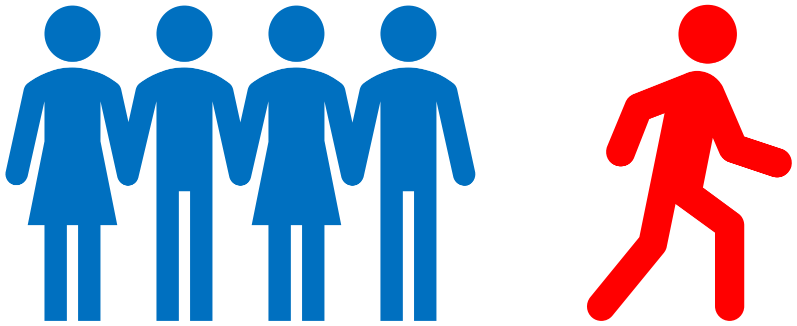
Exploratory Data Analysis

EDA Problem Statement



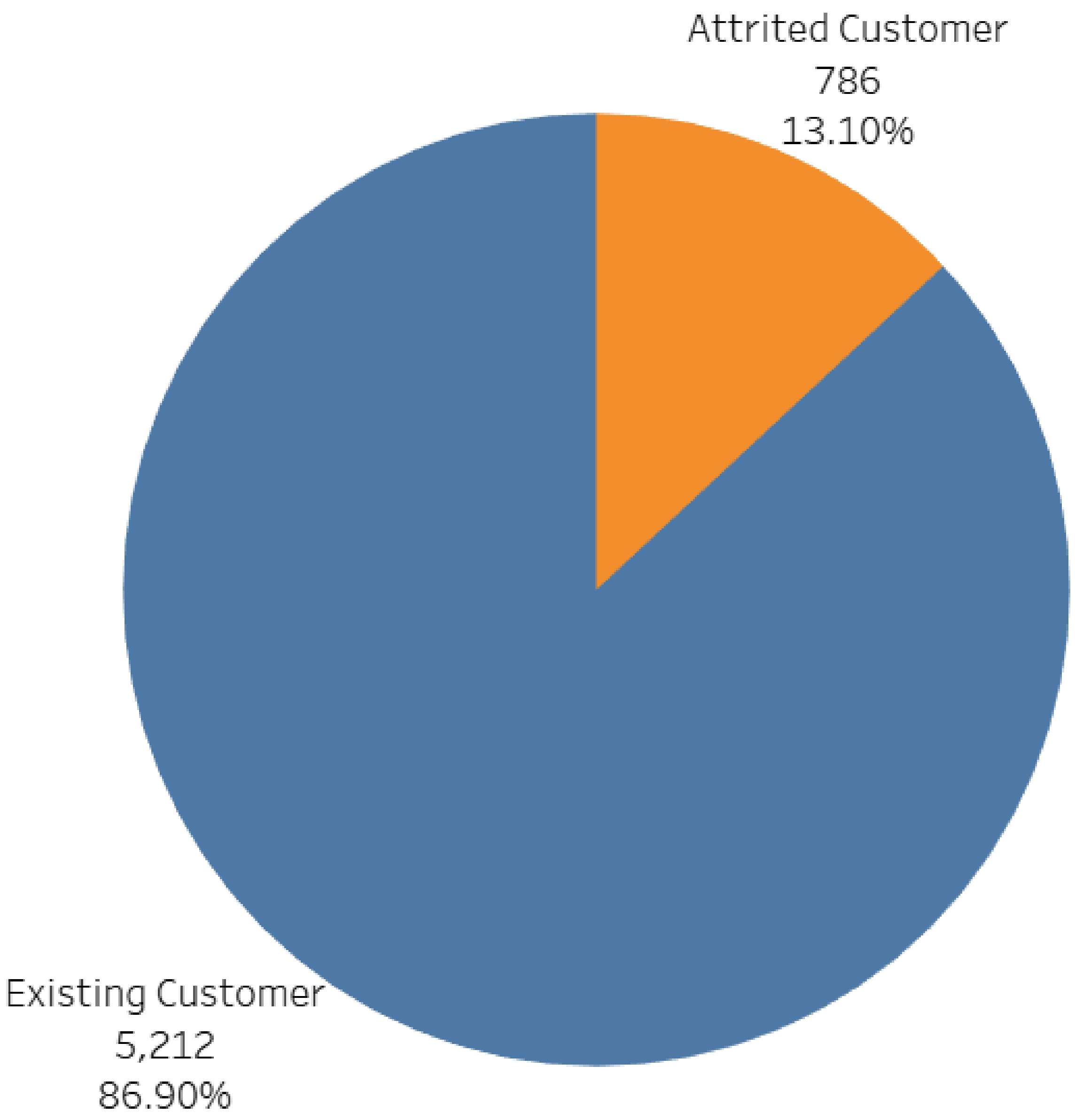
- The CEO of Bank is concerned about customer attrition in their credit card services and wants to proactively address the issue by predicting which customers are most likely to cancel their accounts.
- How can we address the issue of customer attrition in the credit card services provided by Bank Inc?

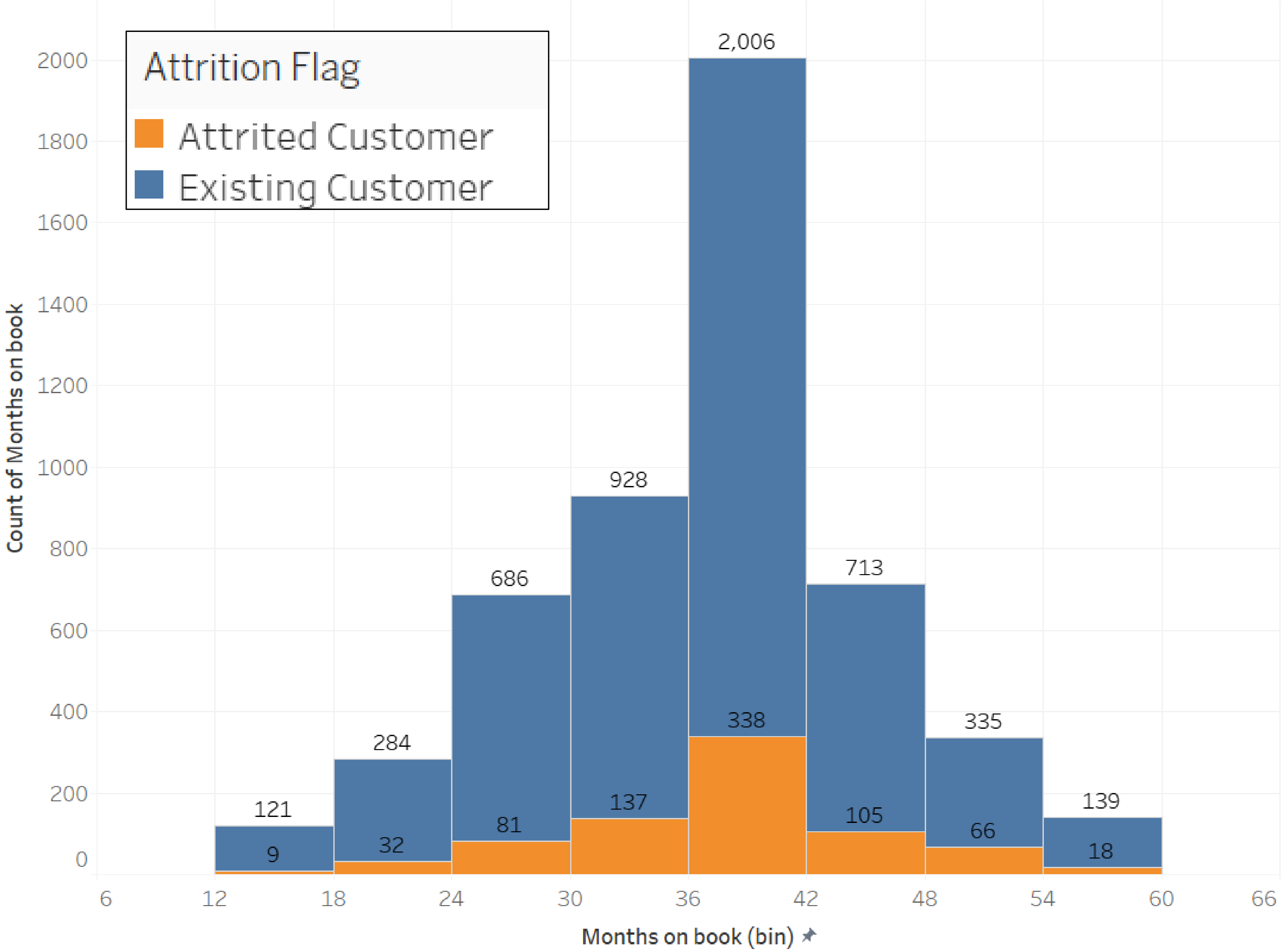
Goals of EDA

- The objective is to offer an initial collection of predictor variables and machine learning models that could be employed for predictive modeling.
- This would involve identifying a set of features that may potentially have a significant impact on predicting customer attrition.
- Additionally, exploring various machine learning techniques that could be utilized to generate accurate predictions based on the identified features.
- The end goal is to create a reliable and effective model that can be used to predict which customers are likely to cancel their credit card services with Very Nice Bank.

Attrition Count

- Data set contains 86.9% of the customers who are using the bank services
- It contains 13.1% of the customers who are not using the services of the bank
- Are there any specific feature/category/sub-category in which the customers are dropping off the bank services?



