

PROYEK SAINS DATA

“Analisis Data Kaggle menggunakan Power BI pada data World Air Quality Index”

Dosen Pengampu : Eka Mala Sari Rochman, S.Kom.M.Kom.



Disusun Oleh :

Azriel Christofer Saragih 210411100156

**PRODI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS TRUNOJOYO MADURA
2023**

Indeks Kualitas Udara (AQI) adalah indikator pemerintah yang dimaksudkan untuk memberi tahu masyarakat tentang jumlah tingkat polusi udara yang ada atau yang diperkirakan. AQI terkait erat dengan masalah kesehatan masyarakat. AQI mengukur polusi ozon dan partikel, serta empat kontaminan udara umum lainnya.

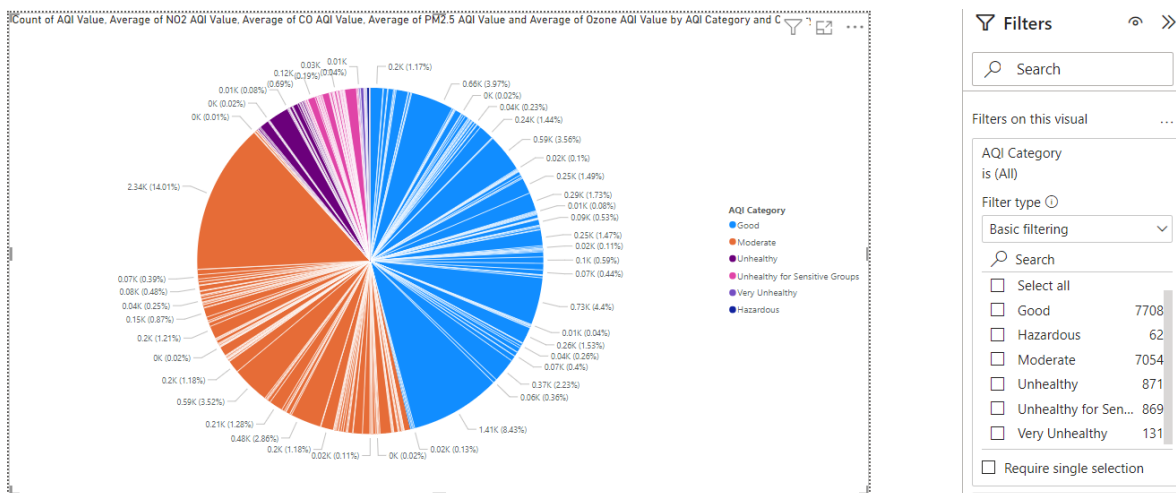
AQI diukur pada skala 1 hingga 500, dengan nilai rendah yang menunjukkan kualitas udara yang dapat diterima dan nilai tinggi yang menunjukkan kualitas udara yang buruk. AQI juga memberikan kategori kualitas udara berdasarkan rentang numerik, di mana setiap kategori dan rentang numerik setara dengan peningkatan risiko kesehatan karena konsentrasi kontaminan udara yang lebih besar.

Peningkatan emisi udara dapat berkontribusi pada peningkatan AQI. Contohnya termasuk jam-jam sibuk lalu lintas, kebakaran hutan yang dipicu oleh angin kencang, dan kurangnya pengenceran polutan udara. Udara yang tergenang di suatu daerah, yang disebabkan oleh antisiklon, inversi suhu, atau kecepatan angin yang rendah, dapat menyebabkan konsentrasi polusi yang tinggi dan kondisi berkabut.

Ketika AQI tinggi, pemerintah menyarankan individu untuk membatasi aktivitas fisik di luar ruangan atau bahkan tetap berada di dalam ruangan. Ketika terjadi kebakaran hutan dengan AQI tinggi, penggunaan masker seperti respirator N95 atau pembersih udara dengan filter HEPA mungkin disarankan.

