

# Lista 1

## Eksploracja danych

Olaf Masłowski nr ablumu 277543

2025-07-08

### Spis treści

1	Analiza opisowa - Telco Customer churn	2
2	Etap 1 - przygotowanie danych	2
3	Etap 2 - analiza opisowa	3
4	Etap 3 - analiza ze względu na Churn	6
5	Etap 4 - podsumowanie	8

# 1 Analiza opisowa - Telco Customer churn

Będziemy analizować zbiór danych WA\_Fn-UseC\_-Telco-Customer-Churn.csv zawierający informacje o klientach pewnej firmy telekomunikacyjnej. Wykorzystamy metody analizy opisowej do poznania podstawowych własności danych. Przede wszystkim zbadamy miary położenia i rozproszenia poszczególnych zmiennych, oraz przedstawimy dane na histogramach i wykresach pudełkowych (zmienne ilościowe), lub wykresach słupkowych (zmienne jakościowe),

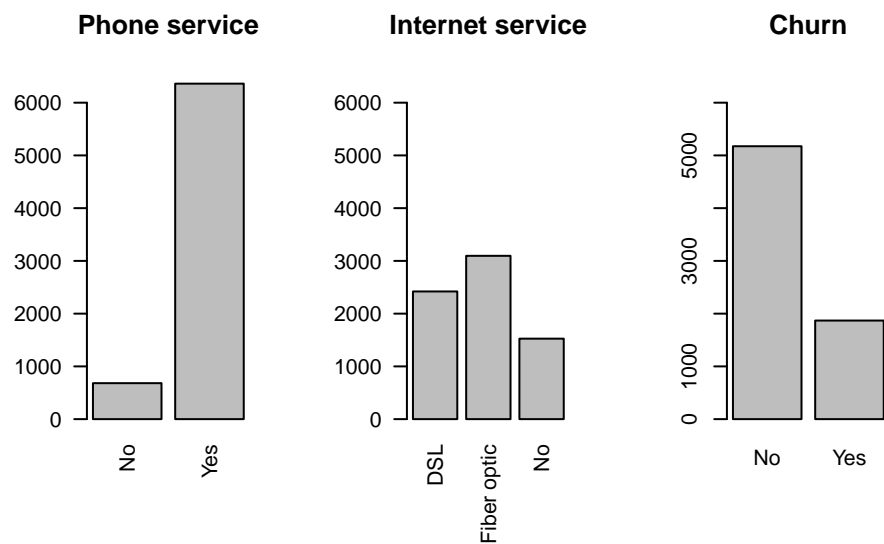
Dane zawierają informacje takie jak:

- płeć
- czy klient jest osobą starszą?
- czy klient ma partnera/partnerkę?
- czy klient utrzymuje inne osoby?
- od jak dawna dostarczana jest usługa (Tenure)?
- jakie usługi otrzymuje?
- metoda płatności
- rodzaj umowy
- wartość miesięcznych transakcji
- Łączna wartość transakcji
- czy klient oszedł w ostatnim miesiącu (Churn)?

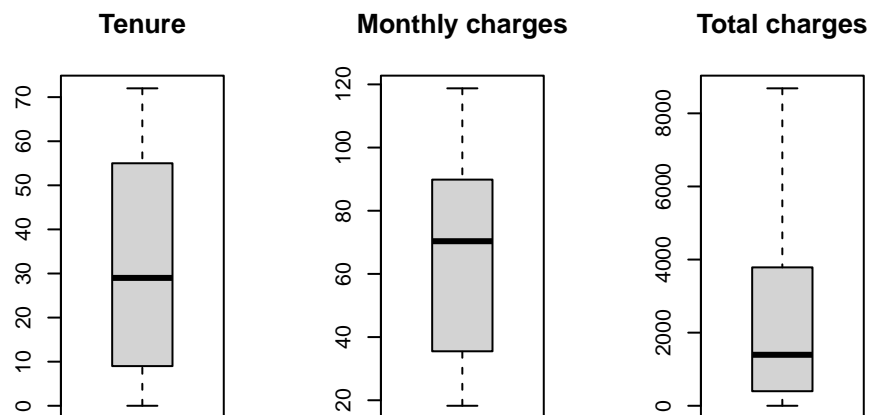
## 2 Etap 1 - przygotowanie danych

Dane zawierają 7043 przypadków i 21 cech. Braki danych występują dla cechy **Charges** i wynikają z faktu, że dany klient jest klientem zbyt krótko i jeszcze się nie rozliczał - wartość Tenure wynosi 0. Oprócz tego za braki danych można uznać np “no internet service” w przypadku cech odnoszących się do dodatkowych usług internetowych.

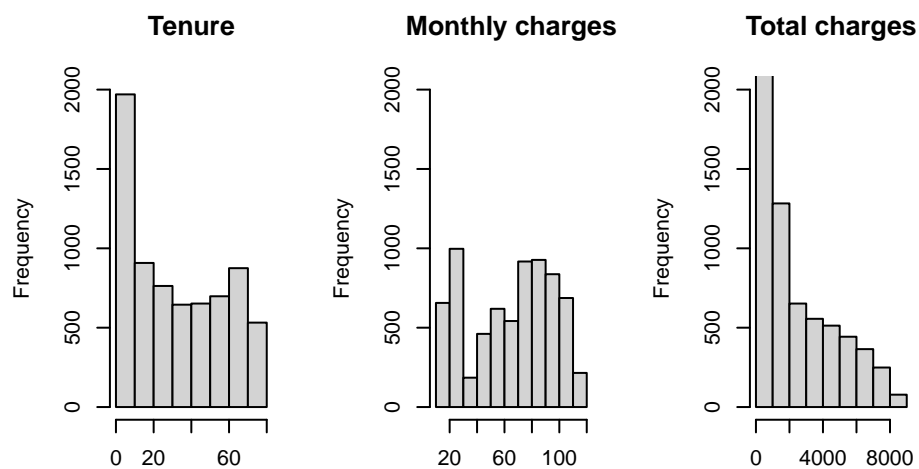
### 3 Etap 2 - analiza opisowa



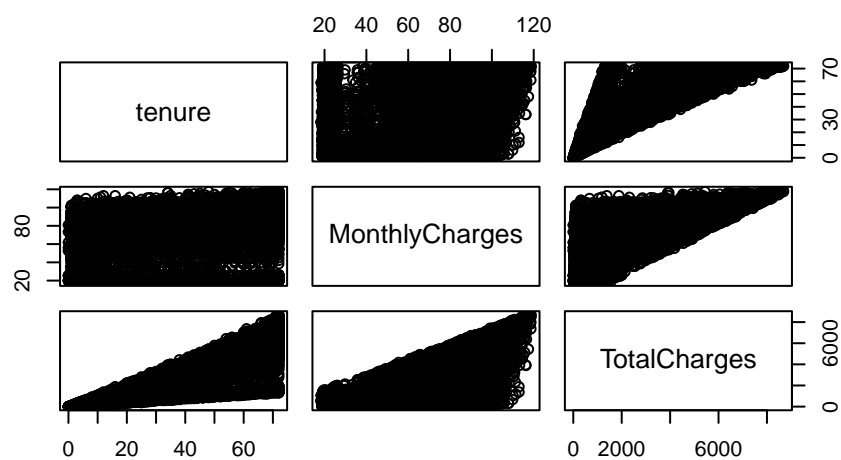
Rysunek 1: Wykresy słupkowe interesujących zmiennych



Rysunek 2: Wykresy pudełkowe zmiennych ciągłych



Rysunek 3: Histogramy zmiennych ciągłych



Rysunek 4: Wykresy rozrzutu

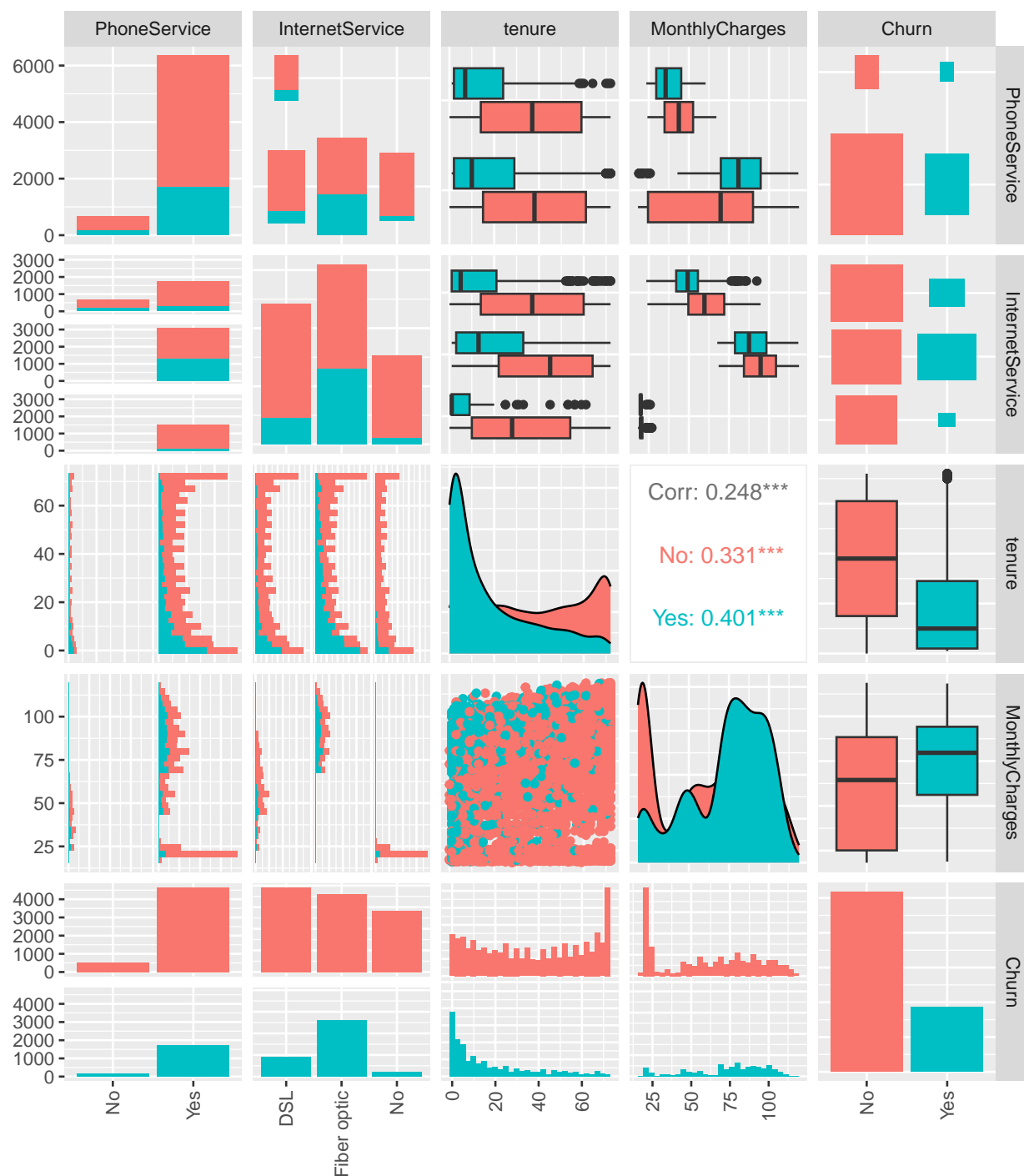
	tenure	MonthlyCharges	TotalCharges
min	0.000	18.250	0.000
Q1	9.000	35.500	398.550
median	29.000	70.350	1394.550
mean	32.371	64.762	2279.734
Q3	55.000	89.850	3786.600
max	72.000	118.750	8684.800
var	603.168	905.411	5138357.168
sd	24.559	30.090	2266.794
IQR	46.000	54.350	3388.050

