THERA BANK PROJECT

2021

OBJECTIVES

- To predict variables responsible for Credit Card churning through the following;
- Explore and visualize the dataset
- Build a classification model to predict if the customer is going to churn or not
- Optimize the model using appropriate techniques
- Generate a set of insights and recommendations that will help the bank
- To determine which segment of customers should be targeted more.

BUSINESS PROBLEM OVERVIEW

BACKGROUND

- Thera Bank's Executive Management seeks to establish factors responsible for its Credit Card Churning as well as facilitating Customer Retention while expanding their customer base and ultimatley maximising revenue and profit through processing fees and loan charges. and equally establish a viable business model using existing data of customers and potential ones as well.
- As a senior data scientist at Thera Bank,, I am tasked with coming up with a predictive classification model using available data and through indepth analysis, profer insightful recommendations to the Policy Makers of Thera Bank to identify strong indices that will influence retention

SOLUTION APPROACH (MACHINE LEARNING)

- Define the problem and perform an Exploratory Data Analysis
- Illustrate the insights based on EDA
- Data pre-processing/Feature Engineering
- Model building – Logistics Regression (OverSampling and Under-Sampling, Bagging and Boosting Models
- HyperParameter Tuning: GridSearchCv and RandomSearchCv and comparing Sytem processing times per tuning method
- Model Performance Metrics Comparison
- Return the Most Import Features/Variables influencing Bank CreditCard Churn
- Actionable Insights & Recommendations

BUSINESS IMPLICATIONS

- Identify major influencers of Attrition among Credit card customers
- Optimize processes for ease of subscription while improving contact and retaltionship with potential churners.
- Maximize Revenue and effectively introduce perks , loyalty points or coupons to endear potential churners through more robust and targeted marketing
- Increased Customer patronage, retention, referral and traffic which ultimately impacts on Thera Bank's bottomline
- Tracking and Trapping potential churners with customised product offerings along with additional alternatives
- Optimize administrative and overhead costs by applying the Predictive model smartly on perculiarity basis
- Optimize Sales and Profit Margin per new card subscriptions
- Effective Allocation and redistribution of resources especially for product re-design, pricing and Ad-campaigns
- Developing potential markets for more revenue

DATA MANIPULATION

- Conversion of Data types
- Identification of Unknown(Missing) Values
- Fixing and Dropping Columns: (CLIENTNUM and Avg_Open_To_Buy) were dropped ultimately prior to Modelling
- Fixing Missingness using KNN_Imputer
- As a result of the perfect correlation with Credit_Limit, Avg_Open_To_Buy was dropped for a cleaner model prediction

DATA INFORMATION

Variable	Description
CLIENTNUM	Client number. Unique identifier for the customer holding the account
Attrition_Flag	Internal event (customer activity) variable - if the account is closed then "Attrited Customer" else "Existing Customer"
Customer_Age	Age in Years
Gender	Gender of the account holder
Dependent_count	Number of dependents
Education_Level	Educational Qualification of the account holder - Graduate, High School, Unknown, Uneducated, College(refers to a college student), Post-Graduate, Doctorate.
Marital_Status:	Marital Status of the account holder
Income_Category	Annual Income Category of the account holder
Card_Category	Type of Card
Months_on_book	Period of relationship with the bank
Total_Relationship_Count	Total no. of products held by the customer
Months_Inactive_12_mon	No. of months inactive in the last 12 months
Contacts_Count_12_mon	No. of Contacts between the customer and bank in the last 12 months

Observations	Variables
10127	21

Rows	Columns
5	21

Float64	Int64
5	9
Credit_Limit Avg_Open_To_Buy Total_Ct_Chng_Q4_Q1 Total_Amt_Chng_Q4_Q1 Avg_Utilization_Ratio	CLIENTNUM Customer_Age Dependent_count Months_on_book Total_Relationship_Count Months_Inactive_12_mon Contacts_Count_12_mon Total_Revolving_Bal Total_Trans_Amt Total_Trans_Ct

Object (6)

Attrition_Flag	Marital_Status
Gender	Card_Category
Educational_Level	Income_Category

DATA INFORMATION

Variable	Description
Credit_Limit	Credit Limit on the Credit Card
Total_Revolving_Bal	The balance that carries over from one month to the next is the revolving balance
Avg_Open_To_Buy	Open to Buy refers to the amount left on the credit card to use (Average of last 12 months)
Total_Trans_Amt	Total Transaction Amount (Last 12 months)
Total_Trans_Ct	Total Transaction Count (Last 12 months)
Total_Ct_Chng_Q4_Q1	Ratio of the total transaction count in 4th quarter and the total transaction count in 1st quarter
Total_Amt_Chng_Q4_Q1	Ratio of the total transaction amount in 4th quarter and the total transaction amount in 1st quarter
Avg_Utilization_Ratio	Represents how much of the available credit the customer spent

'Unknown' Values in Data (Treated)

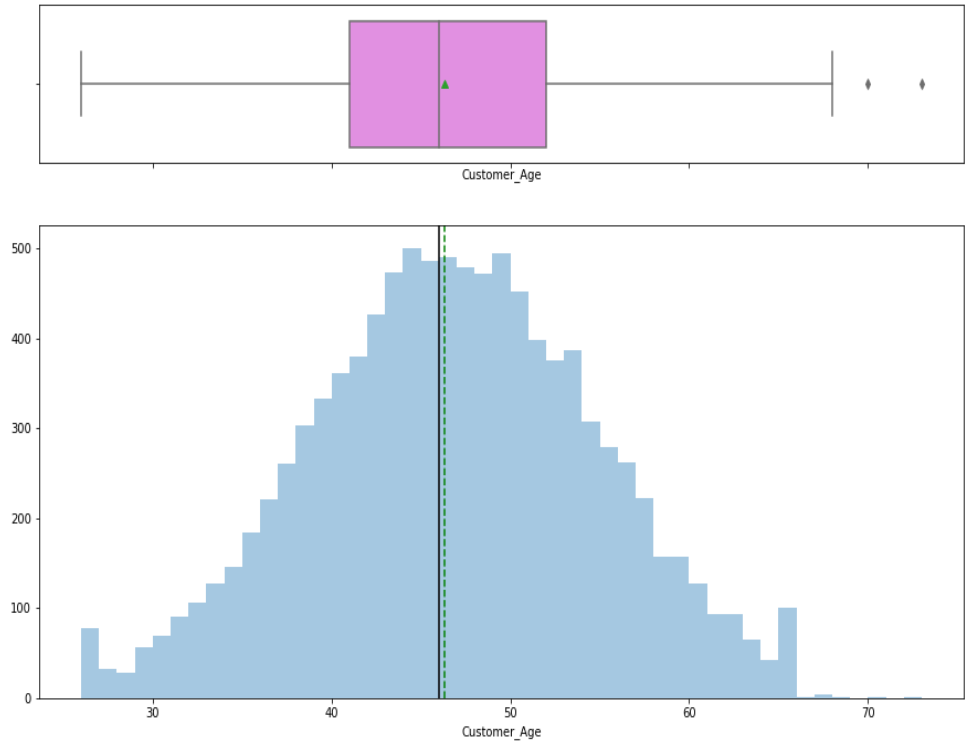
Education_Level

Income_Category

Marital_Status

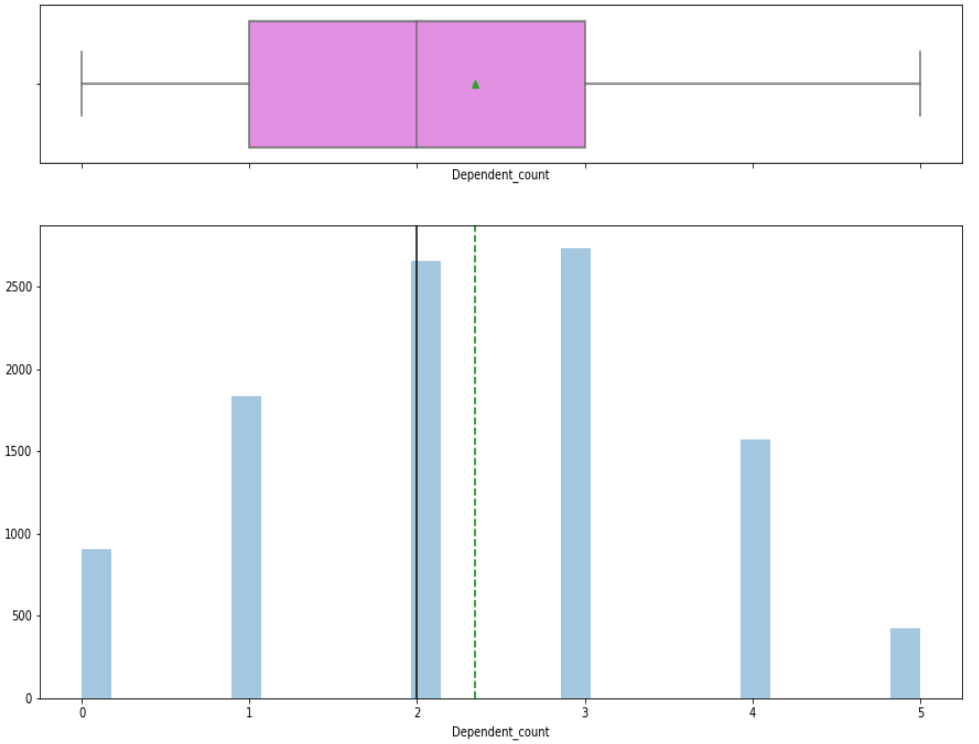
EXPLORATORY DATA ANALYSIS

CUSTOMER_AGE



- Age is looking normally distributed, with a hint of right skew
- The customer distribution has a mean age of 37 yrs and a median of 36 yrs
- 75% of the customers are 43 yrs or less with the oldest at 61 years

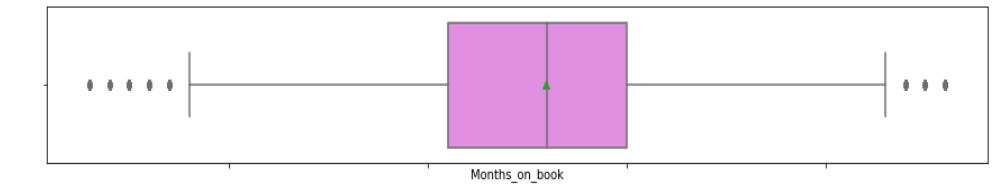
DEPENDENT_COUNT

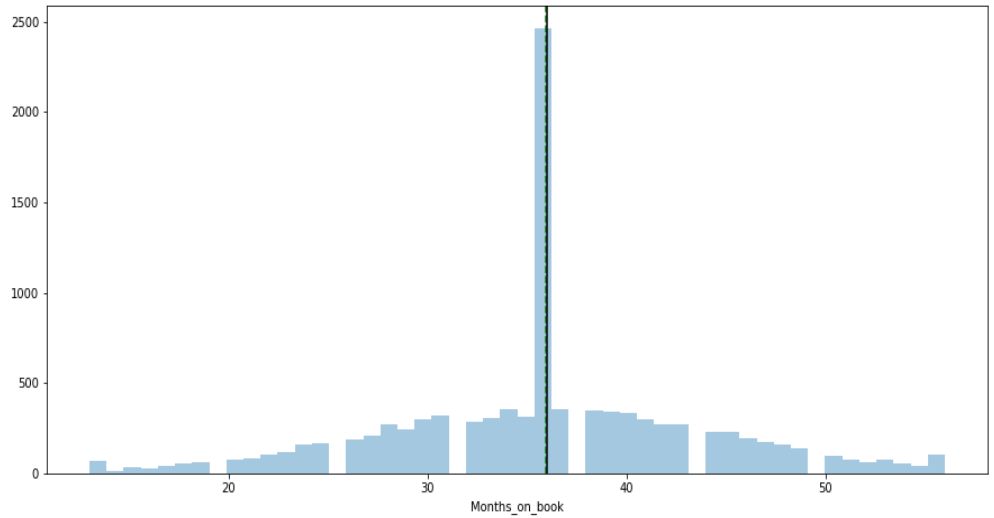


- The Duration of Pitch shows a presence of Outliers
- Some customers were pitched to for far longer periods than others
- The mean duration was 15 while 75% of the customers were accorded 19 or less with maximum at 127

EXPLORATORY DATA ANALYSIS

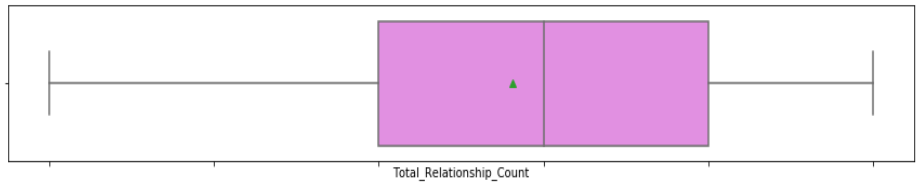
MONTHS_ON_BOOK

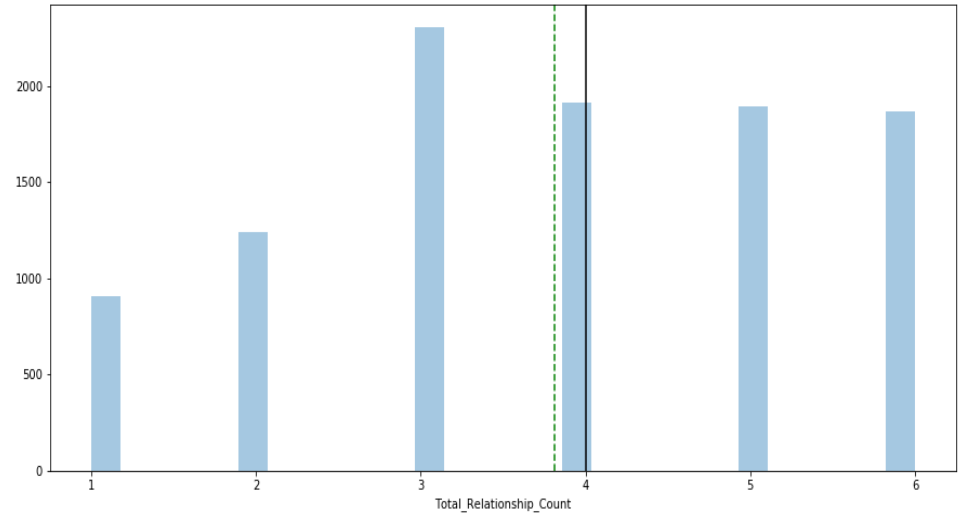




- This distribution has 3 peaks at 2,3 and 4.
- This clearly indicates that majority of the customers have between 2 to 4 people willing to go on a trip with them
- There is a presence of an outlier

TOTAL_RELATIONSHIP_COUNT

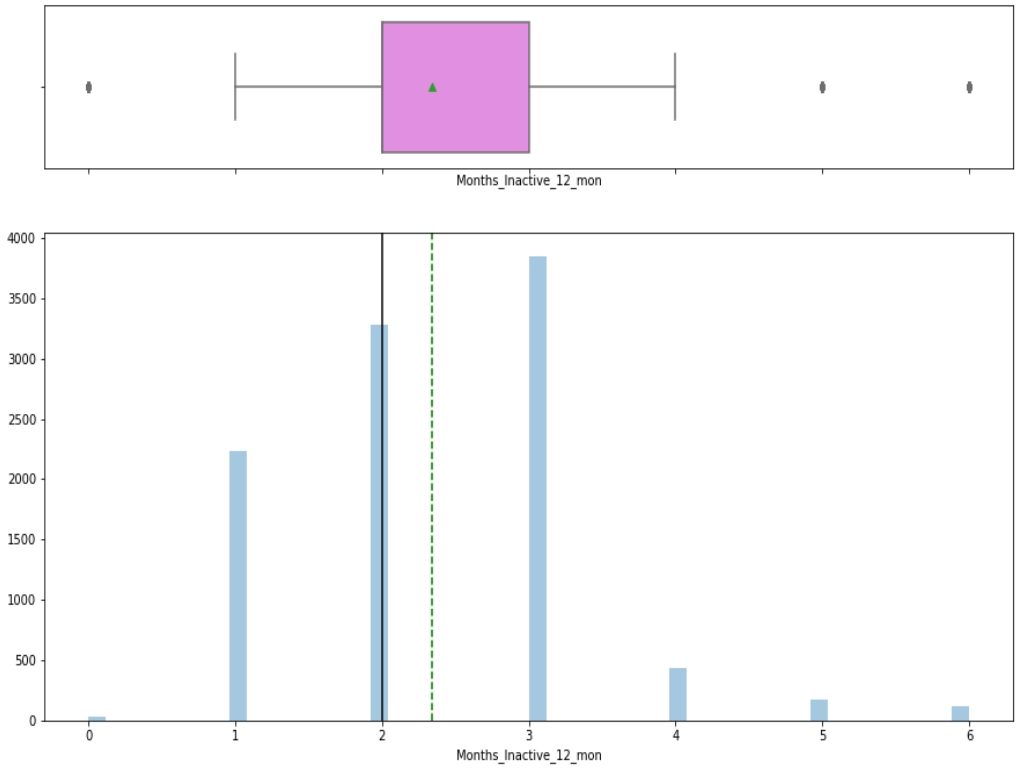




- The distribution for Number of followups has 3 peaks at 3, 4 and 5 with approximately 1400, 2000 and 750 customers respectively
- There is a presence of outliers

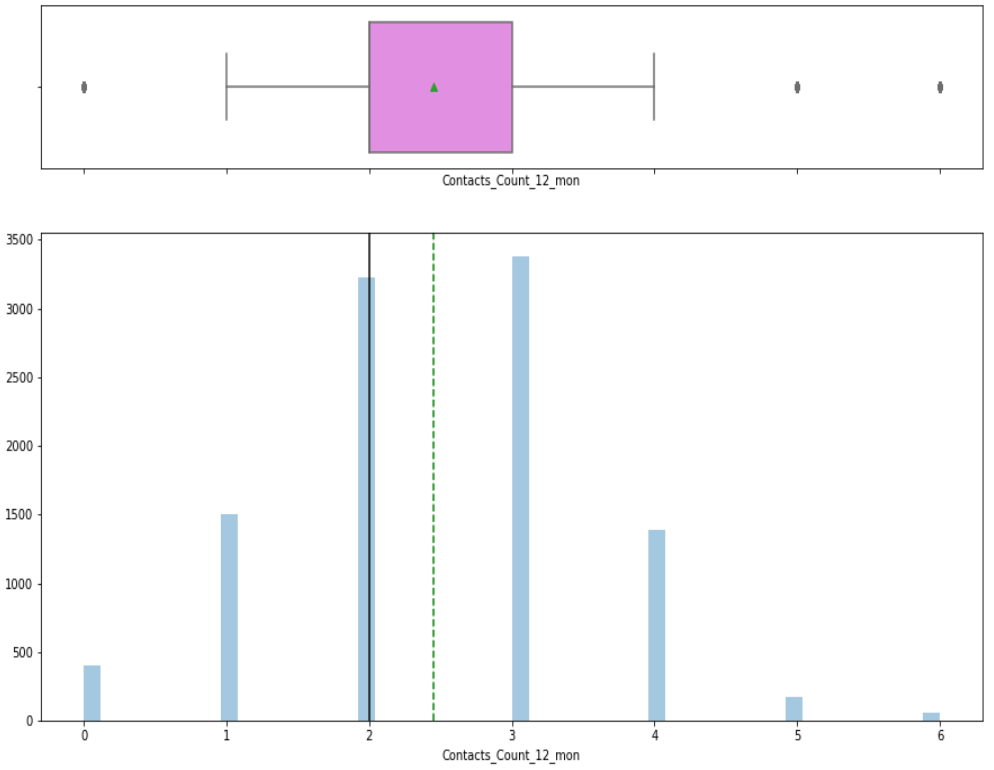
EXPLORATORY DATA ANALYSIS

MONTHS_INACTIVE_12_MON



- This distribution has 3 peaks at 2,3 and 4.
- This clearly indicates that majority of the customers have between 2 to 4 people willing to go on a trip with them
- There is a presence of an outlier

