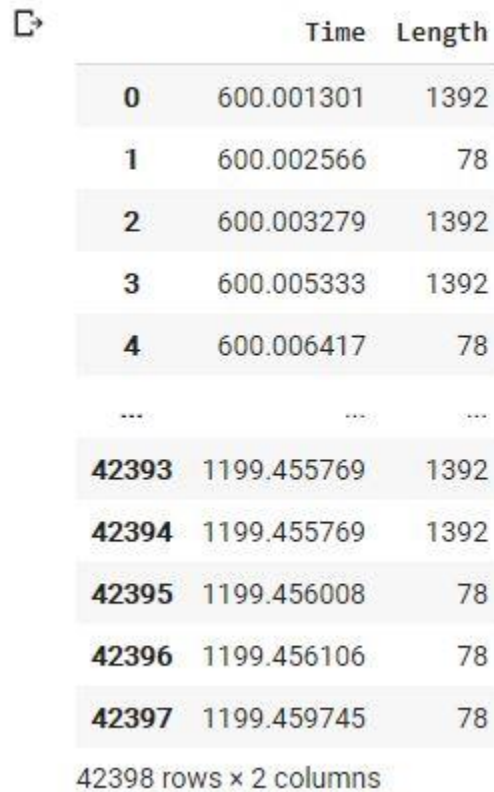


## Laporan Analisis Statistik Data Streaming Multimedia (Youtube)

1. Dari data capturing selama 30 menit, ambilah **sample** selama 10 menit, pada Assignment kali ini saya mengambil data dari 600.0013 sampai 1199.46 sekon karena waktu yang dibutuhkan adalah 10 menit.



The image shows a screenshot of a data table with two columns: 'Time' and 'Length'. The table contains 42398 rows. The first few rows are highlighted in light blue. The 'Time' column shows values ranging from 600.001301 to 1199.459745. The 'Length' column shows values of 1392 and 78. The table is truncated with ellipses in the middle. Below the table, it says '42398 rows x 2 columns'.

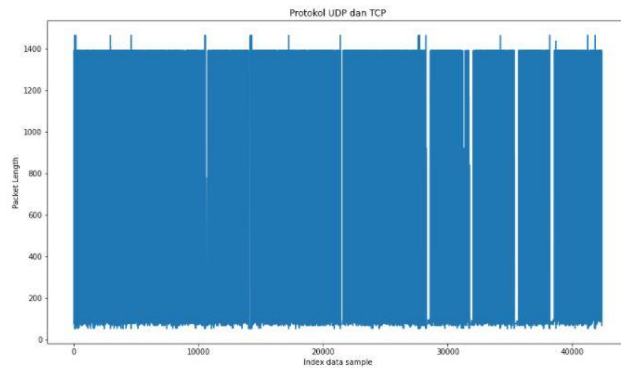
	Time	Length
0	600.001301	1392
1	600.002566	78
2	600.003279	1392
3	600.005333	1392
4	600.006417	78
...	...	...
42393	1199.455769	1392
42394	1199.455769	1392
42395	1199.456008	78
42396	1199.456106	78
42397	1199.459745	78

42398 rows x 2 columns

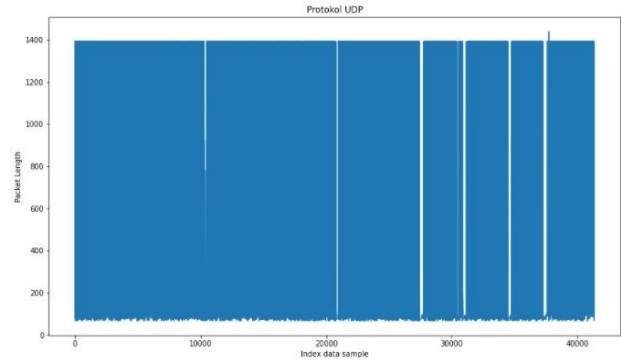
Gambar 1. Import data with pandas

Import dataset dengan library pandas untuk membaca file berformat .xlsx yang telah difilter hanya kolom protocol UDP dan TCP saja. Kolom yang akan digunakan adalah kolom Time dan Packet.

2. Buatlah plot dari data sample tersebut, dimana sumbu Y adalah packet length dan



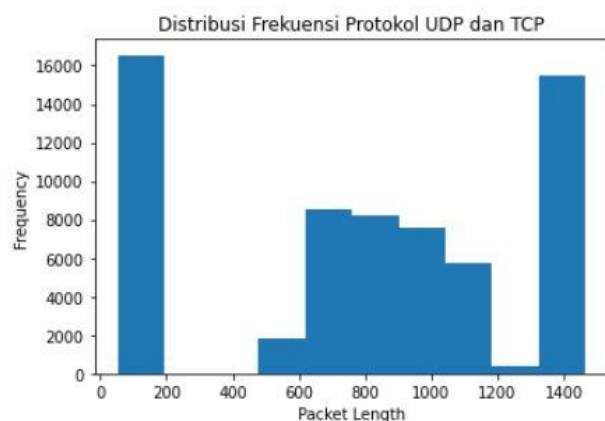
Gambar 3. Plot data sample pada protokol UDP dan TCP



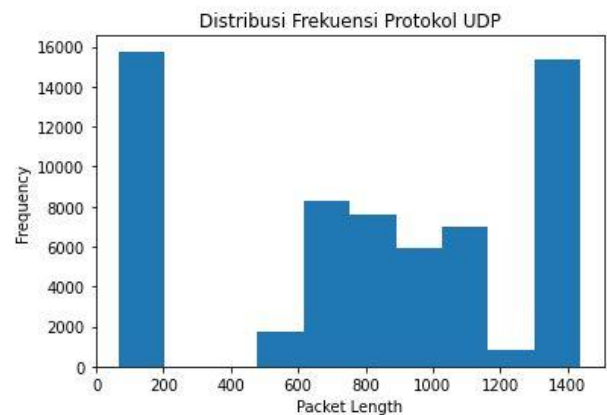
Gambar 2. Plot data Sample protokol UDP

dimana pada sumbu x adalah index data sample dan sumbu y adalah paket length.

3. Buatlah plot Distribusi Frekuensi dan lakukan analisis terhadap sample data traffic tersebut. Berdasarkan distribusi frekuensi tersebut, **packet length manakah yang cenderung memiliki frekuensi paling banyak?**



Gambar 4. Plot data Distribusi Frekuensi Protokol TCP dan UDP



Gambar 5. Plot data Distribusi Frekuensi Protokol UDP

Pada gambar tersebut dapat dilihat perbandingan protocol layer transport antara kedua paket tersebut. Sebelah kiri merupakan paket yang transport layer-nya menggunakan protocol TCP dan UDP, sedangkan gambar sebelah kanan menggunakan protocol UDP. Pada gambar tersebut terlihat jika packet length yang menggunakan protocol TCP/UDP lebih banyak frekuensinya dibandingkan dengan yang hanya menggunakan protocol UDP saja, walaupun data tidak jauh berbeda tetapi jika dilihat dari perbandingan antara frekuensi dengan packet Length lebih tinggi dari pada yang menggunakan protocol UDP saja dan dengan adanya protocol TCP grafik lebih stabil karena bisa digunakan untuk aktivitas secara bersamaan seperti memutar video streaming, sambal browsing dan lain-lain.

4. Berapakah **mean, variance serta standard deviasi** dari data sample anda (tuliskan persamaannya).

Sebelum menghitung terlebih dahulu saya mengimport library yang dibutuhkan yaitu library

```
import pandas as pd
import matplotlib.pyplot as plt
import statistics
from statistics import variance
import numpy as np
```

statistics untuk mengimport variance untuk menghitung variasi dari data sample

**Persamaan matematika menghitung mean, variance dan standar deviasi :**

### 1. Mean/rata-rata :

$$(\bar{x} + a)^n = \sum_{i=1}^n X_i$$

Mean

```
[10] paket.mean(axis = 0)
```

```
714.9970743853097
```

Gambar 6. Mean dari data sample dengan protocol UDP dan TCP

Mean

```
[19] paket1.mean(axis = 0)
```

```
729.882561595245
```

Gambar 7. Mean dari data sample dengan protocol UDP dan TCP

### 2. Variance :

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Variance

```
[11] variance(paket)
```

```
427244.1781482459
```

Gambar 8. Variansi dari data sample dengan protocol UDP dan TCP

Variance

```
[20] variance(paket1)
```

```
426841.12991863297
```

Gambar 9. Variansi dari data sample dengan protocol UDP dan TCP

### 3. Standar Deviasi :

$$\sqrt{\frac{\sum |x - \mu|^2}{N}}$$

Standar Deviasi

```
[12] paket.std()
```

```
653.6391803956072
```

Gambar 9. Standar Deviasi dari data sample dengan protocol UDP dan TCP

Standar Deviasi

```
[21] paket.std()
```

```
653.6391803956072
```

Gambar 10. Standar Deviasi dari data sample dengan protocol UDP

**5. Apakah terdapat perbedaan antara hasil pengolahan data langkah 2 sd. 4 dan yang anda peroleh dari langkah 5 ?** Jika ya mengapa dan jika tidak mengapa (jawaban tidak boleh atau tidak cukup dengan kalimat singkat 'Ya' atau 'Tidak'?)

Ya terdapat perbedaan hasil pengolahan data Mean dan Variance karena jumlah rows data dari protocol TCP dan UDP sebesar 42398 rows x 2 columns lebih banyak dari protocol UDP sebesar 41342 rows x 2columns.

42398 rows x 2 columns

Gambar 11. Data sample dengan protocol UDP dan TCP

41342 rows x 2 columns

Gambar 12. Data sample dengan protocol UDP dan TCP

Karena perbedaan jumlah data ini menyebabkan perhitungan Mean dan Variance juga akan berbeda. Varian dari set data mengukur dispersi matematis dari data relatif terhadap rata-rata/mean. Sedangkan mean merupakan rata-rata dari banyaknya set data.

**6. Hasil capture traffic Youtube dari Wireshark tersebut masih mengandung protocol TCP. Mengapa atau digunakan untuk apa protocol TCP tersebut?**

Karena saat kita memutar video streaming kita sambal melakukan browsing pada sebuah website, untuk mengakses website tersebut kita memerlukan protocol TCP karena grafik lebih stabil karena bisa digunakan untuk aktivitas secara bersamaan seperti memutar video streaming dan juga browsing. Dan juga protocol TCP bersifat VOD (Video On Demand) karena alasan sebagai berikut

1. Menyesuaikan gambar terbaik dengan mengirimkan setiap frame
2. Cukup mudah memonitor bandwidth antara sumber dan klien
3. Streaming TCP dapat dienkripsi untuk mencegah pencurian video karena penerimaan segmen dijamin dengan benar.
4. Karena mekanisme clocking sendiri, TCP lebih baik untuk bandwidth variabel yang terjadi di Internet.
5. TCP menyediakan pemulihan kesalahan dengan mentransmisikan kembali data yang hilang

Link Data dan Codingan : <https://drive.google.com/drive/folders/1hjzaE3tulXoUMhEf8-1QmBDaowIv2rrt?usp=sharing>