

BANK CUSTOMER CHURN REDUCTION

CZ1016 PROJECT - CALISTA, CLARISSA, WEI MAY





OUR TASK

identify likely-to-attrite customers before they attrite
and provide insights to reduce attrition rate

Our DataSet

Source

**fictional
dataset from
Kaggle**

[https://www.kaggle.com/sakshi
goyal7/credit-card-customers?
select=BankChurners.csv](https://www.kaggle.com/sakshi goyal7/credit-card-customers?select=BankChurners.csv)

Size

**21 COLUMNS
10127 ROWS**

About

Details about
Customers & their
**ATTRITION
FLAG**





CUSTOMERS' BACKGROUND

Age, Gender, Income Category, Education

CUSTOMERS' TYPE OF PRODUCT

Card Type, Credit Limit, Relationship Count

CUSTOMERS' SPENDING HABITS

Transaction Amount, Transaction Count etc.

DATA CLEANING



Dropped Rows >2 Unknowns

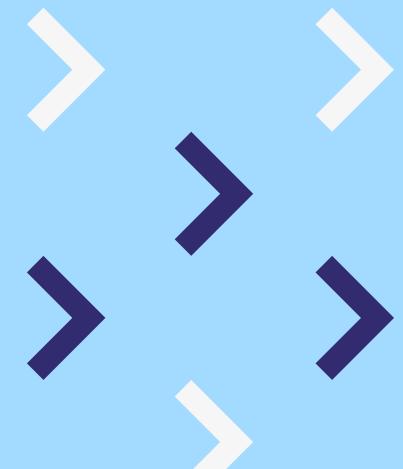
10127 rows to 9800 rows

Dropped Avg Open To Buy

Credit Limit & Avg Open To Buy have perfect correlation

Separate Platinum dataset

only 19/9800 data points and a lot of unknowns



- 5 / 19 attrited
- 4 / 5 are female
- 3 below \$60 K ; 2 Unknown
- Belong in lowest 6 for Total Revolving Balance, Avg Utilisation Ratio

ATTRITED PLATINUM CARDHOLDERS

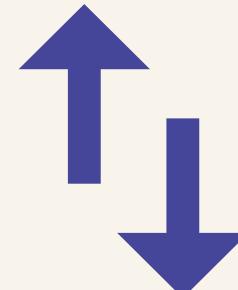


Dealing with Categorical Variables

Education Level



Replaced
'Unknown' with the
mode, 'Graduate'

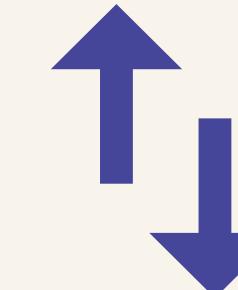


Ordinal Encoding

Income Category



Replaced with mode
depending on
customer's card type

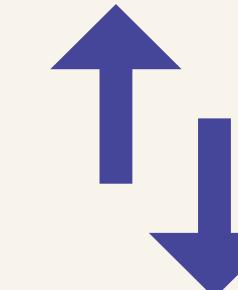


Ordinal Encoding

Marital Status

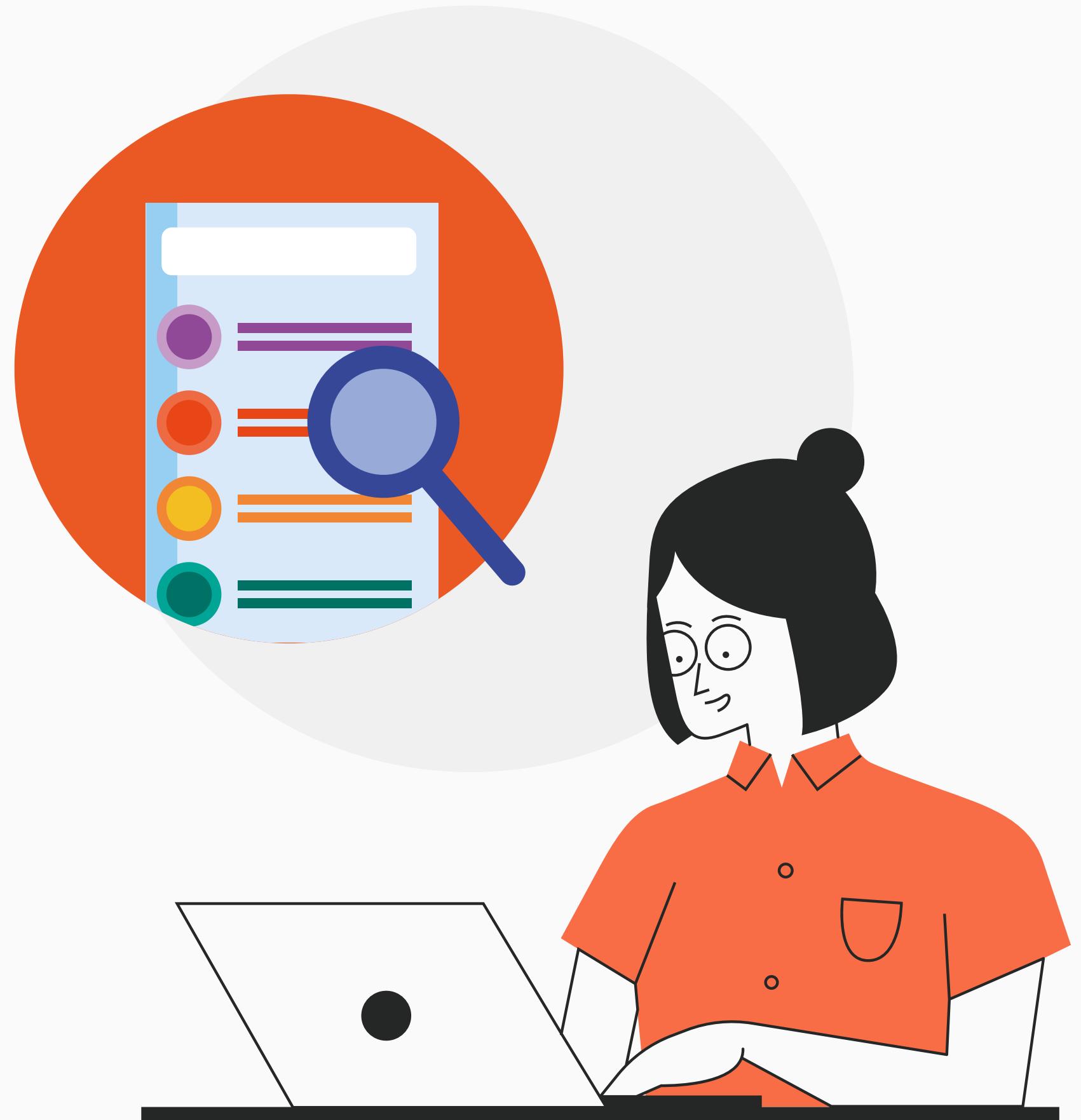


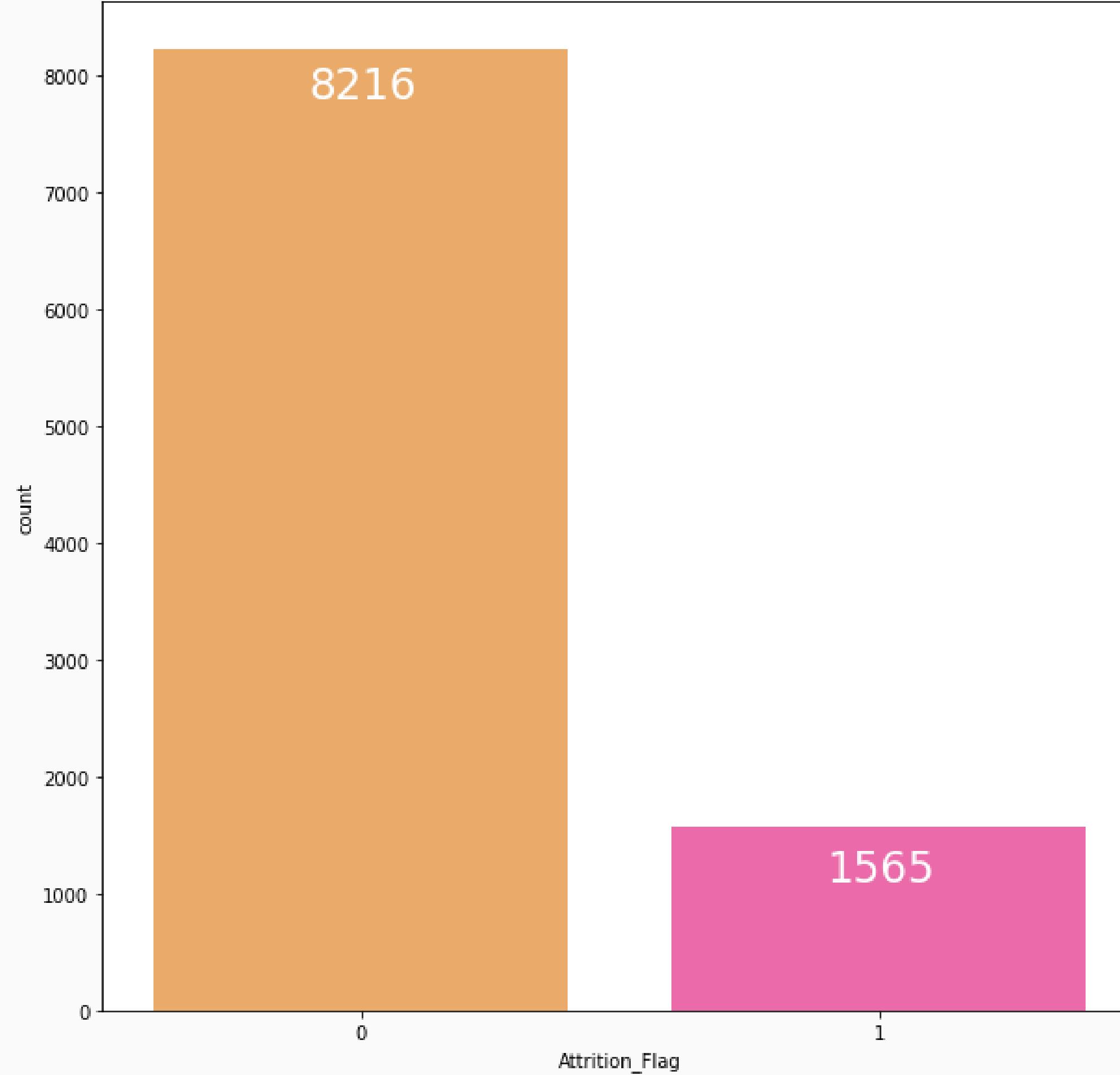
Unable to infer
reasonable link to
replace 'Unknown's



One Hot Encoding

Exploratory Data Analysis





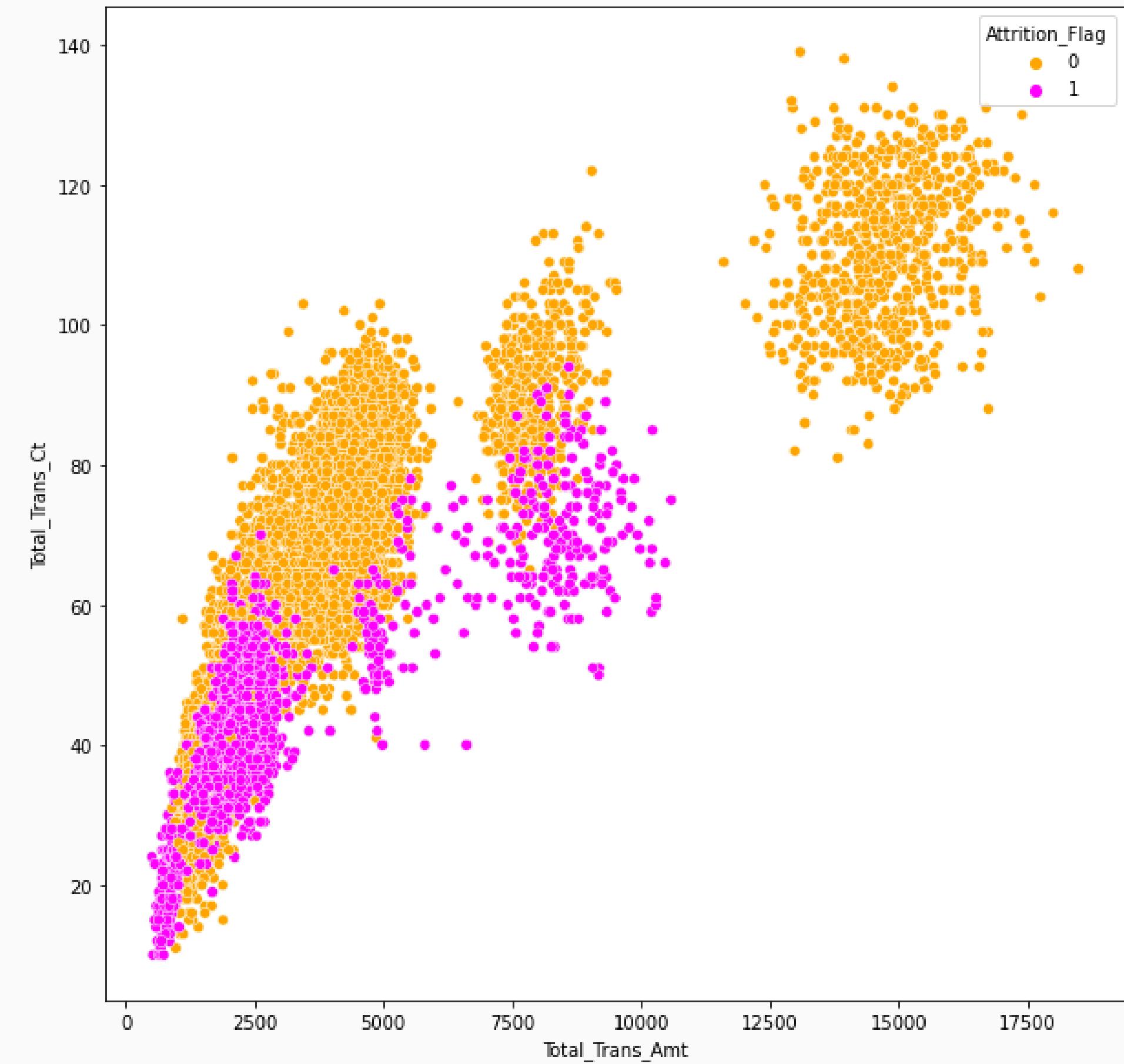
01

Imbalanced Data
16% Attrition

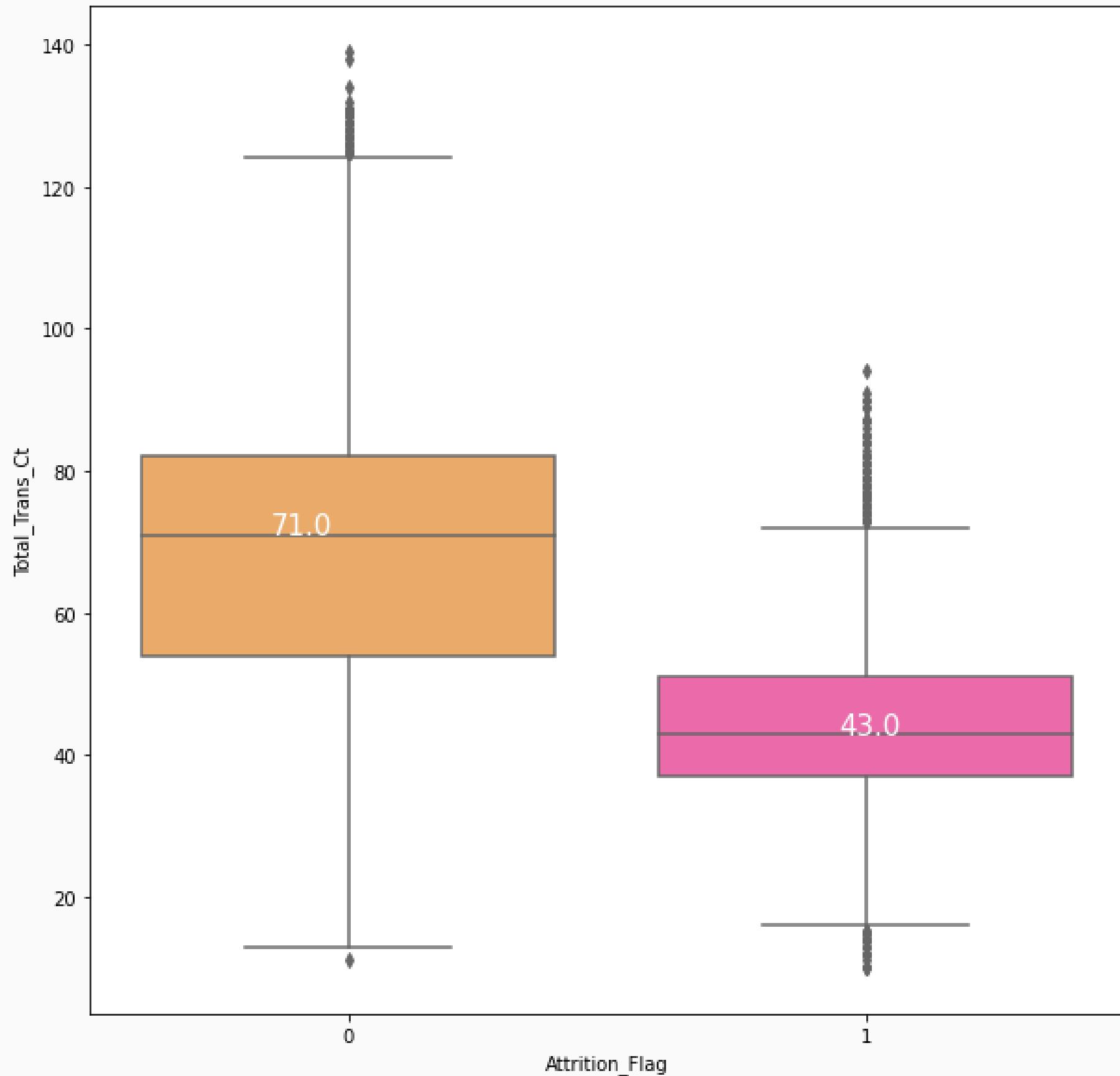


02

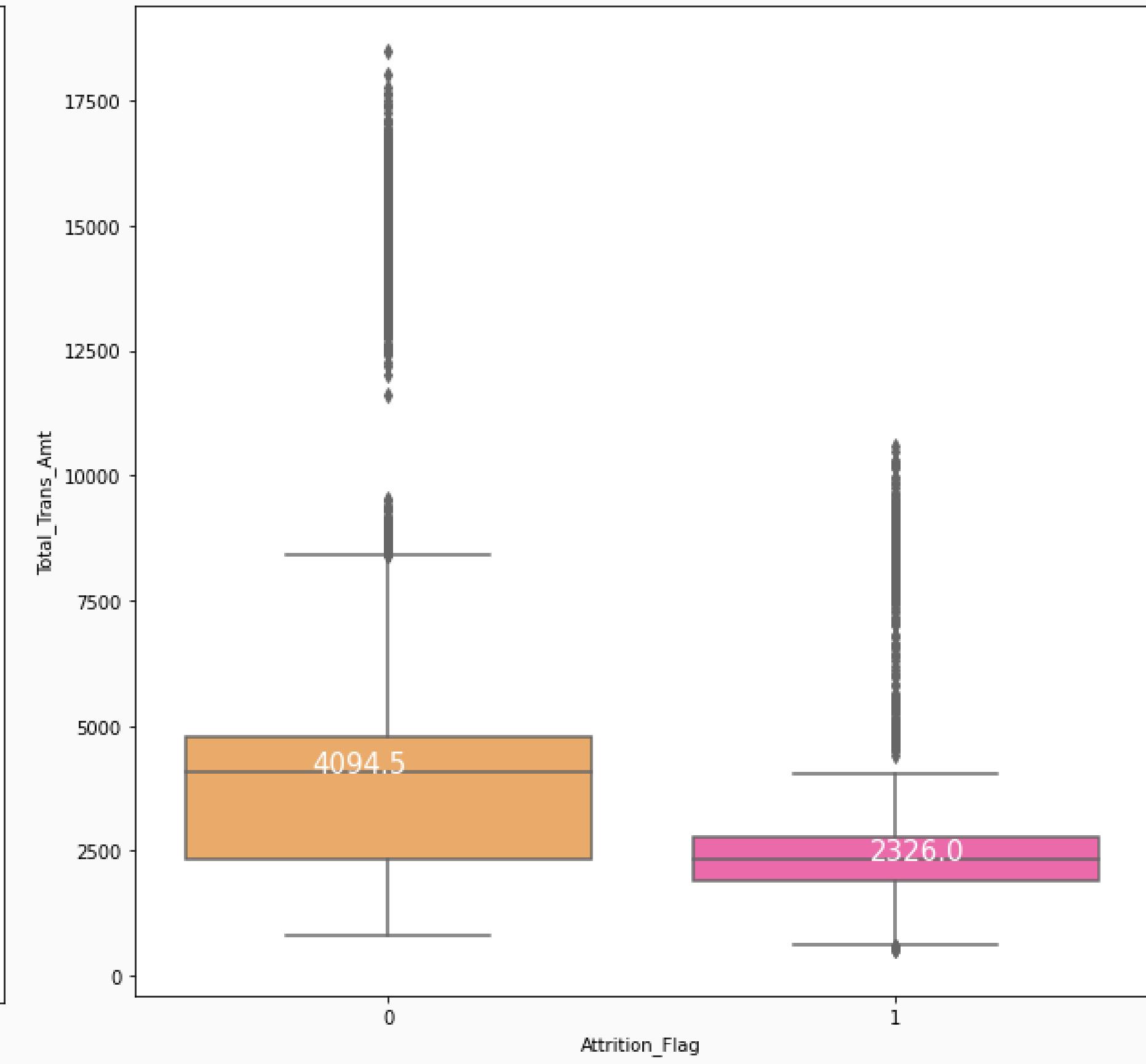
Total Trans Amt
& Total Trans Ct



Total Trans Ct

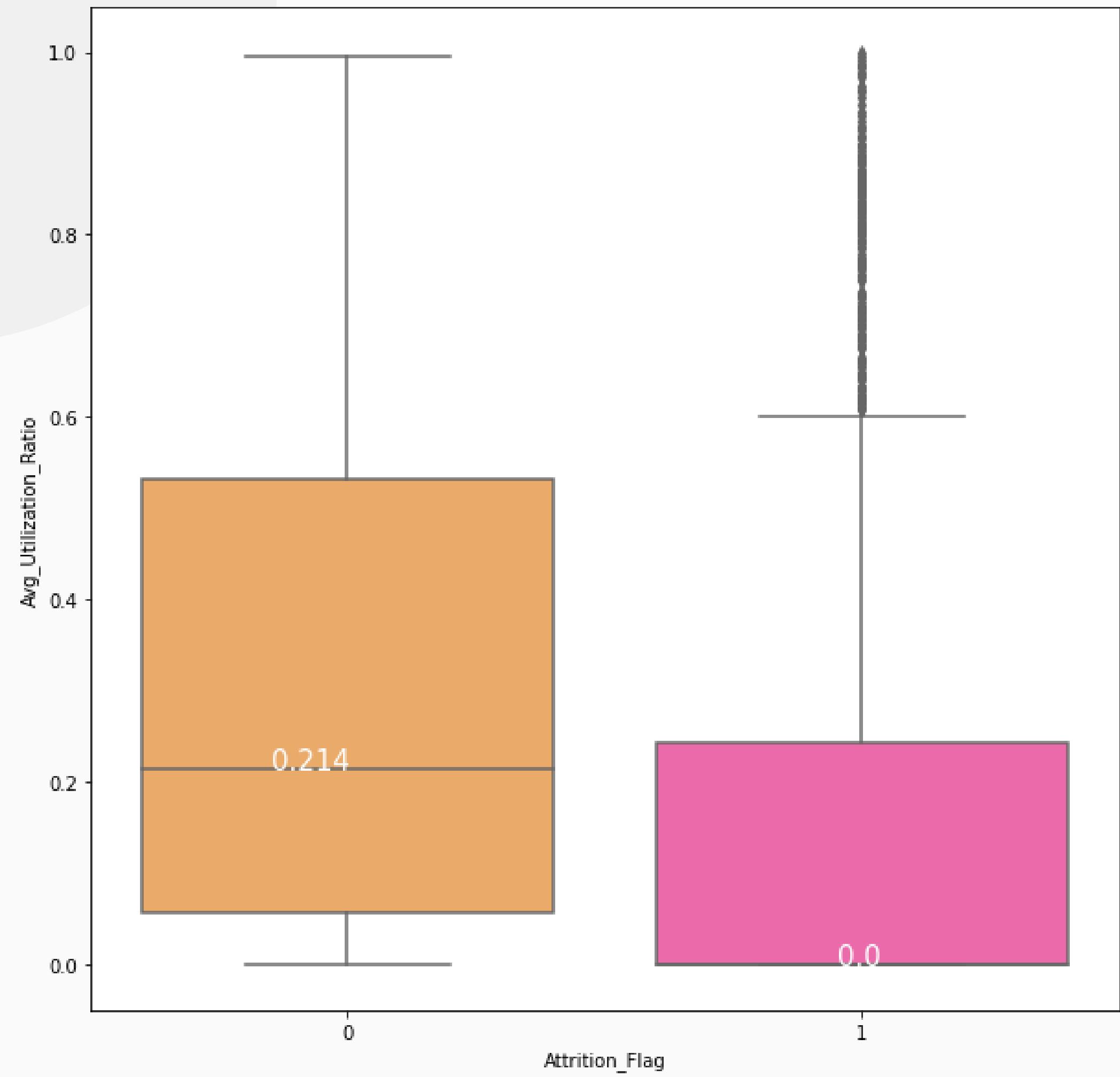
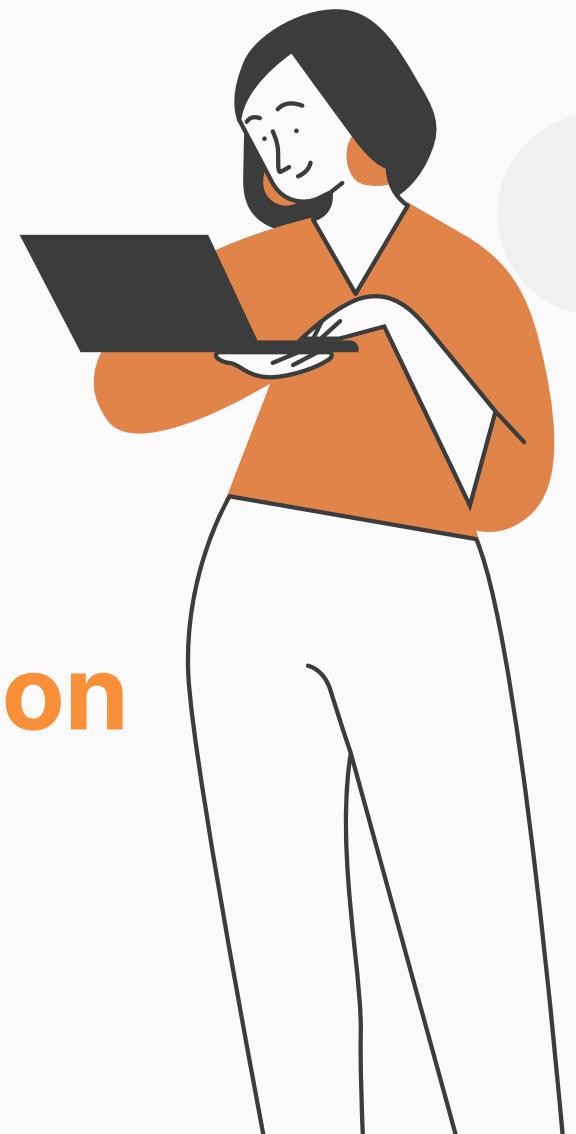


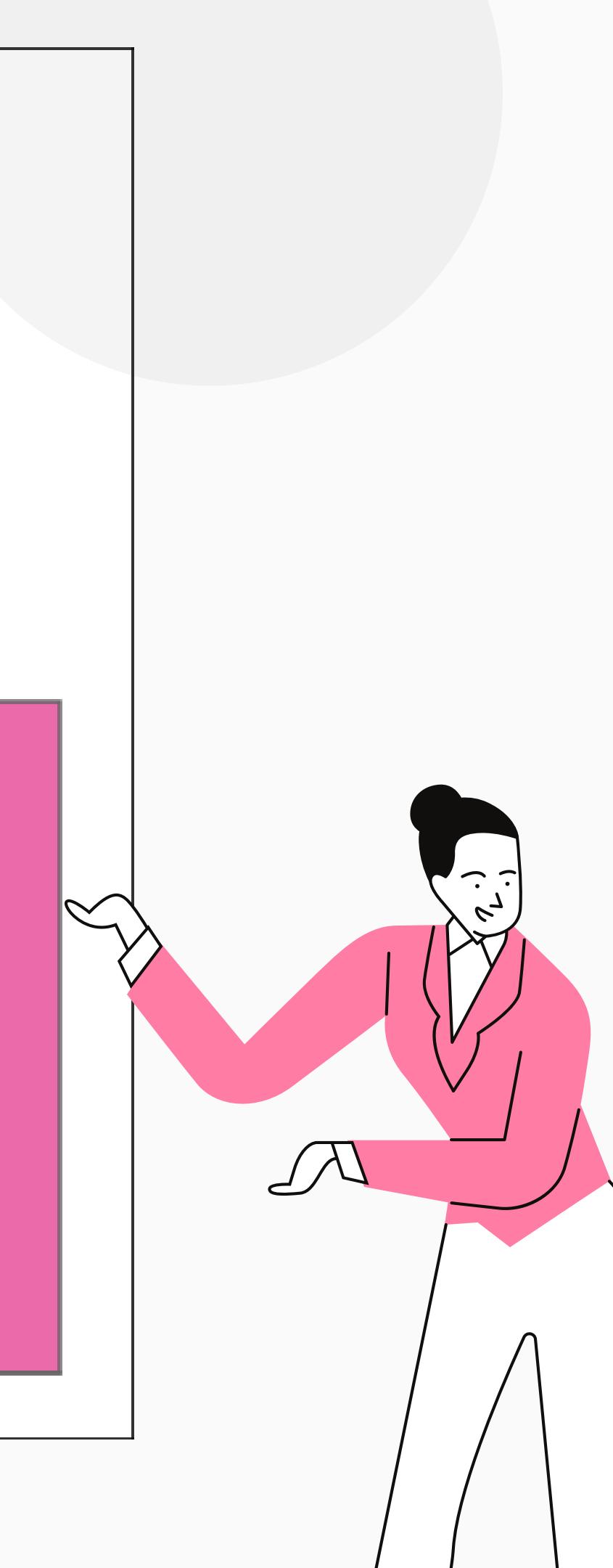
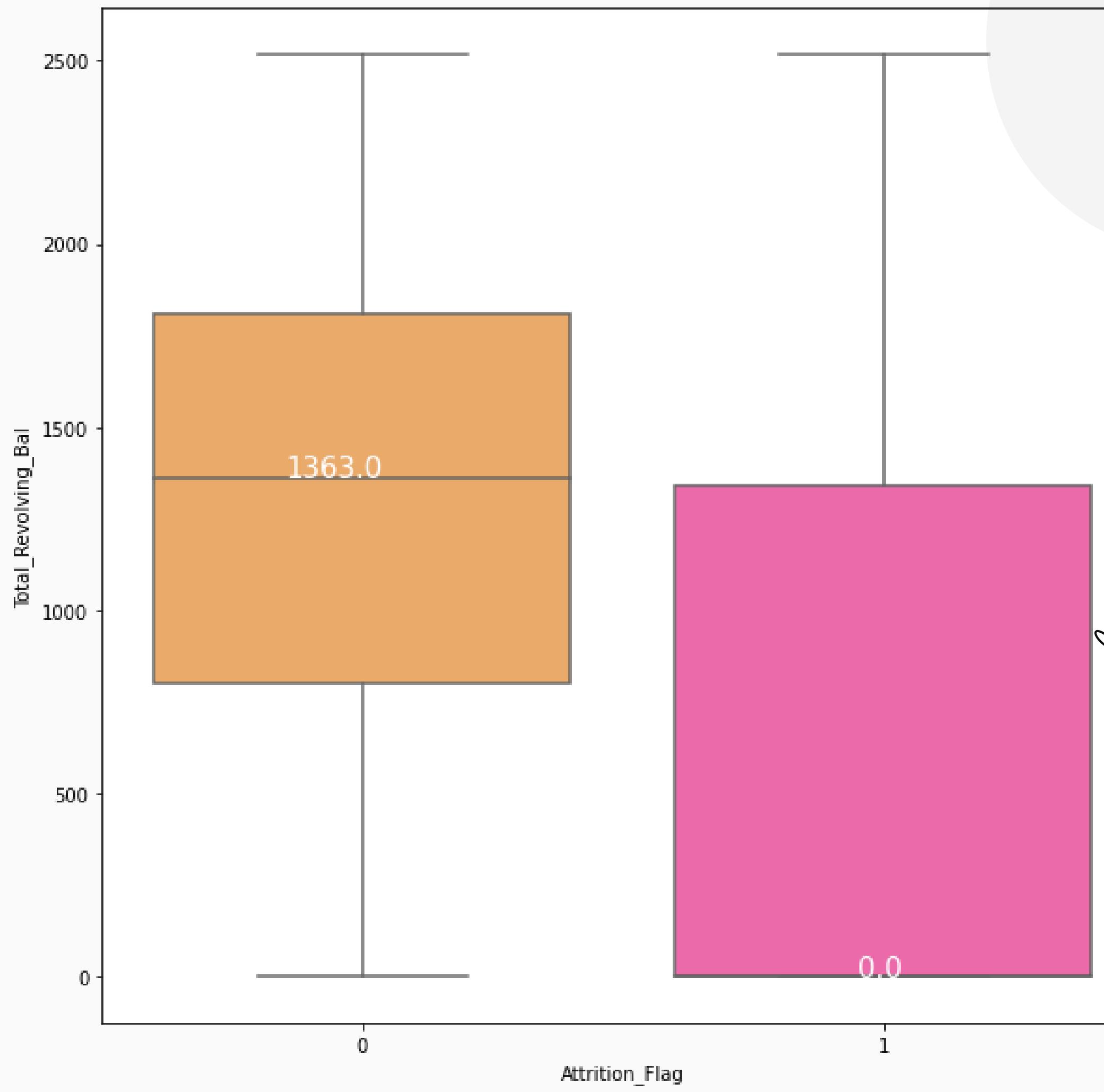
Total Trans Amt



03

Avg Utilization
Ratio





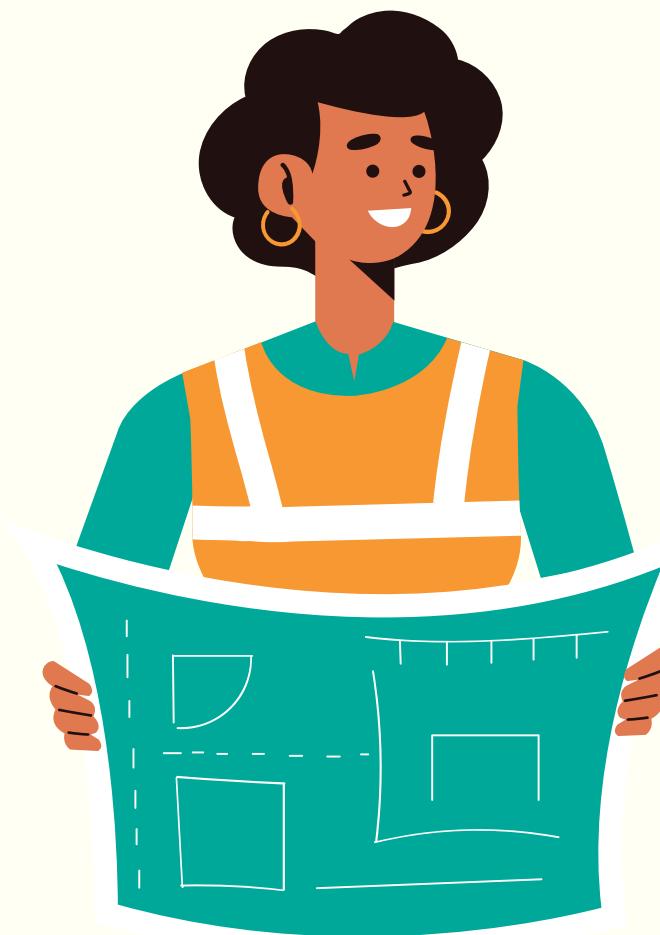
04
**Total Revolving
Balance**

Hypothesis

- Transaction Count & Amount
- Avg Utilisation Ratio
- Revolving Balance



BUILDING the PREDICTION MODEL



random
forest

logistic
regression

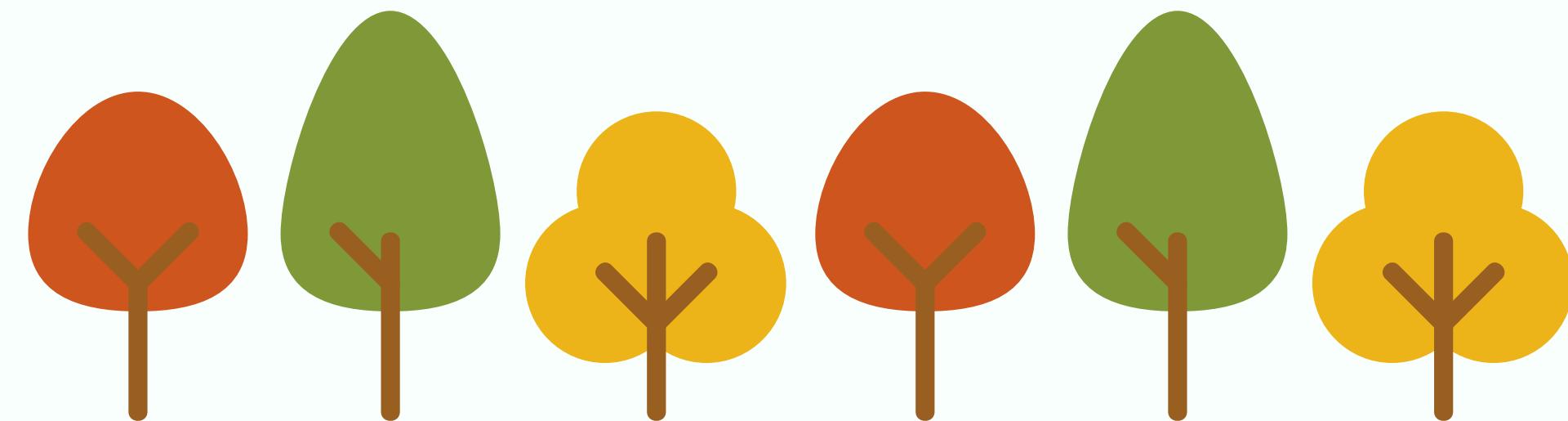
with F1 weighted scoring



Random Forest 1.0

O1

Passed in all
variables except
Avg Open To Buy



O2

Perform
RandomSearchCV

O3

Perform
GridSearchCV

O4

Test model &
extract important
features

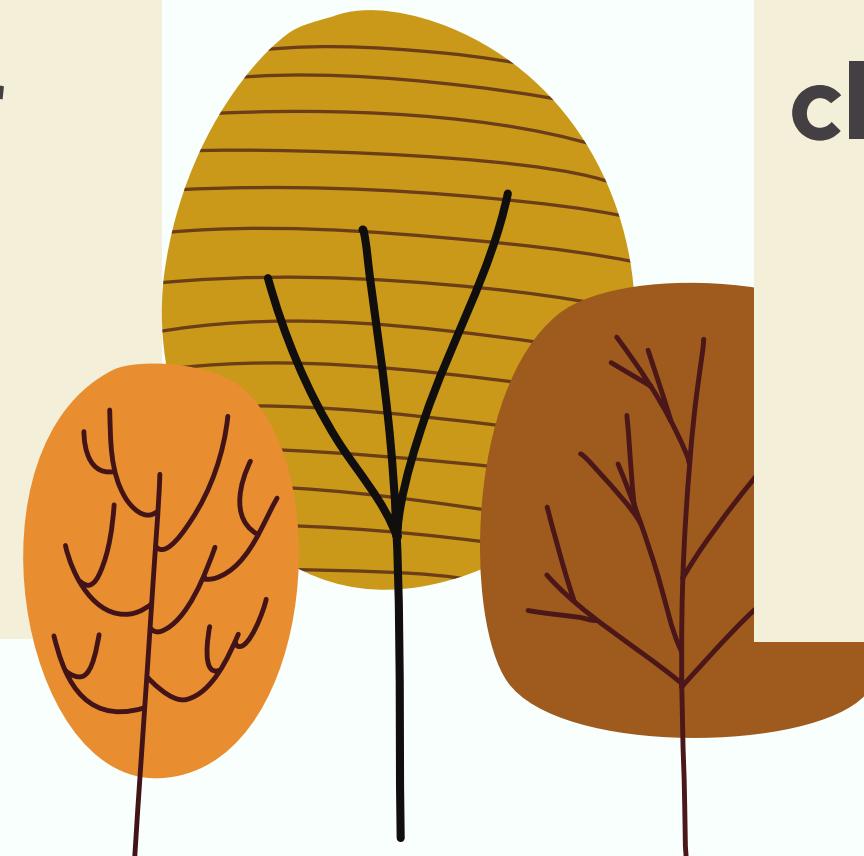
Dealing with Imbalanced Data

UPSAMPLING

`RandomOverSampler`

F1 Weighted

0.96052



CLASS WEIGHTS

`class_weights = 'balanced'`

F1 Weighted

0.95623

FEATURE IMPORTANCE

Total_Trans_Amt	0.186
Total_Trans_Ct	0.166
Total_Revolving_Bal	0.112
Total_Ct_Chng_Q4_Q1	0.106
Avg_Utilization_Ratio	0.074
Total_Amt_Chng_Q4_Q1	0.069
Total_Relationship_Count	0.061
Credit_Limit	0.039

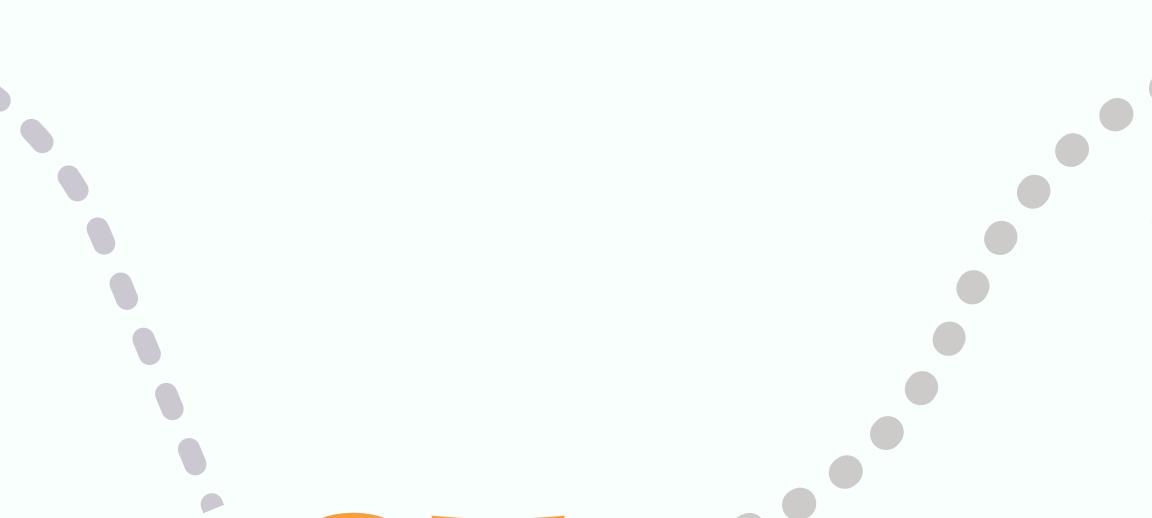
F1 WEIGHTED

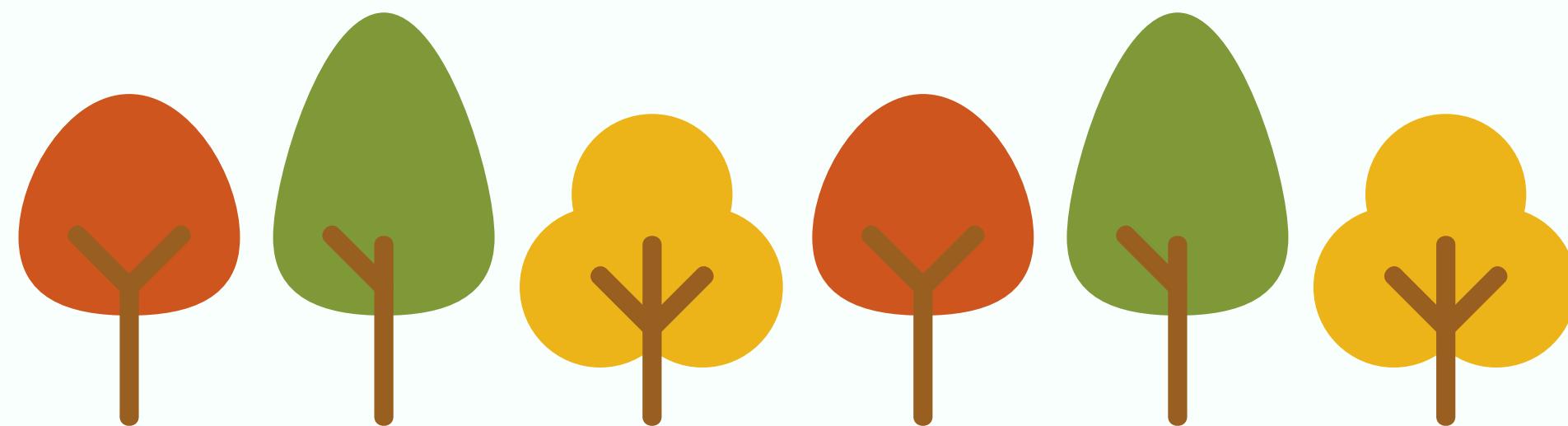
0.9605

ACCURACY

0.9570

Random Forest 2.0

- 
- 05 → Passed in top 7 features
- 06 Perform RandomSearchCV
- 
- 07 Perform GridSearchCV
- 08 Test model



FEATURE IMPORTANCES

	eee
Total_Trans_Amt	0.281
Total_Trans_Ct	0.233
Total_Revolving_Bal	0.136
Total_Ct_Chng_Q4_Q1	0.108
Total_Amt_Chng_Q4_Q1	0.105
Avg_Utilization_Ratio	0.079
Total_Relationship_Count	0.058

Accuracy

eee

0.95707

F1 weighted

eee

0.95709

ROC AUC

eee

0.91994





**Pass in all
variables**

**GridSearch
CV**

Test Model

**LOGISTIC
REGRESSION
1.0**

Accuracy : 0.85659

Weighted F1 : 0.84808

ROC AUC : 0.84137

Perform Feature Selection

Recursive Feature Elimination (RFE)

Accuracy : 0.78536

Weighted F1 : 0.81057

ROC AUC : 0.81061

GridSearchCV

Compare Scores

Select From Model (SFM)

Accuracy : 0.77782

Weighted F1 : 0.74735

ROC AUC : 0.74775



LOGISTIC REGRESSION 2.0

Random Forest

Accuracy :
0.95707

Weighted F1 :
0.95709

ROC AUC :
0.91994



Accuracy :
0.78536

Weighted F1 :
0.81057

ROC AUC :
0.81061

Logistic Regression

Ways to Decrease the % of Customers Attriting

Total Relationship Count
for those with Attrition probability >0.5

+ 1

7%



+ 2

12%



+ 3

13%



Total Transaction Count

+ 10%

11%



+ 20%

26%



+ 30%

44%



Our General Recommendations



Increase Total Relationship Count

by promoting more
products



Increase Total Trans Count

by offering more
spending incentives



Balancing Budget & Ability to Prevent Attrition

Budget required/Associated costs

Low costs vs High costs involved in customer retention expenses

'Ability to Prevent Attrition'

- Change in % of likely-to-attrite customers
- Change in attrition probability for each customer



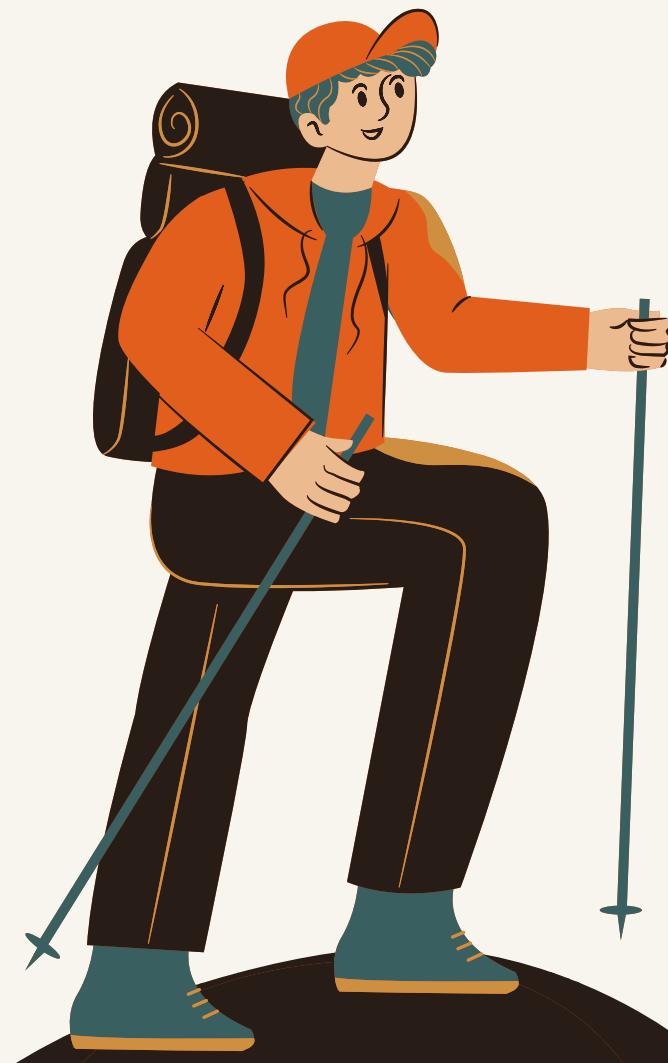
- Assign Attrition Risk Category
Low / Medium / High

- Work on Retaining Which Customers?
Customers with 'reversible' attrition



Further Recommendations for Targeted Attrition Prevention

Further Exploration



1. Review Common Financial Products

Investigate influence on customer attrition

2. Investigate Attrition 'Realness'

Investigate whether customers attrited sign up for new card soon

