

K-means



Kadir, R. A., Shima, Y., Sulaiman, R., & Ali, F. (2018). Clustering of public transport operation using K-means. In 2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA) (pp. 427–432). IEEE.

Introduction

...

Introduction

Efficient performance of transportation systems is questioned in many cities.

The main problem of the bus services usually refers to the bus arrival (Suwardo, 2010).

The objective of big data analysis is to extract new knowledge by searching, for example, for patterns in the data (Gavin, 2015).



Data Collection

Universiti Kebangsaan Malaysia (UKM)

Bus operation routes and frequency

Routes: Zone 2, 3U and 6

Time: 8:00 to 22:00

Frequency: 1 times/sec

| Time | Distance | km/h | Stop |
|---------|-------------|------|------|
| 0:11:16 | 0.000657731 | 2.4 | - |
| 0:11:17 | 0.000519982 | 1.9 | - |
| 0:11:18 | 0.000413813 | 1.5 | - |
| 0:11:19 | 0.001078258 | 3.9 | - |
| 0:11:20 | 0 | 0.0 | Stop |
| 0:11:21 | 0 | 0.0 | Stop |
| 0:11:22 | 0 | 0.0 | Stop |
| 0:11:23 | 0 | 0.0 | Stop |
| 0:11:24 | 0 | 0.0 | Stop |
| 0:11:25 | 0 | 0.0 | Stop |
| 0:11:26 | 0 | 0.0 | Stop |
| 0:11:27 | 0 | 0.0 | Stop |
| 0:11:28 | 0.001905813 | 6.9 | - |
| 0:11:29 | 0.001189536 | 4.3 | - |
| 0:11:30 | 0.002112177 | 7.6 | - |

Data Processing and Analysis

Primary data: Bus coordinates and timestamps (one second intervals).

Extended data: By calculating the number of stops made by bus through each trip
The duration of each stop is derived and added to the dataset.

Data Processing and Analysis

| Zon 2 | | | Zon 3U | | | Zon 6 | | |
|--------|----------|-----------|--------|----------|-----------|--------|----------|-----------|
| X:time | Y:volume | Z:quality | X:time | Y:volume | Z:quality | X:time | Y:volume | Z:quality |
| TIME | STOP | WAIT | TIME | STOP | WAIT | TIME | STOP | WAIT |
| 8:00 | 16 | 0:05:55 | 8:00 | 3 | 0:00:51 | 8:00 | 3 | 0:01:14 |
| 9:00 | 4 | 0:00:59 | 9:00 | 15 | 0:05:35 | 9:00 | 5 | 0:01:55 |
| 10:00 | 16 | 0:06:26 | 10:00 | 17 | 0:06:40 | 10:00 | 4 | 0:00:59 |
| 11:00 | 12 | 0:04:40 | 11:00 | 3 | 0:00:50 | 11:00 | 12 | 0:03:59 |
| 12:00 | 11 | 0:04:38 | 12:00 | 5 | 0:01:43 | 12:00 | 6 | 0:01:40 |
| 13:00 | 7 | 0:01:13 | 13:00 | 5 | 0:01:40 | 13:00 | 10 | 0:02:19 |
| 14:00 | 6 | 0:01:40 | 14:00 | 18 | 0:07:10 | 14:00 | 8 | 0:02:32 |
| 15:00 | 6 | 0:01:12 | 15:00 | 13 | 0:04:34 | 15:00 | 6 | 0:01:39 |
| 16:00 | 10 | 0:02:56 | 16:00 | 11 | 0:03:16 | 16:00 | 8 | 0:02:21 |
| 17:00 | 5 | 0:01:00 | 17:00 | 10 | 0:04:54 | 17:00 | 8 | 0:02:20 |
| 18:00 | 6 | 0:01:38 | 18:00 | 7 | 0:02:18 | 18:00 | 8 | 0:03:23 |
| 19:00 | 5 | 0:01:10 | 19:00 | 6 | 0:01:12 | 19:00 | 6 | 0:01:15 |
| 20:00 | 9 | 0:02:05 | 20:00 | 16 | 0:04:32 | 20:00 | 3 | 0:01:07 |
| 21:00 | 4 | 0:00:53 | 21:00 | 7 | 0:01:46 | 21:00 | 5 | 0:01:17 |
| 22:00 | 3 | 0:00:52 | 22:00 | 12 | 0:03:33 | 22:00 | 8 | 0:02:55 |