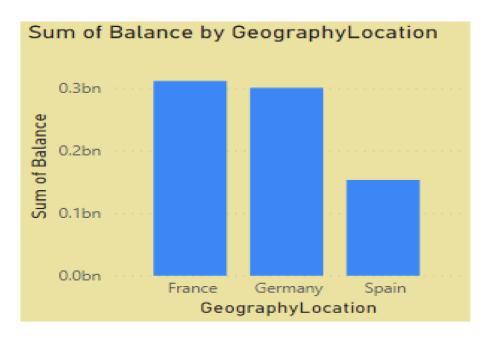
Objective Questions:

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



The Column Chart below illustrates the distribution of account balances across various regions, providing insight into how the total account balance is allocated among three specific regions.

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

CustomerId	join_year	join_quarter	max_salary
15634359	2016	4	199970.74
15804211	2019	4	199841.32
15687913	2018	4	199805.63
15763065	2019	4	199753.97
15599792	2018	4	199638.56

SELECT

CustomerId, YEAR(BankDOJ) AS join_year, QUARTER(BankDOJ) AS join_quarter, MAX(EstimatedSalary) AS max_salary

FROM CustomerInfo

WHERE QUARTER (BankDOJ) = 4

GROUP BY CustomerId, join_year, join_quarter

ORDER BY max_salary DESC LIMIT 5;

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

Custome	erId avg(NumOfProduc	ets)
1563460	2 1.0000	
1561930	4 3.0000	
1573788	8 1.0000	
1557401	2 2.0000	
1559253	1 2.0000	
1565614	8 4.0000	
1559238	9 1.0000	
1573717	3 2.0000	
1563226	2.0000	
1560088	2.0000	
1573745	1.0000	
1578821	8 2.0000	
1556898	2.0000	
1557765	2.0000	
1559794	5 2.0000	

Query:

```
select CustomerId, avg(NumOfProducts) from Bank_churn
where HasCrCard = 1
group by customerId;
```

Total count of customerId who have CreditCard is 7055

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

GenderCategory	total_custom	churned_custom	churn_rate
Male	1776	273	0.1537
Female	1537	385	0.2505

SELECT g.GenderCategory, COUNT(*) AS total_customers,

SUM(CASE WHEN bc.Exited = 1 THEN 1 ELSE 0 END) AS churned_customers,

(SUM(CASE WHEN bc.Exited = 1 THEN 1 ELSE 0 END) / COUNT(*)) AS churn_rate

FROM BANK_CHURN bc

JOIN CustomerInfo ci ON bc.CustomerId = ci.CustomerId

JOIN Gender g ON ci.GenderId = g.GenderId

JOIN (SELECT MAX(YEAR(BankDOJ)) AS max_year FROM CustomerInfo) AS max_year_subquery

WHERE YEAR(ci.BankDOJ) = max_year_subquery.max_year

GROUP BY g.GenderCategory;

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

exit_status	avg_credit_sc
	645.3515
fornonexited	651.8532

SELECT

CASE WHEN Exited = 1 THEN 'forexited' ELSE 'fornonexited' END AS exit_status,

AVG(CreditScore) AS avg_credit_score

FROMBANK_CHURN

GROUP BY CASE WHEN Exited = 1 THEN 'forexited' ELSE 'fornonexited' END;

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

GenderCategory	avg_salary	active_accou
	100601.54	2284
Male	99664.58	2867

Query 😀

WITH AvgSalary AS (SELECT GenderId, ROUND(AVG(EstimatedSalary), 2) AS avg_salary

FROM CustomerInfo GROUP BY GenderId)

SELECT g.GenderCategory, a.avg_salary, SUM(bc.IsActivemember) AS active_accounts

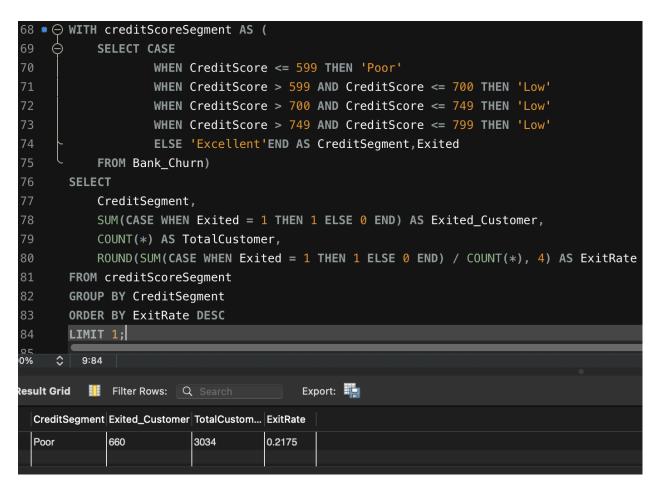
FROM Gender g JOIN AvgSalary a ON g.GenderId = a.GenderId

JOIN CustomerInfo ci ON g.GenderId = ci.GenderId

JOIN BANK_CHURN bc ON ci.CustomerId = bc.CustomerId

GROUP BY g.GenderCategory, a.avg_salary;

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



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

```
select geo.GeographyLocation, sum(bank.IsActivemember) from Geography geo
join CustomerInfo cust on
geo.GeographyID = cust.GeographyID
join BANK_CHURN bank on
cust.CustomerId = bank.CustomerId
where bank.IsActivemember = 1
and bank.Tenure > 5
group by geo.GeographyLocation
order by 2 desc
limit 1;
```

GeographyLocati	sum(bank.lsActivememb
France	797

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

```
CASE

WHEN bank.HasCrCard = 0 THEN 'No_Credit'
ELSE 'Credit_Card'
END AS Creditcard_Status,
COUNT(bank.exited) AS ChurnedCustomers

FROM bank_churn bank
WHERE bank.exited = 1

GROUP BY bank.HasCrCard;
```

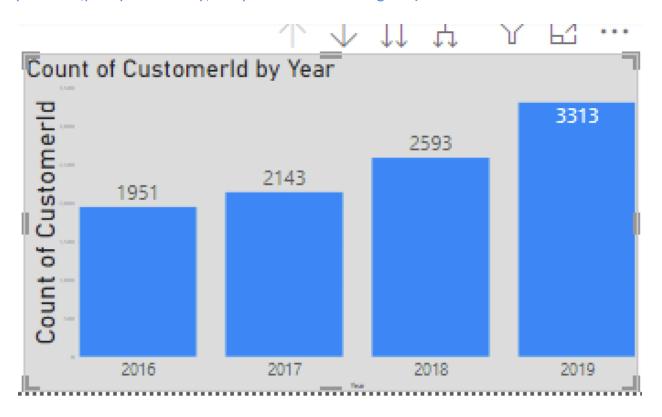
Creditcard_Stat	ChurnedCustome
	613
Credit_Card	1424

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

```
select NumOfProducts as MostCommonProduct, count(*) as CountProducts
from BANK_CHURN where Exited=1
group by NumOfProducts
order by count(*) desc limit 1;
```

MostCommonProduct	CountProducts
1	1409

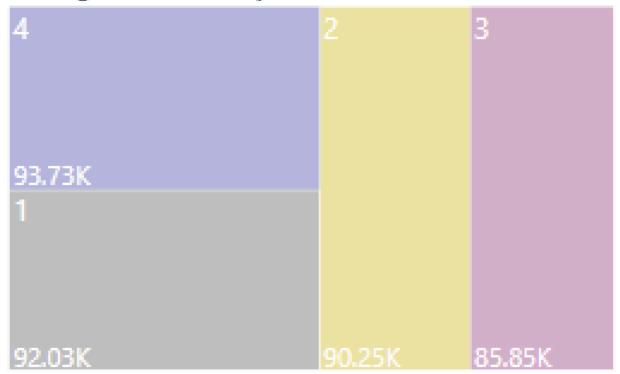
11. Examine the trend of customers joining over time and identify any seasonal patterns (yearly or monthly). Prepare the data through SQL and then visualize it.



We can in above column chart Every year Customer increasing and maximum in 2019

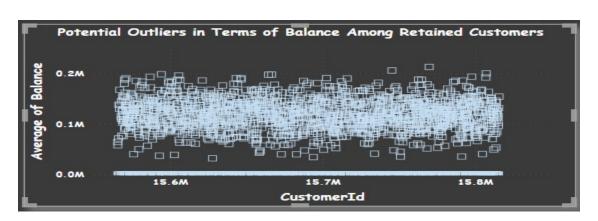
12. Analyze the relationship between the number of products and the account balance for customers who have exited.

Average of Balance by NumOfProducts



As depicted in the chart above, customers who have made purchases of 4 products tend to maintain a notably higher average balance in their accounts compared to other customer segments.

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



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

Total 7 different tables present in database where 5 tables has categorical data

Tables_in_newton_bank_crm
ActiveCustomer
Bank_churn
CreditCard
CustomerInfo
ExitCustomer
Gender
Geography

SHOW TABLES;

Where ActiveCustomer, CreditCard, ExitedCustomer, Gender, Geography which can divide into two or more category

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

```
select distinct case when GenderID = 1 then "Male" else "Female" end as "Gender",
    avg(case when GeographyID=1 then EstimatedSalary else 0 end) as "France",
    avg(case when GeographyID=2 then EstimatedSalary else 0 end) as "Spain",
    avg(case when GeographyID=3 then EstimatedSalary else 0 end) as "Germany",
    rank() over ( order by avg(EstimatedSalary) )as "Rank_Gender"
from CustomerInfo
group by GenderID;
```

Gender	France	Spain	Germany	Rank_Gender
Male	50536.87321238767	25034.791002382262	24092.912715777868	1
Female	49552.00869029278	24146.916803874108	26902.6158881796	2

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+).

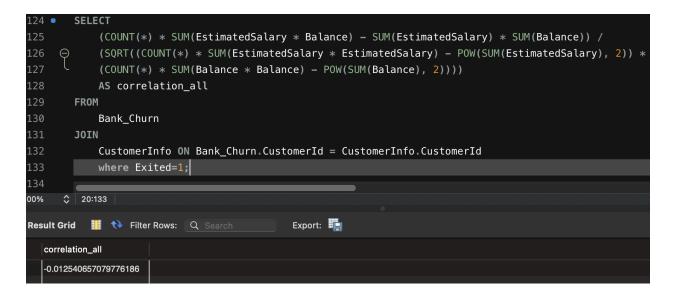
Below query is finding the average tenure of the people who have exited in each age bracket (18-30, 30-50, 50+).

```
    when age between 18 and 30 then "18-30"
    when age between 31 and 50 then "31-50"
    else "50+" end as AgeGroup,
        avg(EstimatedSalary) as avg_salary
    from CustomerInfo
    join Bank_Churn on CustomerInfo.CustomerId=Bank_Churn.CustomerId
    where Exited=1
    group by case
        when age between 18 and 30 then "18-30"
        when age between 31 and 50 then "31-50"
        else "50+" end;
```

AgeGroup	avg_salary		
31-50	102095.68050527928		
18-30	101109.35317567564		
50+	100075.53909413857		

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

```
SELECT
            (COUNT(*) * SUM(EstimatedSalary * Balance) - SUM(EstimatedSalary) * SUM(Balance)) /
            (SQRT((COUNT(*) * SUM(EstimatedSalary * EstimatedSalary) - POW(SUM(EstimatedSalary), 2)) *
            (COUNT(*) * SUM(Balance * Balance) - POW(SUM(Balance), 2))))
           AS correlation_all
       FROM
           Bank_Churn
       JOIN
           CustomerInfo ON Bank_Churn.CustomerId = CustomerInfo.CustomerId
133
           where Exited=0;
134
00%
                                             Export:
Result Grid
          Filter Rows: Q Search
  correlation_all
  0.017198580691270192
```



As we can see in above two query one for Exited customers and second one for non exited customer-

positive for non exited, direct relation between balance and salary

Negative for exited, Inverse relation between balance and salary

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

```
SELECT
138
              (COUNT(*) * SUM(EstimatedSalary * CreditScore) - SUM(EstimatedSalary) * SUM(CreditScore)) /
              (SQRT((COUNT(*) * SUM(EstimatedSalary * EstimatedSalary) - POW(SUM(EstimatedSalary), 2)) *
140
              (\texttt{COUNT}(*) \ * \ \texttt{SUM}(\texttt{CreditScore} \ * \ \texttt{CreditScore}) \ - \ \texttt{POW}(\texttt{SUM}(\texttt{CreditScore}), \ \textcolor{red}{\textbf{2}}))))
141
              AS correlation_all
142
         FROM
143
              Bank_Churn
144
         JOIN
145
              CustomerInfo ON Bank_Churn.CustomerId = CustomerInfo.CustomerId;
00%
                                                       Export:
Result Grid 🏢 💎 Filter Rows: 🔍 Search
   correlation_all
   -0.0013842928679876046
```

Mild inverse relation between salary and Credit Score

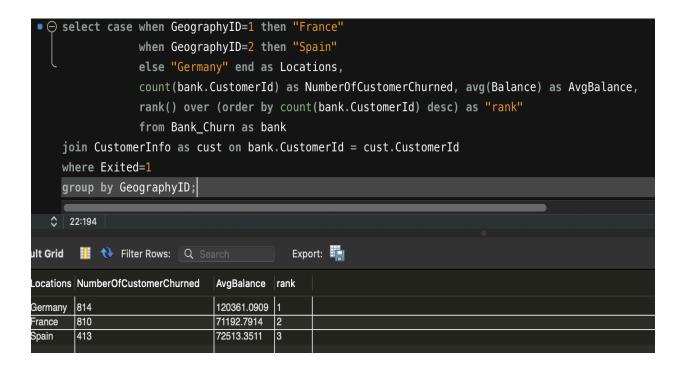
19. Rank each bucket of credit score as per the number of customers who have churned the bank.

```
149 • ⊖ with ChurnCustomers as (select case
           when CreditScore<=599 then 'Poor'
           when CreditScore>599 and CreditScore<=700 then 'Low'
           when CreditScore>700 and CreditScore<=749 then 'Fair'
           when CreditScore>749 and CreditScore<=799 then 'Low'
           else 'Excellent'end as Credit_Category, count(*) as churned_count
            from Bank_Churn where Exited= 1
           group by case
           when CreditScore<=599 then 'Poor'
           when CreditScore>599 and CreditScore<=700 then 'Low'
           when CreditScore>700 and CreditScore<=749 then 'Fair'
           when CreditScore>749 and CreditScore<=799 then 'Low'
           else 'Excellent'end),
           RankBucket as (select Credit_Category,Churned_count, rank() over (order by churned_count desc)as Rank_Bucket
            from ChurnCustomers)
            select Credit_Category, Churned_count,Rank_Bucket from RankBucket;
     $ 36:158
00%
                                         Export:
Result Grid III Filter Rows: Q Search
  Credit_Catego... Churned_count Rank_Bucket
              951
 Poor
              298
              128
```

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

```
-- Question 20
166 • ⊖ WITH CreditCardCounts AS (
            SELECT
     \Diamond
                CASE WHEN cust.Age BETWEEN 18 AND 30 THEN '18-30'
                     WHEN cust.Age BETWEEN 31 AND 50 THEN '31-50' ELSE '50+' END AS ageBucket,
170
                COUNT(bank.HasCrCard) AS CreditCardCount
171
            FROM Bank_Churn bank
            JOIN CustomerInfo cust ON bank.customerid = cust.customerid
            WHERE HasCrCard = 1
174
            GROUP BY ageBucket
176
     SELECT AVG(CreditCardCount) AS avgCreditCard
178
            FROM CreditCardCounts
        SELECT ageBucket,CreditCardCount, avgCreditCard
        FROM CreditCardCounts
        CROSS JOIN AvgCreditCards
        WHERE CreditCardCount < avgCreditCard;</pre>
100%
      $ 35:180
Result Grid Filter Rows: Q Search
                                          Export:
   ageBucket CreditCardCou... avgCreditCard
          1400
                       2351.6667
   50+
          874
                       2351.6667
```

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



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
    cust.*,
    concat(cust.CustomerId, ' ', cust.SurName) as Customerid_Surname
from
    CustomerInfo cust
join
    Bank_Churn bank on cust.CustomerId = bank.customerId;
```

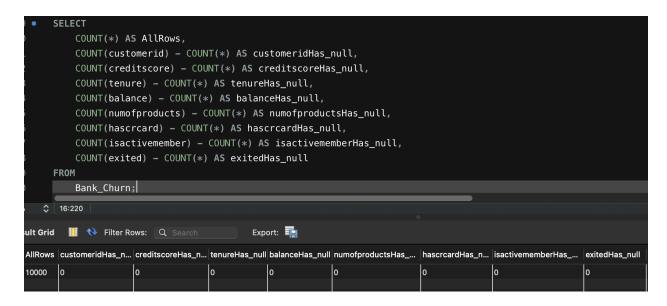
CustomerId	Surname	Age	GenderID	EstimatedSalary	GeographyID	BankDOJ	Customerid_Surname
15565701	Ferri	39	2	90212.38	2	2018-12-11	15565701 Ferri
15565706	Akobundu	35	1	83256.26	2	2019-09-10	15565706 Akobundu
15565714	Cattaneo	47	1	96517.97	1	2019-10-02	15565714 Cattaneo
15565779	Kent	30	2	188258.49	3	2019-08-27	15565779 Kent
15565796	Docherty	48	1	74510.65	3	2018-10-21	15565796 Docherty
15565806	Toosey	38	1	30583.95	1	2018-06-28	15565806 Toosey
15565878	Bates	29	1	197963.46	2	2019-08-25	15565878 Bates
15565879	Riley	28	2	56185.98	1	2017-11-23	15565879 Riley
15565891	Dipietro	39	1	56214.09	1	2018-08-14	15565891 Dipietro

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

```
select
    bank.*,
    (select exi.ExitCategory
from ExitCustomer exi
  where exi.ExitId = bank.Exited) as ExitCategory
from Bank_Churn bank;
```

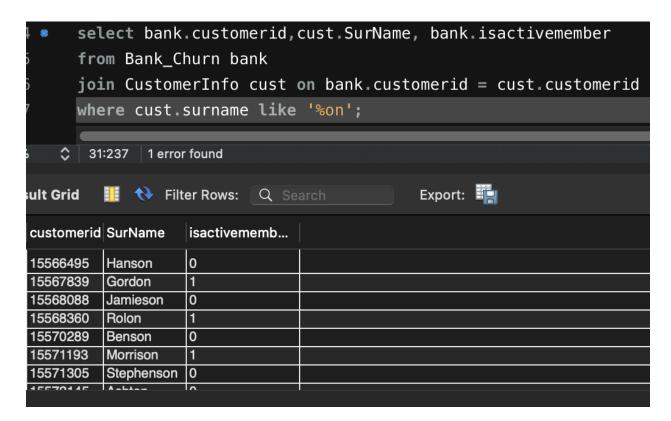
CustomerId	CreditScore	Tenure	Balance	NumOfProduc	HasCrCard	IsActiveMemb	Exited	ExitCategory
15634602	619	7	О	1	1	1	1	Exit
15647311	608	4	83808	1	0	1	0	Retain
15619304	502	4	159661	3	1	0	1	Exit
15701354	699	3	0	2	0	0	0	Retain
15737888	850	3	125511	1	1	1	0	Retain
15574012	645	4	113756	2	1	0	1	Exit
15592531	822	6	0	2	1	1	0	Retain
15656148	376	5	115047	4	1	0	1	Exit
15792365	501	5	142051	2	0	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?



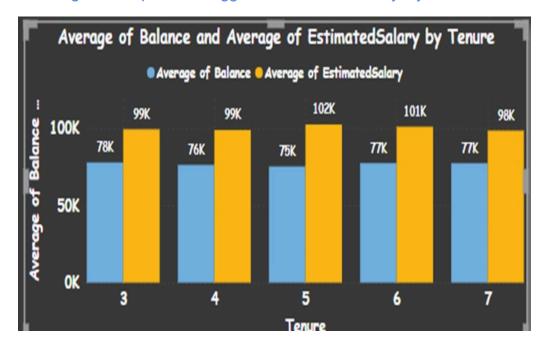
We can see no null values in Bank Churn table

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".



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?

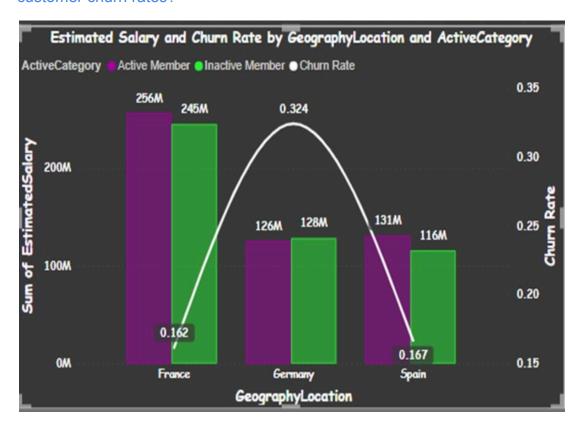


The mean estimated salary offers an insight into the income range of customers, while the mean balance portrays the funds remaining in their accounts post-expenses. These metrics shed light on customer expenditure tendencies. Intriguingly, there seems to be no clear trend in spending habits between new and established customers, as their account balances exhibit a comparable pattern.

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

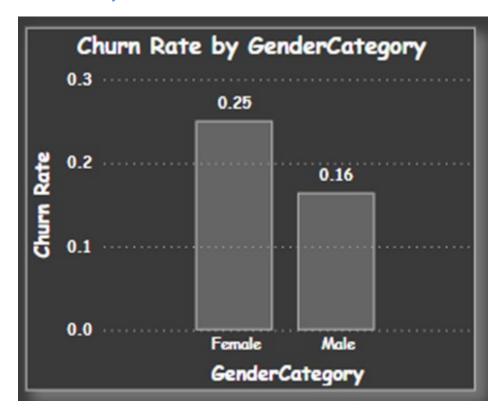
The dataset provides only the number of products purchased by each customer, without any specific information about individual products or their combinations. As a result, it is not feasible to glean insights on product affinity or propose cross-selling strategies based on the provided data.

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



The chart illustrates fluctuations in churn rates across various geographic locations, highlighting Germany as having the highest churn rate among the three countries analyzed. Interestingly, France stands out with a higher total estimated salary and a larger count of active members compared to the other countries.

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



Analysis of customer profiles suggests that the gender category represents the greatest financial risk to the bank, as female customers exhibit a notably higher churn rate compared to their male counterparts. The churn rate increases by nearly 10% when transitioning from male to female customers.

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

Credit Category By Tenure						
Tenure	credit card holder	non credit card holder				
3	963	385				
4	2056	915				
5	1791	719				
6	1444	588				
7	801	338				

Based on the visual representation and the data provided, it can be deduced that customers holding a credit card are instrumental in generating interest income for the bank. This group of customers is likely to contribute significantly to the bank's revenue stream over time.

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 conduct a thorough analysis of the impact of marketing campaigns on customer retention and acquisition, it's essential to gather comprehensive information. This includes data on the cost of marketing campaigns, the churn rate post-campaign, and the acquisition rate both before and after the campaigns. While the dataset may provide information on the churn rate, it's crucial to ascertain whether this churn rate pertains to periods before or after the marketing campaigns were executed. This distinction is vital for accurately evaluating the effectiveness of the campaigns in retaining existing customers and acquiring new ones.

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

Relying solely on attributes such as credit score and active membership status makes it difficult to pinpoint specific reasons for customer attrition

within the bank. While these metrics may offer some insights into customer behavior, they don't directly address the root causes of churn.

Customer churn is influenced by a diverse array of factors, ranging from satisfaction with products or services to experiences with customer service and even external market dynamics.

Therefore, to gain a deeper understanding of why customers leave and to extract actionable insights, it's essential to complement the existing dataset with additional sources of information. This could entail gathering feedback directly from customers, administering surveys, analyzing qualitative data, and conducting thorough evaluations of customer interactions and experiences.

By integrating quantitative data with qualitative findings, organizations can more effectively identify recurring patterns, emerging trends, and pain points contributing to customer attrition. This comprehensive approach enables businesses to develop targeted strategies and initiatives aimed at mitigating churn, enhancing overall customer satisfaction, and fostering improved retention rates.

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

Determining the underlying reasons for customer departure based solely on the provided attributes presents a considerable challenge. The dataset highlights instances where even active members and customers with long-standing relationships have opted to discontinue their association with the bank. Additionally, attributes such as the number of products used and estimated salary fail to yield definitive insights into the drivers of customer churn.

To enhance the accuracy of predicting customer attrition, soliciting feedback directly from customers regarding their motivations for leaving becomes indispensable. This feedback serves as a critical tool in evaluating whether these attributes indeed hold significant predictive value in discerning customer churn. By incorporating customer input, a clearer comprehension of the factors influencing churn can be achieved, facilitating more precise predictions and informed strategic decision-making.

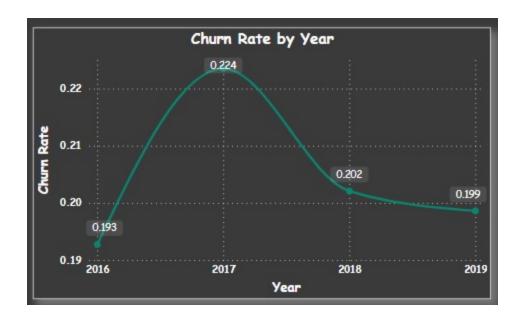
9. Utilize SQL queries to segment customers based on demographics and account details.

Objective questions involved the utilization of SQL queries to categorize customers according to different demographic factors and account particulars such as gender, age, credit score, etc. These categorized segments were subsequently utilized for additional analysis and exploration.

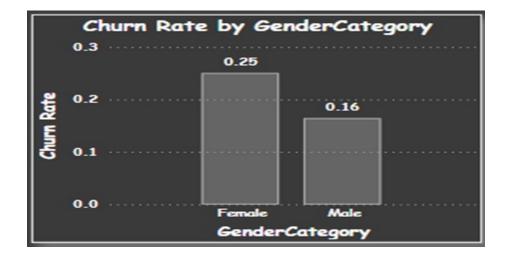
10. How can we create a conditional formatting setup to visually highlight customers at risk of churn and to evaluate the impact of credit card rewards on customer retention?

Though we can establish conditional formatting to identify potentially churn-prone customers, its application to charts is restricted. This is because our current focus is on scrutinizing the causes of customer churn via churn rates categorized by locations and gender. Furthermore, significant elements like credit card rewards, crucial for evaluating their impact on customer retention, are not included in the dataset.

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



The churn rate per annum is depicted in the provided image, with an overall churn rate standing at 0.20 throughout the bank. From the data analysis, it appears that female customers exhibit a higher likelihood of churning compared to their male counterparts, as evidenced by their elevated churn rate.



To counteract this trend and mitigate customer attrition, potential strategies could include offering credit rewards to credit card users, reassessing interest rates, or extending loan durations for these customers. These measures aim to incentivize customer retention and cultivate enduring relationships with the bank.

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

Utilizing additional data such as product specifications, interest income contributions, and customer feedback, my strategy entails assessing the value of customers and formulating retention strategies tailored to retain high-value clientele and minimize churn. By scrutinizing feedback provided by customers who have exited the bank, I aim to pinpoint areas of dissatisfaction and address these specific concerns to bolster retention efforts.

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

In Power BI, it's convenient to rename a column by simply right-clicking on its name and choosing the "Rename" option from the menu that appears. This intuitive process enables users to easily update column names according to their requirements.

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

I've created a thorough dashboard using Power BI, featuring all pertinent key performance indicators (KPIs) and crucial visuals. Moreover, I've integrated slicers into the dashboard to streamline selection processes, offering improved user engagement and analytical capabilities.