

LAPORAN ANALISIS DATA SET EMISI CO₂



Kelompok 8

Mochammad Qussay Alhindi Achmadi (1103213087)

Muhammad Naufal Afif (1103210089)

Mochamad Arief Dermawan (1103210128)

Rangga Khalid (1103210027)

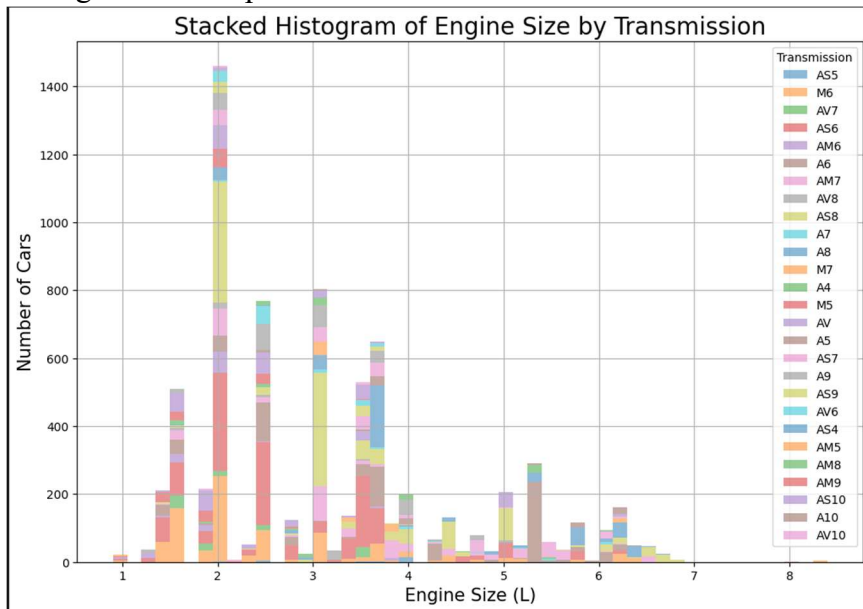
PRODI S1 TEKNIK KOMPUTER

FAKULTAS TEKNIK ELEKTRO

UNIVERSITAS TELKOM

2025

1. Histogram Bertumpuk



Menunjukkan distribusi ukuran mesin (Engine Size dalam liter) berdasarkan jenis transmisi yang berbeda.

Sumbu X (Horizontal):

- Menampilkan ukuran mesin dalam liter (Engine Size (L)).
- Data dikelompokkan ke dalam beberapa **bin (interval)** untuk menunjukkan frekuensi.

Sumbu Y (Vertikal):

- Menunjukkan jumlah mobil (Number of Cars) dalam dataset dengan ukuran mesin tertentu.

Legenda (Kanan):

- Menampilkan berbagai jenis **transmisi** yang ada dalam dataset.
- Setiap warna dalam histogram mewakili satu jenis transmisi.

Bentuk Histogram:

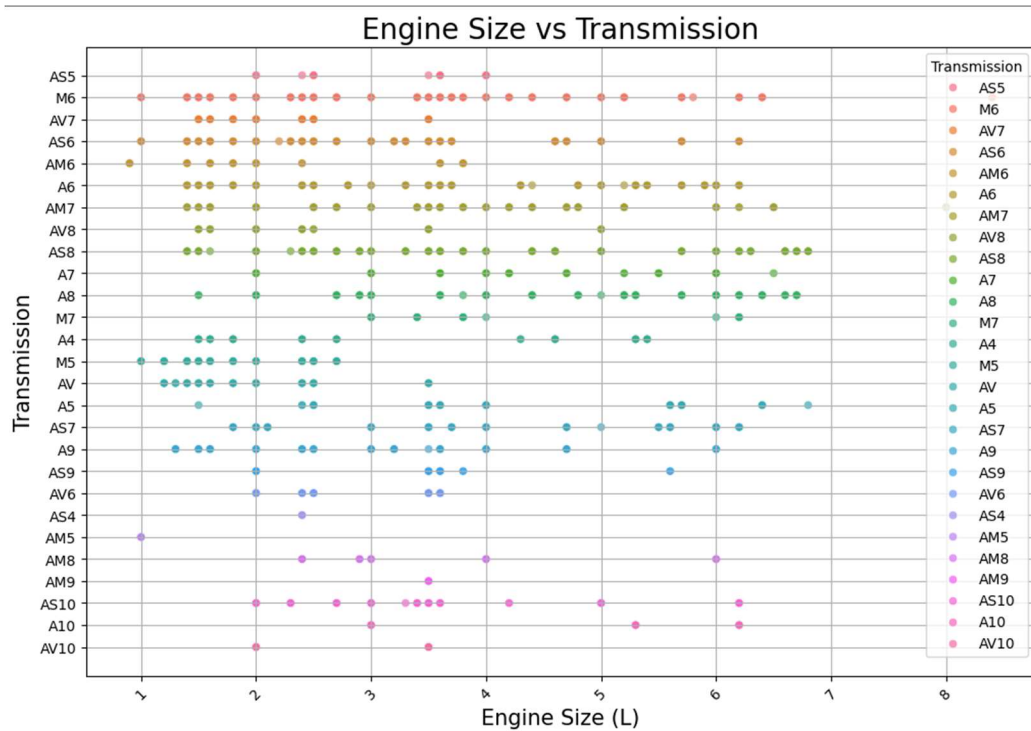
- Stacked histogram artinya setiap batang terdiri dari beberapa lapisan warna, yang merepresentasikan **kontribusi setiap jenis transmisi** terhadap total jumlah mobil dengan ukuran mesin tertentu.
- Ada puncak tinggi pada sekitar **2.0L, 3.0L, dan 4.0L**, yang menunjukkan bahwa mayoritas mobil dalam dataset memiliki ukuran mesin di kisaran ini.

Kesimpulan dari Grafik:

- Sebagian besar mobil memiliki ukuran mesin sekitar **2.0L - 3.0L**.
- Banyak variasi transmisi untuk ukuran mesin tertentu.

- Mobil dengan mesin besar (di atas 5.0L) lebih jarang muncul.

2. Scatter Plot



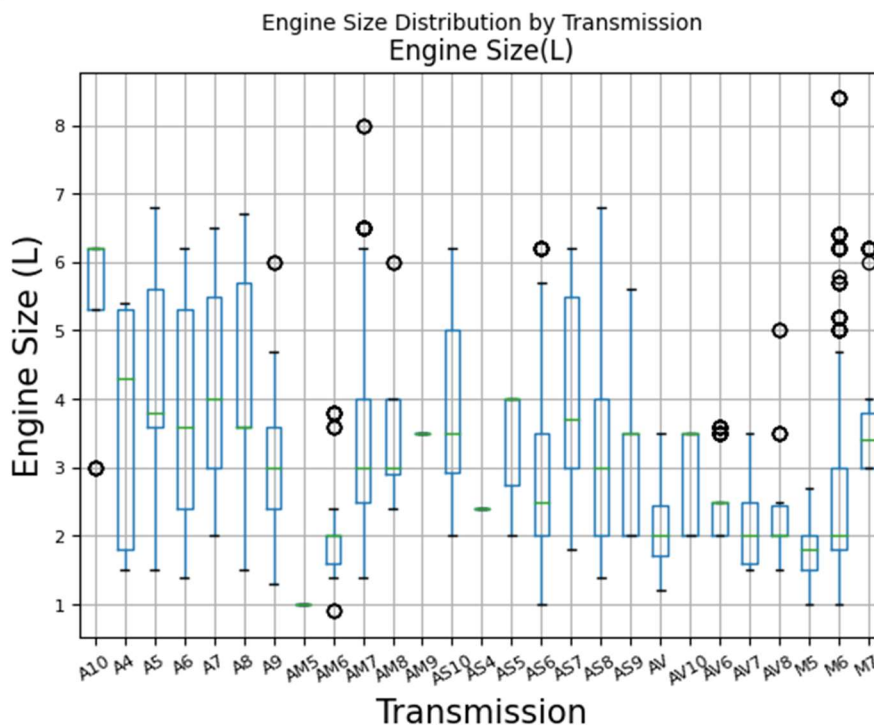
Yang menunjukkan hubungan antara **ukuran mesin (Engine Size)** dan **jenis transmisi (Transmission)**.

- Sumbu X (Horizontal):**
 - Menampilkan **Engine Size (L)** atau ukuran mesin dalam liter.
 - Data tersebar dari sekitar **1L hingga 7L**.
- Sumbu Y (Vertikal):**
 - Menampilkan berbagai jenis **transmisi** yang ada dalam dataset.
 - Setiap titik dalam grafik mewakili satu mobil dengan kombinasi tertentu dari ukuran mesin dan jenis transmisi.
- Warna dalam Grafik:**
 - Setiap warna mewakili satu jenis transmisi, sesuai dengan legenda di sebelah kanan.
 - Data dikelompokkan berdasarkan transmisi untuk melihat pola distribusi ukuran mesin dalam setiap kategori.
- Distribusi Data:**
 - Ukuran mesin **2L hingga 3L** memiliki banyak variasi transmisi dan lebih padat.
 - Mobil dengan ukuran mesin besar (5L ke atas) lebih jarang dan memiliki jenis transmisi yang lebih spesifik.
 - Beberapa jenis transmisi lebih dominan pada ukuran mesin tertentu.

Kesimpulan dari Grafik:

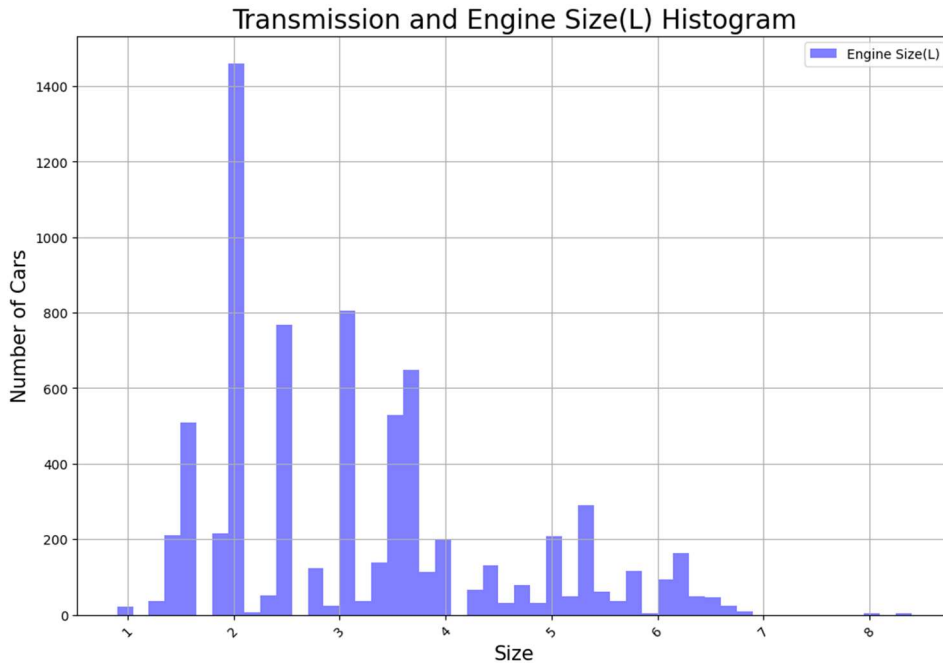
- Jenis transmisi memiliki distribusi yang bervariasi berdasarkan ukuran mesin.
- Mesin kecil lebih umum dengan berbagai jenis transmisi, sedangkan mesin besar memiliki lebih sedikit variasi transmisi.
- Scatter plot ini mempermudah analisis tentang bagaimana jenis transmisi tersebar di berbagai ukuran mesin.

3. Box Plot



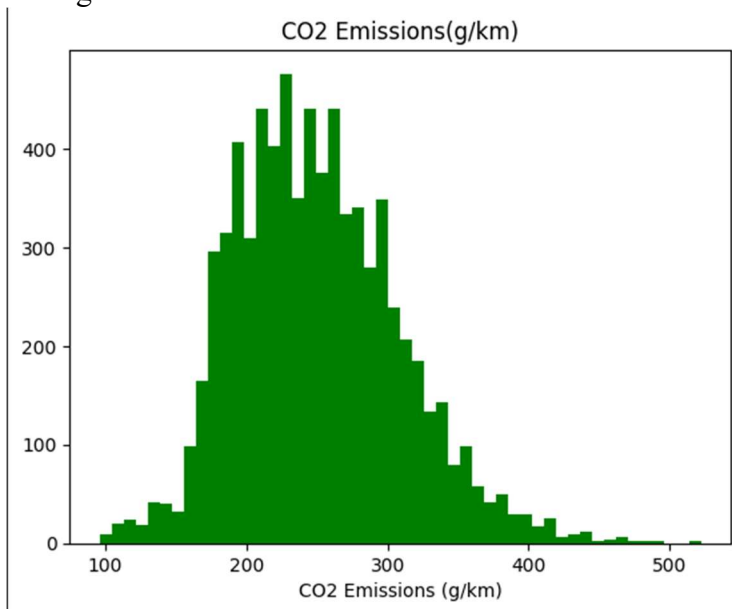
Menggambarkan distribusi ukuran mesin (Engine Size) berdasarkan jenis transmisi (Transmission). Setiap kotak dalam diagram mewakili **distribusi interkuartil** (IQR) dari ukuran mesin untuk masing-masing jenis transmisi, dengan garis horizontal di dalam kotak menunjukkan **median**. Garis vertikal (whiskers) menunjukkan rentang data utama, sementara titik-titik di luar whiskers menunjukkan **outliers**, yaitu nilai yang sangat berbeda dari mayoritas data. Dari grafik ini, terlihat bahwa ukuran mesin sangat bervariasi antar jenis transmisi; beberapa kategori memiliki rentang yang luas dengan banyak outlier, sementara yang lain lebih terkonsentrasi pada ukuran mesin tertentu. Selain itu, beberapa transmisi, seperti **A10**, **A4**, dan **M6**, memiliki ukuran mesin yang relatif besar, sementara yang lain lebih terpusat pada kisaran yang lebih kecil.

4. Histogram



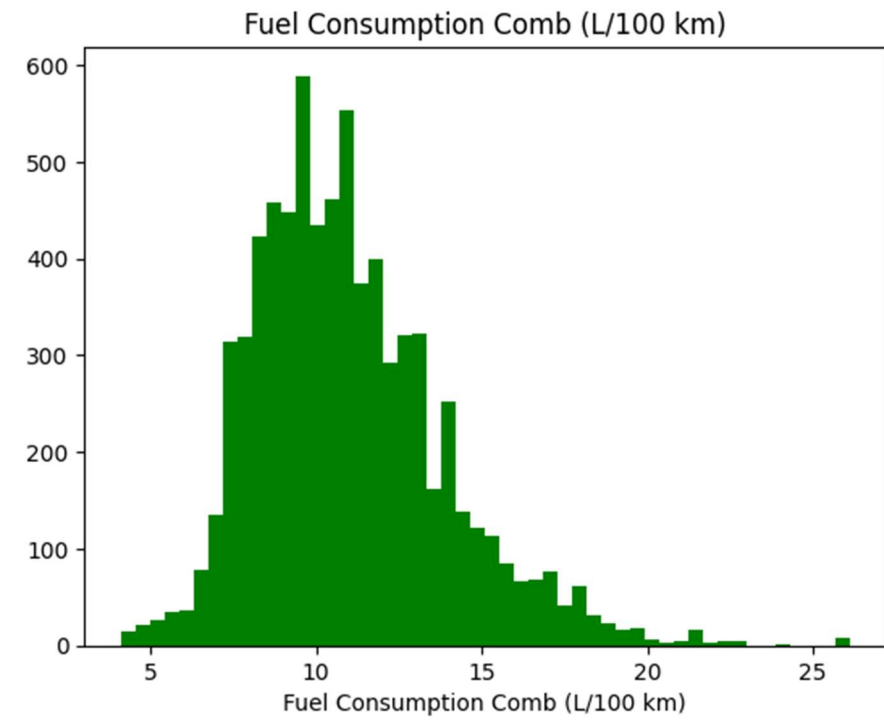
Menunjukkan distribusi jumlah mobil berdasarkan **ukuran mesin (Engine Size)**. Pada sumbu **X**, ditampilkan ukuran mesin dalam liter (L), sementara sumbu **Y** menunjukkan jumlah mobil dalam dataset dengan ukuran mesin tertentu. Histogram ini menggambarkan bahwa sebagian besar mobil memiliki ukuran mesin di kisaran **2L hingga 3L**, dengan puncak tertinggi sekitar **2L**, yang berarti ukuran mesin ini adalah yang paling umum dalam dataset. Selain itu, ada beberapa mobil dengan ukuran mesin lebih besar, tetapi jumlahnya jauh lebih sedikit. Grafik ini membantu dalam memahami distribusi mesin dalam populasi kendaraan yang dianalisis, menunjukkan bahwa mesin kecil hingga menengah lebih dominan dibandingkan dengan mesin berkapasitas besar.

5. Histogram



Menunjukkan distribusi **emisi CO2 (dalam gram per kilometer)** dari kendaraan dalam dataset. Sumbu **X** menunjukkan tingkat emisi CO2 (g/km), sedangkan sumbu **Y** menunjukkan jumlah kendaraan dalam dataset yang memiliki tingkat emisi tertentu. Histogram ini memiliki bentuk **distribusi normal**, dengan mayoritas kendaraan memiliki emisi CO2 berkisar antara **150 hingga 300 g/km**, yang merupakan puncak distribusi. Seiring dengan meningkatnya emisi CO2, jumlah kendaraan yang termasuk dalam kategori tersebut semakin berkurang, menunjukkan bahwa kendaraan dengan emisi tinggi lebih jarang ditemukan. Grafik ini memberikan wawasan tentang bagaimana emisi kendaraan terdistribusi dan dapat digunakan untuk menganalisis tren dalam efisiensi bahan bakar serta dampaknya terhadap lingkungan.

6. Histogram



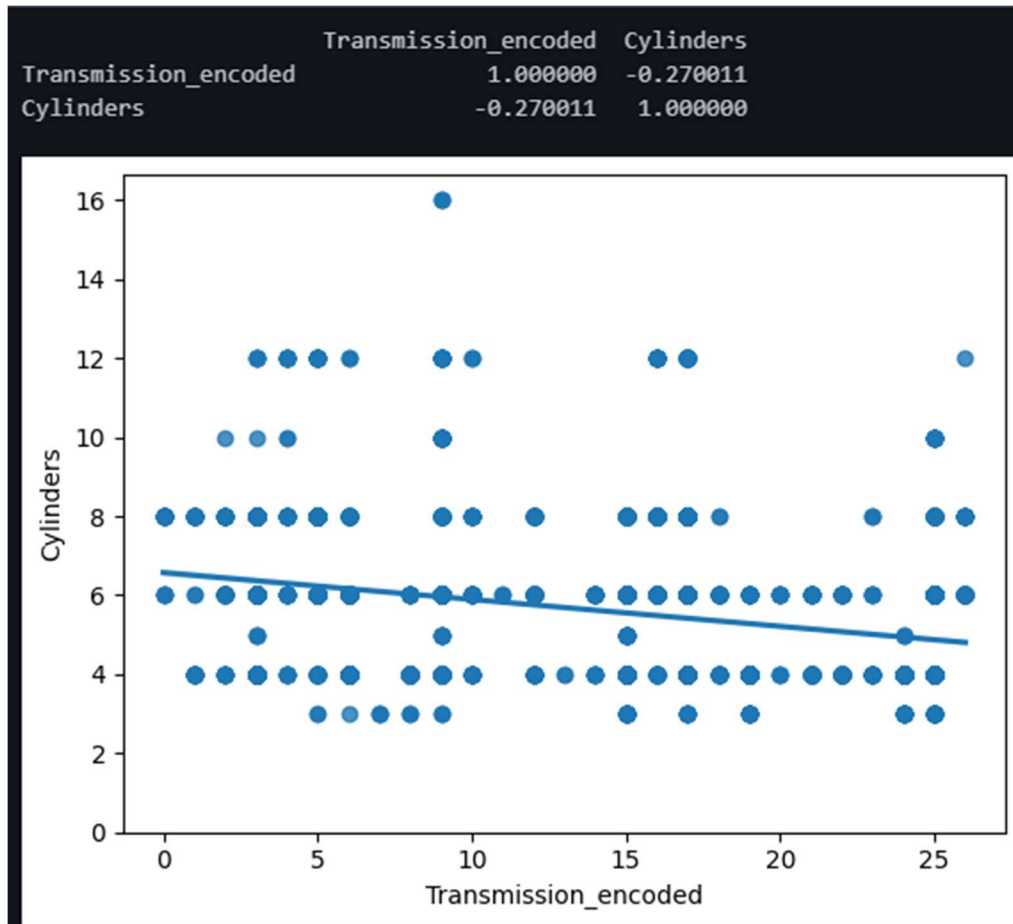
Menunjukkan distribusi konsumsi bahan bakar kendaraan dalam satuan liter per 100 kilometer (L/100 km). Sumbu **X** mewakili konsumsi bahan bakar kombinasi, sementara sumbu **Y** menunjukkan jumlah kendaraan yang memiliki tingkat konsumsi bahan bakar tertentu dalam dataset.

Dari grafik, terlihat bahwa mayoritas kendaraan memiliki konsumsi bahan bakar antara **7 hingga 15 L/100 km**, dengan puncak distribusi berada di sekitar **10 L/100 km**. Seiring dengan meningkatnya konsumsi bahan bakar, jumlah kendaraan dalam kategori tersebut semakin berkurang, menunjukkan bahwa kendaraan dengan konsumsi bahan bakar tinggi lebih jarang ditemukan.

Histogram ini memberikan wawasan penting terkait efisiensi bahan bakar kendaraan dan dapat digunakan untuk memahami tren konsumsi bahan bakar, membandingkan efisiensi berbagai jenis

kendaraan, serta mengevaluasi dampak konsumsi bahan bakar terhadap emisi CO2 dan biaya operasional.

7. Scatter plot dengan regresi linear.

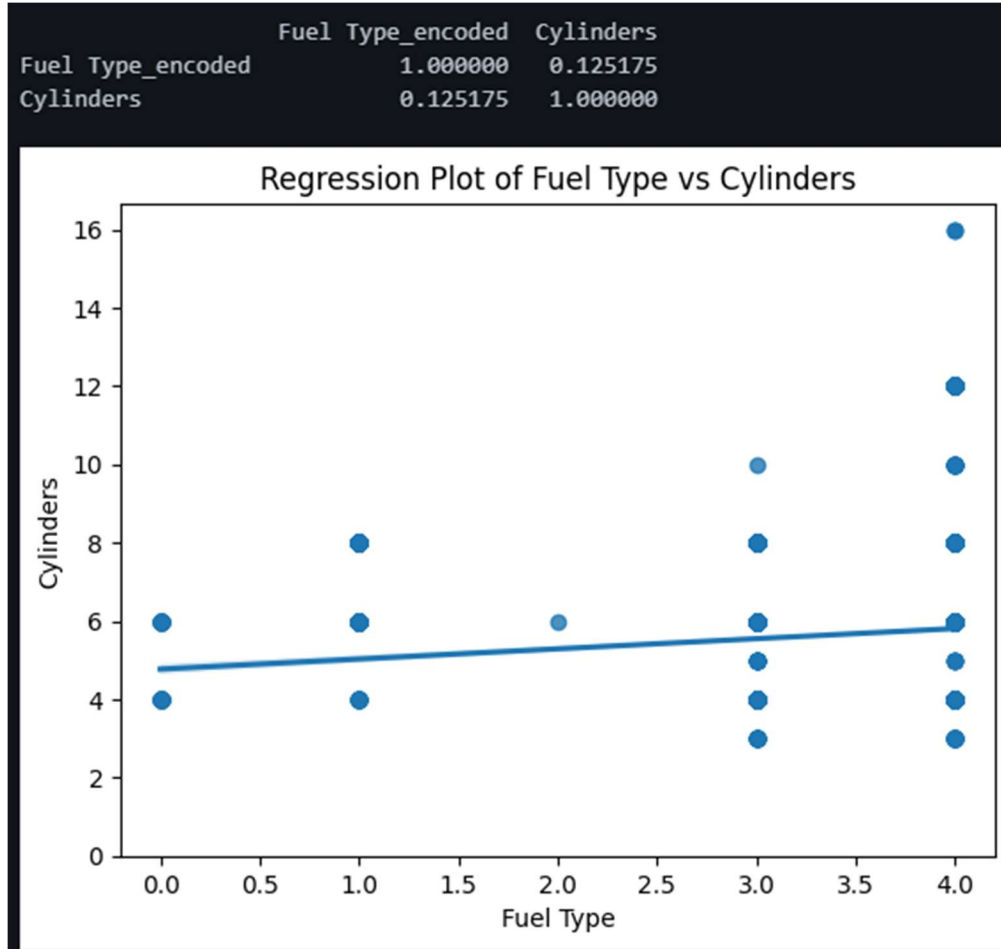


Menunjukkan hubungan antara jumlah silinder mesin kendaraan (Cylinders) dengan jenis transmisi yang telah dikodekan secara numerik (Transmission_encoded) menggunakan scatter plot dengan regresi linear.

Bagian atas grafik menampilkan **matriks korelasi** antara dua variabel, dengan nilai korelasi sebesar **-0.270011**. Korelasi negatif ini menunjukkan bahwa semakin tinggi nilai **Transmission_encoded** (yang merepresentasikan jenis transmisi yang berbeda), umumnya jumlah silinder kendaraan cenderung lebih rendah.

Scatter plot di bawahnya memperlihatkan sebaran data dengan garis regresi yang memiliki kemiringan negatif, yang mengonfirmasi adanya tren penurunan jumlah silinder seiring dengan perubahan nilai transmisi. Hal ini dapat mengindikasikan bahwa jenis transmisi tertentu lebih umum digunakan pada kendaraan dengan jumlah silinder lebih sedikit, misalnya transmisi otomatis modern lebih sering ditemukan pada mobil dengan mesin lebih kecil.

8. Scatter Plot dengan regresi Linear



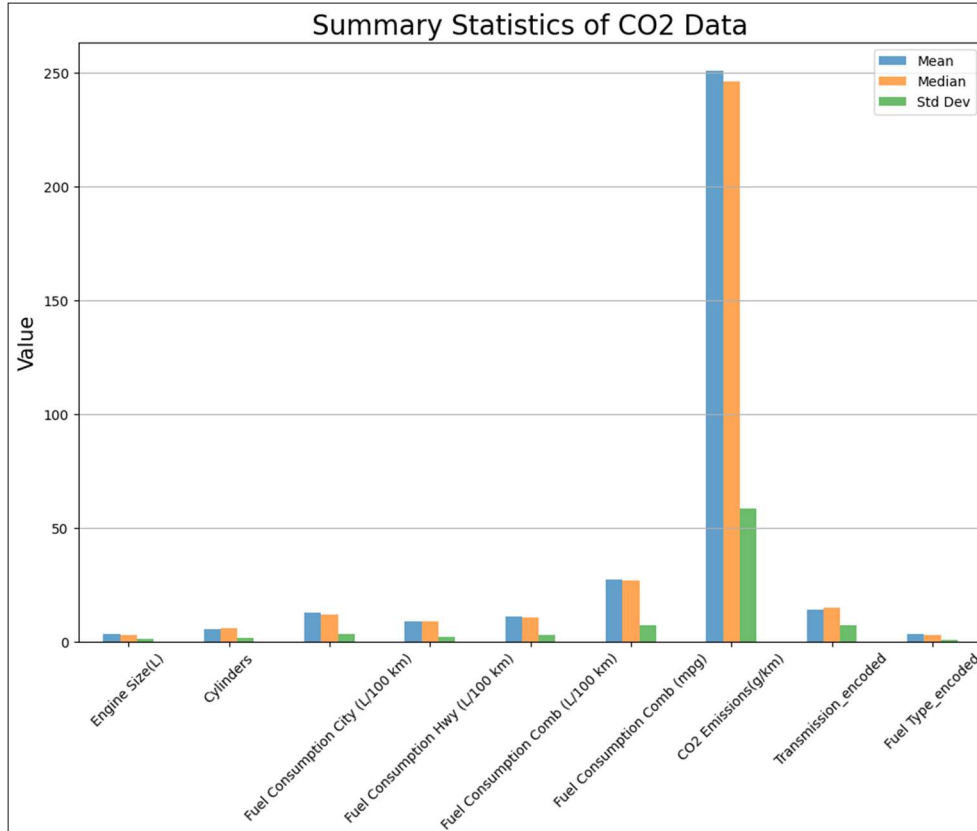
- **Matriks Korelasi** (di bagian atas):

- Nilai korelasi antara **Fuel Type_encoded** (jenis bahan bakar yang dikodekan) dan **Cylinders** adalah **0.125175**.
- Korelasi ini positif tetapi sangat lemah, yang berarti bahwa **jenis bahan bakar memiliki sedikit hubungan dengan jumlah silinder mesin**.

- **Scatter Plot dengan Garis Regresi:**

- Sumbu **X** menunjukkan **Fuel Type_encoded** (bahan bakar yang telah dikodekan menjadi angka).
- Sumbu **Y** menunjukkan **jumlah silinder mesin**.
- Titik-titik mewakili kendaraan dengan jenis bahan bakar dan jumlah silinder tertentu.
- **Garis regresi memiliki kemiringan sedikit ke atas**, menunjukkan bahwa kendaraan dengan bahan bakar tertentu mungkin memiliki sedikit lebih banyak silinder, tetapi hubungannya sangat lemah.

9. Statistik ringkasan



1. Sumbu X:

- Menampilkan berbagai variabel dalam dataset, seperti **ukuran mesin (Engine Size)**, **jumlah silinder (Cylinders)**, **konsumsi bahan bakar (Fuel Consumption)**, **efisiensi bahan bakar dalam mpg (miles per gallon)**, **emisi CO2 (CO2 Emissions)**, serta **fitur kategorikal yang dikodekan (Transmission_encoded & Fuel Type_encoded)**.

2. Sumbu Y:

- Menunjukkan nilai dari masing-masing statistik (Mean, Median, dan Std Dev) untuk setiap variabel.

3. Interpretasi Statistik:

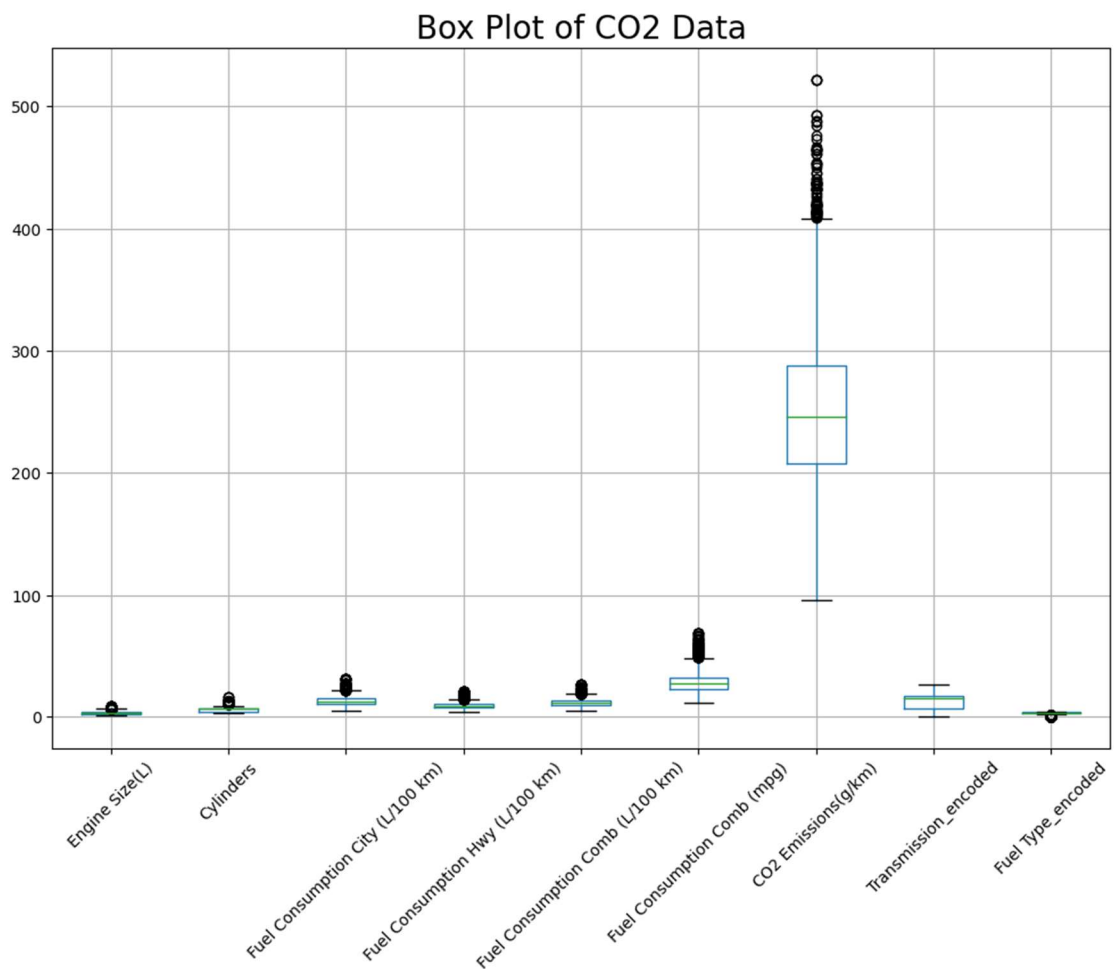
- Mean (Rata-rata):** Rata-rata nilai untuk masing-masing variabel.
- Median:** Nilai tengah dari data untuk masing-masing variabel, yang berguna untuk melihat distribusi data dan apakah ada skewness (kecenderungan miring).
- Std Dev (Standar Deviasi):** Mengukur seberapa besar variasi atau sebaran data. Jika standar deviasi tinggi, berarti datanya lebih tersebar.

Analisis Grafik:

- CO2 Emissions (g/km)** memiliki nilai tertinggi dibanding variabel lain, dengan rata-rata dan median yang hampir sama, tetapi standar deviasi cukup besar. Ini menunjukkan bahwa ada variasi signifikan dalam emisi CO2 antar kendaraan.

- **Fuel Consumption (L/100 km)** juga memiliki distribusi yang serupa dengan CO2 Emissions karena keduanya berkaitan erat.
- **Cylinders & Engine Size** memiliki distribusi yang lebih merata dengan standar deviasi yang relatif kecil dibanding variabel lainnya.
- **Variabel kategorikal seperti Transmission_encoded & Fuel Type_encoded** memiliki nilai yang lebih kecil karena merupakan hasil encoding dari kategori.

10. Box Plot Co2



1. **Setiap kotak (box) dalam plot mewakili distribusi nilai untuk setiap variabel.**
2. **Komponen utama dalam box plot:**
 - **Garis tengah dalam kotak (median):** Mewakili nilai tengah dari dataset.
 - **Bagian atas dan bawah kotak (Q1 dan Q3):** Menunjukkan kuartil pertama (25%) dan kuartil ketiga (75%).
 - **Whiskers (garis vertikal di luar kotak):** Menunjukkan rentang data dalam 1.5 kali Interquartile Range (IQR).
 - **Titik hitam di luar whiskers (outliers):** Nilai yang jauh lebih besar atau lebih kecil dibandingkan dengan distribusi utama.

Analisis dari Box Plot

- **CO2 Emissions (g/km)** memiliki banyak **outlier** yang sangat tinggi, menunjukkan bahwa ada beberapa kendaraan dengan emisi yang jauh lebih besar dibandingkan yang lain.
- **Fuel Consumption (City, Hwy, Comb)** juga memiliki outlier, yang berarti ada kendaraan dengan konsumsi bahan bakar yang jauh lebih boros dibandingkan mayoritas kendaraan lainnya.
- **Engine Size dan Cylinders** memiliki distribusi yang lebih stabil tanpa banyak outlier.
- **Transmission_encoded dan Fuel Type_encoded** terlihat memiliki distribusi yang lebih terkonsentrasi karena variabel ini dikodekan sebagai kategori.