**Comprehensive Study of Credit Card Fraud for Enhancing Financial Transaction Security**

**CHAPTER-8**

**Source code**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import warnings

warnings.filterwarnings('ignore')

df = pd.read\_csv("card\_transdata.csv")

df.head()

df.shape

df.info()

df.describe()

df.isnull().sum()

df.columns

# Exploratory Data Analysis

#Distribution of Distance from Home

#Sample 1000 random rows

sampled\_df = df.sample(n=1000)

sns.histplot(sampled\_df['distance\_from\_home'], bins=10, kde=True)

plt.title('Distribution of Distance from Home (1000 Random Rows)')

plt.xlabel('Distance from Home')

plt.ylabel('Frequency')

plt.show()

#Distribution of Distance from Last Transaction

sampled\_df = df.sample(n=1000)

sns.histplot(sampled\_df['distance\_from\_last\_transaction'], bins = 15,kde=True)

plt.title('Distribution of Distance from Last Transaction (1000 Random Rows)')

plt.xlabel('Distance from Last Transaction')

plt.ylabel('Frequency')

plt.show()

# Box Plot

#Boxplot of Ratio to Median Purchase Price by Fraud

sampled\_df = df.sample(n=1000)

sns.boxplot(x='fraud', y='ratio\_to\_median\_purchase\_price', data=sampled\_df)

plt.title('Boxplot of Ratio to Median Purchase Price by Fraud (1000 Random Rows)')

plt.xlabel('Fraud')

plt.ylabel('Ratio to Median Purchase Price')

plt.show()

# Count Plot

# Count Plot of Repeat Retailer

sampled\_df = df.sample(n=1000)

sns.countplot(x='repeat\_retailer', data=sampled\_df)

plt.title('Count of Repeat Retailer (1000 Random Rows)')

plt.xlabel('Repeat Retailer')

plt.ylabel('Count')

plt.show()

# Count Plot of Used Chip

sampled\_df = df.sample(n=1000)

sns.countplot(x='used\_chip', data=sampled\_df)

plt.title('Count of Used Chip (1000 Random Rows)')

plt.xlabel('Used Chip')

plt.ylabel('Count')

plt.show()

# Count Plot of Used Pin Number

sampled\_df = df.sample(n=1000)

sns.countplot(x='used\_pin\_number', data=sampled\_df)

plt.title('Count of Used Pin Number (1000 Random Rows)')

plt.xlabel('Used Pin Number')

plt.ylabel('Count')

plt.show()

# Count Plot of Online Order

sampled\_df = df.sample(n=1000)

sns.countplot(x='online\_order', data=sampled\_df)

plt.title('Count of Online Order (1000 Random Rows)')

plt.xlabel('Online Order')

plt.ylabel('Count')

plt.show()

# Count Plot of Fraud

sampled\_df = df.sample(n=1000)

sns.countplot(x='fraud', data=sampled\_df)

plt.title('Count of Fraudulent Transactions (1000 Random Rows)')

plt.xlabel('Fraud')

plt.ylabel('Count')

plt.show()

# Correlation Heatmap

sampled\_df = df.sample(n=1000)

sns.heatmap(sampled\_df.corr(), annot=True, cmap='coolwarm')

plt.title('Correlation Matrix (1000 Random Rows)')

plt.show()

# Pairplot

# Pairplot of Selected Features

sampled\_df = df.sample(n=1000)

sns.pairplot(sampled\_df[['distance\_from\_home', 'distance\_from\_last\_transaction', 'ratio\_to\_median\_purchase\_price', 'fraud']], hue='fraud')

plt.title('Pairplot of Selected Features (1000 Random Rows)')

plt.show()

**CHAPTER-9**

**RESULTS**

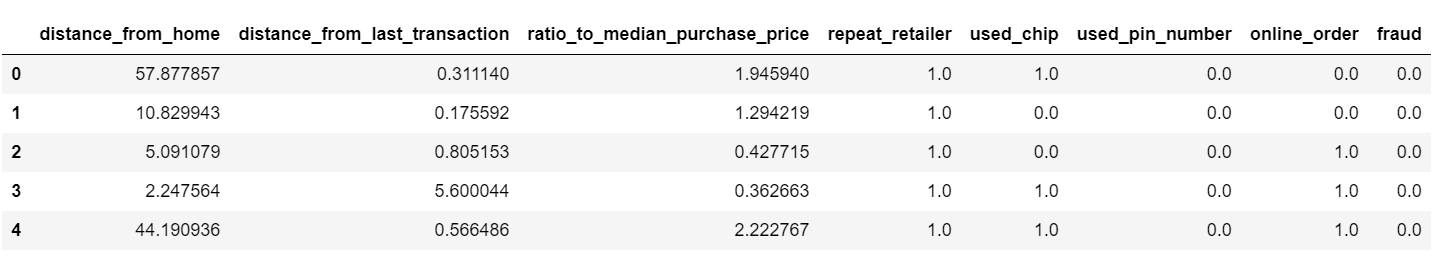


Figure 1: First 5 rows of the dataset.

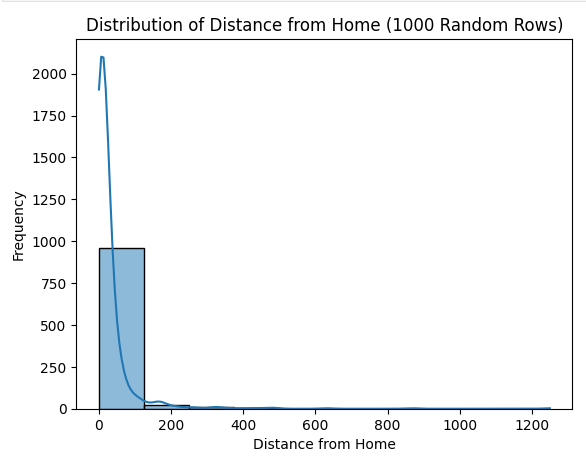


Figure 2: Histplot for the Distance from home column of the dataset.

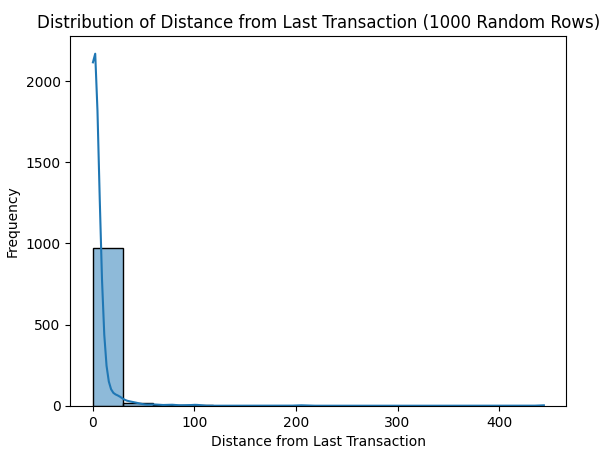


Figure 3: Histplot of the Distance From the Last Transaction column.



Figure 4: Boxplot between the Fraud and Ratio to Median purchase price

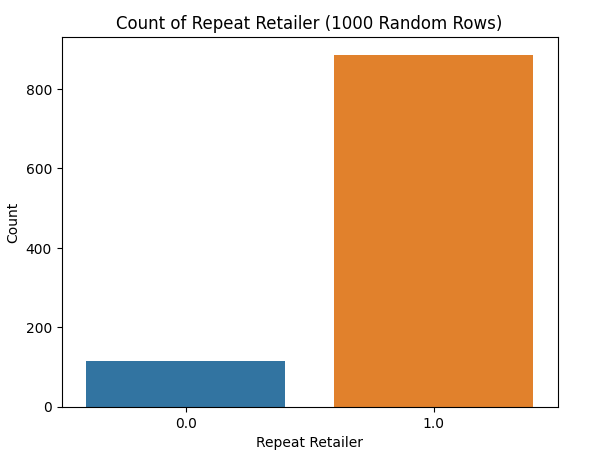


Figure 5: Count plot of the Repeat retailer column of the dataset.

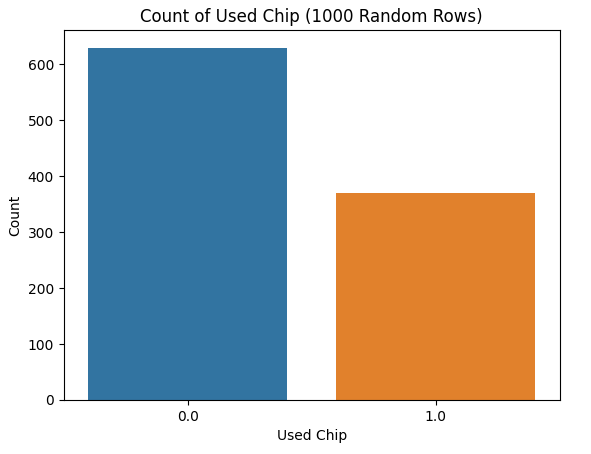


Figure 6: Count plot of the used chips column of the dataset.

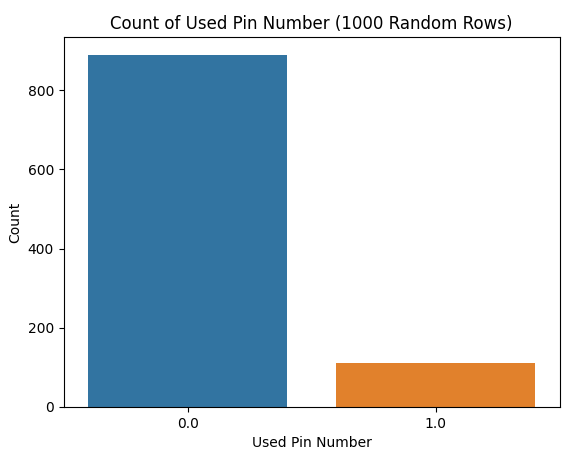


Figure 7: Count plot of the Used pin number column of the dataset.

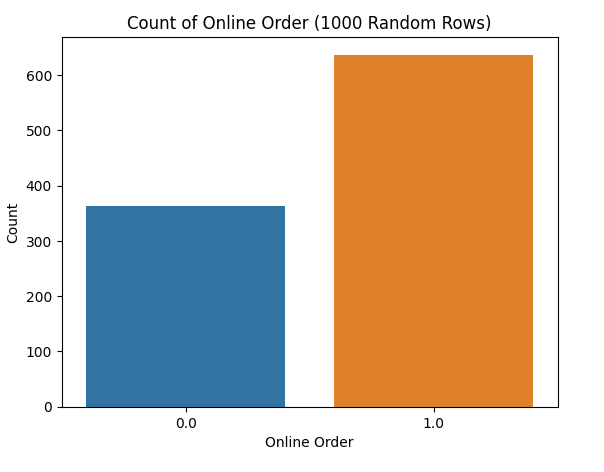


Figure 8: Count plot of the Online order column of the dataset.

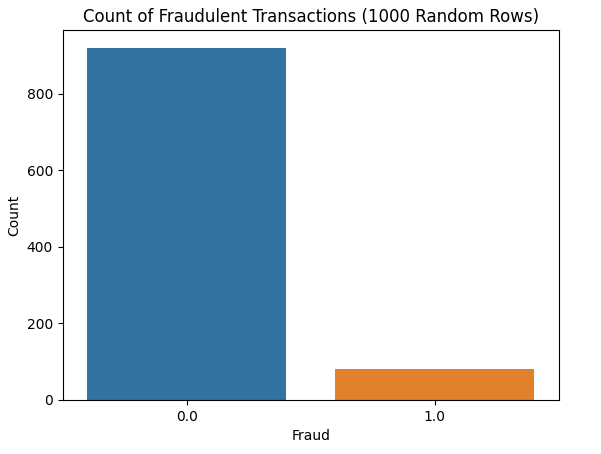


Figure 9: Count plot of the Fraud column of the dataset.

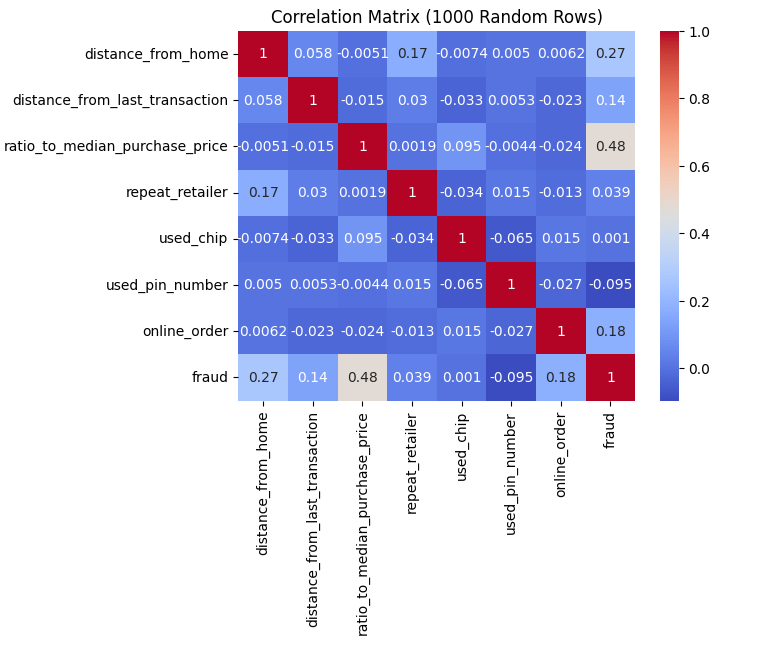


Figure 10: Coreleation Heatmap of the dataset.

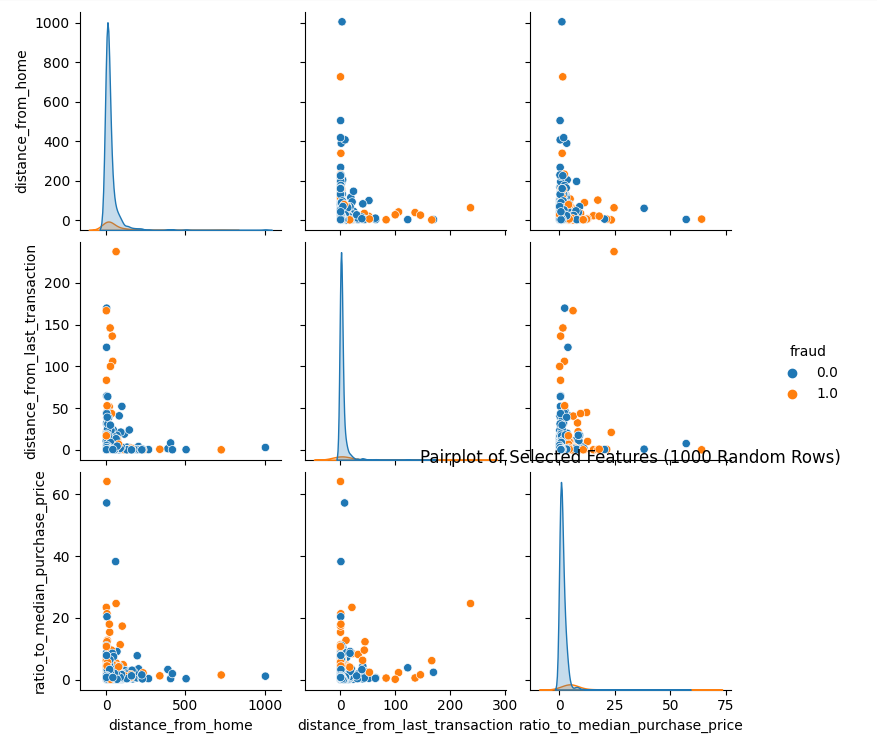


Figure 11: Pair plot between several columns of the dataset.