BankChurners Report

Wuji Shan

12/01/2021

Abstract

Credit Card has taken up a significant part of people's lives in the current society. The "Credit Card Customers" dataset is data of the consumer credit card portfolio of a bank, including 10,000 customers with their age, gender, income, marital status, education level, and etc. The manager wants to know the reason behind customer attrition.

To deal with this problem, I made exploratory data analysis and built a multilevel model. This report are consisted of 5 main parts: Introduction, Method, Result and Discussion. Other explorations besides what in Method part are all put in Appendix.

Introduction

Method

Data Cleaning and Processing

```
##   Attrition_Flag       Customer_Age       Gender           Education_Level
##   Length:10127      Min.   :26.00     Length:10127      Length:10127
##   Class :character  1st Qu.:41.00     Class :character  Class :character
##   Mode  :character  Median :46.00     Mode  :character  Mode  :character
##                     Mean   :46.33
##                     3rd Qu.:52.00
##                     Max.   :73.00
##   Marital_Status    Income_Category   Card_Category
##   Length:10127      Length:10127      Length:10127
##   Class :character  Class :character  Class :character
##   Mode  :character  Mode  :character  Mode  :character
##
##
##
##   Total_Relationship_Count Months_Inactive_12_mon Total_Trans_Amt
##   Min.   :1.000            Min.   :0.000          Min.   :  510
##   1st Qu.:3.000            1st Qu.:2.000          1st Qu.: 2156
##   Median :4.000            Median :2.000          Median : 3899
##   Mean   :3.813            Mean   :2.341          Mean   : 4404
##   3rd Qu.:5.000            3rd Qu.:3.000          3rd Qu.: 4741
##   Max.   :6.000            Max.   :6.000          Max.   :18484
##   Total_Trans_Ct
```

```
##  Min.   : 10.00
##  1st Qu.: 45.00
##  Median : 67.00
##  Mean   : 64.86
##  3rd Qu.: 81.00
##  Max.   :139.00
```

Exploratory Data Analysis

```r
# Create theme parameters
theme <- theme_bw() +
      theme(plot.title = element_text(face = "bold", color = "black", size=14),
        plot.subtitle = element_text(face = "italic", color = "black", size=12),
        axis.text = element_text(color = "black"), legend.text = element_text(size=10),
        legend.title = element_text(size = 12), legend.position = "none",
        strip.background =element_rect(fill="#666666"), strip.text = element_text(color="white", face="
        plot.caption = element_text(face = "italic"))
```

Distributions of all numeric variables
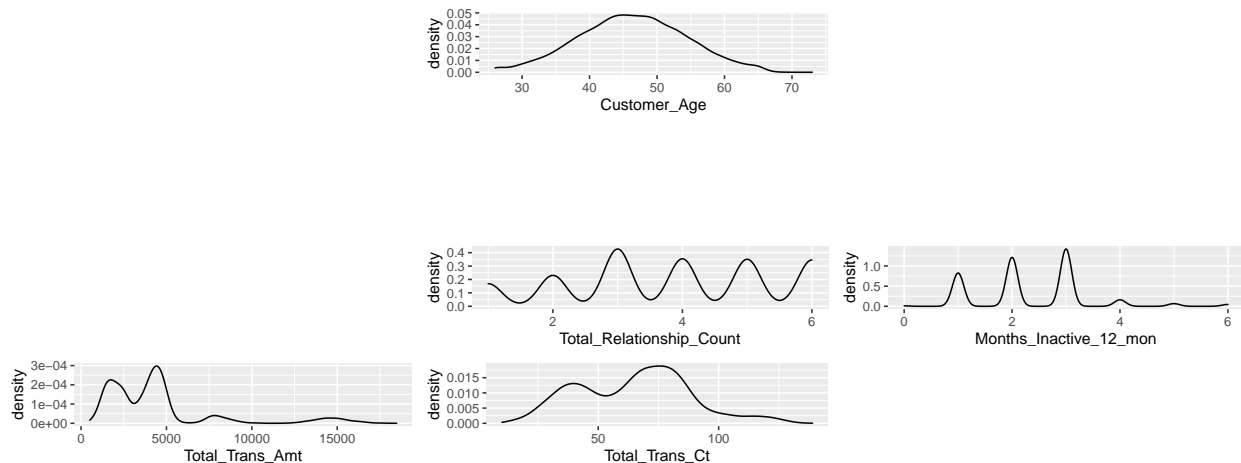
Distributions of all categorical variables
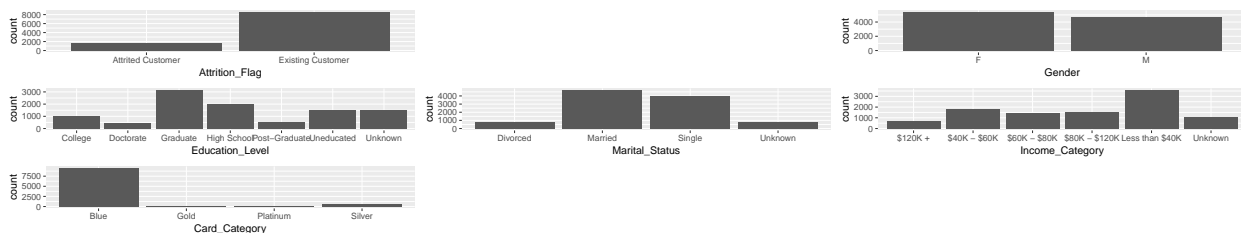
Pie Charts of Attrition_Flag

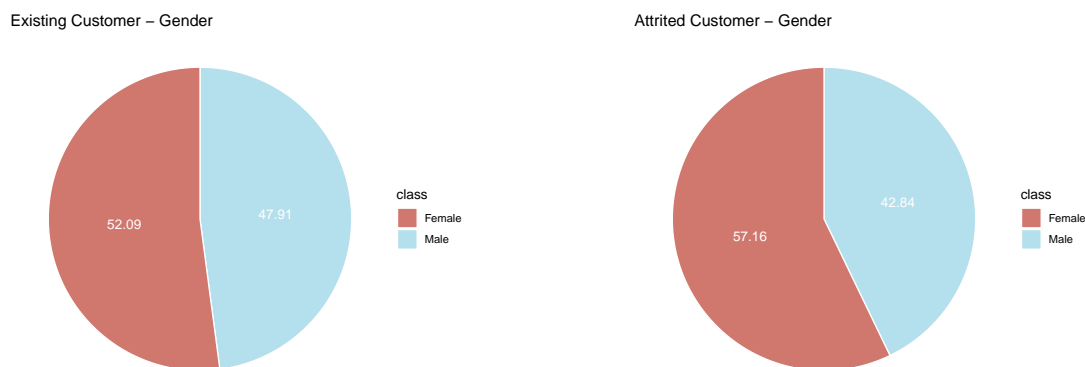a. Pie charts of Gender Proportion comparison for Existing and Attrited customers

Existing Customer – Gender

Attrited Customer – Gender

Figure 1: Gender Proportion Comparison

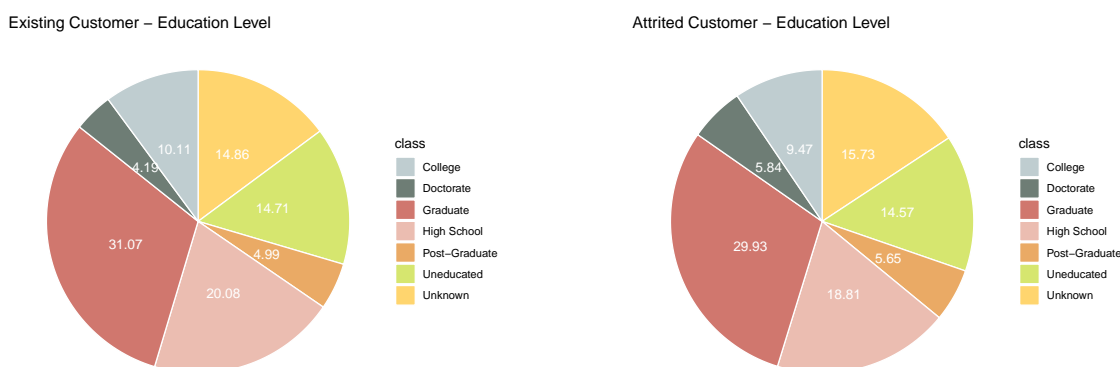b. Pie charts of Education Level Proportion comparison for Existing and Attrited customers

Existing Customer – Education Level

Attrited Customer – Education Level

Figure 2: Education Level Proportion Comparison

Figure 2:

Assuming that customers with "unknown" education level did not receive any education, we can observe that more than 70% of the customers have a formal education level for both existing and attrited customers. Moreover, about 40% have a higher level of education for two groups.
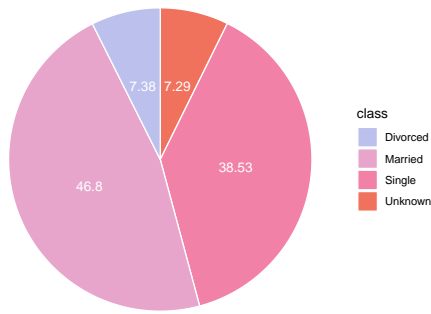
c. Pie charts of Marital Status Proportion comparison for Existing and Attrited customers

Figure 3:

"Almost half of the bank customers are married, and interestingly enough, almost the entire other half are single customers. only about 7% of the customers are divorced, which is surprising considering the worldwide divorce rate statistics!"

The proportion of married status in attrited customers is slightly smaller than that in existing customers; correspondingly, the proportion of single status in attrited customers is slightly larger than that in existing customers.
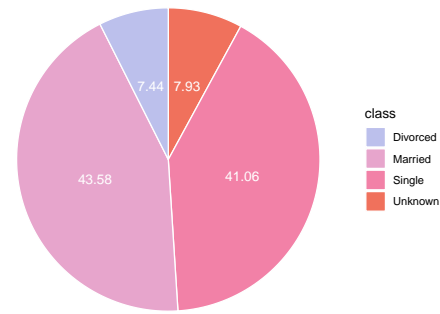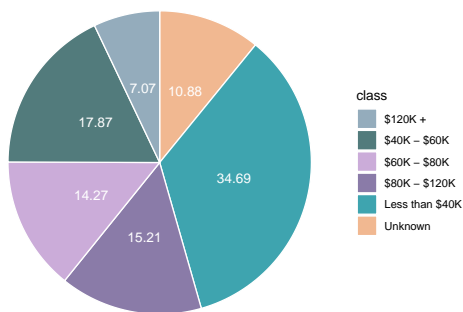
Figure 3: Marital Status Proportion Comparison

d. Pie charts of Income Category Proportion comparison for Existing and Attrited customers
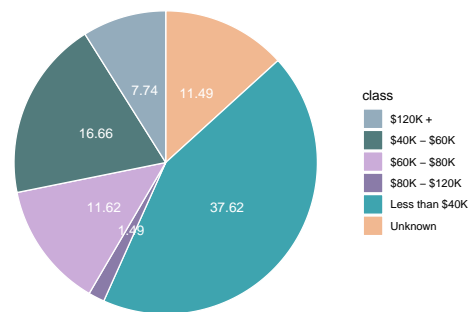

Figure 4: Income Category Proportion Comparison

Income Category

Card Category

"Platinum cards have the highest attrition rates."

Level of Inactivity

Picking joint bandwidth of 0.151

"Attrition Customers have higher levels of Inactivity (3 months vs 2 months Median)."

Total Customer Transactions

Correlation Analysis

Warning: ggrepel: 13 unlabeled data points (too many overlaps). Consider
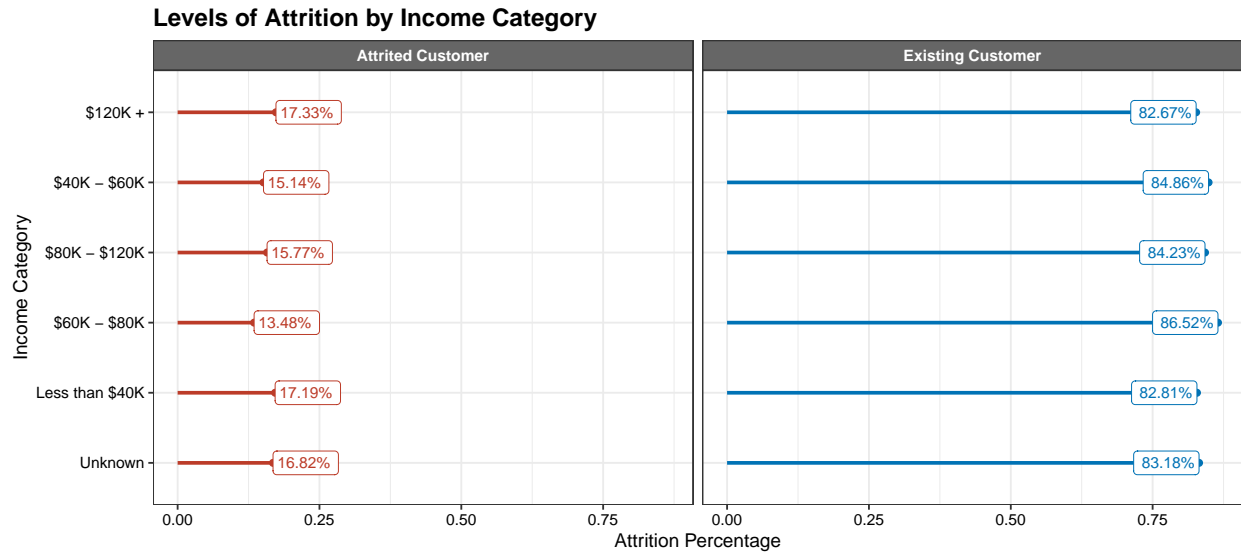## increasing max.overlaps
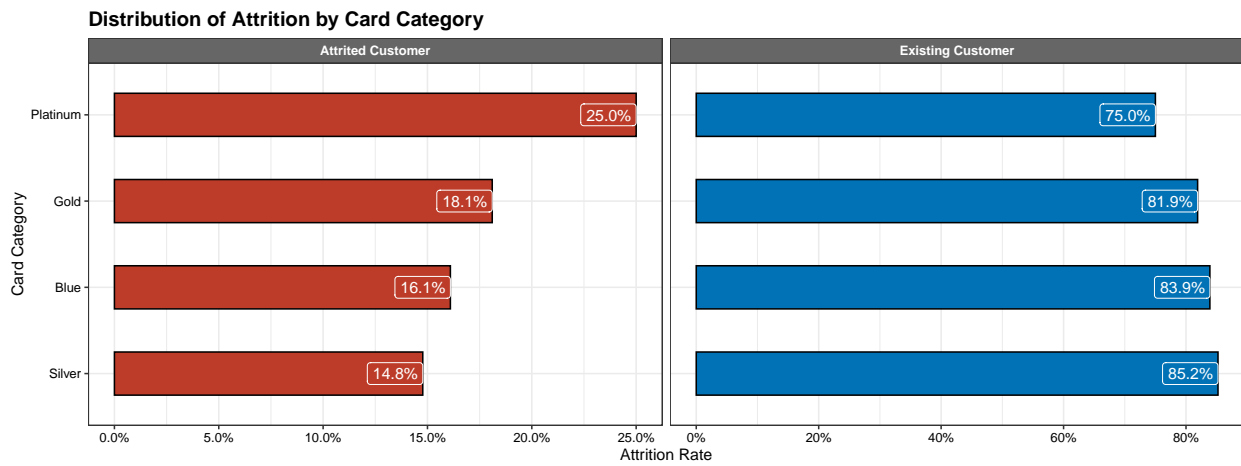
Figure 5: Income Category Proportion Analysis

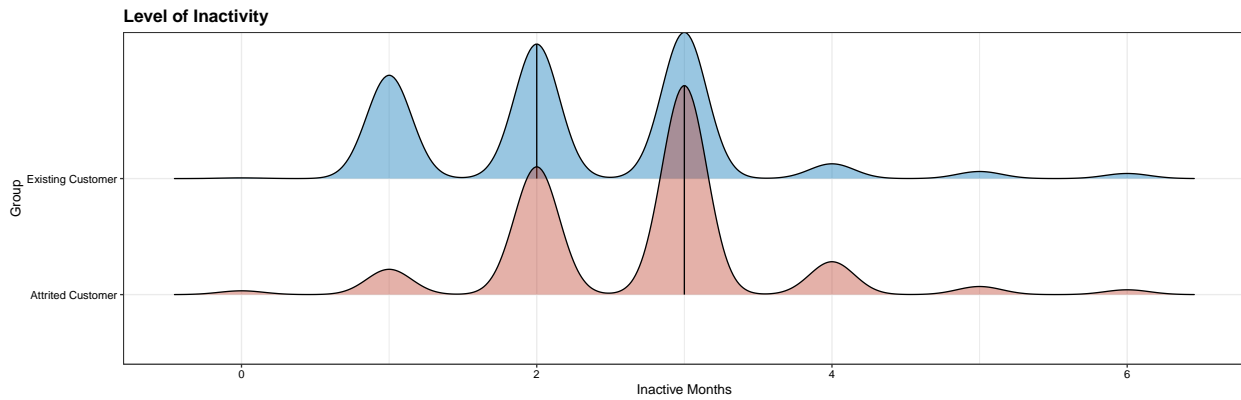Figure 6: Card Category Proportion Analysis

Figure 7: Level of Inactivity

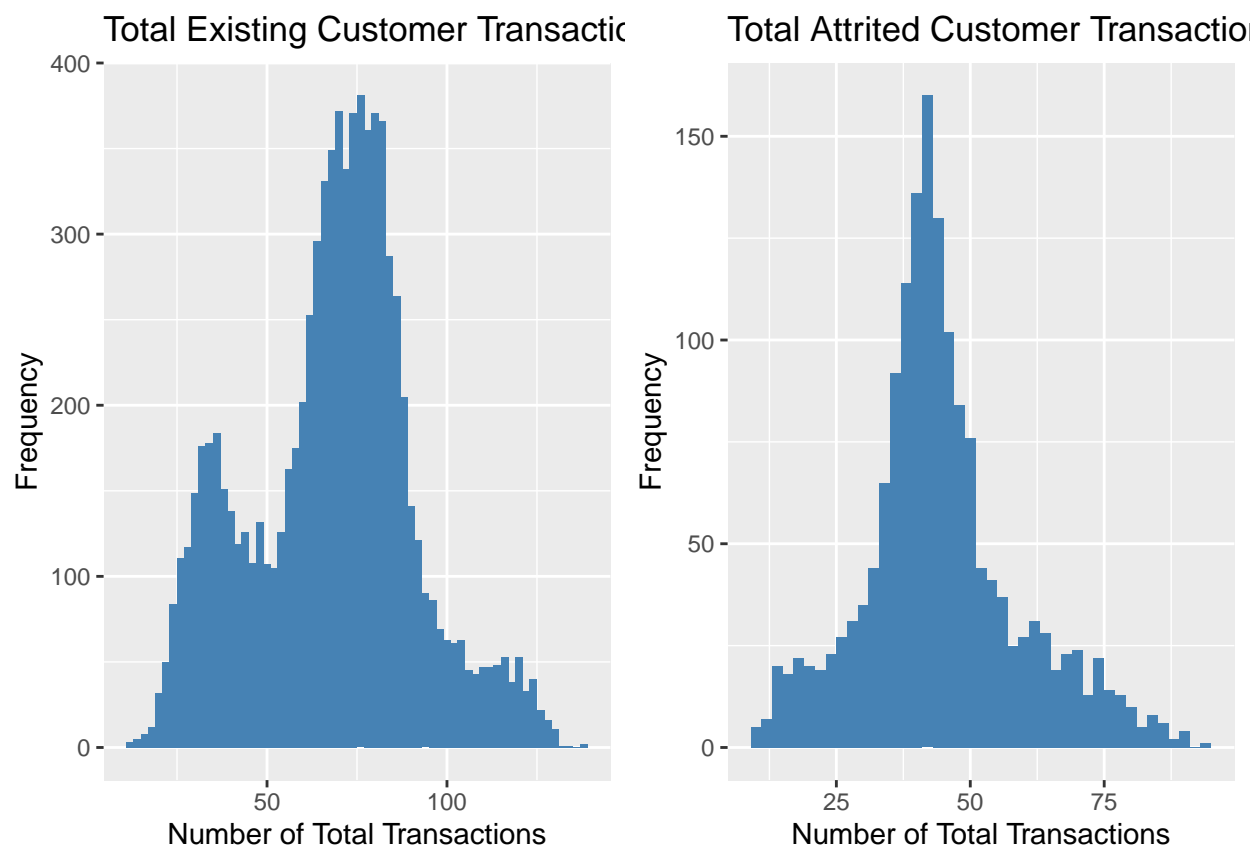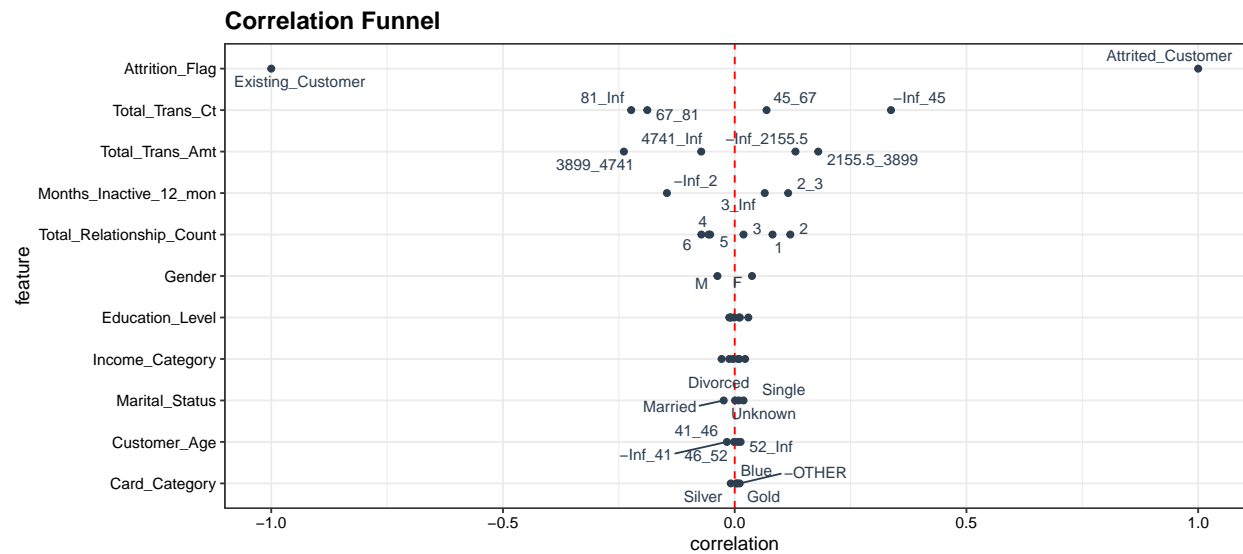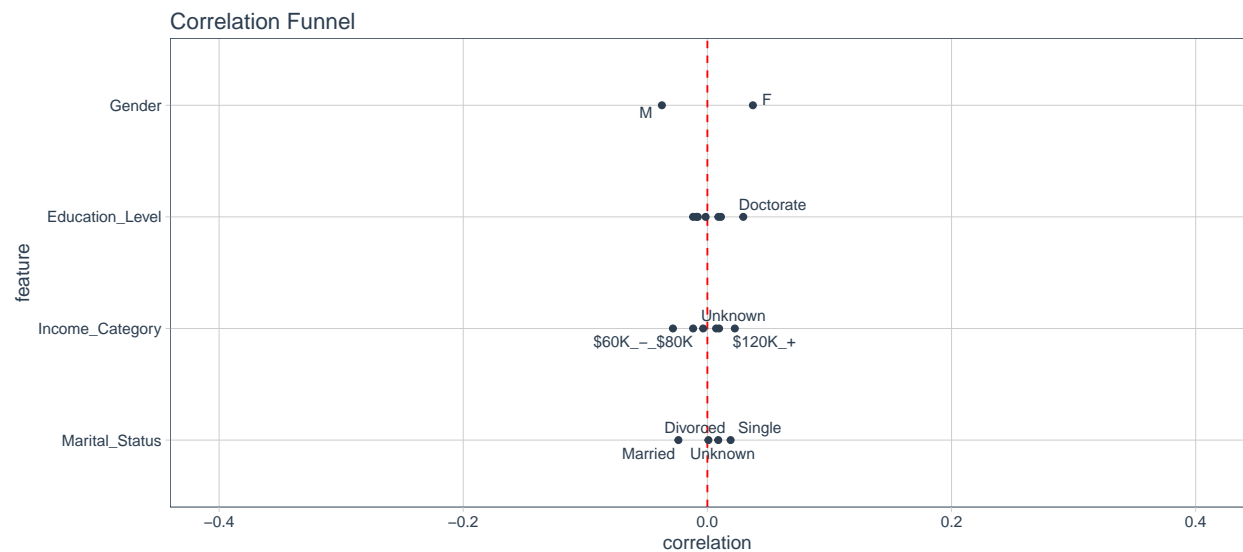Figure 8: Total Customer Transactions Comparison

