



TREN KUALITAS UDARA DUNIA DAN KAITANNYA DENGAN PEMBANGUNAN EKONOMI SERTA PERTUMBUHAN PENDUDUK PADA TAHUN 2010-2019

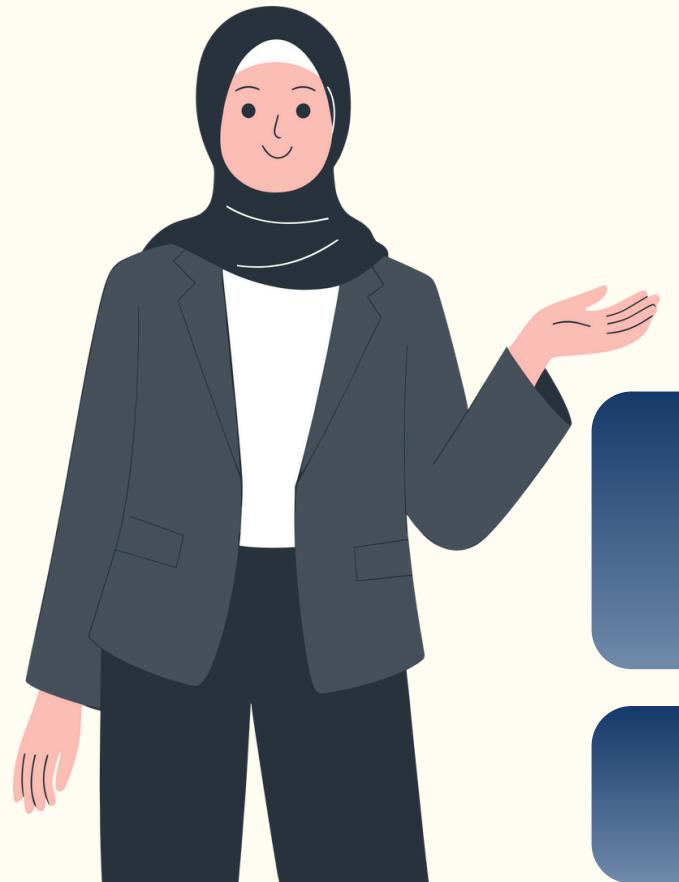
KELOMPOK 6



Anggota Kelompok

**Nagatan Alief Putra
Silahen**

24031554086



Khansa Nadhifa

1314622032

Latar Belakang

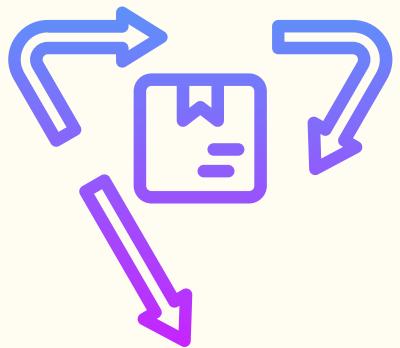
Kualitas udara adalah isu lingkungan yang semakin mendapatkan perhatian serius karena memiliki dampak langsung terhadap kesehatan manusia yang apabila ditelusuri lebih lanjut memiliki dampak terhadap ekonomi. Partikulat halus seperti PM2.5 (partikel dengan diameter $\leq 2,5$ mikrometer) merupakan polutan paling berbahaya karena dapat menembus jauh ke dalam paru-paru dan aliran darah, menyebabkan berbagai penyakit pernapasan dan kardiovaskular seperti asma, penyakit jantung, stroke, dan bahkan kematian dini.

Beberapa penelitian telah menunjukkan hubungan antara tingkat pendapatan GPD per kapita, kepadatan populasi, dan polusi udara. Penelitian yang telah dilakukan oleh (Kim et al., 2021) menemukan bahwa peningkatan GDP per kapita secara statistik berkorelasi dengan penurunan konsentrasi PM2.5, sementara kepadatan populasi (*population density*) berkorelasi dengan peningkatan PM2.5.

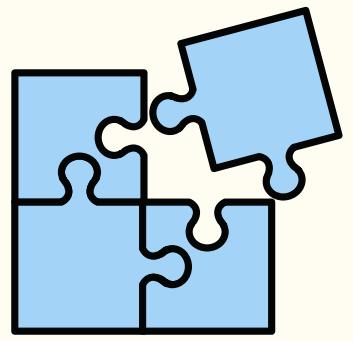
Berdasarkan temuan-temuan tersebut, sangat relevan untuk mengeksplorasi keterkaitan antara kualitas udara (PM2.5), pembangunan ekonomi (GDP per kapita), dan pertumbuhan penduduk dalam skala global. Dengan analisis data wrangling yang menggabungkan data polusi udara, GDP per kapita, dan populasi dari banyak negara



Tujuan Penelitian



Menerapkan teknik wrangling pada dataset PM2.5, GDP, dan populasi menjadi satu dataset yang siap dianalisis.

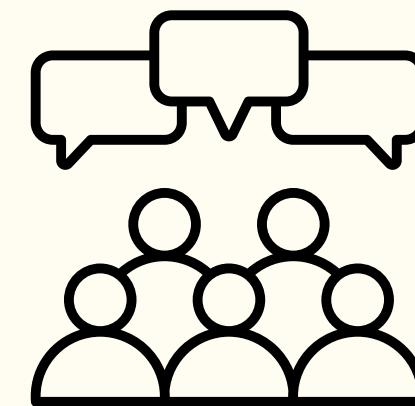


Mengukur hubungan antara kualitas udara (PM2.5) dengan pembangunan ekonomi (GDP per kapita) untuk mengetahui apakah peningkatan ekonomi berpengaruh pada tingkat polusi udara.

