

LAPORAN PROJECT AKHIR DATA WRANGLING
ANALISIS KETERSEDIAAN FASILITAS KESEHATAN DAN JUMLAH
PENDUDUK DI JAWA BARAT TAHUN 2024



Dosen Pengampu:
Ulfa Siti Nuraini, S.Stat., M.Stat.

Disusun Oleh:
Kelompok 17

1. Eka Putri Maharani (24031554121)
2. Muslim Fazlur Rohman (24031554154)

PROGRAM STUDI SAINS DATA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI SURABAYA

2025

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Pelayanan kesehatan merupakan salah satu indikator penting dalam menilai kualitas kesejahteraan masyarakat. Ketersediaan fasilitas kesehatan yang memadai dan merata sangat menentukan akses masyarakat terhadap pelayanan kesehatan yang cepat, aman, dan berkualitas. Ketidakseimbangan antara jumlah penduduk dan fasilitas kesehatan dapat menyebabkan tingginya beban pelayanan, antrian panjang, hingga penurunan mutu layanan (Kemenkes RI, 2024). Provinsi Jawa Barat merupakan provinsi dengan jumlah penduduk terbesar di Indonesia, yaitu lebih dari 49 juta jiwa pada tahun 2024 (BPS, 2024). Sebaran penduduk yang besar dan heterogen di 27 kabupaten/kota membuat kebutuhan fasilitas kesehatan di provinsi ini sangat tinggi. Namun, berbagai laporan resmi menunjukkan bahwa pemerataan fasilitas kesehatan di Jawa Barat masih menghadapi tantangan. Beberapa wilayah dengan jumlah penduduk besar, seperti Kabupaten Bogor, Kabupaten Bandung, dan Kabupaten Garut, mengalami keterbatasan fasilitas kesehatan jika dibandingkan dengan beban penduduknya (Dewi et al., 2023).

Untuk menganalisis kondisi tersebut, diperlukan data yang bersih, konsisten, dan terintegrasi. Akan tetapi, data yang berasal dari berbagai sumber, seperti Open Data Jabar, Kementerian Kesehatan, dan dataset publik lainnya, sering kali memiliki format berbeda, ketidaksamaan penamaan wilayah, perbedaan satuan, serta struktur kolom yang tidak seragam. Oleh karena itu, proses data wrangling menjadi tahap penting agar data dapat digunakan untuk analisis secara akurat (McKinney, 2017). Dalam project ini, tiga dataset utama digunakan: data jumlah penduduk Jawa Barat tahun 2024, data fasilitas kesehatan berdasarkan jenis, dan data rumah sakit seluruh Indonesia yang kemudian di filter khusus Jawa Barat. Setelah dilakukan proses pembersihan dan integrasi data, analisis dilakukan menggunakan metode Exploratory Data Analysis (EDA) untuk menggambarkan persebaran penduduk, pemerataan fasilitas kesehatan, serta tingkat kecukupan fasilitas dibandingkan standar rasio nasional, yaitu 1 rumah sakit per 100.000 penduduk dan 1 fasilitas kesehatan dasar per 30.000 penduduk (Permenkes No. 75 Tahun 2014).

Melalui pendekatan ini, penelitian diharapkan dapat memberikan gambaran yang komprehensif mengenai kondisi pemerataan fasilitas kesehatan di Jawa Barat, mengidentifikasi wilayah yang mengalami kekurangan fasilitas, serta memberikan dasar analisis bagi kebijakan pembangunan kesehatan di masa depan.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, rumusan masalah penelitian ini adalah sebagai berikut:

1. Bagaimana proses pengumpulan, pembersihan, normalisasi, dan integrasi data dilakukan untuk menghasilkan dataset final yang siap dianalisis?
2. Bagaimana persebaran jumlah penduduk di setiap kabupaten/kota di Jawa Barat tahun 2024?

3. Bagaimana kondisi persebaran fasilitas kesehatan dan rumah sakit di Jawa Barat?
4. Apakah ketersediaan fasilitas kesehatan tersebut telah proporsional terhadap jumlah penduduk berdasarkan standar rasio nasional?
5. Wilayah mana saja yang mengalami kekurangan fasilitas kesehatan?
6. Bagaimana hasil data wrangling dapat membantu menjawab permasalahan ketimpangan fasilitas kesehatan?

1.3 Tujuan

Penelitian ini bertujuan untuk:

1. Mendeskripsikan jumlah penduduk pada setiap kabupaten/kota di Jawa Barat tahun 2024.
2. Menganalisis kondisi jumlah fasilitas kesehatan dan rumah sakit di setiap wilayah.
3. Menghitung kebutuhan rumah sakit dan fasilitas kesehatan dasar berdasarkan standar nasional.
4. Mengidentifikasi wilayah yang mengalami kekurangan fasilitas kesehatan.
5. Menyajikan hasil analisis dalam bentuk visualisasi yang informatif untuk mendukung interpretasi.

1.4 Manfaat

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Menjadi referensi bagi pemerintah daerah dalam perencanaan penyediaan fasilitas kesehatan.
2. Memberikan contoh penerapan data wrangling dalam analisis kesehatan masyarakat.
3. Menyediakan dataset yang lebih bersih dan terstruktur untuk penelitian lanjutan.
4. Mendukung upaya pengambilan keputusan berbasis data.

BAB II

METODOLOGI PENELITIAN

2.1 Deskripsi Dataset

Dalam project kali ini kami menggunakan 3 dataset dari sumber berbeda yang menjadi sumber informasi mengenai jumlah penduduk, fasilitas kesehatan, dan data rumah sakit. ketiga dataset ini masing-masing memiliki struktur dan format yang berbeda-beda sehingga diperlukan proses wrangling sebelum dilakukan analisis. berikut adalah deskripsi dari masing-masing dataset:

- Dataset penduduk - “jumlah penduduk menurut kabupaten/kota di provinsi jawa barat, 2024.csv”
dataset ini kami ambil dari BPS kabupaten bandung yang berisi informasi jumlah penduduk pada setiap kabupaten/kota di jawa barat tahun 2024. data ini menjadi dasar perhitungan kebutuhan fasilitas kesehatan di jawa barat. file ini memiliki format csv dengan 2 kolom utama, yaitu kabupaten/kota dan jumlah penduduk. dataset ini terdapat nama kabupaten/kota dengan format yang berbeda, terdapat baris yang perlu dihilangkan, dan adanya kemungkinan memiliki nilai kosong.
- Dataset faskes -
“dinkes-od_15936_jumlah_fasilitas_kesehatan_berdasarkan_jenis_v2_data.csv”
dataset ini kami ambil dari open data jabar, berisi informasi tentang mengenai jumlah fasilitas kesehatan berdasarkan jenisnya, seperti puskesmas, klinik, lab, dll. dataset ini memiliki format csv dimana setiap baris merepresentasikan jenis fasilitas tertentu pada daerah tertentu. data ini memuat beberapa tahun tidak hanya tahun 2024. penulisan nama kabupaten tidak sama sehingga perlu dilakukan normalisasi. selain itu, tipe data dari kolom jumlah fasilitas kesehatan perlu diubah ke tipe numerik.
- Dataset rumah sakit - “Hospital_indonesia_datasets.csv”
dataset ini kami ambil dari kaggle, berisi seluruh data rumah sakit di indonesia. dataset ini memiliki format csv dan menggunakan pemisah ‘ ; ’ data ini memiliki ribuan baris sehingga diperlukan memfilter agar mendapatkan data rumah sakit yang berada di jawa barat, penulisan nama kabupaten/kota di data ini juga tidak konsisten sehingga diperlukan proses cleaning dan penyamaan agar bisa digabung dengan dataset lain, dan ada beberapa rumah sakit dengan data yang tidak lengkap/nilai kosong.

Secara keseluruhan ketiga dataset ini menunjukkan berbagai tantangan dalam pengolahan data, seperti ketidakkonsistenan format, penulisan nama wilayah, perbedaan delimiter atau pemisah, serta adanya baris yang tidak relevan. oleh karena itu proses wrangling menjadi sangat penting untuk menyatukan ketiga dataset agar menjadi data akhir yang bersih, rapi dan siap digunakan untuk analisis ketersediaan jumlah fasilitas kesehatan di jawa barat.

2.2 Proses Wrangling

2.2.1 Teknik Pengambilan dan Integrasi Data

2.2.1.1 pengambilan data

Proses pengambilan data dilakukan secara manual dengan mengunduh file csv dari 3 sumber berbeda dan disimpan tanpa modifikasi. teknik pengambilan data dilakukan dengan memastikan bahwa setiap file memiliki struktur asli agar proses wrangling dapat dilakukan secara konsisten.

Proses pengambilan data menggunakan bahasa pemrograman python dan menggunakan library pandas sebagai alat utama untuk membaca dan mengolah data berbasis tabel. pada tahap ini ketiga dataset dimuat dari file csv, setiap file memiliki karakter yang berbeda sehingga diperlukan parameter yang sesuai.

```
penduduk_raw = pd.read_csv('Jumlah Penduduk Menurut Kabupaten_Kota di Provinsi Jawa Barat, 2024.csv',
skiprows=2)
faskes_raw = pd.read_csv('dinkes-od_15936_jumlah_fasilitas_kesehatan_berdasarkan_jenis_v2_data.csv')
hospital_raw = pd.read_csv('Hospital_Indonesia_datasets.csv', sep=';')
```

dataset penduduk menggunakan parameter skiprows=2 karena file memiliki 2 baris awal berupa header kolom. Dataset faskes bisa dibaca secara langsung tanpa parameter khusus. Dataset hospital dibaca dengan parameter sep=';' karena file ini menggunakan titik koma sebagai pemisah kolom, berbeda dengan 2 file sebelumnya.

2.2.1.2 integrasi data

Integrasi data dilakukan untuk menggabungkan informasi jumlah penduduk, jumlah fasilitas kesehatan, dan jumlah rumah sakit ke dalam satu tabel terpadu berdasarkan kolom kabupaten_kota. sebelum dilakukan integrasi, ketiga dataset harus melalui tahap cleaning dan normalisasi nama wilayah hingga siap untuk digabung.

Proses integrasi menggunakan fungsi *pd.merge()*. dataset pertama yang digabung adalah penduduk dan faskes, kemudian hasilnya digabung lagi dengan dataset hospital.

```
df = pd.merge(penduduk, faskes, on='kabupaten_kota', how='outer') \
.merge(df_RS, on='kabupaten_kota', how='outer')
```

how='outer' untuk memastikan seluruh kabupaten/kota tetap muncul dalam tabel akhir meskipun terdapat wilayah yang hanya muncul pada salah satu dataset. setelah proses merge ada kemungkinan muncul nilai kosong, karena tidak semua dataset memuat informasi lengkap untuk setiap wilayah.

```
df['jumlah_faskes'] = df['jumlah_faskes'].fillna(0).astype(int)
df['jumlah_penduduk'] = df['jumlah_penduduk'].fillna(0).astype(int)
df['jumlah_rs'] = df['jumlah_rs'].fillna(0).astype(int)
```

fungsi *fillna(0)* digunakan untuk mengganti nilai kosong dengan nol dan memastikan semua kolom memiliki tipe data yang sama menggunakan *.astype(int)*

2.2.2 Data Cleaning

Setelah proses pengambilan dan integrasi data selesai, selanjutnya adalah melakukan data cleaning untuk memastikan seluruh dataset memiliki struktur yang sama, bebas dari nilai yang tidak sesuai, serta siap digunakan pada proses analisis dan integrasi data. proses cleaning dilakukan melalui beberapa tahapan, mulai dari penyeragaman nama kolom, penghapusan data yang tidak sesuai, normalisasi, dll.

2.2.2.1 cleaning data penduduk

```
penduduk = penduduk_raw.copy()
```

```

# rename kolom
penduduk = penduduk.rename(columns={penduduk.columns[0]: 'kabupaten_kota', '2024': 'jumlah_penduduk'}) \
    .loc[:, ['kabupaten_kota', 'jumlah_penduduk']]

# drop total row jika ada
penduduk = penduduk[penduduk['kabupaten_kota'] != "Jawa Barat"].copy()
# konversi tipe data ke int64
penduduk["jumlah_penduduk"] = pd.to_numeric(penduduk["jumlah_penduduk"], errors="coerce")
penduduk["jumlah_penduduk"] = (penduduk["jumlah_penduduk"] * 1000).astype("Int64")
# Normalisasi
penduduk['kabupaten_kota'] = penduduk['kabupaten_kota'].astype(str).str.strip().str.replace(r'\s+', ' ',
regex=True)
penduduk['kabupaten_kota'] = penduduk['kabupaten_kota'].str.title()
# transformasi
penduduk = penduduk.groupby('kabupaten_kota', as_index=False)['jumlah_penduduk'].sum()
penduduk

```

Pada tahap awal, proses cleaning dimulai dengan membuat salinan dari dataset asli menggunakan fungsi *copy()*. lalu dilakukan penyeragaman struktur kolom dengan proses rename, kolom pertama diubah menjadi kabupaten_kota dan kolom 2024 diubah menjadi jumlah_penduduk. setelah itu kolom dipilih menggunakan fungsi *.loc[]* . lalu menghapus keseluruhan baris “jawa barat” untuk mencegah double count. Tahap berikutnya, kolom jumlah penduduk dikonversi ke tipe numerik. setelah memastikan data numerik sudah benar, nilai jumlah penduduk dikali dengan 1000 agar interpretasi data tetap jelas dan sesuai dengan konteks sumber aslinya. nilai kemudian dikonversi ke tipe *int64*, yaitu tipe integer khusus dalam pandas yang mampu menangani nilai hilang tanpa mengubah struktur data. Tahap selanjutnya adalah normalisasi, yang mencakup penghapusan spasi berlebih, penyeragaman format huruf serta menghapus karakter yang tidak diperlukan. Lalu langkah terakhir yaitu mengelompokkan berdasarkan nama kabupaten//kota menggunakan *groupby()* kemudian dijumlah dengan *.sum()* . Proses agregasi ini memastikan bahwa setiap kabupaten/kota hanya memiliki satu baris data dengan total jumlah penduduk yang akurat.

