CREDIT CARD CUSTOMER CHURN ANALYSIS

Team 4:

- Will Cutrone
- Edyson Katuuo
- Jess Hunter
- Subhashree Mallick

MOTIVATION AND SUMMARY

- Customer churn is one of the most important and challenging problems for businesses today

- Many businesses are utilizing customer churn metrics in order to try and predict churning and improve customer retention.

QUESTIONS

What factors may or may not be **associated with churning**?

What is the profile of someone **who is likely to churn**?

After determining the profile, which percent of existing customers are now **at risk for churning**?

DATA CLEANUP & EXPLORATION

- Data Source: BankChurners.csv via Kaggle.com
- Pandas was used to create a dataframe from the csv
- In terms of cleaned data, the most significant edit to the data was that we eliminated all customers with a utilization ratio of 0 (aka they never used the card)
- Keeping zero utilization customer data led to misleading representations
- Two extraneous columns were deleted as they did not contribute to the analysis

DATA ANALYSIS

- Each variable (or column in the datasheet) was investigated whether it did or did not have a statistically significant effect in predicting those at risk for churning via using Matplotlib to create various diversified plots

- After plots were created by using the independent t-test on all factors, we determined which factors were statistically significant and came up with a basic profile of a customer that we believe is at risk for churning

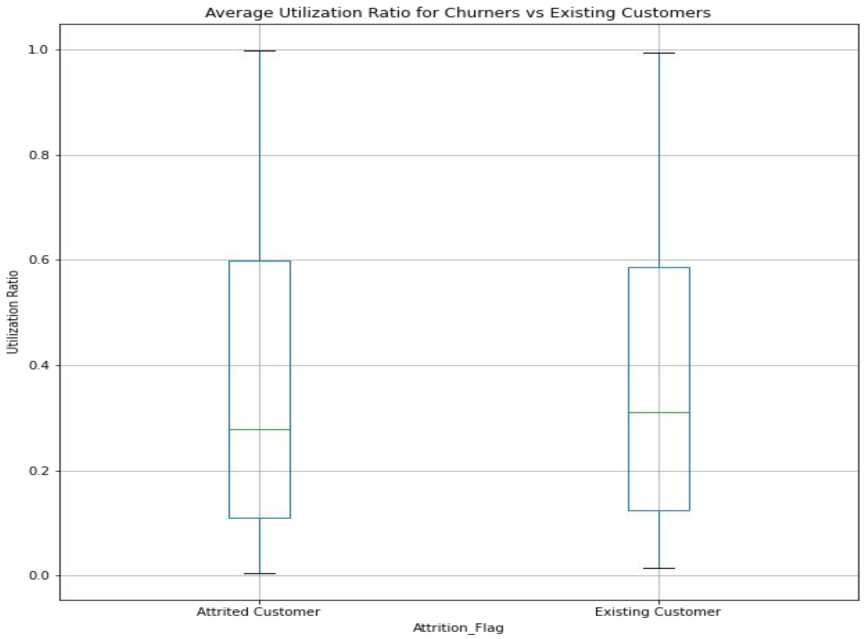
VARIABLES EXAMINED FOR INVESTIGATION

- Age
- Income
- Utilization Ratio
- Transaction Amount
- Transaction Count
- Months Inactive
- Gender
- Marital Status
- Dependents
- Monthly Credit Limit
- Months with the Company

