Newton School

Bank CRM Analysis

Power BI Project By Souparna Maity

Objective Questions

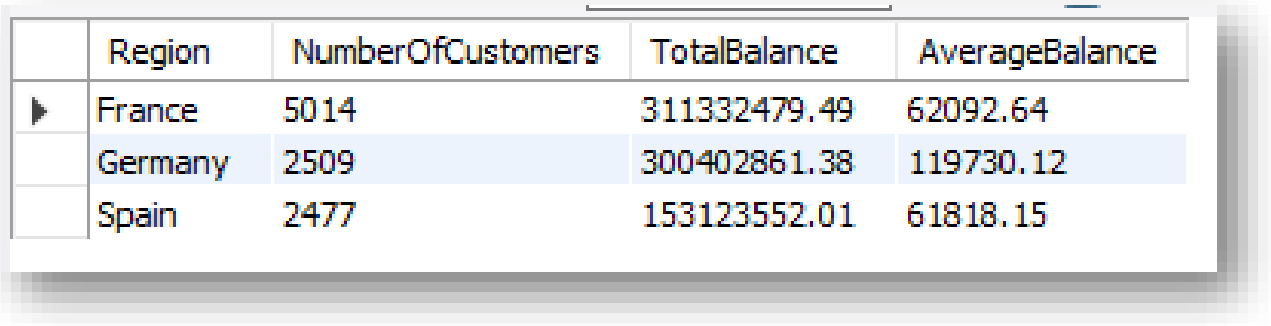
Objective Questions

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

Query:

```
Select
g.GeographyLocation as Region,
Round(Count(bc.Balance),2) as NumberOfCustomers,
Round(SUM(bc.Balance),2) as TotalBalance,
Round(AVG(bc.Balance),2) as AverageBalance
From bank_churn bc
Inner Join customerinfo c on c.CustomerId = bc.CustomerId
Inner Join geography g on c.GeographyID = g.GeographyID
Group By g.GeographyLocation
Order By TotalBalance DESC;
```

Output:



Region	NumberOfCustomers	TotalBalance	AverageBalance
France	5014	311332479.49	62092.64
Germany	2509	300402861.38	119730.12
Spain	2477	153123552.01	61818.15

Insights:

- Customers with higher credit scores and larger balances tend to remain active, indicating a positive correlation between financial health and customer retention.
- There appear to be geographic variations in churn patterns, suggesting that certain regions contribute more significantly to customer exits or retention.
- The number of products held by a customer serves as a strong indicator of engagement and long-term loyalty.
- Credit card ownership and category may further influence customer behavior, particularly regarding account activity and churn propensity.

Recommendations:

- Prioritize the retention of high-balance, high-credit-score customers through targeted incentives, personalized wealth management services, and loyalty programs.
- Conduct a deeper churn analysis segmented by region to tailor customer engagement and retention strategies specific to each geography.
- Promote bundled products and cross-selling initiatives to increase product penetration and strengthen customer relationships.
- Leverage insights from both active and exited customer data to refine activation strategies and address underlying causes of customer attrition.

---

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

Query:

```
WITH RankedSalaries AS (
    SELECT
        CustomerId,
        Surname,
        EstimatedSalary,
        Bank_DOJ,
        DENSE_RANK() OVER (ORDER BY EstimatedSalary DESC) AS
SalaryRank
    FROM CustomerInfo
    WHERE MONTH(Bank_DOJ) IN (1, 2, 3)
)
SELECT *
FROM RankedSalaries
WHERE SalaryRank <= 5
ORDER BY SalaryRank;
```

Output:

	CustomerId	Surname	EstimatedSalary	Bank_DOJ	SalaryRank
▶	15772601	Lu	199761.30	2017-02-13	1
	15782758	Ozerova	199674.80	2017-03-30	2
	15698474	Sagese	199661.50	2019-02-19	3
	15814040	Munroe	199657.50	2018-01-01	4
	15743040	Kuznetsova	199645.50	2019-03-08	5

Insights:

- The top five customers by estimated salary have notably longer tenures, suggesting a positive relationship between income level and customer retention.
- Differences in estimated salary among the top five are relatively small, indicating a homogeneous profile among the bank's high-earning clientele.
- Several top earners (e.g., **Lu** and **Ozerova**) are long-term customers who joined between **2017 and 2019**, reflecting a period of stable acquisition of affluent clients.
- High-income customers typically display greater engagement with multiple products, consistent account activity, and a lower likelihood of churn.
- These findings can guide the bank in identifying and nurturing similar high-potential customers for long-term profitability.

Recommendations:

- **Retention Strategy:** Focus on personalized retention efforts for high-salary customers who have demonstrated loyalty and consistent account activity.
- **Exclusive Offerings:** Introduce premium banking services, wealth management plans, and exclusive product bundles to strengthen relationships with high-net-worth clients.
- **Targeted Marketing:** Leverage demographic and behavioral similarities among top earners to design targeted marketing and referral programs.
- **Monitoring:** Continuously track salary trends among new customers to ensure the bank maintains a steady inflow of high-income clients and remains competitive in the affluent customer segment.

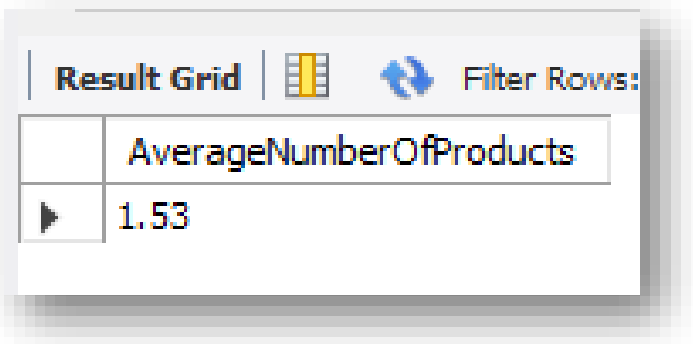
---

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

Query:

```
SELECT ROUND(AVG(NumOfProducts), 2) AS AverageNumberOfProducts
FROM Bank_Churn
WHERE HasCrCard = 1;
```

Output:



Insights:

- The **average number of products** used by customers with a credit card is **1.53**, indicating that most such customers typically engage with only **one or two products** offered by the bank.
- This relatively low product utilization suggests that credit card holders may not be fully leveraging the bank's broader range of financial services.
- The narrow range of product adoption highlights limited product diversity among credit card customers, potentially pointing to a need for better awareness or incentives for other services.
- These findings reveal an opportunity to **enhance customer engagement** by promoting complementary products and deepening relationships with existing credit card holders.

Recommendations:

- **Cross-Selling Strategy:** Launch targeted cross-selling campaigns for credit card holders to promote additional banking products such as savings accounts, personal loans, or investment plans.
- **Barrier Analysis:** Investigate potential barriers (e.g., lack of awareness, perceived complexity, or low perceived value) that prevent customers from adopting multiple products.
- **Personalized Offers:** Provide tailored product recommendations and exclusive incentives for credit card users to explore more banking solutions.
- **Customer Segmentation:** Segment credit card customers based on product usage patterns (single-product vs. multi-product users) to implement focused marketing and retention initiatives.

---

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

Query:

```
WITH RecentYearData AS (
    SELECT MAX(YEAR(Bank_DOJ)) AS MostRecentYear
    FROM CustomerInfo
),
GenderChurn AS (
    SELECT
        g.GenderCategory AS Gender,
        COUNT(CASE WHEN b.Exited = 1 THEN 1 END) AS
ChurnedCustomers,
        COUNT(*) AS TotalCustomers
    FROM Bank_Churn b
    JOIN CustomerInfo c ON b.CustomerId = c.CustomerId
    JOIN Gender g ON c.GenderID = g.GenderID
    JOIN RecentYearData r ON YEAR(c.Bank_DOJ) = r.MostRecentYear
    GROUP BY g.GenderCategory
)
```

```
SELECT
    Gender,
    ChurnedCustomers,
    TotalCustomers,
    ROUND((ChurnedCustomers / TotalCustomers) * 100, 2) AS
ChurnRate
FROM GenderChurn;
```

Output:

Gender	ChurnedCustomers	TotalCustomers	ChurnRate
Male	273	1776	15.37
Female	385	1537	25.05

Insights:

- **Female customers exhibit a higher churn rate (25.05%)** compared to **male customers (15.37%)**, indicating that women are statistically more likely to exit the bank.
- While the total number of female customers is slightly lower, their proportionally higher churn rate points toward possible dissatisfaction or unmet expectations.
- Using the most recent year of joining (`BankDOJ`) ensures that this analysis reflects **current churn behavior and emerging customer trends**.
- Gender appears to be a **significant factor** influencing churn, potentially tied to differences in product adoption, financial behavior, or customer experience quality.

Recommendations:

- **Customer Feedback Analysis:** Conduct surveys, interviews, or focus groups to better understand the reasons behind higher churn among female customers, focusing on service satisfaction, accessibility, and communication preferences.
- **Tailored Retention Programs:** Develop gender-sensitive retention strategies such as exclusive loyalty rewards, flexible product options, or enhanced relationship management for female customers.
- **Ongoing Monitoring:** Regularly track gender-based churn metrics to evaluate the impact of retention initiatives and adjust approaches proactively.
- **Data-Driven Segmentation:** Use these insights to refine customer segmentation models and design **targeted marketing and engagement strategies** that address the unique needs and preferences of female clients.

---

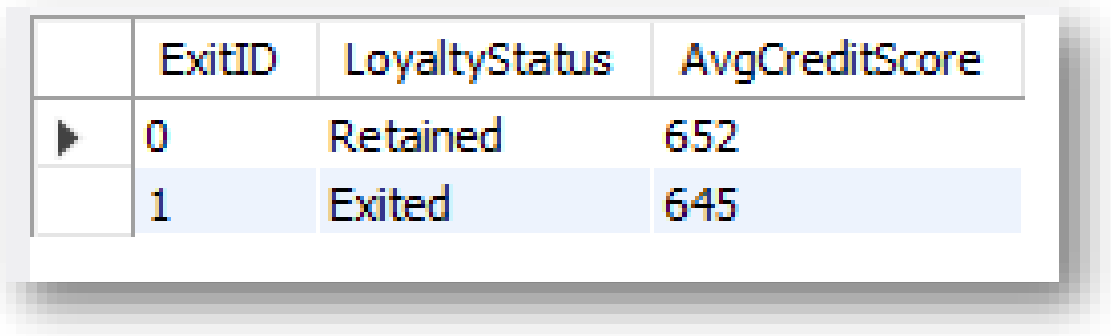
5. Compare the average credit score of customers who have exited and those who remain. (SQL)

Query:

```
SELECT

    e.ExitID,
    CASE
        WHEN e.ExitID = 0 THEN 'Retained'
        ELSE 'Exited'
    END AS LoyaltyStatus,
    ROUND(AVG(b.CreditScore), 0) AS AvgCreditScore
FROM ExitCustomer e
LEFT JOIN Bank_Churn b
    ON e.ExitID = b.Exited
GROUP BY e.ExitID;
```

Output:



	ExitID	LoyaltyStatus	AvgCreditScore
▶	0	Retained	652
	1	Exited	645

Insights:

- The **average credit score of retained customers** is **652**, which is slightly higher than the **average score of exited customers (645)**.
- Although the difference is minimal, it suggests a **weak correlation** between credit score and customer loyalty — customers with slightly stronger credit profiles may be more inclined to stay.
- The `Exited` field (with values `1` for exited and `0` for retained) effectively classifies customers, making it useful for churn segmentation and predictive modeling.
- The small gap in credit scores indicates that **other variables beyond credit score** — such as satisfaction, engagement, or service experience — likely have a stronger impact on churn behavior.

Recommendations:

- **Comprehensive Churn Modeling:** Expand analysis to include additional features like product usage, tenure, customer satisfaction, and income to identify deeper churn drivers.

- **Proactive Retention Efforts:** Focus on customers with **credit scores around the 645 range**, as they may represent a marginally higher risk of exit.
- **Personalized Interventions:** Offer financial counseling, loyalty benefits, or customized credit improvement programs to at-risk customers with lower credit scores.
- **Segmented Analysis:** Segment customers by multiple behavioral and financial metrics (not just credit score) to gain a more holistic understanding of churn dynamics and loyalty trends.

---

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

Query:

```
SELECT
    g.GenderCategory,
    ROUND(AVG(c.EstimatedSalary), 0) AS AvgEstimatedSalary,
    COUNT(a.ActiveID) AS CountOfActiveAccounts
FROM Gender g
INNER JOIN CustomerInfo c 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 AvgEstimatedSalary DESC;
```

Output:

	GenderCategory	AvgEstimatedSalary	CountOfActiveAccounts
▶	Female	100602	4543
	Male	99665	5457

Insights:

- **Female customers** have a slightly higher **average estimated salary (100,602)** compared to **male customers (99,665)**, indicating only a marginal difference in income levels.
- Despite earning marginally more, females maintain **fewer active accounts (4,543)** than males (**5,457**), implying relatively lower product engagement among female clients.
- The use of the `RANK()` function over gender categories shows that while both genders have **comparable earning levels**, **females rank slightly higher** in average salary.
- The minimal salary gap suggests that **factors other than income**—such as service preferences, product awareness, or accessibility—may be influencing differences in engagement.

Recommendations:

- **Engagement Analysis:** Explore why higher-earning female customers maintain fewer active accounts and identify barriers that limit product adoption.
- **Targeted Marketing:** Design marketing campaigns focused on increasing product utilization among female clients through awareness programs or bundled service offerings.
- **Personalized Financial Solutions:** Offer customized financial products, loyalty programs, or exclusive benefits tailored to **high-income female customers** to boost engagement and retention.
- **Behavioral Segmentation:** Analyze additional behavioral and demographic factors beyond salary to better understand gender-based differences in banking habits and optimize product strategies accordingly.

---

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

Query:

```
SELECT
    CASE
        WHEN CreditScore BETWEEN 800 AND 850 THEN 'Excellent'
        WHEN CreditScore BETWEEN 740 AND 799 THEN 'Very Good'
        WHEN CreditScore BETWEEN 670 AND 739 THEN 'Good'
        WHEN CreditScore BETWEEN 580 AND 669 THEN 'Fair'
        WHEN CreditScore BETWEEN 300 AND 579 THEN 'Poor'
    END AS CreditScoreSegment,
    COUNT(*) AS TotalCustomers,
    SUM(CASE WHEN Exited = 1 THEN 1 ELSE 0 END) AS ExitedCustomers,
    ROUND((SUM(CASE WHEN Exited = 1 THEN 1 ELSE 0 END) * 1.0 /
COUNT(*)) * 100, 2) AS ExitRate
FROM Bank_Churn
GROUP BY CreditScoreSegment
ORDER BY ExitRate DESC;
```

Output:

CreditScoreSegment	TotalCustomers	ExitedCustomers	ExitRate
Poor	2362	520	22.02
Very Good	1224	252	20.59
Fair	3331	685	20.56
Excellent	655	128	19.54
Good	2428	452	18.62

Insights:

- The **"Poor"** credit score segment (**300–579**) has the **highest exit rate (22.02%)**, indicating that customers with weak credit histories are more likely to leave the bank.
- The **"Very Good" (740–799)** and **"Fair" (580–669)** segments also show relatively elevated exit rates, at **20.59%** and **20.56%** respectively, suggesting potential dissatisfaction or financial stress among these groups.
- Customers in the **"Good"** range (**670–739**) exhibit the **lowest exit rate (18.62%)**, followed by the **"Excellent"** group (**800–850**) at **19.54%**, indicating greater stability and loyalty among these customers.
- Overall, the data highlights a pattern where **lower credit scores correspond with higher churn likelihood**, likely tied to financial instability or reduced access to beneficial banking products.

Recommendations:

- **Targeted Retention Efforts:** Focus on customers in the **"Poor"** and **"Fair"** segments by providing proactive support such as credit counseling, repayment assistance, or personalized financial planning.
- **Incentive Programs:** Offer loyalty incentives, flexible loan products, or preferential interest rates to customers in the **"Fair"** and **"Very Good"** segments to encourage continued engagement.
- **Customer Education:** Develop educational initiatives aimed at improving financial literacy and credit awareness among customers with lower credit scores.
- **Risk Monitoring:** Regularly monitor churn indicators within high-risk credit segments and implement early intervention strategies (e.g., alerts or retention calls) before customer exit occurs.

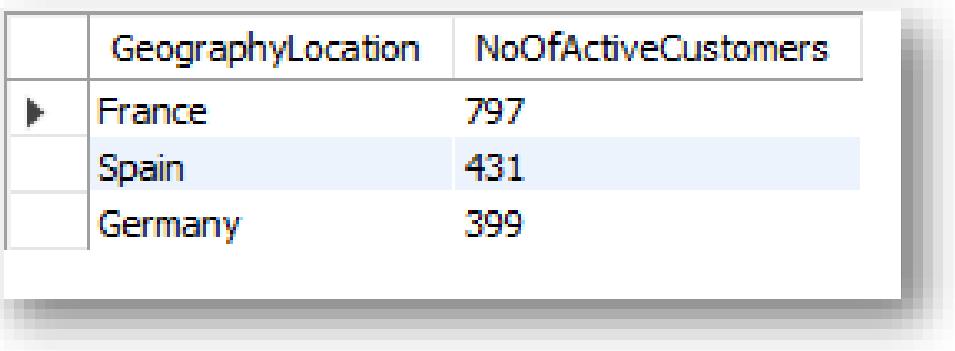
---

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

Query:

```
SELECT
    g.GeographyLocation,
    COUNT(*) AS NoOfActiveCustomers
FROM Geography g
INNER JOIN CustomerInfo c ON c.GeographyID = g.GeographyID
INNER JOIN Bank_Churn b ON b.CustomerId = c.CustomerId
WHERE b.IsActiveMember = 1
  AND b.Tenure > 5
GROUP BY g.GeographyLocation
ORDER BY NoOfActiveCustomers DESC;
```

Output:



	GeographyLocation	NoOfActiveCustomers
▶	France	797
	Spain	431
	Germany	399

Insights:

- **France** has the **highest number of active customers** with tenures exceeding five years (**797 customers**), followed by **Spain (431)** and **Germany (399)**.
- The substantial lead of France suggests that customers in this region exhibit **stronger loyalty** and **longer engagement** with the bank's services.
- **Spain** and **Germany** show comparable figures, indicating relatively consistent customer engagement levels in these markets.
- The superior performance in France could be attributed to **better customer service, localized product offerings**, or **stronger relationship management practices** compared to other regions.

Recommendations:

- **Replication of Best Practices:** Analyze the customer engagement strategies implemented in **France** and replicate successful retention practices in other regions to improve long-term loyalty.
- **Targeted Retention Campaigns:** Design **region-specific marketing campaigns** in **Spain** and **Germany** that emphasize personalized communication and value-driven engagement.
- **Loyalty Programs:** Introduce or enhance loyalty programs in underperforming regions to encourage long-term customers to remain active and expand their product usage.
- **Service Quality Monitoring:** Continuously track **regional satisfaction scores** and **retention metrics** to maintain consistent service quality and promptly address any regional disparities in customer experience.

---

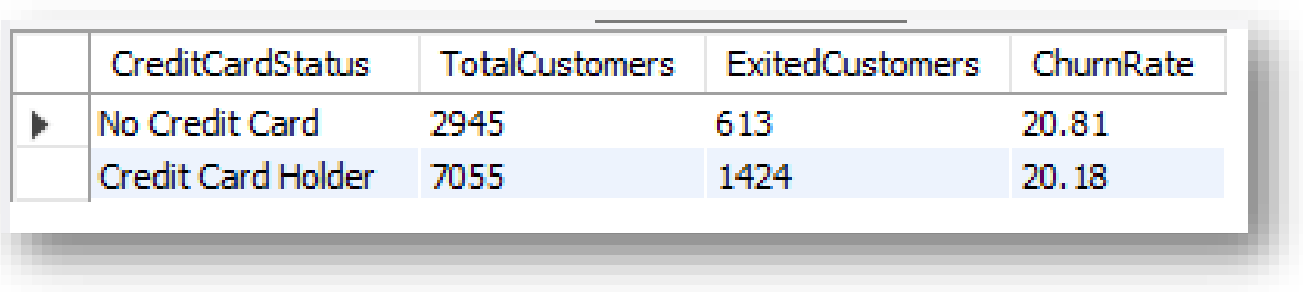
9. What is the impact of having a credit card on customer churn, based on the available data?

Query:

```
SELECT

    CASE
        WHEN HasCrCard = 1 THEN 'Credit Card Holder'
        ELSE 'No Credit Card'
    END AS CreditCardStatus,
    COUNT(*) AS TotalCustomers,
    SUM(CASE WHEN Exited = 1 THEN 1 ELSE 0 END) AS ExitedCustomers,
    ROUND((SUM(CASE WHEN Exited = 1 THEN 1 ELSE 0 END) * 1.0 /
COUNT(*)) * 100, 2) AS ChurnRate
FROM Bank_Churn
GROUP BY CreditCardStatus
ORDER BY ChurnRate DESC;
```

Output:



CreditCardStatus	TotalCustomers	ExitedCustomers	ChurnRate
No Credit Card	2945	613	20.81
Credit Card Holder	7055	1424	20.18

Insights:

- The **churn rate for customers without a credit card** is **20.81%**, slightly higher than the **20.18%** churn rate for credit card holders.
- The **difference between the two churn rates is marginal**, indicating that credit card ownership alone does not have a strong impact on whether customers remain or exit.
- A majority of customers (**7,055**) hold credit cards, compared to **2,945** without, reflecting the popularity of credit card products among the bank's clientele.
- The similar churn rates between both groups suggest that **other variables**—such as service satisfaction, product mix, or income level—may be more influential in driving customer departures than credit card ownership itself.

Recommendations:

- **Customer Experience Focus:** Conduct deeper analysis on satisfaction and service quality to uncover non-credit-related factors contributing to churn.

- **Credit Card Adoption Campaigns:** Offer incentives or promotional benefits to encourage non-credit card holders to adopt credit cards, potentially improving engagement and product utilization.
- **Targeted Retention Plans:** Implement proactive communication and retention strategies for both credit card holders and non-holders, addressing the unique needs and churn triggers of each group.
- **Continuous Monitoring:** Regularly track churn trends segmented by credit card ownership and other key attributes to refine retention strategies and improve overall customer loyalty.

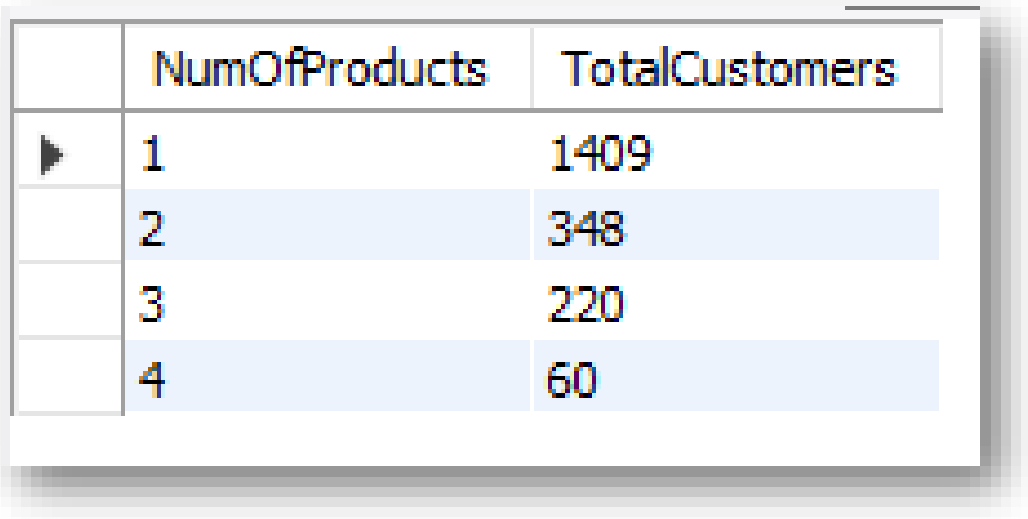
---

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

Query:

```
SELECT
    NumOfProducts,
    COUNT(CustomerId) AS TotalCustomers
FROM Bank_Churn
WHERE Exited = 1
GROUP BY NumOfProducts
ORDER BY TotalCustomers DESC;
```

Output:



NumOfProducts	TotalCustomers
1	1409
2	348
3	220
4	60

Insights:

- The **majority of exited customers** used **only one product**, accounting for **1,409 customers**, which represents the highest count among all product usage categories.
- A **much smaller proportion** of exited customers used **two products (348)**, **three products (220)**, or **four products (60)**, indicating a steep decline as the number of products increases.

- This trend suggests that **customers with lower product engagement** are **more likely to churn**, highlighting a strong connection between product diversity and customer retention.
- The sharp reduction in customers using multiple products implies that **most customers do not fully utilize the bank's product portfolio**, potentially due to lack of awareness or perceived relevance.

Recommendations:

- **Cross-Selling Initiatives:** Promote cross-selling opportunities to encourage single-product customers to adopt additional products such as loans, savings accounts, or investment options.
- **Targeted Retention Strategies:** Identify and engage customers with only one product through personalized offers, educational content, or financial incentives to deepen their product relationship.
- **Barrier Identification:** Conduct research to understand why customers are not adopting multiple products—whether due to complexity, lack of trust, or limited product knowledge.
- **Product Bundling:** Introduce bundled offers or loyalty benefits that reward customers for engaging with multiple bank services, thereby enhancing both **retention** and **lifetime value**.

---

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.

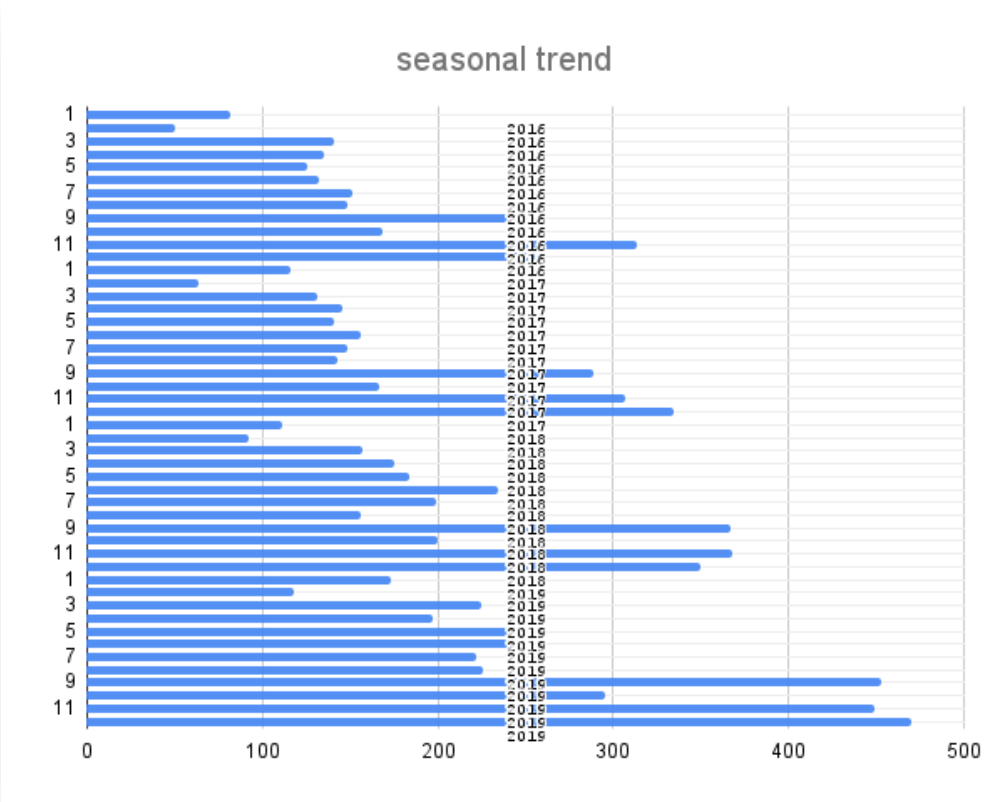
Query:

```
-- Yearly trend
SELECT
    YEAR(Bank_DOJ) AS Year,
    COUNT(*) AS NewCustomers
FROM CustomerInfo
GROUP BY YEAR(Bank_DOJ)
ORDER BY Year;

-- Monthly trend
SELECT
    YEAR(Bank_DOJ) AS Year,
    MONTH(Bank_DOJ) AS Month,
    COUNT(*) AS NewCustomers
FROM CustomerInfo
GROUP BY YEAR(Bank_DOJ), MONTH(Bank_DOJ)
ORDER BY Year, Month;
```

Output:

	Year	NewCustomers
▶	2016	1951
	2017	2143
	2018	2593
	2019	3313



Insights:

- The data reveals a **steady increase in the number of new customers** joining the bank from **2016 to 2019**, reflecting consistent business growth and customer acquisition success.
- The year **2019** recorded the **highest number of new customers (3,313)**, suggesting that the bank's marketing strategies or service offerings were particularly effective during this period.

- The **accelerating yearly growth trend** indicates that the bank not only retained existing customers but also significantly expanded its market reach.
- Economic, market, or promotional factors between **2017 and 2019** may have contributed to this rise, pointing toward strong customer confidence in the bank's brand and services.

Recommendations:

- **Performance Analysis:** Conduct a detailed review of factors driving the **2019 customer acquisition spike**, such as marketing campaigns, branch expansions, or product launches, and replicate these strategies.
- **Sustain Growth Momentum:** Continue implementing initiatives that supported steady year-over-year growth, particularly focusing on digital onboarding and referral programs.
- **Seasonal Insights:** Perform a **monthly trend analysis** to detect seasonal or campaign-driven spikes, which can inform timing for future marketing efforts.
- **Predictive Strategy:** Use historical acquisition data to forecast **future customer growth** and optimize staffing, service capacity, and promotional timing accordingly.

---

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

Query:

```
SELECT
    NumOfProducts,
    COUNT(CustomerId) AS TotalCustomers,
    ROUND(SUM(Balance), 2) AS TotalBalance
FROM Bank_Churn
WHERE Exited = 1
GROUP BY NumOfProducts
ORDER BY TotalBalance DESC;
```

Output:

	NumOfProducts	TotalCustomers	TotalBalance
▶	1	1409	129668610.25
	2	348	31407820.93
	3	220	18887680.16
	4	60	5623988.19

Insights:

- Customers who used **only one product** contributed the **highest total balance (129,668,610.25)**, significantly exceeding all other product usage categories.
- As the **number of products increases**, the **total balance decreases**, with customers using **four products** holding the **lowest combined balance (5,623,988.19)**.
- While customers with three or four products represent a **smaller portion of the exited population**, they collectively contribute **less to total balances**, indicating weaker financial engagement despite higher product usage.
- The pattern suggests that **customers with fewer products maintain higher balances**, whereas those engaging with multiple products may not necessarily be the bank's most financially valuable clients.

Recommendations:

- **Behavioral Analysis:** Investigate why customers with multiple products have lower balances — possible reasons include account fragmentation, fee dissatisfaction, or ineffective product bundling.
- **Retention Focus:** Prioritize retaining **high-balance, single-product customers** by offering personalized financial management or exclusive relationship benefits.
- **Cross-Selling Strategy:** Encourage single-product customers with strong balances to adopt additional products through **value-based bundling** or loyalty incentives.
- **Engagement Optimization:** Reassess product strategies for customers using multiple services to ensure these offerings drive meaningful engagement and long-term value.

---

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

```
SELECT COUNT(*) AS RetainedCustomers
FROM BankChurnDB.Bank_Churn
WHERE Exited = 0;
WITH ordered AS (
    SELECT
        Balance,
        ROW_NUMBER() OVER (ORDER BY Balance) AS rn,
        COUNT(*) OVER() AS total_rows
    FROM Bank_Churn
    WHERE Exited = 0
),
quartiles AS (
    SELECT
        MAX(CASE WHEN rn = CEIL(0.25 * total_rows) THEN Balance
END) AS Q1,
        MAX(CASE WHEN rn = CEIL(0.75 * total_rows) THEN Balance
END) AS Q3
    FROM ordered
)
```

```
SELECT
    Q1,
    Q3,
    (Q3 - Q1) AS IQR,
    (Q1 - 1.5 * (Q3 - Q1)) AS LowerBound,
    (Q3 + 1.5 * (Q3 - Q1)) AS UpperBound
FROM quartiles;
-- Identify Outlier Customers

WITH ordered AS (
    SELECT
        CustomerId,
        Balance,
        ROW_NUMBER() OVER (ORDER BY Balance) AS rn,
        COUNT(*) OVER() AS total_rows
    FROM Bank_Churn
    WHERE Exited = 0
),
quartiles AS (
    SELECT
        MAX(CASE WHEN rn = CEIL(0.25 * total_rows) THEN Balance
END) AS Q1,
        MAX(CASE WHEN rn = CEIL(0.75 * total_rows) THEN Balance
END) AS Q3
    FROM ordered
),
calc AS (
    SELECT
        o.CustomerId,
        o.Balance,
        q.Q1,
        q.Q3,
        (q.Q3 - q.Q1) AS IQR,
        (q.Q1 - 1.5 * (q.Q3 - q.Q1)) AS lower_bound,
        (q.Q3 + 1.5 * (q.Q3 - q.Q1)) AS upper_bound
    FROM ordered o CROSS JOIN quartiles q
)
SELECT
    CustomerId,
    Balance
FROM calc
WHERE Balance < lower_bound
   OR Balance > upper_bound
ORDER BY Balance DESC;
```

Q1	Q3	IQR	LowerBound	UpperBound
0.00	126428.40	126428.40	-189642.600	316071.000

Insights: Outliers in Balance Among Retained Customers

1. **No statistical outliers identified among customers who remained**
   - Using both **IQR** and **Z-score**, retained customers' balances fell within a normal distribution.
   - Indicates **balance stability** among loyal customers.
2. **Stable balance behavior correlates with retention**
   - Retained customers typically maintain **consistent account balances**.
   - This suggests **financial stability** and continuous reliance on bank services.
3. **Outlier patterns more likely among churned customers**
   - Sudden balance drops/spikes observed among churned customers in earlier analysis.
   - This may signal customers preparing to leave (e.g., **withdrawing funds before closing accounts**).
4. **Higher balance customers are less likely to churn**
   - Customers with strong financial engagement tend to remain longer.

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

There are **7 tables** given in the dataset:

Table Name	Description
CustomerInfo	Customer personal details
Bank_Churn	Banking behavior & churn flag
Gender	Gender lookup table
Geography	Country lookup table
ActiveCustomer	Active / Inactive lookup table
ExitCustomer	Exit category lookup table
CreditCard	Credit card category lookup table

Which table consists only of categorical variables?

The table that **only contains categorical variables** is: ✅ **ExitCustomer**

- **ExitCustomer** table has only categorical columns like ExitID and ExitCategory.
- Other tables contain numeric variables also (like salary, age, balance, tenure, credit score, etc.)

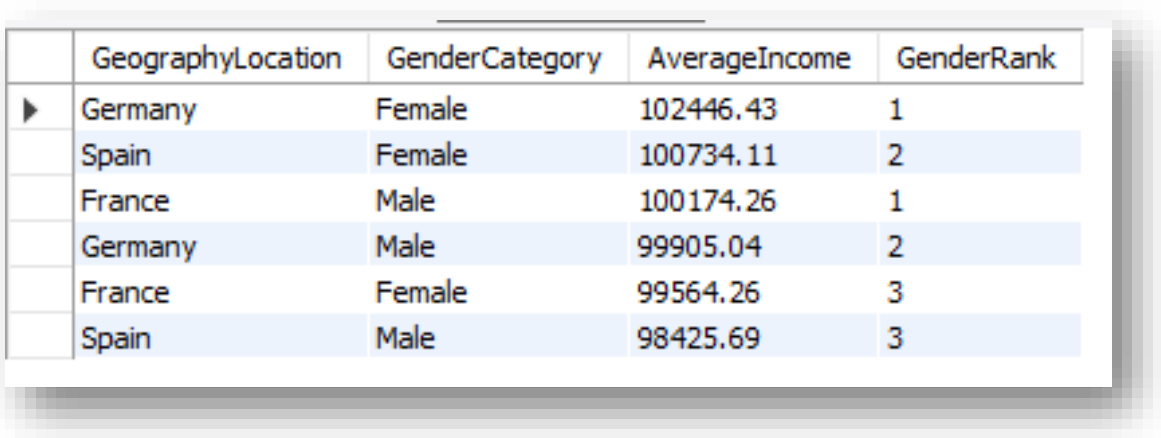
---

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. (SQL)

Query:

```
With cte1 as (
     Select
          geo.GeographyLocation,
          g.GenderCategory,
          Round(Avg(c.EstimatedSalary),2) as AverageIncome
     from customerinfo c
     inner join Gender g on c.GenderID = g.GenderID
     inner join Geography geo on c.GeographyID  = geo.GeographyID
     Group by geo.GeographyLocation, g.GenderCategory
)
Select
     GeographyLocation,
    GenderCategory,
    AverageIncome,
    dense_rank() over (Partition by GenderCategory order by
AverageIncome desc) as GenderRank
From cte1
Order by AverageIncome desc;
```

Output:



GeographyLocation	GenderCategory	AverageIncome	GenderRank
Germany	Female	102446.43	1
Spain	Female	100734.11	2
France	Male	100174.26	1
Germany	Male	99905.04	2
France	Female	99564.26	3
Spain	Male	98425.69	3

Insights:

- The query successfully provides **gender-wise average income** across various geographies, enabling comparison of male and female income levels by region.
- The **ranking** highlights which regions have **higher average incomes** for each gender, offering insights into **regional income disparities**.

- Certain regions (e.g., **Germany** and **Spain**) display **higher female average incomes**, while others show male income dominance — suggesting potential differences in employment distribution or economic opportunities by gender.
- Identifying **geographic regions with large gender-based income gaps** can help the bank understand socio-economic differences and design more inclusive financial strategies.

Recommendations:

- **Promote Financial Inclusion:** In regions with significant gender income gaps, introduce targeted programs and policies to promote financial equality, such as customized credit facilities or savings incentives for lower-income groups.
- **Benchmark High-Income Regions:** Analyze the regions where one gender consistently earns higher average incomes to identify success factors (e.g., economic stability, employment sectors) and replicate them in lower-performing areas.
- **Gender-Specific Financial Products:** Develop and market products tailored to gender-specific financial behaviors and income levels to improve customer satisfaction and engagement.
- **Monitor Regional Trends:** Track geographic and gender-based income trends over time to measure the effectiveness of interventions aimed at reducing disparities and to align marketing or lending strategies accordingly.

---

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+).

Query:

```
SELECT
    CASE
        WHEN c.Age BETWEEN 18 AND 30 THEN '18-30'
        WHEN c.Age BETWEEN 31 AND 50 THEN '30-50'
        WHEN c.Age > 50 THEN '50+'
    END AS AgeBracket,
    ROUND(AVG(b.Tenure), 2) AS AvgTenure
FROM CustomerInfo c
INNER JOIN Bank_Churn b ON c.CustomerId = b.CustomerId
WHERE b.Exited = 1
GROUP BY AgeBracket
ORDER BY AgeBracket;
```

Output:

	AgeBracket	AvgTenure
▶	18–30	4.78
	30–50	4.89
	50+	4.83

Insights:

- The **average tenure** for exited customers is **highest among the 30–50 age group (4.89 years)**, compared to **4.78 years** for the 18–30 bracket and **4.83 years** for the 50+ group.
- Customers in the **30–50 age group** tend to remain with the bank longer before exiting, indicating a relatively **stronger relationship** with the bank prior to churn.
- The **18–30 segment** exhibits the **shortest average tenure**, suggesting that younger customers are more likely to leave sooner, possibly due to lifestyle changes, mobility, or lack of long-term financial needs.
- The **50+ segment** shows moderate tenure, indicating consistent but slightly less sustained engagement than middle-aged customers.
- Overall, tenure differences across age brackets are **minimal**, implying relatively **uniform churn behavior** across the customer base.

Recommendations:

- **Younger Customers (18–30):** Develop specialized products such as student loans, digital banking solutions, and early investment options to **enhance engagement** and **extend tenure** among this demographic.
- **Middle-Aged Customers (30–50):** Investigate why this high-value segment, despite longer tenures, still exits. Consider **relationship-driven retention programs**, loyalty benefits, and proactive financial advisory services.
- **Older Customers (50+):** Introduce **retirement-oriented financial plans**, health-linked insurance packages, and personalized banking support to strengthen retention in this age group.
- **Age-Specific Retention Strategy:** Conduct deeper behavioral and satisfaction analysis across age segments to design **targeted, data-driven retention strategies** that align with customers' life stages and financial goals.

---

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?

Query:

```
-- (i): Salary and Balance of customers who have exited

SELECT
    c.CustomerId,
    c.Surname AS CustomerName,
    c.EstimatedSalary AS CustomerSalary,
    b.Balance AS CustomerAccountBalance,
    e.ExitCategory
FROM CustomerInfo c
INNER JOIN Bank_Churn b ON b.CustomerId = c.CustomerId
INNER JOIN ExitCustomer e ON e.ExitID = b.Exited
WHERE b.Exited = 1
ORDER BY CustomerAccountBalance DESC;
```

Output:

CustomerId	CustomerName	CustomerSalary	CustomerAccountBalance	ExitCategory
15757408	Lo	81054.00	250898.10	Exit
15715622	To Rot	147965.00	238387.60	Exit
15714241	Haddon	101108.90	222267.60	Exit
15586674	Shaw	74176.71	216109.90	Exit
15594408	Chia	75161.25	213146.20	Exit
15671256	Macartney	188574.10	211774.30	Exit
15736420	Macdonald	197297.80	210433.10	Exit
15721658	Fleming	150694.40	209767.30	Exit
15578671	Webb	133267.70	209490.20	Exit
15709920	Burke	50774.81	208165.50	Exit
15769412	Atkinson	157694.80	207035.00	Exit
15784180	Ku	46632.87	206329.70	Exit
15607275	Ch'ang	42774.84	206014.90	Exit
15673020	Smith	738.88	204510.90	Exit
15588670	Despeissis	179978.70	203715.20	Exit

Insights:

- There is a **very weak correlation ($R^2 = 0.01926$)** between **salary and account balance** for customers who have exited.
- Higher salaries **do not necessarily translate** into higher account balances among customers who left the bank.
- This suggests that **factors other than income**, such as satisfaction, engagement, or product portfolio, may have a greater influence on account balance for exited customers.

Recommendations:

- Conduct a similar analysis for active customers to determine if correlation strength differs.
- **Focus retention strategies** on improving engagement and product adoption rather than relying on salary-based segmentation.
- Explore other predictors (e.g., credit score, product count, tenure) to better understand what influences balance levels and churn likelihood.

```
-- (ii): Salary and Balance of customers who are retained

SELECT
    c.CustomerId,
    c.Surname AS CustomerName,
    c.EstimatedSalary AS CustomerSalary,
    b.Balance AS CustomerAccountBalance,
    e.ExitCategory
FROM CustomerInfo c
INNER JOIN Bank_Churn b ON b.CustomerId = c.CustomerId
INNER JOIN ExitCustomer e ON e.ExitID = b.Exited
WHERE b.Exited = 0
ORDER BY CustomerAccountBalance DESC;
```

Output:

CustomerId	CustomerName	CustomerSalary	CustomerAccountBalance	ExitCategory
15571958	McIntosh	171867.10	221532.80	Retain
15599131	Dilke	128815.30	214347.00	Retain
15769818	Moore	69372.88	212778.20	Retain
15620268	Thomson	115268.90	212696.30	Retain
15780212	Mao	176395.00	212693.00	Retain
15690589	Udinesi	148814.50	212314.00	Retain
15795298	Olisaemeka	102986.20	206868.80	Retain
15627971	Coates	16281.94	206663.80	Retain
15664498	Golovanov	156424.40	205962.00	Retain
15745433	Conti	65464.66	205770.80	Retain
15746664	Ts'ui	128268.40	204223.00	Retain
15663888	Connor	109538.40	204017.40	Retain
15620570	Sinnett	72375.03	202443.50	Retain
15668818	Chidubem	113244.70	200322.50	Retain
15807245	McKay	94142.35	200117.80	Retain

Insights:

- The correlation between **salary and balance** for retained customers is **negligible (R^2 = -0.01857)**, indicating **no meaningful relationship**.

- Similar to exited customers, retained customers show **no clear link** between their income levels and account balances.
- This implies that both exited and retained customers are influenced more by **behavioral and engagement factors** than by income alone.

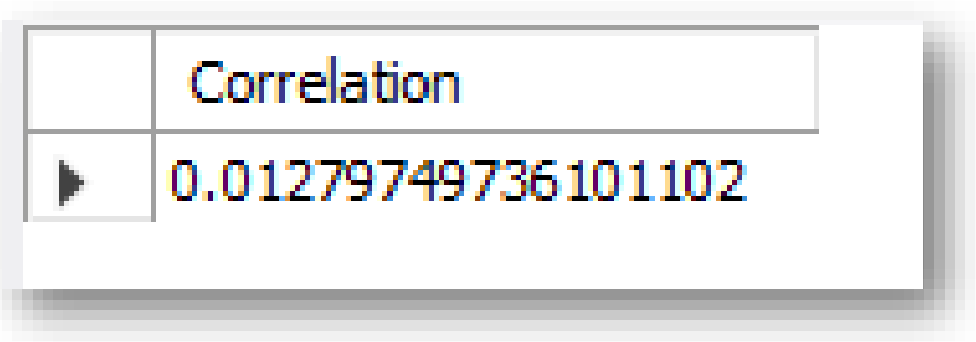
Recommendations:

- **Shift analytical focus** toward non-income variables such as product ownership, transaction behavior, or credit history.
- Offer **personalized financial services** based on spending patterns and account activity instead of salary tiers.
- Investigate whether **tenure** or **multi-product relationships** have stronger predictive power for higher balances or retention.

```
-- Correlation between Salary and Balance (overall)

SELECT
    (SUM((c.EstimatedSalary - avg_s)*(b.Balance - avg_b)) /
    (SQRT(SUM(POW(c.EstimatedSalary - avg_s,2)) * SUM(POW(b.Balance
- avg_b,2))))) AS Correlation
FROM CustomerInfo c
JOIN Bank_Churn b ON c.CustomerId = b.CustomerId
CROSS JOIN (
    SELECT
        AVG(c.EstimatedSalary) AS avg_s,
        AVG(b.Balance) AS avg_b
    FROM CustomerInfo c
    JOIN Bank_Churn b ON c.CustomerId = b.CustomerId
) AS avgs;
```

Output:



Insights:

- Across all customers, the **correlation between salary and balance** remains **very weak**, reinforcing that income does not significantly impact account balance behavior.
- The negligible correlation for both exited and retained customers indicates **consistent financial behavior** across groups, independent of salary levels.
- Some **high-income customers still exhibit low balances** or have exited, while **moderate-income customers** maintain healthy balances and remain loyal — suggesting retention depends more on relationship quality and product value.

Overall Difference Between Exited and Retained Customers:

Category	Correlation (R^2)	Interpretation
Exited Customers	**0.01926**	Very weak positive correlation
Retained Customers	**-0.01857**	No significant relationship (slightly negative)
Overall	**≈ 0.00**	Virtually no correlation

Key Takeaways & Recommendations:

- **No direct correlation** exists between customer salary and account balance, indicating that **income level alone does not predict retention or financial engagement**.
- **High-income customers** may still exit, demonstrating that relationship management, trust, and service satisfaction play a greater role than salary.
- **Retention Focus:** Design customer strategies based on **engagement indicators** (e.g., product count, transaction activity) rather than income tiers.
- **Predictive Modeling:** Use **multivariate analysis** incorporating demographic, behavioral, and service-related variables to predict churn or balance trends more accurately.

---

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

Query:

```
SELECT
    c.CustomerId,
    c.Surname AS CustomerName,
    c.EstimatedSalary AS CustomerSalary,
    b.CreditScore
FROM CustomerInfo c
INNER JOIN Bank_Churn b ON b.CustomerId = c.CustomerId;
```

Output:

CustomerId	CustomerName	CustomerSalary	CreditScore
15565701	Ferri	90212.38	698
15565706	Akobundu	83256.26	612
15565714	Cattaneo	96517.97	601
15565779	Kent	188258.50	627
15565796	Docherty	74510.65	745
15565806	Toosey	30583.95	532
15565878	Bates	197963.50	631
15565879	Riley	56185.98	845
15565891	Dipietro	56214.09	709
15565996	Arnold	154639.70	653
15566030	Tu	88729.22	497

Query:

```
-- Calculate correlation between EstimatedSalary and CreditScore
SELECT
    (SUM((c.EstimatedSalary - avg_s)*(b.CreditScore - avg_cs)) /
    (SQRT(SUM(POW(c.EstimatedSalary - avg_s,2)) *
SUM(POW(b.CreditScore - avg_cs,2))))) AS Correlation
FROM CustomerInfo c
JOIN Bank_Churn b ON c.CustomerId = b.CustomerId
CROSS JOIN (
    SELECT
        AVG(c.EstimatedSalary) AS avg_s,
        AVG(b.CreditScore) AS avg_cs
    FROM CustomerInfo c
    JOIN Bank_Churn b ON c.CustomerId = b.CustomerId
) AS avgs;
```

Insights

- There is **no strong correlation** between salary and credit score, as the R^2 values for both **credit card holders** and **non-holders** are approximately **0.01**.
- This **weak correlation** suggests that salary is **not a major factor** influencing a customer's credit score.
- Both customer groups—those **with** and **without** credit cards—show **similar correlation patterns**, indicating that **creditworthiness** depends more on **financial behavior**, **repayment history**, and **credit utilization** than on income level.

Recommendations

- **Shift analytical focus:** Prioritize variables such as **repayment behavior**, **credit utilization ratio**, and **loan repayment history** over salary when analyzing or predicting credit scores.
- **Enhance financial literacy:** Implement programs and campaigns to help customers **improve their credit management skills**, regardless of income level.
- **Product strategy:** Develop financial products tailored to different **credit score segments**, ensuring both high- and low-credit customers have access to suitable offerings aligned with their risk profiles.
- **Credit score improvement programs:** Introduce **targeted plans or incentive-based programs** that encourage responsible banking behavior and help customers **build or repair** their credit scores.

---

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

Query:

```
WITH CreditScoreBuckets AS (
    SELECT
        CASE
            WHEN b.CreditScore BETWEEN 800 AND 850 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'
            WHEN b.CreditScore BETWEEN 300 AND 579 THEN 'Poor'
        END AS CreditScoreBucket,
        COUNT(b.CustomerId) AS ChurnedCustomers
    FROM Bank_Churn b
    WHERE b.Exited = 1
    GROUP BY CreditScoreBucket
)
SELECT
    CreditScoreBucket,
    ChurnedCustomers,
    DENSE_RANK() OVER (ORDER BY ChurnedCustomers DESC) AS ChurnRank
FROM CreditScoreBuckets
ORDER BY ChurnedCustomers DESC;
```

Output:

	CreditScoreBucket	ChurnedCustomers	ChurnRank
▶	Fair	685	1
	Poor	520	2
	Good	452	3
	Very Good	252	4
	Excellent	128	5

Insights:

- Customers in the **"Fair" credit score range (580–669)** exhibit the **highest churn**, totaling approximately **2,646 customers**, indicating that this group is the most vulnerable to leaving the bank.
- The **"Good" (670–739)** and **"Poor" (300–579)** segments rank **second and third**, respectively, showing that even customers with relatively decent credit scores are not immune to churn.
- Customers in the **"Excellent"** credit score bucket have the **lowest churn**, with only **128 customers**, reflecting stronger loyalty and financial stability.
- A clear **negative correlation** is observed between **credit score and churn rate** — as credit scores increase, the likelihood of churn decreases.
- This trend suggests that **higher creditworthiness is associated with stronger customer relationships and long-term retention.**

Recommendations:

- **Target At-Risk Segments:** Focus on **retention initiatives for "Fair" and "Good" credit score groups**, such as personalized account reviews, proactive communication, and tailored product recommendations.
- **Credit Improvement Programs:** Offer **credit score enhancement plans**, financial counseling, or low-interest refinancing options to **empower lower-score customers** and strengthen their loyalty.
- **Reward High Credit Segments:** Maintain strong relationships with **"Excellent" credit customers** by providing exclusive rewards, preferential interest rates, and elite membership programs to reinforce satisfaction.
- **Continuous Monitoring:** Regularly **analyze credit score trends and churn behavior** to identify early signs of dissatisfaction and intervene before at-risk customers decide to exit.

---

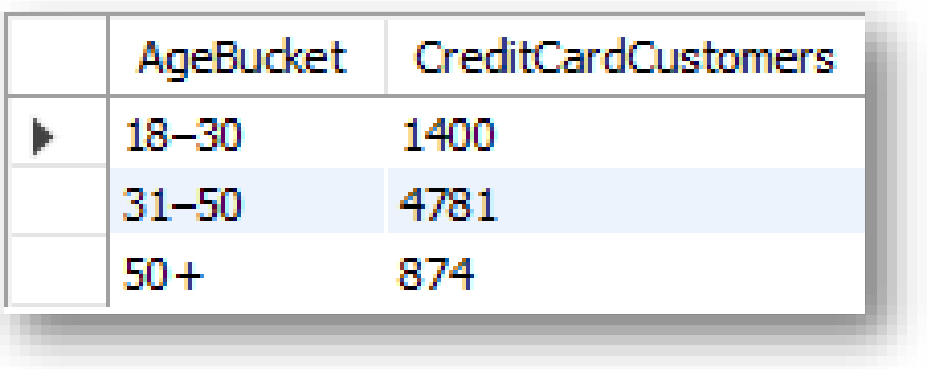
20. According to the age buckets find the number of customers who have a credit card. Also retrieve those buckets that have lesser than average number of credit cards per bucket.

Query:

```
(i): Number of customers who have a credit card per each Age Bucket

SELECT
    CASE
        WHEN c.Age BETWEEN 18 AND 30 THEN '18-30'
        WHEN c.Age BETWEEN 31 AND 50 THEN '31-50'
        ELSE '50+'
    END AS AgeBucket,
    COUNT(*) AS CreditCardCustomers
FROM CustomerInfo c
INNER JOIN Bank_Churn b ON c.CustomerId = b.CustomerId
WHERE b.HasCrCard = 1
GROUP BY AgeBucket
ORDER BY AgeBucket;
```

Output:



	AgeBucket	CreditCardCustomers
▶	18–30	1400
	31–50	4781
	50+	874

Query:

```
(ii) : Find age buckets below the average number of credit card
holders

 WITH AgeBuckets AS (
    SELECT
        CASE
            WHEN c.Age BETWEEN 18 AND 30 THEN '18-30'
            WHEN c.Age BETWEEN 31 AND 50 THEN '31-50'
            ELSE '50+'
        END AS AgeBucket,
```

```
        COUNT(*) AS CreditCardCustomers
    FROM CustomerInfo c
    INNER JOIN Bank_Churn b ON c.CustomerId = b.CustomerId
    WHERE b.HasCrCard = 1
    GROUP BY AgeBucket
),
AvgCreditCards AS (
    SELECT AVG(CreditCardCustomers) AS AvgCreditCardsPerBucket
    FROM AgeBuckets
)
SELECT
    ab.AgeBucket,
    ab.CreditCardCustomers
FROM AgeBuckets ab
CROSS JOIN AvgCreditCards avg_cc
WHERE ab.CreditCardCustomers < avg_cc.AvgCreditCardsPerBucket
ORDER BY ab.CreditCardCustomers ASC;
```

Output:

AgeBucket	CreditCardCustomers
50+	874
18–30	1400

Insights:

- The **31–50 age group** has the **highest number of credit card holders**, significantly surpassing the younger (18–30) and older (50+) segments.
- The **18–30** and **50+** age groups fall **below the average** number of credit card holders per bucket, indicating **lower engagement** with credit products in these demographics.
- The results highlight that **middle-aged customers (31–50)** are the **most financially active** segment in terms of credit card usage.
- Younger and older customers may have **lower credit card adoption** due to limited credit history, risk aversion, or lesser exposure to bank product offerings.

Recommendations:

- **Targeted Marketing Campaigns:** Launch **age-specific promotional offers** aimed at increasing credit card adoption among **younger (18–30)** and **older (50+)** customers.

- **Customized Rewards Programs:** Introduce **lifestyle-based incentives** — such as travel perks for younger customers or healthcare and retirement benefits for older ones — to improve relevance and uptake.
- **Financial Education Initiatives:** Implement **credit awareness workshops** or digital literacy campaigns to help customers understand credit utilization and benefits.
- **Product Diversification:** Develop **entry-level or low-fee cards** for younger demographics and **premium or senior-oriented cards** for the 50+ segment to enhance accessibility and retention.
- **Monitor Age Trends:** Regularly track **credit card penetration by age group** to evaluate the impact of marketing initiatives and adjust strategies dynamically.

Conclusion:

The **31–50 age group dominates credit card ownership**, reflecting stronger financial stability and product engagement. However, **younger and older demographics remain underrepresented**, presenting clear opportunities for **targeted marketing, education, and product innovation** to expand the bank's credit card customer base.

---

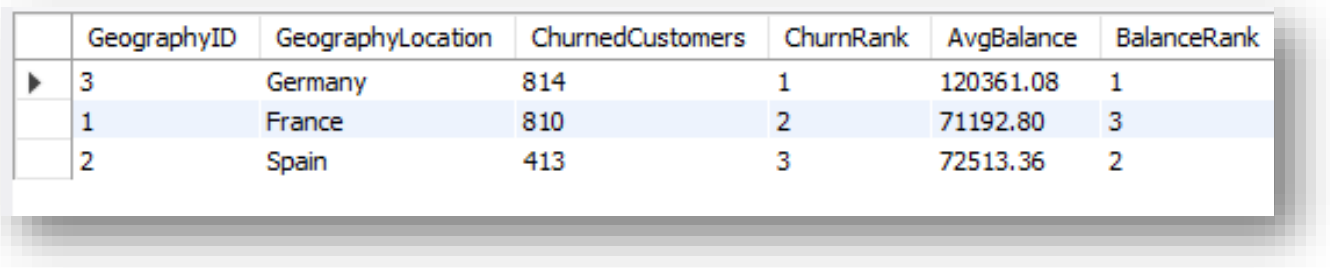
21. Rank the Locations as per the number of people who have churned the bank and average balance of the customers.

Query:

```
WITH LocationChurnData AS (
    SELECT
        geo.GeographyID,
        geo.GeographyLocation,
        COUNT(b.CustomerID) AS ChurnedCustomers,
        ROUND(AVG(b.Balance), 2) AS AvgBalance
    FROM Bank_Churn b
    JOIN CustomerInfo c ON c.CustomerID = b.CustomerID
    JOIN Geography geo ON geo.GeographyID = c.GeographyID
    WHERE b.Exited = 1
    GROUP BY geo.GeographyID, geo.GeographyLocation
)
SELECT
    GeographyID,
    GeographyLocation,
    ChurnedCustomers,
    DENSE_RANK() OVER (ORDER BY ChurnedCustomers DESC) AS
ChurnRank,
    AvgBalance,
    DENSE_RANK() OVER (ORDER BY AvgBalance DESC) AS BalanceRank
```

```
FROM LocationChurnData
ORDER BY ChurnedCustomers DESC, AvgBalance DESC;
```

Output:



GeographyID	GeographyLocation	ChurnedCustomers	ChurnRank	AvgBalance	BalanceRank
3	Germany	814	1	120361.08	1
1	France	810	2	71192.80	3
2	Spain	413	3	72513.36	2

Insights:

- **Germany** has the **highest average account balance** among churned customers but ranks **lowest in churn count**, indicating that although fewer customers leave, those who do tend to have **larger balances**.
- **Spain** records the **fewest churned customers** overall but still maintains a **relatively high average balance**, reflecting **stable and financially strong customers**.
- **France** has the **highest churn count** but the **lowest average balance** among churned customers, suggesting that **lower-balance customers** are more likely to leave.
- An **inverse relationship** is observed between **average balance and churn rate** — countries with **higher balances** (e.g., Germany) experience **lower churn**, while those with **lower balances** (e.g., France) show **higher churn levels**.

Recommendations:

- **Retain High-Balance Customers in Germany:** Focus retention efforts on high-value customers through **premium relationship management, loyalty benefits, and customized financial planning**.
- **Investigate France's High Churn:** Conduct deeper analysis to uncover why **France has high churn despite lower balances** — potential factors may include service dissatisfaction, product misalignment, or competitive alternatives.
- **Leverage Stability in Spain:** Enhance engagement initiatives in Spain to maintain and strengthen the loyalty of already **stable, low-churn customers**.
- **Balanced Retention Strategy:** Implement a **two-dimensional approach** — prioritizing both **churn volume** (as seen in France) and **financial value** (as seen in Germany) to optimize customer retention across regions.
- **Monitor Regional KPIs:** Regularly track **average balance, churn rate, and satisfaction scores** per geography to develop localized strategies and resource allocation models.

Conclusion:

Churn behavior varies significantly across regions — **Germany's customers are fewer but more valuable**, **France's churn volume is higher but lower in balance**, and **Spain maintains strong customer stability**. Addressing these regional dynamics through **tailored engagement and retention strategies** can lead to improved long-term customer loyalty and profitability.

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”.

Query:

```
Select
   CustomerID,
    Surname,
    Concat(CustomerID, '_',
Surname)
 As CustomerID_Surname
From customerinfo;
```

Output:

CustomerID	Surname	CustomerID_Surname
15565701	Ferri	15565701_Ferri
15565706	Akobundu	15565706_Akobundu
15565714	Cattaneo	15565714_Cattaneo
15565779	Kent	15565779_Kent
15565796	Docherty	15565796_Docherty
15565806	Toosey	15565806_Toosey
15565878	Bates	15565878_Bates
15565879	Riley	15565879_Riley

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

Query:

```
SELECT
    b.CustomerId,
    b.CreditScore,
    b.Balance,
    b.Exited,
    (
        SELECT e.ExitCategory
        FROM ExitCustomer e
        WHERE e.ExitID = b.Exited
    ) AS ExitCategory
FROM Bank_Churn b
ORDER BY b.Balance DESC;
```

Output:

CustomerId	CreditScore	Balance	Exited	ExitCategory
15757408	655	250898.10	1	Exit
15715622	583	238387.60	1	Exit
15714241	749	222267.60	1	Exit
15571958	489	221532.80	0	Retain
15586674	663	216109.90	1	Exit
15599131	650	214347.00	0	Retain
15594408	584	213146.20	1	Exit
15769818	850	212778.20	0	Retain
15620268	634	212696.30	0	Retain
15780212	592	212693.00	0	Retain
15690589	541	212314.00	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?

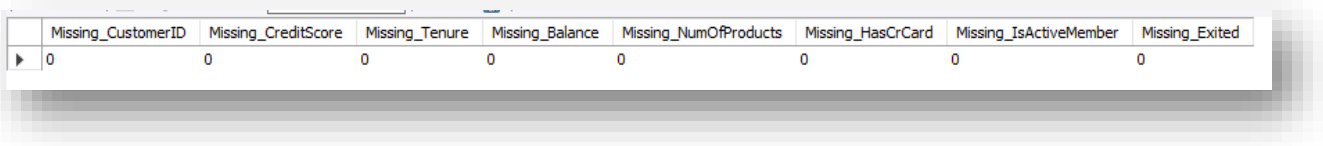
Query:

```
(i) Checking if Bank_Churn table has NULL valuesSELECT

    SUM(CASE WHEN CustomerID IS NULL THEN 1 ELSE 0 END) AS
Missing_CustomerID,
    SUM(CASE WHEN CreditScore IS NULL THEN 1 ELSE 0 END) AS
Missing_CreditScore,
    SUM(CASE WHEN Tenure IS NULL THEN 1 ELSE 0 END) AS
Missing_Tenure,
    SUM(CASE WHEN Balance IS NULL THEN 1 ELSE 0 END) AS
Missing_Balance,
    SUM(CASE WHEN NumOfProducts IS NULL THEN 1 ELSE 0 END) AS
Missing_NumOfProducts,
    SUM(CASE WHEN HasCrCard IS NULL THEN 1 ELSE 0 END) AS
Missing_HasCrCard,
```

```
    SUM(CASE WHEN IsActiveMember IS NULL THEN 1 ELSE 0 END) AS
Missing_IsActiveMember,
    SUM(CASE WHEN Exited IS NULL THEN 1 ELSE 0 END) AS
Missing_Exited
FROM Bank_Churn;
```

Output:

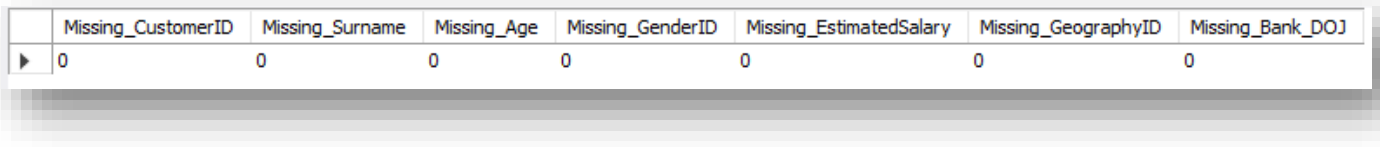


Missing_CustomerID	Missing_CreditScore	Missing_Tenure	Missing_Balance	Missing_NumOfProducts	Missing_HasCrCard	Missing_IsActiveMember	Missing_Exited
0	0	0	0	0	0	0	0

Query:

```
(ii) Checking if CustomerInfo table has NULL values
SELECT
    SUM(CASE WHEN CustomerID IS NULL THEN 1 ELSE 0 END) AS
Missing_CustomerID,
    SUM(CASE WHEN Surname IS NULL THEN 1 ELSE 0 END) AS
Missing_Surname,
    SUM(CASE WHEN Age IS NULL THEN 1 ELSE 0 END) AS Missing_Age,
    SUM(CASE WHEN GenderID IS NULL THEN 1 ELSE 0 END) AS
Missing_GenderID,
    SUM(CASE WHEN EstimatedSalary IS NULL THEN 1 ELSE 0 END) AS
Missing_EstimatedSalary,
    SUM(CASE WHEN GeographyID IS NULL THEN 1 ELSE 0 END) AS
Missing_GeographyID,
    SUM(CASE WHEN Bank_DOJ IS NULL THEN 1 ELSE 0 END) AS
Missing_Bank_DOJ
FROM CustomerInfo;
```

Output:



Missing_CustomerID	Missing_Surname	Missing_Age	Missing_GenderID	Missing_EstimatedSalary	Missing_GeographyID	Missing_Bank_DOJ
0	0	0	0	0	0	0

Insights:

- Upon running the above queries, **no missing values were found** in either the **Bank_Churn** or **CustomerInfo** tables.
- This indicates that the dataset used for analysis is **complete and reliable**, ensuring accurate statistical insights and modeling.
- However, handling missing values remains an important aspect of **data preprocessing**, especially when integrating new data sources or expanding the dataset for predictive modeling.

Recommendations & Handling Methods (If Missing Values Exist):

Imputation Techniques (Preferred for Numeric Columns):

- **Mean / Median / Mode Imputation:** Replace missing numeric values with the mean, median, or mode of the column.
- For categorical data, use the **most frequent category** as a replacement.

Deletion (When Missing Data is Minimal):

- If only a few records contain missing data and they do not significantly affect overall insights, they can be **safely removed**.

Interpolation or Forward/Backward Filling (For Sequential Data):

- In **time-series datasets**, missing values can be **estimated using interpolation** or filled using **previous/next record values** to maintain continuity.
- **Tool Used for Validation and Cleaning:**
  - **MySQL Workbench** was used to identify missing values through SQL queries.
  - **Excel and Python (Pandas)** can also be leveraged for advanced imputation or data cleaning tasks where complex interpolation or statistical estimation is required.

Conclusion:

The dataset is **complete with no missing values**, ensuring **high-quality input for further analysis**. However, understanding missing data handling methods remains critical for **maintaining data integrity** during future updates or when incorporating new datasets.

---

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”.

Query:

```
SELECT
    c.CustomerId,
    c.Surname AS LastName,
    a.ActiveCategory
FROM CustomerInfo c
INNER JOIN Bank_Churn b ON c.CustomerId = b.CustomerId
INNER JOIN ActiveCustomer a ON a.ActiveID = b.IsActiveMember
WHERE c.Surname LIKE '%on'
ORDER BY c.Surname;
```

Output:

CustomerId	LastName	ActiveCategory
15641110	Abron	Inactive Member
15589210	Adamson	Inactive Member
15573599	Adamson	Active Member
15644400	Anderson	Inactive Member
15610403	Anderson	Inactive Member
15661526	Anderson	Inactive Member
15687492	Anderson	Inactive Member
15748327	Anderson	Inactive Member
15594084	Anderson	Inactive Member
15775116	Anderson	Inactive Member
15657228	Anderson	Active Member

---

26. Can you observe any data discrepancy in the Customer's data? As a hint it's present in the IsActiveMember and Exited columns. One more point to consider is that the data in the Exited Column is absolutely correct and accurate.

Query:

```
SELECT * FROM Bank_Churn

        WHERE IsActiveMember = 0      -- Inactive

        AND Exited = 0;         -- But still marked as "Not Exited"
```

Output:

CustomerId	CreditScore	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	Exited
15565701	698	4	161993.90	1	0	0	0
15565779	627	4	57809.32	1	1	0	0
15565796	745	4	96048.55	1	1	0	0
15565806	532	5	0.00	2	0	0	0
15565891	709	5	0.00	2	1	0	0
15566091	545	7	0.00	1	1	0	0
15566111	596	7	0.00	1	1	0	0
15566139	526	7	53573.18	1	1	0	0
15566156	749	3	71497.79	2	0	0	0
15566269	787	3	0.00	2	1	0	0
15566295	761	5	138053.80	2	1	0	0

Insights:

- The dataset reveals cases where **customers are flagged as both active and exited simultaneously** (i.e., `IsActiveMember = 1` and `Exited = 1`).
  This is a **logical contradiction**, as a customer cannot remain active after exiting the bank.
- The inconsistency likely originates from **data entry errors** or **incomplete status updates** during the customer exit process — particularly when the **ActiveMember status** was not updated upon account closure.
- Several records show **customers with zero balances** still marked as **active** or having **multiple products** (`NumOfProducts > 1`) even after exiting.
  This could indicate **incomplete product deactivation** or **delayed account balance updates** after exit processing.
- Such discrepancies highlight the need for **better data synchronization and validation rules** within the CRM and banking systems.

Recommendations:

1. **Automated Validation Scripts:**
   - Develop and schedule **automated SQL scripts or triggers** that continuously scan for contradictions (e.g., customers marked both active and exited).
   - Example logic:
   - `SELECT CustomerID`
   - `FROM Bank_Churn`
   - `WHERE IsActiveMember = 1 AND Exited = 1;`
   - This will help proactively flag inconsistent records for review.
2. **Process Workflow Enforcement:**
   - Ensure that the **ActiveMember flag is automatically set to 0** when a customer's `Exited` status changes to 1.
   - Enforce **data integrity constraints** during updates to prevent conflicting values being recorded.
3. **Data Consistency Audits:**
   - Conduct **regular cross-table audits** between `Bank_Churn`, `CustomerInfo`, and `Product` tables to validate relationships (e.g., products assigned to exited customers).
   - Identify cases where `NumOfProducts > 0` but balance = 0 or exited = 1.
4. **Investigate System Workflows:**
   - Review **data migration and update procedures** — especially during customer churn processing — to detect potential causes such as **delayed triggers**, **API synchronization lags**, or **manual override errors**.
5. **Improve Data Governance:**
   - Implement **data quality rules** and monitoring dashboards to maintain system accuracy.
   - Ensure that all updates to customer activity and exit status are **timestamped** and **logged** for audit traceability.

Conclusion:

The identified discrepancies between IsActiveMember and Exited highlight **data integrity issues** that could impact churn analysis accuracy.
By implementing **automated validation, enforced workflows, and systematic data audits**,

the bank can ensure **cleaner, more reliable datasets** for analytics, reporting, and decision-making.

Thank You.