

Data Collection and Preprocessing Phase

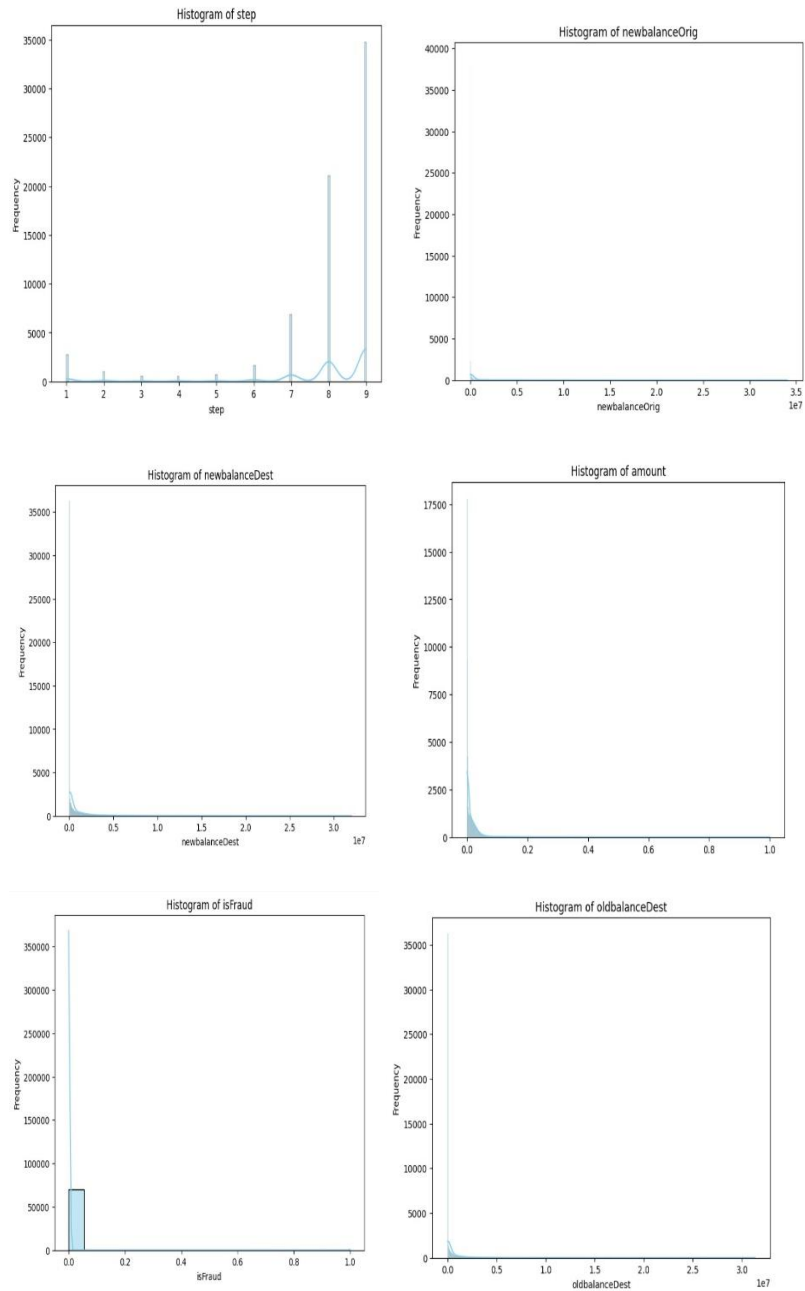
| | |
|---------------|---|
| Date | 26 April 2024 |
| Team ID | Team-738315 |
| Project Title | Online Payment Fraud Detection using Machine Learning |
| Maximum Marks | 6 Marks |

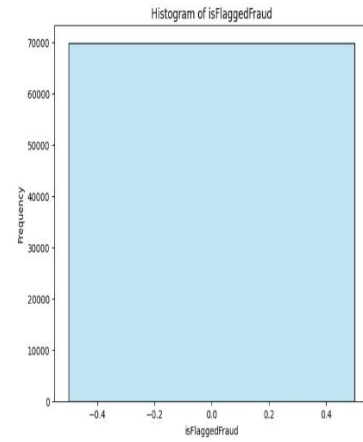
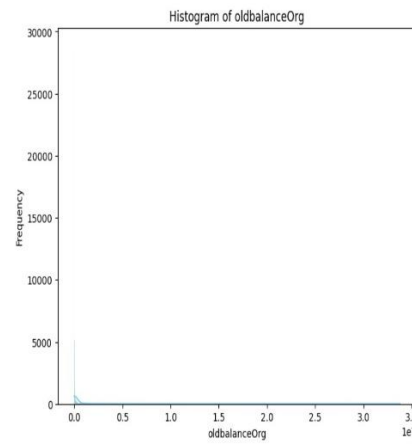
Data Exploration and Preprocessing Report

Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.

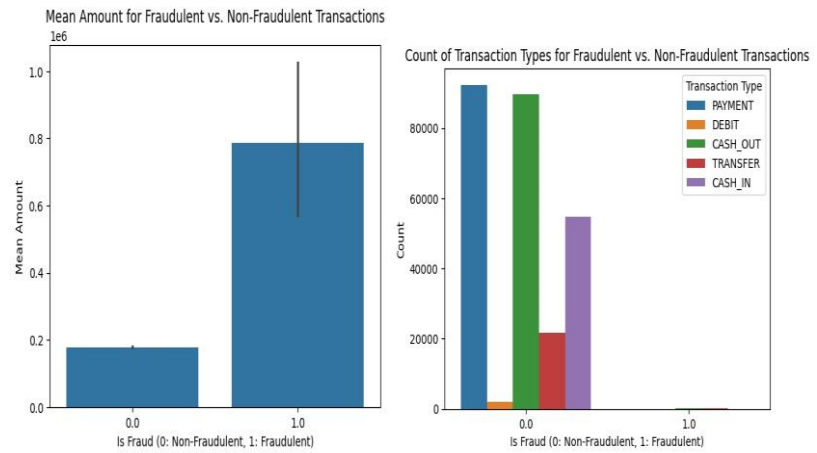
| Section | Description |
|---------------|---|
| Data Overview | <p><u>Dimension:</u> 69858 rows × 11 columns</p> <p><u>Descriptive statistics:</u></p> <pre> Descriptive statistics: step amount oldbalanceOrg newbalanceOrig oldbalanceDest \ count 69858.00 69858.00 69858.00 69858.00 69858.00 mean 7.89 164017.62 904306.61 920297.98 854352.17 std 1.87 331373.01 2791616.42 2829881.42 2403000.41 min 1.00 0.63 0.00 0.00 0.00 25% 8.00 9011.01 0.00 0.00 0.00 50% 8.00 44049.18 19908.50 0.00 14160.50 75% 9.00 199130.74 172117.50 194020.54 518607.29 max 9.00 10000000.00 33797391.55 34008736.98 31306920.44 newbalanceDest isFraud isFlaggedFraud count 69857.00 69857.00 69857.00 mean 1185140.93 0.00 0.00 std 2871790.71 0.04 0.00 min 0.00 0.00 0.00 25% 0.00 0.00 0.00 50% 21130.68 0.00 0.00 75% 982411.35 0.00 0.00 max 31976992.09 1.00 0.00 ... 10 isFlaggedFraud 69857 non-null float64 dtypes: float64(7), int64(1), object(3) memory usage: 5.9+ MB None </pre> |

Univariate Analysis

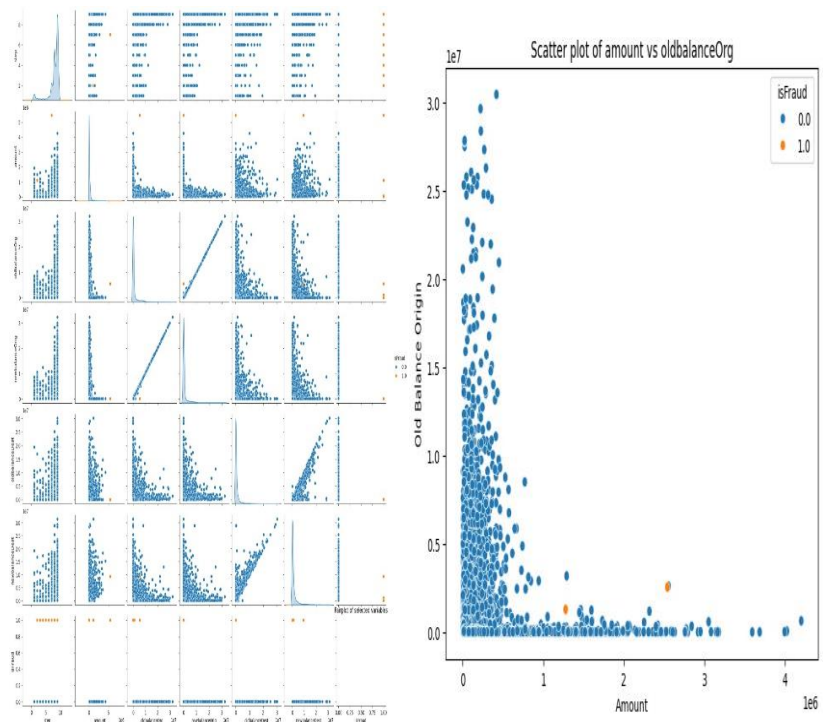




Bivariate Analysis



Multivariate Analysis



Outliers and Anomalies

-

Data Preprocessing Code Screenshots

Loading Data

```
1 import pandas as pd
2 df=pd.read_csv('/content/PS_20174392719_1491204439457_log.csv')
3 dfdi
```

| | step | type | amount | nameOrig | oldbalanceOrg | newbalanceOrig | nameDest | oldbalanceDest | n |
|--------|------|----------|-----------|-------------|---------------|----------------|-------------|----------------|-----|
| 0 | 1 | PAYMENT | 9839.64 | C1231006815 | 170136.00 | 160296.36 | M1979787155 | 0.00 | |
| 1 | 1 | PAYMENT | 1864.28 | C1666544295 | 21249.00 | 19384.72 | M2044282225 | 0.00 | |
| 2 | 1 | TRANSFER | 181.00 | C1305486145 | 181.00 | 0.00 | C553264065 | 0.00 | |
| 3 | 1 | CASH_OUT | 181.00 | C840083671 | 181.00 | 0.00 | C38997010 | 21182.00 | |
| 4 | 1 | PAYMENT | 11668.14 | C2048537720 | 41554.00 | 29885.86 | M1230701703 | 0.00 | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 954362 | 44 | CASH_OUT | 203106.56 | C1477557128 | 614921.99 | 411815.43 | C745656793 | 2619063.89 | |
| 954363 | 44 | CASH_OUT | 21536.67 | C1592775172 | 108619.00 | 87082.33 | C1861308004 | 0.00 | |
| 954364 | 44 | TRANSFER | 19870.84 | C1771376526 | 87082.33 | 67211.50 | C156502301 | 199338.34 | |
| 954365 | 44 | PAYMENT | 733.41 | C478697749 | 16403.00 | 15669.59 | M1234201847 | 0.00 | |
| 954366 | 44 | PAYMENT | 8615.03 | C1766848563 | 15669.59 | 7054.50 | NaN | NaN | |

954367 rows x 11 columns

Handling Missing Data

```
1 # Drop rows with missing values
2 df.dropna(inplace=True)
3
4 # Verify missing values after handling
5 print("\nMissing values after handling:")
6 print(df.isnull().sum())
```

```
Missing values after handling:
step          0
type          0
amount        0
nameOrig      0
oldbalanceOrg 0
newbalanceOrig 0
nameDest      0
oldbalanceDest 0
newbalanceDest 0
isFraud       0
isFlaggedFraud 0
dtype: int64
```

| Data Transformation | <pre>1 import pandas as pd 2 from sklearn.preprocessing import StandardScaler 3 4 # Drop rows with missing values (if any) 5 df.dropna(inplace=True) 6 7 scaler = StandardScaler() 8 numeric_columns = df.select_dtypes(include=['number']) 9 df[numeric_columns.columns] = scaler.fit_transform(numeric_columns) 10 11 print("\nTransformed dataset:") 12 print(df.head())</pre> <p>Transformed dataset:</p> <table><thead><tr><th></th><th>step</th><th>type</th><th>amount</th><th>nameOrig</th><th>oldbalanceOrig</th><th>newbalanceOrig</th><th>\</th></tr></thead><tbody><tr><td>0</td><td>-4.067035</td><td>PAYMENT</td><td>-0.475440</td><td>C1231006815</td><td>-0.263652</td><td>-0.269566</td><td></td></tr><tr><td>1</td><td>-4.067035</td><td>PAYMENT</td><td>-0.498762</td><td>C1666544295</td><td>-0.319003</td><td>-0.321223</td><td></td></tr><tr><td>2</td><td>-4.067035</td><td>TRANSFER</td><td>-0.503685</td><td>C1305486145</td><td>-0.326836</td><td>-0.328329</td><td></td></tr><tr><td>3</td><td>-4.067035</td><td>CASH_OUT</td><td>-0.503685</td><td>C840083671</td><td>-0.326836</td><td>-0.328329</td><td></td></tr><tr><td>4</td><td>-4.067035</td><td>PAYMENT</td><td>-0.470093</td><td>C2048537720</td><td>-0.311454</td><td>-0.317373</td><td></td></tr></tbody></table> <table><thead><tr><th></th><th>nameDest</th><th>oldbalanceDest</th><th>newbalanceDest</th><th>isFraud</th><th>isFlaggedFraud</th></tr></thead><tbody><tr><td>0</td><td>M1979787155</td><td>-0.365855</td><td>-0.421829</td><td>-0.034263</td><td>0.0</td></tr><tr><td>1</td><td>M2044282225</td><td>-0.365855</td><td>-0.421829</td><td>-0.034263</td><td>0.0</td></tr><tr><td>2</td><td>C553264065</td><td>-0.365855</td><td>-0.421829</td><td>29.186334</td><td>0.0</td></tr><tr><td>3</td><td>C38997010</td><td>-0.357041</td><td>-0.421829</td><td>29.186334</td><td>0.0</td></tr><tr><td>4</td><td>M1230701703</td><td>-0.365855</td><td>-0.421829</td><td>-0.034263</td><td>0.0</td></tr></tbody></table> | | step | type | amount | nameOrig | oldbalanceOrig | newbalanceOrig | \ | 0 | -4.067035 | PAYMENT | -0.475440 | C1231006815 | -0.263652 | -0.269566 | | 1 | -4.067035 | PAYMENT | -0.498762 | C1666544295 | -0.319003 | -0.321223 | | 2 | -4.067035 | TRANSFER | -0.503685 | C1305486145 | -0.326836 | -0.328329 | | 3 | -4.067035 | CASH_OUT | -0.503685 | C840083671 | -0.326836 | -0.328329 | | 4 | -4.067035 | PAYMENT | -0.470093 | C2048537720 | -0.311454 | -0.317373 | | | nameDest | oldbalanceDest | newbalanceDest | isFraud | isFlaggedFraud | 0 | M1979787155 | -0.365855 | -0.421829 | -0.034263 | 0.0 | 1 | M2044282225 | -0.365855 | -0.421829 | -0.034263 | 0.0 | 2 | C553264065 | -0.365855 | -0.421829 | 29.186334 | 0.0 | 3 | C38997010 | -0.357041 | -0.421829 | 29.186334 | 0.0 | 4 | M1230701703 | -0.365855 | -0.421829 | -0.034263 | 0.0 |
|---------------------|---|----------------|----------------|-------------|----------------|----------------|----------------|----------------|---|---|-----------|---------|-----------|-------------|-----------|-----------|--|---|-----------|---------|-----------|-------------|-----------|-----------|--|---|-----------|----------|-----------|-------------|-----------|-----------|--|---|-----------|----------|-----------|------------|-----------|-----------|--|---|-----------|---------|-----------|-------------|-----------|-----------|--|--|----------|----------------|----------------|---------|----------------|---|-------------|-----------|-----------|-----------|-----|---|-------------|-----------|-----------|-----------|-----|---|------------|-----------|-----------|-----------|-----|---|-----------|-----------|-----------|-----------|-----|---|-------------|-----------|-----------|-----------|-----|
| | step | type | amount | nameOrig | oldbalanceOrig | newbalanceOrig | \ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | -4.067035 | PAYMENT | -0.475440 | C1231006815 | -0.263652 | -0.269566 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | -4.067035 | PAYMENT | -0.498762 | C1666544295 | -0.319003 | -0.321223 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 | -4.067035 | TRANSFER | -0.503685 | C1305486145 | -0.326836 | -0.328329 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3 | -4.067035 | CASH_OUT | -0.503685 | C840083671 | -0.326836 | -0.328329 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4 | -4.067035 | PAYMENT | -0.470093 | C2048537720 | -0.311454 | -0.317373 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | nameDest | oldbalanceDest | newbalanceDest | isFraud | isFlaggedFraud | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | M1979787155 | -0.365855 | -0.421829 | -0.034263 | 0.0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | M2044282225 | -0.365855 | -0.421829 | -0.034263 | 0.0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 | C553264065 | -0.365855 | -0.421829 | 29.186334 | 0.0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3 | C38997010 | -0.357041 | -0.421829 | 29.186334 | 0.0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4 | M1230701703 | -0.365855 | -0.421829 | -0.034263 | 0.0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Feature Engineering | Attached the codes in final submission. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Save Processed Data | - | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |