

# Consolidated Segmentation and Churn Analysis of Bank Clients

By: Tamjid Ahsan

Date: July 22, 2021, 04:00 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

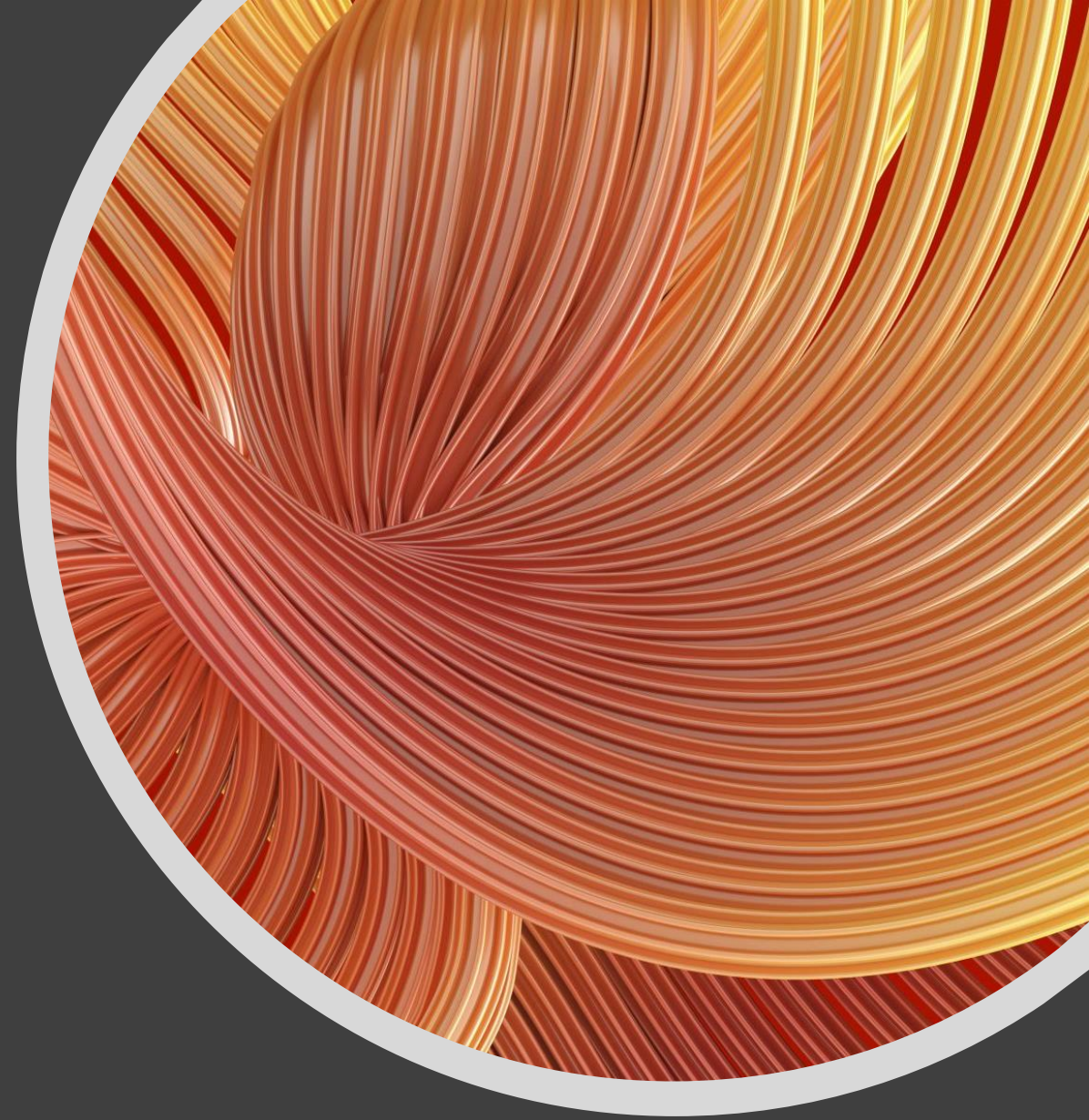
Importance of existing customer management is well recognized.

- 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 know whom of them are in risk of churning.



# METHODOLOGY



This analysis is combining churn prediction and customer segmentation.



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

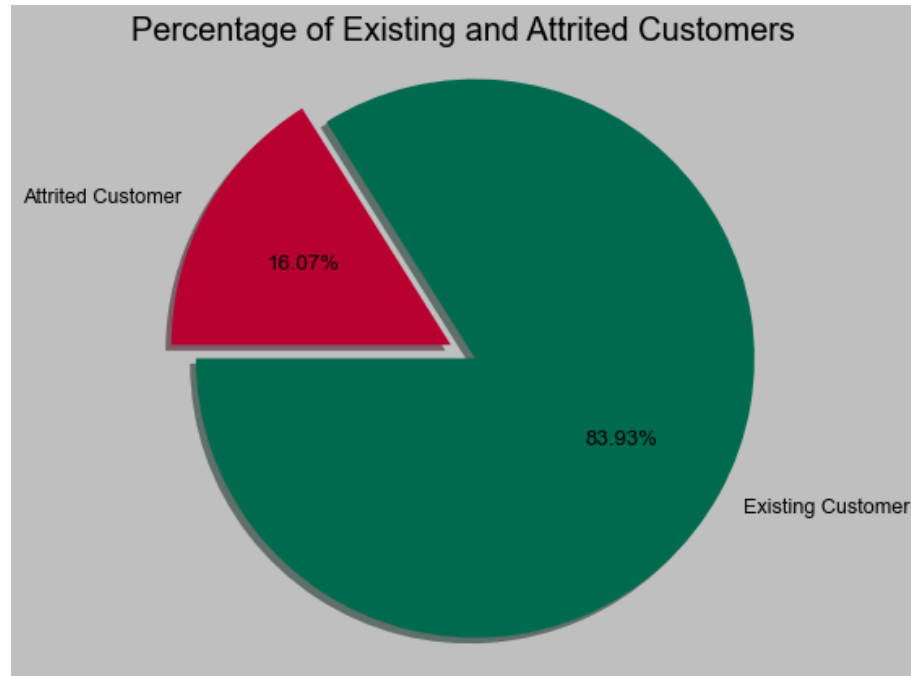
- 0.97 model accuracy
- 0.90 precision for churn class.

# CLUSTERS

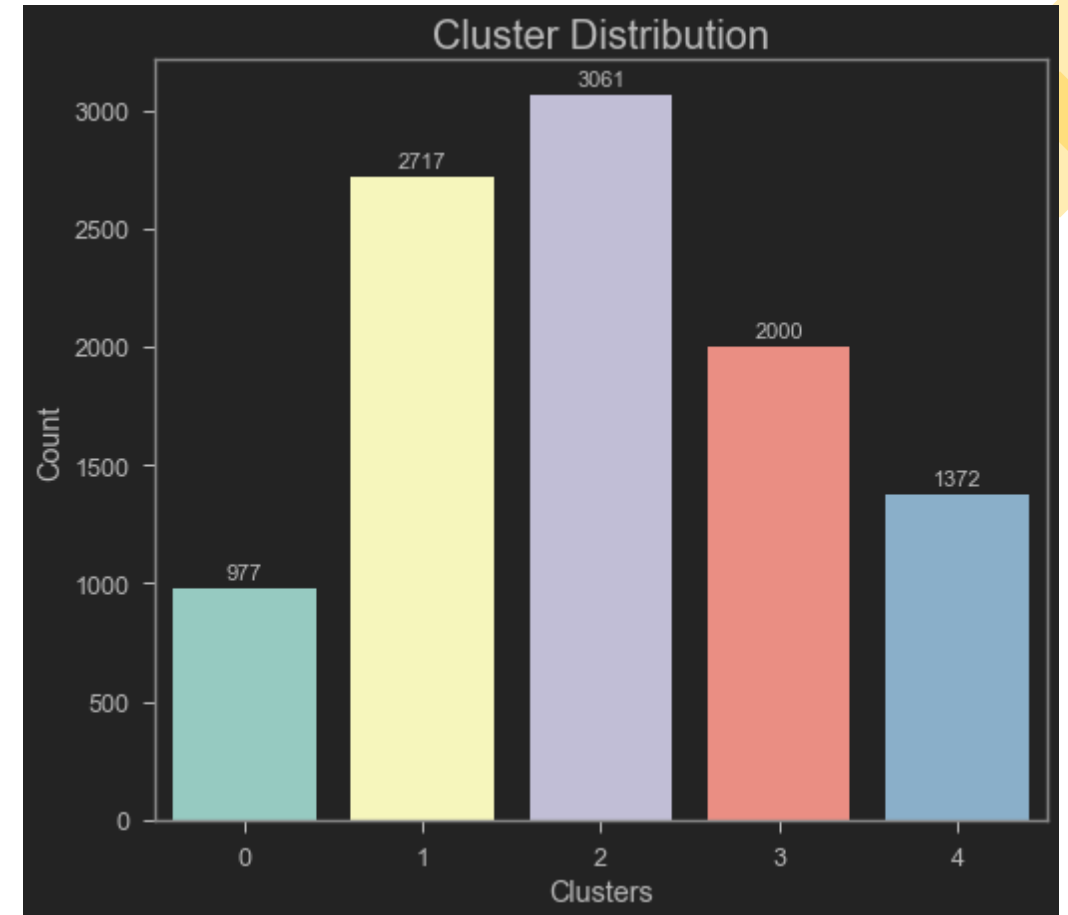
Peeking into clusters



# Cluster size



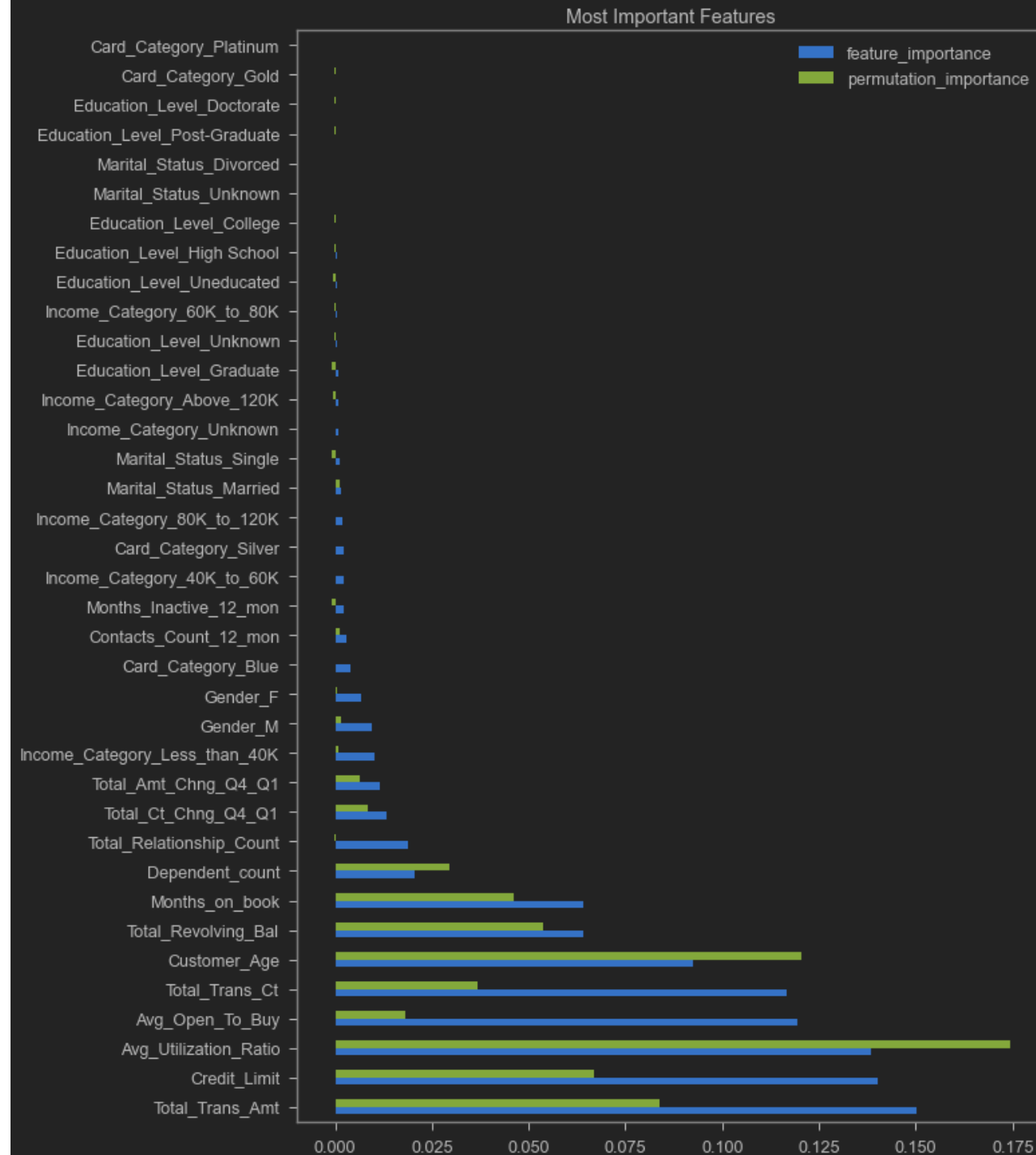
- 16% of customers terminated their relationship with the bank



- Cluster 2 is the largest.
- Cluster 1 is next.
- Rest are close to 1k.

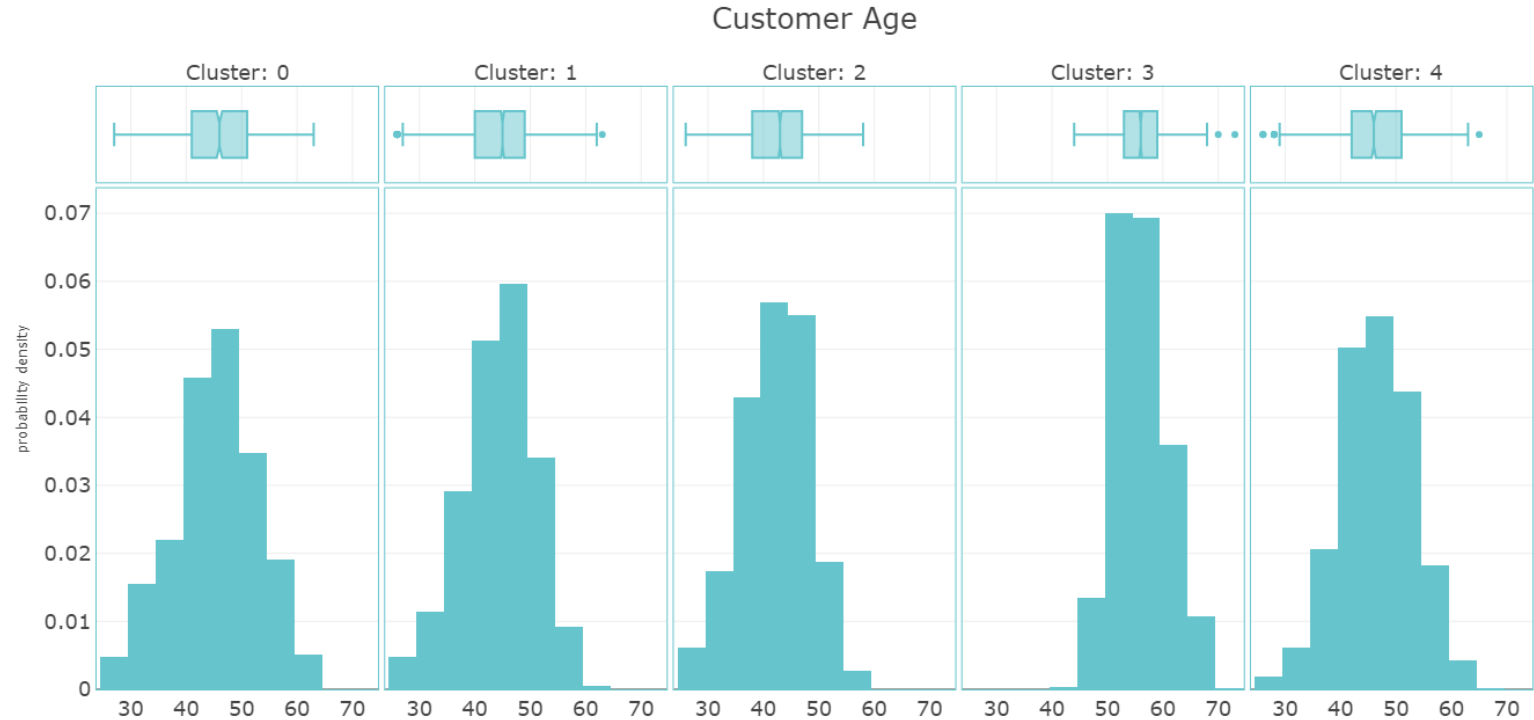
# Features that isolates client segmentation

- 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



# Customer Age

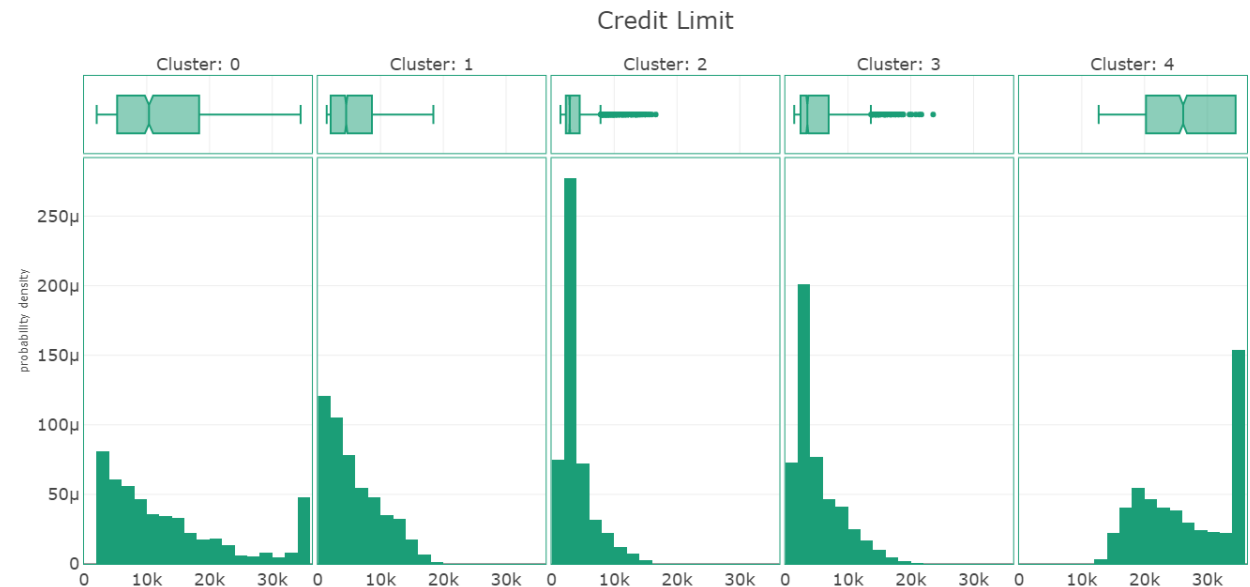
- Cluster 4 and 1 has similar distribution.
- Cluster 0 is slightly younger.
- Cluster 3 is distinct as it is mostly comprised of older clients.
- Others have similar distribution.





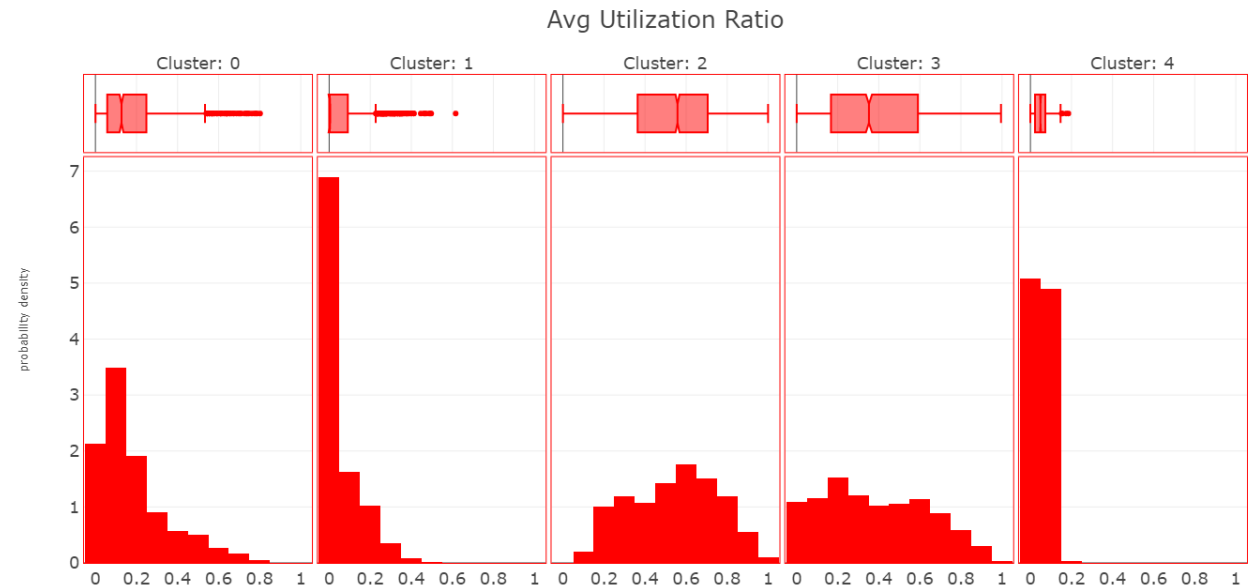
# Credit Limit

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



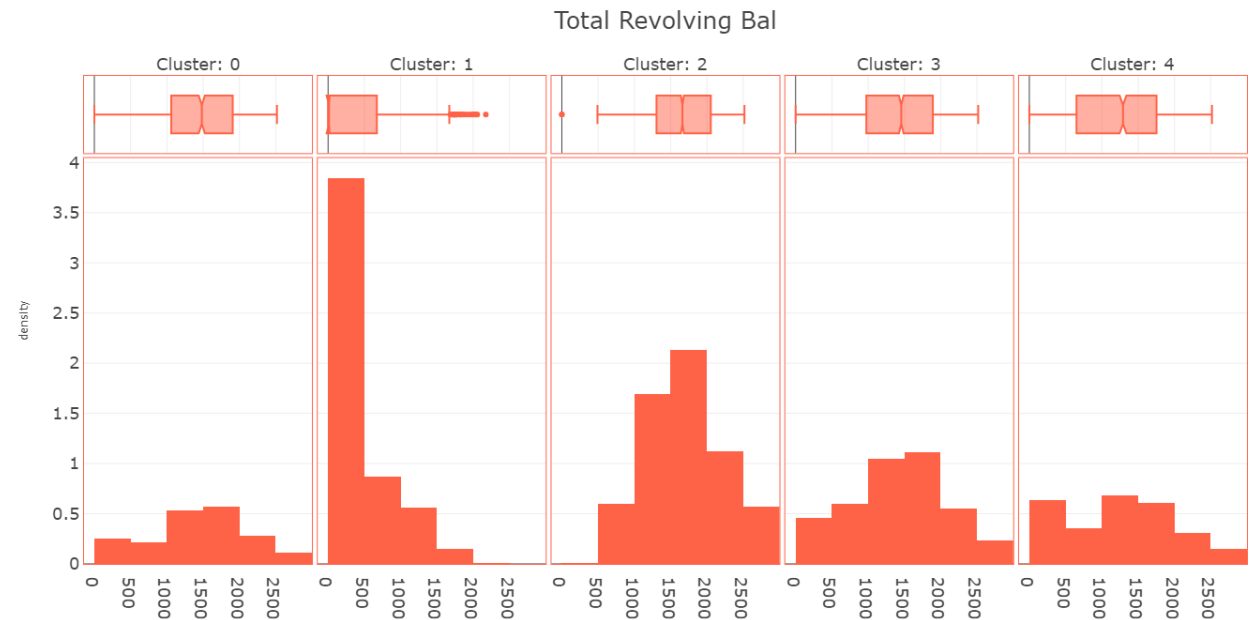
# Average Utilization Ratio

- Cluster 0 shows good utilization ratio, with some 0.
- Cluster 1 has mostly less utilization ratio.
- 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 0 has even distribution.
- Cluster 1 has mostly low revolving balance.
- Cluster 2 does not include low revolving balance clients.
- Cluster 3 and 4 has similar distribution.



