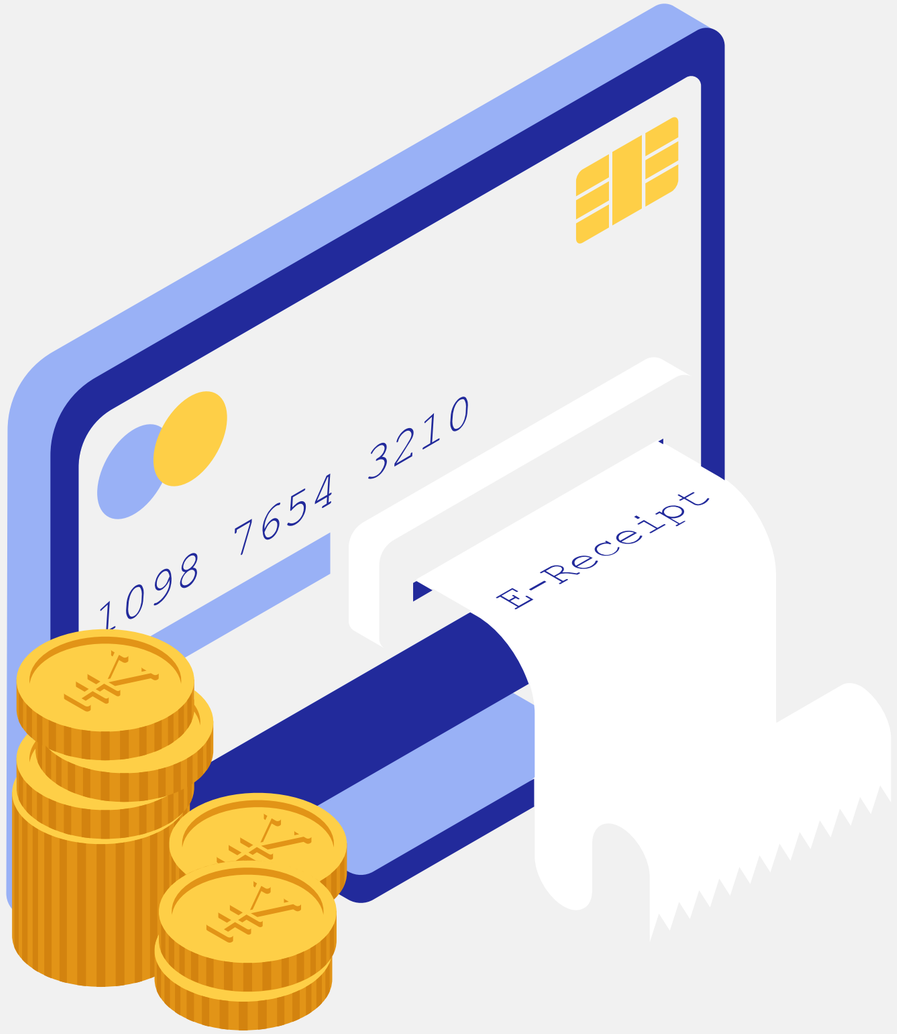


Purwadhika Digital Technology School

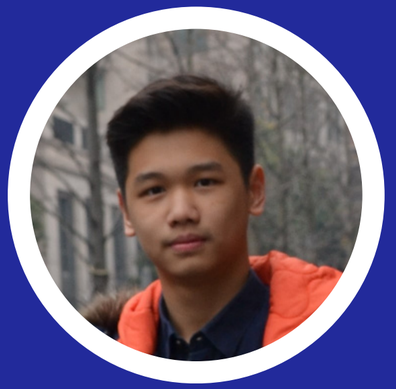
E-Commerce Customer Churn Analysis and Prediction

by: Seaborn Squad (Job Connector Data Science)

seaborn squad profile

Rayhan Romy Syahputra

Nadia Puspitasari

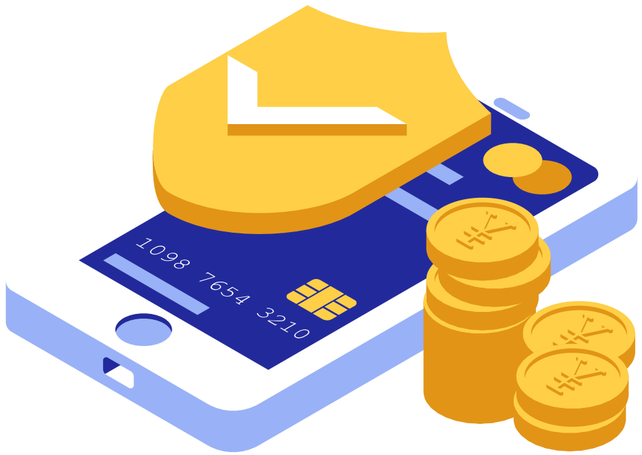


Matthew Nicholas Lengkey

Background and Business Problem

- An e-commerce company receives profits from customer transactions, so customer growth is needed so that the company can get more profits.
- The latest data shows an increasing number of customer churns which resulted in stagnant/decreasing company profits.
- Customer acquisition is 4-5 times more costly than customer retention.

Goals



✓	Predict and analyze which customers have the potential to churn
✓	Provide the right treatment for these customers so they don't churn
✓	Maintain the company profits that have been obtained
✓	Minimize the retention cost required for customers who want to churn

Analytic Approach & Consequences

We will build a classification model to predict customer churn and this model will be used by the e-commerce company.

Consequences:

False Positive:
e-commerce company will be lost customer retention costs (money and material).

False Negative:
e-commerce company will be lost customers and not do anything because it is predicted as not-churn customers.

Cost Assumption
Cost $100 per customer.

to Acquire New Customer
cost $450 per customer.

Based on the consequences, we need to reduce FN because the cost is greater than FP. We use F-beta score (beta = 2) metric because we consider FN and FP important but we put more attention to reducing FN.

Data Understanding

- The dataset is from USA, 2020.
- Dataset has 19 features and 1 target.
- The target on the dataset is an imbalance (83.2% are active customers, and 16.8% are churned customers).
- Feature description is divided into 2: customer demographic data & customer transaction data.
    - Tenure: How long the customer has using the E-Commerce Platform (in months).
    - OrderAmountHikeFromlastYear: increases orders in percentage from last year.

Attribute	Data Type, Length	Description
CustomerID	Integer	Unique customer ID
CityTier	Integer	City tier
Gender	Text	Gender of customer
MaritalStatus	Object	Marital status of customer
NumberOfAddress	Integer	Total number of address added on particular customer
WarehouseToHome	Float	Distance in between warehouse to home of customer
NumberOfDeviceRegistered	Integer	Total number of devices registered on a particular customer

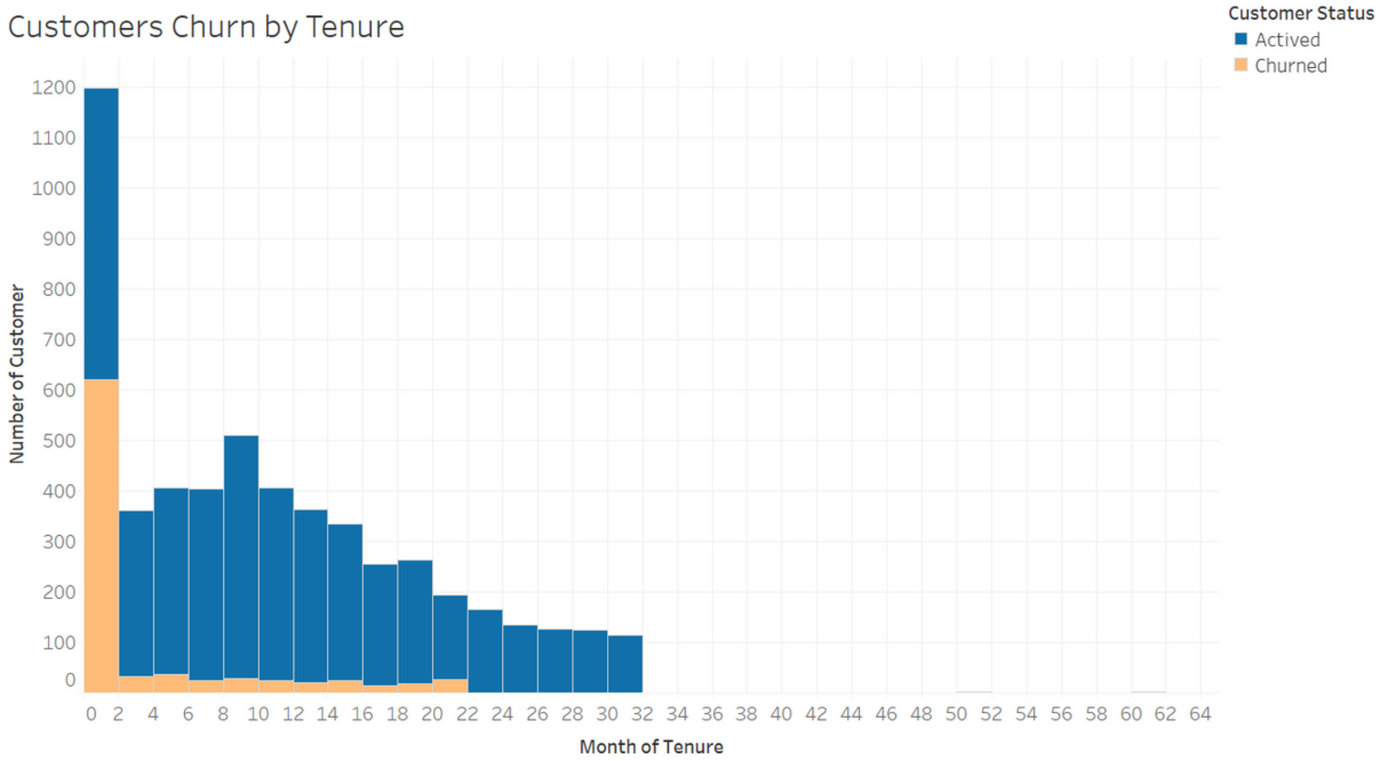
Attribute	Data Type, Length	Description
Tenure	Float	Tenure of customer in organization
PreferredLoginDevice	Text	Preferred login device of customer
PreferredPaymentMode	Text	Preferred payment method of customer
HourSpendOnApp	Float	Number of hours spend on mobile application or website
PreferedOrderCat	Object	Preferred order category of customer in last month
SatisfactionScore	Integer	Satisfactory score of customer on service
Complain	Integer	Any complaint has been raised in last month
OrderAmountHikeFromlastYear	Float	Percentage increases in order from last year
CouponUsed	Float	Total number of coupon has been used in last month
OrderCount	Float	Total number of orders has been places in last month
DaySinceLastOrder	Float	Day since last order by customer
CashbackAmount	Float	Average cashback in last month
Churn	Integer	0 - customer who have not churned; 1 customer who churned

Business Questions

Before we analyze the data, we have some assumptions and we make them into business questions.

- Do customers who stop using e-commerce stop at the beginning of the month of tenure?
- Do complain customers tend to stop using e-commerce services?
- Will a low satisfaction score mean a high churn rate?
- Have customers who left e-commerce services stopped placing orders for purchases in the past week?
- How is the product purchased from churned customers? Does it affect from the cashback they get?
- Are there specific payment methods related to customer churn?
- Do customers who log in using mobile phones more often churned from e-commerce?
- Are male customers and married customers more likely to churn from e-commerce?

Customers Churn by Tenure



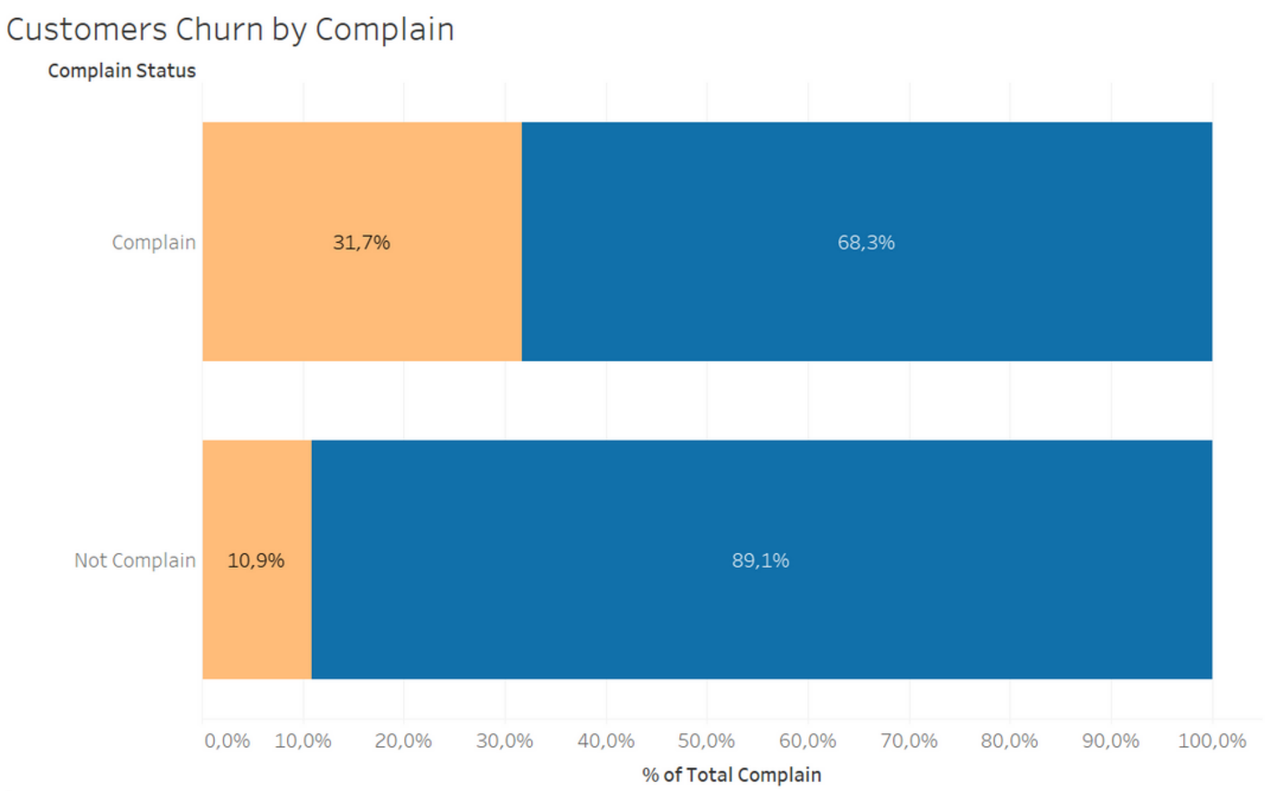
Churn	0	1	Total	Churn %
Tenure				
0.0	236.0	272.0	508.0	53.5
1.0	341.0	349.0	690.0	50.6
2.0	153.0	14.0	167.0	8.4
3.0	177.0	18.0	195.0	9.2
4.0	183.0	20.0	203.0	9.9

Data of the first 5 periods of Tenure

Insights:

- There are an increasing number of users below 2 months of tenure.
- If specified on the churned customers, the number of churned users below 2 months is higher than the number of churned users in other periods (more than 50%; it is called early-life churn).
- To increase customer tenure, the company can consider implementing gamification (e.g. point system, leveling system, etc).

