**ASSIGNMENT IX**

**Project Title:**

Predicting Customer Churn Rate Towards Credit Card services

**Team Leader:**

- Naveen Rampa

**Team members:**

1. J. Sinduja

2. A. Bhavya Sri

3. S. Lakshanya

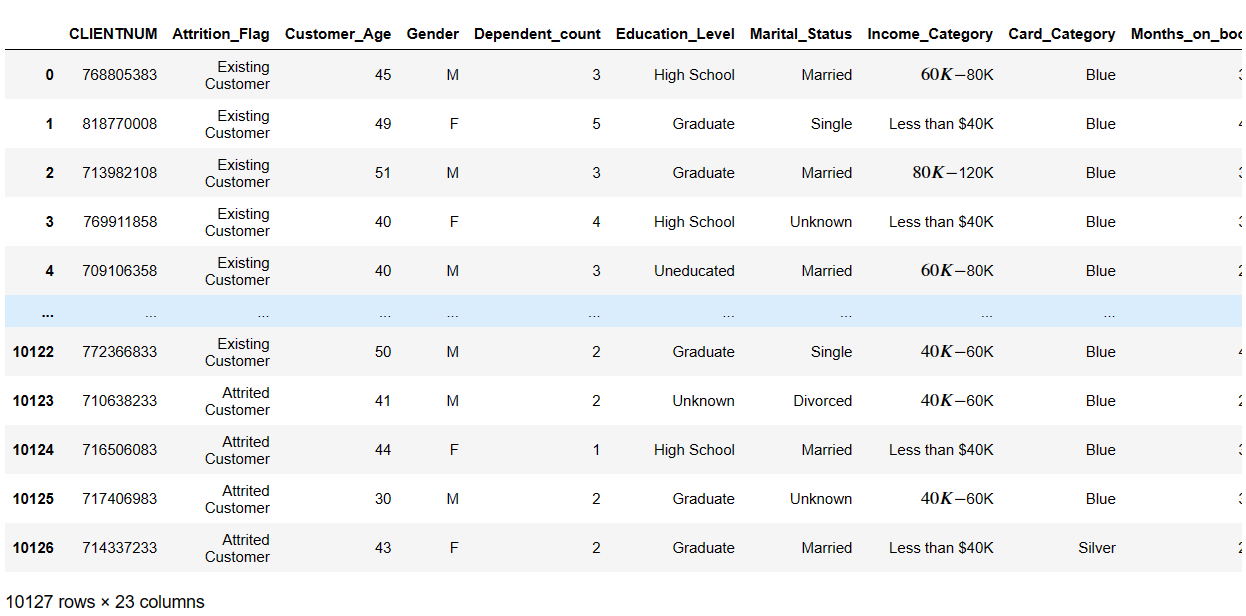
4. Ejumalla Saikiran

**Submission Date:**

**09/03/2024**

**1. Dataset & Objective**

The dataset contains banking transactions records of the customers with 23 features consisting of customer profile and credit card usage. It has 10,127 rows.



Preview of the Imported Dataset

The objective is to build a prediction model for credit card churn prediction using Logistic Regression.

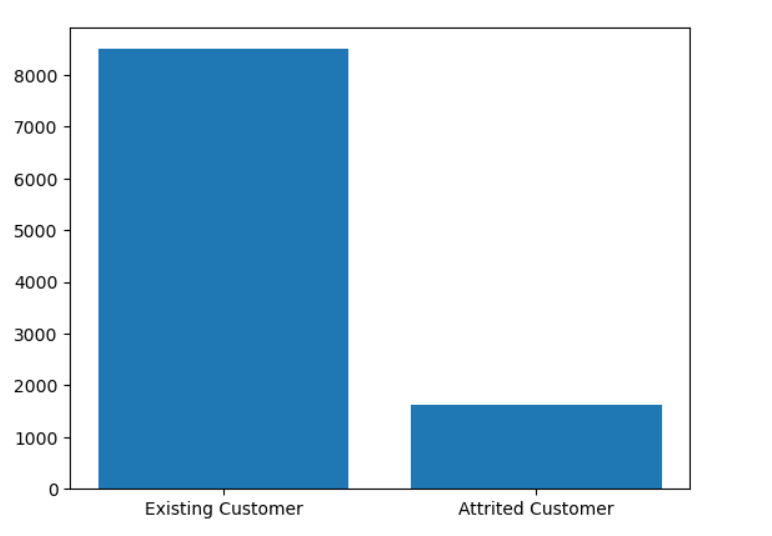
**2. Exploratory Data Analysis(EDA)**

Before proceeding to analyse the data column ‘CLIENTNUM’ was removed from the dataset as it is just acts as an identifier and doesn’t contribute to predict the target.

**2.1 Target Variable(Attrition\_Flag)**

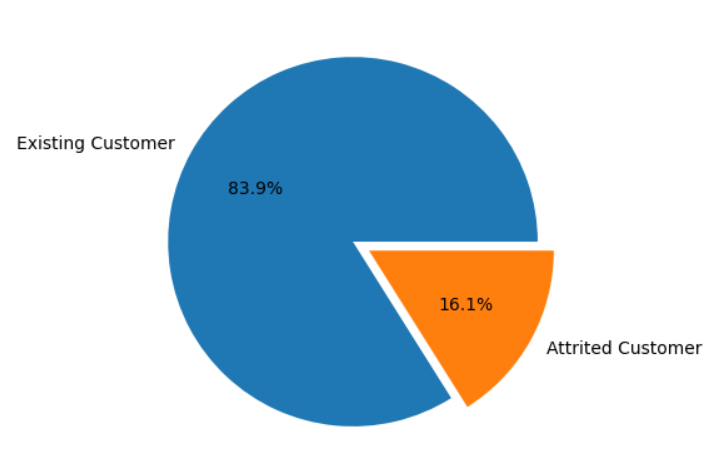
After conducting a thorough analysis of the dataset's column labels and accompanying information, it was determined that the "attrition flag" serves as the pivotal target variable within our dataset.

