Customer Churn Analytics: Churn Profiling, Retention analysis, and Prediction

Introduction:-

Customer churn/attrition is defined as the loss of customers in a particular period of time. For subscription-based and service-based companies, retaining a customer is much more cost effective and then holding campaigns for acquiring new customers. Higher churn does not only mean reduction in revenue but also customer dissatisfaction with the products and services.

The project is an end-to-end churn analysis. It is divided into 3 main parts:-

1. **Churn profiling** - Understanding how customer attributes (such as income, card type, tenure, engagement, etc) influence churn.
2. **Retention Analysis** – Segmenting customers into high-risk and low-risk groups, quantifying their churn rates, and highlighting the disproportionate risk distribution.
3. **Churn Prediction** – Predictive modelling to identify at-risk customers in advance.

Data Preparation:-

For making the numerical data much more interpretable for churn profiling, they must be grouped. I have transformed the numerical features into categorical bins.

1. **Uniform binning**:- Features where equally sized bins made more sense (Eg:- Age, Credit limit, etc)
2. **Business logic binning**:- Used domain understanding for setting ranges.

Baseline Churn Rate Computation:-

- Overall churn rate
- Tells us the average probability of a customer to churn
- Baseline churn rate = 16.065%

Churn Profiling:-

Univariate Profiling

i) Demographics:-
1. gender
2. customer age
3. dependents
4. marital status
5. income category
6. education level

- While the distribution of males and females is almost similar, females show a churn rate of 17.32% which is slightly greater than baseline, indicating higher risk of churn among females but little overall impact. Males show lower risk than average (14.61%) suggesting lesser risk and relatively stable gender.
  Recommendations:
  Focus on the retention offers and campaigns particularly for females.

- Most of the customers have their age lying between 37 to 50 years, and this bin shows churn 16.68% which is slightly greater than average. The age group of 26 to 38 years show a churn of 13% which indicates lesser churning risk compared to average, and stable group.
  Recommendations:
  Middle aged customers show higher churn and hence can be targeted with age specific campaigns (for eg: family oriented offers, retirement benefits, etc.)

- customers with at least one dependent show a slightly higher churn rate than average. Customers with no dependents are less likely to churn (14.9%)
  Recommendations:
  The customers with no dependents could just be monitored and less effort applied here. For customers with dependents, the focus could be more on family based offers for promoting retention.

- churn rate is not affected by the marital status, as it shows slight fluctuations (+-1%), indicating it's not relevant.
  Recommendations:
  As marital status alone is not a key driver, it may not be considered while holding campaigns.

- Customers with low income (<40k) and high incomes (>80k) are having churn rates slightly more than average, indicating they are at slightly more risk of churning. Low income customers may have faced unmanageable credit debts and high income customers may not be dependent on credit cards much may be due to in hand cash or they prefer debit cards. Customers with mediocre income (40k-80k) show lower churn rate (14.4%) suggesting that credit debts are manageable for them while they might not have much in-hand cash.
  Recommendations:
  For low income customers, if credit debts are a problem, educational campaigns can be held where they can be taught about debt management plans. For high income customers, premium cards like gold and platinum could be promoted so that they get luxury perks.

- Customers with high education levels show the highest churn rate (19%) which could indicate them having financial literacy and using credit cards within manageable limits. But, there are very few customers in this segment hence overall impact is less. On the other hand, customers with medium and low education levels exhibit a churn rate of 15.4% and 15.9% respectively, slightly lower than average.
  Recommendations:
  For highly educated individuals, premium cards can be marketed revealing advanced premium benefits and lifestyle perks.

ii) account features:-

1. card_category
2. months_on_book
3. total_relationship_count

- There are just 20 customers who hold the platinum card. The churn rate here is significantly high (25%) which indicates higher churn risk than average. But, due to minor contributions to the card category, the impact of such a high churn rate won't be significant.
  Recommendations:
  Gold cards can be promoted as they have lower annual fees (higher fees of platinum might be the churn driver) and almost equitable lifestyle perks are provided.

- For the tenure > 48 months, the churn rate is slightly above average (17.35), suggesting that it's nominal and not impacting overall churn rate. The tenure of

13-24 months has seen relatively loyal customers as indicated by their churn rate (14.8) moderately below the average.

Recommendations:

For high tenure customers (>4 years), retention campaigns like 5-year onwards membership benefits can be held.

- Customers with just 1 product have a churn rate of 25.6% which is significantly higher than average. As they just have one product with the bank, it shows lesser engagement and it indicates that they are more likely to leave. Customers with 2-3 products also have a high churn rate (21%) which shows that they are less engaged with the bank and likely to leave. Customers who have 4+ products show stickiness with the bank (11.4% churn rate) marking themselves as highly loyal customers. As most of the customers have 4+ products, there will be a positive impact on the overall churn rate.

  Recommendations:

  For customers holding lesser products, campaigns describing the benefits of all other products shall be held in order to promote stickiness with the bank. For customers holding 4+ products, attractive offers shall also be introduced which can attract customers who hold less products.

iii) customer engagement:-

1. months_inactive_12_mon
2. contacts_count_12_mon

- There are just 29 customers who show no inactivity at all. their churn rate is pretty high (51%) which indicates that half of the customers from this segment tend to churn. But as they are just a few in number, the overall impact to churn isn't considerable. It may imply that those customers are new-comers who stayed close to a year. Those who are inactive for 4+ months show significantly high churn rate (24.55%). These customers might be disengaged. Customers who are inactive just for 1-3 months show 15.2% churn rate, slightly lower than the average churn rate. The majority of customers lie in this segment, hence it could suggest that the majority of customers show moderate to active engagement with banking activities.

  Recommendations:

  Customers shall be encouraged to transact more using credit cards in order to unlock more benefits. The benefits should be linked to the number of transactions for a particular time frame (say a month from expiry). The more the transactions, the more the benefits unlocked.

- Those customers who contacted the bank 4+ times have a churn rate of 26.3% indicating that they may be dissatisfied with the services and highly likely to close the account. customers who contacted just 1-3 times show a churn of 14.7% indicating that they are moderately satisfied than an average customer and hence less likely to close. There are a very few customers who never contacted the bank and show 1.7% churn which could suggest that they are satisfied by the products and services and least likely to churn.
  Recommendations:
  Employees of the bank who are on the consulting side shall be given incentives if they are able to address the customer's problems completely and efficiently. This way, the problems which root dissatisfaction among customers can be resolved much faster, promoting retention.

iv) credit and balance features:-

1. credit_limit
2. total_revolving_bal
3. avg_open_to_buy

- customers with credit limit between 17977 and 26246 have a churn rate of 12%, indicating that customers from this segment are considerably loyal. For rest segments, the churn rate is nearer to the average churn rate, hence aren't affecting the overall churn rate much.
  Recommendations:
  Credit limit is not a key driver for churn rate. However, customers of each segment need to be monitored and campaigns to be held whenever churning risk increases.

- Customers with the lowest revolving balance show the highest churn rate (37.9%) which suggests that they aren't much dependent on the credit card and are more likely to cancel it. Customers with higher revolving balances have a churn rate between 4-13% which is significantly low, suggesting that these customers are heavily dependent on credit cards and hence less likely to close it.
  Recommendations:
  Customers should be encouraged to transact more frequently again with the transaction based reward system. It could be a point system where the more you transact using the card, the more points you gather, which could later be redeemed to unlock exciting offers.

- The range of 17.2k-25.8k has the lowest churn rate (12.5%) suggesting that these customers are actively using credit cards but still maintaining a healthy balance. Customers with higher open-to-buy (25k+) have the highest churn (16.7%) suggesting that they don't use credit cards much and hence are more likely to close it. customers with lowest open-to-buy (<8.6k) suggest that they use credit cards heavily but don't maintain a healthy balance, which suggests that they aren't loyal.

Recommendations:

For customers who aren't maintaining a healthy open to buy average, can be informed with the debt management plans so that they can manage the credit debts. For customers not using the cards sufficiently, they should either be encouraged to buy a low-cost card (blue or silver with not much credit limit) or should be encouraged to transact more with the reward system.

v) transaction behaviour features:-

1. total_amt_chng_q4_q1
2. total_trans_amt
3. total_trans_ct
4. total_ct_chng_q4_q1

- Customers which have ratio < 0.7 show highest churn rate (19.7%) indicating that they have reduced their usage of credit cards hence more likely to close it. The ratio between 1-1.3 also shows a high churning rate (18.2%) but that is not significantly higher than average churn rate. Also there are fewer customers in this segment hence less impact to the overall churn rate. Customers with ratio 0.7-1 show low churn rate (13.4%) which suggests that despite reduced dependency on credit cards, customers still are loyal. It might be that they spend more during discounts/offers leading to low transaction amounts. customers with ratio > 1.3 show lowest churn rate (1.9%) indicating that they have actively used a credit card and are significantly dependent on it and hence not likely to close it. But, there are very few customers in this segment, hence overall impact on churn rate will be less.

  Recommendations:

  Customers with ratio < 0.7, should be encouraged to spend more with credit cards through targeted promotions, rewards, and offers. For customers with a ratio between 0.7-1.0, they should be given more discounts if they use credit cards frequently in order to maintain the usage.

- Customers with moderate cumulative transactions (5k-9.5k) show the highest churn rate (21.1%) indicating that their spend is moderate and therefore more

likely to churn. customers with higher spendings (>9.4k) having a churn rate <9% indicating that they use credit cards actively and hence are less likely to close it.
Recommendations:
Customers with moderate spendings should be encouraged to use credit cards more via targeted campaigns (offers, rewards). As the risk for customers spending higher amounts with credit cards is minimal, retention efforts can be minimal for this segment.

- Customers with the least number of transactions (<42) show the highest churn rate (34%) indicating lesser dependency on credit cards. customers with a higher number of transactions (>74) show a churn rate < 2%, which suggests that they are significantly dependent on credit cards and less likely to close it than the average customer group.
  Recommendations:
  Here again, the usage of credit cards is showing somewhat proportionality with the churn. As the usage decreases, customers are more likely to disengage and hence those customers should be encouraged to use credit cards through offers, discounts, and rewards like limited time cashbacks bound with credit card usage.

- Customers with ratio < 0.7 have a significant churn rate of 24.6%. Moreover, there are many customers belonging to this segment, making considerable impact to the overall churn rate. This segment suggests that customers have reduced usage of credit cards and hence are more likely to close it. Customers who have ratio 1-1.3 have a churn rate of 8% but not the lowest even though spending increased which could imply that customers who have slightly increased the usage may not be loyal. For the rest segments, the churn rates are below 8% marking them as loyal and they altogether contribute a higher percentage, hence greater positive impact to the churn rate.
  Recommendations:
  For customers with ratio below 0.7, targeted campaigns which include offers, and rewards like limited time cashbacks linked with credit card usage shall be done. The customers with ratio 1-1.3 shall have their usage monitored as the increase might be for a shorter time frame. The retention efforts for this segment may still be minimal as the churn risk is still much lower.

vi) card features:-
1. avg_utilization_ratio

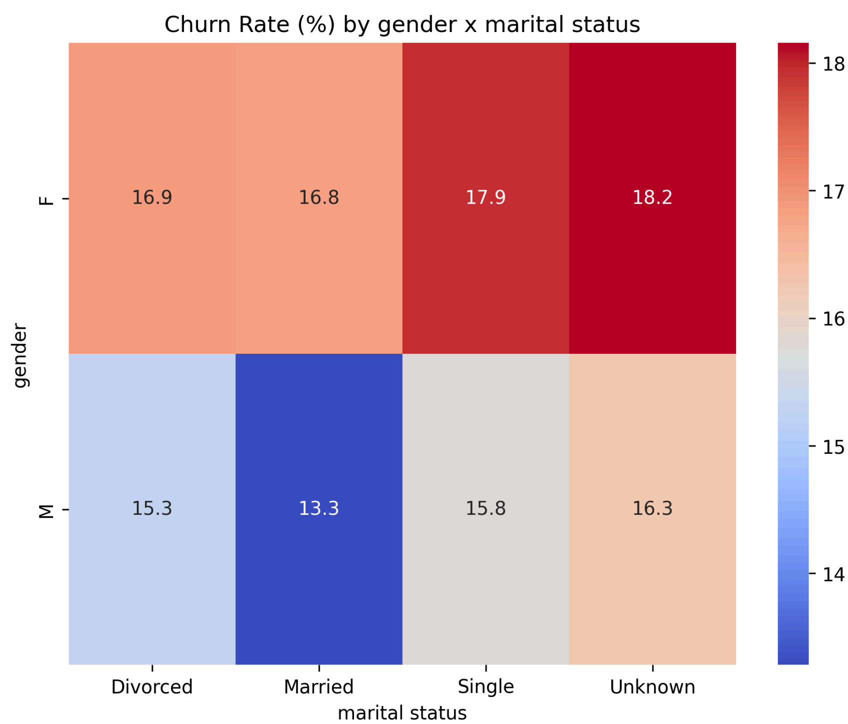
- Customers who have a ratio between 0-19% have the highest churn rate of 22.23%. These customers are also high in number hence may significantly impact the overall churn rate. Such customers have only utilized their credit cards below 20%, which means they are either dissatisfied or aren't much dependent on it hence more likely to close it. Customers who have overly utilized (60+) and utilized in a healthy way (20-59%) have churn below 10% indicating that they use the cards heavily. They might be satisfied with the services too.

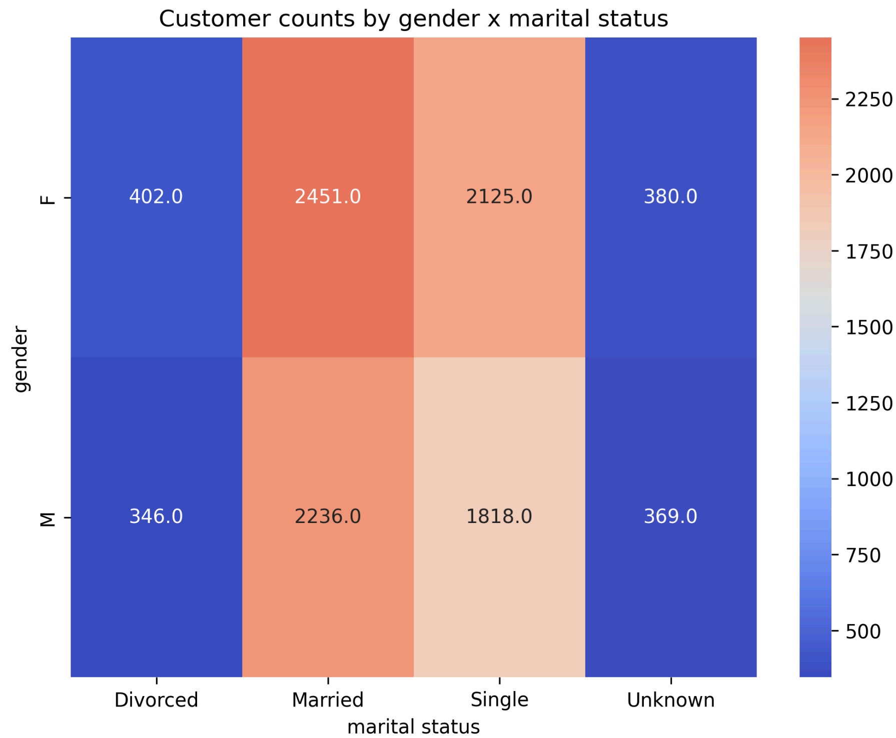
Recommendations:

Customers who have utilized below 20% shall frequently be reminded about the limited time offers and rewards linked with credit card usage. Customers who have utilized their cards more than 20% are not posing the risk to churn hence efforts for this segment shall remain minimal until required.

