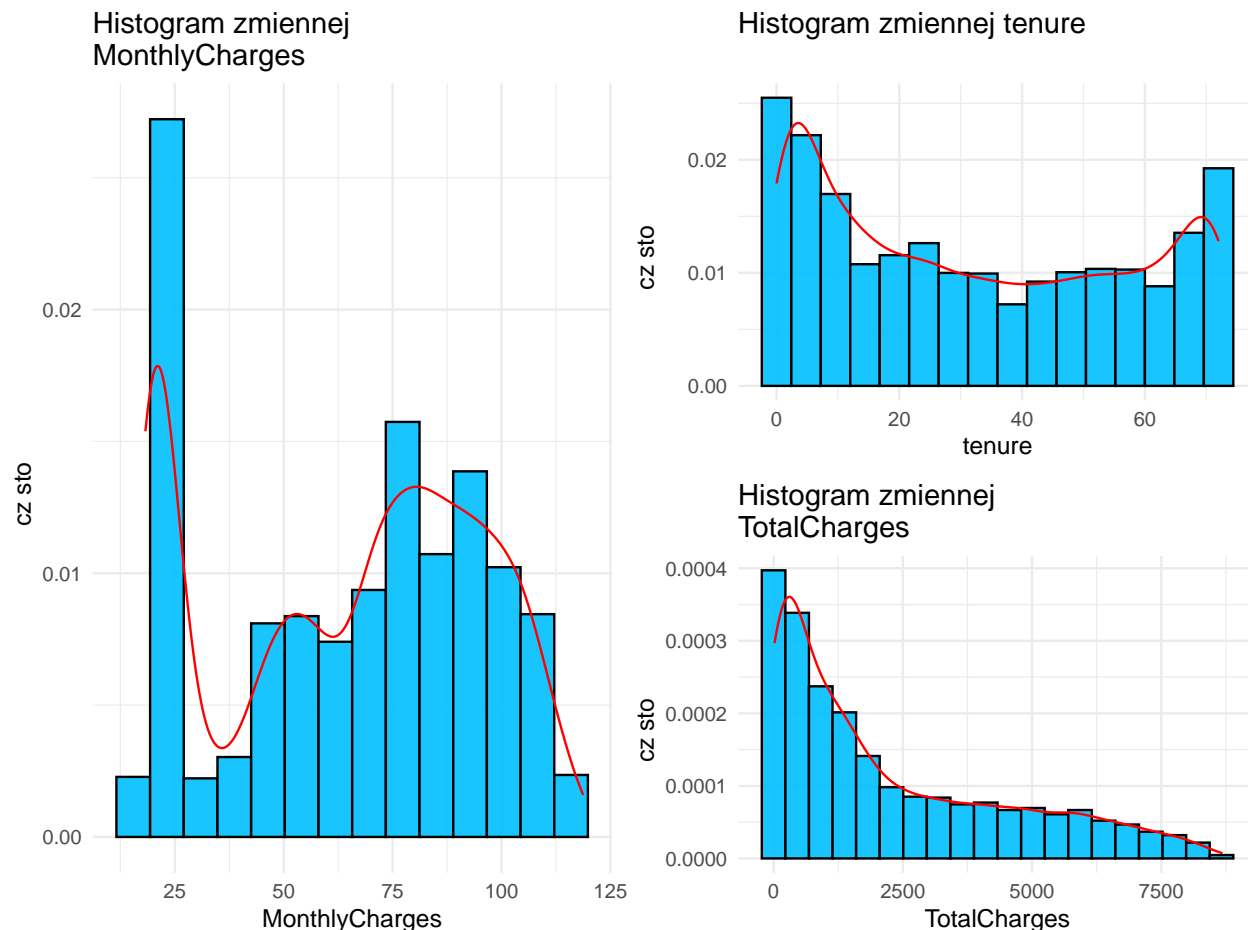
Wskaźnik	tenure	MonthlyCharges	TotalCharges
Min	0.00	18.25	18.80
Max	72.00	118.75	8684.80
Rozstęp	72.00	100.50	8666.00
Q1	9.00	35.50	401.45
Mediana	29.00	70.35	1397.47
Q3	55.00	89.85	3794.74
Var	603.17	905.41	5138252.41
Odchylenie standardowe	24.56	30.09	2266.77
IQR	46.00	54.35	3393.29
Współczynnik skośności	0.24	-0.22	0.96
Kurtoza	-1.39	-1.26	-0.23

Tabela 5: Wskaźniki sumaryczne zmiennych ilościowych

Z tabeli 5 możemy odczytać, że klienci o najdłuższym stażu korzystają z usług firmy przez 72 miesiące. Połowa klientów pozostała z firmą przez co najwyżej 29 miesięcy. Wartość odchylenia standardowego dla zmiennej **tenure** (**24.56**) oznacza duże rozproszenie danych. 50% klientów płaci między 35.5 a 89.85 dolarów miesięcznie. Pomimo dużego zróżnicowania ofert, klienci z reguły decydują się na droższe pakiety usług.

3.2 Wykresy poszczególnych zmiennych

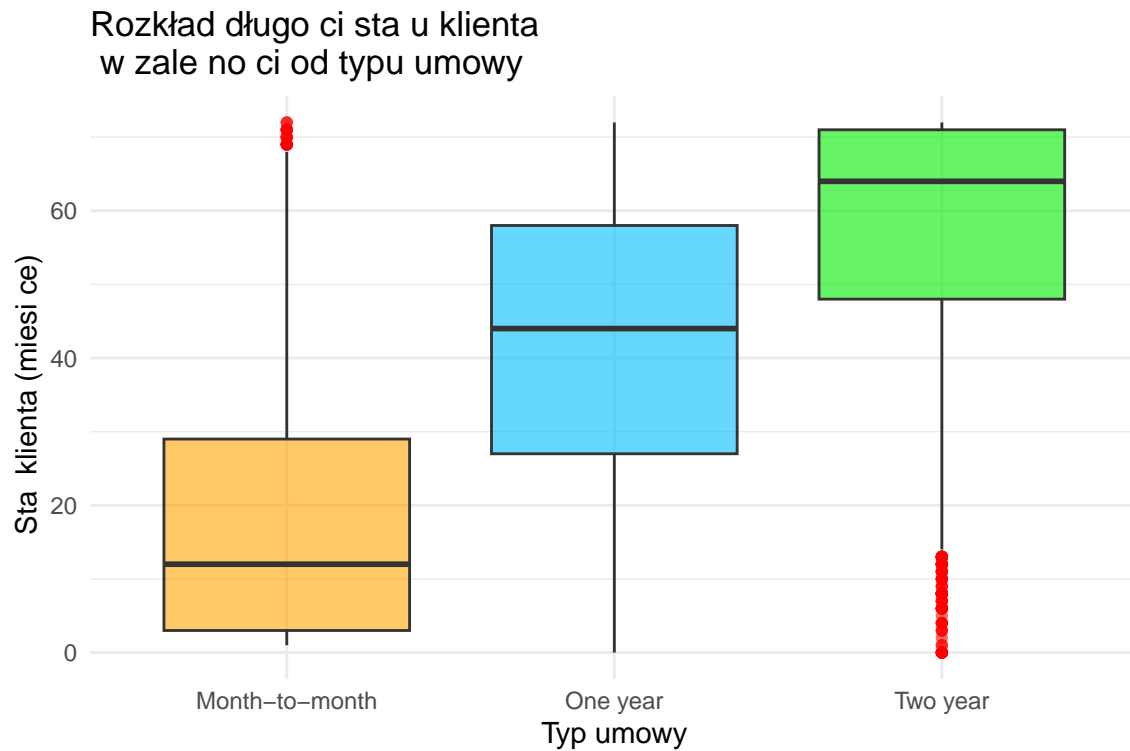
3.2.1 Histogramy



Wykres 1: Histogramy zmiennych tenure, MonthlyCharges, TotalCharges wraz z gęstościami empirycznymi

Z rysunku 1 możemy zauważyć, że rozkład zmiennej **TotalCharges** jest prawostronnie skośny, co potwierdza współczynnik skośności z tabeli 5 wynoszący **0.96**. Dodatkowo rozkład ten cechuje się ujemną wartością kurtozy (**-0.23**), co oznacza, że rozkład jest lekko spłaszczony. Jest ona mniejsza od wartości kurtozy dla standardowego rozkładu normalnego, dla której wartość kurtozy wynosi 0. To znaczy, że rozkład ten ma "łżejsze" ogony w porównaniu do ogonów rozkładu normalnego, a wartości odstające są przyjmowane z mniejszym prawdopodobieństwem. Zatem jest mniejsza szansa na to, że największe kwoty będzie płaciła większa liczba klientów niż np. dla rozkładu normalnego. Zauważalne jest spore rozproszenie danych. Rozstęp kwartylowy wynosi **3393.29**, co wskazuje na ekstremalne różnice w generowanych przychodach od klientów. Histogram zmiennej **tenure** wskazuje na to, że najwięcej jest nowo przybyłych klientów. Drugą, najbardziej liczną grupą, są klienci o długim stażu, przekraczającym 60 mies. Rozkład zmiennej **MonthlyCharges** jest trójmodalny (maksyma lokalne są osiągane w pobliżu wartości **20**, **50** i **80**). Oferty w granicach 20-25 \$ cieszą się dużą popularnością wśród klientów. Rozkład stażu klientów jest bimodalny. Najwięcej jest nowo przybyłych klientów (staż około od 0 do 5 miesięcy) i długoterminowych (65-70 miesięcy).

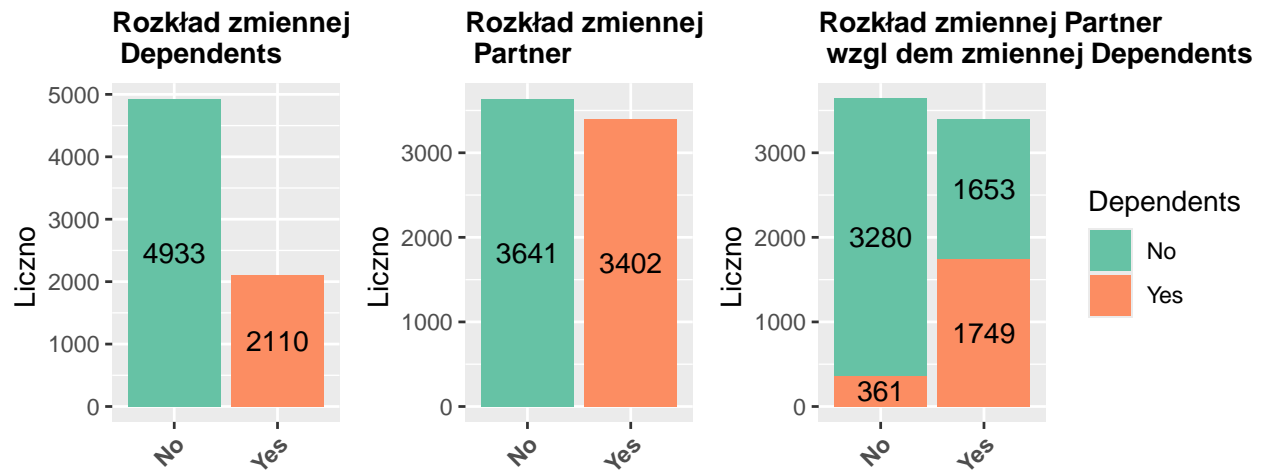
3.2.2 Wykresy pudełkowe



Wykres 2: Wykresy pudełkowe zmiennej tenure z podziałem na typ umowy

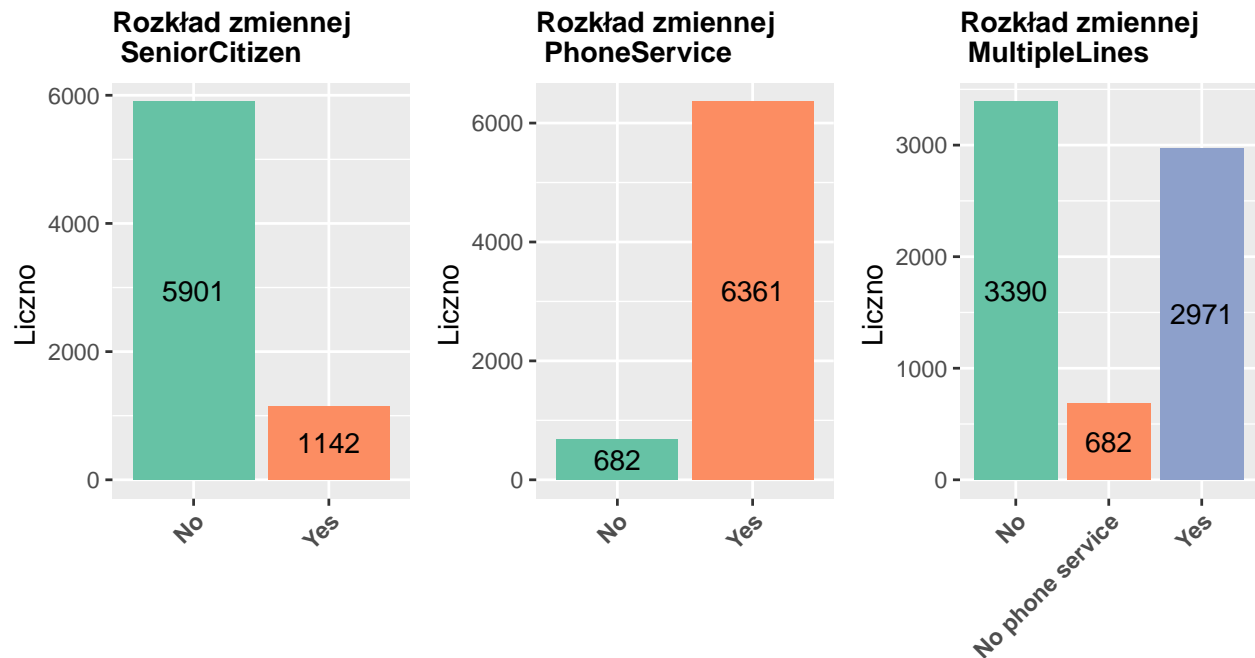
Klienci, którzy korzystają z umowy miesięcznej, mają zdecydowanie krótszy staż niż klienci korzystający z pozostałych rodzajów. Może to być spowodowane tym, że są to nowi klienci lub rezygnują z usług po krótszym czasie. Mediana dla umów rocznych **tenure** jest wyraźnie wyższa niż dla umów miesięcznych, co sugeruje, że klienci roczni zostają średnio dłużej. Rozkład ten jest symetryczny (co obrazuje wykres 2). Najdłuższy staż mają klienci, którzy są na umowie dwuletniej. Zatem wybór umowy przez klienta ma zdecydowany wpływ na długość korzystania ze świadczonych usług.

3.2.3 Wykresy słupkowe



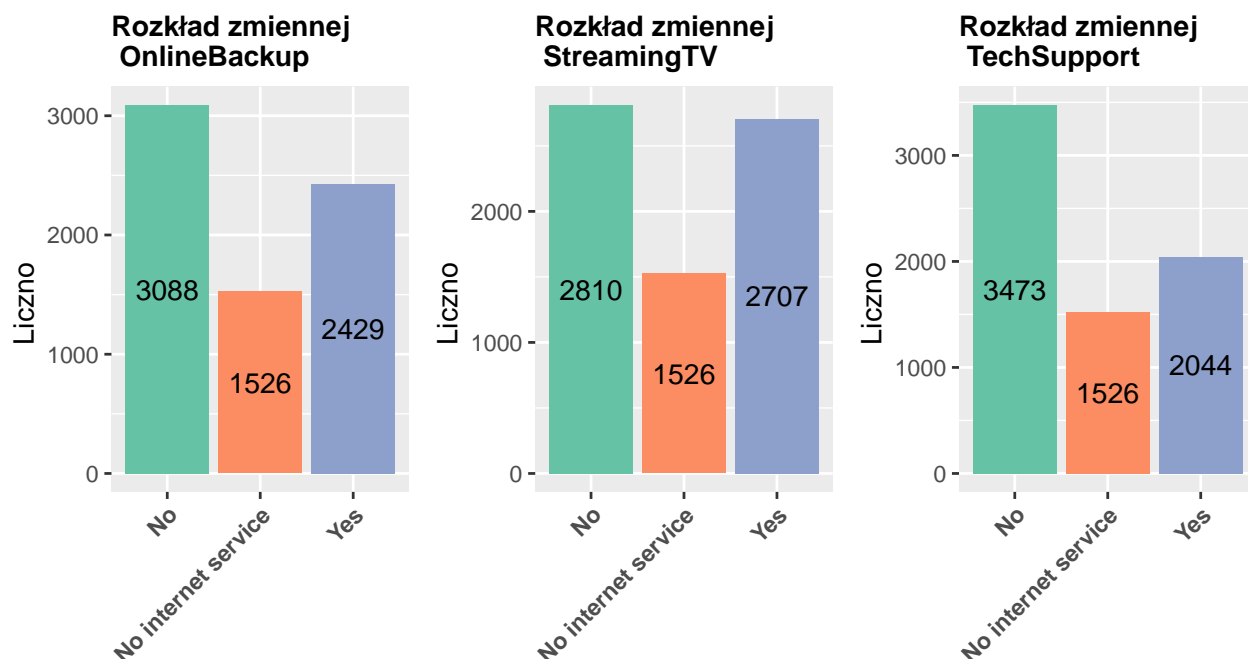
Wykres 3: Wykresy zmiennych Dependents, Partner oraz zmiennej Partner pogrupowanej względem zmiennej Dependents

Z wykresu 3 zmiennej Dependents możemy odczytać, że większość klientów nie ma żadnych osób na utrzymaniu. Możemy również zauważyć, w grupie, które mają kogoś na utrzymaniu, ponad połowa osób ma partnera.



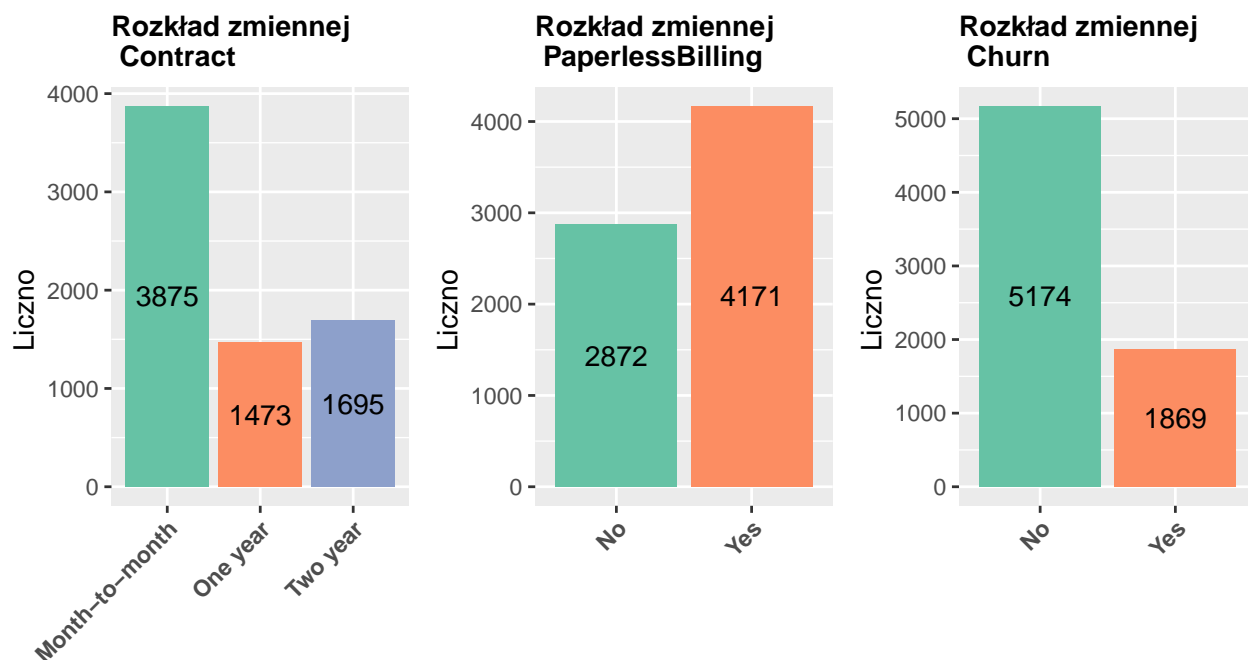
Wykres 4: Wykresy zmiennych SeniorCitizen, PhoneService oraz MultipleLines

Z wykresów 4 zauważyć można, że znaczna większość klientów jest przed wiekiem emerytalnym. Z usług telefonicznych korzysta prawie 90% klientów, lecz nieco ponad połowa z nich nie zdecydowała się na korzystanie z wielu linii.



Wykres 5: Wykresy zmiennych OnlineBackup, StreamingTV oraz TechSupport

Wśród wszystkich klientów, jedynie nieco ponad $\frac{1}{3}$ z nich zdecydowało się na usługę **OnlineBackup**. Z telewizji internetowej oraz wsparcia technicznego korzysta odpowiednio **2707** oraz **2044** osób, co prezentują wykresy 5. Z wykresów zmiennych **DeviceProtection**, **StreamingMovies** i **OnlineSecurity** możemy wyciągnąć odpowiednio podobne wnioski.

