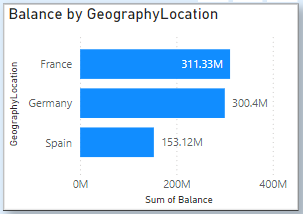
**Bank CRM Analysis Project Document**

**Objective Questions:**

**Q.1 What is the distribution of account balances across different regions?**

**Ans :-** Here we can see as per the geography location, France got the highest amount of account balance that is 311.33M and Spain with the lowest with 153.12M



**Approach :-**

Here we created Brar Chart with Currently there are three regions in dataset: France, Germany and Spain.

With the Balance as France – 311.33M, Gremany – 300.4M, Spain - 153.12M

**Q2. Identify the top 5 customers with the highest Estimated Salary in the last quarter of the year. (SQL)**

**Ans :-** In MySQL we can use following Code to get the answer

Code :-

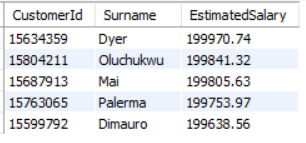
SELECT CustomerId, Surname, EstimatedSalary FROM customerinfo

WHERE QUARTER(Bank\_DOJ) = 4

ORDER BY EstimatedSalary DESC

LIMIT 5;

**Result** :-



**Approach :-**

To find top 5 customers with highest Est.Salary in Last Quarter, we use Quarter function and the sort data by Salary in descending order, then limit the rows to 5.

**Q3. Calculate the average number of products used by customers who have a credit card. (SQL)**

**Ans :-** In MySQL we can use following Code to get the answer

**Code** :-

SELECT AVG(NumOfProducts) as product\_avg

from bank\_churn

where HasCrCard = 1;

**Result** :-



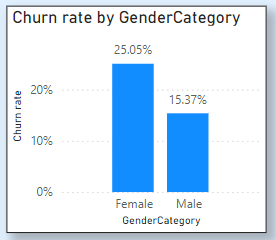
**Approach :-**

The average number of products used by customers who have a credit card = 1.53

To get this, we calculate the average NumOfProducts used by customers, after filtering data by customers who are CreditCard holders.

**Q4. Determine the churn rate by gender for the most recent year in the dataset.**

**Ans :-** Here from the diagram, we can see the churn rate as per gender category for the most recent year from the below mentioned diagram for 2019 which is 25.05% for Females and 15.37% for Males.

**Approach :-**

To find this we simple plot chart of churn rate by Gender Category and with Slicer /filter we can get insights by most recent year i.e 2019

**Q.5 Compare the average credit score of customers who have exited and those who remain. (SQL)**

**Ans :-** In MySQL we can use following Code to get the answer

Code :-

SELECT

AVG(CASE WHEN Exited = 1 THEN CreditScore ELSE Null END) as Avg\_Exited\_Score,

AVG(CASE WHEN Exited = 0 THEN CreditScore ELSE Null END) as Avg\_Remain\_Score

FROM bank\_churn;

**Result** :-



**Approach :-**

For this, we simply calculate average Credit Score of customers wile grouping them by Exit.

**Q.6 Which gender has a higher average estimated salary, and how does it relate to the number of active accounts? (SQL)**

**Ans :-** In MySQL we can use following Code to get the answer

Code :- SELECT g.GenderCategory,

ROUND(AVG(c.EstimatedSalary),2) AS Avg\_Salary,

COUNT(DISTINCT b.CustomerId ) AS Num\_Active\_Accounts

FROM bank\_churn b

JOIN customerinfo c ON b.CustomerId=c.CustomerId

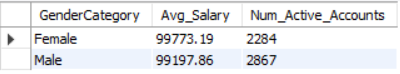
JOIN gender g ON c.GenderID=g.GenderID

WHERE IsActiveMember = 1

GROUP BY g.GenderCategory

ORDER BY Avg\_Salary DESC;

**Result** :-



**Approach :-**

Female customers have the higher average estimated salary.

No, there’s no strong correlation with number of Active members and Average estimated salary.

**Q.7 Segment the customers based on their credit score and identify the segment with the highest exit rate. (SQL)**

**Ans :-** In MySQL we can use following Code to get the answer

Code :-

WITH Customer\_Segments AS (

SELECT

b.customerid,

CASE

WHEN b.CreditScore < 580 THEN 'Deep Subprime'

WHEN b.CreditScore >= 580 AND b.CreditScore < 620 THEN 'Subprime'

WHEN b.CreditScore >= 620 AND b.CreditScore < 660 THEN 'Near Prime'

WHEN b.CreditScore >= 660 AND b.CreditScore < 720 THEN 'Prime'

ELSE 'Superprime'

END AS Credit\_Segment

FROM bank\_churn b

)

SELECT

Credit\_Segment,

ROUND(AVG(Exited)\*100,2) AS Exit\_Rate

FROM Customer\_Segments cs

JOIN bank\_churn b ON cs.customerid = b.customerid

GROUP BY Credit\_Segment

ORDER BY Credit\_Segment ASC

LIMIT 1;

**Result** :-



**Approach :-**

Credit Segment with the highest exit rate is Deep Subprime(<579) of credit score.

**Q8. Find out which geographic region has the highest number of active customers with a tenure greater than 5 years. (SQL)**

**Ans :-**

Code :- In MySQL we can use following Code to get the answer

SELECT c.GeographyID,

COUNT(c.CustomerId) AS Active\_Customers

FROM customerinfo c

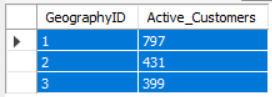
JOIN bank\_churn b ON c.CustomerId = b.CustomerId

WHERE b.IsActiveMember = 1 AND b.Tenure > 5

GROUP BY c.GeographyID

ORDER BY Active\_Customers DESC;

**Result** :-



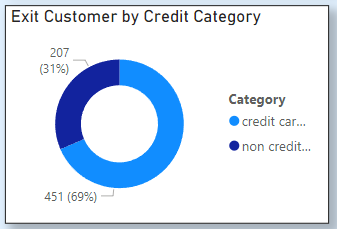
**Approach** :-

France (GeographyID = 1) has the highest number of active customers with tenure > 5 year

For this, we calculate the count of customers with tenure>5 and group by geography.

**Q.9 What is the impact of having a credit card on customer churn, based on the available data?**

**Ans :-** From data exit customers are higher in customers having credit card (69.91%) rather than not having credit card (30.09%).

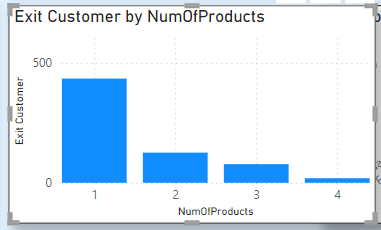


**Approach** :-

By plotting above chart we get that from data exit customers are higher in customers having credit card (69.91%) rather than not having credit card (30.09%).

**Q.10 For customers who have exited, what is the most common number of products they have used?**

**Ans** :- From that who have exited mostly they have used single product



**Approach :-**

From chart we get that for Customers who have Exited,

Most common NumOfProduct is 1

**Q11. Examine the trend of customers joining over time and identify any seasonal patterns (yearly or monthly). Prepare the data through SQL and then visualize it.**

**Ans :-**

Code:-

select year(Bank\_DOJ) as year,

count(c.CustomerId) as count\_customer\_churn

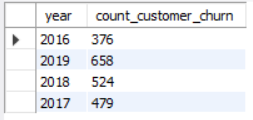
from bank\_churn b

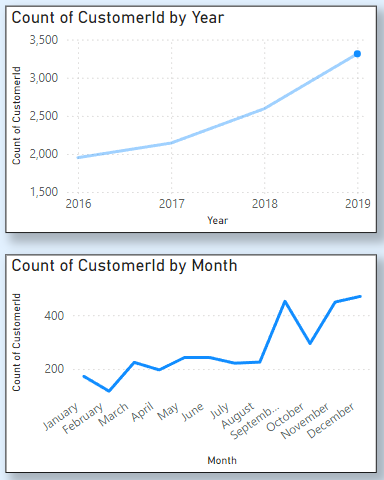
inner join customerinfo c ON b.CustomerId= c.CustomerId

where Exited= 1

group by year(Bank\_DOJ);

**Result** :-





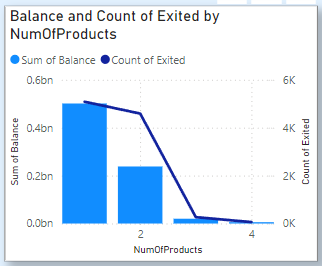
**Approach :-**

Yearly Count of customers is smoothly increasing over years, without any dramatic fluctuations.

Monthly Count of customers is overall increasing, but with some dramatic exits and the joining of customers during the last quarters of the year(Aug-Dec).

