

Inbox (1, x

phase 2.ipynb x

Mokshan x

Untitled0 x

phase 2.ipynb x

Welcome x

Inbox (1, x

Ebpl-DS x

Credit Co x

+

colab.research.google.com/drive/1DPyzuy-aPBq\_ahO4NXVdSAreuZakgJci

phase 2.ipynb

File Edit View Insert Runtime Tools Help


Commands + Code + Test Copy to Drive

# E. Save Preprocessed Data (Optional)  
# E. Save Preprocessed Data (Optional)  
df.to\_csv("preprocessed\_credit\_card\_data.csv", index=False)  
print("Preprocessed data saved to 'preprocessed\_credit\_card\_data.csv'")

First 5 Rows:  
Transaction\_ID Amount Time Is\_Fraud  
0 T0001 105.50 120 0  
1 T0002 220.10 2400 0  
2 T0003 15.75 350 0  
3 T0004 500.00 7200 1  
4 T0005 60.40 150 0

Dataset Info:  
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 15 entries, 0 to 14  
Data columns (total 4 columns):  
# Column Non-Null Count Dtype  
---  
0 Transaction\_ID 15 non-null object  
1 Amount 15 non-null float64  
2 Time 15 non-null int64  
3 Is\_Fraud 15 non-null int64  
dtypes: float64(1), int64(2), object(1)  
memory usage: 612.0+ bytes  
None

Summary Statistics:  
Amount Time Is\_Fraud  
count 15.000000 15.000000 15.000000  
mean 382.363333 3708.400000 0.333333  
std 353.713605 6167.266578 0.487050  
min 15.710000 120.000000 0.000000  
25% 40.910000 445.000000 0.000000  
50% 195.500000 1111.000000 0.000000  
75% 452.510000 6700.000000 1.000000  
max 1190.000000 13000.000000 1.000000

Distribution of Transaction Amounts  


Variables Terminal



colab.research.google.com/drive/1DPyzuy-aPBq\_ahO4NXVdSAreuZakgJci

phase 2.ipynb

File Edit View Insert Runtime Tools Help

Commands + Code + Test Copy to Drive

```
# 6. Save Preprocessed Data (Optional)
#
et.to_csv("preprocessed_credit_card_data.csv", index=False)
print("Preprocessed data saved to 'preprocessed_credit_card_data.csv'")
```

Amount

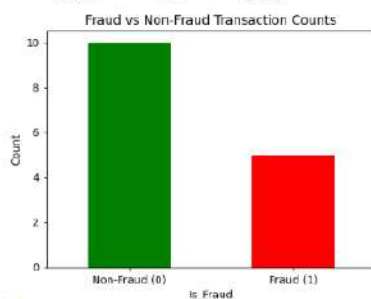
Is Fraud

Time vs Amount by Fraud Status

is\_fraud

Amount

The figure displays two visualizations from a Jupyter Notebook. The top visualization is a box plot showing the distribution of 'Amount' for two categories of 'Is Fraud' (0 and 1). The y-axis represents 'Amount' ranging from 0 to 1000. For 'Is Fraud' = 0, the median amount is approximately 50, with a few outliers around 200 and 300. For 'Is Fraud' = 1, the median amount is approximately 700, with a range from about 400 to 1000. The bottom visualization is a scatter plot titled 'Time vs Amount by Fraud Status' showing 'Amount' on the y-axis (200 to 1000) against 'is\_fraud' on the x-axis. Blue dots represent 'is\_fraud' = 0, and orange dots represent 'is\_fraud' = 1. The scatter plot shows a clear upward trend for 'is\_fraud' = 1, with amounts ranging from approximately 400 to 1000, while 'is\_fraud' = 0 has much lower amounts, mostly below 400.



Preprocessed data saved to 'preprocessed\_credit\_card\_data.csv'