

Import

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler

df = pd.read_csv("/content/kelayakan-pendidikan-indonesia.csv")

df.head()
```

↕

	Provinsi	Sekolah	Siswa	Mengulang	Putus Sekolah	Tenaga Kependidikan(<S1)	Tenaga Kependidikan(>S1)	Tenaga Kependidikan(<SM)	Tenaga Kependidikan(>SM)	Rombel
0	Prov. D.K.I. Jakarta	2224	756455	1202	869	1089	38305	6967	1569	1
1	Prov. Jawa Barat	19553	4487262	4778	5675	5007	209252	27487	3353	1
2	Prov. Jawa Tengah	18618	2595050	6909	2399	2337	167535	21325	2473	1
3	Prov. D.I. Yogyakarta	1846	270610	357	75	326	19771	4536	866	1
4	Prov. Jawa Timur	18980	2530849	4694	3327	2950	180465	21116	3818	1

◀ ▶

Visualiasi Data

+ Code

+ Text

1. Bar Chart

```
[ ] 1 cell hidden
```

2. Horizontal Bar Chart

```
[ ] 1 cell hidden
```

3. Shorted Bar Chart

```
[ ] 1 cell hidden
```

4. Scatter Plot

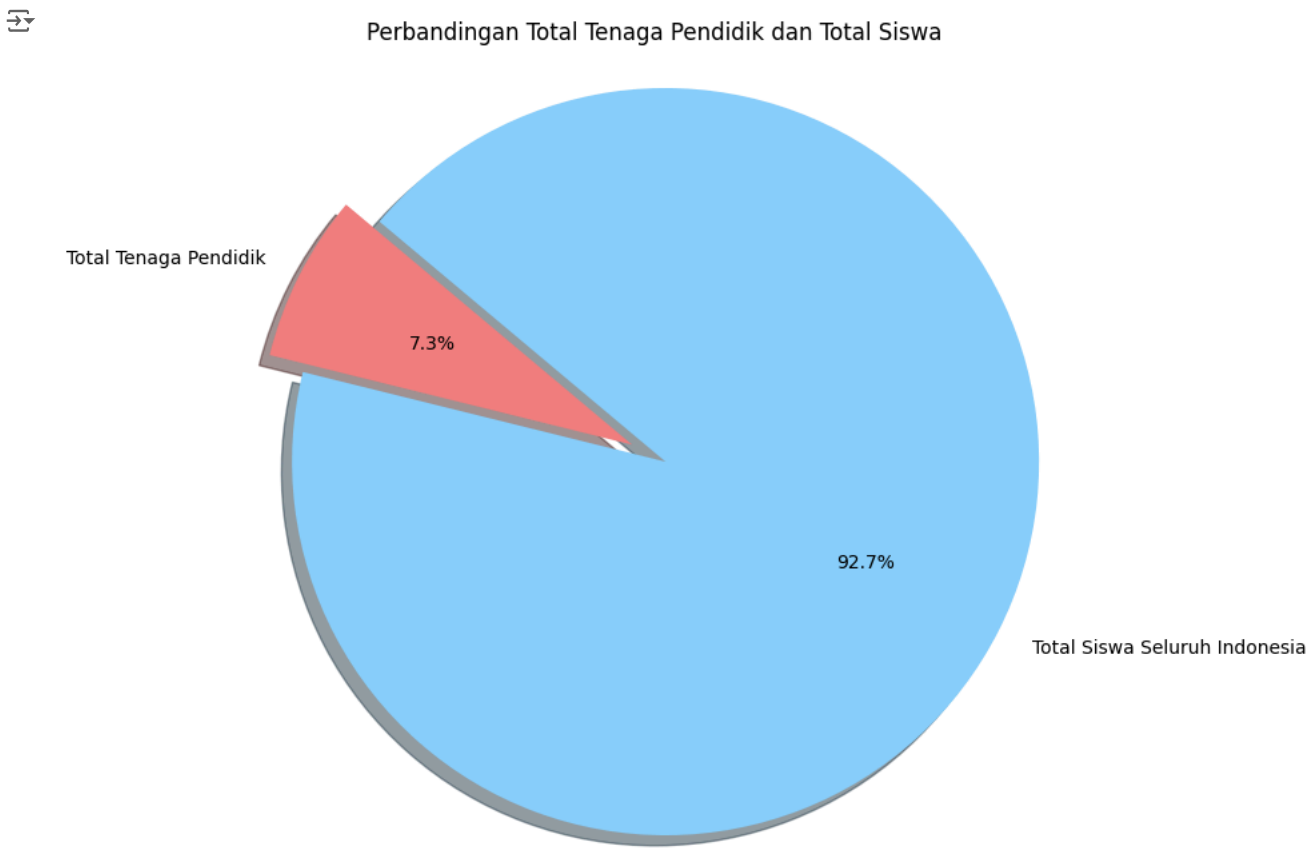
```
[ ] 2 cells hidden
```

5. Pie Chart

```
labels = ['Total Tenaga Pendidik', 'Total Siswa Seluruh Indonesia']
sizes = [total_tenaga_pendidik, total_siswa_seluruh_indonesia]
colors = ['lightcoral', 'lightskyblue']
explode = (0.1, 0) # explode 1st slice

plt.figure(figsize=(8, 8))
plt.pie(sizes, explode=explode, labels=labels, colors=colors,
        autopct='%1.1f%%', shadow=True, startangle=140)
plt.axis('equal') # Equal aspect ratio ensures that pie is drawn as a circle.
```

```
plt.title('Perbandingan Total Tenaga Pendidik dan Total Siswa')
plt.show()
```



Statistika Deskriptif

```
pd.set_option('display.float_format', lambda x: '%.2f' % x)
df.describe()
```

	Sekolah	Siswa	Mengulang	Putus Sekolah	Tenaga Kependidikan(<S1)	Tenaga Kependidikan(>S1)	Tenaga Kependidikan(<SM)	Tenaga Kependidikan(>SM)	Rombonga Belajar
count	39.00	39.00	39.00	39.00	39.00	39.00	39.00	39.00	39.0
mean	3817.49	617043.87	2840.82	1155.05	1311.92	40251.85	5773.79	907.38	29318.2
std	4862.82	865035.15	2312.42	1167.78	1016.53	48430.71	6128.03	933.15	37231.3
min	124.00	17789.00	56.00	75.00	76.00	450.00	15.00	5.00	913.0
25%	956.50	143176.00	1257.50	466.00	593.00	10399.50	1453.00	174.50	7653.5
50%	2347.00	312352.00	2049.00	835.00	1022.00	25737.00	4336.00	709.00	17694.0
75%	4326.50	709881.00	4212.00	1279.50	1997.00	48900.00	7093.00	1265.00	33300.5
max	19553.00	4487262.00	9801.00	5675.00	5007.00	209252.00	27487.00	3818.00	169704.0

Clustering

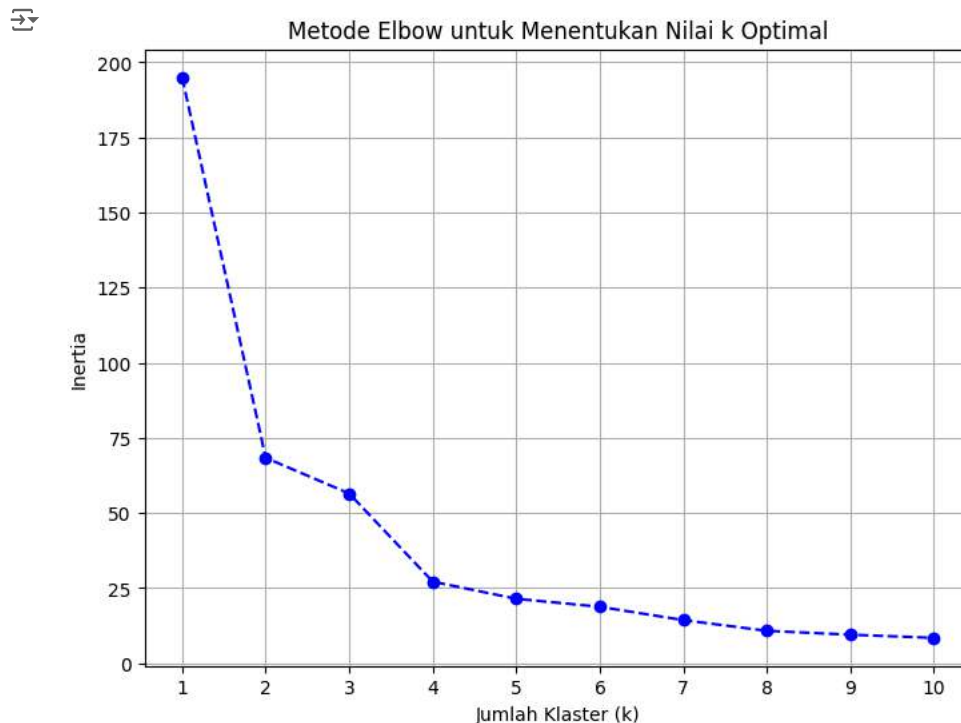
```
# Pilih fitur yang akan digunakan untuk clustering (contoh: 'Siswa' dan 'Tenaga Kependidikan (SM)')
fitur_clustering = df[['Siswa', 'Tenaga Kependidikan(<S1)', 'Tenaga Kependidikan(>S1)',
                        'Tenaga Kependidikan(<SM)', 'Tenaga Kependidikan(>SM)']]

# Standarisasi data
scaler = StandardScaler()
fitur_terstandarisasi = scaler.fit_transform(fitur_clustering)
```

```
# Menentukan nilai k menggunakan Metode Elbow
inertia = []
rentang_k = range(1, 11)

for k in rentang_k:
    kmeans = KMeans(n_clusters=k, random_state=42)
    kmeans.fit(fitur_terstandarisasi)
    inertia.append(kmeans.inertia_)

# Visualisasi Grafik Elbow
plt.figure(figsize=(8, 6))
plt.plot(rentang_k, inertia, marker='o', linestyle='--', color='b')
plt.title('Metode Elbow untuk Menentukan Nilai k Optimal')
plt.xlabel('Jumlah Klaster (k)')
plt.ylabel('Inertia')
plt.xticks(rentang_k)
plt.grid()
plt.show()
```



## Hasil EDA (Exploratory Data Analysis) yang Manusiawi

### 1 Data Awal

Index	Provinsi	Sekolah	Siswa	Mengulang	Putus Sekolah	Tenaga Kependidikan (<S1)	Tenaga Kependidikan (>S1)	Tenaga Kependidikan (<SM)	Tenaga Kependidikan (>SM)
0	Prov. D.K.I. Jakarta	2,224	756,455	1,202	869	1,089	38,305	6,967	1,569
1	Prov. Jawa Barat	19,553	4,487,262	4,778	5,675	5,007	209,252	27,487	3,353
2	Prov. Jawa Tengah	18,618	2,595,050	6,909	2,399	2,337	167,535	21,325	2,473
3	Prov. D.I. Yogyakarta	1,846	270,610	357	75	326	19,771	4,536	866
4	Prov. Jawa Timur	18,980	2,530,849	4,694	3,327	2,950	180,465	21,116	3,818

Data memiliki 39 baris (provinsi) dan 14 kolom, termasuk informasi tentang sekolah, siswa, tenaga kependidikan, dan ruang kelas.

### 2 Informasi Data

- Total kolom: 14
- Total baris: 39
- Tipe data:
  - 1 kolom teks (object)

- 13 kolom angka (int64)

- **Data kosong:** Tidak ada data kosong di kolom mana pun.
- 

### 3 Statistik Deskriptif

#### Jumlah Sekolah

- **Rata-rata:** 3.817 sekolah per provinsi
- **Min:** 124 sekolah (provinsi dengan sekolah paling sedikit)
- **Max:** 19.553 sekolah (provinsi dengan sekolah paling banyak)

#### Jumlah Siswa

- **Rata-rata:** 617.043 siswa per provinsi
- **Min:** 17.789 siswa
- **Max:** 4.487.262 siswa

#### Jumlah Mengulang dan Putus Sekolah

- **Rata-rata siswa mengulang:** 2.841 siswa/provinsi
  - **Rata-rata siswa putus sekolah:** 1.155 siswa/provinsi
- 

### 4 Data Kosong dan Duplikat

- **Data kosong:** Tidak ada data kosong di kolom mana pun
  - **Data duplikat:** Tidak ada data duplikat di dataset ini
- 

### 5 Visualisasi Data

#### 1. Bar Chart: Jumlah Siswa per Provinsi

- Plot ini menunjukkan jumlah siswa di setiap provinsi menggunakan diagram batang vertikal.
- Provinsi disusun dalam sumbu x, sedangkan jumlah siswa pada sumbu y.
- Memberikan pandangan yang jelas tentang distribusi jumlah siswa antar provinsi.

#### 2. Horizontal Bar Chart: Jumlah Siswa per Provinsi

- Diagram batang horizontal memudahkan pembacaan data untuk daftar provinsi yang panjang.
- Mempermudah analisis visual terhadap provinsi dengan jumlah siswa tertinggi dan terendah.

#### 3. Sorted Bar Chart: Jumlah Siswa per Provinsi di Indonesia

- Data jumlah siswa diurutkan dari provinsi dengan siswa terbanyak hingga yang paling sedikit.
- Memberikan penekanan pada perbedaan signifikan antara provinsi.

#### 4. Scatter Plot: Hubungan Mengulang dan Putus Sekolah

- Scatter plot digunakan untuk melihat hubungan antara jumlah siswa yang mengulang dan putus sekolah.
- Grafik ini membantu mengidentifikasi pola atau anomali dalam data.

#### 5. Pie Chart: Perbandingan Total Tenaga Pendidik dan Total Siswa

- Diagram lingkaran menunjukkan proporsi total tenaga pendidik dibandingkan dengan total siswa di seluruh Indonesia.
  - Memberikan gambaran tentang keseimbangan antara sumber daya pendidik dan jumlah siswa.
-