In [1]:
```
import pandas as pd

data = pd.read_csv("RT_IOT2022.csv")
data = pd.DataFrame(data)
data.head()
```

Out[1]:

	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 [2]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 123117 entries, 0 to 123116
Data columns (total 85 columns):
 #   Column                    Non-Null Count   Dtype
---  ------                    --------------   -----
 0   no                        123117 non-null  int64
 1   id.orig_p                 123117 non-null  int64
 2   id.resp_p                 123117 non-null  int64
 3   proto                     123117 non-null  object
 4   service                   123117 non-null  object
 5   flow_duration             123117 non-null  float64
 6   fwd_pkts_tot              123117 non-null  int64
 7   bwd_pkts_tot              123117 non-null  int64
 8   fwd_data_pkts_tot         123117 non-null  int64
 9   bwd_data_pkts_tot         123117 non-null  int64
 10  fwd_pkts_per_sec          123117 non-null  float64
 11  bwd_pkts_per_sec          123117 non-null  float64
 12  flow_pkts_per_sec         123117 non-null  float64
 13  down_up_ratio             123117 non-null  float64
 14  fwd_header_size_tot       123117 non-null  int64
 15  fwd_header_size_min       123117 non-null  int64
 16  fwd_header_size_max       123117 non-null  int64
 17  bwd_header_size_tot       123117 non-null  int64
 18  bwd_header_size_min       123117 non-null  int64
 19  bwd_header_size_max       123117 non-null  int64
 20  flow_FIN_flag_count       123117 non-null  int64
 21  flow_SYN_flag_count       123117 non-null  int64
 22  flow_RST_flag_count       123117 non-null  int64
 23  fwd_PSH_flag_count        123117 non-null  int64
 24  bwd_PSH_flag_count        123117 non-null  int64
 25  flow_ACK_flag_count       123117 non-null  int64
 26  fwd_URG_flag_count        123117 non-null  int64
 27  bwd_URG_flag_count        123117 non-null  int64
 28  flow_CWR_flag_count       123117 non-null  int64
 29  flow_ECE_flag_count       123117 non-null  int64
 30  fwd_pkts_payload.min      123117 non-null  float64
 31  fwd_pkts_payload.max      123117 non-null  float64
 32  fwd_pkts_payload.tot      123117 non-null  float64
 33  fwd_pkts_payload.avg      123117 non-null  float64
 34  fwd_pkts_payload.std      123117 non-null  float64
 35  bwd_pkts_payload.min      123117 non-null  float64
 36  bwd_pkts_payload.max      123117 non-null  float64
 37  bwd_pkts_payload.tot      123117 non-null  float64
 38  bwd_pkts_payload.avg      123117 non-null  float64
 39  bwd_pkts_payload.std      123117 non-null  float64
 40  flow_pkts_payload.min     123117 non-null  float64
 41  flow_pkts_payload.max     123117 non-null  float64
 42  flow_pkts_payload.tot     123117 non-null  float64
 43  flow_pkts_payload.avg     123117 non-null  float64
 44  flow_pkts_payload.std     123117 non-null  float64
 45  fwd_iat.min               123117 non-null  float64
 46  fwd_iat.max               123117 non-null  float64
 47  fwd_iat.tot               123117 non-null  float64
 48  fwd_iat.avg               123117 non-null  float64
 49  fwd_iat.std               123117 non-null  float64
 50  bwd_iat.min               123117 non-null  float64
```

```
 51  bwd_iat.max                123117 non-null  float64
 52  bwd_iat.tot                123117 non-null  float64
 53  bwd_iat.avg                123117 non-null  float64
 54  bwd_iat.std                123117 non-null  float64
 55  flow_iat.min               123117 non-null  float64
 56  flow_iat.max               123117 non-null  float64
 57  flow_iat.tot               123117 non-null  float64
 58  flow_iat.avg               123117 non-null  float64
 59  flow_iat.std               123117 non-null  float64
 60  payload_bytes_per_second   123117 non-null  float64
 61  fwd_subflow_pkts           123117 non-null  float64
 62  bwd_subflow_pkts           123117 non-null  float64
 63  fwd_subflow_bytes          123117 non-null  float64
 64  bwd_subflow_bytes          123117 non-null  float64
 65  fwd_bulk_bytes             123117 non-null  float64
 66  bwd_bulk_bytes             123117 non-null  float64
 67  fwd_bulk_packets           123117 non-null  float64
 68  bwd_bulk_packets           123117 non-null  float64
 69  fwd_bulk_rate              123117 non-null  float64
 70  bwd_bulk_rate              123117 non-null  float64
 71  active.min                 123117 non-null  float64
 72  active.max                 123117 non-null  float64
 73  active.tot                 123117 non-null  float64
 74  active.avg                 123117 non-null  float64
 75  active.std                 123117 non-null  float64
 76  idle.min                   123117 non-null  float64
 77  idle.max                   123117 non-null  float64
 78  idle.tot                   123117 non-null  float64
 79  idle.avg                   123117 non-null  float64
 80  idle.std                   123117 non-null  float64
 81  fwd_init_window_size       123117 non-null  int64
 82  bwd_init_window_size       123117 non-null  int64
 83  fwd_last_window_size       123117 non-null  int64
 84  Attack_type                123117 non-null  object
dtypes: float64(56), int64(26), object(3)
memory usage: 79.8+ MB
```

In [3]:
```
data["flow_duration"].mean()
```

Out[3]:
```
3.8095657699992693
```

In [4]:
```
data["Attack_type"].mode()
```

Out[4]:
```
0    DOS_SYN_Hping
Name: Attack_type, dtype: object
```

In [5]:
```
data[["flow_duration", "fwd_pkts_per_sec", "bwd_pkts_per_sec", "flow_pkts_per_sec",
data.head()
```

Out[5]:

	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 [13]:
```
# Get the time duration each attack types

data_attack_type = data.groupby(["Attack_type"])["flow_duration"].mean()
data_attack_type
```

Out[13]:
```
Attack_type
ARP_poisioning                 15.893538
DDOS_Slowloris                 14.699148
DOS_SYN_Hping                   0.000003
MQTT_Publish                   43.397013
Metasploit_Brute_Force_SSH      3.006557
NMAP_FIN_SCAN                   0.023614
NMAP_OS_DETECTION               0.000008
NMAP_TCP_scan                   0.000019
NMAP_UDP_SCAN                   0.737766
NMAP_XMAS_TREE_SCAN             0.001171
Thing_Speak                     0.934471
Wipro_bulb                    586.845727
Name: flow_duration, dtype: float64
```

In [7]:
```
# Get the number of each attack types

data_attack_type_1 = data.groupby(["Attack_type"])["Attack_type"].count()
data_attack_type_1
```

Out[7]:
```
Attack_type
ARP_poisioning                 7750
DDOS_Slowloris                  534
DOS_SYN_Hping                 94659
MQTT_Publish                   4146
Metasploit_Brute_Force_SSH       37
NMAP_FIN_SCAN                    28
NMAP_OS_DETECTION              2000
NMAP_TCP_scan                  1002
NMAP_UDP_SCAN                  2590
NMAP_XMAS_TREE_SCAN            2010
Thing_Speak                    8108
Wipro_bulb                      253
Name: Attack_type, dtype: int64
```

In [ ]: