***Objective Questions:***

**1. What is the distribution of account balances across different regions?(SQL)**

*select avg(b.Balance) as Avgbalance,g.GeographyLocation from bank\_churn b*

*join customerinfo c*

*on b.CustomerId=c.CustomerId*

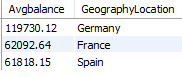
*join geography g*

*on c.GeographyID=g.GeographyID*

*group by g.GeographyLocation*

*order by Avgbalance desc;*

Joining the three tables, customerinfo, Bankchurn, and Geography, yields the desired output. The Bankchurn table facilitates obtaining the average balance through the **AVG function**, while the Geography table is linked to the customerinfo table via the GeographyID using the **JOIN function**. The **GROUP BY** function is employed to group the location-based data, and the **ORDER BY** function arranges the output in descending order of the average balance. From the resulting output, it is observed that ***Germany*** has the highest average customer account balance, amounting to ***119730.11613391782***.



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

*SELECT CustomerId,Surname,Age,GenderID,EstimatedSalary,GeographyID,BankDOJ*

*FROM customerinfo*

*WHERE MONTH(BankDOJ) IN (10, 11, 12)*

*ORDER BY EstimatedSalary DESC*

*LIMIT 5;*

In order to retrieve the highest estimated salary in the last quarter (October, November, and December), the WHERE clause filters data based on the months. Utilizing the MONTH function, the month is extracted, and the 'IN' operator specifies the months. Subsequently, the estimated salaries are sorted in descending order using the ORDER BY clause. Finally, limiting the output to the top 5 highest estimated salaries is achieved through the LIMIT clause.

***Output:***

|  |  |  |
| --- | --- | --- |
| **Surname** | **EstimaedSalary** | **BankDOJ** |
| Dyer | 199970.74 | 2016-11-29 |
| Oluchukwu | 199841.32 | 2019-12-25 |
| Mai | 199805.63 | 2018-11-12 |
| Palerma | 199753.97 | 2019-10-01 |
| Dimauro | 199638.56 | 2018-11-16 |

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

*SELECT round(avg(NumofProducts),2) as Avg\_num\_of\_products from bank\_churn*

*where HasCrCard='1';*

The average number of products can be calculated using the AVERAGE function, with a WHERE condition applied to filter customers who have a credit card.

***Output:***

|  |  |  |  |
| --- | --- | --- | --- |
|  | |  | | --- | | **Avg\_num\_of\_products** | | 1.53 | |
|  | **4. Determine the churn rate by gender for the most recent year in the dataset.(Power Bi)**  In the analysis conducted using a table chart, the churn rate by gender for the most recent year available in the dataset, 2019, was determined. Notably, the churn rate for females (25.05%) exceeded that of males (15.37%), despite the total count of exited customers being 658. It's noteworthy that the total number of male customers is higher than females, yet the count of exited customers is lower compared to females.  ***Output:*** |
|  |  |
|  | **5. Compare the average credit score of customers who have exited and those who remain. (SQL)**  *SELECT ec.ExitCategory,*  *AVG(bc.CreditScore) AS AvgCreditScore*  *FROM Exitcustomer ec*  *JOIN bank\_churn bc*  *ON ec.ExitID = bc.Exited*  *GROUP BY ec.ExitCategory;*  This SQL query calculates the average credit score of customers grouped by their exit category. It selects the exit category from the *Exitcustomer* table and the average credit score from the *CreditScore* column of the *bank\_churn* table, **aliasing** it as *AvgCreditScore*. The query **joins** the *Exitcustomer* table with the *bank\_churn* table based on the *ExitID* column from *Exitcustomer* and the Exited column from *bank\_churn*. Finally, the results are *grouped by* the exit category.  ***Output:***   |  |  | | --- | --- | | **ExitCategory** | **AvgCreditScore** | | Exit | 645.35 | | Retain | 651.85 |   While the difference in average credit scores between the two groups is relatively small, it appears that customers who remain have a slightly higher average credit score compared to those who have exited.  **6. Which gender has a higher average estimated salary, and how does it relate to the number of active accounts? (SQL)**  *SELECT g.GenderCategory,*  *AVG(c.EstimatedSalary) AS AvgEstimatedSalary,*  *round(SUM(CASE WHEN b.IsActiveMember = 1 THEN 1 ELSE 0 END)) AS ActiveAccounts*  JOIN bank\_churn b ON c.CustomerId = b.CustomerId  GROUP By g.GenderCategory; |
|  | *FROM customerinfo c*  *JOIN gender g*  *ON c.GenderID = g.GenderID* |

It calculates the average estimated salary and the count of active accounts for each gender category. It achieves this by **joining three tables**: customerinfo, gender, and bank\_churn. The gender table is linked to the customerinfo table via the GenderID column, providing the gender category for each customer. Meanwhile, the bank\_churn table is connected to the customerinfo table through the CustomerId column, allowing for the retrieval of churn-related data. The query utilizes aggregate functions such as **AVG and SUM**, along with conditional statements within the **CASE expression**, to compute the average estimated salary and count the active accounts. Finally, the results are **grouped by** gender category.

***Output:***

|  |  |  |
| --- | --- | --- |
| **GenderCategory** | **AvgEstimatedSalary** | **ActiveAccounts** |
| Female | 100601.54 | 2284 |
| Male | 99664.57 | 2867 |

The data shows males have more active accounts compared to females. But still, the AvgEstimatedSalary for females is higher than that of males, which indicates that gender affects the AvgEstimatedSalary.

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

*WITH CreditScoreSegments AS (*

*SELECT*

*CASE*

*WHEN CreditScore >= 800 AND CreditScore <=850 THEN 'Excellent'*

*WHEN CreditScore >= 740 AND CreditScore < 800 THEN 'Very Good'*

*WHEN CreditScore >= 670 AND CreditScore < 740 THEN 'Good'*

*WHEN CreditScore >= 580 AND CreditScore < 670 THEN 'Fair'*

*WHEN CreditScore >= 300 AND CreditScore < 580 THEN 'Poor'*

*ELSE 'Unknown'*

*END AS CreditScoreSegment,*

*CustomerId*

*FROM bank\_churn)*

*SELECT CreditScoreSegment,*

*COUNT(CASE WHEN ExitCategory = 'Exit' THEN 1 END) AS ChurnedCustomers,*

*COUNT(\*) AS TotalCustomers,*

*ROUND(COUNT(CASE WHEN ExitCategory = 'Exit' THEN 1 END) \* 100.0 / COUNT(\*), 2) AS ExitRate*

*FROM CreditScoreSegments cs*

*JOIN bank\_churn b ON cs.CustomerId = b.CustomerId*

*join exitcustomer ec on b.Exited=ec.ExitID*

*GROUP BY CreditScoreSegment*

*ORDER BY ExitRate DESC;*

This SQL query segments customers based on their credit score into predefined categories, such as Excellent, Very Good, Good, Fair, and Poor. It calculates the count of churned customers, total customers, and exit rate for each credit score segment. The "*CreditScoreSegments*" common table expression (*CTE*) categorizes customers into segments based on their credit score ranges. The main query then joins this CTE with the "*bank\_churn*" table to retrieve customer exit information and with the "*exitcustomer*" table to obtain exit categories. Finally, it groups the results by credit score segment and calculates the exit rate, ordering the output by exit rate in descending order.

***Output:***

|  |  |  |  |
| --- | --- | --- | --- |
| **CreditScoreSegment** | **ExitedCustomers** | **TotalCustomers** | **ExitRate** |
| Poor | 520 | 2362 | 22.02 |
| Very Good | 252 | 1224 | 20.59 |
| Fair | 685 | 3331 | 20.56 |
| Excellent | 128 | 655 | 19.54 |
| Good | 452 | 2428 | 18.62 |

The data suggests that customers in the "Poor" credit score segment have the highest exit rate at 22.02%. (The number of exited customer is maximum for Fair segment but it has high total customers count.)

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

*SELECT g.GeographyLocation,*

*count(CASE WHEN b.IsActiveMember = 1 THEN 1 ELSE 0 END) AS ActiveAccounts from geography g*

*join customerinfo c*

*on g.GeographyID=c.GeographyID*

*join bank\_churn b*

*on c.CustomerId=b.CustomerId*

*where b.Tenure>5*

*group by g.GeographyLocation;*

The SQL query identifies the geographic region with the highest number of active customers who have a tenure greater than 5 years. It achieves this by **joining** the geography, customerinfo, and bank\_churn tables based on their corresponding IDs. Within the **WHERE clause**, the query filters customers with a tenure exceeding 5 years. Using a **COUNT function** with a conditional statement, the query determines the count of active accounts in each geographic region. The results are then **grouped by** geography location.

***Output:***

|  |  |
| --- | --- |
| **GeographyLocation** | **ActiveAccounts** |
| France | 1575 |
| Spain | 796 |
| Germany | 800 |

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

*SELECT cc.Category AS CreditCardCategory,*

*COUNT(CASE WHEN ec.ExitCategory = 'Exit' THEN 1 END) AS ChurnedCustomers,*

*COUNT(\*) AS TotalCustomers,*

*ROUND(COUNT(CASE WHEN ec.ExitCategory = 'Exit' THEN 1 END) \* 100.0 / COUNT(\*), 2) AS ChurnRate*

*FROM Creditcard cc*

*JOIN bank\_churn ci ON cc.CreditID = ci.HasCrCard*

*JOIN Exitcustomer ec ON ci.Exited = ec.ExitID*

*GROUP BY cc.Category;*

By joining the Creditcard, Customerinfo, and Exitcustomer tables, it categorizes customers based on their credit card status and calculates the churn rate for each category. Using conditional statements, the query identifies churned customers within each credit card category and computes the total number of customers. By dividing the number of churned customers by the total and expressing the result as a percentage, it determines the churn rate for each category.

***Output:***

|  |  |  |  |
| --- | --- | --- | --- |
| **CreditCardCategory** | **ChurnedCustomers** | **TotalCustomers** | **ChurnRate** |
| Credit card holder | 1424 | 7055 | 20.18 |
| Non credit card holder | 613 | 2945 | 20.81 |

In this case, both credit card holders and non-credit card holders exhibit relatively similar churn rates, with credit card holders at 20.18% and non-credit card holders at 20.81%. This suggests that, based on the available data, there is no substantial difference in churn rates between customers who hold credit cards and those who do not.

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

*select b.NumOfProducts,*

*COUNT(CASE WHEN ec.ExitCategory = 'Exit' THEN 1 END) AS ExitedCustomers*

*from bank\_churn b*

*join exitcustomer ec*

*on b.Exited=ec.ExitID*

*group by b.NumOfProducts*

*order by ExitedCustomers desc;*

It retrieves data from the bank\_churn and exitcustomer tables and groups it by the number of products held by customers. Using a conditional statement, it counts the number of customers who have churned within each product category. The results are then ordered in descending order based on the count of exited customers.

***Output:***

|  |  |
| --- | --- |
| **NumOfProducts** | **ExitedCustomers** |
| 1 | 1409 |
| 2 | 348 |
| 3 | 220 |
| 4 | 60 |

It indicates that among exiting customers, the most common number of products used is 1, with 1409 customers.

**11. 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.(SQL)**

*SELECT*

*EXTRACT(YEAR FROM BankDOJ) AS JoinYear,*

*EXTRACT(MONTH FROM BankDOJ) AS JoinMonth,*

*COUNT(\*) AS JoinCount*

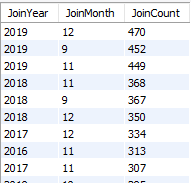
*FROM customerinfo*

*GROUP BY JoinYear,JoinMonth*

*ORDER BY joinCount desc;*

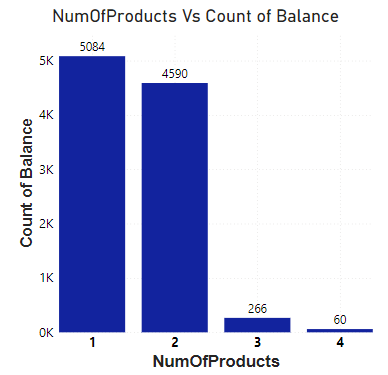
It examines the trend of customers joining over time by extracting the year and month from the "BankDOJ" (Bank Date of Joining) column in the "customerinfo" table. It counts the number of customers who joined in each year-month combination and then groups the results by year and month. Finally, it orders the results based on the join count in descending order.

***Output:***

****

It appears that in the year 2019(470), December had the highest number of customer joining’s, followed by November in 2018(368), December in 2017(334), and November in 2016(313). This analysis suggests that there is seasonal pattern in customer acquisition, with higher joining’s typically observed in the months of September, November, and December.

**12. Analyze the relationship between the number of products and the account balance for customers who have exited.(Power BI)**

****

Based on the analysis, it shows that the majority of customer who purchased 1 or 2 products exit more i.e., *5086 and 4590 for 1 and 2 product*s respectively. In the above chart, a filter ExitID = 1 is used.

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

SELECT

Exited,

COUNT(\*) AS count\_retained,

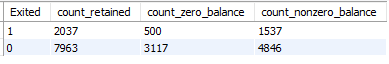
SUM(CASE WHEN Balance = 0 THEN 1 ELSE 0 END) AS count\_zero\_balance,

SUM(CASE WHEN Balance <> 0 THEN 1 ELSE 0 END) AS count\_nonzero\_balance

FROM bank\_churn

GROUP BY Exited;

***Output:***

****

Out of the 7963 retained customers, 3117 have a zero balance. This indicates that some customers who have remained with the bank have fully utilized their funds or closed their accounts. 4846 have a non-zero balance. This suggests that the majority of customers who have remained with the bank still have funds in their accounts.

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

There are seven different tables given in the dataset. Among these tables, "Gender", "Geography", "ExitCustomer", and "ActiveCstomer" are likely categorical variables.

**15. 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)**

*SELECT c.GeographyID, g.GenderCategory,*

*AVG(c.EstimatedSalary) AS AvgIncome,*

*RANK() OVER (PARTITION BY c.GeographyID ORDER BY AVG(c.EstimatedSalary) DESC) AS GenderRank*

*FROM customerinfo c*

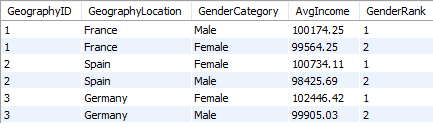
*JOIN gender g ON c.GenderID = g.GenderID*

*GROUP BY c.GeographyID, g.GenderCategory*

*ORDER BY c.GeographyID, GenderRank;*

It directly joins the customerinfo table with the gender table using the JOIN clause based on the GenderID column and then it calculates the average income (AVG(c.EstimatedSalary)) for each combination of GeographyID and GenderCategory using the GROUP BY clause. The RANK() window function is used to rank the genders within each geographic location based on their average income. The PARTITION BY clause ensures that ranking is done separately for each geographic location and gender category. Finally, the results are ordered by GeographyID and GenderRank.

