

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

```
In [68]: iot = pd.read_csv('RT_IOT2022.csv')
```

```
In [69]: iot
```

```
Out[69]:
```

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

123117 rows × 85 columns



```
In [99]: # Filter necessary data
loaded = iot[['no', 'id.orig_p', 'proto', 'service', 'flow_duration', 'Attack_type']]
loaded.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 123117 entries, 0 to 123116
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   no               123117 non-null  int64
1   id.orig_p        123117 non-null  int64
2   proto            123117 non-null  object
3   service          123117 non-null  object
4   flow_duration    123117 non-null  float64
5   Attack_type      123117 non-null  object
dtypes: float64(1), int64(2), object(3)
memory usage: 5.6+ MB
```

```
In [102... loaded
```

Out[102...

	no	id.orig_p	proto	service	flow_duration	Attack_type
0	0	38667	tcp	mqtt	32.011598	MQTT_Publish
1	1	51143	tcp	mqtt	31.883584	MQTT_Publish
2	2	44761	tcp	mqtt	32.124053	MQTT_Publish
3	3	60893	tcp	mqtt	31.961063	MQTT_Publish
4	4	51087	tcp	mqtt	31.902362	MQTT_Publish
...
123112	2005	59247	tcp	-	0.000006	NMAP_XMAS_TREE_SCAN
123113	2006	59247	tcp	-	0.000007	NMAP_XMAS_TREE_SCAN
123114	2007	59247	tcp	-	0.000006	NMAP_XMAS_TREE_SCAN
123115	2008	59247	tcp	-	0.000006	NMAP_XMAS_TREE_SCAN
123116	2009	59247	tcp	-	0.000006	NMAP_XMAS_TREE_SCAN

123117 rows × 6 columns

In [139...

```
loaded = loaded.dropna()  
loaded
```

Out[139...

	no	id.orig_p	proto	service	flow_duration	Attack_type
0	0	38667	tcp	mqtt	32.011598	MQTT_Publish
1	1	51143	tcp	mqtt	31.883584	MQTT_Publish
2	2	44761	tcp	mqtt	32.124053	MQTT_Publish
3	3	60893	tcp	mqtt	31.961063	MQTT_Publish
4	4	51087	tcp	mqtt	31.902362	MQTT_Publish
...
123112	2005	59247	tcp	-	0.000006	NMAP_XMAS_TREE_SCAN
123113	2006	59247	tcp	-	0.000007	NMAP_XMAS_TREE_SCAN
123114	2007	59247	tcp	-	0.000006	NMAP_XMAS_TREE_SCAN
123115	2008	59247	tcp	-	0.000006	NMAP_XMAS_TREE_SCAN
123116	2009	59247	tcp	-	0.000006	NMAP_XMAS_TREE_SCAN

123117 rows × 6 columns

In [127...

```
iot['Attack_type'].unique()
```

```
Out[127...] array(['MQTT_Publish', 'Thing_Speak', 'Wipro_bulb', 'ARP_poisoning',  
      '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 [125...] MQTT_Publish = loaded[loaded['Attack_type'] == 'MQTT_Publish']  
MQTT_Publish.count()
```

```
Out[125...] no                4146  
id.orig_p      4146  
proto          4146  
service        4146  
flow_duration  4146  
Attack_type    4146  
dtype: int64
```

```
In [126...] thing_speak = loaded[loaded['Attack_type'] == 'Thing_Speak']  
thing_speak.count()
```

```
Out[126...] no                8108  
id.orig_p      8108  
proto          8108  
service        8108  
flow_duration  8108  
Attack_type    8108  
dtype: int64
```

```
In [124...] Wipro_bulb = loaded[loaded['Attack_type'] == 'Wipro_bulb']  
Wipro_bulb.count()
```

```
Out[124...] no                253  
id.orig_p      253  
proto          253  
service        253  
flow_duration  253  
Attack_type    253  
dtype: int64
```

```
In [119...] ARP_poisoning = loaded[loaded['Attack_type'] == 'ARP_poisoning']  
ARP_poisoning.count()
```

```
Out[119...] no                7750  
id.orig_p      7750  
proto          7750  
service        7750  
flow_duration  7750  
Attack_type    7750  
dtype: int64
```

```
In [120...] DDOS_Slowloris = loaded[loaded['Attack_type'] == 'DDOS_Slowloris']  
DDOS_Slowloris.count()
```

```
Out[120...] no          534
            id.orig_p  534
            proto      534
            service    534
            flow_duration 534
            Attack_type 534
            dtype: int64
```

```
In [128...] DOS_SYN_Hping = loaded[loaded['Attack_type'] == 'DOS_SYN_Hping']
DOS_SYN_Hping.count()
```

```
Out[128...] no          94659
            id.orig_p  94659
            proto      94659
            service    94659
            flow_duration 94659
            Attack_type 94659
            dtype: int64
```

```
In [129...] Metasploit_Brute_Force_SSH = loaded[loaded['Attack_type'] == 'Metasploit_Brute_Forc
Metasploit_Brute_Force_SSH.count()
```

```
Out[129...] no          37
            id.orig_p  37
            proto      37
            service    37
            flow_duration 37
            Attack_type 37
            dtype: int64
```

```
In [130...] NMAP_FIN_SCAN = loaded[loaded['Attack_type'] == 'NMAP_FIN_SCAN']
NMAP_FIN_SCAN.count()
```

```
Out[130...] no          28
            id.orig_p  28
            proto      28
            service    28
            flow_duration 28
            Attack_type 28
            dtype: int64
```

```
In [143...] NMAP_OS_DETECTION = loaded[loaded['Attack_type'] == 'NMAP_OS_DETECTION']
NMAP_OS_DETECTION.count()
```

```
Out[143...] no          2000
            id.orig_p  2000
            proto      2000
            service    2000
            flow_duration 2000
            Attack_type 2000
            dtype: int64
```

```
In [144...] NMAP_TCP_scan = loaded[loaded['Attack_type'] == 'NMAP_TCP_scan']
NMAP_TCP_scan.count()
```

Out[144... no 1002
id.orig_p 1002
proto 1002
service 1002
flow_duration 1002
Attack_type 1002
dtype: int64

```
In [145... NMAP_UDP_SCAN = loaded[loaded['Attack_type'] == 'NMAP_UDP_SCAN']  
NMAP_UDP_SCAN.count()
```

Out[145... no 2590
id.orig_p 2590
proto 2590
service 2590
flow_duration 2590
Attack_type 2590
dtype: int64

```
In [146... NMAP_XMAS_TREE_SCAN = loaded[loaded['Attack_type'] == 'NMAP_XMAS_TREE_SCAN']  
NMAP_XMAS_TREE_SCAN.count()
```

Out[146... no 2010
id.orig_p 2010
proto 2010
service 2010
flow_duration 2010
Attack_type 2010
dtype: int64

```
In [141... loaded.describe()
```

Out[141...

	no	id.orig_p	flow_duration
count	123117.000000	123117.000000	123117.000000
mean	37035.089248	34639.258738	3.809566
std	30459.106367	19070.620354	130.005408
min	0.000000	0.000000	0.000000
25%	6059.000000	17702.000000	0.000001
50%	33100.000000	37221.000000	0.000004
75%	63879.000000	50971.000000	0.000005
max	94658.000000	65535.000000	21728.335578

```
In [142... loaded.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 123117 entries, 0 to 123116
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   no               123117 non-null  int64
1   id.orig_p        123117 non-null  int64
2   proto            123117 non-null  object
3   service          123117 non-null  object
4   flow_duration    123117 non-null  float64
5   Attack_type      123117 non-null  object
dtypes: float64(1), int64(2), object(3)
memory usage: 5.6+ MB
```