



TEAM FIVE DATA SCIENTIST TEAM





TEAM FIVE

Team Five merupakan **konsultan data scientist** yang menawarkan jasa kepada perusahaan berupa **penyelesaian business problem** dengan **data dan machine learning**.

Anggota :



TRIE SONY K



AMRU HANIF M



M. REZA SYAHZIAR



M. SHOHIH ALWI



GILANG FITRAR

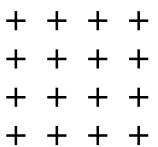
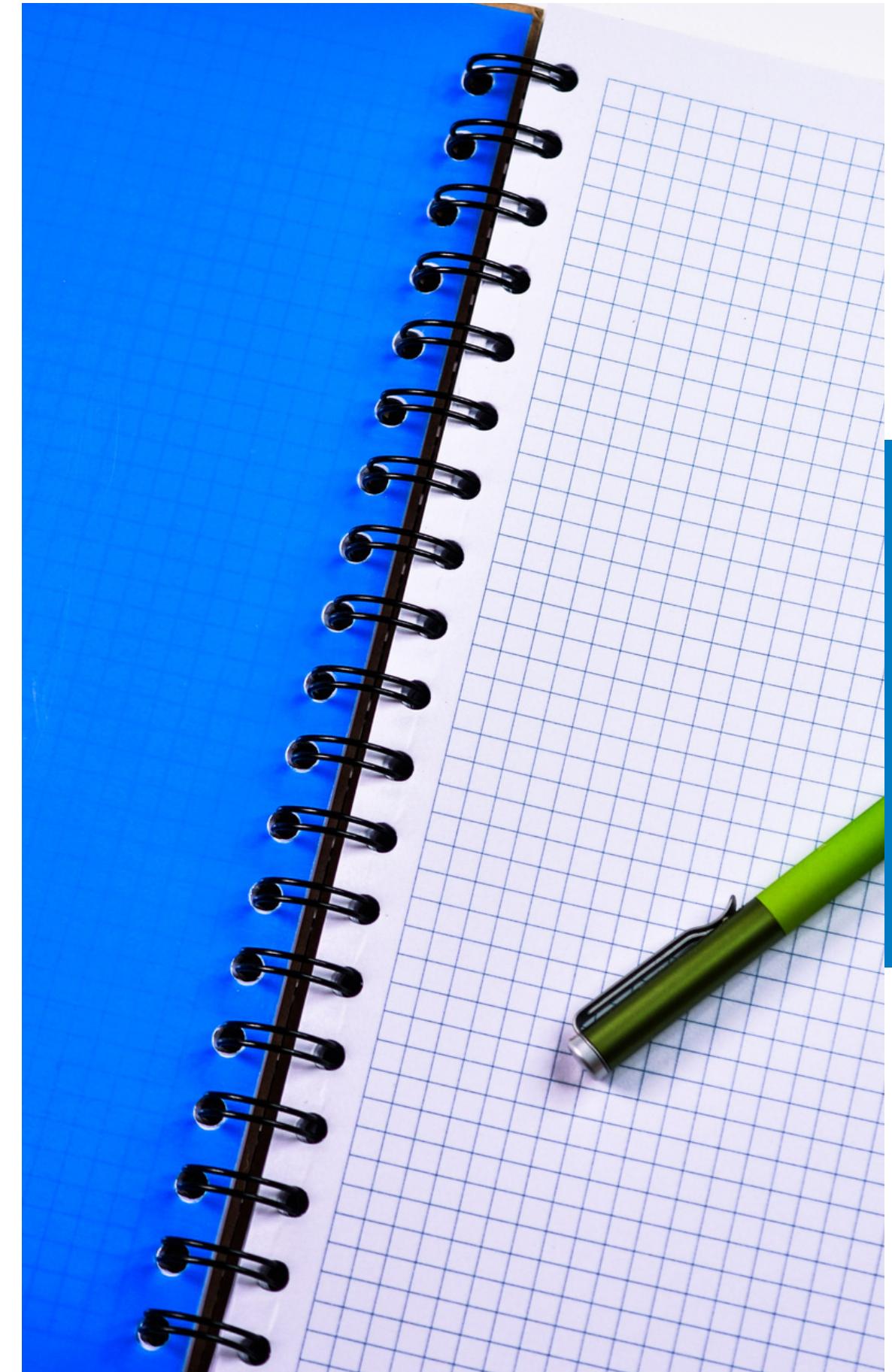


TABLE OF CONTENTS

- Background
- EDA
- Business Insight
- Data Preprocessing
- Modeling and Evaluation
- Business Recomendation

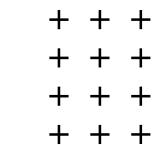




BACKGROUND STORY



BACKGROUND STORY



PT Otokar merupakan sebuah perusahaan yang bergerak di **bidang penjualan mobil bekas**. Perusahaan ini **kesulitan menentukan harga** untuk mobil bekas yang akan mereka jual.

Perusahaan ingin menggunakan **MMR** untuk menentukan harga jual kendaraan. MMR adalah perkiraan harga mobil bekas yang disediakan oleh perusahaan lain. Namun, setelah dilakukan analisis ternyata **prediksi MMR** sering memprediksi harga jual jauh di bawah harga jual sebenarnya, yang artinya **keuntungan yang diterima perusahaan tidak maksimal**.

Analisis lanjut dari tim kami menunjukkan bahwa rata - rata selisih prediksi MMR dan harga jual kendaraan sebenarnya mencapai **-156\$/mobil atau sekitar -2,25 juta rupiah/mobil**. Ini merupakan keuntungan yang tidak dimaksimalkan atau bisa disebut **opportunity cost**.

Mobil	Prediksi MMR (\$)	Harga Jual (\$)
Toyota Sequoia	27,400	28,500
Cadillac CTS	2,300	4,100
Hyundai Sonata	7,875	9,200
Ford F-150	23,800	24,900

BACKGROUND STORY

Selain masalah penentuan harga mobil bekas, PT Otokar juga memiliki masalah terkait kapasitas penyimpanan atau **showroom mobil yang terbatas dengan kapasitas 100 mobil**. Sedangkan banyak seller yang ingin menjual kendaraannya di Otokar Showroom. **Kendaraan seperti apa** yang harus mereka **prioritaskan** masuk ke dalam Otokar Showroom untuk **meningkatkan revenue perusahaan?**



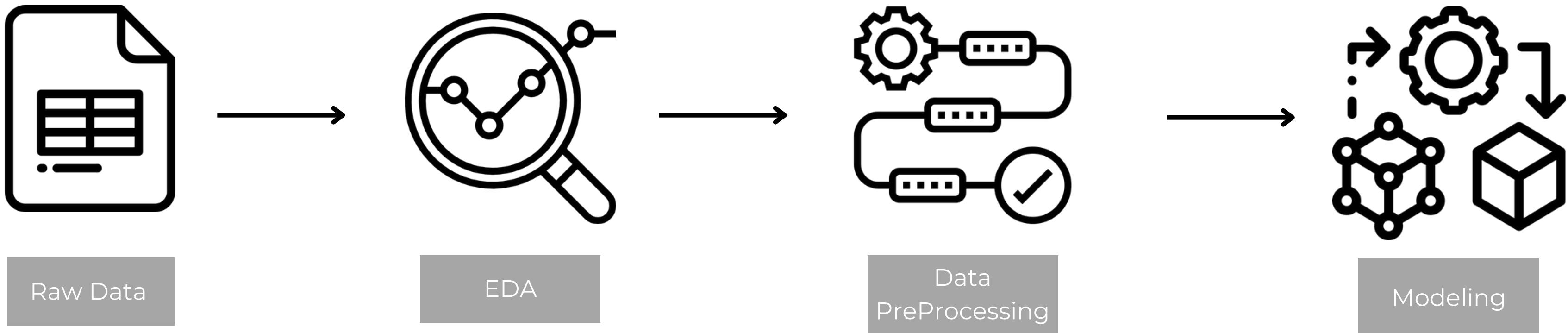
GOALS, OBJECTIVE, BUSINESS METRICS

Goals	Objective	Business Metrcis
Menurunkan nilai opportunity cost per kendaraan dengan machine learning.	Membuat model Machine Learning untuk memprediksi harga mobil bekas berdasarkan spesifikasi dan kondisi kendaraan.	Opportunity cost per mobil.
Meningkatkan revenue dengan rekomendasi kriteria kendaraan		Total Revenue.



FLOW MACHINE LEARNING

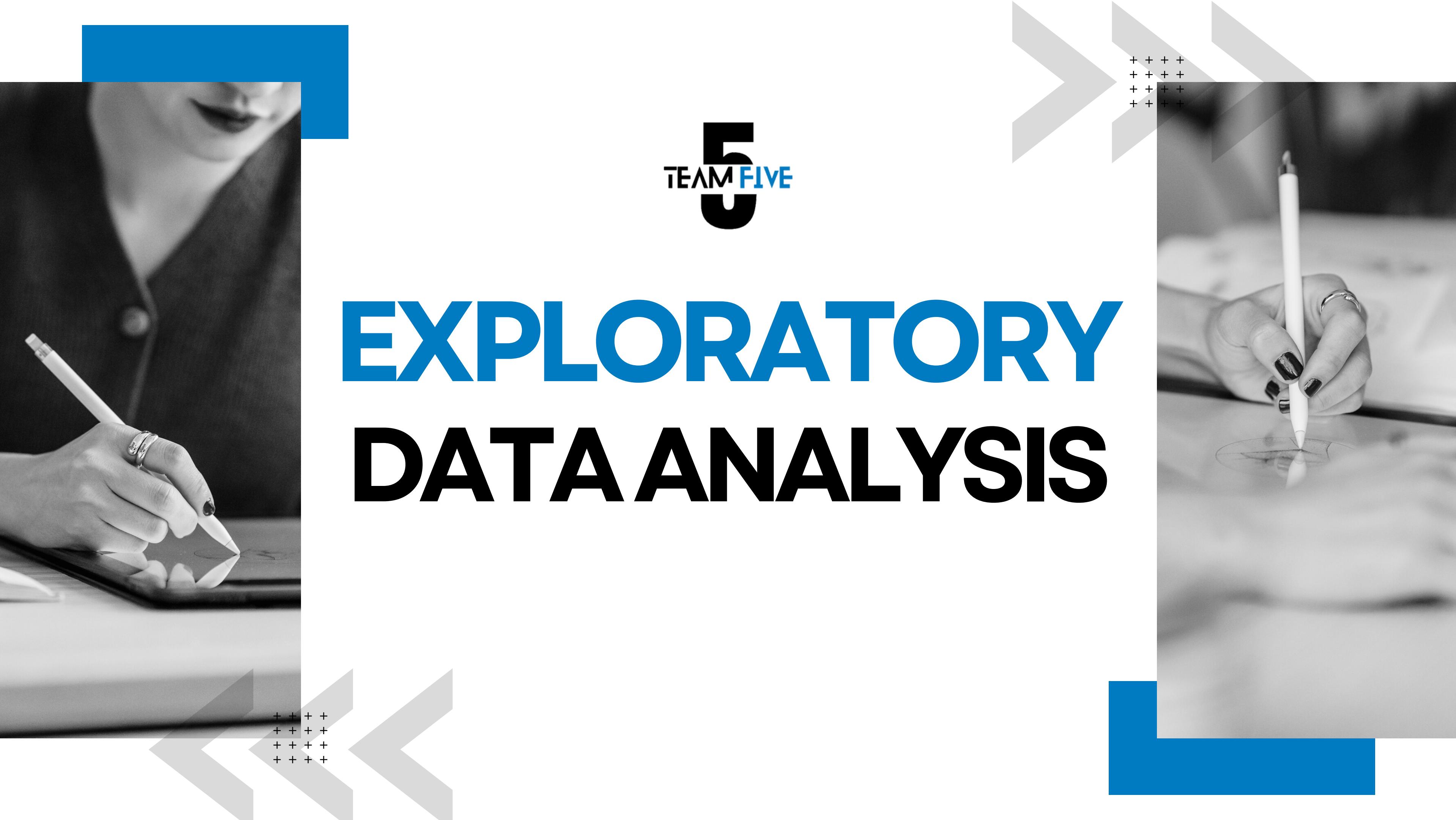
+++
+++
+++
+++



- Data Car_prices diunggah dari Keggle
- Menentukan Goal, Objective, dan Bisnis metric dari data yang dimiliki
- Analisis data apakah terdapat masalah pada data yang dimiliki
- Menganalisis perkolom dan juga hubungan antar kolom
- Melakukan data cleansing terhadap data bermasalah
- Melakukan Feature Encoding untuk memaksimalkan feature yang dipunya
- Menentukan algoritma yang paling tepat pada modeling dengan metric evaluation
- Evaluasi bagaimana hasil model yang didapat bisa mendukung bisnis yang ditargetkan



EXPLORATORY DATA ANALYSIS



EXPLORATORY DATA ANALYSIS

COLUMN	DTYPE
year	Numeric
make	Categoric
model	Categoric
trim	Categoric
body	Categoric
transmission	Categoric
vin	Categoric
state	Categoric
condition	Numeric
odometer	Numeric
color	Categoric
interior	Categoric
seller	Categoric
mmr	Numeric
sellingprice	Numeric (y)
saledate	Categoric

car_prices.csv

Data info

- Terdapat **558.811 baris** dan **16 kolom** pada dataset
- Semua kolom **tipenya sudah sesuai**
- Terdapat **5 kolom Numeric** dan **11 kolom Categoric**
- **sellingprice** merupakan **target(y)** dengan value harga terjualnya mobil
- Terdapat **missing value** pada **9 kolom** yaitu : make, model, trim, body, transmission, condition, odometer, color, and interior
- **Tidak terdapat data duplikat**

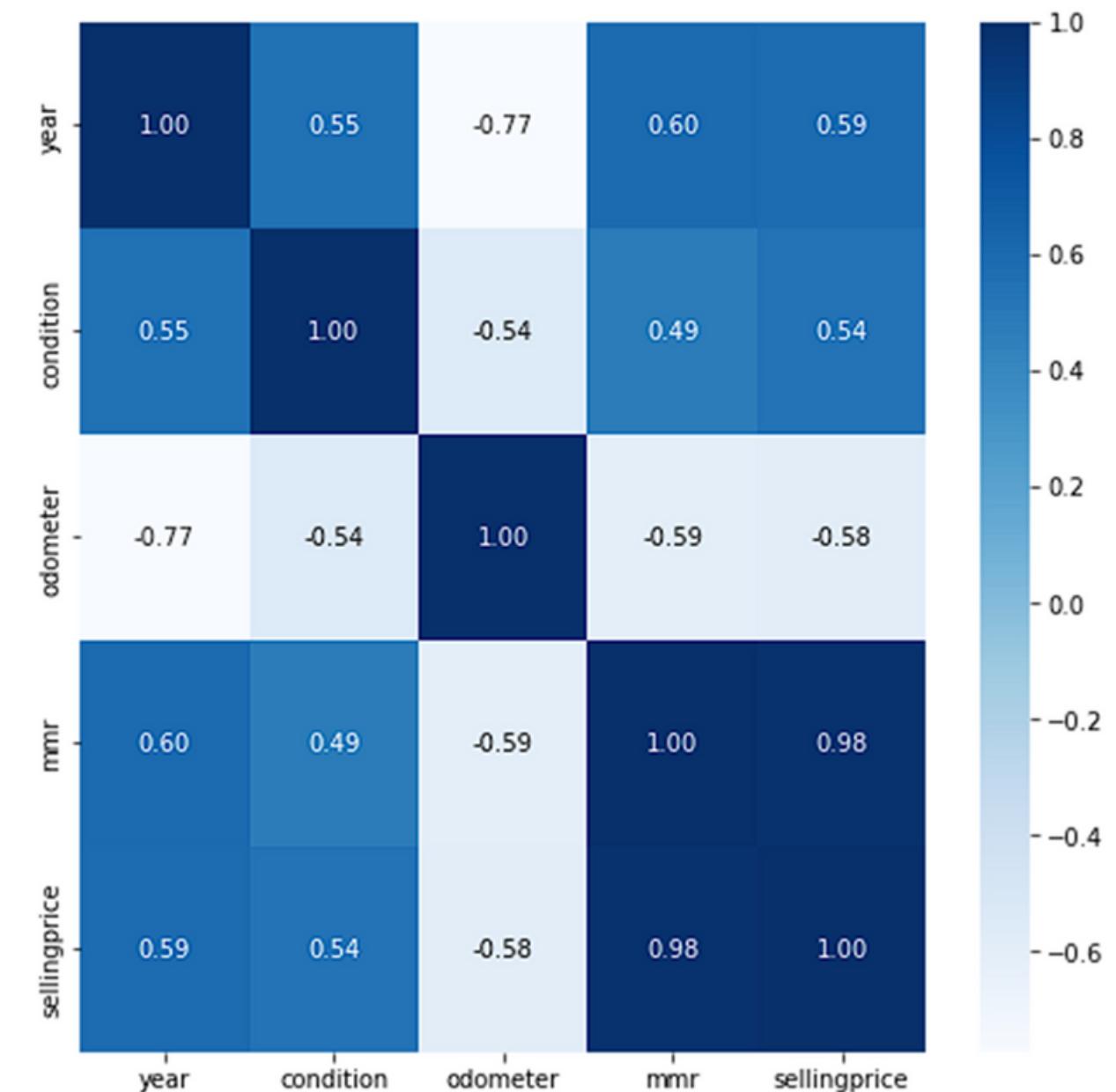
EXPLORATORY DATA ANALYSIS

UNIVARIATE ANALYSIS

- **odometer** memiliki distribusi yang **positively skewed**, datanya banyak di nilai kecil kebesar
- **year** memiliki distribusi yang **negatively skewed**, datanya banyak di nilai besar kekecil
- **Condition** merupakan **distribusi multinomial**, datanya normal ada yang nilainya kecil dan besar
- Terdapat beberapa **kolom** dengan **kardinalitas yang tinggi** (nilai unique yang tinggi), seperti **body, state, make, dan model**. yang akan di proses pada Feature encoding agar dapat dilanjutkan pada proses modeling

MULTIVARIATE ANALYSIS

- Table **Sellingprice** dengan **MMR** mempunyai **korelasi** yang **kuat (0.98)** karena MMR merupakan model yang sudah ada pada data yang nanti akan didrop
- Feature **Condition** yang paling **lemah (0.54)** yang berarti merupakan feature yang paling sedikit pengaruhnya terhadap target
- **Odometer** dan **Year** cenderung **redundan**





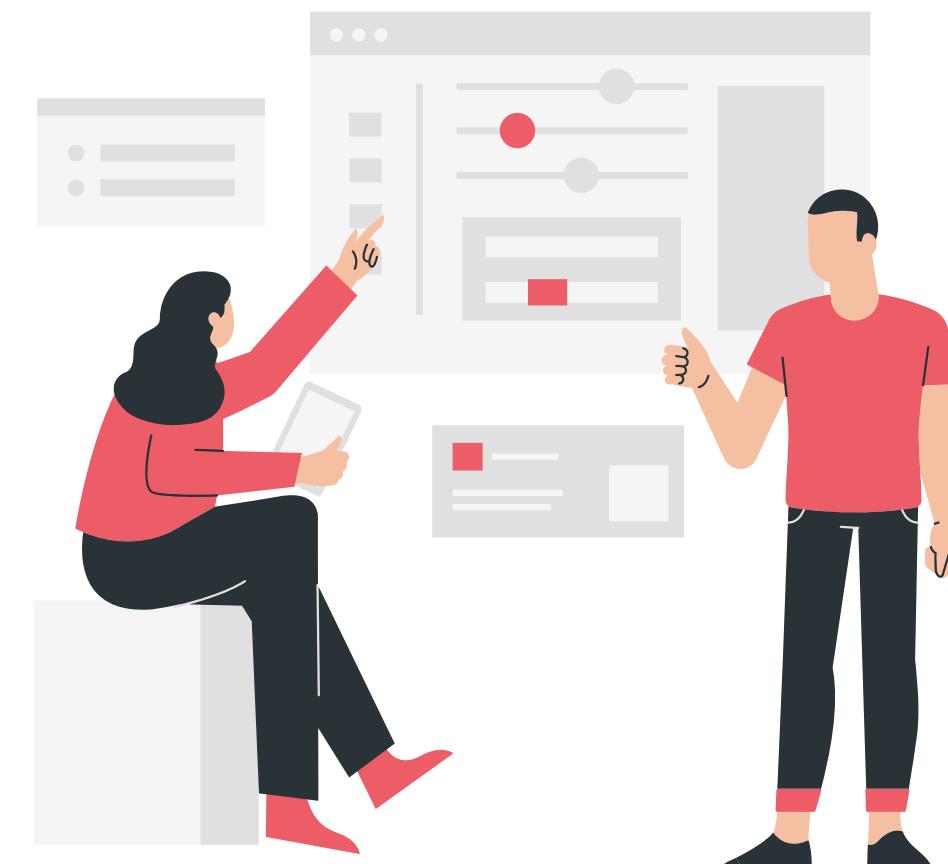
DATA PRE-PROCESSING



DATA CLEANING

+++
+++
+++
+++

Drop Columns	Missing Data	Handling Outliers
Drop kolom trim, vin, seller	Table make, model, odometer, condition, body, transmission, color, interior	Handling outliers menggunakan Z-Score



+++
+++
+++
+++

Feature	Encode	Result
Body	Dikelompokkan berdasarkan <i>body</i> mobil sejenis	10 feature yaitu convertible, cab, van, wagon, coupe, hatchback, minivan, sedan, supercrew dan suv
State	Dikelompokkan berdasarkan negara bagian	5 feature yaitu west, south, northeast, midwest, outside us
Make	Perbaikan penulisan, ex: Bmw -> bmw	47 feature
Model	Digabungkan dengan harga penjualan untuk mengurutkan model terbaik berdasarkan harganya	Rank 1 sampai 10

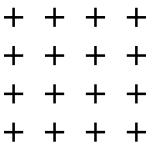
FEATURE ENCODING

+++
+++
+++
+++



MODELLING





MODEL LING



Jenis Model : Model regresi yang akan digunakan untuk memprediksi harga mobil bekas

Algoritma :

- Linear Regression (Ridge, Lasso dan ElasticNet)
- Decision Tree
- XGB Regressor

Metrics Evaluation :

- R²
- RMSE dan MAE

MODEL COMPARISON

Underfitting

Best Fit

Overfitting

Model	R2 (test)	R2 (train)	RMSE (\$)	MAE (\$)
Linear Regression	78.81%	78.67%	3775.33	2737.21
Ridge	78.81%	78.67%	3775.50	2737.30
Lasso	78.75%	78.63%	3780.73	2738.93
ElasticNet	60.03%	60.08%	5185.75	3879.32
XGB Regressor	90.24%	90.65%	2562.58	1733.96
Decision Tree Regressor	83.31%	100%	3350.93	2138.81

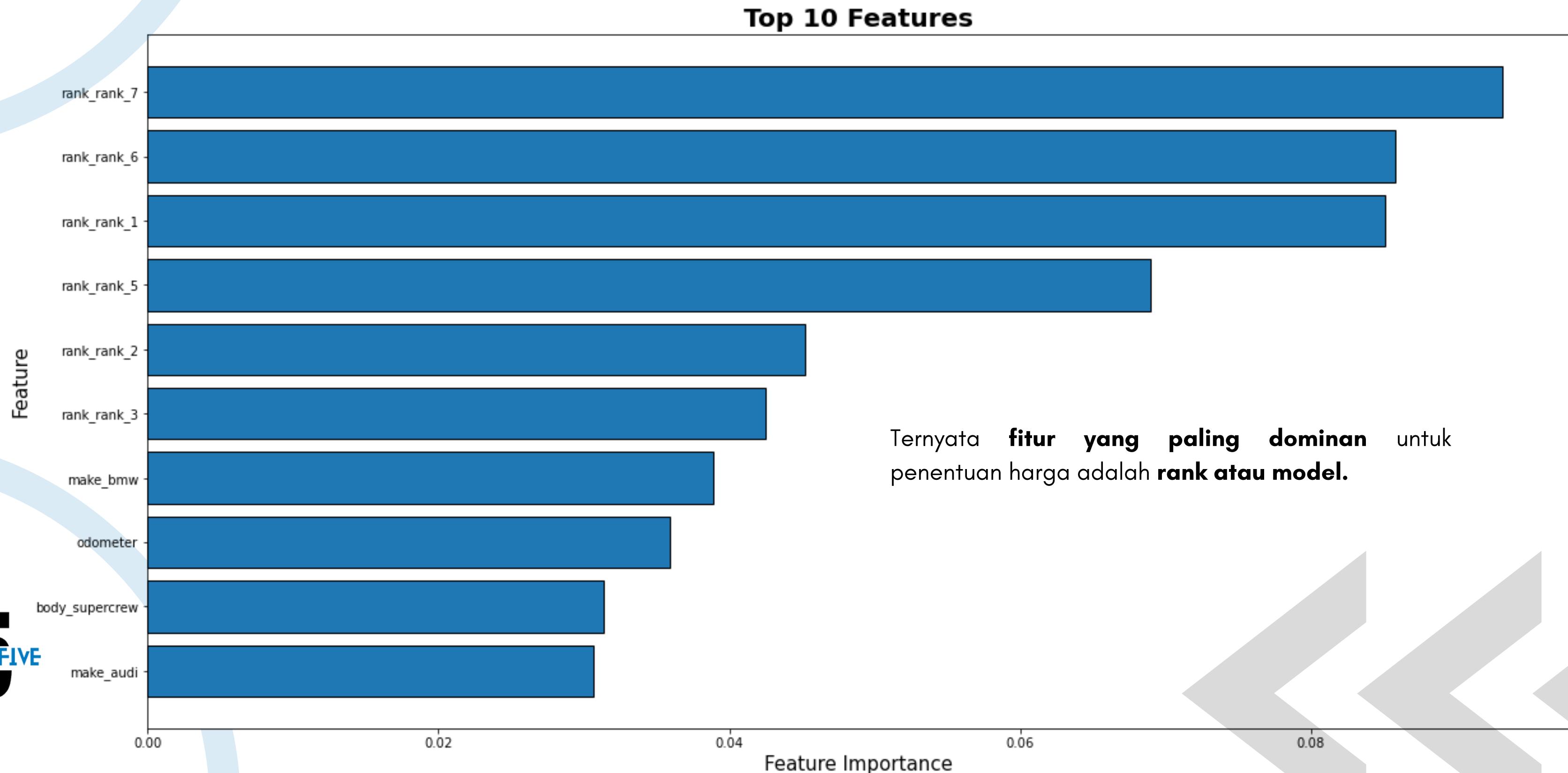


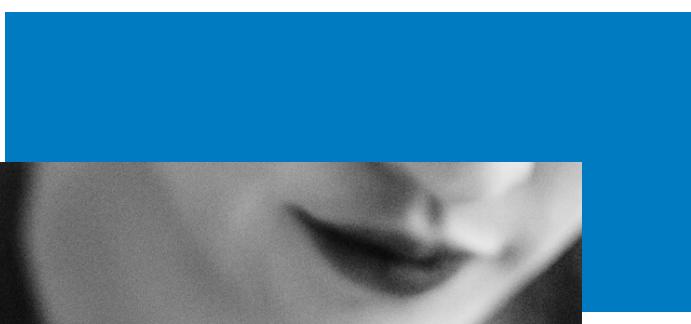
HYPERPARAMETER TUNING

- Metode Tuning : RandomizedSearchCV
- Cross Validation : 5
- Number of Iteration : 30

Model	R2 (Test)	R2 (Train)	RMSE	MAE
Before Tuning	90.24%	90.65%	2562.58	1733.96
After Tuning	91.15%	91.82%	2440.31	1610.64

FEATURE IMPORTANCE





BUSINESS RECOMMENDATION

REKOMENDASI

Gunakan model baru yang telah tim kami buat sebagai acuan **penentuan harga mobil bekas**

Menggunakan model yang telah kami buat **opportunity cost per kendaraan turun dari 156.3 USD menjadi 0.37 USD**

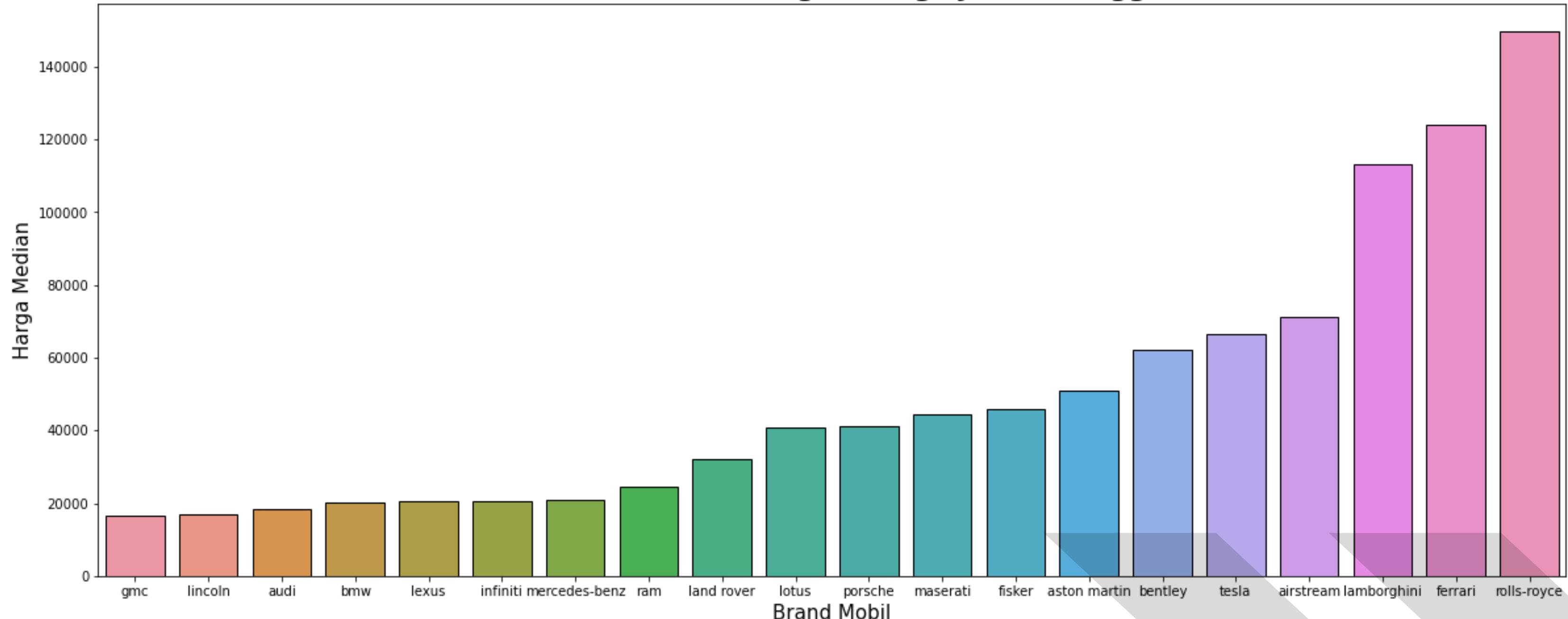
Car	MMR (\$)	New Model (\$)	Selling Price (\$)
Chrysler PT Cruiser	3,675	5,912	6,000
Chevrolet Impala	6,425	6,422	7,200
Ford Fusion	15,650	16,598	17,200
Nissan Altima	11,650	12,712	13,600
Ford Focus	10,450	10,279	11,400

Model	Opportunity Cost per Car (\$)
MMR	156.30 (2.26 juta rupiah)
New Model	0.37 (5,352 rupiah)

BUSINESS INSIGHT

TEAM **5**

20 Brand Mobil dengan Harga Jual Tertinggi



BUSINESS INSIGHT

20 Brand Mobil Paling Laku

Untuk **memaksimalkan revenue**, perusahaan perlu menjual kendaraan yang **harga jualnya tinggi**, tapi...

Yang termahal bukan berarti yang paling laku,
TOP 10 brand termahal memiliki penjualan **kurang dari 10,000 kendaraan**

Begitu juga sebaliknya, maka dari itu untuk **memaksimalkan revenue** dipilih 3 brand kendaraan yang masuk ke dalam kriteria **top 20 brand termahal dan terlaku** yaitu **BMW, Mercedes-Benz dan Lexus** untuk diberi **kuota khusus pada Otocar Showroom**



REKOMENDASI

Untuk **memaksimalkan revenue** dengan **kapasitas showroom yang terbatas**, kami merekomendasikan perusahaan untuk memberi **kuota khusus** sebesar **10%** untuk kendaraan dengan brand **BMW, Mercedes-Benz dan Lexus**

Kombinasi	Revenue (\$)					Average (100 sample)
	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	
100 Random Car	1,212,600	1,381,101	1,430,200	1,205,350	1,348,175	1,360,318
10 Rec. Brand + 90 Random Car	1,384,450	1,380,550	1,550,451	1,421,075	1,381,300	1,457,570
20 Rec. Brand + 80 Random Car	1,404,250	1,549,250	1,520,000	1,520,150	1,477,925	1,512,736



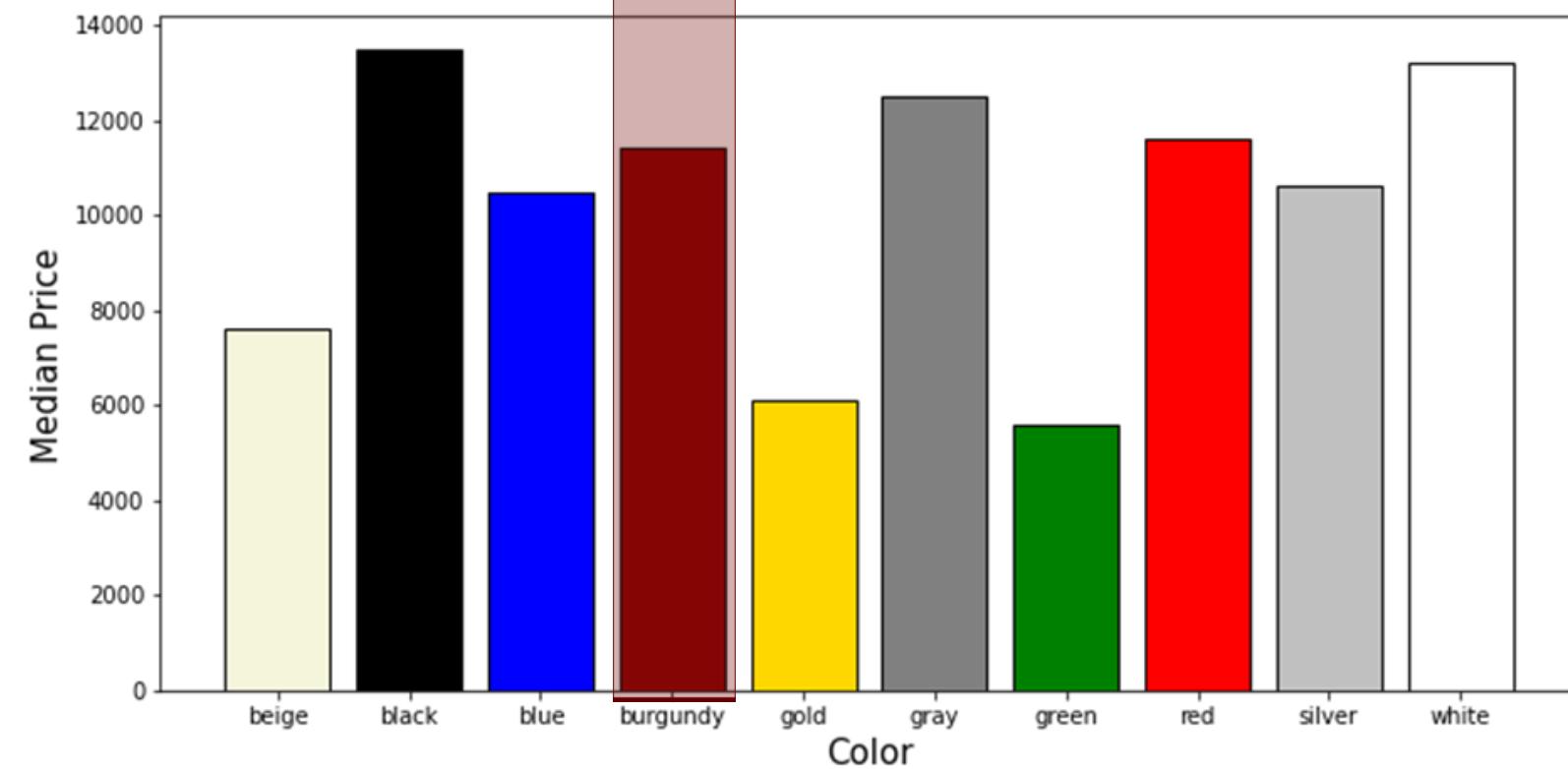
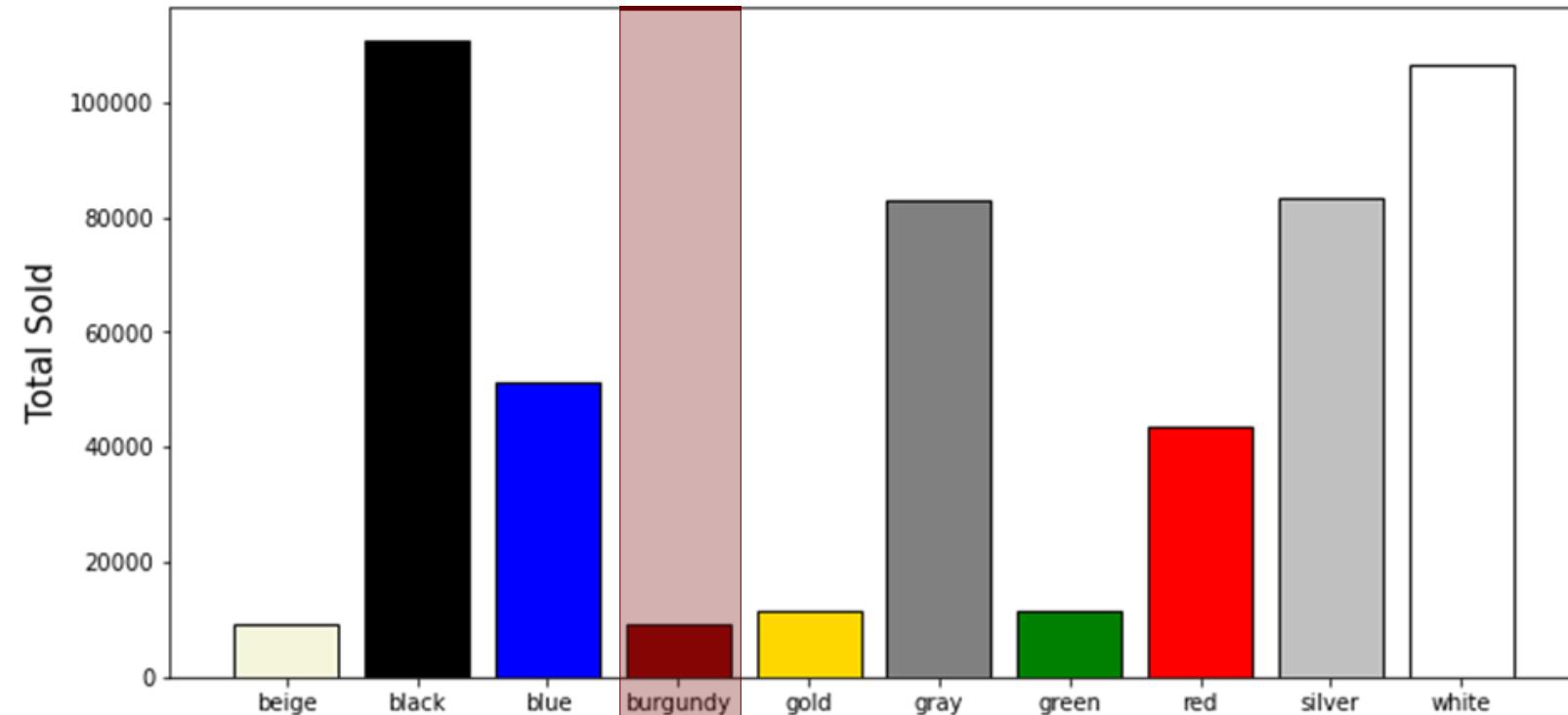
THANK YOU

BUSINESS INSIGHT

+++
+++
+++
+++

+++
+++
+++
+++

Does color influence the sales and prices of Used Cars?



- Warna umum seperti black, white, gray, and silver punya sales dan price yang relatif lebih tinggi.
- Burgundy adalah salah satu warna yang sales-nya sedikit namun prices-nya relatif tinggi,
- Menggunakan teknik marketing berupa tagline "Unique Product" kita dapat melakukan negosiasi harga yang lebih tinggi
- Perusahaan dapat melakukan branding sebagai perusahaan yang menjual mobil bekas dengan warna langka