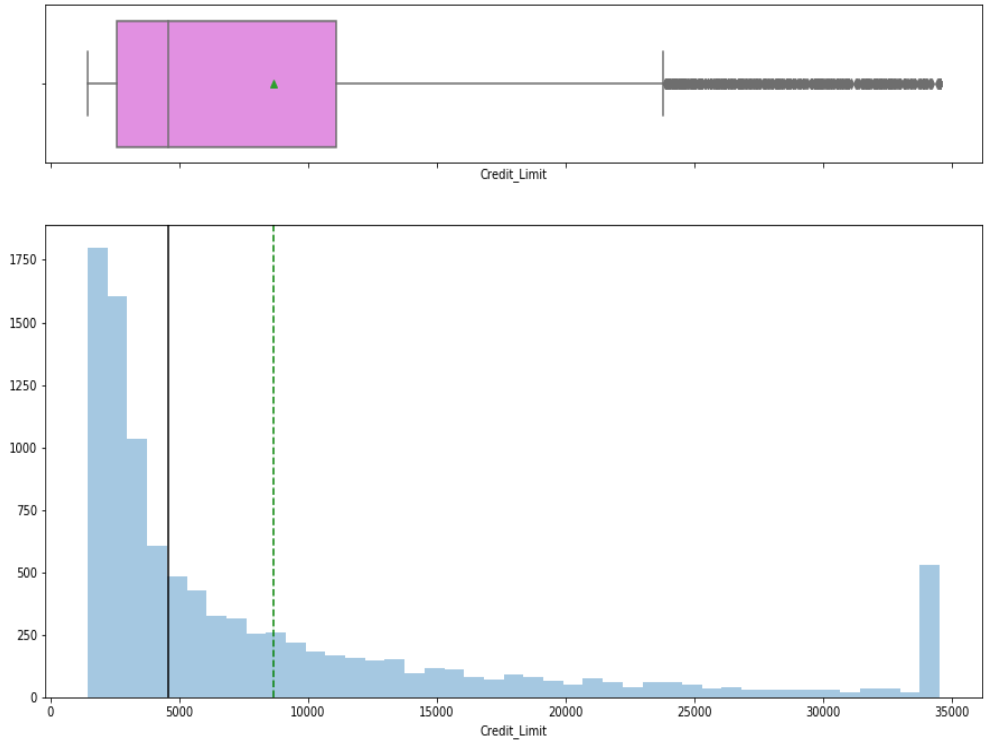
CONTACTS_COUNT_12_MON



- The distribution for Number of followups has 3 peaks at 3, 4 and 5 with approximately 1400, 2000 and 750 customers respectively
- There is a presence of outliers

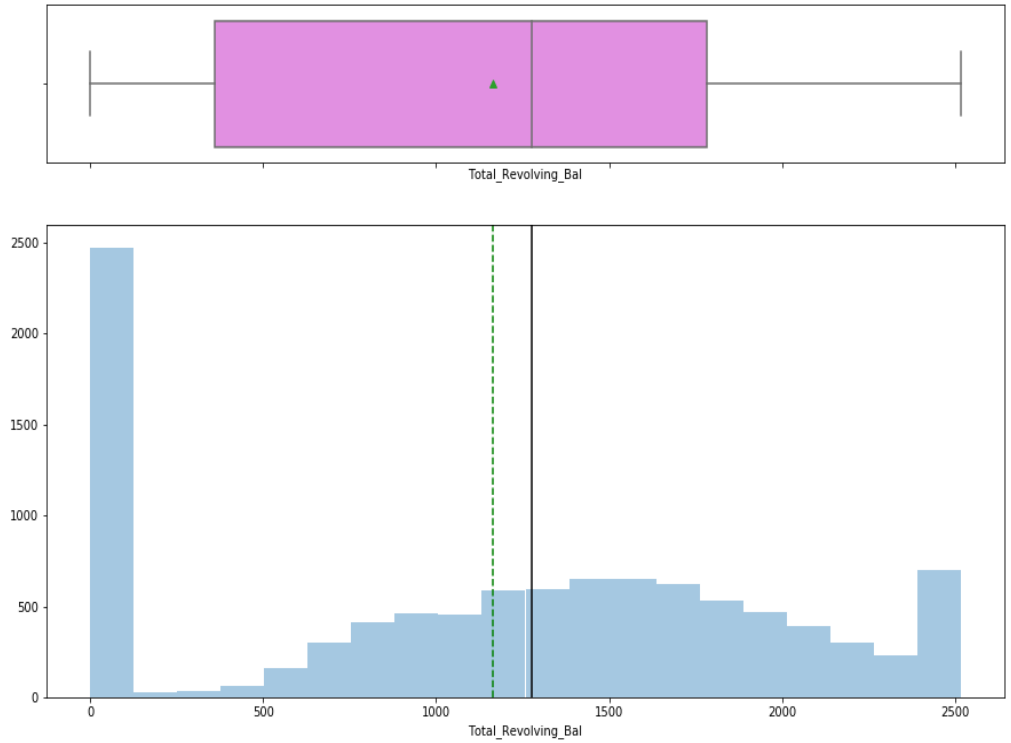
EXPLORATORY DATA ANALYSIS

CREDIT_LIMIT



- Monthly Income is a normal ditribution but is Right skewed.
- Tremendous presence of outliers.This will be treated as we proceed.
- Mean Income is 23473 compared to median at 22347
- 75% of the customers earn a monthly income of 25374 or less

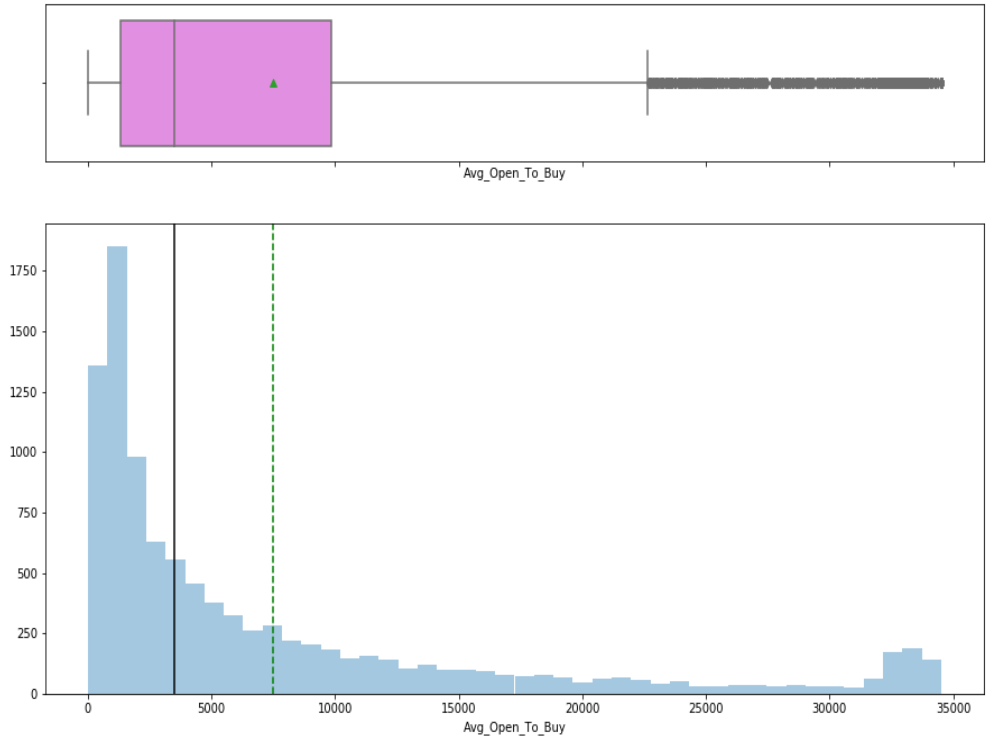
TOTAL_REVOLVING_BAL



- The distribution peaks at 0, 1 and 2 respectively.
- On the average, 1 child is allowed to go on a visit with a customer.
- 75% of the distribution have 2 children or less who plan to go on a trip with customers

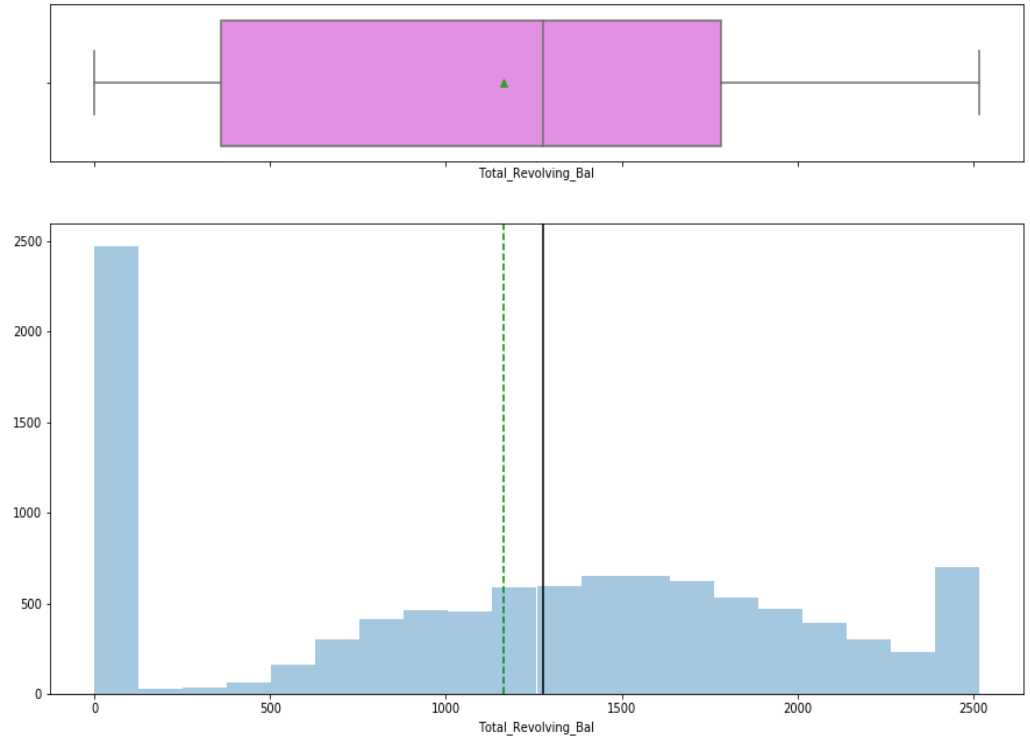
EXPLORATORY DATA ANALYSIS

AVG_OPEN_TO_BUY



- Monthly Income is a normal ditribution but is Right skewed.
- Tremendous presence of outliers.This will be treated as we proceed.
- Mean Income is 23473 compared to median at 22347
- 75% of the customers earn a monthly income of 25374 or less

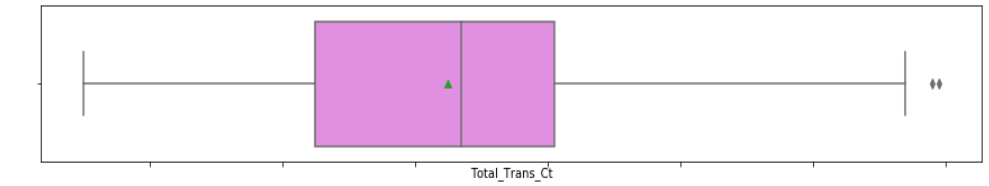
TOTAL_AMT_CHNG_Q4_Q1

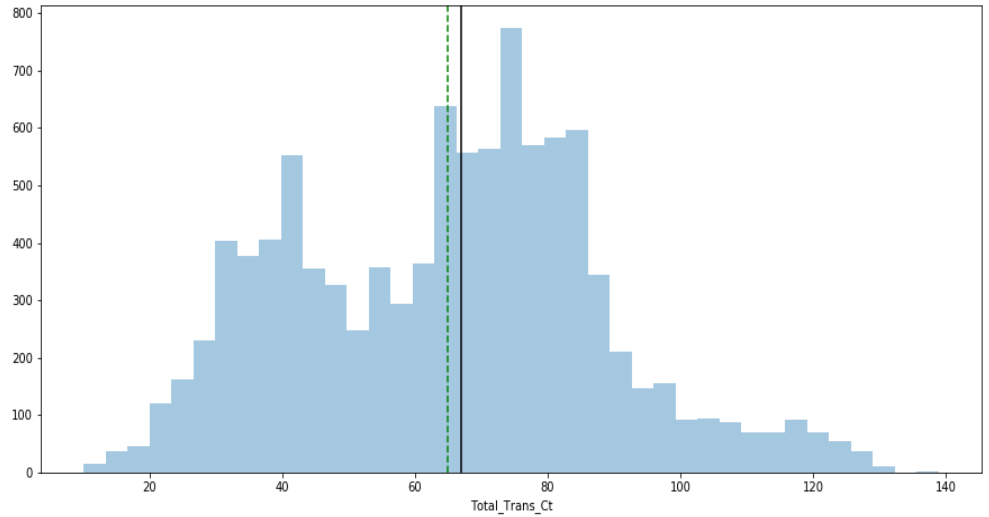


- The distribution peaks at 0, 1 and 2 respectively.
- On the average, 1 child is allowed to go on a visit with a customer.
- 75% of the distribution have 2 children or less who plan to go on a trip with customers

EXPLORATORY DATA ANALYSIS

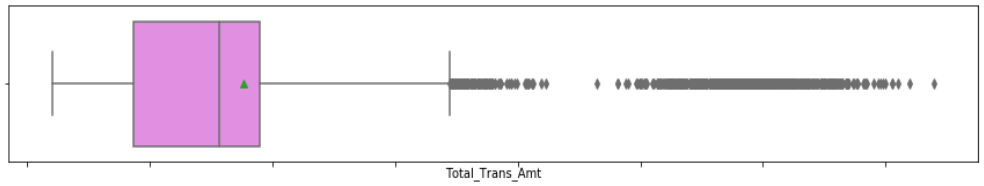
TOTAL_TRANS_CT

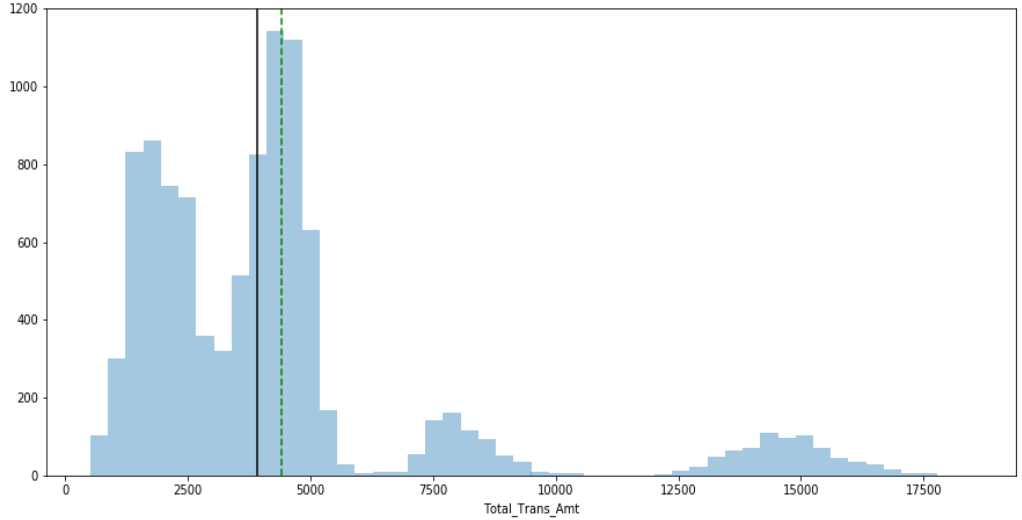




- Monthly Income is a normal ditribution but is Right skewed.
- Tremendous presence of outliers.This will be treated as we proceed.
- Mean Income is 23473 compared to median at 22347
- 75% of the customers earn a monthly income of 25374 or less

TOTAL_TRANS_AMT

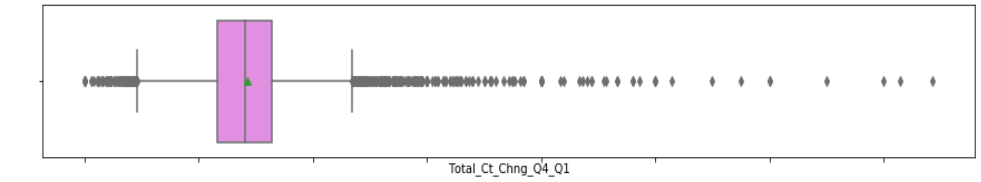


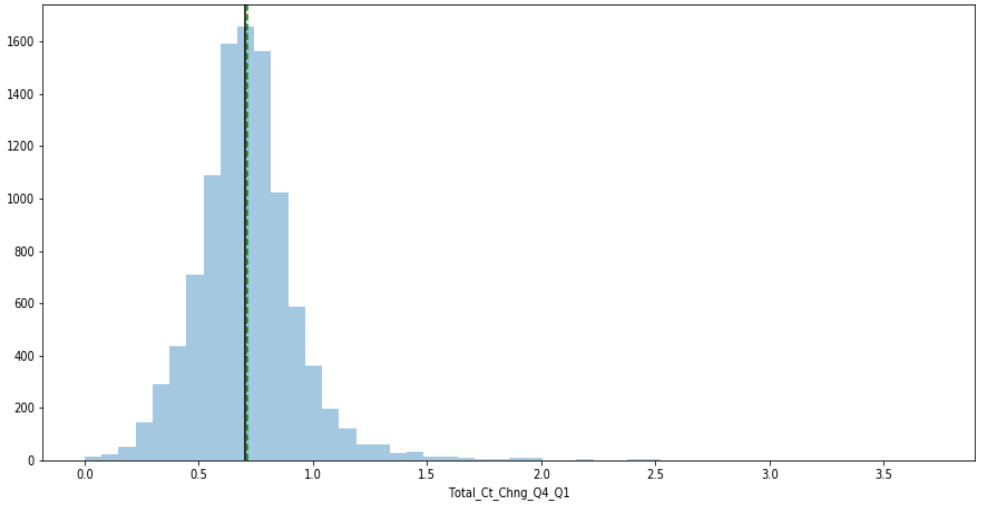


- The distribution peaks at 0, 1 and 2 respectively.
- On the average, 1 child is allowed to go on a visit with a customer.
- 75% of the distribution have 2 children or less who plan to go on a trip with customers

EXPLORATORY DATA ANALYSIS

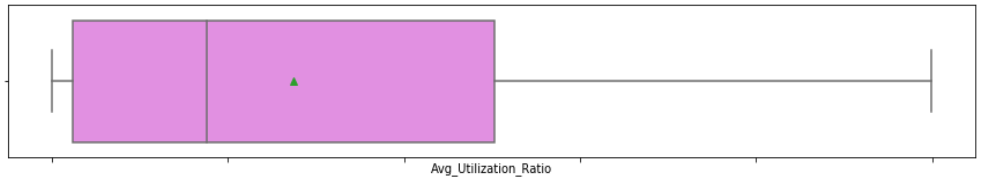
TOTAL_CT_CHNG_Q4_Q1

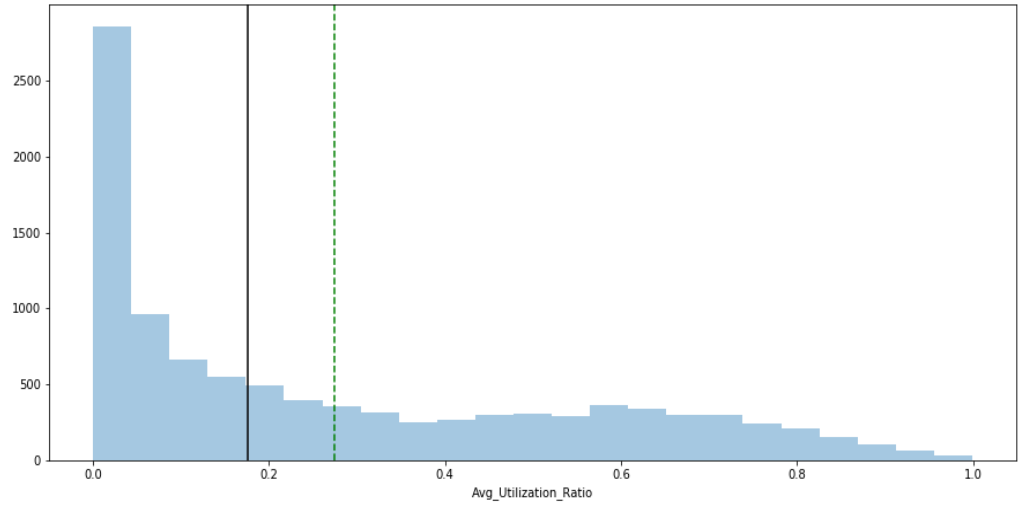




- Monthly Income is a normal ditribution but is Right skewed.
- Tremendous presence of outliers.This will be treated as we proceed.
- Mean Income is 23473 compared to median at 22347
- 75% of the customers earn a monthly income of 25374 or less

AVG_UTILIZATION_RATIO

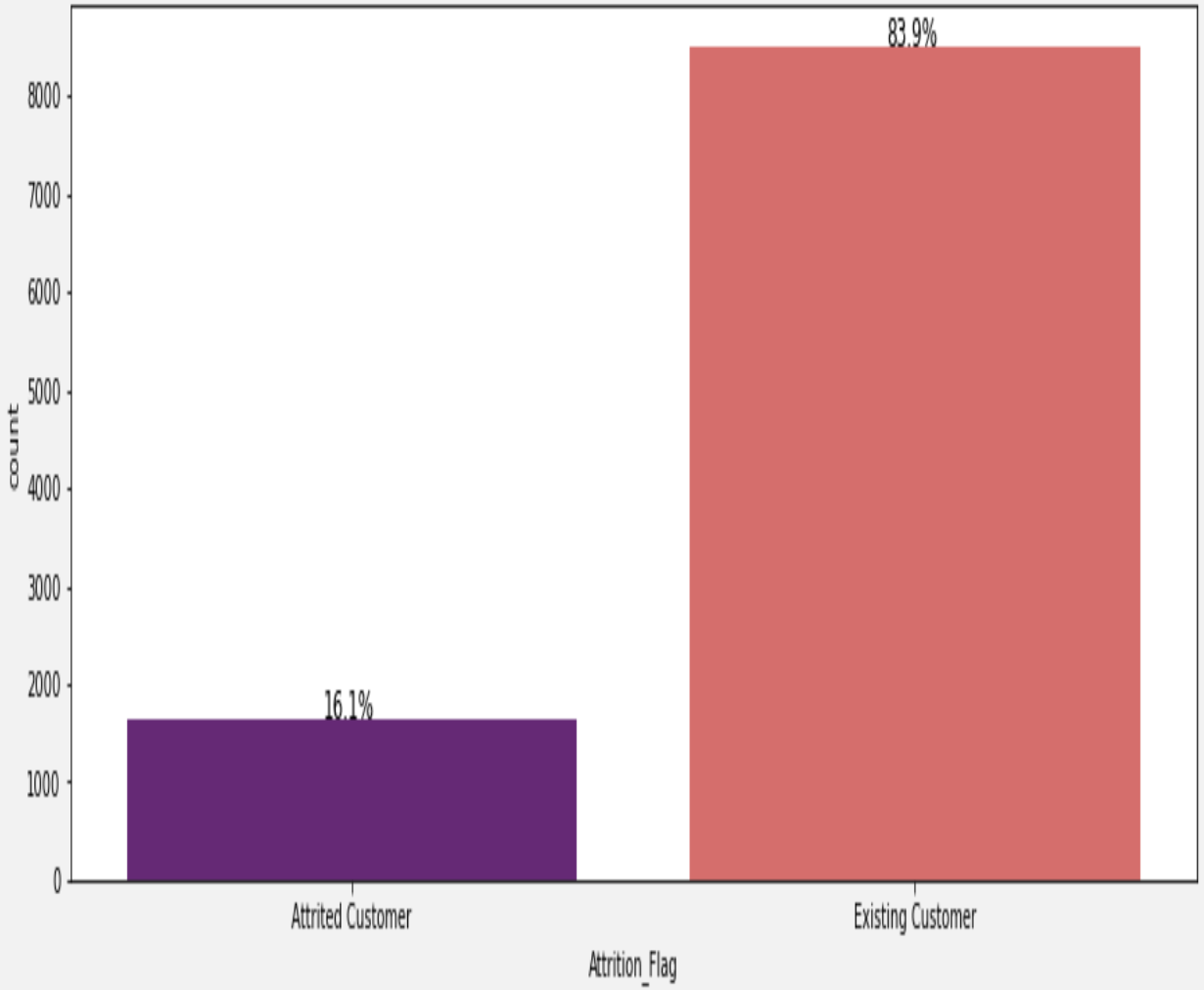




- The distribution peaks at 0, 1 and 2 respectively.
- On the average, 1 child is allowed to go on a visit with a customer.
- 75% of the distribution have 2 children or less who plan to go on a trip with customers

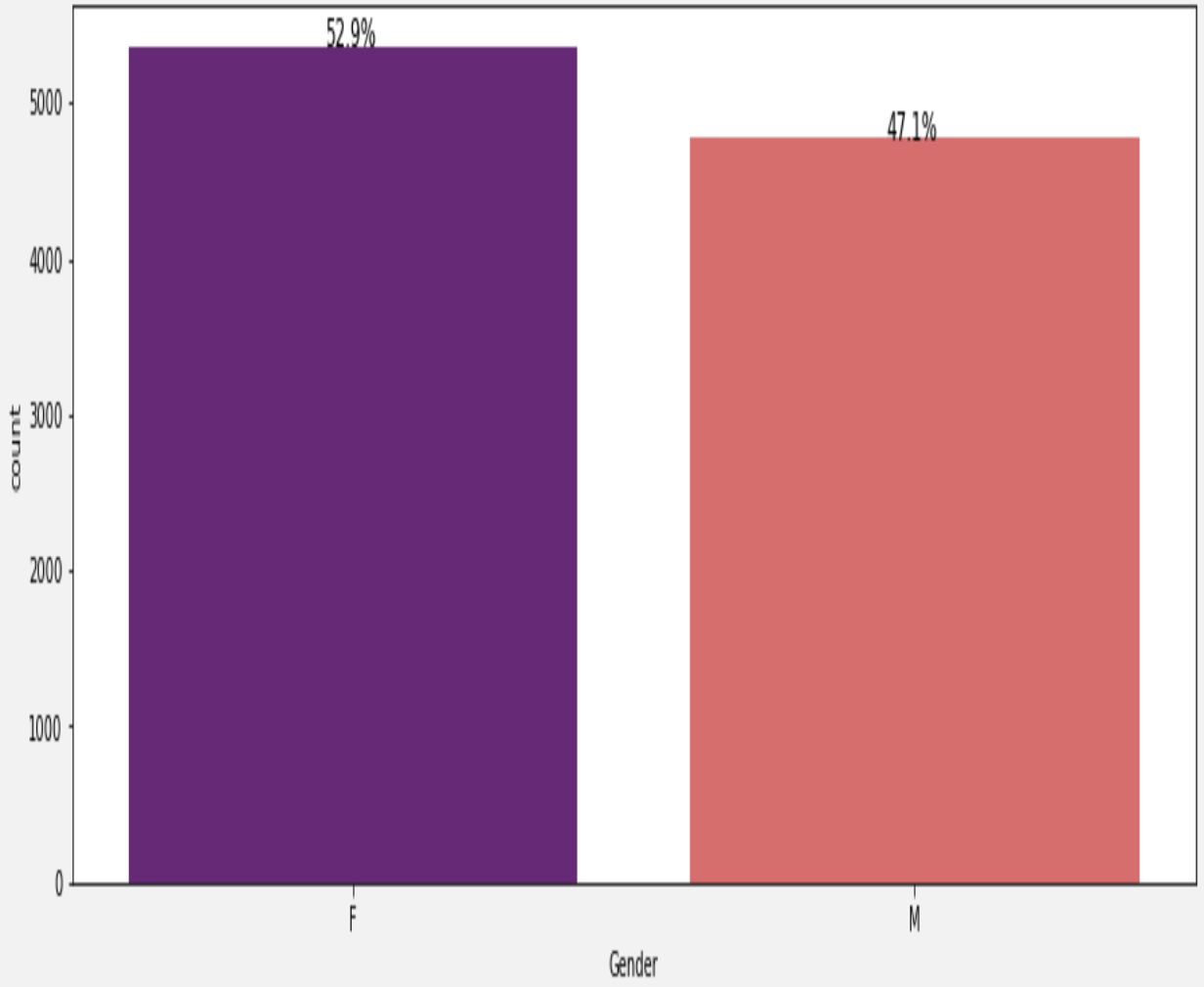
EXPLORATORY DATA ANALYSIS

ATTRITION_FLAG



- 83.9% are Existing customers while 16.1% Attrited.

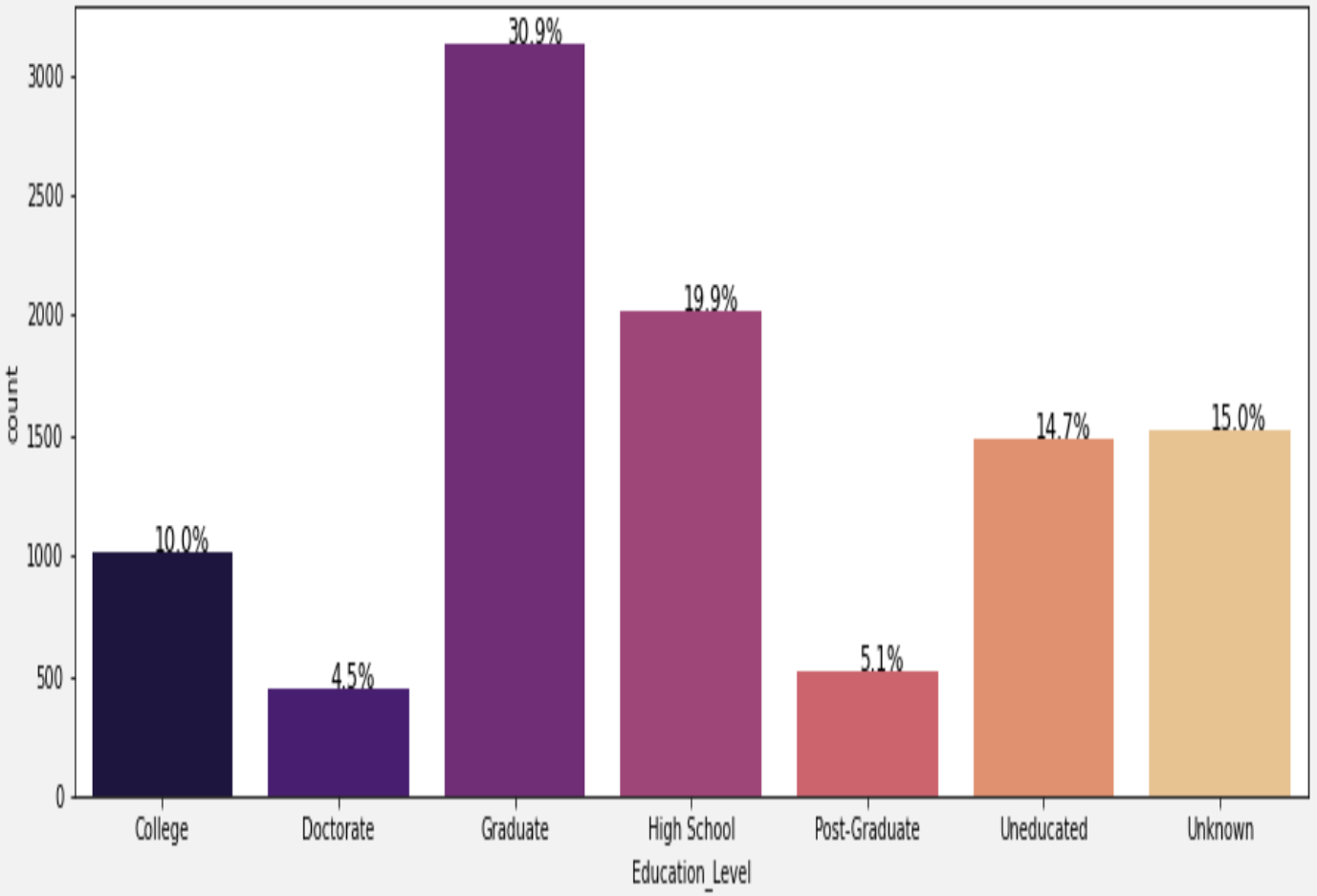
GENDER



- 52.9% are Female while 47.1% are Males

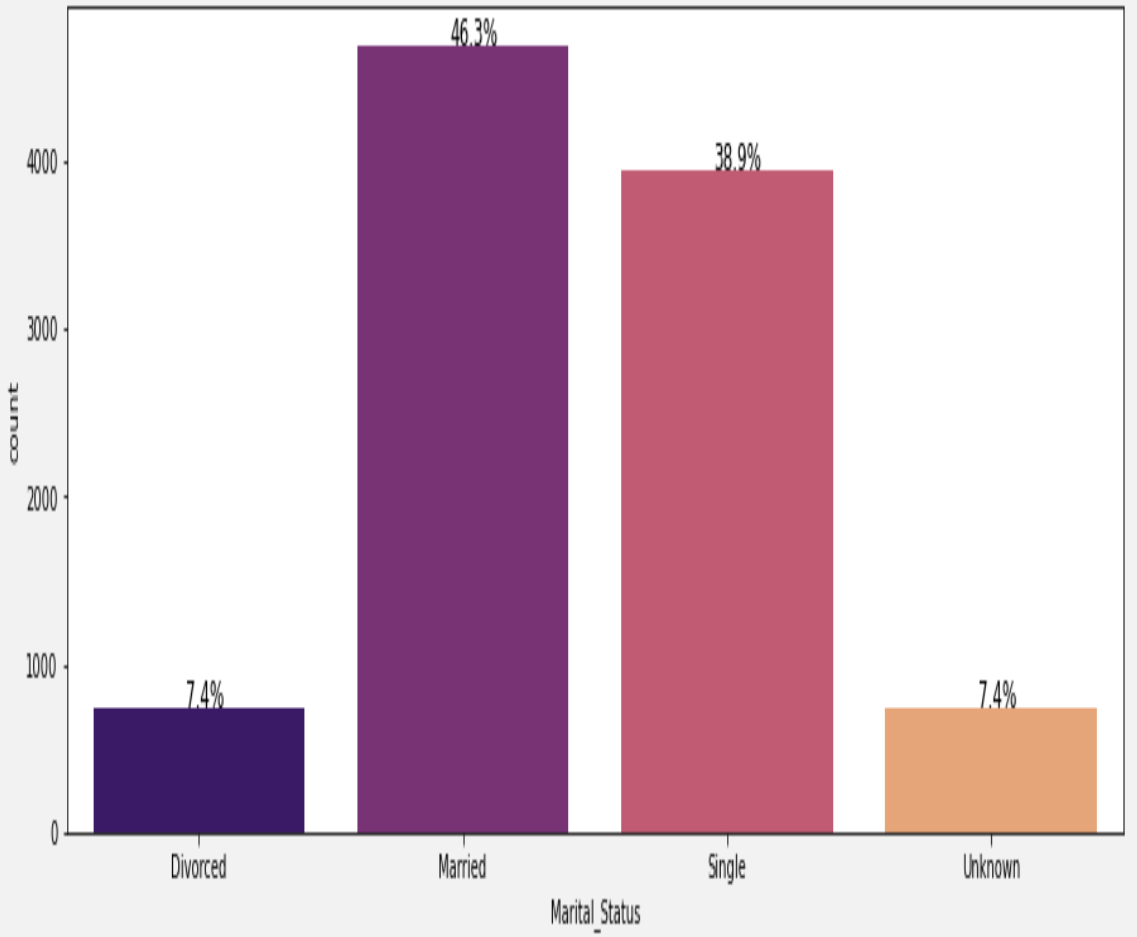
EXPLORATORY DATA ANALYSIS

EDUCATION_LEVEL



- 30.9% of the Graduate customers trump the rest of the customer in other levels of Education
- The High school students follow next with the unknown and uneducated inching closely at 15.0% and 14.7% respectively.

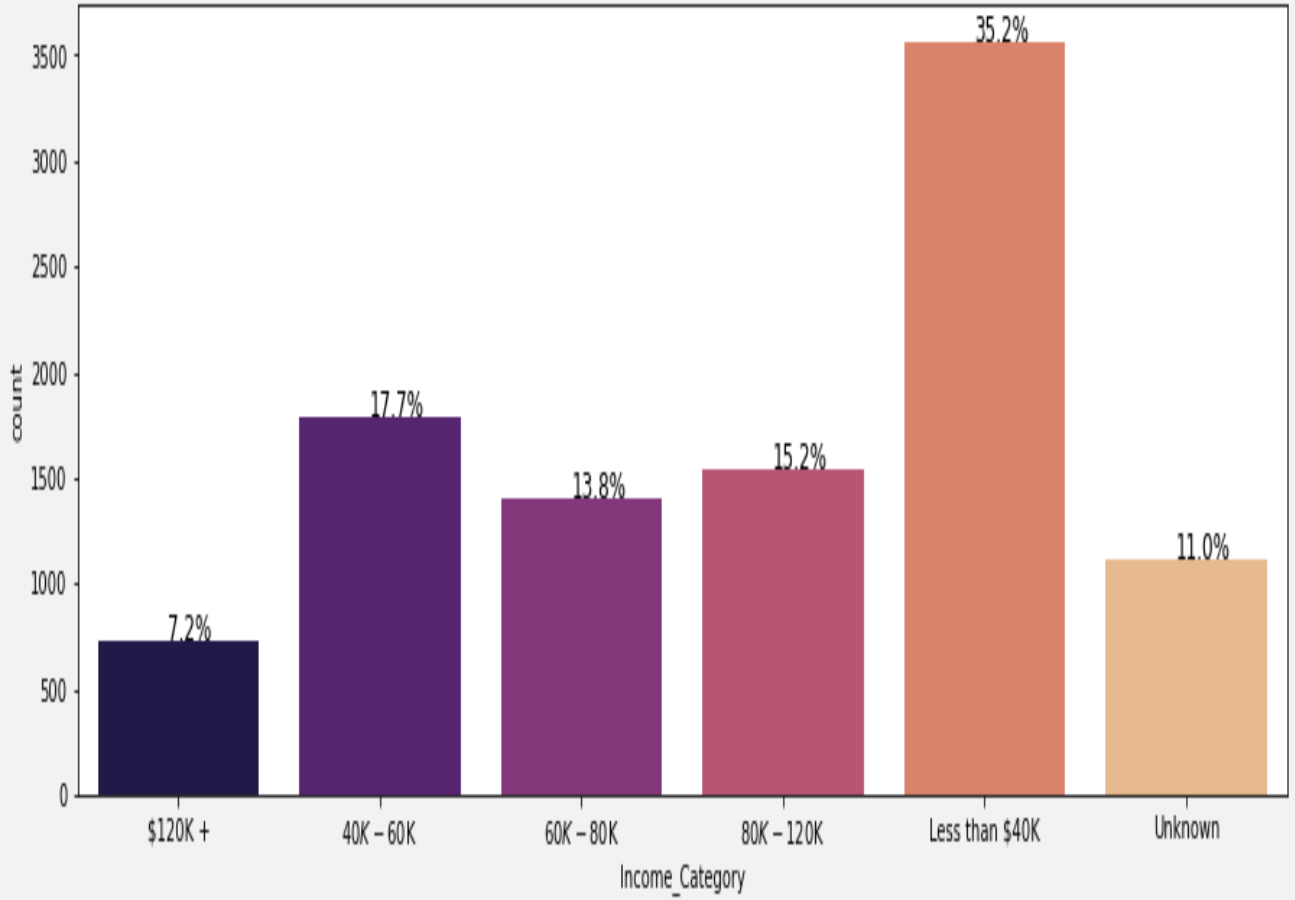
MARITAL_STATUS



- As can be inferred from the plot, 46.3% of the customers are married with Singles following closely at 38.9%
- Divorced and Unknown remain at par with 7.4%

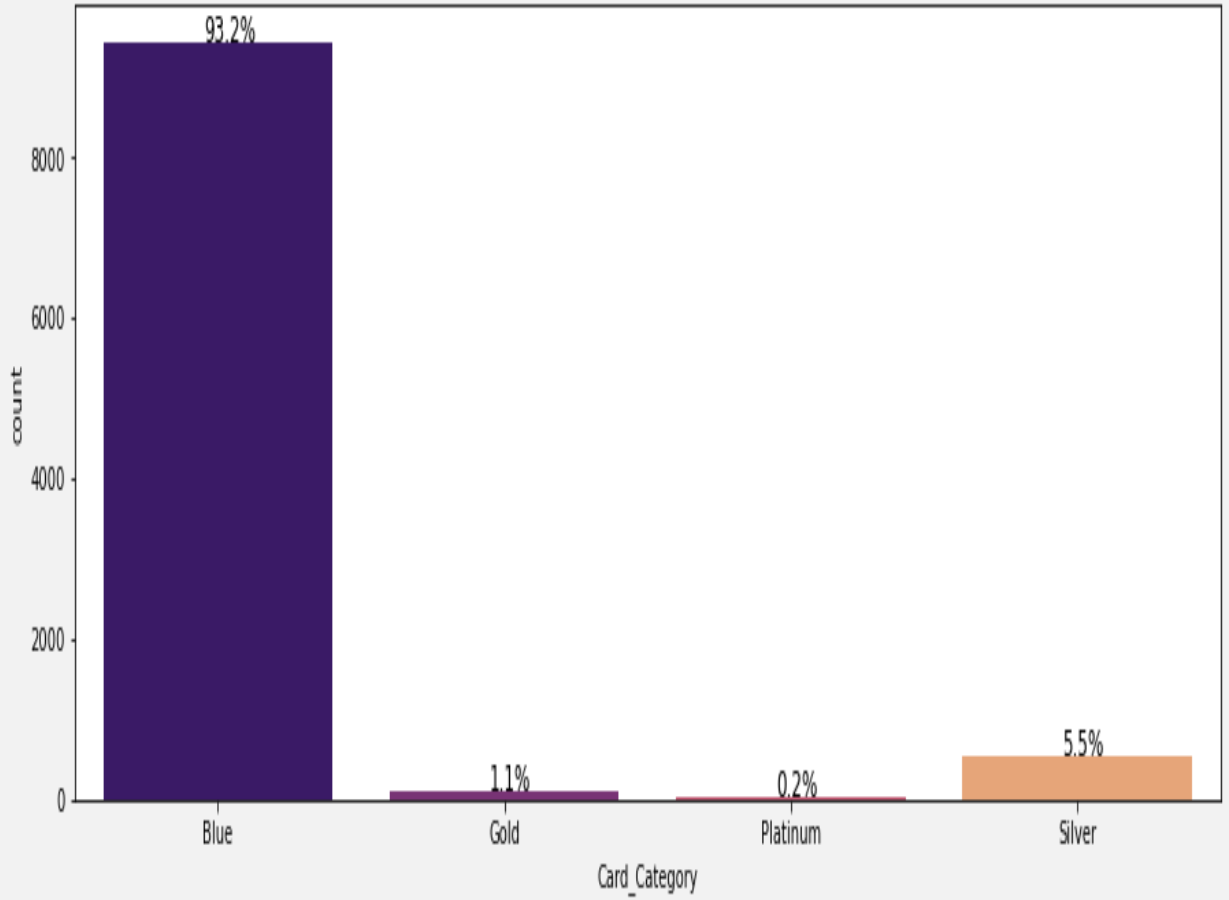
EXPLORATORY DATA ANALYSIS

INCOME_CATEGORY



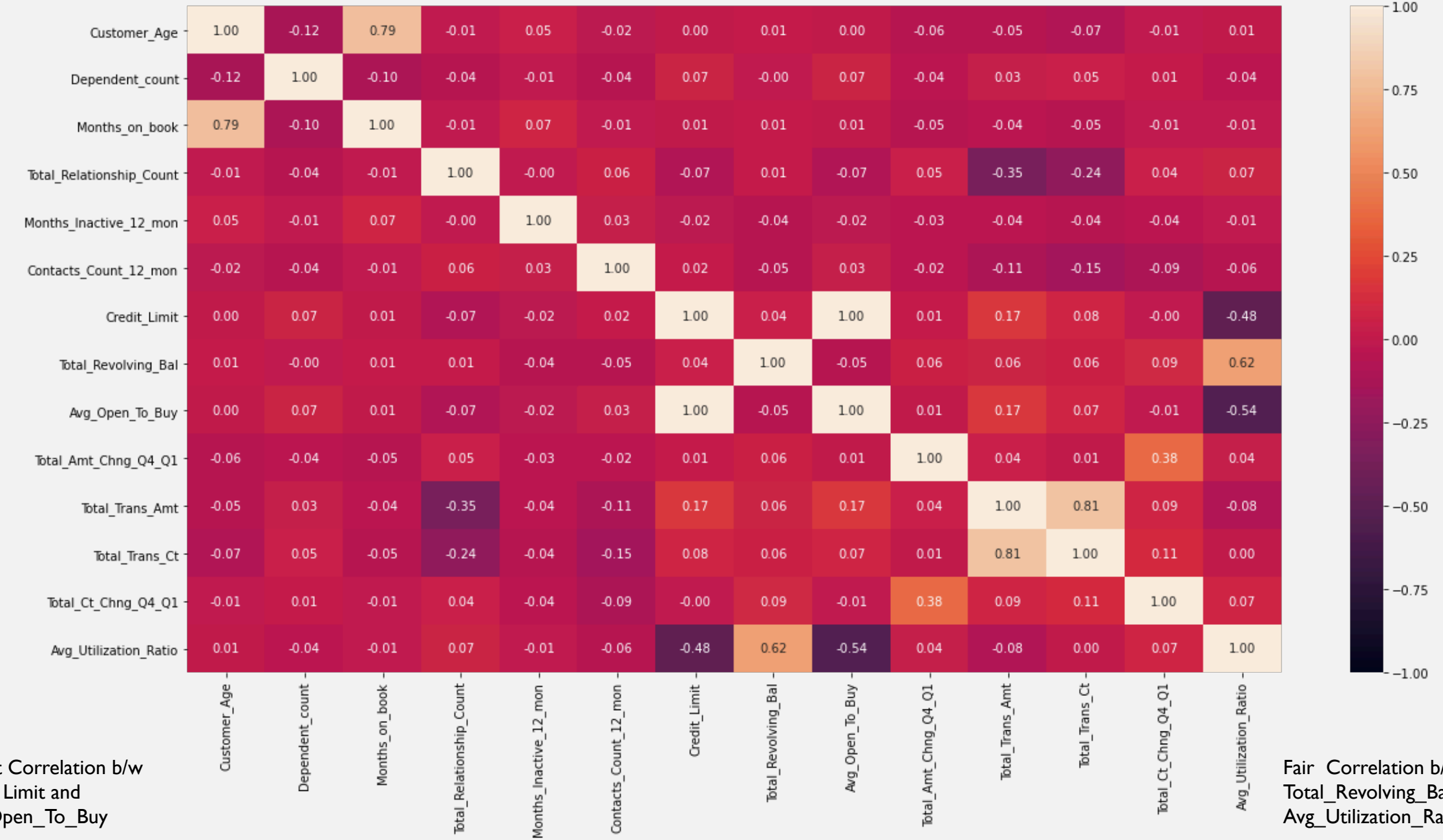
- 35.2% of the customers earn less than 40k as well as 17.7% who earn between 40k-60k p.a
- 7.2% of the customers eanr the highest at 120k+ p.a followed closely by those who earn between 80k-120k

CARD_CATEGORY



- The Blue Card_Category enjoys the most subscription with about 93.2% of the distribution
- The Silver card comes at a distant second with 5.55 and Gold and Platinum respectively attracting the lowest patronage at 1.1% and 0.2%

EXPLORATORY DATA ANALYSIS- HEAT MAP

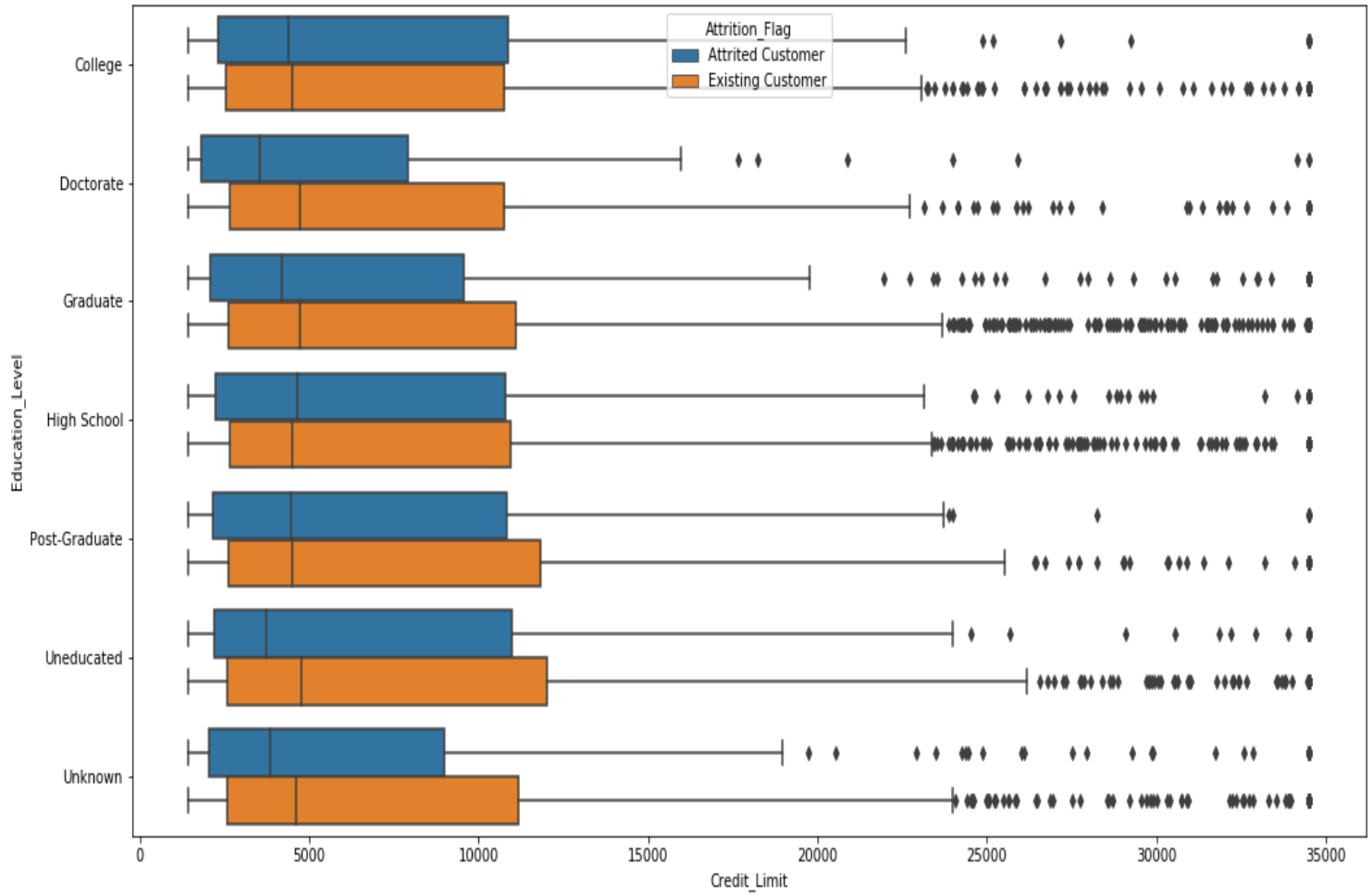


Perfect Correlation b/w Credit Limit and Avg_Open_To_Buy

Fair Correlation b/w Total_Revolving_Bal and Avg_Utilization_Ratio

EXPLORATORY DATA ANALYSIS

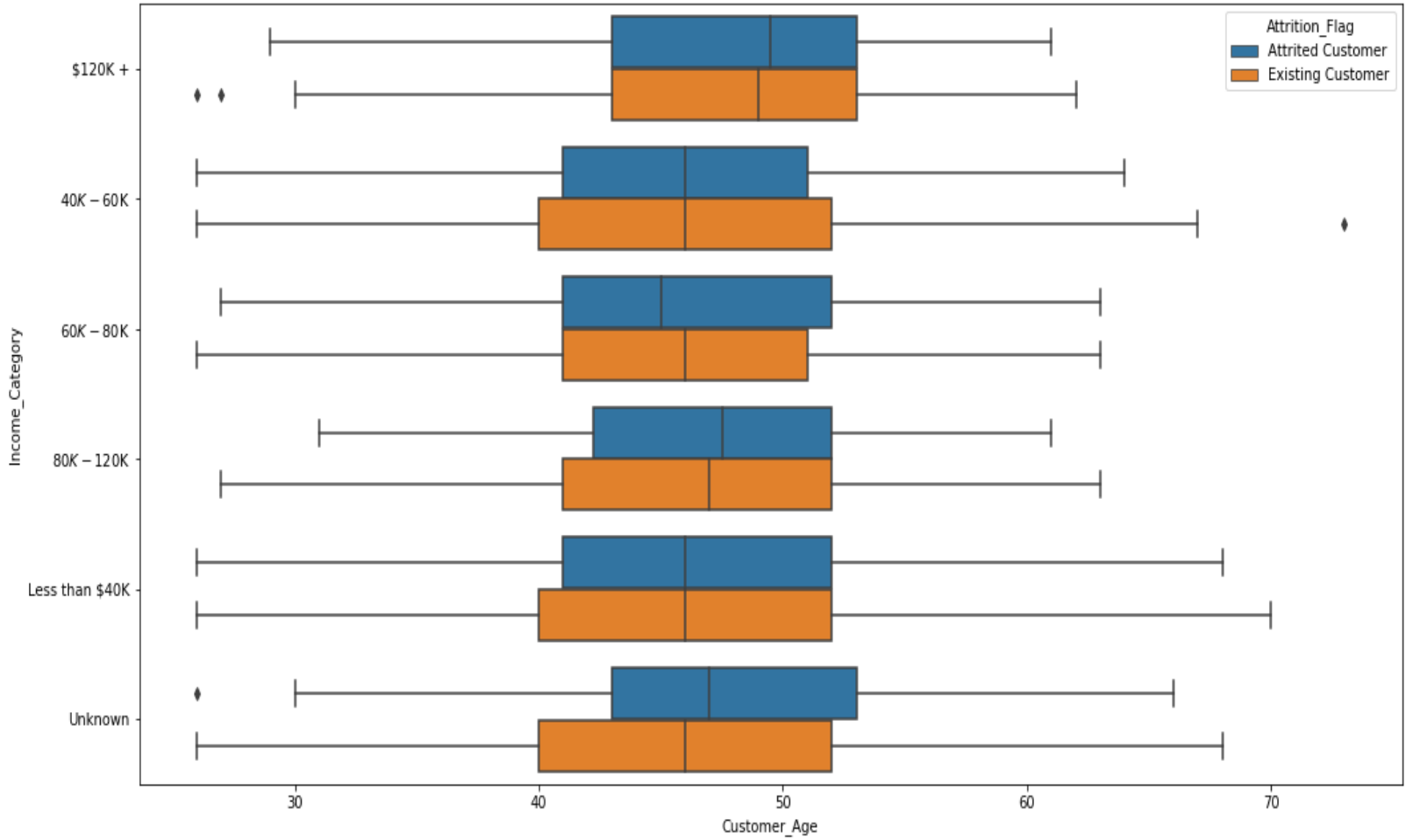
ATTRITION VS EDUCATION VS CREDIT LIMIT



- There are more Existing College customers than there are Attrited College customers but with the latter having access to more credit limits
- There are more Existing Docorate customers than there are Attrited Doctorates with the former enjoying more credit limits
- Thesame goes for the Existing Graduate customers against the Attrited Graduate customers
- Attrited High school customers had access to more credit limites compared to the Existing High school customers evn though the former is higher in population
- Existing Post Graduate customers are both higher in count and credit limites compared to their counterparts
- The exact same sequence speaks to the relationship between the two Uneducated classes
- The Existing Unknown far outweighs the Attrited unknown in population and Credit_limits

EXPLORATORY DATA ANALYSIS

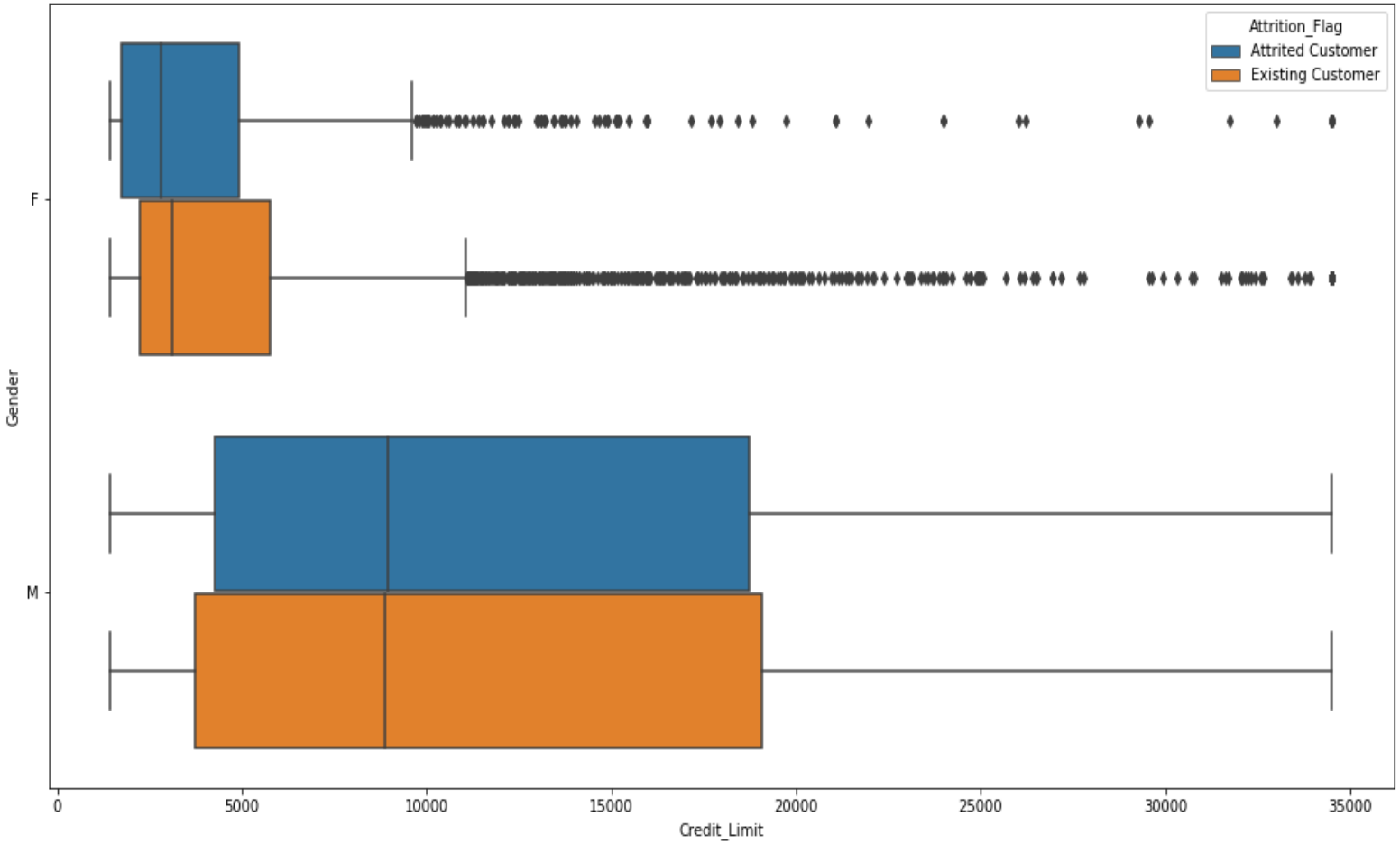
ATTRITION_FLAG VS INCOME_CATEGORY VS CUSTOMER_AGE



- As seen, not much can be inferred concerning the impact of Customer Age on Attrition.
- Most of the customers are between 40 and 54 years of age.
- Oldest Attrition group is from the Unknown group
- Majority of the highest income earning class(120k+) who Attrited are indeed older (near 50 yrs) than the existing customers in thesame income bracket (about 47 yrs)
- Attrited customers in the 60k-80k income bracket are much younger (44yrs) than existing customers (about 46yrs)in the same category.

EXPLORATORY DATA ANALYSIS

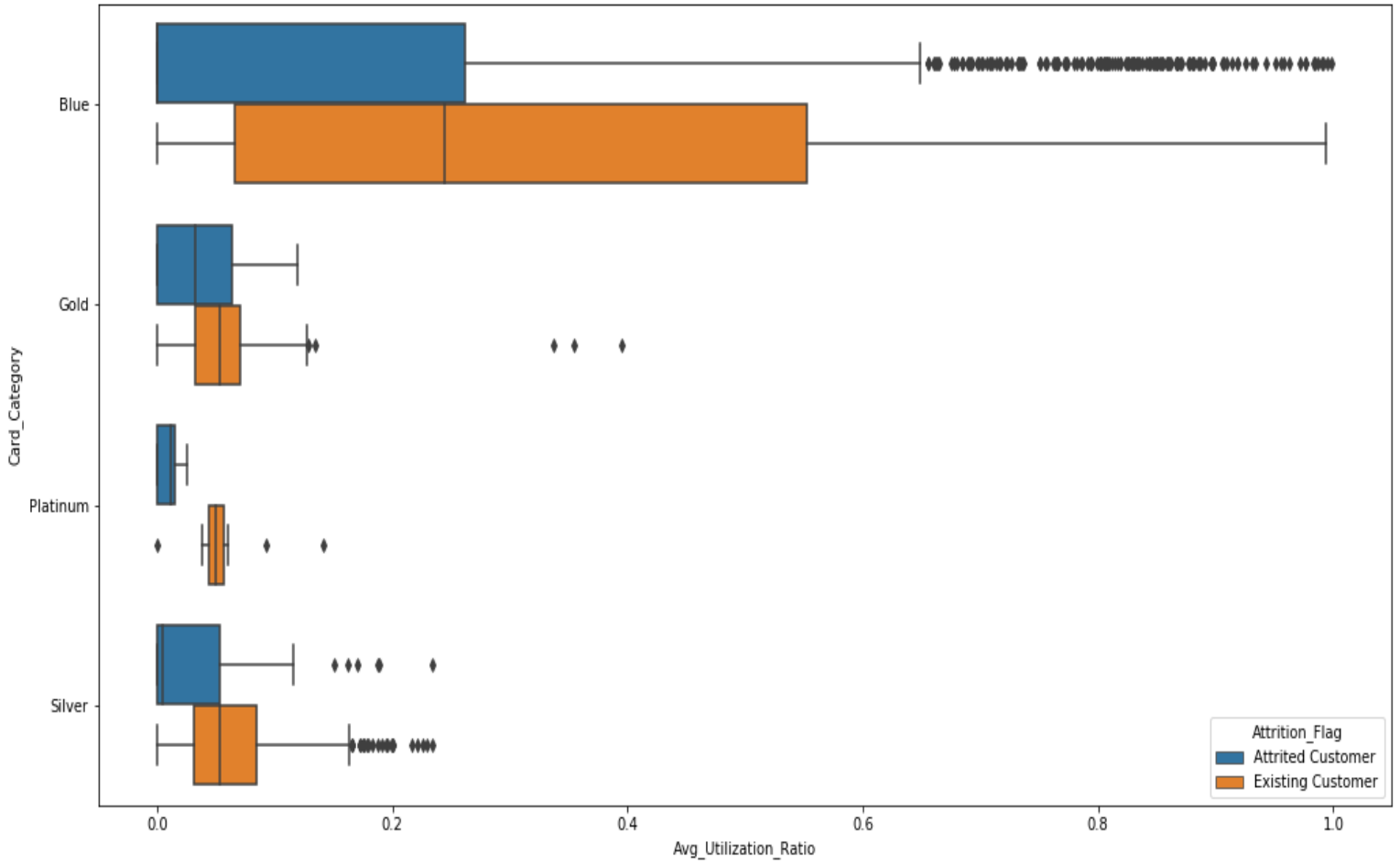
ATTRITION_FLAG VS GENDER VS TOTAL_REVOLVING_BAL



- By and Large, Females enjoy lower credit limits than males.
- Efforts should be made to target High income earning females for increased patronage and retention leading to more fees and profit for the bank
- Attrited females enjoy far lower credit limits compared to the Existing female customers. This strongly speaks to the possible increased attrition rate among females.
- Attrited male customers enjoy a slighlty higher credit limit than their existing partners.

EXPLORATORY DATA ANALYSIS

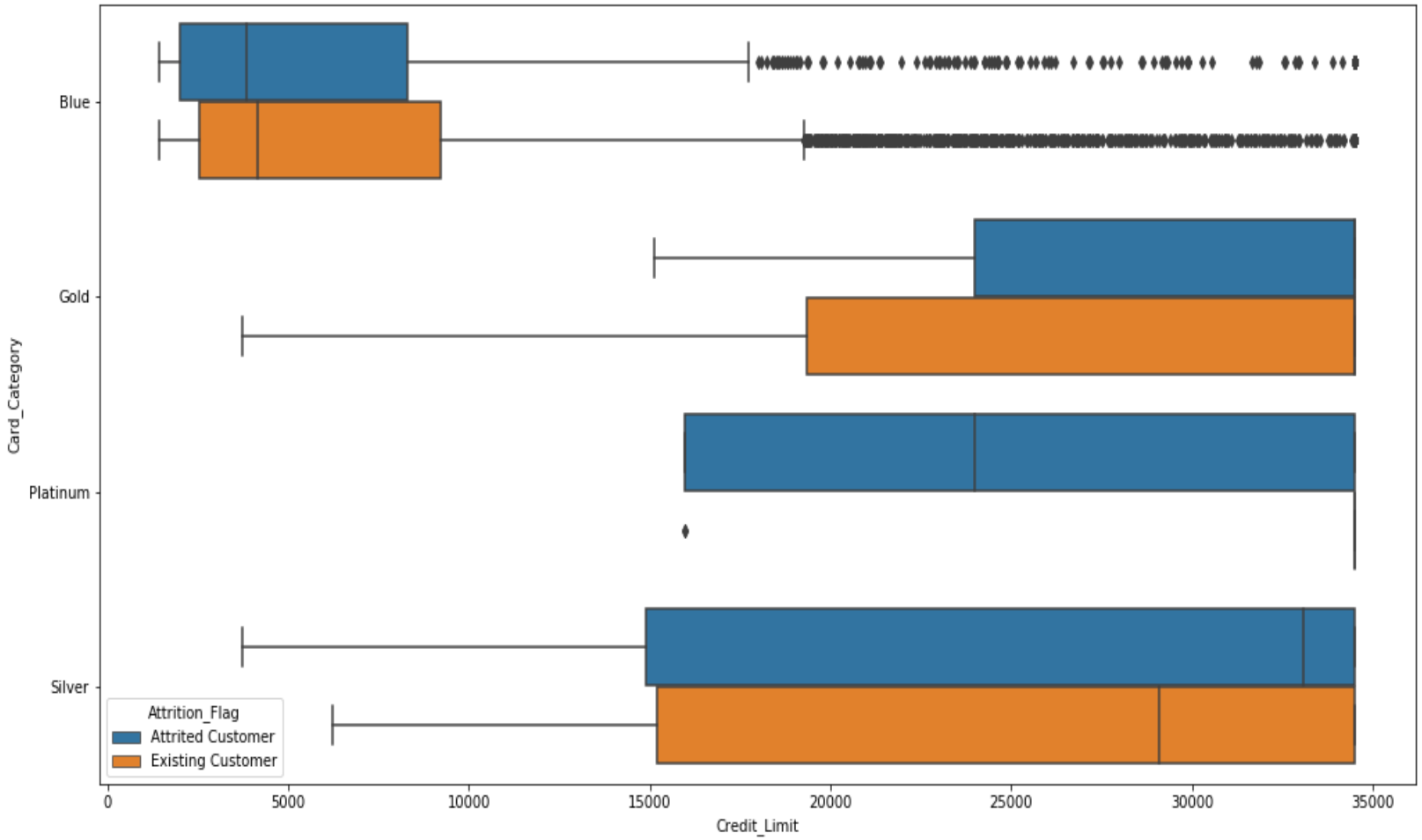
ATTRITION_FLAG VS CARD_CATEGORY VS AVG_UTILIZATION_RATIO



- Across all Card_ Categories, it can be seen that all attrited customers returned very low Average utilization ratios.
- This simply speaks to low Avg_Open_To_Buy as a result of low credit limits , hence the low utilization.
- The Attrited Blue_Card Category customers have the highest Average utilization ratio amongst Attrited customers
- The Existing customers with Blue Card also enjoyed the highest Utilization ratio across the distribution
- The Platinum card holders returned the least Average Utilization Ratio across all card categories

EXPLORATORY DATA ANALYSIS

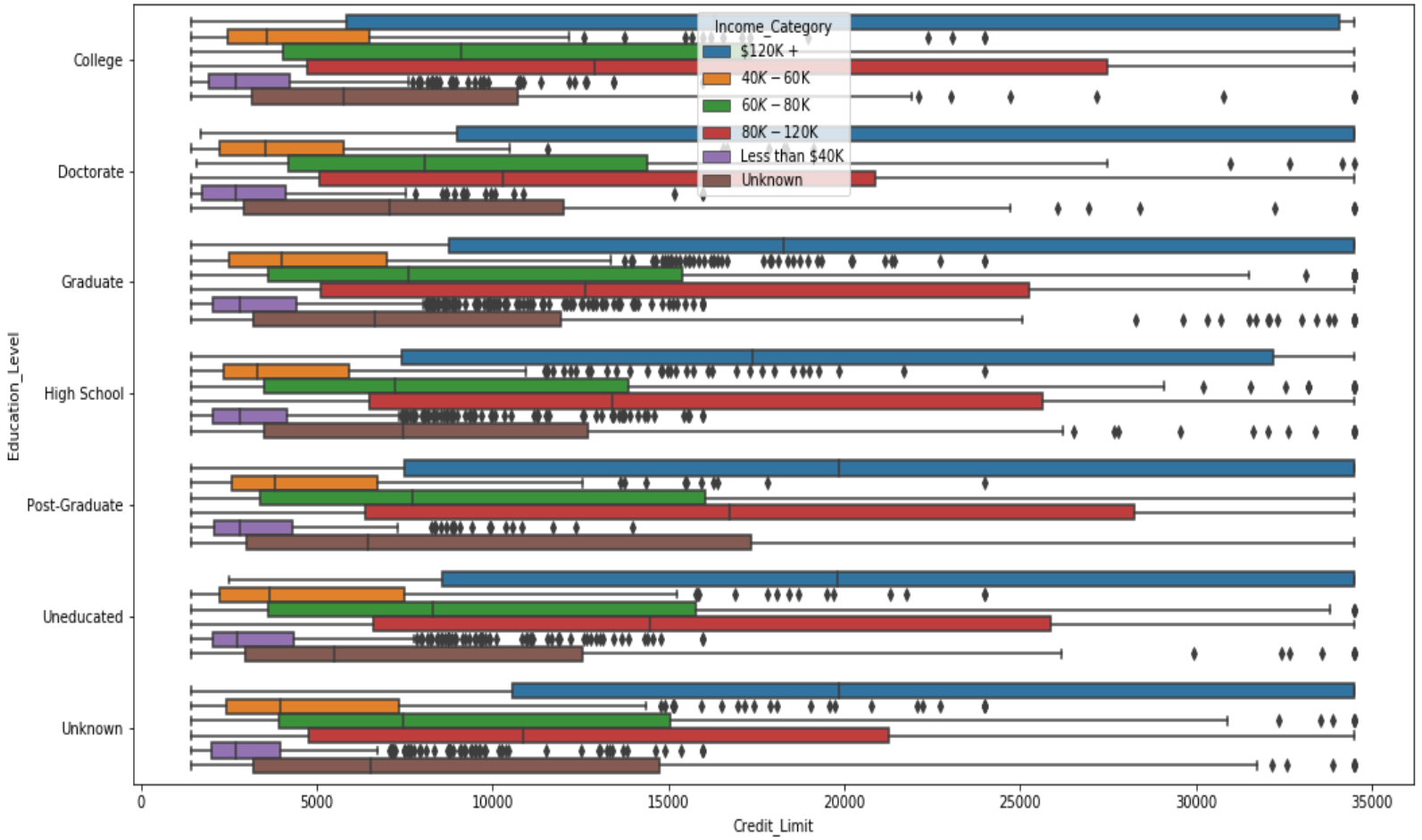
ATTRITION_FLAG VS CARD_CATEGORY VS CREDIT_LIMIT



- The Blue Card category customers attracted the least Credit_Limits but had the highest population of subscribers.
- The Gold card category existing customers got the highest credit limits with very few attritions
- The Platinum Card holders comes next even though they had very high incomes and high credit limits
- The Silver card holders who are existing customers have higher credit limits than their Attrited partners.

EXPLORATORY DATA ANALYSIS

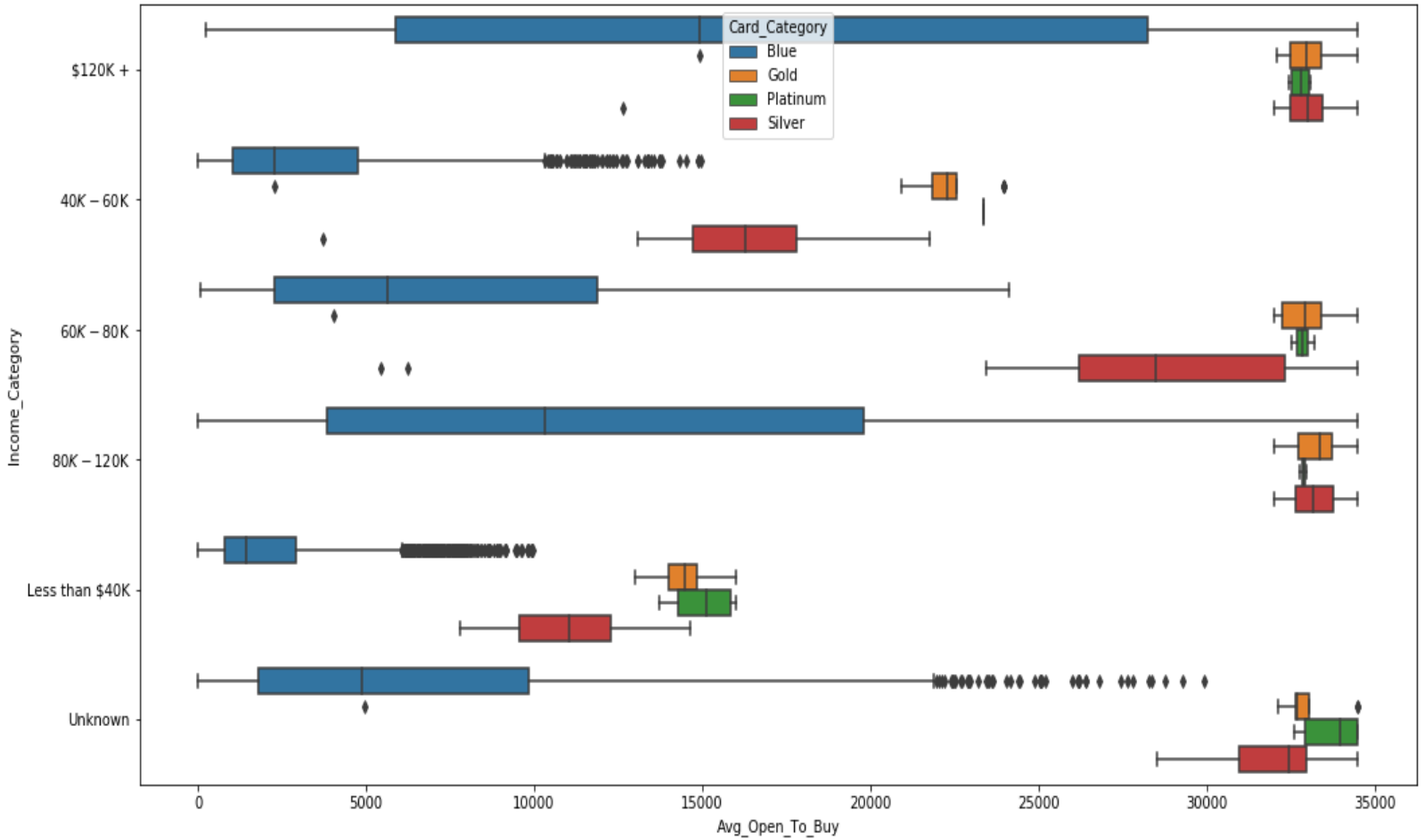
INCOME_CATEOGRY VS CREDIT_LIMIT VS EDUCATION_LEVEL



- Across all Educational Levels, the Highest income category of $120k+ attracts the highest credit limits. This clearly speaks to the higher your income, the higher your credit limit as your
- To a very large extent, income secures the exposure of the bank.
- This is followed by the 80k-120k > 60k-80k >Unknown> 40k-60k >less than 40K

EXPLORATORY DATA ANALYSIS

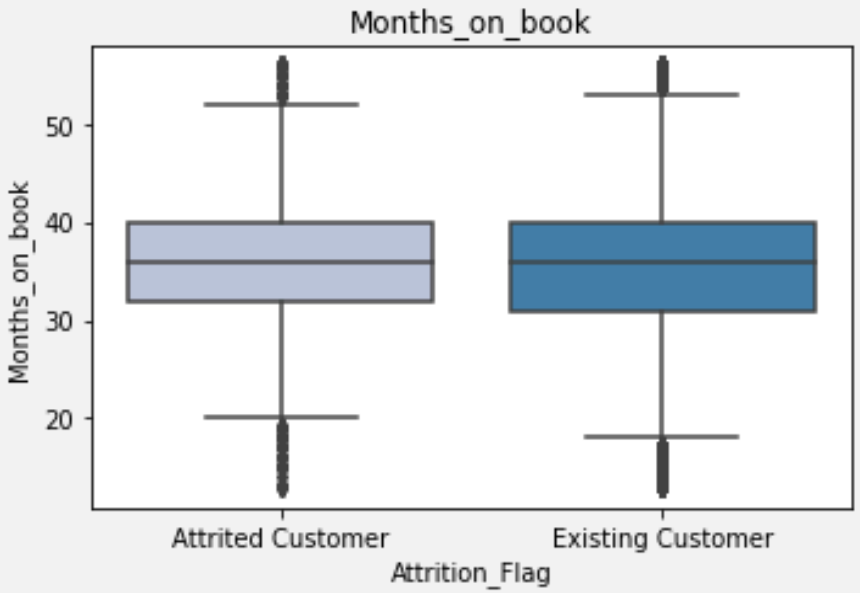
INCOME CATEGORY VS CARD CATEGORY VS AVG_OPEN_TO_BUY

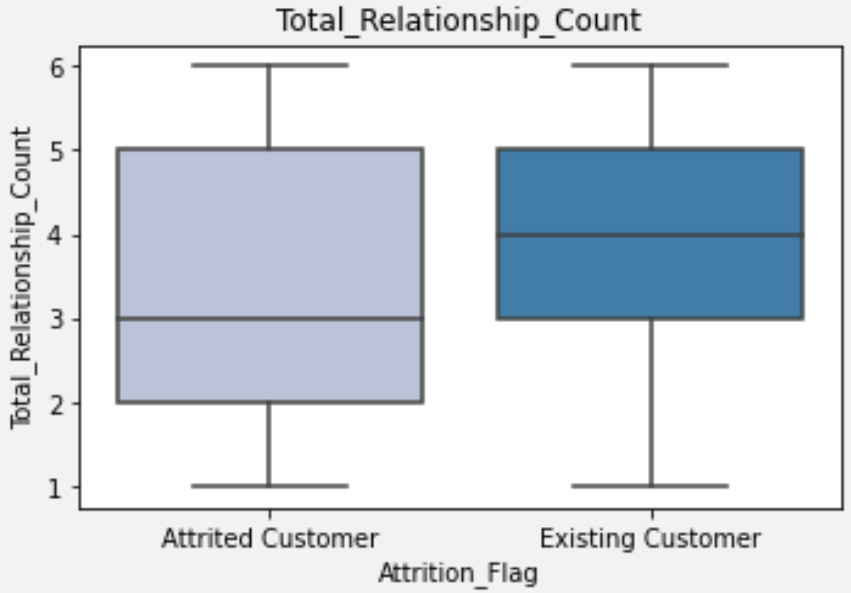


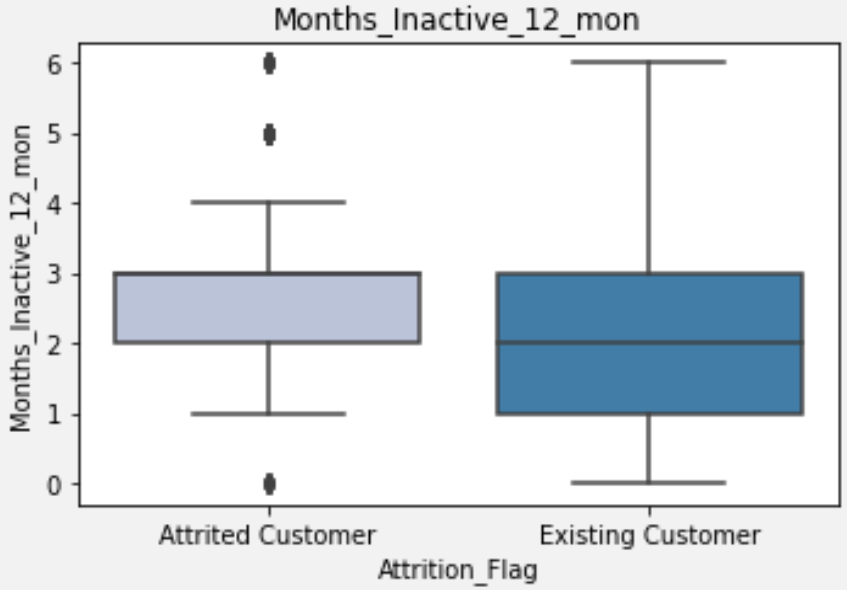
- Gold Card Category with an income of 40k-60k. They have the highest Avg_Open_to_Buy implying higher Total Revolving balances
- Customers who earn an income of 120k+ with a Blue Card come second with the highest Avg_Open_To_Buy implying higher Credit limits and Total Transaction Amounts and Revolving Balances
- Blue card holders who earn less than $40k maintain the lowest AvG_Open_to_Buy and hence the lowest credit limits amongst all card holders

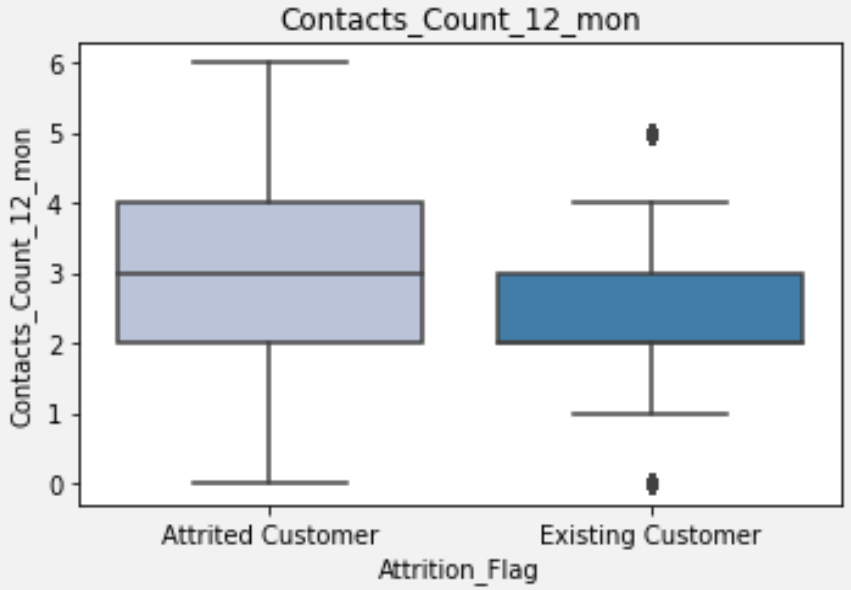
EXPLORATORY DATA ANALYSIS

ATTRITION_FLAG VS BANK RELATIONSHIP





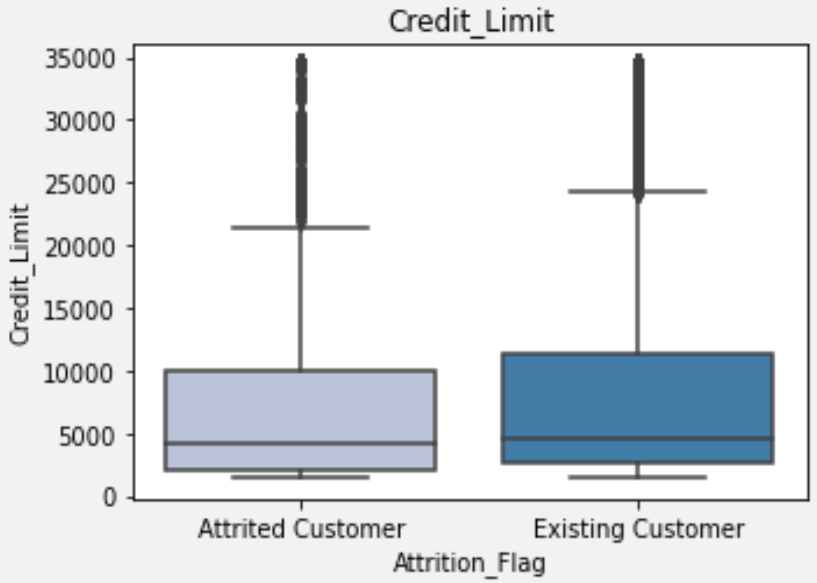


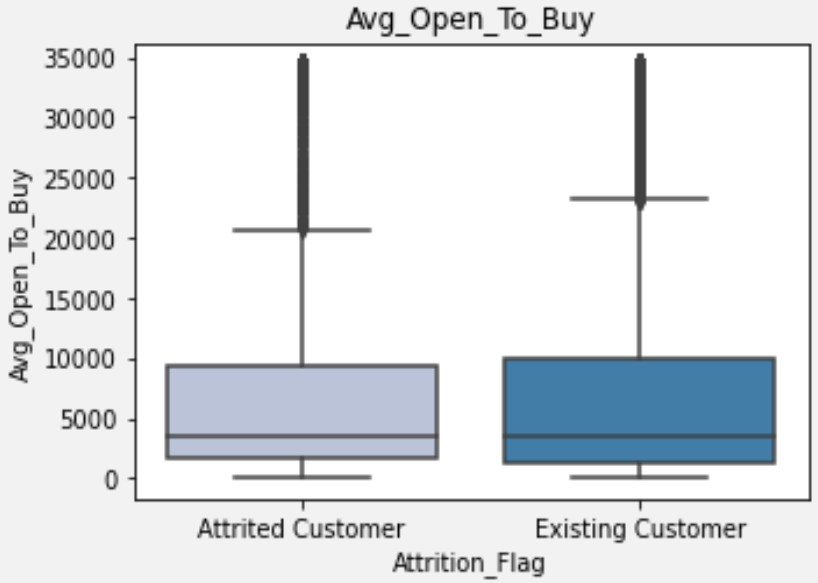


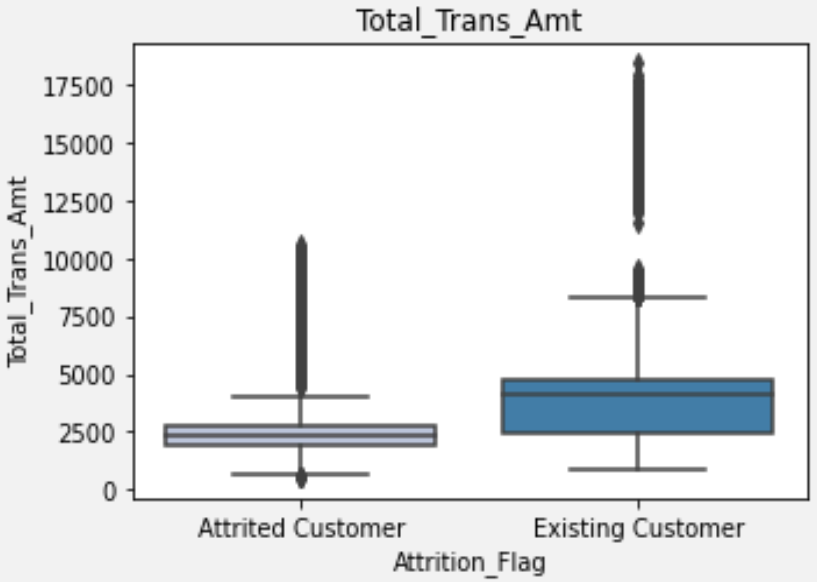
- It would appear as though selling alternate bank products to customers influences retention.
- The more Months_inactive_12_months , the more the likelihood of attrition
- It also appears that the Contacts_Count_12_mon had no impact in customer retention due to other factors

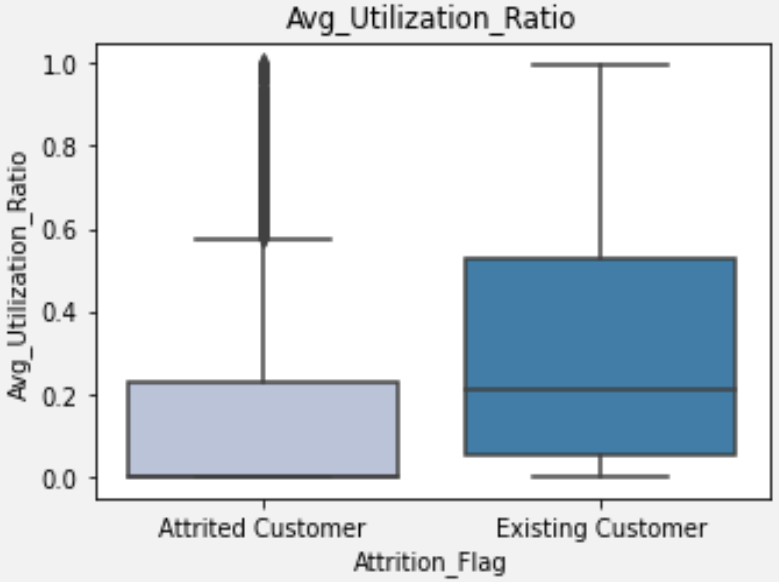
EXPLORATORY DATA ANALYSIS

ATTRITION_FLAG VS TRANSACTION VOLUME VS CREDIT DETAILS





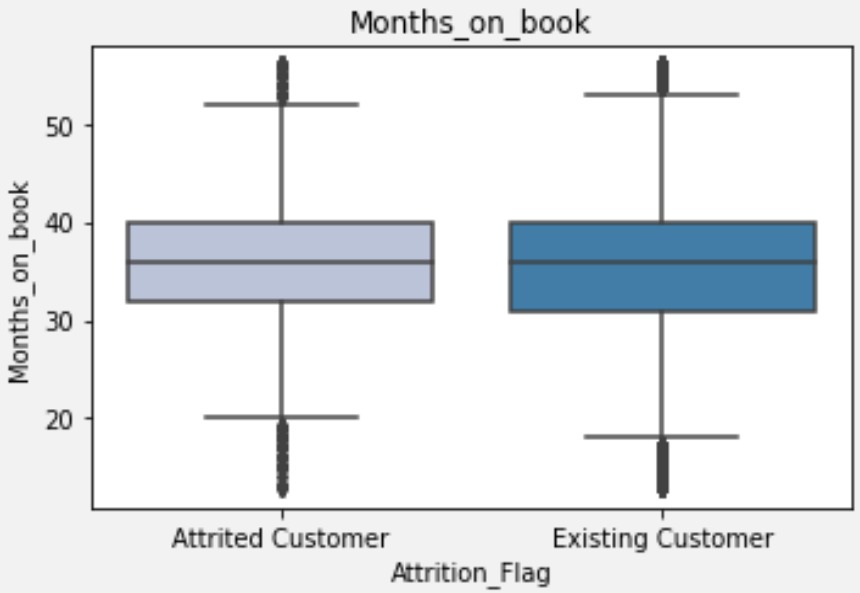


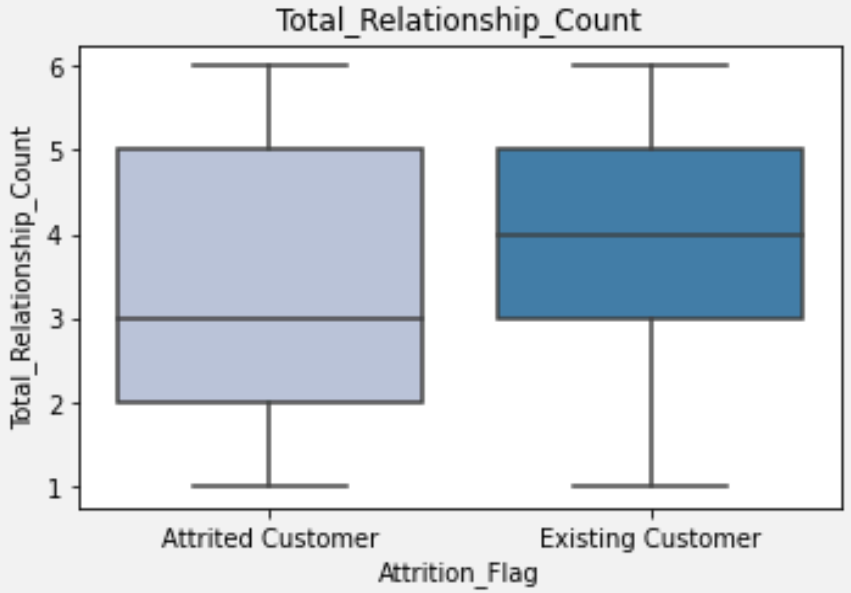


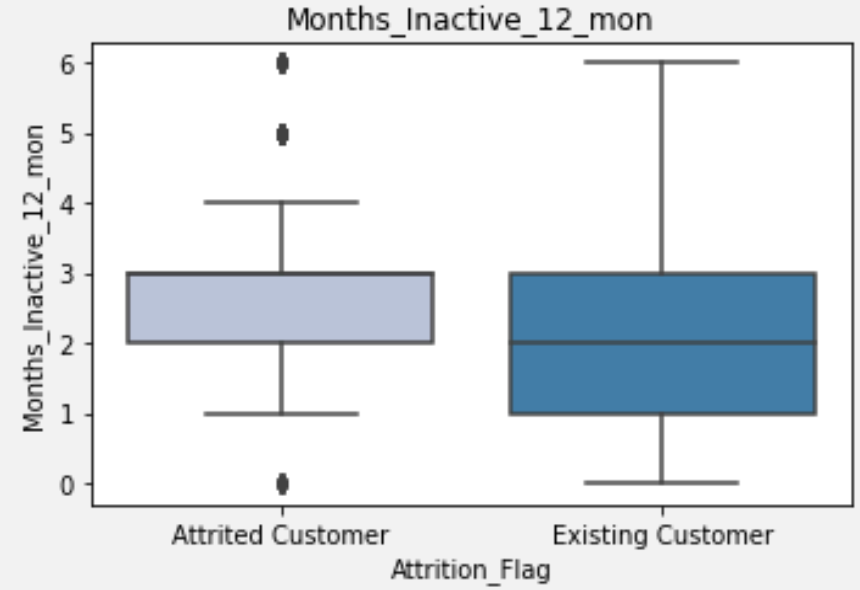
- Existing customers were availed more credit limits compared to attrited customers
- As has been established, a clear correlation between Credit Limit and Avg_open_to_Buy
- Clearly Existing customers have more Avg_open_To_Buy and thus a higher Avg_Utilization_Ratio and hence a higher Total Transaction amount as a result of a Higher Total_Revolving Balance

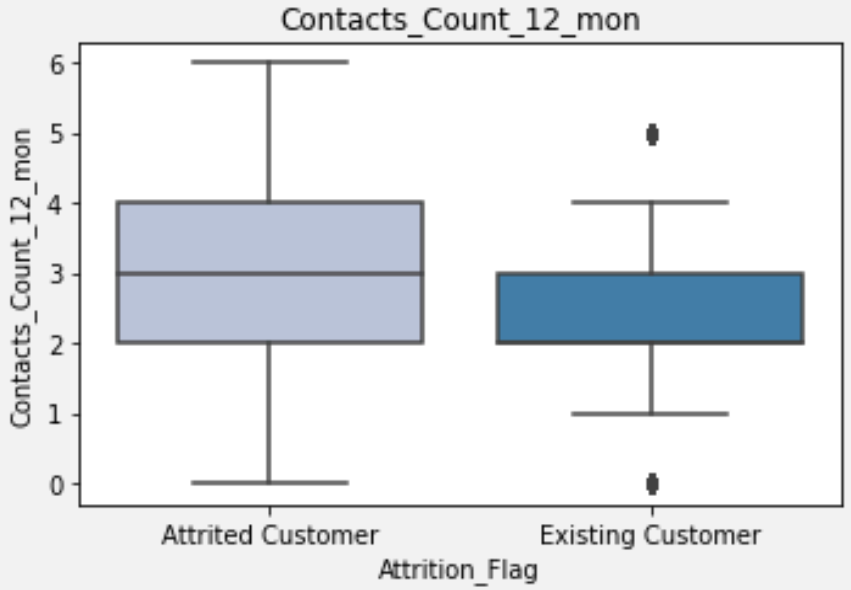
EXPLORATORY DATA ANALYSIS

ATTRITION_FLAG VS BANK RELATIONSHIP





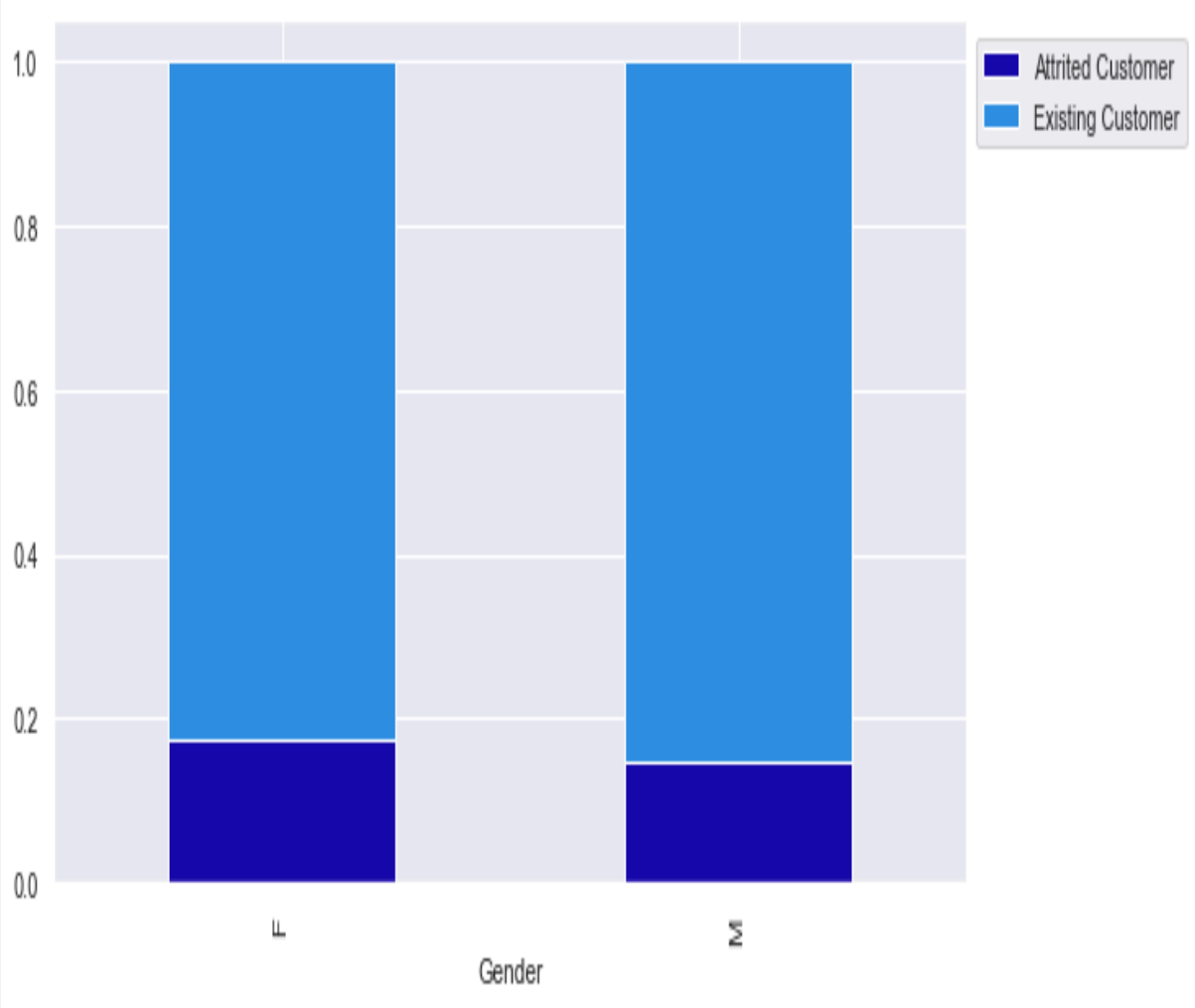




- These visualizations speak to the fact that A customer with a low Total_transaction count implies quite a low Total_Transaction amount and hence a below Avg_Open_To_Buy as a result of a low Total Revolving Balance.
- These all speak to ultimatly a low credit limit
- It would also appear that the older the customer, the higher the likelihood to attrite
- Attrited customers equally have a much lower Total Count as well as Amount Ratio change compared to Existing customers

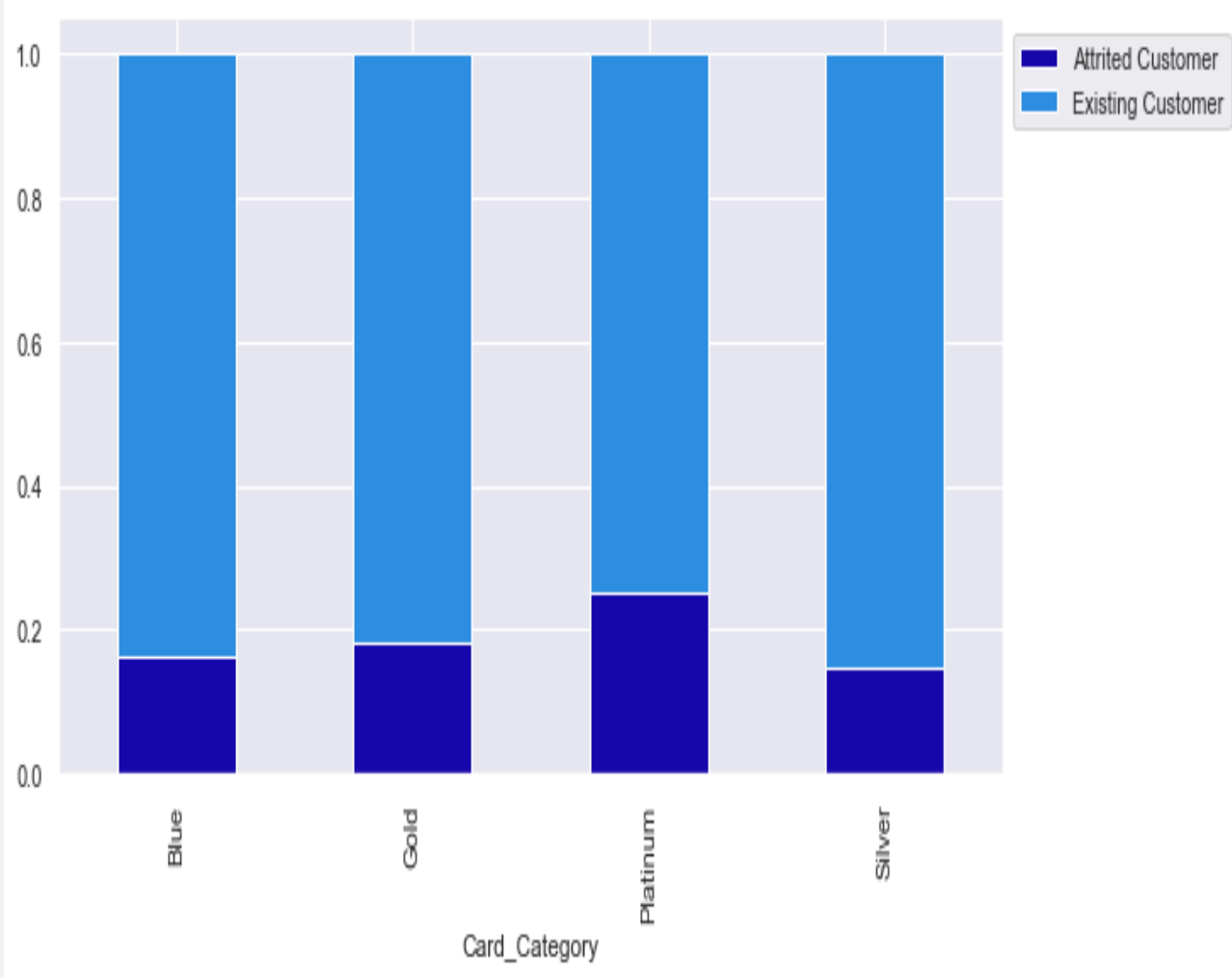
EXPLORATORY DATA ANALYSIS

STACKED GENDER



- About 18% of Female customers attrited while roughly 16% of Male Customers attrited

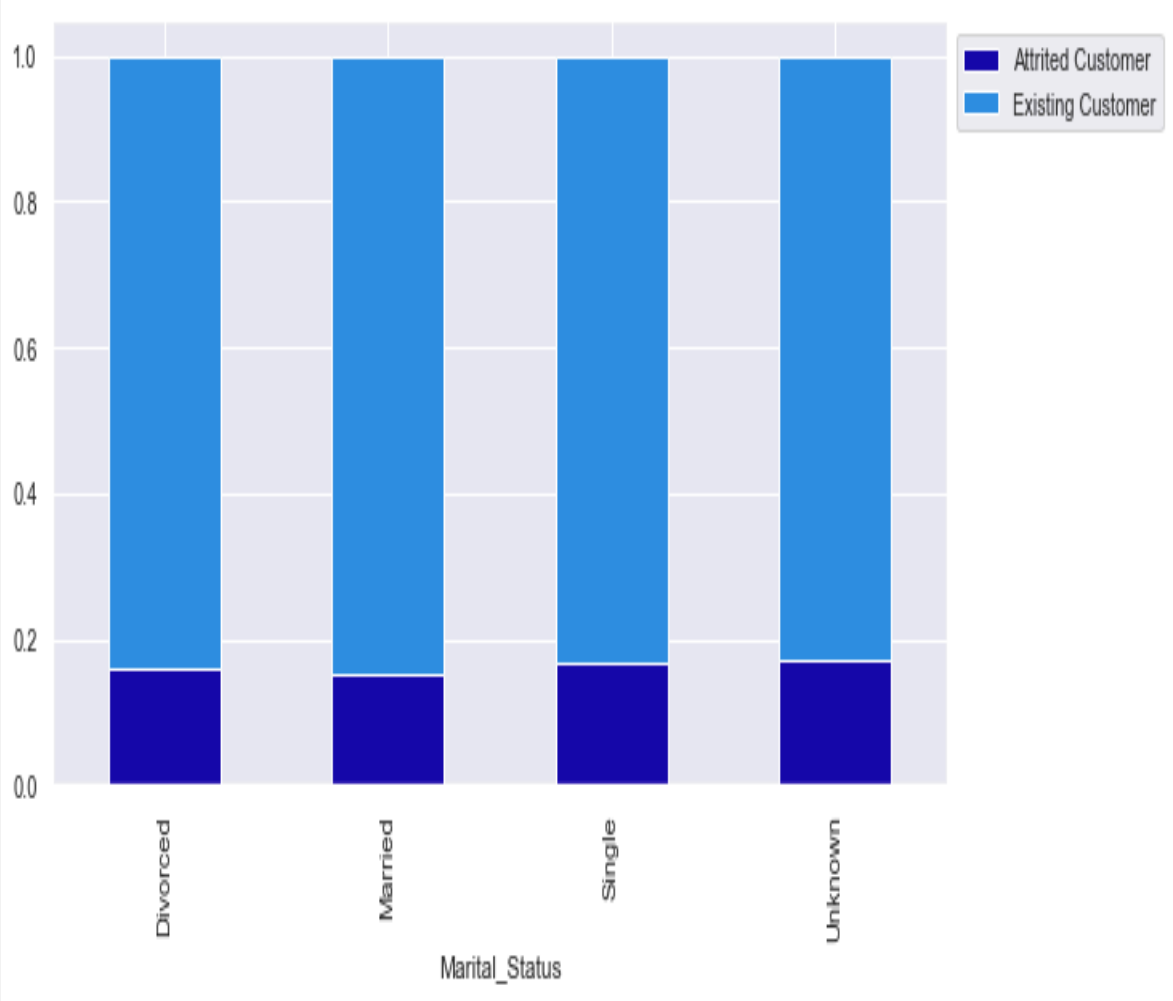
STACKED CARD_CATEGORY



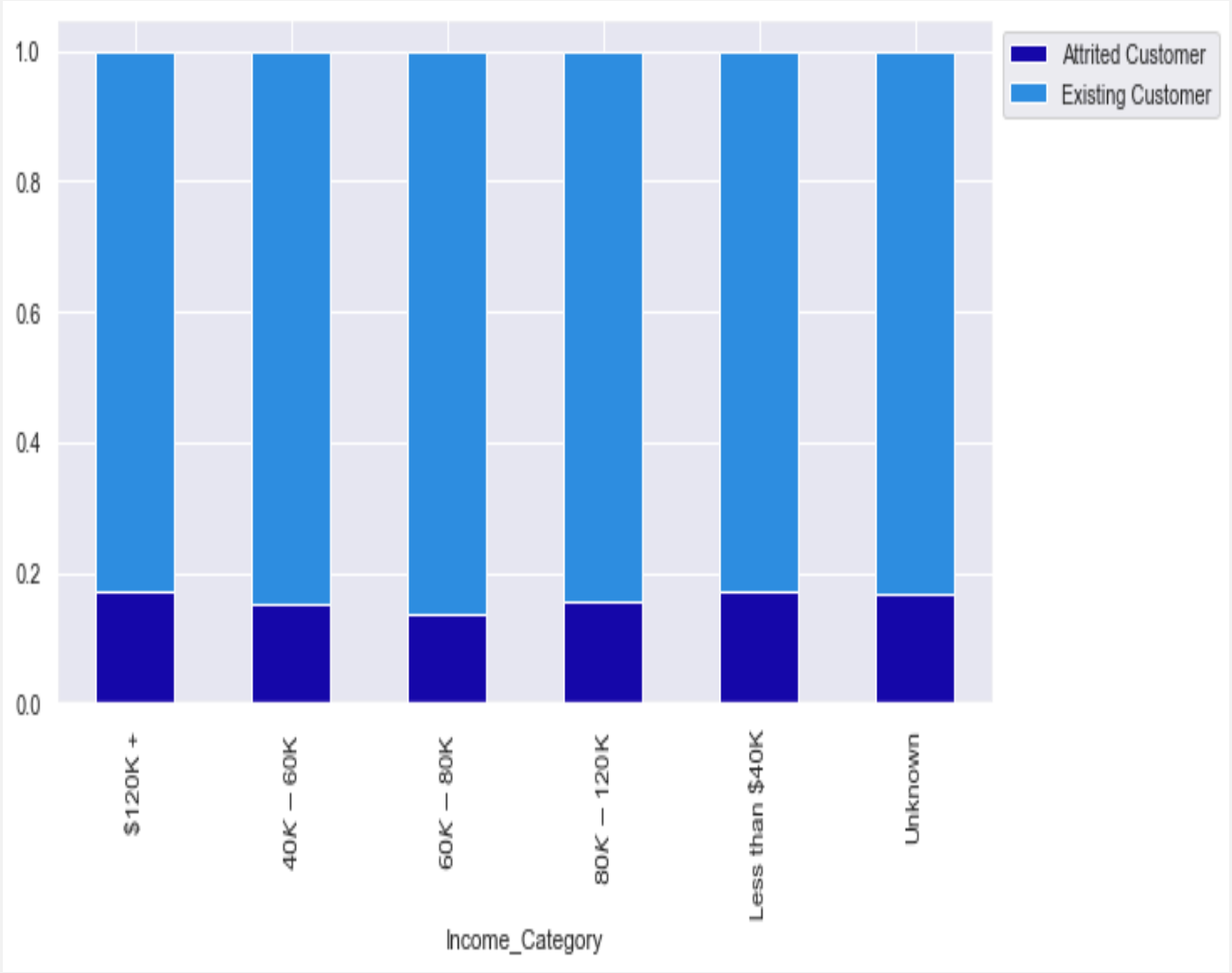
- Roughly 28% of the Platinum card holders attrited the most.
- The Gold card category comes next with 18%

EXPLORATORY DATA ANALYSIS

STACKED MARITAL_STATUS



STACKED INCOME CATEGORY

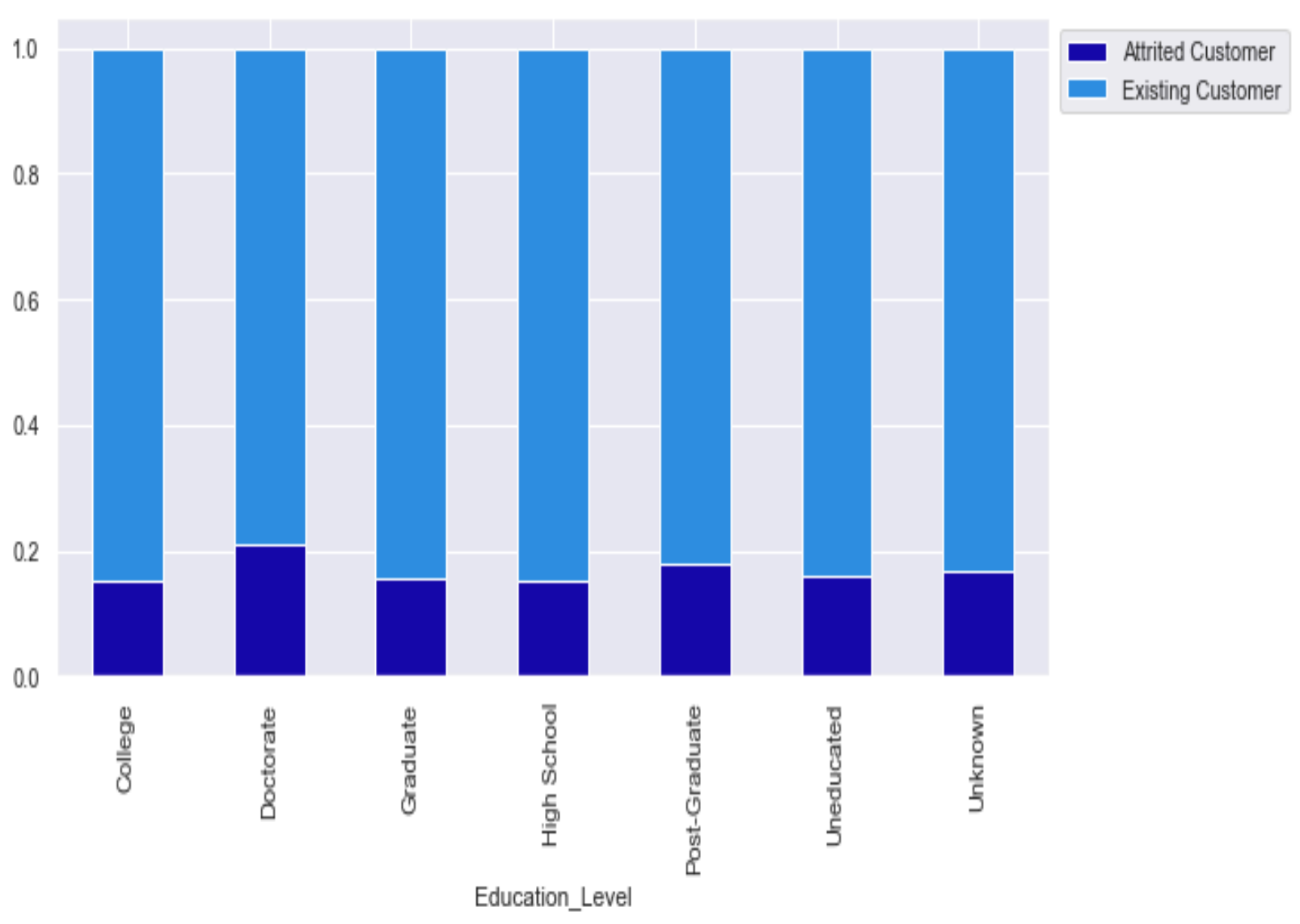


- Unknowns attrited the most at about 18% closely followed by Singles at 17% with Divorced and married at par

- Customers who earn less than $40k + attrited the most at 18.5%.
- The next most attrited group is the 120K group at 18%

EXPLORATORY DATA ANALYSIS

STACKED EDUCATION_LEVEL



- Roughly 21% of the Doctorate Customers attrited which by far trumps the rest of the other categories. This is closely followed by the Post_Graduates at roughly 19%.
- Unknown comes immediately after with about 18% attrition

CUSTOMER LEVEL ANALYSIS BY CARD CATEGORY

BLUE

- There are 7,917 Existing customers
- The Blue Card customers average 46 years in age with 5101 females and 4335 Males
- They have an average of 2 dependents per customer with 4433 who are Married.
- There are 2899 Graduate Blue card holders with 3403 of these class of card holders earning less than $40k pa
- There are 9436 Blue card holders with an average of 36 months in the Bank's books with an avg. of approx. 4 other products
- This category has an average of 2 contacts and 2 inactive months in a year
- They have an average credit limit of 7364 with a Total revolving Balance of 1157.8 per month
- With an Average Open to Buy of 6205 and a Total mean Transaction amount of 4225 amounting to 64 transaction counts in total, the Blue Card category indeed is quite a dominant feature

SILVER

- There are 473 Existing customers
- The Silver Card customers average 45.6 years in age with 210 females and 345 Males
- They have an average of 2 dependants per customer with 251 who are Singles.
- There are 185 Graduate Silver card holders with 3403 of these class of card holders earning less than $40k pa
- There are 555 Silver card holders with an average of 35.5 months in the Bank's books with an avg. of approx. 3 other products
- This category has an average of 2 contacts and 2 inactive months in a year
- They have an average credit limit of 25278 with a Total revolving Balance of 1206 per month
- With an Average Open to Buy of 24072 and a Total mean Transaction amount of 6590 amounting to 75 transaction counts in total, the Silver Card category indeed is quite a promising pool of young prospects.

CUSTOMER LEVEL ANALYSIS BY CARD CATEGORY

GOLD

- There are 95 Exisitng customers
- The Gold Card customers average 45 years in age with 38 females and 78 Males
- They have an average of 2.6 dependants per customer with 58 who are Singles.
- There are 36 Graduate Gold card holders with 29 of these class of card holders earning 60k-80k pa
- There are 116 Gold card holders with an average of 35.5 months in the Bank's books with an avg. of approx. 3 other products
- This category has an average of 2 contacts and 2 inactive months in a year
- They have an average credit limit of 28416 with a Total revolving Balance of 1344.32 per month
- With an Average Open to Buy of 27,072 and a Total mean Transaction amount of 7686 amounting to 38 transaction counts in total, the Gold Card category is indeed is quite a very good prospect basket to harvest.

PLATINUM

- There are 15 Existing customers
- The Platinum customers average 47.5 years in age with 4 females and 11 Males
- They have an average of 2 dependants per customer with 10 who are Singles
- There are 8 Graduate Platinum card holders with 5 of these class of card holders earning Unknown class of income
- There are 20 Platinum card holders with an average of 36 months in the Bank's books with an avg. of approx. 2 other products
- This category has an average of 2 contacts and 2.5 inactive months in a year
- They have an average credit limit of 30284 with a Total revolving Balance of 1268 per month
- With an Average Open to Buy of 29016 and a Total mean Transaction short of 9000 amounting to 87 transaction counts in total, the Platinum Card category is indeed quite Exclusive for a select few

MODEL EVALUATION CRITERION

Model can make wrong predictions as:

Predicting a customer will attrite and the customer doesn't
Predicting a customer will not attrite and the customer actually does

Which case is more important?

Predicting that a customer will not attrite but actually does i.e. losing on a valuable customer that would have added value to Thera Bank

How to reduce this loss i.e need to reduce False Negatives?

Thera Bank wants Recall to be maximized, the greater the Recall the higher the chances of minimizing false negatives. Hence, the focus should be on increasing Recall or minimizing the false negatives or in other words identifying the true positives(i.e. Class 1) so that the bank can provide value oriented solutions, perks and incentives to retain customers thereby optimizing the overall initiative to increase credit card applications and ultimately income vis-a-vis turnover in transaction volumes.

MODEL PERFORMANCE SUMMARY

OVERVIEW OF ML MODELS AND PARAMETERS

- Prior to Modeling, Outliers were treated and further Data-Preprocessing was done to identify independent variables viable for the prediction process
- Using KNN_Imputer, Missingness was handled
- Dummy variables using a function that automates On-Hot encoding for a more surgical approach. Education was hot encoded
- Data was Split into Train and Test sets at a 70:30 Ratio
- Modelling was executed using Logistics Regression (OverSampling and UnderSampling), Bagging and Boosting Models
- HyperParameters Tuned using GridSearchCV and RandomSearchCV and results compared for the best 3 models
- Model Comparison of Metrics scores for Bagging Classifier, AdaBoosting Classifier and Gradient Boosting Classifiers being my best 3 models. (XGCBoosting was avoided while Tuning due to computational expense)

MODEL PERFORMANCE SUMMARY- LR

	Model	Train_Accuracy	Test_Accuracy	Train_Recall	Test_Recall	Train_Precision	Test_Precision
0	Logistic Regression	0.895316	0.899967	0.535558	0.545082	0.741191	0.764368
1	Logistic Regression on Oversampled data	0.833165	0.810464	0.836779	0.799180	0.830774	0.449309
2	Logistic Regression-Regularized (Oversampled d...	0.726677	0.761106	0.658598	0.631148	0.762405	0.360656
3	Logistic Regression on Undersampled data	0.824929	0.799276	0.834426	0.805328	0.818872	0.432819

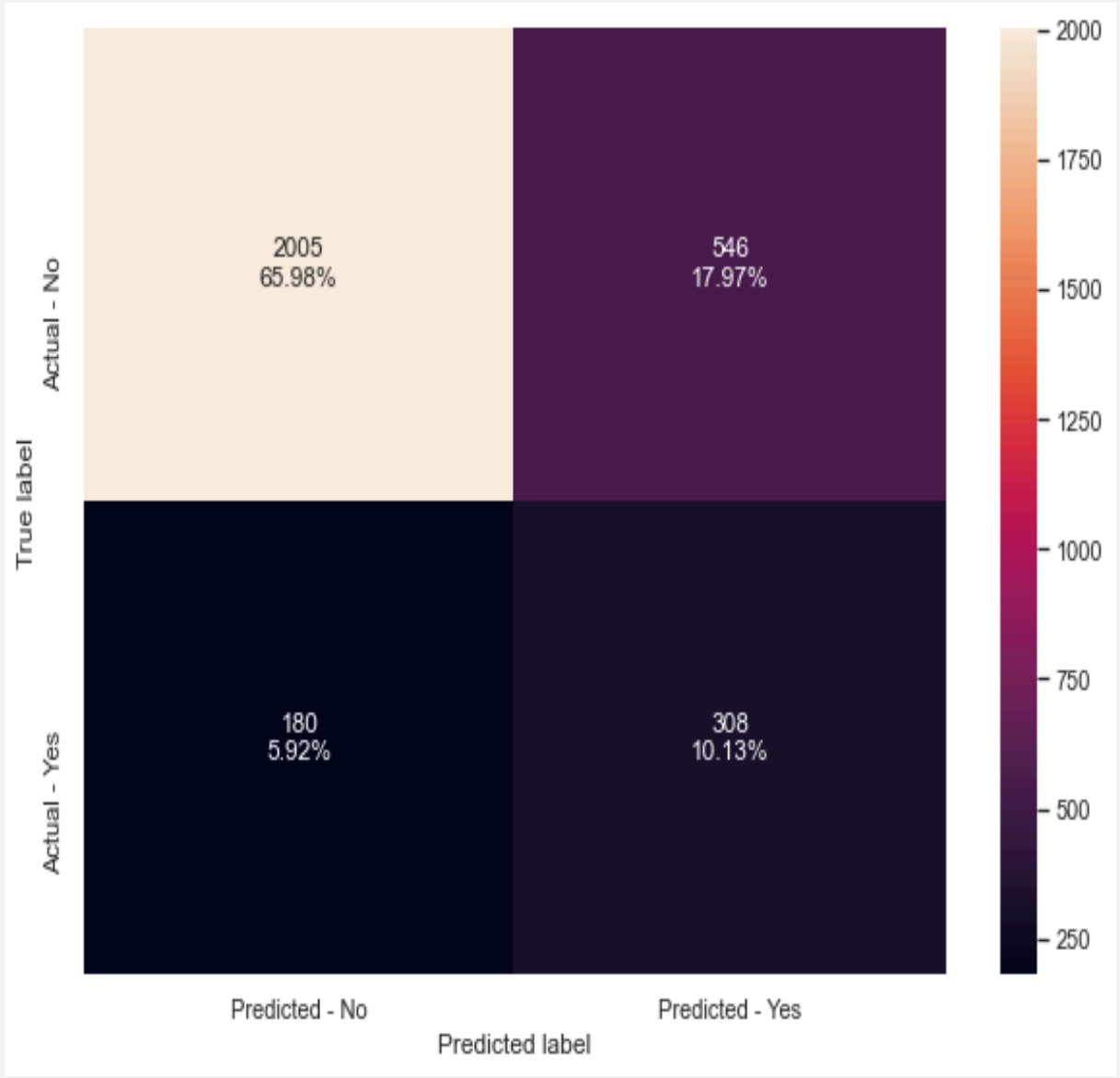
- The Logistics Regression on UnderSampled Data gave the best generalized performance per metrics with the highest Test Recall score.

MODEL PERFORMANCE SUMMARY- POST TUNING

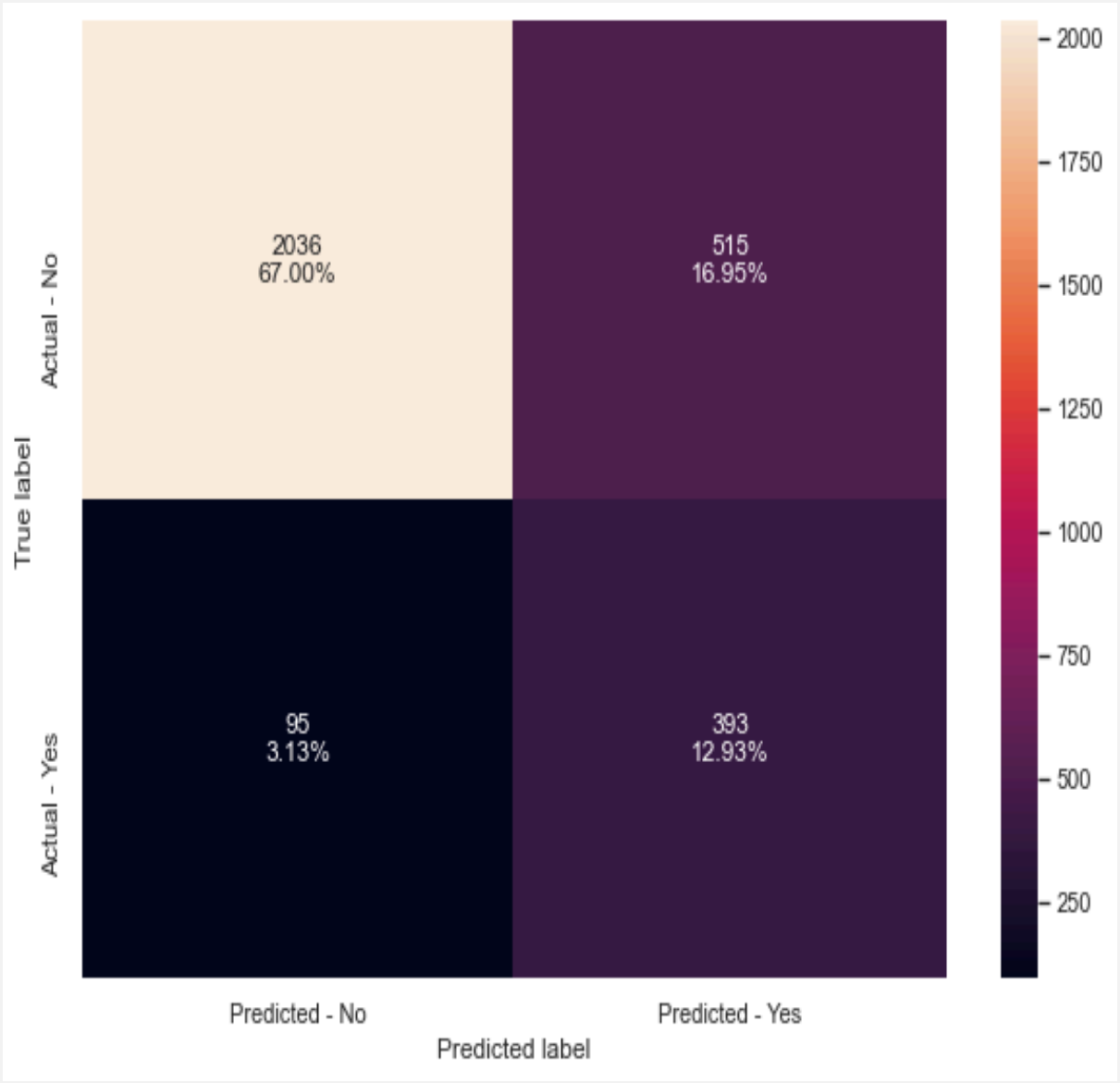
	Model	Train_Accuracy	Test_Accuracy	Train_Recall	Test_Recall	Train_Precision	Test_Precision
4	GridSearch Tuned Adaboost Classifier	0.992240	0.973676	0.971905	0.915984	0.979646	0.919753
5	RandomSearch Tuned Adaboost Classifier	0.992240	0.973676	0.971905	0.915984	0.979646	0.919753
7	GridSearch Tuned GBClassifier	0.992664	0.974663	0.971027	0.909836	0.983111	0.930818
8	RandomSearch Tuned GBClassifier	0.990124	0.976966	0.957858	0.909836	0.980234	0.944681
6	GBClassifier	0.975310	0.968411	0.885865	0.877049	0.957306	0.922414
1	GridSearch Tuned Bagging Classifier	0.993651	0.965449	0.967515	0.858607	0.992793	0.920879
2	RandomSearch Tuned Bagging Classifier	0.993651	0.965449	0.967515	0.858607	0.992793	0.920879
0	Bagging Classifier	0.996896	0.958539	0.983319	0.842213	0.997329	0.893478
3	AdaBoost Classifier	0.957675	0.953603	0.839333	0.838115	0.890960	0.868365

Based on the Comparison Table, Gradient Boosting Classifier with RandomSerach CV gave the best performance on Recall. There may be yet better parameters which may result in a better performance.

CONFUSION MATRIX FOR LR_OVER & LR_UNDER(LOGISTICS REGRESSION)



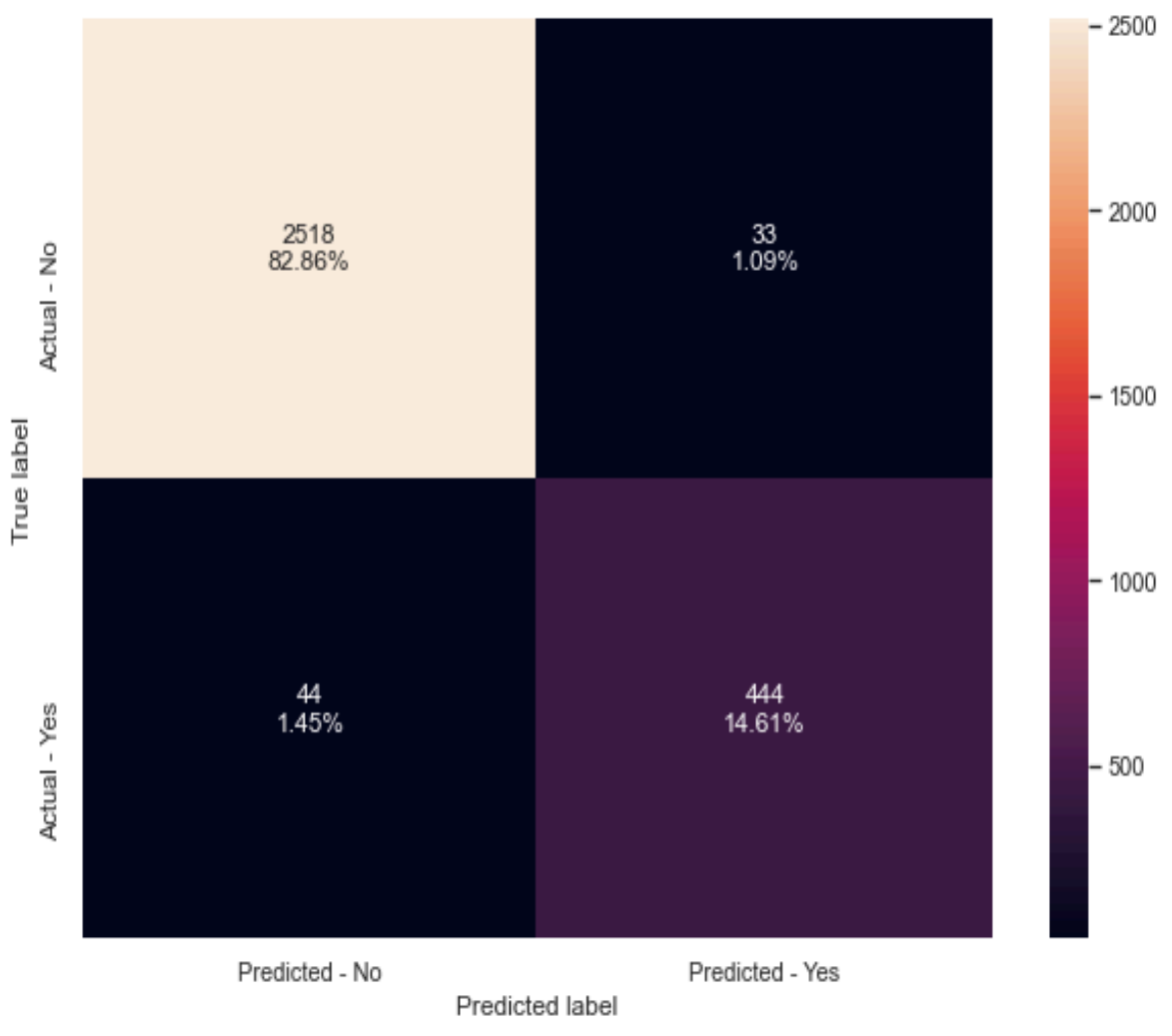
OVER SAMPLING



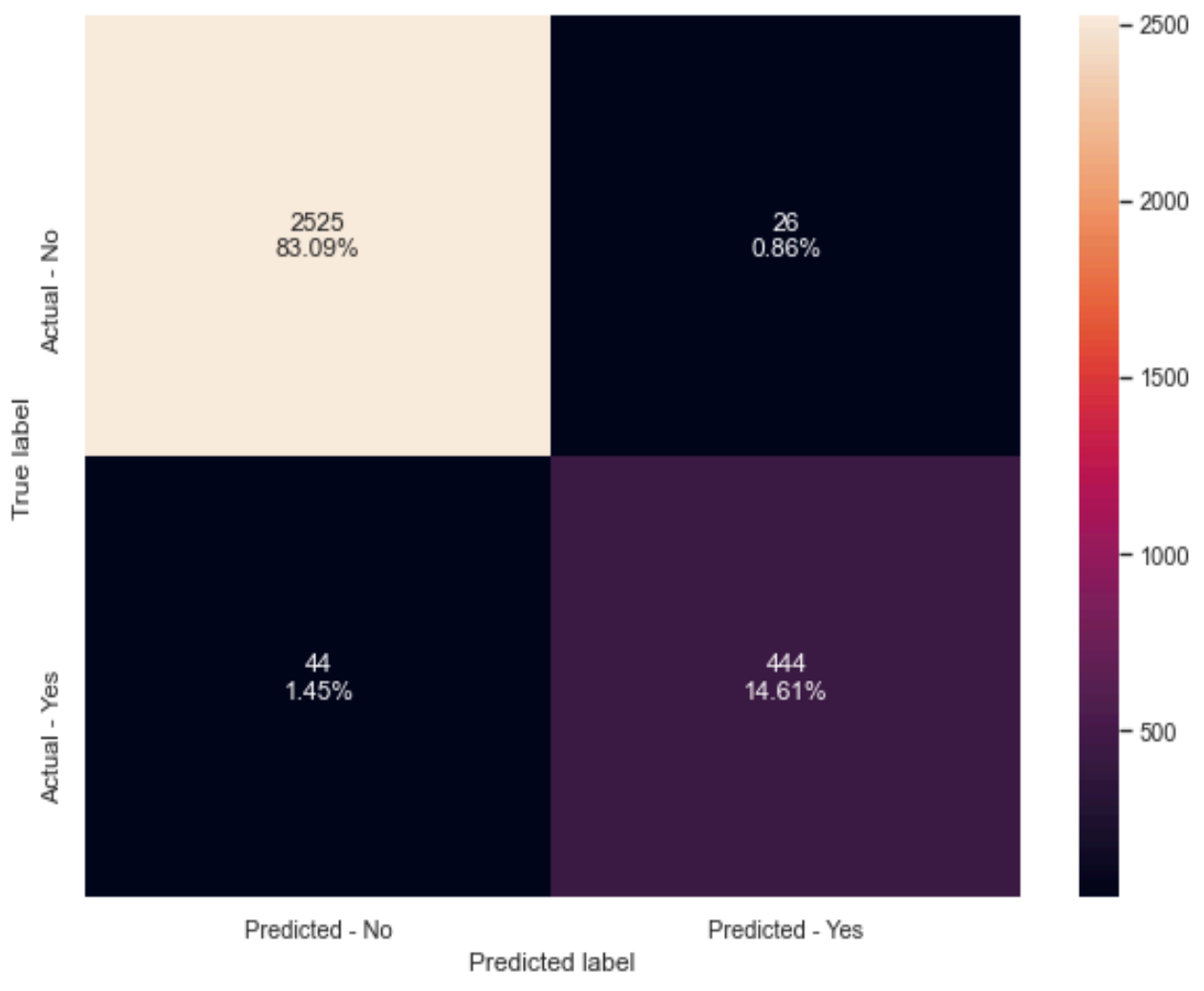
UNDER SAMPLING

Under_Sampling gave a better Test Recall

CONFUSION MATRIX GRIDSEARCHCV & RANDOMSEARCHCV OF BEST MODEL (GRADIENT BOOSTING CLASSIFIER)



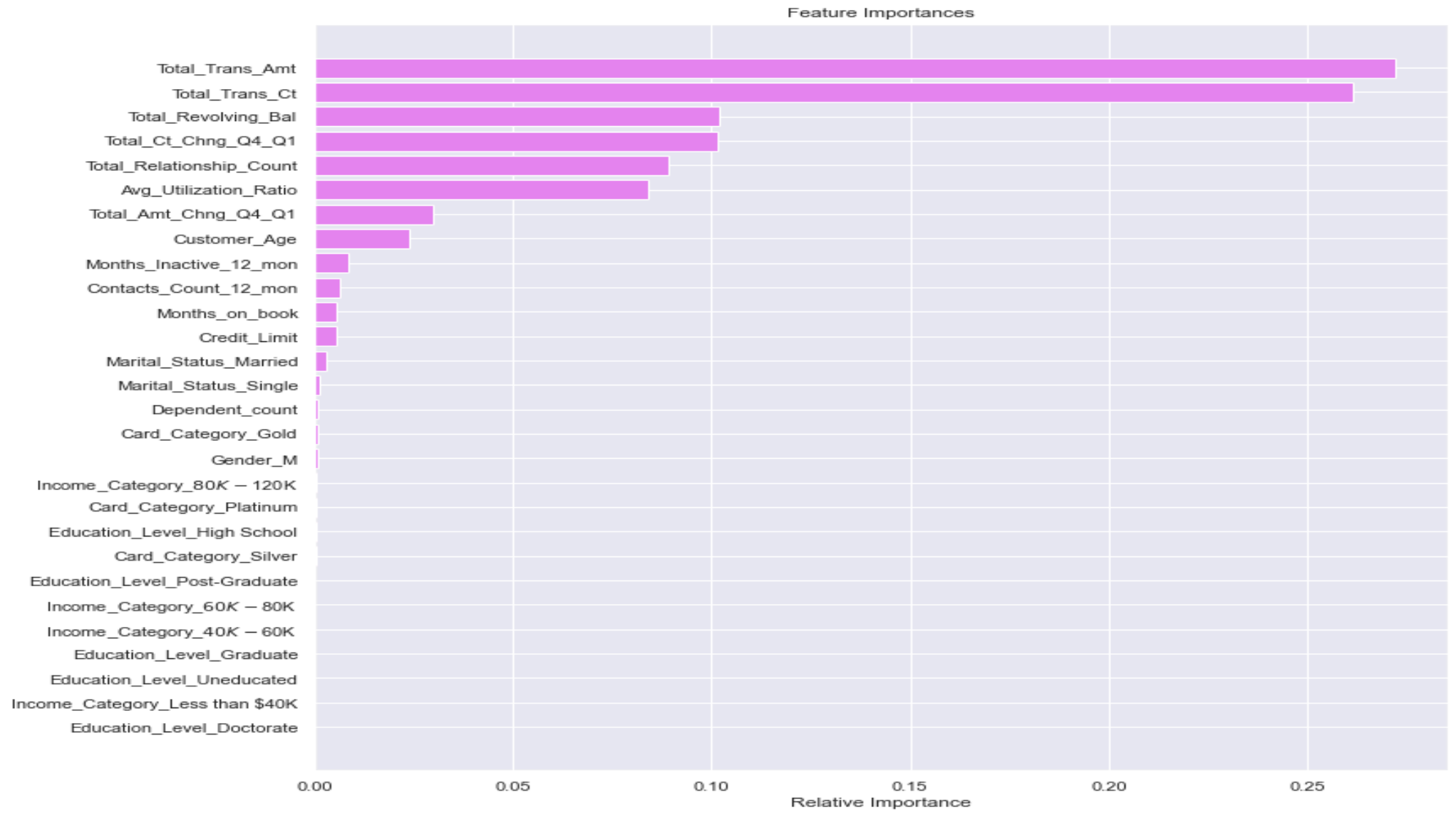
GRIDSEARCHCV (1hr 4mins 47s)



RANDOMSEARCHCV (26 mins 37s)

- GBClassifier with RandomSearchCV gave the best results though quite similar to GridSearchCV
- The Latter was 35mins or more faster than the former indicating RandomSearchCV as the better Tuning Technique

IMPORTANT FEATURES AS RETURNED BY TUNED XGBOOST CLASSIFIER



- The Most important Features that influence the Attrition of Customers as captured by the GradientBoostingClassifier RandomSearch CV are Total_Trans_Amt, Total_Trans_Ct, Total_Revolving_Bal, Total_Ct_Chg_Q4-Q1

CONCLUSIONS-KEY INSIGHTS FROM EDA

- The customer distribution has a mean age of approx.47 yrs and a median of 46 yrs
- 75% of the customers are 52 yrs or less with the oldest at 73 years
- The average dependants across customer distribution is 2
- 75% of the bank customers have 3 dependants or less across the distribution
- Thera Bank has customers who have stayed approx. 36 months on average in their books with 75% of them having stayed 40 months or less
- The Mean number products held by Thera Bank customers are approx. 4 with the highest at 6.
- 75% of the customers were inactive for 3 months or less within 12 monhs of card subscription
- The average customer contact accross the distribution is 2 in a year
- 75% of the distribution have a credit limit of just under $11,100 or less with a mean of $8,632

52.9% are Female while 47.1% are Males of which 46.3% of the customers are married with Singles following closely at 38.9%

30.9% of the Graduate customers trump the rest of the customer in other levels of Education

The mean total revolving balance is 1,163

1,784 or less is the credit limit maintained by 75% the customers sampled

75% of the distribution can access $9859 or less in their respective card subscription

35.2% of the customers earn less than 40k with 7.2% at 120k+

The average transaction amount was 4404 with 75% of the customers dealing a total transaction amount of $4741 or less

The average transaction count is approx. 65 with the highest at 139 with 75% registering 81 counts or less

The average utilization ratio is 0.27 with a media of 0.176.

83.9% are Existing customers while 16.1% Attrited.

CONCLUSIONS-KEY INSIGHTS FROM MODELING

We have been able to build a predictive model that:

a) the company can deploy to identify customers who are at the risk of attrition.
b) the company can use to find the drivers or influencers of attrition.
c) based on which company can take appropriate actions to build better retention policies.

Factors that drive attrition - Total_Trans_Amt, Total_Trans_Ct, Total_Revolving_Bal

Total_Trans_Amt: Customers with a high Total Transaction Amount are very much unlikely to attrite compared to those who return less volumes. This makes a lot of sense as this reflects the credit-card activity of the customer and access to credit as well as Avg_open_to Buy. Customers of this quality tend to have very good credit scores and records and can be tapped for other benefits.

Total_Trans_Ct: Customers who return a very high Total Transaction Count most definitely are much less likely to attrite. This speaks to the idea that the more a customer spends, the more the transaction count and the more the bank would make the required income and equally pen such customers down for a Rollover or credit limit upgrade

Total_Revolving_Bal: The higher the Total_Revolving_Bal of a customer at any given time, the higher his or her affinity to access credit or being evaluated for same. It is quite obvious that the capacity of a customer to hold a high revolving balance is a function of income largely. The higher this value is, the lesser he/she is likely to attrite. Ultimately those with lower income attrite more, which is also logical. Thera Bank can consider offering flexible repayment options and lower fees offered as incentives to endear them.

On the other hand, Singles have a higher chance of attriting than any other class under the Marital_Status category. This is closely followed by Married customers. Thera Bank marketing team can engage a product design that will capture the very youth in this bracket as they have potentials for growth. Loyalty points, coupons and vacations tied to usage could be considered to endear them to the bank. A trade off with other Bank products can aid retention as well

Credit Limit: Those customers who have higher credit limits tend to spend more and have higher transaction volumes. Those with lower Credit limits are the ones who attrite more - the company can provide some additional incentives to such employees to retain them.

.

Our data collection technique is working well, though efforts are to be made to amass more exhaustive data for a surgical and well detailed analysis. The most important features, Total_Trans_Amt, Total_Revolving_Bal and Total_Trans_Ct.shows a significant difference between Existing customer and Attrited customer. These scales can act as a preliminary step to understand the dissatisfaction of Attrited customers. - A Lower rating indicates higher chances of attrition.

□

BUSINESS RECOMMENDATIONS

- The management of Thera Bank should employ coordinated efforts, resources and additional value added offerings to aggressively track and trap all prospects with very reasonable to high income.
- Thera Bank can embark on a Product redesign for Platinum card holders as they are the toast of competition in addition to alternate product offering along with perks, incentives and other value added initiatives to drive the retention of this class
- Efforts should be made to target High income earning females less than 47 yrs for increased patronage and retention leading to more fees and profit for the bank

- These insights can be inferred to form the basis of a formidable marketing vis-à-vis advert campaigns to gain a competitive edge, more market share and ultimately increased revenue
- Resources should effectively be allocated to reign in attrited high income earning customers under 47years to add value and more business to Thera Bank
- The much needed critical mass needed to expand the credit card subscription can be trapped from the populated Blue Card category
- As evidenced in the analysis, an extensive and aggressing marketing initiative ought to be considered to tap into the significant chunk of attrited customers with potentials for subscription.
- Efforts should be galvanized to tap high income (> $60k) attrited Doctorates and Post_Graduates under 47 years as they present opportunities for referrals and huge corporate accounts.
- Attention should be given to High income earning Males and Females of income bracket of 40k-60k below 47 years of who are at the risk of attrition by thoughtful value-oriented engagements, fee review and additional product offerings that can endear their patronage while staying ahead of their peers in the Banking industry.
- Effective and competitive Pricing of different Credit Card products with additional incentives will appeal a lot more to untapped brackets within the customer base and also enable customer retention and referral that will ultimately spike an attendant traffic in patronage and ultimately revenue
- It is quite apparent the training data is not exhaustive. More effort should be geared toward leveraging on a wider observation sample to draw more inclusive, extensive and impactful insights

THANK YOU