```
#import the libraries that will use
import pandas as pd
from google.colab import files
import io
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn import preprocessing
import numpy as np
import copy
import matplotlib.style as style
import os
import math
from scipy import stats
```

Double-click (or enter) to edit

```
pip install dython
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/publ
Collecting dython
  Downloading dython-0.7.2-py3-none-any.whl (22 kB)
Requirement already satisfied: scikit-learn>=0.24.2 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: numpy>=1.19.5 in /usr/local/lib/python3.7/dist-packages (
Requirement already satisfied: seaborn>=0.11.0 in /usr/local/lib/python3.7/dist-packages
Collecting psutil>=5.9.1
  Downloading psutil-5.9.3-cp37-cp37m-manylinux_2_12_x86_64.manylinux2010_x86_64.manylin
     |████████████████████████████████| 291 kB 6.5 MB/s
Collecting scikit-plot>=0.3.7
  Downloading scikit_plot-0.3.7-py3-none-any.whl (33 kB)
```

```
pip install matplotlib==3.1.1
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheel
Collecting matplotlib==3.1.1
  Downloading matplotlib-3.1.1-cp37-cp37m-manylinux1_x86_64.whl (13.1 MB)
     |████████████████████████████████| 13.1 MB 3.1 MB/s
Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/di
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /usr/lo
Requirement already satisfied: numpy>=1.11 in /usr/local/lib/python3.7/dist-packag
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.7/dist-
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages
Installing collected packages: matplotlib
  Attempting uninstall: matplotlib
    Found existing installation: matplotlib 3.5.3
    Uninstalling matplotlib-3.5.3:
      Successfully uninstalled matplotlib-3.5.3
ERROR: pip's dependency resolver does not currently take into account all the pack
dython 0.7.2 requires matplotlib>=3.4.3, but you have matplotlib 3.1.1 which is in
Successfully installed matplotlib-3.1.1
```

```
Successfully installed dython-0.7.2 fonttools-4.38.0 matplotlib-3.5.3 psutil-5.9.3 scik
```

```
from dython.nominal import associations
```

Loading the Dataset

```
# Read the data with the Pandas libray in this stage
data = pd.read_csv('https://raw.githubusercontent.com/AlaaAli968/Bank-Churn/main/BankChurners
```

Describe the Data

```
# To check the datatypes as we can see do not have any null value
```

```
data.info()
```

```
<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   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
 20  Avg_Utilization_Ratio     10127 non-null  float64
dtypes: float64(5), int64(10), object(6)
memory usage: 1.6+ MB
```

```
# To check the data we can use the head() function to see first 5 rows.
data.head()
```

	CLIENTNUM	Attrition_Flag	Customer_Age	Gender	Dependent_count	Education_Lev
0	768805383	Existing Customer	45	M	3	High Sch
1	818770008	Existing Customer	49	F	5	Gradu
2	713982108	Existing Customer	51	M	3	Gradu
3	769911858	Existing Customer	40	F	4	High Sch
4	709106358	Existing Customer	40	M	3	Uneduca

5 rows × 21 columns

```
d=data.describe().T
d
```

	count	mean	std	min	
CLIENTNUM	10127.0	7.391776e+08	3.690378e+07	708082083.0	7.130368e
Customer_Age	10127.0	4.632596e+01	8.016814e+00	26.0	4.100000e
Dependent_count	10127.0	2.346203e+00	1.298908e+00	0.0	1.000000e
Months_on_book	10127.0	3.592841e+01	7.986416e+00	13.0	3.100000e
Total_Relationship_Count	10127.0	3.812580e+00	1.554408e+00	1.0	3.000000e
Months_Inactive_12_mon	10127.0	2.341167e+00	1.010622e+00	0.0	2.000000e
Contacts_Count_12_mon	10127.0	2.455317e+00	1.106225e+00	0.0	2.000000e
Credit_Limit	10127.0	8.631954e+03	9.088777e+03	1438.3	2.555000e
Total_Revolving_Bal	10127.0	1.162814e+03	8.149873e+02	0.0	3.590000e
Avg_Open_To_Buy	10127.0	7.469140e+03	9.090685e+03	3.0	1.324500e
Total_Amt_Chng_Q4_Q1	10127.0	7.599407e-01	2.192068e-01	0.0	6.310000e
Total_Trans_Amt	10127.0	4.404086e+03	3.397129e+03	510.0	2.155500e
Total_Trans_Ct	10127.0	6.485869e+01	2.347257e+01	10.0	4.500000e
Total_Ct_Chng_Q4_Q1	10127.0	7.122224e-01	2.380861e-01	0.0	5.820000e
Avg_Utilization_Ratio	10127.0	2.748936e-01	2.756915e-01	0.0	2.300000e

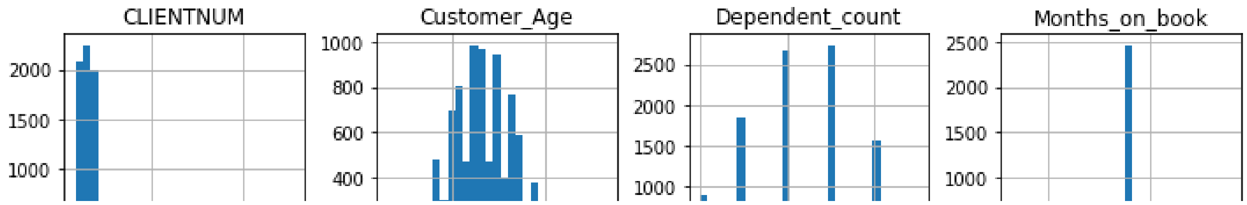
Preparing of Data

```
# Double check for the null-values
data.isnull().sum()
```

```
CLIENTNUM                   0
Attrition_Flag              0
Customer_Age                0
Gender                      0
Dependent_count             0
Education_Level             0
Marital_Status              0
Income_Category             0
```

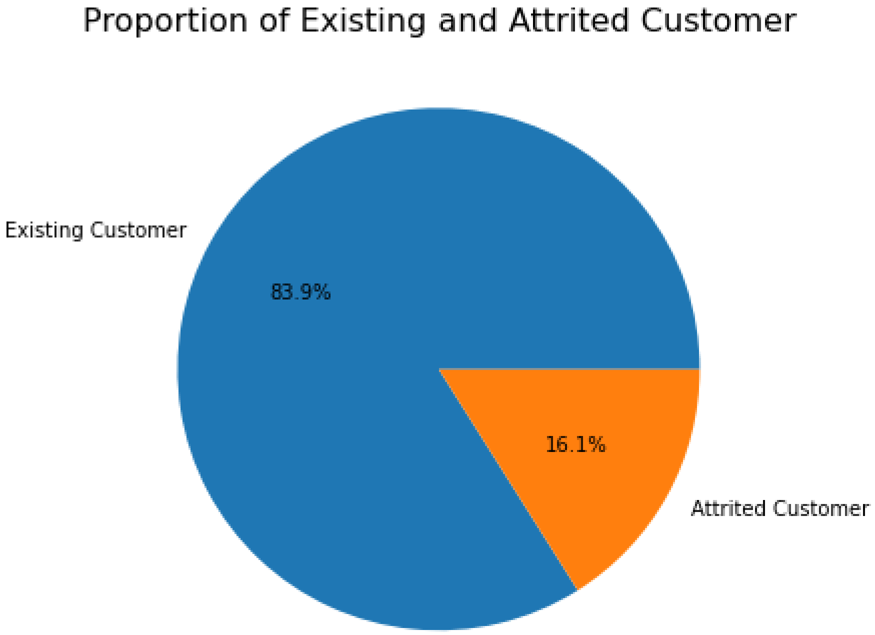
```
Card_Category               0
Months_on_book              0
Total_Relationship_Count    0
Months_Inactive_12_mon      0
Contacts_Count_12_mon       0
Credit_Limit                0
Total_Revolving_Bal         0
Avg_Open_To_Buy             0
Total_Amt_Chng_Q4_Q1        0
Total_Trans_Amt             0
Total_Trans_Ct              0
Total_Ct_Chng_Q4_Q1         0
Avg_Utilization_Ratio       0
dtype: int64
```

```
# distribution of numerical features
axList = data.hist(bins=29, figsize = (12, 12))
plt.savefig("Hist.png")
```



```
table=data['Attrition_Flag'].value_counts(normalize=True) * 100
print(table)
churn=data['Attrition_Flag'].value_counts()
churn
plt.figure(figsize = (6,6))
piechart=plt.pie(x=churn,labels=churn.keys(),autopct="%.1f%%")
plt.title('Proportion of Existing and Attrited Customer', fontsize = 16)
# as we see data is  imbalanced so we will apply some techniques to balance it before running
```

```
Existing Customer    83.934038
Attrited Customer    16.065962
Name: Attrition_Flag, dtype: float64
Text(0.5, 1.0, 'Proportion of Existing and Attrited Customer')
```



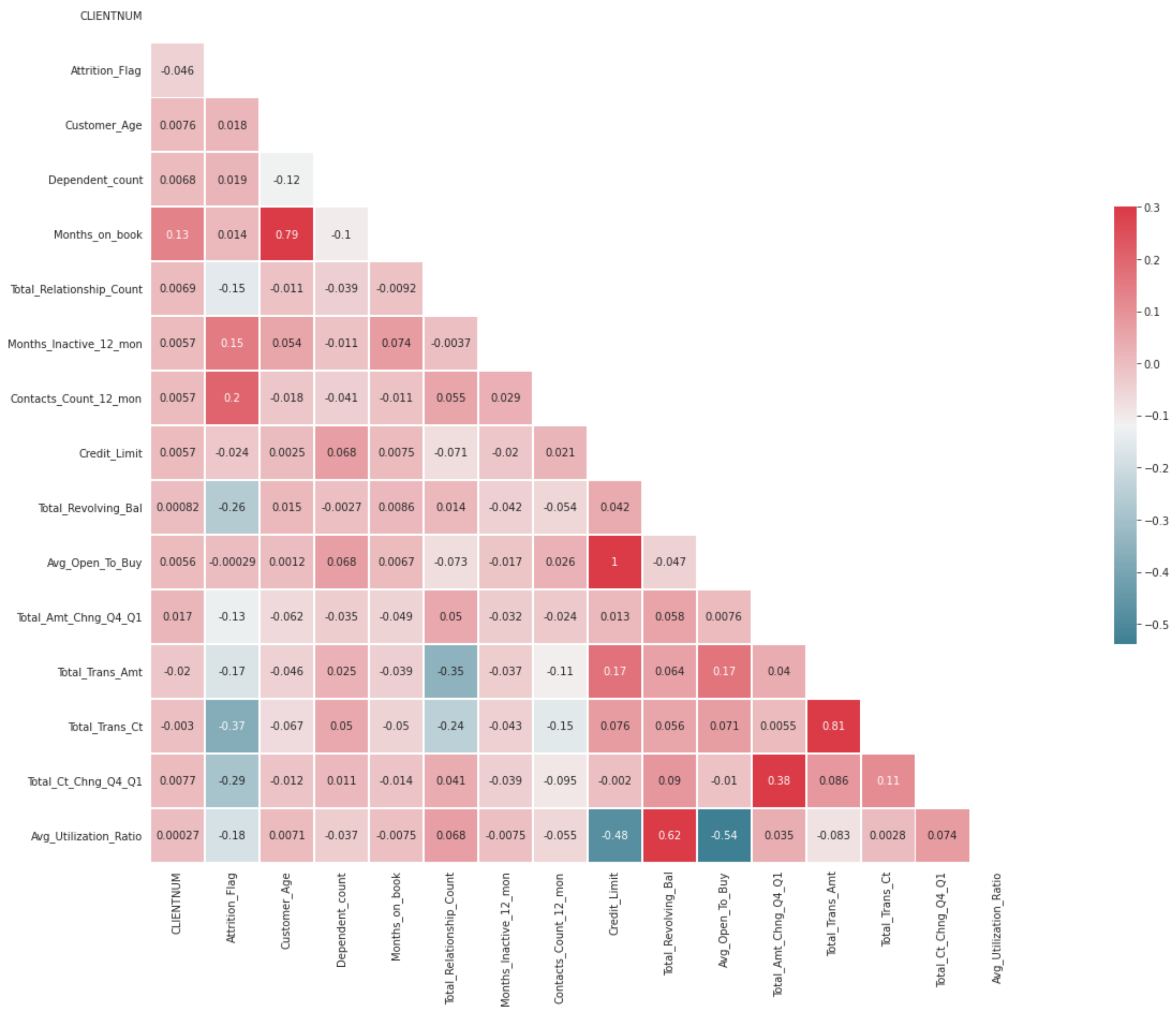
```
#convert Attrition_Flag to int to prepare it for the corrlation
new_data= copy.copy(data)

new_data['Attrition_Flag'].replace({'Existing Customer':0, 'Attrited Customer':1}, inplace=Tr

new_data.head()
```

	CLIENTNUM	Attrition_Flag	Customer_Age	Gender	Dependent_count	Education_Lev
0	768805383	0	45	M	3	High Sch
1	818770008	0	49	F	5	Gradu
2	713982108	0	51	M	3	Gradu
3	769911858	0	40	F	4	High Sch
4	709106358	0	40	M	3	Uneduca

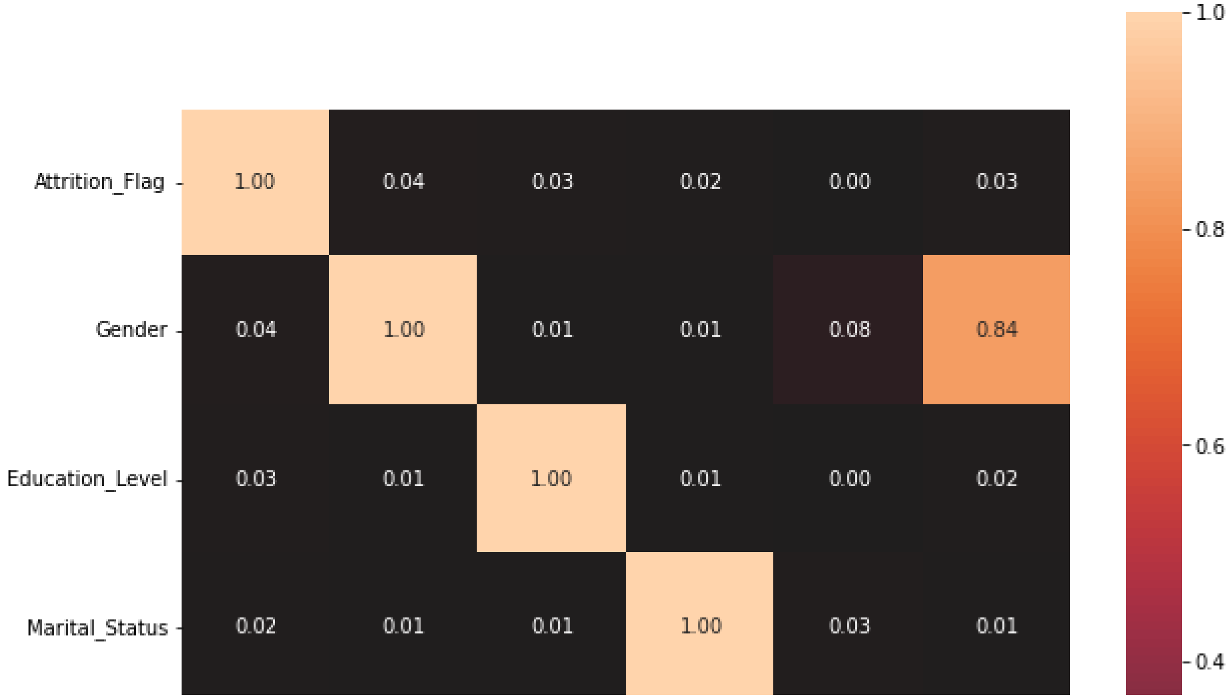
5 rows × 21 columns

New Section

```
corr = new_data.corr()
corr

mask = np.zeros_like(corr)
mask[np.triu_indices_from(mask)] = True

# Generate a custom diverging colormap
cmap = sns.diverging_palette(220, 10, as_cmap=True)

with sns.axes_style("white"):
    # Set up the matplotlib figure
    f, ax = plt.subplots(figsize=(30, 15))
    ax = sns.heatmap(corr, cmap=cmap, mask=mask, vmax=.3, square=True, linewidths=.9, cbar_kw
```

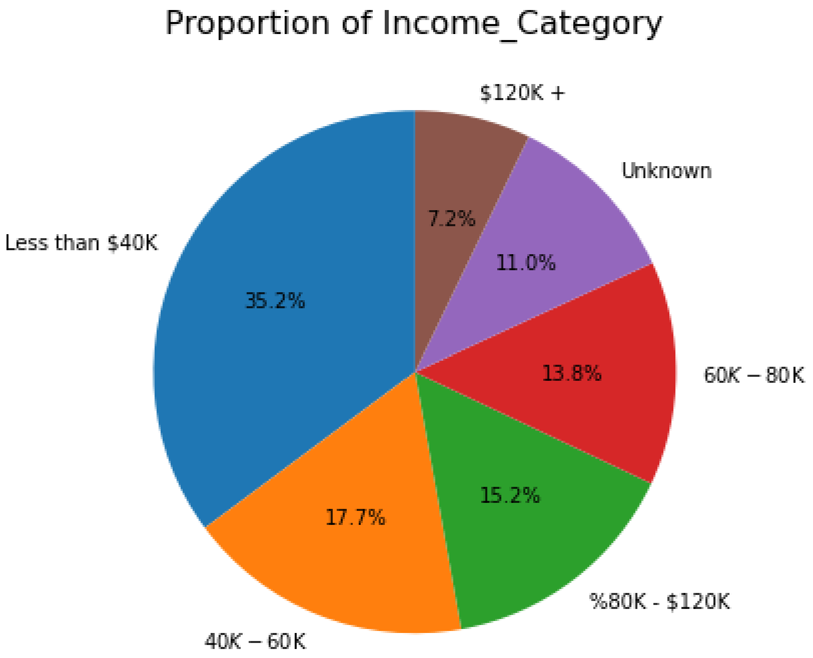


```
#correlation for categorical data
selected_column= data[["Attrition_Flag","Gender","Education_Level","Marital_Status","Card_Cat
categorical_df = selected_column.copy()
categorical_correlation= associations(categorical_df, filename= 'categorical_correlation.png'
```

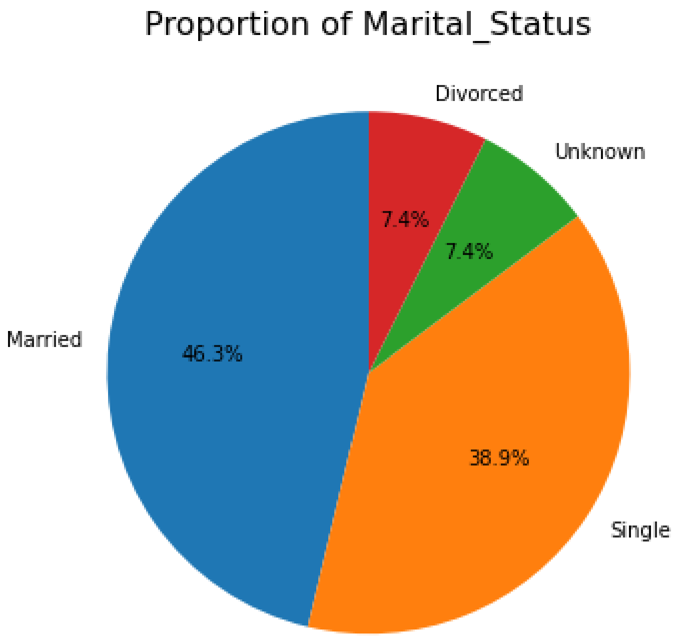


Analysis and Visualization of the Data

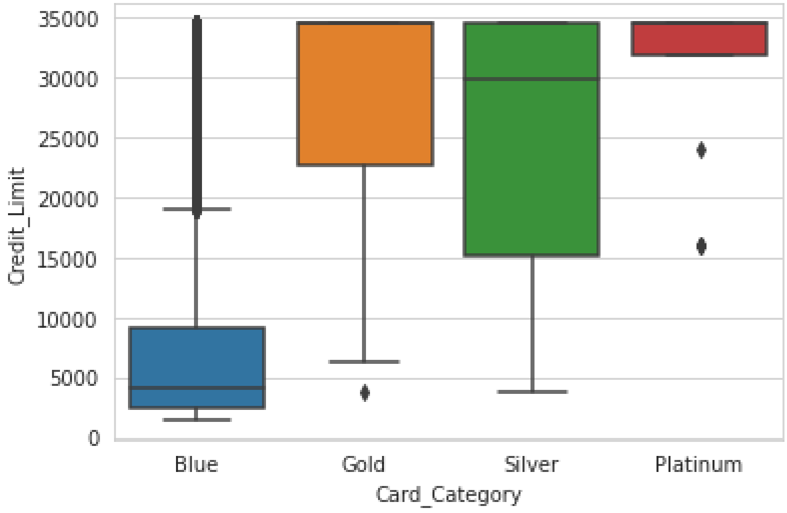
```
# Customers' Distribution  based on Education Level
plt.figure(figsize = (6,6))
plt.pie(data['Income_Category'].value_counts(),
        labels = ['Less than $40K', '$40K - $60K', '%80K - $120K', '$60K - $80K', 'Unknown','
        autopct='%1.1f%%', startangle = 90)
plt.title('Proportion of Income_Category', fontsize = 16)
plt.show()
```



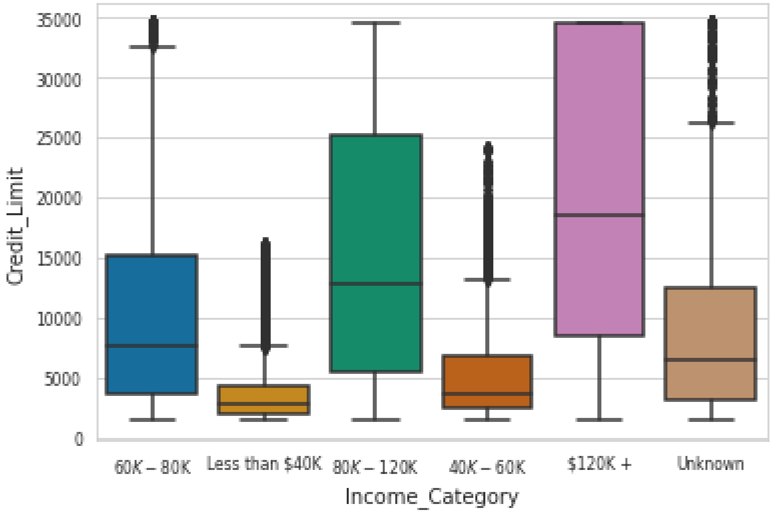
```
# Customers' Distribution  based on Education Level
plt.figure(figsize = (6,6))
plt.pie(data['Marital_Status'].value_counts(),
        labels = ['Married', 'Single', 'Unknown', 'Divorced'],
        autopct='%1.1f%%', startangle = 90)
plt.title('Proportion of Marital_Status', fontsize = 16)
plt.show()
```



```
#Card_Category and Credit_Limit
sns.set_style("whitegrid")
ss=sns.boxplot(x = 'Card_Category', y = 'Credit_Limit', data = data)
```

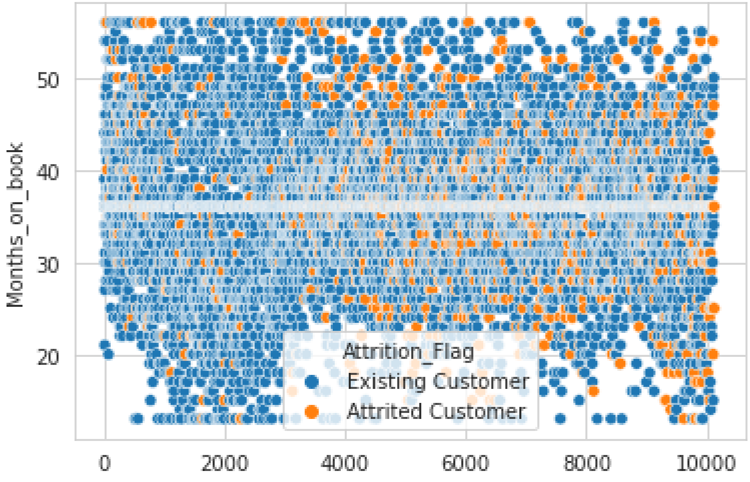


```
#Income_Category and Credit_Limit
sns.set_style("whitegrid")
s=sns.boxplot(x = 'Income_Category', y = 'Credit_Limit', data = data,palette="colorblind")
s.tick_params(labelsize=7.5)
```

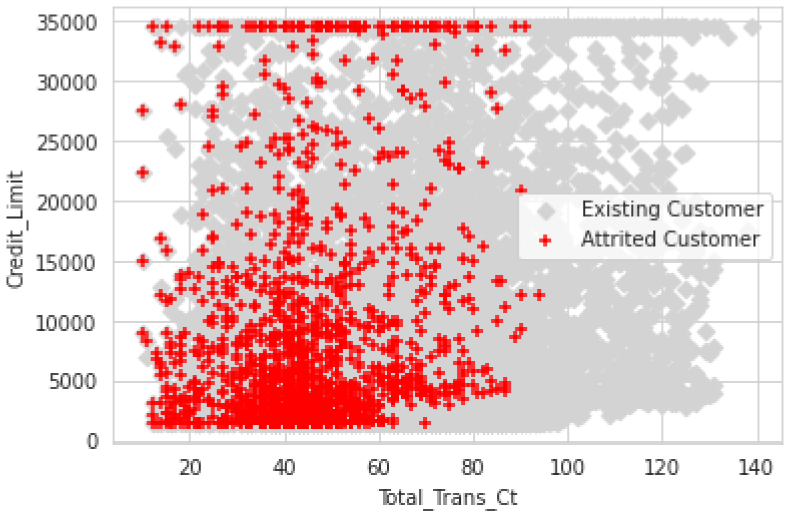


```
import seaborn as sns
sns.scatterplot(data=data, x=data.index, y='Months_on_book', hue='Attrition_Flag')
```

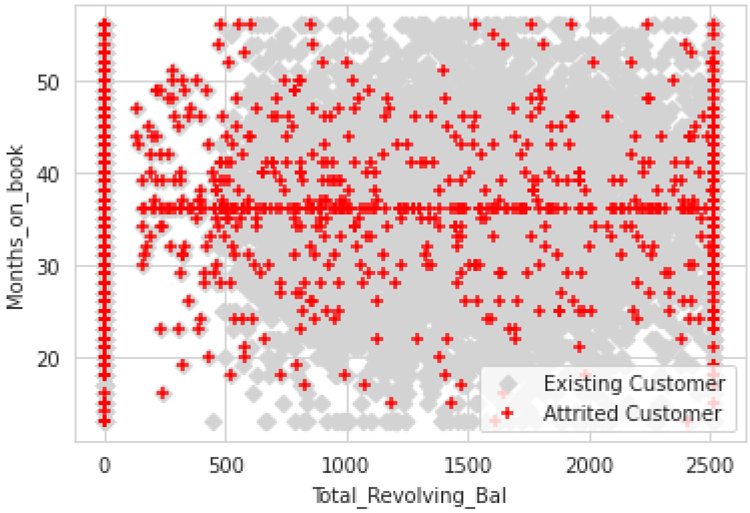
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f29ba462990>
```



```
#Existing Customer Attrited Customer
plt.scatter(data['Total_Trans_Ct'][(data.Attrition_Flag == 'Existing Customer') | (data.Attri
            data['Credit_Limit'][(data.Attrition_Flag == 'Existing Customer') | (data.Attriti
           marker='D',
           color='lightgray',
           label='Existing Customer')
plt.scatter(data['Total_Trans_Ct'][data.Attrition_Flag == 'Attrited Customer'],
            data['Credit_Limit'][data.Attrition_Flag == 'Attrited Customer'],
           marker='+',
           color='red',
           label='Attrited Customer')
plt.xlabel('Total_Trans_Ct')
plt.ylabel('Credit_Limit')
plt.legend()
plt.show()
```



```
plt.scatter(data['Total_Revolving_Bal'][(data.Attrition_Flag == 'Existing Customer') | (data.
            data['Months_on_book'][(data.Attrition_Flag == 'Existing Customer') | (data.Attri
           marker='D',
           color='lightgray',
           label='Existing Customer')
plt.scatter(data['Total_Revolving_Bal'][data.Attrition_Flag == 'Attrited Customer'],
            data['Months_on_book'][data.Attrition_Flag == 'Attrited Customer'],
           marker='+',
           color='red',
           label='Attrited Customer')
plt.xlabel('Total_Revolving_Bal')
plt.ylabel('Months_on_book')
plt.legend()
plt.show()
```


Months_on_book
Total_Revolving_Bal
Existing Customer
Attrited Customer