In [ ]:
```
'''Intrusion Detection Systems (IDSs)
PROBLEM STATEMENT

Scikitlearn

READ ME

=======

Problem Statement: Intrusion Detection Systems (IDSs) and Intrusion Prevention

Network attacks : There are different types of network attacks such as DoS Hul

Features : Total 80 features are taken from network flow data, including Label

ID : record id

Label : BENIGN(Normal), DoS(Attack)

Packet info : 'Total Fwd Packets','Total Backward Packets', 'Total Length of Fu

      'Total Length of Bwd Packets', 'Fwd Packet Length Max',

      'Fwd Packet Length Min', 'Fwd Packet Length Mean',

      'Fwd Packet Length Std', 'Bwd Packet Length Max',

      'Bwd Packet Length Min', 'Bwd Packet Length Mean',

      'Bwd Packet Length Std'

Payload info : 'Flow Bytes/s', 'Flow Packets/s',

      'Flow IAT Mean', 'Flow IAT Std', 'Flow IAT Max', 'Flow IAT Min',

      'Fwd IAT Total', 'Fwd IAT Mean', 'Fwd IAT Std', 'Fwd IAT Max',

      'Fwd IAT Min', 'Bwd IAT Total', 'Bwd IAT Mean', 'Bwd IAT Std',

      'Bwd IAT Max', 'Bwd IAT Min', 'Fwd PSH Flags', 'Bwd PSH Flags',

      'Fwd URG Flags', 'Bwd URG Flags', 'Fwd Header Length',

      'Bwd Header Length', 'Fwd Packets/s', 'Bwd Packets/s',

      'Min Packet Length', 'Max Packet Length', 'Packet Length Mean',

      'Packet Length Std', 'Packet Length Variance', 'FIN Flag Count',

      'SYN Flag Count', 'RST Flag Count', 'PSH Flag Count', 'ACK Flag Count',

      'URG Flag Count', 'CWE Flag Count', 'ECE Flag Count', 'Down/Up Ratio',

      'Average Packet Size', 'Avg Fwd Segment Size', 'Avg Bwd Segment Size',

      'Fwd Header Length.1', 'Fwd Avg Bytes/Bulk', 'Fwd Avg Packets/Bulk',
```

```
      'Fwd Avg Bulk Rate', 'Bwd Avg Bytes/Bulk', 'Bwd Avg Packets/Bulk',

      'Bwd Avg Bulk Rate', 'Subflow Fwd Packets', 'Subflow Fwd Bytes',

      'Subflow Bwd Packets', 'Subflow Bwd Bytes', 'Init_Win_bytes_forward',

      'Init_Win_bytes_backward', 'act_data_pkt_fwd', 'min_seg_size_forward'

Session info : 'Active Mean', 'Active Std', 'Active Max', 'Active Min', 'Idle M

      'Idle Std', 'Idle Max', 'Idle Min'

What to Do?

===========

Build a Machine Learning model which can detect the network attacks.

Submission Details

==================

A zip file containing your code (.ipynb file or .py file or r notebook), with t
'''
```

In [1]:
```
import numpy as np
arr = np.array([1, 2, 3, 4, 5])


print("Original array:", arr)
print("Sum of array:", np.sum(arr))
print("Mean of array:", np.mean(arr))
print("Standard deviation of array:", np.std(arr))
```

```
Original array: [1 2 3 4 5]
Sum of array: 15
Mean of array: 3.0
Standard deviation of array: 1.4142135623730951
```

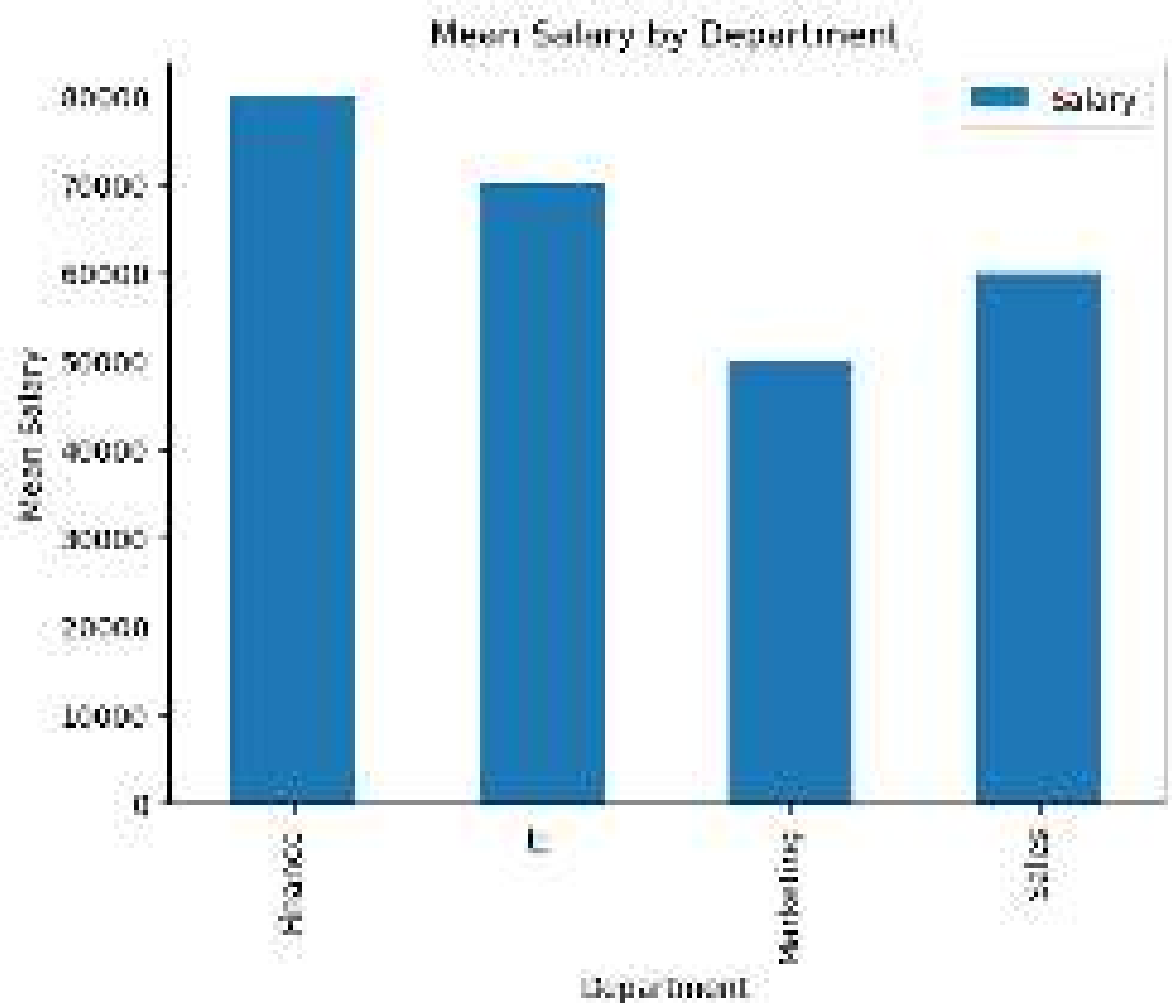
In [2]:
```
import numpy as np
arr2d = np.array([[1, 2, 3], [4, 5, 6], [7, 8, 9]])


print("Original 2D array:", arr2d)
print("Transposed 2D array:", np.transpose(arr2d))
print("Sum of rows in 2D array:", np.sum(arr2d, axis=1))
print("Product of columns in 2D array:", np.prod(arr2d, axis=0))
```

```
Original 2D array: [[1 2 3]
 [4 5 6]
 [7 8 9]]
Transposed 2D array: [[1 4 7]
 [2 5 8]
 [3 6 9]]
Sum of rows in 2D array: [ 6 15 24]
Product of columns in 2D array: [ 28  80 162]
```

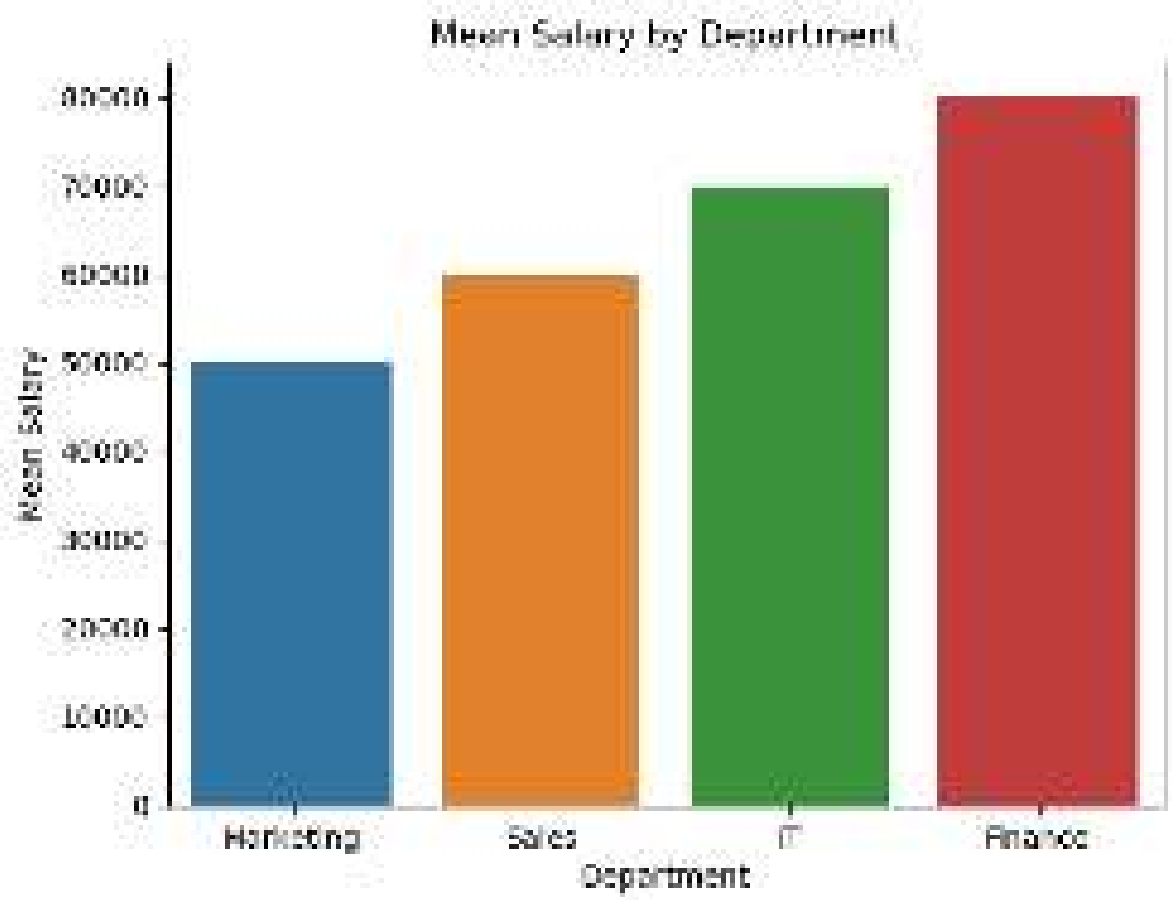
In [83]:
```
import pandas as pd
data = {'Name': ['archana', 'puja', 'abhishek', 'raj'],
        'Age': [26, 25, 21, 40],
        'Salary': [50000, 60000, 70000, 80000]}
df = pd.DataFrame(data)


print(df)
print("=========================================")

print(df['Name'])
print("=========================================")
print(df.loc[1])
print("=========================================")


df['Department'] = ['Marketing', 'Sales', 'IT', 'Finance']
print(df)



grouped_df = df.groupby('Department').mean()
print(grouped_df)
```

```
       Name  Age  Salary
0   archana   26   50000
1      puja   25   60000
2  abhishek   21   70000
3       raj   40   80000
=========================================
0     archana
1        puja
2    abhishek
3         raj
Name: Name, dtype: object
=========================================
Name       puja
Age          25
Salary    60000
Name: 1, dtype: object
=========================================
       Name  Age  Salary Department
0   archana   26   50000  Marketing
1      puja   25   60000      Sales
2  abhishek   21   70000         IT
3       raj   40   80000    Finance
             Age   Salary
Department
Finance     40.0  80000.0
IT          21.0  70000.0
Marketing   26.0  50000.0
Sales       25.0  60000.0
```

In [4]:
```
import matplotlib.pyplot as plt


grouped_df.plot(kind='bar', y='Salary')
plt.title('Mean Salary by Department')
plt.xlabel('Department')
plt.ylabel('Mean Salary')
plt.show()
```



In [5]:
```
import seaborn as sns


sns.barplot(x='Department', y='Salary', data=df)
plt.title('Mean Salary by Department')
plt.xlabel('Department')
plt.ylabel('Mean Salary')
plt.show()
```



In [6]:
```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, confusion_matrix


data = {'Name': ['archana', 'puja', 'Chandana', 'abhishek','rajeshwari','ramesl
        'Age': [25, 30, 35, 40,45,23,26,29],
        'Salary': [90000, 60000, 70000, 90000,50000,60000,20000,40000],
        'Gender': ['F', 'F', 'F', 'F','F','M','M','M']}
df = pd.DataFrame(data)


print(df)


X = df[['Age', 'Salary']]
y = df['Gender']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, rando

dt_model = DecisionTreeClassifier()
dt_model.fit(X_train, y_train)


y_pred_dt = dt_model.predict(X_test)
print("==========================")
y_test




print('Decision Tree Model')
print(classification_report(y_test, y_pred_dt))
print(confusion_matrix(y_test, y_pred_dt))
```

```
      Name  Age  Salary Gender
0    Anitha   25   50000      F
1   Bhavani   30   60000      F
2  Chandana   35   70000      F
3   Deebiga   40   80000      F
4    stella   45   50000      F
5      Balu   23   60000      M
6      Raju   26   20000      M
7      ravi   29   40000      M
==========================
Decision Tree Model
              precision    recall  f1-score   support

           F       1.00      0.50      0.67         2
           M       0.50      1.00      0.67         1

    accuracy                           0.67         3
   macro avg       0.75      0.75      0.67         3
weighted avg       0.83      0.67      0.67         3

[[1 1]
 [0 1]]
```

In [7]:
```
print('Decision Tree Model')
print(classification_report(y_test, y_pred_dt))
print(confusion_matrix(y_test, y_pred_dt))
```

```
Decision Tree Model
              precision    recall  f1-score   support

           F       1.00      0.50      0.67         2
           M       0.50      1.00      0.67         1

    accuracy                           0.67         3
   macro avg       0.75      0.75      0.67         3
weighted avg       0.83      0.67      0.67         3

[[1 1]
 [0 1]]
```

In [8]:
```
rf_model = RandomForestClassifier()
rf_model.fit(X_train, y_train)


y_pred_rf = rf_model.predict(X_test)


print('Random Forest Model')
print(classification_report(y_test, y_pred_rf))
print(confusion_matrix(y_test, y_pred_rf))
```

```
Random Forest Model
              precision    recall  f1-score   support

           F       0.50      0.50      0.50         2
           M       0.00      0.00      0.00         1

    accuracy                           0.33         3
   macro avg       0.25      0.25      0.25         3
weighted avg       0.33      0.33      0.33         3

[[1 1]
 [1 0]]
```

In [9]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn import metrics
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report,confusion_matrix


import warnings
warnings.filterwarnings("ignore")
```

In [13]:
```
df=pd.read_csv("C:/Users/HP/Desktop/pythonproject10thchapter/sklearn test/IDS_
df
```

Out[13]:

	ID	Destination Port	Flow Duration	Total Fwd Packets	Total Backward Packets	Total Length of Fwd Packets	Total Length of Bwd Packets	Fwd Packet Length Max	Fwd Packet Length Min
0	0	22	1420155	37	46	2634	7062	408	0
1	1	80	63122325	7	0	0	0	0	0
2	2	80	85039076	6	6	347	11595	347	0
3	3	443	5379977	5	1	135	46	46	6
4	4	80	84220258	8	5	326	11595	320	0
...	...	...	...	...	...	...	...	...	...
129874	129874	80	101268629	10	6	1074	11595	358	0
129875	129875	53	118727713	4	4	182	493	51	40
129876	129876	443	31246630	3	1	41	41	41	0
129877	129877	80	5183052	3	1	12	0	6	0
129878	129878	80	5280210	4	2	0	0	0	0

129879 rows × 80 columns

In [14]:
```
df.shape
```

Out[14]:
```
(129879, 80)
```

In [15]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 129879 entries, 0 to 129878
Data columns (total 80 columns):
 #   Column                        Non-Null Count   Dtype  
---  ------                        --------------   -----  
 0   ID                            129879 non-null  int64  
 1   Destination Port              129879 non-null  int64  
 2   Flow Duration                 129879 non-null  int64  
 3   Total Fwd Packets             129879 non-null  int64  
 4   Total Backward Packets        129879 non-null  int64  
 5   Total Length of Fwd Packets   129879 non-null  int64  
 6   Total Length of Bwd Packets   129879 non-null  int64  
 7   Fwd Packet Length Max         129879 non-null  int64  
 8   Fwd Packet Length Min         129879 non-null  int64  
 9   Fwd Packet Length Mean        129879 non-null  float64
 10  Fwd Packet Length Std         129879 non-null  float64
 11  Bwd Packet Length Max         129879 non-null  int64  
 12  Bwd Packet Length Min         129879 non-null  int64  
 13  Bwd Packet Length Mean        129879 non-null  float64
 14  Bwd Packet Length Std         129879 non-null  float64
 15  Flow Bytes/s                  129679 non-null  float64
 16  Flow Packets/s                129879 non-null  float64
 17  Flow IAT Mean                 129879 non-null  float64
 18  Flow IAT Std                  129879 non-null  float64
 19  Flow IAT Max                  129879 non-null  int64  
 20  Flow IAT Min                  129879 non-null  int64  
 21  Fwd IAT Total                 129879 non-null  int64  
 22  Fwd IAT Mean                  129879 non-null  float64
 23  Fwd IAT Std                   129879 non-null  float64
 24  Fwd IAT Max                   129879 non-null  int64  
 25  Fwd IAT Min                   129879 non-null  int64  
 26  Bwd IAT Total                 129879 non-null  int64  
 27  Bwd IAT Mean                  129879 non-null  float64
 28  Bwd IAT Std                   129879 non-null  float64
 29  Bwd IAT Max                   129879 non-null  int64  
 30  Bwd IAT Min                   129879 non-null  int64  
 31  Fwd PSH Flags                 129879 non-null  int64  
 32  Bwd PSH Flags                 129879 non-null  int64  
 33  Fwd URG Flags                 129879 non-null  int64  
 34  Bwd URG Flags                 129879 non-null  int64  
 35  Fwd Header Length             129879 non-null  int64  
 36  Bwd Header Length             129879 non-null  int64  
 37  Fwd Packets/s                 129879 non-null  float64
 38  Bwd Packets/s                 129879 non-null  float64
 39  Min Packet Length             129879 non-null  int64  
 40  Max Packet Length             129879 non-null  int64  
 41  Packet Length Mean            129879 non-null  float64
 42  Packet Length Std             129879 non-null  float64
 43  Packet Length Variance        129879 non-null  float64
 44  FIN Flag Count                129879 non-null  int64  
 45  SYN Flag Count                129879 non-null  int64  
 46  RST Flag Count                129879 non-null  int64  
 47  PSH Flag Count                129879 non-null  int64  
 48  ACK Flag Count                129879 non-null  int64  
 49  URG Flag Count                129879 non-null  int64  
 50  CWE Flag Count                129879 non-null  int64  
 51  ECE Flag Count                129879 non-null  int64  
```

```
 52  Down/Up Ratio                129879 non-null  int64
 53  Average Packet Size          129879 non-null  float64
 54  Avg Fwd Segment Size         129879 non-null  float64
 55  Avg Bwd Segment Size         129879 non-null  float64
 56  Fwd Header Length.1          129879 non-null  int64
 57  Fwd Avg Bytes/Bulk           129879 non-null  int64
 58  Fwd Avg Packets/Bulk         129879 non-null  int64
 59  Fwd Avg Bulk Rate            129879 non-null  int64
 60  Bwd Avg Bytes/Bulk           129879 non-null  int64
 61  Bwd Avg Packets/Bulk         129879 non-null  int64
 62  Bwd Avg Bulk Rate            129879 non-null  int64
 63  Subflow Fwd Packets          129879 non-null  int64
 64  Subflow Fwd Bytes            129879 non-null  int64
 65  Subflow Bwd Packets          129879 non-null  int64
 66  Subflow Bwd Bytes            129879 non-null  int64
 67  Init_Win_bytes_forward       129879 non-null  int64
 68  Init_Win_bytes_backward      129879 non-null  int64
 69  act_data_pkt_fwd             129879 non-null  int64
 70  min_seg_size_forward         129879 non-null  int64
 71  Active Mean                  129879 non-null  float64
 72  Active Std                   129879 non-null  float64
 73  Active Max                   129879 non-null  int64
 74  Active Min                   129879 non-null  int64
 75  Idle Mean                    129879 non-null  float64
 76  Idle Std                     129879 non-null  float64
 77  Idle Max                     129879 non-null  int64
 78  Idle Min                     129879 non-null  int64
 79  Label                        129879 non-null  object
dtypes: float64(24), int64(55), object(1)
memory usage: 79.3+ MB
```

In [16]:
```
print(df.head())
```

```
   ID  Destination Port  Flow Duration  Total Fwd Packets  \
0   0                22        1420155                 37
1   1                80       63122325                  7
2   2                80       85039076                  6
3   3               443        5379977                  5
4   4                80       84220258                  8

   Total Backward Packets  Total Length of Fwd Packets  \
0                      46                         2634
1                       0                            0
2                       6                          347
3                       1                          135
4                       5                          326

   Total Length of Bwd Packets  Fwd Packet Length Max  Fwd Packet Length Min
\
0                         7062                    408                      0
1                            0                      0                      0
2                        11595                    347                      0
3                           46                     46                      6
4                        11595                    320                      0

   Fwd Packet Length Mean  ...  min_seg_size_forward  Active Mean  Active Std
\
0               71.189189  ...                    20          0.0         0.0
1                0.000000  ...                    40    7010762.0         0.0
2               57.833333  ...                    32          4.0         0.0
3               27.000000  ...                    20     109172.0         0.0
4               40.750000  ...                    20       1976.0         0.0

   Active Max  Active Min   Idle Mean    Idle Std  Idle Max  Idle Min  \
0           0           0         0.0         0.0         0         0
1     7010762     7010762  18700000.0  12200000.0  32100000   8015920
2           4           4  84900000.0         0.0  84900000  84900000
3      109172      109172   5266022.0         0.0   5266022   5266022
4        1976        1976  84100000.0         0.0  84100000  84100000

             Label
0           BENIGN
1  DoS Slowhttptest
2         DoS Hulk
3           BENIGN
4         DoS Hulk

[5 rows x 80 columns]
```

In [17]:
```
print(df.isnull().sum())
```

```
ID                          0
Destination Port            0
Flow Duration               0
Total Fwd Packets           0
Total Backward Packets      0
                           ..
Idle Mean                   0
Idle Std                    0
Idle Max                    0
Idle Min                    0
Label                       0
Length: 80, dtype: int64
```

In [18]:
```
print(df.describe())
```

	ID	Destination Port	Flow Duration	Total Fwd Packets	\
count	129879.000000	129879.000000	1.298790e+05	129879.000000	
mean	64939.000000	5714.908638	2.808619e+07	8.522710	
std	37492.982143	15752.583680	4.282575e+07	592.895904	
min	0.000000	0.000000	-1.000000e+00	1.000000	
25%	32469.500000	53.000000	2.000000e+02	2.000000	
50%	64939.000000	80.000000	6.131600e+04	2.000000	
75%	97408.500000	443.000000	8.313865e+07	7.000000	
max	129878.000000	65427.000000	1.200000e+08	180892.000000	

	Total Backward Packets	Total Length of Fwd Packets	\
count	129879.000000	1.298790e+05	
mean	8.925400	5.393826e+02	
std	777.866547	5.086162e+03	
min	0.000000	0.000000e+00	
25%	1.000000	1.200000e+01	
50%	2.000000	8.200000e+01	
75%	6.000000	3.650000e+02	
max	230419.000000	1.080532e+06	

	Total Length of Bwd Packets	Fwd Packet Length Max	\
count	1.298790e+05	129879.000000	
mean	1.417065e+04	233.075054	
std	1.815935e+06	596.636120	
min	0.000000e+00	0.000000	
25%	0.000000e+00	6.000000	
50%	1.860000e+02	46.000000	
75%	1.159500e+04	341.000000	
max	5.160000e+08	23360.000000	

	Fwd Packet Length Min	Fwd Packet Length Mean	...	act_data_pkt_fwd	\
count	129879.000000	129879.000000	...	129879.000000	
mean	15.097190	60.403257	...	5.149470	
std	52.801167	157.163962	...	569.744382	
min	0.000000	0.000000	...	0.000000	
25%	0.000000	6.000000	...	0.000000	
50%	0.000000	41.000000	...	1.000000	
75%	32.000000	56.500000	...	2.000000	
max	1983.000000	3491.950000	...	173720.000000	

	min_seg_size_forward	Active Mean	Active Std	Active Max	\
count	129879.000000	1.298790e+05	1.298790e+05	1.298790e+05	
mean	26.770787	8.927445e+04	4.542519e+04	1.568155e+05	
std	6.312032	6.732218e+05	4.730384e+05	1.037478e+06	
min	-1.000000	0.000000e+00	0.000000e+00	0.000000e+00	
25%	20.000000	0.000000e+00	0.000000e+00	0.000000e+00	
50%	32.000000	0.000000e+00	0.000000e+00	0.000000e+00	
75%	32.000000	9.920000e+02	0.000000e+00	9.920000e+02	
max	60.000000	5.060000e+07	7.000000e+07	9.900000e+07	

	Active Min	Idle Mean	Idle Std	Idle Max	Idle Min
count	1.298790e+05	1.298790e+05	1.298790e+05	1.298790e+05	1.298790e+05
mean	6.182056e+04	2.218634e+07	4.709707e+05	2.259182e+07	2.181069e+07
std	5.756609e+05	3.818165e+07	4.478192e+06	3.853675e+07	3.813624e+07
min	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
25%	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00

```
50%    0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00
75%    9.890000e+02  1.620000e+07  0.000000e+00  1.880000e+07  1.000000e+07
max    4.960000e+07  1.200000e+08  7.690000e+07  1.200000e+08  1.200000e+08

[8 rows x 79 columns]
```

In [19]:
```
grouped = df.groupby("Label").mean()
print(grouped)
```

Label	ID	Destination Port	Flow Duration
BENIGN	64961.222449	8958.734156	1.217541e+07
DoS GoldenEye	63730.226233	80.000000	2.403559e+07
DoS Hulk	64994.313875	80.000000	5.712520e+07
DoS Slowhttptest	63711.643418	80.000000	5.707170e+07
DoS slowloris	64365.901608	80.000000	5.590001e+07

Label	Total Fwd Packets	Total Backward Packets
BENIGN	10.355789	11.730783
DoS GoldenEye	5.832232	3.645653
DoS Hulk	5.287636	4.201825
DoS Slowhttptest	5.627701	0.977407
DoS slowloris	6.227058	1.621570

Label	Total Length of Fwd Packets	Total Length of Bwd Packets
BENIGN	675.886313	18081.956507
DoS GoldenEye	421.124555	6400.334520
DoS Hulk	280.987422	7769.870557
DoS Slowhttptest	479.400786	121.881139
DoS slowloris	784.065279	13.654683

Label	Fwd Packet Length Max	Fwd Packet Length Min
BENIGN	233.097843	22.124691
DoS GoldenEye	309.796645	0.000000
DoS Hulk	232.906701	0.205810
DoS Slowhttptest	237.328094	121.335953
DoS slowloris	91.342479	4.408704

Label	Fwd Packet Length Mean	...	act_data_pkt_fwd
BENIGN	67.648009	...	7.339872
DoS GoldenEye	60.284243	...	0.996441
DoS Hulk	44.301794	...	1.312424
DoS Slowhttptest	160.572536	...	1.000982
DoS slowloris	60.440921	...	3.639546

Label	min_seg_size_forward	Active Mean	Active Std
BENIGN	26.056085	7.192734e+04	4.861326e+04
DoS GoldenEye	32.000000	2.057630e+05	3.038283e+04
DoS Hulk	27.489322	2.983491e+03	1.298803e+02
DoS Slowhttptest	36.031434	3.794831e+06	1.902025e+05
DoS slowloris	34.346263	1.200252e+06	1.545561e+06

Label	Active Max	Active Min	Idle Mean	Idle Std
BENIGN	1.621632e+05	4.524307e+04	4.196948e+06	1.601970e+05
DoS GoldenEye	2.272469e+05	1.842791e+05	1.975343e+07	5.601941e+04
DoS Hulk	3.075330e+03	2.891652e+03	5.615286e+07	8.026769e+05
DoS Slowhttptest	3.925266e+06	3.633006e+06	2.942714e+07	6.623034e+06
DoS slowloris	2.293128e+06	1.073749e+05	2.766494e+07	5.930608e+06

Label	Idle Max	Idle Min

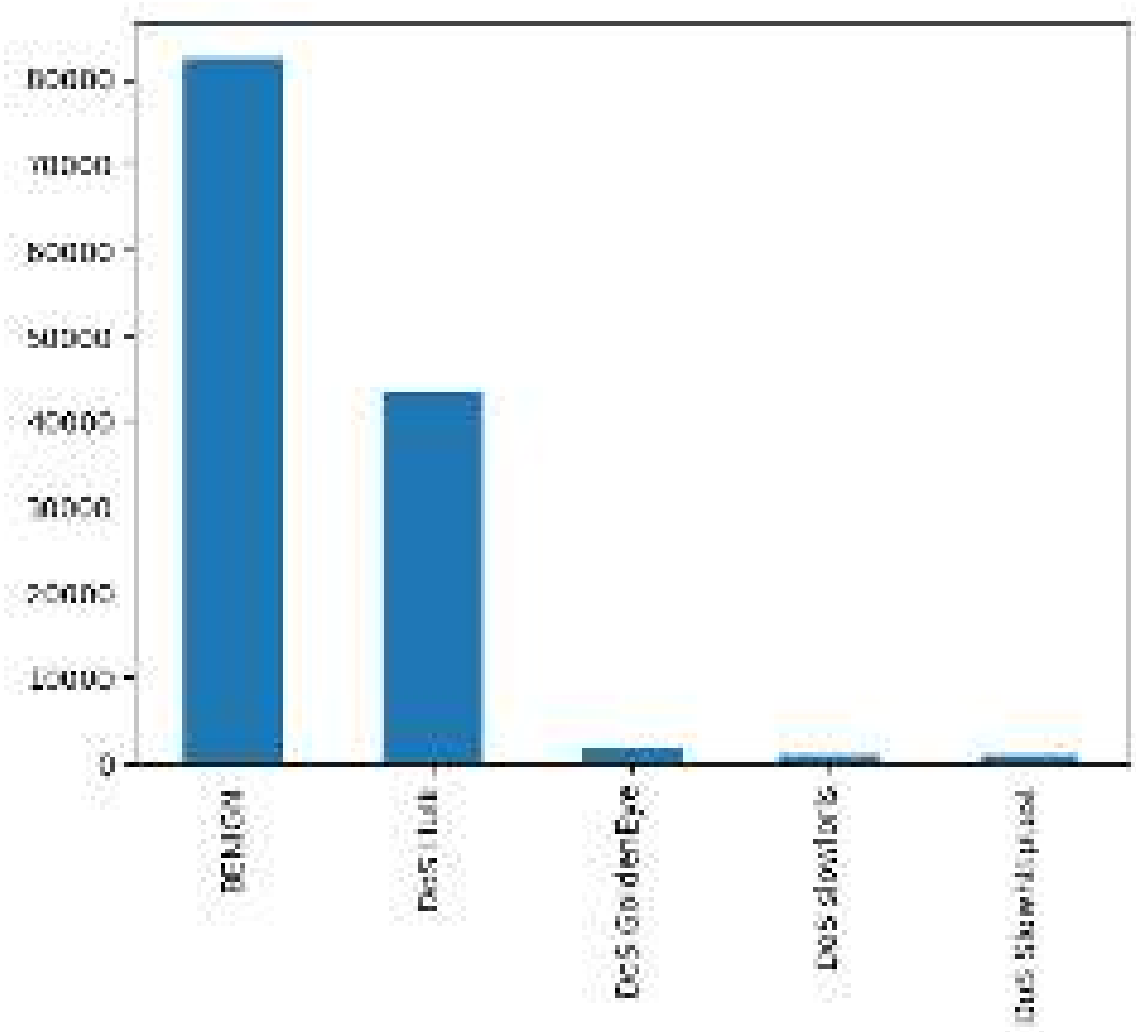
```
Label
BENIGN           4.319841e+06  4.030525e+06
DoS GoldenEye    1.982468e+07  1.970514e+07
DoS Hulk         5.672041e+07  5.558479e+07
DoS Slowhttptest 3.701025e+07  2.368843e+07
DoS slowloris    3.716040e+07  2.343157e+07

[5 rows x 79 columns]
```

In [20]:
```
counts = df["Label"].value_counts()
print(counts)
```

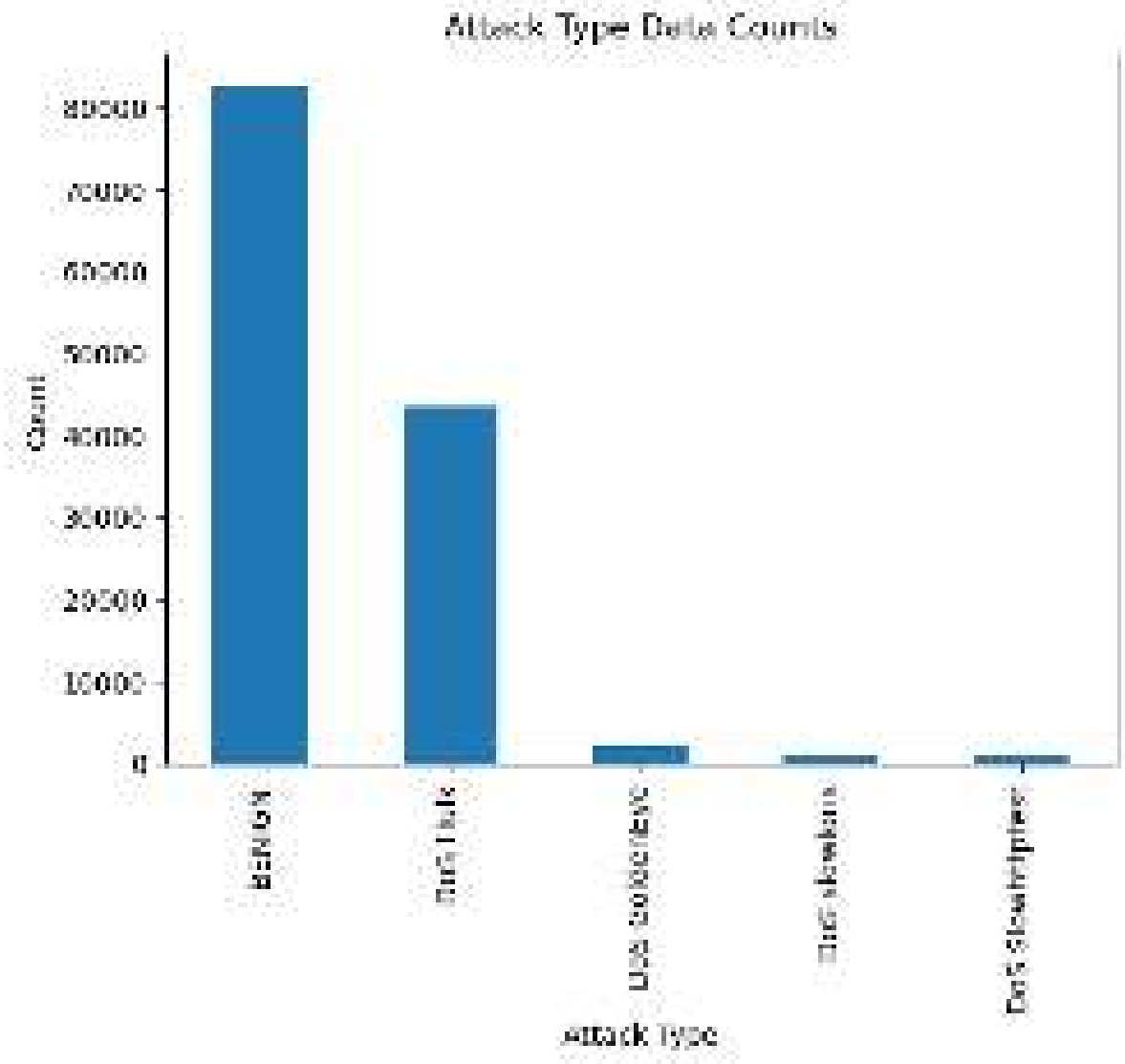
```
BENIGN              82428
DoS Hulk            43409
DoS GoldenEye        1967
DoS slowloris        1057
DoS Slowhttptest     1018
Name: Label, dtype: int64
```

In [21]:
```
counts.plot(kind="bar")
```

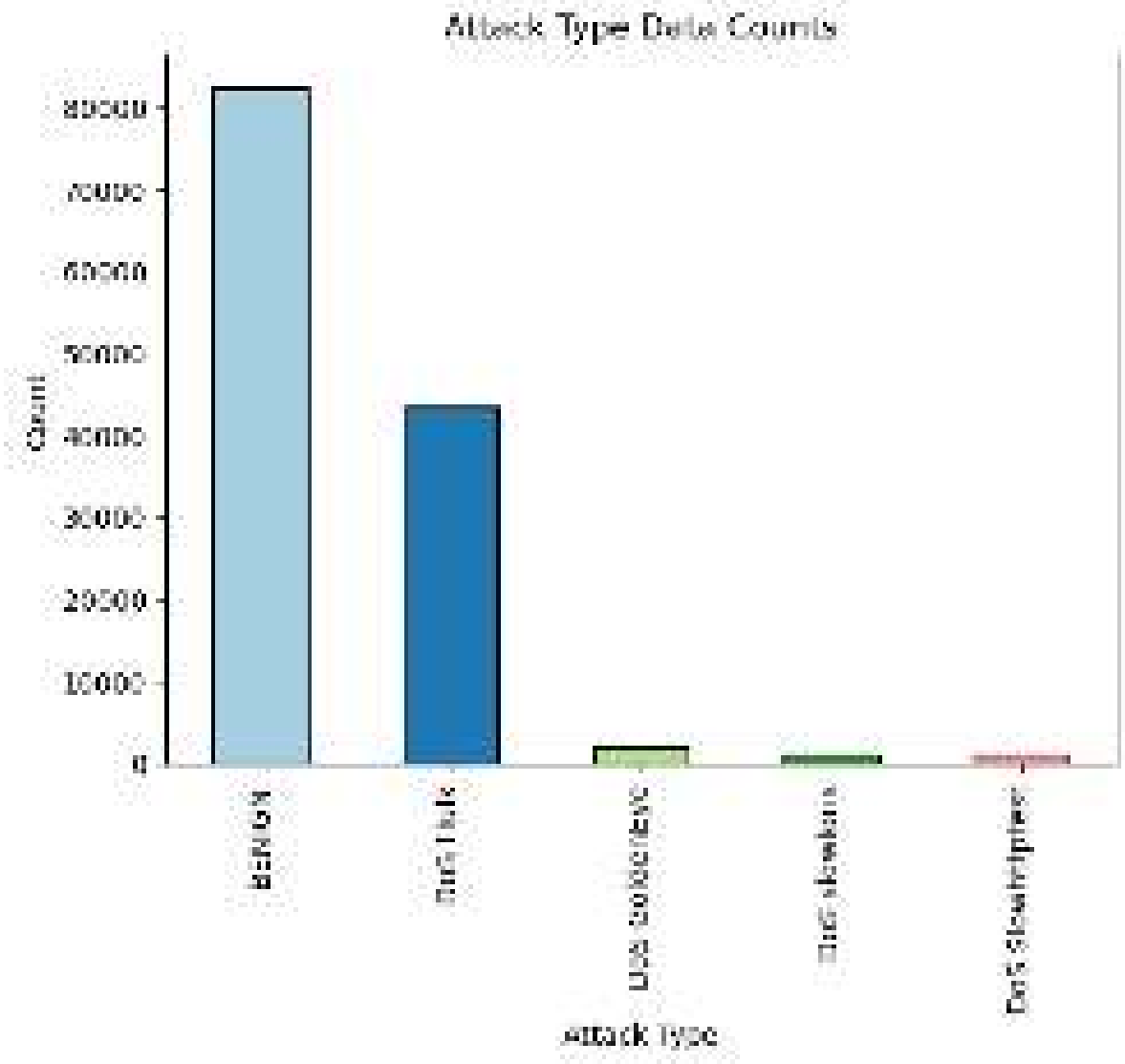
Out[21]: <AxesSubplot:>



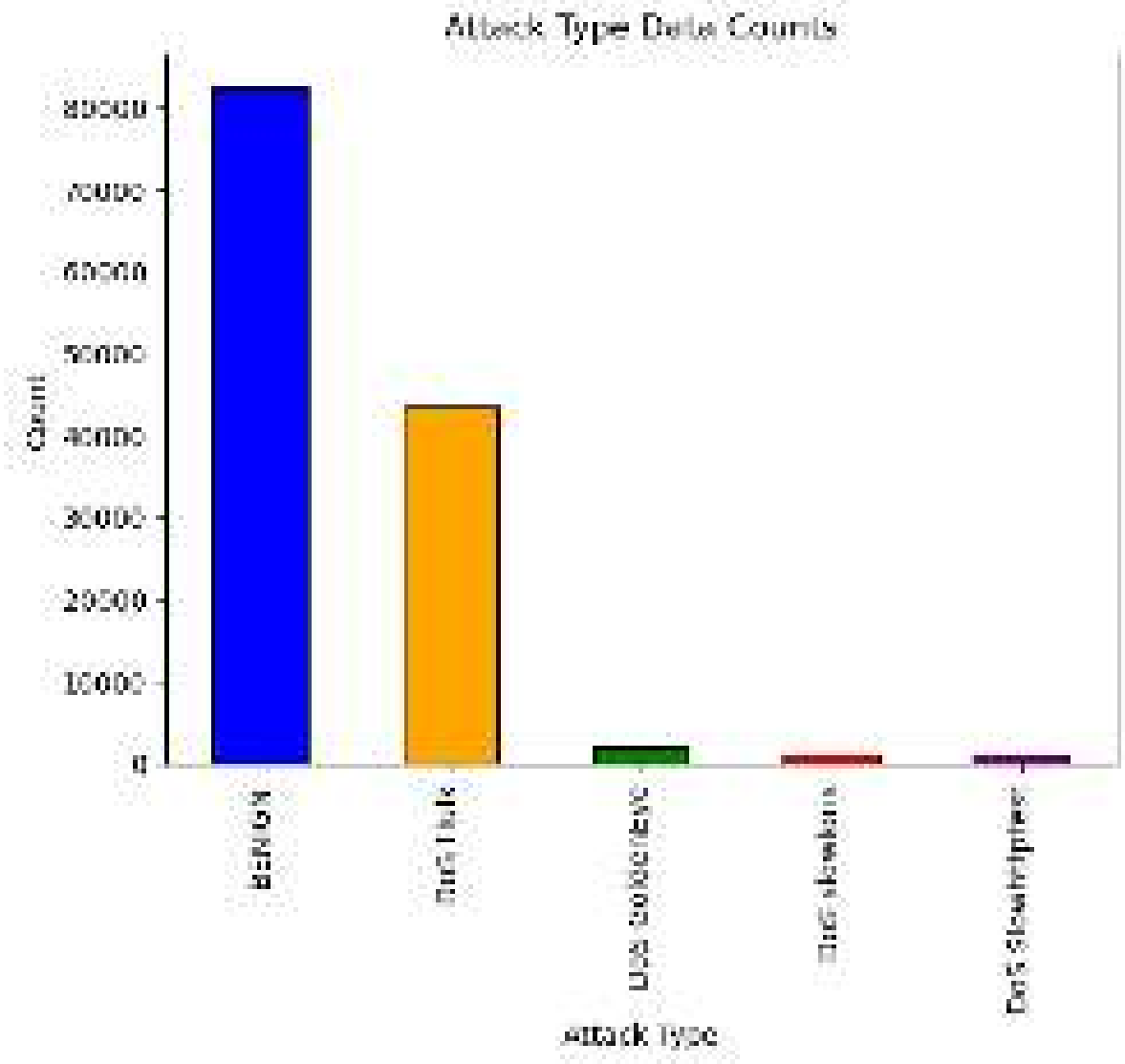
In [22]:
```
ax = counts.plot(kind='bar')


ax.set_xlabel('Attack Type')
ax.set_ylabel('Count')
ax.set_title('Attack Type Data Counts')


plt.show()
```



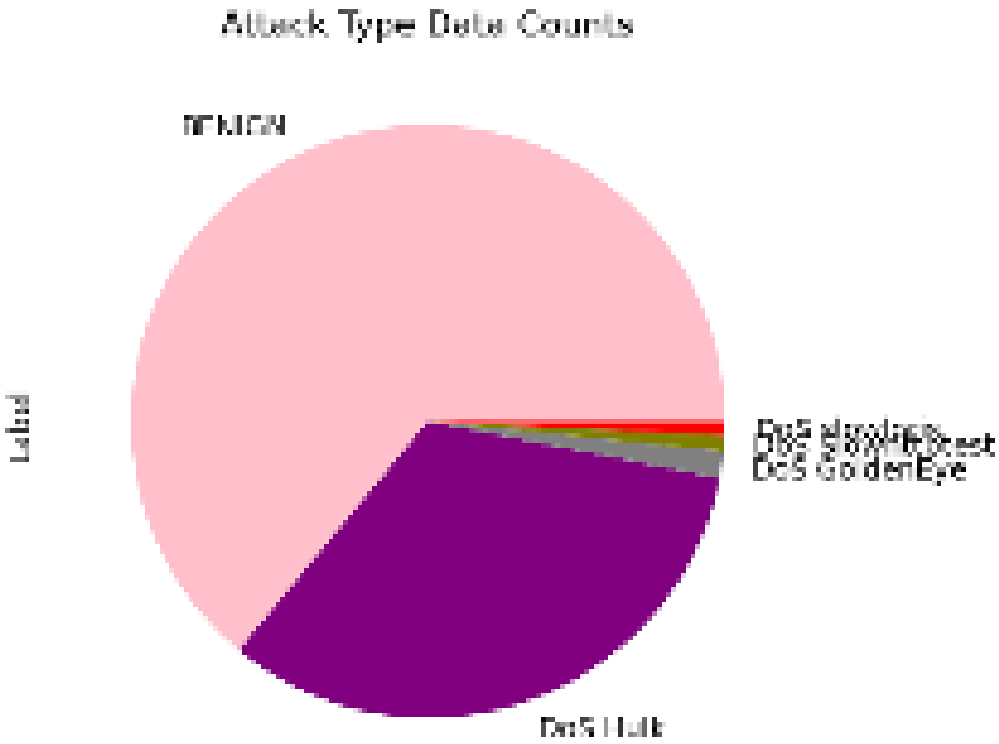
In [23]:
```
counts.plot(kind="bar", edgecolor='black', color=plt.cm.Paired(range(len(count

plt.xlabel("Attack Type")
plt.ylabel("Count")
plt.title("Attack Type Data Counts")


plt.show()
```



In [24]:
```
colors = ['blue', 'orange', 'green', 'red', 'purple', 'brown', 'pink', 'gray',
counts.plot(kind="bar", edgecolor='black',color=colors)

plt.xlabel("Attack Type")
plt.ylabel("Count")
plt.title("Attack Type Data Counts")


plt.show()
```



In [85]:
```
colors=['pink', 'purple', 'gray', 'olive', 'red']
counts.plot(kind='pie', colors=colors)


plt.title("Attack Type Data Counts")


plt.show()
```



In [26]:
```
df.isnull().sum()
```

Out[26]:
```
ID                          0
Destination Port            0
Flow Duration               0
Total Fwd Packets           0
Total Backward Packets      0
                           ..
Idle Mean                   0
Idle Std                    0
Idle Max                    0
Idle Min                    0
Label                       0
Length: 80, dtype: int64
```

In [27]:
```
columns = df.columns
count = 0
for i in df.isnull().sum():
    print("{} : {}".format(columns[count],i))
    count+=1
```

```
ID : 0
Destination Port : 0
Flow Duration : 0
Total Fwd Packets : 0
Total Backward Packets : 0
Total Length of Fwd Packets : 0
Total Length of Bwd Packets : 0
Fwd Packet Length Max : 0
Fwd Packet Length Min : 0
Fwd Packet Length Mean : 0
Fwd Packet Length Std : 0
Bwd Packet Length Max : 0
Bwd Packet Length Min : 0
Bwd Packet Length Mean : 0
Bwd Packet Length Std : 0
Flow Bytes/s : 200
Flow Packets/s : 0
Flow IAT Mean : 0
Flow IAT Std : 0
Flow IAT Max : 0
Flow IAT Min : 0
Fwd IAT Total : 0
Fwd IAT Mean : 0
Fwd IAT Std : 0
Fwd IAT Max : 0
Fwd IAT Min : 0
Bwd IAT Total : 0
Bwd IAT Mean : 0
Bwd IAT Std : 0
Bwd IAT Max : 0
Bwd IAT Min : 0
Fwd PSH Flags : 0
Bwd PSH Flags : 0
Fwd URG Flags : 0
Bwd URG Flags : 0
Fwd Header Length : 0
Bwd Header Length : 0
Fwd Packets/s : 0
Bwd Packets/s : 0
Min Packet Length : 0
Max Packet Length : 0
Packet Length Mean : 0
Packet Length Std : 0
Packet Length Variance : 0
FIN Flag Count : 0
SYN Flag Count : 0
RST Flag Count : 0
PSH Flag Count : 0
ACK Flag Count : 0
URG Flag Count : 0
CWE Flag Count : 0
ECE Flag Count : 0
Down/Up Ratio : 0
Average Packet Size : 0
Avg Fwd Segment Size : 0
Avg Bwd Segment Size : 0
Fwd Header Length.1 : 0
```

```
Fwd Avg Bytes/Bulk : 0
Fwd Avg Packets/Bulk : 0
Fwd Avg Bulk Rate : 0
Bwd Avg Bytes/Bulk : 0
Bwd Avg Packets/Bulk : 0
Bwd Avg Bulk Rate : 0
Subflow Fwd Packets : 0
Subflow Fwd Bytes : 0
Subflow Bwd Packets : 0
Subflow Bwd Bytes : 0
Init_Win_bytes_forward : 0
Init_Win_bytes_backward : 0
act_data_pkt_fwd : 0
min_seg_size_forward : 0
Active Mean : 0
Active Std : 0
Active Max : 0
Active Min : 0
Idle Mean : 0
Idle Std : 0
Idle Max : 0
Idle Min : 0
Label : 0
```

In [28]:
```
df = df.dropna()
```

In [29]:
```
df.shape
```

Out[29]:
```
(129679, 80)
```

In [31]:
```
def attack_encode(value):
    if value == 'BENIGN':
        return 0;
    elif value == "DoS Slowhttptest":
        return 1;
    elif value == 'DoS Hulk':
        return 2;
    elif value == 'DoS GoldenEye':
        return 3;
    else:
        return 4;
```

In [32]:
```
df['intrusion_code'] = df['Label'].apply(attack_encode)
df.iloc[:10, -2:]
```

Out[32]:

	Label	intrusion_code
0	BENIGN	0
1	DoS Slowhttptest	1
2	DoS Hulk	2
3	BENIGN	0
4	DoS Hulk	2
5	BENIGN	0
6	BENIGN	0
7	DoS Hulk	2
8	BENIGN	0
9	BENIGN	0

In [33]:
```
df
```

Out[33]:

	ID	Destination Port	Flow Duration	Total Fwd Packets	Total Backward Packets	Total Length of Fwd Packets	Total Length of Bwd Packets	Fwd Packet Length Max	Fwd Packet Length Min
0	0	22	1420155	37	46	2634	7062	408	0
1	1	80	63122325	7	0	0	0	0	0
2	2	80	85039076	6	6	347	11595	347	0
3	3	443	5379977	5	1	135	46	46	6
4	4	80	84220258	8	5	326	11595	320	0
...	...	...	...	...	...	...	...	...	...
129874	129874	80	101268629	10	6	1074	11595	358	0
129875	129875	53	118727713	4	4	182	493	51	40
129876	129876	443	31246630	3	1	41	41	41	0
129877	129877	80	5183052	3	1	12	0	6	0
129878	129878	80	5280210	4	2	0	0	0	0

129679 rows × 81 columns

In [34]:
```
list_a = [1, 2, 3]
list_b = ['a', 'b', 'c']
zipped = zip(list_a, list_b)
print(list(zipped))
```

```
[(1, 'a'), (2, 'b'), (3, 'c')]
```

In [35]:
```
numerical_cols = [one for each,one in zip(list(df.dtypes),df.dtypes.index)  if
numerical_cols
```

Out[35]:
```
['Fwd Packet Length Mean',
 'Fwd Packet Length Std',
 'Bwd Packet Length Mean',
 'Bwd Packet Length Std',
 'Flow Bytes/s',
 'Flow Packets/s',
 'Flow IAT Mean',
 'Flow IAT Std',
 'Fwd IAT Mean',
 'Fwd IAT Std',
 'Bwd IAT Mean',
 'Bwd IAT Std',
 'Fwd Packets/s',
 'Bwd Packets/s',
 'Packet Length Mean',
 'Packet Length Std',
 'Packet Length Variance',
 'Average Packet Size',
 'Avg Fwd Segment Size',
 'Avg Bwd Segment Size',
 'Active Mean',
 'Active Std',
 'Idle Mean',
 'Idle Std']
```

In [36]:
```
df = df.drop(numerical_cols,axis = 1)
```

In [37]:
```
X = df.drop(['intrusion_code','Label'],axis = 1)
y = df['intrusion_code']
```

In [38]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, rando
```

In [39]:
```
dtree = DecisionTreeClassifier()
dtree.fit(X_train,y_train)
```

Out[39]:
```
DecisionTreeClassifier()
```

In [40]:
```
predictions = dtree.predict(X_test)
print(classification_report(y_test,predictions))
```

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     24884
           1       0.98      0.99      0.98       300
           2       1.00      1.00      1.00     12811
           3       0.98      0.99      0.98       598
           4       0.99      0.99      0.99       311

    accuracy                           1.00     38904
   macro avg       0.99      0.99      0.99     38904
weighted avg       1.00      1.00      1.00     38904
```

In [41]:
```
cm=confusion_matrix(y_test,predictions)
print(cm)
```

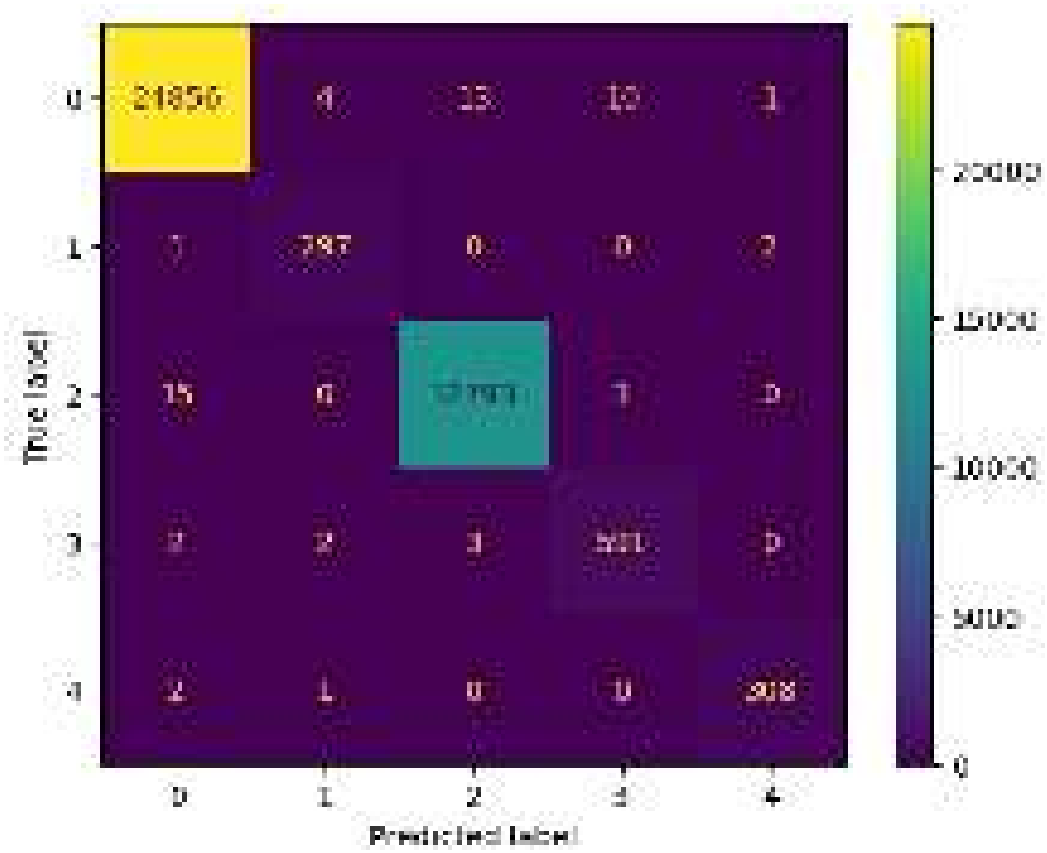
```
[[24856     4    13    10     1]
 [    1   297     0     0     2]
 [   15     0 12793     3     0]
 [    2     2     3   591     0]
 [    2     1     0     0   308]]
```

In [42]:
```
print ("Accuracy of prediction:",round((cm[0,0]+cm[1,1]+cm[2,2]+cm[3,3]+cm[4,4
```

```
Accuracy of prediction: 0.998
```

In [43]:
```
metrics.classification_report(y_test,predictions)
metrics.plot_confusion_matrix(dtree, X_test, y_test)
```

Out[43]:
```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x19cc21c50
d0>
```



In [44]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn import metrics
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report,confusion_matrix


import warnings
warnings.filterwarnings("ignore")
```

In [45]:
```
df1=pd.read_csv("C:/Users/HP/Desktop/pythonproject10thchapter/sklearn test/IDS_
df1
```

Out[45]:

	ID	Destination Port	Flow Duration	Total Fwd Packets	Total Backward Packets	Total Length of Fwd Packets	Total Length of Bwd Packets	Fwd Packet Length Max	Fwd Packet Length Min	
0	0	80	99575743	7	5	401	11595	377	0	5
1	1	80	107429687	2	1	662	6	662	0	33
2	2	80	17291	2	0	12	0	6	6	
3	3	80	85038140	5	7	338	11595	326	0	6
4	4	80	82963302	8	6	375	11595	363	0	4
...	...	...	...	...	...	...	...	...	...	
43289	43289	80	82746195	6	7	324	11595	306	0	5
43290	43290	53	183040	4	2	176	662	44	44	4
43291	43291	443	117811197	12	13	635	4068	198	0	5
43292	43292	80	5023362	3	1	0	0	0	0	
43293	43293	53	272	2	2	56	252	28	28	2

43294 rows × 79 columns

In [46]:
```
df1.shape
```

Out[46]:
```
(43294, 79)
```

In [47]:
```
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 43294 entries, 0 to 43293
Data columns (total 79 columns):
 #   Column                        Non-Null Count  Dtype  
---  ------                        --------------  -----  
 0   ID                            43294 non-null  int64  
 1   Destination Port              43294 non-null  int64  
 2   Flow Duration                 43294 non-null  int64  
 3   Total Fwd Packets             43294 non-null  int64  
 4   Total Backward Packets        43294 non-null  int64  
 5   Total Length of Fwd Packets   43294 non-null  int64  
 6   Total Length of Bwd Packets   43294 non-null  int64  
 7   Fwd Packet Length Max         43294 non-null  int64  
 8   Fwd Packet Length Min         43294 non-null  int64  
 9   Fwd Packet Length Mean        43294 non-null  float64
 10  Fwd Packet Length Std         43294 non-null  float64
 11  Bwd Packet Length Max         43294 non-null  int64  
 12  Bwd Packet Length Min         43294 non-null  int64  
 13  Bwd Packet Length Mean        43294 non-null  float64
 14  Bwd Packet Length Std         43294 non-null  float64
 15  Flow Bytes/s                  43238 non-null  float64
 16  Flow Packets/s                43294 non-null  float64
 17  Flow IAT Mean                 43294 non-null  float64
 18  Flow IAT Std                  43294 non-null  float64
 19  Flow IAT Max                  43294 non-null  int64  
 20  Flow IAT Min                  43294 non-null  int64  
 21  Fwd IAT Total                 43294 non-null  int64  
 22  Fwd IAT Mean                  43294 non-null  float64
 23  Fwd IAT Std                   43294 non-null  float64
 24  Fwd IAT Max                   43294 non-null  int64  
 25  Fwd IAT Min                   43294 non-null  int64  
 26  Bwd IAT Total                 43294 non-null  int64  
 27  Bwd IAT Mean                  43294 non-null  float64
 28  Bwd IAT Std                   43294 non-null  float64
 29  Bwd IAT Max                   43294 non-null  int64  
 30  Bwd IAT Min                   43294 non-null  int64  
 31  Fwd PSH Flags                 43294 non-null  int64  
 32  Bwd PSH Flags                 43294 non-null  int64  
 33  Fwd URG Flags                 43294 non-null  int64  
 34  Bwd URG Flags                 43294 non-null  int64  
 35  Fwd Header Length             43294 non-null  int64  
 36  Bwd Header Length             43294 non-null  int64  
 37  Fwd Packets/s                 43294 non-null  float64
 38  Bwd Packets/s                 43294 non-null  float64
 39  Min Packet Length             43294 non-null  int64  
 40  Max Packet Length             43294 non-null  int64  
 41  Packet Length Mean            43294 non-null  float64
 42  Packet Length Std             43294 non-null  float64
 43  Packet Length Variance        43294 non-null  float64
 44  FIN Flag Count                43294 non-null  int64  
 45  SYN Flag Count                43294 non-null  int64  
 46  RST Flag Count                43294 non-null  int64  
 47  PSH Flag Count                43294 non-null  int64  
 48  ACK Flag Count                43294 non-null  int64  
 49  URG Flag Count                43294 non-null  int64  
 50  CWE Flag Count                43294 non-null  int64  
 51  ECE Flag Count                43294 non-null  int64  
```

```
 52  Down/Up Ratio               43294 non-null  int64
 53  Average Packet Size         43294 non-null  float64
 54  Avg Fwd Segment Size        43294 non-null  float64
 55  Avg Bwd Segment Size        43294 non-null  float64
 56  Fwd Header Length.1         43294 non-null  int64
 57  Fwd Avg Bytes/Bulk          43294 non-null  int64
 58  Fwd Avg Packets/Bulk        43294 non-null  int64
 59  Fwd Avg Bulk Rate           43294 non-null  int64
 60  Bwd Avg Bytes/Bulk          43294 non-null  int64
 61  Bwd Avg Packets/Bulk        43294 non-null  int64
 62  Bwd Avg Bulk Rate           43294 non-null  int64
 63  Subflow Fwd Packets         43294 non-null  int64
 64  Subflow Fwd Bytes           43294 non-null  int64
 65  Subflow Bwd Packets         43294 non-null  int64
 66  Subflow Bwd Bytes           43294 non-null  int64
 67  Init_Win_bytes_forward      43294 non-null  int64
 68  Init_Win_bytes_backward     43294 non-null  int64
 69  act_data_pkt_fwd            43294 non-null  int64
 70  min_seg_size_forward        43294 non-null  int64
 71  Active Mean                 43294 non-null  float64
 72  Active Std                  43294 non-null  float64
 73  Active Max                  43294 non-null  int64
 74  Active Min                  43294 non-null  int64
 75  Idle Mean                   43294 non-null  float64
 76  Idle Std                    43294 non-null  float64
 77  Idle Max                    43294 non-null  int64
 78  Idle Min                    43294 non-null  int64
dtypes: float64(24), int64(55)
memory usage: 26.1 MB
```

In [48]:
```
print(df1.head())
```

```
   ID  Destination Port  Flow Duration  Total Fwd Packets  \
0   0                80       99575743                  7
1   1                80      107429687                  2
2   2                80          17291                  2
3   3                80       85038140                  5
4   4                80       82963302                  8

   Total Backward Packets  Total Length of Fwd Packets  \
0                       5                          401
1                       1                          662
2                       0                           12
3                       7                          338
4                       6                          375

   Total Length of Bwd Packets  Fwd Packet Length Max  Fwd Packet Length Min
\
0                        11595                    377                      0
1                            6                    662                      0
2                            0                      6                      6
3                        11595                    326                      0
4                        11595                    363                      0

   Fwd Packet Length Mean  ...  act_data_pkt_fwd  min_seg_size_forward  \
0               57.285714  ...                 4                    20
1              331.000000  ...                 1                    32
2                6.000000  ...                 1                    20
3               67.600000  ...                 2                    20
4               46.875000  ...                 3                    20

   Active Mean  Active Std  Active Max  Active Min    Idle Mean  Idle Std  \
0      11993.0         0.0       11993       11993   99400000.0       0.0
1          0.0         0.0           0           0  107000000.0       0.0
2          0.0         0.0           0           0          0.0       0.0
3      12007.0         0.0       12007       12007   84900000.0       0.0
4       1983.0         0.0        1983        1983   82800000.0       0.0

    Idle Max   Idle Min
0   99400000   99400000
1  107000000  107000000
2          0          0
3   84900000   84900000
4   82800000   82800000

[5 rows x 79 columns]
```

In [49]:
```
df1.head()
```

Out[49]:

	ID	Destination Port	Flow Duration	Total Fwd Packets	Total Backward Packets	Total Length of Fwd Packets	Total Length of Bwd Packets	Fwd Packet Length Max	Fwd Packet Length Min	Fwd Packe Length Mear
0	0	80	99575743	7	5	401	11595	377	0	57.28571
1	1	80	107429687	2	1	662	6	662	0	331.00000
2	2	80	17291	2	0	12	0	6	6	6.00000
3	3	80	85038140	5	7	338	11595	326	0	67.60000
4	4	80	82963302	8	6	375	11595	363	0	46.87500

5 rows × 79 columns

In [50]:
```
df1.tail()
```

Out[50]:

	ID	Destination Port	Flow Duration	Total Fwd Packets	Total Backward Packets	Total Length of Fwd Packets	Total Length of Bwd Packets	Fwd Packet Length Max	Fwd Packet Length Min	
43289	43289	80	82746195	6	7	324	11595	306	0	54
43290	43290	53	183040	4	2	176	662	44	44	44
43291	43291	443	117811197	12	13	635	4068	198	0	52
43292	43292	80	5023362	3	1	0	0	0	0	0
43293	43293	53	272	2	2	56	252	28	28	28

5 rows × 79 columns

In [51]:
```
df1.describe()
```

Out[51]:

	ID	Destination Port	Flow Duration	Total Fwd Packets	Total Backward Packets	Total Length of Fwd Packets
count	43294.000000	43294.000000	4.329400e+04	43294.000000	43294.000000	4.329400e+04
mean	21646.500000	5706.111886	2.765915e+07	14.791195	17.241650	5.718862e+02
std	12498.045614	15735.386218	4.255524e+07	1226.536554	1629.427964	8.065294e+03
min	0.000000	0.000000	-1.000000e+00	1.000000	0.000000	0.000000e+00
25%	10823.250000	53.000000	1.970000e+02	2.000000	1.000000	1.200000e+01
50%	21646.500000	80.000000	6.117150e+04	2.000000	2.000000	8.000000e+01
75%	32469.750000	443.000000	8.224463e+07	7.000000	6.000000	3.650000e+02
max	43293.000000	63913.000000	1.200000e+08	198590.000000	266796.000000	1.176395e+06

8 rows × 79 columns

In [52]:
```
df1.describe
```

Out[52]:
```
<bound method NDFrame.describe of           ID  Destination Port  Flow Durati
on  Total Fwd Packets  \
0          0                80      99575743                   7
1          1                80     107429687                   2
2          2                80         17291                   2
3          3                80      85038140                   5
4          4                80      82963302                   8
...      ...               ...           ...                 ...
43289  43289                80      82746195                   6
43290  43290                53        183040                   4
43291  43291               443     117811197                  12
43292  43292                80       5023362                   3
43293  43293                53           272                   2

       Total Backward Packets  Total Length of Fwd Packets  \
0                           5                          401
1                           1                          662
2                           0                           12
3                           7                          338
4                           6                          375
...                       ...                          ...
43289                       7                          324
43290                       2                          176
43291                      13                          635
43292                       1                            0
43293                       2                           56

       Total Length of Bwd Packets  Fwd Packet Length Max  \
0                            11595                    377
1                                6                    662
2                                0                      6
3                            11595                    326
4                            11595                    363
...                            ...                    ...
43289                        11595                    306
43290                          662                     44
43291                         4068                    198
43292                            0                      0
43293                          252                     28

       Fwd Packet Length Min  Fwd Packet Length Mean  ...  act_data_pkt_fwd
\
0                          0              57.285714  ...                 4
1                          0             331.000000  ...                 1
2                          6               6.000000  ...                 1
3                          0              67.600000  ...                 2
4                          0              46.875000  ...                 3
...                      ...                    ...  ...               ...
43289                      0              54.000000  ...                 3
43290                     44              44.000000  ...                 3
43291                      0              52.916667  ...                 9
43292                      0               0.000000  ...                 0
43293                     28              28.000000  ...                 1

       min_seg_size_forward  Active Mean   Active Std  Active Max  Active Min
\
0                        20      11993.0      0.00000       11993       11993
```

```
1                       32          0.0      0.00000          0          0
2                       20          0.0      0.00000          0          0
3                       20      12007.0      0.00000      12007      12007
4                       20       1983.0      0.00000       1983       1983
...                    ...          ...          ...        ...        ...
43289                   20      14014.0      0.00000      14014      14014
43290                   32          0.0      0.00000          0          0
43291                   20     186834.5  65208.68025     232944     140725
43292                   32          0.0      0.00000          0          0
43293                   32          0.0      0.00000          0          0

         Idle Mean     Idle Std   Idle Max   Idle Min
0       99400000.0      0.00000   99400000   99400000
1      107000000.0      0.00000  107000000  107000000
2              0.0      0.00000          0          0
3       84900000.0      0.00000   84900000   84900000
4       82800000.0      0.00000   82800000   82800000
...            ...          ...        ...        ...
43289   82600000.0      0.00000   82600000   82600000
43290          0.0      0.00000          0          0
43291   58600000.0  85333.64635   58700000   58600000
43292          0.0      0.00000          0          0
43293          0.0      0.00000          0          0

[43294 rows x 79 columns]>
```

In [53]:
```
df1.isnull()
```

Out[53]:

	ID	Destination Port	Flow Duration	Total Fwd Packets	Total Backward Packets	Total Length of Fwd Packets	Total Length of Bwd Packets	Fwd Packet Length Max	Fwd Packet Length Min	Fv Pack Leng Me
0	False	False	False	False	False	False	False	False	False	Fal
1	False	False	False	False	False	False	False	False	False	Fal
2	False	False	False	False	False	False	False	False	False	Fal
3	False	False	False	False	False	False	False	False	False	Fal
4	False	False	False	False	False	False	False	False	False	Fal
...	...	...	...	...	...	...	...	...	...	
43289	False	False	False	False	False	False	False	False	False	Fal
43290	False	False	False	False	False	False	False	False	False	Fal
43291	False	False	False	False	False	False	False	False	False	Fal
43292	False	False	False	False	False	False	False	False	False	Fal
43293	False	False	False	False	False	False	False	False	False	Fal

43294 rows × 79 columns

In [54]:
```
print(df1.isnull().sum())
```

```
ID                          0
Destination Port            0
Flow Duration               0
Total Fwd Packets           0
Total Backward Packets      0
                           ..
Active Min                  0
Idle Mean                   0
Idle Std                    0
Idle Max                    0
Idle Min                    0
Length: 79, dtype: int64
```

In [55]:
```
columns = df1.columns
count = 0
for i in df1.isnull().sum():
    print("{} : {}".format(columns[count],i))
    count+=1
```

```
ID : 0
Destination Port : 0
Flow Duration : 0
Total Fwd Packets : 0
Total Backward Packets : 0
Total Length of Fwd Packets : 0
Total Length of Bwd Packets : 0
Fwd Packet Length Max : 0
Fwd Packet Length Min : 0
Fwd Packet Length Mean : 0
Fwd Packet Length Std : 0
Bwd Packet Length Max : 0
Bwd Packet Length Min : 0
Bwd Packet Length Mean : 0
Bwd Packet Length Std : 0
Flow Bytes/s : 56
Flow Packets/s : 0
Flow IAT Mean : 0
Flow IAT Std : 0
Flow IAT Max : 0
Flow IAT Min : 0
Fwd IAT Total : 0
Fwd IAT Mean : 0
Fwd IAT Std : 0
Fwd IAT Max : 0
Fwd IAT Min : 0
Bwd IAT Total : 0
Bwd IAT Mean : 0
Bwd IAT Std : 0
Bwd IAT Max : 0
Bwd IAT Min : 0
Fwd PSH Flags : 0
Bwd PSH Flags : 0
Fwd URG Flags : 0
Bwd URG Flags : 0
Fwd Header Length : 0
Bwd Header Length : 0
Fwd Packets/s : 0
Bwd Packets/s : 0
Min Packet Length : 0
Max Packet Length : 0
Packet Length Mean : 0
Packet Length Std : 0
Packet Length Variance : 0
FIN Flag Count : 0
SYN Flag Count : 0
RST Flag Count : 0
PSH Flag Count : 0
ACK Flag Count : 0
URG Flag Count : 0
CWE Flag Count : 0
ECE Flag Count : 0
Down/Up Ratio : 0
Average Packet Size : 0
Avg Fwd Segment Size : 0
Avg Bwd Segment Size : 0
Fwd Header Length.1 : 0
```

```
Fwd Avg Bytes/Bulk : 0
Fwd Avg Packets/Bulk : 0
Fwd Avg Bulk Rate : 0
Bwd Avg Bytes/Bulk : 0
Bwd Avg Packets/Bulk : 0
Bwd Avg Bulk Rate : 0
Subflow Fwd Packets : 0
Subflow Fwd Bytes : 0
Subflow Bwd Packets : 0
Subflow Bwd Bytes : 0
Init_Win_bytes_forward : 0
Init_Win_bytes_backward : 0
act_data_pkt_fwd : 0
min_seg_size_forward : 0
Active Mean : 0
Active Std : 0
Active Max : 0
Active Min : 0
Idle Mean : 0
Idle Std : 0
Idle Max : 0
Idle Min : 0
```

In [56]:
```
df1 = df1.dropna()
```

In [57]:
```
numerical_cols = [one for each,one in zip(list(df1.dtypes),df1.dtypes.index)
df1 = df1.drop(numerical_cols,axis = 1)
```

In [58]:
```
X.shape
```

Out[58]:
```
(129679, 55)
```

In [59]:
```
df1.shape
```

Out[59]:
```
(43238, 55)
```

In [60]:
```
df.Label.unique()
```

Out[60]:
```
array(['BENIGN', 'DoS Slowhttptest', 'DoS Hulk', 'DoS GoldenEye',
       'DoS slowloris'], dtype=object)
```

In [61]:
```
predictions = dtree.predict(df1)
```

In [62]:
```
df1["predictions"] = dtree.predict(df1)
```

In [63]:
```
df1
```

Out[63]:

	ID	Destination Port	Flow Duration	Total Fwd Packets	Total Backward Packets	Total Length of Fwd Packets	Total Length of Bwd Packets	Fwd Packet Length Max	Fwd Packet Length Min	Pa Le
0	0	80	99575743	7	5	401	11595	377	0	
1	1	80	107429687	2	1	662	6	662	0	
2	2	80	17291	2	0	12	0	6	6	
3	3	80	85038140	5	7	338	11595	326	0	
4	4	80	82963302	8	6	375	11595	363	0	
...	...	...	...	...	...	...	...	...	...	
43289	43289	80	82746195	6	7	324	11595	306	0	
43290	43290	53	183040	4	2	176	662	44	44	
43291	43291	443	117811197	12	13	635	4068	198	0	
43292	43292	80	5023362	3	1	0	0	0	0	
43293	43293	53	272	2	2	56	252	28	28	

43238 rows × 56 columns

In [64]:
```
df1.predictions.unique()
```

Out[64]:
```
array([2, 3, 0, 4, 1], dtype=int64)
```

In [65]:
```
def attack_encode(value):
    if value == 0:
        return "BENIGN";
    elif value == 1:
        return "DoS Slowhttptest";
    elif value == 2:
        return 'DoS Hulk';
    elif value == 3:
        return 'DoS GoldenEye';
    else:
        return 'DoS slowloris';
```

In [66]:
```
df1['Label'] = df1['predictions'].apply(attack_encode)
df1.iloc[:10, -2:]
```

Out[66]:

	predictions	Label
0	2	DoS Hulk
1	3	DoS GoldenEye
2	0	BENIGN
3	2	DoS Hulk
4	2	DoS Hulk
5	2	DoS Hulk
6	0	BENIGN
7	2	DoS Hulk
8	2	DoS Hulk
9	2	DoS Hulk

In [67]:
```
df1
```

Out[67]:

	ID	Destination Port	Flow Duration	Total Fwd Packets	Total Backward Packets	Total Length of Fwd Packets	Total Length of Bwd Packets	Fwd Packet Length Max	Fwd Packet Length Min	Pa Le
0	0	80	99575743	7	5	401	11595	377	0	
1	1	80	107429687	2	1	662	6	662	0	
2	2	80	17291	2	0	12	0	6	6	
3	3	80	85038140	5	7	338	11595	326	0	
4	4	80	82963302	8	6	375	11595	363	0	
...	...	...	...	...	...	...	...	...	...	
43289	43289	80	82746195	6	7	324	11595	306	0	
43290	43290	53	183040	4	2	176	662	44	44	
43291	43291	443	117811197	12	13	635	4068	198	0	
43292	43292	80	5023362	3	1	0	0	0	0	
43293	43293	53	272	2	2	56	252	28	28	

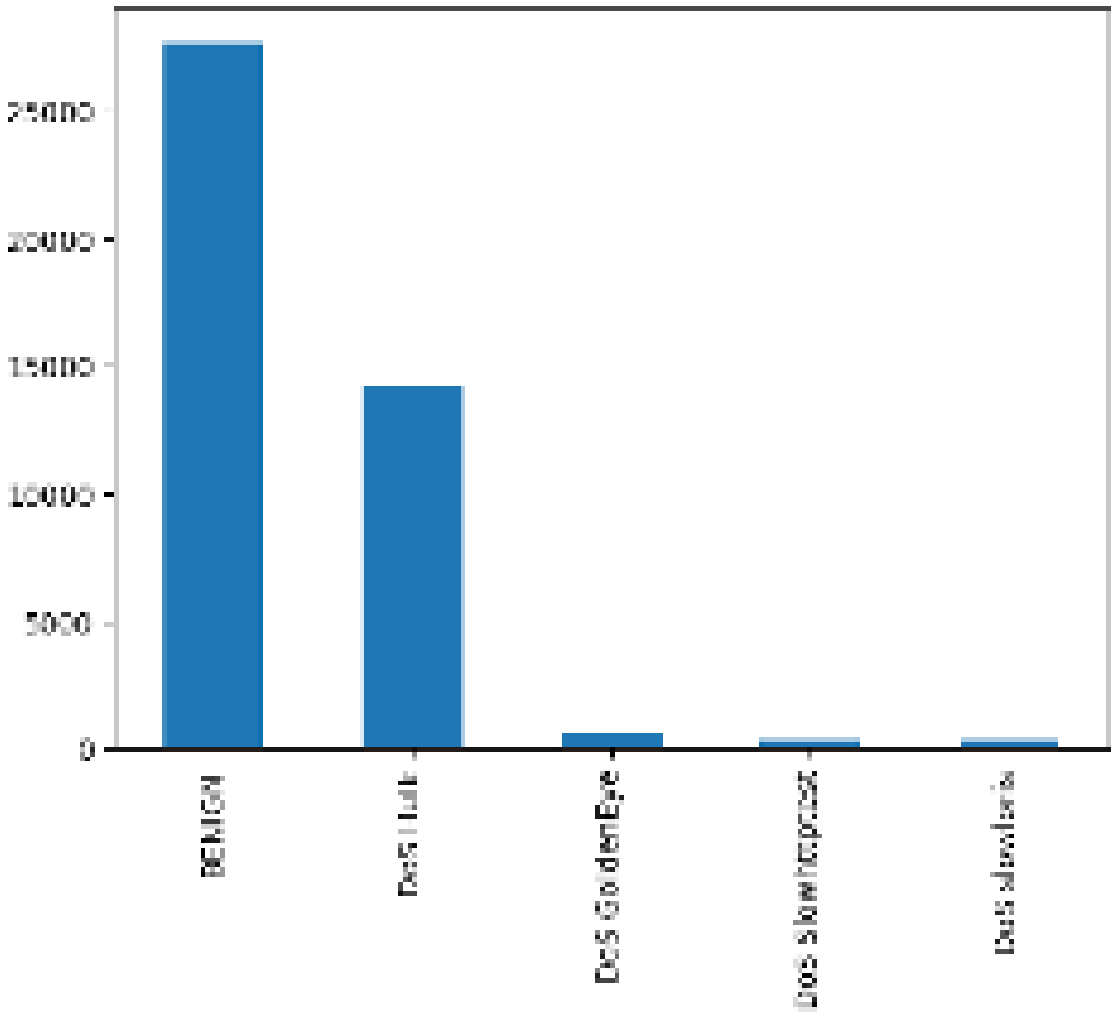
43238 rows × 57 columns

In [68]:
```
counts = df1["Label"].value_counts()
print(counts)
```

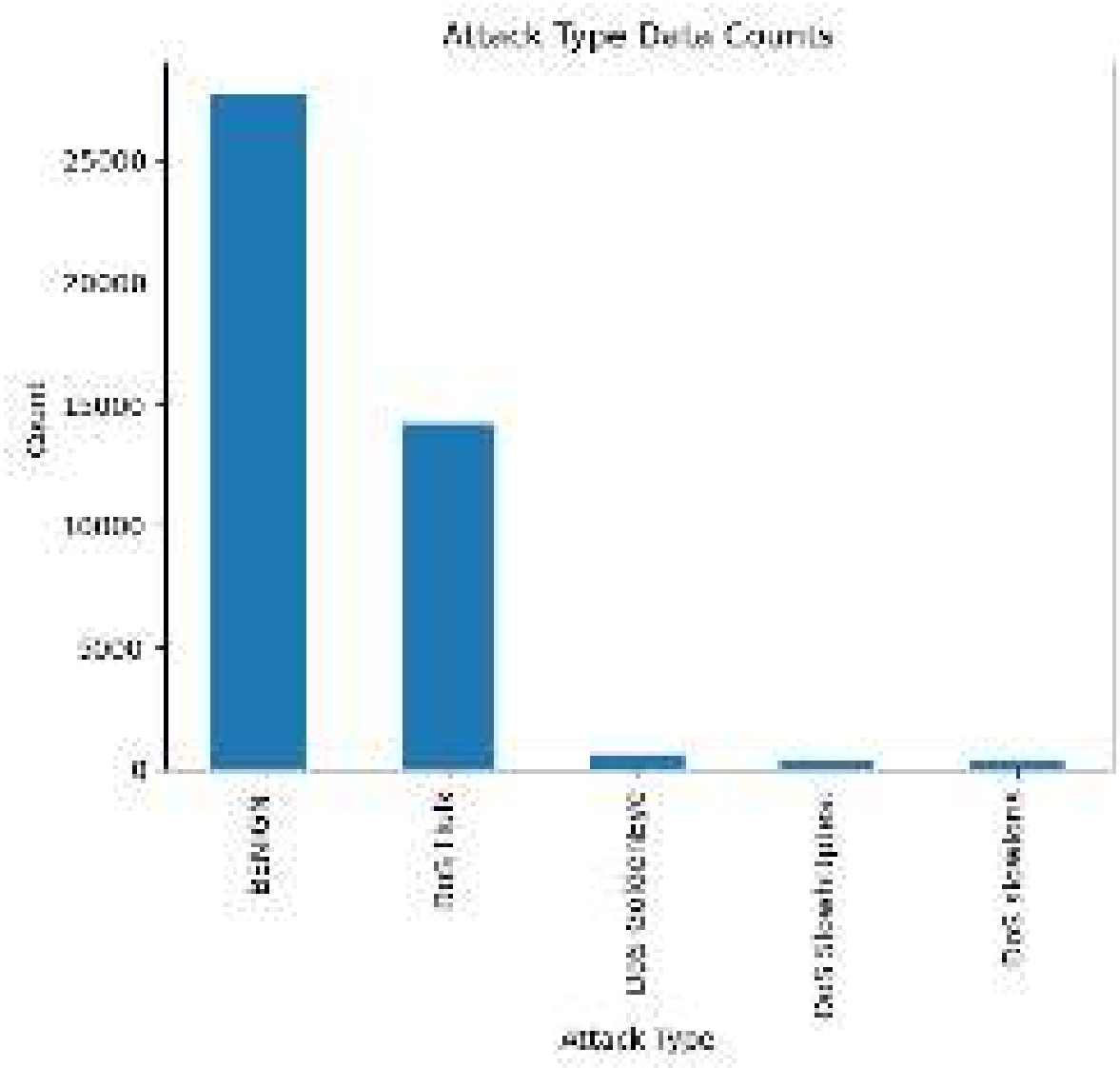
```
BENIGN             27676
DoS Hulk           14238
DoS GoldenEye        632
DoS Slowhttptest     357
DoS slowloris        335
Name: Label, dtype: int64
```

In [69]:
```
counts.plot(kind="bar")
```

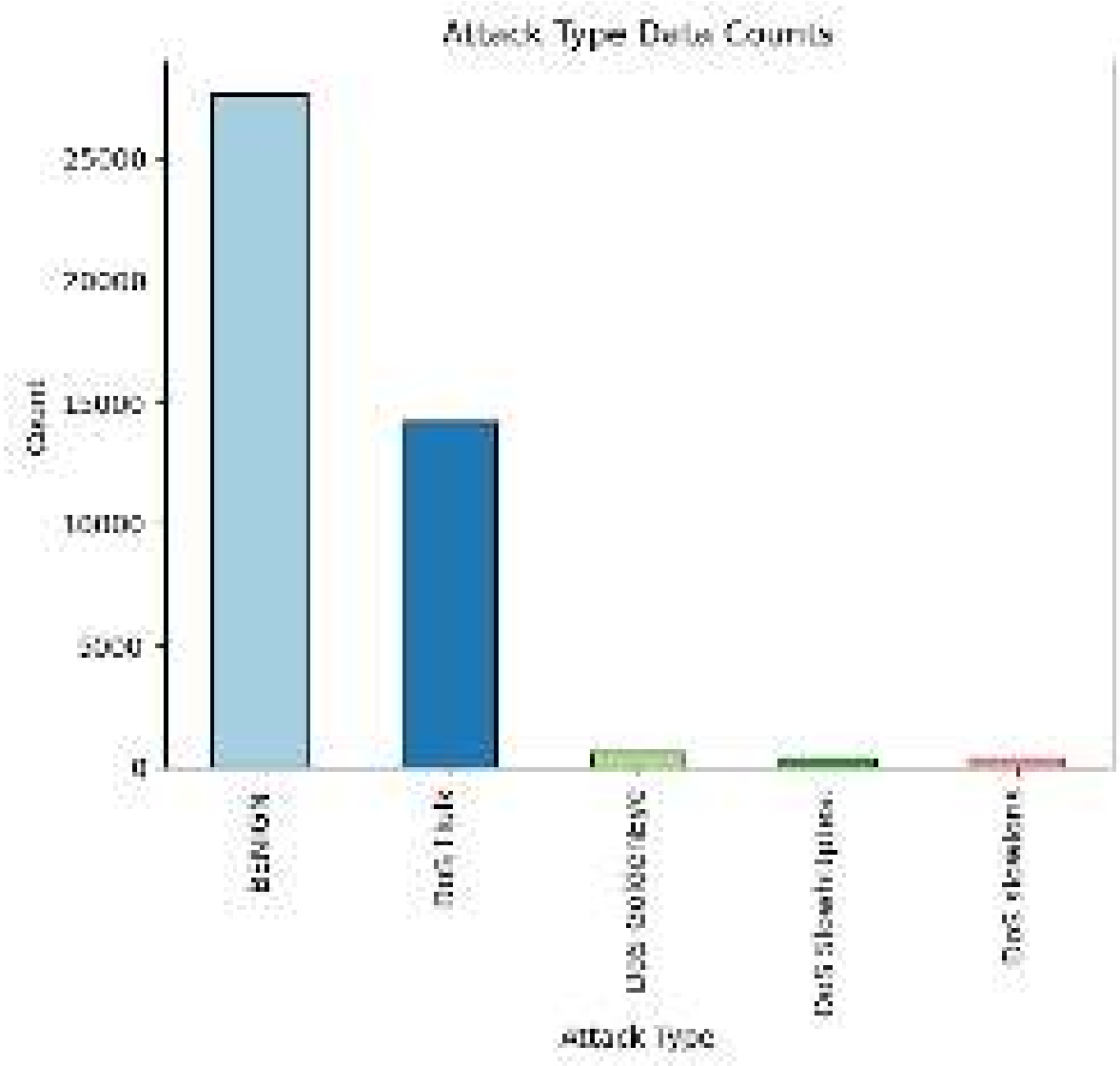
Out[69]: <AxesSubplot:>



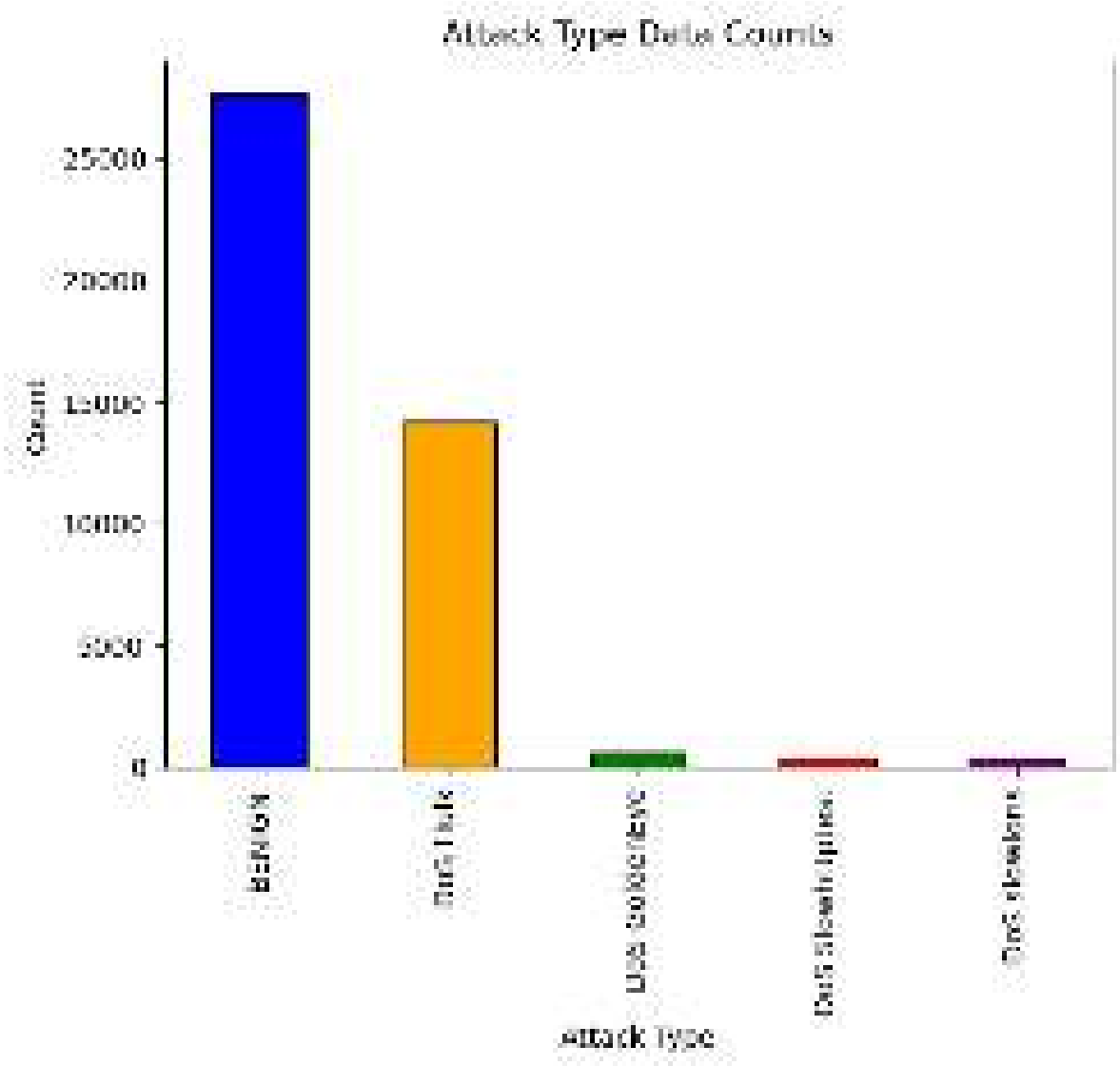
In [70]:
```
ax = counts.plot(kind='bar')


ax.set_xlabel('Attack Type')
ax.set_ylabel('Count')
ax.set_title('Attack Type Data Counts')


plt.show()
```



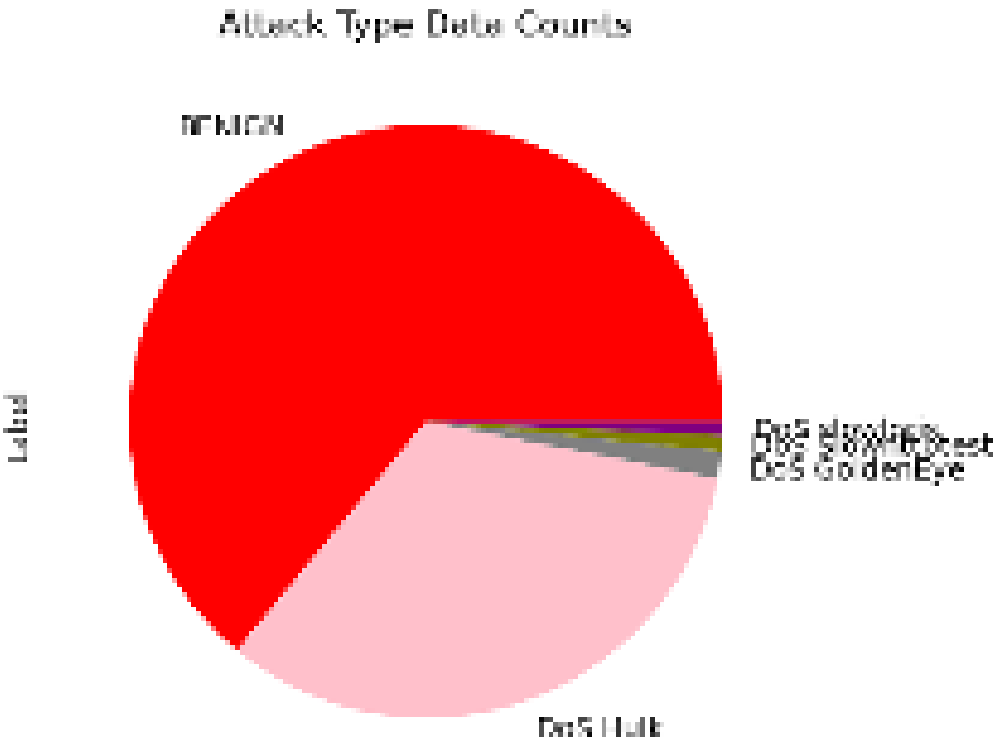
In [71]:
```
counts.plot(kind="bar", edgecolor='black', color=plt.cm.Paired(range(len(count

plt.xlabel("Attack Type")
plt.ylabel("Count")
plt.title("Attack Type Data Counts")


plt.show()
```



In [72]:
```
colors = ['blue', 'orange', 'green', 'red', 'purple', 'brown', 'pink', 'gray',
counts.plot(kind="bar", edgecolor='black',color=colors)

plt.xlabel("Attack Type")
plt.ylabel("Count")
plt.title("Attack Type Data Counts")


plt.show()
```



In [84]:
```
colors=['red', 'pink', 'gray', 'olive', 'purple']
counts.plot(kind='pie', colors=colors)


plt.title("Attack Type Data Counts")


plt.show()
```



In [74]:
```
X = df1.drop(['Label','predictions'],axis = 1)    #    81-80
y = df1['predictions']
```

In [75]:
```
print(X.shape)
print(y.shape)
```

```
(43238, 55)
(43238,)
```

In [76]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, rando
```

In [77]:
```
dtree = DecisionTreeClassifier()
```

In [78]:
```
dtree.fit(X_train,y_train)
```

Out[78]:
```
DecisionTreeClassifier()
```

In [79]:
```
predictions = dtree.predict(X_test)
print(classification_report(y_test,predictions))
```

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      8278
           1       0.99      0.95      0.97       103
           2       1.00      1.00      1.00      4300
           3       0.97      0.99      0.98       183
           4       0.95      0.99      0.97       108

    accuracy                           1.00     12972
   macro avg       0.98      0.99      0.98     12972
weighted avg       1.00      1.00      1.00     12972
```

In [80]:
```
cm=confusion_matrix(y_test,predictions)
print(cm)
```

```
[[8260    1    8    4    5]
 [   5   98    0    0    0]
 [   2    0 4296    1    1]
 [   2    0    0  181    0]
 [   1    0    0    0  107]]
```

In [81]:
```
print ("Accuracy of prediction:",round((cm[0,0]+cm[1,1]+cm[2,2]+cm[3,3]+cm[4,4
```

```
Accuracy of prediction: 0.998
```

In [82]:
```
metrics.classification_report(y_test,predictions)
metrics.plot_confusion_matrix(dtree, X_test, y_test)
```

Out[82]:
```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x19cc216d2
b0>
```