# 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



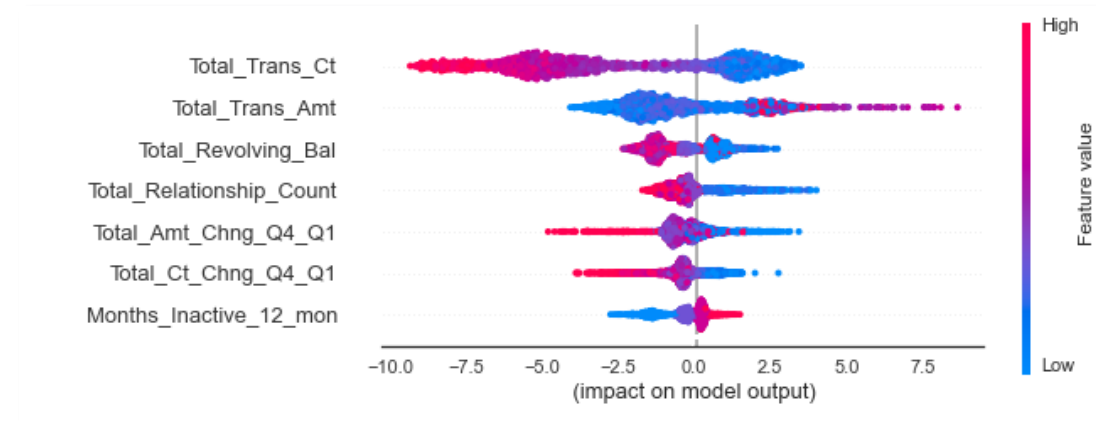
What factors alienates clients

# 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





# RECOMMENDATIONS

---

Cluster 1 is the most 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 products and it is also a sign of churn.

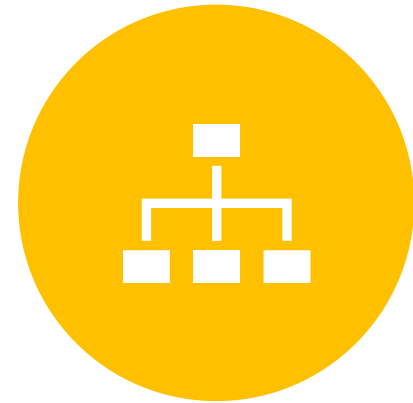
# THANK YOU



GITHUB: @tamjid-ahsan



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



PROJECT REPO:  
[https://github.com/tamjid-ahsan/capstone\\_customer\\_churn](https://github.com/tamjid-ahsan/capstone_customer_churn)



# APPENDIX

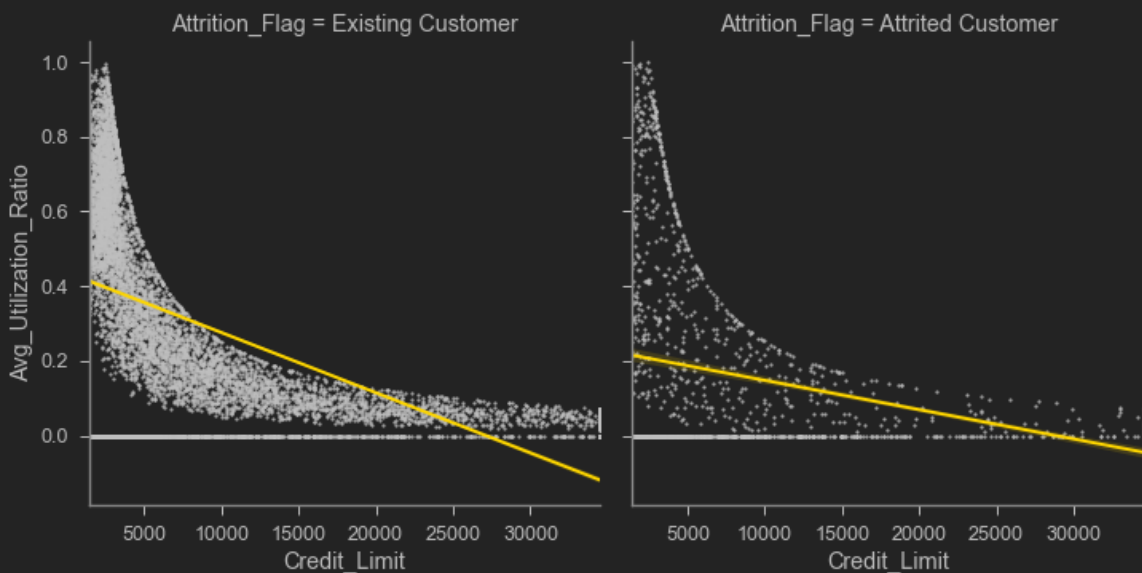
# Features of the dataset

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



# Exploration

Credit Limit & Avg Utilization Ratio on Customer Churn



Distribution of Categorical Features