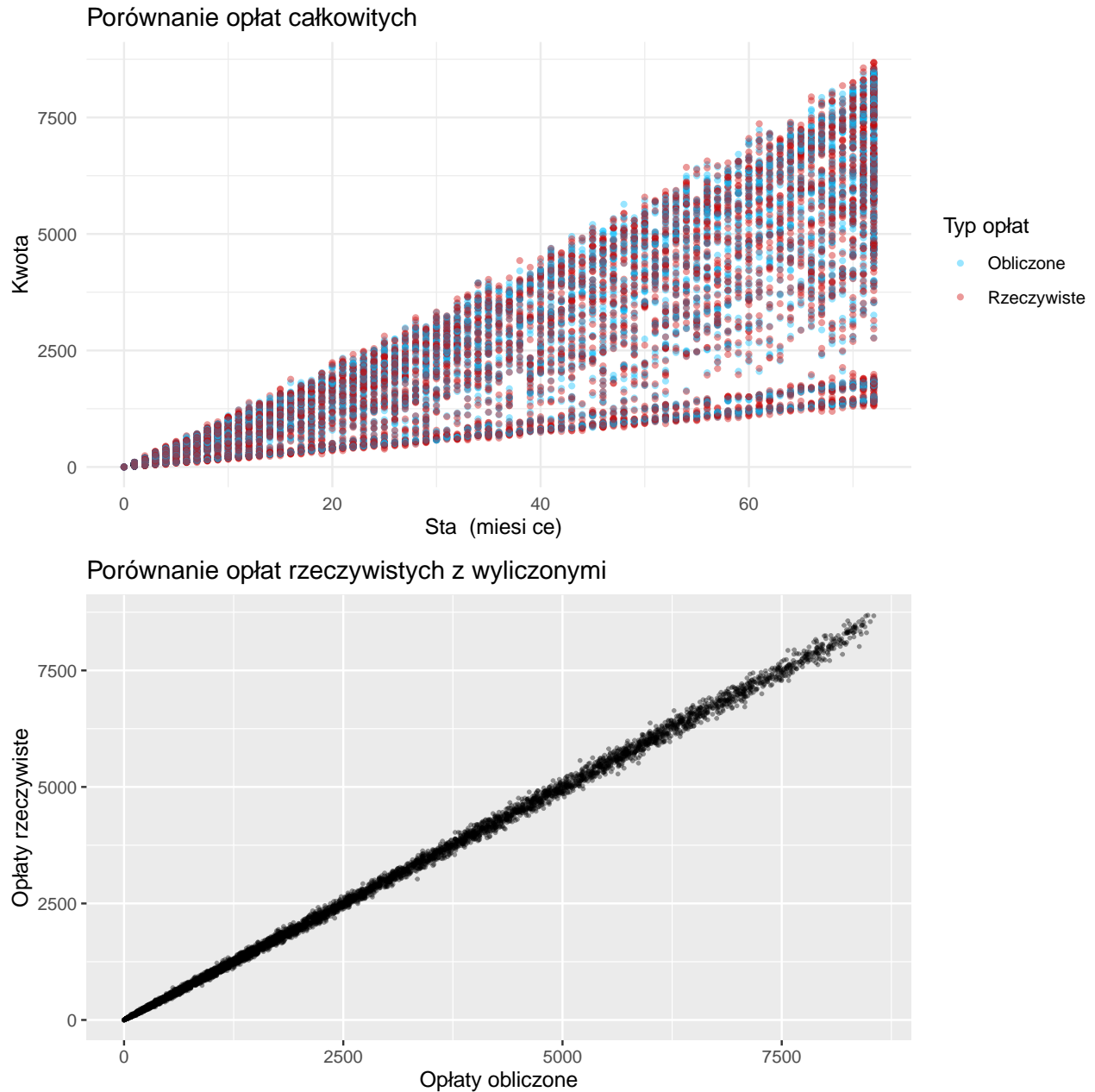


Wykres 6: Wykresy zmiennych Contract, PaperlessBilling oraz Churn

Z wykresów 6 możemy odczytać, że **3875** klientów ma umowę miesięczną z firmą. Przez klientów jest to

najczęściej wybierana umowa. Ponad połowa korzysta także z płatności elektronicznych. Na zrezygnowanie ze wszelkich usług firmy zdecydowało się **1869** klientów (jest to **26.54** % wszystkich klientów).

3.2.4 Wykresy rozrzutu



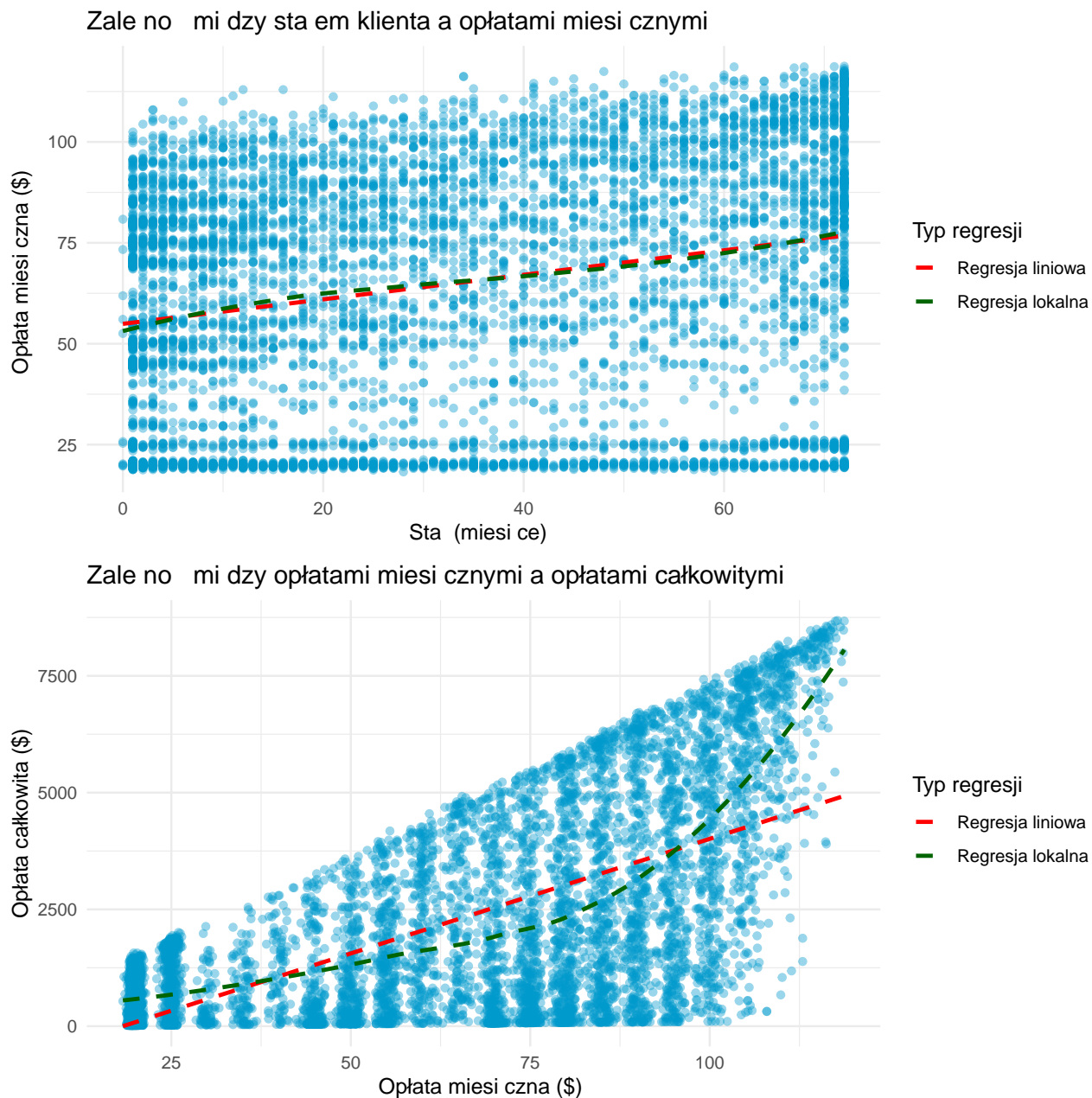
Wykres 7: Wykresy rozrzutu dla zmiennych TotalCharges, tenure oraz opłat całkowitych wyliczonych za pomocą MonthlyCharges i tenure

Na wykresie 7 możemy zaobserwować rozrzut zmiennej TotalCharges względem długości stażu klientów oraz rozrzut opłat całkowitych wyliczonych jako $\text{Opłaty obliczone} = \text{tenure} \times \text{MonthlyCharges}$. W większości przypadków kwoty te są bardzo do siebie zbliżone (współczynnik korelacji wynosi **0.9995606**), jednak nie są identyczne. Może to oznaczać, że opłaty są naliczane w inny sposób, lub też zostały naliczone niepoprawnie.

Powodem może być także zmiana oferty w trakcie trwania umowy (MonthlyCharges pokazuje tylko aktualną wartość).

Tabela 6: Macierz korelacji między zmiennymi ilościowymi

	tenure	MonthlyCharges	TotalCharges
tenure	1.000	0.247	0.826
MonthlyCharges	0.247	1.000	0.651
TotalCharges	0.826	0.651	1.000



Wykres 8: Wykresy rozrzutu dla par zmiennych ciągłych (tenure, MonthlyCharges, TotalCharges)

Występuje słaby związek między długością współpracy a wysokością miesięcznej opłaty (korelacja wynosi około **0.247**). Klienci z długim stażem płacą zarówno niskie, jak i wysokie rachunki. Oczekiwanym wynikiem są coraz większe opłaty całkowite przy coraz dłuższym korzystaniu z usług, co potwierdza odpowiedni wykres 7 oraz wysoki współczynnik korelacji (**0.826**). Oznacza to dość silną zależność liniową między tymi zmiennymi.

Osoby z wyższą miesięczną opłatą szybciej osiągają wyższą wartość **TotalCharges**, ale zależność nie jest liniowa – wpływa na nią zmienna **tenure**. Można zauważyć na wykresie 8, że w okolicy opłat miesięcznych wynoszących około 75 \$ model regresji zaczyna szybciej wzrastać. Może to oznaczać, że klienci z wyższymi opłatami miesięcznymi mają dłuższy staż (gromadzą więcej opłat w czasie) lub korzystają z dodatkowych usług, które znacząco podnoszą całkowite koszty.

4 Analiza opisowa z podziałem na grupy

4.1 Wskaźniki sumaryczne z podziałem na grupy

Tabela 7: Porównanie wskaźników sumarycznych zmiennej **tenure** dla grup klientów lojalnych i odchodzących

Grupa	min	Q1	median	mean	Q3	max	var	sd	IQR
Lojalni	0	15	38	37.57	61	72	581.47	24.11	46
Nielojalni	1	2	10	17.98	29	72	381.46	19.53	27

Cecha **tenure** charakteryzuje się dużą zmiennością dla nielojalnych klientów (CV = **109%**). Natomiast dla lojalnych wynosi CV = **64%**.

$$CV = \frac{\sigma}{\mu} \times 100\%$$

gdzie:

- σ – odchylenie standardowe,
- μ – średnia arytmetyczna.

Tabela 8: Porównanie wskaźników sumarycznych zmiennej **MonthlyCharges** dla grup klientów lojalnych i odchodzących

Grupa	min	Q1	median	mean	Q3	max	var	sd	IQR
Lojalni	18.25	25.10	64.43	61.27	88.4	118.75	966.75	31.09	63.30
Nielojalni	18.85	56.15	79.65	74.44	94.2	118.35	608.41	24.67	38.05

Cecha **MonthlyCharges** charakteryzuje się umiarkowaną zmiennością dla obu grup klientów: nielojalni CV = **33%**, lojalni CV = **51%**.

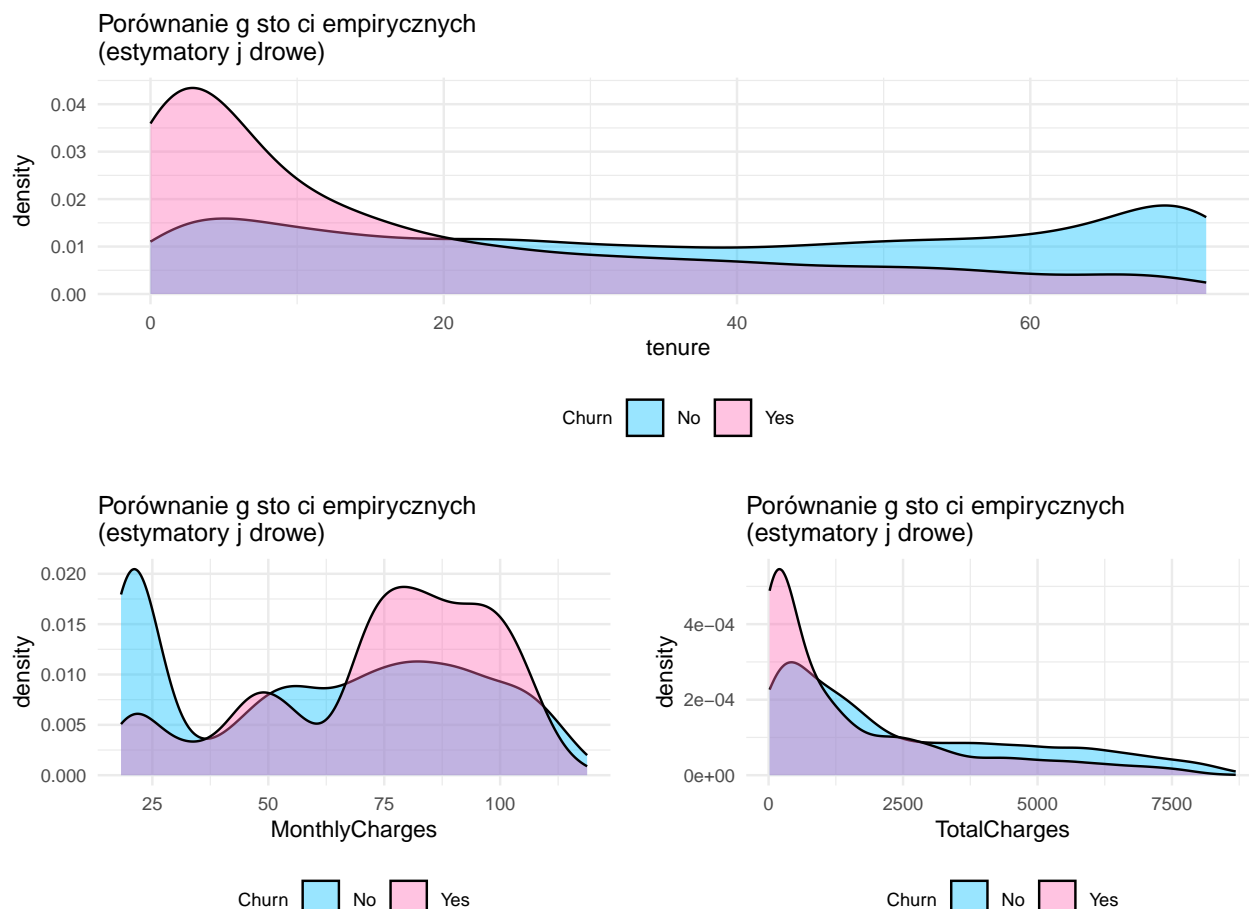
Tabela 9: Porównanie wskaźników sumarycznych zmiennej **TotalCharges** dla grup klientów lojalnych i odchodzących

Grupa	min	Q1	median	mean	Q3	max	var	sd	IQR
Lojalni	18.80	577.82	1683.60	2555.34	4264.12	8672.45	5426370	2329.46	3686.3
Nielojalni	18.85	134.50	703.55	1531.80	2331.30	8684.80	3575212	1890.82	2196.8

Cecha **TotalCharges** charakteryzuje się dużą zmiennością dla obu grup klientów, lecz większą dla nielojalnych CV = **123%**, lojalni CV = **91%**.

4.2 Wykresy z podziałem na grupy

4.2.1 Zmienne ilościowe



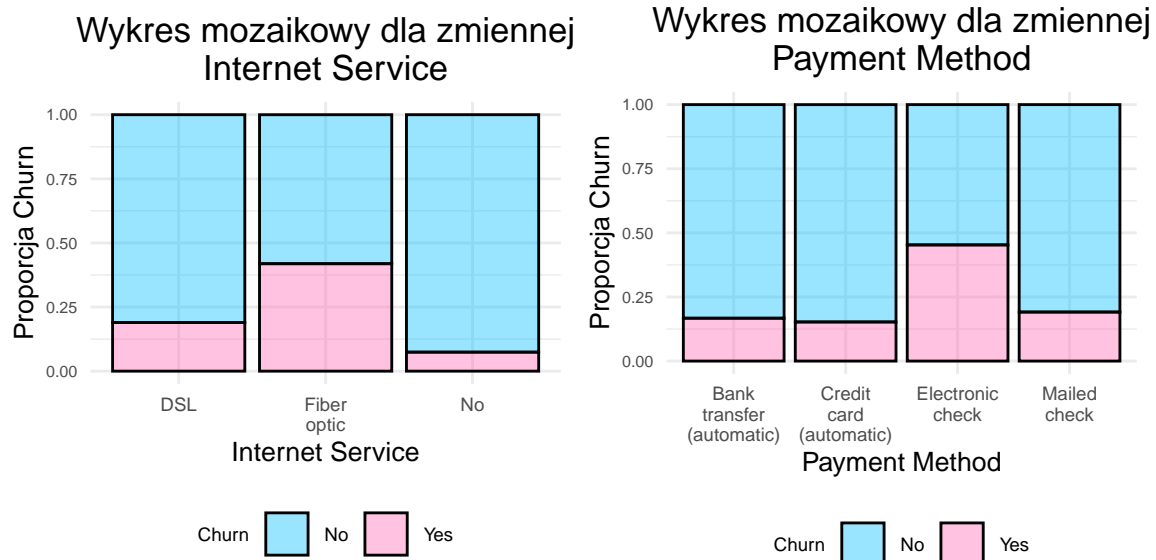
Wykres 9: Porównanie rozkładu zmiennych tenure, MonthlyCharges, TotalCharges dla dwóch grup: lojalnych i odchodzących klientów

