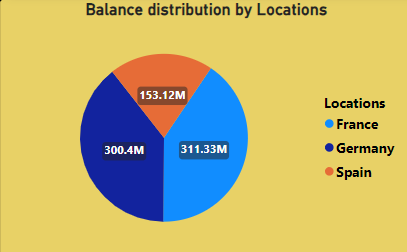
**Objective Questions**

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

**Answer:-**



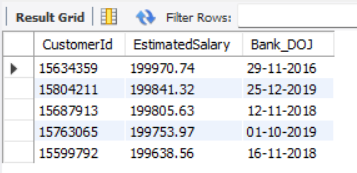
**Approach:**

There are currently three regions in the dataset: France, Germany, and Spain. To determine the distribution of account balances, we calculate the total balances and group them by the regions (Geography).

From this pie chart that we used to visualize the distribution of balances across different regions, we draw an insight that **France** has highest account balance.

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

**Answer:**

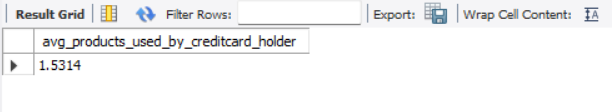
****

**Approach:**

To identify the top 5 customers with the highest estimated salary, we sort the data in descending order by salary and then select the top 5 rows.

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

**Answer:**

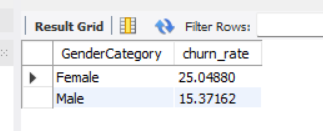
****

**Approach:**

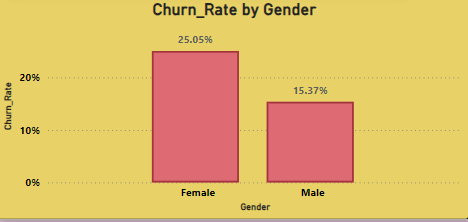
The average number of products used by customers who have a credit card is 1.5314. To determine this, we calculate the average `NumOfProducts` for customers who are credit card holders after filtering the data accordingly.

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

**Answer:**

****

**Using Power BI:**

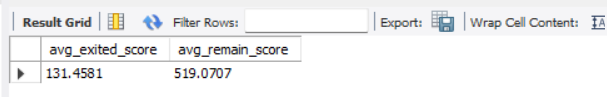


**Approach:**

To determine this, we calculate the average number of churns (Exited) for each gender by grouping the data by gender.

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

**Answer:**

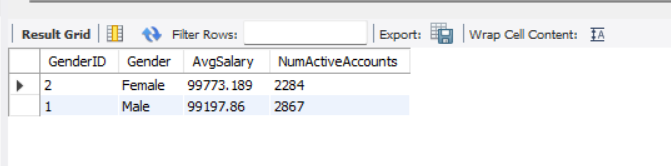
****

**Approach:**

To achieve this, we calculate the average credit score of customers grouped by their exit status.

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

**Answer:**

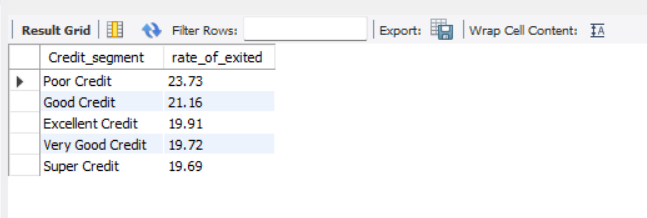
****

**Approach:**

This SQL query calculates the average estimated salary (`AvgSalary`) and counts the number of active accounts (`NumActiveAccounts`) for each gender (`GenderID`) among active members (`IsActiveMember = 1`). Results are grouped by gender and ordered by average salary to determine which gender has a higher average estimated salary and its relation to active accounts.

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

**Answer:**

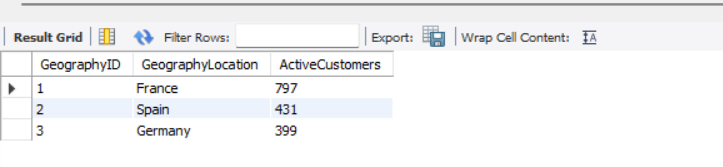
****

**Approach:**

Segment customers by credit score ranges using SQL, compute average exit rates for each segment, and identify the segment with the highest exit rate by ordering results in descending order.

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

**Answer:**

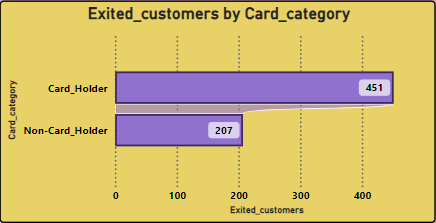
****

**Approach:**

This SQL query identifies the geographic region with the highest number of active customers who have a tenure greater than 5 years. It joins customerinfo, bank\_churn, and geography tables, filters for active members (IsActiveMember = 1) with tenure greater than 5, groups by geography, and counts active customers per region

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

**Answer:**



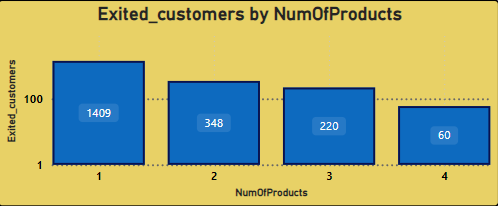
From the chart we can easily get or draw the insight that count of customers with card(card\_holder) id high as compare to customer without card(non-card-holder).

**Approach:**

Based on the data, the count of customers who exited is higher among those who have a credit card (451) compared to those who do not have a credit card (207)

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

**Answer:**



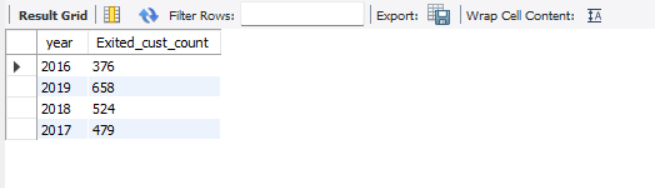
The above graph indicates that the most frequent number of products used by churned or exited customers is 1, with 1,409 out of the total number of churned or exited customers falling into this category.

**Approach:**

To determine the most common number of products used by exited customers, we analyze the dataset focusing on customers who have exited. We count the occurrences of each number of products used and identify the number that appears most frequently among these 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.

**Answer:**

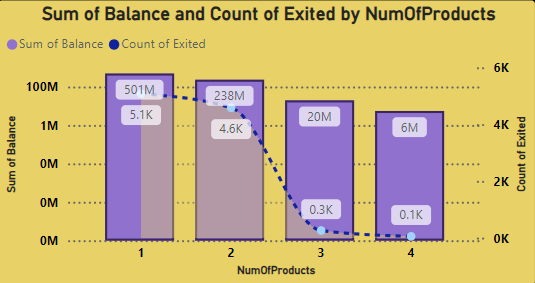
****

**Approach:**

Using SQL, we extract the year from the date of joining (`bank\_DOJ`) and count the number of customers who have exited (`Exited = 1`) grouped by each year. This helps identify trends in customer joinings over time and any potential seasonal patterns, which can be visualized to understand annual or monthly variations in customer behavior.

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

**Answer:**



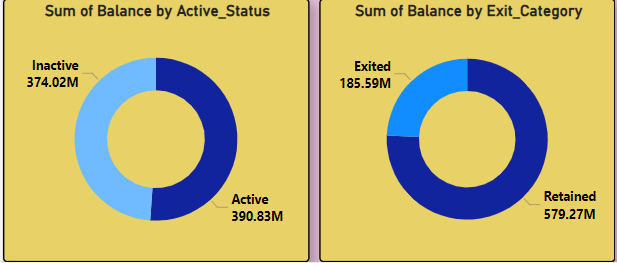
The relationship between the number of products and account balance shows that the total balance of customers decreases as the number of products increases. This suggests that more customers tend to use fewer products. The analysis used the total balance instead of the average balance because the average balance is similar across all segments due to having a similar number of customers in each segment.

**Approach:**

In analyzing the relationship between the number of products used and customer account balances, we observed that as the number of products used increases, the total balance across customers decreases. This observation is based on the total balance sum rather than the average, as average balances were found to be similar across different product usage segments due to comparable customer counts.

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

**Answer:**



From the two visuals, we can infer that customers who are active and retained (not exited) tend to have higher account balances compared to inactive and exited customers.

**Approach:**

Compare account balances between active, retained customers and inactive, exited customers to identify potential outliers based on balance distribution.

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

**Answer:**

The Newton school has provided us 7 different tables(datasets) , and out of these, 5 tables contain only categorical variables. These categorical tables are:

Gender, Geography, Active Customer, Exit Customer, CreditID.

