

USED CAR SALES

IN SAUDI ARABIA

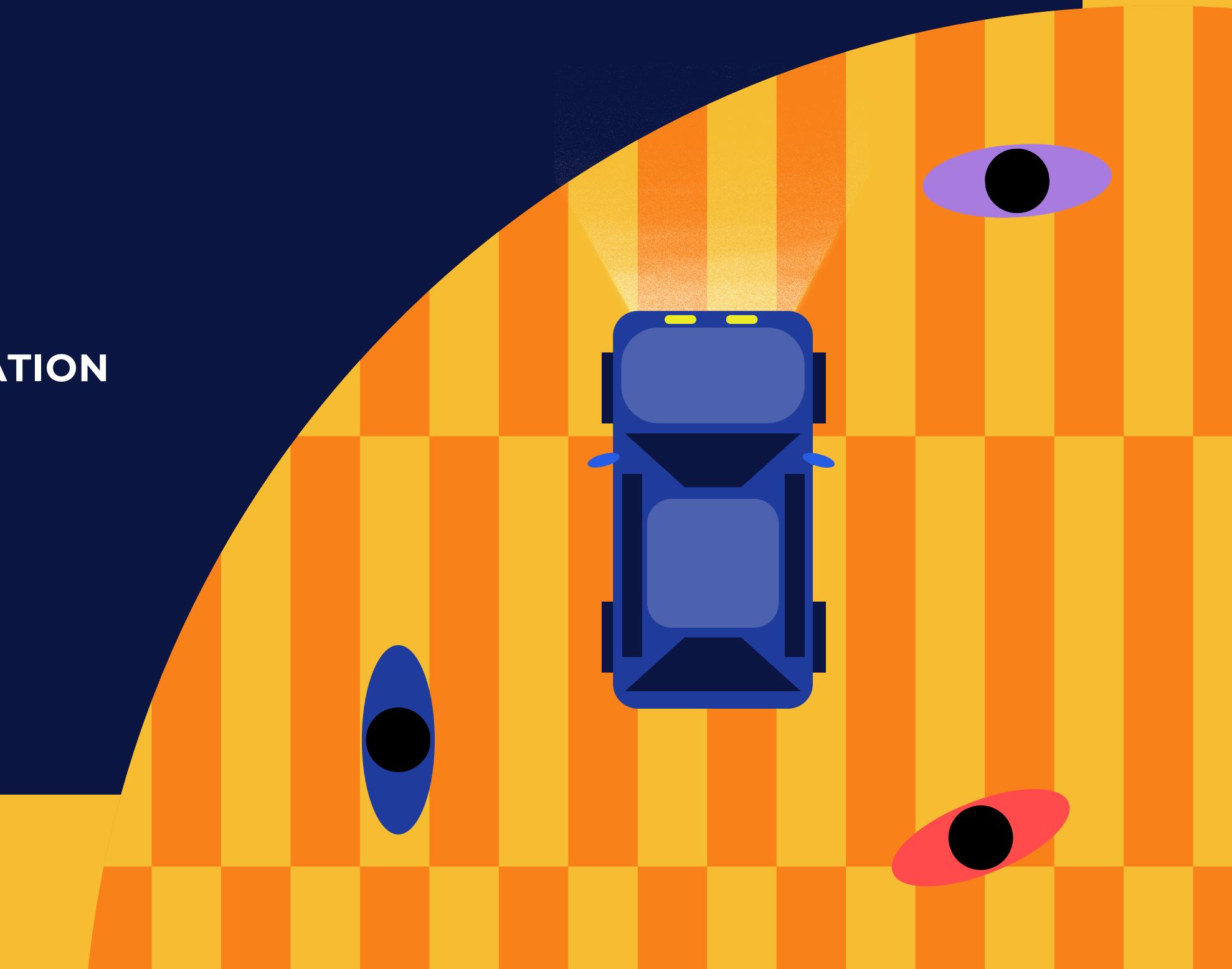
This presentation will complete the Capstone Project Module 3 assignment at Purwadhika.

Presented by:
Justify Ester Pasaribu - JCDS 0406 006



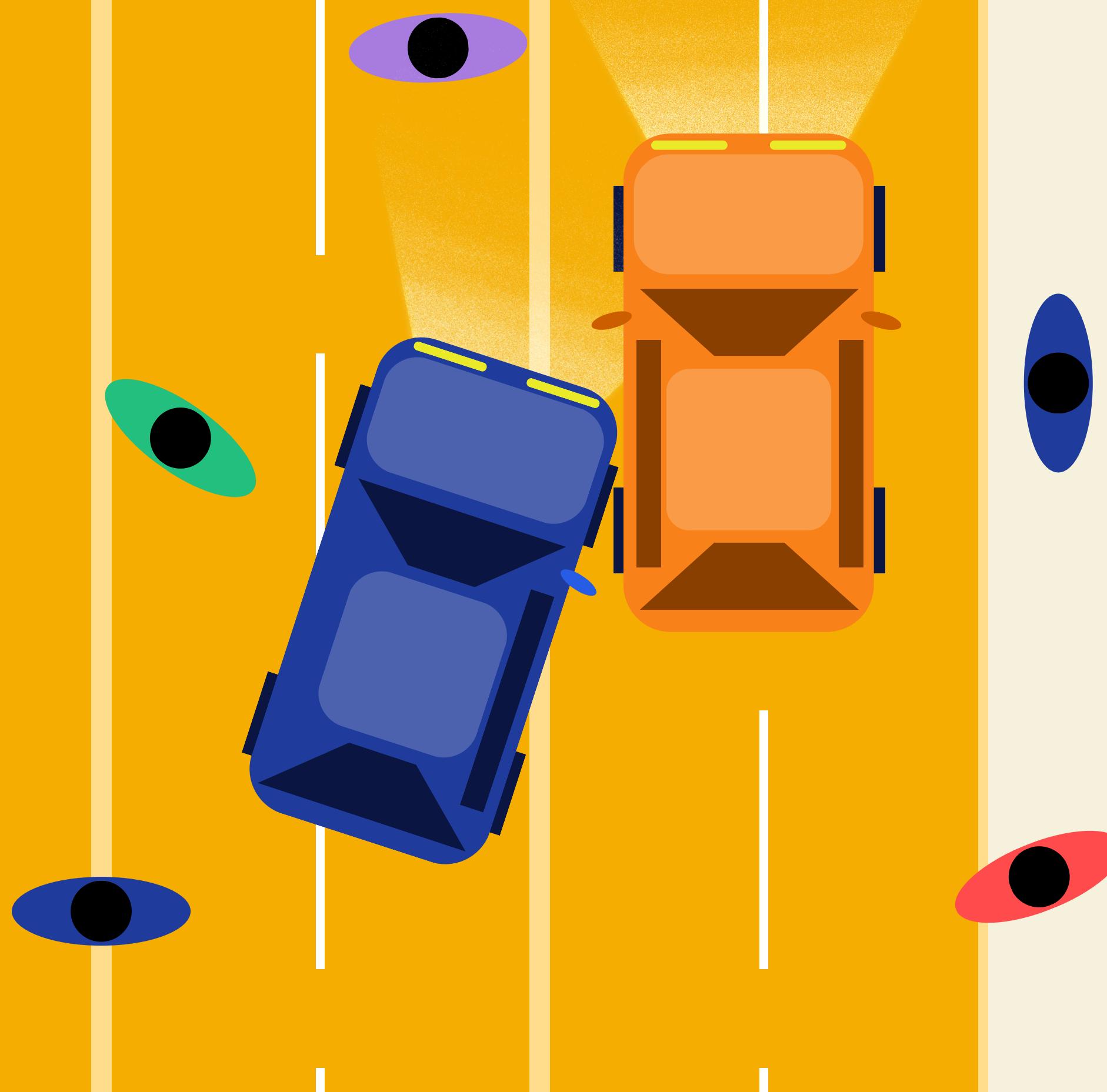
TABLE OF CONTENTS

- 1 BUSINESS PROBLEM
- 2 DATA UNDERSTANDING
- 3 DATA PREPROCESSING
- 4 MODELING
- 5 CONCLUSION
- 6 RECOMMENDATION





01 - **BUSINESS PROBLEM UNDERSTANDING**

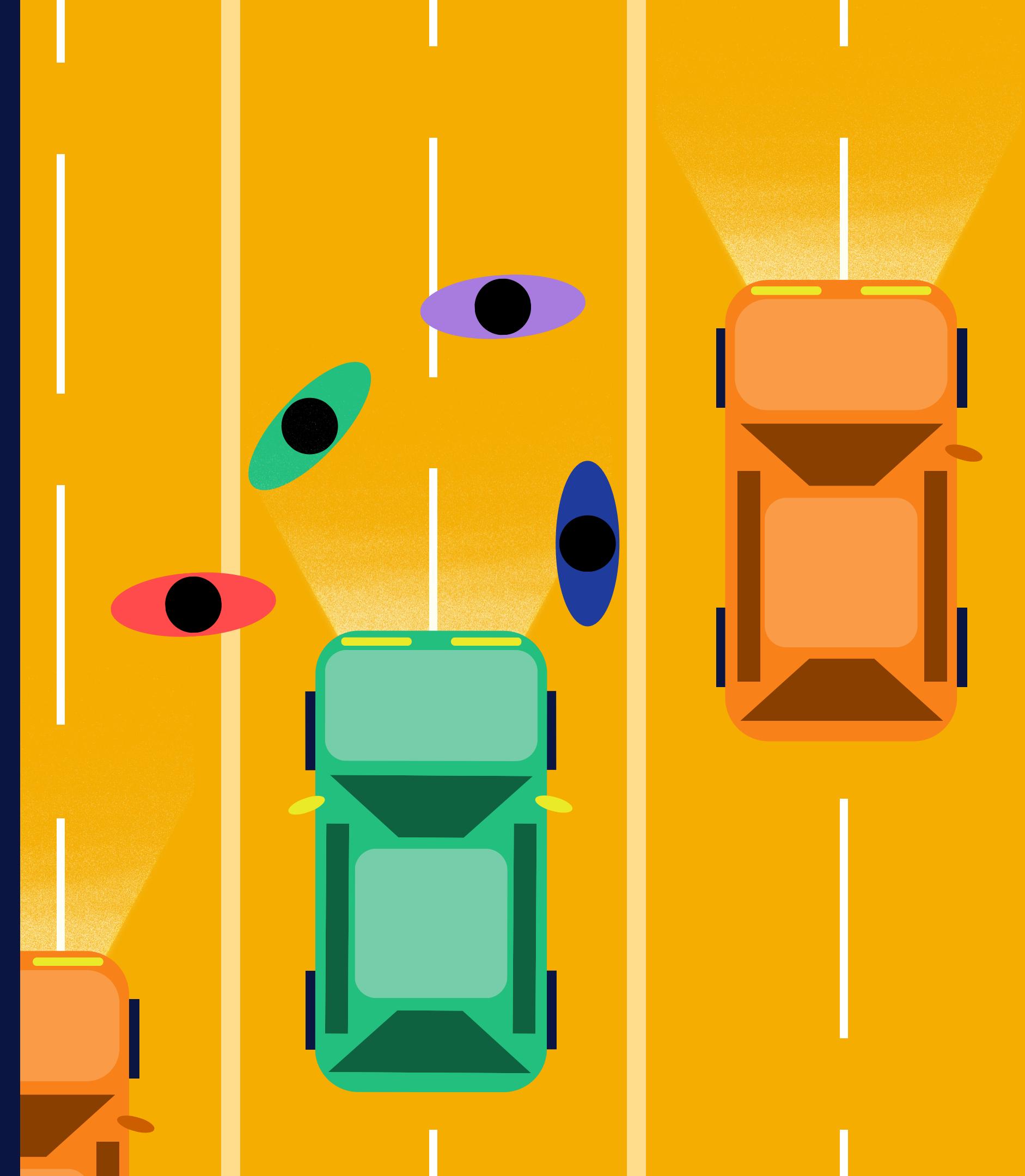




CONTEXT

Arab Saudi, sebagai salah satu negara terkaya di Timur Tengah, memiliki potensi pasar mobil bekas yang signifikan. Dengan peningkatan pendapatan yang tersedia, jumlah pengemudi wanita yang meningkat (sebelumnya dilarang), dan berkembangnya pasar online untuk jual beli mobil bekas, penjualan mobil bekas di negara tersebut diperkirakan akan tumbuh secara substansial. Dari tahun 2019 hingga 2025, nilai (Gross Transaction Value) penjualan mobil bekas diproyeksikan meningkat hingga 4,5%, sementara volume penjualan meningkat hingga 2%. Pandemi COVID-19 telah menyebabkan peningkatan stok mobil bekas dan pengurangan pendapatan masyarakat, khususnya dengan kenaikan pajak pertambahan nilai dari 5% menjadi 15% pada tahun 2020, mendorong calon pembeli untuk lebih tertarik pada mobil bekas.

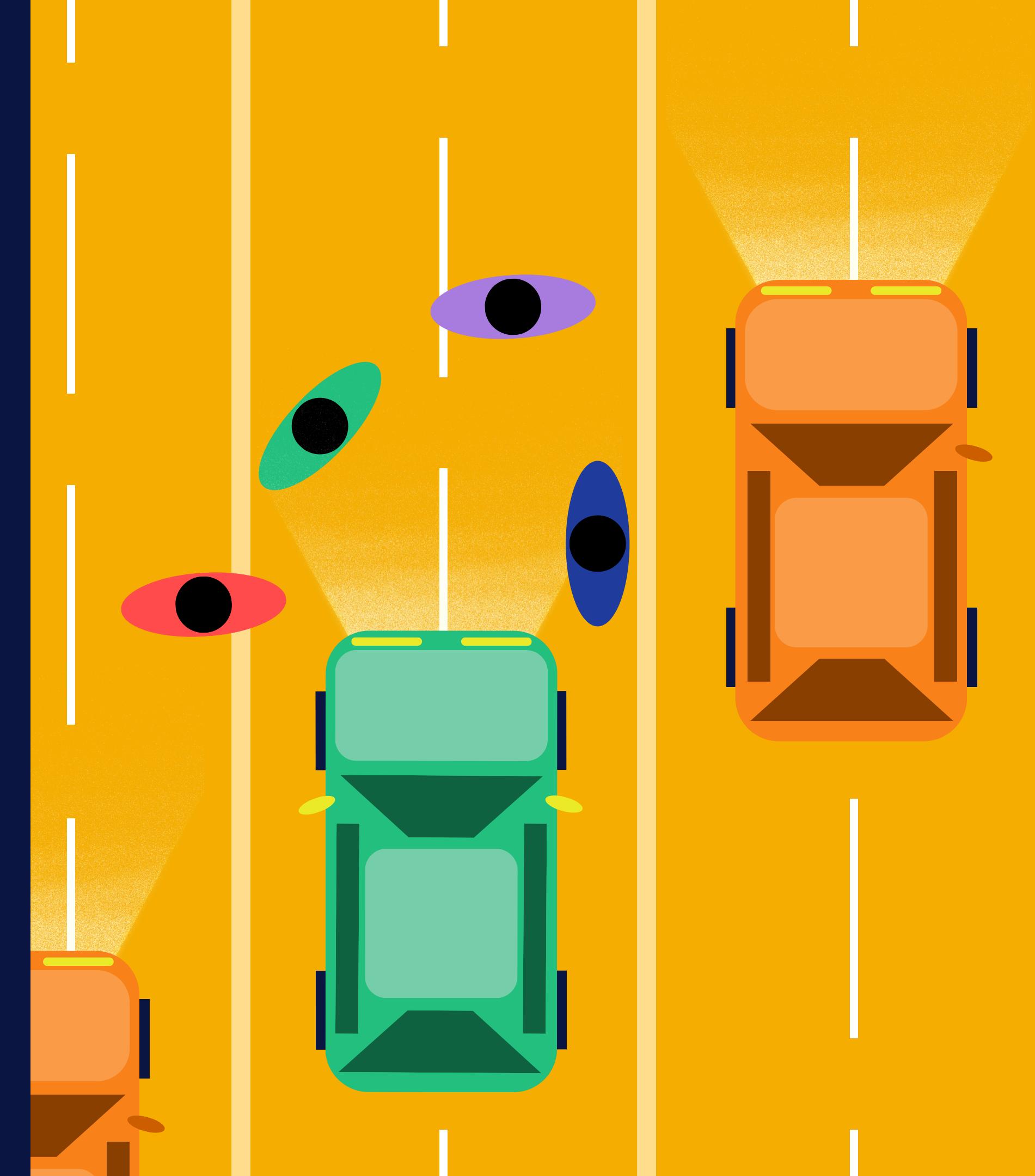
Platform online seperti Sarah.com menjadi semakin penting dalam konteks ini, menyediakan platform bagi penjual dan pembeli mobil bekas. Seperti yang dilaporkan oleh [Saudi Arabia Used Car Market Size & Share Analysis - Growth Trends & Forecasts \(2024 - 2029\)](#), pasar mobil bekas di Arab Saudi mencapai nilai USD 4,91 miliar pada tahun 2021 dan diproyeksikan akan meningkat menjadi USD 8,69 miliar pada tahun 2027, dengan tingkat pertumbuhan tahunan gabungan (CAGR) sebesar 7,36%.

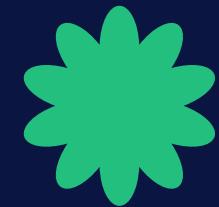




PROBLEM STATEMENT

Tantangan utama adalah menentukan harga mobil bekas dengan akurat. Harga yang terlalu tinggi dapat menghambat penjualan, sementara harga yang terlalu rendah dapat mengurangi keuntungan. Faktor-faktor seperti merk, model, tahun, jarak tempuh, ukuran mesin, dan lain-lain, mempengaruhi harga mobil bekas, namun tidak semua penjual memiliki pengetahuan yang cukup tentang fluktuasi harga ini.

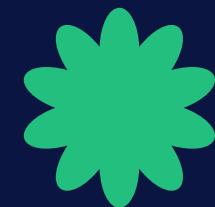




GOALS

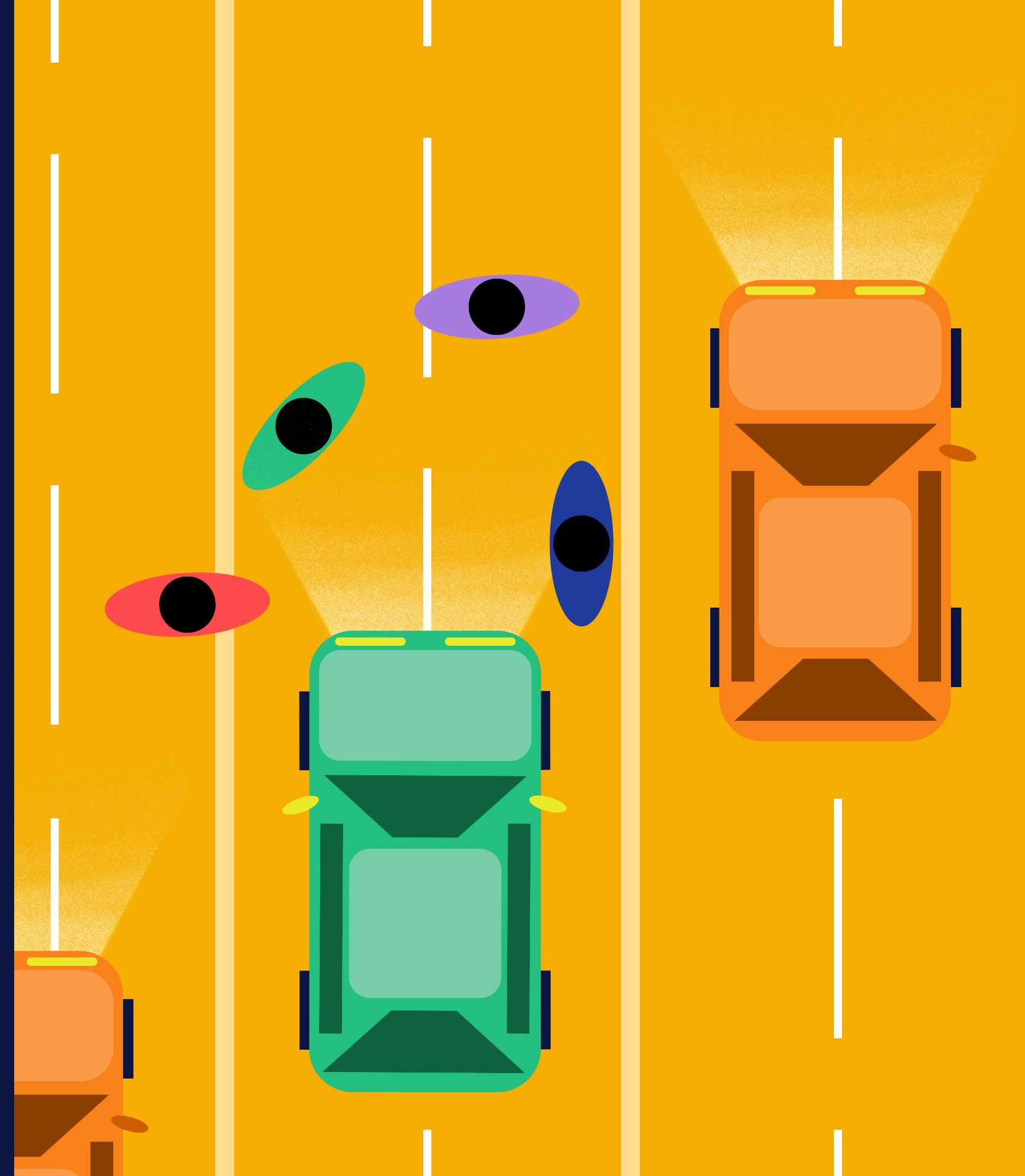
Tujuan utama adalah mengembangkan model yang dapat memprediksi harga mobil bekas berdasarkan spesifikasi mobil. Model ini akan menjadi referensi bagi penjual dan pembeli mobil bekas, membantu mereka membuat keputusan transaksi dengan lebih cepat. Model ini juga bermanfaat bagi platform online seperti Syarah.com untuk meningkatkan jumlah transaksi dan pertumbuhan bisnis mobil bekas.





ANALYTIC APPROACH

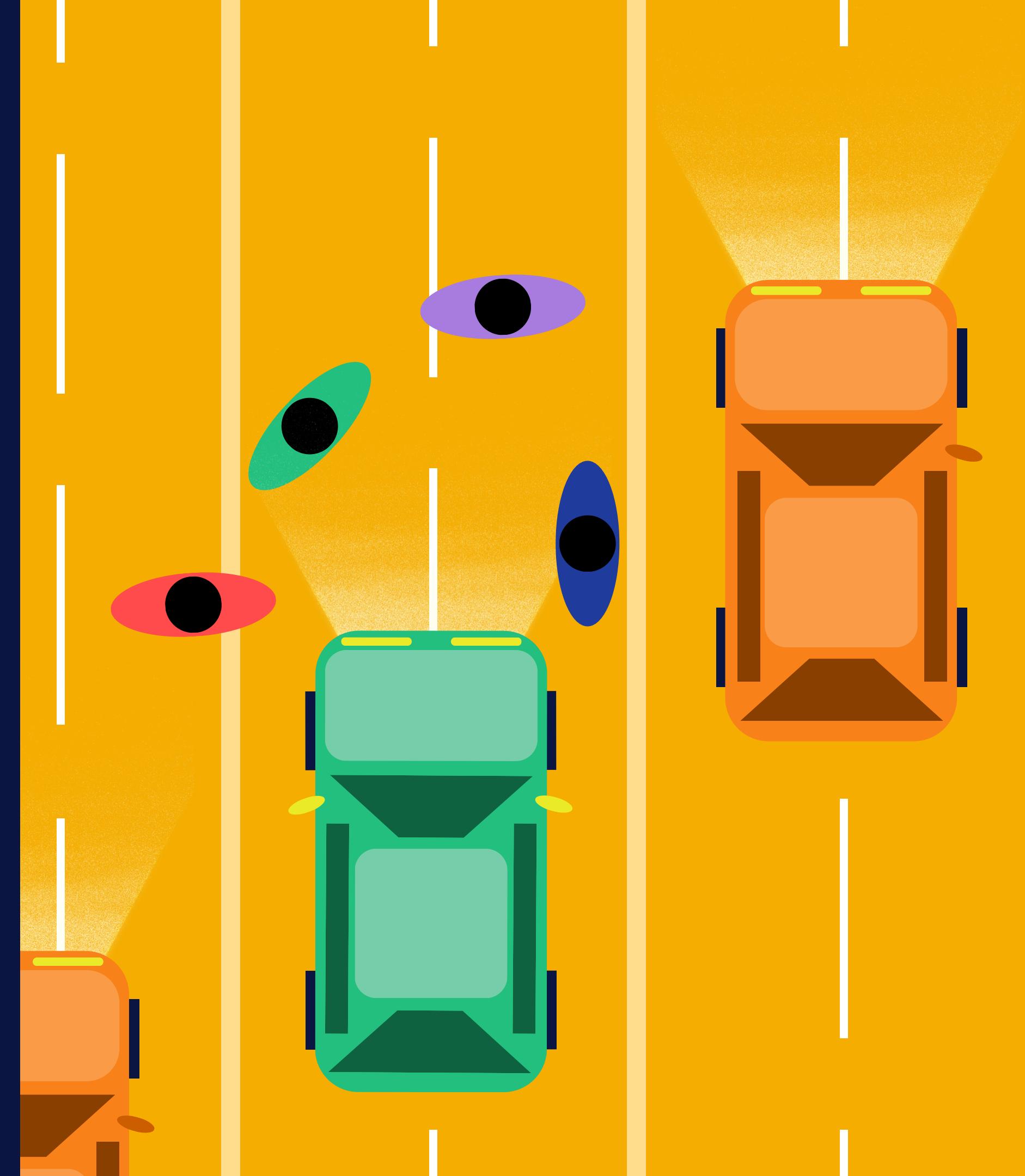
Membangun model regresi yang menganalisis spesifikasi mobil yang mempengaruhi harga. Model ini akan dibangun berdasarkan data historis dan fitur yang relevan.





METRIC EVALUATION

- Root Mean Squared Error (RMSE): Mengukur rata-rata kesalahan kuadrat. R-Squared: Menunjukkan seberapa baik data cocok dengan model regresi.
- Mean Absolute Error (MAE): Mengukur rata-rata kesalahan absolut.
- Mean Absolute Percentage Error (MAPE): Mengukur rata-rata persentase kesalahan.
- Mean Squared Log Error (MSLE): Fokus pada perbedaan logaritmik antara nilai prediksi dan aktual.
- Root Mean Squared Logarithmic Error (RMSLE): Cocok untuk data dengan outlier.
- R-squared (R²): Menunjukkan seberapa besar variasi nilai yang dapat dijelaskan oleh model.



02 - DATA UNDERSTANDING





Attributes	Data Type	Description
Type	Object	Variabel Fitur ini merepresentasikan model mobil
Region	Object	Variabel Fitur ini menunjukkan wilayah di mana mobil bekas tersebut ditawarkan untuk dijual
Make	Object	Variabel Fitur ini merepresentasikan merek mobil
Gear_Type	Object	Variabel Fitur ini menentukan jenis transmisi mobil, yaitu otomatis atau manual
Origin	Object	Variabel Fitur ini mengindikasikan negara asal impor mobil (Gulf / Saudi / Lainnya)
Options	Object	Variabel Fitur ini menunjukkan tingkat kelengkapan mobil, seperti Full Options, Semi-Full, atau Standar
Year	Int	Variabel Fitur ini mencatat tahun pembuatan mobil, seperti 2018, 2015, 2020
Engine_Size	Float	Variabel Fitur ini menggambarkan ukuran mesin mobil, biasanya dalam satuan liter (contoh: 2.5, 3.0, 1.8)
Mileage	Int	Variabel Fitur ini menunjukkan jarak tempuh rata-rata yang telah ditempuh oleh mobil, biasanya dalam kilometer (contoh: 50000km, 75000km, 30000km)
Negotiable	Bool	Variabel Fitur ini menunjukkan apakah harga mobil dapat dinegosiasikan. Jika bernilai 'True', maka harga mobil tersebut belum ditetapkan dan masih dapat dinegosiasikan
Price	Int	Variabel Target ini mencatat harga mobil bekas dalam satuan Riyal Arab Saudi (SAR), seperti 60000, 85000, 45000

03 - DATA PREPROCESSING



DATA PREPROCESSING

Pengecekan dan
Penaganan Missing
Values dan Duplikat
Data



Cek Distribusi pada
Setiap Atribut/Fitur

Menangani perubahan
sample data/values pada
kolom



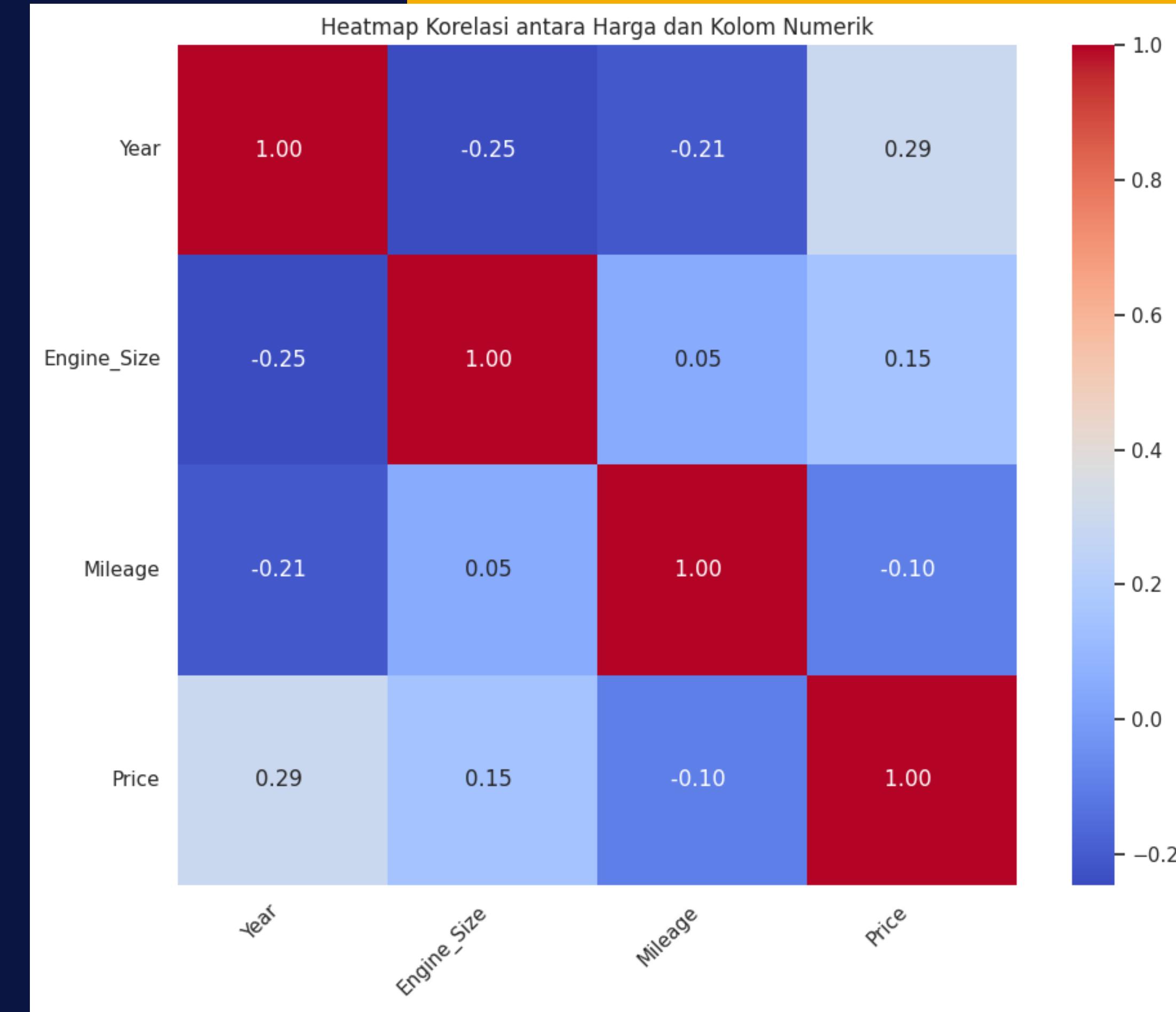
Identifikasi dan
Penaganan Outliers



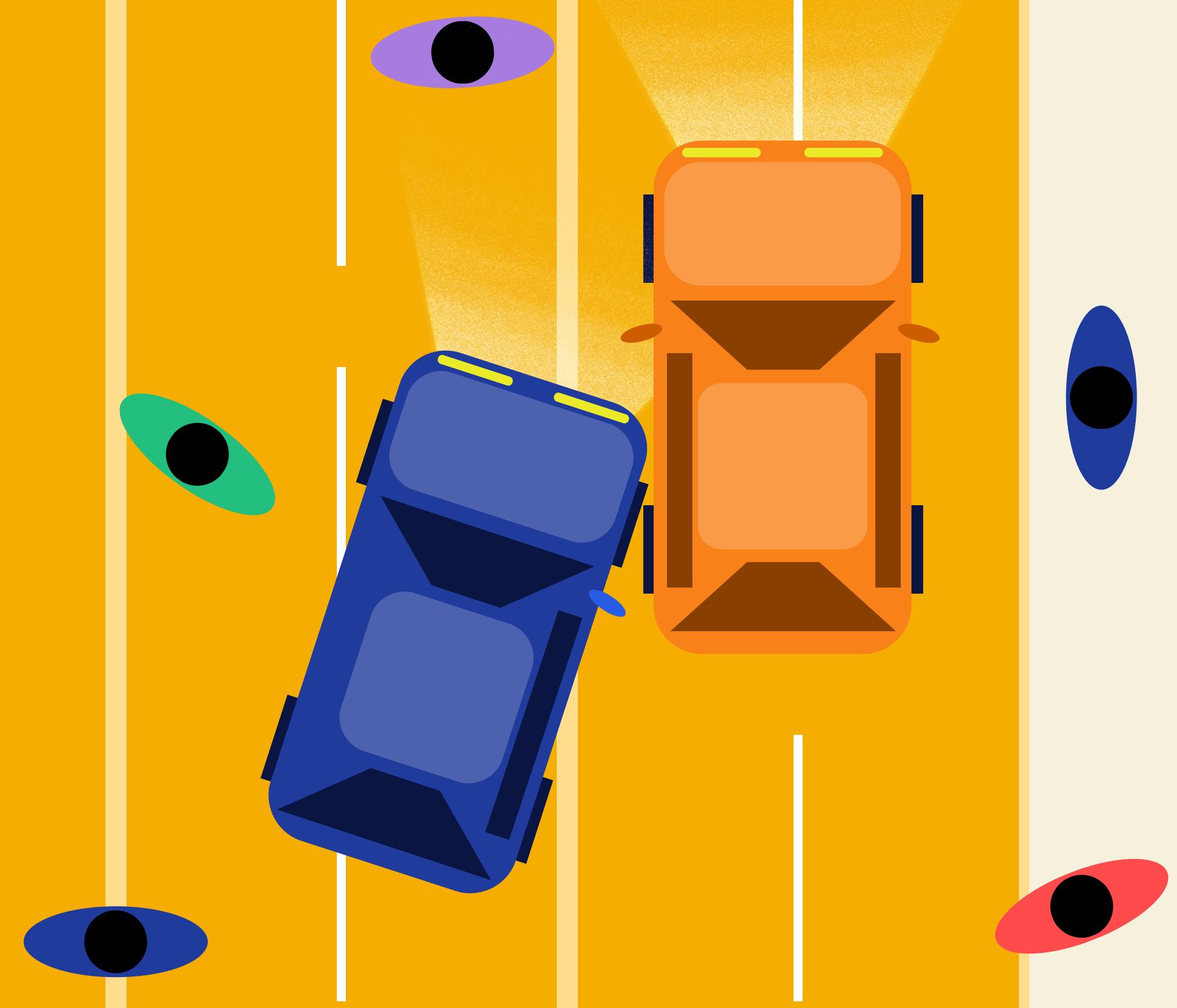
Mengidentifikasi kolom
yang tidak relevan
(feature selection)

Cek Korelasi Setiap
Atribut/Fitur

DATA CORRELATION



04 - MODELING



MODELING



One-Hot Encoding:

- Gear_Type
- Origin
- Options

Binary Encoding:

- Type
- Region
- Make

Robust Scaler:

- Year
- Engine_Size

• Mileage

Splitting data into training and test with proportion 70:30

Base Model:

- Linear Regression
- KNN Regressor
- Decision Tree Regression
- Lasso Regression
- Elastic Net

Ensemble Model:

- Random Forest Regressor
- Ada Boost Regressor
- Gradient Boosting Regressor
- Xtreme Gradient Boosting Regressor

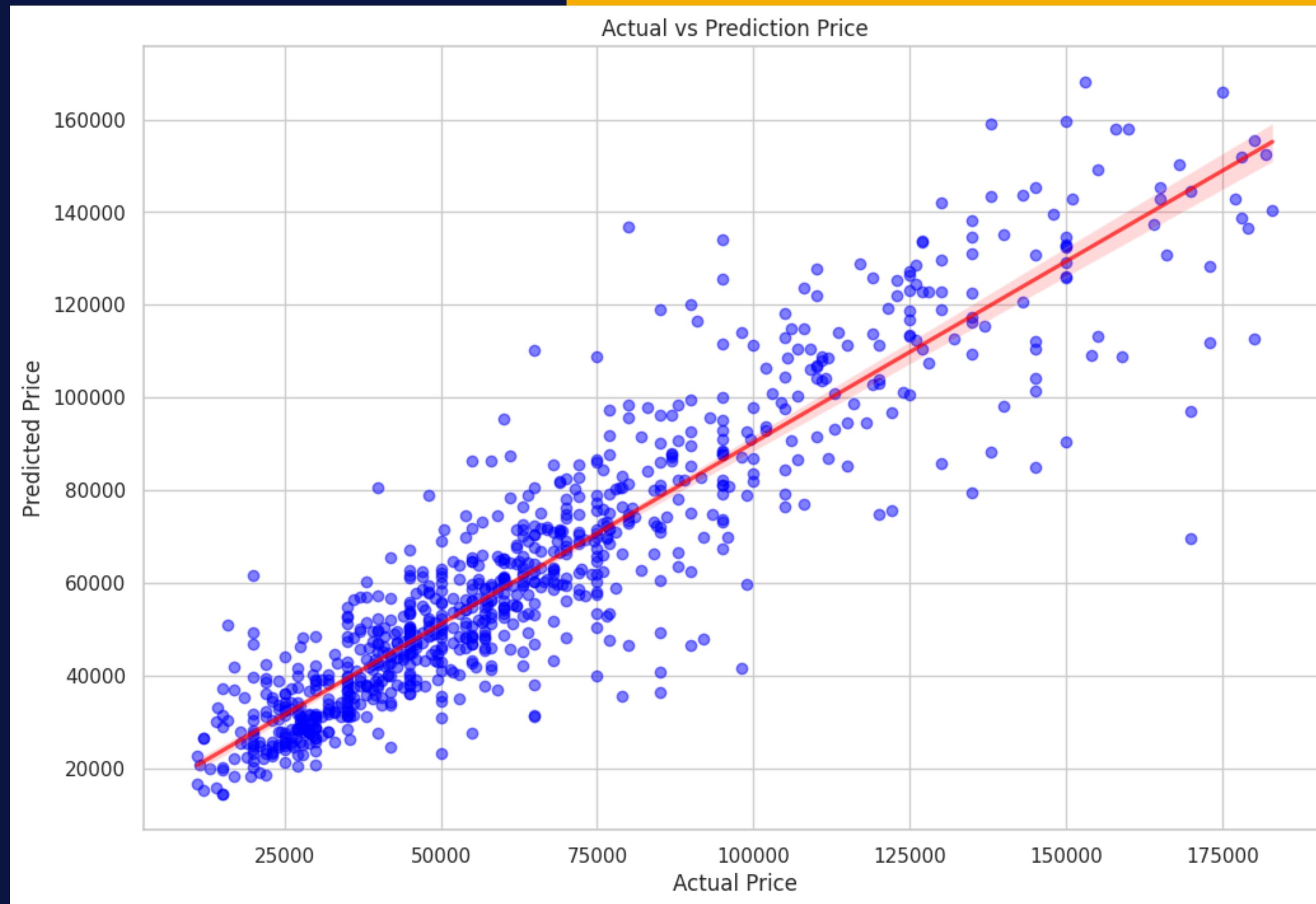
TEST RESULT (SCORE METRIC EVALUATION)

#	Model	Score RMSE	Score MAE	Score MAPE	Score MSLE	Score RMSLE	Score R2
9	Xtreme Gradient Boost Regression	-18084.471550	-11979.854270	-0.204410	-0.077262	0.277960	0.779621
3	Ridge Regression	-18891.857912	-11661.957308	-0.186192	-0.074408	0.272775	0.758264
6	Random Forest Regression	-19126.385326	-12412.585424	-0.209420	-0.082118	0.286558	0.752613
8	Gradient Boost Regression	-19827.270984	-13211.711417	-0.213730	-0.079384	0.281752	0.734282
1	KNN Regression	-24420.545499	-16763.768651	-0.319934	-0.147903	0.384581	0.599125
4	Lasso Regression	-24808.555241	-16920.206801	-0.278956	-0.119473	0.345649	0.583281
5	Elastic Net	-24816.701728	-16924.753214	-0.278958	-0.119480	0.345659	0.583005
0	Linear Regression	-24829.843138	-16930.394313	-0.278935	-0.119491	0.345674	0.582551
7	AdaBoost Regressor	-25578.896928	-18412.758596	-0.303783	-0.125747	0.354608	0.559819
2	Decision Tree Regression	-28559.100106	-16995.436131	-0.300441	-0.156752	0.385919	0.524130

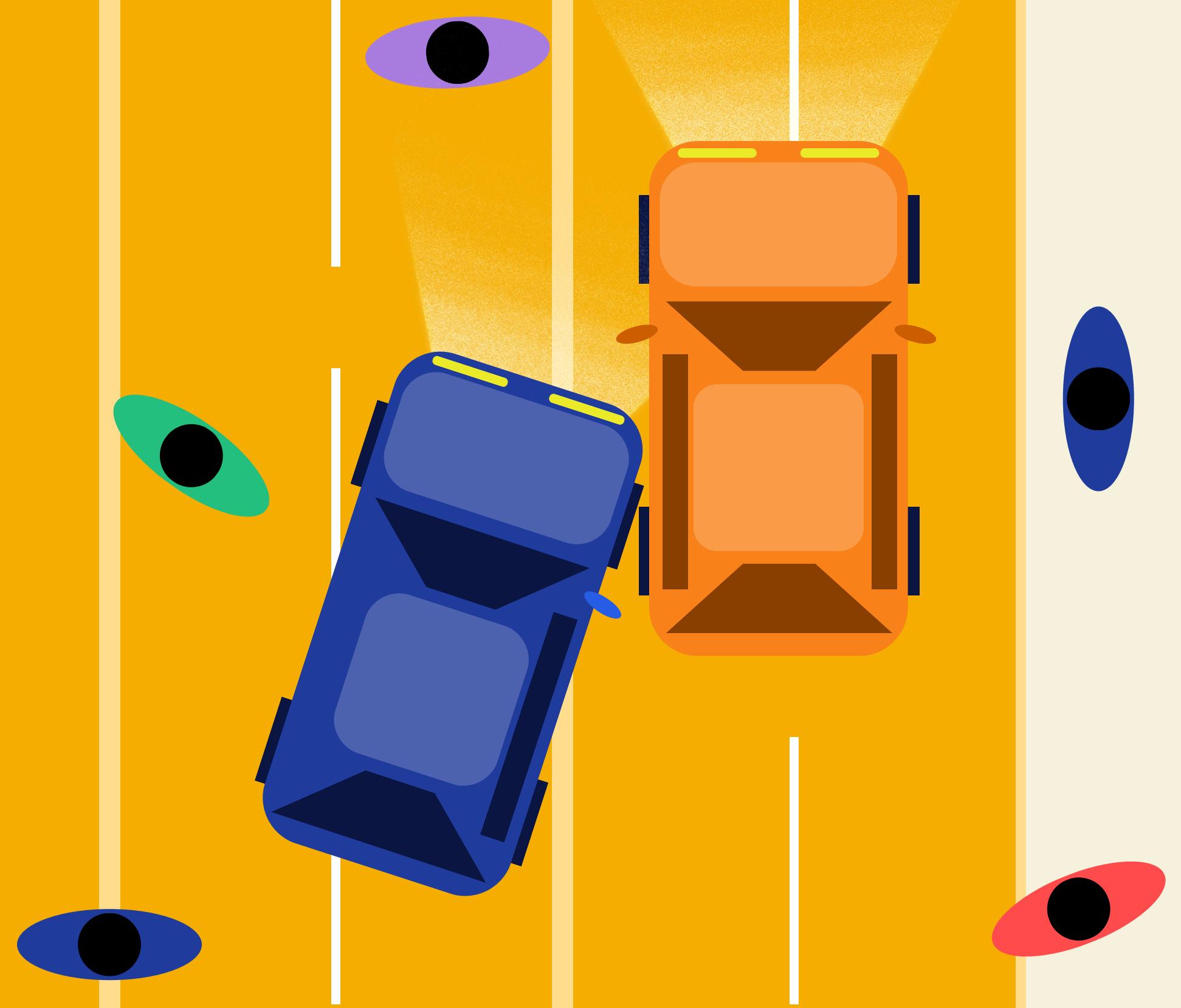
HYPERPARAMETER TUNING

	Score RMSE	Score MAE	Score MAPE	Score MSLE	Score RMSLE	Score R2
Before Hyperparameter Tuning	18802.649703	11019.873302	0.201203	0.073927	0.271895	0.789916
After Hyperparameter Tuning	14947.783852	9977.154877	0.179209	0.058781	0.242448	0.833738

ANALYSIS OF MODEL PREDICTION RESULT



05 - CONCLUSION





CONCLUSION

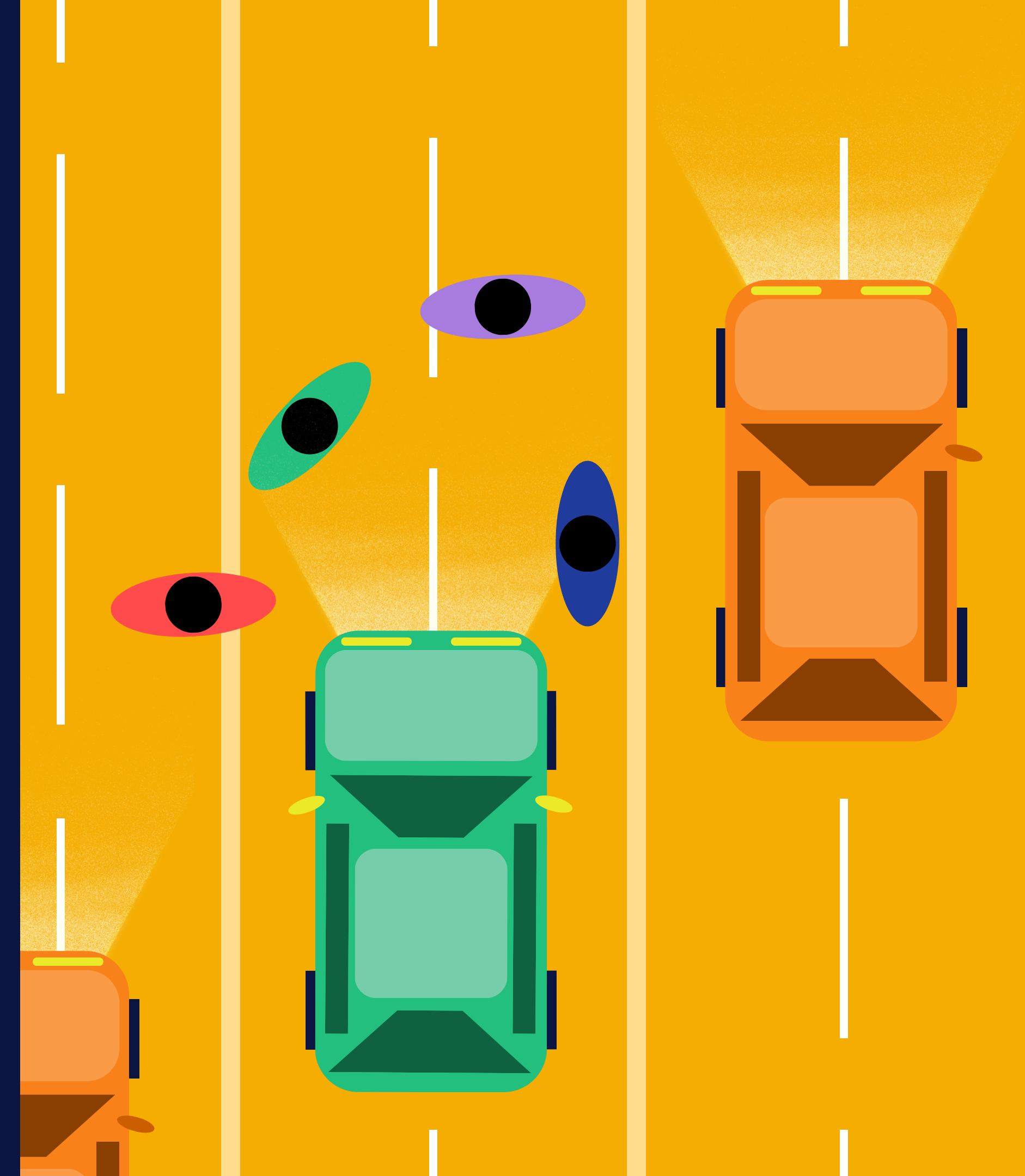
1. Evaluasi Model Sebelum dan Sesudah Hyperparameter Tuning:

- Sebelum penyetelan hyperparameter, model memiliki nilai RMSE sebesar 16,802.65, MAE sebesar 11,019.87, dan MAPE sebesar 20.12%. Setelah penyetelan hyperparameter, terjadi peningkatan kinerja dengan nilai RMSE sebesar 14,947.78, MAE sebesar 9,977.15, dan MAPE sebesar 17.92%. Model yang telah disesuaikan menghasilkan performa yang lebih baik.

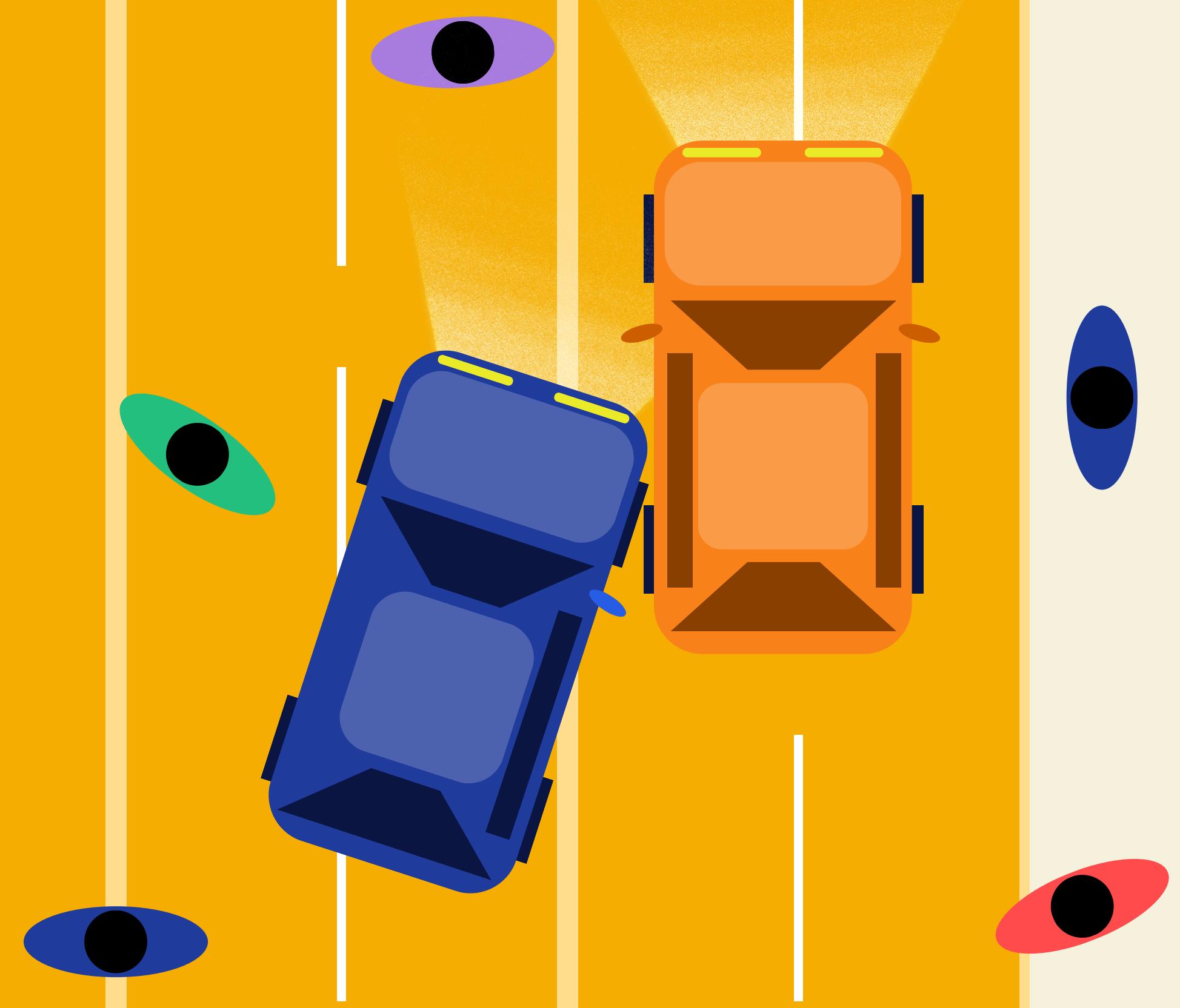
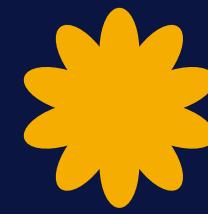
2. Pentingnya Fitur dalam Model:

- Berdasarkan analisis fitur, fitur-fitur yang paling berpengaruh dalam memprediksi harga mobil bekas antara lain:

- 'Make_1' (12.59%)
- 'Engine_Size' (12.56%)
- 'Options_Standard' (9.82%)
- 'Year' (7.44%)
- 'Mileage' (5.82%)



06 - RECOMMENDATION





RECOMMENDATION

- Karena fitur-fitur seperti 'Make_1', 'Engine Size', dan 'Year' memiliki dampak signifikan, perusahaan atau individu dapat lebih memfokuskan perhatian pada informasi dan karakteristik kendaraan terkait. Mungkin perlu dilakukan penelitian lebih lanjut untuk memahami preferensi pasar terkait merek, ukuran mesin, dan tahun pembuatan.
- Fitur 'Options_Standard' memiliki pengaruh signifikan terhadap harga mobil bekas. Oleh karena itu, perusahaan atau individu yang terlibat dalam transaksi jual-beli mobil bekas perlu memperhatikan apakah mobil tersebut memiliki opsi standar atau tidak.
- Meskipun model telah mengalami peningkatan performa setelah penyetelan hyperparameter, perlu dilakukan pemantauan terus-menerus terhadap perilaku model dengan data baru. Dapat dilakukan pembaruan model secara berkala agar tetap akurat dan relevan dengan perubahan tren pasar.

