

CUSTOMER CHURN PREDICTION

Russia Group



SWANDEWI



DAFFA



NABILA



DERAL



FATHI FARHAT



NINDA PUTRI



Problem

Banyaknya user yang tidak aktif melakukan transaksi kembali pada ecommerce Fashion Campus membuat churn rate user **meningkat**.

Memprediksi **probabilitas** customer churn pada bulan selanjutnya

Memberikan **rekomendasi** untuk mencegah dan menanggulangi customer churn

Definisi Customer Churn :

- Tidak ada transaksi pada periode waktu tertentu dan atau
- Tidak menunjukkan aktivitas menggunakan aplikasi Fashion Campus pada periode waktu tertentu



Process



DATA CLEANING



MERGE DATA



**FEATURE
ENGINEERING**

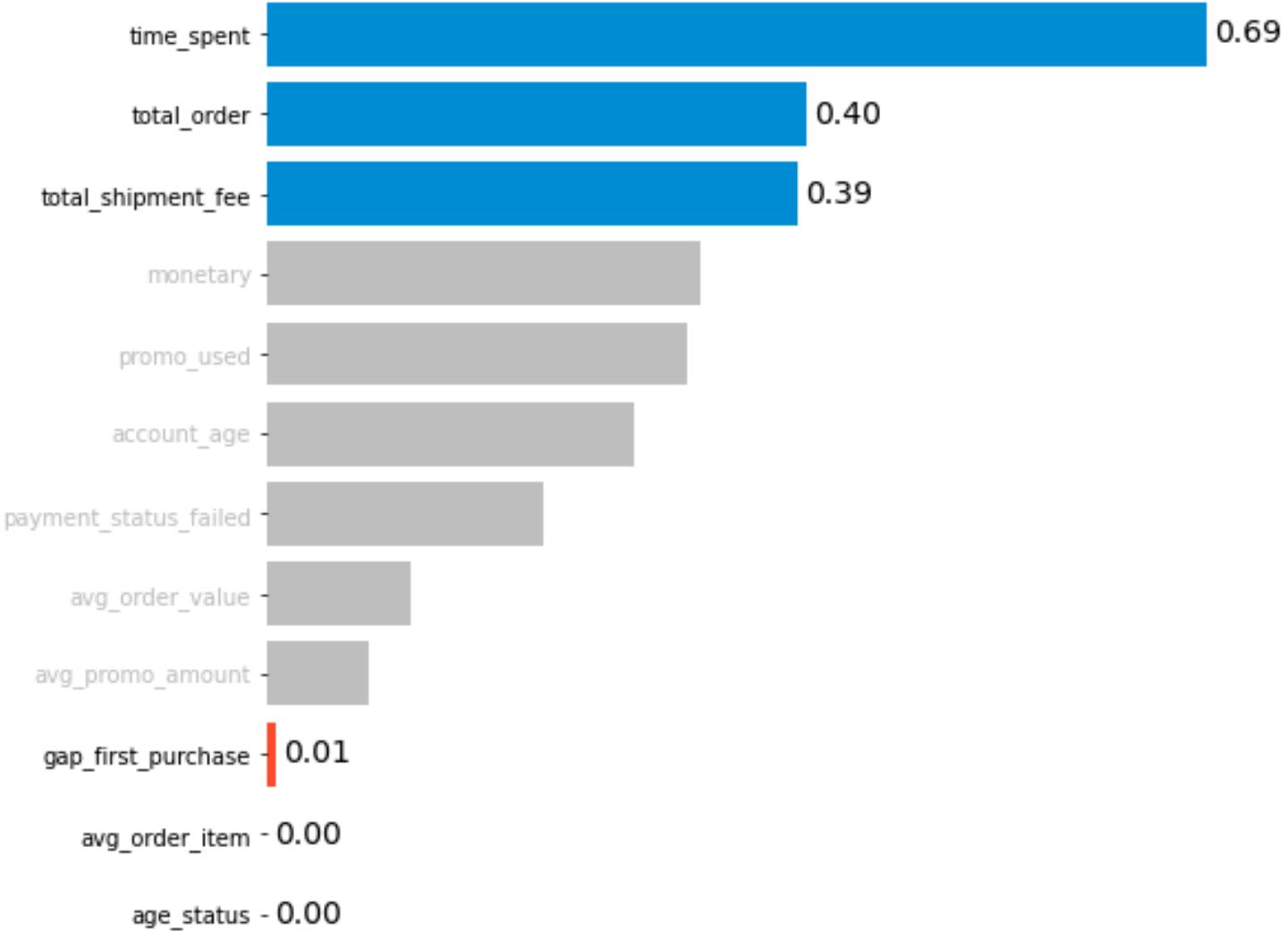


PREPROCESSING



MODELING

Feature Selection



Kami melakukan feature selection menggunakan **Pearson Correlation**. Didapatkan 3 fitur yang memiliki korelasi paling rendah yaitu, '**gap_first_purchase**', '**avg_order_value**', '**age_status**'. Untuk itu, kami drop feature tersebut.

Oversampling and Scaling

Model	Accuracy	Precision	Recall	F1 Score
Gradient Boosting Classifier	0.944463	0.915305	0.91136	0.913315
Hist Gradient Boosting	0.945335	0.913907	0.916884	0.915385
Random Forest	0.944664	0.913191	0.915344	0.914262
Light GBM	0.947616	0.917182	0.920800	0.918975
Decission Tree	0.922798	0.881092	0.878142	0.879605

Before

Model	Accuracy	Precision	Recall	F1 Score
Gradient Boosting Classifier	0.959845	0.977661	0.941196	0.959082
Hist Gradient Boosting	0.964592	0.975170	0.953461	0.964193
Random Forest	0.964970	0.974374	0.955057	0.964619
Light GBM	0.964718	0.975830	0.953041	0.964301
Decission Tree	0.948799	0.950730	0.946657	0.948689

After



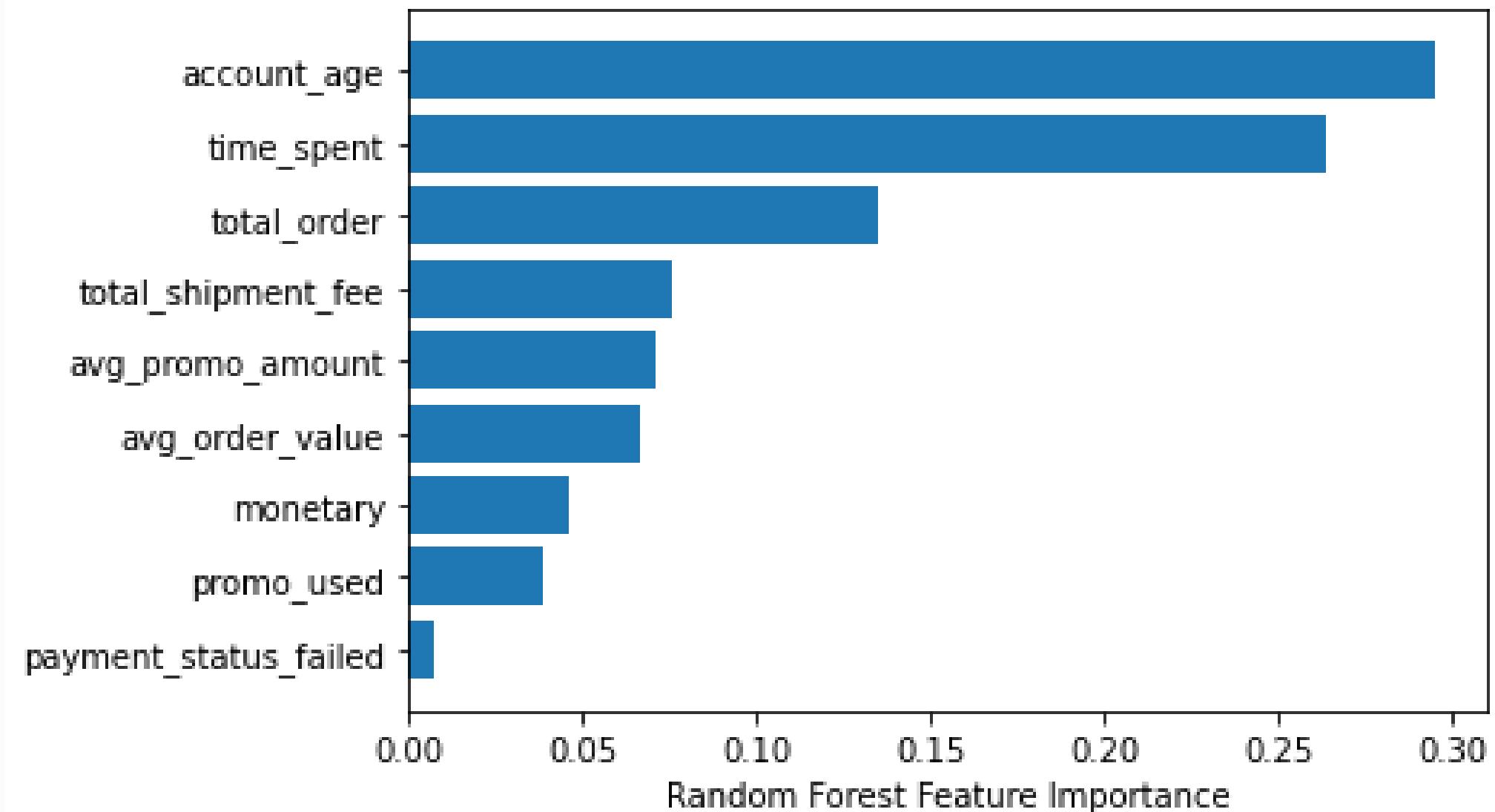
Hyperparameter Tuning

Model	Accuracy	Precision	Recall	F1-Score
Random Forest Baseline	0.964970	0.974374	0.955057	0.964619
Random Forest Tuning	0.965767	0.975392	0.955645	0.965417
Light GBM Baseline	0.964718	0.975830	0.953041	0.964301
Light GBM Tuning	0.965179	0.974466	0.955393	0.964835

Feature Importance Random Forest

Kami mencoba menghilangkan fitur yang memiliki perbedaan signifikan, seperti '**account_age**' dan '**time_spent**' , alhasil akurasi model turun hingga **0.894**

Text(0.5, 0, "Random Forest Feature Importance")

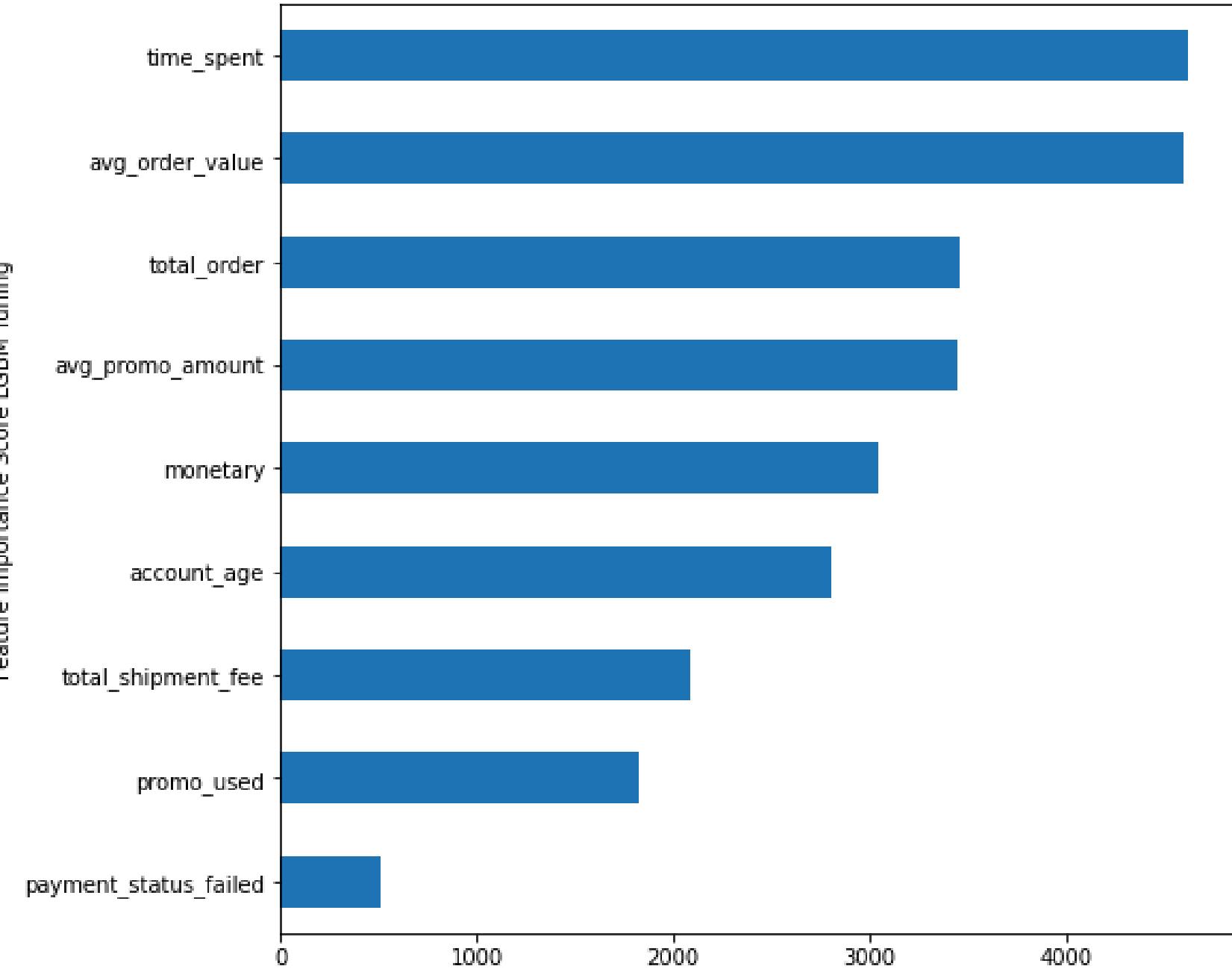


Feature Importance Light GBM

Kami mencoba menghilangkan fitur yang memiliki perbedaan signifikan, seperti '**time_spent**' dan '**avg_order_value**', alhasil akurasi model turun hingga **0.9571**

Text(0, 0.5, 'Feature Importance Score LGBM Tuning')

Feature Importance LGBM Tuning



Validasi Akurasi Light GBM

Dataset	Testing	Validation	Time Window 1	Time Window 2	Time Window 3	Time Window 4
Akurasi	0.961	0.908	0.858	0.860	0.881	0.898

Setelah mempertimbangkan, kami memilih Light GBM sebagai model terbaik & paling stabil. Lalu, kami melakukan validasi akhir akurasi Light GBM pada data validasi yang fresh. Didapat akurasi dengan range **85.8%** hingga **96.1%**.

