# Customer-Churn-Prediction-using-classification-model

ASWIN.G.KUMAR

# ***Introduction***

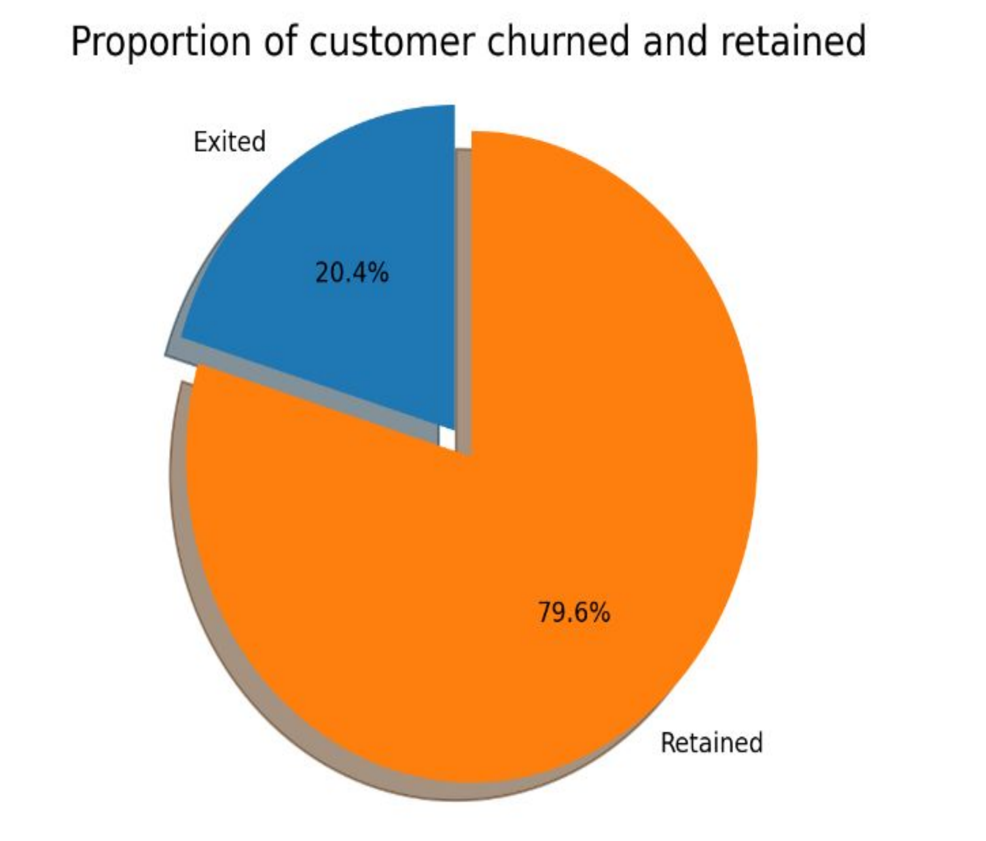
Customer churn, the phenomenon where customers discontinue their relationship with a service provider, is a critical issue for businesses, particularly in competitive industries like banking. Predicting whether a customer will churn enables companies to proactively manage customer retention by identifying at-risk customers and taking timely measures to retain them. This case study focuses on predicting customer churn based on their usage patterns, demographic factors, and interactions with customer service.

The dataset used for this project is sourced from Kaggle and consists of bank customer information, including features such as age, balance, tenure, product usage, and credit scores. The goal is to build a machine learning model to predict whether a customer will leave or stay with the bank. Tools like Scikit-learn for machine learning, Pandas for data manipulation, and Matplotlib for visualization will be utilized in this analysis to gain insights and improve the bank's customer retention strategies.

# **We aim to accomplist the following for this study:** **1. Identify and visualize which factors contribute to customer churn:**

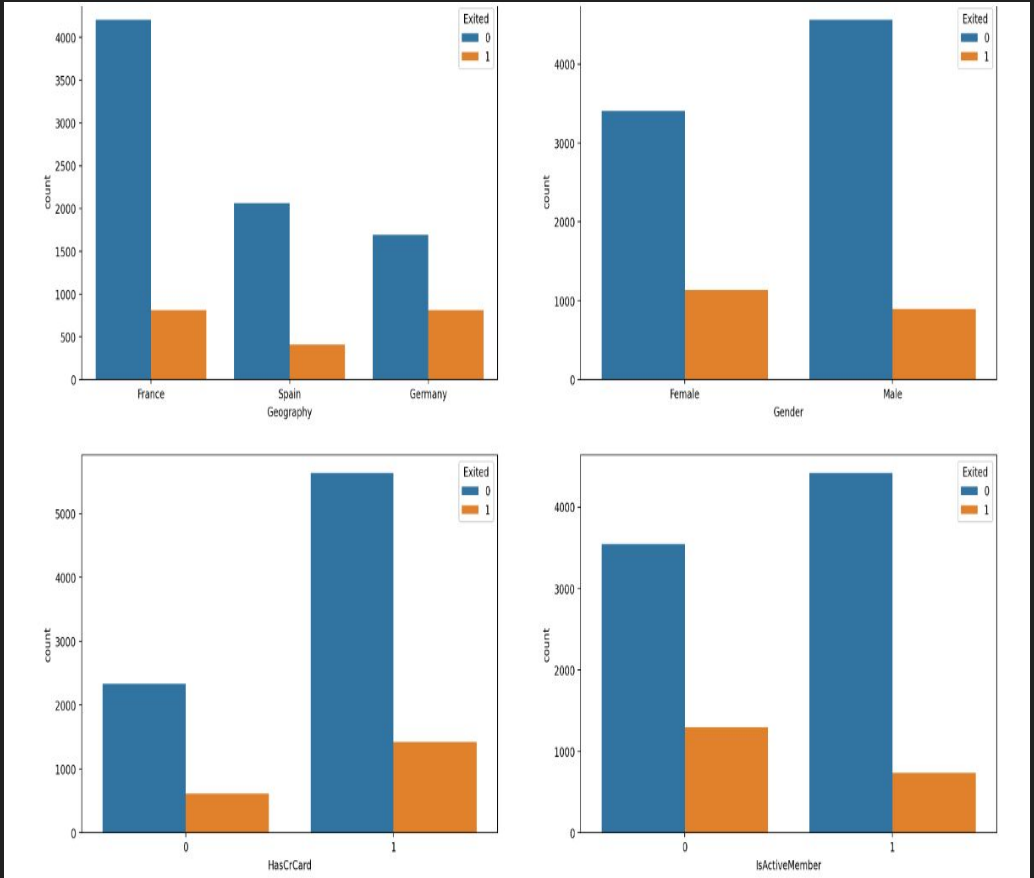
**2. Build a prediction model that will perform the following:**
**- Classify if a customer is going to churn or not**
**- Preferably and based on model performance, choose a model that will attach a probability to the churn to make it easier for customer service to target low hanging fruits in their efforts to prevent churn**

Here our main interest is to get an understanding as to how the given attributes relate to the 'Exit' status




- **So about 20% of the customers have churned. So the baseline model could be to predict that 20% of the customers will churn. Given 20% is a small number, we need to ensure that the chosen model does predict with great accuracy this 20% as it is of interest to the bank to identify and keep this bunch as opposed to accurately predicting the customers that are retained.**

# Here we review the 'Status' relation with categorical variables



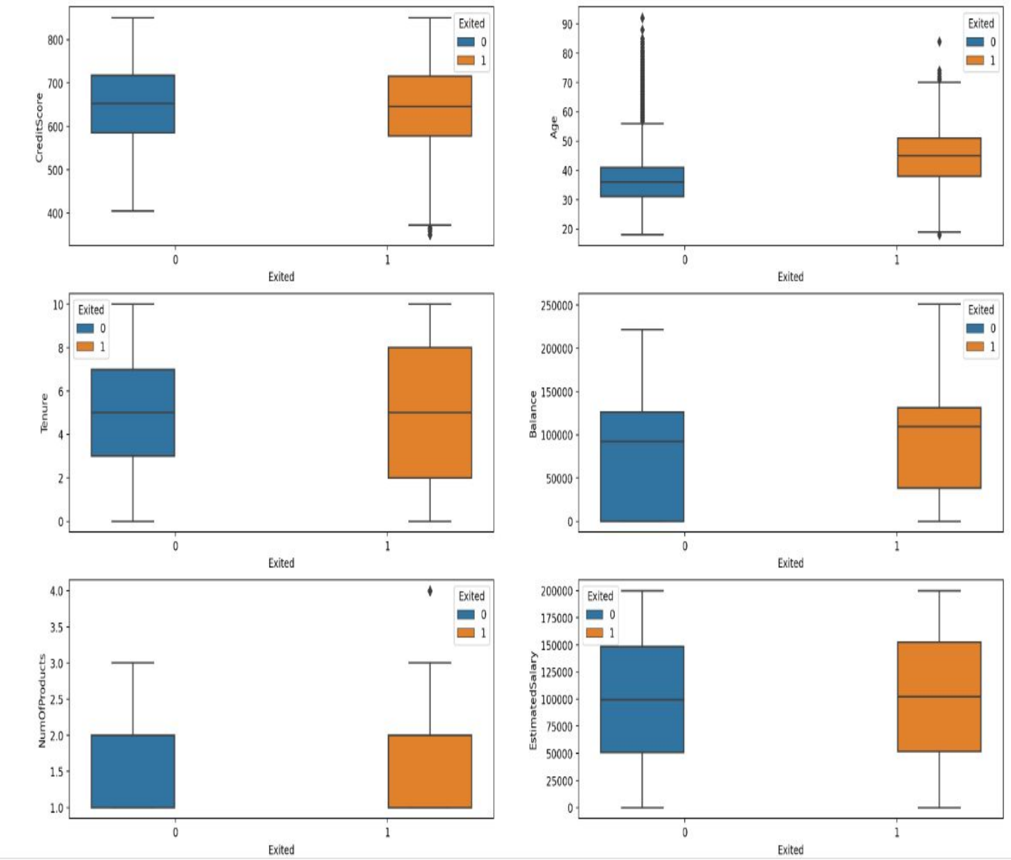

- Majority of the data is from persons from France. However, the proportion of churned customers is with inversely related to the population of customers alluding to the bank possibly having a problem (maybe not enough customer service resources allocated) in the areas where it has fewer clients.

- The proportion of female customers churning is also greater than that of male customers

- Interestingly, majority of the customers that churned are those with credit cards. Given that majority of the customers have credit cards could prove this to be just a coincidence

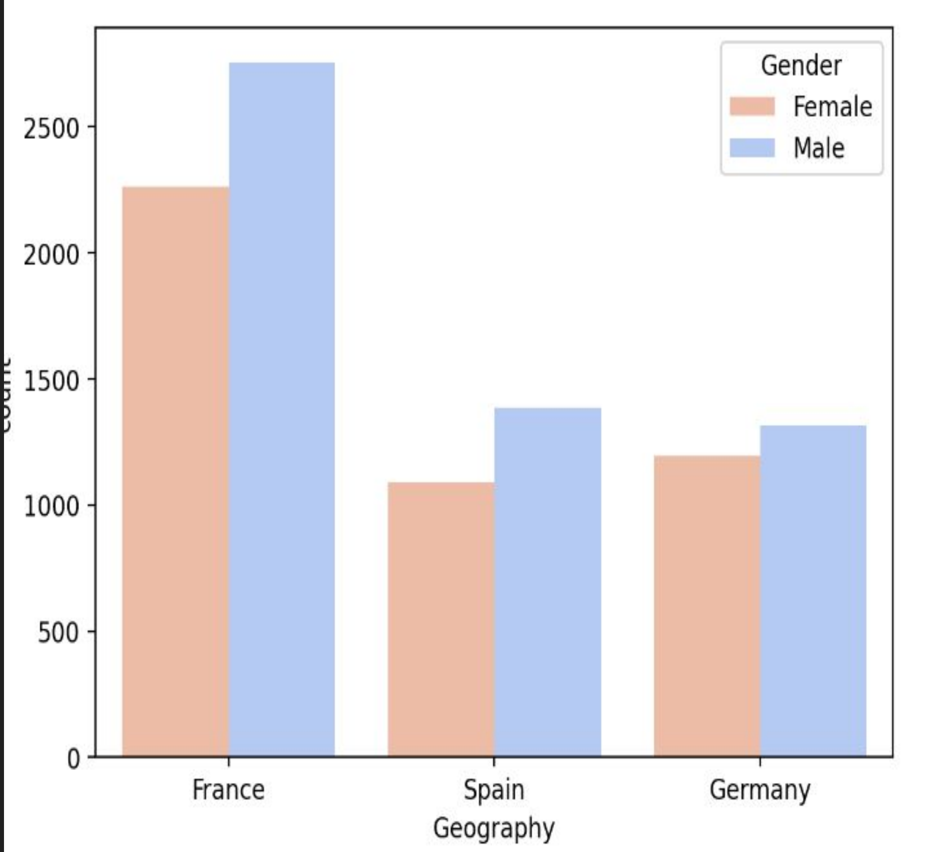
- Unsurprisingly the inactive members have a greater churn. Worryingly is that the overall proportion of inactive mebers is quite high suggesting that the bank may need a program implemented to turn this group to active customers as this will definately have a positive impact on the customer churn.

# Relations based on the continuous data attributes




- This code generates a set of box plots to visualize the relationship between churn (Exited) and continuous features in your dataset. The goal is to explore how features like `Age`, `Balance`, etc., are distributed differently for customers who churned vs. those who didn't, which will help you understand what factors might contribute to churn.
- These visualizations help identify which features are more correlated with churn and might be useful in predicting customer behavior.

**This count plot that visualizes the distribution of customers across different geographical locations (France, Spain, Germany) based on their gender**




- **The plot indicates that, across all three countries, there are generally more male customers than female customers.**
- **The differences in counts across the countries suggest varying gender distribution patterns in the customer base.**
- **This count plot effectively visualizes the distribution of customers by gender across three geographical areas. It shows that males are generally more represented than females in each country, with varying levels of difference. This data can be essential for understanding customer demographics and guiding business strategies accordingly.**

# Conclusion

***The customer churn prediction model built using Random Forest Classifier successfully classifies customers who are likely to churn based on key features such as usage patterns, demographics, and customer service interactions. By leveraging this predictive model, businesses can better understand the drivers of churn and take preemptive actions to retain high-risk customers. The insights derived from the analysis empower companies to design targeted retention strategies, thus reducing churn rates and enhancing overall customer satisfaction. The Random Forest model proved effective in handling complex datasets and delivering valuable results in this churn prediction task.***

** Data set review & preparation
In this section we will seek to explore the structure of this data:

- To understand the input space the data set
- And to prepare the sets for exploratory and prediction tasks as described