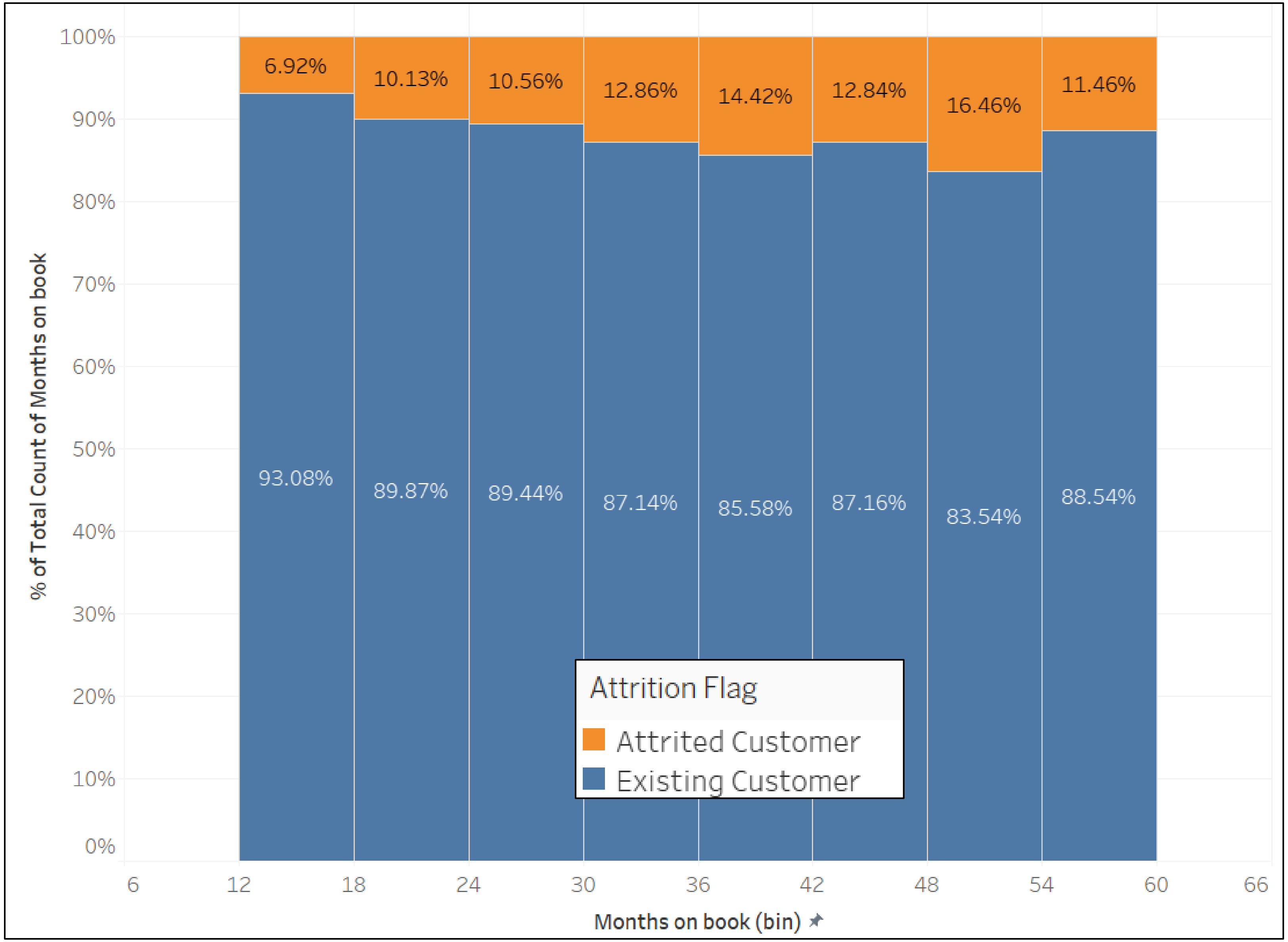


ths on book

relationship with bank)

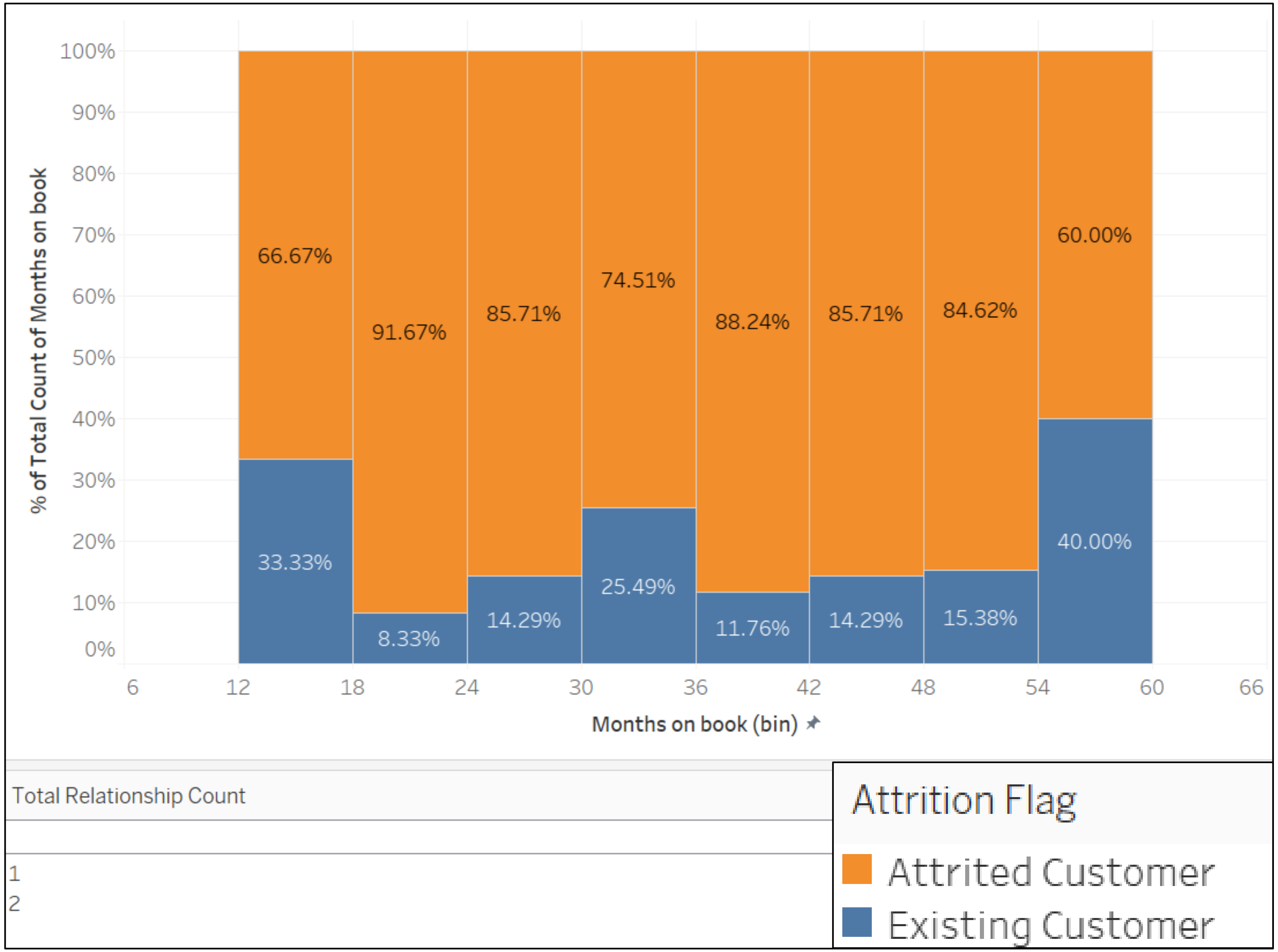
Numerical

ut of 786 (**95%**) s are using bank ; **more than 2 years**
ut of 745 (**45%**) s are the customers ervices b/w **3 to 3-1/2**



Months on book

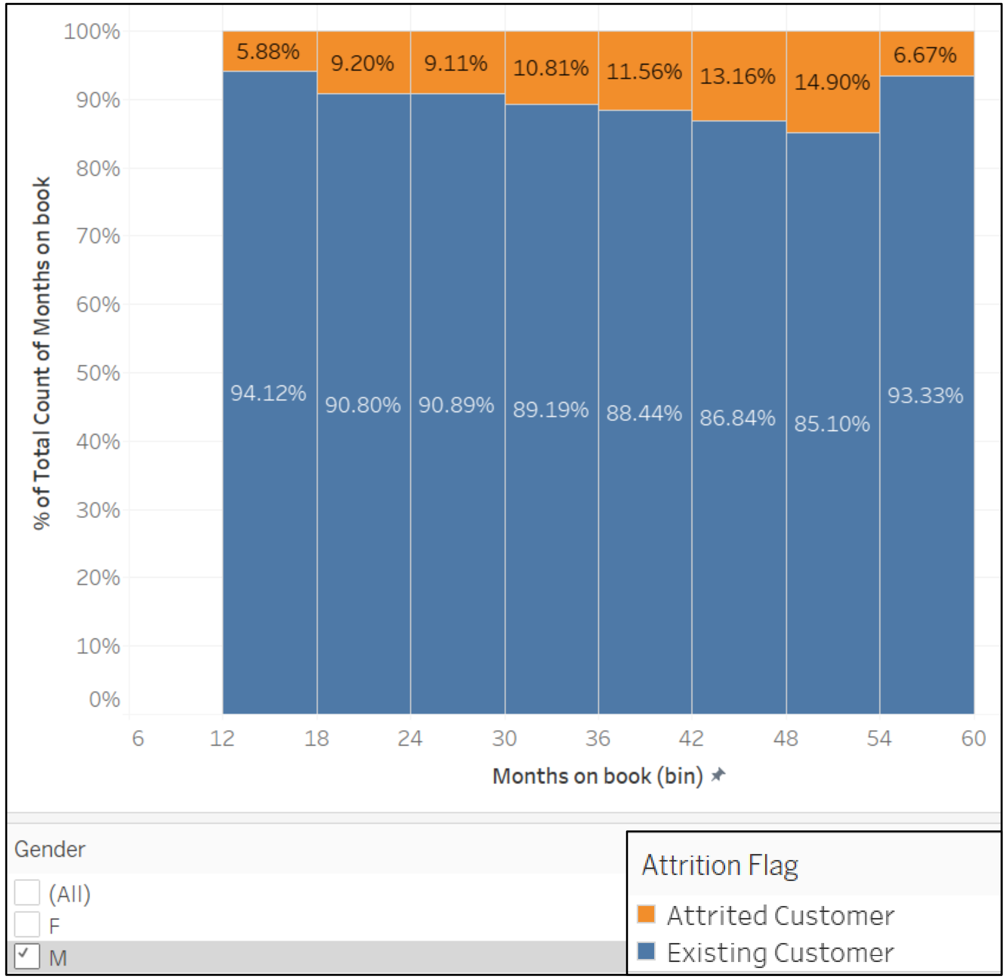
- After normalizing, the bin of **30-36 months** (3 to 3-1/2 yrs), only **14%** of the customers in that period are churners.
- Nearly around **12%** churners in **each** months' bin.

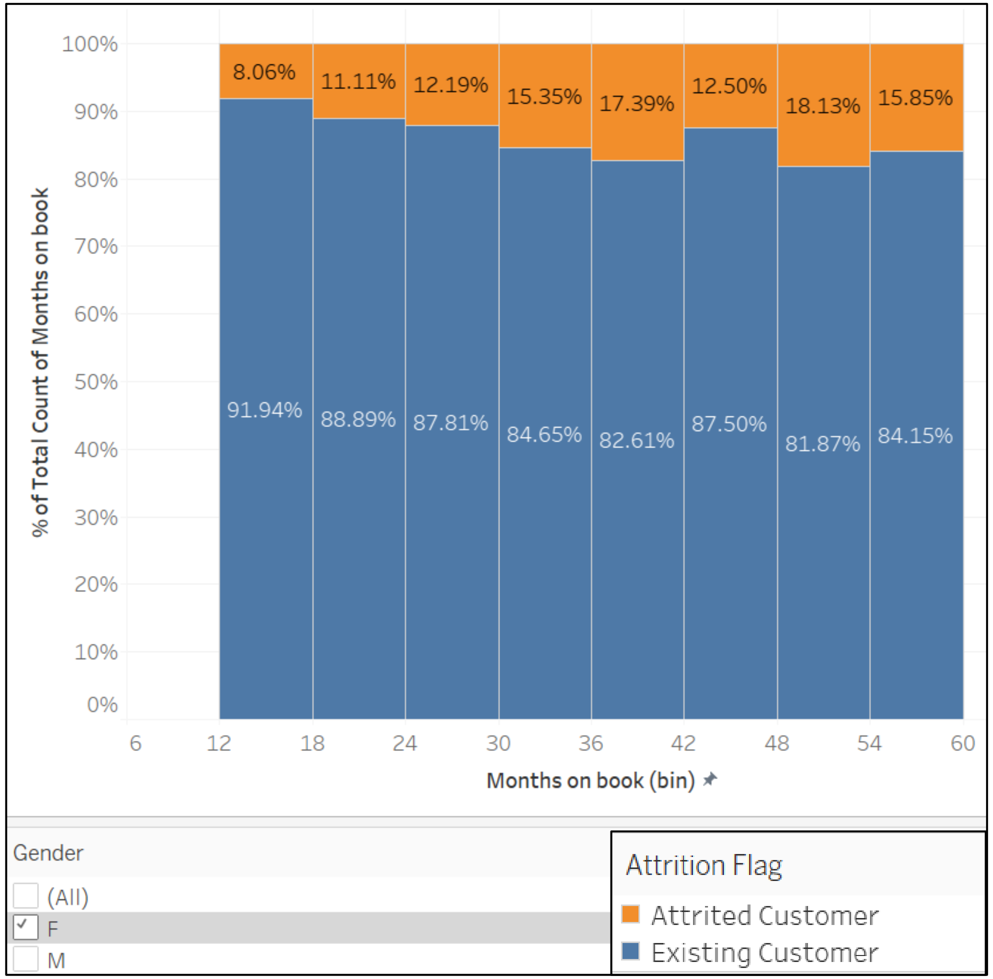


Months on book

After adding filters with Total Relationship Count - 1 & 2, it is clear that customers churn who are almost same.

Months on book



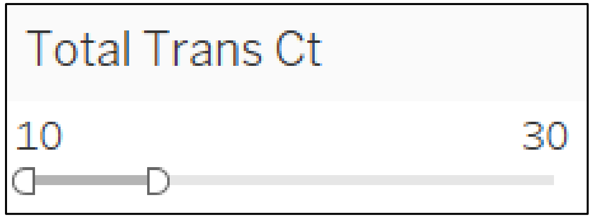


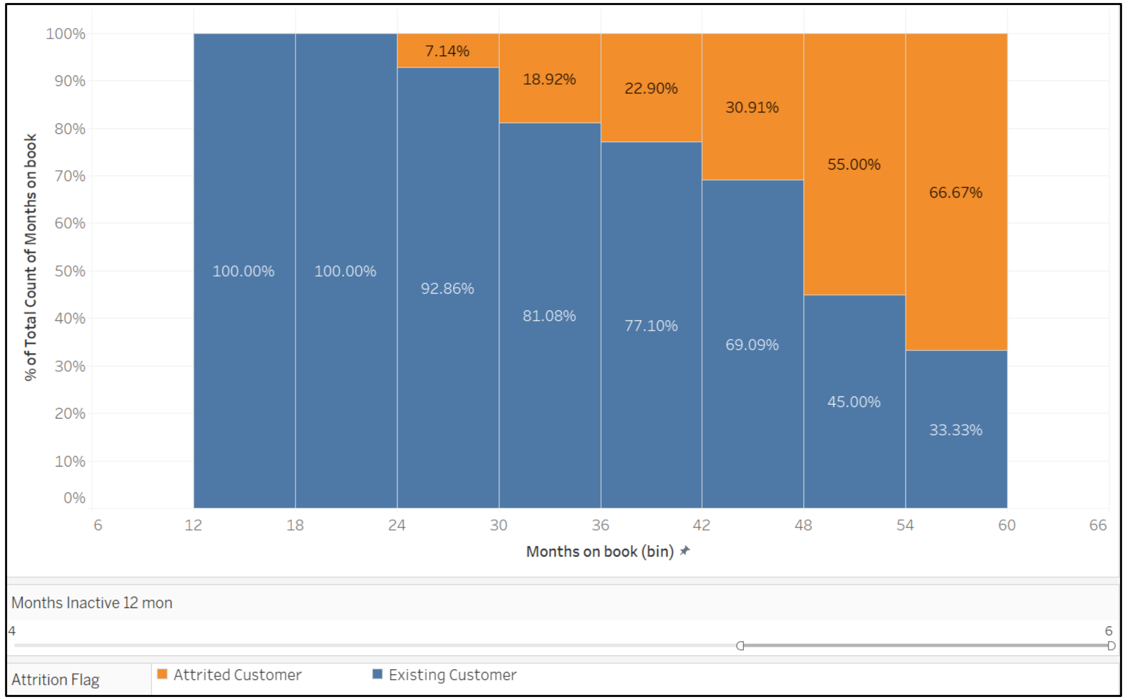
Either Male or Female, customer are churning at almost equal %

Months on book

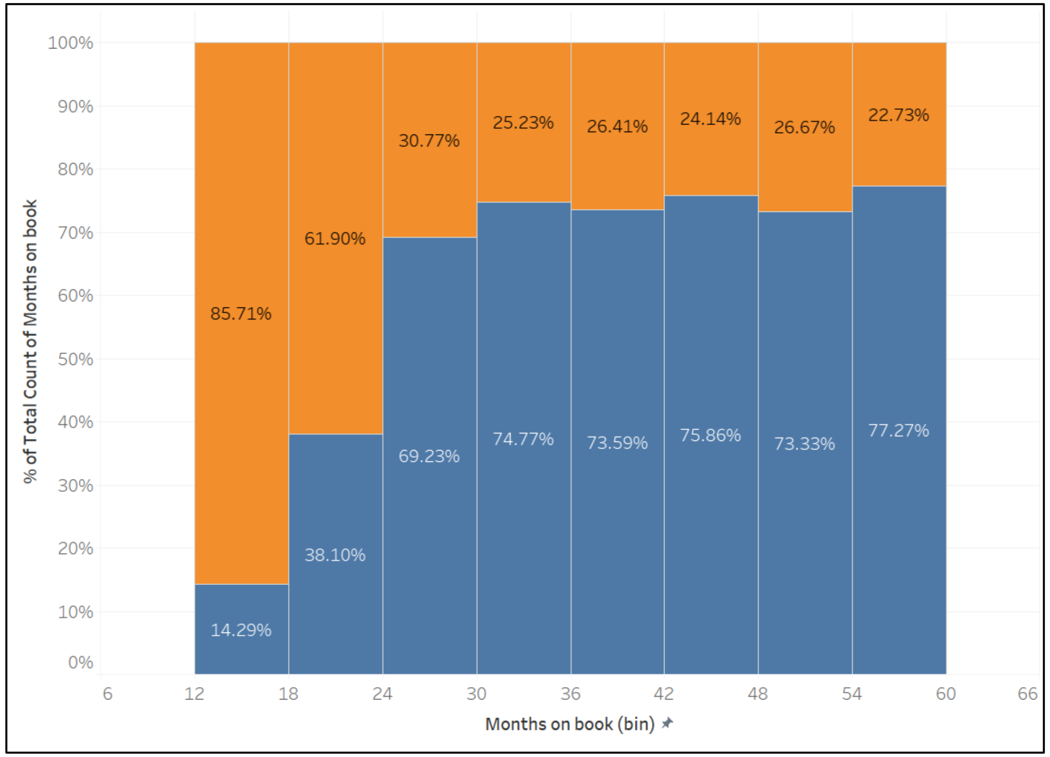
(Period of relationship with bank)

Numerical





After adding filters of No. of Months the customer is inactive during last 12 months, it is clear that customers churn who are inactive for more than 4 months and are joined bank 2.5 years ago



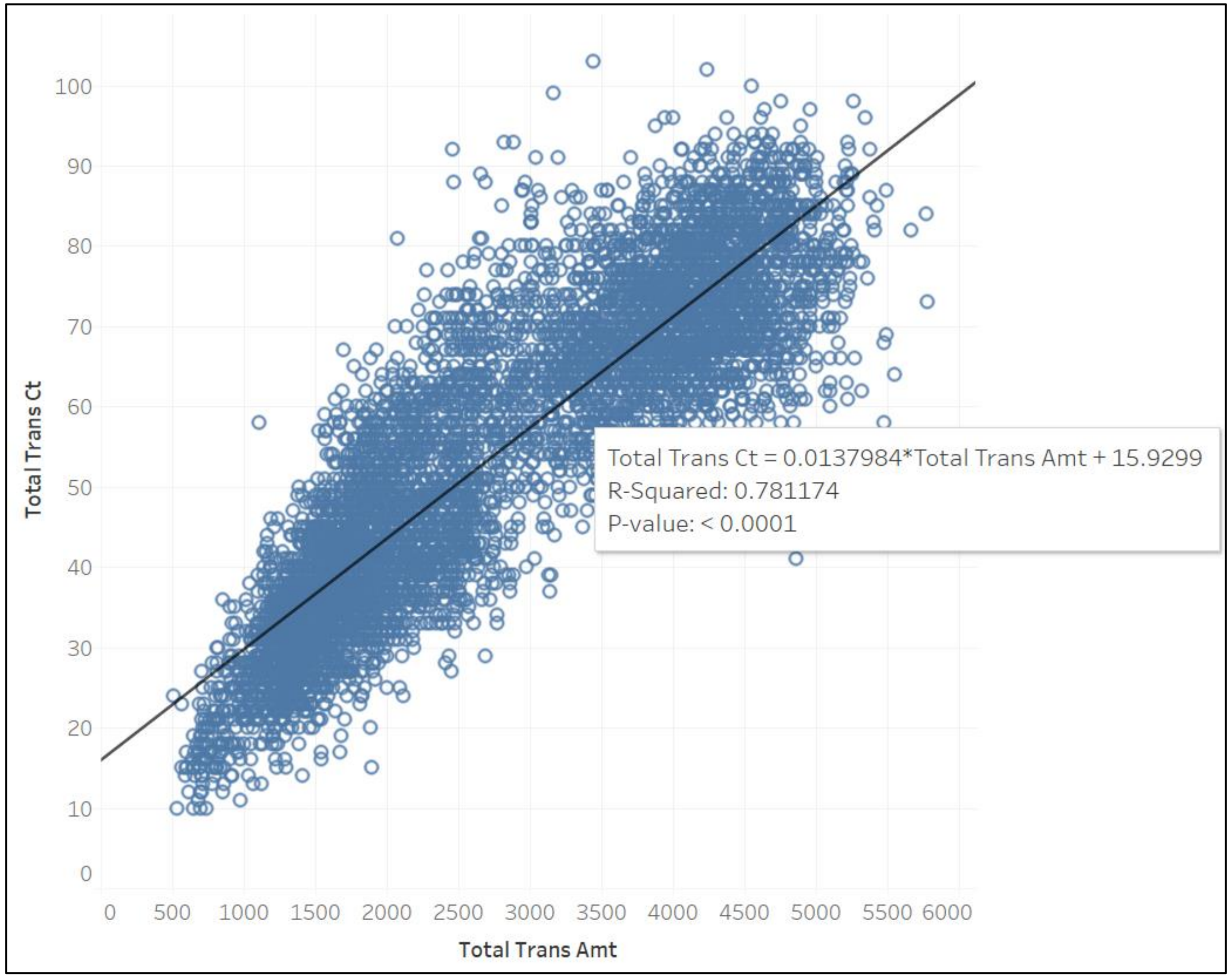
After adding filters of Total Transaction Count in last 12 months, the customers churn % is more on new customers who have made less than 30 transactions