Wykres 9 pozwala odczytać rozkład zmiennej `tenure` z podziałem na klientów lojalnych i tych, którzy odeszli. `Tenure` oznacza czas stażu klienta w zakresie od **0 do 72 miesięcy**. Możemy zauważyć, że osoby, które odeszły były naszymi klientami głównie bardzo krótko. Współczynnik skośności `tenure` bez podziału na `Churn` wynosi **0.24**, lecz z podziałem jest to **1.15** dla klientów, którzy odeszli i **-0.03** dla lojalnych.

Następnie możemy odczytać z wykresu rozkładu zmiennej `MonthlyCharges`, czyli opłat miesięcznych, że osoby, które zrezygnowały ponosiły **bardzo wysokie opłaty miesięczne**. Zakres opłat waha się od **18.25** do **118.75**. Współczynnik skośności `MonthlyCharges` bez podziału na `Churn` wynosi **-0.22**, lecz z podziałem jest to **-0.73** dla klientów, którzy odeszli i **-0.03** dla lojalnych.

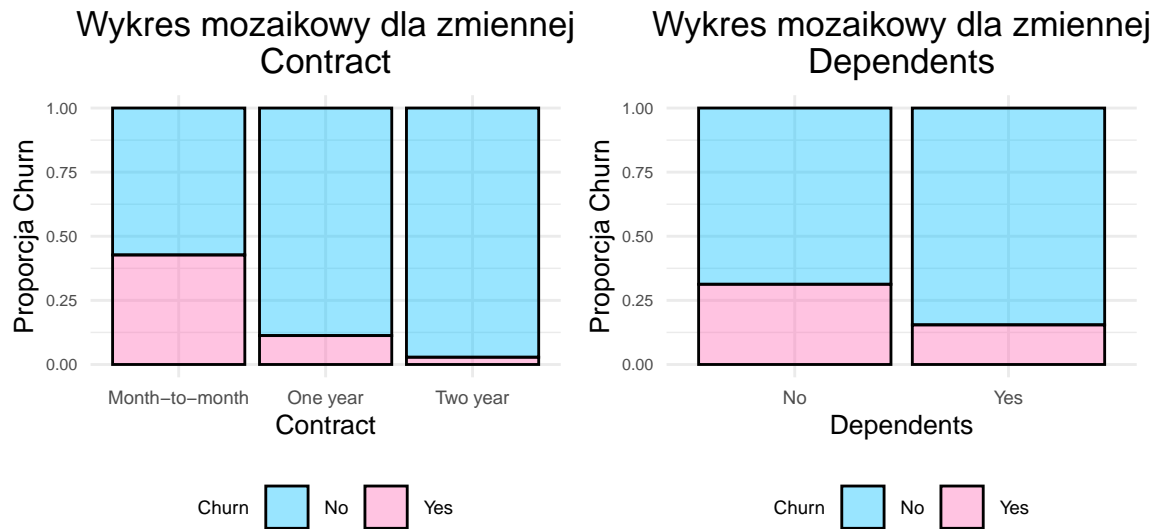
Ostatni wykres `TotalCharges` przedstawia rozkład sumy opłat dla danego klienta. Zakres zmiennych jest z przedziału od **18.80** do **8684.80**. Widzimy, że osoby, które odeszły miały **bardzo niską sumę opłat**. Możemy się domyślać, że wynika to z faktu, że osoby odchodzące były naszymi klientami krótko, więc przez krótki czas zapłaciły nam niewiele.

4.2.2 Zmienne jakościowe



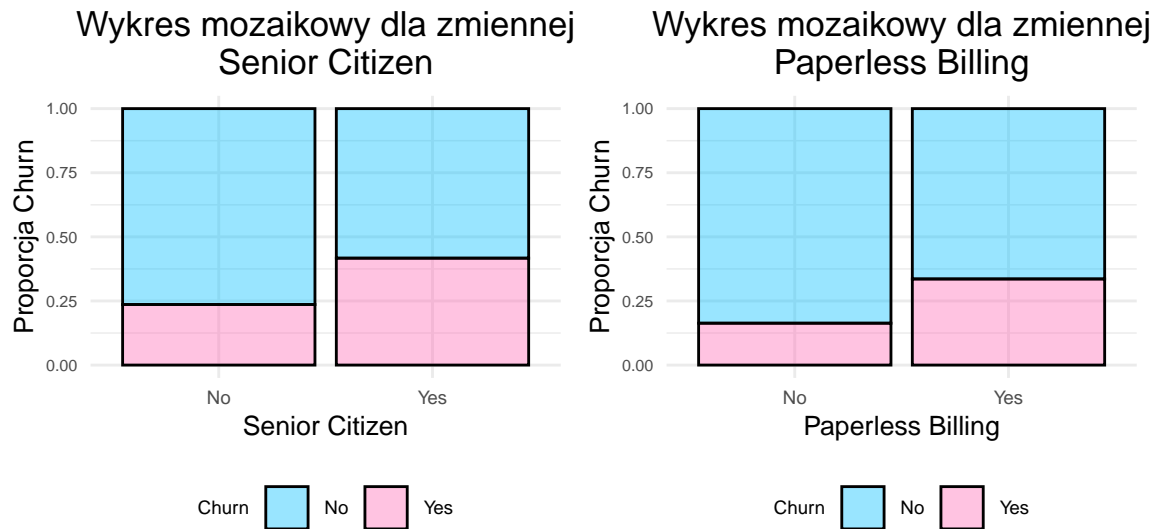
Wykres 10: Wpływ typu usługi internetowej i metody płatności na wskaźnik rezygnacji (Churn)

Na wykresie 10 spośród użytkowników różnych technologii dostarczających internet najczęściej odeszło użytkowników światłowodu. Ponadto patrząc na wykres dla zmiennej Payment Method, czyli rodzaju płatności, widzimy, że najczęściej odchodzi użytkowników płatności elektronicznej.



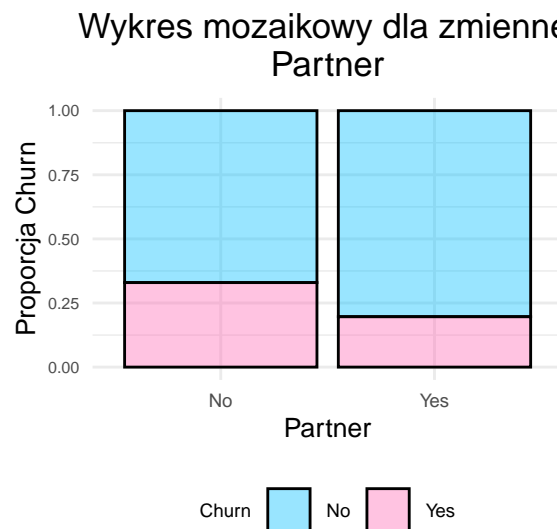
Wykres 11: Wpływ typu umowy i osób na utrzymaniu na wskaźnik rezygnacji (Churn)

Najwięcej odeszło osób, które płaciły za usługi firmy raz na miesiąc co widać na wykresie 11. Ponadto odczytujemy, że trochę częściej odchodzą klienci, którzy nie posiadają osób na utrzymaniu.



