

```
from google.colab import files
uploaded = files.upload()
import pandas as pd
df = pd.read_csv('peer_to_peer.csv')
df.head()
```

 [Show hidden output](#)

✓ **NAME : MOHIT BANKHELE**


**PRN : 202401040059**

DIV : CS2-19


```
# 1. Average value of USD column
print("1. Average USD:", round(df['USD'].mean(), 2))
```

 1. Average USD: 14863.44


```
# 2. Number of unique Threat types
print("2. Unique Threat types:", df['Threats'].nunique())
```

 2. Unique Threat types: 9


```
# 3. Unique protocols used
print("3. Protocols:", df['Protcol'].unique())
```

 3. Protocols: ['TCP' 'UDP' 'ICMP']

```
# 4. Entries with Netflow_Bytes > 10
print("4. Rows with Netflow_Bytes > 10:", (df['Netflow_Bytes'] > 10).sum())
```

 4. Rows with Netflow\_Bytes > 10: 148132


```
# 5. Total BTC transferred
print("5. Total BTC:", round(df['BTC'].sum(), 2))
```

 5. Total BTC: 4553950

```
# 6. Max USD transaction
print("6. Max USD:", round(df['USD'].max(), 2))
```

 6. Max USD: 126379

```
# 7. Top 3 entries with highest USD
print("7. Top 3 USD rows:\n", df.nlargest(3, 'USD'))
```

 7. Top 3 USD rows:

	Time	Protcol	Flag	Family	Clusters	SeddAddress	ExpAddress	BTC	\
2649	29	UDP	AF	DMALocker	3	17dcMo4V	1DiCeTjB	30	

7673	20	UDP	AF	NoobCrypt	3	17dcMo4V	1DiCeTjB	30
13053	29	UDP	AF	JigSaw	3	17dcMo4V	1DiCeTjB	30

	USD	Netflow_Bytes	IPaddress	Threats	Port	Prediction
2649	126379		216	D UDP Scan	5068	S
7673	126379		1440	D UDP Scan	5068	S
13053	126379		216	D Spam	5068	S

# 8. Count of entries with WannaCry threat

```
print("8. WannaCry entries:", df[df['Threats'] == 'WannaCry'].shape[0])
```

➞ 8. WannaCry entries: 0

# 9. Average Netflow\_Bytes per Protocol

```
print("9. Avg Netflow_Bytes by Protocol:\n", df.groupby('Protocol')['Netflow_Bytes'].mean())
```

➞ 9. Avg Netflow\_Bytes by Protocol:

Protocol	
ICMP	2482.097325
TCP	2082.956955
UDP	1525.874084

Name: Netflow\_Bytes, dtype: float64

# 10. Unique prediction labels

```
print("10. Prediction Labels:", df['Prediction'].unique())
```

➞ 10. Prediction Labels: ['SS' 'A' 'S']

# 11. Number of unique IPs (combined source & destination)

```
unique_ips = pd.concat([df['IPaddress'], df['SeddAddress']]).nunique()
print("11. Unique IP addresses:", unique_ips)
```

➞ 11. Unique IP addresses: 10

# 12. Standard deviation of USD

```
print("12. USD Std Dev:", round(df['USD'].std(), 2))
```

➞ 12. USD Std Dev: 26849.43

# 13. Most common Port

```
print("13. Most common Port:", df['Port'].mode()[0])
```

➞ 13. Most common Port: 5066

# 14. Correlation between BTC and USD

```
print("14. BTC↔USD Correlation:", round(df['BTC'].corr(df['USD']), 3))
```

➞ 14. BTC↔USD Correlation: -0.059

# 15. Rows where BTC = 1 and USD > 510

```
print("15. BTC=1 & USD>510 count:", df[(df['BTC'] == 1) & (df['USD'] > 510)].shape[0])
```

➞ 15. BTC=1 & USD>510 count: 210

```
# 16. Median of Time column
print("16. Median Time:", df['Time'].median())
```

➞ 16. Median Time: 19.0

```
#17. Unique export addresses
print("17. Unique Export Addresses:", df['ExpAddress'].nunique())
```

➞ 17. Unique Export Addresses: 7

```
# 18. Entry with minimum Netflow_Bytes
print("18. Min Netflow_Bytes row:\n", df[df['Netflow_Bytes'] == df['Netflow_Bytes'].min()])
```

➞ 18. Min Netflow\_Bytes row:

	Time	Protocol	Flag	Family	Clusters	SeddcAddress	ExpAddress	BTC	\
2439	54	TCP	AF	DMALocker	3	17dcMo4V	1DiCeTjB	13	
2451	20	TCP	AF	DMALocker	3	17dcMo4V	1DiCeTjB	13	
12843	54	TCP	AF	Locky	3	17dcMo4V	1DiCeTjB	13	
12855	20	TCP	AF	Locky	3	17dcMo4V	1DiCeTjB	13	
17868	53	TCP	AF	DMALocker	3	17dcMo4V	1DiCeTjB	13	
17880	20	TCP	AF	DMALocker	3	17dcMo4V	1DiCeTjB	13	
23417	54	TCP	AF	Cryptohitman	3	17dcMo4V	1DiCeTjB	13	
23429	20	TCP	AF	Cryptohitman	3	17dcMo4V	1DiCeTjB	13	
32393	54	TCP	AF	Globe	3	17dcMo4V	1DiCeTjB	13	
32405	20	TCP	AF	Globe	3	17dcMo4V	1DiCeTjB	13	
37311	53	TCP	AF	EDA2	3	17dcMo4V	1DiCeTjB	13	
37323	20	TCP	AF	EDA2	3	17dcMo4V	1DiCeTjB	13	
43438	54	TCP	AF	TowerWeb	3	17dcMo4V	1DiCeTjB	13	
43450	20	TCP	AF	TowerWeb	3	17dcMo4V	1DiCeTjB	13	
51954	54	TCP	AF	WannaCry	3	17dcMo4V	1DiCeTjB	13	
51966	20	TCP	AF	WannaCry	3	17dcMo4V	1DiCeTjB	13	
56768	53	TCP	AF	WannaCry	3	17dcMo4V	1DiCeTjB	13	
56780	20	TCP	AF	WannaCry	3	17dcMo4V	1DiCeTjB	13	
61562	54	TCP	AF	Cryptohitman	3	17dcMo4V	1SYSTEMQ	13	
61574	20	TCP	AF	Cryptohitman	3	17dcMo4V	1SYSTEMQ	13	
68763	54	TCP	AF	Razy	3	17dcMo4V	1DiCeTjB	13	
68775	20	TCP	AF	Razy	3	17dcMo4V	1DiCeTjB	13	
71827	53	TCP	AF	SamSam	3	17dcMo4V	1DiCeTjB	13	
71839	20	TCP	AF	SamSam	3	17dcMo4V	1DiCeTjB	13	
76328	54	TCP	AF	SamSam	3	17dcMo4V	1DiCeTjB	13	
76340	20	TCP	AF	SamSam	3	17dcMo4V	1DiCeTjB	13	
83070	54	TCP	AF	SamSam	3	17dcMo4V	1DiCeTjB	13	
83082	20	TCP	AF	SamSam	3	17dcMo4V	1DiCeTjB	13	
87341	53	TCP	AF	Flyper	3	17dcMo4V	1DiCeTjB	13	
87353	20	TCP	AF	Flyper	3	17dcMo4V	1DiCeTjB	13	
112218	54	TCP	AF	DMALocker	3	17dcMo4V	1DiCeTjB	13	
112230	20	TCP	AF	DMALocker	3	17dcMo4V	1DiCeTjB	13	
122602	53	TCP	AF	Locky	3	17dcMo4V	1DiCeTjB	13	
122614	20	TCP	AF	Locky	3	17dcMo4V	1DiCeTjB	13	
128759	54	TCP	AF	JigSaw	3	17dcMo4V	1DiCeTjB	13	
128771	20	TCP	AF	JigSaw	3	17dcMo4V	1DiCeTjB	13	
133213	53	TCP	AF	Locky	3	17dcMo4V	1DiCeTjB	13	
133225	20	TCP	AF	Locky	3	17dcMo4V	1DiCeTjB	13	
143200	54	TCP	AF	DMALocker	3	17dcMo4V	1DiCeTjB	13	
143212	20	TCP	AF	DMALocker	3	17dcMo4V	1DiCeTjB	13	
147054	53	TCP	AF	JigSaw	3	17dcMo4V	1DiCeTjB	13	
147066	20	TCP	AF	JigSaw	3	17dcMo4V	1DiCeTjB	13	

	USD	Netflow_Bytes	IPaddress	Threats	Port	Prediction
2439	125539	1	B	UDP Scan	5062	S
2451	125587	1	B	UDP Scan	5062	S
12843	125539	1	B	Spam	5062	S

12855	125587	1	B	Spam	5062	S
17868	125539	1	B	Port Scanning	5062	S
17880	125587	1	B	Port Scanning	5062	S
23417	125539	1	B	UDP Scan	5062	S
23429	125587	1	B	UDP Scan	5062	S
32393	125539	1	B	Spam	5062	S
32405	125587	1	B	Spam	5062	S
37311	125539	1	B	Port Scanning	5062	S
37323	125587	1	B	Port Scanning	5062	S

# 19. Average USD per Flag

```
print("19. Avg USD by Flag:\n", df.groupby('Flag')['USD'].mean().round(2))
```



19. Avg USD by Flag:

Flag

A	1678.17
AF	27357.39
AP	13701.89
APRSF	506.08
APS	1260.48
APSF	7469.28
ARF	13838.73
ASF	618.00
R	15989.32

Name: USD, dtype: float64

# 20. Rows with Port = 5061 and Prediction = 'SS'

```
print("20. Port=5061 & Prediction='SS' rows:", df[(df['Port'] == 5061) & (df['Prediction'] == 'SS')])
```



20. Port=5061 & Prediction='SS' rows: 4306