

Final Project

E-COMMERCE CHURN PREDICTION

Bootcamp Data Science
Rakamin Academy
Batch 22





MARVEL STUDIOS

AVENGERS (KELOMPOK 2)

Mentor: Mirza Fahmi



Adi Herianto R. (Leader)



Ryana Tammi Putri



Dian Maulida



Susi Ninik Mahudi



Virginia Wulansari



Andini Febriati



TABLE OF CONTENTS



Project Background



**EDA
(Exploratory Data Analysis)**



Data Pre-Processing



Modeling & Evaluation



**Business Insight &
Recommendation**

PROJECT BACKGROUND

ROLE



Avengers merupakan tim Data Scientist di sebuah e-commerce bernama Marvel Shop yang bertugas untuk mencari solusi dari masalah yang ada dan memberikan rekomendasi bisnis menggunakan dataset yang tersedia.

Dataset: [Churn Prediction](#)



PROBLEM IN MARVEL SHOP??



946

Customer Churn

16,8%

Churn Rate

Customer Churn rate =

Jumlah pelanggan yang berhenti berlangganan

x 100

Jumlah pelanggan dalam satu periode

Dilansir dari artikel gorgias.com, menurut Recurly Research, pada perusahaan dengan bisnis consumer good angka churn rate yang standar itu ada di angka **8,56%**



WHY??



CHURN RATE IS AN IMPORTANT PART OF GROWING A SUSTAINABLE ECOMMERCE BUSINESS.

- Dilansir dari artikel gorgias.com, menurut VWO, **43%** keuntungan untuk e-commerce berasal dari pelanggan yang dipertahankan (tidak churn).
- Terlebih lagi, menurut Outbound Engine, “memperoleh pelanggan baru dapat menghabiskan biaya lima kali lebih banyak daripada mempertahankan pelanggan yang sudah ada.”

GOAL



Membuat model machine learning yang dapat memprediksi customer yang akan churn.

OBJECTIVE



- Mengurangi tingkat churn customer dengan target churn rate < 16,8%
- Memberikan solusi dan rekomendasi bisnis berdasarkan insight data naik

BUSINESS METRICS



- Churn Rate
- CRC (Customer Retention Cost)

EXPLORATORY DATA ANALYSIS

DATA OVERVIEW



5630 Baris

20 Kolom

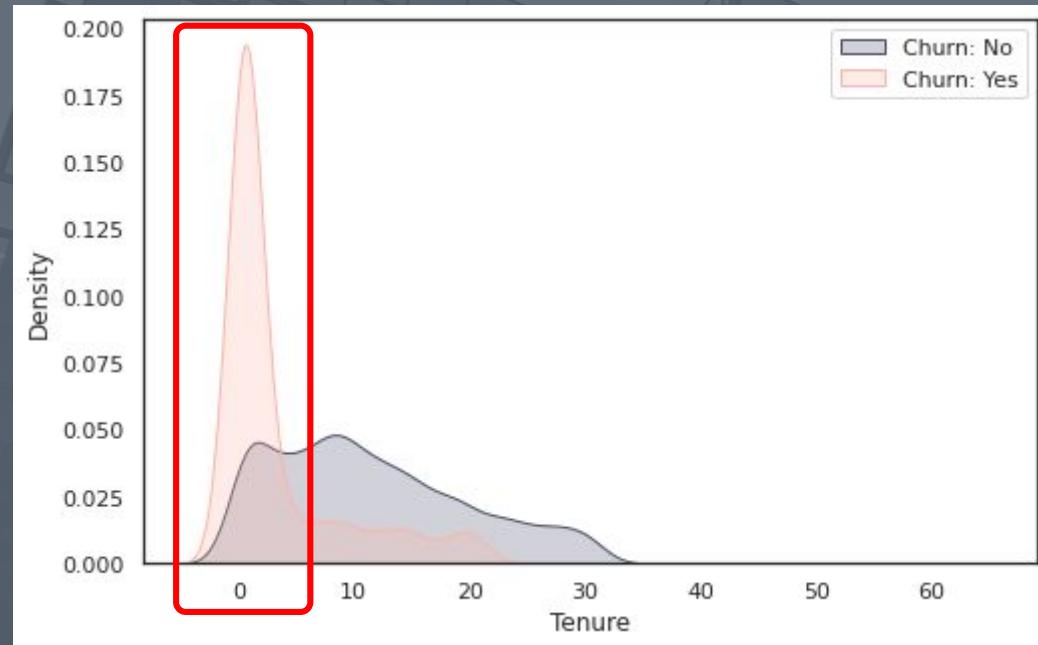
7 Kolom mengandung missing values

0 Data duplikat



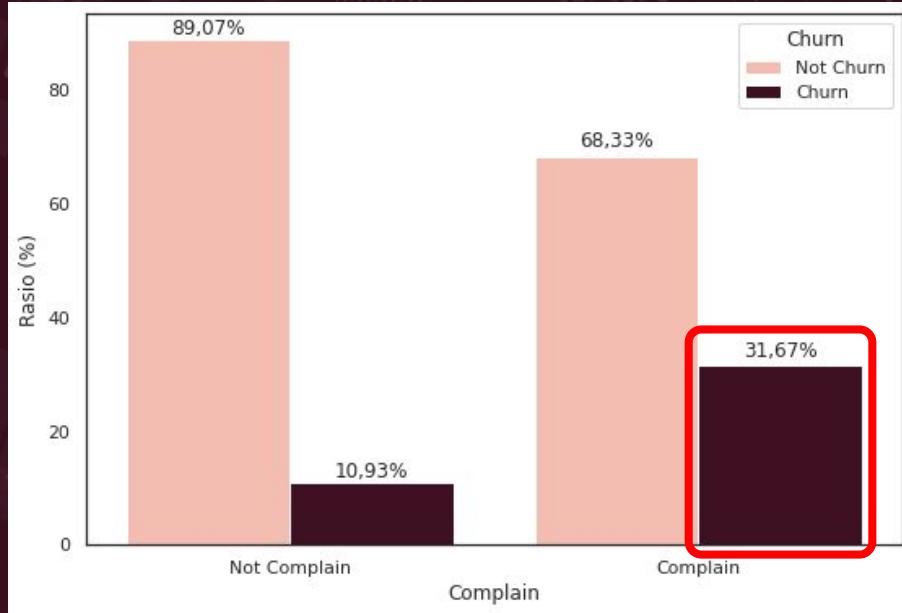
Tipe Data	Nama Feature	Deskripsi
Target Feature	Churn	Churn flag
Numerical	CustomerID	ID unik dari pelanggan
	Tenure	Berapa lama customer menggunakan layanan (loyalitas) (bulan)
	CityTier	Tingkatan kota
	WarehouseToHome	Jarak gudang (warehouse) e-commercenya ke rumah pelanggan
	HourSpendOnApp	Waktu yang dihabiskan untuk menjelajah di aplikasi (dalam jam)
	NumberOfDeviceRegistered	Jumlah total perangkat yang terdaftar
	SatisfactionScore	Tingkat kepuasan pelanggan
	DaySinceLastOrder	Jarak hari ini dengan hari terakhir/rentang waktu customer order pesanan
	NumberOfAddress	Jumlah alamat yang ditambahkan customer
	OrderAmountHikeFromlastYear	Kenaikan Jumlah Pesanan Dari Tahun Lalu
	CouponUsed	Jumlah total Kupon yang telah digunakan bulan lalu
	OrderCount	Jumlah orderan bulan lalu
	CashbackAmount	Jumlah cashback bulan lalu
Categorical	PreferredLoginDevice	Perangkat login yang digunakan pelanggan
	PreferredPaymentMode	Metode pembayaran yang digunakan oleh pelanggan
	Gender	Jenis kelamin pelanggan
	PreferredOrderCat	Kategori produk yang dibeli bulan lalu
	MaritalStatus	Status pernikahan pelanggan
	Complain	Complain atau tidaknya pelanggan bulan lalu

CHURN BY TENURE

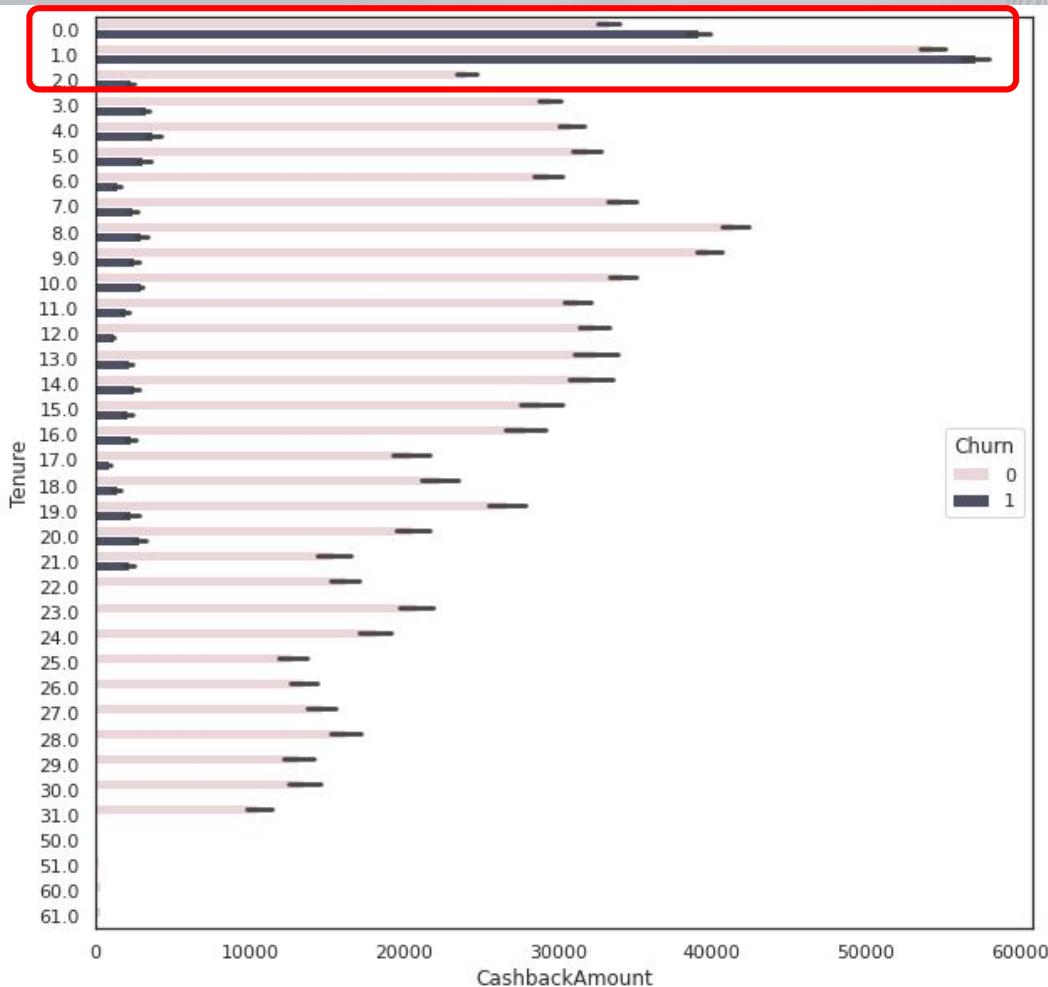


Churn cenderung terjadi pada customer dengan loyalitas yang rendah atau belum terlalu lama menggunakan layanan e-commerce.

CHURN BY COMPLAIN



31.67% dari customer yang churn merupakan customer yang complain pada bulan lalu.

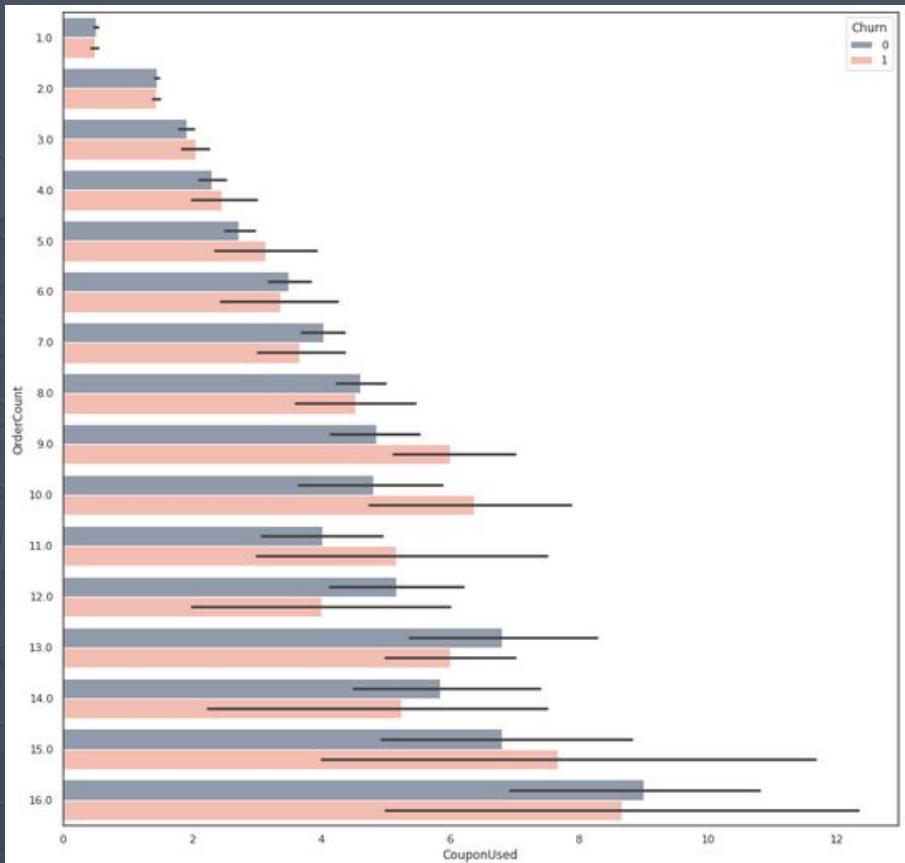


TENURE VS CASHBACK AMOUNT

Customer dengan Tenure yang rendah cenderung paling banyak churn sekaligus banyak menggunakan cashback. Kemungkinan besar customer tersebut merupakan customer baru yang hanya ingin menikmati cashback yang disediakan oleh e-commerce saja.

ORDER COUNT VS COUPON USED

Semakin banyak kupon yang diberikan, semakin banyak pula jumlah pesanan/order customer. Jadi, jumlah order produk itu sangat ditentukan oleh kupon yang kita berikan.



DATA PRE-PROCESSING



HANDLING MISSING VALUES

- imputasi data dengan mean & median

NO DUPLICATED DATA

HANDLING OUTLIERS

- metode Z-Score

FEATURE TRANSFORMATION

- normalization
- standardization

FEATURE ENCODING

- label encoding
- one hot encoding



FEATURE SELECTION

- drop irrelevant feature

FEATURE EXTRACTION

SPLIT DATA 70:30

HANDLING IMBALANCE DATA

- metode SMOTE 1:1





MODELING & EVALUATION

METRIC EVALUATION



Main Metrics: **F1-Score**

Supporting Metrics: **Recall, Precision, Accuracy, AUC**

Tujuan:

- Menghindari False Negative (Customer yang diprediksi tidak Churn, aktualnya Churn).
- Menghindari False Positive (Customer yang diprediksi Churn aktualnya tidak Churn)

Akibat :

- FN besar -> Banyak customer yang churn, akan banyak cost yang terbuang karena acquisition cost 5x lebih besar dari retention cost
- FP besar -> cost yang dialirkan salah sasaran sehingga akan berdampak pada keseimbangan biaya operasional perusahaan.



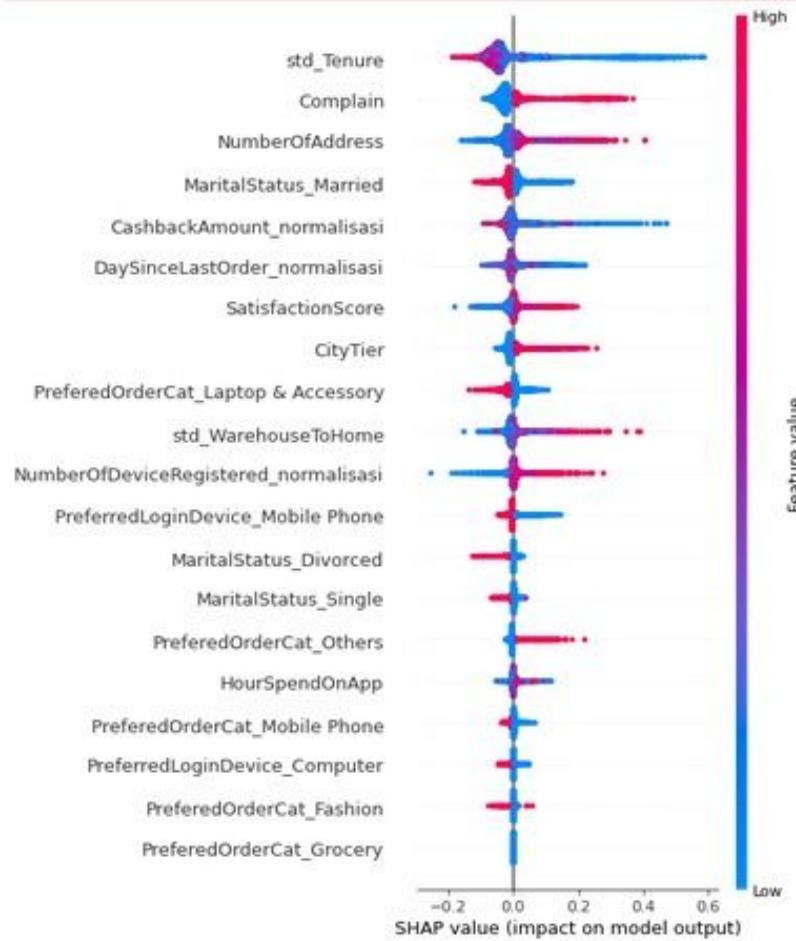
MODEL EVALUATION

Model Evaluation	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	0.77	0.41	0.88	0.56	0.81
K-Nearest Neighbor	0.87	0.57	0.84	0.68	0.86
Decision Tree	0.88	0.64	0.65	0.64	0.79
Random Forest	0.95	0.87	0.84	0.86	0.91
AdaBoost	0.90	0.66	0.81	0.73	0.86
XGBoost	0.96	0.89	0.86	0.87	0.92

METRIC EVALUATION

	Predicted to Not Churn	Predicted to Churn
Actually Not Churn	1310 TN	38 FP
Actually Churn	29 FN	228 TP

FEATURE IMPORTANCE





**BUSINESS INSIGHT
& RECOMMENDATION**

BEFORE MODELING

Data pelanggan
Marvel Shop



Semua pelanggan diberi perlakuan yang sama yang bertujuan agar mereka tidak churn (retention)

BIAYA RETENSI SANGAT TINGGI

AFTER MODELING

Data pelanggan
Marvel Shop



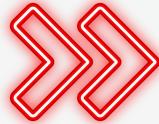
Melakukan prediksi untuk mengetahui pelanggan yang akan churn



Memberikan perlakuan khusus pada pelanggan yang diprediksi akan churn

**BIAYA RETENSI MENJADI LEBIH RENDAH,
SEHINGGA BISA DIALIHKAN UNTUK
KEPENTINGAN LAINNYA**

IMPLEMENTATION RESULT



- Membuat fitur baru, fitur gaining point/gamifikasi
- Memperbaiki penanganan complain
- Memperbaiki strategi marketing → diskon, gratis ongkir, kupon, membership, flash sale, dll.



- Tenure meningkat
- Complain turun < 17%
- HourSpendOnApp naik 82%



SIMULASI KETIKA FITUR GAINING POINT/GAMIFIKASI BERLANGSUNG SELAMA 6 BULAN



HOUR SPEND ON APP

BEFORE

15.656,8
Hours

AFTER

28.496,8
Hours

NAIK 82%



SIMULASI SETELAH 6 BULAN



BEFORE

9,8
%

TENURE



AFTER

10,09
%



SIMULASI SETELAH 6 BULAN

BEFORE

16.8%

Churn Rate

\$10Mio

Retention Cost/Year



AFTER

16,4%

Churn Rate

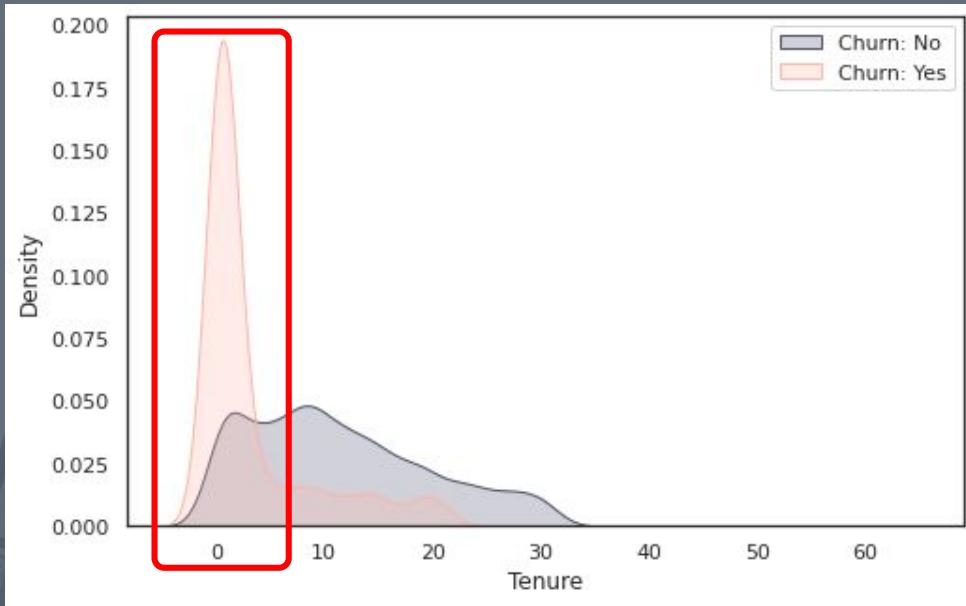
\$1.64Mio

Retention Cost/Year

Assumption:

- total customer = 100.000
- Average CRC = \$100/year





LOYALITAS RENDAH

(Tenure < 9 bulan)



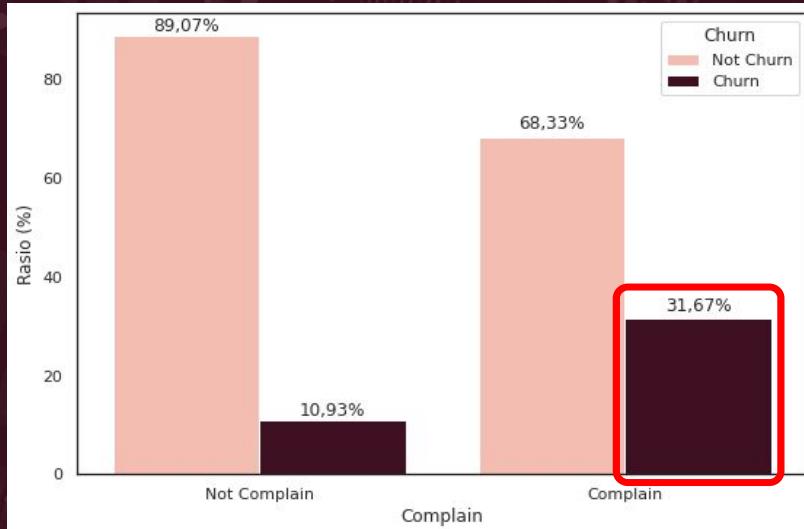
cenderung

CHURN



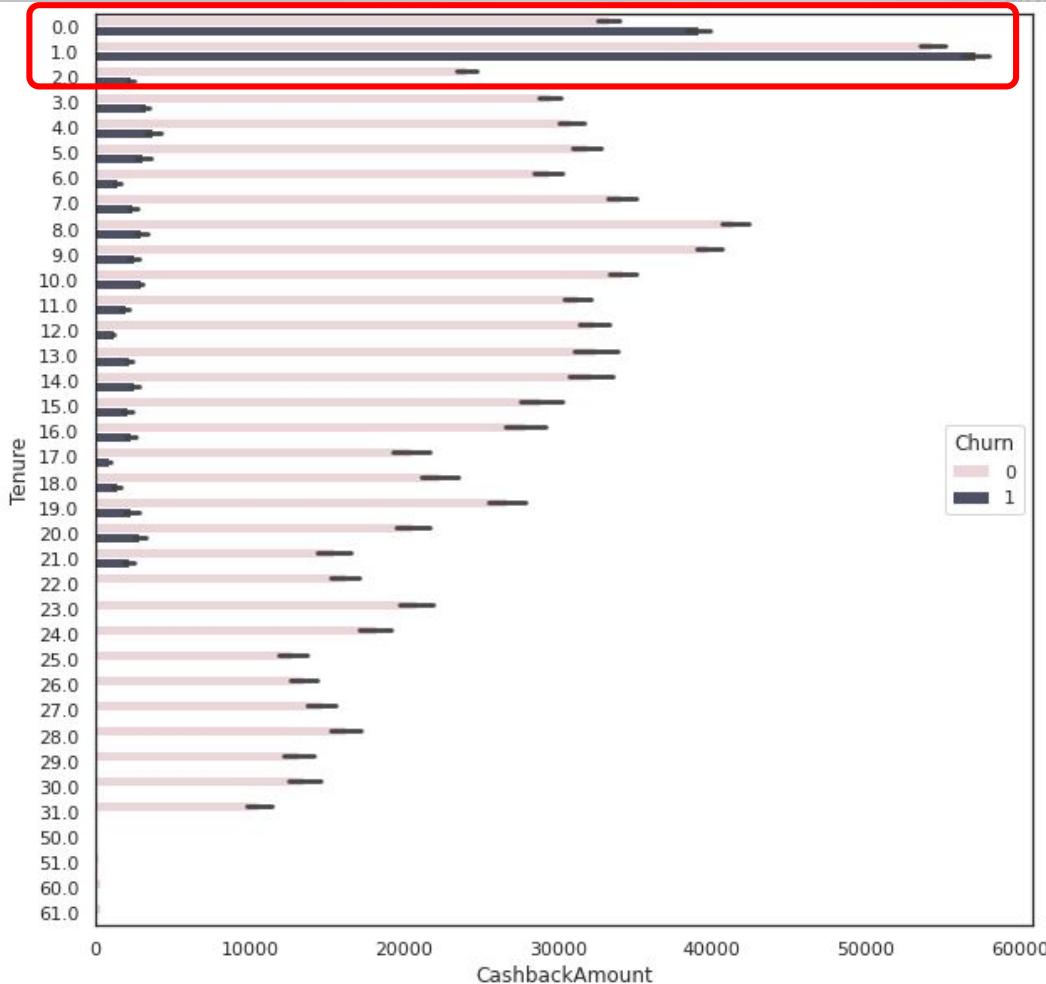
- Membuat fitur baru dalam layanan e-commerce seperti game mengumpulkan coin/poin tersebut yang nantinya dapat ditukar dengan reward berupa voucher diskon, gratis ongkir, dan lain sebagainya.