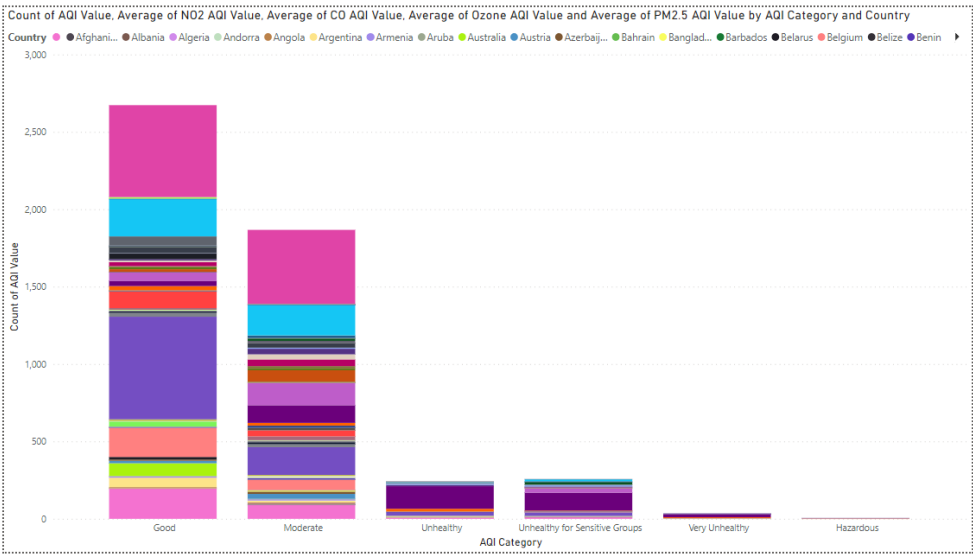
1. SEBARAN DATA PADA POWER BI

Dalam statistik, distribusi data mengacu pada bagaimana data tersebar atau disebarkan di sekitar nilai tertentu dalam kumpulan data. Distribusi data menggambarkan bagaimana data dalam kumpulan data tersebar atau diposisikan di sepanjang berbagai nilai. Memahami distribusi data sangat penting dalam analisis statistik karena membantu kita memahami sifat dan pola data, yang kemudian dapat digunakan untuk membuat penilaian statistik atau menarik kesimpulan statistik.



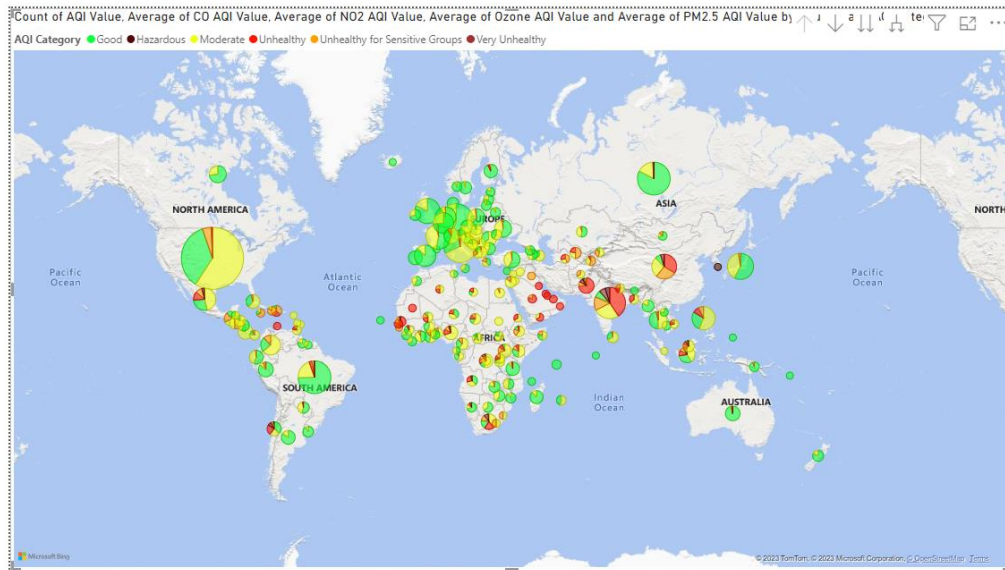
Gambar Sebaran Data menggunakan Pie Chart

Grafik di atas menggambarkan distribusi data menggunakan diagram lingkaran berdasarkan data indeks kualitas udara dunia. Pada gambar tersebut, saya menggunakan parameter yang diperlukan untuk mewakili data, dimana tujuannya adalah kategori indeks kualitas udara, yaitu *good*, *moderate*, *unhealthy for sensitive groups*, *unhealthy*, *very unhealthy*, dan *hazardous*. Kemudian terdapat hasil indeks *good* hingga 7708 data, *moderate* hingga 7054 data, *unhealthy for sensitive groups* hingga 869 data, *unhealthy* hingga 871 data, *very unhealthy* hingga 131 data, dan *hazardous* hingga 62 data. Data-data tersebut didistribusikan secara berurutan sebagai data pendukung untuk menentukan kategori indeks kualitas udara di dunia.



Gambar Sebaran Data menggunakan Stacked Column Chart

Diagram batang diatas adalah sebaran data berdasarkan kategori indeks kualitas udara(AQI). Diagram tersebut menggunakan parameter seperti *Country Name*, *City Name*, *Carbon Monoxide Value*, *Ozone Value*, *Nitrogen Dioxide Value*, dan *PM2.5 AQI Value* untuk menentukan kategori indeks kualitas udara di setiap kota dalam sebuah negara. Sehingga terlihat hasil seperti diagram tersebut.

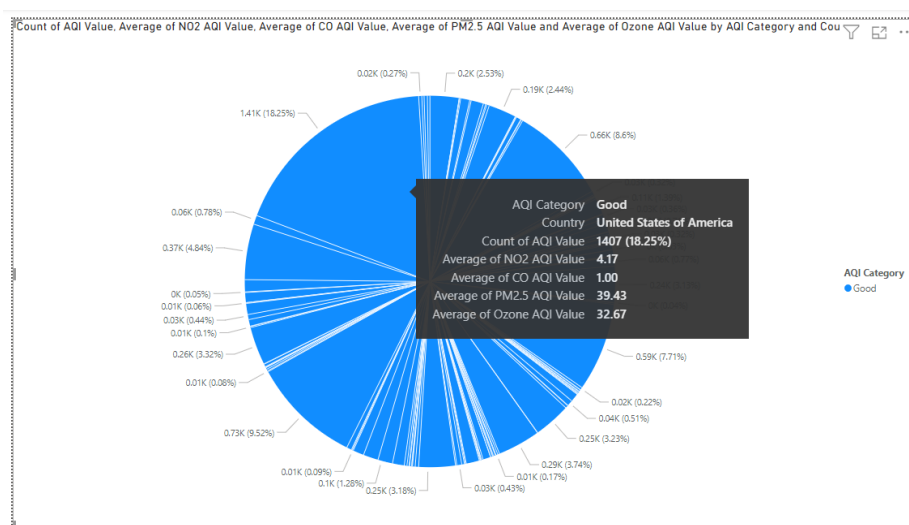


Gambar yang Menampilkan Sebaran Data menggunakan Map

Seperti yang diketahui, gambar diatas adalah sebaran data menggunakan fitur map pada *software* Power BI yang berfungsi untuk menampilkan indeks kualitas udara di setiap negara untuk mengetahui kategori kualitas udara negara-negara tersebut.

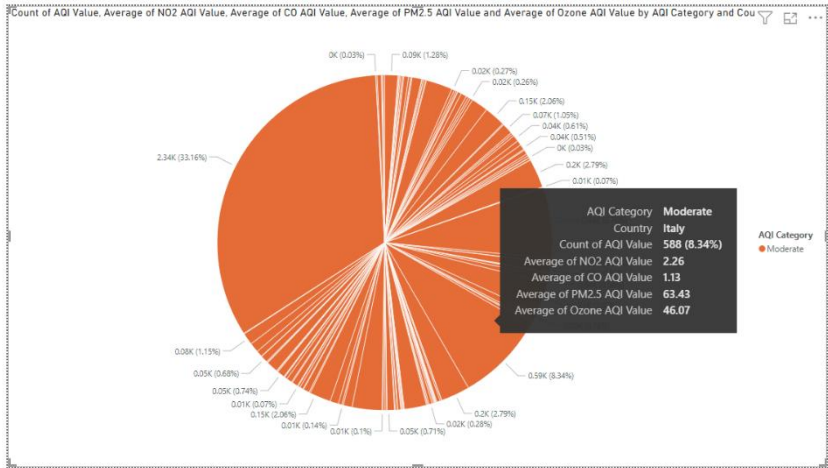
2. ANALISIS DATA

Berdasarkan dari data yang diambil pada situs [www.Kaggle.com/datasets](https://www.kaggle.com/datasets) yang berjudul “Air Quality Index” terdapat 14 parameter dan 1 target, yaitu : *Country*, *City*, *AQI Value*, *AQI Category*, *CO (Carbon Monoxide) AQI Value*, *CO AQI Category*, *Ozone AQI Value*, *Ozone AQI Category*, *NO2 AQI Value (Nitrogen Dioxide)*, *NO2 AQI Category*, *PM2.5 AQI Value (Fine particulate matter less than 2.5 micrometers in diameter)*, *M2.5 AQI Category*, *lat (Latitude of the city)*, dan *lng (Longitude of the city)*. Data yang diperoleh sebanyak 16695 data dan tidak ada ditemukan *missing value*, sehingga saya tidak melakukan proses mengatasi *missing value*.

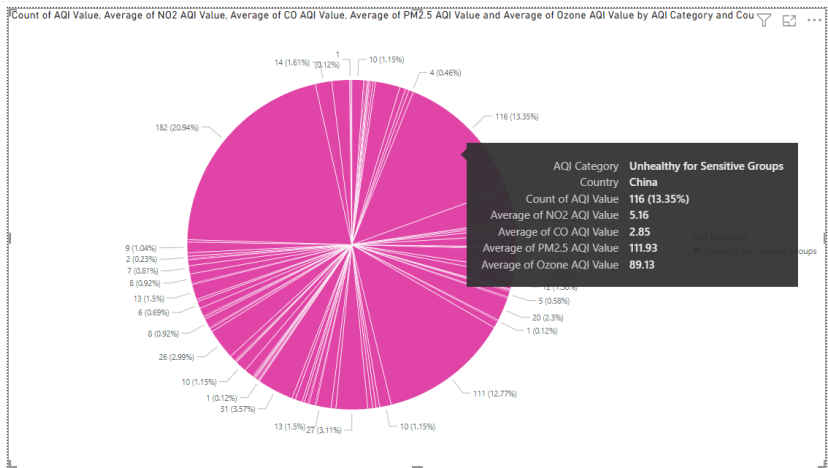


Gambar Hasil Analisis Kategori Indeks Kualitas Udara Good

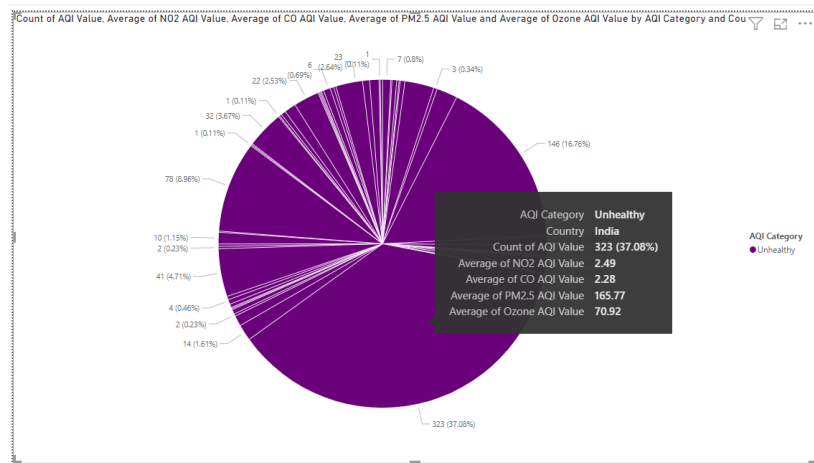
Berikut merupakan hasil analisis kategori indeks kualitas udara *good* yang memiliki jumlah 7708 data yang didapatkan dengan menentukan rata-rata dari parameter NO2, CO, PM2,5, dan Ozone dari suatu kota di negara tertentu. Begitupula dengan kategori indeks kualitas udara lainnya.



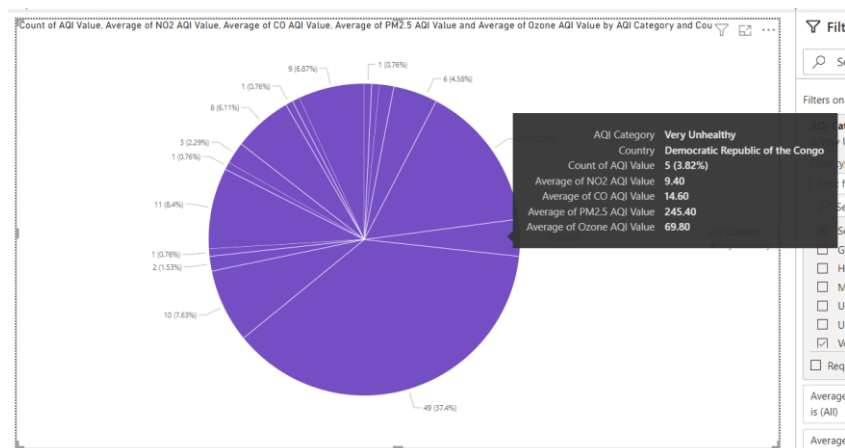
Gambar Hasil Analisis Kategori Indeks Kualitas Udara Moderate



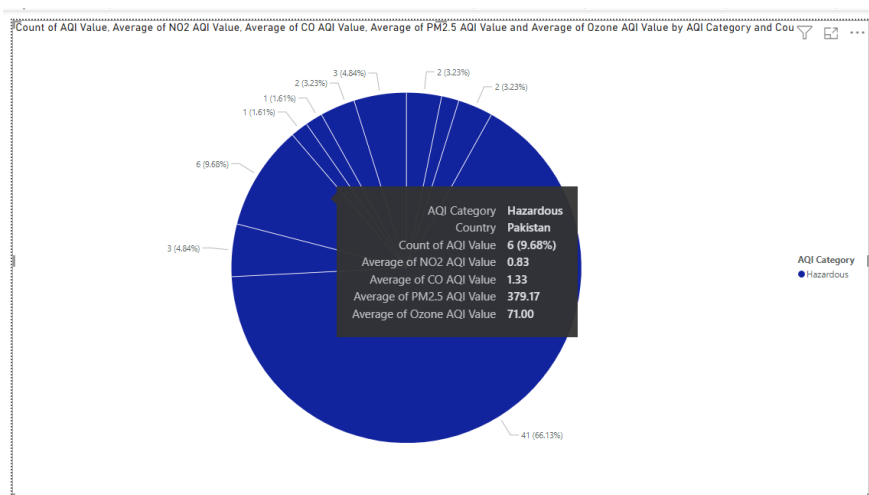
Gambar Hasil Analisis Kategori Indeks Kualitas Udara Unhealthy For Sensitive Groups



Gambar Hasil Analisis Kategori Indeks Kualitas Udara Unhealthy



Gambar Hasil Analisis Kategori Indeks Kualitas Udara Very Unhealthy

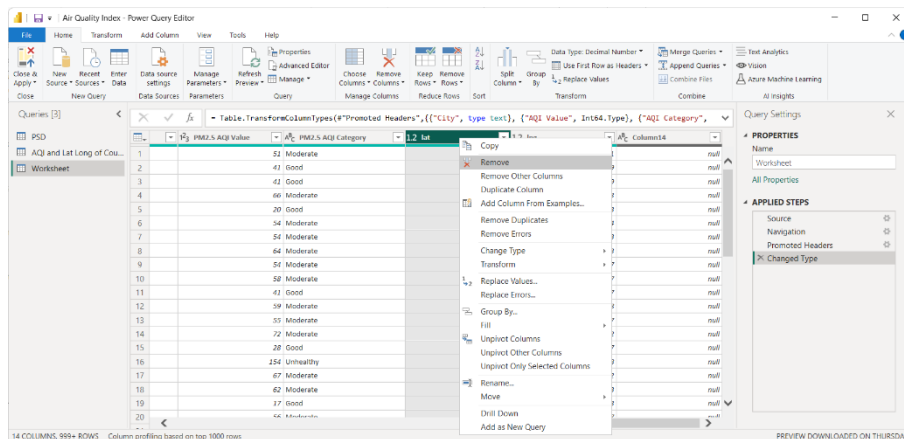


Gambar Hasil Analisis Kategori Indeks Kualitas Udara Hazardous

3. TRANSFORMASI DATA

Transformasi data adalah proses penting dalam menyusun data mentah dan membuatnya tersedia secara bebas untuk analisis informasi, sehingga dapat meningkatkan pengambilan keputusan organisasi.

Data Cleansing adalah tindakan menghilangkan atau memperbaiki data yang tidak akurat, rusak, tidak diformat dengan benar, duplikat, atau tidak lengkap dari kumpulan data. Adapun parameter yang dihapus adalah *lat*, dan *lng*.



Gambar proses Cleansing Data

Pada gambar diatas dilakukan proses Data Cleaning yaitu untuk menghapus atau menghilangkan parameter yang tidak digunakan dalam melakukan identifikasi data yang akan dianalisis.

