

LAPORAN PROYEK AKHIR

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

Mata Kuliah: Data Wrangling



Disusun oleh:

Atya Azharina Aziz (24031554194)

Muhamad Fadly Saifullah (24031554173)

Kelas: 2024 D

Dosen Pengampu

Dinda Galuh Guminta, M.Stat. (0011129602)

Belgis Ainatul Iza, S.Si., M.Mat. (202509237)

Program Studi S1 Sains Data

Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Negeri Surabaya

2024-2025

BAB 1 PENDAHULUAN

1.1 Latar Belakang

Pendidikan merupakan pondasi utama dalam membangun peradaban bangsa. Namun, di Indonesia, kualitas dan pemerataan akses pendidikan dasar masih menjadi tantangan, khususnya di wilayah-wilayah tertentu seperti Jawa Timur yang memiliki karakter geografis dan sosial ekonomi yang beragam. Ketimpangan jumlah sekolah dasar di tiap kabupaten atau kota dapat berdampak pada kesempatan belajar anak-anak, terutama di daerah pedesaan atau tertinggal.

Upaya pemerataan pendidikan ini sejalan dengan Tujuan Pembangunan Berkelanjutan (SDGs) poin ke-4, yaitu Quality Education, yang menekankan pentingnya pendidikan berkualitas dan merata bagi semua anak. Selain itu, semangat ini sejalan dengan nilai Asta Cita pertama dan keempat, terutama dalam memperjuangkan hak memperoleh pendidikan dan memperkuat kualitas sumber daya manusia.

Pemerintah Indonesia melalui berbagai kebijakan strategis, seperti program Merdeka Belajar, pemerataan guru, dan peningkatan fasilitas pendidikan, menegaskan komitmen dalam menciptakan sistem pendidikan yang adil dan merata. Proyek ini mendukung arah kebijakan tersebut dengan menggunakan pendekatan data wrangling untuk menganalisis sebaran sekolah dasar di Jawa Timur. Dengan analisis data yang bersih, akurat, dan mudah dipahami, diharapkan hasilnya dapat menjadi dasar bagi perumusan strategi pemerataan pendidikan yang lebih efektif.

1.2 Rumusan Masalah

Analisis ini dibuat untuk menjawab pertanyaan-pertanyaan berikut:

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?

1.3. Tujuan Analisis

Berdasarkan rumusan masalah tersebut, tujuan dari analisis ini adalah:

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.

1.4 Manfaat Proyek

Pelaksanaan proyek ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Manfaat Akademik: Mahasiswa dapat memahami bagaimana teknik pengolahan, pembersihan, dan integrasi data digunakan untuk mengidentifikasi pola hubungan antara ketersediaan sekolah dan IPM di Jawa Timur.
2. Manfaat Praktik: Hasil analisis ini dapat menjadi bahan masukan bagi pemerintah daerah maupun instansi pendidikan dalam merumuskan kebijakan pemerataan infrastruktur sekolah dasar di wilayah Jawa Timur.
3. Manfaat Sosial: Masyarakat dapat memperoleh gambaran yang lebih jelas tentang kesenjangan IPM dan ketersediaan infrastruktur pendidikan dasar di provinsi Jawa Timur.

4. Manfaat Teknologis: Proyek ini menunjukkan bagaimana Python dan pustaka sains data dapat digunakan untuk mengolah, membersihkan, dan memvisualisasikan data pendidikan.

BAB 2 TINJAUAN PUSTAKA

2.1 Konsep Pemerataan Pendidikan di Indonesia

Pemerataan pendidikan masih menghadapi tantangan di Jawa Timur, khususnya terkait distribusi guru dan sarana sekolah antara wilayah pedesaan dan perkotaan. Menurut BPS Jawa Timur (2024), ketimpangan fasilitas dan tenaga pendidik masih terlihat di beberapa kabupaten/kota, sehingga pemerintah perlu memahami pola ketimpangan agar kebijakan pendidikan dapat dirumuskan secara lebih akurat.

2.2 Dasar Hukum Pemerataan Pendidikan

Dasar hukum pemerataan pendidikan tercantum dalam Pasal 31 UUD 1945 yang menyatakan bahwa setiap warga negara berhak memperoleh pendidikan dan pemerintah wajib menyelenggarakan pendidikan dasar. Selain itu, Peraturan Pemerintah Nomor 19 Tahun 2005 tentang Standar Nasional Pendidikan dan Peraturan Menteri Pendidikan dan Kebudayaan Nomor 23 Tahun 2013 menguraikan upaya peningkatan partisipasi siswa, pemerataan guru, dan penyediaan fasilitas pendidikan di seluruh Indonesia.

2.3 Pemerataan Pendidikan Berdasarkan SDGs

Menurut Perserikatan Bangsa-Bangsa (PBB), Indonesia berkomitmen terhadap Tujuan Pembangunan Berkelanjutan (SDGs), khususnya Tujuan 4: Pendidikan Berkualitas, yang bertujuan memastikan pendidikan yang inklusif dan berkualitas tinggi. SDGs mendorong setiap negara untuk meningkatkan kualitas pendidikan, memperkuat kompetensi guru, serta memperluas akses pembelajaran sepanjang hayat untuk mendukung pembangunan berkelanjutan.

2.4 Asta Cita Nomor 4

Asta Cita Nomor 4 menekankan peningkatan kualitas pendidikan di seluruh Indonesia melalui penguatan karakter, pemerataan infrastruktur pendidikan, peningkatan kompetensi guru, dan digitalisasi pembelajaran (Kementerian Sekretariat Negara RI, 2023). Dalam konteks Jawa Timur, relevansi Asta Cita terlihat pada upaya pemerintah provinsi meningkatkan kompetensi tenaga pendidik melalui program Guru Penggerak, serta memperluas fasilitas sekolah di wilayah yang masih tertinggal.

2.5 Definisi Data Wrangling

Data wrangling atau data munging adalah proses mengubah, membersihkan, dan menyiapkan data mentah ke dalam format yang siap dianalisis (Kandel et al., 2011). Proses ini mencakup identifikasi kesalahan, pembersihan data, standarisasi format, integrasi data, serta penanganan inkonsistensi agar data dapat digunakan secara optimal.

2.6 Pengolahan Data pada Tahapan Penelitian

Menurut Wickham (2014), komponen utama dari pengolahan data meliputi:

1. Pengumpulan Data, yaitu memperoleh informasi dari berbagai sumber seperti file CSV, web scraping, atau basis data.
2. Pembersihan Data, yaitu menghapus duplikasi, memperbaiki data yang hilang, dan mengatasi inkonsistensi.
3. Integrasi Data, yaitu menggabungkan data dari berbagai sumber agar menjadi satu set data yang utuh.
4. Transformasi Data, yaitu mengubah struktur data, melakukan agregasi, atau mengkonversi format.
5. Visualisasi Data, yaitu menyajikan data dalam bentuk grafik atau diagram untuk membantu pemahaman pola dan tren.

Dalam penelitian ini, pengolahan data dilakukan menggunakan Python dengan memanfaatkan pustaka seperti pandas, numpy, dan matplotlib untuk

mengorganisasikan data dari BPS Jawa Timur dan sumber pendidikan resmi lainnya.

BAB 3 METODOLOGI

3.1 Sumber Data, Alat, dan Bahasa Pemrograman

Proyek ini menggunakan metode Data Wrangling sebagai pendekatan utama dalam mempersiapkan dan menganalisis data. Data yang digunakan dalam proyek ini berasal dari tiga sumber, yaitu:

1. Sumber 1 (PDF)

File dari Pusat Data dan Teknologi Informasi Sekretariat Jenderal Kementerian Pendidikan Dasar dan Menengah yang berisi data Indeks Pembangunan Manusia (IPM) dan indikator sosial-ekonomi lainnya, seperti APK, APM, Harapan Lama Sekolah, dan variabel pendidikan lainnya per kabupaten/kota.

2. Sumber 2 (CSV)

Data BPS berjudul Jumlah Sekolah, Guru, dan Murid Sekolah Dasar (SD) di Bawah Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi, berisi jumlah sekolah, guru, dan murid SD (negeri dan swasta) per kabupaten/kota.

3. Sumber 3 (WEB)

Data Angka Partisipasi Sekolah (APS) Provinsi Jawa Timur (nilai 99,34%) yang diekstraksi secara manual dari publikasi daring resmi.

4. Sumber 4 (CSV)

Data Angka Partisipasi Sekolah (APS) Provinsi Jawa Timur per kabupaten/kota.

Tahapan data wrangling meliputi data collection, data cleaning , data integration , serta data transformation. Visualisasi data kemudian dilakukan menggunakan matplotlib.

Bahasa pemrograman yang digunakan adalah Python dengan bantuan pustaka utama seperti pandas, numpy, dan matplotlib.

3.2 Tahapan Data Wrangling

Tahapan data wrangling pada proyek ini adalah:

1. Data Collection (Pengumpulan Data)

a. Mengambil data dari 4 sumber:

- Sumber 1: Jatim.xlsx data dari hasil scraping file pdf Pusat Data dan Teknologi Informasi Sekretariat Jenderal Kementerian Pendidikan Dasar dan Menengah
- Sumber 2: Jumlah Sekolah, Guru, dan Murid Sekolah Dasar (SD) di Bawah Kementerian Pendidikan, Kebudayaan, Riset.csv. File data BPS yang berisi data kuantitas sekolah, guru, dan murid SD (negeri dan swasta) per kabupaten/kota.
- Sumber 3: Data Angka Partisipasi Sekolah (APS) Provinsi Jawa Timur (99.34%) yang diekstraksi secara manual dari publikasi online.
- Sumber 4 (CSV): Data Angka Partisipasi Sekolah (APS) Provinsi Jawa Timur per kabupaten/kota.

2. Data Cleaning (Pembersihan Data)

a. Memeriksa isi data

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

file_excel = f"{raw_path}\\jatim.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, dan Inovasi 2024.csv"
df_csv = pd.read_csv(file_csv)

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

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

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 Riil per Kapita per Tahun yang Disesuaikan 2024	Indeks Pembangunan Manusia 2024	Kuadrat IPM - APN
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	

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)	Jumlah Murid SD (Negeri+Swasta)
0	Pacitan	408	11	419	3614	161	3775	29126	2777	31903
1	Ponorogo	550	21	571	4031	390	4421	40943	6588	47531
2	Trenggalek	417	21	438	3405	239	3644	35871	3742	39613
3	Tulungagung	569	63	632	5587	810	6397	52132	11076	63208
4	Blitar	634	32	666	5629	371	6000	54552	6581	61133

Data yang ditampilkan berisi indikator pendidikan dan kesejahteraan dari berbagai kabupaten/kota di Indonesia untuk tahun 2024. Beberapa kolom utama dalam dataset meliputi:

- Kabupaten/Kota: Nama wilayah administratif.
- Angka Partisipasi Kasar (APK): Persentase anak usia sekolah yang terdaftar, termasuk yang tidak sesuai usia ideal.
- Angka Partisipasi Murni (APM): Persentase anak usia sekolah yang terdaftar sesuai usia ideal.
- Usia Harapan Hidup: Estimasi rata-rata usia hidup penduduk.
- Harapan Lama Sekolah: Estimasi jumlah tahun pendidikan yang diharapkan ditempuh oleh anak-anak.
- Rata-rata Lama Sekolah: Jumlah tahun pendidikan yang telah ditempuh oleh penduduk usia 15 tahun ke atas.

- Pengeluaran Riil per Kapita per Tahun yang Disesuaikan: Indikator ekonomi yang mencerminkan daya beli dan kesejahteraan masyarakat.

b. 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. Hal ini diterapkan pada kedua dataset, yaitu df_excel (data dari Excel) dan df_csv (data dari CSV). Proses ini mempermudah pemanggilan kolom dan menghindari potensi error saat manipulasi data selanjutnya. Nama-nama kolom yang sudah diubah ditampilkan menggunakan fungsi print() untuk memastikan konsistensi.

c. 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

d. Memeriksa data duplikat (Ditemukan 0)

```
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 terduplikasi. Duplikasi data dapat menyebabkan bias dalam analisis statistik, perhitungan agregat yang tidak akurat, dan hasil model yang tidak valid. Dengan hasil nol duplikasi

- e. 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)
```

Python

paten/_kota	object
a_partisipasi_kasar_2024	float64
a_partisipasi_murni_2024	float64
_harapan_hidup_2024	float64
pan_lama_sekolah_2024	float64
-rata_lama_sekolah_2024	float64
eluaran_rill_per_kapita_per_tahun_yang_disesuaikan_2024	float64
ks_pembangunan_manusia_2024	float64
ran_ipm_-_apm	object

e: object

[sers\atv\aa\AppData\Local\Temp\ipykernel_8108\904512835.py:2](#): FutureWarning: errors='ignore' is deprecated and will raise in a future version. Use to_numeric(df_excel[col].astype(str).str.replace(',', '.'), errors='ignore')

[sers\atv\aa\AppData\Local\Temp\ipykernel_8108\904512835.py:2](#): FutureWarning: errors='ignore' is deprecated and will raise in a future version. Use to_numeric(df_excel[col].astype(str).str.replace(',', '.'), errors='ignore')

[sers\atv\aa\AppData\Local\Temp\ipykernel_8108\904512835.py:2](#): FutureWarning: errors='ignore' is deprecated and will raise in a future version. Use to_numeric(df_excel[col].astype(str).str.replace(',', '.'), errors='ignore')

[sers\atv\aa\AppData\Local\Temp\ipykernel_8108\904512835.py:2](#): FutureWarning: errors='ignore' is deprecated and will raise in a future version. Use to_numeric(df_excel[col].astype(str).str.replace(',', '.'), errors='ignore')

[sers\atv\aa\AppData\Local\Temp\ipykernel_8108\904512835.py:2](#): FutureWarning: errors='ignore' is deprecated and will raise in a future version. Use to_numeric(df_excel[col].astype(str).str.replace(',', '.'), errors='ignore')

[sers\atv\aa\AppData\Local\Temp\ipykernel_8108\904512835.py:2](#): FutureWarning: errors='ignore' is deprecated and will raise in a future version. Use to_numeric(df_excel[col].astype(str).str.replace(',', '.'), errors='ignore')

[sers\atv\aa\AppData\Local\Temp\ipykernel_8108\904512835.py:2](#): FutureWarning: errors='ignore' is deprecated and will raise in a future version. Use to_numeric(df_excel[col].astype(str).str.replace(',', '.'), errors='ignore')

[sers\atv\aa\AppData\Local\Temp\ipykernel_8108\904512835.py:2](#): FutureWarning: errors='ignore' is deprecated and will raise in a future version. Use to_numeric(df_excel[col].astype(str).str.replace(',', '.'), errors='ignore')

[sers\atv\aa\AppData\Local\Temp\ipykernel_8108\904512835.py:2](#): FutureWarning: errors='ignore' is deprecated and will raise in a future version. Use to_numeric(df_excel[col].astype(str).str.replace(',', '.'), errors='ignore')

[sers\atv\aa\AppData\Local\Temp\ipykernel_8108\904512835.py:2](#): FutureWarning: errors='ignore' is deprecated and will raise in a future version. Use to_numeric(df_excel[col].astype(str).str.replace(',', '.'), errors='ignore')

- 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.

- f. Mengidentifikasi dan menangani *outlier* menggunakan metode IQR dan teknik *Winsorization*

```
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)
```

outlier_summary

	column	index	value	lower_bound	upper_bound	kabupaten_kota
0	angka_partisipasi_kasar_2024	25	107.740	96.482500	107.382500	None
1	angka_partisipasi_kasar_2024	26	110.700	96.482500	107.382500	None
2	angka_partisipasi_kasar_2024	27	108.240	96.482500	107.382500	None
3	angka_partisipasi_kasar_2024	28	108.020	96.482500	107.382500	None
4	angka_partisipasi_kasar_2024	29	113.140	96.482500	107.382500	None
5	harapan_lama_sekolah_2024	29	15.710	11.261250	15.671250	None
6	harapan_lama_sekolah_2024	31	15.790	11.261250	15.671250	None
7	pengeluaran_rill_per_kapita_per_tahun_yang_dis...	36	19.666	7.216625	17.821625	None

```
def winsorize_series(series, lower_percentile=0.01, upper_percentile=0.99):
    lower = series.quantile(lower_percentile)
    upper = series.quantile(upper_percentile)
    return series.clip(lower, upper)

df_excel['angka_partisipasi_kasar_2024_winsorized'] = winsorize_series(df_excel['angka_partisipasi_kasar_2024'])
```

```
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.

3.3 Data Integration (Integrasi Data)

Tahapan data wrangling pada proyek ini adalah:

1. Membuat kolom kunci `norm_name` untuk standarisasi nama wilayah

```
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))
```

- a. Menyatukan format penamaan wilayah agar konsisten.
 - b. Mempermudah proses penggabungan antar dataset berdasarkan wilayah.
 - c. Menghindari kesalahan pencocokan data akibat perbedaan penulisan nama kabupaten/kota.
2. Menggabungkan (merge) dua data bersih (Data 1 dan Data 2) menjadi satu data utama

```
# 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)

```
# Save ke file processed
output_path = r"C:\Users\atyaa\OneDrive\Pictures\MAIN\SEM 3\UAS.DW\project_wrangle\processed_data"
df_final_clean.to_csv(f"{output_path}\\jatim_merged_final.csv", index=False)
print("Data berhasil disimpan!")
```

Data berhasil disimpan!

```
# 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
```

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

3.4 Data Analysis & Exploration (Analisis dan Eksplorasi)

Langkah eksplorasi dilakukan untuk memahami karakteristik data. Prosesnya meliputi:

- Distribusi dan Pemeriksaan Variabel
 - Pemeriksaan distribusi IPM (histogram, boxplot).
 - Pemeriksaan nilai partisipasi pendidikan (APK, APM, APS).
 - Pemeriksaan distribusi jumlah sekolah, guru, dan murid.
 - Pemeriksaan rasio murid per guru.
- Analisis Perbandingan dan Hubungan Antar Variabel
 - Analisis *Top 10* IPM kabupaten/kota.

- Scatter plot hubungan:
 - Jumlah sekolah vs IPM
 - Jumlah guru vs IPM
 - Jumlah murid vs IPM
 - Rasio murid-guru vs IPM
 - APK vs rasio murid-guru
 - APM vs rasio murid-guru
 - HLS vs rasio murid-guru
- c. Analisis APK wilayah vs APS Provinsi.
- d. Heatmap korelasi antara:
- APK, APM, IPM
 - Jumlah sekolah, guru, murid
 - Rasio murid-guru
- e. Analisis Kuantitatif Lain
- Perbandingan rasio guru-murid antara wilayah urban, suburban, dan rural.
 - Perbandingan rasio guru-murid metro vs non-metro.
 - Identifikasi wilayah dengan rasio guru-murid tertinggi dan terendah.

3.5 Data Visualization & Publishing (Visualisasi dan Publikasi)

Tahap visualisasi dilakukan untuk menyajikan hasil analisis secara informatif dan mudah dipahami. Kegiatan yang dilakukan meliputi:

1. Membuat berbagai jenis visualisasi

Termasuk:

- a. Scatter plot untuk melihat hubungan dua variabel.
- b. Histogram untuk melihat distribusi data,
- c. Boxplot untuk melihat persebaran dan outlier,
- d. Heatmap korelasi untuk memetakan kekuatan hubungan antar variabel.

2. Memperbaiki estetika visualisasi
Seperti menambahkan judul, label, ukuran figur, serta pemilihan warna yang konsisten dengan standar visual laporan.
3. Mempublikasikan data hasil wrangling
Menyimpan dataset gabungan yang telah dibersihkan dalam file `jatim_merged_final.csv`, yang kemudian digunakan sebagai dasar analisis visual, korelasi, dan pembahasan pada BAB selanjutnya.
4. Menyiapkan visualisasi untuk laporan akhir
Semua grafik yang relevan dimasukkan ke dalam laporan untuk mendukung pembahasan pada BAB Hasil dan Pembahasan.

BAB 4 HASIL DAN PEMBAHASAN

Bab ini membahas temuan yang diperoleh dari proses data wrangling dan analisis eksplorasi data (EDA) yang dilakukan dalam notebook.

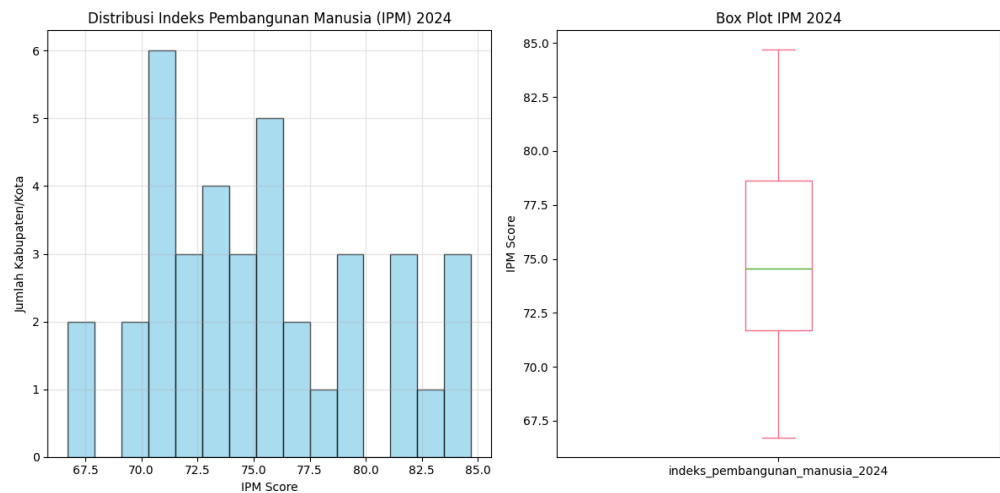
4.1 Hasil Pembersihan dan Integrasi Data

Proses data wrangling dimulai dengan pembersihan dan persiapan dua dataset utama

1. Hasil Pembersihan dan Integrasi Data:
 - a. Data berhasil digabung menjadi dataset final berisi 38 kabupaten/kota.
 - b. Tipe data numerik berhasil dikonversi.
 - c. Outlier pada jumlah murid dan rasio murid-guru teridentifikasi di beberapa wilayah.
 - d. Dataset final siap untuk dianalisis.

4.2 Analisis Data dan Visualisasi

Analisis eksplorasi data (EDA) dari data yang telah dibersihkan memberikan wawasan berikut:



Langkah ini bertujuan untuk menyajikan hasil analisis data secara visual agar lebih mudah dipahami dan diinterpretasikan. Beberapa jenis visualisasi yang digunakan meliputi:

1. Histogram

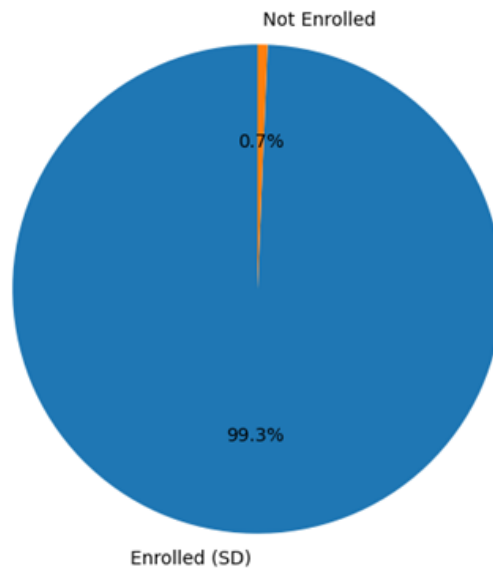
Digunakan untuk menampilkan distribusi nilai dari variabel kuantitatif seperti Indeks Pembangunan Manusia (IPM), pengeluaran per kapita, atau rata-rata lama sekolah. Histogram membantu mengidentifikasi pola sebaran data, seperti apakah data bersifat normal, miring (skewed), atau memiliki outlier.

2. Boxplot

Digunakan untuk mendeteksi nilai pencilan dan memahami sebaran data berdasarkan kuartil. Visualisasi ini sangat berguna dalam membandingkan indikator antar wilayah atau kelompok.

3. Tingkat Partisipasi Pendidikan

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



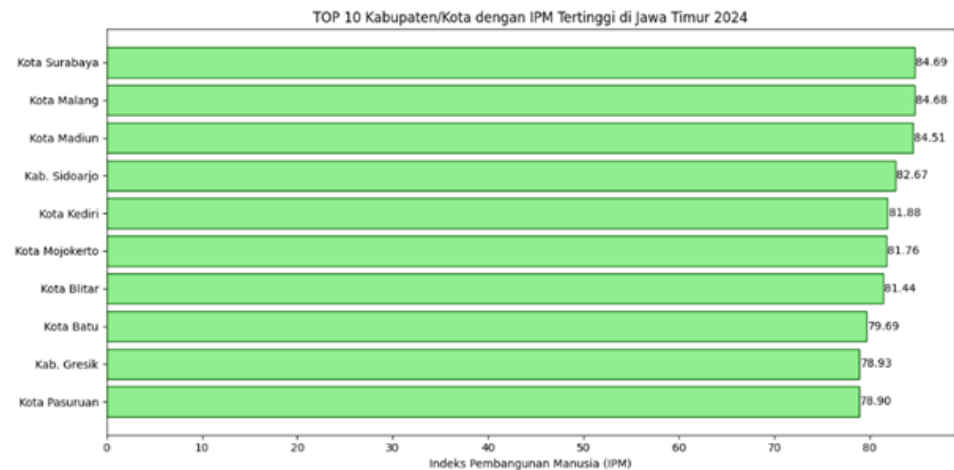
Data ketiga (statistik APS Provinsi) menunjukkan Angka Partisipasi Sekolah (APS) SD di Jawa Timur adalah 99.34%. Angka ini divisualisasikan dalam Pie Chart menunjukkan bahwa partisipasi dan akses pendidikan dasar di Jatim secara umum sudah sangat baik dan hampir sempurna (merata).

4. Kesenjangan IPM

Meskipun partisipasi sekolah merata, kondisi pembangunan manusia (IPM) menunjukkan kesenjangan.

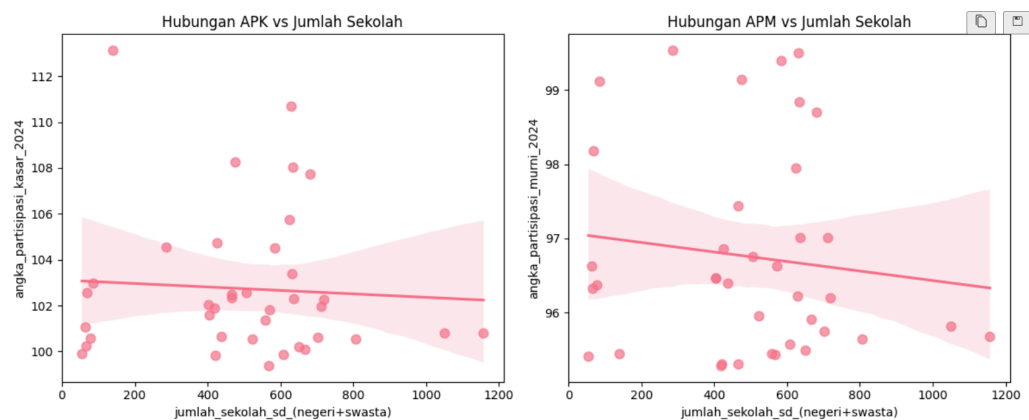
- a. Distribusi IPM: Histogram dan Boxplot menunjukkan bahwa nilai IPM di Jatim bervariasi, dengan sebaran data yang cukup luas dan beberapa nilai tinggi yang menonjol.

b. Top 10 IPM:



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.

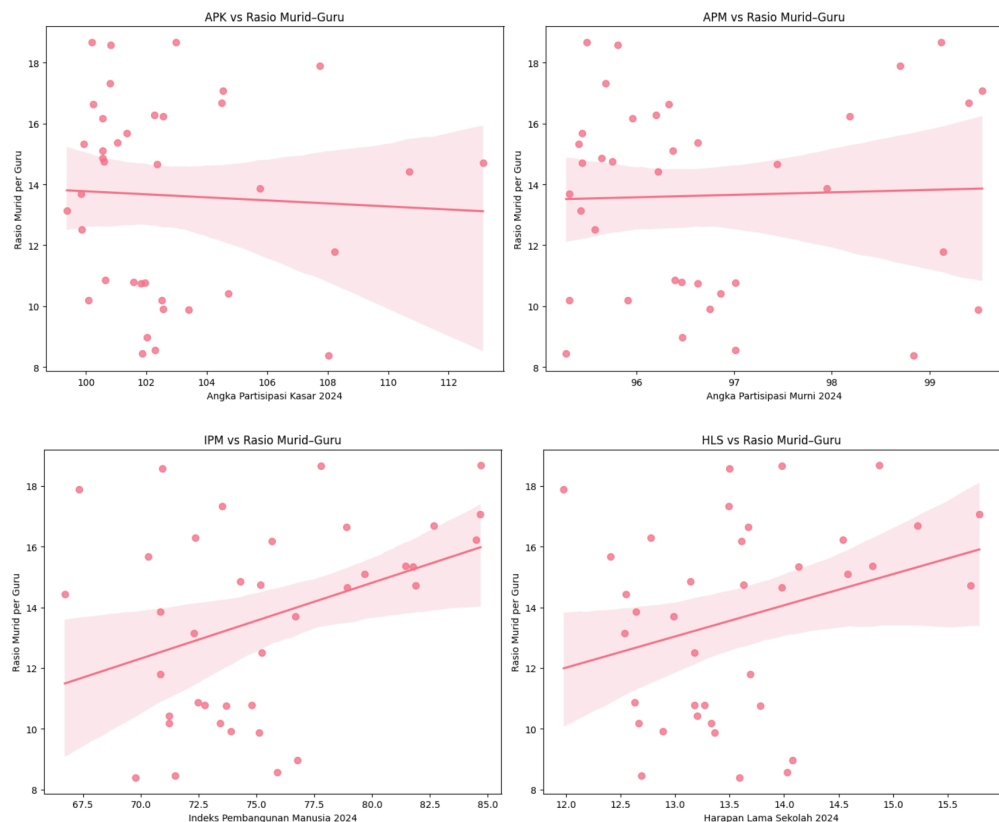
5. Hubungan APK dan APM vs Jumlah sekolah



Plot pertama menunjukkan sebaran data dengan tren menurun, yang mengindikasikan bahwa peningkatan jumlah sekolah tidak selalu berbanding lurus dengan peningkatan APK. Hal ini dapat disebabkan oleh faktor lain seperti distribusi penduduk, akses geografis, atau kualitas fasilitas pendidikan

Plot kedua juga menunjukkan tren menurun, yang mengindikasikan bahwa meskipun jumlah sekolah meningkat, tidak semua anak usia sekolah terdaftar sesuai usia ideal. Hal ini dapat mencerminkan tantangan dalam pemerataan pendidikan atau ketidaksesuaian antara kapasitas sekolah dan kebutuhan lokal

6. Hubungan APK, APM, IPM, HLS vs Rasio Murid-Guru



a. APK vs Rasio Murid-Guru

Plot menunjukkan tren negatif ringan, yang mengindikasikan bahwa wilayah dengan APK tinggi cenderung memiliki rasio murid-guru yang lebih rendah. Hal ini dapat mencerminkan efisiensi distribusi tenaga pengajar di daerah dengan partisipasi pendidikan yang lebih luas.

b. APM vs Rasio Murid-Guru

Plot menunjukkan tren yang hampir datar, menandakan tidak adanya korelasi yang signifikan antara APM dan rasio murid-guru. Ini menunjukkan bahwa jumlah guru relatif tidak berubah meskipun partisipasi murni meningkat atau menurun.

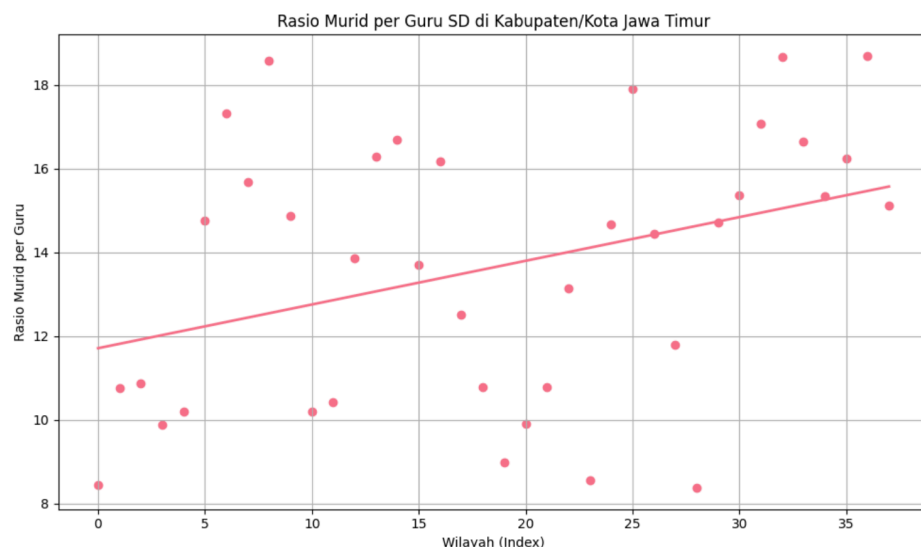
c. IPM vs Rasio Murid-Guru

Plot menunjukkan tren positif, yang mengindikasikan bahwa wilayah dengan IPM lebih tinggi cenderung memiliki rasio murid-guru yang lebih besar. Hal ini dapat menunjukkan bahwa daerah dengan pembangunan manusia yang lebih baik memiliki kapasitas pendidikan yang lebih besar, meskipun rasio ini juga bisa mencerminkan tantangan dalam pemerataan tenaga pengajar.

d. HLS vs Rasio Murid-Guru

Plot menunjukkan tren positif ringan, yang mengindikasikan bahwa wilayah dengan harapan lama sekolah yang lebih tinggi cenderung memiliki rasio murid-guru yang lebih besar. Ini bisa mencerminkan peningkatan permintaan terhadap pendidikan formal di daerah dengan aspirasi pendidikan yang tinggi.

7. Hubungan Rasio Murid per Guru SD di kabupaten jawa timur

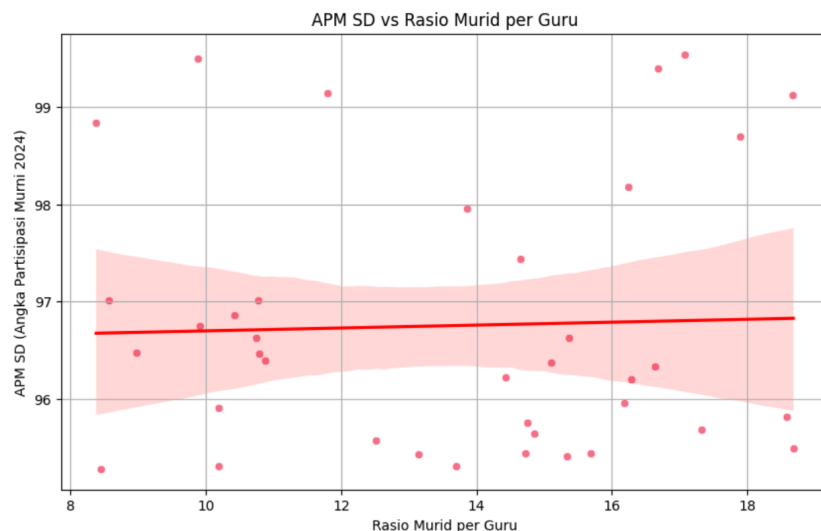


Grafik menunjukkan adanya variasi rasio murid-guru antar daerah, yang dapat mencerminkan ketimpangan distribusi tenaga pengajar di tingkat SD.

Garis tren yang cenderung naik mengindikasikan bahwa beberapa wilayah memiliki rasio murid-guru yang lebih tinggi, yang berpotensi berdampak pada kualitas pembelajaran akibat beban kerja guru yang lebih besar.

Visualisasi ini membantu mengidentifikasi daerah yang membutuhkan intervensi kebijakan pendidikan, seperti penambahan tenaga pengajar atau redistribusi sumber daya pendidikan.

8. Hubungan APM vs Rasio Murid per Guru SD di kabupaten jawa timur



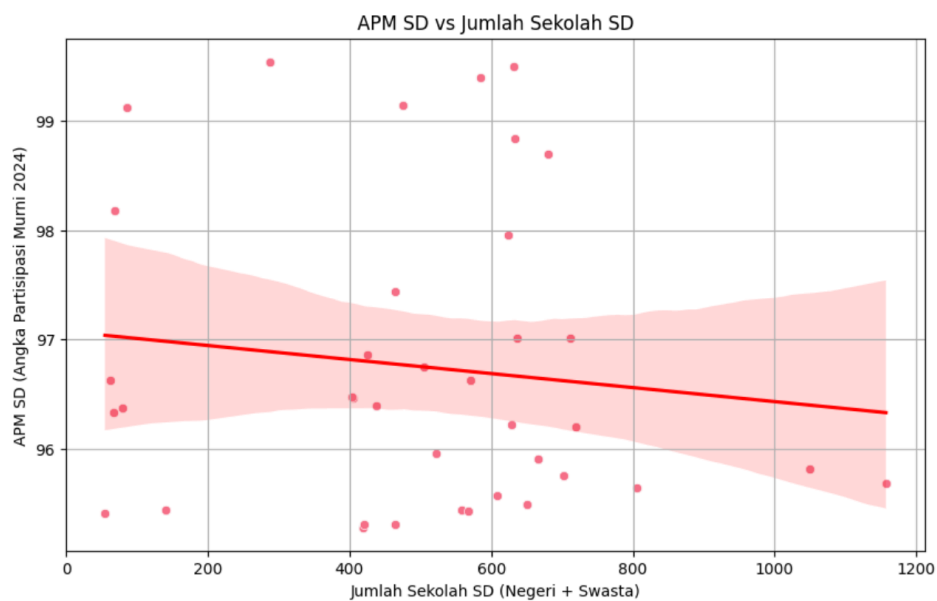
Grafik menunjukkan tren positif ringan, yang mengindikasikan bahwa wilayah dengan rasio murid-guru lebih tinggi cenderung memiliki APM SD yang sedikit lebih tinggi.

Hal ini dapat diinterpretasikan sebagai indikasi bahwa meskipun jumlah murid per guru meningkat, partisipasi murni tetap terjaga atau

bahkan meningkat, kemungkinan karena faktor lain seperti persebaran sekolah, motivasi orang tua, atau kebijakan lokal.

Namun, tren ini juga perlu dianalisis lebih lanjut karena rasio murid-guru yang terlalu tinggi berpotensi menurunkan kualitas pembelajaran jika tidak diimbangi dengan kapasitas dan kompetensi tenaga pengajar.

9. Hubungan APM SD vs Jumlah Sekolah di kabupaten jawa timur



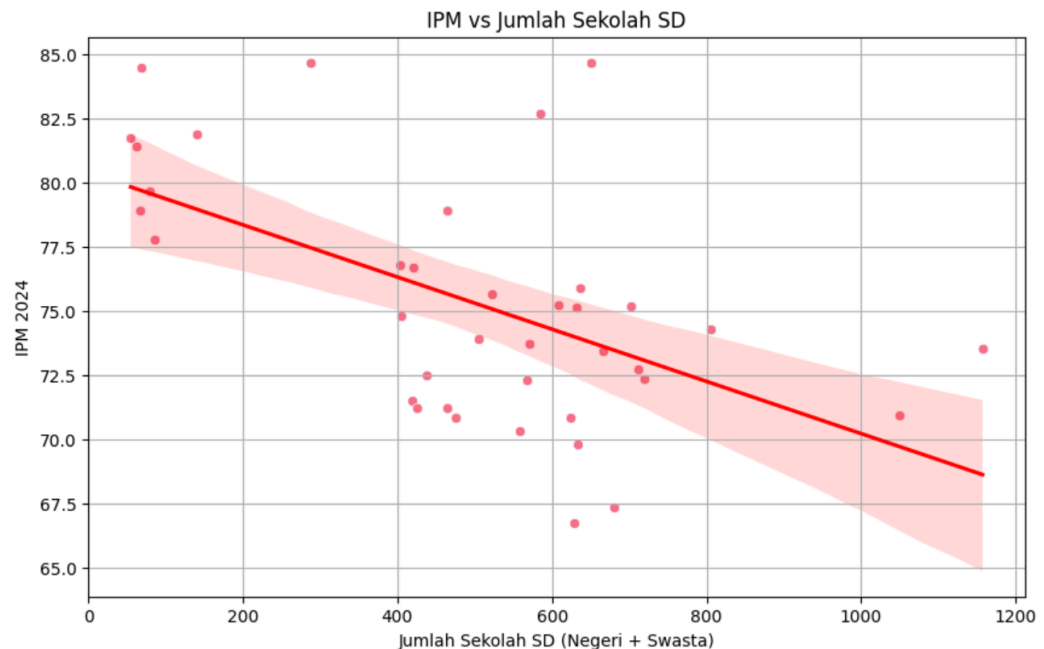
Grafik menunjukkan tren negatif ringan, yang mengindikasikan bahwa peningkatan jumlah sekolah tidak selalu berbanding lurus dengan peningkatan APM SD.

Fenomena ini dapat disebabkan oleh berbagai faktor, seperti:

- a. Ketidakseimbangan antara kapasitas sekolah dan jumlah penduduk usia sekolah
- b. Persebaran sekolah yang tidak merata secara geografis
- c. Faktor sosial-ekonomi yang mempengaruhi keputusan orang tua untuk menyekolahkan anak

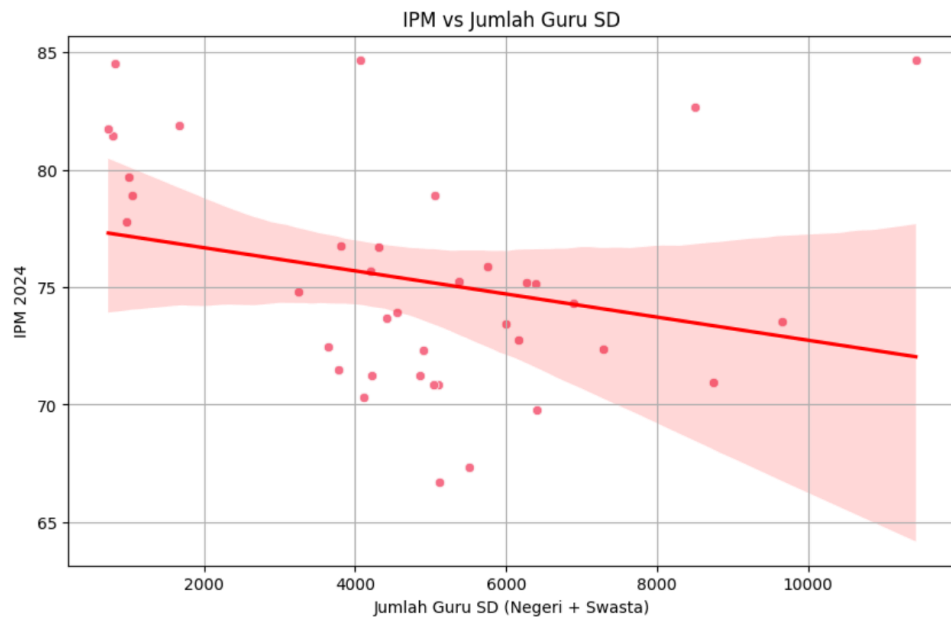
Hasil ini menunjukkan bahwa kuantitas sekolah bukan satu-satunya penentu partisipasi murni, dan perlu dikaji bersama variabel lain seperti kualitas pendidikan, aksesibilitas, dan kebijakan lokal.

10. IPM vs Jumlah Sekolah SD



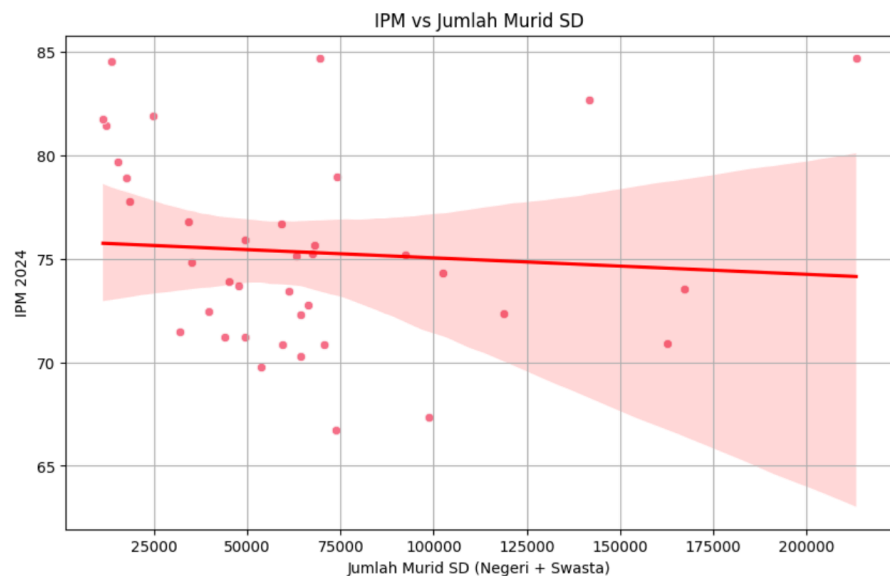
- a. Grafik menunjukkan tren negatif, yang mengindikasikan bahwa wilayah dengan jumlah sekolah SD yang lebih banyak cenderung memiliki nilai IPM yang lebih rendah.
- b. Fenomena ini dapat mencerminkan bahwa kuantitas sekolah tidak selalu berbanding lurus dengan kualitas pembangunan manusia, atau bahwa daerah dengan IPM rendah cenderung membangun lebih banyak sekolah sebagai upaya peningkatan akses pendidikan.
- c. Korelasi negatif ini perlu dianalisis lebih lanjut dengan mempertimbangkan faktor lain seperti kualitas tenaga pengajar, distribusi sekolah, kondisi sosial-ekonomi, dan efektivitas kebijakan pendidikan.

11. IPM vs Jumlah Guru SD



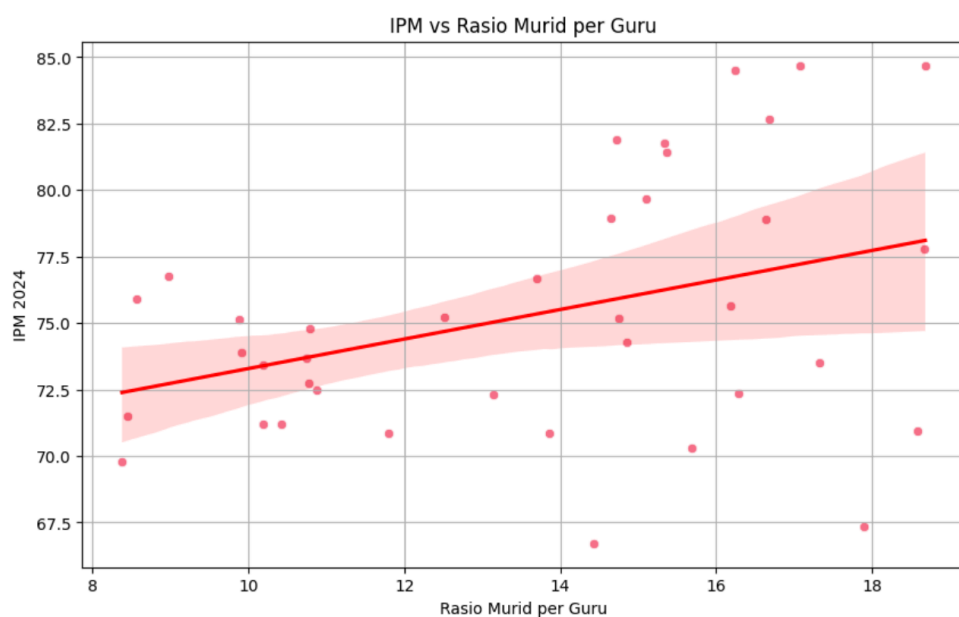
- Grafik menunjukkan tren negatif, yang mengindikasikan bahwa wilayah dengan jumlah guru SD yang lebih banyak cenderung memiliki nilai IPM yang lebih rendah.
- Korelasi ini dapat mencerminkan bahwa daerah dengan IPM rendah berupaya meningkatkan akses pendidikan melalui penambahan tenaga pengajar, atau bahwa jumlah guru belum sebanding dengan kualitas layanan pendidikan yang diberikan.
- Fenomena ini perlu dianalisis lebih lanjut dengan mempertimbangkan variabel lain seperti rasio murid-guru, distribusi geografis, dan efektivitas kebijakan pendidikan.

12. IPM vs Jumlah Murid SD



- Grafik menunjukkan tren negatif ringan, yang mengindikasikan bahwa wilayah dengan jumlah murid SD yang lebih tinggi cenderung memiliki nilai IPM yang sedikit lebih rendah.
- Korelasi ini dapat mencerminkan tantangan dalam pemerataan kualitas pendidikan di daerah dengan populasi usia sekolah yang besar, seperti keterbatasan fasilitas, rasio murid-guru yang tinggi, atau tekanan terhadap kapasitas layanan pendidikan.
- Temuan ini menunjukkan bahwa peningkatan jumlah murid perlu diimbangi dengan peningkatan kualitas dan kapasitas pendidikan agar dapat berkontribusi positif terhadap pembangunan manusia.

13. IPM vs Rasio Murid per Guru

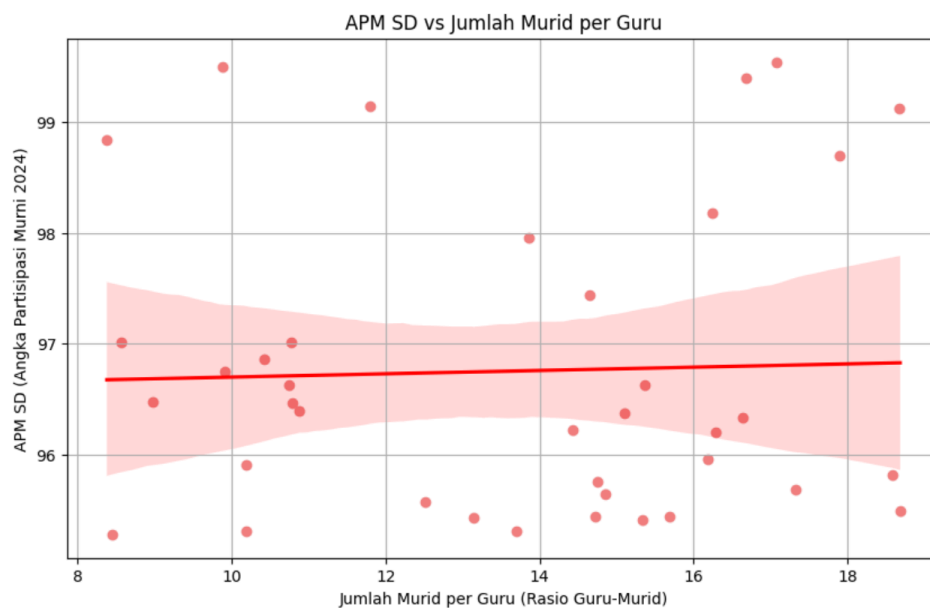


- Grafik menunjukkan tren positif, yang mengindikasikan bahwa wilayah dengan rasio murid-guru yang lebih tinggi cenderung memiliki nilai IPM yang lebih tinggi.
- Korelasi ini dapat diinterpretasikan sebagai indikasi bahwa daerah dengan pembangunan manusia yang lebih baik mampu menampung

lebih banyak peserta didik per guru, atau bahwa tingginya rasio murid-guru tidak selalu menjadi hambatan dalam pencapaian IPM.

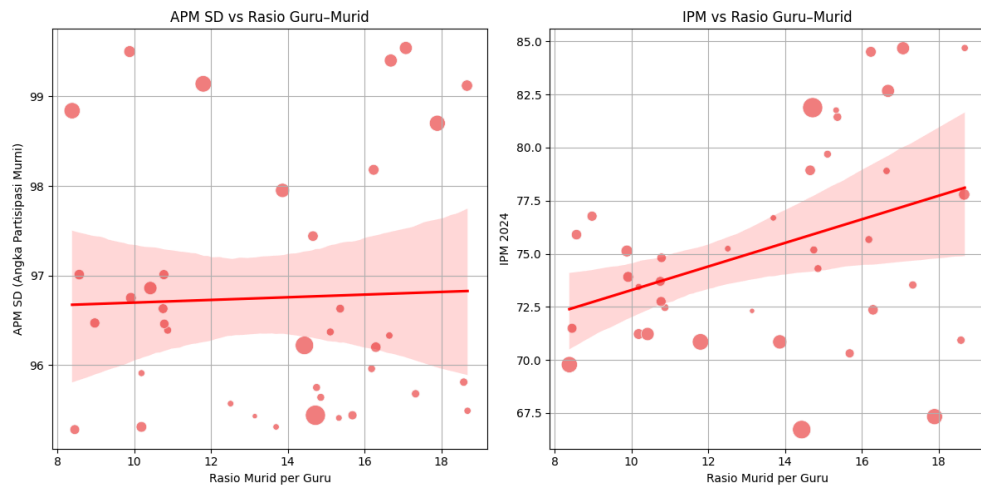
- c. Temuan ini perlu dianalisis lebih lanjut dengan mempertimbangkan faktor lain seperti kualitas tenaga pengajar, efisiensi sistem pendidikan, dan dukungan kebijakan lokal.

14. APM SD vs Jumlah Murid per Guru



- a. Grafik menunjukkan tren positif ringan, yang mengindikasikan bahwa wilayah dengan rasio murid-guru yang lebih tinggi cenderung memiliki APM SD yang sedikit lebih tinggi.
- b. Temuan ini dapat diinterpretasikan sebagai indikasi bahwa peningkatan jumlah murid per guru tidak serta-merta menurunkan partisipasi murni, dan bisa jadi mencerminkan efisiensi sistem pendidikan atau keberhasilan kebijakan pemerataan akses sekolah dasar.
- c. Namun, tren ini juga perlu dianalisis lebih lanjut dengan mempertimbangkan kualitas pengajaran, kapasitas sekolah, dan dukungan sosial-ekonomi di masing-masing wilayah.

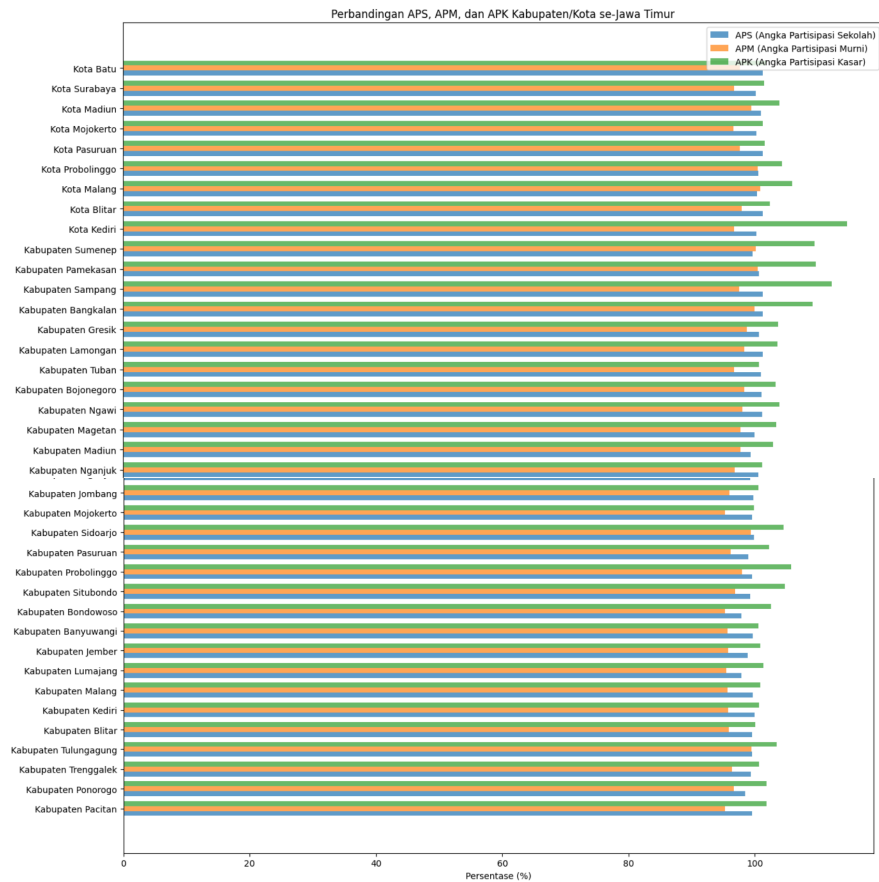
15. Perbandingan APM dan IPM vs Rasio Guru-Murid



Grafik APM menunjukkan tren positif ringan, yang mengindikasikan bahwa wilayah dengan rasio murid-guru lebih tinggi cenderung memiliki APM SD yang sedikit lebih tinggi. Hal ini dapat mencerminkan efisiensi sistem pendidikan atau keberhasilan kebijakan pemerataan akses pendidikan dasar, meskipun perlu dikaji lebih lanjut dampaknya terhadap kualitas pembelajaran.

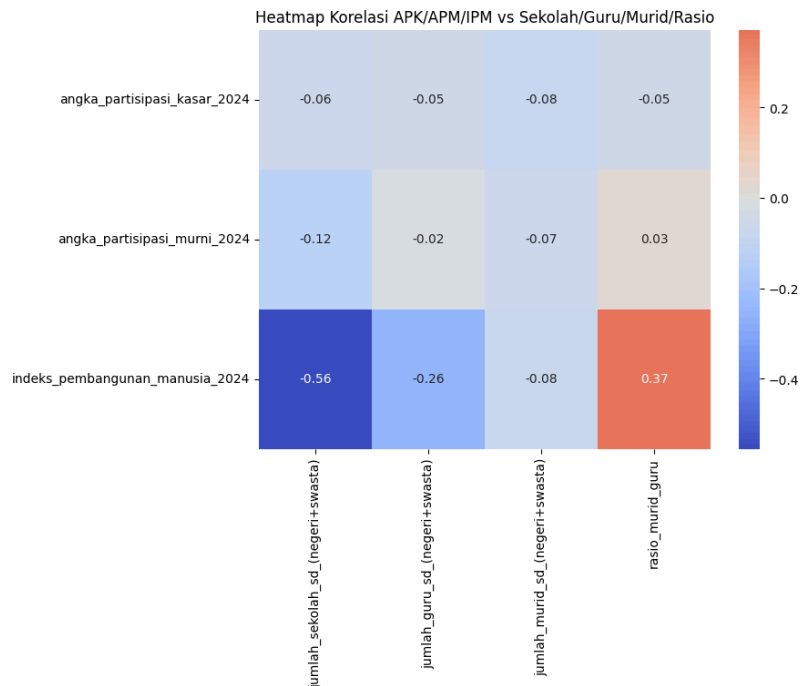
Grafik IPM menunjukkan tren positif yang lebih kuat, yang mengindikasikan bahwa wilayah dengan rasio murid-guru lebih tinggi cenderung memiliki IPM yang lebih tinggi. Temuan ini dapat menunjukkan bahwa daerah dengan pembangunan manusia yang lebih baik mampu menampung lebih banyak peserta didik per guru, atau bahwa rasio tinggi tidak selalu menjadi hambatan dalam pencapaian pembangunan manusia.

16. Perbandingan APS, APM, dan APK Kabupaten /kota se Jawa timur



- Visualisasi ini memberikan gambaran menyeluruh mengenai tingkat partisipasi pendidikan formal di masing-masing wilayah.
- APS menunjukkan proporsi penduduk usia sekolah yang sedang bersekolah, tanpa memperhatikan usia ideal.
- APM mengukur partisipasi siswa yang bersekolah sesuai dengan usia jenjang pendidikan yang seharusnya.
- APK mencerminkan total partisipasi pendidikan pada jenjang tertentu, termasuk mereka yang bersekolah di luar usia ideal.
- Perbedaan panjang batang antar indikator dan antar wilayah menunjukkan ketimpangan partisipasi pendidikan, baik dari segi akses, pemerataan, maupun efisiensi sistem pendidikan.

17. Heatmap Korelasi APK/APM/IPM vs Sekolah/Guru/Murid/Rasio



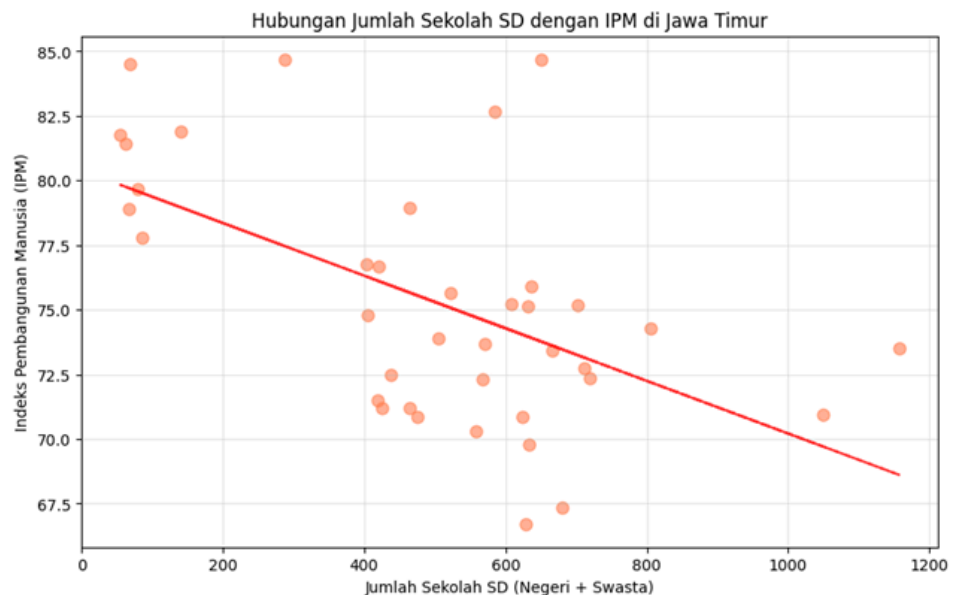
Berdasarkan visualisasi heatmap di atas, terlihat pola hubungan antar variabel sebagai berikut:

1. Korelasi pada APK dan APM: Indikator Angka Partisipasi Kasar (APK) dan Angka Partisipasi Murni (APM) tahun 2024 menunjukkan hubungan korelasi yang sangat lemah (mendekati nol) terhadap seluruh variabel independen (jumlah sekolah, guru, murid, dan rasio), dengan nilai berkisar antara -0.12 hingga 0.03. Hal ini menandakan bahwa kuantitas infrastruktur sekolah dan guru tidak secara langsung mempengaruhi persentase partisipasi siswa secara linier.
2. Korelasi pada IPM (Indeks Pembangunan Manusia): Pola yang lebih signifikan terlihat pada variabel IPM:
 - a. Terdapat korelasi negatif yang cukup kuat (-0.56) antara IPM dengan jumlah total sekolah SD (negeri dan swasta). Hal ini

mengindikasikan bahwa wilayah dengan IPM tinggi cenderung memiliki jumlah unit sekolah yang lebih sedikit. Fenomena ini kemungkinan terjadi karena wilayah perkotaan (IPM tinggi) memiliki sekolah-sekolah besar dengan daya tampung tinggi, sedangkan wilayah pedesaan (IPM rendah) memiliki banyak unit sekolah kecil yang tersebar.

- b. Sebaliknya, terdapat korelasi positif moderat (0.37) antara IPM dengan rasio murid-guru. Ini menunjukkan bahwa daerah dengan pembangunan manusia yang lebih maju cenderung memiliki rasio murid per guru yang lebih tinggi (kelas yang lebih padat/efisien), dibandingkan daerah dengan IPM rendah yang mungkin memiliki rasio lebih kecil karena sebaran murid yang sedikit.
- c. Korelasi negatif lemah juga terlihat antara IPM dengan jumlah guru (-0.26), yang memperkuat dugaan efisiensi tenaga pengajar di wilayah dengan IPM tinggi.

18. Perbandingan IPM vs Jumlah Sekolah

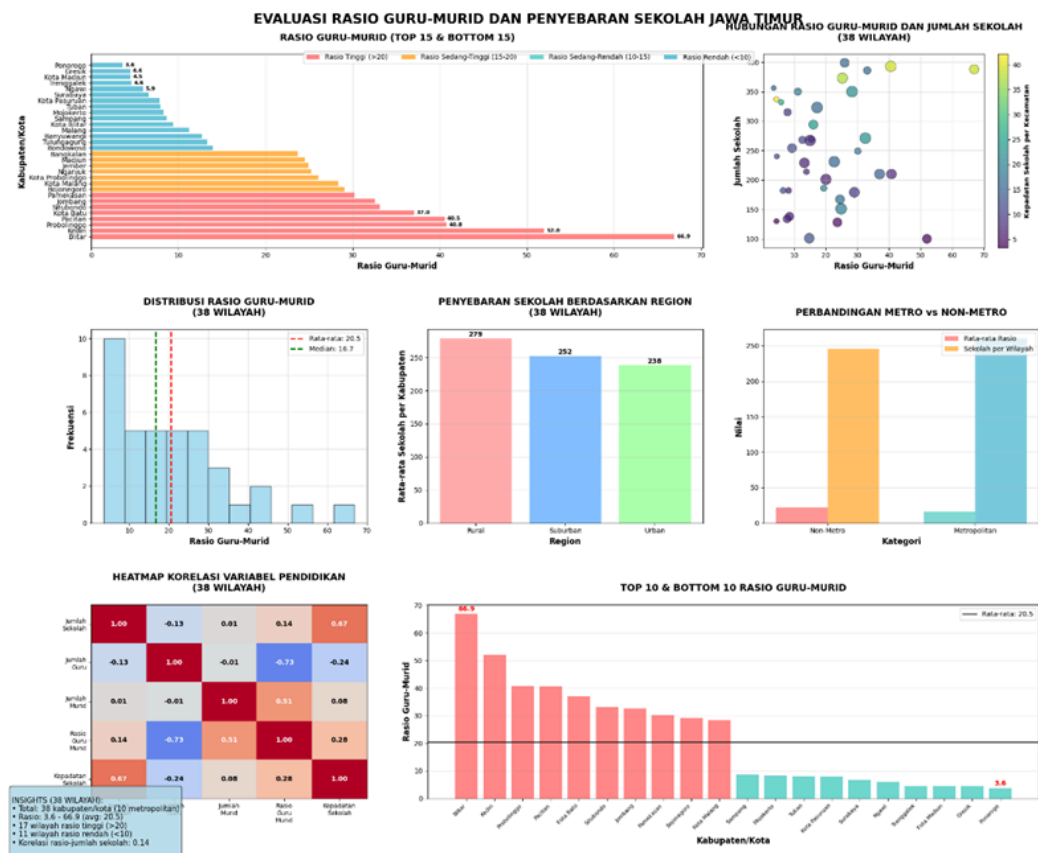


Berdasarkan visualisasi scatter plot di atas, terlihat adanya tren negatif yang cukup jelas antara jumlah sekolah SD dan IPM. Hal ini terkonfirmasi secara

statistik oleh nilai korelasi pada heatmap yang menunjukkan angka sebesar -0.56 (korelasi negatif sedang).

Grafik memperlihatkan bahwa wilayah dengan jumlah sekolah yang sangat banyak justru cenderung memiliki skor IPM yang lebih rendah. Sebaliknya, wilayah-wilayah dengan skor IPM tinggi (umumnya wilayah perkotaan) memiliki jumlah unit sekolah yang lebih sedikit. Fenomena ini mengindikasikan bahwa kuantitas jumlah sekolah bukan penentu utama tingginya Indeks Pembangunan Manusia. Wilayah dengan IPM tinggi kemungkinan memiliki sistem pendidikan yang lebih efisien dengan sekolah-sekolah berkapasitas besar, sedangkan wilayah dengan IPM rendah memiliki banyak unit sekolah kecil yang tersebar luas untuk menjangkau akses geografis, namun belum tentu diiringi dengan kualitas pembangunan manusia yang merata.

19. Evaluasi Rasio Guru-Murid dan Penyebaran Sekolah



Gambar di atas menampilkan beberapa analisis terkait rasio guru-murid dan sebaran sekolah di 38 kabupaten/kota Jawa Timur:

- a. Rasio Guru-Murid (Top 15 & Bottom 15): Grafik batang menunjukkan kabupaten/kota dengan rasio guru-murid tertinggi dan terendah. Terlihat bahwa beberapa wilayah pedesaan memiliki rasio yang sangat tinggi (>50), menandakan jumlah murid per guru sangat banyak, sedangkan kota-kota besar cenderung memiliki rasio lebih rendah.
- b. Hubungan Rasio Guru-Murid dan Jumlah Sekolah: Scatter plot menampilkan hubungan antara jumlah sekolah dan rasio guru-murid. Tidak terlihat hubungan yang kuat (korelasi rendah: 0,14), menunjukkan bahwa jumlah sekolah tidak selalu menentukan rasio guru-murid.
- c. Distribusi Rasio Guru-Murid: Histogram menampilkan frekuensi rasio guru-murid di 38 wilayah, dengan rata-rata 20,5 dan median 16,7. Sebagian besar wilayah memiliki rasio <20 , tetapi ada beberapa outlier dengan rasio sangat tinggi.
- d. Penyebaran Sekolah Berdasarkan Region: Grafik batang memperlihatkan rata-rata jumlah sekolah di wilayah rural, suburban, dan urban. Wilayah rural cenderung memiliki jumlah sekolah lebih banyak dibanding suburban dan urban.
- e. Perbandingan Metro vs Non-Metro: Grafik menunjukkan bahwa meskipun wilayah metropolitan memiliki jumlah sekolah yang tinggi, rasio guru-murid tetap rendah, menandakan kualitas pengajaran lebih tersebar merata.
- f. Heatmap Korelasi Variabel Pendidikan: Korelasi antara jumlah sekolah, jumlah guru, jumlah murid, rasio guru-murid, dan kepadatan sekolah menunjukkan bahwa rasio guru-murid memiliki korelasi rendah terhadap jumlah sekolah (0,14), menandakan faktor lain memengaruhi kualitas distribusi guru.

- g. Top 10 & Bottom 10 Rasio Guru-Murid: Grafik ini menegaskan bahwa wilayah dengan rasio guru-murid tinggi sebagian besar berada di pedesaan, sedangkan wilayah kota memiliki rasio rendah.

4.3 Kendala dan Rencana Tindak Lanjut

1. Kendala Proses Pengerjaan

Selama proses data wrangling, proyek ini tidak menghadapi kendala besar yang menghambat analisis. Beberapa hal yang perlu dicatat antara lain:

- a. Koordinasi Tim: Koordinasi antar anggota tim terkadang sedikit menantang karena perbedaan jadwal, namun berhasil diatasi melalui komunikasi intensif.
- b. Ekstraksi Data Manual: Data awal yang dai pdf menyambungkan kedua excel untuk Jawa Timur dihubungkan secara manual.

2. Rencana Tindak Lanjut

Setelah data bersih dan terintegrasi, beberapa langkah analisis lanjutan yang dapat dilakukan antara lain:

- a. Analisis Kualitas Pendidikan: Menyelidiki faktor-faktor yang mempengaruhi IPM, seperti rasio guru-murid, jumlah sekolah, dan indikator sosial-ekonomi lain, untuk memberikan rekomendasi kebijakan berbasis data.
- b. Analisis Klaster (Clustering): Mengelompokkan 38 kabupaten/kota di Jawa Timur berdasarkan karakteristik pendidikan menggunakan metode clustering (misal: K-Means) untuk mengidentifikasi pola kesenjangan pendidikan.
- c. Visualisasi Lanjutan: Membuat visualisasi interaktif untuk memudahkan interpretasi data dan mendukung pembuatan kebijakan.

BAB 5 KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan proses data wrangling, integrasi data, dan analisis eksplorasi data (EDA) terhadap data pendidikan dasar di Provinsi Jawa Timur, diperoleh beberapa kesimpulan sebagai berikut:

1. **Kondisi Partisipasi Pendidikan Tinggi dan Merata**

Angka Partisipasi Sekolah (APS) Provinsi Jawa Timur pada jenjang SD berada pada nilai 99,34%, menunjukkan bahwa hampir seluruh anak usia sekolah dasar telah mengakses pendidikan. Selain itu, sebagian besar wilayah memiliki APK di atas 100%, sehingga aksesibilitas pendidikan dasar tergolong sangat baik dan merata.

2. **Terdapat Kesenjangan dalam Indeks Pembangunan Manusia (IPM)**

Meskipun akses pendidikan merata, distribusi IPM antar kabupaten/kota menunjukkan kesenjangan. Wilayah urban seperti Kota Surabaya, Kota Malang, dan Kota Madiun mendominasi posisi IPM tertinggi, sedangkan wilayah rural dan kepulauan cenderung memiliki nilai lebih rendah. Hal ini menandakan bahwa keterjangkauan pendidikan tidak selalu paralel dengan kualitas pembangunan manusia.

3. **Hubungan Antara Jumlah Sekolah dan Indikator Pendidikan Tidak Selalu Linear**

Scatter plot yang membandingkan jumlah sekolah dengan indikator seperti APK, APM, dan IPM menunjukkan bahwa bertambahnya jumlah sekolah tidak otomatis meningkatkan kualitas atau partisipasi pendidikan. Faktor geografis, distribusi penduduk, dan kualitas fasilitas lebih berperan dalam mempengaruhi capaian pendidikan daerah.

4. Variasi Rasio Murid-Guru Mengindikasikan Ketimpangan Tenaga Pengajar

Analisis rasio murid per guru menunjukkan variasi signifikan antar kabupaten/kota. Beberapa wilayah memiliki rasio tinggi yang berpotensi menurunkan kualitas pembelajaran karena beban kerja guru yang berat. Hal ini menjadi indikator penting perlunya pemerataan tenaga pendidik.

5. Terdapat Korelasi yang Cukup Relevan antar Variabel Pendidikan

Heatmap korelasi memperlihatkan hubungan moderat–kuat antara variabel seperti APK, APM, jumlah murid, dan jumlah guru. Namun, hubungan antara IPM dan variabel infrastruktur (jumlah sekolah) cenderung lemah hingga negatif, menegaskan bahwa kualitas pembangunan manusia lebih dipengaruhi faktor kompleks seperti ekonomi, kualitas pengajaran, dan lingkungan sosial.

Secara keseluruhan, proyek ini menunjukkan bahwa meskipun akses pendidikan dasar di Jawa Timur sangat baik, kualitas pendidikan dan ketimpangan pembangunan manusia masih menjadi tantangan yang perlu diperhatikan dalam kebijakan pendidikan.

5.2 Saran

Berdasarkan temuan analisis pada penelitian ini, beberapa saran yang dapat diberikan adalah:

1. Saran Kebijakan Pendidikan

a. Pemerataan Distribusi Guru

Pemerintah daerah perlu memprioritaskan redistribusi tenaga pengajar ke wilayah dengan rasio murid-guru tinggi agar beban kerja guru lebih

b. proporsional dan kualitas pembelajaran lebih terjaga.

Peningkatan Kualitas Sekolah di Wilayah Rural

Meskipun jumlah sekolah banyak, beberapa wilayah IPM rendah membutuhkan peningkatan kualitas sarana pendidikan, seperti fasilitas kelas, perpustakaan, dan akses teknologi.

- c. Intervensi Terarah pada Wilayah dengan IPM Rendah
Kebijakan khusus seperti bantuan sosial pendidikan, peningkatan transportasi sekolah, dan pelatihan guru dapat membantu memperbaiki capaian pembangunan manusia di daerah yang tertinggal.

2. Saran untuk Penelitian Selanjutnya

- a. Menggunakan Data yang Lebih Luas dan Beragam
Penelitian berikutnya dapat menggabungkan data kualitas guru, akreditasi sekolah, atau faktor ekonomi sebagai variabel untuk melihat hubungan yang lebih komprehensif dengan IPM.
- b. Menggunakan Teknik Analitik yang Lebih Lanjut
Model statistik atau machine learning (misalnya regresi linier, clustering, atau random forest) dapat digunakan untuk memberikan prediksi dan insight yang lebih mendalam terkait faktor yang paling mempengaruhi IPM atau rasio murid-guru.
- c. Integrasi Data Spasial
Analisis geospasial (GIS) dapat ditambahkan untuk memetakan ketimpangan pendidikan secara visual berdasarkan letak geografis, sehingga kebijakan pemerataan dapat lebih tepat sasaran.

3. Saran untuk Implementasi Teknologi

- a. Pengembangan Dashboard Pemantauan Pendidikan
Data yang telah dibersihkan dapat digunakan untuk membuat dashboard interaktif agar pemerintah dan publik dapat memantau kualitas pendidikan secara real time.
- b. Automasi Data Wrangling
Tahapan pengolahan data dapat diotomatisasi menggunakan pipeline sehingga proses pembaruan data tahunan menjadi lebih cepat dan efisien.

BAB 6 KONTRIBUSI ANGGOTA KELOMPOK

1. Atya Azharina Aziz (24031554194)

- Merumuskan ide proyek.
- Mengumpulkan 3 sumber data
- Menyusun latar belakang, rumusan masalah, dan tujuan analisis.
- Mengumpulkan data dari berbagai sumber (Excel, CSV, PDF/Web).
- Membersihkan, mengintegrasikan, dan mentransformasi data.
- Melakukan analisis eksplorasi data (EDA)
- Membuat semua visualisasi: scatter plot, histogram, boxplot, bar chart.
- Menyusun dan merapikan laporan akhir.

2. Muhamad Fadly Saifullah (24031554173)

- Mengumpulkan 1 sumber data
- Melakukan web scraping.
- Merumuskan ide proyek.
- Menyusun dan merapikan laporan akhir.
- Menyusun PPT.

Daftar Pustaka

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.

Kandel, S., Paepcke, A., Hellerstein, J., & Heer, J. (2011). *Wrangler: Interactive visual specification of data transformation scripts*. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3363–3372). ACM.

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.

Peraturan Pemerintah Republik Indonesia Nomor 19 Tahun 2005 tentang Standar Nasional Pendidikan.

Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 23 Tahun 2013 tentang Standar Penyelenggaraan Pendidikan.

LAMPIRAN

1. Pipeline Data Wrangling proyek Analisis SD di Jawa Timur