Correlation Funnel

The most important features are towards the top. We can investigate these.

```
## Warning: ggrepel: 9 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

Model Fitting

```
## Warning in glmer(Attrition_Flag ~ Customer_Age + Education_Level +
## Income_Category + : calling glmer() with family=gaussian (identity link) as a
## shortcut to lmer() is deprecated; please call lmer() directly


## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00431519 (tol = 0.002, component 1)


## Linear mixed model fit by REML ['lmerMod']
## Formula: Attrition_Flag ~ Customer_Age + Education_Level + Income_Category +
##     Card_Category + Total_Relationship_Count + Months_Inactive_12_mon +
##     (1 + Gender | Marital_Status)
##    Data: BankChurners
##
## REML criterion at convergence: 8077.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9205  0.1496  0.3653  0.5372  1.4459
##
## Random effects:
##  Groups         Name        Variance  Std.Dev. Corr
##  Marital_Status (Intercept) 1.277e-06 0.00113
##                 GenderM     7.128e-04 0.02670  -1.00
##  Residual                   1.284e-01 0.35838
## Number of obs: 10127, groups:  Marital_Status, 4
##
## Fixed effects:
##                                Estimate Std. Error t value
## (Intercept)                   0.8193322  0.0310515  26.386
## Customer_Age                 -0.0002888  0.0004463  -0.647
## Education_LevelDoctorate     -0.0563985  0.0203144  -2.776
## Education_LevelGraduate      -0.0066179  0.0129630  -0.511
## Education_LevelHigh School   -0.0037815  0.0138122  -0.274
## Education_LevelPost-Graduate -0.0346631  0.0193995  -1.787
## Education_LevelUneducated    -0.0093186  0.0146109  -0.638
## Education_LevelUnknown       -0.0199620  0.0145466  -1.372
## Income_Category$40K - $60K    0.0348897  0.0167417   2.084
## Income_Category$60K - $80K    0.0362753  0.0164055   2.211
## Income_Category$80K - $120K   0.0132171  0.0161642   0.818
## Income_CategoryLess than $40K 0.0266788  0.0171538   1.555
## Income_CategoryUnknown        0.0323617  0.0194649   1.663
## Card_CategoryGold             0.0045756  0.0336096   0.136
## Card_CategoryPlatinum        -0.0332984  0.0804051  -0.414
## Card_CategorySilver           0.0228184  0.0157514   1.449
## Total_Relationship_Count      0.0353707  0.0023027  15.361
## Months_Inactive_12_mon       -0.0549151  0.0035317 -15.549


##
## Correlation matrix not shown by default, as p = 18 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it
```
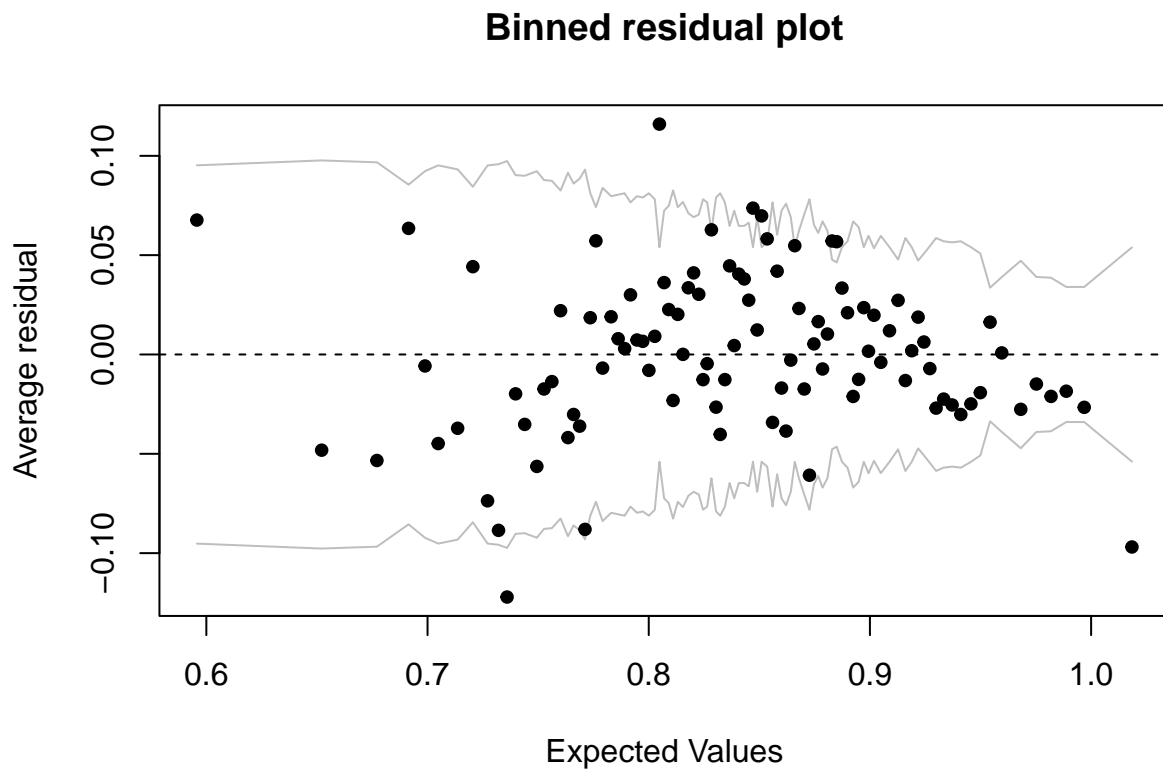
```
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.00431519 (tol = 0.002, component 1)


## (Intercept)
##        0.819


##           (Intercept) GenderM
## Divorced        0.000   0.010
## Married        -0.002   0.039
## Single         -0.001   0.016
## Unknown         0.000   0.011
```
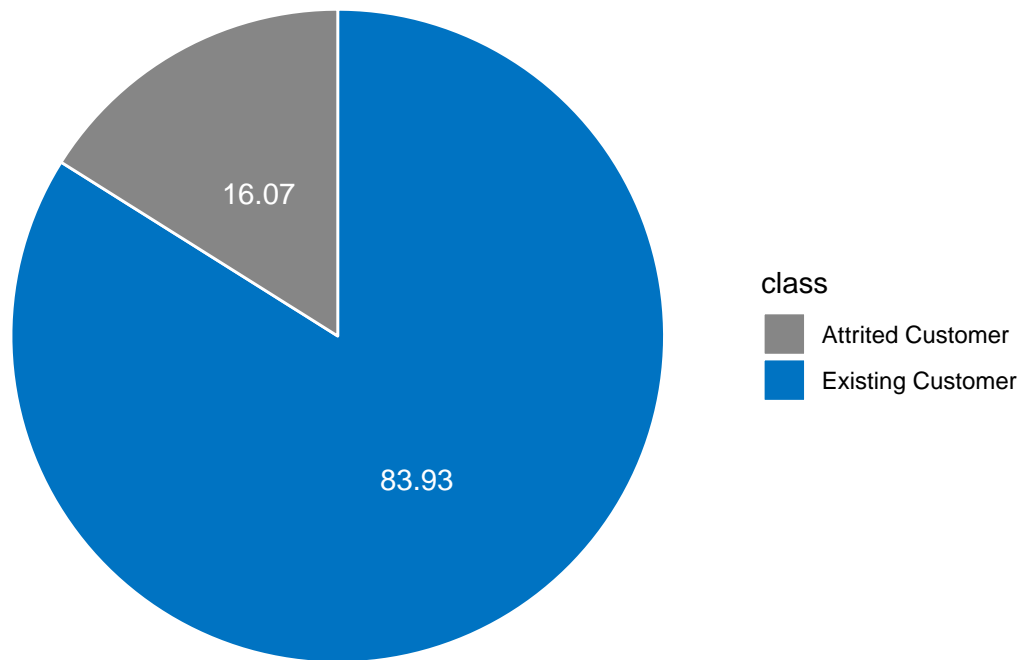
Binned residual plot

Reference

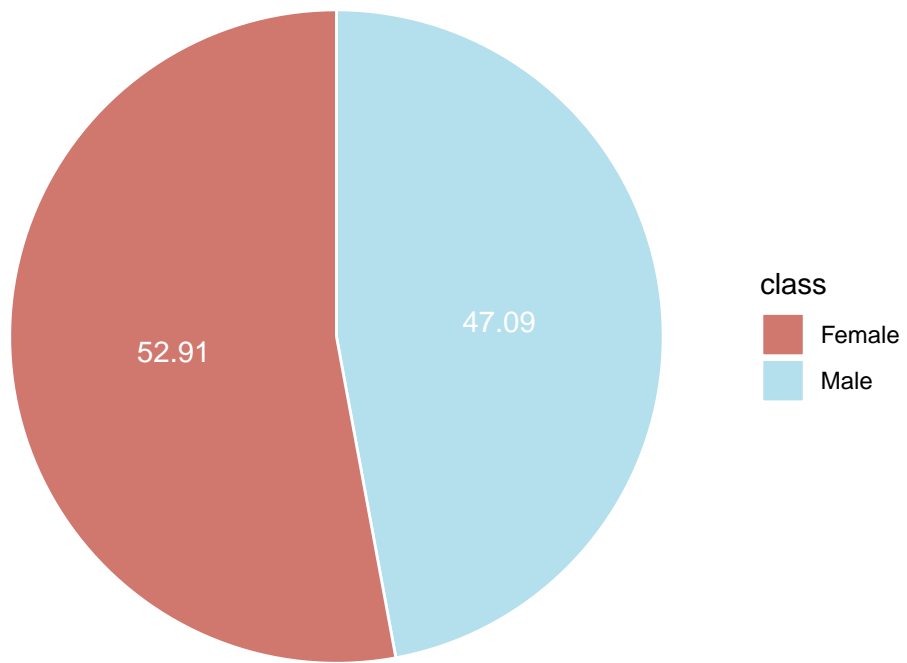1. https://cran.r-project.org/web/packages/correlationfunnel/vignettes/introducing_correlation_ funnel.html

Appendix
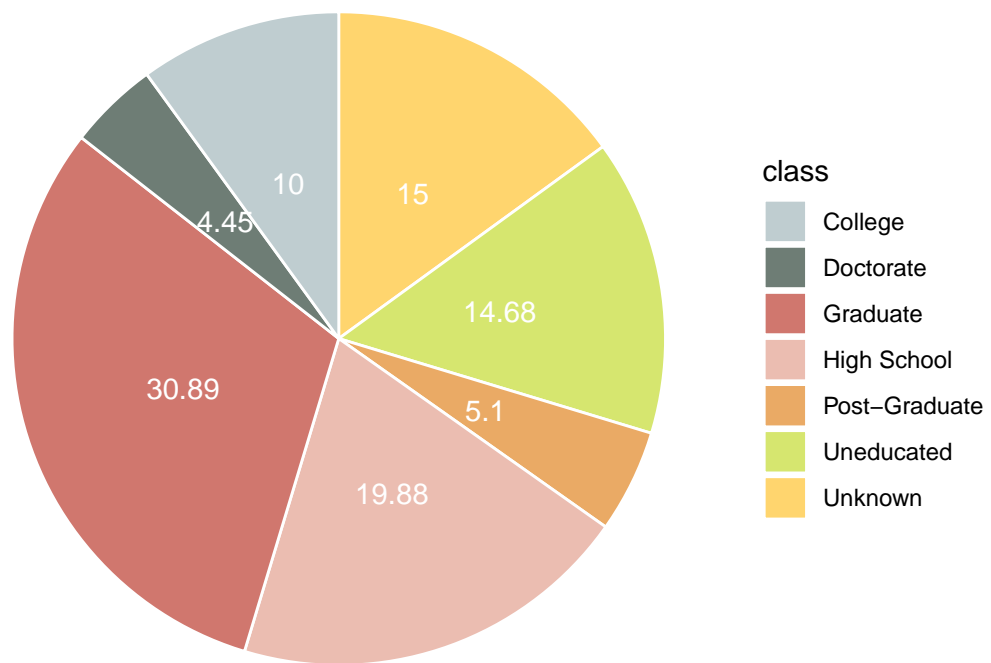
Showing pie charts of overall analysis of the whole data.

Attrition Flag

Gender

class
- Female (52.91)
- Male (47.09)

Education Level

Education Level