31,7% customer yang **CHURN**
merupakan
customer yang **COMPLAIN**



Membuat perbaikan dalam penanganan complain:

- membuat sistem bot untuk mengarahkan customer secara otomatis dan realtime
- membuat fast call service untuk penanganan masalah yang kompleks dengan cepat sekaligus sebagai upaya untuk membuat komunikasi dengan customer.
- Target resolved complain : min 50% customer perbulan



Pada tenure < 2 bulan, penggunaan

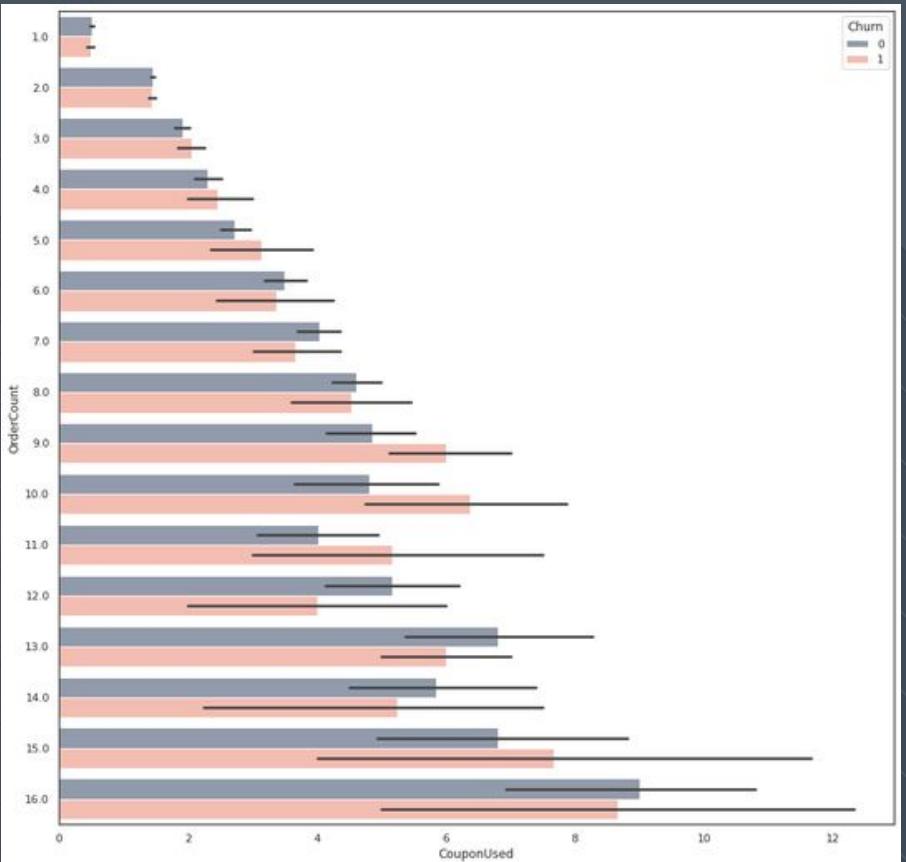
CASHBACK

SANGAT TINGGI

dan customer cenderung **CHURN**



- Membuat klasifikasi user berdasarkan tenure agar lebih mudah menargetkan pemberian cashback terhadap customer
- Jika ingin melakukan akuisisi calon customer, bisa menggunakan strategi marketing lain yang memungkinkan new user benar-benar betah dengan layanan kita dan tidak mengincar promo saja.



Semakin banyak
KUPON yang digunakan,
semakin banyak pula
JUMLAH ORDER



Jika ingin meningkatkan jumlah order, maka kita bisa lebih sering memberikan kupon kepada customer dengan mempertimbangkan ROI, BEP, dll. atau kupon diberikan kepada pelanggan berdasarkan cluster RFM.

THANK YOU