**Analytical CRM Development for a Bank**

**Introduction:**

This comprehensive project aims to analyse customer data provided by a bank to understand customer churn (customer loss). We'll identify key factors driving churn, develop actionable insights to improve customer retention, and ultimately enhance customer satisfaction. By leveraging various tools like Excel, Power BI, and SQL, we'll gain a deeper understanding of customer behaviour and preferences.

**Objective:**

The primary objective of this project is to reduce customer churn and enhance customer satisfaction for the bank. This will be achieved through:

1. Identifying key factors contributing to customer churn.

2. Developing insights to improve customer retention strategies.

3. Enhancing service delivery based on customer preferences and behavior.

**Introduction to Data**

The bank has provided several datasets related to customers, including:

* Customerid**:** A unique identifier for each customer.
* CreditScore**:** A numerical representation of the customer's creditworthiness.
  + **Credit score:** 
    - Excellent: 800–850
    - Very Good: 740–799
    - Good: 670–739
    - Fair: 580–669
    - Poor: 300–579
* GeographyID**:** A numerical identifier that likely corresponds to a geographical location, such as a country or region.
* GenderID**:** A numerical identifier for the customer's gender, where for example, '1' could represent male and '2' could represent female.
* Age**:** The age of the customer.
* Tenure**:** The number of years the customer has been with the bank.
* Balance**:** Current balance in the customer's account.
* Num Of Products: refers to the number of products that a customer has purchased through the bank.
* HasCrCard: denotes whether or not a customer has a credit card. This column is also relevant, since people with a credit card are less likely to leave the bank.
  + - 1 represents credit card holder
    - 0 represents non credit card holder
* **IsActiveMember:** active customers are less likely to leave the bank (as per the criteria defined by the bank for identifying the activeness).
  + - 1 represents Active Member
    - 0 represents Inactive Member
* **Estimated Salary:** as with balance, people with lower salaries are more likely to leave the bank compared to those with higher salaries.
* **Exited:** whether or not the customer left the bank.
  + - 0 represents Retain
    - 1 represents Exit
* **Bank DOJ:** date when the Customer associated/joined with the bank.

**Data Exploration:**

● **Data Sources:** We will carefully review the datasets that have been provided, which include account activity, demographics about the customers, and possibly additional relevant information on churn.

● **Data Understanding:** We will examine the importance and meaning of each data part found in the datasets. This include recognizing potential problems such discrepancies or missing values and understanding all of the forms of data (numerical, categorical, etc.).

● **Data Visualization**: An initial look at data distribution and any correlations between variables can be obtained by using tools such as scatter plots and histograms.

**Data Processing:**

1. **Cleaning Data with Power BI and Excel**

● **Missing Value Imputation**: We will use methods such as mean/median imputation for numerical data and mode imputation for categorical data to proactively handle missing data points.

● **Error Correction**: If there are any mistakes, outliers, or inconsistent date formats or currency representations, we'll carefully find and fix them.

● **Data Standardization**: By standardizing formats among datasets, we'll guarantee data consistency. This could include translating currencies to a single unit for analysis as well as date formats (MM/DD/YYYY, for example).

● **Feature Engineering**: If needed, we can extract additional features from the data that already exists. Examples include figuring out client age groups and duration (years) since joining the bank by using past transaction data.

1. **Data Transformation in SQL**

● Database Schema develop: To effectively store and manage the combined customer data, we will develop a relational database schema in SQL that is well-organized. The retrieval and manipulation of data for analysis will be optimized by this template.

● Data Cleaning Queries: To further clean and modify data, we'll make use of strong SQL queries. This involves changing data types (e.g., converting strings to dates), handling missing values using techniques like AVG (), MEDIAN (), or COALESCE (), and using aggregation functions like COUNT () and GROUP BY to calculate derived variables like churn rate by demographics.

**Creating Columns using conditional Columns (Power Query Editor):**

**Conditional Columns:** In Power BI Query Editor, conditional columns allow us to create new columns based on the logical conditions applied to existing columns. This feature is useful for adding calculated columns that reflect certain criteria or business rules directly within our data transformation process.

**Using conditional column to create Age-Group column:**

This column segments costumers into age groups using conditional column logic,

**A screenshot of a computer

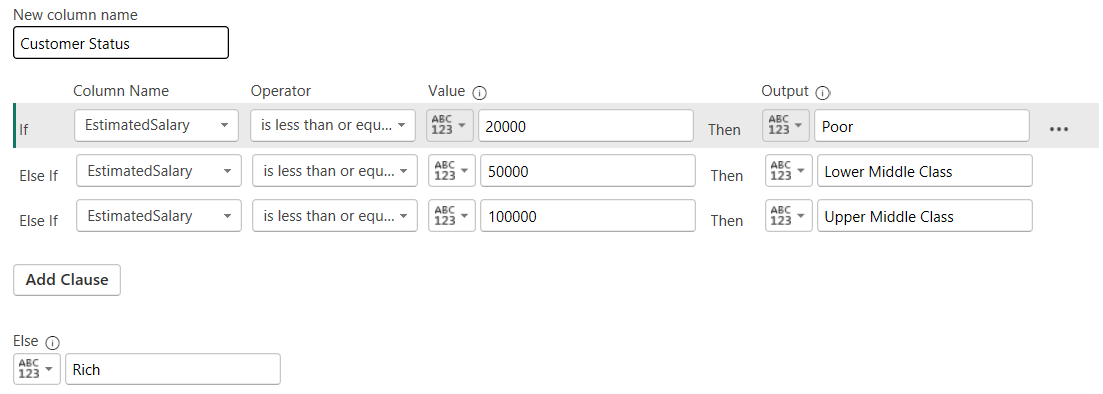
Description automatically generated**

**Using Conditional column to create Salary-Segments and Customer status:**

These columns segments costumers estimated salary into salary segments and customer status, such as

A screenshot of a computer

Description automatically generated



**Using Conditional column to create Balance segment:**

This column segments customers account balance into balance segments:

A screenshot of a computer

Description automatically generated

**Calculated Columns:**

The capabilities of your data are increased by calculated columns, which generate new data points from existing information. They make analysing data easier and provide insightful information. Below is a summary of the calculated columns that are available:

**CreditType( crm bank\_churn):**

This column segments customers based on their credit score using nested if statement:

CreditType =

SWITCH(

    TRUE(),

'crm bank\_churn'[CreditScore] >= 800 &&'crm bank\_churn'[CreditScore] <= 850, "Excellent",

'crm bank\_churn'[CreditScore] >= 740 && 'crm bank\_churn'[CreditScore]<= 799, "Very Good",

'crm bank\_churn'[CreditScore] >= 670 && 'crm bank\_churn'[CreditScore] <= 739, "Good",

'crm bank\_churn'[CreditScore] >= 580 && 'crm bank\_churn'[CreditScore] <= 669, "Fair",

    "Poor"

)

**CustomerType(crm customerinfo):**

This column classifies customer category as “New Customers”, “Old Customers”, using if statement:

Customer Type = IF(YEAR('crm customerinfo'[BankDOJ])=2019,"New Customer", "Old Customer")

**Objective Questions:**

1. A graph on a computer screen

   Description automatically generated**What is the distribution of account balances across different regions?**

Key Insights:

* **France (311.33M)** has the highest balance among the three countries, with over 40% of the total balance.
* **Germany (300.4M)** closely follows France, with nearly 39.28% of the total balance, making it almost equal to France.
* **Spain (153.12M)** has the smallest share of the balance, with only 20.02%.

Conclusion:

As we can see, France, Germany, and Spain have a large client base. The distribution is somewhat even between France and Germany, with Spain having a significantly smaller share.

1. **Identify the top 5 customers with the highest Estimated Salary in the last quarter of the year. (SQL)**

This SQL query will look for the top 5 customers with highest estimated salary who joined in the last quarter (quarter 4) of the year.

QUERY:

**A screenshot of a computer code

Description automatically generated**

RESULT:

A screenshot of a computer

Description automatically generated

1. **Calculate the average number of products used by customers who have a credit card. (SQL)**

This SQL query will filter the average number of products used by customers based on their credit card status i.e. customers with credit card.

QUERY:

A screen shot of a computer

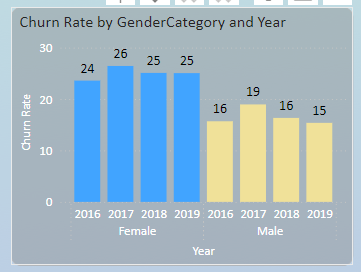
Description automatically generated

A white rectangular box with black text

Description automatically generatedRESULT:

1. **Determine the churn rate by gender for the most recent year in the dataset.**

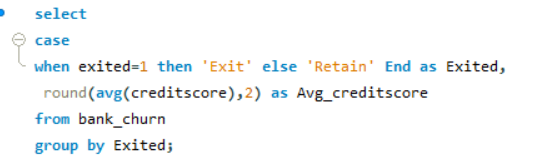
Interact with a column chart to view churn rates specifically for mala and female customers within the selected year.

This provides insights into the potential gender disparities in churn rate.

1. Compare the average credit score of customers who have exited and those who remain.

This SQL query compares the average credit score of customers who exited the bank (Exited = 1) with those who remain (Exited = 0).

QUERY:



RESULT:

A screenshot of a computer

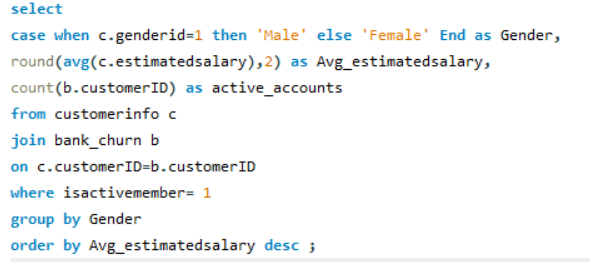
Description automatically generated

The Exit customers have higher avg. credit score than those who have remained.

1. Which gender has a higher average estimated salary, and how does it relate to the number of active accounts?

This query compares average estimated salary between genders and explores its relation to the number of active accounts

QUERY:

****

RESULT:

A screenshot of a computer

Description automatically generated

Here Female category has higher avg. estimated salary with less active accounts than male who has a less avg. estimated salary compared to female but has more active accounts than female.

1. Segment the customers based on their credit score and identify the segment with the highest exit rate.

QUERY:

A screenshot of a computer code

Description automatically generated

RESULT:

A screen shot of a computer

Description automatically generated

After identifying the credit score segment, the segment with the highest exit rate is the poor segment of credit scores.

1. Find out which geographic region has the highest number of active customers with a tenure greater than 5 years.

This query finds the geographic region with the highest number of active customers who have been with the bank for more than 5 years (tenure).

QUERY:

A screenshot of a computer program

Description automatically generated

RESULT:

A screenshot of a computer

Description automatically generated France has the highest count of active customers, with 797 having a tenure greater than 5 years.

1. What is the impact of having a credit card on customer churn, based on the available data?

A screenshot of a computer

Description automatically generatedFrom the available data, customers who have credit card have a higher churn rate than those who don’t

have a credit card. 69.91% of customers with credit card have churned the bank. This might be due to mismanagement of credit card by customers, leading to higher churn rates.

1. For customers who have exited, what is the most common number of products they had used?

A screenshot of a graph

Description automatically generatedThe most common number of products used by customers who have exited is number 1, with 1409 customers using this product.

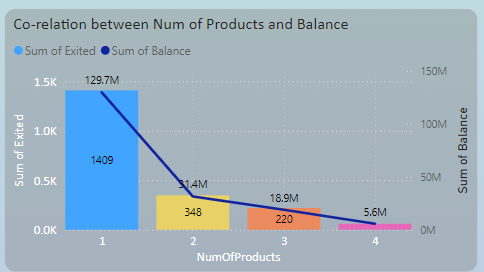
1. Examine the trend of customer joining over time and identify any seasonal patterns (yearly or monthly). Prepare the data through SQL and then visualize it.

A graph on a computer screen

Description automatically generated

From the above graph we can see that, there’s a gradual increase in number of customers joining the bank.

1. Analyse the relationship between the number of products and the account balance for customers who have exited.

Product number 1 has the highest account balance of 130M compared to other products for exited customers. Product 1 also has the highest count of exited customers.

1. Identify any potential outliers in terms of balance among customers who have remained with the bank.

A blue dot pattern on a grey background

Description automatically generated

As we can see from the above scatter plot graph:

The customers with the balance 0 are potential outliers as Outlier are the one which is abnormally have distance compared to the other value in our sample dataset

As we can see that more customer with balance 0 are more far from the customer having 50k, 100k, 150, 200k balance

3617 are the customers having 0 balance.

1. How many different tables are given in the dataset, out of these tables which table only consist of categorical variables?

Answer-

The dataset contains seven distinct tables.   
Five of the seven tables—active, exiting, credit card, gender, and geography—have categorical variables.   
Since the active customer table contains a categorical active category column. The exit category, category, gender category, and geographical location columns in the exitcustomer, creditcard, gender, and geography tables are also categorical.

1. Using SQL, write a query to find out the gender wise average income of male and female in each geography id. Also rank the gender according to the average value.

QUERY:

A screenshot of a computer code

Description automatically generated

RESULT:

A screenshot of a computer

Description automatically generated

The table provides a breakdown of average income by gender across three geographic locations: France, Spain, and Germany. In France, females have the highest average income at 99,564.25, whereas in Spain, males have the highest average income at 98,425.69. In Germany, females again lead with an average income of 102,446.42. The Gender\_Rank column indicates the ranking of average incomes within each country.

1. Using SQL, write a query to find out the average tenure of the people who have exited in each age bracket (18-30, 30-50, 50+).

QUERY:

A screenshot of a computer code

Description automatically generated

RESULT:

A screenshot of a computer

Description automatically generatedFrom the image we can conclude that people in age group 30-50 tend to stay with the bank for long term as their avg\_tenure is highest followed by 50 above with avg\_tenure of 4.83.

1. Is there any direct correlation between salary and balance of the customers? And is it different for people who have exited or not?

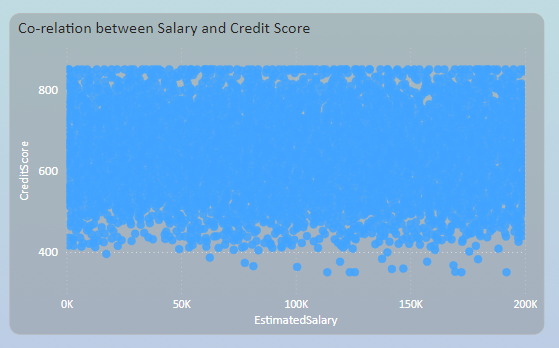
A screen shot of a graph

Description automatically generatedThis scatterplot illustrates that there is a wide distribution of salaries and balance with no strong co-relation between salary and balance.

Yes but it’s different as balance range (up to 250k)is high for customer who remains in bank as compare to customers who exited the bank have range of balance up to 220k.

1. Is there any correlation between salary and Credit score of customers?

The scatter plot shows the relationship between estimated salary and credit score for customers. The data points are densely populated across all salary ranges from 0 to 200K and credit scores from 400 to 900, indicating no strong correlation between salary and credit score. This suggests that credit scores are relatively independent of salary levels among the customers in this dataset.

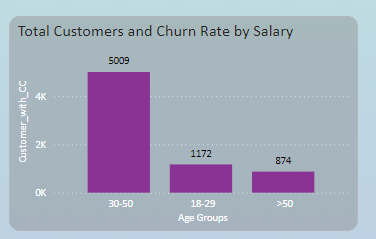


1. . Rank each bucket of credit score as per the number of customers who have churned the bank.

A screenshot of a graph

Description automatically generatedThe customers with fair credit type with churned customers 685 has been ranked 1 followed by Poor credit type with churned customers 520 has been ranked 2 followed by good credit type with 452 churned customers has been ranked 3 then Very Goof credit type ranked 4 with 252 churned customers and lastly excellent has been ranked 5 with 128 churned customers.

1. According to the age buckets find the number of customers who have a credit card. Also retrieve those buckets who have lesser than average number of credit cards per bucket.



As from the above graph:

5009 customers from the age bracket 30-50 are having credit card

1172 customers from the age bracket 18-29 are having credit card

874 customers from the age bracket 50+ are having credit card

As 18-29 and 50+ these are two age groups having less average credit card.

1. A graph of different colored squares

   Description automatically generatedRank the Locations as per the number of people who have churned the bank and average balance of the learners.

From the image,

As Germany is rank 1 as number of people who churn the bank is 814 and average balance is 120361.08

France is rank 2 as number of people who churn the bank is 810 and average balance is 71192.8

Spain is rank 3 as number of people who churn the bank is 413 and average balance is 72513.35

1. As we can see that the “CustomerInfo” table has the CustomerID and Surname, now if we have to join it with a table where the primary key is also a combination of CustomerID and Surname, come up with a column where the format is “CustomerID\_Surname”.

QUERY:

A close-up of a computer screen

Description automatically generated

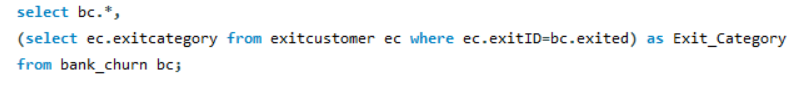
RESULT:

A screenshot of a computer

Description automatically generated

1. Without using “Join”, can we get the “ExitCategory” from ExitCustomers table to Bank\_Churn table? If yes do this using SQL.

QUERY:



RESULT:

A screenshot of a data

Description automatically generated

1. Were there any missing values in the data, using which tool did you replace them and what are the ways to handle them?

No there were no missing value in the data set.

If in case the value will be missing then there are indeed several ways to handle missing values in SQL, and different tools may have their own specific functions or methods for dealing with them. Here are some common approaches in SQL

1. Use of IS NULL or IS NOT NULL: we can filter rows where a column is NULL or is not NULL using the IS NULL and IS NOT NULL operators, respectively.

SELECT \* FROM table\_name WHERE column\_name IS NULL;

SELECT \* FROM table\_name WHERE column\_name IS NOT NULL;

1. Use of COALESCE: The COALESCE function returns the first non-NULL value in a list of expressions.

SELECT COALESCE(column\_name, replacement\_value) FROM table\_name;

1. Use of CASE statements: You can use CASE statements to replace NULL values with specific values based on conditions.

SELECT CASE WHEN column\_name IS NULL THEN replacement\_value ELSE column\_name END AS new\_column\_name FROM table\_name;

1. Write the query to get the customer ids, their last name and whether they are active or not for the customers whose surname ends with “on”.

QUERY:

A computer screen shot of a computer screen

Description automatically generatedRESULT:

A screenshot of a computer

Description automatically generated

DATA ANALYSIS AND VISUALISATION (SUBJECTIVE QUESTIONS)

1. **Customer Behaviour Analysis:** What patterns can be observed in the spending habits of long-term customers compared to new customers, and what might these patterns suggest about customer loyalty?

A graph of a bar chart

Description automatically generated with medium confidence

**Potentially Higher Spending by Older Customers:** The column bar chart illustrates that long-term clients with tenures longer than four years have a higher spending habit than new clients with tenures of three or four years. Long-term clients also have higher average estimated salaries than new clients, but they also maintain average account balances that are equal to those of new clients, indicating that long-term clients have a higher spending habit than new clients.

**Customer Loyalty:** The graph above illustrates that the most loyal customers are those who use the bank's services to the fullest, stay with the company for longer periods of time (more than four), and have average balances in their accounts. These devoted clients have confidence in the bank and even express interest in any potential new services the bank may offer.

1. **Product Affinity Study:** Which bank products or services are most commonly used together, and how might this influence cross-selling strategies?

A screenshot of a graph

Description automatically generated

As we can see from the pie chart, NumOfProducts **1** has most commonly used together among all NumOfProduct as **NumOfProducts 1** is used by 5084 customers

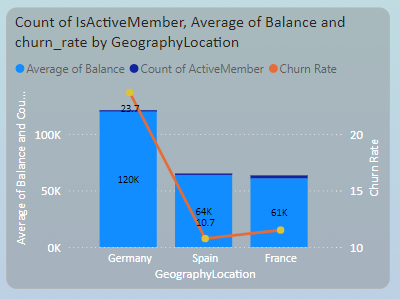
**NumOfProducts 2** is used by 4590 customers

**NumOfProducts 3** is used by 266 customers

**NumOfProducts 4** is used by 60

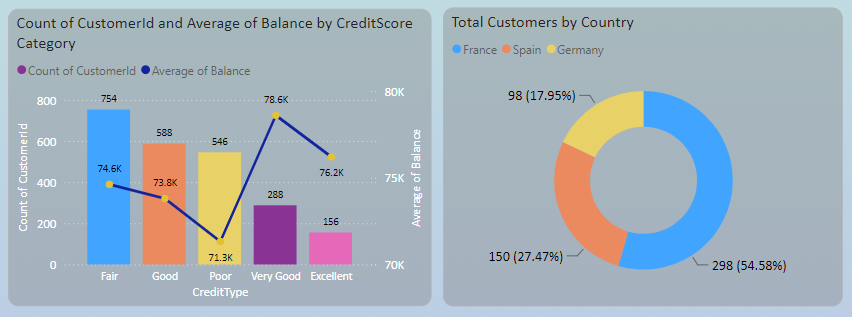
These observations may impact cross-selling strategies:   
1. Identifying Related Products: Banks can identify natural pairings or complementary offerings by analysing which bank products or services are frequently utilized when combined with one another.   
  
2. Customizing Product Bundles: Combining similar products into one offering might draw in more clients and raise the possibility of cross selling several items in one transaction.   
  
3. Personalized suggestions: By utilizing product loyalty analytics, banks are able to provide specific consumers with suggestions for products that are customized to their banking habits.   
  
4. Cross-Selling Campaigns: Banks are able to create focused cross-selling campaigns that showcase supplementary goods and services for particular client groups.

5. Improving Customer Experience: By providing smooth connection and compatibility between related products and services, banks can also improve the overall customer experience by considering product connections.

1. **Geographic Market Trends:** How do economic indicators in different geographic regions correlate with the number of active accounts and customer churn rates?

According to the stacked column table, Germany has the highest average balance of 120K, the lowest number of active accounts (1248), and the highest churn rate (23.72%).   
  
• France has the second-highest average balance of 61K, the most active accounts (2591), and the second-highest churn rate (11.50%).   
  
• Spain has the second-highest number of active accounts (1312), the lowest churn rate (10.75%), and an average balance of 64K.

1. **Risk Management Assessment:** Based on customer profiles, which demographic segments appear to pose the highest financial risk to the bank, and why?



As from above images Based on customer profile (Credit score category):

* According to the credit score category and customer profile, a customer falling into the poor credit score range (350–579) poses a significant financial risk to the bank due to their low average balance of 71.3K, non-carrying status, and inability to make timely payments on any debt they may have taken on.
* Of the bank's customers, 546 fall into the category of having low credit scores and do not possess a credit card.
* Of all customers, 298 are from France and fall into the category of having a poor credit score. This means that customers in France who do not use credit cards and have either a zero balance or a low average balance pose a financial risk to the bank.

1. **Customer Lifetime Value Forecast:** How would you use the available data to model and predict the lifetime (tenure) value of different customer segments?

A screenshot of a graph

Description automatically generated

From the above graph:

* Customer lifetime (tenure) based on age bracket, credit score, and average balance maintained.
* The customer, who is young (18–29 years old), recently opened an account with the bank, keeps an extremely low average balance, and despite this, has a low credit score.
* Customers who are older than thirty to fifty have higher tenure and stay with the bank for longer periods of time since they have good credit scores and keep high average account balances.
* When we get past the 50+ age group, these customers' credit scores decline once more, which results in shorter account tenure and decreased bank loyalty, even when they keep their average amount high.

1. **Marketing Campaign Effectiveness:** How could you assess the impact of marketing campaigns on customer retention and acquisition within the dataset? What extra information would you need to solve this?

**Answer:** • By dividing the consumer base into segments according to their behaviour, demographics, and product usage, we may examine how marketing efforts affect various customer segments. This segmentation will assist in determining which segments respond best to the campaigns and where possible changes should be made.   
•Additional data, such as detailed information on marketing campaign spend, channels employed, and messaging, may be required in order to undertake a thorough study of the effectiveness of marketing campaigns.   
• Information on customer response, such as campaign attribution, click-through and conversion rates.   
• Sentiment analysis and customer feedback from social media, reviews, and customer service exchanges.   
• External variables including the state of the economy, the level of competition, and market trends that could affect consumer behaviour.

1. **Customer Exit Reasons Exploration:** Can you identify common characteristics or trendsamong customers who have exited that could explain their reasons for leaving?

* HasCredCard: Since the bank is not providing any services, customers without credit cards are more inclined to quit.
* Tenure: Clients with tenures of three or four years are more likely to quit the bank because, as new customers, they are unsure of the services offered by the bank and find it difficult to comprehend its policies.
* NumOfProduct: According to the dataset that is currently available, the bank has NumOfProducts 1, 2, 3, and 4, so customers who are using fewer services from those four products are more likely to depart the bank than those who are using none at all.
* **IsActiveMember:** Customers who are not so active are more likely to leave bank.
* **EstimatedSalary**: Customers who have low average estimated salary are more likely to leave the bank.
* **Balance**: Customers with 0 balance are more like to leave the bank

1. Is 'Tenure', 'NumOfProducts', 'IsActiveMember', and 'EstimatedSalary' important for predicting if a customer will leave the bank?

Yes, "EstimatedSalary," "NumOfProducts," "Tenure," and "IsActiveMember" are all important factors in whether or not a consumer will leave the bank. These columns offer important information on a customer's capacity to make purchases within their projected pay, make payments on time, and keep their bank membership active.

1. Utilize SQL queries to segment customers based on demographics and account details.

QUERY:

A screen shot of a computer

Description automatically generated

RESULT:

A screenshot of a computer

Description automatically generated

1. How can we create a conditional formatting setup to visually highlight customers at risk of churn and to evaluate the impact of credit card rewards on customer retention?

• The bank can rapidly identify customers who might be at risk of churn by highlighting those who have no credit cards and no balance. This makes it possible to implement focused retention strategies to boost account activity and reduce churn. It is indicated by the colour red.

• One way to evaluate how well credit card rewards work to retain consumers is to compare the churn rates of customers who have and do not have credit cards. A reduced percentage of customer attrition among credit cardholders signifies a favourable effect on retention, providing insights for improving credit card incentive schemes and cultivating patronage.

A screenshot of a computer

Description automatically generated

1. What is the current churn rate per year and overall, as well in the bank. Can you suggest some insights to the bank about which kind of customers are more likely to churn and what are the different strategies that can be used to decrease the churn rate.

A graph with different colored bars

Description automatically generated

The above graphic displays the bank's annual churn rate as well as its overall churn rate based on the most recent data analysis.

Data to the bank regarding the kinds of clients who are most likely to leave:

• Customers who are more likely to leave the bank include those with shorter tenure, fewer products, inactive memberships, lower anticipated salaries, and no credit card.

• Tailored marketing campaigns to re-engage inactive customers, tailored offers to encourage loyalty, enhanced customer service, and proactive retention initiatives like loyalty or rewards programs are some examples of strategies to lower the churn rate. Furthermore, conducting satisfaction surveys and evaluating consumer feedback can yield insightful data about areas in need of development and assist in customizing retention methods.

1. Create a dashboard incorporating all the KPIs and visualization related metrics. Use a slicer in order to assist in selection in the dashboard.

A screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generated

A screenshot of a computer

Description automatically generated

1. How would you approach this problem, if the objective and subjective questions weren't given?

Answer- If the objective and subjective questions were not provided, and I will approach the problem of analyzing customer churn and related factors in a banking churn, I would follow a structured approach to gain insights and create a meaningful analysis. Here's how I would approach it:

• Get to know the dataset and comprehend its variables, structure, and meanings.

• Using statistical summaries and visualizations to investigate data distributions, patterns, and correlations between variables.

• Improving analysis by adding new features or changing already-existing ones.

• Create dynamic dashboards to efficiently visualize important trends and insights.

• Condense results and offer doable suggestions to lower attrition and raise client satisfaction.

1. In the “Bank\_Churn” table how can you modify the name of “HasCrCard” column to “Has\_creditcard”?

We will use following synatax by using sql queries

Alter table bank\_churn

Rename column HasCrCard to Has\_creditcard;