Bivariate Profiling

i) Gender x Marital status:-



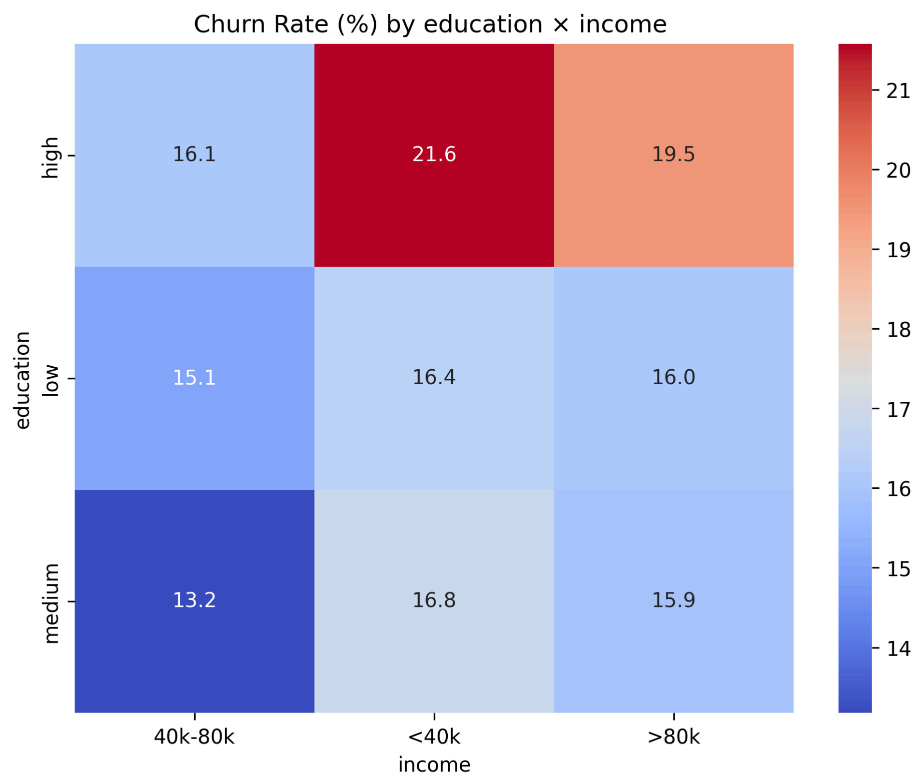


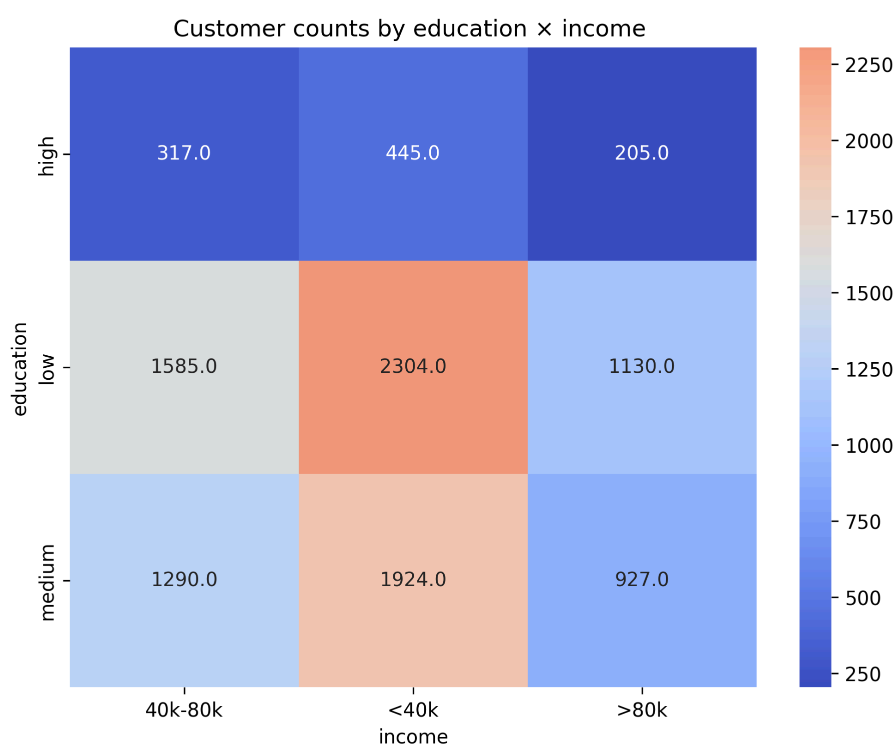
- Female customers, irrespective of marital status, show higher churn than average. Women with unknown status show the highest churn (18.2%) but they are less in number hence less impact. Women who are single show moderate churn (17.9%).
- males have shown a mild churn rate (16.3%) with unknown status. For other statuses, the churn rates are lower than average, with married men showing lowest churn (13.3%) indicating that married men will show most loyalty. with married men being the most among all statuses, the impact to overall churn rate might be significantly positive.

  Recommendations:

  There shall be targeted campaigns personalized for female customers like offers and discounts linked with fashion, wellness, travel, etc and rewards linked with credit card usage as females show higher risk comparatively.

ii) Education x Income:-





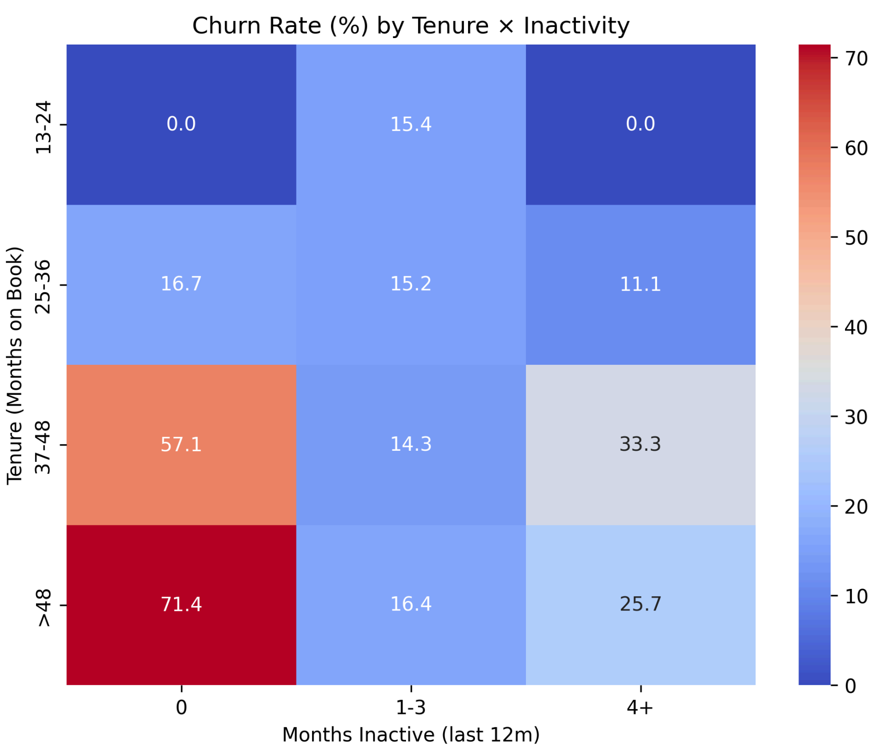
- customers who received higher education are most likely to churn (21.6%) when they have low incomes (<40k). it could be because they may be financially literate and know their capabilities of paying back and hence not in high credit debt. With them having income > 80k, they still show high churn (19.5%) maybe because they might show dissatisfaction with minor inconveniences which they may spot due to them being highly literate. But, both the income segments combined, still have lesser customers and hence less impact to the overall churn rate.

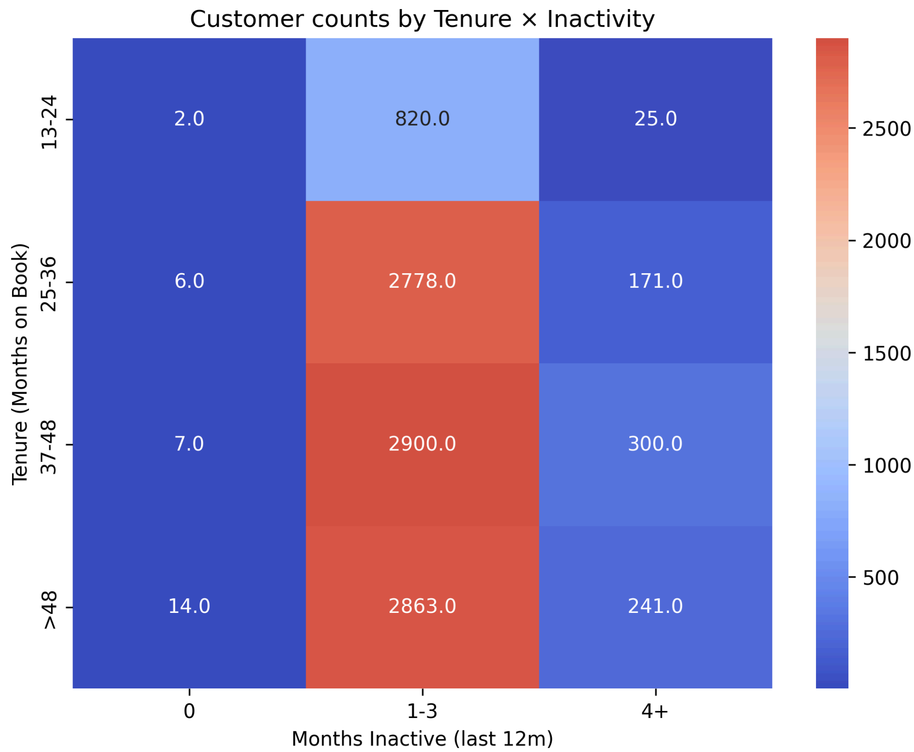
- Irrespective of incomes, the customers with low education have a churn rate nearer to the average churn rate, suggesting that these customers more or less share the same characteristics with an average customer.
- Customers with medium level education show low churn (13.2%) at mediocre incomes indicating they may have moderate financial literacy and may have healthy credit card usage and therefore less likely to close it.

Recommendations:

Customers who are highly educated but have lower incomes should be targeted with affordable value driven perks linked with credit card usage. Customers who are highly educated and have high incomes should be encouraged to use gold or platinum cards and shall be informed everything about its premium benefits and lifestyle/luxury perks.

iii) tenure x inactivity:-



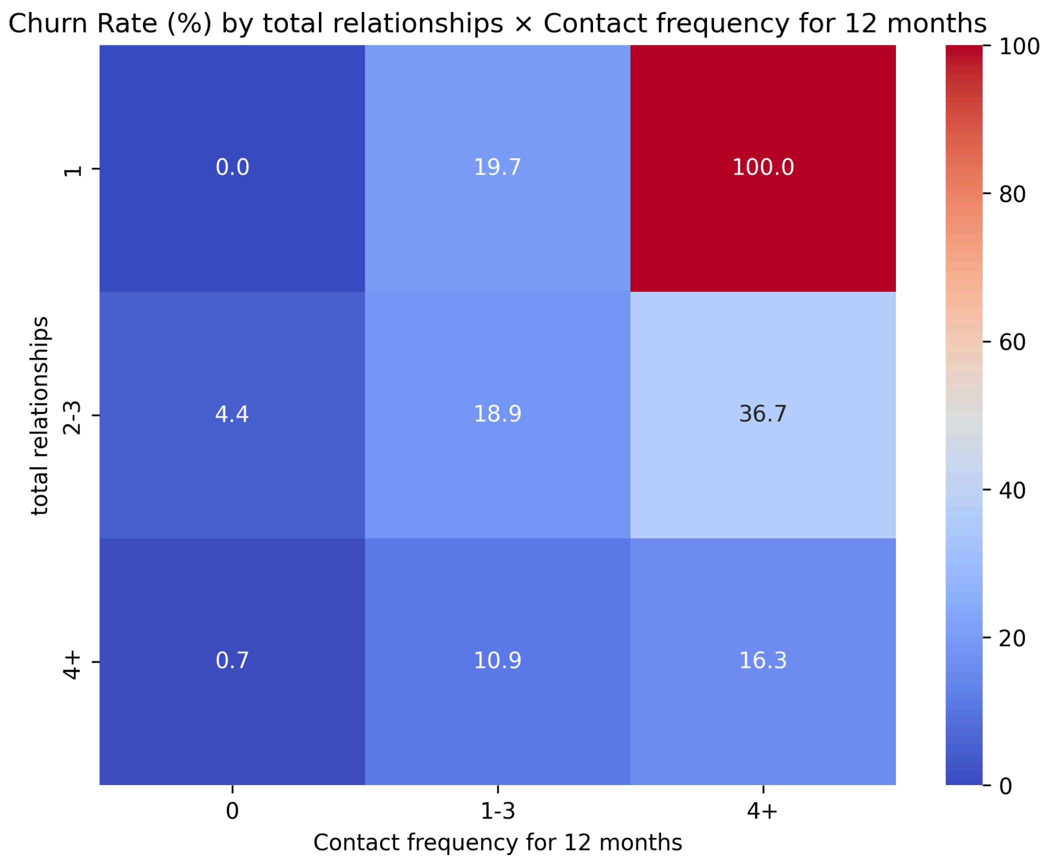


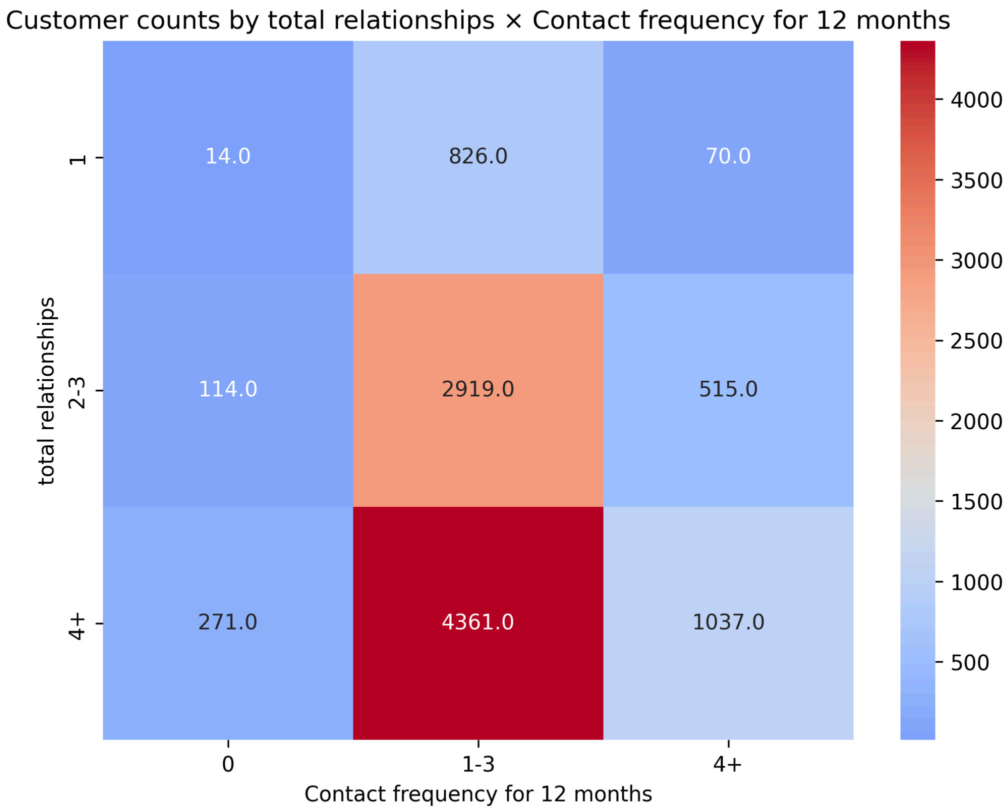
- customers with a small tenure (13-24 months) show 0% churn rate for 0 months and 4+ months. This is not a general business trend as the sample size is negligible compared to other segments. Hence it is creating sample bias.
- For the rest of the segments too, irrespective of the tenure, churn rates are higher than average for 0 months of inactivity again due to sampling bias.
- With tenure >37 months, the churn rate is greater than 25% for 4+ months of inactivity suggesting that customers in these segments are highly inactive and might not be engaged enough with the banking activities. These customers are more likely to churn than an average customer. For 1-3 months of inactivity, the churn rates are lower than average for customers less than 4 years of tenure. Above 4 years, customers showing the same inactivity show slightly higher churn rate suggesting that most loyal customers are more likely to churn if their inactivity is between 1-3 months and very likely to churn (25.7%) if the inactivity goes beyond 4 months. However, as customers with 48+ months of tenure and 4+ months of inactivity are in small numbers and hence less overall impact.

  Recommendations:

  As the loyal customers (who stayed with the bank for more than 3 years) are more likely to churn if they are inactive for 4+ months, this segment should give timely reminders bucketed with limited time offers and returning rewards. They shall also be given long-term membership offers, encouraging them to stay longer. New customers (tenure of 1 year) shall be given more offers and habit-forming incentives like cashbacks or milestone rewards which could encourage them to stay.

iv) total relationships x contact frequency (last year):-



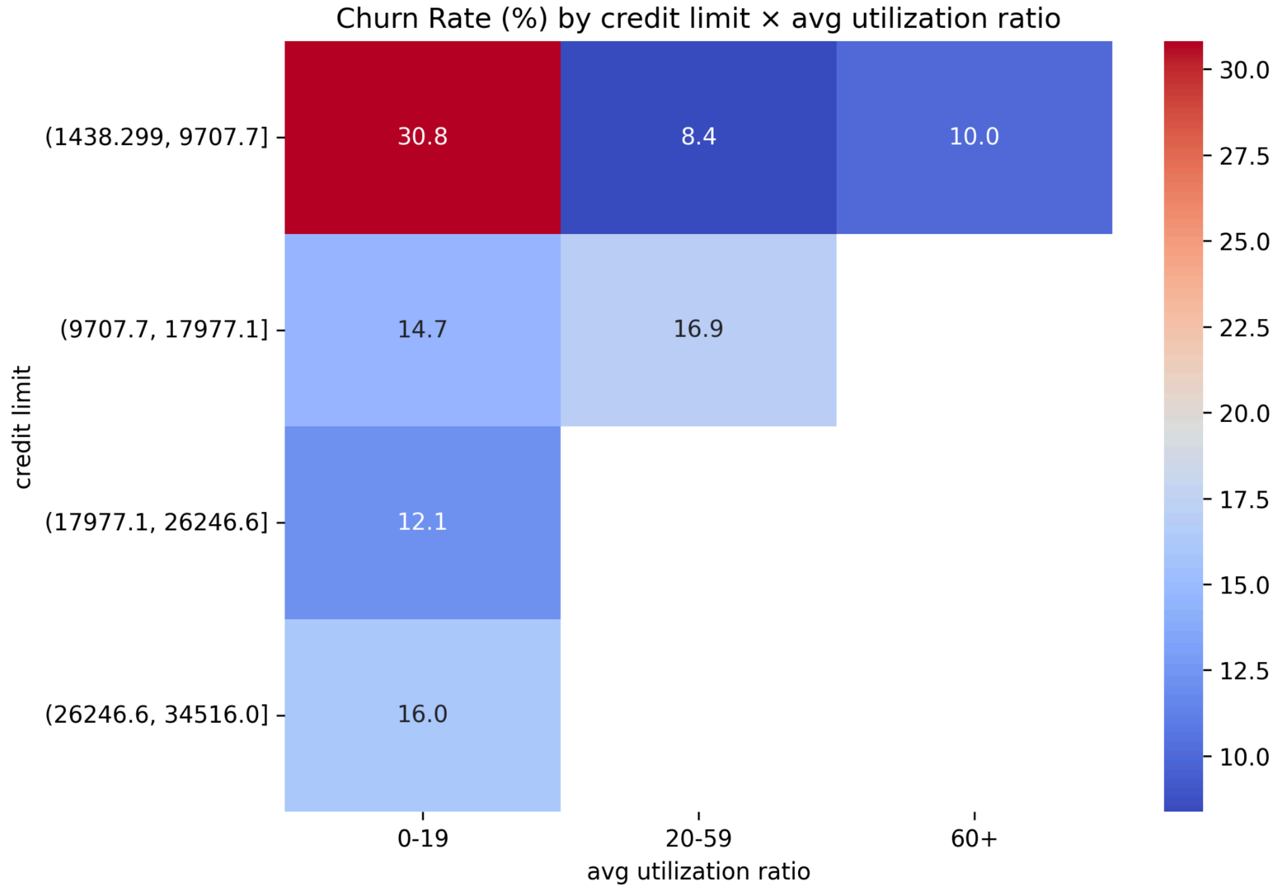


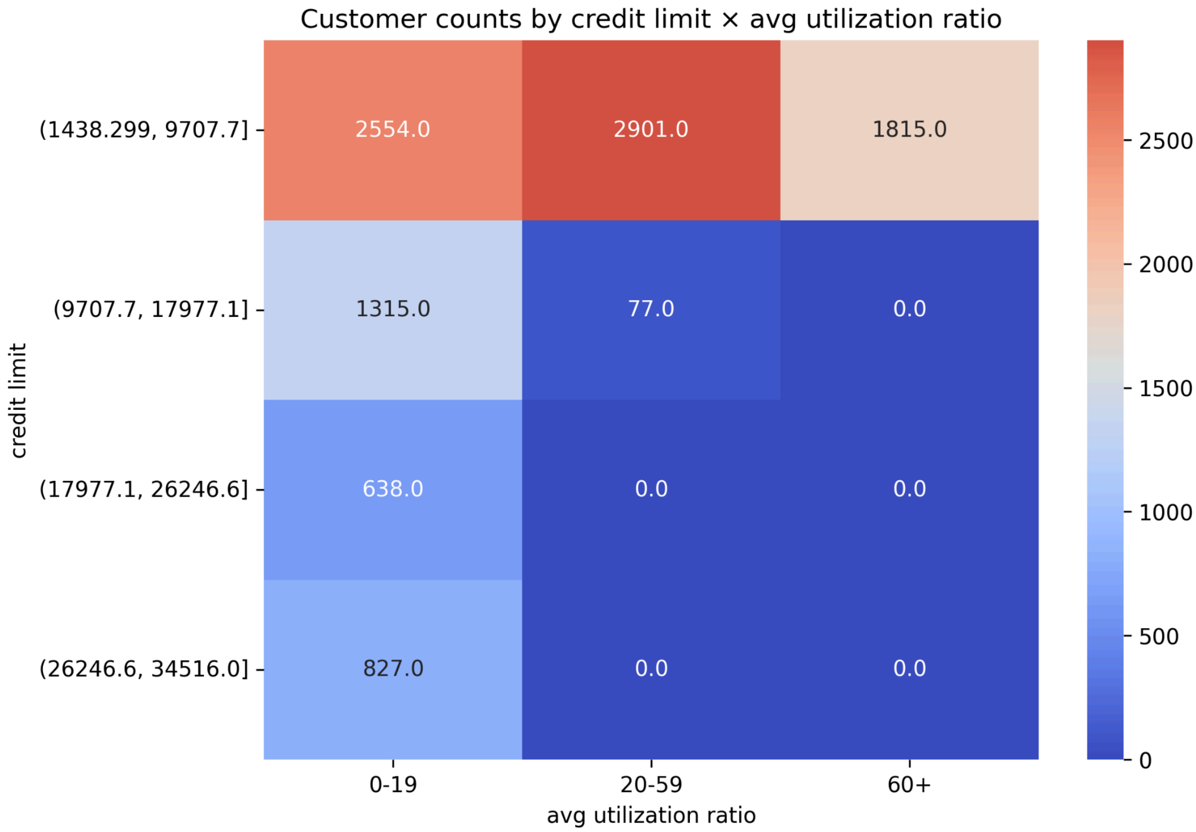
- irrespective of products they hold, higher the contacts to the bank, higher the churn rate seen suggesting dissatisfaction from the products and services might be a root cause.
- customers who have just 1 product with the bank and contacted 4+ times show 100% indicating that they aren't sticky enough as they have just 1 product and moreover they contacted 4+ times showing high dissatisfaction with just 1 product, making the customer leave. However, the sample size is just 70, which indicates this is not a general pattern and its impact on overall churn rate is still less due to size.
- Customers with 0 contacts irrespective of products they hold have a churn rate <5%, which suggests that they are highly satisfied with products they have or they might be disengaged.
- customers showing moderately high satisfaction (1-3 times contacted), show decrease in churn rate with increase in products they have, suggesting that even with moderate satisfaction, they tend to leave less if they have more products with the bank. This relationship is the same for highly dissatisfied customers (4+ contacts) but with much higher churn rate when they have below 4 products.

  Recommendations:

  Customers who have contacted the bank 4+ times and have less than 4 products, are most likely to leave and hence their complaints shall be looked after on priority, as the more quickly the problems the customers are facing get resolved, the less they have to contact and the more satisfied will they be with the services and products. This segment shall also be encouraged to buy more products and marketing campaigns for all products must be held so that customers become aware about each product's benefits. Customers who have never contacted the bank and customers who have 4+ products pose very less churn risk. For these 2 segments, retention efforts shall be minimal.

v) credit limit x avg utilization ratio:-



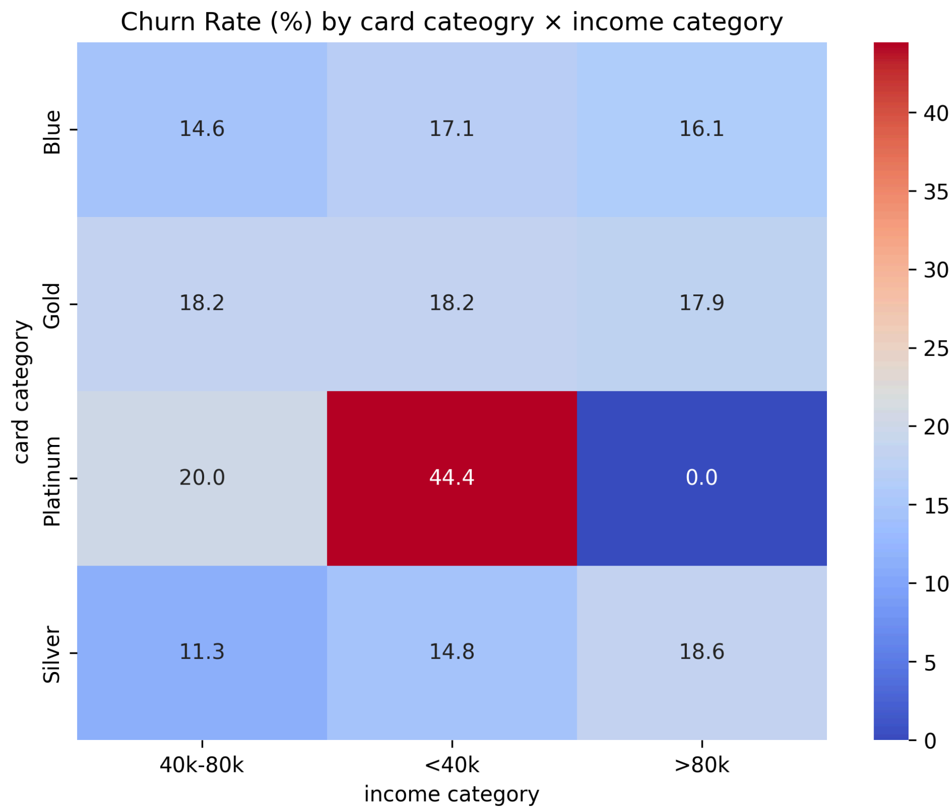


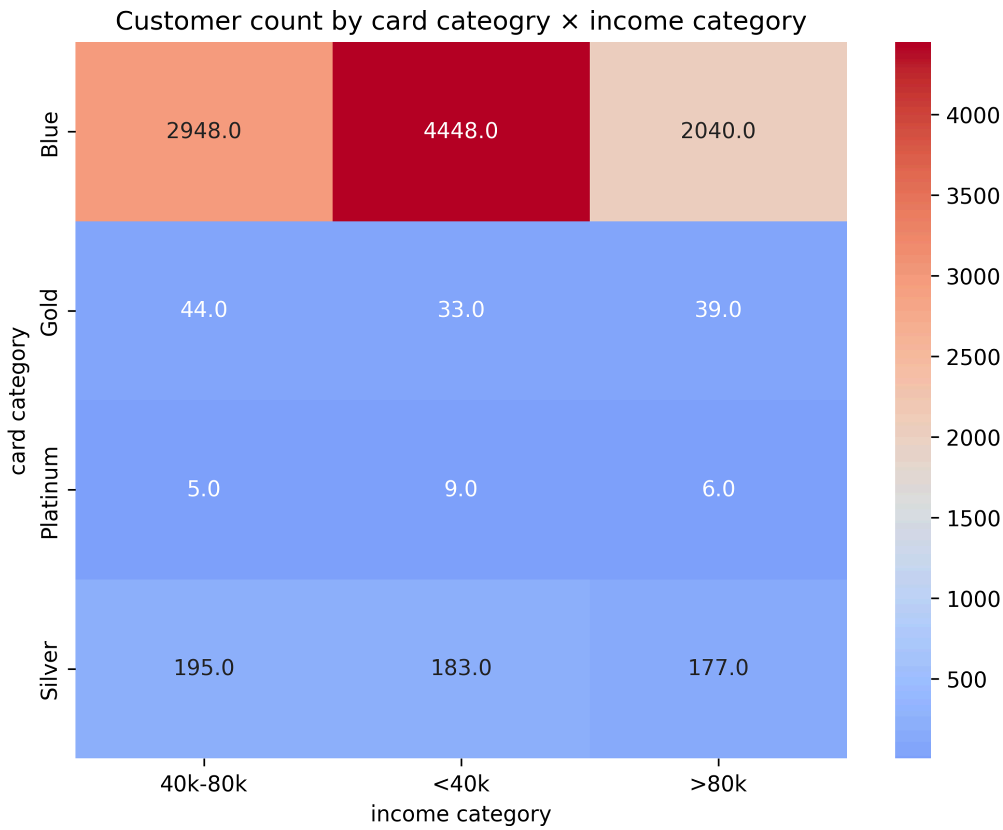
- Customers with lowest credit limits (1.4k - 9.7k) show the highest churn rate (30.8%) if only 20% of it is used. This indicates that such customers may want to execute frequent and heavy transactions but aren't able to due to low credit limits and hence not properly utilized. Such customers are highly likely to close the credit card. But, with utilization > 20%, the churn rates drop really low (<10%), suggesting that most of the customers may execute small and fewer transactions and are okay with lesser credit limits.
- Only the customers with cards with credit limit of 1.4k - 9.7k and 9.7k - 17.9k are utilizing it till 60%+ and below 60% respectively. For rest credit limits, utilization is below 20% only.
- A limit of 17.9k - 26.2k is seen as a high as well as optimum credit limit where churn rate at below 20% utilization is 12.1%. This suggests that customers may not be frequently transacting or transacting huge amounts but still keep higher credit limits because they might need more money anytime.
- With below 20% utilization only, credit limits above 17.9k are generally not used efficiently and may be suited only for loyal customers who have frequent transactions.

  Recommendations:

Only the customers with lowest credit limits and utilization below 20% show high churn risk. If customers in this segment have a history of executing high amount transactions in their past which are close to their credit limits, it could suggest that they are wanting to transact more than the limit they are getting offered with. These customers shall be encouraged to buy silver, gold, or platinum cards based on their lifestyle and income, as these cards have credit limit extension along with other lifestyle perks. For the rest segments, risk is much less regardless of credit limits or utilization of cards, hence retention efforts for these segments may remain minimal until required.

vi) card category x income category:-





- Most of the customers irrespective of incomes prefer blue cards as they have basic/low annual fee. This suggests that most of the customers are okay with basic non-luxury benefits with low credit limits, indicating that they transact less often and they transact lesser amounts. Customers with low incomes (<40k) churn slightly more (17.1%) and customers with moderate incomes (40k-80k) churn less (14.6%).
- A few customers of all income categories prefer silver cards indicating that they might need an extended credit limit and might be satisfied with some more niche benefits.
- There are very few customers holding gold cards in all 3 income categories. Again, churn rate is moderately high among low income and moderate income customers (18.2%) suggesting that fee might outweigh the benefits as customers closing the cards might be moderate spenders.
- A negligible amount of customers have platinum cards where churn rates in low income and moderate income categories are high (20% and 44.4% respectively). Such a high churn rate of 44.4% for low income customers suggests that they may be unable to pay the high annual fees and extract all benefits it gives. It also indicates the low income tier customers may not be luxury spenders hence annual fee not justified for them. People with high income show 0% churn for

platinum suggesting that they are luxury spenders who are able to afford high annual fees. However, even the total customers holding platinum cards is quite less, causing sampling bias and not enough to draw out business patterns.
Recommendations:
Churn rates are higher for low income and moderate income customers who are platinum card holders, which suggests that the annual fees outweigh the benefits. Customers with moderate incomes should be encouraged to try silver or gold cards while the customers with low incomes should be encouraged to try out blue cards. This way the annual fees might easily come under their budget while enjoying the card perks. Among silver card holders, the churn risk is highest for customers with high incomes. They should be encouraged to try out gold or platinum cards where they can get premium benefits and luxury perks for lifestyle, travel, etc. As most of the customers are blue card holders with low churn rates, they should get loyalty bonuses.

Retention Analysis:-

Risk Segmentation

Customers are classified into ‘High Risk’ or ‘Low Risk’ depending on the results derived from Bivariate analysis which show churn rates of combinations of interdomain and intradomain features.

- **High Risk:** Customers having segment churn rate >= 20%
- **Low Risk:** Customers having segment churn rate < 20%
- **Unclassified:** Customers who did not satisfy any conditions

Segment Churn Rate

The average churn rate is calculated for each category:-

Risk Category	Churn Rate (%)
High Risk	27.82
Low Risk	9.19
Unclassified	9.09

From this table above, we can clearly see that the High risk segment average has almost 3 times higher churn risk than the Low risk segment. The Unclassified segment has risk very close to the Low risk.

Note:- As the size of the 'Unclassified' segment is also negligible, and with its churn rate being really close to the Low Risk segment, it is combined with the Low risk segment.

Distribution Check (Overall vs Churned Customers)

Overall Distribution:-
% of customers in each risk segment regardless of attrition.
Churned Distribution:-
% of customers in each risk segment only considering customers who churned.

Risk Segment	Overall Distribution (%)	Churned Distribution (%)
Low Risk	63.09	36.08
High Risk	36.91	63.92

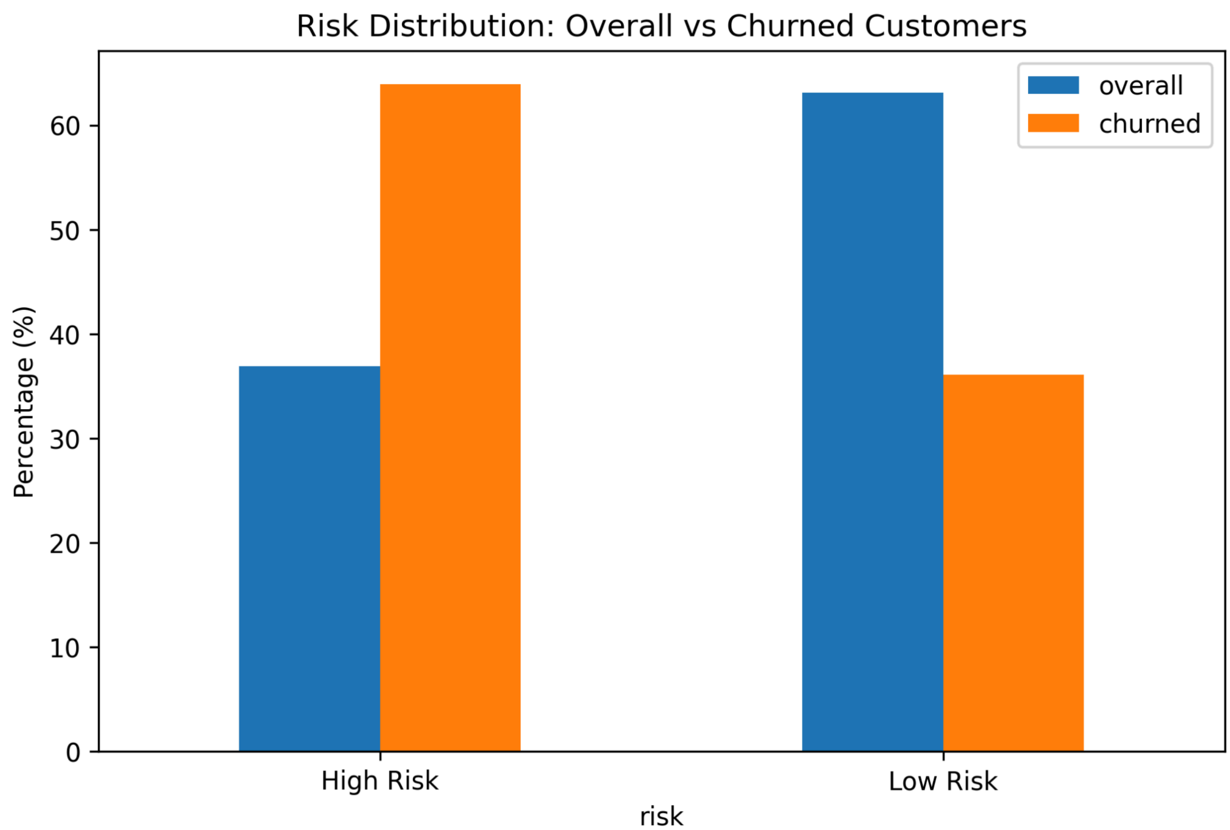
Insights:-

- When we look at the overall distribution, we can see that almost 2/3rd of the customers are low risk customers and the rest 1/3rd are high risk customers.
- We may assume that even if we considered the churned-only customers distribution, the contribution of low risk would be higher than high risk.
- But in reality, 2/3rd of the churned customers are high risk and rest 1/3rd are low risk.

Reasoning:-

- High risk customers carry characteristics that make them more prone to churn. This makes their churn probability significantly higher.

Hence, even though low risk customers may churn occasionally, due to lower churning probability, their share in churned customers is relatively small.



Churn Prediction:-

Feature Engineering

In order to prepare the data for predictive modelling, the categorical features must be encoded as the computer just understands numbers. Moreover, the numerical columns which are continuous or having many discrete numbers, must be scaled to bring everything to a common scale.

i) Encoding:-

- **Ordinal Encoding:** This is done for features like income_category and education_level where the order of categories matter as it is hierarchical.
- **Label Encoding:** This is done for features having just binary values or if it's a target variable and where order doesn't matter, like gender and attrition_flag.
- **One hot Encoding:** This is done for variables where there are more unique values (3+) and where order doesn't matter.

ii) Feature scaling:-

- **Standardization:** Also called Z-score normalization, this is the safest scaling method which can be used if using linear/non-linear models. For non-linear models like decision trees, scaling doesn't play a huge role, but for linear models like linear and logistic regression, scaling plays a crucial role.

Class Imbalance Check

Checking the distribution of each class (here, the attrition/churn), is a very important step before starting the modelling. It can give us a caution beforehand regarding the upcoming biases towards the class which dominates the most.

Attrition Flag	Proportion
0 (Not Churn)	83.9%
1 (Churn)	16.1%

From the above table, customers who have not churned clearly dominate and form a majority. This imbalance can worsen the performance of models (like Logistic Regression) which are highly sensitive to it.

Predictive Modelling

3 models are used to predict the churn to see which of them gives the most accurate result:-

(a) Logistic regression: It was chosen as it is an easily interpretable baseline classification model.

Results:

Class	Prec	Rec	F1	Sup
0	0.84	1.00	0.91	1701
1	0.00	0.00	0.00	325
Acc			0.84	2026

Interpretation:

- Prec (Precision): Out of 1701 negatives (not churned) captured, 84% of them are those customers who are predicted truly to haven't churned. Out of 325 people who did churn, 0% of them were correctly predicted to churn.
- This suggests that the model is completely biased towards customers who didn't churn because of heavy class imbalance inclined towards them.
- Overall, it shows 84% accuracy which is also false because it is predicting 0, regardless of whether the customer shows churning characteristics or not. The model didn't learn well mainly due to class imbalance as the logistic regression model is very sensitive to it.

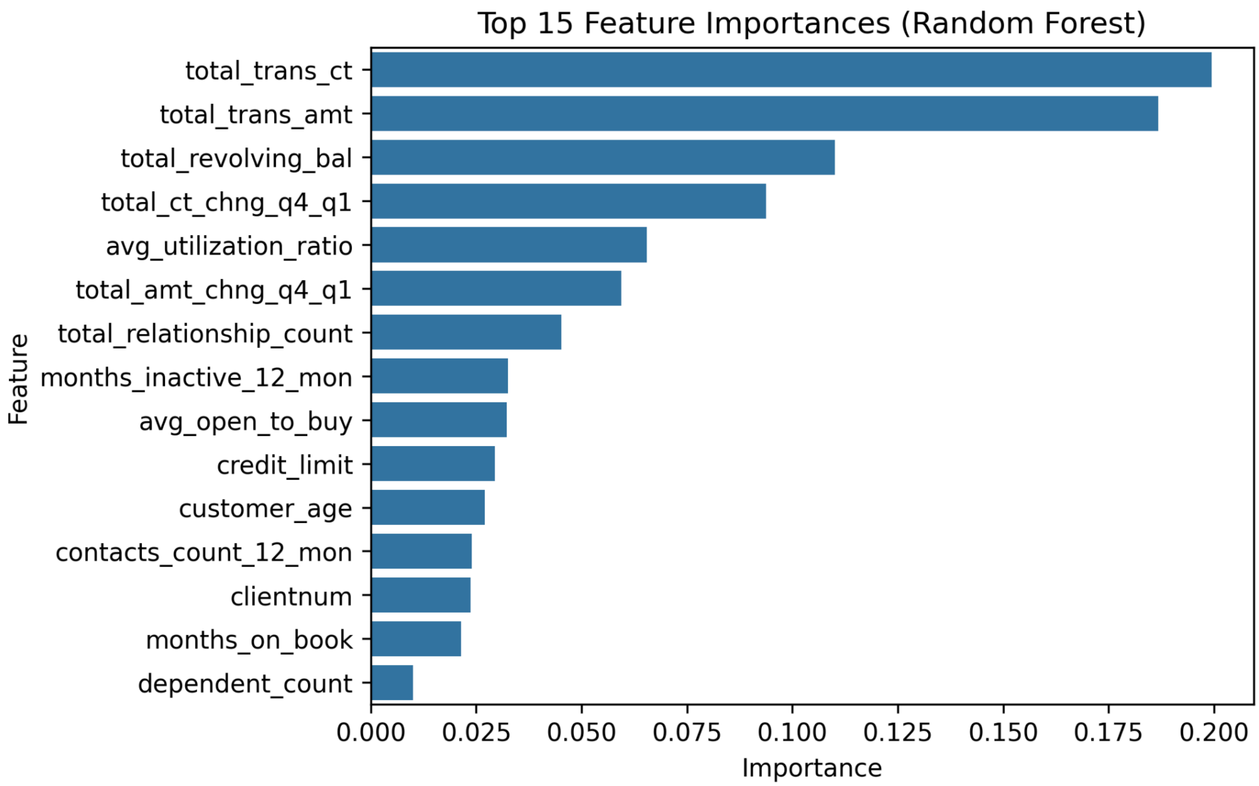
(b) Random Forest: It handles categorical features well and it captures non-linear relationships well.

