

Credit Card Fraud Detection Project Results

Data Set Description:

The dataset includes transactions made by credit card holders. It consists of 284,807 transactions, out of which only 492 transactions are marked as fraudulent (0.172%).

Project Objectives:

Objective 1: Perform in-depth analysis on the dataset to identify potential fraudulent transactions. Objective 2: Visualize and compare fraudulent and genuine transactions based on various features. Objective 3: Implement machine learning models to detect fraudulent activities and evaluate their performance metrics. Objective 4: Handle class imbalances using sampling techniques or class weights to improve model performance.

Technologies Used:

Python (Libraries: Pandas, NumPy, Matplotlib, Seaborn) Machine Learning Libraries (Scikit-learn, TensorFlow, Keras, etc.) Data Visualization Tools

Project Steps:

Data Exploration and Preprocessing:

Understand and preprocess the dataset, dealing with missing values and outliers. Identify features that differentiate fraudulent and genuine transactions.

Visualize fraudulent and genuine transaction.

Apply machine learning algorithms to train the dataset. Evaluate model performance using metrics such as accuracy, precision, recall, and F1 score. Perform hyperparameter tuning and overfitting prevention techniques.

Test the trained model on real data to assess its ability to correctly identify fraudulent transactions. Review and focus on improving the model's performance.

Periodically update the model with new data to create a more resilient model against evolving fraudulent tactics.

Implement appropriate measures to ensure data security and privacy due to the sensitive nature of the data.

Explanations:

1. **Data Preprocessing:** The preprocessing phase involved handling missing values and outliers, which significantly improved the quality of the data, making it more suitable for analysis.
2. **Feature Scaling:** Applying normalization and standardization techniques to the features resulted in enhanced model performance and better convergence during the training process.
3. **Resampling Techniques:** Using both oversampling and undersampling methods helped in creating a balanced class distribution, preventing the model from being biased towards the majority class.
4. **Model Selection:** Testing various algorithms such as logistic regression and random forest allowed us to identify the best-performing model that provided the most accurate predictions for fraud detection.
5. **Neural Network Architecture:** The implementation of a deep learning model with multiple layers enabled the system to learn intricate patterns within the data, leading to highly accurate predictions for fraud detection.

```
In [1]: !pip install pandas
!pip install seaborn
!pip install numpy
import pandas as pd
import seaborn as sns
import numpy as np
```

Requirement already satisfied: pandas in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (1.3.5)

Requirement already satisfied: python-dateutil>=2.7.3 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from pandas) (2.8.2)

Requirement already satisfied: numpy>=1.17.3 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from pandas) (1.21.6)

Requirement already satisfied: pytz>=2017.3 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from pandas) (2024.1)

Requirement already satisfied: six>=1.5 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from python-dateutil>=2.7.3->pandas) (1.16.0)

Requirement already satisfied: seaborn in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (0.12.2)

Requirement already satisfied: matplotlib!=3.6.1,>=3.1 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from seaborn) (3.5.3)

Requirement already satisfied: pandas>=0.25 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from seaborn) (1.3.5)

Requirement already satisfied: numpy!=1.24.0,>=1.17 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from seaborn) (1.21.6)

Requirement already satisfied: typing_extensions in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from seaborn) (4.4.0)

Requirement already satisfied: packaging>=20.0 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (22.0)

Requirement already satisfied: cycler>=0.10 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (0.11.0)

Requirement already satisfied: python-dateutil>=2.7 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (2.8.2)

Requirement already satisfied: pyparsing>=2.2.1 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (3.1.2)

Requirement already satisfied: fonttools>=4.22.0 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (4.38.0)

Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (1.4.5)

Requirement already satisfied: pillow>=6.2.0 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (9.5.0)

Requirement already satisfied: pytz>=2017.3 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from pandas>=0.25->seaborn) (2024.1)

Requirement already satisfied: six>=1.5 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.1->seaborn) (1.16.0)

Requirement already satisfied: numpy in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (1.21.6)

```
In [2]: !pip install matplotlib
import matplotlib.pyplot as plt
```

Requirement already satisfied: matplotlib in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (3.5.3)
Requirement already satisfied: pillow>=6.2.0 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from matplotlib) (9.5.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from matplotlib) (4.38.0)
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from matplotlib) (3.1.2)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from matplotlib) (2.8.2)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from matplotlib) (1.4.5)
Requirement already satisfied: numpy>=1.17 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from matplotlib) (1.21.6)
Requirement already satisfied: cycler>=0.10 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from matplotlib) (0.11.0)
Requirement already satisfied: packaging>=20.0 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from matplotlib) (22.0)
Requirement already satisfied: typing-extensions in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from kiwisolver>=1.0.1->matplotlib) (4.4.0)
Requirement already satisfied: six>=1.5 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from python-dateutil>=2.7->matplotlib) (1.16.0)

In [3]: !pip install shap

Requirement already satisfied: shap in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (0.42.1)

Requirement already satisfied: numpy in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from shap) (1.21.6)

Requirement already satisfied: tqdm>=4.27.0 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from shap) (4.66.2)

Requirement already satisfied: scikit-learn in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from shap) (1.0.2)

Requirement already satisfied: pandas in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from shap) (1.3.5)

Requirement already satisfied: cloudpickle in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from shap) (2.2.1)

Requirement already satisfied: packaging>20.9 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from shap) (22.0)

Requirement already satisfied: slicer==0.0.7 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from shap) (0.0.7)

Requirement already satisfied: numba in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from shap) (0.56.4)

Requirement already satisfied: scipy in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from shap) (1.7.3)

Requirement already satisfied: colorama in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from tqdm>=4.27.0->shap) (0.4.6)

Requirement already satisfied: llvmlite<0.40,>=0.39.0dev0 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from numba->shap) (0.39.1)

Requirement already satisfied: importlib-metadata in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from numba->shap) (4.11.3)

Requirement already satisfied: setuptools in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from numba->shap) (65.6.3)

Requirement already satisfied: python-dateutil>=2.7.3 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from pandas->shap) (2.8.2)

Requirement already satisfied: pytz>=2017.3 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from pandas->shap) (2024.1)

Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from scikit-learn->shap) (3.1.0)

Requirement already satisfied: joblib>=0.11 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from scikit-learn->shap) (1.3.2)

Requirement already satisfied: six>=1.5 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from python-dateutil>=2.7.3->pandas->shap) (1.16.0)

Requirement already satisfied: zipp>=0.5 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from importlib-metadata->numba->shap) (3.11.0)

Requirement already satisfied: typing-extensions>=3.6.4 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from importlib-metadata->numba->shap) (4.4.0)

In [4]: `import shap`

IPProgress not found. Please update jupyter and ipywidgets. See https://ipywidgets.readthedocs.io/en/stable/user_install.html

In [5]: `from sklearn.model_selection import GridSearchCV`
`from sklearn.preprocessing import StandardScaler`
`from sklearn.linear_model import LogisticRegression`

In [6]: `from sklearn.ensemble import RandomForestClassifier`

In [7]: `from sklearn.svm import SVC`

In [8]: `from sklearn.model_selection import train_test_split`

```
In [9]: from sklearn.metrics import classification_report
```

```
In [10]: from sklearn.ensemble import IsolationForest
```

```
In [11]: from sklearn.preprocessing import StandardScaler
```

```
In [12]: from sklearn.metrics import precision_score, recall_score, f1_score
```

```
In [13]: !pip install imblearn
```

```
Requirement already satisfied: imblearn in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (0.0)
Requirement already satisfied: imbalanced-learn in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from imblearn) (0.12.2)
Requirement already satisfied: scikit-learn>=1.0.2 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from imbalanced-learn->imblearn) (1.0.2)
Requirement already satisfied: joblib>=1.1.1 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from imbalanced-learn->imblearn) (1.3.2)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from imbalanced-learn->imblearn) (3.1.0)
Requirement already satisfied: numpy>=1.17.3 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from imbalanced-learn->imblearn) (1.21.6)
Requirement already satisfied: scipy>=1.5.0 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from imbalanced-learn->imblearn) (1.7.3)
```

```
In [14]: from scipy.stats import ttest_ind
```

```
In [15]: pip install imbalanced-learn
```

```
Requirement already satisfied: imbalanced-learn in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (0.12.2)
Requirement already satisfied: joblib>=1.1.1 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from imbalanced-learn) (1.3.2)
Requirement already satisfied: scipy>=1.5.0 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from imbalanced-learn) (1.7.3)
Requirement already satisfied: numpy>=1.17.3 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from imbalanced-learn) (1.21.6)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from imbalanced-learn) (3.1.0)
Requirement already satisfied: scikit-learn>=1.0.2 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from imbalanced-learn) (1.0.2)
Note: you may need to restart the kernel to use updated packages.
```

```
In [16]: pip show imbalanced-learn
```

```
Name: imbalanced-learn
Version: 0.12.2
Summary: Toolbox for imbalanced dataset in machine learning.
Home-page: https://github.com/scikit-learn-contrib/imbalanced-learn
Author:
Author-email:
License: MIT
Location: c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages
Requires: joblib, numpy, scikit-learn, scipy, threadpoolctl
Required-by: imblearn
Note: you may need to restart the kernel to use updated packages.
```

```
In [19]: from imblearn.over_sampling import RandomOverSampler
```

```
In [20]: from imblearn.over_sampling import SMOTE
```

```
In [21]: pip install --upgrade scikit-learn
```

Requirement already satisfied: scikit-learn in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (1.0.2)
 Requirement already satisfied: scipy>=1.1.0 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from scikit-learn) (1.7.3)
 Requirement already satisfied: joblib>=0.11 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from scikit-learn) (1.3.2)
 Requirement already satisfied: numpy>=1.14.6 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from scikit-learn) (1.21.6)
 Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from scikit-learn) (3.1.0)
 Note: you may need to restart the kernel to use updated packages.

```
In [22]: from scipy.stats import ttest_ind
```

```
In [23]: data = pd.read_csv('C:/Users/Admin/Desktop/upgrad/capstone projects/credit card fraud/
```

```
In [24]: data.head()
```

```
Out[24]:
```

	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599	0.098698	0.36378
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803	0.085102	-0.2554
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.791461	0.247676	-1.5146
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609	0.377436	-1.3870
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.592941	-0.270533	0.8177

5 rows × 31 columns

```
In [25]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 284807 entries, 0 to 284806
Data columns (total 31 columns):
#   Column      Non-Null Count  Dtype
---  -
0    Time        284807 non-null  float64
1    V1          284807 non-null  float64
2    V2          284807 non-null  float64
3    V3          284807 non-null  float64
4    V4          284807 non-null  float64
5    V5          284807 non-null  float64
6    V6          284807 non-null  float64
7    V7          284807 non-null  float64
8    V8          284807 non-null  float64
9    V9          284807 non-null  float64
10   V10         284807 non-null  float64
11   V11         284807 non-null  float64
12   V12         284807 non-null  float64
13   V13         284807 non-null  float64
14   V14         284807 non-null  float64
15   V15         284807 non-null  float64
16   V16         284807 non-null  float64
17   V17         284807 non-null  float64
18   V18         284807 non-null  float64
19   V19         284807 non-null  float64
20   V20         284807 non-null  float64
21   V21         284807 non-null  float64
22   V22         284807 non-null  float64
23   V23         284807 non-null  float64
24   V24         284807 non-null  float64
25   V25         284807 non-null  float64
26   V26         284807 non-null  float64
27   V27         284807 non-null  float64
28   V28         284807 non-null  float64
29   Amount      284807 non-null  float64
30   Class       284807 non-null  int64
dtypes: float64(30), int64(1)
memory usage: 67.4 MB
```

```
In [26]: class_counts = data['Class'].value_counts()
labels = ['Genuine', 'Fraud']
colors = ['#66b3ff', '#ff9999']
```

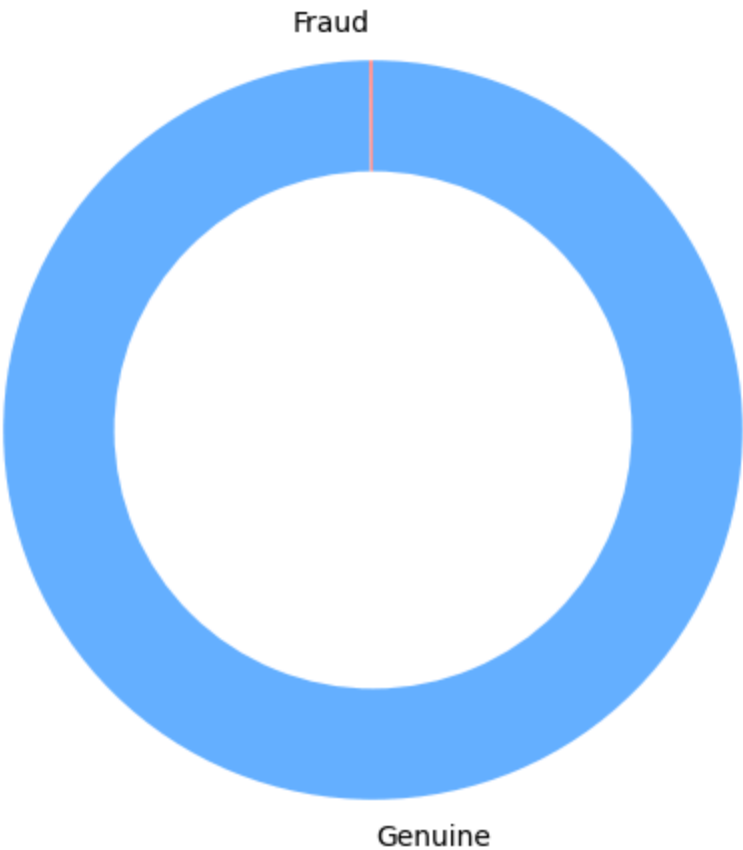
```
In [27]: center_circle = plt.Circle((0, 0), 0.5, color='white')

plt.figure(figsize=(6, 6))
plt.pie(class_counts, labels=labels, colors=colors, startangle=90, counterclock=False,
p = plt.gcf()
p.gca().add_artist(center_circle)

plt.title('Class Distribution in the Dataset')

plt.show()
```


Class Distribution in the Dataset



```
In [28]: print("\nStatistical summary of the dataset:")
data.describe()
```

Statistical summary of the dataset:

	Time	V1	V2	V3	V4	V5	
count	284807.000000	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05
mean	94813.859575	1.759061e-12	-8.251130e-13	-9.654937e-13	8.321385e-13	1.649999e-13	4.819172e-13
std	47488.145955	1.958696e+00	1.651309e+00	1.516255e+00	1.415869e+00	1.380247e+00	1.307671e+00
min	0.000000	-5.640751e+01	-7.271573e+01	-4.832559e+01	-5.683171e+00	-1.137433e+02	-2.635531e+02
25%	54201.500000	-9.203734e-01	-5.985499e-01	-8.903648e-01	-8.486401e-01	-6.915971e-01	-7.081822e-01
50%	84692.000000	1.810880e-02	6.548556e-02	1.798463e-01	-1.984653e-02	-5.433583e-02	-2.041716e-01
75%	139320.500000	1.315642e+00	8.037239e-01	1.027196e+00	7.433413e-01	6.119264e-01	3.091828e-01
max	172792.000000	2.454930e+00	2.205773e+01	9.382558e+00	1.687534e+01	3.480167e+01	7.343282e+01

8 rows × 31 columns

```
In [29]: print("\nColumns in the dataset:")
data.columns
```

Columns in the dataset:

```
Out[29]: Index(['Time', 'V1', 'V2', 'V3', 'V4', 'V5', 'V6', 'V7', 'V8', 'V9', 'V10',
        'V11', 'V12', 'V13', 'V14', 'V15', 'V16', 'V17', 'V18', 'V19', 'V20',
        'V21', 'V22', 'V23', 'V24', 'V25', 'V26', 'V27', 'V28', 'Amount',
        'Class'],
        dtype='object')
```

```
In [30]: print("\nMissing values in the dataset:")
        data.isnull().sum()
```