Tabela 1: Miary położenia rozproszenia

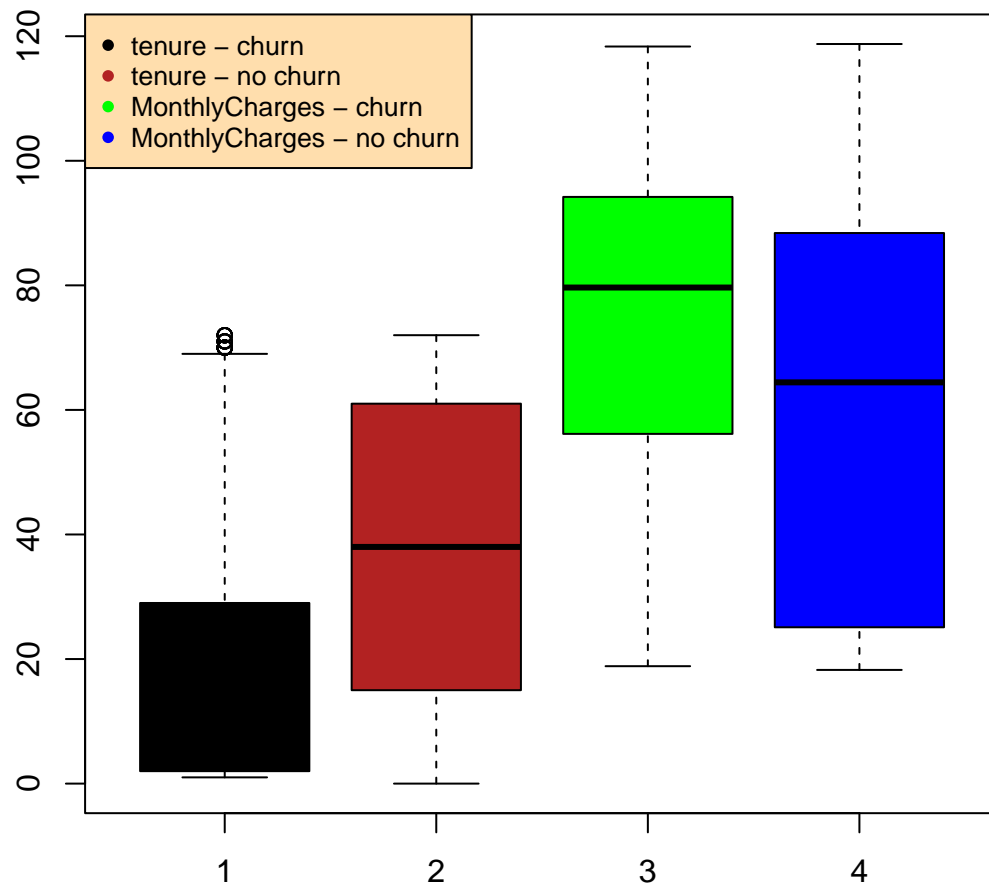
Cechy tenure, MonthlyCharges i TotalCharges mieszczą się w przedziałach od 0, do odpowiednio 72, 118.75 i 8684.8. Zmienne ciągłe nie rozkładają się symetrycznie, najbliżej symetrii jest MonthlyCharges. TotalCharges charakteryzuje się dużą zmiennością z największymi obserwacjami wielokrotnie większymi, niż mediana.

Można zaobserwować, że popularniejsze wśród klientów są usługi związane z dostarczaniem telefonii, niż internetu. Istotna część klientów zrezygnowała w ciągu ostatniego miesiąca. Łączne zyski rosną zależnie od czasu dostarczania usługi i miesięcznych płatności, co nie powinno być zaskoczeniem. “Stali” klienci mogą mieć nieznacznie wyższe płatności niż nowi.

## 4 Etap 3 - analiza ze względu na Churn



Rysunek 5: zestawienie wykresów dla istotnych zmiennych



Rysunek 6: wykresy pudełkowe dla zmiennych tenure i monthly charges ze względu na churn

	tenure	MonthlyCharges	TotalCharges
min	1.000	18.850	18.850
Q1	2.000	56.150	134.500
median	10.000	79.650	703.550
mean	17.979	74.441	1531.796
Q3	29.000	94.200	2331.300
max	72.000	118.350	8684.800
var	381.465	608.414	3575211.596
sd	19.531	24.666	1890.823
IQR	27.000	38.050	2196.800

Tabela 2: Miary położenia rozproszenia - churn

	tenure	MonthlyCharges	TotalCharges
min	1.000	18.850	18.850
Q1	2.000	56.150	134.500
median	10.000	79.650	703.550
mean	17.979	74.441	1531.796
Q3	29.000	94.200	2331.300
max	72.000	118.350	8684.800
var	381.465	608.414	3575211.596
sd	19.531	24.666	1890.823
IQR	27.000	38.050	2196.800

Tabela 3: Miary położenia rozproszenia - no churn

## 5 Etap 4 - podsumowanie

Firma oferuje szeroki wachlarz usług w różnych cenach.

Z wcześniejszych wykresów i analiz możemy zaobserwować, że znacząca część klientów, to klienci wieloletni. Większą część usług stanowią usługi telefoniczne

Z wykresów możemy odczytać, że odchodzili (Churn) przede wszystkim osoby nie będące klientami od dawna i mające wysokie płatności. Ponadto szczególnie dużo z rezygnujących klientów korzystało ze światłowodu. Zatem firma powinna skupić się przede wszystkim na jakości high-endowych usług i utrzymaniu nowych klientów.