

VISUALISASI & MACHINE LEARNING

Dipersembahkan oleh Nurul Fadil



Hello i'am Nurul Fadil.

i'am data science antusiast form FGA
Binar Academy

TABLE CONTAINS



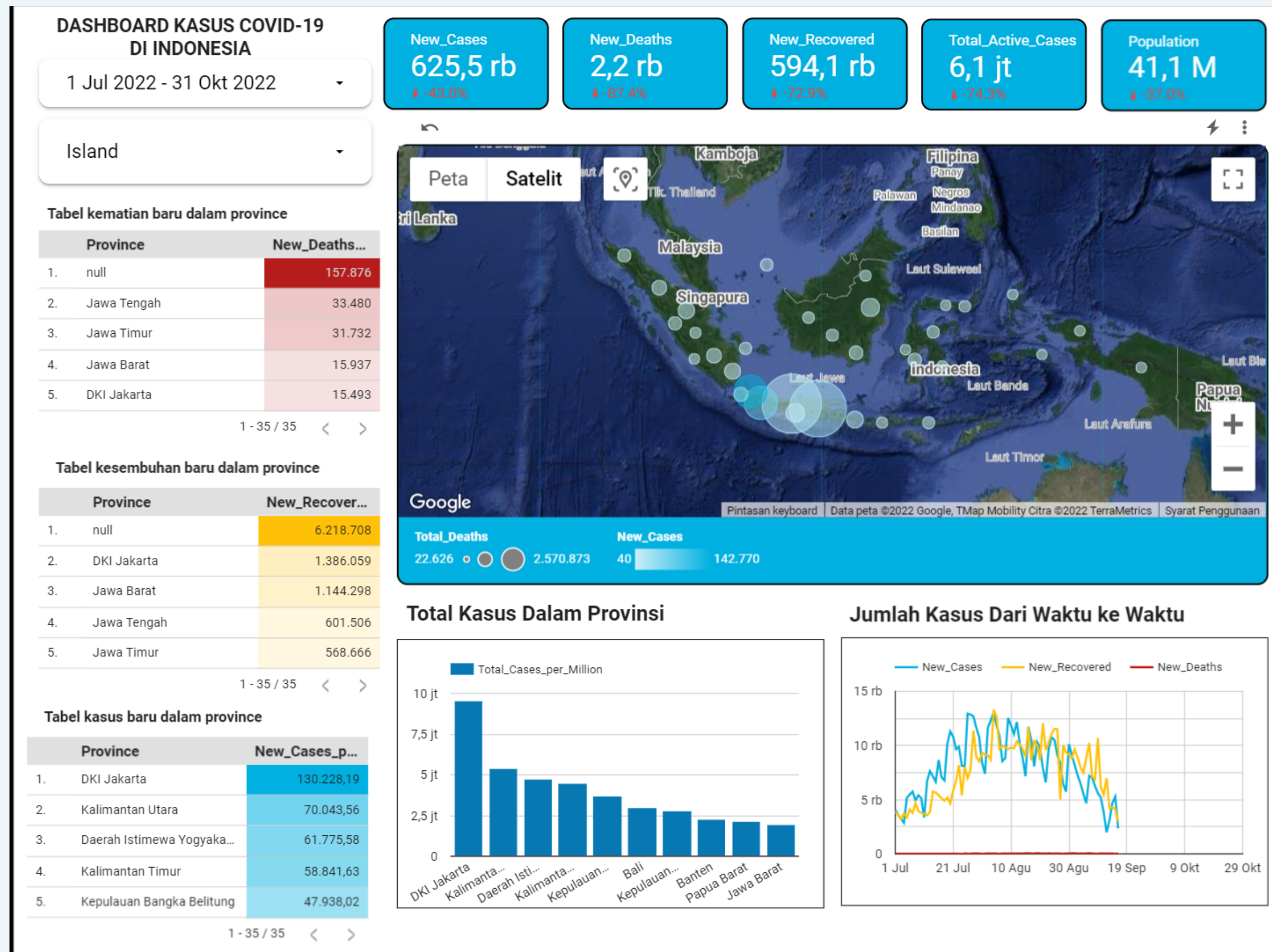
Visualisasi Dashboard



Machine Learning

- Business Problem
- Data Understanding
- EDA
- Feature Engineering
- Modeling
- Predict With Data Test
- Conclusion
- Appendix

Dashboard Visualisasi Kasus Covid-19



Penjelasan

Dashboard interaktif untuk mengetahui jumlah kasus covid-19 dengan rentang waktu yang diinginkan pembaca dan kolom pencarian informasi berdasarkan provinsi.

Dashboard ini digunakan untuk mengetahui lebih cepat informasi terkini kasus covid-19 di berbagai wilayah dan untuk memudahkan pembaca dalam mencari informasi tentang kasus covid di berbagai wilayah

Predict Customer Churn With Machine Learning Python

BUSINESS PROBLEM

Perkembangan industri telekomunikasi sangatlah cepat, hal ini dapat dilihat dari perilaku masyarakat yang menggunakan internet dalam berkomunikasi.

Perilaku ini menyebabkan banyaknya perusahaan telekomunikasi dan meningkatnya internet service provider yang dapat menimbulkan persaingan antar provider.

Pelanggan memiliki hak dalam memilih provider yang sesuai dan dapat beralih dari provider sebelumnya yang diartikan sebagai Customer Churn.

Peralihan ini dapat menyebabkan berkurangnya pendapatan bagi perusahaan telekomunikasi sehingga penting untuk ditangani.



DATA UNDERSTANDING



pada kolom churn sebagai variable dependent / kolom target

Data Understanding

terlihat terdapat 20 kolom dengan keterangannya pada prediksi kali ini akan digunakan target pada kolom churn untuk prediksi dan melihat pengaruh feature terhadap target

variable dependent

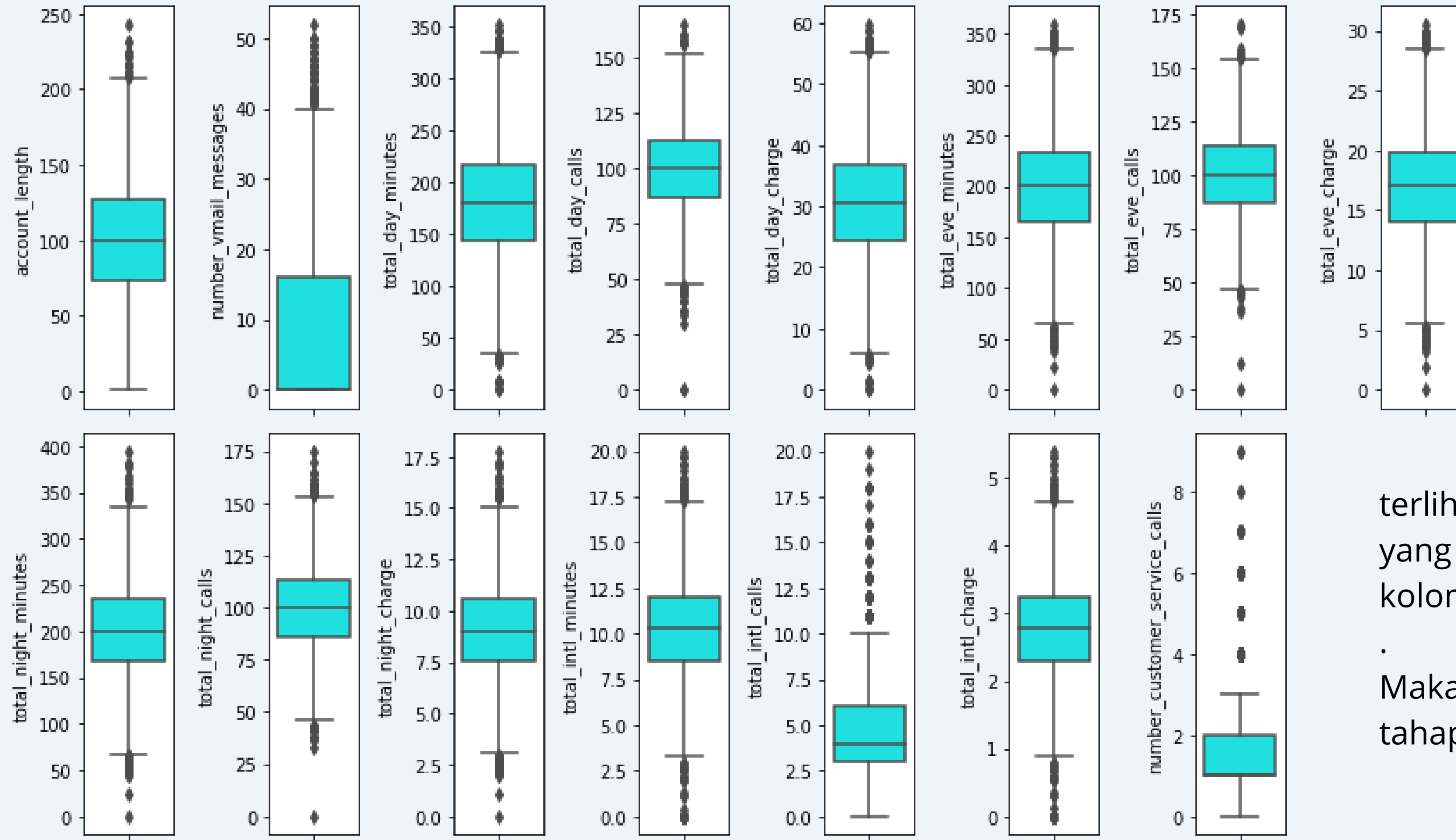
- churn

variable independent

- state
- account_length
- area_code
- international_plan
- voice_mail_plan
- number_vmail_messages
- total_day_minutes
- total_day_calls
- total_day_charge
- total_eve_minutes
- total_eve_calls
- total_eve_charge
- total_night_minutes
- total_night_calls
- total_night_charge
- total_intl_minutes
- total_intl_calls
- total_intl_charge
- number_customer_service_calls

Kolom	Definisi
state	US State
account_length	Total bulan customer menjadi user telco provider
area_code	Kode Area
international_plan	Customer memiliki plan international
voice_mail_plan	Customer memiliki plan voice mail
number_vmail_messages	Total pesan voice mail
total_day_minutes	Total minutes pada day calls
total_day_calls	Total day calls
total_day_charge	Total charge dari day calls
total_eve_minutes	Total menit pada evening call
total_eve_calls	Total evening call
total_eve_charge	Total charge pada evening call
total_night_minutes	Total menit pada night call
total_night_calls	Total night call
total_night_charge	Total charge pada night call
total_intl_minutes	Total menit pada international call
total_intl_calls	Total international call
total_intl_charge	Total charge pada international call
number_customer_service_calls	Total call kepada customer service
churn	Customer churn

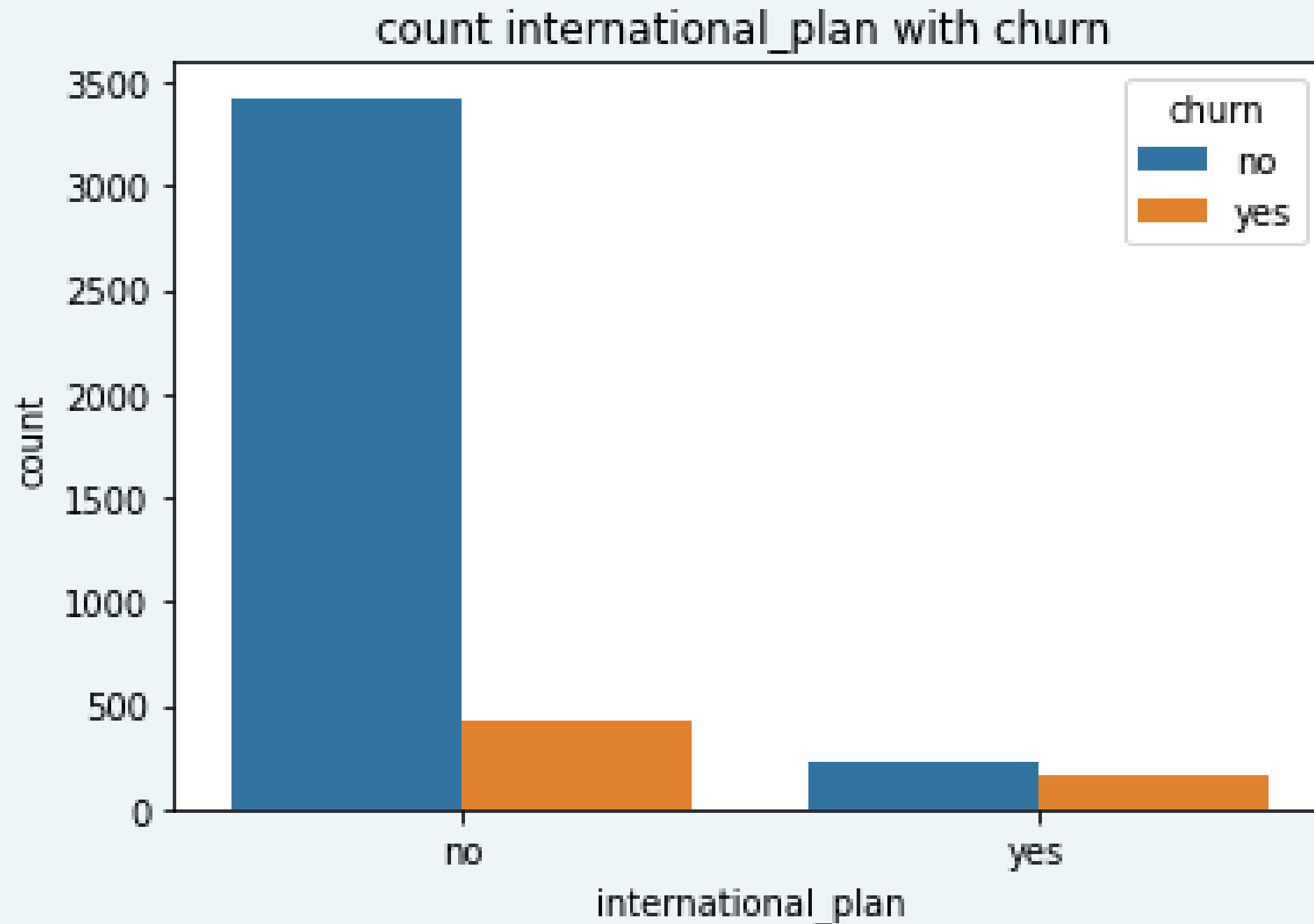
Exploratory Data Analysis



terlihat terdapat 15 kolom numerikal yang memiliki outlier dan terdapat 3 kolom dengan distribusi skew positif

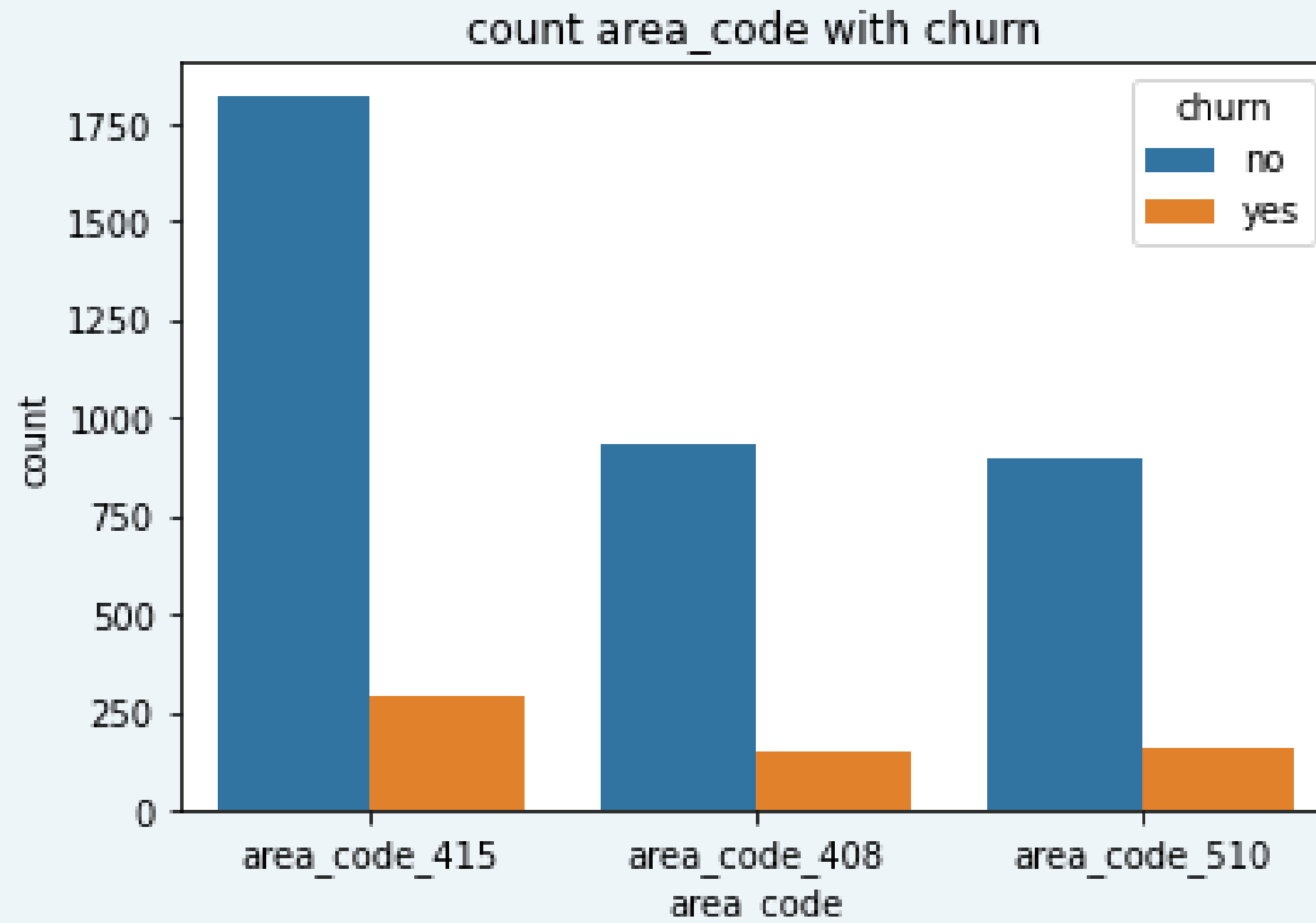
Maka akan dilakukan scaling pada tahapan preprosesing

Exploratory Data Analysis



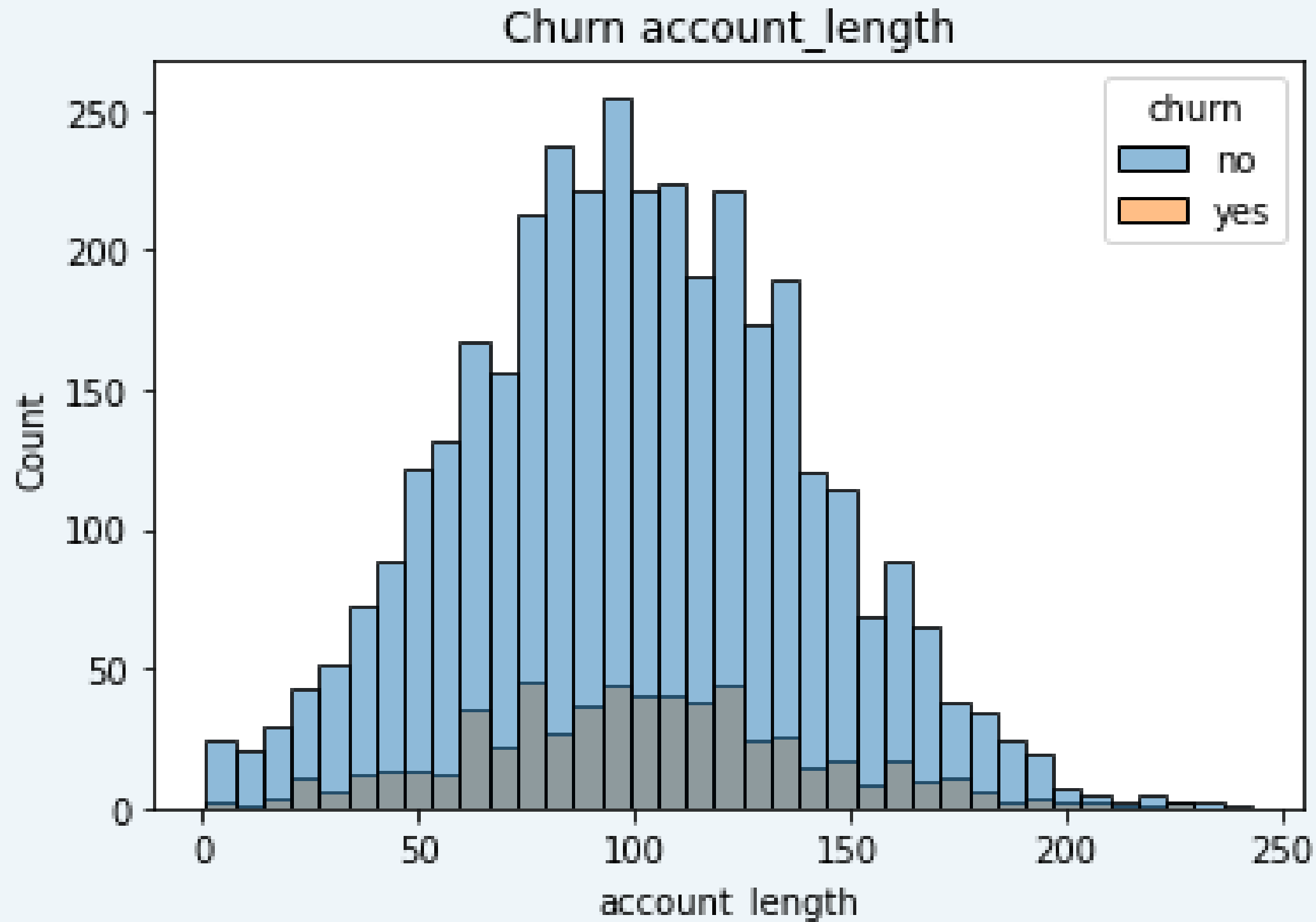
terlihat pada pengguna jaringan internasional memiliki selisih lebih kecil maka untuk pengguna jaringan internasional mempertimbangkan untuk pemberian promo atau melakukan penyesuaian harga agar pelanggan tidak mudah dalam churn

Exploratory Data Analysis



terlihat pada area dengan code 415 lebih banyak customer churn dengan area lainnya maka perlu di adakan campaign untuk meminimalisir customer churn pada area tersebut

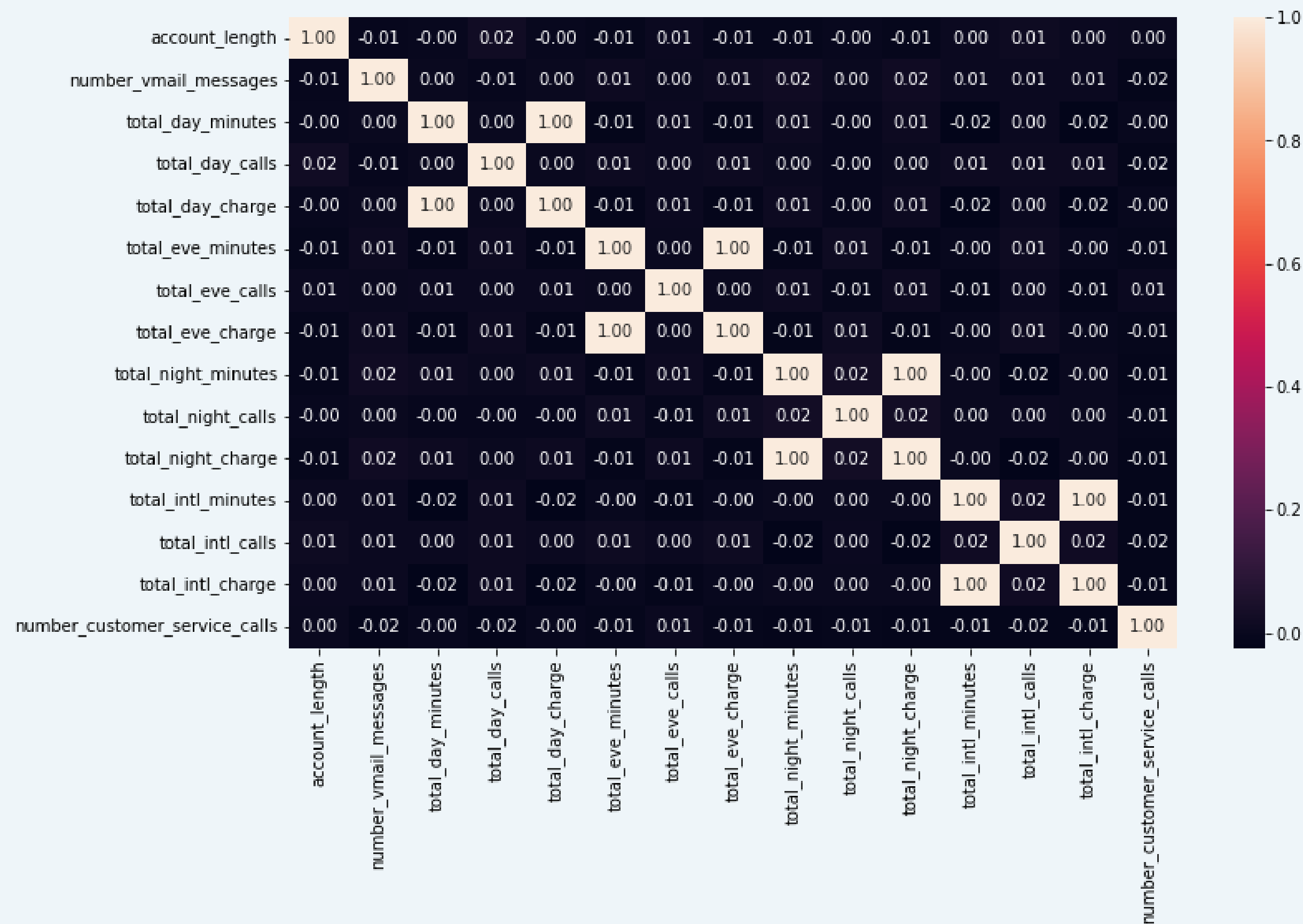
Exploratory Data Analysis



terlihat user mulai melakukan churn pada account_length di atas 50

maka perlu mempertimbangkan kualitas pelayanan bagi pengguna lama

Exploratory Data Analysis



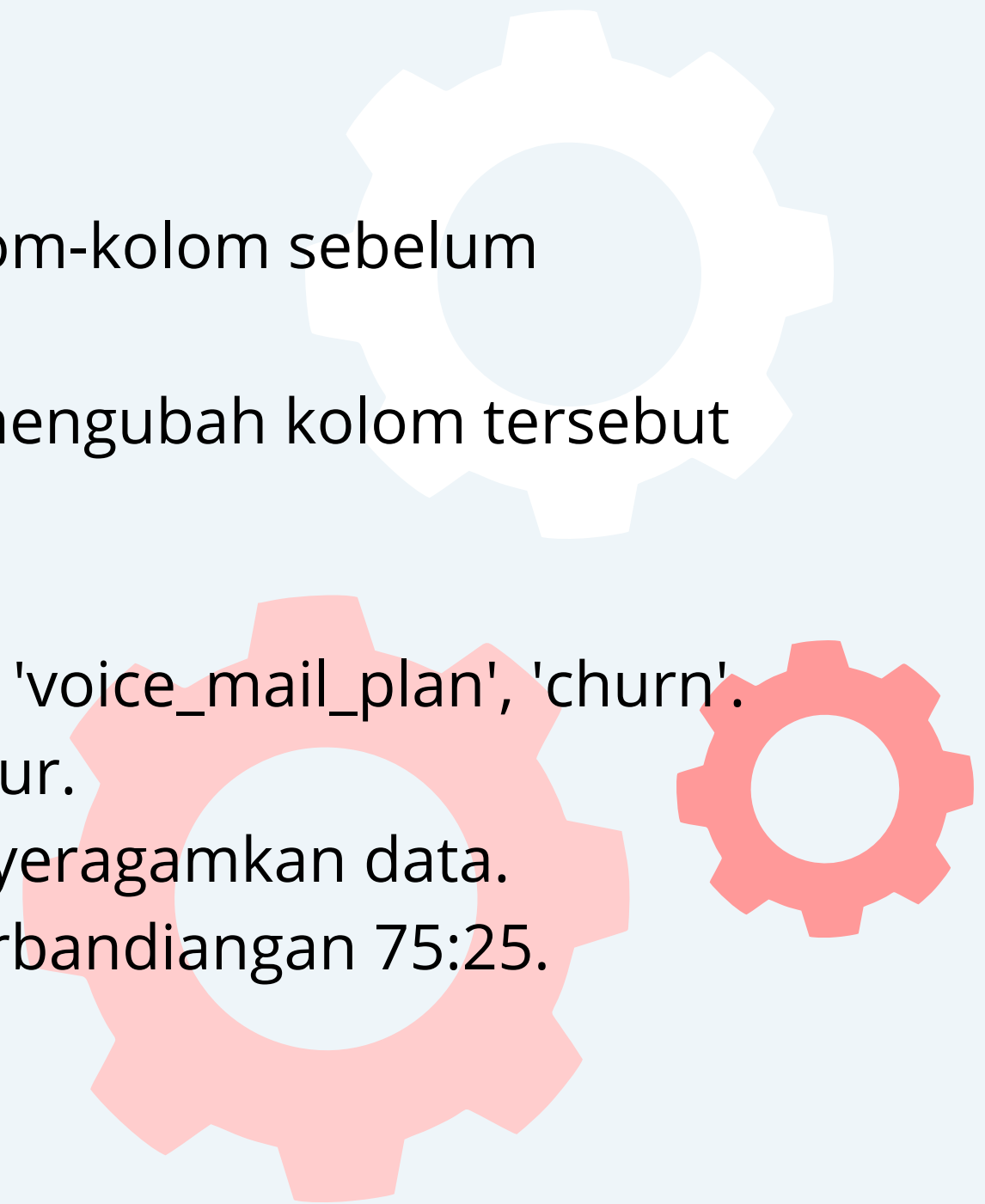
terlihat terdapa feature yang saling berkolerasi maka akan dilakukan drop pada kolom yang berkolerasi dan menetapkan 1 feature

feature yang akan di drop ialah
'total_day_charge',
'total_eve_charge',
'total_night_charge',
'total_intl_charge']\

Feature Engineering

pada tahapan ini akan dilakukan penyesuaian data pada kolom-kolom sebelum dilakukan permodelan diantaranya:

- 1.pada kolom area_code akan diambil di 3 digit angka dan mengubah kolom tersebut menjadi integer.
- 2.melakukan onehot encoding pada kolom state.
- 3.melakukan label encoding pada kolom 'international_plan', 'voice_mail_plan', 'churn'.
- 4.melakukan drop pada data yang memiliki korelasi antar fitur.
- 5.melakukan scaling data dengan standarization untuk menyeragamkan data.
- 6.melakukan pemisahan data train dan data test dengan perbandingan 75:25.
- 7.memisahkan kolom target dengan kolom fitur.