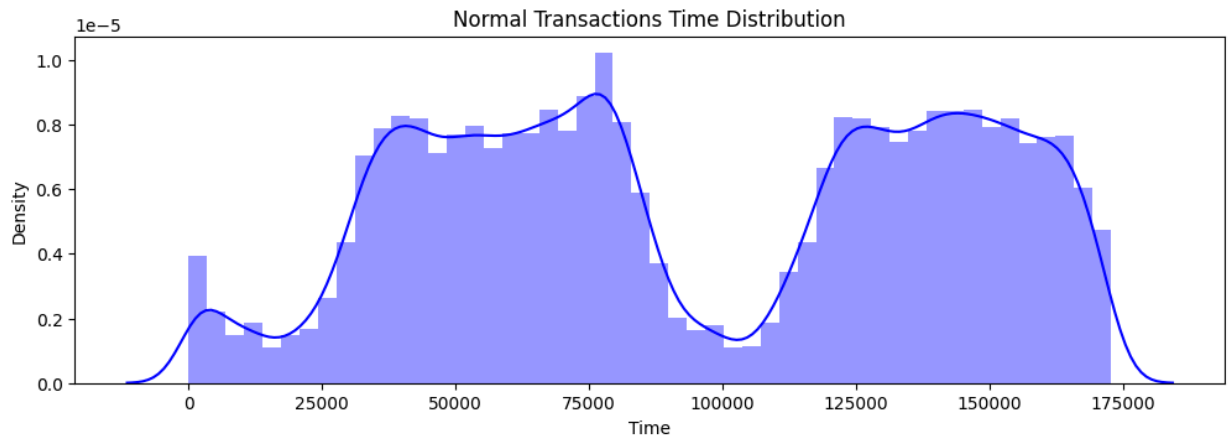
```
Out[30]: Missing values in the dataset:
Time      0
V1         0
V2         0
V3         0
V4         0
V5         0
V6         0
V7         0
V8         0
V9         0
V10        0
V11        0
V12        0
V13        0
V14        0
V15        0
V16        0
V17        0
V18        0
V19        0
V20        0
V21        0
V22        0
V23        0
V24        0
V25        0
V26        0
V27        0
V28        0
Amount     0
Class      0
dtype: int64
```

```
In [31]: plt.figure(figsize=(12, 8))
        plt.subplot(2, 1, 1)
        sns.distplot(data[data['Class'] == 0]["Time"], color='b')
        plt.title('Normal Transactions Time Distribution')
        plt.show()
```

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see <https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

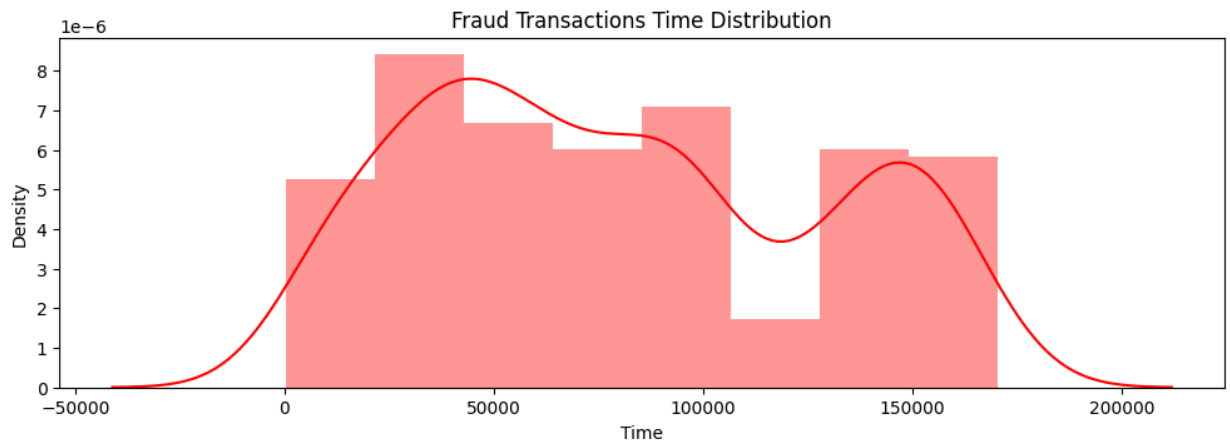


```
In [32]: plt.figure(figsize=(12, 8))
plt.subplot(2, 1, 2)
sns.distplot(data[data['Class'] == 1]["Time"], color='r')
plt.title('Fraud Transactions Time Distribution')
plt.show()
```

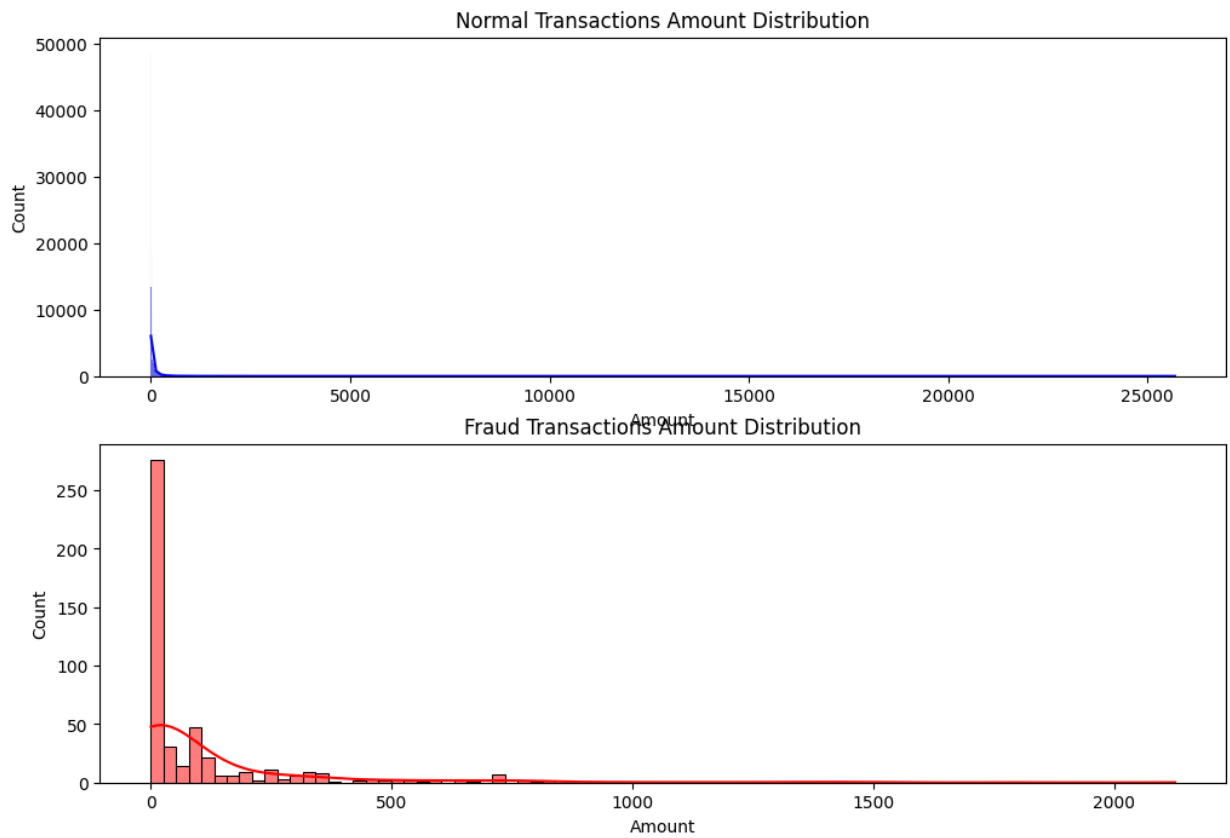
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

