PRESENTASI UAS MACHINE LEARNING

# Penanganan Genangan Daerah di Surabaya Dengan Metode Clustering K-Means


Oleh Kelompok 4 :

| | |
|---|---|
| Anas Ittaqullah | **1202180011** |
| Nabila Nur Amalia | **1202190013** |
| Grahito Ardani Bimanta | **1202199006** |
| Andreas Juand Pratama | **1202192047** |

# PRESENTATION TOPICS

- Latar Belakang
- Rumusan Masalah
- Tujuan & Manfaat
- Batasan Masalah
- Landasan Teori
- Metodologi
- Sumber Data
- Langkah Analisis
- Flow Chart
- Analisis Pembahasan
- Karakteristik Genangan Air
- Parameter
- Data Preprocessing
- Kesimpulan dan Saran

# LATAR BELAKANG

Indonenisa memiliki 2 musim yaitu musim penghujan dan kemarau

Setiap musimnya memiliki jangka waktu yang panjang -+ 6 bulan

Musim hujan dengan curah yang tinggi memiliki beberapa dampak salah satunya yaitu genangan air

Metode Clustering K-Means untuk mengelompokkan daerah dengan genangan air paling tinggi untuk menjadi fokus penanganan pemerintah Kota Surabaya

# RUMUSAN MASALAH

Bagaimana cara mengelompokkan daerah genangan air di Kota Surabaya dengan optimal menggunakan metode clustering K-Means?

Bagaimana cara mengetahui tingkat genangan air pada tiap daerah di Kota Surabaya?

Daerah mana di Kota Surabaya yang membutuhkan penanganan terlebih dahulu terhadap genangan?

# TUJUAN PENELITIAN

Untuk mengelompokkan daerah di Kota Surabaya yang terdampak genangan air sehingga dapat termonitoring dengan baik.

Untuk menganalisis dan mengolah data dengan metode clustering K-Means.

# MANFAAT PENELITIAN

Dari hasil penelitian yang dilakukan diharapkan dapat memberikan masukan dan solusi terhadap fenomena banjir atau genangan pada ruas jalan di Surabaya.

Dari hasil penelitian yang dilakukan dapat digunakan sebagai bahan acuan dalam mengembangkan ilmu pengetahuan, dan dapat digunakan sebagai bahan kajian untuk penelitian yang akan datang.

# BATASAN MASALAH

Clustering ini hanya mencakup daerah di Kota Surabaya.

Daerah yang dikelompokkan yaitu hanya berdasarkan data genangan air dengan parameter luas genangan, tinggi genangan, dan lama genangan.

# DASAR TEORI

Data Mining

Genangan Air

Clustering

K-Means

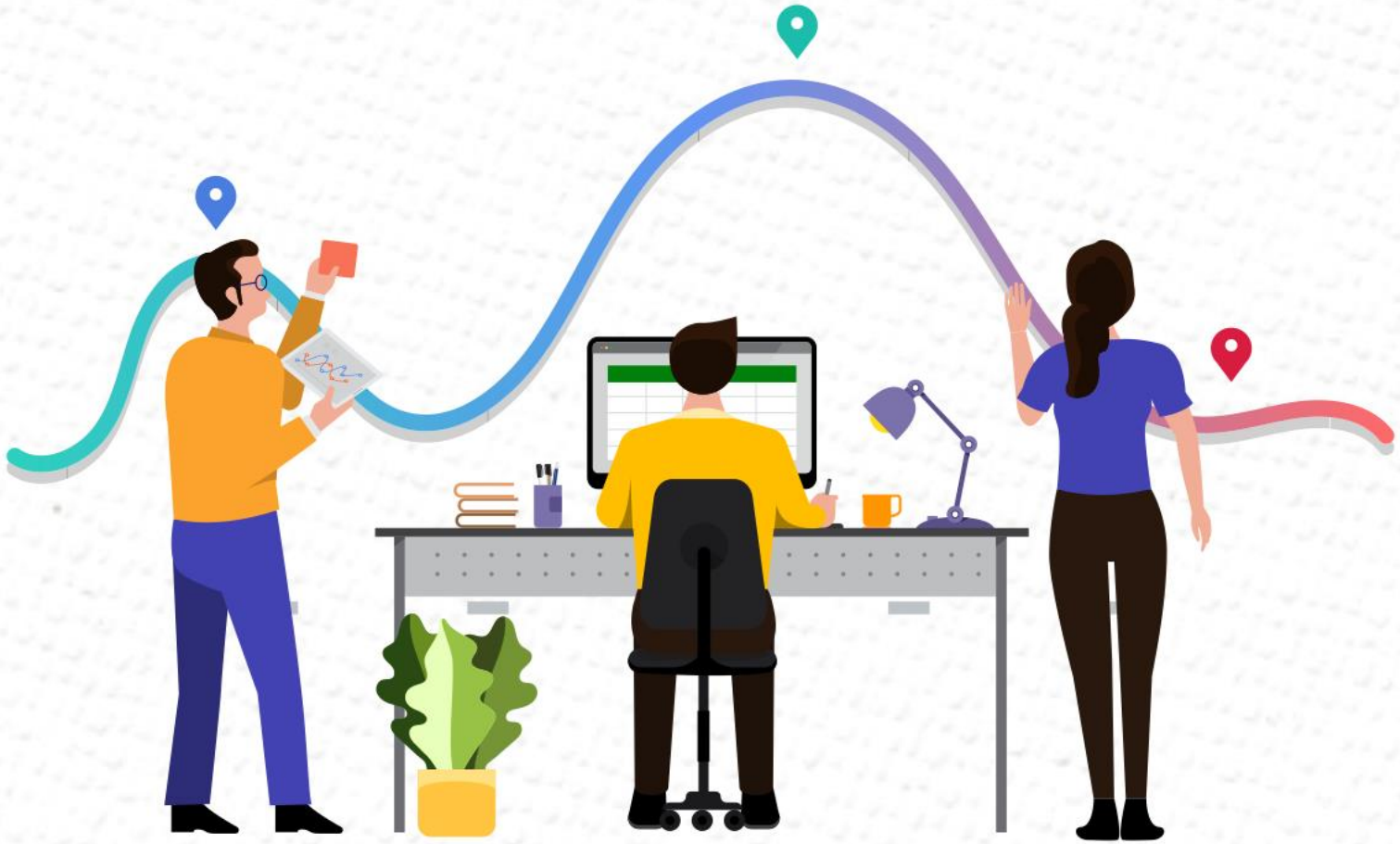
Colab

# METODOLOGI PENELITIAN

# SUMBER DATA

Dataset penelitian ini diperoleh dari data Dinas Sumber Daya Air dan Bina Marga kota Surabaya tahun 2022. Data didapatkan langsung oleh dosen pengampu perkuliahan secara langsung. Terdapat kurang lebih 244 dataset pada case genangan air daerah di kota Surabaya ini.

# LANGKAH ANALISA



# LANGKAH ANALISA

01 | Identifikasi Masalah

02 | Mendapatkan Data

03 | Normalisasi Data

04 | Eksplorasi Data

05 | Persiapan Data

06 | Pemodelan K-Means

07 | Eksperimen

# ALUR PENELITIAN

5. Studi Literature

↓

4. Pencarian dan Pengumpulan Data

↓

3. Identifikasi Masalah

↓

2. Implementasi Clustering dan Pengujian

↓

1. Hasil

# SEDIKIT DATA GENANGAN AIR (DAR( 244 DATASET)

| | Lokasi | Kecamatan | Kelurahan | Rayon | Luas genangan | Tinggi genangan | Lama genangan |
|---|---|---|---|---|---|---|---|
| 1 | Jl. Pakis Agosari (Yai | Dukuh Pakis | Dukuh Pakis | Rayon Wiyung | 10.8 | 10 | 40 |
| 2 | Jl. Mayjend Sungkor | Dukuh Pakis | Dukuh Pakis | Rayon Wiyung | 0.816236 | 25 | 58 |
| 3 | Jl. Mayjend Sungkor | Dukuh Pakis | Dukuh Pakis | Rayon Wiyung | 0.509047 | 15 | 16 |
| 4 | Kp. Kebraon | Karangpilang | Warugunung | Rayon Wiyung | 0.785111 | 10 | 40 |
| 5 | Jl. Lakar Santri RW.0 | Lakarsantri | Bangkingan | Rayon Wiyung | 0.749717 | 30 | 16 |
| 6 | Lontar Lidah Kulon | Lakarsantri | Lidah Kulon | Rayon Wiyung | 1.175965 | 4 | 16 |
| 7 | Jl. Babatan Indah | Wiyung | Babatan | Rayon Wiyung | 8.1 | 10 | 16 |
| 8 | Perumahan Lembah | Lakarsantri | Lidah Wetan | Rayon Wiyung | 0.182205 | 10 | 16 |
| 9 | Jl. Simo Hilir | Sukomanunggal | Simomulyo | Rayon Tandes | 0.205258 | 4 | 16 |
| 10 | Jl. Kupang Baru | Dukuh Pakis | Dukuh Kupang | Rayon Tandes | 0.258649 | 25 | 16 |
| 11 | Kp. Babatan dan sek | Wiyung | Babatan | Rayon Wiyung | 0.22065 | 25 | 16 |
| 12 | Jl. Wiyung (Gg. DPR | Wiyung | Wiyung | Rayon Wiyung | 0.1855 | 25 | 16 |
| 13 | Jl. Hayam Wuruk | Wonokromo | Sawunggaling | Rayon Wiyung | 0.394008 | 15 | 40 |
| 14 | Jl. Adityawaman | Wonokromo | Wonokromo | Rayon Wiyung | 1.120918 | 10 | 40 |
| 15 | Jl. Jajar Tunggal | Wiyung | Jajar Tunggal | Rayon Wiyung | 0.546206 | 50 | 80 |
| 16 | Jl. H. Abdul Wahab S | Dukuh Pakis | Dukuh Pakis | Rayon Wiyung | 0.69076 | 30 | 50 |
| 17 | Jl. Gajah Mada | Wonokromo | Sawunggaling | Rayon Wiyung | 0.224743 | 25 | 40 |
| 18 | Jl. Simo Rejo | Sukomanunggal | Simomulyo | Rayon Tandes | 10 | 10 | 31 |
| 19 | Jl. Simo Langit | Sukomanunggal | Simomulyo | Rayon Tandes | 0.166687 | 10 | 16 |
| 20 | Jl. Dukuh Menangga | Gayungan | Dukuh Menanggal | Rayon Jambangan | 0.308184 | 10 | 40 |
| 21 | Jl. Bangkingan | Lakarsantri | Bangkingan | Rayon Wiyung | 0.488274 | 15 | 16 |
| 22 | Jl. Lakarsantri (Jemb | Lakarsantri | Bangkingan | Rayon Wiyung | 9.6 | 25 | 16 |
| 23 | Jl. Dukuh Pakis | Dukuh Pakis | Dukuh Pakis | Rayon Wiyung | 0.36091 | 25 | 31 |
| 24 | Jl. Waru Gunung (Ke | Karangpilang | Warugunung | Rayon Wiyung | 0.212653 | 25 | 16 |
| 25 | Jl. Waru Gunung (Ru | Karangpilang | Warugunung | Rayon Wiyung | 0.413372 | 25 | 16 |
| 26 | Kp. Lidah Kulon RW | Lakarsantri | Lidah Kulon | Rayon Wiyung | 0.13233 | 30 | 16 |
| 27 | Jl. Raya Simo Jawar | Sukomanunggal | Simomulyo | Rayon Tandes | 0.762466 | 4 | 16 |
| 28 | Jl. Suko Manunggal | Sukomanunggal | Simomulyo | Rayon Tandes | 1.018027 | 4 | 16 |

# PERHITUNGAN ALGORITMA K-MEANS

Algoritma K-means merupakan teknik data mining yaitu metode clustering atau pengelompokkan yang proses pemodelannya tanpa supervisi/pembelajaran serta metode pengelompokkan datanya dilakukan secara partisi. Pada metode yang digunakan dalam algoritma K-means, data akan dikelompokkan menjadi beberapa bagian kelompok dan setiap kelompok memiliki ciri-ciri yang mirip dengan satu sama lain, namun memiliki ciri-ciri yang berbeda dengan kelompok lain.

bertujuan untuk meminimalisir perbedaan antara satu cluster dan memaksimalkan perbedaan data dengan cluster lain. Setiap data yang ada di dalam satu kelompok tersebut, mempunyai karakteristik yang sama maupun karakteristik yang berbeda dengan data di kelompok yang lain.

# PROSES PENGOLAHAN DATA K-MEANS

1. Menentukan jumlah cluster

2. Menentukan k yang menjadi titik pusat cluster (centeroid) secara acak untuk awal perhitungan, selanjutnya akan ditentukan centeroid cluster untuk iterasi berikutnya menggunakan persamaan

$$v = \frac{\sum_{i=1}^{n} xi}{n}; i - 1,2,3 \dots \dots n$$

v : titik pusat kelompok
xi : objek data ke=i
n : jumlah objek yang menjadi anggota cluster

3. Menghitung jarak setiap objek data ke setiap centeroid yang telah ditentukan menggunakan Euclidean distance [7] pada persamaan (4). Tempatkan setiap objek sesuai dengan titik pusat cluster terdekat.

$$D(x_2, x_1) = \|x_2, x_1\|_2 = \sqrt{\sum_{j=1}^{p} (x_{2j} - x_{1j})^2}$$

D : jarak kuadrat Euclidean antara objek ke x2 dan x1
P : jumlah variabel cluster X2j : data ke-2 pada variabel ke-j
X1j : data ke-1 pada variabel ke-j

# KARAKTERISTIK GENANGAN AIR

Dalam menentukan karakteristik genangan air didapatkan dengan mengolah data menggunakan Google Colab dengan mengimplementasikan metode clustering K-Means dengan memakai tiga atribut, yaitu Luas Genangan, Tinggi Genangan, Lama Genangan yang ada di wilayah Surabaya.

Dalam menganalisis karakteristik genangan air diuji dengan menggunakan Google Collab dan mengimplementasikan metode clustering K-Means, sehingga genangan pada daerah di Surabaya per tahun 2022 dapat dikelompokkan menjadi tiga cluster, yaitu daerah dengan genangan tinggi, sedang, dan rendah.

# KRITERIA ATRIBUT GENANGAN AIR

## Luas Genangan

Semakin besar luas genangan, maka kerugian dan kerusakan yang diperoleh akan semakin besar pula. Begitu juga dengan tingkat kerugian genangan air di Surabaya semakin besar. Klasifikasi luas genangan akan dibagi menjadi tiga kelas yaitu :

< 1 (Rendah)

2 - 9 (Sedang)

> 10 (Tinggi)

# KRITERIA ATRIBUT GENANGAN AIR

## Tinggi Genangan

Semakin besar tinggi genangan, maka kerugian dan kerusakan yang diperoleh akan semakin besar pula. Begitu juga dengan tingkat bahaya genangan banjir di daerah Surabya akan semakin besar. Klasifikasi tinggi genangan akan dibagi menjadi tiga kelas yaitu :

- Kedalaman < 15 cm (Rendah)
- Kedalaman 15 – 50 cm (Sedang)
- Kedalaman > 50 cm (Tinggi).

# KRITERIA ATRIBUT GENANGAN AIR

## Lama Genangan

Semakin besar lama genangan, maka kerugian dan kerusakan yang diperoleh akan semakin besar pula. Begitu juga dengan tingkat bahaya yang ditimbulkan genangan air di Surabaya akan semakin besar. Klasifikasi lama genangan akan dibagi menjadi tiga kelas yaitu :

- < 6 jam (Rendah)
- 6 – 12 jam (Sedang)
- > 12 jam (Tinggi)

# DATA PREPROCESSING

Data set yang digunakan pada penelitian ini adalah data dengan format csv.
Berikut contoh tabel sampel data :

| | _id | Lokasi | Kecamatan | Kelurahan | Rayon | Luas genangan | Tinggi genangan | Lama genangan | Skor Luas Genangan | Skor Tinggi Genangan | Skor Lama Genangan |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | Jl. Pakis Agosari (Yamaha) | Dukuh Pakis | Dukuh Pakis | Rayon Wiyung | 10.800000 | 10 | 40 | 100 | 25 | 25 |
| 1 | 2 | Jl. Mayjend Sungkono - Darmo Park I | Dukuh Pakis | Dukuh Pakis | Rayon Wiyung | 0.816236 | 25 | 58 | 75 | 50 | 25 |
| 2 | 3 | Jl. Mayjend Sungkono (Pom Bensin) | Dukuh Pakis | Dukuh Pakis | Rayon Wiyung | 0.509047 | 15 | 16 | 75 | 50 | 0 |
| 3 | 4 | Kp. Kebraon | Karangpilang | Warugunung | Rayon Wiyung | 0.785111 | 10 | 40 | 75 | 25 | 25 |
| 4 | 5 | Jl. Lakar Santri RW 001 | Lakarsantri | Bangkingan | Rayon Wiyung | 0.749717 | 30 | 16 | 75 | 50 | 0 |

# DATA PREPROCESSING

Kemudian melakukan tahap preprocessing data untuk melakukan pengecekkan pada data yang akan diolah.

Tahap Preprocessing data meliputi :
- Pengecekkan nilai null
  pada dataset = tidak terdapat missing value pada dataset yang diolah

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 244 entries, 0 to 243
Data columns (total 11 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   _id                   244 non-null    int64
 1   Lokasi                244 non-null    object
 2   Kecamatan             244 non-null    object
 3   Kelurahan             244 non-null    object
 4   Rayon                 244 non-null    object
 5   Luas genangan         244 non-null    float64
 6   Tinggi genangan       244 non-null    int64
 7   Lama genangan         244 non-null    int64
 8   Skor Luas Genangan    244 non-null    int64
 9   Skor Tinggi Genangan  244 non-null    int64
 10  Skor Lama Genangan    244 non-null    int64
dtypes: float64(1), int64(6), object(4)
memory usage: 21.1+ KB
```

# DATA PREPROCESSING

Lanjutan Tahap Preprocessing data meliputi :

- Pengecekan kepada nilai duplikat dataset = tidak terdapat nilai duplikat pada dataset yang diolah.

- Pengecekan data outliers, apakah data mempunyai value atau karakteristik yang berbeda jauh dengan data lainnya = tidak terdapat data yang berbeda, dengan rata-rata data masih dalam range yang sama dan hasil yang didapatkan stabil.

# NORMALISASI DATA

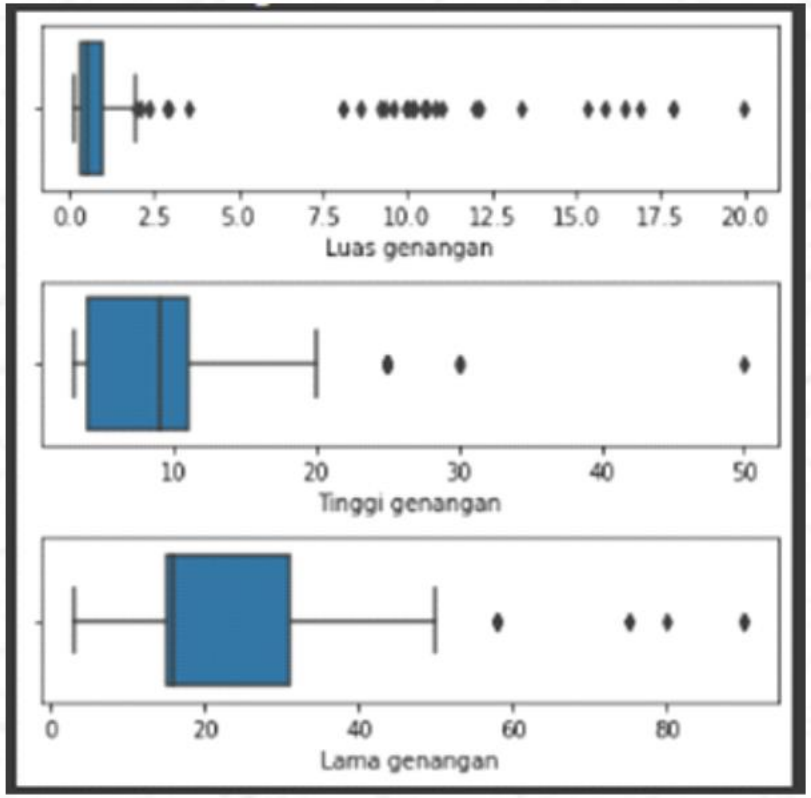
Untuk memeriksa apakah pada dataframe ada nilai null atau tidaknya = tidak ada, seperti pada gambar di bawah ini :

```
_id                    0
Lokasi                 0
Kecamatan              0
Kelurahan              0
Rayon                  0
Luas genangan          0
Tinggi genangan        0
Lama genangan          0
Skor Luas Genangan     0
Skor Tinggi Genangan   0
Skor Lama Genangan     0
dtype: int64
```

# CEK DATA OUTLIER

Pengecekan data outliers, apakah data mempunyai value atau karakteristik yang berbeda jauh dengan data lainnya = Tidak terdapat data yang berbeda, rata-rata data masih dalam range yang sama dan hasil yang didapatkan stabil. Dapat dilihat hasil dari pengecekan data outliers seperti gambar dibawah ini :

Dapat dilihat dari hasil boxplot bahwa hasil data stabil atau normal, tidak ada data yang memiliki value yang berbeda jauh dari data lainnya. Sehingga dapat disimpulkan dalam data ini bahwa tidak ada data outlier.

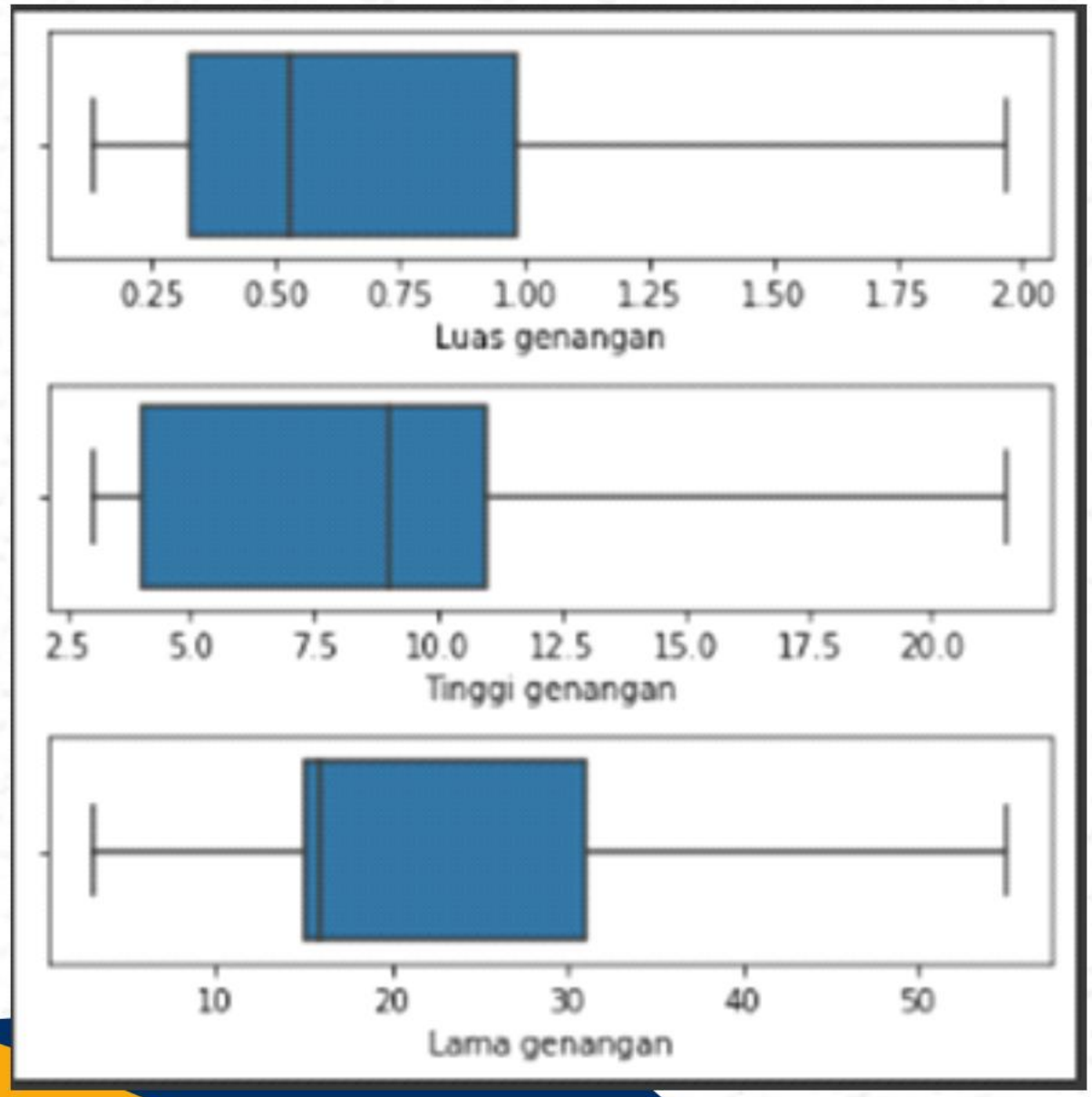



Hasil Box Plot

# CEK PERSEBARAN DATA

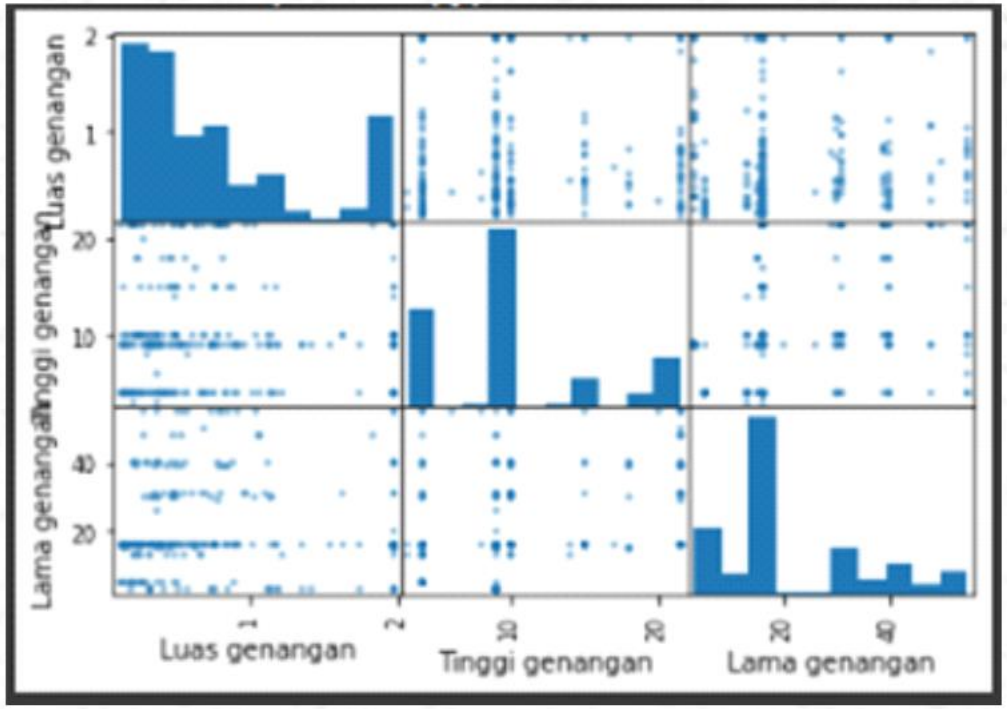
| | _id | Luas genangan | Tinggi genangan | Lama genangan | Skor Luas Genangan | Skor Tinggi Genangan | Skor Lama Genangan |
|---|---|---|---|---|---|---|---|
| count | 244.000000 | 244.000000 | 244.000000 | 244.000000 | 244.000000 | 244.000000 | 244.000000 |
| mean | 122.500000 | 1.854777 | 10.922131 | 22.040984 | 67.315574 | 24.795082 | 7.172131 |
| std | 70.580923 | 3.741298 | 7.146187 | 16.531749 | 18.421974 | 18.143207 | 12.619596 |
| min | 1.000000 | 0.131661 | 3.000000 | 3.000000 | 50.000000 | 0.000000 | 0.000000 |
| 25% | 61.750000 | 0.324975 | 4.000000 | 15.000000 | 50.000000 | 0.000000 | 0.000000 |
| 50% | 122.500000 | 0.527731 | 9.000000 | 16.000000 | 75.000000 | 25.000000 | 0.000000 |
| 75% | 183.250000 | 0.983013 | 11.000000 | 31.000000 | 75.000000 | 31.250000 | 25.000000 |
| max | 244.000000 | 19.980000 | 50.000000 | 90.000000 | 100.000000 | 75.000000 | 50.000000 |

Kemudian pengecekan terhadap Nilai Maximum dan minimum dari masing-masing Variabel, serta persebaran data setelah di tangani outlier.


0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
Luas genangan
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Tinggi genangan
10
20
30
40
50
Lama genangan

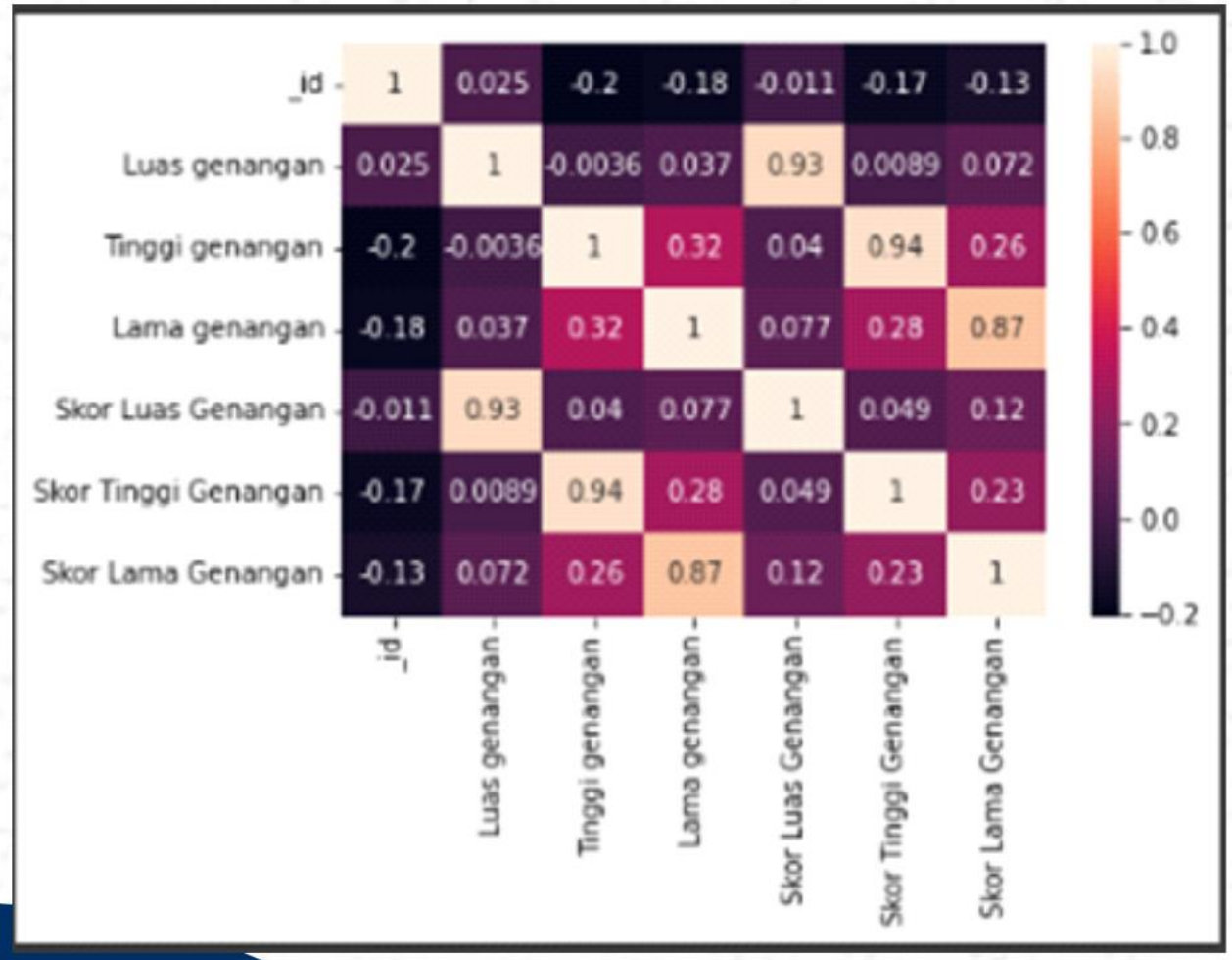
# PERBAIKAN PENYEBARAN DATA

- Kemudian melakukan perbaikan penyebaran data dengan metode interquartile range berikut hasilnya :



# HEATMAP

-Heatmap yang dihasilkan pada tahap ini berupa :



# CEK PERSEBARAN DATA

| | _id | Luas genangan | Tinggi genangan | Lama genangan | Skor Luas Genangan | Skor Tinggi Genangan | Skor Lama Genangan |
|---|---|---|---|---|---|---|---|
| count | 244.000000 | 244.000000 | 244.000000 | 244.000000 | 244.000000 | 244.000000 | 244.000000 |
| mean | 122.500000 | 1.854777 | 10.922131 | 22.040984 | 67.315574 | 24.795082 | 7.172131 |
| std | 70.580923 | 3.741298 | 7.146187 | 16.531749 | 18.421974 | 18.143207 | 12.619596 |
| min | 1.000000 | 0.131661 | 3.000000 | 3.000000 | 50.000000 | 0.000000 | 0.000000 |
| 25% | 61.750000 | 0.324975 | 4.000000 | 15.000000 | 50.000000 | 0.000000 | 0.000000 |
| 50% | 122.500000 | 0.527731 | 9.000000 | 16.000000 | 75.000000 | 25.000000 | 0.000000 |
| 75% | 183.250000 | 0.983013 | 11.000000 | 31.000000 | 75.000000 | 31.250000 | 25.000000 |
| max | 244.000000 | 19.980000 | 50.000000 | 90.000000 | 100.000000 | 75.000000 | 50.000000 |

Kemudian pengecekan terhadap Nilai Maximum dan minimum dari masing-masing Variabel, serta persebaran data setelah di tangani outlier.

# K-Means Clustering (Mix Max Scaler)

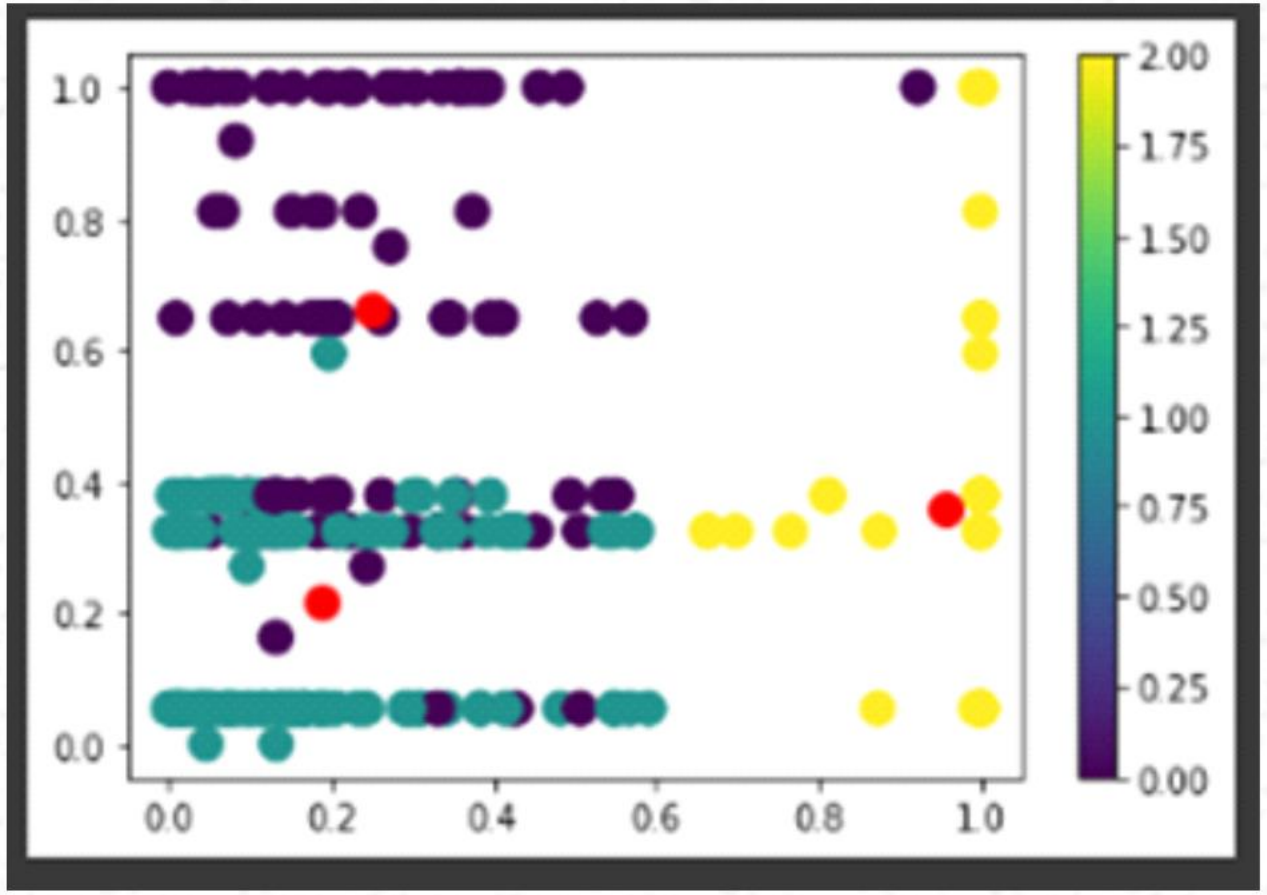
Mencari cluster center dan didapatkan hasil :

```
[[0.25116326 0.66151866 0.58150183]
 [0.1878503  0.21713239 0.20257497]
 [0.95603304 0.35714286 0.30769231]]
```

Mencari K-Means label :

```
[2 0 0 0 0 1 2 1 1 0 0 0 0 0 0 0 0 2 1 0 0 2 0 0 0 0 1 1 1 1 1 2 0 1 1 1 2
 0 0 2 0 1 2 0 0 0 0 0 0 1 2 1 2 1 1 1 0 0 1 1 1 1 0 1 0 1 0 1 2 2 0 1 1 2
 1 0 0 2 2 2 1 0 0 0 2 2 0 0 1 1 1 1 1 0 1 2 2 1 0 2 2 1 1 2 1 2 0 0 0 0 1
 1 1 0 0 1 1 2 2 0 1 1 1 1 0 0 1 1 0 0 1 1 1 1 1 1 2 1 2 1 1 1 1 1 2 1 1 0
 1 0 0 1 0 1 1 1 1 2 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 0 1 1 1 1 1 2 1 1 1 0
 0 0 0 0 2 0 0 0 0 0 1 0 0 0 1 0 1 1 1 0 1 1 0 0 0 0 0 0 0 1 0 0 1 1 1 1 2
 2 1 1 1 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 2 2 2]
```

# Hasil Scatter Plot



# Hasil Proses Min - Max Scaler

```
     _id                              Lokasi      Kecamatan      Kelurahan  \
0      1          Jl. Pakis Agosari (Yamaha)    Dukuh Pakis    Dukuh Pakis
1      2  Jl. Mayjend Sungkono - Darmo Park I    Dukuh Pakis    Dukuh Pakis
2      3    Jl. Mayjend Sungkono (Pom Bensin)    Dukuh Pakis    Dukuh Pakis
3      4                          Kp. Kebraon   Karangpilang     Warugunung
4      5              Jl. Lakar Santri RW.001    Lakarsantri     Bangkingan
..   ...                                  ...            ...            ...
239  240                            Jl. Kanwa      WONOKROMO          DARMO
240  241                Per. Rungkut Harapan        Rungkut  Rungkut Kidul
241  242                       TANJUNGSARI I  SUKOMANUNGGAL    TANJUNGSARI
242  243                      Jl. Karangrejo      Wonokromo      Wonokromo
243  244                      Kp. Tambakdono         Benowo     Tambakdono

               Rayon  Luas genangan  Tinggi genangan  Lama genangan  \
0       Rayon Wiyung       1.970070             10.0             40
1       Rayon Wiyung       0.816236             21.5             55
2       Rayon Wiyung       0.509047             15.0             16
3       Rayon Wiyung       0.785111             10.0             40
4       Rayon Wiyung       0.749717             21.5             16
..               ...            ...              ...            ...
239    Rayon Genteng       0.334177              9.0              3
240  Rayon Jambangan       0.851371              9.0              3
241     Rayon Tandes       1.970070              9.0             15
242  Rayon Jambangan       1.970070              9.0             20
243     Rayon Tandes       1.970070              4.0             15

     Skor Luas Genangan  Skor Tinggi Genangan  Skor Lama Genangan  kluster
0                   100                    25                  25        2
1                    75                    50                  25        0
2                    75                    50                   0        0
3                    75                    25                  25        0
4                    75                    50                   0        0
..                  ...                   ...                 ...      ...
239                  50                    25                   0        1
240                  75                    25                   0        1
241                 100                    25                   0        2
242                 100                    25                   0        2
243                 100                     0                   0        2

[244 rows x 12 columns]
```

# Hasil Standard Scaler

| | Luas genangan | Tinggi genangan | Lama genangan |
|---|---|---|---|
| 0 | 1.970070 | 10.0 | 40 |
| 1 | 0.816236 | 21.5 | 55 |
| 2 | 0.509047 | 15.0 | 16 |
| 3 | 0.785111 | 10.0 | 40 |
| 4 | 0.749717 | 21.5 | 16 |
| ... | ... | ... | ... |
| 239 | 0.334177 | 9.0 | 3 |
| 240 | 0.851371 | 9.0 | 3 |
| 241 | 1.970070 | 9.0 | 15 |
| 242 | 1.970070 | 9.0 | 20 |
| 243 | 1.970070 | 4.0 | 15 |

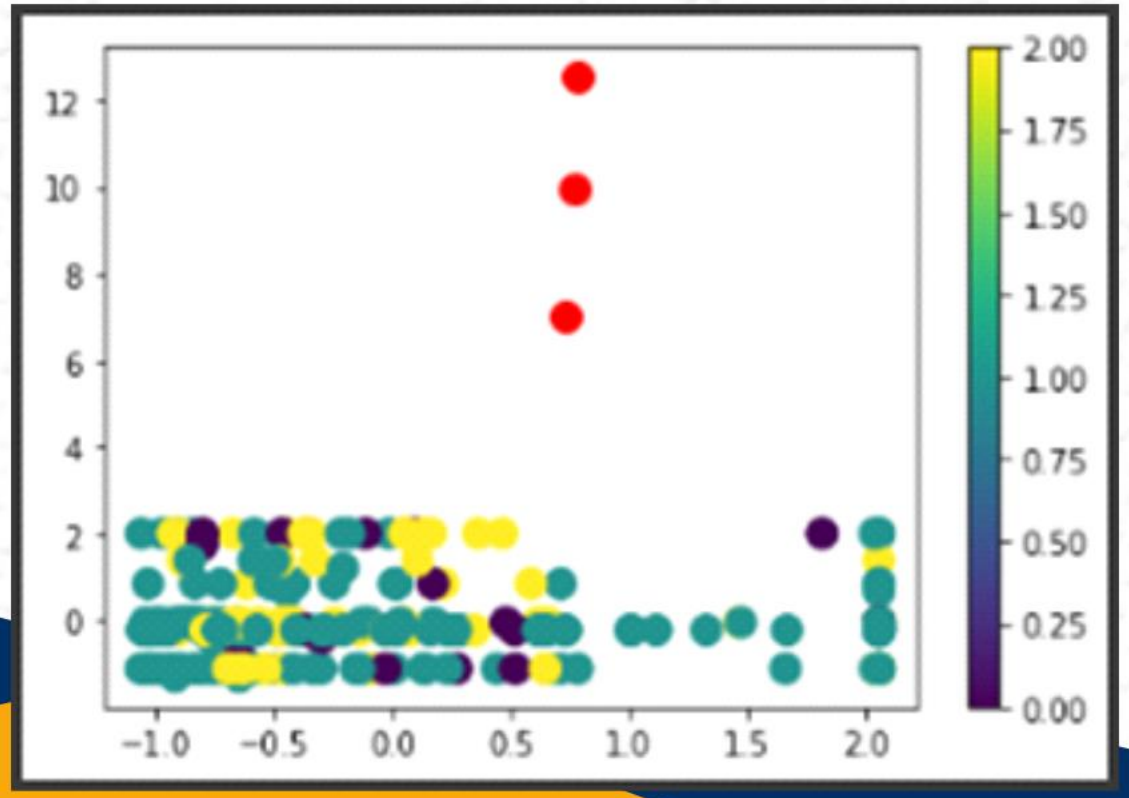
244 rows × 3 columns

# Identifikasi Cluster

```
array([1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 1, 0, 0,
       1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0,
       0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0,
       1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 1,
       0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1,
       0, 2, 2, 1, 1, 2, 2, 0, 0, 0, 0, 2, 0, 2, 1, 1, 0, 0, 1, 1, 0, 0,
       2, 0, 0, 0, 0, 0, 1, 2, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 1, 1, 1, 0,
       0, 0, 0, 0, 0, 0, 2, 0, 2, 0, 0, 2, 0, 2, 2, 2, 0, 0, 0, 0, 1, 0,
       2, 0, 0, 0, 0, 0, 2, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0,
       1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 0,
       0, 0], dtype=int32)
```

# Scatter Plot dari Target

Berikut Scatter Plot dari ketiga target, yakni :
Luas genangan, Tinggi genangan, Lama genangan.




Pada scatter plot diatas sudah terlihat jelas jika ketiga target berhasil di kelompokkan sesuai dengan warna masing-masing.

# Hasil Clustering Berdasar Target, Rayon, Skor

| | _id | Lokasi | Kecamatan | Kelurahan | \ |
|---|---|---|---|---|---|
| 0 | 1 | Jl. Pakis Agosari (Yamaha) | Dukuh Pakis | Dukuh Pakis | |
| 1 | 2 | Jl. Mayjend Sungkono - Darmo Park I | Dukuh Pakis | Dukuh Pakis | |
| 2 | 3 | Jl. Mayjend Sungkono (Pom Bensin) | Dukuh Pakis | Dukuh Pakis | |
| 3 | 4 | Kp. Kebraon | Karangpilang | Warugunung | |
| 4 | 5 | Jl. Lakar Santri RW.001 | Lakarsantri | Bangkingan | |
| .. | ... | ... | ... | ... | |
| 239 | 240 | Jl. Kanwa | WONOKROMO | DARMO | |
| 240 | 241 | Per. Rungkut Harapan | Rungkut | Rungkut Kidul | |
| 241 | 242 | TANJUNGSARI I | SUKOMANUNGGAL | TANJUNGSARI | |
| 242 | 243 | Jl. Karangrejo | Wonokromo | Wonokromo | |
| 243 | 244 | Kp. Tambakdono | Benowo | Tambakdono | |

| | Rayon | Luas genangan | Tinggi genangan | Lama genangan | \ |
|---|---|---|---|---|---|
| 0 | Rayon Wiyung | 1.970070 | 10.0 | 40 | |
| 1 | Rayon Wiyung | 0.816236 | 21.5 | 55 | |
| 2 | Rayon Wiyung | 0.509047 | 15.0 | 16 | |
| 3 | Rayon Wiyung | 0.785111 | 10.0 | 40 | |
| 4 | Rayon Wiyung | 0.749717 | 21.5 | 16 | |
| .. | ... | ... | ... | ... | |
| 239 | Rayon Genteng | 0.334177 | 9.0 | 3 | |
| 240 | Rayon Jambangan | 0.851371 | 9.0 | 3 | |
| 241 | Rayon Tandes | 1.970070 | 9.0 | 15 | |
| 242 | Rayon Jambangan | 1.970070 | 9.0 | 20 | |
| 243 | Rayon Tandes | 1.970070 | 4.0 | 15 | |

| | Skor Luas Genangan | Skor Tinggi Genangan | Skor Lama Genangan | kluster | \ |
|---|---|---|---|---|---|
| 0 | 100 | 25 | 25 | 2 | |
| 1 | 75 | 50 | 25 | 0 | |
| 2 | 75 | 50 | 0 | 0 | |
| 3 | 75 | 25 | 25 | 0 | |
| 4 | 75 | 50 | 0 | 0 | |
| .. | ... | ... | ... | ... | |
| 239 | 50 | 25 | 0 | 1 | |
| 240 | 75 | 25 | 0 | 1 | |
| 241 | 100 | 25 | 0 | 2 | |
| 242 | 100 | 25 | 0 | 2 | |
| 243 | 100 | 0 | 0 | 2 | |

| | TipePenanganan |
|---|---|
| 0 | Rendah |
| 1 | Tinggi |
| 2 | Sedang |
| 3 | Rendah |
| 4 | Sedang |
| .. | ... |
| 239 | Sedang |
| 240 | Sedang |
| 241 | Sedang |
| 242 | Sedang |

# Hasil Clustering Berdasar Target, Rayon, Skor

Tipe Genangan = Tinggi

| | _id | Lokasi | Kecamatan | \ |
|---|---|---|---|---|
| 1 | 2 | Jl. Mayjend Sungkono - Darmo Park I | Dukuh Pakis | |
| 14 | 15 | Jl. Jajar Tunggal | Wiyung | |
| 15 | 16 | Jl. H. Abdul Wahab Siamin | Dukuh Pakis | |
| 69 | 70 | Kp. Sumber Rejo | Pakal | |
| 83 | 84 | Perumahan Kutisari Indah | Tenggilis Mejoyo | |
| 98 | 99 | Kp. Merisi Selatan | Wonocolo | |
| 108 | 109 | Jl. Jetis Kulon | Wonokromo | |
| 124 | 125 | Jl. Randu | Kenjeran | |
| 128 | 129 | Jl. Kalijudan (Perbatasan Ploso) | Mulyorejo | |
| 129 | 130 | Perumahan Mulyosari | Mulyorejo | |
| 138 | 139 | Jl. Hang Tuah | Semampir | |
| 147 | 148 | Kp. Ploso Baru | Tambaksari | |
| 149 | 150 | Jl. Salak dan Sekitarnya | Tambaksari | |
| 150 | 151 | Kp. Karang Empat | Tambaksari | |
| 185 | 186 | Perempatan Jalan Dupak,Jalan Demak | Bubutan | |
| 187 | 188 | Jalan Raya Dupak,PGS s/d Pom Bensin | Bubutan | |
| 188 | 189 | Jl. Semarang | Bubutan | |
| 190 | 191 | Jl. Tanjung Sadari | Krembangan | |
| 198 | 199 | Jl. Demak | Bubutan | |
| 204 | 205 | Jl. Patemon Barat | Sawahan | |

| | Kelurahan | Rayon | Luas genangan | Tinggi genangan | \ |
|---|---|---|---|---|---|
| 1 | Dukuh Pakis | Rayon Wiyung | 0.816236 | 21.5 | |
| 14 | Jajar Tunggal | Rayon Wiyung | 0.546206 | 21.5 | |
| 15 | Dukuh Pakis | Rayon Wiyung | 0.690760 | 21.5 | |
| 69 | Sumber Rejo | Rayon Tandes | 1.970070 | 9.0 | |
| 83 | Kutisari | Rayon Jambangan | 0.917751 | 4.0 | |
| 98 | Bendul Merisi | Rayon Jambangan | 1.040533 | 10.0 | |
| 108 | Wonokromo | Rayon Jambangan | 0.509372 | 10.0 | |
| 124 | Sidotopo Wetan | Rayon Gubeng | 0.792657 | 21.5 | |
| 128 | Kalijudan | Rayon Gubeng | 0.375170 | 6.0 | |
| 129 | Kalisari | Rayon Gubeng | 0.858418 | 15.0 | |
| 138 | Ujung | Rayon Gubeng | 1.970070 | 10.0 | |
| 147 | Pacar Keling | Rayon Gubeng | 0.739720 | 4.0 | |
| 149 | Ploso | Rayon Gubeng | 0.284249 | 20.0 | |
| 150 | Ploso | Rayon Gubeng | 0.581005 | 8.0 | |
| 185 | Gundih | Rayon Genteng | 1.063949 | 4.0 | |
| 187 | Gundih | Rayon Genteng | 1.062551 | 9.0 | |
| 188 | Tembok Dukuh | Rayon Genteng | 0.483196 | 21.5 | |
| 190 | Perak Barat | Rayon Genteng | 0.285382 | 21.5 | |
| 198 | Gundih | Rayon Genteng | 1.828212 | 21.5 | |
| 204 | Petemon | Rayon Genteng | 0.539814 | 9.0 | |

| | Lama genangan | Skor Luas Genangan | Skor Tinggi Genangan | \ |
|---|---|---|---|---|
| 1 | 55 | 75 | 50 | |
| 14 | 55 | 75 | 75 | |
| 15 | 50 | 75 | 50 | |
| 69 | 55 | 100 | 25 | |
| 83 | 55 | 75 | 0 | |
| 98 | 55 | 75 | 25 | |
| 108 | 55 | 75 | 25 | |
| 124 | 55 | 75 | 50 | |
| 128 | 55 | 50 | 25 | |
| 129 | 55 | 75 | 50 | |
| 138 | 55 | 100 | 25 | |
| 147 | 55 | 75 | 0 | |
| 149 | 55 | 50 | 50 | |
| 150 | 55 | 75 | 25 | |
| 185 | 48 | 75 | 0 | |
| 187 | 48 | 75 | 25 | |
| 188 | 48 | 50 | 50 | |
| 190 | 48 | 50 | 50 | |
| 198 | 48 | 100 | 50 | |
| 204 | 48 | 75 | 25 | |

| | Skor Lama Genangan | kluster | TipePenanganan |
|---|---|---|---|
| 1 | 25 | 0 | Tinggi |
| 14 | 50 | 0 | Tinggi |
| 15 | 25 | 0 | Tinggi |
| 69 | 50 | 2 | Tinggi |
| 83 | 25 | 0 | Tinggi |
| 98 | 25 | 0 | Tinggi |
| 108 | 25 | 0 | Tinggi |
| 124 | 50 | 0 | Tinggi |
| 128 | 25 | 0 | Tinggi |
| 129 | 25 | 0 | Tinggi |
| 138 | 50 | 2 | Tinggi |
| 147 | 50 | 0 | Tinggi |
| 149 | 25 | 0 | Tinggi |
| 150 | 50 | 0 | Tinggi |
| 185 | 25 | 0 | Tinggi |
| 187 | 25 | 0 | Tinggi |
| 188 | 25 | 0 | Tinggi |
| 190 | 25 | 0 | Tinggi |
| 198 | 25 | 0 | Tinggi |
| 204 | 25 | 0 | Tinggi |

# Hasil

Dari Hasil yang didapatkan, berikut rincian dari beberapa data dengan tipe penanganan :

Tinggi, Sedang dan Rendah

-> Terdapat 20 data atau daerah dengan tipe genangan tinggi
-> Terdapat 166 data atau daerah dengan tipe genangan sedang
-> Terdapat 58 data atau daerah dengan tipe genangan rendah

# Kesimpulan

1. Berdasarkan hasil pengujian data menggunakan Google Collab dan mengimplementasikan metode clustering K-Means, genangan pada daerah di Surabaya per tahun 2022 dapat dikelompokkan menjadi tiga cluster, yaitu daerah dengan genangan tinggi, sedang, dan rendah.

2. Dengan mengimplementasikan dua metode clustering yang berbeda maka didapatkan hasil yang berbeda pula.

# Saran

Penelitian selanjutnya dapat dilakukan secara lebih detail dengan menggunakan metode clustering yang lebih kompleks. Dengan itu, maka hasil keluaran daerah yang terdampak genangan air akan terkelompokkan lebih jelas dan baik.

# Terimakasih.