In [42]:
```
# Loading the dataset
import pandas as pd

df = pd.read_csv('RT_IOT2022.csv')
```

In [2]:
```
df.head() # checking of initial valuesS
```

Out[2]:

	no	id.orig_p	id.resp_p	proto	service	flow_duration	fwd_pkts_tot	bwd_pkts_tot	fwd
0	0	38667	1883	tcp	mqtt	32.011598	9	5	
1	1	51143	1883	tcp	mqtt	31.883584	9	5	
2	2	44761	1883	tcp	mqtt	32.124053	9	5	
3	3	60893	1883	tcp	mqtt	31.961063	9	5	
4	4	51087	1883	tcp	mqtt	31.902362	9	5	

5 rows × 85 columns

In [3]:
```
df.dtypes
```

Out[3]:
```
no                       int64
id.orig_p                int64
id.resp_p                int64
proto                   object
service                 object
                         ...
idle.std               float64
fwd_init_window_size     int64
bwd_init_window_size     int64
fwd_last_window_size     int64
Attack_type             object
Length: 85, dtype: object
```

In [4]:
```
df.isnull().sum()
```

Out[4]:
```
no                      0
id.orig_p               0
id.resp_p               0
proto                   0
service                 0
                       ..
idle.std                0
fwd_init_window_size    0
bwd_init_window_size    0
fwd_last_window_size    0
Attack_type             0
Length: 85, dtype: int64
```

In [5]:
```
df['proto'].unique()
```

Out[5]:
```
array(['tcp', 'udp', 'icmp'], dtype=object)
```

In [6]:
```
df['service'].unique()
```

Out[6]:
```
array(['mqtt', '-', 'http', 'dns', 'ntp', 'ssl', 'dhcp', 'irc', 'ssh',
       'radius'], dtype=object)
```

In [8]:
```
df['Attack_type'].unique()
```

Out[8]:
```
array(['MQTT_Publish', 'Thing_Speak', 'Wipro_bulb', 'ARP_poisioning',
       'DDOS_Slowloris', 'DOS_SYN_Hping', 'Metasploit_Brute_Force_SSH',
       'NMAP_FIN_SCAN', 'NMAP_OS_DETECTION', 'NMAP_TCP_scan',
       'NMAP_UDP_SCAN', 'NMAP_XMAS_TREE_SCAN'], dtype=object)
```

In [9]:
```
df.nlargest(5, 'fwd_pkts_per_sec')
```

Out[9]:

	no	id.orig_p	id.resp_p	proto	service	flow_duration	fwd_pkts_tot	bwd_pkts_tc
28823	8032	10874	21	tcp	-	0.000001	1	
28824	8033	10875	21	tcp	-	0.000001	1	
28825	8034	10876	21	tcp	-	0.000001	1	
28829	8038	10880	21	tcp	-	0.000001	1	
28830	8039	10881	21	tcp	-	0.000001	1	

5 rows × 85 columns

In [11]:
```
(df['service'] == '-').count()
```

Out[11]:
```
np.int64(123117)
```

In [14]:
```
df = df['service'].apply(lambda x: None if x == '-' else continue)
```

```
  Cell In[14], line 1
    df = df['service'].apply(lambda x: None if x == '-' else continue)
                                                            ^
SyntaxError: invalid syntax
```

In [17]:
```
df.nlargest(5, 'flow_pkts_per_sec')
```

Out[17]:

	no	id.orig_p	id.resp_p	proto	service	flow_duration	fwd_pkts_tot	bwd_pkts_tc
28823	8032	10874	21	tcp	-	0.000001	1	
28824	8033	10875	21	tcp	-	0.000001	1	
28825	8034	10876	21	tcp	-	0.000001	1	
28829	8038	10880	21	tcp	-	0.000001	1	
28830	8039	10881	21	tcp	-	0.000001	1	

5 rows × 85 columns

In [21]:
```
df['fwd_pkts_payload.std'].nunique()
```

Out[21]: 1739

In [25]:
```
dos = df[df['Attack_type'] == 'DDOS_Slowloris']
```

In [26]:
```
dos['proto'].unique()
```

Out[26]: array(['tcp', 'udp'], dtype=object)

In [27]:
```
dos[dos['proto'] == 'tcp']
```

Out[27]:

	no	id.orig_p	id.resp_p	proto	service	flow_duration	fwd_pkts_tot	bwd_pkts_tot
20257	0	51312	80	tcp	http	0.000181	8	0
20258	1	51314	80	tcp	http	0.000536	8	0
20259	2	51316	80	tcp	http	0.000551	8	0
20260	3	51318	80	tcp	http	0.000546	8	0
20261	4	51320	80	tcp	http	0.000539	8	0
...	...	...	...	...	...	...	...	...
20785	528	49928	80	tcp	http	0.003540	4	3
20786	529	49930	80	tcp	http	0.003119	4	3
20787	530	49932	80	tcp	http	0.003120	4	3
20788	531	49934	80	tcp	http	0.004047	4	3
20789	532	49936	80	tcp	http	0.003871	4	3

529 rows × 85 columns

In [28]:
```
df[df['fwd_bulk_packets'] > 0]
```

Out[28]:

	no	id.orig_p	id.resp_p	proto	service	flow_duration	fwd_pkts_tot	bwd_pkts_t
4520	374	52375	443	tcp	ssl	21.379392	92	1
4535	389	52372	443	tcp	ssl	11.049312	40	
4536	390	52374	443	tcp	ssl	1.106636	19	
4540	394	52370	443	tcp	ssl	65.690455	18	
4602	456	58116	443	tcp	ssl	1.100566	32	
...	...	...	...	...	...	...	...	
20224	7717	48848	443	tcp	ssl	184.457190	56	
20227	7720	57772	443	tcp	ssl	171.688333	31	
20639	382	49634	80	tcp	http	26.774708	11	
121105	2588	3	3	icmp	-	883.106953	882	
121106	2589	3	3	icmp	-	905.964201	903	

246 rows × 85 columns

In [36]:
```
df = round(df['flow_duration'], 6)
df.head()
```

Out[36]:
```
0    32.011598
1    31.883584
2    32.124053
3    31.961063
4    31.902362
Name: flow_duration, dtype: float64
```

In [45]:
```
df[df['down_up_ratio'] > df['down_up_ratio'].mean()]
```

Out[45]:

	no	id.orig_p	id.resp_p	proto	service	flow_duration	fwd_pkts_tot	bwd_pkts_t
560	560	51123	1883	tcp	mqtt	31.249999	7	
581	581	33557	1883	tcp	mqtt	54.177752	7	
1094	1094	38745	1883	tcp	mqtt	61.977579	9	
2868	2868	52589	1883	tcp	mqtt	17.921505	9	
3227	3227	50797	1883	tcp	-	0.001747	1	
...	...	...	...	...	...	...	...	
123112	2005	59247	63331	tcp	-	0.000006	1	
123113	2006	59247	64623	tcp	-	0.000007	1	
123114	2007	59247	64680	tcp	-	0.000006	1	
123115	2008	59247	65000	tcp	-	0.000006	1	
123116	2009	59247	65129	tcp	-	0.000006	1	

99139 rows × 85 columns

In [44]:
```
df.columns
```

Out[44]:
```
Index(['no', 'id.orig_p', 'id.resp_p', 'proto', 'service', 'flow_duration',
       'fwd_pkts_tot', 'bwd_pkts_tot', 'fwd_data_pkts_tot',
       'bwd_data_pkts_tot', 'fwd_pkts_per_sec', 'bwd_pkts_per_sec',
       'flow_pkts_per_sec', 'down_up_ratio', 'fwd_header_size_tot',
       'fwd_header_size_min', 'fwd_header_size_max', 'bwd_header_size_tot',
       'bwd_header_size_min', 'bwd_header_size_max', 'flow_FIN_flag_count',
       'flow_SYN_flag_count', 'flow_RST_flag_count', 'fwd_PSH_flag_count',
       'bwd_PSH_flag_count', 'flow_ACK_flag_count', 'fwd_URG_flag_count',
       'bwd_URG_flag_count', 'flow_CWR_flag_count', 'flow_ECE_flag_count',
       'fwd_pkts_payload.min', 'fwd_pkts_payload.max', 'fwd_pkts_payload.tot',
       'fwd_pkts_payload.avg', 'fwd_pkts_payload.std', 'bwd_pkts_payload.min',
       'bwd_pkts_payload.max', 'bwd_pkts_payload.tot', 'bwd_pkts_payload.avg',
       'bwd_pkts_payload.std', 'flow_pkts_payload.min',
       'flow_pkts_payload.max', 'flow_pkts_payload.tot',
       'flow_pkts_payload.avg', 'flow_pkts_payload.std', 'fwd_iat.min',
       'fwd_iat.max', 'fwd_iat.tot', 'fwd_iat.avg', 'fwd_iat.std',
       'bwd_iat.min', 'bwd_iat.max', 'bwd_iat.tot', 'bwd_iat.avg',
       'bwd_iat.std', 'flow_iat.min', 'flow_iat.max', 'flow_iat.tot',
       'flow_iat.avg', 'flow_iat.std', 'payload_bytes_per_second',
       'fwd_subflow_pkts', 'bwd_subflow_pkts', 'fwd_subflow_bytes',
       'bwd_subflow_bytes', 'fwd_bulk_bytes', 'bwd_bulk_bytes',
       'fwd_bulk_packets', 'bwd_bulk_packets', 'fwd_bulk_rate',
       'bwd_bulk_rate', 'active.min', 'active.max', 'active.tot', 'active.avg',
       'active.std', 'idle.min', 'idle.max', 'idle.tot', 'idle.avg',
       'idle.std', 'fwd_init_window_size', 'bwd_init_window_size',
       'fwd_last_window_size', 'Attack_type'],
      dtype='object')
```

In [ ]: