

```
D:\SentinelNet\scripts>python main.py
```

NO OF ROWS:

125972

NO OF UNIQUE LABELS:

```
['normal' 'neptune' 'warezclient' 'ipsweep' 'portsweep' 'teardrop' 'nmap'  
'satan' 'smurf' 'pod' 'back' 'guess_passwd' 'ftp_write' 'multihop'  
'rootkit' 'buffer_overflow' 'imap' 'warezmaster' 'phf' 'land'  
'loadmodule' 'spy' 'perl']
```

Top 5 Frequent attack types:

outcome

normal 67342

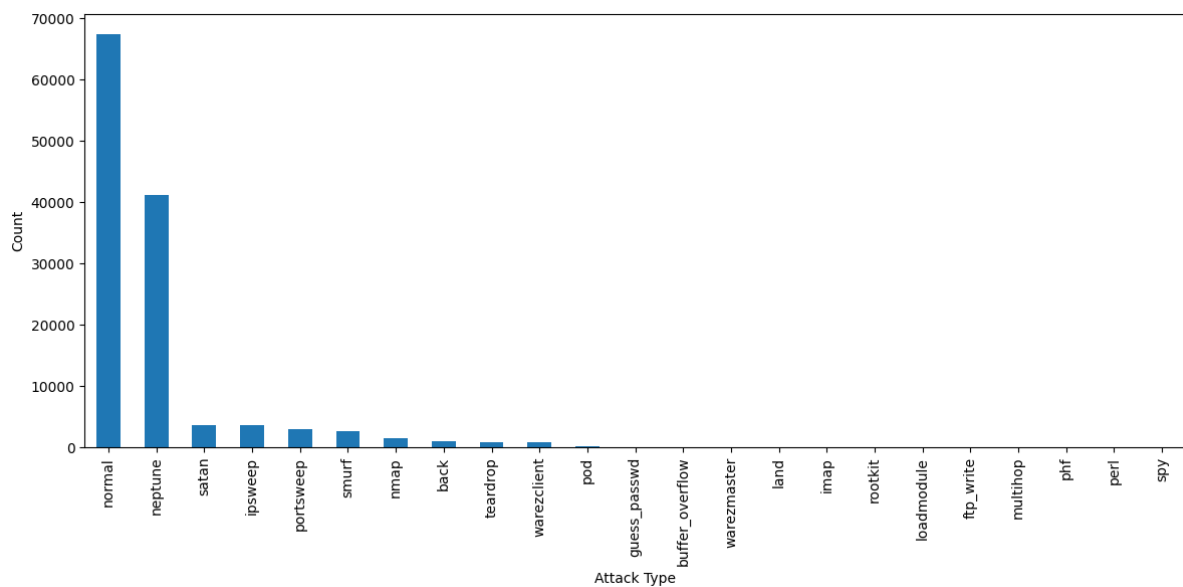
neptune 41214

satan 3633

ipsweep 3599

portsweep 2931

Name: count, dtype: int64



First 5 rows:

```
duration protocol_type service flag src_bytes dst_bytes ... dst_host_serror_rate  
dst_host_srv_serror_rate dst_host_error_rate dst_host_srv_error_rate outcome level  
0 0 udp other SF 146 0 ... 0.00 0.00 0.0  
0.00 normal 15
```

1	0	tcp	private	S0	0	0 ...	1.00	1.00	0.0
0.00 neptune 19									
2	0	tcp	http	SF	232	8153 ...	0.03	0.01	0.0
0.01 normal 21									
3	0	tcp	http	SF	199	420 ...	0.00	0.00	0.0
0.00 normal 21									
4	0	tcp	private	REJ	0	0 ...	0.00	0.00	1.0
1.00 neptune 21									

[5 rows x 43 columns]

Last 5 rows:

	duration	protocol_type	service	flag	src_bytes	...	dst_host	srv_serror_rate	dst_host_rerror_rate	dst_host_srv_rerror_rate	outcome	level
125967	0	tcp	private	S0	0 ...		1.0	0.00			0.0	
neptune 20												
125968	8	udp	private	SF	105 ...		0.0	0.00			0.0	
normal 21												
125969	0	tcp	smtp	SF	2231 ...		0.0	0.01			0.0	
normal 18												
125970	0	tcp	klogin	S0	0 ...		1.0	0.00			0.0	neptune
20												
125971	0	tcp	ftp_data	SF	151 ...		0.0	0.00			0.0	
normal 21												

[5 rows x 43 columns]

Structure of dataset

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 125972 entries, 0 to 125971

Data columns (total 43 columns):

#	Column	Non-Null Count	Dtype
0	duration	125972 non-null	int64
1	protocol_type	125972 non-null	object
2	service	125972 non-null	object

3	flag	125972 non-null object
4	src_bytes	125972 non-null int64
5	dst_bytes	125972 non-null int64
6	land	125972 non-null int64
7	wrong_fragment	125972 non-null int64
8	urgent	125972 non-null int64
9	hot	125972 non-null int64
10	num_failed_logins	125972 non-null int64
11	logged_in	125972 non-null int64
12	num_compromised	125972 non-null int64
13	root_shell	125972 non-null int64
14	su_attempted	125972 non-null int64
15	num_root	125972 non-null int64
16	num_file_creations	125972 non-null int64
17	num_shells	125972 non-null int64
18	num_access_files	125972 non-null int64
19	num_outbound_cmds	125972 non-null int64
20	is_host_login	125972 non-null int64
21	is_guest_login	125972 non-null int64
22	count	125972 non-null int64
23	srv_count	125972 non-null int64
24	serror_rate	125972 non-null float64
25	srv_serror_rate	125972 non-null float64
26	error_rate	125972 non-null float64
27	srv_error_rate	125972 non-null float64
28	same_srv_rate	125972 non-null float64
29	diff_srv_rate	125972 non-null float64
30	srv_diff_host_rate	125972 non-null float64
31	dst_host_count	125972 non-null int64
32	dst_host_srv_count	125972 non-null int64
33	dst_host_same_srv_rate	125972 non-null float64

34 dst_host_diff_srv_rate 125972 non-null float64
35 dst_host_same_src_port_rate 125972 non-null float64
36 dst_host_srv_diff_host_rate 125972 non-null float64
37 dst_host_serror_rate 125972 non-null float64
38 dst_host_srv_serror_rate 125972 non-null float64
39 dst_host_rerror_rate 125972 non-null float64
40 dst_host_srv_rerror_rate 125972 non-null float64
41 outcome 125972 non-null object
42 level 125972 non-null int64

dtypes: float64(15), int64(24), object(4)

memory usage: 41.3+ MB

None

Statistical summary

	duration	src_bytes	dst_bytes	land ...	dst_host_srv_serror_rate
	dst_host_rerror_rate	dst_host_srv_rerror_rate	level		
count	125972.000000	1.259720e+05	1.259720e+05	125972.000000 ...	125972.000000
	125972.000000	125972.000000	125972.000000		
mean	287.146929	4.556710e+04	1.977927e+04	0.000198 ...	0.278487
	0.118832	0.120241	19.504056		
std	2604.525522	5.870354e+06	4.021285e+06	0.014086 ...	0.445670
	0.306559	0.319460	2.291512		
min	0.000000	0.000000e+00	0.000000e+00	0.000000 ...	0.000000
	0.000000	0.000000	0.000000		
25%	0.000000	0.000000e+00	0.000000e+00	0.000000 ...	0.000000
	0.000000	0.000000	18.000000		
50%	0.000000	4.400000e+01	0.000000e+00	0.000000 ...	0.000000
	0.000000	0.000000	20.000000		
75%	0.000000	2.760000e+02	5.160000e+02	0.000000 ...	1.000000
	0.000000	0.000000	21.000000		
max	42908.000000	1.379964e+09	1.309937e+09	1.000000 ...	1.000000
	1.000000	1.000000	21.000000		

[8 rows x 39 columns]

Feature types:

duration	int64
protocol_type	object
service	object
flag	object
src_bytes	int64
dst_bytes	int64
land	int64
wrong_fragment	int64
urgent	int64
hot	int64
num_failed_logins	int64
logged_in	int64
num_compromised	int64
root_shell	int64
su_attempted	int64
num_root	int64
num_file_creations	int64
num_shells	int64
num_access_files	int64
num_outbound_cmds	int64
is_host_login	int64
is_guest_login	int64
count	int64
srv_count	int64
serror_rate	float64
srv_serror_rate	float64
rerror_rate	float64
srv_rerror_rate	float64
same_srv_rate	float64
diff_srv_rate	float64
srv_diff_host_rate	float64

```
dst_host_count          int64
dst_host_srv_count      int64
dst_host_same_srv_rate  float64
dst_host_diff_srv_rate  float64
dst_host_same_src_port_rate float64
dst_host_srv_diff_host_rate float64
dst_host_serror_rate    float64
dst_host_srv_serror_rate float64
dst_host_rerror_rate    float64
dst_host_srv_rerror_rate float64
outcome                 object
level                   int64
dtype: object
```

Attack Category Distribution:

category

Normal 67342

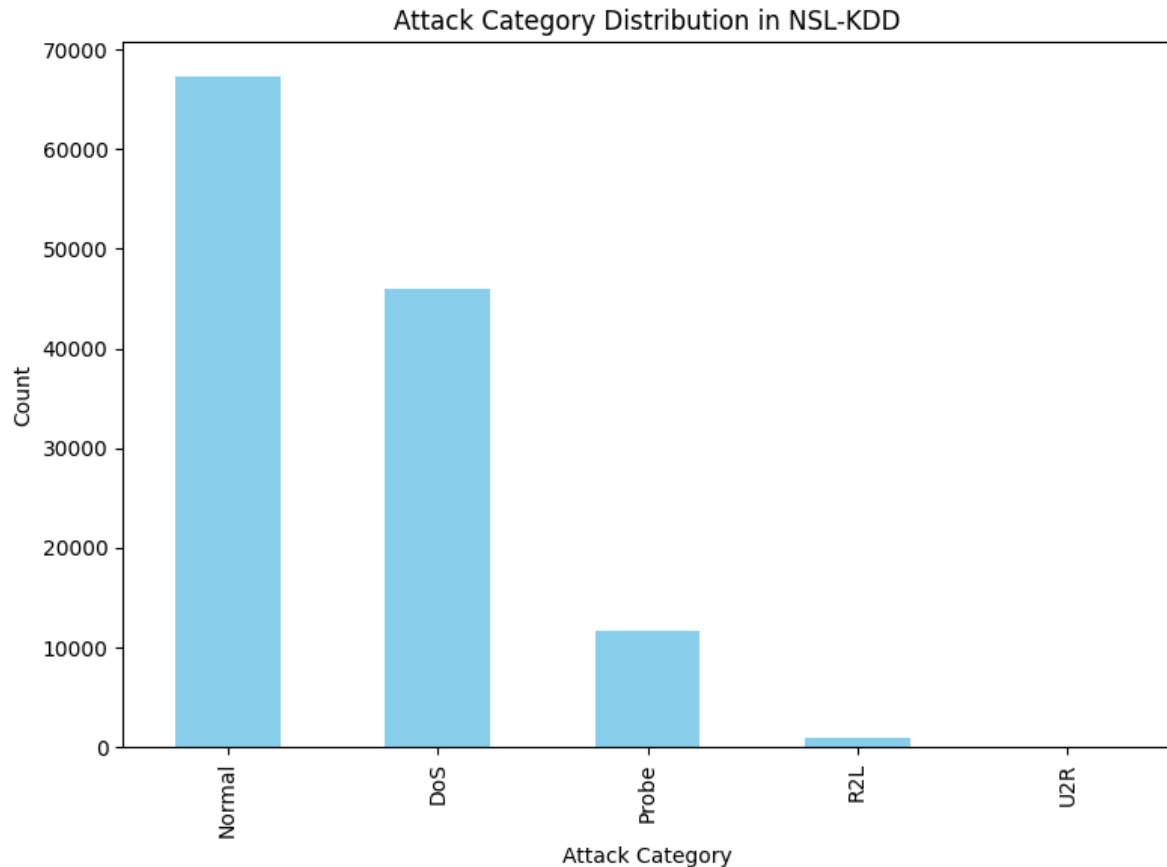
DoS 45927

Probe 11656

R2L 995

U2R 52

Name: count, dtype: int64



Categorical Features: ['protocol_type', 'service', 'flag', 'outcome', 'category']

Numerical Features: ['duration', 'src_bytes', 'dst_bytes', 'land', 'wrong_fragment', 'urgent', 'hot', 'num_failed_logins', 'logged_in', 'num_compromised', 'root_shell', 'su_attempted', 'num_root', 'num_file_creations', 'num_shells', 'num_access_files', 'num_outbound_cmds', 'is_host_login', 'is_guest_login', 'count', 'srv_count', 'serror_rate', 'srv_serror_rate', 'error_rate', 'srv_error_rate', 'same_srv_rate', 'diff_srv_rate', 'srv_diff_host_rate', 'dst_host_count', 'dst_host_srv_count', 'dst_host_same_srv_rate', 'dst_host_diff_srv_rate', 'dst_host_same_src_port_rate', 'dst_host_srv_diff_host_rate', 'dst_host_serror_rate', 'dst_host_srv_serror_rate', 'dst_host_error_rate', 'dst_host_srv_error_rate', 'level']

Missing values per column:

duration	0
protocol_type	0
service	0
flag	0
src_bytes	0
dst_bytes	0
land	0

wrong_fragment	0
urgent	0
hot	0
num_failed_logins	0
logged_in	0
num_compromised	0
root_shell	0
su_attempted	0
num_root	0
num_file_creations	0
num_shells	0
num_access_files	0
num_outbound_cmds	0
is_host_login	0
is_guest_login	0
count	0
srv_count	0
serror_rate	0
srv_serror_rate	0
rerror_rate	0
srv_rerror_rate	0
same_srv_rate	0
diff_srv_rate	0
srv_diff_host_rate	0
dst_host_count	0
dst_host_srv_count	0
dst_host_same_srv_rate	0
dst_host_diff_srv_rate	0
dst_host_same_src_port_rate	0
dst_host_srv_diff_host_rate	0
dst_host_serror_rate	0


```
dst_host_srv_error_rate    0
dst_host_error_rate        0
dst_host_srv_error_rate    0
outcome                    0
level                      0
category                   0
dtype: int64
```

Number of duplicate rows: 0

	Feature	Type	Preprocessing Semantic
0	duration	Numerical	Normalize/Standardize Feature
1	protocol_type	Categorical	One-hot encode / Label encode Feature
2	service	Categorical	One-hot encode / Label encode Feature
3	flag	Categorical	One-hot encode / Label encode Feature
4	src_bytes	Numerical	Log-transform + Normalize Feature
5	dst_bytes	Numerical	Log-transform + Normalize Feature
6	land	Binary	Keep as is (0/1) Feature
7	wrong_fragment	Numerical	Normalize/Standardize Feature
8	urgent	Numerical	Normalize/Standardize Feature
9	hot	Numerical	Normalize/Standardize Feature
10	num_failed_logins	Numerical	Normalize/Standardize Feature
11	logged_in	Binary	Keep as is (0/1) Feature
12	num_compromised	Numerical	Normalize/Standardize Feature
13	root_shell	Binary	Keep as is (0/1) Feature
14	su_attempted	Binary	Keep as is (0/1) Feature
15	num_root	Numerical	Normalize/Standardize Feature
16	num_file_creations	Numerical	Normalize/Standardize Feature
17	num_shells	Numerical	Normalize/Standardize Feature
18	num_access_files	Numerical	Normalize/Standardize Feature
19	num_outbound_cmds	Numerical	Drop Feature
20	is_host_login	Binary	Drop Feature
21	is_guest_login	Binary	Keep as is (0/1) Feature

22	count	Numerical	Normalize/Standardize	Feature
23	srv_count	Numerical	Normalize/Standardize	Feature
24	serror_rate	Numerical	Already normalized	Feature
25	srv_serror_rate	Numerical	Already normalized	Feature
26	rerror_rate	Numerical	Already normalized	Feature
27	srv_rerror_rate	Numerical	Already normalized	Feature
28	same_srv_rate	Numerical	Already normalized	Feature
29	diff_srv_rate	Numerical	Already normalized	Feature
30	srv_diff_host_rate	Numerical	Already normalized	Feature
31	dst_host_count	Numerical	Normalize/Standardize	Feature
32	dst_host_srv_count	Numerical	Normalize/Standardize	Feature
33	dst_host_same_srv_rate	Numerical	Already normalized	Feature
34	dst_host_diff_srv_rate	Numerical	Already normalized	Feature
35	dst_host_same_src_port_rate	Numerical	Already normalized	Feature
36	dst_host_srv_diff_host_rate	Numerical	Already normalized	Feature
37	dst_host_serror_rate	Numerical	Already normalized	Feature
38	dst_host_srv_serror_rate	Numerical	Already normalized	Feature
39	dst_host_rerror_rate	Numerical	Already normalized	Feature
40	dst_host_srv_rerror_rate	Numerical	Already normalized	Feature
41	outcome	Categorical	Label encode	Label
42	level	Numerical	Drop	Feature
43	category	Categorical	Label encode	Label

Missing values handled.

Duplicates removed. Current shape: (125972, 44)

Categorical features encoded. Current shape: (125972, 122)

Numerical features scaled.