Euro Bank – Exploratory Data Analysis

17th August 2021

Table of Contents

- Engagement Background
- Data Science Process
- Identify Business Problem – Project Charter
- Explore Data – Key Components of EDA Considered
- Explore Data – Understand the data
- Explore Data – Proportion of Customers Churned
- Explore Data – Customer Age wise Balance distribution
- Explore Data – Categorical Variables Vs Churn
- Categorical Variables Vs Churn – Findings & Recommendations
- Explore Data – Continuous Variables Vs Churn
- Continuous Variables Vs Churn – Findings & Recommendations
- Explore Data – Continuous Variables Vs Churn (Additional Plots)
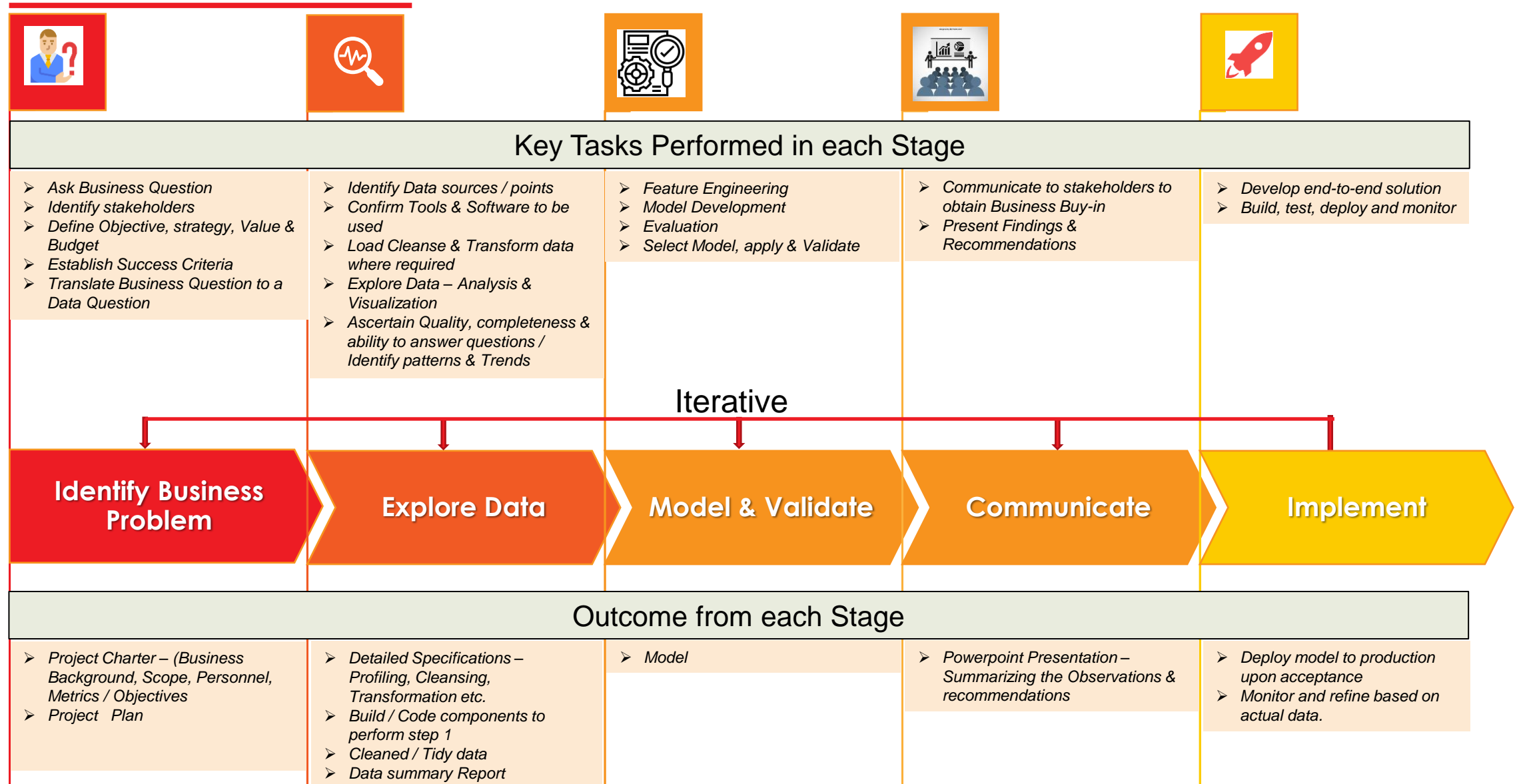- Wrap-up

Engagement Background

Euro Bank, a leading European Bank is facing an issue of Customer churn and would like to engage the services of Sai Science Pte Ltd, a Data Science Vendor firm to analyze a representative set of data and provide insights, inferences and recommendations.

Given the competitive market, the Bank wants things to move swiftly in a progressive manner and hence the Vendor firm has proposed a phased approach wherein the initial discovery phase would kick-off once the data set is made available and will last for 2 weeks.

As part of the discovery phase, the Vendor firm would do an **exploratory data analysis** on the data set provided using Python and would present the findings along with some potential recommendations in 2 weeks time.

Data Science Process

Key Tasks Performed in each Stage

Identify Business Problem	Explore Data	Model & Validate	Communicate	Implement
➤ Ask Business Question ➤ Identify stakeholders ➤ Define Objective, strategy, Value & Budget ➤ Establish Success Criteria ➤ Translate Business Question to a Data Question	➤ Identify Data sources / points ➤ Confirm Tools & Software to be used ➤ Load Cleanse & Transform data where required ➤ Explore Data – Analysis & Visualization ➤ Ascertain Quality, completeness & ability to answer questions / Identify patterns & Trends	➤ Feature Engineering ➤ Model Development ➤ Evaluation ➤ Select Model, apply & Validate	➤ Communicate to stakeholders to obtain Business Buy-in ➤ Present Findings & Recommendations	➤ Develop end-to-end solution ➤ Build, test, deploy and monitor

Iterative

Outcome from each Stage

Identify Business Problem	Explore Data	Model & Validate	Communicate	Implement
➤ Project Charter – (Business Background, Scope, Personnel, Metrics / Objectives ➤ Project Plan	➤ Detailed Specifications – Profiling, Cleansing, Transformation etc. ➤ Build / Code components to perform step 1 ➤ Cleaned / Tidy data ➤ Data summary Report	➤ Model	➤ Powerpoint Presentation – Summarizing the Observations & recommendations	➤ Deploy model to production upon acceptance ➤ Monitor and refine based on actual data.

Identify Business Problem – Project Charter

Customer: _Euro Bank, Leading European Bank_

Business Problem / Statement: Bank has noticed increased number of customers leaving the bank. To tackle this alarming situation, Bank has decided to engage the services of our company to Explore data as part of discovery phase and present the findings. 6 Months of Data pertaining to close to 10000 customers are selected randomly across three countries - France, Germany , Spain and has been shared. The bank wants to understand and get the insights and patterns around customer churning so that they can devise / adapt new policies and change existing processes if required

Stakeholders:

Key Client Stakeholders
Client Engagement Director
Client Project Manager
Client BA / SME
Business Sponsor - Head of Business - Retail & Consumer Banking
Technology Sponsor - Head of Technology - Retail & Consumer Banking

Vendor Stakeholders
Engagement Director
Project Manager
Lead Data Scientist / Solution Director
Solution Architect
Data Architect
Developers

Business Objectives / Outcomes expected:
- ✓ _To identify data insights that could help to reduce the Customer Churn_
- ✓ _Based on the data provided, perform Exploratory Analysis & Visualization to_
- ✓ _Handle Missing, Duplicate, Null records & perform any transformations as required_
- ✓ _Identify Relationship (if any) between Categorical Variables like Geography, Gender, Is Active Member, Has Credit Card on the Customer Churn_
- ✓ _Identify Relationship (if any) between Numerical variables like Age, Credit Score, Balances, Tenure, Number of Products , Estimated Salary on the Customer Churn_
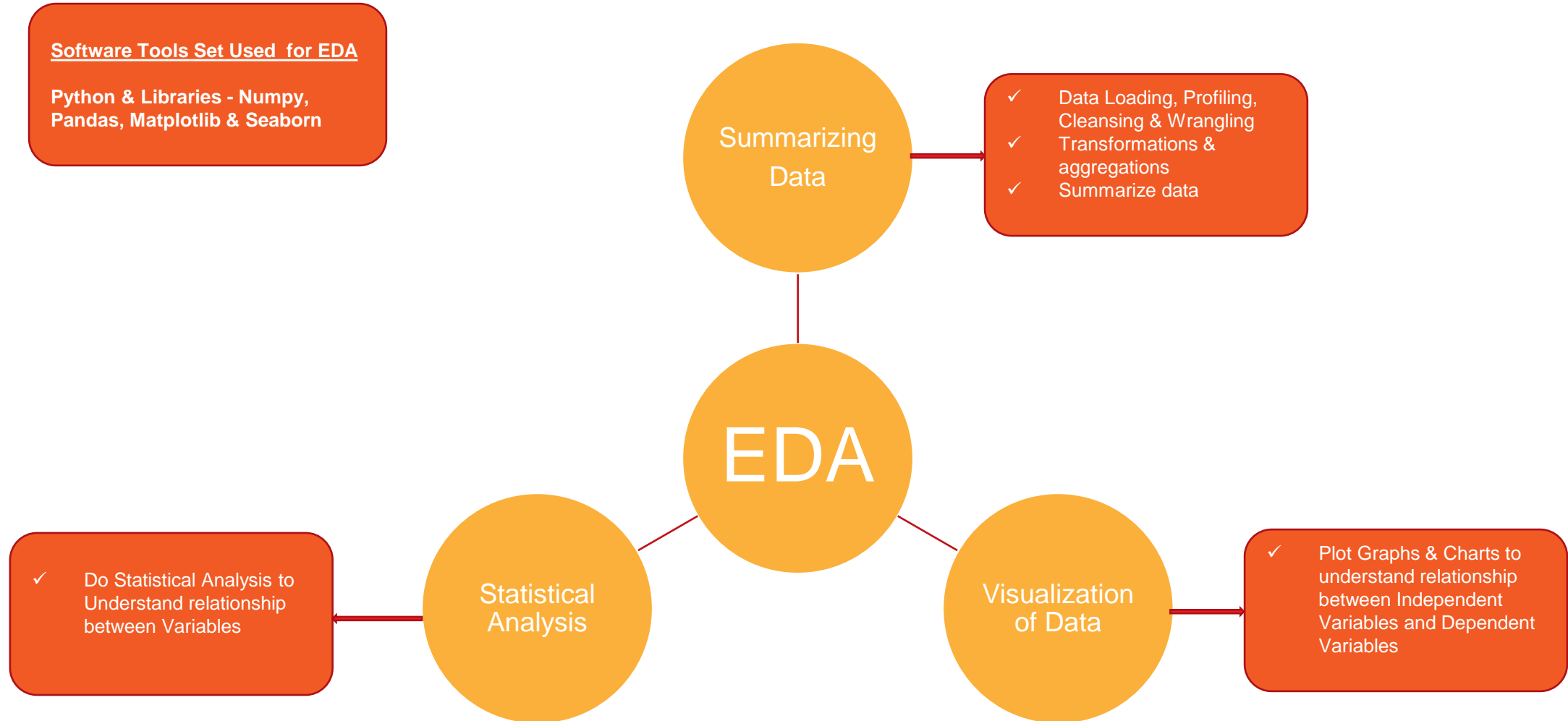- ✓ _Summarize Findings & Recommendations_

The Bank has also mentioned that customers with 0 account balance are more likely to leave the Bank.
We will consider the above statement as a **Hypothesis** and Validate the same as part of the Analysis

Key Assumptions:
a) _This is an Exploratory data analysis project & the outcome will be in the form of findings & recommendations. There will not be any Model generated as part of this exercise_
b) _The data set is complete and a significant representation of the customer data across the 3 geographies._
c) _The client team would make themselves available to clarify any questions on the data set. (2 sessions of 2 hours each have been planned to tackle such questions)_
d) _It is possible that the recommendations may not be conclusive and require a deep dive on additional data sets / source systems in the case of some variables._
e) _Since it is an exploratory project with a pre-defined timeline, there will not be a detailed project plan and Project Charter would summarize the details_

Explore Data – Key Components of EDA Considered

Software Tools Set Used for EDA

Python & Libraries - Numpy, Pandas, Matplotlib & Seaborn

Summarizing Data

- ✓ Data Loading, Profiling, Cleansing & Wrangling
- ✓ Transformations & aggregations
- ✓ Summarize data

EDA

Statistical Analysis

- ✓ Do Statistical Analysis to Understand relationship between Variables

Visualization of Data

- ✓ Plot Graphs & Charts to understand relationship between Independent Variables and Dependent Variables

Explore Data – Understand the data

Data Source – CSV file (Comma separated Values)
# of records – 10015
# of Features / Variables – 14
Type of Data – Customer Churn along with some other Variables to analyze, so as to understand the impact of these variables on the Churn

Key Variables to be Analyzed

Independent Variables	Description
CreditScore	Credit Score of the Customer
Geography	Country information of Customer
Gender	Gender of Customer
Age	Age of Customer
Tenure	How long the Customer has been with the Bank
Balance	Balance of the Customer (Amount)
NumOfProducts	Number of products held by the Customer
HasCrCard	Whether the Customer has a Credit Card or Not
IsActiveMember	Where the customer is Active or Not
EstimatedSalary	Estimated Salary of the Customer

Knowledge Discovery Methodology used
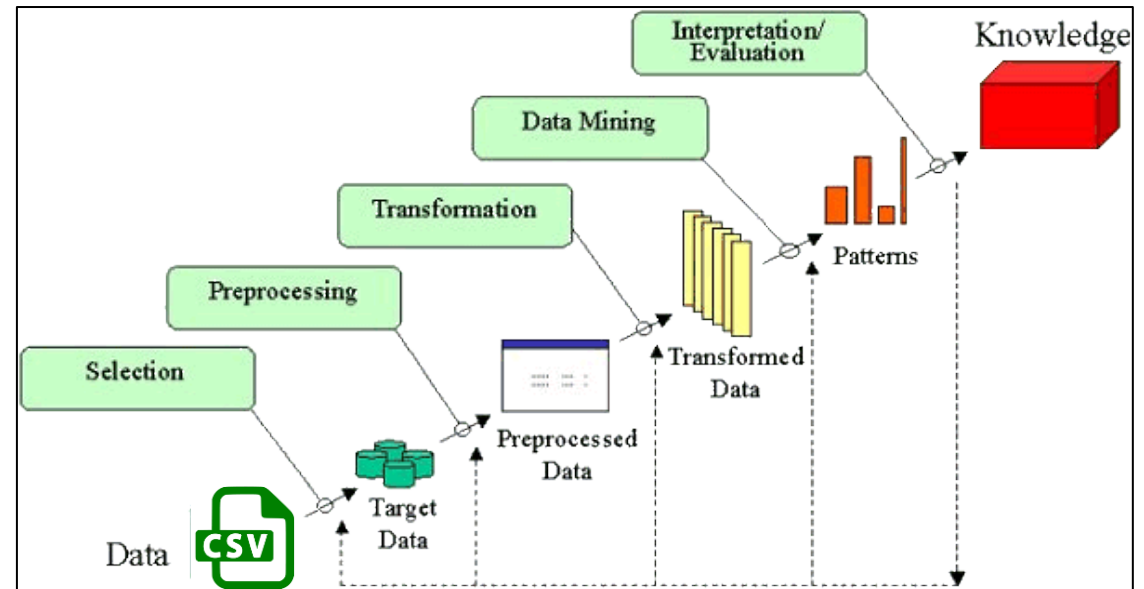
The above Knowledge Discovery Method was used to
a) *Load the CSV file*
b) *Check for null values and fill the same*
c) *Check for duplicate records and remove the same*
d) *Standardize data with different values for the same column etc. Geography had "France" and "FR" Values.*
e) *There was no need to any complex transformations but above standardization involved some simple transformations.*
f) *Following the above steps, the Data set was loaded, inspected & corrected for null values, deduplication and Standardization there by resulting in a target data set of 10,000 rows.*

Explore Data – Proportion of Customers Churned

Proportion of customer churned and retained
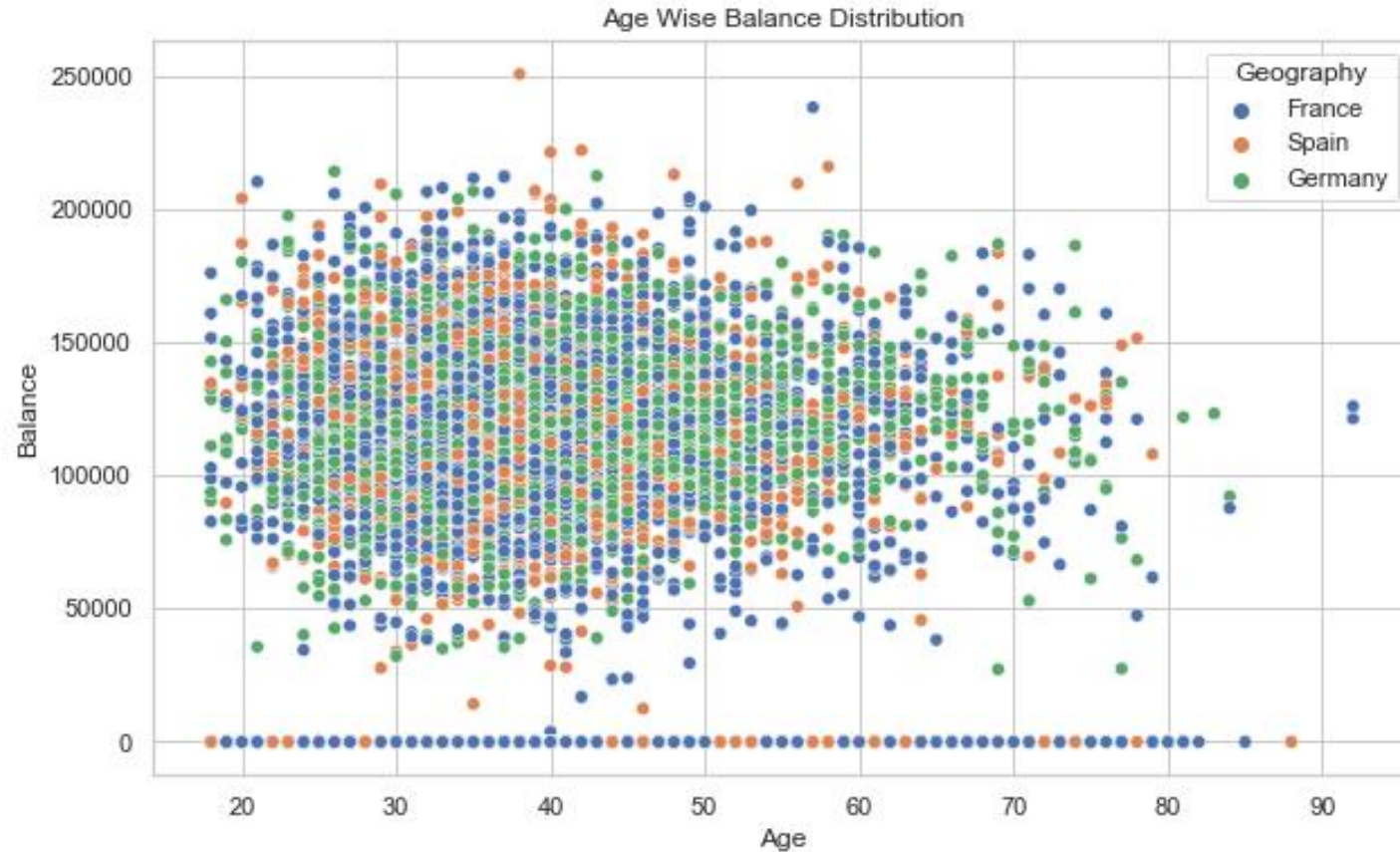

Exited — 20.4%

Retained — 79.6%
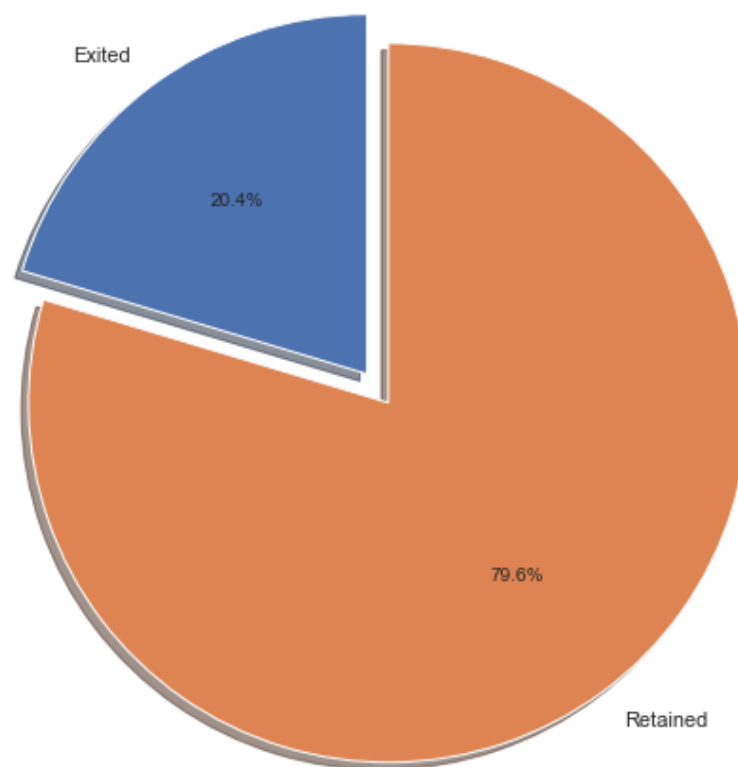
- ➢ Close to 20% churn based on the 6-Month random data set provided based on 10,000 customers & 3 Geographies

- ➢ This seems to be a significant number warranting further analysis

# Explore Data – Customer Age wise Balance distribution

Age Wise Balance Distribution

Observation :

a) There does not seem to be any correlation between Age of Customer and Balance of Customer.

b) One Interesting fact is that France & Spain has Accounts with 0 balances. This is ironic as Germany has maximum number of Churned Customers.

This observation invalidates the hypothesis that customers with 0 account balance are most likely to leave the bank.
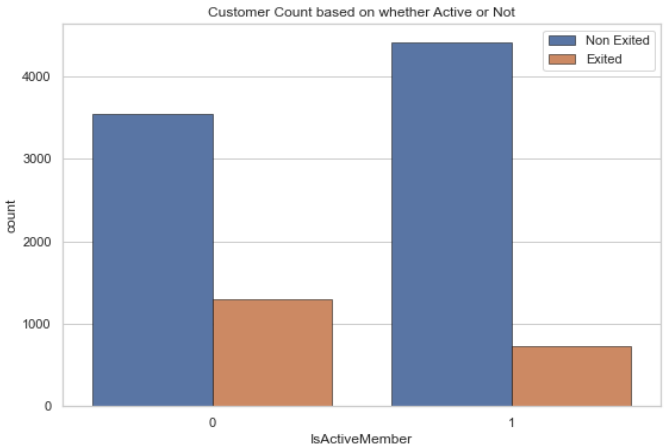
Explore Data – Categorical Variables Vs Churn
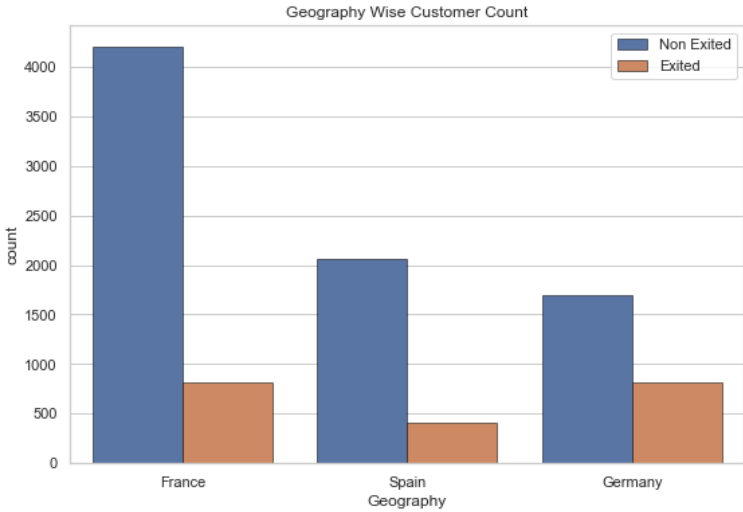
Country wise Churn – Plot 1
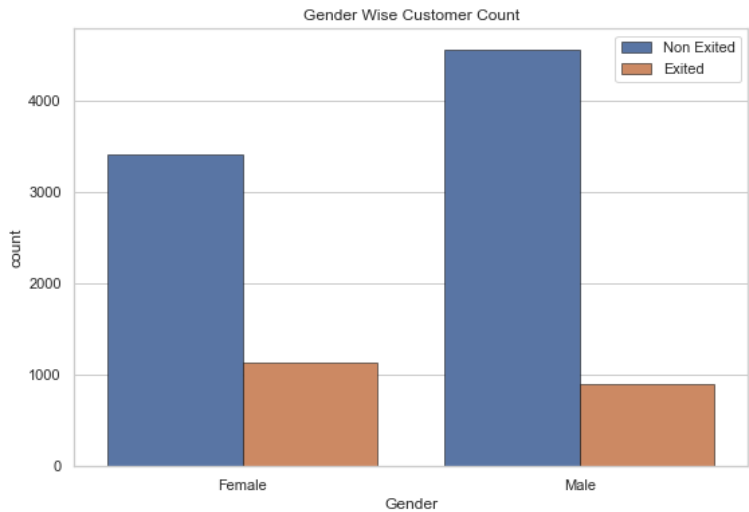
Gender wise Churn – Plot 2
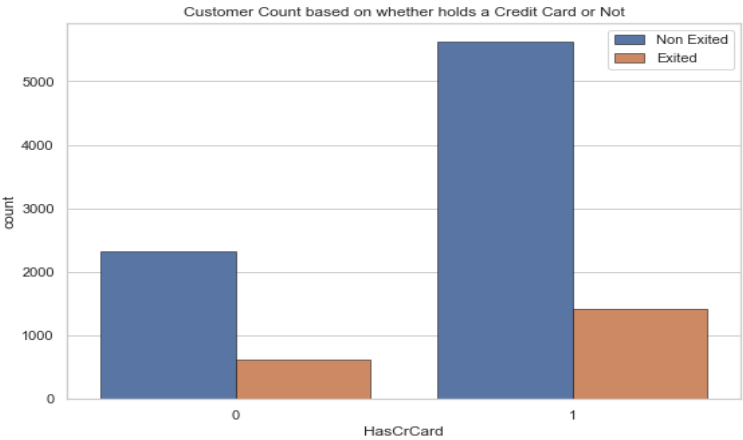
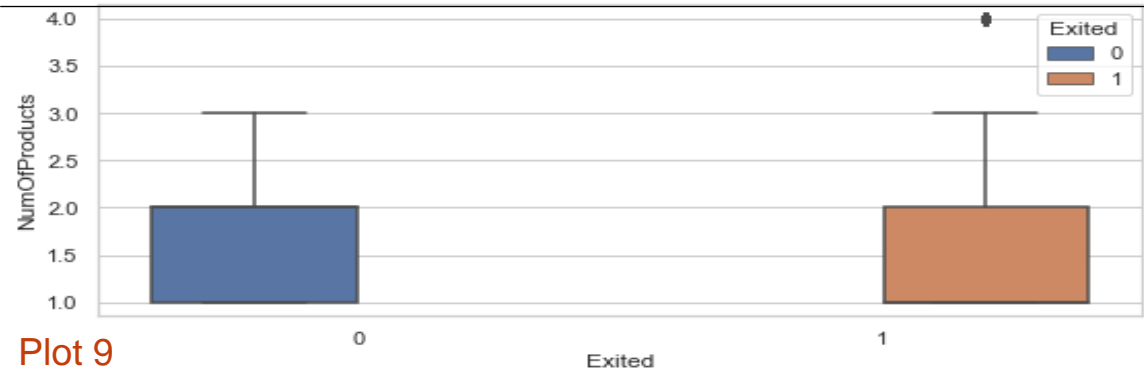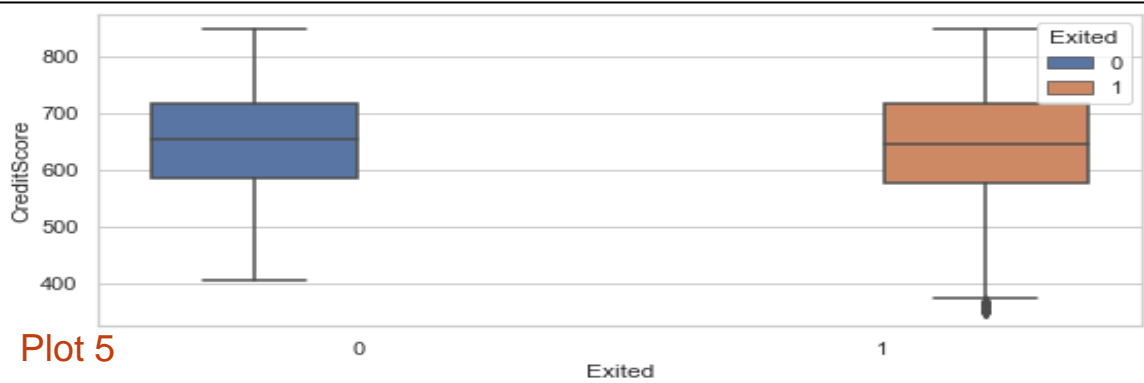**Credit Card holding Customers Vs Churn – Plot 4**

Is Active Customers Vs Churn – Plot 3

Categorical Variables Vs Churn – Findings & Recommendations

S#	Variables Analyzed	Findings	Conclusion & Recommendations	Supporting Data

1	Countrywise Customer Churn - Plot 1 (Slide 7)	Observation 1: The plot shows the count of customers present in each country against customer churned from the same country. Here we can notice France and Germany has maximum number of customers as compared to Spain. But the maximum number of customers Exited are from Germany.	Potential reasons for higher churn in Germany could be a) The recent digitization drive in Germany resulting in closure of branches and promoting self-service mechanisms could have some customers leave. Further analysis is required b) More banks with better strategies and competitive market c) Current service levels / schemes provided in Germany have not been visited for a while	`France      5014`
				`Germany      2509`
				`Spain        2477`

			Recommendations - 1) Bank to revisit and enhance existing schemes / Service Levels and bring in new schemes to cater to the Competition. 2) Need to do a separate Survey to understand the impact of Digitization.	
2	Genderwise Customer Churn - Plot 2 (Slide 7)	Observation 2: The proportion of female customers churning is also greater than that of male customers.	Potential Schemes could be 1) Banks policies & schemes towards Females have become more competitive and competitors are offering better schemes / incentives. 2) Digitization could have had an impact. Survey mentioned in Rec 1) should provide insights	
			Recommendations - Bank to revisit the schemes / products offered to Female customers and look for potential enhancement / incentivisation based on Market offerings	
3	Is Active Customer Churn - Plot 3 (Slide 7)	Observation 3: As expected the inactive members have a greater churn. Observation is overall portion of the inactive members is quite high.	This is a worrying Trend as Overall portion of the Inactive members is quite high **Recommendations**- Bank might have to devise a new or upgrade existing Schemes / Programs implemented in this area to turn this group to active customers. This could have a positive impact on the Customer Churn	
4	Credit Card holding Customer Churn - Plot 4 (Slide 7)	Observation 4: 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 or It's possible that the Credit Card promotions / offerings pertaining to the Bank is very good compared to contributors and hence people are just availaing that facility	

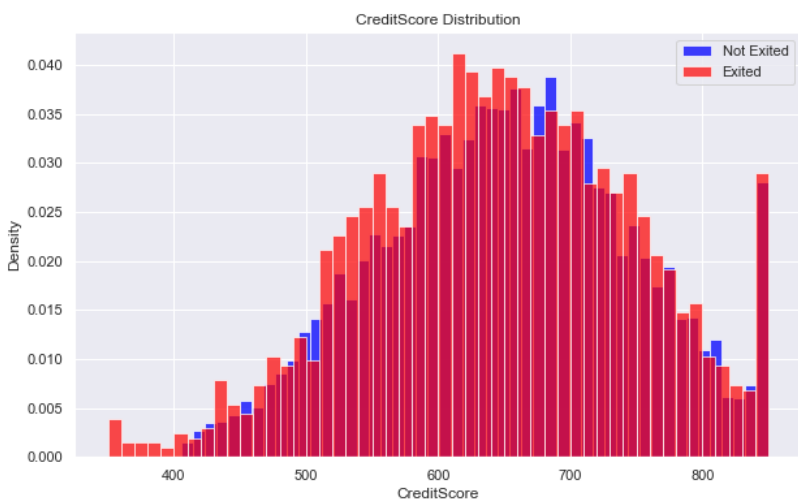# Explore Data – Continuous Variables Vs Churn

Plot 5
Plot 6
Plot 7
Plot 8
Plot 9
Plot 10

Continuous Variables Vs Churn – Findings & Recommendations

S#	Variables Analyzed	Findings	Conclusion & Recommendations
1	Credit Score Vs Churn - Plot 5 (Slide 9)	Observation 5: Credit Score Vs Churn - Credit Score of both exited and non exited customers seems to be more or less same. No significant difference. Can try another plot to see if could be any observations.	Need to explore the data further using other types of plots to check if can get any conclusion
2	Agewise Customer Churn - Plot 6 (Slide 9)	Observation 6: Age Vs Churn - older customers are churning more than the younger ones	This Finding alludes to a potential difference in service preference in the age categories. **Recommendations -** The bank may need to review their target market as well as review the existing schemes / offerings for the different age groups so as to offer better incentives in line with the competition. (Need to validate against Market offerings)
3	Tenure Vs Customer Churn - Plot 7 (Slide 9)	Observation 7: Tenure Vs Churn - Customers at either end (spent little time with the bank or a lot of time with the bank) are more likely to churn compared to those that are of average tenure.	**Recommendations-** The bank might need to review it's existing promotions / offers for customers based on their Tenure so as to provide different levels of service / products based on the years with the Bank such as better interest rates, new products etc. Loyalty programs for long duration customers need to be introduced / revised based on Market standards
4	Customer Balance Vs Churn - Plot 8 (Slide 9)	Observation 8: Balance Vs Churn - Worryingly, the bank is losing customers with significant bank balances which is likely to hit their available capital for lending.	**Recommendations -** The bank might have to revisit their interest rate policies based on balances as well as the other instruments like FD etc. that can be made more attractive & flexible with increasing customer balance maintained so that customers can be retained. Higher balance customers (above certain amounts) might be given Privileged services and customized products
5	# of Products held Vs Churn - Plot 9 (Slide 7)	Observation 9: The Number of Products does not seem to impact the churn based on the data available.	However, based on market intelligence, a) Holding to more products could potentialy indicate longer tenure with the bank, active meber, Increased Investments and thereby higher balance etc. More data might be required to draw a conclusion.
6	Customer Estimated Salary Vs Churn - Plot 9 (Slide 7)	Observation 9: The Estimated Salary does not seem to impact the churn based on the data available.	However, a) Higher Salary could mean some privileged services & products could be made available. More data is required in this area to draw a conclusion.

Explore Data – Continuous Variables Vs Churn (Additional Plots)
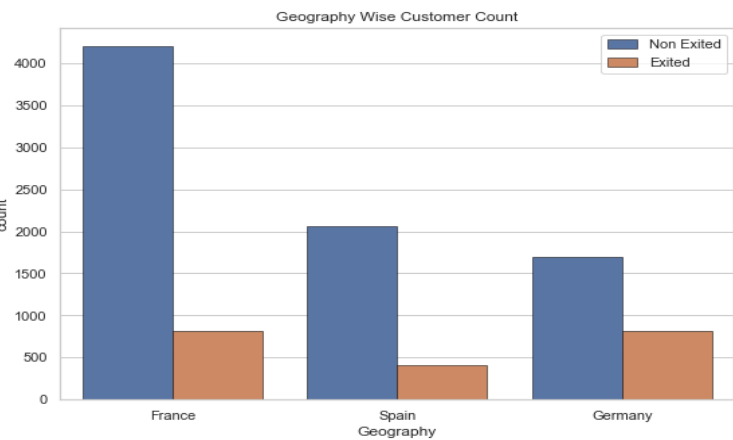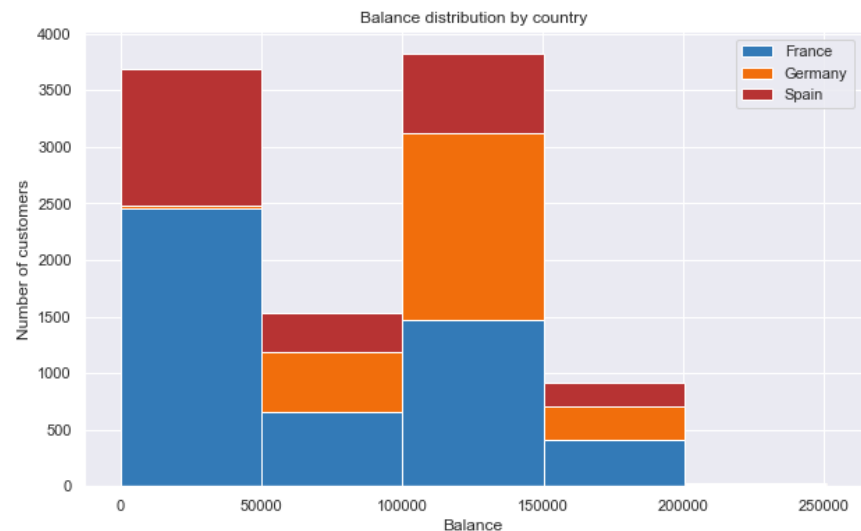

Age wise Churn – Plot 6a

Credit Score wise Churn – Plot 5a

Variables Analyzed	Findings	Conclusion & Recommendations
Agewise Customer Churn - Plot 6a (Slide 11)	Observation 6a : Age Vs Churn (Histogram) - a)The number of customers in the bank decrease as the age increases alluding to a potential difference in service preference in the age categories. This reconfirms Plot 6 Observation 6. b) After the age of 40 there is an increase in the rate of customers leaving the bank.	This Finding alludes to a potential difference in service preference in the age categories. **Recommendations -** The bank may need to review their target market as well as review the existing schemes / offerings for the different age groups so as to offer better incentives in line with the competition. (Need to validate against Market offerings)

Variables Analyzed	Findings	Conclusion & Recommendations
Credit Score Vs Churn - Plot 5a (Slide 11)	Observation 5a: Credit Score Vs Churn - Credit Score of both exited and non exited customers seems to be more or less the same, but Histogram gives additional information.	The Histogram clearly indicates that those customers with a score < 400 are churning. **Recommendations** - Bank might have to take a look at the credit scoring parameters and the scores itself to understand the cause

Balance Distribution by Country

Country wise Churn – Plot 1

Observations:
The plot above shows the Balance of customers of different countries.
In the first bar you notice, It contains the customers who have a balance between (0 - 50000). France has the highest number of customers and then comes Spain and Germany has very less customers in this segment.

In the third bar, you notice Germany (100000 - 150000) has the maximum number of customers.

Based on the above plot we can conclude that the accounts from Germany are very rich as compared to the other two countries. But if you see Plot 1(Pasted below for comparison), then you'll understand the maximum number of customers leaving the bank are from Germany.

<<<On this note, we can infer that Germany being a strong Economy, has multiple banks which offer better incentives to the customers as compared to Euro Bank. That's why there is a significant churn as customers might be opting for other banks. Another potential reason could be the latest digitalization drive. We don't have enough though to prove it upfront>>>

Wrap up

In addition to the recommendations provided, the Bank, in general,

a) Needs to analyze the call details to understand the reasons for Churn. The CRM solution should be able to provide with the details on the same.

b) Enhance the existing survey parameters as well as frequency of the surveys so as to understand customer movements well in advance.

c) Invest in schemes in bringing the customers up to speed and comfortable with the digital drive.

d) Churning of customers with a long-standing relationship needs to be looked in to in detail as these could be Bank's "VIP" customers and need to be retained.

e) Churning of customers with higher balances as well poses a serious threat as it impacts the capital. Bank might need to mobilize a special task force to focus on points d) and e) immediately to bring the situation to control and gain the confidence of these "VIP" customers

While the Bank has been very active in bringing new customers, Retention mechanisms need to be strengthened as it's always easy to sell to existing customers.