**Objective Questions**

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

**Ans ->**

select ci.GeographyID, g.GeographyLocation, round(sum(bc.Balance),2) as Balance

from customerinfo ci

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

join geography g ON ci.GeographyID= g.GeographyID

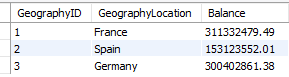
group by 1,2

ORDER BY ci.GeographyID;

**Approach:**

Currently there are three regions in dataset: France, Germany and Spain.

To find distribution of account balances we calculate the total balances and group by regions(Geography).



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

**Ans ->**

**Approach:**

To find top 5 customers with highest Est.Salary, we sort data by Salary in descending order, then limit the rows to 5.

select CustomerId, Surname, EstimatedSalary from customerinfo

where year(bankDOJ)= 2019 and quarter(bankDOJ)= 4

order by EstimatedSalary desc limit 5;



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

**Ans ->**

**Approach:**

The average number of products used by customers who have a credit card = 1.53

To get this, we calculate the average NumOfProducts used by customers, after filtering data by customers who are CreditCard holders.

select avg(NumOfProducts) as avg\_product\_cc from bank\_churn

where HasCrCard= 1;



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

**Ans ->**

**Approach:**

To find this, we calculate the average number of Churns(Exited) for each Gender using group by.

select g.GenderCategory,

cast(count(case when exited= 1 then b.CustomerId end)\*100/ count(b.CustomerId) as decimal(10,2))

as churn\_rate

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

join gender g ON g.GenderID= c.GenderID

where year(bankDOJ)= 2019

group by 1;



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

**Ans ->**

**Approach:**

For this, we simply calculate average Credit Score of customers wile grouping them by Exit.

select

Avg(Case when exited= 1 then creditscore end) as avg\_credit\_exited,

Avg(case when exited= 0 then CreditScore end) as avg\_credit\_remain

from bank\_churn;



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

**Ans ->**

**Approach:**

Female customers have the higher average estimated salary.

No, there’s no strong correlation with number of Active members and Average estimated salary.

select g.GenderCategory,round(avg(c.EstimatedSalary),2) as avg\_salary,

round(avg( case when a.ActiveID=1 then c.EstimatedSalary end),2) as avg\_salary\_active,

round(avg( case when a.ActiveID=0 then c.EstimatedSalary end),2) as avg\_salary\_inactive

from customerinfo c

inner join gender g

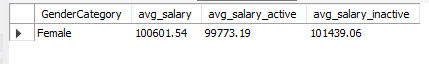
ON c.genderid= g.genderid

inner join bank\_churn b ON b.CustomerId= c.CustomerId

inner join activecustomer a ON b.IsActiveMember= a.ActiveID

group by g.GenderCategory

Order by avg\_salary desc limit 1 ;



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

**Ans ->**

**Approach:**

Segment with the highest exit rate Poor(300-579) of credit score.

with cte as(

select c.CustomerID, b.Exited, case

when b.CreditScore >= 800 then "Excellent"

when b.CreditScore between 740 and 799 then "Very Good"

when b.CreditScore between 670 and 739 then "Good"

when b.CreditScore between 580 and 669 then "Fair"

else "Poor"

end as Segment

from customerinfo c

join bank\_churn b

on c.CustomerID = b.CustomerID

)

select Segment, round(count(case when Exited = 1 then 1 end)/count(\*) \* 100, 2) as Exit\_Rate

from cte

group by Segment

order by count(case when Exited = 1 then 1 end)/count(\*) desc;

For this,

● we group by customers based on credit score

● count total customers of that segment

● count total exited customers of that segment

● and take ratio of it, we get exit rate

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



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

**Ans ->**

**Approach:**

France has the highest number of active customers (with tenure > 5 year

For this, we calculate the count of customers with tenure>5 and group by geography.

select g.GeographyLocation,count(case when a.ActiveCategory= 'Active Member' then 1 end )

as count\_active\_member from customerinfo c

inner join bank\_churn b ON c.CustomerId= b.CustomerId

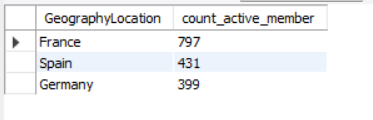
inner join geography g ON c.GeographyID= g.GeographyID

inner join activecustomer a ON b.IsActiveMember= a.ActiveID

where b.Tenure>5

group by g.GeographyLocation

order by count\_active\_member desc limit 3;

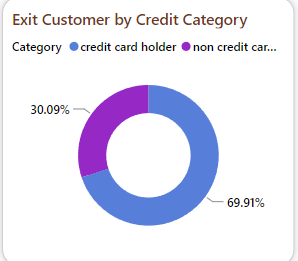


1. **What is the impact of having a credit card on customer churn, based on the available data?**

**Ans ->**

**Approach:**

From data exit customers are higher in customers having credit card (69.91%) rather than not having credit card (30.09%).

****

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

**Ans ->**

**Approach:**

For Customers who have Exited,

Most common NumOfProduct is 1

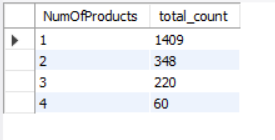
select NumOfProducts, count(NumOfProducts) as total\_count

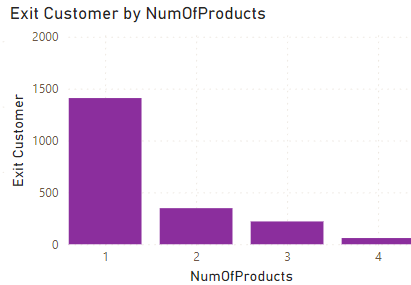
from bank\_churn

where exited= 1

group by NumOfProducts

order by total\_count desc limit 1;





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

**Ans ->**

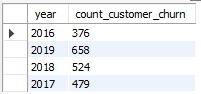
select year(bankDOJ) as year, count(c.CustomerId) as count\_customer\_churn

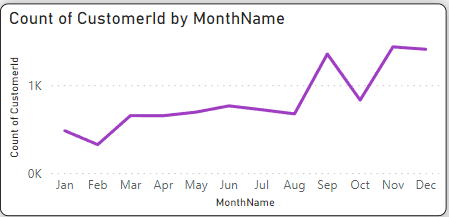
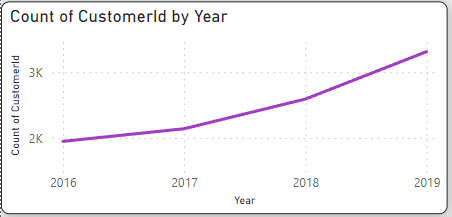
from bank\_churn b

inner join customerinfo c ON b.CustomerId= c.CustomerId

where Exited= 1

group by year(bankdoj);





**Approach:**

● Yearly Count of customers is smoothly increasing over years, without any dramatic fluctuations.

● Monthly Count of customers is overall increasing, but with some dramatic exits and the joining of customers during the last quarters of the year(Aug-Dec).

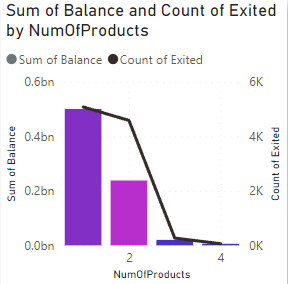
1. **Analyze the relationship between the number of products and the account balance for customers who have exited.**

**Ans ->**

**Approach:**

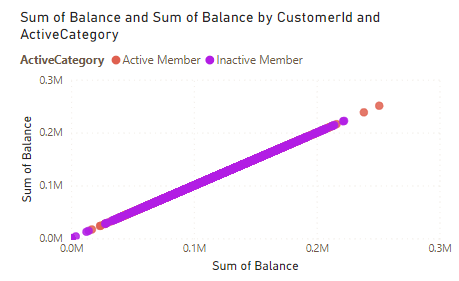
Relation that we can see between Num of Products and account balance of customers is: Sum of Balance of customers decreases with increase in Number of Products that indicates more customers use lesser number of products.

It was analysed on the Sum of balance rather than Average because average is almost the same for every segment because of similar count of customers.



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

**Ans ->** In the below scatter plot there are four outliers who have remained with bank.

****

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

**Ans ->**

**There are 7 different tables in dataset. And among them 5 tables consists of categorical variables.**

|  |
| --- |
| **Categorical Tables** |
| Gender |
| Geography |
| Active Customer |
| Exit Customer |
| CreditID |

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

**Ans ->**

with temp as

(

select c.GeographyID,g.GenderCategory ,

round(AVG(c.EstimatedSalary),2) as avg\_salary

from customerinfo c

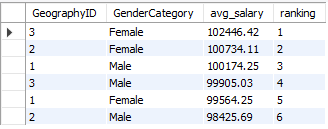
inner join gender g ON c.GenderID= g.GenderID

group by c.GeographyID,g.GenderCategory

)

select \*, rank() over(order by avg\_salary desc)

as ranking from temp ;



1. **Using SQL, write a query to find out the average tenure of the people who have exited in each age bracket (18-30, 30-50, 50+).**

**Ans ->**

**Approach:**

For this, we first segment ages in bracket then calculate average tenure and grouping by age bracket

with AgeBucket as

(

select c.CustomerId,c.surname,c.age,c.GenderID,

c.EstimatedSalary,c.GeographyID,c.bankDOJ,

b.CreditScore,b.tenure,b.balance,b.NumOfProducts,

b.HasCrCard,b.IsActiveMember,b.Exited,

case when c.age between 18 and 30 then '18-30'

when c.age between 31 and 50 then '30-50'

else '50+'

end as age\_bracket

from bank\_churn b

inner join customerinfo c ON b.CustomerId= c.CustomerId

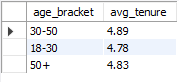
where exited=1

)

select age\_bracket, round(avg(tenure),2) avg\_tenure from AgeBucket

group by age\_bracket

;



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

**Ans ->**

**Approach:**

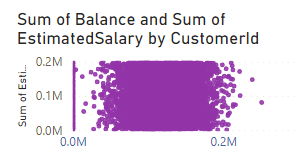
As we can see in the scatter charts below:

● No direct correlation of salary and balance is seen.

● The average balance (of all customers) lies in the range of 70K - 150K, despite salary difference.

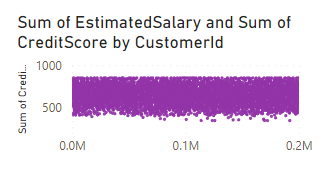
● As for the exited customers, the correlation is still negligible.

● The average balance (of exited customers) lies in the range of 90K - 150K, despite salary difference.

****

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

**Ans ->** **From following scatter plot it is clear that there is no correlation between salary and credit score**

****

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

**Ans ->**

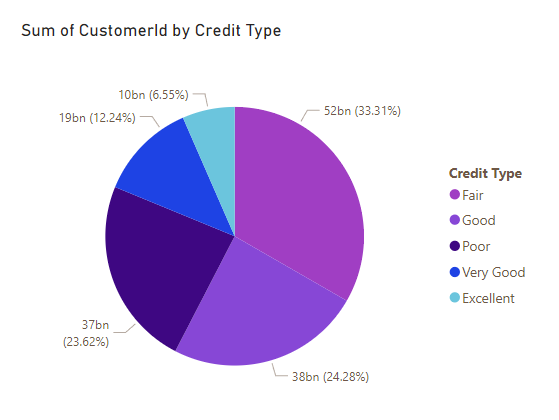
As per below graph 33.31% customers has fair credit score.

24.28% customers Good fair credit score.

23.62% customers has Poor credit score.

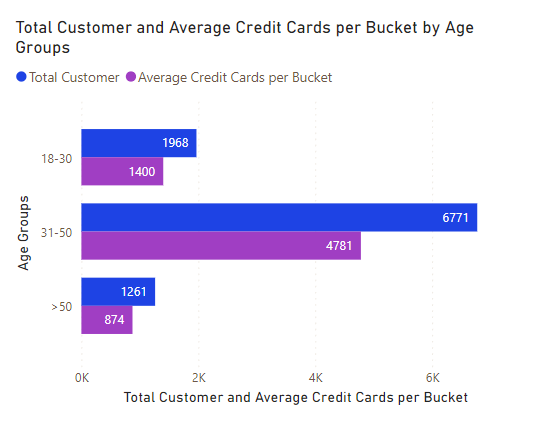
12.24% customers has Very Good credit score.

6.55% customers has Excellent credit score.

****

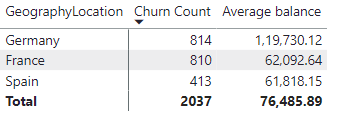
1. **According to the age buckets find the number of customers who have a credit card. Also retrieve those buckets that have lesser than average number of credit cards per bucket.**

**Ans ->** According to the age buckets the number of customers who have a credit card. And buckets those have lesser than number of credit cards per bucket. For age bucket **18-30** total customers are **1968** average number of credit cards per bucket is **1400**. For age bucket **31-50** total customers are **6771** average number of credit cards per bucket is **4781**. For age bucket 50 and above total customers are 1261 average number of credit cards per bucket is **874**.

****

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

**Ans ->** As per following table Germany is at 1st rank. 2nd one is France and last is Spain.

****

1. **As we can see that the “CustomerInfo” table has the CustomerID and Surname, now if we have to join it with a table where the primary key is also a combination of CustomerID and Surname, come up with a column where the format is “CustomerID\_Surname”.**

**Ans ->** select

ci.CustomerId,

ci.Surname,

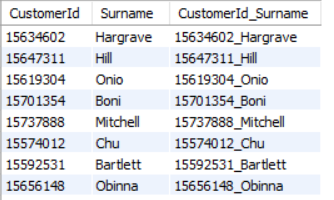
concat(ci.CustomerId,'\_',ci.Surname) as CustomerId\_Surname

from

customerinfo ci

join

bank\_churn ot on ci.CustomerId = ot.CustomerId;



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

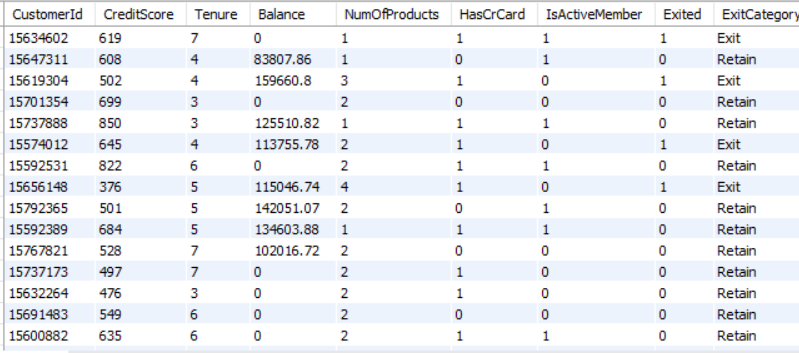
**Ans ->** SELECT

bc.\*,

(SELECT ExitCategory FROM exitcustomer ec WHERE ec.ExitID = bc.Exited) AS ExitCategory

FROM

bank\_churn bc;

****

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

**Ans ->**No there are no missing values but there are some null rows in the data. By using Power Query editor I handled them. And after that when I am getting the data to SQL then the data is getting corrupt so I made CSV file of that data (Separately) then I imported the data into SQL and then I made the connections.

1. **Write the query to get the customer IDs, their last name, and whether they are active or not for the customers whose surname ends with “on”.**

**Ans ->**

**Approach:**

● For this, we will select customers with a filter (‘where’ clause in SQL) of their surname that ends with “on” and their active status.

● Not all customers are shown.

● Limited sample output has been shown due to large number of output rows (i.e. large number of customers have surname ending with “on”)

● We can see all customers in SQL Workbench after running the query.

SELECT

bc.CustomerId,

ci.Surname,

CASE WHEN bc.IsActiveMember = 1 THEN 'Active' ELSE 'Inactive' END AS ActiveStatus

FROM

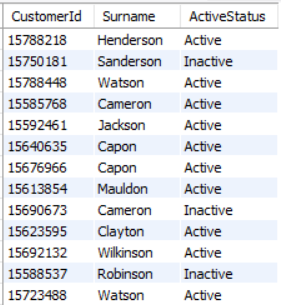
bank\_churn bc

JOIN

customerinfo ci ON bc.CustomerId = ci.CustomerId

WHERE

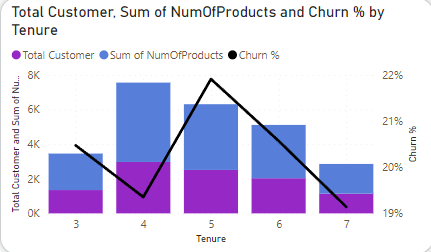
ci.Surname LIKE '%on';

****

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

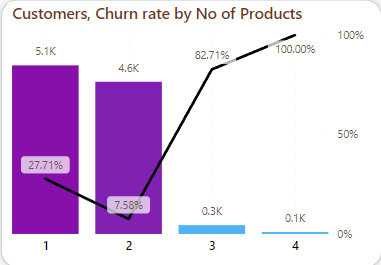
1. **Customer Behavior Analysis: What patterns can be observed in the spending habits of long-term customers compared to new customers, and what might these patterns suggest about customer loyalty?**

**Ans ->** From data it is Observed that Customers with Tenure 4 has spend more and customers with tenure 5 has higher churn rate and loyalty of customers is proportional to tenure, so with higher tenure customers are more loyal.

****

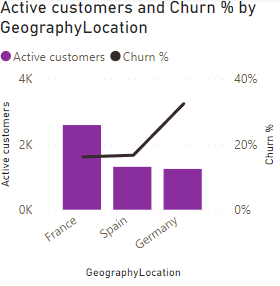
1. **Product Affinity Study: Which bank products or services are most commonly used together, and how might this influence cross-selling strategies?**

**Ans ->** From data it is been observed that Customers with less no of products has less churn rate compare to the customers who bought more no of products.

****

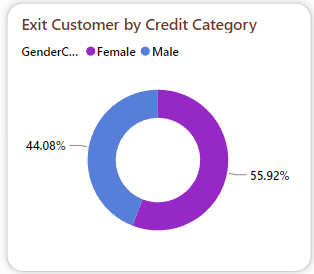
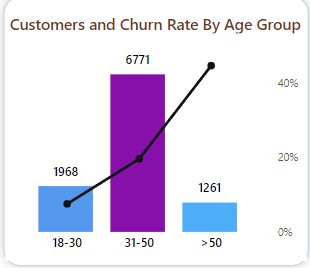
1. **Geographic Market Trends: How do economic indicators in different geographic regions correlate with the number of active accounts and customer churn rates?**

**Ans ->** From graph it is clear that France has more active customers and churn rate is max in Germany.

****

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

**Ans ->** From data it is clear that customer churn rate is proportional to the customer age. Female has higher churn rate compared to male.

****

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

**Ans ->** - Provide tailored offers and incentives based on customer preferences and behavior to encourage retention

- Implement loyalty programs that reward customers for their continued business and encourage them to stay with

the bank

1. **Marketing Campaign Effectiveness: How could you assess the impact of marketing campaigns on customer retention and acquisition within the dataset? What extra information would you need to solve this?**

**Ans ->**

Enhance the customer service experience service experiences by providing and personalized assistance, resolving issues efficiently, and addressing customer feedback

Marketing team should give a good offers and more security to customers age above 50, Also should offer more to customers who buys more no of products. They should also give more offers to credit card holders.

● Customer Segmentation: Segment customers based on age, geography, and number of products used. This allows for targeted analysis of campaign performance within specific groups.

● Trend Analysis: Analyse changes in active counts, exit rates, and product usage over time. This can reveal trends potentially influenced by marketing campaigns.

● Control Groups: If possible, identify control groups not exposed to specific campaigns. Comparing their behaviour with exposed groups helps isolate campaign impact.

Extra Information for Absolute Assessment:

● Campaign Details & Timing: Details on campaign content, channels used (online, offline), and launch dates are crucial. This allows for direct correlation between campaign exposure and changes in customer behaviour.

● Customer Acquisition Channel: Understanding how customers were initially acquired helps assess if campaigns are effective at retaining customers acquired through different channels (referral, online ad, etc).

● Customer Lifetime Value: Calculate customer lifetime value to assess the long-term impact of campaigns on customer retention and overall revenue generation.

By enriching your data with these additional elements, you can move beyond basic correlations and establish causation.

This empowers you to:

● Optimise Campaigns: Identify which campaigns resonate best with specific customer segments and adjust strategies for maximum impact.

● Measure ROI: Calculate the return on investment for marketing campaigns by comparing acquisition costs to customer lifetime value.

Conclusion:

The existing data provides a decent foundation for assessing marketing campaign effectiveness.

However, incorporating campaign details, customer acquisition channels, and customer lifetime value unlocks a deeper understanding of campaign impact.

By segmenting your data, conducting trend analysis, and utilising

control groups when possible, you can gain valuable insights.

Ultimately, a data-driven approach that combines existing data with additional information will empower optimise campaigns, maximise ROI, and drive long-term business growth.

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

**Ans ->** Customers with age above 50 are not satisfied with bank terminology or schemes. Customers having credit score below 700 are also leaving. Germany and France are having high churn rates.

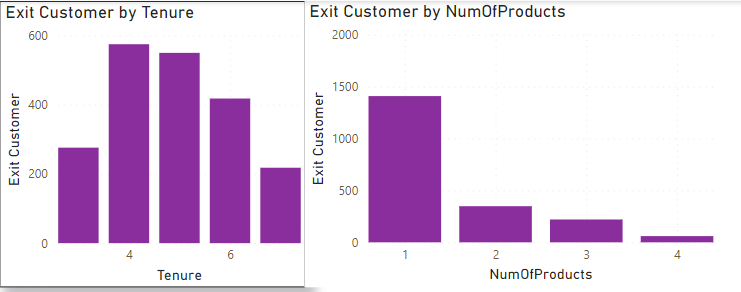
Low levels of customer satisfaction can lead to churn.

Dissatisfaction with customer Service, long wait times, and Unresolved issues are common reasons why customers churn banks

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

**Ans ->** From given dataset Customers with tenure 4 and 5 are more likely to exit. Customers with less no of products are

more likely to exit and we can not predict it from estimated salary.

****

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

**Ans ->** with AgeBucket as

(

select c.CustomerId,c.surname,c.age,c.GenderID,

c.EstimatedSalary,c.GeographyID,c.bankDOJ,

b.CreditScore,b.tenure,b.balance,b.NumOfProducts,

b.HasCrCard,b.IsActiveMember,b.Exited,

case when c.age between 18 and 30 then '18-30'

when c.age between 31 and 50 then '30-50'

else '50+'

end as age\_bracket

from bank\_churn b

inner join customerinfo c ON b.CustomerId= c.CustomerId

)

select age\_bracket, count(customerId) total\_customers from AgeBucket

group by age\_bracket

;

;

By Credit Score- with creditbucket as

(

select \*,

case when creditscore between 0 and 579 then 'Poor'

when creditscore between 580 and 669 then 'Fair'

when creditscore between 670 and 739 then 'Good'

when creditscore between 740 and 800 then 'Very Good'

else 'Excellent'

end as creditBucket

from bank\_churn

where exited = 1

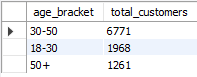
)

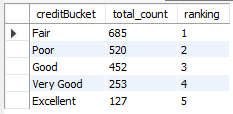
select creditbucket, count(customerid) as total\_count,

dense\_rank() over(order by count(customerid) desc) as ranking

from creditbucket

group by creditbucket





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?**

**Ans ->** Created Churn Risk Score: Created a calculated column on basis of currently 3 reasonable columns, i.e.

Churn Risk Score (CRS) = Balance / (Credit Score \* NumOfProducts)

that implies Churn Risk is directly proportional to Balance and inversely proportional to Credit Score & Number of Products

Note: Customer with Balance 0 will have CRS equal to 0

Created the table visual: Build a table incorporating factors like:

● Number of products

● Credit scores

● Balance

● Tenure

● other account activity

Conditional Formatting:

● Select the churn risk score column in your table

● Right-click and choose Conditional formatting

● Highlight risk levels using background colours:

○ Low risk (score below a defined threshold)

○ Moderate risk (score within a specific range)

○ High risk (score exceeding a defined threshold).

○ Too Low value/ Null to be judged

Benefits:

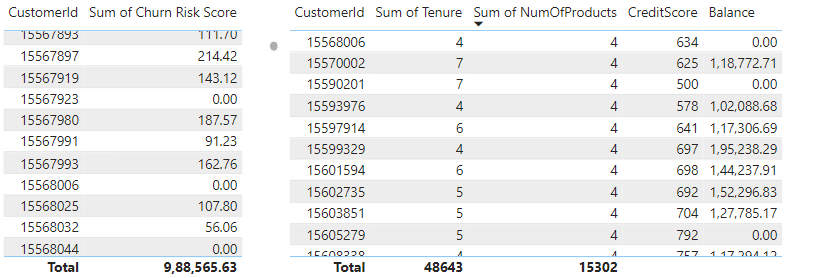
This colour-coded table instantly flags at-risk customers, allowing you to prioritise retention efforts. Think of it as a heatmap within your table, guiding your focus towards potential churners.

Remember:

● Judging must be for Active Customers.

● Adjustment of thresholds and colours based on data and risk assessment should be done periodically.

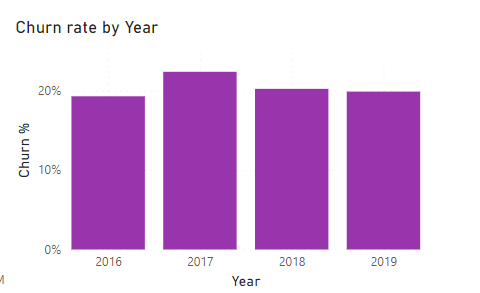
● We can consider exploring additional formatting options for even more visual cues as our data would grow



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.**

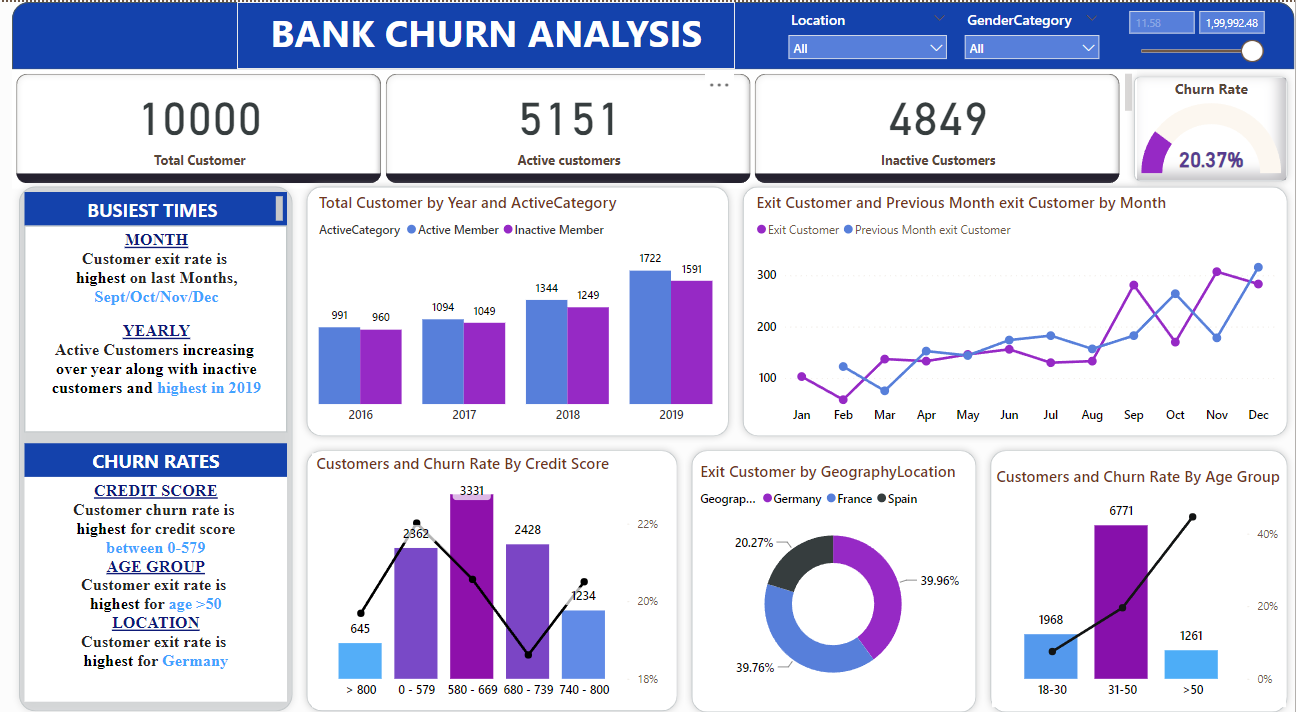
**Ans ->** Overall churn rate is 20.47 % . In year 2016- 19.27%, 2017- 22.35%, 2018- 20.21% and in 2019- 19.86%.

Customers above age 50 and customers having less no of product are more likely to churn. So if bank provides some good rewards, Offers and securities to customers from this category they might stay with bank for long period.

****

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

**Ans ->**

****

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

**Ans ->**

Approaching a data analysis problem without specific objective and subjective questions requires a flexible and exploratory approach. Here's a solution tailored to this scenario:

**1. Understand the Context:** Begin by gaining a thorough understanding of the context and purpose of the data analysis. This might involve consulting with stakeholders or reviewing project documentation to identify any implicit objectives or expectations.

**2. Explore the Data:** Conduct exploratory data analysis to familiarise yourself with the dataset. Examine the structure, distributions, and relationships between variables. Look for any patterns, anomalies, or interesting trends that emerge.

**3. Identify Potential Insights:** Based on your exploration of the data, brainstorm potential insights or hypotheses that could be explored further. Consider both quantitative patterns and qualitative aspects that may be relevant to the problem domain.

**4.** **Generate Hypotheses:** Formulate hypotheses or conjectures based on your initial observations. These hypotheses can serve as guiding principles for further analysis and experimentation.

**5. Iterative Analysis:** Iteratively apply various analytical techniques to test and refine your hypotheses. This could include statistical analysis, machine learning algorithms, or qualitative methods depending on the nature of the data and the problem, but in this project we would run multiple different queries and play with data in Excel, PowerBI & SQL.

**6. Visualisation and Interpretation:** Use data visualisation techniques to communicate your findings effectively. Visualisations can help uncover patterns, highlight relationships, and convey insights to stakeholders in a meaningful way.

**7. Synthesise Findings:** Synthesise your findings into coherent narratives or themes that provide a holistic understanding of the data. Look for overarching patterns or trends that emerge across different aspects of the analysis.

**8. Seek Validation and Feedback** Validate your findings through peer review, domain experts, or comparison with external sources if possible. Seek feedback from stakeholders to ensure that your analysis resonates with their expectations and contributes to their understanding of the problem.\

**9. Iterate and Refine:** Iterate on your analysis as needed, refining your approach based on new insights or feedback from stakeholders. Be open to exploring alternative hypotheses or perspectives to deepen your understanding of the problem.

**10. Document and Communicate:** Document your analysis process, methodologies, and findings in a clear and transparent manner. Communicate your results effectively to stakeholders, highlighting key insights, implications, and recommendations for further action.

By following this approach, we can systematically explore and analyse the data even in the absence of specific objective and subjective questions, ultimately deriving valuable insights that can inform decision-making and drive positive outcomes for the project.

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

**Ans ->**

ALTER TABLE Bank\_Churn

RENAME COLUMN HasCrCard TO Has\_creditcard;

**This will modify the name of the “HasCrCard” column to “Has\_creditcard” column.**