**Approach:**

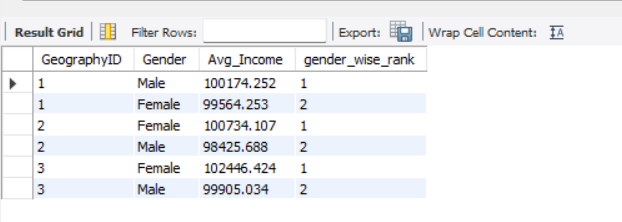
To determine the number of different tables and identify those containing only categorical variables:

1. Review the dataset to count the total number of tables provided.

2. Identify tables that exclusively contain categorical variables by examining their structure and contents, such as Gender, Geography, Active Customer, Exit Customer, and CreditID tables.

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)

**Answer:**

****

**Approach:**

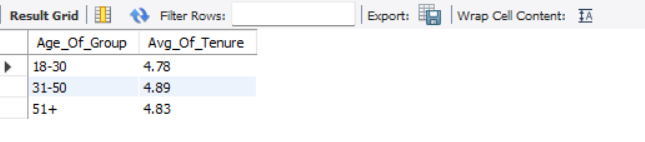
1. Calculate Average Income by Gender and Geography: Use SQL to group customers by `GeographyID` and `GenderID`, calculating the rounded average `EstimatedSalary` for each group.

2. Rank Gender by Average Income: Utilize the `RANK()` window function to rank genders within each `GeographyID` based on their average income, ordering results by `GeographyID` and the rank.

This approach efficiently computes and ranks the gender-wise average income across different geography IDs using SQL.

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+).

**Answer:**

****

**Approach:**

1. Categorize Age Groups: Use SQL to classify customers into age brackets (18-30, 31-50, 50+) based on their `Age`.

2. Calculate Average Tenure: Calculate the rounded average `Tenure` of customers who have exited (`Exited = 1`) within each age group.

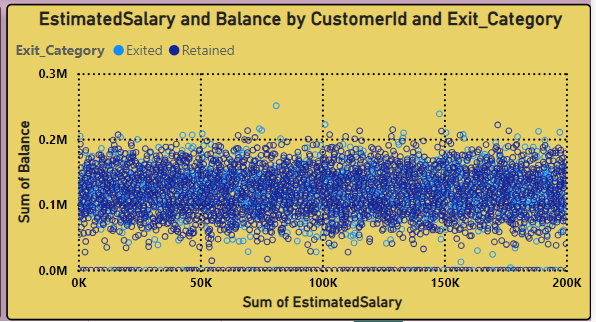
3. Group and Order Results: Group the results by age bracket and order them accordingly to display the average tenure for each category.

This approach effectively segments customers by age, computes their average tenure among those who have exited, and presents the results in SQL.

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?

**Answer:**

Customers who exited tend to have a lower sum of estimated salary and balance compared to retained customers. There are however some outliers.There is a positive correlation between the sum of estimated salary and the sum of balance. This means that customers with a higher estimated salary also tend to have a higher balance.



**Approach:**

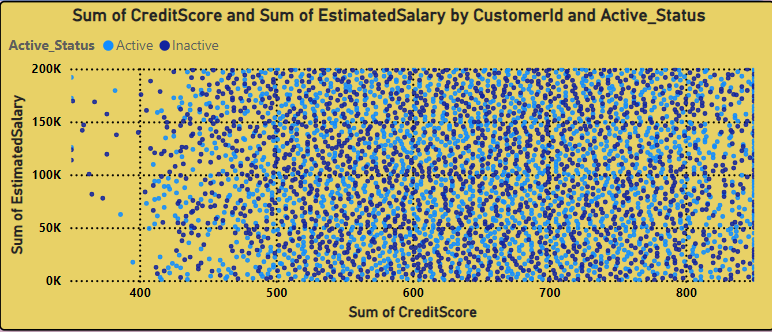
Comparison by Exit Status: Compare the sum of estimated salary (EstimatedSalary) and balance (Balance) between customers who have exited (Exited = 1) and those who have not.

Observation of Outliers: Identify any outliers where the relationship between salary and balance deviates significantly.

Correlation Analysis: Determine the correlation between the sum of estimated salary and balance using statistical methods or visualization tools to confirm a positive correlation where applicable.

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

**Answer:**



Yes, there appears to be a positive correlation between the sum of estimated salary and credit score in the data you provided. This means that customers with a higher estimated salary tend to also have a higher credit score.

**Approach:**

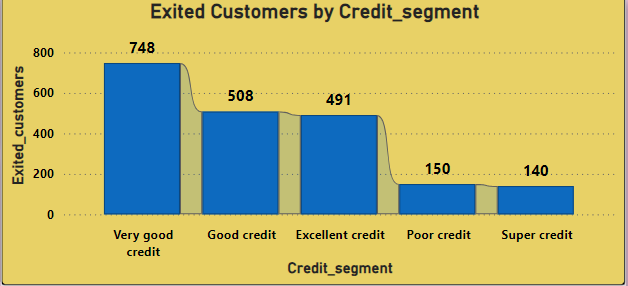
Data Analysis: Examine the provided dataset to calculate the sum of estimated salary (EstimatedSalary) and credit score (CreditScore) for customers.

Correlation Assessment: Use statistical methods or visualization techniques to determine the relationship between salary and credit score.

Observation: Identify patterns or trends that indicate a positive correlation, where higher estimated salary aligns with higher credit scores among customers.

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

**Answer:**

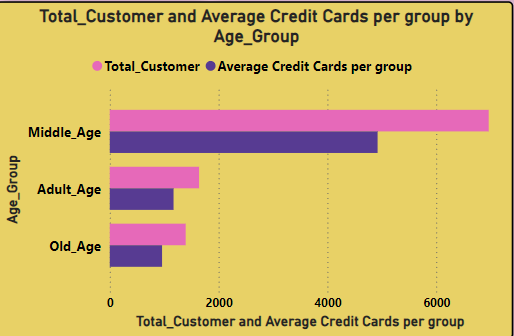


**Approach:**

Group the different segments of credit score based on range of score the groups are very good credit, good credit, excellent credit, poor credit, super credit so that we can easily get the which segment or groups have more or less customers that are exited or retained.

1. 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**.**

**Answer:**



According to the visual to we can say that the highest customer that have credit cards or high average of credit card have is the customers belong to middle age.

**Approach:**

Age Bucket Calculation: Group customers into age categories (e.g., 18-30, 31-50, 50+).

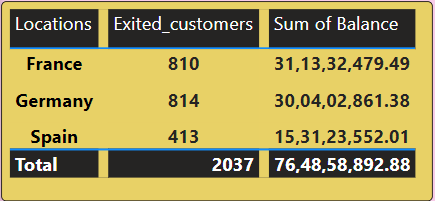
Count Customers with Credit Cards: Use SQL to count customers with credit cards in each age bucket.

Compute Average Credit Cards: Calculate the average number of credit cards per age bucket.

Identify Below-Average Buckets: Retrieve age buckets where the number of credit cards is less than the average.

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

**Answer:**



From this table we can easily get to knows that France has highest balance and high exited customers from Germany.

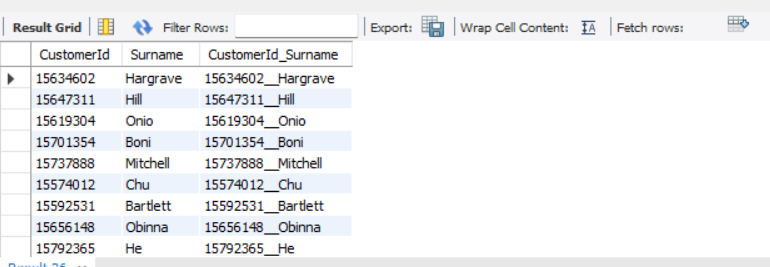
**Approach:**

Group by Location: Group customers by their geographical locations.

Rank by Churn Count and Average Balance: Calculate the number of churned customers and the average balance per location, then rank locations based on these metrics to identify which location has the highest average balance.

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”**.**

**Answer:**

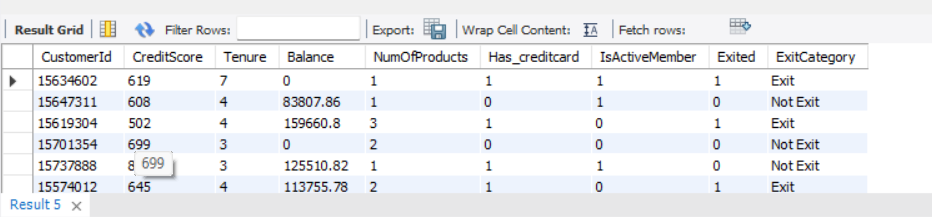
****

**Approach:**

To create a column named `CustomerId\_Surname` that combines `CustomerID` and `Surname` in the format "CustomerID\_Surname", join the `CustomerInfo` table (`c`) with another table (`bc`) on `CustomerID`. Use the `CONCAT` function in SQL to concatenate `CustomerID` and `Surname`, aliasing it as `CustomerId\_Surname` in the select statement for output.

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

**Answer:**

****

**Approach:**

The subquery (SELECT ExitCategory FROM ExitCustomers WHERE ExitCustomers.CustomerId = bank\_churn.CustomerId) is used to fetch the ExitCategory for each customer in the Bank\_Churn table based on their CustomerId.

This approach avoids using a traditional JOIN operation and directly fetches the ExitCategory using a correlated subquery within the SELECT statement.

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

**Answer:**

This section discusses the issue of missing values, a common challenge when data points are not recorded or become unavailable. Fortunately, our current dataset has no missing values, sparing us from the complications of dealing with them and avoiding potential biases introduced by imputation techniques.

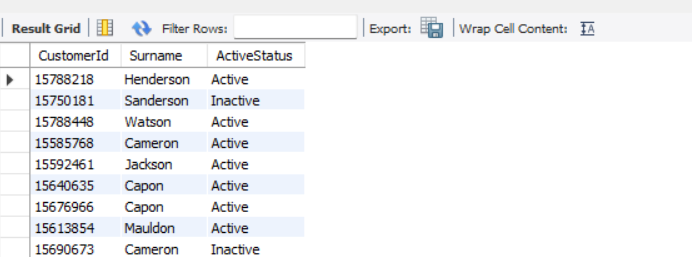
### **Handling Missing Values: A Reference for Future Analyses**

While we do not face the problem of missing values in this dataset, it's useful to know the common strategies for managing them in future analyses:

* **Deletion**: This method involves removing rows or columns that contain missing values. It is suitable when the amount of missing data is negligible and its removal does not significantly affect the analysis. However, it comes with the drawback of losing some information.
* **Imputation**: This approach fills in missing values with estimated ones. Techniques include imputing with the mean, median, or mode, using k-Nearest Neighbors (KNN), or applying more advanced methods. The choice of method should align with the data type and its distribution characteristics.
* **Modeling Techniques**: Certain statistical models are capable of handling missing values inherently. Nonetheless, it's crucial to understand why data is missing, as this can influence the appropriate handling method.

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”.

**Answer:**

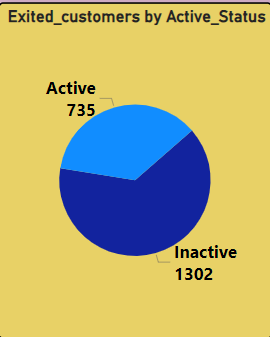


**Approach:**

Use SQL to join `bank\_churn` and `customerinfo` tables on `CustomerID`, then filter results where `Surname` ends with "on". Select `CustomerID`, `Surname`, and use a `CASE` statement to determine `ActiveStatus` based on `IsActiveMember` in `bank\_churn` table.

1. 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.

**Answer:**



customers who've left the bank should no longer be considered active members. However, our graph shows that 735 of these exited customers are still marked as "Active Members." This doesn't match our logic that churned customers should be inactive.

**Approach:**

Identify discrepancies by querying customers marked as both "Exited" and "Active Members". Validate data integrity by ensuring that all customers marked as "Exited" are correctly categorized as "Inactive" in the `IsActiveMember` column to align with logical expectations.

**Subjective Question**

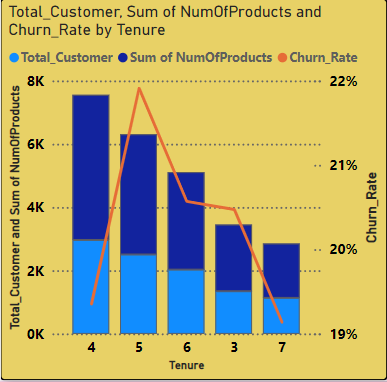
1. Customer Behavior Analysis: What patterns can be observed in the spending habits of long-term customers compared to new customers, and what might these patterns suggest about customer loyalty?

**Answer:**

Looking at the data, we found that customers who have been with us for 4 years tend to spend more. Interestingly, those with a tenure of 5 years show a higher rate of leaving, suggesting they're more likely to churn. Overall, it appears that customer loyalty increases with tenure, meaning those who have been with us longer tend to stick around more.

**Approach:**

Analyze spending data to identify trends in long-term versus new customer behavior, noting higher spending among long-term customers. Investigate churn rates at different tenures to understand and address loyalty dynamics.



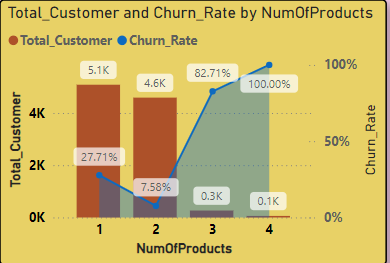
1. Product Affinity Study: Which bank products or services are most commonly used together, and how might this influence cross-selling strategies?

**Answer:**

The data shows that customers with fewer products tend to have a lower churn rate compared to those who have purchased more products.

**Approach:**

Identify commonly paired products and services to uncover usage patterns, noting that fewer product holdings correlate with lower churn rates. Use these insights to refine cross-selling strategies to enhance customer retention.



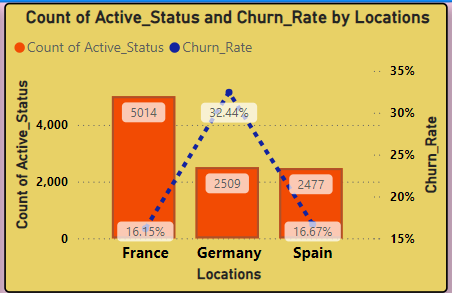
1. Geographic Market Trends: How do economic indicators in different geographic regions correlate with the number of active accounts and customer churn rates?

**Answer:**

From the above visual we can easily get to know that the France has high number of active members as compare to other two location but churn rate of Germany is high.

**Approach:**

Analyze economic indicators by region to correlate with active account numbers and churn rates, highlighting France's high membership and Germany's high churn. Utilize these findings to tailor regional strategies for retention and growth.



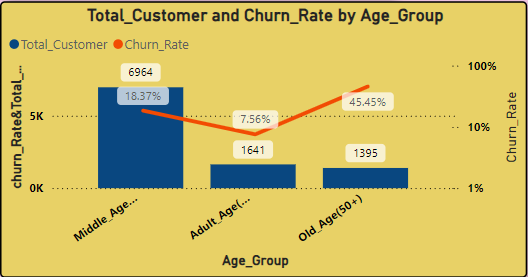
1. Risk Management Assessment: Based on customer profiles, which demographic segments appear to pose the highest financial risk to the bank, and why?

**Answer:**

Highest churn (45.45%) among busy Middle-Aged Adults (35-49). Young Adults (18-34) churn at a moderate rate (18.37%). Customer loyalty is strong among Old Age (50+) with churn below 1%.

**Approach:**

Target busy Middle-Aged Adults (35-49) & Young Adults (18-34) with incentives and improved service. Focus on increasing customer satisfaction for these age groups. Leverage CRM to understand their needs and personalize your strategy.



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?

**Answer:**

To enhance customer retention, the bank should offer personalized incentives and promotions that cater to individual preferences and behaviors. Additionally, implementing loyalty programs that reward customers for their ongoing patronage can further motivate them to remain with the bank.

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:**

To evaluate how marketing campaigns affect customer retention and acquisition, we need to consider several key strategies and additional data points:

**Enhanced Customer Service:** Improve customer service by offering personalized assistance, resolving issues quickly, and addressing customer feedback efficiently.

**Targeted Offers:** The marketing team should create special offers and provide additional security features for customers over the age of 50, as well as offer incentives to customers who purchase multiple products. Special promotions for credit card holders should also be considered.

### Key Strategies:

**Customer Segmentation:** Group customers by age, location, and the number of products they use. This segmentation allows for a more focused analysis of how different campaigns perform within specific groups.

**Trend Analysis:** Study changes in active customer numbers, exit rates, and product usage over time. This helps identify trends that might be influenced by marketing efforts.

**Control Groups:** If possible, establish control groups that haven't been exposed to certain campaigns. Comparing these groups to those that have been exposed can help determine the true impact of the campaigns.

**Campaign Details & Timing:** Detailed information about campaign content, channels used (e.g., online, offline), and launch dates. This data is essential for linking campaign exposure to changes in customer behavior.

**Customer Acquisition Channel:** Information on how customers were originally acquired (e.g., referral, online advertisement). This helps assess whether campaigns are effective in retaining customers acquired through different channels.

**Customer Lifetime Value (CLV):** Calculate the CLV to understand the long-term impact of marketing campaigns on customer retention and revenue.

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

**Answer:**

Customers over the age of 50 find the bank's terminology and schemes unsatisfactory. Additionally, those with credit scores under 700 are also departing. Both Germany and France exhibit high churn rates. Poor customer satisfaction often results in churn, with common reasons including dissatisfaction with customer service, long wait times, and unresolved issues.

Low customer satisfaction is a key driver of churn. Common reasons include poor customer service, long wait times, and unresolved issues.

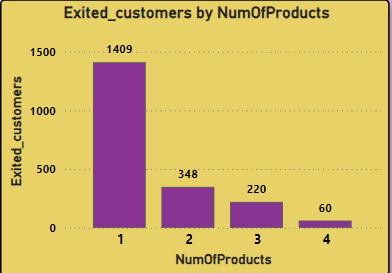
1. Are 'Tenure', 'NumOfProducts', 'IsActiveMember', and 'EstimatedSalary' important for predicting if a customer will leave the bank?

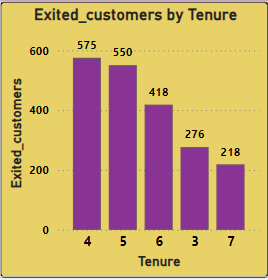
**Answer:**

According to the dataset, customers with a tenure of 4 or 5 years are more prone to leaving. Those who use fewer products also have a higher likelihood of exiting, and their likelihood of departure cannot be predicted based on their estimated salary.

**Approach:**

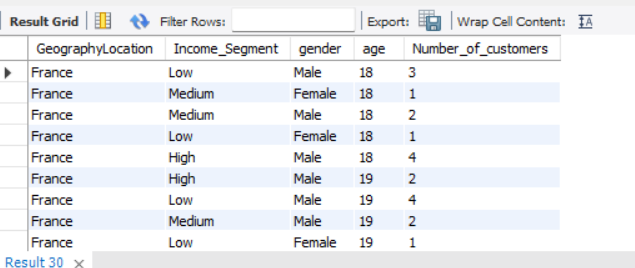
Evaluate the influence of 'Tenure', 'NumOfProducts', 'IsActiveMember', and 'EstimatedSalary' on customer churn, noting higher exit rates for 4-5 year tenures and fewer product users. Use these insights to refine predictive models for customer retention.





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

**Answer:**



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?

**Answer:**

Created a Churn Risk Score: Added a new calculated column based on three relevant columns: Churn Risk Score (CRS) = Balance / (Credit Score NumOfProducts). This indicates that the Churn Risk is directly proportional to the Balance and inversely proportional to the Credit Score and Number of Products. Note: Customers with a Balance of 0 will have a CRS of 0.

Created a Table Visual: Developed a table that includes 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

- Use background colors to highlight different risk levels:

- 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 evaluated

Benefits:

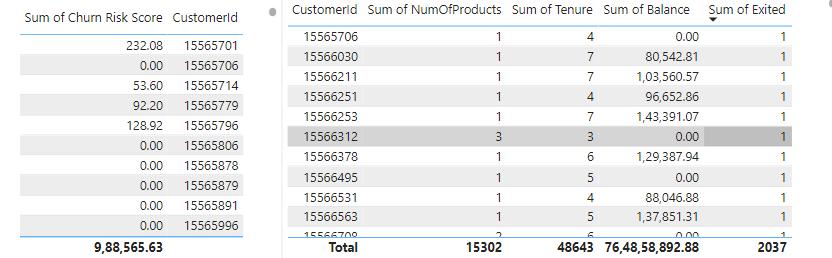
This color-coded table quickly identifies at-risk customers, helping you prioritize retention efforts. It acts like a heatmap within your table, directing attention to potential churners.

Remember:

- Judgments should be made for Active Customers.

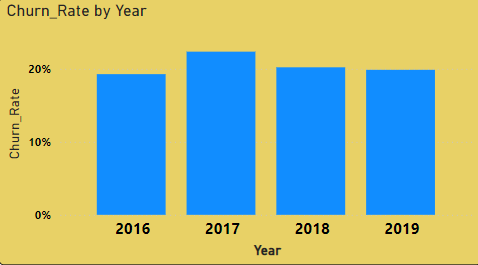
- Thresholds and colors should be periodically adjusted based on data and risk assessment.

- Consider exploring additional formatting options for enhanced visual cues as the data grows.



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?

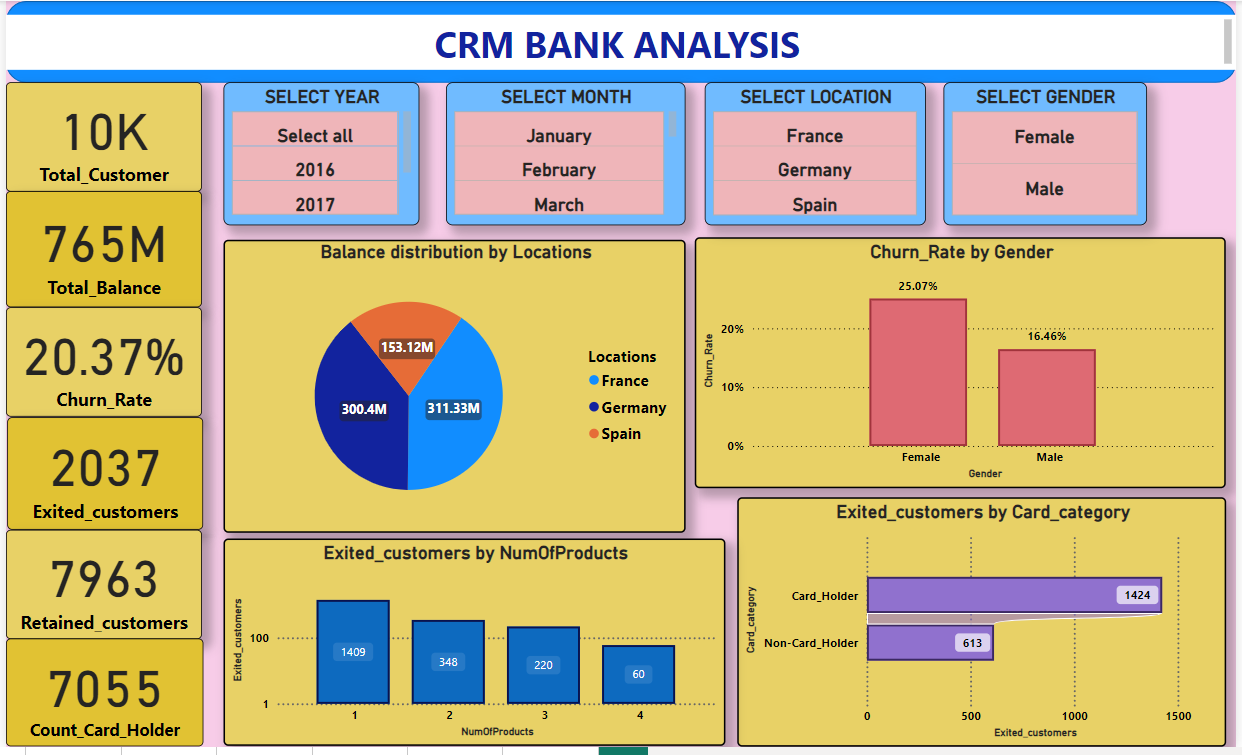
**Answer:**



The overall churn rate stands at 20.47%. In 2016, it was 19.27%, rising to 22.35% in 2017, then decreasing to 20.21% in 2018, and slightly increasing to 19.86% in 2019. Customers over the age of 50 and those with fewer products are more prone to leaving. By offering attractive rewards, promotions, and security features to these customer segments, the bank could potentially retain them for a longer duration.

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

**Answer:**



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

**Answer:**

Contextual Understanding: Start by thoroughly comprehending the context and aim of the data analysis. Engage with stakeholders or review project materials to uncover any implicit goals or expectations.

Data Exploration: Perform exploratory data analysis to become familiar with the dataset. Analyze its structure, distributions, and variable relationships, noting any patterns, anomalies, or trends.

Insight Identification: Brainstorm potential insights or hypotheses based on your initial data exploration. Consider both quantitative trends and qualitative factors relevant to the problem domain.

Hypothesis Generation: Develop hypotheses or conjectures from your preliminary observations. Use these as guiding principles for further analysis and testing.

Iterative Analysis Process: Apply a variety of analytical methods iteratively to test and refine your hypotheses. This may involve statistical analysis, machine learning, or qualitative techniques, as well as experimenting with data using tools like Excel, PowerBI, and SQL.

Visualization and Communication: Use data visualization techniques to effectively present your findings. Visuals can reveal patterns, relationships, and insights to stakeholders in an understandable way.

Synthesis of Findings: Integrate your findings into coherent narratives or themes, providing a comprehensive understanding of the data. Identify overarching patterns or trends that span different aspects of the analysis.

Feedback and Validation: Validate your results through peer reviews, expert opinions, or external comparisons if feasible. Seek stakeholder feedback to ensure your analysis aligns with their expectations and enhances their understanding of the problem.

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

**Answer:**

