

Phase	Activity	Code and Output																																																																																																																									
Extract	Counting the number of observations	<pre>IoT.count()  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</pre>																																																																																																																									
	Displaying the first 10 observations	<pre>IoT.head(10)</pre> <table><thead><tr><th></th><th>no</th><th>id.orig_p</th><th>id.resp_p</th><th>proto</th><th>service</th><th>flow_duration</th><th>fwd_pkts_tot</th><th>bwd_pkts_tot</th><th>fwd_data_pkts_tot</th><th>bwd_data_pkts_tot</th></tr></thead><tbody><tr><td>0</td><td>0</td><td>38667</td><td>1883</td><td>tcp</td><td>mqtt</td><td>32.011598</td><td>9</td><td>5</td><td>3</td><td></td></tr><tr><td>1</td><td>1</td><td>51143</td><td>1883</td><td>tcp</td><td>mqtt</td><td>31.883584</td><td>9</td><td>5</td><td>3</td><td></td></tr><tr><td>2</td><td>2</td><td>44761</td><td>1883</td><td>tcp</td><td>mqtt</td><td>32.124053</td><td>9</td><td>5</td><td>3</td><td></td></tr><tr><td>3</td><td>3</td><td>60893</td><td>1883</td><td>tcp</td><td>mqtt</td><td>31.961063</td><td>9</td><td>5</td><td>3</td><td></td></tr><tr><td>4</td><td>4</td><td>51087</td><td>1883</td><td>tcp</td><td>mqtt</td><td>31.902362</td><td>9</td><td>5</td><td>3</td><td></td></tr><tr><td>5</td><td>5</td><td>48579</td><td>1883</td><td>tcp</td><td>mqtt</td><td>31.869686</td><td>9</td><td>5</td><td>3</td><td></td></tr><tr><td>6</td><td>6</td><td>54063</td><td>1883</td><td>tcp</td><td>mqtt</td><td>32.094711</td><td>9</td><td>5</td><td>3</td><td></td></tr><tr><td>7</td><td>7</td><td>33457</td><td>1883</td><td>tcp</td><td>mqtt</td><td>32.104011</td><td>9</td><td>5</td><td>3</td><td></td></tr><tr><td>8</td><td>8</td><td>52181</td><td>1883</td><td>tcp</td><td>mqtt</td><td>32.026967</td><td>9</td><td>5</td><td>3</td><td></td></tr><tr><td>9</td><td>9</td><td>53469</td><td>1883</td><td>tcp</td><td>mqtt</td><td>32.048637</td><td>9</td><td>5</td><td>3</td><td></td></tr></tbody></table> <p>10 rows × 85 columns</p>		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	0	0	38667	1883	tcp	mqtt	32.011598	9	5	3		1	1	51143	1883	tcp	mqtt	31.883584	9	5	3		2	2	44761	1883	tcp	mqtt	32.124053	9	5	3		3	3	60893	1883	tcp	mqtt	31.961063	9	5	3		4	4	51087	1883	tcp	mqtt	31.902362	9	5	3		5	5	48579	1883	tcp	mqtt	31.869686	9	5	3		6	6	54063	1883	tcp	mqtt	32.094711	9	5	3		7	7	33457	1883	tcp	mqtt	32.104011	9	5	3		8	8	52181	1883	tcp	mqtt	32.026967	9	5	3		9	9	53469	1883	tcp	mqtt	32.048637	9	5	3	
		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																																																																																																																
	0	0	38667	1883	tcp	mqtt	32.011598	9	5	3																																																																																																																	
1	1	51143	1883	tcp	mqtt	31.883584	9	5	3																																																																																																																		
2	2	44761	1883	tcp	mqtt	32.124053	9	5	3																																																																																																																		
3	3	60893	1883	tcp	mqtt	31.961063	9	5	3																																																																																																																		
4	4	51087	1883	tcp	mqtt	31.902362	9	5	3																																																																																																																		
5	5	48579	1883	tcp	mqtt	31.869686	9	5	3																																																																																																																		
6	6	54063	1883	tcp	mqtt	32.094711	9	5	3																																																																																																																		
7	7	33457	1883	tcp	mqtt	32.104011	9	5	3																																																																																																																		
8	8	52181	1883	tcp	mqtt	32.026967	9	5	3																																																																																																																		
9	9	53469	1883	tcp	mqtt	32.048637	9	5	3																																																																																																																		
	Displaying the last 10 observations	<pre>IoT.tail(10)</pre> <table><thead><tr><th></th><th>no</th><th>id.orig_p</th><th>id.resp_p</th><th>proto</th><th>service</th><th>flow_duration</th><th>fwd_pkts_tot</th><th>bwd_pkts_tot</th><th>fwd_data_pkts_tot</th><th>bwd_data_pkts_tot</th></tr></thead><tbody><tr><td>123107</td><td>2000</td><td>59247</td><td>55600</td><td>tcp</td><td>-</td><td>0.000007</td><td>1</td><td>1</td><td>0</td><td></td></tr><tr><td>123108</td><td>2001</td><td>59247</td><td>57797</td><td>tcp</td><td>-</td><td>0.000006</td><td>1</td><td>1</td><td>0</td><td></td></tr><tr><td>123109</td><td>2002</td><td>59247</td><td>60020</td><td>tcp</td><td>-</td><td>0.000007</td><td>1</td><td>1</td><td>0</td><td></td></tr><tr><td>123110</td><td>2003</td><td>59247</td><td>60443</td><td>tcp</td><td>-</td><td>0.000006</td><td>1</td><td>1</td><td>0</td><td></td></tr><tr><td>123111</td><td>2004</td><td>59247</td><td>61900</td><td>tcp</td><td>-</td><td>0.000007</td><td>1</td><td>1</td><td>0</td><td></td></tr><tr><td>123112</td><td>2005</td><td>59247</td><td>63331</td><td>tcp</td><td>-</td><td>0.000006</td><td>1</td><td>1</td><td>0</td><td></td></tr><tr><td>123113</td><td>2006</td><td>59247</td><td>64623</td><td>tcp</td><td>-</td><td>0.000007</td><td>1</td><td>1</td><td>0</td><td></td></tr><tr><td>123114</td><td>2007</td><td>59247</td><td>64680</td><td>tcp</td><td>-</td><td>0.000006</td><td>1</td><td>1</td><td>0</td><td></td></tr><tr><td>123115</td><td>2008</td><td>59247</td><td>65000</td><td>tcp</td><td>-</td><td>0.000006</td><td>1</td><td>1</td><td>0</td><td></td></tr><tr><td>123116</td><td>2009</td><td>59247</td><td>65129</td><td>tcp</td><td>-</td><td>0.000006</td><td>1</td><td>1</td><td>0</td><td></td></tr></tbody></table> <p>10 rows × 85 columns</p>		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	123107	2000	59247	55600	tcp	-	0.000007	1	1	0		123108	2001	59247	57797	tcp	-	0.000006	1	1	0		123109	2002	59247	60020	tcp	-	0.000007	1	1	0		123110	2003	59247	60443	tcp	-	0.000006	1	1	0		123111	2004	59247	61900	tcp	-	0.000007	1	1	0		123112	2005	59247	63331	tcp	-	0.000006	1	1	0		123113	2006	59247	64623	tcp	-	0.000007	1	1	0		123114	2007	59247	64680	tcp	-	0.000006	1	1	0		123115	2008	59247	65000	tcp	-	0.000006	1	1	0		123116	2009	59247	65129	tcp	-	0.000006	1	1	0	
	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																																																																																																																	
123107	2000	59247	55600	tcp	-	0.000007	1	1	0																																																																																																																		
123108	2001	59247	57797	tcp	-	0.000006	1	1	0																																																																																																																		
123109	2002	59247	60020	tcp	-	0.000007	1	1	0																																																																																																																		
123110	2003	59247	60443	tcp	-	0.000006	1	1	0																																																																																																																		
123111	2004	59247	61900	tcp	-	0.000007	1	1	0																																																																																																																		
123112	2005	59247	63331	tcp	-	0.000006	1	1	0																																																																																																																		
123113	2006	59247	64623	tcp	-	0.000007	1	1	0																																																																																																																		
123114	2007	59247	64680	tcp	-	0.000006	1	1	0																																																																																																																		
123115	2008	59247	65000	tcp	-	0.000006	1	1	0																																																																																																																		
123116	2009	59247	65129	tcp	-	0.000006	1	1	0																																																																																																																		
	Displaying the data types of each column	<pre>IoT.dtypes</pre>																																																																																																																									

		<pre>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</pre>																																																																																																																																																																								
	Identifying the Attack types within the dataset	<pre>IoT['Attack_type'].values array(['Wipro_bulb', 'Wipro_bulb', 'Wipro_bulb', ..., 'DOS_SYN_Hping', 'DOS_SYN_Hping', 'DOS_SYN_Hping'], dtype=object)</pre>																																																																																																																																																																								
Transform	Sorting the dataset by 'flow_duration'	<pre>IoT = IoT.sort_values(by='flow_duration',ascending=False)</pre> <table><thead><tr><th></th><th>no</th><th>id.orig_p</th><th>id.resp_p</th><th>proto</th><th>service</th><th>flow_duration</th><th>fu</th></tr></thead><tbody><tr><td><b>12506</b></td><td>252</td><td>59766</td><td>6667</td><td>udp</td><td>-</td><td>21728.335578</td><td></td></tr><tr><td><b>12505</b></td><td>251</td><td>40434</td><td>8886</td><td>tcp</td><td>ssl</td><td>18761.401291</td><td></td></tr><tr><td><b>12420</b></td><td>166</td><td>62366</td><td>6667</td><td>udp</td><td>-</td><td>17747.121108</td><td></td></tr><tr><td><b>12419</b></td><td>165</td><td>40261</td><td>8886</td><td>tcp</td><td>ssl</td><td>17732.696970</td><td></td></tr><tr><td><b>12317</b></td><td>63</td><td>62969</td><td>6667</td><td>udp</td><td>-</td><td>9433.886888</td><td></td></tr><tr><td><b>12285</b></td><td>31</td><td>62969</td><td>6667</td><td>udp</td><td>-</td><td>9433.886888</td><td></td></tr><tr><td><b>12316</b></td><td>62</td><td>40545</td><td>8886</td><td>tcp</td><td>-</td><td>9379.541767</td><td></td></tr><tr><td><b>12284</b></td><td>30</td><td>40545</td><td>8886</td><td>tcp</td><td>-</td><td>9379.541767</td><td></td></tr><tr><td><b>12353</b></td><td>99</td><td>40533</td><td>8886</td><td>tcp</td><td>-</td><td>6950.994837</td><td></td></tr><tr><td><b>12342</b></td><td>88</td><td>62969</td><td>6667</td><td>udp</td><td>-</td><td>6828.456767</td><td></td></tr><tr><td><b>12438</b></td><td>184</td><td>40856</td><td>8886</td><td>tcp</td><td>ssl</td><td>5678.531132</td><td></td></tr><tr><td><b>12437</b></td><td>183</td><td>57069</td><td>6667</td><td>udp</td><td>-</td><td>5676.684995</td><td></td></tr><tr><td><b>20154</b></td><td>7647</td><td>39242</td><td>443</td><td>tcp</td><td>ssl</td><td>5341.392332</td><td></td></tr><tr><td><b>20152</b></td><td>7645</td><td>44058</td><td>443</td><td>tcp</td><td>ssl</td><td>5340.386892</td><td></td></tr><tr><td><b>12468</b></td><td>214</td><td>40140</td><td>8886</td><td>tcp</td><td>ssl</td><td>2943.393571</td><td></td></tr><tr><td><b>15208</b></td><td>2701</td><td>46602</td><td>443</td><td>tcp</td><td>ssl</td><td>2379.349634</td><td></td></tr><tr><td><b>15183</b></td><td>2676</td><td>37242</td><td>443</td><td>tcp</td><td>ssl</td><td>2083.013699</td><td></td></tr><tr><td><b>14995</b></td><td>2488</td><td>59400</td><td>443</td><td>tcp</td><td>ssl</td><td>1401.219948</td><td></td></tr><tr><td><b>121106</b></td><td>2589</td><td>3</td><td>3</td><td>icmp</td><td>-</td><td>905.964201</td><td></td></tr><tr><td><b>14908</b></td><td>2401</td><td>33802</td><td>443</td><td>tcp</td><td>-</td><td>900.887802</td><td></td></tr></tbody></table>		no	id.orig_p	id.resp_p	proto	service	flow_duration	fu	<b>12506</b>	252	59766	6667	udp	-	21728.335578		<b>12505</b>	251	40434	8886	tcp	ssl	18761.401291		<b>12420</b>	166	62366	6667	udp	-	17747.121108		<b>12419</b>	165	40261	8886	tcp	ssl	17732.696970		<b>12317</b>	63	62969	6667	udp	-	9433.886888		<b>12285</b>	31	62969	6667	udp	-	9433.886888		<b>12316</b>	62	40545	8886	tcp	-	9379.541767		<b>12284</b>	30	40545	8886	tcp	-	9379.541767		<b>12353</b>	99	40533	8886	tcp	-	6950.994837		<b>12342</b>	88	62969	6667	udp	-	6828.456767		<b>12438</b>	184	40856	8886	tcp	ssl	5678.531132		<b>12437</b>	183	57069	6667	udp	-	5676.684995		<b>20154</b>	7647	39242	443	tcp	ssl	5341.392332		<b>20152</b>	7645	44058	443	tcp	ssl	5340.386892		<b>12468</b>	214	40140	8886	tcp	ssl	2943.393571		<b>15208</b>	2701	46602	443	tcp	ssl	2379.349634		<b>15183</b>	2676	37242	443	tcp	ssl	2083.013699		<b>14995</b>	2488	59400	443	tcp	ssl	1401.219948		<b>121106</b>	2589	3	3	icmp	-	905.964201		<b>14908</b>	2401	33802	443	tcp	-	900.887802	
	no	id.orig_p	id.resp_p	proto	service	flow_duration	fu																																																																																																																																																																			
<b>12506</b>	252	59766	6667	udp	-	21728.335578																																																																																																																																																																				
<b>12505</b>	251	40434	8886	tcp	ssl	18761.401291																																																																																																																																																																				
<b>12420</b>	166	62366	6667	udp	-	17747.121108																																																																																																																																																																				
<b>12419</b>	165	40261	8886	tcp	ssl	17732.696970																																																																																																																																																																				
<b>12317</b>	63	62969	6667	udp	-	9433.886888																																																																																																																																																																				
<b>12285</b>	31	62969	6667	udp	-	9433.886888																																																																																																																																																																				
<b>12316</b>	62	40545	8886	tcp	-	9379.541767																																																																																																																																																																				
<b>12284</b>	30	40545	8886	tcp	-	9379.541767																																																																																																																																																																				
<b>12353</b>	99	40533	8886	tcp	-	6950.994837																																																																																																																																																																				
<b>12342</b>	88	62969	6667	udp	-	6828.456767																																																																																																																																																																				
<b>12438</b>	184	40856	8886	tcp	ssl	5678.531132																																																																																																																																																																				
<b>12437</b>	183	57069	6667	udp	-	5676.684995																																																																																																																																																																				
<b>20154</b>	7647	39242	443	tcp	ssl	5341.392332																																																																																																																																																																				
<b>20152</b>	7645	44058	443	tcp	ssl	5340.386892																																																																																																																																																																				
<b>12468</b>	214	40140	8886	tcp	ssl	2943.393571																																																																																																																																																																				
<b>15208</b>	2701	46602	443	tcp	ssl	2379.349634																																																																																																																																																																				
<b>15183</b>	2676	37242	443	tcp	ssl	2083.013699																																																																																																																																																																				
<b>14995</b>	2488	59400	443	tcp	ssl	1401.219948																																																																																																																																																																				
<b>121106</b>	2589	3	3	icmp	-	905.964201																																																																																																																																																																				
<b>14908</b>	2401	33802	443	tcp	-	900.887802																																																																																																																																																																				

Load

Summary statistics of the dataset

IoT.describe()

:

IoT.describe()

	no	id.orig_p	id.resp_p	flow_duration	fwd_pkts_tot	bwd_pkts_tot	fwd_data_pkts_tot	bwd_data_pkts_tot	fwd_pi
count	123117.000000	123117.000000	123117.000000	123117.000000	123117.000000	123117.000000	123117.000000	123117.000000	1.2
mean	37035.089248	34639.258738	1014.305092	3.809566	2.268826	1.909509	1.471218	0.820260	3.5
std	30459.106367	19070.620354	5256.371994	130.005408	22.336565	33.018311	19.635196	32.293948	3.7
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
25%	6059.000000	17702.000000	21.000000	0.000001	1.000000	1.000000	1.000000	0.000000	7.4
50%	33100.000000	37221.000000	21.000000	0.000004	1.000000	1.000000	1.000000	0.000000	2.4
75%	63879.000000	50971.000000	21.000000	0.000005	1.000000	1.000000	1.000000	0.000000	5.2
max	94658.000000	65535.000000	65389.000000	21728.335578	4345.000000	10112.000000	4345.000000	10105.000000	1.0

8 rows × 82 columns

The activities performed aim to analyze the type of attacks that happen to IoT devices based on ‘flow\_duration’. To extract information about the contents of the dataset, the number of observations were taken, reading the first and last 10 observations, then crucially, identifying the data types and the values within the ‘Attack\_type’ column.

For the intended purpose, the dataset needs little transformation. As such, sorting the dataframe in descending order based on flow\_duration is enough.

In analyzing the dataset, summary statistics were taken, and in terms of flow\_duration, the standard deviation was considerably high, suggesting that a lot of the values are very far from the mean, which was evident just by looking at the highest and lowest records in terms of flow\_duration.

	no	id.orig_p	id.resp_p	proto	service	flow_duration
12506	252	59766	6667	udp	-	21728.335578
12505	251	40434	8886	tcp	ssl	18761.401291
12420	166	62366	6667	udp	-	17747.121108
12419	165	40261	8886	tcp	ssl	17732.696970
12317	63	62969	6667	udp	-	9433.886888
12285	31	62969	6667	udp	-	9433.886888
12316	62	40545	8886	tcp	-	9379.541767
12284	30	40545	8886	tcp	-	9379.541767
12353	99	40533	8886	tcp	-	6950.994837
12342	88	62969	6667	udp	-	6828.456767
12438	184	40856	8886	tcp	ssl	5678.531132
12437	183	57069	6667	udp	-	5676.684995
20154	7647	39242	443	tcp	ssl	5341.392332
20152	7645	44058	443	tcp	ssl	5340.386892
12468	214	40140	8886	tcp	ssl	2943.393571
15208	2701	46602	443	tcp	ssl	2379.349634
15183	2676	37242	443	tcp	ssl	2083.013699
14995	2488	59400	443	tcp	ssl	1401.219948
121106	2589	3	3	icmp	-	905.964201
14908	2401	33802	443	tcp	-	900.887802

Highest records in flow\_duration

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

#### Lowest records in flow\_duration

The quantile values for flow\_duration also suggest that most records in the dataset may have values hovering around those quantiles, which may have affected the mean to be as low as 3.809566.