This data is from Kaggle, used to predict churn. The origin is https://www.kaggle.com/datasets/sakshigoyal7/credit-card-customers/data

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns


# Load bank_churners data
churners_data = pd.read_csv('BankChurners.csv')


# Display the first few rows of the dataframe to understand its structure
print("First few rows of the churners data:")
print(churners_data.head())
```

```
First few rows of the churners data:
   CLIENTNUM       Attrition_Flag Customer_Age Gender  Dependent_count  \
0  768805383  Existing Customer            45      M                3
1  818770008  Existing Customer            49      F                5
2  713982108  Existing Customer            51      M                3
3  769911858  Existing Customer            40      F                4
4  709106358  Existing Customer            40      M                3

  Education_Level Marital_Status Income_Category Card_Category  \
0     High School        Married      $60K - $80K          Blue
1        Graduate         Single  Less than $40K          Blue
2        Graduate        Married     $80K - $120K          Blue
3     High School        Unknown  Less than $40K          Blue
4      Uneducated        Married      $60K - $80K          Blue

   Months_on_book  ...  Credit_Limit  Total_Revolving_Bal  Avg_Open_To_Buy  \
0              39  ...       12691.0                  777          11914.0
1              44  ...        8256.0                  864           7392.0
2              36  ...        3418.0                    0           3418.0
3              34  ...        3313.0                 2517            796.0
4              21  ...        4716.0                    0           4716.0

   Total_Amt_Chng_Q4_Q1  Total_Trans_Amt  Total_Trans_Ct  Total_Ct_Chng_Q4_Q1  \
0                 1.335             1144              42                1.625
1                 1.541             1291              33                3.714
2                 2.594             1887              20                2.333
3                 1.405             1171              20                2.333
4                 2.175              816              28                2.500

   Avg_Utilization_Ratio  \
0                  0.061
1                  0.105
2                  0.000
3                  0.760
4                  0.000

   Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_Level_Months_Inactive_12_mon
0                                             0.000093
1                                             0.000057
2                                             0.000021
3                                             0.000134
4                                             0.000022

   Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_Level_Months_Inactive_12_mon
0                                              0.99991
1                                              0.99994
2                                              0.99998
3                                              0.99987
4                                              0.99998

[5 rows x 23 columns]
```

⌄ Statistics Summary

```python
# Summary statistics of numerical columns
print("Summary statistics of numerical columns:")
print(churners_data.describe())
```

```
Summary statistics of numerical columns:
           CLIENTNUM  Customer_Age  Dependent_count  Months_on_book  \
count   1.012700e+04  10127.000000     10127.000000    10127.000000
mean    7.391776e+08     46.325960         2.346203       35.928409
std     3.690378e+07      8.016814         1.298908        7.986416
min     7.080821e+08     26.000000         0.000000       13.000000
25%     7.130368e+08     41.000000         1.000000       31.000000
```

```
50%     7.179264e+08     46.000000        2.000000       36.000000
75%     7.731435e+08     52.000000        3.000000       40.000000
max     8.283431e+08     73.000000        5.000000       56.000000

        Total_Relationship_Count  Months_Inactive_12_mon  \
count              10127.000000            10127.000000
mean                   3.812580                2.341167
std                    1.554408                1.010622
min                    1.000000                0.000000
25%                    3.000000                2.000000
50%                    4.000000                2.000000
75%                    5.000000                3.000000
max                    6.000000                6.000000

        Contacts_Count_12_mon  Credit_Limit  Total_Revolving_Bal  \
count            10127.000000  10127.000000         10127.000000
mean                 2.455317   8631.953698          1162.814061
std                  1.106225   9088.776650           814.987335
min                  0.000000   1438.300000             0.000000
25%                  2.000000   2555.000000           359.000000
50%                  2.000000   4549.000000          1276.000000
75%                  3.000000  11067.500000          1784.000000
max                  6.000000  34516.000000          2517.000000

        Avg_Open_To_Buy  Total_Amt_Chng_Q4_Q1  Total_Trans_Amt  Total_Trans_Ct  \
count      10127.000000          10127.000000     10127.000000    10127.000000
mean        7469.139637              0.759941      4404.086304       64.858695
std         9090.685324              0.219207      3397.129254       23.472570
min            3.000000              0.000000       510.000000       10.000000
25%         1324.500000              0.631000      2155.500000       45.000000
50%         3474.000000              0.736000      3899.000000       67.000000
75%         9859.000000              0.859000      4741.000000       81.000000
max        34516.000000              3.397000     18484.000000      139.000000

        Total_Ct_Chng_Q4_Q1  Avg_Utilization_Ratio  \
count          10127.000000           10127.000000
mean               0.712222               0.274894
std                0.238086               0.275691
min                0.000000               0.000000
25%                0.582000               0.023000
50%                0.702000               0.176000
75%                0.818000               0.503000
max                3.714000               0.999000

        Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_Level_Months_Inactive
count                                        10127.000000
mean                                             0.159997
std                                              0.365301
min                                              0.000008
```

Summary based on the provided numerical columns:

Client Statistics:

- Mean client number: 7.391776e+08
- Mean customer age: 46.33 years
- Mean dependent count: 2.35
- Mean months on book: 35.93
- Mean total relationship count: 3.81
- Mean months inactive in the last 12 months: 2.34
- Mean contacts count in the last 12 months: 2.46

Credit Statistics:

- Mean credit limit: 8631.95
- Mean total revolving balance: 1162.81
- Mean average open to buy: 7469.14
- Mean total amount change Q4-Q1: 0.76
- Mean total transaction amount: 4404.09
- Mean total transaction count: 64.86
- Mean total count change Q4-Q1: 0.71
- Mean average utilization ratio: 0.27

Naive Bayes Classifier Statistics:

- Mean Naive Bayes Classifier (Attrition Flag - Card Category - Contacts Count 12 mon - Dependent count - Education Level - Months Inactive 12 mon - 1): 0.159997
- Mean Naive Bayes Classifier (Attrition Flag - Card Category - Contacts Count 12 mon - Dependent count - Education Level - Months Inactive 12 mon - 2): 0.840003

These statistics provide an overview of the numerical features in the dataset, including client demographics, credit-related information, and classifier statistics.

```
# Check for missing values
print("Missing values:")
print(churners_data.isnull().sum())
```

```
        Missing values:
        CLIENTNUM
        Attrition_Flag
        Customer_Age
        Gender
        Dependent_count
        Education_Level
        Marital_Status
        Income_Category
        Card_Category
        Months_on_book
        Total_Relationship_Count
        Months_Inactive_12_mon
        Contacts_Count_12_mon
        Credit_Limit
        Total_Revolving_Bal
        Avg_Open_To_Buy
        Total_Amt_Chng_Q4_Q1
        Total_Trans_Amt
        Total_Trans_Ct
        Total_Ct_Chng_Q4_Q1
        Avg_Utilization_Ratio
        Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_Level_Months_Inactive_12_mon_1
        Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_Level_Months_Inactive_12_mon_2
        dtype: int64
```

˅ Correlations Matrix

```
# Correlation matrix
correlation_matrix = churners_data.corr()
print("Correlation matrix:")
print(correlation_matrix)
```

```
Months_Inactive_12_mon                                                    0.155492
Contacts_Count_12_mon                                                    -0.205772
Credit_Limit                                                              0.023853
Total_Revolving_Bal                                                       0.263032
Avg_Open_To_Buy                                                           0.000267
Total_Amt_Chng_Q4_Q1                                                      0.131207
Total_Trans_Amt                                                          0.168642
Total_Trans_Ct                                                           0.371403
Total_Ct_Chng_Q4_Q1                                                      0.290115
Avg_Utilization_Ratio                                                    0.178405
Naive_Bayes_Classifier_Attrition_Flag_Card_Cate...                      -1.000000
Naive_Bayes_Classifier_Attrition_Flag_Card_Cate...                       1.000000
<ipython-input-10-ca4bdf0fda51>:2: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future
  correlation_matrix = churners_data.corr()
```

The provided correlation matrix represents the pairwise correlations between numerical features in the dataset. Here's a summary:

Positive Correlations:

- Customer Age is positively correlated with Months on Book (0.788912).
- Total Transaction Amount is positively correlated with Total
- Transaction Count (0.807192).
- Total Revolving Balance is positively correlated with Avg Utilization Ratio (0.624022).

Negative Correlations:

- Avg Utilization Ratio is negatively correlated with Credit Limit (-0.482965).
- Total Relationship Count is negatively correlated with Total
- Transaction Amount (-0.347229).
- Total Transaction Count is negatively correlated with Naive Bayes Classifier (Attrition Flag - Card Category - Contacts Count 12 mon - Dependent count - Education Level - Months Inactive 12 mon - 1) (-0.371403).
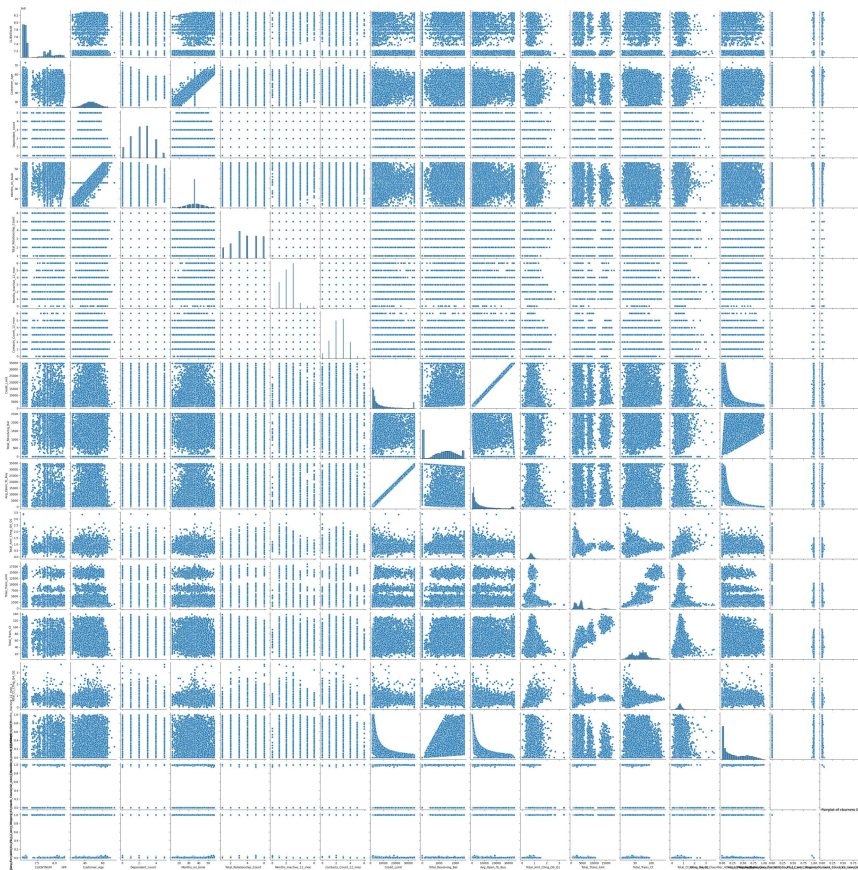
Weak Correlations:

There are weak correlations (close to 0) between many pairs of features, indicating little linear relationship between them. This correlation matrix provides valuable insights into how features are related to each other within the dataset. However, it's important to note that correlation doesn't imply causation, and further analysis may be needed to understand the relationships between variables more deeply.

```
# Visualization: Pairplot
sns.pairplot(churners_data)
plt.title('Pairplot of churners Data')
plt.show()
```

```
# Visualization: Correlation Heatmap
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Heatmap of churners Data')
plt.show()
```

Correlation Heatmap of churners Data