***Output:***



In two geographic region(Spain,Germany), the average income of females is greater than that of male, it suggests that female income is higher than male income across two locations.

**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+).**

*SELECT CASE*

*WHEN Age BETWEEN 18 AND 30 THEN '18-30'*

*WHEN Age BETWEEN 31 AND 50 THEN '31-50'*

*ELSE '50+'*

*END AS AgeBracket,*

*AVG(Tenure) AS AvgTenure*

*FROM customerinfo ci*

*JOIN bank\_churn bc ON ci.CustomerId = bc.CustomerId*

*JOIN exitcustomer ec ON bc.Exited = ec.ExitID*

*GROUP BY AgeBracket;*

The customers are categorized into age brackets ('18-30', '31-50', and '50+') using a CASE statement based on their age. The average tenure of customers who have exited is then calculated by joining the necessary tables (customerinfo, bank\_churn, and exitcustomer) and utilizing the AVG() function. Finally, the results are grouped by the AgeBracket column to obtain the average tenure for each bracket.

***Output:***

|  |  |
| --- | --- |
| **AgeBracket** | **AvgTenure (Years)** |
| 31-50 | 4.87 |
| 18-30 | 4.84 |
| 50+ | 4.85 |

**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? (Power Bi)**

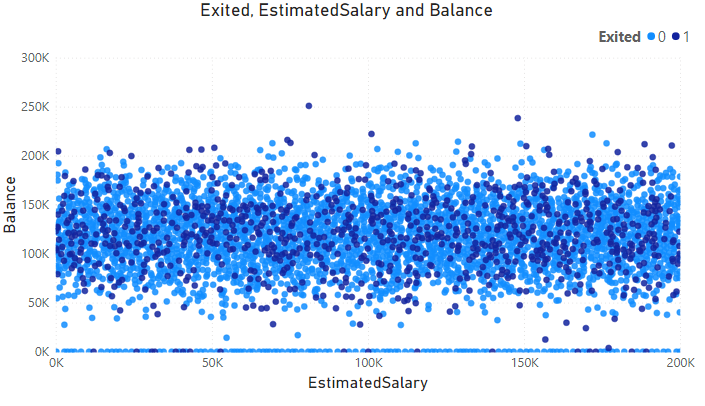
From the chart it is understood that

* Number of Exited is lesser than Number of Retained.
* Balance is lies between 50K and 200K.
* There are a maximum number of retained customers with zero balance compare to exited customers.

**0- Retained**

**1- Exited**

***Output:***

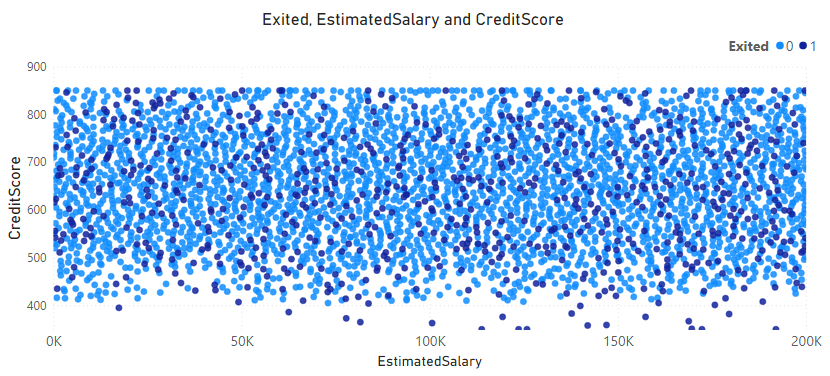
****

**18. Is there any correlation between the salary and the Credit score of customers? (Power Bi)**

* From the chart it is understood that Number of Exited is lesser than Number of Retained.
* Maximum number of customers those who possess highest credit score are retained customers.
* It is also shows that the customers having least credit score tends to exit the bank.

**0- Retained**

**1- Exited**



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

*SELECT CreditScoreBucket,*

*COUNT(CASE WHEN ec.ExitCategory = 'Exit' THEN 1 END) AS ChurnedCustomers,*

*RANK() OVER (ORDER BY COUNT(CASE WHEN ec.ExitCategory = 'Exit' THEN 1 END) DESC) AS Ranks*

*FROM (*

*SELECT CASE*

*WHEN CreditScore BETWEEN 800 AND 850 THEN 'Excellent'*

*WHEN CreditScore BETWEEN 740 AND 799 THEN 'Very Good'*

*WHEN CreditScore BETWEEN 670 AND 739 THEN 'Good'*

*WHEN CreditScore BETWEEN 580 AND 669 THEN 'Fair'*

*WHEN CreditScore BETWEEN 300 AND 579 THEN 'Poor'*

*END AS CreditScoreBucket,*

*CustomerId,Exited*

*FROM bank\_churn*

*) AS ScoreBuckets*

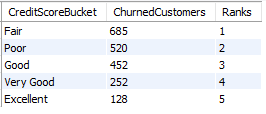
*LEFT JOIN exitcustomer ec ON ScoreBuckets.Exited = ec.ExitID*

*GROUP BY CreditScoreBucket*

*ORDER BY Ranks;*

This SQL query categorizes customers into credit score buckets (**'Poor', 'Fair', 'Good', 'Excellent', 'Unknown')** based on their credit scores. It then calculates the number of churned customers within each bucket. Additionally, it **ranks** the buckets according to the count of churned customers, with higher ranks indicating more churn. The query utilizes a subquery to assign credit score buckets to customers and then joins it with the exitcustomer table to identify churned customers. Finally, it **groups** the results by credit score bucket and orders them by rank in descending order.

***Output:***

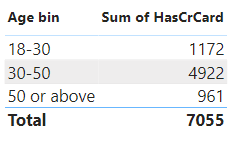


The "Fair" credit score bucket is ranked first because it has the highest number of churned customers compared to other credit score buckets, with 685 customers exiting the bank. It is noteworthy that a significant portion of customers fall within the 600-700 credit score range, which corresponds to the "Fair" credit score category.

**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.**

The given table contains the number of customers who have a credit card across three age buckets. The Average number of credit cards per age bucket is 2351.66 (7055/3). So, the age buckets 18-30 and 50 or above have lesser values than the average number of credit cards.

***Output:***

****

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

*SELECT GeographyLocation,Num\_Churned\_Customers,Avg\_Balance,*

*RANK() OVER (ORDER BY Num\_Churned\_Customers DESC, Avg\_Balance DESC) AS Location\_Rank*

*FROM(SELECT*

*geo.GeographyLocation,*

*COUNT(\*) AS Num\_Churned\_Customers,*

*ROUND(AVG(bc.Balance),2) AS Avg\_Balance*

*FROM bank\_churn bc*

*JOIN CustomerInfo ci ON bc.CustomerId = ci.CustomerId*

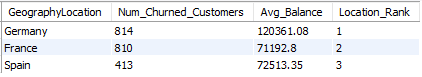
*JOIN Geography geo ON ci.GeographyID = geo.GeographyID*

*WHERE bc.Exited = 1*

*GROUP BY geo.GeographyLocation) AS LocationStats;*

This SQL query ranks the geographic locations based on the number of customers who have churned the bank (Num\_Churned\_Customers) and their average balance (Avg\_Balance). It utilizes the RANK() function to assign a rank to each location, ordering them by the number of churned customers in descending order, and then by the average balance in descending order.

***Output:***



Germany has the highest number of churned customers (814) and the highest average balance among them (120,361.08), thus it is ranked first.

**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”.**

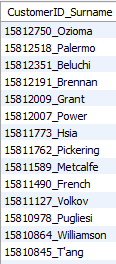
*SELECT CONCAT(ci.CustomerID, '\_', ci.Surname) AS CustomerID\_Surname*

*FROM CustomerInfo ci*

*JOIN OtherTable ot ON ci.CustomerID = ot.CustomerID AND ci.Surname = ot.Surname;*

This query joins the "CustomerInfo" table with the "bankchurn" based on the CustomerID column. The CONCAT function is then used to concatenate the CustomerID and Surname columns with an underscore (\_) separator to create the new column "CustomerID\_Surname".

***Output:***



Each row represents a unique customer identified by their CustomerID and Surname. For example, "15634602\_Hargrave" indicates a customer with a CustomerID of 15634602.

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

Yes, we get the ExitCategory” from ExitCustomers table to Bank\_Churn table by using subquery

SELECT \*,(SELECT ExitCategory FROM exitcustomer WHERE ExitID = bc.Exited) AS ExitCategory FROM bank\_churn bc;

This SQL query selects all columns from the table bank\_churn. Additionally, it includes a subquery that retrieves the ExitCategory from the exitcustomer table based on a condition matching ExitID with Exited from the bank\_churn table. The result includes an extra column named ExitCategory. ***Output:*** A screenshot of a computer

Description automatically generated

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

The dataset does not contain any missing values. The only change made to the dataset was converting the datatype of the BankDOJ column to a date datatype.

**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”.**

*SELECT ci.CustomerId,ci.Surname,*

*MAX(ac.ActiveCategory) AS ActiveCategory*

*FROM customerinfo ci*

*JOIN bank\_churn b ON ci.CustomerId = b.CustomerId*

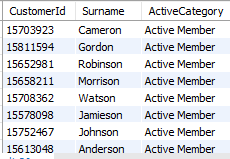
*JOIN activecustomer ac ON b.IsActiveMember = ac.ActiveID*

*WHERE ci.Surname LIKE '%on'*

*GROUP BY ci.CustomerId,ci.Surname;*

This SQL query selects unique combinations of CustomerId and Surname from the customerinfo table where the Surname contains "on". It retrieves the maximum ActiveCategory associated with each combination from the activecustomer table.

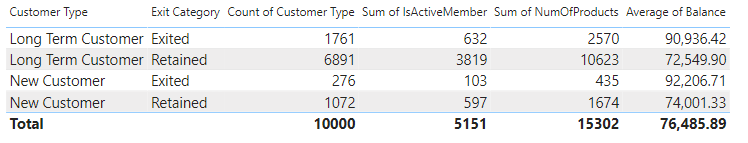
***Sample Output:***



This query groups results by CustomerId and Surname, selecting the highest ActiveCategory for each group, ensuring unique CustomerId entries in the output.

***Subjective Questions:***

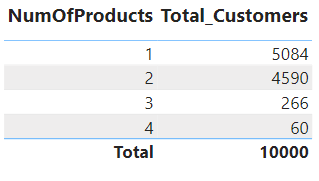
**1. Customer Behavior Analysis: What patterns can be observed in the spending habits of long-term customers compared to new customers, and what might these patterns suggest about customer loyalty?**

****

* Long term customers who have retained in the bank have purchased more number of products when compared to new customers.
* It is to be noted that the Number of customers and their Activeness are more in this category.
* The reason behind the Low average balance among other categories might be because they have purchased more products.

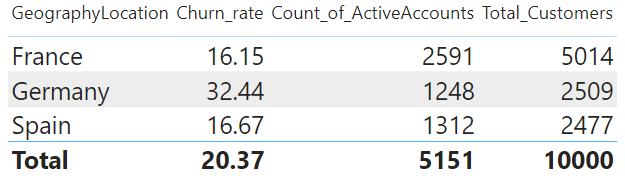
This indicates that long-term customers have developed deeper relationships with the bank over time and are more likely to utilize multiple products and services, which suggests a deeper level of engagement and loyalty to the bank.

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

****

There are high numbers of customers with only one or two products. This suggests that there is a strong association between the number of products a customer holds and their possession with the bank. Targeted marketing campaigns and personalized recommendations can be tailored based on customers' existing product holdings to encourage them to expand their relationships with the bank.

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

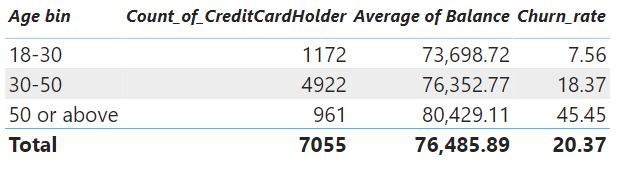
****

* Germany has the highest churn rate at 32.44%, followed by Spain at 16.67%, and France at 16.15%. Churn rates vary across geographic regions, indicating potential differences in economic conditions, customer behavior, or competitive landscapes.
* France and Spain have relatively similar counts of active accounts despite differences in churn rates.

In regions with high churn rates, bank may need to focus on improving customer retention strategies, such as enhancing customer service, offering competitive rates, or introducing loyalty programs.

Similarly, in regions with lower counts of active accounts, bank may need to invest in marketing efforts, product innovation, or partnerships to attract new customers and expand their market share.

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

****

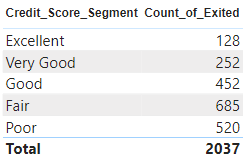
* Customers aged 50 or above demonstrate a significantly higher churn rate compared to younger age groups, indicating that they pose a higher financial risk to the bank.
* Despite having a higher average balance, older customers are more likely to churn, suggesting that factors beyond account balance influence their decision to leave the bank.
* It suggests that older customers warrant special attention in risk management strategies to mitigate potential losses associated with churn.

**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?**

|  |  |
| --- | --- |
|  |  |

* There are noticeable differences in the sum of estimated salaries across France, Germany, and Spain. France and Germany generally have higher total estimated salaries compared to Spain, indicating potential economic disparities between these regions.
* While France and Germany have higher total estimated salaries, the tenure patterns vary across different geographic locations. France and Germany show a similar trend of increasing estimated salaries up to a certain tenure (around 5 years) before declining for longer tenures. In contrast, Spain exhibits a similar trend but with lower estimated salaries overall.
* The analysis of tenure distribution suggests that customers in France and Germany tend to have longer relationships with the bank, as indicated by their tenure extending up to 7 years. This may imply higher customer loyalty or satisfaction levels in these regions compared to Spain.

**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?**

****

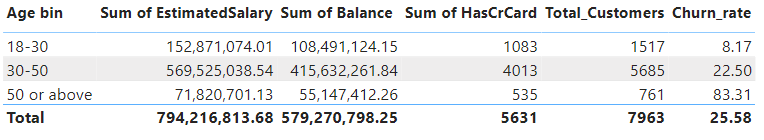
From the table, it is observed that the credit score segment of *Excellent and Very good* has the least count of exited customers which shows that the improvement of credit score impacts in the retention rate of the customers. By conducting Marketing Campaigns the Credit score of the customer will increase which will directly impact on the retention rates.

Toeffectively assess the impact of marketing campaigns on customer retention and acquisition within the dataset, additional information that would be helpful includes:

* Data on the customer journey from initial awareness to conversion, including touchpoints and interactions with marketing channels.
* Information on the cost associated with each marketing campaign, including ad spend, creative production costs, and other expenses.
* Feedback from customers regarding their experience with marketing campaigns, perceived value, and factors influencing their decision-making.
* Insights into competitor marketing strategies, messaging, and market positioning to understand the competitive landscape and potential impact on campaign effectiveness.

By combining these additional information, a comprehensive analysis can be conducted to evaluate the effectiveness of marketing campaigns on customer retention and acquisition.

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

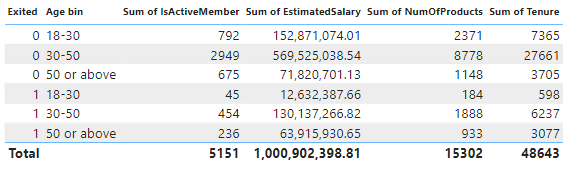


* The churn rate is notably higher for customers aged 50 or above compared to other age groups. This indicates that older customers may be more likely to leave the bank.
* Customers who exited the bank have a significant total balance and estimated salary compared to others. This suggests that financial factors may play a role in their decision to leave.

Some of the possible reasons for leaving:

* Dissatisfaction with banking services or customer experiences.
* Inadequate credit card benefits or rewards.
* Financial instability or changes in personal circumstances.
* Better offers or services from competitors.
* Lack of tailored products or solutions for older customers.

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



* The sum of tenure for retained customers (0) is significantly higher compared to exited customers (1). This suggests that customers who have been with the bank for longer periods are less likely to leave.
* The sum of the number of products for retained customers (0) is higher compared to exited customers (1) across all age groups. This indicates that customers who use more products/services offered by the bank are less likely to churn.
* The sum of IsActiveMember for retained customers (0) is higher compared to exited customers (1) across all age groups. This suggests that customers who are active members are less likely to churn.
* The sum of estimated salary for retained customers (0) is significantly higher compared to exited customers (1) across all age groups. This implies that customers with higher estimated salaries are less likely to leave the bank.
* Based on the provided data, 'Tenure', 'NumOfProducts', 'IsActiveMember', and 'EstimatedSalary' appear to be important factors for predicting if a customer will leave the bank. Customers with longer tenure, higher usage of products/services, active membership status, and higher estimated salaries are less likely to churn.

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

SELECT g.GenderCategory,

CASE

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

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

ELSE '50 or above'

END AS AgeBin,

COUNT(\*) AS CustomerCount,

AVG(bc.balance) AS AverageBalance

FROM customerinfo c

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

JOIN activecustomer ac ON bc.IsActiveMember = ac.ActiveID

JOIN Gender g ON c.GenderId = g.GenderId

JOIN creditcard cc ON bc.HasCrCard = cc.CreditID

JOIN exitcustomer ec ON bc.Exited = ec.ExitID

GROUP BY g.GenderCategory,

CASE

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

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

ELSE '50 or above'

END;

SELECT

-- Demographic Segmentation

CASE

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

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

WHEN ci.Age > 50 THEN '50+'

ELSE 'Unknown'

END AS Age\_Group,

g.GeographyLocation,

gi.GenderCategory AS Gender,

-- Account Details Segmentation

CASE

WHEN bc.Balance < 10000 THEN 'Low Balance'

WHEN bc.Balance >= 10000 AND bc.Balance < 50000 THEN 'Medium Balance'

WHEN bc.Balance >= 50000 THEN 'High Balance'

ELSE 'Unknown'

END AS Balance\_Category,

cc.Category AS Credit\_Card\_Status

FROM customerinfo ci

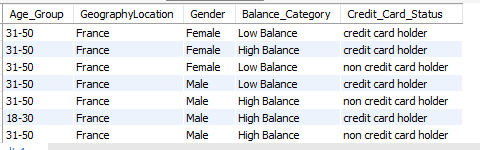
JOIN bank\_churn bc ON ci.CustomerId = bc.CustomerId

JOIN geography g ON ci.GeographyID = g.GeographyID

JOIN gender gi ON ci.GenderID = gi.GenderID

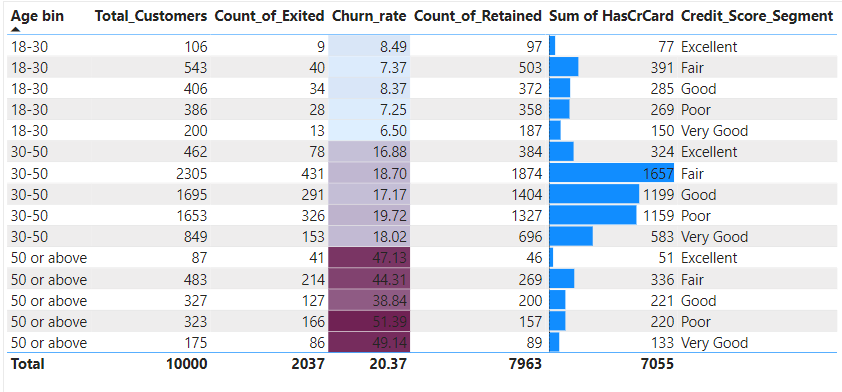
JOIN creditcard cc ON bc.HasCrCard = cc.CreditID;

The SQL query segments customers based on demographic attributes like age group, geography, and gender, as well as account details such as balance category and credit card status. It utilizes CASE statements to categorize customers into appropriate groups and performs joins across multiple tables to gather necessary information. The resulting output provides a segmented view of customers, facilitating analysis of their characteristics and behaviors across different segments.

****

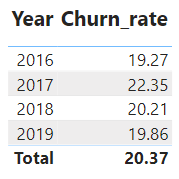
**10. How can we create a conditional formatting setup to visually highlight customers at risk of churn and to evaluate the impact of credit card rewards on customer retention?**

* In Power Bi table visualization is selected to apply conditional formatting. Right-click on the field or measure to format and choose "Conditional formatting" from the context menu.
* In the Conditional formatting pane, the rules are defined based on Chun rate and Credit card criteria’s.
* Specify the formatting options for each rule. Different formatting styles are chosen such as colors and Data bar is used to represent different levels of risk or impact.
* Once the rules are defined in the formatting options, click "Apply" or "OK" to apply the conditional formatting to the visualization.



* The visualization enables comparison across different age groups to understand churn rates, credit card ownership, and credit score distributions. It highlights potential trends or correlations between age, credit card usage, credit scores, and churn rates, providing valuable insights for decision-making and targeting specific customer segments for retention efforts or marketing campaigns.
* From the table chart, it is evident that the agebin 50+ has the highest churn rate which implies that they are likely prone to leave the bank. Also, this age groups has the least number of credit cards.
* Similarly, customers having credit cards are less likely to leave the bank which shows the impact of credit card rewards in the customers.

**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?**

**A screenshot of a table

Description automatically generated**

The churn rates vary across different age groups each year. Generally, older customers (50 or above) consistently exhibit higher churn rates compared to younger age groups (18-30 and 30-50).While there are fluctuations in churn rates from year to year, there isn't a significant upward or downward trend in any specific age group.

**Strategies to Decrease Churn Rate:**

* The bank can conduct marketing and engagement campaigns tailored to the unique needs and preferences of different age groups, such as offering retirement planning services and personalized financial advice for older customers.
* Specialized customer support services, including dedicated helplines for senior citizens or personalized assistance for complex financial transactions, can be provided to improve satisfaction and retention among older customers.
* Conducting financial literacy workshops and educational seminars targeted at different age groups can empower customers with knowledge and skills to manage their finances effectively, reducing the likelihood of churn.
* Implementing loyalty programs and retention incentives customized for each age group, including rewards, discounts, and exclusive benefits, can enhance customer loyalty and decrease churn rates.
* Establishing feedback mechanisms to gather insights directly from customers about their needs, preferences, and pain points can inform continuous refinement and optimization of retention strategies for each age group.

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

|  |
| --- |
|  |

|  |
| --- |
|  |

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

If the objective and subjective questions weren't provided, I would approach the problem by :

* First understanding the dataset thoroughly, including its structure, variables, and relationships between them.
* I would perform exploratory data analysis to find out the uncover patterns, trends, and insights within the data.
* I would involve visualizing data through charts and graphs, and identifying any correlations or relationships between variables.
* I would identify customer churn rate, customer retention rate, average account balance, demographics distribution, and product usage metrics.
* I would then use SQL queries to calculate these metrics and present them in a meaningful way.
* After that, I would conduct deeper analysis to answer specific questions or address particular objectives, such as identifying factors influencing customer churn, understanding customer segmentation based on demographics or behavior.

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

In Power BI, you can rename columns in the query editor. Here's how you can modify the name of the "HasCrCard" column to "Has\_creditcard":

Open your Power BI desktop file.

Go to the "Home" tab and click on "Transform Data" to open the query editor.

In the query editor, find the table "Bank\_Churn" from the list of queries on the left-hand side.

Select the "Bank\_Churn" query to display its columns and data.

Right-click on the "HasCrCard" column header and select "Rename".

Replace "HasCrCard" with "Has\_creditcard" and press Enter to confirm.

Once you've renamed the column, click on "Close & Apply" to save your changes and close the query editor.