

Consolidated Segmentation and Churn Analysis of Bank Clients

By: Tamjid Ahsan

Date: July 22, 2021, 05:00 PM [DST]



OVERVIEW

- Customer churn has direct impact on profit.
 - Cost of maintaining existing customers is significantly less than the cost of obtaining a new one.
- The financial crisis of 2008 changed the banking sector's strategy.
 - Previously focused on acquiring more and more clients.
 - Technology and laws making things easier than ever to transfer assets and money between institutions. Which introduced new competitors in market
 - open banking
 - neo-banks
 - fin-tech businesses (Banking as a Service (BaaS))
 - Banks can use existing data to tackle client turnover challenge.

BUSINESS PROBLEM

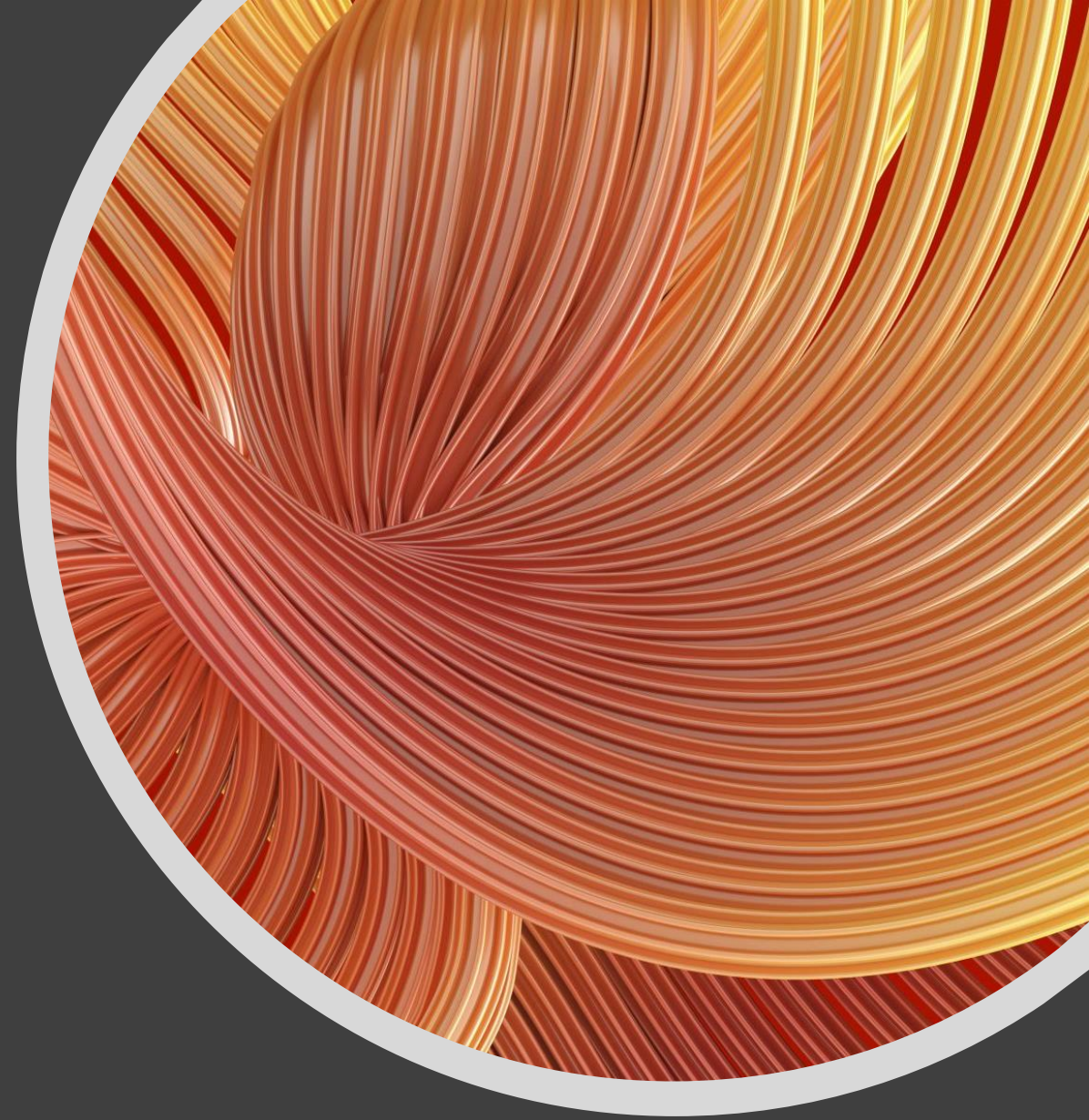
Importance of existing customer management is well recognized.

- Hard to anticipate

Feedback from customers is hard to obtain.

XYZ Bank (read: fictional) is a mature financial institution based in North America.

- New competitors in market.
- Have existing data of their clients.
- Based on the data available, the bank wants to predict risk of churning.



METHODOLOGY



This analysis is combining churn prediction and customer segmentation.

- Using clusters from segmentation model as input for prediction model.



Customer data of the bank is used for this analysis.

- 10127 unique client information.
- 18 features.



Customer segmentation: 'K-means' clustering.

- Divided into five clusters.



Churn prediction: Using the predictions from the customer segmentation model, a 'XGBClassifier' model is used

- 0.97 model accuracy
- 0.90 precision for churn class.

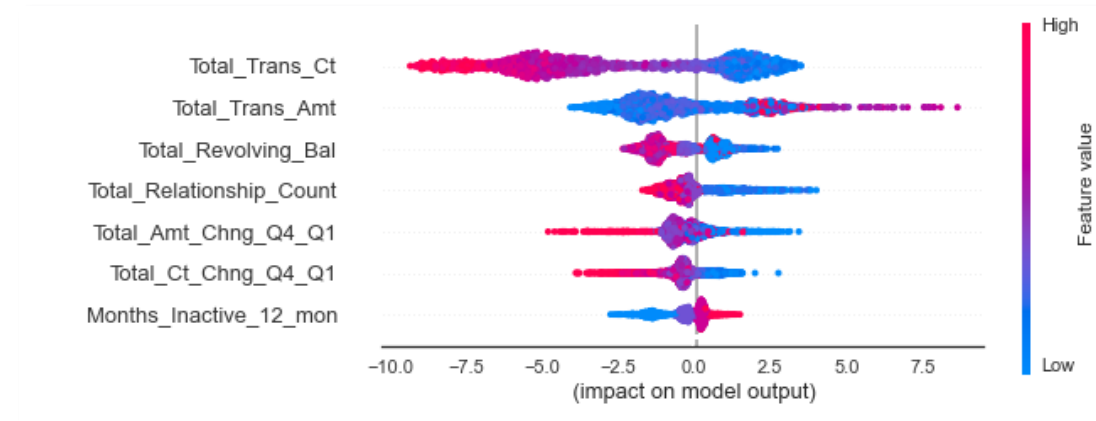


What factors alienates clients

CHURN PREDICTION

Impact of feature values

Feature	Observation
Total_Trans_Ct	Low value means higher risk of churning
Total_Trans_Amt	Above average value means higher risk of churning
Total_Revolving_Bal	Low value means higher risk of churning
Total_Relationship_Count	More relationship indicates higher chance of churning
Total_Amt_Chng_Q4_Q1	Low value means higher risk of churning
Total_Ct_Chng_Q4_Q1	Low value means higher risk of churning
Months_Inactive_12_mon	Higher value means higher risk of churning
Contacts_Count_12_mon	Higher value means higher risk of churning



