



NUSANTARA

USED CAR PRICE PREDICTION

USING MACHINE LEARNING



MEET OUR CREW

DATA AND BUSINESS CONTROLLING DEPARTMENT

 **DWI AYU NOUVALINA**

Leader

 **IRHAM MAULANA AHMADI**

 **DONNY RACHMANDA BAHAR**

 **AHMAD GHULAM SHOLIH**

 **MUHAMMAD ABY WICAKSONO**

 **FERNANDA ZUROTUL HIKMAH**

 **DANTA BARUNA RISTIAWAN**

 **ALIF RIZKY DARMAWAN**

ABOUT COMPANY

We are an **e-commerce** company specializing in **selling used cars** to provide the best experience for customers looking for used cars at prices that match their specifications.

The business model used is C2C, where we want to bring together car sellers and buyers to make transactions on our platform.





PROBLEM STATEMENT

- The number of transactions in the 2015 sales period tended to be unstable.
- Previously, there was no appropriate reference that sellers could use to determine the appropriate selling price according to the specifications of the car they wanted to sell.
- The company has not been able to provide car-selling price recommendations that could help sellers.

OBJECTIVES

Increase transactions by providing sellers with more accurate selling price recommendations so they can sell their cars at more competitive prices.

KEY RESULT

01. Create a machine learning model to predict the appropriate selling price more accurately.

02. Improve the seller's experience in using our platform by getting appropriate selling price recommendations.

03. Increase the number of transactions





BUSINESS METRICS



PTM

PRICE TO MARKET

Price adjustments for different markets.

Measuring PTM allows sellers to find opportunities for maximum profit by measuring the value of their cars.



CTR

CLICK-THROUGH RATE

Measures the level of success in implementing the model created by calculating the number of users through clicks made when using the selling price obtained from the recommendations provided.

PROJECT STAGES

1

Analyze and look for insights by exploring data and then visualizing it.

2

Cleaning and processing data to make it ready to use for modeling

3

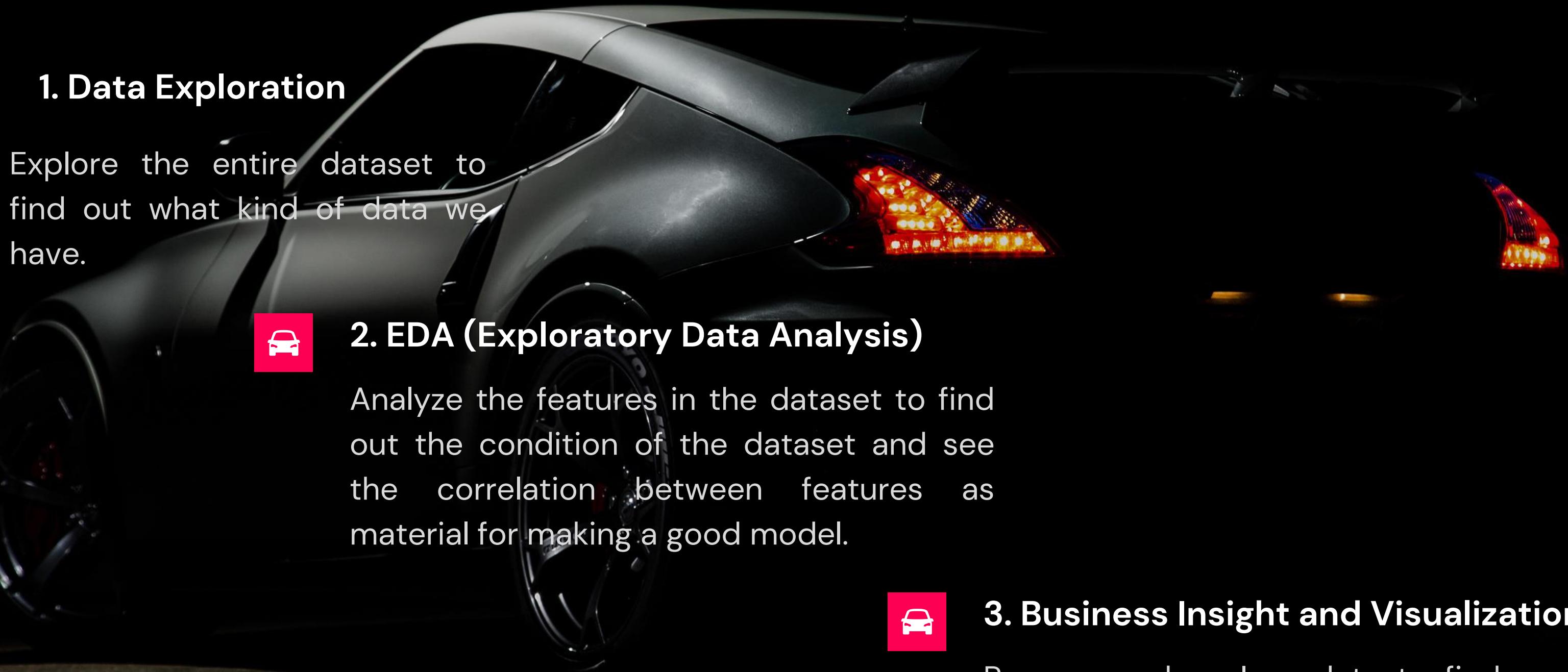
Carry out modelling and evaluation to get the expected output.

STAGE 1



1. Data Exploration

Explore the entire dataset to find out what kind of data we have.



2. EDA (Exploratory Data Analysis)

Analyze the features in the dataset to find out the condition of the dataset and see the correlation between features as material for making a good model.



3. Business Insight and Visualization

Process and analyze data to find business insights, then visualize them to understand current business conditions.

STAGE 1



DATA EXPLORATION

Feature:

1. Year : The year of production of the cars.
2. Make : The brand of the car.
3. Model : The edition of the car of a specific brand.
4. Trim : The trim levels for a car are just different versions of the model.
5. Body : he body style of a vehicle refers to the shape and model of a particular car make.
6. Transmission : The mechanism that moves the power from the engine to the wheels.
7. VIN : Vehichel identification number.
8. State : The state in which the car is auctioned.
9. Condition : The condition of the cars being at the time of auction.
10. Odometer : The distance the car has travelled since manufactured.
11. Color : Exterior color of the car.
12. Interior : Interior color of the car.
13. Seller : The seller of the car (car dealers).
- 14. MMR : Manheim Market Report, Manhiem market record, the market estimated price of the cars.**
15. Sellingprice : The price a car was sold at auctions.
16. Saledate : The date on which the car has been sold.



STAGE 1

DATA EXPLORATION

1. In total, there are 558,811 entries in the data with 16 features
2. Several rows of data are incomplete, and the value writing is still messy.
3. Statistically, there is odd data in the numerical feature where the odometer value is 1, the selling price is \$1, and mmr is \$25.
4. The target feature (label) is an MMR feature.

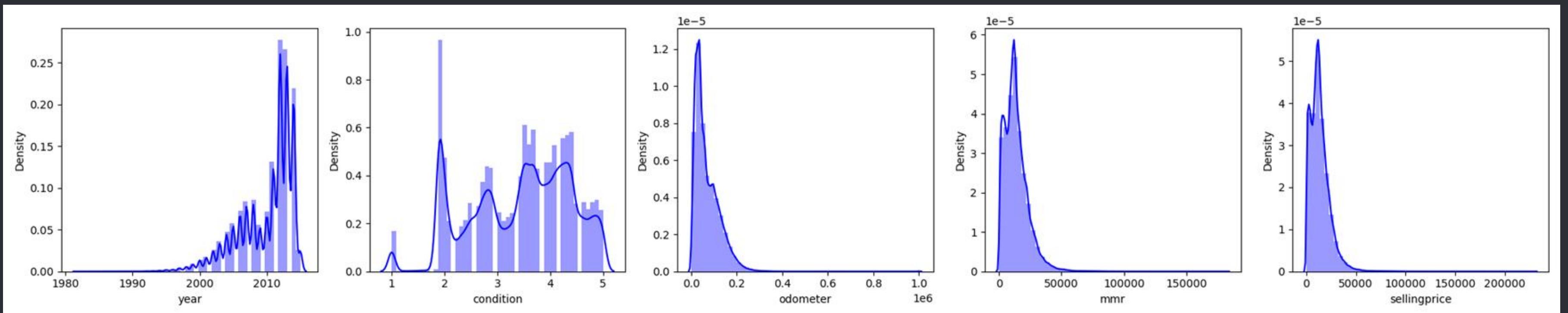


STAGE 1



EDA (EXPLORATORY DATA ANALYSIS)

- Data Distribution

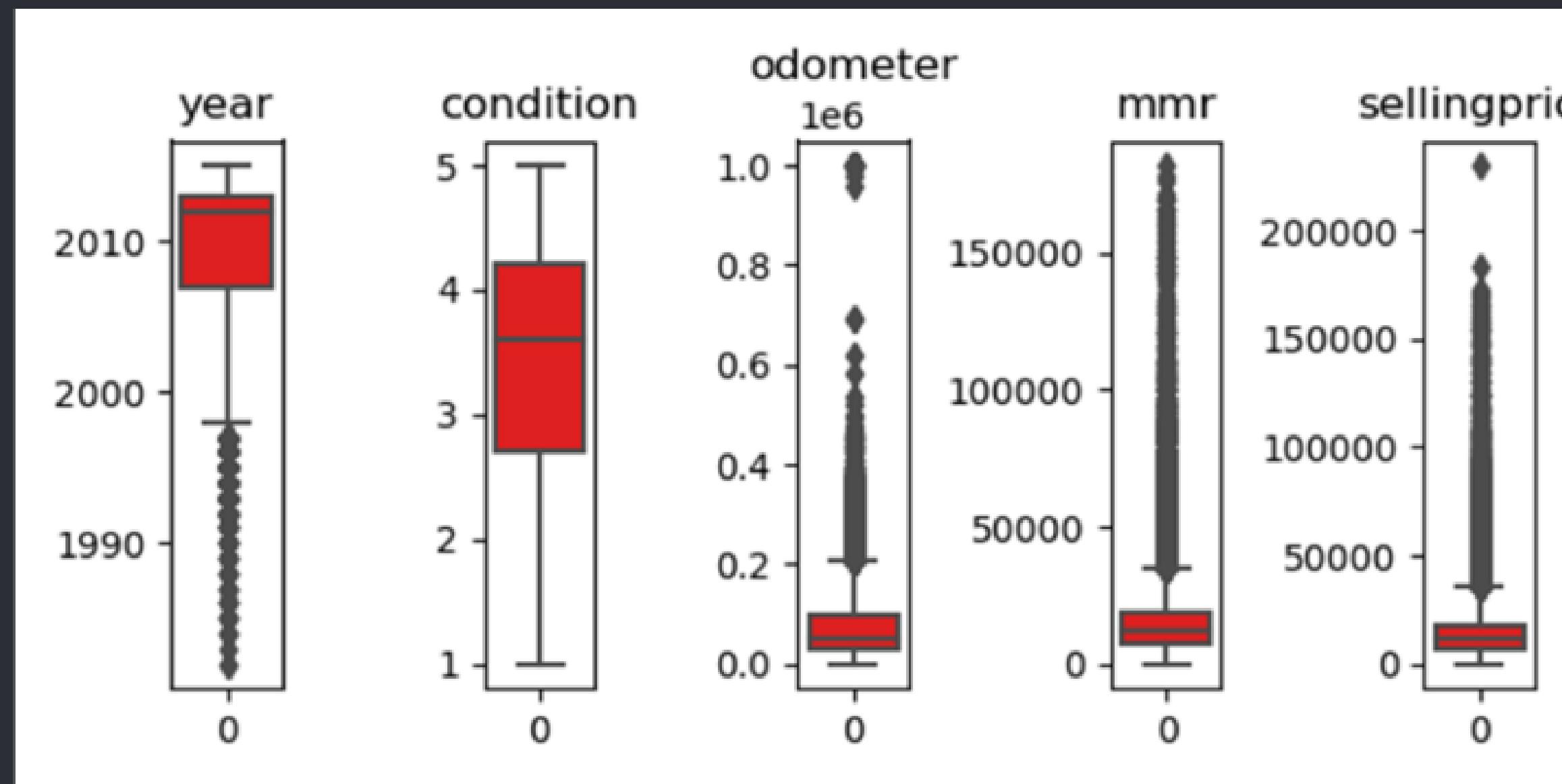


STAGE 1



EDA (EXPLORATORY DATA ANALYSIS)

- UNIVARIATE ANALYSIS
 - Outliers



~~STAGE 1~~

- UNIVARIATE ANALYSIS
 - Unreasonable Value Checking

	year	condition	odometer	mmr	sellingprice
166160	2002	1.900	1.000	5025	1200
137181	2002	2.000	1.000	3100	800
324897	2005	3.700	1.000	8325	1000
459941	1998	2.000	1.000	1475	300
30028	1996	1.000	1.000	1700	425

STAGE 1

- UNIVARIATE ANALYSIS
 - Data Insight

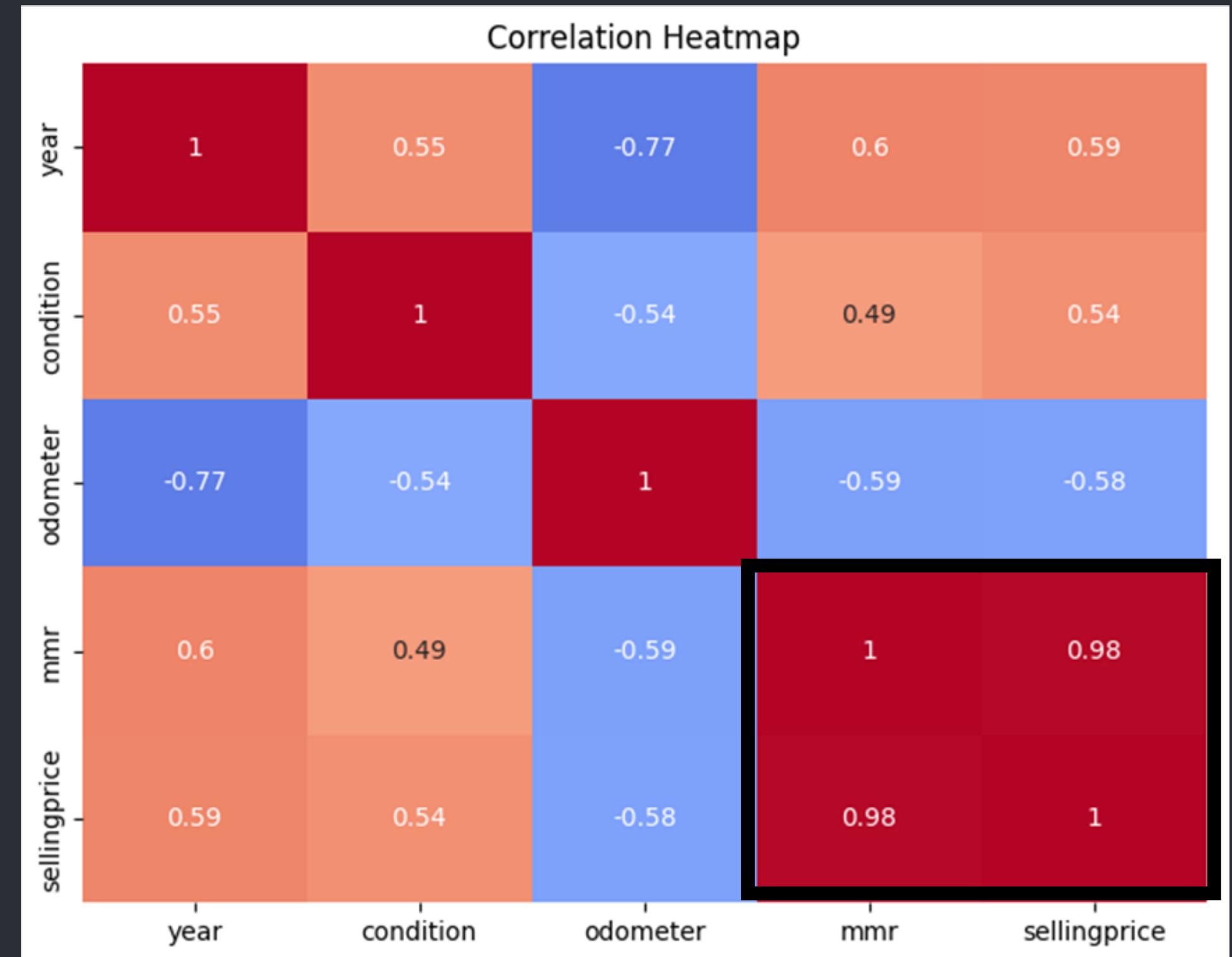
Make	Model	Body	Trim	State	Color	Interior
Ford	Altima	Sedan	Base	California	Black	Black
Chevrolet	F-150	SUV	SE	Texas	White	Gray
Nissan	Fusion	Hatchback	LX	Pennysvalnia	Silver	Beige

STAGE 1



EDA (EXPLORATORY DATA ANALYSIS)

- Correlation Between Numerical Features

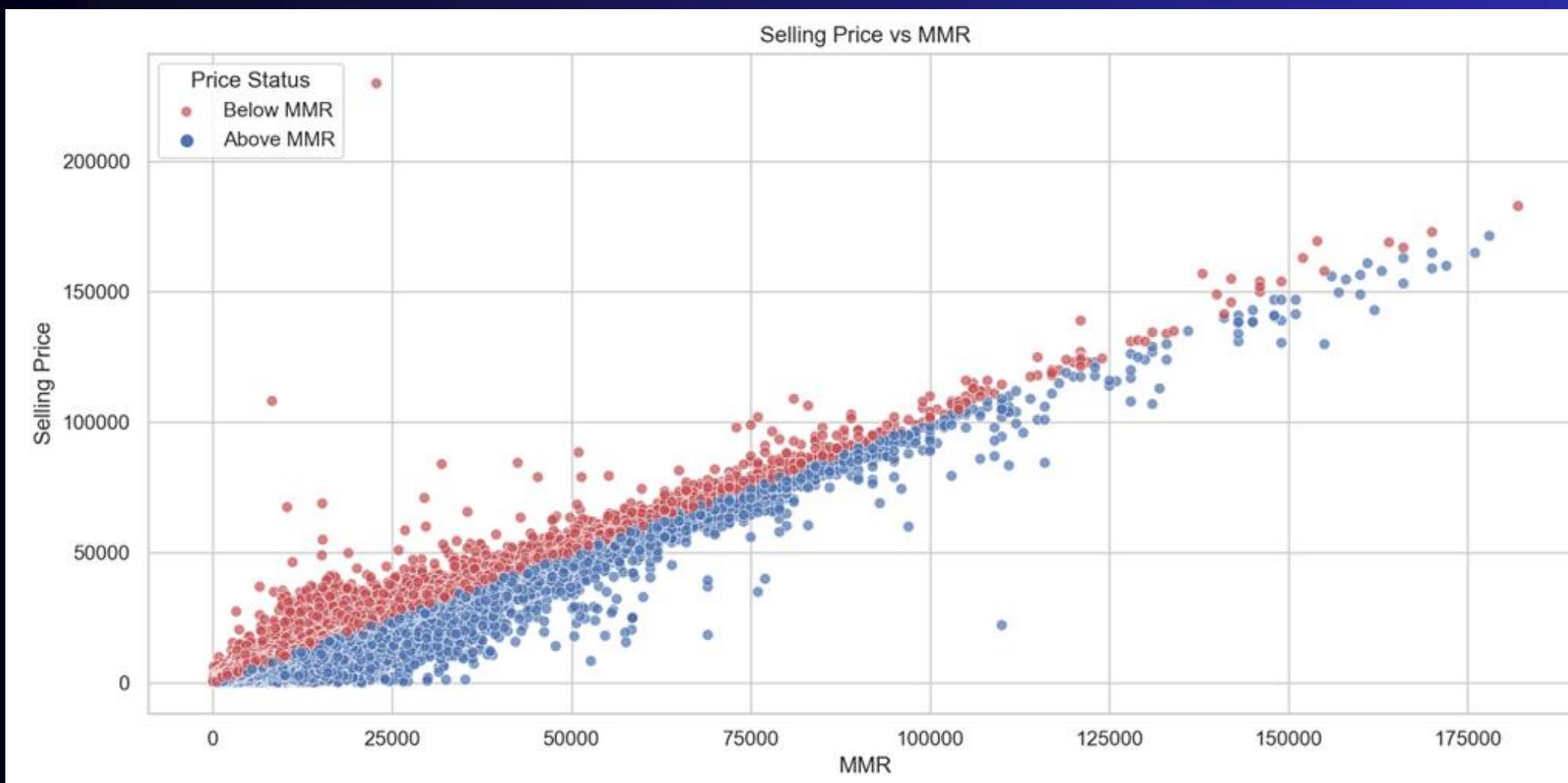


STAGE 1



BUSINESS INSIGHT AND VISUALIZATION

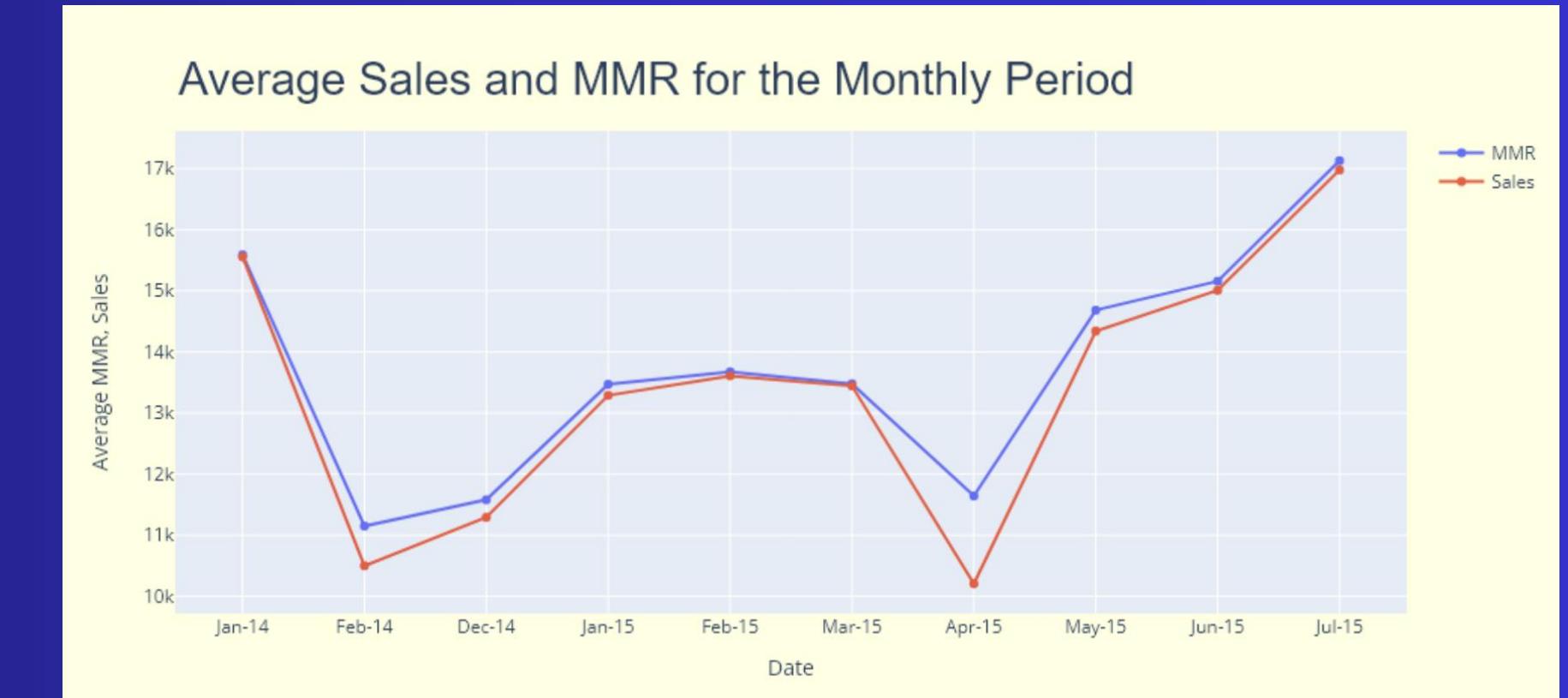
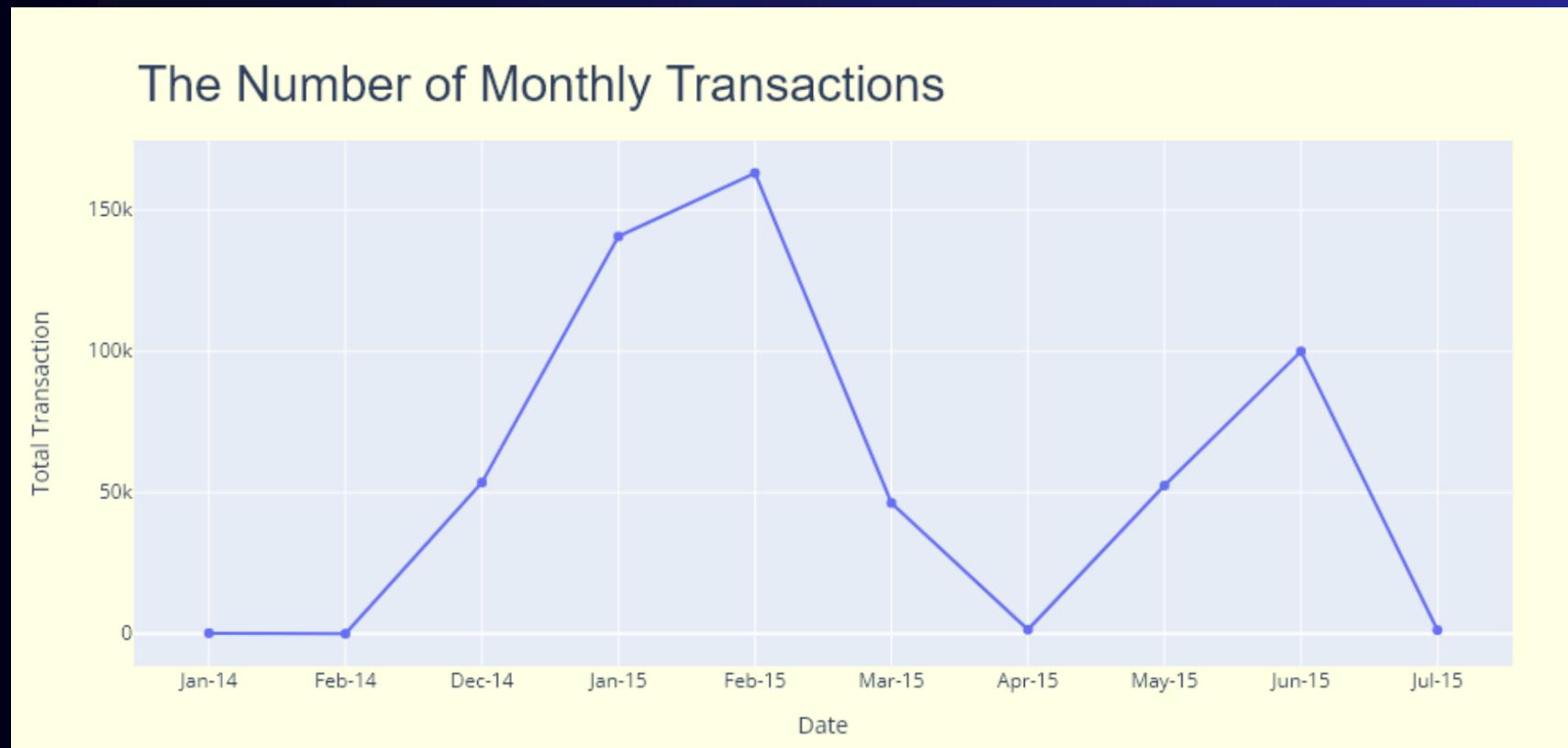
- Sales percentage



Total Transaction	Percentage	Average Car Price
Above mmr	272386	\$14754
Below mmr	286425	\$12523

STAGE 1

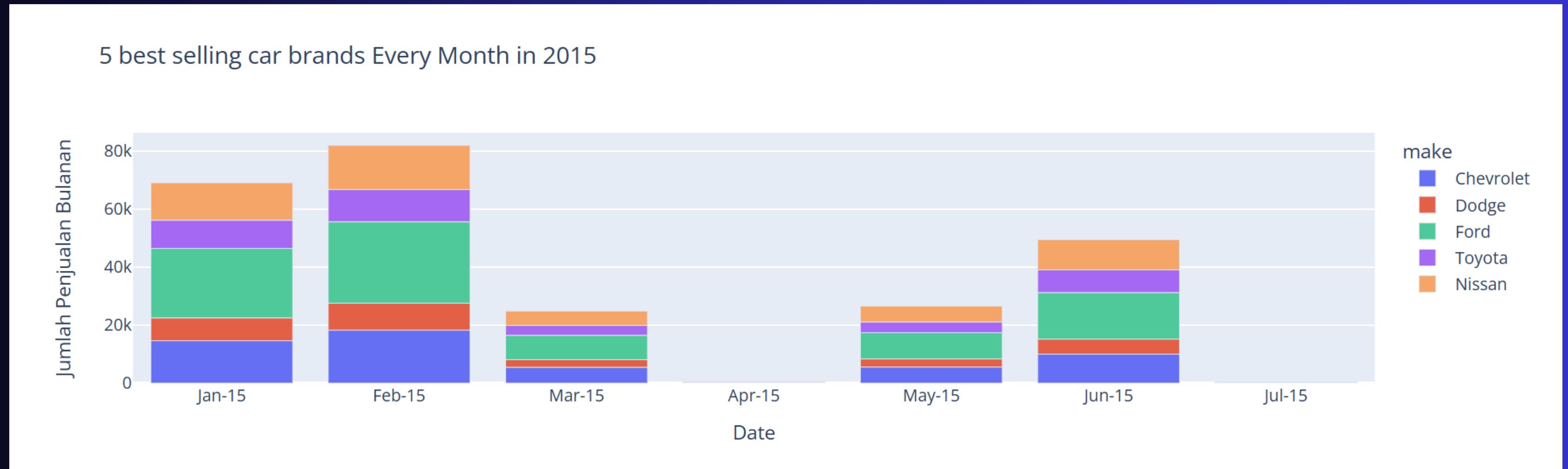
- Total Transaction And Average Sales in Monthly Periods



From the graph above, it can be seen that the peak monthly transactions occurred in February 2015 (around 160k transactions). Still, after that, there was a decline in transactions until April 2015 and increased again until June 2015. The average number of sales each month also seems always lower than the mmr value.

STAGE 1

- Top 5 Car Brands Based on Number of Sales



Some information obtained from the Barplot above is:

1. February has the most sales compared to other months.
2. Ford is the brand that sells the most, especially in January and February 2015. This shows that Ford has a strong market share.
3. April and July had the lowest sales compared to other months in 2015.

STAGE 2

Data Cleansing

Data Preparation





Data Cleaning

Dropping

- Rows with abnormal values
- Irrelevant Features

Unique Values

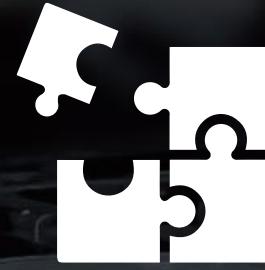
Generalized string writing format

Missing Values

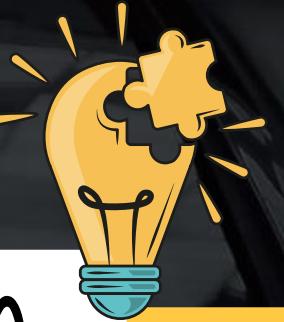
- Categorical
- Numerical
- Certain features

Outliers

Removing outliers based on IQR



Data Preparation



Feature Extraction

- **Overall**

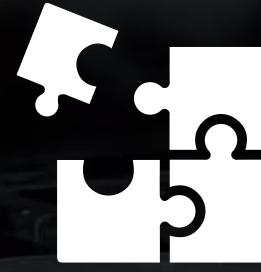
The 'condition' and 'odometer' features are linked to 'mmr', resulting in Poor, Fair, Good, and Very Good.

- **Made In**

The 'make' feature is based on the country of origin of the car maker

- **Top Make**

Fitur 'year' dan 'make' untuk menghasilkan 10 top merk mobil yang terjual berdasarkan tahun pembuatan



Data Preparation

Feature Selection



Overall



Made In



Top Make



Seller



Model



Trim



Body



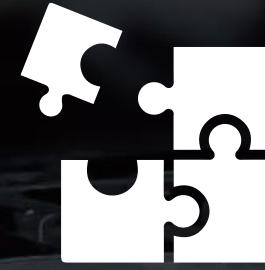
State



Interior



Color



Data Preparation



Feature Encoding

- One Hot Encoding
Overall, Made In, Top Make

- Target Encoding
Model, Body, Trim, dan State

- Count Encoding
Seller, Color, dan Interior



STAGE 3

Modelling and Evaluation



ML ALGORITHMS



LINEAR REGRESSION

MAE : 2468
RMSE : 3340
R2 : 0.80

LASSO REGRESSION

MAE : 2468
RMSE : 3340
R2 : 0.80

RIDGE REGRESSION

MAE : 2468
RMSE : 3340
R2 : 0.80

The three algorithms above have the same model performance and the differences only lie in the coefficients and intercepts.

ML ALGORITHMS



RANDOM FOREST

MAE : 675
RMSE : 1223
R2 : 0.97

CAT BOOST

MAE : 930
RMSE : 1470
R2 : 0.96

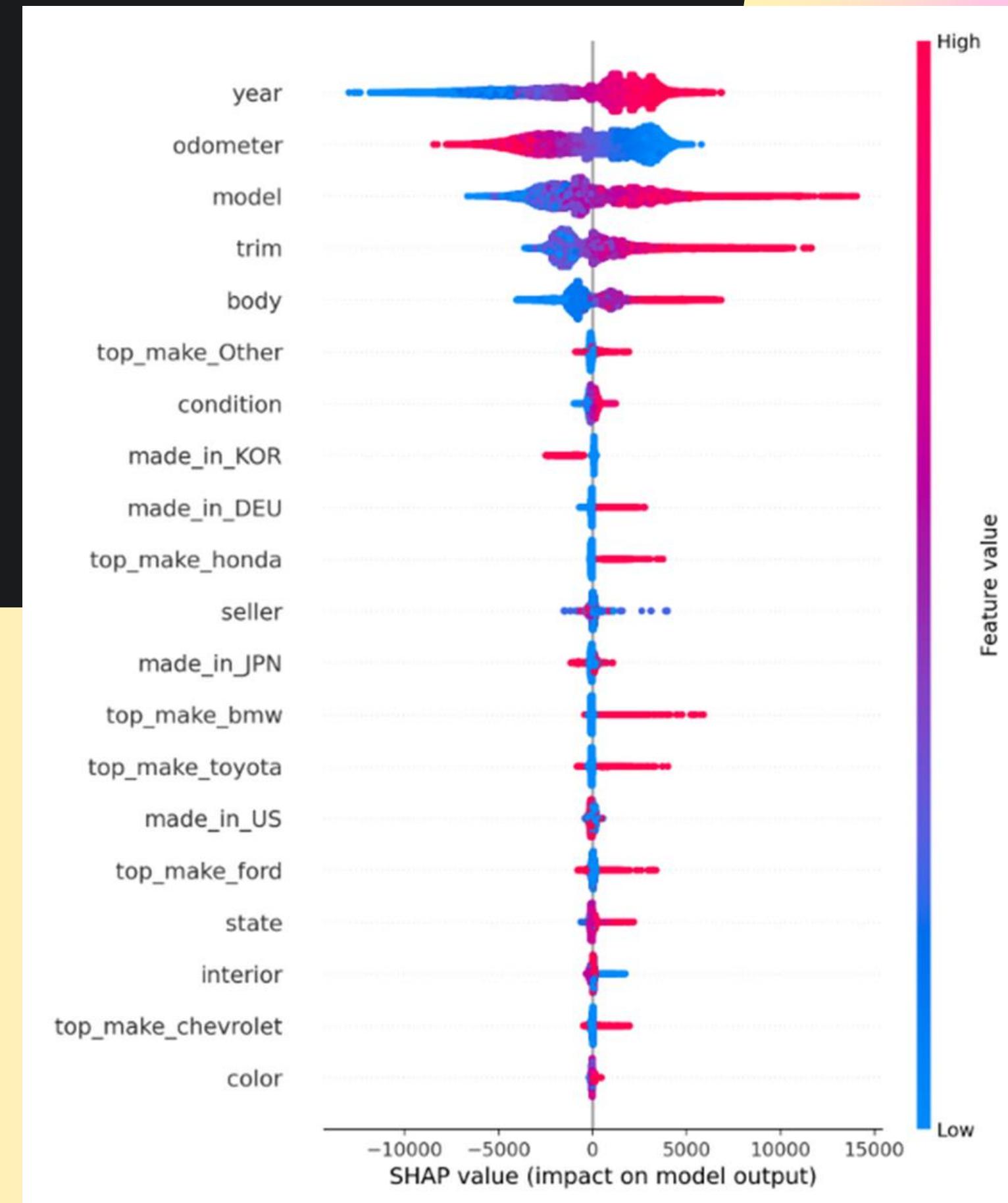
XG BOOST

MAE : 1000
RMSE : 1552
R2 : 0.95

Based on several modeling algorithms that have been tried, we chose Random Forest as the best model because it has the highest R2 and the smallest MAE and RMSE.

Next, we carried out cross validation with 5 k-folds and obtained a mean and STD score of 97%.

SHAP VALUE



NEW PRICE

	mmr	sellingprice_predicted
262193	8775	8931.500
392466	8100	7763.750
524832	20300	19847.000
286417	2500	2865.250
323786	16050	15851.000
...
274895	14800	14576.000
382225	14700	18625.500
495101	11600	12117.500
175086	5425	5277.500
181557	19100	18892.500

105562 rows × 2 columns

Comparison using the same data sample

		Percentage	Total Transaction	Average Car Price
Before Modelling	Above MMR	48.44%	51136	\$14,816
	Below MMR	51.56%	54426	\$12,506
After Modelling	Above MMR	51.88%	54761	\$12,775
	Below MMR	48.12%	50801	\$13,422

BUSINESS RECOMMENDATION



PRICING

● Before Modelling

Average price above MMR = \$14,816

Average price below MMR = \$12,506

● After Modelling

Average price above MMR = \$12,775

Average price below MMR = \$13,422

It can be concluded that the price of a car below the recommended MMR is \$916 (6.8%) higher, so the seller can get more profit.

It is recommended that cars priced above MMR can be sold for \$2041 (13.7%) less so that the price is not too expensive.





RECOMMENDATION SYSTEM

Create a recommendation system that can provide car recommendations to customers so that they are more personalized. This also intends to increase sales by combining the right target and the appropriate selling price.

QUALITY INVENTORY

Based on the insights gained from the condition and odometer of the car, we can carry out Quality Control (QC) to sort the cars to be sold on the platform, for example, only choosing cars with conditions above 2 or those with odometers that are not too high. This aims to increase customer trust in buying used cars through our platform.





**THANK
YOU**