**Q.12 Analyze the relationship between the number of products and the account balance for customers who have exited.**

**Ans** :-

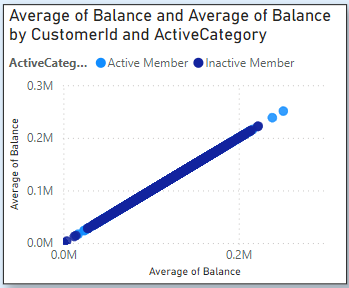


**Approach :-**

Relation that we can see between Num of Products and account balance of customers is: Sum of Balance of customers decreases with increase in Number of Products that indicates more customers use lesser number of products.

**Q.13 Identify any potential outliers in terms of balance among customers who have remained with the bank.**

**Ans** :- There are four outliers who have remained with bank.



**Approach :-**

By plotting above chart we can get that there are four outliers who have remained with bank

**Q14. How many different tables are given in the dataset, out of these tables which table only consists of categorical variables?**

**Ans** :-There are 7 different tables in dataset and among them 5 tables consists of categorical variables.

|  |
| --- |
| **Categorical Tables** |
| Gender |
| Geography |
| Active Customer |
| Exit Customer |
| CreditID |

**Q15. Using SQL, write a query to find out the gender-wise average income of males and females in each geography id. Also, rank the gender according to the average value. (SQL)**

**Ans** :-

Code :-

WITH Gender\_Income AS (

SELECT

c.GeographyID,

CASE WHEN c.GenderID = 1 THEN 'Male' ELSE 'Female' END AS Gender,

ROUND(AVG(c.EstimatedSalary),2) AS Avg\_Income

FROM customerinfo c

GROUP BY c.GeographyID, Gender

)

SELECT

GeographyID,

Gender,

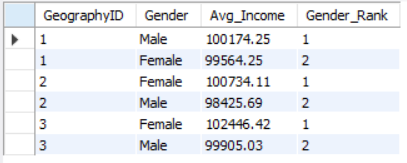
Avg\_Income,

RANK() OVER(PARTITION BY GeographyID ORDER BY Avg\_Income DESC) AS Gender\_Rank

FROM Gender\_Income

ORDER BY GeographyID, Gender\_Rank;

**Result** :-



**Approach** :-

Here from above SQL query we can get the GeoghraphyID,Gender,Avg\_Income, and Gender\_Rank

Here we use CTE ,Case When and Rank fuctions for the Insights

**Q.16 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+).**

**Ans :-**

Code :-

SELECT

CASE

WHEN c.Age BETWEEN 18 AND 30 THEN '18-30'

WHEN c.Age BETWEEN 31 AND 50 THEN '31-50'

ELSE '51+'

END AS Age\_Group,

ROUND(AVG(Tenure),2) AS Avg\_Tenure

FROM customerinfo c

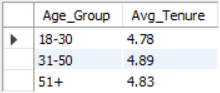
JOIN bank\_churn b ON c.CustomerId = b.CustomerId

WHERE Exited = 1

GROUP BY Age\_Group

ORDER BY Age\_Group;

**Result :-**



**Approach :-**

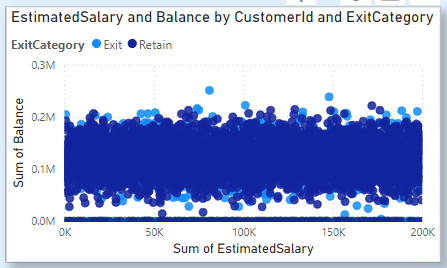
For this, we first segment ages in groups then calculate average tenure and grouping by Age\_Group

**Q.17 Is there any direct correlation between salary and the balance of the customers? And is it different for people who have exited or not?**

**Ans :-**

**Approach :-**

* No direct correlation of salary and balance is seen.
* The average balance (of all customers) lies in the range of 70K - 150K, despite salary difference.
* As for the exited customers, the correlation is still negligible.
* The average balance (of exited customers) lies in the range of 90K - 150K, despite salary difference.

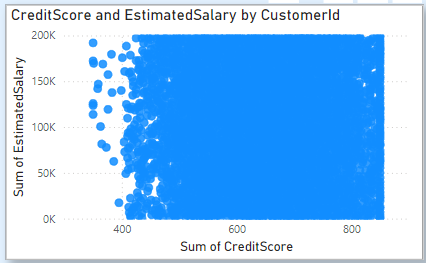


**Q.18 Is there any correlation between the salary and the Credit score of customers?**

**Ans :-**

**Approach :-**

From following scatter plot it is clear that there is no correlation between salary and credit score



**Q.19 Rank each bucket of credit score as per the number of customers who have churned the bank.**

**Ans :-**

**Approach :-**

As per below graph

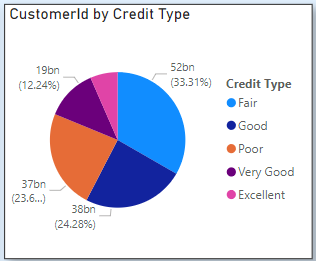
33.31% customers has fair credit score.

24.28% customers Good fair credit score.

23.62% customers has Poor credit score.

12.24% customers has Very Good credit score.

6.55% customers has Excellent credit score.

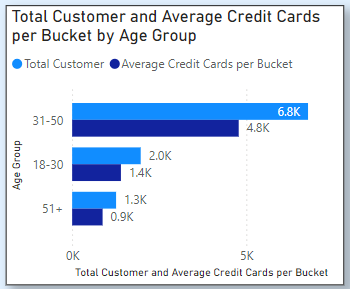


**Q.20 According to the age buckets find the number of customers who have a credit card. Also retrieve those buckets that have lesser than average number of credit cards per bucket.**

**Ans** :-

**Approach** :-

According to the age buckets the number of customers who have a credit card. And buckets those have lesser than number of credit cards per bucket. For age bucket 18-30 total customers are 1968 average number of credit cards per bucket is 1400. For age bucket 31-50 total customers are 6771 average number of credit cards per bucket is 4781. For age bucket 50 and above total customers are 1261 average number of credit cards per bucket is 874.

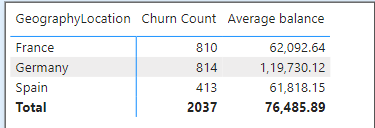


**Q.21 Rank the Locations as per the number of people who have churned the bank and average balance of the customers.**

**Ans** :-

**Approach :-**

As per following table Germany is at 1st rank. 2nd one is France and last is Spain.



**Q.22 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”.**

**Ans** :-

Code :-

select

c.CustomerId,

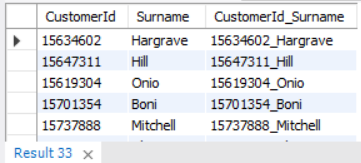
c.Surname,

concat(c.CustomerId,'\_',c.Surname) as CustomerId\_Surname

from customerinfo c

join bank\_churn b on c.CustomerId = b.CustomerId;

Result :-



**Approach :-**

For the above table we get the CustomerID\_Surname by simply using Concat Function in SQL workbench.

**Q.23 Without using “Join”, can we get the “ExitCategory” from ExitCustomers table to Bank\_Churn table? If yes do this using SQL.**

**Ans :-**

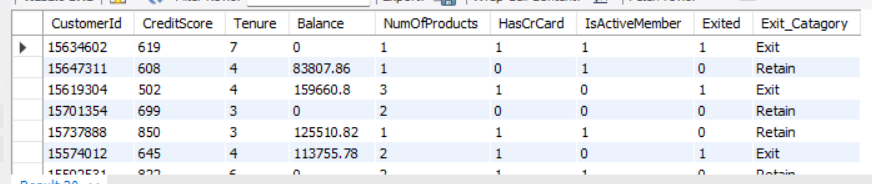
Code :-

SELECT \*,

CASE WHEN Exited = 0 THEN 'Retain' ELSE 'Exit' END AS Exit\_Catagory

FROM bank\_churn;

Result :-



**Approach :-**

Here we get above table using Case When Statement without using join abd created Exit\_Catagory column in Table.

**Q.24 Were there any missing values in the data, using which tool did you replace them and what are the ways to handle them?**

**Ans :-** We are fortunate to have a dataset with no missing values, avoiding the need for imputation techniques that might introduce biases.

### **Common Approaches to Handling Missing Values:**

1. **Deletion**: Remove rows or columns with missing values. Suitable if minimal and doesn't significantly impact the analysis but can lead to information loss.
2. **Imputation**: Replace missing values with estimates, such as mean, median, mode, or k-Nearest Neighbors (KNN). The method should match the data type and distribution.
3. **Modeling Techniques**: Use statistical models that handle missing values directly, while understanding the reasons for the missing data remains important.

**Q.25 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”.**

**Ans :-**

Code:-

SELECT c.CustomerID, c.Surname AS last\_name,

CASE WHEN b.IsActiveMember = 1 THEN 'Active' ELSE 'Inactive' END AS Active\_status

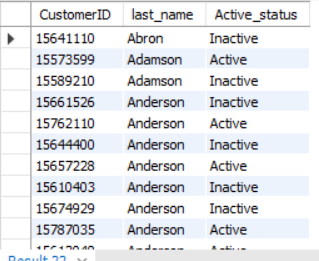
FROM customerinfo c

JOIN bank\_churn b ON c.CustomerId=b.CustomerId

WHERE c.Surname LIKE '%on'

ORDER BY c.Surname;

Result :-



**Approach :-**

For this, we will select customers with a filter (‘where’ clause in SQL) of their surname that ends with “on” and their active status.

Not all customers are shown.

Limited sample output has been shown due to large number of output rows (i.e. large number of customers have surname ending with “on”)

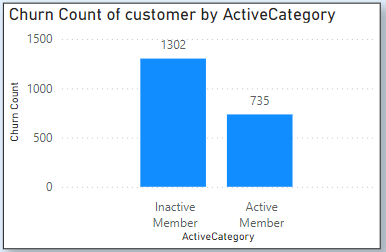
We can see all customers in SQL Workbench after running the query.

**Q.26 Can you observe any data disrupency in the Customer’s data? As a hint it’s present in the IsActiveMember and Exited columns. One more point to consider is that the data in the Exited Column is absolutely correct and accurate.**

**Ans** :-

**Approach** :-

The graph below highlights a clear discrepancy in the Customers data. It shows that 735 customers who have already exited the bank are still indicated as "Active Member." This is a substantial error, as logically, churned customers should be inactive.



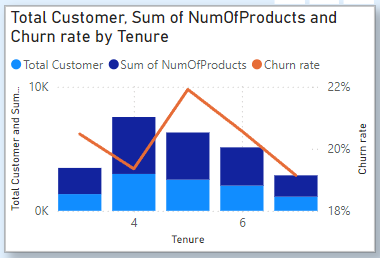
**Subjective Question:**

**Q.1 Customer Behavior 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?**

Ans :-

Approach :-

From data it is Observed that Customers with Tenure 4 has spend more and customers with tenure 5 has higher churn rate and loyalty of customers is proportional to tenure, so with higher tenure customers are more loyal.

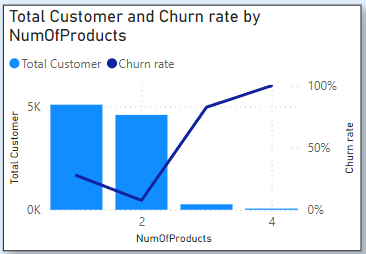


**Q.2 Product Affinity Study: Which bank products or services are most commonly used together, and how might this influence cross-selling strategies?**

**Ans :-**

**Approach :-**

From data it is been observed that Customers with less no of products has less churn rate compare to the customers who bought more no of products.

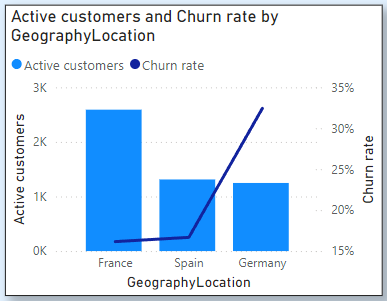


**Q.3 Geographic Market Trends: How do economic indicators in different geographic regions correlate with the number of active accounts and customer churn rates?**

**Ans :-**

**Approach:-**

From graph it is clear that France has more active customers and churn rate is max in Germany.

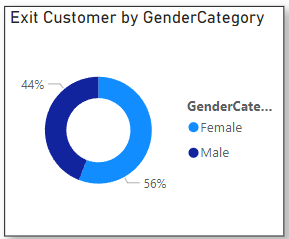
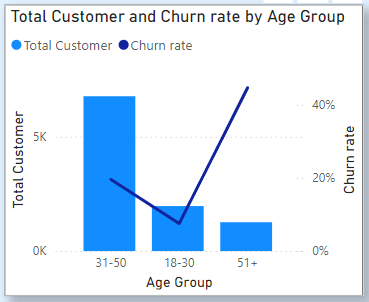


**Q.4 Risk Management Assessment: Based on customer profiles, which demographic segments appear to pose the highest financial risk to the bank, and why?**

**Ans :-**

**Approach :-**

From data it is clear that customer churn rate is proportional to the customer age. Female has higher churn rate compared to male.



**Q.5 Customer Tenure Value Forecast: How would you use the available data to model and predict the lifetime (tenure) value in the bank of different customer segments?**

**Ans :-**

* Provide custom offers and rewards tailored to customer behavior and preferences to promote retention.
* Establish loyalty programs that reward customers for their continued engagement and persuade them to stay with the bank.

**Q.6 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?**

**Ans :-**

Enhance customer service by providing personalized help, quickly resolving issues, and responding to feedback.

The marketing team should give special offers and better security to customers over 50. They should also provide more incentives to those who buy many products and to credit card holders.

Customer Segmentation:

Group customers by age, location, and number of products used to better analyze campaign results.

Trend Analysis:

Track changes in active customers, exit rates, and product usage over time to see how marketing campaigns affect these trends.

Control Groups:

If possible, use groups not exposed to specific campaigns to compare their behavior with those who were, helping to identify the campaign's impact.

Extra Information for Better Analysis:

Campaign Details & Timing: Knowing what the campaigns were about, where they were run, and when they started helps link them to changes in customer behavior.

Customer Acquisition Channel: Understanding how customers were first acquired helps see if campaigns are keeping these customers, whether they came from referrals, ads, etc.

Customer Lifetime Value: Calculating the long-term value of customers helps assess how campaigns impact overall revenue and retention.

Adding this extra information to data helps move from just seeing correlations to understanding cause and effect. This helps you:

Optimize Campaigns:

Identify which campaigns work best for different customer groups and adjust strategies accordingly.

Measure ROI:

Compare the cost of acquiring customers to their long-term value to measure the return on investment for marketing campaigns.

Conclusion:

The current data is a good start for evaluating marketing campaigns. Adding details about the campaigns, how customers were acquired, and their long-term value gives a deeper understanding of their impact. By grouping customers, analyzing trends, and using control groups,we can get valuable insights. A data-driven approach combining current and additional information will help us improve campaigns, maximize ROI, and drive long-term growth.

**Q.7 Customer Exit Reasons Exploration: Can you identify common characteristics or trends among customers who have exited that could explain their reasons for leaving?**

**Ans :-**

Customers over 50 are dissatisfied with the bank's terminology and schemes. Additionally, customers with credit scores below 700 are also leaving. Germany and France have high churn rates.

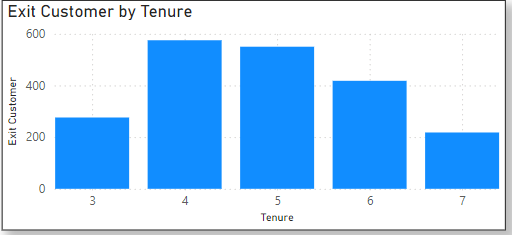
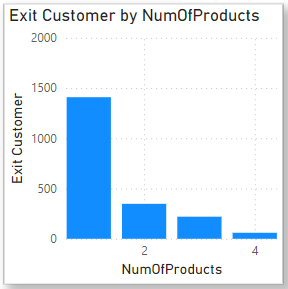
Low customer satisfaction can result in churn. Common reasons for customers leaving banks include dissatisfaction with customer service, long wait times, and unresolved issues.

**Q.8 Are 'Tenure', 'NumOfProducts', 'IsActiveMember', and 'EstimatedSalary' important for predicting if a customer will leave the bank?**

**Ans :-**

From given dataset Customers with tenure 4 and 5 are more likely to exit.

Customers with less no of products are more likely to exit and we can not predict it from estimated salary.

**Approach :-**

Here we created the column charts to get insights from Exit Customer by Tenure and Exit Customer by NumOfProducts

**Q.9 Utilize SQL queries to segment customers based on demographics and account details.**

**Ans :-**

Code :-

SELECT g.GeographyLocation,

CASE WHEN EstimatedSalary < 50000 THEN 'Low'

WHEN EstimatedSalary < 100000 THEN 'Medium'

ELSE 'High'

END AS Income\_Segment,

CASE WHEN GenderID = 1 THEN 'Male'

ELSE 'Female'

END AS gender, age,

COUNT(c.CustomerId) as Number\_of\_customers

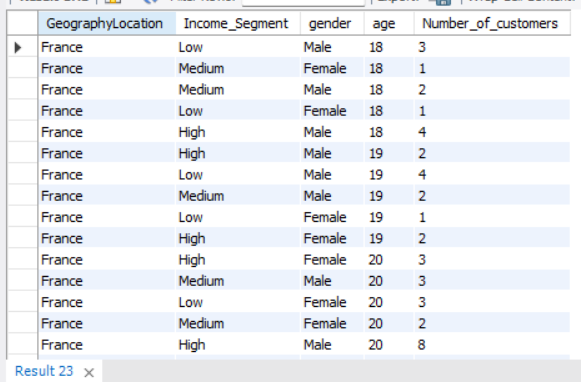
FROM customerinfo c

JOIN geography g on c.GeographyID= g.GeographyID

GROUP BY Income\_Segment,g.GeographyLocation,gender,age

ORDER BY g.GeographyLocation,age;

Result :-



**Approach :-**

Here in above table we use case when statement to get segment customers based on demographics and account details.

**Q.10 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?**

**Ans :-**

**Approach :-**

Created Churn Risk Score: We developed a calculated column based on three key factors:

Churn Risk Score (CRS) = Balance / (Credit Score \* NumOfProducts)

This indicates that churn risk is directly proportional to balance and inversely proportional to credit score and number of products.

Note: Customers with a balance of 0 will have a CRS of 0.

Created the table visual: Constructed a table incorporating factors such as:

* Number of products
* Credit scores
* Balance
* Tenure
* Other account activities

Conditional Formatting:

* Select the churn risk score column in your table.
* Right-click and choose Conditional formatting.
* Highlight risk levels using background colours:
* Low risk (score below a defined threshold)
* Moderate risk (score within a specific range)
* High risk (score exceeding a defined threshold)
* Too low value/Null to be assessed.

Benefits:

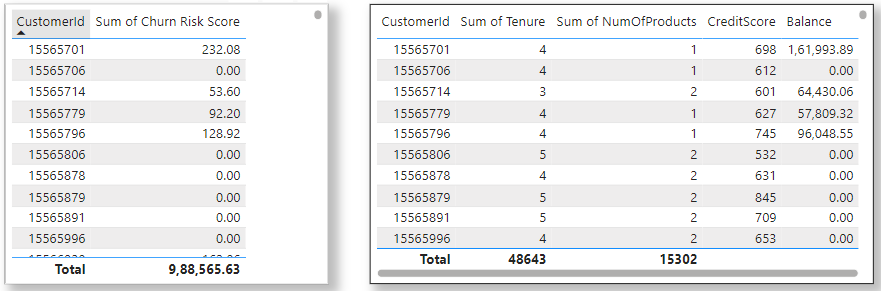
This colour-coded table instantly identifies at-risk customers, enabling you to prioritize retention efforts. It acts as a heatmap within your table, directing your attention to potential churners.

Remember:

● Evaluation should focus on active customers.

● Periodically adjust thresholds and colours based on data and risk assessments.

● Explore additional formatting options for more visual cues as our data grows.



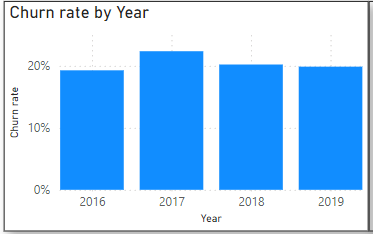
**Q.11 What is the current churn rate per year and overall as well in the bank? Can you suggest some insights to the bank about which kind of customers are more likely to churn and what different strategies can be used to decrease the churn rate?**

**Ans :-**

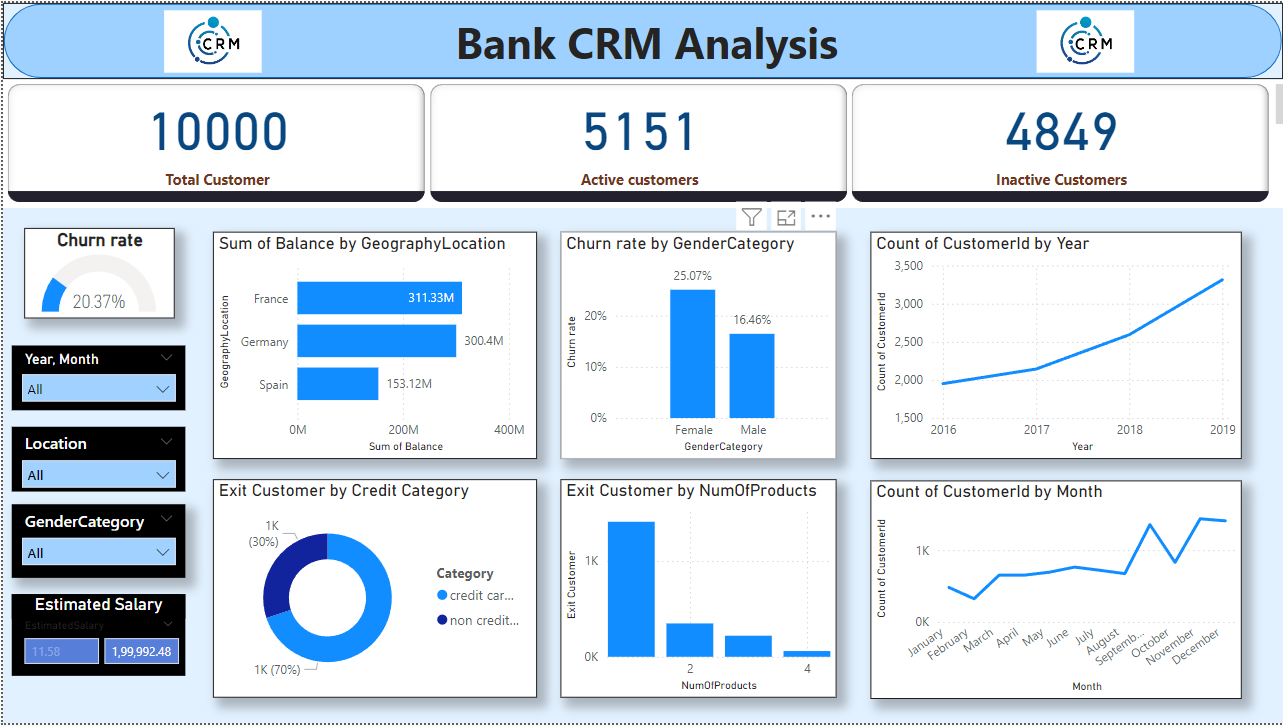
**Approach :-**

Overall churn rate is 20.47 % . In year 2016- 19.27%, 2017- 22.35%, 2018- 20.21% and in 2019- 19.86%.

Customers above age 50 and customers having less no of product are more likely to churn. So if bank provides some good rewards, Offers and securities to customers from this category they might stay with bank for long period.



**Q.12 Create a dashboard incorporating all the KPIs and visualization-related metrics. Use a slicer in order to assist in selection in the dashboard.**

**Ans :-** 

**Q.13 How would you approach this problem, if the objective and subjective questions weren't given?**

**Ans :-**

**Approach:-**

Approaching a data analysis problem without specific questions requires a flexible and exploratory approach. Here's how to do it:

1. Understand the Context: Learn about the purpose and goals of the data analysis by talking to stakeholders or reviewing project documents.

2. Explore the Data: Get to know the dataset by examining its structure and looking for patterns, anomalies, or trends.

3. Identify Potential Insights: Think about potential insights or interesting points based on your data exploration.

4. Generate Hypotheses: Create guesses or theories based on your observations that you can test further.

5. Iterative Analysis: Use different analytical techniques to test and refine your hypotheses, such as running queries and analyzing data in Excel, PowerBI, and SQL.

6. Visualize and Interpret: Use charts and graphs to show your findings clearly and uncover patterns.

7. Synthesize Findings: Combine your findings into clear themes or stories that explain the data.

8. Seek Validation and Feedback: Get feedback from peers, experts, or stakeholders to confirm your findings and ensure they meet expectations.

9. Iterate and Refine: Keep refining your analysis based on new insights or feedback.

10. Document and Communicate: Clearly document your process and findings, and effectively share your results with stakeholders, highlighting key insights and recommendations.

By following these steps, you can systematically explore and analyze data, even without specific questions, to gain valuable insights and drive better decision-making.

**Q.14 In the “Bank\_Churn” table how can you modify the name of the “HasCrCard” column to “Has\_creditcard”?**

**Ans** :-

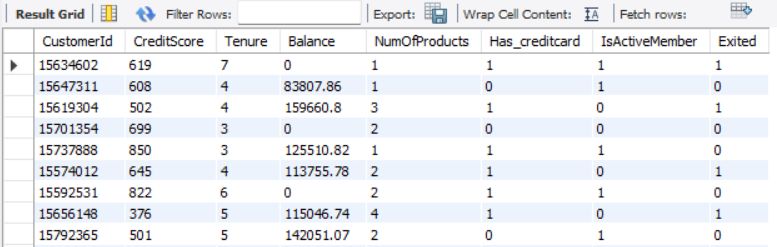
Code :-

ALTER TABLE bank\_churn

RENAME COLUMN HasCrCard TO Has\_creditcard;

SELECT \* FROM bank\_churn;

Result :-



**Approach** :-

This will modify the name of the “HasCrCard” column to “Has\_creditcard” column.