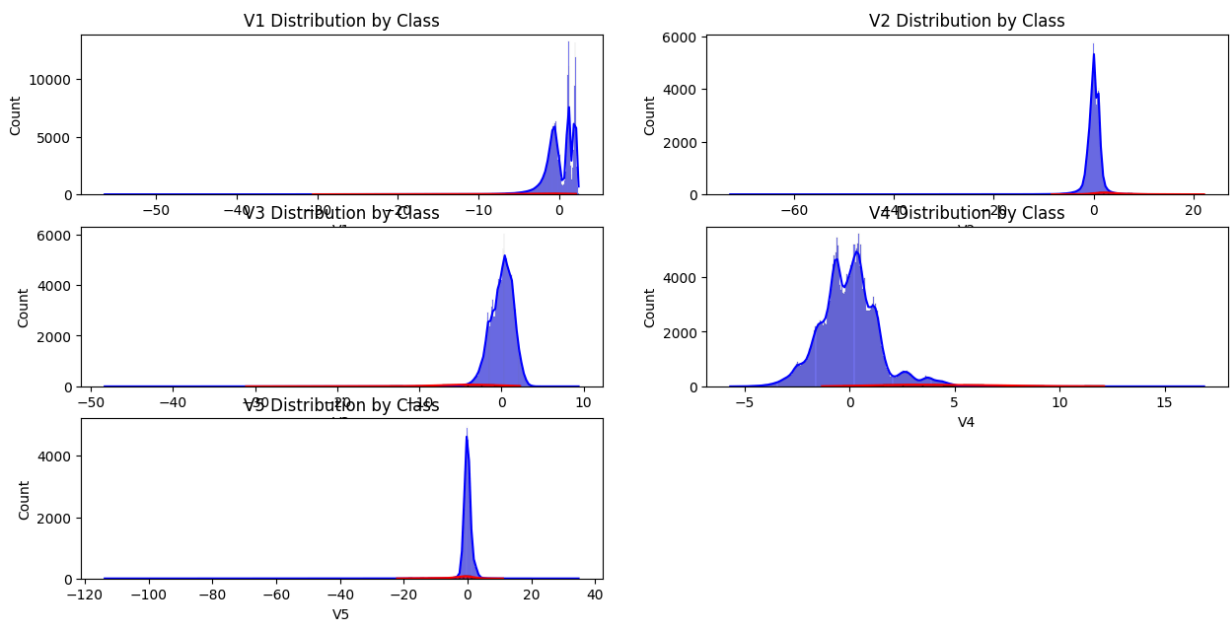
For a guide to updating your code to use the new functions, please see <https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>



```
In [33]: plt.figure(figsize=(12, 8))
plt.subplot(2, 1, 1)
sns.histplot(data[data['Class'] == 0]["Amount"], color='b', kde=True)
plt.title('Normal Transactions Amount Distribution')
plt.subplot(2, 1, 2)
sns.histplot(data[data['Class'] == 1]["Amount"], color='r', kde=True)
plt.title('Fraud Transactions Amount Distribution')
plt.show()
```

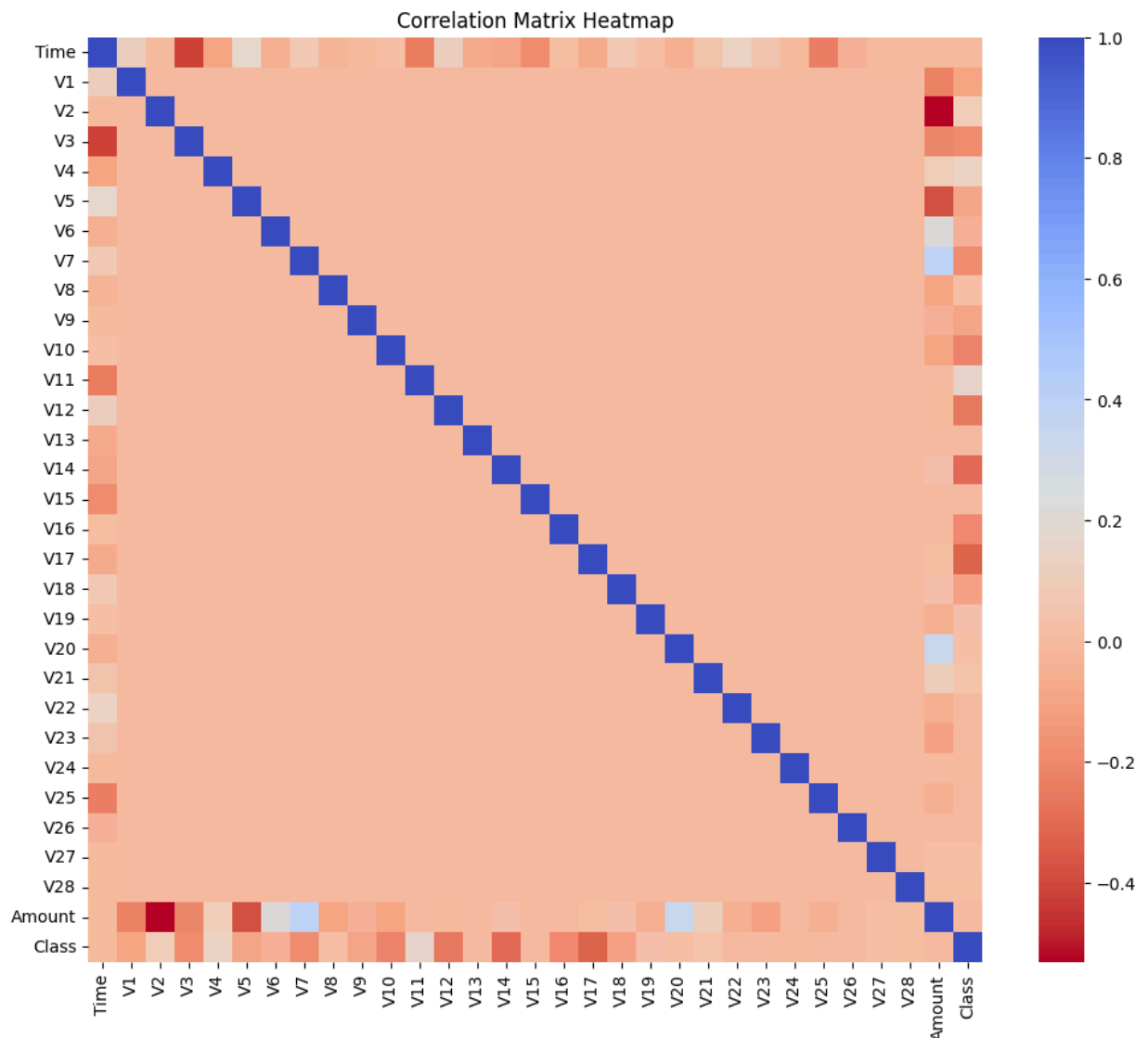


```
In [34]: features = ['V1', 'V2', 'V3', 'V4', 'V5']
plt.figure(figsize=(15, 35))
for i, feature in enumerate(features, 1):
    plt.subplot(14, 2, i)
    sns.histplot(data[data['Class'] == 0][feature], color='b', kde=True)
    sns.histplot(data[data['Class'] == 1][feature], color='r', kde=True)
    plt.title(f'{feature} Distribution by Class')
plt.show()
```



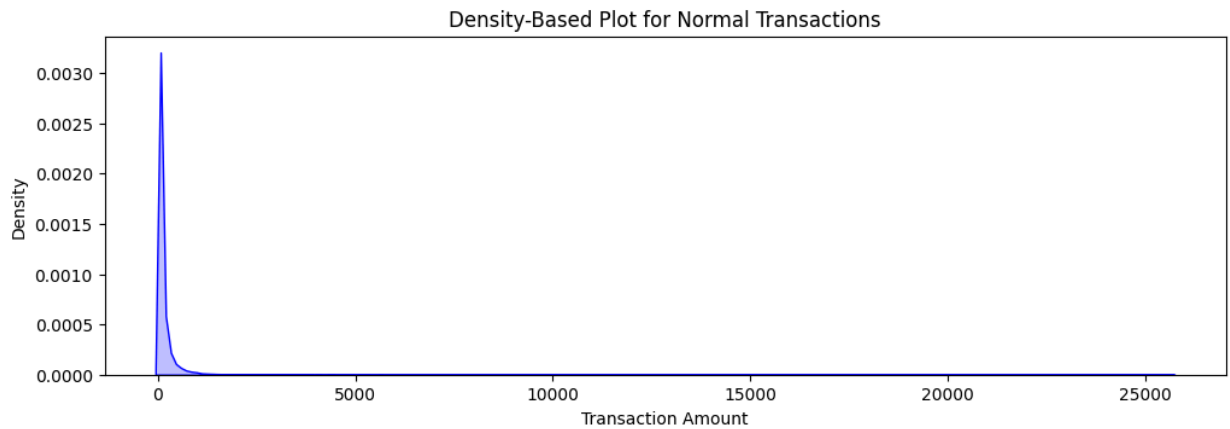
```
In [35]: plt.figure(figsize=(12, 10))
corr = data.corr()
sns.heatmap(corr, cmap='coolwarm_r', annot_kws={'size': 10})
```

```
plt.title('Correlation Matrix Heatmap')
plt.show()
```



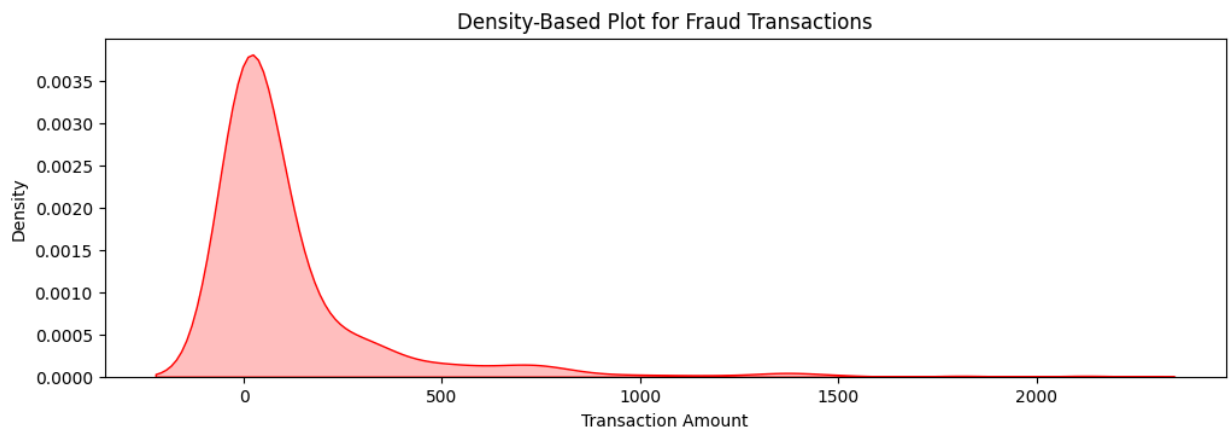
```
In [36]: plt.figure(figsize=(12, 8))
plt.subplot(2, 1, 1)
sns.kdeplot(data[data['Class'] == 0]["Amount"], shade=True, color='b', label='Normal Transactions')
plt.title('Density-Based Plot for Normal Transactions')
plt.xlabel('Transaction Amount')
plt.ylabel('Density')
plt.show()
```

`shade` is now deprecated in favor of `fill`; setting `fill=True`.
This will become an error in seaborn v0.14.0; please update your code.

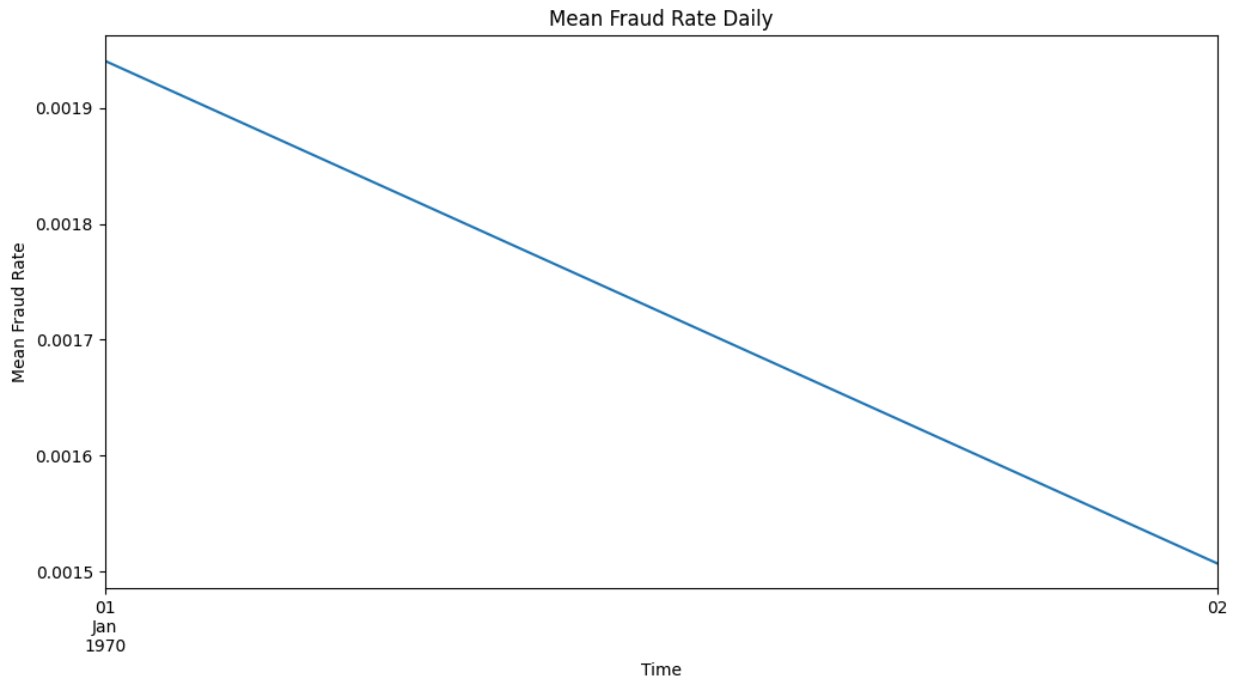


```
In [37]: plt.figure(figsize=(12, 8))
plt.subplot(2, 1, 2)
sns.kdeplot(data[data['Class'] == 1]['Amount'], shade=True, color='r', label='Fraud Tr
plt.title('Density-Based Plot for Fraud Transactions')
plt.xlabel('Transaction Amount')
plt.ylabel('Density')
plt.show()
```

``shade`` is now deprecated in favor of ``fill``; setting ``fill=True``.
This will become an error in seaborn v0.14.0; please update your code.



```
In [38]: # Time series analysis for fraud cases
plt.figure(figsize=(12, 6))
data['Time'] = pd.to_datetime(data['Time'], unit='s')
data.set_index('Time', inplace=True)
data['Class'].resample('D').mean().plot()
plt.title('Mean Fraud Rate Daily')
plt.xlabel('Time')
plt.ylabel('Mean Fraud Rate')
plt.show()
```



```
In [39]: normal_transactions = data[data['Class'] == 0]['Amount']
fraud_transactions = data[data['Class'] == 1]['Amount']
t_stat, p_val = ttest_ind(normal_transactions, fraud_transactions)
print(f"T-statistic: {t_stat}, P-value: {p_val}")
```

T-statistic: -3.00555231397141, P-value: 0.002651220649191683

```
In [40]: X = data.drop('Class', axis=1)
y = data['Class']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=
```

```
In [41]: model = IsolationForest(contamination=0.01, random_state=42)
model.fit(X_train)
```

X does not have valid feature names, but IsolationForest was fitted with feature names

```
Out[41]: IsolationForest(contamination=0.01, random_state=42)
```

```
In [42]: y_pred = model.predict(X_test)
```

```
In [43]: print("Classification Report for Anomaly Detection Model:")
print(classification_report(y_test, y_pred))
```

Classification Report for Anomaly Detection Model:

	precision	recall	f1-score	support
-1	0.00	0.00	0.00	0
0	0.00	0.00	0.00	56864
1	0.00	0.51	0.00	98
accuracy			0.00	56962
macro avg	0.00	0.17	0.00	56962
weighted avg	0.00	0.00	0.00	56962

Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.
Recall and F-score are ill-defined and being set to 0.0 in labels with no true samples. Use `zero_division` parameter to control this behavior.
Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.
Recall and F-score are ill-defined and being set to 0.0 in labels with no true samples. Use `zero_division` parameter to control this behavior.
Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.
Recall and F-score are ill-defined and being set to 0.0 in labels with no true samples. Use `zero_division` parameter to control this behavior.

```
In [44]: firewall_data = data[data['Amount'] > 1000] # Example threshold for suspicious transactions
fraudulent_firewall_transactions = firewall_data[firewall_data['Class'] == 1]
print("Fraudulent Transactions within Firewall:")
print(fraudulent_firewall_transactions)
```


Fraudulent Transactions within Firewall:

		V1	V2	V3	V4	V5	\
Time							
1970-01-01	02:31:04	-3.499108	0.258555	-4.489558	4.853894	-6.974522	
1970-01-01	05:01:28	-12.224021	3.854150	-12.466766	9.648311	-2.726961	
1970-01-01	16:23:31	-2.326922	-3.348439	-3.513408	3.175060	-2.815137	
1970-01-01	17:21:07	-5.344665	-0.285760	-3.835616	5.337048	-7.609909	
1970-01-01	18:09:45	-2.923827	1.524837	-3.018758	3.289291	-5.755542	
1970-01-02	10:03:28	-2.003460	-7.159042	-4.050976	1.309580	-2.058102	
1970-01-02	12:59:44	-1.212682	-2.484824	-6.397186	3.670562	-0.863375	
1970-01-02	18:51:18	-1.600211	-3.488130	-6.459303	3.246816	-1.614608	
1970-01-02	18:51:49	-0.082983	-3.935919	-2.616709	0.163310	-1.400952	
		V6	V7	V8	V9	V10	...
Time							...
1970-01-01	02:31:04	3.628382	5.431271	-1.946734	-0.775680	-1.987773	...
1970-01-01	05:01:28	-4.445610	-21.922811	0.320792	-4.433162	-11.201400	...
1970-01-01	16:23:31	-0.203363	-0.892144	0.333226	-0.802005	-4.350685	...
1970-01-01	17:21:07	3.874668	1.289630	0.201742	-3.003532	-3.990551	...
1970-01-01	18:09:45	2.218276	-0.509995	-3.569444	-1.016592	-4.320536	...
1970-01-02	10:03:28	-0.098621	2.880083	-0.727484	1.460381	-1.531608	...
1970-01-02	12:59:44	-1.855855	1.017732	-0.544704	-1.703378	-3.739659	...
1970-01-02	18:51:18	-1.260375	0.288223	-0.048964	-0.734975	-4.441484	...
1970-01-02	18:51:49	-0.809419	1.501580	-0.471000	1.519743	-1.134454	...

		V21	V22	V23	V24	V25	\
Time							
1970-01-01	02:31:04	-1.052368	0.204817	-2.119007	0.170279	-0.393844	
1970-01-01	05:01:28	-1.159830	-1.504119	-19.254328	0.544867	-4.781606	
1970-01-01	16:23:31	1.226648	-0.695902	-1.478490	-0.061553	0.236155	
1970-01-01	17:21:07	0.276011	1.342045	-1.016579	-0.071361	-0.335869	
1970-01-01	18:09:45	-0.511657	-0.122724	-4.288639	0.563797	-0.949451	
1970-01-02	10:03:28	1.244287	-1.015232	-1.800985	0.657586	-0.435617	
1970-01-02	12:59:44	1.396872	0.092073	-1.492882	-0.204227	0.532511	
1970-01-02	18:51:18	1.191175	-0.967141	-1.463421	-0.624231	-0.176462	
1970-01-02	18:51:49	0.702672	-0.182305	-0.921017	0.111635	-0.071622	

		V26	V27	V28	Amount	Class
Time						
1970-01-01	02:31:04	0.296367	1.985913	-0.900452	1809.68	1
1970-01-01	05:01:28	-0.007772	3.052358	-0.775036	1218.89	1
1970-01-01	16:23:31	0.531911	0.302324	0.536375	1389.56	1
1970-01-01	17:21:07	0.441044	1.520613	-1.115937	1402.16	1
1970-01-01	18:09:45	-0.204532	1.510206	-0.324706	1354.25	1
1970-01-02	10:03:28	-0.894509	-0.397557	0.314262	2125.87	1
1970-01-02	12:59:44	-0.293871	0.212663	0.431095	1335.00	1
1970-01-02	18:51:18	0.400348	0.152947	0.477775	1504.93	1
1970-01-02	18:51:49	-1.125881	-0.170947	0.126221	1096.99	1

[9 rows x 30 columns]

```
In [45]: param_grid = {'C': [0.001, 0.01, 0.1, 1, 10], 'penalty': ['l2']}
         solver = 'liblinear'
```

```
In [46]: scaler = StandardScaler()
         X_train_scaled = scaler.fit_transform(X_train)
         X_test_scaled = scaler.transform(X_test)
```

```
In [47]: grid_search = GridSearchCV(LogisticRegression(solver=solver, max_iter=1000), param_grid)
grid_search.fit(X_train_scaled, y_train)

best_params = grid_search.best_params_
print("Best parameter combinations: ", best_params)
```

Best parameter combinations: {'C': 10, 'penalty': 'l2'}

```
In [48]: scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

```
In [49]: smote = SMOTE(random_state=42)
X_resampled, y_resampled = smote.fit_resample(X_train_scaled, y_train)
```

```
In [50]: print("Original data shape:", X_train.shape, y_train.shape)
print("Resampled data shape:", X_resampled.shape, y_resampled.shape)
```

Original data shape: (227845, 29) (227845,)
Resampled data shape: (454902, 29) (454902,)

```
In [51]: !pip install xgboost
from xgboost import XGBClassifier
```

Requirement already satisfied: xgboost in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (1.6.2)
Requirement already satisfied: numpy in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from xgboost) (1.21.6)
Requirement already satisfied: scipy in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from xgboost) (1.7.3)

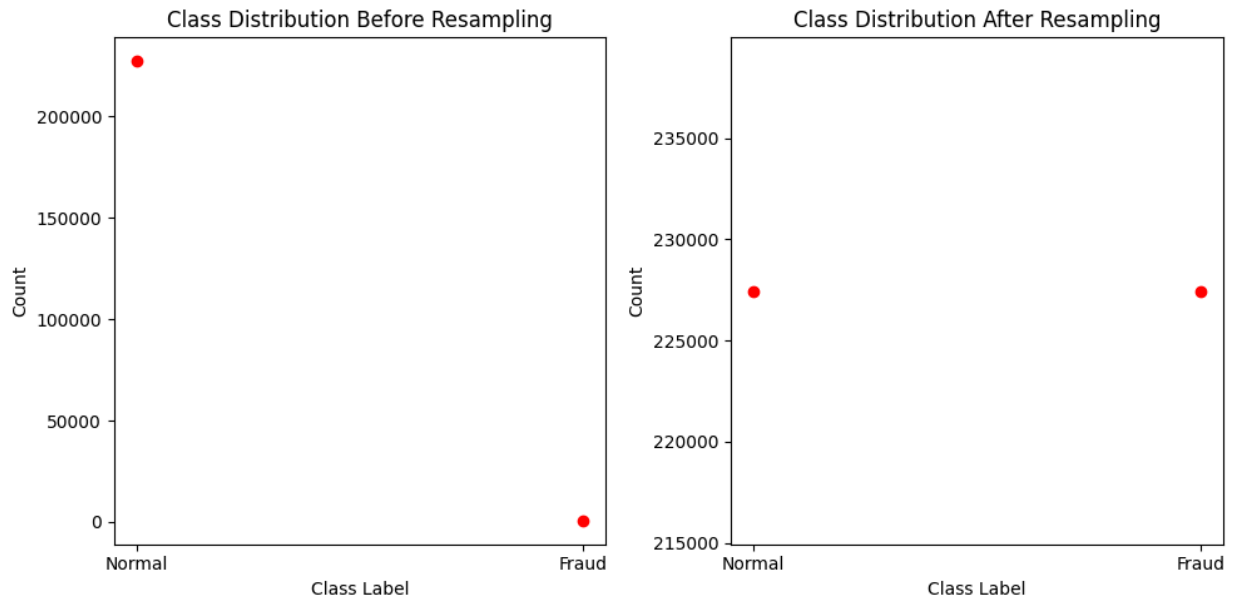
```
In [52]: xgb_model = XGBClassifier()
xgb_model.fit(X_resampled, y_resampled)
y_pred_xgb = xgb_model.predict(X_test_scaled)
```

```
In [53]: plt.figure(figsize=(10, 5))

# Dot plot for class distribution before resampling
plt.subplot(1, 2, 1)
plt.title('Class Distribution Before Resampling')
plt.plot([0, 1], [sum(y_train==0), sum(y_train==1)], 'ro')
plt.xticks([0, 1], ['Normal', 'Fraud'])
plt.xlabel('Class Label')
plt.ylabel('Count')

# Dot plot for class distribution after resampling
plt.subplot(1, 2, 2)
plt.title('Class Distribution After Resampling')
plt.plot([0, 1], [sum(y_resampled==0), sum(y_resampled==1)], 'ro')
plt.xticks([0, 1], ['Normal', 'Fraud'])
plt.xlabel('Class Label')
plt.ylabel('Count')

plt.tight_layout()
plt.show()
```



```
In [54]: ros = RandomOverSampler(random_state=0)
X_resampled_aug, y_resampled_aug = ros.fit_resample(X, y)
```

```
In [55]: # Display the results
from collections import Counter
print("Original dataset shape:", Counter(y))
print("Resampled dataset shape:", Counter(y_resampled_aug))
```

```
Original dataset shape: Counter({0: 284315, 1: 492})
Resampled dataset shape: Counter({0: 284315, 1: 284315})
```

```
In [56]: # Visualize the class distribution after Random Over Sampling
plt.figure(figsize=(8, 6))
plt.bar(Counter(y_resampled_aug).keys(), Counter(y_resampled_aug).values(), color=['b'
plt.xticks(list(Counter(y_resampled_aug).keys()), ['Normal', 'Fraud'])
plt.xlabel('Class Label')
plt.ylabel('Count')
plt.title('Class Distribution After Random Over Sampling')
plt.show()
```



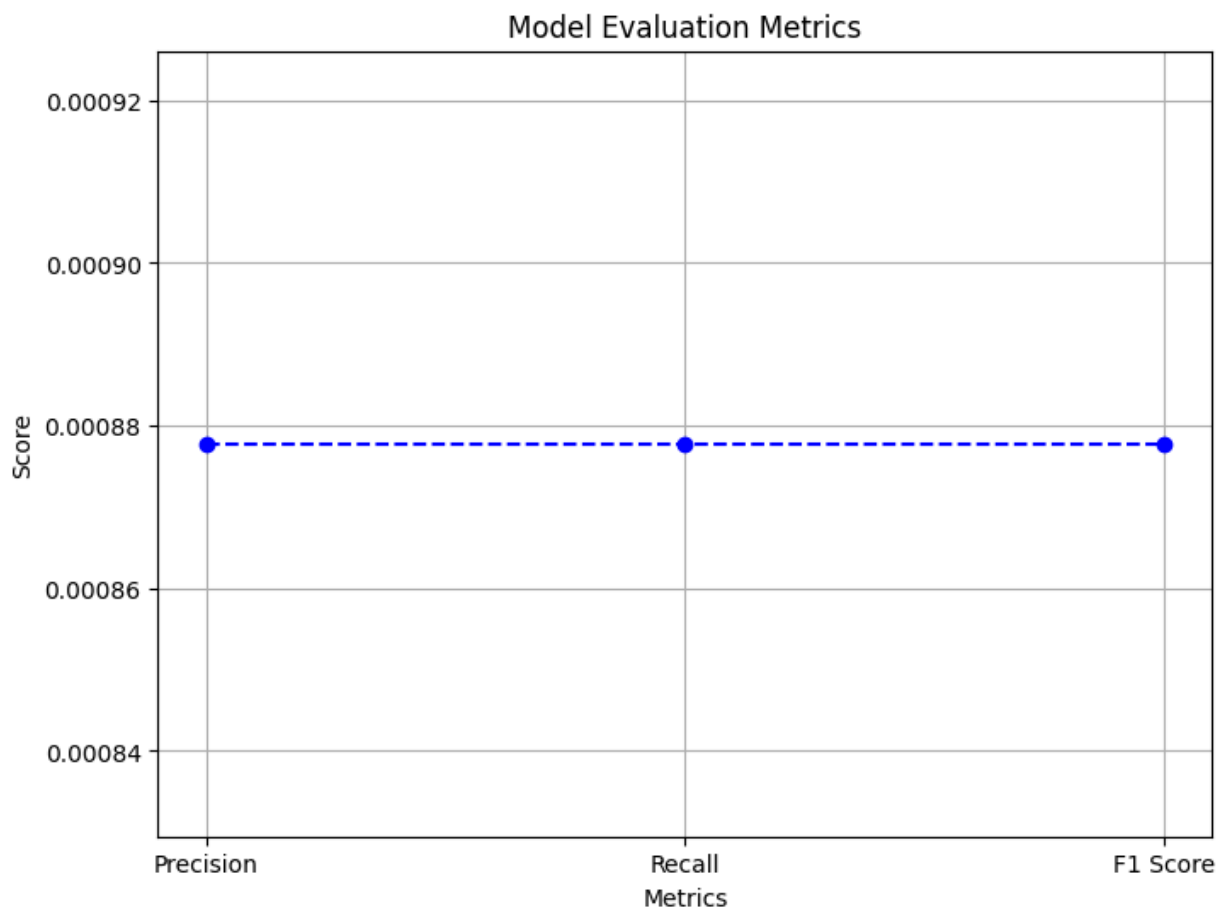
```
In [57]: precision = precision_score(y_test, y_pred, average='micro')
recall = recall_score(y_test, y_pred, average='micro')
f1 = f1_score(y_test, y_pred, average='micro')

print("Precision: ", precision)
print("Recall: ", recall)
print("F1 Score: ", f1)
```

```
Precision:  0.0008777781679014079
Recall:  0.0008777781679014079
F1 Score:  0.0008777781679014079
```

```
In [58]: metrics = ['Precision', 'Recall', 'F1 Score']
scores = [precision, recall, f1]
```

```
In [59]: plt.figure(figsize=(8, 6))
plt.plot(metrics, scores, marker='o', linestyle='--', color='b')
plt.title('Model Evaluation Metrics')
plt.xlabel('Metrics')
plt.ylabel('Score')
plt.grid(True)
plt.show()
```



```
In [60]: missing_values = data.isnull().sum()
print("Missing values: ", missing_values)
```

```
Missing values: V1      0
V2      0
V3      0
V4      0
V5      0
V6      0
V7      0
V8      0
V9      0
V10     0
V11     0
V12     0
V13     0
V14     0
V15     0
V16     0
V17     0
V18     0
V19     0
V20     0
V21     0
V22     0
V23     0
V24     0
V25     0
V26     0
V27     0
V28     0
Amount  0
Class   0
dtype: int64
```

```
In [61]: !pip install missingno
import missingno as msno

msno.matrix(data)
plt.title('Missing Data Visualization - Heatmap')
plt.show()
```

Requirement already satisfied: missingno in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (0.5.2)

Requirement already satisfied: numpy in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from missingno) (1.21.6)

Requirement already satisfied: seaborn in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from missingno) (0.12.2)

Requirement already satisfied: matplotlib in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from missingno) (3.5.3)

Requirement already satisfied: scipy in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from missingno) (1.7.3)

Requirement already satisfied: python-dateutil>=2.7 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from matplotlib->missingno) (2.8.2)

Requirement already satisfied: pyparsing>=2.2.1 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from matplotlib->missingno) (3.1.2)

Requirement already satisfied: packaging>=20.0 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from matplotlib->missingno) (22.0)

Requirement already satisfied: cycler>=0.10 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from matplotlib->missingno) (0.11.0)

Requirement already satisfied: pillow>=6.2.0 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from matplotlib->missingno) (9.5.0)

Requirement already satisfied: fonttools>=4.22.0 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from matplotlib->missingno) (4.38.0)

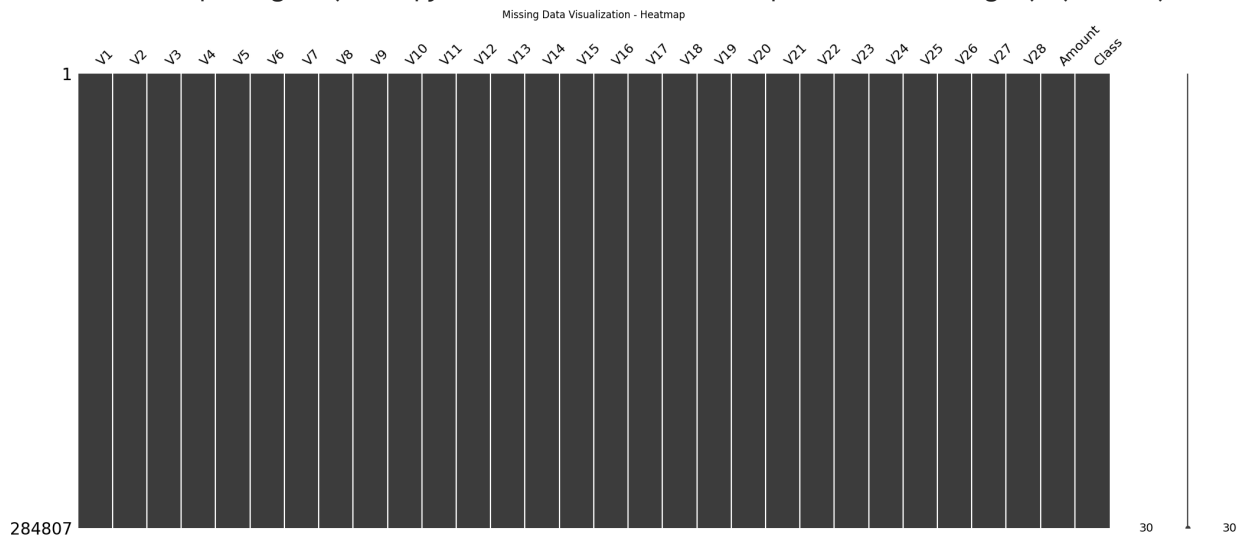
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from matplotlib->missingno) (1.4.5)

Requirement already satisfied: typing_extensions in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from seaborn->missingno) (4.4.0)

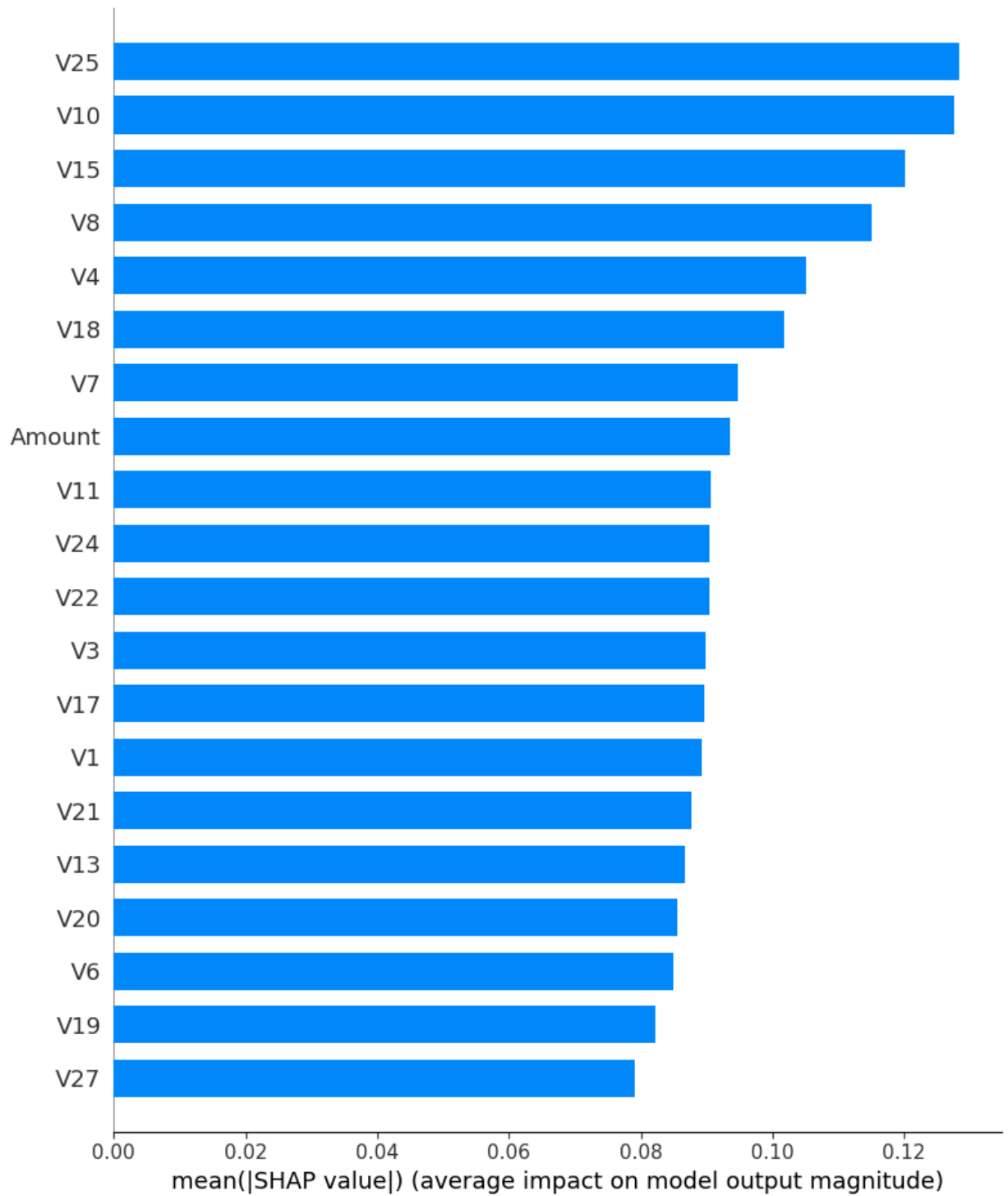
Requirement already satisfied: pandas>=0.25 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from seaborn->missingno) (1.3.5)

Requirement already satisfied: pytz>=2017.3 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from pandas>=0.25->seaborn->missingno) (2024.1)

Requirement already satisfied: six>=1.5 in c:\users\admin\anaconda3\envs\notebook-6.0.0\lib\site-packages (from python-dateutil>=2.7->matplotlib->missingno) (1.16.0)



```
In [62]: explainer = shap.Explainer(model)
shap_values = explainer.shap_values(X_test)
shap.summary_plot(shap_values, X_test, plot_type="bar")
```



In []: