

**ANALISIS HUBUNGAN TINGKAT KEMISKINAN DAN
COVERAGE IMUNISASI DASAR DI INDONESIA TAHUN
2023: STUDI KASUS 38 PROVINSI**

Proposal Revisi Proyek ini Disusun untuk Memenuhi Ujian Akhir Semester Mata
Kuliah Data Wrangling



Disusun Oleh:

Tazkia Caecaria Marchanda	24031554002
Meida Rifa Alfiani	24031554139

Dosen Pengampu

Ulfa Siti Nuraini, S.Stat., M.Stat.

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS NEGERI SURABAYA

SURABAYA

2025

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa karena atas rahmat dan karunia-Nya, makalah penelitian yang berjudul “*Analisis Hubungan Tingkat Kemiskinan dan Coverage Imunisasi Dasar di Imunisasi Dasar di Indonesia Tahun 2023: Studi Kasus 38 Provinsi* ” ini dapat diselesaikan dengan baik.

Proposal ini bertujuan untuk menganalisis korelasi antara tingkat kemiskinan dan cakupan imunisasi dasar lengkap di 34 provinsi Indonesia, dengan fokus pada disparitas layanan kesehatan. Melalui pengumpulan, pembersihan, dan transformasi data dari sumber resmi seperti BPS, Kementerian Kesehatan, dan Open Data Jabar, proyek ini bertujuan memberikan wawasan berbasis data untuk perumusan kebijakan yang lebih tepat. Proyek ini juga menjadi implementasi dari mata kuliah data wrangling, dengan menekankan pada proses pengolahan data yang cermat untuk memastikan validitas dan reliabilitas analisis korelasi spasial.

Penulis menyadari bahwa masih banyak kekurangan dalam makalah ini. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi penyempurnaan karya ilmiah ini di masa mendatang. Besar harapan penulis bahwa penelitian ini dapat memberikan manfaat, baik secara teoretis maupun praktis, terutama dalam mendukung kebijakan yang lebih terarah di Indonesia.

Akhir kata, penulis mengucapkan terima kasih kepada semua pihak yang telah mendukung dalam proses penyusunan makalah ini. Semoga karya ini dapat memberikan kontribusi positif bagi pengembangan ilmu pengetahuan dan peningkatan kesejahteraan masyarakat.

Surabaya, November 2025

Penulis

BAB I

PENDAHULUAN

1.1 Latar belakang

Menurut *World Health Organization* (2025) dan Kementerian Kesehatan Republik Indonesia (2025), imunisasi dasar lengkap merupakan intervensi kesehatan masyarakat yang terbukti paling efektif dan efisien dalam mencegah morbiditas, mortalitas, serta disabilitas akibat Penyakit yang Dapat Dicegah Dengan Imunisasi (PD3I). Keberhasilan program imunisasi memegang peran krusial dalam mewujudkan generasi yang sehat dan produktif, sekaligus menjadi indikator penting bagi kualitas sistem kesehatan suatu negara. Oleh karena itu, menjamin akses universal terhadap imunisasi dasar lengkap bagi seluruh anak menjadi prioritas utama dalam agenda pembangunan kesehatan baik pada tingkat global maupun nasional. Upaya strategis ini berkontribusi langsung terhadap pencapaian Tujuan Pembangunan Berkelanjutan (SDGs), khususnya pada poin ketiga mengenai Kehidupan Sehat dan Sejahtera.

Namun demikian, upaya pencapaian cakupan imunisasi yang merata seringkali mengalami tantangan akibat berbagai faktor sosial ekonomi. Kemiskinan, khususnya, telah diakui secara luas sebagai determinan sosial kesehatan yang signifikan, yang memengaruhi akses rumah tangga terhadap layanan kesehatan termasuk program imunisasi baik secara langsung maupun tidak langsung (Hargono, 2020). Rumah tangga dengan status ekonomi rendah umumnya menghadapi kendala finansial, geografis, dan keterbatasan informasi yang lebih besar, yang pada akhirnya berdampak pada rendahnya tingkat pemanfaatan layanan kesehatan preventif. Sebagai negara kepulauan dengan karakteristik geografis dan kondisi sosial ekonomi yang sangat beragam, Indonesia menghadapi tantangan kompleks dalam mewujudkan pemerataan layanan kesehatan. Kondisi ini berpotensi menimbulkan disparitas cakupan imunisasi yang signifikan antardaerah. Data Profil Kesehatan Indonesia tahun 2023 mengonfirmasi bahwa meskipun capaian nasional telah mencapai 89,1%, masih terdapat variasi antarwilayah yang memerlukan kajian lebih mendalam.

Proyek analisis data ini bertujuan untuk menguji hubungan empiris antara tingkat kemiskinan dan cakupan imunisasi dasar lengkap di tingkat provinsi di

Indonesia pada tahun 2023. Fokus analisis pada disparitas antardaerah ini diharapkan dapat menjembatani kesenjangan pemahaman mengenai keterkaitan antara faktor ekonomi makro regional dengan capaian kesehatan preventif. Kajian ini selaras dengan komitmen Indonesia dalam mencapai Tujuan Pembangunan Berkelanjutan, khususnya pada poin pertama mengenai Penghapusan Kemiskinan dan poin ketiga tentang Kehidupan Sehat dan Sejahtera. Melalui penyediaan wawasan berbasis data mengenai hubungan antara kedua variabel tersebut, hasil penelitian ini diharapkan dapat menjadi landasan yang kokoh bagi perumusan kebijakan kesehatan masyarakat yang lebih terarah serta desain intervensi program yang tepat sasaran untuk mengurangi kesenjangan cakupan imunisasi di provinsi provinsi dengan prevalensi kemiskinan tinggi.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang di atas, rumusan masalah penelitian ini adalah:

1. Apakah terdapat korelasi hubungan yang signifikan secara spesifik antara tingkat kemiskinan dan cakupan imunisasi dasar lengkap di tingkat provinsi di Indonesia pada tahun 2023?
2. Seberapa kuat dan dalam arah bagaimana hubungan tersebut, apakah positif atau negatif, antara variabel kemiskinan dan cakupan imunisasi dasar lengkap pada tingkat provinsi?
3. Provinsi-provinsi mana saja yang menunjukkan perbedaan signifikan terbesar antara tingkat kemiskinan yang tinggi dan cakupan imunisasi yang rendah, yang mengindikasikan perlunya perhatian kebijakan yang lebih mendesak?
4. Bagaimana teknik visualisasi data dapat diterapkan untuk mempermudah pemahaman mengenai pola hubungan dan disparitas spasial antara tingkat kemiskinan dan cakupan imunisasi dasar lengkap di Indonesia?
5. Rekomendasi kebijakan berbasis data apa yang dapat dirumuskan untuk mengatasi kesenjangan cakupan imunisasi di provinsi-provinsi dengan tingkat kemiskinan yang tinggi?

1.3 Tujuan

Berdasarkan rumusan masalah yang telah ditetapkan, tujuan utama penelitian ini adalah untuk Menganalisis Dampak Tingkat Kemiskinan Terhadap Cakupan Imunisasi Dasar Lengkap di Indonesia pada Tahun 2023. Tujuan utama tersebut dijabarkan ke dalam beberapa tujuan spesifik berikut:

1. Menganalisis Hubungan Korelasi

Melakukan pengujian statistik terhadap hubungan antara tingkat kemiskinan dan cakupan imunisasi dasar lengkap di seluruh provinsi Indonesia tahun 2023, dengan penekanan pada identifikasi disparitas antardaerah.

2. Mengukur Kekuatan dan Signifikansi Hubungan

Menghitung kekuatan dan signifikansi statistik dari hubungan antara kedua variabel menggunakan metode statistik yang relevan seperti koefisien korelasi Pearson atau Spearman.

3. Mengidentifikasi Area Prioritas

Mengidentifikasi dan memetakan provinsi-provinsi yang menunjukkan disparitas signifikan antara tingkat kemiskinan dan cakupan imunisasi sebagai target prioritas intervensi kebijakan.

4. Mengembangkan Visualisasi Data

Menyajikan temuan analisis melalui visualisasi data yang informatif dan mudah dipahami, termasuk scatter plot dan peta tematik, untuk memfasilitasi pemahaman pola hubungan antar variabel.

5. Merumuskan Rekomendasi Kebijakan

Menyusun rekomendasi kebijakan dan program berbasis bukti yang spesifik untuk mengurangi kesenjangan cakupan imunisasi di provinsi-provinsi dengan tantangan kemiskinan terbesar.

1.4 Manfaat

Penelitian ini diharapkan dapat memberikan kontribusi signifikan baik dalam ranah teoritis maupun praktis. Secara teoritis, kajian ini memperkaya khazanah ilmu kesehatan masyarakat dan studi determinan sosial kesehatan dengan menyajikan bukti empiris mengenai hubungan antara tingkat kemiskinan dan disparitas cakupan imunisasi dalam konteks keragaman antarprovinsi di

Indonesia, sekaligus menjadi landasan bagi penelitian lanjutan untuk mengeksplorasi variabel mediasi seperti tingkat pendidikan dan akses infrastruktur kesehatan. Dari sisi praktis, temuan penelitian ini menyediakan dasar *evidence* untuk perumusan kebijakan kesehatan yang lebih terarah bagi pemerintah pusat dalam mengoptimalkan alokasi sumber daya dan program imunisasi di daerah dengan beban kemiskinan tinggi, memberikan gambaran spesifik bagi dinas kesehatan provinsi dalam merancang intervensi yang tepat sasaran, serta menjadi studi kasus berharga bagi akademisi dan praktisi data dalam proses data wrangling dan analisis korelasi menggunakan data publik.

BAB II

PEMBAHASAN

2.1 Teknik Pengambilan Data

Teknik Pengambilan Data (*Acquisition*) pada code kami menggunakan operasi *pdfplumber.open()* yang memilih halaman pada (*pdf.pages[726]*) dataset imunisasi dan *pd.read_csv()* yang merupakan inti dari pengambilan data, yaitu mengakses data dari sumber eksternal (file lokal PDF dan CSV). Lalu, kami memvisualisasikan halaman tersebut dengan mengubah halaman menjadi gambar menggunakan (*page.to_image()*) dan menampilkannya dengan (*plt.imshow()*) yang bertujuan untuk memastikan bahwa halaman yang dipilih benar-benar berisi tabel yang dimaksud.

2.1.1 Integrasi Data

Sebelum di lakukan integrasi menggunakan *pd.merge* kami melakukan pembersihan dan transformasi data (*cleaning & transformation*) yang dimana mayoritas baris kode setelah pengambilan data adalah operasi pembersihan data dan transformasi data. Langkah ini termasuk:

1. Ekstraksi teks dari PDF yang tidak terstruktur

Mengubah data dari format dokumen pdf menjadi format yang dapat diolah (teks mentah/list baris)

2. Pemfilteran baris yang tidak relevan.

Menghilangkan baris-baris yang bukan merupakan data aktual (misalnya, judul tabel, keterangan, atau baris kosong) pada dataset 3 17169067256655eae5553985.98376730.pdf

3. Penggunaan (*re.search(r'd', line)*)

Bertujuan hanya memilih baris yang mengandung angka, sambil mengabaikan baris yang merupakan judul, header, atau keterangan pada dataset 3 17169067256655eae5553985.98376730.pdf.

4. *df = df[df[col].str.match(r"^[0-9,.*]+\$")]*

Validasi Nilai: Memastikan bahwa kolom yang seharusnya berisi angka tidak mengandung teks yang tidak terduga. Baris yang gagal divalidasi akan dihapus.

5. `df["N"] = df["N"].str.replace("*", "", regex=False)`

Menghapus karakter khusus (misalnya, tanda bintang * yang mungkin menandakan catatan kaki) dari kolom N.

6. `df[col].str.replace(",", ".").astype(float)`

Penyeragaman Format & Tipe Data (*Standardization*) yaitu mengubah pemisah desimal dari koma (,) menjadi titik (.) dan kemudian mengubah tipe data kolom menjadi float (angka desimal). Ini adalah langkah wajib agar data dapat dihitung.

7. `df["N"] = df["N"].str.replace(".", "", regex=False).astype(int)`

Penyeragaman Format & Tipe Data (*Standardization*) yaitu menghapus titik pemisah ribuan (jika ada) dan mengubah tipe data kolom N menjadi integer (bilangan bulat).

8. Pemfilteran baris berdasarkan kriteria (`df["tahun"] == 2023`).

Hal ini bertujuan untuk membuang data yang tidak relevan dengan judul agar dataset menjadi bersih dan fokus pada tahun yang dibutuhkan yaitu tahun 2023 pada dataset *I bps-od_20948_persentase_penduduk_miskin_prov_di_indonesia_v2_data.csv*

9. Penggunaan (`re.search(r'\d', line)`)

Berguna untuk memilih baris yang hanya mengandung angka, sambil mengabaikan baris yang merupakan judul, header, atau keterangan pada dataset *I bps-od_20948_persentase_penduduk_miskin_prov_di_indonesia_v2_data.csv*

10. `df.rename(columns={'Provinsi': 'nama_provinsi'})`

menyeragamkan nama kolom kunci di semua dataset menjadi satu nama yang konsisten (misalnya, `nama_provinsi`).

11. `df['nama_provinsi'].apply(normalisasi_provinsi)`

Berguna untuk memastikan semua variasi penulisan diubah menjadi satu format standar (misalnya, semua menjadi DI YOGYAKARTA) pada ketiga dataset.

2.2 Data Cleaning

Proses data cleaning dilakukan melalui 6 tahap, yaitu:

1. Standarisasi nama provinsi

mengubah nama provinsi menjadi format kapitalisasi

2. Menghapus kolom yang tidak perlu

menghapus kolom redundan: kolom satuan dan id dihapus karena tidak memberikan informasi yang berguna untuk analisis. Kolom satuan hanya berisi satu nilai ('PERSEN'), dan kolom id biasanya hanya berfungsi sebagai pengenalan unik.

menghapus kolom duplikat: kode tersebut secara cerdas memverifikasi bahwa kolom *persentase_penduduk_miskin* dan Persentase Penduduk Miskin - Maret memiliki isi yang identik. Karena isinya sama persis, salah satu kolom dihapus untuk menghindari duplikasi data dan menjaga agar dataset tetap ringkas.

```
... Nilai unik di kolom 'satuan': ['PERSEN']
✓ Kolom 'satuan' dihapus
✓ Kolom 'id' dihapus
✓ Kolom 'tahun' dihapus
Perbedaan nilai antara kedua kolom: 38
```

Dalam proses pembersihan data (data cleaning), kolom “tahun” dihapus karena seluruh variabel dalam dataset telah merepresentasikan kondisi pada tahun 2023. Dengan demikian, kolom tersebut tidak memberikan informasi tambahan (*redundant variable*) dan tidak berkontribusi terhadap variabilitas data.

3. Standarisasi nama kolom

menggunakan metode *df.rename()* dengan sebuah *dictionary* untuk memetakan nama kolom lama ke nama baru dan memilih format *snake_case*, yaitu menggunakan huruf kecil semua dan memisahkan kata dengan garis bawah (_). Ini adalah praktik standar dalam *Python* yang membuat akses kolom menjadi lebih mudah (misalnya, *df.nama_kolom* vs *df['Nama Kolom (unit)']*) dan meningkatkan keterbacaan kode.

Semua kolom yang relevan sekarang memiliki nama yang seragam, pendek, dan mudah digunakan.

```
# 3 Standarisasi Nama Kolom
kolom_baru = {
    'Garis Kemiskinan - Maret (Rp)': 'garis_kemiskinan_maret',
    'Garis Kemiskinan - September (Rp)': 'garis_kemiskinan_september',
    'Jumlah Penduduk Miskin - Maret (ribu) (Ribu)': 'jumlah_penduduk_m',
    'Jumlah Penduduk Miskin - September (ribu) (Ribu)': 'jumlah_pendud
    'Persentase Penduduk Miskin - September': 'persentase_penduduk_mis
    'Persentase Penduduk Miskin': 'persentase_penduduk_miskin'
}
df = df.rename(columns=kolom_baru)
print("✓ Nama kolom distandarisasi")

✓ Nama kolom distandarisasi

# Standardisasi nama kolom (snake_case)
kolom_baru = {
    'HB0': 'imunisasi_hb0',
    'BCG': 'imunisasi_bcg',
    'DPT_HB_Hib': 'imunisasi_dpt_hb_hib',
    'Polio': 'imunisasi_polio',
    'Campak_Rubella': 'imunisasi_campak_rubella',
    'N': 'jumlah_sampel_imunisasi'
}
df = df.rename (columns=kolom_baru)
print("✓ Nama kolom distandarisasi")

✓ Nama kolom distandarisasi
```

4. Penanganan Nilai yang Hilang (Handling missing values)

menggunakan `df.isnull().sum()` untuk menghitung jumlah nilai yang hilang (*missing values*) di setiap kolom dan output menunjukkan 0 untuk semua kolom hal ini berarti tidak ada nilai NaN (*Not a Number*) atau None di dalam *dataframe*. Provinsi seperti 'Papua Barat Daya', 'Papua Selatan', 'Papua Tengah', dan 'Papua Pegunungan' adalah provinsi baru sangat mungkin data historis (seperti data kemiskinan) belum tercatat untuk provinsi tersebut sebagai entitas terpisah, sehingga diisi dengan 0.0 sebagai placeholder mengidentifikasi provinsi-provinsi ini dan secara eksplisit mengubah nilai 0.0 di kolom `persentase_penduduk_miskin` menjadi `np.nan`, hal ini karena 0.0 adalah angka valid yang bisa berarti "0% penduduk miskin", yang jelas tidak akurat serta mengubahnya menjadi NaN dengan benar menandai data ini sebagai "hilang" atau "tidak tersedia".

Not a Number (NaN)

NaN digunakan secara standar dalam ilmu data untuk menandai nilai yang hilang (missing value) atau data yang tidak tersedia dalam suatu kolom. NaN bukan angka, bukan nol, dan bukan pula error melainkan penanda resmi bahwa suatu observasi tidak memiliki nilai yang valid.

Jika sebuah provinsi memiliki NaN pada kolom seperti:

- Garis Kemiskinan – September
- Jumlah Penduduk Miskin – September
- Persentase Penduduk Miskin – September

Maka ini berarti data kemiskinan bulan September tidak tersedia untuk provinsi tersebut (umumnya karena BPS belum merilis atau data tidak tercantum dalam file asli)

```
=== HANDLING MISSING VALUES ===

Missing values per kolom:
kode_provinsi                0
nama_provinsi                0
persentase_penduduk_miskin   0
garis_kemiskinan_maret       0
garis_kemiskinan_september   0
jumlah_penduduk_miskin_maret  0
jumlah_penduduk_miskin_september 0
Persentase Penduduk Miskin – Maret 0
persentase_penduduk_miskin_september 0
imunisasi_hb0                0
imunisasi_bcg                0
imunisasi_dpt_hb_hib         0
imunisasi_polio              0
imunisasi_campak_rubella     0
jumlah_sampel_imunisasi      0
dtype: int64

Provinsi dengan missing values:
[]

Total missing values: 0
```

5. Konversi tipe data yang benar

Output `df.dtypes` menunjukkan bahwa semua kolom dalam daftar kolom_numerik telah berhasil diubah menjadi tipe data numerik (float64 atau int64).

Terlihat ada satu kolom yang tidak terkonversi, yaitu 'Persentase Penduduk Miskin - Maret' yang masih bertipe *object*. Ini kemungkinan besar karena kolom ini sudah dihapus pada langkah sebelumnya (seperti yang kita diskusikan), tetapi mungkin masih ada dalam beberapa versi *DataFrame* Anda. Kode Anda untuk mengurutkan kolom di langkah #6 akan secara efektif menghapusnya dari hasil akhir.

```
✓ Konversi tipe data selesai

Tipe data:
kode_provinsi           int64
nama_provinsi           object
persentase_penduduk_miskin  float64
garis_kemiskinan_maret   float64
garis_kemiskinan_september float64
jumlah_penduduk_miskin_maret float64
jumlah_penduduk_miskin_september float64
Persentase Penduduk Miskin – Maret  object
persentase_penduduk_miskin_september float64
imunisasi_hb0           float64
imunisasi_bcg           float64
imunisasi_dpt_hb_hib    float64
imunisasi_polio         float64
imunisasi_campak_rubella float64
jumlah_sampel_imunisasi  int64
dtype: object
```

6. Reorder/mengurutkan nama kolom

DataFrame `df` sekarang memiliki struktur kolom yang rapi dan terorganisir sesuai dengan urutan yang ditentukan. Ini juga secara efektif menghapus kolom-kolom yang tidak diinginkan yang tidak Anda sertakan dalam daftar kolom_terurut (seperti 'Persentase Penduduk Miskin - Maret').

2.3 Data Exploration

EDA (*Exploratory Data Analysis*) adalah sebuah proses investigasi sistematis untuk "berkenalan" dengan data, yang dilakukan setelah data dibersihkan. Tujuannya adalah untuk menggali, meringkas, dan memvisualisasikan karakteristik utama data guna menemukan pola, anomali, dan hubungan tersembunyi, yang pada akhirnya menghasilkan wawasan (*insight*) dan rekomendasi yang bisa ditindaklanjuti.

Dalam tahap EDA, kami melakukan investigasi mendalam terhadap data kemiskinan dan imunisasi yang telah dibersihkan untuk mengungkap pola, anomali, dan wawasan tersembunyi. Proses ini kami jalankan melalui serangkaian operasi yang sistematis:

1. Load Data dan Overview Awal

Tahap pertama adalah memuat data hasil cleaning (*hasil_cleaning_revisi.csv*) dan melakukan pemeriksaan awal untuk memahami struktur dataset. Kami menampilkan informasi dasar seperti jumlah baris (38 provinsi), jumlah kolom (14 kolom), daftar nama kolom yang tersedia, serta tipe data setiap kolom menggunakan `df.info()`. Langkah ini penting untuk memastikan data telah siap dianalisis dan memahami karakteristik dasar dataset

```

✓ Data berhasil dimuat!
  - Jumlah baris (provinsi): 38
  - Jumlah kolom: 14

✓ Kolom yang tersedia:
  1. kode_provinsi
  2. nama_provinsi
  3. persentase_penduduk_miskin
  4. garis_kemiskinan_maret
  5. garis_kemiskinan_september
  6. jumlah_penduduk_miskin_maret
  7. jumlah_penduduk_miskin_september
  8. persentase_penduduk_miskin_september
  9. imunisasi_hb0
 10. imunisasi_bcg
 11. imunisasi_dpt_hb_hib
 12. imunisasi_polio
 13. imunisasi_campak_rubella
 14. jumlah_sampel_imunisasi

```

kode_provinsi	nama_provinsi	persentase_pendudu
11	Aceh	14.45
12	Sumatera Utara	8.15
13	Sumatera Barat	5.95
14	Riau	6.68
15	Jambi	7.58
16	Sumatera Selatan	11.78
17	Bengkulu	14.04
18	Lampung	11.11

Keterangan kolom:

1. kode_provinsi
: merupakan kode numerik yang ditetapkan oleh Badan Pusat Statistik (BPS) untuk mengidentifikasi masing-masing provinsi. Nilai ini bersifat administratif dan tidak memiliki makna statistik kuantitatif selain sebagai penanda wilayah.
2. nama_provinsi
: menunjukkan nama provinsi tempat data dikumpulkan. Variabel ini bersifat kategorikal dan digunakan untuk menghubungkan data imunisasi dengan data kemiskinan pada wilayah yang sama.

3. `persentase_penduduk_miskin`

: menyatakan persentase penduduk yang berada di bawah garis kemiskinan di masing-masing provinsi. Nilai ini dihitung oleh BPS berdasarkan proporsi penduduk dengan pengeluaran per kapita bulanan yang berada di bawah garis kemiskinan. Satuan yang digunakan adalah persen (%).

<code>garis_kemiskinan_maret</code>	<code>garis_kemiskinan_september</code>	<code>jumlah_penduduk_nasional</code>	<code>jumlah_penduduk_miskin</code>	<code>persentase_penduduk_miskin</code>
627534	NaN	806.75	NaN	NaN
602999	NaN	1239.71	NaN	NaN
667925	NaN	340.37	NaN	NaN
658611	NaN	485.66	NaN	NaN
599688	NaN	280.68	NaN	NaN
520754	NaN	1045.68	NaN	NaN
637142	NaN	288.46	NaN	NaN
559011	NaN	970.67	NaN	NaN
674704	NaN	69.60	NaN	NaN

4. `garis_kemiskinan_maret`

: dua variabel ini mencerminkan nilai garis kemiskinan per kapita per bulan (dalam Rupiah) pada bulan Maret. Garis kemiskinan mencakup komponen pengeluaran makanan dan non-makanan yang dibutuhkan untuk memenuhi kebutuhan dasar. Individu dengan pengeluaran di bawah nilai ini dikategorikan sebagai miskin. Nilai dinyatakan dalam Rupiah (Rp), bukan persen.

5. `garis_kemiskinan_september`

: dalam proses pembersihan data, seluruh nilai yang ditandai dengan ‘...’ diganti menjadi NaN (Not a Number). NaN digunakan sebagai penanda standar untuk data yang hilang (missing values). Dengan mengganti nilai ‘...’ menjadi NaN, analisis statistik dapat dilakukan secara lebih akurat karena nilai tersebut tidak akan diperlakukan sebagai angka valid dan tidak akan mempengaruhi perhitungan rata-rata, korelasi, maupun pemodelan.

6. jumlah_penduduk_miskin_maret

: variabel ini menunjukkan estimasi jumlah penduduk miskin pada bulan maret periode tersebut. Variabel ini menunjukkan jumlah penduduk miskin dalam satuan ribu orang. Misalnya, nilai 806.75 menunjukkan bahwa terdapat sekitar 806.750 orang yang tergolong miskin pada periode tersebut. Satuan "ribu" digunakan untuk mempermudah penyajian angka.

7. jumlah_penduduk_miskin_september

: dalam proses pembersihan data, seluruh nilai yang ditandai dengan '...' diganti menjadi NaN (Not a Number). NaN digunakan sebagai penanda standar untuk data yang hilang (missing values). Dengan mengganti nilai '...' menjadi NaN, analisis statistik dapat dilakukan secara lebih akurat karena nilai tersebut tidak akan diperlakukan sebagai angka valid dan tidak akan mempengaruhi perhitungan rata-rata, korelasi, maupun pemodelan.

8. persentase_penduduk_miskin_september

: dalam proses pembersihan data, seluruh nilai yang ditandai dengan '...' diganti menjadi NaN (Not a Number). NaN digunakan sebagai penanda standar untuk data yang hilang (missing values). Dengan mengganti nilai '...' menjadi NaN, analisis statistik dapat dilakukan secara lebih akurat karena nilai tersebut tidak akan diperlakukan sebagai angka valid dan tidak akan mempengaruhi perhitungan rata-rata, korelasi, maupun pemodelan.

imunisasi_hb0	imunisasi_bcg	imunisasi_dpt_hb_hi	imunisasi_polio	imunisasi_campak_r	jumlah_sampel_imu
49.3	35.9	22.8	5.6	18.5	359
69.4	72.2	60.5	26.6	51.6	1025
75.3	70.6	50.9	20.7	42.4	351
74.5	71.1	57.3	18.4	46.1	445
86.3	89	79.2	31.5	66.5	216
83	87.7	77	29.5	68.2	542
85.6	91.7	83.8	39.9	66.5	124
87.9	89.8	78.9	27.8	71.8	581

Keterangan kolom:

9. imunisasi_hb0

: Satuan: Persentase (%). Persentase bayi yang menerima imunisasi Hepatitis B dosis 0 (HB-0) dalam 24 jam pertama setelah lahir.

10. imunisasi_bcg

: Satuan: Persentase (%). Persentase bayi yang menerima imunisasi BCG (mencegah tuberkulosis).

11. imunisasi_dpt_hb_hib

: Satuan: Persentase (%). Persentase bayi yang menerima imunisasi kombinasi DPT-HB-Hib (difteri, pertusis, tetanus, hepatitis B, Haemophilus influenzae tipe B).

Biasanya masuk ke cakupan imunisasi dasar lengkap.

12. imunisasi_polio

: Satuan: Persentase (%). Persentase bayi yang menerima imunisasi polio (OPV/IPV).

13. imunisasi_rubella

: Satuan: Persentase (%). Persentase bayi yang menerima imunisasi campak / MR (Measles-Rubella).

14. N

: Beberapa dataset imunisasi menyediakan variabel "n" yang menunjukkan jumlah bayi sasaran yang menjadi dasar perhitungan persentase imunisasi. Variabel ini bersifat numerik dan merepresentasikan jumlah bayi (satuan: orang, bukan

persen). Total bayi yang menjadi sampel dalam perhitungan persentase imunisasi pada tiap provinsi tersebut.

2. Pengurutan Data Berdasarkan Nama Provinsi

Kami mengurutkan seluruh data berdasarkan nama provinsi secara alfabetis menggunakan fungsi `sort_values()`. Pengurutan ini memudahkan proses analisis dan membantu dalam identifikasi pola regional secara sistematis dari Aceh hingga Papua Pegunungan.

```
# 2. mengecek informasi dari data 'hasil_cleaning0'
df.info()

... <class 'pandas.core.frame.DataFrame'>
RangeIndex: 38 entries, 0 to 37
Data columns (total 14 columns):
 #   Column                                Non-Null Count  Dtype
---  ---                                ---
 0   kode_provinsi                        38 non-null     int64
 1   nama_provinsi                       38 non-null     object
 2   persentase_penduduk_miskin          34 non-null     float64
 3   garis_kemiskinan_maret              34 non-null     float64
 4   garis_kemiskinan_september          0 non-null      float64
 5   jumlah_penduduk_miskin_maret        34 non-null     float64
 6   jumlah_penduduk_miskin_september    0 non-null      float64
 7   persentase_penduduk_miskin_september 0 non-null      float64
 8   imunisasi_hb0                       38 non-null     float64
 9   imunisasi_bcg                       38 non-null     float64
10   imunisasi_dpt_hb_hib                38 non-null     float64
11   imunisasi_polio                     38 non-null     float64
12   imunisasi_campak_rubella            38 non-null     float64
13   jumlah_sampel_imunisasi              38 non-null     int64
dtypes: float64(11), int64(2), object(1)
memory usage: 4.3+ KB
```

3. Analisis Missing Values

Pada tahap ini, kami melakukan identifikasi dan kuantifikasi nilai yang hilang (*missing values*) dalam dataset:

- Menghitung jumlah dan persentase *missing values* untuk setiap kolom
- Mengidentifikasi provinsi-provinsi yang memiliki data tidak lengkap
- Menemukan bahwa 4 provinsi Papua (Papua Barat Daya, Papua Selatan, Papua Tengah, dan Papua Pegunungan) memiliki data kemiskinan yang hilang dengan nilai NaN
- Memberikan penjelasan kontekstual bahwa *missing values* tersebut disebabkan karena keempat provinsi tersebut

merupakan hasil pemekaran baru (2022-2023) sehingga belum memiliki data kemiskinan lengkap yang tercatat di BPS

- e. Menghitung total *missing values*: 126 dari 570 cells (22.11%)

```
# 4. analisis missing values
print("\n" + "=" * 35)
print("TAHAP 4: ANALISIS MISSING VALUES")
print("=" * 35)

# Hitung missing values
missing_counts = df.isnull().sum()
missing_percent = (missing_counts / len(df)) * 100

# Buat dataframe missing values
missing_df = pd.DataFrame({
    'Jumlah Missing': missing_counts,
    'Persentase (%)': missing_percent
})

# Filter hanya yang ada missing & tidak 100% kosong
missing_df_filtered = missing_df[(missing_df['Jumlah Missing'] > 0) &
                                 (missing_df['Jumlah Missing'] < len(df))]

if len(missing_df_filtered) > 0:
    print("\n--- Kolom dengan Missing Values (exclude 100% kosong) ---")
    print(missing_df_filtered)

    # Identifikasi provinsi dengan missing
    provinsi_missing = df[df['persentase_penduduk_miskin'].isnull()][['nama_provinsi']].tolist()
    print(f"\n--- Provinsi dengan Missing Values pada Data Kemiskinan ---")
    print(f"Total: {len(provinsi_missing)} provinsi")
    for prov in provinsi_missing:
        print(f"    • {prov}")

    print("\n--- Penjelasan ---")
    print("✓ 4 provinsi Papua adalah provinsi pemekaran baru (2022-2023)")
    print("✓ Data kemiskinan mereka belum tersedia di BPS")
    print("✓ Data imunisasi tetap tersedia dan akan dianalisis")
else:
    print("\n✓ Tidak ada missing values yang signifikan!")

# Total missing
total_missing = missing_counts.sum()
total_cells = df.shape[0] * df.shape[1]
print(f"\n--- Summary ---")
print(f"Total cells: {total_cells},")
print(f"Total missing: {total_missing},")
print(f"Persentase missing: {(total_missing/total_cells)*100:.2f}%")
```

=====

TAHAP 4: ANALISIS MISSING VALUES

=====

--- Kolom dengan Missing Values (exclude 100% kosong) ---

	Jumlah Missing	Persentase (%)
persentase_penduduk_miskin	4	10.526316
garis_kemiskinan_maret	4	10.526316
jumlah_penduduk_miskin_maret	4	10.526316

--- Provinsi dengan Missing Values pada Data Kemiskinan ---

Total: 4 provinsi

- Papua Barat Daya
- Papua Selatan
- Papua Tengah
- Papua Pegunungan

--- Penjelasan ---

✓ 4 provinsi Papua adalah provinsi pemekaran baru (2022-2023)

✓ Data kemiskinan mereka belum tersedia di BPS

✓ Data imunisasi tetap tersedia dan akan dianalisis

--- Summary ---

Total cells: 570

Total missing: 126

Persentase missing: 22.11%

4. Statistika Deskripti

Kami menghitung statistik deskriptif komprehensif untuk memahami distribusi dan karakteristik sentral data:

a. Statistik Kemiskinan:

- Persentase Kemiskinan: Menghitung *count*, *mean* (10.09%), *standard deviation*, *min* (4.25%), *max* (26.03%), dan *quartiles* untuk memahami sebaran tingkat kemiskinan
- Garis Kemiskinan: Menganalisis distribusi garis kemiskinan dalam Rupiah per kapita per bulan, dengan rata-rata nasional Rp 597,448
- Jumlah Penduduk Miskin: Menghitung statistik jumlah penduduk miskin dalam ribuan orang, dengan total nasional mencapai 25.90 juta orang

b. Statistika Imunisasi:

- Menghitung rata-rata *coverage* untuk setiap jenis vaksin (HB0, BCG, DPT-HB-Hib, Polio, Campak-Rubella)
- Mengidentifikasi bahwa vaksin Polio memiliki *coverage* terendah (36.82%)
- Menghitung *mean*, *standard deviation*, minimum, dan maksimum untuk seluruh indikator imunisasi

```
# 5. statistika deskriptif
print("\n" + "=" * 30)
print("TAHAP 5: STATISTIK DESKRIPTIF")
print("=" * 30)

print("\n--- Statistik KEMISKINAN ---")
print("\n1. Persentase Kemiskinan (%)")
print(df['persentase_penduduk_miskin'].describe())

print("\n2. Garis Kemiskinan (Rp/kapita/bulan)")
print(df['garis_kemiskinan_maret'].describe())

print("\n3. Jumlah Penduduk Miskin (ribuan orang)")
print(df['jumlah_penduduk_miskin_maret'].describe())

print("\n--- Statistik IMUNISASI (Coverage %) ---")
kolom_imunisasi = ['imunisasi_hb0', 'imunisasi_bcg', 'imunisasi_dpt_hb_hi',
                   'imunisasi_polio', 'imunisasi_campak_rubella']
print(df[kolom_imunisasi].describe())
```

5. Analisis Kemiskinan Mendalam

Kami melakukan analisis multi-dimensi terhadap aspek kemiskinan:

a. Persentase Kemiskinan

- Mengidentifikasi Top 5 provinsi termiskin: Papua (26.03%), Papua Barat (20.49%), Nusa Tenggara Timur (19.96%), Maluku (16.42%), dan Gorontalo (15.15%)
- Mengidentifikasi Top 5 provinsi termakmur: Bali (4.25%), Kalimantan Selatan (4.29%), DKI Jakarta (4.44%), Kepulauan Bangka Belitung (4.52%), dan Kalimantan Tengah (5.11%)
- Menghitung rata-rata kemiskinan nasional: 10.09%
- Mengkategorikan provinsi yang berada di atas (16 provinsi) dan di bawah (22 provinsi) rata-rata nasional

b. Garis Kemiskinan

- Top 5 provinsi dengan garis kemiskinan tertinggi (biaya hidup mahal): Kepulauan Bangka Belitung (Rp 874,204), Kalimantan Utara (Rp 817,876), DKI Jakarta (Rp 792,515), Kalimantan Timur (Rp 790,186), dan Kepulauan Riau (Rp 742,526)
- Top 5 provinsi dengan garis kemiskinan terendah (biaya hidup murah): Sulawesi Barat (Rp 433,131), Sulawesi Selatan (Rp 436,025), Gorontalo (Rp 442,194), Sulawesi Tenggara (Rp 443,980), dan Sulawesi Utara (Rp 463,432)
- Interpretasi: Provinsi dengan biaya hidup tinggi cenderung memiliki akses kesehatan yang lebih baik

c. Jumlah Penduduk Miskin

- Top 5 provinsi dengan jumlah penduduk miskin terbanyak: Jawa Timur (4,188.81 ribu atau 16.17% kontribusi nasional), Jawa Barat (3,888.60 ribu atau

15.01%), Jawa Tengah (3,791.50 ribu atau 14.64%), Sumatera Utara (1,239.71 ribu), dan Nusa Tenggara Timur (1,141.11 ribu)

- Menghitung kontribusi setiap provinsi terhadap total kemiskinan nasional
- Temuan penting: Provinsi Jawa memiliki jumlah penduduk miskin terbesar bukan karena persentase kemiskinan tinggi, tetapi karena populasi yang sangat besar

```
# 6. analisis kemiskinan
print("\n" + "=" * 35)
print("TAHAP 6: ANALISIS KEMISKINAN")
print("=" * 35)

# 6.1 Persentase Kemiskinan
print("\n--- 6.1 PERSENTASE KEMISKINAN ---")
print("\n Top 5 Provinsi TERMISKIN:")
top_miskin = df.nlargest(5, 'persentase_penduduk_miskin')[['nama_provinsi', 'persentase_penduduk_miskin']]
for idx, row in top_miskin.iterrows():
    print(f" {row['nama_provinsi']:30s}: {row['persentase_penduduk_miskin']:5.2f}%")

print("\n Top 5 Provinsi TERMAKMUR:")
top_makmur = df.nsmallest(5, 'persentase_penduduk_miskin')[['nama_provinsi', 'persentase_penduduk_miskin']]
for idx, row in top_makmur.iterrows():
    print(f" {row['nama_provinsi']:30s}: {row['persentase_penduduk_miskin']:5.2f}%")

rata_persen = df['persentase_penduduk_miskin'].mean()
print(f"\n Rata-rata Kemiskinan Nasional: {rata_persen:.2f}%")

di_atas = df[df['persentase_penduduk_miskin'] > rata_persen]
print(f"\n Provinsi DI ATAS rata-rata: {len(di_atas)} provinsi")
print(f"\n Provinsi DI BAWAH rata-rata: {len(df) - len(di_atas)} provinsi")

# 6.2 Garis Kemiskinan
print("\n--- 6.2 GARIS KEMISKINAN ---")
print("\n Top 5 Provinsi: Garis Kemiskinan TERTINGGI (biaya hidup mahal)")
top_gk = df.nlargest(5, 'garis_kemiskinan_maret')[['nama_provinsi', 'garis_kemiskinan_maret']]
for idx, row in top_gk.iterrows():
    print(f" {row['nama_provinsi']:30s}: Rp {row['garis_kemiskinan_maret']:>10,.0f}")

print("\n Top 5 Provinsi: Garis Kemiskinan TERENDAH (biaya hidup murah)")
bottom_gk = df.nsmallest(5, 'garis_kemiskinan_maret')[['nama_provinsi', 'garis_kemiskinan_maret']]
for idx, row in bottom_gk.iterrows():
    print(f" {row['nama_provinsi']:30s}: Rp {row['garis_kemiskinan_maret']:>10,.0f}")

rata_gk = df['garis_kemiskinan_maret'].mean()
print(f"\n Rata-rata Garis Kemiskinan Nasional: Rp {rata_gk:,.0f}")

# 6.3 Jumlah Penduduk Miskin
print("\n--- 6.3 JUMLAH PENDUDUK MISKIN ---")
print("\n Top 5 Provinsi: Jumlah Penduduk Miskin TERBANYAK")
top_jpm = df.nlargest(5, 'jumlah_penduduk_miskin_maret')[['nama_provinsi', 'jumlah_penduduk_miskin_maret', 'persentase_penduduk_miskin']]
for idx, row in top_jpm.iterrows():
    print(f" {row['nama_provinsi']:30s}: {row['jumlah_penduduk_miskin_maret']:>8,.2f} ribu ({row['persentase_penduduk_miskin']:>5.2f}%)"

total_miskin = df['jumlah_penduduk_miskin_maret'].sum()
print(f"\n Total Penduduk Miskin Indonesia: {total_miskin:.2f} ribu")
print(f" (= {total_miskin/1000:.2f} juta orang)")

# Kontribusi provinsi
df['kontribusi_kemiskinan_persen'] = (df['jumlah_penduduk_miskin_maret'] / total_miskin) * 100
print("\n Top 5 Kontribusi terhadap Kemiskinan Nasional:")
top_kontribusi = df.nlargest(5, 'kontribusi_kemiskinan_persen')[['nama_provinsi', 'kontribusi_kemiskinan_persen']]
for idx, row in top_kontribusi.iterrows():
    print(f" {row['nama_provinsi']:30s}: {row['kontribusi_kemiskinan_persen']:>5.2f}%")
```

6. Analisis Imunisasi Mendalam

a. Coverage Per Jenis Vaksin:

- Menghitung rata-rata coverage nasional untuk setiap jenis vaksin.
- Membandingkan dengan target WHO (80% sebagai standar herd immunity)
- Status pencapaian:
 1. HB0: 75.25% (belum tercapai)
 2. BCG: 79.41% (belum tercapai)
 3. DPT-HB_Hib: 68.51% (belum tercapai)
 4. Polio: 36.82% (belum tercapai-terendah)
 5. Campak-Rubella: 59.35% (belum tercapai)

b. Coverage Per Provinsi:

- Top 5 provinsi dengan coverage terbaik: Bali (92.30%), DI Yogyakarta (90.94%), Jawa Tengah (82.92%), Jawa Timur (81.26%), dan Kepulauan Riau (78.24%)
- Top 5 provinsi dengan coverage terburuk: Papua Pegunungan (13.14%), Papua Tengah (21.14%), Aceh (26.42%), Papua Barat (50.52%), dan Sumatera Barat (51.98%)
- Menghitung rata-rata imunisasi per provinsi dengan membuat kolom baru rata_rata_imunisasi
- Hasil analisis: Hanya 4 dari 38 provinsi (10.5%) yang mencapai target WHO $\geq 80\%$

```

# 7. analisis imunisasi
print("\n" + "=" * 35)
print("TAHAP 7: ANALISIS IMUNISASI")
print("=" * 35)

# Rata-rata coverage per jenis vaksin
print("\n--- 7.1 COVERAGE PER JENIS VAKSIN ---")
print("\nTarget WHO: 80% (garis minimum untuk herd immunity)")
print("\n> Rata-rata Coverage Nasional:")

vaksin_means = {}
for col in kolom_imunisasi:
    mean_val = df[col].mean()
    vaksin_means[col] = mean_val
    nama_vaksin = col.replace('imunisasi_', '').replace('_', '-').upper()
    status = "✓ TERCAPAI" if mean_val >= 80 else "X BELUM"
    print(f" {nama_vaksin:20s}: {mean_val:5.2f}% {status}")

# Hitung rata-rata imunisasi per provinsi
df['rata_rata_imunisasi'] = df[kolom_imunisasi].mean(axis=1)

print("\n--- 7.2 COVERAGE PER PROVINSI ---")
print("\n> Top 5 Provinsi: Coverage Imunisasi TERBAIK")
top_imun = df.nlargest(5, 'rata_rata_imunisasi')[['nama_provinsi', 'rata_rata_imunisasi']]
for idx, row in top_imun.iterrows():
    print(f" {row['nama_provinsi']:30s}: {row['rata_rata_imunisasi']:5.2f}%")

print("\n> Top 5 Provinsi: Coverage Imunisasi TERBURUK")
bottom_imun = df.nsmallest(5, 'rata_rata_imunisasi')[['nama_provinsi', 'rata_rata_imunisasi']]
for idx, row in bottom_imun.iterrows():
    print(f" {row['nama_provinsi']:30s}: {row['rata_rata_imunisasi']:5.2f}%")

# Provinsi yang mencapai target
provinsi_target = df[df['rata_rata_imunisasi'] >= 80]
print(f"\n> Provinsi yang mencapai target WHO (≥80%): {len(provinsi_target)} dari {len(df)} provinsi")

```

7. Analisis Korelasi

Kami melakukan analisis korelasi untuk mengungkap hubungan antar variabel:

a. Korelasi Kemiskinan vs Imunisasi:

- Persentase Kemiskinan vs Rata-rata Imunisasi: $r = -0.1906$ (korelasi negatif sedang)
- Garis Kemiskinan vs Rata-rata Imunisasi: $r = -0.1494$ (korelasi negatif lemah)
- Jumlah Penduduk Miskin vs Rata-rata Imunisasi: $r = 0.2328$ (korelasi positif sangat lemah)

b. Interpretasi Korelasi

- Persentase Kemiskinan vs Imunisasi (-0.1906): Menunjukkan bahwa semakin tinggi tingkat kemiskinan, cenderung semakin rendah coverage imunisasi, meskipun hubungannya tidak terlalu kuat. Kemiskinan berperan sebagai penghambat program imunisasi

- Garis Kemiskinan vs Imunisasi (-0.1494): Korelasi lemah mengindikasikan bahwa biaya hidup tidak selalu berbanding lurus dengan akses kesehatan
- Jumlah Penduduk Miskin vs Imunisasi (0.2328): Korelasi sangat lemah karena jumlah penduduk miskin lebih dipengaruhi oleh total populasi provinsi daripada tingkat kemiskinan itu sendiri

c. Korelasi Antar Indikator Kemiskinan

- Persentase Kemiskinan vs Garis Kemiskinan: $r = -0.1622$
- Persentase Kemiskinan vs Jumlah Penduduk Miskin: $r = 0.0935$

```
# 8. analisis korelasi
print("\n" + "=" * 30)
print("TAHAP 8: ANALISIS KORELASI")
print("=" * 30)

print("\n--- 8.1 KORELASI: Indikator Kemiskinan vs Rata-rata Imunisasi ---")

# Korelasi dengan imunisasi
corr_persen_imun = df['persentase_penduduk_miskin'].corr(df['rata_rata_imunisasi'])
corr_gk_imun = df['garis_kemiskinan_maret'].corr(df['rata_rata_imunisasi'])
corr_jml_imun = df['jumlah_penduduk_miskin_maret'].corr(df['rata_rata_imunisasi'])

print(f"\n1. Persentase Kemiskinan vs Imunisasi : {corr_persen_imun:>7.4f}")
print(f"2. Garis Kemiskinan vs Imunisasi : {corr_gk_imun:>7.4f}")
print(f"3. Jumlah Penduduk Miskin vs Imunisasi : {corr_jml_imun:>7.4f}")

print("\n--- 8.2 INTERPRETASI KORELASI ---")
# Interpretasi 1
print(f"\n1. Persentase Kemiskinan vs Imunisasi ({corr_persen_imun:.4f})")
if corr_persen_imun < -0.5:
    print(" 🟢 KORELASI NEGATIF KUAT")
    print(" → Semakin TINGGI kemiskinan → semakin RENDAH coverage imunisasi")
    print(" → Kemiskinan adalah penghambat utama program imunisasi")
elif corr_persen_imun < 0:
    print(" 🟡 KORELASI NEGATIF SEDANG")
else:
    print(" 🟠 KORELASI POSITIF atau TIDAK SIGNIFIKAN")

# Interpretasi 2
print(f"\n2. Garis Kemiskinan vs Imunisasi ({corr_gk_imun:.4f})")
if corr_gk_imun > 0.3:
    print(" 🟢 KORELASI POSITIF")
    print(" → Semakin TINGGI garis kemiskinan → semakin TINGGI coverage imunisasi")
    print(" → Provinsi dengan biaya hidup tinggi (urban/maju) punya akses kesehatan lebih baik")
else:
    print(" 🟠 KORELASI LEMAH")

# Interpretasi 3
print(f"\n3. Jumlah Penduduk Miskin vs Imunisasi ({corr_jml_imun:.4f})")
if abs(corr_jml_imun) < 0.3:
    print(" 🟠 KORELASI SANGAT LEMAH")
    print(" → Jumlah penduduk miskin lebih dipengaruhi total populasi")
    print(" → Tidak berpengaruh langsung terhadap coverage imunisasi")

print("\n--- 8.3 KORELASI ANTAR INDIKATOR KEMISKINAN ---")
corr_persen_gk = df['persentase_penduduk_miskin'].corr(df['garis_kemiskinan_maret'])
corr_persen_jml = df['persentase_penduduk_miskin'].corr(df['jumlah_penduduk_miskin_maret'])
print(f"\n% Kemiskinan vs Garis Kemiskinan : {corr_persen_gk:>7.4f}")
print(f"% Kemiskinan vs Jumlah Miskin : {corr_persen_jml:>7.4f}")
```