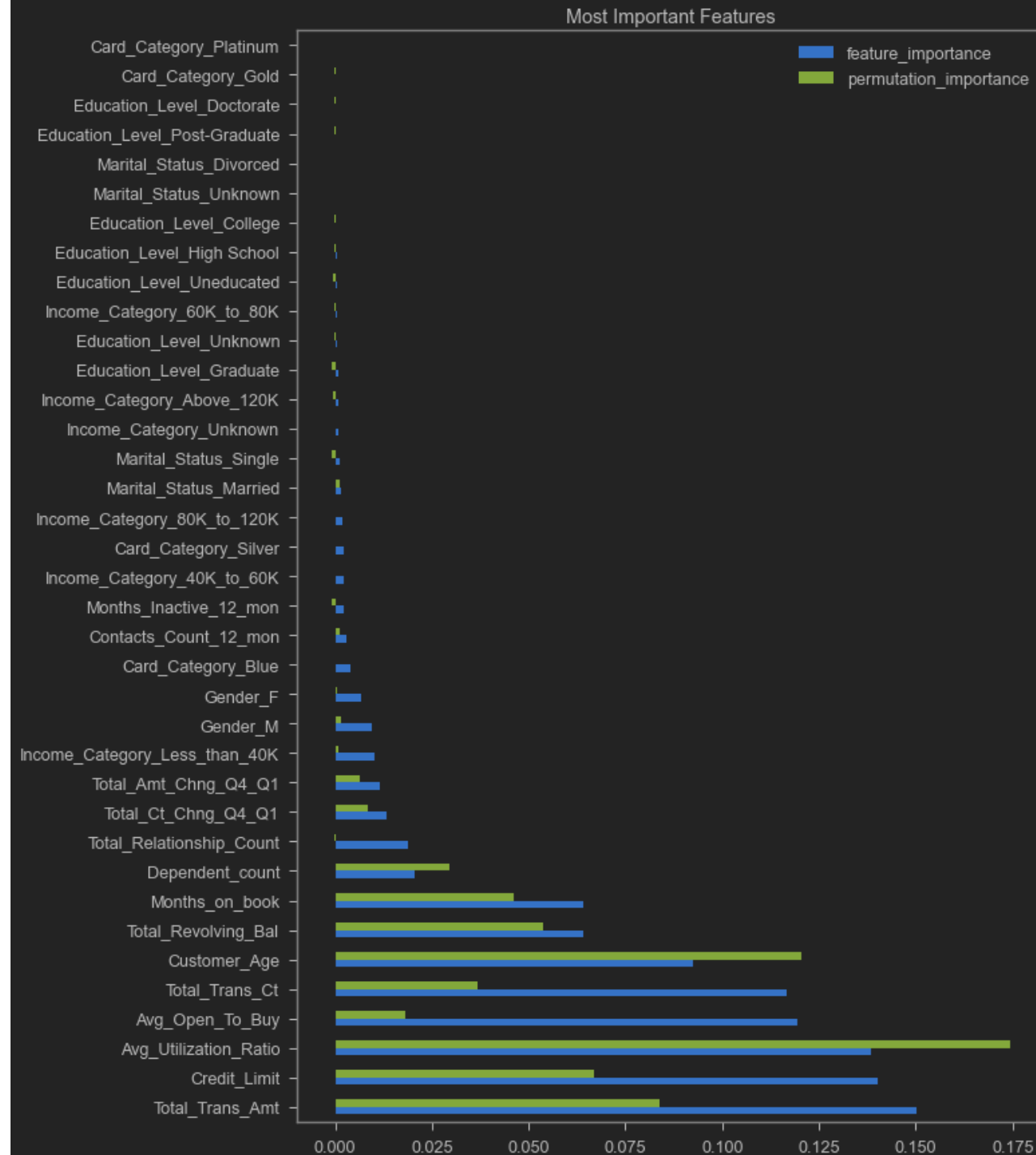
CLUSTERS

Peeking into clusters



Features that isolates client segmentation

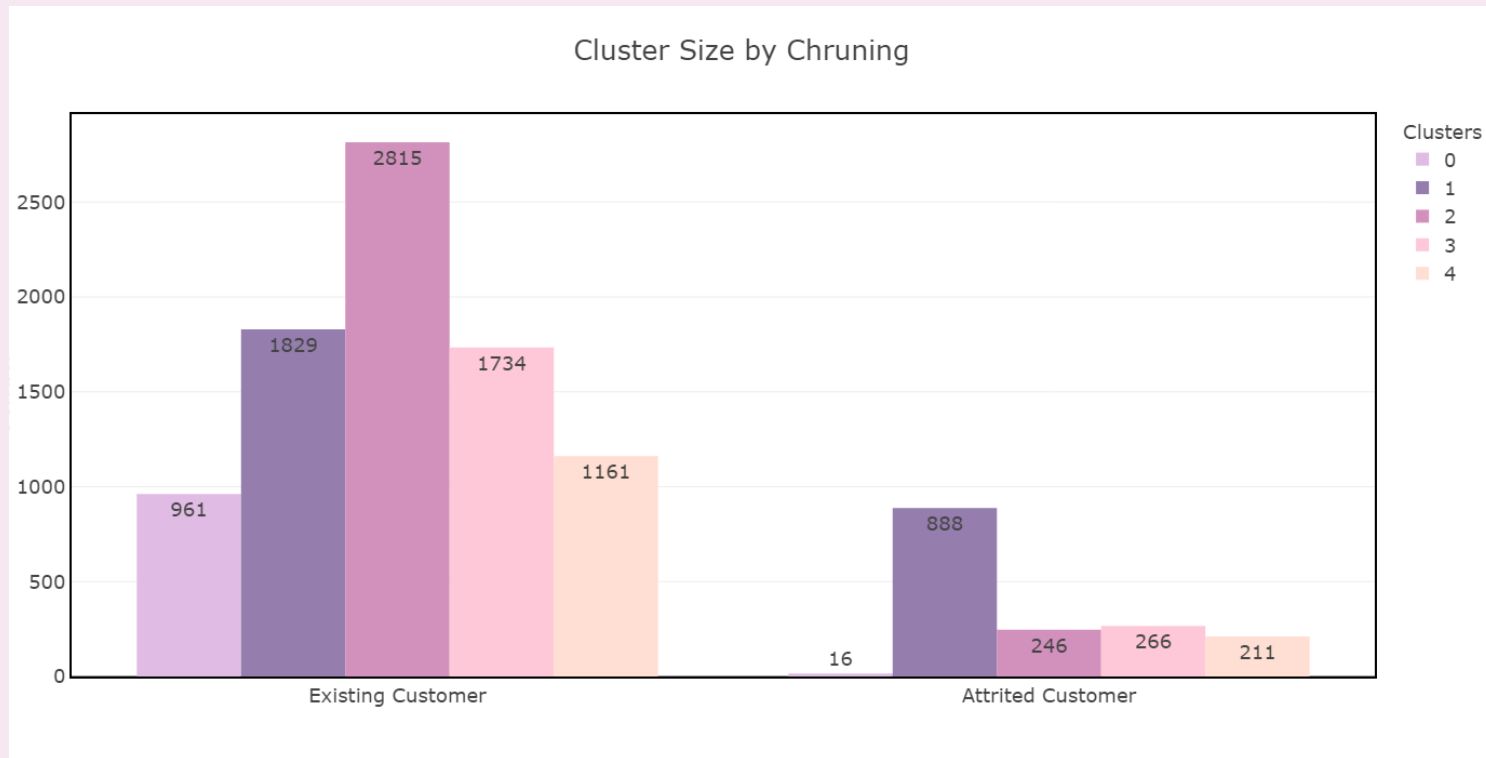
- With insights from model, these are the topmost important features.
 - Total Transaction Amount
 - Credit Limit
 - Avg Utilization Ratio
 - Avg Open To Buy
 - Total Transaction Count
 - Customer Age
 - Total Revolving Balance
 - Months on book
 - Dependent count
 - Total Relationship Count



Identifying Clusters

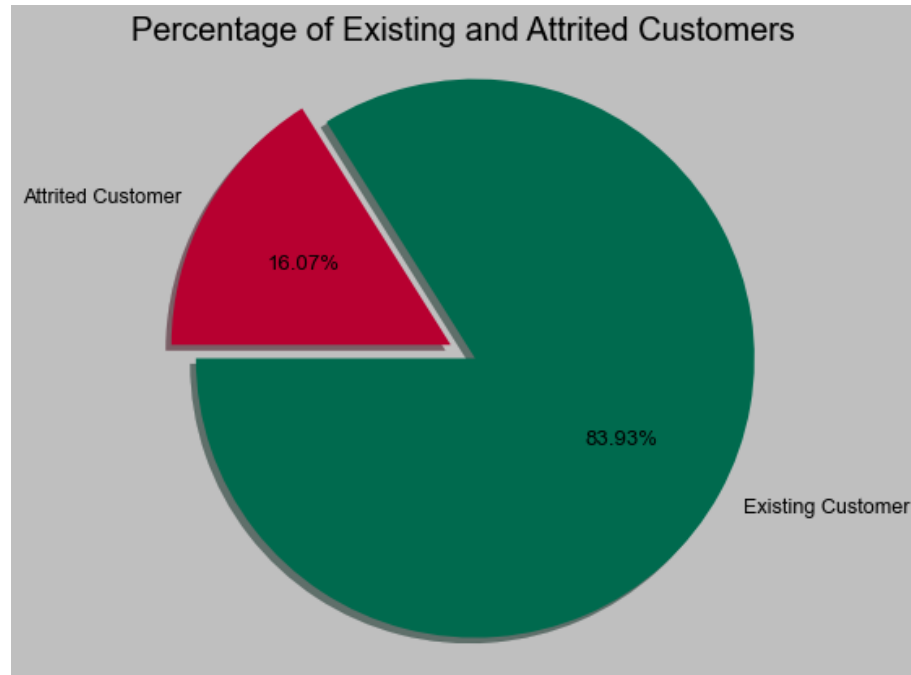
CLUSTER	LABEL
Cluster 0	Low value frequent users of services
Cluster 1	High risk clients
Cluster 2	Regular clients
Cluster 3	Most loyal clients. (mostly consists of older clients)
Cluster 4	High value clients

Clusters & Churn

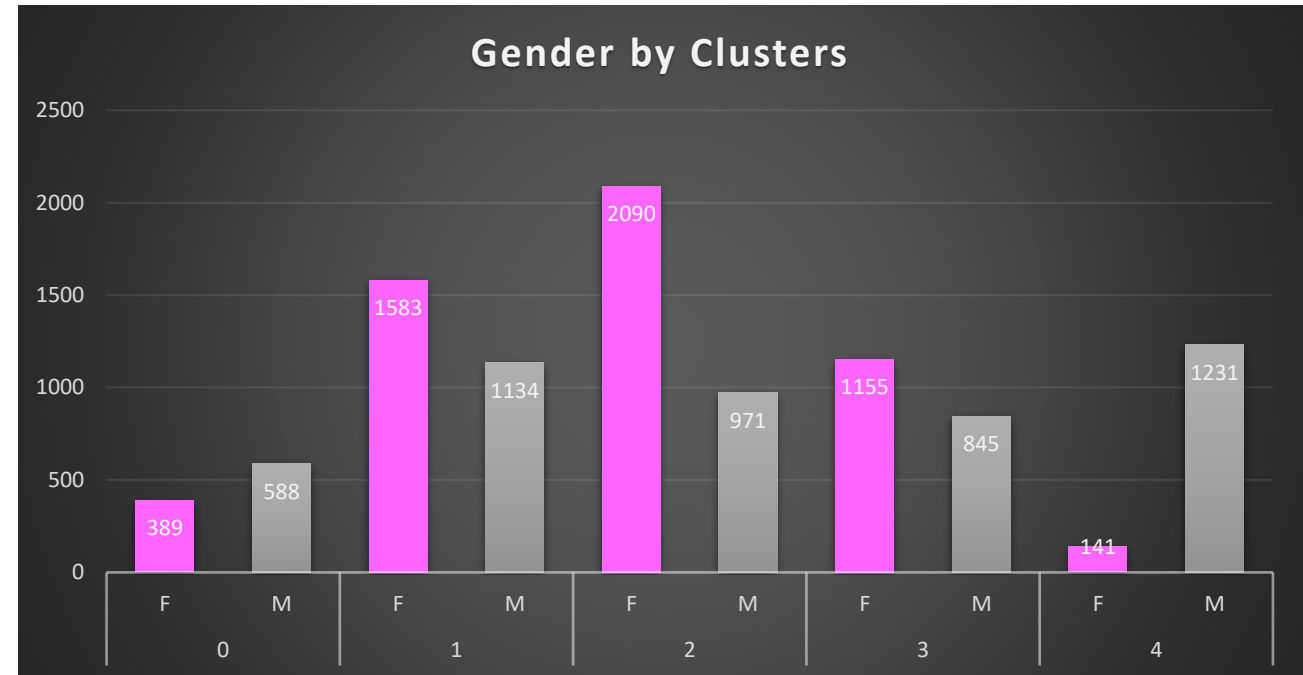


- Cluster 0 has low churn rate.
- Cluster 2, labeled as High Risk, has large number of churned clients.

Cluster size



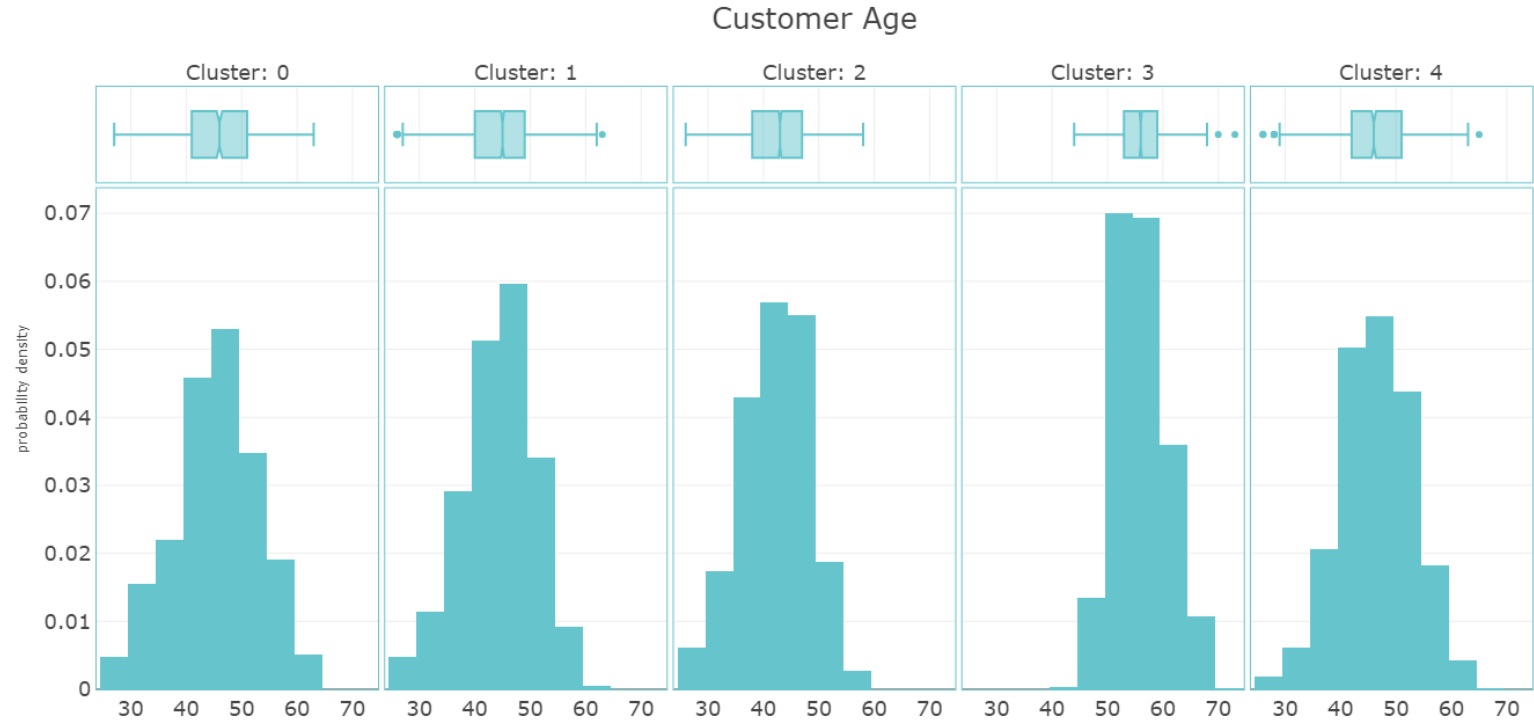
- 16% of customers terminated their relationship with the bank.



- Female clients dominates cluster 1 and 2.
 - They can be targeted with tailored offer.

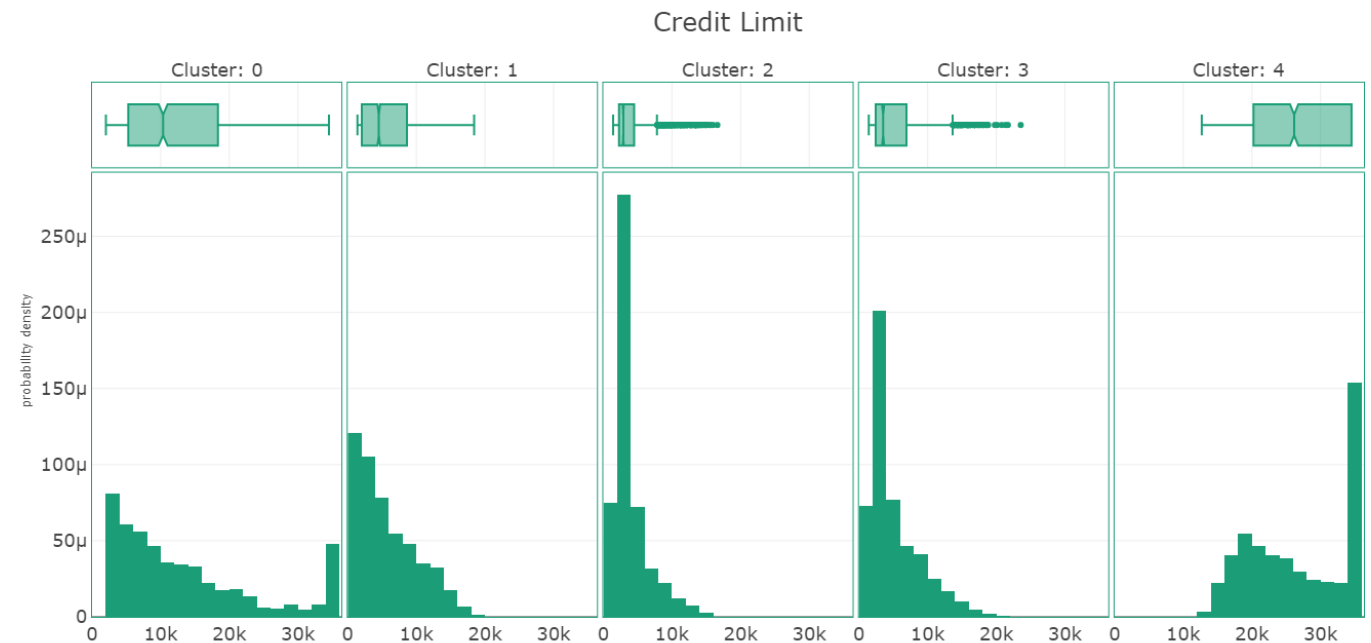
Customer Age

- Cluster 1 and 4 has similar distribution.
- Cluster 0 is slightly younger.
- Cluster 3 is distinct as it is mostly comprised of older clients.
- Others have similar distribution.



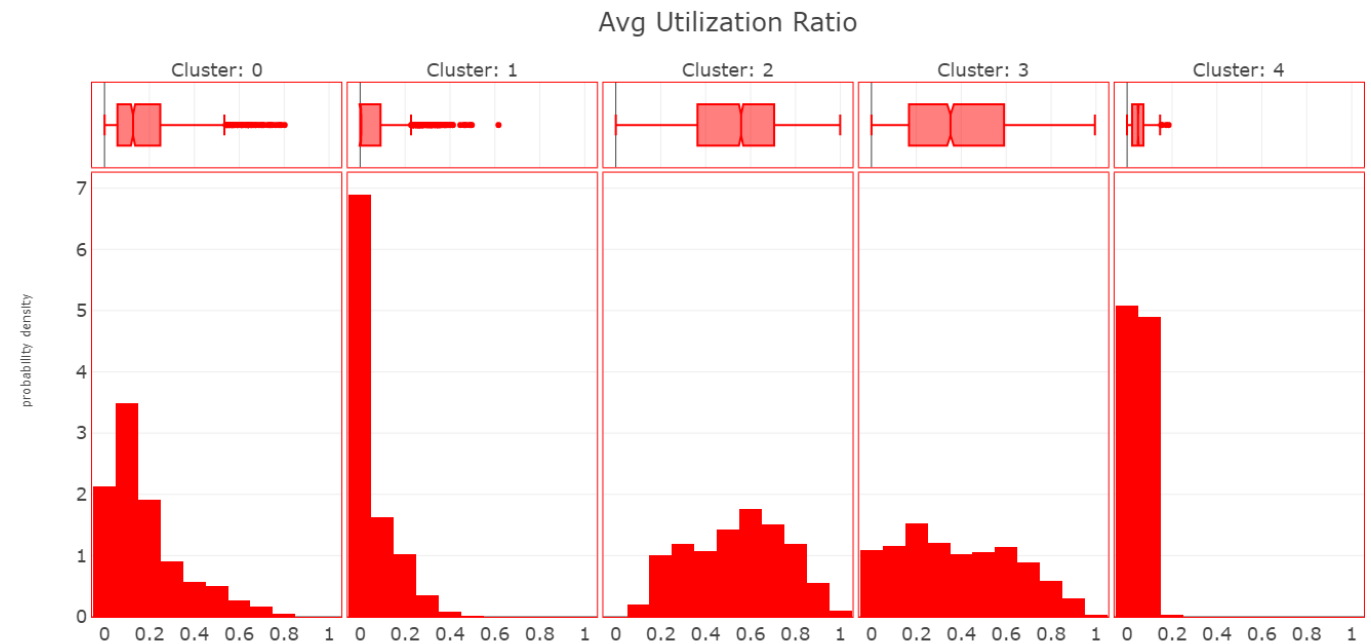
Credit Limit

- Cluster 1 has mostly lower credit limit clients.
- Cluster 0 has a well-balanced distribution; it does not have lower credit limit clients.
- Cluster 2 and 3 has mostly same characteristics.
- Cluster 4 has the clients with mostly high credit limit.



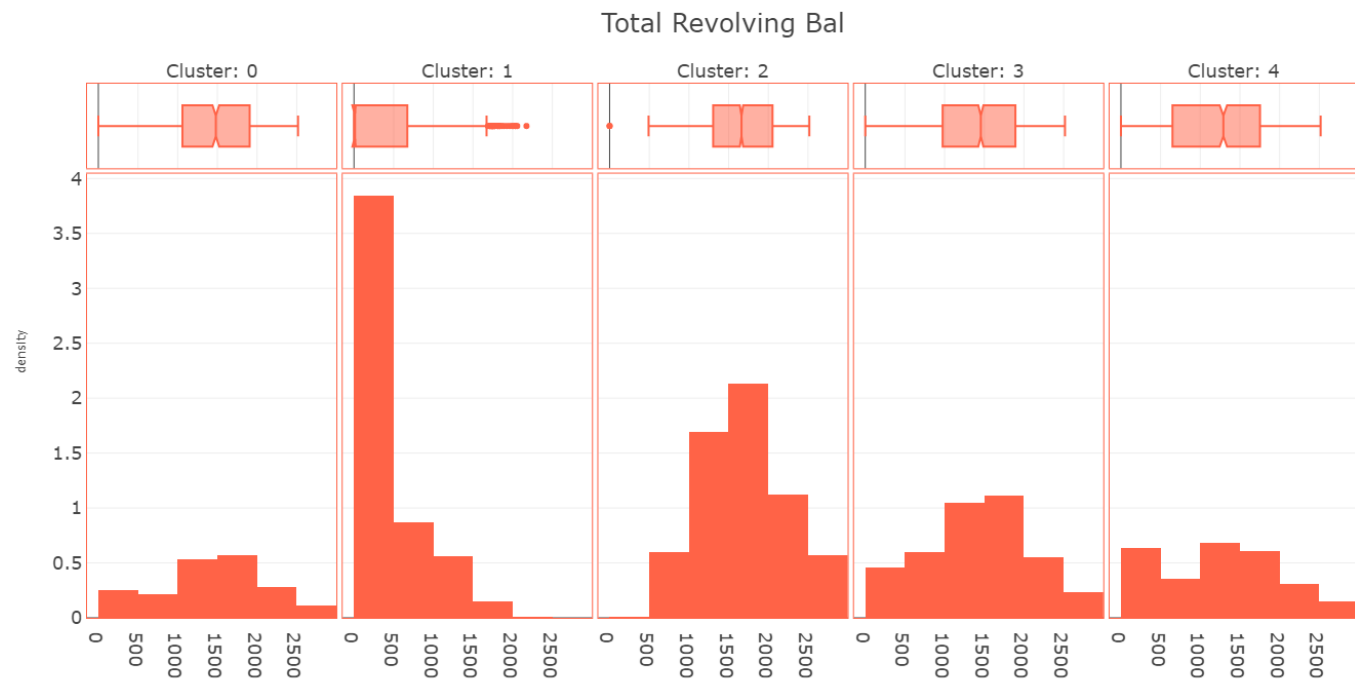
Average Utilization Ratio

- Cluster 1 has mostly less utilization ratio.
- Cluster 0 shows good utilization ratio, with some 0.
- Cluster 2 and 3 has similar utilization. Cluster 2 does not have many 0's.
- Cluster 4 has low utilization of credit.



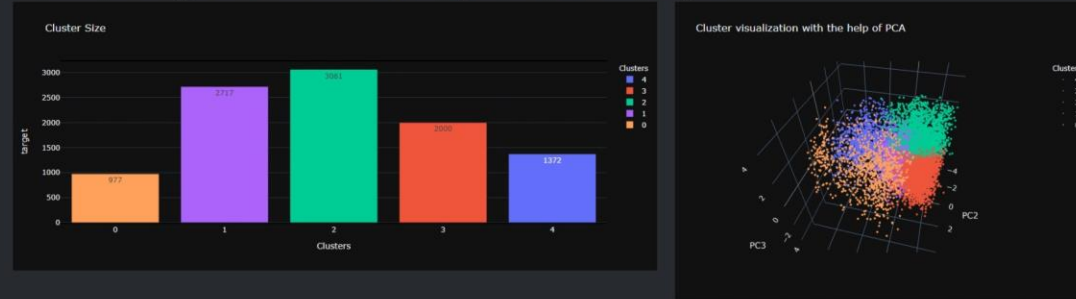
Total Revolving Balance

- Cluster 1 has mostly low revolving balance.
- Cluster 0 has even distribution.
- Cluster 2 does not include low revolving balance clients.
- Cluster 3 and 4 has similar distribution.



Dashboard

Consolidated Segmentation and Churn Analysis of Bank Clients Report



Intracuster analysis



All the features with churning



Prediction Model

Please input features (press ENTER after inputting)

F								
4								
Graduate								
Married								
80K_to_120K								
Blue								
30	40	2	0	2	10000	2000	2000	1
1500	20	1	2					

Prediction:
Client is identified as "Regular client" with a prediction of Continuing.

BY Tamejd Ahsan, Contact: [LinkedIn](#), [GitHub](#)

- Additional insights
- Predict new clients churn and cluster group.



RECOMMENDATIONS

Cluster 1 is the riskiest client segmentation.

- Their **utilization ratio is low**.
 - offering incentives like **cash back**.
- Their **credit limits are low**. Based on their credit habit, they can be offered a larger credit limit.

As a rule of thumb:

- Marketers should target **female clients** with specific package.
- **Frequent smaller amount of transaction** can be perceived as a red flag. When spotted, customer relationship team must act on it.
- **Large expenditure** can be a signal for cross selling opportunity, and it is also a sign of churn.

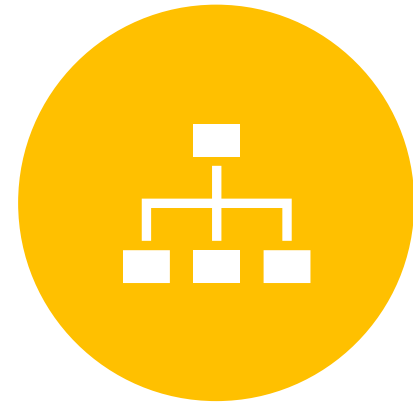
THANK YOU



GITHUB: @tamjid-ahsan



LINKEDIN:
[linkedin.com/in/tamjidahsan/](https://www.linkedin.com/in/tamjidahsan/)



PROJECT REPO:
https://github.com/tamjid-ahsan/capstone_customer_churn



APPENDIX

Features of the dataset

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 1 else 0
Customer_Age	Demographic variable - Customer's Age in Years
Gender	Demographic variable - M=Male, F=Female
Dependent_count	Demographic variable - Number of dependents
Education_Level	Demographic variable - Educational Qualification of the account holder (example: high school, college graduate, etc.)
Marital_Status	Demographic variable - Married, Single, Divorced, Unknown
Income_Category	Demographic variable - Annual Income Category of the account holder (< 40K, 40K - 60K, 60K—80K, 80K—120K, > \$120K, Unknown)
Card_Category	Product Variable - Type of Card (Blue, Silver, Gold, Platinum)
Months_on_book	Months on book (Time of Relationship)
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 in the last 12 months
Credit_Limit	Credit Limit on the Credit Card
Total_Revolving_Bal	Total Revolving Balance on the Credit Card
Avg_Open_To_Buy	Open to Buy Credit Line (Average of last 12 months)
Total_Amt_Chng_Q4_Q1	Change in Transaction Amount (Q4 over Q1)
Total_Trans_Amt	Total Transaction Amount (Last 12 months)
Total_Trans_Ct	Total Transaction Count (Last 12 months)
Total_Ct_Chng_Q4_Q1	Change in Transaction Count (Q4 over Q1)
Avg_Utilization_Ratio	Average Card Utilization Ratio

Summary of exploring

Variable	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Churn	Comment
Avg_Open_To_Buy	spread	low	low	low	high value	1	Majority values are low
Avg_Utilization_Ratio	low utilization	minimal low utilization	no low utilization ratio	med utilization	low utilization	1	Majority values are low
Card_Category						1	High class imbalance to comment
Credit_Limit	all clients from 2k	mostly low limit	2k to 4k, no high limit		high limit, above 14k	1	
Customer_Age	similar	similar	similar	older	similar	3	
Dependent_count	spread	spread	spread	low	spread	1	count 3 and 4 is risky
Education_Level	Graduate	Graduate	College	College	Uneducated	1	Graduates > HS >= Unknown >= Uneducated, PG and PhD is less likely to churn
Gender	M	F	F	F	M	1	Females is risky
Income_Category	Less_than_40K	40K_to_60K	40K_to_60K	Less_than_40K	Unknown	1	Less than 40K
Marital_Status	Unknown	Single	Married	Married	Unknown	1	Majority values is Married
Months_on_book	good	similar	similar	loyal customer	similar	3	
Total_Relationship_Count	low	high	high	high	high	1	2 and 3 are most frequent
Total_Revolving_Bal	spread	low	mod	spread	spread	1	Majority values are low
Total_Trans_Amt	High transaction amount	low	mid amount till 5k high freq transaction	mid amount till 5k high freq transaction	mid amount till 5k med freq transaction	1	low amounts
Total_Trans_Ct	heavy user	moderate user	moderate user	moderate user	moderate user	1	Majority values are between 30 to 50

Exploration

Credit Limit & Avg Utilization Ratio on Customer Churn

