TUGAS PROJECT 1

Model Linier (A)



KELOMPOK H

Adriaan Aylsworth 2106725072

Angelina Putri 2006464152

Aulia Shinta 2106724984

Ricky 2106724952

Rifqah Majidah 2106725122

Departemen Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Indonesia 2022

Regresi Linier Emisi CO_2

Tujuan:

Tugas ini bertujuan untuk memahami lebih dalam dan menerapkan model regresi linear untuk menyelesaikan permasalahan jumlah emisi ${\cal CO}_2$

Cakupan tugas:

- 1. Pemahaman konstekstual, berupa interpretasi kuantitas yang dipelajari dalam konteks permasalahan yang ada
- 2. Pengolahan data, analisis dan interpretasi hasil

Anggota kelompok:

No	Nama	NPM	Kontribusi	Tingkat kontribusi	
1	Adriaan Aylsworth Waleleng	2106725072	Aktif dalam diskusi, mencari sumber dan topik, dan membuat koding.	100%	
2	Angelina Putri	2006464152	Mencari sumber lain untuk klarifikasi pemahaman topik, menulis ringkasan hasil diskusi	100%	
3	Aulia Shinta Madani	2106724984	Aktif dalam diskusi, mengerjakan powerpoint, dan presentasi	100%	
4	Ricky	2106724952	Aktif dalam berdiskusi memilih topik, membuat laporan, serta powerpoint	100%	
5	Rifqah Majidah Avisenna Nasution	2106725122	Aktif dalam diskusi dan mengerjakan laporan	100%	

Bagian 1

Pendahuluan

Meningkatnya jumlah kendaraan mobil memberikan pengaruh yang besar terhadap peningkatan polusi udara. Gas sisa yang keluar melalui knalpot berperan besar terhadap pencemaran udara. Zat-zat yang ada di udara tersebut dapat berbahaya bagi tubuh jika terhirup dalam jangka waktu yang lama. Salah satu zat yang dikeluarkan dari sisa pembakaran kendaraan mobil adalah gas karbon dioksida. Karbon dioksida jika diabaikan maka konsentrasinya akan terakumulasi di atmosfer dan berpotensi menyebabkan pemanasan global dan dalam jangka panjang akan mengakibatkan perubahan iklim yang berbahaya bagi kehidupan manusia.

Permasalahan yang dibahas kali ini adalah catatan emisi karbon dioksida yang dihasilkan dari berbagai mobil dengan fitur yang berbeda. Akan dicari variabel-variabel yang menyumbang besar terhadap emisi karbon dioksida. Data diambil dari CO2 Emissions Data by Car in Canada (https://www.kaggle.com/datasets/debajyotipodder/co2-emission-by-vehicles). Dataset terdiri dari 7385 baris dan 12 kolom data.

Berikut adalah variabel dan tipe data yang ada, antara lain:

: Perusahaan dari kendaraan bertipe data object Make

Model : Model mobil bertipe data object

: Kelas kendaraan berdasarkan utilitas bertipe data object Vehicle Class

 Engine Size : Ukuran dari mesin dalam liter bertipe data float64

 Cylinders : Banyaknya silinder bertipe data int64

 Transmission : Tipe transmisi beserta jumlah gigi bertipe data object

Fuel_Type : Tipe bensin yang digunakan bertipe data object

Fuel_Cons City : Konsumsi bensin di jalanan kota (L/100 km) bertipe data float64 Fuel Cons Hwy (I/100 km): Konsumsi bensin di jalan tol (L/100 km) bertipe data float64 Fuel Cons Comb : Konsumsi bensin gabungan (L/100 km) bertipe data float64

 Fuel Cons Comb mpg : Konsumsi bensin gabungan (mpg) bertipe data int64

 CO2_Emissions : Emisi karbon dioksida yang dihasilkan (g/km) untuk gabungan

Variabel Target : CO2 Emissions

Kode:

https://www.kaggle.com/adriaanwaleleng/project-molin-group-h

Bagian 2

Pre-processing dan Analisis Deskriptif

Pada tahap-tahap awal pemrosesan data yang kami lakukan adalah mencari statistik deskriptif dan melakukan eksplorasi untuk melihat pola-pola dalam data agar dapat melakukan pengolahan yang sesuai dengan karakteristik data. Salah satu hasil statistik deskriptif yang berguna dalam tahap preprocessing menuju pembangunan model adalah identifikasi jarak interkuartil serta kuartil 1 dan kuartil 2, yang dimana hal tersebut digunakan untuk menemukan dan menghilangkan outlier. Hal tersebut sangat diperlukan sebab regresi linier sensitif terhadap outlier. Cara kami melakukan eksplorasi adalah dengan melakukan visualisasi dengan berbagai macam jenis. Bagian awal visualisasi adalah melakukan plotting hubungan-hubungan antar variabel yang ada dalam dataset untuk melihat hubungan variabel mana yang dapat dilakukan regresi linier sederhana serta juga menggunakan heatmap untuk melihat korelasi antar variabel.

Kemudian ada juga visualisasi barplot untuk membuat 10 besar setiap objek yang ada untuk mendapatkan gambaran besar tentang mobil-mobil di Kanada. Adapula plot distribusi untuk CO2_Emissions untuk melihat apakah variabel tersebut berdistribusi normal. Kami juga membuat barplot mengenai hubungan variabel-variabel kategorik dengan CO2_Emissions untuk melihat objek mana yang mengeluarkan CO2 terbanyak. Hal tersebut berguna untuk membangun model regresi berganda karena dengan visualisasi tersebut kami dapat melihat hubungan yang mungkin signifikan. Kami mendapatkan bahwa beberapa objek yang sedikit dalam kuantitas menyumbang CO2 dengan proporsi yang besar dibandingkan objek yang lebih banyak kuantitasnya.

Setelah itu kami melakukan feature engineering untuk membuat sebuah fitur baru yang menjadikan satu mobil-mobil yang secara ukuran dan desain sama kedalam variabel kategorik yang sama. Begitupula kami lakukan hal tersebut untuk merek-merek mobil, dimana kami menyatukan merek mobil sesuai dengan rentang harga mobil. Mengingat banyak sekali jenis pada setiap objek untuk, pembuatan fitur esensial dalam proses menuju pembuatan model karena akan mereduksi variabel dengan kepada level yang membuat pembuatan regresi mudah. Selain itu, fitur yang kami buat memberikan gambaran kepada kami tentang bagaimana setiap jenis mobil menyumbang kepada emisi CO2.

Sesuai hasil eksplorasi dan visualisasi kami pada tahap preprocessing, kami berhipotesa bahwa Fuel_Cons_City_(L/100 km), Fuel_Cons_Hwy_(L/100km), Fuel_Cons_Comb_ (L/100km), Vehicle_Class, dan Fuel_Type akan memiliki p-value yang signifikan. Kami juga berpendapat bahwa beberapa variabel kategorik hasil pembuatan fitur seperti Make_Type_Luxury dan Make_Type_Sports akan menyumbang lebih banyak terhadap CO2 dibandingkan dengan variabel lain pada kelasnya yaitu Make_Type. Hal serupa juga kami harapkan terjadi pada objek-objek lain.

Bagian 3

Pemodelan

Kami melakukan regresi linier sederhana univariat untuk: Fuel_Cons_City dan Fuel_Cons_Comb_km, Fuel_Cons_Hwy_(I/100km) dan CO2_Emissions, Fuel_Cons_City dan CO2_Emissions. Kami memilih variabel-variabel ini karena pada hasil visualisasi kami mendapatkan adanya hubungan linier antara variabel-variabel tersebut. Kemudian kami melakukan regresi berganda dimana prediktor adalah semua variabel selain model dan variabel target adalah CO2_Emissions. Kami membuang kolom model karena terlalu banyak model mobil yang ada sehingga pada pembuatan model akan menjadi lebih rumit. Regresi berganda dilakukan dua kali, yang pertama dilakukan tanpa memasukkan variabel kategorik sementara dalam regresi kedua variabel kategorik dimasukkan.

Asumsi-asumsi model:

- 1. Hubungan antar variabel linier
- Error berdistribusi normal
- 3. Homoskedatisitas error
- 4. Observasi yang independen satu dengan lainnya

Model regresi univariat:

- 1. Fuel Cons Comb km = β_0 + β_1 (Fuel Cons City)
- 2. Fuel_Cons_Comb_km = β_0 + β_1 (Fuel_Cons_Hwy_(I/100 km)
- 3. CO2_Emissions = β_0 + β_1 (Fuel_Cons_City)
- 4. CO2_Emissions = β_0 + β_1 (Fuel_Cons_Hwy_(I/100km))

Bagian 4. Pengolahan data dan analisis hasil

Pada dasarnya kami ingin mengetahui jenis mobil, bahan bakar, dan komponen mobil yang menjadi penyumbang terbesar dalam emisi CO2 berdasarkan dataset yang kami dapatkan. Analisis kami berpusat kepada signifikansi objek dalam dataset dalam emisi CO2. Sebelum kami mendapatkan hasil dari regresi, kami sudah bisa memperikirakan hubungan antara variabel-variabel prediktor dan variabel target seperti di Fig.5 yang ada di lampiran dibawah. Kami mendapatkan bahwa sebagian kecil jenis mobil, jenis bahan bakar, transmisi, merk mobil, dan model mobil bertanggung jawab atas sebagian besar emisi CO2. Kami memperkirakan

fakta tersebut terefleksikan pada model regresi linier berganda yang kami buat. Dapat dilihat bahwa beberapa fitur yang kami buat jauh lebih banyak memuat jenis objek yang memiliki sumbangan besar terhadap emisi CO2 dibandingkan fitur lainnya. Misalnya, kebanyakan merek mobil (variabel 'Make') yang masuk kedalam 10 besar merek dengan emisi terbanyak, termasuk dalam kategori mobil sports dan luxury (kategori setelah adanya Make Type).

Hasil regresi linier berganda yang pertama dan kedua memiiliki perbedaan dalam nilai koefisien determinasi (R-Squared), yang berarti adanya variabel kategorik dalam model membuat variabel-variabel independen lebih dapat menjelaskan variasi dalam variabel target yaitu emisi CO2. Pada awalnya nilai R-Squared adalah 0.864 namun setelah dimasukkan variabel kategorik menjadi 0.994. Hal tersebut dapat diartikan bahwa variabel-variabel kategorik yang ditambahkan memiliki dampak yang cukup signifikan dalam menjelaskan emisi CO2. Dapat dilihat pada Fig.9 bahwa hasil regresi pertama menunjukkan bahwa Fuel_Cons_City, Fuel_Cons_Hwy_(I/100 km), Fuel_Cons_Comb_km tidak signifikan (meninjau p value) tetapi pada regresi kedua ketiga variabel tersebut menjadi signifikan. Artinya efek dari ketiga variabel tersebut diberikan variabel-variabel lain berdampak signifikan dalam menjelaskan variabel target.

Hipotesa kami pada Bagian 2 juga terpenuhi semuanya, semua variabel untuk Vehicle_Class dan Fuel_type memiliki p-value<0.05 yang berari mereka signifikan. Semua merek mobil (Make_Type) memiliki efek yang signifikan terhadap variabel target. Hanya saja perkiraan awal kami bahwa sebagian dari koefisien untuk kategori merek mobil yang berada diantara penyumbang emisi CO2 terbesar tidak lebih besar dari koefisien kategori merek mobil yang bukan diantara penyumbang emisi CO2 terbesar. Bahkan, koefisien variabel untuk mobil yang kelas menengah (common) lebih besar daripada untuk mobil-mobil dengan kelas harga yang diatas yaitu luxury, premium, dan sports. Semua jenis bahan bakar kecuali N memiliki efek signifikan. Lihat Fig.10 untuk hasil regresi linier berganda kedua.

Pada hasil kedua dituliskan bahwa salah satu nilai eigen ada yang mendekati nol sehingga ada kemungkinan sangat besar adanya multikolinearitas antar variabel. Untuk permasalahan ini, kami belom bisa memberikan jawaban yang pasti. Walaupun perkiraan kasar kami adalah karena kami sedang mengolah data mobil, setiap komponen akan kemungkinan besar memiliki keterkaitan dengan komponen lain. Misalnya, mobil sports akan menggunakan jenis bahan bakar dan transmisi tertentu yang bisa jadi mengeluarkan CO2 lebih banyak dari tipe mobil lainnya sehingga bahan bakar yang diabiskan di jalanan juga akan berbeda. Begitupula mobil yang termasuk dalam kelas luxury akan dibuat menggunakan komponen berbeda dan membutuhkan bahan bakar berbeda dengan mobil yang ada dalam kelas common. Hal ini karena mobil adalah satu kesatuan barang yang disusun oleh banyak sekali komponen yang disesuaikan dengan harga jual dan pasar dari mobil tersebut.

Bagian 5. Penutup

Masalah emisi CO2 oleh mobil adalah masalah yang multivariabel, tidak mungkin kami ingin mengambil kesimpulan mengenai permasalahan emisi tanpa mempertimbangkan keseluruhan variabel yang ada pada dataset. Hal ini tercermin dari perbedaan yang kami dapatkan pada hasil regresi berganda pertama dan kedua. Salah satu tren dari dataset yang penting untuk diperhatikan secara seksama untuk menjawab pertama adalah bagaimana sebagian kecil objek dalam data dapat menyumbang emisi CO2 jauh lebih besar dibandingkan dengan objek lain. Dengan meninjau hal tersebut bisa kami simpulkan bahwa untuk menanggulangi banyaknya emisi CO2 maka penggunaan sebagian kecil dari jenis mobil yang mengemisi CO2 secara berlebih dibanding jenis mobil lainnya sebaiknya harus dikontrol atau dikurangi.

Bagian 6. Lampiran

Fig.1 Statistik Deskriptif

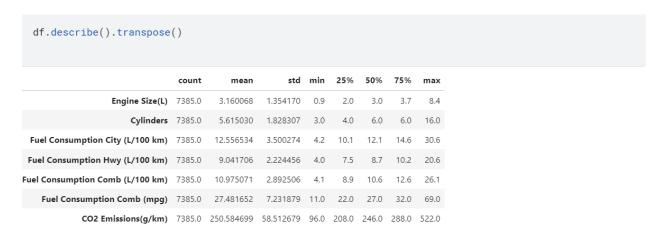


Fig.2 Kode untuk plot hubungan antar variabel

```
sns.pairplot(df)
```

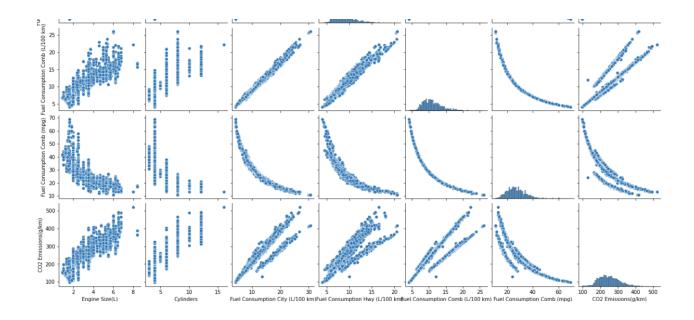


Fig.3 Statistik deskriptif variabel kategorik

df.describe(include = object)

	Make	Model	Vehicle_Class	Transmission	Fuel_Type
count	7385	7385	7385	7385	7385
unique	42	2053	16	27	5
top	FORD	F-150 FFV 4X4	SUV - SMALL	AS6	Х
freq	628	32	1217	1324	3637

Fig.4 10 besar kuantitas terbanyak untuk setiap objek

```
fig, ax=plt.subplots(nrows=2, ncols=3, figsize=(28,10))
for variable, subplot in zip(df_cat_features1.columns, ax.flatten()):
    cat_count=df[variable].value_counts()
    cat_count10=cat_count[:10,]
    d=sns.barplot(cat_count10.values, cat_count10.index, alpha=0.8, ax=subplot)
    if cat_count.size>10:
        d.set_title("Top10{}".format(variable))
    else:
        d.set_title(variable)
    d.set_xlabel("Number of Cars",fontsize=10)
fig.delaxes(ax[1][2])
```

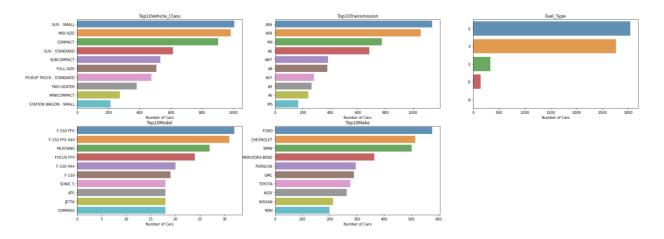


Fig.5 Distribusi CO2

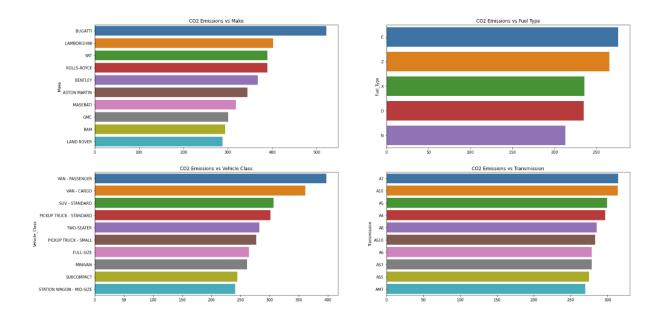
```
sns.distplot(df["CO2_Emissions"], bins=30, kde=True, axlabel="Carbon Dioxide Emissions(30 bins)")

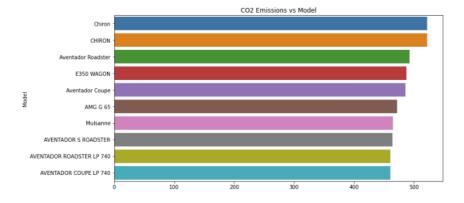
<AxesSubplot:xlabel='Carbon Dioxide Emissions(30 bins)', ylabel='Density'>
0.007
0.006
0.005
```

0.006 - 0.005 - 0.004 - 0.003 - 0.002 - 0.001 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000 - 0.000

Fig.6 Analisa hubungan variabel prediktor dengan variabel target

```
#Analyse relationship between independent and target variables
make_co2=df.groupby('Make')['CO2_Emissions'].mean().sort_values(ascending=False).head(10)
model_co2=df.groupby('Model')['CO2_Emissions'].mean().sort_values(ascending=False).head(10)
vehicle_class_co2=df.groupby('Vehicle_class')['CO2_Emissions'].mean().sort_values(ascending=False).head
transmission_co2=df.groupby('Transmission')['CO2_Emissions'].mean().sort_values(ascending=False).head('fuel_type_co2=df.groupby('Fuel_Type')['CO2_Emissions'].mean().sort_values(ascending=False).head(10)
```





Kode untuk Feature Engineering

#Feature Engineering

df["Vehicle_Class_Type"]=df["Vehicle_Class"].replace(["COMPACT",'MINICOMPACT','SUBCOM PACT'],'Hatchback')

df["Vehicle_Class_Type"]=df["Vehicle_Class_Type"].replace(['MID-SIZE','TWO-SEATER','FULL-SIZE','STATION WAGON - SMALL','STATION WAGON - MID-SIZE'], 'Sedan')

df["Vehicle_Class_Type"]=df["Vehicle_Class_Type"].replace(['SUV - SMALL', 'SUV - STANDARD', 'MINIVAN'], 'SUV')

df["Vehicle_Class_Type"]=df["Vehicle_Class_Type"].replace(['Van - CARGO','VAN - PASSENGER', 'PICKUP TRUCK - STANDARD', 'SPECIAL PURPOSE VEHICLE', 'PICKUP TRUCK - SMALL'], 'Truck')

df['Make_Type']=df["Make"].replace(['BUGATTI','PORSCHE','MASERATI','ASTON
MARTIN','LAMBORGHINI','JAGUAR','SRT'], 'Sports')
df['Make_Type'] = df['Make_Type'].replace(['ALFA ROMEO', 'AUDI', 'BMW', 'BUICK',
'CADILLAC', 'CHRYSLER', 'DODGE', 'GMC','INFINITI', 'JEEP', 'LAND ROVER', 'LEXUS',
'MERCEDES-BENZ','MINI', 'SMART', 'VOLVO'],'Premium')
df['Make_Type'] = df['Make_Type'].replace(['ACURA', 'BENTLEY', 'LINCOLN', 'ROLLS-ROYCE',
'GENESIS'], 'Luxury')
df['Make_Type'] = df['Make_Type'].replace(['CHEVROLET', 'FIAT', 'FORD', 'KIA', 'HONDA',
'HYUNDAI', 'MAZDA', 'MITSUBISHI','NISSAN', 'RAM', 'SCION', 'SUBARU',
'TOYOTA','VOLKSWAGEN'],'Common')

Fig.7 Hasil Feature Engineering

```
df.drop(['Vehicle_Class'], inplace=True, axis=1)

df['Vehicle_Class_Type'].unique()

array(['Hatchback', 'SUV', 'Sedan', 'VAN - CARGO', 'Truck'], dtype=object)

df.drop(['Make'], inplace=True, axis=1)

df['Make_Type'].unique()

array(['Luxury', 'Premium', 'Sports', 'Common'], dtype=object)
```

Fig.8 Identfikasi dan penghapusan outlier

```
##Identify outliers using interquartile range
  Q1 = df_num_features.quantile(0.25)
  Q3 = df_num_features.quantile(0.75)
  IQR = Q3 - Q1
  print(IQR)
Engine Size
                           1.7
                           2.0
Cylinders
Fuel_Cons_City_(1/100km)
                          4.6
Fuel_Cons_Hwy_(1/100km)
                          2.8
Fuel Cons Comb (1/100km)
                          3.8
Fuel Cons Comb mpg
                          10.0
CO2 Emissions
                          81.0
dtype: float64
```

```
outlier = pd.DataFrame((df_num_features < (Q1 - 1.5 * IQR)) | (df_num_features > (Q3 + 1.5 * IQR)))

for i in outlier.columns:
    print('Total number of Outliers in column {} are {}'.format(i, (len(outlier[i] == True][i])))

Total number of Outliers in column Engine_Size are 121
Total number of Outliers in column Cylinders are 177
Total number of Outliers in column Fuel_cons_City_(I/100km) are 116
Total number of Outliers in column Fuel_cons_Comb_(I/100km) are 150
Total number of Outliers in column Fuel_cons_Comb_(I/100km) are 115
Total number of Outliers in column Fuel_cons_Comb_mpg are 95
Total number of Outliers in column CO2_Emissions are 74

##Outlier Removal
df = df[~((df < (Q1 - 1.5 * IQR)) | (df > (Q3 + 1.5 * IQR))).any(axis=1)]
df.reset_index(inplace=True, drop=True)
```

Fig.9 Hasil regresi berganda pertama

Dep. Variable:	CO2_Emissions	R-squared:	0.846
Model:	OLS	Adj. R-squared:	0.846
Method:	Least Squares	F-statistic:	1.068e+04
Date:	Fri, 11 Nov 2022	Prob (F-statistic):	0.00
Time:	05:20:11	Log-Likelihood:	-25580.
No. Observations:	5816	AIC:	5.117e+04
Df Residuals:	5812	BIC:	5.119e+04
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Fuel_Cons_City	1.4959	3.374	0.443	0.657	-5.118	8.109
Fuel_Cons_Hwy_(l/100km)	1.1656	2.781	0.419	0.675	-4.287	6.618
Fuel_Cons_Comb_km	16.7421	6.114	2.738	0.006	4.756	28.729
const	38.2199	1.256	30.432	0.000	35.758	40.682

Fig.10 Hasil regresi berganda kedua

OLS Regression Results

Dep. Variable: CO2_Emissions R-squared: 0.994 Model: Adj. R-squared: OLS 0.994 **F-statistic:** 1.752e+04 Method: Least Squares Date: Fri, 11 Nov 2022 Prob (F-statistic): 0.00 Log-Likelihood: Time: 05:20:11 -11160. No. Observations: **AIC:** 2.240e+04 4071 Df Residuals: **BIC:** 2.267e+04 4029

41

Covariance Type: nonrobust

Df Model:

	coef	std err	t	P> t	[0.025	0.975]
const	17.4378	1.458	11.960	0.000	14.579	20.296
Engine_Size	0.4907	0.169	2.910	0.004	0.160	0.821
Cylinders	-0.1571	0.129	-1.217	0.224	-0.410	0.096
Fuel_Cons_City	5.7088	0.803	7.106	0.000	4.134	7.284
Fuel_Cons_Hwy_(l/100km)	5.6272	0.669	8.413	0.000	4.316	6.939
Fuel_Cons_Comb_km	9.7668	1.458	6.700	0.000	6.909	12.625
Fuel_Cons_Comb_mpg	-0.8497	0.044	-19.161	0.000	-0.937	-0.763
Vehicle_Class_Type_Hatchback	3.7817	0.422	8.971	0.000	2.955	4.608
Vehicle_Class_Type_SUV	3.0063	0.398	7.558	0.000	2.226	3.786
Vehicle_Class_Type_Sedan	3.6887	0.416	8.860	0.000	2.872	4.505
Vehicle_Class_Type_Truck	3.4029	0.425	8.006	0.000	2.570	4.236
Vehicle_Class_Type_VAN - CARGO	3.5582	1.352	2.632	0.009	0.907	6.209
Make_Type_Common	4.7072	0.400	11.774	0.000	3.923	5.491
Make_Type_Luxury	4.6821	0.476	9.830	0.000	3.748	5.616
Make_Type_Premium	3.9201	0.378	10.384	0.000	3.180	4.660
Make_Type_Sports	4.1284	0.425	9.723	0.000	3.296	4.961
Fuel_Type_D	53.8406	0.484	111.333	0.000	52.893	54.789
Fuel_Type_E	-80.3913	0.582	-138.146	0.000	-81.532	-79.250

Fuel_Type_N	2.763e-17	8.3e-16	0.033	0.973	-1.6e-15	1.65e-15
Fuel_Type_X	22.2686	0.339	65.715	0.000	21.604	22.933
Fuel_Type_Z	21.7200	0.352	61.728	0.000	21.030	22.410
Transmission_A10	3.9844	0.829	4.805	0.000	2.359	5.610
Transmission_A4	-0.4602	0.809	-0.569	0.570	-2.047	1.126
Transmission_A5	-0.7299	0.539	-1.353	0.176	-1.787	0.328
Transmission_A6	-0.5835	0.292	-1.996	0.046	-1.157	-0.010
Transmission_A7	1.4000	0.794	1.763	0.078	-0.157	2.957
Transmission_A8	1.5705	0.327	4.797	0.000	0.929	2.212
Transmission_A9	2.4645	0.347	7.100	0.000	1.784	3.145
Transmission_AM5	3.2718	2.135	1.533	0.125	-0.913	7.457
Transmission_AM6	1.0674	0.509	2.098	0.036	0.070	2.065
Transmission_AM7	0.9876	0.334	2.959	0.003	0.333	1.642
Transmission_AM8	1.2544	0.695	1.806	0.071	-0.108	2.616
Transmission_AM9	-8.392e-15	4.36e-15	-1.923	0.055	-1.69e-14	1.63e-16
Transmission_AS10	1.6629	0.427	3.891	0.000	0.825	2.501
Transmission_AS4	-3.3211	3.639	-0.913	0.361	-10.455	3.813
Transmission_AS5	-1.8590	0.913	-2.037	0.042	-3.648	-0.070
Transmission_AS6	0.3523	0.252	1.400	0.162	-0.141	0.846