Manfaat Penelitian

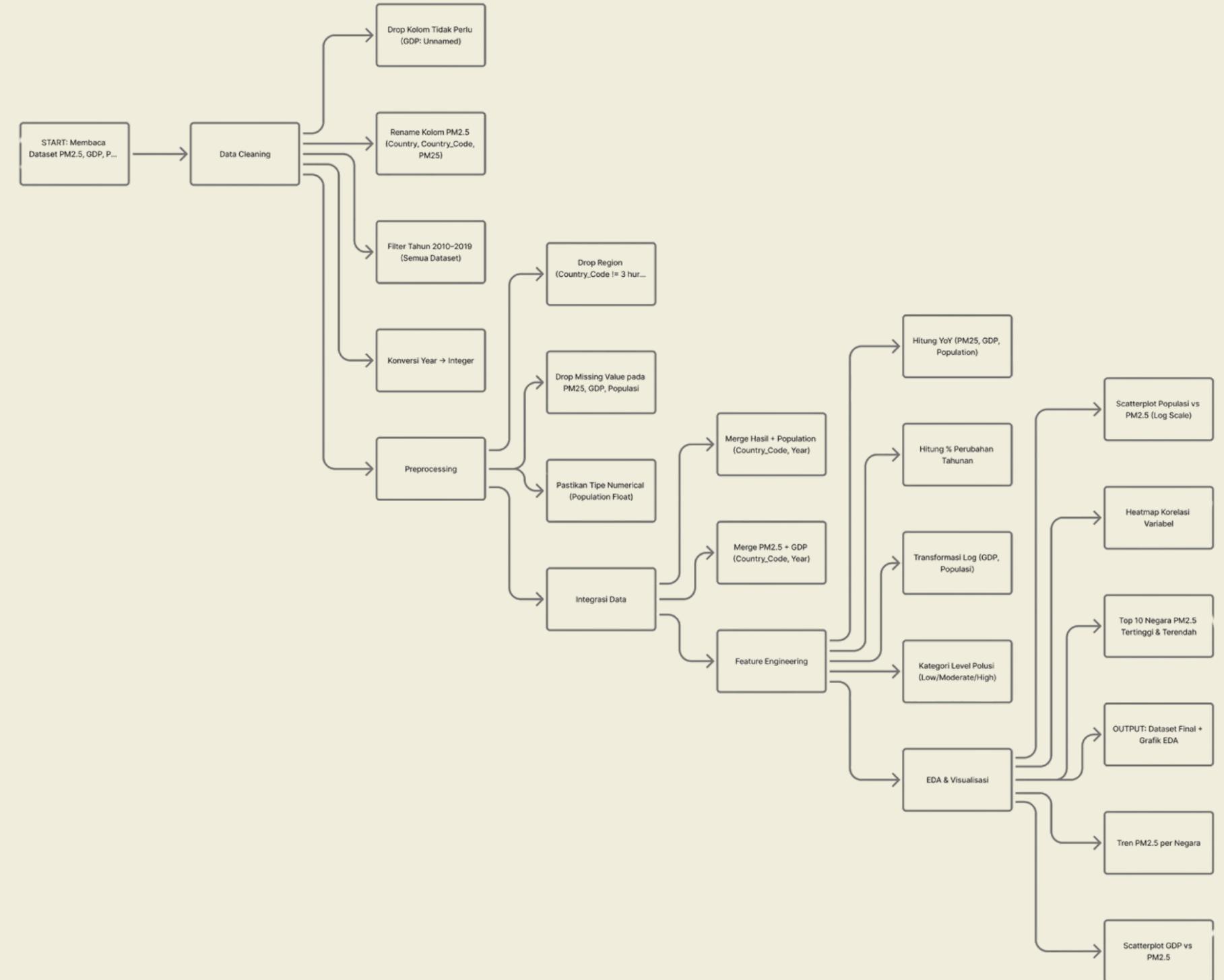


Menghasilkan dataset terintegrasi PM2.5, GDP, dan Populasi yang bersih dan siap digunakan untuk analisis lanjutan juga menunjukkan pentingnya pengolahan data (wrangling), integrasi, dan visualisasi dalam memecahkan masalah global



Memberikan gambaran bagi pemerintah atau lembaga lingkungan tentang negara atau wilayah yang memiliki tren polusi tinggi dan faktor apa yang mungkin memengaruhinya juga dasar pertimbangan dalam merancang kebijakan pembangunan berkelanjutan yang mempertimbangkan aspek ekonomi dan lingkungan.

Diagram Alur





Dataset

Dataset yang digunakan dalam penelitian ini terdiri dari tiga sumber utama. Data PM2.5 diperoleh dari Our World in Data yang menyediakan informasi tahunan mengenai tingkat polusi udara PM2.5 berbagai negara. Data GDP per kapita diunduh dari World Bank API, berisi nilai GDP per kapita dalam US Dollar sebagai indikator tingkat pembangunan ekonomi setiap negara. Sementara itu, data jumlah penduduk diperoleh dari DataHub sebagai mirror dari dataset resmi UN/World Bank, yang memuat total populasi tiap negara per tahun. Ketiga dataset tersebut kemudian digabungkan berdasarkan negara dan tahun untuk mendukung analisis hubungan antara kualitas udara, kondisi ekonomi, dan pertumbuhan penduduk.

Proses Wrangling

Cleaning

Pada tahap data cleaning, masing-masing dataset dibersihkan dengan mengganti nama kolom agar konsisten, menghapus data yang tidak relevan, memfilter rentang tahun 2010–2019, dan mengatur ulang format tabel (misalnya GDP yang awalnya berbentuk wide diubah menjadi long).

Pre-Processing

Tahap pre-processing dilakukan untuk menstandarisasi struktur data, seperti memastikan Country_Code valid, mengonversi tipe tahun menjadi integer, menghapus data non-negara, dan menyimpan kembali dataset yang telah siap diproses lebih lanjut.

Proses Wrangling

Integrasi Data

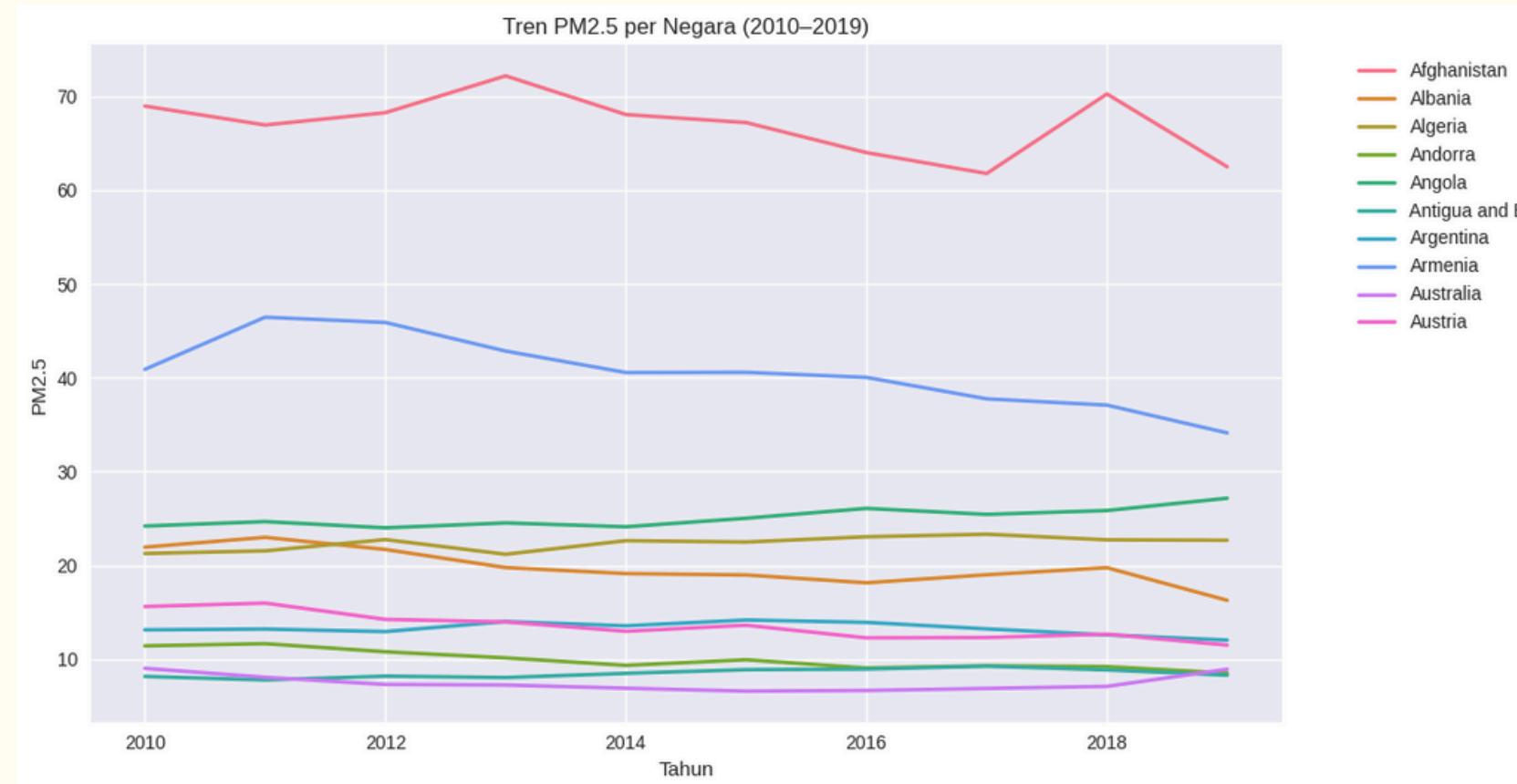
Tahap berikutnya adalah integrasi data, yaitu menggabungkan ketiga dataset menggunakan kolom Country_Code dan Year. Proses merge dilakukan dua kali: pertama untuk menggabungkan PM2.5 dengan GDP, kemudian digabungkan lagi dengan populasi. Setelah seluruh data menyatu, baris yang masih memiliki nilai kosong pada kolom penting dihapus sehingga menghasilkan dataset final yang lengkap.

Feature Engineering

Tahap terakhir adalah feature engineering, yaitu menambahkan fitur-fitur baru untuk mendukung analisis. Langkah ini mencakup menghapus kolom duplikat, mengurutkan data berdasarkan negara dan tahun, menghitung perubahan tahun-ke-tahun (YoY) dan persentase perubahan, mengelompokkan tingkat polusi menjadi kategori, serta melakukan transformasi log pada GDP dan populasi agar distribusinya lebih stabil. Seluruh hasil engineering kemudian disimpan sebagai dataset baru untuk digunakan pada tahap analisis selanjutnya.

