OBJECTIVE QUESTIONS

-Firstly, creating database in **MySQL Workbench** and
Given excel data converted to CSV format and imported into CAPSTONE database.

SQL Query:

```
CREATE DATABASE CAPSTONE;
```

In SQL, the **USE** statement is used to select a particular database before implementing the answered queries.

SQL Query:

```
USE CAPSTONE;
```

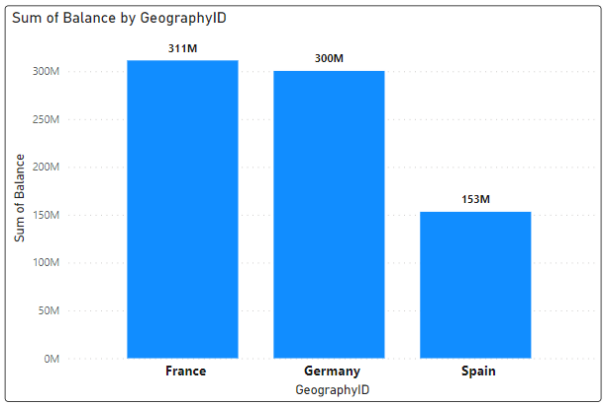
1.What is the distribution of account balance across different regions?

SQL Query:

```
SELECT
C.GeographyID,ROUND(SUM(Bc.balance),2)AS Account_balance
FROM Bank_Churn Bc
JOIN Customerinfo C ON
Bc.CustomerId = C.CustomerId
GROUP BY C.GeographyID;
```

GeographyID	Account_balance
France	311332479.49
Spain	153123552.01
Germany	300402861.38

VISUALISATION:



-The graph shows the distribution of balance for the regions of
FRANCE-**311M**
GERMANY-**300M**
SPAIN-**153M**

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

SQL Query:

```
SELECT EstimatedSalary,CustomerId
FROM Customerinfo
WHERE QUARTER(Bank_DOJ) = 4
ORDER BY EstimatedSalary DESC
LIMIT 5;
```

EstimatedSalary	CustomerId
199970.74	15634359
199841.32	15804211
199805.63	15687913
199753.97	15763065
199638.56	15599792

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

SQL Query:

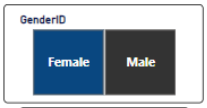
```
SELECT AVG(NumOfProducts) AS AverageProducts
FROM Bank_Churn bc
WHERE HASCrCard = '1';
```

AverageProducts
1.5314

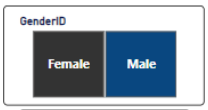
4. Determine the churn rate by gender for the most recent year in the dataset.
SQL Query:

```
With CTE AS
(SELECT c.GenderID,SUM(Exited) AS churncount
FROM Bank_Churn bc
JOIN Customerinfo c ON c.CustomerId=bc.CustomerId
GROUP BY GenderID)

SELECT GenderID,SUM(Exited) * 100.0 / COUNT(*) AS churnrate
FROM Bank_Churn bc
JOIN Customerinfo c ON c.CustomerId=bc.CustomerId
GROUP BY GenderID;
```



GenderID	churnrate
Female	25.07154
Male	16.45593



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

SQL Query:

```
SELECT Exited,AVG(CreditScore) AS Avg_credit_score
FROM Bank_Churn
GROUP BY Exited;
```

Exited	Avg_credit_score
1	645.3515
0	651.8532

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

SQL Query:

```
SELECT GenderID,
Round(AVG(EstimatedSalary),2) AS Avg_estimated_salary,
COUNT(IsActiveMember) AS Active_members
FROM Customerinfo c
JOIN Bank_Churn bc ON bc.CustomerId=c.CustomerId
WHERE IsActiveMember ='1'
GROUP BY GenderID
Order by Avg_estimated_salary DESC;
```

GenderID	Avg_estimated_salary	Active_members
Female	99773.19	2284
Male	99197.86	2867

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

SQL Query:

```
SELECT
CASE
WHEN CreditScore BETWEEN 350 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 799 THEN 'Very good'
ELSE 'Excellent'
END AS Credit_bucket,
ROUND((SUM(Exited)/ COUNT(*)) * 100,2) AS ExitRate
FROM Bank_churn
GROUP BY Credit_bucket
ORDER BY ExitRate DESC;
```

Credit_Bucket	ExitRate
Poor	22.02
Very good	20.59
Fair	20.56
Excellent	19.54
Good	18.62

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

SQL Query:

```
SELECT c.GeographyID,COUNT(IsActiveMember) AS Highest_Activecustomer
FROM Customerinfo c
JOIN Bank_Churn bc ON c.CustomerId=bc.CustomerId
WHERE IsActiveMember ='1' AND Tenure >5
GROUP BY c.GeographyID
ORDER BY Highest_Activecustomer DESC;
```

GeographyID	Highest_Activecustomer
France	797
Spain	431
Germany	399

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

SQL Query:

```
SELECT HASCrCard,
   SUM(Exited) * 100 / COUNT(*) AS ChurnRate
FROM bank_churn
GROUP BY HasCrCard
ORDER BY ChurnRate DESC;
```

HasCrCard	ChurnRate
0	20.8149
1	20.1843

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

SQL Query:

```
SELECT NumOfProducts,COUNT(Exited)AS Exitedcust_count
FROM Bank_Churn
WHERE Exited = '1'
GROUP BY NumOfProducts
ORDER BY NumOfProducts;
```

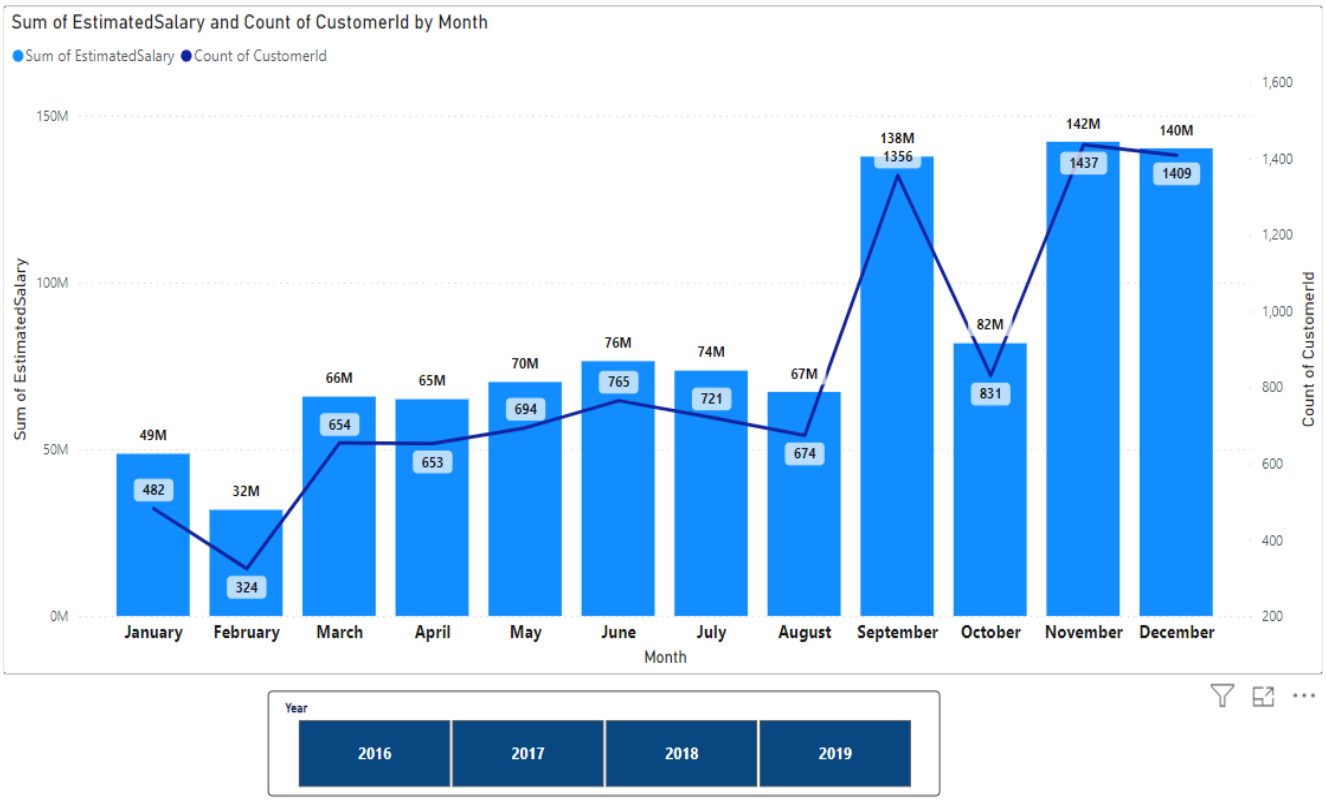
NumOfProducts	Exitedcust_count
1	1409
2	348
3	220
4	60

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.

SQL Query:

```
SELECT YEAR(Bank_DOJ) AS 'Yearly',
COUNT(customerID) AS CustomerCount,
ROUND(SUM(EstimatedSalary), 0) AS Salary
FROM customerinfo
GROUP BY YEAR(Bank_DOJ)
ORDER BY CustomerCount DESC, Salary DESC;
```

	Yearly	CustomerCount	Salary
▶	2019	3313	328272363
	2018	2593	261815573
	2017	2143	218398849
	2016	1951	192415614



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

SQL Query:

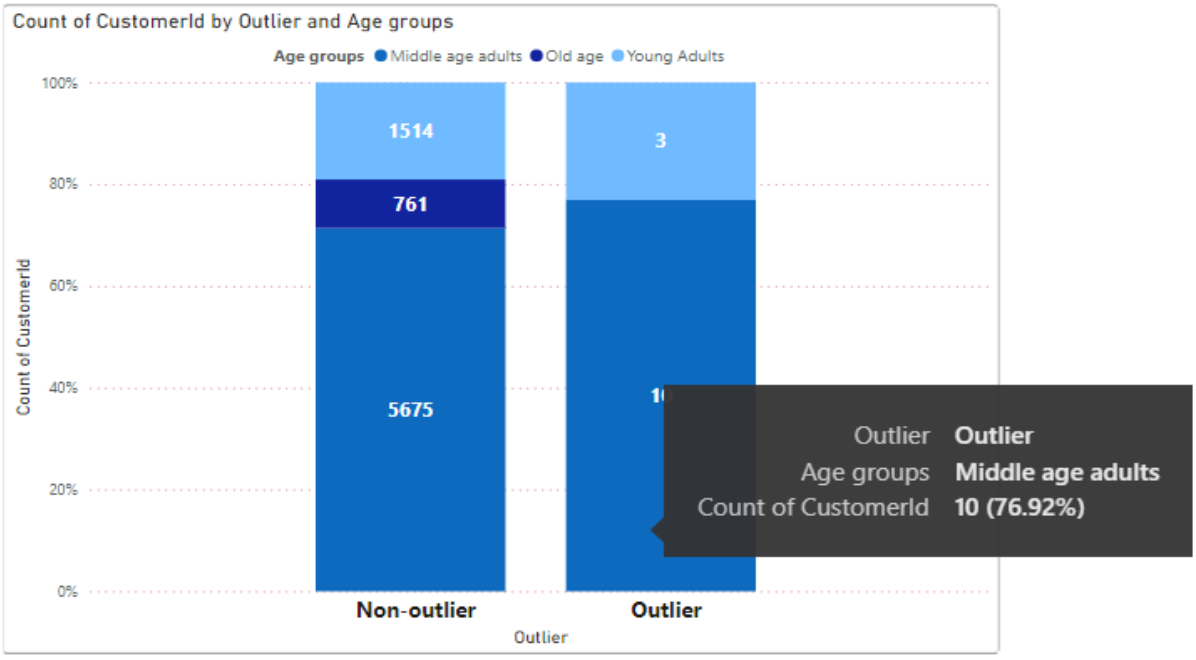
```
SELECT NumOfProducts,ROUND(AVG(Balance),2) AS Avg_acc_balance
FROM Bank_Churn
WHERE Exited='1'
GROUP BY NumOfProducts
ORDER BY NumOfProducts ASC;
```

	NumOfProducts	Avg_acc_balance
▶	1	92028.82
	2	90252.36
	3	85853.09
	4	93733.14

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

- There are 3 types of statistical methods to identify outliers such as IQR (interquartile Range), Standard deviation and Z-score.

- I Utilized Z-score statistical method among customer balance and different age groups who remained with bank.
- From the below visualisation we can observe few outliers in Middle age adults and young adults.



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

- From the given data there are 7 tables. Out of those 5 tables - Activecustomers, Creditcard, Exitcustomer, Gender, Geography are tables consisting of categorical variables.

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)

SQL Query:

```
SELECT
RANK () OVER (PARTITION BY GeographyID ORDER BY AVG(EstimatedSalary) DESC)
 AS 'Salary_rank',
GenderID,GeographyID,
ROUND(AVG(EstimatedSalary),2) AS Average_income
FROM Customerinfo
GROUP BY GenderID,GeographyID;
```

Salary_rank	GenderID	GeographyID	Average_income
1	Male	France	100174.25
2	Female	France	99564.25
1	Female	Germany	102446.42
2	Male	Germany	99905.03
1	Female	Spain	100734.11
2	Male	Spain	98425.69

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+).

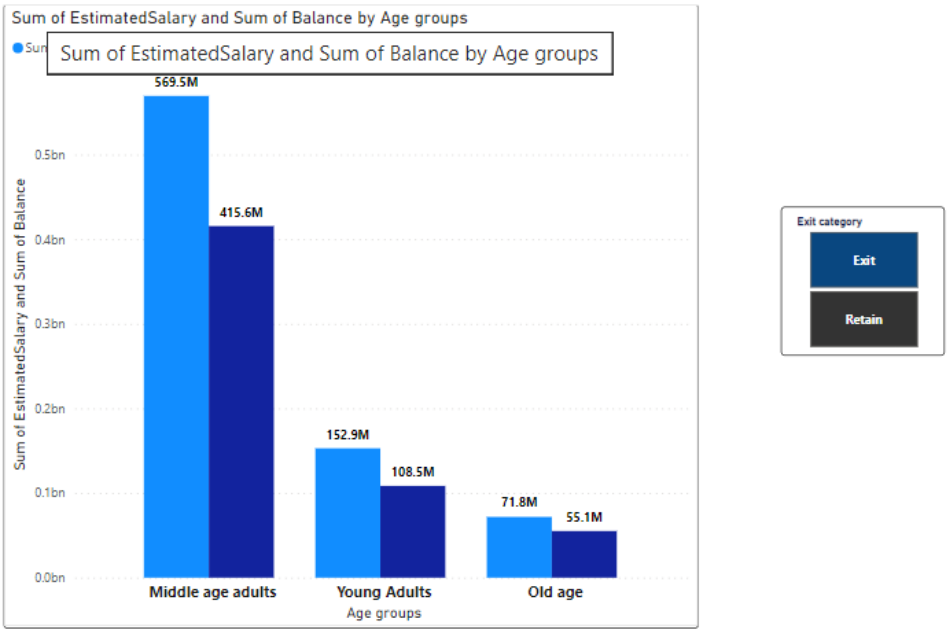
SQL Query:

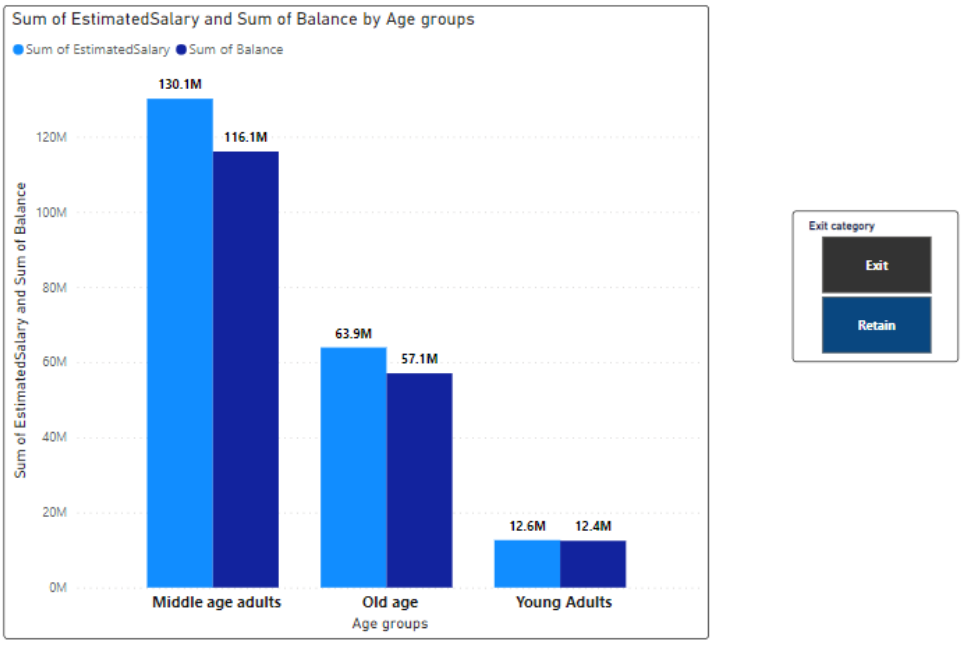
```
WITH CTE AS (SELECT bc.Tenure,
CASE
WHEN Age BETWEEN 18 AND 30 THEN 'Young Adults'
WHEN Age BETWEEN 31 AND 50 THEN 'Middle-aged Adults'
ELSE 'Old-aged Adults'
END AS Age_brackets
FROM Customerinfo c
JOIN Bank_Churn bc ON c.CustomerId=bc.CustomerId
WHERE Exited='1')

SELECT AVG(Tenure) AS Avg_tenure,Age_brackets
FROM CTE
GROUP BY Age_brackets
ORDER BY Age_brackets;
```

	Avg_tenure	Age_brackets
▶	4.8899	Middle-aged Adults
	4.8330	Old-aged Adults
	4.7770	Young Adults

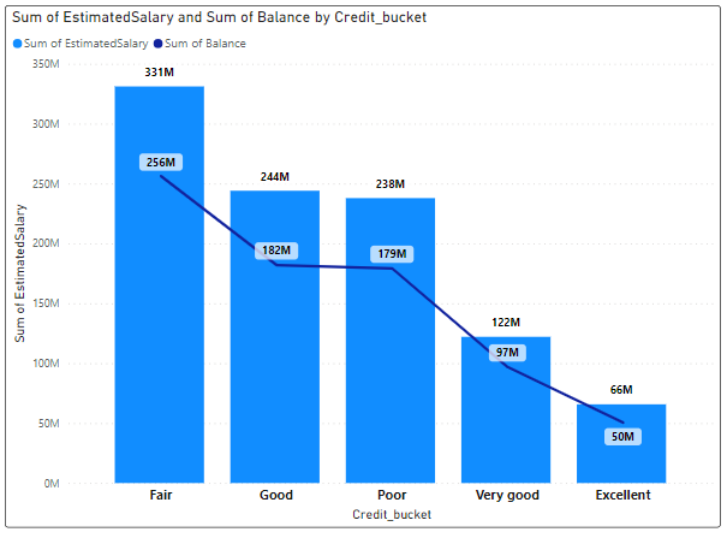
17. Is there any direct correlation between salary and balance of the customers? And is it different for people who have exited or not?





-From the above graph's relation between salary and balance of customers for customers who have exited and retained shows that there is a linear relation between Balance and Salary for customers who have exited the bank.

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



From the above chart, we can see that customer with
-Excellent Credit score have the lowest salary and least customer count.
-Fair Credit Score Customers have the Highest Salary and highest customer count.

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

SQL Query:

```
WITH CTE AS (
SELECT CustomerID,
CASE
WHEN CreditScore BETWEEN 350 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 799 THEN 'Very good'
ELSE 'Excellent'
END AS Credit_bucket
FROM Bank_Churn)

SELECT COUNT(Exited) AS Exit_rate,CTE.Credit_bucket,
RANK() OVER( ORDER BY COUNT(Exited) DESC) AS 'Rank'
FROM Bank_churn bc
JOIN CTE ON CTE.CustomerId=bc.CustomerId
WHERE Exited = '1'
GROUP BY CTE.Credit_bucket;
```

	Exit_rate	Credit_bucket	Rank
▶	685	Fair	1
	520	Poor	2
	452	Good	3
	252	Very good	4
	128	Excellent	5

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.

SQL Query:

```
WITH CTE AS
(SELECT CustomerID,
CASE
WHEN c.Age BETWEEN 18 AND 30 THEN 'Young Adults'
WHEN c.Age BETWEEN 30 AND 50 THEN 'Middle-aged Adults'
ELSE 'Old-aged Adults'
END AS Agebucket
FROM Customerinfo c)

SELECT
COUNT(c.CustomerId) AS Customercount,CTE.Agebucket
FROM  Customerinfo c
JOIN CTE ON c.CustomerId=CTE.CustomerId
JOIN Bank_churn bc ON CTE.CustomerId=bc.CustomerId
WHERE HASCrCard ='1'
GROUP BY CTE.Agebucket,HASCrCard;
```

	Customercount	Agebucket
▶	4781	Middle-aged Adults
	1400	Young Adults
	874	Old-aged Adults

--Also retrieve those buckets that have lesser than average number of credit cards per bucket.

```
WITH CTE AS
(SELECT CustomerID,
CASE
WHEN c.Age BETWEEN 18 AND 30 THEN 'Young Adults'
WHEN c.Age BETWEEN 30 AND 50 THEN 'Middle-aged Adults'
ELSE 'Old-aged Adults'
END AS Agebucket
FROM Customerinfo c)

SELECT
    Agebucket,
    SUM(bc.HASCrCard) AS Customercount
FROM  Bank_Churn bc
INNER JOIN
    CTE ON CTE.CustomerId = bc.CustomerId
GROUP BY  AgeBucket
HAVING SUM(bc.HASCrCard) < (SELECT AVG(NumCards)
                            FROM (SELECT Agebucket, SUM(HASCrCard) AS NumCards
                                  FROM Bank_Churn bc INNER JOIN CTE ON
                                  CTE.CustomerId = bc.CustomerId
```

```
GROUP BY Agebucket) AS subquery);
```

	Agebucket	Customercount
▶	Young Adults	1400
	Old-aged Adults	874

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

SQL Query:

```
SELECT
RANK()OVER(ORDER BY COUNT(c.CustomerID) DESC) AS 'Location_rank',
GeographyID,
COUNT(c.CustomerID) AS Customer_Count,
ROUND(AVG(Balance),2) AS AvgBalance
FROM CustomerInfo c
INNER JOIN Bank_Churn b
ON c.CustomerId = b.CustomerId
WHERE Exited = 1
GROUP BY GeographyID;
```

	Location_rank	GeographyID	Customer_Count	AvgBalance
▶	1	Germany	814	120361.08
	2	France	810	71192.8
	3	Spain	413	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”.

SQL Query:

```
SELECT
CustomerId,
Surname,
CONCAT(CustomerId,'_',Surname) AS CustomerId_Surname
FROM CustomerInfo;
```

CustomerId	Surname	CustomerId_Surname
15634602	Hargrave	15634602_Hargrave
15647311	Hill	15647311_Hill
15619304	Onio	15619304_Onio
15701354	Boni	15701354_Boni
15737888	Mitchell	15737888_Mitchell
15574012	Chu	15574012_Chu
15592531	Bartlett	15592531_Bartlett

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

SQL Query:

```
SELECT *,
(SELECT ExitCategory
FROM ExitCustomer E
WHERE E.ExitID = bc.Exited) AS ExitCategory
FROM Bank_Churn bc;
```

CustomerId	CreditScore	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	Exited	ExitCategory
15634602	619	7	0	1	1	1	1	Exit
15647311	608	4	83807.86	1	0	1	0	Retain
15619304	502	4	159660.8	3	1	0	1	Exit
15701354	699	3	0	2	0	0	0	Retain
15737888	850	3	125510.82	1	1	1	0	Retain
15574012	645	4	113755.78	2	1	0	1	Exit
15592531	822	6	0	2	1	1	0	Retain

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

--Using Power Query Editor in Power BI and Excel, I looked for duplicates and missing values. Prior to doing an analysis, I ensured that the dataset was optimized and that all data types were in the correct format with no missing values.

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”.

SQL Query:

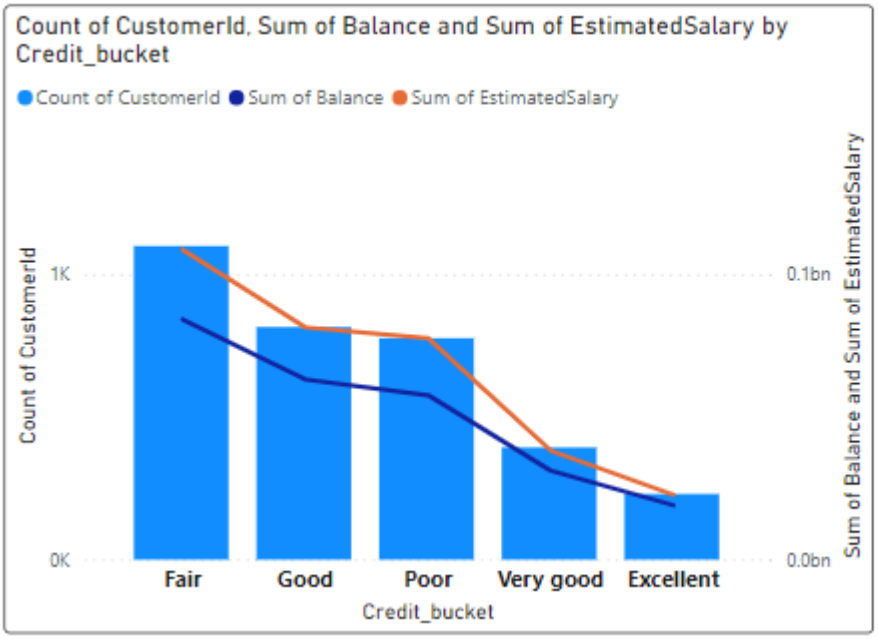
```
SELECT c.CustomerId,Surname,IsActiveMember,
```

```
(SELECT ActiveCategory FROM ActiveCustomer A WHERE
A.ActiveID=bc.IsActiveMember) AS 'Member_status'
FROM CustomerInfo c
JOIN Bank_Churn bc
ON c.CustomerId = bc.CustomerId
WHERE Surname LIKE '%on';
```

CustomerId	Surname	IsActiveMember	Member_status
15788218	Henderson	1	Active Member
15750181	Sanderson	0	Inactive Member
15788448	Watson	1	Active Member
15585768	Cameron	1	Active Member
15592461	Jackson	1	Active Member
15640635	Capon	1	Active Member
15676966	Capon	1	Active Member

SUBJECT QUESTIONS

1. Customer Behaviour 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?



-The above graph the distribution of Total Customers across different Credit Score buckets, revealing intriguing patterns and correlations. The highest number of customers are found in the Fair, Good, and Poor credit score segments and Very Good and Excellent segments has low number of customers.

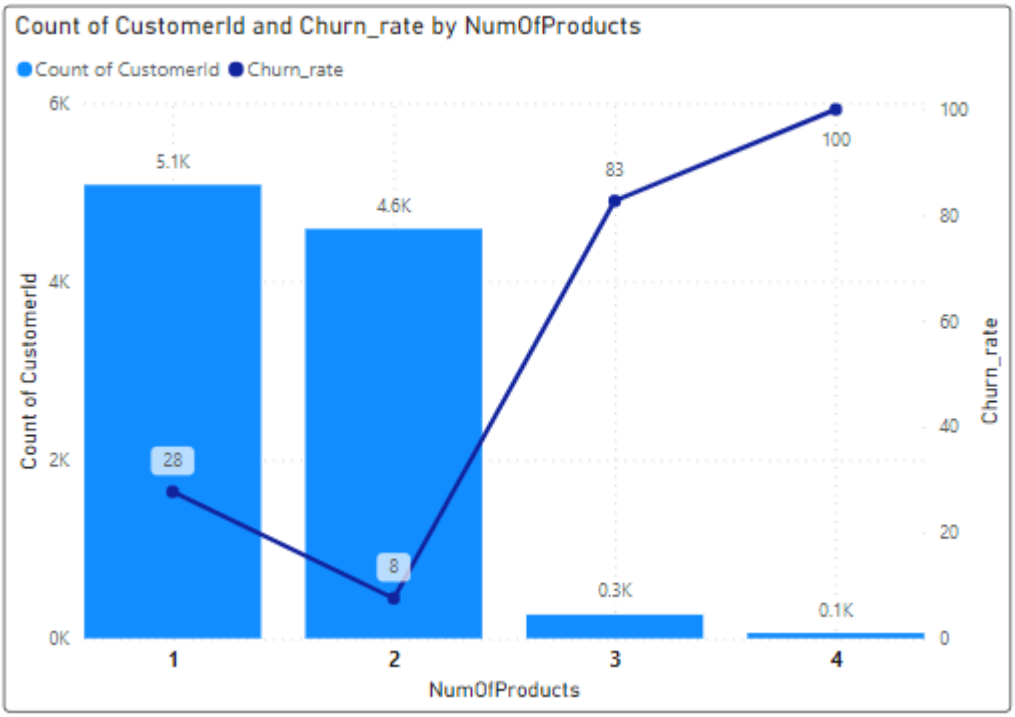
-Utilizing the slicer for a year-by-year analysis of the relationships among these variables, this phenomenon suggests that individuals with higher credit scores tend to manage their finances more prudently.

-Their balances closely align with their salaries, indicating wise spending habits.

-For other segments we notice significant gaps between the amounts in their balances and their salaries, implying the presence of potentially detrimental spending behaviours.

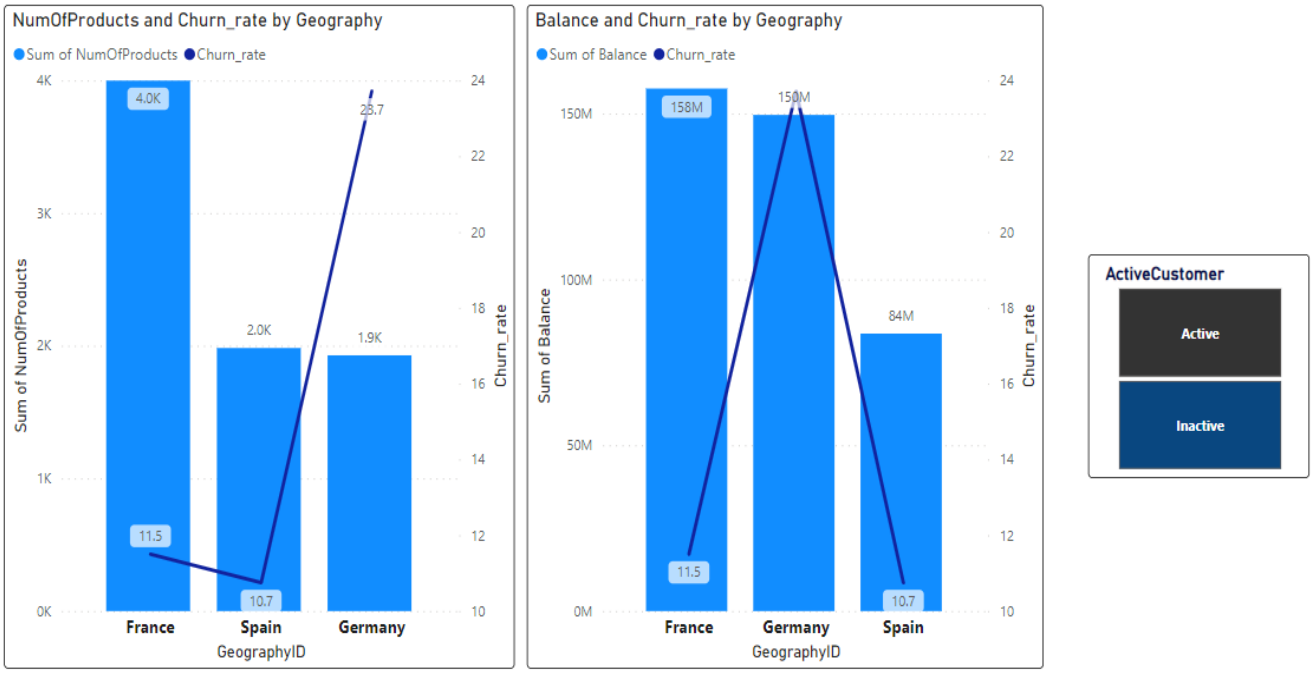
-These habits likely contribute to their lower credit scores.

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



The maximum customers use 1 and 2 products. No. of customers using 3 and 4 products are less comparatively. The churn rates of customers using 3 and 4 products are also high with 83% and 100% respectively.

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



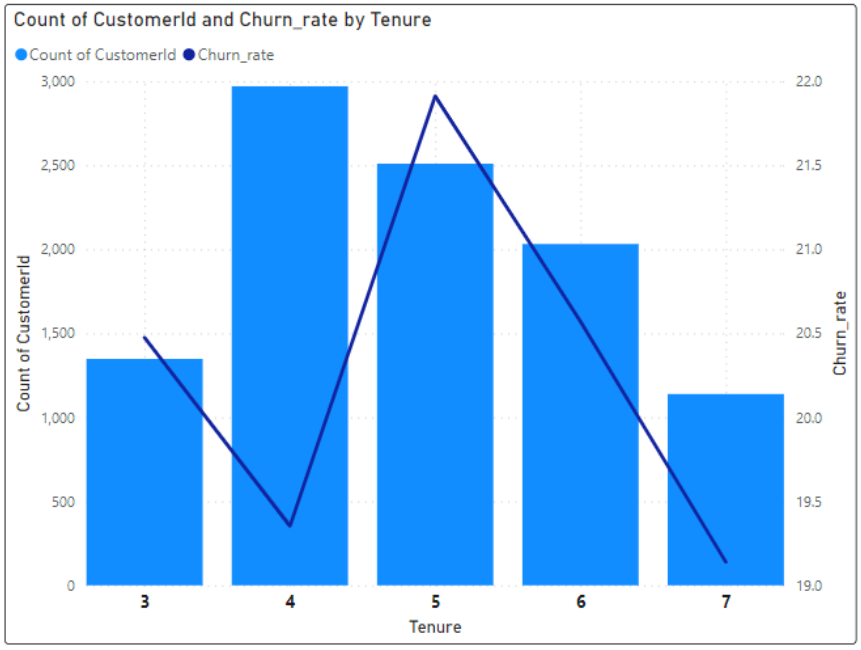
- In terms of account balances, France and Germany are leading, while Spain is lagging behind.

- And in terms of Number of products France is leading while Spain and Germany are lagging.

- While comparing churn rates with Number of products and Balance, I found out that churn rates are significantly higher in Germany for Active customers.

-This can be a key indicator for RISK and should be treated with priority

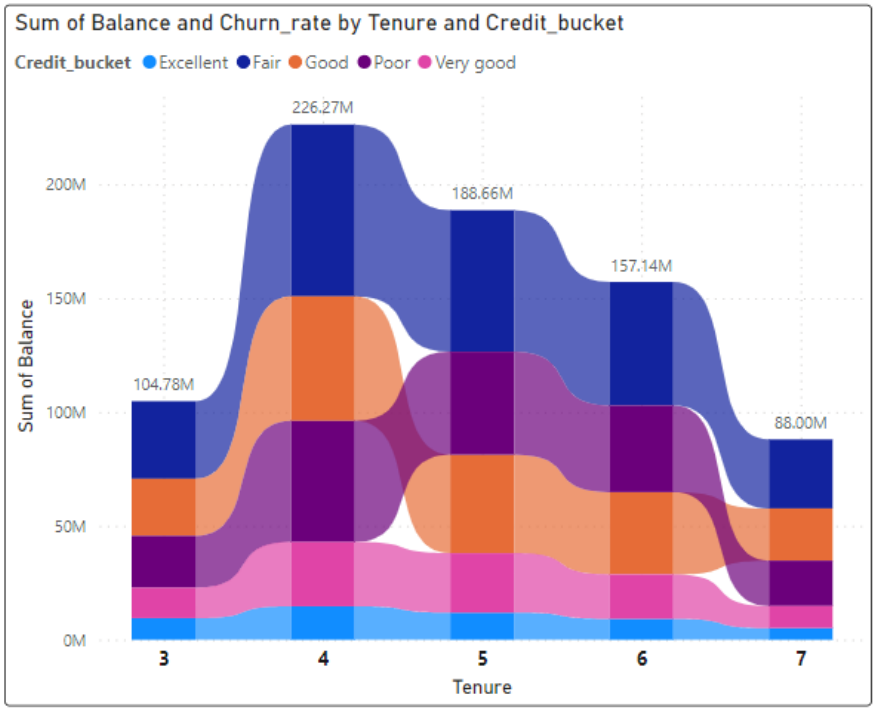
4. Risk Management Assessment: Based on customer profiles, which demographic segments appear to pose the highest financial risk to the bank, and why?



- The churn rates decline from the 3 to 4 tenures.
- Also, we observe that the churn rate rises again in tenure 5, and that this decline continues as tenure grows.

- This suggests that customer will remain with the bank for 4 to 5 years, after which they might leave.
- This indicates a high risk for the bank because there is a good probability the customer will leave once their five-year tenure is over.

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



- The largest balance is found in the Fair Segment of customers when we look at the balance distribution across all of the credit card segments.
- The Excellent and Very Good segments have the least amount of balance.
- There are noticeable balances in Poor segment as well.

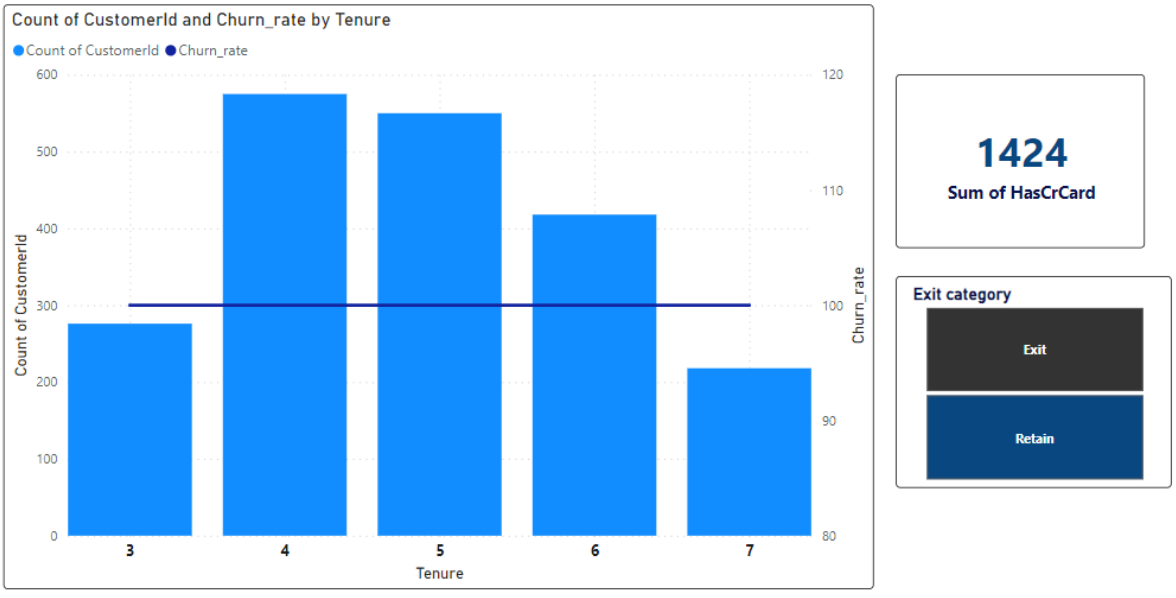
- As we see other tenures, this pattern continues.

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

   - To assess the impact of marketing campaigns on customer retention and acquisition within the dataset, we can measure changes in customer activity status and acquisition rates over time, comparing periods before and after the campaigns.

   - Continuous Monitoring: Establish a system for continuous monitoring and evaluation of marketing campaign effectiveness. Iterate and refine analysis based on new data and insights gained over time.

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



- About 1424 of the 7055 customers that left having credit cards, out of the total number of credit card holders. This demonstrates that customers without credit cards are leaving the bank.

- As tenure increases, we observe customers leaving the bank. This indicates that we are unable to keep customers for longer periods of time.

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

    - Of course, these are significant demographics that highlight several aspects of bank and customer, as we've covered in the questions all of these factors are crucial for analysing customer behaviour and can help estimate if a customer will leave the bank or not.

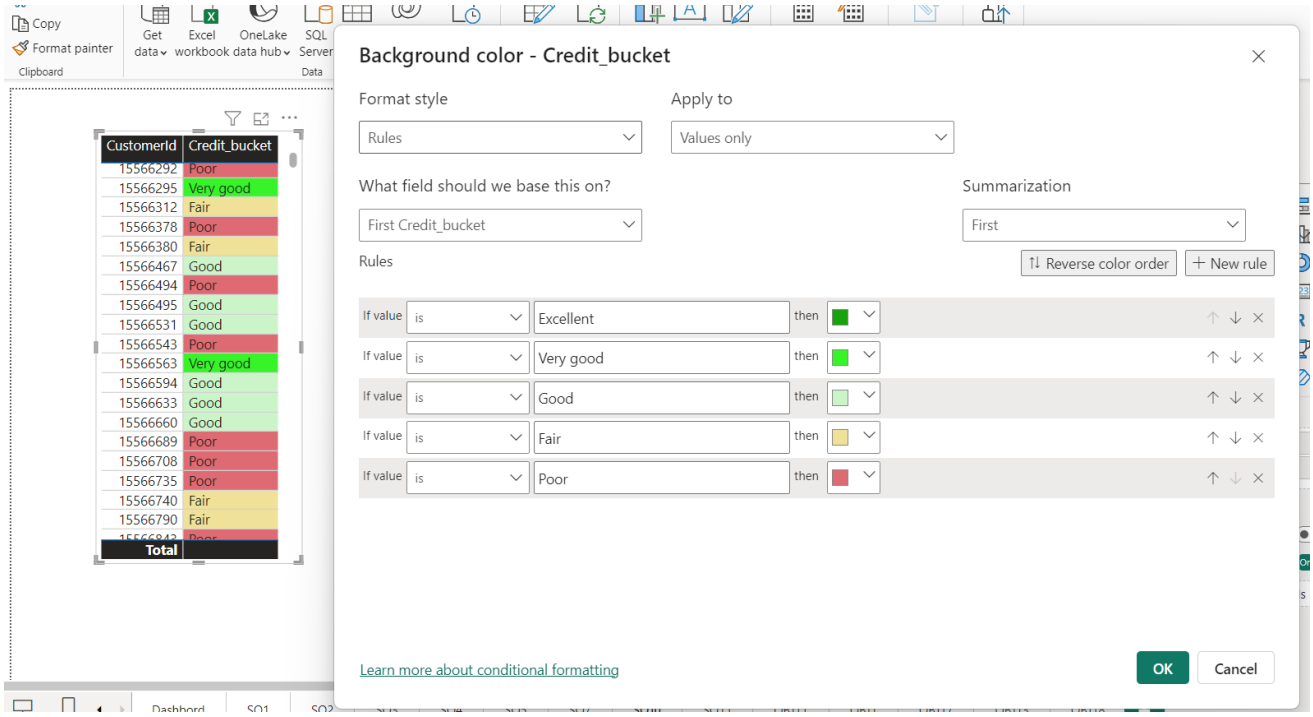
9. Utilize SQL queries to segment customers based on demographics and account details.

```sql
WITH CTE AS (
SELECT CustomerID,
CASE
WHEN CreditScore BETWEEN 350 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 799 THEN 'Very good'
ELSE 'Excellent'
END AS Credit_bucket
FROM Bank_Churn)

SELECT Tenure,
NumOfProducts,Credit_bucket,
COUNT(CTE.CustomerId) AS CustomerCount,
ROUND(AVG(Balance), 2) AS AverageBalance
FROM Bank_Churn bc
JOIN CTE ON   bc.CustomerId=CTE.CustomerId
GROUP BY Credit_bucket,Tenure,NumOfProducts
ORDER BY Tenure,NumOfProducts;
```

	Tenure	NumOfProducts	Credit_bucket	CustomerCount	AverageBalance
▶	3	1	Good	163	98427.27
	3	1	Excellent	54	107421.23
	3	1	Very good	77	98937.91
	3	1	Poor	153	95729.85
	3	1	Fair	203	97806.88
	3	2	Very good	85	63060.37
	3	2	Good	147	54750.68

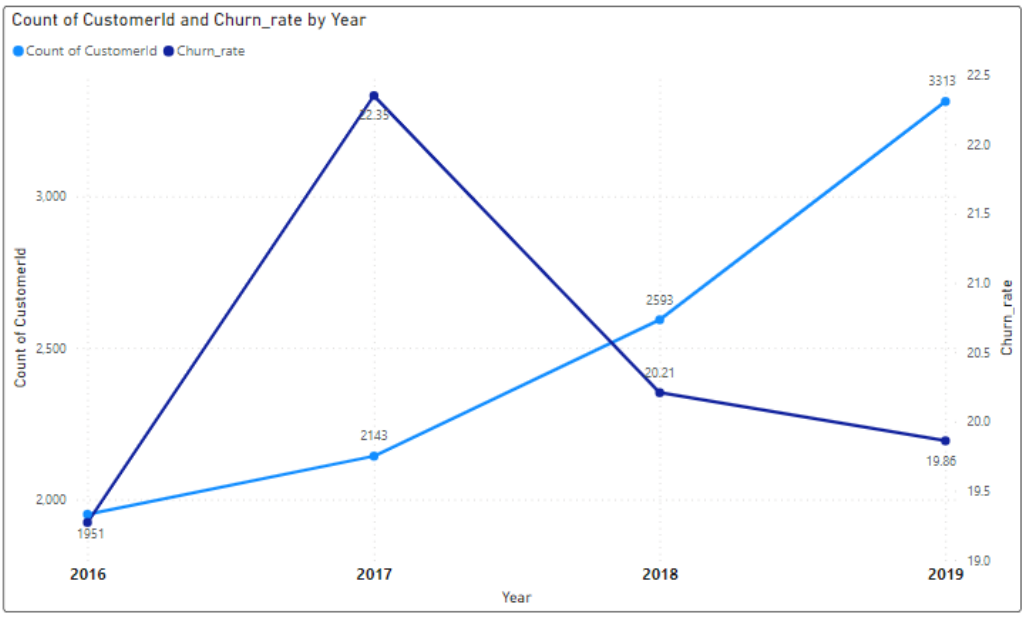
10. How can we create a conditional formatting setup to visually highlight customers at risk of churn and to evaluate the impact of credit card rewards on customer retention?



-By creating a table view and after inserting the data. By clicking on the data, we can see the conditional formatting option in which we can colour each data points based on the conditions.

And customer with poor credit score is high risk of leaving the bank.

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



- Customers with Poor credit scores and those in the young age adults and old age adult groups are the most likely to leave the bank because churn rates are highest in these categories.

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

    - Created a Power bi Dashboard to visualise all questions, KPIs and related metrics and submitted the .pbix file with in the zip file which I have submitted.

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

    - I would approach analysing this problem similarly, whether it involves objective or subjective questions. My initial step would involve identifying all KPIs within the data. Subsequently, I would structure my reports to comprehensively address these KPIs, ensuring coverage across all relevant aspects.

    - Furthermore, I would look into various demographics and examine their impact on our KPIs. By incorporating demographic breakdowns, I aim to understand how different segments influence our metrics.

    To enhance the insights, I would utilize slicers to segment the data based on relevant parameters. This approach facilitates a more detailed exploration of the data, allowing for thorough analysis and actionable insights.

14. In the “Bank_Churn” table how can you modify the name of “HasCrCard” column to “Has_creditcard”?

    SQL query:

    ```
    ALTER TABLE bank_churn RENAME COLUMN HasCrCard TO Has_creditcard;
    ```

    -We can modify the column name using the above query but changing column name after analysing the data could affect the previous queries.