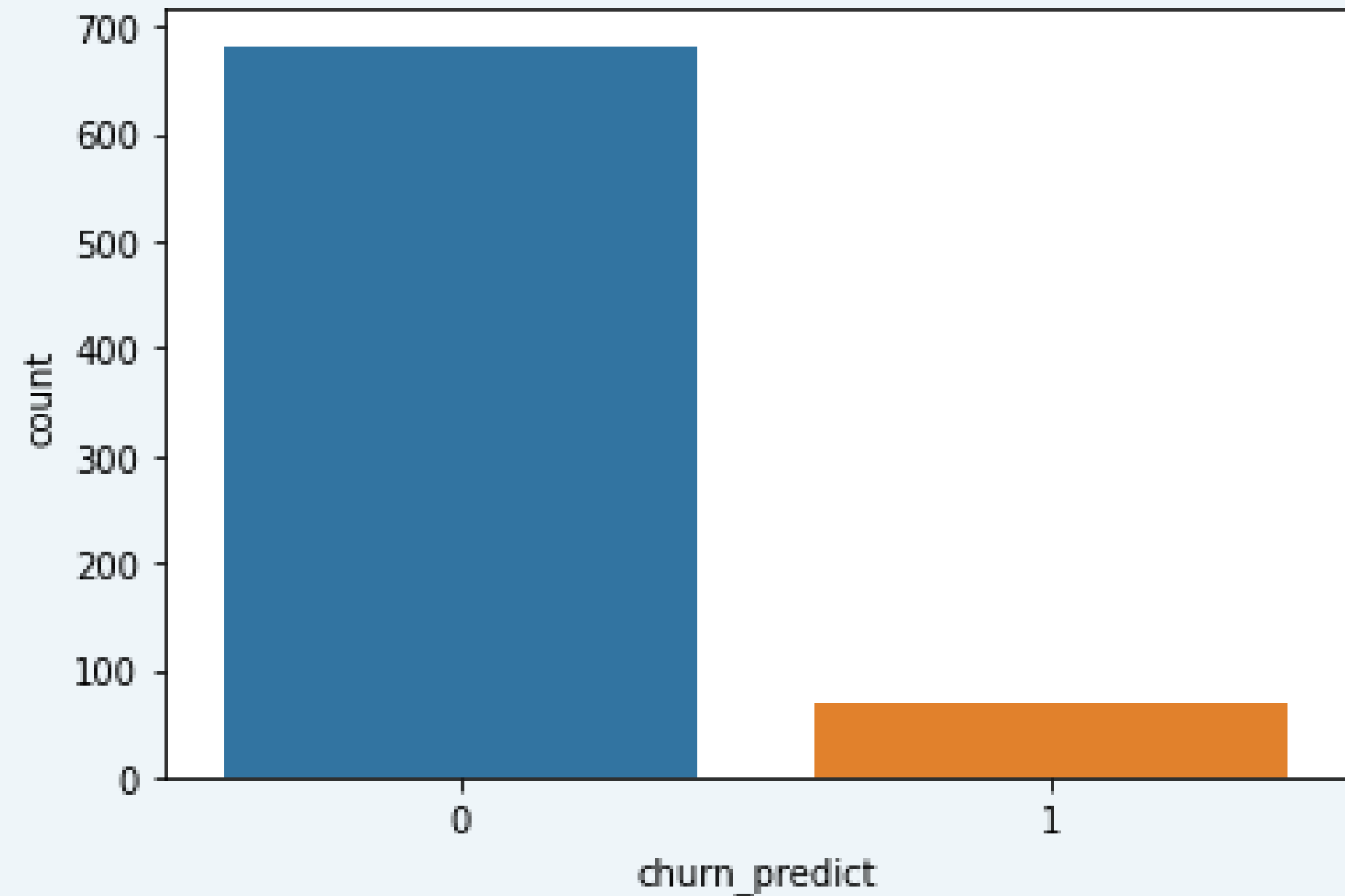
MODELLING RESULT

Model	Accuracy	Precision	Recall	F1-Score	ROC
Random Forest Classifier	0.9397	0.9411	0.6233	0.7499	0.9240
Decision Tree Classifier	0.9238	0.7212	0.7727	0.7460	0.8610

Ditemukan hasil bahwa pada model random forest menjadi best model dengan akurasi model lebih tinggi sebesar 0.9397

Predict With Data Test

Pada data test dilakukan prediksi dimana customer yang melakukan churn sebanyak 68 orang yang diprediksi dan 682 customer yang tidak diprediksi churn



Conclusion

- kasus bisnis yang memerlukan tinggi akurasi model dengan random forest menjadi pilihan terbaik dengan beberapa catatan untuk melakukan peningkatan performa pada nilai recall
- data yang tersedia memiliki ketidakseimbangan pada data target maka untuk melakukan improvisasi perlu adanya data yang seimbang
- penambahan fitur label pada masing-masing customer akan lebih baik untuk menentukan target pemasaran yang lebih spesifik

Appendix

Google Colab

<https://colab.research.google.com/drive/1Xdj26oPc2wOQaHAODnzzcTXhvC3E1ZLX?usp=sharing>

Dashboard Google Studio

<https://datastudio.google.com/reporting/c5480907-ca44-43cb-8e0b-7cb82a1ece66>

Thank You