Exploratory Data Analysis

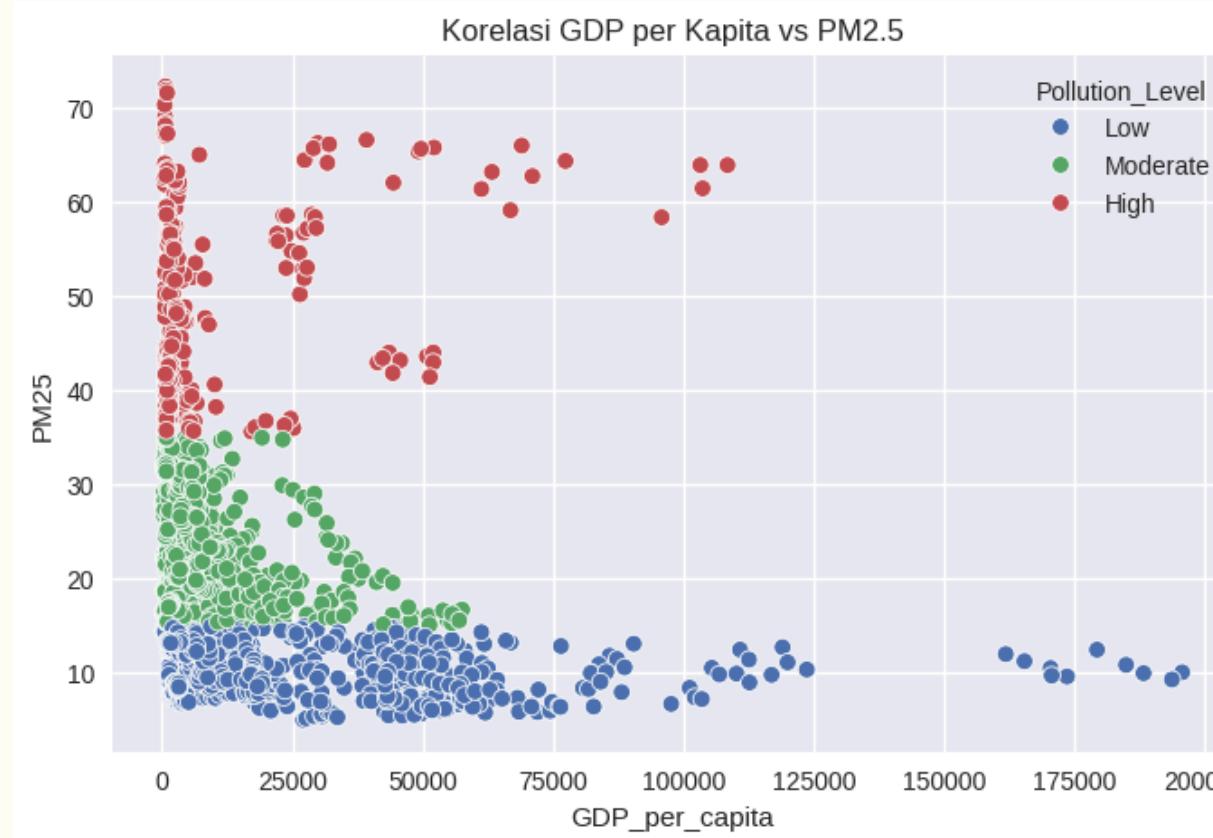
Tren Global PM2.5, GDP, dan Populasi



Grafik tren PM2.5 per negara dari tahun 2010 hingga 2019 digunakan untuk melihat perubahan kualitas udara di berbagai negara selama satu dekade. Setiap garis pada grafik mewakili satu negara, sehingga pola peningkatan atau penurunan polusi dapat terlihat dengan jelas. Dari visualisasi ini, beberapa negara tampak berhasil menurunkan tingkat polusinya secara konsisten, sementara negara lain masih menunjukkan fluktuasi atau bahkan kenaikan PM2.5 pada tahun tertentu. Grafik ini memberikan gambaran menyeluruh mengenai dinamika polusi udara antarnegara dan memudahkan perbandingan tren kualitas udara di tingkat global.

Exploratory Data Analysis

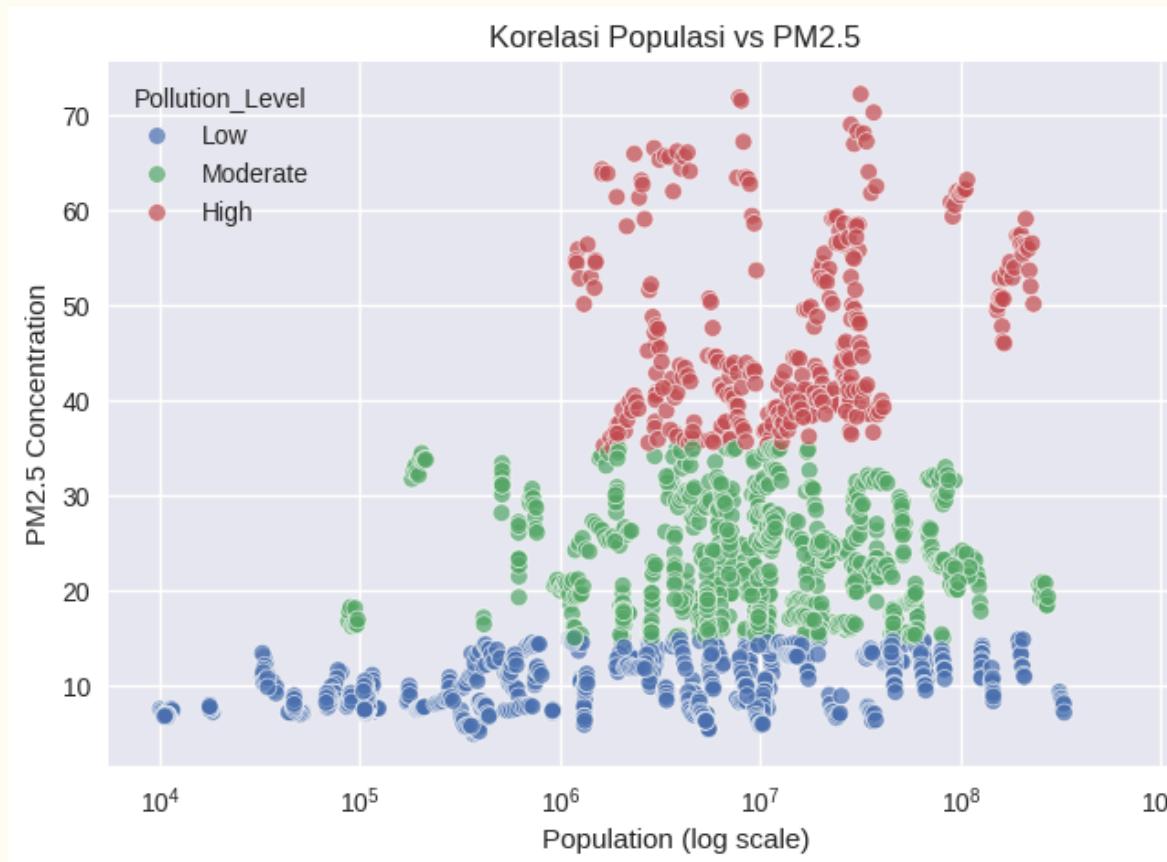
Korelasi GDP vs PM2.5



Grafik menunjukkan bahwa wilayah dengan GDP per kapita tinggi umumnya memiliki tingkat PM2.5 yang lebih rendah, terlihat dari banyaknya titik biru (kategori polusi rendah) pada rentang GDP tinggi. Sebaliknya, wilayah dengan GDP per kapita rendah lebih sering berada pada kategori polusi sedang hingga tinggi, ditandai oleh dominasi titik hijau dan merah pada GDP yang rendah. Pola ini menggambarkan bahwa peningkatan kesejahteraan ekonomi biasanya berkaitan dengan kualitas udara yang lebih baik, kemungkinan karena adanya investasi pada teknologi bersih, regulasi lingkungan, serta infrastruktur yang lebih efisien dalam mengendalikan polusi.

Exploratory Data Analysis

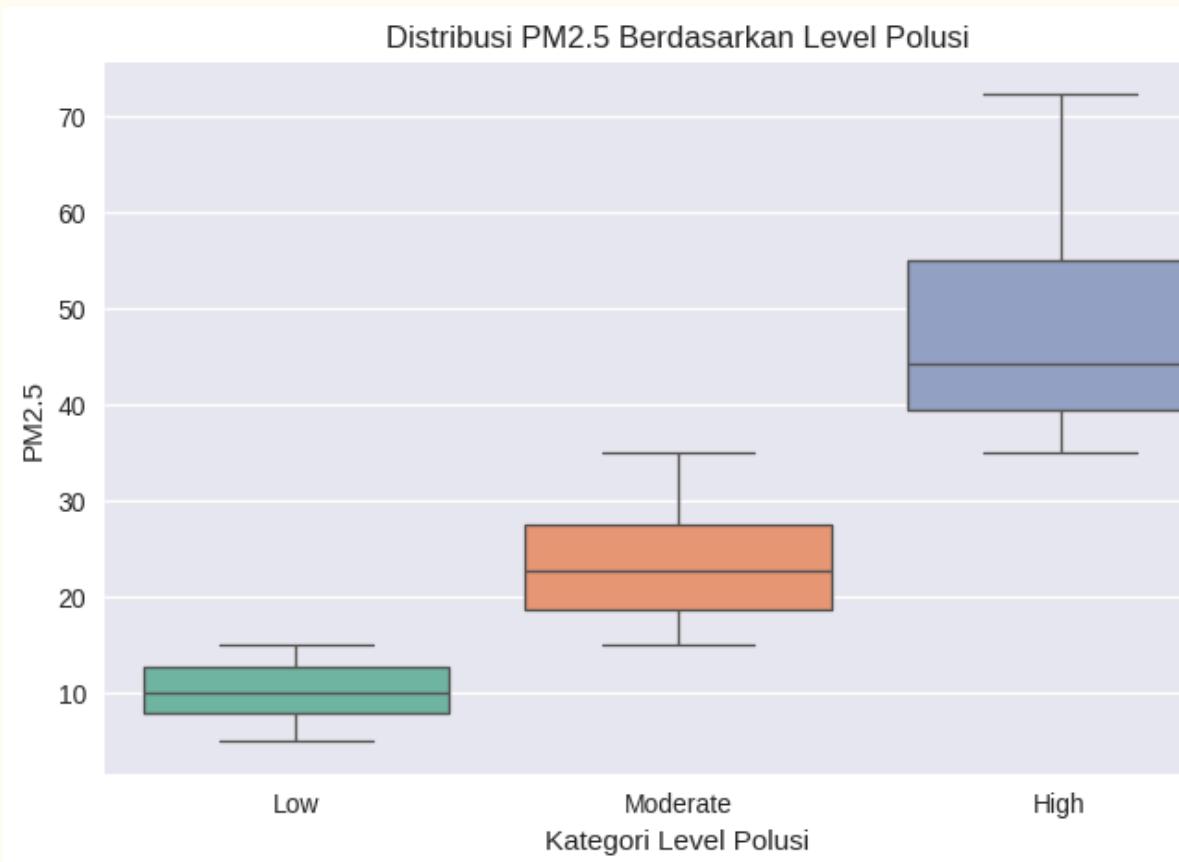
Korelasi Population vs PM2.5



Scatterplot hubungan antara populasi & PM2.5 diatas kami gunakan untuk melihat kecenderungan apakah negara dengan jumlah penduduk yang besar memiliki tingkat polusi udara lebih tinggi. Karena populasi tiap negara berbeda jauh skala angkanya, grafik menggunakan skala logaritmik agar pola penyebaran data lebih terlihat. Visualisasi ini menunjukkan apakah pertumbuhan penduduk berkaitan dengan peningkatan PM2.5, sekaligus mengungkap pola khusus seperti negara kecil dengan polusi tinggi atau negara besar yang mampu menjaga tingkat polusi tetap rendah. Dengan demikian, grafik ini membantu memahami peran populasi terhadap kualitas udara secara lebih menyeluruh.

Exploratory Data Analysis

Bar plot negara dengan polusi tertinggi



Grafik boxplot distribusi PM2.5 berdasarkan level polusi diatas kami gunakan untuk membandingkan sebaran nilai PM2.5 pada tiga kategori yaitu Low, Moderate, dan High. Grafik ini memperlihatkan bagaimana setiap kategori memiliki median, rentang nilai, dan outlier yang berbeda. Kelompok Low menunjukkan nilai PM2.5 yang rendah dan cenderung stabil, sedangkan kategori Moderate memiliki sebaran yang lebih lebar dengan nilai tengah yang lebih tinggi. Pada kategori High, terlihat bahwa nilai PM2.5 jauh lebih tinggi serta memiliki rentang yang lebih besar, menunjukkan variasi polusi yang signifikan antar negara dalam kelompok ini. Melalui visualisasi ini, perbedaan level polusi antar kategori dapat dengan jelas terlihat, sehingga mempermudah pemahaman mengenai seberapa jauh kualitas udara bervariasi antar kelompok negara.

Exploratory Data Analysis

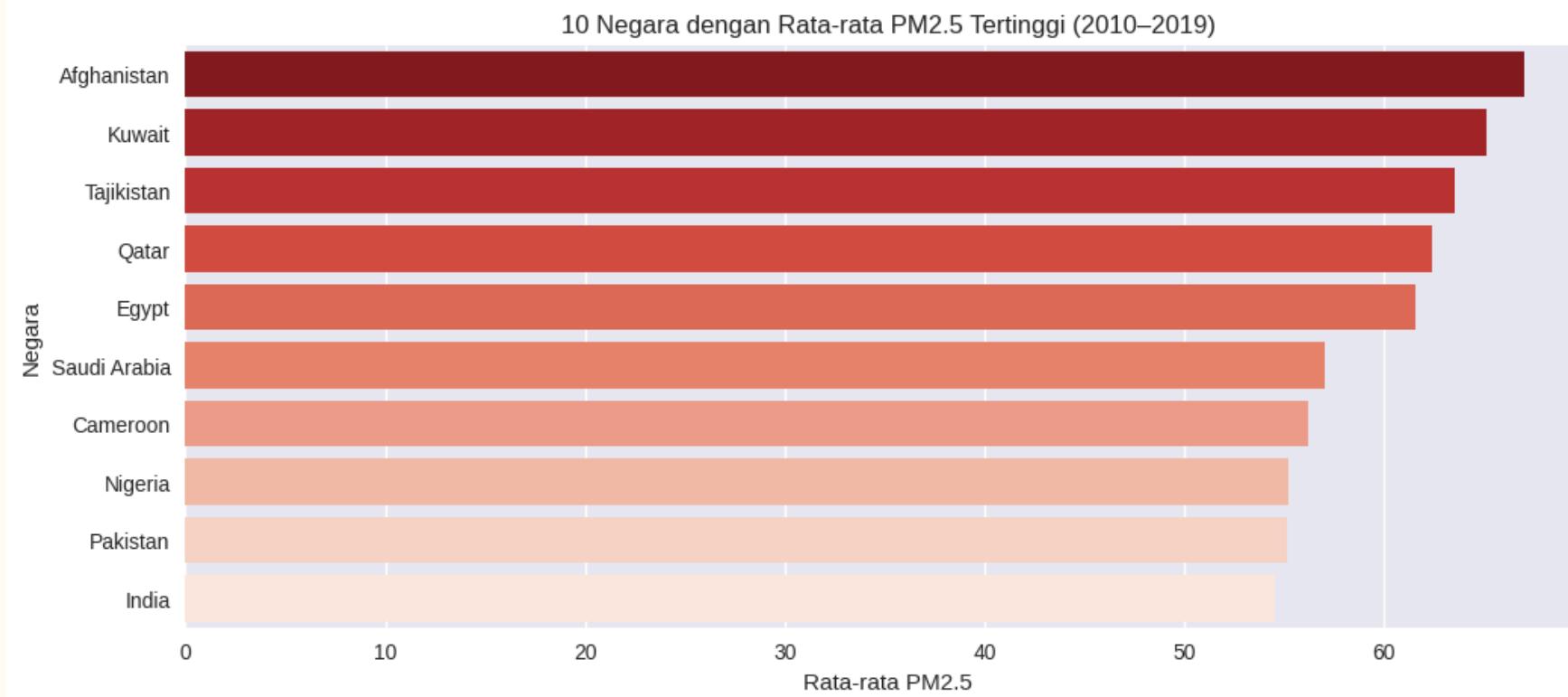
Heatmap hubungan semua variabel



Heatmap korelasi diatas kami gunakan untuk melihat kekuatan dan arah hubungan antara tiga variabel utama, yaitu PM2.5, GDP per kapita, dan populasi. Nilai korelasi menunjukkan bahwa PM2.5 memiliki hubungan negatif dengan GDP per kapita, yang berarti semakin tinggi tingkat perekonomian suatu negara, biasanya kualitas udaranya lebih baik atau tingkat PM2.5 lebih rendah. Sebaliknya, hubungan antara PM2.5 dan populasi bernilai positif, menunjukkan bahwa negara dengan jumlah penduduk lebih besar cenderung memiliki polusi udara lebih tinggi, meskipun pengaruhnya tidak terlalu kuat. Adapun hubungan populasi dengan GDP per kapita bernilai negatif dan sangat lemah, mengindikasikan bahwa jumlah penduduk tidak berhubungan langsung dengan tingkat kemakmuran ekonomi. Melalui visualisasi ini, pola hubungan antar variabel dapat dipahami secara cepat dan membantu memperkuat interpretasi pada analisis berikutnya.

Exploratory Data Analysis

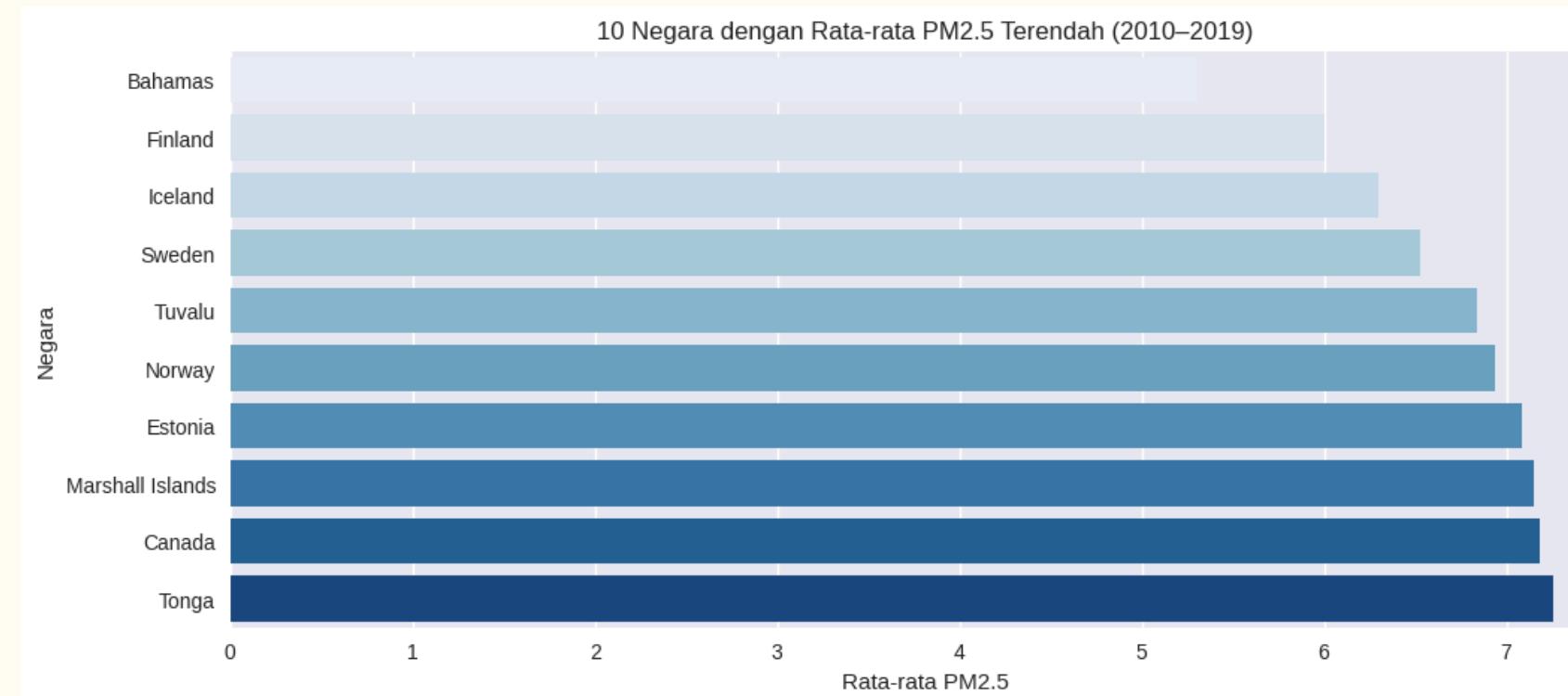
Plot negara dengan PM2.5 tertinggi



Grafik bar chart ini menampilkan sepuluh negara dengan rata-rata PM2.5 tertinggi selama periode 2010–2019, sehingga dapat terlihat negara mana yang mengalami tingkat polusi udara paling parah. Dari visualisasi tersebut terlihat bahwa Afghanistan memiliki nilai PM2.5 tertinggi, diikuti oleh Kuwait, Tajikistan, Qatar, dan Mesir. Negara-negara ini umumnya memiliki kondisi lingkungan yang dipengaruhi oleh faktor seperti aktivitas industri berat, urbanisasi yang cepat, penggunaan energi fosil, hingga kondisi geografis yang membuat polusi lebih mudah terperangkap. Grafik ini membantu mengidentifikasi negara-negara yang berada dalam kategori polusi ekstrem dan menjadi dasar penting untuk analisis lanjutan, seperti kaitannya dengan faktor ekonomi, populasi, atau kebijakan lingkungan yang diterapkan di masing-masing negara.

Exploratory Data Analysis

Plot negara dengan PM2.5 terendah



Grafik bar chart ini menampilkan sepuluh negara dengan rata-rata PM2.5 terendah pada periode 2010–2019, sehingga menggambarkan negara-negara yang memiliki kualitas udara terbaik. Negara seperti Bahamas, Finland, Iceland, Sweden, dan Tuvalu berada pada posisi teratas dengan tingkat polusi yang sangat rendah. Umumnya, negara-negara ini memiliki regulasi lingkungan yang kuat, tingkat industrialisasi yang tidak terlalu tinggi, penggunaan energi bersih yang dominan, serta kondisi geografis yang mendukung kualitas udara lebih baik. Visualisasi ini memberikan gambaran yang jelas mengenai negara mana yang berhasil menjaga udara tetap bersih, sekaligus menjadi pembanding yang bermanfaat bagi negara-negara yang masih menghadapi tingkat polusi tinggi.

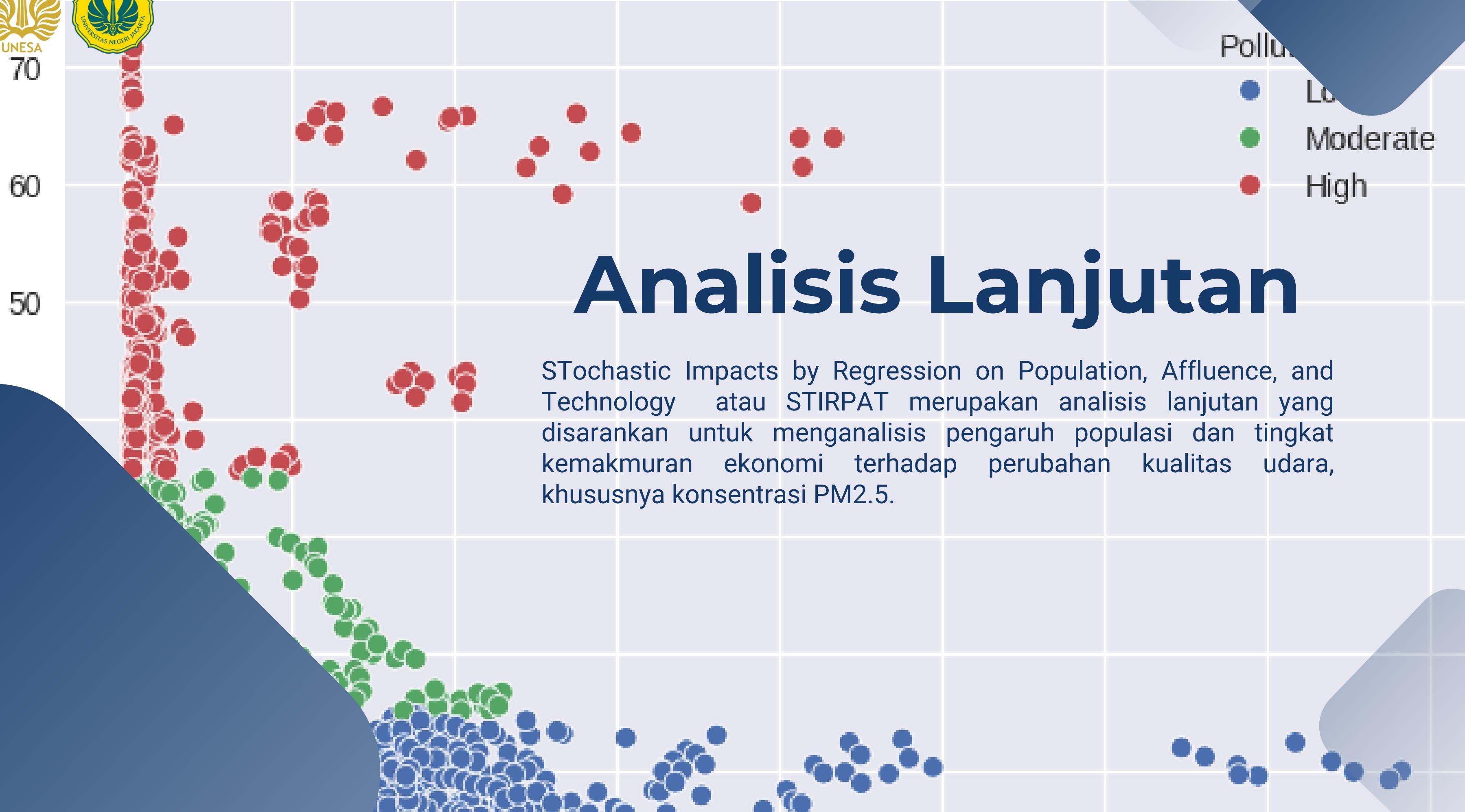


Data Publishing

Sebagai tahap akhir dari proses data wrangling, seluruh hasil pengolahan data, mulai dari dataset yang telah dibersihkan (cleaning), hasil pre-processing, hasil integrasi data, dataset feature engineering, hingga seluruh visualisasi EDA telah dipublikasikan (data publishing) ke dalam repository GitHub. Penerbitan data ini dilakukan agar seluruh proses dan output proyek dapat diakses, diverifikasi, serta direplikasi oleh pihak lain.



Korelasi GDP per Kapita vs PM2.5





TERIMA KASIH