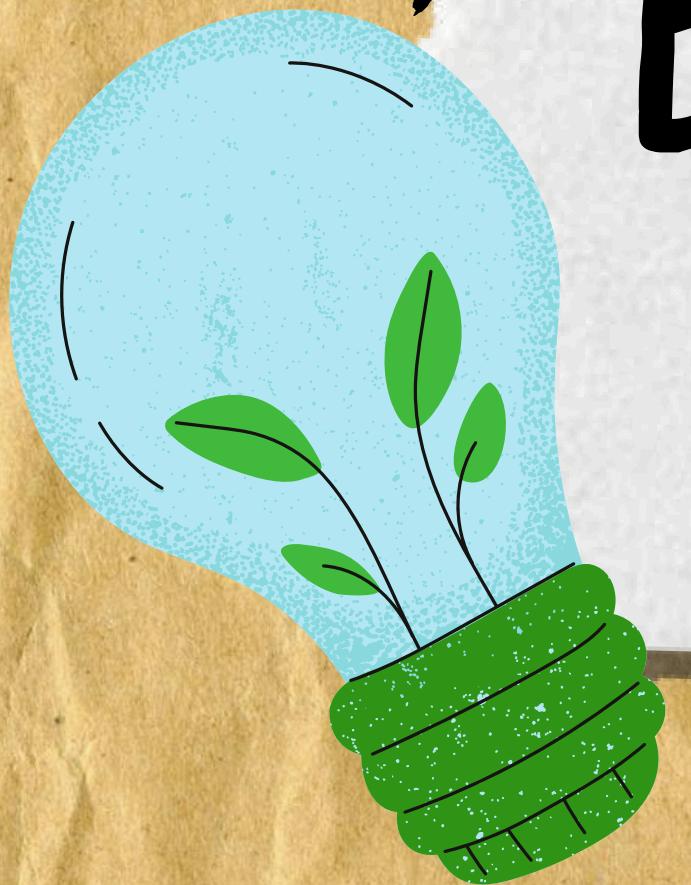




Clustering

BENCANA BANJIR DI DKI JAKARTA



Kelompok 8

2502047190 - Muhammad Irsyad S.K

2540122281 - Bryan Wijanarko

2540117804 - Vincentius Gunawan



Latar Belakang

Bencana banjir merupakan salah satu masalah serius yang dihadapi oleh banyak kota besar di seluruh dunia, termasuk DKI Jakarta. Untuk mengatasi masalah ini, diperlukan pemahaman yang lebih mendalam mengenai daerah-daerah yang paling sering terdampak banjir. Pada penelitian ini, pendekatan clustering digunakan untuk mengelompokkan kecamatan-kecamatan di DKI Jakarta berdasarkan dampak yang diakibatkan dari banjir itu sendiri. Dengan melakukan clustering, dapat diidentifikasi kecamatan-kecamatan yang memerlukan perhatian khusus dalam upaya mitigasi bencana dan perencanaan kota yang lebih baik.



Latar Belakang



Acuan dari penelitian yang kami lakukan adalah penelitian-penelitian sebelumnya yang berkaitan dengan bencana banjir ataupun teknik clustering. Terdapat penelitian yaitu memprediksi banjir di DKI Jakarta menggunakan algoritma K-Means dan Random Forest. Pada penelitian ini, algoritma K-Means digunakan untuk mengelompokan data atau membuat cluster berdasarkan kemiripan karakteristiknya. Kemudian algoritma Random Forest digunakan untuk memproses data hasil clustering dari algoritma k-means untuk membuat prediksi bencana banjir. Penelitian ini dilakukan oleh mahasiswa magister Teknik informatika, universitas amikom Yogyakarta.

- Wasis Haryo Sasoko, Eka Wahyu Pujihartono, Ruby Haris, Adela Yuza Kania, Kusrini, Kusnawi, 2024, "Prediksi Banjir Di Dki Jakarta Dengan Menggunakan Algoritma K-Means Dan Random Forest", Vol.05 No.01, <https://ejurnalunsam.id/index.php/jicom/article/view/8153/4697>

Sustainable Development Goals



SDG 11 : Bertujuan untuk membuat kota dan permukiman inklusif, aman, tangguh, dan berkelanjutan

- Dapat mengurangi jumlah kematian dan orang yang terdampak yang diakibatkan oleh bencana
- Dapat Mengurangi dampak lingkungan negatif per kapita dari kota-kota
- Meningkatkan jumlah kota dan permukiman manusia yang mengadopsi dan mengimplementasikan kebijakan dan rencana yang mengarah pada mitigasi dan adaptasi perubahan iklim dan bencana

About Dataset

Kami mengambil dataset dari website satu data jakarta dan sumber dari dataset ini adalah dari Badan Penanggulangan Bencana Daerah

Source Data : https://satudata.jakarta.go.id/open-data/detail?kategori=dataset&page_url=data-kejadian-bencana-banjir-tahun-2023&data_no=1



Satu Data
JAKARTA



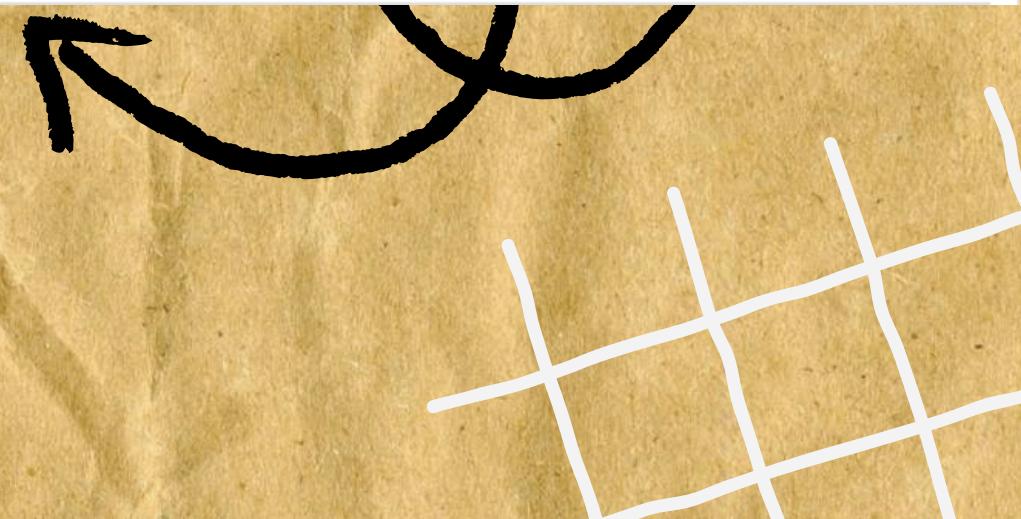
Komponen Dataset

- periode_data : Penjelasan Periode Data 3 Bulan Sekali
- bulan : bulan
- wilayah : wilayah
- kecamatan : kecamatan
- kelurahan : kelurahan
- jumlah_rata_rata_ketinggian_air : jumlah rata rata ketinggian air
- jumlah_rw_terdampak : jumlah rw terdampak
- jumlah_kk_terdampak : jumlah kk terdampak
- jumlah_jiwa_terdampak : jumlah jiwa terdampak
- jumlah_kejadian : jumlah kejadian
- jumlah_korban_meninggal : jumlah korban meninggal
- jumlah_korban_luka : jumlah korban luka
- jumlah_pengungsi : jumlah pengungsi
- jumlah_tempat_pengungsian : jumlah tempat pengungsian
- nilai_kerugian : nilai kerugian

periode_data	bulan	wilayah	kecamatan	kelurahan
3 Triwulan	2	JAKARTA	KEBAYORAN	GROGOL
		SELATAN	LAMA	SELATAN
3 Triwulan	2	JAKARTA	KEBAYORAN	KEBAYORAN
		SELATAN	LAMA	LAMA SELATAN
3 Triwulan	2	JAKARTA	MAMPANG	PELA
		SELATAN	PRAPATAN	MAMPANG
3 Triwulan	2	JAKARTA	MAMPANG	TEGAL
		SELATAN	PRAPATAN	PARANG
3 Triwulan	2	JAKARTA	MAMPANG	KUNINGAN
		SELATAN	PRAPATAN	BARAT
3 Triwulan	2	JAKARTA	SELATAN	PESANGGRAHAN
		TEBET	ULUJAMI	
3 Triwulan	2	JAKARTA	SELATAN	BUKIT DURE
		TEBET		MANGGAR
3 Triwulan	2	JAKARTA	SELATAN	TEBET
		SELATAN	TEBET	KEBON BAR
3 Triwulan	2	JAKARTA	SELATAN	PANCORAN
		SELATAN	PANCORAN	RAWAJATI
3 Triwulan	2	JAKARTA	SELATAN	PENGADE
		TIMUR	KRAMAT JATI	KAMPUNG TENGAH
3 Triwulan	2	JAKARTA	TIMUR	KRAMAT JATI
		TIMUR	KRAMAT JATI	CILILITAN
3 Triwulan	2	JAKARTA	TIMUR	KRAMAT JATI
		TIMUR	KRAMAT JATI	CAWANG
3 Triwulan	2	JAKARTA	TIMUR	KRAMAT JATI
		TIMUR	KRAMAT JATI	DUKUH
3 Triwulan	2	JAKARTA	TIMUR	KRAMAT JATI
		TIMUR	KRAMAT JATI	BALEKAMB
3 Triwulan	2	JAKARTA	TIMUR	JATINEGARA
		TIMUR	JATINEGARA	BIDARA CI
3 Triwulan	2	JAKARTA	TIMUR	JATINEGARA
		TIMUR	JATINEGARA	KAMPUNG MELAYU
3 Triwulan	2	JAKARTA	TIMUR	JATINEGARA
		TIMUR	JATINEGARA	CIPINANG MUARA
3 Triwulan	2	JAKARTA	TIMUR	CIRACAS
		TIMUR	CIRACAS	CIRACAS
3 Triwulan	2	JAKARTA	TIMUR	CIRACAS
		TIMUR	CIRACAS	SUSUKAN
3 Triwulan	2	JAKARTA	TIMUR	CIRACAS
		TIMUR	CIRACAS	RAMBUTAI
3 Triwulan	2	JAKARTA	TIMUR	CAKUNG
		TIMUR	CAKUNG	CAKUNG TIMUR
3 Triwulan	2	JAKARTA	TIMUR	CAKUNG
		CAKUNG	CAKUNG	RAWA TER

periode_data	triwulan	bulan	wilayah	kecamatan	kelurahan	jumlah_rata_rata_ketinggian_air	jumlah_rw_terdampak	jumlah_kk_terdampak
2024 Triwulan 1	1	1	jakarta barat	cengkareng	cengkareng timur	30 s.d 40 cm	2	0
2024 Triwulan 1	1	1	jakarta barat	kembangan	joglo	30 s.d 40 cm	1	0
2024 Triwulan 1	1	1	jakarta barat	kali deres	kamal	30 cm	1	0
2024 Triwulan 1	1	1	jakarta barat	cengkareng	kapuk	30 cm	1	0
2024 Triwulan 1	1	1	jakarta barat	cengkareng	kedaung kali angke	30 s.d 40 cm	2	0
2024 Triwulan 1	1	1	jakarta barat	kebon jeruk	kedoya selatan	30 cm	1	0
2024 Triwulan 1	1	1	jakarta barat	kembangan	kembangan selatan	30 s.d 100 cm	3	0
2024 Triwulan 1	1	1	jakarta barat	kembangan	kembangan utara	30 s.d 100 cm	1	24
2024 Triwulan 1	1	1	jakarta barat	kali deres	pegadungan	40 cm	1	0
2024 Triwulan 1	1	1	jakarta barat	cengkareng	rawa buaya	30 s.d 100 cm	3	0

jumlah_jiwa_terdampak	jumlah_kejadian	jumlah_korban_meninggal	jumlah_korban_luka	jumlah_pengungsi	jumlah_tempat_pengungsian	nilai_kerugian
0	2	0	0	0	0	0
0	1	0	0	0	0	0
0	1	0	0	0	0	0
0	1	0	0	0	0	0
0	3	0	0	0	0	0
0	1	0	0	0	0	0
0	2	0	0	0	0	0
95	1	0	0	30	1	0

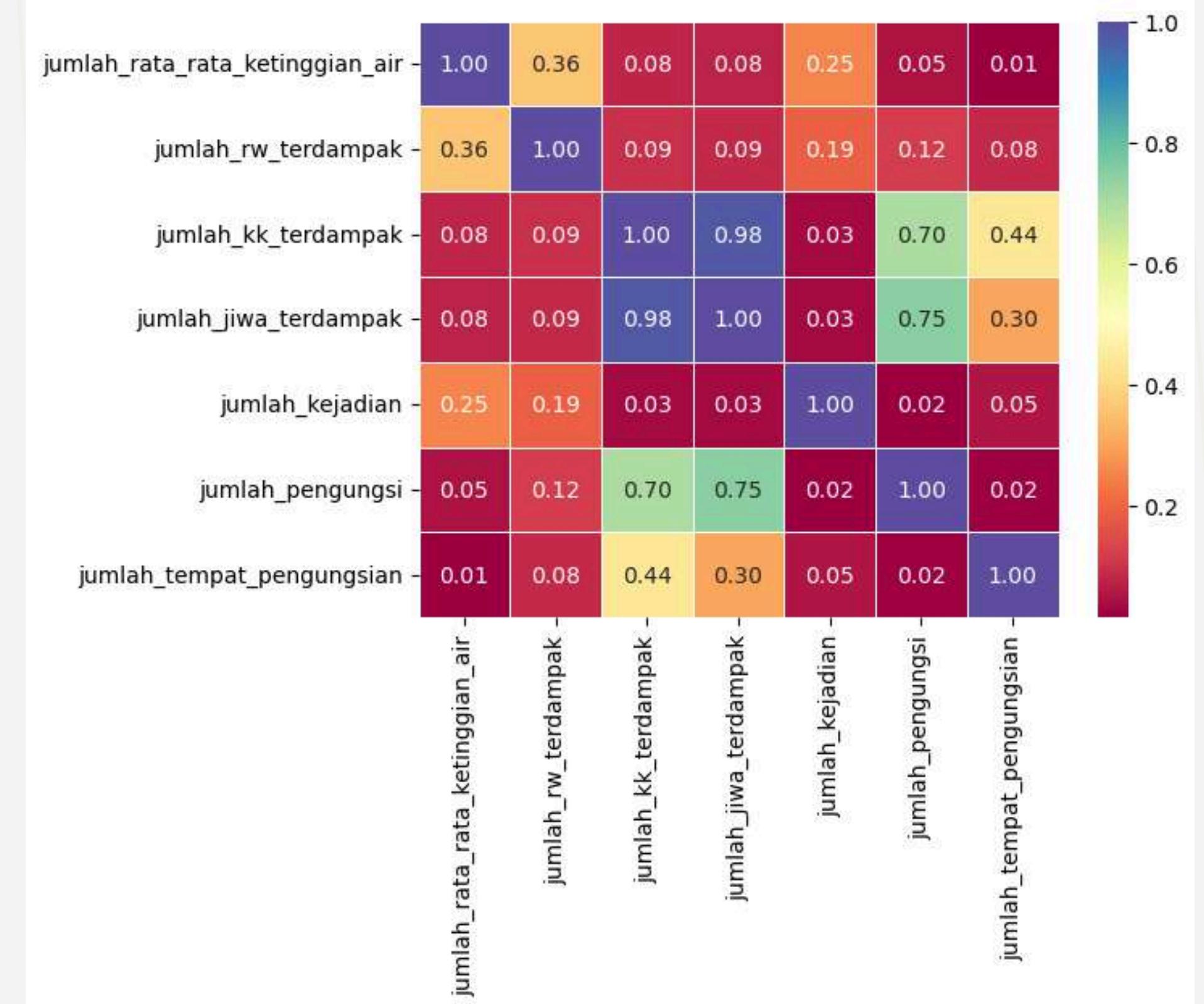


Data Banjir 2023-2024

- Heatmap

Heatmap menunjukkan korelasi antara data-data yang dimiliki dalam dataset.

Dari heatmap terdapat data yang memiliki value korelasi yang besar terhadap satu samalain.



- periode_data : Penjelasan Periode Data 3 Bulan Sekali
- bulan : bulan
- wilayah : wilayah
- kecamatan : kecamatan
- kelurahan : kelurahan
- jumlah_rata_rata_ketinggian_air : jumlah rata rata ketinggian air
- jumlah_rw_terdampak : jumlah rw terdampak
- jumlah_kk_terdampak : jumlah kk terdampak
- jumlah_jiwa_terdampak : jumlah jiwa terdampak
- jumlah_kejadian : jumlah kejadian
- jumlah_korban_meninggal : jumlah korban meninggal
- jumlah_korban_luka : jumlah korban luka
- jumlah_pengungsi : jumlah pengungsi
- jumlah_tempat_pengungsian : jumlah tempat pengungsian
- nilai_kerugian : nilai kerugian

Drop Kolom ↪

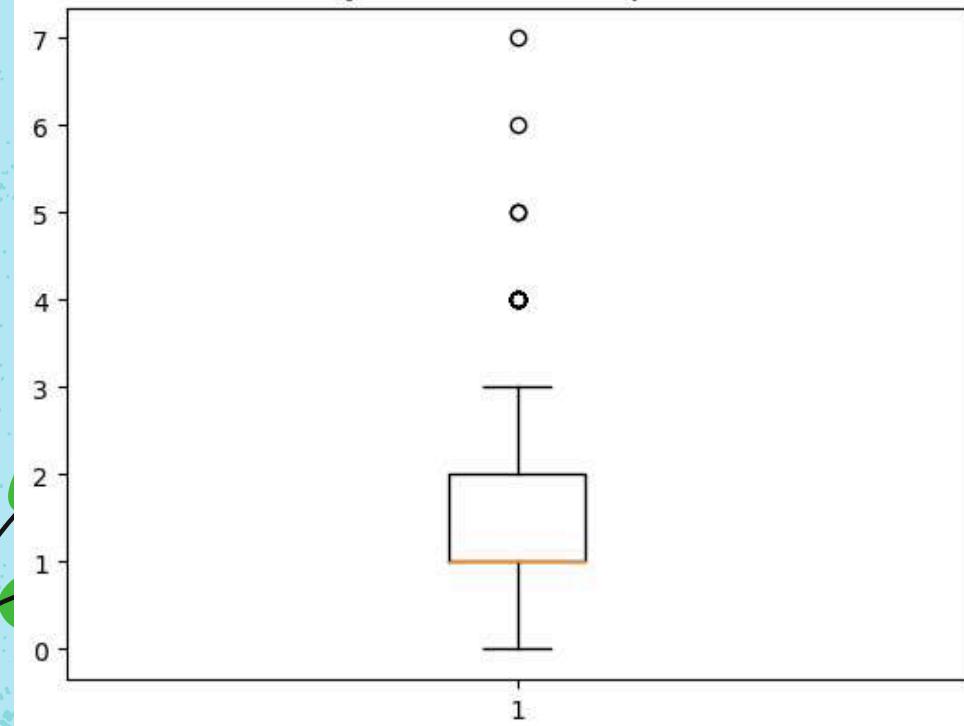
- Group By Kecamatan

Tabel nantinya akan di group by kecamatan, sehingga wilayah, kelurahan tidak diperlukan. Di sisi lain, bulan tidak diperlukan karena periode data sudah cukup

- Drop Kolom Kosong

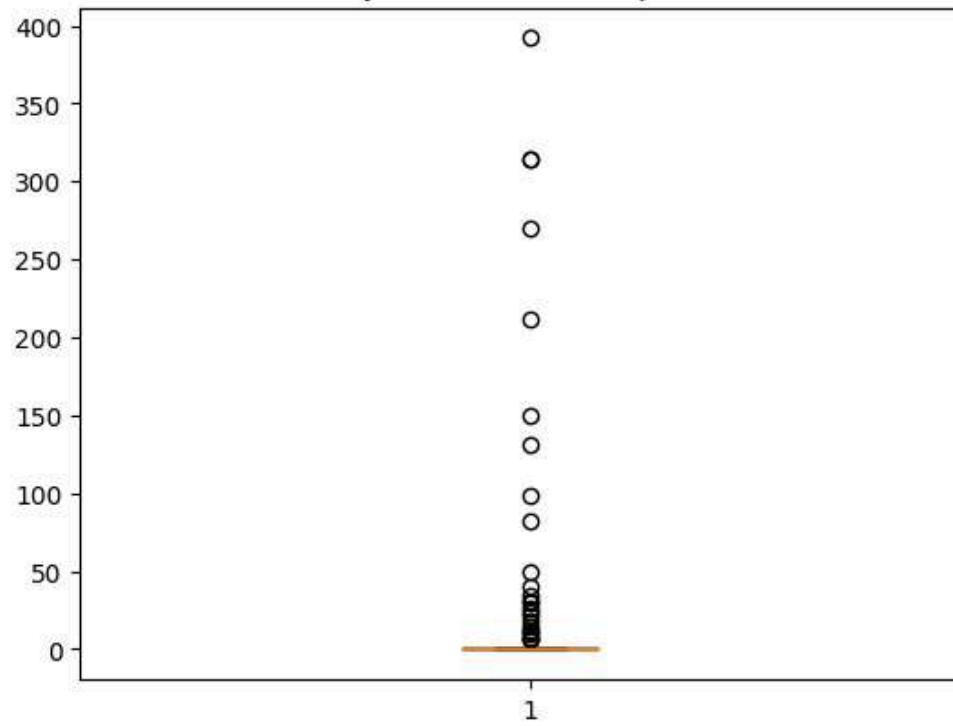
```
print("Sum kolom jumlah meninggal: ", banjir_df['jumlah_korban_meninggal'].sum())
print("Sum kolom jumlah meninggal: ", banjir_df['jumlah_korban_luka'].sum())
print("Sum kolom jumlah meninggal: ", banjir_df['nilai_kerugian'].sum())
✓ 0.0s
Sum kolom jumlah meninggal: 0
Sum kolom jumlah meninggal: 0
Sum kolom jumlah meninggal: 0
```

Jumlah RW Terdampak



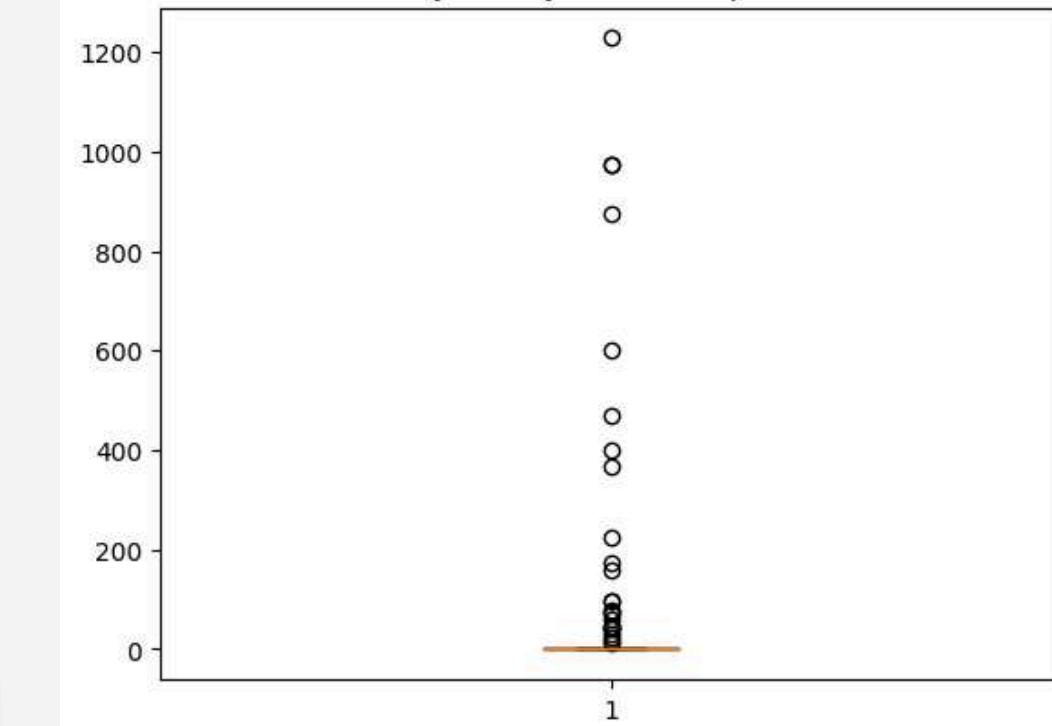
Jumlah RW Terdampak

Jumlah KK Terdampak



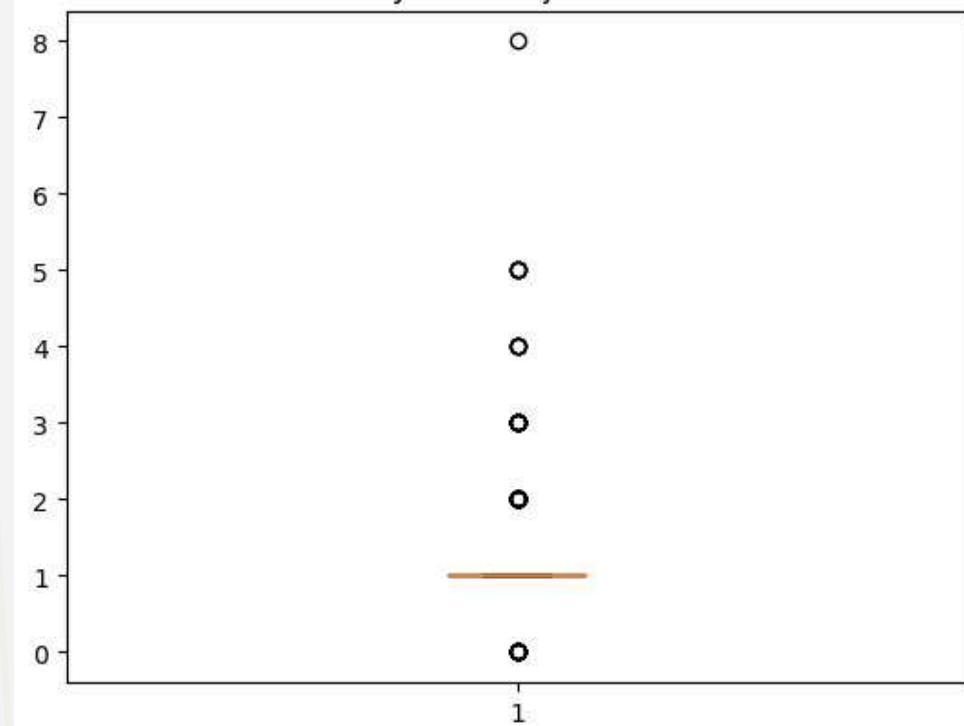
Jumlah KK Terdampak

Jumlah Jiwa Terdampak



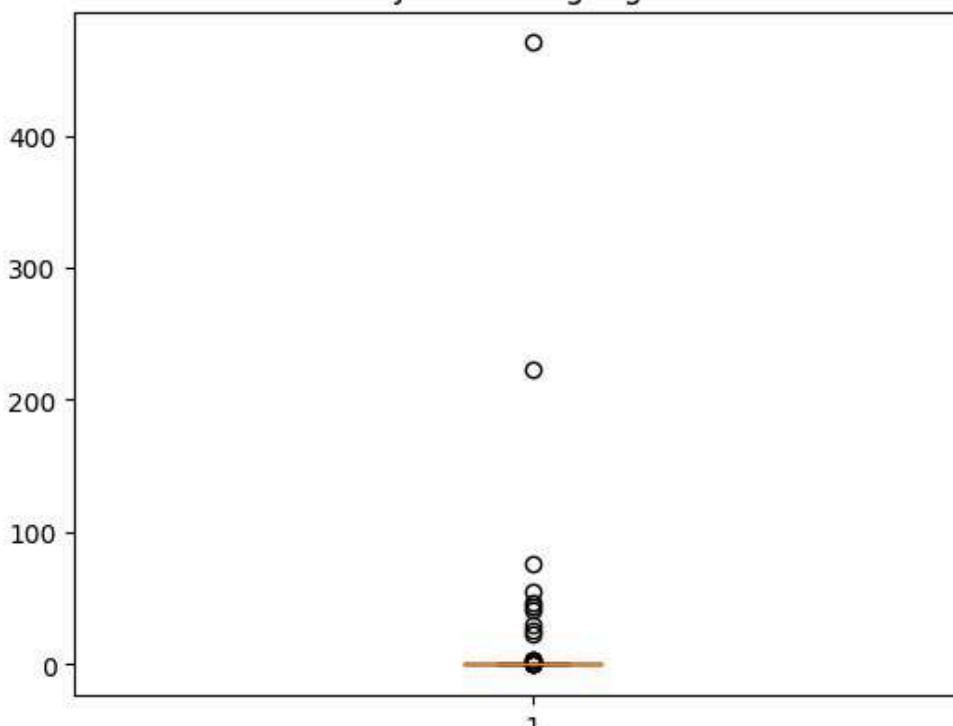
Jumlah Jiwa Terdampak

Jumlah Kejadian



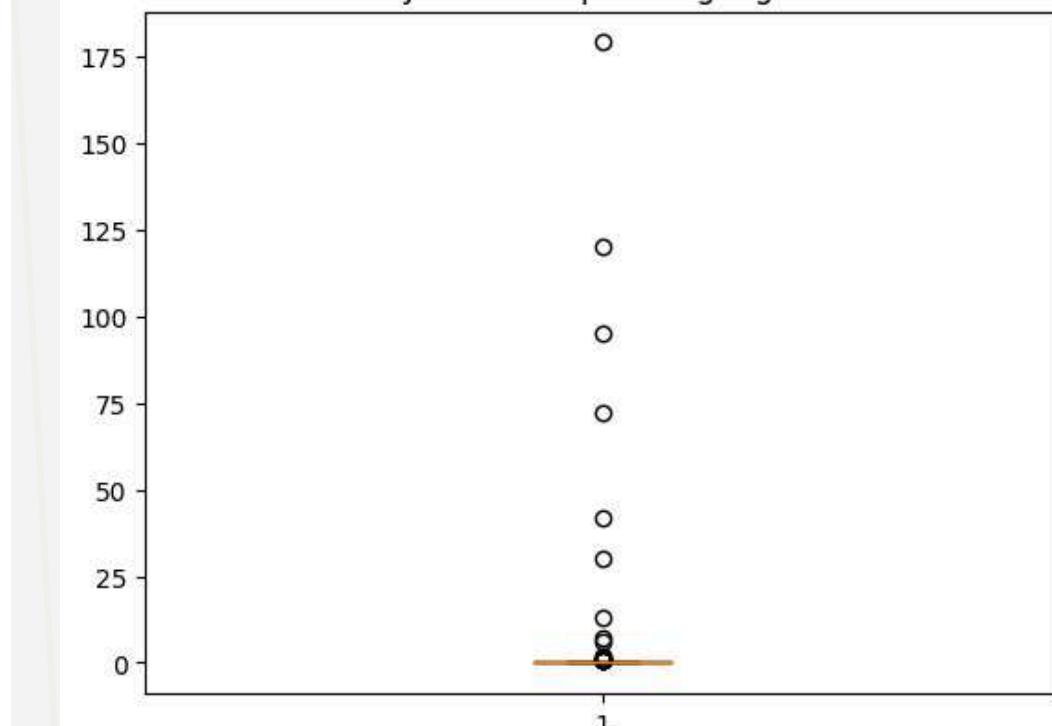
Jumlah Kejadian

Jumlah Pengungsi

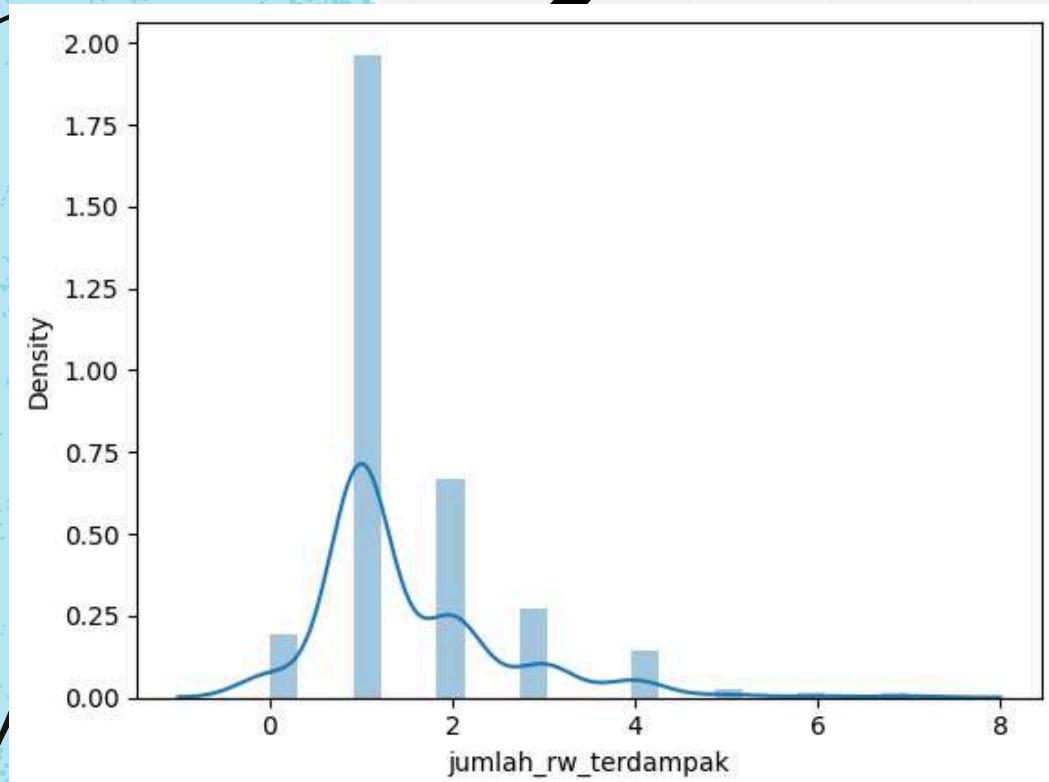


Jumlah Pengungsi

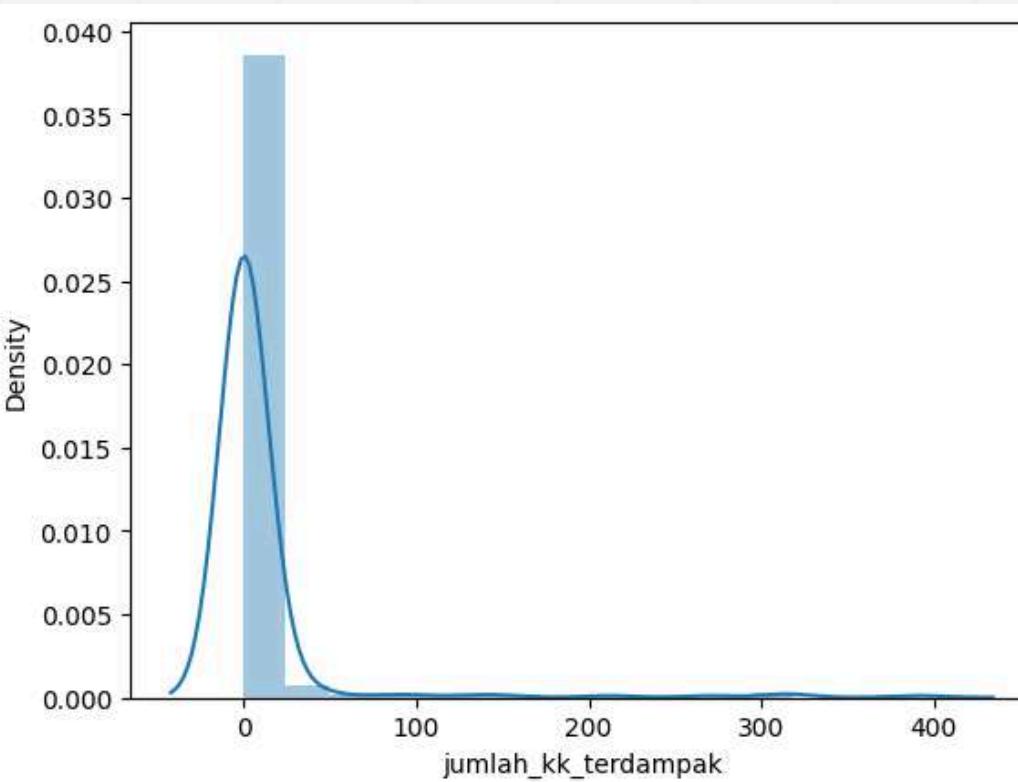
Jumlah Tempat Pengungsian



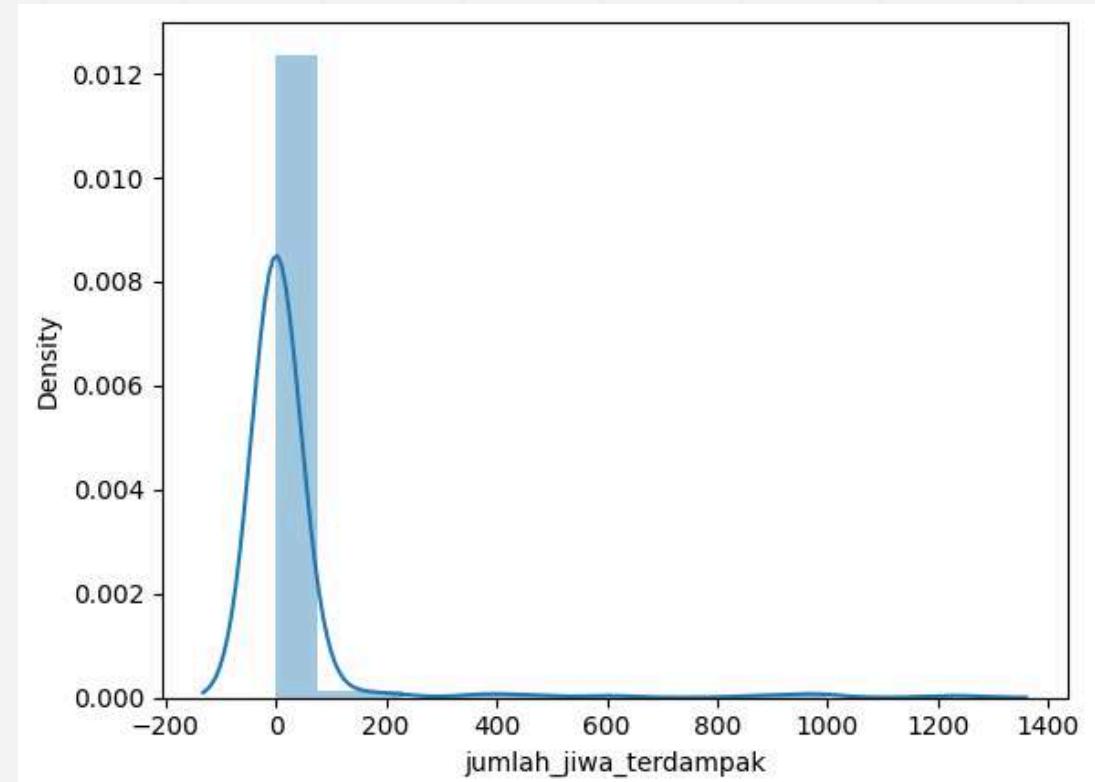
Jumlah Tempat Pengungsian



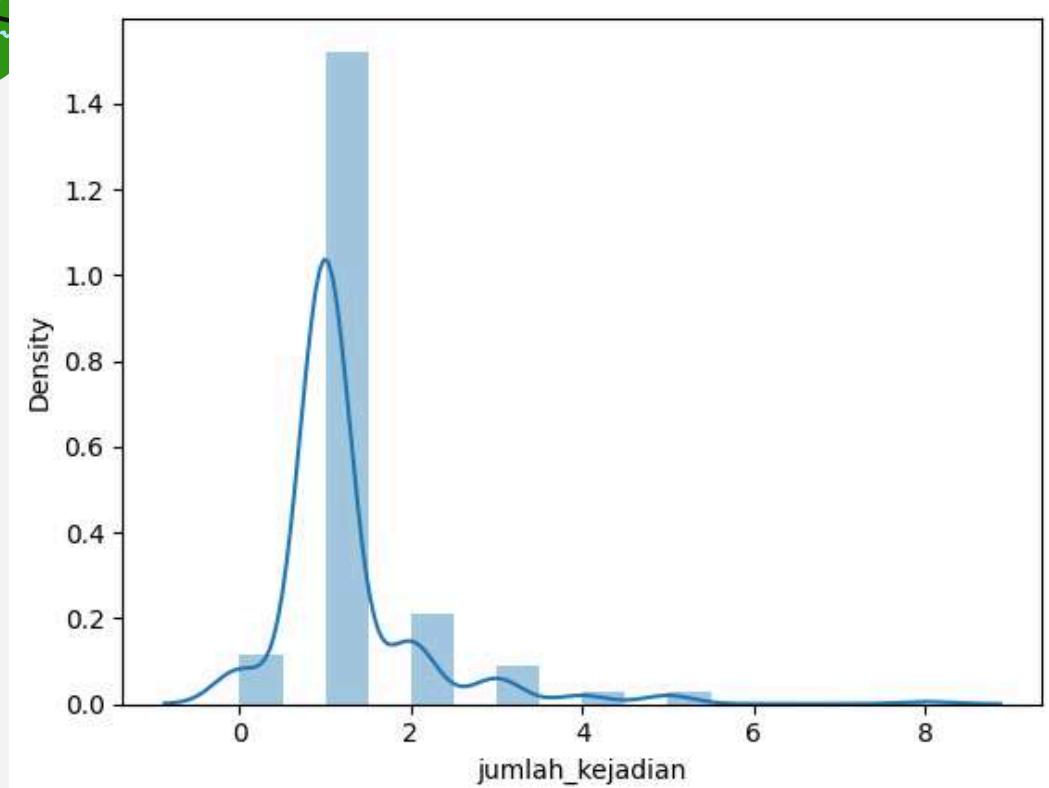
Jumlah RW Terdampak



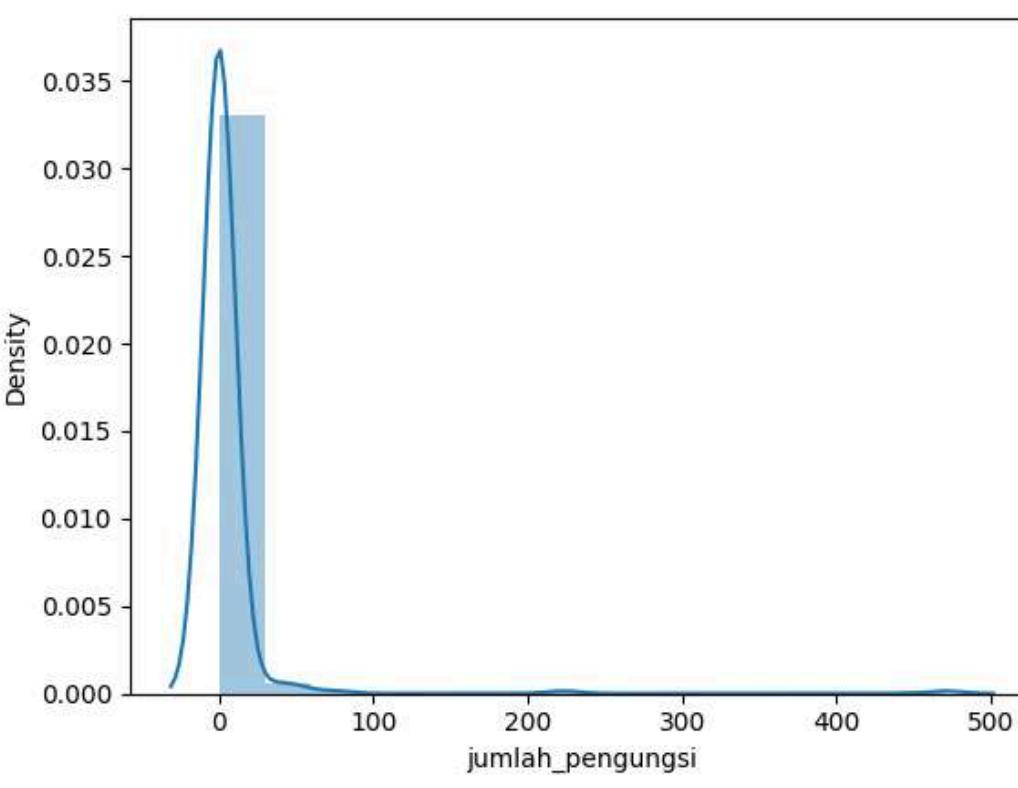
Jumlah KK Terdampak



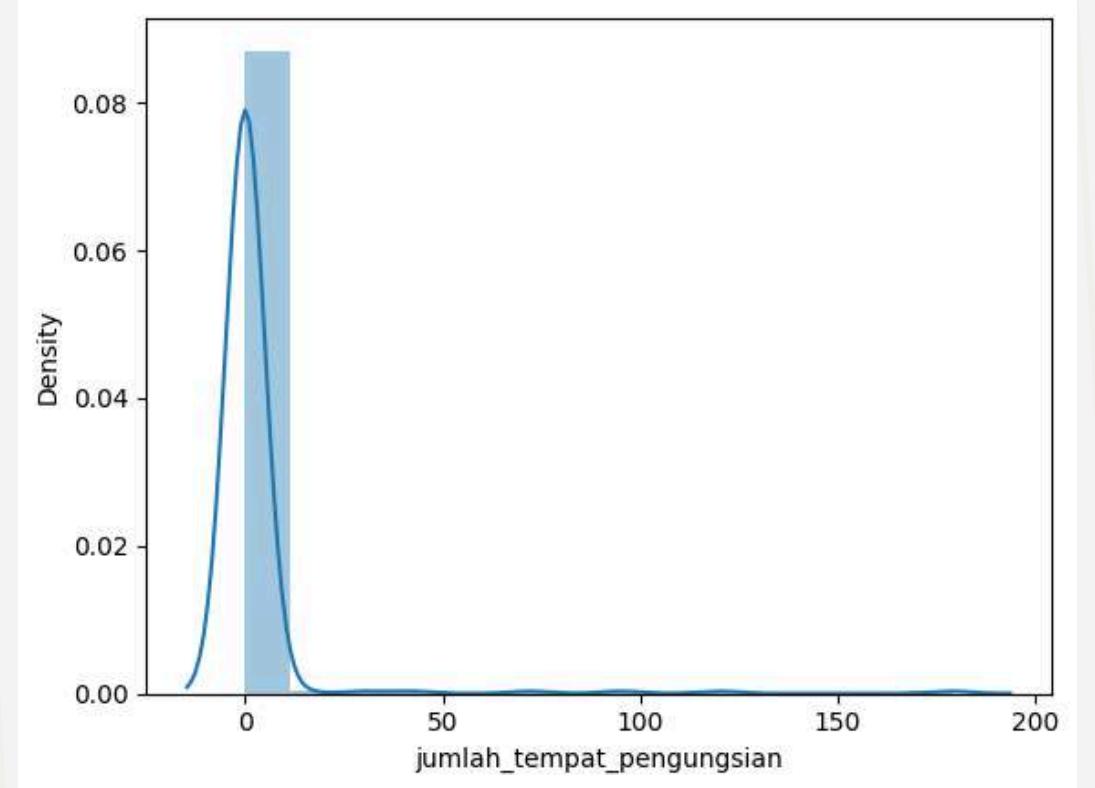
Jumlah Jiwa Terdampak



Jumlah Kejadian



Jumlah Pengungsian



Komponen Dataset Untuk Clustering

	kecamatan	jumlah_rw_terdampak	jumlah_kk_terdampak	jumlah_jiwa_terdampak	jumlah_kejadian	jumlah_pengungsi	jumlah_tempat_pengungsian
0	CAKUNG	9.0	497.0	1524.0	8.0	47.0	1.0
1	CENGKARENG	38.0	153.0	491.0	28.0	227.0	276.0
2	CILANDAK	13.0	18.0	54.0	11.0	0.0	0.0
3	CILINGCING	11.0	162.0	530.0	8.0	121.0	44.0
4	CIPAYUNG	8.0	0.0	0.0	5.0	0.0	0.0

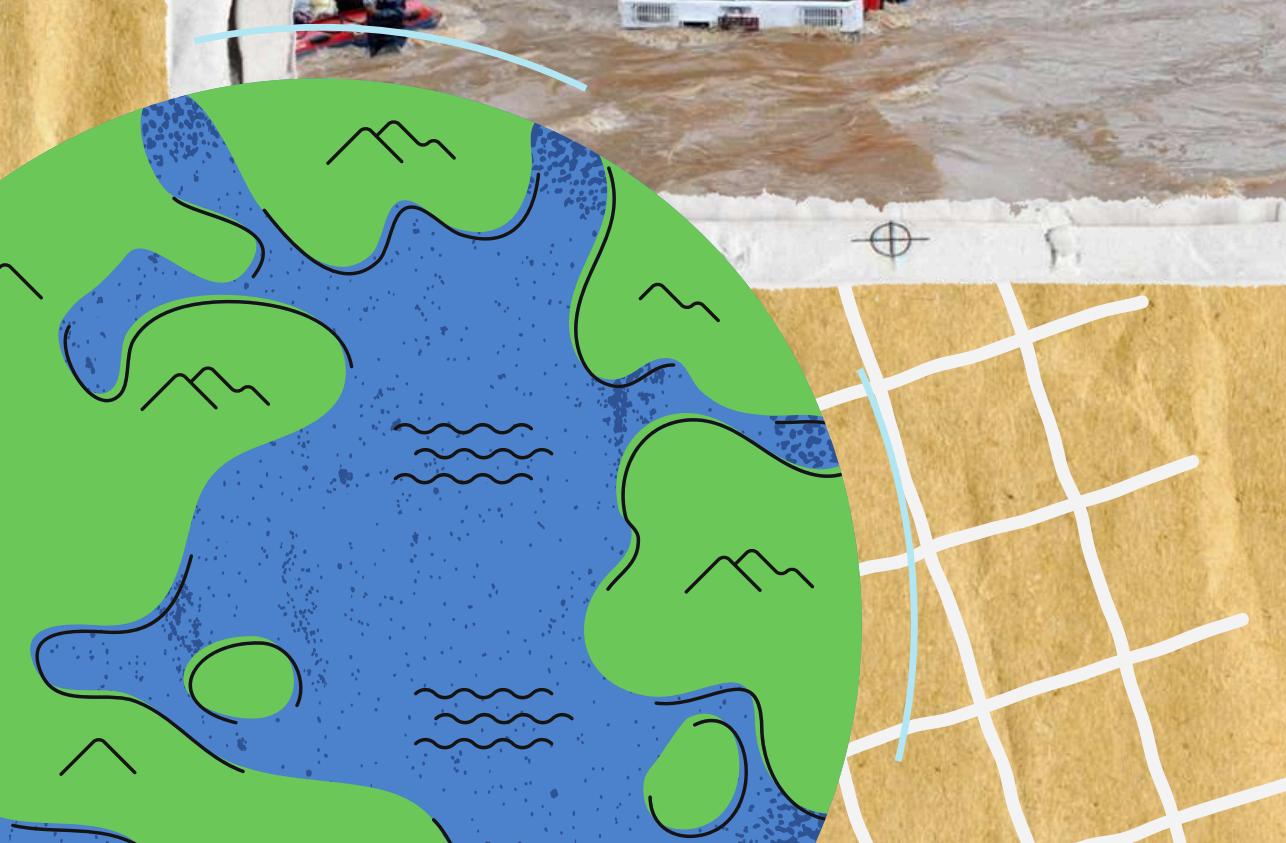
- kecamatan : berisi unique value, dengan jumlah kecamatan 32
- jumlah_rw_terdampak : jumlah dari tahun 2023 sampai 2024 Q1
- jumlah_kk_terdampak : jumlah dari tahun 2023 sampai 2024 Q1
- jumlah_jiwa_terdampak : jumlah dari tahun 2023 sampai 2024 Q1
- jumlah_kejadian : jumlah dari tahun 2023 sampai 2024 Q1
- jumlah_pengungsi : jumlah dari tahun 2023 sampai 2024 Q1
- jumlah_tempat_pengungsian : jumlah dari tahun 2023 sampai 2024 Q1

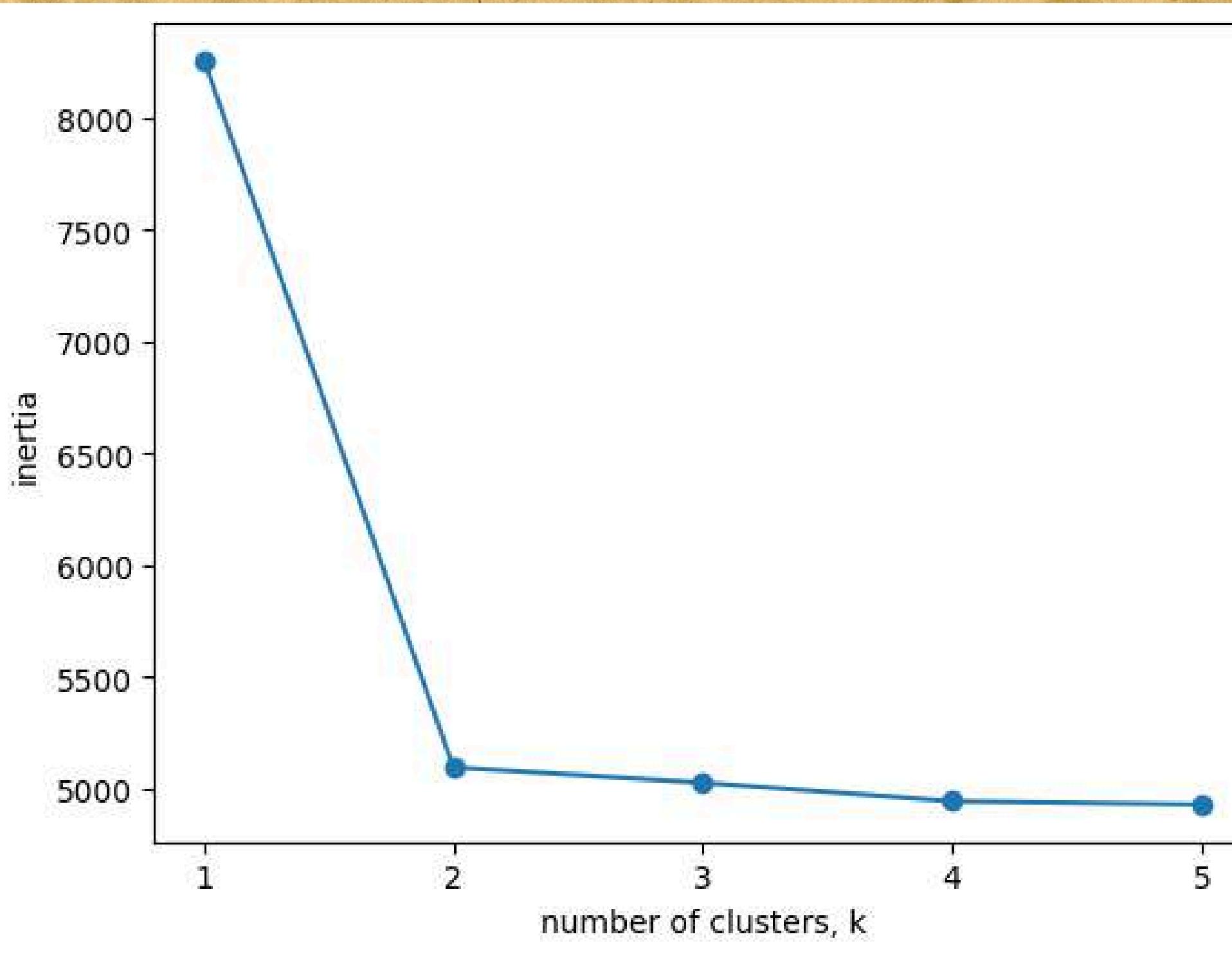


Methodology

Kami menggunakan K-Medoids sebagai algoritma dari penelitian clustering kami. K-Medoids adalah algoritma klustering yang mirip dengan K-Means, tetapi alih-alih menggunakan mean dari titik-titik dalam klaster sebagai pusat klaster. K-Medoids menggunakan titik yang paling mewakili (medoid) yang memiliki jarak total terkecil ke titik-titik lain dalam klaster.

Keuntungan utama K-Medoids dibandingkan K-Means adalah ketahanannya terhadap outlier, karena medoid dipilih dari titik data yang ada, sehingga outlier tidak memiliki dampak yang signifikan seperti pada K-Means yang menggunakan mean yang dapat dipengaruhi oleh nilai ekstrem.

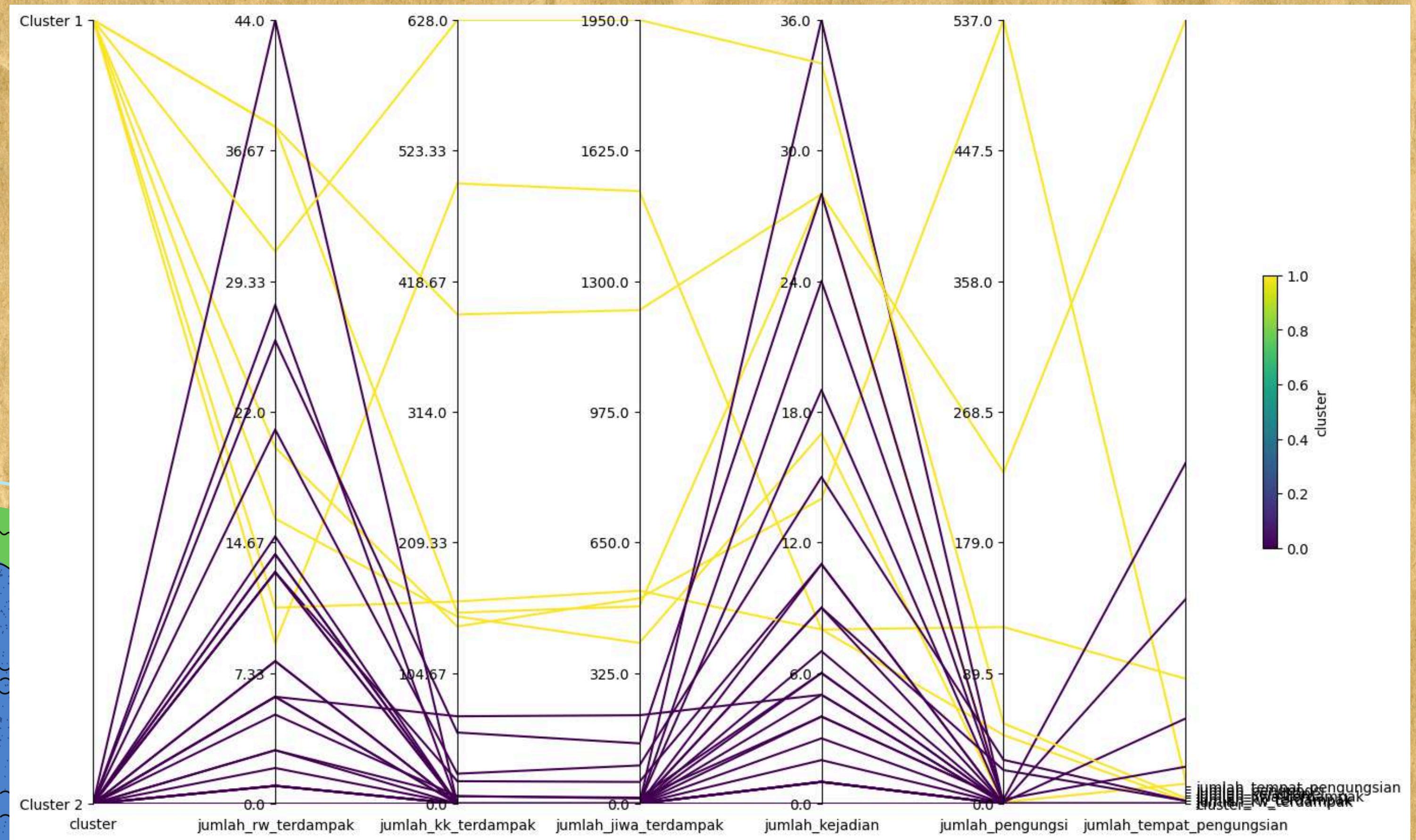




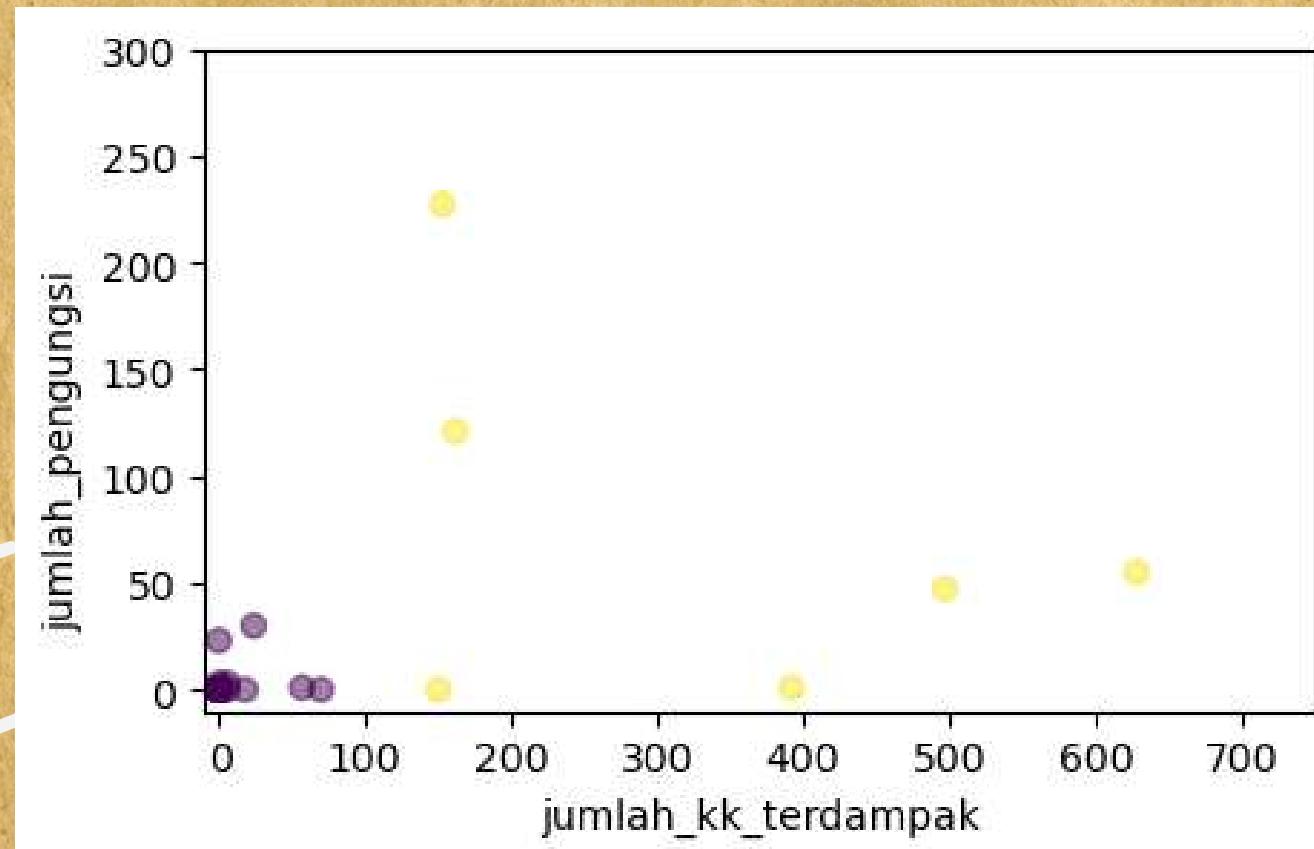
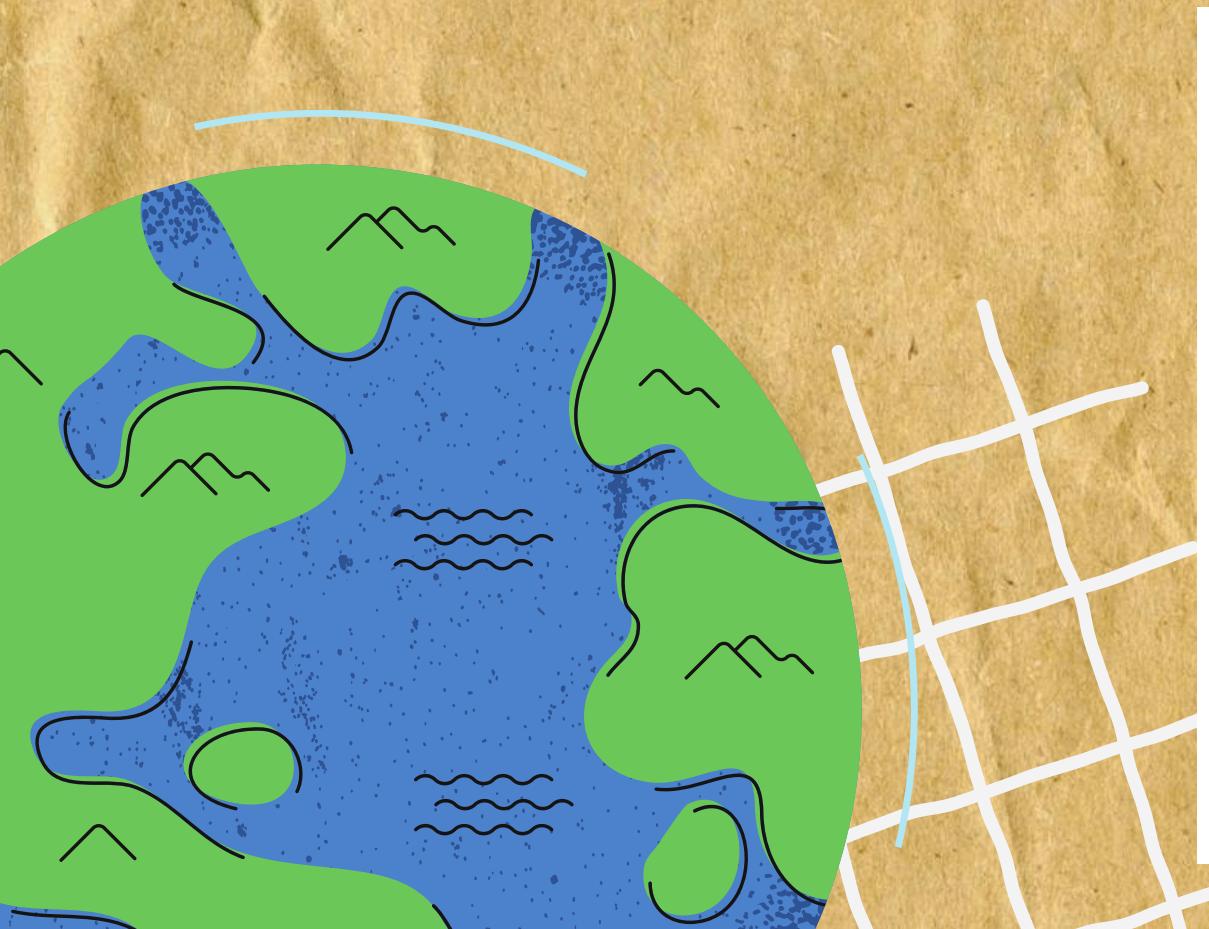
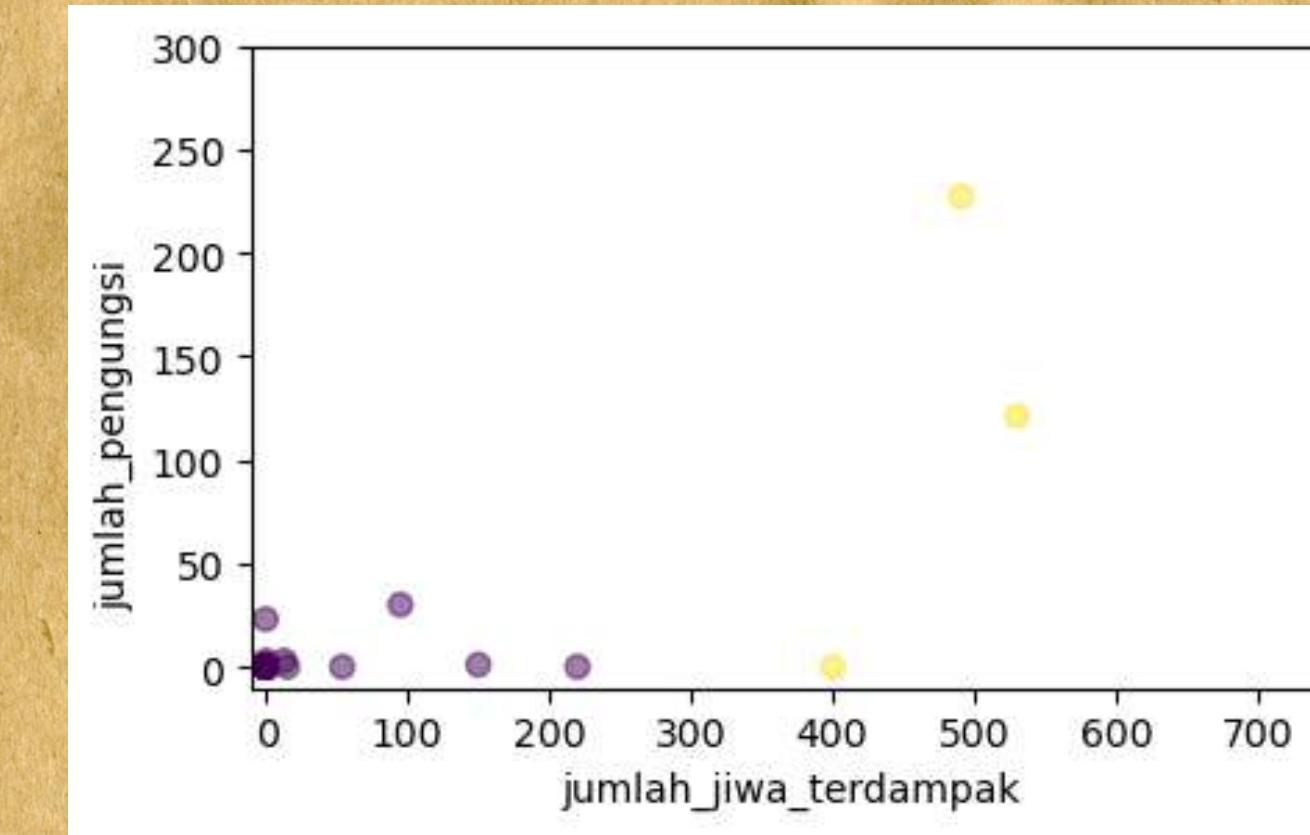
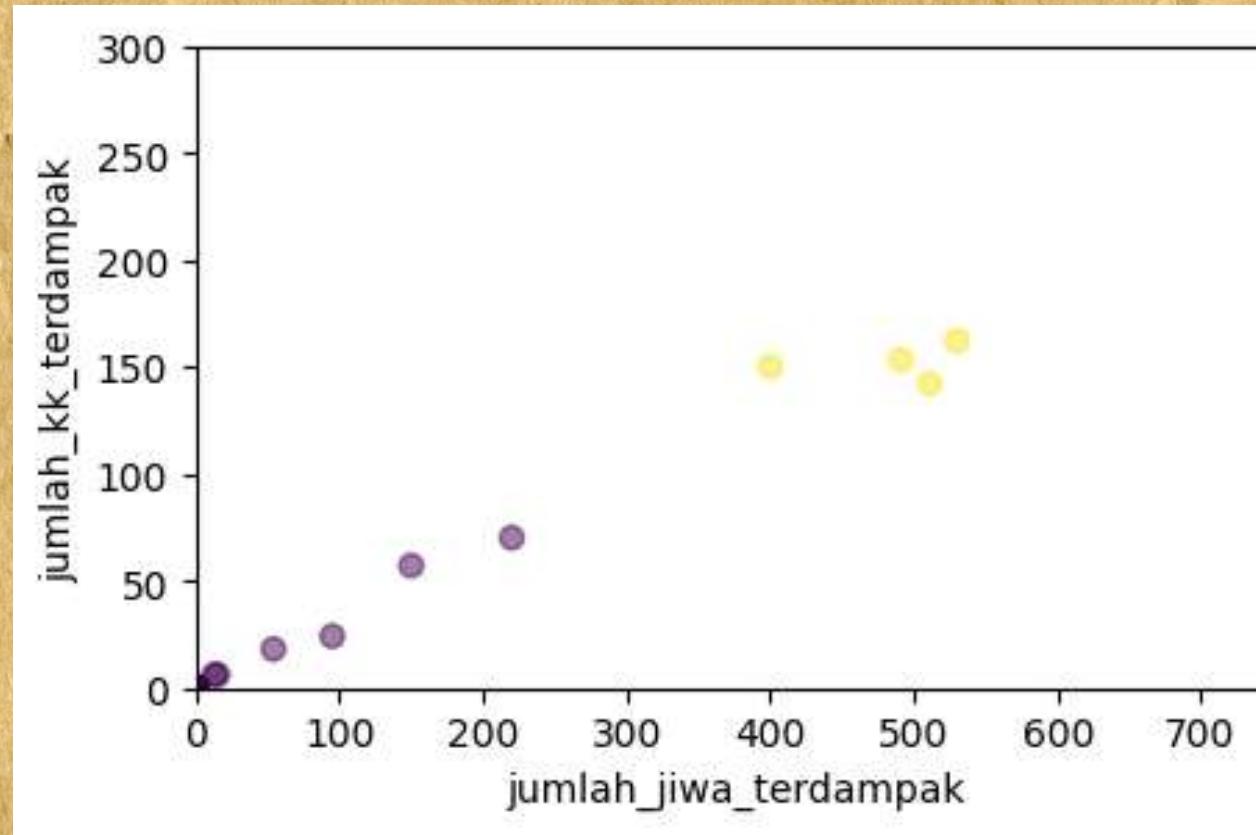
Banyak Cluster

Menggunakan inertia, kami memutuskan menggunakan 2 cluster

Parallel Coordinates

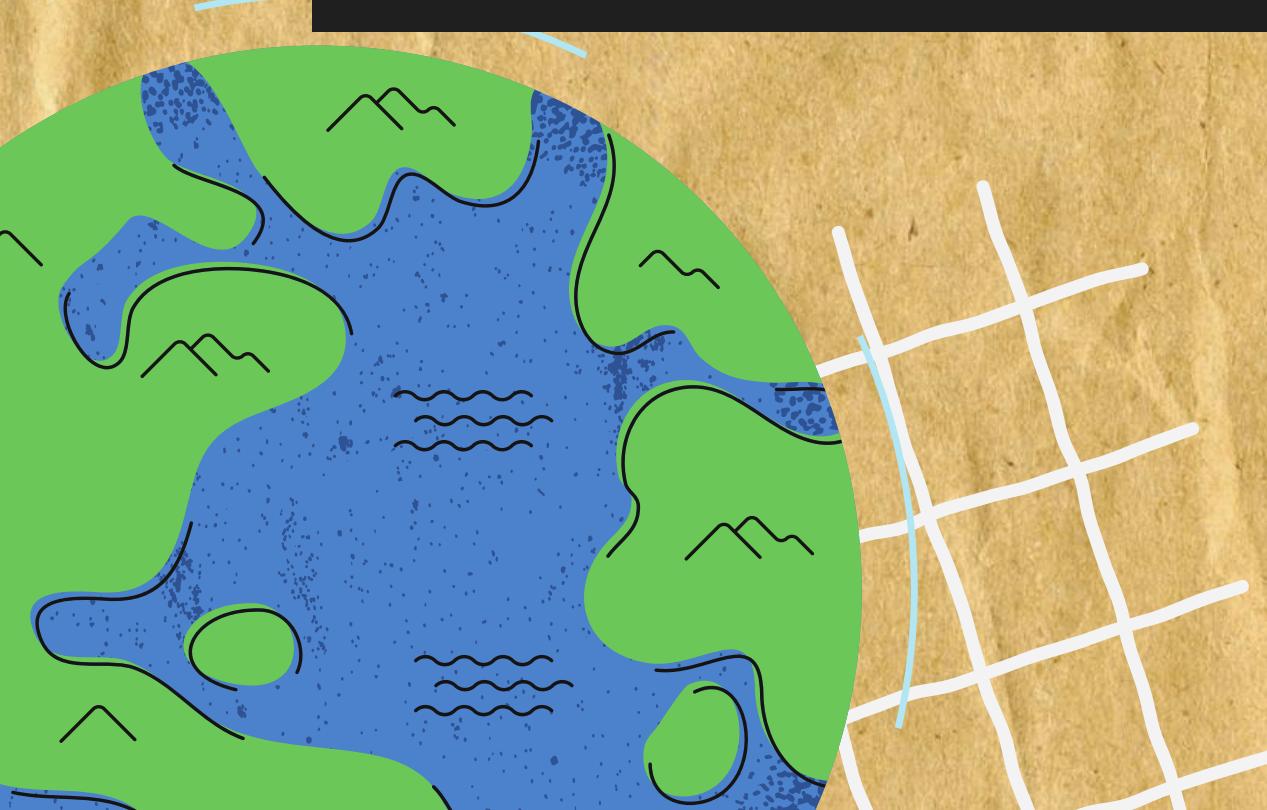


Scatter Plot



Hasil Cluster

	cluster	kecamatan
0	Cluster 1	CAKUNG
1	Cluster 1	CENGKARENG
3	Cluster 1	CILINCING
9	Cluster 1	JATINEGARA
10	Cluster 1	KALI DERES
13	Cluster 1	KEBAYORAN LAMA
24	Cluster 1	PASAR MINGGU



	cluster	kecamatan
2	Cluster 2	CILANDAK
4	Cluster 2	CIPAYUNG
5	Cluster 2	CIRACAS
6	Cluster 2	DUREN SAWIT
7	Cluster 2	GROGOL PETAMBURAN
8	Cluster 2	JAGAKARTA
11	Cluster 2	KALIDERES
12	Cluster 2	KEBAYORAN BARU
14	Cluster 2	KEBON JERUK
15	Cluster 2	KELAPA GADING
16	Cluster 2	KEMBANGAN
17	Cluster 2	KOJA
18	Cluster 2	KRAMAT JATI
19	Cluster 2	MAKASAR

20	Cluster 2	MAMPANG PRAPATAN
21	Cluster 2	NIHIL
22	Cluster 2	PALMERAH
23	Cluster 2	PANCORAN
25	Cluster 2	PENJARINGAN
26	Cluster 2	PESANGGRAHAN
27	Cluster 2	PULOGADUNG
28	Cluster 2	TAMAN SARI
29	Cluster 2	TANAH ABANG
30	Cluster 2	TANJUNG PRIOK
31	Cluster 2	TEBET

Training Model

Summary LR

Data yang digunakan sebagai variabel dependen adalah Cluster 1



K-Neighbour

	precision	recall	f1-score	support
False	0.66	0.67	0.67	70
True	0.41	0.40	0.41	40
accuracy			0.57	110
macro avg	0.54	0.54	0.54	110
weighted avg	0.57	0.57	0.57	110

Logistic Regression

	precision	recall	f1-score	support
False	0.69	0.87	0.77	70
True	0.57	0.30	0.39	40
accuracy				0.66
macro avg	0.63	0.59	0.58	110
weighted avg	0.64	0.66	0.63	110

Naive Bayes

	precision	recall	f1-score	support
False	0.67	0.91	0.78	70
True	0.60	0.23	0.33	40
accuracy			0.66	110
macro avg	0.64	0.57	0.55	110
weighted avg	0.65	0.66	0.61	110

Support Vector Machine

	precision	recall	f1-score	support
False	0.67	1.00	0.80	70
True	1.00	0.12	0.22	40
accuracy				0.68
macro avg	0.83	0.56	0.51	110
weighted avg	0.79	0.68	0.59	110

Decision Tree

Depth-5

	precision	recall	f1-score	support
False	0.67	0.94	0.78	70
True	0.64	0.17	0.27	40
accuracy			0.66	110
macro avg	0.65	0.56	0.53	110
weighted avg	0.66	0.66	0.60	110

Random Forest

	precision	recall	f1-score	support
False	0.69	0.89	0.78	70
True	0.60	0.30	0.40	40
accuracy			0.67	110
macro avg	0.64	0.59	0.59	110
weighted avg	0.66	0.67	0.64	110

Depth-10

	precision	recall	f1-score	support
False	0.67	0.94	0.79	70
True	0.67	0.20	0.31	40
accuracy			0.67	110
macro avg	0.67	0.57	0.55	110
weighted avg	0.67	0.67	0.61	110

Hasil Training Model

	precision	recall	f1-score	accuracy
Regresi Logistik	0.63	0.59	0.58	0.66
Kneighbours	0.54	0.54	0.54	0.57
Naive bayes	0.64	0.57	0.55	0.66
Decision Tree D-5	0.65	0.56	0.53	0.66
Decision Tree D-10	0.67	0.57	0.55	0.67
Random Forest	0.64	0.59	0.59	0.67
SVM	0.83	0.56	0.51	0.68

Thank you!