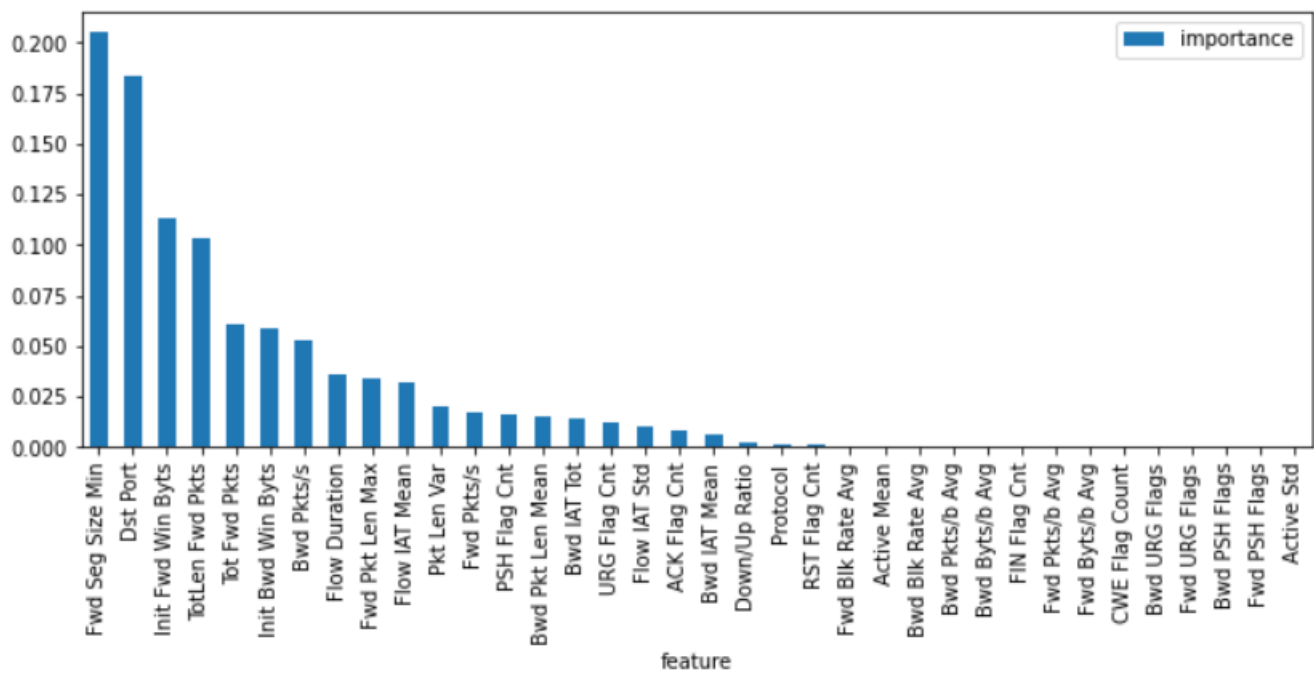


# Minimum Viable Product (MVP)

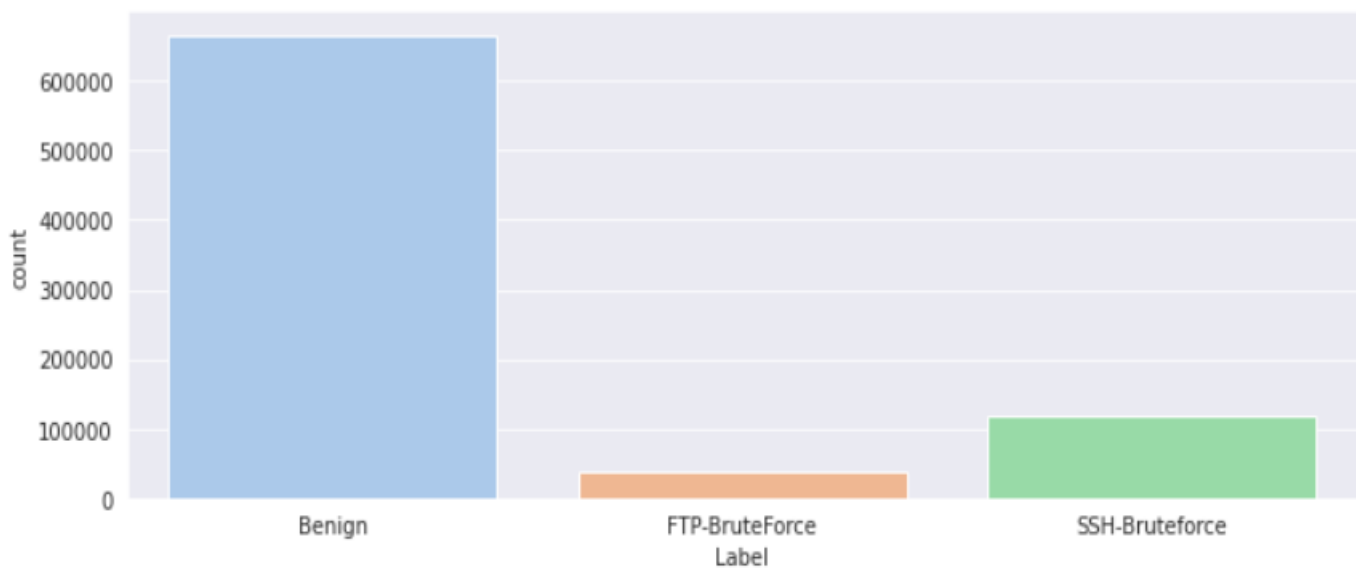


## Most Important Features





## Visualizing the distribution of the Target





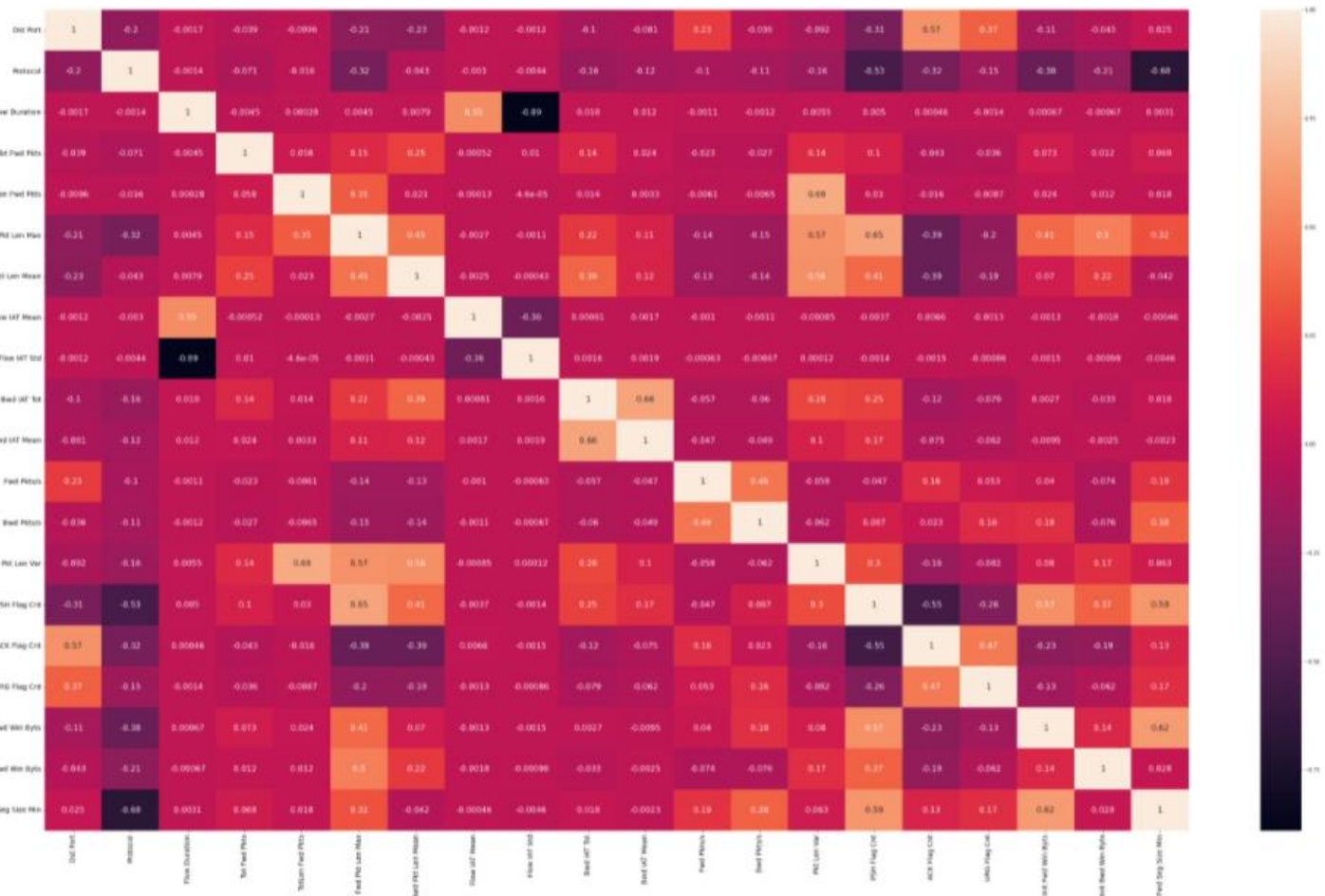
## Display Correlations between features

	Dst Port	Protocol	Flow Duration	Tot Fwd Pkts	Tot Bwd Pkts	TotLen Fwd Pkts	TotLen Bwd Pkts	Fwd Pkt Len Max	Fwd Pkt Len Min	Fwd Pkt Len Mean	...	Fwd Act Data Pkts	Fwd Seg Size Min	Active Mean	Acti S
Dst Port	1.000000	-0.202655	-0.001697	-0.039387	-0.024921	-0.009621	-0.012106	-0.205624	-0.177340	-0.240438	...	-0.188173	0.024604	-0.012843	-0.0387
Protocol	-0.202655	1.000000	-0.001422	-0.071419	-0.035996	-0.016018	-0.018136	-0.319890	0.839669	-0.059431	...	-0.300379	-0.677461	-0.053187	-0.0646
Flow Duration	-0.001697	-0.001422	1.000000	-0.004514	0.002075	0.000276	0.001668	0.004462	-0.002425	0.002254	...	0.006278	0.003057	0.003623	0.0070
Tot Fwd Pkts	-0.039387	-0.071419	-0.004514	1.000000	0.884620	0.058497	0.876916	0.145987	-0.063392	0.069405	...	0.247404	0.069493	0.087864	0.0946
Tot Bwd Pkts	-0.024921	-0.035996	0.002075	0.884620	1.000000	0.060777	0.996948	0.085757	-0.032784	0.032500	...	0.173029	0.030254	0.095460	0.0833
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Active Min	-0.005269	-0.041386	0.002297	0.069989	0.080108	0.006465	0.074444	0.050540	-0.033344	0.051298	...	0.061586	0.005230	0.972349	0.0327
Idle Mean	-0.001950	-0.006798	-0.741638	0.012485	0.001152	0.000030	0.001080	-0.000533	-0.002527	-0.001803	...	0.000958	-0.005986	0.003252	0.0038
Idle Std	-0.000773	-0.003511	-0.733111	0.011327	-0.000036	-0.000047	0.000025	-0.001240	-0.001042	-0.001577	...	-0.000882	-0.004072	0.000456	0.0005
Idle Max	-0.001168	-0.004615	-0.727980	0.011746	0.000337	-0.000026	0.000359	-0.001049	-0.001525	-0.001705	...	-0.000334	-0.004746	0.001185	0.0013
Idle Min	-0.035799	-0.101693	-0.672193	0.044952	0.036332	0.002379	0.032277	0.020813	-0.045734	-0.006544	...	0.055314	-0.060205	0.092014	0.1077

76 rows × 76 columns



# Display Correlations between selected features





## Key Finding After Data Visualization at this point:

- ❖ The highly most Important feature is (Fwd seg Size Min) , which refer to Minimum segment size observed in the forward direction
- ❖ Benign packets are more than the two malicious packets
- ❖ There is a high correlation between some columns, and this may cause a problem so, the columns with a correlation higher than 0.7 have been deleted

After that I will apply the classification model, and I will try more than one algorithm