

```
In [1]: import pandas as pd
d = pd.read_csv(r"D:\New folder (4)\log2.csv")
```

```
In [2]: m = d.isnull().sum()
print("Missing Values:\n", m)
```

```
Missing Values:
Source Port          0
Destination Port     0
NAT Source Port      0
NAT Destination Port 0
Action              0
Bytes               0
Bytes Sent           0
Bytes Received        0
Packets             0
Elapsed Time (sec)   0
pkts_sent            0
pkts_received        0
dtype: int64
```

```
In [3]: d.fillna(0, inplace=True)
print("Updated Dataset:\n", d.head())
```

```
Updated Dataset:
Source Port  Destination Port  NAT Source Port  NAT Destination Port  \
0          57222              53          54587              53
1          56258             3389          56258             3389
2           6881            50321          43265            50321
3          50553             3389          50553             3389
4          50002             443          45848             443

Action  Bytes  Bytes Sent  Bytes Received  Packets  Elapsed Time (sec)  \
0  allow   177         94           83         2           30
1  allow  4768        1600          3168        19           17
2  allow   238         118           120         2          1199
3  allow  3327        1438          1889        15           17
4  allow 25358        6778         18580        31           16

pkts_sent  pkts_received
0          1             1
1         10             9
2          1             1
3          8             7
4         13            18
```

```
In [4]: d_dummies = pd.get_dummies(d, drop_first=True)
print("Dataset with Dummy Variables:\n", d_dummies.head())
```

Dataset with Dummy Variables:

| | Source Port | Destination Port | NAT Source Port | NAT Destination Port | \ |
|---|-------------|------------------|-----------------|----------------------|---|
| 0 | 57222 | 53 | 54587 | 53 | |
| 1 | 56258 | 3389 | 56258 | 3389 | |
| 2 | 6881 | 50321 | 43265 | 50321 | |
| 3 | 50553 | 3389 | 50553 | 3389 | |
| 4 | 50002 | 443 | 45848 | 443 | |

| | Bytes | Bytes Sent | Bytes Received | Packets | Elapsed Time (sec) | pkts_sent | \ |
|---|-------|------------|----------------|---------|--------------------|-----------|---|
| 0 | 177 | 94 | 83 | 2 | 30 | 1 | |
| 1 | 4768 | 1600 | 3168 | 19 | 17 | 10 | |
| 2 | 238 | 118 | 120 | 2 | 1199 | 1 | |
| 3 | 3327 | 1438 | 1889 | 15 | 17 | 8 | |
| 4 | 25358 | 6778 | 18580 | 31 | 16 | 13 | |

| | pkts_received | Action_deny | Action_drop | Action_reset-both |
|---|---------------|-------------|-------------|-------------------|
| 0 | 1 | 0 | 0 | 0 |
| 1 | 9 | 0 | 0 | 0 |
| 2 | 1 | 0 | 0 | 0 |
| 3 | 7 | 0 | 0 | 0 |
| 4 | 18 | 0 | 0 | 0 |

```
In [5]: s = d[d['Action'] == 'allow']
print("Subset of Rows where Action is 'allow':\n", s.head())
```

Subset of Rows where Action is 'allow':

| | Source Port | Destination Port | NAT Source Port | NAT Destination Port | \ |
|---|-------------|------------------|-----------------|----------------------|---|
| 0 | 57222 | 53 | 54587 | 53 | |
| 1 | 56258 | 3389 | 56258 | 3389 | |
| 2 | 6881 | 50321 | 43265 | 50321 | |
| 3 | 50553 | 3389 | 50553 | 3389 | |
| 4 | 50002 | 443 | 45848 | 443 | |

| | Action | Bytes | Bytes Sent | Bytes Received | Packets | Elapsed Time (sec) | \ |
|---|--------|-------|------------|----------------|---------|--------------------|---|
| 0 | allow | 177 | 94 | 83 | 2 | 30 | |
| 1 | allow | 4768 | 1600 | 3168 | 19 | 17 | |
| 2 | allow | 238 | 118 | 120 | 2 | 1199 | |
| 3 | allow | 3327 | 1438 | 1889 | 15 | 17 | |
| 4 | allow | 25358 | 6778 | 18580 | 31 | 16 | |

| | pkts_sent | pkts_received |
|---|-----------|---------------|
| 0 | 1 | 1 |
| 1 | 10 | 9 |
| 2 | 1 | 1 |
| 3 | 8 | 7 |
| 4 | 13 | 18 |

```
In [6]: samp = d.sample(n=5)
print("Random Sample:\n", samp)
```

Random Sample:

| | Source Port | Destination Port | NAT Source Port | NAT Destination Port | \ |
|-------|-------------|------------------|-----------------|----------------------|---|
| 54325 | 57904 | 445 | 0 | 0 | |
| 52986 | 1049 | 54369 | 0 | 0 | |
| 29297 | 34357 | 443 | 37129 | 443 | |
| 49734 | 54776 | 445 | 0 | 0 | |
| 42998 | 62408 | 3478 | 62290 | 3478 | |

| | Action | Bytes | Bytes Sent | Bytes Received | Packets | Elapsed Time (sec) | \ |
|-------|--------|-------|------------|----------------|---------|--------------------|---|
| 54325 | drop | 70 | 70 | 0 | 1 | 0 | |
| 52986 | deny | 62 | 62 | 0 | 1 | 0 | |
| 29297 | allow | 17372 | 10352 | 7020 | 91 | 422 | |
| 49734 | drop | 66 | 66 | 0 | 1 | 0 | |
| 42998 | allow | 140 | 140 | 0 | 1 | 30 | |

| | pkts_sent | pkts_received |
|-------|-----------|---------------|
| 54325 | 1 | 0 |
| 52986 | 1 | 0 |
| 29297 | 52 | 39 |
| 49734 | 1 | 0 |
| 42998 | 1 | 0 |

```
In [7]: g = d.groupby('Action')['Bytes'].sum().reset_index()
print("Grouped Data by Action:\n", g)
```

Grouped Data by Action:

| | Action | Bytes |
|---|------------|------------|
| 0 | allow | 6362588560 |
| 1 | deny | 1247094 |
| 2 | drop | 882546 |
| 3 | reset-both | 8497 |

```
In [8]: t = d.sample(frac=0.8, random_state=1)
te = d.drop(t.index)
print("Training Set Shape:", t.shape)
print("Test Set Shape:", te.shape)
```

Training Set Shape: (52426, 12)

Test Set Shape: (13106, 12)

```
In [9]: d2 = pd.DataFrame({
'Source Port': [80, 443],
'Protocol': ['HTTP', 'HTTPS']
})
m_df = pd.merge(d, d2, on='Source Port', how='left')
print("Merged DataFrame:\n", m_df.head())
```

Merged DataFrame:

| | Source Port | Destination Port | NAT Source Port | NAT Destination Port | \ |
|---|-------------|------------------|-----------------|----------------------|---|
| 0 | 57222 | 53 | 54587 | 53 | |
| 1 | 56258 | 3389 | 56258 | 3389 | |
| 2 | 6881 | 50321 | 43265 | 50321 | |
| 3 | 50553 | 3389 | 50553 | 3389 | |
| 4 | 50002 | 443 | 45848 | 443 | |

| | Action | Bytes | Bytes Sent | Bytes Received | Packets | Elapsed Time (sec) | \ |
|---|--------|-------|------------|----------------|---------|--------------------|---|
| 0 | allow | 177 | 94 | 83 | 2 | 30 | |
| 1 | allow | 4768 | 1600 | 3168 | 19 | 17 | |
| 2 | allow | 238 | 118 | 120 | 2 | 1199 | |
| 3 | allow | 3327 | 1438 | 1889 | 15 | 17 | |
| 4 | allow | 25358 | 6778 | 18580 | 31 | 16 | |

| | pkts_sent | pkts_received | Protocol |
|---|-----------|---------------|----------|
| 0 | 1 | 1 | NaN |
| 1 | 10 | 9 | NaN |
| 2 | 1 | 1 | NaN |
| 3 | 8 | 7 | NaN |
| 4 | 13 | 18 | NaN |

```
In [10]: ij = pd.merge(d, d2, on='Source Port', how='inner')
lj = pd.merge(d, d2, on='Source Port', how='left')
rj = pd.merge(d, d2, on='Source Port', how='right')
oj = pd.merge(d, d2, on='Source Port', how='outer')
print("Inner Join:\n", ij.head())
print("Left Join:\n", lj.head())
print("Right Join:\n", rj.head())
print("Outer Join:\n", oj.head())
```

Inner Join:

| | Source Port | Destination Port | NAT Source Port | NAT Destination Port | \ |
|---|-------------|------------------|-----------------|----------------------|---|
| 0 | 443 | 47582 | 0 | 0 | |
| 1 | 443 | 47582 | 0 | 0 | |
| 2 | 443 | 47582 | 0 | 0 | |
| 3 | 443 | 49988 | 0 | 0 | |
| 4 | 443 | 18368 | 0 | 0 | |

| | Action | Bytes | Bytes Sent | Bytes Received | Packets | Elapsed Time (sec) | \ |
|---|--------|-------|------------|----------------|---------|--------------------|---|
| 0 | deny | 60 | 60 | 0 | 1 | 0 | |
| 1 | deny | 85 | 85 | 0 | 1 | 0 | |
| 2 | deny | 100 | 100 | 0 | 1 | 0 | |
| 3 | deny | 82 | 82 | 0 | 1 | 0 | |
| 4 | deny | 60 | 60 | 0 | 1 | 0 | |

| | pkts_sent | pkts_received | Protocol |
|---|-----------|---------------|----------|
| 0 | 1 | 0 | HTTPS |
| 1 | 1 | 0 | HTTPS |
| 2 | 1 | 0 | HTTPS |
| 3 | 1 | 0 | HTTPS |
| 4 | 1 | 0 | HTTPS |

Left Join:

| | Source Port | Destination Port | NAT Source Port | NAT Destination Port | \ |
|---|-------------|------------------|-----------------|----------------------|---|
| 0 | 57222 | 53 | 54587 | 53 | |
| 1 | 56258 | 3389 | 56258 | 3389 | |
| 2 | 6881 | 50321 | 43265 | 50321 | |
| 3 | 50553 | 3389 | 50553 | 3389 | |
| 4 | 50002 | 443 | 45848 | 443 | |

| | Action | Bytes | Bytes Sent | Bytes Received | Packets | Elapsed Time (sec) | \ |
|---|--------|-------|------------|----------------|---------|--------------------|---|
| 0 | allow | 177 | 94 | 83 | 2 | 30 | |
| 1 | allow | 4768 | 1600 | 3168 | 19 | 17 | |
| 2 | allow | 238 | 118 | 120 | 2 | 1199 | |
| 3 | allow | 3327 | 1438 | 1889 | 15 | 17 | |
| 4 | allow | 25358 | 6778 | 18580 | 31 | 16 | |

| | pkts_sent | pkts_received | Protocol |
|---|-----------|---------------|----------|
| 0 | 1 | 1 | NaN |
| 1 | 10 | 9 | NaN |
| 2 | 1 | 1 | NaN |
| 3 | 8 | 7 | NaN |
| 4 | 13 | 18 | NaN |

Right Join:

| | Source Port | Destination Port | NAT Source Port | NAT Destination Port | \ |
|---|-------------|------------------|-----------------|----------------------|---|
| 0 | 80 | 13987 | 0 | 0 | |
| 1 | 80 | 18811 | 0 | 0 | |
| 2 | 80 | 56289 | 0 | 0 | |
| 3 | 80 | 61486 | 0 | 0 | |
| 4 | 80 | 8562 | 0 | 0 | |

| | Action | Bytes | Bytes Sent | Bytes Received | Packets | Elapsed Time (sec) | \ |
|---|--------|-------|------------|----------------|---------|--------------------|---|
| 0 | deny | 66 | 66 | 0 | 1 | 0 | |
| 1 | deny | 60 | 60 | 0 | 1 | 0 | |
| 2 | deny | 578 | 578 | 0 | 1 | 0 | |
| 3 | deny | 66 | 66 | 0 | 1 | 0 | |
| 4 | deny | 74 | 74 | 0 | 1 | 0 | |

| | pkts_sent | pkts_received | Protocol |
|---|-----------|---------------|----------|
| 0 | 1 | 0 | HTTP |
| 1 | 1 | 0 | HTTP |
| 2 | 1 | 0 | HTTP |
| 3 | 1 | 0 | HTTP |
| 4 | 1 | 0 | HTTP |

Outer Join:

| | Source Port | Destination Port | NAT Source Port | NAT Destination Port | \ |
|---|-------------|------------------|-----------------|----------------------|---|
| 0 | 57222 | 53 | 54587 | 53 | |
| 1 | 57222 | 53 | 51448 | 53 | |
| 2 | 57222 | 445 | 0 | 0 | |
| 3 | 56258 | 3389 | 56258 | 3389 | |
| 4 | 56258 | 53 | 45112 | 53 | |

| | Action | Bytes | Bytes Sent | Bytes Received | Packets | Elapsed Time (sec) | \ |
|---|--------|-------|------------|----------------|---------|--------------------|---|
| 0 | allow | 177 | 94 | 83 | 2 | 30 | |
| 1 | allow | 177 | 94 | 83 | 2 | 30 | |
| 2 | drop | 66 | 66 | 0 | 1 | 0 | |
| 3 | allow | 4768 | 1600 | 3168 | 19 | 17 | |
| 4 | allow | 166 | 85 | 81 | 2 | 31 | |

| | pkts_sent | pkts_received | Protocol |
|---|-----------|---------------|----------|
| 0 | 1 | 1 | NaN |
| 1 | 1 | 1 | NaN |
| 2 | 1 | 0 | NaN |
| 3 | 10 | 9 | NaN |
| 4 | 1 | 1 | NaN |

In []: