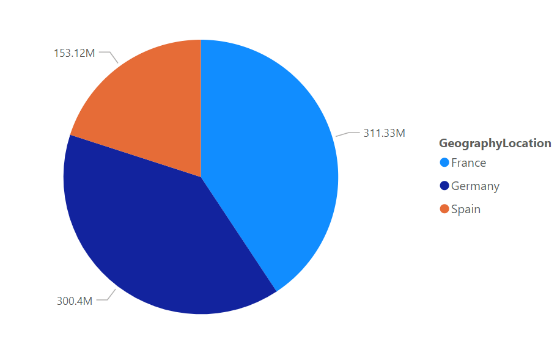
## **Capstone Project:**

## Analytical CRM Development for a Bank

Objective Questions

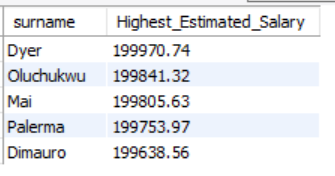
1. What is the distribution of account balance across different regions?

As we can see from the pie chart as France has highest account balance among all the country which is 311Million and Germany has second highest account balance which is 300million.

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

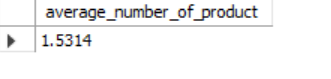
Answer- As we can see below in image that Dyer has highest Estimated salary among all the customers.

Dyer , Oluchukwu, Mai, Palerma, Dimaruro are the top 5 customers which has highest estimated salary.

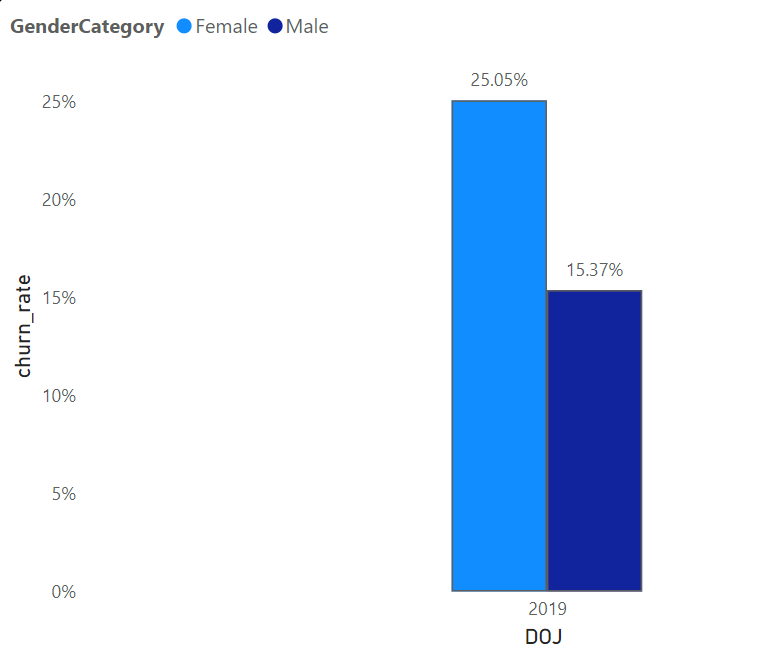


1. Calculate the average number of products used by customers who have a credit card.

Answer- The average number of product used by customer who have a credit card is 1.5314

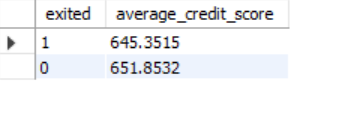


1. Determine the churn rate by gender for the most recent year in the dataset

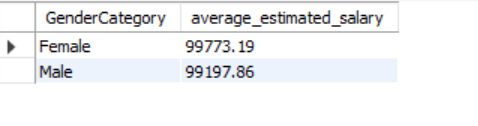


As from the from image, for the most recent year(2019) we have churn rate 25.05% for the male and 15.35% for female.

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

As we can see from the image customer who have remain in bank has higher average credit score (651.854) then customer who exited the bank(645.351).

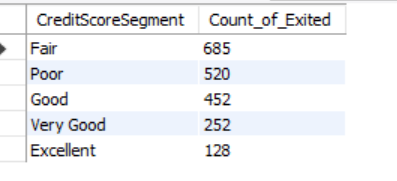
1. Which gender has a higher average estimated salary, and how does it relate to the number of active accounts? (SQL)

As we can see that female has highest estimated salary than male who has active account.

Male has estimated salary 99197.86

Female has estimated salary 99773.19

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

As we can see from the image we have 5 segment based on the credit score:

Excellent: 800–850

Very Good: 740–799

Good: 670–739

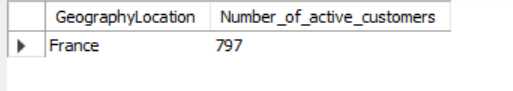
Fair: 580–669

Poor: 300–579

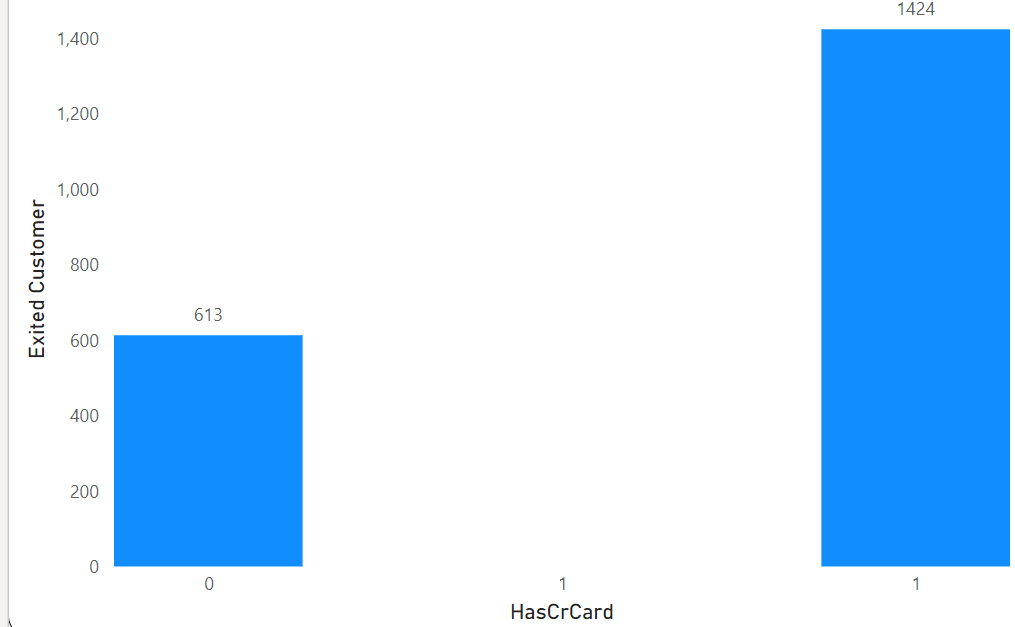
Fair Segment (580-669) has highest number of the customers who exited which is 685.

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

Answer- France has highest number of active customer with the tenure greater than 5 year.

d

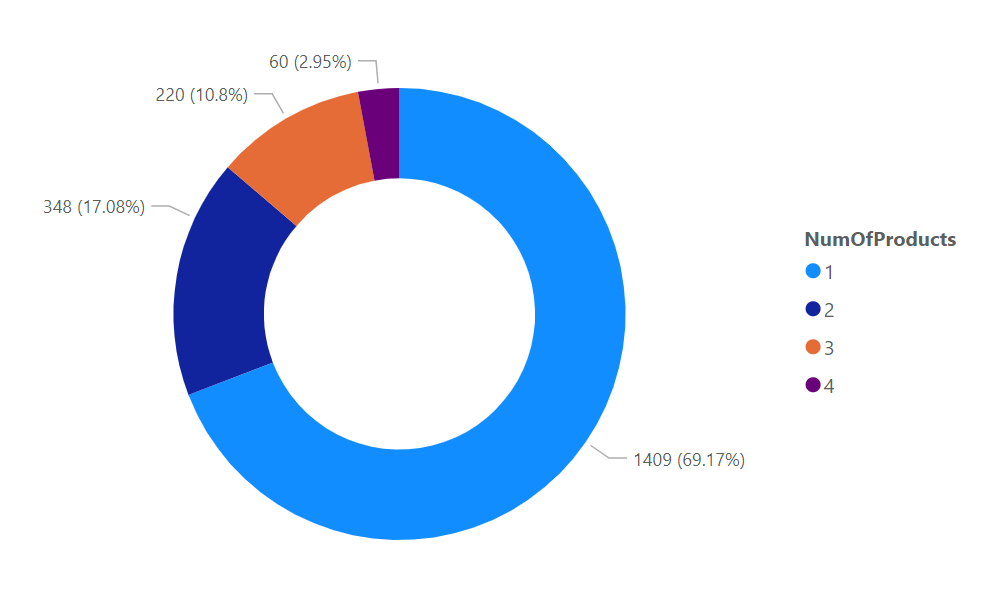
1. What is the impact of having a credit card on customer churn, based on the available data?



As we can see above image :

Churn rate is more for the customer having credit card as 1424 customers exited the bank which has credit card as compare to 613 customers which has not have the credit card

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

As we can see from the pie chart product number 1 is used by maximum number of the customers who exited.

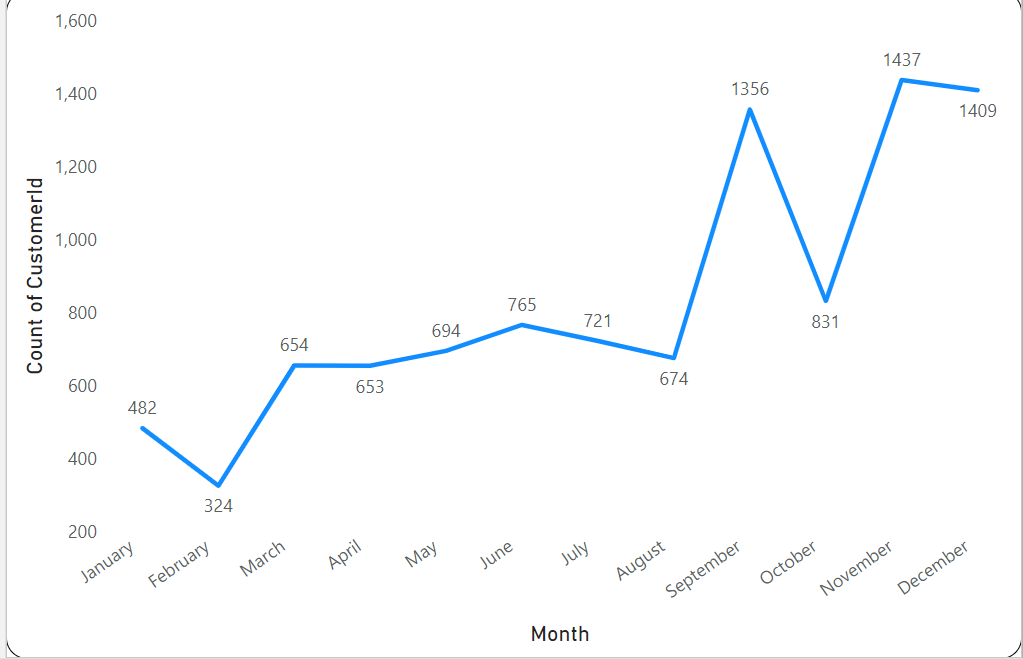
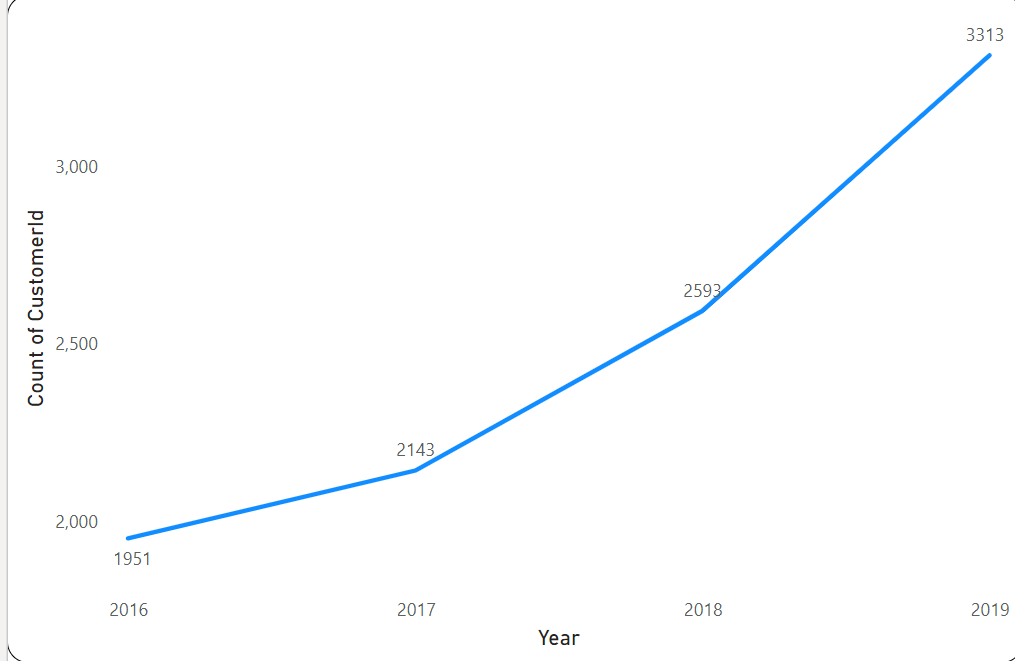
Product Number 1 used by 1409customers

Product Number 2 used by 348customers

Product Number 3 used by 220ustomers

Product Number 4 used by 60customers

11. 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.

As we can see from above line chart that count of customer joining the bank increase with increase in year

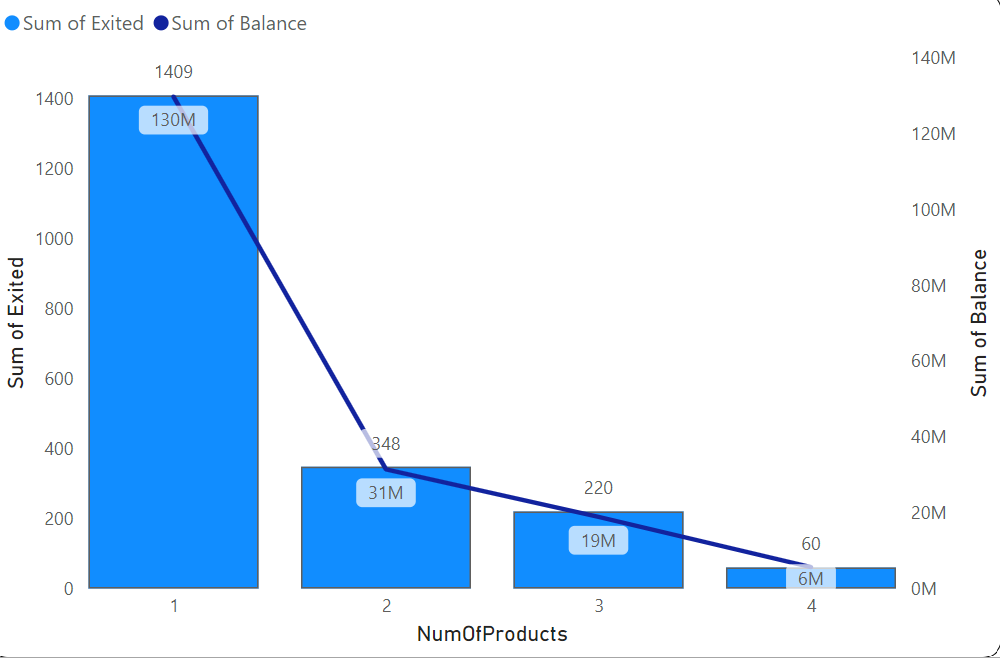
As 2016 has 1951 customers who joined the bank

As in 2017 , 2143 customers joined the bank

As in 2018, 2593 customers joined the bank

As in 2019 , 3313 customers joined the bank

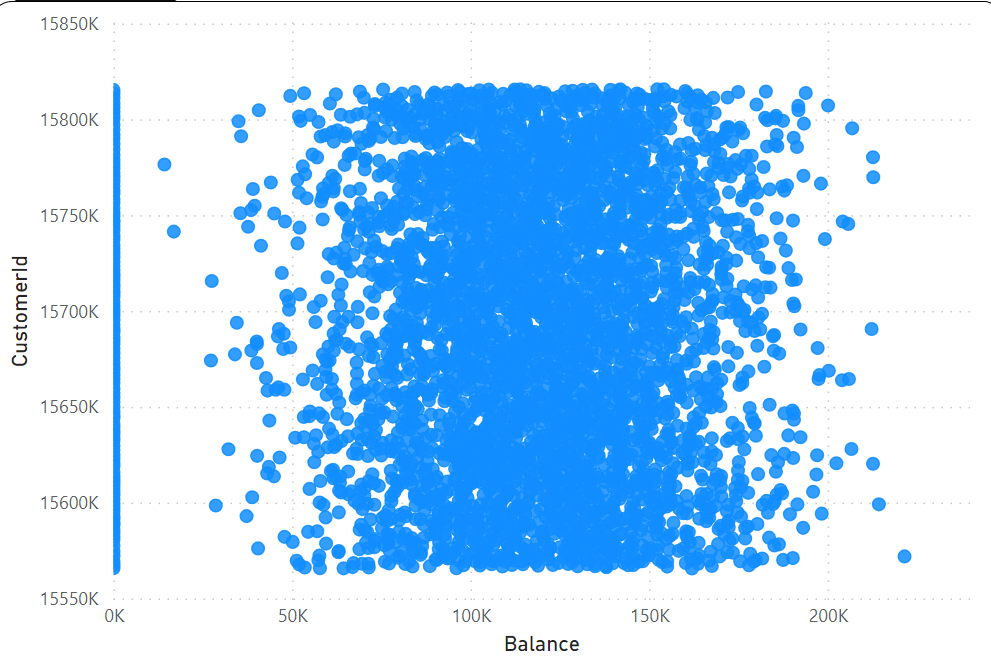
12. Analyze the relationship between the number of products and the account balance for customers who have exited.



As we can see from above image;

1. 1409 customers who exited the bank has used product nummber 1 and has total balance of 130million
2. 348 customers who exited the bank has used product nummber 1 and has total balance of 31million
3. 220 customers who exited the bank has used product nummber 1 and has total balance of 19million
4. 60 customers who exited the bank has used product nummber 1 and has total balance of 6million

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



As we can see from the above scatter plot graph:

The customers with the balance 0 is potential outliers as Outlier are the one which is abnormally have distance compare to the other value in our sample dataset

As we can see that more customer with balance 0 are more far from the customer having 50k, 100k, 150, 200k balance

3617 are the customers having 0 balance

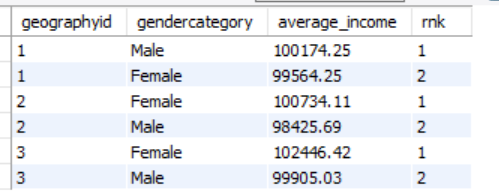
14. How many different tables are given in the dataset, out of these tables which table only consist of categorical variables?

Answer- 7 different table are given in dataset.

Out of this 7 tables, 5 tables( activecustomer, exitcustomer, creditcard, gender, geography) have categorical variable.

As activecustomer table has active category column which is categorical. Similarly in exitcustomer table, creditcard table, gender table, geography table has exit category, category, gender category, geography location columns respectively which are categorical.

15. 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)



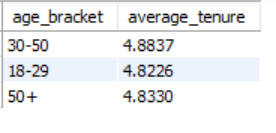
As we can see in above image:

for geography id 1, male has rank 1 with average income of 100174.25

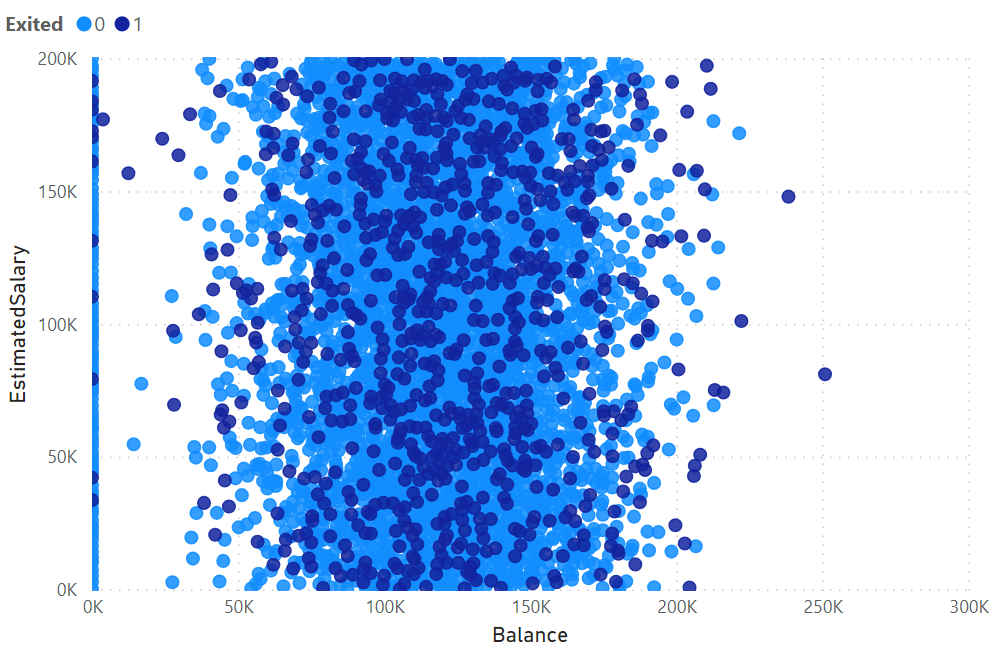
for geography id 2, female has rank 1 with average income of 100734.11

for geography id 3, female has rank 1 with average income of 99905.03

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+).

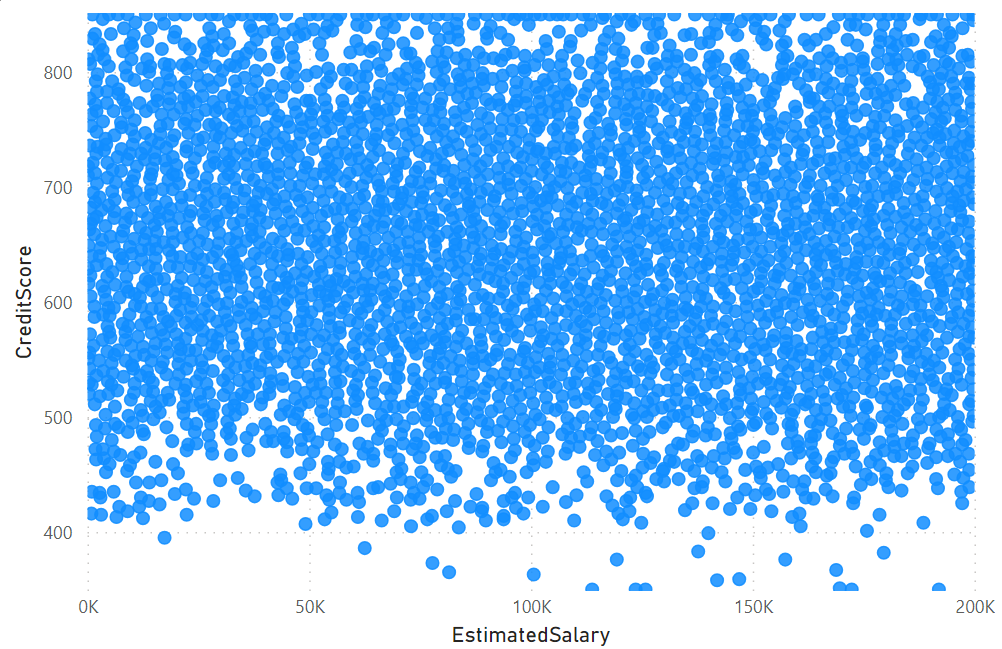
As we can see in the image customer from the age bracket 30-50 will stay more with the bank as they have high average tenure which helps in understanding that the customer with age between 30 to 50 are more likely to be with bank with maximum tenure.

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

As from the above scatter plot chart there no such correlation between salary and balance

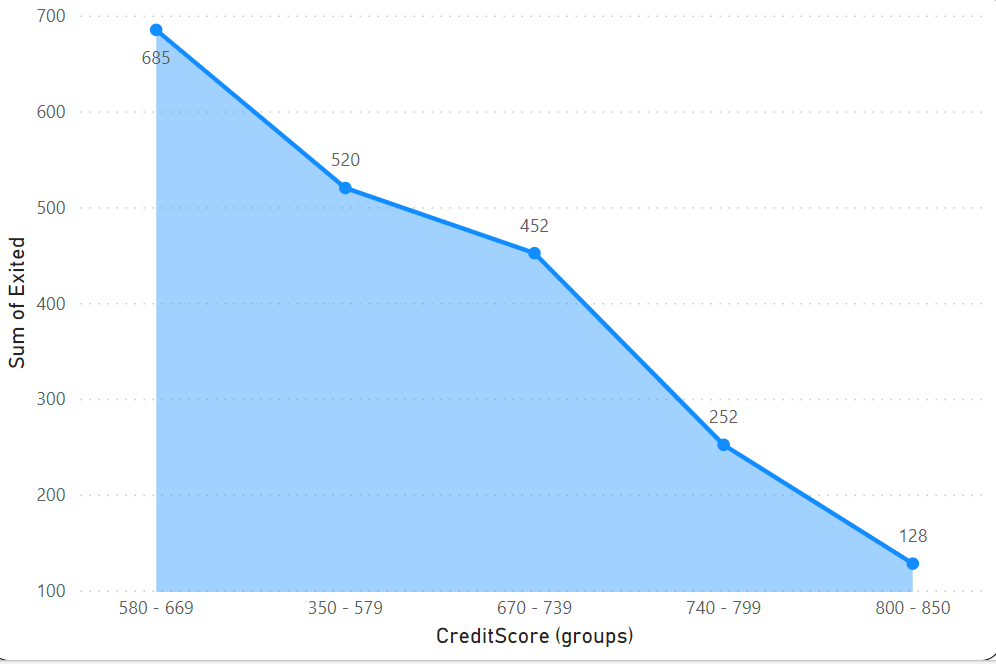
Yes but its different as balance range (upto 250k)is high for customer who remains in bank as compare to customers who exited the bank have range of balance upto 220k.

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



As from the above scatter plot graph as with increasing of the Estimated salary more customers are having good credit score or credit score of the customer also increasing but there are many customers with credit score between 600 – 700

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

As we can see in graph:

Credit score bucket 580-669 has rank 1 as it has 685 customer who exited the bank.

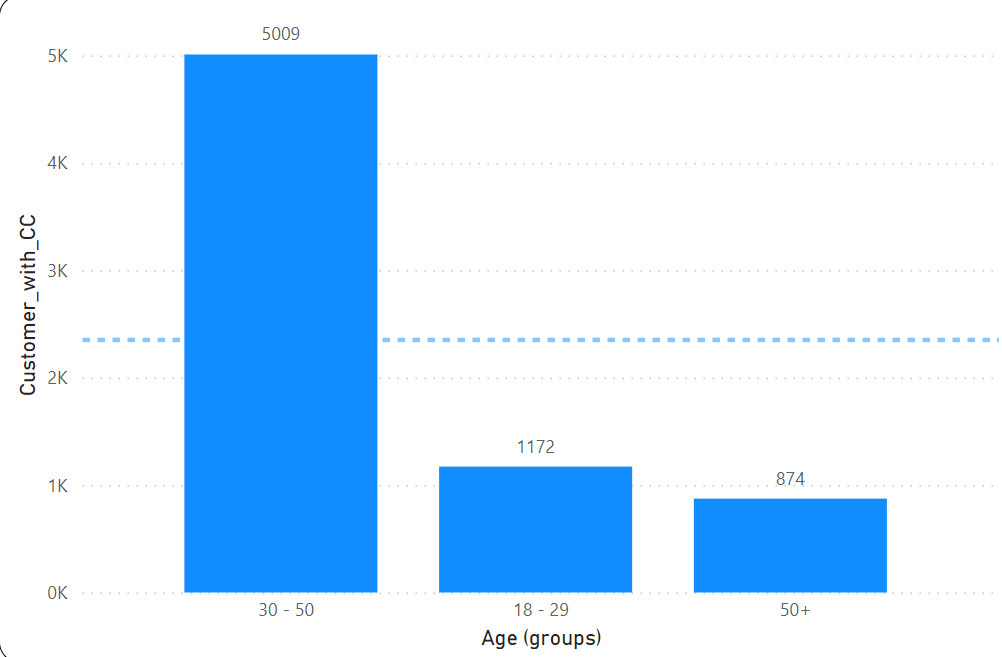
Credit score bucket 350-579 has rank 2 as it has 520 customer who exited the bank.

Credit score bucket 670-739 has rank 3 as it has 452 customer who exited the bank.

Credit score bucket 740-799 has rank 4 as it has 252 customer who exited the bank.

Credit score bucket 800-850 has rank 5 as it has 120 customer who exited the bank.

20. 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.



As from the above graph:

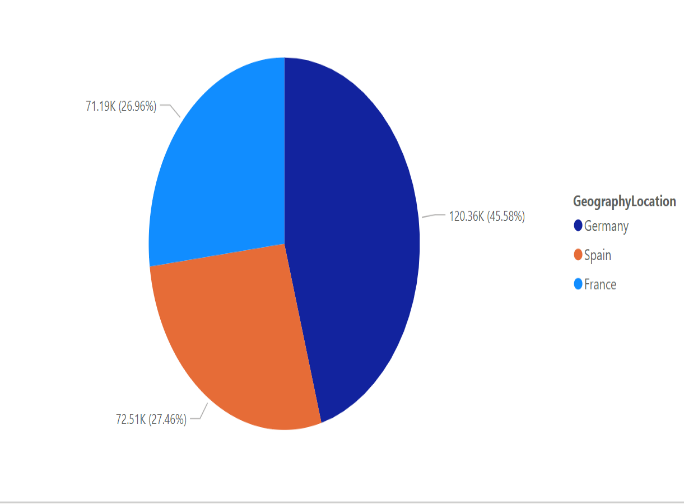
5009 customers from the age bracket 30-50 are having credit card

1172 customers from the age bracket 18-29 are having credit card

874 customers from the age bracket 50+ are having credit card

As 18-29 and 50+ these are two age bracket having less average credit card.

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

From the image;

As Germany is rank 1 as number of people who churn the bank is 814 and average balance is 120361.08

France is rank 2 as number of people who churn the bank is 810 and average balance is 71192.8

Spain is rank 3 as number of people who churn the bank is 413 and average balance is 72513.35

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

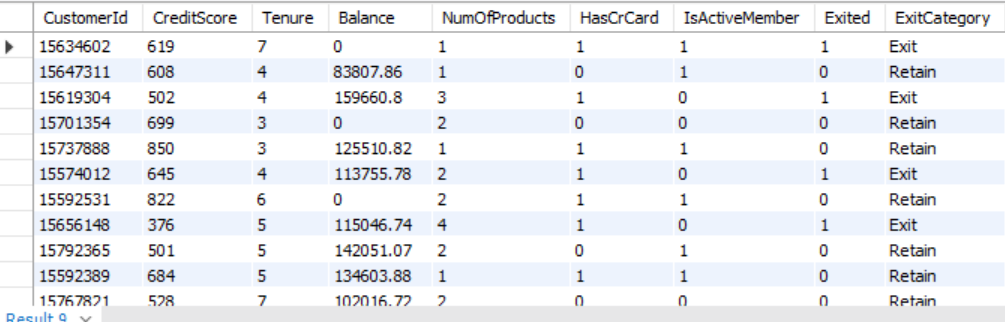
CONCAT(c1.CustomerID, '\_', c2.Surname) AS CustomerID\_Surname

FROM customerinfo c1

JOIN

customerinfo c2 ON c1.CustomerID = c2.CustomerID AND c1.Surname = c2.Surname;

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



select b.\*, e.ExitCategory from bank\_churn b, exitcustomer e

where b.exited = e.exitid;

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

Answer- No there were no missing value in the data set.

If in case the value will be missing then there are indeed several ways to handle missing values in SQL, and different tools may have their own specific functions or methods for dealing with them. Here are some common approaches in SQL

1. Use of IS NULL or IS NOT NULL: we can filter rows where a column is NULL or is not NULL using the IS NULL and IS NOT NULL operators, respectively.

SELECT \* FROM table\_name WHERE column\_name IS NULL;

SELECT \* FROM table\_name WHERE column\_name IS NOT NULL;

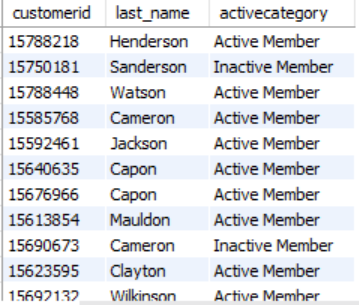
1. Use of COALESCE: The COALESCE function returns the first non-NULL value in a list of expressions.

SELECT COALESCE(column\_name, replacement\_value) FROM table\_name;

1. Use of CASE statements: You can use CASE statements to replace NULL values with specific values based on conditions.

SELECT CASE WHEN column\_name IS NULL THEN replacement\_value ELSE column\_name END AS new\_column\_name FROM table\_name;

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 b.customerid, c.surname as last\_name, a.activecategory

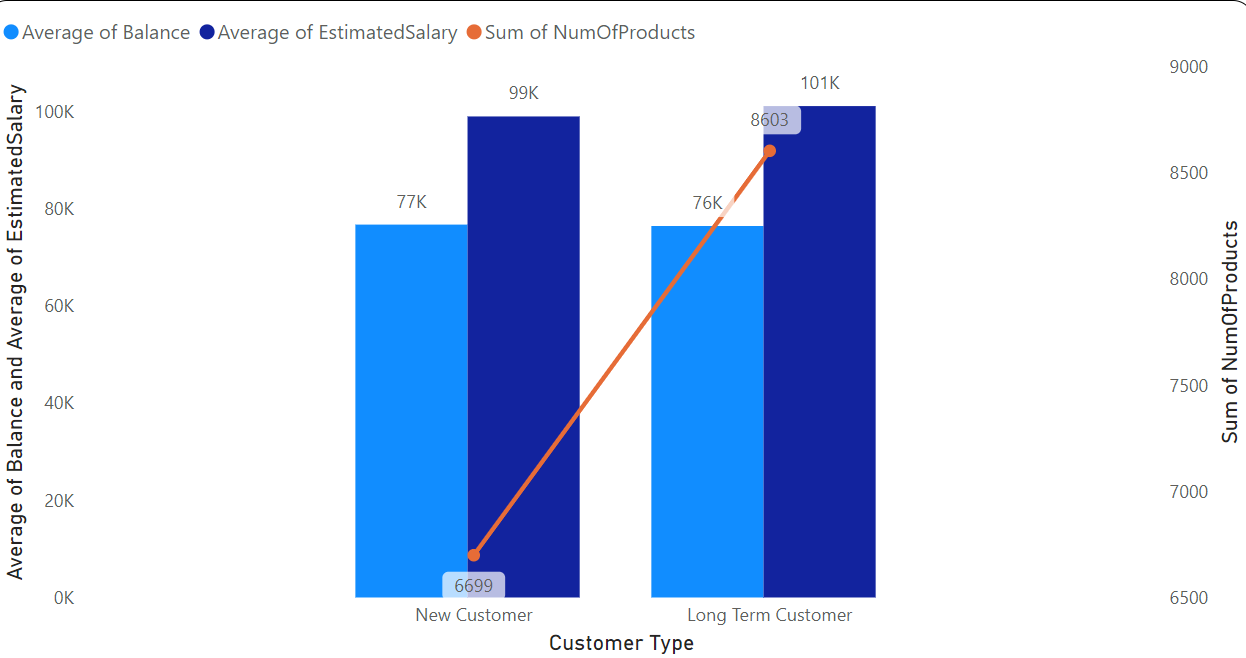
from customerinfo c join bank\_churn b on c.customerid = b.customerid

join activecustomer a on b.isactivemember = a.activeid

where c.surname like "%on";

**Data Analysis and Visualizations (Subjective Questions):**

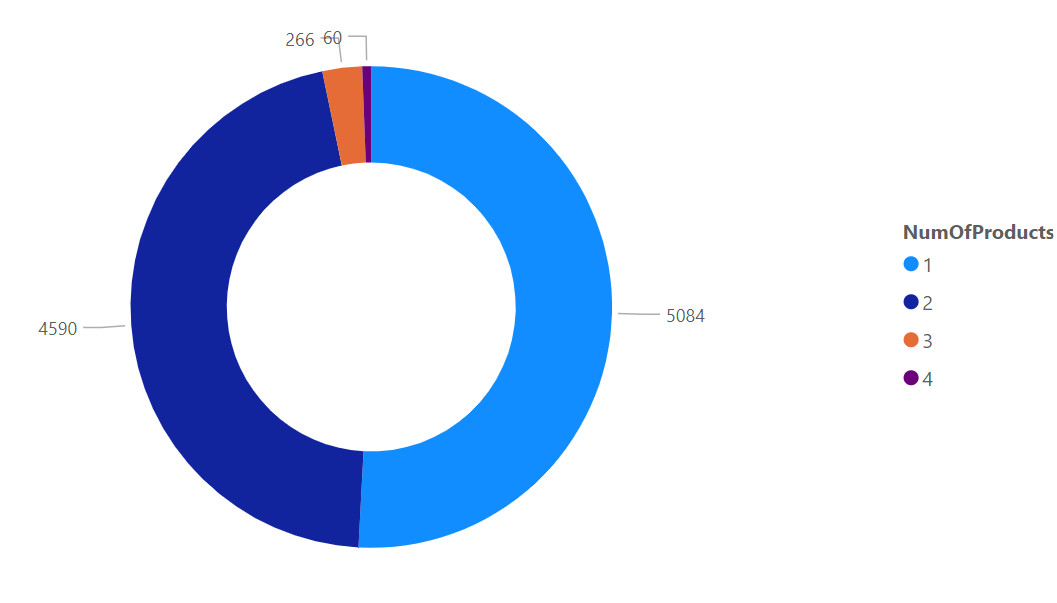
1. **Customer Behavior Analysis:** What patterns can be observed in the spending habits of long-term customers compared to new customers, and what might these patterns suggest about customer loyalty?



**Potentially Higher Spending by Long-Term Customers:** As we can see from the Column bar chart as the long term customers (tenure greater than 4 year) has high spending habit compared to new customers (tenure equals to 3 or 4 year) as the long term customers have higher average estimated salary than the new customers but even they maintain their average balance in account equals to average balance account of new customers which comes at point that long term customers have high spending habit compare to new customers

**Customer Loyalty:** As above graph suggests as the customer who stay more with the bank and has maximum tenure (greater than 4) and have more average balance in their account and use maximum of the services provided by bank are the loyal customers. These loyal customers have faith in bank and even show interest in new services if bank provide in future.

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



As we can see from the pie chart, **NumOfProducts 1** has most commonly used together among all NumOfProduct as **NumOfProducts 1** is used by 5084 customers

**NumOfProducts 2** is used by 4590 customers

**NumOfProducts 3** is used by 266 customers

**NumOfProducts 4** is used by 60 customers

These insights might influence cross-selling strategies:

1. **Identifying Complementary Products:** By understanding which bank products or services are frequently used together, banks can identify natural pairings or complementary offerings

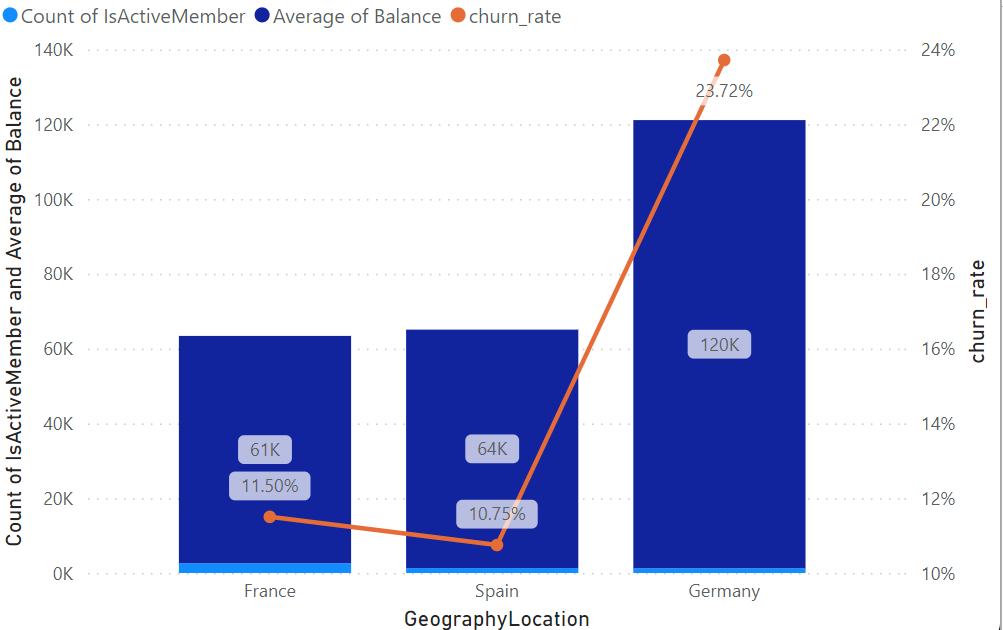
2. **Tailoring Product Bundles:** Bundling related products together can make the offering more attractive to customers and increase the likelihood of cross-selling multiple products in a single transaction.

3. **Personalized Recommendations:** Banks can leverage product affinity insights to deliver personalized product recommendations to individual customers based on their banking behavior.

4. **Cross-Selling Campaigns:** Banks can design targeted cross-selling campaigns that highlight complementary products and services to specific customer segments.

5.**Improving Customer Experience:** Understanding product affinities can also help banks enhance the overall customer experience by offering seamless integration and interoperability between related products and services.

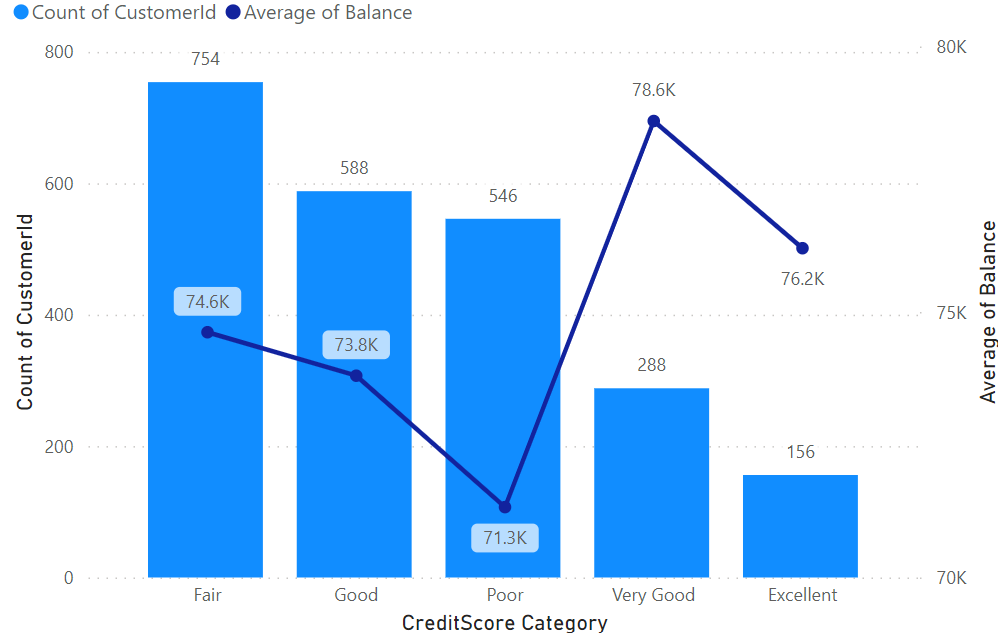
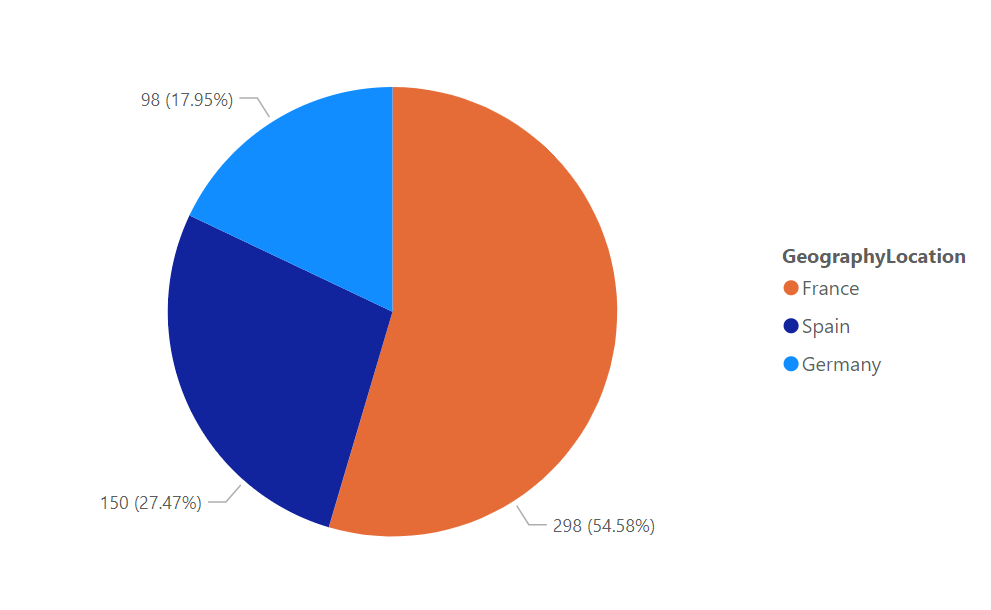
1. **Geographic Market Trends:** How do economic indicators in different geographic regions correlate with the number of active accounts and customer churn rates?



As from the stacked column chart:

* Germany has highest churn rate 23.72% and has lowest number of active accounts 1248 and has highest average balance of 120K.
* France has second highest churn rate 11.50% and has highest number of active account 2591 and has lowest average balance of 61K.
* Spain has lowest churn rate 10.75% and has second highest number of active account 1312 and has average balance of 64K

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

As from above images Based on customer profile (Credit score category):

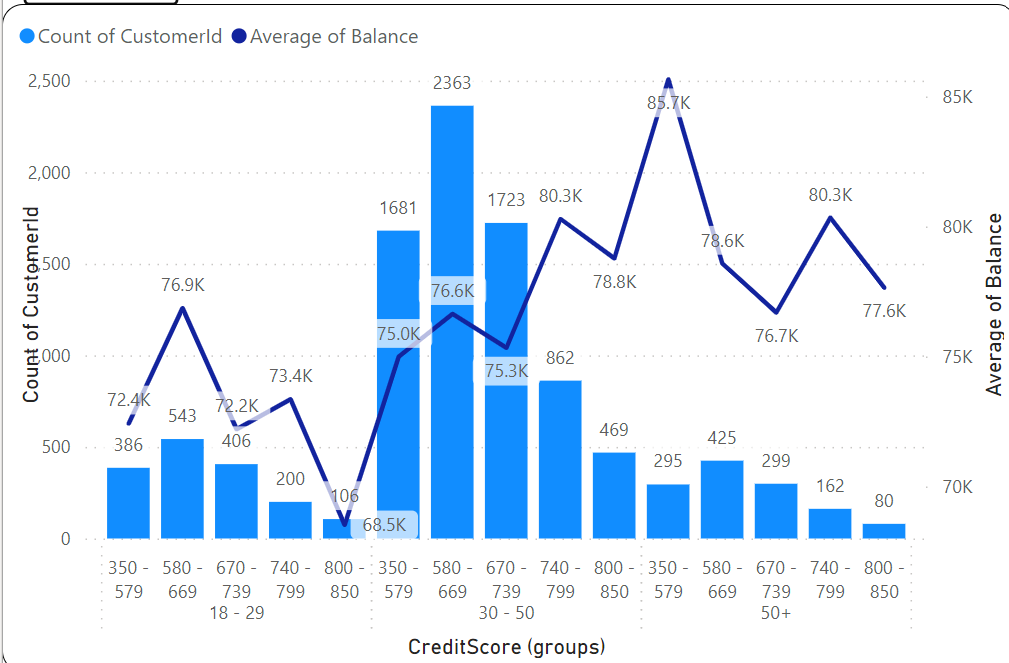
* Customer that belongs to Poor credit score category(350-579) has high financial risk to the bank as they have lowest average balance 71.3K, and even they are non- card holder

and they fail to pay the debt amount if they have taken debt.

* There are 546 customers belongs to poor credit score category who are with the bank and has non credit card holder
* **France** has highest number of customer 298 who belong to Poor Credit score category,

So they customer of **France** with non-credit card and having 0 balance or having low average balance are finically risk to bank.

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



From the above graph:

* Lifetime (tenure) of the customers based on the Age Bracket, Credit Score, Average balance they are maintaining.
* As Young in age (18-29) customer just joined the bank and maintain very low average balance and even they have low credit score as they have just joined the bank.
* **As we move further from age group 30- 50 these are customers who have high tenure and stay long with bank as they have good credit score and even they maintain high average balance in their account.**
* As we move from 50+ age bracket these customer again decrease there credit score so they less tenure and stay less with bank even if they maintain high average balance in account.

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 -**

* Segmentation of the customer base based on demographics, behavior, and product usage will allow us to analyze the impact of marketing campaigns on different customer groups. This segmentation will help identify which segments are most responsive to the campaigns and where adjustments may be needed.
* To conduct a comprehensive analysis of marketing campaign effectiveness, additional information may be needed, such as:
* Detailed data on marketing campaign spend, channels used, and messaging.
* Customer response data, including click-through rates, conversion rates, and campaign attribution.
* Customer feedback and sentiment analysis from social media, reviews, and customer support interactions.
* External factors that may influence customer behavior, such as economic conditions, competitive landscape, and industry trends.

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-

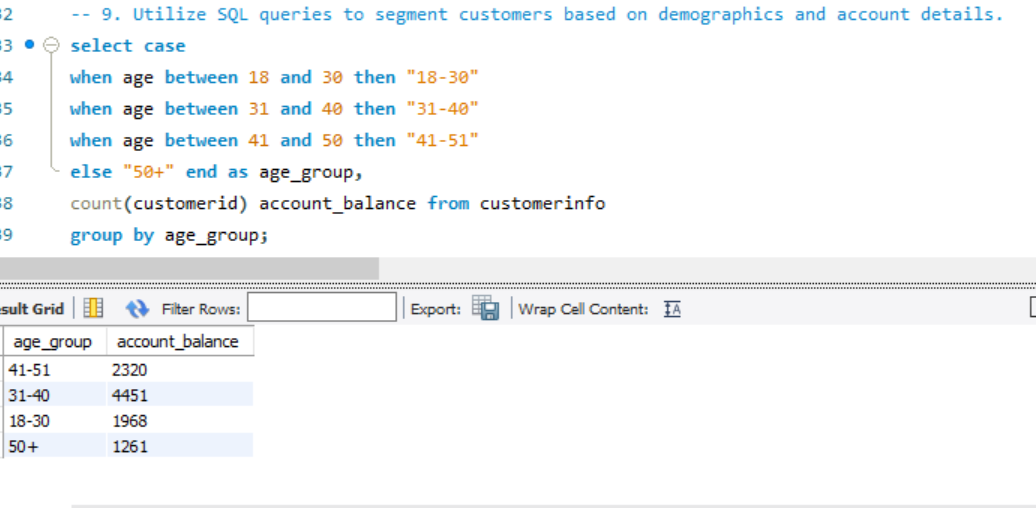
Some of characteristics among the customers that explain reason for leaving:

1. **HasCrCard**: Customers with no credit card are more likely to leave bank as no service is being provided by the bank.
2. **Tenure**: Customers who are having tenure of 3 or 4 year are more likely to leave bank as these customers are new one, they are unaware of the service provided by bank so they are uncomfortable in understanding the rules and regulations of the bank.
3. **NumOfProduct**: Customers who are using no service or no product or less service provided by bank are more likely to leave the bank as in available dataset bank has NumOfProduct 1, 2, 3, 4 so the customers who are using less number of service from four of the product are more likely to leave the bank.
4. **IsActiveMember:** Customers who are not so active are more likely to leave bank.
5. **EstimatedSalary**: Customers who have low average estimated salary are more likely to leave the bank .
6. **Balance**: Customers with 0 balance are more like to leave the bank
7. Are 'Tenure', 'NumOfProducts', 'IsActiveMember', and 'EstimatedSalary' important for predicting if a customer will leave the bank?

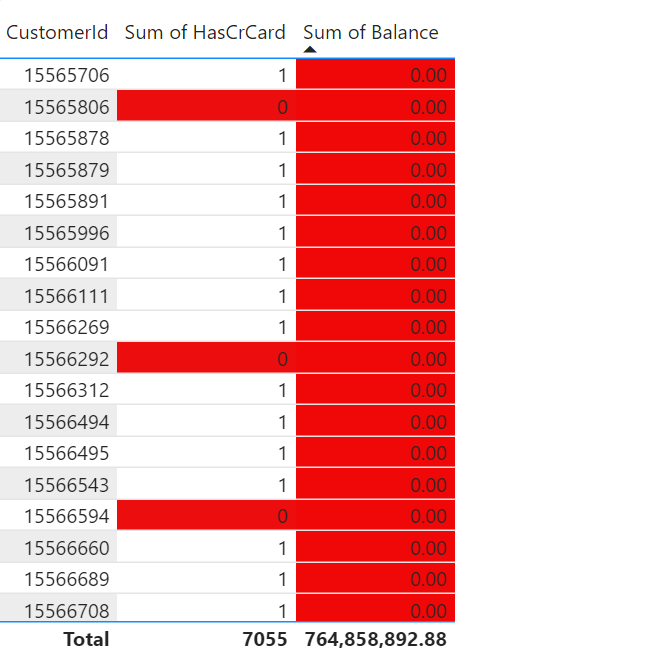
Answer-

Yes, 'Tenure', 'NumOfProducts', 'IsActiveMember', and 'EstimatedSalary' are important predictors for determining if a customer will leave the bank. These columns provide valuable insights into a customer's ability to purchase products within their estimated salary, settle payments within their tenure, and maintain active membership status in the bank.

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

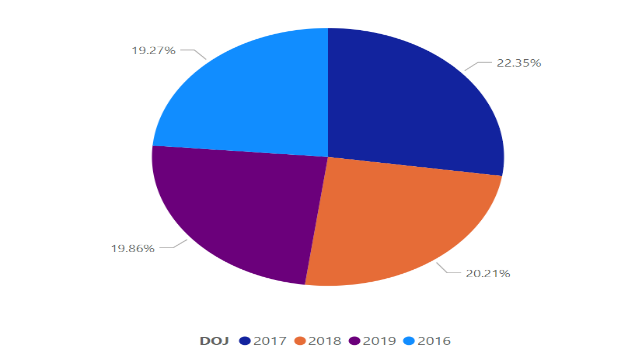


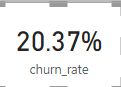
1. How can we create a conditional formatting setup to visually highlight customers at risk of churn and to evaluate the impact of credit card rewards on customer retention?



* By visually highlighting customers with 0 credit cards and 0 balance, the bank can quickly identify individuals who may be at risk of churn. This allows for targeted retention efforts to encourage account activity and mitigate churn. It indicating by red color.
* Comparing churn rates between customers with and without credit cards helps assess the effectiveness of credit card rewards in retaining customers. A lower churn rate among customers with credit cards indicates a positive impact on retention, informing strategies to enhance credit card rewards programs and foster customer loyalty.

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 are the different strategies that can be used to decrease the churn rate.



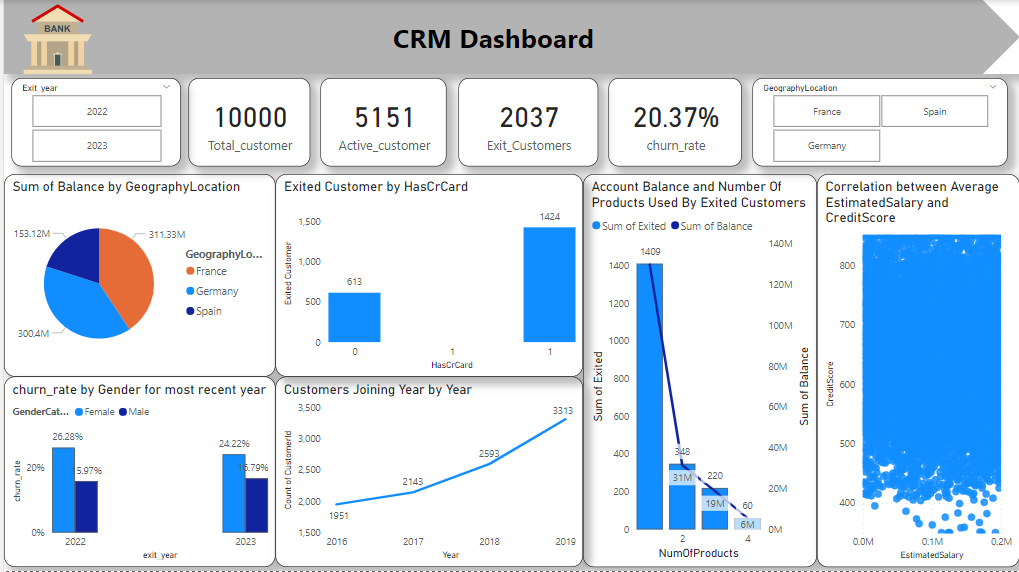
------Overall Churn Rate

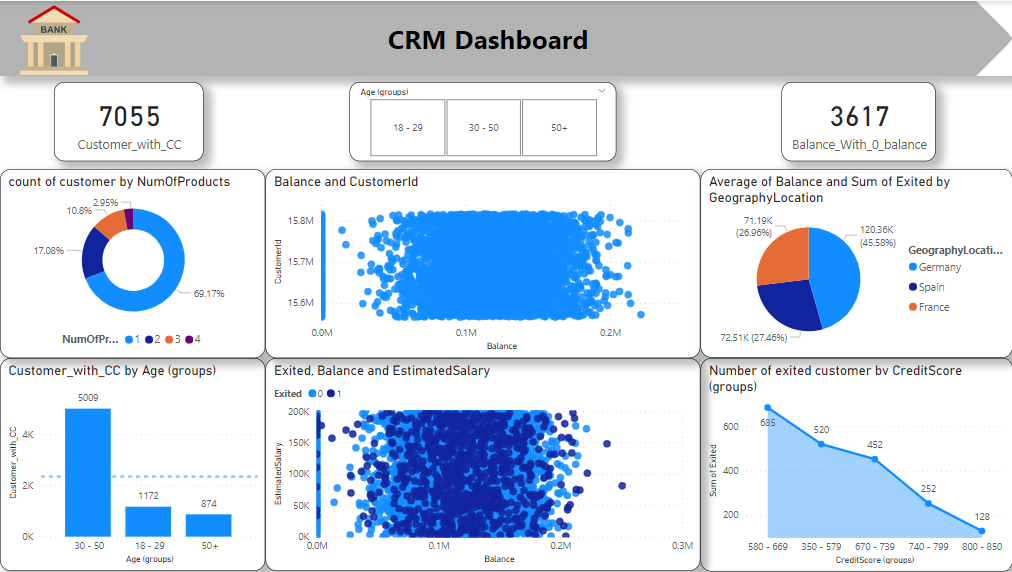
Based on the current data analysis, the churn rate per year and overall, in the bank can be seen from the above visual.

Insights to the bank about the types of customers more likely to churn:

* For instance, customers with lower tenures, fewer products, inactive memberships, or lower estimated salaries and not having credit card are likely to churn the bank.
* Strategies to decrease the churn rate could include targeted marketing campaigns to re-engage inactive customers, personalized offers to incentivize loyalty, improved customer service experiences, and proactive retention efforts such as loyalty programs or rewards programs. Additionally, analyzing customer feedback and conducting satisfaction surveys can provide valuable insights into areas for improvement and help tailor retention strategies accordingly.

1. 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?

Answer- If the objective and subjective questions were not provided, and I will approach the problem of analyzing customer churn and related factors in a banking churn, I would follow a structured approach to gain insights and create a meaningful analysis. Here's how I would approach it:

* Familiarize with the dataset and understand its structure, variables, and meanings.
* Exploring data distributions, patterns, and relationships between variables using visualizations and statistical summaries.
* Creating new features or transforming existing ones to enhance analysis.
* Analyze churn rates, trends, and factors influencing churn using statistical tests or machine learning models.
* Develop interactive dashboards to visualize key insights and trends effectively.
* Summarize findings and provide actionable recommendations to reduce churn and improve customer satisfaction.
* Iterate on the analysis based on feedback and new insights, updating the dashboard and recommendations accordingly.

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

We will use following synatax by using sql queries

Alter table bank\_churn

Rename column HasCrCard to Has\_creditcard;