Bring in Data and Import Libraries

```python
# Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import re
import plotly.graph_objects as go

from sklearn.base import BaseEstimator, TransformerMixin
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import FunctionTransformer, OneHotEncoder, MultiLabelBinarizer, LabelEncoder
from sklearn.pipeline import Pipeline
from sklearn.ensemble import RandomForestClassifier, BaggingClassifier, GradientBoostingClassifier
from sklearn.model_selection import train_test_split, cross_val_score, validation_curve
from sklearn.metrics import accuracy_score, classification_report
from scipy import sparse
from sklearn.decomposition import PCA
from sklearn.naive_bayes import BernoulliNB
from sklearn.multiclass import OneVsRestClassifier
from sklearn.linear_model import SGDClassifier
```

```python
# Bring in Data (https://www.kaggle.com/datasets/tannubarot/cybersecurity-attack-and-defence-dataset)
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```python
# file_path = '/content/Attack_Dataset.csv'
file_path = '/content/drive/MyDrive/info_4604/homework/Final_Project/Attack_Dataset.csv'
df = pd.read_csv(file_path)
```

Preprocessing

2. Complete Preprocessing Pipeline (20 pts) Must use:

- SimpleImputer for missing values
- OneHotEncoder or OrdinalEncoder for categorical features
- StandardScaler or MinMaxScaler for numeric features
- ColumnTransformer to manage preprocessing
- Wrap preprocessing + model in a Pipeline

```python
# Rename columns to be more usable later on
df = df.rename(columns={
    'Attack Type': 'Attack_Type',
    'Tools Used': 'Tools_Used',
    'Target Type': 'Target_Type',
    'Detection Method': 'Detection_Method',
    'MITRE Technique': 'MITRE_Technique',
    })

## Define Features and Target - Select actual data from DataFrame
features = ['Category', 'Attack_Type', 'Tools_Used_items', 'Target_Type_items', 'Detection_Method_items', 'Tags_items', 'Source_items']
target = 'MITRE_Technique_items' # Target is the column containing the list of techniques
```

Clean and Split Columns into Lists for the MultiLabelBinarizer

```python
def clean_and_split_column(arr):
    """
    Expects a 1D numpy array of strings.
    Returns a list-of-lists (of tokens) of same length.
    Drops any parenthetical comments, then removes
    ' – ' (or '-' or '—') plus everything after up to newline,
    then splits on commas, semicolons, slashes, or just before '#'.
    """
    # split on comma, semicolon, slash, or before '#'
    split_pattern = re.compile(r',|;|(?=#)|\s*/\s*')
    # match ASCII hyphen (-), en-dash (–) or em-dash (—) with spaces around
    dash_pattern = re.compile(r'\s[-–—]\s.*?(?=\n|$)', flags=re.DOTALL)
    # match any parentheses and their contents
    paren_pattern = re.compile(r'\s*\(.*?\)')

    cleaned = []
    for val in arr:
        s = val if isinstance(val, str) else ''
        # 1) strip out parentheses comments
        s = paren_pattern.sub('', s)
        # 2) remove dash + trailing text
        s = dash_pattern.sub('', s)
        # 3) split into tokens
        parts = split_pattern.split(s)
        toks = [p.strip() for p in parts if p and p.strip()]
        cleaned.append(toks)

    return np.array(cleaned, dtype=object)
```

```python
'''
Here I am applying the clean_and_split_column function to certain columns because
There are multiple items in one cell, separated by commas, semicolons, slashes, etc.

This is so we can use MultiLabelBinarizer to binarize the data for exploration later on.
'''

# Drop rows in MITRE_Technique that are empty
df.dropna(subset=['MITRE_Technique'], inplace=True)

# Apply clean_and_split_column function to columns with multiple inputs to convert into list
for col in ['Tools_Used', 'Target_Type', 'MITRE_Technique', 'Detection_Method', 'Tags', 'Source']:
    df[f'{col}_items'] = clean_and_split_column(df[col])
    # `df[f'{col}_items']` is now a list-of-tokens column,
    # `item_dict` maps each token to its row indices.

# Drop raw data rows we converted and useless features
useless_feat = ['Tools_Used', 'Target_Type',
                'MITRE_Technique', 'Detection_Method',
                'Tags', 'Source', 'ID', 'Title',
                'Scenario Description', 'Attack Steps ',
                'Impact', 'Solution', 'Vulnerability']
df = df.drop(columns=useless_feat)
```

OneHotEncode and MultiLabelBinarizer

```python
class MultiLabelBinarizerTransformer(BaseEstimator, TransformerMixin):
    def __init__(self):
        self.mlb = MultiLabelBinarizer()

    def fit(self, X, y=None):
        return self.mlb.fit(X.iloc[:, 0])

    def transform(self, X):
        return self.mlb.transform(X.iloc[:, 0])

    def get_feature_names_out(self, input_features=None):
        return [f"{input_features[0]}__{cls}" for cls in self.mlb.classes_]


class MultiColumnMultiLabelBinarizer(BaseEstimator, TransformerMixin):
    def __init__(self, columns):
        self.columns = columns
        # tell each MLBI to give us sparse CSR output
        self.mlbs = {
            col: MultiLabelBinarizer() # Removed sparse_output=True
            for col in columns
        }

    def fit(self, X, y=None):
        # X might be a DF or 2D array
        df = X if hasattr(X, 'loc') else pd.DataFrame(X, columns=self.columns)
        for col in self.columns:
            self.mlbs[col].fit(df[col])
        return self

    def transform(self, X):
        df = X if hasattr(X, 'loc') else pd.DataFrame(X, columns=self.columns)
        blocks = []
        for col in self.columns:
            # each arr is a scipy CSR matrix of shape (n_samples, n_classes_for_col)
            arr = self.mlbs[col].transform(df[col])
            blocks.append(arr)
        # hstack into one big CSR
        # Need to hstack dense arrays now
        return np.hstack(blocks)


    def get_feature_names_out(self, input_features=None):
        names = []
        for col in self.columns:
            for cls in self.mlbs[col].classes_:
                names.append(f"{col}__{cls}")
        return names
```

Pipeline

```python
list_feats = ['Tools_Used_items',
              'Target_Type_items',
              'Detection_Method_items',
              'Tags_items',
              'Source_items'
              ]
# Make sure each cell is actually a Python list, not a string
for col in list_feats:
    df[col] = df[col].apply(lambda s: ast.literal_eval(s) if isinstance(s, str) else s)

solo_feats = ['Category', 'Attack_Type']
```

```python
# OneHotEncode Pipeline
ohe_pipe = Pipeline([
    ('encoder', OneHotEncoder(handle_unknown='ignore'))
])

# MultiLabelBinarizerTransformer Pipeline
list_column_pipe = Pipeline([
    ('multi_label_binarizer', MultiLabelBinarizerTransformer())
])


# Preprocess Pipeline
preprocess = ColumnTransformer([
    ('cat',  ohe_pipe, solo_feats),
    ('lists', MultiColumnMultiLabelBinarizer(list_feats), list_feats)
])
```

Explore Data

1. Full Exploratory Data Analysis (20 pts)

- Implement data exploration methods to understand the data
- At least 5 visualizations, with titles for readability

Markdown summary of:

- What the dataset contains
- Target variable + feature types
- Any data quality issues (missing, skewed, imbalanced)

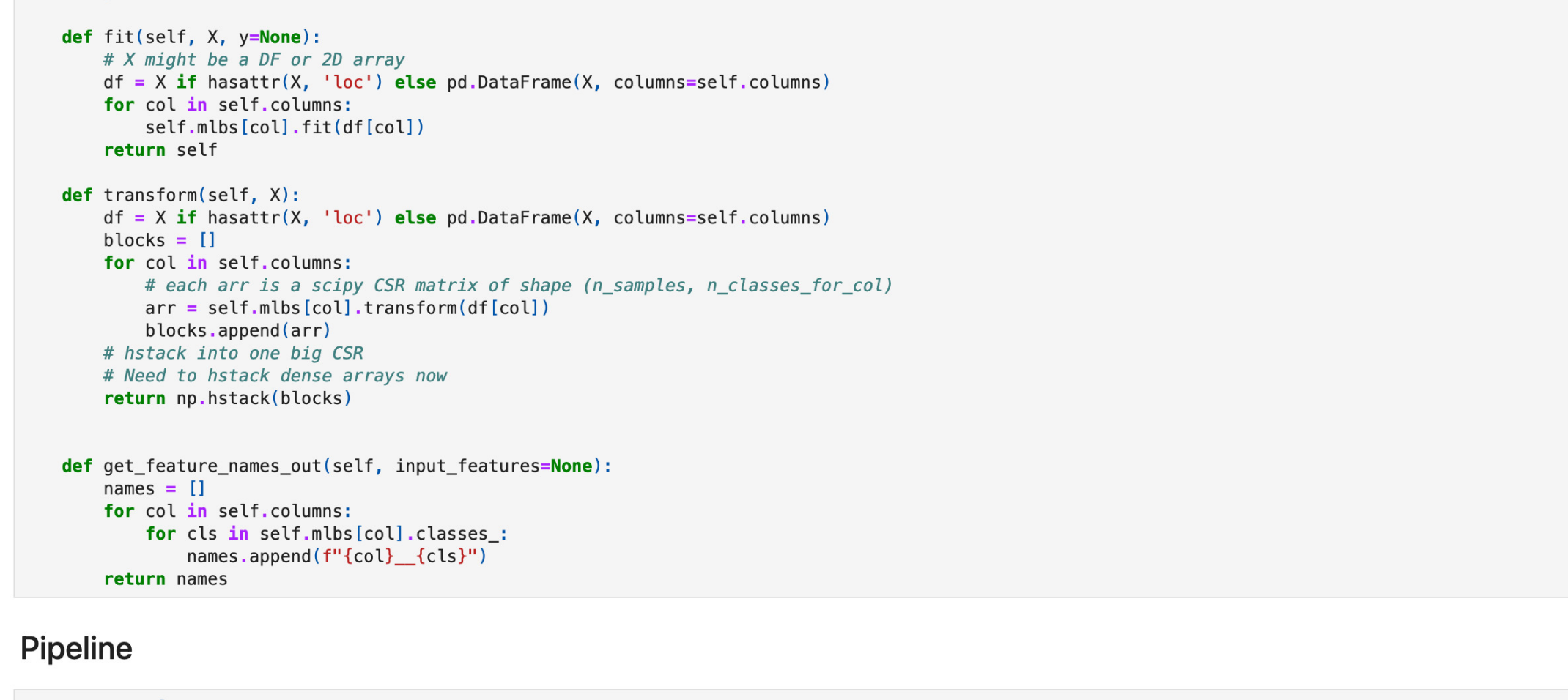
```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 14109 entries, 0 to 14132
Data columns (total 9 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Category                14109 non-null  object
 1   Attack_Type             14109 non-null  object
 2   Unnamed: 15             45 non-null     object
 3   Tools_Used_items        14109 non-null  object
 4   Target_Type_items       14109 non-null  object
 5   MITRE_Technique_items   14109 non-null  object
 6   Detection_Method_items  14109 non-null  object
 7   Tags_items              14109 non-null  object
 8   Source_items            14109 non-null  object
dtypes: object(9)
memory usage: 1.1+ MB
```

```python
df.describe(include='all').T
```

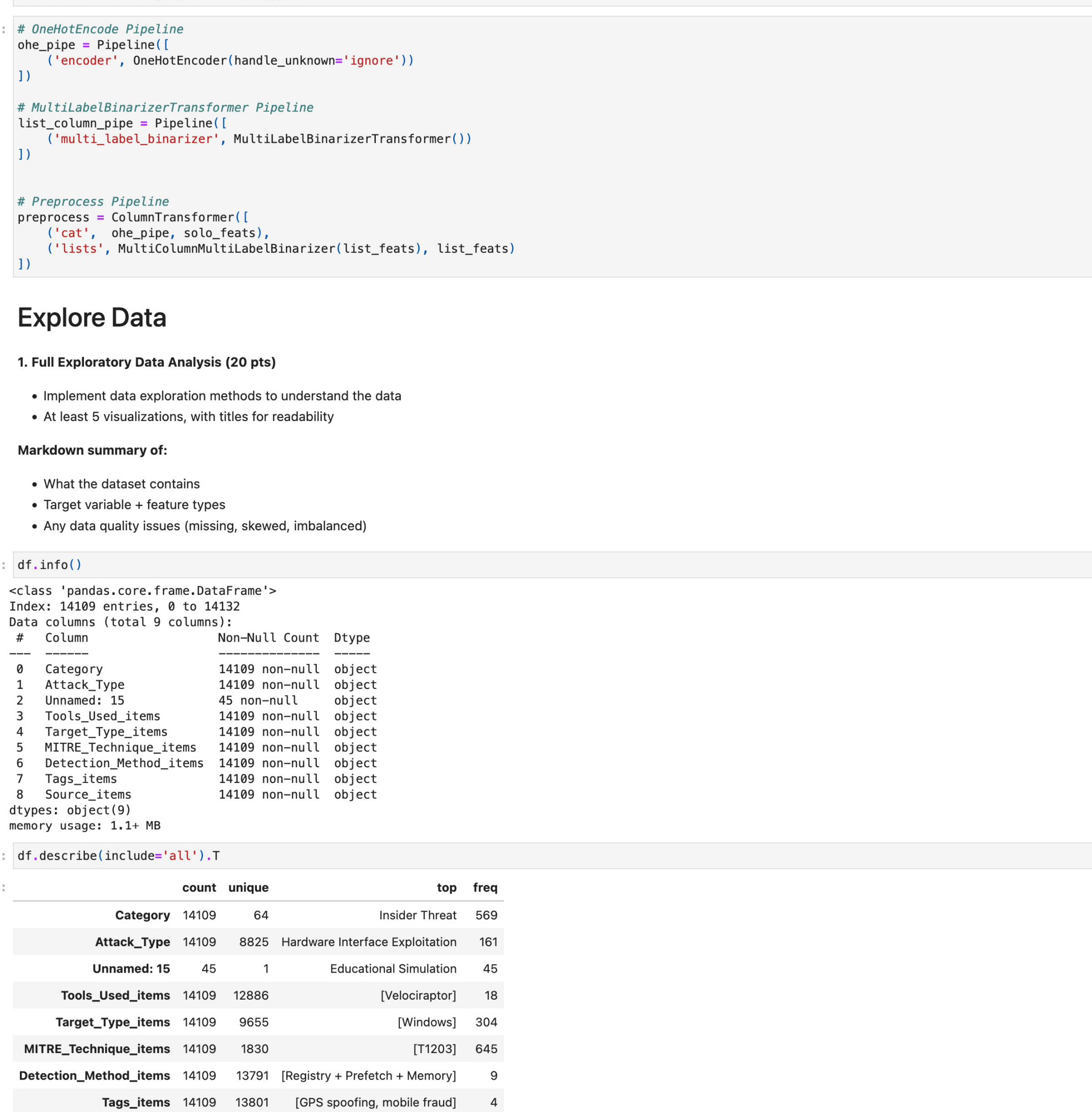
	count	unique	top	freq
Category	14109	64	Insider Threat	569
Attack_Type	14109	8825	Hardware Interface Exploitation	161
Unnamed: 15	45	1	Educational Simulation	45
Tools_Used_items	14109	12886	[Velociraptor]	18
Target_Type_items	14109	9655	[Windows]	304
MITRE_Technique_items	14109	1830	[T1203]	645
Detection_Method_items	14109	13791	[Registry + Prefetch + Memory]	9
Tags_items	14109	13801	[GPS spoofing, mobile fraud]	4
Source_items	14109	7000	[Simulated]	2347

```python
# Distribution of MITRE Techniques (Top N)
mlb_techniques = MultiLabelBinarizer()
tech_matrix = mlb_techniques.fit_transform(df['MITRE_Technique_items'])
tech_counts = pd.Series(tech_matrix.sum(axis=0), index=mlb_techniques.classes_).sort_values(ascending=False)

top_n_techs = 20
top_tech_counts = tech_counts.head(top_n_techs)

plt.figure(figsize=(12, 6))
top_tech_counts.plot(kind='bar')
plt.title(f'Top {top_n_techs} Most Frequent MITRE Techniques')
plt.xlabel('MITRE Technique')
plt.ylabel('Frequency')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```



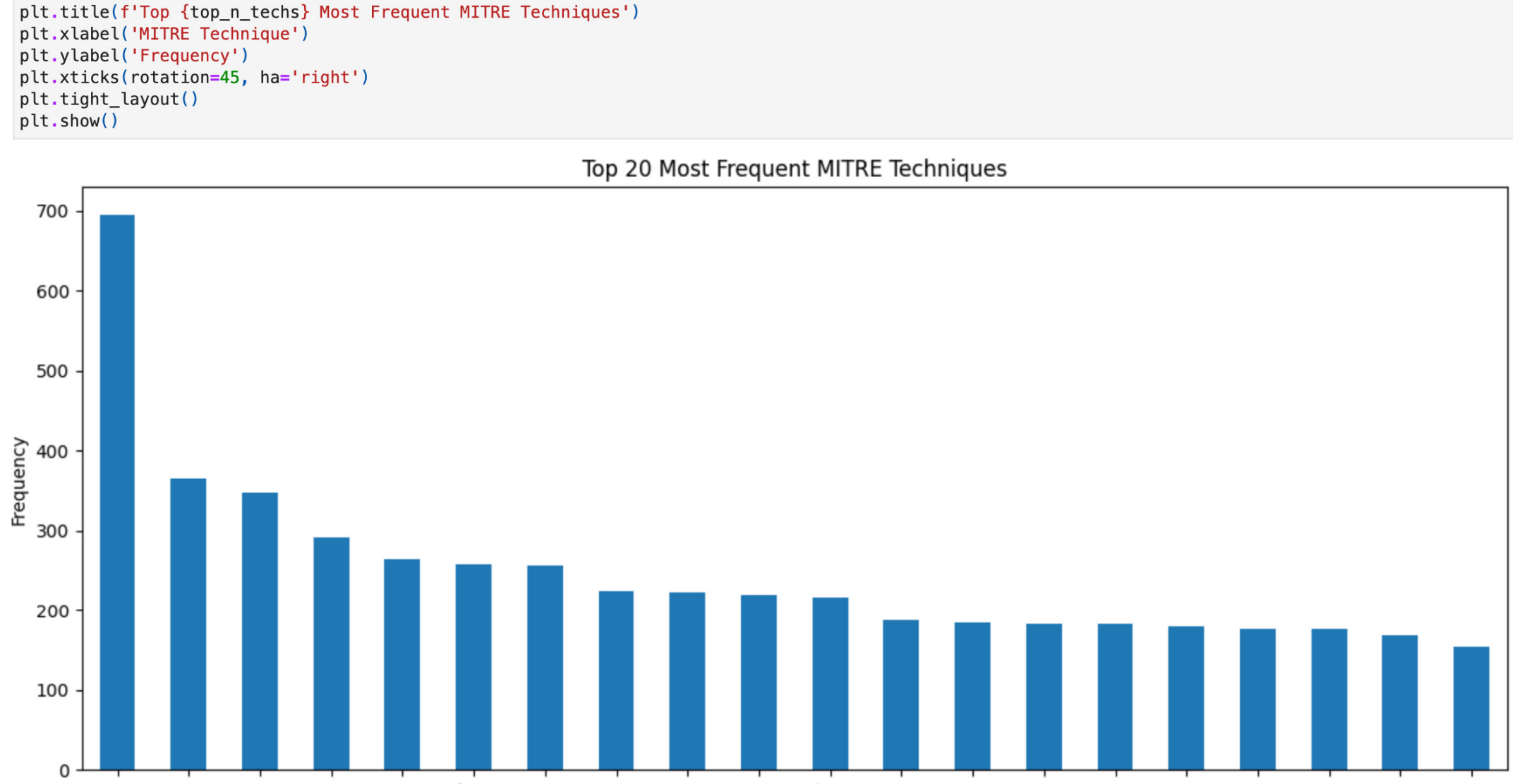
```python
# Category Frequency Bar Chart
'''
Gives a quick view of class imbalance—critical for classification tasks—so you can
decide whether to re-sample, weight classes, or aggregate rare categories
'''
category_counts = df['Category'].value_counts()

fig, ax = plt.subplots(figsize=(12, 6))
ax.bar(category_counts.index, category_counts.values)
ax.set_xticklabels(category_counts.index, rotation=45, ha='right')
ax.set_xlabel('Category')
ax.set_ylabel('Count')
ax.set_title('Category Frequency')
plt.tight_layout()
plt.show()
```

```
/tmp/ipython-input-2633509385.py:10: UserWarning: set_ticklabels() should only be used with a fixed number of ticks, i.e. after set_ticks() or using a FixedL
ocator.
  ax.set_xticklabels(category_counts.index, rotation=45, ha='right')
/tmp/ipython-input-2633509385.py:14: UserWarning: Glyph 128737 (\N{SHIELD}) missing from font(s) DejaVu Sans.
  plt.tight_layout()
/tmp/ipython-input-2633509385.py:14: UserWarning: Tight layout not applied. The bottom and top margins cannot be made large enough to accommodate all Axes de
corations.
  plt.tight_layout()
/usr/local/lib/python3.11/dist-packages/IPython/core/pylabtools.py:151: UserWarning: Glyph 128737 (\N{SHIELD}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
```

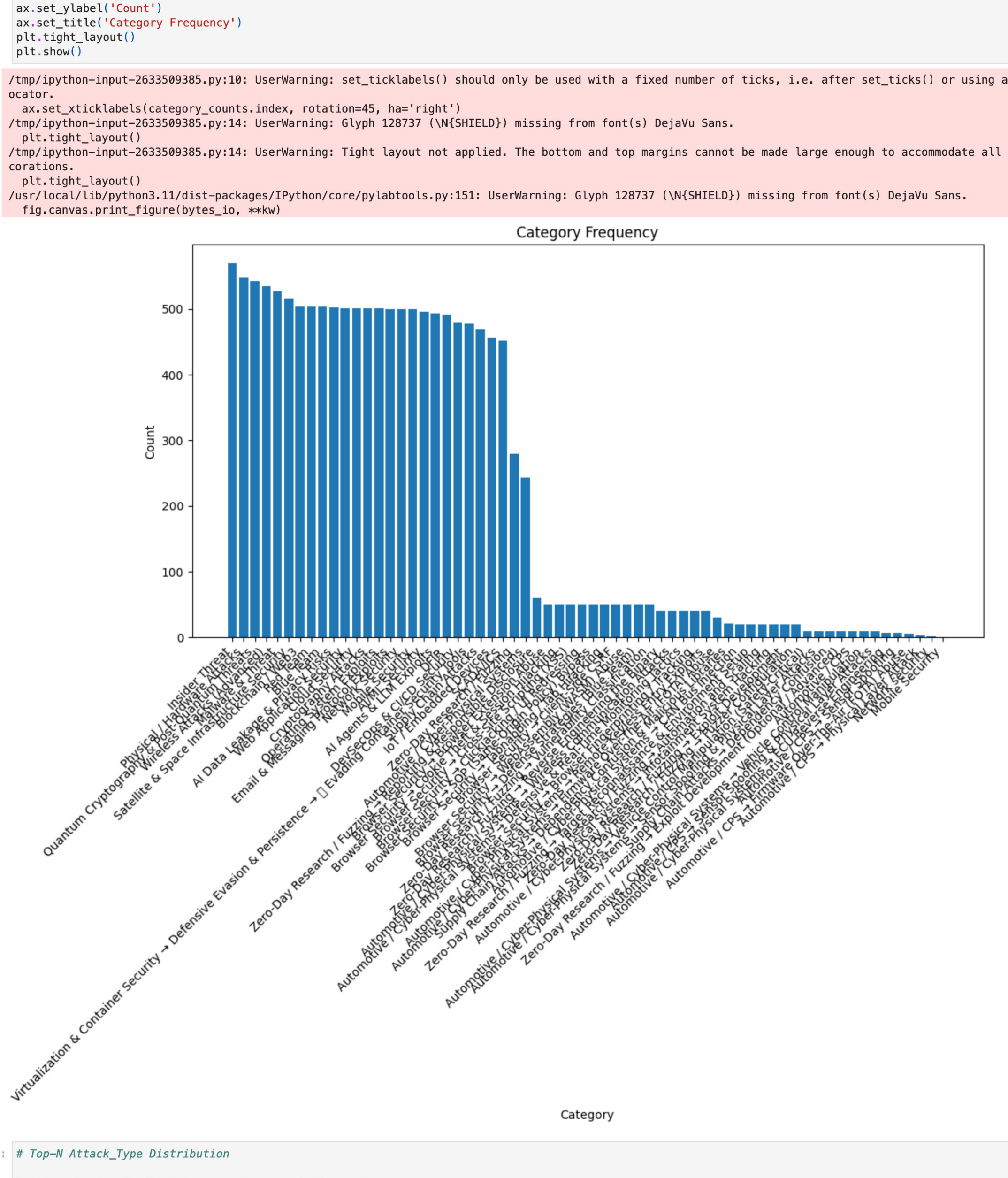


```python
# Top-N Attack_Type Distribution

# Determine top N attack types and aggregate the rest
top_n = 20
attack_counts = df['Attack_Type'].value_counts()
top_counts = attack_counts.nlargest(top_n)
other_count = attack_counts.iloc[top_n:].sum()
top_counts['Other'] = other_count

# Plot
fig, ax = plt.subplots(figsize=(12, 6))
ax.bar(top_counts.index, top_counts.values)
ax.set_xticklabels(top_counts.index, rotation=45, ha='right')
ax.set_xlabel('Attack Type')
ax.set_ylabel('Count')
ax.set_title(f'Top {top_n} Attack Types (with Others Aggregated)')
plt.tight_layout()
plt.show()
```

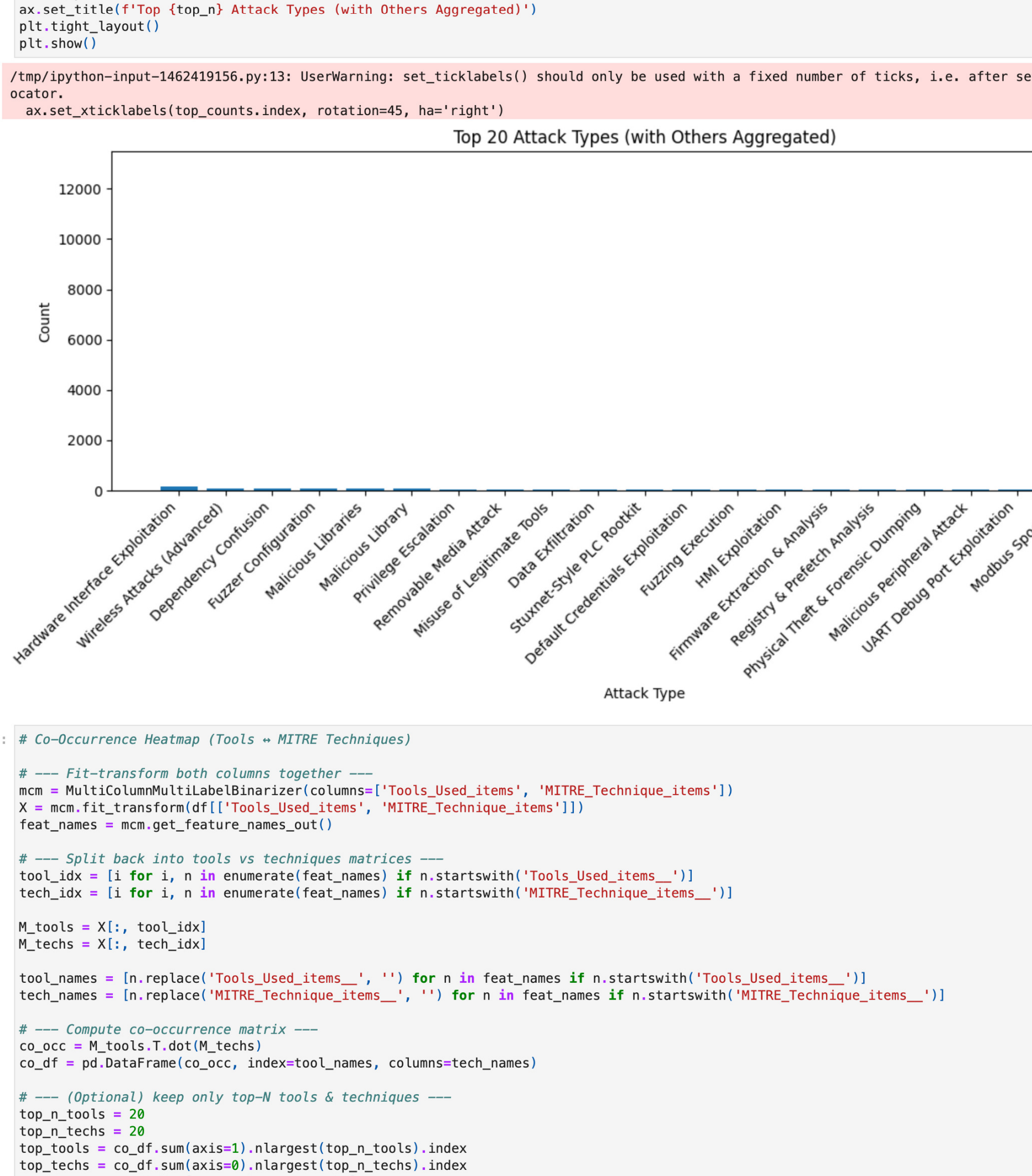
```
/tmp/ipython-input-1462419156.py:13: UserWarning: set_ticklabels() should only be used with a fixed number of ticks, i.e. after set_ticks() or using a FixedL
ocator.
  ax.set_xticklabels(top_counts.index, rotation=45, ha='right')
```



```python
# Co-Occurrence Heatmap (Tools ↔ MITRE Techniques)

# --- Fit-transform both columns together ---
mcm = MultiColumnMultiLabelBinarizer(columns=['Tools_Used_items', 'MITRE_Technique_items'])
X = mcm.fit_transform(df[['Tools_Used_items', 'MITRE_Technique_items']])
feat_names = mcm.get_feature_names_out()

# --- Split back into tools vs techniques matrices ---
tool_idx = [i for i, n in enumerate(feat_names) if n.startswith('Tools_Used_items__')]
tech_idx = [i for i, n in enumerate(feat_names) if n.startswith('MITRE_Technique_items__')]

M_tools = X[:, tool_idx]
M_techs = X[:, tech_idx]

tool_names = [n.replace('Tools_Used_items__', '') for n in feat_names if n.startswith('Tools_Used_items__')]
tech_names = [n.replace('MITRE_Technique_items__', '') for n in feat_names if n.startswith('MITRE_Technique_items__')]

# --- Compute co-occurrence matrix ---
co_occ = M_tools.T.dot(M_techs)
co_df = pd.DataFrame(co_occ, index=tool_names, columns=tech_names)

# --- (Optional) keep only top-N tools & techniques ---
top_n_tools = 20
top_n_techs = 20
top_tools = co_df.sum(axis=1).nlargest(top_n_tools).index
top_techs = co_df.sum(axis=0).nlargest(top_n_techs).index
co_df_sub = co_df.loc[top_tools, top_techs]

# --- Plot heatmap ---
fig, ax = plt.subplots(figsize=(12, 10))
cax = ax.imshow(co_df_sub.values, aspect='auto')
ax.set_xticks(range(len(co_df_sub.columns)))
ax.set_xticklabels(co_df_sub.columns, rotation=90, ha='right')
ax.set_yticks(range(len(co_df_sub.index)))
ax.set_yticklabels(co_df_sub.index)
ax.set_xlabel('MITRE Technique')
ax.set_ylabel('Tool')
ax.set_title('Co-Occurrence Heatmap: Tools vs. MITRE Techniques')
fig.colorbar(cax, ax=ax)
plt.tight_layout()
plt.show()
```



```python
# 3) Apply PCA to reduce to 2 components
pca = PCA(n_components=2)
X_pca = pca.fit_transform(X)

# 4) Encode Category labels for coloring
y = LabelEncoder().fit_transform(df['Category'])

# 5) Scatter plot of the first two principal components
plt.figure(figsize=(10, 8))
plt.scatter(X_pca[:, 0], X_pca[:, 1], c=y, alpha=0.5)
plt.xlabel('Principal Component 1')
plt.ylabel('Principal Component 2')
plt.title('2D PCA Projection of Incident Features')
plt.tight_layout()
plt.show()
```



Models

3. Train and Compare Multiple Models (25 pts)

- Must train at least three distinct models (e.g., logistic regression vs. random forest)
- Must evaluate with:
  - train_test_split
  - Appropriate metrics:
    - Regression: MSE, RMSE, R²
    - Classification: accuracy, precision, recall, f1, confusion_matrix
  - Use cross_val_score to improve evaluation robustness
  - Discuss:
    - Which model performed best and why?
    - Tradeoffs between accuracy and interpretability

Bernoulli (Naive Bayes)

https://scikit-learn.org/stable/modules/naive_bayes.html

```python
# # Full BernoulliNB pipeline (with preprocessing in the pipeline for cross-validation)
# nb_clf = OneVsRestClassifier(BernoulliNB())

# # Train/test split
# mlb = MultiLabelBinarizer()
# y_nb = mlb.fit_transform(df[target])

# X_nb = df[solo_feats + list_feats]
# X_train_nb, X_test_nb, y_train_nb, y_test_nb = train_test_split(
#     X_nb, y_nb, train_size=0.5, random_state=42
# )

# # Create a pipeline with preprocessing and the classifier
# nb_pipeline = Pipeline([
#     ('preprocess', preprocess),
#     ('clf', nb_clf)
# ])

# # Fit & predict
# nb_pipeline.fit(X_train_nb, y_train_nb)
# y_pred_nb = nb_pipeline.predict(X_test_nb)

# # Results
# print("Classification Report:")
# print(classification_report(y_test_nb, y_pred_nb, target_names=mlb.classes_))

# # Cross-validation
# # Use the pipeline directly with cross_val_score
# scores = cross_val_score(nb_pipeline, X_nb, y_nb, cv=3, scoring='accuracy', n_jobs=1)
# print(f"3-Fold CV Accuracy: {scores.mean():.3f} ± {scores.std():.3f}")
```

```
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 0 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 1 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 3 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 9 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 10 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 12 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 13 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 14 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 16 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 18 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 20 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 21 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 22 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 24 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 26 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 28 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 33 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 35 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 37 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 38 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 41 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 42 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 43 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 46 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 49 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 50 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 51 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 53 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 60 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 62 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 65 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 66 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 69 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 70 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 71 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 73 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 74 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 75 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 84 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 87 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 88 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 91 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 95 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 96 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 98 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 100 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 101 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 102 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 103 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 105 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 108 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 110 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 111 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 113 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 115 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 116 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 117 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 118 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 119 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 121 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 127 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 133 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 136 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 141 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 142 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 143 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 146 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 147 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 154 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 161 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 164 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 169 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 171 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 172 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 174 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 183 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 186 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 190 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 191 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 192 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 197 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 201 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 202 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 207 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 211 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 214 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 219 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 220 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 222 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 225 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 226 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 228 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 232 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 235 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 236 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 237 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 242 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 248 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 261 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 265 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 267 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 271 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 278 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 279 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 280 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 281 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 283 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 288 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 290 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 297 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 299 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 300 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 303 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 305 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 306 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 307 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 308 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 311 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 314 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 319 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 324 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 325 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 329 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 336 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 343 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 344 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 345 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 348 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 354 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 355 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 360 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 361 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 364 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 365 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 376 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 377 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 381 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 384 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 388 is present in all training examples.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/multiclass.py:90: UserWarning: Label not 391 is present in all training examples.
```