Customer Churn Prediction

Customer's Churn Analysis and Modeling is a comprehensive project focused on understanding and predicting customer churn in the Banking industry. Utilizing advanced data analysis and machine learning techniques, this project aims to provide insights into customer behavior and help develop effective strategies for customer.

**Problem Statement**

Customer churn or customer attrition is a tendency of clients or customers to abandon a brand and stop being a paying client of a particular business or organization. The percentage of customers that discontinue using a company’s services or products during a specific period is called a customer churn rate. Several bad experiences (or just one) are enough, and a customer may quit. And if a large chunk of unsatisfied customers churn at a time interval, both material losses and damage to reputation would be enormous.

A reputed bank “ABC BANK” wants to predict the Churn rate. Create a model by using different machine learning approaches that can predict the best result.

## Dataset Description:

The details about all columns are given in the following data dictionary.

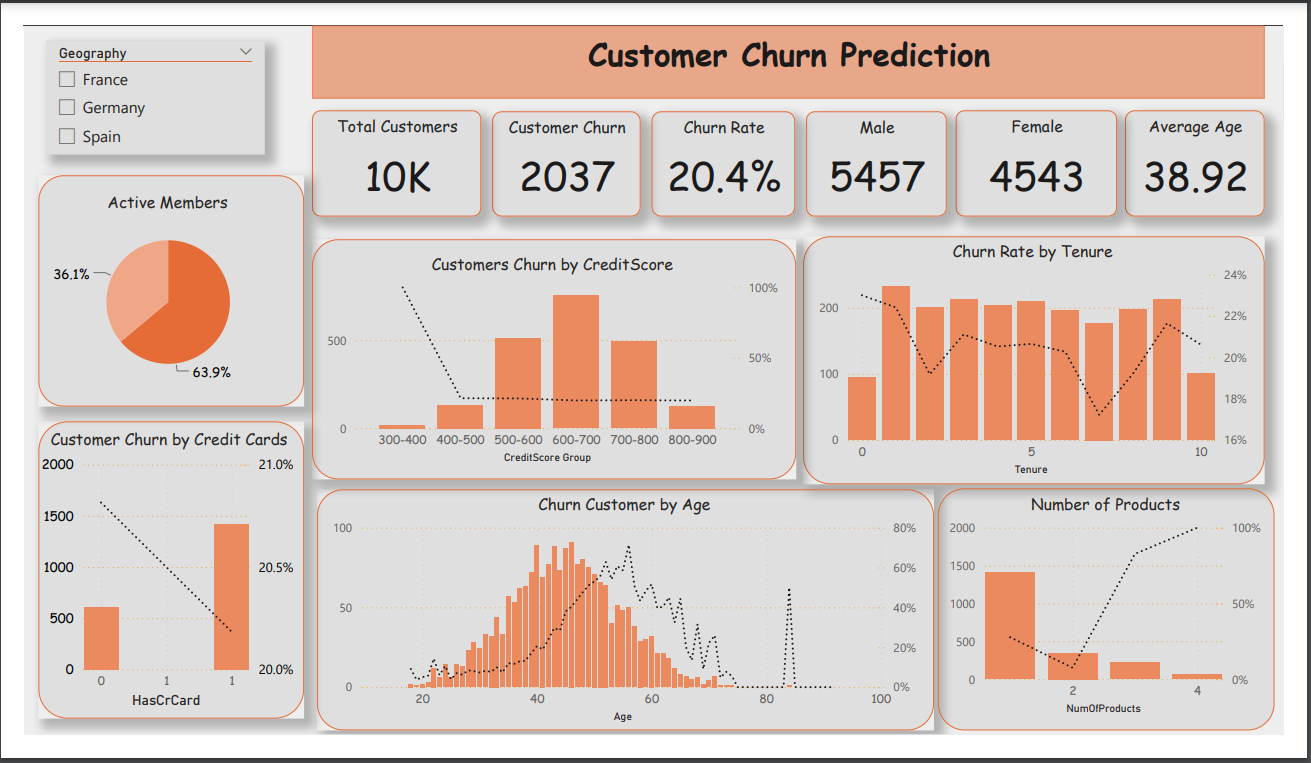
|  |  |
| --- | --- |
| Variable | Definition |
| RowNumber | Unique Row Number |
| CustomerId | Unique Customer Id |
| Surname | Surname of a Customer |
| CreditScore | Credit Score of each Customer |
| Geography | Geographical Location of Customers |
| City\_Category | Category of the City |
| Gender | Sex of Customers |
| Age | Age of each Customer |
| Tenure | Number of years as member |
| Balance | Current Balance of Customers |
| NumofProducts | Number of Products Customers are using |
| HasCreditCard | If a customer has credit card |
| IsActiveMember | If a customer is active member or not |
| EstimatedSalary | Estimated Salary of each Customer |
| Exit | Customer left company or not |

## Working Flow:

In order to create a model these are the following procedure -

* Split the dataset in 70% of Train set and 30% of Test Set
* Feature engineering
* Check the accuracy score for both Training and Test Set
* Compare the accuracies for both Training and Test set, in order to check for the overfitting issues

Dashboard



Firstly, I took my time familiarizing myself with the dataset. I checked the shape of the dataset. What column headers I have, what datatypes the columns hold, what values are present in each column. Then I checked number of unique customers, and calculated churn rate, average number of product customer have with bank, average age of these customers, number of male and female customer to provide more context.

After getting summary metrics, I decided to dig deeper by looking at churn rates across age and tenure. I could instantly see the people between age 45 and 60 tend to leave a bank at a much higher rate than other age groups. I could also see that tenure, the years the customers have been using the bank for, doesn't impact the churn rate as they are relatively constant across all tenures. Then I tried to explore where there is any geographical differences based on the country column. The higher churn rate in Germany stood out immediately basically double out of France and Spain. Active Customers are tend to leave to bank less than in-active members.

Finally, I analyzed churn rate across number of products the customer holds, has any credit card and points earned by customers. Finding here were that the customers with three of four products left more often. Maybe they didn't like the fact that they have been sold so many products. In term of churn rate by points earned by customers, I could see customers with very little points left the bank more often with the rest of them churning at a pretty even rate. I built a story in Power BI to communicate these findings.