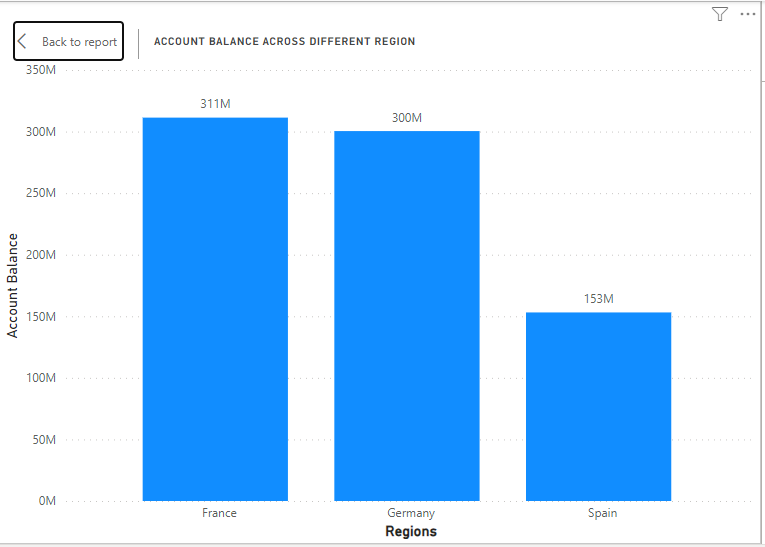
**Learners have to come up with a Report to support the answers to the following questions and suggestions**

**Objective Questions:**

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

**ANS** - France has the highest account balance of 311 M followed by Germany with 300 M and then Spain has 153 M.



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

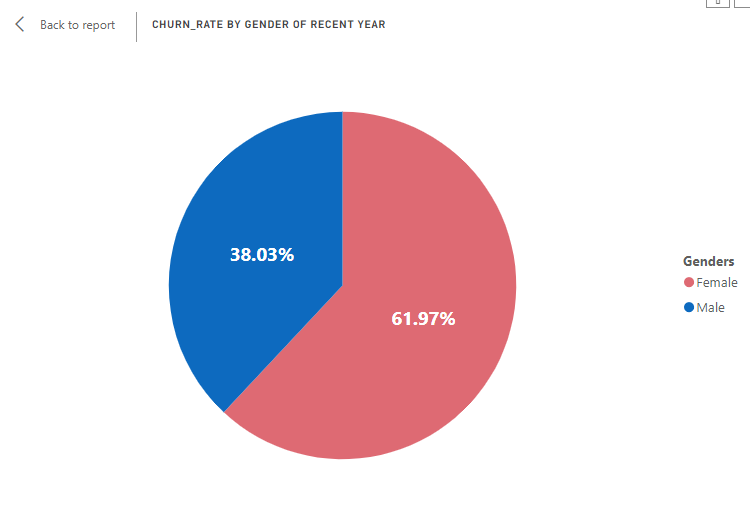
**ANS** – CustomerId – 15804211 , 15763065, 15687913, 15634359, 15599792

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

**ANS** – avg(NumOfProducts) - '1.5314'

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

ANS – The Churn rate of females is more (62%) than Males (38%) in recent years 2019



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

ANS - AverageCreditScoreExited --'645.3515', AverageCreditScoreremain -- '651.8532'

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

GenderID, GenderCategory, AverageEstimatedSalary, ActiveAccounts

2 Female 100601.54138234643 2284

1 Male 99664.57693054793 2867

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

ANS - CreditScoreSegment, TotalCustomers, ExitedCustomers, ExitRate

Low 632 150 23.73

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

GeographyID, GeographyLocation, CUSTOMER\_COUNT

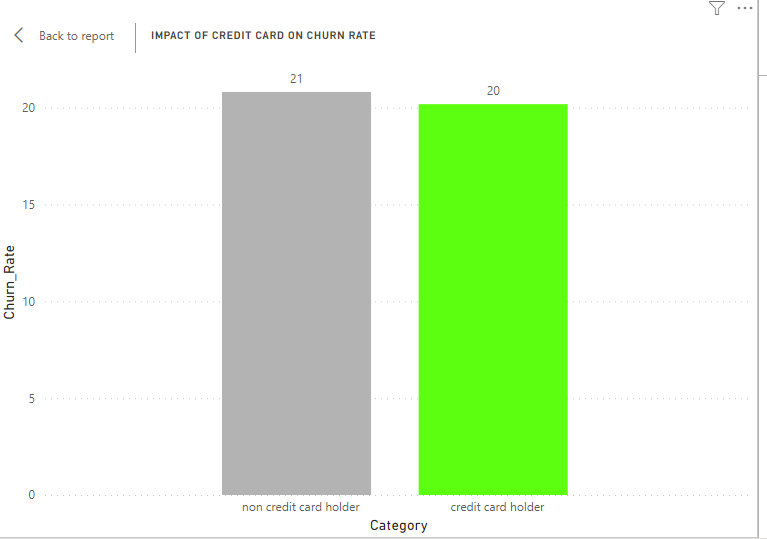
1 France 797

2 Spain 431

3 Germany 399

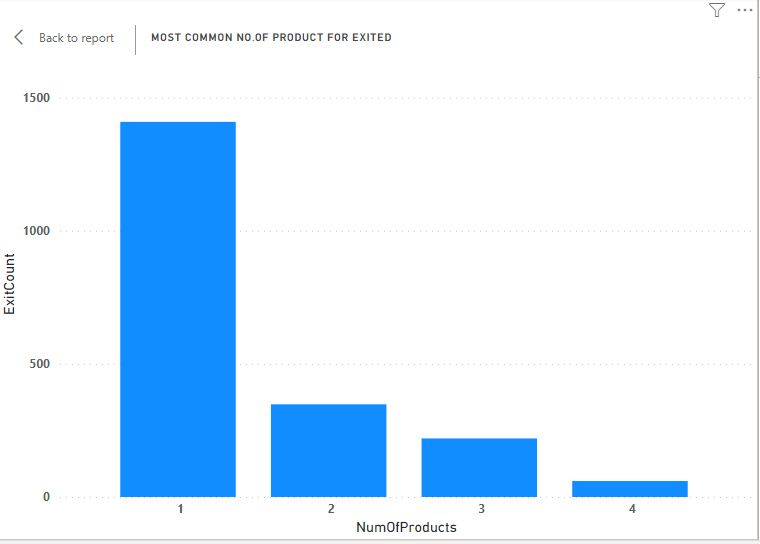
1. What is the impact of having a credit card on customer churn, based on the available data?

The churn rate of credit card holders is lower than the churn rate of non-credit card holders.



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

Product 1 is the most commonly used by exited customers



1. 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.

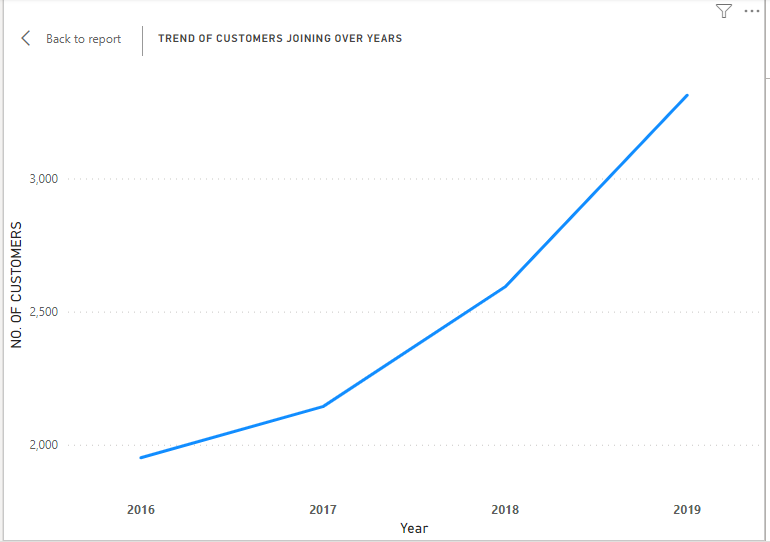
JoiningYEAR, Customer\_count

2019 3313

2018 2593

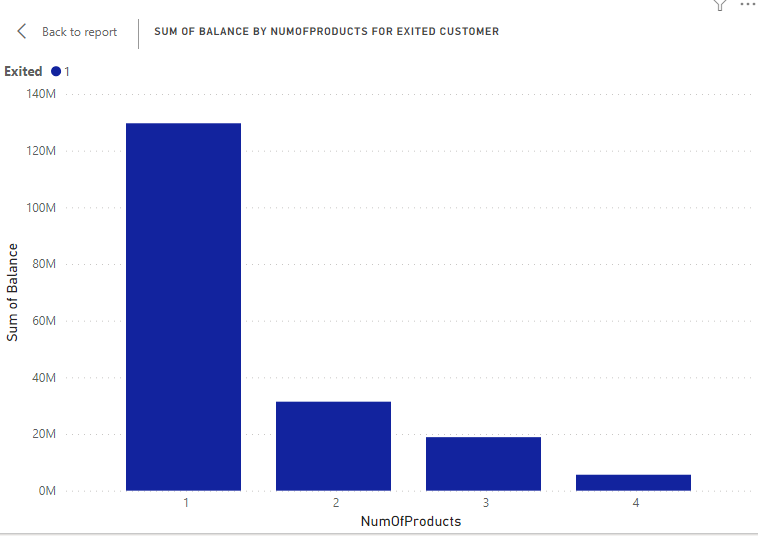
2017 2143

2016 1951



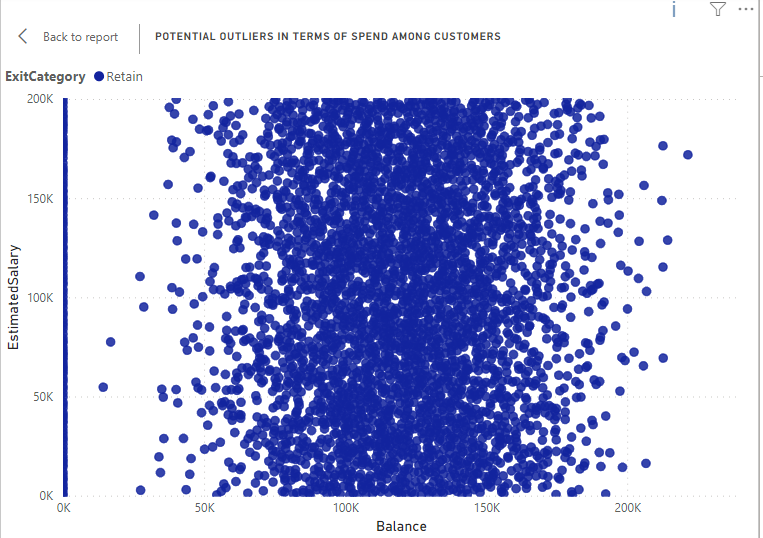
1. Analyze the relationship between the number of products and the account balance for customers who have exited.

Balance is highest for the customer who exited and used a number of products 1.



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

We can find outliers with a scatter plot chart that shows the distribution of salaries.



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

There are 7 different tables in the data set, the tables that consist of only categorical values are

1). ActiveCustomer 2).CreditCard 3). ExitCustomer 4). Gender 5).Geography

1. 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)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| GeographyID | GeographyLocation | Gender | AverageIncome | GenderRank |
| 1 | France | Male | 100174.3 | 1 |
| 1 | France | Female | 99564.25 | 3 |
| 2 | Spain | Female | 100734.1 | 2 |
| 2 | Spain | Male | 98425.69 | 3 |
| 3 | Germany | Female | 102446.4 | 1 |
| 3 | Germany | Male | 99905.03 | 2 |

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+).

AgeBracket, AverageTenure

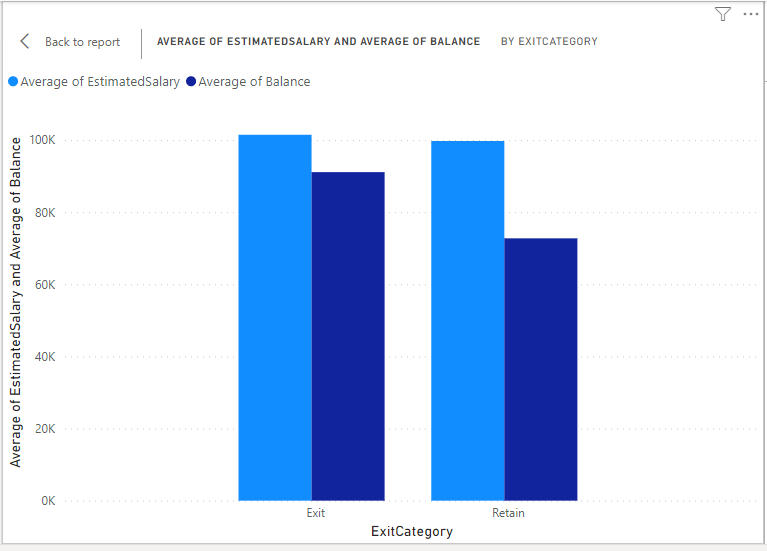
18 TO 30 4.7770

31 TO 50 4.8899

50 ABOVE 4.8330

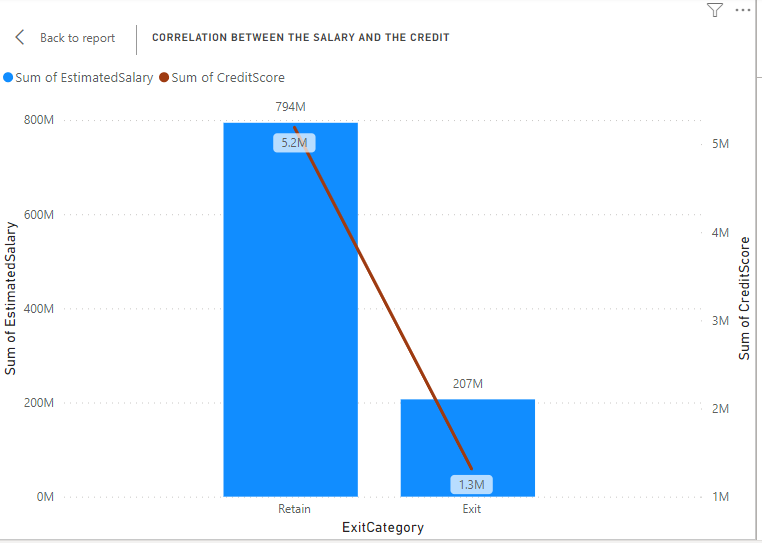
1. Is there any direct correlation between salary and the balance of the customers? And is it different for people who have exited or not?

People who exited have the highest average salary and highest average balance compared to the people who retain



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

ANS – people with the highest salary have the highest credit score it directly proportional and people who retain has the highest than exited



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

We get the same visual in power bi also using table visualization as we get here using SQL query

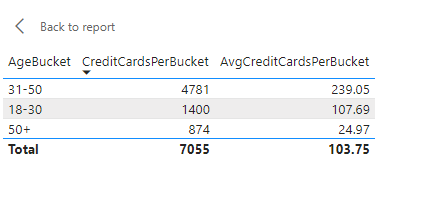
CreditScoreBucket, NumberOfCustomersChurned, CreditScoreRank

501-700 1266 1

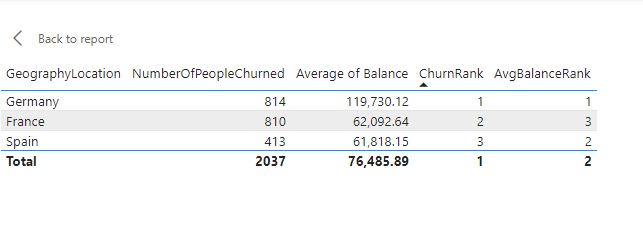
701-850 619 2

301-500 152 3

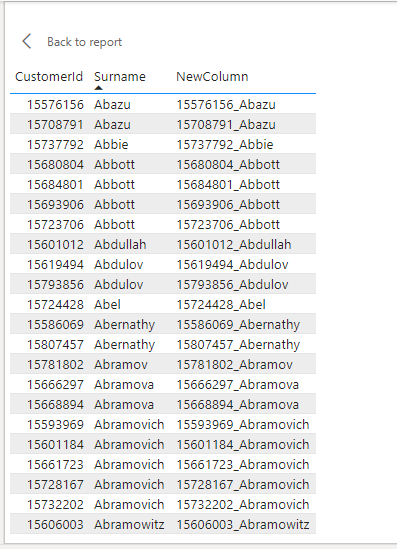
1. According to the age buckets find the number of customers who have a credit card. Also retrieve those buckets who have lesser than average number of credit cards per bucket.



1. Rank the Locations as per the number of people who have churned the bank and average balance of the learners.



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”.



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

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **CustomerId** | **CreditScore** | **Tenure** | **Balance** | **NumOfProducts** | **HasCreditCard** | **IsActiveMember** | **Exited** | **ExitCategory** |
| 15634602 | 619 | 7 | 0 | 1 | 1 | 1 | 1 | Exit |
| 15647311 | 608 | 4 | 83807.86 | 1 | 0 | 1 | 0 | Retain |
| 15619304 | 502 | 4 | 159660.8 | 3 | 1 | 0 | 1 | Exit |
| 15701354 | 699 | 3 | 0 | 2 | 0 | 0 | 0 | Retain |
| 15737888 | 850 | 3 | 125510.82 | 1 | 1 | 1 | 0 | Retain |
| 15574012 | 645 | 4 | 113755.78 | 2 | 1 | 0 | 1 | Exit |
| 15592531 | 822 | 6 | 0 | 2 | 1 | 1 | 0 | Retain |
| 15656148 | 376 | 5 | 115046.74 | 4 | 1 | 0 | 1 | Exit |
| 15792365 | 501 | 5 | 142051.07 | 2 | 0 | 1 | 0 | Retain |
| 15592389 | 684 | 5 | 134603.88 | 1 | 1 | 1 | 0 | Retain |
| 15767821 | 528 | 7 | 102016.72 | 2 | 0 | 0 | 0 | Retain |
| 15737173 | 497 | 7 | 0 | 2 | 1 | 0 | 0 | Retain |
| 15632264 | 476 | 3 | 0 | 2 | 1 | 0 | 0 | Retain |
| 15691483 | 549 | 6 | 0 | 2 | 0 | 0 | 0 | Retain |
| 15600882 | 635 | 6 | 0 | 2 | 1 | 1 | 0 | Retain |

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

ANS- In the data cleaning process we used replace error in the transform group tab in the power query editor there I removed the error from the table.

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”.

|  |  |  |
| --- | --- | --- |
| **CustomerId** | **Surname** | **IsActiveMember** |
| 15700255 | Robson | 0 |
| 15619955 | Bevington | 0 |
| 15787835 | Simpson | 1 |
| 15708916 | Paterson | 0 |
| 15575581 | Dickson | 0 |
| 15612525 | Preston | 0 |
| 15648800 | Paterson | 0 |
| 15692132 | Wilkinson | 1 |
| 15673481 | Morton | 0 |
| 15725141 | Whiddon | 0 |
| 15736112 | Walton | 1 |
| 15685226 | Morrison | 1 |
| 15788218 | Henderson | 1 |
| 15759381 | Johnson | 1 |
| 15622834 | Stevenson | 0 |

**Subjective Question:**

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**

**1. Tenure Classification:**

1. Long-term customers have a tenure time greater than 4 because Information is not available for customers so I take average

2. Long-term customers spend more than new customers.

3. There's a spending disparity between long-term and new customers, with long-term customers having higher spending habits.

**2. Churn Rate Observation:**

1. The churn rate is higher in long-term customers compared to new customers.

2. Long-term customers are experiencing a higher rate of churn than new customers, indicating potential challenges in retaining these customers**.**

**3. Product Usage and Balance:**

1. Both long-term and new customers predominantly use a single product, with a balance associated with this product.

2. The number of products used does not seem to significantly differ between the two customer categories.

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

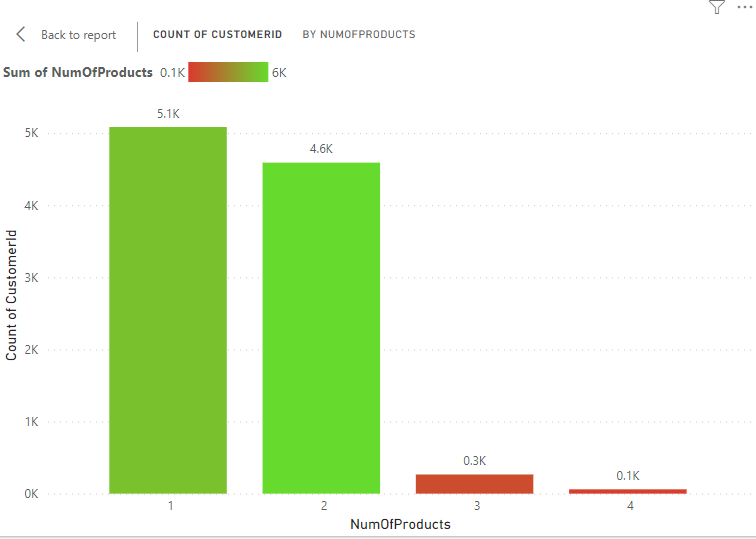
ANS - the analysis reveals that the majority of customers primarily utilize one bank product or service, with a smaller proportion opting for two services. Specifically, most customers are observed to use a single product or service, while a subset of customers engage with two distinct services.

**Cross-selling strategies**

1. leveraging insights from the distribution of product usage among customers, the bank can design more effective cross-selling strategies that are tailored to individual preferences, thereby increasing customer satisfaction, loyalty, and overall profitability of the bank's product portfolio.

2. By maintaining regular contact and providing relevant updates, the bank can position itself as a trusted financial partner and advisor, making customers more inclined to explore additional offerings.

3. For customers using a single service, personalized offers and recommendations can be provided based on their transaction history, demographics, and behavioral patterns to encourage the adoption of additional products.

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

**ANS –**

**1. Germany's Churn Rate:**

1. Observation: Germany has the highest churn rate at 32%.

2. Insight: Customers from Germany are churning at a higher rate compared to other countries.

**Germany's Active Customers:**

1. Observation: Germany has fewer active customers compared to other countries.

2. Implication: The lower number of active customers might contribute to the higher churn rate in Germany.

**3. France's Active Customers:**

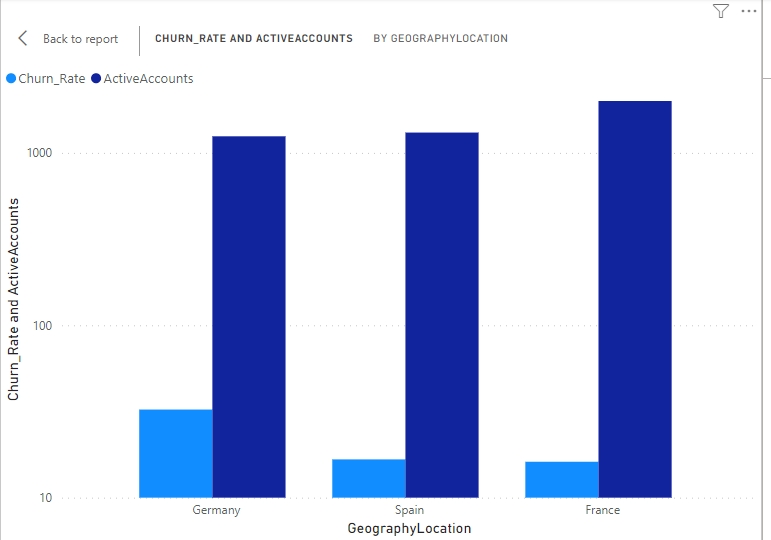
1. Observation: France has the highest count of active customers, totaling 2591.

2. Insight: France, with a strong presence of active customers, exhibits the lowest churn rate.

**4. Relationship Between Active Accounts and Churn Rate:**

1. Observation: There appears to be an inverse relationship between active account counts and churn rates.

Insight: As the count of active customers increases (e.g., in France), the churn rate tends to decrease. This suggests that a stronger presence of active accounts may contribute to lower churn rates.



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

**1. Financial Risk Metrics:**

1. Metrics: High credit utilization, low credit scores, etc.

2. High credit utilization is identified as a financial risk metric, indicating a potential risk of financial loss for the bank.

**Gender-Based Insights:**

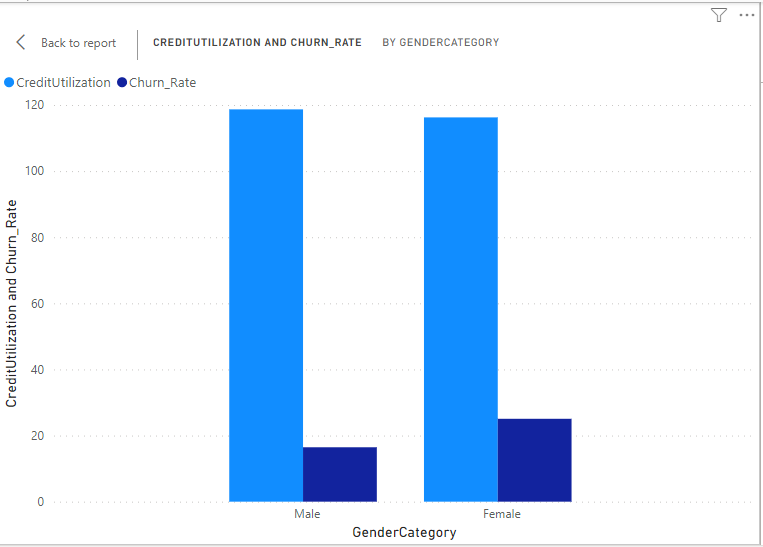
1. Observation: Churn rates are nearly equal for both genders.

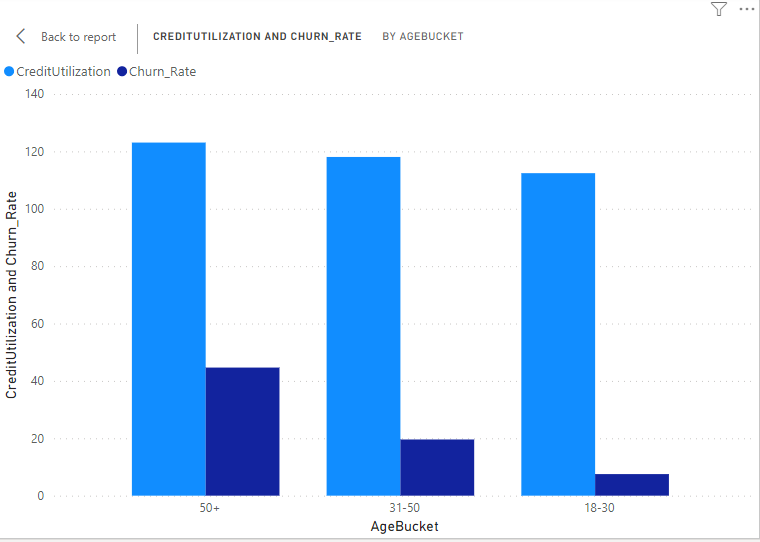
2. Insight: Credit utilization is higher for females compared to males. This suggests that while churn rates are similar, there may be a gender-specific financial risk associated with higher credit utilization among females.

**Age-Based Insights:**

1. Observation: Churn rate and credit utilization are higher for the age category of 50 and above.

2. Insight: This age category is identified as falling under financial risk to the bank. The higher churn rate and credit utilization in this category may indicate potential financial challenges and increased risk for bank**.**





1. 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?

**1. Tenure Value Forecast Based on Age Category and Gender:**

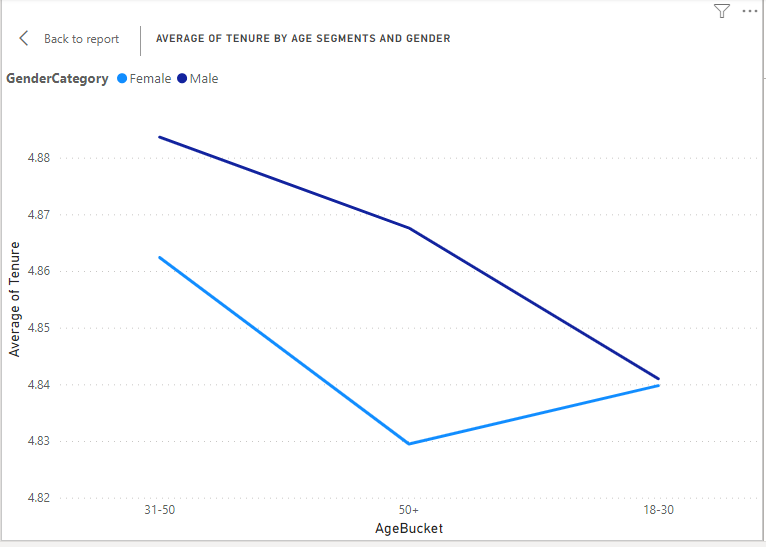
1. Observation: The average tenure time is highest among males in the age group of 31 to 50.

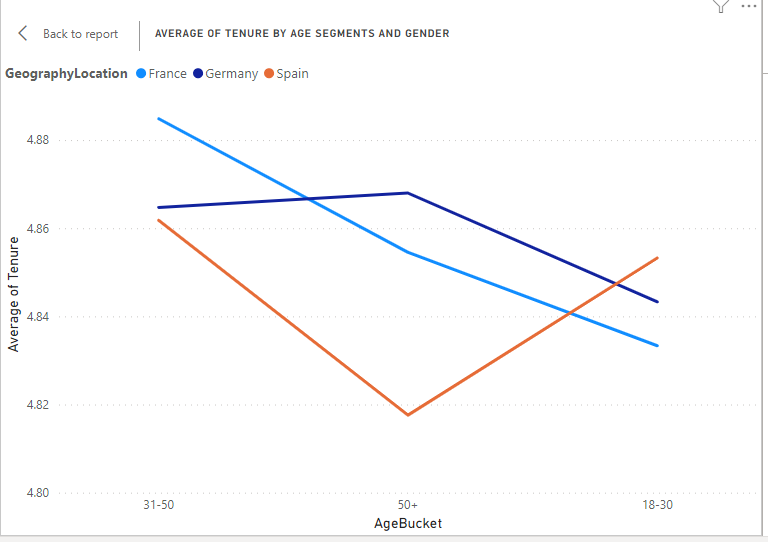
2. Insight: This suggests that, on average, male customers in the age range of 31 to 50 tend to have a longer tenure with the company compared to other age categories and genders.

**2. Tenure Value Forecast Based on Age Category and Location:**

1. Observation: The average tenure time is highest in France for the age group of 31 to 50.

2. Insight: Customers in the age group of 31 to 50 located in France, on average, have the longest tenure compared to other age categories and locations.





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?

We can identify impact using various segments like churn rate, age group, locations, and retention to see that we need some more information

1. Details about marketing campaigns: Type, duration, channels used, etc.

2. Customer interaction data during the campaign period.

3. Information on any changes in products or services during the analysis period.

4. Feedback or survey data from customers regarding their experience with the bank and its marketing efforts.

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

**1. Gender Disparity**:

More females tend to exit than males.

**2. Single Product Usage:**

Those who exit are primarily using only one product.

**3. Age-Related Churn:**

The churn rate is higher for individuals aged 50 and above.

**4. Trend in ExitCount**:

ExitCount has shown an upward trend since 2016, with a significant increase of 75.00% (282 exits) in the last three years.

**5. Churn Rate Discrepancy Between Customer Categories:**

Churn\_Rate is higher for Older Customers (age category 50 and above), specifically mentioned as 21, compared to New Customers (age category 20), which is denoted as 20.

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

**1. Number of Products:**

Customers using only one product have the highest chance of leaving, as most exited customers fall into this category**.**

**2. Tenure:**

Tenure is crucial for understanding customer loyalty. It helps differentiate between older and new customers, providing insights into which category has the highest churn rate.

**3.IsActiveMember:**

The activity level of customers, as indicated by 'IsActiveMember,' is inversely proportional to the churn rate. Higher activeness suggests a lower churn rate for that category.

**4. Estimated Salary:**

Estimated salary is identified as an important factor. The average estimated salary for both older and new customers is approximately the same, potentially influencing the churn rate**.**

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

I used Geographic location as a customer segment for demographics and account details I used the isactivemember count

|  |  |  |  |
| --- | --- | --- | --- |
| GeographyID | GeographyLocation | number\_of\_customers | active\_count |
| 1 | France | 5014 | 2591 |
| 2 | Spain | 2477 | 1312 |
| 3 | Germany | 2509 | 1248 |

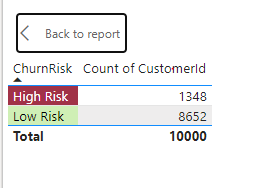
1. How can we create a conditional formatting setup to visually highlight customers at risk of churn and to evaluate the impact of credit card rewards on customer retention?

We can go to format visual tab then in values then in cell element we can do conditional formatting to a column, used various charts to show conditional formatting

**Churn Risk Analysis (Table Visualization):**

Formatting Approach: Utilized the "Format Visual" tab and applied conditional formatting within the "Values" section.

Key Metrics: Displayed counts of customers categorized into "HIGH" and "LOW" churn risk.

Insights: Easily identified and compared the number of customers at different levels of churn risk.

**Credit Card Impact on Reward (Bar Chart):**

Formatting Approach: Implemented conditional formatting to distinguish between "Positive Impact" and "No Impact" of credit cards.

Key Metrics: Presented counts of customers affected by credit card rewards.

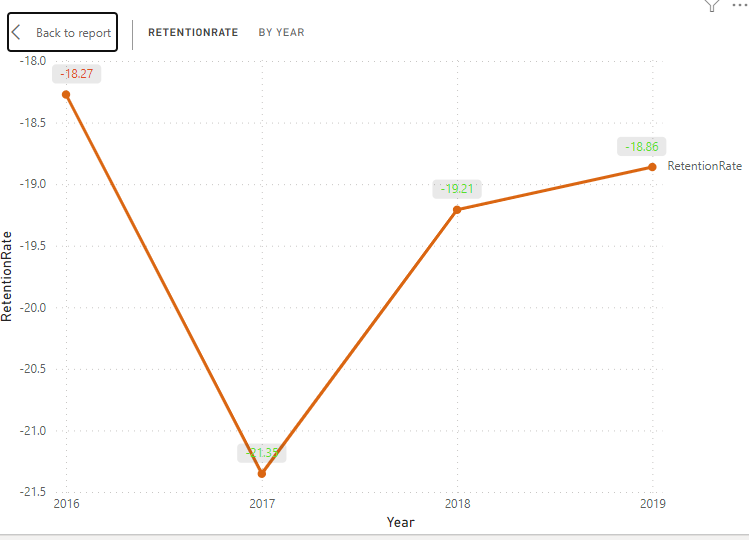
Insights: Clearly visualized the impact of credit cards on customer retention, categorizing customers based on their response to credit card rewards.

**Retention Rate Over Time (Line Chart):**

Formatting Approach: Utilized the "Lines" chart type to represent fluctuations in retention rates over different years.

Key Metrics: Displayed retention rates for each year, showing trends and changes.

Insights: Illustrated the dynamic nature of customer retention, allowing for the identification of patterns and fluctuations.



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 different strategies can be used to decrease the churn rate?

**Observation**

Churn rate according to year, we see that churn is highest for the year 2017 and least for the 2016

**Insights**

**1. Gender churn**

More females tend to exit than males.

**2. Age-Related Churn:**

The churn rate is higher for individuals aged 50 and above.

**3. location Churn Rate:**

Observation: Germany has the highest churn rate at 32%.

**Strategies used to decrease churn rate**

**1. Improved Communication:**

Strengthen communication channels to keep customers informed about products, services, and promotions.

**2. Customer Feedback:**

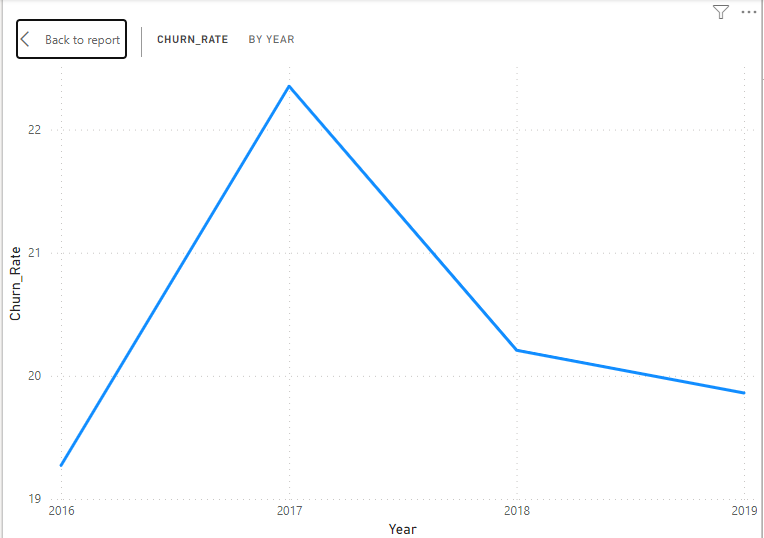
Gather feedback from customers who churned to identify pain points and areas for improvement**.**

**3. Enhanced Customer Support:**

Provide proactive and responsive customer support to address concerns and inquiries promptly.

**4. Retention Campaigns:**

Implement targeted retention campaigns for specific customer segments identified as high-risk.



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

Created a Dashboard in Power BI that is in Power BI file

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

**Data Exploration:**

Explore the structure and content of the 'bank\_churn' and 'customerinfo' tables.

Understand the relationships between different columns and how they may be relevant to churn analysis.

**Data Cleaning and Preparation:**

Clean the data by handling missing values, removing duplicates, and ensuring consistency in formatting.

Merge or join the two tables based on the 'CustomerId' column to create a unified dataset for analysis.

**Descriptive Analysis:**

Calculate basic statistics and visualize distributions of key variables such as age, credit score, tenure, balance, etc.

Identify any patterns or anomalies in the data that may require further investigation.

**Churn Rate Calculation:**

Calculate the churn rate overall and potentially per year by counting the number of customers who have exited ('Exited' = 1) divided by the total number of customers**.**

Explore trends in churn rates over time, if applicable.

**Segmentation Analysis:**

Segment the customer base based on various demographic and behavioral attributes such as age, gender, geography, credit score, tenure, etc.

Analyze churn rates and other relevant metrics within each segment to identify patterns and differences.

**Correlation Analysis:**

Conduct correlation analysis to understand the relationships between different variables and churn rate.

Identify factors that may be positively or negatively correlated with churn.

**Insights and Recommendations:**

Summarize key findings from the analysis, including factors influencing churn rate and characteristics of customers more likely to churn.

Provide actionable recommendations to reduce churn rate, such as targeted marketing campaigns, improved customer service, product/service enhancements, etc.

**Visualization and Reporting:**

Create visualizations (e.g., charts, graphs, dashboards) to communicate insights effectively.

Prepare a comprehensive report summarizing the analysis, insights, and recommendations for stakeholders.

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

**1.Access Power Query Editor:**

Open Power BI Desktop.

Go to the "Home" tab.

Click on "Transform data" to access Power Query Editor.

**2. Source Table:**

Navigate to the relevant table or data source in the Power Query Editor.

**3. Edit Column Name:**

Double-click on the header of the column you want to rename.

**4. Enter New Column Name:**

A text box will appear, allowing you to enter the new name for the column.

**5. Apply Changes:**

Press Enter to confirm the new name.

**6. Close & Apply:**

Once you've made all necessary changes, click on "Close & Apply" to save the modifications and exit the Power Query Editor.