Customers Churn by Number of Complain

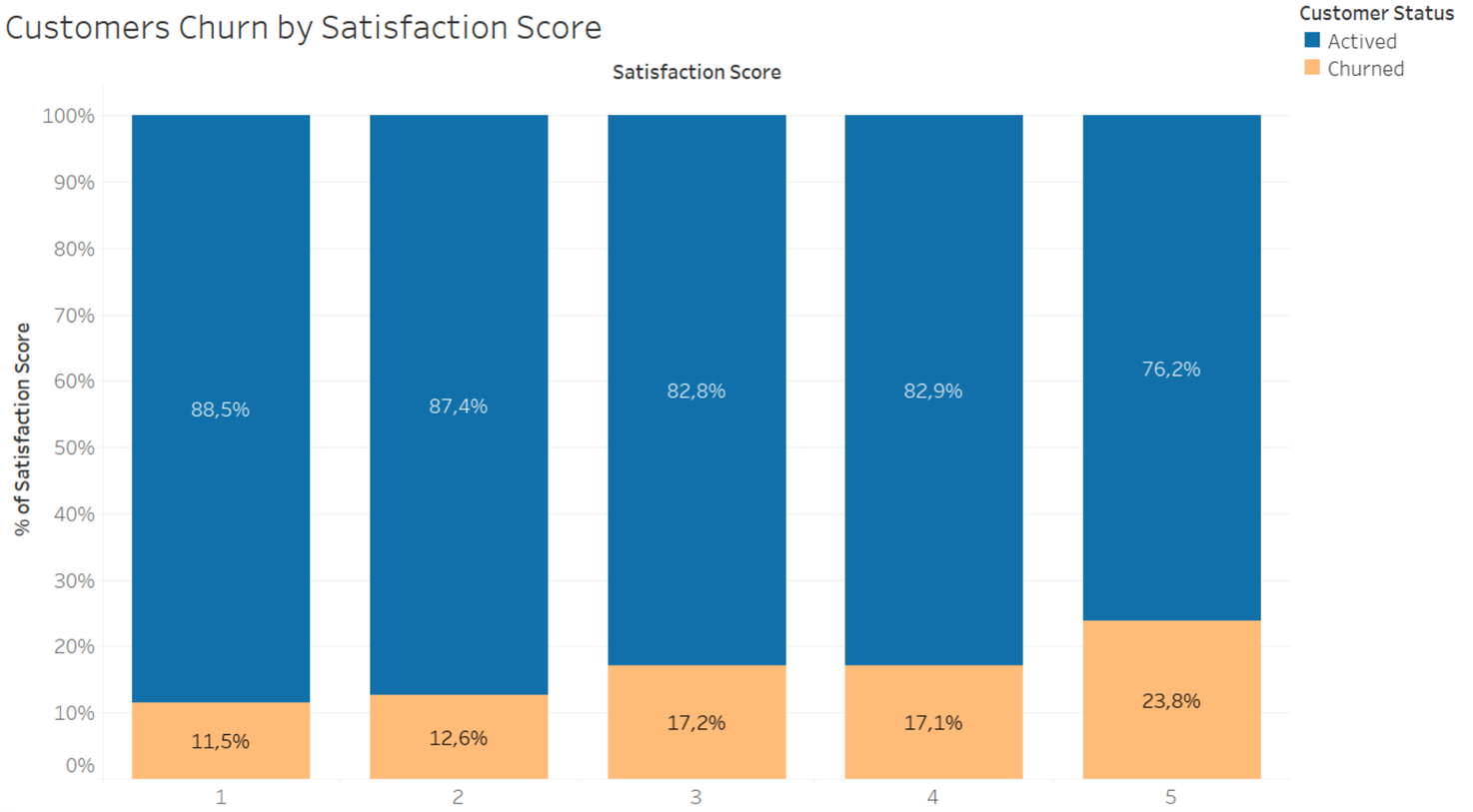


Complain	0	1
Churn		
0	3586.0	1096.0
1	440.0	508.0
Total	4026.0	1604.0
Churn %	10.9	31.7

Insights:

- Complain customers tend to churn 3 times higher (by proportion), than not-complain customers.
- Complain customers are unsatisfied customers who are prone to churn and seek other e-commerce services (competitors).
- The company needs to review customer support procedures and improve them by creating FAQs about buying and shipping process, using chatbots, and clear steps to escalate customers' issues.

Customers Churn by Satisfaction Score

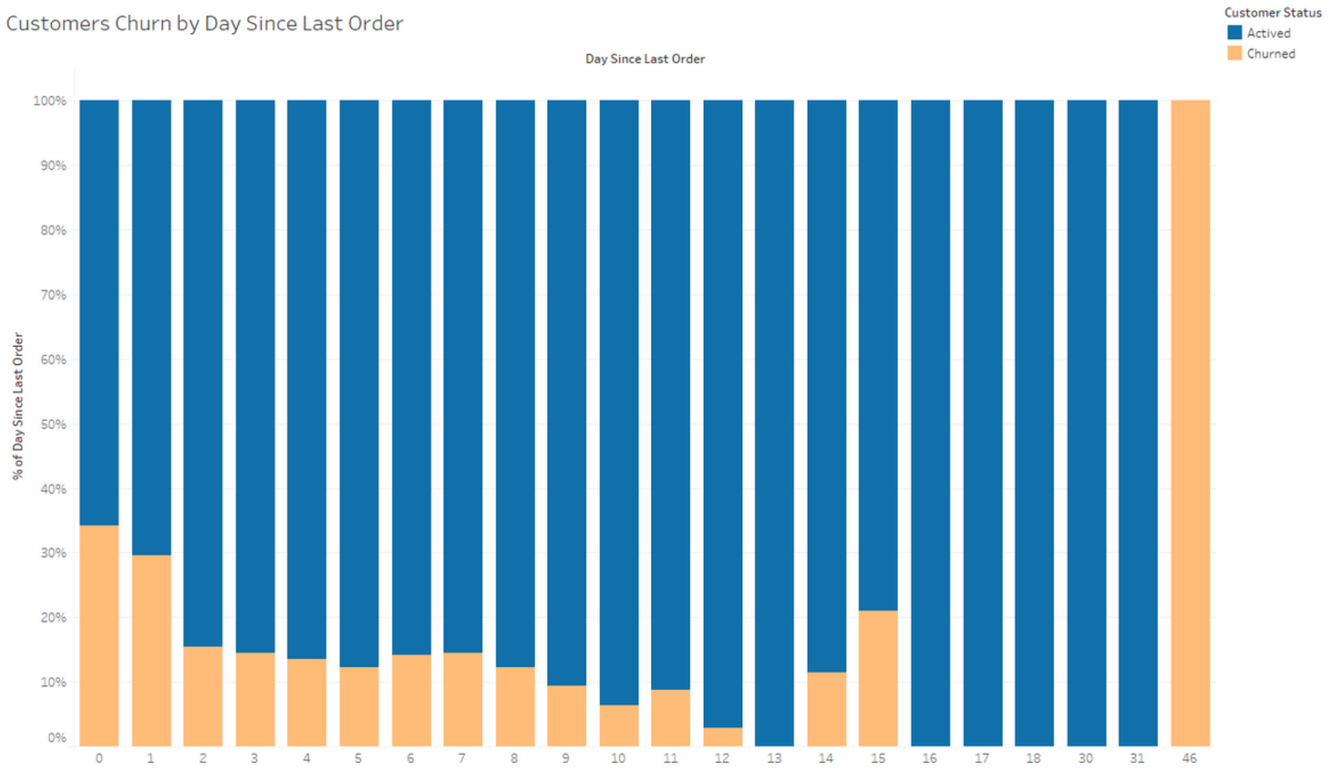


SatisfactionScore	1	2	3	4	5
Churn					
0	1030.0	512.0	1406.0	890.0	844.0
1	134.0	74.0	292.0	184.0	264.0
Total	1164.0	586.0	1698.0	1074.0	1108.0
Churn %	11.5	12.6	17.2	17.1	23.8

Insights:

- 23.8% of customers were churned even though they give the highest satisfaction score.
- It is possible to customers with high satisfaction scores will churn from the services, and vice versa.
- The company needs to explore customer reviews on the app and consider using NPS (net promotor score).

Customers Churn by Number of Day Since Last Order

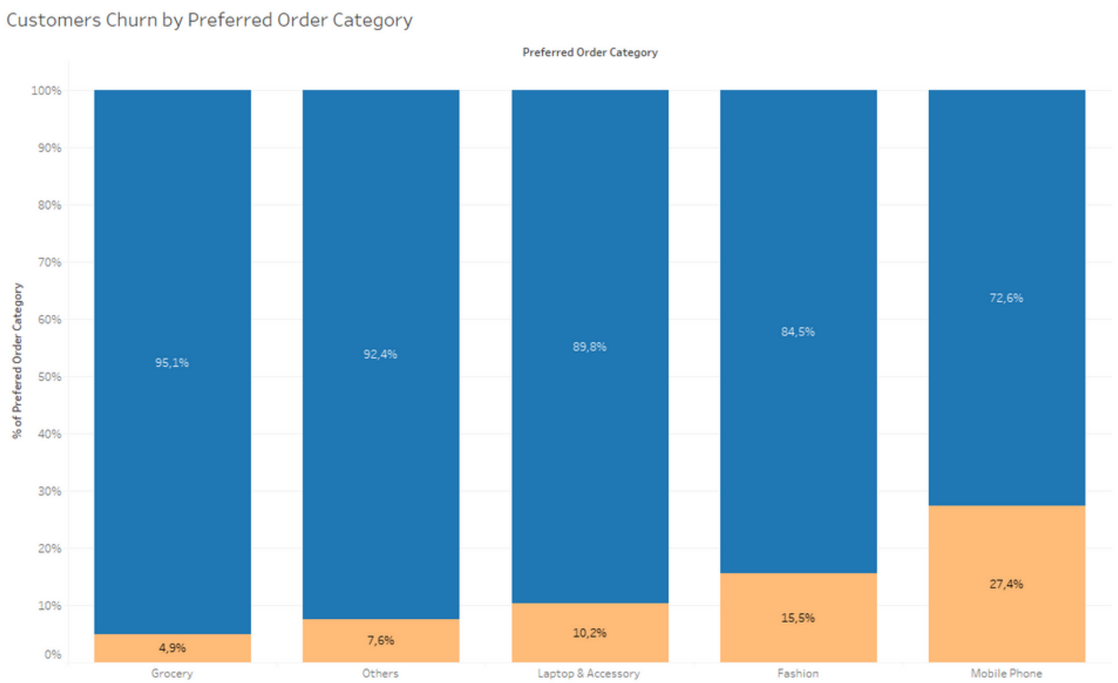


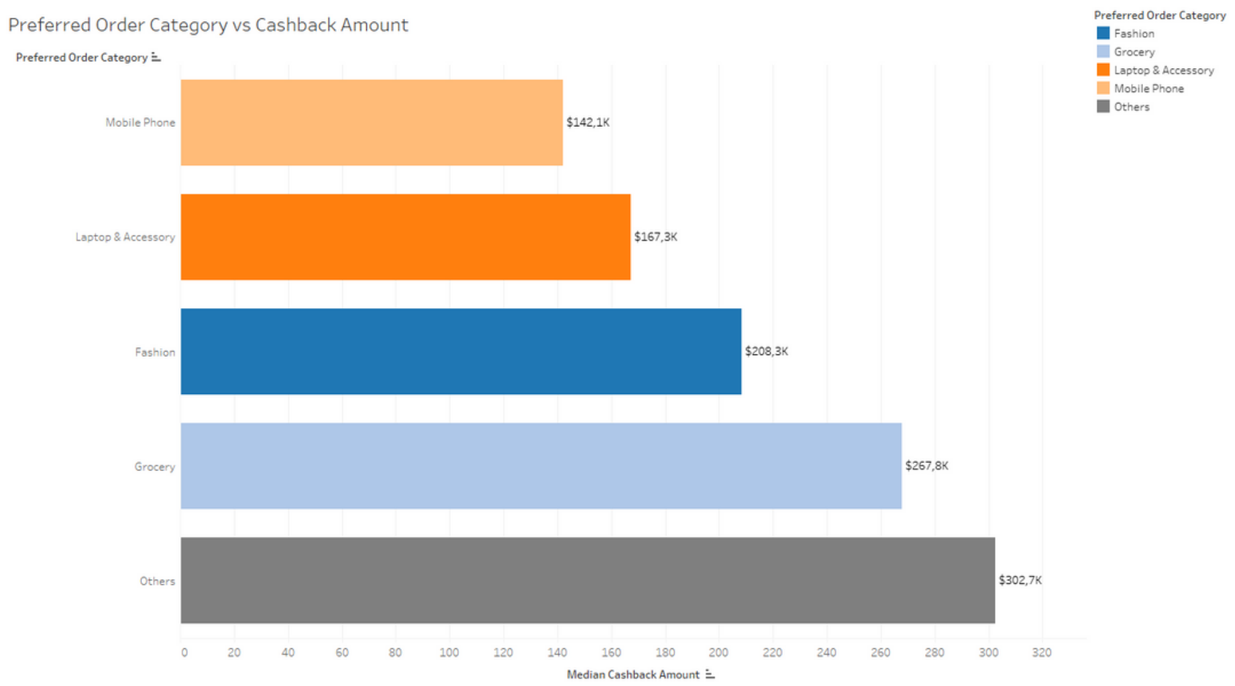
DaySinceLastOrder	0.0	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0	11.0	12.0	13.0	14.0	15.0	16.0	17.0	18.0	30.0	31.0	46.0
Churn																						
0	326.0	432.0	670.0	770.0	373.0	200.0	97.0	382.0	472.0	271.0	147.0	83.0	67.0	51.0	31.0	15.0	13.0	17.0	10.0	1.0	1.0	0.0
1	170.0	182.0	122.0	130.0	58.0	28.0	16.0	65.0	66.0	28.0	10.0	8.0	2.0	0.0	4.0	4.0	0.0	0.0	0.0	0.0	0.0	1.0
Total	496.0	614.0	792.0	900.0	431.0	228.0	113.0	447.0	538.0	299.0	157.0	91.0	69.0	51.0	35.0	19.0	13.0	17.0	10.0	1.0	1.0	1.0
Churn %	34.3	29.6	15.4	14.4	13.5	12.3	14.2	14.5	12.3	9.4	6.4	8.8	2.9	0.0	11.4	21.1	0.0	0.0	0.0	0.0	0.0	100.0

Insights:

- The churned customers still do some transactions until day 7.
- The previous assumption is the day since the last order from churned customer will be more than 7 days.
- We can assume that day since last order is not the reason for the customer's churn.

Preferred Order Category by Churn & Cashback Amount

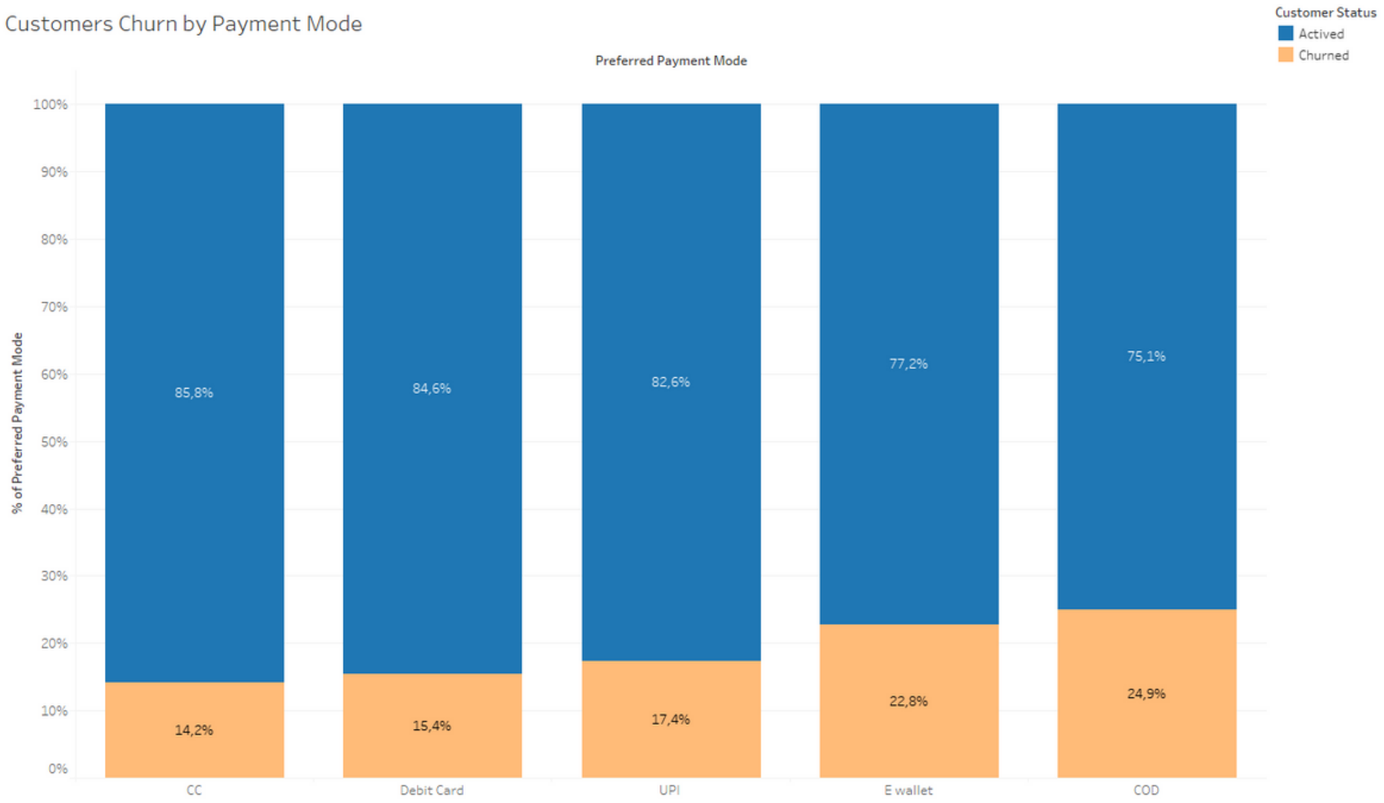




Insights:

- Customers who buy mobile phone has a higher churn rate (27.4%) than customers who buy other items and cashback from buying mobile phones is the lowest than other categories.
- The probability is customers will churn and move to competitors due to the lower cashback (on mobile phones) given by our company.
- The company can consider giving more cashback (especially for mobile phone buyers).

