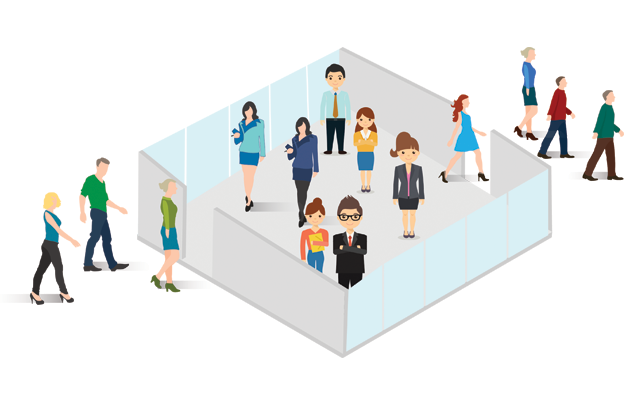
**CUSTOMER CHURN ANALYSIS — Power BI Project**



Customer Churn: Image from Kaggle

**Overview**

What is Customer churn? This is essentially the rate at which customers leave a business against the total customers that are actively in business. This is also known as *churn rate*or *customer attrition*.

**Project Objective:**

The objective of this analysis is to discover various factors contributing to increased customer churn rate at the bank, and provide the business users with these insights which they can use to make informed decisions and strategize on how to improve customer retention and reduce churn rate.

**Tools:**I used Power BI and Power Query for this analysis.

**About the data**

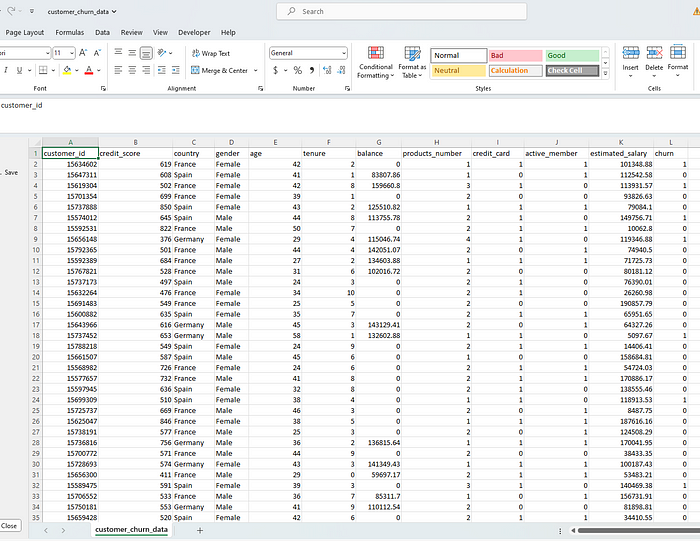
The raw data was provided in the form of a [*CSV file*](https://github.com/technwa/PortfolioProjects/blob/main/customer_churn_data.csv) with 10,000 rows of data. Data was sourced from a commercial bank who was looking to improve its customer churn rate.

***Note***: all personally Identifiable information has been removed from this data to protect customer’s private and sensitive information.

**Data Dictionary**

* **customer\_id**: This is the customers’ unique id which identifies them bankwide.
* **credit\_score:** this is the rating of the customers’ credit worthiness, generated based on the information on their credit report.
* **country**: This is the country the customer hails from.
* **gender**: This is the sex of the customer, grouped into male and female.
* **age**: This refers to the customer’s age.
* **tenure**: The number of years of relationship this customer has maintained with the bank.
* **balance:** The deductible balance on their bank account as at time this data was generated.
* **products\_number:** This shows the bank account type the customer has.
* **credit\_card:** This column shows if a customer has a credit card or not, represented as either 1 or 0 respectively.
* **active\_member:** These numerical values show whether an account is active or not, represented as either 1 or 0 respectively.
* **estimated\_ salary**: This is the average annual salary received into the account.
* **churn:** This field shows if the customer is still with the bank or not, represented by either 0 or 1 respectively.

**Data Discovery**



Raw data

I opened the data in excel and inspected the columns for relevance to the business task (This can equally be done directly on Power Query), I was able to confirm that there is sufficient data to aid my analysis. However, data needed to be cleaned and properly transformed in the format that would be useful for the analysis.

**Limitations**: The data provided by the business user is only that of salary-earning customers.

**Analytical Steps**

1. Data Preparation
2. Data Categorization and Grouping
3. Formatting
4. Data Transformation
5. Data Modeling
6. Visualisation — Report
7. Insights
8. Recommendation
9. Conclusion

**Note**: Step 2- 4 are iterative as I had to repeat them as needed throughout this activity.

1. **Data Preparation**

I imported the data into Power BI and transformed it on Power Query as follows:

* I renamed the data from *customer\_churn\_data* to *Customer Data*for clarity. The new name better captures what the content of this file is. .
* Next, I eliminated the auto-generated header, by promoting the first row of my data to become the header. The option for this can be found on both the home tab and the transform tab of Power Query Editor.
* With the project objective in mind, I sought to eliminate any data that would be irrelevant to my analysis. I removed the *estimated\_salary* column, since they are estimates anyways and there is another column with account balance information for each customer and that would be more useful for analysis. To remove the column, I right-clicked on the column header and clicked on*remove colum*.

**Note**: Power Query provides an *applied steps* pane, where any change made on the editor is automatically recorded. To customise this feature which greatly aided my documentation, I renamed each step to better describe the action I took. To rename, simply right-click any of the steps and select *rename.*

**2. Data categorization and Grouping.**

* In order to clearly label the data in the *products\_number* column, I utilized the *replace values* option. Since it indicates the account type a customer owns, I replaced product 1,2,3, and 4, with Prod 1, Prod 2, Prod 3, and Prod 4 respectively. I followed it up with deleting the original column.
* For the *age*, *credit\_score, balance*columns, I looked at the column profile from the *View*tab, and saw that there was a wide range between the minimum and maximum values.
* The *Age* column had 60 distinct value with a minimum of 18 and a maximum of 82
* The *credit\_score*column had 354 distinct values with a minimum of 376 and a maximum of 850.
* The **balance**column had 649 distinct values. With a minimum of 0 and a maximum of 213146.2

I then grouped each of them using *conditional column* function.

**3. Formatting**

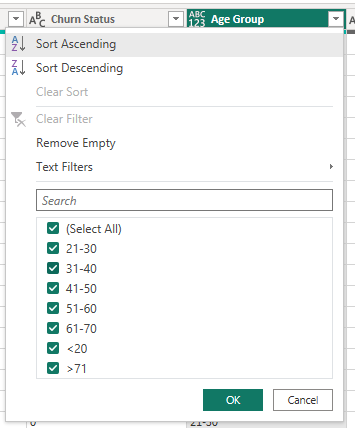
The following columns; *churn*, *active\_member* and *credit\_card* had boolean datatypes of 1 and 0, essentially representing true or false. However, this lacks clarity and would be confusing on the report. I formatted them by using the *replace values* option to change them to more intuitive data types.

* For active\_member column, 0 = dormant, 1 = active
* For churn column, 0 = not churned, 1 = churned
* For credit\_card column, 0 = no, 1 = yes

**4. Data Transformation**

Here, I created Queries to organise data for better analysis and reporting.

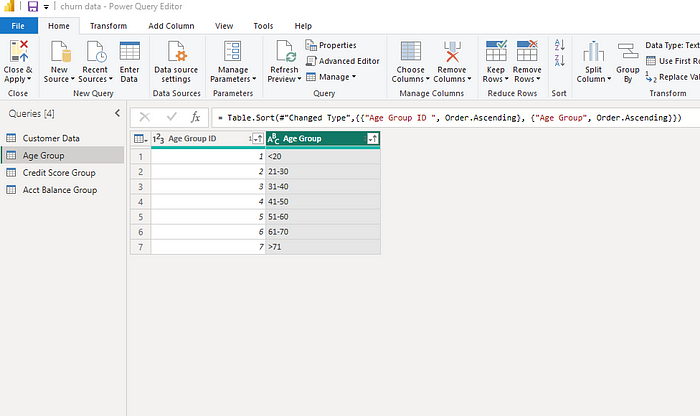
Because I had grouped some of these columns used symbols like <, >=, etc, and changed the groups to them to text data type, I was not able to filter it successfully in ascending or descending order. This is because the system would place all the groups with numbers only first, before the ones with symbols as shown below.



Before custom Sorting

In order to get them in chronological order, I created a reference table from the original one. In this new table, I equally created an index column using the *Conditional column*option to represent each unique entry in order, and allow me to filter them chronologically.

* To achieve this, I started by selecting the column of interest. On the queries pane, I right-clicked on the data and selected the *reference* option which created a new table. I removed all other columns except the one I wanted to index, then renamed this new table to reflect the column accordingly.
* Next, I removed duplicate rows leaving behind only the distinct values that I would index. Since it was a custom sorting, I selected the *conditional column* option and entered the conditions as desired. I then changed the data type of the index columns to whole numbers.



After Custom Sorting

**Note:** The above steps were repeated for the *balance*and *credit\_score* columns.

Next, I renamed the columns where needed, so they are meaningful enough when they appear as title in the charts.

active\_member = Activity Status

customer\_id = Customer ID

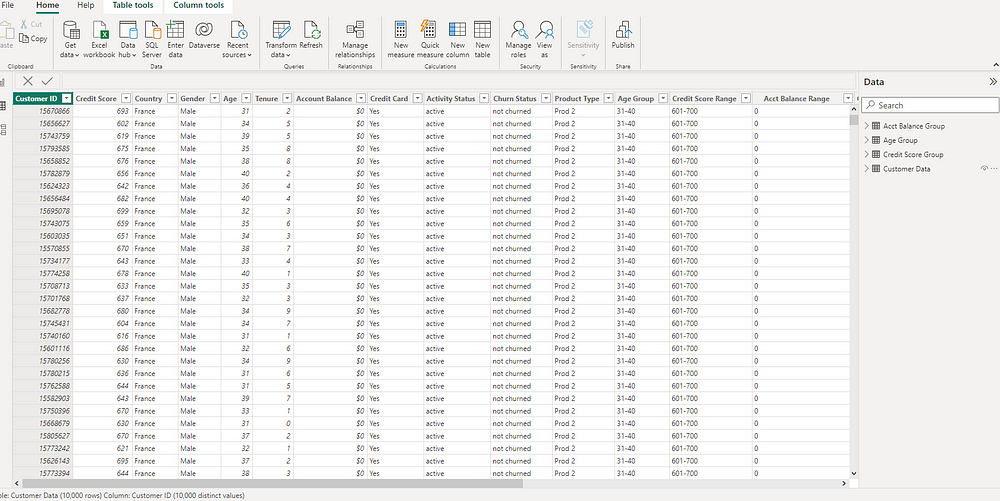
churn = Churn Status

balance = Account Balance

products\_number = Product Type

I also ensured all the columns had the right data type by checking the type indicated at the left hand side of the column name, and making changes where necessary.

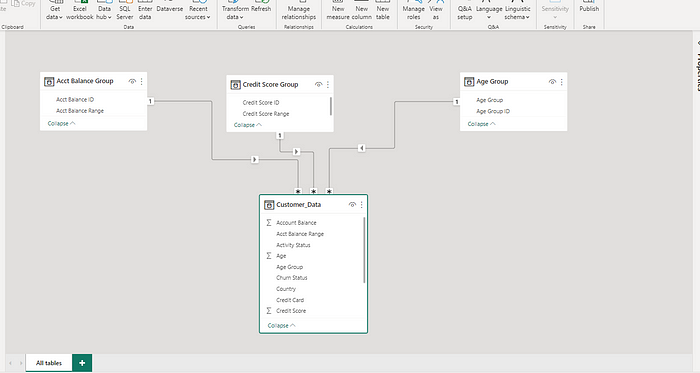
In the end, I had 4 tables for this report. I clicked on *close and apply* which loaded the cleaned data (see below) onto the Power BI window.



cleaned data

5. **Data Modelling**

Here, I explored the relationship between these tables and how they are connected. I also arranged them so that all the *fact tables* were at the top of the *dimension table*. This is to make visualisation easy. Power BI pre-detected the cardinality of the model relationship as many-to-one which is correct.



table

6. **Visualization — Report**

I started the report by creating a few DAX measures.

***DAX*** simply means Data Analysis Expression. It is an expression that helps with calculations while building report. The three I created were:

* Customers(Number of customers)
* Churn rate (%)
* Customers Churned

**Step**:

On the right pane where the tables are listed, I right clicked on the *Customer Data* table, then selected *new measure*. I labeled the measure as *Customers* and used the formular below.

Customers = DISTINCTCOUNT('Customer Data'[Customer ID])

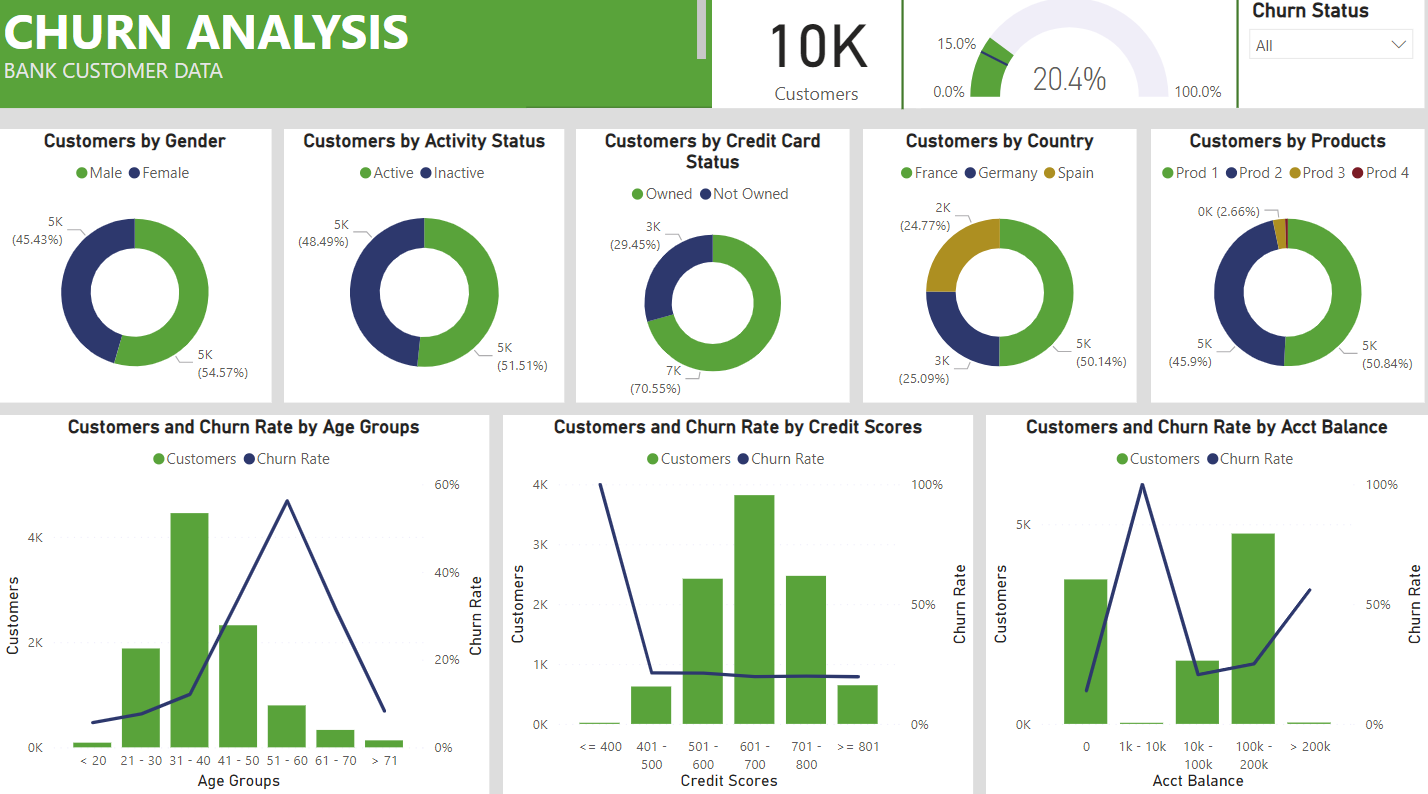
This formula above counted all the unique ID in the *Customer ID* column.

I repeated the above steps for the*Customers Churned*and *Churn Rate* measure with the respective formulars below:

Customers Churned = CALCULATE(COUNT('Customer Data'[Churn Status]), 'Customer Data'[Churn Status] = "churned")  
  
Churn Rate = 'Customer Data'[Customers Churned]/'Customer Data'[Customers]

1. Customers: 10K
2. Customers churned: 2037
3. Churn Rate: 20.4% ( Changed to % from the measures tool tab)

Finally, I built the report in the screenshot below, taking into account the key factors like account balance, age group, activity status, gender, credit score, e.t.c. and their effects on the churn rate. This helped me gain insights into where the bank could make neccessary improvements.



7. **Insights**

1. Churn rate for male customers is 65% but the overall male churn rate is at 25%.
2. In terms of those with credit card facility, the churn rate is highest amongst lowest credit score group of >400
3. Surprisingly, the customers whose account balance are <=200k are the most churned in terms of account balance.
4. Overall, Churn rate is very high at 56.2% for middle aged customers between 51–60 years of age.
5. For the products, churn rate is high @ 27.7% amongst customers in the Prod 1 group. The age account and credit score factors are same here. Similar situation for prod 2 =, however in prod 2, the churn rate is at all time 100% for customers between the 1k-1ok account balance.

7. **Recommendation**

1. The stakeholders could consider creating products that target seniors close to their early retirement age to avoid losing them.
2. An incentive program could be considered for customers who have maintained longer relationship with the bank.
3. In the same way, an exclusive package such as travel and vacation packages, subsidized investment portfolio, e.t.c., could be considered for the customers in the >200k account balance group, to reduce the rate at which they leave the bank.
4. The business needs to understand why there is 100% churn rate for Prod 2 customers with account balance of 1k-10k and how to get them back in business.

In Summary, age group of 51–60, people with low credit scores and those with high account balance (>200k) are most likely to churn. The bank should consider the recommendations above and other effective strategies that can address the findings, to better customer retention and reduce chrun rate.

8. **Conclusion**

In this Power BI analysis, I went through some major data analysis process to get the data from raw to cleaned and onto building the report, ensuring data quality across board. Most importantly, it provided valuable insights into some of the major factors that encourage customer churn at the bank.

With this report, the stakeholders can better manage the churn situation (specifically for salaried customers), by focusing on the key factors and evaluating the results of their changes periodically.