

DS - ASSIGNMENT REPORT

Prepared by : Omkar Vilas Narkar

Date: 10th July 2024

Executive Summary

This report provides a comprehensive analysis of the provided datasets (Bank Statement, Office Supplies Data, Churn Modeling Data). The analysis includes transaction trends, balance trends, spending patterns, income sources, sales analysis, customer analysis, time series analysis, geographical sales trends, profit analysis, customer demographics, churn analysis, product usage, financial analysis, and predictive modeling for customer churn.

1. Bank Statement Data Analysis

a) Transactional Analysis

1. Total number of transactions made over the year.

Total Number of Transactions in 2023: 431

Total Number of Transactions in 2024: 554

2. Distribution of transaction amounts.

Medium	344
Small	329
Large	312

Reasoning:

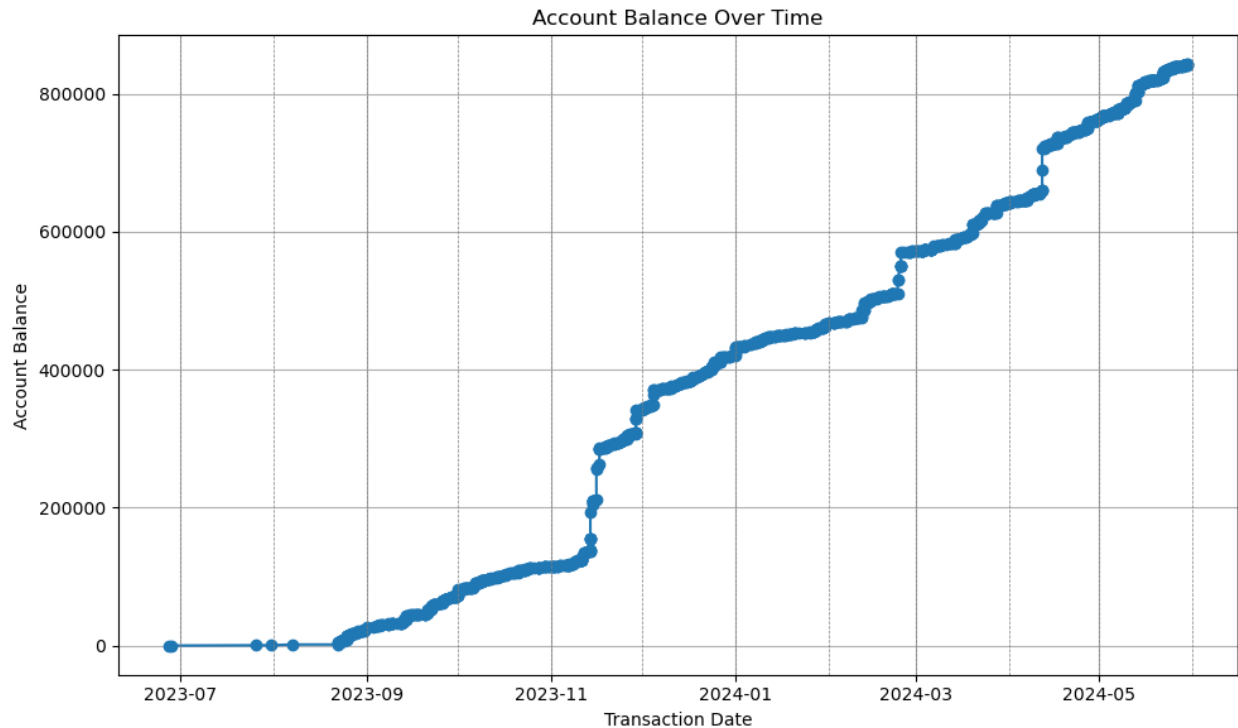
Quantile method is used for distributing transition amount, where Bottom & Top 33% data will be considered as Small & Large Transactions respectively.