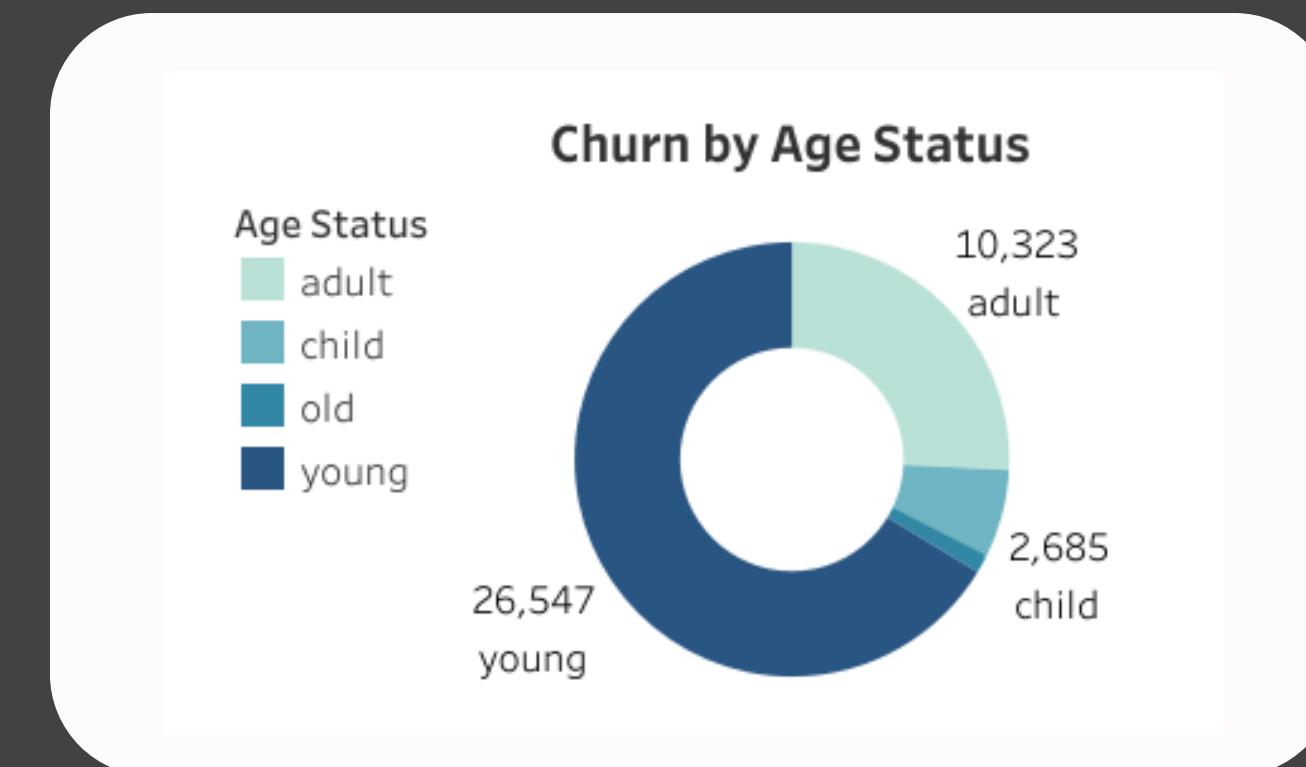
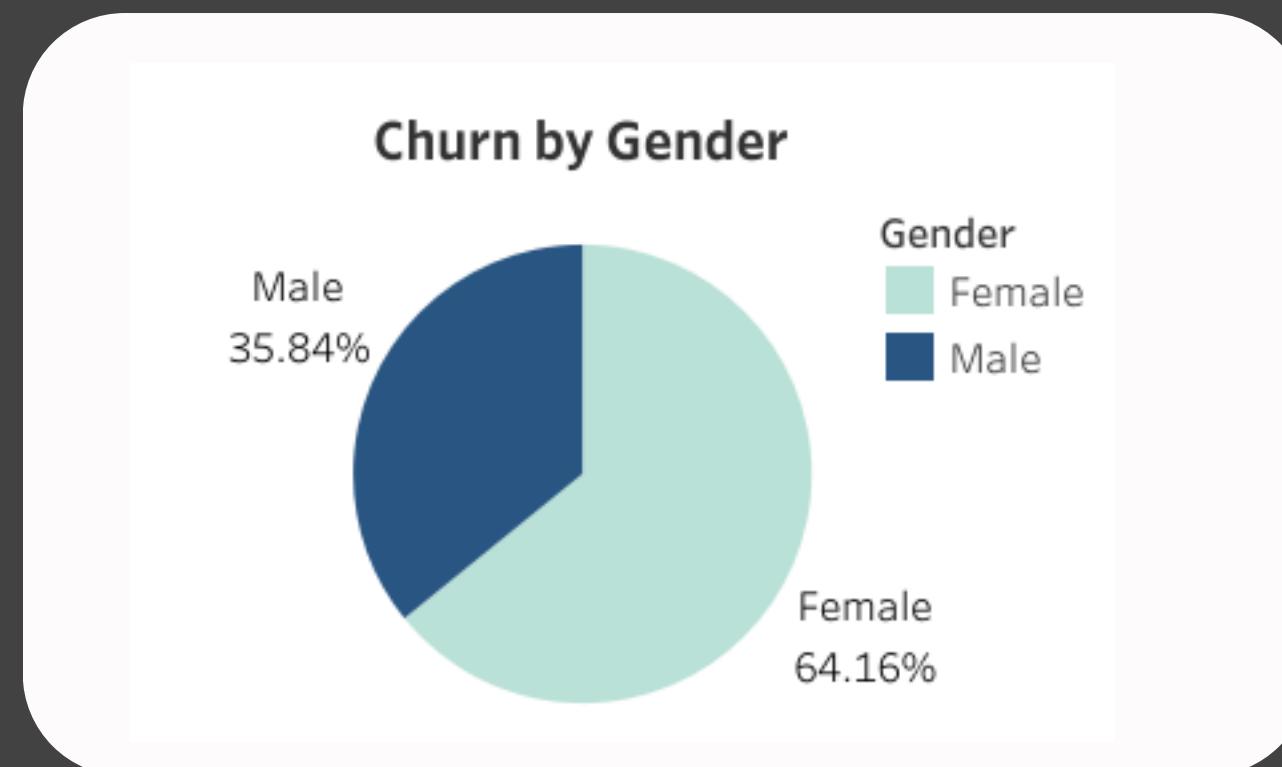


Result (1)



