|  |  |
| --- | --- |
| Pascual, Ken Leonard | 8 April 2024 |
| CPE22S3 | CPE311 – Quiz 2 |

|  |  |  |
| --- | --- | --- |
| Phase | Activity | Code and Output |
| Extract | Counting the number of observations | IoT.count() |
| Displaying the first 10 observations | IoT.head(10) |
| Displaying the last 10 observations | IoT.tail(10) |
| Displaying the data types of each column | IoT.dtypes |
| Identifying the Attack types within the dataset | IoT['Attack\_type'].values |
| Transform | Sorting the dataset by ‘flow\_duration’ | IoT = IoT.sort\_values(by='flow\_duration',ascending=False) |
| Load | Summary statistics of the dataset | IoT.describe() |

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.

A table of numbers and letters

AI-generated content may be incorrect.

Highest records in flow\_duration

A table with numbers and letters

AI-generated content may be incorrect.

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.