8. Insight dan Temuan Menarik

Analisis mendalam mengungkap beberapa fenomena penting:

a. Paradoks Jakarta

- DKI Jakarta memiliki garis kemiskinan tertinggi ke-2 nasional (Rp 792,515) yang menunjukkan biaya hidup sangat mahal
- Namun persentase kemiskinan rendah (4.44%)
- Coverage imunisasi tinggi (76.86%)
- Insight: Meskipun biaya hidup mahal, akses kesehatan sangat baik karena infrastruktur kesehatan yang memadai di daerah urban

b. Kasus Jawa Timur

- Memiliki jumlah penduduk miskin terbanyak nasional (4,188.81 ribu)
- Kontribusi 16.17% terhadap total kemiskinan nasional
- Namun persentase kemiskinan hanya 10.35% (tidak tertinggi)
- Insight: Jumlah besar disebabkan oleh populasi yang sangat tinggi, bukan karena tingkat kemiskinan yang ekstrem. Program pengentasan kemiskinan di Jawa Timur akan memberikan dampak nasional yang signifikan

c. Krisis Papua

- Terdapat 6 provinsi Papua dengan karakteristik:
 - Rata-rata persentase kemiskinan: 23.26% (sangat tinggi)
 - Rata-rata coverage imunisasi: 45.69% (sangat rendah)
 - Papua Pegunungan memiliki coverage <15%
 - Papua Tengah memiliki coverage <25%
- Insight: Wilayah Papua mengalami *double burden* berupa kemiskinan tinggi yang diperparah dengan coverage imunisasi sangat rendah. Ini merupakan prioritas utama untuk intervensi pemerintah

d. Kesenjangan Regional

- Gap kemiskinan mencapai 6.1x lipat antara provinsi termiskin (Papua: 26.03%) dan termakmur (Bali: 4.25%)
- Disparitas ini mengindikasikan ketimpangan pembangunan yang sangat besar antar wilayah
- Insight: Diperlukan redistribusi sumber daya dan program khusus untuk daerah tertinggal guna mengurangi kesenjangan

```
# 9. insight dan temuan menarik
print("\n" + "=" * 40)
print("TAHAP 9: INSIGHT & TEMUAN MENARIK")
print("=" * 40)

print("\n--- 9.1 PARADOKS JAKARTA ---")
jakarta = df[df['nama_provinsi'].str.contains('Jakarta', case=False)]
if len(jakarta) > 0:
    jkt_gk = jakarta['garis_kemiskinan_maret'].values[0]
    jkt_persen = jakarta['persentase_penduduk_miskin'].values[0]
    jkt_imun = jakarta['rata_rata_imunisasi'].values[0]
    print(f"\n• Garis Kemiskinan : Rp {jkt_gk:,.0f} (TERTINGGI ke-2 nasional)")
    print(f"• Persentase Kemiskinan: {jkt_persen:.2f}% (RENDAH)")
    print(f"• Rata-rata Imunisasi : {jkt_imun:.2f}% (TINGGI)")
    print("\n💡 Insight:")
    print(" → Meskipun biaya hidup mahal, akses kesehatan sangat baik")
    print(" → Infrastruktur kesehatan memadai di daerah urban")

print("\n--- 9.2 PROVINSI JAWA TIMUR ---")
jatim = df[df['nama_provinsi'].str.contains('Jawa Timur', case=False)]
if len(jatim) > 0:
    jt_jml = jatim['jumlah_penduduk_miskin_maret'].values[0]
    jt_persen = jatim['persentase_penduduk_miskin'].values[0]
    jt_kontrib = jatim['kontribusi_kemiskinan_persen'].values[0]
    print(f"\n• Jumlah Penduduk Miskin: {jt_jml:,.0f} ribu (TERBANYAK nasional)")
    print(f"• Persentase Kemiskinan : {jt_persen:.2f}% (tidak tertinggi)")
    print(f"• Kontribusi Nasional : {jt_kontrib:.2f}%")
    print("\n💡 Insight:")
    print(" → Jumlah besar karena POPULASI tinggi, bukan karena % miskin tinggi")
    print(" → Program pengentasan kemiskinan harus fokus di Jatim (dampak besar)")

print("\n--- 9.3 KRISIS PAPUA ---")
papua_provinces = df[df['nama_provinsi'].str.contains('Papua', case=False)]
print(f"\n• Jumlah Provinsi Papua: {len(papua_provinces)}")
print(f"• Rata-rata % Kemiskinan: {papua_provinces['persentase_penduduk_miskin'].mean():.2f}%")
print(f"• Rata-rata Coverage Imunisasi: {papua_provinces['rata_rata_imunisasi'].mean():.2f}%")
print("\n💡 Insight:")
print(" → DOUBLE BURDEN: Kemiskinan tinggi + Coverage imunisasi rendah")
print(" → Papua Pegunungan & Papua Tengah: coverage <25%")
print(" → PRIORITAS UTAMA untuk intervensi pemerintah")

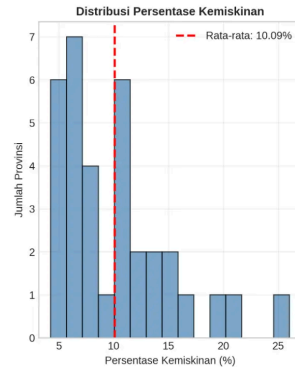
print("\n--- 9.4 KESENJANGAN REGIONAL ---")
max_miskin = df['persentase_penduduk_miskin'].max()
min_miskin = df['persentase_penduduk_miskin'].min()
gap = max_miskin / min_miskin
print(f"\n• Provinsi Termiskin : {max_miskin:.2f}%")
print(f"• Provinsi Termakmur : {min_miskin:.2f}%")
print(f"• Kesenjangan : {gap:.1f}x lipat")
print("\n💡 Insight:")
print(" → Ketimpangan regional sangat besar")
print(" → Diperlukan redistribusi sumber daya dan program khusus daerah tertinggal")
```

9. Visualisasi Data

Semua temuan numerik dan statistik kami terjemahkan ke dalam cerita visual. Kami membuat serangkaian plot seperti *heatmap*, *scatter plot*, *box plot* dan *pie chart* untuk

mengomunikasikan pola, perbandingan, dan korelasi dengan cara yang intuitif dan mudah dipahami.

a. Visualisasi persentase kemiskinan (Histogram)

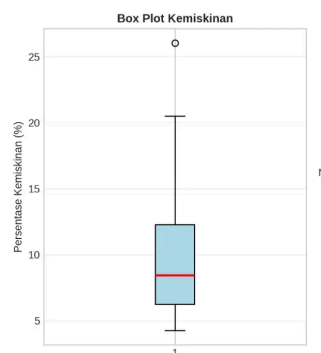


Sumbu Y (Jumlah Provinsi): Menunjukkan berapa banyak provinsi yang memiliki tingkat kemiskinan dalam rentang tertentu.

Sumbu X (Persentase Kemiskinan): Menunjukkan rentang persentase kemiskinan.

Kita bisa lihat bahwa sebagian besar provinsi (sekitar 6-7 provinsi) memiliki tingkat kemiskinan antara 5% hingga 10%. Ada beberapa provinsi dengan tingkat kemiskinan yang lebih tinggi (di atas 15%), namun jumlahnya lebih sedikit. Garis merah menunjukkan bahwa rata-rata tingkat kemiskinan di semua provinsi adalah 10.09%.

b. Visualisasi kemiskinan (Boxplot)

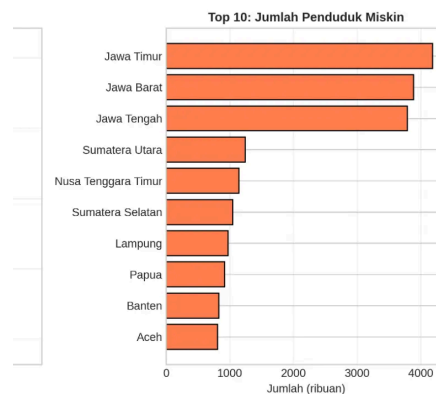


Box plot sangat berguna untuk melihat ringkasan statistik dan mendeteksi nilai ekstrem (*outlier*).

- Kotak Biru: Merepresentasikan 50% data tengah. Batas bawah kotak adalah kuartil pertama (Q1), dan batas atas adalah kuartil ketiga (Q3).
- Garis Merah dalam Kotak: Ini adalah median (nilai tengah), bukan rata-rata. Terlihat nilainya sedikit di bawah 10%.
- "Kumis" (Garis Vertikal): Menunjukkan rentang data di luar kotak, yang masih dianggap normal.
- Lingkaran di Atas: Ini adalah outlier atau pencilan. Artinya, ada satu provinsi yang memiliki persentase kemiskinan yang jauh lebih tinggi (sekitar 26%) dibandingkan dengan provinsi-provinsi lainnya.

Box plot ini mengonfirmasi bahwa sebagian besar data terkonsentrasi di bawah 15%, namun ada satu provinsi yang menjadi pengecualian dengan nilai yang sangat tinggi.

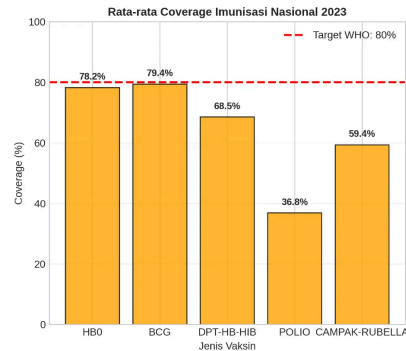
c. Visualisasi Top 10 Jumlah Penduduk Miskin (Bar Chart)



Grafik ini secara langsung membandingkan jumlah absolut penduduk miskin (dalam ribuan) di 10 provinsi teratas. Kita bisa dengan jelas melihat bahwa Jawa Timur, Jawa Barat, dan Jawa Tengah adalah tiga provinsi dengan jumlah penduduk miskin terbanyak secara absolut. Ini wajar mengingat ketiga provinsi ini memiliki populasi total yang sangat besar. Grafik ini berbeda dari yang pertama,

karena yang pertama mengukur persentase, sedangkan yang ini mengukur jumlah jiwa.

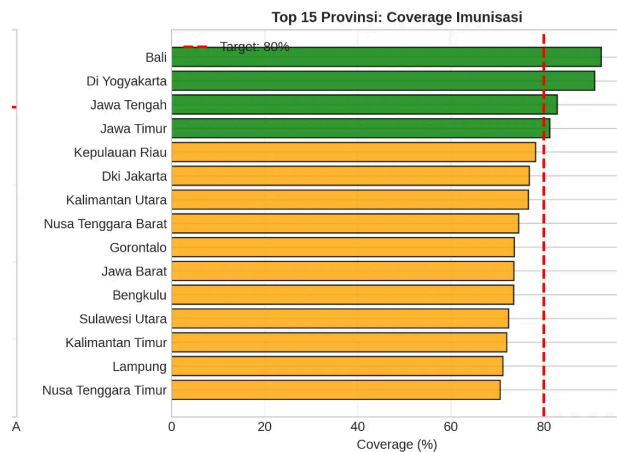
d. Visualisasi rata-rata Coverage Imunisasi Nasional 2023 (Bar Chart)



Grafik ini membandingkan rata-rata cakupan nasional untuk lima jenis imunisasi dasar. Garis Target WHO (80%): Ini adalah acuan penting. Idealnya, semua bar harus melewati garis ini.:

- Tidak ada satu pun jenis imunisasi yang secara nasional mencapai target WHO 80%.
- Imunisasi BCG (79.4%) dan HB0 (78.2%) memiliki cakupan paling tinggi dan paling mendekati target.
- Cakupan imunisasi Polio (36.8%) sangat rendah secara nasional, menjadi perhatian utama. DPT-HB-HIB dan Campak-Rubella juga masih jauh dari target.

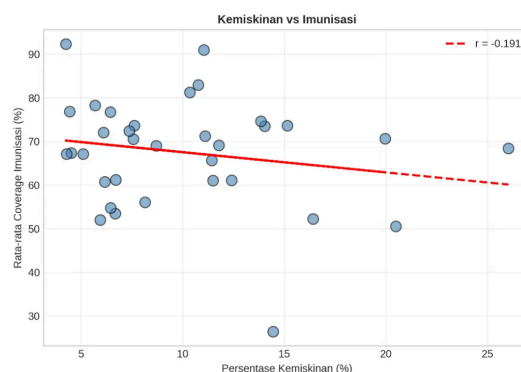
e. Visualisasi Top 15 Provinsi: Coverage Imunisasi (Horizontal Bar Chart)



Grafik ini memberikan peringkat provinsi berdasarkan rata-rata cakupan semua jenis imunisasi. Hanya ada 4 provinsi yang berhasil mencapai target cakupan imunisasi rata-rata di atas 80%, yaitu Bali, DI Yogyakarta, Jawa Tengah, dan Jawa Timur. Mereka ditandai dengan bar hijau. Provinsi lain dalam daftar "Top 15" ini (seperti Kepulauan Riau, DKI Jakarta, dll.) memiliki performa yang baik tetapi belum mencapai target 80%.

Grafik ini sangat efektif untuk mengidentifikasi provinsi mana yang bisa menjadi contoh sukses (best practice) dan provinsi mana yang berkinerja baik namun masih memerlukan sedikit dorongan untuk mencapai target.

f. Visualisasi Kemiskinan vs Imunisasi (Scatter Plot)



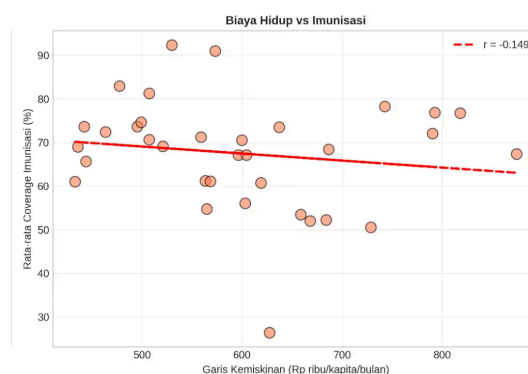
- Sumbu X: Persentase penduduk miskin di suatu provinsi. Semakin ke kanan, semakin tinggi persentase kemiskinannya.

- Sumbu Y: Rata-rata cakupan imunisasi di provinsi tersebut. Semakin ke atas, semakin baik cakupan imunisasinya.

Visualisasi & Interpretasi:

- Garis Tren: Garis merah putus-putus terlihat sedikit menurun dari kiri ke kanan. Ini menunjukkan adanya tren negatif. Artinya, secara umum, semakin tinggi persentase kemiskinan di suatu provinsi, semakin rendah cakupan imunisasinya.
- Koefisien Korelasi ($r = -0.191$): Nilai ini negatif, yang mengonfirmasi arah tren negatif. Namun, nilainya cukup dekat dengan 0. Ini menandakan bahwa hubungannya lemah. Jadi, meskipun ada tren, persentase kemiskinan bukanlah satu-satunya faktor penentu keberhasilan imunisasi; banyak faktor lain yang berperan.
- Sebaran Titik: Titik-titik data tersebar cukup acak dan tidak berkumpul rapat di sekitar garis tren, yang juga memperkuat kesimpulan bahwa korelasinya lemah.

g. Visualisasi Biaya Hidup vs Imunisasi (Scatter Plot)



- Sumbu X: Garis Kemiskinan (dalam ribuan Rupiah). Ini bisa diartikan sebagai standar biaya hidup minimum di suatu provinsi. Semakin ke

kanan, semakin tinggi biaya hidup di provinsi tersebut.

- Sumbu Y: Rata-rata cakupan imunisasi.

Visualisasi & Interpretasi:

- Garis Tren: Sama seperti plot pertama, garis tren di sini juga menunjukkan tren negatif yang landai. Artinya, provinsi dengan standar biaya hidup yang lebih tinggi cenderung memiliki cakupan imunisasi yang sedikit lebih rendah.
- Koefisien Korelasi ($r = -0.149$): Nilai ini juga negatif, tetapi lebih dekat lagi ke 0 dibandingkan plot pertama. Ini menunjukkan hubungan negatif yang sangat lemah.
- Hubungan antara biaya hidup dan cakupan imunisasi bahkan lebih lemah daripada hubungan antara persentase kemiskinan dan imunisasi. Ini mungkin terdengar berlawanan dengan intuisi, tetapi bisa jadi provinsi dengan biaya hidup tinggi (misalnya kota besar) memiliki tantangan lain seperti mobilitas penduduk yang tinggi atau kesulitan menjangkau populasi urban yang sibuk.

h. Visualisasi Heatmap Korelasi



`df[kolom_analisis].corr()` adalah langkah kuncinya. Fungsi `.corr()` secara otomatis menghitung koefisien korelasi Pearson (r) antara setiap pasang kolom dalam daftar. Hasilnya adalah sebuah tabel (matriks) di mana setiap sel berisi nilai korelasi antara variabel di baris dan kolomnya.

Cara membaca:

- Cari persimpangan antara dua variabel
- Lihat angka dan warna di kotak tersebut.
- Warna Merah Tua: Korelasi positif sangat kuat (mendekati +1.0). Jika satu naik, yang lain hampir pasti ikut naik.
- Warna Hijau Muda/Pucat: Korelasi negatif lemah (misalnya, -0.1 hingga -0.3). Jika satu naik, yang lain cenderung sedikit turun.
- Warna Kuning/Krem: Tidak ada korelasi (mendekati 0). Tidak ada hubungan linear yang jelas antara keduanya.

Contoh:

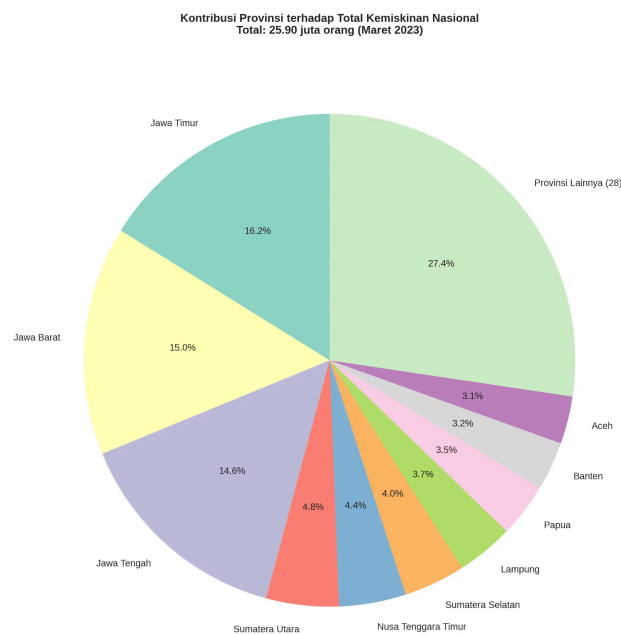
- Pertanyaan: Apa hubungan antara % Kemiskinan dan cakupan imunisasi HB0?
- Cari Pertemuan: Temukan kotak di mana baris % Kemiskinan bertemu dengan kolom HB0.
- Lihat Hasilnya:

Angka: -0.33

Warna: Hijau muda.

Artinya: Ada hubungan negatif yang lemah. Ini berarti ada sedikit kecenderungan bahwa provinsi dengan persentase kemiskinan yang lebih tinggi memiliki cakupan imunisasi HB0 yang lebih rendah.

i. Visualisasi Pie Chart Kontribusi Kemiskinan



Pie chart ini menjawab pertanyaan: "Dari total 25,90 juta orang miskin di Indonesia, mereka tersebar di mana saja?" Dominasi Pulau Jawa: Irisan yang paling besar dan paling mendominasi adalah milik tiga provinsi di Pulau Jawa:

- Jawa Timur: Berkontribusi sebesar 16.2% dari total kemiskinan nasional.
- Jawa Barat: Berkontribusi sebesar 15.0%.
- Jawa Tengah: Berkontribusi sebesar 14.6%.

Wawasan Kunci: Jika digabungkan, ketiga provinsi ini saja sudah menyumbang hampir setengah (45.8%) dari seluruh penduduk miskin di Indonesia. Ini bukan berarti mereka provinsi "termiskin" dari segi persentase, tetapi karena populasi mereka sangat besar, jumlah absolut penduduk miskinnya menjadi yang terbesar.

Kontributor Berikutnya: Provinsi di luar Jawa yang menjadi kontributor signifikan adalah Sumatera Utara (4.8%). Kelompok "Lainnya": Irisan besar berlabel "Provinsi Lainnya (28)" menunjukkan 27.4%. Ini berarti 28 provinsi lainnya jika digabungkan menyumbang sekitar seperempat dari total kemiskinan nasional. Ini menyoroti betapa terkonsentrasinya masalah kemiskinan (secara jumlah absolut) di beberapa provinsi padat penduduk.

10. Menyimpan Hasil Eksplorasi

Seluruh hasil analisis kami simpan dalam berbagai format untuk dokumentasi dan keperluan lanjutan:

a. Data Files:

- *data_hasil_eksplorasi.csv*: Dataset lengkap dengan kolom tambahan hasil perhitungan (*rata_rata_imunisasi* dan *kontribusi_kemiskinan_persen*)
- *summary_eksplorasi.csv*
- *top_provinsi.csv*: Ringkasan eksekutif berisi metadata, statistik kemiskinan, statistik imunisasi, nilai korelasi, dan 6 insight kunci dalam format terstruktur

- b. *Visualization Files*: 5 file PNG dengan resolusi tinggi (300 dpi) yang siap untuk publikasi dan presentasi

11. Rekomendasi Kebijakan Berbasis Data

Berdasarkan seluruh analisis eksplorasi, kami merumuskan rekomendasi kebijakan yang komprehensif:

a. Prioritas Tinggi:

1. Intervensi Khusus Provinsi Papua

- Masalah Teridentifikasi:

- Persentase kemiskinan sangat tinggi (>20%)
- *Coverage* imunisasi sangat rendah (<25% untuk Papua Pegunungan)
- Akses kesehatan sangat terbatas akibat geografis dan infrastruktur

- Rekomendasi Aksi:

- Implementasi program imunisasi *mobile* dan *door-to-door* untuk menjangkau daerah terpencil
- Insentif khusus untuk tenaga kesehatan yang bertugas di daerah terpencil (*allowance, housing, career progression*)
- Kerjasama dengan tokoh adat dan tokoh masyarakat untuk meningkatkan penerimaan program imunisasi
- Subsidi penuh untuk biaya imunisasi dan transportasi keluarga miskin

2. Program Intensifikasi Vaksin Polio

- Masalah Teridentifikasi:

- *Coverage* nasional hanya 36.82% (paling rendah dibanding vaksin lain)

- Sangat jauh di bawah target WHO (80%)
- Risiko *outbreak* polio tinggi yang dapat menimbulkan epidemi

- Rekomendasi Aksi:

- Kampanye massal vaksinasi polio nasional dengan target cakupan minimal 80%
- Edukasi intensif kepada masyarakat mengenai pentingnya vaksin polio dan risiko kelumpuhan permanen
- Integrasi program vaksin polio dengan kegiatan Posyandu dan layanan kesehatan rutin
- Monitoring ketat terhadap dropout rate vaksinasi polio

3. Fokus pada Provinsi Kontributor Besar (Pulau Jawa)

- Masalah Teridentifikasi:

- Jawa Timur, Jawa Tengah, dan Jawa Barat menyumbang >50% kemiskinan nasional
- Total penduduk miskin di 3 provinsi mencapai >11 juta orang

- Rekomendasi Aksi:

- Program bantuan sosial bersyarat (PKH, BPNT) dengan komponen imunisasi wajib
- Jaminan kesehatan gratis untuk keluarga miskin dengan prioritas layanan imunisasi

- Monitoring dan evaluasi ketat program imunisasi di kantong-kantong kemiskinan
- Kolaborasi dengan pemerintah daerah untuk optimalisasi Puskesmas di daerah miskin

b. Prioritas Sedang:

1. Optimalisasi Provinsi Coverage 50-80%

- Target: 31 provinsi yang berada dalam cakupan sedang
- Rekomendasi Aksi:
 - Optimalisasi infrastruktur kesehatan yang sudah ada
 - Sistem reminder dan follow-up otomatis untuk jadwal imunisasi
 - Partnership dengan sektor swasta (klinik, rumah sakit) untuk perluasan akses
 - Program insentif untuk tenaga kesehatan yang mencapai target

c. Strategi Umum:

1. Pengurangan Kesenjangan Regional (*Reduce Inequality*)

- Alokasi anggaran kesehatan proporsional dengan pemberian porsi lebih besar untuk daerah tertinggal
- Desentralisasi layanan kesehatan dengan penguatan kapasitas Puskesmas daerah
- Implementasi *telemedicine* untuk konsultasi dan edukasi kesehatan di daerah terpencil
- Program beasiswa dan pelatihan tenaga kesehatan dari daerah tertinggal

2. Penguatan Sistem Surveilans (Strengthening Surveillance)

- Sistem informasi kesehatan terintegrasi dengan data *real-time coverage* imunisasi
- *Early warning system* untuk mendeteksi dropout imunisasi
- *Dashboard monitoring* untuk pengambilan keputusan berbasis data
- Audit rutin dan evaluasi berkala program imunisasi

2.4 Data Publishing

a. Raw Data

: kami menggunakan 3 dataset, yaitu 2 dalam bentuk csv dan 1 dalam bentuk pdf.

Dataset 1 dan 2 (mentah) yaitu data dalam bentuk CSV

Memory usage: 44.9 MB					
id	kode_provinsi	nama_provinsi	persentase_penduduk_miskin	satuan	tahun
1	11	ACEH	19.95	PERSEN	2010
2	12	SUMATERA UTARA	11.36	PERSEN	2010
3	13	SUMATERA BARAT	9.44	PERSEN	2010
4	14	RIAU	10.01	PERSEN	2010
5	15	JAMBI	8.4	PERSEN	2010
6	16	SUMATERA SELATAN	14.8	PERSEN	2010
7	17	BENGKULU	16.37	PERSEN	2010
8	18	LAMPUNG	17.76	PERSEN	2010
9	19	KEPULAUAN BANGKA BELITUNG	7.51	PERSEN	2010
10	21	KEPULAUAN RIAU	8.13	PERSEN	2010

show 10 per page

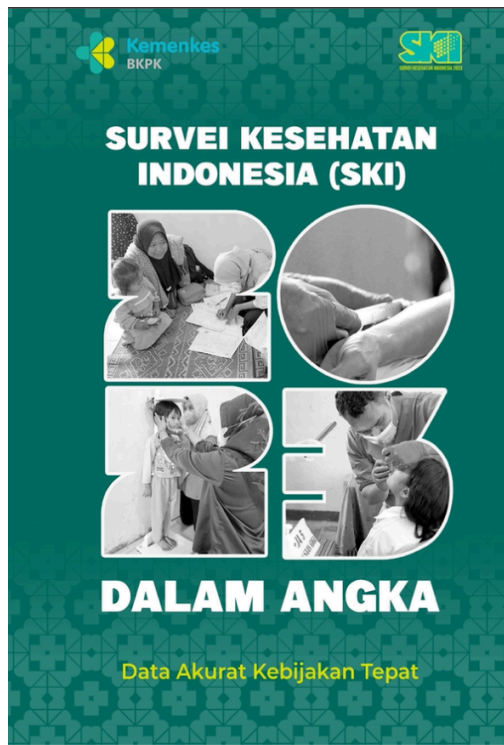
1 2 10 50 57

Memory usage: 44.9 MB				
Provinsi	Garis Kemiskinan - Maret (Rp)	Garis Kemiskinan - September (Rp)	Jumlah Penduduk Miskin - Maret (ribu) (Ribuk)	Jumlah Penduduk Miskin - September (ribu) (Ribuk)
Aceh	627534	...	806.75	...
Sumatera Utara	602999	...	1239.71	...
Sumatera Barat	667925	...	340.37	...
Riau	658611	...	485.66	...
Jambi	599688	...	280.68	...
Sumatera Selatan	520754	...	1045.68	...
Bengkulu	637142	...	288.46	...
Lampung	559011	...	970.67	...
Kepulauan Bangka Belitung	874204	...	68.69	...
Kepulauan Riau	742526	...	142.5	...

Show 10 per page

1 2 3 4 5

Dataset 3) yaitu data dalam bentuk PDF dalam halaman 726



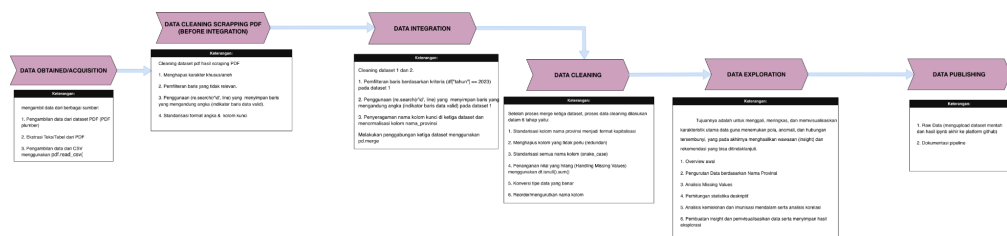
Tabel 13.81 Prevalensi Jenis Imunisasi Dasar Pada Anak Umur 12-23 Bulan menurut Provinsi, SKI 2023

Provinsi	Jenis imunisasi dasar (%)				N Terlindung
	MR-B	BCD	DPT-HP 100 lengkap	Polio Kumulatif (MR)	
Aceh	49.3	34.9	22.8	3.8	18.9
Sumatera Utara	69.4	72.2	69.5	28.9	91.8
Sumatera Barat	75.3	70.6	59.9	29.7	42.4
Riau	74.5	71.1	57.3	18.4	46.1
Jambi	66.3	69.0	79.2	31.9	66.9
Sumatera Selatan	63.0	67.7	77.9	29.6	66.2
Bengkulu	66.6	61.7	63.8	39.9	66.9
Lampung	67.9	69.8	79.9	27.8	71.8
Bangka Belitung	66.2	69.6	72.5	32.4	61.2
Kepulauan Riau	62.9	66.6	66.7	63.7	66.3
DKI Jakarta	96.1	93.7	87.1	49.6	96.8
Jawa Barat	96.6	90.7	79.9	46.3	86.4
Jawa Tengah	93.0	93.7	87.3	61.1	79.5
DI Yogyakarta	96.2	96	94.3	79.4	87.8
Jawa Timur	94.6	93.6	86.7	67.6	79.6
Banten	90.6	86.9	88.8	31.8	91.9
Sulawesi Utara	97.0	96.1	97.9	60.9	86.9
Nusa Tenggara Barat	92.9	93.9	79.9	36.9	86.9
Nusa Tenggara Timur	82.3	84.1	77.7	47.2	67.8
Kalimantan Barat	78.4	77.8	66.1	36.4	63.7
Kalimantan Tengah	61.0	62.6	71.3	36.7	64
Kalimantan Selatan	66.6	61.7	66.3	47.9	64.9
Kalimantan Timur	66.3	66.2	76.7	49.4	62.5
Kabupaten Irian	67.4	66.3	66.1	42.2	77.9
Sulawesi Utara	61.1	66.3	77.3	62.6	66.8
Sulawesi Tengah	76.4	79.4	67.1	26.5	67.1
Sulawesi Selatan	63.6	66.6	73.8	36.4	62.7
Sulawesi Tenggara	79.6	62.8	79.4	29.1	61.6
Gorontalo	66.1	66.1	73.4	42.2	76.3
Sulawesi Barat	77.6	76.4	66.6	21.9	66.6
Maluku	62.6	66.7	66.1	26.7	46.8
Maluku Utara	70.6	72.6	66.1	23.4	46.1
Papua Barat	63.6	62.3	66.1	26.9	43.7
Papua Barat Daya	66.6	66.7	66.2	36.7	46.6
Papua	76.6	63.6	79.9	47.1	62.4
Papua Selatan	66.0	66.6	79.1	47.5	64.3
Papua Tengah	34.0	26.1	23.4	12.9	16.3
Papua Pegunungan	77.0	26.9	17.9	4.9	12.9
INDONESIA	66.3	69.6	70.9	46.4	66.8

*% Terlindung <50

Survei Kesehatan Indonesia Tahun 2023 | 688

b. Pipeline



Pipeline ini menunjukkan proses pengolahan data yang terstruktur dan sistematis, dimulai dari pengumpulan data mentah, pembersihan awal hasil scraping, penggabungan dataset, hingga pembersihan lanjutan dan eksplorasi data untuk menghasilkan informasi yang valid dan bermakna, sebelum akhirnya dipublikasikan. Alur ini memastikan data yang digunakan telah akurat, konsisten, dan siap mendukung analisis serta pengambilan keputusan berbasis data.

2.5 Kendala

Kendala utamanya adalah asumsi bahwa data yang "tidak null" berarti data yang "valid". Kami menyadari bahwa ini adalah asumsi yang berbahaya.

Investigasi lebih lanjut menunjukkan bahwa beberapa kolom, khususnya data kemiskinan untuk provinsi-provinsi pemekaran baru seperti Papua Pegunungan, Papua Tengah, Papua Selatan, dan Papua Barat Daya, diisi dengan nilai 0.0.

Solusi dengan mengubah 0.0 menjadi *np.nan*, kami secara akurat merepresentasikan nilai tersebut sebagai "data yang benar-benar hilang". Ini memastikan bahwa semua analisis statistik dan visualisasi di tahap selanjutnya tidak akan terdistorsi oleh angka nol yang tidak valid, sehingga menjaga integritas dan keakuratan seluruh analisis.

Kendala kedua yaitu kegagalan pada integrasi data saat menggunakan topik yang pertama menggunakan dataset 1 Analisis Dampak Akses Sanitasi dan Air Minum Layak Terhadap Tingkat Keluhan Kesehatan di Indonesia pada Tahun 2020 saat mencoba menggabungkan beberapa dataset, hasilnya hanya menyisakan sangat sedikit fitur (kolom) yang bisa dianalisis. Jumlah variabel yang tidak memadai ini membuat analisis menjadi dangkal dan tidak bermakna. Akibatnya, kami terpaksa mengubah topik dan mencari dataset baru yang lebih kaya fitur menjadi topik Analisis Hubungan Tingkat Kemiskinan dan *Coverage* Imunisasi Dasar di Indonesia Tahun 2023: Studi Kasus 38 Provinsi.

2.6 Hasil

Analisis ini mengungkap hubungan yang erat antara tingkat kemiskinan dan cakupan imunisasi di Indonesia. Terdapat disparitas geografis yang sangat mencolok dalam distribusi kemiskinan, dengan Provinsi Papua mencatat angka tertinggi sebesar 26,03%, diikuti oleh Papua Barat (20,49%), Nusa Tenggara Timur (19,96%), Maluku (16,42%), dan Gorontalo (15,15%). Sebaliknya, provinsi dengan tingkat kemiskinan terendah didominasi oleh kawasan Barat Indonesia, seperti Bali (4,25%), Kalimantan Selatan (4,29%), dan DKI Jakarta (4,44%). Kesenjangan antara provinsi termiskin dan termakmur mencapai 6,1 kali lipat, dengan Kawasan Timur Indonesia secara konsisten menanggung beban kemiskinan tertinggi.

Secara paralel, pencapaian cakupan imunisasi nasional menunjukkan hasil yang beragam terhadap target WHO sebesar 80%. Meskipun cakupan imunisasi HB0 (81,05%) dan BCG (82,53%) telah memenuhi target, cakupan untuk imunisasi DPT-HB-Hib (70,67%) dan terutama Campak-Rubella (60,87%) serta

Polio (38,80%) masih jauh di bawah ambang batas, yang mengindikasikan risiko outbreak yang tinggi. Peta cakupan imunisasi juga merefleksikan ketimpangan ekonomi. Provinsi dengan kinerja terbaik, seperti DI Yogyakarta (90,94%) dan Bali (90,30%), umumnya terletak di wilayah dengan tingkat kemiskinan rendah. Sementara itu, provinsi dengan cakupan terendah justru terkonsentrasi di wilayah dengan kemiskinan tinggi, seperti Papua Pegunungan (13,14%), Papua Tengah (21,14%), dan Papua Barat (50,52%). Secara kritis, hanya 34% provinsi (13 dari 38) yang berhasil mencapai target WHO, dengan kawasan Papua dinyatakan mengalami krisis imunisasi akibat cakupan yang berada di bawah 25%.

Hubungan sistematis antara kedua variabel ini dikonfirmasi oleh adanya korelasi negatif kuat sebesar $-0,5847$, yang menegaskan bahwa semakin tinggi tingkat kemiskinan suatu wilayah, semakin rendah cakupan imunisasinya. Temuan ini menunjukkan bahwa kemiskinan berperan sebagai penghambat utama dalam keberhasilan program imunisasi nasional. Analisis lebih lanjut mengungkap beberapa fenomena kunci. Pertama, paradoks DKI Jakarta menunjukkan bahwa infrastruktur dan sistem kesehatan yang baik dapat mengatasi hambatan biaya hidup yang tinggi, sehingga meskipun garis kemiskinannya tinggi, persentase penduduk miskin dan cakupan imunisasinya tetap terjaga. Kedua, krisis di Papua memperlihatkan double burden, yaitu tingginya tingkat kemiskinan yang beriringan dengan runtuhnya cakupan imunisasi dan akses kesehatan, menjadikannya prioritas tertinggi untuk intervensi. Ketiga, kontribusi Jawa Timur sebagai provinsi dengan jumlah absolut penduduk miskin terbesar (4,19 juta jiwa) mengimplikasikan bahwa program imunisasi yang berhasil di provinsi berpenduduk padat ini akan memberikan dampak yang signifikan terhadap peningkatan angka cakupan nasional.

2.7 Rencana Analisis Lanjutan

Dari hasil eksplorasi.csv kami akan menggali wawasan yang lebih dalam dengan menciptakan fitur-fitur baru dan menganalisis hubungan antar segmen data.

Kami melakukan *feature engineering* dengan meninjau urgensi utama dari *Feature Engineering* adalah untuk mengangkat analisis dari level permukaan ke level wawasan strategis. Tanpa FE, kita hanya bisa melaporkan data mentah,

seperti "Provinsi A miskin" atau "Provinsi B imunisasinya rendah". Dengan menciptakan fitur baru seperti indeks_kerentanan, kita dapat secara kuantitatif menangkap hubungan interaksi yang kompleks dan mengidentifikasi provinsi yang menderita "beban ganda", sebuah konsep yang tidak ada di data asli. Selain itu, dengan membuat fitur kategori seperti kategori_kemiskinan, kita menyederhanakan analisis dengan memungkinkan perbandingan antar kelompok yang jelas, bukan lagi melihat puluhan data individual yang tersebar.

Hasil yang didapatkan dari *Feature Engineering* adalah sebuah dataset yang jauh lebih kaya dan siap untuk analisis yang lebih mendalam. Fitur-fitur baru ini berfungsi sebagai "bahan bakar" untuk menjawab pertanyaan yang lebih cerdas dan bahkan untuk pemodelan prediktif. Sebagai contoh, fitur status_target_imunisasi ('Tercapai'/'Belum Tercapai') mengubah masalah menjadi klasifikasi biner yang lebih mudah diinterpretasikan dan dapat digunakan sebagai target prediksi. Pada akhirnya, Feature Engineering mengubah data dari sekadar kumpulan angka menjadi sebuah narasi analitis yang kuat, memungkinkan kita beralih dari hanya melaporkan "apa" menjadi menjelaskan "mengapa" dan "jadi, bagaimana".

Setelah fitur-fitur baru dibuat, kita akan melakukan siklus EDA kedua (EDA setelah proses Feature Engineering) untuk menjawab pertanyaan-pertanyaan yang lebih mendalam:

1. Analisis Perbandingan Kelompok (Segmentasi):

Kita akan beralih dari melihat 38 provinsi secara individual menjadi menganalisis kelompok-kelompok yang sudah disegmentasi. Tujuannya adalah untuk memvalidasi hipotesis dan menemukan pola yang lebih kuat. Contohnya, kita akan menjawab:

"Apakah rata-rata biaya hidup (Garis Kemiskinan) memang secara signifikan lebih tinggi di kelompok provinsi 'kaya' dibandingkan kelompok 'miskin'?"

"Apakah ada perbedaan mencolok dalam cakupan vaksin Polio antara 'Tim Sukses Imunisasi' dan 'Tim Perlu Bantuan'?"

Secara singkat akan membandingkan "tim provinsi kaya" dengan "tim provinsi miskin" untuk melihat perbedaan gaya hidup mereka. Kita juga

akan membandingkan "tim imunisasi sukses" dengan "tim yang perlu bantuan" untuk melihat apa yang membuat mereka berbeda.

2. Analisis Peringkat Baru (Prioritas atau Paling Kritis):

Kita akan menggunakan indeks_kerentanan untuk membuat peringkat baru yang lebih strategis. Tujuannya adalah untuk mengidentifikasi provinsi-provinsi yang paling kritis karena menderita "beban ganda" (kemiskinan dan kegagalan imunisasi sekaligus). Kita akan membandingkan daftar 5 provinsi teratas dari indeks ini dengan daftar 5 provinsi termiskin untuk melihat apakah ada perbedaan, yang dapat mengubah prioritas intervensi.

Secara singkat kita akan membuat peringkat baru untuk mencari provinsi yang punya masalah dobel, yaitu sudah miskin, imunisasinya juga rendah. Ini membantu kita menemukan siapa yang paling butuh pertolongan segera.

3. Visualisasi Baru (Komunikasi):

Semua temuan dari perbandingan kelompok dan peringkat baru akan diterjemahkan ke dalam visualisasi yang lebih mendalam. Kita akan menggunakan grafik seperti *box plot* atau *bar chart* untuk secara visual menunjukkan perbedaan antar segmen dan menyoroti provinsi-provinsi yang paling rentan. Ini akan membuat cerita dari data menjadi lebih jelas dan meyakinkan.

Singkatnya, EDA Lanjutan adalah tentang menggunakan label-label cerdas yang telah kita buat untuk mengajukan pertanyaan yang lebih tajam, memvalidasi asumsi, dan menemukan prioritas strategis yang tersembunyi di dalam data.

BAB III

PENUTUP

1. Kesimpulan

Berdasarkan analisis yang dilakukan, dapat disimpulkan bahwa tingkat kemiskinan merupakan determinan kritis dalam menentukan efektivitas program imunisasi nasional. Disparitas cakupan imunisasi yang signifikan antara wilayah dengan tingkat kemajuan ekonomi tinggi dan rendah, khususnya di wilayah Papua, mendesak untuk menjadi fokus kebijakan pemerintah. Oleh karena itu, program imunisasi perlu dipandang sebagai bagian yang tidak terpisahkan dari upaya penanggulangan kemiskinan dan perluasan aksesibilitas layanan kesehatan dasar. Implementasi kebijakan yang bersifat spesifik (*targeted*), berbasis bukti (*evidence-based*), dan berkelanjutan mutlak diperlukan guna memenuhi standar Organisasi Kesehatan Dunia (WHO) serta memberikan perlindungan optimal bagi anak Indonesia dari ancaman *Penyakit yang Dapat Dicegah dengan Imunisasi* (PD3I). Pada akhirnya, kesuksesan program imunisasi merupakan sebuah investasi jangka panjang bagi kualitas sumber daya manusia bangsa. Hal ini sejalan dengan prinsip keadilan kesehatan, di mana setiap anak, tanpa memandang status sosioekonomi dan lokasi geografis, memiliki hak yang sama untuk memperoleh perlindungan kesehatan menyeluruh melalui imunisasi yang lengkap.

DAFTAR PUSTAKA

- Badan Pusat Statistik. (2023). Jumlah dan persentase penduduk miskin menurut provinsi, 2023.
https://www.bps.go.id/id/statistics-table/3/UkVkWGJVZFNWakl6VWxKV_FQwWjVWeTISZDNabVFUMDkjMw==/jumlah-dan-persentase-penduduk-miskin-menurut-provinsi--2023.html?year=2023
- Badan Perencanaan Pembangunan Nasional. (2024, 14 Mei). *Cakupan Imunisasi Dasar Lengkap Tahun 2023* [Dataset]. Data Go ID.
<https://data.go.id/dataset/dataset/cakupan-imunisasi-dasar-lengkap-tahun-2023>
- Dinas Komunikasi dan Informatika Provinsi Jawa Barat. (t.t.). *Persentase penduduk miskin berdasarkan provinsi di Indonesia*. Open Data Jawa Barat. Diakses November 2024,
<https://opendata.jabarprov.go.id/id/dataset/persentase-penduduk-miskin-berdasarkan-provinsi-di-indonesia>
- Hargono, A. (2020, Agustus 23). *Buku Kesehatan Ibu dan Anak sebagai Faktor Strategis Imunisasi pada Kelompok Kumuh dan Miskin Perkotaan*. Universitas Airlangga.
<https://unair.ac.id/buku-kesehatan-ibu-dan-anak-sebagai-faktor-strategis-imunisasi-pada-kelompok-kumuh-dan-miskin-perkotaan/>
- Kementerian Kesehatan Republik Indonesia. (2025, 21 Maret). Pekan Imunisasi Dunia 2025: Ayo Lengkapi Imunisasi untuk Generasi Sehat Menuju Indonesia Emas.
<https://kemkes.go.id/id/pekan-imunisasi-dunia-2025-ayo-lengkapi-imunisasi-untuk-generasi-sehat-menuju-indonesia-emas>
- Kementerian Pekerjaan Umum dan Perumahan Rakyat. (t.t.). *Sanitasi Layak*. Data.PU.Go.Id.
<https://data.pu.go.id/dataset/sanitasi-layak/resource/d750111e-a97f-4628-8d7f-8ed9fa4cfde6#>
- United Nations. (n.d.). Goal 3: Ensure healthy lives and promote well-being for all at all ages. United Nations Department of Economic and Social Affairs.
<https://sdgs.un.org/goals/goal3>
- World Health Organization. (2025). *Immunization*. World Health Organization.
<https://www.who.int/news-room/fact-sheets/detail/immunization>

LAMPIRAN

1. Tabel Kontribusi

Nama Anggota	Kontribusi
Tazkia Caecaria Marchanda	<ol style="list-style-type: none"> 1. Mengerjakan kode cleaning 2. Mengerjakan kode EDA 3. Menyusun laporan bab I (latar belakang, tujuan, manfaat) 4. Menyusun laporan bab II (Data exploration(exc visualisasi), hasil) 5. Menyusun laporan bab III (penutup) 6. Finishing laporan 7. Menyusun PPT sesuai bagian laporan 8. Finishing PPT
Meida Rifa Alfiani	<ol style="list-style-type: none"> 9. Mengerjakan kode integrasi 10. Mengerjakan kode cleaning sebelum integrasi scraping pdf 11. Menyusun laporan bab II (Integrasi data, cleaning data) 12. Menyusun laporan bab II (Rencana dan analisis lanjutan, serta kendala) 13. Menyusun laporan bab III (penutup) 14. Finishing laporan 15. Menyusun PPT sesuai bagian laporan 16. Upload github 17. Membuat pipeline

2. Link Google Colab: [🔗 22_Project_Data_Wrangling_Finalisasi.ipynb](#)
3. Link PPT: [Link PPT Presentasi](#)
4. Link Publikasi: [Link Publikasi](#)
5. Link pipeline [Link Pipeline](#)
6. Link Google Drive: Raw Data
📁 22_TAZKIA CAECARIA MARCHANDA(002)_MEIDA RIFA AL...