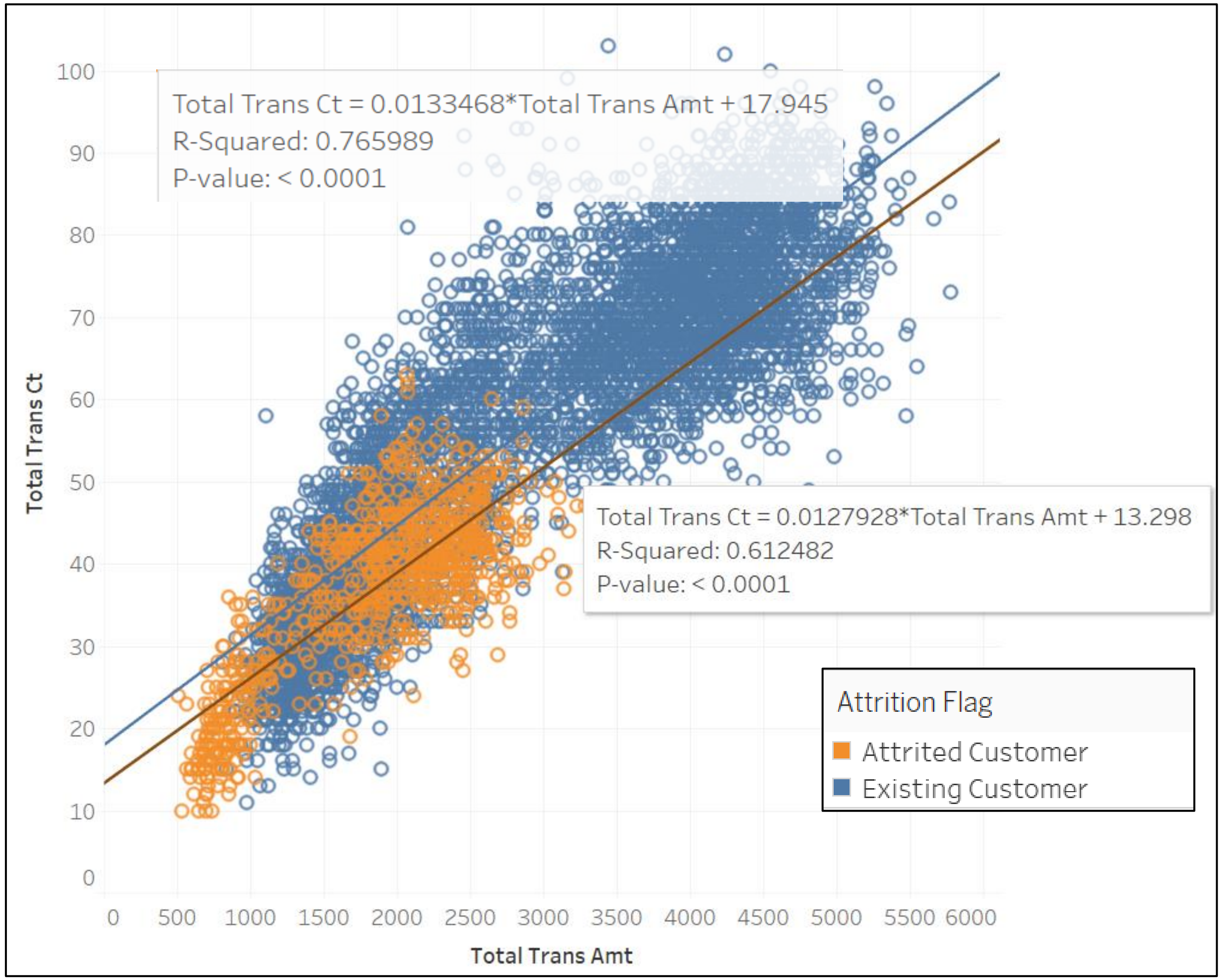
Numerical

Total Transaction Amt Vs Total Transaction Ct

Numerical

There is strong relation - R-Squared value is 0.78



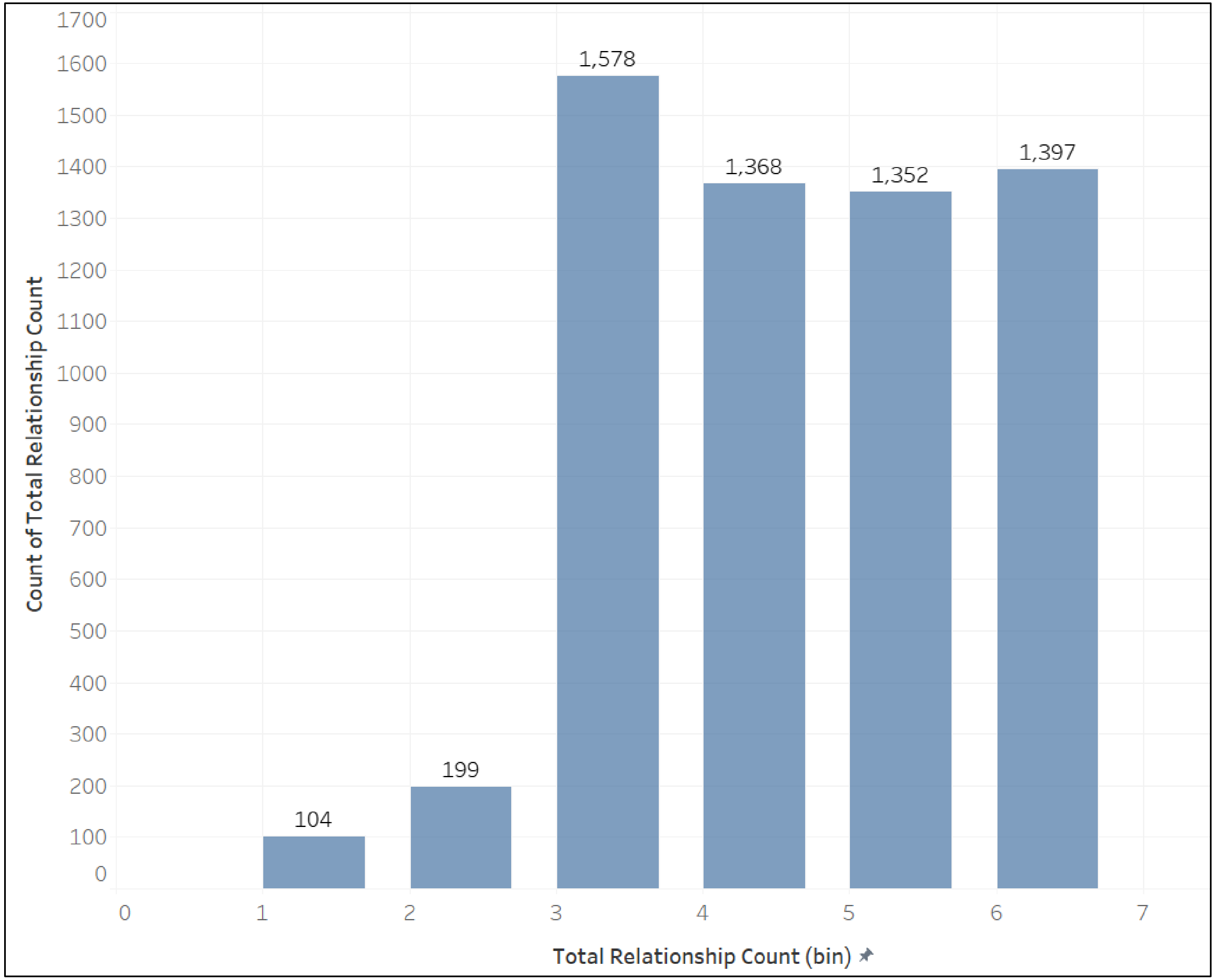


Total Relationship Count

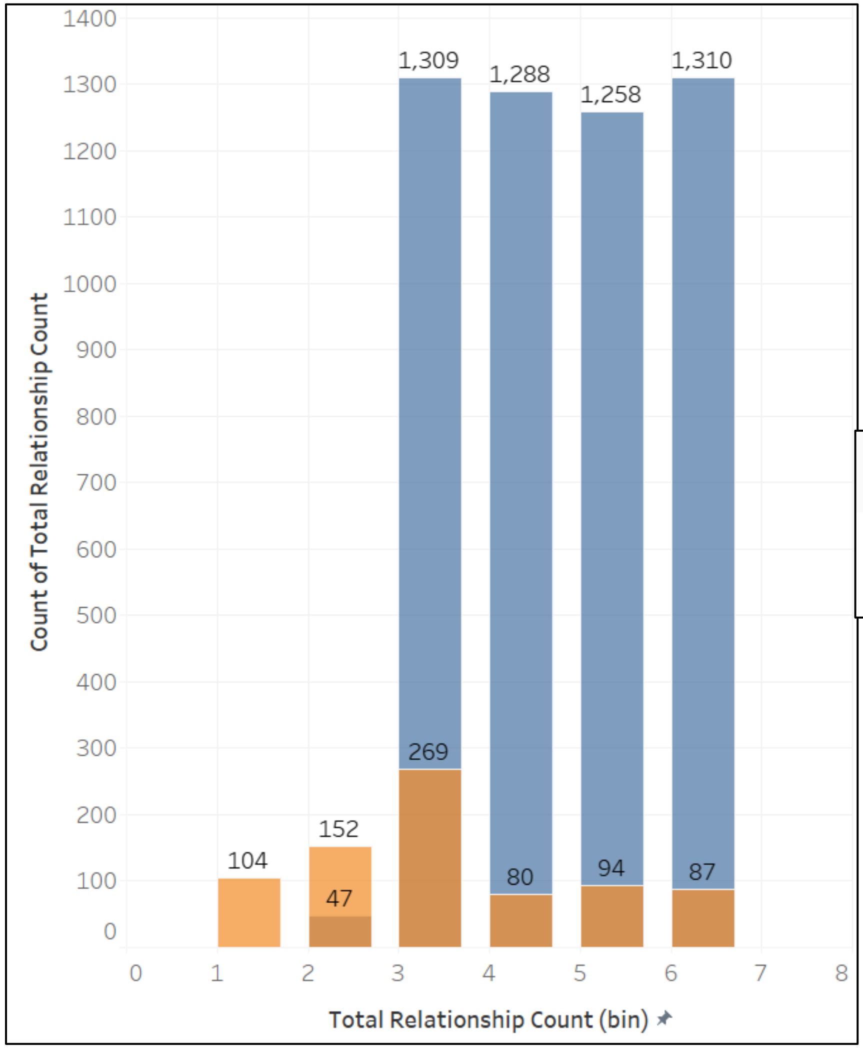
(No. of products/services the customer/s is/are using)

