**Bank Customer Churn Dataset Documentation**

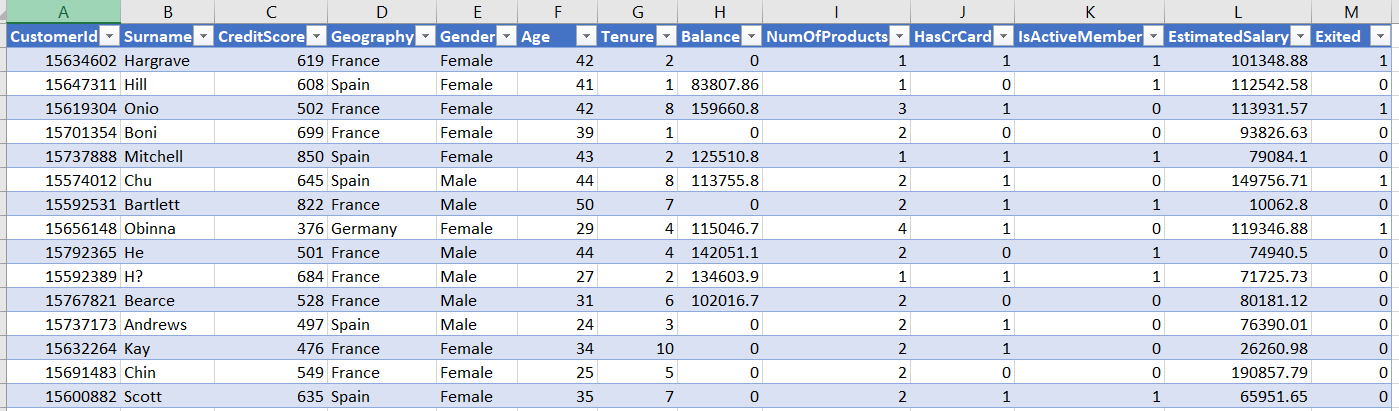
**Dataset Overview**

The Bank Customer Churn dataset contains data on bank customers and their relationship with the bank, including demographic information, account balance, services used, and whether they have churned . The goal of analyzing this dataset is to understand customer behavior and identify patterns that may lead to customer churn, enabling the bank to take proactive actions to reduce churn and improve customer retention.

**Dataset Structure**

The dataset contains the following columns:

| **Column Name** | **Description** | **Data Type** |
| --- | --- | --- |
| **CustomerID** | Unique identifier for each customer. | Categorical |
| **Surname** | Last name of the customer. | Categorical |
| **CreditScore** | Customer's credit score, a numeric measure of creditworthiness. | Numeric |
| **Geography** | The country or region where the customer resides (e.g., France, Germany, Spain). | Categorical |
| **Gender** | Customer’s gender (Male, Female). | Categorical |
| **Age** | Customer’s age. | Numeric |
| **Tenure** | Number of years the customer has been with the bank. | Numeric |
| **Balance** | The customer’s account balance. | Numeric |
| **NumOfProducts** | Number of bank products that the customer is using (e.g., 1, 2, 3, or 4 products). | Numeric |
| **HasCrCard** | Whether the customer has a credit card (1 = Yes, 0 = No). | Categorical |
| **IsActiveMember** | Whether the customer is an active member (1 = Active, 0 = Inactive). | Categorical |
| **EstimatedSalary** | The customer’s estimated yearly salary. | Numeric |
| **Exited** | Whether the customer has churned (1 = Yes, 0 = No). | Categorical |



### ****Key Metrics and Concepts****

* **Customer Churn**: Refers to customers who have left the bank, indicated by the **Exited** column. Understanding churn is a primary focus of this dataset.
* **Credit Score**: A numeric representation of a customer's creditworthiness, often influencing their likelihood of staying with or leaving the bank.
* **Tenure**: The number of years the customer has been with the bank. Customers with longer tenures might have different behaviors compared to newer customers.
* **Balance**: The total balance in a customer's account. Customers with high balances might behave differently than those with lower balances.
* **Active Member**: Whether the customer is still actively using the bank’s services. This feature may correlate with customer engagement and churn.

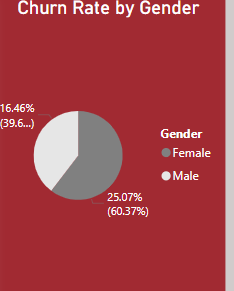
### ****Objective of the Analysis****

The primary objective of analyzing this dataset is to understand the factors contributing to customer churn and identify actionable insights that the bank can use to reduce churn. Some of the key questions to explore include:

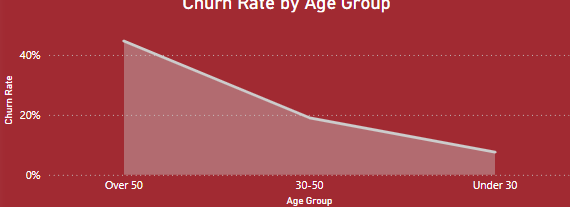
1. What is the overall customer churn rate?



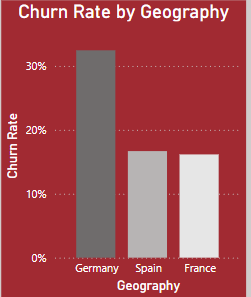
1. Which gender has a higher churn rate?



1. How does customer age affect churn? Over 50 is the highest churn rate



1. Which regions or cities have the highest churn rate? Germany



### ****Key Data Analysis Questions****

The following questions can guide your analysis of the dataset:

1. **What is the overall churn rate?**

20.37%

1. **Which gender has a higher churn rate?**

**Female**

1. **How does customer age affect churn?**

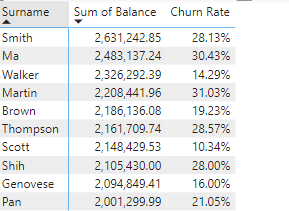
Over 50 is the highest churn rate

1. **Does credit score have an impact on churn?**

Customers with a credit score 350 might churn rate 100%

1. **Do customers with higher balances churn less?**

**NO**



1. **Which regions or cities have the highest churn rate? Germany**

### ****Data Cleaning and Preprocessing Steps****

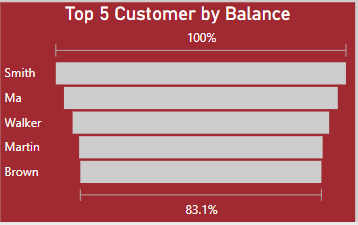
Before performing analysis on the dataset, it is essential to preprocess and clean the data for any inconsistencies or missing values.

1. **Handling Missing Values**: Ensure that all missing values are addressed. Depending on the column, you may choose to remove rows with missing values or fill them with appropriate values (e.g., mean or median for numeric columns).
2. **Outliers Detection**: Check for outliers in numeric columns (such as **CreditScore**, **Balance**, and **Age**) that may distort the analysis. Use appropriate techniques such as box plots for outlier detection.
3. **Encoding Categorical Variables**: Convert categorical columns like **Gender**, **Geography**, and **IsActiveMember** into numerical format if necessary (e.g., one-hot encoding).
4. **Standardizing Data**: Standardize numerical columns like **CreditScore**, **Balance**, and **EstimatedSalary** to bring them to a comparable scale if using machine learning algorithms.

### ****ample Analysis Visualizations****

You can use various charts and visualizations to explore and analyze the data:

1. **Churn Rate by Geography** (Bar Chart): Identify regions with the highest churn rate.
2. **Churn by Gender** (Pie Chart): Visualize the gender breakdown of churned customers.
3. **Customer Age Distribution** (Histogram): Show the age distribution of churned vs. active customers.
4. **Tenure vs. Churn** (line chart): Explore the relationship between customer tenure and churn rate.
5. **Churn Rate by age group** (Line Chart): Track churn rates by age group
6. **top 5 customer by balance**



### ****Slicers for Dynamic Filtering****

Add **slicers** to enable dynamic filtering and enhance your analysis:

* **Gender Slicer**: Filter visualizations by gender (Male/Female).
* **Active Member Slicer**: Filter based on whether customers are active members or not.
* **Age group** : Filter visualizations by Age group( over 50 ,30-50,under 30)

### ****Conclusion****

This dataset offers valuable insights into customer behavior and the factors contributing to churn. By analyzing this data, the bank can develop strategies to reduce churn, improve customer retention, and increase overall customer satisfaction