

**Judul :Pengelompokan Wilayah Berdasarkan Luas dan  
Kepadatan Penduduk Menggunakan Clustering**

**Anggota Kelompok : 1. Baretta Panuju Luhur (123220204)**

**2. Dainovien Marchmaurel Dirangga Al Sherard  
(123220215)**

**3. Dhiya Aulya Achsa Fitriani (123220055)**

### **1. Business Understanding & Analytic Approach**

Pengelolaan sumber daya dan perencanaan pembangunan yang efektif memerlukan pemahaman tentang distribusi penduduk berdasarkan luas wilayah dan kepadatan penduduk. Data yang akurat membantu pemerintah dalam merumuskan kebijakan yang tepat. Dengan menggunakan metode clustering, kita dapat mengelompokkan wilayah untuk memahami karakteristik demografis dan kebutuhan layanan publik secara lebih baik.

- Manfaat Data Science dalam Masalah Ini:
  - a) Optimalisasi Sumber Daya: Mengidentifikasi wilayah yang memerlukan perhatian lebih dalam pelayanan publik.
  - b) Perencanaan Berbasis Data: Membantu pemerintah dalam pengambilan keputusan strategis.
  - c) Mitigasi Risiko: Mengurangi ketidakmerataan dalam distribusi layanan.
- Usulan Task:

Task 1: Pengelompokan Wilayah:

Tujuan: Mengelompokkan wilayah berdasarkan luas dan kepadatan penduduk. Metode: K-Means untuk mengidentifikasi kelompok dengan karakteristik serupa.

Task 2: Pemetaan Wilayah:

Tujuan: Membuat peta interaktif yang menunjukkan kategori wilayah berdasarkan hasil clustering. Metode: Visualisasi menggunakan GIS untuk menampilkan hasil analisis.

## 2. Data Requirement & Data Collection

Data yang digunakan adalah data kepadatan penduduk menurut Kabupaten/Kota di D.I. Yogyakarta pada tahun 2007-2015. Berikut adalah sumber data yang digunakan:

Badan Pusat Statistik D.I. Yogyakarta, Kepadatan Penduduk menurut Kabupaten/Kota di D.I. Yogyakarta Tahun 2007-2009.

<https://yogyakarta.bps.go.id/id/statistics-table/2/MTMylzl=/kepadatan-penduduk-menurut-kabupaten-kota-di-d-i-yogyakarta.html>

Badan Pusat Statistik D.I. Yogyakarta, Kepadatan Penduduk menurut Kabupaten/Kota di D.I. Yogyakarta Tahun 2010-2012.

<https://yogyakarta.bps.go.id/id/statistics-table/2/MTMylzl=/kepadatan-penduduk-menurut-kabupaten-kota-di-d-i-yogyakarta.html>

Badan Pusat Statistik D.I. Yogyakarta, Kepadatan Penduduk menurut Kabupaten/Kota di D.I. Yogyakarta Tahun 2013-2015.

<https://yogyakarta.bps.go.id/id/statistics-table/2/MTMylzl=/kepadatan-penduduk-menurut-kabupaten-kota-di-d-i-yogyakarta.html>

Kabupaten/Kota	Kepadatan Penduduk menurut Kabupaten/Kota di D.I. Yogyakarta					
	luas			kepadatan penduduk		
	2007	2008	2009	2007	2008	2009
D.I. Yogyakarta	3.186	3.186	3.186	1.054	1.065	1.076
Kulonprogo	586	586	586	656	658	661
Bantul	507	507	507	1.722	1.748	1.774
Gunungkidul	1.485	1.485	1.485	455	455	455
Sleman	575	575	575	1.801	1.835	1.870
Kota Yogyakarta	32	32	32	12.056	12.024	11.990

