Fraud Payment Data Analysis:

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
payfrd = pd.read_csv('payment_fraud.csv')

import warnings
warnings.filterwarnings("ignore")

payfrd.head(15)
```

Out[1]:

	accountAgeDays	numItems	localTime	paymentMethod	paymentMethodAgeDays	label
0	29	1	4.745402	paypal	28.204861	0
1	725	1	4.742303	storecredit	0.000000	0
2	845	1	4.921318	creditcard	0.000000	0
3	503	1	4.886641	creditcard	0.000000	0
4	2000	1	5.040929	creditcard	0.000000	0
5	119	1	4.962055	paypal	0.000000	0
6	2000	1	4.921349	paypal	0.000000	0
7	371	1	4.876771	creditcard	0.000000	0
8	2000	1	4.748314	creditcard	0.000000	0
9	4	1	4.461622	creditcard	0.000000	0
10	26	1	4.745402	paypal	0.000000	0
11	2000	1	4.505662	creditcard	0.000694	0
12	3	1	4.962055	creditcard	2.690972	0
13	3	1	4.921318	creditcard	2.069444	0
14	23	1	4.745402	creditcard	0.001389	0

In [2]:
```
print(payfrd.dtypes)
```

```
accountAgeDays            int64
numItems                  int64
localTime               float64
paymentMethod            object
paymentMethodAgeDays    float64
label                     int64
dtype: object
```

In [3]:
```
payfrd.isnull().sum()
```

Out[3]:
```
accountAgeDays          0
numItems                0
localTime               0
paymentMethod           0
paymentMethodAgeDays    0
label                   0
dtype: int64
```

In [4]:
```
from sklearn import preprocessing
label_encoder = preprocessing.LabelEncoder()
payfrd['paymentMethod'] = label_encoder.fit_transform(payfrd['paymentMethod'])
payfrd.head(15)
```

Out[4]:

	accountAgeDays	numItems	localTime	paymentMethod	paymentMethodAgeDays	label
0	29	1	4.745402	1	28.204861	0
1	725	1	4.742303	2	0.000000	0
2	845	1	4.921318	0	0.000000	0
3	503	1	4.886641	0	0.000000	0
4	2000	1	5.040929	0	0.000000	0
5	119	1	4.962055	1	0.000000	0
6	2000	1	4.921349	1	0.000000	0
7	371	1	4.876771	0	0.000000	0
8	2000	1	4.748314	0	0.000000	0
9	4	1	4.461622	0	0.000000	0
10	26	1	4.745402	1	0.000000	0
11	2000	1	4.505662	0	0.000694	0
12	3	1	4.962055	0	2.690972	0
13	3	1	4.921318	0	2.069444	0
14	23	1	4.745402	0	0.001389	0

In [5]:
```
payfrd = pd.get_dummies(payfrd, columns=['paymentMethod'])
payfrd.sample(3)
```

Out[5]:

	accountAgeDays	numItems	localTime	paymentMethodAgeDays	label	paymentMethod_0
13566	107	1	4.895263	0.000000	0	1
15496	1766	1	5.017904	1.945833	0	1
22895	1961	1	4.965339	0.000000	0	0

In [6]:
```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(payfrd.drop('label', axis=1),
```

In [7]:
```
X_train.head(5)
```

Out[7]:

	accountAgeDays	numItems	localTime	paymentMethodAgeDays	paymentMethod_0	paym
13272	4	1	4.886641	0.000000	1	
22716	1554	1	4.461622	0.256944	0	
30255	2	1	4.742303	1.840278	1	
7676	2000	1	4.748314	0.584028	1	
15105	157	1	5.017904	156.022222	1	

In [8]:
```
y_train.head(5)
```

Out[8]:
```
13272    0
22716    0
30255    0
7676     0
15105    0
Name: label, dtype: int64
```

In [9]:
```
print(X_train.shape)
print(X_test.shape)
print(y_train.shape)
print(y_test.shape)
```

```
(26278, 7)
(12943, 7)
(26278,)
(12943,)
```

Logistic Regression:

In [10]:
```
from sklearn.linear_model import LogisticRegression
trmodel = LogisticRegression()
trmodel.fit(X_train, y_train)
```

Out[10]:
```
LogisticRegression()
```

In [11]:
```
y_pred = trmodel.predict(X_test)
```

In [12]:
```
from sklearn.metrics import accuracy_score
print(accuracy_score(y_pred, y_test))
```

```
1.0
```

In [13]:
```
from sklearn.metrics import confusion_matrix
print(confusion_matrix(y_test, y_pred))
```

```
[[12753     0]
 [    0   190]]
```