

zenius

Kampus  
Merdeka  
INDONESIA JAYA

# Final Project Presentation

Nomor Kelompok: 2

Nama Mentor: Rachmadio Noval

Nama:

- Damar Amarsa Ajiputra
- Dea Salsa Nisrina Hanuradi

Machine Learning Class

---

Program Studi Independen Bersertifikat  
Zenius Bersama Kampus Merdeka



- **Latar Belakang**
- **Explorasi Data dan Visualisasi**
- **Modelling**
- **Kesimpulan**

# Latar Belakang

# Latar Belakang Project

Sumber Data:

<https://www.kaggle.com/datasets/barun2104/telecom-churn?datasetId=567482>

Problem: **classification**

Tujuan:

- Mengurangi Churn dari para pelanggan dan mengetahui faktor apa saja yang dapat mempengaruhi Churn



# Explorasi Data dan Visualisasi

# Business Understanding

- Dataset yang digunakan adalah dataset mengenai *customer churn* pada suatu perusahaan telekomunikasi.
- **Churn** adalah kondisi dalam dunia industri yang menunjukkan adanya customer berhenti menggunakan layanan suatu perusahaan ataupun customer yang berpindah layanan ke perusahaan lain.
- Dengan adanya churn, perusahaan tentunya akan rugi sangat besar karena kehilangan pelanggan. Dengan begitu, *Growth Rate* dari perusahaan pun juga akan terhambat.

# Understanding Dataset

Dataset berisi 11 Kolom

Nama Kolom	Deskripsi	Tipe Data	Isi Data
Churn	Apakah Customer Churn atau tidak	Kategorikal	1=Customer Churn 0=Tidak
AccountWeeks	Jumlah mingguan customer memiliki akun aktif	Numerikal	
ContractRenewal	Apakah Customer baru saja memperbarui kontrak	Kategorikal	1=Ya 0=Tidak
DataPlan	Apakah Customer memiliki Paket Data	Kategorikal	1=Ya 0=Tidak

# Understanding Dataset

Nama Kolom	Deskripsi	Tipe Data	Isi Data
DataUsage	Penggunaan data (dalam Gigabyte) dalam 1 bulan	Numerikal	
CustServCalls	Jumlah panggilan ke customer service	Numerikal	
DayMins	Rata-rata penggunaan menit harian dalam 1 bulan	Numerikal	
DayCalls	Rata-rata jumlah panggilan harian	Numerikal	



# Understanding Dataset

Nama Kolom	Deskripsi	Tipe Data	Isi Data
MonthlyCharge	Rata-rata tagihan pembayaran bulanan	Numerikal	
OverageFee	Biaya berlebih terbesar dalam 12 bulan terakhir	Numerikal	
RoamMins	Rata-rata jumlah menit roaming	Numerikal	

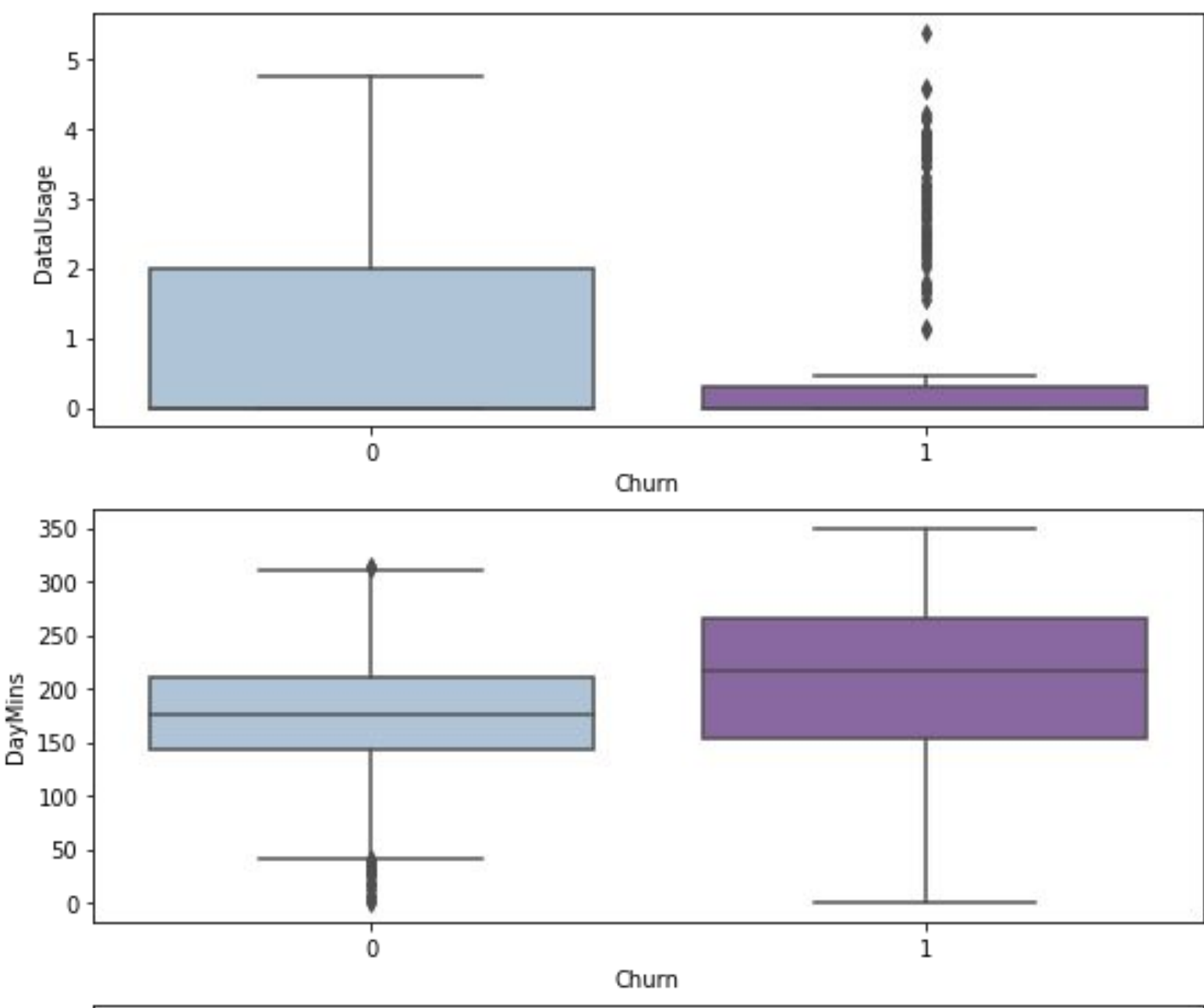
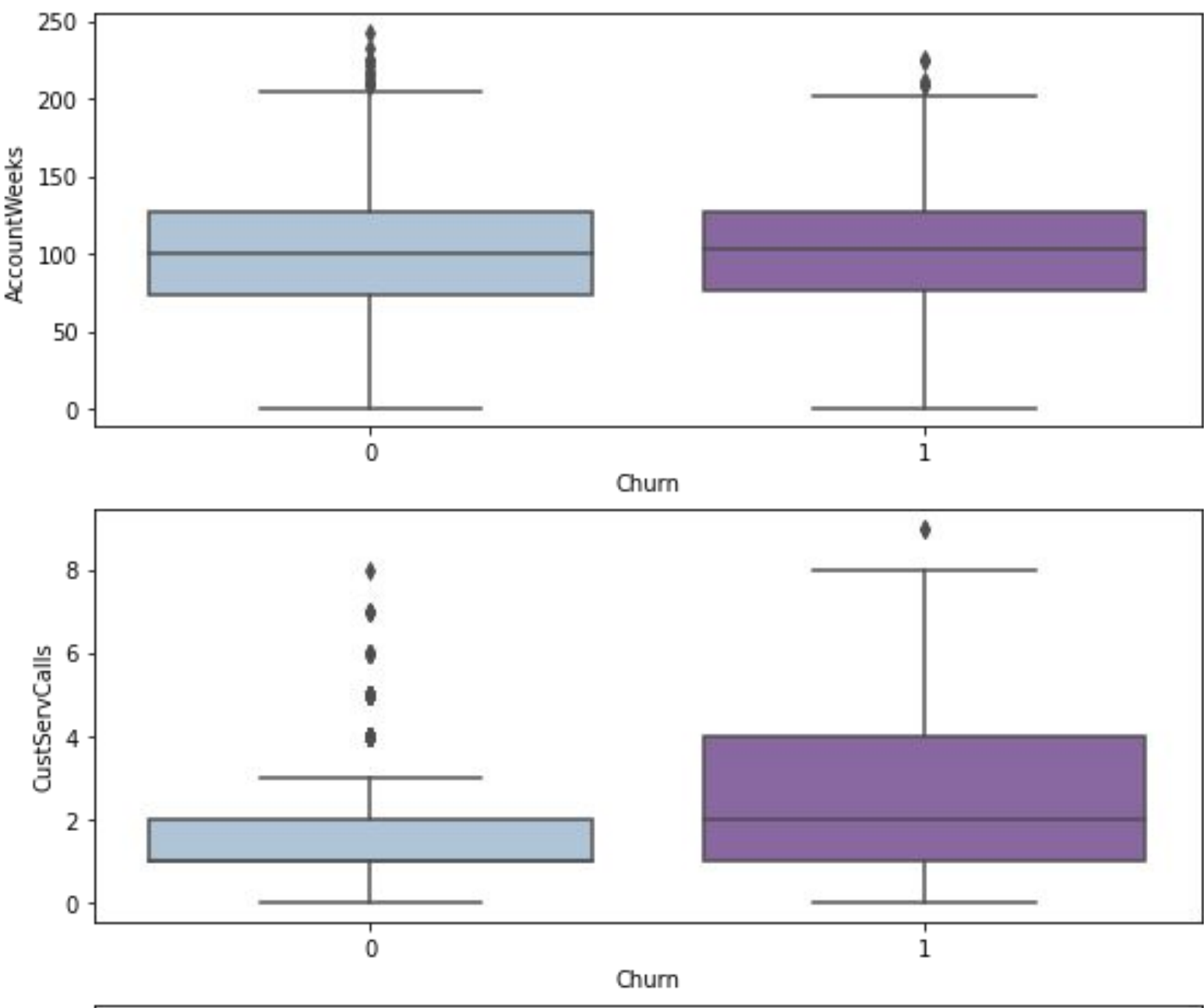
# Data Cleansing

```
RangeIndex: 3333 entries, 0 to 3332
Data columns (total 11 columns):
#   Column              Non-Null Count  Dtype
---  -
0   Churn                3333 non-null   int64
1   AccountWeeks         3333 non-null   int64
2   ContractRenewal      3333 non-null   int64
3   DataPlan              3333 non-null   int64
4   DataUsage             3333 non-null   float64
5   CustServCalls         3333 non-null   int64
6   DayMins               3333 non-null   float64
7   DayCalls              3333 non-null   int64
8   MonthlyCharge         3333 non-null   float64
9   OverageFee            3333 non-null   float64
10  RoamMins              3333 non-null   float64
```

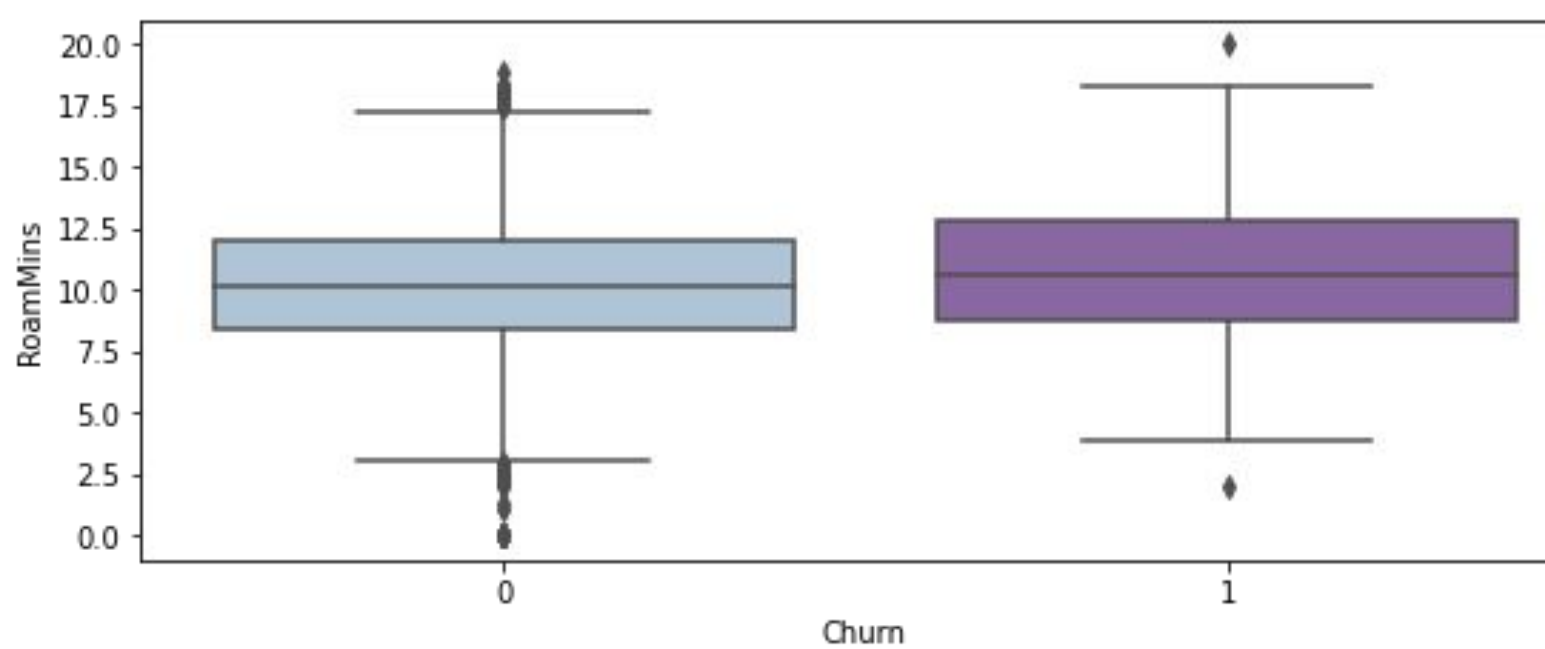
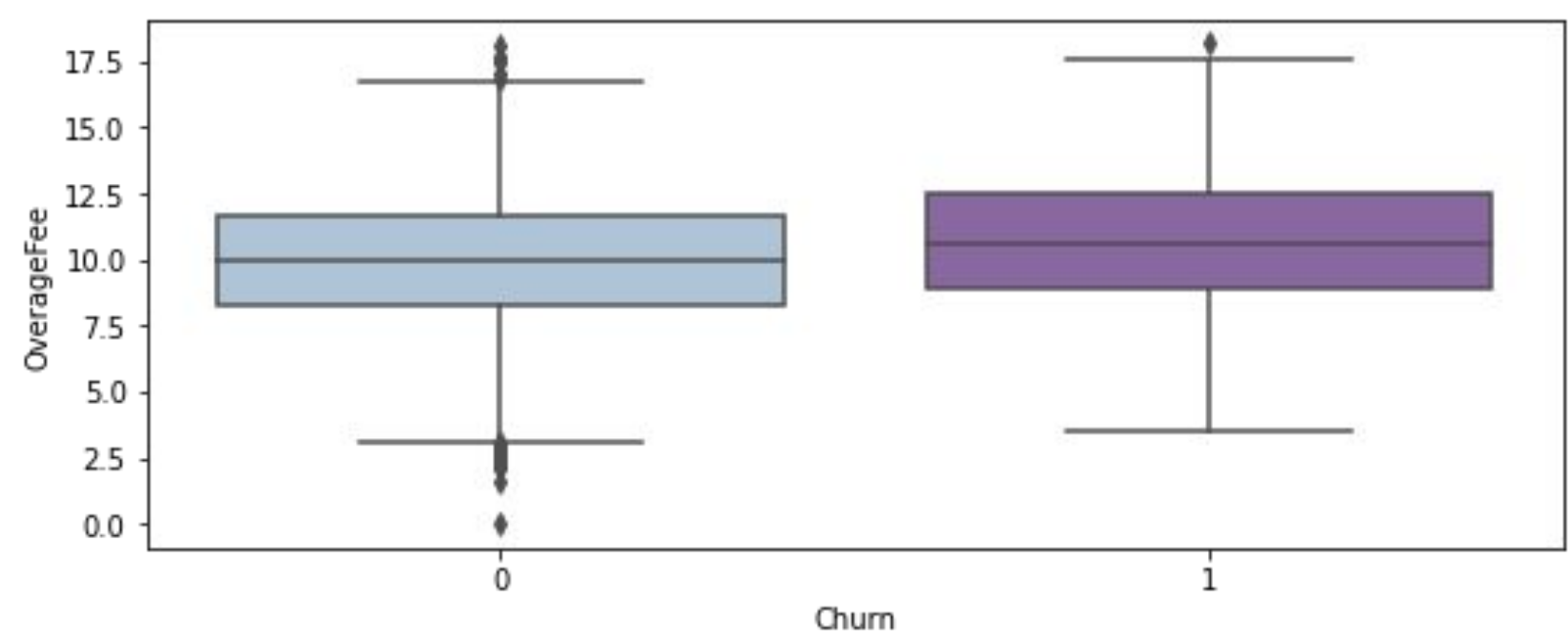
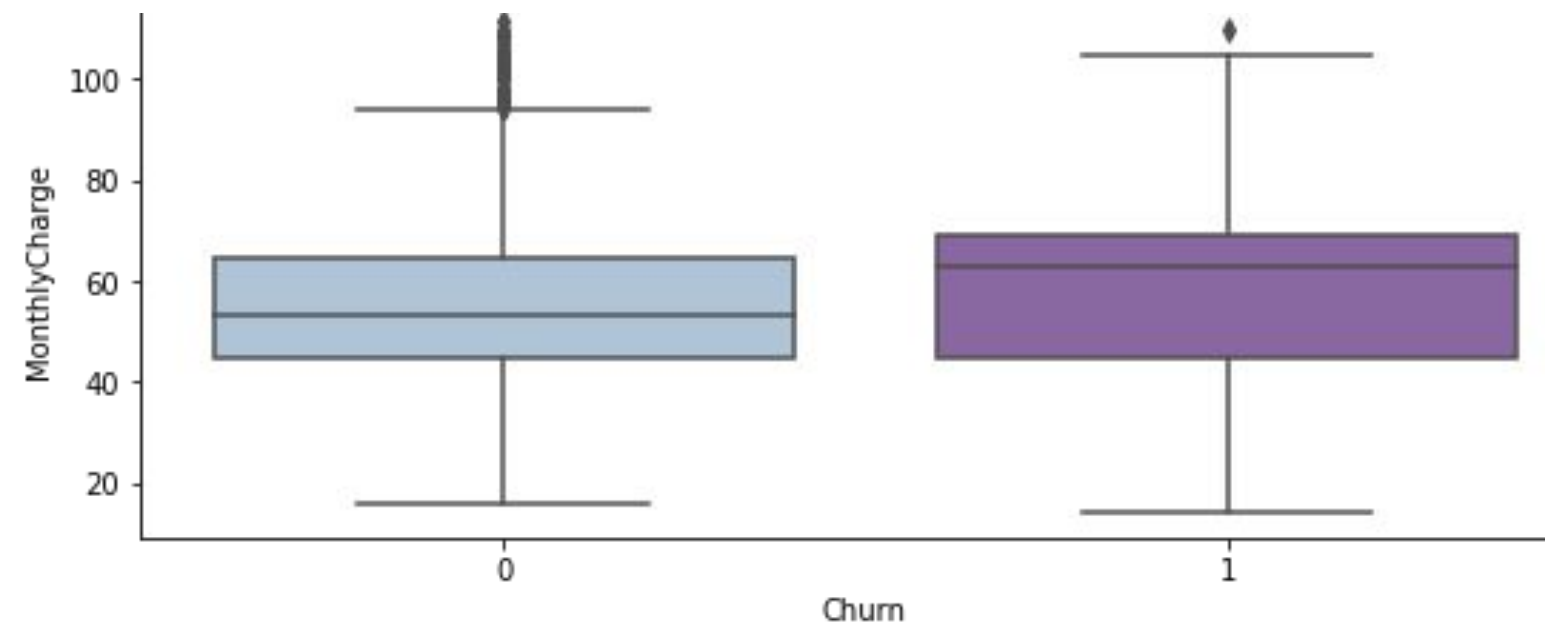
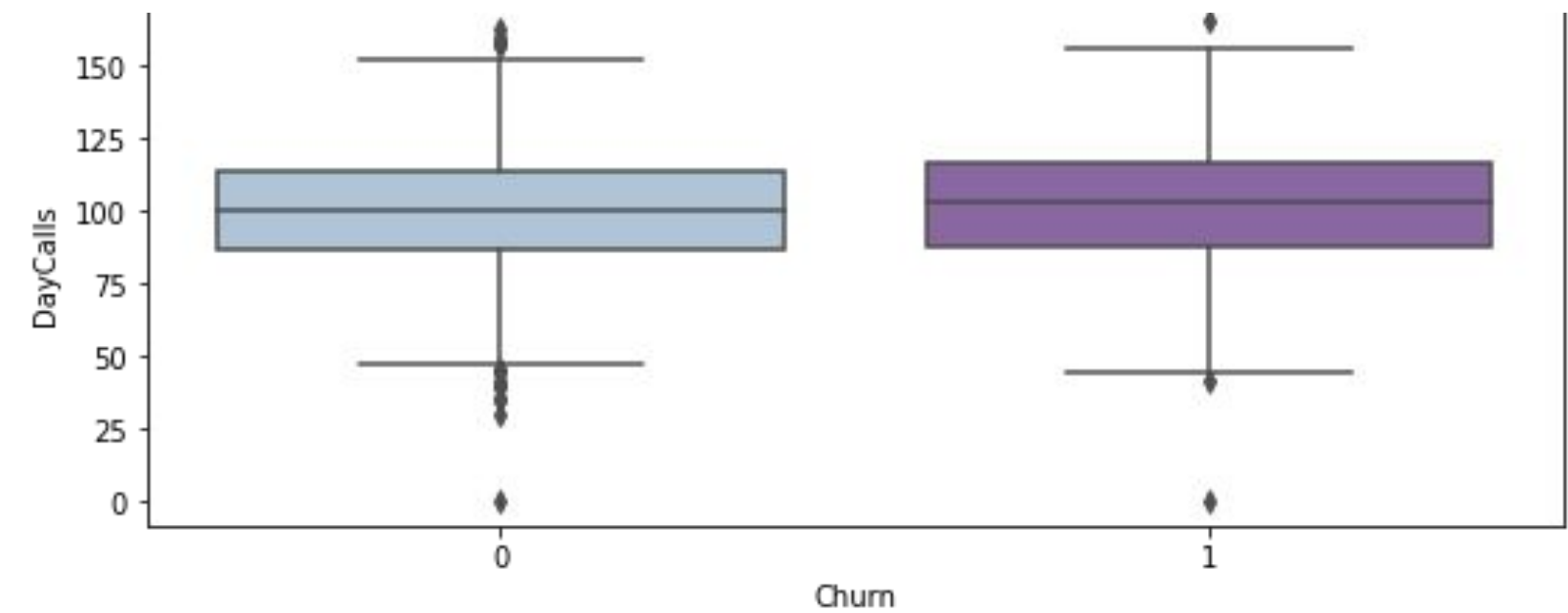
Dataset memiliki 3333 baris

Tidak ada Missing Value dalam dataframe, dan data berisi data Kategorikal dan Numerical

# Data Cleansing

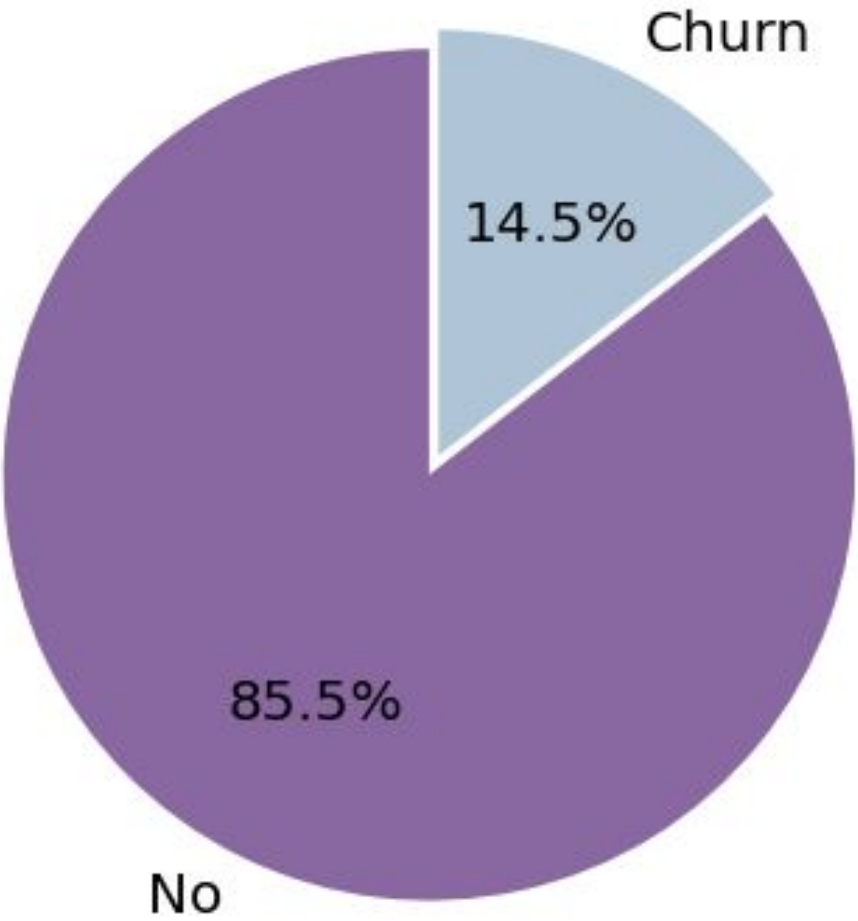
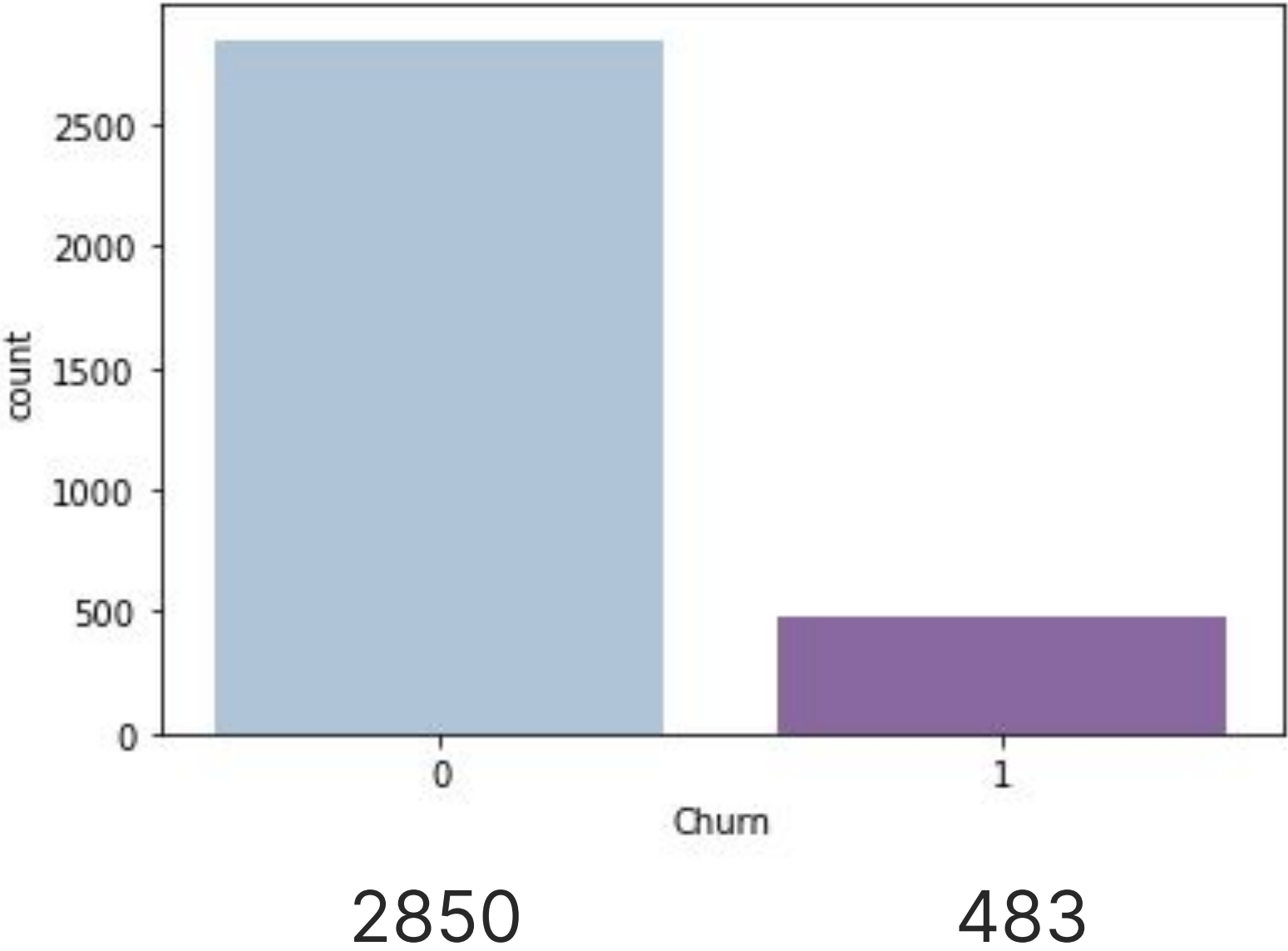


# Data Cleansing



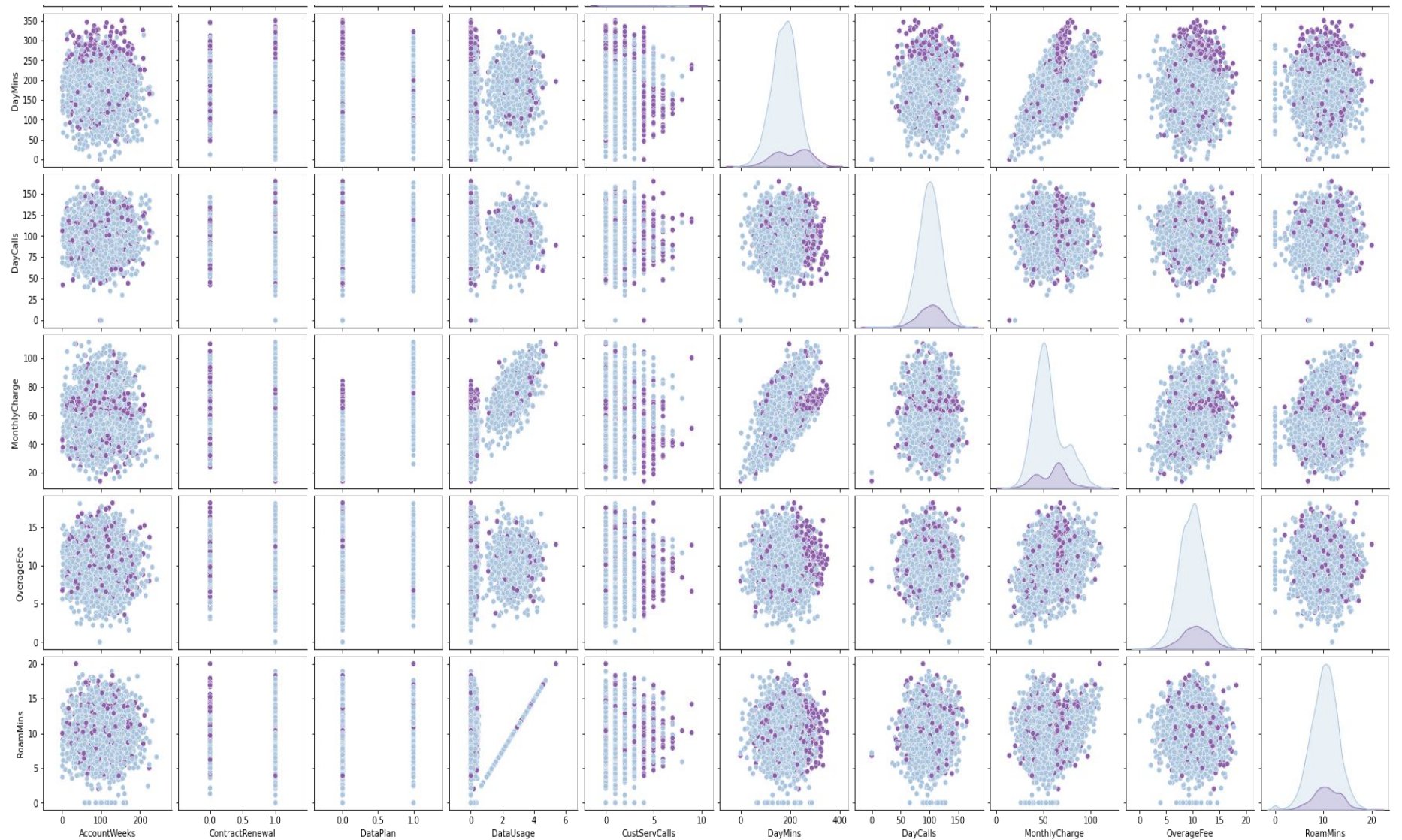
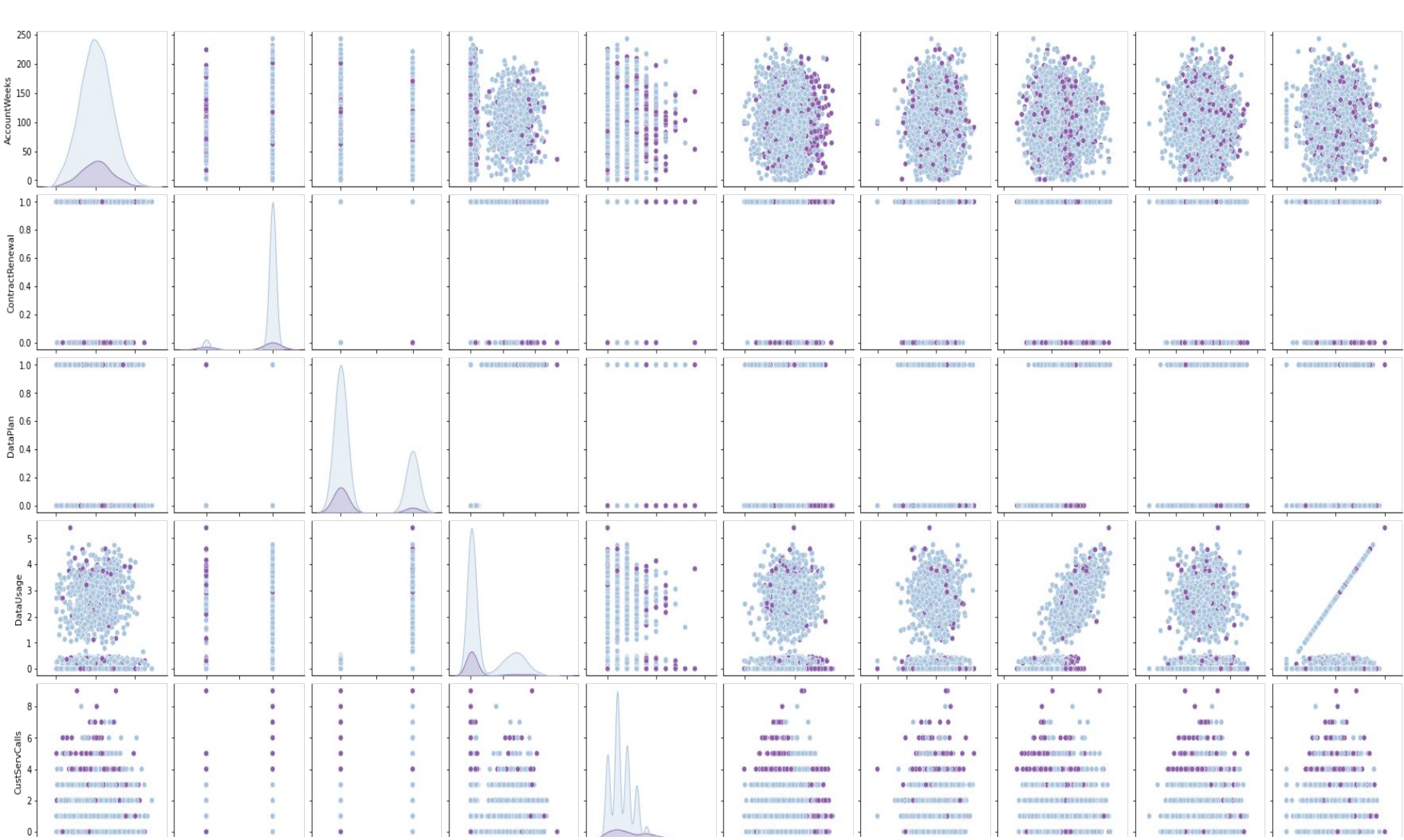
Outliers tidak perlu dibersihkan karena data tersebut masih mungkin terjadi

# Exploratory Data Analysis

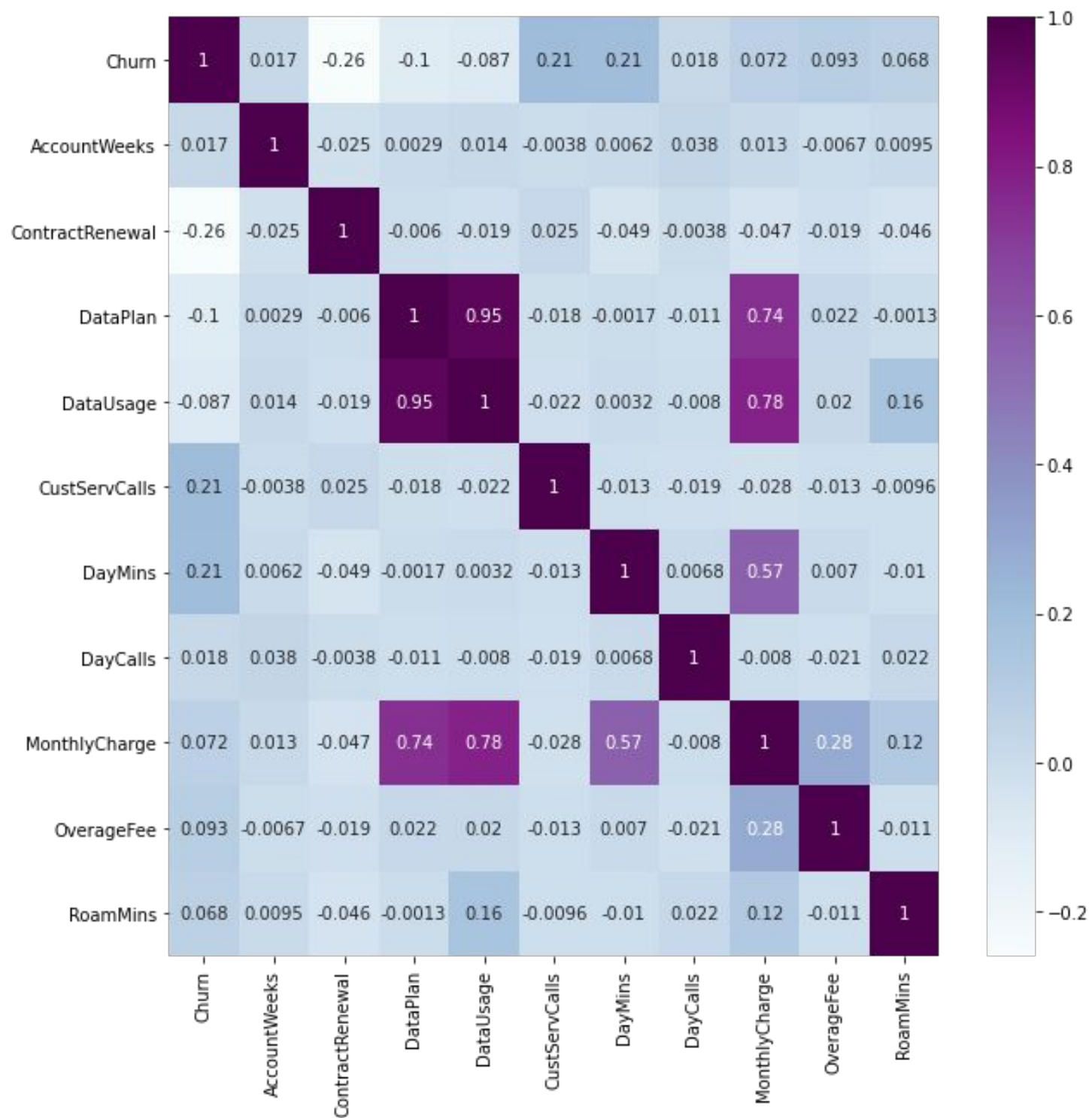




# Exploratory Data Analysis



# Exploratory Data Analysis



Terdapat korelasi yang sangat kuat dengan DataUsage dan DataPlan. Korelasi terkuat berikutnya adalah antara Penggunaan Data (DataPlan, DataUsage, DayMins) dan MonthlyCharge. Korelasi kuat antara OverageFee dan MonthlyCharge disebabkan oleh pelanggan dengan pembelanjaan yang lebih tinggi menjadi lebih mungkin untuk dikenakan dan overage charge.

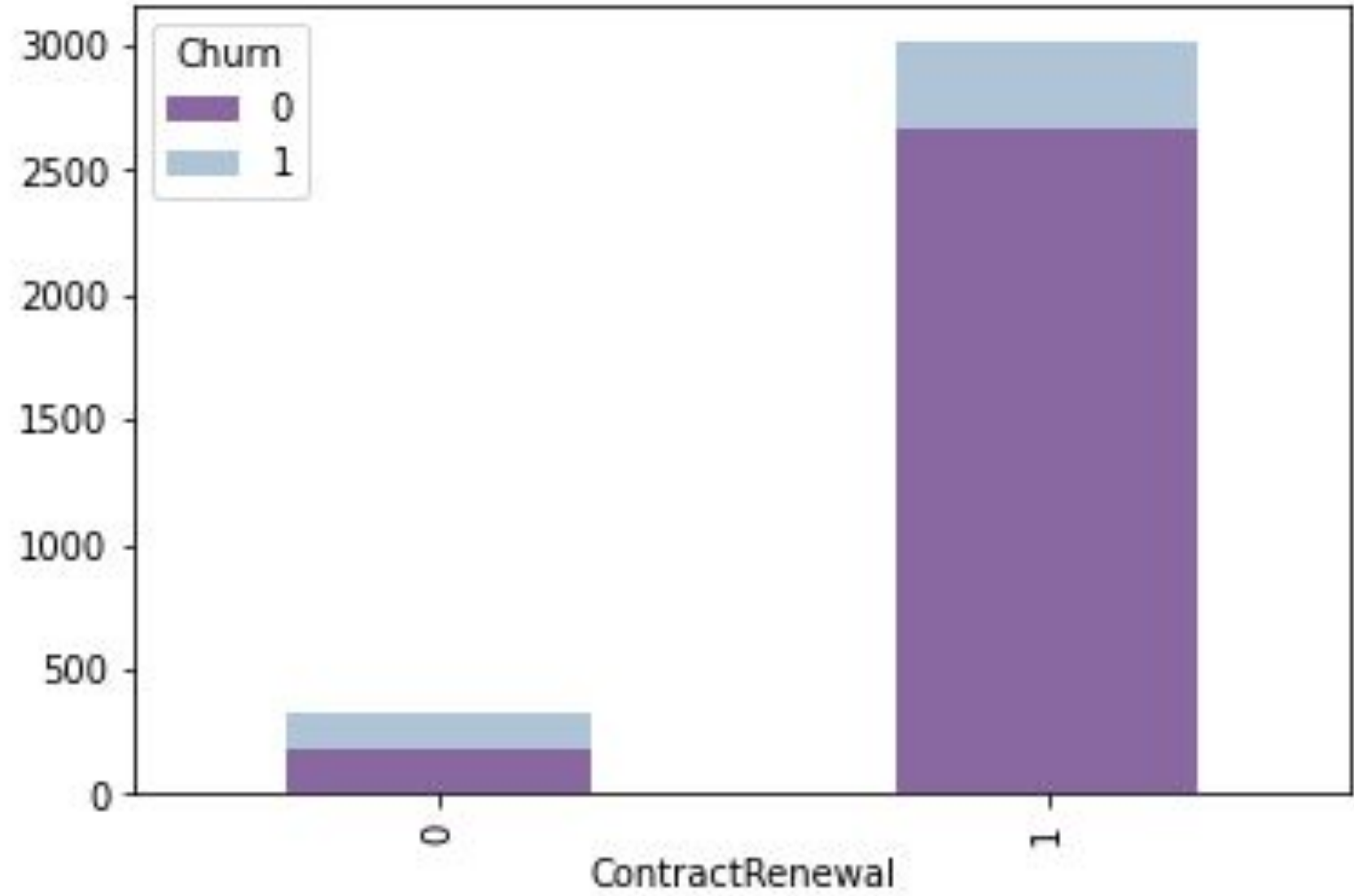


# Exploratory Data Analysis

## Data Biner (Kategorikal)

Contract Renewal

Churn		0	1
ContractRenewal	0	186	137
	1	2664	346

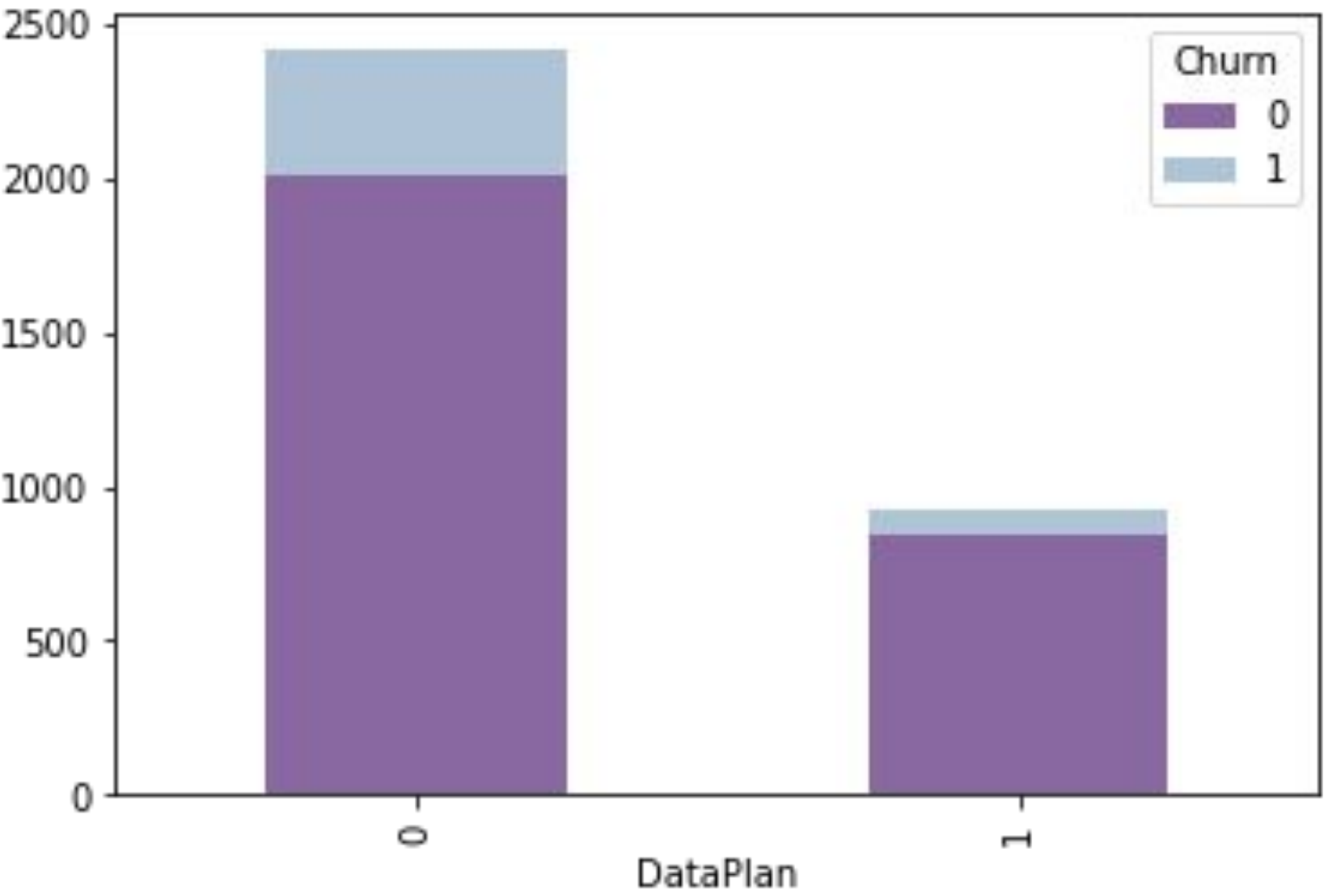


# Exploratory Data Analysis

## Data Biner (Kategorikal)

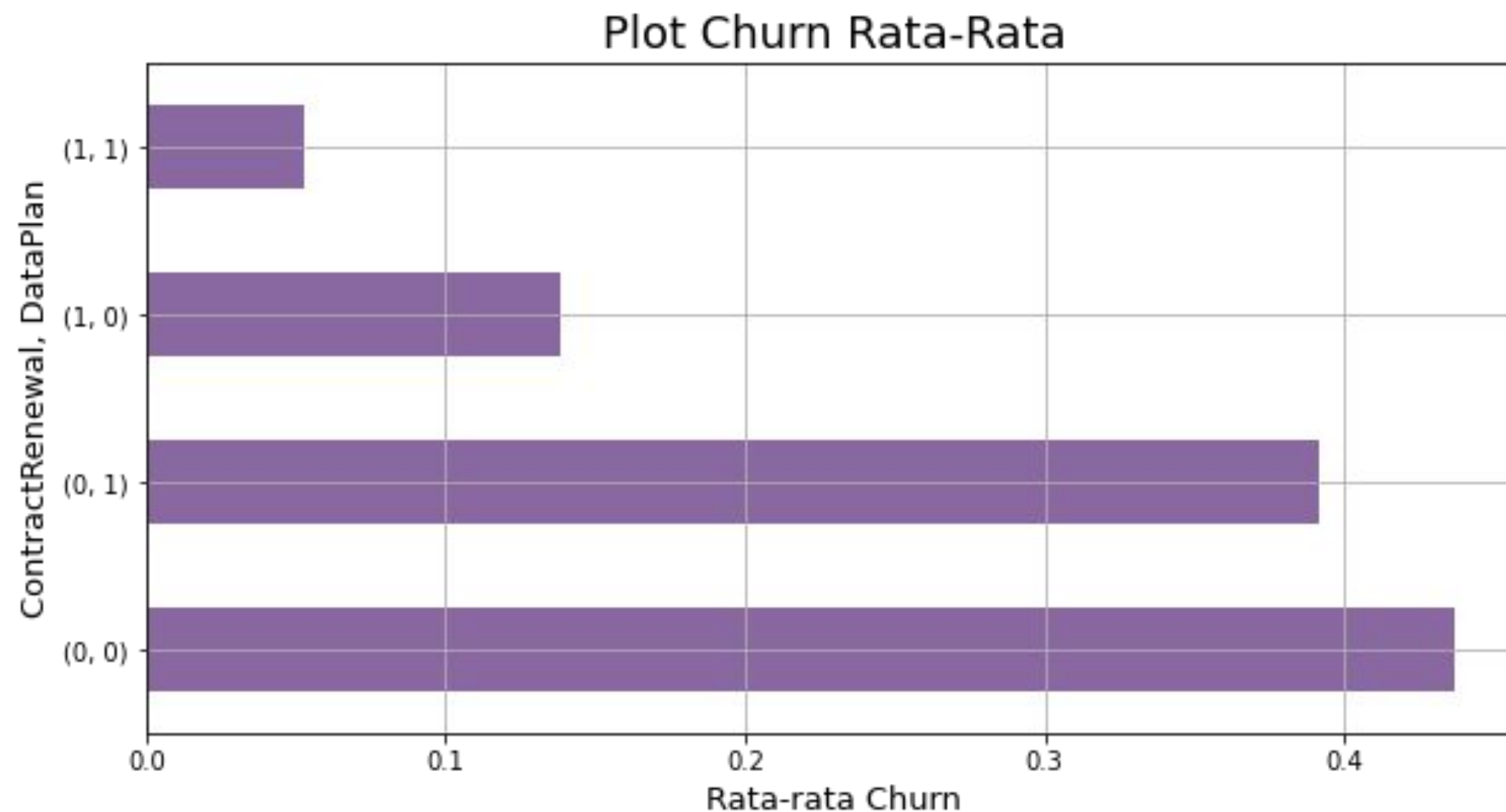
DataPlan

Churn		0	1
DataPlan			
0	2008	403	
1	842	80	



# Exploratory Data Analysis

## Data Biner (Kategorikal)



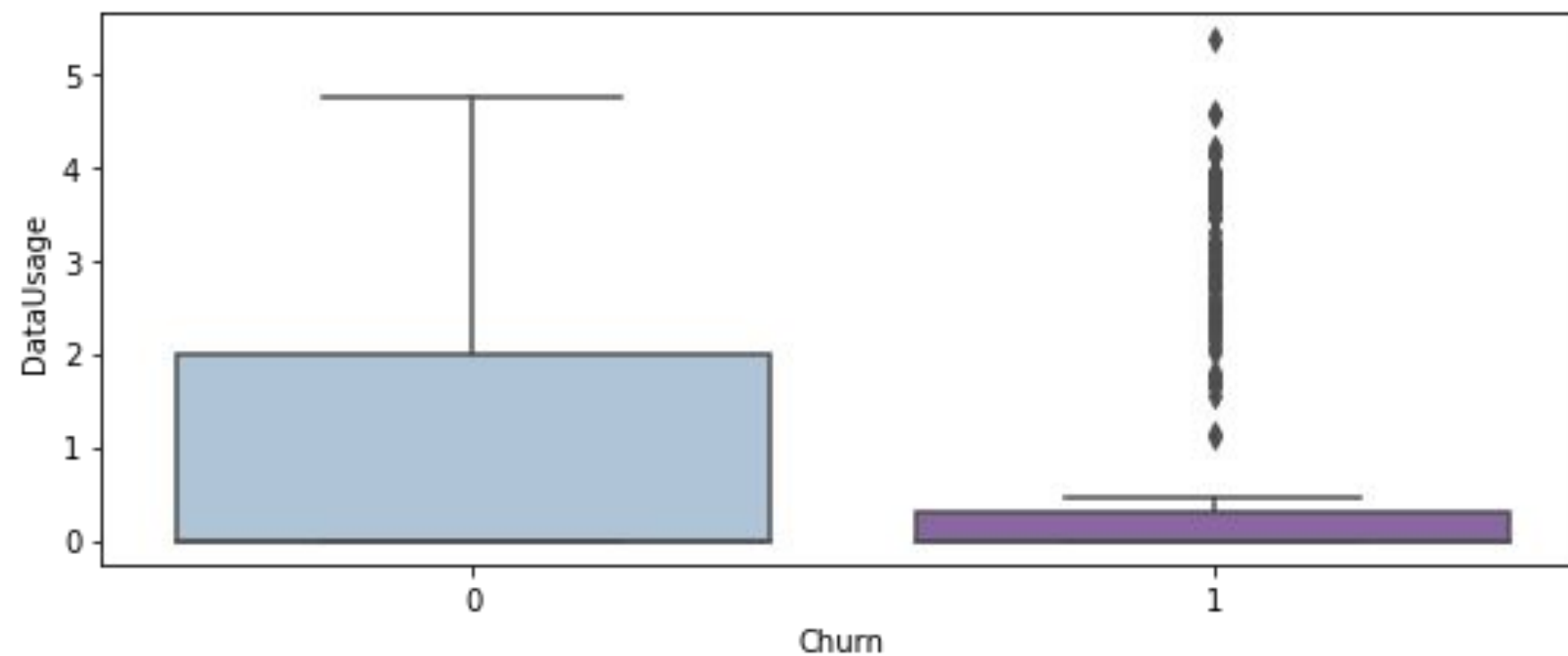
Atribut ContractRenewal dan DataPlan adalah dua atribut penting dalam churn pelanggan. Jika kedua atribut ini adalah "1", kemungkinan churn pelanggan rendah.

**Dampak ContractRenewal lebih besar dari DataPlan**



# Exploratory Data Analysis

## Data Numerikal

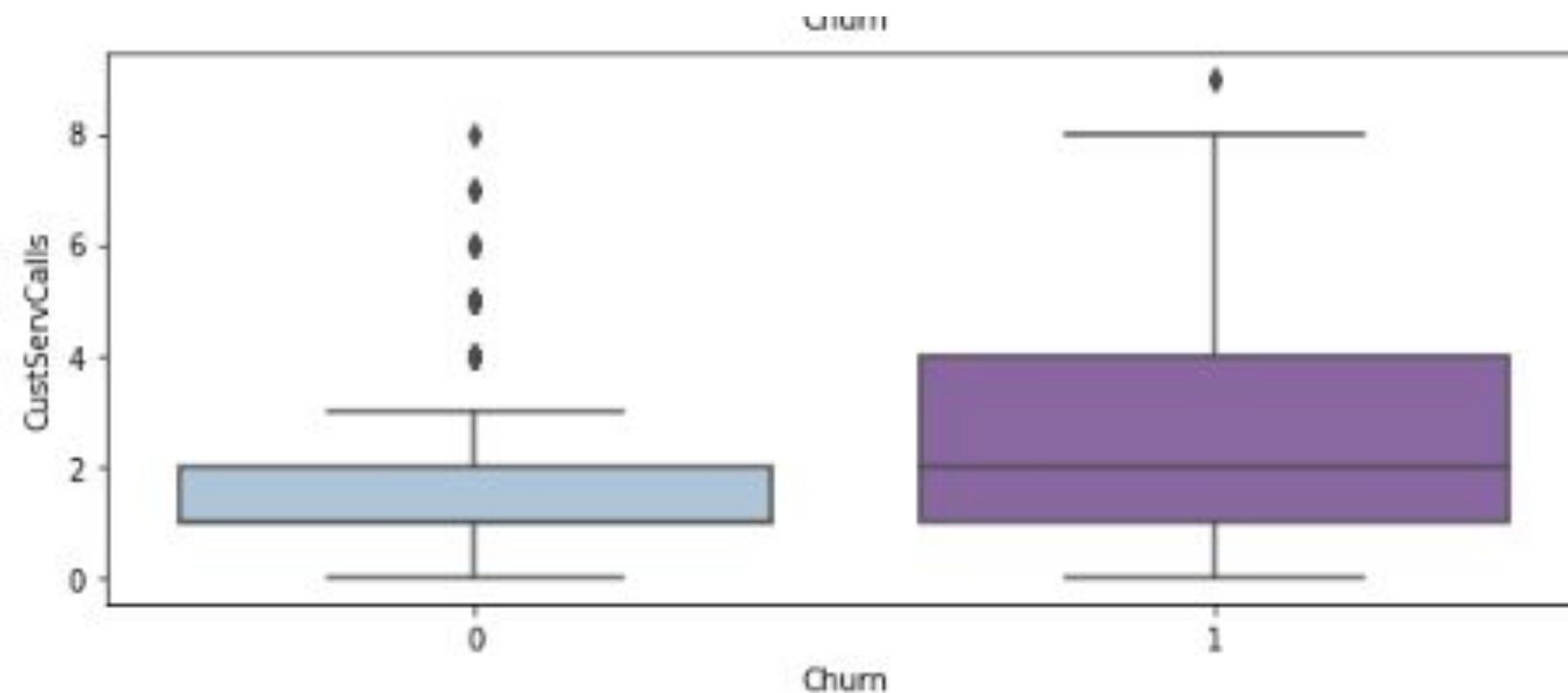


### DataUsage:

- Median dari grup churn maupun no churn bernilai 0.
- Dalam grup churn, rentang interquartile tidak sebesar dari grup yang no churn.
- Hal tersebut berarti data usage lebih menyebar untuk pelanggan yang kemudian tinggal di perusahaan (No churn).

# Exploratory Data Analysis

## Data Numerikal

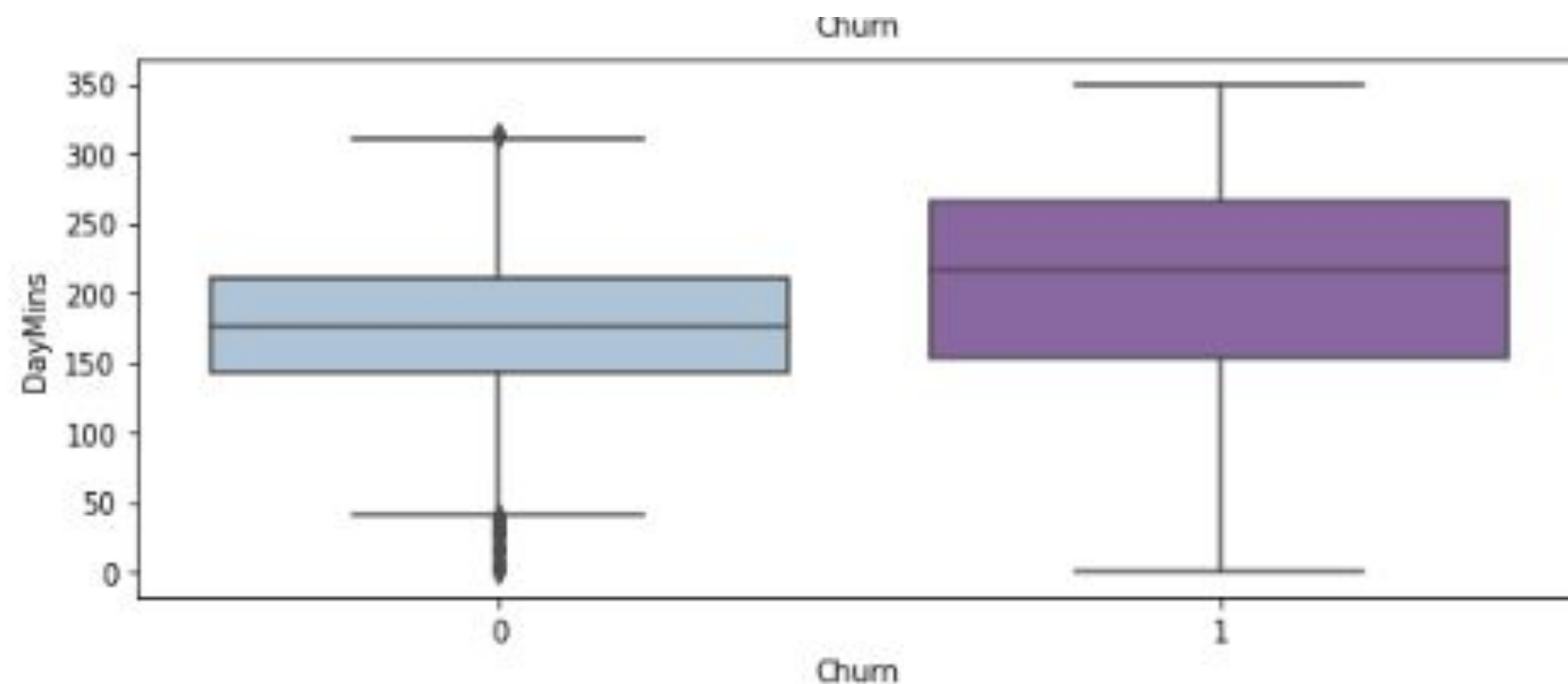


### CustServCalls:

- Customer yang churn rata-rata memiliki satu panggilan Customer Service daripada mereka yang tetap tinggal di perusahaan.

# Exploratory Data Analysis

## Data Numerikal

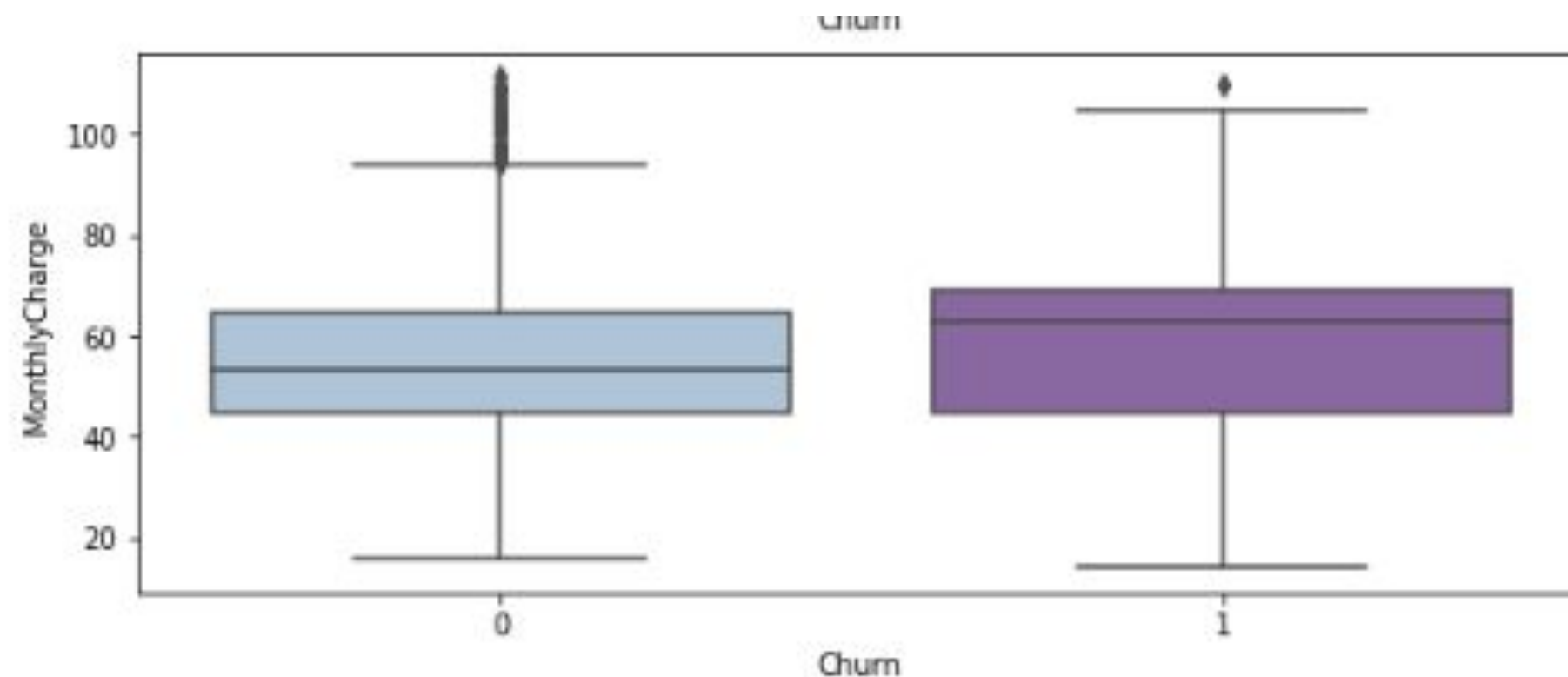


DayMins:

- Memiliki rata-rata menit harian yang lebih tinggi per bulan untuk customer yang churn.

# Exploratory Data Analysis

## Data Numerikal



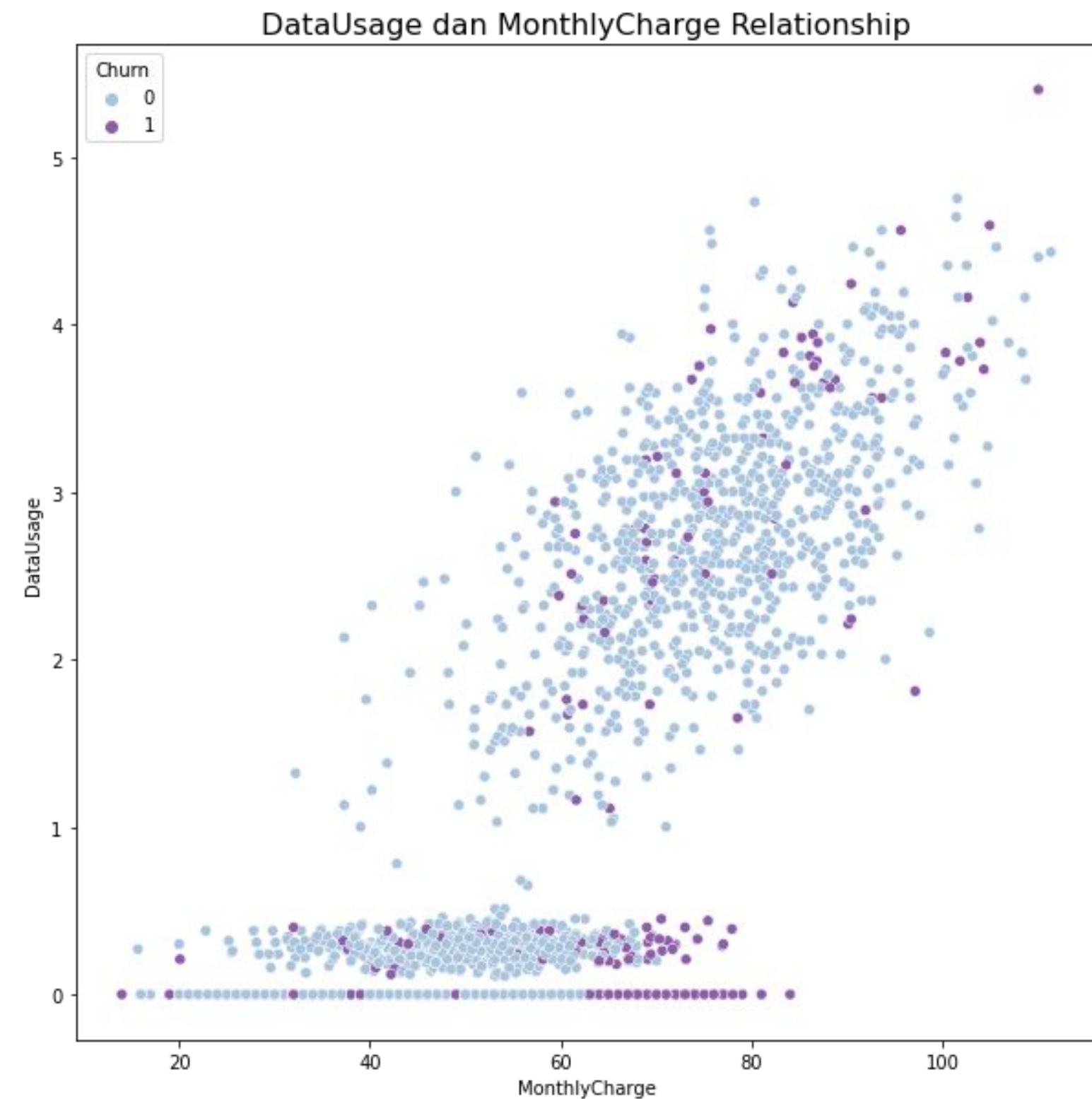
### MonthlyCharge:

- Customer yang churn memiliki rata-rata Monthly Charge yang lebih tinggi daripada yang tinggal di perusahaan.

# Exploratory Data Analysis

## Melihat Relationship

Relationship terbaik antara 2 data numerikal (MonthlyCharge dan DataUsage)

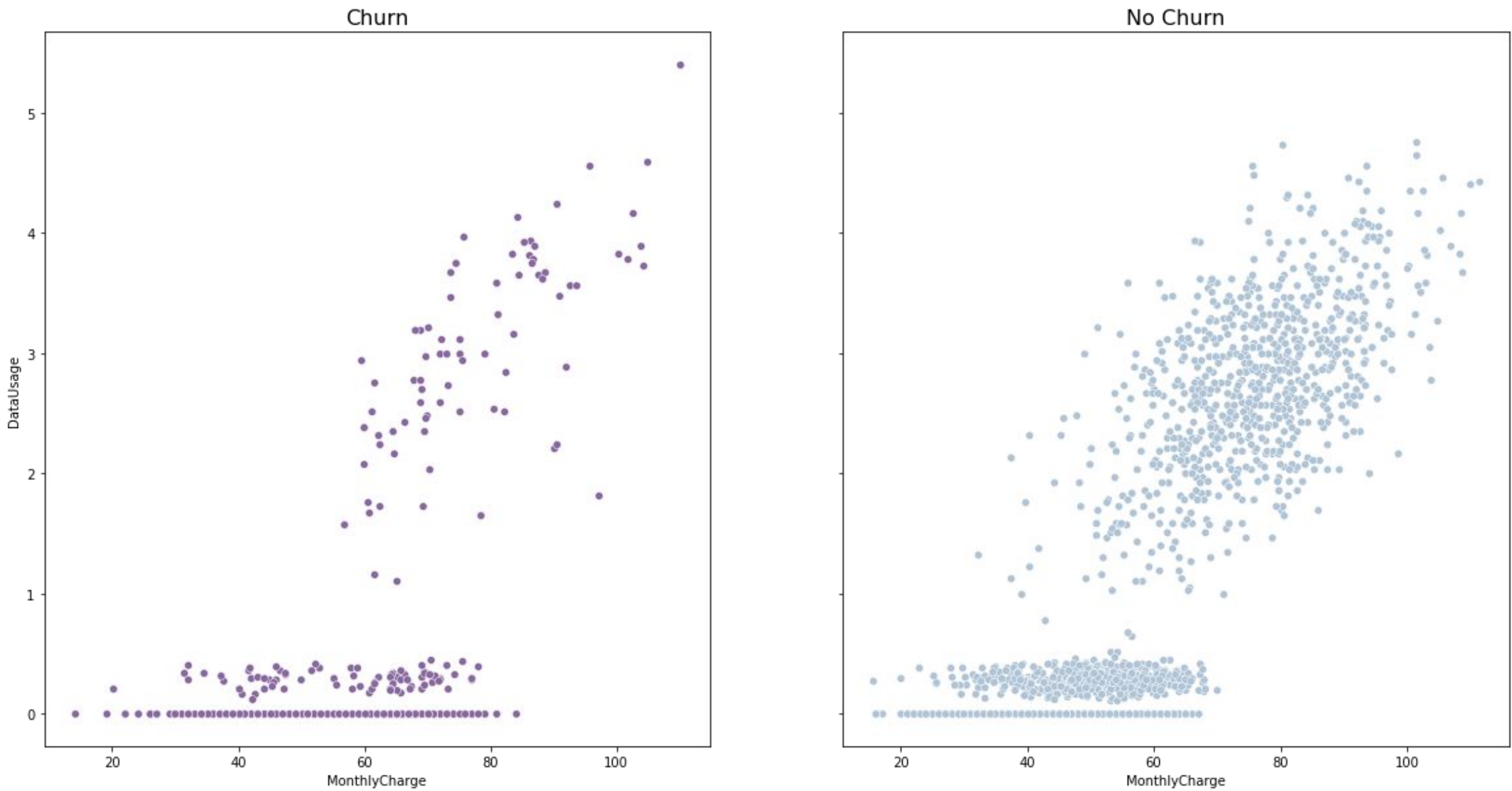




# Exploratory Data Analysis

## Melihat Relationship

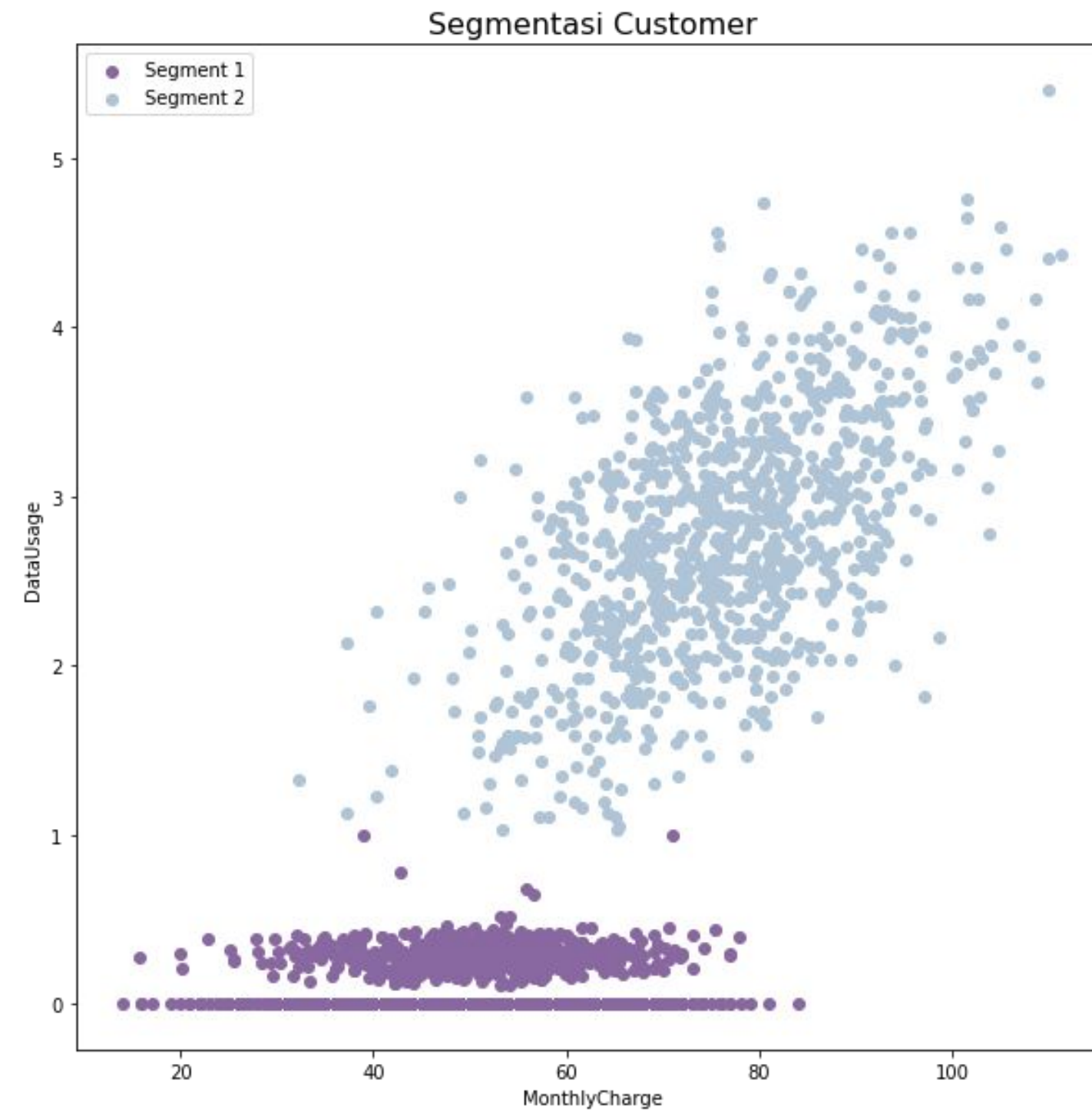
Pemisahan Churn dan No Churn



# Exploratory Data Analysis

## Segmentasi

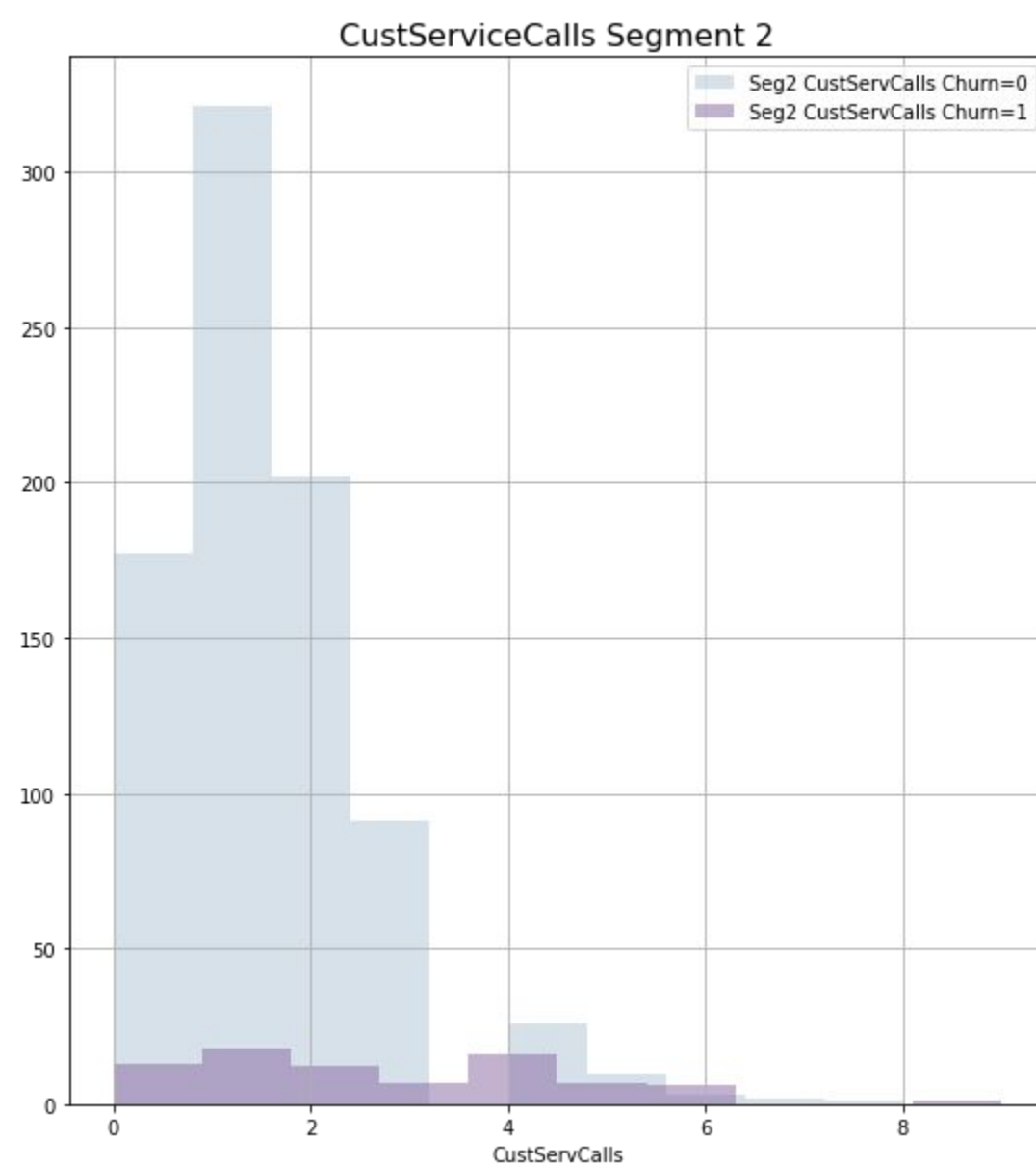
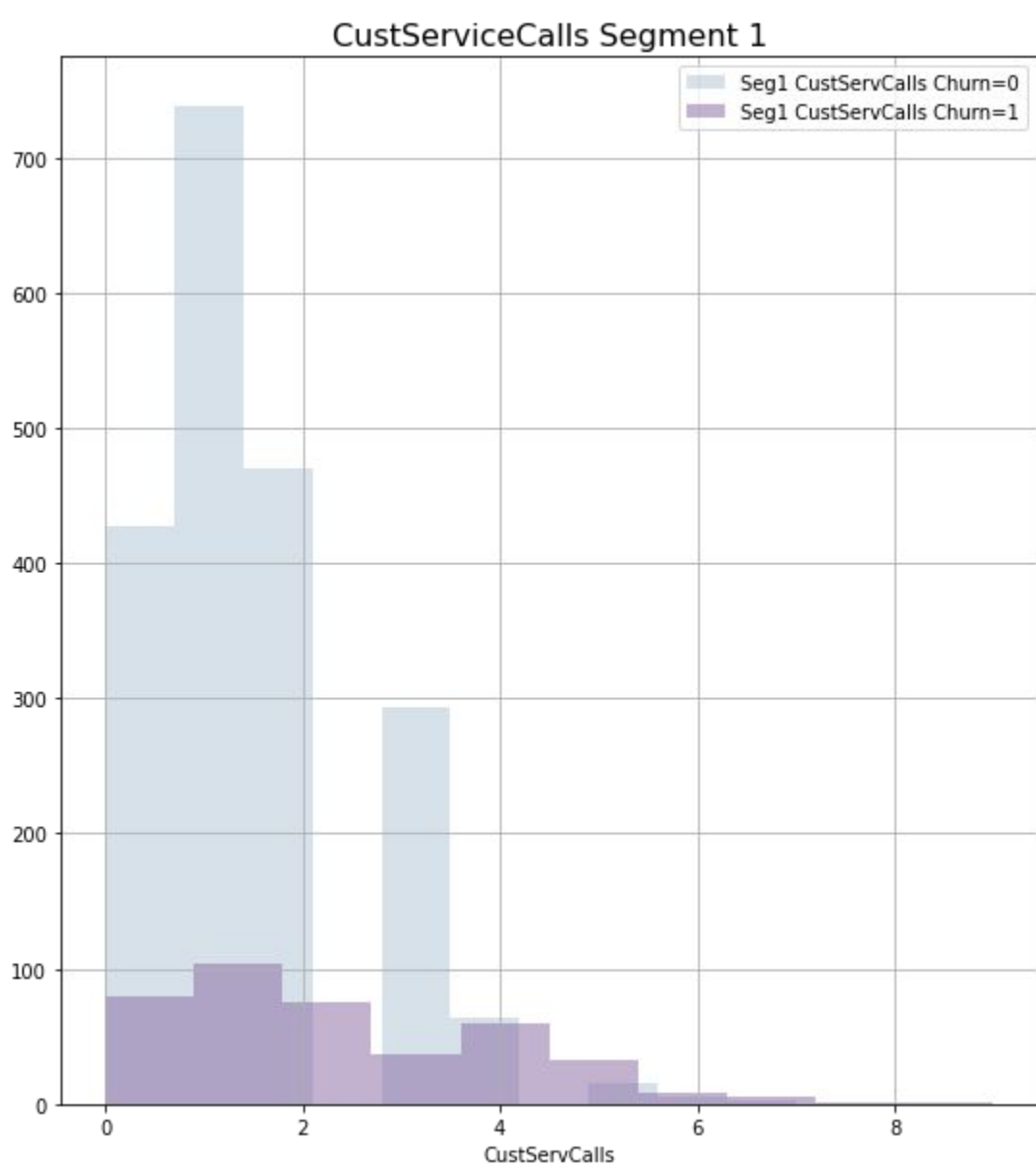
- DataUsage 0-1 = Segment 1
- DataUsage >1 = Segment 2