Results:

Class	Prec	Rec	F1	Sup
0	0.95	0.99	0.97	1701
1	0.94	0.72	0.81	325
Acc			0.95	2026

Interpretation:

- The performance of this model is far better than the logistic regression model. Here, this model is able to identify 72% of the total churners (325) correctly, unlike logistic regression where no churner was identified.
- Out of 72% of positive predictions, 94% of them are true positives, indicating that the model just falsely predicted 6% of non-churners to be churners.
- The bias towards non-churners still prevails but in much lesser magnitude as it is not much affecting the predictions of churners.



Insights:

- Count of transactions show the highest importance, suggesting that fewer the transactions, higher the churn risk. Hence, if customers are less engaged with the credit cards, they are very likely to close it.
- It is followed by importance of total transaction amount and total revolving balance, suggesting the lower the transaction amounts and higher the customer is on credit debts, the lower the customer feels the need to use the card, and hence might close it.
- Frequency of contacts to the bank, clientnum (unique identifier), tenure, and dependent count show the least importance, suggesting that even if the customer contacted the bank frequently, regardless of his/her id and number of dependents he/she has, no matter how long the customer stayed, churn risk is least affected by these features.
- Churn is more about how the customer is using the card and less about his/her demographics and direct relationship with the bank.

(c) XGBoost: It is robust to class imbalance.

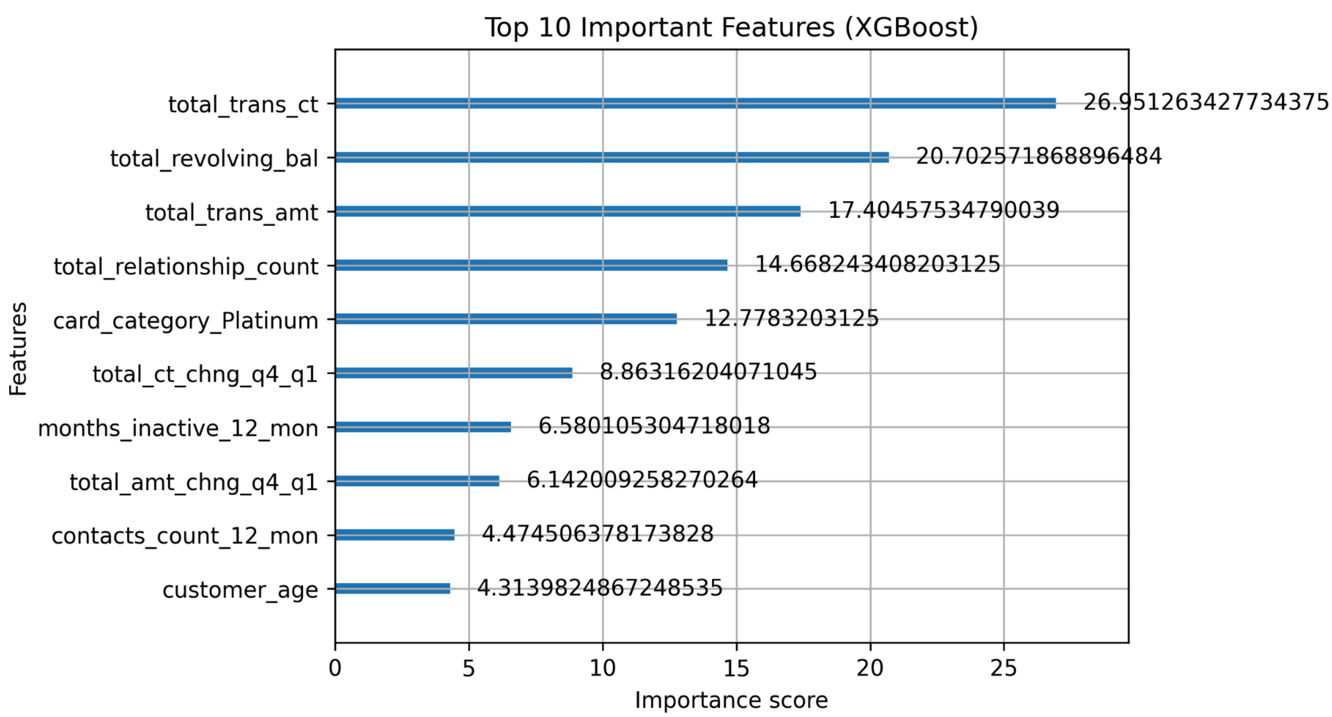
<u>Results:</u>

Class	Prec	Rec	F1	Sup
0	0.98	0.99	0.98	1701

1	0.94	0.89	0.91	325
Acc			0.97	2026

<u>Interpretation:</u>

- A higher number of churned customers (89%) are identified this time by using this model. It shows the same precision.
- Precision is higher for non-churners this time (98%).
- This marks a further decrease in the bias caused by non-churners.



Insights:

- Here again, the importance of features is very similar to that of random forest. Customer transaction frequency remains on the top, showing how crucial the risk factor it is, and how important it is to track the transaction counts of the customers in order to predict whether the customer will close the credit card or not.
- It is followed by total revolving balance and total transaction amount, again indicating them being crucial features which increase the risk of churn when

credit debts of a customer increases and transactions amounts decrease, making the customer disengaged.
- This graph again shows that customer demographics are least important as shown with customer age getting minor importance and it not being a major churn risk driver.
- The majority of the top 10 features according to importance are related to how actively the customer uses the card.

<u>Model performance comparison:</u>

Model	Prec(0)	Recall (0)	F1(0)	Prec(1)	Recall (1)	F1 (1)	Acc
Logistic Regression	0.84	1.00	0.91	0.00	0.00	0.00	0.84
Random Forest	0.95	0.99	0.97	0.94	0.72	0.81	0.95
XGBoost	0.98	0.99	0.98	0.94	0.89	0.91	0.97

Conclusion:-

The end to end churn analytics highlight how different customer attributes, engagement and credit usage influence churn risk. During the churn profiling, several high risk segments were identified which possibly affected the overall churn rate. At the same time, loyal segments were also identified which depicted low churn risk.

The retention analysis showed how the distribution of the churners and non churners is disproportionate and how it significantly varies among 2 different groups, overall vs churned. While high risk customers formed one-third of the entire dataset, they contributed two-thirds among churned customers. This suggested that high risk customers contributed the most to overall churn rate when compared to low risk customers and hence the retention campaigns should focus more on the high risk segment in order to reduce the overall churn risk.

Churn prediction was done to get to know if a customer will churn, in advance. The logistic regression model failed to capture the churners and predicted the whole test data as non churners due to heavy class imbalance, hence failing. Random forest models showed significantly improved performance by identifying 72% of churners. The XGBoost model gave the best performance by identifying 89% of churners and gave an accuracy of 97%. Plotting the feature importance graph showed that transaction related features (transaction counts, transaction amounts, revolving balance) are the strongest churn drivers, while customer demographics are not easily predictable. This confirmed that churn is mainly based on customer behaviour.