Wykres 12: Wpływ bycia seniorem i płatności bezgotówkowej na wskaźnik rezygnacji (Churn)

Z wykresu 12 zauważamy, że z naszych usług częściej rezygnują seniorzy. Na rysunku obok widać, że trochę częściej rezygnują klienci płacący bezgotówkowo.

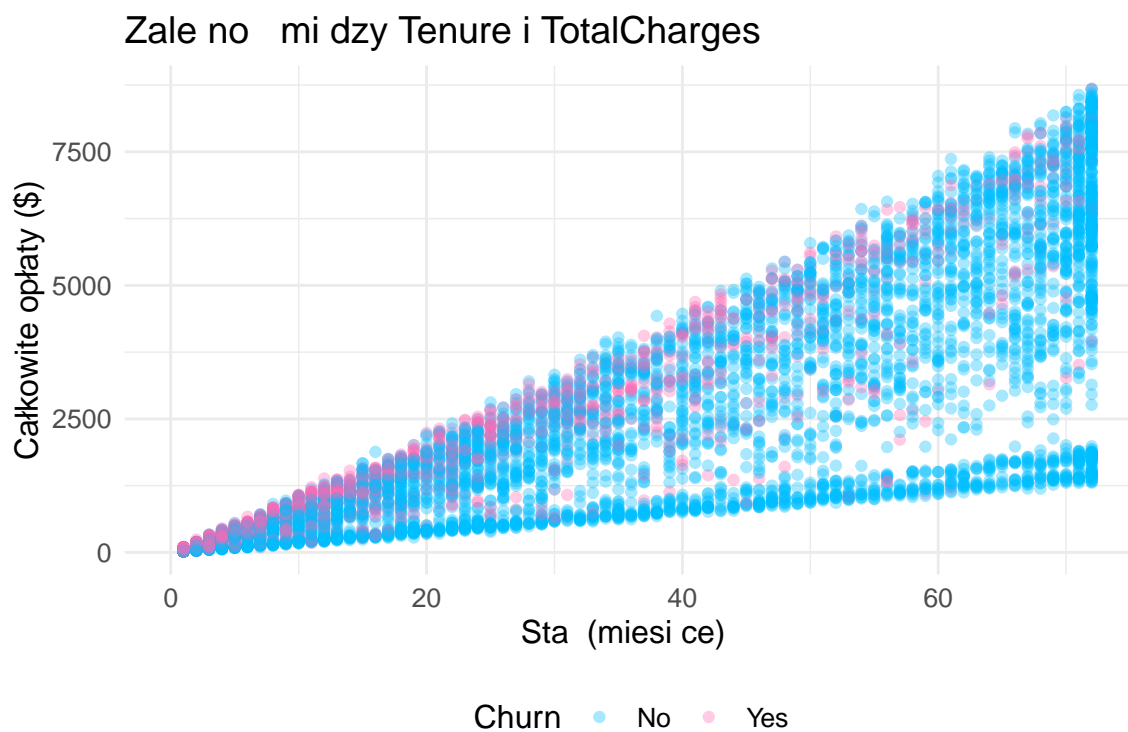


Wykres 13: Wpływ bycia w związku na wskaźnik rezygnacji (Churn)

Rysunek 13 przedstawia zależność churnu od posiadania partnera. Widać, że częściej odchodzą osoby bez partnera.

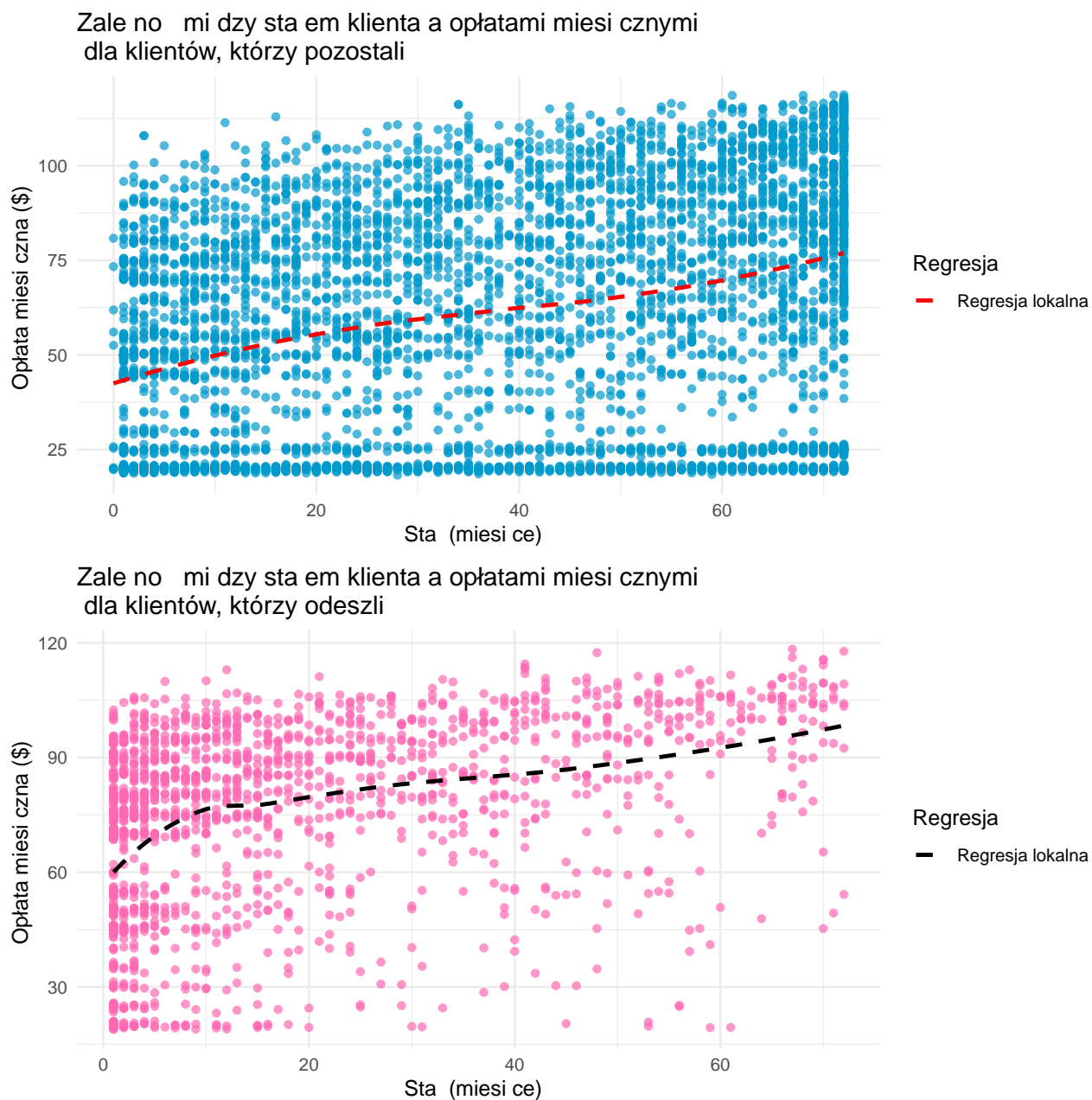
Wykresy proporcji churnu dla zmiennych `gender`, `Phone service` i `MultipleLines` są bardzo podobne dla wszystkich wartości i nie pozwalają zauważyć zależności, więc nie są zamieszczone w zestawieniu. Dla zmiennych `OnlineSecurity`, `OnlineBackup`, `DeviceProtection`, `TechSupport` wykresy proporcji churnu są bardzo zbliżone do siebie i można z nich odczytać, że najczęściej odchodzą osoby, które posiadają internet service, ale nie posiadają wyżej wymienionych udogodnień.