79.8%
CHURN

20.2% 
TIDAK CHURN

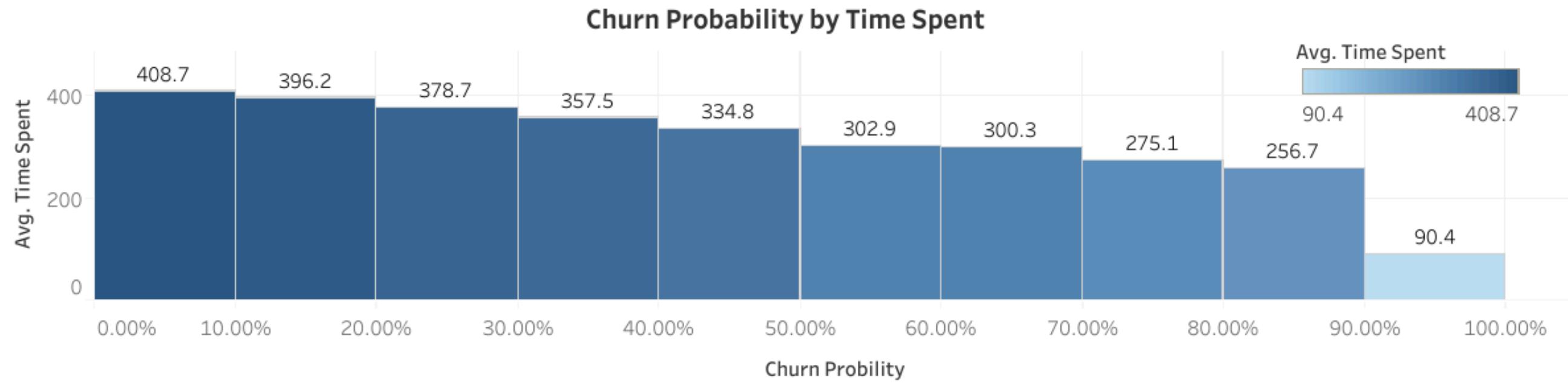


GENERAL DATA VISUALIZATION

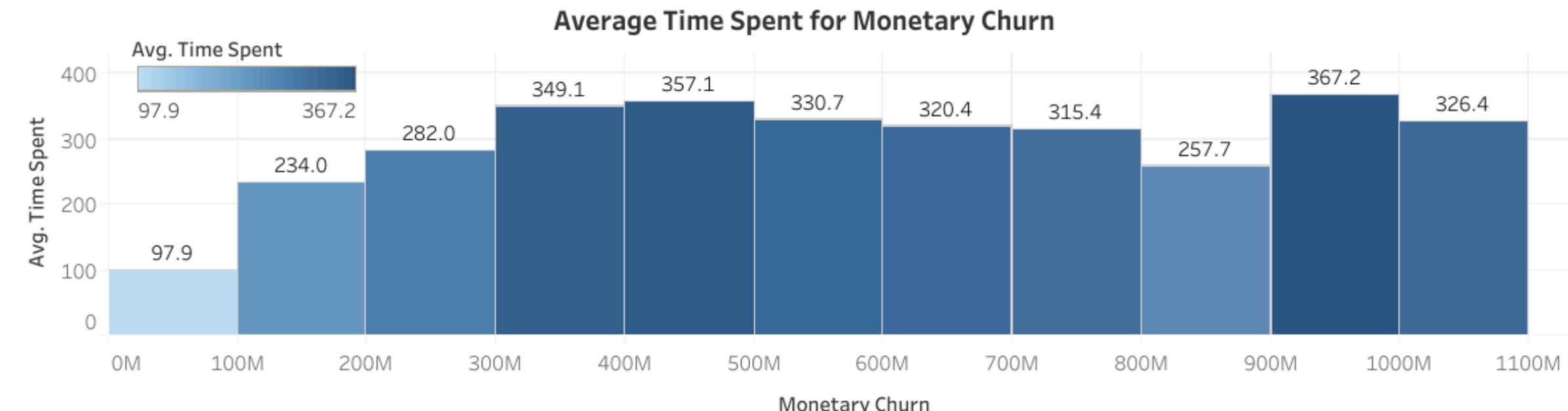




Result (2)