3. The frequency of different transaction types (debit vs. credit).

DEBIT	695 transactions
CREDIT	290 transactions

b) Balance Analysis

1. The trend of the account balance over time



2. Periods with significant changes in the account balance.

- Up until late August 2023, the account balance consistently remained at a low level.
- During mid-November 2023, the account balance surged sharply from 150,000 to 300,000.
- By late February 2024, the balance had increased from 500,000 to 580,000.
- In mid-April 2024, the balance further grew from 625,000 to 725,000.

c) Spending Pattern

1. The main categories of expenses & Frequency.

Categories	Frequency	Total Spending
1. UPI	789	587092.90
2. NEFT	159	172430.51
3. BY CASH	1	45000.00
4. IMPS	25	23672.00
5. ATM	3	13500.00
6. PRCR	4	830.00
7. 51888100006461IntPd01	4	135.00

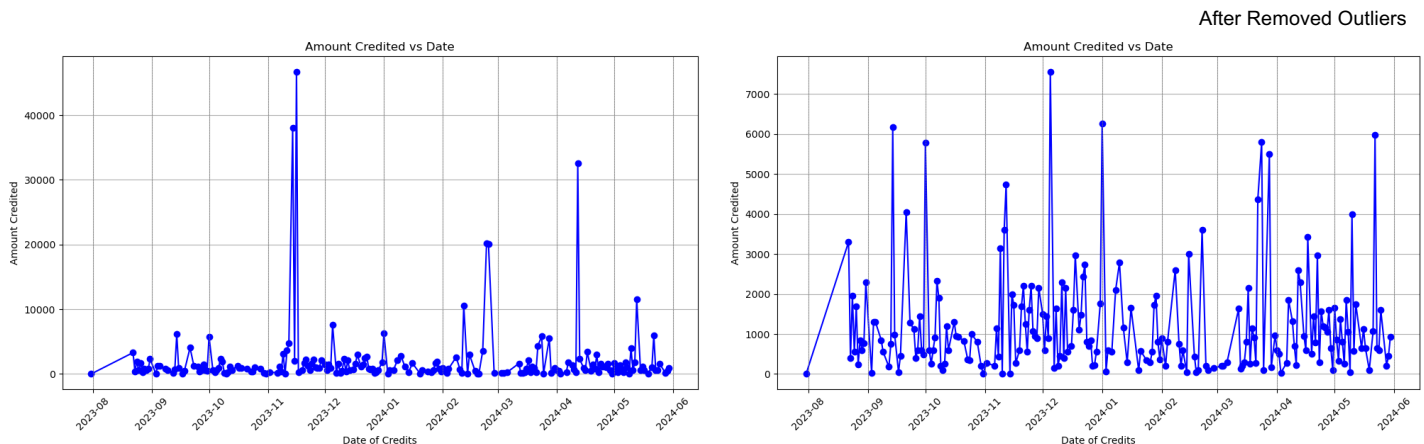
d) Income Analysis

1. The main categories of expenses

Categories	Total Amount
1 UPI	179333.00
2 CASH	45000.00

The Analysis shows the UPI credits is the main source of Income.

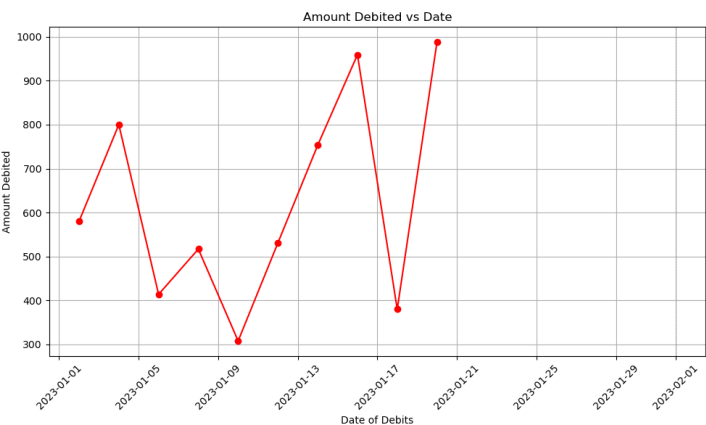
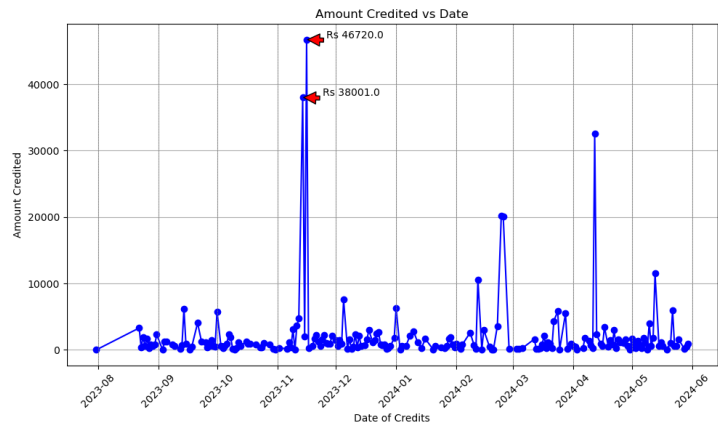
e) Patterns in the timing and amount of income received



- In the month of November and December highest amount is credited which is over 27,00,000.
- Majority of large amount credited in end of the month.

d) Alert Generation

1. Unusual or suspicious transactions



Reason:

- During mid-November, two unusually large credit transactions were observed, each exceeding seven times the average credited amount.
- In mid-April, a credit transaction of \$32,500 was recorded, which was more than five times higher than the average credited amount.

2. Alerts for Low Balance and High Expenditure Periods

Low Balance Periods:

	valueDate	currentBalance	low_balance_alert
769	2024-04-20	1426.21	True
830	2024-05-04	1423.21	True
797	2024-04-28	1418.21	True
38	2023-08-28	1405.80	True
39	2023-08-28	1395.80	True
..
73	2023-09-13	29.80	True
980	2024-05-29	23.41	True
864	2024-05-09	19.21	True
749	2024-04-14	1.21	True
51	2023-09-01	0.80	True

High Expenditure Periods:

	valueDate	amount	high_expenditure_alert
291	2023-11-16	45000.0	True
275	2023-11-14	37999.0	True
743	2024-04-12	30000.0	True
744	2024-04-12	30000.0	True
295	2023-11-17	21000.0	True
..
303	2023-11-20	1700.0	True
266	2023-11-12	1660.0	True
813	2024-05-01	1650.0	True
621	2024-03-12	1630.0	True
359	2023-12-08	1630.0	True

[99 rows x 3 columns]

3. Office Supplies Data Analysis

a) Sales Analysis

1. Total sales for each product category

	Item	Total sales	Total Price
1	Binder	15	9577.65
2	Pencil	13	2135.14
3	Pen Set	7	4169.87
4	Pen	5	2045.22
5	Desk	3	1700.00

2. Product category has the highest sales

Binder product category has a highest Sales of Rs9577.65.

b) Customer Analysis

1) The top 10 customers by sales

	Customers	Sales
1.	Richard	396
2.	Alex	281
3.	Bill	213
4.	Matthew	193
5.	Rachel	183
6.	Morgan	173
7.	Susan	170
8.	Smith	156
9.	James	142
10.	Nick	125

2) The total number of unique customers

There are **11** Unique customers:

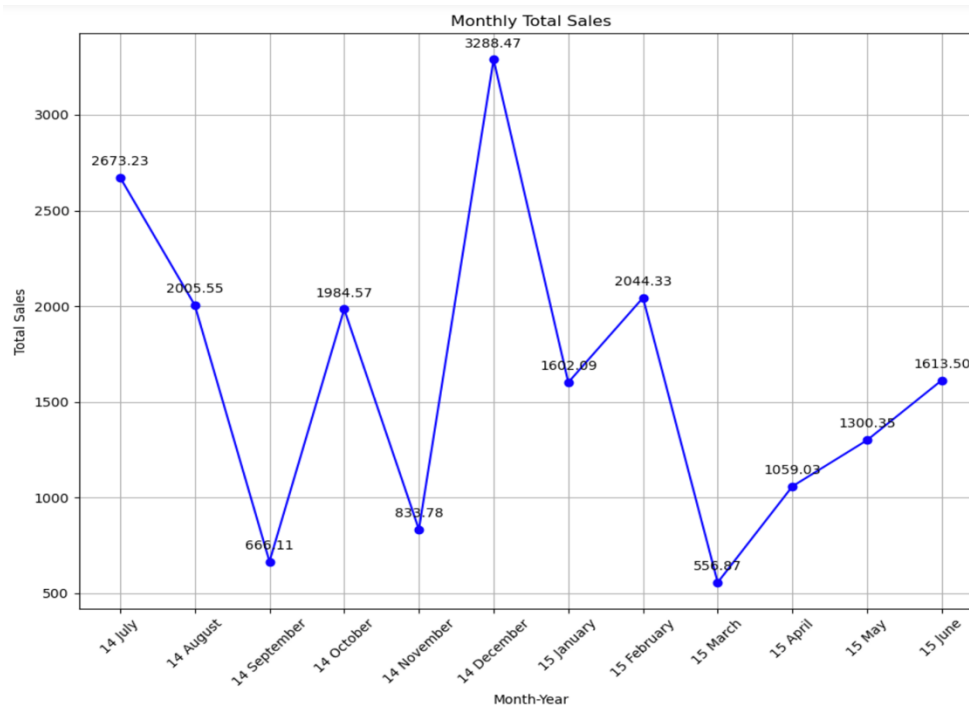
Richard, Nick, Morgan, Susan, Matthew, James, Smith, Bill, Thomas
, Rachel, Alex

3. The customer purchase frequency

	Customer	Frequency
0	Alex	5
1	Bill	5
2	James	4
3	Matthew	4
4	Morgan	3
5	Nick	2
6	Rachel	4
7	Richard	8
8	Smith	3
9	Susan	3
10	Thomas	2

c) Time Series Analysis

1. The monthly sales trends over the past year



2. Seasonal patterns in the sales data

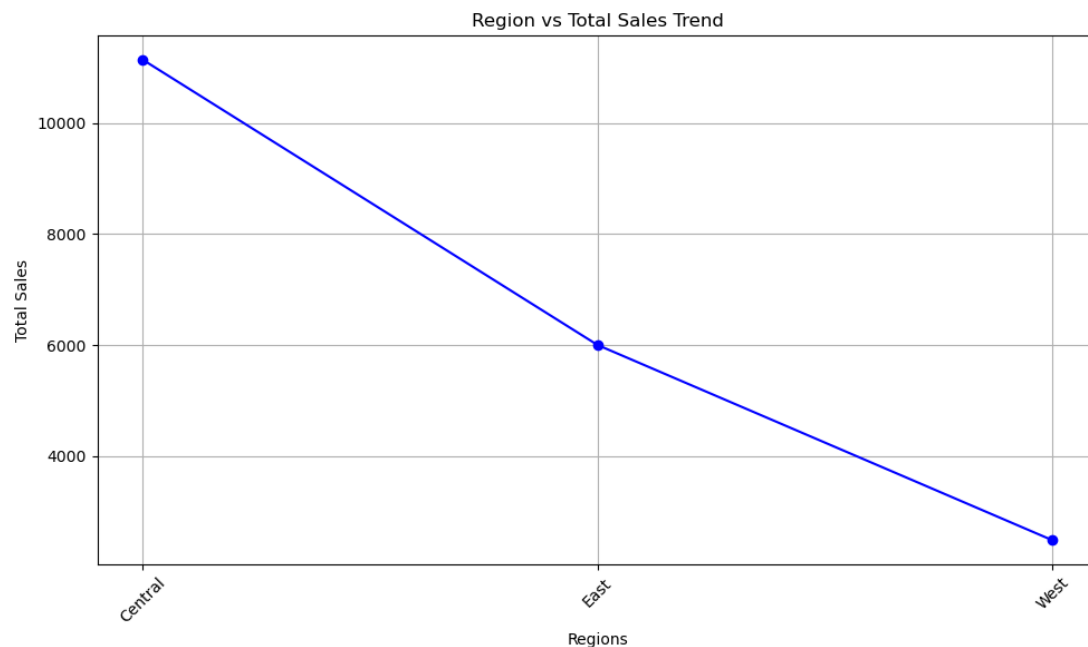
- In August 2014 and February 2015, there was a notable decline in sales performance.
- Conversely, November 2014 marked the highest sales volume observed during the period.

d) Geographical Analysis

1. Regions generate the most sales

Central region has most sales of 11139.07.

2. The sales trends across different regions



e) Profit Analysis

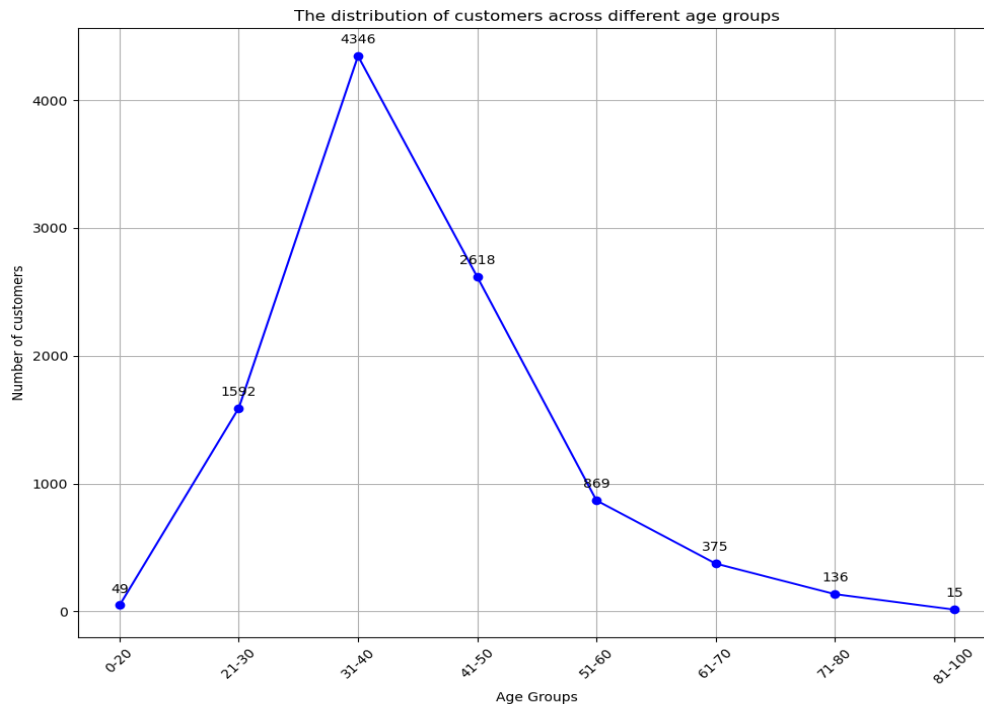
Reasoning:

To calculate profit both Selling and Cost price is needed. But one of the prices is missing in the data.

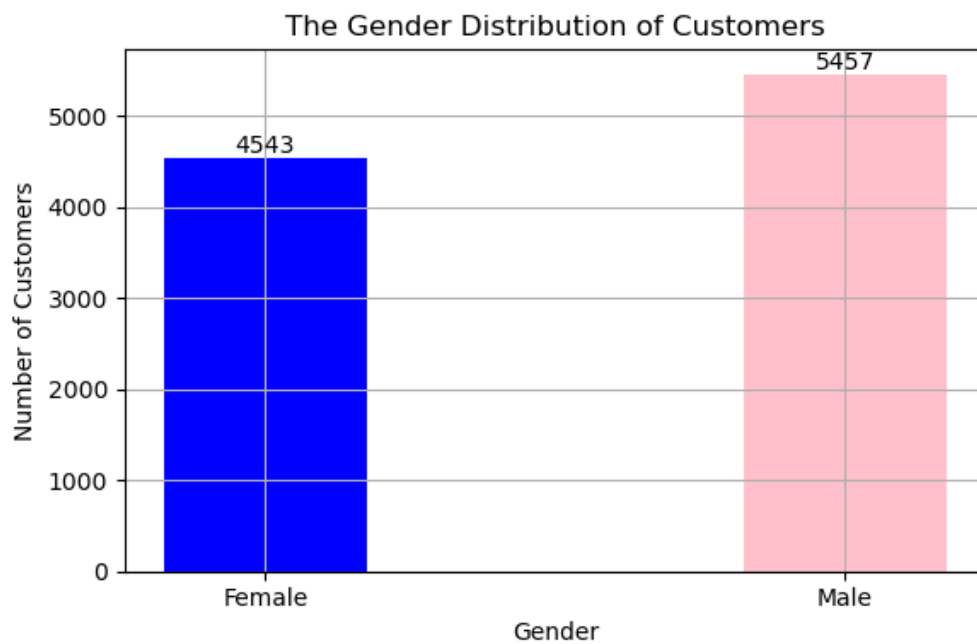
3. Customer Churn Data Analysis

a) Customer Demographics

1. The distribution of customers across different age groups



2. The gender distribution of customers



b) Churn Analysis

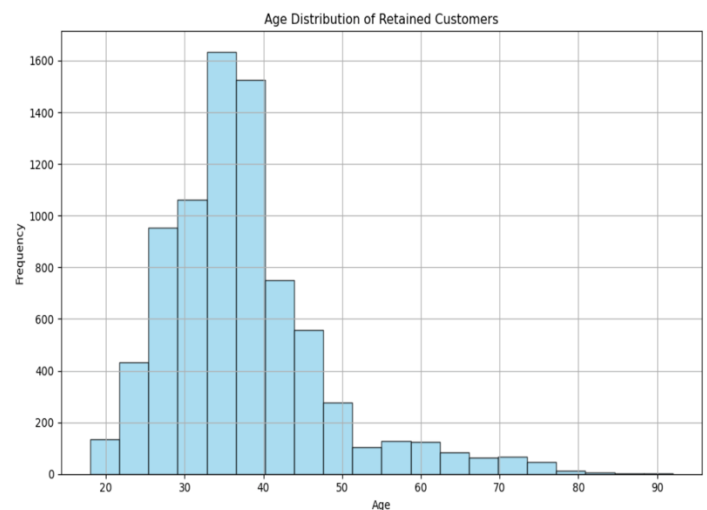
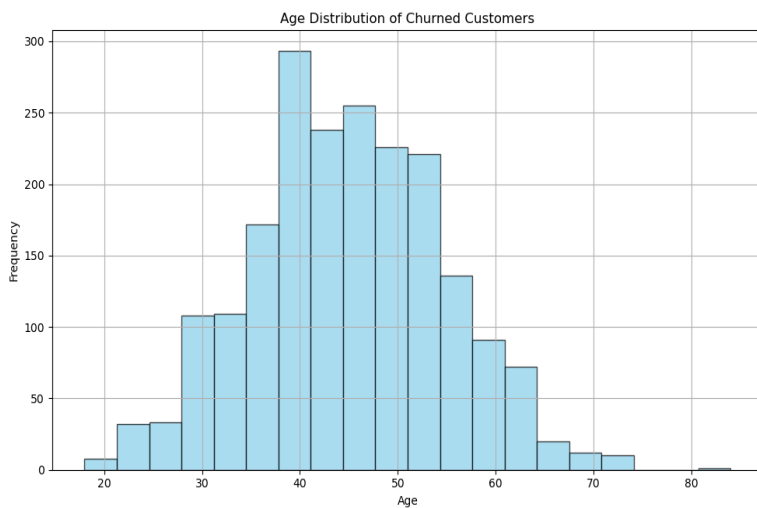
1. Percentage of customers have churned

Percentage of customers have churned: 20.37

2. The main reasons for customer churn

- The churn rate is notably higher among customers **aged 35–55**.
- A significant portion of churned customers, **52%**, are from **France**.
- The average balance of churned customers stands at \$91,108.54, which is **12%** higher than that of retained customers.

3. Patterns or trends among customers who have churned



- The churn rate is notably higher among customers aged **35 to 55**.

c) Product Analysis

1. The most commonly used products or services.

Credit card:

Total customers with credit card: 7055

Total customers without credit card: 2945

Active Members:

Number of active members: 5151

Number of inactive members: 4849

d) Financial Analysis

1. What is the average account balance of customers?

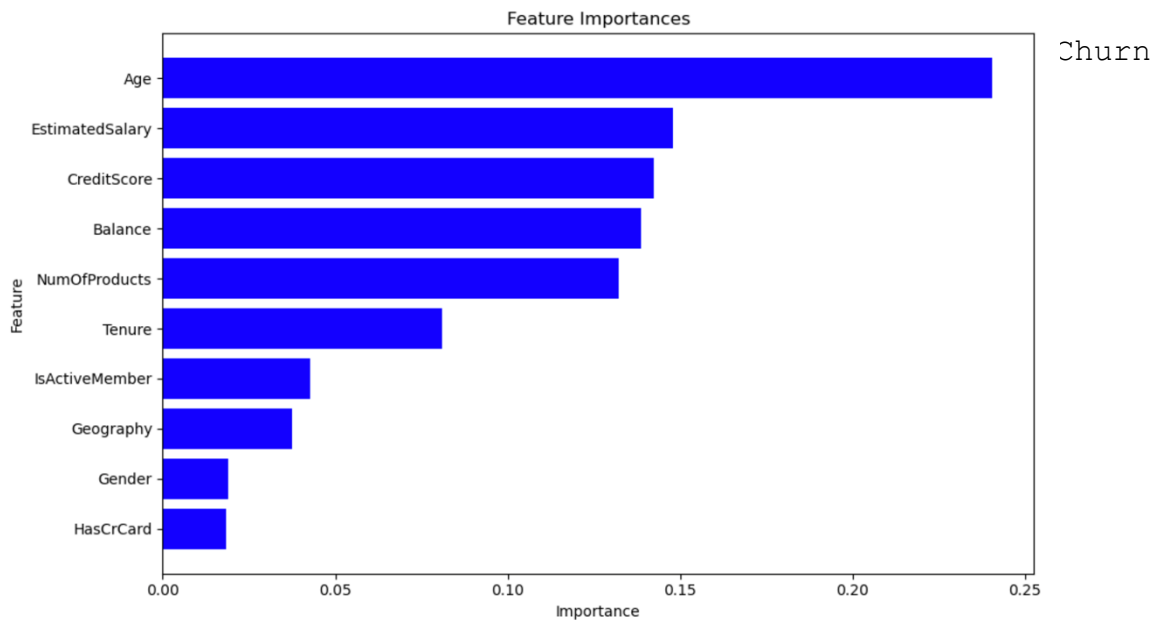
The average account balance of customers: 76485.89

2. The financial characteristics of churned vs. non-churned customers

	Balance	EstimatedSalary	CreditScore
Non-Churned	72745.296779	99738.391772	651.853196
Churned	91108.539337	101465.677531	645.351497

e) Predictive Modeling

1. The most significant predictors of customer churn

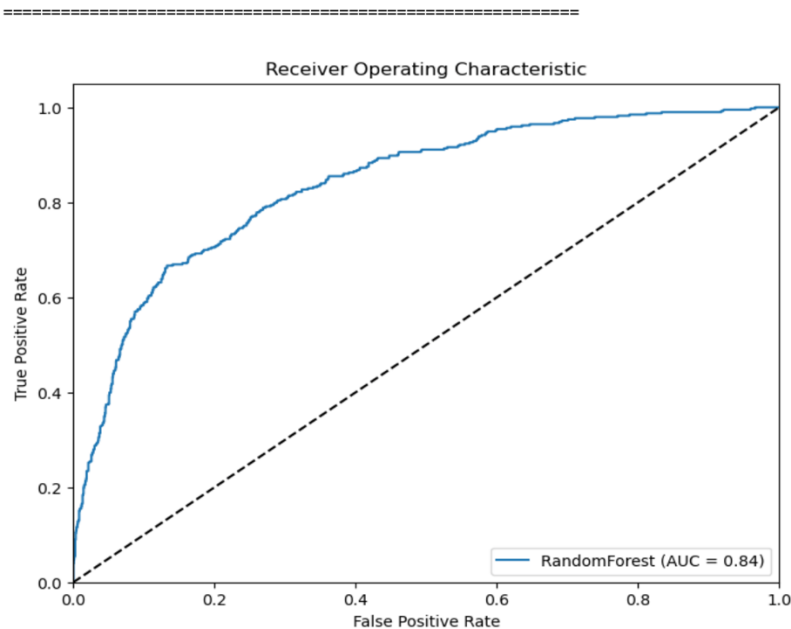


2. A predictive model to identify at-risk customers

Imbalance Dataset: Performed SMOTE Analysis

Logistic Regression: Best performed Model

	precision	recall	f1-score	support
Retained Customer	0.91	0.86	0.89	1607
Churned Customer	0.54	0.67	0.60	393
accuracy			0.82	2000
macro avg	0.73	0.77	0.74	2000
weighted avg	0.84	0.82	0.83	2000



Conclusion:

Based on the comprehensive analysis of bank statements, office supplies data, and customer churn modeling, several key insights emerge. Transaction trends reveal varying distribution patterns of transaction amounts, with notable alerts for unusual transactions and low balance periods. Sales analysis of office supplies identifies product and customer trends, while churn analysis highlights demographic factors influencing customer retention. These insights equip decision-makers with actionable strategies to enhance financial monitoring, optimize sales strategies, and mitigate customer churn risks effectively.