Eksplorasi Data Kepadatan Penduduk dan Kondisi Sosial Ekonomi Antarprovinsi di Indonesia Menggunakan Python

Emon Sejahtera Zendrato, Cika Satrio Putri, Yelipe Laia*, Dwika Ardani Fadillah, Rahmadi Pasha Ramadhan, Kania Restya Diva, Hanna Dewi M. Hutabarat
Universitas Negeri Medan (UNIMED)

Received: 27 September 2026 Accepted: 12 October 2026 Published: 1 December 2026

*yelipelaia69@gmail.com

Kata Kunci: *Exploratory Data Analysis*, kepadatan penduduk, kemiskinan, TPT, IPM, PDRB per kapita.

DSI: Jurnal Data Science Indonesia is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

Abstrak : Penelitian bertujuan mengeksplorasi karakteristik kepadatan penduduk dan kondisi sosial ekonomi antarprovinsi di Indonesia serta mengidentifikasi kestabilan hubungan antarindikator terhadap keberadaan nilai ekstrem. Data yang digunakan merupakan data sekunder Badan Pusat Statistik tahun 2021 dengan unit analisis sebanyak 34 provinsi. Variabel yang dianalisis meliputi kepadatan penduduk, tingkat kemiskinan, Tingkat Pengangguran Terbuka (TPT), Indeks Pembangunan Manusia (IPM), dan Produk Domestik Regional Bruto (PDRB) per kapita. Analisis dilakukan menggunakan Exploratory Data Analysis (EDA) melalui statistik deskriptif, histogram, boxplot, scatter plot, serta korelasi Pearson dan Spearman. Analisis korelasi dilakukan pada seluruh provinsi dan diulang setelah DKI Jakarta dikeluarkan untuk mengevaluasi sensitivitas hasil terhadap nilai ekstrem. Hasil menunjukkan bahwa kepadatan penduduk memiliki variasi yang sangat besar antarprovinsi. Hubungan kepadatan penduduk dengan IPM relatif konsisten positif pada Pearson dan Spearman serta tetap signifikan setelah DKI Jakarta dikeluarkan. Sebaliknya, hubungan kepadatan penduduk dengan PDRB per kapita menunjukkan perubahan arah yang besar ketika DKI Jakarta dikeluarkan. Temuan ini menunjukkan bahwa nilai ekstrem dapat memengaruhi interpretasi hubungan antarvariabel, sehingga penggunaan lebih dari satu metode korelasi dan pemeriksaan sensitivitas terhadap outlier penting dalam analisis data antarprovinsi.

PENDAHULUAN

Indonesia merupakan negara dengan karakteristik wilayah dan jumlah penduduk yang beragam, sehingga kondisi kependudukan dan sosial ekonomi antarprovinsi tidak selalu menunjukkan pola yang sama. Salah satu indikator yang dapat menggambarkan karakteristik kependudukan suatu wilayah adalah kepadatan penduduk, yaitu perbandingan antara jumlah penduduk dengan luas wilayah. Data BPS menunjukkan bahwa kepadatan penduduk antarprovinsi pada tahun 2021 memiliki perbedaan yang sangat besar[1] [2]. Perbedaan kepadatan penduduk tersebut dapat dilihat bersama berbagai indikator sosial ekonomi, seperti tingkat kemiskinan, Tingkat Pengangguran Terbuka (TPT), Indeks Pembangunan Manusia (IPM), dan Produk Domestik Regional Bruto (PDRB) per kapita. Agustina, menunjukkan bahwa kepadatan penduduk merupakan salah satu indikator yang dapat digunakan untuk membandingkan karakteristik kependudukan antarprovinsi[3], sedangkan Permatasari, menggunakan kemiskinan, TPT, IPM, dan PDRB per kapita untuk menggambarkan perbedaan kondisi sosial ekonomi antarprovinsi[4].

Perbedaan karakteristik antarprovinsi tersebut perlu dieksplorasi untuk memperoleh gambaran mengenai pola dan variasi kepadatan penduduk serta kondisi sosial ekonomi di Indonesia. Analisis terhadap beberapa indikator secara bersamaan dapat membantu menunjukkan wilayah dengan karakteristik yang relatif berbeda. Azrai, menemukan bahwa kepadatan penduduk dan kemiskinan berkaitan dengan IPM dalam analisis pembangunan manusia di Indonesia[5]. Sementara itu, penelitian mengenai TPT

menunjukkan bahwa kondisi ketenagakerjaan daerah dapat berkaitan dengan faktor seperti investasi, pendidikan, pertumbuhan penduduk, dan karakteristik ekonomi wilayah[6] [7].

Exploratory Data Analysis (EDA) dapat digunakan untuk memperoleh pemahaman awal terhadap karakteristik data melalui statistik deskriptif, pemeriksaan distribusi, visualisasi, dan eksplorasi hubungan antarvariabel. Srihith, menunjukkan bahwa Python dapat digunakan untuk mengotomatisasi berbagai tahapan EDA dan memperoleh ringkasan informasi statistik dari dataset.[8] Clemente, juga menjelaskan bahwa proses profiling data dapat membantu mengidentifikasi distribusi, korelasi, nilai ekstrem, duplikasi, serta masalah kualitas data.[9] Penggunaan Python dengan pustaka pengolahan dan visualisasi data karena itu sesuai untuk mendukung eksplorasi dataset antarprovinsi secara sistematis.

Penelitian memiliki fokus yang berbeda dari eksplorasi deskriptif yang hanya menggambarkan distribusi dan hubungan antarvariabel. Analisis diarahkan pada pemeriksaan kestabilan hubungan antara kepadatan penduduk dan indikator sosial ekonomi ketika terdapat nilai ekstrem antarprovinsi. Secara khusus, penelitian membandingkan korelasi Pearson dan Spearman serta membandingkan hasil korelasi sebelum dan sesudah DKI Jakarta dikeluarkan dari analisis. Pendekatan tersebut digunakan untuk melihat apakah pola hubungan yang diperoleh tetap konsisten atau berubah akibat karakteristik ekstrem suatu provinsi. Dengan demikian, kontribusi penelitian terletak pada evaluasi sensitivitas hasil korelasi dalam konteks ketimpangan karakteristik antarprovinsi, bukan pada penggunaan Python sebagai alat analisis.

Berdasarkan hal tersebut, penelitian ini dilakukan untuk mengeksplorasi karakteristik kepadatan penduduk dan kondisi sosial ekonomi antarprovinsi di Indonesia tahun 2021 serta mengevaluasi pola hubungan antarindikator. Kondisi sosial ekonomi direpresentasikan oleh tingkat kemiskinan, TPT, IPM, dan PDRB per kapita. Analisis dilakukan melalui statistik deskriptif dan visualisasi, kemudian dilanjutkan dengan korelasi Pearson dan Spearman. Untuk mengetahui sensitivitas hasil terhadap nilai ekstrem, korelasi dihitung menggunakan seluruh 34 provinsi dan kembali dihitung setelah DKI Jakarta dikeluarkan. Pendekatan tersebut digunakan untuk mengidentifikasi hubungan yang relatif konsisten dan hubungan yang berubah akibat keberadaan pengamatan ekstrem. Pertanyaan adalah bagaimana karakteristik dan variasi kepadatan penduduk serta kondisi sosial ekonomi antarprovinsi di Indonesia tahun 2021, bagaimana pola hubungan kepadatan penduduk dengan indikator sosial ekonomi, dan sejauh mana keberadaan DKI Jakarta memengaruhi hasil korelasi antarindikator.

TINJAUAN LITERATUR

Kepadatan Penduduk

Kepadatan penduduk adalah ukuran yang menggambarkan jumlah penduduk yang menempati suatu luas wilayah tertentu. Indikator ini dapat digunakan untuk melihat konsentrasi penduduk dan membandingkan karakteristik antarwilayah. Agustina, menggunakan kepadatan penduduk bersama sejumlah indikator kependudukan untuk membandingkan karakteristik 34 provinsi pada beberapa periode[3] Data BPS juga menunjukkan adanya perbedaan kepadatan penduduk yang besar antarprovinsi pada tahun 2021[2]. Kepadatan penduduk digunakan untuk menggambarkan variasi antarprovinsi dan mengeksplorasi keterkaitannya dengan indikator sosial ekonomi.

Kondisi Sosial Ekonomi

Kondisi sosial ekonomi merupakan gambaran mengenai keadaan sosial dan ekonomi masyarakat dalam suatu wilayah yang dapat dilihat melalui beberapa indikator. Kondisi tersebut direpresentasikan oleh tingkat kemiskinan, TPT, IPM, dan PDRB per kapita. Permatasari, menggunakan keempat indikator tersebut untuk mengelompokkan provinsi berdasarkan karakteristik sosial ekonomi, sehingga indikator-indikator tersebut dapat digunakan secara bersama-sama untuk melihat perbedaan antarwilayah[4]

Tingkat Kemiskinan

Tingkat kemiskinan merupakan indikator yang menggambarkan proporsi penduduk yang berada dalam kondisi miskin pada suatu wilayah. Indikator digunakan untuk melihat perbedaan kesejahteraan antarwilayah. BPS menyediakan data jumlah dan persentase penduduk miskin menurut provinsi, termasuk untuk tahun 2021[10] Dalam kajian Azrai, kemiskinan juga ditempatkan sebagai salah satu variabel yang berkaitan dengan pembangunan manusia[5]. Oleh karena itu, tingkat kemiskinan dalam penelitian ini digunakan untuk membandingkan kondisi kesejahteraan antarprovinsi dan mengeksplorasi polanya bersama kepadatan penduduk serta indikator lainnya.

Tingkat Pengangguran Terbuka (TPT)

Tingkat Pengangguran Terbuka (TPT) merupakan persentase angkatan kerja yang tidak bekerja dan termasuk dalam kelompok pengangguran sesuai definisi statistik ketenagakerjaan. TPT dapat digunakan untuk menggambarkan kondisi pasar kerja suatu wilayah. Data BPS menunjukkan adanya perbedaan TPT antarprovinsi pada 2021[11]. Penelitian Sinambela dan Wenagama menunjukkan bahwa TPT di wilayah Bali berkaitan dengan beberapa karakteristik ekonomi dan pendidikan, sedangkan Nurhijah, menemukan bahwa pertumbuhan penduduk, partisipasi angkatan kerja, dan upah minimum termasuk variabel yang berpengaruh dalam model TPT di Jawa Barat [6] [7] [12] [13]. TPT digunakan untuk mengeksplorasi variasi kondisi ketenagakerjaan antarprovinsi.

Indeks Pembangunan Manusia (IPM)

Indeks Pembangunan Manusia (IPM) merupakan indikator yang menggambarkan capaian pembangunan manusia melalui dimensi dasar yang berkaitan dengan kehidupan manusia. BPS menerbitkan data IPM menurut provinsi dan publikasi IPM 2021 sebagai sumber informasi mengenai capaian pembangunan manusia.[14] IPM juga digunakan sebagai salah satu indikator sosial ekonomi dalam penelitian Permatasari et al., sedangkan Azrai et al. mengkaji IPM bersama kepadatan penduduk dan kemiskinan[5] [4]. IPM digunakan untuk membandingkan capaian pembangunan manusia antarprovinsi dan mengeksplorasi hubungannya dengan indikator lainnya.

PDRB per Kapita

PDRB per kapita merupakan ukuran yang diperoleh dengan membandingkan nilai PDRB suatu wilayah dengan jumlah penduduknya dan dapat digunakan untuk memberikan gambaran mengenai kondisi ekonomi rata-rata suatu wilayah. BPS menyediakan data PDRB dan PDRB per kapita dalam publikasi PDRB kabupaten/kota di Indonesia.[15] Permatasari, juga menggunakan PDRB per kapita sebagai salah satu indikator untuk membedakan karakteristik sosial ekonomi provinsi[4]. PDRB per kapita digunakan untuk mengeksplorasi variasi kondisi ekonomi antarprovinsi serta pola hubungannya dengan kepadatan penduduk dan indikator sosial ekonomi lainnya.

Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) adalah proses eksplorasi yang digunakan untuk memahami karakteristik, pola, distribusi, dan hubungan dalam suatu dataset sebelum analisis lebih lanjut. Srihith, membahas penggunaan Python untuk membantu otomatisasi proses EDA dan menghasilkan ringkasan informasi statistik dari data[8]. Clemente, menunjukkan bahwa data profiling dapat membantu meninjau statistik ringkas, distribusi variabel, korelasi, nilai ekstrem, duplikasi, dan beberapa persoalan kualitas data[9]. EDA digunakan untuk melihat karakteristik, perbedaan, distribusi, dan hubungan antarvariabel pada 34 provinsi.

Python dalam Eksplorasi Data

Python merupakan bahasa pemrograman yang dapat digunakan untuk pengolahan dan analisis data karena memiliki berbagai pustaka untuk manipulasi data, perhitungan statistik, dan visualisasi. Penggunaan Python dalam Exploratory Data Analysis (EDA) dapat membantu proses pengolahan data, penyajian ringkasan statistik, serta visualisasi untuk mengidentifikasi pola dan karakteristik data[8] [9]. Python digunakan untuk mengolah dataset, menghitung statistik deskriptif dan korelasi, serta menghasilkan histogram, boxplot, heatmap, dan scatter plot.

METODE PENELITIAN

Penelitian menggunakan pendekatan kuantitatif deskriptif dengan metode Exploratory Data Analysis (EDA) untuk mengeksplorasi karakteristik, persebaran, dan pola hubungan data kependudukan serta kondisi sosial ekonomi antarprovinsi di Indonesia. Data yang digunakan merupakan data sekunder yang bersumber dari Badan Pusat Statistik (BPS) tahun 2021. Pemilihan tahun 2021 dilakukan agar seluruh variabel yang dianalisis berada pada periode yang sama sehingga perbandingan antarprovinsi dapat dilakukan secara lebih konsisten.

Unit analisis dalam penelitian ini adalah 34 provinsi di Indonesia. Variabel yang digunakan terdiri atas lima variabel, yaitu kepadatan penduduk, tingkat kemiskinan, Tingkat Pengangguran Terbuka (TPT), Indeks Pembangunan Manusia (IPM), dan Produk Domestik Regional Bruto (PDRB) per kapita. Kepadatan penduduk dinyatakan dalam jiwa/km^2, tingkat kemiskinan dan TPT dinyatakan dalam persen, IPM dinyatakan dalam indeks, sedangkan PDRB dinyatakan dalam ribu rupiah per kapita. Kelima variabel

tersebut digunakan untuk memberikan gambaran yang lebih menyeluruh mengenai karakteristik kependudukan, kesejahteraan, ketenagakerjaan, pembangunan manusia, dan kondisi ekonomi antarprovinsi.

Tahap pengolahan data diawali dengan mengumpulkan data dari sumber BPS, kemudian melakukan penyesuaian nama provinsi, satuan, dan format angka agar seluruh variabel dapat digabungkan ke dalam satu dataset. Selanjutnya dilakukan pemeriksaan data untuk memastikan tidak terdapat data kosong, duplikasi wilayah, maupun ketidaksesuaian format. Dataset yang telah tersusun kemudian diolah menggunakan Python dengan bantuan Pandas dan NumPy untuk pengelolaan serta perhitungan data, sedangkan Matplotlib dan Seaborn digunakan untuk menghasilkan visualisasi[8] [16]. Analisis dilakukan menggunakan Python versi [3.12.1]. Pengolahan data menggunakan Pandas [2.2.0] dan NumPy [1.26.4], sedangkan visualisasi data dibuat menggunakan Matplotlib [3.5.2] dan Seaborn [0.13.2]. Pencantuman versi perangkat lunak digunakan untuk mendukung reproduksibilitas proses analisis.

Analisis EDA dilakukan melalui beberapa tahap. Pertama, dilakukan statistik deskriptif untuk memperoleh nilai minimum, maksimum, rata-rata, median, kuartil, dan standar deviasi dari kelima variabel. Kedua, distribusi data dianalisis menggunakan histogram dan boxplot untuk melihat pola persebaran, tingkat variasi, serta kemungkinan adanya nilai pencilan. Ketiga, dilakukan perbandingan karakteristik antarprovinsi berdasarkan nilai tertinggi dan terendah pada setiap variabel. Keempat, hubungan antarvariabel dieksplorasi menggunakan scatter plot, matriks korelasi, dan heatmap. Analisis korelasi dilakukan menggunakan korelasi Pearson untuk mengukur hubungan linear dan korelasi Spearman untuk mengukur hubungan berdasarkan peringkat. Masing-masing koefisien korelasi dilengkapi dengan p-value pada taraf signifikansi 5% dan interval kepercayaan 95%. Interval kepercayaan digunakan untuk memberikan informasi mengenai ketidakpastian estimasi koefisien korelasi. Analisis dilakukan menggunakan korelasi Pearson untuk mengevaluasi hubungan linear berdasarkan nilai asli variabel dan korelasi Spearman untuk mengevaluasi hubungan monoton berdasarkan peringkat data. Untuk mengevaluasi sensitivitas hasil korelasi terhadap nilai ekstrem, analisis dilakukan pada seluruh 34 provinsi dan diulang setelah DKI Jakarta dikeluarkan dari dataset. Hasil kedua analisis dibandingkan berdasarkan arah, besaran koefisien, p-value, dan interval kepercayaan 95%. Perbandingan tersebut digunakan untuk menilai kestabilan pola hubungan antara kepadatan penduduk dan indikator sosial ekonomi.

HASIL PENELITIAN

Penelitian menggunakan data 34 provinsi di Indonesia tahun 2021 yang terdiri atas lima indikator, yaitu kepadatan penduduk, tingkat kemiskinan, Tingkat Pengangguran Terbuka (TPT), Indeks Pembangunan Manusia (IPM), dan Produk Domestik Regional Bruto (PDRB) per kapita. Analisis dilakukan menggunakan pendekatan Exploratory Data Analysis (EDA) untuk mengetahui karakteristik, persebaran, serta hubungan antarvariabel.

Tabel 1. Statistik Deskriptif Kepadatan Penduduk dan Indikator Sosial Ekonomi Provinsi di Indonesia Tahun 2021

Variabel	Minimum	Q1	Median	Mean	Q3	Maksimum	Std. Deviasi
Kepadatan Penduduk (jiwa/km^2)	9	54,25	103,5	744,26	261	15.978	2.721,06
Kemiskinan (%)	4,56	6,41	8,51	10,43	12,64	27,38	5,41
TPT (%)	3,01	4,45	5,06	5,49	6,32	9,91	1,82
IPM	60,62	69,73	71,65	71,35	72,6	81,11	3,93
PDRB per Kapita (ribu Rp)	20.555	40.414,75	53.235	67.053,18	62.122,50	274.660	51.796,12

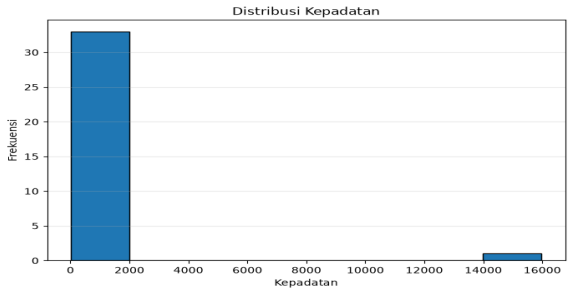
Hasil statistik deskriptif menunjukkan bahwa karakteristik kepadatan penduduk antarprovinsi memiliki ketimpangan yang jauh lebih besar dibandingkan indikator sosial ekonomi lainnya. Perbedaan antara mean dan median menunjukkan bahwa distribusi kepadatan penduduk tidak simetris dan dipengaruhi oleh beberapa provinsi dengan nilai yang sangat tinggi. Kondisi ini penting diperhatikan pada tahap analisis hubungan karena nilai ekstrem dapat memengaruhi ukuran korelasi yang menggunakan nilai asli data. PDRB per kapita juga menunjukkan variasi yang cukup besar, sedangkan kemiskinan, TPT, dan IPM memiliki penyebaran yang relatif lebih terbatas. Perbedaan karakteristik tersebut menunjukkan bahwa setiap

indikator memiliki pola distribusi yang berbeda sehingga interpretasi hubungan antarvariabel perlu mempertimbangkan bentuk distribusi dan keberadaan nilai ekstrem.

Tabel 2 Nilai Tertinggi dan Terendah Antarprovinsi

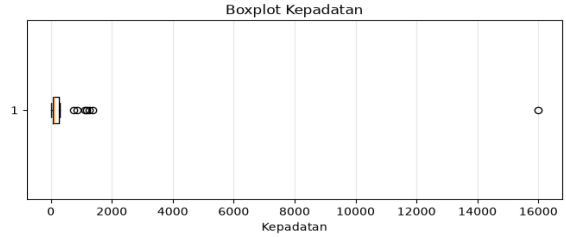
Variabel	Provinsi Tertinggi	Nilai	Provinsi Terendah	Nilai
Kepadatan Penduduk	DKI Jakarta	15.978	Kalimantan Utara	9
Kemiskinan	Papua	27,38%	Kalimantan Selatan	4,56%
TPT	Kepulauan Riau	9,91%	NTB & Gorontalo	3,01%
IPM	DKI Jakarta	81,11	Papua	60,62
PDRB per Kapita	DKI Jakarta	274.660	Nusa Tenggara Timur	20.555

Perbandingan nilai ekstrem memperlihatkan bahwa DKI Jakarta memiliki karakteristik yang berbeda dari sebagian besar provinsi, terutama pada kepadatan penduduk, IPM, dan PDRB per kapita. Sebaliknya, beberapa provinsi memiliki nilai yang jauh lebih rendah pada indikator tertentu. Perbedaan tersebut menunjukkan adanya heterogenitas antarprovinsi yang perlu diperhitungkan ketika menganalisis hubungan antarindikator. Keberadaan DKI Jakarta menjadi perhatian khusus karena nilainya pada beberapa indikator berada jauh di atas sebagian besar provinsi lain. Selain perbedaan nilai statistik deskriptif dan matriks korelasi, hasil eksplorasi data juga mengidentifikasi adanya disparitas sosial ekonomi yang tajam antarprovinsi. Provinsi Papua dan Nusa Tenggara Timur (NTT) secara konsisten muncul sebagai wilayah dengan indikator kesejahteraan dan pembangunan ekonomi yang tertinggal dibandingkan rata-rata nasional, seperti tingkat kemiskinan tertinggi dan IPM terendah. Temuan deskriptif ini memiliki implikasi kebijakan yang krusial bagi pemerintah daerah dan pusat dalam merumuskan arah perencanaan pembangunan wilayah yang lebih afirmatif dan tepat sasaran. Alokasi anggaran serta program pengentasan kemiskinan perlu diprioritaskan pada provinsi-provinsi dengan deviasi ekstrem tersebut guna mempercepat pemerataan kesejahteraan sosial di Indonesia.



Gambar 1. Histogram Kepadatan Penduduk dan Indikator Sosial Ekonomi

Berdasarkan Gambar 1, distribusi kepadatan penduduk menunjukkan penyebaran yang tidak merata. Terdapat nilai kepadatan yang jauh lebih tinggi dibandingkan sebagian besar provinsi. Kondisi tersebut terlihat dari perbedaan yang cukup besar antara nilai rata-rata dan median kepadatan penduduk. Variasi juga terlihat pada PDRB per kapita, sedangkan indikator kemiskinan, TPT, dan IPM memiliki rentang yang lebih terbatas.



Gambar 2. Boxplot Kepadatan Penduduk dan Indikator Sosial Ekonomi

Berdasarkan Gambar 2, terlihat adanya perbedaan tingkat penyebaran antarvariabel. Kepadatan penduduk menunjukkan penyebaran yang sangat lebar dibandingkan variabel lainnya. Kondisi tersebut berkaitan dengan adanya provinsi yang memiliki kepadatan sangat tinggi dibandingkan sebagian besar provinsi lainnya.

Untuk menguji hubungan antarvariabel, digunakan korelasi Pearson dan Spearman. Korelasi Pearson digunakan untuk melihat hubungan linear berdasarkan nilai asli variabel, sedangkan korelasi Spearman digunakan untuk melihat hubungan berdasarkan peringkat data. Analisis dilakukan pada seluruh 34 provinsi

dan kembali dilakukan setelah DKI Jakarta dikeluarkan untuk mengetahui sensitivitas hasil terhadap nilai ekstrem.

Tabel 3. Hasil Korelasi Kepadatan Penduduk dengan Indikator Sosial Ekonomi

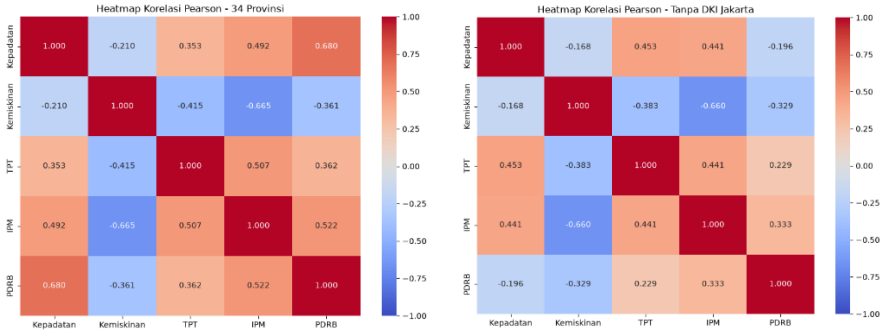
Hubungan	Pearson	p-value	Spearman	p-value
Kepadatan–Kemiskinan	-0,210	0,233	-0,167	0,344
Kepadatan–TPT	0,353	0,041	0,329	0,057
Kepadatan–IPM	0,492	0,003	0,559	<0,001
Kepadatan–PDRB	0,680	<0,001	-0,185	0,294

Keterangan: p-value < 0,05 menunjukkan hubungan signifikan secara statistik. CI = confidence interval 95%. Berdasarkan Tabel 3, hubungan kepadatan penduduk dengan kemiskinan menunjukkan arah negatif, tetapi tidak signifikan berdasarkan Pearson maupun Spearman. Hubungan kepadatan penduduk dengan TPT menunjukkan korelasi Pearson positif dan signifikan, sedangkan korelasi Spearman tidak signifikan. Hubungan kepadatan penduduk dengan IPM menunjukkan arah positif dan signifikan berdasarkan kedua metode. Sementara itu, hubungan kepadatan penduduk dengan PDRB per kapita menunjukkan perbedaan paling besar antara Pearson dan Spearman. Pearson menghasilkan korelasi positif sebesar 0,680 dan signifikan secara statistik ($p < 0,001$), sedangkan Spearman menghasilkan korelasi negatif sebesar -0,185 dan tidak signifikan ($p = 0,294$). Perbedaan tersebut menunjukkan bahwa hubungan kepadatan penduduk dan PDRB per kapita sangat sensitif terhadap karakteristik distribusi data dan keberadaan nilai ekstrem.

Tabel 4. Hasil Korelasi Setelah DKI Jakarta Dikeluarkan

Hubungan	Pearson	p-value	Spearman	p-value
Kepadatan–Kemiskinan	-0,168	0,349	-0,098	0,589
Kepadatan–TPT	0,453	0,008	0,271	0,128
Kepadatan–IPM	0,441	0,010	0,517	0,002
Kepadatan–PDRB	-0,196	0,274	-0,296	0,094

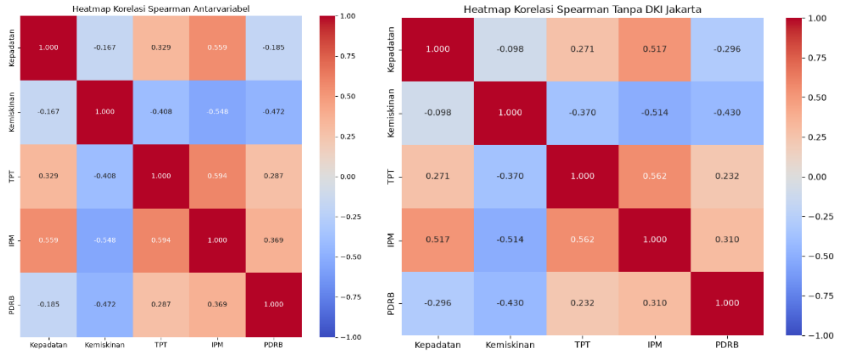
Keterangan: p-value < 0,05 menunjukkan hubungan signifikan secara statistik. CI = confidence interval 95%. Setelah DKI Jakarta dikeluarkan, beberapa hubungan menunjukkan perubahan dibandingkan hasil pada seluruh 34 provinsi. Hubungan kepadatan penduduk dengan kemiskinan tetap menunjukkan arah negatif dan tidak signifikan. Pada TPT, korelasi Pearson meningkat dari 0,353 menjadi 0,453 dan tetap signifikan, sedangkan Spearman berubah menjadi 0,271 dan tidak signifikan. Hubungan kepadatan penduduk dengan IPM tetap positif dan signifikan pada Pearson maupun Spearman, masing-masing sebesar 0,441 dan 0,517. Perubahan paling besar terlihat pada PDRB per kapita. Korelasi Pearson berubah dari 0,680 menjadi -0,196 dan menjadi tidak signifikan, sedangkan korelasi Spearman berubah dari -0,185 menjadi -0,296 dan tetap tidak signifikan. Hasil tersebut menunjukkan bahwa hubungan positif antara kepadatan penduduk dan PDRB per kapita pada analisis seluruh provinsi sangat dipengaruhi oleh keberadaan DKI Jakarta sebagai pengamatan ekstrem.



Gambar 3. Perbandingan Heatmap Pearson

Berdasarkan Gambar 3, terlihat adanya perubahan intensitas warna pada beberapa pasangan variabel setelah DKI Jakarta dikeluarkan dari analisis. Hubungan kepadatan penduduk dengan PDRB per kapita yang sebelumnya ditunjukkan oleh warna merah (korelasi positif) berubah menjadi biru muda (korelasi negatif). Sementara itu, hubungan kepadatan penduduk dengan IPM tetap didominasi warna merah, menunjukkan bahwa arah hubungan positif masih konsisten. Pada hubungan dengan kemiskinan, warna biru tetap terlihat meskipun intensitasnya berkurang, sedangkan hubungan dengan TPT tetap ditunjukkan oleh warna merah

dengan perubahan intensitas yang relatif kecil. Secara visual, perubahan warna paling mencolok terjadi pada pasangan kepadatan penduduk dan PDRB per kapita.



Gambar 4. Perbandingan Heatmap Spearman

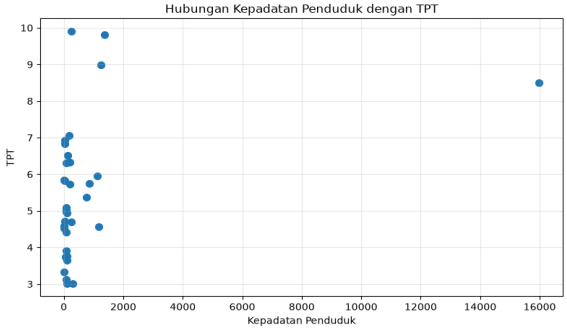
Pada Gambar 4, pola warna sebelum dan sesudah DKI Jakarta dikeluarkan cenderung lebih stabil dibandingkan heatmap Pearson. Hubungan kepadatan penduduk dengan IPM tetap ditunjukkan oleh warna merah yang cukup kuat, sedangkan hubungan dengan kemiskinan tetap berwarna biru dengan intensitas yang rendah. Warna pada hubungan kepadatan penduduk dengan TPT masih menunjukkan korelasi positif meskipun sedikit memudar setelah DKI Jakarta dikeluarkan. Sementara itu, hubungan kepadatan penduduk dengan PDRB per kapita tetap berada pada spektrum warna biru, yang menunjukkan arah hubungan negatif. Secara keseluruhan, perubahan visual pada heatmap Spearman tidak terlalu besar, sehingga pola hubungan antarvariabel terlihat lebih konsisten.

Untuk memperlihatkan pola hubungan antara kepadatan penduduk dan masing-masing indikator sosial ekonomi secara lebih jelas, digunakan scatter plot. Visualisasi ini menunjukkan persebaran setiap provinsi berdasarkan dua variabel yang dibandingkan.



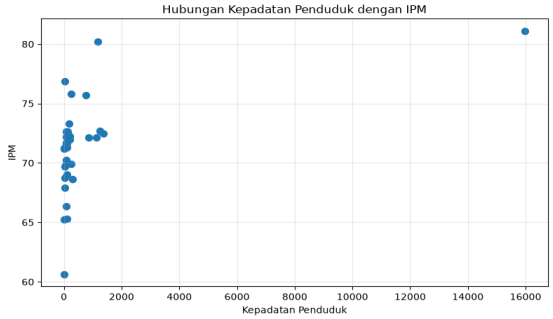
Gambar 5. Hubungan Kepadatan Penduduk dengan Kemiskinan

Berdasarkan Gambar 5, Scatter plot menunjukkan kecenderungan hubungan negatif antara kepadatan penduduk dan kemiskinan. Hubungan tersebut relatif lemah dan tidak signifikan berdasarkan hasil pengujian korelasi.



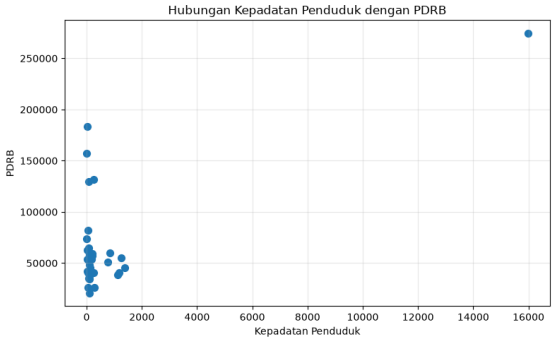
Gambar 6. Hubungan Kepadatan Penduduk dengan TPT

Berdasarkan Gambar 6, Scatter plot menunjukkan kecenderungan hubungan positif antara kepadatan penduduk dan TPT. Hasil pengujian menunjukkan bahwa signifikansi hubungan bergantung pada metode korelasi yang digunakan.



Gambar 7. Hubungan Kepadatan Penduduk dengan IPM

Gambar 7 menunjukkan kecenderungan hubungan positif antara kepadatan penduduk dan IPM. Hasil Pearson dan Spearman sama-sama menunjukkan hubungan yang signifikan.



Gambar 8. Hubungan Kepadatan Penduduk dengan PDRB per Kapita

Berdasarkan Gambar 8, Scatter plot menunjukkan pola positif ketika seluruh 34 provinsi digunakan. Namun, pola tersebut berubah setelah DKI Jakarta dikeluarkan. Hasil tersebut menunjukkan adanya pengaruh nilai ekstrem terhadap hubungan yang diamati.

Hasil analisis korelasi dalam penelitian ini menunjukkan koefisien korelasi antara tingkat kemiskinan dan Indeks Pembangunan Manusia (IPM) sebesar -0,210 (dengan Pearson) dan -0,167 (dengan Spearman) pada seluruh 34 provinsi, yang tidak signifikan secara statistik. Temuan ini melengkapi studi terdahulu oleh Azrai et al. yang menyoroti pentingnya keterkaitan variabel pembangunan manusia, di mana penelitian ini memperjelas bahwa hubungan linier langsung antara kepadatan penduduk dan indikator kesejahteraan di tingkat provinsi cenderung sangat lemah atau tidak signifikan secara statistik apabila dipengaruhi oleh nilai ekstrem seperti DKI Jakarta.

PEMBAHASAN

Hasil penelitian menunjukkan bahwa kepadatan penduduk memiliki distribusi yang sangat tidak merata antarprovinsi. Perbedaan karakteristik tersebut menjadi faktor penting dalam interpretasi hubungan antara kepadatan penduduk dan indikator sosial ekonomi. Keberadaan DKI Jakarta sebagai provinsi dengan karakteristik ekstrem membuat hasil korelasi perlu ditinjau menggunakan lebih dari satu pendekatan.[3] [2] Perbandingan Pearson dan Spearman menunjukkan bahwa hubungan kepadatan penduduk dengan IPM merupakan pola yang paling konsisten. Hubungan tersebut tetap positif dan signifikan pada kedua metode, baik ketika seluruh provinsi dianalisis maupun ketika DKI Jakarta dikeluarkan. Temuan ini menunjukkan bahwa hubungan positif antara kepadatan penduduk dan IPM relatif tidak berubah akibat penghapusan satu pengamatan ekstrem. Namun, hasil tersebut tetap menunjukkan hubungan statistik dan tidak dapat diinterpretasikan sebagai hubungan sebab-akibat.

Hasil penelitian sejalan dengan penelitian Azrai et al[5] . yang menunjukkan adanya hubungan positif dan signifikan antara kepadatan penduduk dan Indeks Pembangunan Manusia (IPM). Meskipun demikian, hasil kedua penelitian tidak dapat dibandingkan secara langsung berdasarkan nilai koefisien karena metode analisis yang digunakan berbeda. Penelitian ini menggunakan korelasi Pearson dan Spearman pada data 34 provinsi tahun 2021, sedangkan Azrai et al. menggunakan regresi linear berganda dengan data periode 2015-2024. Perbedaan periode pengamatan, unit analisis, dan metode statistik tersebut dapat menghasilkan ukuran hubungan yang berbeda. Kesamaan arah hubungan menunjukkan bahwa kepadatan

penduduk dan IPM memiliki pola keterkaitan yang positif dalam kedua penelitian, tetapi hasil tersebut tetap perlu dipahami sesuai dengan karakteristik data dan metode analisis yang digunakan.

Berbeda dengan IPM, hubungan kepadatan penduduk dengan PDRB per kapita menunjukkan sensitivitas yang tinggi terhadap keberadaan DKI Jakarta. Ketika seluruh 34 provinsi digunakan, korelasi Pearson menunjukkan hubungan positif sebesar 0,680 dan signifikan secara statistik ($p < 0,001$). Namun, korelasi Spearman menunjukkan hubungan negatif sebesar -0,185 dan tidak signifikan ($p = 0,294$). Setelah DKI Jakarta dikeluarkan, korelasi Pearson berubah menjadi -0,196 dan tidak signifikan ($p = 0,274$), sedangkan Spearman menjadi -0,296 dan tetap tidak signifikan ($p = 0,094$). Perubahan arah korelasi Pearson dari positif menjadi negatif menunjukkan bahwa hubungan yang terlihat pada analisis seluruh provinsi sangat dipengaruhi oleh nilai ekstrem DKI Jakarta. Dengan demikian, hasil tersebut menunjukkan pentingnya pemeriksaan sensitivitas terhadap nilai ekstrem ketika menganalisis hubungan antarprovinsi.Perbedaan antara Pearson dan Spearman juga menunjukkan pentingnya mempertimbangkan bentuk data ketika menginterpretasikan korelasi. Pearson menggunakan nilai asli variabel sehingga lebih sensitif terhadap pengamatan ekstrem. Spearman menggunakan peringkat sehingga perubahan akibat nilai ekstrem dapat lebih terbatas. Dalam penelitian ini, perbandingan kedua metode memberikan informasi tambahan mengenai kestabilan pola hubungan yang tidak dapat diperoleh jika hanya menggunakan satu ukuran korelasi.

Hasil scatter plot mendukung temuan tersebut. Hubungan kepadatan penduduk dengan kemiskinan menunjukkan pola yang lemah, sedangkan hubungan dengan TPT dan IPM cenderung positif. Pada hubungan dengan PDRB per kapita, posisi DKI Jakarta terlihat berbeda dari sebagian besar provinsi sehingga dapat memengaruhi arah hubungan ketika seluruh data dianalisis. Oleh karena itu, pemeriksaan visual terhadap data ekstrem perlu dilakukan sebelum menarik kesimpulan berdasarkan koefisien korelasi. Secara keseluruhan, hasil penelitian menunjukkan bahwa hubungan antarindikator sosial ekonomi tidak bersifat seragam. Beberapa hubungan relatif konsisten setelah pengamatan ekstrem dikeluarkan, sedangkan hubungan lainnya berubah secara substansial. Temuan ini memperlihatkan bahwa analisis antarprovinsi perlu mempertimbangkan heterogenitas wilayah dan sensitivitas ukuran statistik terhadap nilai ekstrem.

Temuan penelitian memiliki implikasi bagi perencanaan pembangunan daerah. Perbedaan kepadatan penduduk dan indikator sosial ekonomi antarprovinsi menunjukkan bahwa kebijakan pembangunan perlu mempertimbangkan karakteristik masing-masing wilayah. Provinsi dengan tingkat kemiskinan yang tinggi memerlukan perhatian pada penguatan akses terhadap pendidikan, kesehatan, dan kesempatan ekonomi, sedangkan wilayah dengan PDRB per kapita yang relatif rendah memerlukan strategi peningkatan produktivitas dan aktivitas ekonomi daerah. Pada wilayah dengan kepadatan penduduk yang sangat tinggi, seperti DKI Jakarta, perencanaan pembangunan juga perlu mempertimbangkan tekanan terhadap layanan publik, perumahan, transportasi, dan infrastruktur. Namun, hasil korelasi dalam penelitian ini bersifat eksploratif sehingga tidak dapat digunakan untuk menetapkan hubungan sebab-akibat secara langsung. Temuan ini lebih tepat digunakan sebagai informasi awal untuk mengidentifikasi perbedaan karakteristik wilayah dan menentukan aspek yang memerlukan kajian lebih lanjut.

KESIMPULAN

Berdasarkan hasil Exploratory Data Analysis terhadap 34 provinsi di Indonesia tahun 2021, penelitian ini menunjukkan adanya variasi yang besar pada kepadatan penduduk dan indikator sosial ekonomi antarprovinsi. Kepadatan penduduk memiliki penyebaran yang sangat tidak merata dan menunjukkan keberadaan nilai ekstrem, terutama pada DKI Jakarta. Kondisi tersebut perlu diperhatikan karena dapat memengaruhi hasil analisis hubungan antarvariabel.

Hasil korelasi menunjukkan bahwa hubungan kepadatan penduduk dengan kemiskinan relatif lemah dan tidak signifikan. Hubungan dengan TPT menunjukkan arah positif, tetapi signifikansinya berbeda antara Pearson dan Spearman. Hubungan dengan IPM menunjukkan pola yang paling konsisten karena tetap positif dan signifikan pada kedua metode, baik ketika seluruh provinsi dianalisis maupun setelah DKI Jakarta dikeluarkan. Sebaliknya, hubungan kepadatan penduduk dengan PDRB per kapita menunjukkan sensitivitas yang tinggi terhadap DKI Jakarta. Ketika seluruh provinsi digunakan, Pearson menghasilkan hubungan positif yang signifikan. Namun, hubungan tersebut berubah menjadi negatif dan tidak signifikan setelah

DKI Jakarta dikeluarkan. Hasil Spearman juga tidak menunjukkan hubungan yang signifikan. Temuan ini menunjukkan bahwa hubungan positif antara kepadatan penduduk dan PDRB per kapita pada seluruh provinsi tidak cukup stabil dan sangat dipengaruhi oleh keberadaan pengamatan ekstrem.

Secara keseluruhan, penelitian menunjukkan bahwa EDA dapat digunakan untuk mengidentifikasi variasi, pola distribusi, nilai ekstrem, dan hubungan antarindikator sosial ekonomi. Perbandingan Pearson dan Spearman serta analisis dengan dan tanpa DKI Jakarta memberikan informasi mengenai sensitivitas hasil korelasi. Dengan demikian, interpretasi hubungan antarprovinsi perlu mempertimbangkan distribusi data, keberadaan nilai ekstrem, dan metode korelasi yang digunakan. Hasil penelitian bersifat eksploratif dan tidak dapat digunakan untuk menyimpulkan hubungan sebab-akibat. Penelitian selanjutnya dapat menggunakan data beberapa tahun dan metode statistik yang lebih mendalam untuk menguji faktor-faktor yang berkaitan dengan perbedaan kondisi kependudukan dan sosial ekonomi antarprovinsi.

REFERENCES

[1] B. P. Statistik, "Statistik Indonesia 2022," Jakarta. Accessed: Oct. 11, 2026. [Online]. Available: https://www.bps.go.id/id/publication/2022/02/25/0a57b3b1f4a4758fa28f7dcb/statistik-indonesia-2022.html

[2] B. P. Statistik, "Penduduk, Laju Pertumbuhan Penduduk, Distribusi Persentase Penduduk, Kepadatan Penduduk, Rasio Jenis Kelamin Penduduk Menurut Provinsi, 2021," Jakarta. Accessed: Oct. 11, 2026. [Online]. Available: https://www.bps.go.id/id/statistics-table/3/V1ZSbFRUY3lTbFpEYTNsVWNGcDZjek53YkhsNFFUMDkjMw==/penduduk--laju-pertumbuhan-penduduk--distribusi-persentase-penduduk--kepadatan-penduduk--rasio-jenis-kelamin-penduduk-menurut-provinsi.html?year=2021

[3] L. Agustina, Irmeilyana, S. I. Maiyanti, and Ngudiantoro., "Perbandingan Biplot Indikator Kependudukan di Indonesia Tahun 2020, 2023, dan 2024 dengan Menggunakan Analisis Procrustes," *Jurnal Penelitian Sains*, vol. 28, no. 1, pp. 1–8, Apr. 2026, doi: 10.56064/jps.v28i1.1314.

[4] E. P. Permatasari, L. A. Iriani, and E. Widodo, "Identification of Indonesian Provinces Based on Socioeconomic Indicators in 2024 Using K-Means," *Jurnal Sintak*, vol. 4, no. 2, pp. 57–70, 2026, doi: 10.62375/jsintak.v4i2.789.

[5] I. Azrai, Budimansyah, and T. Rahman, "Pengaruh Tingkat Kepadatan Penduduk, Kemiskinan Dan Air Bersih Terhadap Indeks Pembangunan Manusia Di Indonesia," *Jurnal Program Studi Pendidikan Ekonomi*, vol. 14, no. 2, pp. 650–667, 2026, [Online]. Available: https://doi.org/10.24127/jp

[6] M. H. Sinambela and I. W. Wenagama, "Determinan Tingkat Pengangguran Terbuka (Tpt) Kabupaten/Kota Di Provinsi Bali," *E-Jurnal EP Unud*, vol. 12, no. 3, pp. 199–218, 2023.

[7] F. Nurhijah, A. Alamudi, and B. Susetyo, "Identifikasi Faktor yang Memengaruhi Tingkat Pengangguran Terbuka di Jawa Barat Menggunakan Regresi Data Panel," *Xplore: Journal of Statistics*, vol. 12, no. 3, pp. 230–241, Dec. 2023, doi: 10.29244/xplore.v12i3.1173.

[8] I. V. D. Srihith, A. D. Donald, T. A. S. Srinivas, G. Thippanna, and P. V. Lakshmi, "Exploratory Data Analysis on Autopilot: Python's Automatic Solutions," *Recent Trends in Androids and IOS Applications*, vol. 5, no. 3, pp. 20–26, 2023.

[9] M. Rahmany, A. M. Zin, and E. A. Sundararajan, "Comparing Tools Provided By Python And R For Exploratory Data Analysis," *International Journal Information System and Computer Science (IJISCS)*, vol. 4, no. 3, pp. 131–142, 2020.

[10] B. P. Statistik, "Jumlah dan Persentase Penduduk Miskin Menurut Provinsi, 2021," jakarta, 2021. Accessed: Oct. 11, 2026. [Online]. Available: https://www.bps.go.id/id/statistics-table/1/ZUNac1lXUTJOM2h6VmtwUWRrMXNNV0U1YW1KcVlVazk=/jumlah-dan-persentase-penduduk-miskin-menurut-provinsi.html

penduduk-miskin-menurut-provinsi.html)

[11] B. P. Statistik, "[REVISI per 09/11/2021]Tingkat Pengangguran Terbuka (TPT) sebesar 6,49 persen," Jakarta, 2021. Accessed: Oct. 11, 2026. [Online]. Available: https://www.bps.go.id/id/pressrelease/2021/11/05/1857/-revisi-per-09-11-2021--agustus-2021-tingkat-pengangguran-terbuka-tpt-sebesar-6-49-persen.html [cite: 7]

[12] A. Pamuji, E. P. Setiawan, and A. M. Munir, "Regresi Data Panel dalam Analisis Faktor-Faktor yang Mempengaruhi Tingkat Pengangguran Terbuka di Provinsi Daerah Istimewa Yogyakarta Tahun 2017-2021," *Indonesian Journal of Applied Statistics*, vol. 6, no. 2, pp. 162–171, Apr. 2024, doi: 10.13057/ijas.v6i2.80010.

[13] A. T. Romdhani, P. N. Manik, N. Nabilah, A. Mafiroh, F. Nabila, and R. D. Oktaviyuda, "Open Unemployment Rate in 35 Indonesian Provinces During Covid-19 Affected by Population Growth Rate per Year, School Participation Rate, Life Expectancy Rate, and Classification of Javanese-Non-Java Regions," *Indonesian Journal of Development and Economics*, vol. 1, no. 1, pp. 20–33, 2025, Accessed: Oct. 11, 2026. [Online]. Available: [https://journal.karyarisetnusantara.com

[14] B. P. Statistik, "Indeks Pembangunan Manusia 2021," Jakarta, 2021. Accessed: Oct. 11, 2026. [Online]. Available: https://www.bps.go.id/id/publication/2022/05/11/48b6466dcf14b562df9f17e2/indeks-pembangunan-manusia-2021.html

[15] B. P. Statistik, "Produk Domestik Regional Bruto KabupatenKota di Indonesia 2017-2021," Jakarta, 2021. Accessed: Oct. 11, 2026. [Online]. Available: [https://www.bps.go.id/id/publication/2022/10/07/==28bc38029d20c3261904a29a/produk-domestik-regional-bruto-kabupaten-kota-di-indonesia-2017-2021.html](https://www.bps.go.id/id/publication/2022/10/07/28bc38029d20c3261904a29a/produk-domestik-regional-bruto-kabupaten-kota-di-indonesia-2017-2021.html)

[16] F. A. Maresti, M. A. Rahman, and S. N. Anisah, "Analisis Anova Two-Way: Pengaruh Kesehatan Mental Dan Tahun Akademik Terhadap Nilai Ipk Mahasiswa Dengan Bahasa Pemrograman Python," *Data Sciences Indonesia (DSI)*, vol. 4, no. 2, pp. 126–136, Feb. 2025, doi: 10.47709/dsi.v4i2.5403.