**CSE523 Machine Learning**

**Prof. Mehul Raval**

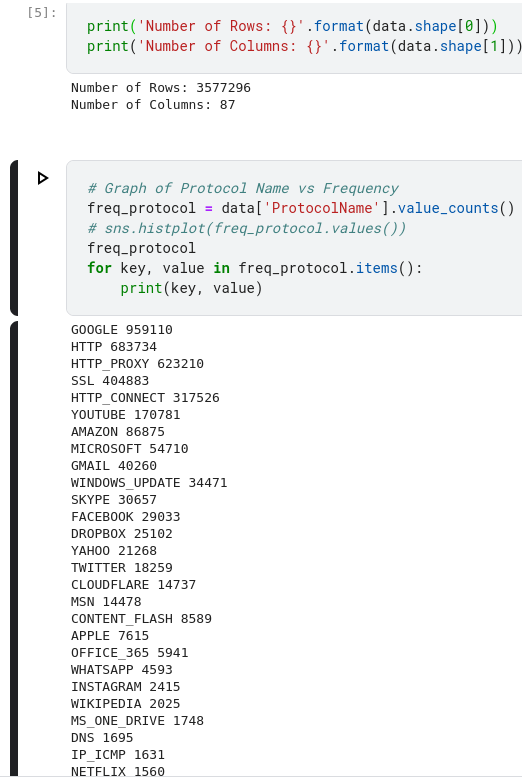
**Analyzing TCP Network Packets**

**Weekly Report**

| **Name** | **Enrolment Number** |
| --- | --- |
| Nityam Dixit | AU2040140 |
| Tej Thakar | AU2040262 |
| Yagnesh Patel | AU2040025 |
| Kavan Gondalia | AU2040030 |

Kaggle Link: <https://www.kaggle.com/code/kavangondalia/predict-protocol-name-with-ml>

1. Fetching the No. of different protocols and no. of rows of that protocol.



From the 78 classes, we will be selecting only those classes which have more than 10000 occurrences.

1. Fixing a minimum threshold as 10000 rows to get the protocol names which are used more.



1. Combining the features with almost the same correlation and removing the features which have correlation between 0.01 to -0.01.

