Jams Delay Clustering

This report was created to fulfil Mini Project Data Scientist in the Data Science Learning Studio Bootcamp 2023

Reported by **Ayub Indra**

Project Overview

Background

Bogor as one of the connecting cities of West Java and DKI Jakarta is one of the cities with dense traffic in Indonesia, it is undeniable that traffic jams can have an impact on environmental economic factors.

Project Goals

As a stakeholder, the Bogor city government needs a modelling in the form of road clustering to be able to find out which streets have the potential for high traffic jams to be able to make a traffic engineering strategy in order to expedite jams traffic.

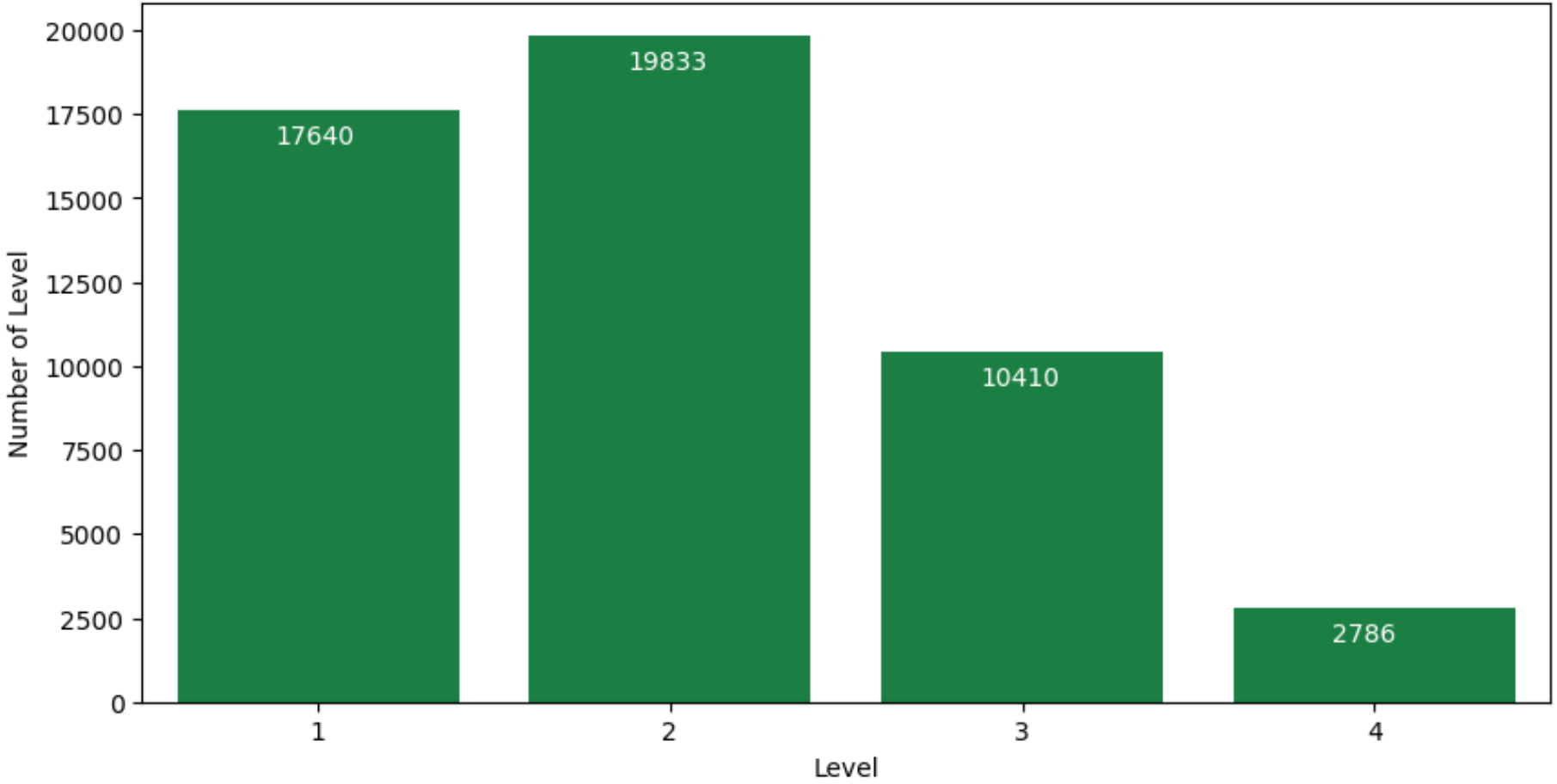
Data Understanding

To solve the problem, a dataset from satellite navigation software on smartphones (Waze) is provided. The dataset has three types of data:

- **Jams:** Traffic jam data based on current condition
- **Irregularities:** Street data based on historical data
- **Alerts:** Flags to indicate whether a special occasions occur or not on a street

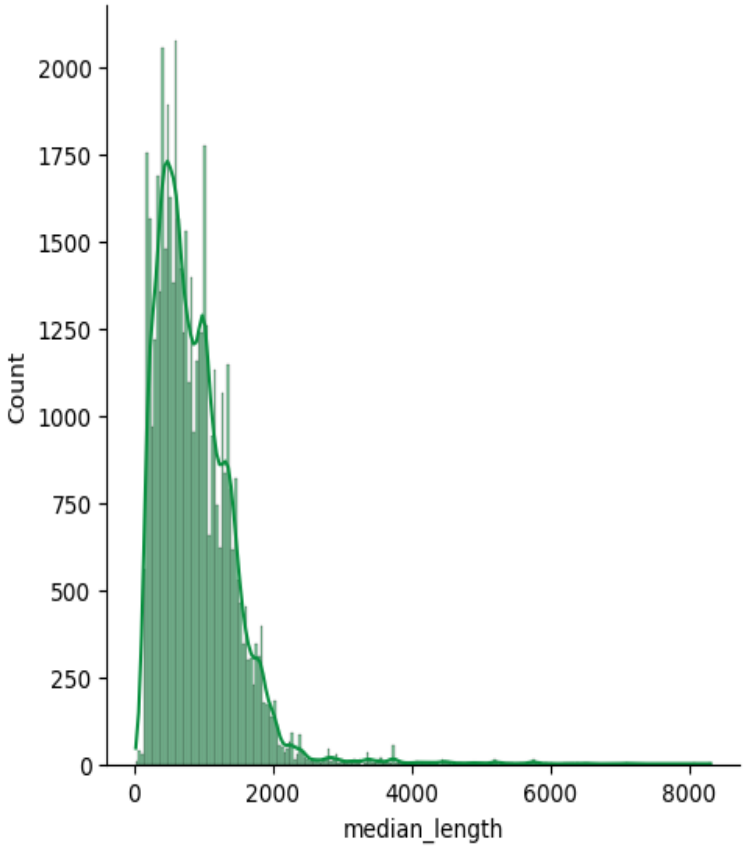
Street Level 2 Dominate the Dataset

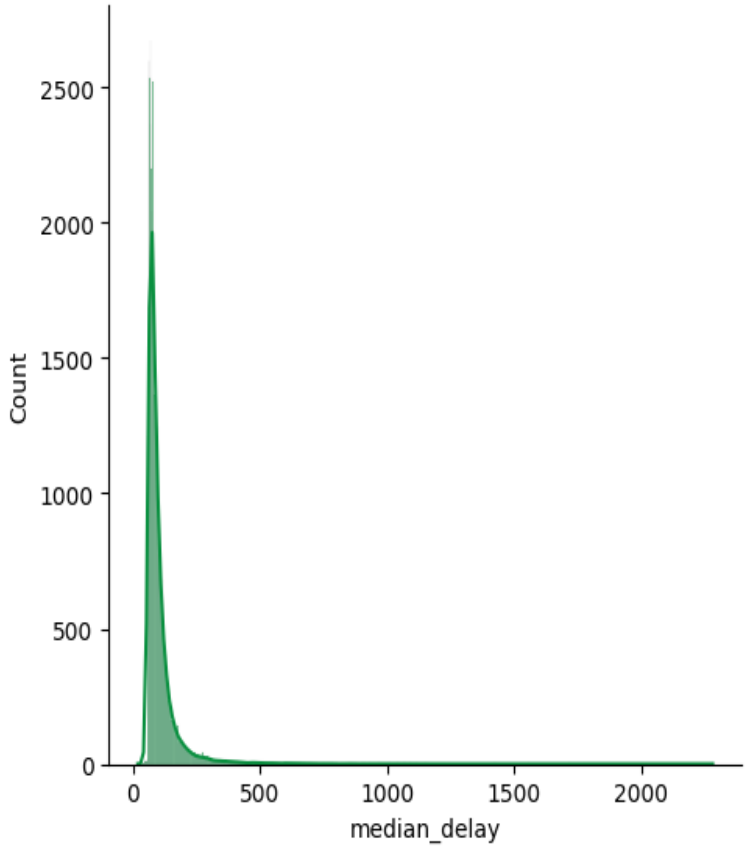
With almost 20,000 records for level 2, and then followed by street level 1, 3, 4

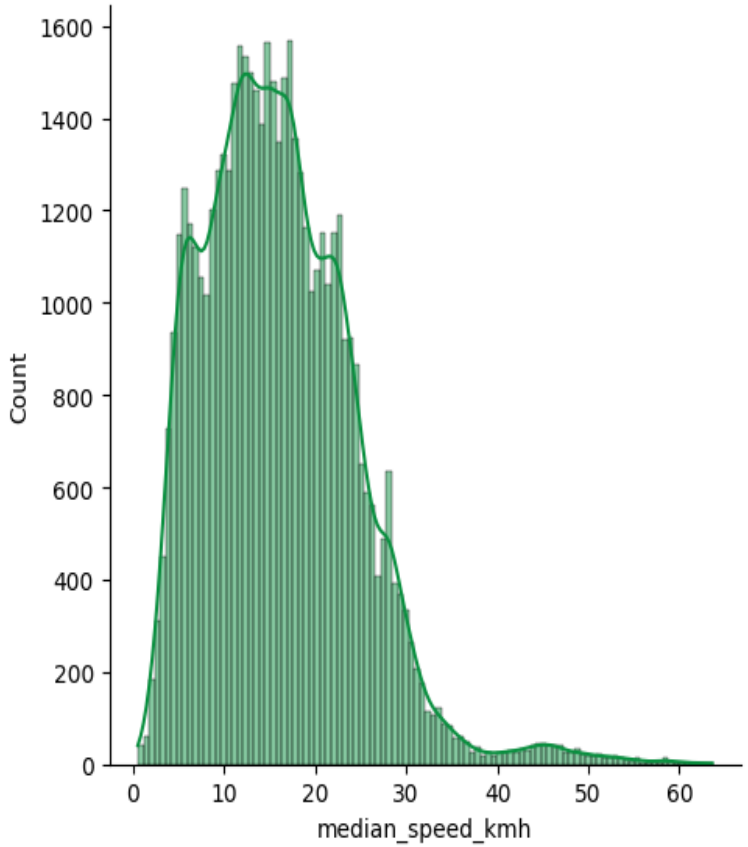


All Three Features Have Right Skewed Distribution

Median Speed have the most spread distribution than median length and median delay

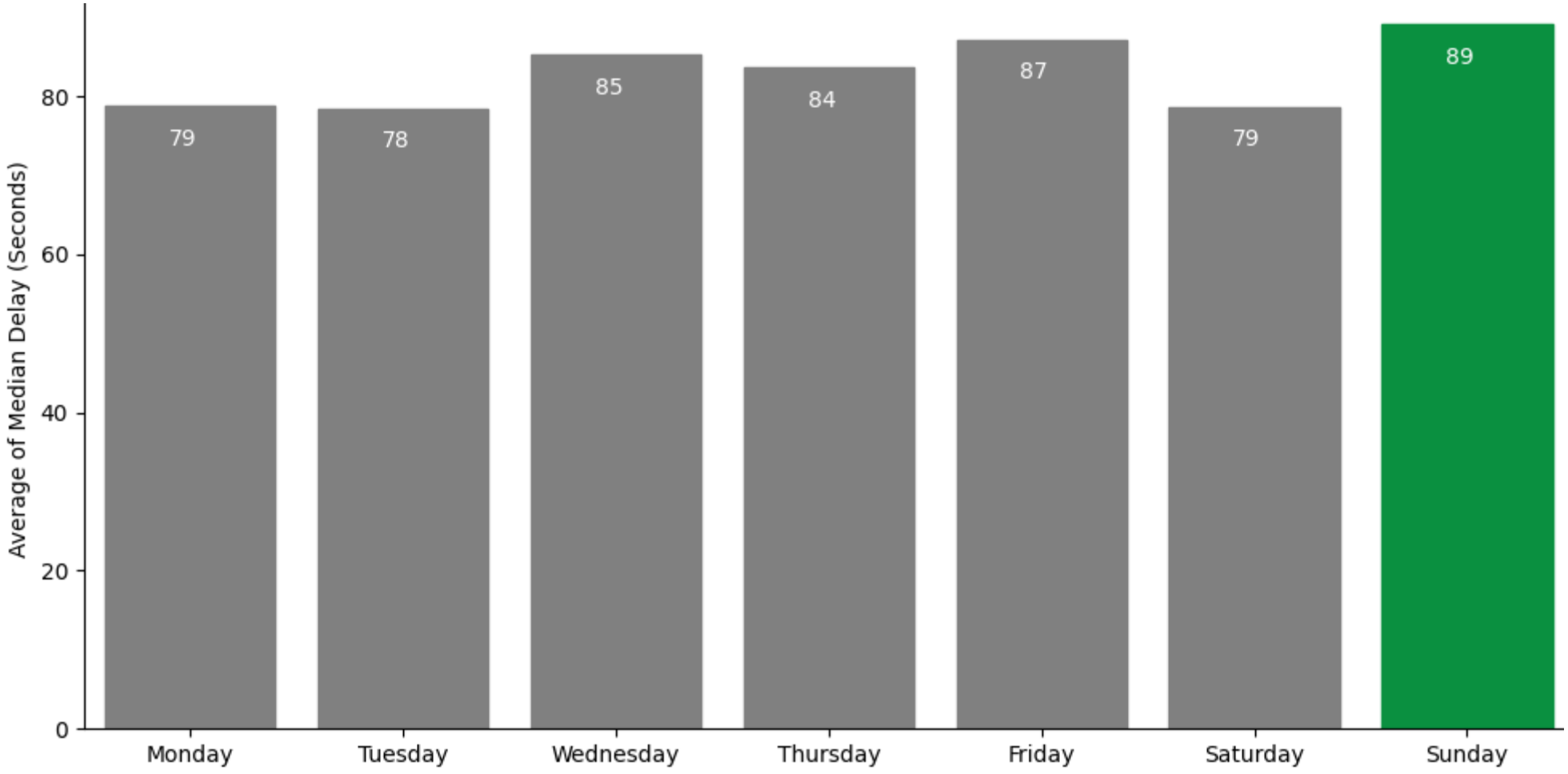




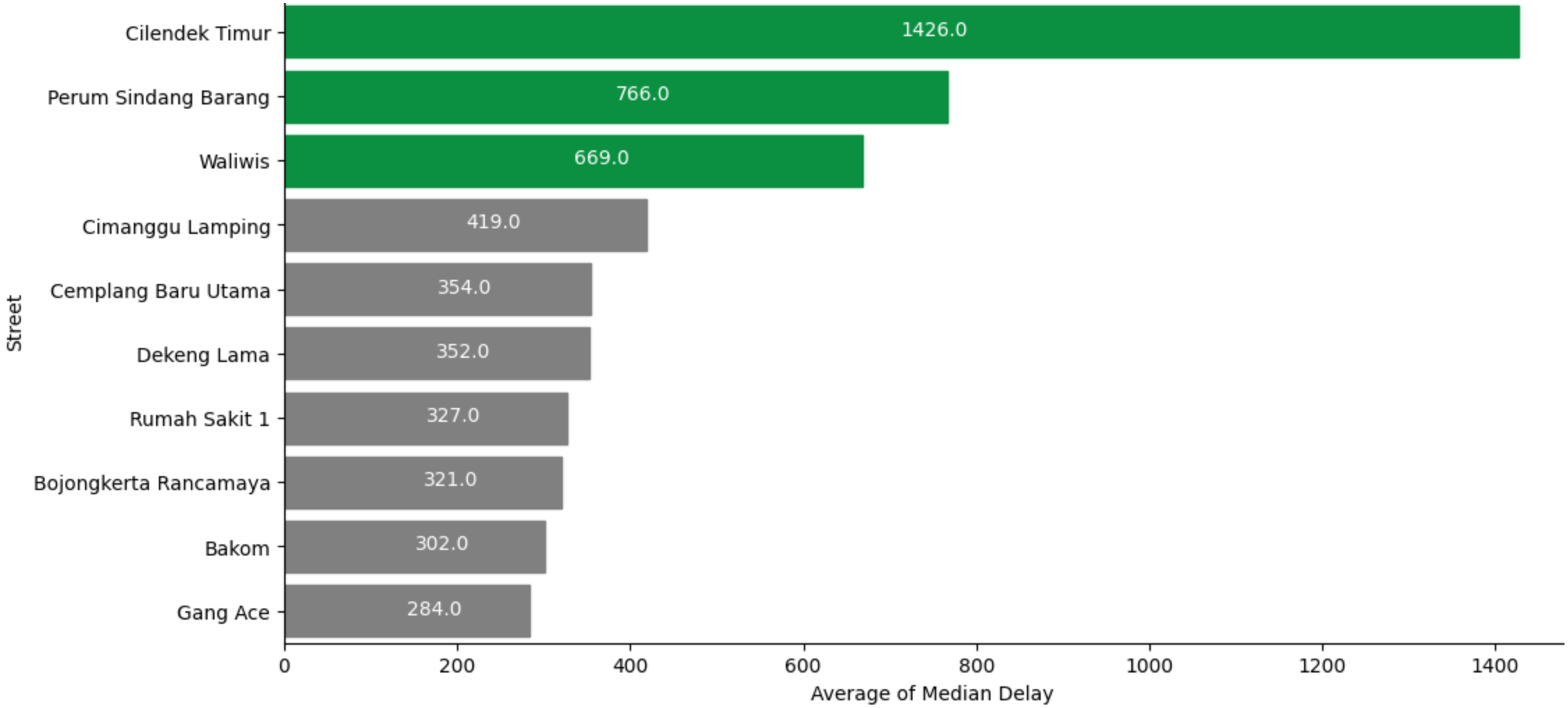


In The One Week, The Weekend of Sunday Has The Highest Delay

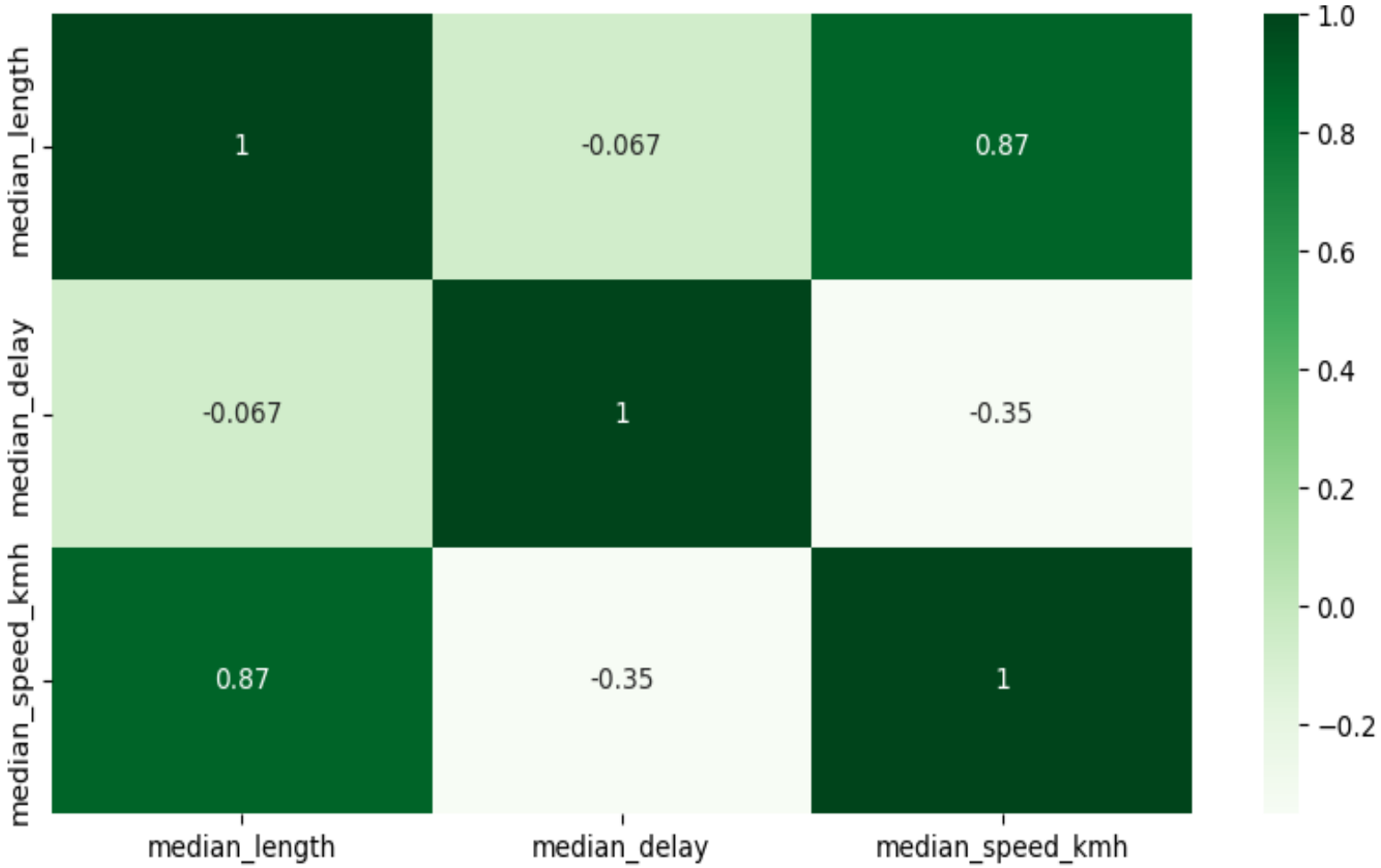
I use N9 Jalan Raya Pajajaran as a sample because it is have highest number of data records 1187.



Top 10 Street With Highest Median Delay



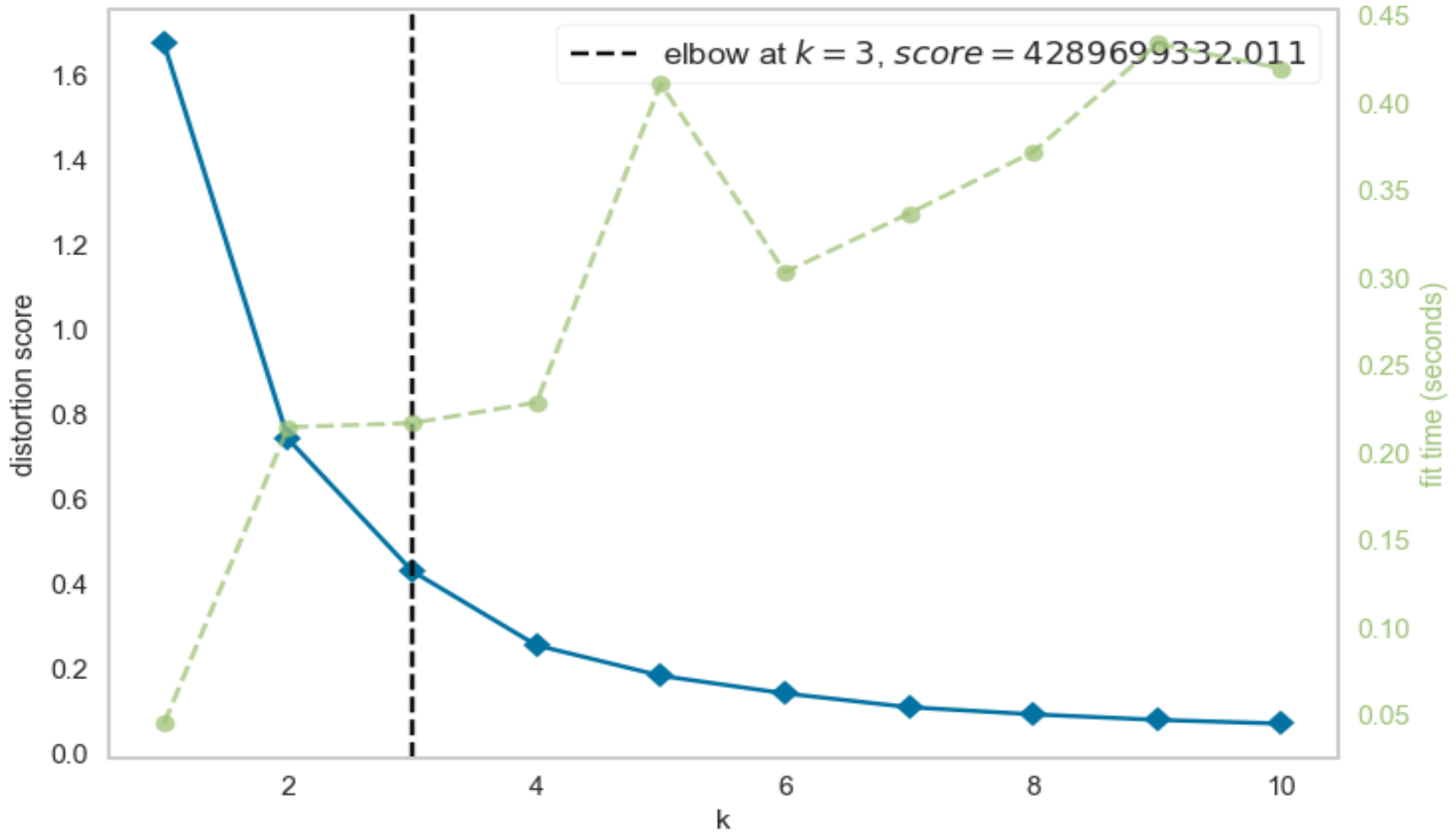
Numerical Variable Correlation Profile



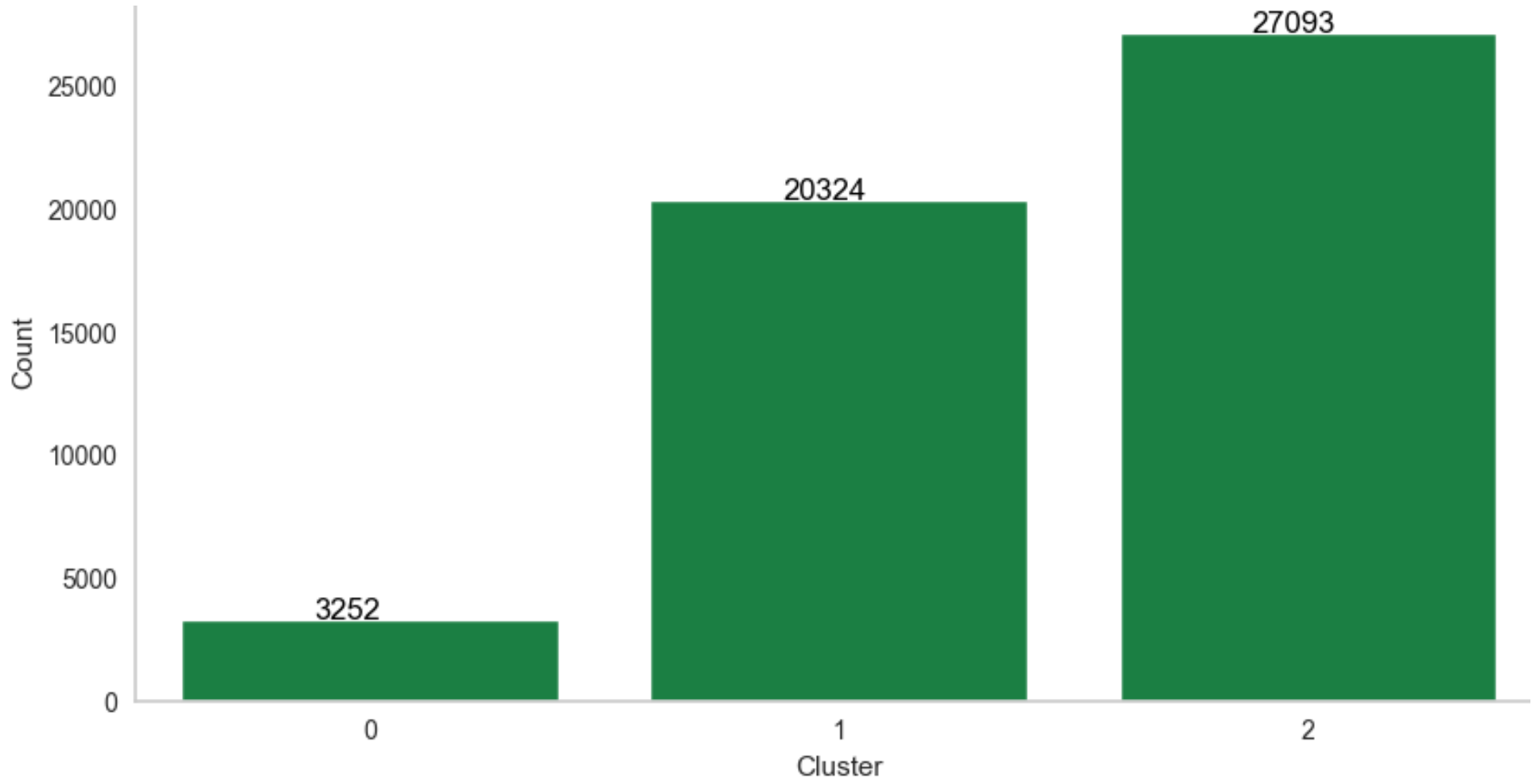
- Median Length ~ Median Delay = Weak
- Median Length ~ Median Speed = Strong
- Median Delay ~ Median Speed = Weak

K = 3 is the best optimum clusters

Using distortion score I decide to use k = 3 as shown from K Elbow Visualizer



Cluster 2 Dominate The Number of Data Records



Cluster Profile

Cluster 0 and 2 have the dangerous impact and cluster 3 has the lowest danger impact

Cluster	Average Level	Median Length	Median Delay	Median Speed
0	3,05	271,61	126,39	5,82
1	1,85	1.068,67	112,95	18,54
2	2,40	577,83	136,74	10,99
3	1,55	3.020,04	118,32	41,62

- Cluster 0 :
  lower length, higher delay, lower speed
- Cluster 1:
  moderate length, delay, and speed
- Cluster 2:
  Similar with cluster 0 but have higher length and higher speed
- Cluster 3:
  higher length, lower delay, high speed

Thank You

Contact me

ayubindra8@gmail.com

LinkedIn Profile

Github Link