**Gambar 2. 1** Kepadatan Penduduk Menurut Kabupaten/Kota di D.I. Yogyakarta Tahun 2007-2009

Kabupaten/Kota	Kepadatan Penduduk menurut Kabupaten/Kota di D.I. Yogyakarta					
	luas			kepadatan penduduk		
	2010	2011	2012	2010	2011	2012
D.I. Yogyakarta	3.186	3.186	3.186	1.085	1.095	1.103
Kulonprogo	586	586	586	663	666	670
Bantul	507	507	507	1.798	1.818	1.831
Gunungkidul	1.485	1.485	1.485	455	456	461
Sleman	575	575	575	1.902	1.926	1.939
Kota Yogyakarta	32	32	32	11.958	12.017	12.123

**Gambar 2. 2** Kepadatan Penduduk Menurut Kabupaten/Kota di D.I. Yogyakarta Tahun 2010-2012

Kabupaten/Kota	Kepadatan Penduduk menurut Kabupaten/Kota di D.I. Yogyakarta					
	luas			kepadatan penduduk		
	2013	2014	2015	2013	2014	2015
D.I. Yogyakarta	3.186	3.186	3.186	1.128	1.142	1.155
Kulonprogo	586	586	586	685	691	703
Bantul	507	507	507	1.884	1.911	1.917
Gunungkidul	1.485	1.485	1.485	467	470	482
Sleman	575	575	575	1.995	2.025	2.031
Kota Yogyakarta	32	32	32	12.241	12.322	12.699

**Gambar 2. 3** Kepadatan Penduduk Menurut Kabupaten/Kota di D.I. Yogyakarta Tahun 2013-2015

Penggunaan data tergantung pada tujuan analisis. Jika Anda ingin menganalisis seluruh DIY, maka semua data kabupaten/kota relevan. Namun, jika fokus hanya pada satu kabupaten/kota atau tren spesifik, maka data dapat disaring.

Alasan pemilihan data

- Seluruh data digunakan jika tujuan analisis mencakup perbandingan antar kabupaten/kota atau studi regional.
- Sebagian data digunakan jika fokus pada pola lokal atau variabel tertentu, seperti hanya mempelajari kabupaten dengan kepadatan tinggi.

### 3. Data Preparation

sebelum :

	Kota	tahun	luas	kepadatan_penduduk
1	D.I. Yogyakarta	2007	3186	1054
2	Kulonprogo	2007	586	656
3	Bantul	2007	507	1722
4	Gunungkidul	2007	1485	455
5	Sleman	2007	575	1801
6	Kota Yogyakarta	2007	32	12056
7	D.I. Yogyakarta	2008	3186	1065
8	Kulonprogo	2008	586	658
9	Bantul	2008	507	1748
10	Gunungkidul	2008	1485	455
11	Sleman	2008	575	1835
12	Kota Yogyakarta	2008	32	12024
13	D.I. Yogyakarta	2009	3186	1076
14	Kulonprogo	2009	586	661
15	Bantul	2009	507	1774
16	Gunungkidul	2009	1485	455
17	Sleman	2009	575	1870
18	Kota Yogyakarta	2009	32	11990
19	D.I. Yogyakarta	2010	3186	1085
20	Kulonprogo	2010	586	663
21	Bantul	2010	507	1798
22	Gunungkidul	2010	1485	455
23	Sleman	2010	575	1902
24	Kota Yogyakarta	2010	32	11958
25	D.I. Yogyakarta	2011	3186	1095
26	Kulonprogo	2011	586	666
27	Bantul	2011	507	1818
28	Gunungkidul	2011	1485	456
29	Sleman	2011	575	1926
30	Kota Yogyakarta	2011	32	12017
31	D.I. Yogyakarta	2012	3186	1103
32	Kulonprogo	2012	586	670

sesudah :

	luas	kepadatan_penduduk
1	2.0182847	-0.4733382
2	-0.4521148	-0.5689912
3	-0.5271769	-0.3127950
4	0.4020733	-0.6172984
5	-0.4625665	-0.2938086
6	-0.9784999	2.1708186
7	2.0182847	-0.4706945
8	-0.4521148	-0.5685105
9	-0.5271769	-0.3065463
10	0.4020733	-0.6172984
11	-0.4625665	-0.2856372
12	-0.9784999	2.1631279
13	2.0182847	-0.4680508
14	-0.4521148	-0.5677895
15	-0.5271769	-0.3002976
16	0.4020733	-0.6172984
17	-0.4625665	-0.2772255
18	-0.9784999	2.1549566
19	2.0182847	-0.4658878
20	-0.4521148	-0.5673089
21	-0.5271769	-0.2945296
22	0.4020733	-0.6172984
23	-0.4625665	-0.2695348
24	-0.9784999	2.1472659
25	2.0182847	-0.4634845
26	-0.4521148	-0.5665879
27	-0.5271769	-0.2897229
28	0.4020733	-0.6170580
29	-0.4625665	-0.2637668
30	-0.9784999	2.1614456
31	2.0182847	-0.4615618
32	-0.4521148	-0.5656265

## 4. Modelling & Evaluation

### Jenis Model yang Dipilih dan Alasannya

Model yang digunakan adalah **K-Means Clustering**, yang merupakan metode unsupervised learning.

#### Alasan Pemilihan Model:

- **K-Means** cocok untuk data numerik dan berfokus pada segmentasi berdasarkan kemiripan pola.
- Dataset memiliki variabel numerik seperti luas wilayah dan kepadatan penduduk, yang relevan untuk dikelompokkan menjadi beberapa cluster.
- Hasil clustering dapat digunakan untuk mengidentifikasi pola atau tren kepadatan penduduk dalam wilayah yang dianalisis.
- K-Means sederhana dan efisien untuk dataset berukuran sedang hingga besar.

### Pemilihan Algoritma dan Alasan

#### Algoritma yang Dipilih: K-Means

- **Studi Literatur:** K-Means banyak digunakan dalam penelitian terkait clustering geografis atau demografis. Sebagai contoh:
  - Penelitian mengenai analisis pola distribusi penduduk menggunakan clustering ([Sumber: IEEE](#)).
  - Studi segmentasi wilayah berdasarkan data kepadatan penduduk di jurnal *Springer*.
- **Kelebihan K-Means:**
  - Mudah diimplementasikan.
  - Waktu komputasi cepat, bahkan untuk dataset yang cukup besar.
  - Dapat menangani data numerik yang sudah dinormalisasi dengan baik.
- **Kekurangan:**
  - Sensitif terhadap inisialisasi awal dan outlier.
  - Memerlukan pemilihan jumlah cluster (k) yang optimal.



	Kota	tahun	luas	kepadatan_penduduk	cluster
1	D.I. Yogyakarta	2007	3186	1054	1
2	Kulonprogo	2007	586	656	3
3	Bantul	2007	507	1722	3
4	Gunungkidul	2007	1485	455	3
5	Sleman	2007	575	1801	3
6	Kota Yogyakarta	2007	32	12056	2
7	D.I. Yogyakarta	2008	3186	1065	1
8	Kulonprogo	2008	586	658	3
9	Bantul	2008	507	1748	3
10	Gunungkidul	2008	1485	455	3
11	Sleman	2008	575	1835	3
12	Kota Yogyakarta	2008	32	12024	2
13	D.I. Yogyakarta	2009	3186	1076	1
14	Kulonprogo	2009	586	661	3
15	Bantul	2009	507	1774	3

## Rencana Evaluasi Model

### Metode Evaluasi yang Digunakan

- **Elbow Method:**
  - Menentukan jumlah cluster optimal berdasarkan nilai *Within-Cluster Sum of Squares (WSS)*.
  - WSS mengevaluasi seberapa baik data dalam cluster yang sama berkumpul (semakin kecil, semakin baik).
- **Interpretasi Visual Clustering:**
  - Memvisualisasikan distribusi data dalam dimensi 2D (contohnya, berdasarkan luas wilayah dan kepadatan penduduk) untuk memahami pola cluster.
- **Statistik Deskriptif Cluster:**
  - Melakukan analisis statistik untuk melihat rata-rata kepadatan penduduk dan luas wilayah di setiap cluster.

### Referensi Studi Literatur:

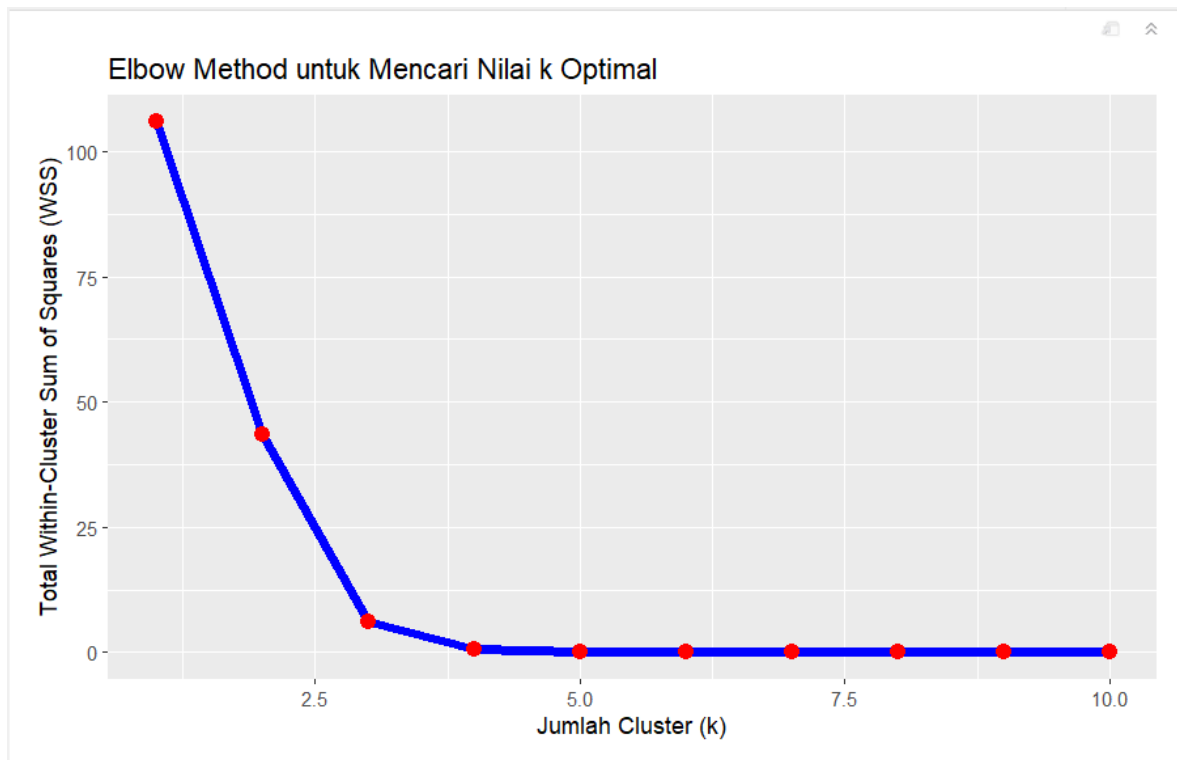
1. "Evaluating Clustering Results Using the Elbow Method" (Jurnal IEEE).
2. "Visualizing K-Means Clustering Using Dimensional Reduction" (Springer).

### Proses Clustering:

- Dataset dinormalisasi menggunakan `scale()` untuk memastikan variabel numerik setara.
- K-Means diterapkan dengan parameter:
  - `centers = 3` (berdasarkan hasil Elbow Method).
  - `nstart = 54` untuk mencegah hasil clustering bergantung pada inisialisasi awal.

**Visualisasi Elbow Method:** Grafik Elbow menunjukkan nilai WSS menurun signifikan hingga cluster ke-3, setelah itu penurunan melambat. Oleh karena itu, jumlah cluster optimal adalah **3**.

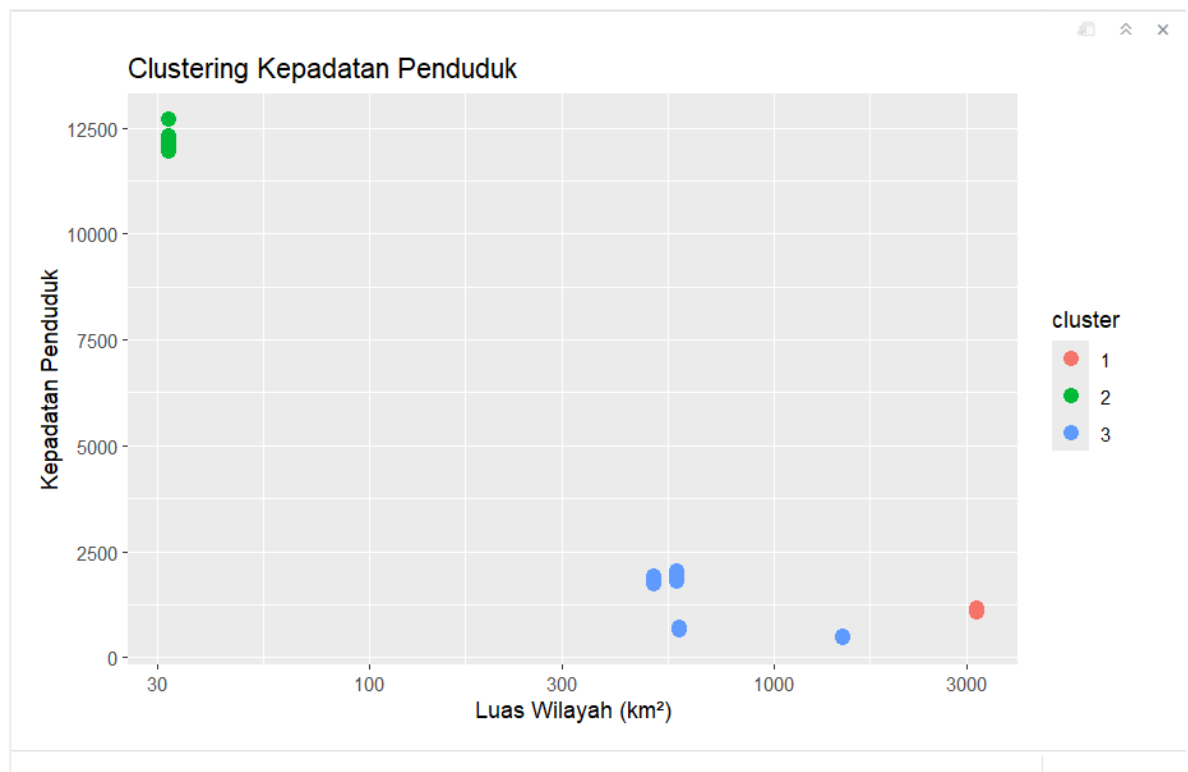
### Hasil Visualisasi Elbow Method:



### Visualisasi Cluster:

- Hasil clustering divisualisasikan dengan plot scatter untuk memetakan data berdasarkan luas wilayah dan kepadatan penduduk.

### Hasil Visualisasi Clustering:



**Statistik Deskriptif:** Setiap cluster dianalisis untuk mendapatkan rata-rata luas wilayah dan kepadatan penduduk:

cluster <fctr>	avg_kepadatan <dbl>	avg_luas <dbl>
1	1100.333	3186.00
2	12158.889	32.00
3	1220.444	788.25

3 rows

### Hasil Evaluasi Sesuai Rencana

1. **Jumlah Cluster Optimal:** Berdasarkan Elbow Method, jumlah cluster optimal adalah 3.
2. **Visualisasi Clustering:** Pola clustering menunjukkan kelompok yang terpisah secara visual, mengindikasikan bahwa K-Means bekerja baik untuk dataset ini.
3. **Statistik Deskriptif:** Cluster dengan rata-rata kepadatan penduduk tertinggi (Cluster 3) memiliki luas wilayah terkecil, menunjukkan pola konsentrasi populasi.
4. **Kesimpulan Evaluasi:**
  - Model K-Means memberikan hasil yang cukup baik dalam memisahkan data ke dalam cluster berdasarkan pola.

- Visualisasi dan statistik deskriptif mendukung interpretasi hasil yang dapat digunakan untuk perencanaan wilayah.

## 5. Kesimpulan

### Hasil Clustering

- **Cluster 1** mencakup wilayah dengan luas besar namun kepadatan penduduk relatif rendah. Wilayah ini cenderung memiliki potensi untuk pengembangan infrastruktur dan program pemerintah yang bertujuan meningkatkan populasi atau aktivitas ekonomi.
- **Cluster 2** terdiri dari daerah dengan luas sedang dan kepadatan penduduk moderat. Wilayah ini cocok untuk strategi pembangunan berimbang yang mengutamakan pertumbuhan populasi dan ekonomi secara paralel.
- **Cluster 3** berisi wilayah dengan luas kecil tetapi memiliki kepadatan penduduk tinggi, seperti Kota Yogyakarta. Wilayah ini membutuhkan fokus pada optimalisasi layanan publik, pengelolaan infrastruktur, serta mitigasi tekanan sosial akibat kepadatan.

Hasil analisis clustering ini sangat berguna untuk mendukung pengambilan keputusan strategis, baik oleh pemerintah daerah maupun pelaku bisnis:

- **Bagi Pemerintah:**
  - Menerapkan kebijakan pemerataan penduduk untuk mengurangi tekanan di daerah padat penduduk (Cluster 3).
  - Alokasi sumber daya yang tepat, misalnya pembangunan fasilitas pendidikan, layanan kesehatan, atau transportasi umum, sesuai dengan kebutuhan tiap cluster.
- **Bagi Bisnis:**
  - Daerah dengan kepadatan tinggi (Cluster 3) adalah peluang utama untuk bisnis ritel, layanan makanan cepat saji, dan sektor logistik.
  - Daerah dengan kepadatan rendah (Cluster 1) dapat menjadi fokus ekspansi jangka panjang melalui investasi di infrastruktur digital atau layanan berbasis teknologi.

### Strategi Bisnis dan Rekomendasi

- **Optimalisasi Wilayah Padat Penduduk:** Fokuskan pengembangan bisnis di Kota Yogyakarta (Cluster 3) dengan strategi pemasaran langsung ke konsumen dan ekspansi layanan berbasis kebutuhan.
- **Ekspansi Infrastruktur Digital:** Untuk wilayah berpenduduk rendah (Cluster 1), investasi pada infrastruktur digital seperti layanan e-commerce dapat menjadi

solusi untuk meningkatkan penetrasi pasar tanpa memerlukan infrastruktur fisik besar.

- **Meningkatkan Daya Saing Wilayah Sedang:** Wilayah di Cluster 2 dapat dijadikan prioritas untuk pengembangan fasilitas pendukung seperti pendidikan atau kawasan industri yang dapat menarik penduduk dan investasi.

#### **Kekurangan dan Kelebihan Model**

- **Kelebihan:**
  - Clustering memberikan gambaran yang jelas mengenai pola kepadatan penduduk per wilayah.
  - Visualisasi hasil clustering mempermudah interpretasi, baik untuk keperluan analisis maupun pengambilan keputusan.
- **Kekurangan:**
  - Dataset yang hanya mencakup dua variabel utama (luas dan kepadatan penduduk) dapat membatasi kompleksitas analisis.
  - Sensitivitas terhadap outlier seperti perubahan drastis pada data tahunan.

#### **Kesimpulan**

Dengan menggunakan data kepadatan penduduk D.I. Yogyakarta tahun 2007-2015, analisis clustering ini memberikan rekomendasi strategis yang sesuai dengan kebutuhan pemerintah dan bisnis. Dengan pemanfaatan hasil analisis, dapat tercipta perencanaan pembangunan yang lebih terarah dan berdampak nyata dalam mendukung pertumbuhan ekonomi dan kesejahteraan masyarakat di wilayah D.I. Yogyakarta.