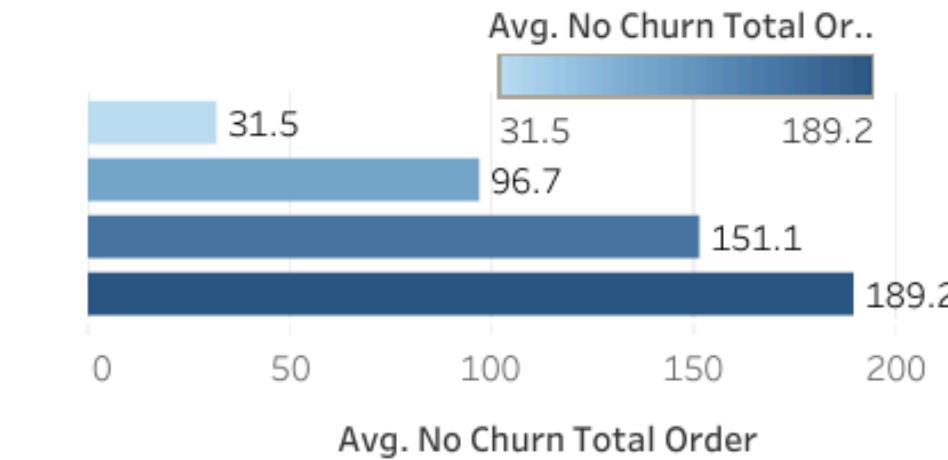
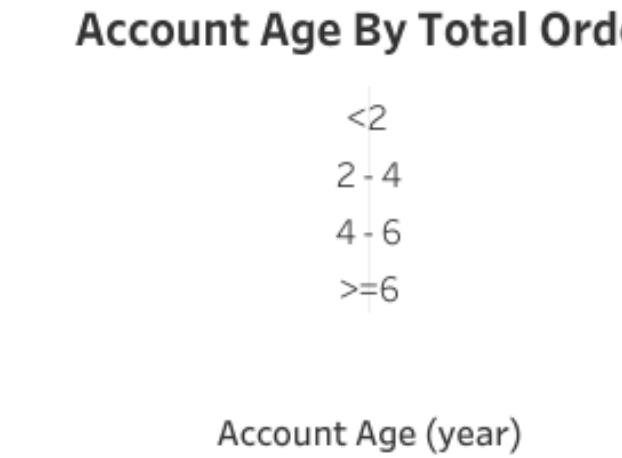
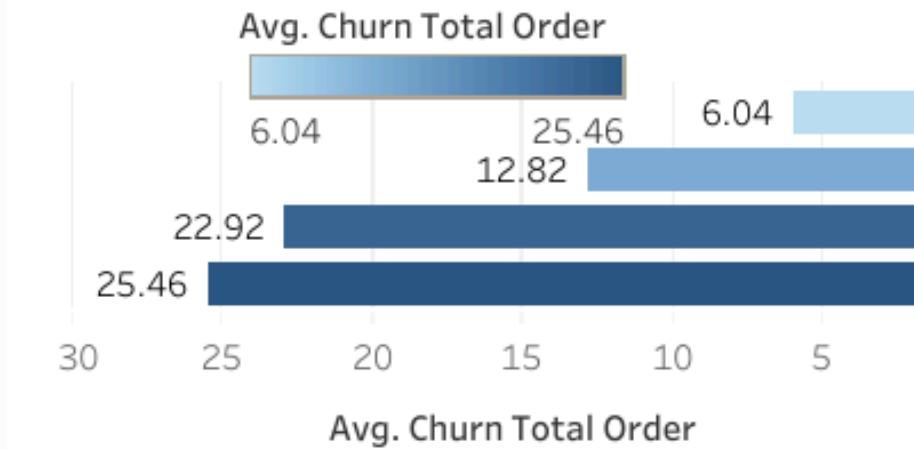


CHURN PROBABILITY BY TIME SPENT

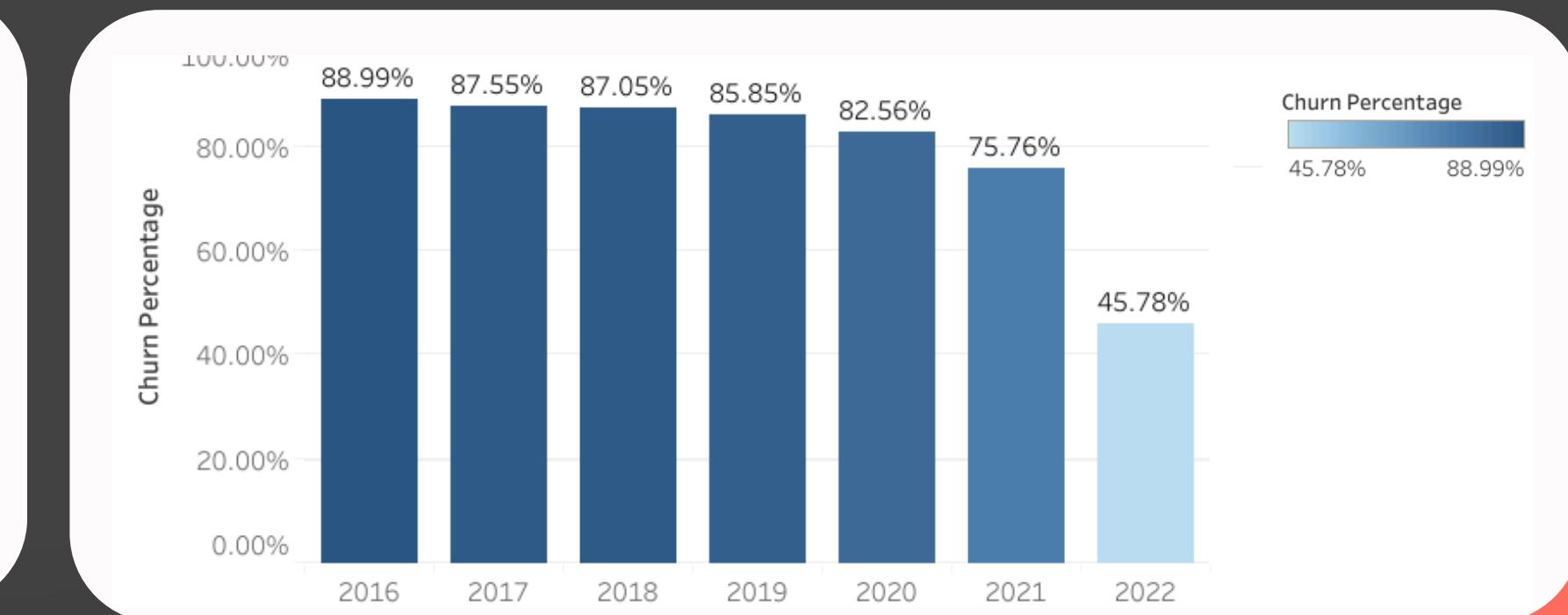
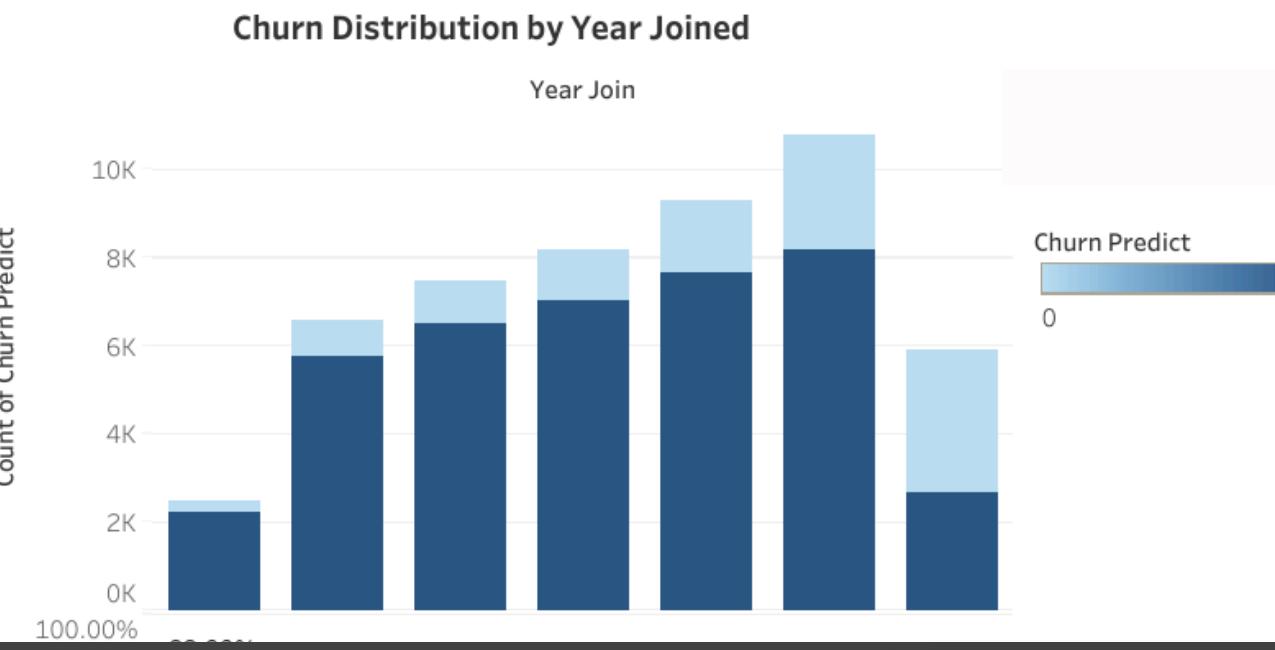


RELATION BETWEEN TIME SPENT & MONETARY CHURN

Result (3)



ACCOUNT AGE BY TOTAL ORDER



CHURN DISTRIBUTION BY YEAR JOINED



Recommendation

Gender & Age :

- Fokus pada segmen female dan young

Time Spent Using App :

- Meningkatkan UI/UX dari aplikasi (platform) agar aplikasi lebih user friendly sehingga user nyaman dan berujung pada peningkatan monetary customer dan profit perusahaan
- Menyediakan informasi berkaitan dengan fashion dengan berkolaborasi dengan platform lain seperti Female Daily (API)

Rate Old Customer Churn :

- Menyediakan layanan keluhan customer
- Mengadakan program loyalty untuk customer berdasarkan umur akun dan monetary



**"It's better to maintain your old customer
than always try to get new customer"**

