

# **ANALISIS DATA SEKOLAH DASAR DI DAERAH JAWA TIMUR UNTUK EVALUASI RASIO GURU-MURID DAN PENYEBARAN SEKOLAH**

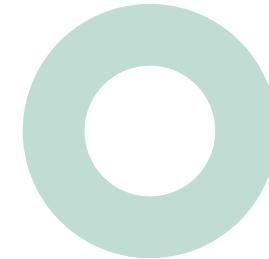
**Muhamad Fadly syaifullah (24031554173)  
Atya Azharina Aziz (24031554194)**

MATA KULIAH: DATA WRANGLING

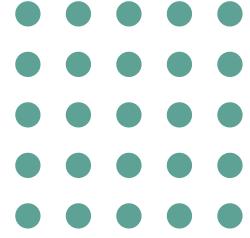




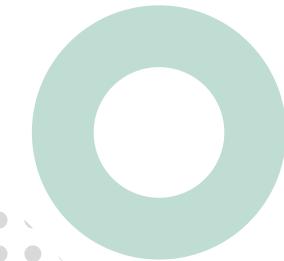
# Latar Belakang



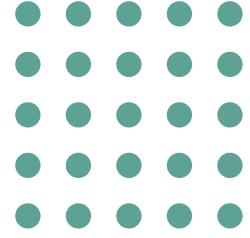
Pemerataan pendidikan dasar di Indonesia masih menjadi tantangan, khususnya di Jawa Timur yang memiliki kondisi geografis dan sosial ekonomi yang beragam sehingga menimbulkan ketimpangan jumlah sekolah dasar antar kabupaten/kota. Ketimpangan ini berdampak pada akses dan kesempatan belajar anak, terutama di daerah pedesaan dan tertinggal



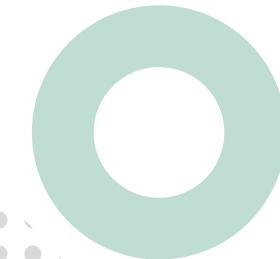
# Rumusan Masalah



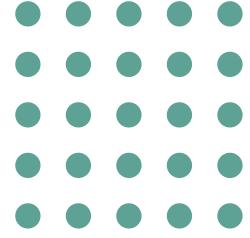
1. Bagaimana sebaran nilai Indeks Pembangunan Manusia (IPM) di Jawa Timur pada tahun 2024?
2. Kabupaten/Kota mana saja yang memiliki nilai IPM tertinggi di Jawa Timur?
3. Apakah ada hubungan visual antara jumlah Sekolah Dasar (SD) dengan nilai IPM di setiap wilayah?
4. Bagaimana perbandingan Angka Partisipasi Kasar (APK) pendidikan di tiap kabupaten/kota terhadap Angka Partisipasi Sekolah (APS) rata-rata provinsi?



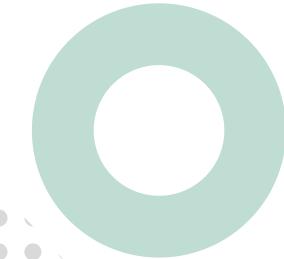
# Tujuan Analisis



1. Menampilkan distribusi IPM tahun 2024 menggunakan histogram dan boxplot
2. Menyajikan 10 kabupaten/kota dengan IPM tertinggi dalam bentuk diagram batang.
3. Memvisualisasikan potensi hubungan antara jumlah sekolah SD dan IPM menggunakan diagram sebar (scatter plot).
4. Membandingkan APK setiap kabupaten/kota dengan nilai APS provinsi menggunakan diagram batang.



# Manfaat



1. Akademik: Membantu mahasiswa memahami penerapan data wrangling untuk menganalisis hubungan ketersediaan sekolah dan IPM.
2. Praktis: Menjadi bahan masukan bagi pemerintah daerah dalam kebijakan pemerataan SD di Jawa Timur.
3. Sosial: Memberikan gambaran kesenjangan IPM dan infrastruktur pendidikan dasar kepada masyarakat.
4. Teknologis: Menunjukkan pemanfaatan Python dan pustaka sains data dalam pengolahan dan visualisasi data pendidikan.

# TINJAUAN PUSTAKA

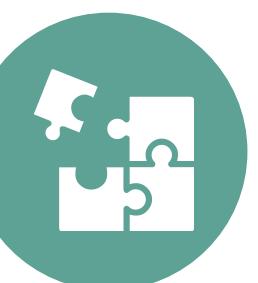
## Konsep Pemerataan Pendidikan di Indonesia

perlu memahami pola ketimpangan agar kebijakan pendidikan dapat dibahas secara lebih akurat(BPS Jawa Timur, 2024).



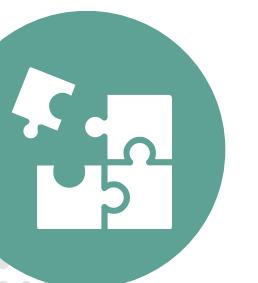
## Pemerataan Pendidikan Berdasarkan SDGs

- SDGs 4: Pendidikan berkualitas dan inklusif.
- Fokus: meningkatkan kualitas pendidikan, kompetensi guru, dan akses pembelajaran sepanjang hayat.



## Asta Cita Nomor 4

- Fokus: kualitas pendidikan, pemerataan infrastruktur, peningkatan kompetensi guru, digitalisasi pembelajaran.



## Dasar Hukum Pemerataan Pendidikan

- Pasal 31 UUD 1945: setiap warga berhak memperoleh pendidikan.
- PP No. 19/2005 & Permendikbud No. 23/2013: peningkatan partisipasi siswa, pemerataan guru, penyediaan fasilitas.



## Data Wrangling

- Proses mengubah, membersihkan, dan menyiapkan data mentah agar siap dianalisis.
- Meliputi: identifikasi kesalahan, pembersihan, standarisasi, integrasi, dan penanganan inkonsistensi.



## Pengolahan Data pada Tahapan Penelitian

- Tahapan utama: pengumpulan, pembersihan, integrasi, transformasi, visualisasi.
- Tools: Python (pandas, numpy, matplotlib).
- Sumber: BPS Jawa Timur & data pendidikan resmi lainnya.



# Metode Penelitian



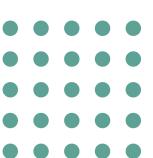
## Sumber Data

- Analisis Keterkaitan IPM dan APK APM (Sekretariat Jenderal Kementerian).
- Jumlah sekolah, guru, murid SD (BPS).
- APS Provinsi(Surabaya, Harian Disway)
- APS per kabupaten/Kota (BPS)



## Data Wrangling

- Data Collection
- Data Cleaning (rename kolom, outlier).
- Data Integration (merge dataset 38 kabupaten/kota).
- Data Analysis (EDA: distribusi IPM, top 10, scatter plot, APK vs APS).
- Data Visualization (scatter, histogram, bar chart).



# Pipeline Data Wrangling Proyek Analisis SD di Jawa Timur

## Pengumpulan Data (Data Collection)

- Mencari dan mengumpulkan data dari berbagai sumber: PDF, CSV, dan web.
- Menentukan dataset yang relevan untuk analisis, misal IPM, jumlah sekolah, guru, murid, APK, APM, APS.

## Pembersihan Data (Data Cleaning)

- Memeriksa data yang kosong (null/NaN) dan memastikan tidak ada nilai hilang.
- Menghapus duplikasi (duplicate data).
- Mengubah format angka (dari koma ke titik) dan mengubah tipe data menjadi numerik agar bisa dianalisis.
- Mengidentifikasi outlier dan menangani nilai ekstrem dengan metode IQR atau Winsorization.
- Menstandarisasi nama kolom agar konsisten dan mudah diakses.

## Integrasi Data (Data Integration)

- Menstandarisasi nama wilayah menjadi kolom kunci (norm\_name) agar bisa digabungkan antar dataset.
- Menggabungkan semua dataset menjadi satu dataset final.
- Menghapus kolom yang tidak diperlukan (redundant) dan menyimpan hasilnya (jatim\_merged\_final.csv)

## Pembahasan Hasil

- Menafsirkan temuan: kesenjangan IPM, variasi rasio murid-guru, hubungan jumlah sekolah dengan kualitas pendidikan.
- Menjelaskan korelasi antar variabel.
- Memberikan insight untuk kebijakan dan tindak lanjut.

## Visualisasi dan Publikasi (Data Visualization & Publishing)

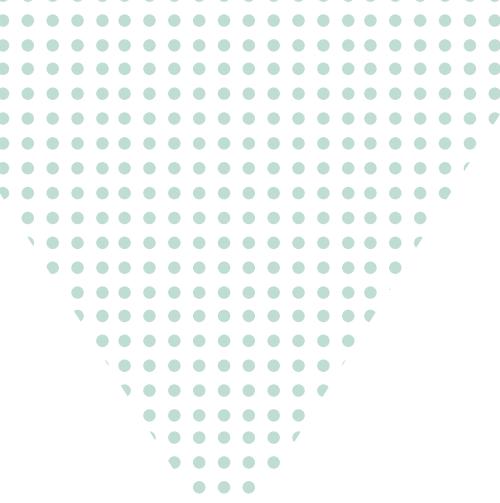
- Membuat visualisasi yang informatif dan estetis (judul, label, ukuran figur, warna konsisten).
- Menyimpan dataset final yang telah bersih dan terintegrasi.
- Menyusun visualisasi untuk mendukung laporan akhir.

## Analisis dan Eksplorasi Data (Data Analysis & Exploration)

- Melihat distribusi variabel utama: IPM, jumlah sekolah, guru, murid, rasio murid-guru, APK, APM, APS.
- Membuat visualisasi: histogram, boxplot, scatter plot, diagram batang, heatmap korelasi.
- Mengeksplorasi hubungan antar variabel, misal:
- Jumlah sekolah vs IPM
- Rasio murid-guru vs IPM/APM/APK
- Perbandingan wilayah urban vs rural



# Tahapan Data Wrangling



```
import pandas as pd
raw_path = r"C:\Users\atyaa\OneDrive\Pictures\RAIN\SEM 3\UAS.DK\project_wrangle\raw_data"

file_excel = f"{raw_path}\jatin.xlsx"
df_excel = pd.read_excel(file_excel)

file_csv = f"{raw_path}\Jumlah Sekolah, Guru, dan Murid Sekolah Dasar (SD) di Bawah Kementerian Pendidikan, Kebudayaan, Riset,.csv"
df_csv = pd.read_csv(file_csv)

print("Excel Preview:")
display(df_excel.head())

print("\nCSV Preview:")
display(df_csv.head())
```

CSV Preview:

Kabupaten/Kota	Jumlah Sekolah SD (Negeri)	Jumlah Sekolah SD (Swasta)	Jumlah Sekolah SD (Negeri+Swasta)	Jumlah Guru SD (Negeri)	Jumlah Guru SD (Swasta)	Jumlah Guru SD (Negeri+Swasta)	Jumlah Murid SD (Negeri)	Jumlah Murid SD (Swasta)
0 Pacitan	408	11	419	3614	161	3775	29126	27
1 Ponorogo	550	21	571	4031	390	4421	40943	65
2 Trenggalek	417	21	438	3405	239	3644	35871	37
3 Tulungagung	569	63	632	5587	810	6397	52132	110
4 Blitar	634	31	666	5629	371	6000	54552	65

Excel Preview:

Kabupaten / Kota	Angka Partisipasi Kasar 2024	Angka Partisipasi Murni 2024	Usia Harapan Hidup 2024	Harapan Lama Sekolah 2024	Rata-Rata Lama Sekolah 2024	Pengeluaran Rill per Kapita per Tahun yang Disesuaikan 2024	Indeks Pembangunan Manusia 2024	Kuadra IPM - APA
0 Kab. Pacitan	101,87	95,28	74,74	12,69	7,90	10.099	71,49	
1 Kab. Ponorogo	101,81	96,63	75,28	13,78	7,80	11.065	73,70	
2 Kab. Trenggalek	100,64	96,39	75,35	12,63	7,92	10.872	72,47	
3 Kab. Tulungagung	103,40	99,50	75,20	13,36	8,68	11.966	75,13	
4 Kab. Blitar	100,08	95,91	75,32	12,67	7,87	12.020	73,44	

Data yang ditampilkan berisi indikator pendidikan dan kesejahteraan dari berbagai kabupaten/kota di Indonesia untuk tahun 2024.

# Memulai Pre-Processing

```
df_excel.columns = df_excel.columns.str.lower().str.replace(' ', '_')
df_csv.columns = df_csv.columns.str.lower().str.replace(' ', '_')

print(df_excel.columns)
print(df_csv.columns)

Index(['kabupaten/_kota', 'angka_partisipasi_kasar_2024',
       'angka_partisipasi_murni_2024', 'usia_harapan_hidup_2024',
       'harapan_lama_sekolah_2024', 'rata-rata_lama_sekolah_2024',
       'pengeluaran_rill_per_kapita_per_tahun_yang_disesuaikan_2024',
       'indeks_pembangunan_manusia_2024', 'kuadran_ipm_-_apm'],
      dtype='object')
Index(['kabupaten/kota', 'jumlah_sekolah_sd_(negeri)',
       'jumlah_sekolah_sd_(swasta)', 'jumlah_sekolah_sd_(negeri+swasta)',
       'jumlah_guru_sd_(negeri)', 'jumlah_guru_sd_(swasta)',
       'jumlah_guru_sd_(negeri+swasta)', 'jumlah_murid_sd_(negeri)',
       'jumlah_murid_sd_(swasta)', 'jumlah_murid_sd_(negeri+swasta)'],
      dtype='object')
```

Langkah pertama dalam pra-pemrosesan data adalah menstandarisasi nama kolom pada dataset agar konsisten dan mudah diakses. Pada kode ini, semua nama kolom diubah menjadi huruf kecil dan spasi diganti dengan underscore.

# Memeriksa nilai null/NaN

```
print("Missing values in Excel:")
print(df_excel.isna().sum())

print("\nMissing values in CSV:")
print(df_csv.isna().sum())
```

```
Missing values in Excel:
kabupaten/_kota          0
angka_partisipasi_kasar_2024 0
angka_partisipasi_murni_2024 0
usia_harapan_hidup_2024    0
harapan_lama_sekolah_2024   0
rata-rata_lama_sekolah_2024 0
pengeluaran_rill_per_kapita_per_tahun_yang_disesuaikan_2024 0
indeks_pembangunan_manusia_2024 0
kuadran_ipm_-_apm          0
dtype: int64

Missing values in CSV:
kabupaten/kota           0
jumlah_sekolah_sd_(negeri) 0
jumlah_sekolah_sd_(swasta) 0
jumlah_sekolah_sd_(negeri+swasta) 0
jumlah_guru_sd_(negeri)    0
jumlah_guru_sd_(swasta)   0
jumlah_guru_sd_(negeri+swasta) 0
jumlah_murid_sd_(negeri)   0
jumlah_murid_sd_(swasta)  0
jumlah_murid_sd_(negeri+swasta) 0
dtype: int64
```

Langkah ini merupakan bagian dari data cleaning, yaitu proses validasi kelengkapan data sebelum dilakukan analisis lebih lanjut. Dengan memastikan tidak ada nilai kosong, pengguna dapat melanjutkan ke tahap analisis statistik, visualisasi, atau pemodelan tanpa risiko error akibat data yang tidak lengkap

# Memeriksa data duplikat

```
print("Duplicate rows in Excel:", df_excel.duplicated().sum())
print("Duplicate rows in CSV:", df_csv.duplicated().sum())
```

Duplicate rows in Excel: 0

Duplicate rows in CSV: 0

yaitu validasi integritas data dengan memastikan tidak ada baris yang terduplicasi. Duplikasi data dapat menyebabkan bias dalam analisis statistik, perhitungan agregat yang tidak akurat, dan hasil model yang tidak valid. Dengan hasil nol duplikasi

# Mengkonversi format angka (menggunakan koma) ke tipe data numerik (float)

```
for col in df_excel.columns:  
    df_excel[col] = pd.to_numeric(df_excel[col].astype(str).str.replace(',', '.'), errors='ignore')  
  
print(df_excel.dtypes)  
  
paten/_kota  
a_partisipasi_kasar_2024  
a_partisipasi_murni_2024  
_harapan_hidup_2024  
pan_lama_sekolah_2024  
-rata_jama_sekolah_2024  
eluaran_rill_per_kapita_per_tahun_yang_disesuaikan_2024  
ks_pembangunan_manusia_2024  
ran_ipm_-_spm  
e: object  
serslatya\AppData\Local\Temp\ipykernel_8188\904512835.py:2: FutureWarning: errors='ignore' is deprecated and will raise in a future version. Use to_num  
_excel[col] = pd.to_numeric(df_excel[col].astype(str).str.replace(',', '.'), errors='ignore')  
serslatya\AppData\Local\Temp\ipykernel_8188\904512835.py:2: FutureWarning: errors='ignore' is deprecated and will raise in a future version. Use to_num  
_excel[col] = pd.to_numeric(df_excel[col].astype(str).str.replace(',', '.'), errors='ignore')  
serslatya\AppData\Local\Temp\ipykernel_8188\904512835.py:2: FutureWarning: errors='ignore' is deprecated and will raise in a future version. Use to_num  
_excel[col] = pd.to_numeric(df_excel[col].astype(str).str.replace(',', '.'), errors='ignore')  
serslatya\AppData\Local\Temp\ipykernel_8188\904512835.py:2: FutureWarning: errors='ignore' is deprecated and will raise in a future version. Use to_num  
_excel[col] = pd.to_numeric(df_excel[col].astype(str).str.replace(',', '.'), errors='ignore')  
serslatya\AppData\Local\Temp\ipykernel_8188\904512835.py:2: FutureWarning: errors='ignore' is deprecated and will raise in a future version. Use to_num  
_excel[col] = pd.to_numeric(df_excel[col].astype(str).str.replace(',', '.'), errors='ignore')  
serslatya\AppData\Local\Temp\ipykernel_8188\904512835.py:2: FutureWarning: errors='ignore' is deprecated and will raise in a future version. Use to_num  
_excel[col] = pd.to_numeric(df_excel[col].astype(str).str.replace(',', '.'), errors='ignore')
```

Python

- Menstandarkan format angka agar sesuai dengan sistem desimal mengganti koma dengan titik.
- Mengubah tipe data menjadi numerik agar bisa digunakan dalam analisis statistik, visualisasi, atau pemodelan.
- Menghindari error saat melakukan operasi matematis atau agregasi data.

# Mengidentifikasi outlier dengan teknik winzorization

```

import pandas as pd

def detect_and_display_outliers(df):
    numeric_cols = df.select_dtypes(include='number').columns
    outlier_reports = []

    for col in numeric_cols:
        Q1 = df[col].quantile(0.25)
        Q3 = df[col].quantile(0.75)
        IQR = Q3 - Q1
        lower = Q1 - 1.5 * IQR
        upper = Q3 + 1.5 * IQR

        outliers = df[(df[col] < lower) | (df[col] > upper)]

        if not outliers.empty:
            for idx, row in outliers.iterrows():
                outlier_reports.append({
                    'column': col,
                    'index': idx,
                    'value': row[col],
                    'lower_bound': lower,
                    'upper_bound': upper,
                    'kabupaten_kota': row.get('kabupaten_kota')
                })

    if outlier_reports:
        return pd.DataFrame(outlier_reports)
    else:
        return "Tidak ada outlier yang terdeteksi di kolom numerik."

outlier_summary = detect_and_display_outliers(df_excel)

```

	column	index	value	lower_bound	upper_bound	kabupaten_kota
0	angka_partisipasi_kasar_2024	25	107.740	96.482500	107.382500	Kab. Bangkalan
1	angka_partisipasi_kasar_2024	26	110.700	96.482500	107.382500	Kab. Sampang
2	angka_partisipasi_kasar_2024	27	108.240	96.482500	107.382500	Kab. Pamekasan
3	angka_partisipasi_kasar_2024	28	108.020	96.482500	107.382500	Kab. Sumenep
4	angka_partisipasi_kasar_2024	29	113.140	96.482500	107.382500	Kota Kediri
5	harapan_lama_sekolah_2024	29	15.710	11.261250	15.671250	Kota Kediri
6	harapan_lama_sekolah_2024	31	15.790	11.261250	15.671250	Kota Malang
7	pengeluaran_rill_per_kapita_per_tahun.yang_dis...	36	19.666	7.216625	17.821625	Kota Surabaya

```

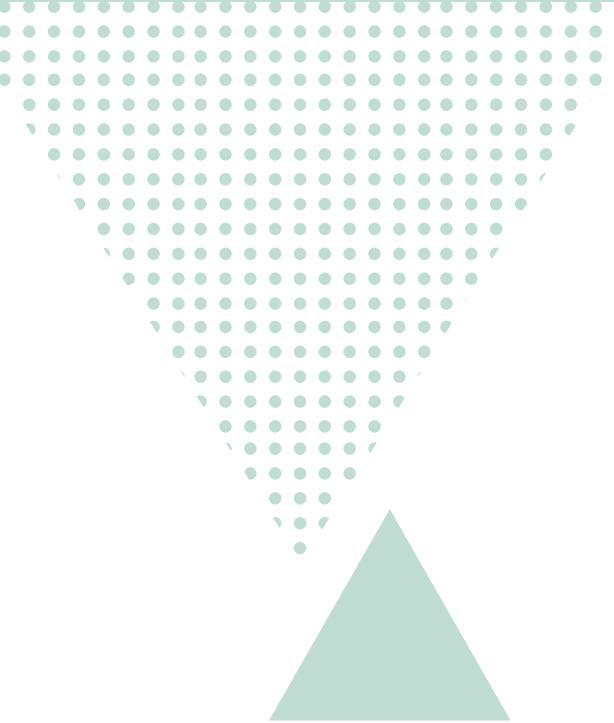
print(df_excel[['angka_partisipasi_kasar_2024', 'angka_partisipasi_kasar_2024_winsorized']].describe())

```

	angka_partisipasi_kasar_2024	angka_partisipasi_kasar_2024_winsorized
count	38.000000	38.000000
mean	102.734211	102.714932
std	3.165067	3.082422
min	99.370000	99.540200
25%	100.570000	100.570000
50%	101.915000	101.915000
75%	103.295000	103.295000
max	113.140000	112.237200

- Mengidentifikasi nilai ekstrem yang dapat mempengaruhi analisis.
- Menentukan apakah perlu dilakukan penanganan seperti penghapusan, imputasi, atau transformasi data.
- Menjaga kualitas dan keandalan hasil analisis statistik atau pemodelan.