2.2.2.2 cleaning data faskes

```

faskes = faskes_raw.copy()

# Filter tahun 2024
df_faskes = faskes[faskes["tahun"] == 2024].copy()

# Rename kolom 'nama_kabupaten_kota' menjadi 'kabupaten_kota'
df_faskes = df_faskes.rename(columns={"nama_kabupaten_kota": "kabupaten_kota"})

# Memastikan jumlah_faskes numeric
df_faskes["jumlah_faskes"] = pd.to_numeric(df_faskes["jumlah_faskes"], errors="coerce")

# normalisasi
df_faskes['kabupaten_kota'] = df_faskes['kabupaten_kota'].astype(str).str.strip().str.replace(r'\s+', ' ',
regex=True).str.title()

# transformasi
faskes = df_faskes.groupby('kabupaten_kota', as_index=False)['jumlah_faskes'].sum()
faskes

```

Proses cleaning pada data faskes diawali dengan memfilter berdasarkan tahun 2024. setelah itu, kolom yang memuat nama wilayah di rename menjadi kabupaten_kota lalu dikonversi

menjadi numerik. Selanjutnya, nama wilayah dinormalisasi dengan merapikan penulisan agar konsisten. tahap terakhir adalah melakukan agregasi data berdasarkan kabupaten/kota untuk menjumlahkan total fasilitas kesehatan apabila satu wilayah tercatat lebih dari satu kali.

2.2.2.3 cleaning data hospital

```
hospital = hospital_raw.copy()

# filter provinsi Jawa Barat
df_hospital = hospital[hospital["propinsi"] == "Jawa Barat"].copy()

df_hospital['kabupaten_kota'] = (
    df_hospital['kab']
    .astype(str)
    .str.strip()
    .str.replace(r'\s+', ' ', regex=True)
    .str.title()
)

# count RS per kabupaten menggunakan id
if 'id' in df_hospital.columns:
    df_RS = df_hospital.groupby('kabupaten_kota', as_index=False).agg(jumlah_rs=('id', 'nunique'))
else:
    df_RS = df_hospital.groupby('kabupaten_kota', as_index=False).size().reset_index(name='jumlah_rs')
df_RS
```

Proses cleaning pada data hospital ini dimulai dengan melakukan filter berdasarkan provinsi untuk mengambil data yang berasal dari Jawa Barat. selanjutnya kolom 'kab' diubah menjadi 'kabupaten_kota' agar penamaannya sesuai dengan dataset lainnya. proses menghitung jumlah rumah sakit dilakukan menggunakan kolom id dengan fungsi *nunique('id')* untuk memastikan setiap rumah sakit hanya dihitung 1 kali meskipun muncul dalam beberapa baris.

2.2.2.4 mapping nama wilayah

```
set_p = set(penduduk['kabupaten_kota'].astype(str))
set_f = set(faskes['kabupaten_kota'].astype(str))
set_r = set(df_RS['kabupaten_kota'].astype(str))
all_names = sorted(set_p | set_f | set_r)

# Fungsi base name (tanpa prefix)
def base_name(name):
    s = str(name).strip()
    low = s.lower()
    if low.startswith('kota '):
        return s[5:].strip()
    if low.startswith('kabupaten '):
        return s[10:].strip()
    return s

from collections import defaultdict
variants = defaultdict(set)
for n in all_names:
    variants[base_name(n)].add(n)

mapping = {}
for base, vars in variants.items():
    # find canonical existing
    kota = next((v for v in vars if v.lower().startswith('kota ')), None)
```

```

kab = next((v for v in vars if v.lower().startswith('kabupaten ')), None)
bare = next((v for v in vars if not v.lower().startswith('kota ') and not
v.lower().startswith('kabupaten ')), None)

# normalize existing kota/kab names capitalization
if kota:
    mapping[kota] = 'Kota ' + base.title()
if kab:
    mapping[kab] = 'Kabupaten ' + base.title()
if bare:
    # rule: jika kota & kab ada -> map bare -> 'Kabupaten X'
    if kab:
        mapping[bare] = 'Kabupaten ' + base.title()
    elif kota and not kab:
        mapping[bare] = 'Kota ' + base.title()
    else:
        mapping[bare] = 'Kabupaten ' + base.title()

# terapkan mapping function
def apply_map_series(s):
    s = s.astype(str).str.strip()
    s_mapped = s.replace(mapping)
    s_mapped = s_mapped.str.replace(r'\s+', ' ', regex=True).str.strip()
    return s_mapped

penduduk['kabupaten_kota'] = apply_map_series(penduduk['kabupaten_kota'])
faskes['kabupaten_kota'] = apply_map_series(faskes['kabupaten_kota'])
df_RS['kabupaten_kota'] = apply_map_series(df_RS['kabupaten_kota'])

# transformation
penduduk = penduduk.groupby('kabupaten_kota', as_index=False)['jumlah_penduduk'].sum()
faskes = faskes.groupby('kabupaten_kota', as_index=False)['jumlah_faskes'].sum()
df_RS = df_RS.groupby('kabupaten_kota', as_index=False)['jumlah_rs'].sum()

```

Pada proses penyatuan nama wilayah, dilakukan proses normalisasi untuk mengatasi model penulisan seperti “bandung”, “kota bandung”, atau “kabupaten bandung”. fungsi *base_name* digunakan untuk mencari kata asli tanpa prefix sehingga dapat dikelompokkan ke dalam satu baris yang sama. berdasarkan kelompok tersebut dibuatlah mapping otomatis yang menentukan nama yang akan digunakan. jika terdapat prefix “kota” maka wilayah tersebut dianggap kota, jika terdapat prefix “kabupaten” maka dianggap kabupaten, jika tanpa prefix maka dipilih bentuk yang dianggap paling masuk akal berdasarkan data yang tersedia, dengan kecenderungan diarahkan ke “kabupaten” ketika tidak ada informasi lain. Setelah menentukan mapping, fungsi *apply_map_series* digunakan untuk mengganti semua nama lama menjadi nama yang sudah dinormalkan. langkah ini diakhiri dengan menggabungkan baris-baris yang sebelumnya terpisah dengan *groupby().sum()*.

2.2.3 Data Eksplorasi

Tahap Exploratory Data Analysis (EDA) dilakukan untuk memahami pola distribusi data setelah proses pembersihan, penyamaan nama kabupaten/kota, dan penggabungan dataset. Pada tahap ini dilakukan analisis statistik deskriptif dan visualisasi grafik untuk melihat hubungan antar variabel, mendeteksi ketimpangan, serta menemukan pola penyebaran fasilitas kesehatan di Jawa Barat.

2.2.3.1 Perhitungan Kebutuhan Fasilitas Kesehatan

Berdasarkan standar nasional, kebutuhan rumah sakit dihitung dengan rumus 1 rumah sakit per 100.000 penduduk dan 1 fasilitas kesehatan per 30.000 penduduk dihitung dengan membagi jumlahnya dengan rasio dan untuk kekurangan rumah sakit dan fasilitas kesehatan dihitung dengan mengurangi jumlah kebutuhannya dengan jumlahnya, sebagai berikut.

```
rasio_rs = 100000
df['jumlah_rs_dibutuhkan'] = df['jumlah_penduduk'] / rasio_rs
rasio_faskes = 30000
df['jumlah_faskes_dibutuhkan'] = df['jumlah_penduduk'] / rasio_faskes
df['kekurangan_rs'] = df['jumlah_rs_dibutuhkan'] - df['jumlah_rs']
df['kekurangan_faskes'] = df['jumlah_faskes_dibutuhkan'] - df['jumlah_faskes']

print(df[['kabupaten_kota', 'jumlah_penduduk', 'jumlah_rs', 'jumlah_rs_dibutuhkan',
'kekurangan_rs', 'jumlah_faskes', 'jumlah_faskes_dibutuhkan', 'kekurangan_faskes']].head())
```

2.2.3.2 Visualisasi Data

Berbagai grafik dibuat untuk menggambarkan hubungan antara jumlah penduduk dengan jumlah fasilitas kesehatan dan rumah sakit yang tersedia, serta kekurangan fasilitas. Untuk tahap visualisasi ini kami menggunakan pustaka matplotlib, sebagai berikut.

- **Perbandingan Jumlah Rumah Sakit Tersedia dan Dibutuhkan**

Grafik ini menunjukkan ketersediaan rumah sakit dibandingkan dengan kebutuhan di setiap kabupaten/kota di Jawa Barat tahun 2024.

```
df.plot(x="kabupaten_kota", y=["jumlah_rs", "jumlah_rs_dibutuhkan"], kind="bar",
figsize=(12, 6), color=["green", "red"])
plt.title("Perbandingan Jumlah Rumah Sakit yang Dibutuhkan dan Tersedia")
plt.ylabel("Jumlah Penduduk")
plt.xlabel("Kabupaten/Kota")
plt.xticks(rotation=90)
plt.savefig('Grafik Perbandingan Jumlah Rumah Sakit yang Dibutuhkan dan Tersedia.png')
plt.show()
```

- **Perbandingan Jumlah Fasilitas Kesehatan Tersedia dan Dibutuhkan**

Grafik ini menunjukkan jumlah fasilitas kesehatan yang dibutuhkan dibandingkan dengan jumlah yang tersedia.

```
df.plot(x="kabupaten_kota", y=["jumlah_faskes", "jumlah_faskes_dibutuhkan"], kind="bar",
figsize=(12, 6), color=["blue", "orange"])
plt.title("Perbandingan Jumlah Fasilitas Kesehatan yang Dibutuhkan dan Tersedia")
plt.ylabel("Jumlah Fasilitas Kesehatan")
plt.xlabel("Kabupaten/Kota")
plt.xticks(rotation=90)
plt.savefig('Grafik Perbandingan Jumlah Fasilitas Kesehatan yang Dibutuhkan dan Tersedia.png')
plt.show()
```

- **Kekurangan Rumah Sakit per Kabupaten/Kota**

Grafik ini memperlihatkan kekurangan rumah sakit per kabupaten/kota, memberikan gambaran jelas tentang wilayah yang membutuhkan lebih banyak rumah sakit.

```
df.plot(x="kabupaten_kota", y="kekurangan_rs", kind="bar", figsize=(12, 6),
color="purple")
```

```
plt.title("Kekurangan Rumah Sakit per Kabupaten/Kota")
plt.ylabel("Kekurangan Rumah Sakit")
plt.xlabel("Kabupaten/Kota")
plt.xticks(rotation=90)
path = 'Kekurangan Rumah Sakit per Kabupaten-Kota.png'
plt.savefig(path, dpi=300)
plt.show()
```

- **Kekurangan Fasilitas Kesehatan Dasar**

Grafik ini memperlihatkan kekurangan fasilitas kesehatan yang dibutuhkan berdasarkan rasio kebutuhan fasilitas kesehatan dibandingkan dengan yang tersedia.

```
df.plot(x="kabupaten_kota", y="kekurangan_faskes", kind="bar", figsize=(12, 6),
color="orange")
plt.title("Kekurangan Fasilitas Kesehatan per Kabupaten/Kota")
plt.ylabel("Kekurangan Fasilitas Kesehatan")
plt.xlabel("Kabupaten/Kota")
plt.xticks(rotation=90)

path = 'Kekurangan Fasilitas Kesehatan per Kabupaten-Kota.png'
plt.savefig(path, dpi=300)
plt.show()
```

- **Grafik 10 Wilayah dengan Penduduk Terbesar**

Memperlihatkan 10 kabupaten/kota dengan jumlah penduduk terbesar di Jawa Barat tahun 2024.

```
top_penduduk = df.sort_values("jumlah_penduduk", ascending=False).head(10)
plt.figure(figsize=(12,6))
plt.bar(top_penduduk['kabupaten_kota'], top_penduduk['jumlah_penduduk'], color='blue')
plt.title("10 Wilayah dengan Penduduk Terbesar")
plt.xlabel("Kabupaten/Kota")
plt.ylabel("Jumlah Penduduk")
plt.xticks(rotation=60)
plt.savefig('Grafik 10 Wilayah dengan Penduduk Terbesar.png')
plt.show()
```

- **Grafik 10 Wilayah dengan Jumlah Rumah Sakit Terbanyak**

Menunjukkan 10 kabupaten/kota dengan jumlah rumah sakit terbanyak di Jawa Barat tahun 2024.

```
top_rs = df.sort_values("jumlah_rs", ascending=False).head(10)
plt.figure(figsize=(12,6))
plt.bar(top_rs['kabupaten_kota'], top_rs['jumlah_rs'], color='orange')
plt.title("10 Wilayah dengan Jumlah Rumah Sakit Terbanyak")
plt.xlabel("Kabupaten/Kota")
plt.ylabel("Jumlah Rumah Sakit")
plt.xticks(rotation=45)
plt.savefig('Grafik 10 Wilayah dengan Jumlah Rumah Sakit Terbanyak.png')
plt.show()
```

- **Grafik 10 Wilayah dengan Fasilitas Kesehatan Terbanyak**

Grafik ini untuk menampilkan 10 kabupaten/kota yang memiliki fasilitas kesehatan terbanyak di Jawa Barat tahun 2024

```
top_faskes = df.sort_values("jumlah_faskes", ascending=False).head(10)
plt.figure(figsize=(12,6))
plt.bar(top_faskes['kabupaten_kota'], top_faskes['jumlah_faskes'], color='green')
plt.title("10 Wilayah dengan Fasilitas Kesehatan Terbanyak")
plt.xlabel("Kabupaten/Kota")
plt.ylabel("Jumlah Fasilitas Kesehatan")
plt.xticks(rotation=45)
plt.savefig('Grafik 10 Wilayah dengan Fasilitas Kesehatan Terbanyak.png')
plt.show()
```

2.2.4 Data Publishing

Pada penelitian ini, data *publishing* mencakup penyajian komponen utama sebagai berikut. semua file telah kami upload di github: [link](#)

1. Raw Data

Data asli yang kami gunakan dalam project ini tanpa proses pembersihan atau modifikasi. Raw data yang digunakan meliputi:

- Dataset jumlah penduduk per kabupaten/kota.
- Dataset fasilitas kesehatan berdasarkan jenis.
- Dataset rumah sakit Indonesia.

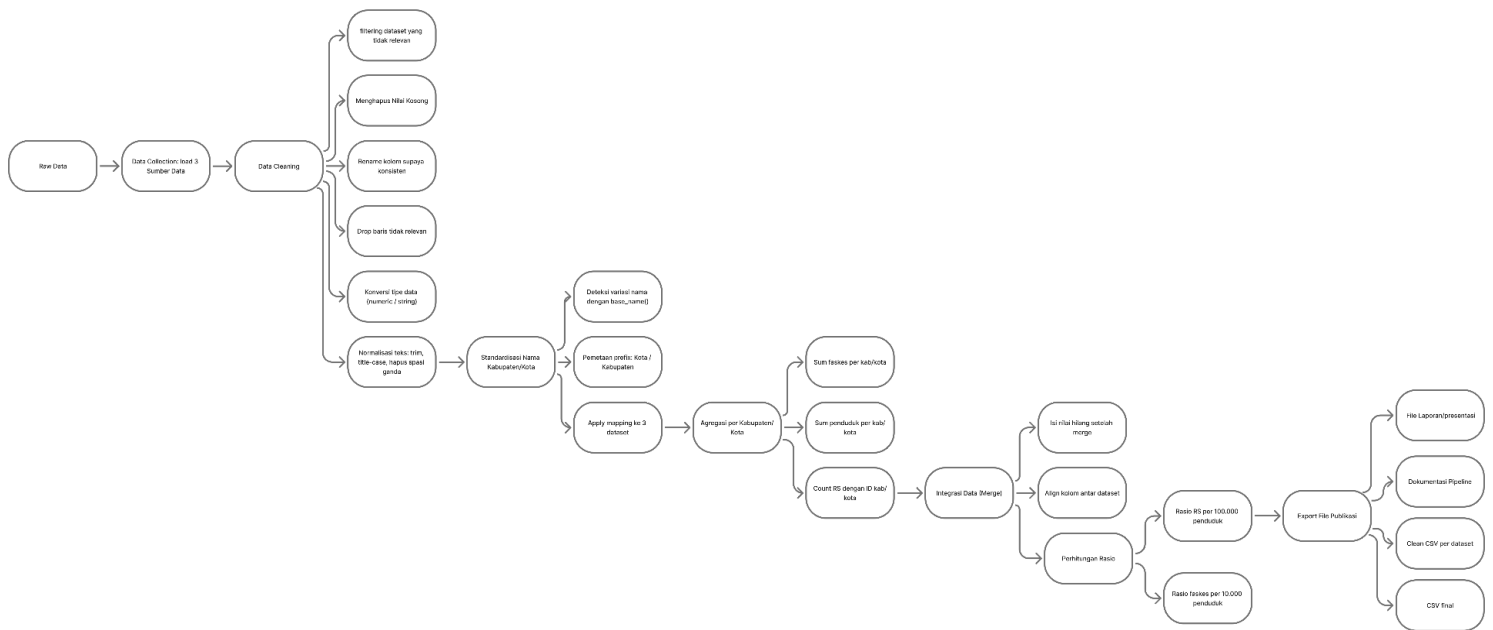
2. Data Hasil Wrangling

Data yang telah melalui proses cleaning, standarisasi nama wilayah, agregasi, integrasi antar dataset, serta perhitungan rasio fasilitas kesehatan dan rumah sakit per 10.000 dan 100.000 penduduk.

Data hasil wrangling terdiri dari:

- Dataset bersih dari masing-masing sumber.
- Dataset terintegrasi per kabupaten/kota.
- Dataset final berisi rasio fasilitas kesehatan dan rumah sakit.

3. Dokumentasi Pipeline



dari diagram alur diatas dapat dijelaskan sebagai berikut:

1. Kami memulai project kali ini dengan mengumpulkan 3 dataset dari 3 sumber yang berbeda. lalu membaca file tersebut untuk diproses
2. Setiap dataset dilakukan cleaning awal yang meliputi filtering, menghapus nilai kosong, menstandarkan nama kolom, mengkonversi tipe data, dll.
3. Selanjutnya dilakukan standarisasi nama kabupaten/kota dengan membuat mapping otomatis yang menentukan prefix kota/kabupaten lalu menerapkan ke seluruh dataset.
4. Ketiga dataset kemudian dilakukan agregasi per wilayah, seperti menjumlahkan fasilitas kesehatan, jumlah penduduk, serta jumlah rumah sakit berdasarkan ID wilayah.
5. Dataset yang sudah dilakukan agregasi kemudian digabungkan/merge dan dilakukan pengecekan missing value, kolom diselaraskan, menghitung rasio rumah sakit per penduduk, dan rasio faskes per penduduk.
6. Hasil akhirnya adalah mengupload dokumentasi project ke github.

2.3 Kendala

Dalam proses pengerjaan project kali ini, kendala utama yang dihadapi dalam project ini adalah ketidakkonsistenan penulisan nama wilayah, khususnya terkait penggunaan prefix “Kota” dan “Kabupaten”. Ketiga dataset yang digunakan memiliki variasi penamaan yang berbeda-beda, misalnya ada yang menuliskan “Kota Bandung”, “Bandung Kota”, “Kabupaten Bandung”, “Bandung Kab.”, atau hanya “Bandung” tanpa prefix. Perbedaan ini menyebabkan proses integrasi data antar dataset menjadi sulit.

2.4 Rencana Tindak Lanjut

Beberapa langkah lanjutan dapat diterapkan untuk memperdalam analisis dan meningkatkan kualitas data sebagai berikut.

1. Perluasan Dataset

Menambahkan variabel lain, seperti jumlah tempat tidur rumah sakit, jumlah tenaga kesehatan, dan anggaran kesehatan untuk analisis yang lebih mendalam dan komprehensif.

2. Analisis Spasial

Menggunakan Geographic Information System (GIS) untuk memetakan distribusi fasilitas kesehatan secara geografis. Ini akan memperjelas pemerataan fasilitas di wilayah yang lebih kecil (kecamatan/desa).

3. Model Statistik dan Machine Learning

Menerapkan model prediksi untuk kebutuhan fasilitas di masa depan menggunakan regresi, serta metode clustering untuk mengelompokkan wilayah berdasarkan demografi dan fasilitas kesehatan.

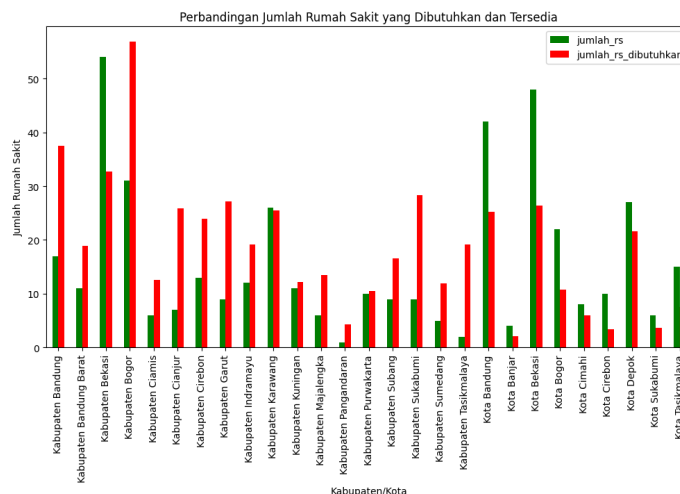
BAB III

HASIL & PEMBAHASAN

3.1 Hasil Dataset Terintegrasi

Setelah melalui proses data wrangling, ketiga dataset utama berhasil digabungkan menjadi satu dataset terintegrasi yang memuat data jumlah penduduk, jumlah fasilitas kesehatan dasar, jumlah rumah sakit, serta perhitungan kebutuhan ideal berdasarkan rasio nasional. Dataset penduduk awalnya bersatuan “ribu”, sehingga dilakukan koreksi skala dengan mengalikan seluruh nilai dengan 1.000 agar menjadi satuan jiwa. Dataset fasilitas kesehatan dasar yang semula terpecah berdasarkan jenis kemudian diagregasi untuk memperoleh total fasilitas per kabupaten/kota. Sementara itu, dataset rumah sakit yang mencakup seluruh Indonesia difilter khusus untuk Provinsi Jawa Barat dan dihitung berdasarkan jumlah ID rumah sakit unik di tiap wilayah. Ketiga dataset ini berhasil disatukan setelah dilakukan standarisasi nama kabupaten/kota menggunakan fungsi *base_name()* dan kamus pemetaan nama sehingga variasi penulisan seperti “Kota Bekasi”, “BEKASI”, atau “Kab. Bekasi” menjadi konsisten. Proses merge dilakukan dengan metode *outer join*, yang memastikan seluruh wilayah tetap muncul meskipun ada data fasilitas yang tidak ditemukan pada salah satu dataset. Hasil integrasi selanjutnya diisi untuk nilai kosong sehingga dataset final siap digunakan sebagai dasar analisis. Dataset akhir ini memberikan gambaran lengkap mengenai kondisi penduduk dan fasilitas kesehatan di seluruh kabupaten/kota Jawa Barat pada tahun 2024.

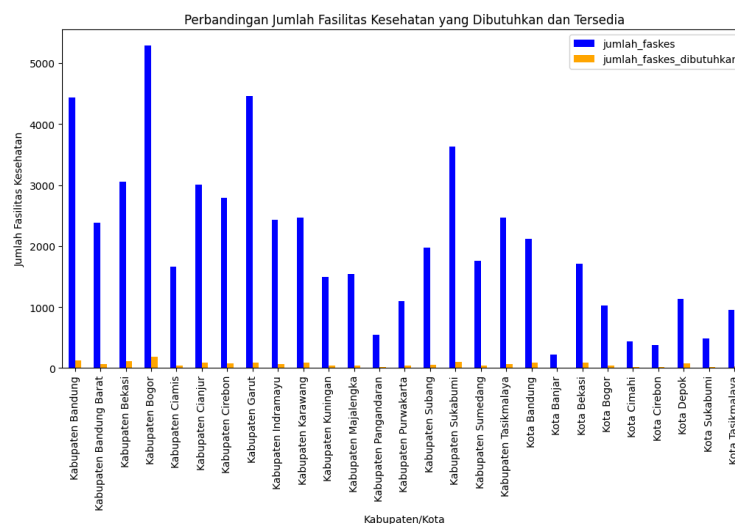
3.2 Perbandingan Jumlah Rumah Sakit Tersedia dan Dibutuhkan



Grafik ini membandingkan jumlah rumah sakit yang tersedia dengan jumlah yang dibutuhkan berdasarkan standar nasional, yaitu satu rumah sakit per seratus ribu penduduk. Grafik perbandingan memperlihatkan dua batang untuk setiap kabupaten/kota, dengan batang hijau menunjukkan jumlah rumah sakit yang tersedia dan batang merah menunjukkan jumlah rumah sakit ideal berdasarkan perhitungan rasio.

Hasilnya menunjukkan bahwa hampir semua wilayah mengalami kekurangan rumah sakit karena batang merah terlihat lebih tinggi daripada batang hijau. Kabupaten Bogor merupakan wilayah dengan kekurangan paling besar karena jumlah penduduknya yang sangat tinggi, yakni lebih dari 5,6 juta jiwa, sehingga kebutuhan rumah sakit ideal mencapai sekitar 56 unit sementara jumlah yang tersedia hanya sekitar 32 unit. Pola serupa juga terlihat di Kabupaten Bekasi dan Kabupaten Bandung yang sama-sama memiliki jumlah penduduk besar tetapi fasilitas yang tersedia tidak berkembang sebanding. Sebaliknya, kota seperti Kota Bandung dan Kota Bekasi memiliki kesenjangan yang lebih kecil, menunjukkan bahwa wilayah perkotaan cenderung lebih baik dalam pemenuhan fasilitas rujukan dibandingkan kabupaten besar. Hal ini menunjukkan bahwa pemerataan rumah sakit masih menjadi tantangan signifikan, terutama di kabupaten yang luas dan memiliki jumlah penduduk tinggi. Ketidakseimbangan ini berpotensi menimbulkan antrean panjang, bottleneck pada layanan IGD, dan peningkatan tekanan terhadap tenaga kesehatan.

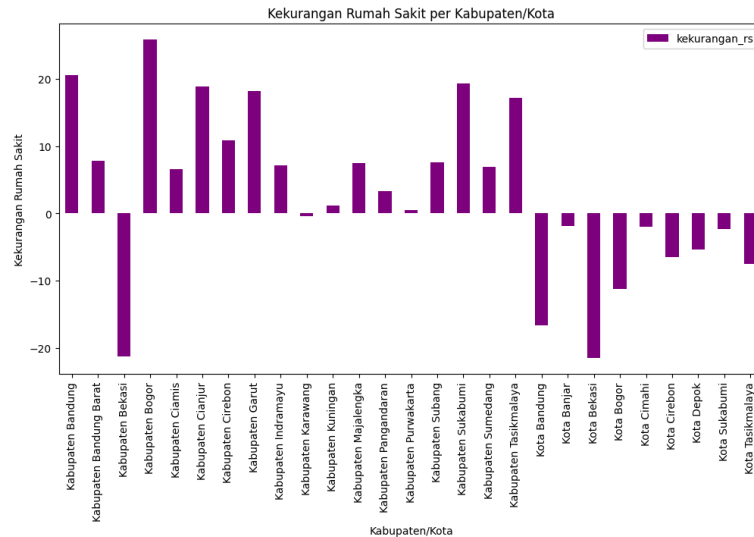
3.3 Perbandingan Jumlah Fasilitas Kesehatan Tersedia dan Dibutuhkan



Grafik ini membandingkan jumlah fasilitas kesehatan yang tersedia dengan kebutuhan ideal berdasarkan rasio satu fasilitas kesehatan per 30.000 penduduk. Grafik ini menampilkan batang biru untuk jumlah fasilitas kesehatan yang tersedia dan batang oranye untuk jumlah fasilitas kesehatan yang dibutuhkan.

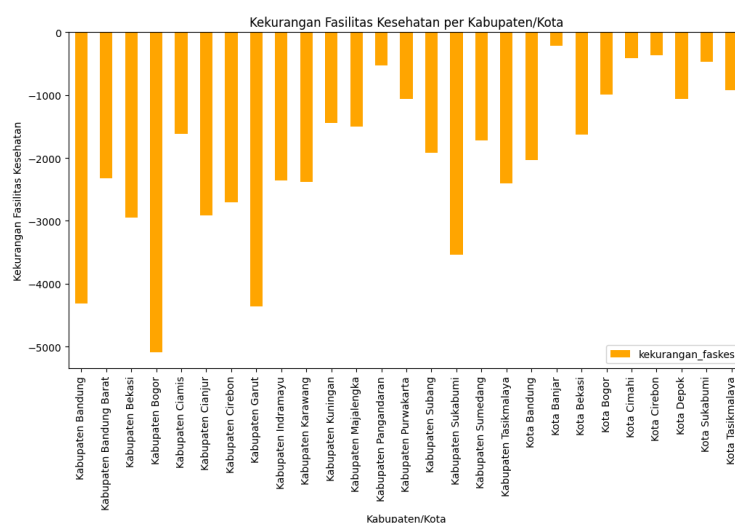
Hasil grafik dan data menunjukkan adanya kelebihan ketersediaan fasilitas kesehatan secara signifikan di hampir semua wilayah yang dianalisis. Secara umum, jumlah fasilitas kesehatan yang tersedia jauh melampaui kebutuhan ideal yang dihitung berdasarkan rasio 1:30.000. Contohnya, di Kabupaten Bogor, ketersediaan fasilitas mencapai 5.284 unit, sementara kebutuhan ideal hanya sekitar 189 unit, menghasilkan surplus kelebihan fasilitas sebesar lebih dari 5.000 unit. Kelebihan jumlah fasilitas kesehatan ini mengindikasikan bahwa secara kuantitas.

3.4 Grafik Kekurangan Rumah Sakit per Kabupaten/Kota



Grafik Kekurangan Rumah Sakit (RS) per Kabupaten/Kota menunjukkan selisih antara kebutuhan ideal dan ketersediaan RS. Dalam konteks ini, nilai positif menunjukkan kekurangan, sedangkan nilai negatif menunjukkan kelebihan RS. Berdasarkan data dan visualisasi, wilayah dengan kekurangan/defisit RS terbesar adalah Kabupaten Bogor ditunjukkan oleh batang positif terpanjang, dengan kekurangan lebih dari 25 unit, diikuti oleh Kabupaten Bandung dan Kabupaten Sukabumi. Kekurangan ini disebabkan oleh jumlah penduduk yang sangat besar di wilayah tersebut yang tidak diimbangi oleh ketersediaan RS yang memadai sesuai rasio 1:100.000. Sebaliknya, wilayah yang menunjukkan kelebihan RS terbesar ditunjukkan oleh batang negatif terpanjang adalah Kabupaten Bekasi, diikuti oleh Kota Bandung dan Kota Bekasi. Hal ini mengindikasikan adanya ketidakseimbangan pemerataan, di mana kabupaten padat penduduk masih sangat kekurangan, sementara wilayah perkotaan tertentu sudah kelebihan fasilitas Rumah Sakit.

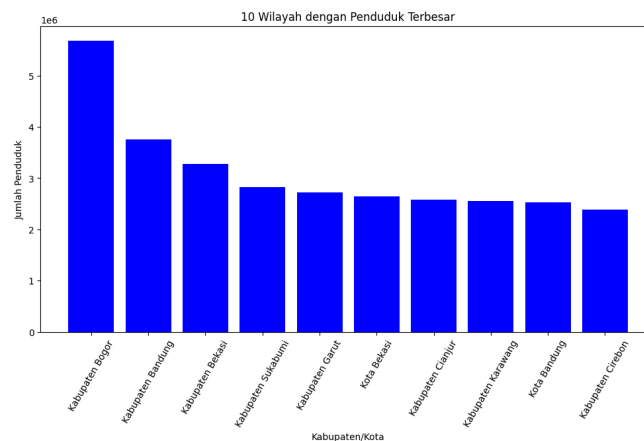
3.5 Grafik Kekurangan Fasilitas Kesehatan Dasar



Grafik Kekurangan Fasilitas Kesehatan Dasar menunjukkan nilai selisih antara kebutuhan ideal dan ketersediaan, namun hasilnya berbanding terbalik dengan rumah sakit.

Semua batang pada grafik ini berada di nilai negatif, mencapai ratusan hingga ribuan unit. Nilai negatif ini mengindikasikan adanya surplus/kelebihan fasilitas kesehatan yang sangat signifikan di semua kabupaten/kota di Jawa Barat. Hal ini menunjukkan bahwa ketersediaan fasilitas kesehatan secara kuantitas jauh melampaui kebutuhan ideal yang ditetapkan oleh rasio 1 faskes per 30.000 penduduk. Kabupaten Bogor, Bandung, Bekasi, Sukabumi, dan Garut merupakan lima wilayah dengan kelebihan fasilitas kesehatan dasar terbesar. Dengan demikian, dapat disimpulkan bahwa secara kuantitas, pelayanan fasilitas kesehatan di Jawa Barat sudah sangat merata dan tersedia dalam jumlah berlebih terhadap kebutuhan minimal.

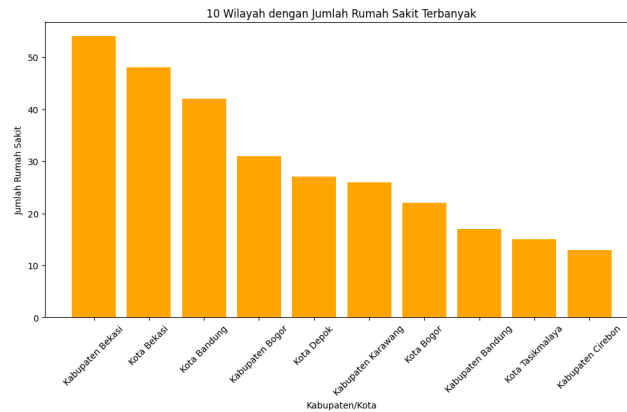
3.6 Grafik 10 Wilayah dengan Penduduk Terbesar



Grafik 10 wilayah dengan jumlah penduduk terbesar menampilkan kabupaten/kota di Jawa Barat yang memiliki populasi paling tinggi pada tahun 2024. Pola yang terlihat pada grafik menunjukkan bahwa Kabupaten Bogor menempati urutan pertama dengan jumlah penduduk lebih dari 5,6 juta jiwa, disusul oleh Kabupaten Bandung, Kabupaten Bekasi, Kabupaten Sukabumi, dan Kabupaten Garut. Di bawahnya terdapat beberapa kabupaten lain yang juga berpenduduk besar, sementara kota seperti Kota Bandung dan Kota Bekasi memiliki jumlah penduduk yang relatif lebih kecil dibandingkan kabupaten besar tersebut. Grafik ini menegaskan bahwa struktur demografi Jawa Barat didominasi oleh kabupaten yang luas dengan jumlah penduduk yang sangat besar.

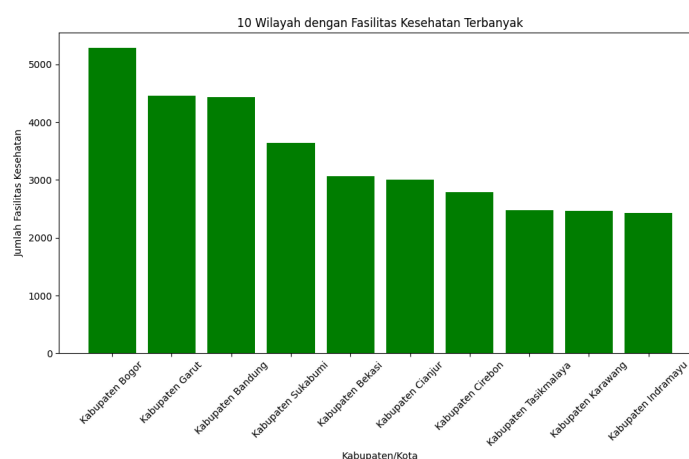
Secara analitis, grafik ini berfungsi sebagai dasar untuk memahami mengapa beberapa kabupaten mengalami kekurangan fasilitas kesehatan yang jauh lebih besar dibandingkan wilayah lain. Semakin besar jumlah penduduk, semakin besar pula kebutuhan rumah sakit dan fasilitas kesehatan. Ketika grafik ini dibandingkan dengan grafik kekurangan fasilitas, terlihat bahwa wilayah dengan penduduk terbesar cenderung menjadi wilayah dengan kekurangan fasilitas kesehatan tertinggi. Dengan demikian, grafik ini bukan hanya menggambarkan distribusi penduduk, tetapi juga menjelaskan mengapa kabupaten-kabupaten tertentu muncul berulang kali sebagai daerah yang paling membutuhkan intervensi dalam hal penyediaan fasilitas kesehatan.

3.7 Grafik 10 Wilayah dengan Jumlah Rumah Sakit Terbanyak



Grafik 10 wilayah dengan jumlah rumah sakit terbanyak memberikan gambaran mengenai distribusi rumah sakit di Jawa Barat. Grafik ini memperlihatkan bahwa Kabupaten Bekasi dan Kota Bekasi merupakan wilayah dengan jumlah rumah sakit terbanyak. Dominasi wilayah ini menunjukkan bahwa layanan rujukan tingkat lanjut cenderung terkonsentrasi di wilayah perkotaan. Meskipun demikian, ketika grafik ini dikaitkan dengan jumlah penduduk dan grafik kekurangan rumah sakit, terlihat bahwa kabupaten berpenduduk sangat besar seperti Kabupaten Bogor tetap mengalami kekurangan rumah sakit. Artinya, meskipun ada banyak rumah sakit di Kabupaten Bogor, angka tersebut masih belum proporsional dengan jumlah penduduk yang dilayani. Di sisi lain, tingginya jumlah RS di Kabupaten Bekasi dan Kota Bekasi justru menciptakan kondisi surplus di wilayah tersebut, karena ketersediaannya telah melampaui rasio ideal. Dengan demikian, grafik sepuluh wilayah dengan rumah sakit terbanyak membantu menjelaskan konsentrasi fasilitas rujukan, tetapi sekaligus menegaskan kembali bahwa indikator kecukupan layanan harus selalu dihubungkan dengan besarnya populasi.

3.8 Grafik 10 Wilayah dengan Fasilitas Kesehatan Terbanyak



Grafik 10 Wilayah dengan Fasilitas Kesehatan Terbanyak menunjukkan dominasi wilayah berpenduduk besar Kabupaten Bogor, Kabupaten Bandung, Kabupaten Bekasi, Kabupaten Garut dalam jumlah absolut fasilitas kesehatan dasar yang tersedia. Jumlah fasilitas kesehatan yang besar secara absolut ini, ketika diuji dengan rasio kebutuhan ideal,

menghasilkan kesimpulan bahwa ketersediaan fasilitas kesehatan dasar di wilayah-wilayah ini telah menciptakan kelebihan yang besar. Hal ini berbeda dengan kondisi Rumah Sakit di mana jumlah yang besar pun masih menghasilkan kekurangan. Dengan adanya surplus fasilitas kesehatan yang melampaui kebutuhan minimal, fokus kebijakan selanjutnya dapat dialihkan pada peningkatan kualitas dan distribusi fasilitas agar manfaat dari ketersediaan yang melimpah ini dapat dioptimalkan.

BAB IV

KESIMPULAN

4.1 Kesimpulan

Berdasarkan proses data wrangling dan analisis eksploratif terhadap tiga dataset yaitu jumlah penduduk tahun 2024, fasilitas kesehatan dasar, dan rumah sakit di Jawa Barat, dapat disimpulkan bahwa:

1. Proses wrangling berhasil menghasilkan dataset yang bersih dan terintegrasi. Setelah normalisasi nama kabupaten/kota, konversi satuan penduduk, agregasi fasilitas kesehatan, dan penggabungan data menggunakan *outer join*, diperoleh dataset final yang mencakup 27 kabupaten/kota dengan kolom-kolom yang lengkap, seperti jumlah penduduk, fasilitas kesehatan, rumah sakit, kebutuhan ideal, dan selisih kekurangan.
2. Terdapat ketidakseimbangan signifikan antara ketersediaan Rumah Sakit dan Fasilitas Kesehatan. Ketersediaan Fasilitas Kesehatan secara kuantitas berada dalam kondisi surplus yang besar di seluruh wilayah, menunjukkan bahwa rasio minimal telah terlampaui. Namun, ketidakseimbangan terjadi pada ketersediaan Rumah Sakit, di mana kabupaten berpenduduk padat seperti Bogor dan Bandung mengalami defisit Rumah Sakit yang signifikan.
3. Analisis terhadap grafik 10 wilayah dengan penduduk terbesar, fasilitas terbanyak, dan rumah sakit terbanyak menunjukkan pola konsisten: semakin besar jumlah penduduk, semakin besar beban fasilitas kesehatan yang harus dipenuhi. Meskipun beberapa kota memiliki fasilitas lebih banyak, kebutuhan mereka lebih rendah karena jumlah penduduk yang lebih sedikit, sehingga kota lebih mendekati rasio ideal.

Secara keseluruhan, ketersediaan fasilitas kesehatan di Jawa Barat pada tahun 2024 belum proporsional dengan beban penduduk, terutama di kabupaten besar. Proses data wrangling yang dilakukan berhasil mengidentifikasi kesenjangan ini, dan hasil EDA memberikan gambaran yang jelas mengenai wilayah-wilayah yang membutuhkan prioritas pembangunan fasilitas kesehatan.

4.2 Kontribusi Anggota

Nama	NIM	Kontribusi
Eka Putri Maharani	24031554121	<ul style="list-style-type: none">• code(cleaning, EDA),• laporan(pendahuluan, EDA, hasil, dan kesimpulan),• ppt(pembahasan grafik, kendala, kesimpulan)
Muslim Fazlur Rohman	24031554154	<ul style="list-style-type: none">• code(dataset, mapping, integrasi, agregasi),• laporan(pengambilan dan integrasi, cleaning, publishing),• ppt(pendahuluan, diagram)

REFERENSI

- Dewi, F. R., Marthias, T., Riyani, I. S., Kusumawardani, N., & Mahendradhata, Y. (2023). *Spatial evaluation of healthcare accessibility across archipelagic communities of Indonesia*. PLOS Global Public Health, 3(5), e0001600. <https://doi.org/10.1371/journal.pgph.0001600>
- Kementerian Kesehatan Republik Indonesia. (2024). *Profil Kesehatan Indonesia 2024*. Kemenkes RI. <https://kemkes.go.id/id/profil-kesehatan-indonesia-2024>
- McKinney, W. (2017). *Python for data analysis* (2nd ed.). O'Reilly Media. <https://www.lkhibra.ma/books/Python-for-Data-Analysis.pdf>
- Kementerian Kesehatan Republik Indonesia. (2014). *Peraturan Menteri Kesehatan No. 75 Tahun 2014 tentang Pusat Kesehatan Masyarakat (Puskesmas)*. Kementerian Kesehatan Republik Indonesia.
- Badan Pusat Statistik Kota Bandung. (2024). Jumlah penduduk menurut kabupaten/kota di Provinsi Jawa Barat. Badan Pusat Statistik Kota Bandung. <https://bandungkab.bps.go.id/id/statistics-table/2/MjgwIzI=/jumlah-penduduk-menurut-kabupaten-kota-di-provinsi-jawa-barat.html>
- Habibna, M. (2022). Hospital data in Indonesia. Kaggle. <https://www.kaggle.com/datasets/muhammadhabibna/hospital-data-in-indonesia>
- Pemerintah Provinsi Jawa Barat. (t.t.). Jumlah fasilitas kesehatan berdasarkan jenis di Jawa Barat [Kumpulan data]. Open Data Jawa Barat. Diperoleh dari <https://opendata.jabarprov.go.id/id/dataset/jumlah-fasilitas-kesehatan-berdasarkan-jenis-di-jawa-barat>