

eda-kelompok-2

October 4, 2024

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

Tanggal Pembuatan Awal : 4 Oktober 2024

Tanggal Perbaikan : -

Dibuat Oleh Kelompok 3 :

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

##Pendahuluan

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 keterse- diaan fasilitas serta sumber daya tersebut cukup untuk meningkatkan Indeks Pendidikan. Analisis ini menggunakan data dari Badan Pusat Statistik (BPS) dan Dinas Pendidikan Jawa Barat untuk mengevaluasi pengaruh fasilitas dan sumber daya pendidikan terhadap Indeks Pendidikan di berbagai wilayah di Jawa Barat. Dalam menghadapi kebutuhan pendidikan yang terus berkem- bang, 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 kabu- paten/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.
- **indeks_pendidikan_provinsi_indonesia.csv**: dataset Indeks Pendidikan ditingkat Provinsi di Indonesia.
- **jumlah_smp.csv**: dataset Jumlah Sekolah SMP ditingkat kabupaten/kota di Jawa Barat.
- **jumlah_sma.csv**: dataset Jumlah Sekolah SMA ditingkat kabupaten/kota di Jawa Barat.
- **jumlah_guru_sma.csv**: dataset Jumlah Guru SMA ditingkat kabupaten/kota di Jawa Barat.
- **jumlah_guru_smp.csv**: dataset Jumlah Guru SMP ditingkat kabupaten/kota di Jawa Barat.
- **rata_rata_lama_sekolah_berdasarkan_kabupatenkota_data.csv**: dataset rata rata lama sekolah siswa kabupaten/kota di Jawa Barat
- **harapan_lama_sekolah.csv**: dataset Harapan Lama Sekolah ditingkat kabupaten/kota di Jawa Barat.
- **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.
- **tingkat_pengangguran_terbuka.csv**: dataset Tingkat Pengangguran Terbuka ditingkat Kabupaten/Kota di Jawa Barat.

```
[ ]: %matplotlib inline
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
[ ]: #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-0zG4vsScfBOFCNQ7uHiLZ")
#jumlah smp
smp = pd.read_csv("https://drive.google.com/uc?
    ↪export=download&id=1jQF3oIe5sEIW50JTgUnmGHWRVURR4Wuq")
#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=1YnzkRFlor3NW3ItupNPJS0Qh3jQSt6sG")
#jumlah guru sma
```

```

gsma = pd.read_csv("https://drive.google.com/uc?
    ↪export=download&id=1Bt0v097V4dSgE-qeSC7QpZm5Q1DuHK1K")
#data rata_rata_lama_sekolah
rrl = pd.read_csv("https://drive.google.com/uc?
    ↪export=download&id=14RhvxatwOKtdmHu0MfItAELZCyKRxuHB")
#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 Eksplorasi Dataset Indeks Pendidikan

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

[]: ipp.head()

| | 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 |

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

[]: 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

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

```
[ ]: 0
```

Cek apakah ada record data yang null/belum diisi

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

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

Cek apakah tipe data dari tiap kolom/atribut

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

```
[ ]: 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)

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

```
[ ]: count    306.000000
mean      60.704869
std       5.387991
min       46.820000
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

1. List item
2. List item

```
[ ]: # 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

```
[ ]: # 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          ...
302          59.97
303          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

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

```
[ ]:   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 info dari dataset yang dimiliki (column, types, jmlh record)

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 348 entries, 0 to 347
Data columns (total 6 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   id               348 non-null    int64  
 1   kode_kabupaten_kota  348 non-null    int64  
 2   nama_kabupaten_kota  348 non-null    object  
 3   indeks_pendidikan    348 non-null    float64 
 4   satuan             348 non-null    object  

```

```
5 tahun           348 non-null    int64
dtypes: float64(1), int64(3), object(2)
memory usage: 16.4+ KB
```

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

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

ipk.head()
```

```
[ ]:   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

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

```
[ ]: 0
```

Cek apakah ada record data yang null/belum diisi

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

```
[ ]: 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

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

```
[ ]: 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)

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

```
[ ]: 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

```
[ ]: # 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]}")
```

Empty DataFrame
Columns: [id, kode_provinsi, nama_provinsi, kode_kabupaten_kota,

```
nama_kabupaten_kota, indeks_pendidikan, satuan, tahun]
Index: []
Jumlah outlier: 0
```

```
[ ]:
```

0.0.2 Eksplorasi Data Jumlah SMP & SMA di Jawa Barat

Cek Data yang sudah di Import

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

```
[ ]:   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
```

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

```
[ ]:   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

```
[ ]: 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
```

```
[ ]: 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

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

```
[ ]: 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
```

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

```
[ ]: 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

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

```
[ ]: 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
```

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

```
[ ]: 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

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

```
[ ]: 0
```

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

```
[ ]: 0
```

Mengahpus kolom ‘id_provinsi’ & ‘nama_provinsi’ karena jelas data nya hanya berasal dari 1 provinsi

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

smp.head()
```

```
[ ]:   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
```

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

sma.head()
```

```
[ ]:   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

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

```
[ ]: 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
```

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

```
[ ]: 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 Eksplorasi Data Jumlah Guru di SMP dan SMA di Jawa Barat

Jumlah Guru SMP di Jawa Barat 2015 - 2023

```
[ ]: # 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
```

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

```
[ ]:      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

```
[ ]: # 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
```

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

| | 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 Eksplorasi Data Harapan Lama Sekolah

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

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

| | 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)

```
[ ]: 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

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

```
[ ]: 0
```

Cek apakah ada record data yang null/belum diisi

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

```
[ ]: 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

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

```
[ ]: 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 jelas data nya hanya berasal dari 1 provinsi

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

hls.head()
```

```
[ ]:   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

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

```
[ ]: 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

```
[ ]: # 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

```
[ ]: # 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 dataset Angka Melek Huruf Penduduk Usia 15 Tahun ke Atas

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

[]: amh.head()

| | 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
```

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

```
[ ]: 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

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

```
[ ]: 0
```

Cek apakah ada record data yang null/belum diisi

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

```
[ ]: id                 0
kode_kabupaten_kota  0
nama_kabupaten_kota  0
angka_melek_huruf    0
satuan               0
tahun                0
angka_melek_huruf_robust 0
dtype: int64
```

Cek apakah tipe data dari tiap kolom/atribut

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

```
[ ]: 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 jelas data nya hanya berasal dari 1 provinsi

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

amh.head()
```

```
[ ]:    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)

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

```
[ ]: 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

```
[ ]: # 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

```
[ ]: # 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]

0.0.6 Mengeksplorasi dataset Tingkat Pengangguran Terbuka

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

[]: jpt.head()

| | id | kode_provinsi | nama_provinsi | kode_kabupaten_kota | nama_kabupaten_kota | tingkat_pengangguran_terbuka | satuan | tahun |
|---|----|---------------|---------------|---------------------|---------------------|------------------------------|--------|-------|
| 0 | 1 | 32 | JAWA BARAT | 3201 | KABUPATEN BOGOR | 14.26 | PERSEN | 2007 |
| 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 | | | |

```
1          10.85  PERSEN  2007
2          13.82  PERSEN  2007
3          17.37  PERSEN  2007
4          12.18  PERSEN  2007
```

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

```
[ ]: 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

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

```
[ ]: 0
```

Cek apakah ada record data yang null/belum diisi

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

```
[ ]: 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

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

```
[ ]: 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 jelas data nya hanya berasal dari 1 provinsi

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

jpt.head()
```

```
[ ]:   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)

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

```
[ ]: 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

```
[ ]: # 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

```
[ ]: # 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]

0.0.7 Explorasi Data Rata Rata Lama Sekolah Di Provinsi Jawa Barat

Cek informasi dataset

[]: 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

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

```
[ ]: 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
```

Menghilangkan Kolom kode_provinsi dan nama_provinsi

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

rrl.head()
```

```
[ ]:   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
```

0.1 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

```
[ ]: 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()
```

```
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `*(name,)` instead of `name` to
silence this warning.

    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `*(name,)` instead of `name` to
silence this warning.

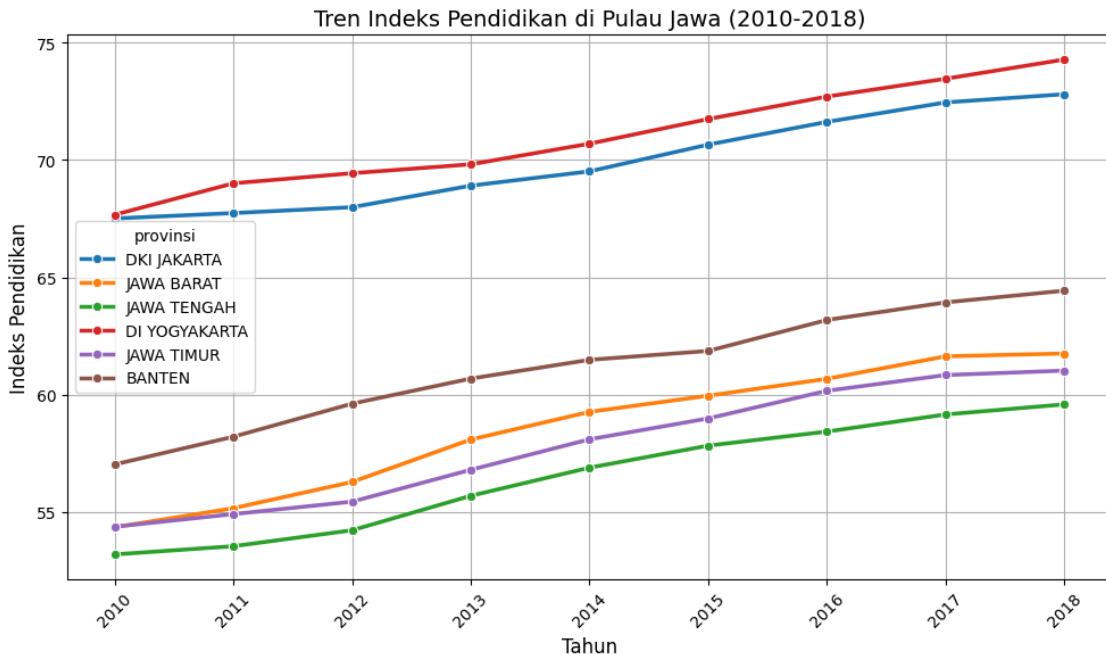
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `*(name,)` instead of `name` to
silence this warning.

    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `*(name,)` instead of `name` to
silence this warning.

    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `*(name,)` instead of `name` to
silence this warning.
```

When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
```



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

```
[ ]: provinsi_psemua = ['DI YOGYAKARTA', 'ACEH', 'KALIMANTAN TIMUR', 'SULAWESI  
↔TENGGARA', 'PAPUA BARAT',  
    'JAWA BARAT', 'BALI', 'NUSA TENGGARA TIMUR', 'NUSA  
↔TENGGARA 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()

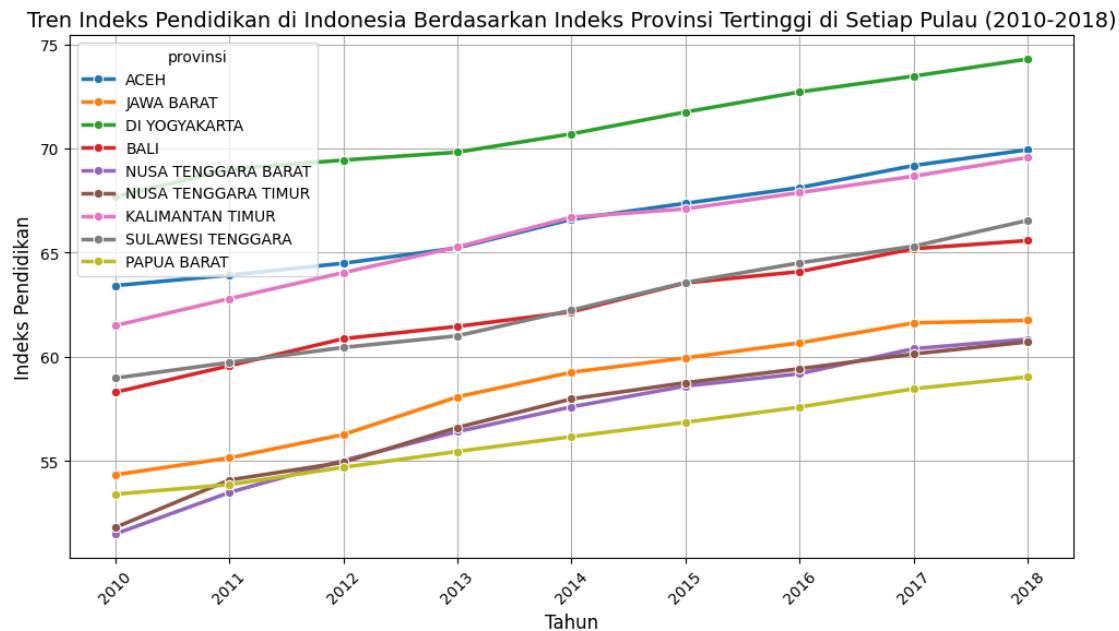
plt.show()

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `*(name,)` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `*(name,)` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `*(name,)` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `*(name,)` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `*(name,)` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `*(name,)` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `*(name,)` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `*(name,)` instead of `name` to
silence this warning.

```

to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
```



KESIMPULAN

Dari data yang sudah diekslorasi 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

```
[ ]: # 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 |

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

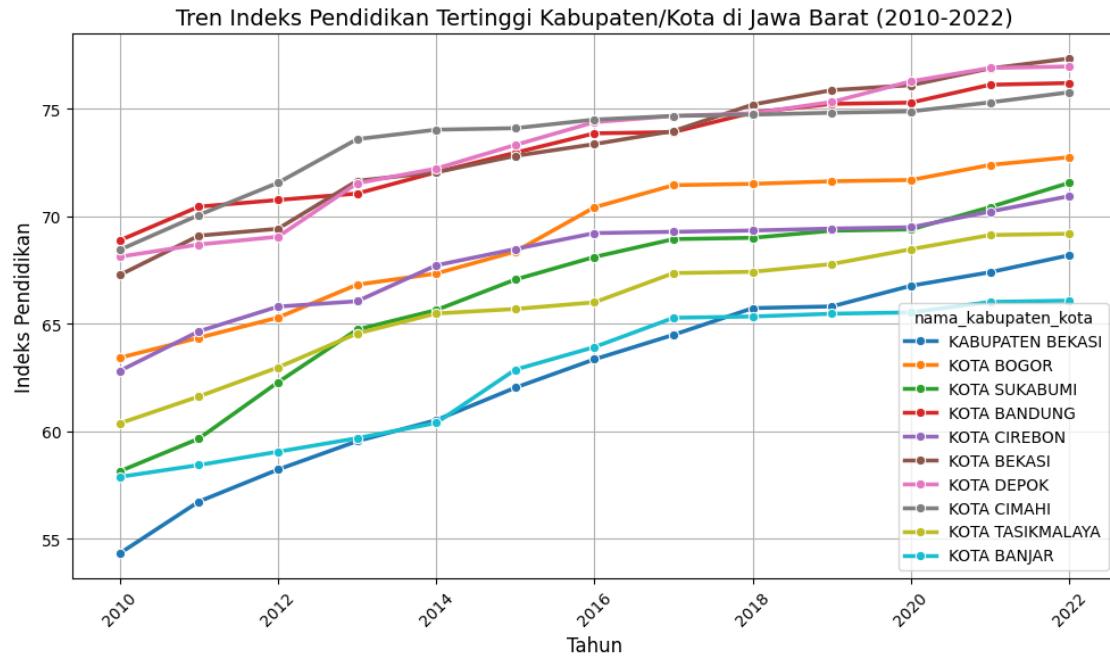
plt.title('Tren Indeks Pendidikan Tertinggi Kabupaten/Kota di Jawa Barat (2010-2022)', fontsize=14)
plt.xlabel('Tahun', fontsize=12)
plt.ylabel('Indeks Pendidikan', fontsize=12)
plt.xticks(rotation=45)
plt.grid(True)
plt.tight_layout()

plt.show()
```

```
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  
When grouping with a length-1 list-like, you will need to pass a length-1 tuple  
to get_group in a future version of pandas. Pass `*(name,)` instead of `name` to  
silence this warning.  
    data_subset = grouped_data.get_group(pd_key)  
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  
When grouping with a length-1 list-like, you will need to pass a length-1 tuple  
to get_group in a future version of pandas. Pass `*(name,)` instead of `name` to  
silence this warning.  
    data_subset = grouped_data.get_group(pd_key)  
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  
When grouping with a length-1 list-like, you will need to pass a length-1 tuple  
to get_group in a future version of pandas. Pass `*(name,)` instead of `name` to  
silence this warning.  
    data_subset = grouped_data.get_group(pd_key)  
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  
When grouping with a length-1 list-like, you will need to pass a length-1 tuple  
to get_group in a future version of pandas. Pass `*(name,)` instead of `name` to  
silence this warning.  
    data_subset = grouped_data.get_group(pd_key)  
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  
When grouping with a length-1 list-like, you will need to pass a length-1 tuple  
to get_group in a future version of pandas. Pass `*(name,)` instead of `name` to  
silence this warning.  
    data_subset = grouped_data.get_group(pd_key)  
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  
When grouping with a length-1 list-like, you will need to pass a length-1 tuple  
to get_group in a future version of pandas. Pass `*(name,)` instead of `name` to  
silence this warning.  
    data_subset = grouped_data.get_group(pd_key)  
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  
When grouping with a length-1 list-like, you will need to pass a length-1 tuple  
to get_group in a future version of pandas. Pass `*(name,)` instead of `name` to  
silence this warning.
```

```
silence this warning.
```

```
data_subset = grouped_data.get_group(pd_key)
```



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

```
[ ]: # 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 |

```
[ ]: 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()
```

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `name,` instead of `name` to
silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  

When grouping with a length-1 list-like, you will need to pass a length-1 tuple  

to get_group in a future version of pandas. Pass `name,` instead of `name` to  

silence this warning.
```

```
data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  

When grouping with a length-1 list-like, you will need to pass a length-1 tuple  

to get_group in a future version of pandas. Pass `name,` instead of `name` to  

silence this warning.
```

```
data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  

When grouping with a length-1 list-like, you will need to pass a length-1 tuple  

to get_group in a future version of pandas. Pass `name,` instead of `name` to  

silence this warning.
```

```
data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  

When grouping with a length-1 list-like, you will need to pass a length-1 tuple
```

```

to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

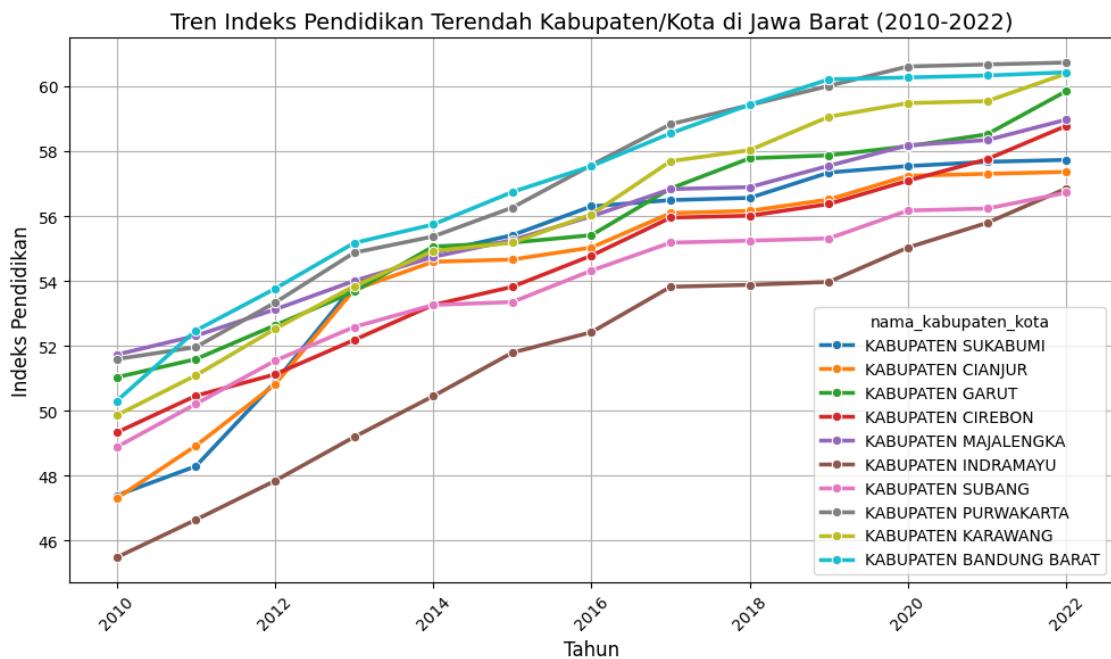
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

    data_subset = grouped_data.get_group(pd_key)

```



Kesimpulan

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

- Kabupaten/kota yang mendapat perhatian lebih dari pemerintah seperti cenderung mendapat 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 nya dan cirebon dengan hasil lautnya, diasumsikan ini lah yang menjadi pusat perhatian pemerintah.
- Kabupaten/kota yang kurang mendapat perhatian dan cenderung berimbang 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.

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

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

```
[ ]: daerah_list = [
    'KABUPATEN BOGOR', 'KABUPATEN SUKABUMI', 'KABUPATEN CIANJUR',
    'KABUPATEN BANDUNG', 'KABUPATEN GARUT', 'KABUPATEN TASIKMALAYA',
    'KABUPATEN CIAMIS', 'KABUPATEN KUNINGAN', 'KABUPATEN CIREBON',
    'KABUPATEN MAJALENGKA', 'KABUPATEN SUMEDANG', 'KABUPATEN INDRAMAYU',
    'KABUPATEN SUBANG', 'KABUPATEN PURWAKARTA', 'KABUPATEN KARAWANG',
    'KABUPATEN BEKASI', 'KABUPATEN BANDUNG BARAT', 'KABUPATEN PANGANDARAN',
    'KOTA BOGOR', 'KOTA SUKABUMI', 'KOTA BANDUNG', 'KOTA CIREBON',
    'KOTA BEKASI', 'KOTA DEPOK', 'KOTA CIMAH', 'KOTA TASIKMALAYA', 'KOTA
    ↵BANJAR'
]

tahun_ajaran_list = ['2015/2016', '2016/2017', '2017/2018', '2018/2019',
                     '2019/2020', '2020/2021', '2021/2022', '2022/2023']

# Perulangan untuk setiap daerah
for daerah in daerah_list:
    # Filter data berdasarkan daerah dan tahun ajaran
    daerah_data = gsmp[(gsmp['nama_kabupaten_kota'] == daerah) &
    ↵(gsmp['tahun_ajaran'].isin(tahun_ajaran_list))]

    # Jika data untuk daerah tersebut tersedia, buat plot
    if not daerah_data.empty:
```

```

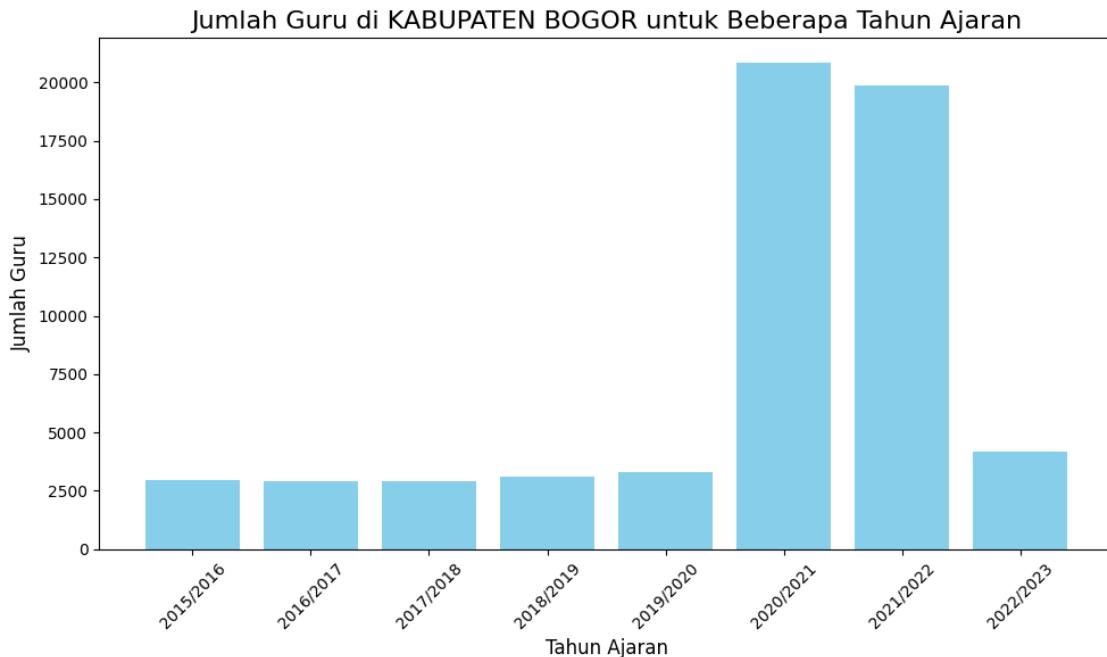
plt.figure(figsize=(10,6))
plt.bar(daerah_data['tahun_ajaran'], daerah_data['jumlah_guru'], color='skyblue')

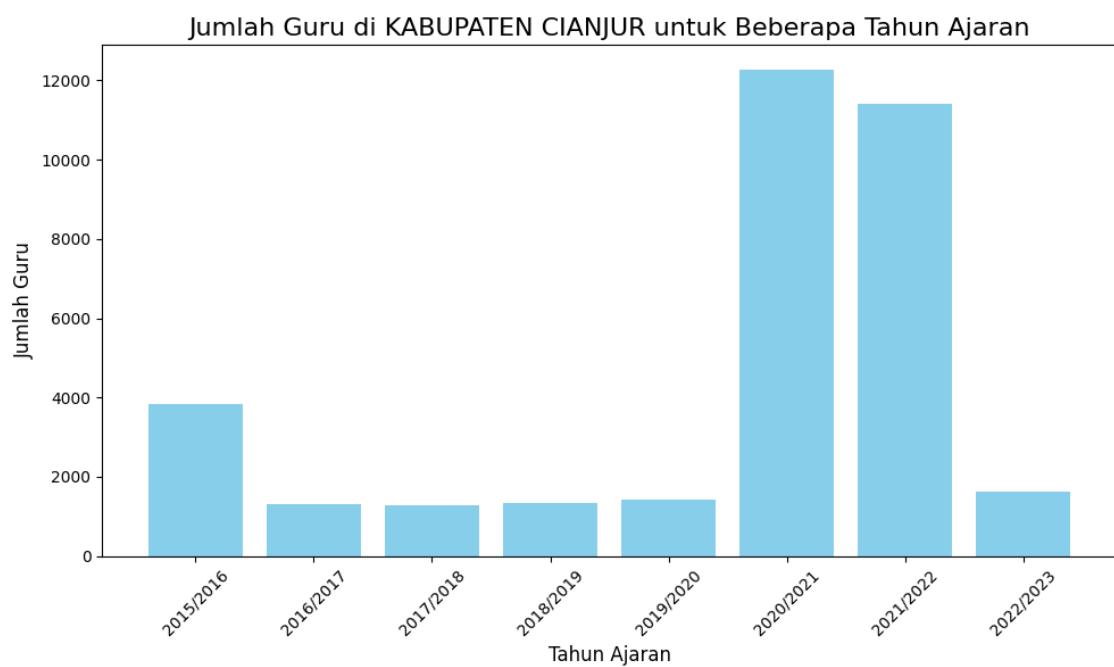
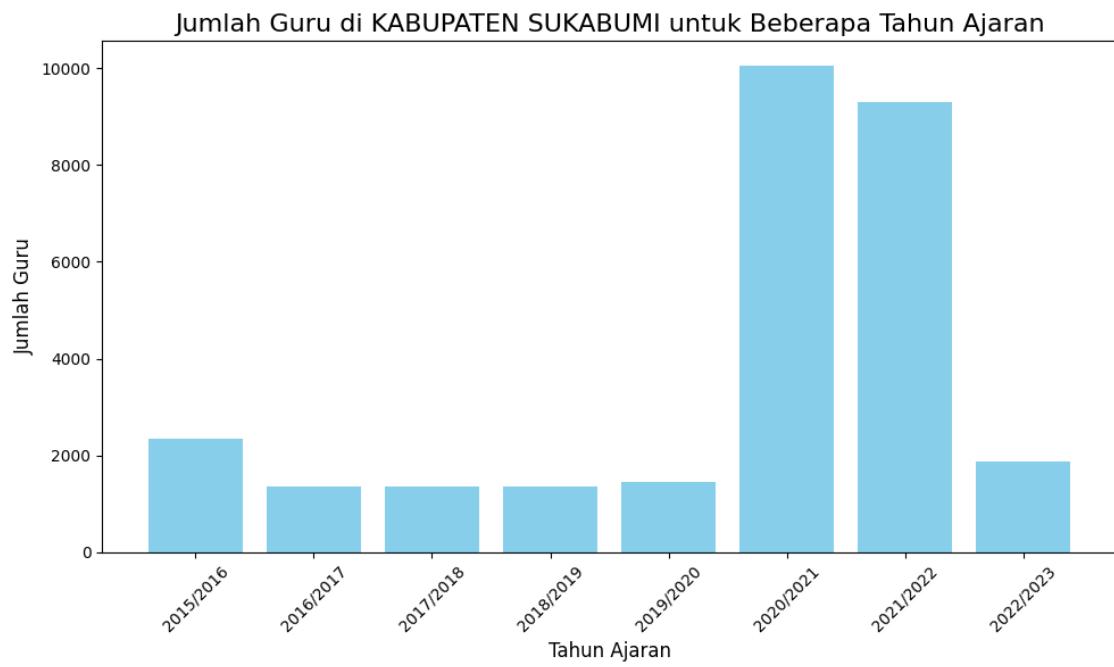
# Tambahkan judul dan label
plt.title(f'Jumlah Guru di {daerah} untuk Beberapa Tahun Ajaran', fontsize=16)
plt.xlabel('Tahun Ajaran', fontsize=12)
plt.ylabel('Jumlah Guru', fontsize=12)
plt.xticks(rotation=45)

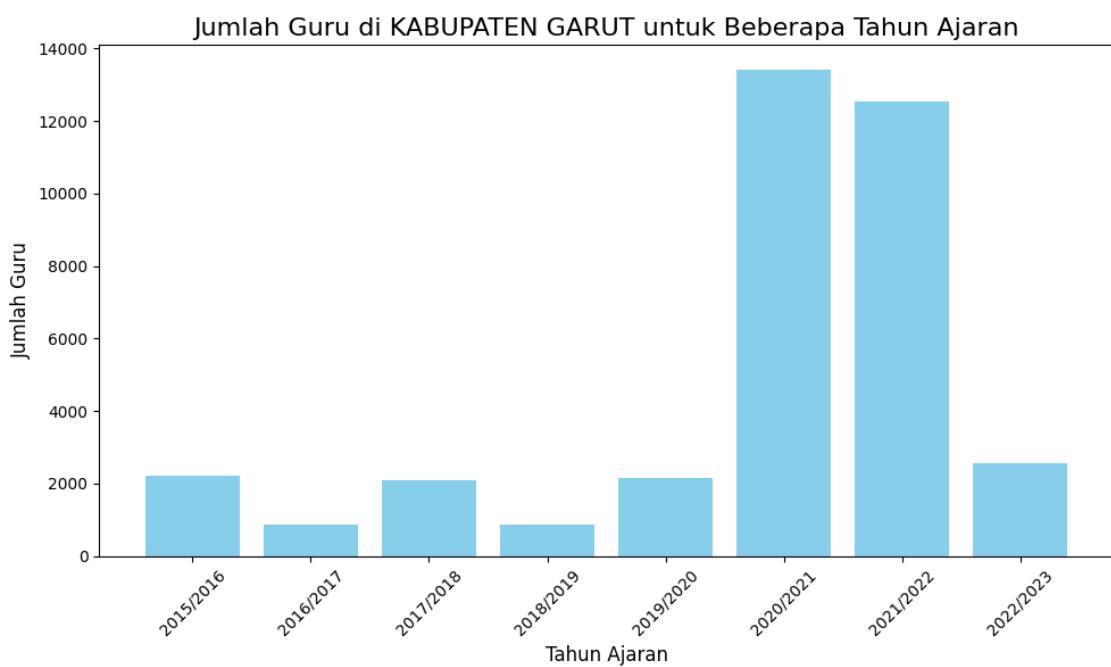
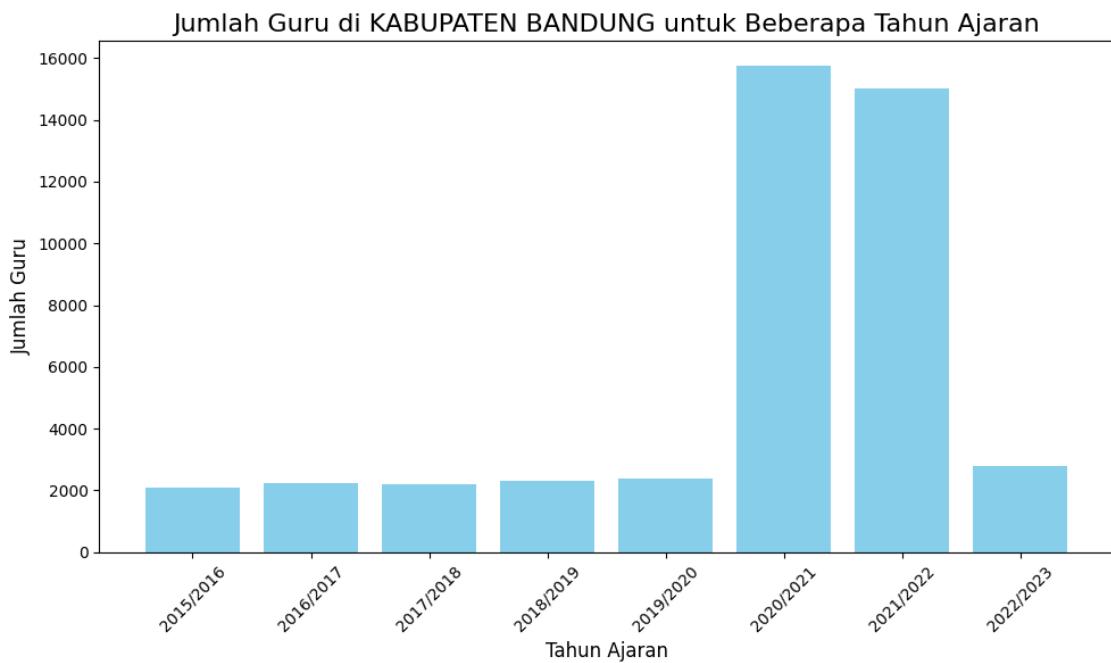
# Tampilkan diagram
plt.tight_layout()
plt.show()

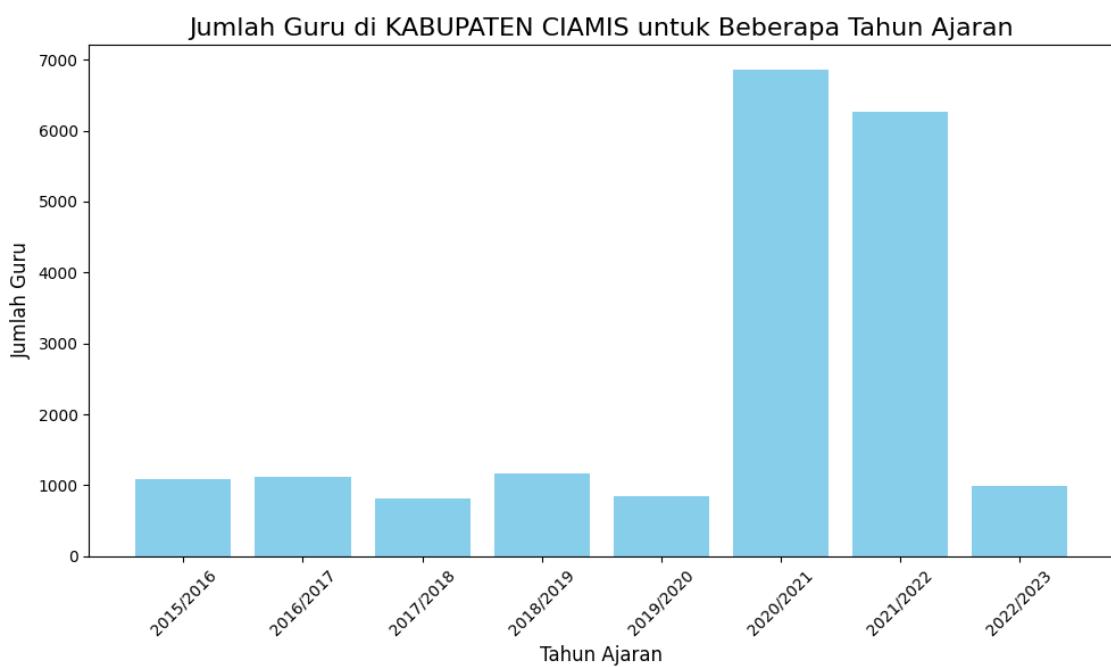
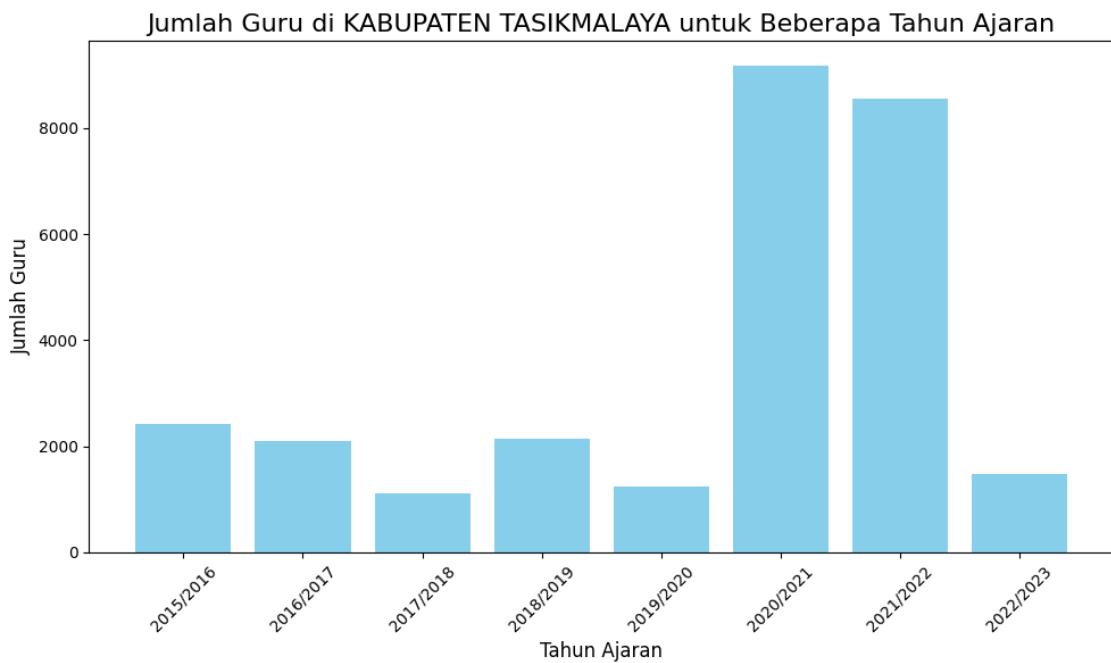
else:
    print(f"Tidak ada data untuk {daerah}")

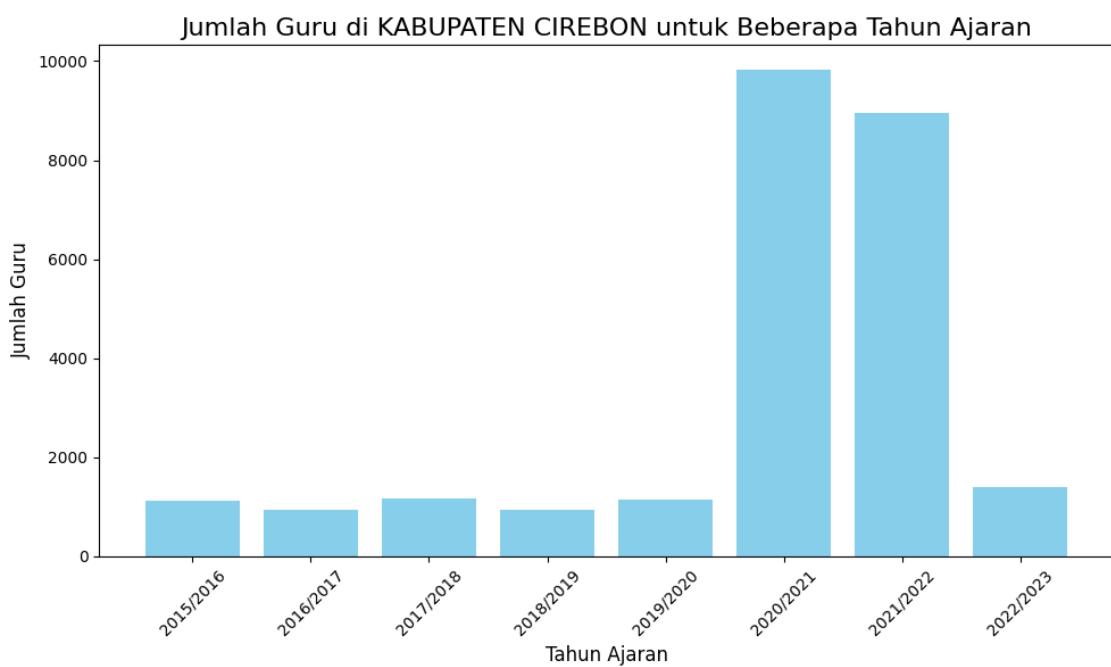
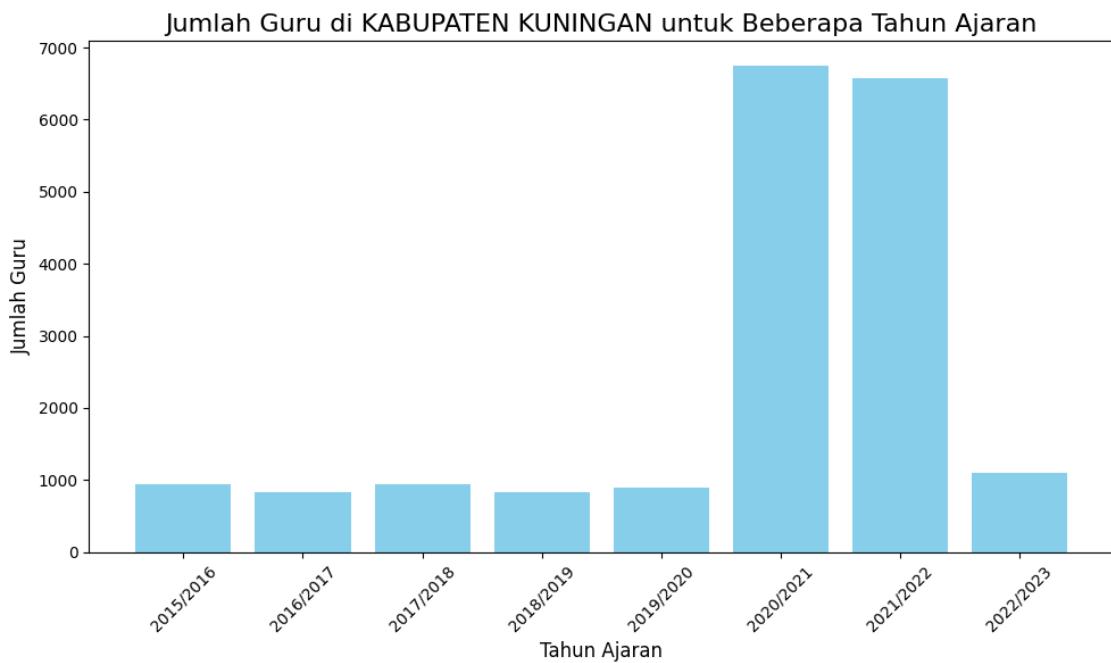
```

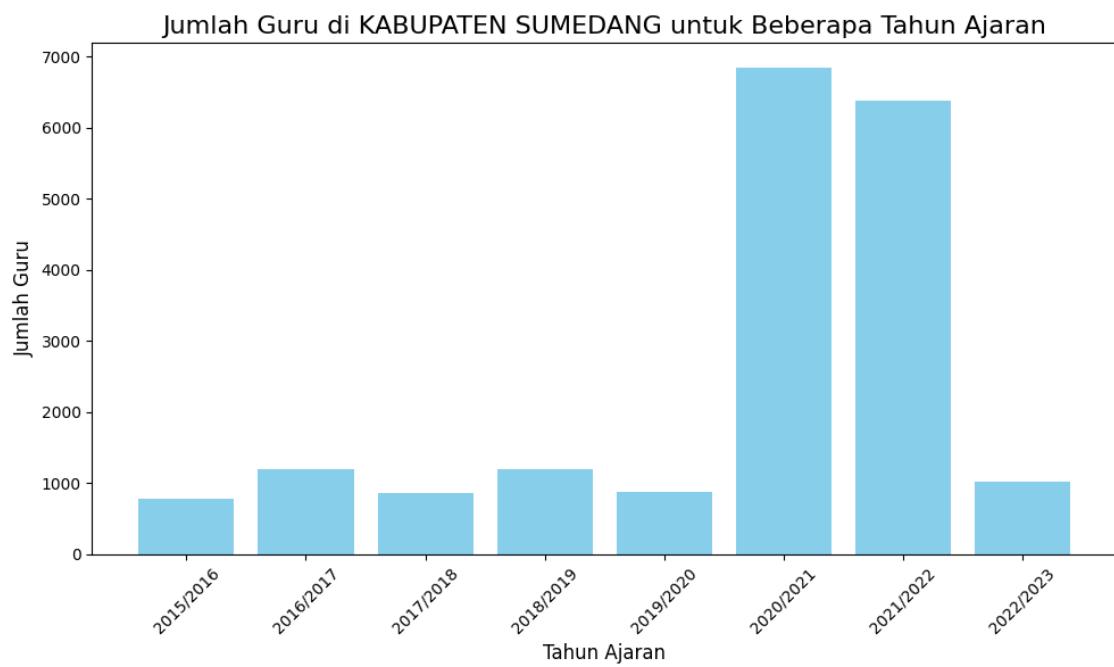
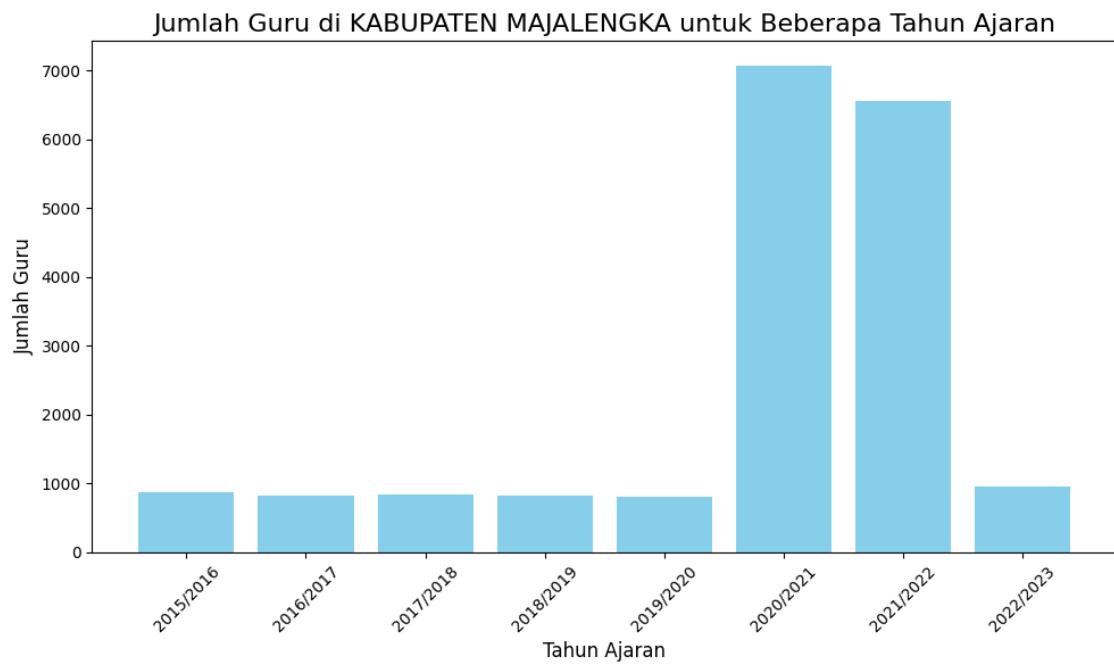




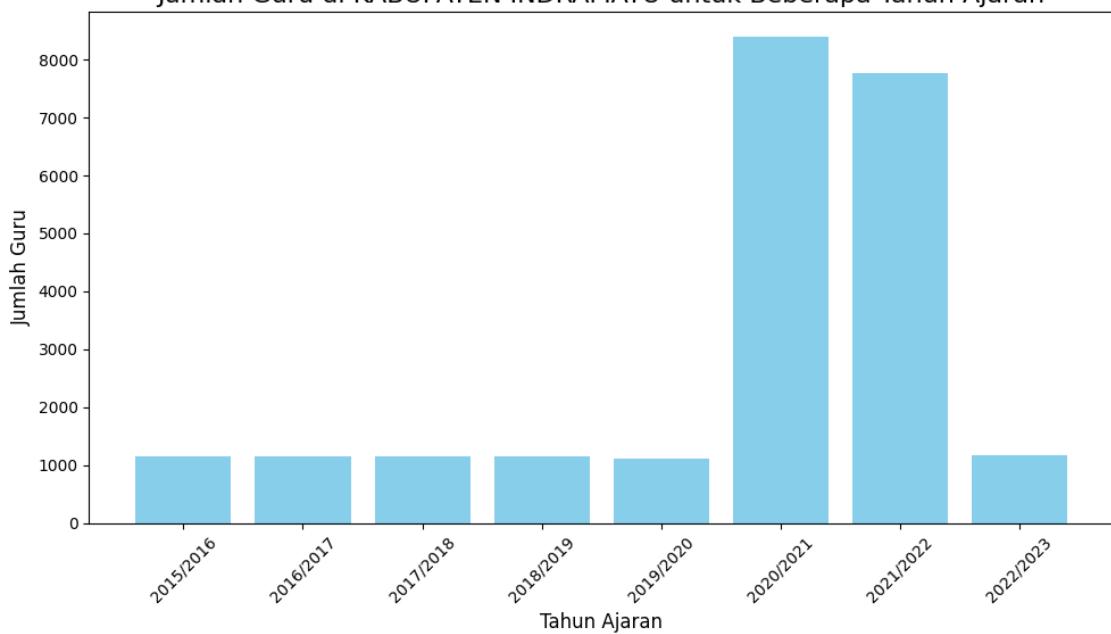




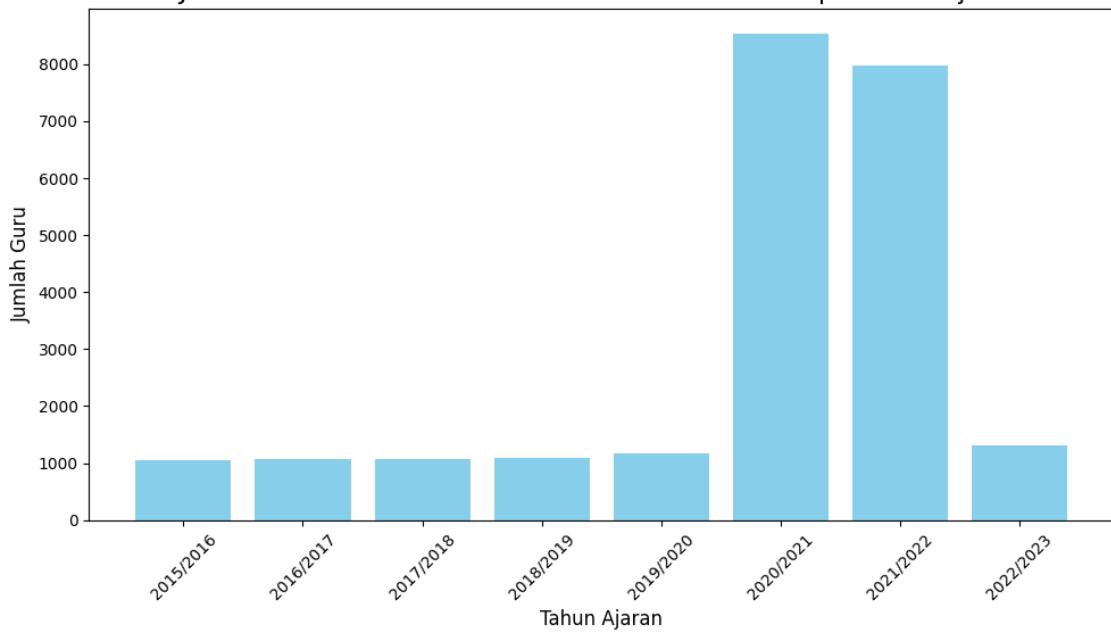


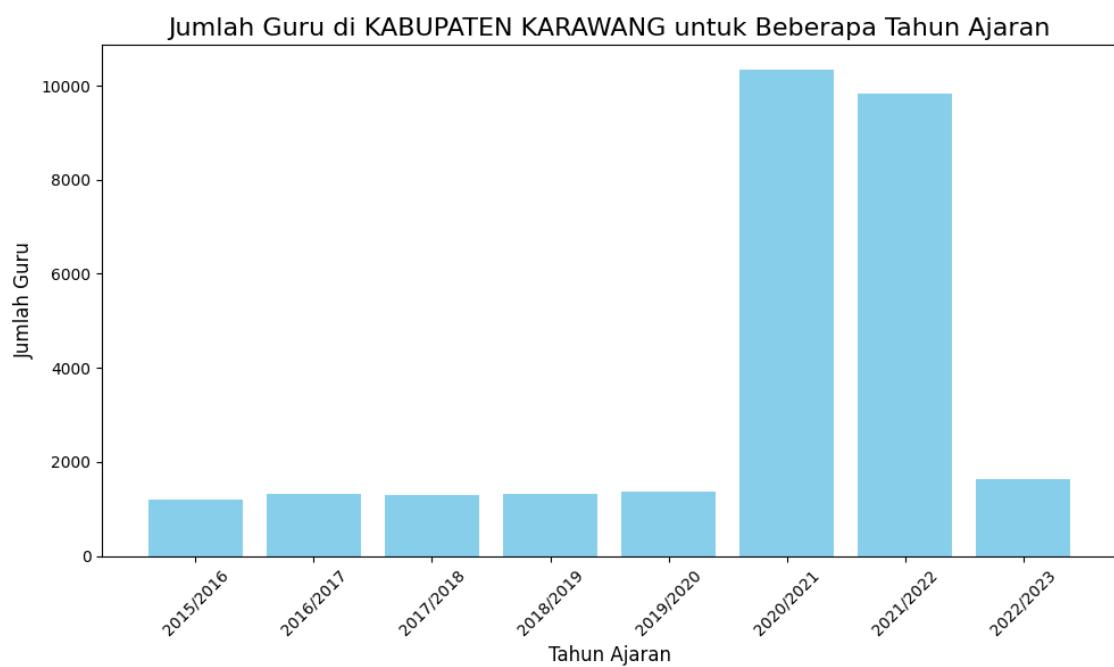
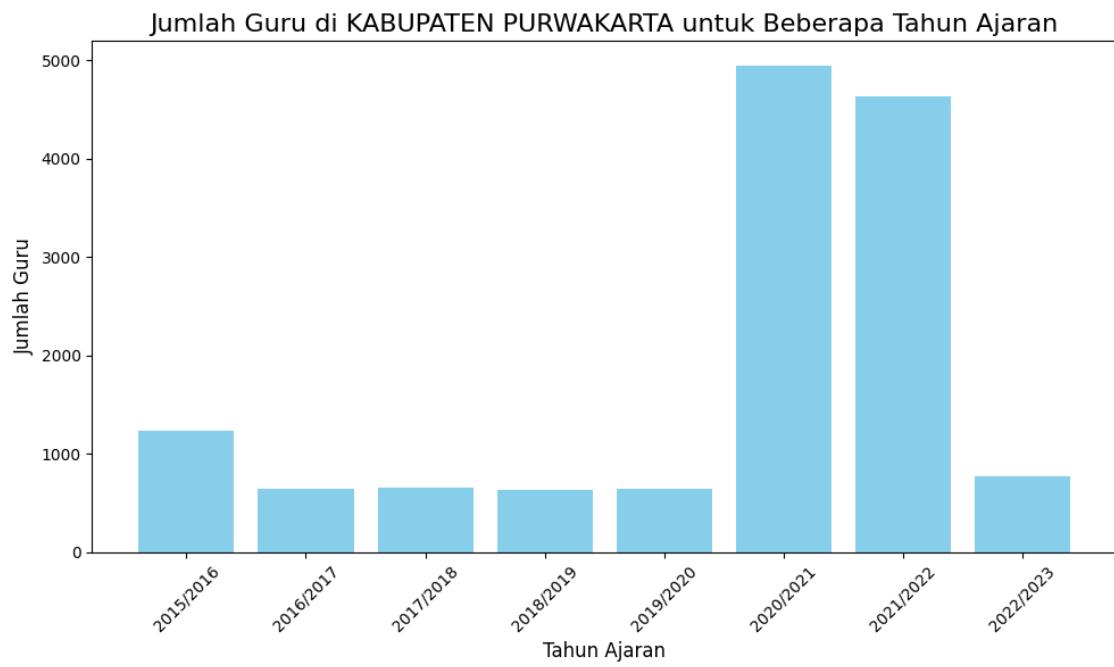


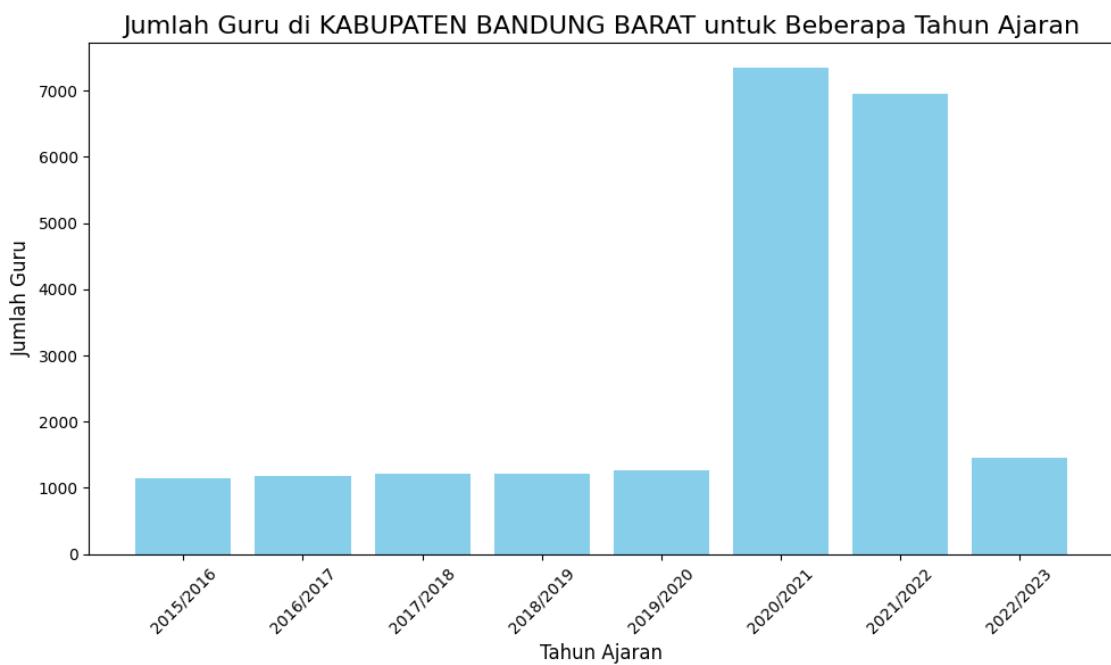
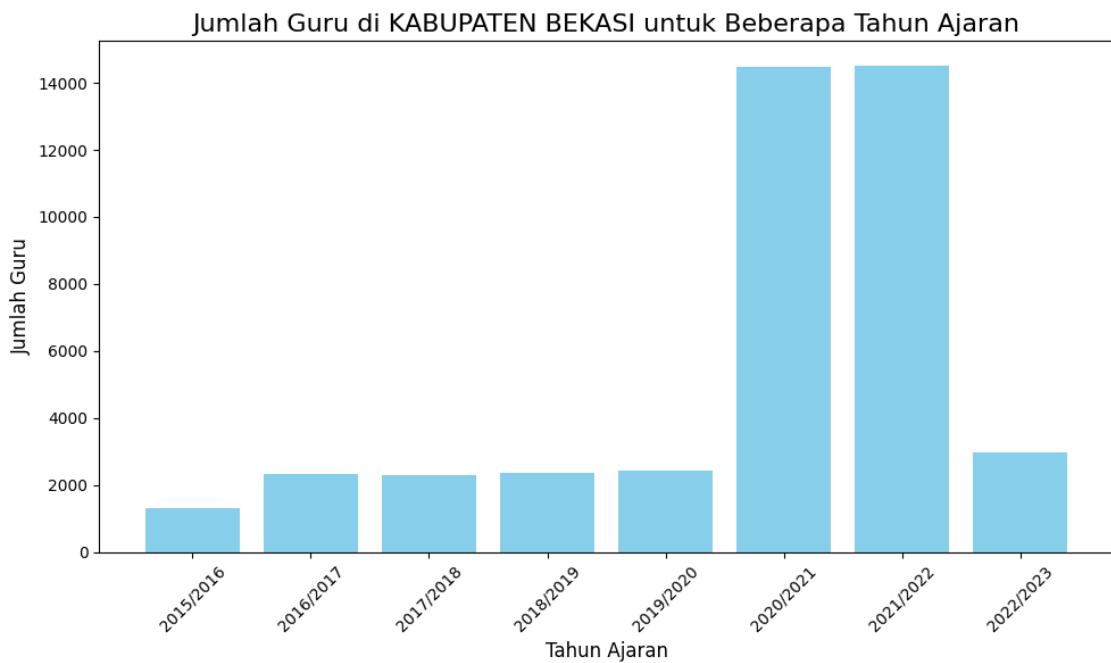
Jumlah Guru di KABUPATEN INDRAMAYU untuk Beberapa Tahun Ajaran

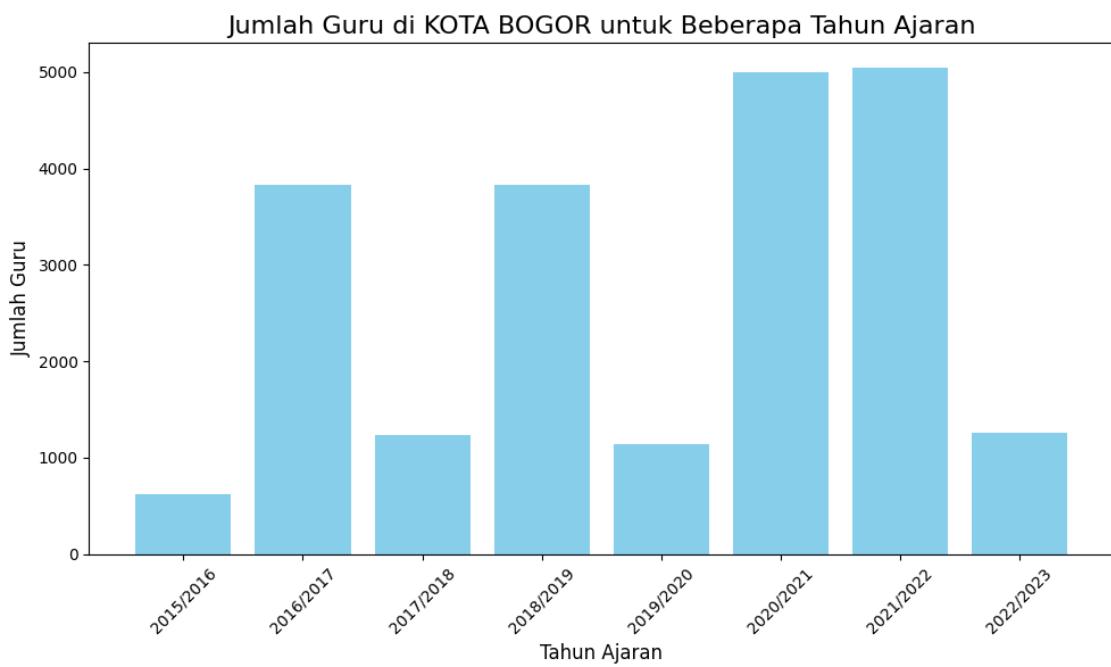
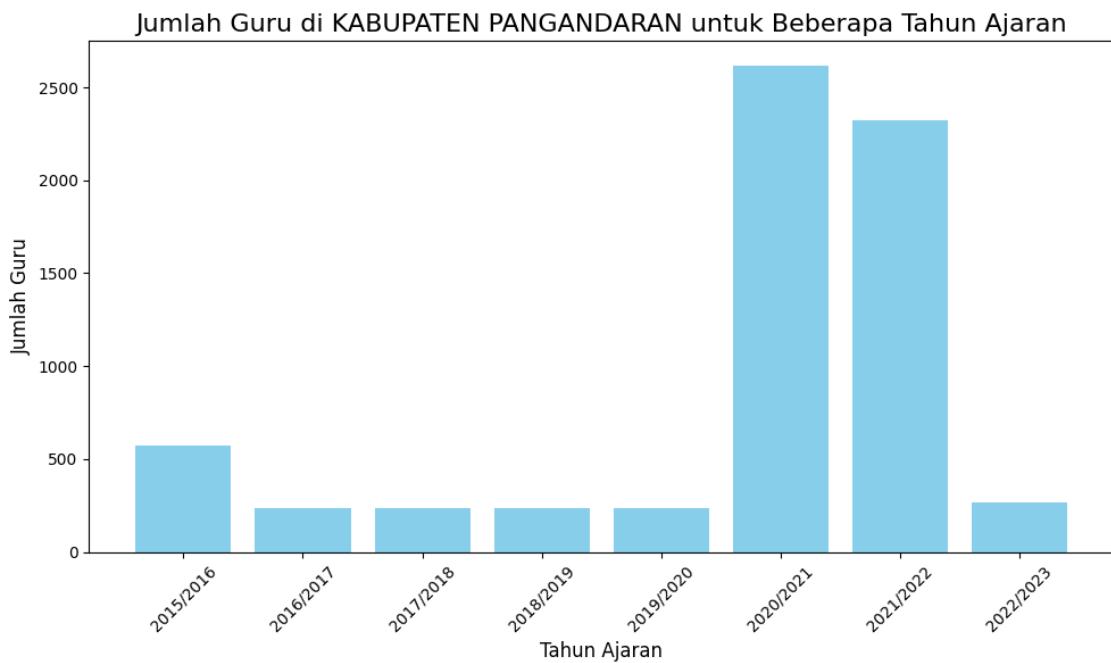


