

ANALISIS HARGA MOBIL BEKAS DI ARAB SAUDI

● BUSINESS PRESENTATION

TABLE OF CONTENT

PITCH DECK PRESENTATION

01 Business Problem

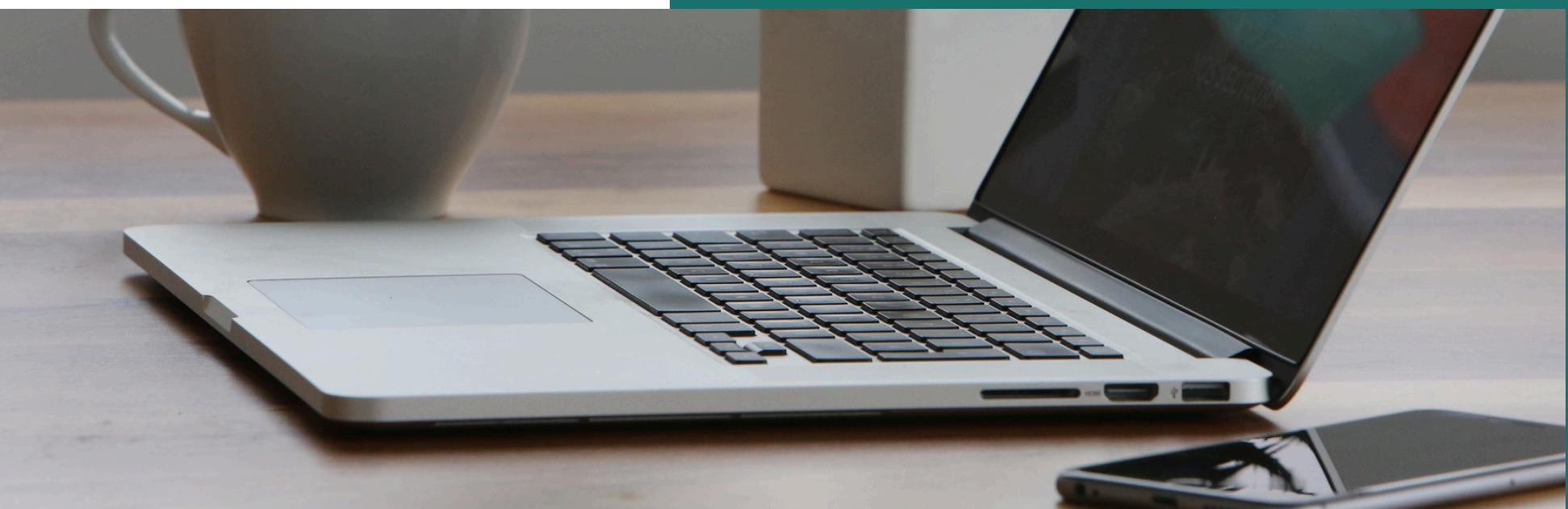
02 Data Understanding

03 Data Preprocessing

04 Modelling

05 Conclusion

06 Recommendation





TOPIK PERMASALAHAN

REFERENSI DATA : PURWADHIKA

Penggunaan mobil di Arab Saudi terkait dengan analisis dan prediksi harga mobil bekas yang diiklankan di syarah.com. Tujuan Analisis:

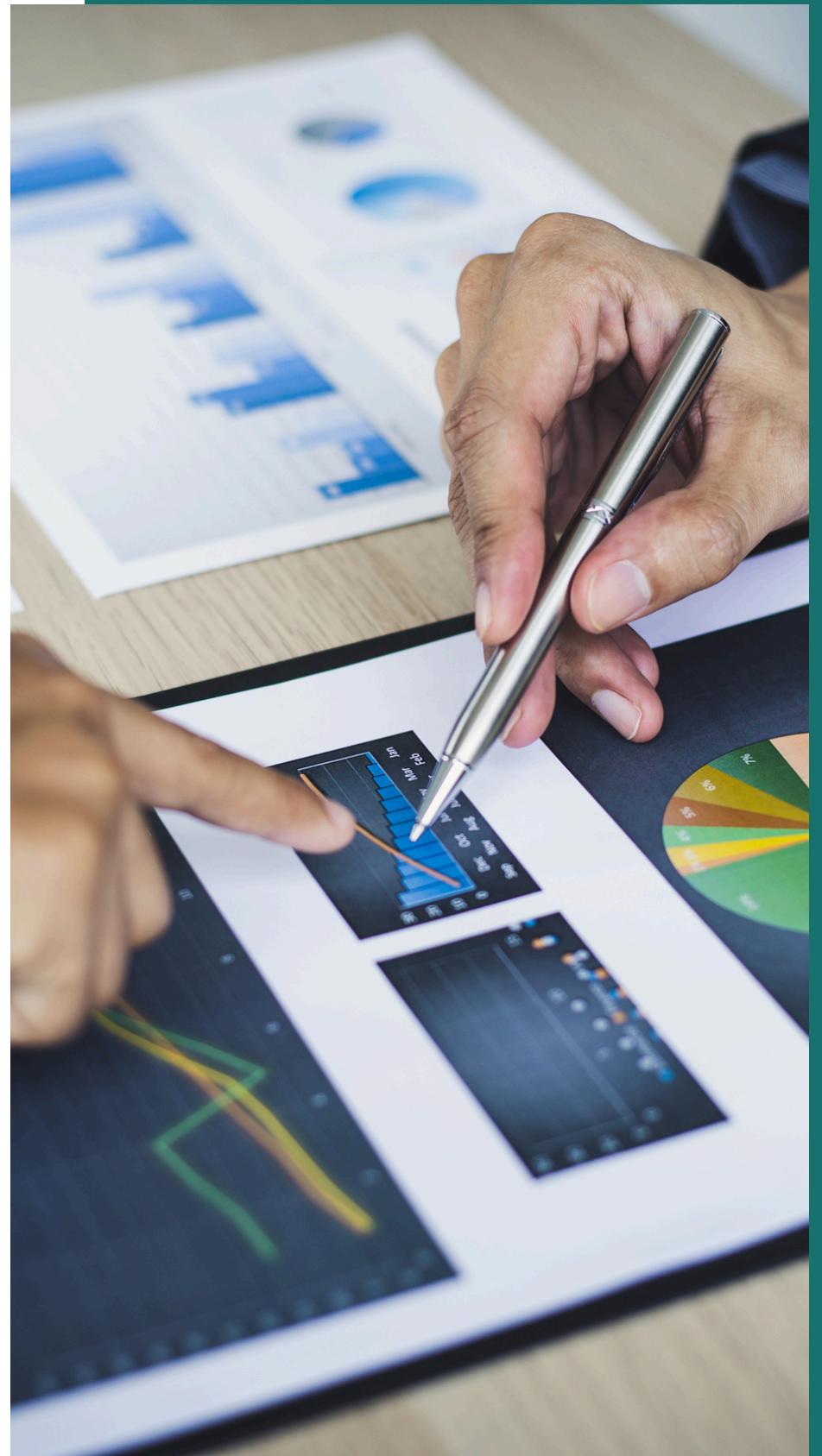
1. Penetapan Harga yang Kompetitif
2. Identifikasi Faktor yang Mempengaruhi Harga
3. Optimasi Stok dan Penawaran
4. Negosiasi Harga
5. Segmentasi Pasar

Thank you

DATA UNDERSTANDING

Attribute	Data Type	Description
Type	Object	jenis mobil bekas yang dicatat dalam dataset
Region	Object	lokasi di mana mobil bekas tersebut ditawarkan untuk dijual
Make	Object	merek mobil bekas
Gear_Type	Object	tipe transmisi mobil
Origin	Object	asal negara atau wilayah tempat mobil diproduksi
Options	Object	opsi tambahan atau fitur khusus yang ada pada mobil
Years	Integer	tahun di mana mobil diproduksi
Engine_Size	Float	ukuran mesin mobil dalam liter atau cc
Mileage	Integer	jumlah jarak yang telah ditempuh oleh mobil dalam satuan kilometer atau mil
Negotiable	Boolean	indikator biner (True/False) yang menunjukkan apakah harga mobil bisa dinegosiasikan atau tidak
Price	Integer	harga jual dari mobil bekas dalam dataset

GET TO KNOW US BETTER



DATASHEET

BY ; PURWADHIKA

DESKRIPSI DATASHEET

Deskripsi Dataset:

- Jumlah Entri: 5624
- Kolom: 11
- Fitur:
 - Kategorikal: Type, Region, Make, Gear_Type, Origin, Options
 - Numerik: Year, Engine_Size, Mileage, Price

EKSPLORASI DATA

Statistik Deskriptif:

- Harga Mobil: Rata-rata, median, dan rentang harga
- Mileage: Rentang dan distribusi
- Tahun Pembuatan: Rentang tahun pembuatan

FILTERING DATA

PROSES SELEKTIF DALAM MEMILIH, MENYARING,
ATAU MEMODIFIKASI DATASET.

Data yang Difilter:

- Kriteria: Harga > 100.000 dan Mileage < 50.000
- Jumlah Data Setelah Filtering: 340

Contoh Mobil yang Ditemukan:

- Model: Toyota, Mercedes, BMW, dll.
- Tahun Pembuatan: 2013-2021

Mobil bekas yang relatif baru dan dengan harga tinggi, namun memiliki jarak tempuh yang rendah. Ini mungkin mengindikasikan mobil-mobil premium atau dalam kondisi sangat baik, dan dapat menarik minat pembeli yang mencari mobil bekas dengan harga tinggi tetapi dengan penggunaan yang minimal.

FEATURE SELECTION AND ENGINEER

Feature Selection

Year: Memiliki skor tertinggi, menunjukkan bahwa tahun pembuatan mobil adalah faktor yang sangat penting dalam menentukan harga mobil.

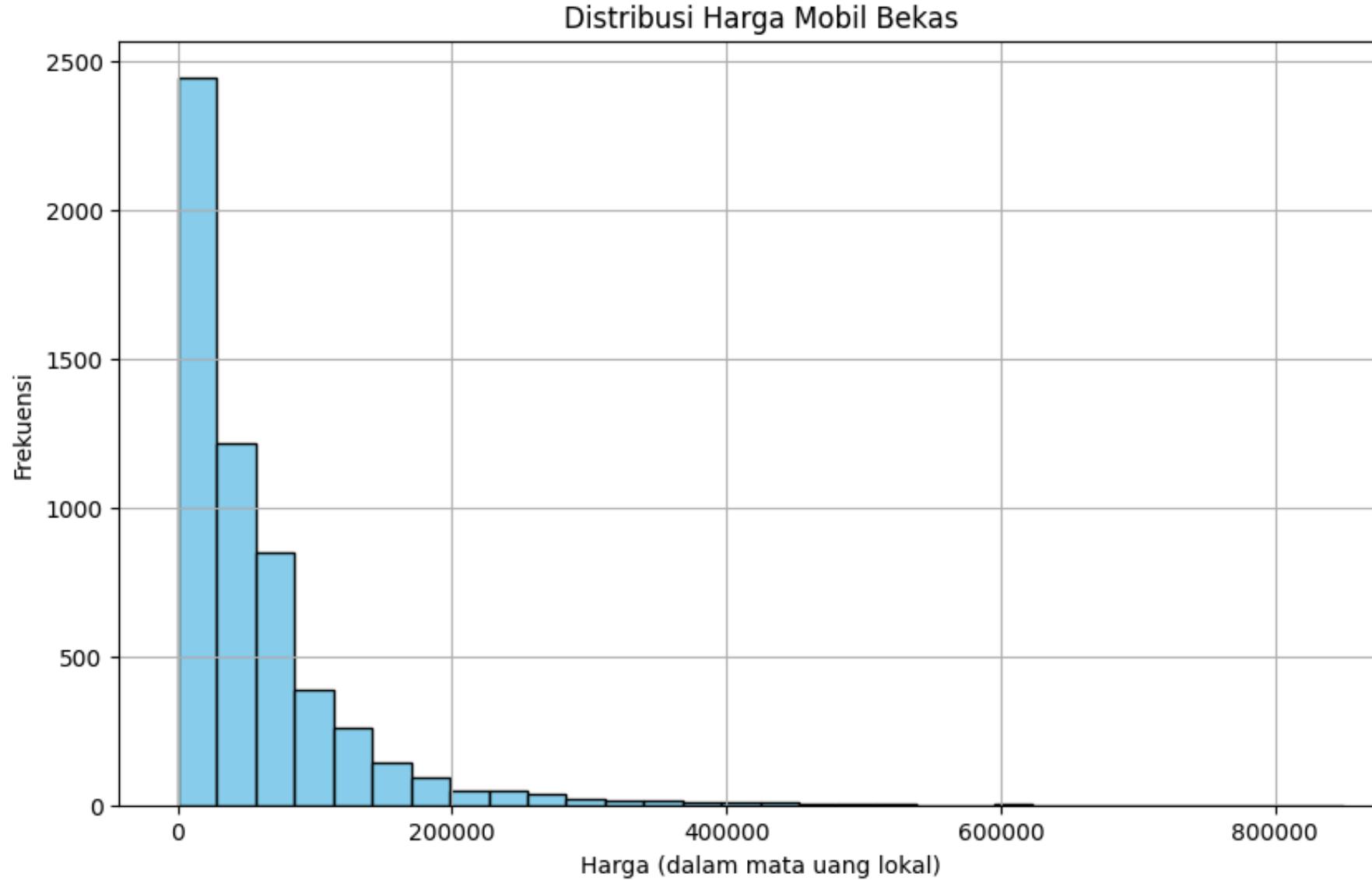
Make_Mercedes, Make_Land Rover, dan fitur-fitur lain terkait merek mobil juga memiliki skor yang tinggi, menunjukkan bahwa merek mobil sangat mempengaruhi harga.

Feature Engineering

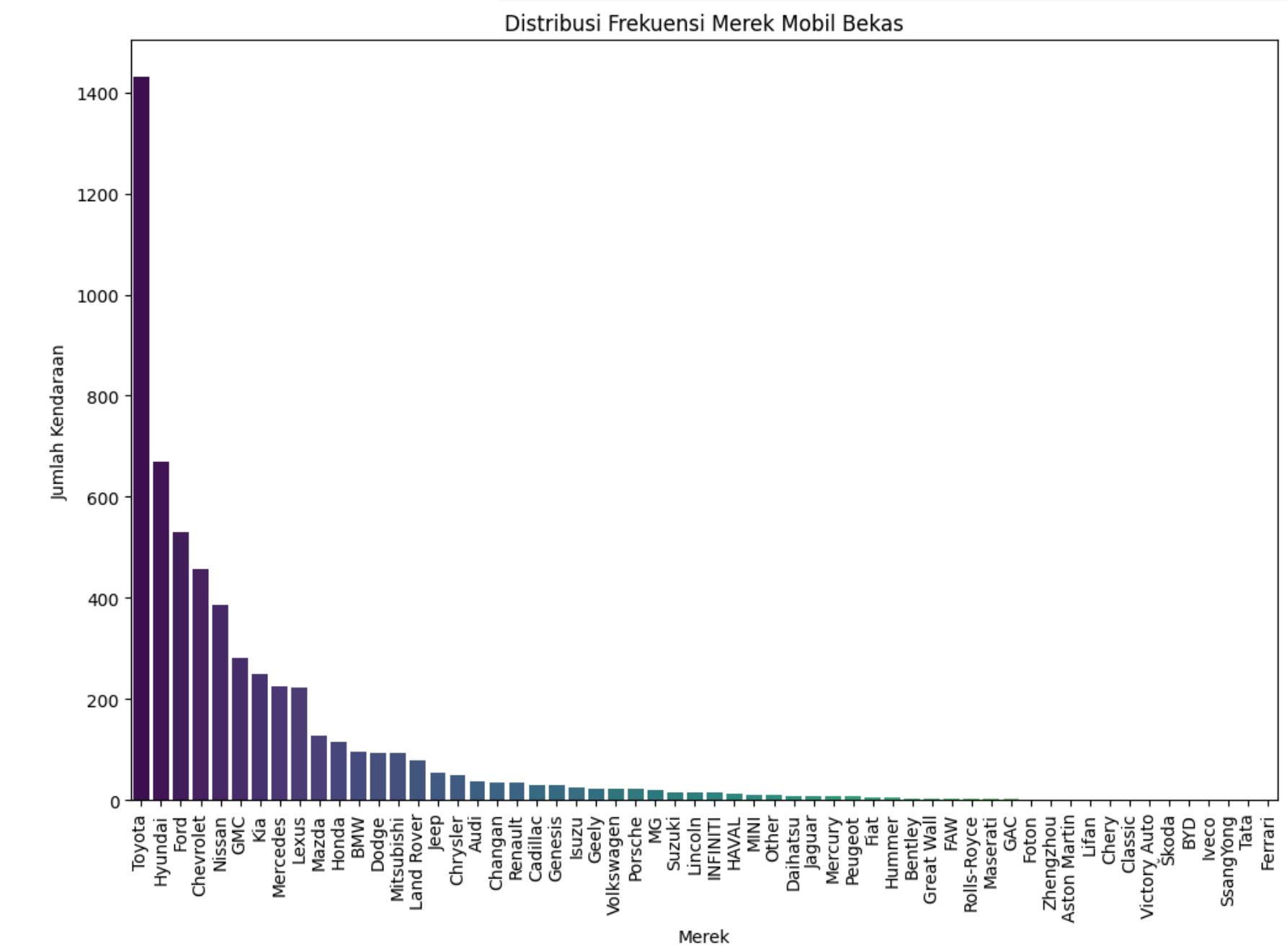
Year dan Car_Age: Kedua fitur ini memiliki skor yang sangat tinggi dan sama (420.87)

DATA ANALYSIS

01 Distribusi Harga Mobil Bekas



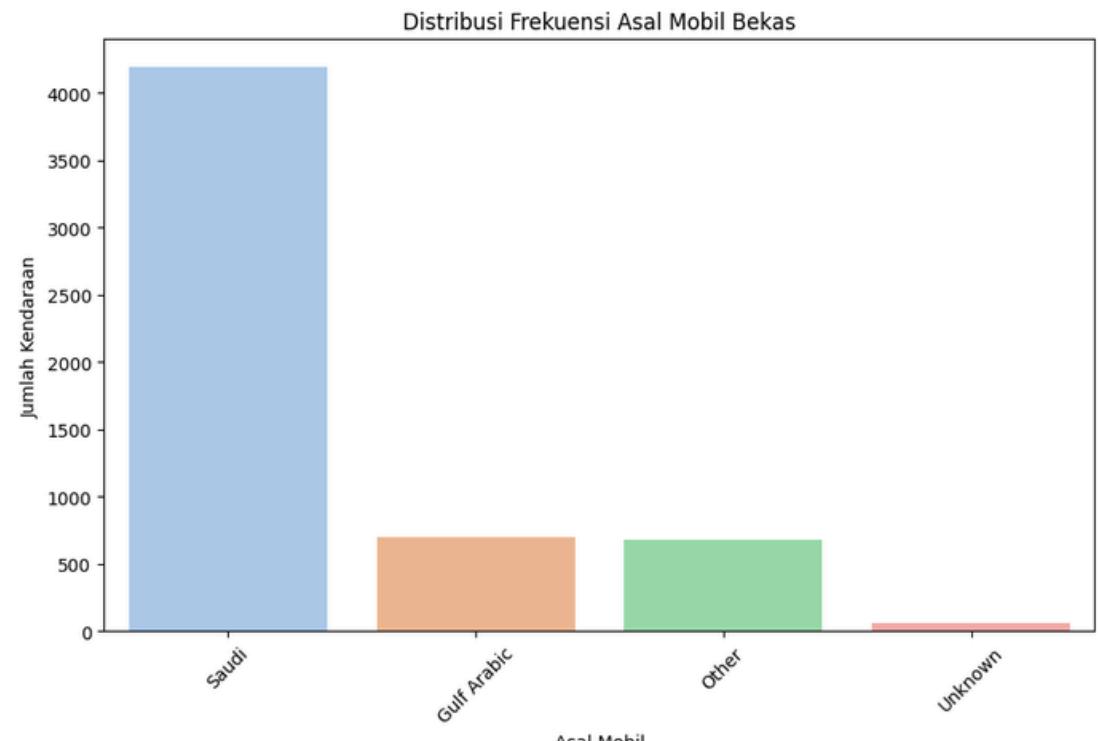
02 Distribusi Merk Mobil



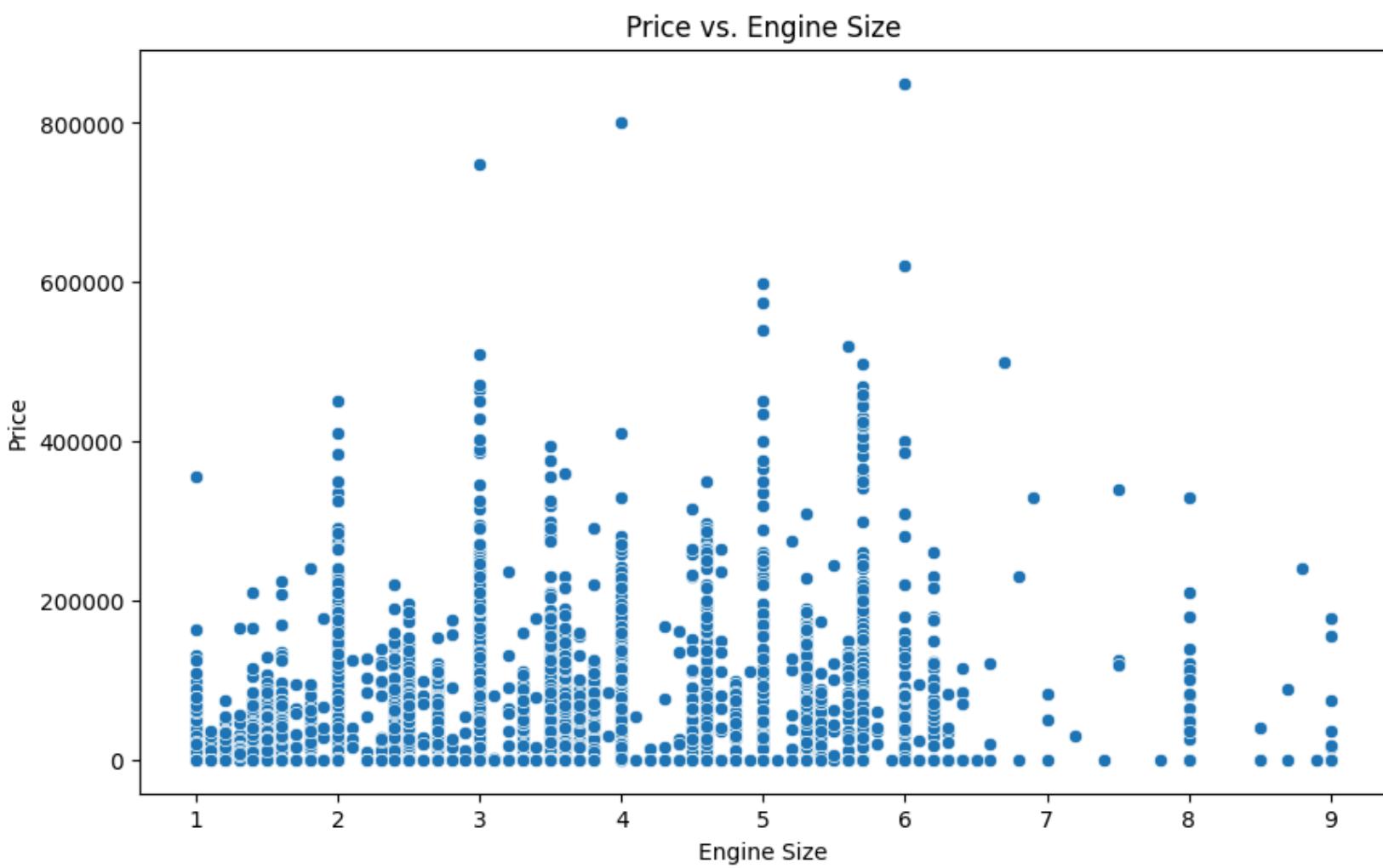
DATA ANALYSIS



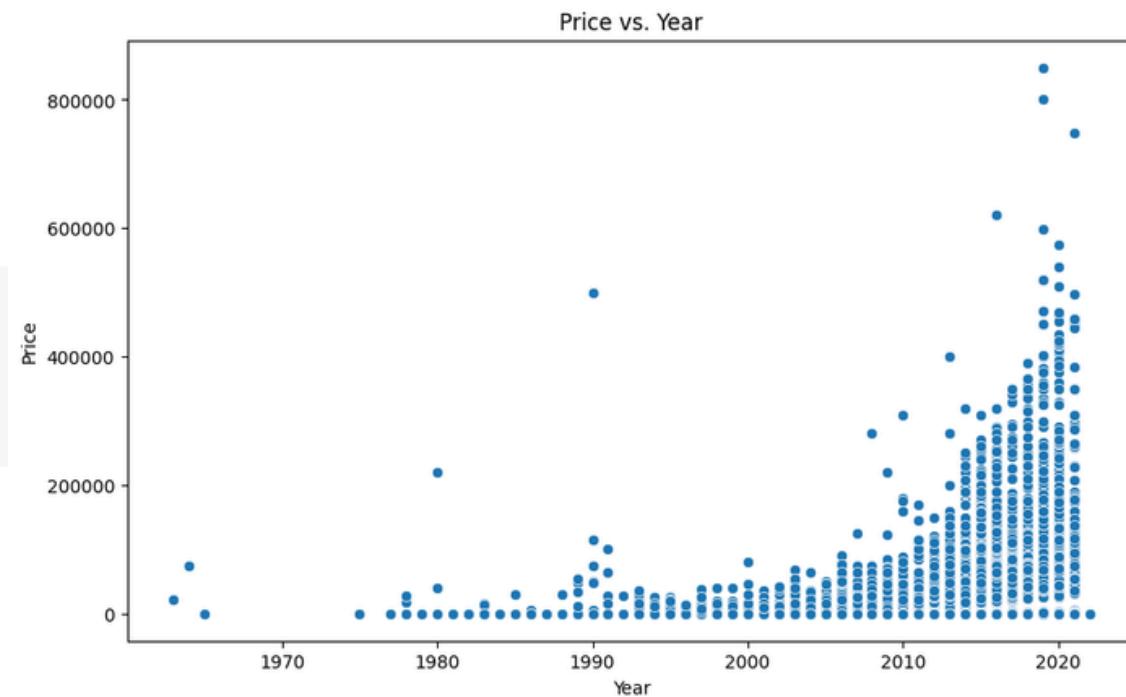
03 Distribusi Frekuensi Asal Mobil



04 Price vs Engine Size



05 Price vs Year

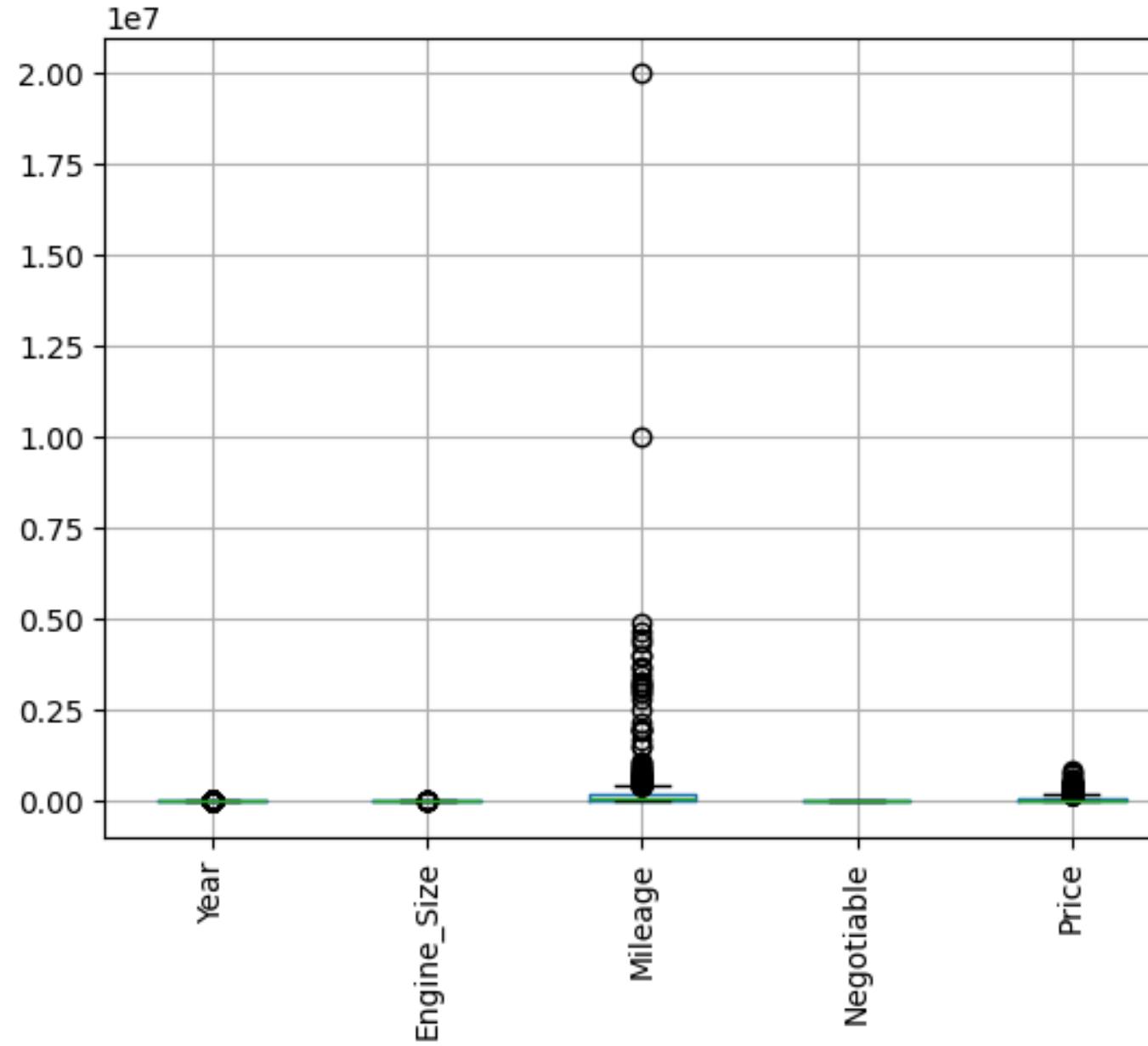


DATA OUTLIER

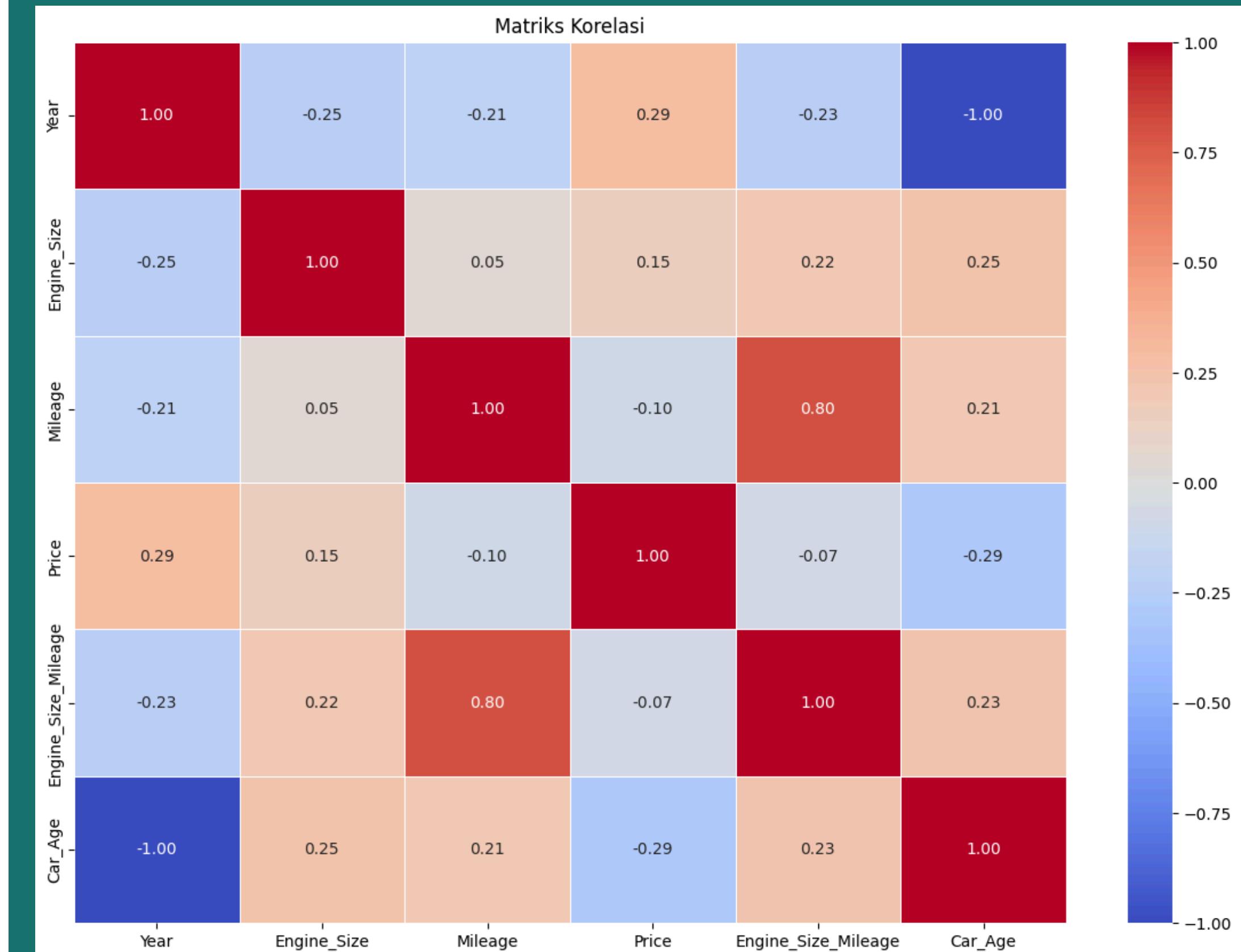
Analisis Outliers

Outliers dalam Fitur Mileage:

- Jumlah Outliers: 175
- Contoh: Mobil dengan mileage yang sangat tinggi



MATRIC KORELASI





MODEL BENCHMARKING

PEMODELAN YANG DILAKUKAN UNTUK EVALUASI TEHADAP MODEL YANG LEBIH KOMPLEKS. SEHINGGA PADA PROSESINI AKAN MEMBERIKAN TITIK REFERENSI YANG KEMUDIAN DILAKUKAN PERBANDINGAN

Hasil Benchmark Model:

- Model yang Diuji: Linear Regression, Decision Tree, Random Forest, SVR, KNN
- Metrik: RMSE, MAE, MAPE
- Hasil Terbaik: Model Random Forest (dengan MAPE sangat tinggi menunjukkan potensi masalah dalam model)

Model	RMSE	MAE	MAPE
Linear Regression	86734.022929	51356.095851	inf
Decision Tree	85400.840018	49723.382989	inf
Random Forest	81401.544438	47529.409720	inf
Support Vector Machine	63603.620979	36680.322218	inf
K-Nearest Neighbors	80675.693651	45864.323918	inf

HYPERPARAMETER TUNING

PROSES UNTUK MENGOPTIMALKAN PARAMETER MODEL GUNA MENDAPATKAN KINERJA TERBAIK.

Tuning Model BaggingRegressor:

- Model: Bagging dengan Linear Regression sebagai estimator
- Parameter yang Diuji:
 - n_estimators
 - max_samples
 - max_features
- Hasil: Parameter terbaik dan metrik evaluasi (RMSE dan MAE)



Parameter Terbaik	RMSE	MAE	MAPE
{'max_features': 0.5, 'max_samples': 0.7, 'n_estimators': ...}	5.05296	4.635694	1.053792e+1 6

max_features: nilainya adalah 0.5, artinya 50% dari fitur digunakan.

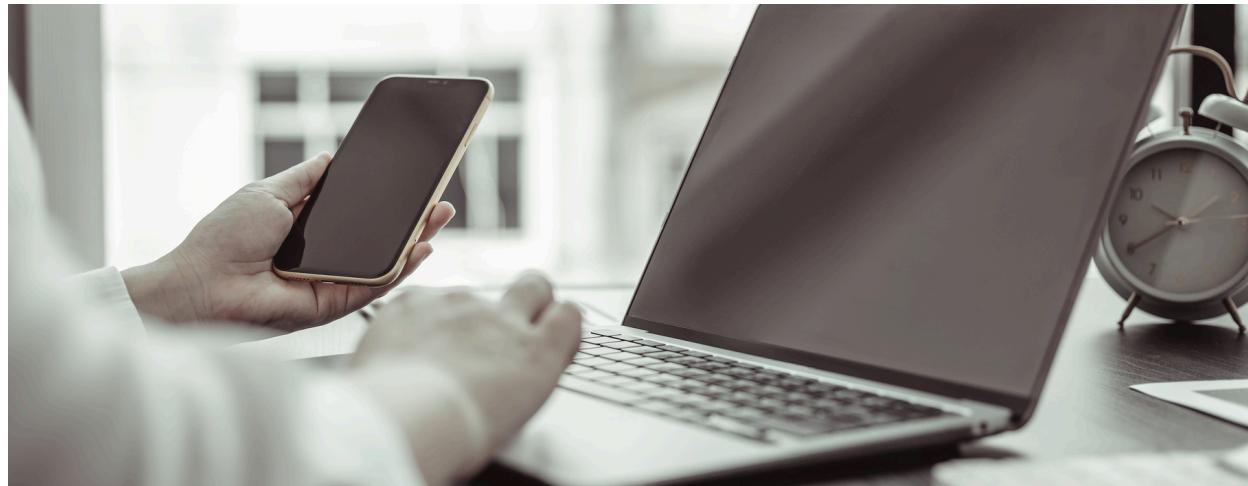
max_samples: nilainya adalah 0.7, artinya 70% dari data digunakan untuk setiap model.

n_estimators: Jumlah model dasar (regresi linier) dalam ensemble.

BaggingRegressor dilakukan metode ensemble yang menggunakan teknik Bootstrap Aggregating (Bagging) untuk mengurangi variabilitas dan overfitting

Tuning hyperparameter memberikan insight yang lebih baik tentang parameter optimal untuk BaggingRegressor.

CONCLUSION



- Benchmarking menggunakan 5 model regresi (Linier, pohon keputusan, random forest, support vector dan K-Nearest). Hasil nilai dari Regresi Random forest didapatkan nilai RMSE sebesar 81.401,54 dan MAE 47.529,41. Kondisi ini termasuk dalam kondisi akurat data
- Adanya fitur Engine_size_Mileage dan Car_age mampu meningkatkan kinerja keseluruhan model dalam hubungan atribut matrik
- Year dan Car_age berpengaruh terhadap prediksi harga jual mobil
- Oulier tertinggi pada Mileage, dimana perlu dilakukan pemodelan prediksi dan penghapusan sebagian outlier
- Bagging dengan Regresi Linier didapatkan hasil pada Grid search mendapat parameter max_features = 0.5 dan max_samples=0.7 serta N-estimators=30. Hal ini nilai RMSE = 5.05, MAE = 4.64
- Rekayasa prediksi terhadap pemilihan model, rekayasa fitur, dan setelan hyperparameter dinilai lebih prediktif dalam menentukan estimasi harga mobil

RECOMENDATION



- Untuk meningkatkan kualitas Data, maka perlu adanya pengaturan terhadap outlier data. dalam kasus ini adalah Mileage. serta melakukan filtering dan cleaning data agar tidak terjadi data ganda dan data 0
- Penggunaan rekayasa fitur tambahan dalam mobil dapat meningkatkan harga. Dalam konteks ini bisa menawarkan servis tambahan berupa servis atau penambahan hal lainnya
- Perlu dilakukan uji terhadap model gradient boosting dan XGboost agar menambah keakuratan data
- Menetapkan harga mobil yang kompetitif agar menarik pelanggan dan menjaga keuntungan. Dalam hal ini bisa dilakukan dengan prediksi harga
- Berfokus pada Year, Car_age, dan Make dalam melakukan pembelian dan penjualan karena hal ini dirasa mampu memberikan gambaran secara umum kebutuhan
- Melakukan analisis lebih lanjut dengan mempertimbangkan pada wilayah dan segmen konteks ini disesuaikan dengan strategi harga berdasarkan wilayah