4.3 Zależności liniowe z podziałem na grupy



Wykres 14: Zależność łącznej sumy opłat od czasu subskrypcji z podziałem na klientów lojalnych i nielojalnych.

Możemy zauważyć, że wraz z czasem subskrypcji rośnie suma pieniędzy jaką zapłacił nam klient, ale co ważne widać, że lojalni klienci płacą zazwyczaj mniej, a klienci, którzy zrezygnowali płacili nam więcej.



Wykres 15: Wykres rozrzutu zmiennej MonthlyCharges od zmiennej z podziałem na Churn

Klienci z krótkim stażem (0-10 miesięcy), którzy odeszli, mają zdecydowanie wyższe opłaty miesięczne niż klienci lojalni (dane są bardziej rozproszone). Wykres 15 sugeruje, że klienci, którzy zdecydowali się zrezygnować z usług, bardzo szybko zdecydowali się na droższe pakiety.

4.4 Które ze zmiennych wykazują największe zróżnicowanie?

Po analizie z podziałem na grupy można stwierdzić, że na najlepsze rozróżnienie klientów z obu grup pozwala zmienna **tenure** i **MonthlyCharges**. Możemy to zauważyć na wykresie 9 i zamieszczonych poniżej współczynnikach skośności. Różnica w skośności dla **tenure** (**1.15** vs **-0.03**) wskazuje na fundamentalnie różne wzorce zachowań. Nowi klienci są znacznie bardziej narażeni na rezygnację. Natomiast dla klientów lojalnych czas współpracy nie jest istotnym predyktorem ryzyka rezygnacji. W przypadku **MonthlyCharges** grupa

rezygnująca koncentruje się na górnym przedziale cenowym, podczas gdy lojalni klienci wykazują równomierny rozkład we wszystkich przedziałach. Sugeruje to, że wysokie ceny są istotnym czynnikiem rezygnacji.

Wykresy zmiennych jakościowych (10 i 11) takich jak **Internet Service**, **Payment Method** i **Contract** pozwalają zauważyć, że proporcjonalnie z naszych usług rezygnuje dużo klientów korzystający ze światłowodu (blisko **50%**) i bardzo mało klientów, którzy nie korzystają z naszego internetu wcale. Rezygnuje również prawie **50%** użytkowników płatności elektronicznej i prawie **50%** użytkowników mających umowę month-to-month. Natomiast użytkownicy innych metod płatności i innych umów rezygnują rzadziej.

5 Podsumowanie – wnioski z przeprowadzonej analizy

5.1 Krótkie podsumowanie

Analiza danych pozwoliła na identyfikację kluczowych cech klientów oraz czynników wpływających na ich decyzje dotyczące korzystania z usług firmy. Stwierdzono, że duża część klientów wybiera podstawowe pakiety usług i unika dodatkowych opcji, takich jak wsparcie techniczne czy ochrona urządzeń. Ponadto zauważono zależność między typem umowy a skłonnością do rezygnacji – klienci z umową miesięczną znacznie częściej odchodzą. Analiza wskazała również na istotne różnice w preferencjach płatności oraz związek między stażem klienta a skłonnością do poszerzania zakresu usług. Wyniki te dostarczają cennych informacji na temat grup klientów najbardziej narażonych na odejście oraz potencjalnych strategii mających na celu zwiększenie ich lojalności.

5.2 Co wiemy o klientach firmy?

- Klienci to głównie osoby:
 - nie będące seniorami,
 - bez osób na utrzymaniu,
 - korzystające z pojedynczej linii usług telefonicznych,
 - korzystające ze światłowodu,
 - płacące bezgotówkowo,
 - nie korzystające z usług dodatkowych takich jak: wsparcie techniczne lub ochrona urządzeń.
- Duża część klientów ma wykupiony u nas najtańszy i najbardziej podstawowy pakiet usług co można wywnioskować po bardzo licznej grupie osób mających niskie opłaty miesięczne i po tym, że wiele osób nie korzysta z dodatkowych usług. Jest to również grupa klientów mająca najdłuższy staż.
- Klienci z umową miesięczną są bardziej skłonni do rezygnacji.
- Seniorzy oraz klienci bez partnera preferują niższe opłaty miesięczne.
- Istnieje korelacja między stażem i opłatami miesięcznymi, więc być może wraz z upływem czasu klienci przekonują się do wykupienia dodatkowych usług.
- Zazwyczaj rezygnują klienci, których opłaty miesięczne są wyższe.
- Dużo klientów, którzy płacą miesięcznie więcej, bardzo szybko rezygnują, więc ich opłaty całkowite nie są wysokie.
- Klienci korzystający z płatności elektronicznej odchodzą częściej.

5.3 Analiza etapu trzeciego

5.3.1 Przyczyny odchodzenia

- Często odchodzą osoby, będące klientami krótko i mające wysokie opłaty miesięczne. Sugeruje to, że klienci wykupują bardzo dużą liczbę pakietów i po przetestowaniu oferty, rezygnują. Być może nasze dodatkowe opcje nie funkcjonują na dostatecznym poziomie lub nie są warte swojej ceny.

- Bardzo często rezygnują osoby korzystające ze światłowodu, być może nie zapewnia on wystarczającej prędkości internetu, stabilności połączenia lub jest zbyt drogi.
- W ostatnim miesiącu zrezygnowało blisko 50% osób płacących elektronicznie. Ta metoda wymaga ręcznego potwierdzania płatności, co może prowadzić do opóźnień lub pomyłek. Ponadto brak automatycznego obciążenia konta oznacza, że klient ma większą kontrolę nad decyzją o płatności, a tym samym łatwiej mu przerwać subskrypcję.
- Umowa miesięczna nie wiąże klienta na dłuższy czas, więc klienci wybierający umowę miesięczną mogą być bardziej skłonni do zmiany dostawcy usług, jeśli znajdą lepszą ofertę.

5.3.2 Propozycje przeciwdziałania

- Umowa miesięczna pozwala klientom na zbyt łatwą rezygnację, przez co ta grupa klientów najczęściej odchodzi. Być może firma powinna zastąpić ją poprzez umowę 6-miesięczną.
- Duża rotacja wśród klientów o krótkim stażu. Firma powinna stworzyć oferty, dzięki którym klienci zdecydują się na dłuższe korzystanie z usług firmy. Klienci, których staż jest mniejszy niż 5 miesięcy, są najbardziej skłonni do zrezygnowania z propozycji firmy.
- Lepsze oferty, które zachęcą do korzystania z OnlineSecurity, OnlineBackup, DeviceProtection, Tech-Support, ponieważ osoby korzystające z tych usług rzadziej rezygnują. Średnio tylko połowa klientów korzysta z tych opcji.
- Rezygnowanie z możliwości zapłaty poprzez czek elektroniczny lub zachęta do wybierania automatycznych metod płatności.
- Lepsze oferty dla seniorów i tych klientów, którzy nie mają partnera ponieważ te grupy osób odchodzą częściej.
- Firma powinna się skupić na promowaniu ofert wielu linii telefonicznych.
- Oferta dotycząca światłowodu powinna zostać także zmieniona (duża grupa klientów, która z nich korzysta, decyduje się na odejście).
- Bardziej korzystne umowy dla nowo przybyłych klientów. Firma powinna rozważyć dodanie okresu próbnego lub ulg dla nowych użytkowników.
- Problemem wydają się być zbyt wysokie opłaty miesięczne. Dobrym pomysłem mogło być zniżki na zestawy usług oferowanych przez firmę. Im większy pakiet, tym większa zniżka.
- W grupie klientów, którzy odeszli znajdują się również tacy, których staż wynosi ponad 60 miesięcy. Być może powinny być rozpatrzone bonusy dla lojalnych klientów.