# Exploratory Data Analysis

## Segmentasi

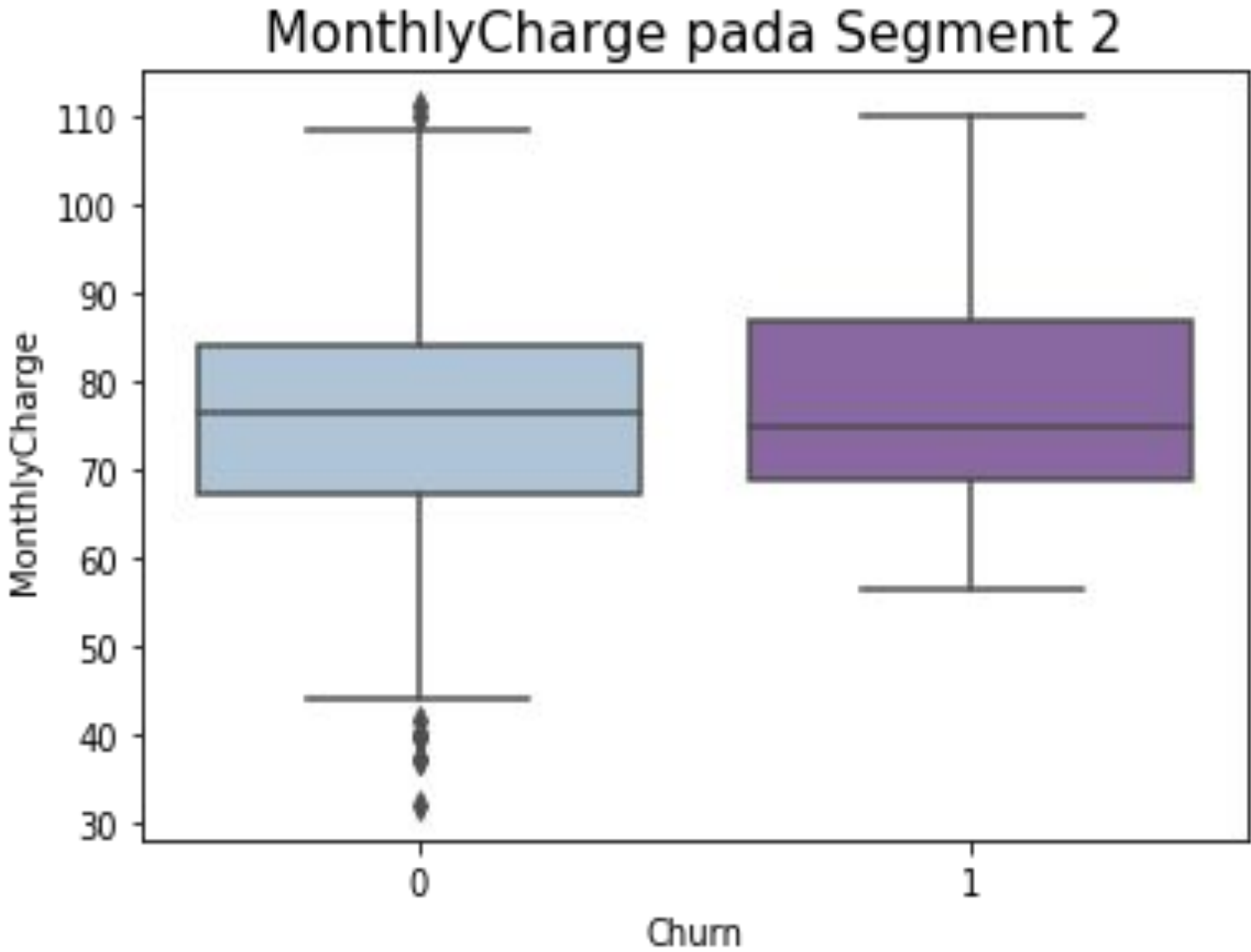
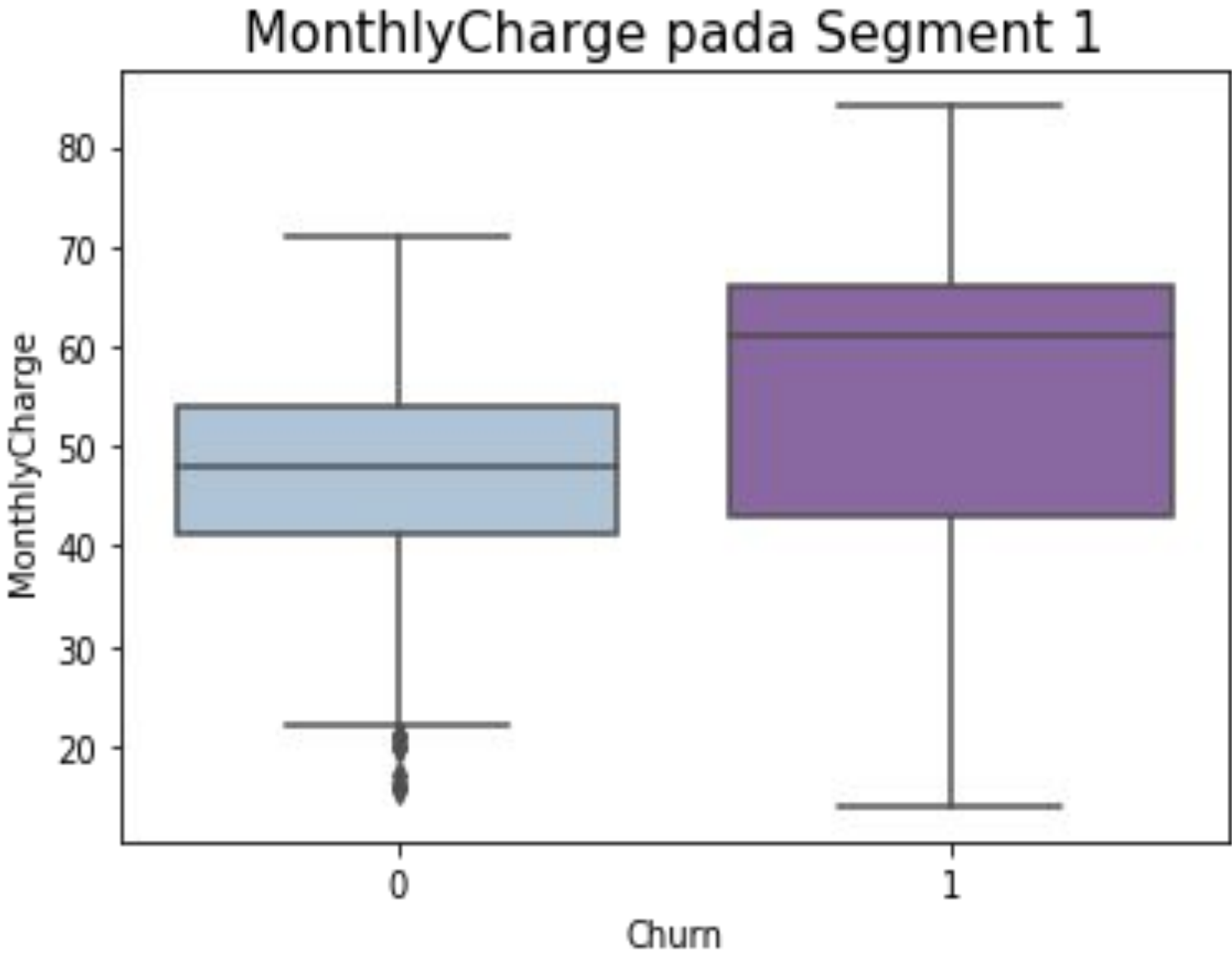
## CustServCalls



# Exploratory Data Analysis

## Segmentasi

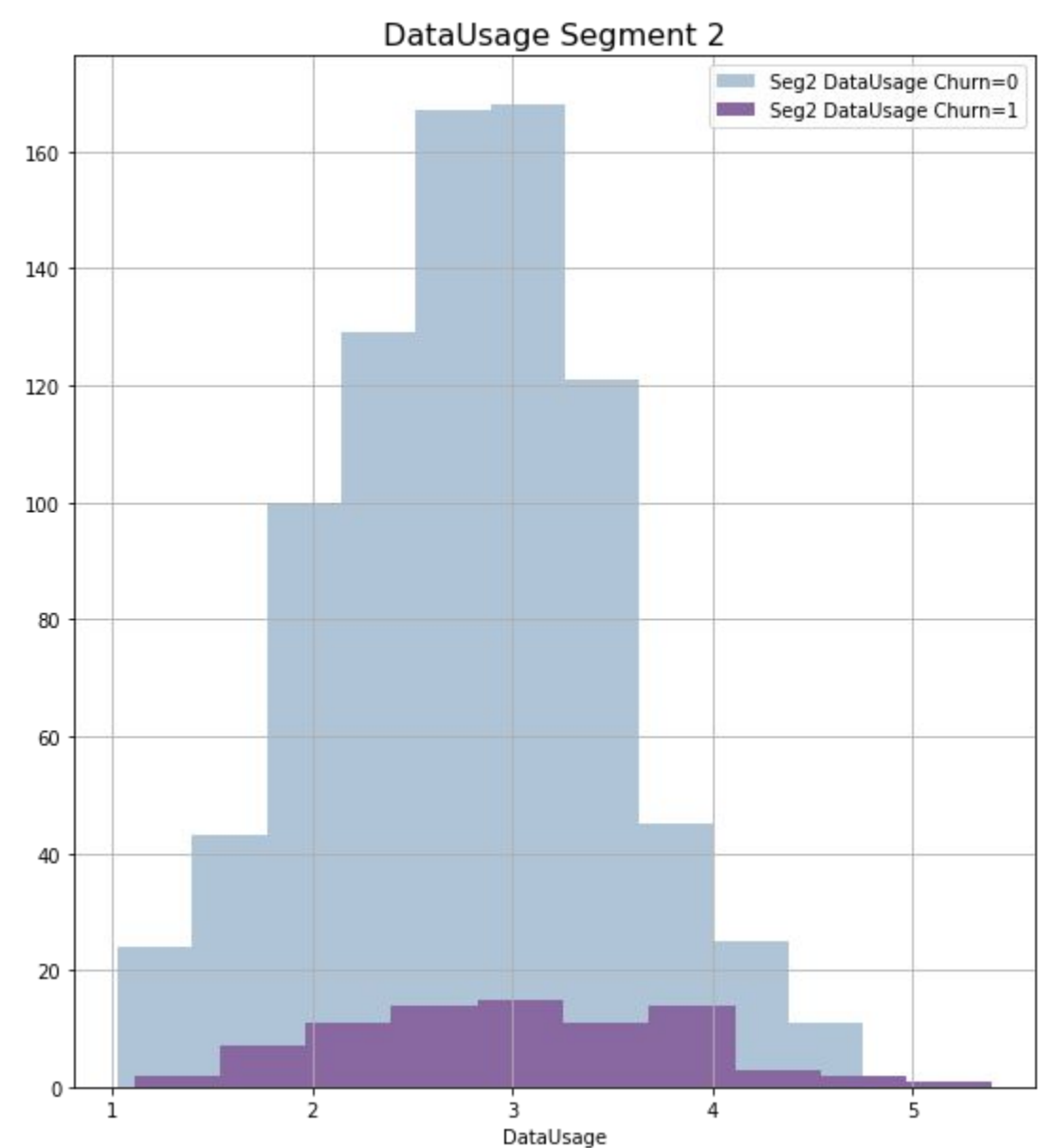
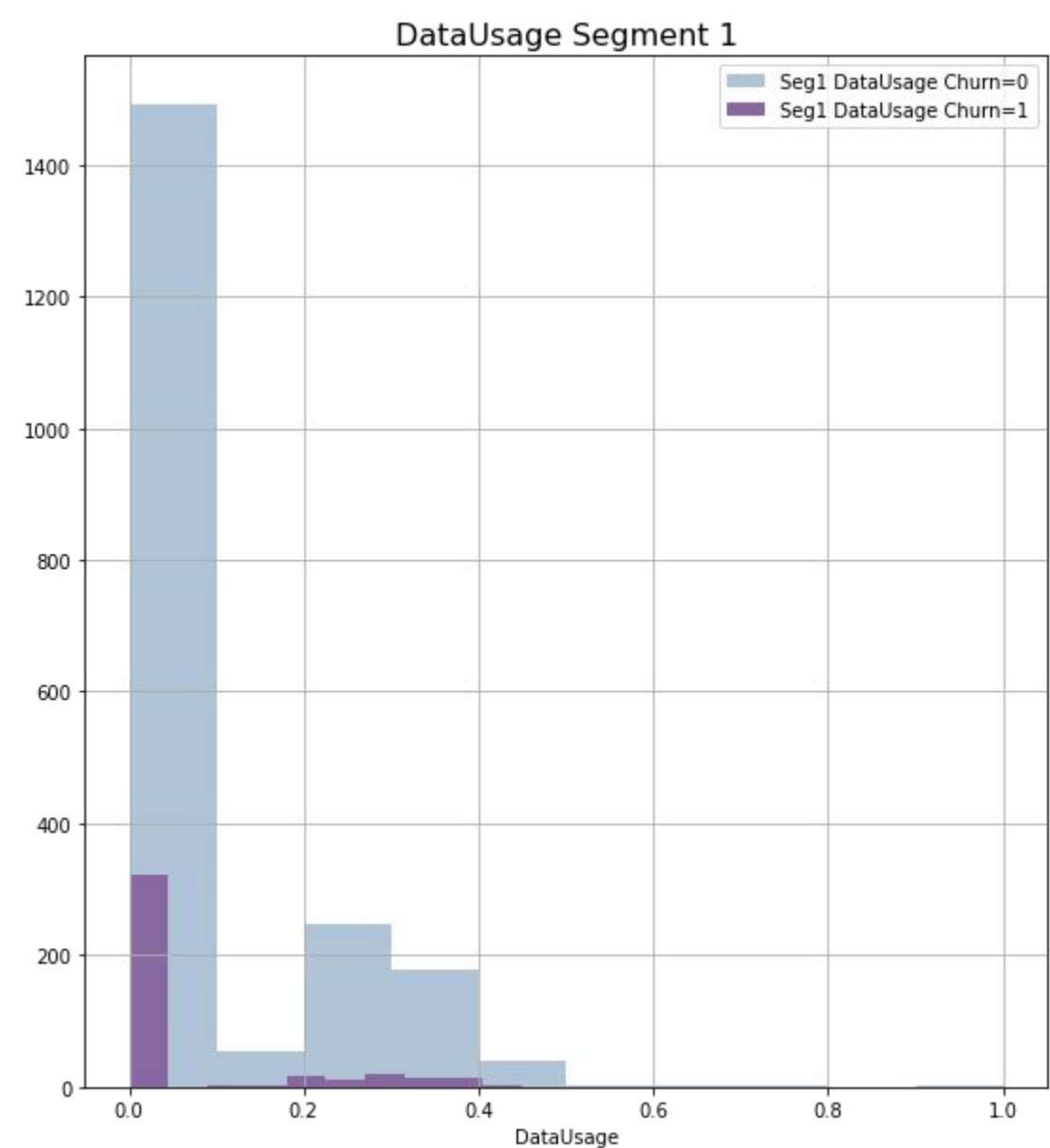
### MonthlyCharge



