Quiz 2 - Midterm

Extract

In [22]:
```
import pandas as pd
import numpy as np
import matplotlib as plt
IoT = pd.read_csv('RT_IOT2022.csv')
IoT.count()
```

Out[22]:
```
no                      123117
id.orig_p               123117
id.resp_p               123117
proto                   123117
service                 123117
                         ...
idle.std                123117
fwd_init_window_size    123117
bwd_init_window_size    123117
fwd_last_window_size    123117
Attack_type             123117
Length: 85, dtype: int64
```

In [23]:
```
IoT.head(10)
```

Out[23]:

	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	5	48579	1883	tcp	mqtt	31.869686	9	5	
6	6	54063	1883	tcp	mqtt	32.094711	9	5	
7	7	33457	1883	tcp	mqtt	32.104011	9	5	
8	8	52181	1883	tcp	mqtt	32.026967	9	5	
9	9	53469	1883	tcp	mqtt	32.048637	9	5	

10 rows × 85 columns

In [24]:
```
IoT.tail(10)
```

Out[24]:

	no	id.orig_p	id.resp_p	proto	service	flow_duration	fwd_pkts_tot	bwd_pkts_t
123107	2000	59247	55600	tcp	-	0.000007	1	
123108	2001	59247	57797	tcp	-	0.000006	1	
123109	2002	59247	60020	tcp	-	0.000007	1	
123110	2003	59247	60443	tcp	-	0.000006	1	
123111	2004	59247	61900	tcp	-	0.000007	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	

10 rows × 85 columns

In [25]:
```
IoT.dtypes
```

Out[25]:
```
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 [35]:
```
print(IoT['Attack_type'].values)
```

```
['Wipro_bulb' 'Wipro_bulb' 'Wipro_bulb' ... 'DOS_SYN_Hping'
 'DOS_SYN_Hping' 'DOS_SYN_Hping']
```

Transform

In [31]:
```
#Sorting by flow_duration
IoT = IoT.sort_values(by='flow_duration',ascending=False)
IoT.head(20)
```

Out[31]:

	no	id.orig_p	id.resp_p	proto	service	flow_duration	fwd_pkts_tot	bwd_pkts_t
12506	252	59766	6667	udp	-	21728.335578	4345	
12505	251	40434	8886	tcp	ssl	18761.401291	704	4
12420	166	62366	6667	udp	-	17747.121108	3549	
12419	165	40261	8886	tcp	ssl	17732.696970	671	3
12317	63	62969	6667	udp	-	9433.886888	1887	
12285	31	62969	6667	udp	-	9433.886888	1887	
12316	62	40545	8886	tcp	-	9379.541767	440	2
12284	30	40545	8886	tcp	-	9379.541767	440	2
12353	99	40533	8886	tcp	-	6950.994837	398	2
12342	88	62969	6667	udp	-	6828.456767	1344	
12438	184	40856	8886	tcp	ssl	5678.531132	274	1
12437	183	57069	6667	udp	-	5676.684995	1136	
20154	7647	39242	443	tcp	ssl	5341.392332	812	5
20152	7645	44058	443	tcp	ssl	5340.386892	365	1
12468	214	40140	8886	tcp	ssl	2943.393571	164	
15208	2701	46602	443	tcp	ssl	2379.349634	106	1
15183	2676	37242	443	tcp	ssl	2083.013699	1453	19
14995	2488	59400	443	tcp	ssl	1401.219948	197	1
121106	2589	3	3	icmp	-	905.964201	903	
14908	2401	33802	443	tcp	-	900.887802	7	

20 rows × 85 columns

In [66]:
```
IoT.tail(20)
```

Out[66]:

	no	id.orig_p	id.resp_p	proto	service	flow_duration	fwd_pkts_tot	bwd_pkts_t
23587	2796	5637	21	tcp	-	0.0	1	
18628	6121	59906	137	udp	dns	0.0	1	
18623	6116	48142	5353	udp	dns	0.0	1	
18624	6117	48789	5353	udp	dns	0.0	1	
18625	6118	60506	5353	udp	dns	0.0	1	
18626	6119	47434	5353	udp	dns	0.0	1	
18627	6120	60834	5353	udp	dns	0.0	1	
56648	35857	38771	21	tcp	-	0.0	1	
56739	35948	38863	21	tcp	-	0.0	1	
56756	35965	38880	21	tcp	-	0.0	1	
56761	35970	38885	21	tcp	-	0.0	1	
23706	2915	5756	21	tcp	-	0.0	1	
56827	36036	38951	21	tcp	-	0.0	1	
23695	2904	5745	21	tcp	-	0.0	1	
56842	36051	38966	21	tcp	-	0.0	1	
18650	6143	52390	5353	udp	dns	0.0	1	
18651	6144	36778	5353	udp	dns	0.0	1	
56865	36074	38989	21	tcp	-	0.0	1	
56871	36080	38995	21	tcp	-	0.0	1	
73172	52381	54911	21	tcp	-	0.0	1	

20 rows × 85 columns

Load

In [65]:
```
IoT.describe()
```

Out[65]:

	no	id.orig_p	id.resp_p	flow_duration	fwd_pkts_tot	bwd_p
count	123117.000000	123117.000000	123117.000000	123117.000000	123117.000000	123117.
mean	37035.089248	34639.258738	1014.305092	3.809566	2.268826	1.
std	30459.106367	19070.620354	5256.371994	130.005408	22.336565	33.
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.
25%	6059.000000	17702.000000	21.000000	0.000001	1.000000	1.
50%	33100.000000	37221.000000	21.000000	0.000004	1.000000	1.
75%	63879.000000	50971.000000	21.000000	0.000005	1.000000	1.
max	94658.000000	65535.000000	65389.000000	21728.335578	4345.000000	10112.

8 rows × 82 columns

In [ ]:
```
#summary statistics for rows with values ARP_poisioning in Attack_type column
```

In [ ]:
```
#summary statistics for rows with values DOS_SYN_Hping in Attack_type column
```

In [ ]:
```
#summary statistics for rows with values NMAP_OS_DETECTION in Attack_type column
```

In [ ]:
```
#summary statistics for rows with values NMAP_TCP_scan in Attack_type column
```

In [ ]:
```
#summary statistics for rows with values NMAP_UDP_SCAN in Attack_type column
```

In [ ]:
```
#summary statistics for rows with values NMAP_XMAS_TREE_SCAN in Attack_type column
```

In [ ]:
```
#summary statistics for rows with values NMAP_FIN_SCAN in Attack_type column
```

In [ ]:
```
#summary statistics for rows with values Metasploit_Brute_Force_SSH in Attack_type
```

In [ ]:
```
#summary statistics for rows with values DDOS_Slowloris in Attack_type column
```