QUANTITATIVE DATA

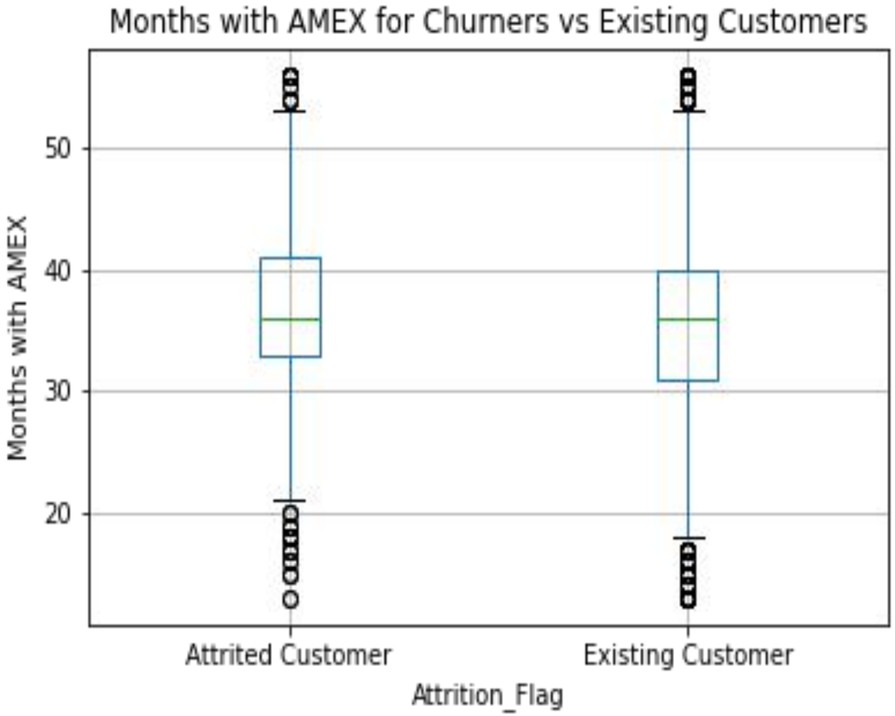
- Utilization Ratio
- Months Spent
- Monthly Credit Limit
- Total Revolving Balance

UTILIZATION RATIO



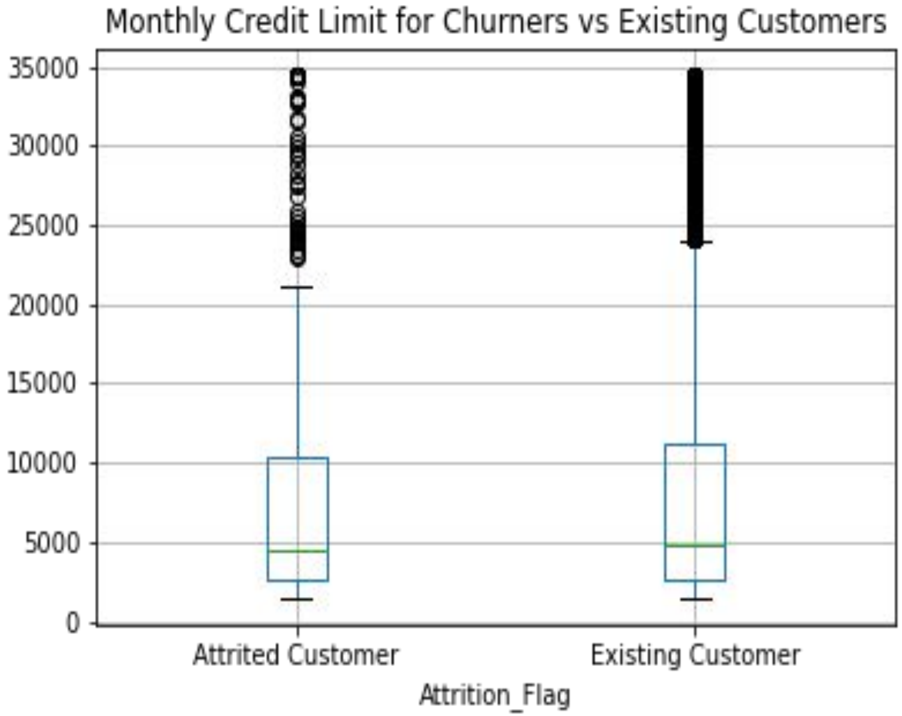
P-value: 0.73

MONTHS SPENT



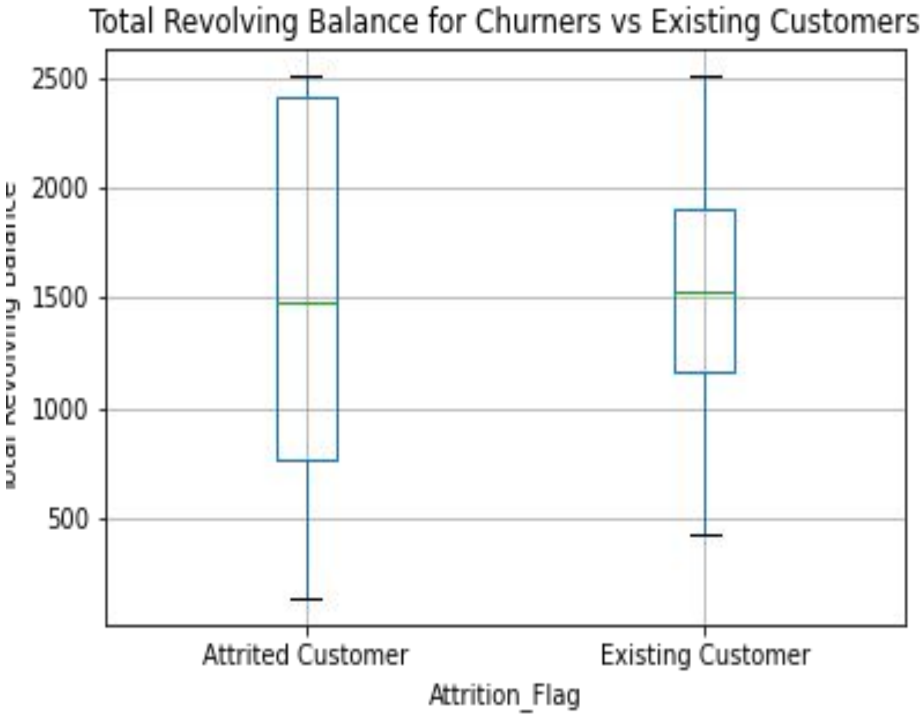
P-value: 0.11

MONTHLY CREDIT LIMIT



P-value: 0.73

TOTAL REVOLVING BALANCE

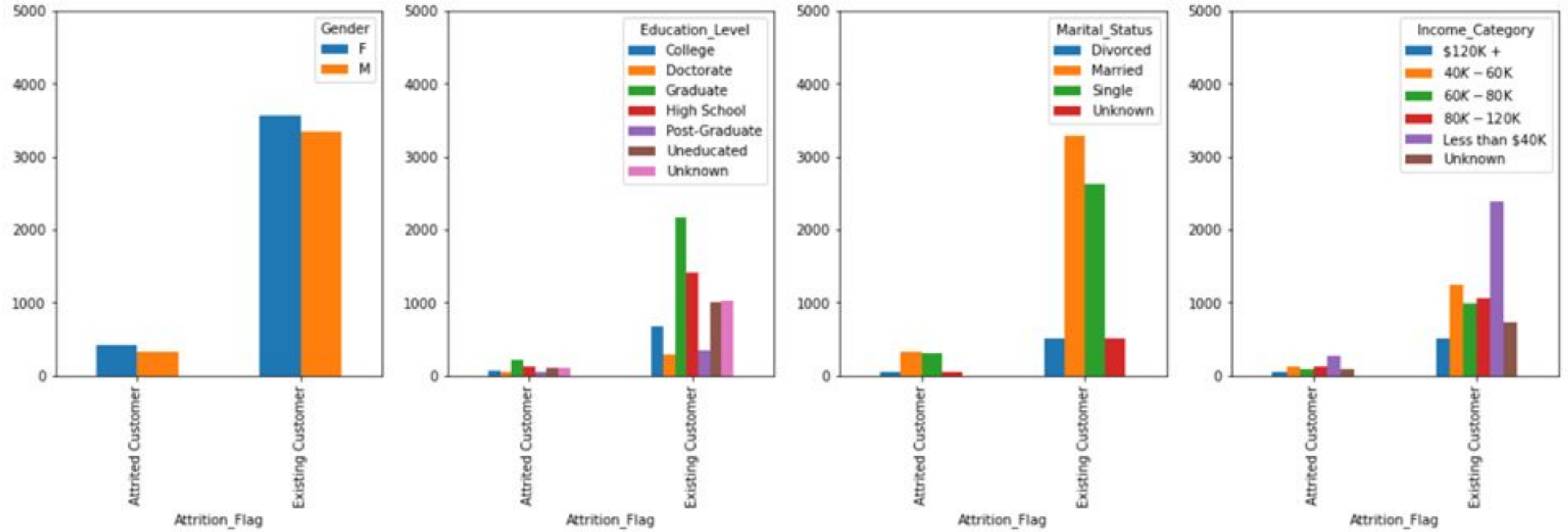


P-value: 0.09

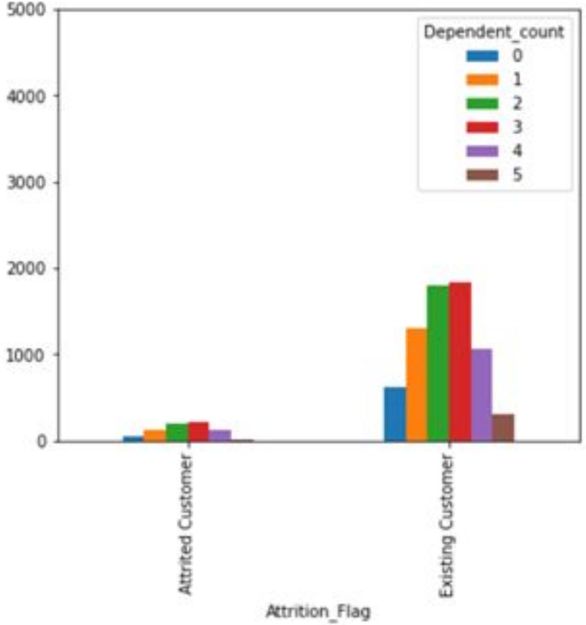
CATEGORICAL DATA

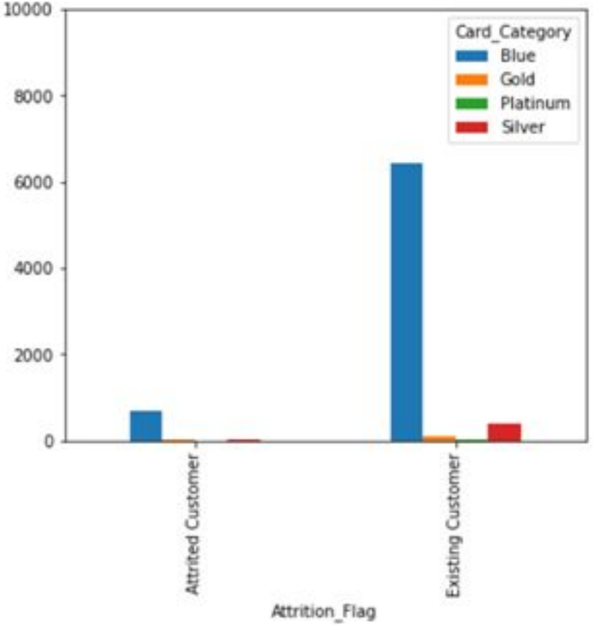
- Gender
- Education Level
- Marital Status
- Income_Category
- Dependent Count
- Card Category
- Customer Age

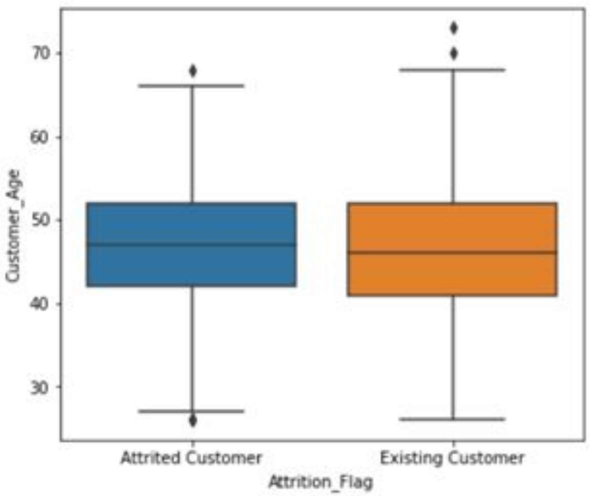
GENDER, EDUCATION LEVEL, MARITAL STATUS & INCOME



DEPENDENT COUNT, CARD CATEGORY & AGE



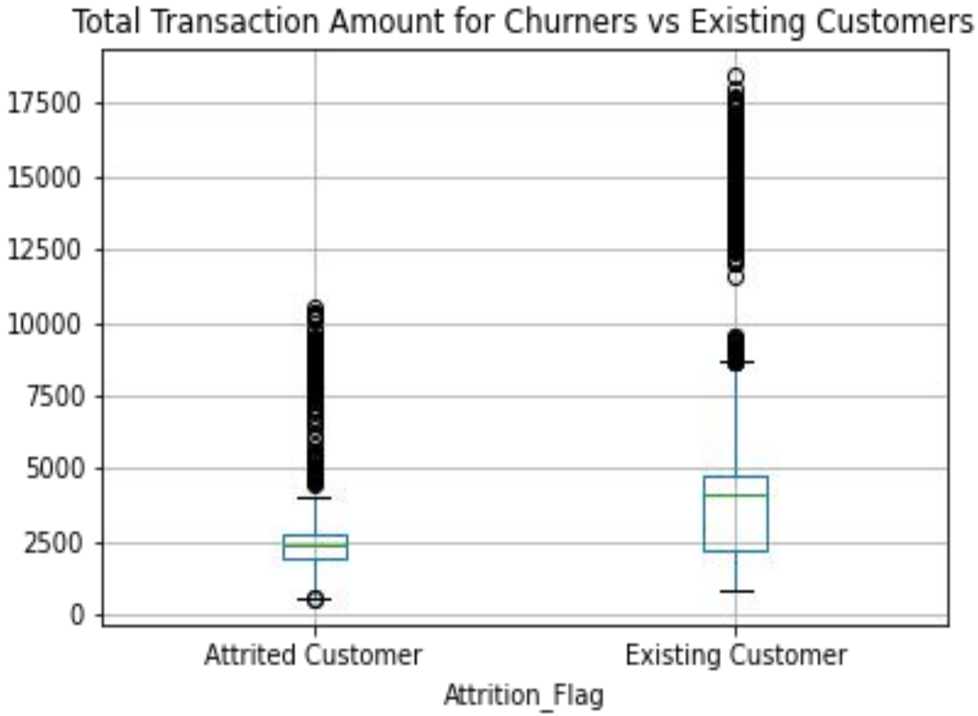




STATISTICALLY SIGNIFICANT AREAS

- Total Transaction Amount
- Total Transaction Count
- Months Inactive Out of 12 Months

TOTAL TRANSACTION AMOUNT



Churners

Mean= $3,148.97

Median= $2,361.00

Std = $2,357

Existing

Mean = $4,861.27

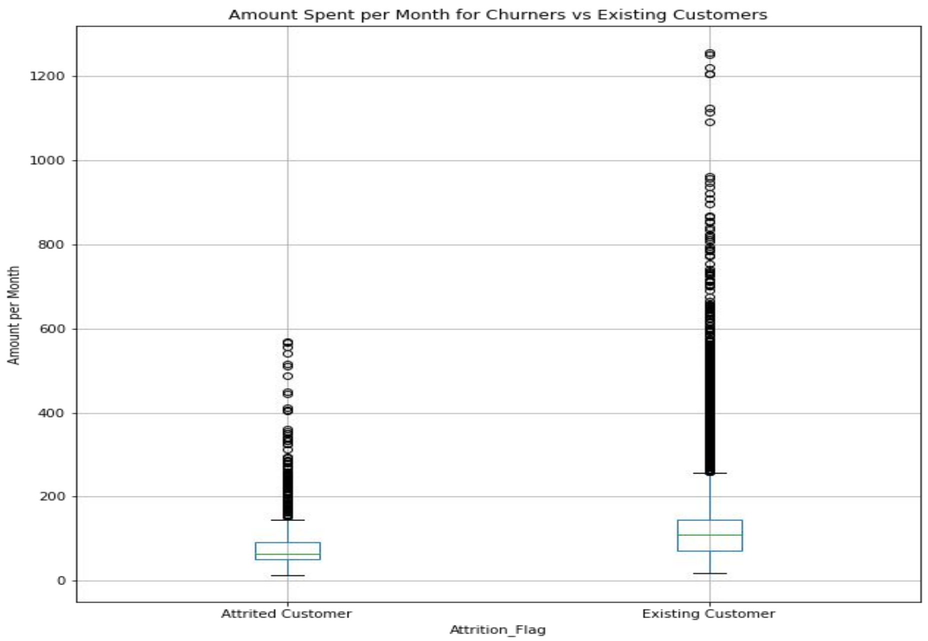
Median = $4,081.00

Std = $3611.78

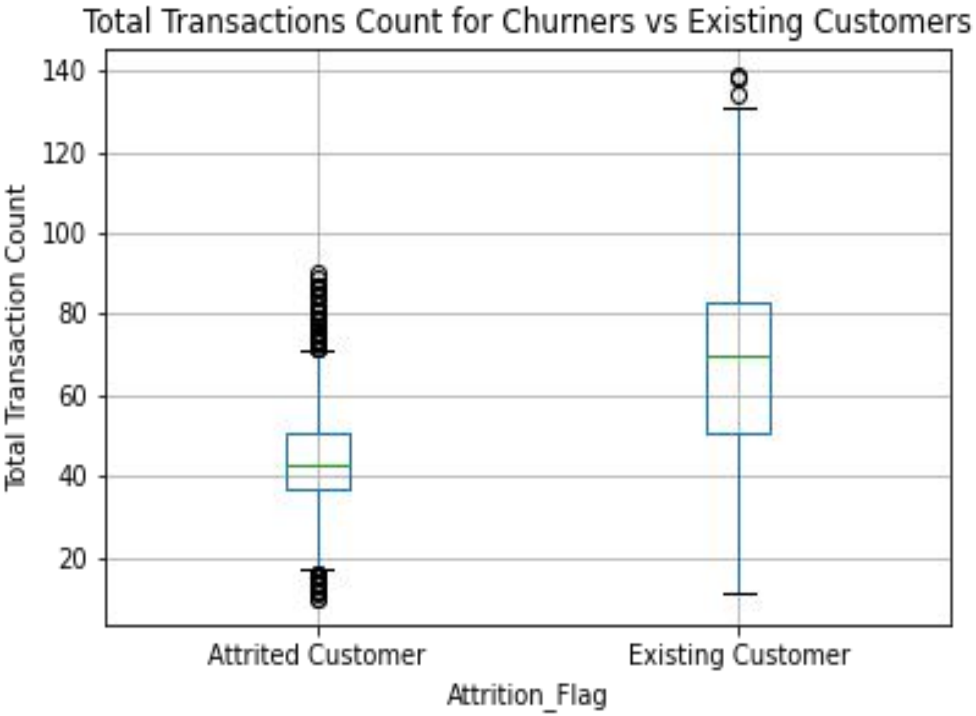
P-value

9.9e-51

TRANSACTION AMOUNT (BY MONTH)



TOTAL TRANSACTION COUNT

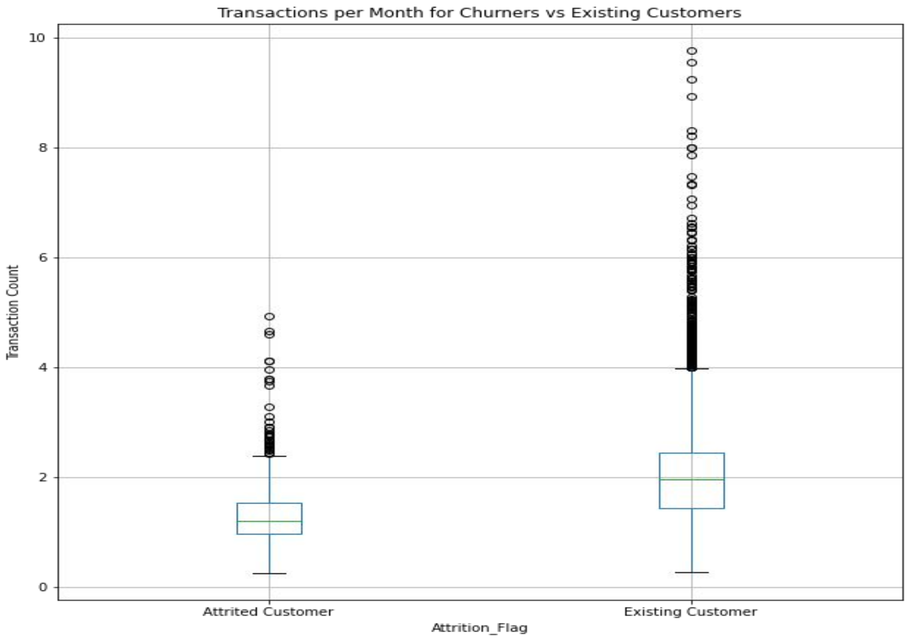


<u>Churners</u>
Mean: 45
Median: 43
Standard Deviation: 14.38

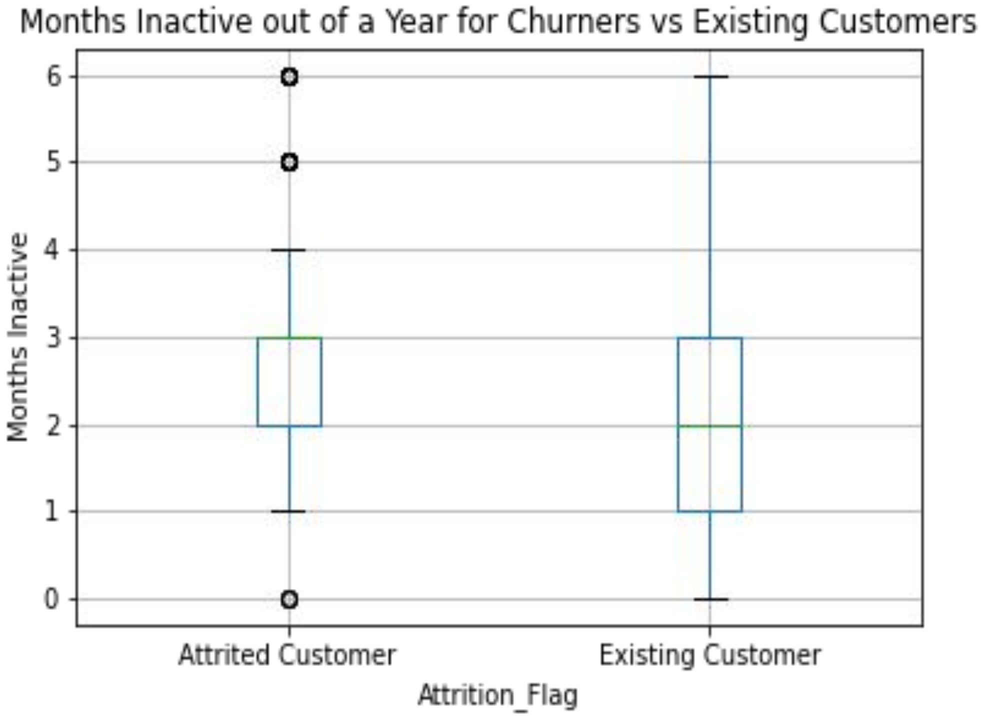
<u>Existing Customers</u>
Mean: 68
Median: 70
Standard Deviation: 24

<u>P-value</u>
6.9e-212

TRANSACTION COUNT (BY MONTH)



MONTHS INACTIVE



Churners
Mean: 2.73
Median: 3
Standard Deviation: 0.9

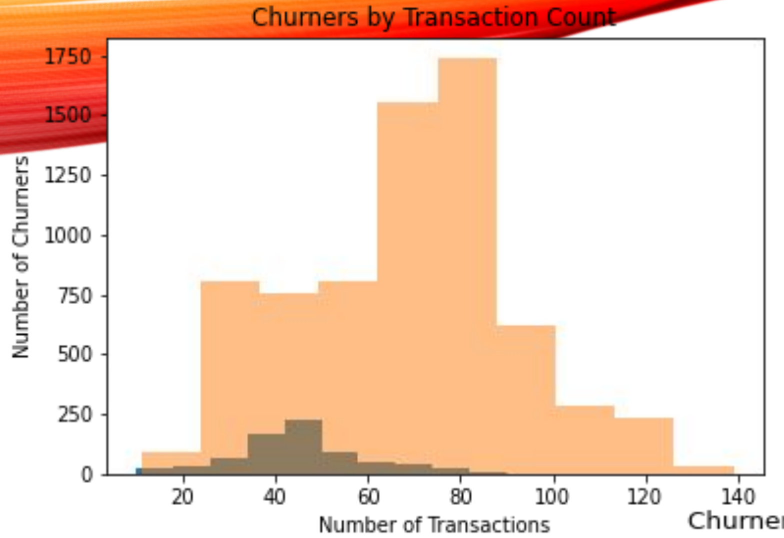
Existing Customers
Mean: 2.27
Median: 2
Standard Deviation: 1.02

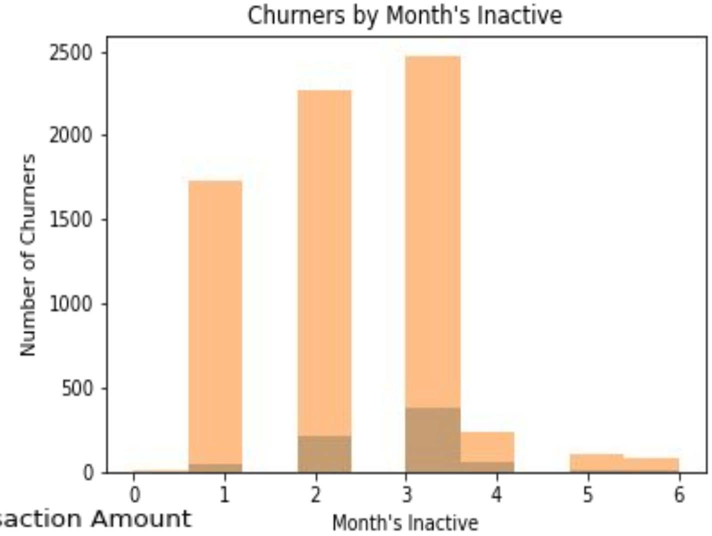
P-value
2.8e-35

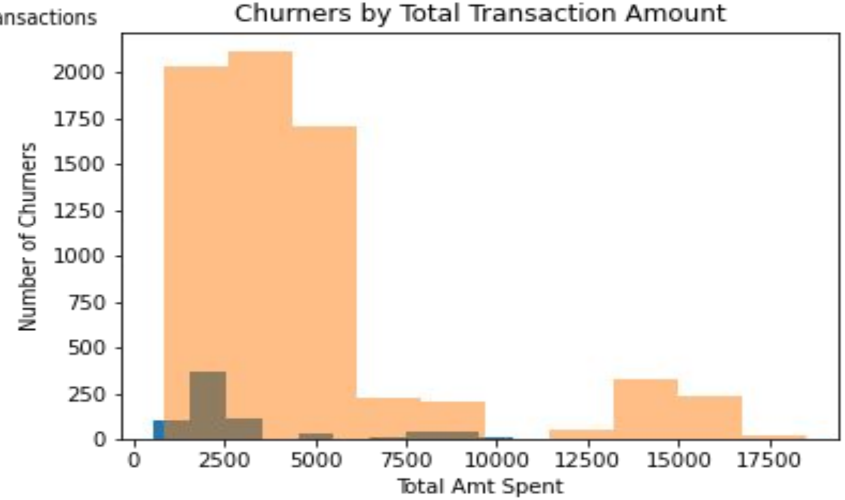
PROFILE OF A CHURNER

Total Trans Count <= 51

Months Inactive >= 3


Churners by Transaction Count
Number of Churners
Number of Transactions
0
250
500
750
1000
1250
1500
1750
20
40
60
80
100
120
140


Churners by Month's Inactive
Number of Churners
Month's Inactive
0
500
1000
1500
2000
2500
0
1
2
3
4
5
6


Churners by Total Transaction Amount
Number of Churners
Total Amt Spent
0
250
500
750
1000
1250
1500
1750
2000
0
2500
5000
7500
10000
12500
15000
17500

RISK OF EXISTING CUSTOMER CHURN

- Existing customers who meet both conditions - 1%

- Existing customers who meet one of the two conditions - 68%

LIMITATIONS

- The dataset is quite large, containing data from more than 10,000 credit card accounts with 19 variables
- Data may be skewed by card category (revolving balance cards vs monthly pay in full cards)
- The dataset is missing variables that would have been valuable in our analysis

ADDITIONAL AREAS WORTH EXPLORING

- Some important variables were not included in the dataset but would be worth exploring:
  - Customer credit score
  - Annual percentage rate per card
  - Time period of dataset collection
  - Geographic location of the customer
  - Special offer such as 0% or low APR for a particular time period

QUESTIONS?