Customers Churn by Payment Method

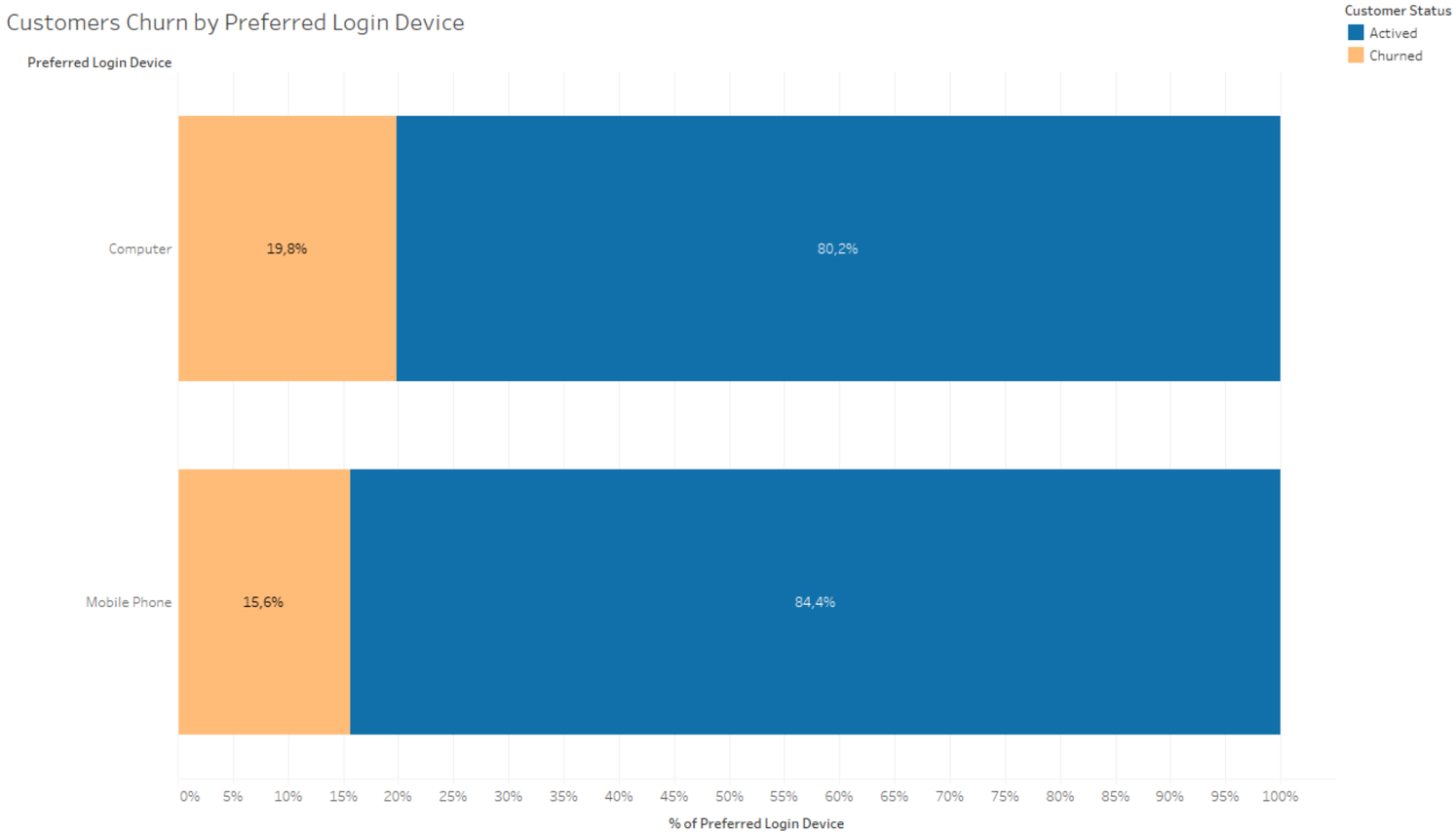


PreferredPaymentMode	CC	COD	Debit Card	E wallet	UPI
Churn					
0	1522.0	386.0	1958.0	474.0	342.0
1	252.0	128.0	356.0	140.0	72.0
Total	1774.0	514.0	2314.0	614.0	414.0
Churn %	14.2	24.9	15.4	22.8	17.4

Insights:

- 24.9% of customers who have payment with Cash on Delivery method tend to churn more than customers with other payment methods.
- The company can review partnerships with third parties regarding COD payment support.

Customers Churn by Login Device

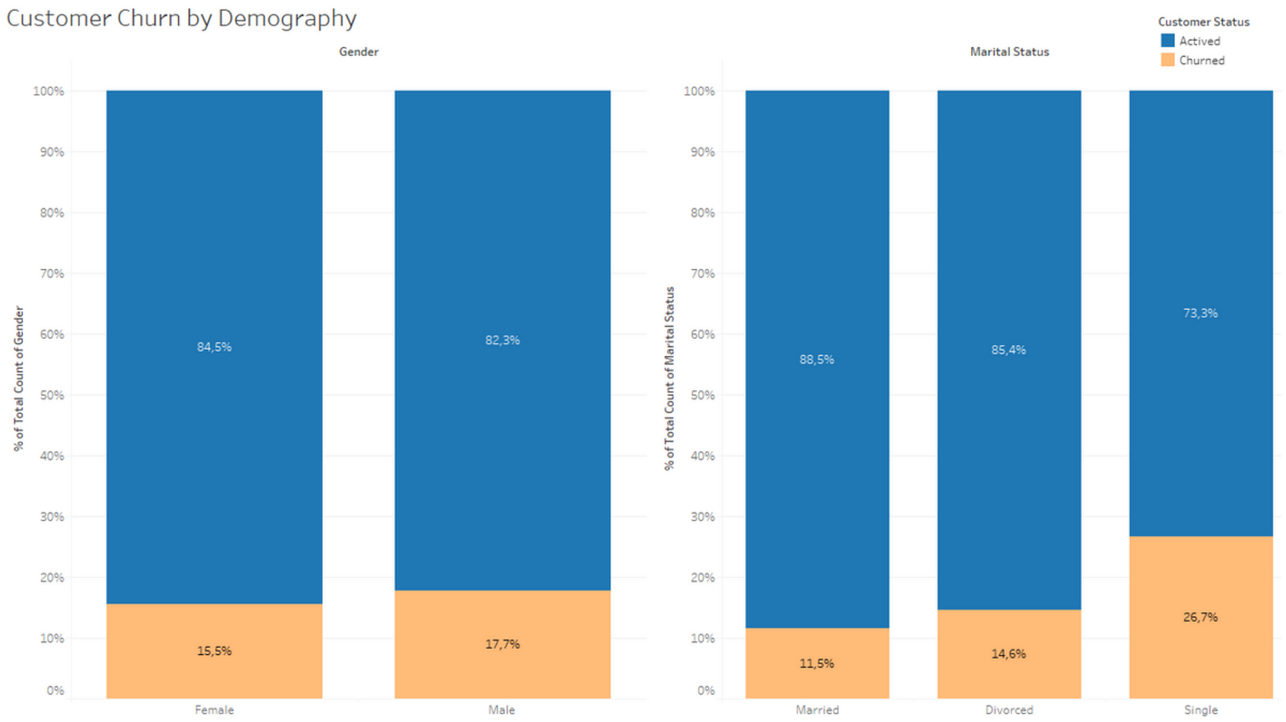


PreferredLoginDevice	Computer	Mobile Phone
Churn		
0	1310.0	3372.0
1	324.0	624.0
Total	1634.0	3996.0
Churn %	19.8	15.6

Insights:

- 19.8% of customers who logged in by computer tend to churn more than customers logged in by mobile phone.
- Company can review the UI/UX of e-commerce applications via computer/mobile phone.

Customers Churn by Demography



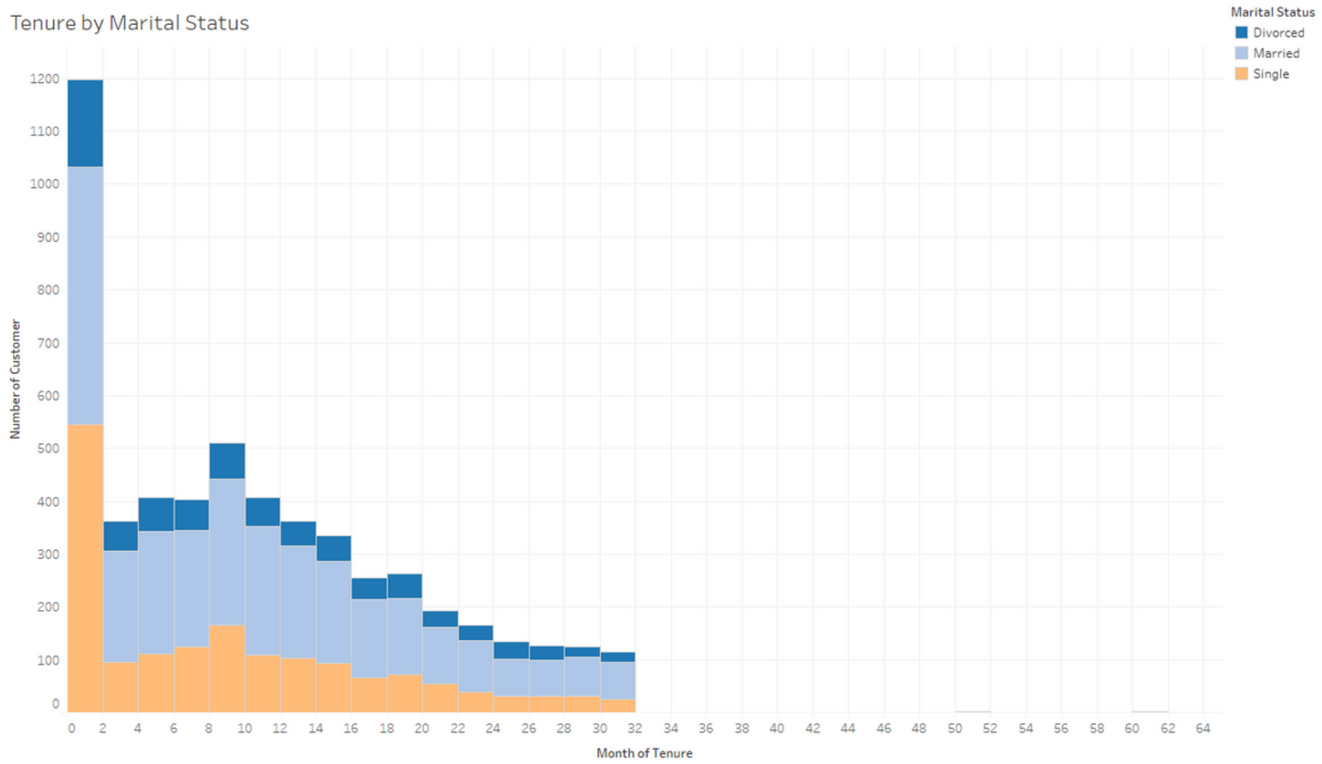
Gender	Female	Male
Churn		
0	1898.0	2784.0
1	348.0	600.0
Total	2246.0	3384.0
Churn %	15.5	17.7

MaritalStatus	Divorced	Married	Single
Churn			
0	724.0	2642.0	1316.0
1	124.0	344.0	480.0
Total	848.0	2986.0	1796.0
Churn %	14.6	11.5	26.7

Insights:

- Male customers tend to churn 2.2% higher (by proportion), than female customers
- 26.7% of single customers tend to churn more than married and divorced customers.

Customers Churn by Demography



MaritalStatus	Divorced	Married	Single	Total
Tenure				
0.0	70	201	237	508
1.0	95	286	309	690
2.0	25	99	43	167
3.0	31	112	52	195
4.0	31	118	54	203

Data of the first 5 periods of Tenure

Insights:

- Single customers also had the highest number of users in below 2 months than married and divorced customers.
- The company can create a more segmented marketing campaign (especially for customers with higher churn characteristics).

Data Cleaning & Preprocessing

Data Cleaning

- We remove 'CustomerID' column.
- Delete 556 duplicated values
- Some numerical features have outliers and we using IQR method to remove the outliers (which resulting into data limitations)
- There are some missing values on 7 features.

Data Preprocessing

- Impute the missing values by using a combination of iterative imputers and simple imputers.
- Simple imputer will be filled with the average column that has a correlation.
- Categorical features will be one hot encoding because these features are not ordered/non-ordinal, and also the amount of unique data is only small.

Clean Dataset

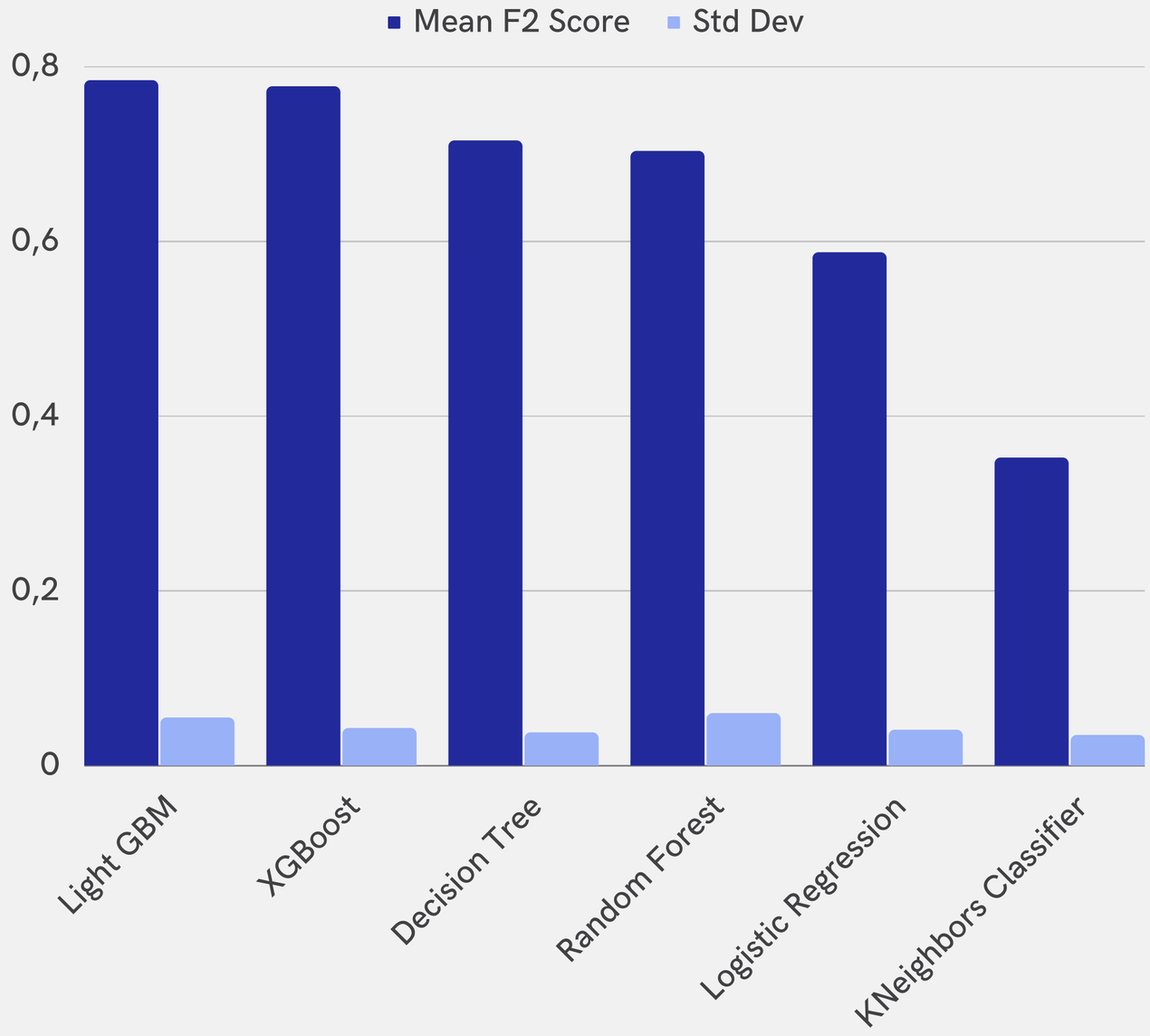
- 2570 rows
- 19 columns

Modeling Machine Learning

Choose a Benchmark Model

Using F2 Score (Mean and Standard Deviation)

model	mean F2	sdev
LightGBM	0.784921	0.055674
XGBoost	0.777521	0.043133
Decision Tree	0.715422	0.038675
Random Forest	0.703324	0.060253
Logistic Regression	0.587486	0.041565
KNN	0.352136	0.035020



Modeling Machine Learning

Hyperparameter Tuning

To improve model performance, we did hyperparameter tuning using RandomizedSearchCV with parameters below:

```
hyperparam_space = {
    'model__max_depth': list(np.arange(1, 31)),
    'model__num_leaves':list(np.arange(2,900,5)),
    'model__min_data_in_leaf': list(np.arange(10,101,2)),
    'model__num_iterations':list(np.arange(10,101,2)),
    'model__learning_rate': list(np.arange(0.1,1,0.01)),
}
```

The result is improved as the score is increased than before tuning.

Condition	Mean F2 Score
Before Tuning	0.784
After Tuning	0.825

Best Parameters from RandomizedSearch:

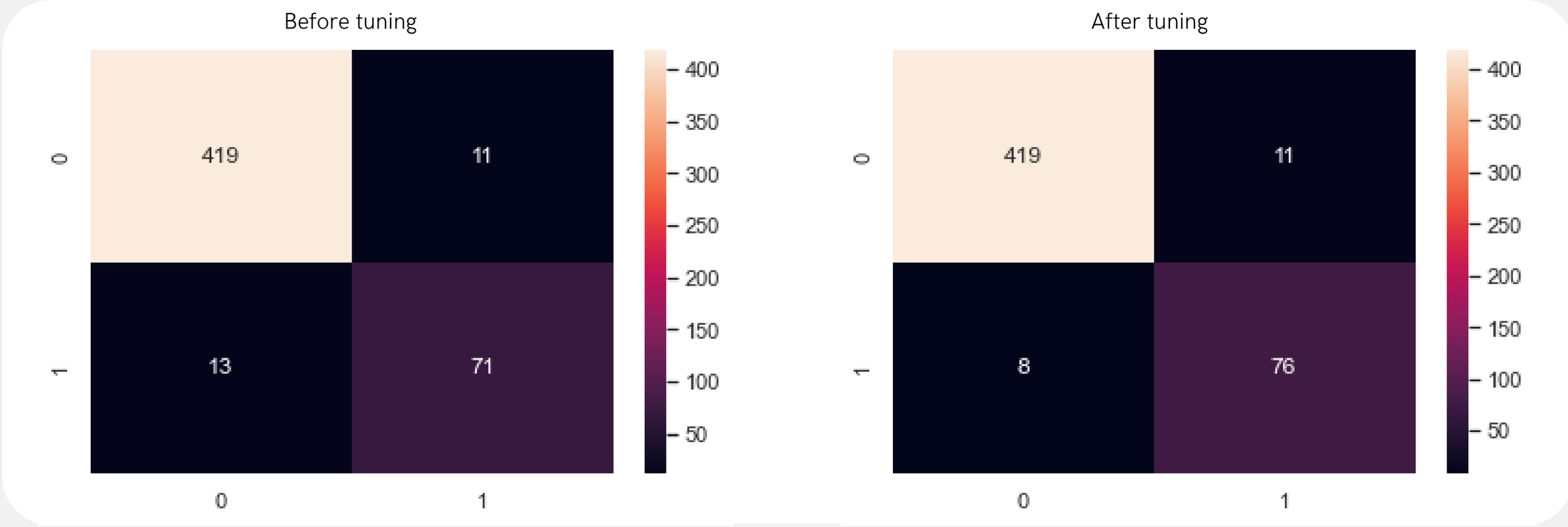
- model__random_state: 0
- model__num_leaves: 137
- model__num_iterations: 28
- model__min_data_in_leaf: 82
- model__max_depth: 25
- model__learning_rate: 0.8199999999999996

Predict to Test Set

After hyperparameter tuning, the best-tuned model predicts the test set, and the result is also better which shows a higher F2 Score compared to the score on the benchmark model.

```
F2 Score Default LGBM :  0.8492822966507177
F2 Score Tuned LGBM :  0.8983451536643026
```

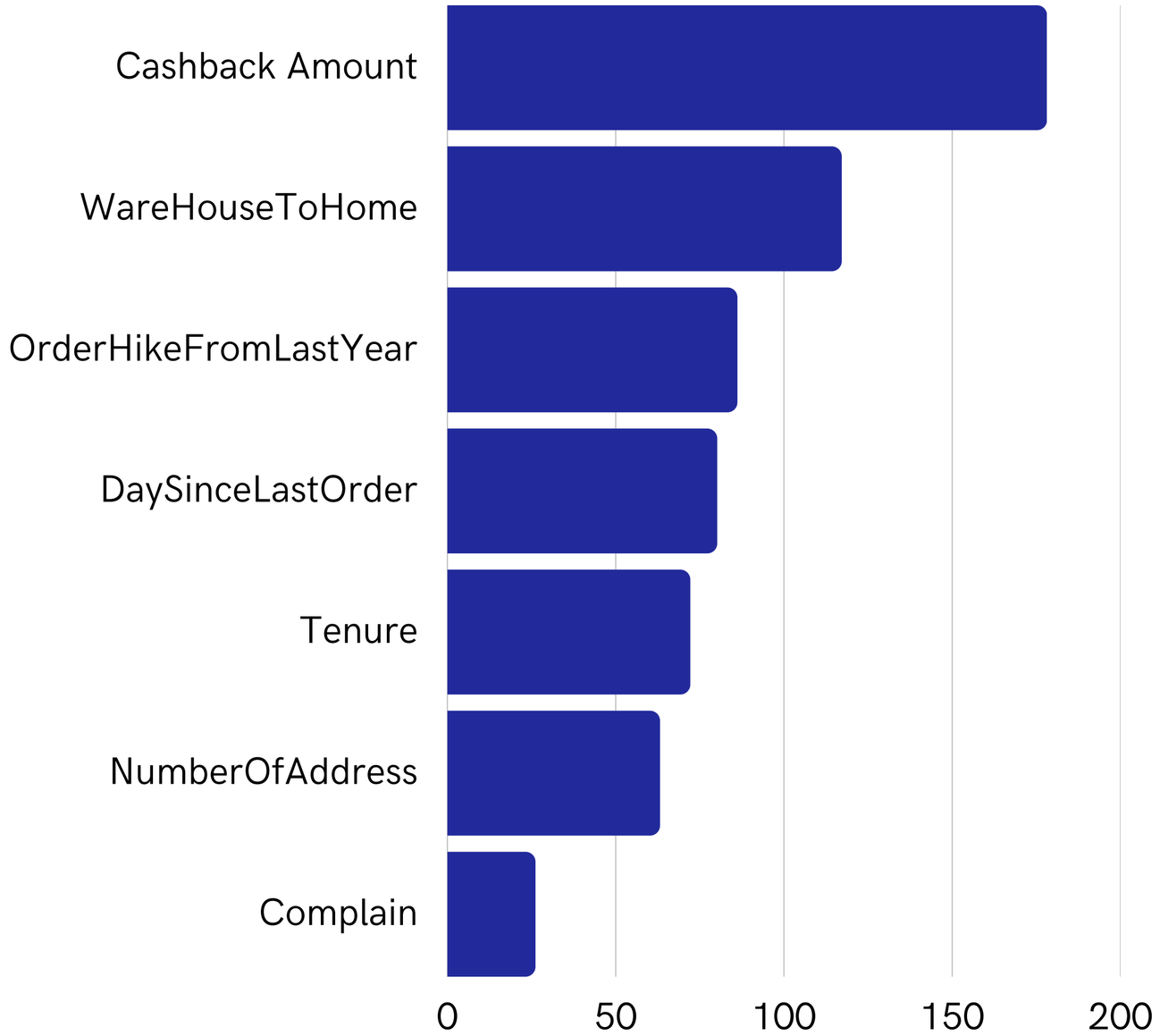
Confusion Matrix



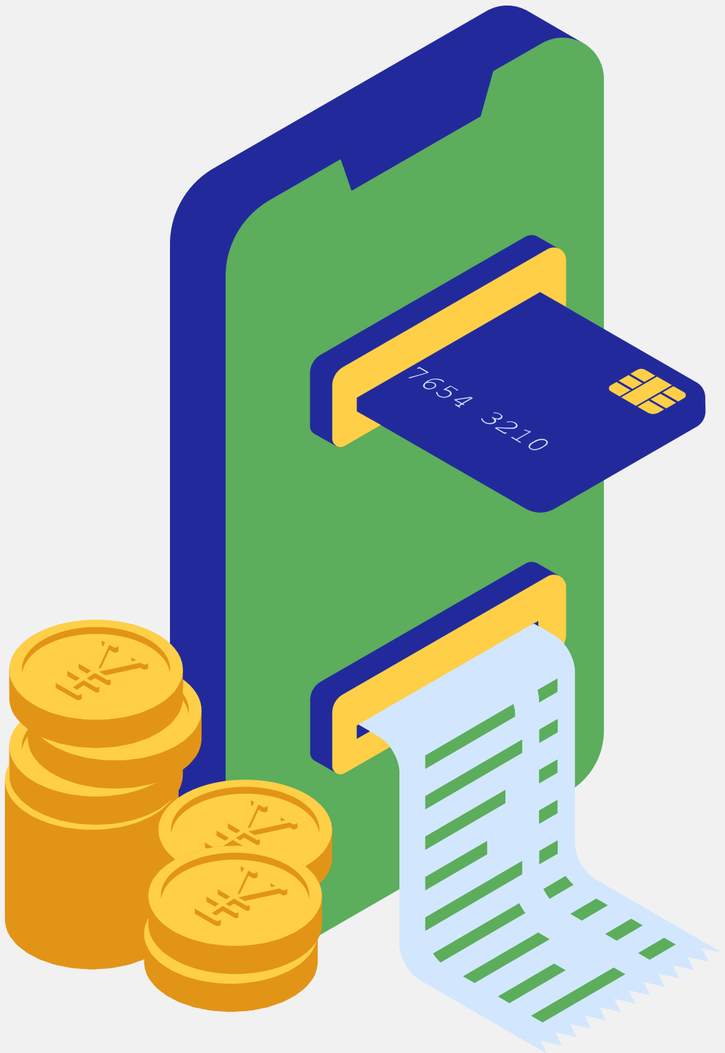
Confusion Matrix Terms

- True Positive (TP): customers who are predicted to churn and actually churn.
- True Negative (TN): customers who are predicted not to churn and actually do not churn.
- False Positive (FP): customers who are predicted to churn and actually do not churn.
- False Negative (FN): customers who are predicted not to churn and actually churn.

Feature Importances



Conclusion & Recommendation



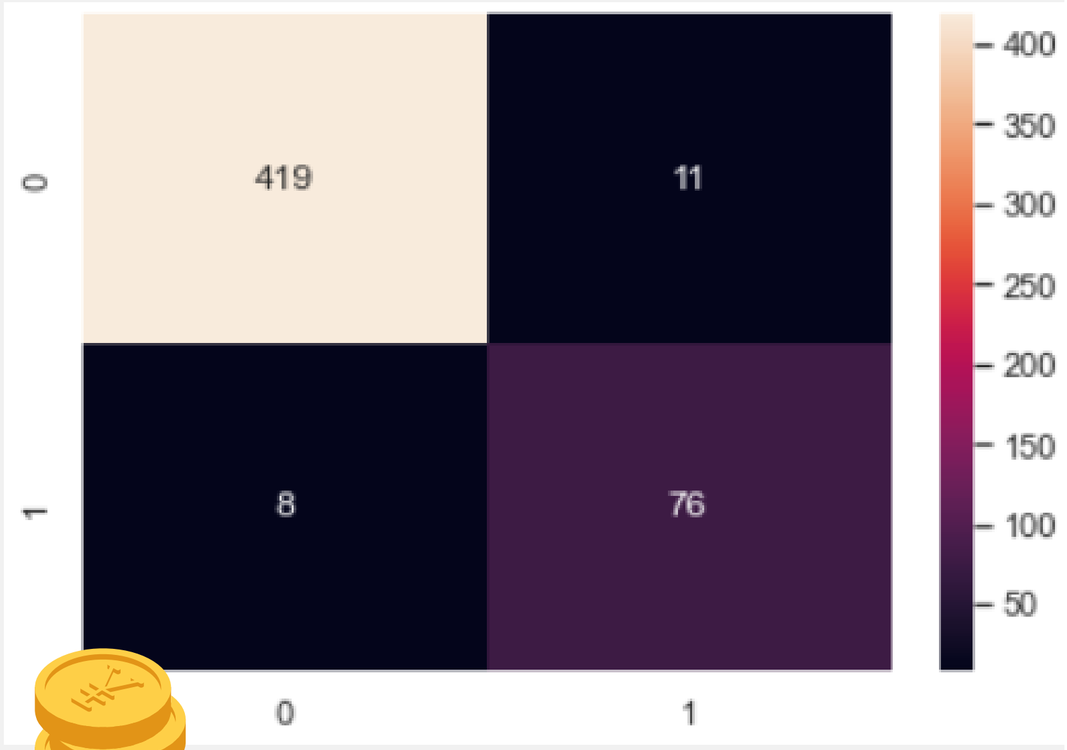
Conclusion (Calculation)

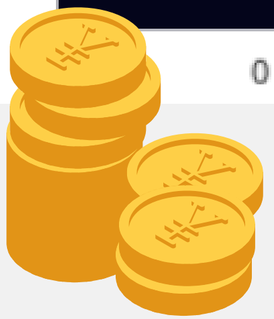
Without Model:
Statistic Churn using y_test (20%):
- 0: 430
- 1: 84
- Total data: 514

Without a model, it is difficult for us to know which customers are churn or not, so the calculation is:

- Total customers who will definitely churn: 84 people
- Acquisition cost to replace churned customers:
  - 84 × 450 USD = 37.800 USD
  - Total Cost = 37.800 USD

Saving amount: 0 USD





Conclusion (Calculation)

Using Model (test set 20%):

Total records = 514

Based on the confusion matrix:

- Cost for promotion:
  - (76+11) × 100 USD = 8.700 USD
- Acquisition cost to replace a churned customer:
  - 8 × 450 USD = 3.600 USD
- Total Cost:
  - 8.700 USD + 3.600 USD = 12.300 USD
- Savings amount:
  - 37.800 USD - 12.300 USD = 25.500 USD

Cost Saving is up to 67%

(% Cost Saving: 25,500 USD / 37,800 USD x 100% = 67%)

Based on the above calculation, utilizing the model that has been made, the company can save more costs incurred.

It is better to pay for retention costs than to potentially lose customers (churn).

Recommendation (ML Model)

- Add new features/columns that are related to potential customers to churn.
- Conduct cohort analysis to see customers who are churning based on the application access period/customer transaction period.
- Adding samples to the dataset so that the model can have many references so that predictions can be more precise.
- Try other machine learning algorithms and perform hyperparameter tuning.
- Analyzing the data that the model predicts incorrectly to find out the reasons and what its characteristics are.

Recommendation (Business)

- Implement gamification to increase customer's tenure (e.g. point system, leveling system, etc.).
- Creating a more segmented marketing campaign, especially for customers with higher churn characteristics.
- Reviewing the UI/UX of e-commerce applications via computer/mobile phone to provide a better experience for customers.
- Providing more promos and good customer support for new customers to increase the customer's tenure.
- Consider giving more cashback (especially to mobile phone buyers) to reduce the number of churn customers and do A/B testing referring to cashback given to customers.
- Reviewing partnerships with third parties regarding COD payment support to improve a better experience for customers.

Thank you.

Do you have any questions?