# Data Integration

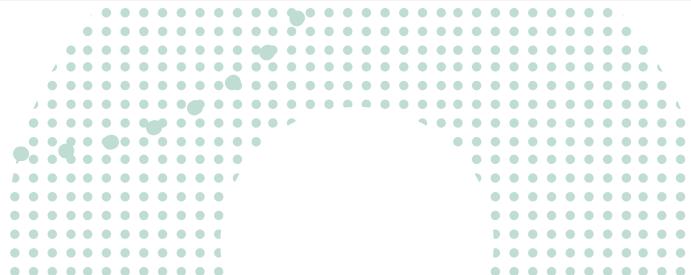


```
def simple_normalize(name):
    name = str(name).strip()
    if name.startswith('Kota'):
        return name.replace('Kota', '').strip() + '_Kota'
    elif name.startswith('Kab.'):
        return name.replace('Kab.', '').strip() + '_Kab'
    else:
        return name + '_Kab' # default ke Kabupaten

# Apply ke kedua dataset
df_ipm_clean['norm_name'] = df_ipm_clean['kabupaten/_kota'].apply(simple_normalize)
df_sd_clean['norm_name'] = df_sd_clean['kabupaten/kota'].apply(simple_normalize)

# Cek hasil
print("Sample normalized names:")
print(df_ipm_clean[['kabupaten/_kota', 'norm_name']].head(40))
print(df_sd_clean[['kabupaten/kota', 'norm_name']].head(40))
```

- Menyatukan format penamaan wilayah agar konsisten.
- Mempermudah proses penggabungan antar dataset berdasarkan wilayah.
- Menghindari kesalahan pencocokan data akibat perbedaan penulisan nama kabupaten/kota.



# Menggabungkan (merge) dua data bersih

```
# Merge
df_final = pd.merge(df_ipm_clean, df_sd_clean, on='norm_name', how='inner')

print(f"Final shape: {df_final.shape}")
print(f"Duplicates: {df_final['norm_name'].duplicated().sum()}")
```

Final shape: (38, 28)  
Duplicates: 0

```
# Hapus kolom redundant
columns_to_drop = [
    'nama_normalized',
    'nama_normalized_v2',
    'norm_name'
]

df_final_clean = df_final.drop(columns=[col for col in columns_to_drop if col in df_final.columns])
print(f"Shape setelah cleanup: {df_final_clean.shape}")
```

Shape setelah cleanup: (38, 27)

- Menghapus kolom redundant
- Menyimpan hasil merged
- Memeriksa ulang hasil validasi akhir

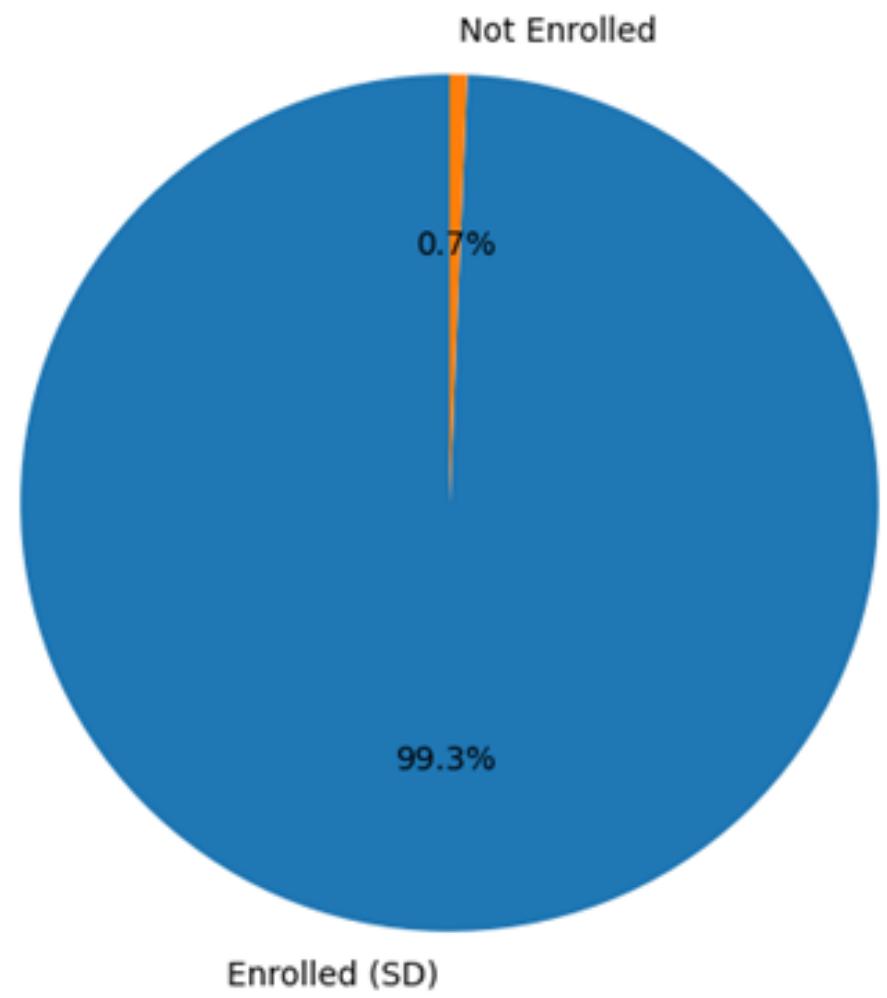
```
# Final validation
print("FINAL VALIDATION")
print(f"Total districts: {len(df_final_clean)}")
print(f"Missing values: {df_final_clean.isnull().sum().sum()}")
print(f"Duplicate districts: {df_final_clean['kabupaten/_kota'].duplicated().sum()}")

print("\nSample final data:")
print(df_final_clean[['kabupaten/_kota', 'indeks_pembangunan_manusia_2024', 'jumlah_sekolah_sd_(negeri+swasta)']].head())
```

FINAL VALIDATION  
Total districts: 38  
Missing values: 0  
Duplicate districts: 0  
  
Sample final data:  
kabupaten/\_kota indeks\_pembangunan\_manusia\_2024 \
0 Kab. Pacitan 71.49
1 Kab. Ponorogo 73.70
2 Kab. Trenggalek 72.47
3 Kab. Tulungagung 75.13
4 Kab. Blitar 73.44  
  
jumlah\_sekolah\_sd\_(negeri+swasta)
0 419
1 571
2 438
3 632
4 666

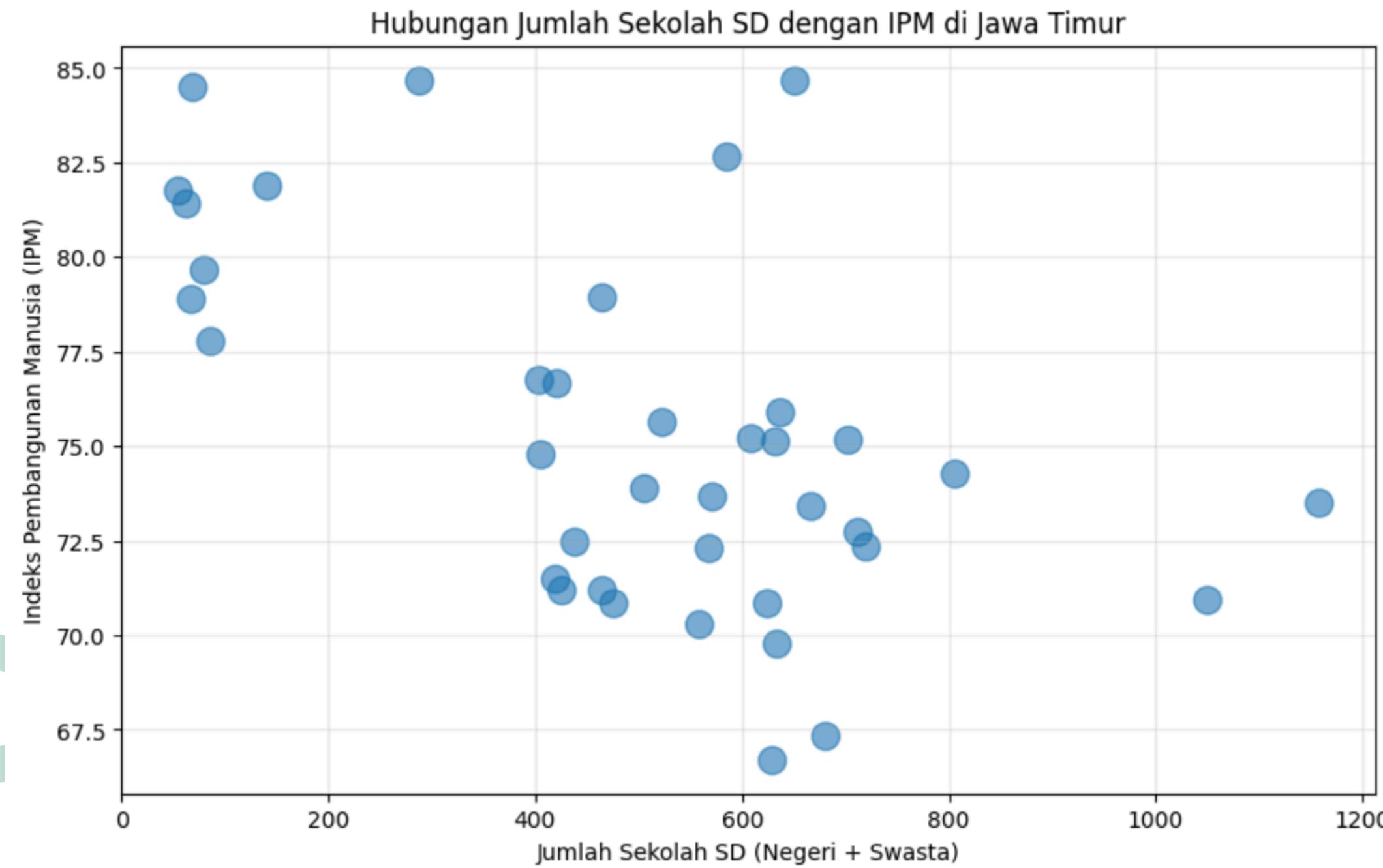
# Hasil Penelitian

Angka Partisipasi Sekolah (APS) Jenjang SD di Jawa Timur (2025)



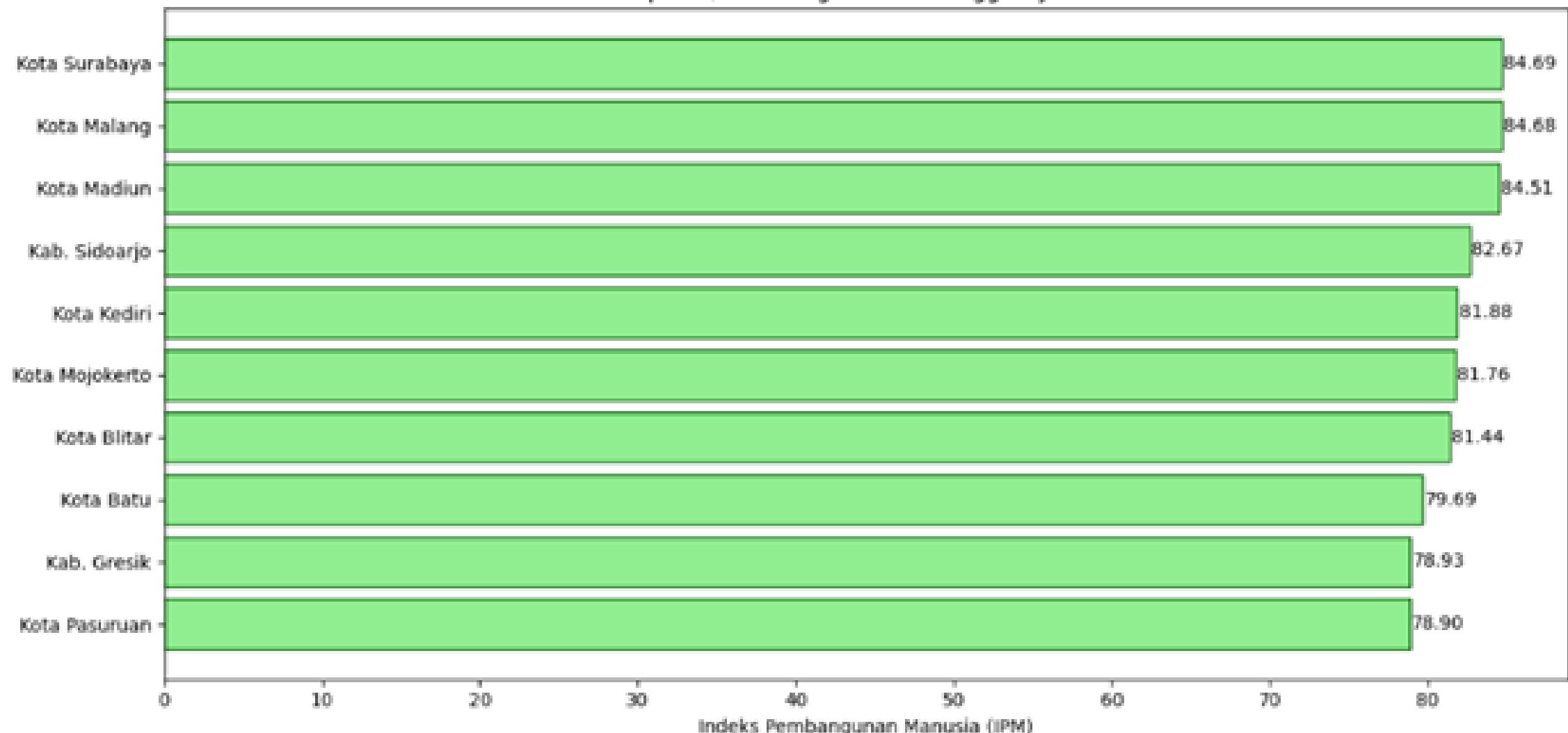
Data pertama (statistik APS Provinsi) menunjukkan Angka Partisipasi Sekolah (APS) SD di Jawa Timur adalah 99.34%

# Hubungan Jumlah sekolah SD dengan IPM



## Top 10 IPM

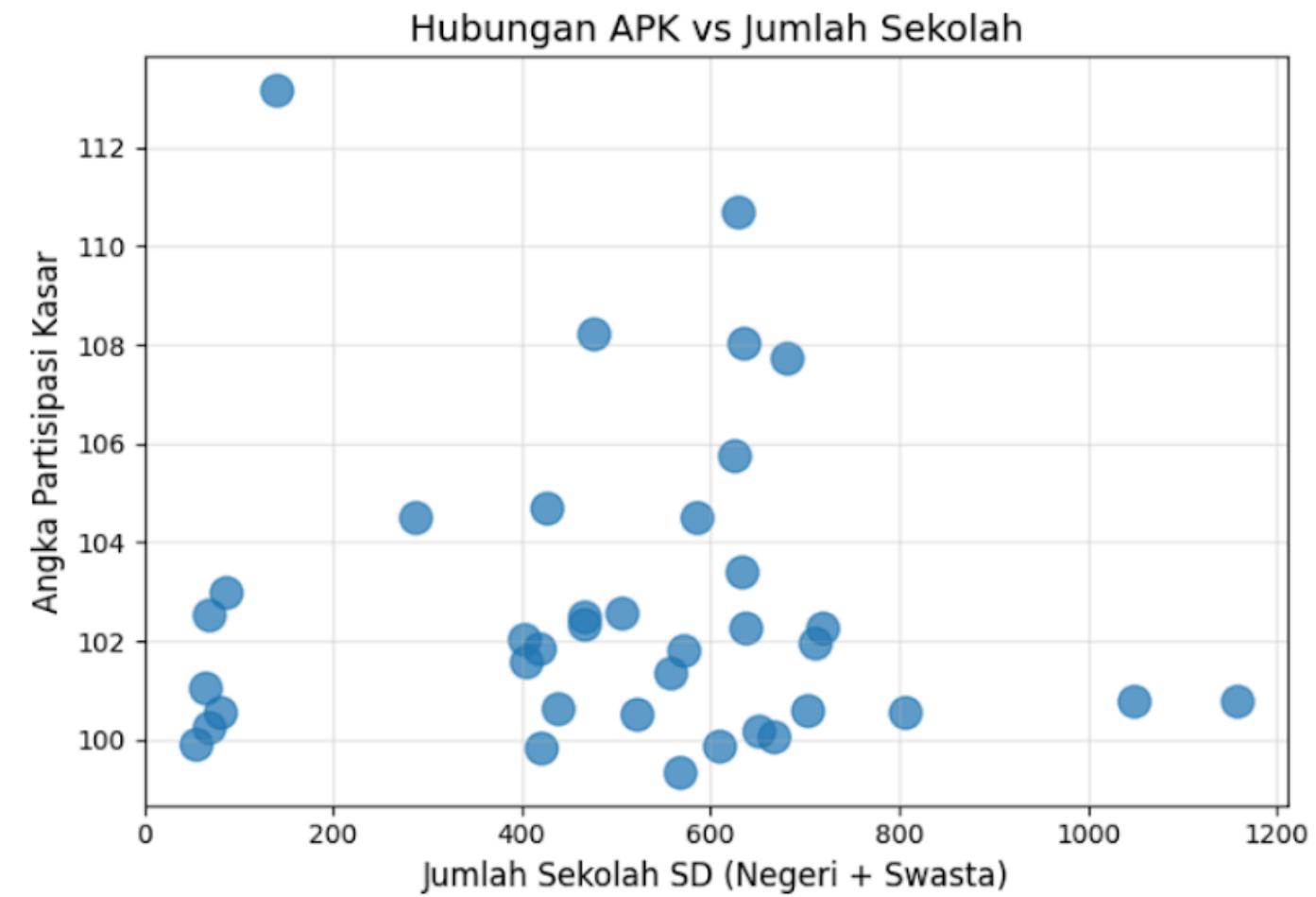
TOP 10 Kabupaten/Kota dengan IPM Tertinggi di Jawa Timur 2024



Grafik batang menunjukkan bahwa 10 besar IPM didominasi oleh wilayah perkotaan (Kota Surabaya, Kota Malang, Kota Madiun, dll.), yang mengindikasikan adanya kesenjangan IPM yang signifikan antara wilayah urban dan rural di Jawa Timur.

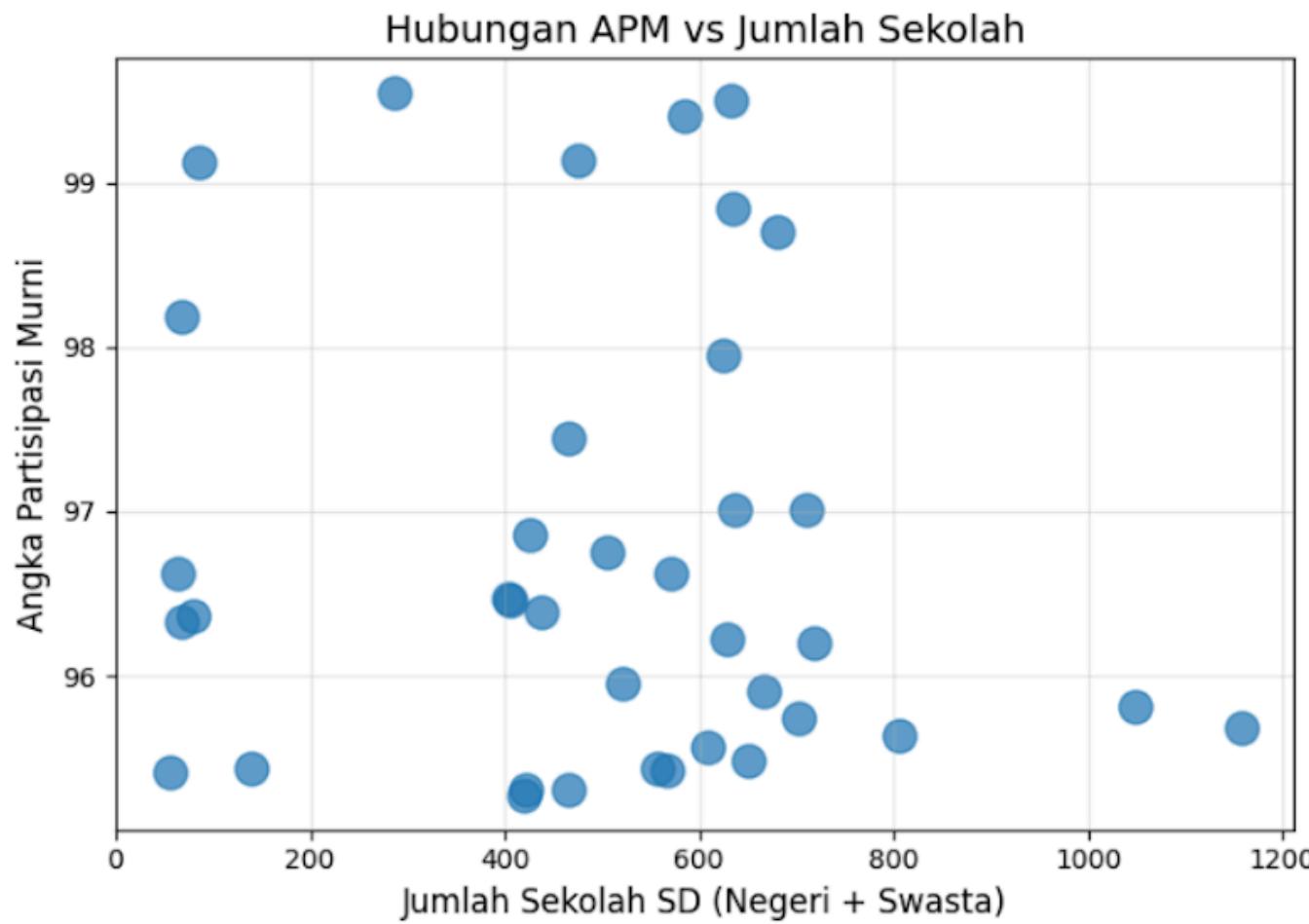


# Hubungan APK dan APM vs Jumlah sekolah



Plot kiri: Hubungan antara Jumlah Sekolah SD dan Angka Partisipasi Kasar (APK)

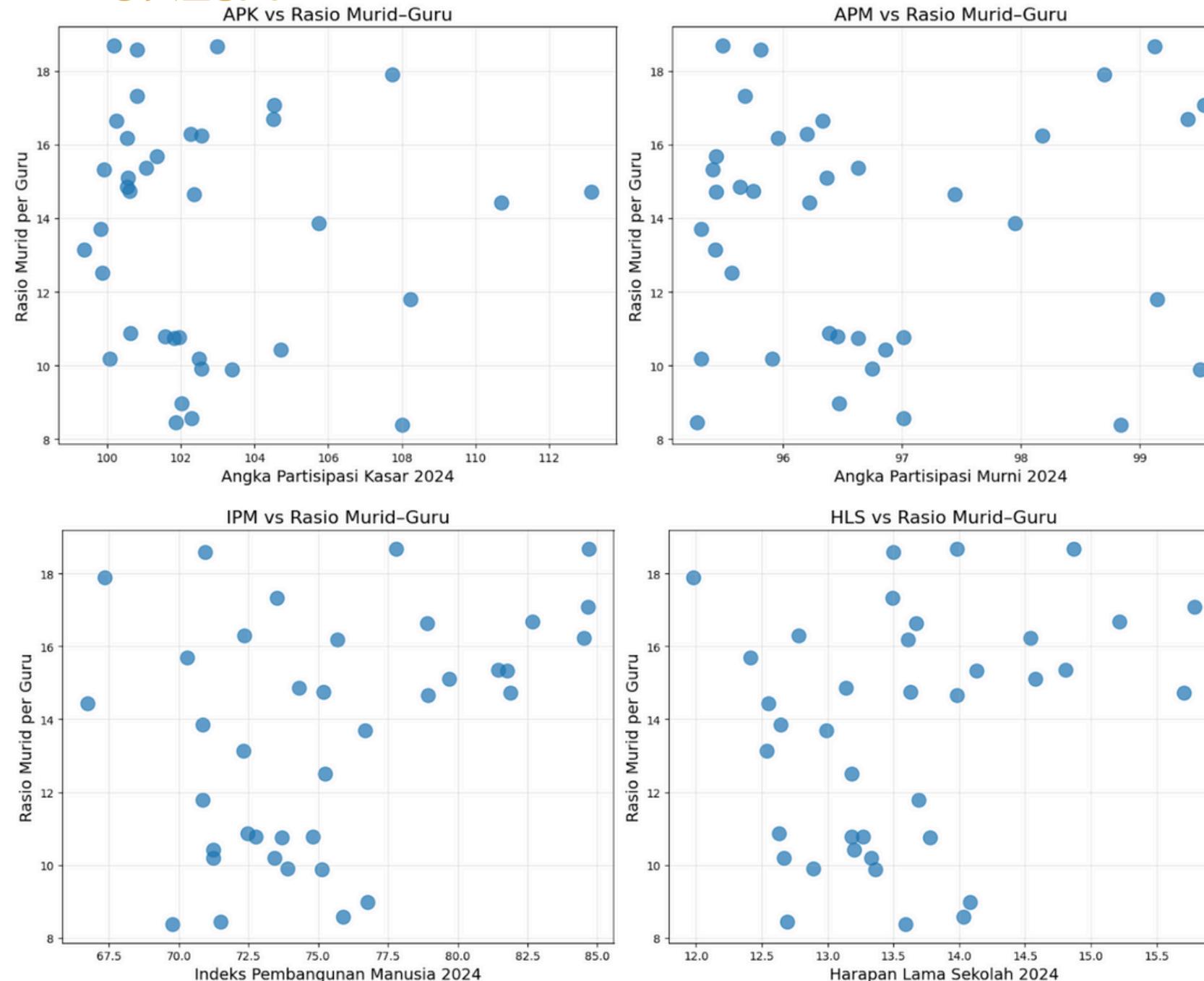
- Titik-titik menunjukkan variasi APK pada setiap kabupaten/kota.
- Polanya menyebar, sehingga hubungan terlihat lemah atau tidak jelas



Plot kanan: Hubungan antara Jumlah Sekolah SD dan Angka Partisipasi Murni (APM)

- Titik-titik juga tersebar acak, menunjukkan hubungan yang kurang kuat.

# Hubungan APK, APM, IPM, HLS vs Rasio Murid-Guru



## 1. APK vs Rasio Murid-Guru

- Menampilkan hubungan antara Angka Partisipasi Kasar (APK) dan rasio murid per guru.
- Titik-titik tersebar acak, menunjukkan hubungan yang lemah.

## 2. APM vs Rasio Murid-Guru

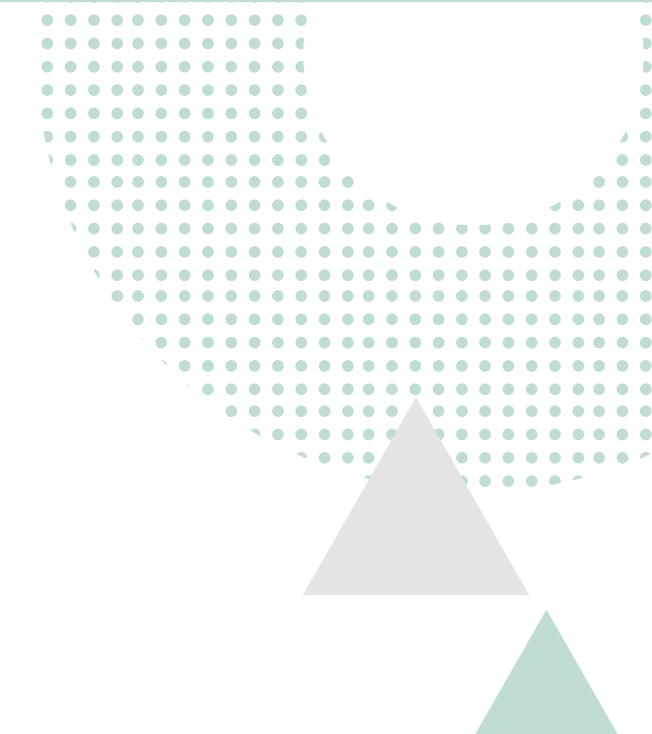
- Menampilkan hubungan Angka Partisipasi Murni (APM) dengan rasio murid per guru.
- Sebaran titik juga acak, tidak menunjukkan pola hubungan kuat.

## 3. IPM vs Rasio Murid-Guru

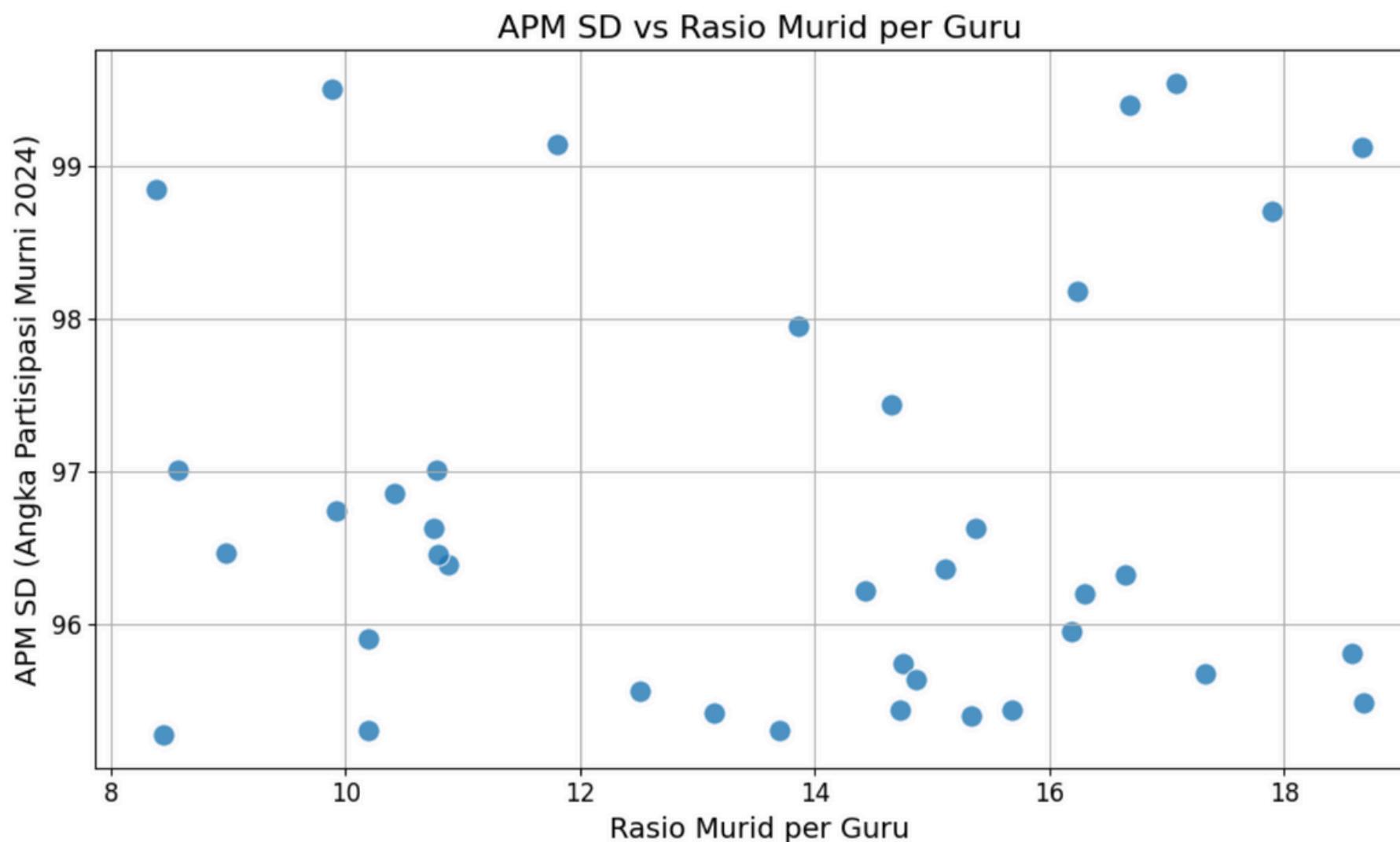
- Menggambarkan hubungan Indeks Pembangunan Manusia (IPM) dan rasio murid per guru.
- Pola sebaran tetap acak, menunjukkan tidak ada korelasi jelas.

## 4. HLS vs Rasio Murid-Guru

- Menghubungkan Harapan Lama Sekolah (HLS) dengan rasio murid per guru.
- Titik tersebar merata tanpa pola tertentu.

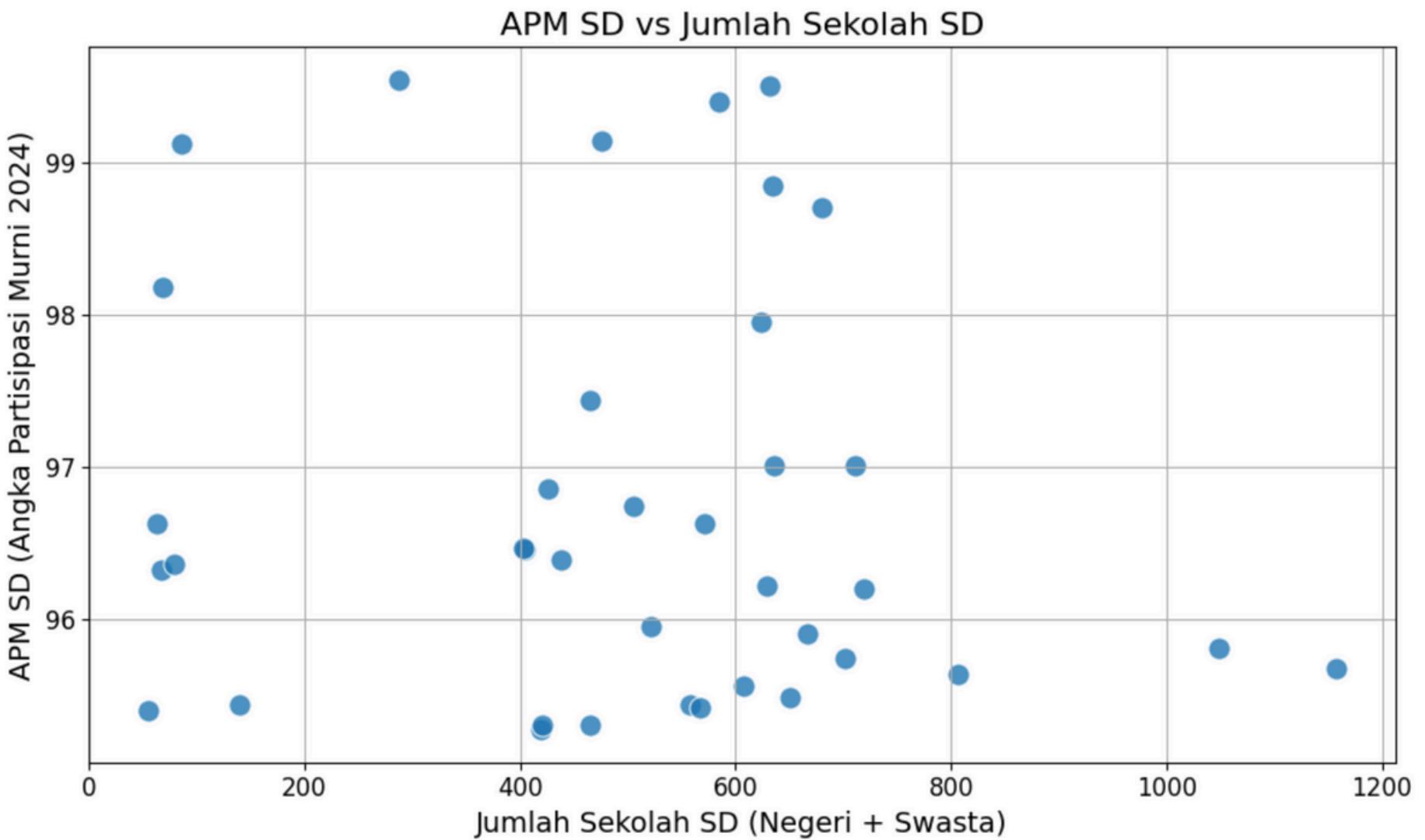


## Hubungan APM vs Rasio Murid per Guru SD



Pola ini mengindikasikan tidak adanya korelasi yang kuat antara rasio murid per guru dengan Angka Partisipasi Murni (APM).

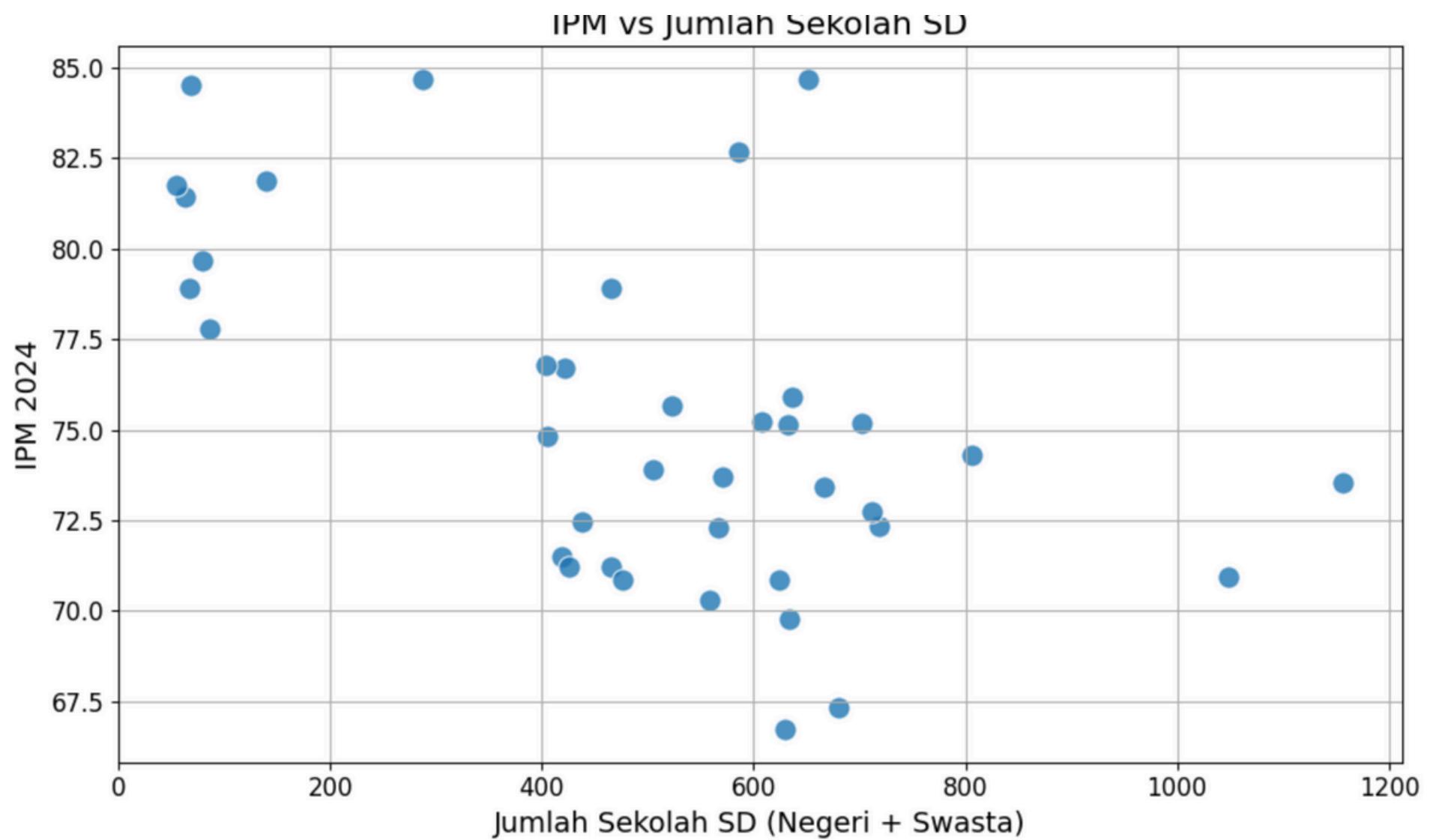
## Hubungan APM SD vs Jumlah Sekolah



Pola sebaran titik sangat acak. Hal ini menunjukkan bahwa tidak ada korelasi yang jelas antara rasio murid per guru dengan APM. Wilayah dengan rasio guru yang ideal, tidak selalu memiliki APM tinggi, begitu pula sebaliknya.

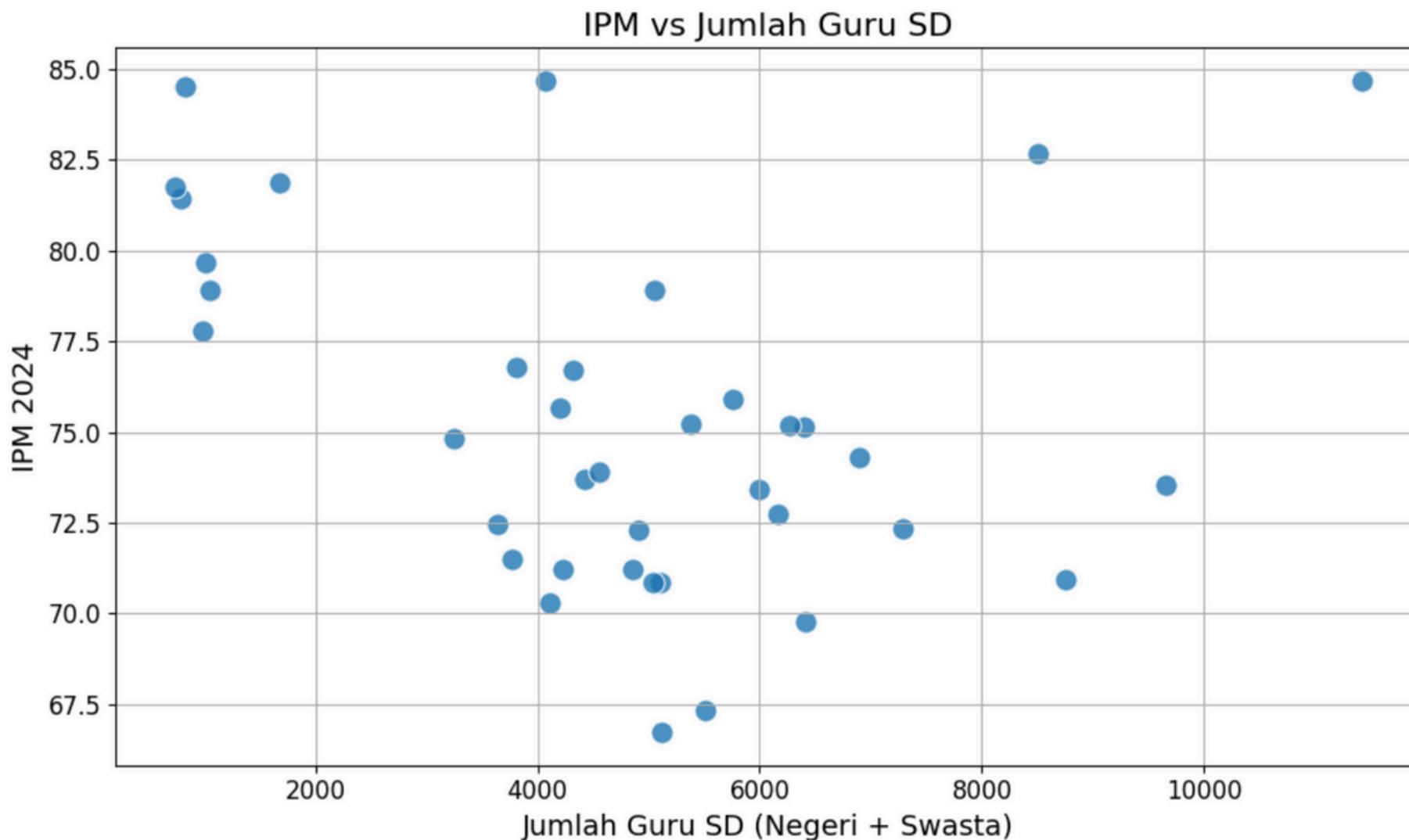


## IPM vs Jumlah Sekolah SD

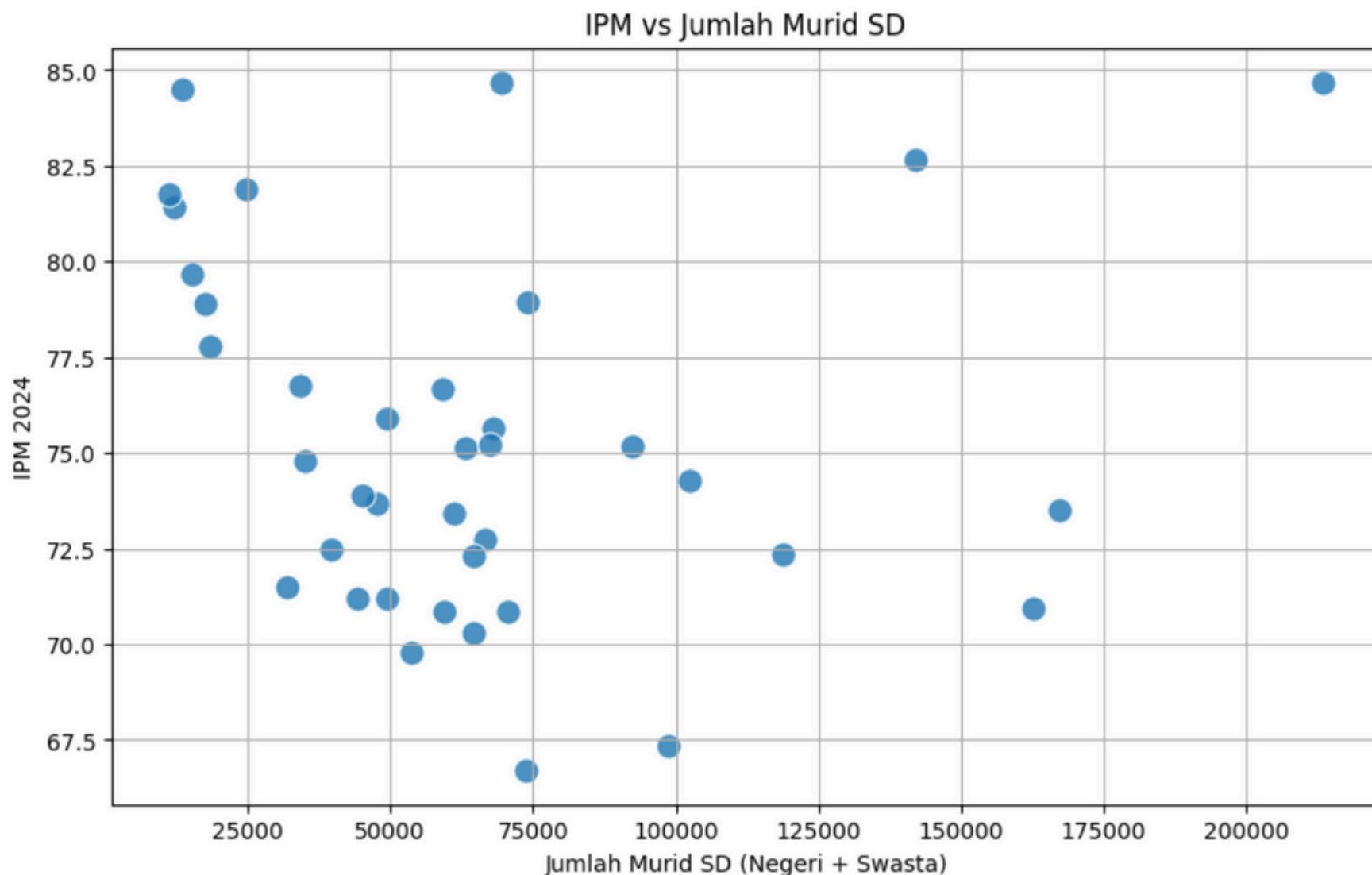


- Grafik menunjukkan bahwa wilayah dengan Jumlah Sekolah SD yang sedikit (di bawah 200) memiliki IPM yang sangat tinggi (sekitar 79 hingga 85, kelompok paling atas).
- Sebaliknya, wilayah dengan jumlah sekolah terbanyak (> 1.000) cenderung memiliki IPM yang lebih rendah (di bawah 75).
- Ini mengindikasikan adanya korelasi negatif yang samar yang memiliki banyak sekolah (yang mungkin mengindikasikan wilayah pedesaan atau luas) dikaitkan dengan IPM yang lebih rendah.

## IPM vs Jumlah Guru SD

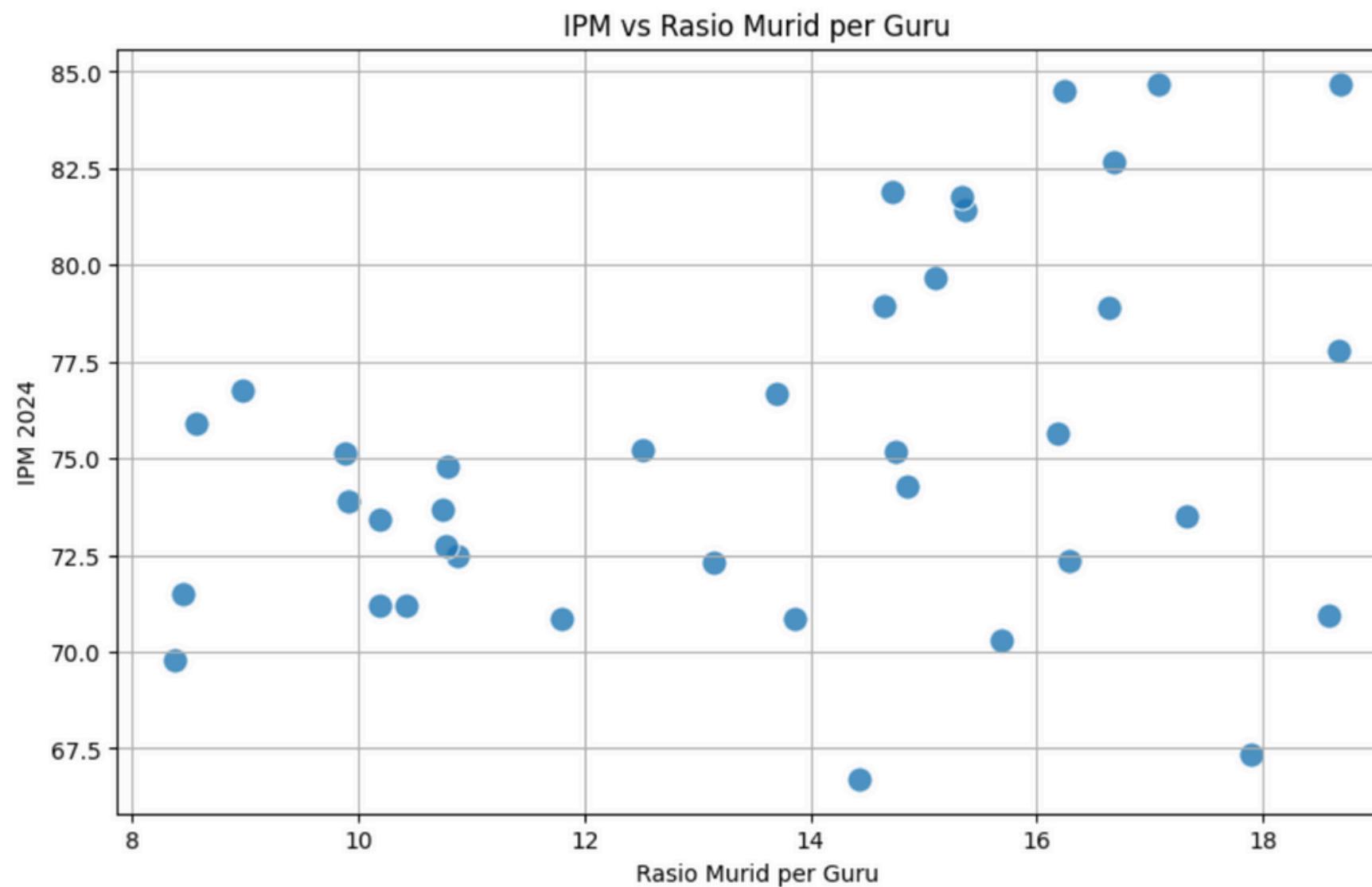


## IPM vs Jumlah Murid SD



Terdapat korelasi negatif yang samar. Wilayah dengan jumlah murid yang sedikit (di bawah 25.000) cenderung memiliki IPM yang sangat tinggi (di atas 81). Sebaliknya, wilayah dengan jumlah murid terbesar (mendekati 210.000) hanya memiliki IPM di sekitar 74. Ada satu wilayah dengan jumlah murid sangat besar yang juga memiliki IPM sangat tinggi.

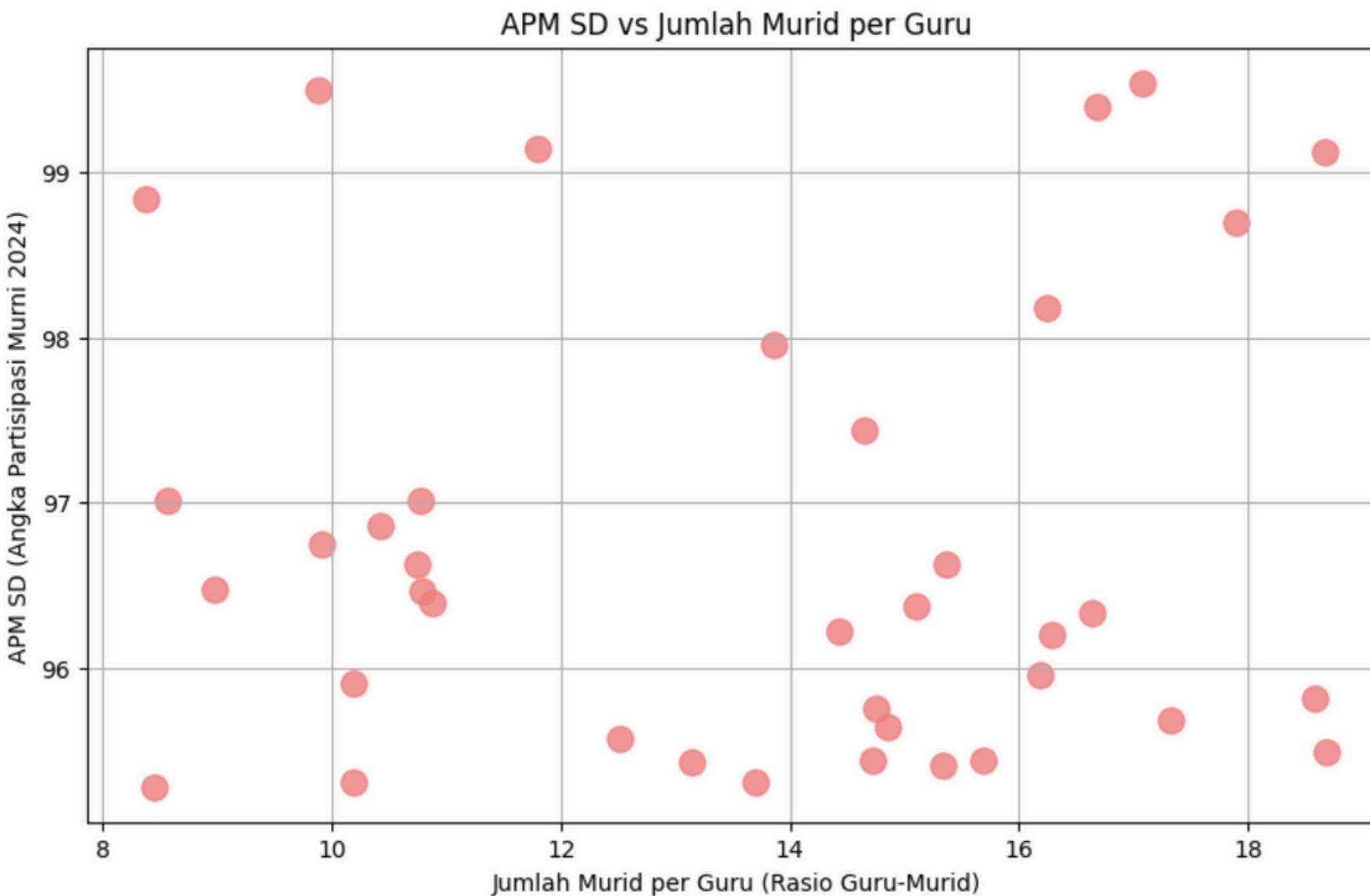
## IPM vs Rasio Murid per Guru



Grafik mengindikasikan bahwa mengurangi rasio murid per guru mungkin berkontribusi pada peningkatan IPM. Ini dapat menjadi dasar kebijakan untuk menambah jumlah guru atau memperbaiki distribusi guru, terutama di daerah dengan IPM rendah dan rasio murid per guru yang tinggi.

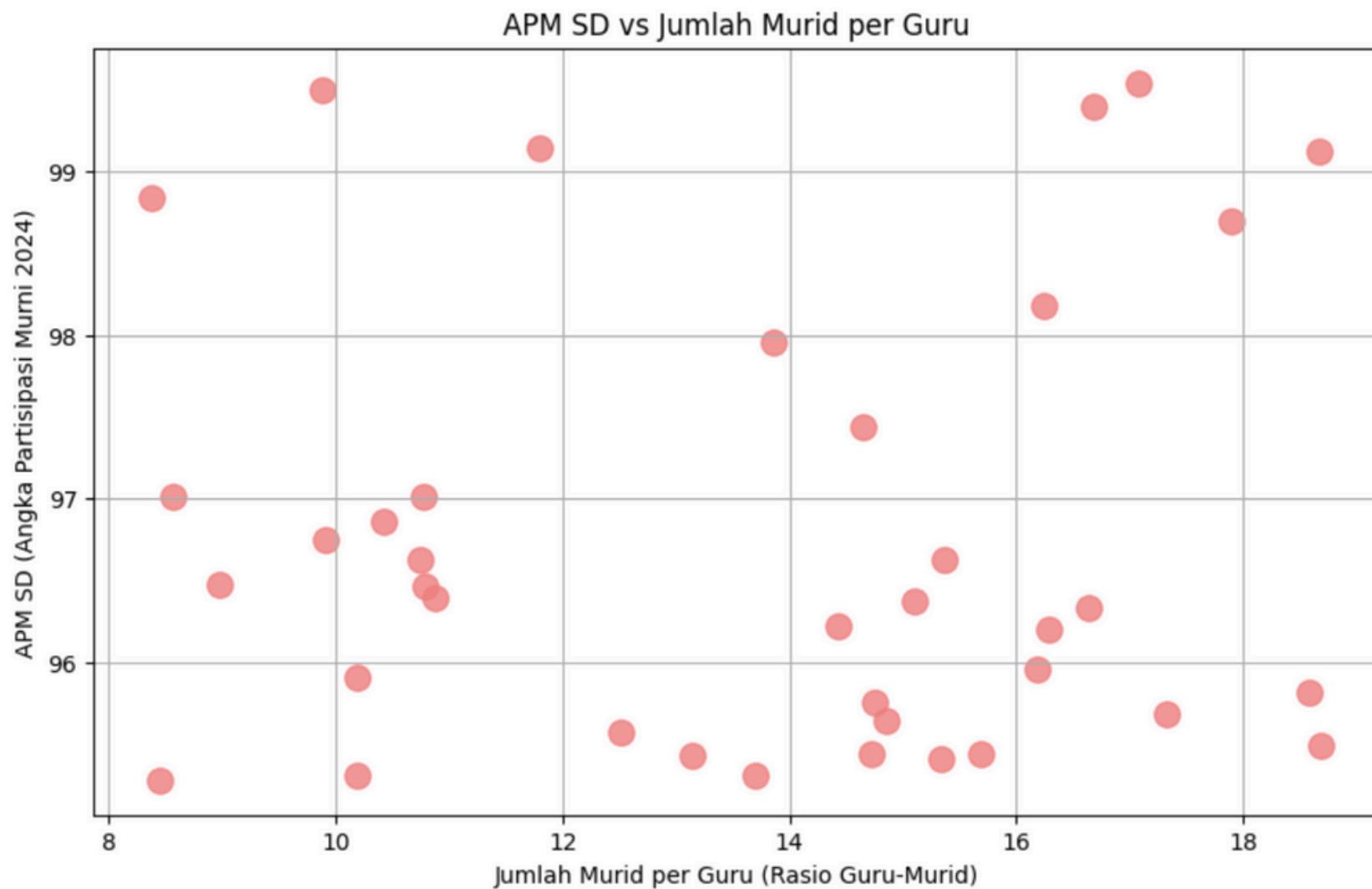


## APM SD vs Jumlah Murid per Guru



Titik data tersebar cukup luas, tidak membentuk pola linier yang jelas. Artinya, rasio murid per guru bukan satu-satunya faktor penentu tingginya Angka Partisipasi Murni (APM) SD.

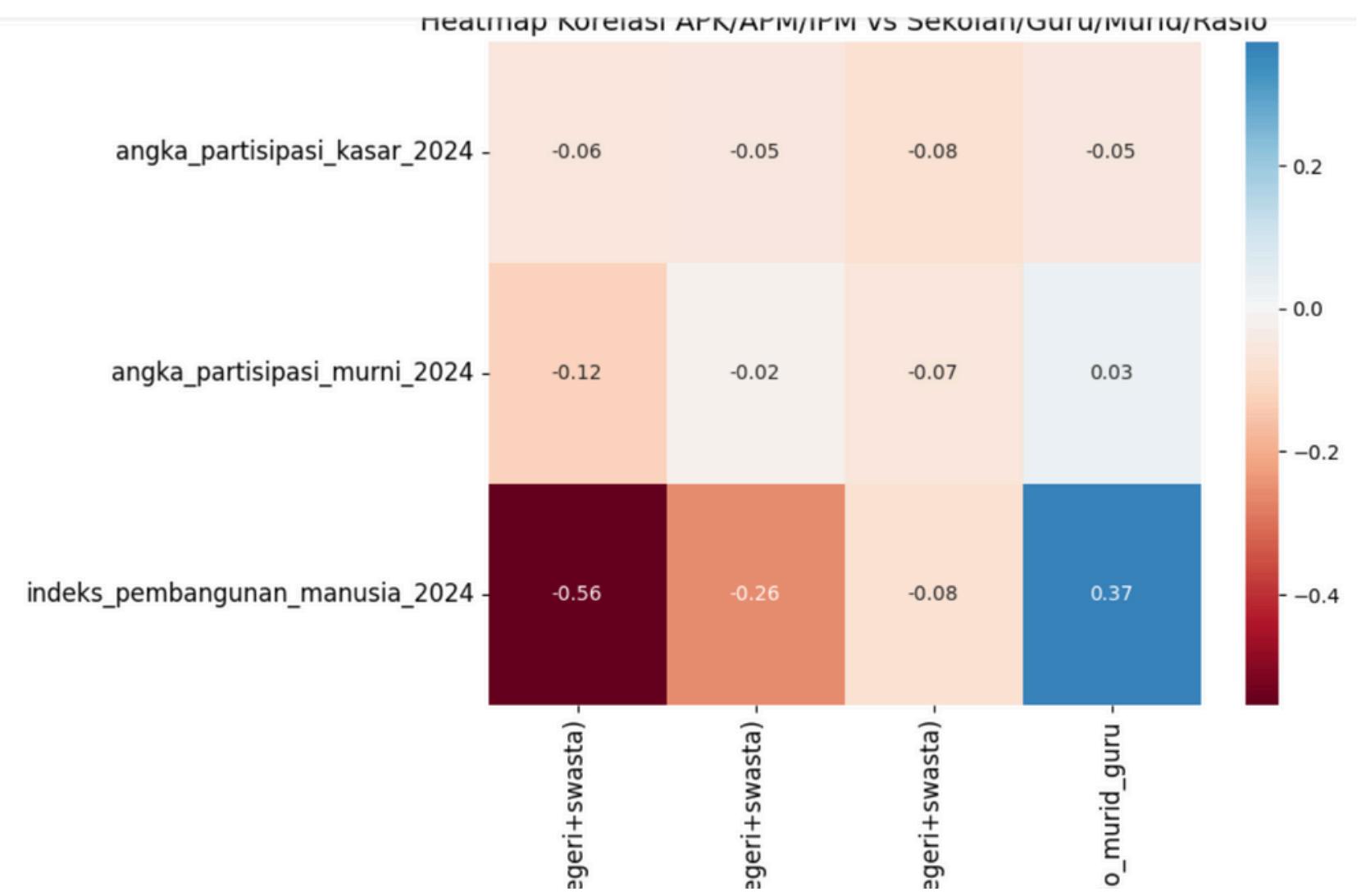
## Perbandingan APM dan IPM vs Rasio Guru-Murid



APM SD cenderung tetap tinggi (>96%) meski rasio murid per guru besar. Hal ini menunjukkan bahwa akses ke sekolah dasar relatif telah tercapai, tetapi IPM yang mencakup kualitas pendidikan, kesehatan, dan daya beli lebih sensitif terhadap rasio guru-murid yang tidak ideal.



## Heatmap Korelasi APK/APM/IPM vs Sekolah/Guru/Murid/Rasio



- IPM memiliki korelasi negatif yang signifikan dengan variabel rasio guru-murid. IPM 2024 memiliki korelasi negatif terkuat dengan jumlah murid per guru (-0.56) dan juga cukup kuat dengan rasio lainnya (-0.37).
- Kebijakan untuk meningkatkan partisipasi sekolah (kehadiran) bisa berjalan tanpa harus menambah guru, tetapi untuk meningkatkan mutu (IPM), rasio guru-murid harus diperbaiki.
- Investasi dalam penambahan guru dan perbaikan rasio murid-guru akan lebih berdampak pada peningkatan IPM daripada sekadar mengejar peningkatan APK/APM

# Interpretasi

**01.**

Pemerataan jumlah sekolah cukup baik, tetapi rasio guru-murid masih bervariasi, terutama di wilayah pedesaan

**02.**

Penambahan sekolah saja tidak otomatis menurunkan rasio guru-murid. Kualitas distribusi guru dan manajemen sekolah juga perlu diperhatikan.

**03.**

Analisis ini dapat menjadi dasar untuk rekomendasi kebijakan pendidikan yang lebih fokus pada penempatan guru dan kualitas pengajaran, bukan hanya jumlah sekolah.



# Kesimpulan

- ▶ Korelasi antar indikator pendidikan cukup kuat, tetapi lemah dengan IPM  
IPM tidak banyak dipengaruhi jumlah sekolah, melainkan faktor ekonomi, kualitas pengajaran, dan lingkungan sosial.
- ▶ Akses pendidikan dasar sangat tinggi dan merata  
APS SD Jawa Timur mencapai 99,34%, dengan mayoritas daerah memiliki APK >100%.
- ▶ Masih terdapat kesenjangan IPM antar wilayah  
Daerah urban memiliki IPM lebih tinggi dibanding wilayah rural dan kepulauan.



- Kebijakan: Fokus pada pemerataan distribusi guru, peningkatan kualitas sekolah di wilayah rural, dan intervensi khusus bagi daerah dengan IPM rendah.
- Penelitian Lanjutan: Perluasan variabel data, penggunaan metode analitik lanjutan, serta integrasi analisis spasial untuk hasil yang lebih komprehensif.
- Implementasi Teknologi: Pengembangan dashboard pemantauan pendidikan dan automasi proses data wrangling untuk mendukung pengambilan kebijakan yang cepat dan tepat.



# Referensi Sumber Penelitian

- Badan Pusat Statistik Jawa Timur. (2024). Jumlah sekolah, guru, dan murid Sekolah Dasar di bawah Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. Surabaya: BPS Jawa Timur.
- Kementerian Sekretariat Negara Republik Indonesia. (2023). Asta Cita: Visi dan misi pembangunan pendidikan nasional. Jakarta: Sekretariat Negara.
- Wickham, H. (2014). Tidy data. *Journal of Statistical Software*, 59(10), 1-23.  
<https://doi.org/10.18637/jss.v059.i10>
- Perserikatan Bangsa-Bangsa. (2022). Sustainable development goals report. United Nations Publications.

# Data Publishing



## Repository GitHub Proyek Data Wrangling SD Jawa Timur

- Nama Repository: data-wrangling-sekolah-dasar-jatim
- Pemilik: ariez-cyber
- Link GitHub: <https://github.com/ariez-cyber/data-wrangling-sekolah-dasar-jatim>

### Deskripsi Repository:

- Analisis data sekolah dasar di Jawa Timur untuk evaluasi rasio guru-murid dan sebaran sekolah menggunakan Python, dengan visualisasi hasil dan data wrangling terstruktur.

### Isi Repository:

- Dataset mentah
- Notebook Python untuk preprocessing, wrangling, dan analisis (DWPDF.ipynb, aps.ipynb, wrangle\_jatim.ipynb, dll.)
- Laporan PDF hasil analisis dan visualisasi
- README.md sebagai panduan penggunaan



# Terima Kasih