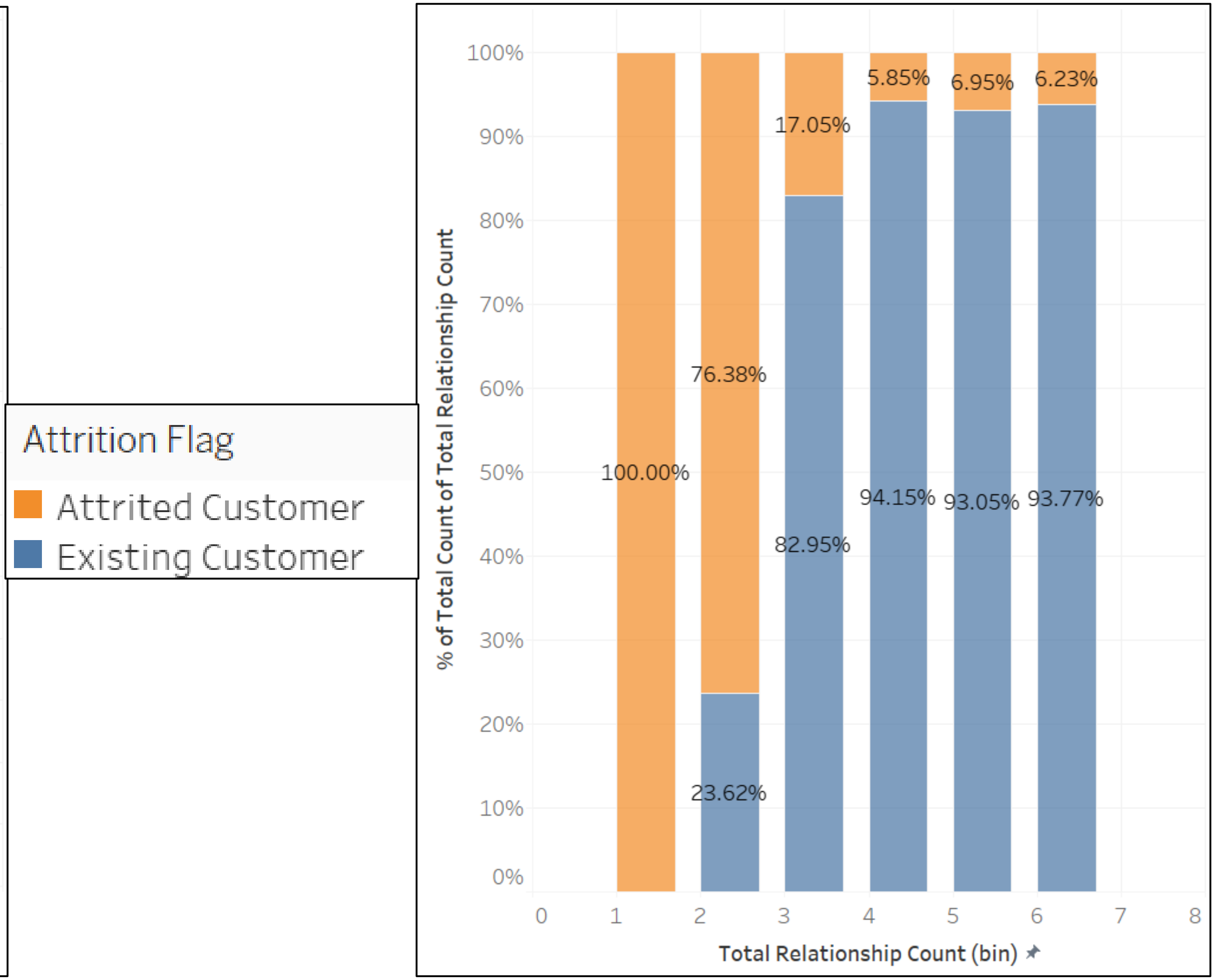
Discrete Categorical

Around **95%** customers use products/services more than 2



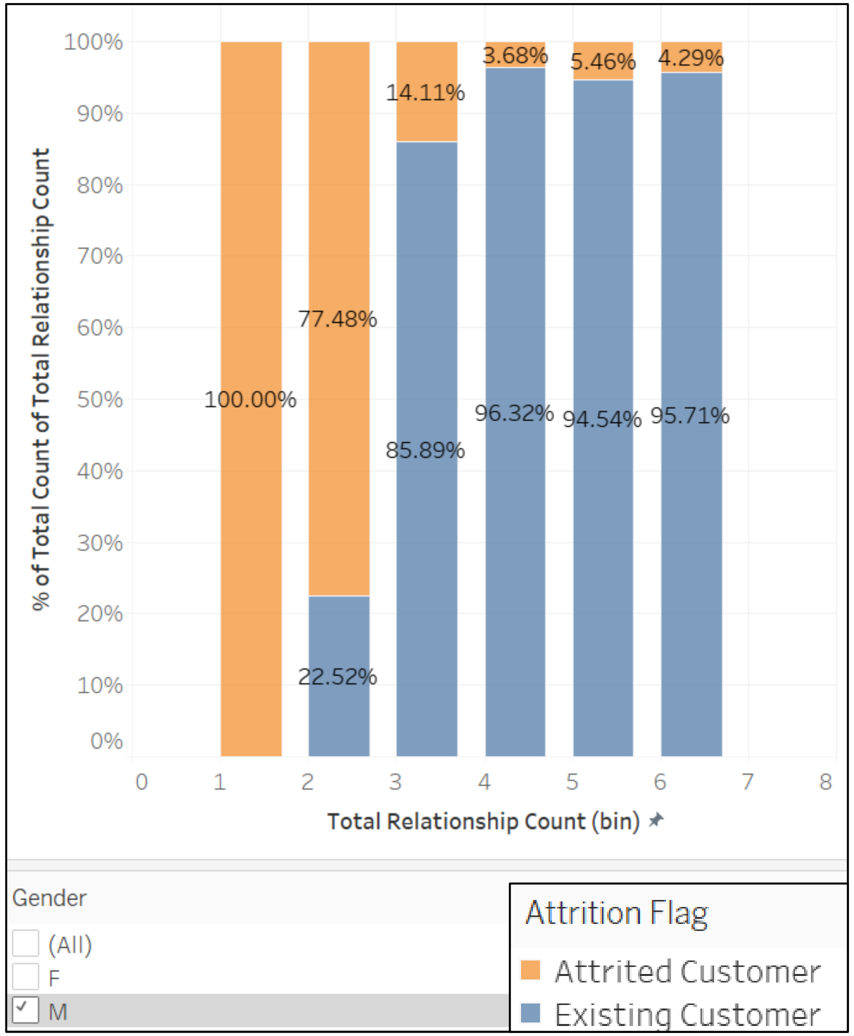
Total Relationship Count

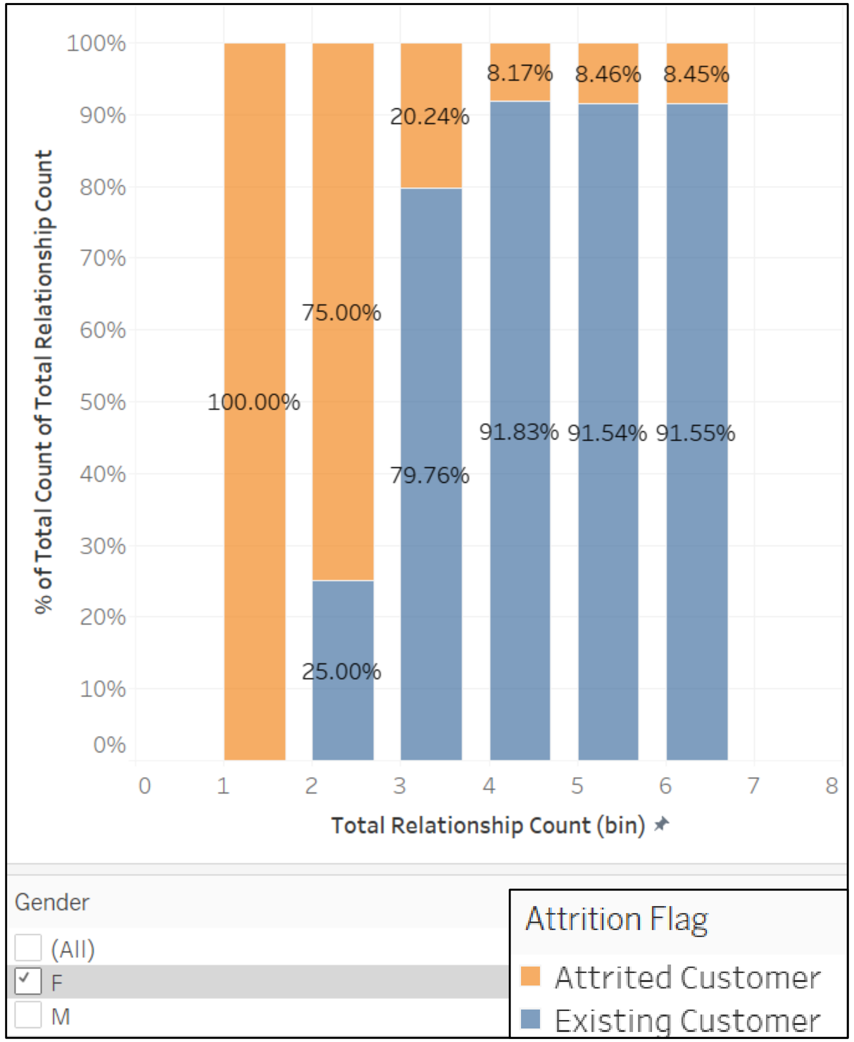




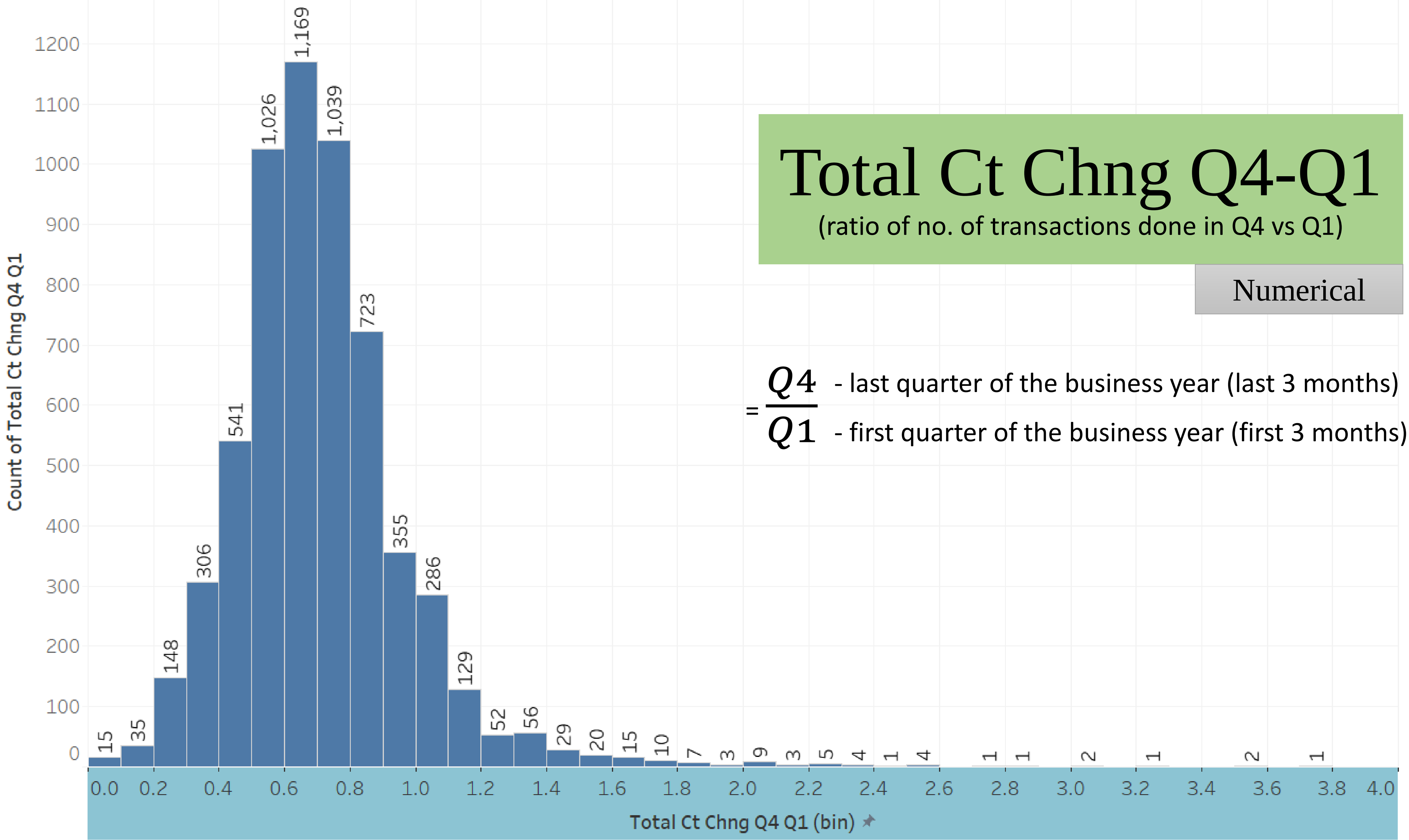
Customers using products/services count-1 or 2 are churning

Total Relationship Count





Either Male or Female customer using products/services count-1 or 2 are churning



Total Ct Chng Q4-Q1

(ratio of no. of transactions done in Q4 vs Q1)

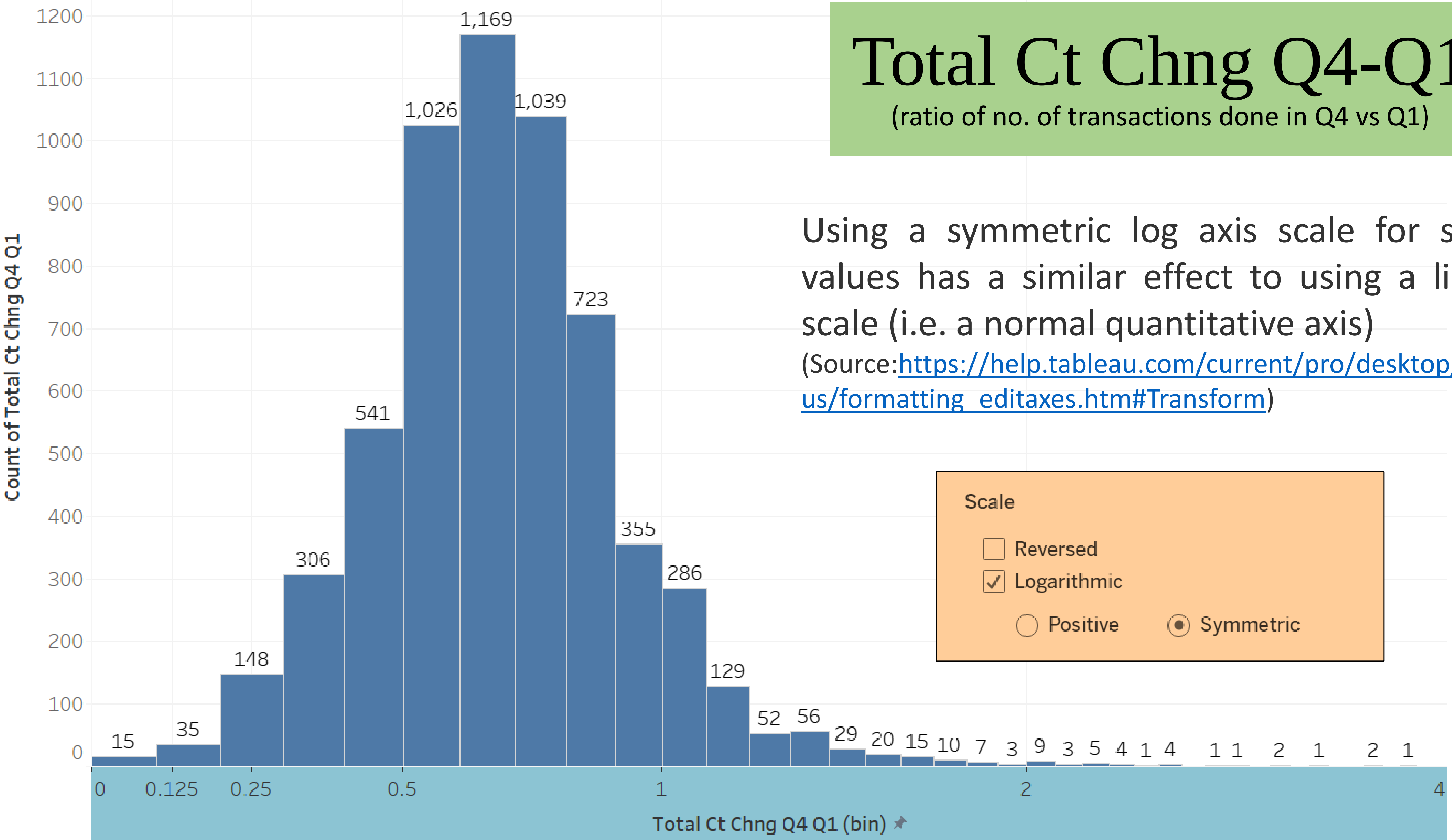
Numerical

$$= \frac{Q4}{Q1}$$

- Q4 - last quarter of the business year (last 3 months)
- Q1 - first quarter of the business year (first 3 months)

Total Ct Chng Q4-Q1

(ratio of no. of transactions done in Q4 vs Q1)



Using a symmetric log axis scale for small values has a similar effect to using a linear scale (i.e. a normal quantitative axis)

(Source:https://help.tableau.com/current/pro/desktop/en-us/formatting_editaxes.htm#Transform)

Scale

Reversed

Logarithmic

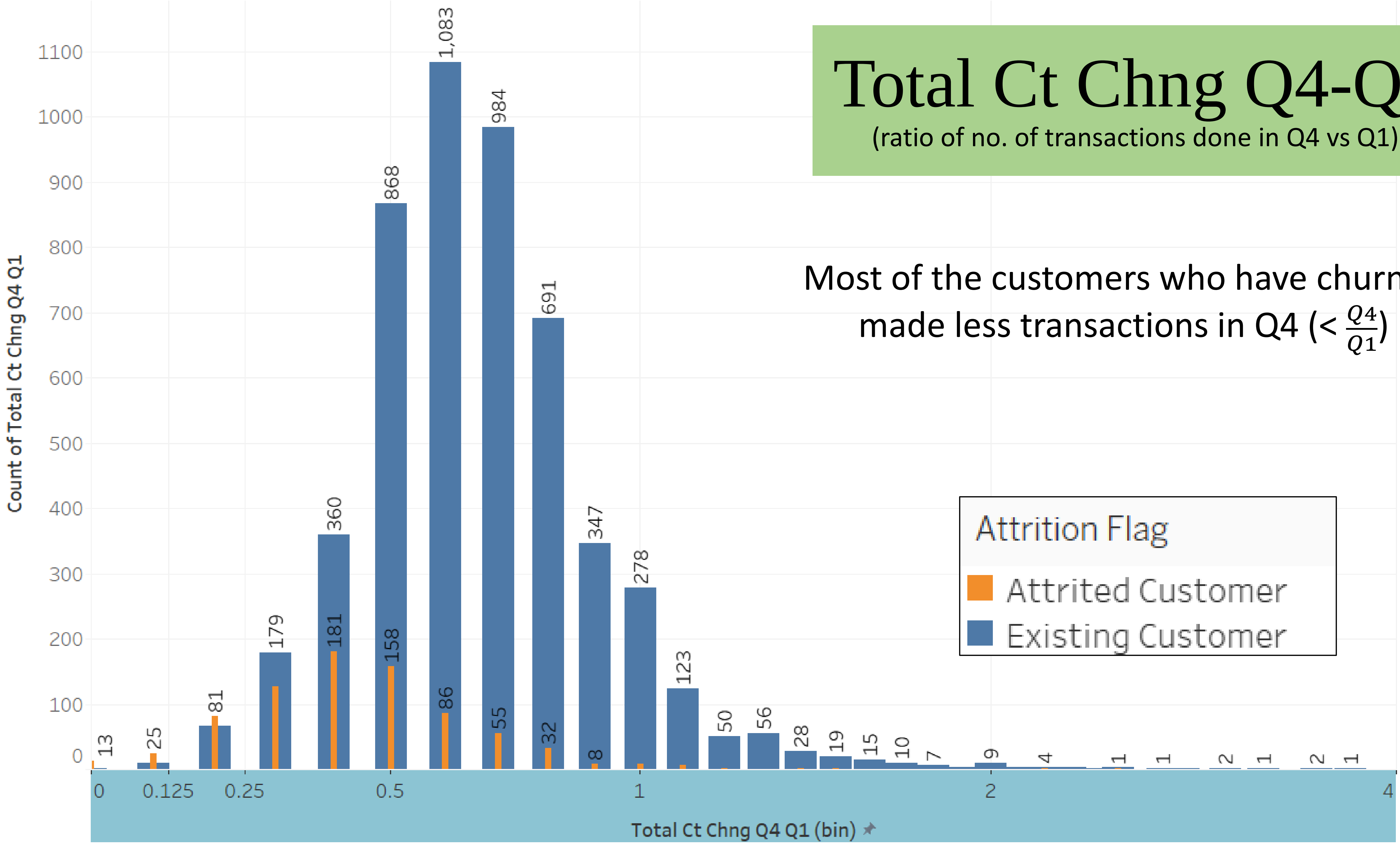
Positive

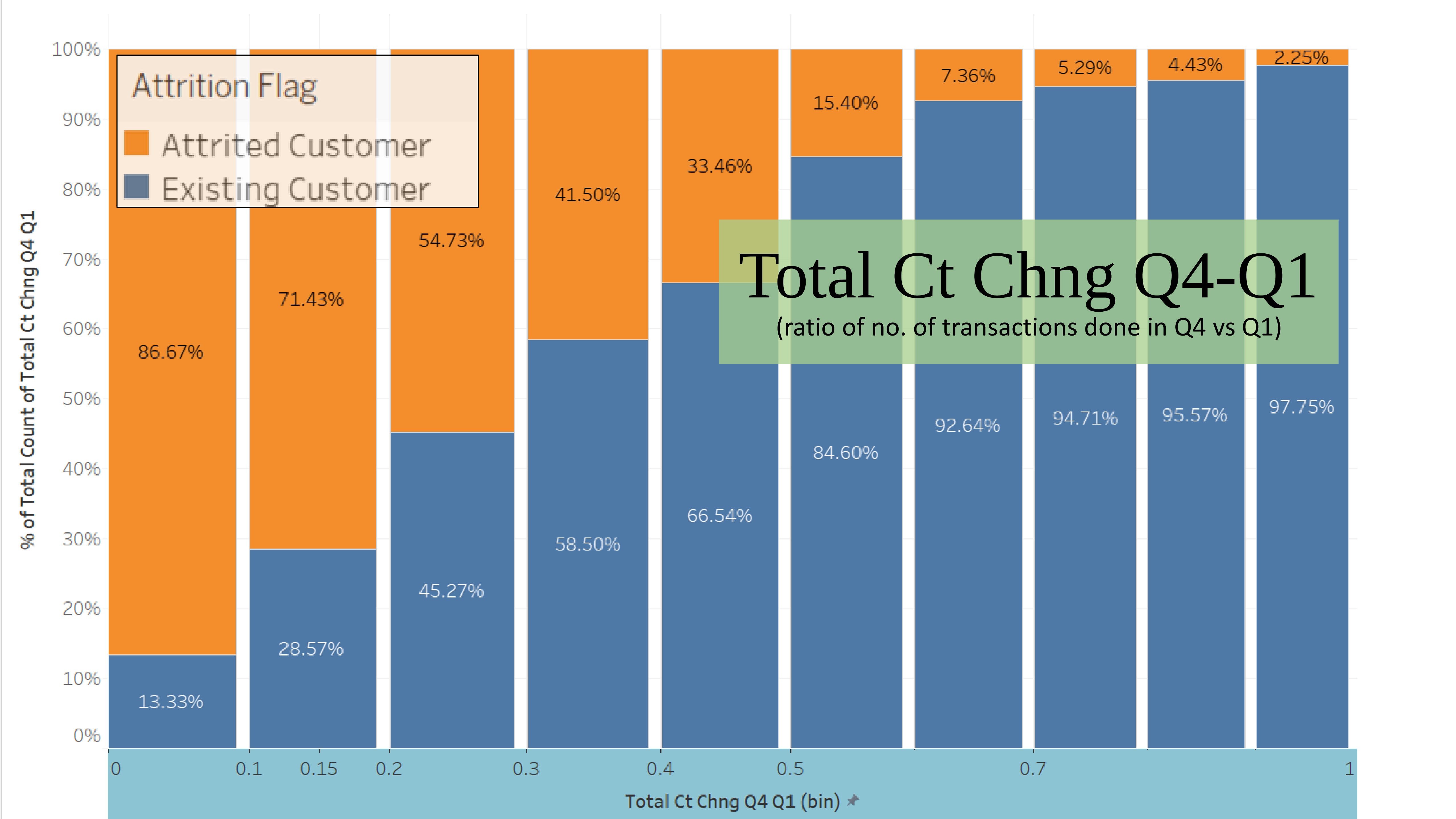
Symmetric

Total Ct Chng Q4-Q1

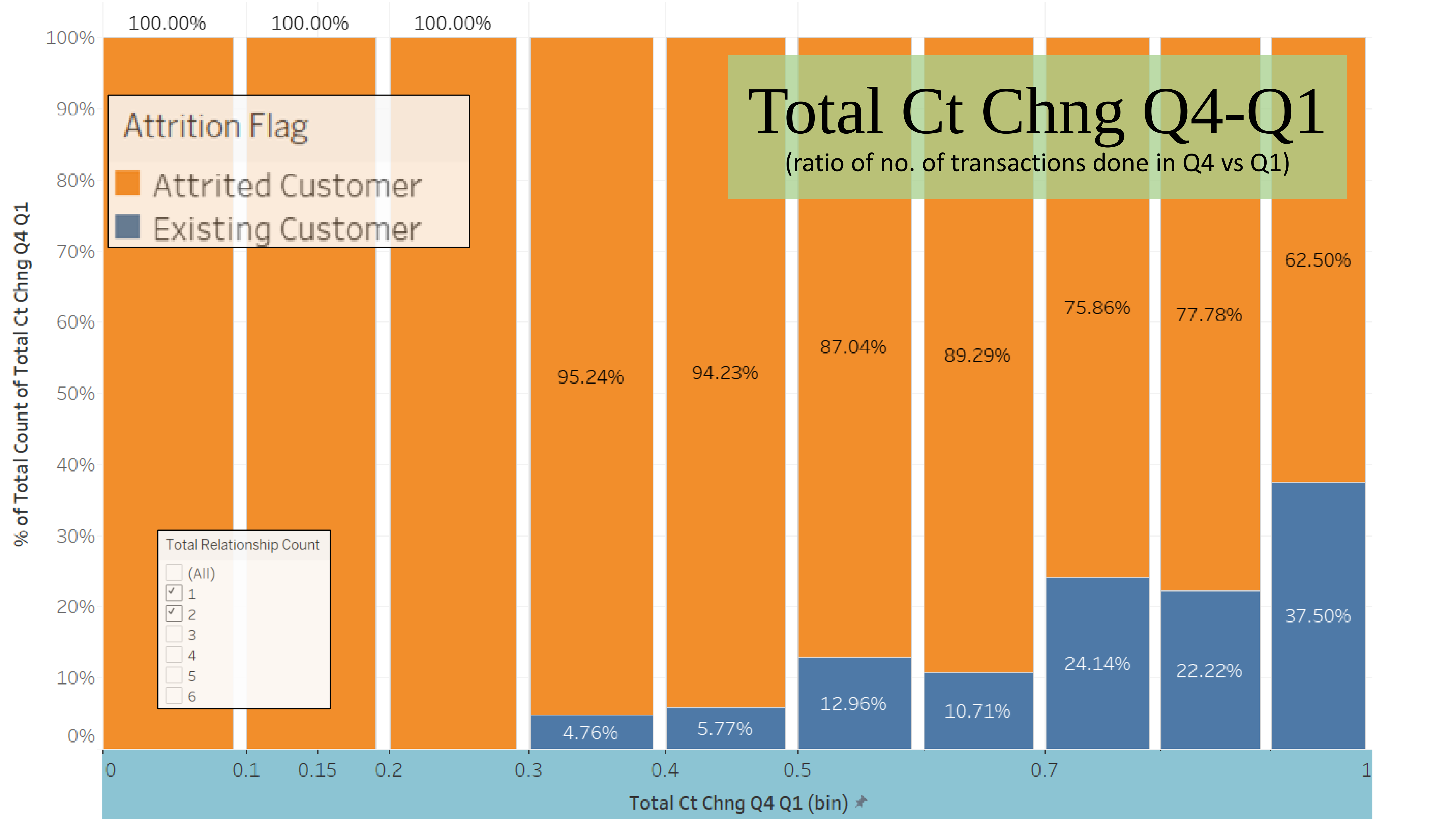
(ratio of no. of transactions done in Q4 vs Q1)

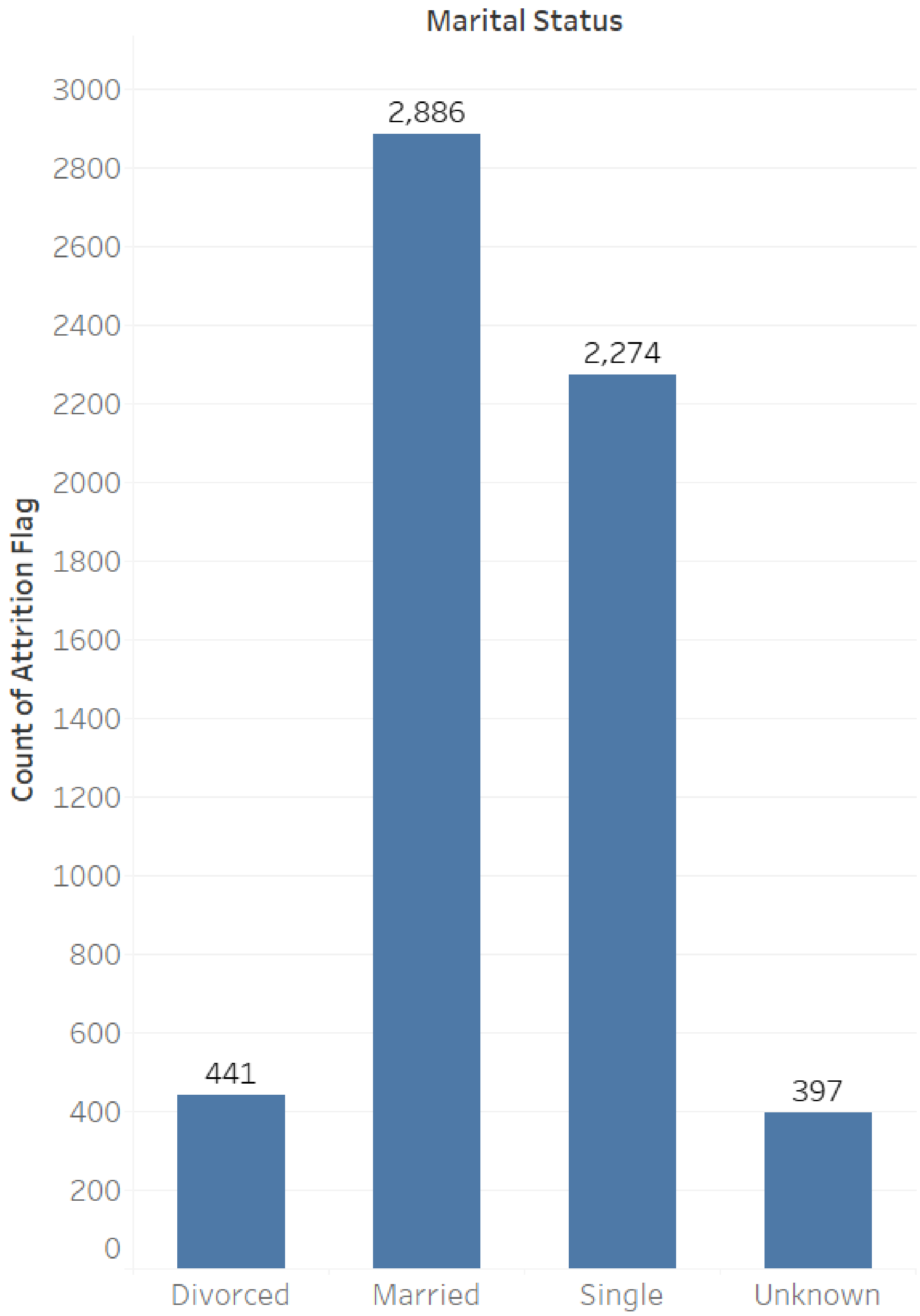
Most of the customers who have churned made less transactions in Q4 ($< \frac{Q4}{Q1}$)





Total Ct Chng Q4-Q1
(ratio of no. of transactions done in Q4 vs Q1)
Attrition Flag
Attrited Customer
Existing Customer
% of Total Count of Total Ct Chng Q4 Q1
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
86.67%
13.33%
71.43%
28.57%
54.73%
45.27%
41.50%
58.50%
33.46%
66.54%
15.40%
84.60%
7.36%
92.64%
5.29%
94.71%
4.43%
95.57%
2.25%
97.75%
0
0.1
0.15
0.2
0.3
0.4
0.5
0.7
1
Total Ct Chng Q4 Q1 (bin)

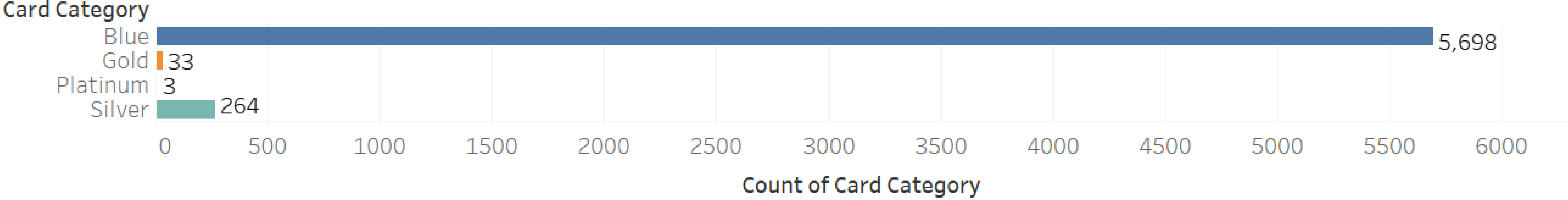

Total Ct Chng Q4-Q1
(ratio of no. of transactions done in Q4 vs Q1)
Attrition Flag
Attrited Customer
Existing Customer
Total Relationship Count
(All)
1
2
3
4
5
6
100.00%
100.00%
100.00%
95.24%
94.23%
87.04%
89.29%
75.86%
77.78%
62.50%
4.76%
5.77%
12.96%
10.71%
24.14%
22.22%
37.50%
% of Total Count of Total Ct Chng Q4 Q1
Total Ct Chng Q4 Q1 (bin)
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
0
0.1
0.15
0.2
0.3
0.4
0.5
0.7
1



Marital Status

Categorical

- Nearly **86%** of the customers are Married or Single
- **6%** belongs to unknown category



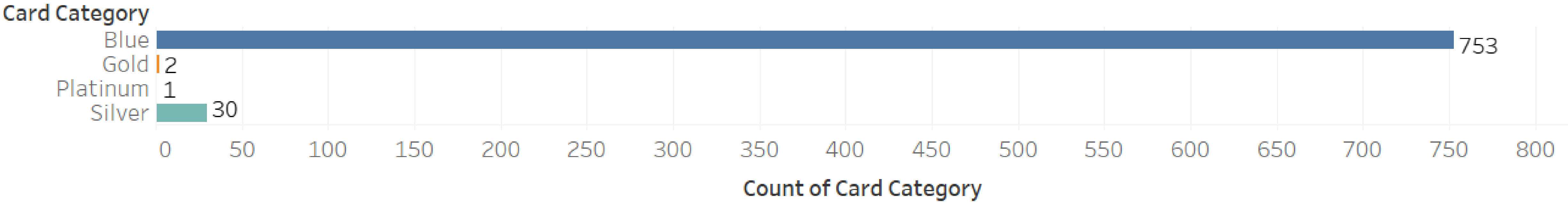
Blue Card Category contributes the highest no. of customers (**95%**) which are churning

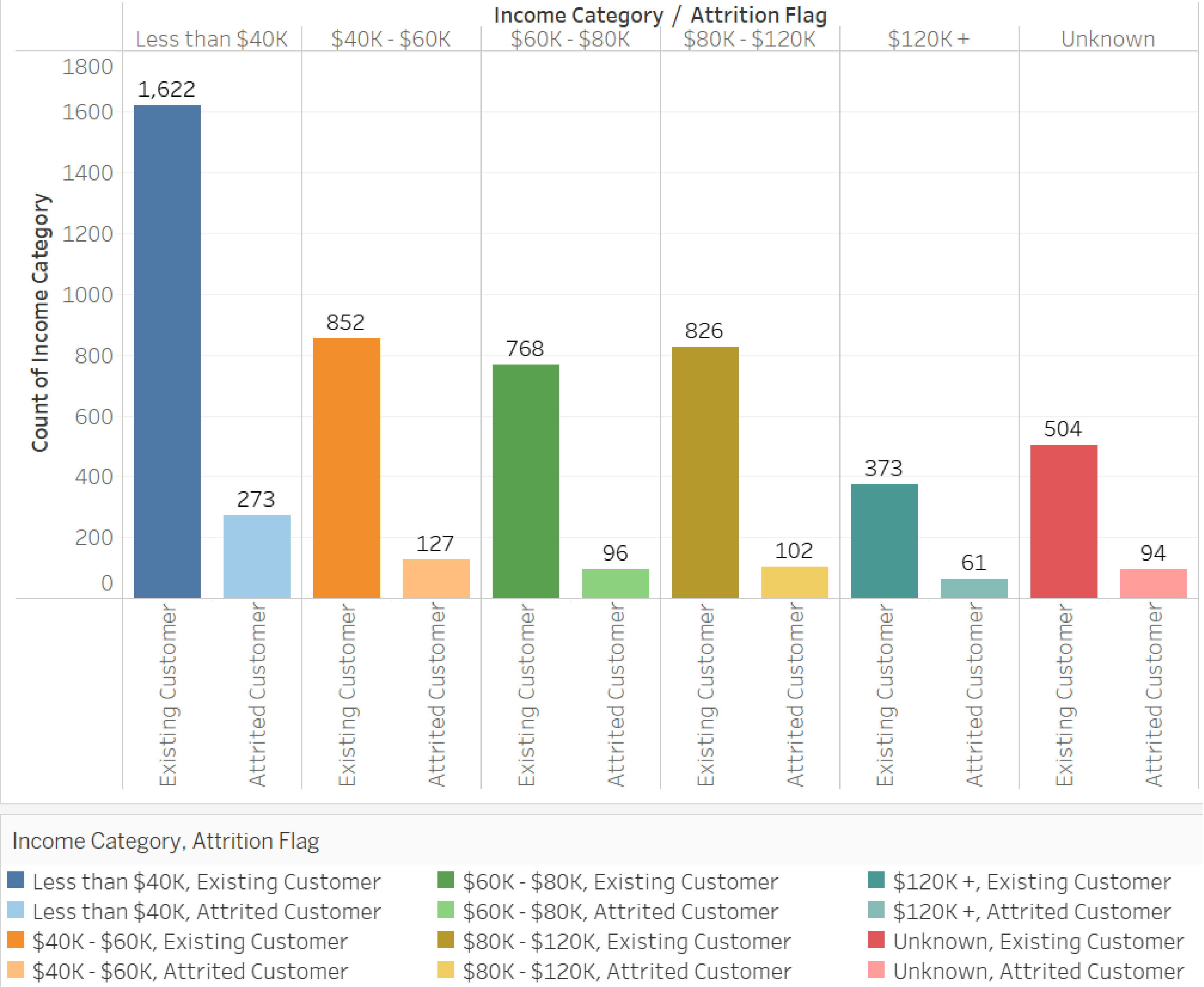
Card Category

Categorical

Attrition Flag

- [ ] (All)
- [x] Attrited Customer
- [ ] Existing Customer





Income Category

Categorical

Conclusion

- The analysis of attrition of customers with the total relationship count, total ct change Q4-Q1 & other features like income category, card category, etc shows the churners leaving the bank services. Thus, the increment in the credit limit of the customers & improving services for customers might help to prevent customers from leaving the credit card service of the bank
- The analysis of this data establishes a trend inferring that the Blue Card Category is the customers favorite, majority of whom are churners in Female category within the Income bracket of less than $40k
- After the analysis as well as through test conducted on Auto model in Rapid Miner, models which predict with high accuracy rate are :
  - Deep Learning (92%)
  - Logistic Regression (90%)
  - Naive Bayes (89%)