

## eda-kelompok-2

November 1, 2024

#Exploratory Data Analysis Pengaruh Fasilitas dan Sumber Daya Pendidikan terhadap Indeks Pendidikan di Jawa Barat

Tanggal Pembuatan Awal : 4 Oktober 2024

Tanggal Perbaikan : 25 Oktober 2024

Dibuat Oleh Kelompok 2 :

- Abdurrahman Al Ghifari (23000456) - Ahmad Izzuddin Azzam (2300492) - Julian Dwi Satrio (2300484) - Meisya Amalia (2309357) - Nuansa Bening Aura Jelita (2301410)

##Pendahuluan

Indeks Pendidikan merupakan ukuran yang mencerminkan kualitas pendidikan di suatu wilayah, menggabungkan berbagai indikator seperti angka partisipasi sekolah, rata-rata lama sekolah, dan tingkat melek huruf. Indeks ini memberikan gambaran umum mengenai seberapa baik sistem pendidikan di suatu daerah memenuhi kebutuhan pembelajaran penduduknya. Analisis yang mendalam terhadap indeks ini dapat membantu mengidentifikasi daerah-daerah yang memerlukan perhatian khusus, sehingga intervensi yang tepat dapat diterapkan untuk meningkatkan kualitas pendidikan secara merata di seluruh provinsi.

Sebagai salah satu provinsi terpadat di Indonesia, Jawa Barat memiliki tantangan besar dalam mengembangkan kualitas pendidikan. Pendidikan yang berkualitas dipengaruhi oleh berbagai faktor, termasuk fasilitas pendidikan yang memadai serta ketersediaan sumber daya pendidikan yang berkualitas, seperti tenaga pengajar dan bahan ajar. Pertanyaan yang sering muncul adalah bagaimana kedua faktor ini memengaruhi kualitas pendidikan di setiap daerah dan apakah ketersediaan fasilitas serta sumber daya tersebut cukup untuk meningkatkan Indeks Pendidikan.

Analisis ini menggunakan data dari Badan Pusat Statistik (BPS) dan Data Jabar untuk mengevaluasi pengaruh fasilitas dan sumber daya pendidikan terhadap Indeks Pendidikan di berbagai wilayah di Jawa Barat. Dalam menghadapi kebutuhan pendidikan yang terus berkembang, pemahaman mendalam mengenai pengaruh kedua faktor ini sangat penting bagi pemangku kebijakan dan tenaga pendidik dalam merencanakan strategi peningkatan kualitas pendidikan di masa depan.

##Pertanyaan

1. Bagaimana hubungan antara indeks pendidikan provinsi dan indeks pendidikan kabupaten/kota di Jawa Barat? (Apakah ada korelasi positif antara kedua indeks ini?)
2. Apakah kabupaten/kota dengan indeks pendidikan yang lebih tinggi juga menunjukkan lebih banyak jumlah sekolah dan guru? (Bagaimana pengaruhnya terhadap kualitas pendidikan di daerah tersebut?)

3. Bagaimana jumlah sekolah (SMP dan SMA) dan jumlah guru berpengaruh terhadap rata-rata lama sekolah di kabupaten/kota di Jawa Barat? (Apakah kabupaten/kota dengan lebih banyak sekolah dan guru memiliki rata-rata lama sekolah yang lebih tinggi?)
4. Bagaimana hubungan antara angka melek huruf penduduk usia 15 tahun ke atas dan tingkat pendidikan di kabupaten/kota di Jawa Barat (Apakah kabupaten/kota dengan tingkat melek huruf yang lebih tinggi memiliki indeks pendidikan yang lebih baik?)

##Import Library dan Dataset

- **indeks\_pendidikan\_berdasarkan\_kabupatenkota.csv:** dataset Indeks Pendidikan ditingkat kabupaten/kota di Jawa Barat dari data.jabarprov.go.id (<https://opendata.jabarprov.go.id/id/dataset/indeks-pendidikan-berdasarkan-kabupatenkota-di-jawa-barat>), diakses pada 4 oktober 2024.
- **indeks\_pendidikan\_provinsi\_indonesia.csv:** dataset Indeks Pendidikan ditingkat Provinsi di Indonesia dari data.jabarprov.go.id (<https://opendata.jabarprov.go.id/id/dataset/indeks-pendidikan-berdasarkan-provinsi-di-indonesia>), diakses pada 4 oktober 2024.
- **jumlah\_smp.csv:** dataset Jumlah Sekolah SMP ditingkat kabupaten/kota di Jawa Barat dari data.jabarprov.go.id (<https://opendata.jabarprov.go.id/id/dataset/jumlah-sekolah-menengah-pertama-smp-berdasarkan-kategori-sekolah-di-jawa-barat>), diakses pada 4 oktober 2024.
- **jumlah\_sma.csv:** dataset Jumlah Sekolah SMA ditingkat kabupaten/kota di Jawa Barat dari data.jabarprov.go.id (<https://opendata.jabarprov.go.id/id/dataset/jumlah-sekolah-menengah-atas-sma-berdasarkan-kategori-sekolah-di-jawa-barat>), diakses pada 4 oktober 2024.
- **jumlah\_guru\_sma.csv:** dataset Jumlah Guru SMA ditingkat kabupaten/kota di Jawa Barat dari data.jabarprov.go.id (<https://opendata.jabarprov.go.id/id/dataset/jumlah-guru-sekolah-menengah-atas-sma-berdasarkan-kabupatenkota-di-jawa-barat>), diakses pada 4 oktober 2024.
- **jumlah\_guru\_smp.csv:** dataset Jumlah Guru SMP ditingkat kabupaten/kota di Jawa Barat dari data.jabarprov.go.id (<https://opendata.jabarprov.go.id/id/dataset/jumlah-guru-sekolah-menengah-pertama-smp-berdasarkan-kabupatenkota-di-jawa-barat>), diakses pada 4 oktober 2024.
- **rata\_rata\_lama\_sekolah\_berdasarkan\_kabupatenkota\_data.csv:** dataset rata rata lama sekolah siswa kabupaten/kota di Jawa Barat dari data.jabarprov.go.id (<https://opendata.jabarprov.go.id/id/dataset/rata-rata-lama-sekolah-berdasarkan-kabupatenkota-di-jawa-barat>), diakses pada 4 oktober 2024.
- **harapan\_lama\_sekolah.csv:** dataset Harapan Lama Sekolah ditingkat kabupaten/kota di Jawa Barat dari data.jabarprov.go.id (<https://opendata.jabarprov.go.id/id/dataset/harapan-lama-sekolah-berdasarkan-kabupatenkota-di-jawa-barat>), diakses pada 4 oktober 2024.
- **angka\_melek\_huruf\_penduduk\_usia\_15\_thn\_ke\_atas.csv:** dataset Angka Melek Huruf Penduduk Usia 15 Tahun ke Atas ditingkat Kabupaten/Kota di Jawa Barat dari data.jabarprov.go.id (<https://opendata.jabarprov.go.id/id/dataset/angka-melek-huruf-penduduk-usia-15-tahun-ke-atas-berdasarkan-kabupatenkota-di-jawa-barat>), diakses pada 4

oktober 2024.

- **tingkat\_pengangguran\_terbuka.csv:** dataset Tingkat Pengangguran Terbuka ditingkat Kabupaten/Kota di Jawa Barat dari data.jabarprov.go.id (<https://opendata.jabarprov.go.id/id/dataset/tingkat-pengangguran-terbuka-berdasarkan-kabupatenkota-di-jawa-barat>), diakses pada 4 oktober 2024.

```
[1]: %matplotlib inline
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')

[2]: #data ipk, Indeks Pendidikan Kabupaten/Kota
ipk = pd.read_csv("https://drive.google.com/uc?
↳export=download&id=1Vqv3_d7w9dmone2khz8MwZh8tUJfVMGA")
#data ipp, Indeks Pendidikan Provinsi
ipp = pd.read_csv("https://drive.google.com/uc?
↳export=download&id=1VethoY8ViZ-OzG4vsScfBOFCNQ7uHiLZ")
#jumlah smp
smp = pd.read_csv("https://drive.google.com/uc?
↳export=download&id=1jQF3oIe5sEIW50JTgUnmGHWVRVURR4Wuq")
#jumlah sma
sma = pd.read_csv("https://drive.google.com/uc?
↳export=download&id=1GBbhfRA0itcSh4VXNRLJ-VnBkmY-02d6")
#jumlah guru smp
gsmp = pd.read_csv("https://drive.google.com/uc?
↳export=download&id=1YnzkrRFlor3NW3ItupNPJS0Qh3jQSt6sG")
#jumlah guru sma
gsma = pd.read_csv("https://drive.google.com/uc?
↳export=download&id=1Bt0v097V4dSgE-qeSC7QpZm5QlDuHKlK")
#data rata_rata_lama_sekolah
rrl = pd.read_csv("https://drive.google.com/uc?
↳export=download&id=14RhvaxtwOKtdmHu0MfItAELZCyKRxuHB")
#data hls, Harapan Lama Sekolah Kabupaten/Kota
hls = pd.read_csv("https://drive.google.com/uc?
↳export=download&id=1abq6HiCH1ouLwbL61Zyiei7frUGP-e3W")
#data amh, Angka Melek Huruf Usia 15 Tahun ke Atas Kabupaten/Kota
amh = pd.read_csv("https://drive.google.com/uc?
↳id=1qrYaBSUd3Z7jHn5tnbSbgvv9B0ubNjN7&export=download")
#data jpt, Jumlah Pengangguran Terbuka
jpt = pd.read_csv("https://drive.google.com/uc?
↳export=download&id=1u8b5sD8mxDlTipH48jWOKfL5hfkUj8bZ")
```

##Data Exploration

### 0.0.1 Mengeksplorasi Dataset Indeks Pendidikan

Mengeksplorasi dataset Indeks Pendidikan Provinsi di Indonesia Cek apakah dataset sudah terkoneksi ke program dengan menampilkan 5 record dari data

```
[3]: ipp.head()
```

```
[3]:
```

	id	kode_provinsi	provinsi	indeks_pendidikan	satuan	tahun
0	1	11	ACEH	63.42	POIN	2010
1	2	12	SUMATERA UTARA	61.20	POIN	2010
2	3	13	SUMATERA BARAT	61.06	POIN	2010
3	4	14	RIAU	60.18	POIN	2010
4	5	15	JAMBI	55.96	POIN	2010

Penjelasan Kolom Tabel:

- Kolom Indeks pendidikan adalah satuan dalam mengukur seberapa bagus pendidikan di Indonesia, parameternya adalah Harapan Lama Sekolah (HLS) dan Rata-rata Lama Sekolah (RLS).
- Kolom Satuan adalah penunjuk bahwa angka yang berada di indeks pendidikan memiliki satuan yang sama yaitu Poin.

Cek info dari dataset yang dimiliki (column, types, jmlh record)

```
[4]: ipp.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 306 entries, 0 to 305
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                    306 non-null   int64
1   kode_provinsi         306 non-null   int64
2   provinsi              306 non-null   object
3   indeks_pendidikan     306 non-null   float64
4   satuan                306 non-null   object
5   tahun                306 non-null   int64
dtypes: float64(1), int64(3), object(2)
memory usage: 14.5+ KB
```

Cek apakah ada record data yg duplikat

```
[5]: ipp.duplicated().sum()
```

```
[5]: 0
```

Cek apakah ada record data yang null/belum diisi

```
[6]: ipp.isnull().sum()
```

```
[6]: id                0
     kode_provinsi     0
     provinsi          0
     indeks_pendidikan 0
     satuan            0
     tahun             0
     dtype: int64
```

Cek apakah tipe data dari tiap kolom/atribut

```
[7]: ipp.dtypes
```

```
[7]: id                int64
     kode_provinsi     int64
     provinsi          object
     indeks_pendidikan float64
     satuan            object
     tahun             int64
     dtype: object
```

Cek statistik dari data indeks\_pendidikan (jmlh, rata2, std deviasi, nilmin, nilmax, q1, q2, q3)

```
[8]: ipp.describe().indeks_pendidikan
```

```
[8]: count    306.000000
     mean      60.209150
     std       8.017469
     min       0.000000
     25%       57.132500
     50%       60.700000
     75%       64.267500
     max       74.290000
     Name: indeks_pendidikan, dtype: float64
```

Cek apakah ada outlier atau record data tidak normal yang terdeteksi

```
[106]: # Menghitung Q1 (25%) dan Q3 (75%)
       Q1 = ipp['indeks_pendidikan'].quantile(0.25)
       Q3 = ipp['indeks_pendidikan'].quantile(0.75)

       # Menghitung IQR (Interquartile Range)
       IQR = Q3 - Q1

       # Menentukan batas bawah dan atas untuk outlier
       batas_bawah = Q1 - 1.5 * IQR
       batas_atas = Q3 + 1.5 * IQR
```

```

# Mendeteksi outlier
outlier = ipp[(ipp['indeks_pendidikan'] < batas_bawah) |
              (ipp['indeks_pendidikan'] > batas_atas)]

# Menampilkan outlier
print(outlier)

# Jumlah outlier
print(f"Jumlah outlier: {outlier.shape[0]}")

# Menampilkan boxplot
plt.figure(figsize=(8, 6))
sns.boxplot(x=ipp['indeks_pendidikan'])
plt.title('Boxplot Indeks Pendidikan', fontsize=14)
plt.xlabel('Indeks Pendidikan', fontsize=12)
plt.show()

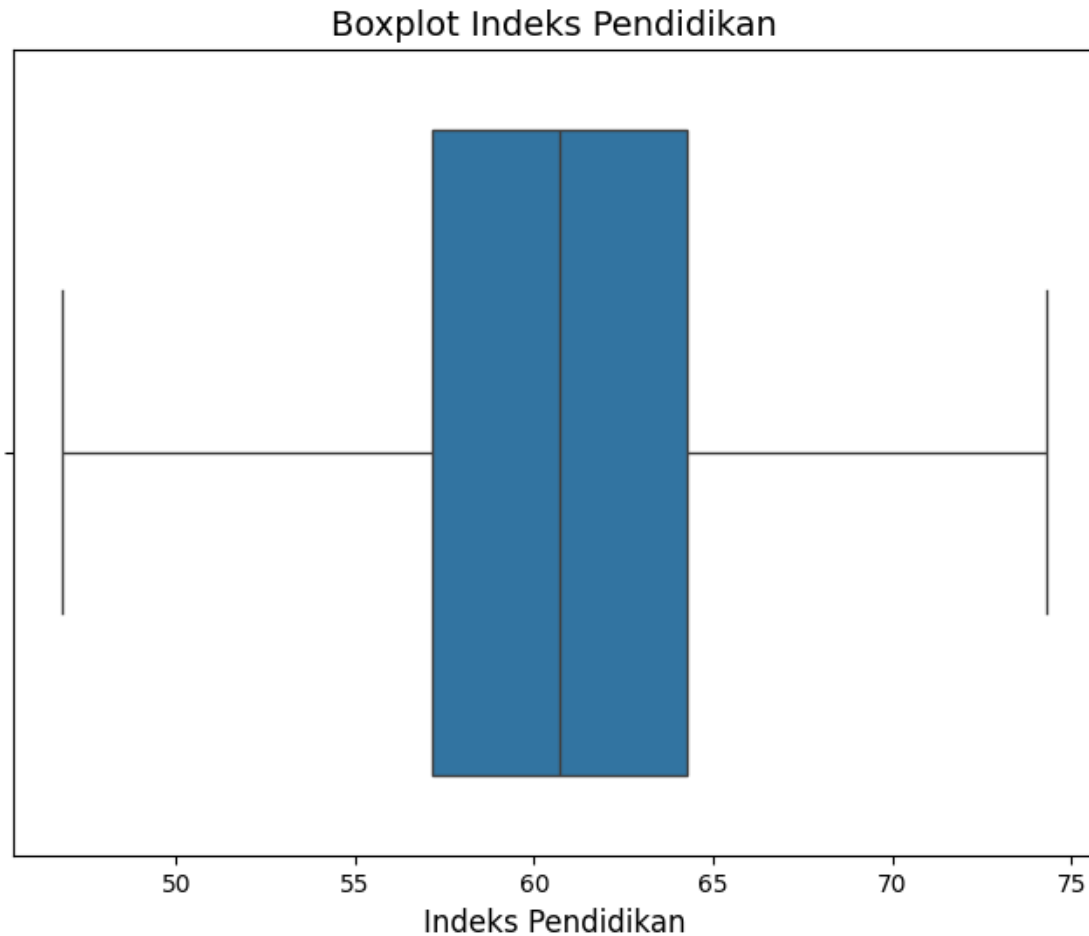
```

Empty DataFrame

Columns: [id, kode\_provinsi, provinsi, indeks\_pendidikan, satuan, tahun]

Index: []

Jumlah outlier: 0



Disini terdeteksi ada, maka melakukan proses menormalisasikan data dengan mengganti nilai outlier dengan nilai minimal karena outlier yang terdeteksi berada di bawah nilai minimal

```
[10]: # Mendapatkan nilai minimal dari data yang berada dalam rentang non-outlier
nilai_minimal = ipp[(ipp['indeks_pendidikan'] >= batas_bawah) &
↳ (ipp['indeks_pendidikan'] <= batas_atas)]['indeks_pendidikan'].min()

# Mengganti outlier dengan nilai minimal
ipp['indeks_pendidikan'] = np.where((ipp['indeks_pendidikan'] < batas_bawah) |
↳ (ipp['indeks_pendidikan'] > batas_atas),
                                nilai_minimal, ipp['indeks_pendidikan'])

# Menampilkan hasil setelah mengganti outlier
print(ipp[['indeks_pendidikan']])
```

```
      indeks_pendidikan
0                63.42
```

```

1          61.20
2          61.06
3          60.18
4          55.96
..         ...
301        59.97
302        70.60
303        66.90
304        59.04
305        51.82

```

[306 rows x 1 columns]

**Mengeksplorasi dataset Indeks Pendidikan Kabupaten/Kota di Jawa Barat Cek apakah dataset sudah terkoneksi ke program dengan menampilkan 5 record dari data**

```
[11]: ipk.head()
```

```

[11]:   id  kode_provinsi nama_provinsi  kode_kabupaten_kota nama_kabupaten_kota \
0    1             32    JAWA BARAT             3201    KABUPATEN BOGOR
1    2             32    JAWA BARAT             3202    KABUPATEN SUKABUMI
2    3             32    JAWA BARAT             3203    KABUPATEN CIANJUR
3    4             32    JAWA BARAT             3204    KABUPATEN BANDUNG
4    5             32    JAWA BARAT             3205    KABUPATEN GARUT

      indeks_pendidikan satuan  tahun
0             51.55    POIN    2010
1             47.39    POIN    2010
2             47.30    POIN    2010
3             56.65    POIN    2010
4             51.03    POIN    2010

```

Penjelasan Kolom Tabel:

- Kolom Indeks pendidikan adalah satuan dalam mengukur seberapa bagus pendidikan di Indonesia, parameternya adalah Harapan Lama Sekolah (HLS) dan Rata-rata Lama Sekolah (RLS).
- Kolom Satuan adalah penunjuk bahwa angka yang berada di indeks pendidikan memiliki satuan yang sama yaitu Poin.

**Cek info dari dataset yang dimiliki (column, types, jmlh record)**

```
[12]: ipk.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 348 entries, 0 to 347
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -

```



```

0    id                348 non-null    int64
1    kode_provinsi     348 non-null    int64
2    nama_provinsi     348 non-null    object
3    kode_kabupaten_kota 348 non-null    int64
4    nama_kabupaten_kota 348 non-null    object
5    indeks_pendidikan  348 non-null    float64
6    satuan            348 non-null    object
7    tahun              348 non-null    int64
dtypes: float64(1), int64(4), object(3)
memory usage: 21.9+ KB

```

Disini terdapat atribut/kolom yang tidak berguna, dimana kolom ini adalah kolom kode dan nama provinsi yang dimana informasi tersebut sudah tergambarkan dari nama datasetnya sendiri, yaitu Indeks Pendidikan di Jawa Barat

```

[13]: ipk = ipk.drop(columns=['kode_provinsi', 'nama_provinsi'])

      ipk.head()

```

```

[13]:   id  kode_kabupaten_kota  nama_kabupaten_kota  indeks_pendidikan  satuan  \
0    1                3201      KABUPATEN BOGOR                51.55    POIN
1    2                3202      KABUPATEN SUKABUMI                47.39    POIN
2    3                3203      KABUPATEN CIANJUR                47.30    POIN
3    4                3204      KABUPATEN BANDUNG                56.65    POIN
4    5                3205      KABUPATEN GARUT                51.03    POIN

      tahun
0    2010
1    2010
2    2010
3    2010
4    2010

```

Cek apakah ada record data yg duplikat

```

[14]: ipk.duplicated().sum()

```

```

[14]: 0

```

Cek apakah ada record data yang null/belum diisi

```

[15]: ipk.isnull().sum()

```

```

[15]: id                0
      kode_kabupaten_kota  0
      nama_kabupaten_kota  0
      indeks_pendidikan    0
      satuan              0
      tahun              0

```

```
dtype: int64
```

Cek apakah tipe data dari tiap kolom/atribut

```
[16]: ipk.dtypes
```

```
[16]: id                int64
      kode_kabupaten_kota    int64
      nama_kabupaten_kota    object
      indeks_pendidikan      float64
      satuan                 object
      tahun                  int64
      dtype: object
```

Cek statistik dari data indeks\_pendidikan (jmlh, rata2, std deviasi, nilmin, nilmax, q1, q2, q3)

```
[17]: ipk.describe().indeks_pendidikan
```

```
[17]: count      348.000000
      mean       61.308563
      std        7.332452
      min       45.480000
      25%       55.980000
      50%       59.560000
      75%       66.772500
      max       77.330000
      Name: indeks_pendidikan, dtype: float64
```

Cek apakah ada outlier atau record data tidak normal yang terdeteksi

```
[18]: # Menghitung Q1 (25%) dan Q3 (75%)
      Q1 = ipk['indeks_pendidikan'].quantile(0.25)
      Q3 = ipk['indeks_pendidikan'].quantile(0.75)

      # Menghitung IQR (Interquartile Range)
      IQR = Q3 - Q1

      # Menentukan batas bawah dan atas untuk outlier
      batas_bawah = Q1 - 1.5 * IQR
      batas_atas = Q3 + 1.5 * IQR

      # Mendeteksi outlier
      outlier = ipk[(ipk['indeks_pendidikan'] < batas_bawah) |
                    (ipk['indeks_pendidikan'] > batas_atas)]

      # Menampilkan outlier
      print(outlier)
```

```

# Jumlah outlier
print(f"Jumlah outlier: {outlier.shape[0]}")

# Menampilkan boxplot
plt.figure(figsize=(8, 6))
sns.boxplot(x=ipp['indeks_pendidikan'])
plt.title('Boxplot Indeks Pendidikan', fontsize=14)
plt.xlabel('Indeks Pendidikan', fontsize=12)
plt.show()

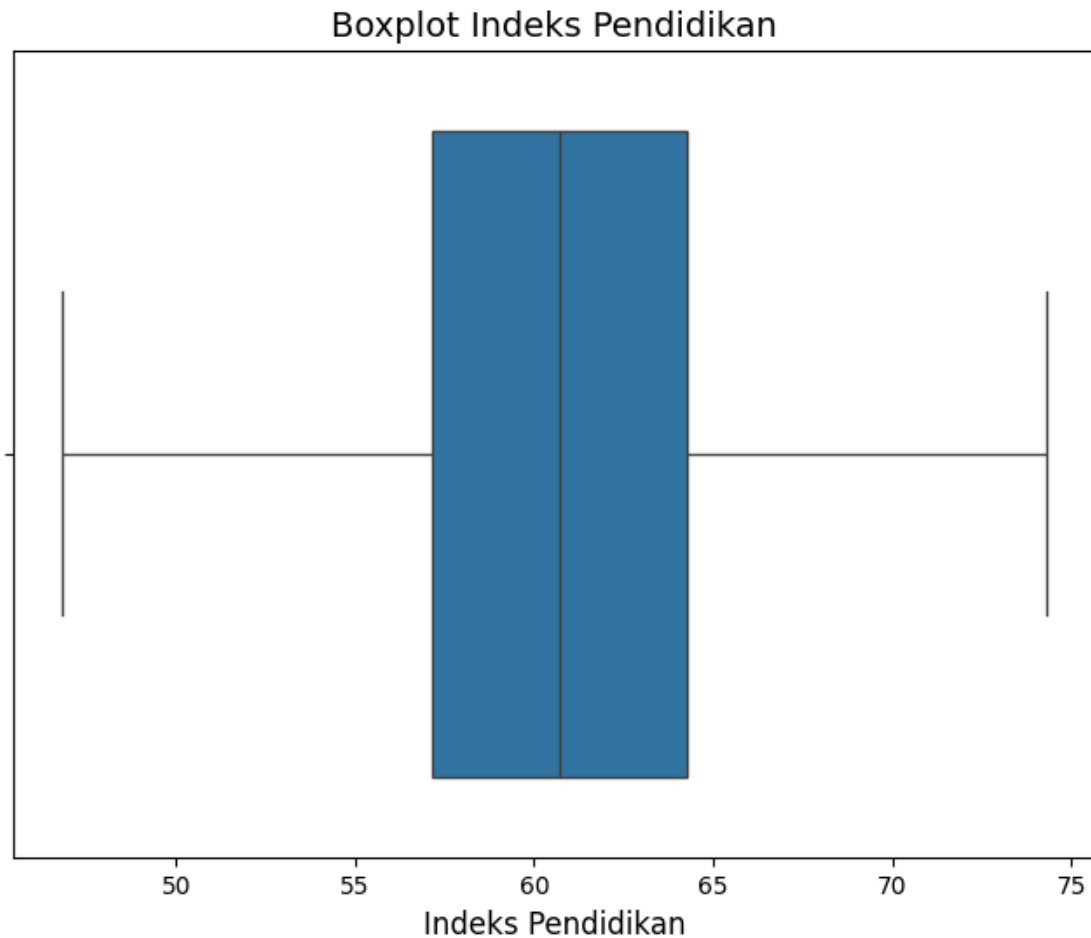
```

Empty DataFrame

Columns: [id, kode\_kabupaten\_kota, nama\_kabupaten\_kota, indeks\_pendidikan, satuan, tahun]

Index: []

Jumlah outlier: 0



## 0.0.2 Mengeksplorasi Data Jumlah SMP & SMA di Jawa Barat

### Cek Data yang sudah di Import

```
[19]: smp.head()
```

```
[19]:   id  kode_provinsi nama_provinsi  kode_kabupaten_kota nama_kabupaten_kota \
0    1             32    JAWA BARAT             3201    KABUPATEN BOGOR
1    2             32    JAWA BARAT             3201    KABUPATEN BOGOR
2    3             32    JAWA BARAT             3202    KABUPATEN SUKABUMI
3    4             32    JAWA BARAT             3202    KABUPATEN SUKABUMI
4    5             32    JAWA BARAT             3203    KABUPATEN CIANJUR
```

```
   kategori_sekolah  jumlah_sekolah satuan tahun_ajaran
0          NEGERI             103    UNIT    2016/2017
1          SWASTA             570    UNIT    2016/2017
2          NEGERI             162    UNIT    2016/2017
3          SWASTA             168    UNIT    2016/2017
4          NEGERI             151    UNIT    2016/2017
```

```
[20]: sma.head()
```

```
[20]:   id  kode_provinsi nama_provinsi  kode_kabupaten_kota nama_kabupaten_kota \
0    1             32    JAWA BARAT             3201    KABUPATEN BOGOR
1    2             32    JAWA BARAT             3201    KABUPATEN BOGOR
2    3             32    JAWA BARAT             3202    KABUPATEN SUKABUMI
3    4             32    JAWA BARAT             3202    KABUPATEN SUKABUMI
4    5             32    JAWA BARAT             3203    KABUPATEN CIANJUR
```

```
   kategori_sekolah  jumlah_sekolah satuan tahun_ajaran
0          NEGERI              44    UNIT    2016/2017
1          SWASTA             132    UNIT    2016/2017
2          NEGERI              26    UNIT    2016/2017
3          SWASTA              50    UNIT    2016/2017
4          NEGERI              18    UNIT    2016/2017
```

### Cek Informasi dari data SMP dan SMA

```
[21]: smp.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 378 entries, 0 to 377
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                    378 non-null   int64
1   kode_provinsi         378 non-null   int64
2   nama_provinsi         378 non-null   object
3   kode_kabupaten_kota   378 non-null   int64
```

```

4  nama_kabupaten_kota  378 non-null  object
5  kategori_sekolah     378 non-null  object
6  jumlah_sekolah       378 non-null  int64
7  satuan               378 non-null  object
8  tahun_ajaran         378 non-null  object
dtypes: int64(4), object(5)
memory usage: 26.7+ KB

```

```
[22]: sma.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 378 entries, 0 to 377
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                    378 non-null   int64
1   kode_provinsi         378 non-null   int64
2   nama_provinsi         378 non-null   object
3   kode_kabupaten_kota   378 non-null   int64
4   nama_kabupaten_kota   378 non-null   object
5   kategori_sekolah      378 non-null   object
6   jumlah_sekolah        378 non-null   int64
7   satuan                378 non-null   object
8   tahun_ajaran          378 non-null   object
dtypes: int64(4), object(5)
memory usage: 26.7+ KB

```

**Cek apakah ada atribut NULL dalam data SMP & SMA**

```
[23]: smp.isnull().sum()
```

```

[23]: id                    0
      kode_provinsi         0
      nama_provinsi         0
      kode_kabupaten_kota    0
      nama_kabupaten_kota    0
      kategori_sekolah       0
      jumlah_sekolah         0
      satuan                 0
      tahun_ajaran           0
      dtype: int64

```

```
[24]: sma.isnull().sum()
```

```

[24]: id                    0
      kode_provinsi         0
      nama_provinsi         0
      kode_kabupaten_kota    0

```

```
nama_kabupaten_kota    0
kategori_sekolah       0
jumlah_sekolah         0
satuan                 0
tahun_ajaran           0
dtype: int64
```

Cek tipe data yang digunakan dalam setiap atribut pada data SMP & SMA

```
[25]: smp.dtypes
```

```
[25]: id                int64
      kode_provinsi     int64
      nama_provinsi     object
      kode_kabupaten_kota int64
      nama_kabupaten_kota object
      kategori_sekolah   object
      jumlah_sekolah     int64
      satuan            object
      tahun_ajaran       object
      dtype: object
```

```
[26]: sma.dtypes
```

```
[26]: id                int64
      kode_provinsi     int64
      nama_provinsi     object
      kode_kabupaten_kota int64
      nama_kabupaten_kota object
      kategori_sekolah   object
      jumlah_sekolah     int64
      satuan            object
      tahun_ajaran       object
      dtype: object
```

Cek apakah ada data yang duplikat dalam data SMP & SMA

```
[27]: smp.duplicated().sum()
```

```
[27]: 0
```

```
[28]: sma.duplicated().sum()
```

```
[28]: 0
```

Menghapus kolom 'id\_provinsi' & 'nama\_provinsi' karena sudah jelas data nya hanya berasal dari 1 provinsi

```
[29]: smp = smp.drop(columns=['kode_provinsi', 'nama_provinsi'])

smp.head()
```

```
[29]:   id  kode_kabupaten_kota  nama_kabupaten_kota  kategori_sekolah  \
0    1                3201      KABUPATEN BOGOR          NEGERI
1    2                3201      KABUPATEN BOGOR          SWASTA
2    3                3202  KABUPATEN SUKABUMI          NEGERI
3    4                3202  KABUPATEN SUKABUMI          SWASTA
4    5                3203  KABUPATEN CIANJUR          NEGERI

      jumlah_sekolah  satuan  tahun_ajaran
0                103    UNIT    2016/2017
1                570    UNIT    2016/2017
2                162    UNIT    2016/2017
3                168    UNIT    2016/2017
4                151    UNIT    2016/2017
```

```
[30]: sma = sma.drop(columns=['kode_provinsi', 'nama_provinsi'])

sma.head()
```

```
[30]:   id  kode_kabupaten_kota  nama_kabupaten_kota  kategori_sekolah  \
0    1                3201      KABUPATEN BOGOR          NEGERI
1    2                3201      KABUPATEN BOGOR          SWASTA
2    3                3202  KABUPATEN SUKABUMI          NEGERI
3    4                3202  KABUPATEN SUKABUMI          SWASTA
4    5                3203  KABUPATEN CIANJUR          NEGERI

      jumlah_sekolah  satuan  tahun_ajaran
0                 44    UNIT    2016/2017
1                132    UNIT    2016/2017
2                 26    UNIT    2016/2017
3                 50    UNIT    2016/2017
4                 18    UNIT    2016/2017
```

### Cek statistik dari data yang SMP & SMA

```
[31]: smp.describe().jumlah_sekolah
```

```
[31]: count      378.000000
mean        101.830688
std         97.904370
min         10.000000
25%         33.000000
50%         79.000000
75%        137.000000
```

```
max      662.000000
Name: jumlah_sekolah, dtype: float64
```

```
[32]: sma.describe().jumlah_sekolah
```

```
[32]: count      378.000000
mean       30.645503
std        30.045787
min         2.000000
25%        11.250000
50%        19.000000
75%        39.750000
max       170.000000
Name: jumlah_sekolah, dtype: float64
```

### 0.0.3 Mengeksplorasi Data Jumlah Guru di SMP dan SMA di Jawa Barat

#### Jumlah Guru SMP di Jawa Barat 2015 - 2023

```
[33]: # Tampilkan info dan hapus kolom yang tidak diperlukan
      gsmp.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 216 entries, 0 to 215
Data columns (total 8 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   id                    216 non-null   int64
 1   kode_provinsi         216 non-null   int64
 2   nama_provinsi         216 non-null   object
 3   kode_kabupaten_kota   216 non-null   int64
 4   nama_kabupaten_kota   216 non-null   object
 5   jumlah_guru           216 non-null   int64
 6   satuan                216 non-null   object
 7   tahun_ajaran          216 non-null   object
dtypes: int64(4), object(4)
memory usage: 13.6+ KB
```

```
[34]: # Hapus atribut yang tidak diperlukan
      gsmp = gsmp.drop(columns=['kode_provinsi', 'nama_provinsi', 'satuan'])
      # Tampilkan 100 baris pertama
      gsmp.head(100)
```

```
[34]:    id  kode_kabupaten_kota  nama_kabupaten_kota  jumlah_guru  \
0     1          3201      KABUPATEN BOGOR          2950
1     2          3202      KABUPATEN SUKABUMI          2353
2     3          3203      KABUPATEN CIANJUR          3832
3     4          3204      KABUPATEN BANDUNG          2100
```



4	5	3205	KABUPATEN GARUT	2208
..	...	...	...	...
95	96	3215	KABUPATEN KARAWANG	1312
96	97	3216	KABUPATEN BEKASI	2352
97	98	3217	KABUPATEN BANDUNG BARAT	1210
98	99	3218	KABUPATEN PANGANDARAN	236
99	100	3271	KOTA BOGOR	3833

	tahun_ajaran
0	2015/2016
1	2015/2016
2	2015/2016
3	2015/2016
4	2015/2016
..	...
95	2018/2019
96	2018/2019
97	2018/2019
98	2018/2019
99	2018/2019

[100 rows x 5 columns]

### Jumlah Guru SMA di Jawa Barat 2015 - 2023

```
[35]: # Tampilkan info dan hapus kolom yang tidak diperlukan
gsma.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 216 entries, 0 to 215
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                     216 non-null   int64
1   kode_provinsi          216 non-null   int64
2   nama_provinsi          216 non-null   object
3   kode_kabupaten_kota    216 non-null   int64
4   nama_kabupaten_kota    216 non-null   object
5   jumlah_guru            216 non-null   int64
6   satuan                 216 non-null   object
7   tahun_ajaran           216 non-null   object
dtypes: int64(4), object(4)
memory usage: 13.6+ KB
```

```
[36]: # Hapus atribut yang tidak diperlukan
gsma = gsma.drop(columns=['kode_provinsi', 'nama_provinsi', 'satuan'])
# Tampilkan 100 baris pertama
gsma.head(100)
```

```
[36]:      id  kode_kabupaten_kota      nama_kabupaten_kota  jumlah_guru  \
0      1      3201      KABUPATEN BOGOR      8372
1      2      3202      KABUPATEN SUKABUMI      5169
2      3      3203      KABUPATEN CIANJUR      5398
3      4      3204      KABUPATEN BANDUNG      5462
4      5      3205      KABUPATEN GARUT      5517
..    ...
95     96      3215      KABUPATEN KARAWANG      3270
96     97      3216      KABUPATEN BEKASI      5107
97     98      3217      KABUPATEN BANDUNG BARAT      2884
98     99      3218      KABUPATEN PANGANDARAN      931
99    100      3271      KOTA BOGOR      5281
```

```
      tahun_ajaran
0      2015/2016
1      2015/2016
2      2015/2016
3      2015/2016
4      2015/2016
..    ...
95     2018/2019
96     2018/2019
97     2018/2019
98     2018/2019
99     2018/2019
```

```
[100 rows x 5 columns]
```

#### 0.0.4 Mengeksplorasi Data Harapan Lama Sekolah

Cek apakah dataset sudah ter-import ke program dengan menampilkan 5 record awal

```
[37]: hls.head()
```

```
[37]:      id  kode_provinsi  nama_provinsi  kode_kabupaten_kota  nama_kabupaten_kota  \
0      1      32      JAWA BARAT      3201      KABUPATEN BOGOR
1      2      32      JAWA BARAT      3202      KABUPATEN SUKABUMI
2      3      32      JAWA BARAT      3203      KABUPATEN CIANJUR
3      4      32      JAWA BARAT      3204      KABUPATEN BANDUNG
4      5      32      JAWA BARAT      3205      KABUPATEN GARUT
```

```
      harapan_lama_sekolah  satuan  tahun
0      10.28      TAHUN      2010
1      10.08      TAHUN      2010
2      9.62      TAHUN      2010
3      10.82      TAHUN      2010
4      10.36      TAHUN      2010
```

Cek info dari dataset yang dimiliki (column, types, jumlah record)

```
[38]: hls.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 378 entries, 0 to 377
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                    378 non-null    int64
1   kode_provinsi         378 non-null    int64
2   nama_provinsi         378 non-null    object
3   kode_kabupaten_kota   378 non-null    int64
4   nama_kabupaten_kota   378 non-null    object
5   harapan_lama_sekolah  378 non-null    object
6   satuan                378 non-null    object
7   tahun                 378 non-null    int64
dtypes: int64(4), object(4)
memory usage: 23.8+ KB
```

Cek apakah ada data yang terduplikat

```
[39]: hls.duplicated().sum()
```

```
[39]: 0
```

Cek apakah ada record data yang null/belum diisi

```
[40]: hls.isnull().sum()
```

```
[40]: id                    0
      kode_provinsi       0
      nama_provinsi       0
      kode_kabupaten_kota  0
      nama_kabupaten_kota  0
      harapan_lama_sekolah 0
      satuan              0
      tahun               0
      dtype: int64
```

Cek apakah tipe data dari tiap kolom/atribut

```
[41]: hls.dtypes
```

```
[41]: id                    int64
      kode_provinsi       int64
      nama_provinsi       object
      kode_kabupaten_kota int64
      nama_kabupaten_kota object
```

```
harapan_lama_sekolah    object
satuan                  object
tahun                   int64
dtype: object
```

Menghapus kolom 'id\_provinsi' & 'nama\_provinsi' karena sudah jelas data nya hanya berasal dari 1 provinsi

```
[42]: hls = hls.drop(columns=['kode_provinsi', 'nama_provinsi'])
      hls.head()
```

```
[42]:   id  kode_kabupaten_kota  nama_kabupaten_kota  harapan_lama_sekolah  satuan \
0    1          3201      KABUPATEN BOGOR          10.28    TAHUN
1    2          3202      KABUPATEN SUKABUMI          10.08    TAHUN
2    3          3203      KABUPATEN CIANJUR           9.62    TAHUN
3    4          3204      KABUPATEN BANDUNG          10.82    TAHUN
4    5          3205      KABUPATEN GARUT           10.36    TAHUN

      tahun
0    2010
1    2010
2    2010
3    2010
4    2010
```

Cek statistik dari data harapan\_lama\_sekolah

```
[43]: hls['harapan_lama_sekolah'].describe()
```

```
[43]: count      378
      unique     232
      top       12.24
      freq        7
      Name: harapan_lama_sekolah, dtype: object
```

Cek apakah ada outlier atau record data tidak normal yang terdeteksi

```
[107]: # Mengganti nilai yang tidak valid dengan NaN
      hls['harapan_lama_sekolah'] = hls['harapan_lama_sekolah'].replace('-', np.nan)

      # Mengonversi kolom 'harapan_lama_sekolah' menjadi tipe float
      hls['harapan_lama_sekolah'] = hls['harapan_lama_sekolah'].astype(float)

      # Menghapus baris yang mengandung NaN (opsional)
      hls = hls.dropna(subset=['harapan_lama_sekolah'])

      # Menghitung Q1 (25%) dan Q3 (75%)
```

```

Q1 = hls['harapan_lama_sekolah'].quantile(0.25)
Q3 = hls['harapan_lama_sekolah'].quantile(0.75)

# Menghitung IQR (Interquartile Range)
IQR = Q3 - Q1

# Menentukan batas bawah dan atas untuk outlier
batas_bawah = Q1 - 1.5 * IQR
batas_atas = Q3 + 1.5 * IQR

# Mendeteksi outlier
outlier = hls[(hls['harapan_lama_sekolah'] < batas_bawah) |
              (hls['harapan_lama_sekolah'] > batas_atas)]

# Menampilkan jumlah outlier
jumlah_outlier = outlier.shape[0]
print(f"Jumlah outlier: {jumlah_outlier}")

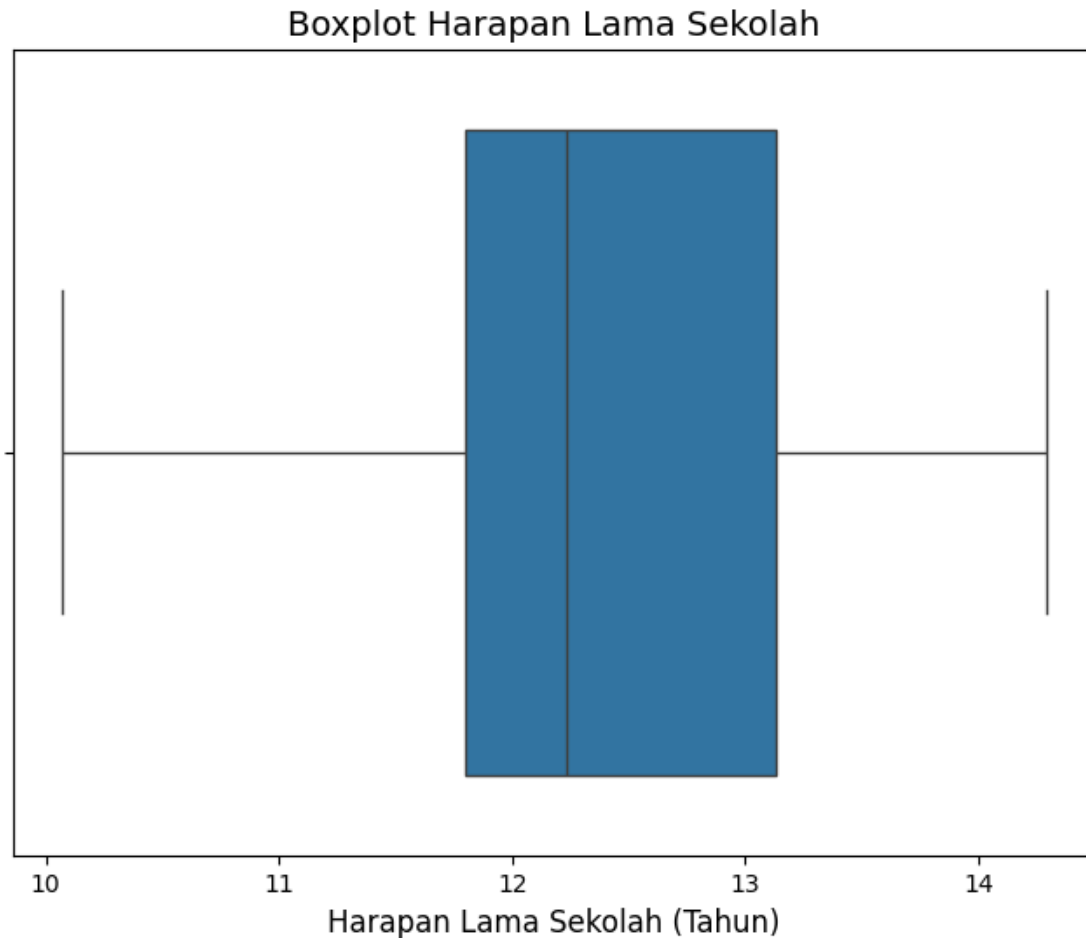
# Menampilkan outlier
if jumlah_outlier > 0:
    print("Outlier yang terdeteksi:")
    print(outlier[['nama_kabupaten_kota', 'harapan_lama_sekolah']])
else:
    print("Tidak ada outlier yang terdeteksi.")

# Menampilkan boxplot untuk visualisasi
plt.figure(figsize=(8, 6))
sns.boxplot(x=hls['harapan_lama_sekolah'])
plt.title('Boxplot Harapan Lama Sekolah', fontsize=14)
plt.xlabel('Harapan Lama Sekolah (Tahun)', fontsize=12)
plt.show()

```

Jumlah outlier: 0

Tidak ada outlier yang terdeteksi.



Disini terdeteksi ada, maka melakukan proses menormalisasikan data dengan mengganti nilai outlier dengan nilai minimal karena outlier yang terdeteksi berada di bawah nilai minimal

```
[45]: # Menghitung Q1 (25%) dan Q3 (75%) untuk harapan_lama_sekolah
Q1 = hls['harapan_lama_sekolah'].quantile(0.25)
Q3 = hls['harapan_lama_sekolah'].quantile(0.75)

# Menghitung IQR (Interquartile Range)
IQR = Q3 - Q1

# Menentukan batas bawah dan atas untuk outlier
batas_bawah = Q1 - 1.5 * IQR
batas_atas = Q3 + 1.5 * IQR

# Mendapatkan nilai minimal dari data yang berada dalam rentang non-outlier
nilai_minimal = hls[(hls['harapan_lama_sekolah'] >= batas_bawah) &
↪ (hls['harapan_lama_sekolah'] <= batas_atas)]['harapan_lama_sekolah'].min()
```

```
# Mengganti outlier dengan nilai minimal
hls['harapan_lama_sekolah'] = np.where((hls['harapan_lama_sekolah'] <
↳batas_bawah) | (hls['harapan_lama_sekolah'] > batas_atas),
                                       nilai_minimal,
↳hls['harapan_lama_sekolah'])

# Menampilkan hasil setelah mengganti outlier
print(hls[['nama_kabupaten_kota', 'harapan_lama_sekolah']])
```

	nama_kabupaten_kota	harapan_lama_sekolah
0	KABUPATEN BOGOR	10.28
1	KABUPATEN SUKABUMI	10.08
2	KABUPATEN CIANJUR	10.07
3	KABUPATEN BANDUNG	10.82
4	KABUPATEN GARUT	10.36
..	...	...
373	KOTA BEKASI	14.12
374	KOTA DEPOK	13.96
375	KOTA CIMAHI	13.84
376	KOTA TASIKMALAYA	13.49
377	KOTA BANJAR	13.27

[375 rows x 2 columns]

### 0.0.5 Mengeksplorasi Data Rata Rata Lama Sekolah Di Provinsi Jawa Barat

Cek apakah dataset sudah terkoneksi ke program

```
[46]: rrl.head()
```

```
[46]:   id  kode_provinsi  nama_provinsi  kode_kabupaten_kota  nama_kabupaten_kota \
0    1             32    JAWA BARAT             3201    KABUPATEN BOGOR
1    2             32    JAWA BARAT             3202    KABUPATEN SUKABUMI
2    3             32    JAWA BARAT             3203    KABUPATEN CIANJUR
3    4             32    JAWA BARAT             3204    KABUPATEN BANDUNG
4    5             32    JAWA BARAT             3205    KABUPATEN GARUT

      rata_rata_lama_sekolah  satuan  tahun
0              6.90  TAHUN    2010
1              5.82  TAHUN    2010
2              6.17  TAHUN    2010
3              7.98  TAHUN    2010
4              6.68  TAHUN    2010
```

Cek informasi dataset

```
[47]: rrl.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 375 entries, 0 to 374
Data columns (total 8 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   id                                     375 non-null    int64
1   kode_provinsi                         375 non-null    int64
2   nama_provinsi                         375 non-null    object
3   kode_kabupaten_kota                  375 non-null    int64
4   nama_kabupaten_kota                  375 non-null    object
5   rata_rata_lama_sekolah                375 non-null    float64
6   satuan                               375 non-null    object
7   tahun                                375 non-null    int64
dtypes: float64(1), int64(4), object(3)
memory usage: 23.6+ KB
```

Cek apakah ada data yang hilang

```
[48]: rrl.isnull().sum()
```

```
[48]: id                0
      kode_provinsi     0
      nama_provinsi     0
      kode_kabupaten_kota 0
      nama_kabupaten_kota 0
      rata_rata_lama_sekolah 0
      satuan            0
      tahun            0
      dtype: int64
```

Cek apakah ada data yang terduplikasi

```
[49]: rrl.duplicated().sum()
```

```
[49]: 0
```

Cek apakah ada outlier atau record data tidak normal yang terdeteksi

```
[108]: # Mengganti nilai yang tidak valid (jika ada) dengan NaN
      rrl['rata_rata_lama_sekolah'] = rrl['rata_rata_lama_sekolah'].replace('-', np.
      ↪nan)

      # Mengonversi kolom menjadi tipe float (jika belum)
      rrl['rata_rata_lama_sekolah'] = rrl['rata_rata_lama_sekolah'].astype(float)

      # Menghapus baris yang mengandung NaN (opsional)
      rrl_cleaned = rrl.dropna(subset=['rata_rata_lama_sekolah'])

      # Menghitung Q1 dan Q3
```



```

Q1 = rrl_cleaned['rata_rata_lama_sekolah'].quantile(0.25)
Q3 = rrl_cleaned['rata_rata_lama_sekolah'].quantile(0.75)

# Menghitung IQR
IQR = Q3 - Q1

# Menentukan batas bawah dan atas untuk outlier
batas_bawah = Q1 - 1.5 * IQR
batas_atas = Q3 + 1.5 * IQR

# Mendeteksi outlier
outlier = rrl_cleaned[(rrl_cleaned['rata_rata_lama_sekolah'] < batas_bawah) |
                      (rrl_cleaned['rata_rata_lama_sekolah'] > batas_atas)]

# Menghitung jumlah outlier
jumlah_outlier = outlier.shape[0]

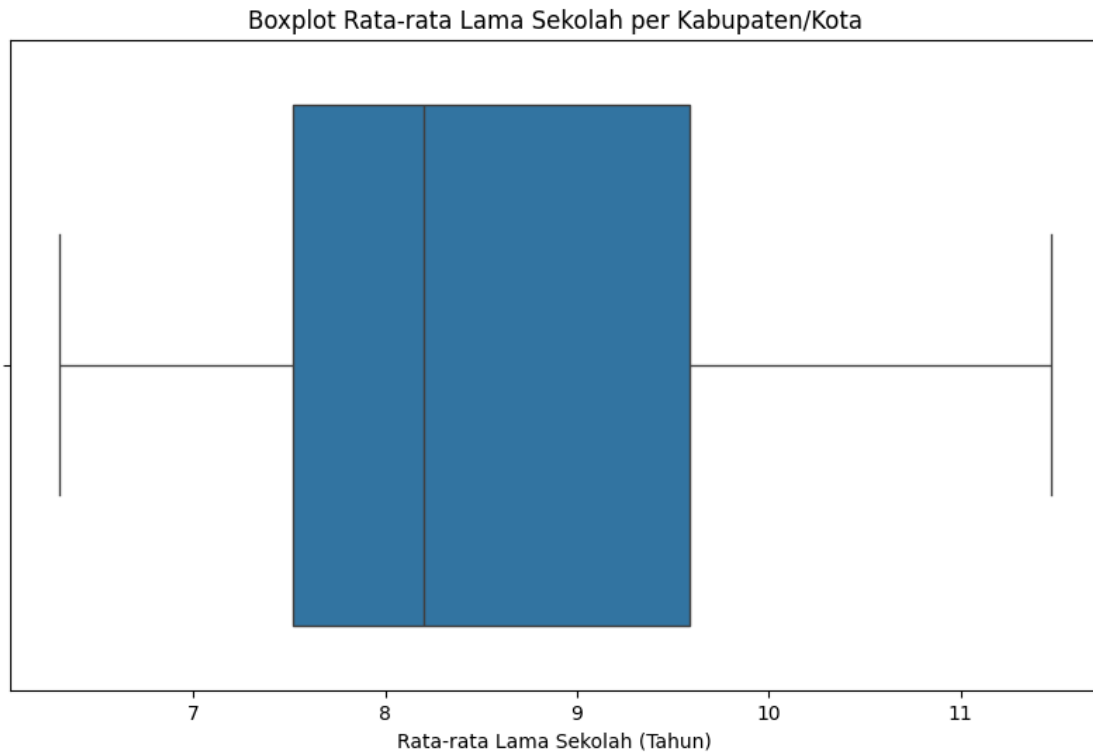
# Menampilkan jumlah outlier
jumlah_outlier = outlier.shape[0]
print(f"Jumlah outlier: {jumlah_outlier}")

# Menampilkan outlier
if jumlah_outlier > 0:
    print("Outlier yang terdeteksi:")
    print(outlier[['nama_kabupaten_kota', 'harapan_lama_sekolah']])
else:
    print("Tidak ada outlier yang terdeteksi.")

# Siapkan plot
plt.figure(figsize=(10, 6))
sns.boxplot(data=rrl, x='rata_rata_lama_sekolah')
plt.title("Boxplot Rata-rata Lama Sekolah per Kabupaten/Kota")
plt.xlabel("Rata-rata Lama Sekolah (Tahun)")
plt.show()

```

Jumlah outlier: 0  
Tidak ada outlier yang terdeteksi.



Tidak ada outlier yang terdeteksi dalam data `rata_rata_lama_sekolah`. Semua data berada dalam rentang yang dianggap normal sesuai dengan perhitungan IQR.

Menghilangkan Kolom `kode_provinsi` dan `nama_provinsi` karena berasal dari 1 provinsi

```
[51]: rrl = rrl.drop(columns=['kode_provinsi', 'nama_provinsi'])
      rrl.head()
```

```
[51]:
```

	id	kode_kabupaten_kota	nama_kabupaten_kota	rata_rata_lama_sekolah	satuan	\
0	1	3201	KABUPATEN BOGOR	6.90	TAHUN	
1	2	3202	KABUPATEN SUKABUMI	5.82	TAHUN	
2	3	3203	KABUPATEN CIANJUR	6.17	TAHUN	
3	4	3204	KABUPATEN BANDUNG	7.98	TAHUN	
4	5	3205	KABUPATEN GARUT	6.68	TAHUN	

	tahun
0	2010
1	2010
2	2010
3	2010
4	2010

### 0.0.6 Mengeksplorasi dataset Angka Melek Huruf Penduduk Usia 15 Tahun ke Atas

Cek apakah dataset sudah ter-import ke program dengan menampilkan 5 record awal

```
[52]: amh.head()
```

```
[52]:
```

	id	kode_provinsi	nama_provinsi	kode_kabupaten_kota	nama_kabupaten_kota	\
0	1	32	JAWA BARAT	3201	KABUPATEN BOGOR	
1	2	32	JAWA BARAT	3202	KABUPATEN SUKABUMI	
2	3	32	JAWA BARAT	3203	KABUPATEN CIANJUR	
3	4	32	JAWA BARAT	3204	KABUPATEN BANDUNG	
4	5	32	JAWA BARAT	3205	KABUPATEN GARUT	

	angka_melek_huruf	satuan	tahun
0	93.0	PERSEN	2004
1	96.0	PERSEN	2004
2	96.5	PERSEN	2004
3	97.8	PERSEN	2004
4	97.7	PERSEN	2004

Penjelasan Kolom Tabel:

- Kolom **angka\_melek\_huruf** adalah persentase penduduk berusia 15 tahun ke atas yang bisa membaca dan menulis, digunakan sebagai ukuran tingkat literasi pada populasi tiap kabupaten/kota di Jawa Barat
- Kolom **satuan** adalah penunjuk bahwa nilai dalam kolom angka\_melek\_huruf diukur dalam persentase (%), yang menunjukkan persentase penduduk yang melek huruf

Cek info dari dataset yang dimiliki (column, types, jmlh record)

```
[53]: amh.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 378 entries, 0 to 377
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                    378 non-null   int64
1   kode_provinsi         378 non-null   int64
2   nama_provinsi         378 non-null   object
3   kode_kabupaten_kota   378 non-null   int64
4   nama_kabupaten_kota   378 non-null   object
5   angka_melek_huruf     378 non-null   float64
6   satuan                378 non-null   object
7   tahun                378 non-null   int64
dtypes: float64(1), int64(4), object(3)
memory usage: 23.8+ KB
```

Cek apakah ada data yang terduplikat

```
[54]: amh.duplicated().sum()
```

```
[54]: 0
```

Cek apakah ada record data yang null/belum diisi

```
[55]: amh.isnull().sum()
```

```
[55]: id                0
      kode_provinsi    0
      nama_provinsi    0
      kode_kabupaten_kota  0
      nama_kabupaten_kota  0
      angka_melek_huruf  0
      satuan           0
      tahun           0
      dtype: int64
```

Cek apakah tipe data dari tiap kolom/atribut

```
[56]: amh.dtypes
```

```
[56]: id                int64
      kode_provinsi    int64
      nama_provinsi    object
      kode_kabupaten_kota  int64
      nama_kabupaten_kota  object
      angka_melek_huruf  float64
      satuan           object
      tahun           int64
      dtype: object
```

Menghapus kolom 'id\_provinsi' & 'nama\_provinsi' karena sudah jelas data nya hanya berasal dari 1 provinsi

```
[57]: amh = amh.drop(columns=['kode_provinsi', 'nama_provinsi'])

      amh.head()
```

```
[57]:   id  kode_kabupaten_kota  nama_kabupaten_kota  angka_melek_huruf  satuan \
0    1          3201      KABUPATEN BOGOR          93.0  PERSEN
1    2          3202      KABUPATEN SUKABUMI          96.0  PERSEN
2    3          3203      KABUPATEN CIANJUR          96.5  PERSEN
3    4          3204      KABUPATEN BANDUNG          97.8  PERSEN
4    5          3205      KABUPATEN GARUT          97.7  PERSEN

      tahun
0    2004
```

```
1 2004
2 2004
3 2004
4 2004
```

Cek statistik dari data `angka_melek_huruf` (jumlah, rata-rata, std deviasi, nilmin, nilmax, q1, q2, q3)

```
[58]: amh.describe().angka_melek_huruf
```

```
[58]: count      378.000000
      mean       94.056481
      std        16.682850
      min         0.000000
      25%        95.307500
      50%        98.010000
      75%        99.140000
      max        99.950000
      Name: angka_melek_huruf, dtype: float64
```

Cek apakah ada outlier atau record data tidak normal yang terdeteksi

```
[109]: # Misalkan amh adalah DataFrame Anda
      # Menghitung Q1 (25%) dan Q3 (75%)
      Q1 = amh['angka_melek_huruf'].quantile(0.25)
      Q3 = amh['angka_melek_huruf'].quantile(0.75)

      # Menghitung IQR (Interquartile Range)
      IQR = Q3 - Q1

      # Menentukan batas bawah dan atas untuk outlier
      batas_bawah = Q1 - 1.5 * IQR
      batas_atas = Q3 + 1.5 * IQR

      # Mendeteksi outlier
      outlier = amh[(amh['angka_melek_huruf'] < batas_bawah) |
                    (amh['angka_melek_huruf'] > batas_atas)]

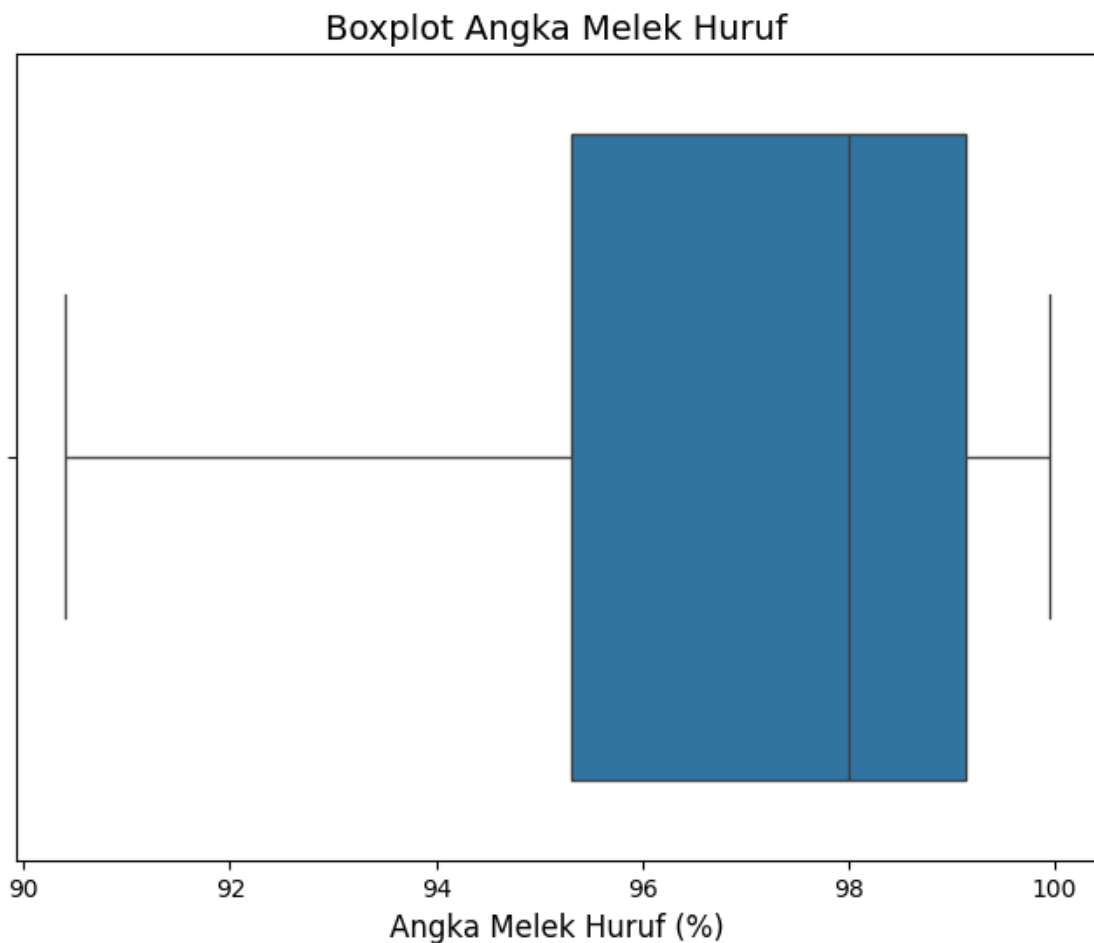
      # Menampilkan jumlah outlier
      jumlah_outlier = outlier.shape[0]
      print(f"Jumlah outlier: {jumlah_outlier}")

      # Menampilkan outlier
      if jumlah_outlier > 0:
          print("Outlier yang terdeteksi:")
          print(outlier[['nama_kabupaten_kota', 'angka_melek_huruf']])
      else:
          print("Tidak ada outlier yang terdeteksi.")
```

```
# Menampilkan boxplot untuk visualisasi
plt.figure(figsize=(8, 6))
sns.boxplot(x=amh['angka_melek_huruf'])
plt.title('Boxplot Angka Melek Huruf', fontsize=14)
plt.xlabel('Angka Melek Huruf (%)', fontsize=12)
plt.show()
```

Jumlah outlier: 0

Tidak ada outlier yang terdeteksi.



Disini terdeteksi ada, maka melakukan proses menormalisasikan data dengan mengganti nilai outlier dengan nilai minimal karena outlier yang terdeteksi berada di bawah nilai minimal

```
[60]: # Menghitung Q1 (25%) dan Q3 (75%) untuk angka_melek_huruf
Q1 = amh['angka_melek_huruf'].quantile(0.25)
Q3 = amh['angka_melek_huruf'].quantile(0.75)
```

```

# Menghitung IQR (Interquartile Range)
IQR = Q3 - Q1

# Menentukan batas bawah dan atas untuk outlier
batas_bawah = Q1 - 1.5 * IQR
batas_atas = Q3 + 1.5 * IQR

# Mendapatkan nilai minimal dari data yang berada dalam rentang non-outlier
nilai_minimal = amh[(amh['angka_melek_huruf'] >= batas_bawah) &
↳ (amh['angka_melek_huruf'] <= batas_atas)]['angka_melek_huruf'].min()

# Mengganti outlier dengan nilai minimal
amh['angka_melek_huruf'] = np.where((amh['angka_melek_huruf'] < batas_bawah) |
↳ (amh['angka_melek_huruf'] > batas_atas),
                                   nilai_minimal, amh['angka_melek_huruf'])

# Menampilkan hasil setelah mengganti outlier
print(amh[['nama_kabupaten_kota', 'angka_melek_huruf']])

```

	nama_kabupaten_kota	angka_melek_huruf
0	KABUPATEN BOGOR	93.00
1	KABUPATEN SUKABUMI	96.00
2	KABUPATEN CIANJUR	96.50
3	KABUPATEN BANDUNG	97.80
4	KABUPATEN GARUT	97.70
..	...	...
373	KOTA BEKASI	99.40
374	KOTA DEPOK	99.19
375	KOTA CIMAHI	99.64
376	KOTA TASIKMALAYA	99.81
377	KOTA BANJAR	98.95

[378 rows x 2 columns]

### 0.0.7 Mengeksplorasi dataset Tingkat Pengangguran Terbuka

Cek apakah dataset sudah ter-import ke program dengan menampilkan 5 record awal

```
[61]: jpt.head()
```

```

[61]:   id  kode_provinsi  nama_provinsi  kode_kabupaten_kota  nama_kabupaten_kota  \
0    1             32    JAWA BARAT             3201    KABUPATEN BOGOR
1    2             32    JAWA BARAT             3202    KABUPATEN SUKABUMI
2    3             32    JAWA BARAT             3203    KABUPATEN CIANJUR
3    4             32    JAWA BARAT             3204    KABUPATEN BANDUNG
4    5             32    JAWA BARAT             3205    KABUPATEN GARUT

```

	tingkat_pengangguran_terbuka	satuan	tahun
0	14.26	PERSEN	2007
1	10.85	PERSEN	2007
2	13.82	PERSEN	2007
3	17.37	PERSEN	2007
4	12.18	PERSEN	2007

Penjelasan Kolom Tabel:

- Kolom **tingkat\_pengangguran\_terbuka** adalah persentase penduduk yang termasuk dalam angkatan kerja namun tidak memiliki pekerjaan, digunakan sebagai ukuran tingkat pengangguran terbuka di wilayah kabupaten/kota Jawa Barat
- Kolom **satuan** adalah penunjuk bahwa nilai dalam kolom tingkat\_pengangguran\_terbuka diukur dalam persentase (%), yang menunjukkan persentase penduduk yang menganggur dalam angkatan kerja.

Cek info dari dataset yang dimiliki (column, types, jumlah record)

```
[62]: jpt.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 397 entries, 0 to 396
Data columns (total 8 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   id                                     397 non-null    int64
1   kode_provinsi                         397 non-null    int64
2   nama_provinsi                         397 non-null    object
3   kode_kabupaten_kota                  397 non-null    int64
4   nama_kabupaten_kota                  397 non-null    object
5   tingkat_pengangguran_terbuka         397 non-null    float64
6   satuan                               397 non-null    object
7   tahun                                397 non-null    int64
dtypes: float64(1), int64(4), object(3)
memory usage: 24.9+ KB
```

Cek apakah ada data yang terduplikat

```
[63]: jpt.duplicated().sum()
```

```
[63]: 0
```

Cek apakah ada record data yang null/belum diisi

```
[64]: jpt.isnull().sum()
```

```
[64]: id                0
      kode_provinsi     0
      nama_provinsi     0
```



```

kode_kabupaten_kota      0
nama_kabupaten_kota      0
tingkat_pengangguran_terbuka  0
satuan                   0
tahun                    0
dtype: int64

```

Cek apakah tipe data dari tiap kolom/atribut

```
[65]: jpt.dtypes
```

```

[65]: id                int64
      kode_provinsi     int64
      nama_provinsi     object
      kode_kabupaten_kota int64
      nama_kabupaten_kota object
      tingkat_pengangguran_terbuka float64
      satuan            object
      tahun             int64
      dtype: object

```

Menghapus kolom 'id\_provinsi' & 'nama\_provinsi' karena sudah jelas data nya hanya berasal dari 1 provinsi

```

[66]: jpt = jpt.drop(columns=['kode_provinsi', 'nama_provinsi'])

      jpt.head()

```

```

[66]:   id  kode_kabupaten_kota  nama_kabupaten_kota  tingkat_pengangguran_terbuka  \
0    1                3201      KABUPATEN BOGOR                        14.26
1    2                3202      KABUPATEN SUKABUMI                       10.85
2    3                3203      KABUPATEN CIANJUR                        13.82
3    4                3204      KABUPATEN BANDUNG                       17.37
4    5                3205      KABUPATEN GARUT                        12.18

      satuan  tahun
0  PERSEN    2007
1  PERSEN    2007
2  PERSEN    2007
3  PERSEN    2007
4  PERSEN    2007

```

Cek statistik dari data tingkat\_pengangguran\_terbuka (jumlah, rata-rata, std deviasi, nilmin, nilmax, q1, q2, q3)

```
[67]: jpt.describe().tingkat_pengangguran_terbuka
```

```
[67]: count      397.000000
      mean        9.114786
      std         2.833911
      min         1.520000
      25%         7.490000
      50%         9.010000
      75%        10.380000
      max        22.150000
      Name: tingkat_pengangguran_terbuka, dtype: float64
```

Cek apakah ada outlier atau record data tidak normal yang terdeteksi

```
[110]: # Menghitung Q1 (25%) dan Q3 (75%)
Q1 = jpt['tingkat_pengangguran_terbuka'].quantile(0.25)
Q3 = jpt['tingkat_pengangguran_terbuka'].quantile(0.75)

# Menghitung IQR (Interquartile Range)
IQR = Q3 - Q1

# Menentukan batas bawah dan atas untuk outlier
batas_bawah = Q1 - 1.5 * IQR
batas_atas = Q3 + 1.5 * IQR

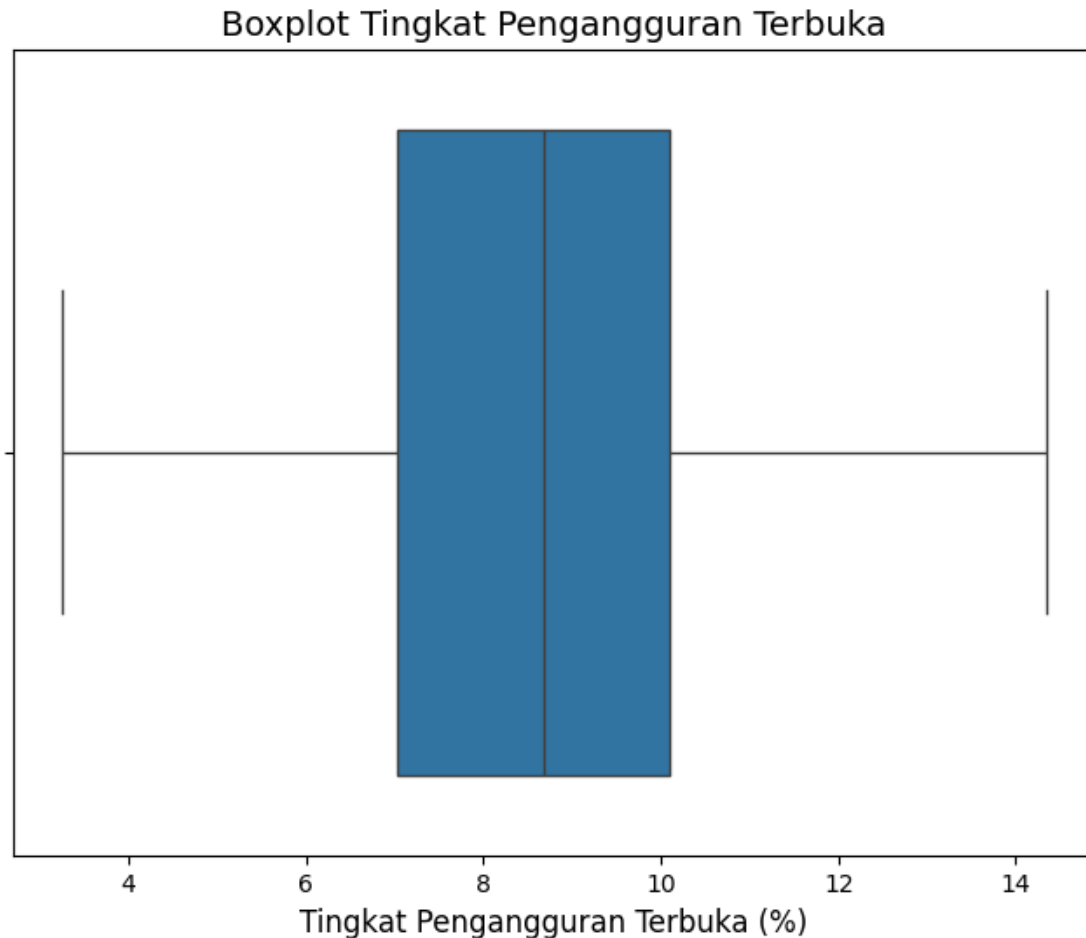
# Mendeteksi outlier
outlier = jpt[(jpt['tingkat_pengangguran_terbuka'] < batas_bawah) |
              (jpt['tingkat_pengangguran_terbuka'] > batas_atas)]

# Menampilkan jumlah outlier
jumlah_outlier = outlier.shape[0]
print(f"Jumlah outlier: {jumlah_outlier}")

# Menampilkan outlier
if jumlah_outlier > 0:
    print("Outlier yang terdeteksi:")
    print(outlier[['nama_kabupaten_kota', 'tingkat_pengangguran_terbuka']])
else:
    print("Tidak ada outlier yang terdeteksi.")

# Menampilkan boxplot untuk visualisasi
plt.figure(figsize=(8, 6))
sns.boxplot(x=jpt['tingkat_pengangguran_terbuka'])
plt.title('Boxplot Tingkat Pengangguran Terbuka', fontsize=14)
plt.xlabel('Tingkat Pengangguran Terbuka (%)', fontsize=12)
plt.show()
```

```
Jumlah outlier: 0
Tidak ada outlier yang terdeteksi.
```



Disini terdeteksi ada, maka melakukan proses menormalisasikan data dengan mengganti nilai outlier dengan nilai minimal karena outlier yang terdeteksi berada di bawah nilai minimal

```
[69]: # Menghitung Q1 (25%) dan Q3 (75%) untuk tingkat_pengangguran_terbuka
Q1 = jpt['tingkat_pengangguran_terbuka'].quantile(0.25)
Q3 = jpt['tingkat_pengangguran_terbuka'].quantile(0.75)

# Menghitung IQR (Interquartile Range)
IQR = Q3 - Q1

# Menentukan batas bawah dan atas untuk outlier
batas_bawah = Q1 - 1.5 * IQR
batas_atas = Q3 + 1.5 * IQR

# Mendapatkan nilai minimal dari data yang berada dalam rentang non-outlier
```

```

nilai_minimal = jpt[(jpt['tingkat_pengangguran_terbuka'] >= batas_bawah) &
↳(jpt['tingkat_pengangguran_terbuka'] <=
↳batas_atas)]['tingkat_pengangguran_terbuka'].min()

# Mengganti outlier dengan nilai minimal
jpt['tingkat_pengangguran_terbuka'] = np.
↳where((jpt['tingkat_pengangguran_terbuka'] < batas_bawah) |
↳(jpt['tingkat_pengangguran_terbuka'] > batas_atas),
                                             nilai_minimal,
↳jpt['tingkat_pengangguran_terbuka'])

# Menampilkan hasil setelah mengganti outlier
print(jpt[['nama_kabupaten_kota', 'tingkat_pengangguran_terbuka']])

```

	nama_kabupaten_kota	tingkat_pengangguran_terbuka
0	KABUPATEN BOGOR	14.26
1	KABUPATEN SUKABUMI	10.85
2	KABUPATEN CIANJUR	13.82
3	KABUPATEN BANDUNG	3.25
4	KABUPATEN GARUT	12.18
..	...	...
392	KOTA BEKASI	7.90
393	KOTA DEPOK	6.97
394	KOTA CIMAHI	10.52
395	KOTA TASIKMALAYA	6.55
396	KOTA BANJAR	5.43

[397 rows x 2 columns]

## 0.1 Analysis & Visualizations

### 0.1.1 Analisis dan Visualisasi dataset Indeks Pendidikan

Analisis dan Visualisasi dataset Indeks Pendidikan Provinsi di Indonesia Visualisasi record data dari Tren Indeks Pendidikan by Java Central atau tersentralisasi di Jawa

```

[70]: provinsi_pjawa = ['BANTEN', 'JAWA BARAT', 'DKI JAKARTA', 'JAWA TENGAH', 'JAWA_
↳TIMUR', 'DI YOGYAKARTA']

pjawa_comparison = ipp[(ipp['provinsi'].isin(provinsi_pjawa)) & (ipp['tahun']
↳>= 2010) & (ipp['tahun'] <= 2018)]

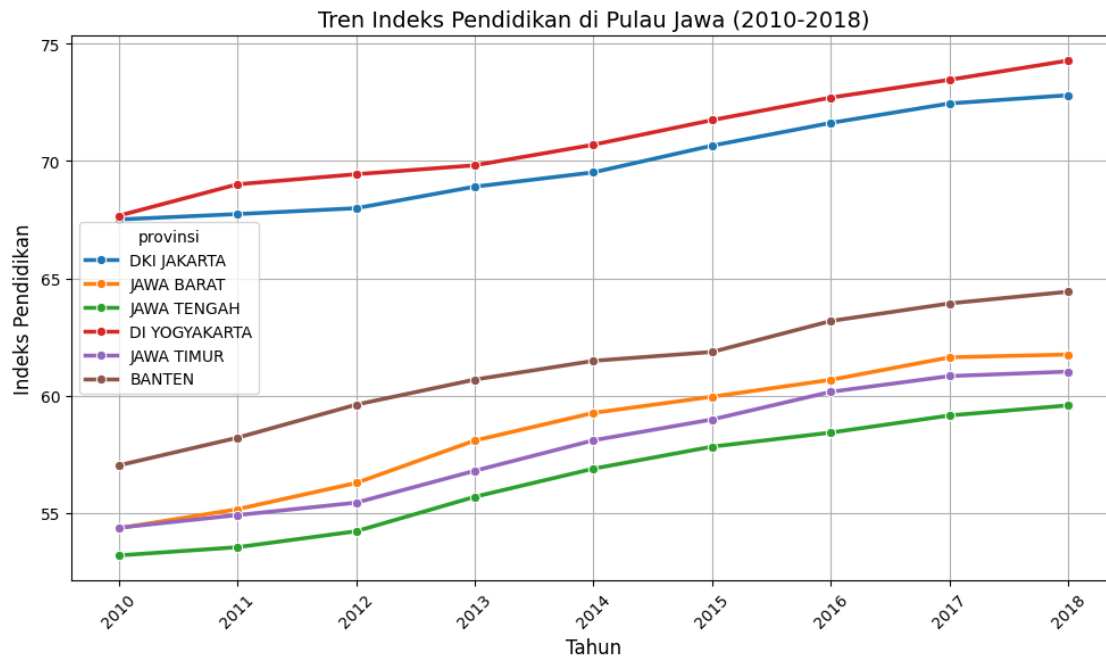
#print(pjawa_comparison)

plt.figure(figsize=(10, 6))
sns.lineplot(x='tahun', y='indeks_pendidikan', hue='provinsi',
↳data=pjawa_comparison, marker='o', linewidth=2.5)

```

```
plt.title('Tren Indeks Pendidikan di Pulau Jawa (2010-2018)', fontsize=14)
plt.xlabel('Tahun', fontsize=12)
plt.ylabel('Indeks Pendidikan', fontsize=12)
plt.xticks(rotation=45)
plt.grid(True)
plt.tight_layout()

plt.show()
warnings.filterwarnings('ignore')
```



Plot diagram diatas adalah lineplot atau diagram garis yang menunjukkan perubahan indeks pendidikan dari masing-masing provinsi yang ada di pulau jawa sebagai pembandingan terhadap provinsi jawa barat berdasarkan timeline waktu per tahunnya.

- Sumbu x menunjukkan rentang waktu atau timeline tahun (2010-2018).
- Sumbu y menunjukkan rentang indeks pendidikan yang diraih tiap provinsi nya.
- Keterangan warna garis menunjukkan label provinsi.

Keseluruhan diagram menampilkan perubahan indeks pendidikan masing-masing provinsi di pulau jawa dilihat dari tiap tahunnya.

Analisis Singkat:

Dari diagram yang sudah di plot dapat dilihat bahwa di pulau jawa sendiri, DI Yogyakarta menempati posisi pertama dalam indeks pendidikan tertinggi, sedangkan jawa barat sendiri berada di peringkat ke 4 yang berarti di pulau jawa sendiri jawa barat tidaklah terlalu bagus indeks pendidikannya karena menjadi golongan yang menengah ke bawah dalam lingkup pulau jawa. Ini dapat menjadi analisis baru melihat apa yang menjadi faktor bagi jawa barat memiliki indeks pendidikan

yang tidak terlalu baik itu.

### Visualisasi record data dari Tren Indeks Pendidikan by Tertinggi dari tiap Pulau

```
[71]: provinsi_psemua = ['DI YOGYAKARTA', 'ACEH', 'KALIMANTAN TIMUR', 'SULAWESI_
↳TENGGERA', 'PAPUA BARAT',
                        'JAWA BARAT', 'BALI', 'NUSA TENGGARA TIMUR', 'NUSA_
↳TENGGERA BARAT']

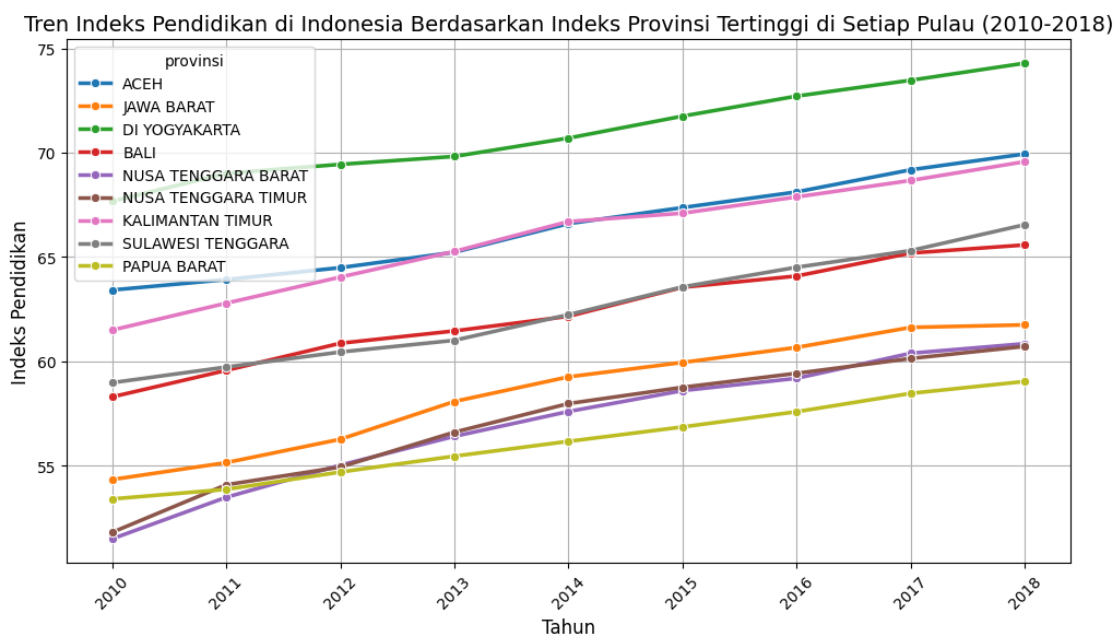
psemua_comparison = ipp[(ipp['provinsi'].isin(provinsi_psemua)) & (ipp['tahun']_
↳>= 2010) & (ipp['tahun'] <= 2018)]

#print(psemua_comparison)

plt.figure(figsize=(10, 6))
sns.lineplot(x='tahun', y='indeks_pendidikan', hue='provinsi',_
↳data=psemua_comparison, marker='o', linewidth=2.5)

plt.title('Tren Indeks Pendidikan di Indonesia Berdasarkan Indeks Provinsi_
↳Tertinggi di Setiap Pulau (2010-2018)', fontsize=14)
plt.xlabel('Tahun', fontsize=12)
plt.ylabel('Indeks Pendidikan', fontsize=12)
plt.xticks(rotation=45)
plt.grid(True)
plt.tight_layout()

plt.show()
warnings.filterwarnings('ignore')
```



Plot diagram diatas adalah lineplot atau diagram garis yang menunjukkan perubahan indeks pendidikan dari masing-masing provinsi yang ada di Indonesia dan diambil yang tertinggi dari tiap pulau nya (Pulau-pulau besar) sebagai pembandingan terhadap provinsi jawa barat berdasarkan timeline waktu per tahunnya.

- Sumbu x menunjukkan rentang waktu atau timeline tahun (2010-2018).
- Sumbu y menunjukkan rentang indeks pendidikan yang diraih tiap provinsi nya.
- Keterangan warna garis menunjukkan label provinsi.

Keseluruhan diagram menampilkan perubahan indeks pendidikan masing-masing provinsi di Indonesia diambil yang tertinggi dari tiap pulau nya (Pulau-pulau besar) dan dilihat perubahan dari tiap tahunnya.

Analisis Singkat:

Dari diagram yang sudah di plot, dapat dilihat indeks pendidikan tertinggi dari masing-masing provinsi tiap pulau besar dan beberapa kelompok pulau pilihan dari seluruh Indonesia dibandingkan dengan provinsi Jawa Barat. Didapat hasil bahwa Jawa Barat masih belum dapat bersaing baik dengan provinsi lain yang memiliki nilai indeks pendidikan tertinggi di pulau nya, padahal secara fakta pembangunan di pulau Jawa adalah yang paling pesat diantara pulau lainnya tapi tidak menutup kemungkinan bahwa masih banyak provinsi yang perlu mendapatkan penanganan salah satunya Jawa Barat yang terletak di pulau Jawa namun tidak dapat bersaing baik dengan provinsi lainnya diluar pulau Jawa, tentunya ini dapat berlanjut ke analisis lain untuk melihat pengaruh internal yang ada di Jabar itu sendiri dan sebagai laporan yang harus ditangani oleh pemangku kebijakan.

## KESIMPULAN

Dari data yang sudah dieksplorasi dan disajikan dalam bentuk diagram garis, dapat disimpulkan:

- Jawa Barat belum menjadi provinsi TOP 5 indeks pendidikan tertinggi di INDONESIA, ini menandakan adanya indikasi serius yang harus ditelaah lebih lanjut mengapa Jawa Barat tergolong cukup rendah indeks pendidikannya dibandingkan indeks tertinggi provinsi dari pulau lain, secara Jawa adalah pulau yang cukup pesat pembangunan dan pengucuran dana dari pemerintah.
- Indeks Pendidikan yang dipengaruhi oleh jumlah partisipasi aktif siswa yang sekolah, mungkin menjadi asumsi mengapa Jawa Barat cukup rendah indeks pendidikannya karena populasi yang lebih banyak.
- Didapat hasil menyeluruh juga, bahwa seluruh indeks pendidikan dari kesepuluh provinsi yang tersebar diberbagai pulau memiliki reputasi line chart yang baik karena mengalami peningkatan yang terus menerus sepanjang tahun 2010 - 2018, tetapi ini tidak menjadikan Jawa Barat bisa menembus TOP 5 indeks pendidikan tertinggi yang paling tinggi nya sekarang masih diduduki oleh DI Yogyakarta.

**Analisis dan Visualisasi dataset Indeks Pendidikan Kabupaten/Kota di Jawa Barat**  
**Visualisasi record data dari Tren Indeks Pendidikan by Top 10 Tertinggi di Jawa Barat**

```
[72]: # Mengelompokkan data berdasarkan kabupaten/kota, menghitung rata-rata indeks_
      ↪ pendidikan
avg_indeks_per_kabkot = ipk.groupby('nama_kabupaten_kota')['indeks_pendidikan'].
      ↪ mean().reset_index()

# Mengurutkan berdasarkan indeks pendidikan tertinggi
top10_kabkot = avg_indeks_per_kabkot.sort_values(by='indeks_pendidikan',
      ↪ ascending=False).head(10)

# Print Top 10 kabupaten/kota dengan indeks pendidikan tertinggi
print("Top 10 Kabupaten/Kota dengan Indeks Pendidikan Tertinggi:")
print(top10_kabkot)

# Masukkan kabupaten/kota tersebut ke dalam list 'kabkot_semua'
kabkot_semua = top10_kabkot['nama_kabupaten_kota'].tolist()

# Menampilkan data untuk top 10 kabupaten/kota tersebut
kabkotsemua_comparison = ipk[(ipk['nama_kabupaten_kota'].isin(kabkot_semua)) &
      ↪ (ipk['tahun'] >= 2010) & (ipk['tahun'] <= 2022)]
```

Top 10 Kabupaten/Kota dengan Indeks Pendidikan Tertinggi:

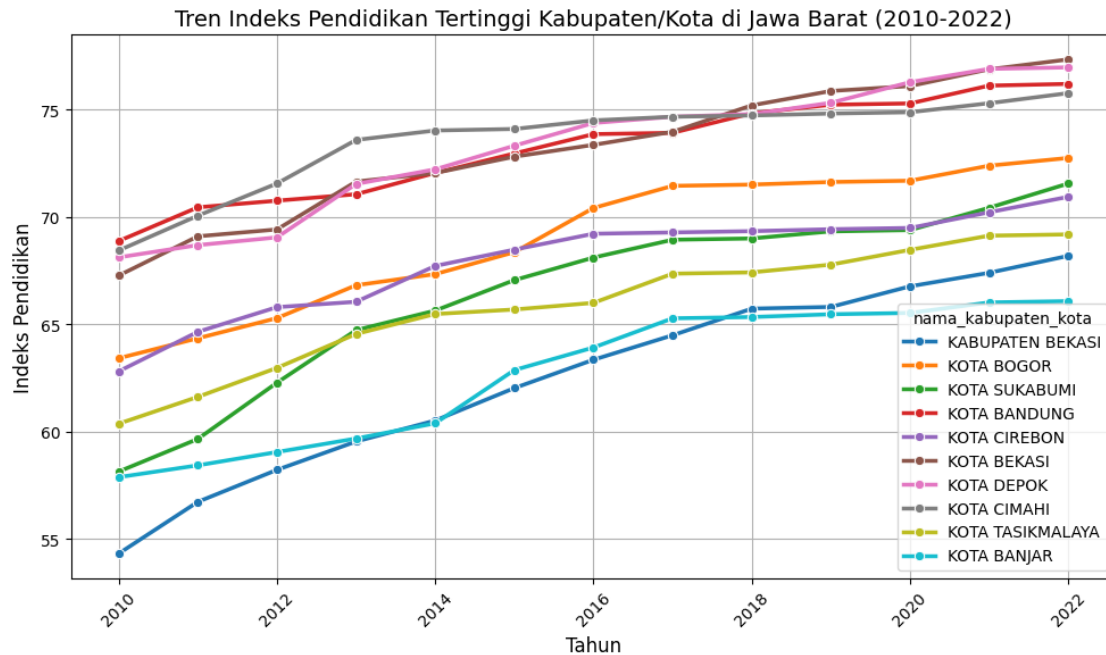
	nama_kabupaten_kota	indeks_pendidikan
22	KOTA CIMAHI	73.563077
24	KOTA DEPOK	73.240000
18	KOTA BANDUNG	73.190769
20	KOTA BEKASI	73.145385
21	KOTA BOGOR	69.022308
23	KOTA CIREBON	67.946154
25	KOTA SUKABUMI	66.472308
26	KOTA TASIKMALAYA	65.838462
19	KOTA BANJAR	62.754615
2	KABUPATEN BEKASI	62.536154

```
[73]: plt.figure(figsize=(10, 6))
      sns.lineplot(x='tahun', y='indeks_pendidikan', hue='nama_kabupaten_kota',
      ↪ data=kabkotsemua_comparison, marker='o', linewidth=2.5)

plt.title('Tren Indeks Pendidikan Tertinggi Kabupaten/Kota di Jawa Barat_
      ↪ (2010-2022)', fontsize=14)
plt.xlabel('Tahun', fontsize=12)
plt.ylabel('Indeks Pendidikan', fontsize=12)
plt.xticks(rotation=45)
plt.grid(True)
plt.tight_layout()

plt.show()
```





Plot diagram diatas adalah lineplot atau diagram garis yang menunjukkan perubahan indeks pendidikan dari top 10 kabupaten/kota yang memiliki indeks pendidikan tertinggi di provinsi jawa barat berdasarkan timeline waktu per tahunnya.

- Sumbu x menunjukkan rentang waktu atau timeline tahun (2010-2022).
- Sumbu y menunjukkan rentang indeks pendidikan yang diraih tiap kabupaten/kota nya.
- Keterangan warna garis menunjukkan label kabupaten/kota.

Keseluruhan diagram menampilkan perubahan indeks pendidikan dari top 10 kabupaten/kota yang memiliki indeks pendidikan tertinggi di provinsi jawa barat berdasarkan timeline waktu per tahunnya.

Analisis Singkat:

Dari diagram yang sudah di plot, dapat dilihat bahwa 10 top indeks pendidikan tertinggi di jawa barat didominasi oleh pemerintahan kota dan hanya ada 1 kabupaten yaitu kabupaten bekasi, dapat dilihat juga 4 dari 10 adalah kabupaten/kota yang berdekatan dengan DKI Jakarta seperti bekasi, depok dan bogor. Tentu ini menjadi analisis yang menarik karena daerah yang dekat ibu kota akan mendapat infrastruktur yang lebih baik ketimbang daerah daerah lain yang jauh dari pusat kota, namun tidak menutup kemungkinan juga yang terpencil bukan berarti tidak bagus, seperti kota tasikmalaya dan kota sukabumi yang masuk sebagai top 10 tertinggi menandakan pengelolaan pemerintahan yang baik dan dapat ditiru dan disebarkan kemajuannya dalam indeks pendidikan ke kabupaten/kota lain yang berada di sekitarnya.

**Visualisasi record data dari Tren Indeks Pendidikan by Top 10 Terendah di Jawa Barat**

[74]: `# Mengelompokkan data berdasarkan kabupaten/kota, menghitung rata-rata indeks pendidikan`

```

avg_indeks_per_kabkot = ipk.groupby('nama_kabupaten_kota')['indeks_pendidikan'].
↳mean().reset_index()

# Mengurutkan berdasarkan indeks pendidikan terendah
bottom10_kabkot = avg_indeks_per_kabkot.sort_values(by='indeks_pendidikan',
↳ascending=True).head(10)

# Print 10 kabupaten/kota dengan indeks pendidikan terendah
print("10 Kabupaten/Kota dengan Indeks Pendidikan Terendah:")
print(bottom10_kabkot)

# Masukkan kabupaten/kota tersebut ke dalam list 'kabkot_semua'
kabkot_semua = bottom10_kabkot['nama_kabupaten_kota'].tolist()

# Menampilkan data untuk 10 kabupaten/kota dengan indeks pendidikan terendah
kabkotsemua_comparisondown = ipk[(ipk['nama_kabupaten_kota'].
↳isin(kabkot_semua)) & (ipk['tahun'] >= 2010) & (ipk['tahun'] <= 2022)]

```

10 Kabupaten/Kota dengan Indeks Pendidikan Terendah:

	nama_kabupaten_kota	indeks_pendidikan
8	KABUPATEN INDRAMAYU	51.780769
14	KABUPATEN SUBANG	53.768462
5	KABUPATEN CIANJUR	54.284615
6	KABUPATEN CIREBON	54.376923
15	KABUPATEN SUKABUMI	54.641538
7	KABUPATEN GARUT	55.660769
11	KABUPATEN MAJALENGKA	55.685385
9	KABUPATEN KARAWANG	55.973077
1	KABUPATEN BANDUNG BARAT	56.995385
13	KABUPATEN PURWAKARTA	57.016923

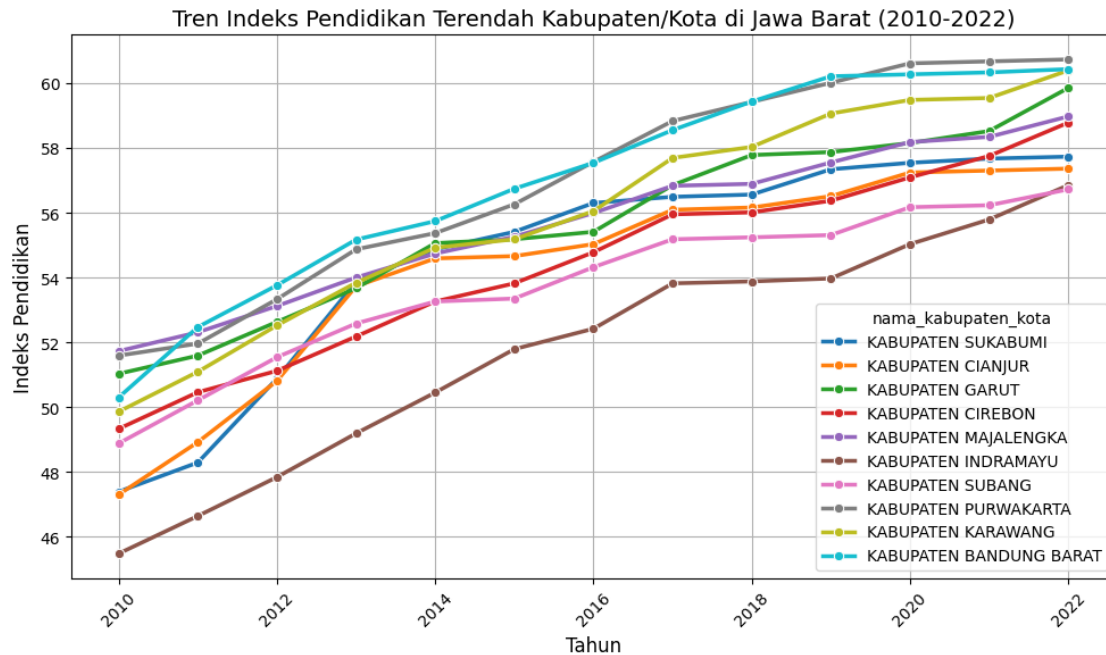
```

[75]: plt.figure(figsize=(10, 6))
sns.lineplot(x='tahun', y='indeks_pendidikan', hue='nama_kabupaten_kota',
↳data=kabkotsemua_comparisondown, marker='o', linewidth=2.5)

plt.title('Tren Indeks Pendidikan Terendah Kabupaten/Kota di Jawa Barat,
↳(2010-2022)', fontsize=14)
plt.xlabel('Tahun', fontsize=12)
plt.ylabel('Indeks Pendidikan', fontsize=12)
plt.xticks(rotation=45)
plt.grid(True)
plt.tight_layout()

plt.show()
warnings.filterwarnings('ignore')

```



Plot diagram diatas adalah lineplot atau diagram garis yang menunjukkan perubahan indeks pendidikan dari top 10 kabupaten/kota yang memiliki indeks pendidikan terendah di provinsi jawa barat berdasarkan timeline waktu per tahunnya.

- Sumbu x menunjukkan rentang waktu atau timeline tahun (2010-2022).
- Sumbu y menunjukkan rentang indeks pendidikan yang diraih tiap kabupaten/kota nya.
- Keterangan warna garis menunjukkan label kabupaten/kota.

Keseluruhan diagram menampilkan perubahan indeks pendidikan dari top 10 kabupaten/kota yang memiliki indeks pendidikan terendah di provinsi jawa barat berdasarkan timeline waktu per tahunnya.

Analisis Singkat:

Dari diagram yang sudah di plot, dapat dilihat hasil yang menunjukkan top 10 kabupaten/kota dengan indeks pendidikan terendah di provinsi jawa barat, dapat dijadikan analisis bahwa perlu nya ada penelusuran lebih lanjut terkait 10 kabupaten/kota ini kenapa memiliki indeks pendidikan yang rendah dibandingkan yang lainnya yang tentu memberi pengaruh yang cukup besar kepada indeks pendidikan di provinsi jawa barat. Ini menjadi hal yang perlu dikhawatirkan juga oleh pemangku kebijakan pasalnya daerah-daerah tersebut didominasi oleh daerah yang cukup terpencil dan jauh dari pusat kota, sehingga ada asumsi kualitas sumber daya pendidik dan infrastruktur yang harus lebih diperhatikan oleh pemerintah, naik Pemprov maupun Pemkot/pemkab.

## Kesimpulan

Dari eksplorasi data dan ditunjukkan dalam bentuk visualisasi yang dapat dianalisis, didapat kesimpulan:

- Kabupaten/kota yang mendapat perhatian lebih dari pemerintah seperti cenderung men-

dapat nilai indeks pendidikan yang lebih tinggi di Jawa Barat, seperti kabupaten dan kota Bekasi, Cirebon, Kota Depok dan lain-lain. Tentunya daerah tersebut memang daerah yang cenderung ramai penduduk sekaligus memiliki aspek penting yang dipegang setiap daerahnya seperti Bekasi dengan industri nyanya dan Cirebon dengan hasil lautnya, diasumsikan ini lah yang menjadi pusat perhatian pemerintah.

- Kabupaten/kota yang kurang mendapat perhatian dan cenderung berimbas pada nilai indeks pendidikannya yang turun adalah daerah yang dikatakan sebagai tempat singgah untuk pergi ke daerah lain atau daerah-daerah yang minim populasi penduduknya, seperti Cianjur, Subang, Indramayu dan Sukabumi. Karena itulah daerah ini diasumsikan kurang mendapat perhatian yang cukup dari pemerintah dalam membenahi kasus pendidikan disana.

Menjawab pertanyaan no. 1

Dapat dilihat dari visualisasi yang ada bahwa indeks pendidikan di Jawa Barat masih tergolong menengah-bawah dengan provinsi lain dari berbagai pulau di seluruh Indonesia, ini menjadikannya memiliki korelasi positif yang dihasilkan karena dilihat dari 10 top kabupaten/kota yang memiliki indeks pendidikan terendah di Jawa Barat menjadi daya pengaruh kenapa Jawa Barat masih tergolong memiliki indeks pendidikan menengah-bawah dari provinsi lainnya di Indonesia. Walaupun provinsi pembandingnya diambil dari yang tertinggi dari tiap pulau nya (pulau-pulau besar), namun ini tidak menjadikan alasan yang konkret Jawa Barat bisa lebih rendah, tentunya melihat pertimbangan infrastruktur yang berkembang pesat di pulau Jawa itu sendiri.

### 0.1.2 Analisis dan Visualisasi dari dataset Jumlah SMP & SMA di Provinsi Jawa Barat

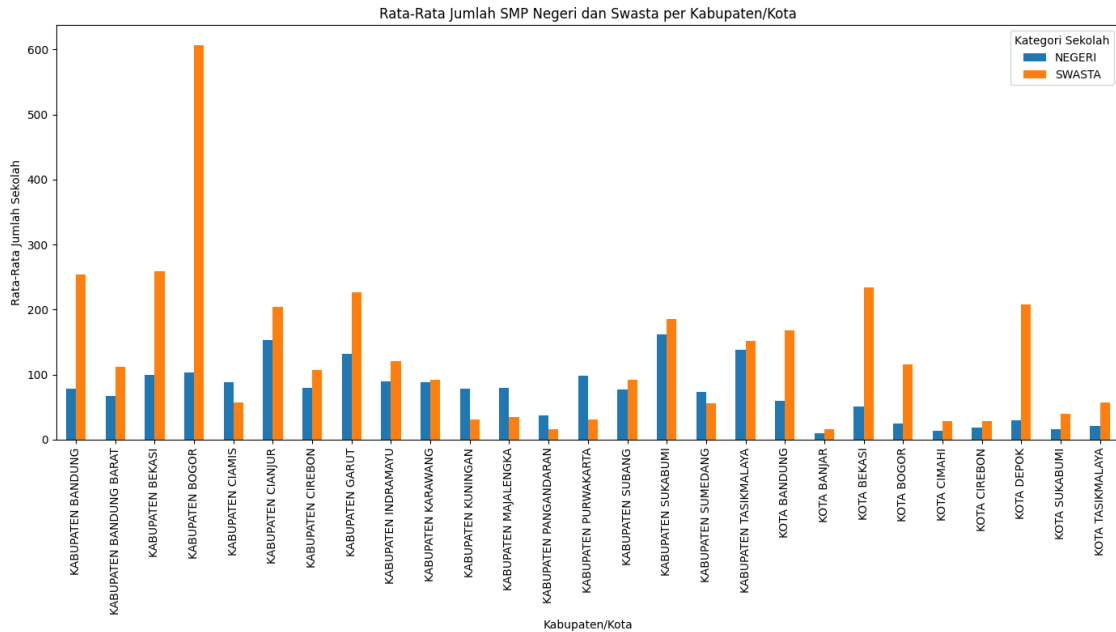
Melihat Perbandingan SMP Negeri & Swasta berdasarkan Kota/Kabupaten

```
[76]: # Mengelompokkan data berdasarkan kabupaten/kota dan kategori sekolah, lalu
      ↪ menghitung rata-rata jumlah sekolah
data_averaged = smp.groupby(['nama_kabupaten_kota',
      ↪ 'kategori_sekolah'])['jumlah_sekolah'].mean().unstack()

# Membuat grafik bar dengan bar yang dipisah
data_averaged.plot(kind='bar', stacked=False, figsize=(14, 8))

# Menambahkan judul dan label
plt.title('Rata-Rata Jumlah SMP Negeri dan Swasta per Kabupaten/Kota')
plt.xlabel('Kabupaten/Kota')
plt.ylabel('Rata-Rata Jumlah Sekolah')
plt.legend(title='Kategori Sekolah')

# Menampilkan grafik
plt.tight_layout()
plt.show()
```



Dari visualisasi di atas, terlihat bahwa Kabupaten Bogor memiliki jumlah SMP swasta yang signifikan dibandingkan dengan kabupaten/kota lainnya. Fenomena ini dapat dikaitkan dengan beberapa faktor utama:

**Luas Wilayah:** Kabupaten Bogor mencakup wilayah yang sangat luas, memberikan ruang lebih untuk pembangunan berbagai fasilitas pendidikan, termasuk sekolah swasta.

**Kepadatan Penduduk:** Tingginya jumlah penduduk di Kabupaten Bogor juga berkontribusi pada meningkatnya kebutuhan akan akses pendidikan. Untuk memenuhi kebutuhan ini, sekolah swasta hadir sebagai pelengkap bagi sekolah negeri yang ada, agar dapat mencakup kebutuhan pendidikan di seluruh wilayah.

Kedua faktor ini, yaitu luas wilayah dan tingginya jumlah penduduk, memberikan pengaruh signifikan terhadap tingginya jumlah sekolah swasta di Kabupaten Bogor. Kesimpulan ini menyoroti keterkaitan antara karakteristik demografi dan distribusi fasilitas pendidikan di wilayah tersebut, menunjukkan bahwa faktor-faktor seperti luas area dan populasi dapat menjadi variabel penting dalam analisis kebutuhan infrastruktur pendidikan.

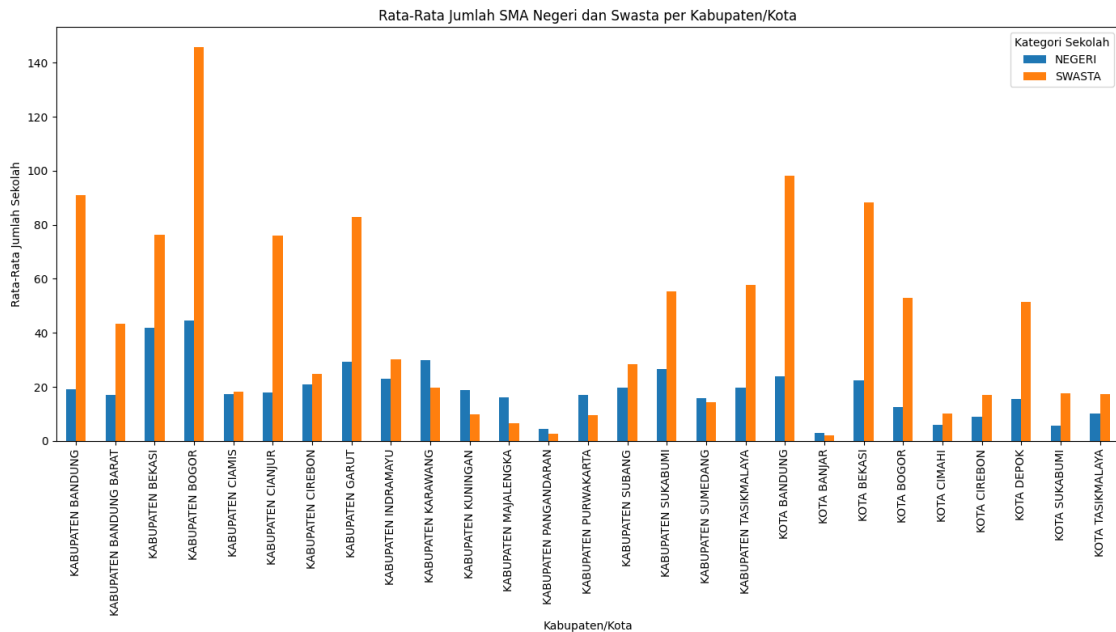
### Melihat Perbandingan SMA Negeri & Swasta berdasarkan Kota/Kabupaten

```
[77]: # Mengelompokkan data berdasarkan kabupaten/kota dan kategori sekolah, lalu
      ↪ menghitung rata-rata jumlah sekolah
data_averaged = sma.groupby(['nama_kabupaten_kota',
      ↪ 'kategori_sekolah'])['jumlah_sekolah'].mean().unstack()

# Membuat grafik bar dengan bar yang dipisah
data_averaged.plot(kind='bar', stacked=False, figsize=(14, 8))
```

```
# Menambahkan judul dan label
plt.title('Rata-Rata Jumlah SMA Negeri dan Swasta per Kabupaten/Kota')
plt.xlabel('Kabupaten/Kota')
plt.ylabel('Rata-Rata Jumlah Sekolah')
plt.legend(title='Kategori Sekolah')

# Menampilkan grafik
plt.tight_layout()
plt.show()
```



Dari visualisasi data SMA, terlihat pola distribusi yang serupa dengan data sekolah sebelumnya, yaitu dominasi jumlah sekolah swasta di wilayah tertentu. Kabupaten Bogor menonjol dengan jumlah sekolah swasta SMA yang jauh lebih tinggi dibandingkan dengan kabupaten/kota lain.

Faktor-faktor yang mendukung tingginya jumlah sekolah di wilayah ini kemungkinan besar mencakup luasnya cakupan wilayah dan tingginya jumlah penduduk yang mendorong peningkatan fasilitas pendidikan, baik negeri maupun swasta.

### 0.1.3 Analisis dan Visualisasi dari Jumlah Guru SMP & SMA di Provinsi Jawa Barat

Menampilkan data guru SMP berdasarkan Tahun Ajaran dari setiap kabupaten dan kota

```
[78]: daerah_list = gsmp['nama_kabupaten_kota'].drop_duplicates().tolist()

# Daftar tahun ajaran yang ingin digunakan
tahun_ajaran_list = ['2015/2016', '2016/2017', '2017/2018', '2018/2019',
                     '2019/2020', '2020/2021', '2021/2022', '2022/2023']
```

```

# Filter data untuk daerah dalam daerah_list dan tahun ajaran yang ditentukan
filtered_data = gsmp[(gsmp['nama_kabupaten_kota'].isin(daerah_list)) &
                     (gsmp['tahun_ajaran'].isin(tahun_ajaran_list))]

# Buat plot
plt.figure(figsize=(12,8))

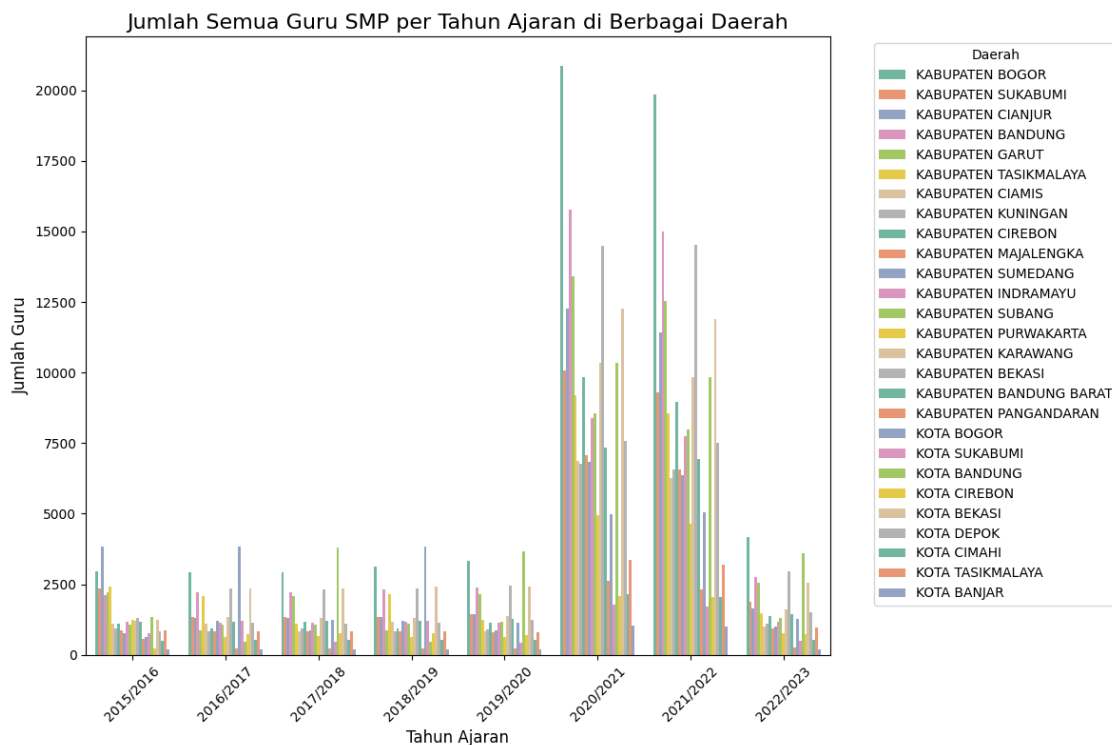
# Menggunakan seaborn untuk membuat barplot
sns.barplot(x='tahun_ajaran', y='jumlah_guru', hue='nama_kabupaten_kota',
            data=filtered_data, palette='Set2')

# Tambahkan judul dan label
plt.title('Jumlah Semua Guru SMP per Tahun Ajaran di Berbagai Daerah',
        fontsize=16)
plt.xlabel('Tahun Ajaran', fontsize=12)
plt.ylabel('Jumlah Guru', fontsize=12)
plt.xticks(rotation=45)

# Tampilkan legenda di luar plot agar lebih jelas
plt.legend(title='Daerah', bbox_to_anchor=(1.05, 1), loc='upper left')

# Tampilkan diagram
plt.tight_layout()
plt.show()

```



Dari visualisasi jumlah guru SMP di berbagai daerah Jawa Barat selama periode 2015/2016 hingga 2022/2023, beberapa poin penting dapat ditarik:

**Kenaikan Jumlah Guru SMP saat Pandemi (2020/2021):** Pada tahun ajaran 2020/2021, jumlah guru SMP mengalami peningkatan signifikan, yang bertepatan dengan awal pandemi Covid-19. Peningkatan ini kemungkinan disebabkan oleh kebutuhan tambahan tenaga pengajar SMP untuk mendukung pembelajaran daring dan penyediaan fasilitas yang memadai untuk pembelajaran jarak jauh.

**Penurunan Jumlah Guru SMP Setelah Pandemi (2022/2023):** Setelah pandemi mulai mereda, di tahun ajaran 2022/2023 jumlah guru SMP di sebagian besar daerah mengalami penurunan. Hal ini mungkin terkait dengan kembalinya metode pembelajaran tatap muka, sehingga kebutuhan tenaga pengajar SMP pun disesuaikan.

**Variasi Berdasarkan Daerah untuk Guru SMP:** Walaupun tren keseluruhan menunjukkan pola serupa—peningkatan selama masa pandemi dan penurunan setelah pandemi—masing-masing daerah memperlihatkan variasi yang berbeda. Beberapa daerah menunjukkan perubahan jumlah guru SMP yang lebih signifikan dibandingkan daerah lain, mengindikasikan adanya perbedaan kebutuhan atau kebijakan lokal yang memengaruhi perekrutan guru SMP.

```
[79]: # Daftar tahun ajaran yang ingin digunakan
tahun_ajaran_list = ['2015/2016', '2016/2017', '2017/2018', '2018/2019',
                    '2019/2020', '2020/2021', '2021/2022', '2022/2023']

# Filter data untuk daerah dalam daerah_list dan tahun ajaran yang ditentukan
filtered_data = gsmp[(gsmp['nama_kabupaten_kota'].isin(daerah_list)) &
                    (gsmp['tahun_ajaran'].isin(tahun_ajaran_list))]

# Group by 'nama_kabupaten_kota' and sum 'jumlah_guru' over all years to find
↳ top 8 regions
top_8_regions = filtered_data.groupby('nama_kabupaten_kota')['jumlah_guru'].
↳ sum().nlargest(8).index

# Filter the original data to include only these top 8 regions
top_8_data = filtered_data[filtered_data['nama_kabupaten_kota'].
↳ isin(top_8_regions)]

# Buat plot
plt.figure(figsize=(12,8))

# Menggunakan seaborn untuk membuat barplot dengan data dari 8 daerah teratas
↳ untuk semua tahun ajaran
sns.barplot(x='tahun_ajaran', y='jumlah_guru', hue='nama_kabupaten_kota',
↳ data=top_8_data, palette='Set2')

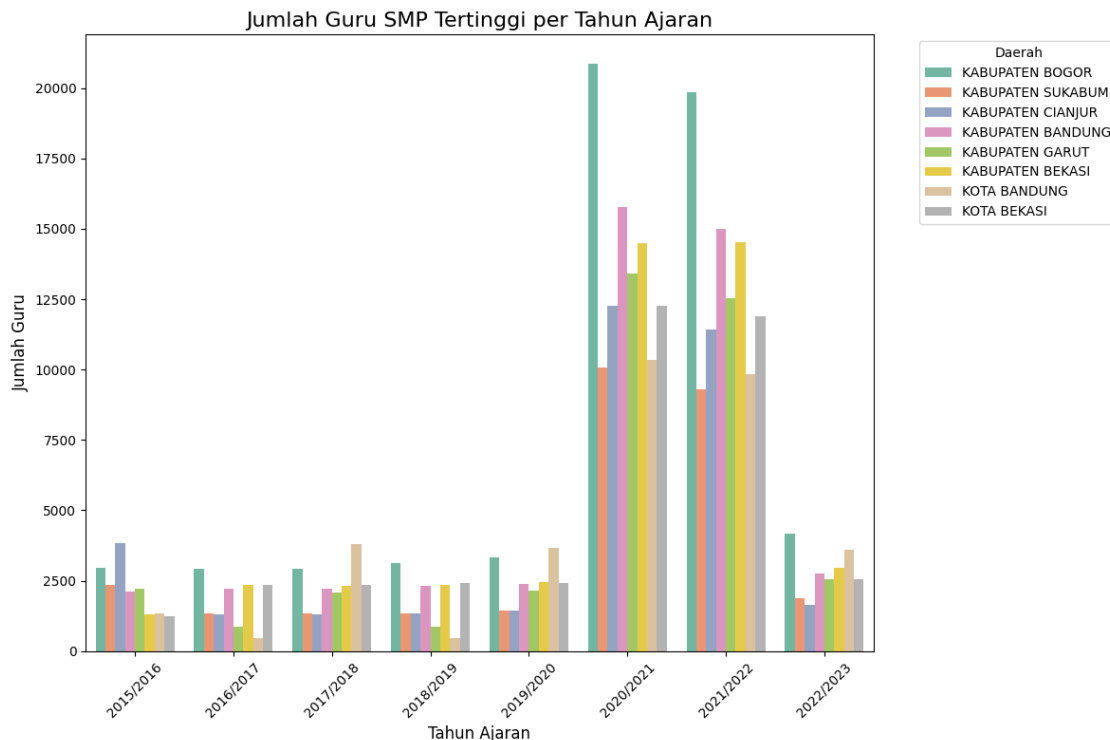
# Tambahkan judul dan label
```



```
plt.title('Jumlah Guru SMP Tertinggi per Tahun Ajaran', fontsize=16)
plt.xlabel('Tahun Ajaran', fontsize=12)
plt.ylabel('Jumlah Guru', fontsize=12)
plt.xticks(rotation=45)

# Tampilkan legenda di luar plot agar lebih jelas
plt.legend(title='Daerah', bbox_to_anchor=(1.05, 1), loc='upper left')

# Tampilkan diagram
plt.tight_layout()
plt.show()
```



Dari diagram mengenai delapan besar Kabupaten/Kota di Jawa Barat, terdapat beberapa poin yang dapat disimpulkan:

1. Kabupaten Bogor menunjukkan tren peningkatan jumlah guru yang stabil hingga masa pandemi Covid-19, tetapi mengalami penurunan tajam setelah pandemi berakhir.
2. Kota Bandung memperlihatkan fluktuasi tajam pada tahun ajaran 2015/2016 hingga 2019/2020, dengan pola kenaikan, penurunan, peningkatan lagi, diikuti oleh penurunan dan kenaikan. Ketika pandemi Covid-19 dimulai pada 2020/2021, terjadi lonjakan signifikan.
3. Kota Bekasi tampak relatif stabil tanpa penurunan besar pada jumlah guru sebelum pandemi berakhir.
4. Kabupaten Bogor menduduki posisi teratas dalam jumlah guru selama masa pandemi Covid-

19.

5. Jumlah guru sebelum dan setelah pandemi terlihat serupa di sebagian besar daerah, dengan hanya beberapa wilayah yang mengalami kenaikan jumlah guru setelah pandemi.

```
[80]: # Daftar tahun ajaran yang ingin digunakan
tahun_ajaran_list = ['2015/2016', '2016/2017', '2017/2018', '2018/2019',
                    '2019/2020', '2020/2021', '2021/2022', '2022/2023']

# Filter data untuk daerah dalam daerah_list dan tahun ajaran yang ditentukan
filtered_data = gsmp[(gsmp['nama_kabupaten_kota'].isin(daerah_list)) &
                    (gsmp['tahun_ajaran'].isin(tahun_ajaran_list))]

# Group by 'nama_kabupaten_kota' and sum 'jumlah_guru' over all years to find
↳bottom 8 regions
bottom_8_regions = filtered_data.groupby('nama_kabupaten_kota')['jumlah_guru'].
↳sum().nsmallest(8).index

# Filter the original data to include only these bottom 8 regions
bottom_8_data = filtered_data[filtered_data['nama_kabupaten_kota'].
↳isin(bottom_8_regions)]

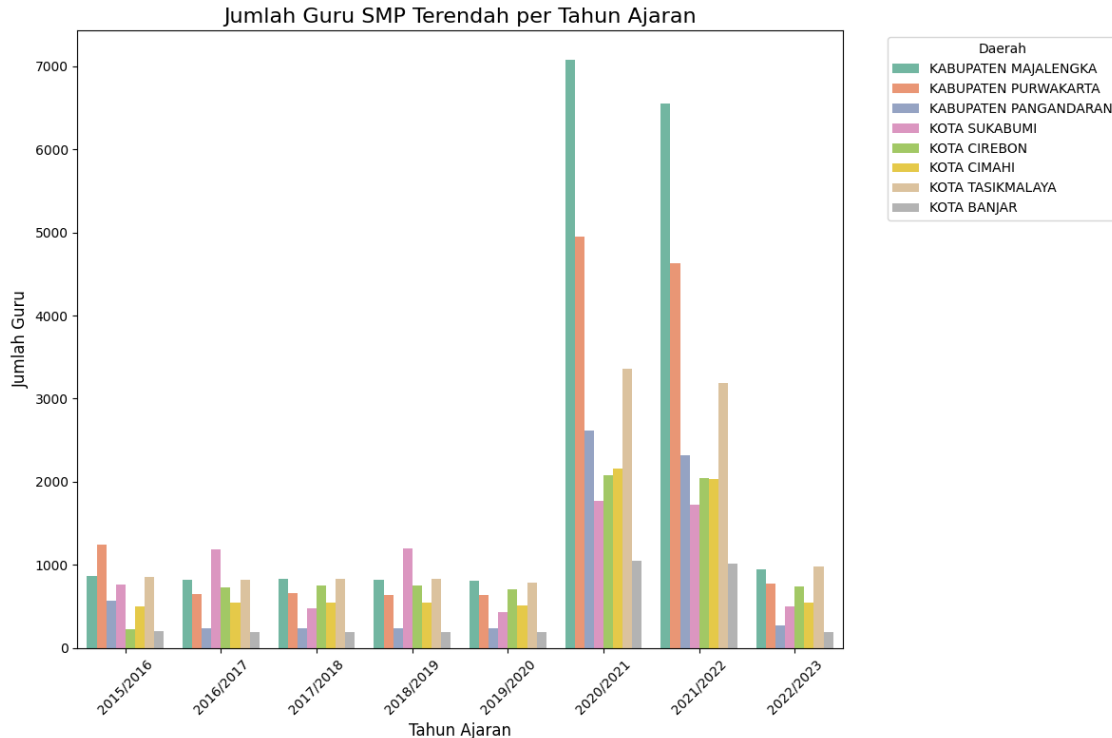
# Buat plot
plt.figure(figsize=(12,8))

# Menggunakan seaborn untuk membuat barplot dengan data dari 8 daerah terendah
↳untuk semua tahun ajaran
sns.barplot(x='tahun_ajaran', y='jumlah_guru', hue='nama_kabupaten_kota',
↳data=bottom_8_data, palette='Set2')

# Tambahkan judul dan label
plt.title('Jumlah Guru SMP Terendah per Tahun Ajaran', fontsize=16)
plt.xlabel('Tahun Ajaran', fontsize=12)
plt.ylabel('Jumlah Guru', fontsize=12)
plt.xticks(rotation=45)

# Tampilkan legenda di luar plot agar lebih jelas
plt.legend(title='Daerah', bbox_to_anchor=(1.05, 1), loc='upper left')

# Tampilkan diagram
plt.tight_layout()
plt.show()
```



Kesimpulan dari diagram di atas adalah sebagai berikut:

1. Kabupaten Majalengka menunjukkan jumlah guru yang stabil, tanpa lonjakan besar atau penurunan tajam setiap tahunnya, kecuali saat pandemi Covid-19.
2. Kabupaten Tasikmalaya menunjukkan pola yang serupa dengan Majalengka, mempertahankan stabilitas jumlah guru hingga tahun-tahun pandemi.
3. Purwakarta sempat mengalami penurunan jumlah guru, namun setelah itu jumlahnya cenderung stabil.
4. Kota Banjar memiliki jumlah guru paling sedikit dibandingkan dengan kota dan kabupaten lainnya.
5. Terjadi peningkatan signifikan dalam jumlah guru pada tahun-tahun awal pandemi Covid-19.

#### Jumlah Guru SMA di Jawa Barat 2015 - 2023

```
[81]: daerah_list = gsma['nama_kabupaten_kota'].drop_duplicates().tolist()

# Daftar tahun ajaran yang ingin digunakan
tahun_ajaran_list = ['2015/2016', '2016/2017', '2017/2018', '2018/2019',
                    '2019/2020', '2020/2021', '2021/2022', '2022/2023']

# Filter data untuk daerah dalam daerah_list dan tahun ajaran yang ditentukan
filtered_data = gsma[(gsma['nama_kabupaten_kota'].isin(daerah_list)) &
                    (gsma['tahun_ajaran'].isin(tahun_ajaran_list))]
```

```

# Buat plot
plt.figure(figsize=(12,8))

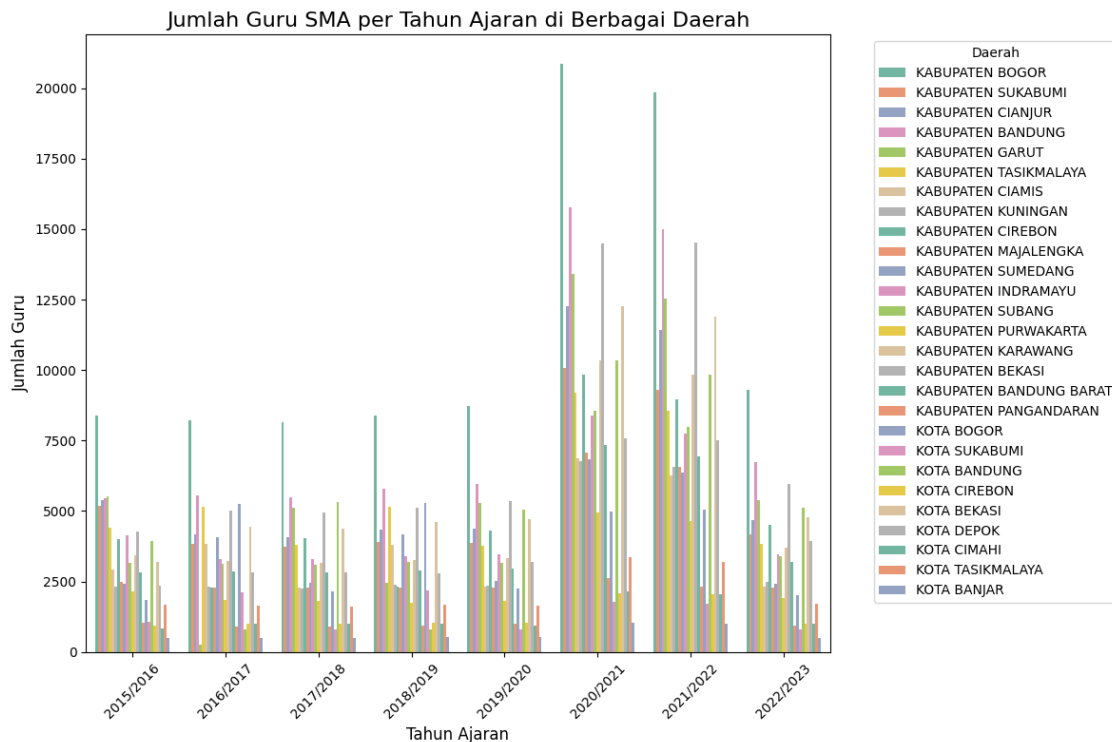
# Menggunakan seaborn untuk membuat barplot
sns.barplot(x='tahun_ajaran', y='jumlah_guru', hue='nama_kabupaten_kota',
            data=filtered_data, palette='Set2')

# Tambahkan judul dan label
plt.title('Jumlah Guru SMA per Tahun Ajaran di Berbagai Daerah', fontsize=16)
plt.xlabel('Tahun Ajaran', fontsize=12)
plt.ylabel('Jumlah Guru', fontsize=12)
plt.xticks(rotation=45)

# Tampilkan legenda di luar plot agar lebih jelas
plt.legend(title='Daerah', bbox_to_anchor=(1.05, 1), loc='upper left')

# Tampilkan diagram
plt.tight_layout()
plt.show()

```



Dari grafik yang menunjukkan jumlah guru SMA per tahun ajaran di berbagai daerah di Jawa Barat, kita dapat menarik beberapa kesimpulan dan analisis yang lebih rinci sebagai

berikut:

### Kenaikan dan Penurunan Jumlah Guru di Jawa Barat:

Secara umum, terdapat fluktuasi jumlah guru dari tahun ke tahun, dengan beberapa daerah menunjukkan pola kenaikan dan penurunan yang jelas. Kenaikan signifikan pada tahun ajaran 2020/2021 bertepatan dengan masa pandemi Covid-19, yang menunjukkan bahwa banyak daerah mungkin menambah jumlah guru untuk memenuhi kebutuhan pembelajaran daring.

### Variasi kenaikan dan penurunan Antar Daerah:

Ada perbedaan mencolok dalam jumlah guru antar daerah. Beberapa daerah seperti Kota Bandung dan Kabupaten Bogor mungkin menunjukkan jumlah guru yang lebih tinggi, sementara daerah lain, seperti Kota Banjar, memiliki jumlah yang lebih rendah.

Kenaikan jumlah guru di tahun ajaran 2020/2021 selama pandemi dapat diartikan sebagai upaya untuk meningkatkan kapasitas pendidikan dalam kondisi darurat, sementara penurunan setelah pandemi menunjukkan adaptasi kembali ke kondisi normal dengan jumlah guru yang lebih sesuai dengan kebutuhan.

```
[82]: # Daftar tahun ajaran yang ingin digunakan (seluruh tahun ajaran)
tahun_ajaran_list = ['2015/2016', '2016/2017', '2017/2018', '2018/2019',
                    '2019/2020', '2020/2021', '2021/2022', '2022/2023']

# Filter data untuk tahun ajaran yang ditentukan
filtered_data = gsma[gsma['tahun_ajaran'].isin(tahun_ajaran_list)]

# Ambil 8 kabupaten/kota dengan jumlah guru terbesar
top_daerah = filtered_data.groupby('nama_kabupaten_kota')['jumlah_guru'].sum().
    ↪nlargest(8).index

# Filter data berdasarkan 8 kabupaten/kota teratas
filtered_data_top = filtered_data[filtered_data['nama_kabupaten_kota'].
    ↪isin(top_daerah)]

# Buat plot
plt.figure(figsize=(12,8))

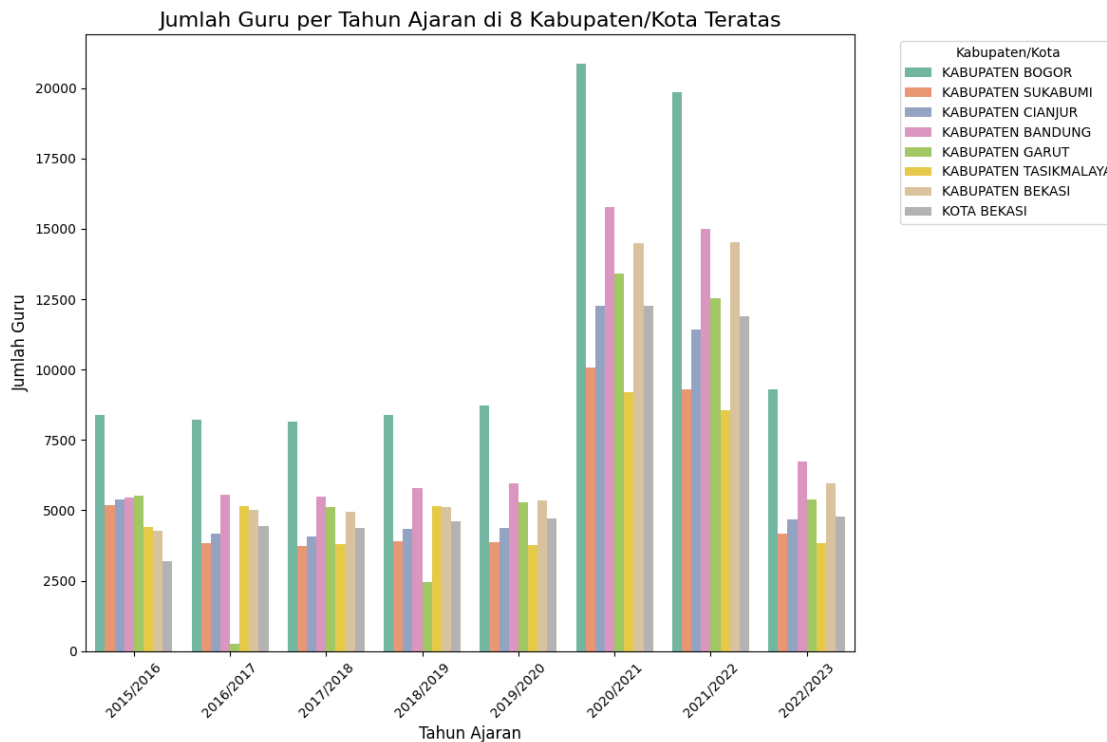
# Menggunakan seaborn untuk membuat barplot
sns.barplot(x='tahun_ajaran', y='jumlah_guru', hue='nama_kabupaten_kota',
    ↪data=filtered_data_top, palette='Set2')

# Tambahkan judul dan label
plt.title('Jumlah Guru per Tahun Ajaran di 8 Kabupaten/Kota Teratas',
    ↪fontsize=16)
plt.xlabel('Tahun Ajaran', fontsize=12)
plt.ylabel('Jumlah Guru', fontsize=12)
plt.xticks(rotation=45)

# Tampilkan legenda di luar plot agar lebih jelas
```

```
plt.legend(title='Kabupaten/Kota', bbox_to_anchor=(1.05, 1), loc='upper left')

# Tampilkan diagram
plt.tight_layout()
plt.show()
```



Berdasarkan analisis dari diagram di atas, berikut adalah beberapa kesimpulan yang dapat diambil:

Dominasi Kabupaten Bogor:

1. Kabupaten Bogor konsisten menjadi daerah dengan jumlah guru SMA tertinggi dari tahun ajaran 2015/2016 hingga 2022/2023. Hal ini menunjukkan keberhasilan daerah tersebut dalam menyediakan tenaga pengajar yang memadai, mungkin karena populasi yang tinggi dan fokus pada pengembangan pendidikan. Anomali di Kabupaten Garut:
2. Terjadi penurunan yang signifikan dan tidak biasa pada jumlah guru di Kabupaten Garut pada tahun ajaran 2016/2017. Dari jumlah lebih dari 5000 guru pada tahun sebelumnya, terjadi pengurangan yang sangat drastis, yang mungkin disebabkan oleh kebijakan pengurangan anggaran, perubahan dalam kebijakan pendidikan, atau faktor lain yang memengaruhi perekrutan guru. Stabilitas di Kabupaten Bandung:
3. Kabupaten Bandung menunjukkan jumlah guru yang relatif stabil sepanjang periode tersebut, kecuali pada tahun ajaran selama pandemi Covid-19. Kenaikan yang terjadi selama tahun ajaran 2020/2021 kemungkinan besar merupakan respons terhadap kebutuhan pendidikan yang mendesak akibat pembelajaran jarak jauh.

```
[83]: # Daftar tahun ajaran yang ingin digunakan (seluruh tahun ajaran)
tahun_ajaran_list = ['2015/2016', '2016/2017', '2017/2018', '2018/2019',
                    '2019/2020', '2020/2021', '2021/2022', '2022/2023']

# Filter data untuk tahun ajaran yang ditentukan
filtered_data = gsma[gsma['tahun_ajaran'].isin(tahun_ajaran_list)]

# Ambil 8 kabupaten/kota dengan jumlah guru terendah
bottom_daerah = filtered_data.groupby('nama_kabupaten_kota')['jumlah_guru'].
    ↪sum().nsmallest(8).index

# Filter data berdasarkan 8 kabupaten/kota terendah
filtered_data_bottom = filtered_data[filtered_data['nama_kabupaten_kota'].
    ↪isin(bottom_daerah)]

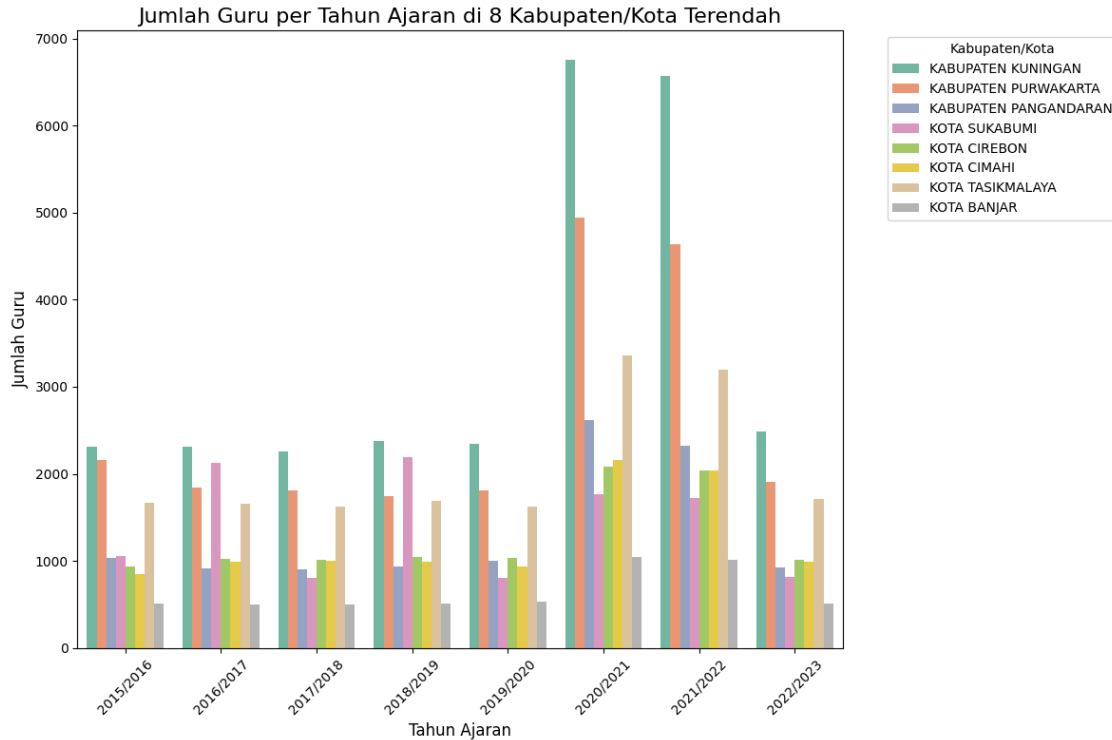
# Buat plot
plt.figure(figsize=(12,8))

# Menggunakan seaborn untuk membuat barplot
sns.barplot(x='tahun_ajaran', y='jumlah_guru', hue='nama_kabupaten_kota',
    ↪data=filtered_data_bottom, palette='Set2')

# Tambahkan judul dan label
plt.title('Jumlah Guru per Tahun Ajaran di 8 Kabupaten/Kota Terendah',
    ↪fontsize=16)
plt.xlabel('Tahun Ajaran', fontsize=12)
plt.ylabel('Jumlah Guru', fontsize=12)
plt.xticks(rotation=45)

# Tampilkan legenda di luar plot agar lebih jelas
plt.legend(title='Kabupaten/Kota', bbox_to_anchor=(1.05, 1), loc='upper left')

# Tampilkan diagram
plt.tight_layout()
plt.show()
```



**Kesimpulan Jumlah Guru SMA dan SMP** 1. Jumlah total guru di tingkat SMA di kabupaten/kota selama periode 2015/2016 hingga 2022/2023 menunjukkan stabilitas dengan fluktuasi yang relatif kecil. Ini menunjukkan bahwa sistem pendidikan setempat tetap mempertahankan jumlah tenaga pengajar yang konsisten selama beberapa tahun.

2. Terjadi peningkatan jumlah guru yang signifikan dari 2019/2020 ke 2020/2021. Kenaikan ini mungkin disebabkan oleh kebijakan pemerintah yang mendorong perekrutan lebih banyak guru, peningkatan alokasi dana untuk pendidikan, atau peningkatan kesadaran akan pentingnya kualitas pendidikan.
3. Tetapi, Terdapat penurunan jumlah guru pada tahun ajaran 2021/2022 dan 2022/2023 dibandingkan dengan tahun ajaran sebelumnya. Penurunan ini dapat menjadi indikator adanya masalah dalam mempertahankan tenaga pengajar, seperti pengunduran diri, pensiun, atau kendala dalam perekrutan guru baru.
4. Masing-masing kabupaten/kota menunjukkan perbedaan dalam jumlah guru, yang bisa diakibatkan oleh berbagai faktor, seperti kepadatan penduduk, tingkat pendidikan masyarakat, dan anggaran daerah untuk pendidikan. Beberapa kabupaten/kota mungkin memiliki lebih banyak sekolah atau jumlah siswa yang lebih tinggi, sehingga memerlukan lebih banyak guru.
5. Periode tahun ajaran 2020/2021 mungkin juga terpengaruh oleh situasi pandemi COVID-19. Meskipun jumlah guru meningkat, ada kemungkinan bahwa pembelajaran jarak jauh telah memengaruhi cara guru berinteraksi dengan siswa dan penyampaian materi pelajaran.
6. Kabupaten Bogor selalu menjadi urutan pertama di Jawa Barat melihat jumlah sekolah (negeri dan swasta) terbanyak juga berada di Bogor



Menjawab Pertanyaan N0. 2

Kabupaten/kota dengan indeks pendidikan lebih tinggi, seperti Kota Cimahi dan Kota Depok, umumnya memiliki jumlah sekolah dan guru yang memadai. Ketersediaan ini mendukung akses pendidikan yang lebih baik, sehingga berdampak positif pada indeks pendidikan. Contohnya, Kabupaten Bogor dengan jumlah guru dan sekolah yang tinggi memiliki indeks pendidikan yang cukup baik, meskipun tidak tertinggi. Jumlah sekolah dan guru yang banyak meningkatkan akses, tetapi kualitas pendidikan tetap bergantung pada kualitas pengajaran dan manajemen sekolah, bukan hanya pada jumlah fasilitas dan tenaga pengajar.

#### 0.1.4 Analisis dan Visualisasi dari dataset Harapan Lama Sekolah

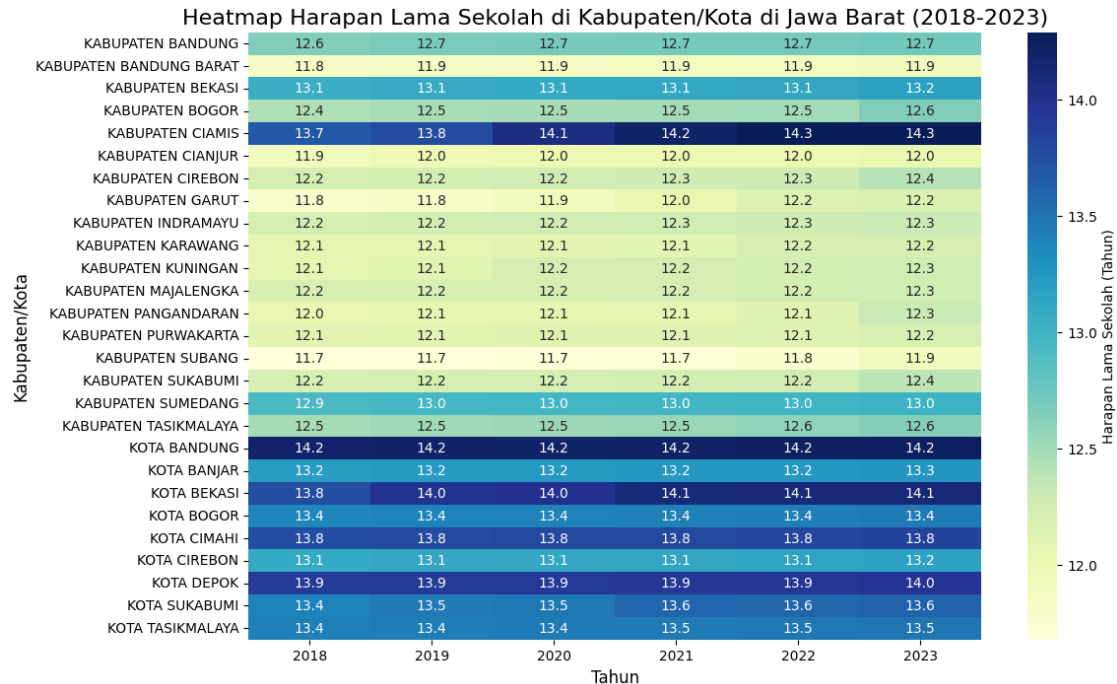
**Visualisasi Heatmap record data dari Harapan Lama Sekolah Per Kabupaten/Kota di Jawa Barat (2018-2023)**

```
[84]: # Mengonversi kolom 'harapan_lama_sekolah' menjadi tipe float
hls['harapan_lama_sekolah'] = hls['harapan_lama_sekolah'].replace('-', np.nan).
    ↪astype(float)

# Memfilter data untuk tahun 2018-2023
hls_filtered = hls[(hls['tahun'] >= 2018) & (hls['tahun'] <= 2023)]

# Pivot DataFrame untuk heatmap menggunakan pivot_table
heatmap_data = hls_filtered.pivot_table(index="nama_kabupaten_kota",
                                         columns="tahun",
                                         values="harapan_lama_sekolah",
                                         aggfunc='mean') # Menggunakan 'mean'
    ↪jika ada duplikasi

# Membuat heatmap
plt.figure(figsize=(12, 8))
sns.heatmap(heatmap_data, cmap='YlGnBu', annot=True, fmt='.1f',
    ↪cbar_kws={'label': 'Harapan Lama Sekolah (Tahun)'})
plt.title('Heatmap Harapan Lama Sekolah di Kabupaten/Kota di Jawa Barat
    ↪(2018-2023)', fontsize=16)
plt.xlabel('Tahun', fontsize=12)
plt.ylabel('Kabupaten/Kota', fontsize=12)
plt.show()
```



## Record data Top 10 Teratas Harapan Lama Sekolah (2018-2023)

```
[85]: # Memfilter data untuk tahun 2018 hingga 2023
hls_filtered = hls[(hls['tahun'] >= 2018) & (hls['tahun'] <= 2023)]

# Menghitung rata-rata harapan lama sekolah per kabupaten/kota
average_hls = hls_filtered.
    ↳groupby('nama_kabupaten_kota')['harapan_lama_sekolah'].mean().reset_index()

# Mengurutkan kabupaten/kota berdasarkan rata-rata harapan lama sekolah secara
    ↳menurun
top_hls = average_hls.sort_values(by='harapan_lama_sekolah', ascending=False).
    ↳head(10)

# Menampilkan hasil
print("10 Kab/Kota dengan Harapan Lama Sekolah Tertinggi (2018-2023):")
print(top_hls)
```

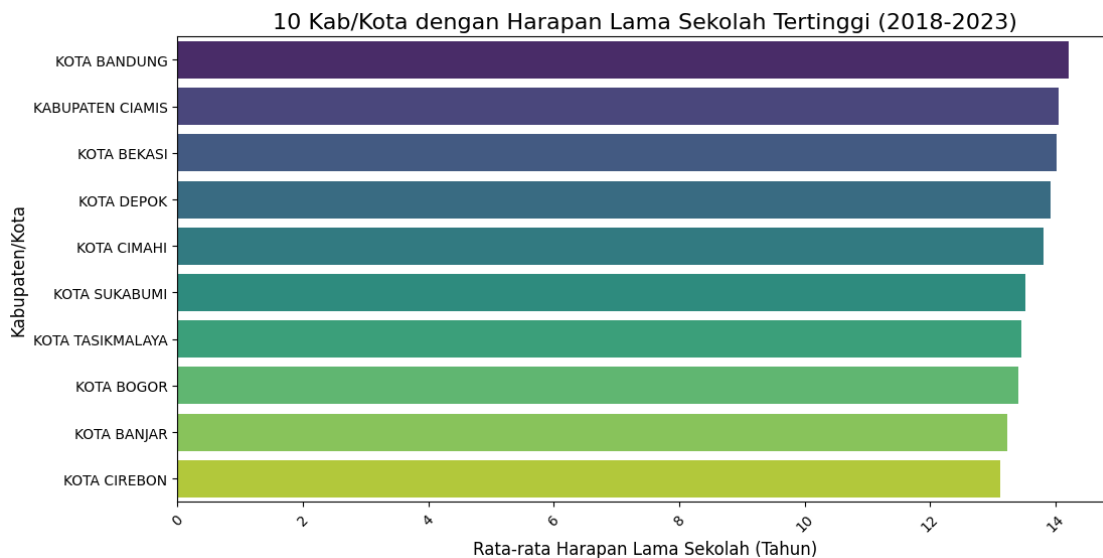
10 Kab/Kota dengan Harapan Lama Sekolah Tertinggi (2018-2023):

	nama_kabupaten_kota	harapan_lama_sekolah
18	KOTA BANDUNG	14.208333
4	KABUPATEN CIAMIS	14.048333
20	KOTA BEKASI	14.013333
24	KOTA DEPOK	13.926667
22	KOTA CIMAHI	13.805000

25	KOTA SUKABUMI	13.516667
26	KOTA TASIKMALAYA	13.455000
21	KOTA BOGOR	13.415000
19	KOTA BANJAR	13.235000
23	KOTA CIREBON	13.125000

### Visualisasi Top 10 Teratas

```
[86]: # Membuat visualisasi grafik batang
plt.figure(figsize=(12, 6))
sns.barplot(x='harapan_lama_sekolah', y='nama_kabupaten_kota', data=top_hls,
           palette='viridis')
plt.title('10 Kab/Kota dengan Harapan Lama Sekolah Tertinggi (2018-2023)',
          fontsize=16)
plt.xlabel('Rata-rata Harapan Lama Sekolah (Tahun)', fontsize=12)
plt.ylabel('Kabupaten/Kota', fontsize=12)
plt.xticks(rotation=45)
plt.show()
```



### Record data Top 10 Terendah Harapan Lama Sekolah (2018-2023)

```
[87]: # Memfilter data untuk tahun 2018 hingga 2023
hls_filtered = hls[(hls['tahun'] >= 2018) & (hls['tahun'] <= 2023)]

# Menghitung rata-rata harapan lama sekolah per kabupaten/kota
average_hls = hls_filtered.
               groupby('nama_kabupaten_kota')['harapan_lama_sekolah'].mean().reset_index()
```

```
# Mengurutkan kabupaten/kota berdasarkan rata-rata harapan lama sekolah secara
↳meningkat
lowest_hls = average_hls.sort_values(by='harapan_lama_sekolah', ascending=True).
↳head(10)

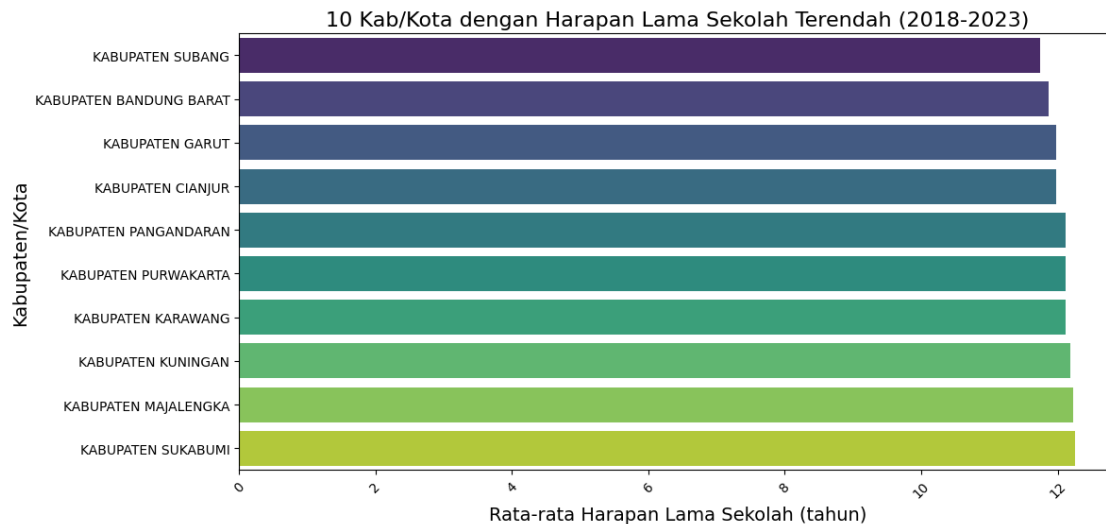
# Menampilkan hasil
print("10 Kab/Kota dengan Harapan Lama Sekolah Terendah (2018-2023):")
print(lowest_hls)
```

10 Kab/Kota dengan Harapan Lama Sekolah Terendah (2018-2023):

	nama_kabupaten_kota	harapan_lama_sekolah
14	KABUPATEN SUBANG	11.745000
1	KABUPATEN BANDUNG BARAT	11.873333
7	KABUPATEN GARUT	11.978333
5	KABUPATEN CIANJUR	11.985000
12	KABUPATEN PANGANDARAN	12.110000
13	KABUPATEN PURWAKARTA	12.121667
9	KABUPATEN KARAWANG	12.121667
10	KABUPATEN KUNINGAN	12.186667
11	KABUPATEN MAJALENGKA	12.225000
15	KABUPATEN SUKABUMI	12.253333

### Visualisasi Top 10 Terendah

```
[88]: # Visualisasi dengan grafik batang
plt.figure(figsize=(12, 6))
sns.barplot(x='harapan_lama_sekolah', y='nama_kabupaten_kota', data=lowest_hls,
↳palette='viridis')
plt.title('10 Kab/Kota dengan Harapan Lama Sekolah Terendah (2018-2023)',
↳fontsize=16)
plt.xlabel('Rata-rata Harapan Lama Sekolah (tahun)', fontsize=14)
plt.ylabel('Kabupaten/Kota', fontsize=14)
plt.xticks(rotation=45)
plt.show()
```



**Kesimpulan** - Kota Bandung menempati posisi teratas dengan harapan lama sekolah rata-rata sebesar 14.21 tahun, diikuti oleh Kabupaten Ciamis dan Kota Bekasi. Hal ini menunjukkan bahwa kota-kota tersebut memiliki sistem pendidikan yang relatif lebih baik dan akses yang lebih besar terhadap pendidikan bagi penduduknya. - Kabupaten Subang tercatat sebagai daerah dengan harapan lama sekolah terendah, yakni 11.75 tahun. Kabupaten Bandung Barat dan Garut juga menunjukkan angka harapan lama sekolah yang rendah, yang menunjukkan bahwa masih terdapat tantangan dalam meningkatkan akses dan kualitas pendidikan di daerah-daerah tersebut. - Terdapat kesenjangan yang signifikan antara daerah dengan harapan lama sekolah tertinggi dan terendah, yaitu sekitar 2.5 tahun. Hal ini mengindikasikan perlunya perhatian lebih dari pemerintah dan pemangku kepentingan untuk meningkatkan kualitas pendidikan, khususnya di kabupaten/kota yang memiliki harapan lama sekolah yang rendah, demi mencapai pemerataan pendidikan di seluruh wilayah Jawa Barat.

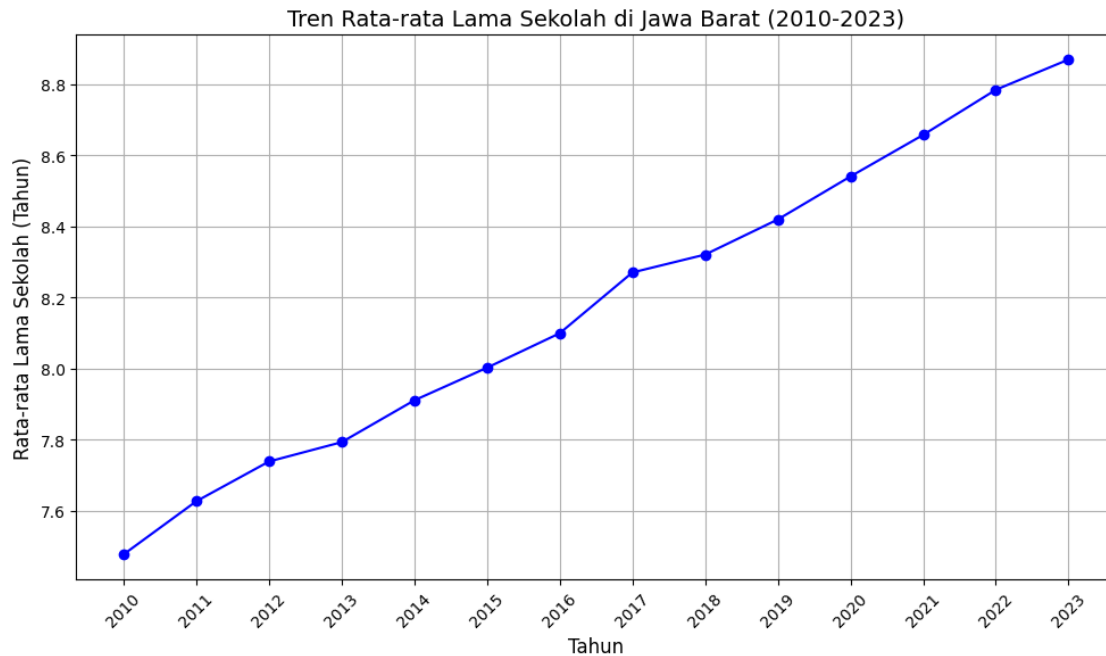
### 0.1.5 Analisis dan Visualisai Data Rata Rata Lama Sekolah di Provinsi Jawa Barat

#### Tren rata-rata lama sekolah per tahun

```
[89]: # Menghitung rata-rata lama sekolah per tahun
rata_rata_sekolah_per_tahun = rrl.groupby('tahun')['rata_rata_lama_sekolah'].
    ↪mean()

# Membuat plot garis untuk tren rata-rata lama sekolah dari tahun 2010 hingga
    ↪2023
plt.figure(figsize=(10, 6))
plt.plot(rata_rata_sekolah_per_tahun.index, rata_rata_sekolah_per_tahun.values,
    ↪marker='o', linestyle='-', color='b')
plt.title('Tren Rata-rata Lama Sekolah di Jawa Barat (2010-2023)', fontsize=14)
plt.xlabel('Tahun', fontsize=12)
plt.ylabel('Rata-rata Lama Sekolah (Tahun)', fontsize=12)
plt.xticks(rata_rata_sekolah_per_tahun.index, rotation=45)
```

```
plt.grid(True)
plt.tight_layout()
plt.show()
```



Plot diagram di atas adalah line chart yang menunjukkan tren rata-rata lama sekolah di Jawa Barat dari tahun 2010 hingga 2023.

- Sumbu y menunjukkan rata-rata lama sekolah dalam satuan tahun.
- Sumbu x menunjukkan tahun, dari 2010 hingga 2023.

Terlihat adanya peningkatan yang stabil dalam rata-rata lama sekolah dari tahun 2010 hingga 2023. Ini menunjukkan bahwa, secara rata-rata, masyarakat di Jawa Barat menghabiskan lebih banyak waktu di bangku sekolah seiring berjalannya waktu.

### Rata-rata Lama Sekolah per Kabupaten/Kota Di Jawa Barat

```
[90]: # Menghitung rata-rata lama sekolah per kabupaten/kota
rrl_grouped = rrl.groupby('nama_kabupaten_kota')['rata_rata_lama_sekolah'].
    ↪mean().reset_index()
print(rrl_grouped)
```

	nama_kabupaten_kota	rata_rata_lama_sekolah
0	KABUPATEN BANDUNG	8.557143
1	KABUPATEN BANDUNG BARAT	7.750714
2	KABUPATEN BEKASI	8.710714
3	KABUPATEN BOGOR	7.795714
4	KABUPATEN CIAMIS	7.534286

5	KABUPATEN CIANJUR	6.758571
6	KABUPATEN CIREBON	6.574286
7	KABUPATEN GARUT	7.178571
8	KABUPATEN INDRAMAYU	5.802857
9	KABUPATEN KARAWANG	7.135000
10	KABUPATEN KUNINGAN	7.297857
11	KABUPATEN MAJALENGKA	6.932143
12	KABUPATEN PANGANDARAN	7.497273
13	KABUPATEN PURWAKARTA	7.572143
14	KABUPATEN SUBANG	6.654286
15	KABUPATEN SUKABUMI	6.668571
16	KABUPATEN SUMEDANG	7.998571
17	KABUPATEN TASIKMALAYA	7.087857
18	KOTA BANDUNG	10.624286
19	KOTA BANJAR	8.214286
20	KOTA BEKASI	10.870000
21	KOTA BOGOR	10.126429
22	KOTA CIMAHI	10.837143
23	KOTA CIREBON	9.757857
24	KOTA DEPOK	10.825714
25	KOTA SUKABUMI	9.187143
26	KOTA TASIKMALAYA	8.865000

```
[91]: # Menghitung rata-rata lama sekolah per kabupaten/kota
rrl_grouped = rrl.groupby('nama_kabupaten_kota')['rata_rata_lama_sekolah'].
    ↪mean().reset_index()

# Plot setelah data diakumulasi atau dirata-ratakan
plt.figure(figsize=(12, 8))
sns.barplot(x='nama_kabupaten_kota', y='rata_rata_lama_sekolah',
    ↪data=rrl_grouped)
plt.title('Rata-rata Lama Sekolah per Kabupaten/Kota Di Jawa Barat (2010 -
    ↪2023)', fontsize=14)
plt.xticks(rotation=90)
plt.xlabel('Kabupaten/Kota', fontsize=12)
plt.ylabel('Rata-rata Lama Sekolah (Tahun)', fontsize=12)
plt.tight_layout()
plt.show()
```

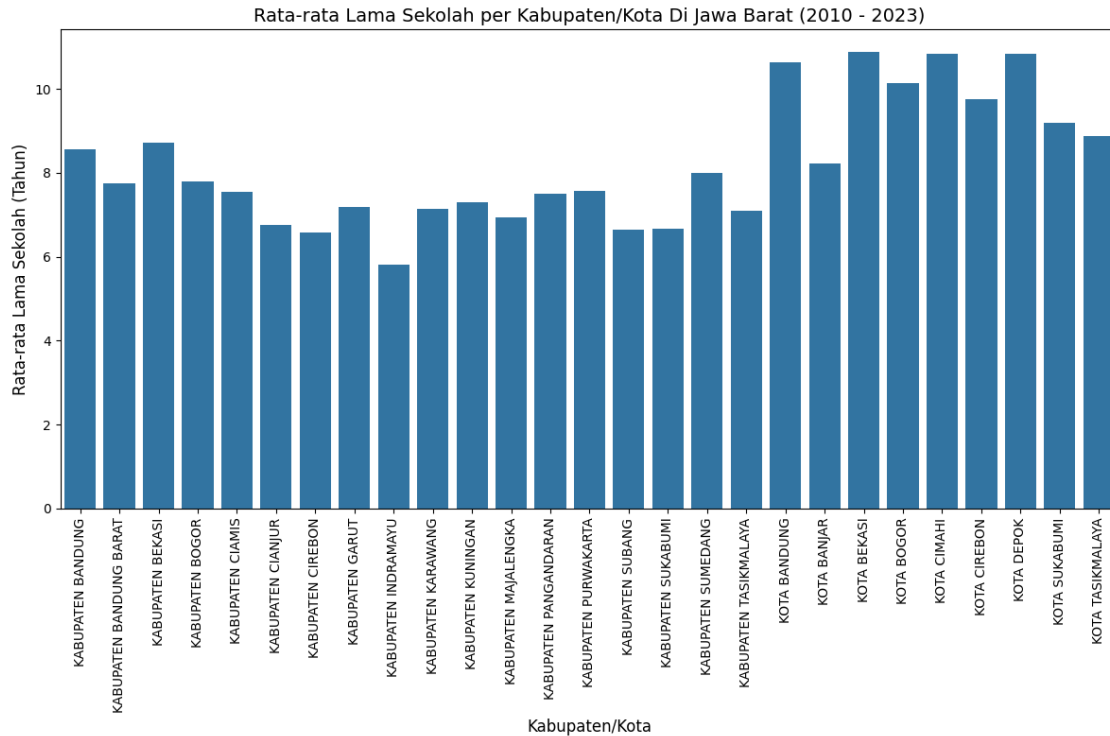


Diagram di atas adalah bar chart yang menunjukkan rata-rata lama sekolah per kabupaten/kota di Jawa Barat dari tahun 2010 hingga 2023.

- Sumbu y menunjukkan rata-rata lama sekolah dalam satuan tahun.
- Sumbu x menunjukkan berbagai kabupaten/kota di Jawa Barat.

### Top 10 Kabupaten/Kota dengan Rata-rata Lama Sekolah Terendah

```
[92]: # Menghitung rata-rata lama sekolah per kabupaten/kota per tahun
rata_rata_per_kabupaten = rrl.groupby(['nama_kabupaten_kota',
    ↳ 'tahun'])['rata_rata_lama_sekolah'].mean().reset_index()

# Menghitung rata-rata lama sekolah total per kabupaten/kota
rata_rata_total = rata_rata_per_kabupaten.
    ↳ groupby('nama_kabupaten_kota')['rata_rata_lama_sekolah'].mean().reset_index()

# Mengambil 10 kabupaten/kota dengan rata-rata lama sekolah terendah
top_10_kabupaten = rata_rata_total.nsmallest(10, 'rata_rata_lama_sekolah')

plt.figure(figsize=(10, 6))
sns.barplot(data=top_10_kabupaten, x='rata_rata_lama_sekolah',
    ↳ y='nama_kabupaten_kota')
plt.title('Top 10 Kabupaten/Kota dengan Rata-rata Lama Sekolah Terendah (2010 -
    ↳ 2023)')
```



```
plt.xlabel('Rata-rata Lama Sekolah (Tahun)')
plt.ylabel('Kabupaten/Kota')
plt.show()
```

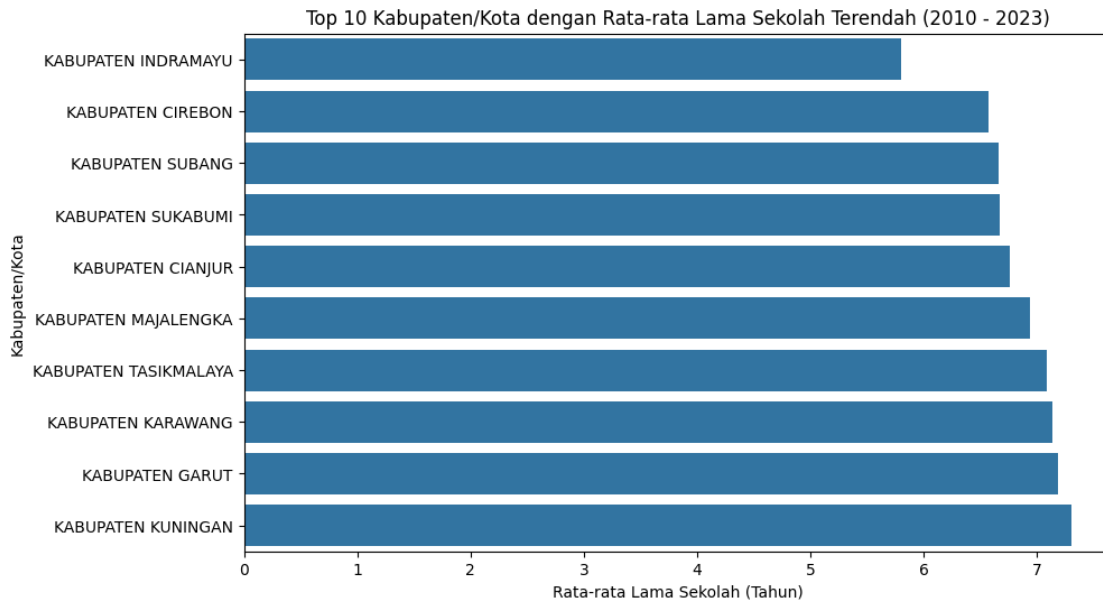


Diagram di atas adalah horizontal bar chart yang menunjukkan 10 kabupaten/kota dengan rata-rata lama sekolah terendah di Jawa Barat dari tahun 2010 hingga 2023.

- Sumbu x menunjukkan rata-rata lama sekolah dalam satuan tahun.
- Sumbu y menunjukkan berbagai kabupaten/kota di Jawa Barat.

### Top 10 Kabupaten/Kota dengan Rata-rata Lama Sekolah Tertinggi

```
[93]: # Menghitung rata-rata lama sekolah per kabupaten/kota per tahun
rata_rata_per_kabupaten = rrl.groupby(['nama_kabupaten_kota', 'tahun'])['rata_rata_lama_sekolah'].mean().reset_index()

# Menghitung rata-rata lama sekolah total per kabupaten/kota
rata_rata_total = rata_rata_per_kabupaten.groupby('nama_kabupaten_kota')['rata_rata_lama_sekolah'].mean().reset_index()

# Mengambil 10 kabupaten/kota dengan rata-rata lama sekolah tertinggi
top_10_kabupaten = rata_rata_total.nlargest(10, 'rata_rata_lama_sekolah')

# Membuat diagram
plt.figure(figsize=(10, 6))
sns.barplot(data=top_10_kabupaten, x='rata_rata_lama_sekolah', y='nama_kabupaten_kota')
```

```
plt.title('Top 10 Kabupaten/Kota dengan Rata-rata Lama Sekolah Tertinggi (2010_
↪- 2023)')
plt.xlabel('Rata-rata Lama Sekolah (Tahun)')
plt.ylabel('Kabupaten/Kota')
plt.show()
```

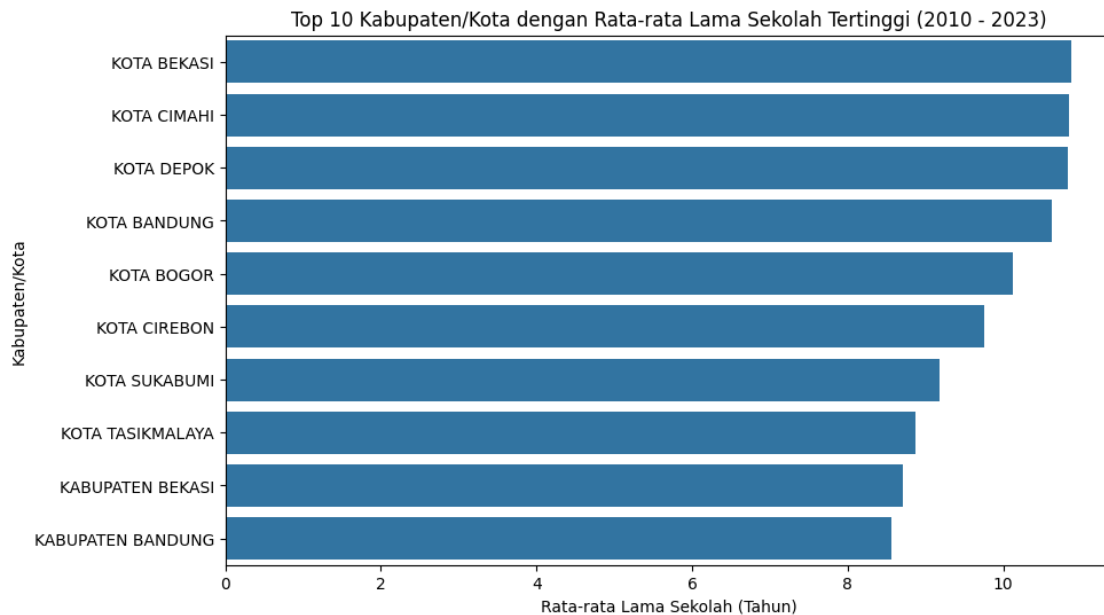


Diagram di atas adalah horizontal bar chart yang menunjukkan 10 kabupaten/kota dengan rata-rata lama sekolah tertinggi di Jawa Barat dari tahun 2010 hingga 2023.

- Sumbu x menunjukkan rata-rata lama sekolah dalam satuan tahun.
- Sumbu y menunjukkan berbagai kabupaten/kota di Jawa Barat.

### Kesimpulan

- Peningkatan umum: Secara keseluruhan, rata-rata lama sekolah di semua kabupaten/kota di Jawa Barat mengalami peningkatan dari tahun 2010 hingga 2023.
- Kesenjangan daerah: Terdapat kesenjangan yang cukup signifikan antara daerah perkotaan dan pedesaan. Kota-kota besar seperti Bandung, Bekasi, dan Depok memiliki rata-rata lama sekolah yang lebih tinggi dibandingkan dengan kabupaten-kabupaten.
- Target pendidikan: Masih ada pekerjaan yang harus dilakukan untuk mencapai rata-rata lama sekolah 12 tahun (setara dengan tamat SMA) di seluruh Jawa Barat, terutama di daerah-daerah yang masih di bawah 8 tahun.

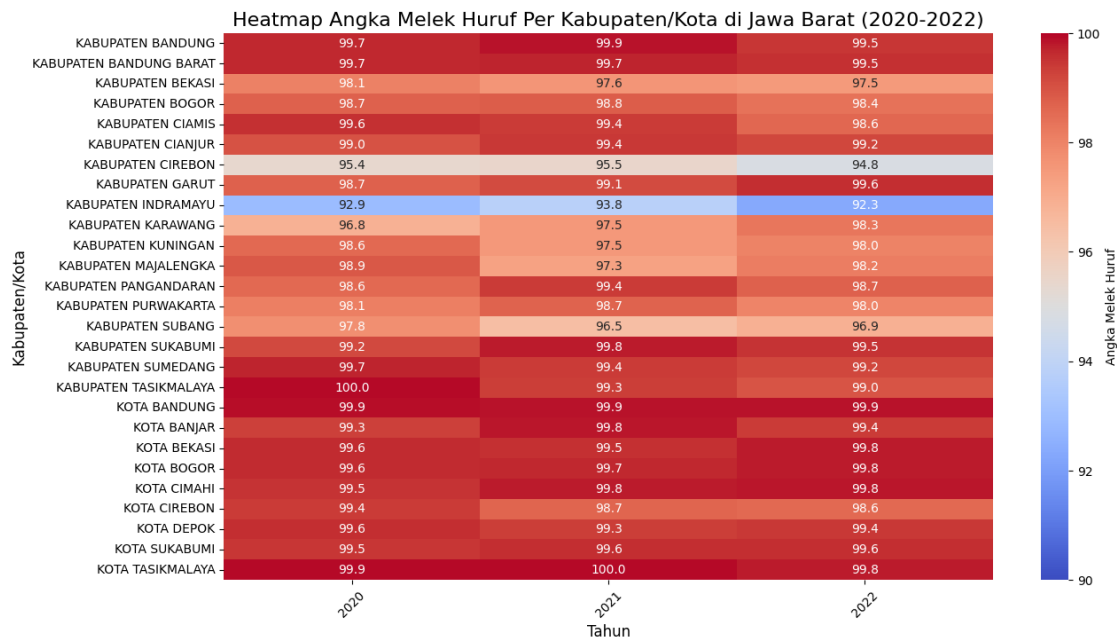
Hal ini menunjukkan bahwa meskipun ada kemajuan dalam pendidikan di Jawa Barat, masih diperlukan upaya lebih lanjut untuk mengurangi kesenjangan dan meningkatkan rata-rata lama sekolah di seluruh wilayah.

### 0.1.6 Analisis dan Visualisasi dari dataset Angka Melek Huruf Penduduk Usia 15 Tahun ke Atas

Visualisasi Heatmap record data dari Angka Melek Huruf Per Kabupaten/Kota di Jawa Barat (2020-2022)

```
[94]: # Membuat pivot table hanya untuk tahun 2020, 2021, dan 2022
tahun_filter = [2020, 2021, 2022]
pivot_table_filtered = amh.pivot_table(values='angka_melek_huruf',
    ↪index='nama_kabupaten_kota', columns='tahun')[tahun_filter]

# Membuat heatmap dengan rentang nilai angka melek huruf 90-100
plt.figure(figsize=(14, 8))
sns.heatmap(pivot_table_filtered, annot=True, fmt=".1f", cmap='coolwarm',
    ↪vmin=90, vmax=100, cbar_kws={'label': 'Angka Melek Huruf'})
plt.title('Heatmap Angka Melek Huruf Per Kabupaten/Kota di Jawa Barat
    ↪(2020-2022)', fontsize=16)
plt.xlabel('Tahun', fontsize=12)
plt.ylabel('Kabupaten/Kota', fontsize=12)
plt.xticks(rotation=45)
plt.show()
```



Plot diagram di atas adalah heatmap yang menunjukkan Angka Melek Huruf (AMH) per kabupaten/kota di Jawa Barat pada rentang waktu tahun 2020 hingga 2022

- Sumbu y menampilkan nama-nama kabupaten/kota di Jawa Barat
- Sumbu x menunjukkan rentang waktu atau timeline berdasarkan tahun (2020-2022)
- Warna dalam heatmap menunjukkan tingkat Angka Melek Huruf, yaitu warna merah tua

menunjukkan tingkat melek huruf yang lebih tinggi dan warna biru menunjukkan tingkat yang lebih rendah

Analisis Singkat:

Dari heatmap yang telah dibuat, terlihat bahwa sebagian besar kabupaten/kota di Jawa Barat memiliki Angka Melek Huruf yang cukup tinggi (di atas 95%), dengan beberapa kabupaten seperti Kota Bandung dan Kabupaten Tasikmalaya yang mencapai hampir 100% pada setiap tahun. Namun, ada beberapa kabupaten, seperti Kabupaten Indramayu dan Kabupaten Cirebon, yang cenderung memiliki tingkat AMH lebih rendah dibandingkan daerah lainnya.

Perbedaan ini mengindikasikan adanya faktor-faktor yang mempengaruhi literasi di tiap kabupaten. Analisis lebih lanjut diperlukan untuk memahami penyebab tingkat literasi rendah di beberapa wilayah, sehingga upaya peningkatan dapat dilakukan secara merata di seluruh Jawa Barat.

### Record data 10 Top Teratas dari Angka Melek Huruf Per Kabupaten/Kota di Jawa Barat (2020-2022)

```
[95]: # Membuat pivot table hanya untuk tahun 2020, 2021, dan 2022
tahun_filter = [2020, 2021, 2022]
pivot_table_filtered = amh.pivot_table(values='angka_melek_huruf',
    ↪index='nama_kabupaten_kota', columns='tahun')[tahun_filter]

# Menghitung rata-rata angka melek huruf untuk setiap kabupaten/kota
rata_rata = pivot_table_filtered.mean(axis=1)

# Mengurutkan kabupaten/kota berdasarkan rata-rata tertinggi
rata_rata_sorted = rata_rata.sort_values(ascending=False)

# Mengambil 10 kabupaten/kota teratas
top_10_kabupaten_kota = rata_rata_sorted.head(10)

# Menampilkan hasil
print("Top 10 Kabupaten/Kota dengan Rata-rata Angka Melek Huruf Tertinggi:")
print(top_10_kabupaten_kota)
```

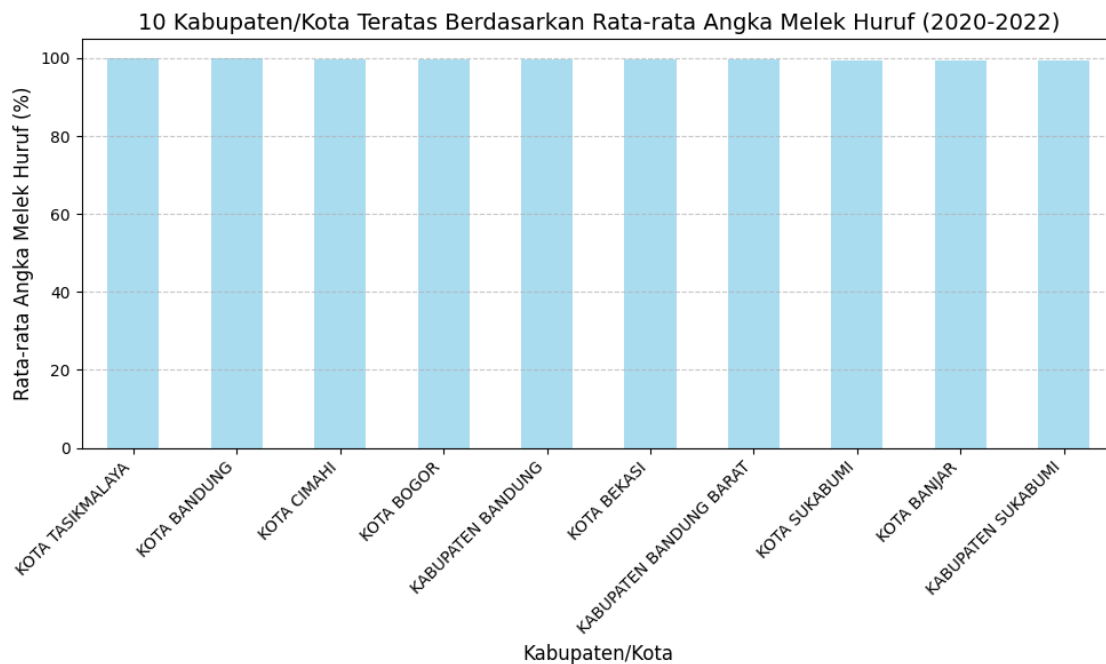
Top 10 Kabupaten/Kota dengan Rata-rata Angka Melek Huruf Tertinggi:

nama_kabupaten_kota	
KOTA TASIKMALAYA	99.893333
KOTA BANDUNG	99.890000
KOTA CIMAHI	99.716667
KOTA BOGOR	99.696667
KABUPATEN BANDUNG	99.676667
KOTA BEKASI	99.650000
KABUPATEN BANDUNG BARAT	99.633333
KOTA SUKABUMI	99.543333
KOTA BANJAR	99.510000
KABUPATEN SUKABUMI	99.496667

dtype: float64

## Visualisasi Top 10 Teratas

```
[96]: # Membuat grafik batang
plt.figure(figsize=(10, 6))
top_10_kabupaten_kota.plot(kind='bar', color='skyblue', alpha=0.7)
plt.title('10 Kabupaten/Kota Teratas Berdasarkan Rata-rata Angka Melek Huruf_
↪(2020-2022)', fontsize=14)
plt.xlabel('Kabupaten/Kota', fontsize=12)
plt.ylabel('Rata-rata Angka Melek Huruf (%)', fontsize=12)
plt.xticks(rotation=45, ha='right')
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```



Plot Diagram di atas adalah bar chart atau diagram batang yang menunjukkan rata-rata Angka Melek Huruf (AMH) untuk 10 kabupaten/kota dengan nilai tertinggi di Jawa Barat dari tahun 2020 hingga 2022.

- Sumbu y menunjukkan rata-rata Angka Melek Huruf dalam persentase (%)
- Sumbu x menunjukkan nama kabupaten/kota dengan nilai AMH tertinggi

## Record data Top 10 Terendah dari Angka Melek Huruf Per Kabupaten/Kota di Jawa Barat (2020-2022)

```
[97]: # Membuat pivot table hanya untuk tahun 2020, 2021, dan 2022
tahun_filter = [2020, 2021, 2022]
```

```

pivot_table_filtered = amh.pivot_table(values='angka_melek_huruf',
    ↪index='nama_kabupaten_kota', columns='tahun')[tahun_filter]

# Menghitung rata-rata angka melek huruf untuk setiap kabupaten/kota
rata_rata = pivot_table_filtered.mean(axis=1)

# Mengurutkan kabupaten/kota berdasarkan rata-rata terendah
rata_rata_sorted = rata_rata.sort_values(ascending=True)

# Mengambil 10 kabupaten/kota terendah
top_10_kabupaten_kota_terendah = rata_rata_sorted.head(10)

# Menampilkan hasil
print("Top 10 Kabupaten/Kota dengan Rata-rata Angka Melek Huruf Terendah:")
print(top_10_kabupaten_kota_terendah)

```

Top 10 Kabupaten/Kota dengan Rata-rata Angka Melek Huruf Terendah:

nama_kabupaten_kota	
KABUPATEN INDRAMAYU	93.006667
KABUPATEN CIREBON	95.243333
KABUPATEN SUBANG	97.033333
KABUPATEN KARAWANG	97.550000
KABUPATEN BEKASI	97.720000
KABUPATEN KUNINGAN	98.030000
KABUPATEN MAJALENGKA	98.116667
KABUPATEN PURWAKARTA	98.253333
KABUPATEN BOGOR	98.636667
KOTA CIREBON	98.873333

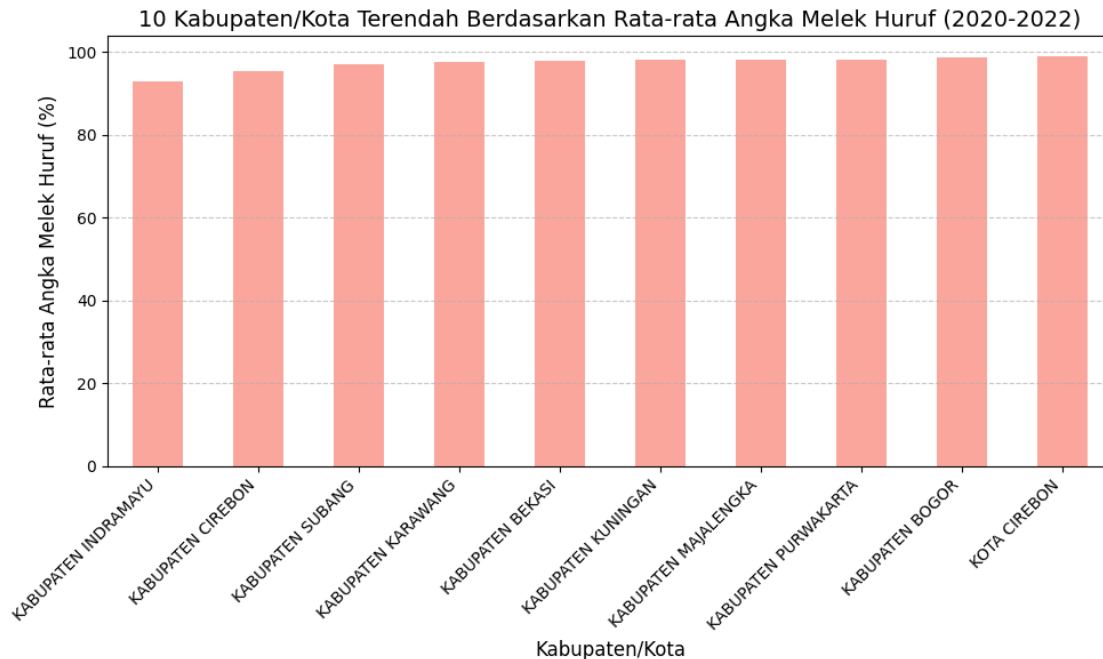
dtype: float64

### Visualisasi Top 10 Terendah

```

[98]: # Membuat grafik batang
plt.figure(figsize=(10, 6))
top_10_kabupaten_kota_terendah.plot(kind='bar', color='salmon', alpha=0.7)
plt.title('10 Kabupaten/Kota Terendah Berdasarkan Rata-rata Angka Melek Huruf
    ↪(2020-2022)', fontsize=14)
plt.xlabel('Kabupaten/Kota', fontsize=12)
plt.ylabel('Rata-rata Angka Melek Huruf (%)', fontsize=12)
plt.xticks(rotation=45, ha='right')
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()

```



Plot Diagram di atas adalah bar chart atau diagram batang yang menunjukkan rata-rata Angka Melek Huruf (AMH) untuk 10 kabupaten/kota dengan nilai terendah di Jawa Barat dari tahun 2020 hingga 2022.

- Sumbu y menunjukkan rata-rata Angka Melek Huruf dalam persentase (%)
- Sumbu x menunjukkan nama kabupaten/kota dengan nilai AMH terendah

### Kesimpulan

- Kota-kota besar seperti Kota Tasikmalaya, Kota Bandung, dan Kota Cimahi memiliki Angka Melek Huruf yang sangat tinggi, mendekati 100%. Ini menunjukkan bahwa pendidikan di area perkotaan cenderung lebih baik dibandingkan dengan daerah-daerah lain.
- Di sisi lain, Kabupaten Indramayu mencatat angka melek huruf terendah di antara kabupaten/kota lain dengan nilai sekitar 93.01%. Hal ini menunjukkan adanya kesenjangan pendidikan yang signifikan antara daerah perkotaan dan pedesaan, serta di antara kabupaten/kota di Jawa Barat.
- Temuan ini dapat memberikan panduan bagi pembuat kebijakan untuk mengembangkan strategi yang lebih efektif dalam meningkatkan literasi di daerah-daerah dengan angka melek huruf yang rendah. Upaya yang bisa dilakukan antara lain penyediaan sumber daya pendidikan, peningkatan kualitas guru, dan program-program literasi masyarakat

#### 0.1.7 Analisis dan Visualisasi dari dataset Tingkat Pengangguran Terbuka

**Visualisasi Heatmap record data dari Tingkat Pengangguran Terbuka Per Kabupaten/Kota di Jawa Barat (2018-2023)**

```
[99]: # Memfilter data untuk tahun 2018 hingga 2023
jpt_filtered = jpt[(jpt['tahun'] >= 2018) & (jpt['tahun'] <= 2023)]

# Mengatur ukuran heatmap
plt.figure(figsize=(14, 8))

# Pivot DataFrame untuk heatmap
heatmap_data = jpt_filtered.pivot(index="nama_kabupaten_kota", columns="tahun",
    ↪ values="tingkat_pengangguran_terbuka")

# Membuat heatmap
sns.heatmap(heatmap_data, annot=True, fmt=".1f", cmap="YlGnBu", linewidths=.5,
    ↪ cbar_kws={"label": "Tingkat Pengangguran Terbuka (%)"})

# Menambahkan judul
plt.title('Heatmap Tingkat Pengangguran Terbuka di Kabupaten/Kota Jawa Barat
    ↪ (2018-2023)', fontsize=18, weight='bold')

# Menampilkan grafik
plt.tight_layout()
plt.show()
```

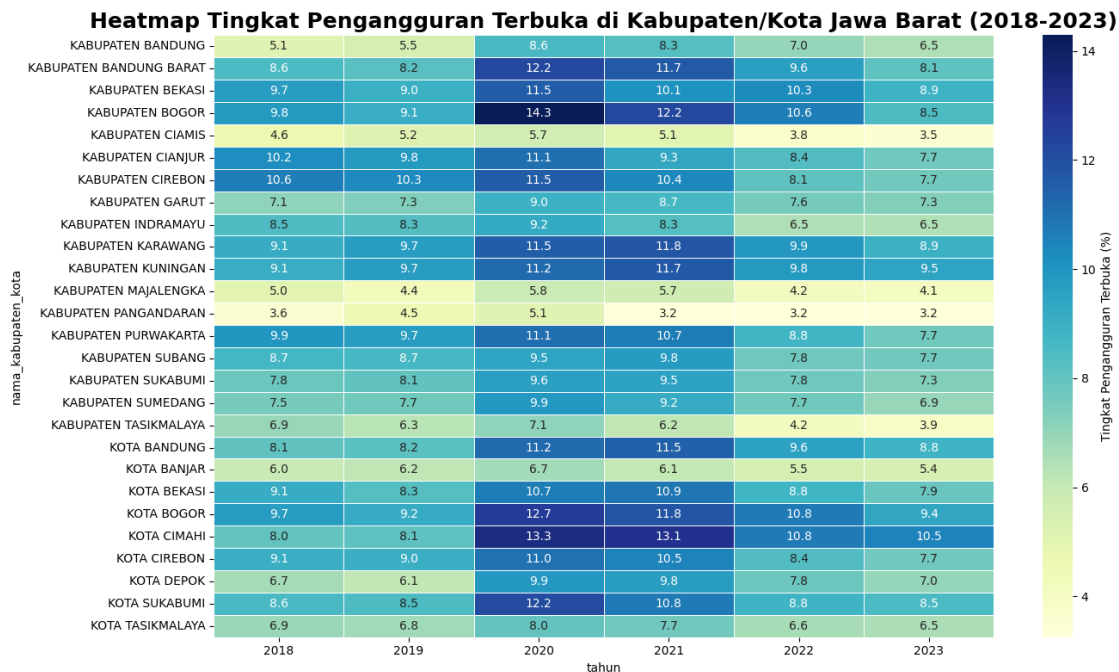


Diagram heatmap yang ditampilkan menunjukkan Tingkat Pengangguran Terbuka (TPT) per Kabupaten/Kota di Jawa Barat dari tahun 2018 hingga 2023

- Sumbu y menampilkan nama-nama kabupaten/kota di Jawa Barat



- Sumbu x menunjukkan rentang waktu atau timeline berdasarkan tahun (2018-2023)
- Warna pada heatmap menunjukkan persentase Tingkat Pengangguran Terbuka, dengan warna lebih gelap mengindikasikan tingkat pengangguran yang lebih tinggi, sementara warna lebih terang menunjukkan tingkat pengangguran yang lebih rendah

Analisis Singkat:

Diagram heatmap ini menunjukkan perubahan Tingkat Pengangguran Terbuka (TPT) di kabupaten/kota Jawa Barat dari tahun ke tahun. Kota Bogor dan Kota Cimahi sering memiliki tingkat pengangguran lebih tinggi dibandingkan rata-rata daerah lain, sedangkan Kabupaten Pangandaran secara konsisten mempertahankan TPT di bawah 6%, yang mungkin menunjukkan keberhasilan program pengurangan pengangguran di wilayah tersebut.

Secara keseluruhan, variasi TPT antar wilayah dan tahun ini dapat memberikan wawasan berharga untuk menganalisis faktor-faktor yang mungkin mempengaruhi perbedaan TPT di setiap kabupaten/kota.

### Record data Top 10 Teratas Tingkat Pengangguran Terbuka (2018-2023)

```
[100]: # Memfilter data untuk tahun 2018 hingga 2023
jpt_filtered = jpt[(jpt['tahun'] >= 2018) & (jpt['tahun'] <= 2023)]

# Menghitung rata-rata tingkat pengangguran terbuka per kabupaten/kota
average_unemployment = jpt_filtered.
    ↳groupby('nama_kabupaten_kota')['tingkat_pengangguran_terbuka'].mean().
    ↳reset_index()

# Mengurutkan kabupaten/kota berdasarkan rata-rata pengangguran terbuka secara_
    ↳menurun
top_unemployment = average_unemployment.
    ↳sort_values(by='tingkat_pengangguran_terbuka', ascending=False).head(10)

# Menampilkan hasil
print("10 Kab/Kota dengan Tingkat Pengangguran Terbuka Tertinggi (2018-2023):")
print(top_unemployment)
```

10 Kab/Kota dengan Tingkat Pengangguran Terbuka Tertinggi (2018-2023):

	nama_kabupaten_kota	tingkat_pengangguran_terbuka
3	KABUPATEN BOGOR	10.760000
22	KOTA CIMAHI	10.625000
21	KOTA BOGOR	10.590000
10	KABUPATEN KUNINGAN	10.163333
9	KABUPATEN KARAWANG	10.161667
2	KABUPATEN BEKASI	9.925000
6	KABUPATEN CIREBON	9.775000
1	KABUPATEN BANDUNG BARAT	9.738333
13	KABUPATEN PURWAKARTA	9.651667
25	KOTA SUKABUMI	9.561667

Visualisasi Top 10 Teratas

```
[101]: # Membuat grafik batang
plt.figure(figsize=(12, 6))
sns.barplot(x='tingkat_pengangguran_terbuka', y='nama_kabupaten_kota',
            data=top_unemployment, palette='viridis')

# Menambahkan judul dan label
plt.title('10 Kab/Kota dengan Tingkat Pengangguran Terbuka Tertinggi_
            (2018-2023)', fontsize=16, weight='bold')
plt.xlabel('Tingkat Pengangguran Terbuka (%)', fontsize=12)
plt.ylabel('Kabupaten/Kota', fontsize=12)

# Menampilkan grafik
plt.tight_layout()
plt.show()
```



Plot Diagram di atas adalah bar chart atau diagram batang yang menunjukkan Tingkat Pengangguran Terbuka (TPT) untuk 10 kabupaten/kota dengan nilai tertinggi di Jawa Barat dari tahun 2018 hingga 2023

- Sumbu y menunjukkan nama kabupaten/kota dengan nilai TPT tertinggi
- sumbu x menunjukkan rata-rata Tingkat Pengangguran Terbuka dalam persentase (%)

#### Record data Top 10 Terendah Tingkat Pengangguran Terbuka (2018 2023)

```
[102]: # Memfilter data untuk tahun 2018 hingga 2023
jpt_filtered = jpt[(jpt['tahun'] >= 2018) & (jpt['tahun'] <= 2023)]

# Menghitung rata-rata tingkat pengangguran terbuka per kabupaten/kota
```

```

average_unemployment = jpt_filtered.
↳groupby('nama_kabupaten_kota')['tingkat_pengangguran_terbuka'].mean().
↳reset_index()

# Mengurutkan kabupaten/kota berdasarkan rata-rata pengangguran terbuka secara
↳menaik
bottom_unemployment = average_unemployment.
↳sort_values(by='tingkat_pengangguran_terbuka', ascending=True).head(10)

# Menampilkan hasil
print("10 Kab/Kota dengan Tingkat Pengangguran Terbuka Terendah (2018-2023):")
print(bottom_unemployment)

```

10 Kab/Kota dengan Tingkat Pengangguran Terbuka Terendah (2018-2023):

	nama_kabupaten_kota	tingkat_pengangguran_terbuka
12	KABUPATEN PANGANDARAN	3.823333
4	KABUPATEN CIAMIS	4.631667
11	KABUPATEN MAJALENGA	4.866667
17	KABUPATEN TASIKMALAYA	5.761667
19	KOTA BANJAR	5.981667
0	KABUPATEN BANDUNG	6.830000
26	KOTA TASIKMALAYA	7.081667
7	KABUPATEN GARUT	7.840000
24	KOTA DEPOK	7.866667
8	KABUPATEN INDRAMAYU	7.878333

### Visualisasi Top 10 Terendah

```

[103]: # Memfilter data untuk tahun 2018 hingga 2023
jpt_filtered = jpt[(jpt['tahun'] >= 2018) & (jpt['tahun'] <= 2023)]

# Menghitung rata-rata tingkat pengangguran terbuka per kabupaten/kota
average_unemployment = jpt_filtered.
↳groupby('nama_kabupaten_kota')['tingkat_pengangguran_terbuka'].mean().
↳reset_index()

# Mengurutkan kabupaten/kota berdasarkan rata-rata pengangguran terbuka secara
↳menaik
bottom_unemployment = average_unemployment.
↳sort_values(by='tingkat_pengangguran_terbuka', ascending=True).head(10)

# Membuat grafik batang
plt.figure(figsize=(12, 6))
sns.barplot(x='tingkat_pengangguran_terbuka', y='nama_kabupaten_kota',
↳data=bottom_unemployment, palette='viridis')

# Menambahkan judul dan label

```

```
plt.title('10 Kab/Kota dengan Tingkat Pengangguran Terbuka Terendah_↵
↵(2018-2023)', fontsize=16, weight='bold')
plt.xlabel('Tingkat Pengangguran Terbuka (%)', fontsize=12)
plt.ylabel('Kabupaten/Kota', fontsize=12)

# Menampilkan grafik
plt.tight_layout()
plt.show()
```



Plot Diagram di atas adalah bar chart atau diagram batang yang menunjukkan Tingkat Pengangguran Terbuka (TPT) untuk 10 kabupaten/kota dengan nilai terendah di Jawa Barat dari tahun 2018 hingga 2023

- Sumbu y menunjukkan nama kabupaten/kota dengan nilai TPT terendah
- sumbu x menunjukkan rata-rata Tingkat Pengangguran Terbuka dalam persentase (%)

## Kesimpulan

- Kabupaten Bogor mencatat tingkat pengangguran terbuka tertinggi sebesar 10,76%, diikuti oleh Kota Cimahi dan Kota Bogor dengan masing-masing 10,63% dan 10,59%.
- Kabupaten Pangandaran memiliki tingkat pengangguran terbuka terendah, yaitu 3,82%, yang menunjukkan kinerja positif dalam penyerapan tenaga kerja.
- Terdapat perbedaan yang signifikan antara kabupaten/kota dengan tingkat pengangguran terbuka tertinggi dan terendah, menunjukkan pentingnya bagi pemerintah daerah untuk menganalisis faktor-faktor penyebab pengangguran di daerah dengan TPT tinggi dan mengimplementasikan program yang bertujuan untuk menciptakan lapangan kerja. Ini bisa termasuk peningkatan pendidikan dan pelatihan.

### 0.1.8 Analisis dan Visualisasi dari Pengaruh Jumlah Sekolah(SMP dan SMA) dan Jumlah Guru(SMP dan SMA) terhadap Rata-Rata Lama Sekolah di kabupaten/kota di Jawa Barat

```
[104]: # Fungsi untuk mengekstrak tahun pertama dari format "YYYY/YYYY"
def extract_year(year_str):
    return int(year_str.split('/')[0])

smp['tahun'] = smp['tahun_ajaran'].apply(extract_year)
sma['tahun'] = sma['tahun_ajaran'].apply(extract_year)
gsmp['tahun'] = gsmp['tahun_ajaran'].apply(extract_year)
gsma['tahun'] = gsma['tahun_ajaran'].apply(extract_year)

# Filter data untuk rentang tahun 2020 hingga 2022
smp = smp[(smp['tahun'] >= 2020) & (smp['tahun'] <= 2022)]
sma = sma[(sma['tahun'] >= 2020) & (sma['tahun'] <= 2022)]
gsmp = gsmp[(gsmp['tahun'] >= 2020) & (gsmp['tahun'] <= 2022)]
gsma = gsma[(gsma['tahun'] >= 2020) & (gsma['tahun'] <= 2022)]
rrl = rrl[(rrl['tahun'] >= 2020) & (rrl['tahun'] <= 2022)]

# Menghitung jumlah sekolah untuk SMP
smp_aggregated = smp.groupby('kode_kabupaten_kota')['jumlah_sekolah'].sum().
    ↪reset_index()
smp_aggregated.columns = ['kode_kabupaten_kota', 'jumlah_smp']

# Menghitung jumlah sekolah untuk SMA
sma_aggregated = sma.groupby('kode_kabupaten_kota')['jumlah_sekolah'].sum().
    ↪reset_index()
sma_aggregated.columns = ['kode_kabupaten_kota', 'jumlah_sma']

# Mengambil data jumlah guru untuk SMP dan SMA
jumlah_guru_smp = gsmp[['kode_kabupaten_kota', 'jumlah_guru']].
    ↪rename(columns={'jumlah_guru': 'jumlah_guru_smp'})
jumlah_guru_sma = gsma[['kode_kabupaten_kota', 'jumlah_guru']].
    ↪rename(columns={'jumlah_guru': 'jumlah_guru_sma'})

# Mengambil data rata-rata lama sekolah
rata_rata_lama_sekolah = rrl[['kode_kabupaten_kota', 'rata_rata_lama_sekolah']]

# Langkah 1: Gabungkan semua data menjadi satu DataFrame berdasarkan
    ↪`kode_kabupaten_kota`
# Gabungkan data SMP dan SMA
schools_data = pd.merge(smp_aggregated, sma_aggregated,
    ↪on='kode_kabupaten_kota', how='outer')

# Gabungkan data guru SMP dan guru SMA
```

```

guru_data = pd.merge(jumlah_guru_smp, jumlah_guru_sma,
    ↪on='kode_kabupaten_kota', how='outer')

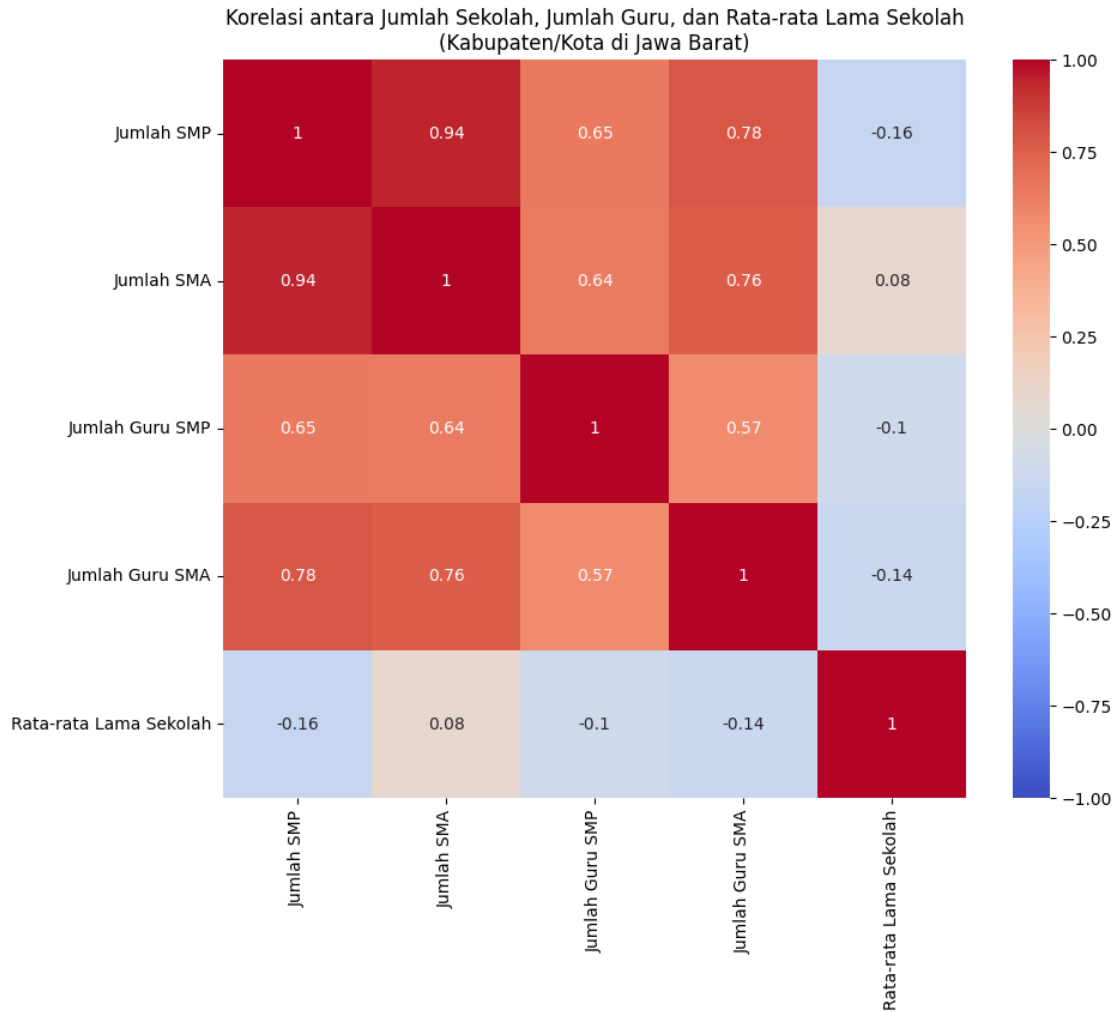
# Gabungkan semua data menjadi satu DataFrame berdasarkan `kode_kabupaten_kota`
combined_data_all = pd.merge(schools_data, guru_data, on='kode_kabupaten_kota',
    ↪how='outer')
combined_data_all = pd.merge(combined_data_all, rata_rata_lama_sekolah,
    ↪on='kode_kabupaten_kota', how='outer')

# Langkah 2: Hitung matriks korelasi untuk kelima variabel yang diinginkan
correlation_matrix_all = combined_data_all[['jumlah_smp', 'jumlah_sma',
    ↪'jumlah_guru_smp', 'jumlah_guru_sma', 'rata_rata_lama_sekolah']].corr()

# Langkah 3: Visualisasikan korelasi menggunakan heatmap
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix_all, annot=True, cmap='coolwarm', vmin=-1,
    ↪vmax=1,
            xticklabels=['Jumlah SMP', 'Jumlah SMA', 'Jumlah Guru SMP', 'Jumlah_
    ↪Guru SMA', 'Rata-rata Lama Sekolah'],
            yticklabels=['Jumlah SMP', 'Jumlah SMA', 'Jumlah Guru SMP', 'Jumlah_
    ↪Guru SMA', 'Rata-rata Lama Sekolah'])

plt.title('Korelasi antara Jumlah Sekolah, Jumlah Guru, dan Rata-rata Lama_
    ↪Sekolah\n(Kabupaten/Kota di Jawa Barat)')
plt.show()

```



1. Jumlah Sekolah (SMP dan SMA):

Jumlah SMP memiliki korelasi negatif lemah dengan rata-rata lama sekolah (-0.16), yang menunjukkan bahwa adanya lebih banyak SMP tidak berkorelasi kuat dengan rata-rata lama sekolah yang lebih tinggi. Jumlah SMA memiliki korelasi yang sedikit positif (0.08) dengan rata-rata lama sekolah, tetapi korelasi ini juga sangat lemah.

2. Jumlah Guru (SMP dan SMA):

Jumlah guru SMP memiliki korelasi negatif sangat lemah dengan rata-rata lama sekolah (-0.1), yang menunjukkan hampir tidak ada hubungan. Jumlah guru SMA juga memiliki korelasi negatif lemah dengan rata-rata lama sekolah (-0.14), menunjukkan sedikit pengaruh negatif, tetapi korelasinya sangat kecil.

**Berdasarkan korelasi yang sangat lemah antara jumlah sekolah dan jumlah guru terhadap rata-rata lama sekolah, bisa kita simpulkan bahwa jumlah sekolah dan jumlah guru (baik di tingkat SMP maupun SMA) tidak memiliki pengaruh signifikan terhadap rata-rata lama sekolah di kabupaten/kota di Jawa Barat.**

### 0.1.9 Analisis dan Visualisasi dari dataset Angka Melek Huruf Penduduk Usia 15 Tahun ke Atas dengan Indeks Pendidikan di Kabupaten/Kota Jawa Barat

Visualisasi keterhubungan record data dari Angka Melek Huruf dengan Indeks Pendidikan Per Kabupaten/Kota di Jawa Barat (2020-2022)

```
[105]: # Gabungkan dataset AMH dan IPK berdasarkan kabupaten/kota
merged_data = pd.merge(amh, ipk, on=['nama_kabupaten_kota', 'tahun'])

# Filter untuk tahun 2020-2022
filtered_data = merged_data[(merged_data['tahun'] >= 2020) &
                             (merged_data['tahun'] <= 2022)]

# Hitung rata-rata angka melek huruf dan indeks pendidikan per kabupaten/kota
avg_data = filtered_data.groupby('nama_kabupaten_kota').agg({
    'angka_melek_huruf': 'mean',
    'indeks_pendidikan': 'mean'
}).reset_index()

# Buat dictionary untuk singkatan
abbreviations = {
    'KABUPATEN BANDUNG': 'K.BDG',
    'KABUPATEN BANDUNG BARAT': 'K.BB',
    'KABUPATEN BEKASI': 'K.BKS',
    'KABUPATEN BOGOR': 'K.BGR',
    'KABUPATEN CIAMIS': 'K.CS',
    'KABUPATEN CIANJUR': 'K.CJ',
    'KABUPATEN CIREBON': 'K.CR',
    'KABUPATEN GARUT': 'K.GR',
    'KABUPATEN INDRAMAYU': 'K.IDM',
    'KABUPATEN KARAWANG': 'K.KRW',
    'KABUPATEN KUNINGAN': 'K.KNG',
    'KABUPATEN MAJALENGKA': 'K.MJL',
    'KABUPATEN PANGANDARAN': 'K.PNG',
    'KABUPATEN PURWAKARTA': 'K.PWK',
    'KABUPATEN SUBANG': 'K.SBG',
    'KABUPATEN SUKABUMI': 'K.SMI',
    'KABUPATEN SUMEDANG': 'K.SMD',
    'KABUPATEN TASIKMALAYA': 'K.TSM',
    'KOTA BANDUNG': 'KT.BDG',
    'KOTA BANJAR': 'KT.BNJ',
    'KOTA BEKASI': 'KT.BKS',
    'KOTA BOGOR': 'KT.BGR',
    'KOTA CIMAHI': 'KT.CMH',
    'KOTA CIREBON': 'KT.CR',
    'KOTA DEPOK': 'KT.DPK',
    'KOTA SUKABUMI': 'KT.SMI',
    'KOTA TASIKMALAYA': 'KT.TSM'
```



```

}

# Buat scatter plot
plt.figure(figsize=(12, 8))
plt.scatter(avg_data['angka_melek_huruf'], avg_data['indeks_pendidikan'],
            color='blue', alpha=0.6)

# Tambahkan label dan judul
plt.title('Hubungan antara Angka Melek Huruf dan Indeks Pendidikan di Kabupaten/
            Kota Jawa Barat (2020-2022)', fontsize=16)
plt.xlabel('Rata-rata Angka Melek Huruf (%)', fontsize=14)
plt.ylabel('Rata-rata Indeks Pendidikan', fontsize=14)

# Menambahkan anotasi nama kabupaten/kota untuk setiap titik dengan singkatan
for i, row in avg_data.iterrows():
    abbreviation = abbreviations.get(row['nama_kabupaten_kota'],
    row['nama_kabupaten_kota']) # Use the full name if not found
    plt.text(row['angka_melek_huruf'] + 0.1, row['indeks_pendidikan'] + 0.1,
            abbreviation, fontsize=9, color='black', ha='right', alpha=0.7)

# Mengatur grid agar lebih rapi
plt.grid(True)

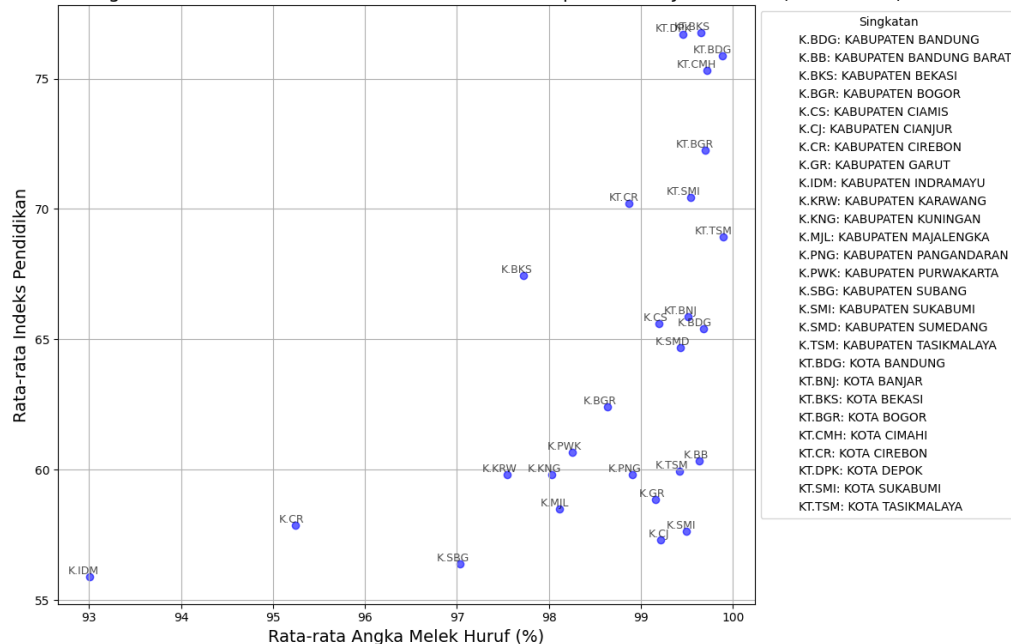
# Menambahkan legend untuk singkatan tanpa titik
handles = [plt.Line2D([0], [0], marker=None, color='w', label=f'{abbrev}:\n{full_name}')]
for full_name, abbrev in abbreviations.items():

plt.legend(handles=handles, bbox_to_anchor=(1, 1), loc='upper left',
            title='Singkatan')

# Tampilkan plot
plt.tight_layout()
plt.show()

```

Hubungan antara Angka Melek Huruf dan Indeks Pendidikan di Kabupaten/Kota Jawa Barat (2020–2022)



#### Menjawab Pertanyaan No. 4

Kesimpulan :

Secara umum, Kabupaten/Kota dengan tingkat literasi tinggi cenderung memiliki indeks pendidikan yang lebih baik. Lalu, daerah dengan Angka Melek Huruf (AMH) yang lebih rendah sering kali juga menunjukkan Indeks Pendidikan (IPK) yang lebih rendah, menandakan bahwa tingginya kemampuan literasi di daerah dapat berkontribusi positif terhadap kualitas pendidikan yang lebih baik.

Contohnya, daerah dengan IPK tertinggi umumnya adalah kota-kota besar seperti Kota Cimahi, Kota Bandung, dan Kota Bekasi yang juga memiliki AMH tertinggi. Ini menunjukkan bahwa kota-kota besar dengan literasi yang kuat juga memiliki pencapaian pendidikan yang baik.

Sebaliknya, untuk daerah-daerah seperti Kabupaten Indramayu, Kabupaten Subang, dan Kabupaten Cirebon tidak hanya memiliki IPK yang rendah tetapi juga AMH yang lebih rendah dibandingkan Kabupaten/kota lainnya. Kabupaten Indramayu yang memiliki IPK dan AMH paling rendah, menggambarkan adanya tantangan signifikan dalam bidang literasi dan pendidikan yang dapat mempengaruhi pembangunan sumber daya manusia di daerah tersebut.

#### KESIMPULAN SECARA GARIS BESAR

Analisis terhadap Indeks Pendidikan (IPK) dan berbagai faktor yang mempengaruhi kualitas pendidikan di Jawa Barat menunjukkan hubungan kompleks antara pendidikan, infrastruktur, dan kondisi sosial ekonomi.

- **Hubungan antara Indeks Pendidikan Provinsi dan Kabupaten/Kota:** Terdapat korelasi positif antara Indeks Pendidikan di tingkat provinsi dan kabupaten/kota, meskipun secara umum IPK Jawa Barat masih tergolong menengah-bawah dibandingkan provinsi lain

di Indonesia. Hal ini menunjukkan bahwa daerah-daerah dengan IPK lebih rendah mempengaruhi IPK provinsi secara keseluruhan.

- **Pengaruh Jumlah Sekolah dan Guru:** Kabupaten/kota dengan IPK tinggi umumnya memiliki jumlah sekolah dan guru yang memadai, seperti Kota Cimahi dan Kota Depok. Namun, ketersediaan fasilitas pendidikan tidak cukup hanya dengan jumlah, tetapi juga kualitas pengajaran dan manajemen yang harus diperhatikan untuk meningkatkan hasil pendidikan.
- **Rata-rata Lama Sekolah dan Fasilitas Pendidikan:** Terdapat korelasi yang sangat lemah antara jumlah sekolah dan guru terhadap rata-rata lama sekolah. Ini menunjukkan bahwa keberadaan fasilitas pendidikan dan jumlah tenaga pengajar tidak berkontribusi signifikan dalam meningkatkan lama pendidikan yang ditempuh oleh siswa di kabupaten/kota.
- **Literasi dan Indeks Pendidikan:** Kabupaten/kota dengan tingkat melek huruf yang lebih tinggi cenderung memiliki IPK yang lebih baik. Hal ini menandakan bahwa literasi berperan penting dalam kualitas pendidikan, di mana daerah dengan angka melek huruf rendah berisiko mengalami pendidikan yang berkualitas rendah.
- **Keterkaitan dengan Pengangguran:** Tingkat pengangguran terbuka di Jawa Barat menunjukkan pola yang berkaitan dengan kualitas pendidikan. Daerah dengan IPK rendah dan angka melek huruf rendah, seperti Kabupaten Indramayu, juga tercatat memiliki tingkat pengangguran yang tinggi. Sebaliknya, daerah dengan IPK tinggi cenderung memiliki pengangguran yang lebih rendah, menunjukkan bahwa pendidikan yang baik dapat berkontribusi pada penciptaan lapangan kerja dan pengurangan pengangguran.

Secara keseluruhan, untuk meningkatkan Indeks Pendidikan dan mengurangi pengangguran di Jawa Barat, diperlukan pendekatan holistik yang tidak hanya fokus pada peningkatan jumlah sekolah dan guru, tetapi juga memperhatikan kualitas pendidikan, fasilitas yang memadai, serta peningkatan literasi masyarakat. Intervensi yang tepat harus dilakukan untuk memastikan akses pendidikan yang berkualitas bagi semua lapisan masyarakat, yang pada gilirannya dapat berkontribusi pada peningkatan kesejahteraan sosial dan ekonomi di provinsi ini.