Jumlah Guru di KABUPATEN SUBANG untuk Beberapa Tahun Ajaran

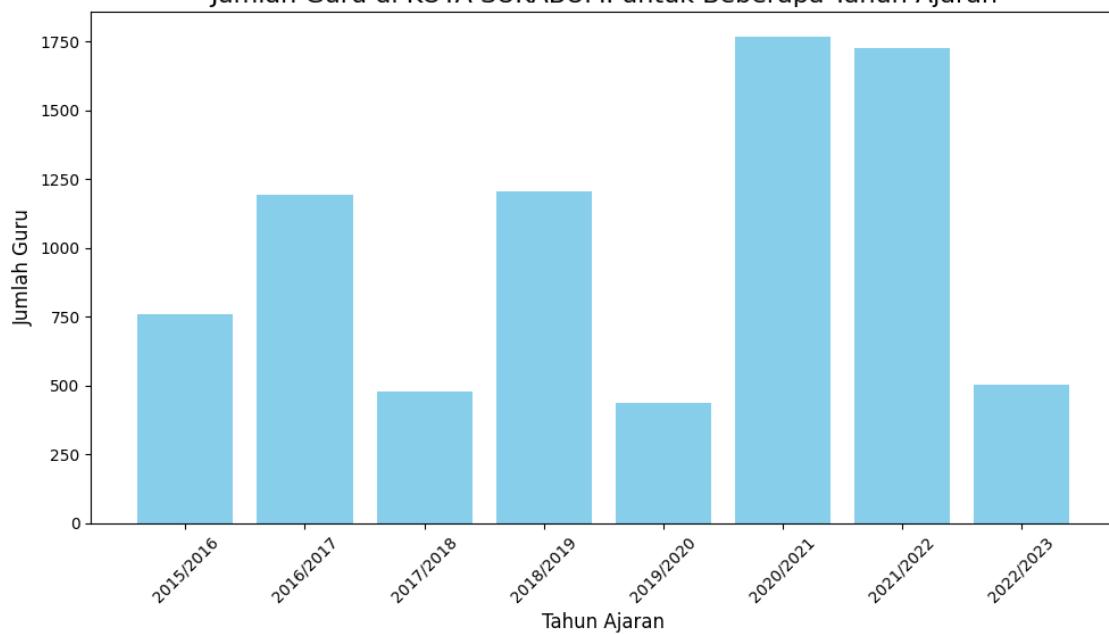




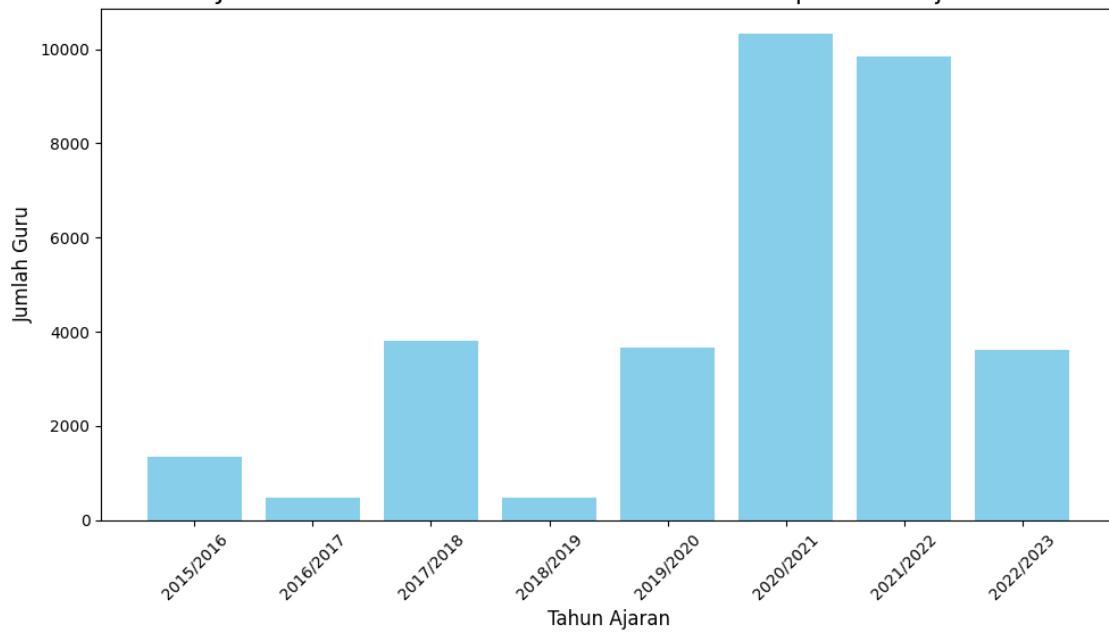




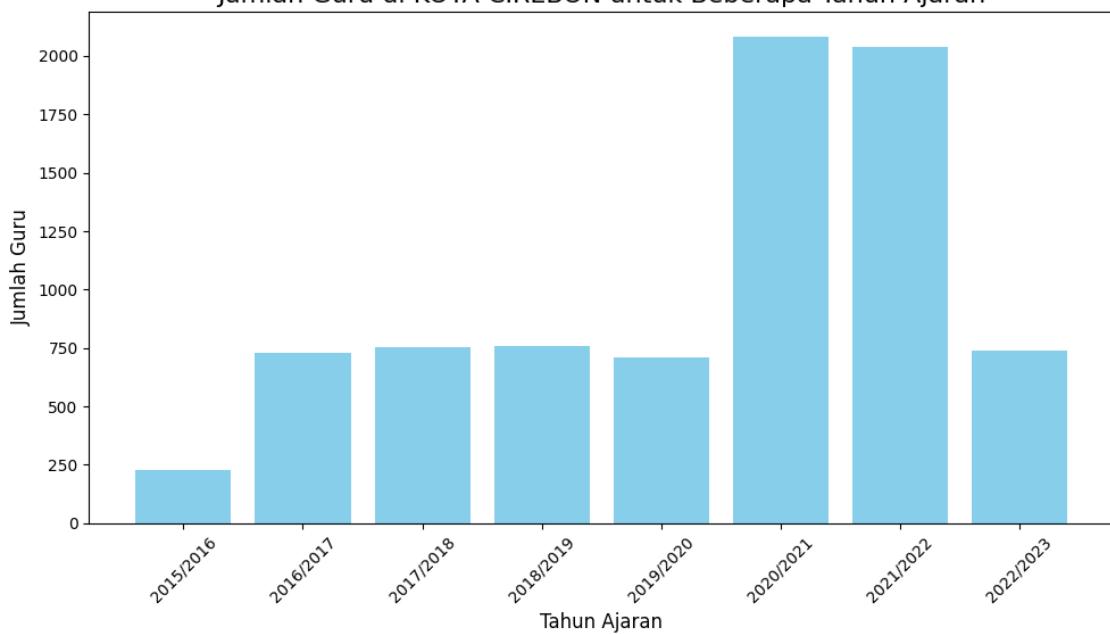
Jumlah Guru di KOTA SUKABUMI untuk Beberapa Tahun Ajaran



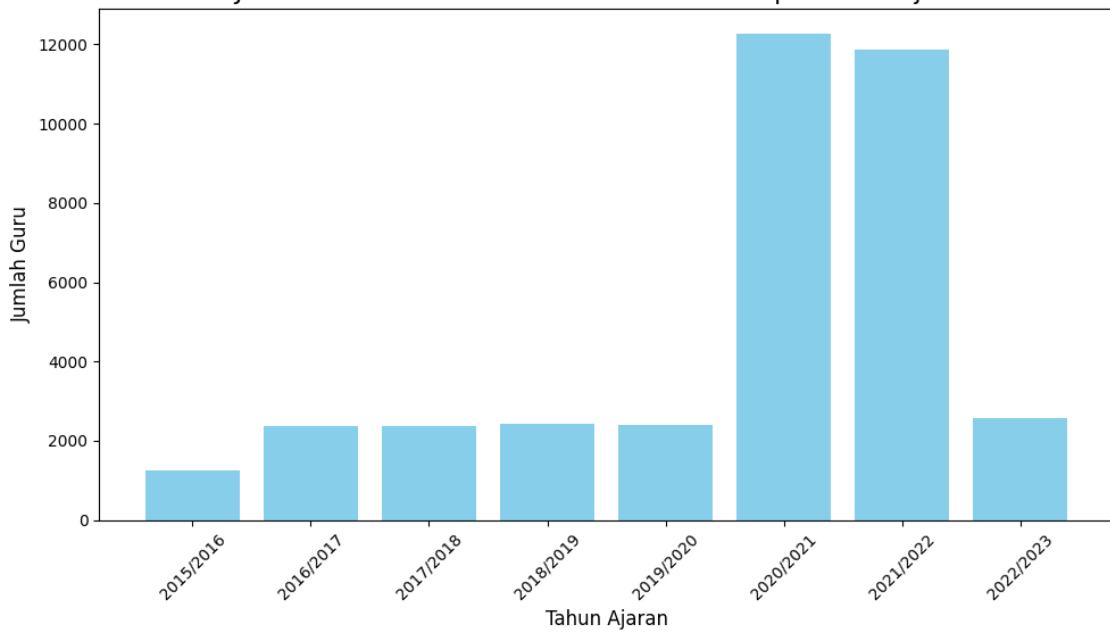
Jumlah Guru di KOTA BANDUNG untuk Beberapa Tahun Ajaran

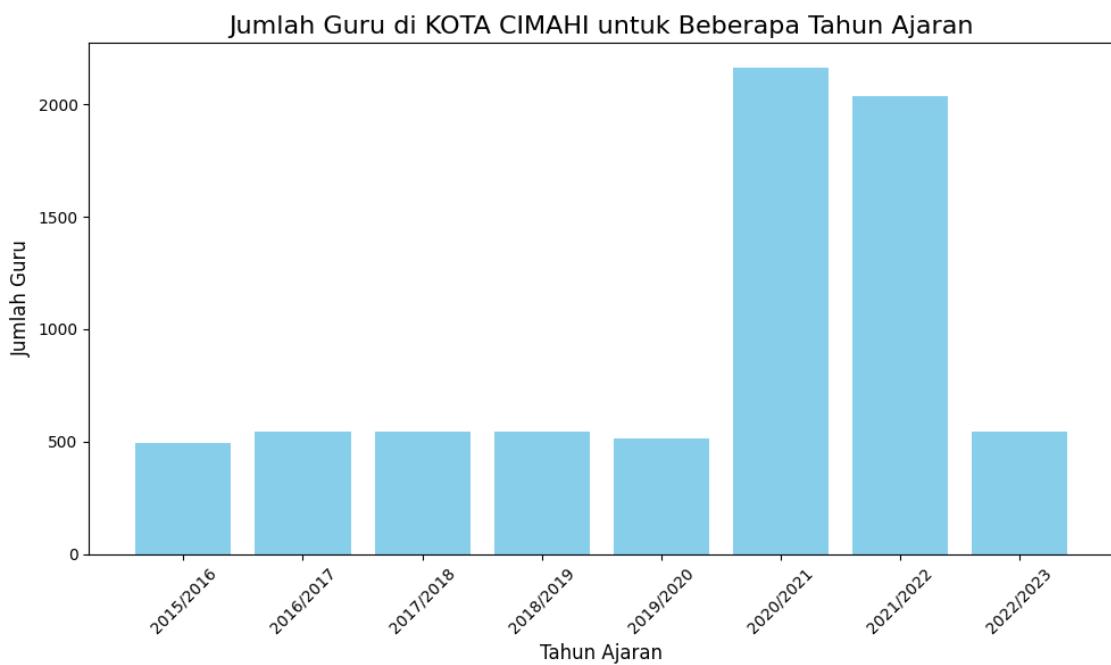
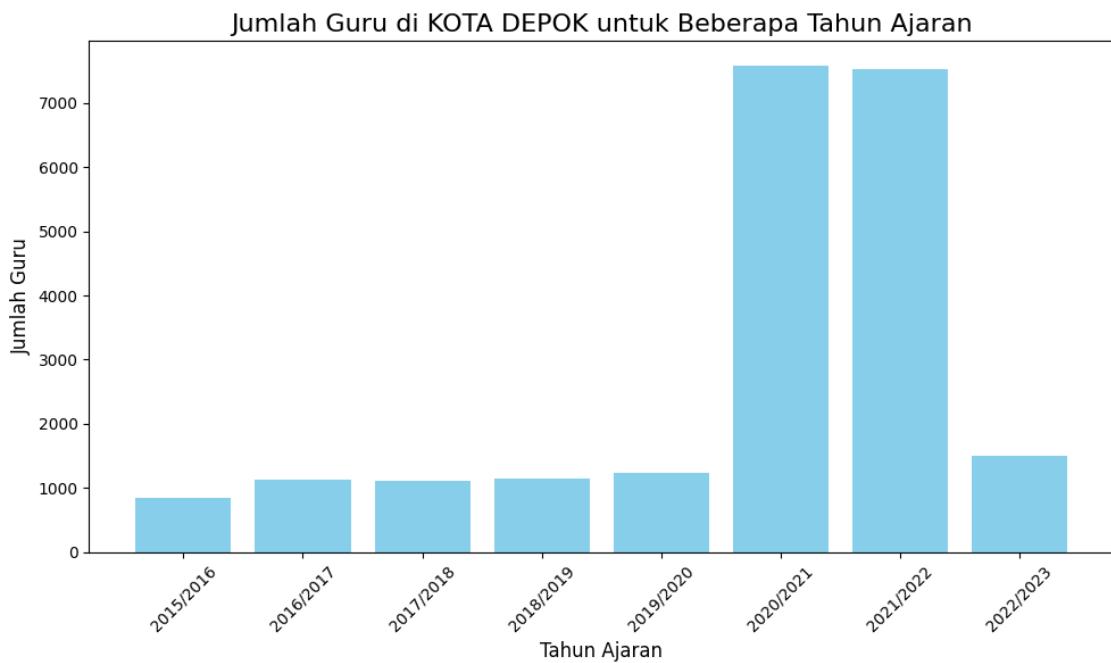


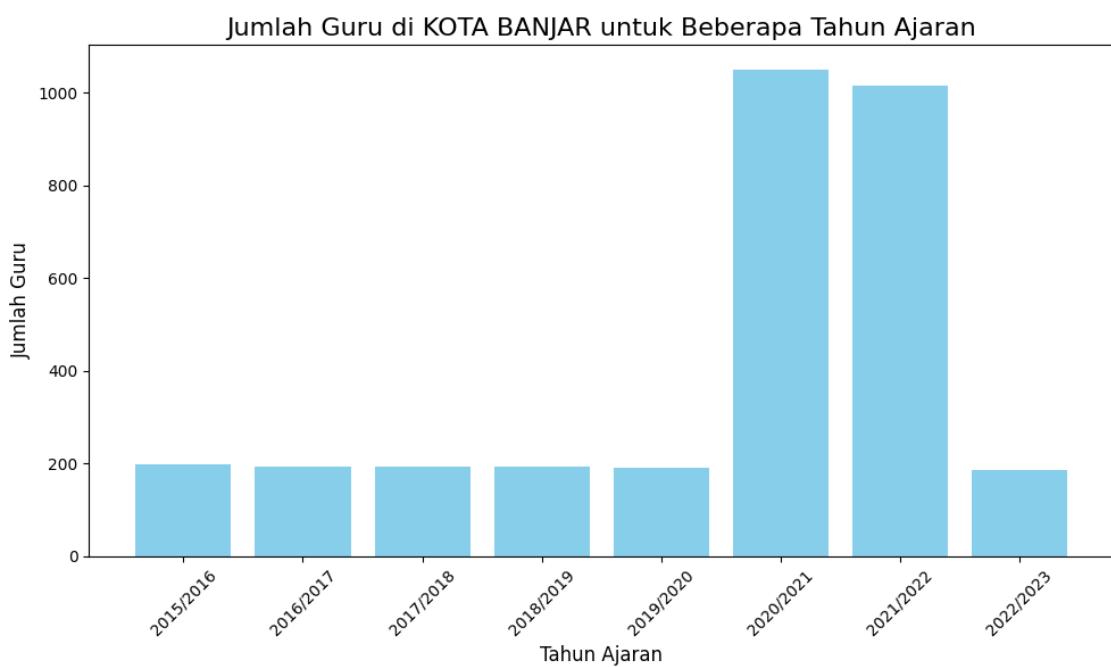
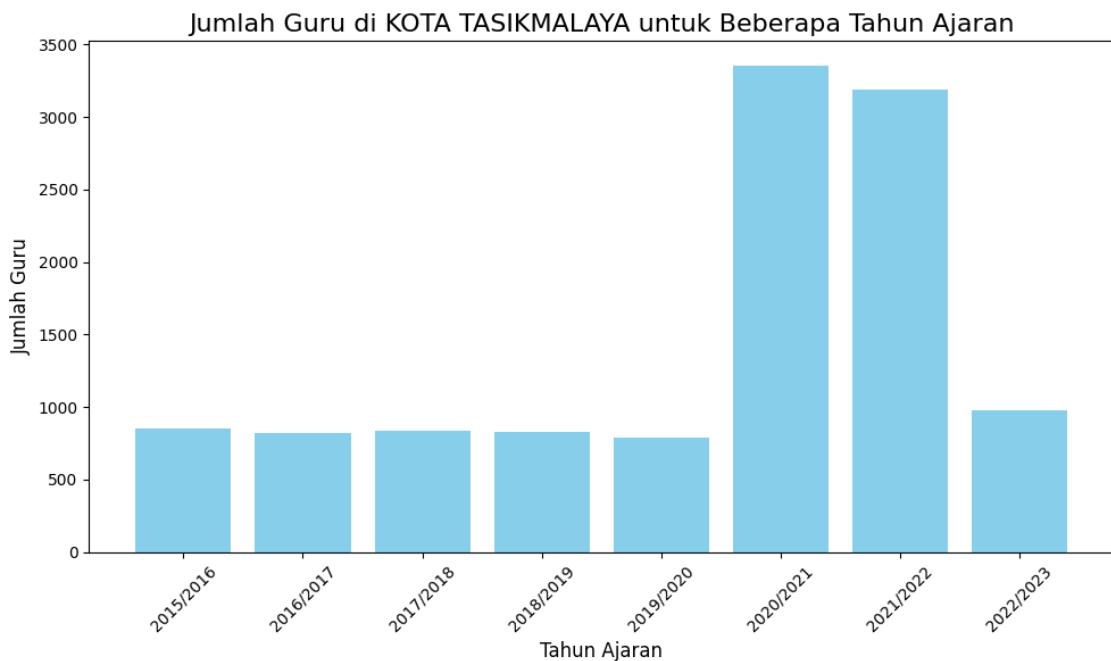
Jumlah Guru di KOTA CIREBON untuk Beberapa Tahun Ajaran



Jumlah Guru di KOTA BEKASI untuk Beberapa Tahun Ajaran







```
[ ]: # List tahun ajaran dari 2015/2016 hingga 2022/2023
tahun_ajaran_list = ['2015/2016', '2016/2017', '2017/2018', '2018/2019',
                     '2019/2020', '2020/2021', '2021/2022', '2022/2023']
```

```

# Looping untuk menampilkan diagram sebanyak jumlah tahun ajarannya kak
for tahun_ajaran in tahun_ajaran_list:
    # kita filter data, sesuai tahun ajaran
    data_per_tahun = gsmp[gsmp['tahun_ajaran'] == tahun_ajaran]

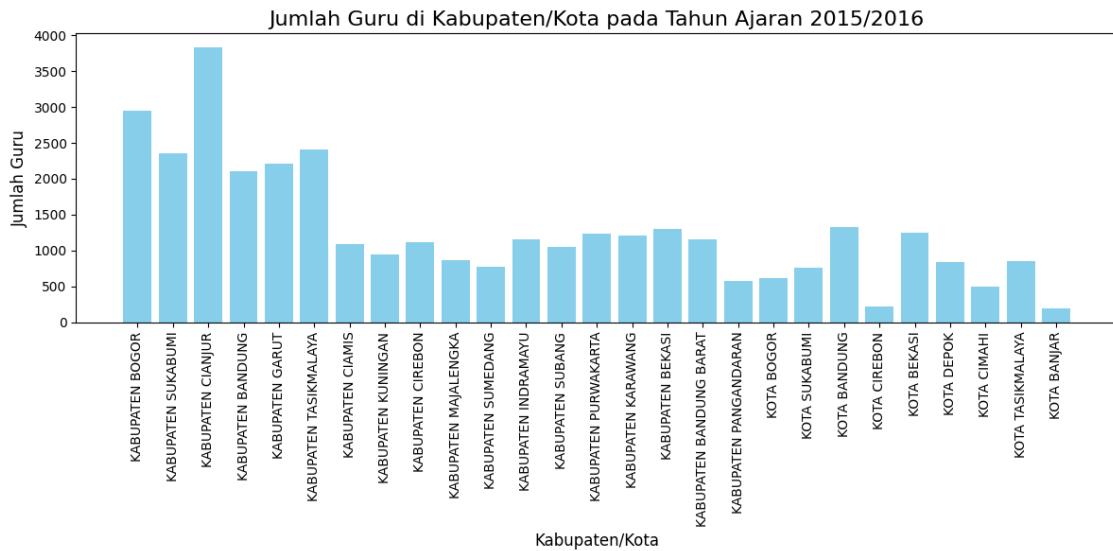
    # Jika data untuk tahun ajaran tersebut tersedia, buat plot
    if not data_per_tahun.empty:
        plt.figure(figsize=(12,6))
        plt.bar(data_per_tahun['nama_kabupaten_kota'], data_per_tahun['jumlah_guru'], color='skyblue')

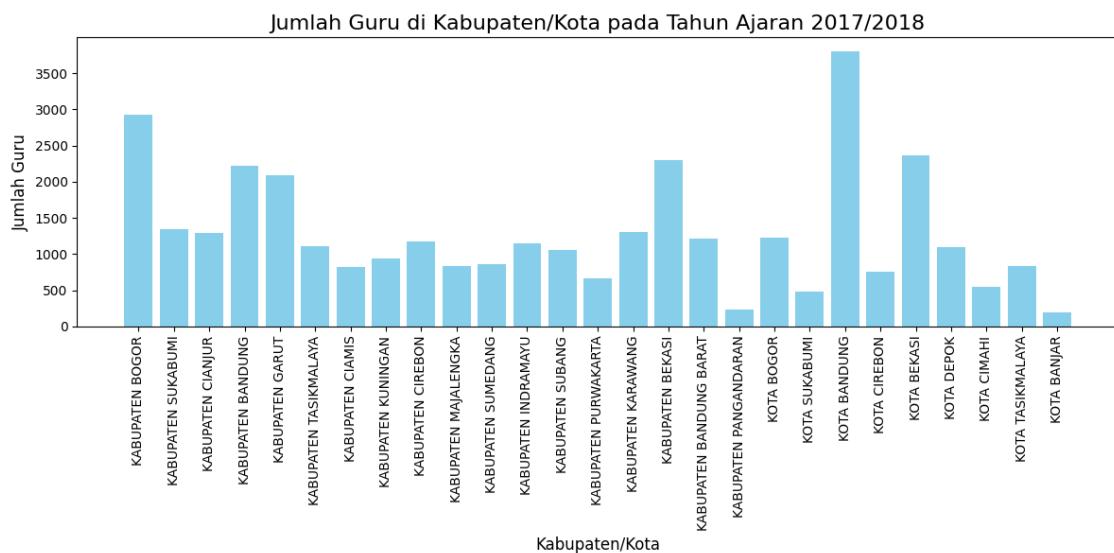
    # Tambahkan judul dan label pada diagram yang kita buat
    plt.title(f'Jumlah Guru di Kabupaten/Kota pada Tahun Ajaran {tahun_ajaran}', fontsize=16)
    plt.xlabel('Kabupaten/Kota', fontsize=12)
    plt.ylabel('Jumlah Guru', fontsize=12)
    plt.xticks(rotation=90) # Rotasi nama kabupaten/kota agar mudah terbaca

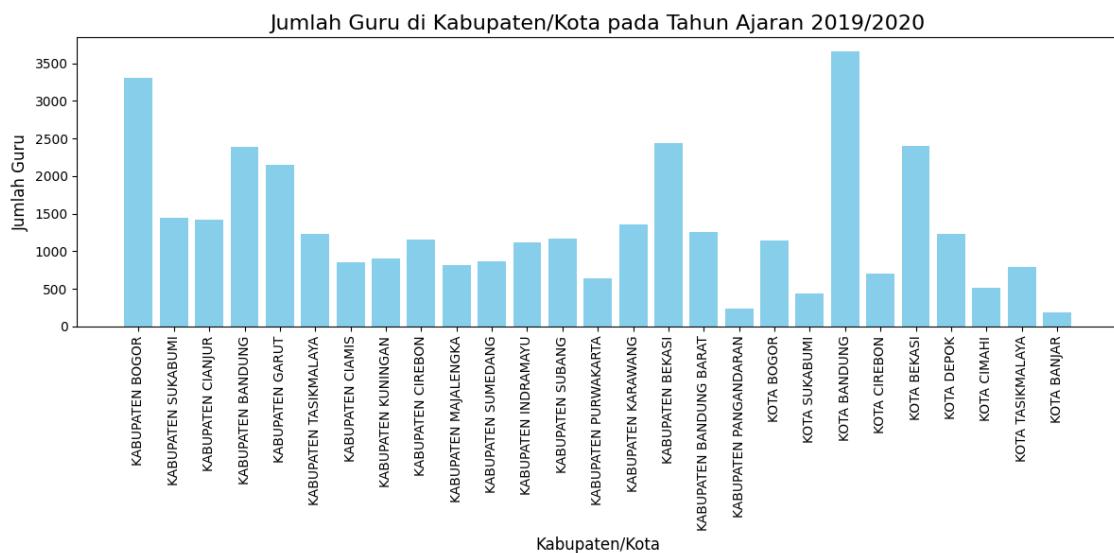
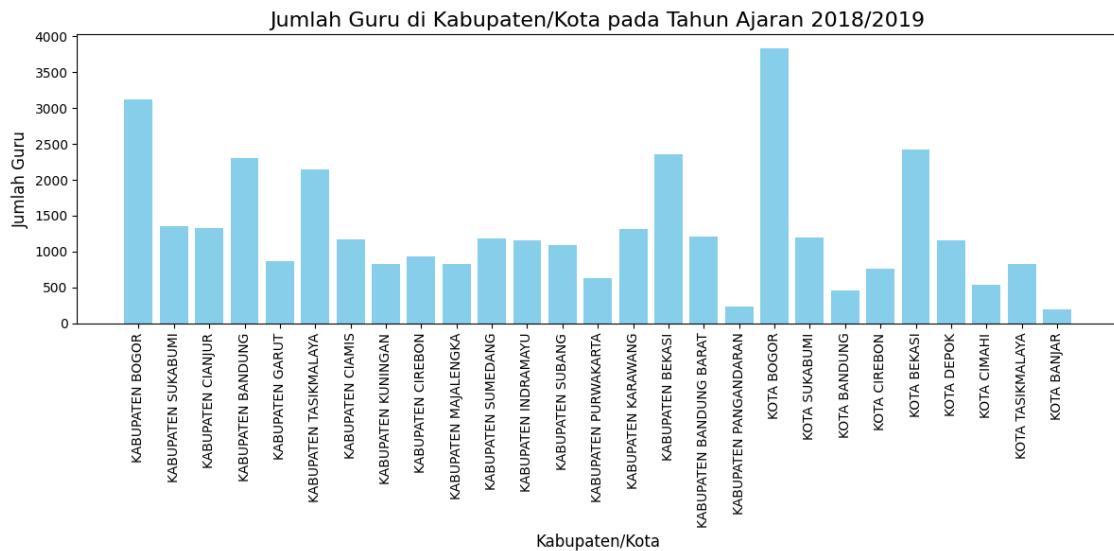
    # Tampilkan diagram
    plt.tight_layout()
    plt.show()

else:
    print(f"Tidak ada data untuk Tahun Ajaran {tahun_ajaran}")

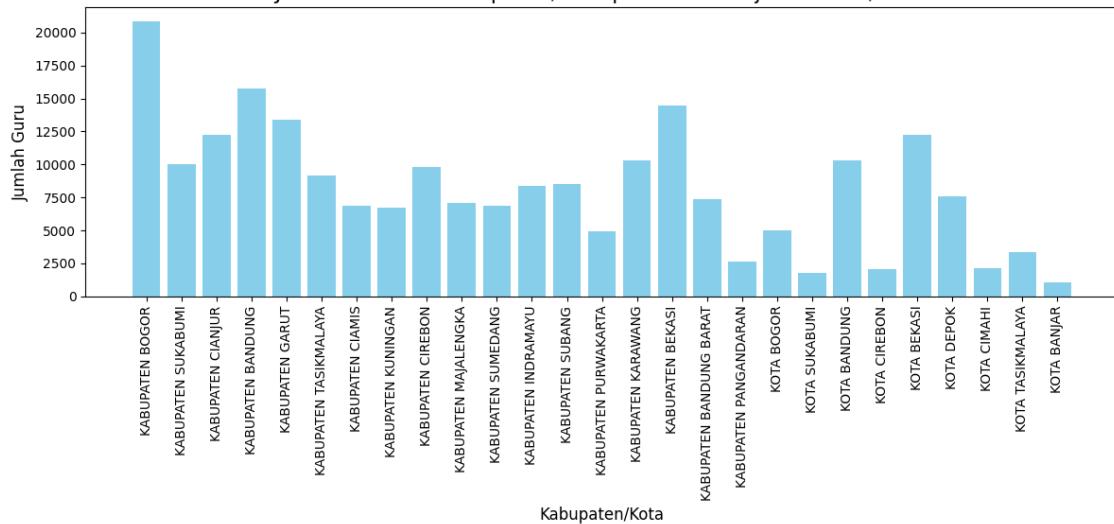
```



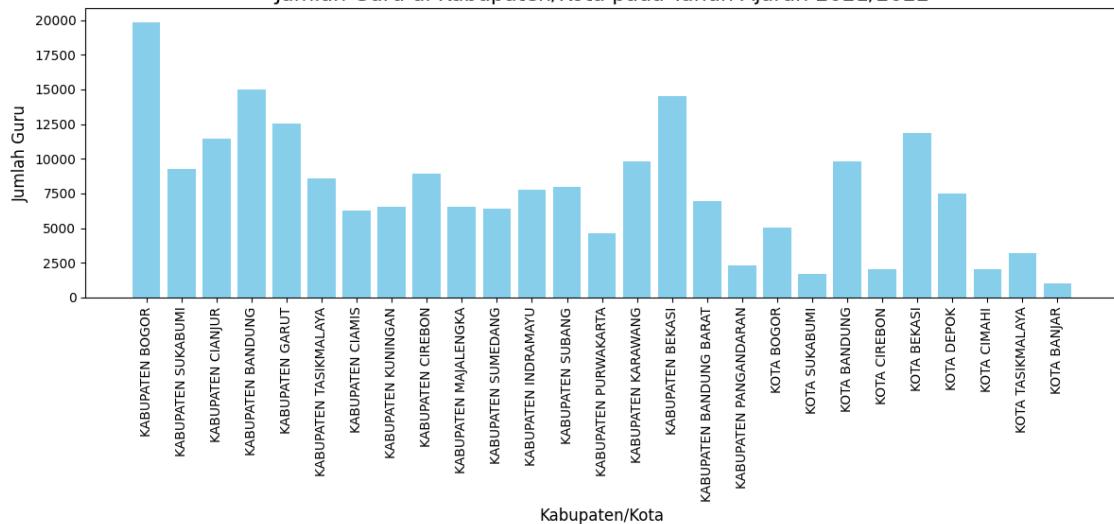


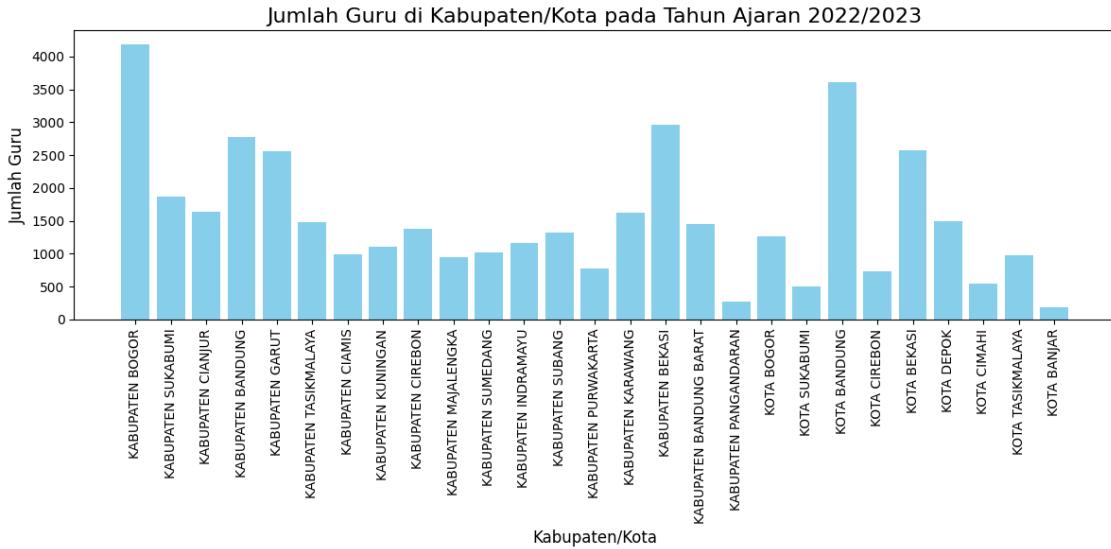


Jumlah Guru di Kabupaten/Kota pada Tahun Ajaran 2020/2021



Jumlah Guru di Kabupaten/Kota pada Tahun Ajaran 2021/2022





Jumlah Guru SMA di Jawa Barat 2015 - 2023

```
[ ]: daerah_list = [
    'KABUPATEN BOGOR', 'KABUPATEN SUKABUMI', 'KABUPATEN CIANJUR',
    'KABUPATEN BANDUNG', 'KABUPATEN GARUT', 'KABUPATEN TASIKMALAYA',
    'KABUPATEN CIAMIS', 'KABUPATEN KUNINGAN', 'KABUPATEN CIREBON',
    'KABUPATEN MAJALENGKA', 'KABUPATEN SUMEDANG', 'KABUPATEN INDRAMAYU',
    'KABUPATEN SUBANG', 'KABUPATEN PURWAKARTA', 'KABUPATEN KARAWANG',
    'KABUPATEN BEKASI', 'KABUPATEN BANDUNG BARAT', 'KABUPATEN PANGANDARAN',
    'KOTA BOGOR', 'KOTA SUKABUMI', 'KOTA BANDUNG', 'KOTA CIREBON',
    'KOTA BEKASI', 'KOTA DEPOK', 'KOTA CIAMHI', 'KOTA TASIKMALAYA', 'KOTA BANJAR'
]

tahun_ajaran_list = ['2015/2016', '2016/2017', '2017/2018', '2018/2019',
                     '2019/2020', '2020/2021', '2021/2022', '2022/2023']

# Perulangan untuk setiap daerah
for daerah in daerah_list:
    # Filter data berdasarkan daerah dan tahun ajaran
    daerah_data = gsma[(gsma['nama_kabupaten_kota'] == daerah) &
                         (gsma['tahun_ajaran'].isin(tahun_ajaran_list))]

    # Jika data untuk daerah tersebut tersedia, buat plot
    if not daerah_data.empty:
        plt.figure(figsize=(10,6))
        plt.bar(daerah_data['tahun_ajaran'], daerah_data['jumlah_guru'],
                color='skyblue')
```

```

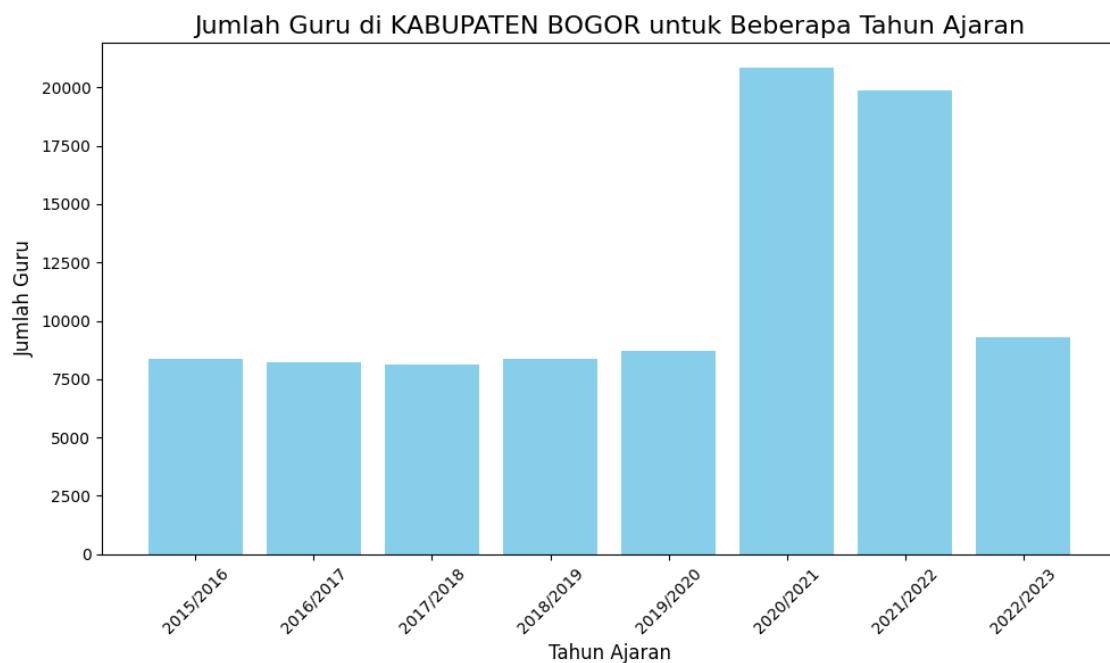
# Tambahkan judul dan label
plt.title(f'Jumlah Guru di {daerah} untuk Beberapa Tahun Ajaran',  

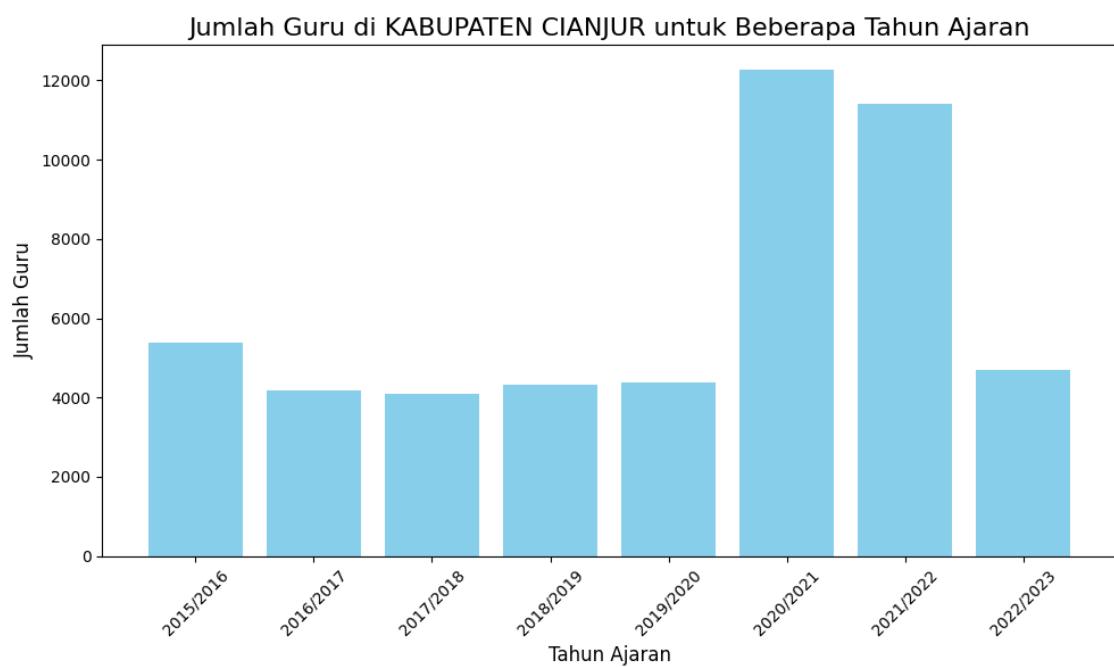
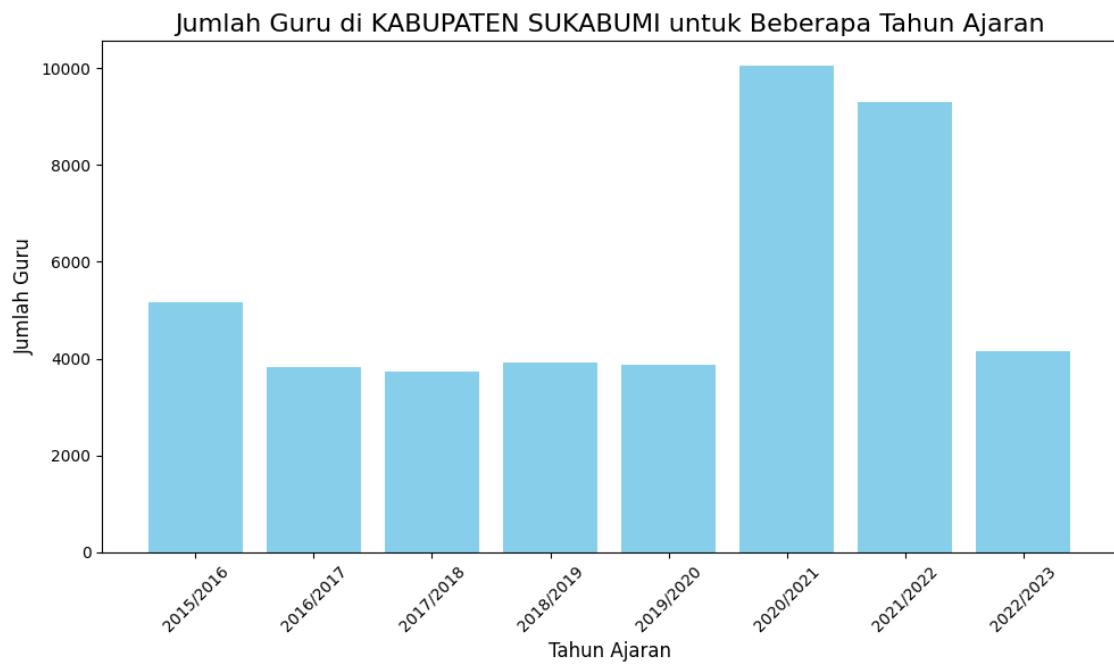
         fontsize=16)
plt.xlabel('Tahun Ajaran', fontsize=12)
plt.ylabel('Jumlah Guru', fontsize=12)
plt.xticks(rotation=45)

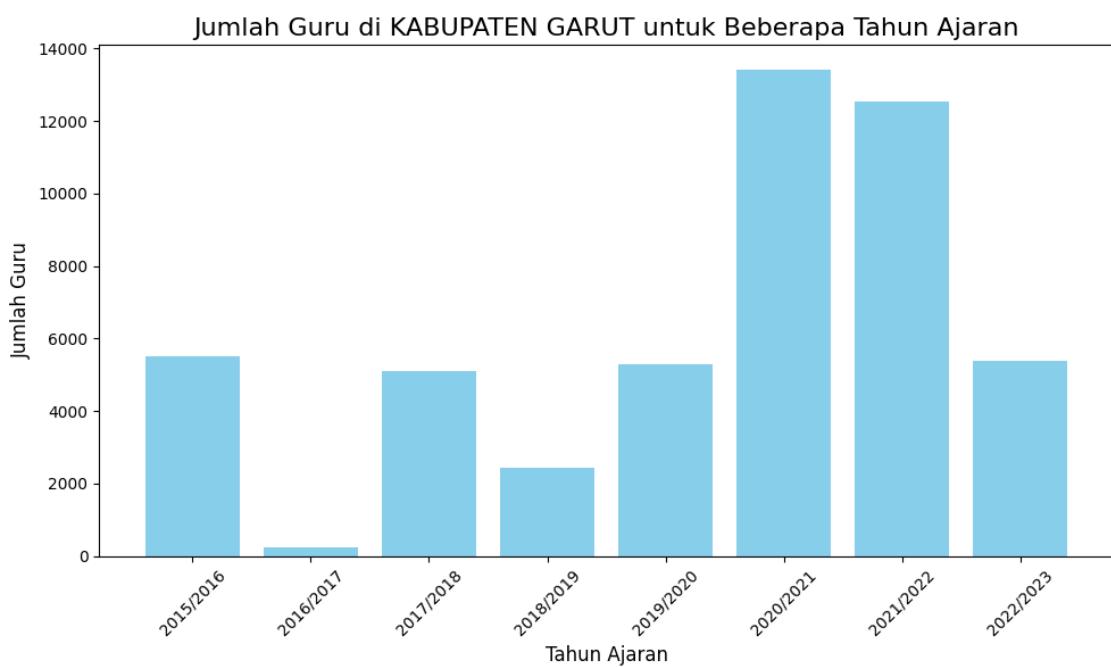
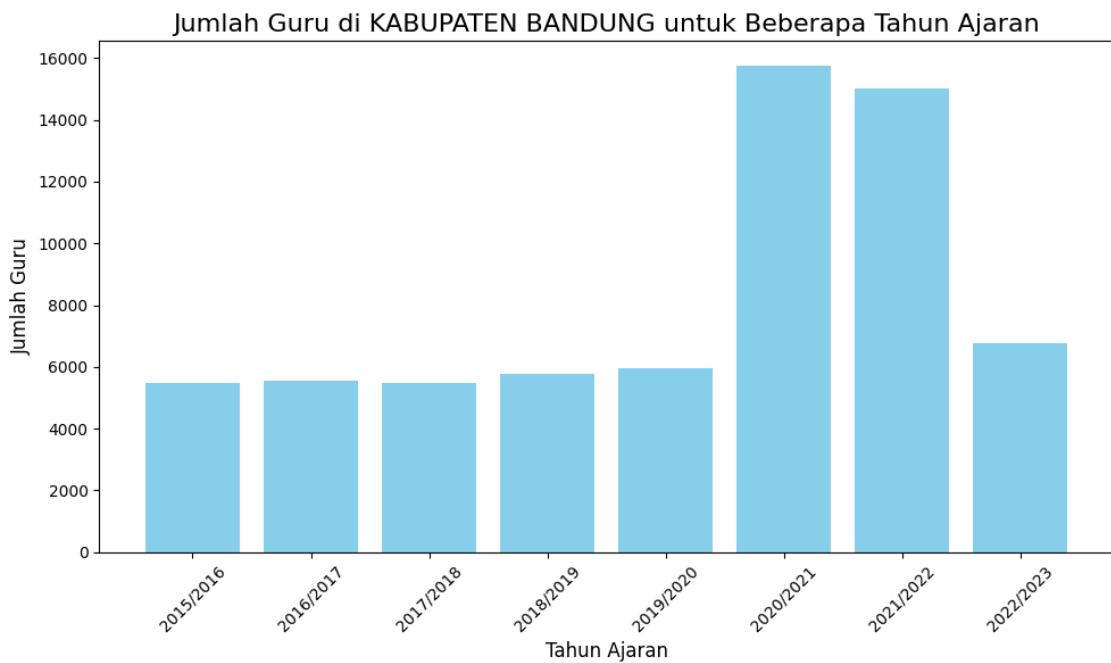
# Tampilkan diagram
plt.tight_layout()
plt.show()

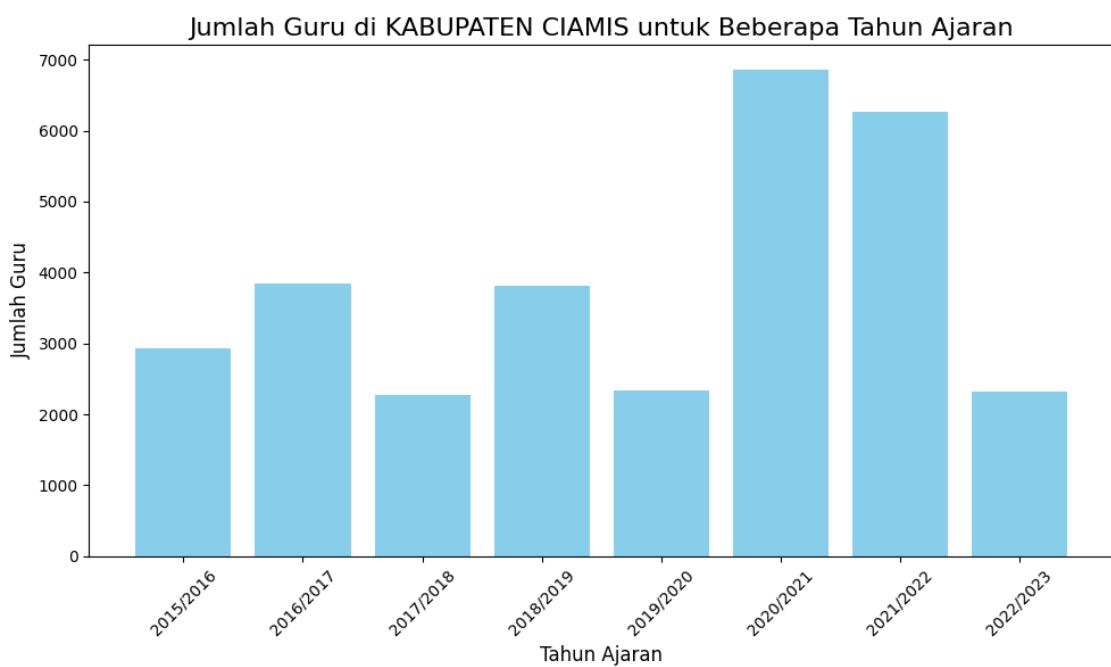
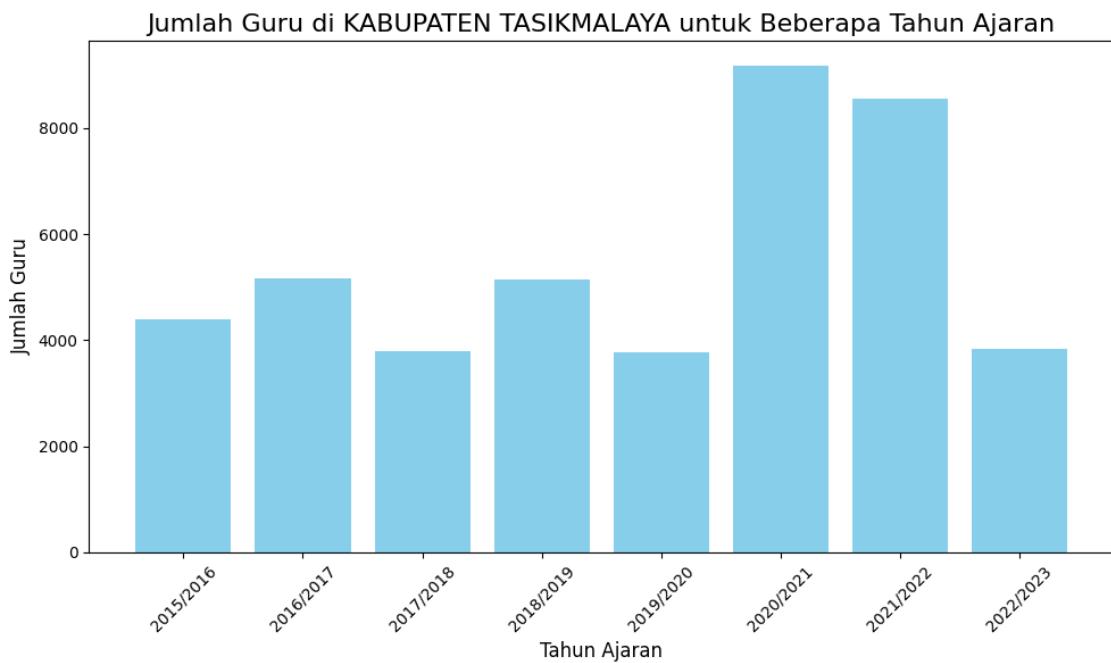
else:
    print(f"Tidak ada data untuk {daerah}")

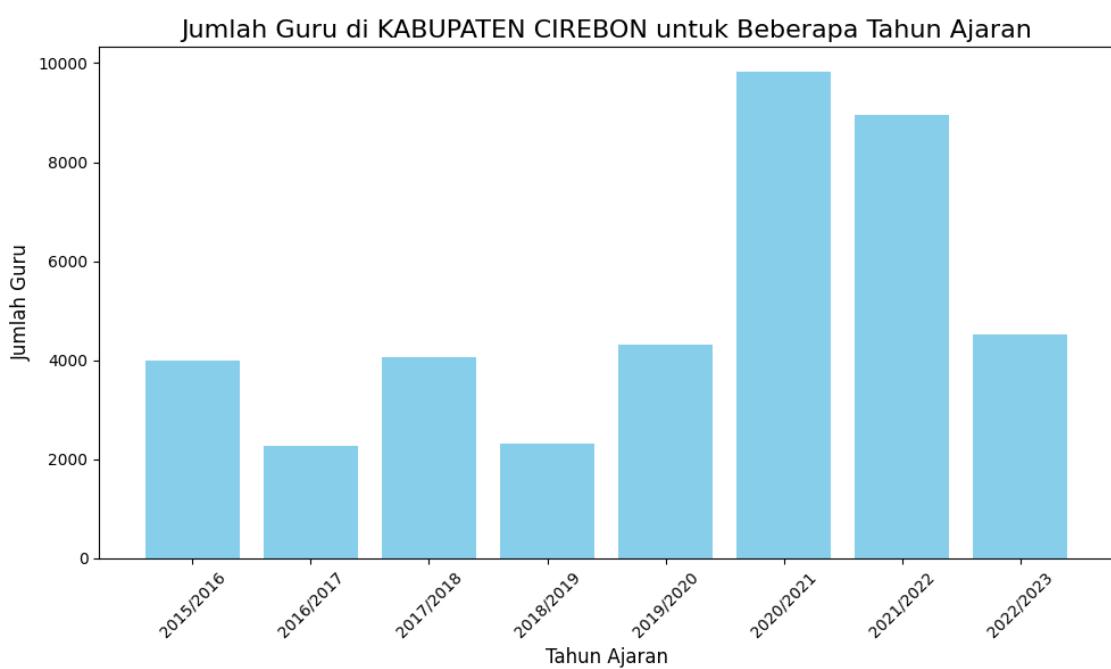
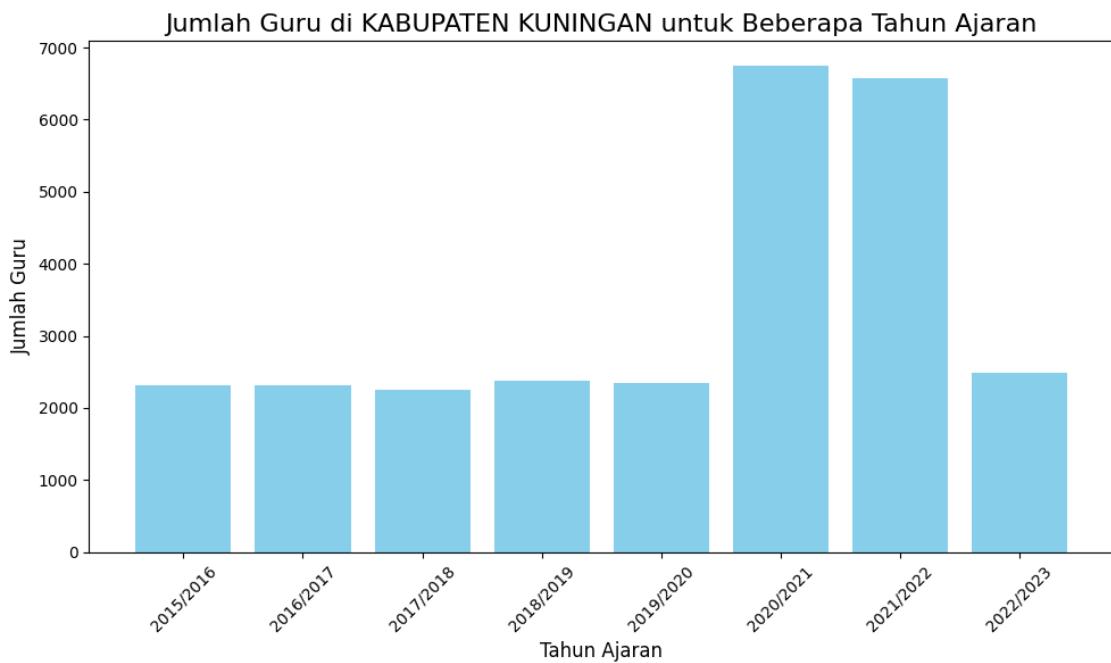
```

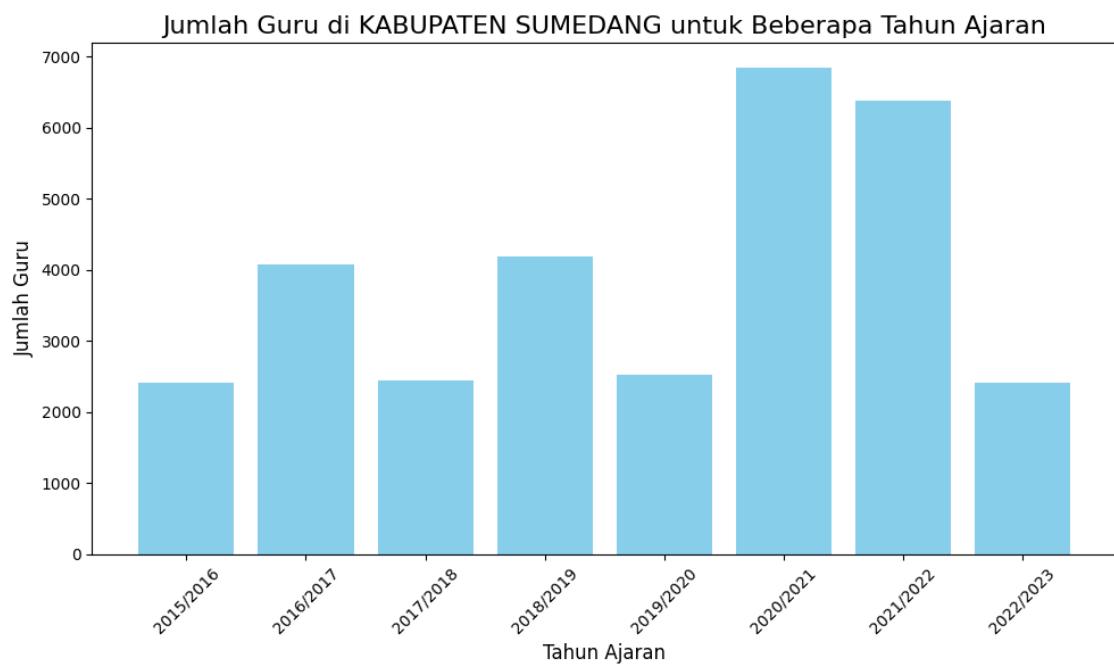
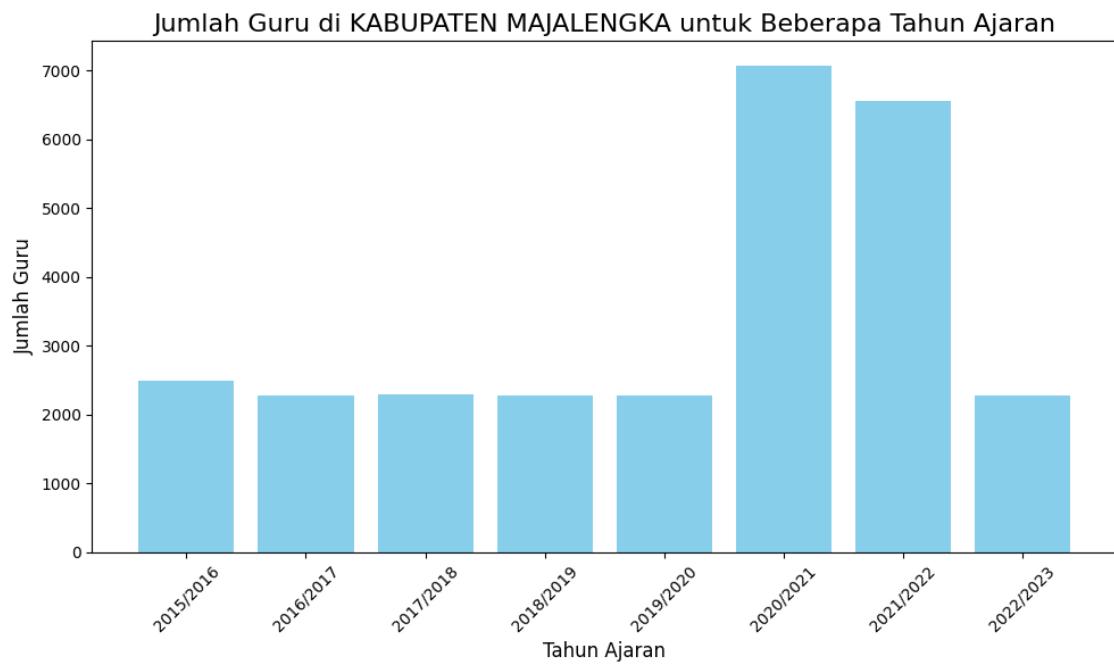


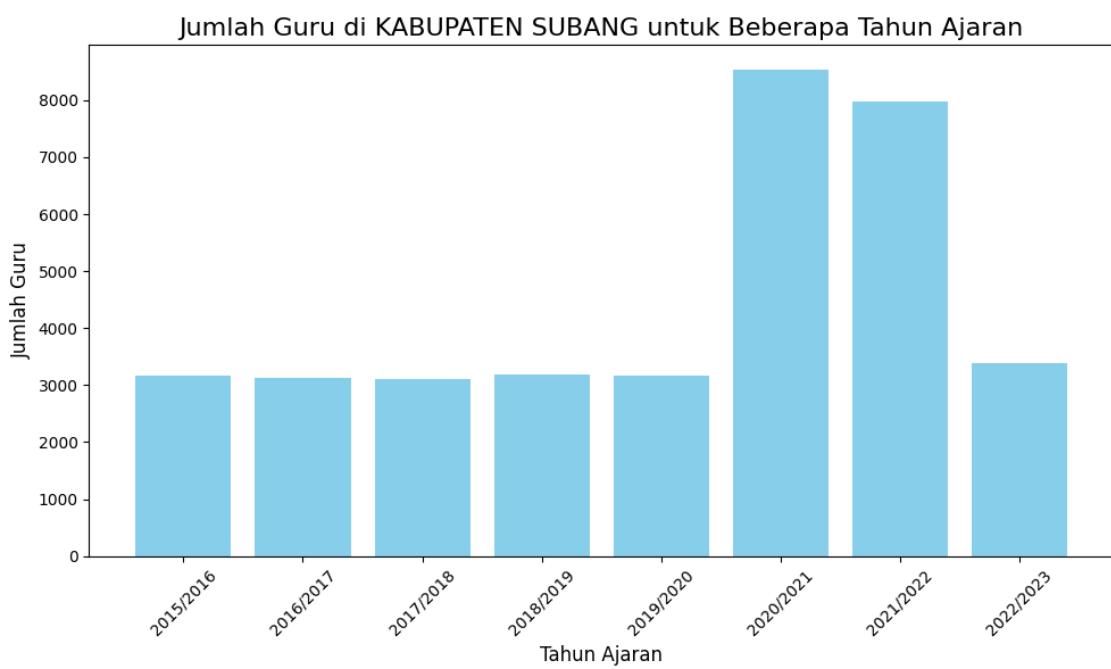
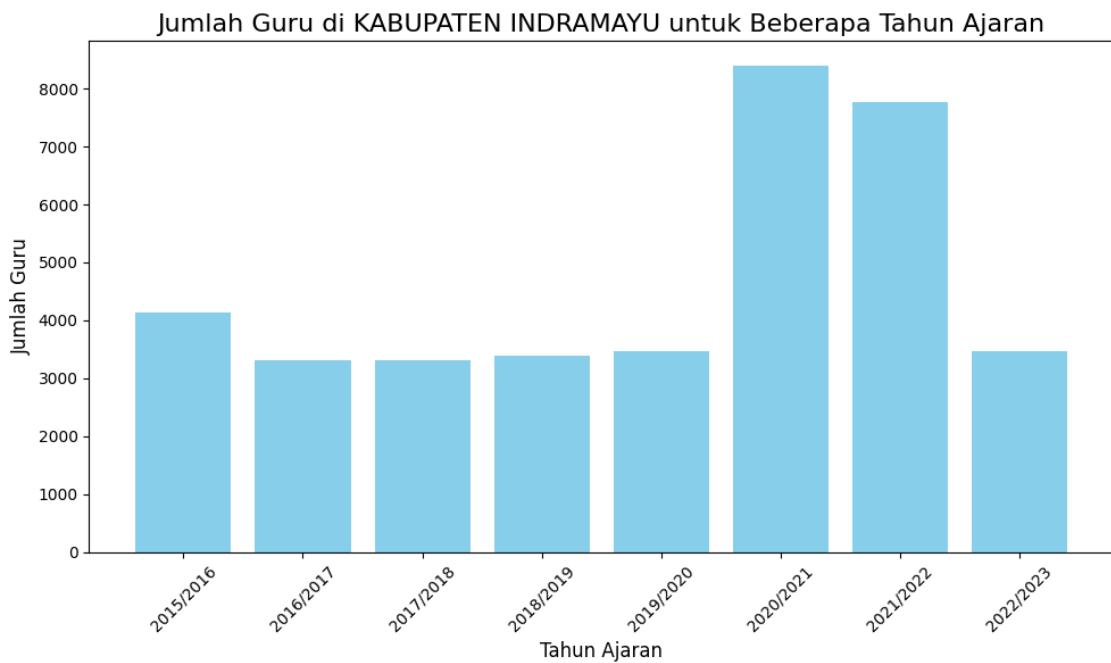


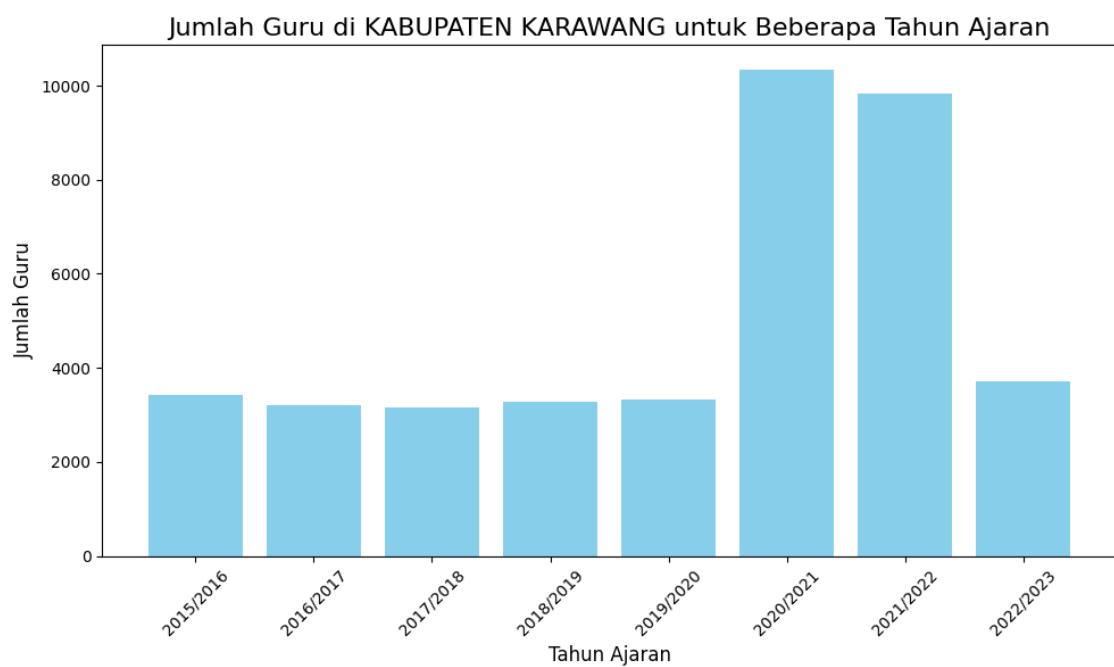
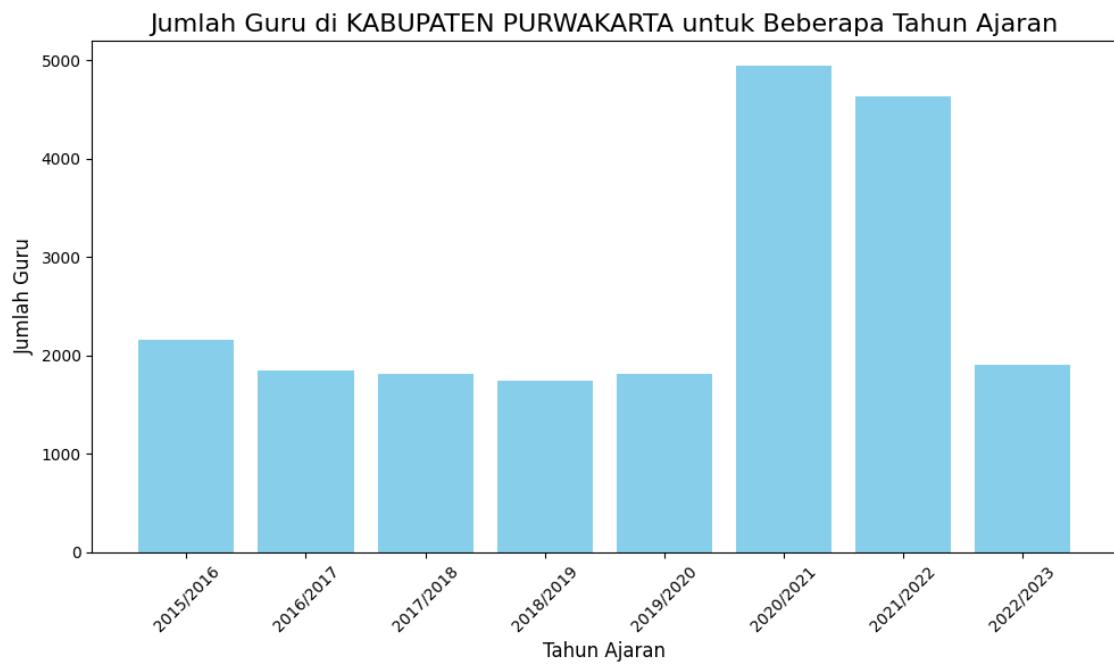


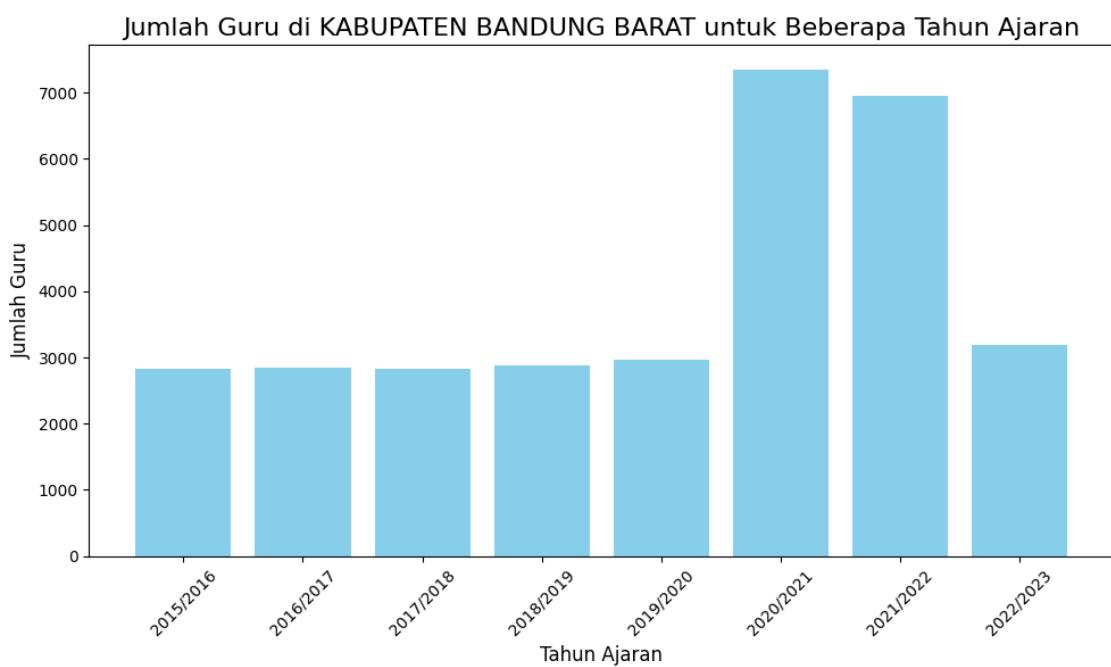
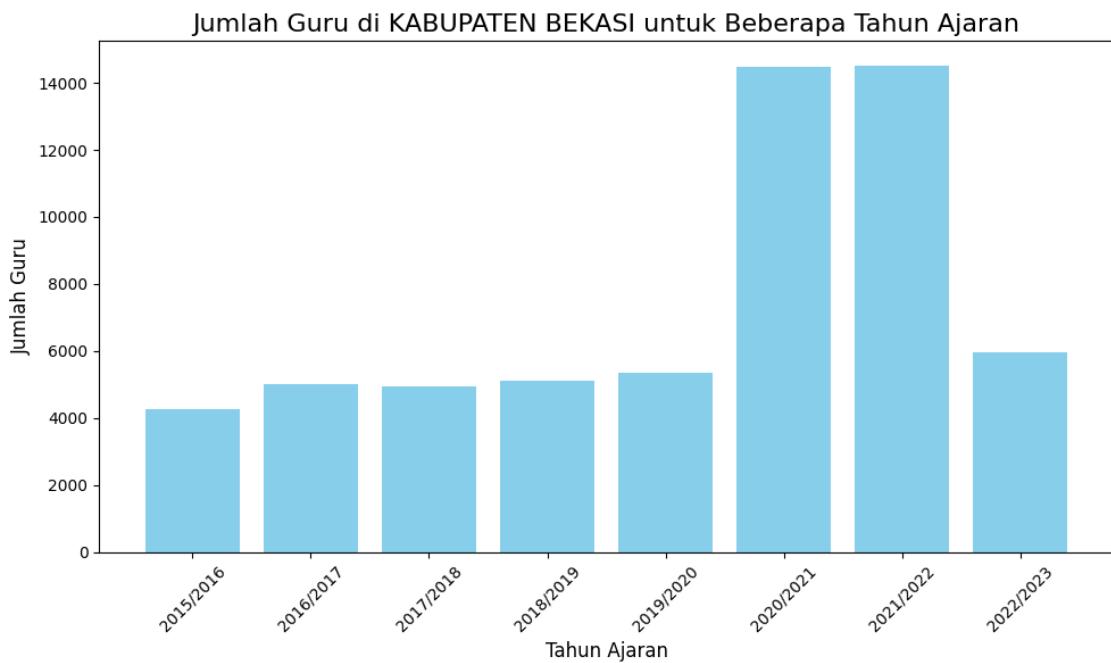


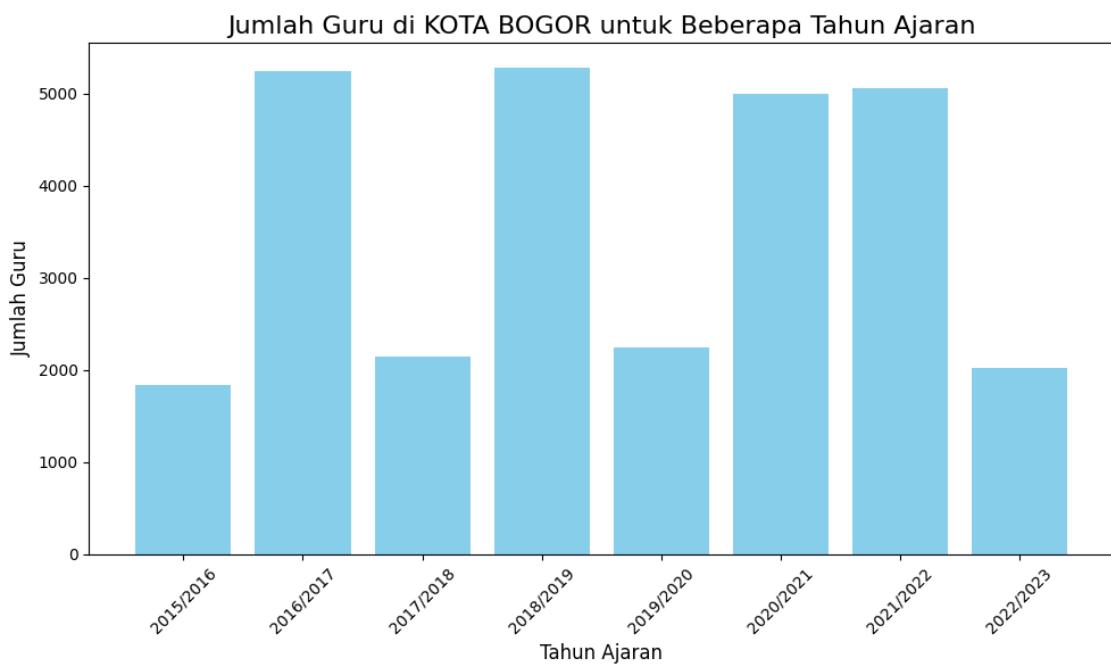
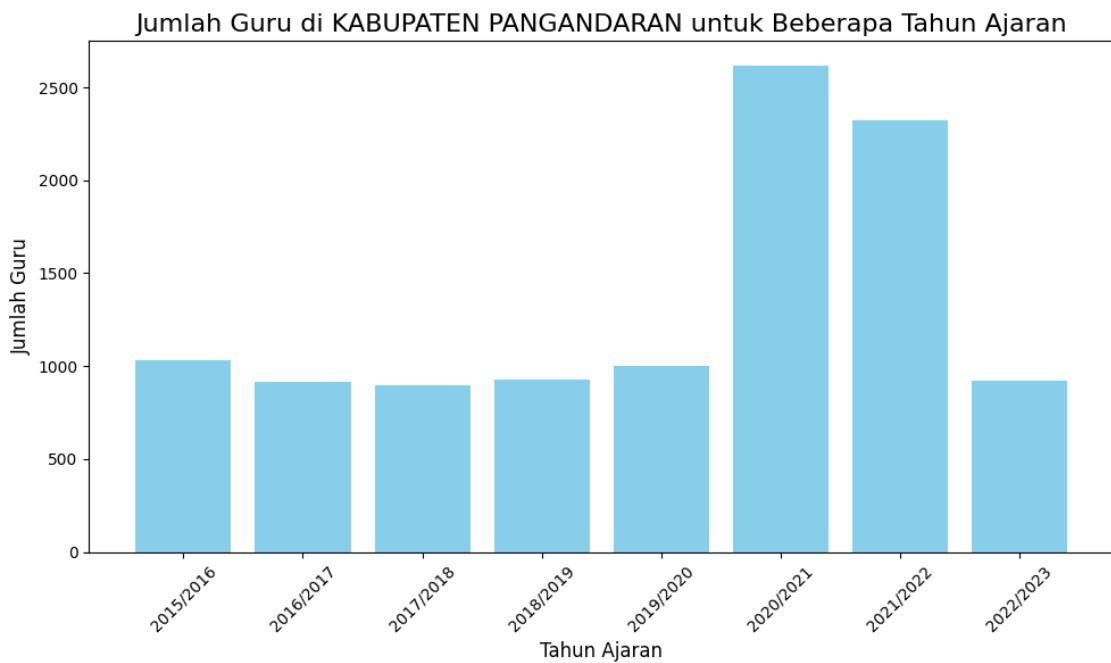




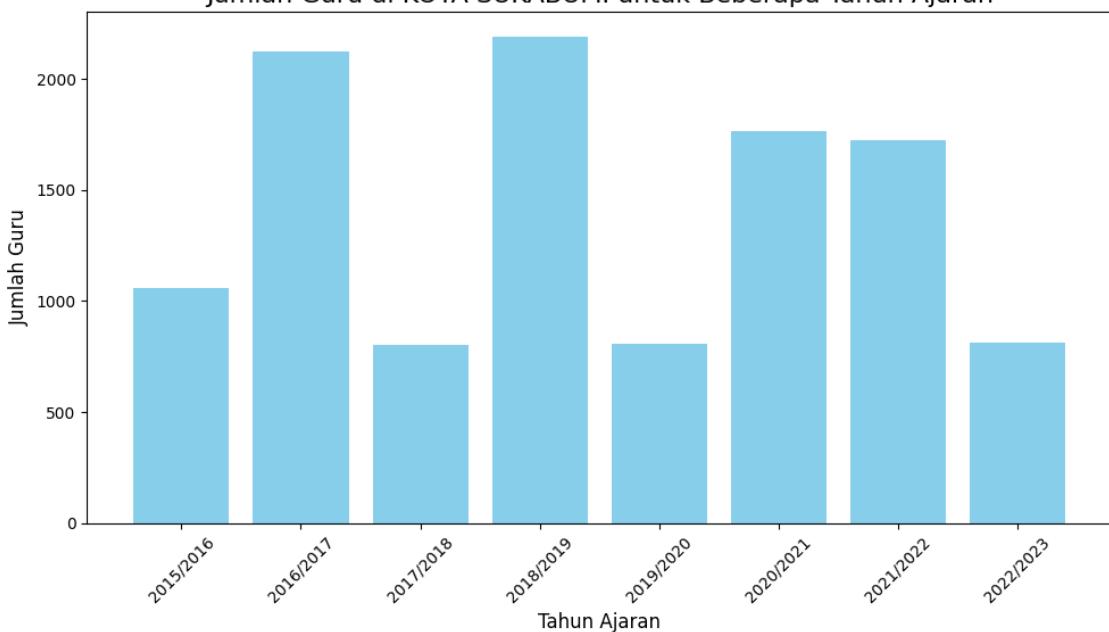




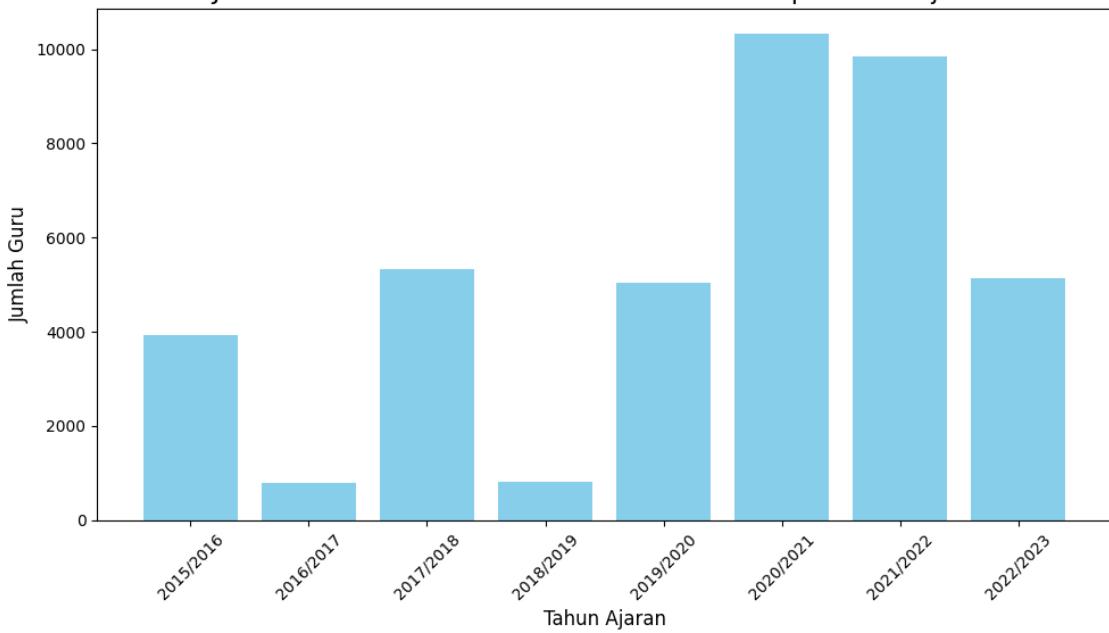




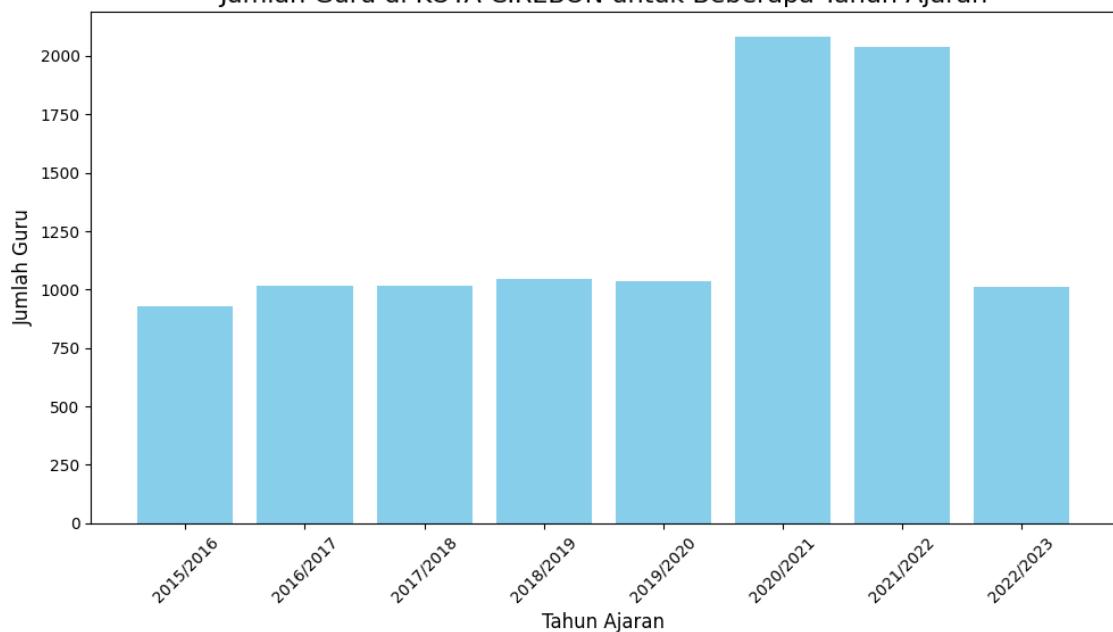
Jumlah Guru di KOTA SUKABUMI untuk Beberapa Tahun Ajaran



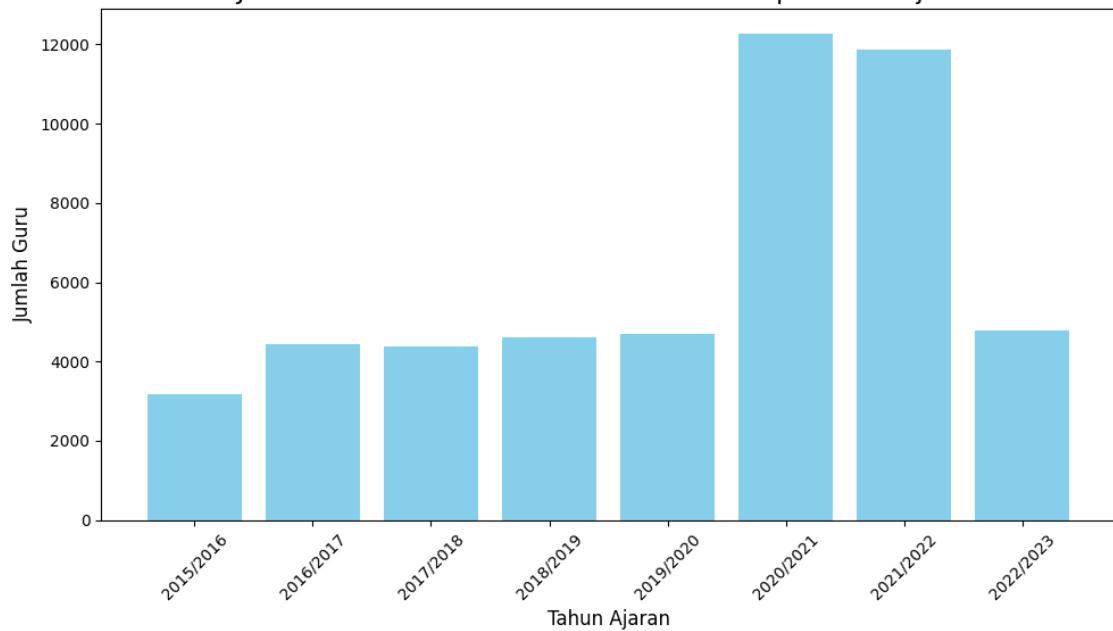
Jumlah Guru di KOTA BANDUNG untuk Beberapa Tahun Ajaran



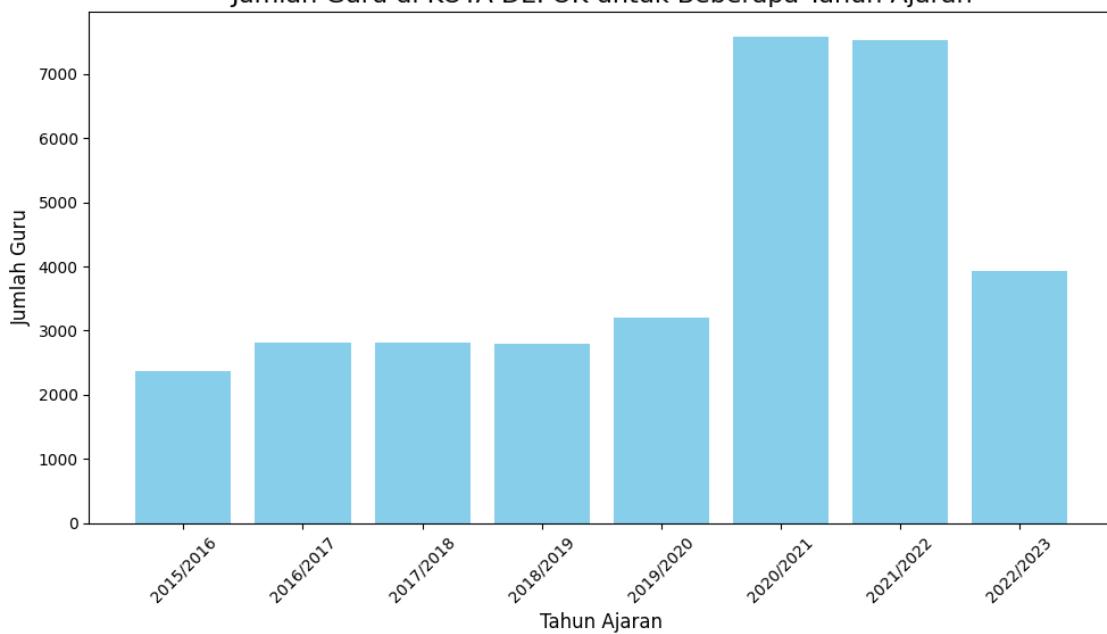
Jumlah Guru di KOTA CIREBON untuk Beberapa Tahun Ajaran



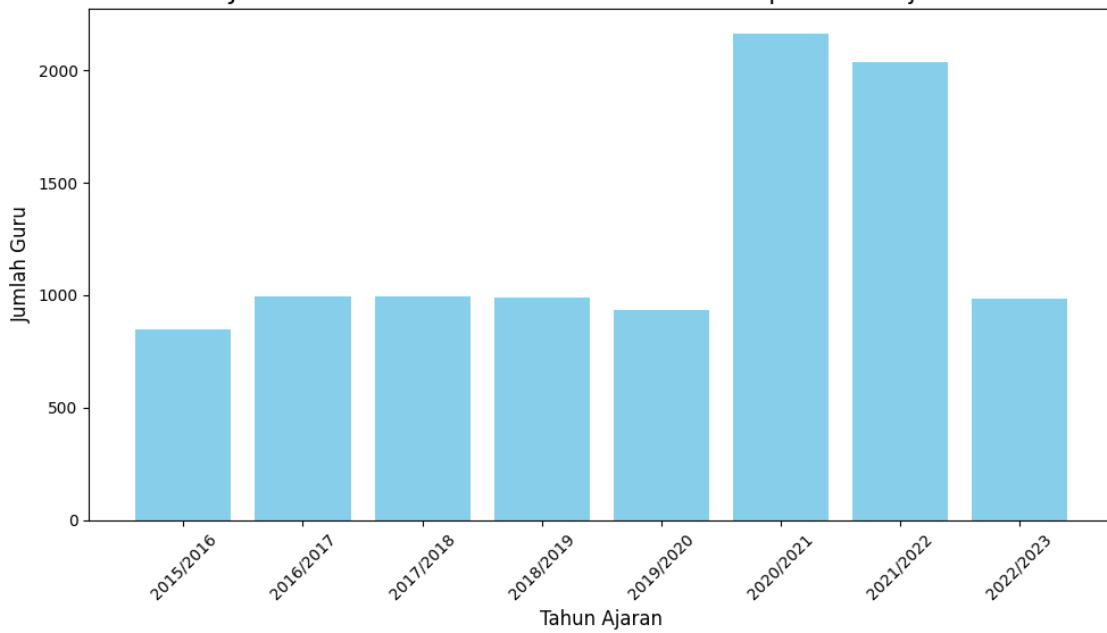
Jumlah Guru di KOTA BEKASI untuk Beberapa Tahun Ajaran

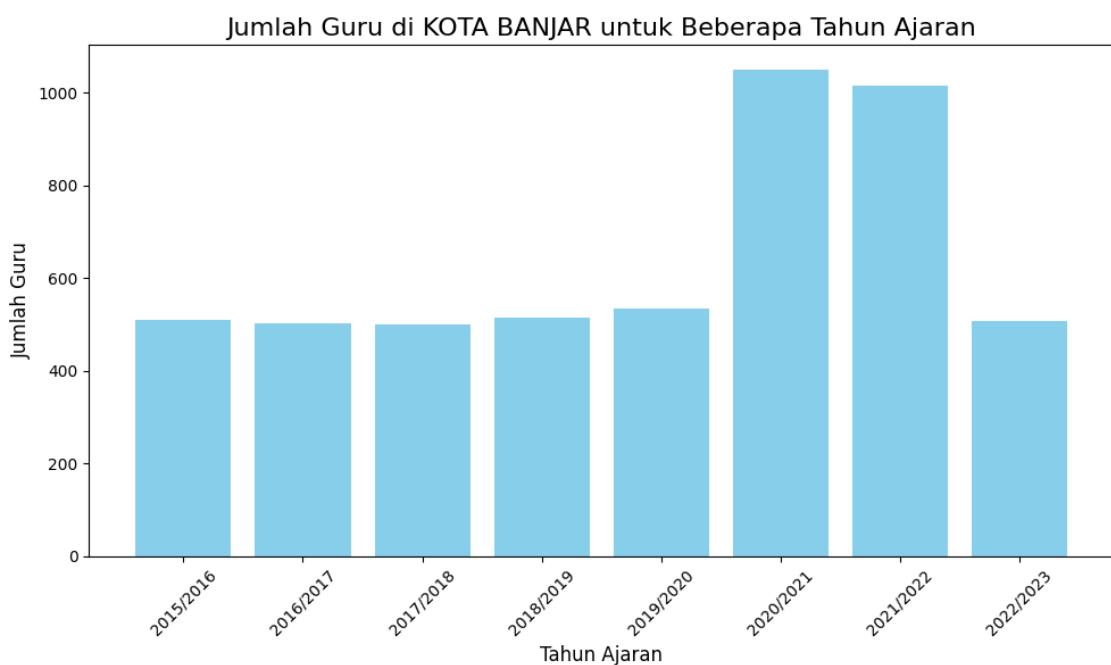
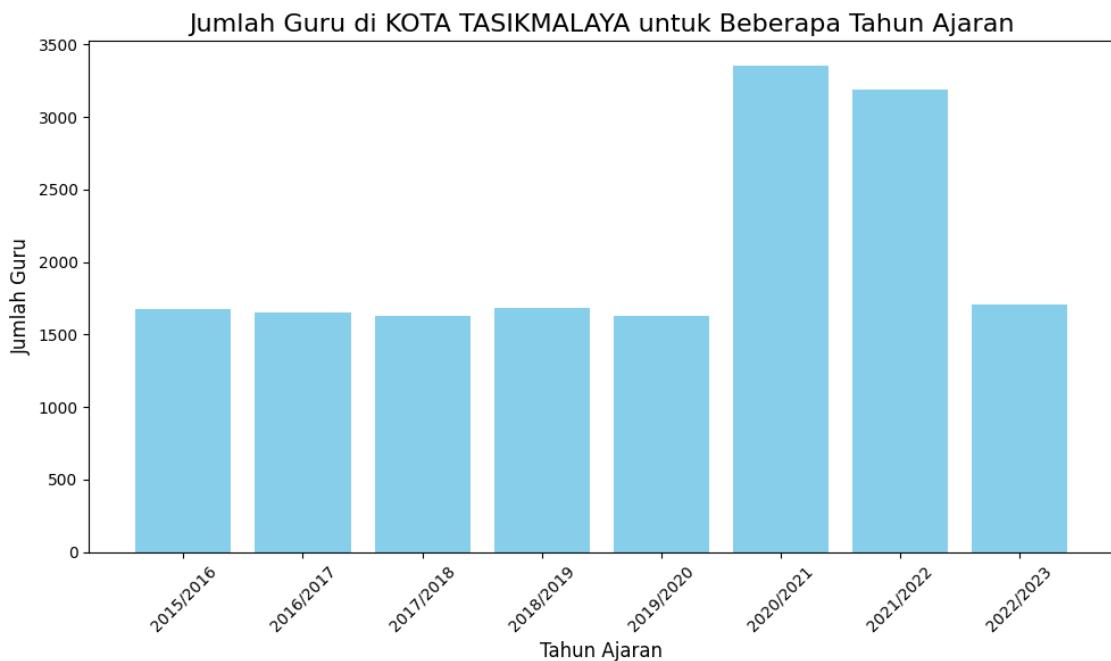


Jumlah Guru di KOTA DEPOK untuk Beberapa Tahun Ajaran



Jumlah Guru di KOTA CIMAHI untuk Beberapa Tahun Ajaran





```
[ ]: # List tahun ajaran dari 2015/2016 hingga 2022/2023
tahun_ajaran_list = ['2015/2016', '2016/2017', '2017/2018', '2018/2019',
                     '2019/2020', '2020/2021', '2021/2022', '2022/2023']
```

```

# Looping untuk menampilkan diagram sebanyak jumlah tahun ajarannya kak
for tahun_ajaran in tahun_ajaran_list:
    # kita filter data, sesuai tahun ajaran
    data_per_tahun = gsma[gsma['tahun_ajaran'] == tahun_ajaran]

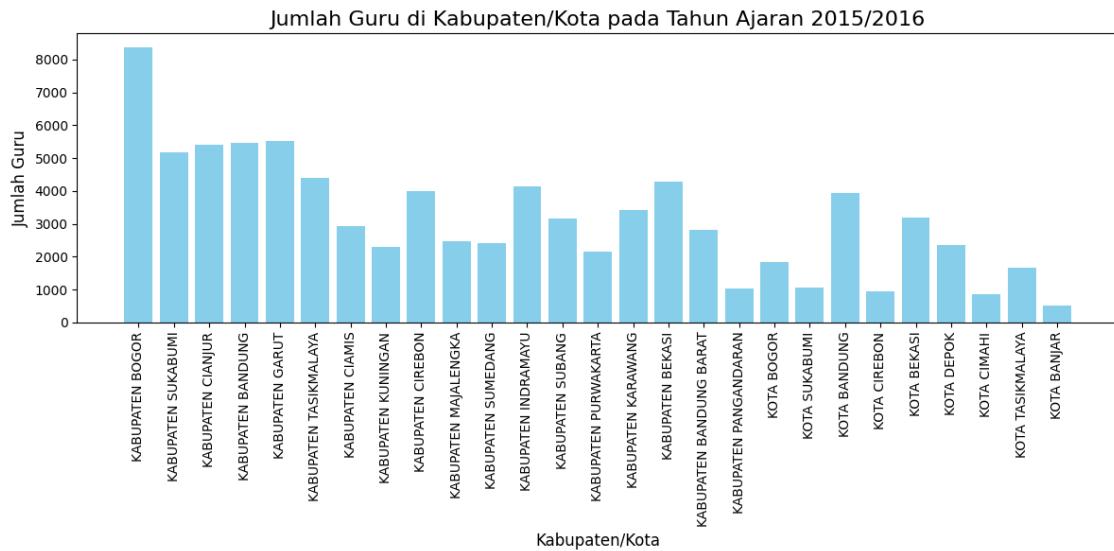
    # Jika data untuk tahun ajaran tersebut tersedia, buat plot
    if not data_per_tahun.empty:
        plt.figure(figsize=(12,6))
        plt.bar(data_per_tahun['nama_kabupaten_kota'], data_per_tahun['jumlah_guru'], color='skyblue')

    # Tambahkan judul dan label pada diagram yang kita buat
    plt.title(f'Jumlah Guru di Kabupaten/Kota pada Tahun Ajaran {tahun_ajaran}', fontsize=16)
    plt.xlabel('Kabupaten/Kota', fontsize=12)
    plt.ylabel('Jumlah Guru', fontsize=12)
    plt.xticks(rotation=90) # Rotasi nama kabupaten/kota agar mudah terbaca

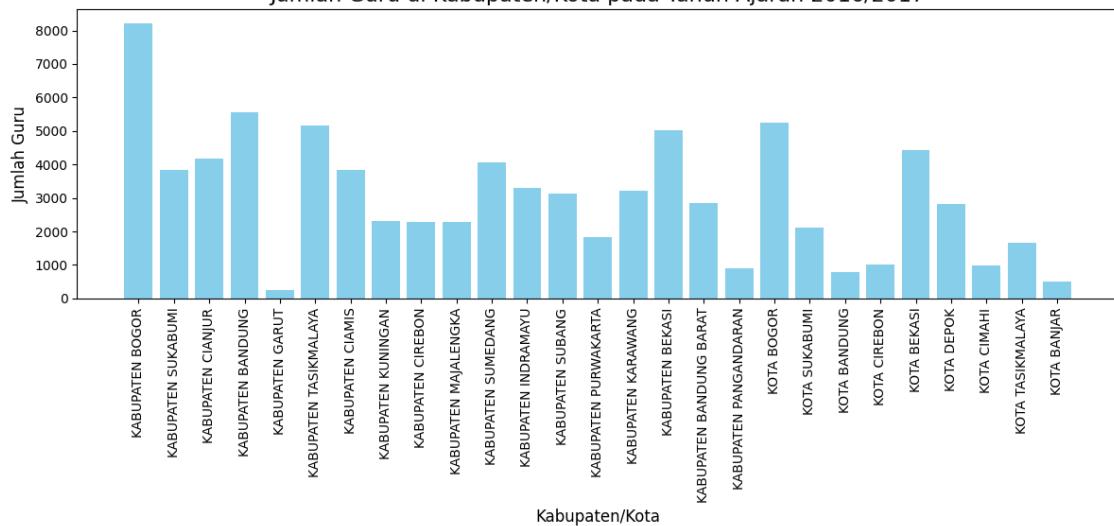
    # Tampilkan diagram
    plt.tight_layout()
    plt.show()

else:
    print(f"Tidak ada data untuk Tahun Ajaran {tahun_ajaran}")

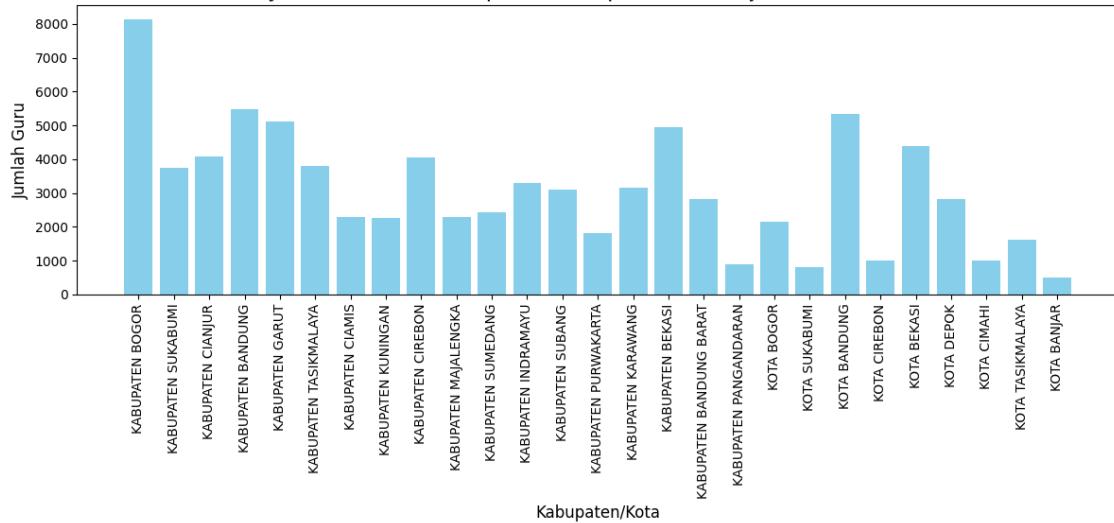
```



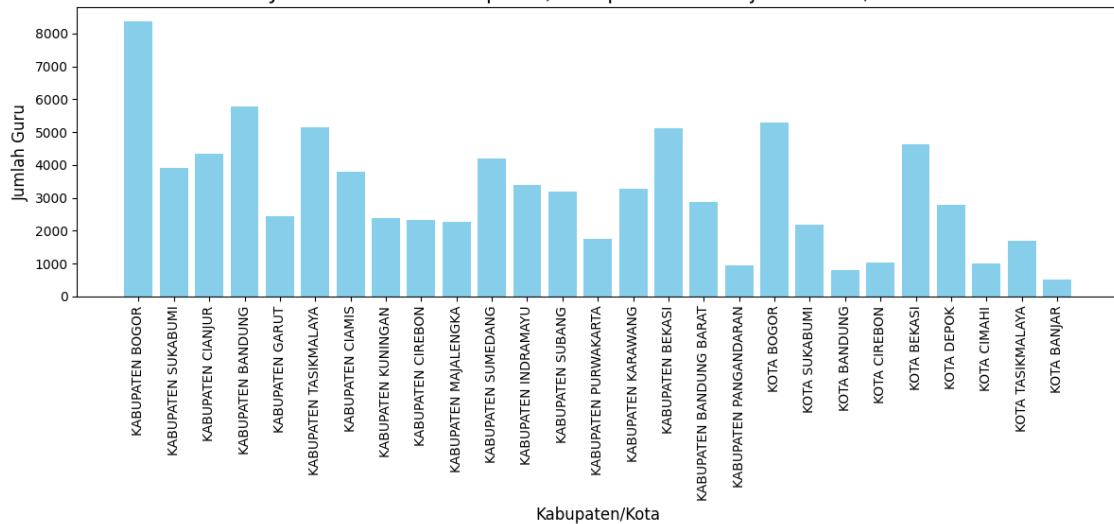
Jumlah Guru di Kabupaten/Kota pada Tahun Ajaran 2016/2017



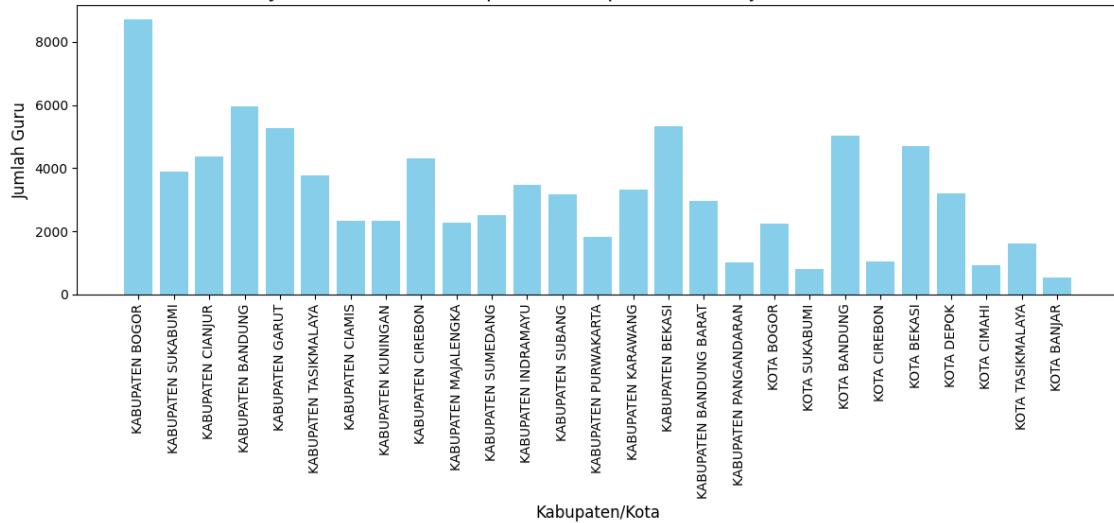
Jumlah Guru di Kabupaten/Kota pada Tahun Ajaran 2017/2018

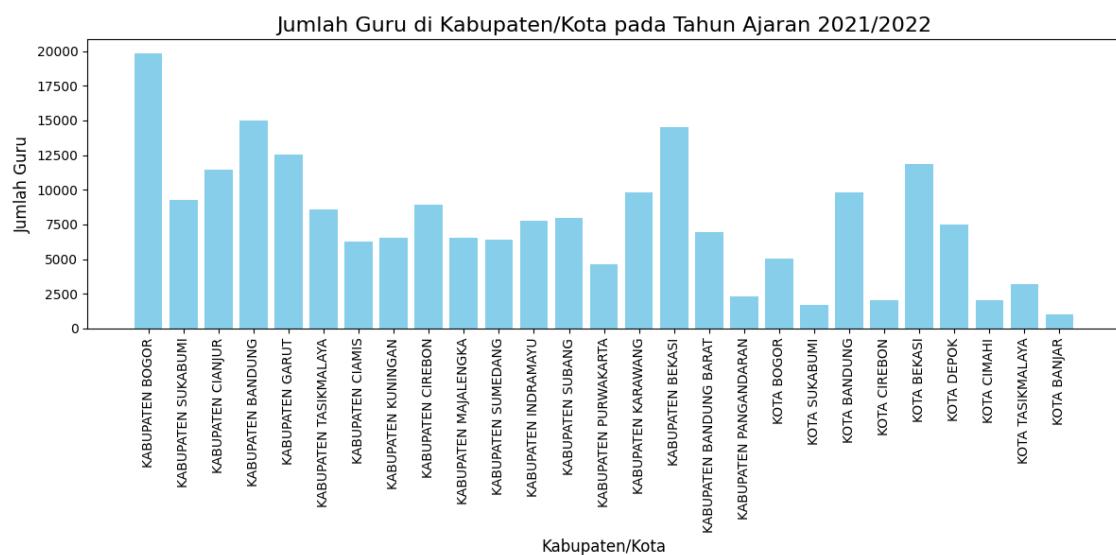
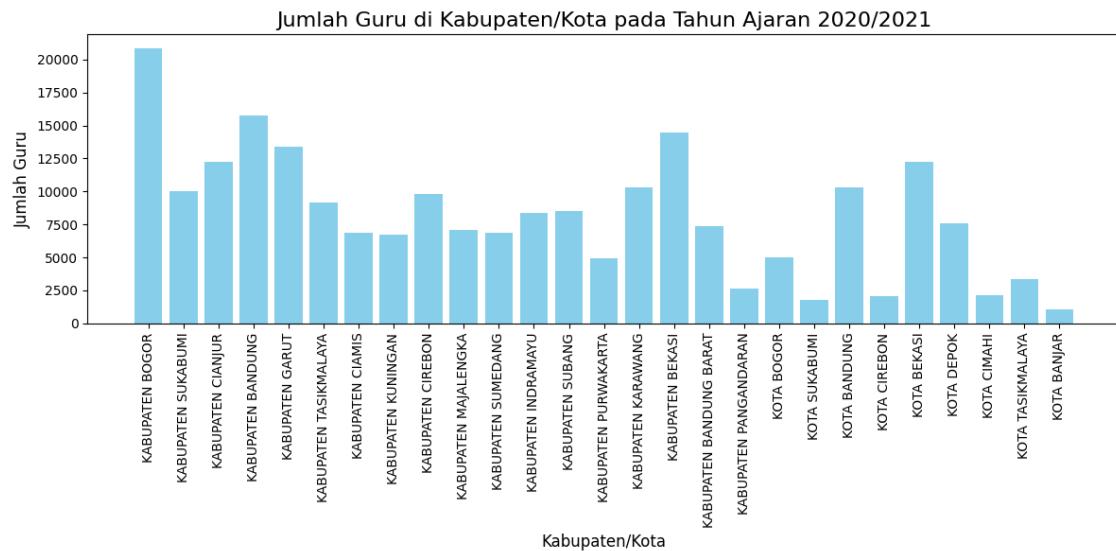


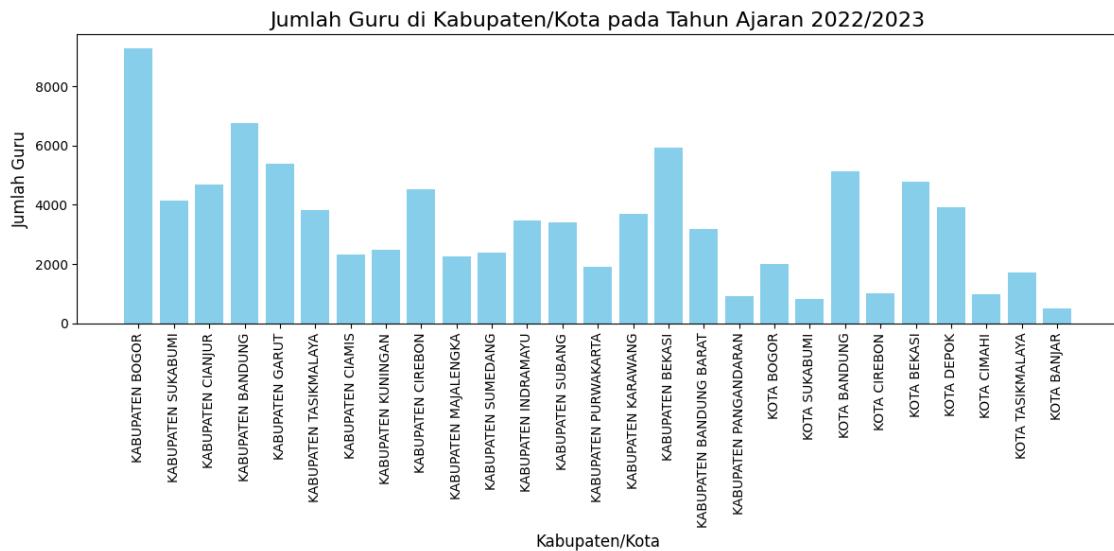
Jumlah Guru di Kabupaten/Kota pada Tahun Ajaran 2018/2019



Jumlah Guru di Kabupaten/Kota pada Tahun Ajaran 2019/2020







Kesimpulan Jumlah Guru SMA 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 perekutan 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 perekutan 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.

Jumlah Guru SMP

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

Melihat Jumlah SMP Negeri berdasarkan Kota/Kabupaten ditampilkan 10 terbanyak

```
[ ]: # Filter data untuk kategori_sekolah 'NEGERI'
smp_negeri = smp[smp['kategori_sekolah'] == 'NEGERI']

# Group by 'nama_kabupaten_kota' dan hitung total jumlah_sekolah di seluruh_
↪tahun ajaran
total_sekolah_negeri = smp_negeri.
↪groupby('nama_kabupaten_kota')['jumlah_sekolah'].sum().reset_index()

# Urutkan berdasarkan jumlah_sekolah terbanyak dan ambil 10 kota/kabupaten_
↪teratas
top_10_kabupaten_kota = total_sekolah_negeri.nlargest(10,_
↪'jumlah_sekolah')['nama_kabupaten_kota']

# Filter kembali data asli untuk hanya menyertakan 10 kota/kabupaten teratas
smp_negeri_top_10 = smp_negeri[smp_negeri['nama_kabupaten_kota'].
↪isin(top_10_kabupaten_kota)]

# Group by 'nama_kabupaten_kota' dan 'tahun_ajaran', lalu hitung total_
↪'jumlah_sekolah'
stats_negeri_top_10 = smp_negeri_top_10.groupby(['nama_kabupaten_kota',_
↪'tahun_ajaran'])['jumlah_sekolah'].sum().reset_index()

# Buat lineplot
plt.figure(figsize=(10, 6))
sns.lineplot(data=stats_negeri_top_10, x='tahun_ajaran', y='jumlah_sekolah',_
↪hue='nama_kabupaten_kota', marker='o')

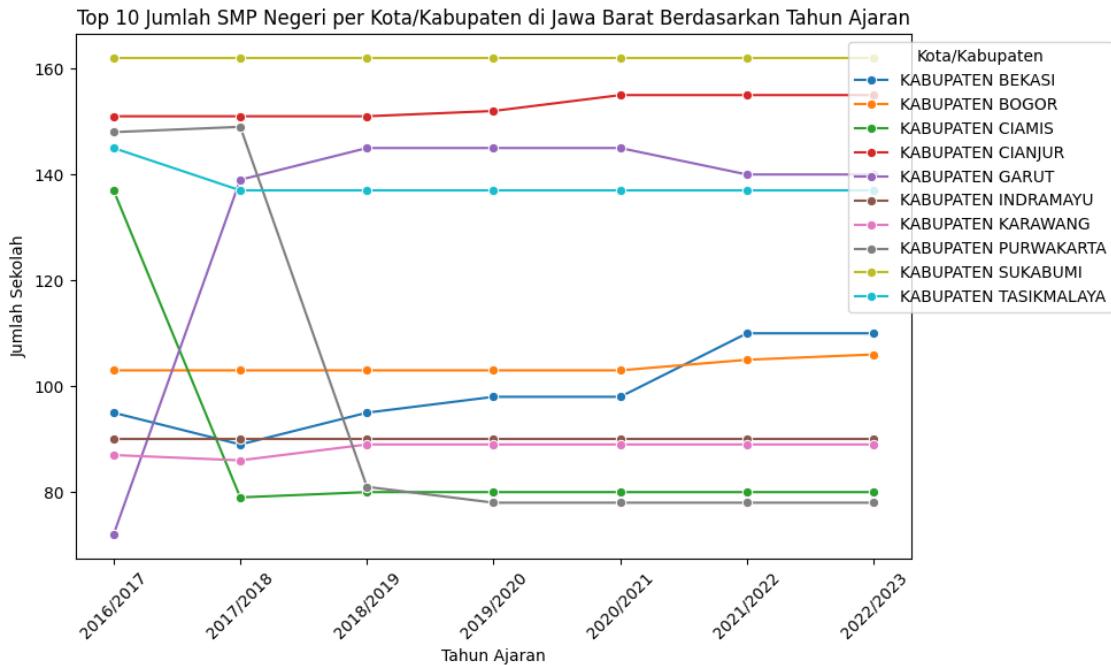
# Menambahkan judul dan label
plt.title('Top 10 Jumlah SMP Negeri per Kota/Kabupaten di Jawa Barat_
↪Berdasarkan Tahun Ajaran')
plt.xlabel('Tahun Ajaran')
plt.ylabel('Jumlah Sekolah')
plt.xticks(rotation=45) # Memutar label tahun ajaran agar lebih mudah dibaca
plt.legend(loc='upper right', bbox_to_anchor=(1.25, 1), title='Kota/Kabupaten')

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

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `^(name,)` instead of `name` to
silence this warning.

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `^(name,)` instead of `name` to

```
silence this warning.  
    data_subset = grouped_data.get_group(pd_key)  
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  
When grouping with a length-1 list-like, you will need to pass a length-1 tuple  
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to  
silence this warning.  
    data_subset = grouped_data.get_group(pd_key)  
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  
When grouping with a length-1 list-like, you will need to pass a length-1 tuple  
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to  
silence this warning.  
    data_subset = grouped_data.get_group(pd_key)  
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  
When grouping with a length-1 list-like, you will need to pass a length-1 tuple  
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to  
silence this warning.  
    data_subset = grouped_data.get_group(pd_key)  
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  
When grouping with a length-1 list-like, you will need to pass a length-1 tuple  
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to  
silence this warning.  
    data_subset = grouped_data.get_group(pd_key)  
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  
When grouping with a length-1 list-like, you will need to pass a length-1 tuple  
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to  
silence this warning.  
    data_subset = grouped_data.get_group(pd_key)  
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  
When grouping with a length-1 list-like, you will need to pass a length-1 tuple  
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to  
silence this warning.  
    data_subset = grouped_data.get_group(pd_key)
```



Melihat 10 Kabupaten/Kota dengan Jumlah SMA Negeri Terbanyak

```
[ ]: # Filter data untuk kategori_sekolah 'NEGERI'
sma_negeri = sma[sma['kategori_sekolah'] == 'NEGERI']

# Group by 'nama_kabupaten_kota' dan hitung total jumlah_sekolah di seluruh
# tahun ajaran
total_sekolah_negeri = sma_negeri.
    ↪groupby('nama_kabupaten_kota')['jumlah_sekolah'].sum().reset_index()

# Urutkan berdasarkan jumlah_sekolah terbanyak dan ambil 10 kota/kabupaten
# teratas
top_10_kabupaten_kota = total_sekolah_negeri.nlargest(10,
    ↪'jumlah_sekolah')['nama_kabupaten_kota']

# Filter kembali data asli untuk hanya menyertakan 10 kota/kabupaten teratas
sma_negeri_top_10 = sma_negeri[sma_negeri['nama_kabupaten_kota'].
    ↪isin(top_10_kabupaten_kota)]

# Group by 'nama_kabupaten_kota' dan 'tahun_ajaran', lalu hitung total
# jumlah_sekolah
stats_negeri_top_10 = sma_negeri_top_10.groupby(['nama_kabupaten_kota',
    ↪'tahun_ajaran'])['jumlah_sekolah'].sum().reset_index()

# Buat lineplot
```

```

plt.figure(figsize=(10, 6))
sns.lineplot(data=stats_negeri_top_10, x='tahun_ajaran', y='jumlah_sekolah',  

             hue='nama_kabupaten_kota', marker='o')

# Menambahkan judul dan label
plt.title('Top 10 Jumlah SMA Negeri per Kota/Kabupaten di Jawa Barat  

           Berdasarkan Tahun Ajaran')
plt.xlabel('Tahun Ajaran')
plt.ylabel('Jumlah Sekolah')
plt.xticks(rotation=45) # Memutar label tahun ajaran agar lebih mudah dibaca
plt.legend(loc='upper right', bbox_to_anchor=(1.25, 1), title='Kota/Kabupaten')

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

```

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.

```
    data_subset = grouped_data.get_group(pd_key)
```

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.

```
    data_subset = grouped_data.get_group(pd_key)
```

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.

```
    data_subset = grouped_data.get_group(pd_key)
```

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.

```
    data_subset = grouped_data.get_group(pd_key)
```

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.

```
    data_subset = grouped_data.get_group(pd_key)
```

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.

```
    data_subset = grouped_data.get_group(pd_key)
```

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:

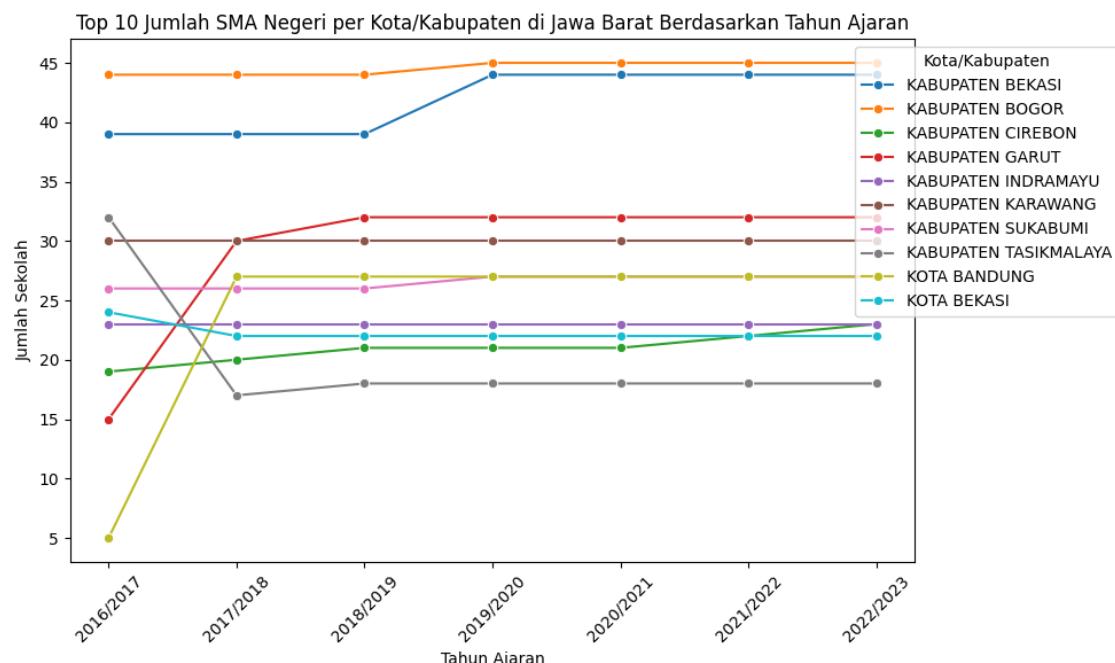
When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

data_subset = grouped_data.get_group(pd_key)
```



Melihat 10 Kota Kabupaten dengan Jumlah SMP Negeri Terendah di Jawa barat

```
[ ]: # Filter data untuk kategori_sekolah 'NEGERI'
smp_negeri = smp[smp['kategori_sekolah'] == 'NEGERI']
```

```

# Group by 'nama_kabupaten_kota' dan hitung total jumlah_sekolah di seluruh
# tahun ajaran
total_sekolah_negeri = smp_negeri.
    ↪groupby('nama_kabupaten_kota')['jumlah_sekolah'].sum().reset_index()

# Urutkan berdasarkan jumlah_sekolah terbanyak dan ambil 10 kota/kabupaten
# teratas
top_10_kabupaten_kota = total_sekolah_negeri.nsmallest(10,
    ↪'jumlah_sekolah')['nama_kabupaten_kota']

# Filter kembali data asli untuk hanya menyertakan 10 kota/kabupaten teratas
smp_negeri_top_10 = smp_negeri[smp_negeri['nama_kabupaten_kota'].
    ↪isin(top_10_kabupaten_kota)]

# Group by 'nama_kabupaten_kota' dan 'tahun_ajaran', lalu hitung total
# 'jumlah_sekolah'
stats_negeri_top_10 = smp_negeri_top_10.groupby(['nama_kabupaten_kota',
    ↪'tahun_ajaran'])['jumlah_sekolah'].sum().reset_index()

# Buat lineplot
plt.figure(figsize=(10, 6))
sns.lineplot(data=stats_negeri_top_10, x='tahun_ajaran', y='jumlah_sekolah',
    ↪hue='nama_kabupaten_kota', marker='o')

# Menambahkan judul dan label
plt.title('10 Kabupaten dengan Jumlah SMP Negeri Terendah di Jawa Barat
    ↪Berdasarkan Tahun Ajaran')
plt.xlabel('Tahun Ajaran')
plt.ylabel('Jumlah Sekolah')
plt.xticks(rotation=45) # Memutar label tahun ajaran agar lebih mudah dibaca
plt.legend(loc='upper right', bbox_to_anchor=(1.25, 1), title='Kota/Kabupaten')

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

```

```

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `^(name,)` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `^(name,)` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:

```

When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

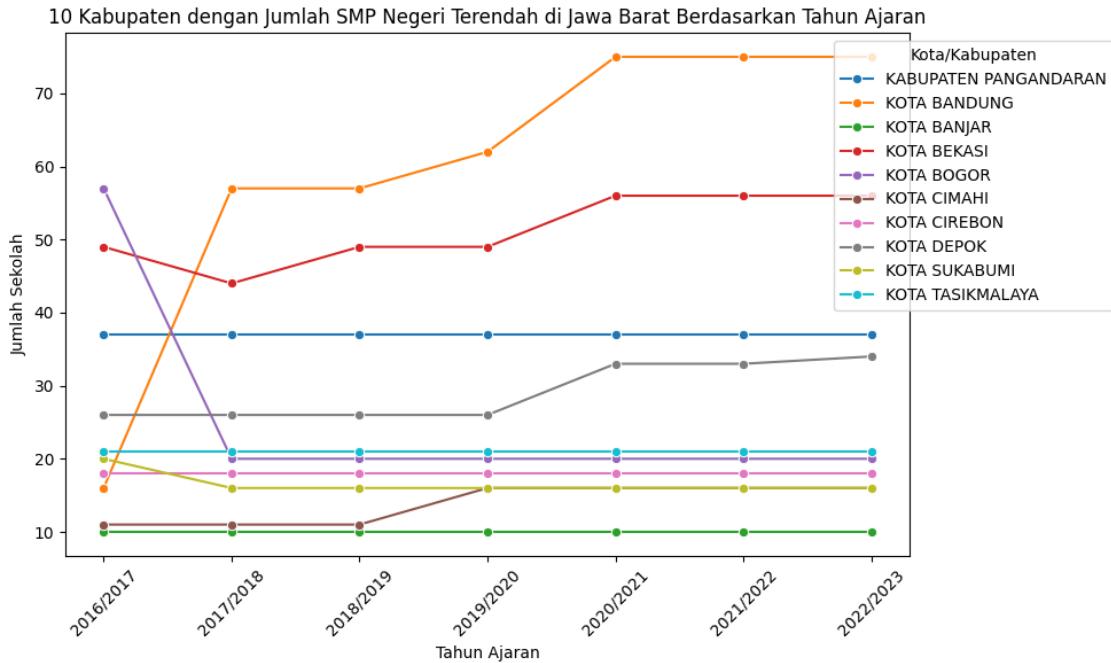
data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

data_subset = grouped_data.get_group(pd_key)
```



Melihat 10 Kabupaten/Kota dengan Jumlah SMA Negeri Terendah di Jawa Barat

```
[ ]: # Filter data untuk kategori_sekolah 'NEGERI'
sma_negeri = sma[sma['kategori_sekolah'] == 'NEGERI']

# Group by 'nama_kabupaten_kota' dan hitung total jumlah_sekolah di seluruh
# tahun ajaran
total_sekolah_negeri = sma_negeri.
    ↪groupby('nama_kabupaten_kota')['jumlah_sekolah'].sum().reset_index()

# Urutkan berdasarkan jumlah_sekolah terbanyak dan ambil 10 kota/kabupaten
# teratas
top_10_kabupaten_kota = total_sekolah_negeri.nsmallest(10,
    ↪'jumlah_sekolah')['nama_kabupaten_kota']

# Filter kembali data asli untuk hanya menyertakan 10 kota/kabupaten teratas
sma_negeri_top_10 = sma_negeri[sma_negeri['nama_kabupaten_kota'].
    ↪isin(top_10_kabupaten_kota)]

# Group by 'nama_kabupaten_kota' dan 'tahun_ajaran', lalu hitung total
# jumlah_sekolah
stats_negeri_top_10 = sma_negeri_top_10.groupby(['nama_kabupaten_kota',
    ↪'tahun_ajaran'])['jumlah_sekolah'].sum().reset_index()

# Buat lineplot
```

```

plt.figure(figsize=(10, 6))
sns.lineplot(data=stats_negeri_top_10, x='tahun_ajaran', y='jumlah_sekolah',  

             hue='nama_kabupaten_kota', marker='o')

# Menambahkan judul dan label
plt.title('10 Kabupaten dengan Jumlah SMA Negeri Terendah di Jawa Barat  

           Berdasarkan Tahun Ajaran')
plt.xlabel('Tahun Ajaran')
plt.ylabel('Jumlah Sekolah')
plt.xticks(rotation=45) # Memutar label tahun ajaran agar lebih mudah dibaca
plt.legend(loc='upper right', bbox_to_anchor=(1.25, 1), title='Kota/Kabupaten')

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

```

```

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  

When grouping with a length-1 list-like, you will need to pass a length-1 tuple  

to get_group in a future version of pandas. Pass `*(name,)` instead of `name` to  

silence this warning.  

    data_subset = grouped_data.get_group(pd_key)  

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  

When grouping with a length-1 list-like, you will need to pass a length-1 tuple  

to get_group in a future version of pandas. Pass `*(name,)` instead of `name` to  

silence this warning.  

    data_subset = grouped_data.get_group(pd_key)  

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  

When grouping with a length-1 list-like, you will need to pass a length-1 tuple  

to get_group in a future version of pandas. Pass `*(name,)` instead of `name` to  

silence this warning.  

    data_subset = grouped_data.get_group(pd_key)  

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  

When grouping with a length-1 list-like, you will need to pass a length-1 tuple  

to get_group in a future version of pandas. Pass `*(name,)` instead of `name` to  

silence this warning.  

    data_subset = grouped_data.get_group(pd_key)  

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  

When grouping with a length-1 list-like, you will need to pass a length-1 tuple  

to get_group in a future version of pandas. Pass `*(name,)` instead of `name` to  

silence this warning.  

    data_subset = grouped_data.get_group(pd_key)

```

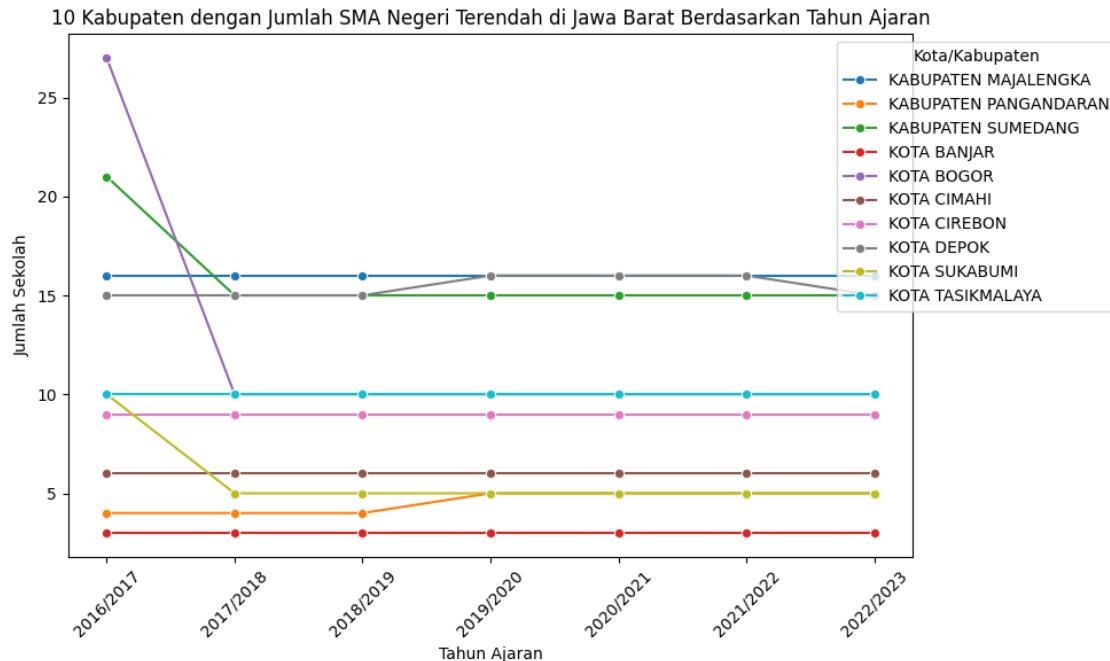
When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.
```

```
data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.
```

```
data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.
```

```
data_subset = grouped_data.get_group(pd_key)
```



Melihat 10 Kota/Kabupaten dengan Jumlah SMP Swasta terbanyak di Jawa Barat

```
[ ]: # Filter data untuk kategori_sekolah 'NEGERI'
smp_negeri = smp[smp['kategori_sekolah'] == 'SWASTA']
```

```

# Group by 'nama_kabupaten_kota' dan hitung total jumlah_sekolah di seluruh
# tahun ajaran
total_sekolah_negeri = smp_negeri.
    ↪groupby('nama_kabupaten_kota')['jumlah_sekolah'].sum().reset_index()

# Urutkan berdasarkan jumlah_sekolah terbanyak dan ambil 10 kota/kabupaten
# teratas
top_10_kabupaten_kota = total_sekolah_negeri.nlargest(10,
    ↪'jumlah_sekolah')['nama_kabupaten_kota']

# Filter kembali data asli untuk hanya menyertakan 10 kota/kabupaten teratas
smp_negeri_top_10 = smp_negeri[smp_negeri['nama_kabupaten_kota'].
    ↪isin(top_10_kabupaten_kota)]

# Group by 'nama_kabupaten_kota' dan 'tahun_ajaran', lalu hitung total
# 'jumlah_sekolah'
stats_negeri_top_10 = smp_negeri_top_10.groupby(['nama_kabupaten_kota',
    ↪'tahun_ajaran'])['jumlah_sekolah'].sum().reset_index()

# Buat lineplot
plt.figure(figsize=(10, 6))
sns.lineplot(data=stats_negeri_top_10, x='tahun_ajaran', y='jumlah_sekolah',
    ↪hue='nama_kabupaten_kota', marker='o')

# Menambahkan judul dan label
plt.title('Top 10 Jumlah SMP Swasta per Kota/Kabupaten di Jawa Barat
    ↪Berdasarkan Tahun Ajaran')
plt.xlabel('Tahun Ajaran')
plt.ylabel('Jumlah Sekolah')
plt.xticks(rotation=45) # Memutar label tahun ajaran agar lebih mudah dibaca
plt.legend(loc='upper right', bbox_to_anchor=(1.25, 1), title='Kota/Kabupaten')

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

```

```

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `^(name,)` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `^(name,)` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:

```

When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

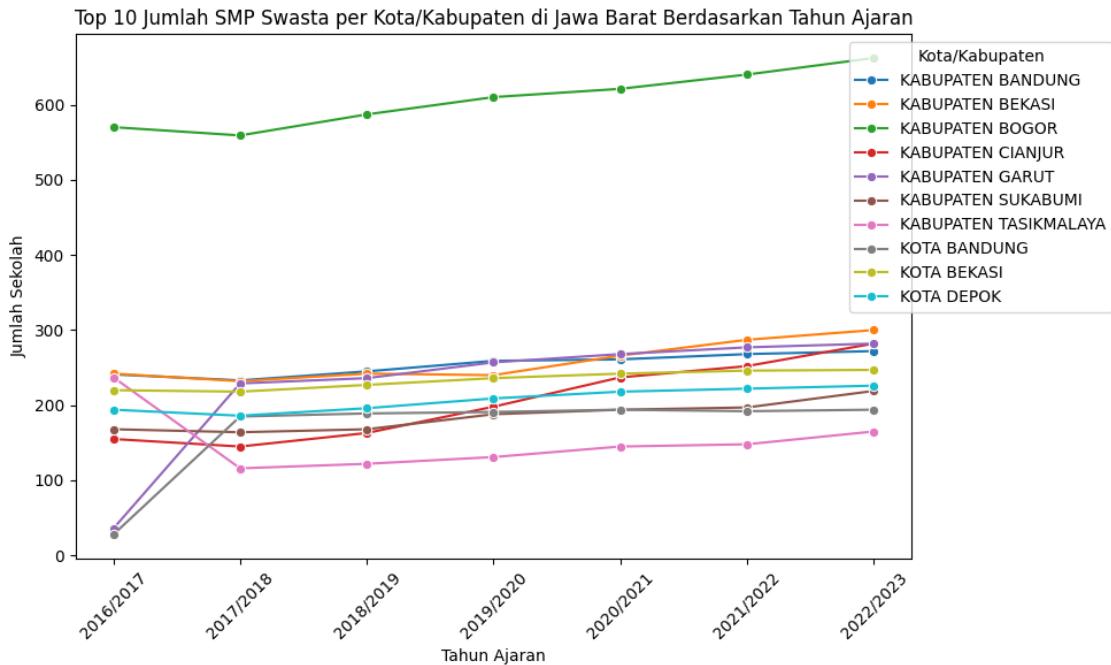
data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

data_subset = grouped_data.get_group(pd_key)
```



Melihat 10 Kota Kabupaten dengan Jumlah SMA Swasta Terbanyak di Jawa Barat

```
[ ]: # Filter data untuk kategori_sekolah 'NEGERI'
sma_negeri = sma[sma['kategori_sekolah'] == 'SWASTA']

# Group by 'nama_kabupaten_kota' dan hitung total jumlah_sekolah di seluruh tahun ajaran
total_sekolah_negeri = sma_negeri.
    ↪groupby('nama_kabupaten_kota')['jumlah_sekolah'].sum().reset_index()

# Urutkan berdasarkan jumlah_sekolah terbanyak dan ambil 10 kota/kabupaten teratas
top_10_kabupaten_kota = total_sekolah_negeri.nlargest(10,
    ↪'jumlah_sekolah')['nama_kabupaten_kota']

# Filter kembali data asli untuk hanya menyertakan 10 kota/kabupaten teratas
sma_negeri_top_10 = sma_negeri[sma_negeri['nama_kabupaten_kota'].isin(top_10_kabupaten_kota)]

# Group by 'nama_kabupaten_kota' dan 'tahun_ajaran', lalu hitung total jumlah_sekolah
stats_negeri_top_10 = sma_negeri_top_10.groupby(['nama_kabupaten_kota',
    ↪'tahun_ajaran'])['jumlah_sekolah'].sum().reset_index()

# Buat lineplot
```

```

plt.figure(figsize=(10, 6))
sns.lineplot(data=stats_negeri_top_10, x='tahun_ajaran', y='jumlah_sekolah',  

             hue='nama_kabupaten_kota', marker='o')

# Menambahkan judul dan label
plt.title('Top 10 Jumlah SMA Swasta per Kota/Kabupaten di Jawa Barat  

           Berdasarkan Tahun Ajaran')
plt.xlabel('Tahun Ajaran')
plt.ylabel('Jumlah Sekolah')
plt.xticks(rotation=45) # Memutar label tahun ajaran agar lebih mudah dibaca
plt.legend(loc='upper right', bbox_to_anchor=(1.25, 1), title='Kota/Kabupaten')

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

```

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.

```
    data_subset = grouped_data.get_group(pd_key)
```

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.

```
    data_subset = grouped_data.get_group(pd_key)
```

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.

```
    data_subset = grouped_data.get_group(pd_key)
```

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.

```
    data_subset = grouped_data.get_group(pd_key)
```

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.

```
    data_subset = grouped_data.get_group(pd_key)
```

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.

```
    data_subset = grouped_data.get_group(pd_key)
```

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:

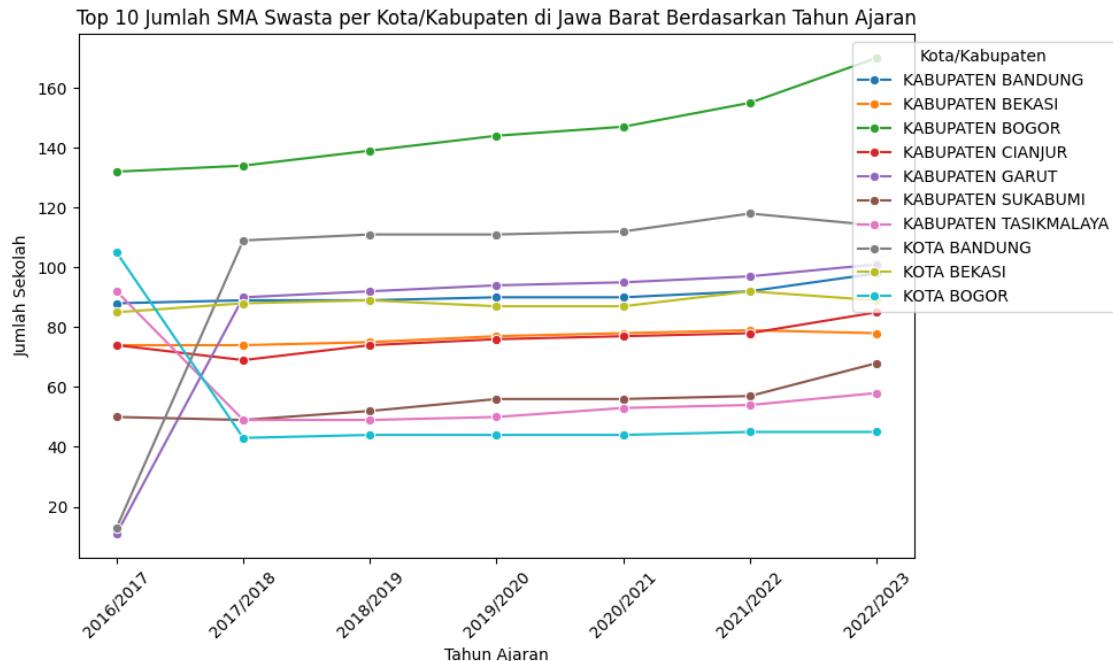
When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.
```

```
data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.
```

```
data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.
```

```
data_subset = grouped_data.get_group(pd_key)
```



Melihat 10 Kota/Kabupaten dengan Jumlah SMP Swasta paling sedikit di jawa Barat

```
[ ]: # Filter data untuk kategori_sekolah 'NEGERI'
smp_negeri = smp[smp['kategori_sekolah'] == 'SWASTA']
```

```

# Group by 'nama_kabupaten_kota' dan hitung total jumlah_sekolah di seluruh
# tahun ajaran
total_sekolah_negeri = smp_negeri.
    ↪groupby('nama_kabupaten_kota')['jumlah_sekolah'].sum().reset_index()

# Urutkan berdasarkan jumlah_sekolah terbanyak dan ambil 10 kota/kabupaten
# teratas
top_10_kabupaten_kota = total_sekolah_negeri.nsmallest(10,
    ↪'jumlah_sekolah')['nama_kabupaten_kota']

# Filter kembali data asli untuk hanya menyertakan 10 kota/kabupaten teratas
smp_negeri_top_10 = smp_negeri[smp_negeri['nama_kabupaten_kota'].
    ↪isin(top_10_kabupaten_kota)]

# Group by 'nama_kabupaten_kota' dan 'tahun_ajaran', lalu hitung total
# 'jumlah_sekolah'
stats_negeri_top_10 = smp_negeri_top_10.groupby(['nama_kabupaten_kota',
    ↪'tahun_ajaran'])['jumlah_sekolah'].sum().reset_index()

# Buat lineplot
plt.figure(figsize=(10, 6))
sns.lineplot(data=stats_negeri_top_10, x='tahun_ajaran', y='jumlah_sekolah',
    ↪hue='nama_kabupaten_kota', marker='o')

# Menambahkan judul dan label
plt.title('Kabupaten/Kota dengan Jumlah SMP Swasta paling sedikit di Jawa Barat
    ↪Berdasarkan Tahun Ajaran')
plt.xlabel('Tahun Ajaran')
plt.ylabel('Jumlah Sekolah')
plt.xticks(rotation=45) # Memutar label tahun ajaran agar lebih mudah dibaca
plt.legend(loc='upper right', bbox_to_anchor=(1.25, 1), title='Kota/Kabupaten')

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

```

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `^(name,)` instead of `name` to
silence this warning.

```

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  

When grouping with a length-1 list-like, you will need to pass a length-1 tuple  

to get_group in a future version of pandas. Pass `^(name,)` instead of `name` to  

silence this warning.

```

```

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:

```

When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

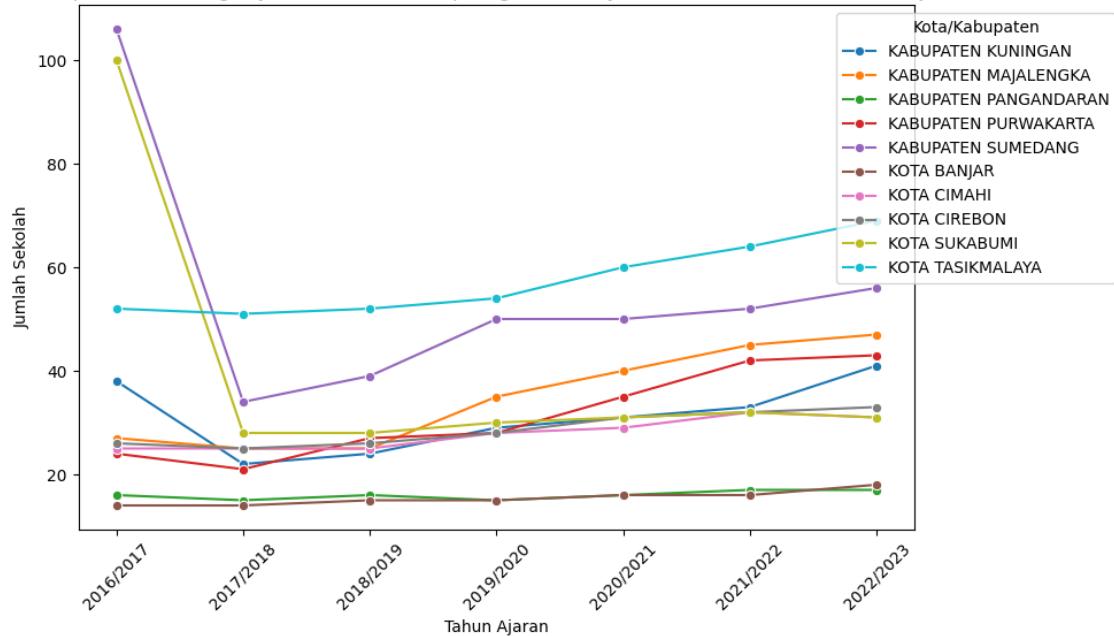
data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

data_subset = grouped_data.get_group(pd_key)
```

Kabupaten/Kota dengan Jumlah SMP Swasta paling sedikit di Jawa Barat Berdasarkan Tahun Ajaran



Melihat 10 Kota/Kabupaten dengan Jumlah SMA Swasta paling sedikit di jawa Barat

```
[ ]: # Filter data untuk kategori_sekolah 'NEGERI'
sma_negeri = sma[sma['kategori_sekolah'] == 'SWASTA']

# Group by 'nama_kabupaten_kota' dan hitung total jumlah_sekolah di seluruh
# tahun ajaran
total_sekolah_negeri = sma_negeri.
    ↪groupby('nama_kabupaten_kota')['jumlah_sekolah'].sum().reset_index()

# Urutkan berdasarkan jumlah_sekolah terbanyak dan ambil 10 kota/kabupaten
# teratas
top_10_kabupaten_kota = total_sekolah_negeri.nsmallest(10,
    ↪'jumlah_sekolah')['nama_kabupaten_kota']

# Filter kembali data asli untuk hanya menyertakan 10 kota/kabupaten teratas
sma_negeri_top_10 = sma_negeri[sma_negeri['nama_kabupaten_kota'].
    ↪isin(top_10_kabupaten_kota)]

# Group by 'nama_kabupaten_kota' dan 'tahun_ajaran', lalu hitung total
# jumlah_sekolah
stats_negeri_top_10 = sma_negeri_top_10.groupby(['nama_kabupaten_kota',
    ↪'tahun_ajaran'])['jumlah_sekolah'].sum().reset_index()

# Buat lineplot
```

```

plt.figure(figsize=(10, 6))
sns.lineplot(data=stats_negeri_top_10, x='tahun_ajaran', y='jumlah_sekolah', u
↳hue='nama_kabupaten_kota', marker='o')

# Menambahkan judul dan label
plt.title('Kabupaten/Kota dengan Jumlah SMA Swasta paling sedikit di Jawa Barat' u
↳Berdasarkan Tahun Ajaran')
plt.xlabel('Tahun Ajaran')
plt.ylabel('Jumlah Sekolah')
plt.xticks(rotation=45) # Memutar label tahun ajaran agar lebih mudah dibaca
plt.legend(loc='upper right', bbox_to_anchor=(1.25, 1), title='Kota/Kabupaten')

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

```

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.

```

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  

When grouping with a length-1 list-like, you will need to pass a length-1 tuple  

to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to  

silence this warning.

```

```

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  

When grouping with a length-1 list-like, you will need to pass a length-1 tuple  

to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to  

silence this warning.

```

```

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  

When grouping with a length-1 list-like, you will need to pass a length-1 tuple  

to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to  

silence this warning.

```

```

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  

When grouping with a length-1 list-like, you will need to pass a length-1 tuple  

to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to  

silence this warning.

```

```

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  

When grouping with a length-1 list-like, you will need to pass a length-1 tuple  

to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to  

silence this warning.

```

```

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:

```

When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

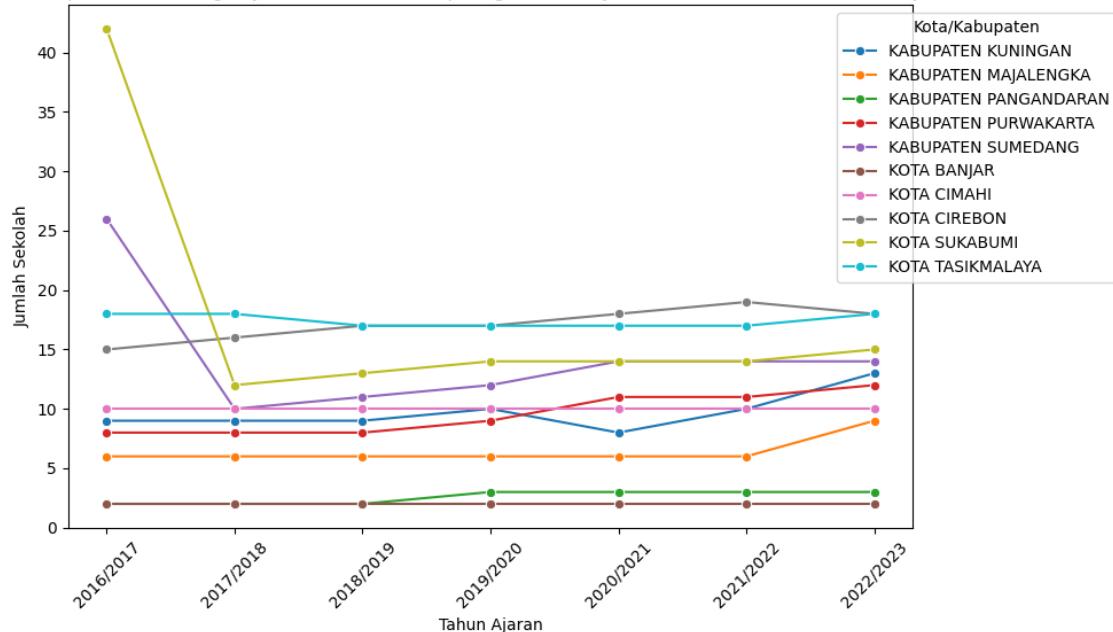
```
data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

data_subset = grouped_data.get_group(pd_key)
```

Kabupaten/Kota dengan Jumlah SMA Swasta paling sedikit di Jawa Barat Berdasarkan Tahun Ajaran



Melihat Perbandingan Jumlah SMP Negeri dan Swasta di tiap Kota/Kabupaten

```
[ ]: plt.figure(figsize=(14, 8))
sns.barplot(x='nama_kabupaten_kota', y='jumlah_sekolah',  
            hue='kategori_sekolah', data=smp, palette='Set2')
plt.title('Jumlah SMP Negeri dan Swasta per Kota/Kabupaten')
```

```

# Memutar label dengan sudut 45 derajat
plt.xticks(rotation=45, ha='right')

# Menggunakan metode pengaturan otomatis untuk menyesuaikan label
plt.tight_layout()

plt.show()

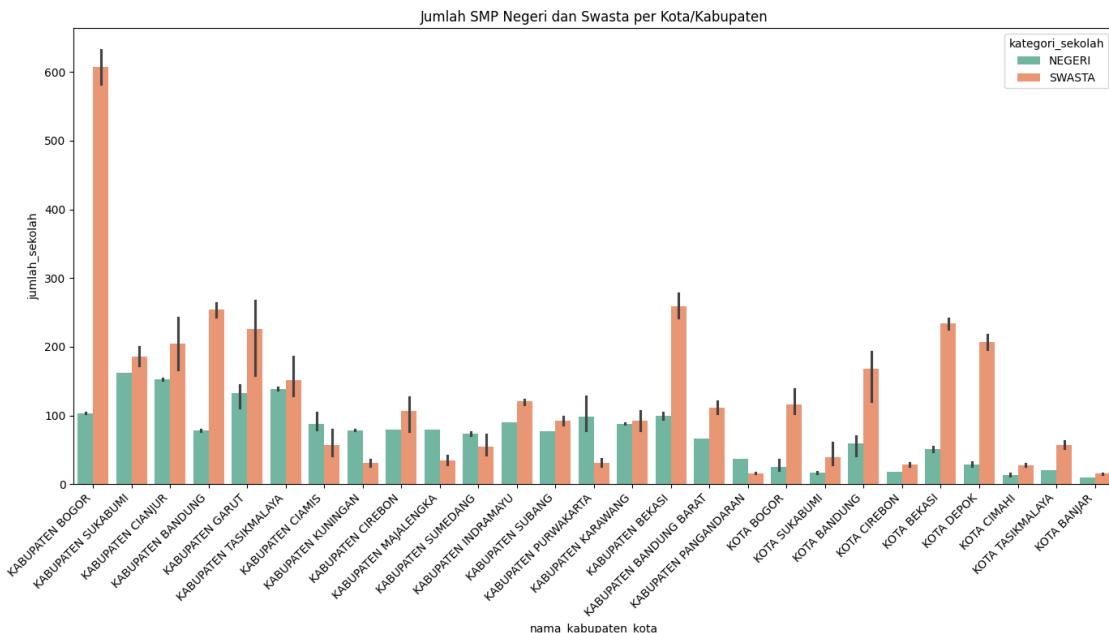
```

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `^(name,)` instead of `name` to
silence this warning.

```

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  
When grouping with a length-1 list-like, you will need to pass a length-1 tuple  
to get_group in a future version of pandas. Pass `^(name,)` instead of `name` to  
silence this warning.
data_subset = grouped_data.get_group(pd_key)

```



Melihat Perbandingan Jumlah SMA Negeri dan Swasta di Jawa Barat

```

[ ]: plt.figure(figsize=(14, 8))
sns.barplot(x='nama_kabupaten_kota', y='jumlah_sekolah',  

            hue='kategori_sekolah', data=sma, palette='Set2')
plt.title('Jumlah SMA Negeri dan Swasta per Kota/Kabupaten')

```

```

# Memutar label dengan sudut 45 derajat
plt.xticks(rotation=45, ha='right')

# Menggunakan metode pengaturan otomatis untuk menyesuaikan label
plt.tight_layout()

plt.show()

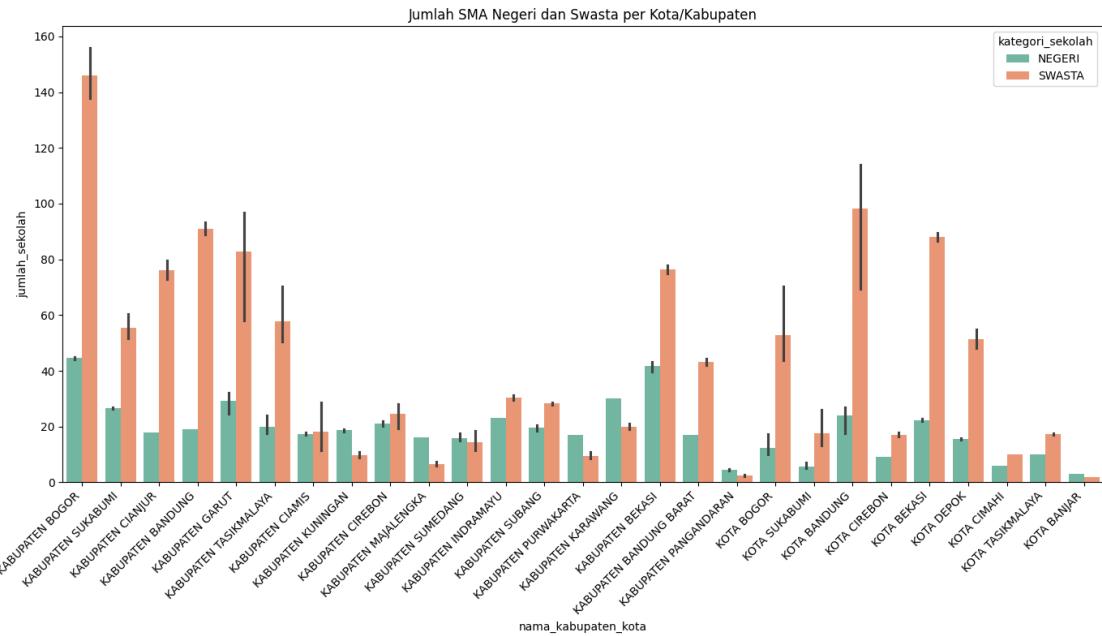
```

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `^(name,)` instead of `name` to
silence this warning.

```
    data_subset = grouped_data.get_group(pd_key)
```

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `^(name,)` instead of `name` to
silence this warning.

```
    data_subset = grouped_data.get_group(pd_key)
```



Perbandingan jumlah sekolah negeri dan swasta di tiap kabupaten/kota (SMP vs SMA)

```
[ ]: # Membuat figure dan subplot
fig, axes = plt.subplots(2, 1, figsize=(16, 12))

# Plot untuk SMP
```

```

sns.barplot(ax=axes[0], x='nama_kabupaten_kota', y='jumlah_sekolah',  

    ↪hue='kategori_sekolah', data=smp, palette='Blues', ci=None)  

axes[0].set_title('Jumlah SMP Negeri dan Swasta per Kota/Kabupaten')  

axes[0].set_xticklabels(axes[0].get_xticklabels(), rotation=45, ha='right')

# Plot untuk SMA
sns.barplot(ax=axes[1], x='nama_kabupaten_kota', y='jumlah_sekolah',  

    ↪hue='kategori_sekolah', data=sma, palette='Greens', ci=None)  

axes[1].set_title('Jumlah SMA Negeri dan Swasta per Kota/Kabupaten')  

axes[1].set_xticklabels(axes[1].get_xticklabels(), rotation=45, ha='right')

# Menambahkan judul umum untuk seluruh plot
fig.suptitle('Perbandingan Jumlah Sekolah Negeri dan Swasta di Setiap Kota/  

    ↪Kabupaten (SMP vs SMA)', fontsize=16)

# Mengatur layout agar tidak overlap
plt.tight_layout(rect=[0, 0, 1, 0.95])

# Menampilkan plot
plt.show()

```

<ipython-input-38-8688615d6e94>:5: FutureWarning:

The `ci` parameter is deprecated. Use `errorbar=None` for the same effect.

```

sns.barplot(ax=axes[0], x='nama_kabupaten_kota', y='jumlah_sekolah',  

hue='kategori_sekolah', data=smp, palette='Blues', ci=None)  

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  

When grouping with a length-1 list-like, you will need to pass a length-1 tuple  

to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to  

silence this warning.  

    data_subset = grouped_data.get_group(pd_key)  

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  

When grouping with a length-1 list-like, you will need to pass a length-1 tuple  

to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to  

silence this warning.  

    data_subset = grouped_data.get_group(pd_key)  

<ipython-input-38-8688615d6e94>:7: UserWarning: FixedFormatter should only be  

used together with FixedLocator  

    axes[0].set_xticklabels(axes[0].get_xticklabels(), rotation=45, ha='right')
<ipython-input-38-8688615d6e94>:10: FutureWarning:

```

The `ci` parameter is deprecated. Use `errorbar=None` for the same effect.

```

sns.barplot(ax=axes[1], x='nama_kabupaten_kota', y='jumlah_sekolah',  

hue='kategori_sekolah', data=sma, palette='Greens', ci=None)  

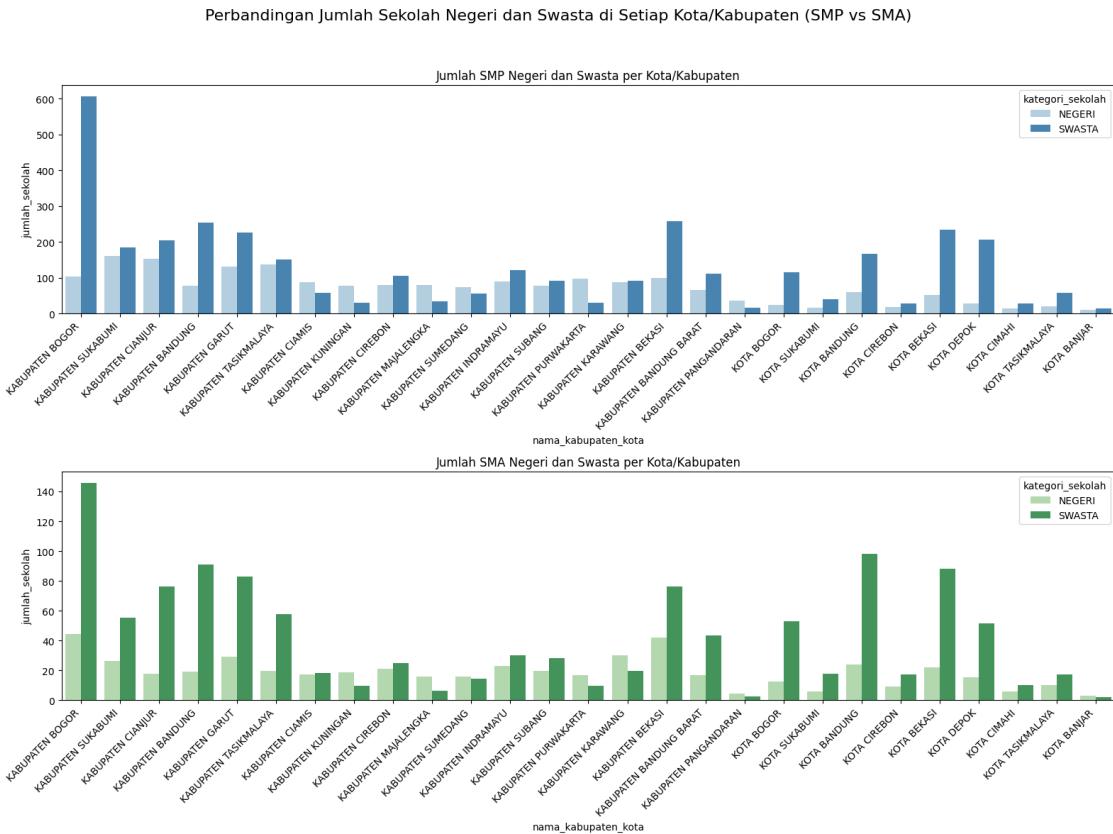
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:

```

When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

data_subset = grouped_data.get_group(pd_key)
<ipython-input-38-8688615d6e94>:12: UserWarning: FixedFormatter should only be
used together with FixedLocator
    axes[1].set_xticklabels(axes[1].get_xticklabels(), rotation=45, ha='right')
```



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)

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

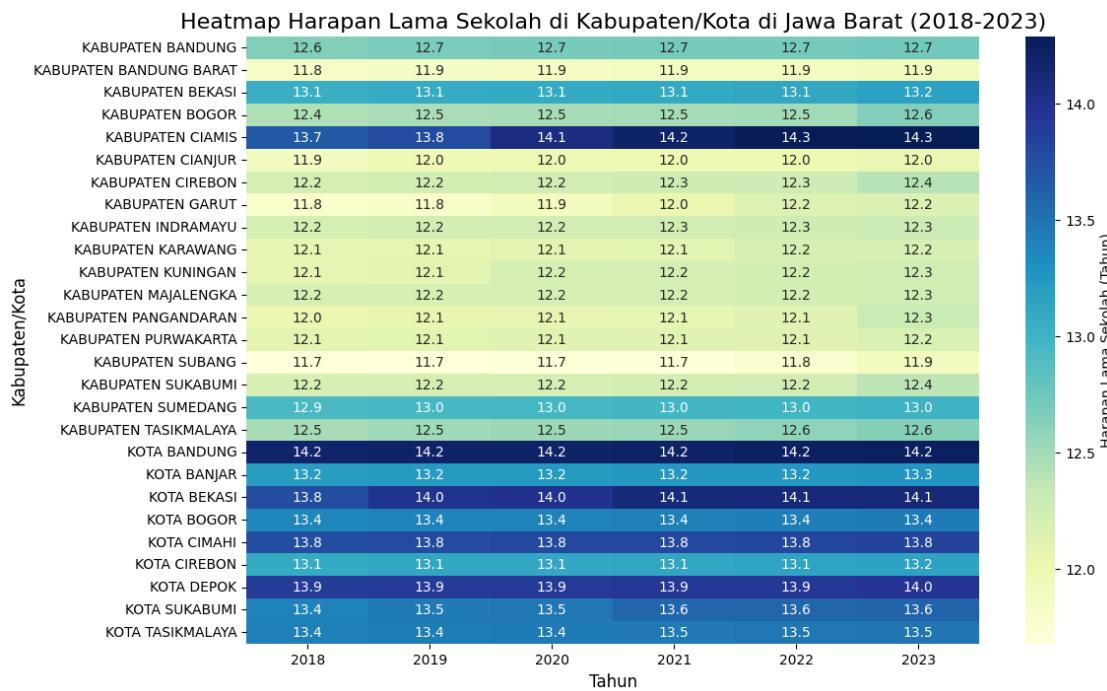
```

# 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)

```
[ ]: # 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

```
[ ]: # 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()
```

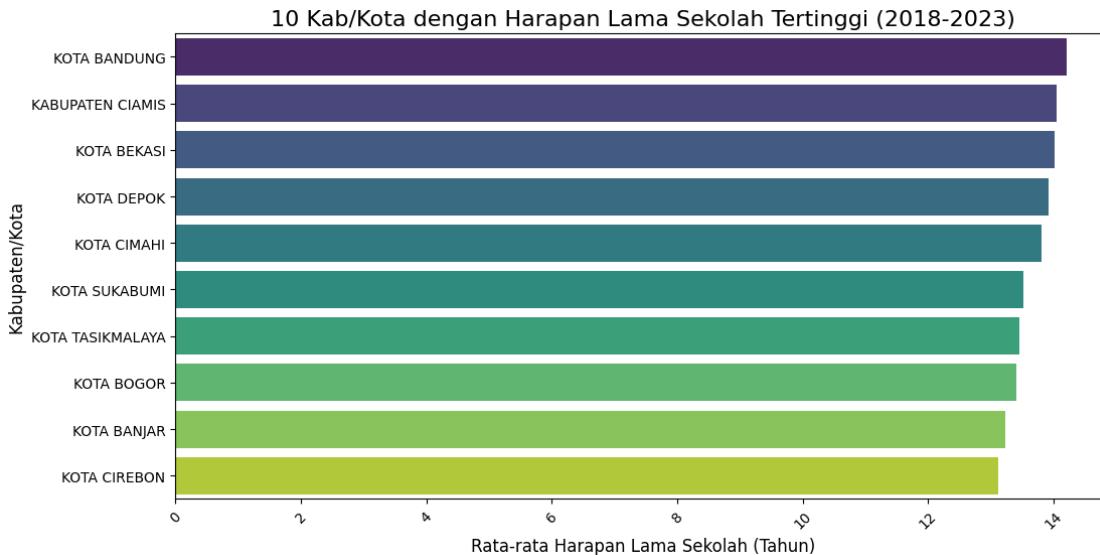
<ipython-input-48-ac3669dc33b2>:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

```
sns.barplot(x='harapan_lama_sekolah', y='nama_kabupaten_kota', data=top_hls,
palette='viridis')
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.
```

When grouping with a length-1 list-like, you will need to pass a length-1 tuple to `get_group` in a future version of pandas. Pass ``(name,)`` instead of ``name`` to silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
```



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

```
[ ]: # 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

```
[ ]: # 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)', fontstyle='italic',
           fontsize=16)
plt.xlabel('Rata-rata Harapan Lama Sekolah (tahun)', fontsize=14)
plt.ylabel('Kabupaten/Kota', fontsize=14)
plt.xticks(rotation=45)
plt.show()
```

<ipython-input-50-6dad83372170>:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

```
sns.barplot(x='harapan_lama_sekolah', y='nama_kabupaten_kota',
            data=lowest_hls, palette='viridis')
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
```

```

/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.

    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.

    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.

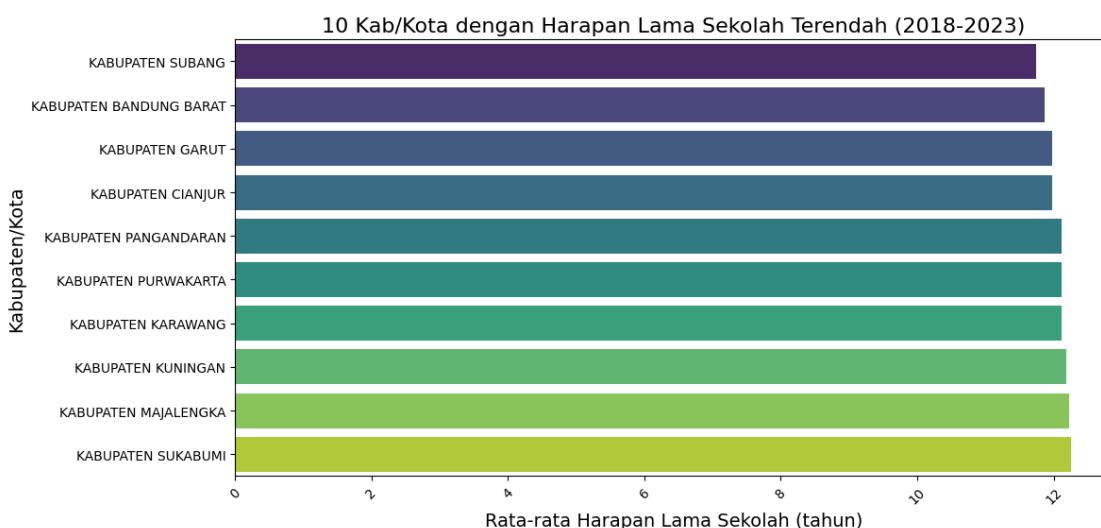
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.

    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.

    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.

    data_subset = grouped_data.get_group(pd_key)

```



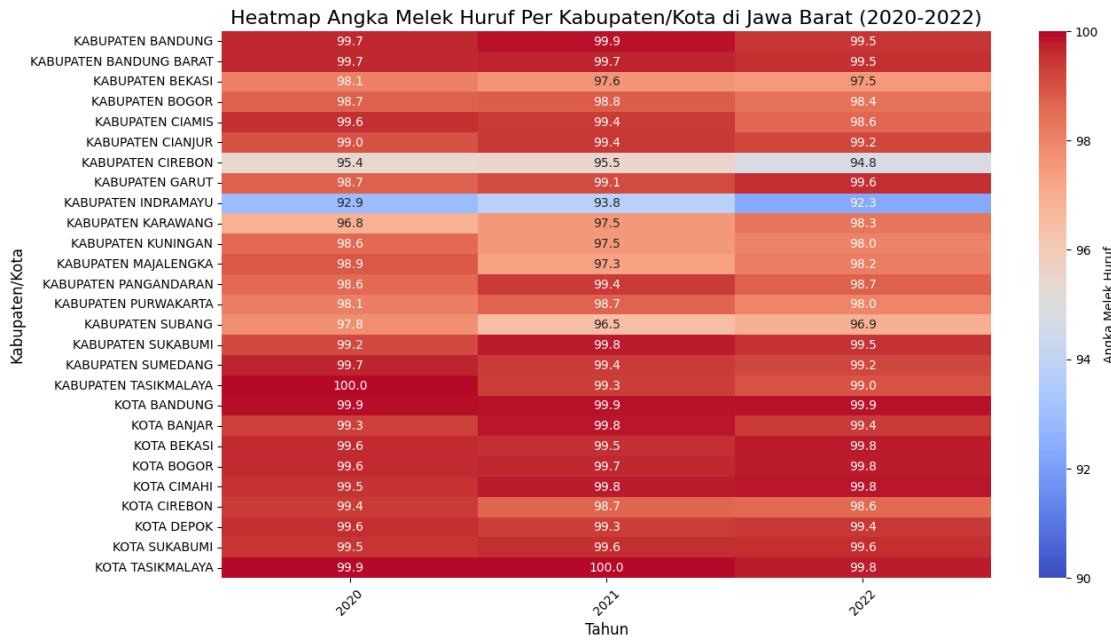
Kesimpulan - Kota Bandung menempati posisi teratas dengan harapan lama sekolah rata-rata sebesar 14.21 tahun, diikuti oleh Kabupaten Ciamis dan Kota Bekasi. Hal ini menunjukkan bahwa kota-kota tersebut memiliki sistem pendidikan yang relatif lebih baik dan akses yang lebih besar terhadap pendidikan bagi 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 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)

```
[ ]: # 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()
```



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

```
[ ]: # 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)
```

| nama_kabupaten_kota | rata-rata |
|---------------------|-----------|
| KOTA TASIKMALAYA | 99.893333 |
| KOTA BANDUNG | 99.890000 |
| KOTA CIMAH | 99.716667 |
| KOTA BOGOR | 99.696667 |
| KABUPATEN BANDUNG | 99.676667 |
| KOTA BEKASI | 99.650000 |

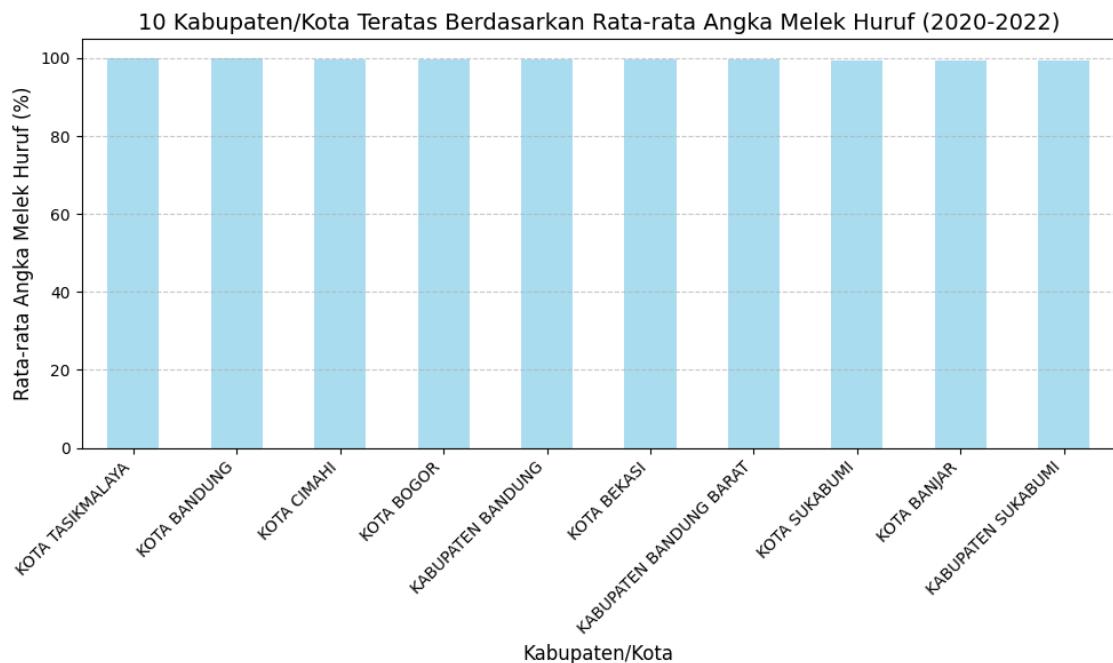
```

KABUPATEN BANDUNG BARAT      99.633333
KOTA SUKABUMI                 99.543333
KOTA BANJAR                   99.510000
KABUPATEN SUKABUMI            99.496667
dtype: float64

```

Visualisasi Top 10 Teratas

```
[ ]: # 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()
```



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

```
[ ]: # 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', u
↳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)

```

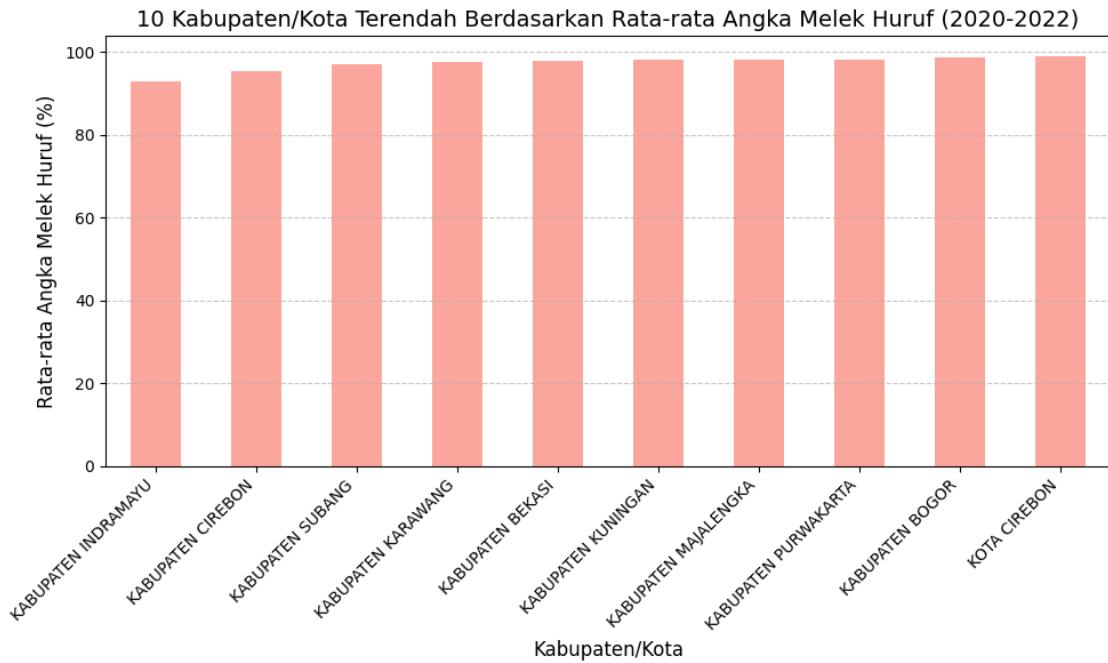
Top 10 Kabupaten/Kota dengan Rata-rata Angka Melek Huruf Terendah:

| | |
|----------------------|-----------|
| 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

```
[ ]: # 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 u
↳(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()
```



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

0.1.6 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)

```
[ ]: # 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 CIMAH'I': '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'], None)
    if abbreviation is None:
        abbreviation = 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)

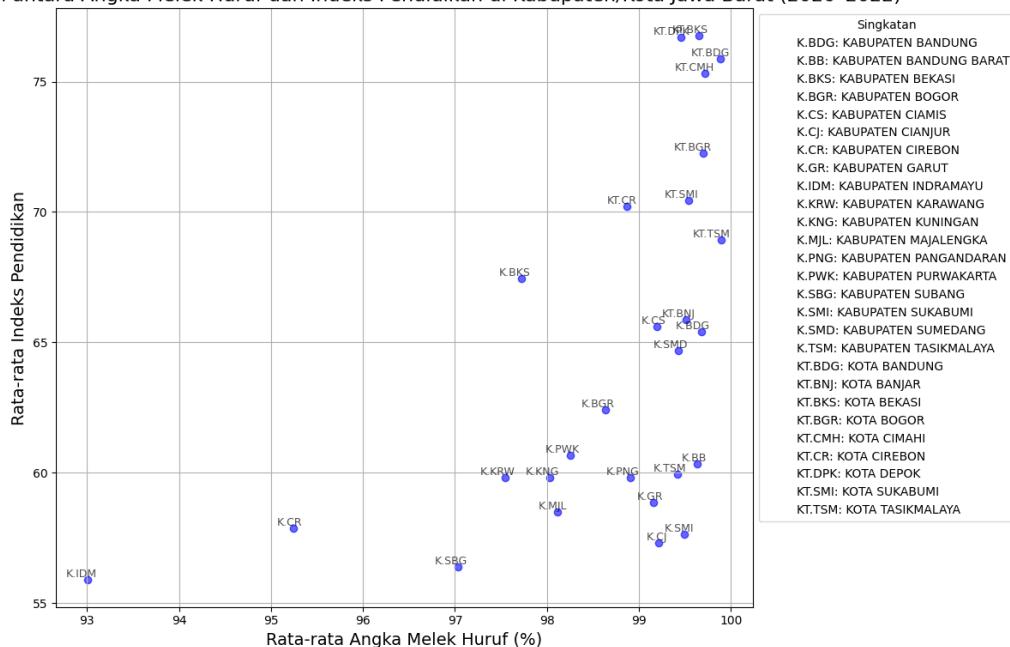
# Menambahkan legend untuk singkatan tanpa titik
handles = [plt.Line2D([0], [0], marker=None, color='w', label=f'{abbreviations[full_name]}:{full_name}') for full_name, abbrev in abbreviations.items()]

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)



Bisa dilihat semakin besar rata-rata angka melek hurufnya, maka semakin besar juga indeks pendidikannya

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)

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

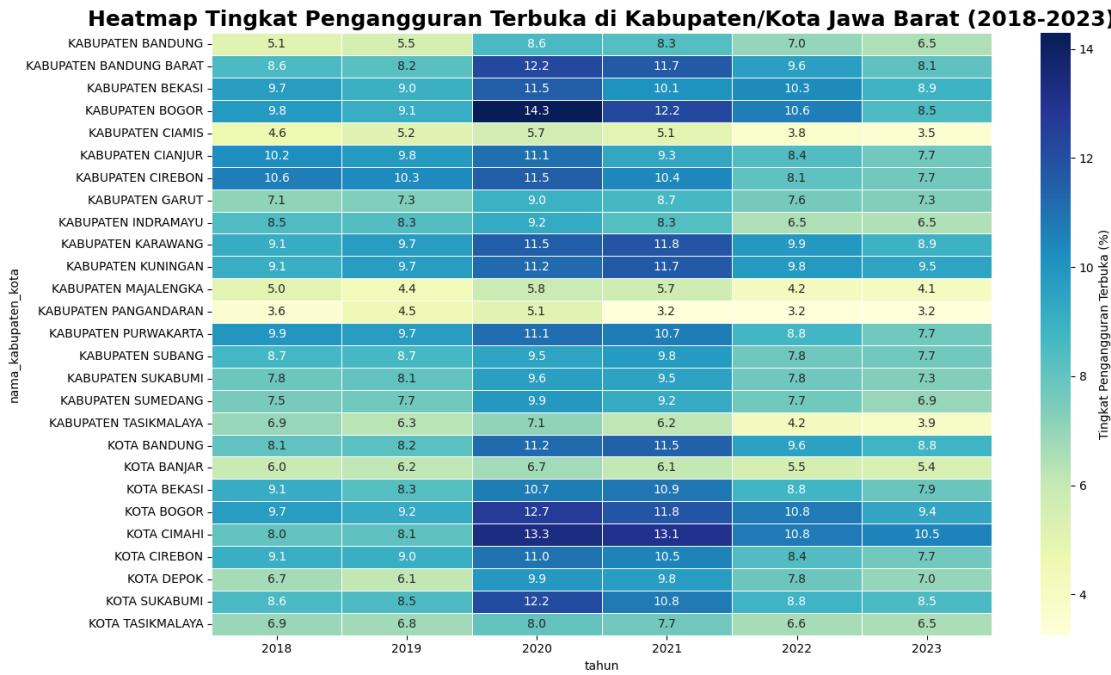
# Mengatur ukuran heatmap
plt.figure(figsize=(14, 8))

# 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()
```



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

```
[ ]: # 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

```
[ ]: # 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()
```

<ipython-input-56-15715b392fa2>:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

```
sns.barplot(x='tingkat_pengangguran_terbuka', y='nama_kabupaten_kota',
            data=top_unemployment, palette='viridis')
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `^(name,)` instead of `name` to
silence this warning.

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `^(name,)` instead of `name` to
silence this warning.

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `^(name,)` instead of `name` to
silence this warning.

data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `^(name,)` instead of `name` to
```

```
to get_group in a future version of pandas. Pass `name` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `name` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `name` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `name` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `name` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `name` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
```



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

```
[ ]: # 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 MAJALENGKA | 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

```
[ ]: # 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()
```

<ipython-input-55-6cfa695a6fca>:12: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

```
sns.barplot(x='tingkat_pengangguran_terbuka', y='nama_kabupaten_kota',
data=bottom_unemployment, palette='viridis')
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `'(name,)` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
```

```
to get_group in a future version of pandas. Pass `name` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `name` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `name` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `name` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `name` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `name` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `name` instead of `name` to
silence this warning.
    data_subset = grouped_data.get_group(pd_key)
```



Kesimpulan

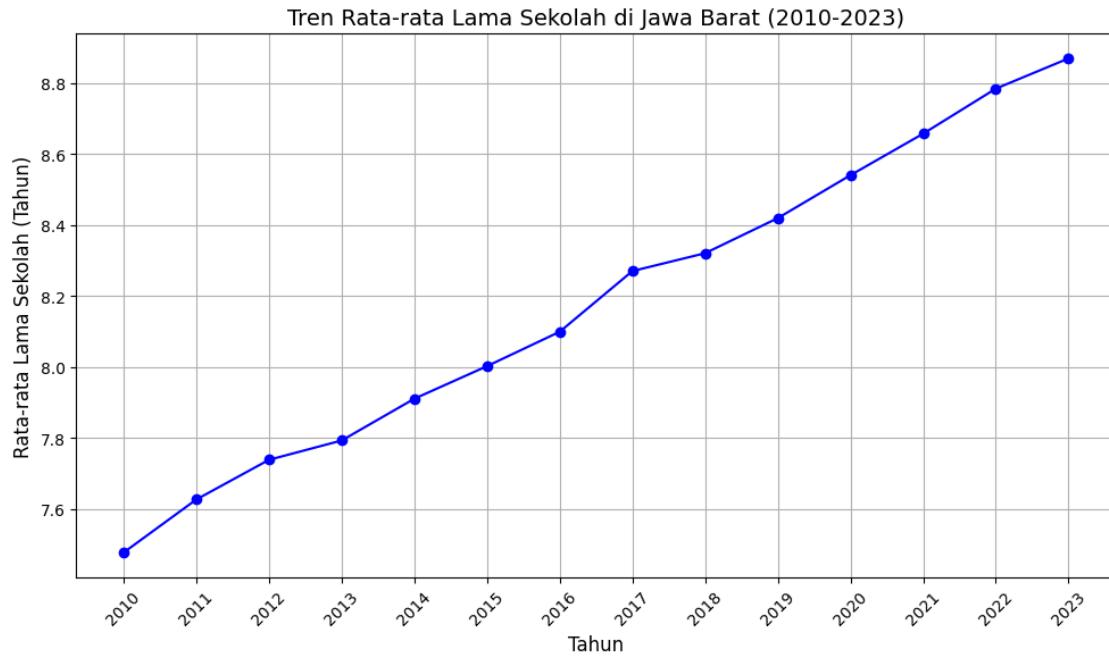
- Kabupaten Bogor mencatat tingkat pengangguran terbuka tertinggi sebesar 10,76%, diikuti oleh Kota Cimahi dan Kota Bogor dengan masing-masing 10,63% dan 10,59%.
- Kabupaten Pangandaran memiliki tingkat pengangguran terbuka terendah, yaitu 3,82%, yang menunjukkan kinerja positif dalam penyerapan tenaga 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 Data Rata Rata Lama Sekolah di Provinsi Jawa Barat

Tren rata-rata lama sekolah per tahun

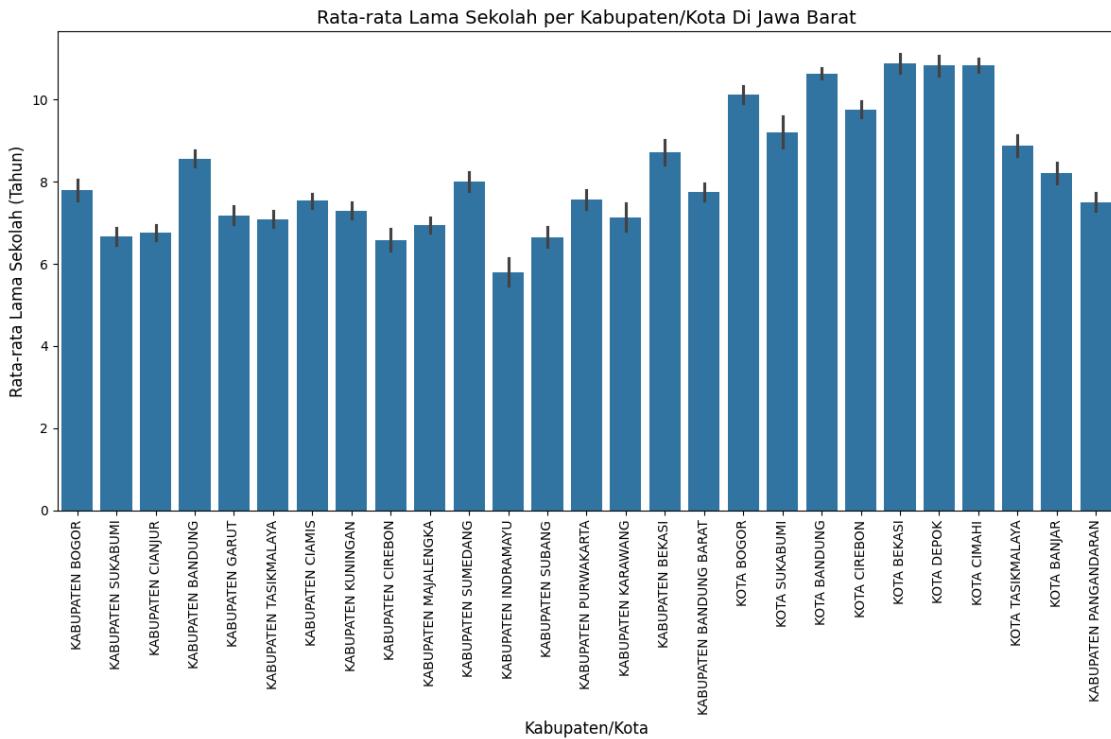
```
[ ]: rata_rata_sekolah_per_tahun = rrl.groupby('tahun')[['rata_rata_lama_sekolah']].  
      ↪mean()  
  
# Membuat plot garis untuk tren rata-rata lama sekolah dari tahun 2010 hingga  
# 2023  
plt.figure(figsize=(10, 6))  
plt.plot(rata_rata_sekolah_per_tahun.index, rata_rata_sekolah_per_tahun.values,  
         ↪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()
```



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

```
[ ]: plt.figure(figsize=(12, 8))
sns.barplot(x='nama_kabupaten_kota', y='rata_rata_lama_sekolah', data=rwl)
plt.title('Rata-rata Lama Sekolah per Kabupaten/Kota Di Jawa Barat', fontstyle='italic', 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()
```



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

```
[ ]: # 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', palette='viridis')
plt.title('Top 10 Kabupaten/Kota dengan Rata-rata Lama Sekolah Tertinggi')
plt.xlabel('Rata-rata Lama Sekolah')
plt.ylabel('Kabupaten/Kota')
plt.show()
```

<ipython-input-16-538095d9922a>:12: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

```
sns.barplot(data=top_10_kabupaten, x='rata_rata_lama_sekolah',
y='nama_kabupaten_kota', palette='viridis')
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.

    data_subset = grouped_data.get_group(pd_key)
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:
When grouping with a length-1 list-like, you will need to pass a length-1 tuple
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to
silence this warning.
```

```
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to  
silence this warning.
```

```
data_subset = grouped_data.get_group(pd_key)  
/usr/local/lib/python3.10/dist-packages/seaborn/_base.py:949: FutureWarning:  
When grouping with a length-1 list-like, you will need to pass a length-1 tuple  
to get_group in a future version of pandas. Pass `(name,)` instead of `name` to  
silence this warning.  
data_subset = grouped_data.get_group(pd_key)
```

