

# ANALISIS HUBUNGAN ANTARA UPAH MINIMUM PROVINSI, POLA MAKAN, DAN PREVALENSI KESEHATAN DI INDONESIA



KELOMPOK 11

---

Salsabila Alike Seftizianka (24031554117)

Naufal Muzaki (24031554061)

An abstract graphic on the left side of the slide. It features a wireframe cube with a glowing cyan point at its center. Several other smaller, glowing points in cyan, magenta, and yellow are scattered around the cube. The background is a gradient of dark blue and purple.

# LATAR BELAKANG

Indonesia memiliki keberagaman kondisi ekonomi dan sosial yang berdampak pada pola konsumsi dan kesehatan masyarakat. Upah Minimum Provinsi (UMP) menjadi indikator penting yang memengaruhi daya beli masyarakat dalam memenuhi kebutuhan pangan bergizi. Namun, tingginya UMP tidak selalu berkorelasi dengan kualitas pola makan ataupun kondisi kesehatan.

Di sisi lain, data konsumsi energi dan protein menunjukkan ketimpangan antar provinsi, yang turut berpengaruh terhadap masalah gizi seperti stunting, wasting, dan overweight. Kerawanan pangan juga menjadi faktor penting yang dapat memperburuk kondisi kesehatan masyarakat, terutama di wilayah dengan akses pangan yang terbatas.

Oleh karena itu, diperlukan analisis data yang mengintegrasikan UMP, pola makan, dan indikator kesehatan untuk memahami hubungan antarvariabel tersebut serta memberikan gambaran yang lebih komprehensif mengenai kondisi gizi dan kesejahteraan masyarakat di Indonesia.

# RUMUSAN MASALAH

Bagaimana mengintegrasikan data UMP, pola makan, dan kesehatan dari berbagai sumber menggunakan teknik data wrangling?

Bagaimana pengaruh pola asupan kalori terhadap prevalensi kesehatan?

Apakah terdapat hubungan antara tingkat UMP dan pola konsumsi pangan masyarakat di Indonesia?

# TUJUAN

1. Mengumpulkan dan mengolah data UMP, asupan kalori, dan kesehatan dari sumber yang diperoleh.
2. Mengidentifikasi hubungan antara UMP dan pola asupan kalori masyarakat.
3. Menganalisis pengaruh pola asupan kalori terhadap prevalensi kesehatan berdasarkan dataset yang sudah digabungkan.

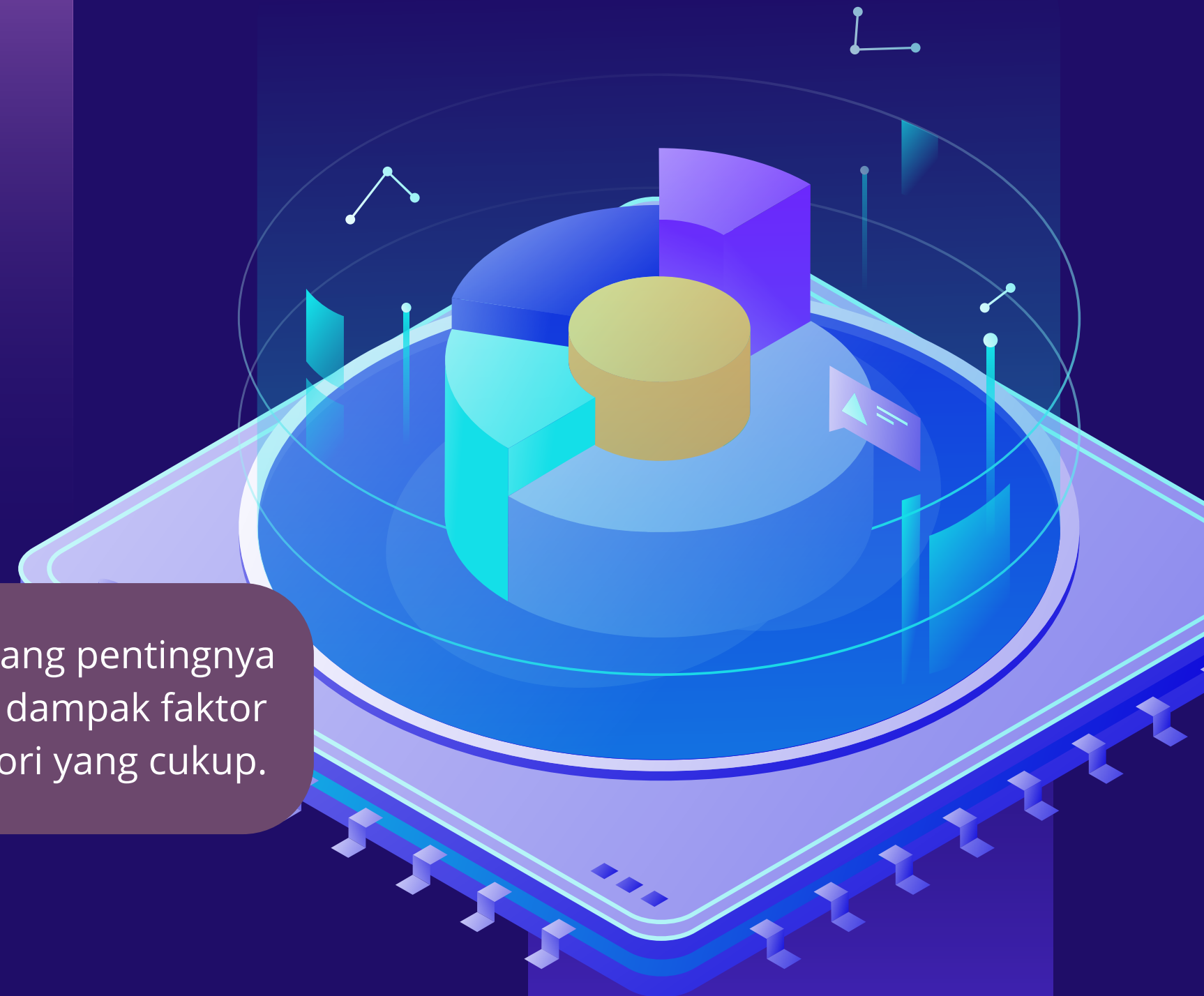


# MANFAAT

Mendukung penyusunan kebijakan terkait UMP, jumlah asupan kalori, dan perbaikan kesehatan masyarakat.

Menambah literatur tentang hubungan ekonomi–jumlah asupan kalori–kesehatan serta melatih penerapan teknik data wrangling dalam analisis nyata.

Memberikan pemahaman tentang pentingnya asupan kalori yang cukup dan dampak faktor ekonomi terhadap asupan kalori yang cukup.





# SUMBER DAN PENGAMBILAN DATA

## 1. UMP 2024 dari website Catapa

```
def rupiah_to_int(s):
    s = s.strip()
    s = s.replace("Rp", "").replace(".", "").replace(",", "")
    try:
        return int(s)
    except:
        return None

def norm_prov(prov):
    prov = prov.strip()
    prov = prov.replace("Daerah Istimewa Yogyakarta", "D.I. Yogyakarta")
    return prov

url = "https://catapa.com/blog/daftar-lengkap-kenaikan-ump-2024-di-38-provinsi-indonesia"
resp = requests.get(url)
resp.raise_for_status()
soup = BeautifulSoup(resp.text, "html.parser")

paras = soup.find_all("p")
text_lines = []
for p in paras:
    text_lines.extend(p.get_text().split("\n"))

pattern = re.compile(r"^[A-Za-z\s\.\-]+\s*:\s*Rp([\d\.,]+)")

data = []
for line in text_lines:
    m = pattern.match(line)
    if m:
```

|   | Provinsi        | UMP_2024 |
|---|-----------------|----------|
| 0 | Aceh            | 3460672  |
| 1 | Sumatera Utara  | 2809915  |
| 2 | Sumatera Barat  | 2811499  |
| 3 | Kepulauan Riau  | 3402492  |
| 4 | Bangka Belitung | 3640000  |
| 5 | Riau            | 3294625  |

metode:  
webscrapping



## 2. Asupan Kalor perkuintil (PDF BPS)

metode: Ekstrak tabel pada PDF (pdfplumber)

```
import pdfplumber
import pandas as pd

with pdfplumber.open("1737605184791-81-23.-konsumsi-kalori-dan-protein-penduduk-indonesia-dan-provinsi--maret-2024.pdf") as pdf:
    page = pdf.pages[29]
    table = page.extract_table()

    df_kalori = pd.DataFrame(table[1:], columns=table[0])

    print(df_kalori)
```

|   | Provinsi\nProvince | Kuintil Pengeluaran\nQuintile of Expenditure | \ |
|---|--------------------|--|---|
| 0 | None               | Pertama\nFirst                               |   |
| 1 | (1)                | (2)  |   |
| 2 | Aceh               | 1.559,07                                     |   |
| 3 | Sumatera Utara     | 1.648,10                                     |   |
| 4 | Sumatera Barat     | 1.645,65                                     |   |
| 5 | Riau               | 1.573,72                                     |   |
| 6 | Jambi              | 1.588,60                                     |   |
| 7 | Sumatera Selatan   | 1.742,56                                     |   |

## 3. Prevalensi Kesehatan (PDF UNICEF)

metode: input CSV manual dan Read CSV

```
df_kesehatan = pd.read_csv("tabel_gizi_lengkap_part2.csv")
df_kesehatan.head(38)
```

|   | Provinsi         | Stunting | Wasting | Kelebihan berat badan | Anak 5-12 tahun | Remaja 13-15 tahun | Remaja 16-18 tahun | Dewasa > 18 tahun | Kerawanan pangan |
|---|------------------|----------|---------|-----------------------|-----------------|--------------------|--------------------|-------------------|------------------|
| 0 | Aceh             | 29.3     | 13.6    | 3.5                   | 17.2            | 15.3               | 12.1               | 39.9              | 4.7              |
| 1 | Banten           | 23.9     | 10.2    | 4.9                   | 19.3            | 15.6               | 10.5               | 36.9              | 5.76             |
| 2 | DKI Jakarta      | 17.6     | 10.1    | 6.7                   | 27.3            | 23.3               | 19.2               | 48.0              | 3.36             |
| 3 | Gorontalo        | 26.8     | 12.7    | 4.4                   | 17.0            | 14.3               | 14.8               | 42.5              | 5.96             |
| 4 | Jawa Barat       | 21.7     | 6.3     | 3.9                   | 18.4            | 17.0               | 13.1               | 39.8              | 4.9              |
| 5 | Kalimantan Barat | 24.5     | 13.3    | 5.5                   | 20.6            | 15.5               | 9.9                | 30.7              | 5.02             |

# CLEANING DATA

## 1. Membersihkan data asupan kalori per-kapita menggunakan regex

```
df_clean = df_kalori.copy()

# 1. Hapus baris kosong total
df_clean = df_clean.replace("", np.nan)
df_clean = df_clean.dropna(how="all")

# 2. Set header manual
df_clean.columns = [
    "Provinsi",
    "Kuintil Pertama",
    "Kuintil Kedua",
    "Kuintil Ketiga",
    "Kuintil Keempat",
    "Kuintil Kelima"
]

# 3. Drop baris "(1)" "(2)"
df_clean = df_clean[~df_clean["Provinsi"].str.contains(r"\(", na=False)]

# 4. Bersihkan kolom angka: HAPUS SEMUA HURUF dan simbol selain angka/.,,
angka_cols = ["Kuintil Pertama", "Kuintil Kedua", "Kuintil Ketiga",
               "Kuintil Keempat", "Kuintil Kelima"]

for col in angka_cols:
    df_clean[col] = df_clean[col].astype(str)
    df_clean[col] = df_clean[col].apply(lambda x: re.sub(r"[A-Za-z/\.:\\:\n\s]", "", x))
    # sekarang hanya tersisa "1661,76" atau "1.660,82"

# 5. Convert format Eropa
def europe(x):
    x = x.replace(".", "").replace(",", ".")
    try:
        return float(x)
    except:
        return np.nan

for col in angka_cols:
    df_clean[col] = df_clean[col].apply(europe)

# 6. Rapikan nama provinsi
df_clean["Provinsi"] = df_clean["Provinsi"].str.replace(r"\d+", "", regex=True)
df_clean["Provinsi"] = df_clean["Provinsi"].str.replace(r"\s+", " ", regex=True).str.strip()
df_clean = df_clean[df_clean["Provinsi"].str.len() > 2]

df_clean = df_clean.reset_index(drop=True)
```

## 2. Mengganti Nama Kolom

```
#GANTI NAMA KOLOM
df_kesehatan.rename(columns={
    'Anak 5-12 tahun': 'Kelebihan BB Anak 5-12 tahun',
    'Remaja 13-15 tahun': 'Kelebihan BB Remaja 13-15 tahun',
    'Remaja 16-18 tahun': 'Kelebihan BB Remaja 16-18 tahun',
    'Dewasa >18 tahun': 'Kelebihan BB Dewasa >18 tahun'
}, inplace=True)
df_kesehatan.head()
```

| Provinsi    | Stunting | Wasting | Kelebihan berat badan | Kelebihan BB Anak 5-12 tahun | Kelebihan BB Remaja 13-15 tahun | Kelebihan BB Remaja 16-18 tahun | Kelebihan BB Dewasa >18 tahun | Kerawanan pangan |
|-------------|----------|---------|-----------------------|------------------------------|---------------------------------|---------------------------------|-------------------------------|------------------|
| Aceh        | 29.3     | 13.6    | 3.5                   | 17.2                         | 15.3                            | 12.1                            | 39.9                          | 4.7              |
| Banten      | 23.9     | 10.2    | 4.9                   | 19.3                         | 15.6                            | 10.5                            | 36.9                          | 5.76             |
| DKI Jakarta | 17.6     | 10.1    | 6.7                   | 27.3                         | 23.3                            | 19.2                            | 48.0                          | 3.36             |

## 3. Menyeragamkan nama provinsi

```
#menghapus spasi awal dan akhir kolom provinsi
for df in [df_ump, df_clean_kalori, df_kesehatan]:
    df["Provinsi"] = df["Provinsi"].str.strip()

#Mengubah nama provinsi yang tidak sesuai
mapping = {
    "DI Yogyakarta": "D.I. Yogyakarta",
    "D I Yogyakarta": "D.I. Yogyakarta",
    "Maluku ": "Maluku",
    "Kep. Bangka Belitung": "Bangka Belitung",
}

for df in [df_ump, df_clean_kalori, df_kesehatan]:
    df["Provinsi"] = df["Provinsi"].replace(mapping)
```

# INTEGRASI DATA

metode: left merge

```
df_gabungan = df_ump.merge(df_clean_kalori, on="Provinsi", how="left").merge(df_kesehatan, on="Provinsi", how="left")
df_gabungan.head(38)
```

✓ 0.0s

Python

|   | Provinsi        | UMP_2024 | Kuintil Pertama | Kuintil Kedua | Kuintil Ketiga | Kuintil Keempat | Kuintil Kelima | Stunting | Wasting | Kelebihan berat badan | Kelebihan BB Anak 5-12 tahun | Kelebihan BB Remaja 13-15 tahun | Kelebihan BB Remaja 16-18 tahun | Kelebihan BB Dewasa >18 tahun | Kerawanan pangan |
|---|-----------------|----------|-----------------|---------------|----------------|-----------------|----------------|----------|---------|-----------------------|------------------------------|---------------------------------|---------------------------------|-------------------------------|------------------|
| 0 | Aceh            | 3460672  | 1559.07         | 1863.50       | 2038.54        | 2228.84         | 2574.43        | 29.3     | 13.6    | 3.5                   | 17.2                         | 15.3                            | 12.1                            | 39.9                          | 4.7              |
| 1 | Sumatera Utara  | 2809915  | 1648.10         | 1898.96       | 2070.20        | 2282.54         | 2568.40        | 18.9     | 7.9     | 3.5                   | 18.3                         | 14.4                            | 10.8                            | 40.9                          | 5.19             |
| 2 | Sumatera Barat  | 2811499  | 1645.65         | 1921.83       | 2103.17        | 2265.57         | 2494.01        | NaN      | NaN     | NaN                   | 16.2                         | 14.7                            | 9.5                             | NaN                           | NaN              |
| 3 | Kepulauan Riau  | 3402492  | 1660.66         | 1963.96       | 2070.37        | 2245.82         | 2297.70        | 16.8     | 6.3     | 4.6                   | 21.6                         | 16.5                            | 16.0                            | 44.2                          | 4.58             |
| 4 | Bangka Belitung | 3640000  | 1596.43         | 1890.34       | 2013.61        | 2261.71         | 2530.82        | NaN      | NaN     | NaN                   | NaN                          | NaN                             | NaN                             | NaN                           | NaN              |

## CLEANING SETELAH MERGE

### 1. Membersihkan Format Angka

```
#mengubah dan membersihkan kolom angka ke numeric
kolom = [
    "Kuintil Pertama", "Kuintil Kedua", "Kuintil Ketiga",
    "Kuintil Keempat", "Kuintil Kelima",
    "Stunting", "Wasting", "Kelebihan berat badan",
    "Kelebihan BB Anak 5-12 tahun", "Kelebihan BB Remaja 13-15 tahun",
    "Kelebihan BB Remaja 16-18 tahun", "Kelebihan BB Dewasa >18 tahun",
    "Kerawanan pangan"
]

for col in kolom:
    df_gabungan[col] = pd.to_numeric(df_gabungan[col], errors="coerce")

df_gabungan.dtypes
```

### 2. Cek duplikasi data

```
df_gabungan['Provinsi'].duplicated().sum()
```

✓ 0.0s

```
np.int64(0)
```



# PENANGANAN MISSING VALUE

1. Menghapus kolom dan baris dengan NaN lebih dari 50%

```
# Hapus kolom < 50% NaN
df_gabungan = df_gabungan.loc[:, df_gabungan.isna().mean() < 0.5]

# Hapus kolom < 50% NaN
df_gabungan = df_gabungan.loc[df_gabungan.isna().mean(axis=1) < 0.5]

# Cek hasilnya
print(df_gabungan.shape)

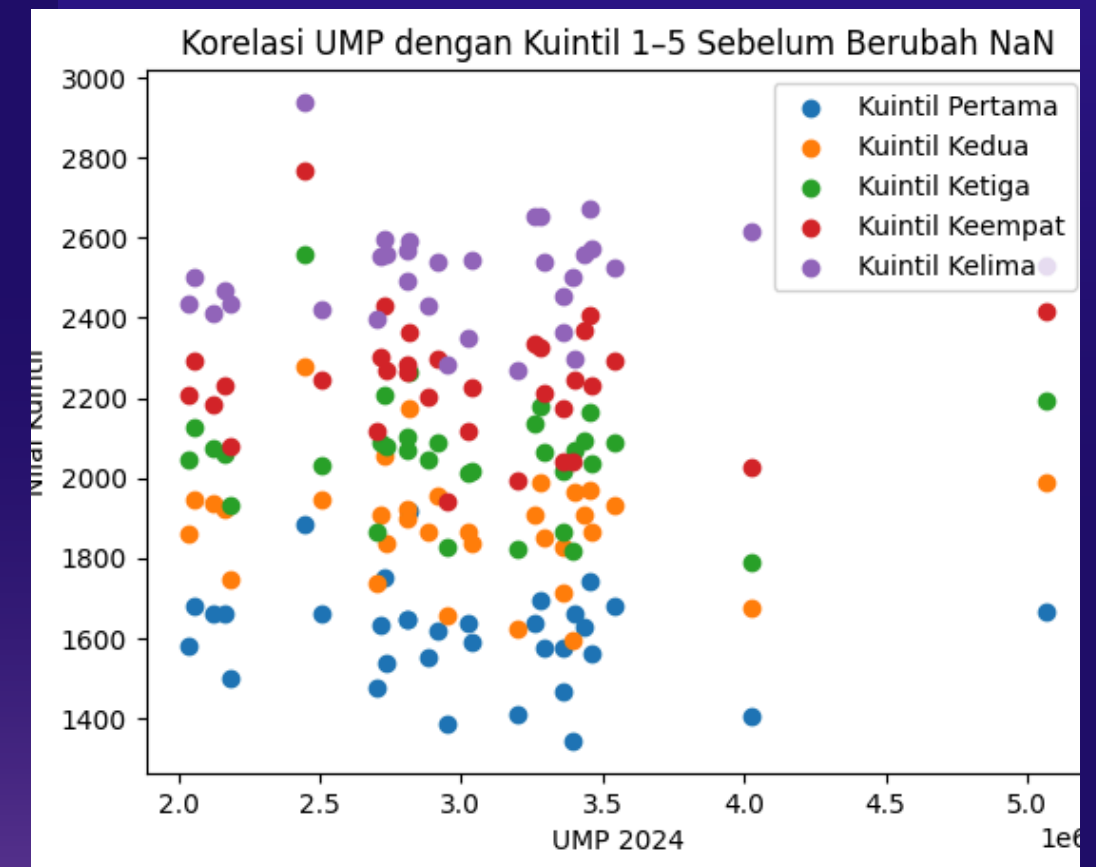
df_gabungan.head(38)
```

✓ 0.0s

(37, 15)

|   | Provinsi | UMP_2024 | Kuintil Pertama | Kuintil Kedua | Kuintil Ketiga | Kuintil Keempat |
|---|----------|----------|-----------------|---------------|----------------|-----------------|
| 0 | Aceh     | 3460672  | 1559.07         | 1863.50       | 2038.54        | 2228.84         |
| 1 | Bali     | 2813672  | 1918.83         | 2175.64       | 2266.56        | 2366.14         |
| 3 | Banten   | 2727812  | 1752.49         | 2053.64       | 2205.32        | 2428.25         |
| 4 | Bengkulu | 2507079  | 1661.21         | 1943.68       | 2030.50        | 2243.65         |

2. NaN Pada Provinsi Papua Pemekaran diisi dengan nilai Papua Induk



# PENANGANAN MISSING VALUE

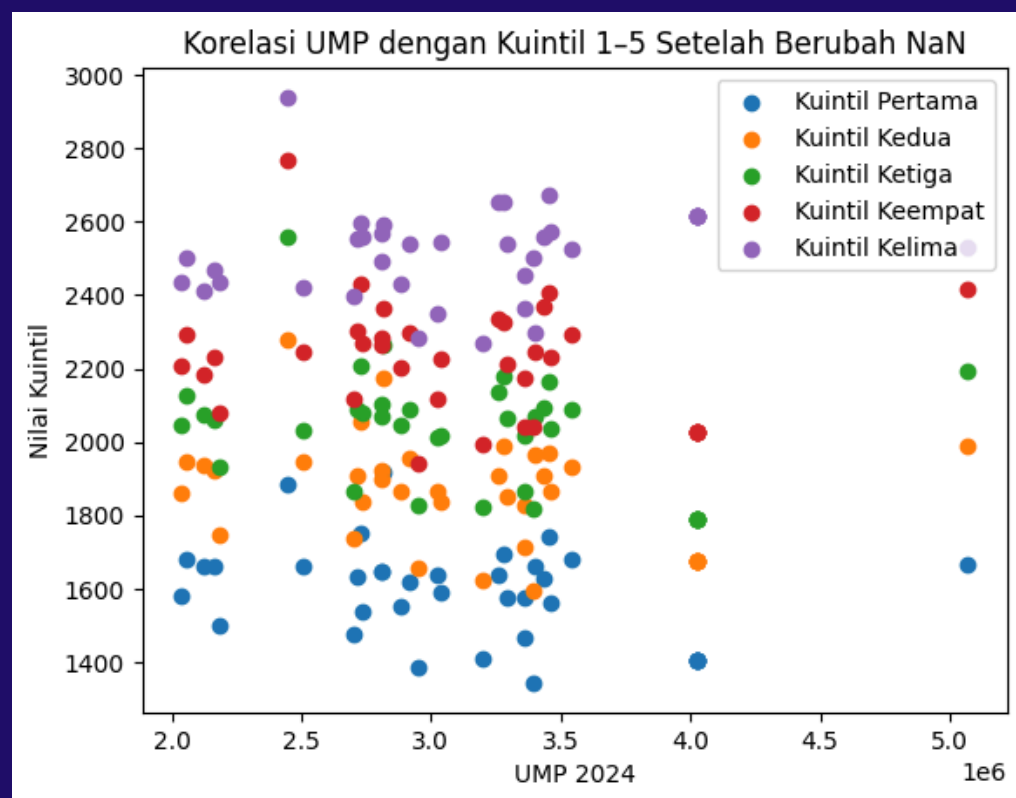
```
#ISI NAN PAPUA PEMEKARAN DENGAN NILAI PAPUA
kolom = ['Kuintil Pertama', 'Kuintil Kedua', 'Kuintil Ketiga', 'Kuintil Keempat', 'Kuintil Kelima']
baris = df_gabungan['Provinsi'].isin(['Papua Pegunungan', 'Papua Barat Daya', 'Papua Selatan', 'Papua Tengah'])

df_gabungan.loc[baris, kolom] = [1402.98, 1674.22, 1790.51, 2028.2, 2617.53]
(variable) kolom: list[str]

df_gabungan.loc[baris, ['Provinsi'] + kolom]
```

✓ 0.0s

|    | Provinsi         | Kuintil Pertama | Kuintil Kedua | Kuintil Ketiga | Kuintil Keempat | Kuintil Kelima |
|----|------------------|-----------------|---------------|----------------|-----------------|----------------|
| 25 | Papua Barat Daya | 1402.98         | 1674.22       | 1790.51        | 2028.2          | 2617.53        |
| 26 | Papua Pegunungan | 1402.98         | 1674.22       | 1790.51        | 2028.2          | 2617.53        |
| 27 | Papua Selatan    | 1402.98         | 1674.22       | 1790.51        | 2028.2          | 2617.53        |
| 28 | Papua Tengah     | 1402.98         | 1674.22       | 1790.51        | 2028.2          | 2617.53        |



## 3. Provinsi lain NaN diisi dengan median kolom

```
#UNTUK PROVINSI PEMEKARAN PAPUA
provinsi_pemekaran_papua = [
    'Papua Pegunungan', 'Papua Barat Daya',
    'Papua Selatan', 'Papua Tengah']

filter_papua = df_gabungan['Provinsi'].isin(provinsi_pemekaran_papua)
filter_non_papua = df_gabungan['Provinsi'].isin(provinsi_pemekaran_papua) == False

# Untuk provinsi pemekaran Papua → samakan dengan Papua induk
nilai_papua_induk = df_gabungan.loc[df_gabungan['Provinsi'] == 'Papua', 'Kerawanan pangan'].values[0]
df_gabungan.loc[filter_papua, 'Kerawanan pangan'] = nilai_papua_induk

# Untuk provinsi lain, isi dengan median nasional
median_nasional = df_gabungan.loc[filter_non_papua, 'Kerawanan pangan'].median()
df_gabungan.loc[df_gabungan['Kerawanan pangan'].isna() & filter_non_papua, 'Kerawanan pangan'] = median_nasional
```

# PENANGANAN OUTLIER

1. Menghapus baris dengan outlier lebih dari tiga nilai

```
# Hitung jumlah outlier per baris
outlier_flags = (df_gabungan[numeric_cols] < lower_bound) | (df_gabungan[numeric_cols] > upper_bound)
df_gabungan['outlier_count'] = outlier_flags.sum(axis=1)

# Cek berapa baris yang outlier_count > 3
print(df_gabungan[df_gabungan['outlier_count'] > 3])
```

✓ 0.0s

Python

|    | Provinsi            | UMP_2024 | Kuintil Pertama | Kuintil Kedua | \ |
|----|---------------------|----------|-----------------|---------------|---|
| 5  | DKI Jakarta         | 5067381  | 1665.07         | 1987.44       |   |
| 20 | Nusa Tenggara Barat | 2444067  | 1883.94         | 2276.32       |   |
| 27 | Papua Tengah        | 4024270  | 1402.98         | 1674.22       |   |

## Dataset Clean

df\_clean.head(38)

✓ 0.0s

Python

|   | Provinsi        | UMP_2024 | Kuintil Pertama | Kuintil Kedua | Kuintil Ketiga | Kuintil Keempat | Kuintil Kelima | Stunting | Wasting | Kelebihan berat badan | Kelebihan BB Anak 5-12 tahun | Kelebihan BB Remaja 13-15 tahun | Kelebihan BB Remaja 16-18 tahun | Kelebihan BB Dewasa >18 tahun | Kerawanan pangan |
|---|-----------------|----------|-----------------|---------------|----------------|-----------------|----------------|----------|---------|-----------------------|------------------------------|---------------------------------|---------------------------------|-------------------------------|------------------|
| 0 | Aceh            | 3460672  | 1559.07         | 1863.50       | 2038.54        | 2228.84         | 2574.43        | 24.6     | 13.60   | 4.7                   | 17.2                         | 15.30                           | 12.1                            | 39.9                          | 4.70             |
| 1 | Bali            | 2813672  | 1918.83         | 2175.64       | 2266.56        | 2366.14         | 2590.63        | 24.6     | 10.55   | 4.7                   | 26.7                         | 20.00                           | 13.2                            | 39.8                          | 4.80             |
| 2 | Banten          | 2727812  | 1752.49         | 2053.64       | 2205.32        | 2428.25         | 2597.54        | 23.9     | 10.20   | 4.9                   | 19.3                         | 15.60                           | 10.5                            | 36.9                          | 5.76             |
| 3 | Bengkulu        | 2507079  | 1661.21         | 1943.68       | 2030.50        | 2243.65         | 2420.84        | 24.6     | 10.55   | 4.7                   | 23.6                         | 15.20                           | 7.7                             | 39.4                          | 4.80             |
| 4 | D.I. Yogyakarta | 2125897  | 1661.76         | 1935.91       | 2072.51        | 2181.61         | 2412.39        | 24.6     | 10.55   | 4.7                   | 21.6                         | 15.25                           | 11.1                            | 39.0                          | 4.80             |
| 6 | Gorontalo       | 3025100  | 1635.97         | 1864.76       | 2014.22        | 2118.37         | 2347.49        | 26.8     | 12.70   | 4.4                   | 17.0                         | 14.30                           | 14.8                            | 42.5                          | 5.96             |

2. isi dengan median jika otlier dalam baris hanya 1 atau 2

```
df_clean = df_gabungan[df_gabungan['outlier_count'] <= 3].copy()
print(f"Jumlah baris setelah penghapusan (>3 outlier): {len(df_clean)}")
print("-" * 50)

# Ganti Outlier Sisanya dengan Median (Menggunakan Batas BARU)
for col in numeric_cols:
    Q1_new = df_clean[col].quantile(0.25)
    Q3_new = df_clean[col].quantile(0.75)
    IQR_new = Q3_new - Q1_new

    lower_new = Q1_new - 1.5 * IQR_new
    upper_new = Q3_new + 1.5 * IQR_new
    median_new = df_clean[col].median()

    outliers_mask = (df_clean[col] < lower_new) | (df_clean[col] > upper_new)

    count_replaced = outliers_mask.sum()
    if count_replaced > 0:
        df_clean.loc[outliers_mask, col] = median_new
        print(f"Kolom '{col}': {count_replaced} nilai outlier diganti dengan median ({median_new:.2f})")

print("-" * 50)

df_clean = df_clean.drop(columns='outlier_count')

print("✅ Penanganan Outlier Selesai.")
```

✓ 0.0s

Jumlah baris setelah penghapusan (>3 outlier): 34

-----  
Kolom 'Stunting': 13 nilai outlier diganti dengan median (24.60)  
Kolom 'Wasting': 2 nilai outlier diganti dengan median (10.55)  
Kolom 'Kelebihan berat badan': 10 nilai outlier diganti dengan median (4.70)  
Kolom 'Kelebihan BB Anak 5-12 tahun': 1 nilai outlier diganti dengan median (18.80)  
Kolom 'Kelebihan BB Remaja 13-15 tahun': 5 nilai outlier diganti dengan median (15.25)  
Kolom 'Kelebihan BB Remaja 16-18 tahun': 1 nilai outlier diganti dengan median (11.10)  
Kolom 'Kerawanan pangan': 6 nilai outlier diganti dengan median (4.80)  
-----

✅ Penanganan Outlier Selesai.

# EKSPLORASI DATA

## 1. Statistika Deskriptif

```
#!/Descriptive statistics
df_clean.describe()
```

✓ 0.0s

Python

|       | UMP_2024     | Kuintil Pertama | Kuintil Kedua | Kuintil Ketiga | Kuintil Keempat | Kuintil Kelima | Stunting  | Wasting   | Kelebihan berat badan | Kelebihan BB Anak 5-12 tahun | Kelebihan BB Remaja 13-15 tahun | Kelebihan BB Remaja 16-18 tahun | Kelebihan BB Dewasa > 18 tahun | Kerawanan pangan |
|-------|--------------|-----------------|---------------|----------------|-----------------|----------------|-----------|-----------|-----------------------|------------------------------|---------------------------------|---------------------------------|--------------------------------|------------------|
| count | 3.400000e+01 | 34.000000       | 34.000000     | 34.000000      | 34.000000       | 34.000000      | 34.000000 | 34.000000 | 34.000000             | 34.000000                    | 34.000000                       | 34.000000                       | 34.000000                      | 34.000000        |
| mean  | 3.051832e+06 | 1579.717353     | 1850.395882   | 2013.317059    | 2200.867059     | 2503.226471    | 24.494118 | 10.841176 | 4.700000              | 19.047059                    | 15.242647                       | 11.138235                       | 38.626471                      | 4.817353         |
| std   | 5.637724e+05 | 127.255918      | 132.637101    | 133.679480     | 128.979863      | 109.793343     | 0.740983  | 1.600638  | 0.161433              | 3.908285                     | 2.120111                        | 3.142284                        | 4.036223                       | 0.646099         |
| min   | 2.036947e+06 | 1343.240000     | 1596.200000   | 1790.510000    | 1943.430000     | 2269.370000    | 21.700000 | 7.300000  | 4.300000              | 9.900000                     | 10.500000                       | 4.700000                        | 30.700000                      | 3.630000         |
| 25%   | 2.730034e+06 | 1481.997500     | 1740.512500   | 1883.325000    | 2088.557500     | 2432.857500    | 24.600000 | 10.287500 | 4.700000              | 16.225000                    | 14.400000                       | 9.525000                        | 35.850000                      | 4.800000         |
| 50%   | 3.031110e+06 | 1602.440000     | 1865.805000   | 2053.260000    | 2227.225000     | 2532.610000    | 24.600000 | 10.550000 | 4.700000              | 19.000000                    | 15.250000                       | 10.950000                       | 39.400000                      | 4.800000         |
| 75%   | 3.400119e+06 | 1660.780000     | 1935.017500   | 2089.130000    | 2293.542500     | 2586.580000    | 24.600000 | 11.950000 | 4.700000              | 21.400000                    | 15.950000                       | 13.050000                       | 40.650000                      | 4.800000         |
| max   | 4.024270e+06 | 1918.830000     | 2175.640000   | 2266.560000    | 2428.250000     | 2670.340000    | 26.800000 | 13.800000 | 5.200000              | 28.200000                    | 20.000000                       | 17.200000                       | 47.500000                      | 6.710000         |

Tabel ini menunjukkan ringkasan seluruh variabel, mulai dari UMP, konsumsi energi, hingga indikator gizi. Terlihat bahwa UMP dan konsumsi energi bervariasi antar provinsi, sementara stunting dan wasting cenderung seragam. Kelebihan berat badan cukup tinggi pada kelompok dewasa, dan kerawanan pangan masih tampak berbeda antar wilayah

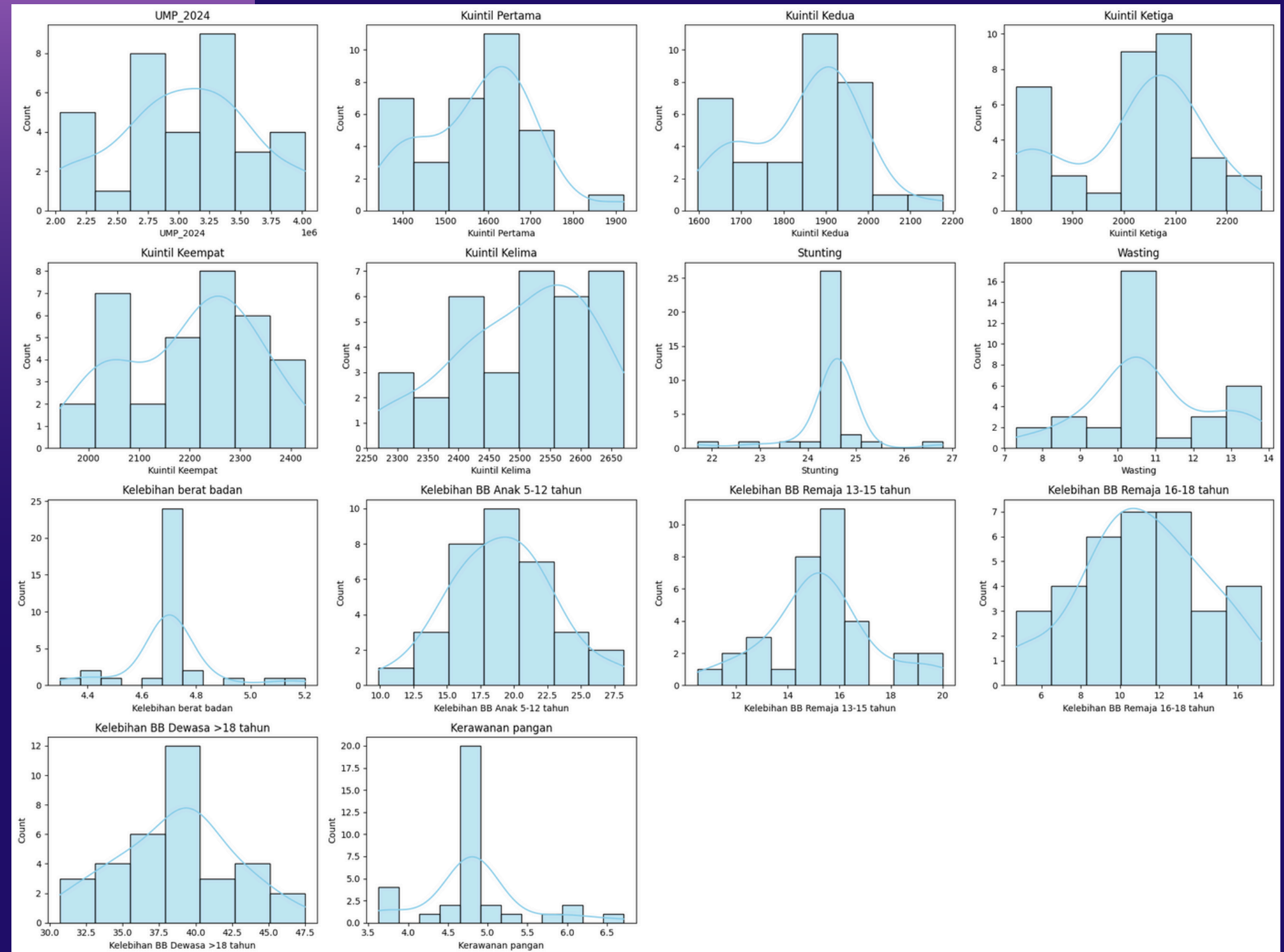


# EKSPLORASI DATA

## 2. Distribusi Data Perkolom

```
plt.figure(figsize=(20,15))
for i, col in enumerate(numeric_cols):
    plt.subplot(4,4,i+1)
    sns.histplot(df_eda[col], kde=True, color='skyblue')
    plt.title(col)
plt.tight_layout()
plt.show()
```

Distribusi UMP terlihat cukup merata namun sedikit condong ke kanan, dengan mayoritas provinsi berada pada rentang 2,7–3,4 juta. Distribusi asupan kalori per kuintil menunjukkan pola yang semakin naik dan semakin stabil, mulai dari kuintil pertama yang terkonsentrasi pada 1.450–1.750 kalori hingga kuintil kelima yang dominan pada 2.400–2.650 kalori. Variabel kesehatan memiliki pola yang berbeda: stunting sangat mengerucut di 24–25%, wasting menyebar di 10–12% dengan sedikit nilai rendah, dan overweight balita berada di kisaran sempit 4,6–4,8%. Pada kelompok usia overweight, anak 5–12 tahun memiliki pola mendekati normal (16–22%), remaja 13–15 tahun menunjukkan variasi lebih besar, remaja 16–18 tahun data lebih banyak menyebar ke kiri, sedangkan dewasa menunjukkan variasi paling lebar (30–47%). Sementara itu, distribusi kerawanan pangan condong ke kanan, dengan banyak provinsi berada pada skor 4–5, namun beberapa mencapai nilai lebih tinggi hingga sekitar 6,7.

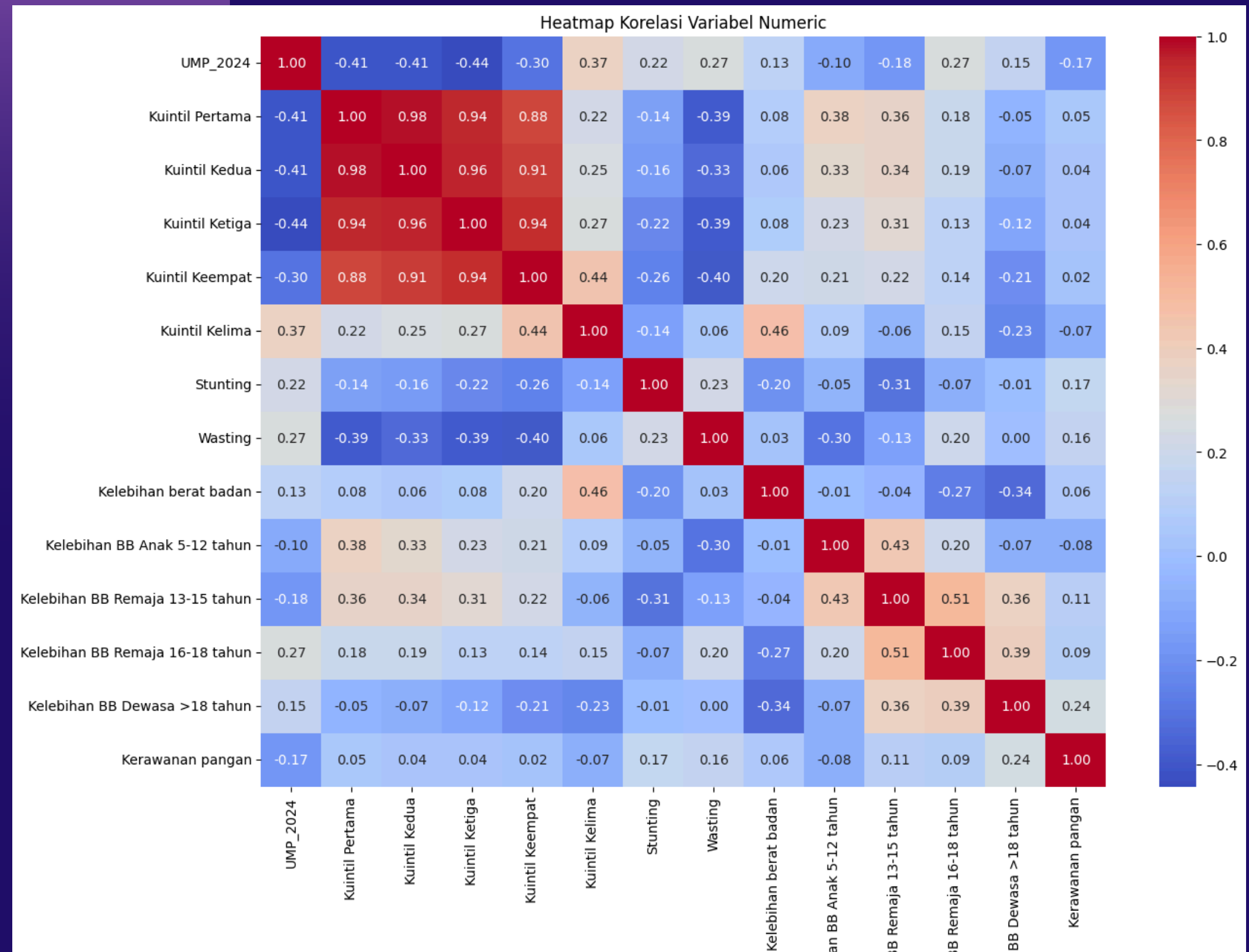


# EKSPLORASI DATA

## 3. Heatmap atau Korelasi Antar Variabel

```
plt.figure(figsize=(14,10))
sns.heatmap(df_eda[numeric_cols].corr(), annot=True, fmt=".2f", cmap="coolwarm")
plt.title("Heatmap Korelasi Variabel Numeric")
plt.show()
```

Heatmap menunjukkan bahwa kelima kuintil asupan kalori memiliki korelasi yang sangat kuat mendekati 1.0. UMP terlihat memiliki korelasi sangat lemah dengan stunting, wasting, maupun kelebihan berat badan mendekati 0. Sementara itu, UMP, stunting, wasting, dan overweight memiliki korelasi sangat lemah, menandakan bahwa upah tidak berkaitan langsung dengan kondisi kesehatan. indikator overweight antar kelompok usia menunjukkan korelasi 0,24–0,5 menunjukkan bahwa kelebihan berat badan pada anak cenderung berlanjut hingga remaja dan dewasa.

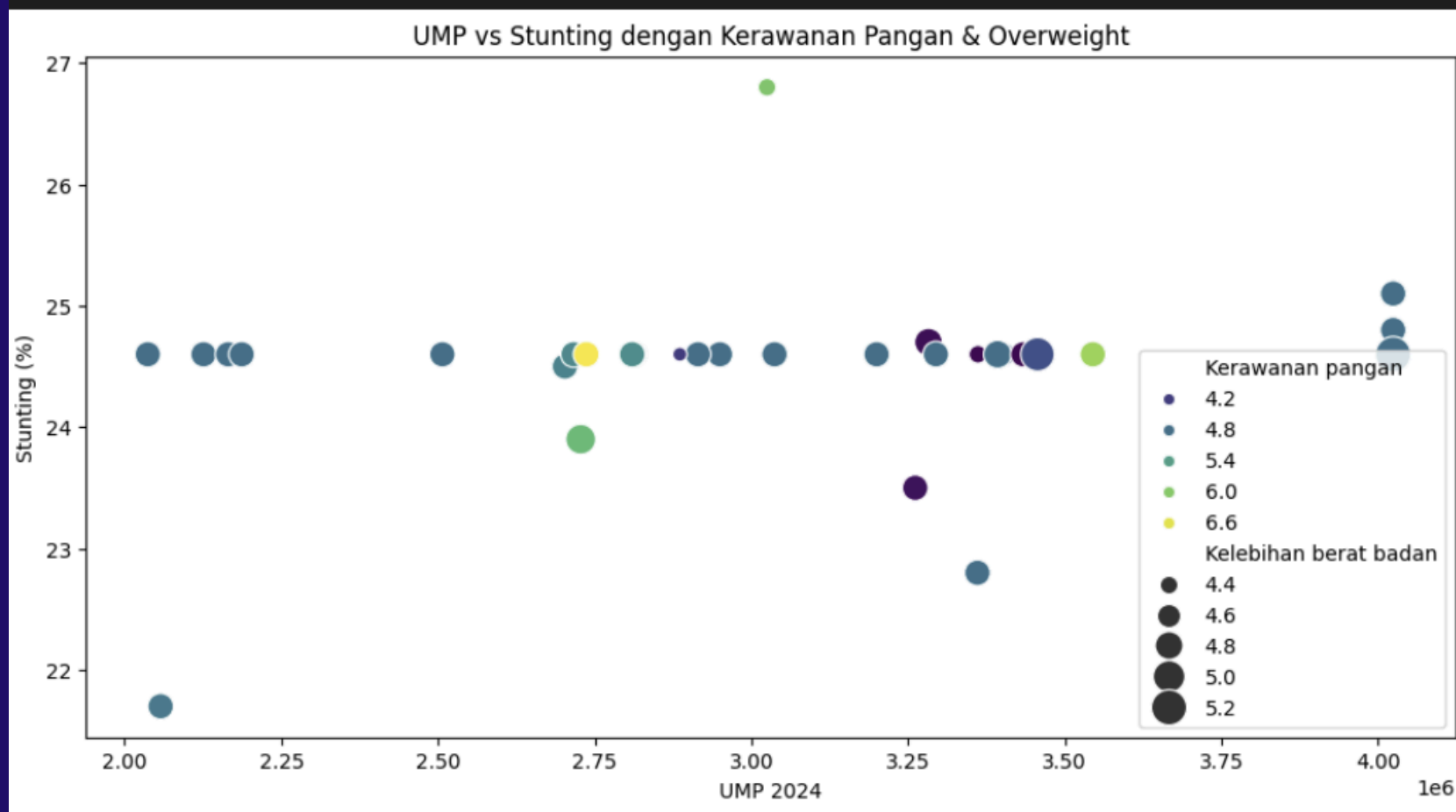


# EKSPLORASI DATA

## 4. Analisis Scatterplot

```
plt.figure(figsize=(12,6))
sns.scatterplot(data=df_eda, x='UMP_2024', y='Stunting', hue='Kerawanan pangan', size='Kelebihan berat badan', palette='viridis', sizes=(50,300))
plt.title("UMP vs Stunting dengan Kerawanan Pangan & Overweight")
plt.xlabel("UMP 2024")
plt.ylabel("Stunting (%)")
plt.show()
```

0.1s



Hasil Scatter plot menunjukkan bahwa UMP tidak berkorelasi secara linear dengan stunting. Sedangkan kerawanan pangan lebih dipengaruhi oleh tingginya stunting. Ukuran bubble yang besar menunjukkan kelebihan berat badan yang tinggi. Nilai UMP yang tinggi menunjukkan bahwa ukuran bubble besar. Sehingga tingkat UMP mempengaruhi kelebihan berat badan

# EKSPLORASI DATA

## 5. Barplot top 5 provinsi

```
# Daftar indikator yang mau ditampilkan
metrics = ['Stunting', 'Wasting', 'Kelebihan berat badan',
           'Kelebihan BB Anak 5-12 tahun', 'Kelebihan BB Remaja 13-15 tahun',
           'Kelebihan BB Remaja 16-18 tahun', 'Kelebihan BB Dewasa >18 tahun',
           'Kerawanan pangan']

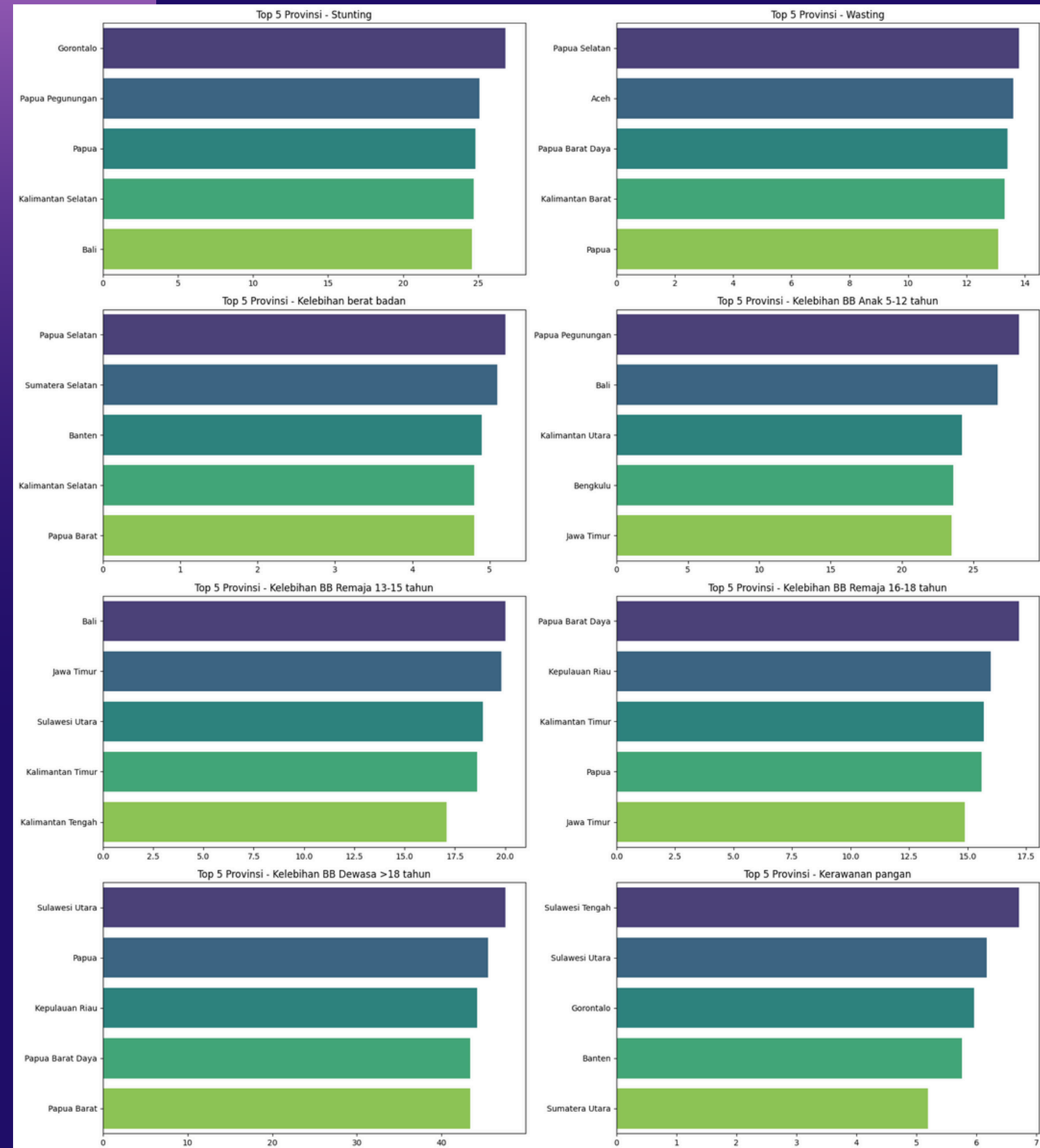
fig, axes = plt.subplots(4, 2, figsize=(18, 20))
axes = axes.flatten()

for i, col in enumerate(metrics):
    # Ambil top 5 provinsi
    top5 = df_eda[['Provinsi', col]].sort_values(by=col, ascending=False).head(5)

    sns.barplot(data=top5, x=col, y='Provinsi', ax=axes[i], palette='viridis')
    axes[i].set_title(f'Top 5 Provinsi - {col}')
    axes[i].set_xlabel('')
    axes[i].set_ylabel('')

plt.tight_layout()
plt.show()
```

Barplot menunjukkan setiap indikator gizi memiliki Provinsi dengan nilai tertinggi yang berbeda beda. Stunting dan wasting banyak terjadi di wilayah timur seperti Papua dan Gorontalo, sedangkan kelebihan berat badan lebih tinggi di provinsi seperti Papua Selatan, Sumatera Selatan, Banten, Bali, dan Jawa Timur. Kerawanan pangan tertinggi muncul di Sulawesi Tengah, Sulawesi Utara, dan Gorontalo. Secara keseluruhan, tiap provinsi menghadapi masalah gizi yang berbeda, mulai dari stunting hingga kelebihan berat badan dan kerawanan pangan





# KESIMPULAN



Analisis hubungan antara UMP, asupan kalori (kuintil), dan indikator kesehatan menunjukkan bahwa faktor ekonomi tidak memiliki pengaruh langsung terhadap penurunan stunting. Meskipun beberapa provinsi memiliki UMP tinggi hingga mendekati 4 juta rupiah, angka stunting tetap berada pada kisaran 24–25 persen, hampir sama dengan provinsi ber-UMP rendah. Korelasi heatmap juga memperlihatkan bahwa hubungan antara UMP dan stunting sangat lemah. Sebaliknya, korelasi antar kuintil asupan kalori sangat kuat (mendekati 1), menandakan bahwa jika kuintil pertama mengalami peningkatan jumlah asupan kalori maka kuintil di atasnya juga mengalami peningkatan. Kehadiran outlier pada indikator kesehatan mengindikasikan bahwa kondisi gizi masyarakat tidak hanya dipengaruhi pendapatan, tetapi juga faktor struktural lain seperti kerawanan pangan.

# RENCANA TINDAK LANJUT

1. Menambahkan variabel lain seperti pendidikan, sanitasi, harga pangan, dan akses kesehatan untuk memperkaya analisis.
2. Menggunakan metode lanjutan seperti regresi, clustering, atau machine learning untuk mengidentifikasi faktor yang paling berpengaruh antar variabel
3. Melakukan validasi dengan data tahun berbeda serta analisis spasial untuk melihat pola persebaran masalah gizi antar provinsi.

...



# TERIMA KASIH