# Exploratory Data Analysis

## Segmentasi

## DataUsage

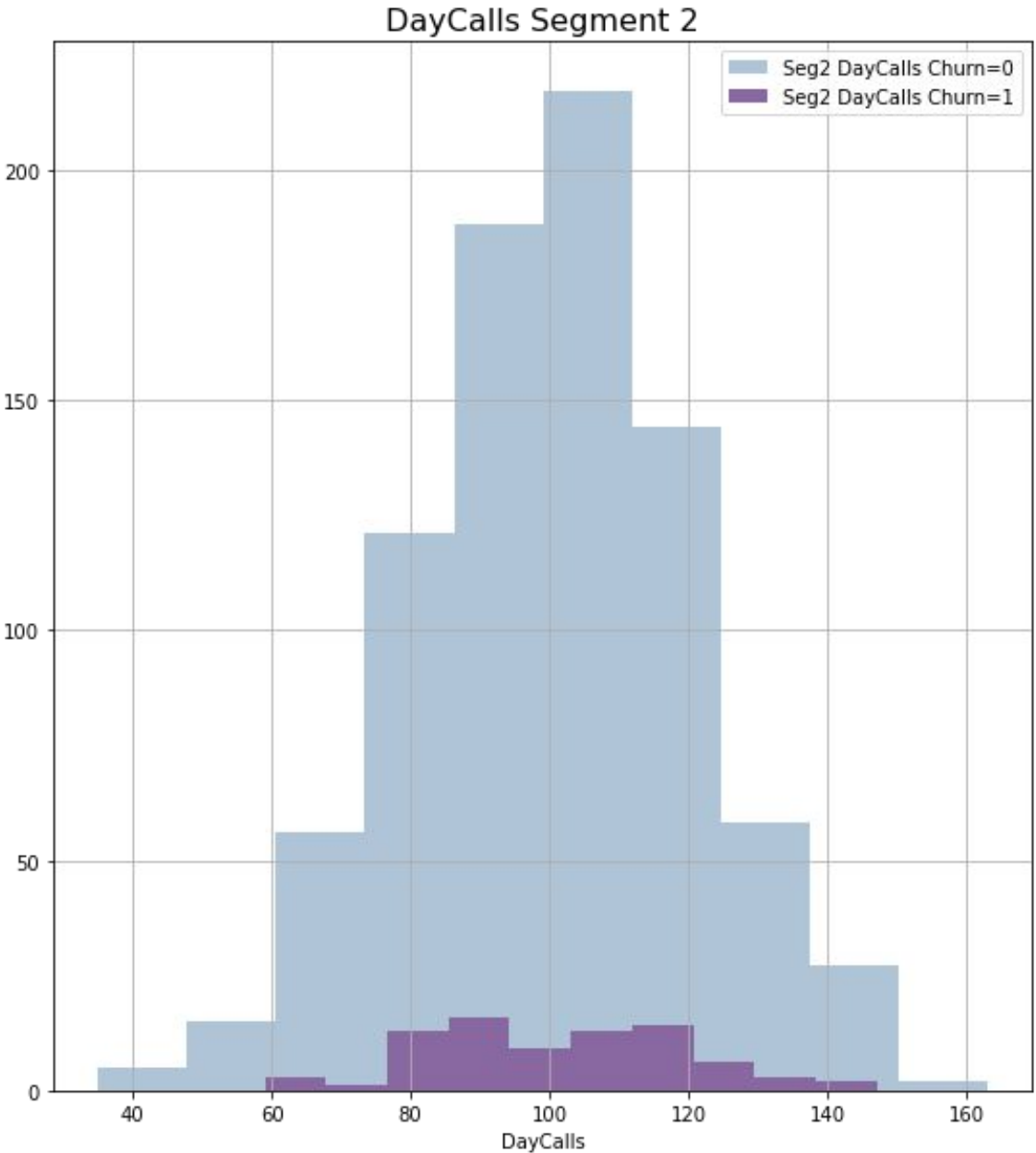
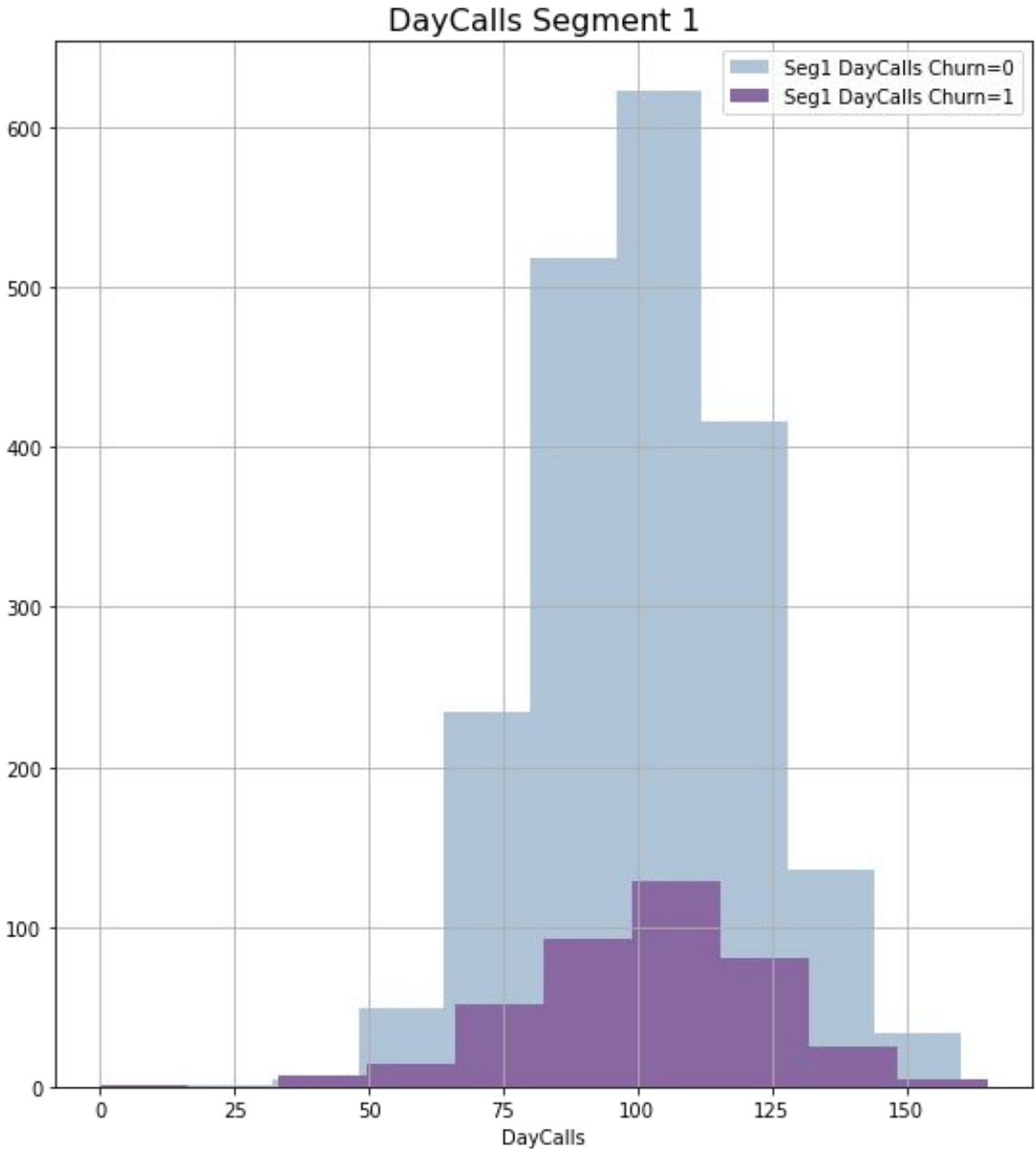




# Exploratory Data Analysis

## Segmentasi

### DayCalls



# Exploratory Data Analysis

## Segmentasi

DataPlan

Segment 1

Churn	0	1
DataPlan		
0	83.284944	16.715056
1	100.000000	0.000000

Segment 2

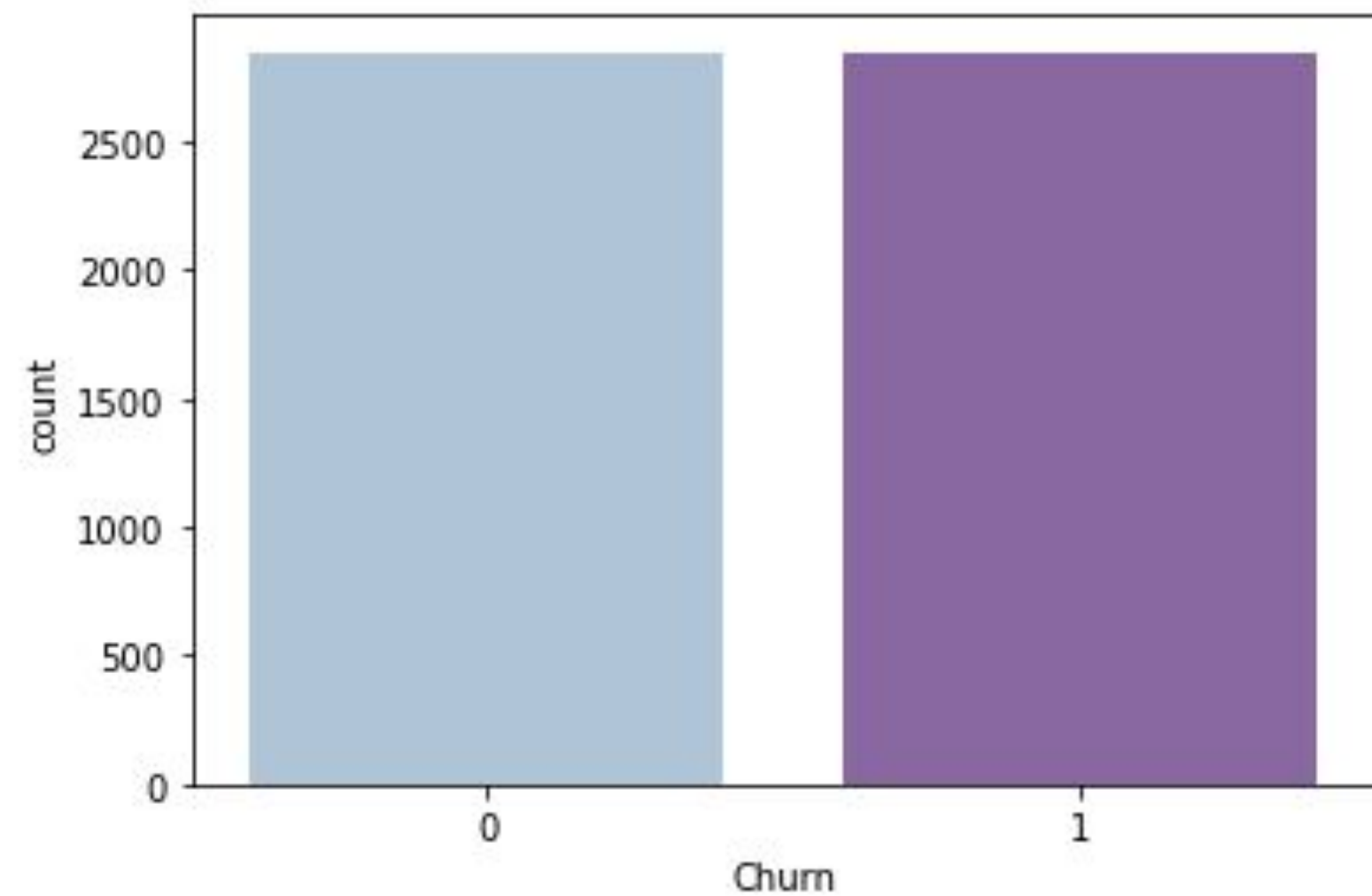
Churn	0	1
DataPlan		
1	91.237678	8.762322

# Modelling

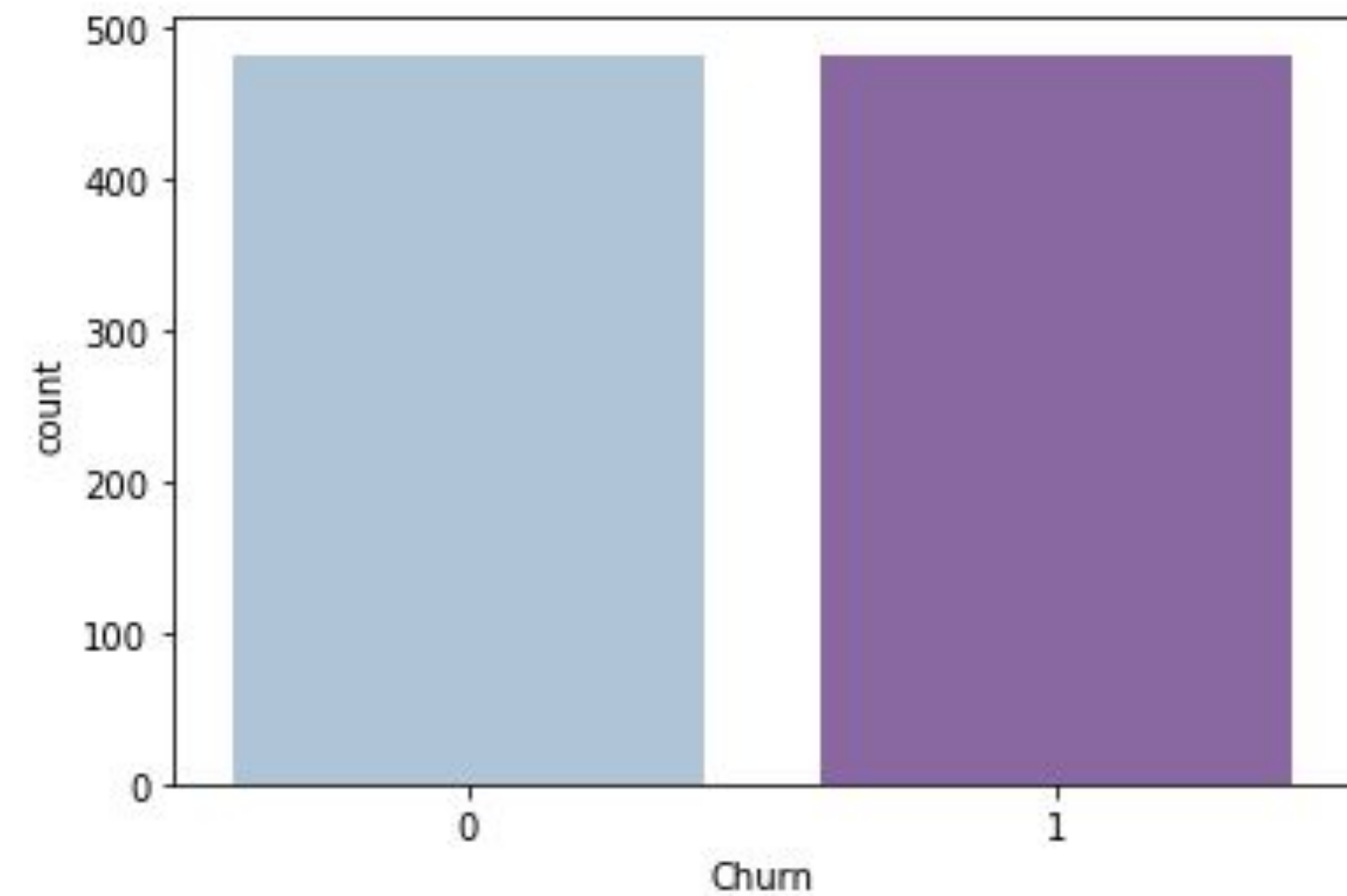
# Persiapan

## 1. Imbalance Problem

- Upper Sampling



- Under Sampling



# Persiapan

## 2. One Hot Encoding

One Hot Encoding tidak ada pengaruh karena sebelumnya dataframe tidak mempunyai tipe data string

## 3. Train Test Split

Data yang sudah di upper sampling dan under sampling kemudian di train untuk langkah selanjutnya yaitu modelling

# Modelling

## 1. Logistic Regression

## 2. Decision Tree

## 3. Random Forest

- Tanpa Hyperparameter Tuning
- Menggunakan Hyperparameter Tuning



# Evaluation

## Data Default (sebelum Upper Sampling/Under Sampling)

Pemodelan	Akurasi	Presisi	Recall
Logistic Regression	0,849	0,669	0,552
Decision Tree	0,865	0,735	0,741
Random Forest (tanpa Hyperparameter Tuning)	0,921	0,923	0,751
Random Forest (menggunakan Hyperparameter Tuning)	0,933	0,908	0,812

Model yang paling bagus untuk tahap ini adalah **Random Forest menggunakan Hyperparameter Tuning**

# Evaluation

## Upper Sampling

Pemodelan	Akurasi	Presisi	Recall
Logistic Regression	0,747	0,747	0,747
Decision Tree	0,952	0,954	0,952
Random Forest (tanpa Hyperparameter Tuning)	0,880	0,880	0,880
Random Forest (menggunakan Hyperparameter Tuning)	0,932	0,933	0,932

Model yang paling bagus untuk tahap ini adalah **Decision Tree**

# Evaluation

## Under Sampling

Pemodelan	Akurasi	Presisi	Recall
Logistic Regression	0,757	0,758	0,756
Decision Tree	0,736	0,736	0,736
Random Forest (tanpa Hyperparameter Tuning)	0,852	0,852	0,852
Random Forest (menggunakan Hyperparameter Tuning)	0,829	0,829	0,829

Model yang paling bagus untuk tahap ini adalah **Random Forest tanpa Hyperparameter Tuning**

# Conclusion

## **Cara untuk Mempertahankan Customer**

- Semakin banyak Customer Service Call, maka semakin rentan untuk Churn
- Optimalkan harga waktu panggilan untuk customer segmen 1
- Perkenalkan paket data kepada customer yang menggunakan data tanpa paket data secepatnya.
- Optimalkan harga paket data untuk customer segmen 2

**Terima  
kasih!**

**Ada pertanyaan?**

**zenius**



**Kampus  
Merdeka**  
INDONESIA JAYA