The Column is categorical in nature due to 2 values ‘Existing Customer’ and ‘Attrited Customer’. The frequency of occurrence of these values is shown below

****Frequency Distribution of Outcome in Target Variable

From the above bar graph it can seen that there is large difference between ‘Existing Customer’ and ‘Attrited Customer’.

So as to visualise the clear difference between the percentages of the 2 values of target variable the below graph is provided

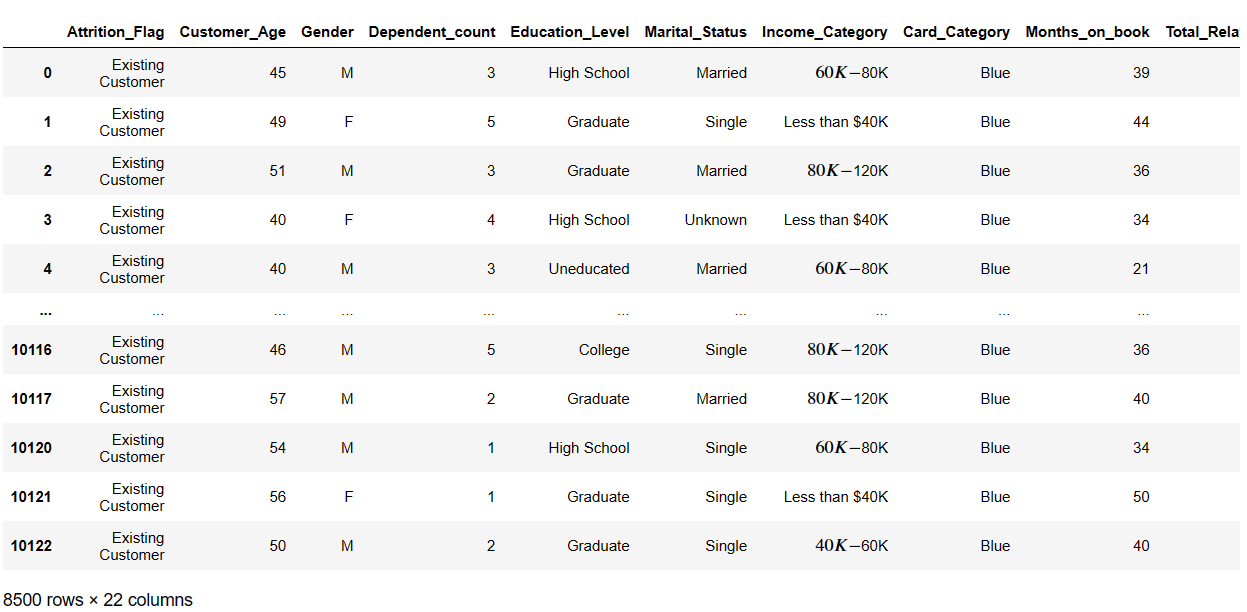


Pie Graph showing the frequency in Percentages

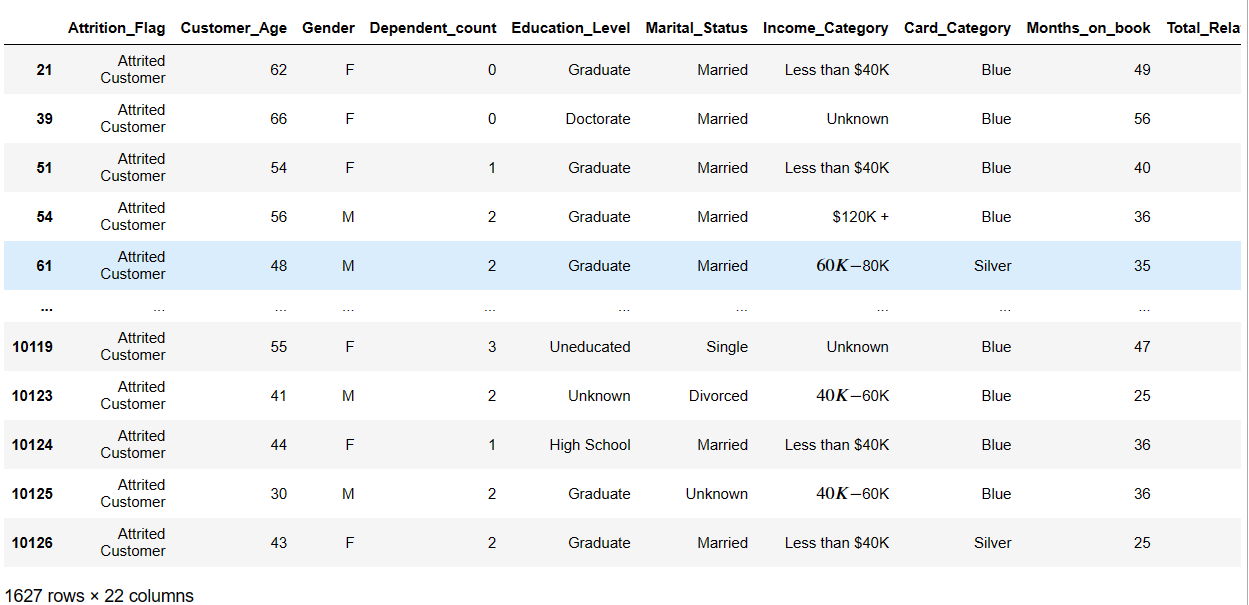
**2.2 Relationship between target variable and all the other factors**

We partitioned the dataset into two distinct segments: ‘Existing Customers’ and ‘Attrited Customers’. This division allowed for a more granular analysis of the data, facilitating the discovery of patterns and insights.

**Existing Customer Dataset**:

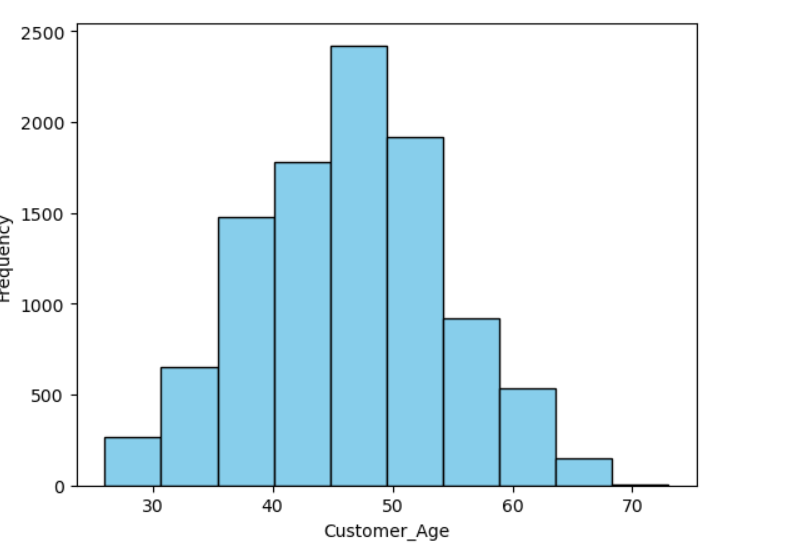


**Attrited Customer Dataset:**

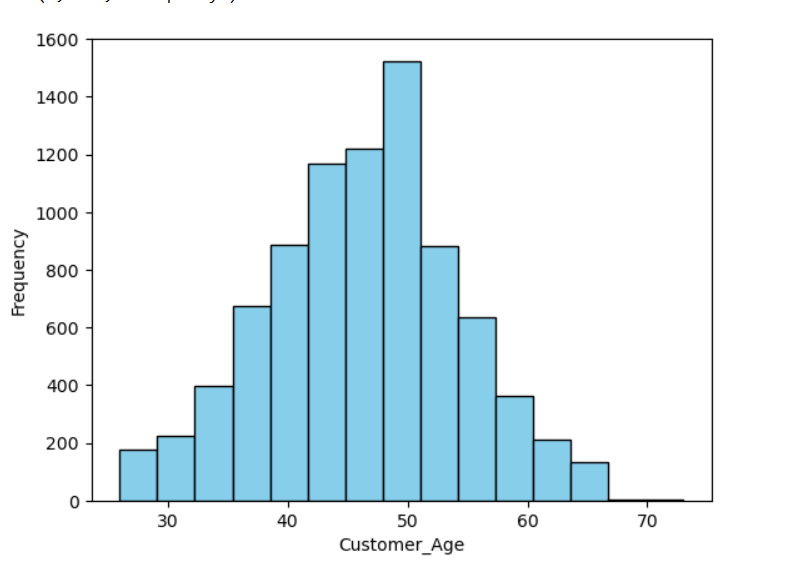


**2.2.1 Customer Age**

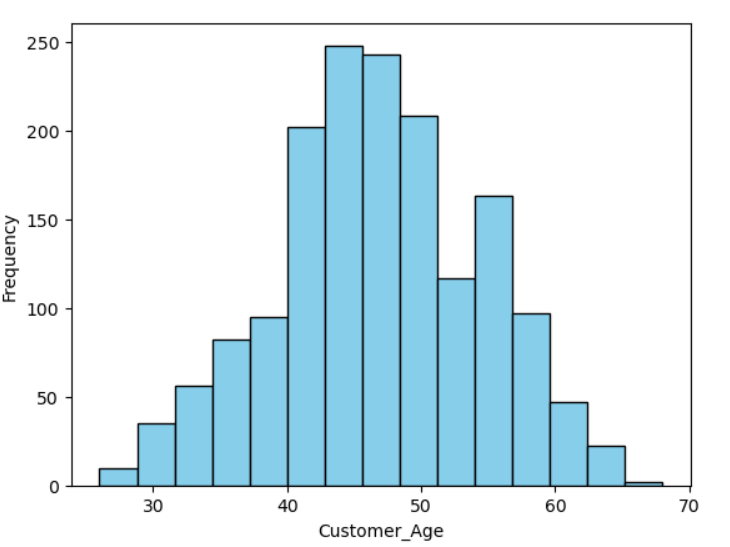
**Distribution of Age in Original Dataset:**

****

**Distribution of age for existing customers:**

****

**Distribution of age for Attrited customers:**

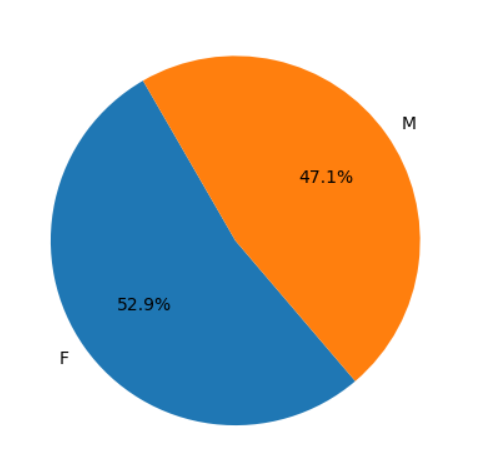
****

**Observations**

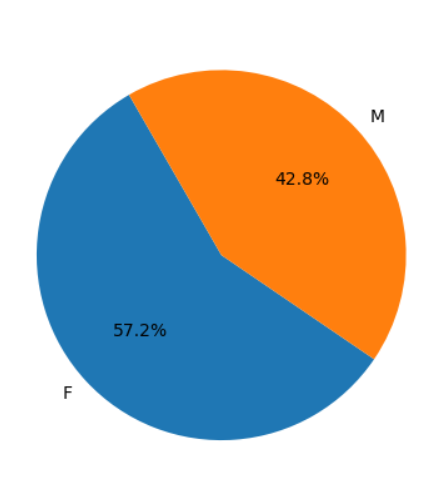
We observed a critical age range, specifically between 42 and 50, where existing customers exhibit a tendency to either maintain their relationship with the bank or transition into attrited customers. Focusing our efforts on this specific demographic segment presents an opportunity to mitigate credit card churn rates effectively.

**2.2.2 Gender Distribution Disparity in Customer Attrition**

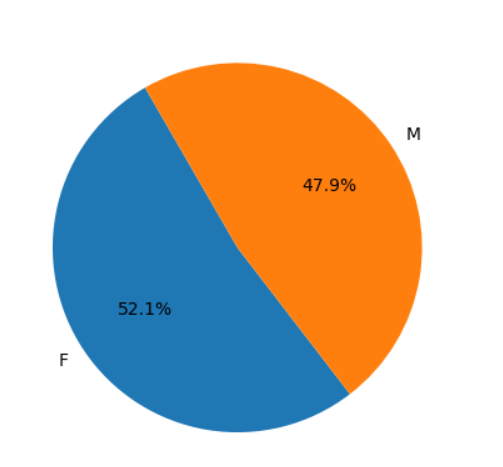
**Frequency Distribution of Gender in Original Data set:**

****

**Frequency Distribution of Gender in Original Data set Attrited Customers:**

****

**Frequency Distribution of Gender in Original Data set in Existing Customers :**

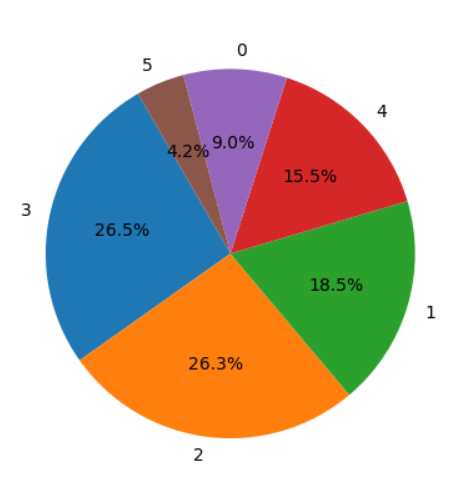
****

**Observations:**

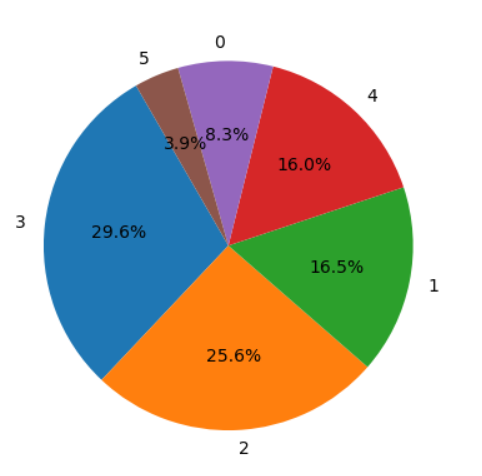
An observation from the gender distribution within the attrited and existing customer segments reveals a notable difference. Among attrited customers, females constitute 57.2%, while males account for 42.8%. In contrast, among existing customers, females make up 52.1%, with males at 47.9%. This suggests a higher attrition rate among female customers compared to males, indicating a potential area to be concentrated on.

**2.2.3 Impact of Dependent\_Count on Customer Attrition**

**Existing Customers:**

****

**Attrited Customers:**

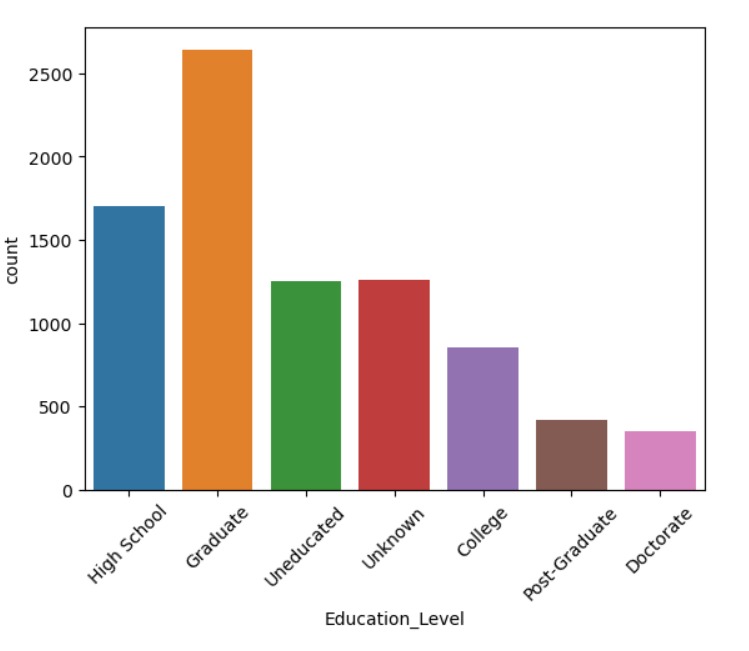
****

**Observations:**

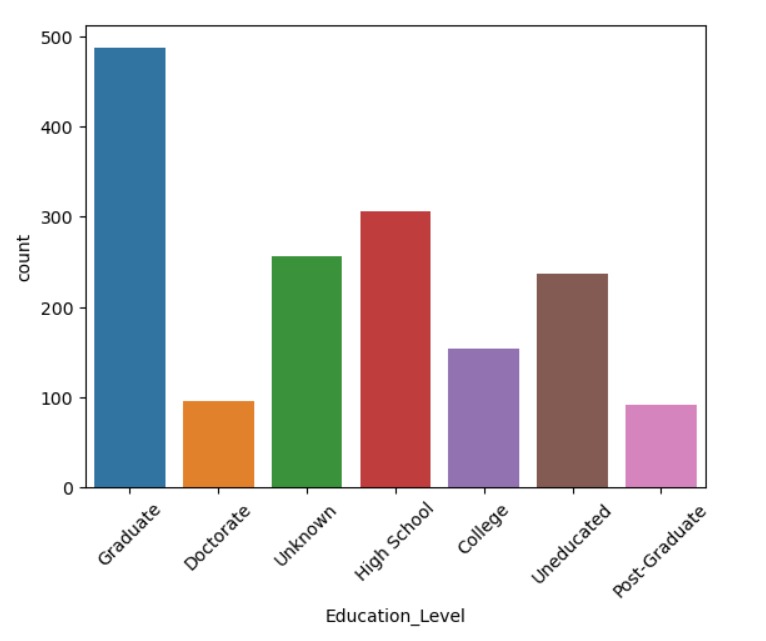
The comparison of percentages based on the number of dependents reveals a slight difference between existing and attrited customers. Among existing customers, 26.5% have three dependents, while among attrited customers, this percentage increases to 29.6%. While the difference may seem modest, it suggests that customers with three dependents are marginally more likely to churn.

**2.2.4 Impact of Education\_level on Customer Attrition**

**Existing Customers:**



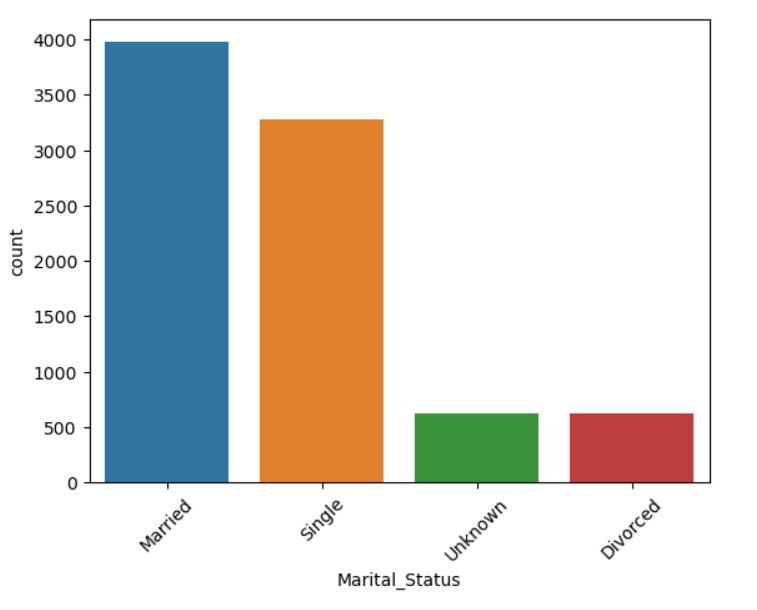
**Attrited Customers:**



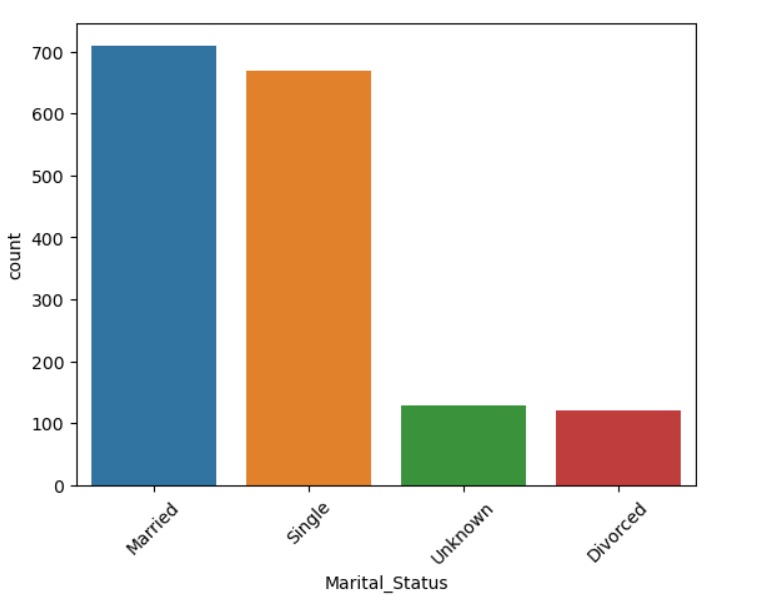
**­­­­­­­­­­­­­Observations:**

On observing the above graphs, which show the frequency distribution of educational levels of existing and attrited customers, we can conclude that the frequency percentage is almost similar for both kind of customers. So, we can conclude that this column does not affect the target variable. Hence, we can ignore this column while building the predictive model.

**2.2.5 Impact of Marital\_Status on Customer Attrition**



Existing Customers

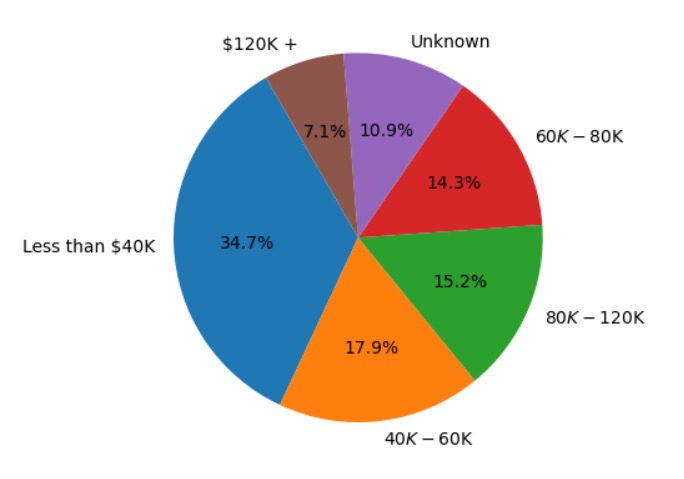


Attrited Customers

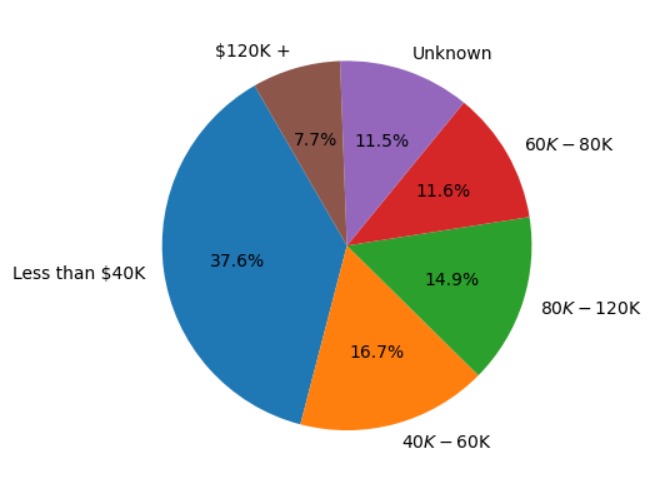
**Observations:**

On observing the above graphs, which depict the marital status of existing and attrited customers, we can see that the customers who are single have higher frequency of churning the credit card as the difference between singles being existing customer and attrited customers is high compared to other categories.

**2.2.6 Impact of Income\_Category on Customer Attrition**



Existing Customer

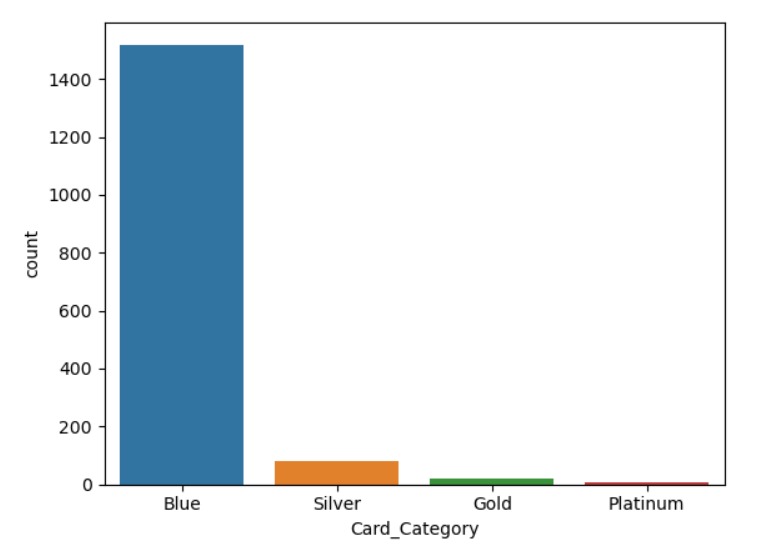


Attrited Customers

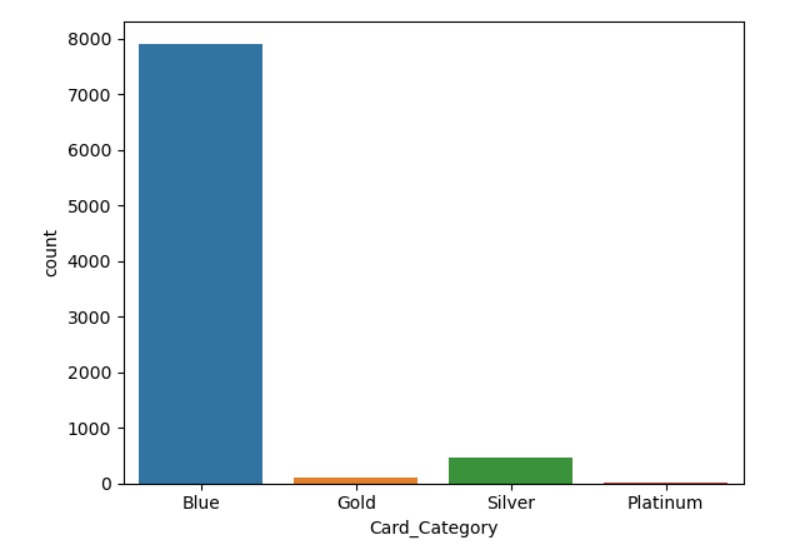
**Observations:**

The comparison of percentages based on the income reveals a slight difference between existing and attrited customers. Among existing customers and attrited customers, customers who have less that $40k income are more likely to churn compared to other income categories.

**2.2.7 Impact of Card\_Category on Customer Attrition**



Existing customers

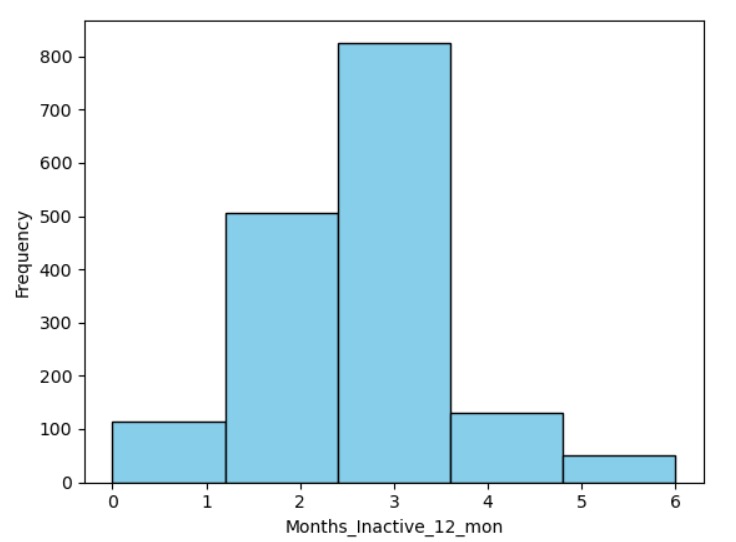


Attrited Customers

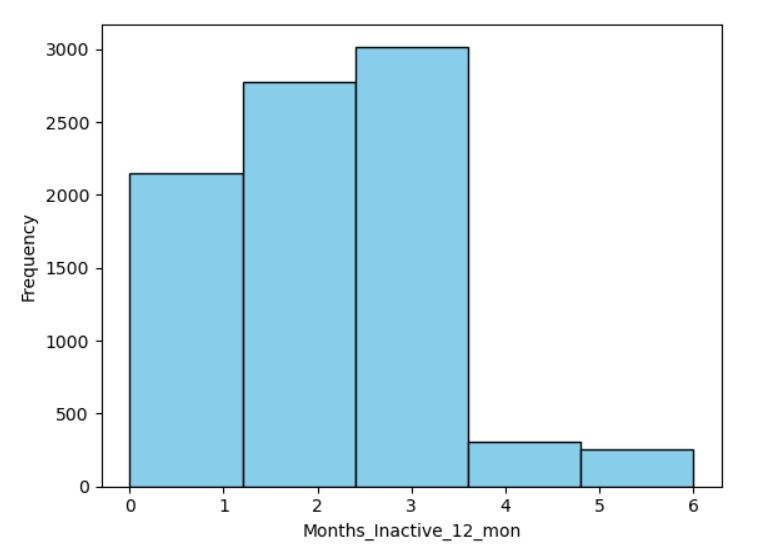
**Observations:**

Upon analysing the type of cards used by existing and attrited customers, we can conclude that both the cases have higher frequency for Blue category. And since there are no specific variations, we can exclude the card category while building the prediction model.

**2.2.7 Impact of Months\_Inactive on Customer Attrition**



Attrited Customers

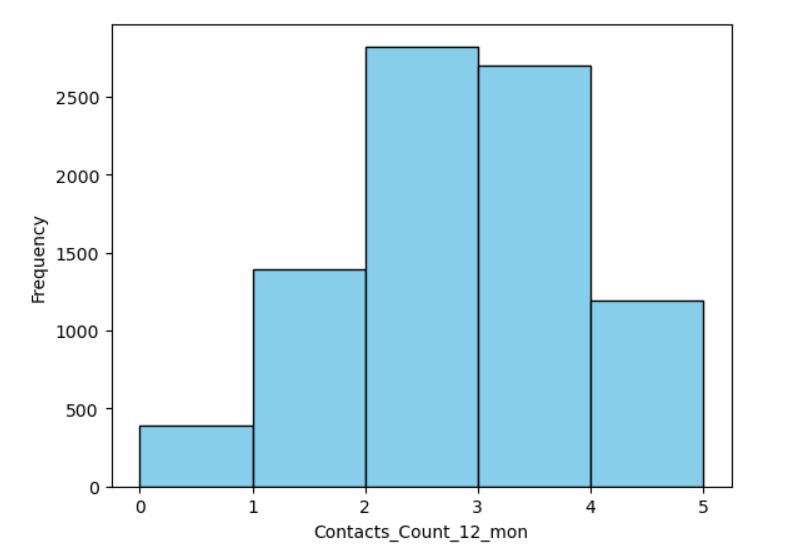


Existing Customers

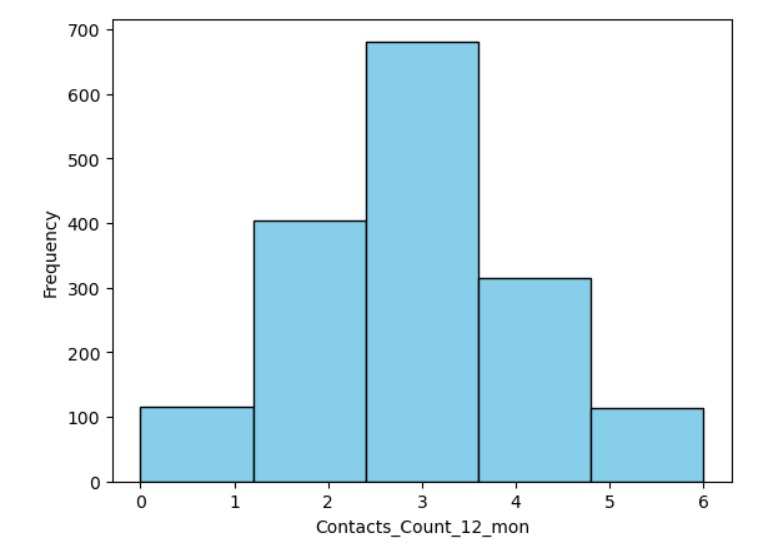
**Observations:**

On observing the above graphs, we can conclude that if a customer is inactive for 3 months, they have a higher chance of churning the card.

**2.2.7 Impact of Months\_Inactive on Customer Attrition**



Existing Customers

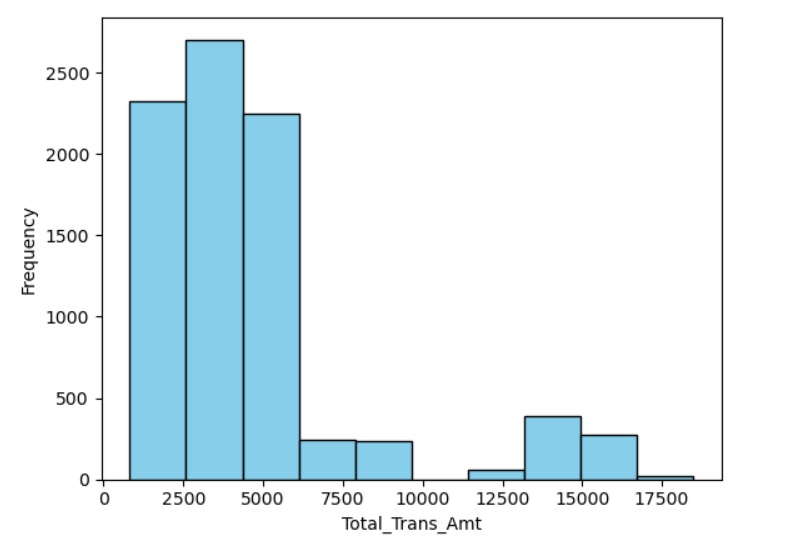


Attrited Customers

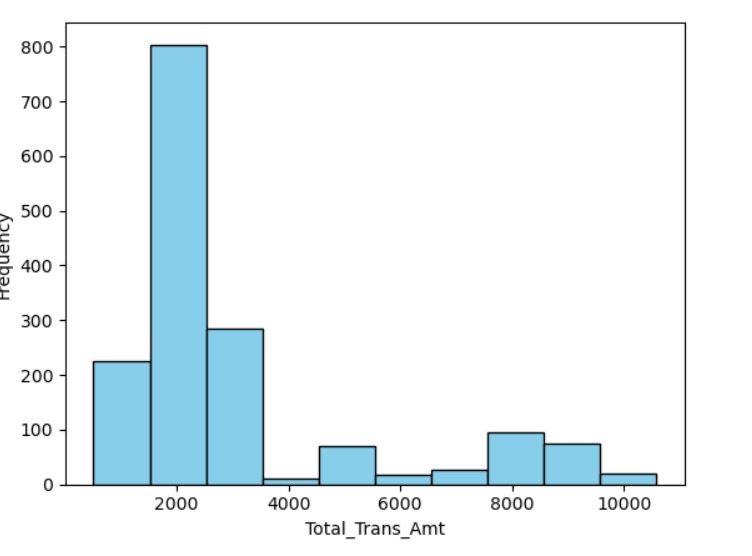
**Observations:**

Upon observing the above graphs, we get to know that the customers who have contacted the bank for 3 times have higher chances of churning compared to other categories.

**2.2.8 Impact of Total\_Transaction\_Amt on Customer Attrition**



Existing Customers



Attrited Customer

**Observations:**

From the above graphs, we see that both the graphs are left skewed. The customers having transaction amount around 2000 have higher chance of churning the credit card compared to the other transaction amounts.

**3. Building the Model using Logistic Regression**

Since the target variable, ‘Attrition\_Flag’ is categorical in nature, we use logistic regression tot build the predictive model.

Based on the observations derived from Exploratory Data Analysis (EDA), we have identified several variables that appear to significantly influence the target variable. These variables include:

1. Customer\_Age

2. Dependent\_count

3. Total\_Trans\_Ct

4. Months\_Inactive\_12\_mon

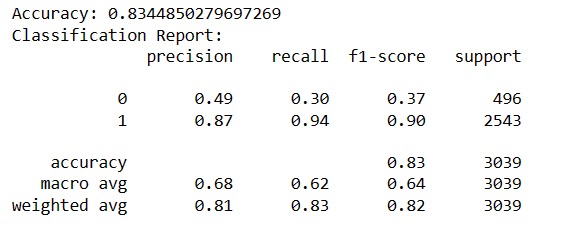
5. Contacts\_Count\_12\_mon

6. Total\_Trans\_Amt

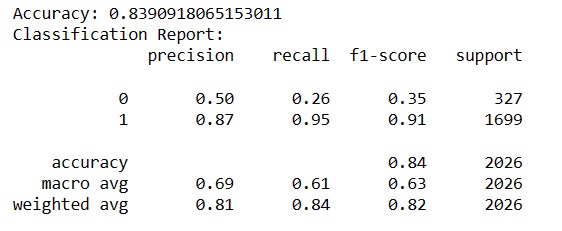
7. Avg\_Utilization\_Ratio

While building the predictive model, we analyzed the accuracy based on the ratio of splitting the data into training and testing set.

When the data is split into 70:30 ratio, the accuracy of the model is 0.834 approximately as shown in the figure:



When split in 80:20 ratio, the accuracy of the model is 0.839 which is almost same as the accuracy of the model with 70:30 split ratio.



Here, 0 represented the attrited customer, meanwhile 1 represents the existing customer.

**4. Conclusion**

Based on the analysis of the data and the construction of the predictive model using logistic regression, we can confidently conclude that the obtained model demonstrates strong performance. This conclusion is primarily based on its high accuracy rate, which serves as a reliable indicator of its effectiveness in making predictions.