```
# Data Handling
import pandas as pd
import numpy as np

# Save/Export Datasets
from google.colab import files


# Load CSVs
features_df = pd.read_csv('datasets/elliptic_txs_features.csv')
classes_df = pd.read_csv('datasets/elliptic_txs_classes.csv')
```

```
/tmp/ipython-input-3-1203509627.py:2: DtypeWarning: Columns (108) have mixe
  features_df = pd.read_csv('elliptic_txs_features.csv')
```

```
# Quick preview
print("Features shape:", features_df.shape)
display(features_df.head())

print("Classes shape:", classes_df.shape)
display(classes_df.head())
```

```
Features shape: (124130, 167)
```

	230425980	1	-0.1714692896288031	-0.18466755143291433	-1.2013688016765
0	5530458	1	-0.171484	-0.184668	-1.201
1	232022460	1	-0.172107	-0.184668	-1.201
2	232438397	1	0.163054	1.963790	-0.646
3	230460314	1	1.011523	-0.081127	-1.201
4	230459870	1	0.961040	-0.081127	-1.201

5 rows × 167 columns

```
Classes shape: (203769, 2)
```

	txId	class
0	230425980	unknown
1	5530458	unknown
2	232022460	unknown
3	232438397	2
4	230460314	unknown

```
# Assign proper names to transaction id and time step features
features_df.rename(columns={features_df.columns[0]: 'txId', features_df.columns

print("Features shape:", features_df.shape)
display(features_df.head())
```

```
Features shape: (124130, 167)
```

	txId	timeStep	-0.1714692896288031	-0.18466755143291433	-1.2013688
0	5530458	1	-0.171484	-0.184668	
1	232022460	1	-0.172107	-0.184668	
2	232438397	1	0.163054	1.963790	
3	230460314	1	1.011523	-0.081127	
4	230459870	1	0.961040	-0.081127	

5 rows × 167 columns

```
# Merge the datasets
transactions_df = pd.merge(features_df, classes_df, on='txId', how='left')

print("Merged data shape:", transactions_df.shape)
transactions_df.head()
```

```
Merged data shape: (124130, 168)
```

	txId	timeStep	-0.1714692896288031	-0.18466755143291433	-1.2013688
0	5530458	1	-0.171484	-0.184668	
1	232022460	1	-0.172107	-0.184668	
2	232438397	1	0.163054	1.963790	
3	230460314	1	1.011523	-0.081127	
4	230459870	1	0.961040	-0.081127	

5 rows × 168 columns

```
# Drop any rows with missing data
transactions_df = transactions_df.dropna()
display(transactions_df)
```

	txId	timeStep	-0.1714692896288031	-0.18466755143291433	-1.20
0	5530458	1	-0.171484	-0.184668	
1	232022460	1	-0.172107	-0.184668	
2	232438397	1	0.163054	1.963790	
3	230460314	1	1.011523	-0.081127	
4	230459870	1	0.961040	-0.081127	
...	...	...	...	...	
124124	335403844	31	-0.150914	-0.103078	
124125	376207023	31	-0.164305	-0.132897	
124126	382945607	31	-0.172439	-0.128652	
124127	382945604	31	-0.172451	-0.172527	
124128	382907521	31	-0.169806	-0.127651	

124129 rows × 168 columns

```
# Find out all the values in class
transactions_df['class'].value_counts()
```

class	count
unknown	97019
2	24126
1	2984

dtype: int64

```
# Filter only transactions where class is 'unknown'
transactions_df = transactions_df[transactions_df['class'] == 'unknown']
display(transactions_df)
transactions_df['class'].value_counts()
```

	txId	timeStep	-0.1714692896288031	-0.18466755143291433	-1.20
0	5530458	1	-0.171484	-0.184668	
1	232022460	1	-0.172107	-0.184668	
3	230460314	1	1.011523	-0.081127	
4	230459870	1	0.961040	-0.081127	
5	230333930	1	-0.171264	-0.184668	
...	...	...	...	...	
124123	341106370	31	-0.171406	-0.127651	
124124	335403844	31	-0.150914	-0.103078	
124125	376207023	31	-0.164305	-0.132897	
124127	382945604	31	-0.172451	-0.172527	
124128	382907521	31	-0.169806	-0.127651	

97019 rows × 168 columns

class	count
unknown	97019

dtype: int64

```
# Drop the class column/feature
transactions_df.drop(columns=['class'], inplace=True)
display(transactions_df)
```

	txId	timeStep	-0.1714692896288031	-0.18466755143291433	-1.20
0	5530458	1	-0.171484	-0.184668	
1	232022460	1	-0.172107	-0.184668	
3	230460314	1	1.011523	-0.081127	
4	230459870	1	0.961040	-0.081127	
5	230333930	1	-0.171264	-0.184668	
...	...	...	...	...	
124123	341106370	31	-0.171406	-0.127651	
124124	335403844	31	-0.150914	-0.103078	
124125	376207023	31	-0.164305	-0.132897	
124127	382945604	31	-0.172451	-0.172527	
124128	382907521	31	-0.169806	-0.127651	

97019 rows × 167 columns

```
# Change column names back to original
transactions_df.rename(columns={transactions_df.columns[0]: '230425980', transa
display(transactions_df)
```

	230425980	1	-0.1714692896288031	-0.18466755143291433	-1.2013688(
0	5530458	1	-0.171484	-0.184668	
1	232022460	1	-0.172107	-0.184668	
3	230460314	1	1.011523	-0.081127	
4	230459870	1	0.961040	-0.081127	
5	230333930	1	-0.171264	-0.184668	
...	...	...	...	...	
124123	341106370	31	-0.171406	-0.127651	
124124	335403844	31	-0.150914	-0.103078	
124125	376207023	31	-0.164305	-0.132897	
124127	382945604	31	-0.172451	-0.172527	
124128	382907521	31	-0.169806	-0.127651	

97019 rows × 167 columns

```
# Save and export entire deployment data
transactions_df.to_csv('all_deployment_data.csv', index=False)
files.download('all_deployment_data.csv')
```

```
# Save and export smaller sample dataset
transactions_df.head(300).to_csv('sample_deployment_data.csv', index=False)
files.download('sample_deployment_data.csv')
```