GHKVJVJH

**Bank CRM Data Analysis**

**By S SHABIN**

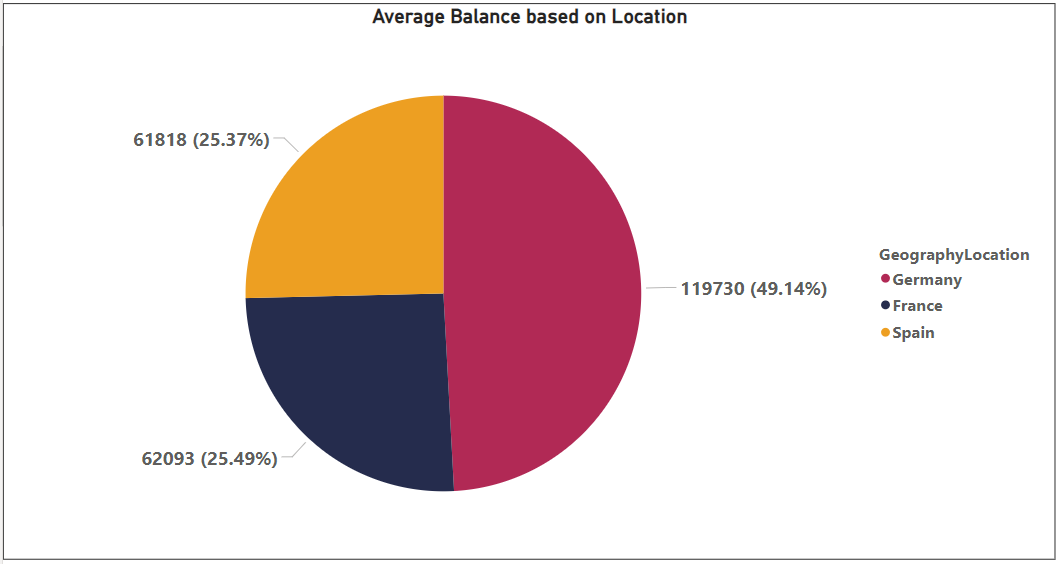
**DS NOVEMBER 2023**

**Objective Questions**

**Q1. What is the distribution of account balance across different regions?**

Answer:

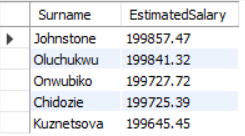
The distribution of average account balance across different regions is shown in the chart below.



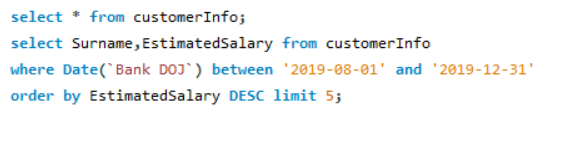
**Q2. Identify the top 5 customers with the highest Estimated Salary in the last quarter of the year. (SQL)**

Answer:

The top 5 customers with the highest Estimated Salary in the last quarter of the year are given in the table given below.



The query used:



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

Answer:

* The average number of products used by customers who have credit cards is 1.5314.

The Query used –

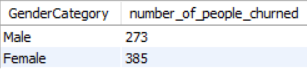
A close-up of a computer screen

Description automatically generated

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

Answer:

* The number of males churned in the most recent year (2019) is 273 and the number of females churned in the most recent year (2019) is 385**.**



The Query used-

**A screenshot of a computer code

Description automatically generated**

**Q5. Compare the average credit score of customers who have exited and those who remain. (SQL)**

Answer:

The average credit score of customers who have exited and those who remain can be compared from the below table .

* The credit score for exited customers is 645.3515 and the credit score for retained customers is 651.8532.

A screenshot of a credit score

Description automatically generated

The Query used is-

**A computer code with black text

Description automatically generated**

**Q6. Which gender has a higher average estimated salary, and how does it relate to the number of active accounts? (SQL)**

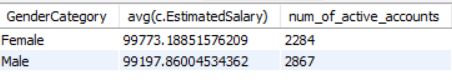
Answer:

From the table below it is evident that

* Female has the highest estimated salary of 99773.1885 than that of males having estimated salary of 99197.8600.

We can also observe that

* females with higher estimated salary have lesser number of active accounts compared to males who has lower estimated salary .females have 2284 active accounts and males have 2867 active accounts.



The Query used is-

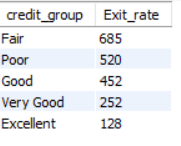
A screenshot of a computer program

Description automatically generated

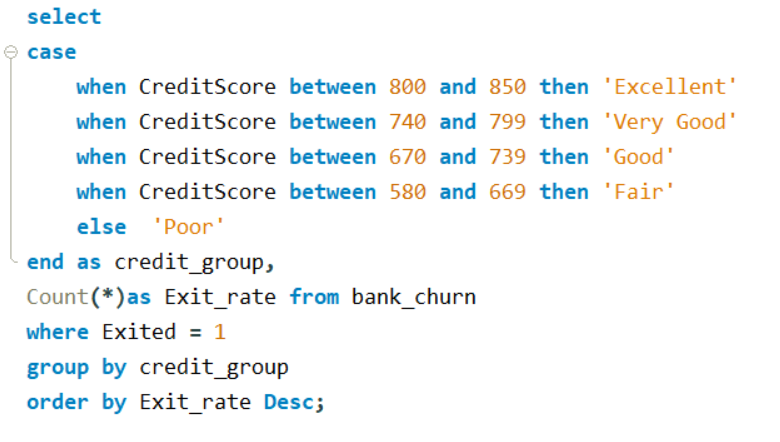
**Q7. Segment the customers based on their credit score and identify the segment with the highest exit rate. (SQL)**

Answer:

The Customers where segmented into 5 categories named Fair, Poor,Good,Very Good, Excellent .The exit rates were calculated based on the categories. From the table the category “Fair” has the highest exit rate of 685 people.



The Query used is-

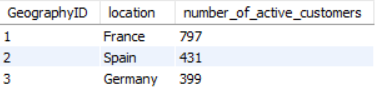
****

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

Answer:

From the table below its clear that

* The highest number of active customers with tenure greater than 5 is for France with 797 active customers.



The Query used is :

**A screenshot of a computer program

Description automatically generated**

**Q9. What is the impact of having a credit card on customer churn, based on the available data?**

Answer:

* The number of customers churned who had a credit card is 1424 and customers who did not have a credit card is 613.

Thus ,we can infer that churning is more frequent among customers with credit card than those without creditcard.

A screenshot of a computer

Description automatically generated

The Query used is :

**A computer screen shot of a credit card

Description automatically generated**

**Q10. For customers who have exited, what is the most common number of products they had used?**

Answer:

From the table below it is clear that

* The most common number of products used by exited customers is ‘1’.
* Number of products 1 has been used 1409 times.

A screenshot of a computer

Description automatically generated

The Query used is :

**A white background with black text

Description automatically generated**

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

Answer:

The trend of customers joining over time can be understood from the following table .The table below visualizes the number of customers joining over each month of each year.

* The maximum number of people are joining in 3 months September,November,December each year .
* The lowest number of people joining is in the 2 months, January, February.

These patterns are visible in all years for which data is available.



A screenshot of a calendar

Description automatically generated

The Query used is :

**A computer code with black text

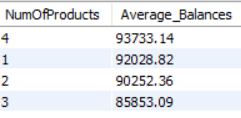
Description automatically generated with medium confidence**

**Q12. Analyze the relationship between the number of products and the account balance for customers who have exited.**

Answer:

From the table below it is understood that

* The churned customers with number of Products ‘4’ has the highest average balance 93733.14 ,followed by number of products ‘1’ with average balance 92028.82.
* The churned customers with number of products ‘3’ has the least average balance of all 85853.09.



The Query used is :

**A screenshot of a computer code

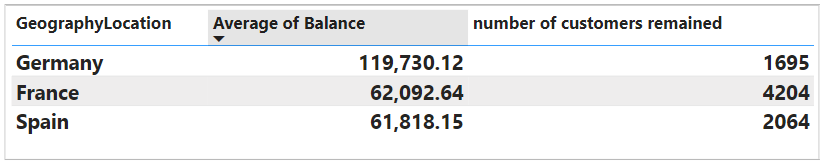
Description automatically generated**

**Q13. Identify any potential outliers in terms of balance among customers who have remained with the bank.**

Answer:

The only potential outlier in terms of balance among customers who have remained with the bank is observed when the data is analyzed in terms of location.

Here the average balance value is abnormally high for Germany alone when compared to other regions.



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

Answer:

There are mainly 7 tables given in the dataset which are given below-

1.ActiveCustomer

2.Bank\_Churn

3.CreditCard

4.CustomerInfo

5.ExitCustomer

6.Gender

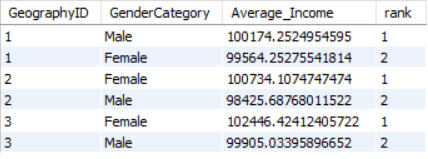
7.Geography

Out of these 7 tables ,the CustomerInfo table contains categorical variables like surname, age,gender,Bank DOJ etc.

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

Answer:

The gender wise average income of male and female in each geography id are shown below in the table . Also, the gender has been ranked according to the average value as well.



The Query used is :

**A screenshot of a computer code

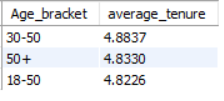
Description automatically generated**

**Q16. 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:

The average tenure of people who have exited in each age bracket has been shown in the table below.

* The age brackets ‘30-50’ has the highest average tenure of 4.8837.



The Query used is :

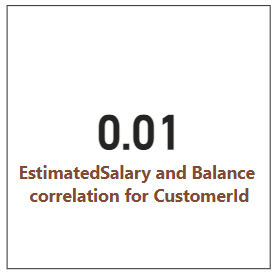
**A screenshot of a computer code

Description automatically generated**

**Q17. Is there any direct correlation between salary and balance of the customers? And is it different for people who have exited or not?**

Answer:

* There is small positive correlation between the salary and the balance of the customers.
* The correlation function for salary and balance over customerId has value 0.01 .



* There is a small negative correlation between salary and balances for customers who have exited.
* The correlation coefficient value is -0.01.

A white background with black text

Description automatically generated

* There is a small positive correlation between salary and balances for customers who has been retained.
* The correlation coefficient value is 0.02.

A white background with black text

Description automatically generated

**Q18. Is there any correlation between salary and Credit score of customers?**

Answer:

There is no correlation between salary and credit score of customers. The value of estimated salary and credit score correlation for customerId is 0.

A close-up of a number

Description automatically generated

**Q19. Rank each bucket of credit score as per the number of customers who have churned the bank.**

Answer:

The credit score has been categorized as

* Excellent: 800–850
* Very Good: 740–799
* Good: 670–739
* Fair: 580–669
* Poor: 300–579

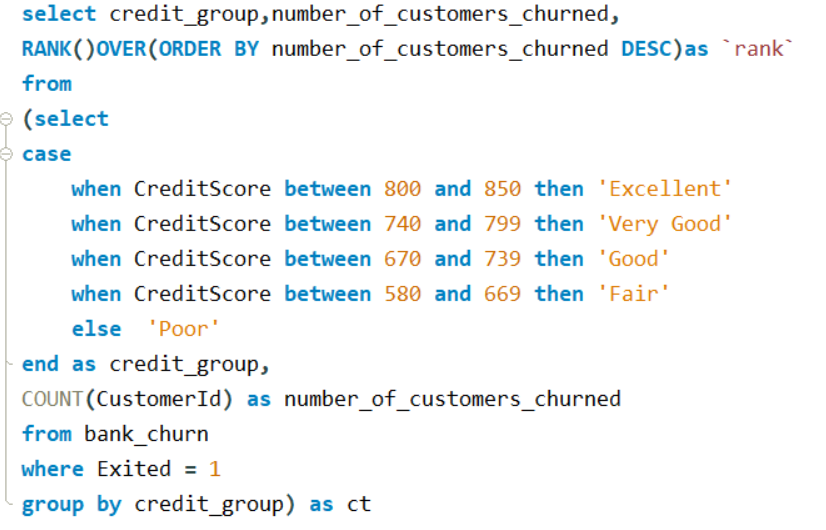
The number of churned customers were found corresponding to each category where ranked as per the number of churned customers .

* It was observed that ‘Fair’ category has the highest number of churned customers 685 .
* ‘Excellent’ category has the lowest number of churned customers 128.

A screenshot of a computer

Description automatically generated

The Query used is :

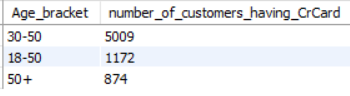
****

**Q20. 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.**

Answer:

Number of customers who had a credit card based on age buckets are shown below in table 1.

* The highest number of customers having a credit card is in the age group of ’30-50’ which equals 5009.



The age group who has less than average number of credit cards per group is shown in another table shown below.

* There are only two age groups whose average is less than the average number of credit cards per group. Which are ’18-50’ and ‘50+’.

A close-up of a computer screen

Description automatically generated

The Query used is :

Table 1

**A screenshot of a computer program

Description automatically generated**

**Table 2**

A screenshot of a computer code

Description automatically generated

**Q21. Rank the Locations as per the number of people who have churned the bank and average balance of the learners.**

Answer:

The table below shows the ranking of locations as per the requirements.

* Germany ranks first with highest no of churned people 814 and highest average balance of 120361.075.
* Spain ranks last with 413 churned people and 72513.35244 average balance



The Query used is :

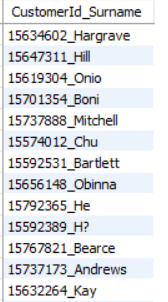
**A screenshot of a computer program

Description automatically generated**

**Q22. As we can see that the “CustomerInfo” table has the CustomerID and Surname, now if we must 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:

The column can be created by using concat function to join 2 different columns. The table below shows the required column “CustomerID\_Surname”.



The Query used is :

****

**Q23. Without using “Join”, can we get the “ExitCategory” from ExitCustomers table to Bank\_Churn table? If yes do this using SQL.**

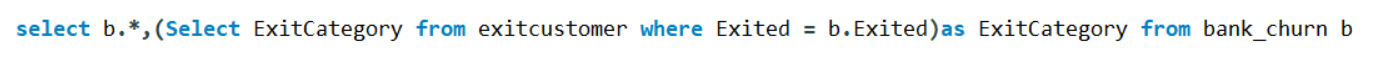
Answer:

Yes, the “ExitCategory” column from ExitCustomers table can be brought into Bank\_Churn table Without using ‘join’.

A screenshot of a data

Description automatically generated

The Query used is :

****

**Q24. Were there any missing values in the data, which tool did you replace them and what are the ways to handle them?**

Answer:

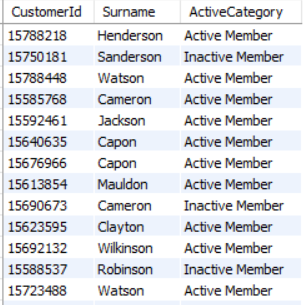
No ,there was no missing data in the given data set.

If there was any missing data, I would have used either Microsoft excel or would have done data transformation in power bi to identify and replace missing values.

**Q25. 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:

The table below shows the required columns.



The Query used is :

**A screenshot of a computer program

Description automatically generated**

**Subjective Questions**

**Q1. 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?**

Solution:

To analyze the customer behavior for long term and new customers , joining year was used . customers with joining year 2016 can be called long-term customers and those with joining year 2019 can be called new customers.

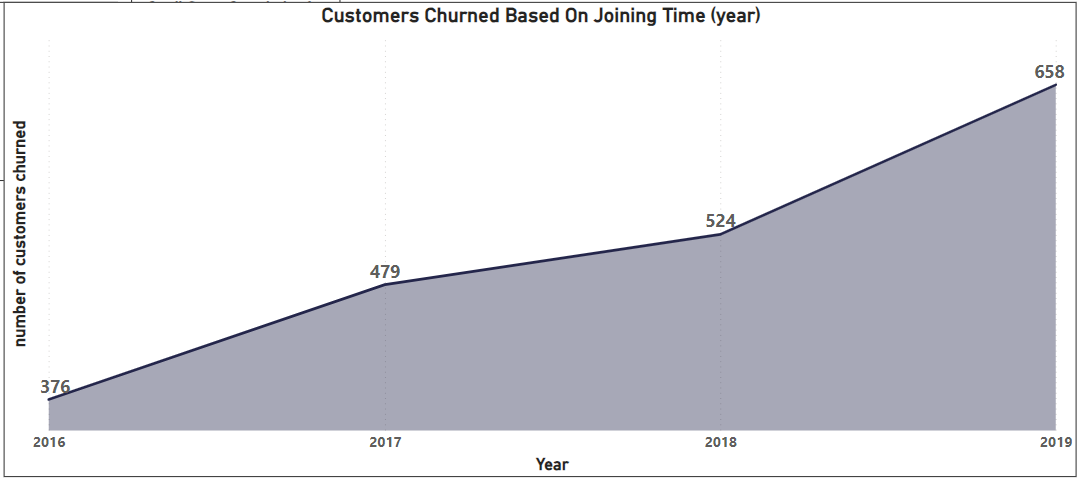
When spending habits corresponding to long-term customers and new customers were analyzed ,

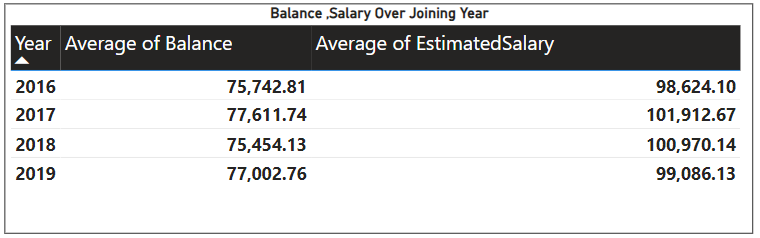
* It was observed that the spending percentage for long term customers(2016) (23%) were higher than that of new customers(2019)(22%).
* The average salary and average balance values of new customers were higher than those of long-term customers.

When the number of customers churned where analyzed ,it shows that

* The new customers had the maximum number of churning (658) when compared to long term-customers(376).
* Thus, it could be inferred that long term customers with higher spending rates have a lower tendency to churn when compared to new customers with lower spending rates.

The following charts and tables give details about the analysis.

****

****

**A graph showing the number of years

Description automatically generated**

**A table with numbers and a number on it

Description automatically generated**

**For further details refer the power bi file : Bank\_CRM\_Data\_Analysis.pbix**

**Q2. Product Affinity Study: Which bank products or services are most commonly used together, and how might this influence cross-selling strategies?**

**Solution:**

**To understand the product affinity, we require data related to the bank products and services, and some specific sales details . Since this data is not sufficient ,the above analysis cannot be done.**

**Q3. Geographic Market Trends: How do economic indicators in different geographic regions correlate with the number of active accounts and customer churn rates?**

**Solution:**

**Geographic market trends are crucial for banks .when various economic indicators were analyzed based on different geographic regions -**

* **It was observed that they have a direct correlation with the customer churning .**

**When the average balance is analyzed -**

* **it is observed that the location with the highest average balance is Germany 119730, and the lowest average balance is for Spain with 61818.**

**Similarly, when average salary is analyzed ,**

* **It shows that highest average salary is for Germany with 101113 and the lowest average salary is for Spain with 99441.**

**When the number of active customers were considered,**

* **Germany was the country with least number of active customers which is 1248.**

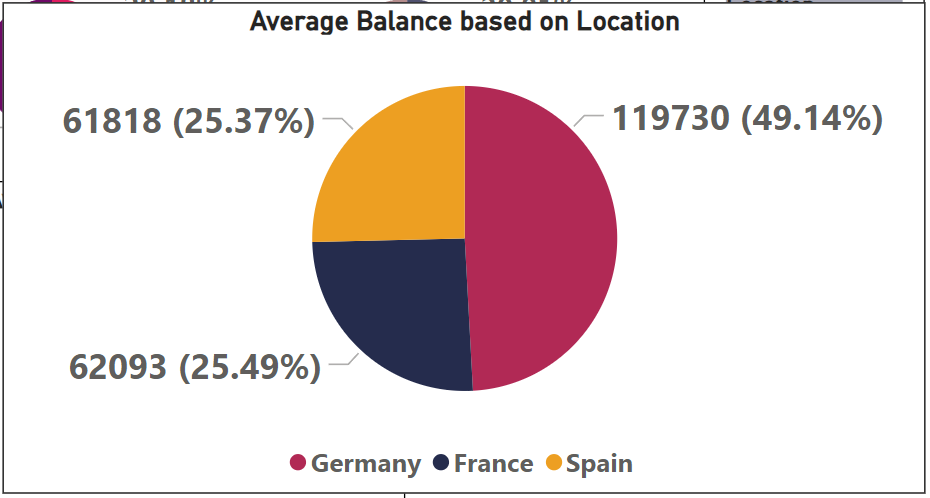
**The scenario becomes clearer when churning rate is analyzed as Germany has the highest number of churned customers.**

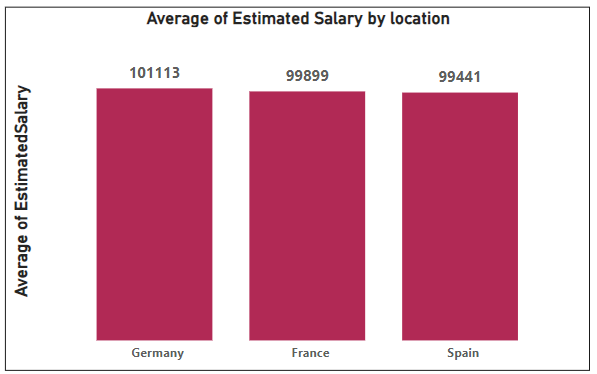
**This shows a trend that the region with highest salary and balance are churning the most where non active members are the most.**

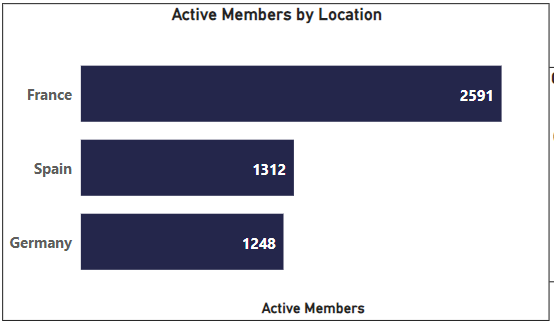
**This trend can be reconfirmed when the specific churning rate for non-active members is observed.**

* **Here also Germany is the location with the highest number of non-active members churning which is 518 members with a churn rate of around 41% of non-active members.**

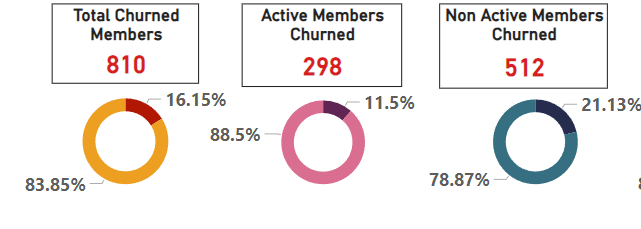
**The charts and tables given below visualize the above analysis.**

****

****

****

**FRANCE**

****

**GERMANY**

**A close-up of a graph

Description automatically generated**

**SPAIN**

**A close-up of a graph

Description automatically generated**

**For further details refer the power bi file :Bank\_CRM\_Data\_Analysis.pbix**

**Q4. Risk Management Assessment: Based on customer profiles, which demographic segments appear to pose the highest financial risk to the bank, and why?**

**Solution:**

**Banks have customers from various segments of society, and they belong to various demographic segments. On analyzing the given data-**

* **The highest number of customers are churned from one segment that is customers in the age group of 30 -50.**
* **There are around 1350 customers churned which is 66.27% of the entire churned customers.**
* **The average salary for the age group ‘30-50’ is 100371.81 and the average balance for age group ‘30-50’ is 76514.60 .**
* **The age group of 30-50 has the second highest spending percentage of 24%.**

**Thus, customers in the age group ‘30-50’ pose the highest financial risk to the bank.**

**A pie chart with numbers and a number of people

Description automatically generated**

**A screenshot of a graph

Description automatically generated**

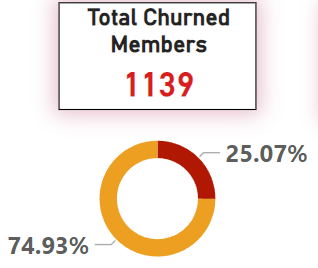
**A table with numbers and text

Description automatically generated**

**When gender wise observations are made**

* **higher number of customers churned are females 1139,which is 25.07% of total churned customers .**
* **whereas there are 898 churned male members which makes 16.46 % of total churned.**

**Female Male**

** A chart with numbers and a circle

Description automatically generated**

**For further details refer the power bi file: Bank\_CRM\_Data\_Analysis.pbix**

**Q5. Customer Lifetime Value Forecast: How would you use the available data to model and predict the lifetime (tenure) value of different customer segments?**

Solution:

The available data was analyzed, and the lifetime value (tenure) was compared with various customer segments and churned rate was calculated.

* It showed that churn rate was maximum for those customers who had a tenure of 5 years (21.91%).
* secondly for customers with tenure 6 years(20.57)
* least churn rate for customers with tenure 7 years(19.14).

Similarly, analysis was done based on estimated salary and balance against tenure and corresponding spending percentage were calculated .

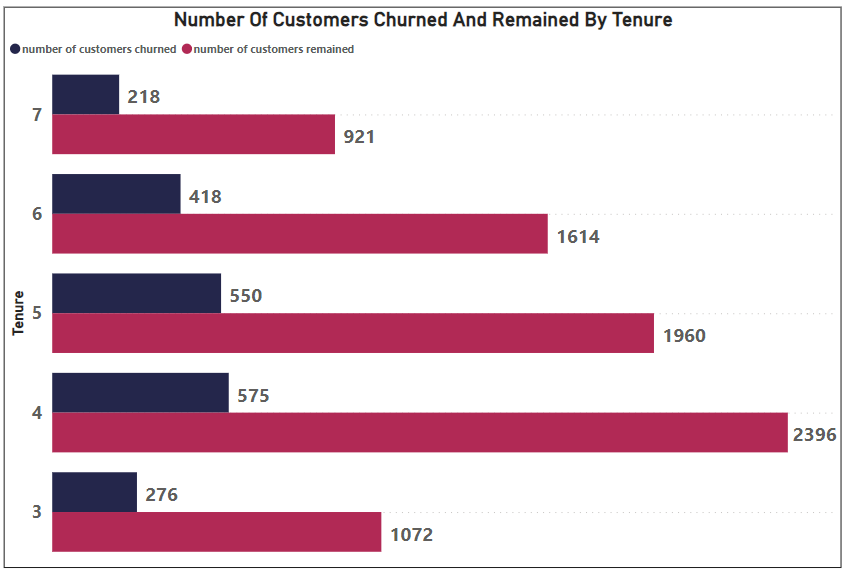
* It was observed that maximum spending percentage was that for customers with tenure of 5 years(27%)
* secondly for customers with tenure of 6 years(23%)
* the lowest for tenure of 7 years( 21%).

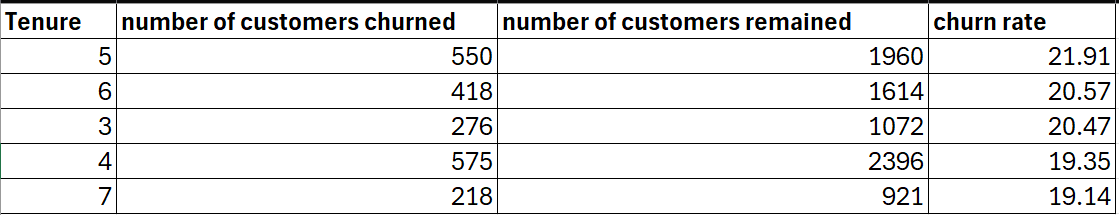
The average salary also shows this type of trend with values in the table below.

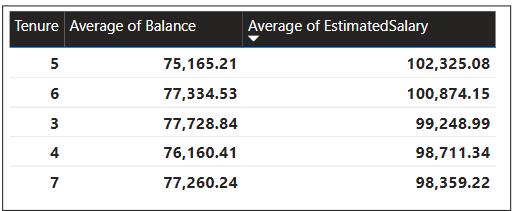
Form the analysis it is understood that

* The churn rate and average salary have a pattern corresponding to the tenure.
* This trend can be utilized to predict the tenure value of different customer segments.

The tables and chart given below contain the outcomes of analysis.

****

****

****

**A table with numbers and letters

Description automatically generated**

**For further details refer the power bi file:Bank\_CRM\_Data\_Analysis.pbix**

**Q6. 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?**

Solution:

To assess the impact of marketing campaigns on customer retention and acquisition, the dataset provided is insufficient. To analyze such trends few more specific data is required such as:

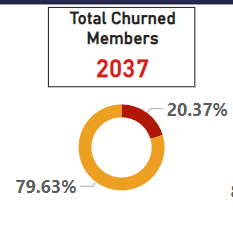
1. Details about the products and services for which the campaign is for.
2. Separate customer data for before and after the campaign.
3. Details about targeted age groups.

**Q7. Customer Exit Reasons Exploration: Can you identify common characteristics or trends among customers who have exited that could explain their reasons for leaving?**

**Solution:**

**When the given data was analyzed , it was observed that churning of customers has been happening significantly for the bank.**

* **Around 2037 customers have churned based on the given data at a churn rate of 20.37%.**

****

**When observed carefully it was found that there are a few common characteristics or trends among the churning customers.**

**These trends are explained below.**

1. **Active and non-active members**

**When churning among active and non-active members are analyzed ,it shows that**

* **The number of non-active members churned (1302) is significantly higher than those with active members (735).**
* **The churn rate among non-active members is 26.85% and churn rate among active members is 14.27%.**

**A close-up of a graph

Description automatically generated**

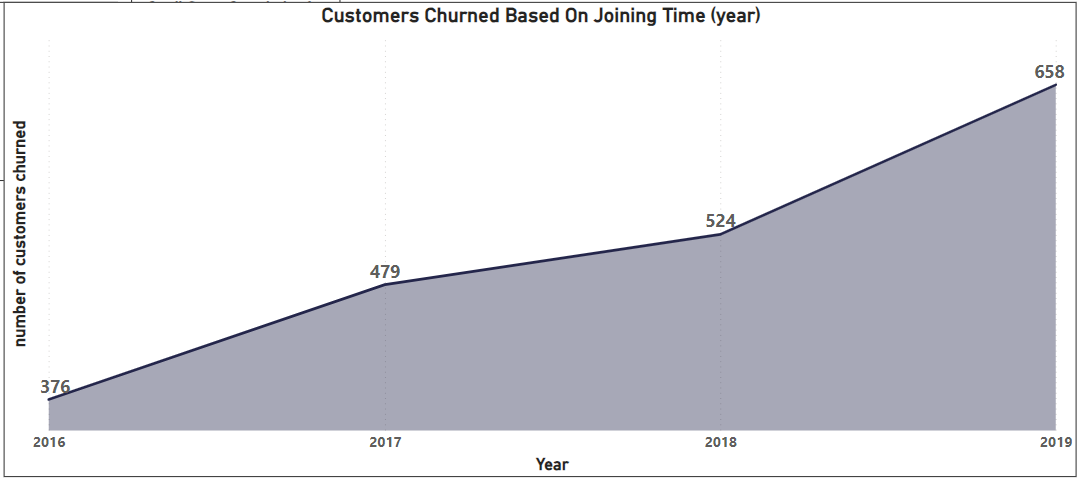
1. **Long term customers and new customers**

**When churning of customers based on their joining year are analyzed, it shows that the number of customers churned are higher in new customers when compared to long term-customers.**

**From the data, a trend is observed where churning has been increasing over the year.**

* **The number of customers churned is highest in 2019 with 658 and lowest in 2016 with 376.**

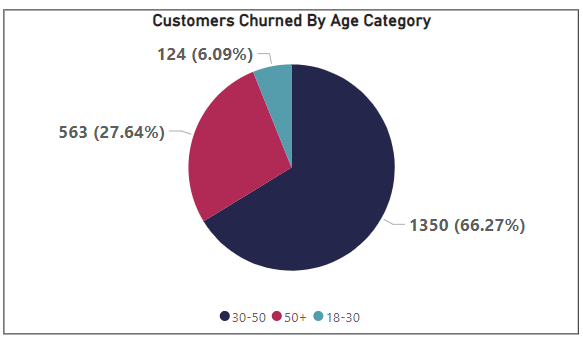
**A probable reason could be the availability of better options in recent years.**

****

**3.Age category**

* **The number of customers churned is highest among the age group of 30-50 years with 1350 churned customers at a churning rate of 66.27%.**

**The probable reason for churning in this age category could be unsatisfaction from the services and products offered.**

****

**4.Location**

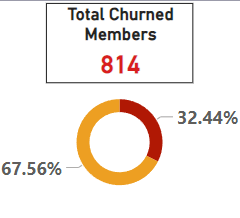
**From the given data analysis, it was understood that**

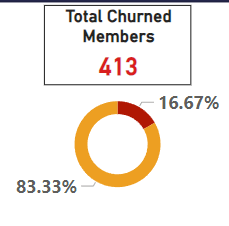
* **The maximum number of churnings takes place in Germany with 814 and churn rate of 32.44% and the lowest for Spain with 423 and churn rate of 16.67%.**

**Thus, we can infer that churning is also location specific.**

**The probable reason for this could be cultural and behavioral factors which are location specific.**

**Germany,France,Spain**

** A chart with numbers and a circle

Description automatically generated **

**5.Credit score category**

**A graph with numbers and text

Description automatically generated with medium confidence**

**From the analysis it is clear that**

* **The maximum churning of 395 customers happened for customers who have credit scores in the category “Fair”, that is they have a credit score between 800 and 850.**

**Q8. Are 'Tenure', 'NumOfProducts', 'IsActiveMember', and 'EstimatedSalary' important for predicting if a customer will leave the bank?**

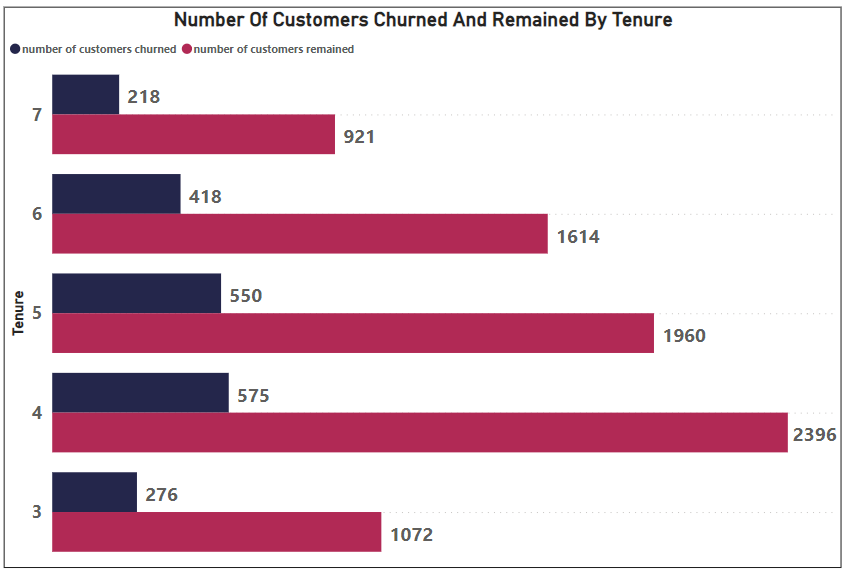
Solution:

Yes , Tenure of the customers, number of products used by customers, whether the customer is active or not and estimated salary all these factors are crucial in predicting if a customer will leave the bank or not.

Let’s consider each one separately .

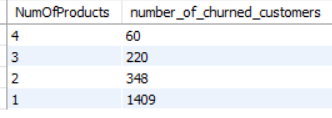
1.Tenure

Analyzing the data based on tenure help us understand churning rates for tenure and predict the probable time a particular segment of customers can exit .

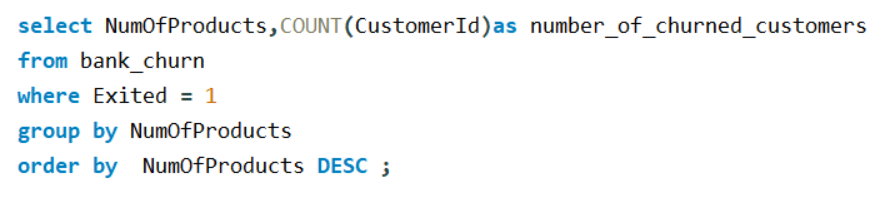
****

2.NumOfProducts

The number of churned customers corresponding to number of products used by them can be helpful in categorizing customers-based churn possibility.

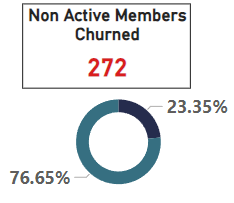


Query used is:

****

3.whether active or not active member

Analyzing the churning of customers based on whether the customer is an active member or not gives clarity about customers mindset. People who are not active are more likely to churn when compared to active customers .The table given below gives quantified value for the trend.

 A purple circle with numbers and a white rectangle

Description automatically generated

4.Estimated salary

Estimated salary is another crucial factor helpful in predicting the churning customer. Estimated salary alone or combined with other performance indicators can be used .

**Q9. Utilize SQL queries to segment customers based on demographics and account details.**

The customers were segmented based on demographics into 3 groups age between 18-30,30-50,50+ and account details corresponding to these categories were obtained using SQL query given below.

A screenshot of a number

Description automatically generated

Query used is:

A screenshot of a computer program

Description automatically generated

**Q10.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?**

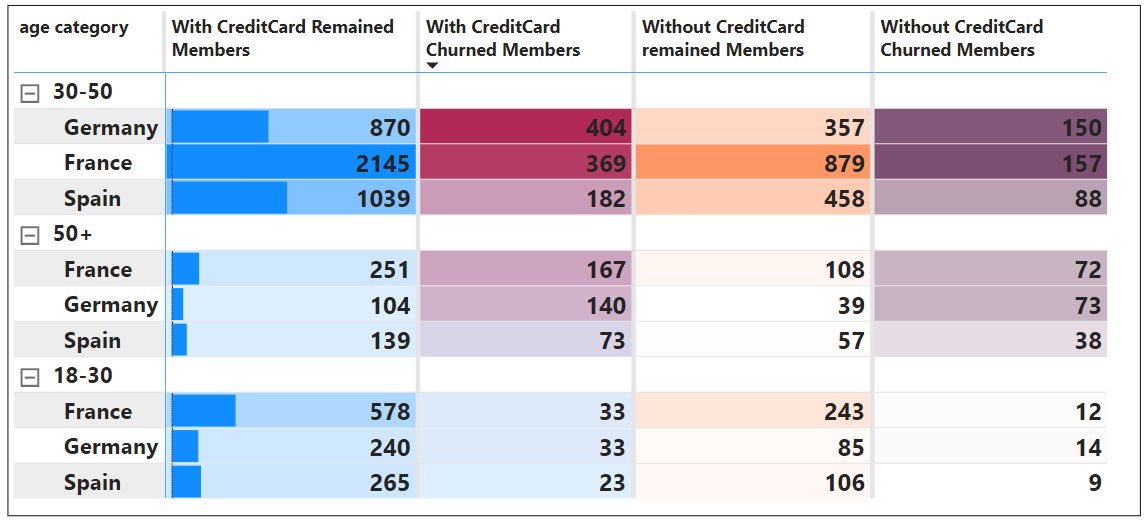
Solution:

A conditional formatting setup to visually highlight customers at risk .

In chart 1 we focus on analyzing the customers with credit cards.

* The red shaded column represents the category of customers who have credit cards and have exited .
* The column in blue contains the category of remaining customers with credit cards.
* From the chart it could be understood that maximum people churned with credit card is from age group of 30-50 and Germany location.
* So, the customers at risk of churning now are from the blue table corresponding to Germany(870) and then France(2145).

Chart 1-To analyze with credit card



In chart 2 we focus on analyzing the customers without credit cards.

* The purple column represents the category of customers who do not have credit cards and have exited.
* The column in orange contains the category of remaining customers without credit cards.
* From the chart it could be understood that maximum people churned without credit card is from age group of 30-50 and France location.
* So, the customers at risk of churning in customers without credit card is from the orange table corresponding to France (879) and then Germany(357).

Chart 2-To analyze without credit card

A screenshot of a data

Description automatically generated

**Q11. What is the current churn rate per year and overall, as well in the bank. Can you suggest some insights to the bank about which kind of customers are more likely to churn and what are the different strategies that can be used to decrease the churn rate.**

Solution:

As per the data, the overall Customers churned is 2037 which is 20.37% percentage of total customers. The number of customers churned in the recent year (2019) is 658 which is 32% of total customers churned.

Some insights on which kind of customers are more likely to churn are mentioned below.

1. Customers who are not active have a higher chance to exit the bank when compared to active customers.
2. Customers in the age group of 30 -50 years have the higher chance of churning.
3. Customer who uses sole product has higher chances of churning when compared to others.
4. Customers with tenure of 5 or 6 years have higher chances of churning when compared to others.
5. Customers in Germany has higher chances of churning in comparison

Strategies to decrease churn rate are as follows :

1. Improve positive customer-bank interactions

2. improving the services and products offered to the customers,

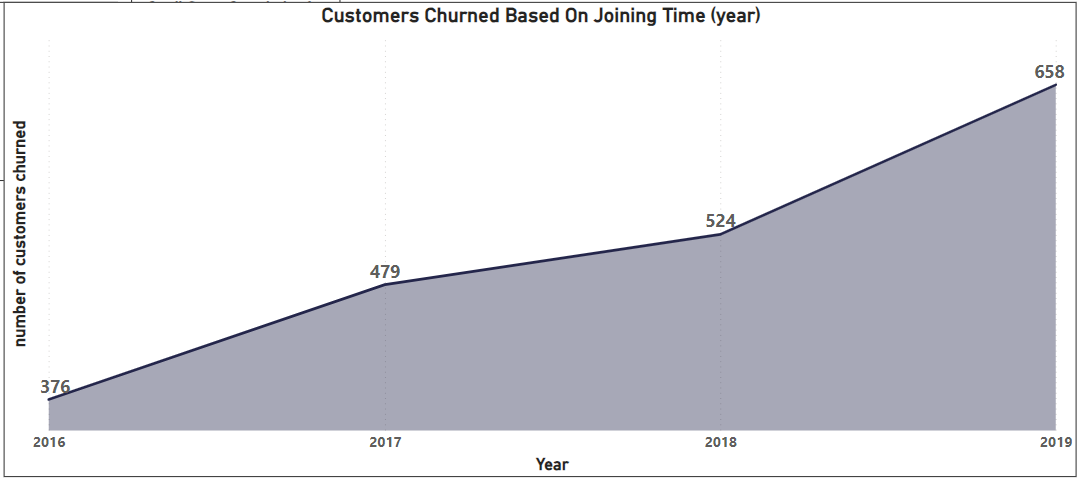
3. Enhancing customer service such that the query redressal of customers happens effortlessly and reduces any dissatisfaction among customers.

4.Introducing location and culturally inclusive banking products which may target customers from a specific locality.

The table and chart below have the details of the analysis.

**A screenshot of a graph

Description automatically generated**





**Q12.Create a dashboard incorporating all the KPIs and visualization related metrics. Use a slicer in order to assist in selection in the dashboard.**

Solution:

The below mentioned power bi file

BANK\_CRM\_DATA\_ANALYSIS.pbix

contains dashboard incorporating all the KPIs and visualization related metrics .This file has been submitted along with the document.

**Q13. How would you approach this problem, if the objective and subjective questions weren't given?**

Solution:

If the objective and subjective questions were not given ,I would have approached the problem in the following way-

* Firstly, I would analyze the data and find out churn rate and number of churned customers corresponding to each performance indicator.
* This would help in understanding how various performance indicators align with the churning rate.
* Once the churn rates are analyzed for basic factors, I would compare financial indicators such as account balance, salary, credit score with churned data and find out what financial decision resulted in churning.
* Next, I would analyze churned data based on location specific parameters and find out what components of the location influenced the churning.
* Now, I would analyze the long term and short-term customers data to find out the relation between tenure and churning rates.
* Finally, I would analyze to find out what effect the various products and services provided by the bank had on the churning.

**Q14. In the “Bank\_Churn” table how can you modify the name of “HasCrCard” column to “Has\_creditcard”?**

**Solution:**

**There are multiple ways to modify the name of the column from “HasCrCard” to “Has\_creditcard”.**

1. **It can be modified in excel /csv file before importing the files to SQL. This can be done by simply renaming the column.**
2. **In SQL we can do this by using the following Query**

**A close-up of a white background

Description automatically generated**