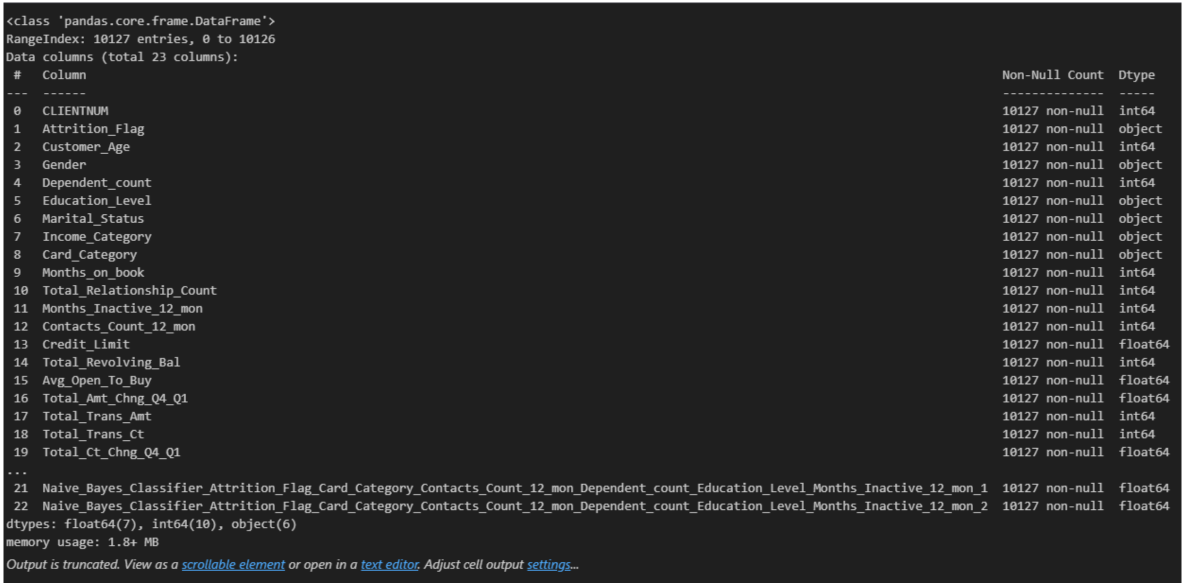
Credit Card Churn Prediction - EDA Report

EDA Overview

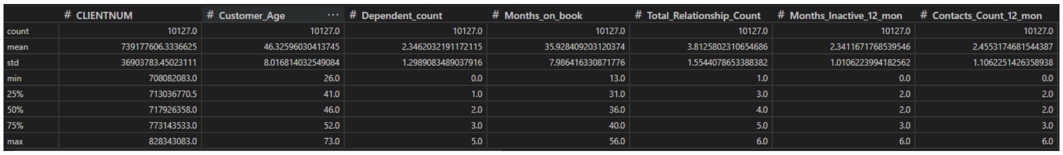
This document presents the Exploratory Data Analysis (EDA) performed on the credit card churn dataset. The purpose of EDA is to understand the data distribution, detect potential issues, and uncover patterns related to churn behavior.

Raw Dataset Overview

Initial dataset before cleaning. Contains original features and Naive Bayes outputs.



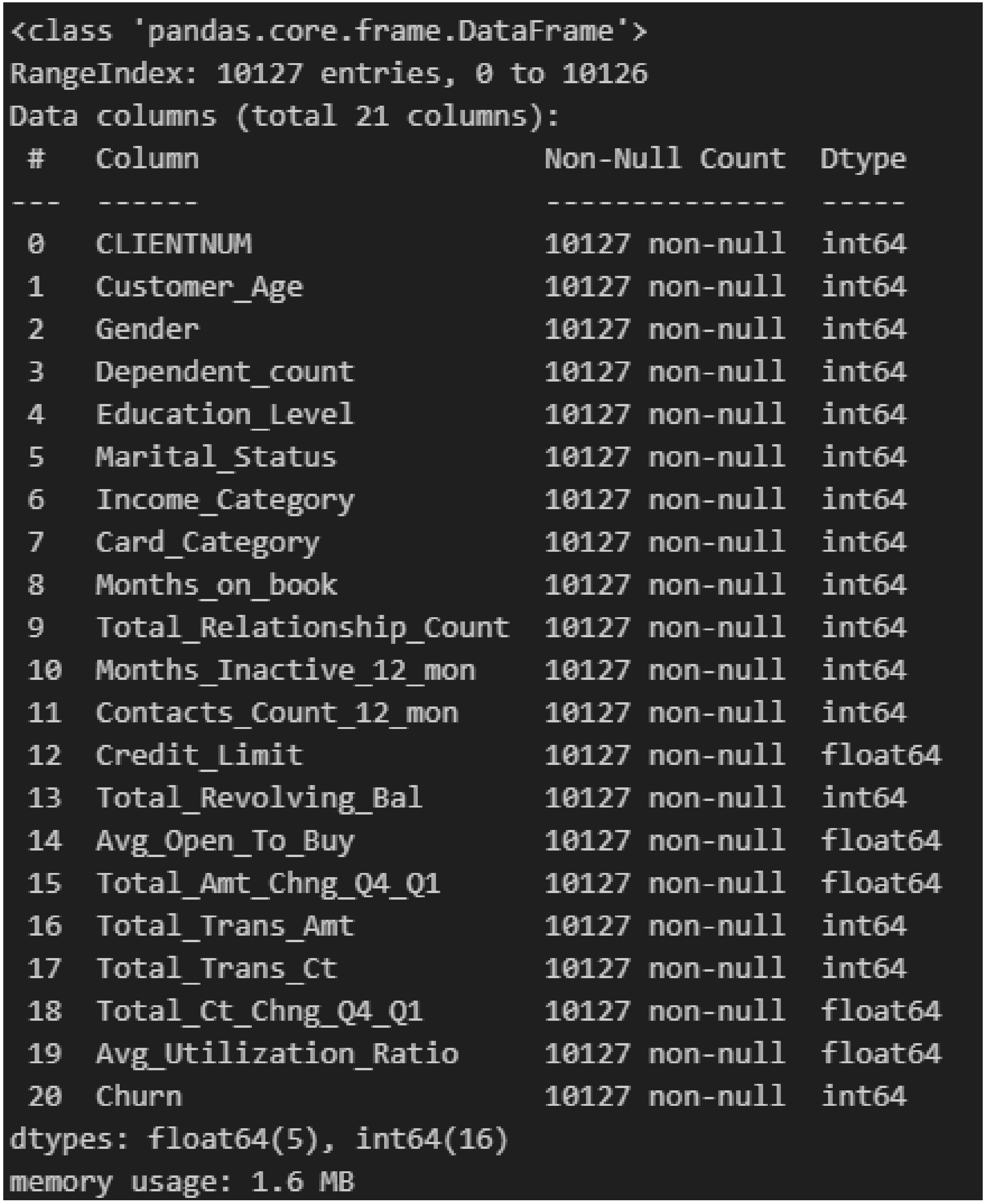
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10127 entries, 0 to 10126
Data columns (total 23 columns):
 #   Column                                                                                                                               Non-Null Count  Dtype
---  ------                                                                                                                               --------------  -----
 0   CLIENTNUM                                                                                                                            10127 non-null  int64
 1   Attrition_Flag                                                                                                                       10127 non-null  object
 2   Customer_Age                                                                                                                         10127 non-null  int64
 3   Gender                                                                                                                               10127 non-null  object
 4   Dependent_count                                                                                                                      10127 non-null  int64
 5   Education_Level                                                                                                                      10127 non-null  object
 6   Marital_Status                                                                                                                       10127 non-null  object
 7   Income_Category                                                                                                                      10127 non-null  object
 8   Card_Category                                                                                                                        10127 non-null  object
 9   Months_on_book                                                                                                                       10127 non-null  int64
 10  Total_Relationship_Count                                                                                                             10127 non-null  int64
 11  Months_Inactive_12_mon                                                                                                               10127 non-null  int64
 12  Contacts_Count_12_mon                                                                                                                10127 non-null  int64
 13  Credit_Limit                                                                                                                         10127 non-null  float64
 14  Total_Revolving_Bal                                                                                                                  10127 non-null  int64
 15  Avg_Open_To_Buy                                                                                                                      10127 non-null  float64
 16  Total_Amt_Chng_Q4_Q1                                                                                                                 10127 non-null  float64
 17  Total_Trans_Amt                                                                                                                      10127 non-null  int64
 18  Total_Trans_Ct                                                                                                                       10127 non-null  int64
 19  Total_Ct_Chng_Q4_Q1                                                                                                                  10127 non-null  float64
...
 21  Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_Level_Months_Inactive_12_mon_1  10127 non-null  float64
 22  Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_Level_Months_Inactive_12_mon_2  10127 non-null  float64
dtypes: float64(7), int64(10), object(6)
memory usage: 1.8+ MB
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...
```



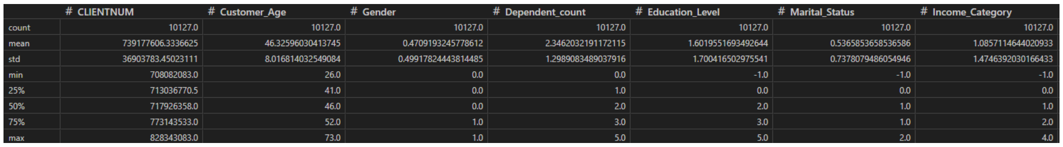
	# CLIENTNUM	# Customer_Age	# Dependent_count	# Months_on_book	# Total_Relationship_Count	# Months_Inactive_12_mon	# Contacts_Count_12_mon
count	10127.0	10127.0	10127.0	10127.0	10127.0	10127.0	10127.0
mean	739177606.3336625	46.32596030413745	2.3462032191172115	35.928409203120374	3.8125802310654686	2.3411671768539546	2.4553174681544387
std	36903783.45023111	8.016814032549084	1.2989083489037916	7.986416330871776	1.5544078653388382	1.0106223994182562	1.1062251426358938
min	708082083.0	26.0	0.0	13.0	1.0	0.0	0.0
25%	713036770.5	41.0	1.0	31.0	3.0	2.0	2.0
50%	717926358.0	46.0	2.0	36.0	4.0	2.0	2.0
75%	773143533.0	52.0	3.0	40.0	5.0	3.0	3.0
max	828343083.0	73.0	5.0	56.0	6.0	6.0	6.0

Cleaned Dataset Overview

After preprocessing and feature selection. Dropped irrelevant columns, mapped categorical values, and created binary churn label.



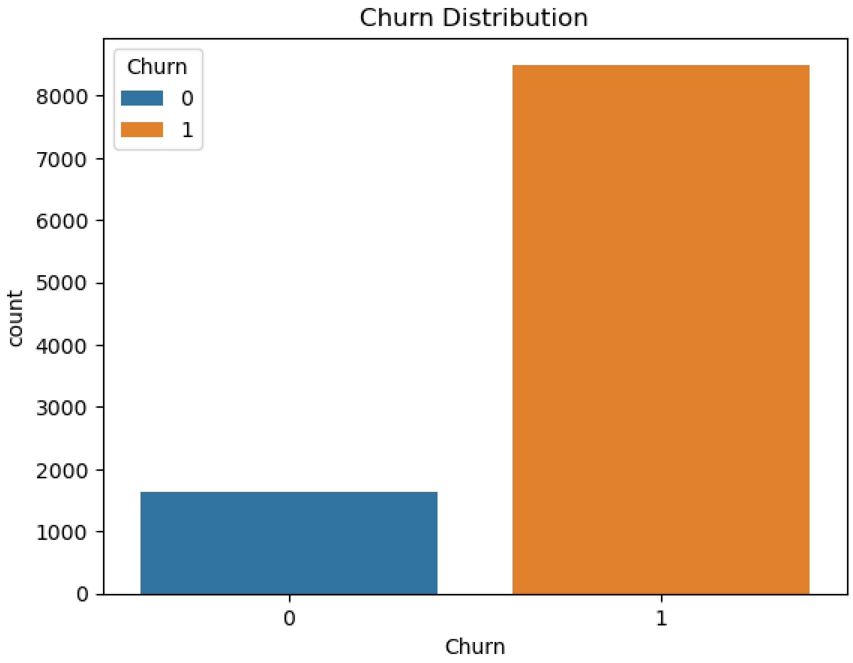
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10127 entries, 0 to 10126
Data columns (total 21 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   CLIENTNUM                 10127 non-null  int64  
 1   Customer_Age              10127 non-null  int64  
 2   Gender                    10127 non-null  int64  
 3   Dependent_count           10127 non-null  int64  
 4   Education_Level           10127 non-null  int64  
 5   Marital_Status            10127 non-null  int64  
 6   Income_Category           10127 non-null  int64  
 7   Card_Category             10127 non-null  int64  
 8   Months_on_book            10127 non-null  int64  
 9   Total_Relationship_Count  10127 non-null  int64  
 10  Months_Inactive_12_mon    10127 non-null  int64  
 11  Contacts_Count_12_mon     10127 non-null  int64  
 12  Credit_Limit              10127 non-null  float64
 13  Total_Revolving_Bal       10127 non-null  int64  
 14  Avg_Open_To_Buy           10127 non-null  float64
 15  Total_Amt_Chng_Q4_Q1      10127 non-null  float64
 16  Total_Trans_Amt           10127 non-null  int64  
 17  Total_Trans_Ct            10127 non-null  int64  
 18  Total_Ct_Chng_Q4_Q1       10127 non-null  float64
 19  Avg_Utilization_Ratio     10127 non-null  float64
 20  Churn                     10127 non-null  int64  
dtypes: float64(5), int64(16)
memory usage: 1.6 MB
```



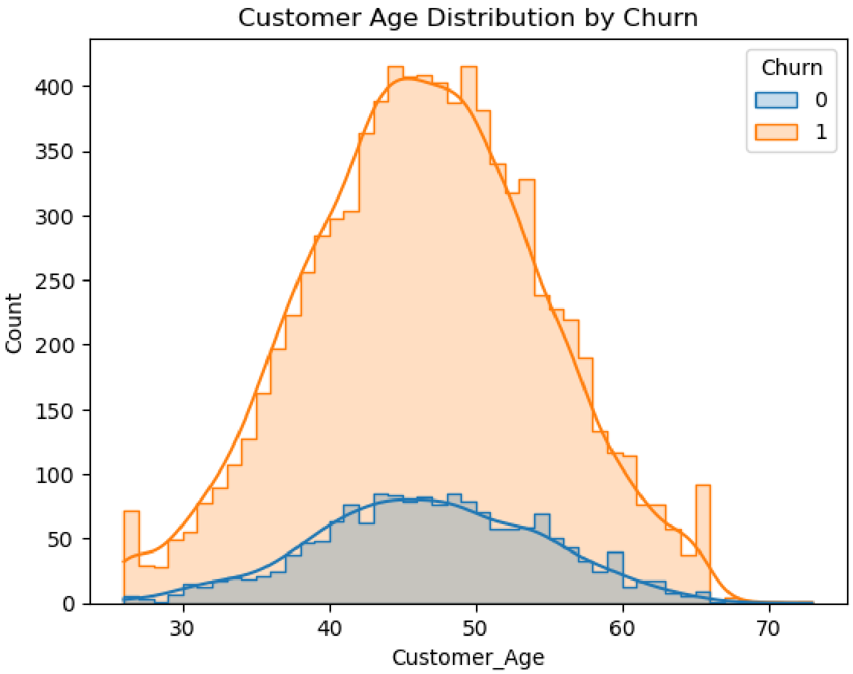
	# CLIENTNUM	# Customer_Age	# Gender	# Dependent_count	# Education_Level	# Marital_Status	# Income_Category
count	10127.0	10127.0	10127.0	10127.0	10127.0	10127.0	10127.0
mean	739177606.3336625	46.32596030413745	0.4709193245778612	2.3462032191172115	1.6019551693492644	0.5365853658536586	1.0857114644020933
std	36903783.45023111	8.016814032549084	0.49917824443814485	1.2989083489037916	1.700416502975541	0.7378079486054946	1.4746392030166433
min	708082083.0	26.0	0.0	0.0	-1.0	-1.0	-1.0
25%	713036770.5	41.0	0.0	1.0	0.0	0.0	0.0
50%	717926358.0	46.0	0.0	2.0	2.0	1.0	1.0
75%	773143533.0	52.0	1.0	3.0	3.0	1.0	2.0
max	828343083.0	73.0	1.0	5.0	5.0	2.0	4.0

Key Visualizations

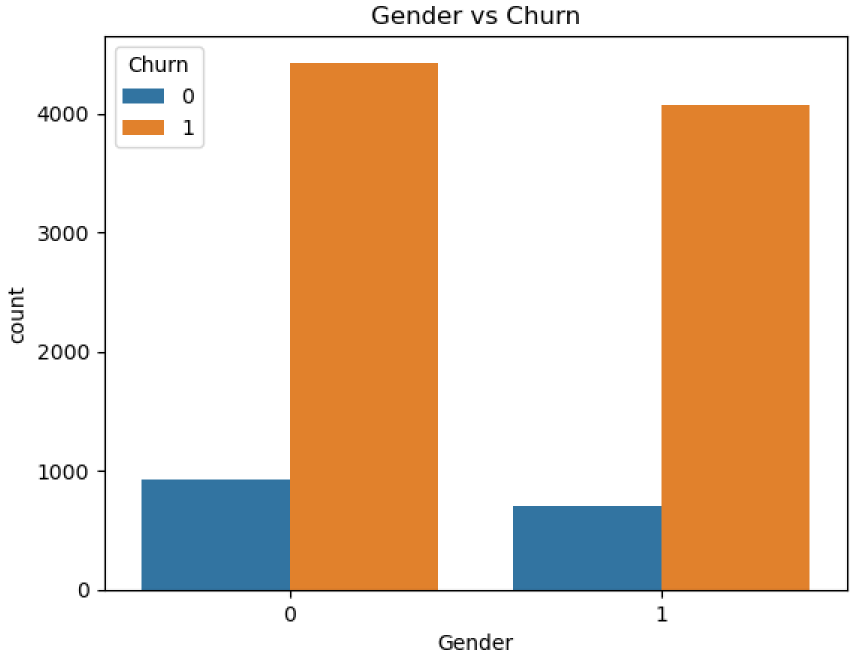
Distribution of Churned vs Non-Churned Customers:
The dataset is imbalanced with more retained customers (1) than churned customers (0).



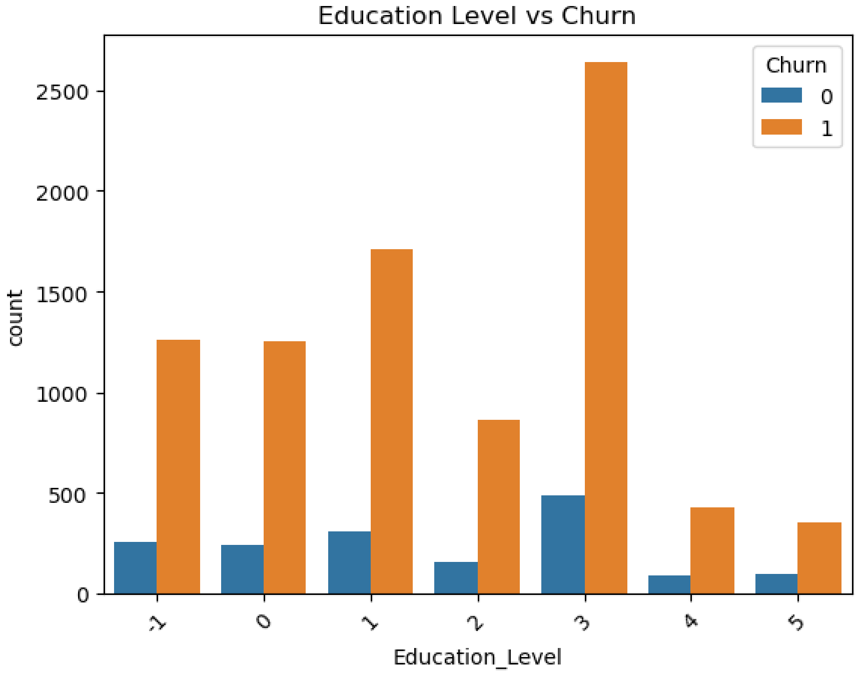
Age Distribution by Churn:
Customers aged 40–60 show a higher tendency to churn, with peaks in the mid-40s.



Gender vs Churn:
Both male and female customers exhibit churn, with slightly higher churn among females.



Education Level vs Churn:
Graduate and college-level customers are more likely to churn compared to other education levels.



AWS Upload

Data was securely uploaded to AWS RDS (PostgreSQL) for remote access and cloud integration.