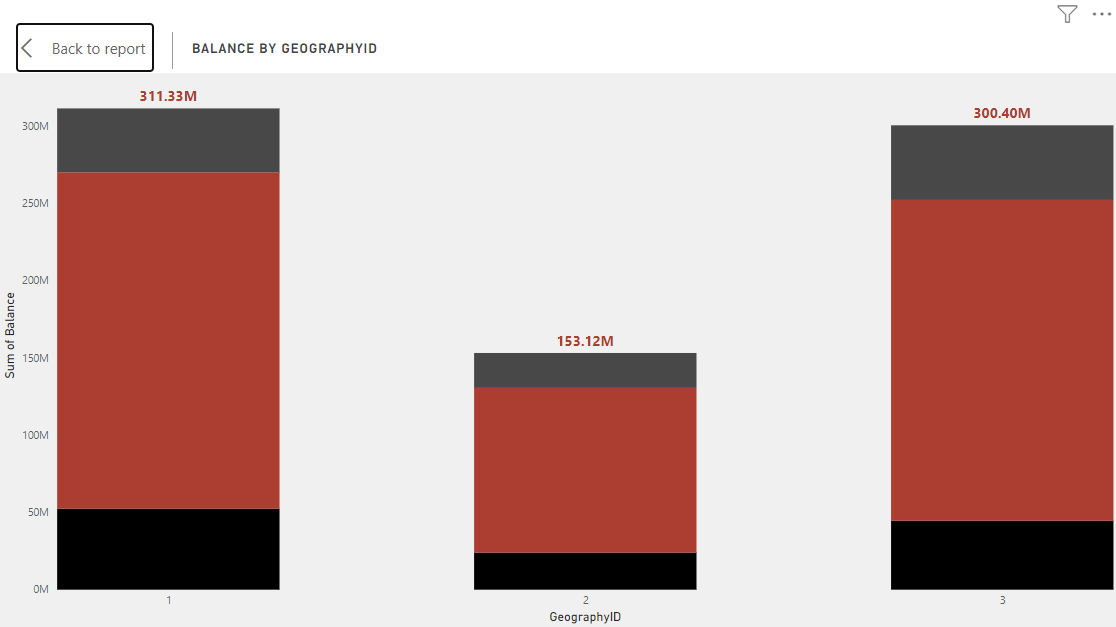
**Learners have to come up with a Report to support the answers to the following questions and suggestions**

**Objective Questions:**

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

****

**Germany: 300.40 million**

**France: 311.33 million**

**Spain: 153.12 million**

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

**SQL QUERY**

Select

customerID,

Surname,

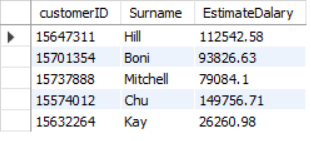
max(EstimatedSalary) as EstimatedSalary

from customerinfo

where quarter(Bank\_DOJ) = 4

group by 1, 2

limit 5;



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

**SQL QUERY**

select

round(avg(NumOfProducts), 0) as AverageNoOfProducts

from bank\_churn

where HasCrCard = 1



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

**DAX FORMULA**

* churn\_rate = divide( calculate( COUNT(Bank\_Churn[Exited]), Bank\_Churn[Exited] = 1), count( Bank\_Churn[Exited]))\*100



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

**SQL QUERY**

select ExitCategory,

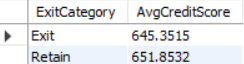
avg(CreditScore) as AvgCreditScore

from bank\_churn b

inner join exitcustomer e

on b.Exited = e.ExitID

group by 1

****

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

**SQL QUERY**

select genderCategory,

Round(avg(EstimatedSalary), 0) as AvgEstimatedSalary,

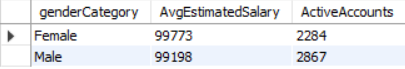
count(IsActiveMember) as ActiveAccounts

from customerinfo c inner join gender g on c.GenderID = g.GenderID

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

where IsActiveMember = 1

group by 1;



* Female Customers have higher average salary than male customers, also female active users are less compared to male active users.

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

**SQL QUERY**

select case

when CreditScore >= 800 then 'Excellent'

when CreditScore < 800 and CreditScore >= 740 then 'Very Good'

when CreditScore < 740 and CreditScore >= 670 then 'Good'

when CreditScore < 670 and CreditScore >= 580 then 'Fair'

else 'Poor'

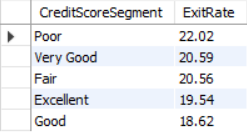
end as CreditScoreSegment,

Round((sum(Exited)/count(\*))\*100, 2) as ExitRate

from bank\_churn

group by 1

Order by 2 desc;



* Credit score segment with the highest Exit Rate is Poor.

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

**SQL QUERY**

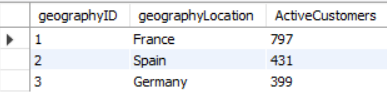
select c.geographyID, g.geographyLocation, count(b.IsActiveMember) as ActiveCustomers

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

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

where IsActiveMember = 1 and Tenure > 5

group by 1, 2;



* France has the highest no. of Active Customers of all Geographic Locations.

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

**SQL QUERY**

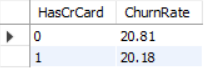
select HasCrCard,

Round((sum(Exited)/count(\*))\*100, 2) as ChurnRate

from bank\_churn

group by 1

order by 2 desc;



* We can see that Customers who don’t have Credit Card has higher Churn Rate than the customers who have Credit Card.

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

**SQL QUERY**

select NumOfProducts,

count(Exited) as ExitCustomers

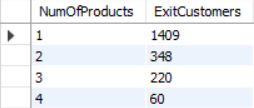
from

bank\_churn

where Exited = 1

Group by 1

order by 2 desc;



* The maximum number of customers who left were using only 1 product. Results from the above image shows us that with increasing numbers of products used, the number of customers who left decreased.

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

**SQL QUERY**

select year(Bank\_DOJ) as 'year',

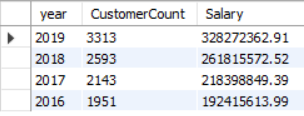
count(customerID) as CustomerCount,

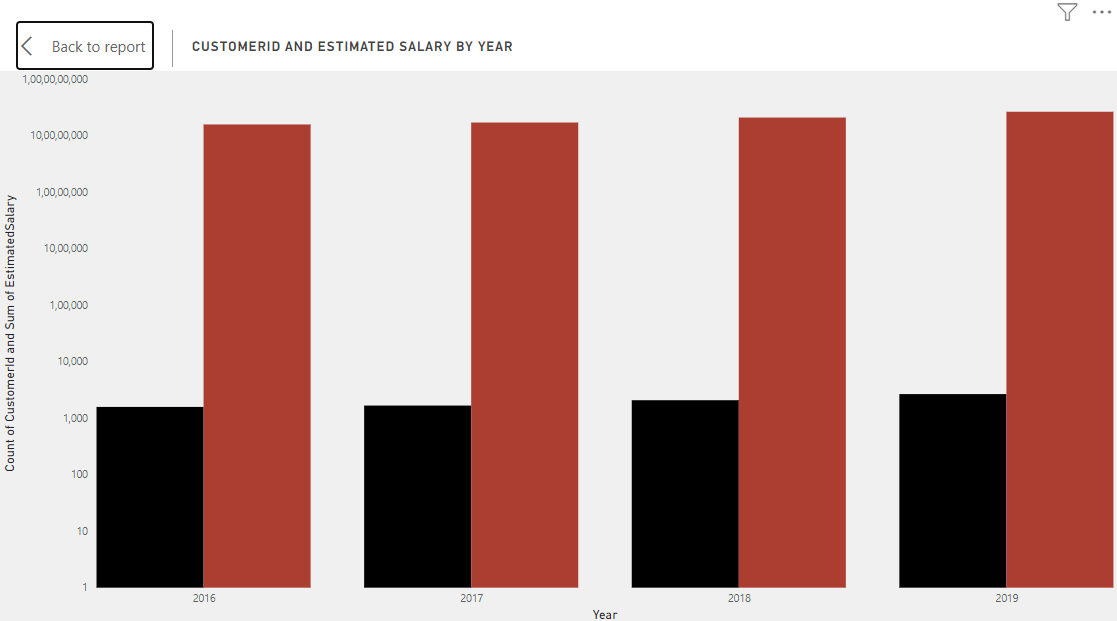
round(sum(EstimatedSalary), 2) as Salary

from customerinfo

group by 1

order by 2 desc, 3 desc;





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

**SQL QUERY**

select NumOfProducts, count(Exited) as ExitCustomers,

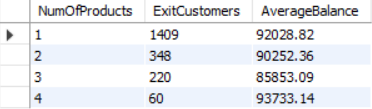
round(avg(Balance), 2) as AverageBalance

from Bank\_churn

where Exited = 1

group by 1

order by 2 desc;



* We can see from this above image, average balances for different numbers of products used by customers, along with customers who have exited.

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

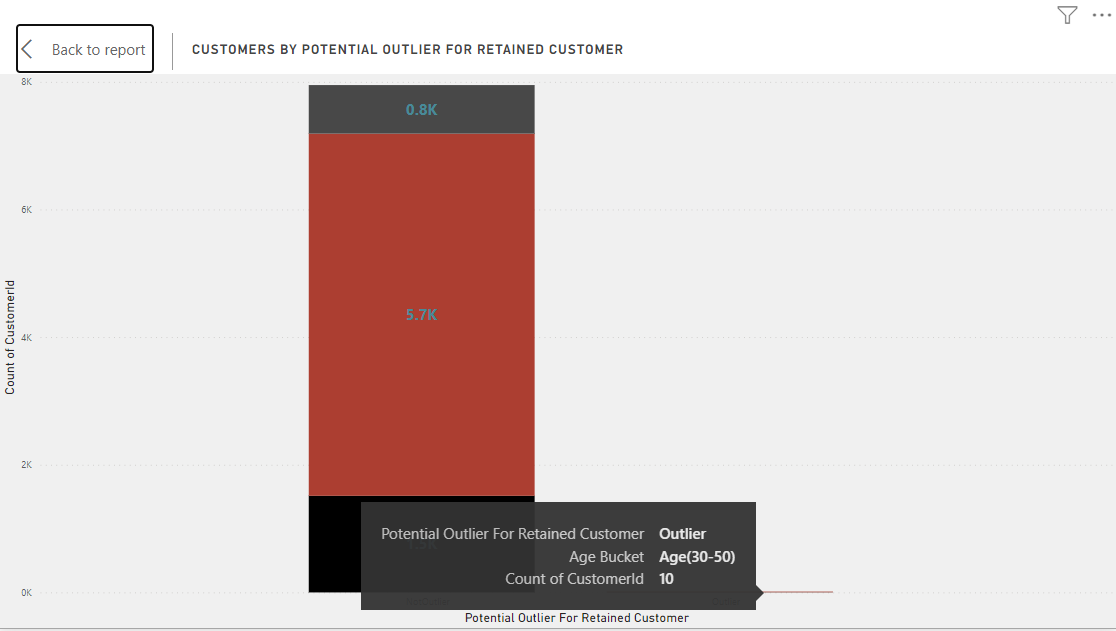
* There are different methods in statistics to identify the outliers in the data. One method which we have used is using Z Scores.
* **Z Scores:** Z-score is a statistical measurement that describes a value's relationship to the mean of a group of values.
* Z = (value - mean)/standard deviation

The DAX formula which I have used in Power BI is as follows:

Z-Score=ROUND((Bank\_Churn[Balance]-AVERAGE(Bank\_Churn[Balance]))/STDEV.P(Bank\_Churn[Balance]), 2)

I also created another column Potential Outlier for retained customers using this DAX formula:

Potential Outlier For Retained Customer = if(Bank\_Churn[Exited] = 0, if(Bank\_Churn[Z-Score] > 2, "Outlier", "NotOutlier") , "Not Applicable")



From this chart we can see, there are very few outliers in the data for customers who are retained by the bank.

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

* There are a total of 7 tables given in the dataset. Out of these 7, 5 tables namely activecustomers, creditcard, exitcustomer, gender, geography are tables consisting of categorical variables

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

**SQL QUERY**

select g.GeographyID,

ge.GenderCategory,

avg(b.Balance) as AverageIncome

rank() over(partition by g.GeographyID order by avg(b.Balance) desc) as 'Rank'

from customerinfo c

inner join bank\_churn b

on b.CustomerId = c.CustomerId

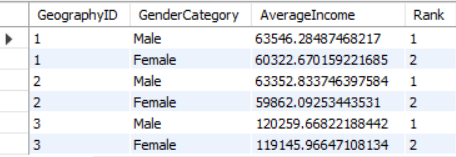
inner join geography g

on c.GeographyID = g.GeographyID

inner join gender ge

on c.GenderID = ge.GenderID

group by 1, 2



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

**SQL QUERY**

select case

when Age between 18 and 30 then 'Young Adults'

when Age between 30 and 50 then 'Adults'

else 'Old'

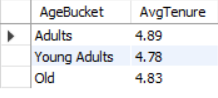
end as AgeBucket,

round(avg(Tenure), 2) as AvgTenure

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

where Exited = 1

group by 1;



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



* This Graph represents relation between salary and balance of customers for customers who have exited. This shows that there is a linear relation between Balance and Salary for customers who have exited the bank.

Similar results can be seen between Balance and Salary for customers who are retained by the bank.



* **From the above results we can say that Balance in Bank Account is directly related to the salaries of customers.**

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



* From the above chart, we can see that customers with Excellent Credit Score have the least Total Salary. While Fair Credit Score Customers have the highest Total Salary.
* But the count of customers is also least for Excellent Credit Score customers. This shows us that we have least no. of customers in Excellent Credit Score Category and most no. of customers in Fair Credit Score category.

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

**SQL QUERY**

select case

when CreditScore >= 800 then 'Excellent'

when CreditScore < 800 and CreditScore >= 740 then 'Very Good'

when CreditScore < 740 and CreditScore >= 670 then 'Good'

when CreditScore < 670 and CreditScore >= 580 then 'Fair'

else 'Poor'

end as CreditScoreSegment,

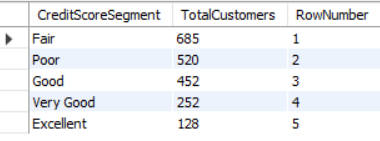
count(CustomerID) as TotalCustomers,

row\_number() over( order by count(CustomerID) desc) as 'RowNumber'

from bank\_churn

where exited = 1

group by 1;



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

(1). According to the age buckets find the number of customers who have a credit card

**SQL QUERY**

select case

when Age between 18 and 30 then 'Young Adults'

when Age between 30 and 50 then 'Adults'

else 'Old'

end as AgeBucket,

count(c.CustomerID) as CustomerCount

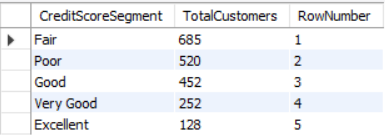
from bank\_churn b

inner join customerinfo c

on b.CustomerId = c.CustomerId

where HasCrCard = 1

group by 1;

****

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

**SQL QUERY**

select case

when Age between 18 and 30 then 'Young Adults'

when Age between 30 and 50 then 'Adults'

else 'Old'

end as AgeBucket,

(sum(HasCrCard)) as TotalCards

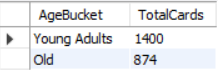
from customerinfo c

inner join bank\_churn b

on c.CustomerId = b.CustomerId

group by 1

having (sum(HasCrCard)) < (select (sum(HasCrCard))/3 from bank\_churn)



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

**SQL QUERY**

select GeographyID,

count(c.CustomerID) as CustomerCount,

row\_number() over( order by count(CustomerID) desc) as 'CustomerRANK',

round(avg(Balance), 2) as AverageBalance,

row\_number() over( order by avg(Balance) desc) as 'BalanceRANK'

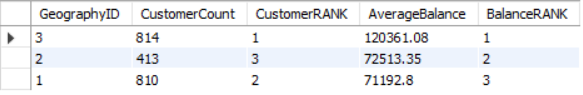
from customerinfo c

inner join bank\_churn b

on c.CustomerId = b.CustomerId

where Exited = 1

group by 1;



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

**SQL QUERY**

select concat(CustomerID,'\_', Surname) as CustomerID\_Surname from customerinfo



**(**NOTE: List of customerID\_Surname is too big to fit in this doc file, hence attaching a small sample from the output of the query for reference.**)**

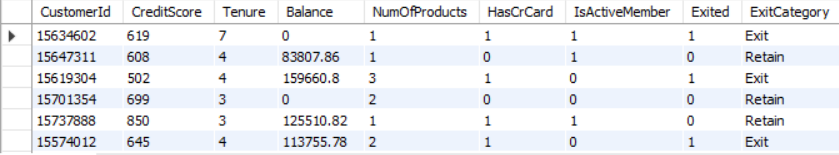
1. Without using “Join”, can we get the “ExitCategory” from ExitCustomers table to Bank\_Churn table? If yes do this using SQL.

**SQL QUERY**

SELECT \*,

(SELECT ExitCategory FROM exitcustomer ec WHERE ec.ExitID = Bank\_Churn.Exited) AS ExitCategory

FROM Bank\_Churn;



**(**NOTE: Output is too big to fit in this doc file, hence attaching a small sample from the output of the query for reference.**)**

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

For Data Cleaning, I used Power Query Editor of Power BI, where I check for missing values, duplicates. I also checked all data types and whether or not they were in the correct format.

**Although there were no missing values in the dataset**, there were few irregularities which I corrected. The Datatype was incorrect for the Bank\_DOJ column in CustomerInfo table. Also I tried removing any duplicates from the data if there were any, to optimise the dataset before analysing it.

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

**SQL QUERY**

select c.CustomerId,

Surname,

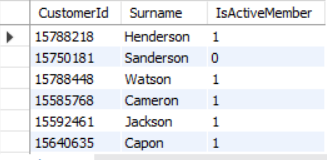
IsActiveMember

from customerinfo c

inner join bank\_churn b

on c.CustomerId = b.CustomerId

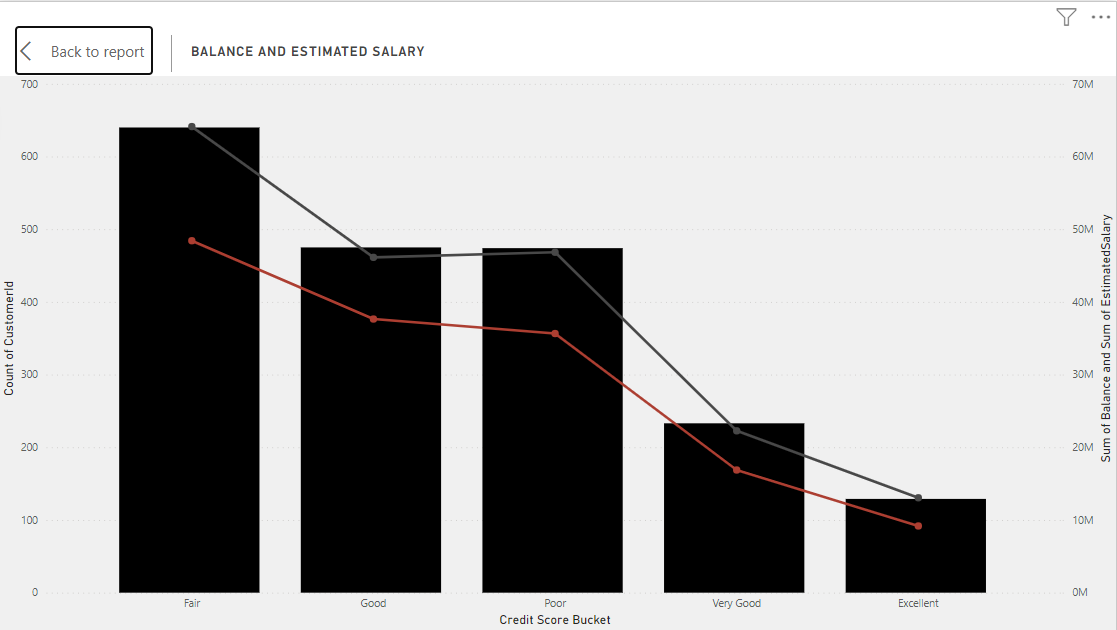
where Surname like '%on'



**(**NOTE: Output is too big to fit in this doc file, hence attaching a small sample from the output of the query for reference.**)**

**Subjective Question:**

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?





* I have prepared this chart of Balance, salary and Total Customers by Credit score buckets to see the patterns and relations.

Here we can see that **maximum no of customers are in Fair, Good and Poor segments of Credit score. While only 1224 and 655 customers are in Very Good and Excellent segments respectively.** I have used the slicer shown in the second Screenshot for year by year relations between all three.

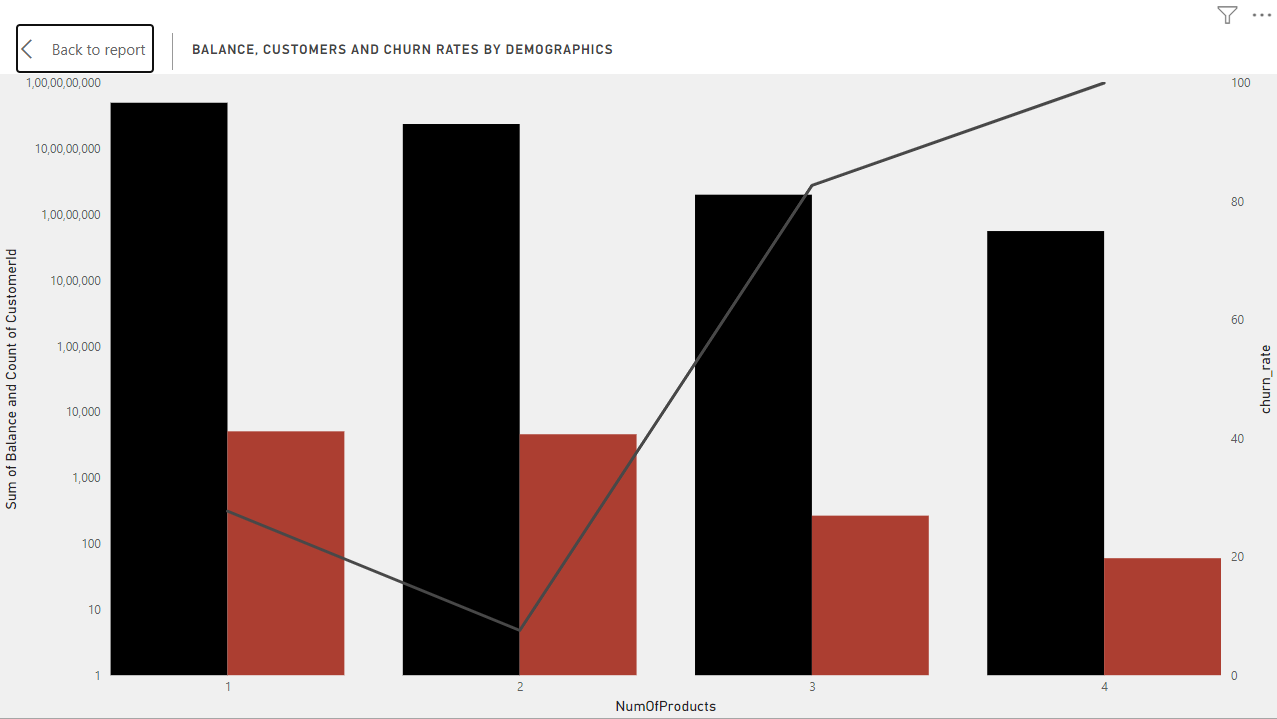
The results here are very fascinating. **We can see from the charts that the distance between lines of Balance and Salary decreases as the Credit Score segment is changed from Poor to Excellent. We can see that it is least for Very Good and Excellent Credit Scores. This shows that The balance is very close to salaries received by customers and they spend the money wisely. For other segments we can see that Amount in their balances and their salaries have big gaps, showing there are chances of bad spending habits of customers and thus their credit scores reflect the same.**

* After analysing this report for different joining years and tenure, I have found similar patterns for years 2016, 2017 and 2018. In 2019 there is a slight decrease in distance between the two lines, showing customers have tried to change their spending habits.



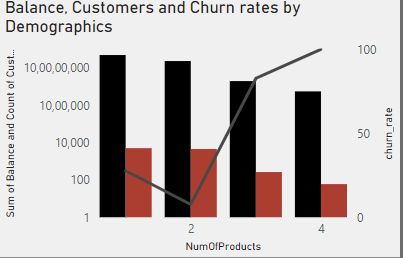
I found churn rates of different Credit score segments for different years. Here I found that as customers have better credit scores as well as the tenure is big. They are less likely to leave the bank and are loyal customers of the banks. Customers with poor credit scores are at risk of leaving the bank.

1. 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 82.71% and 100% respectively.





The Balance is also maximum in the customers who use 1 or 2 products.

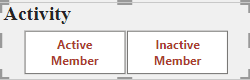




For further analysis of cross selling products and diving deep into the analysis and VIsualizing we need more information on products with all required KPIs.

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

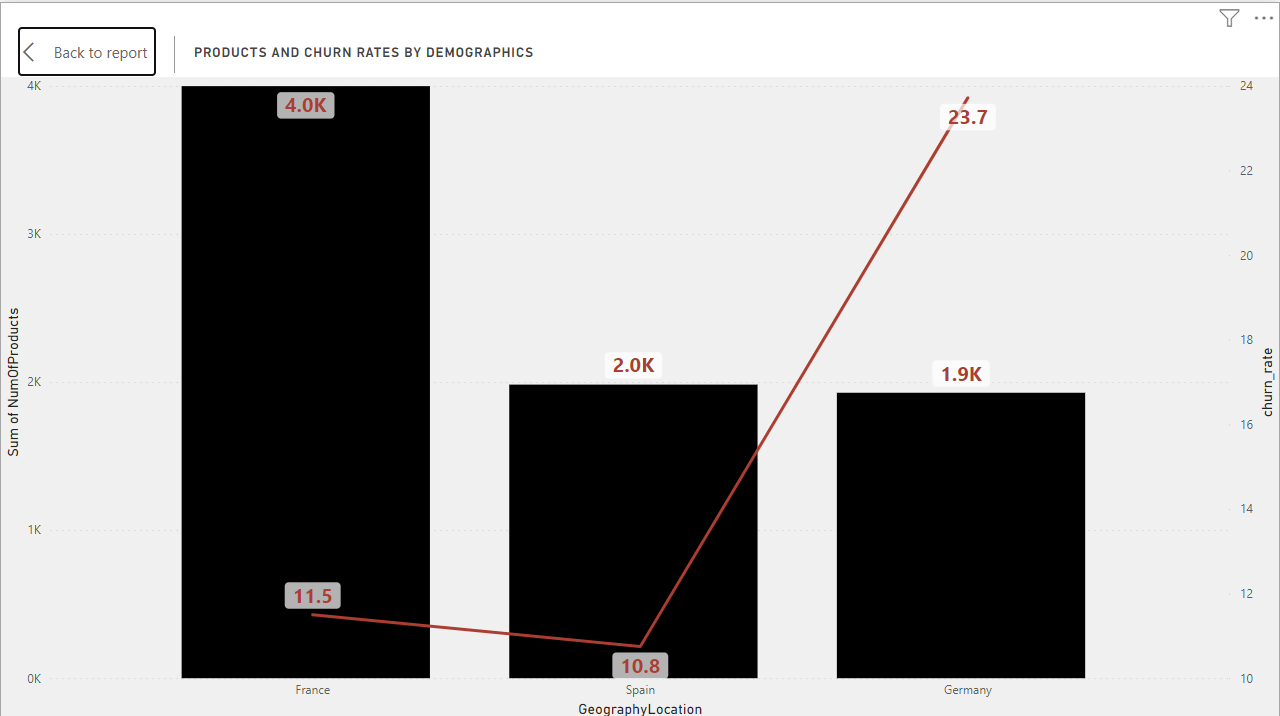
**Geographic Market Trends:**



(NOTE: I HAVE USED THIS SLICER TO BIFURCATE BETWEEN ACTIVE AND INACTIVE CUSTOMERS IN THE CHARTS)

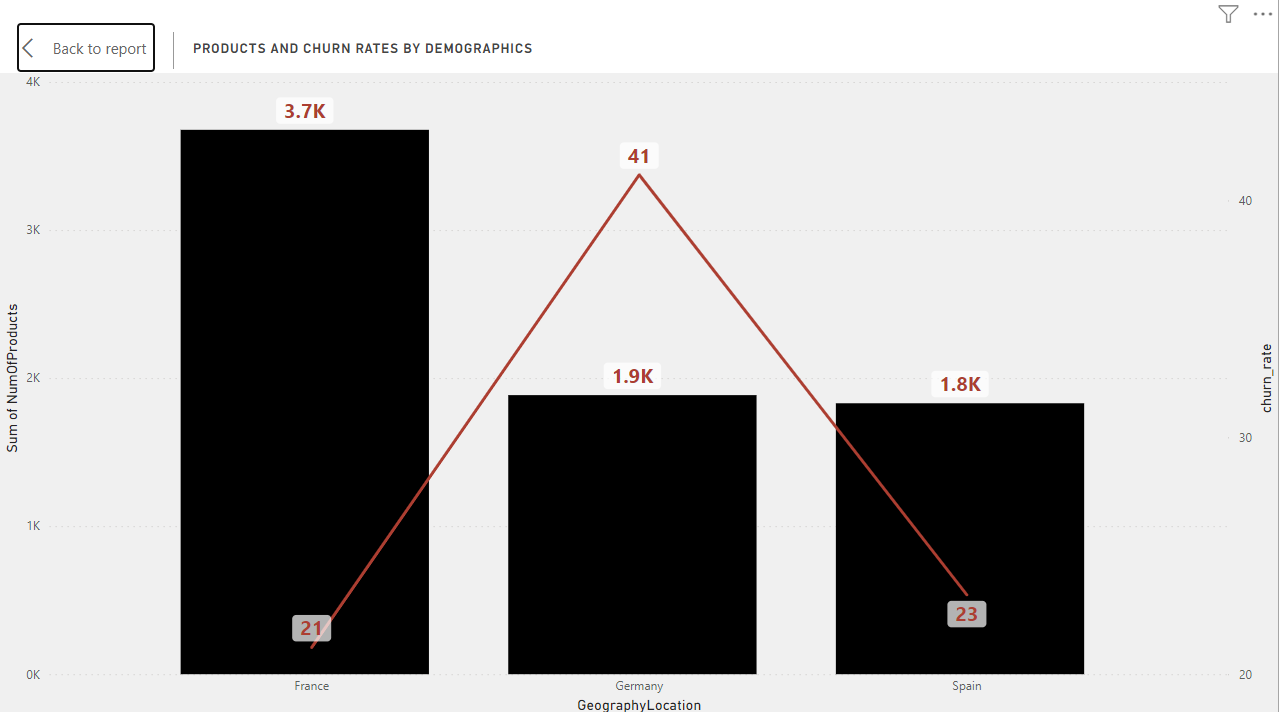
Products:

ACTIVE CUSTOMERS



France consumes most no. of products 4k from total products consumed. Also from the chart we can see France and Spain have lower churn rates with 11.5 and 10.8 respectively Germany has less products as well as high churn rates.

INACTIVE CUSTOMERS

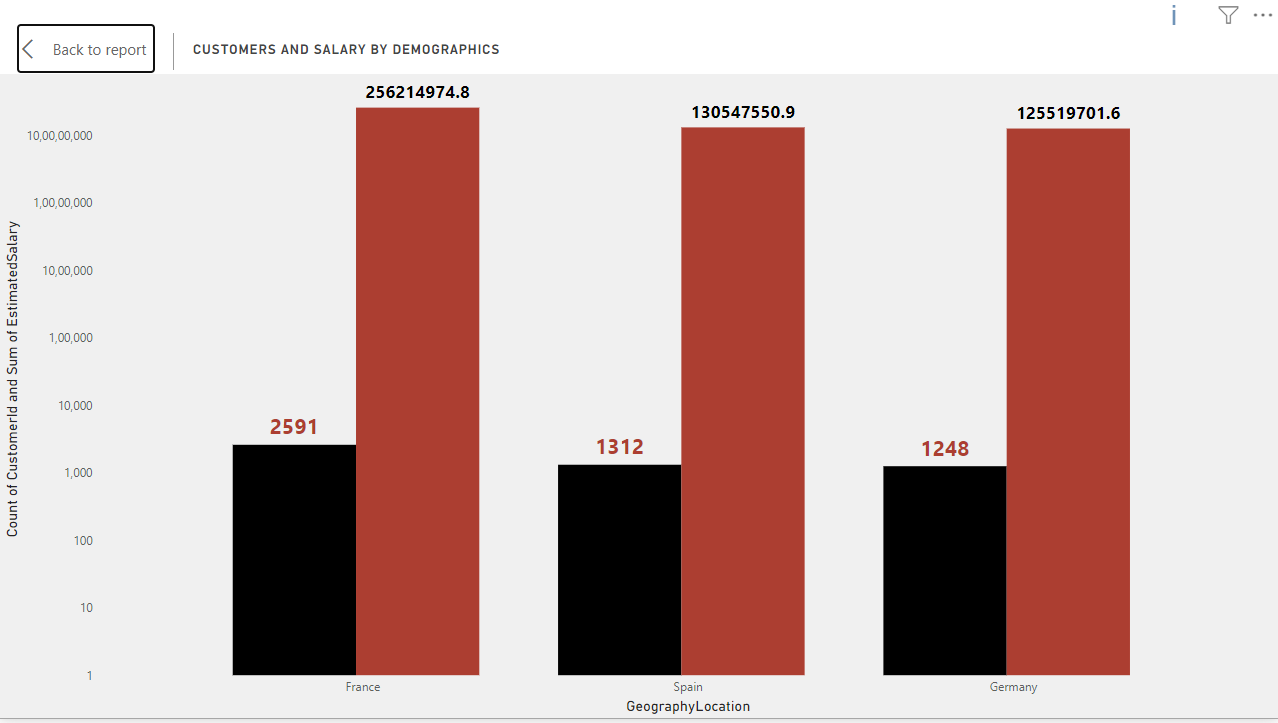


For inactive customers also, products are consumed most by france and patterns are identical to products from Active customers.

**Although while comparing churn rates, I found out that churn rates are significantly higher for all regions for Inactive customers. This can be a key indicator for RISK and should be treated with priority.**

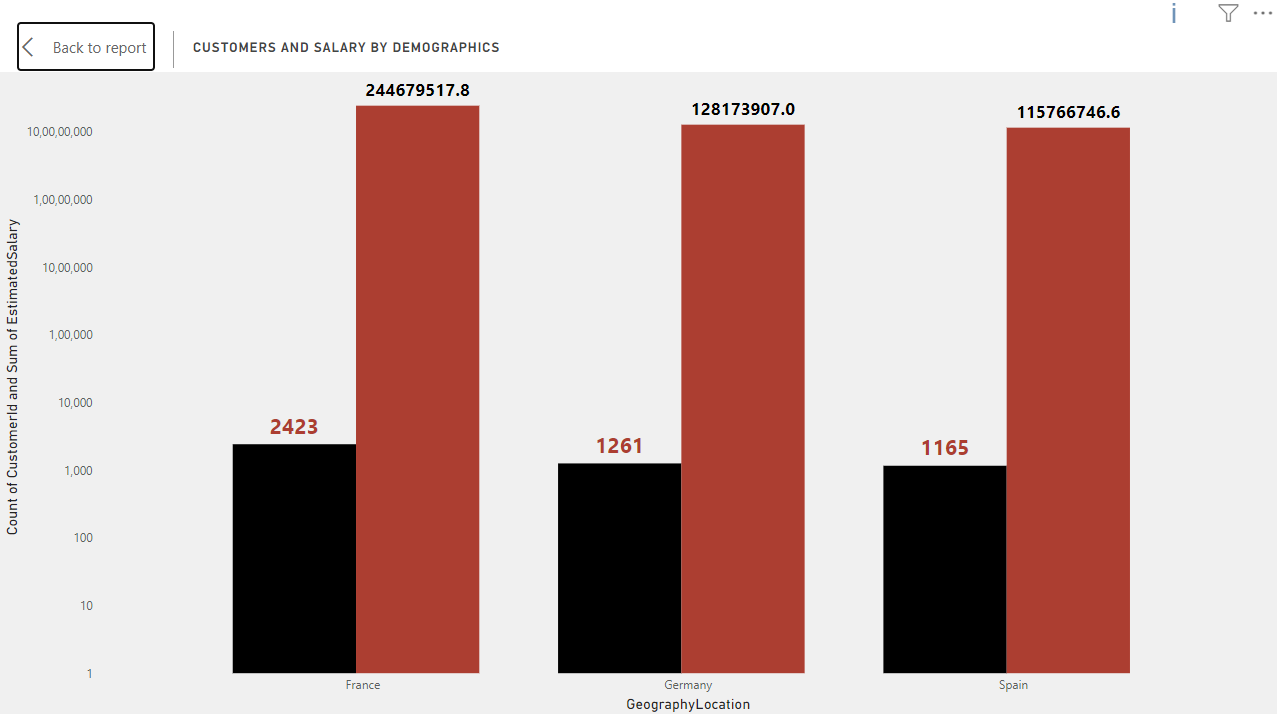
Salaries:

ACTIVE CUSTOMERS



The chart shows total salaries and count of customers for each region. Here also we can see similar trends as we saw in the products section. France is leading in terms of salaries as well. This indicates that our maximum business comes from France and there is room for improvement in other two regions

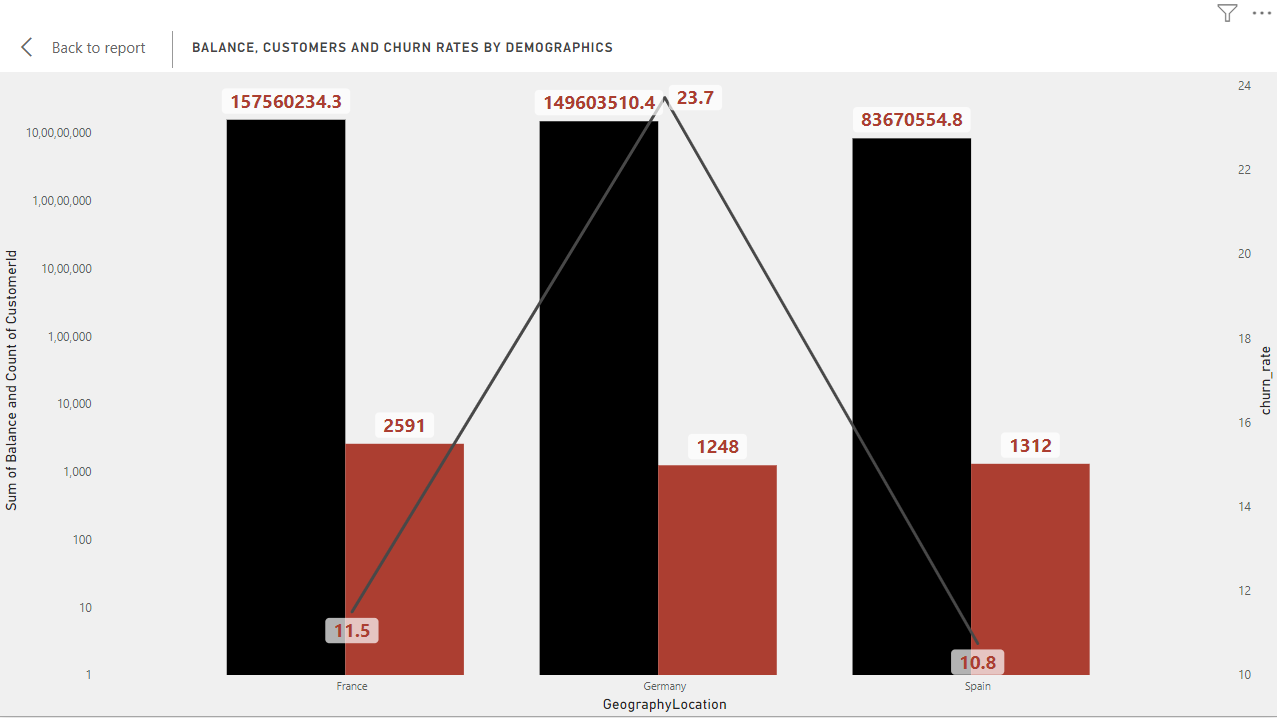
INACTIVE CUSTOMERS



The results are identical for inactive customers as well.

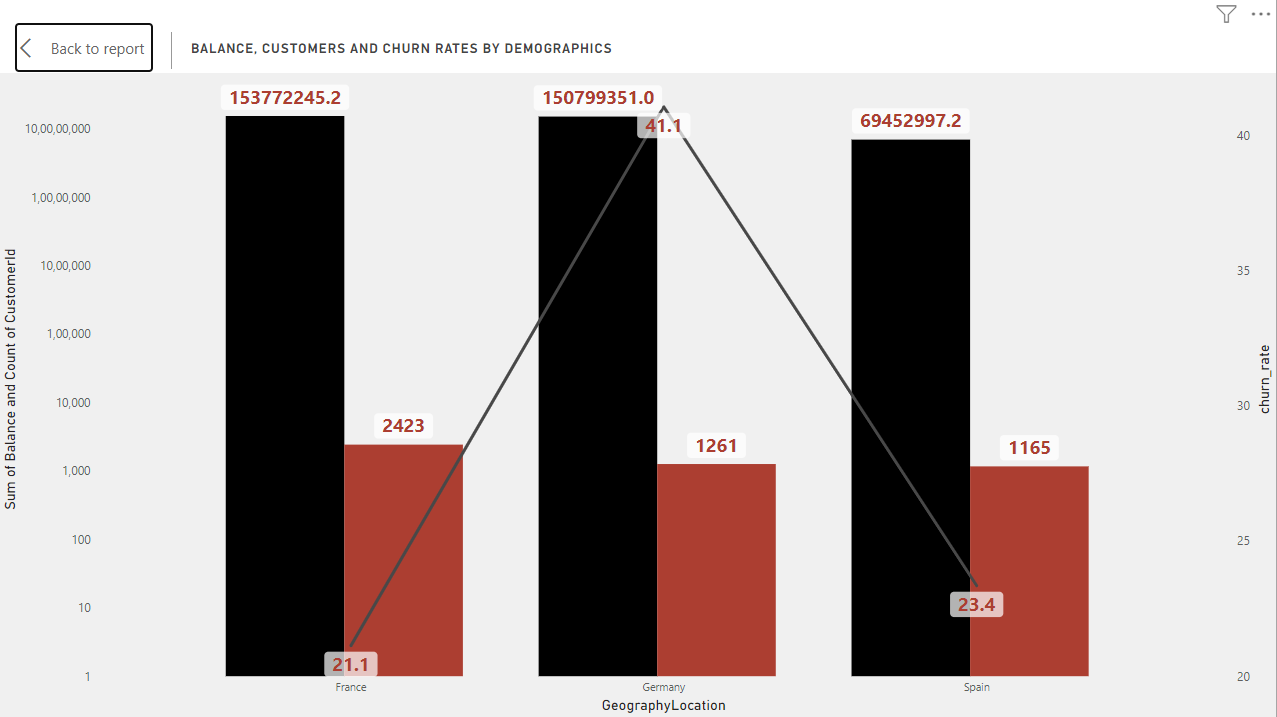
Balance:

ACTIVE CUSTOMERS



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

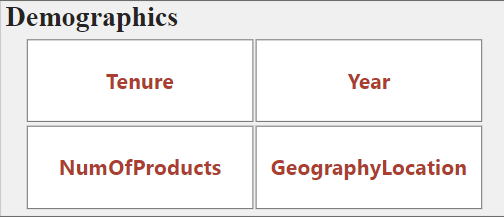
INACTIVE CUSTOMERS



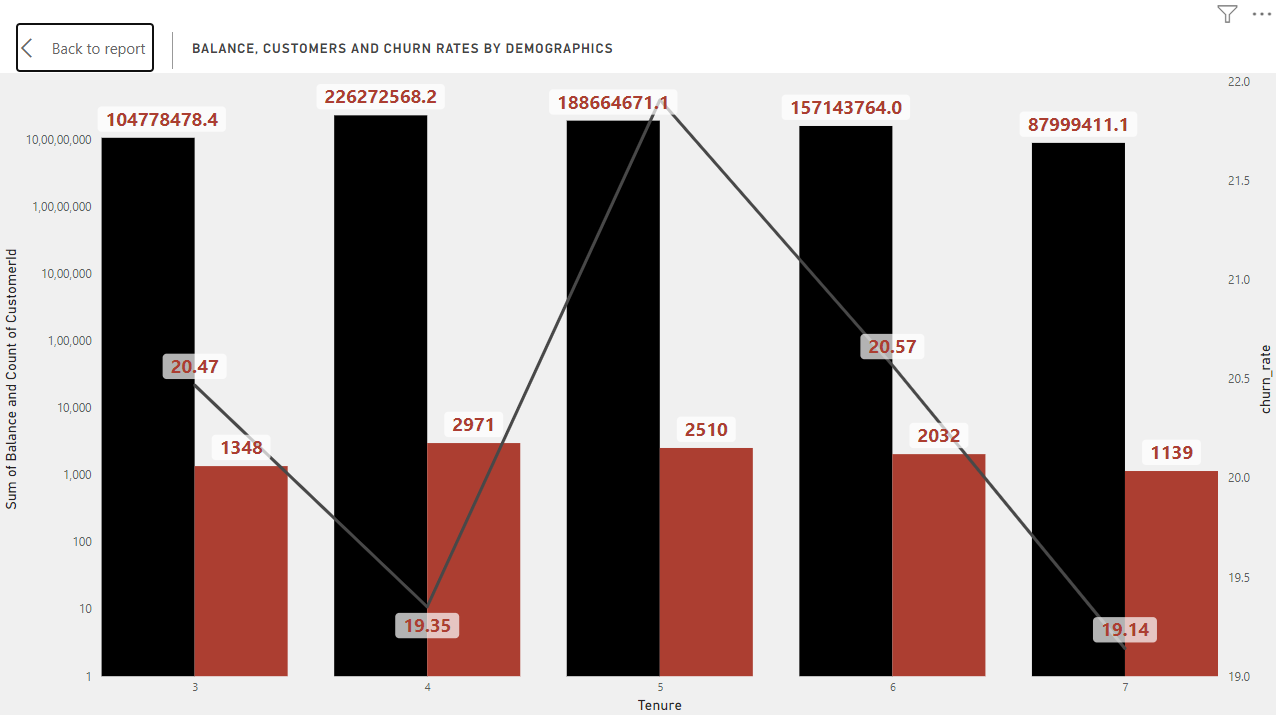
Same trend can be seen for both active and inactive members in terms of balances.

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

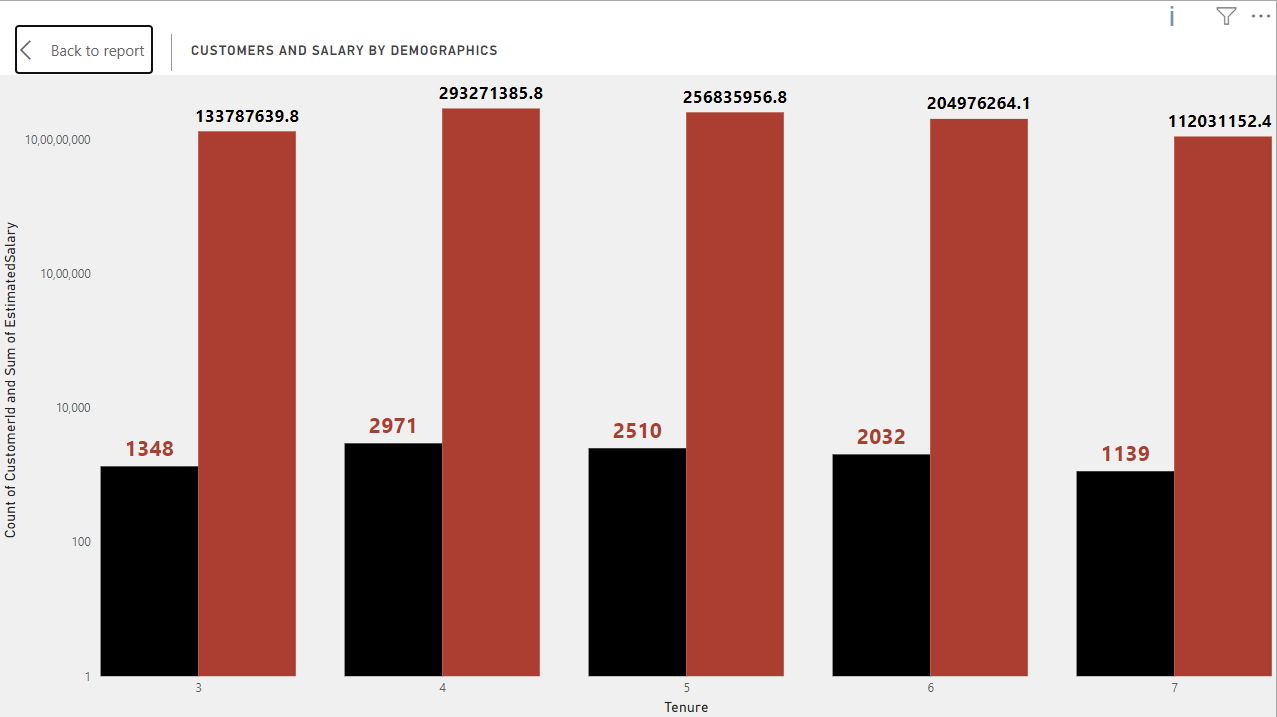
I have created the Demographics parameter from field parameters for this and we will analyse different charts for conclusions.



**BY TENURE:**



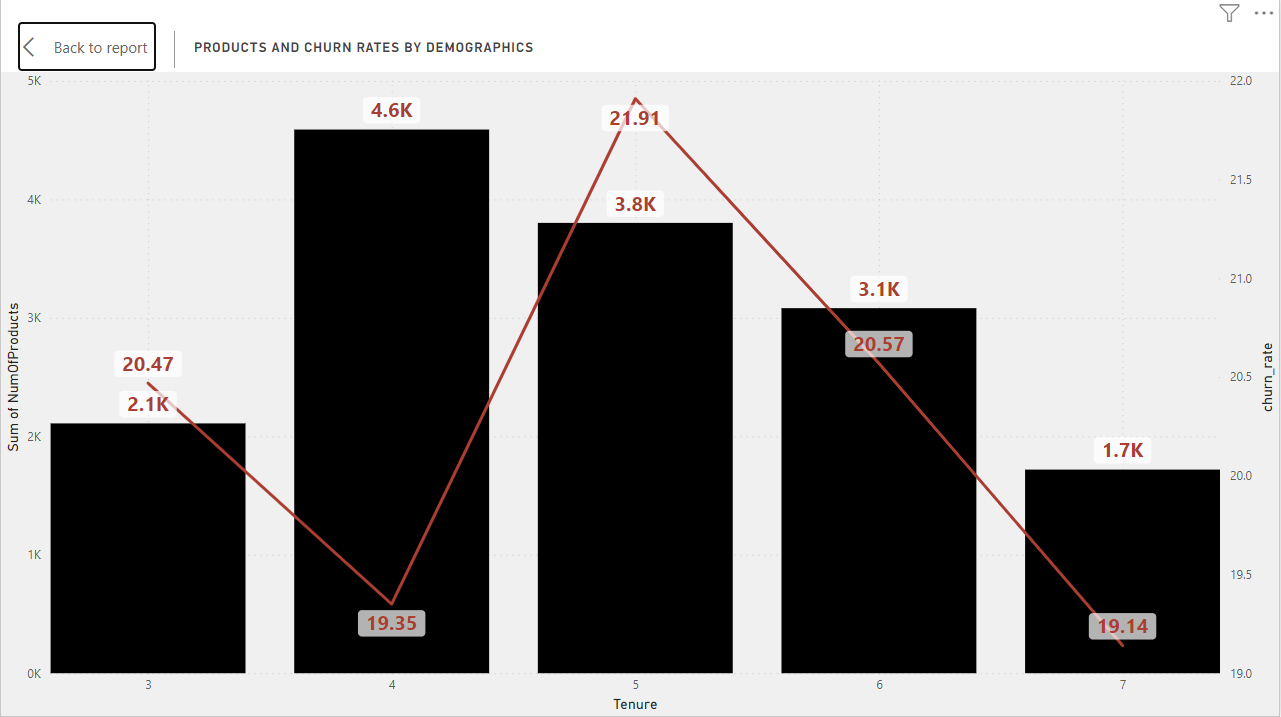
The above chart gives us idea on how balance of accounts varies with tenure



This graph gives us insights on salaries over different tenure.

Both Balance over tenure and salaries over tenure gives us similar insights. We can see that the balance and salaries of customers are less when tenure is small. Then we see a spike in balance and salaries both when tenure is 4 and 5 years. Then there is a drop in both when tenures are 6 and 7.

When we see the churn rates for different tenure we get key insights

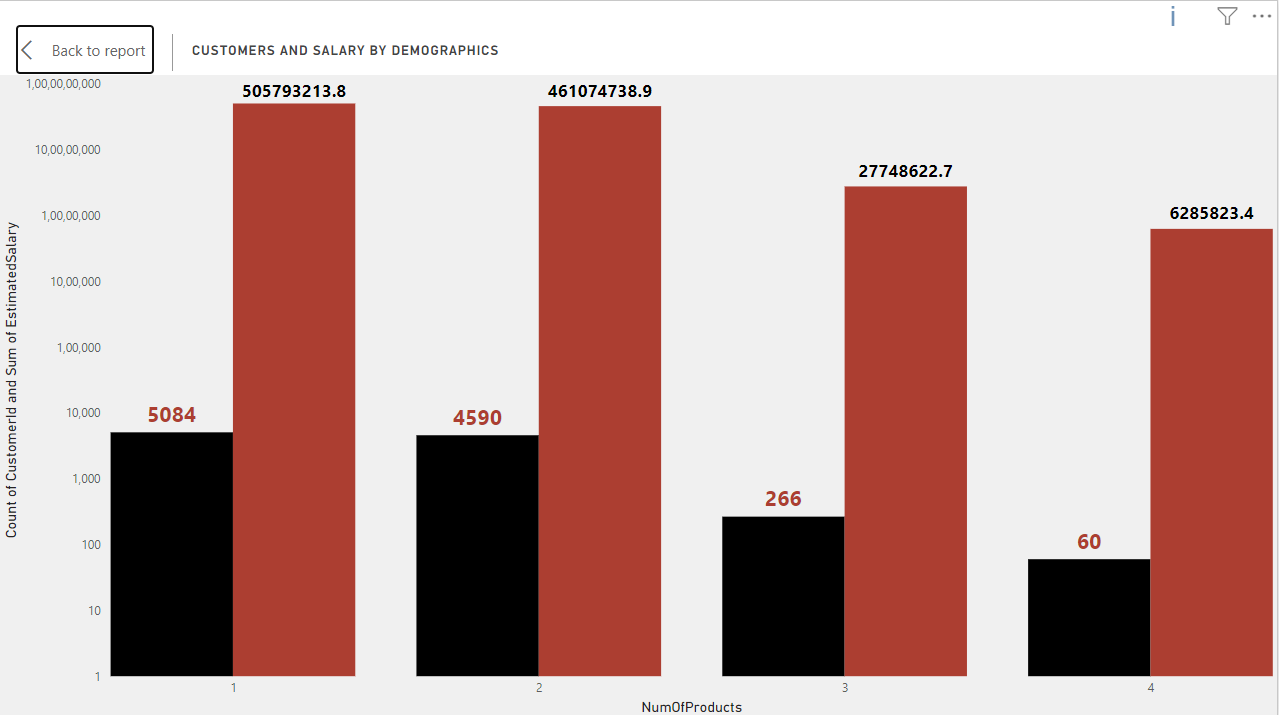


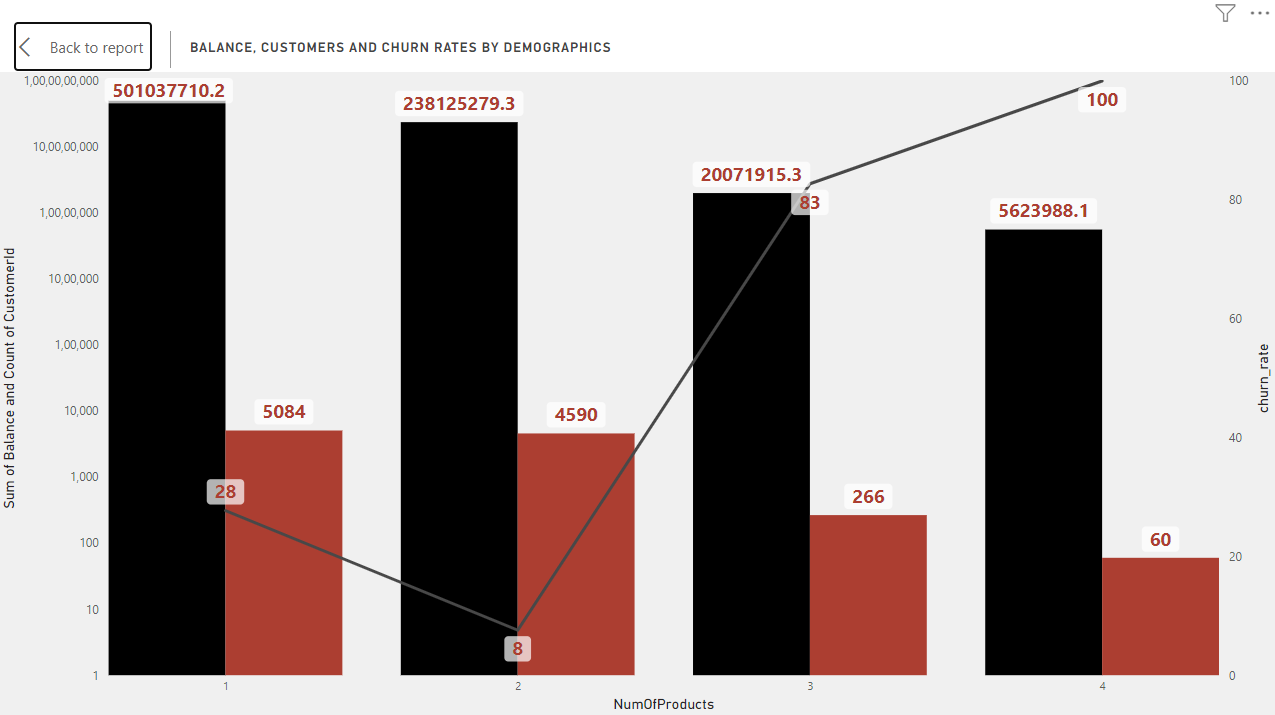
After seeing churn rates over the tenure we see that as tenure increases from 3 to 4 the churn rates decreases and we see that balance and salaries increases this indicates that no. of customers are also increasing. Then we see churn rate again increases in tenure 5 and we see drop in salaries and balance, this decline continues with increase in tenure.

This indicates that an average customer will stay for 4 to 5 years in the bank and then there are chances that he will leave the bank.

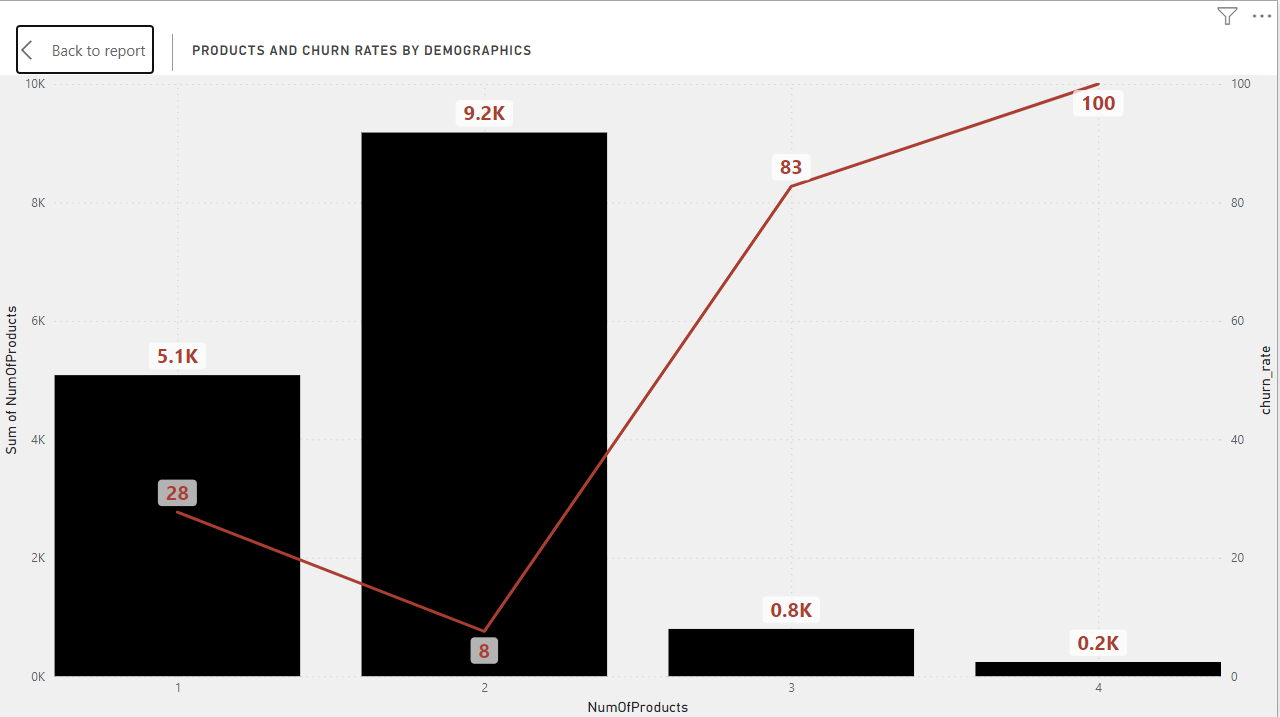
This indicates a High Risk for the bank, because as the customer will reach a tenure of 5 there are high chances that customer will leave the bank.

**BY NO. OF PRODUCTS:**

****

****

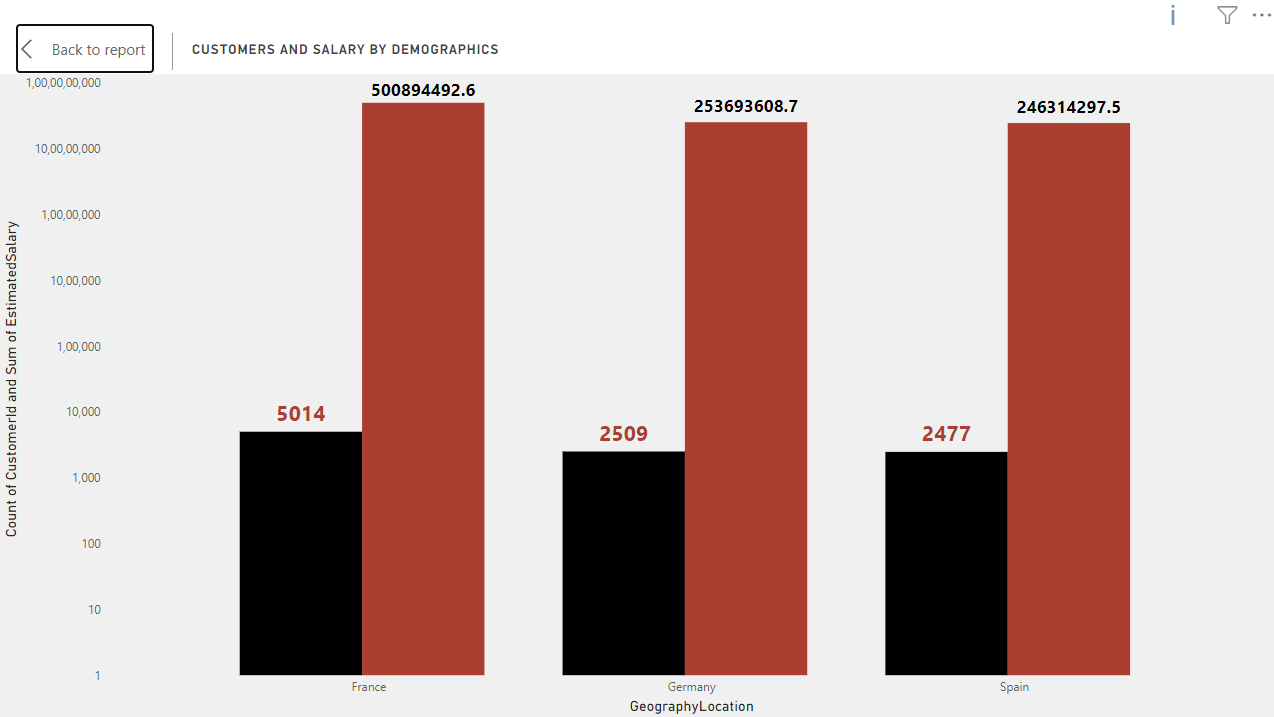
After comparing the charts of balance and salaries by the No. of products. We can see that as no. of products increases, customers are decreasing.

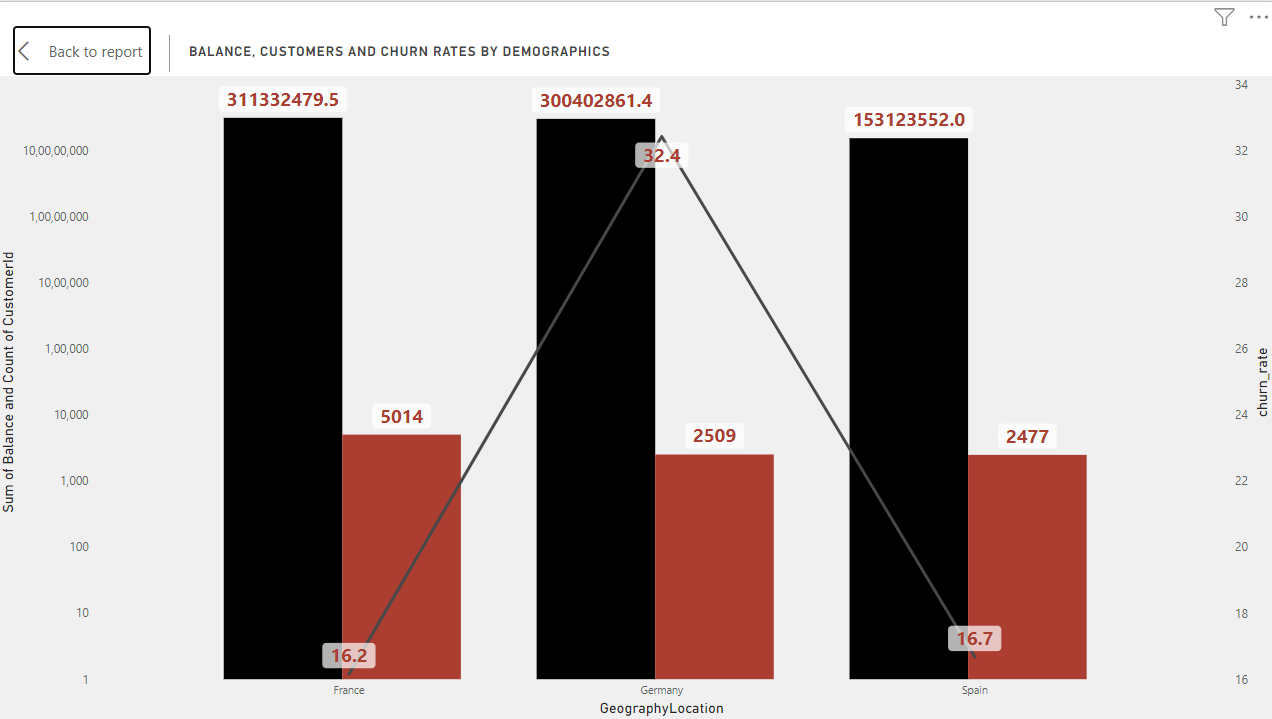


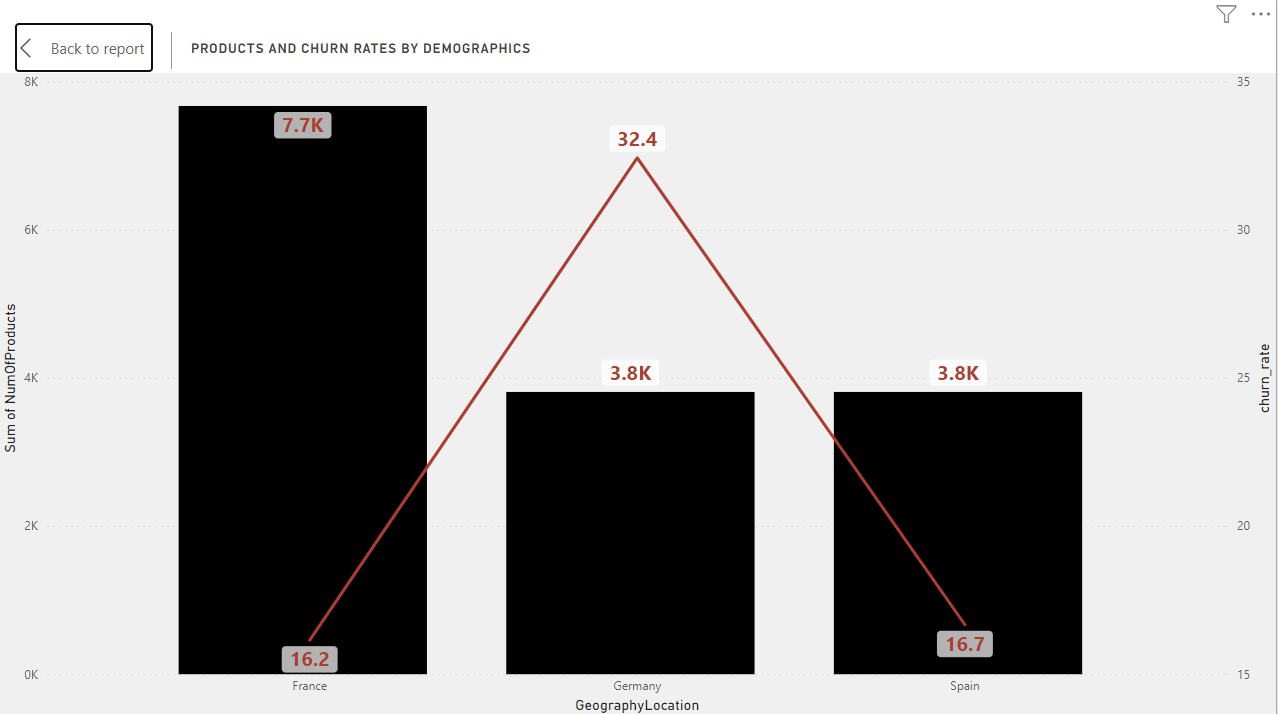
While we see churn rates over products we see similar results too. With increase in no. of products, the churn rates are also increasing.

For further analysis on how No. of Products affect customer churn rates we need further data and information.

**BY GEOGRAPHY:**







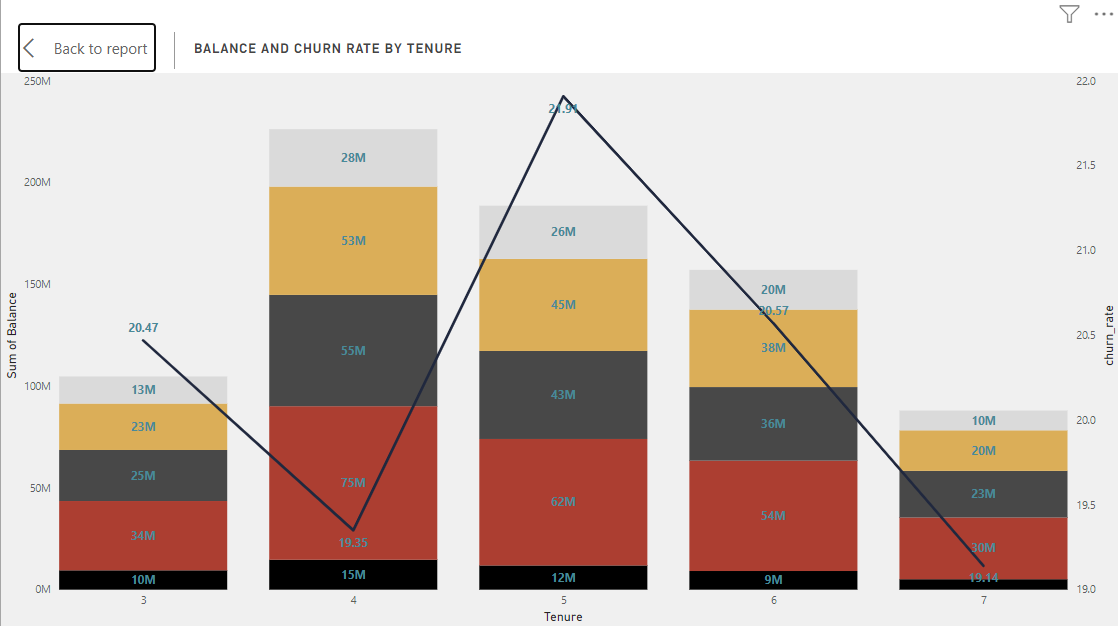
After analysing Balance, salaries and churn rates, we see that France is where our business most comes from. We see that churn rates are most in Germany and we need to make extra efforts to retain customers there.

Also the business is less in Germany and Spain, and we need to focus on capturing more customers from there.

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

I have divided the customers into two segments: Based on credit scores, Based on Age.

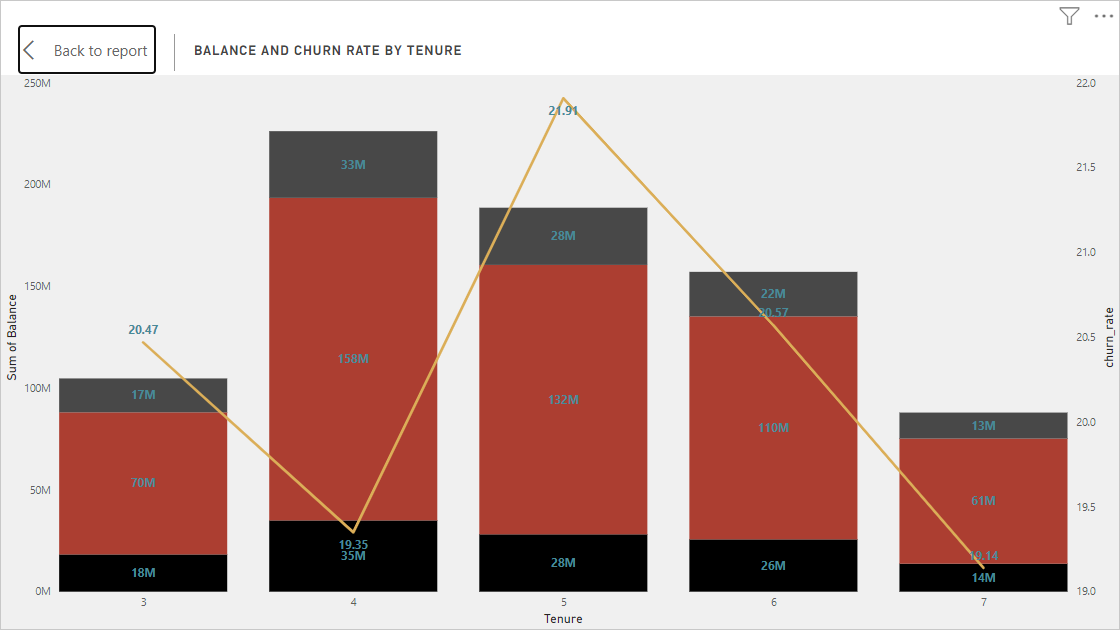
Tenure(Credit Score Segment):



When we see balance distribution over different credit card segments, we see that maximum balance lies in the Fair Segment of customers. The least Balance is found in Excellent and Very Good segments.

In Poor segments too, the balances are significant.

This trend follows as we go through different tenures, and we see this is a general problem. Old as well as new customers both are high in the Fair and Poor segment.



The maximum balance lies in the age group of 30 to 50. We see that young as well as old both customers have comparatively very less balances. Young customers can have long tenures with the bank and we need to give extra focus on young customers.

We need to focus our marketing strategies towards these young customers.

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

After analysing customer segments on question 5. We can find key segments where we need to focus our marketing on.

This segments are

Based on age: Young and OLd

Based on Credit Score: Very Good and Excellent.

For preparing marketing strategies, we will need additional information like what type of products customers use. We will also need relations of these products with geographical regions, Tenures, Activity of the customers based on the products they use.

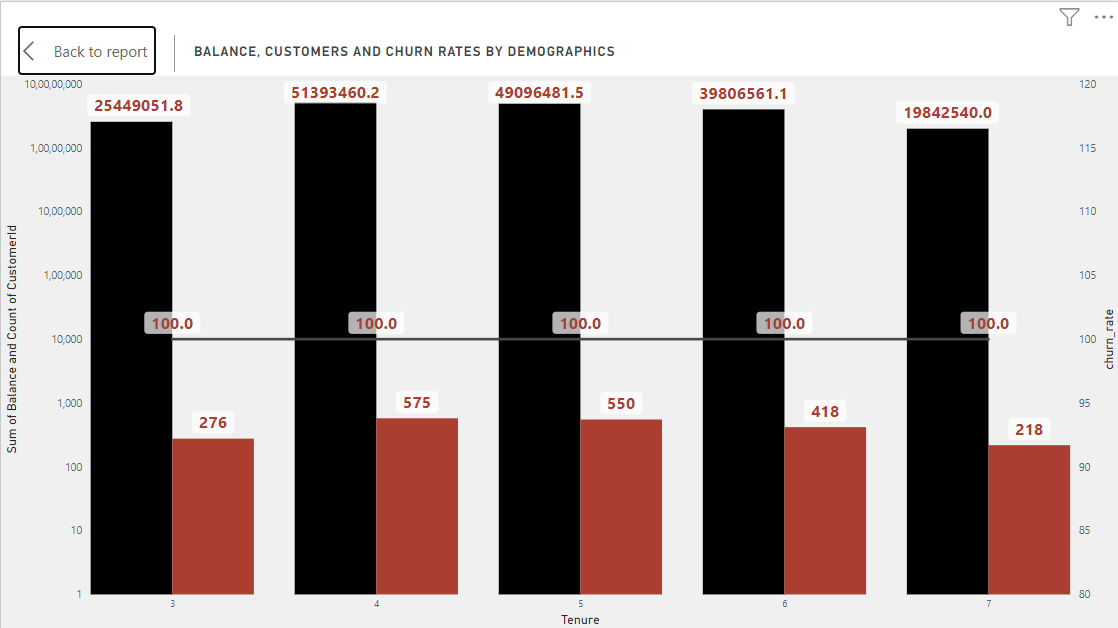
1. Customer Exit Reasons Exploration: Can you identify common characteristics or trends among customers who have exited that could explain their reasons for leaving?

For analysing different factors, I have created a slicer for customers who are retained and who have exited. I used this slicer to filter charts for customers who left the bank.

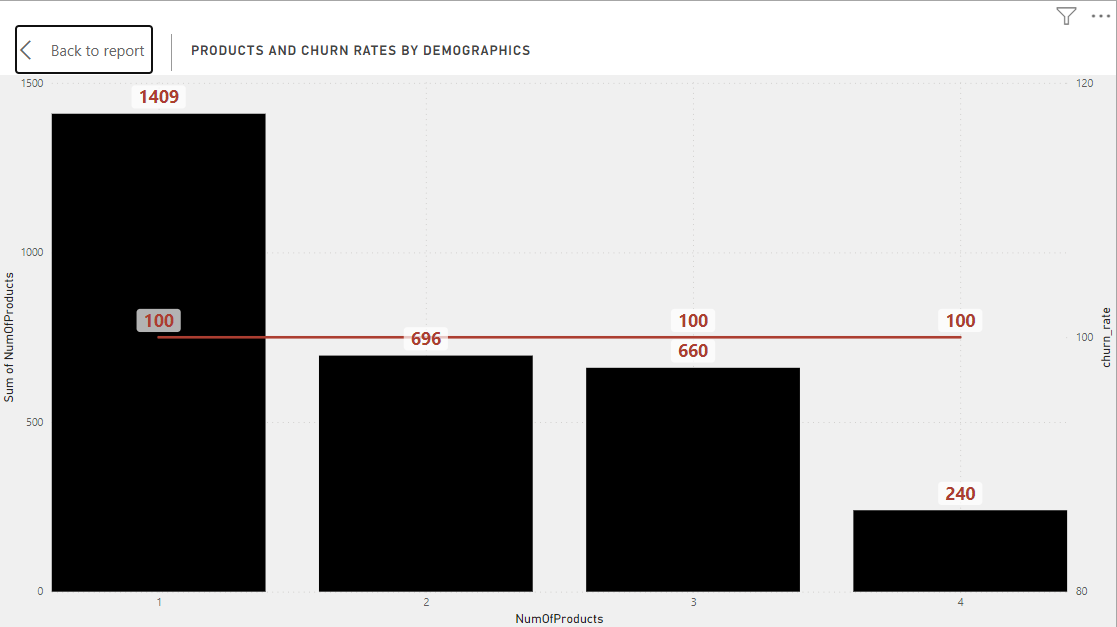




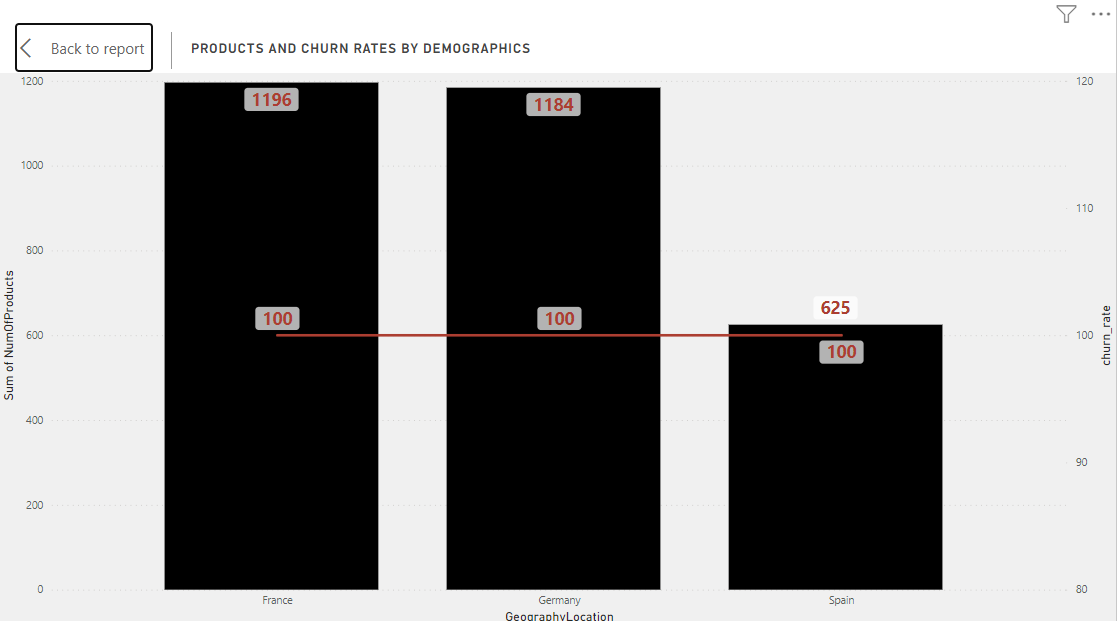
Out of total customers having a Credit card which is 7055, customers who exited only 1424 of those had credit card. This shows that people who don’t have credit cards can leave the bank.



With an increase in tenure we see customers leaving the bank. This shows we are not able to retain our customers for longer tenures. We need to focus on retaining these customers.



By this chart, we can see that with an increase in no. of products, we see customers exit decreases. This shows with the increase in use of our products, customers are more dependent on it and won’t leave the bank easily. We need to increase the no. of products used by customers to retain them.



This graph shows us the customers exit is least from Spain. France and Germany are regions where customers left most. Also no. of customers in germany is already less. This shows we are not able retain our customers in Germany and we need to focus our marketing in that region.

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

Yes, as this are important demographics which shows different aspects of our business, As we have discussed on above questions all this parameters play important role in analysing customer behaviours and can assist in predicting weather or not a customer will leave the bank.

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

**SQL Query**

select Tenure,

NumOfProducts,

case

when CreditScore >= 800 then 'Excellent'

when CreditScore < 800 and CreditScore >= 740 then 'Very Good'

when CreditScore < 740 and CreditScore >= 670 then 'Good'

when CreditScore < 670 and CreditScore >= 580 then 'Fair'

else 'Poor'

end as CreditScoreSegment,

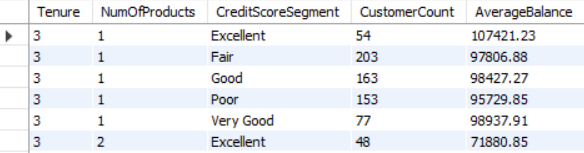
count(customerID) as CustomerCount,

round(avg(Balance), 2) as AverageBalance

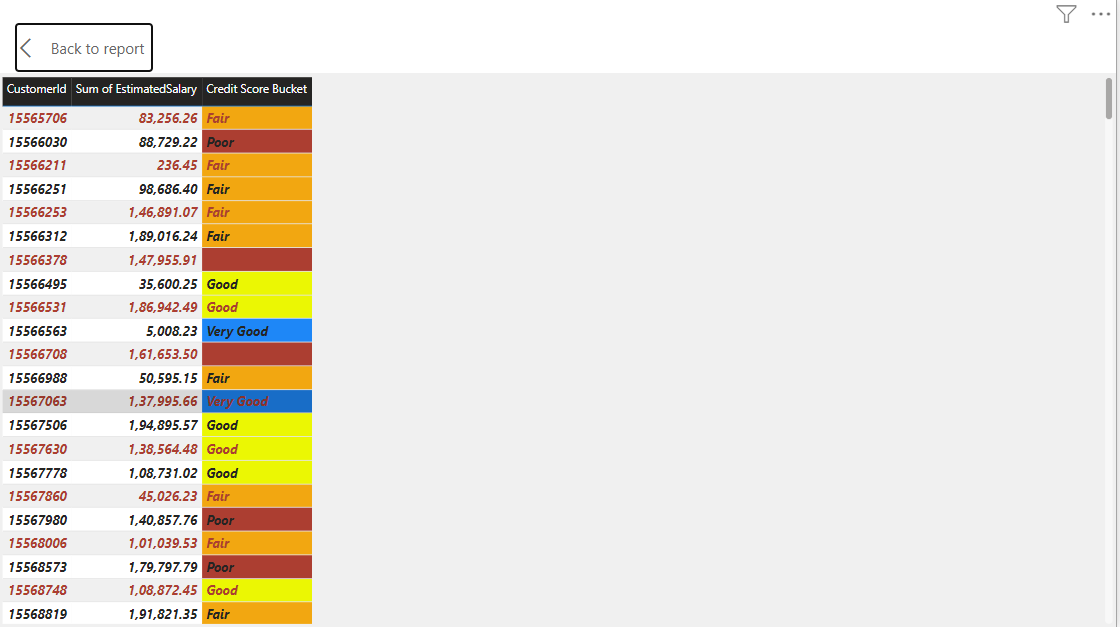
from bank\_churn

group by 1, 2, 3

order by 1;



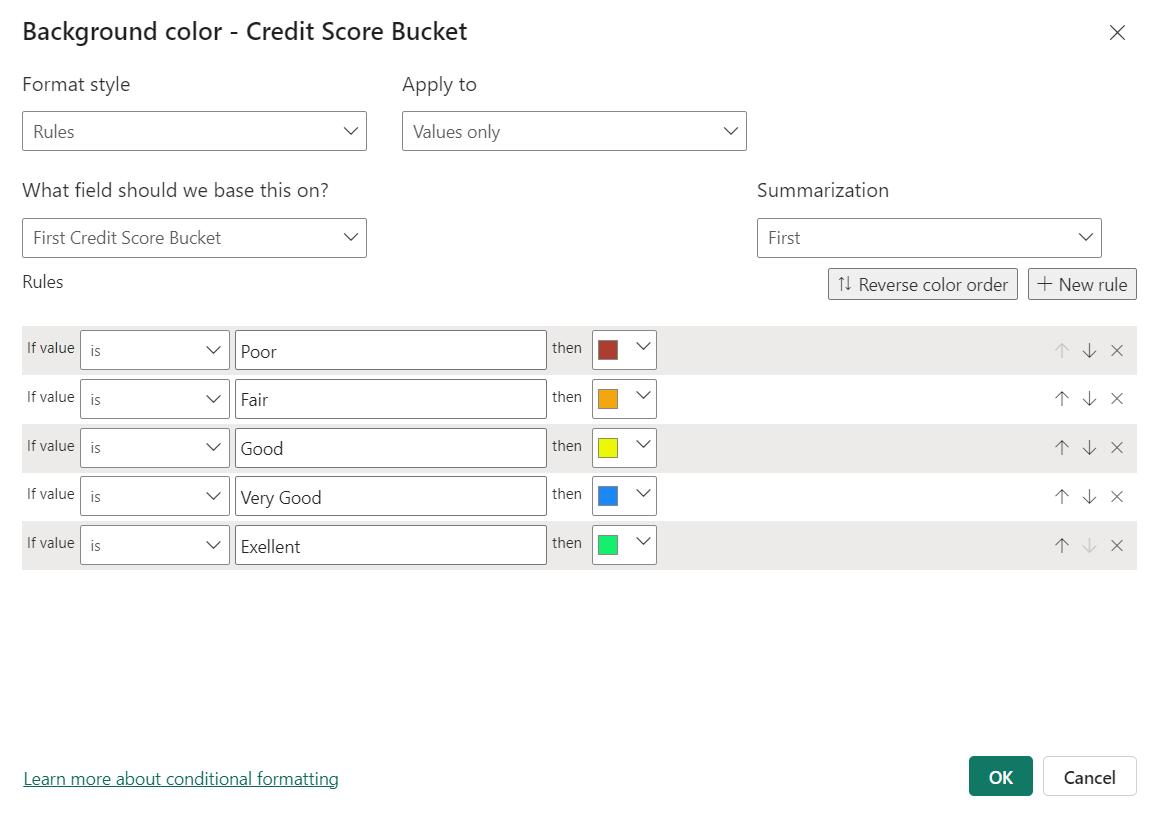
1. How can we create a conditional formatting setup to visually highlight customers at risk of churn and to evaluate the impact of credit card rewards on customer retention?



I have created a table for customer id, balance and credit card segments to highlight customers with bad score segments. Customers with less credit scores are at risk of leaving the bank.

I used the conditional formatting feature to highlight these customers.

Right click on the column > conditional formatting from the visualisation pane > Select Background > put necessary inputs in rule part > click Ok.

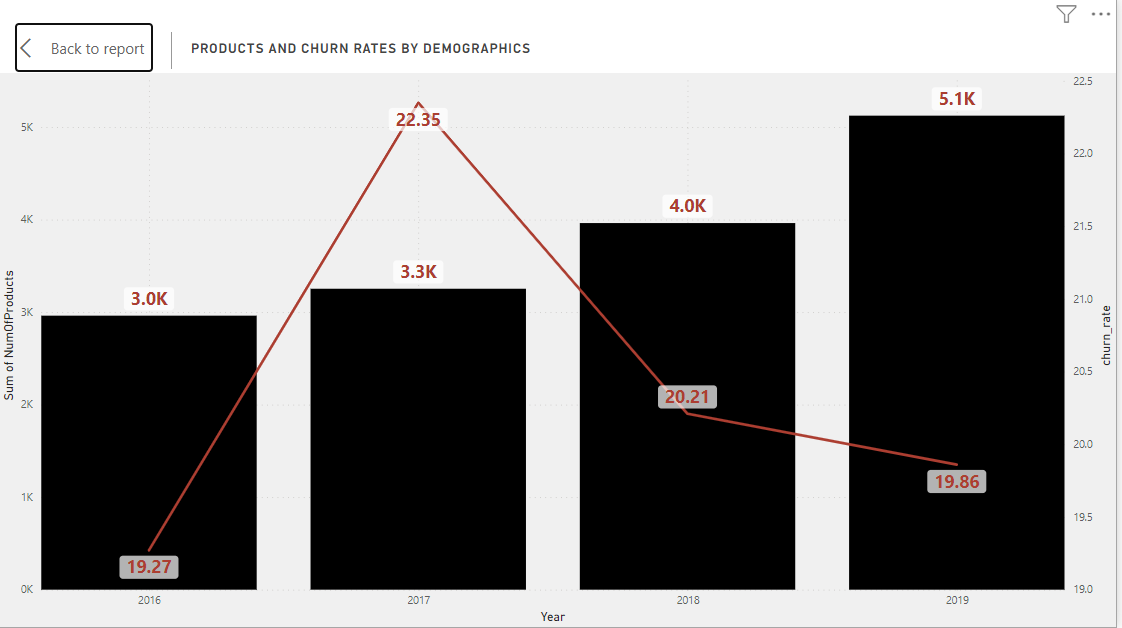


1. What is the current churn rate per year and overall as well in the bank? Can you suggest some insights to the bank about which kind of customers are more likely to churn and what different strategies can be used to decrease the churn rate?

Overall churn rate:



Churn rates by year:



2016: 19.27

2017: 22.35

2018: 20.21

2019: 19.86

Customers, which fall under the (18-30) and (50+) segment by age and customers with Poor, Fair credit score segments are most likely to leave the bank as churn rates are maximum here.

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

I have created a Power bi Dashboard to cater all questions and requirements and added it in the file which I have submitted

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

I would have analysed this problem in almost identical ways, even if the objective and subjective questions.

Objective and subjective questions help in finding a way, based on what we project our analysis.

But if this is not given I would first figure out all important Kpis in data and would build my reports such that I would cover all aspects of Kpis.

I would include all different demographics and how these demographics impact our Kpis.

I would have used suitable slicers which would bifurcate the data based on respective slicers to have more detailed insights.

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

**SQL Query:**

Alter table bank\_churn rename column HasCrCard to Has\_creditcard;