

Laporan Final Project

Kelompok 4 (Nusantara)

Dwi Ayu Nouvalina
Irham Maulana Ahmadi
Ahmad Ghulam Sholih
Fernanda Zurotul Hikmah
Donny Rachmanda Bahar
Muhammad Aby Wicaksono
Danta Baruna Ristiawan
Alif Rizky Darmawan



Latar Belakang Masalah

DATASET: USED CAR AUCTION PRICE

- Kami adalah perusahaan e-commerce dengan model bisnis C2C yang khusus menjual mobil bekas. Setelah dilakukan analisa, data transaksi pada semester pertama tahun 2015 ditemukan jumlah transaksi yang tidak stabil.
- Sebelumnya tidak ada acuan harga yang reliable, sehingga tidak ada dasar dalam menentukan harga jual yang pas untuk customer.
- Belum ada harga yang bisa direkomendasikan perusahaan untuk membantu seller.
- Solusinya adalah dengan mengembangkan model Machine Learning (ML) yang mampu memprediksi harga jual yang paling sesuai dan kompetitif yang nantinya digunakan untuk memberikan harga kepada seller dengan harapan dapat meningkatkan jumlah transaksi.

Data Exploration

List fitur pada dataset adalah sebagai berikut :

1. Year : Tanggal produksi dari mobil.
2. Make : Merk dari mobil.
3. Model : Edisi dari tiap merk mobil.
4. Trim : Versi trim dari mobil.
5. Body : Tipe bentuk dari mobil.
6. Transmission : Transmisi yang digunakan pada mobil.
7. VIN : Vehicle Identification Number.
8. State : Negara bagian tempat mobil dijual.
9. Condition : Kondisi dari mobil pada saat dijual.
10. Odometer : Jarak tempuh mobil semenjak tanggal manufacture.
11. Color : Warna eksterior dari mobil.
12. Interior : Warna interior dari mobil.
13. Seller : Penjual dari mobil (Car dealers).
14. **MMR : Manhiem Market Report, market yang memprediksi harga mobil.**
15. Sellingprice : Nilai jual mobil.
16. Saledate : Tanggal mobil dijual.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 558811 entries, 0 to 558810
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype
---  -
0   year                  558811 non-null  int64
1   make                  548510 non-null  object
2   model                 548412 non-null  object
3   trim                  548160 non-null  object
4   body                  545616 non-null  object
5   transmission          493458 non-null  object
6   vin                   558811 non-null  object
7   state                 558811 non-null  object
8   condition             547017 non-null  float64
9   odometer              558717 non-null  float64
10  color                 558062 non-null  object
11  interior              558062 non-null  object
12  seller                558811 non-null  object
13  mmr                   558811 non-null  int64
14  sellingprice          558811 non-null  int64
15  saledate              558811 non-null  object
dtypes: float64(2), int64(3), object(11)
memory usage: 68.2+ MB
```

a. Data Exploration (Info)

1. Total dari data ada 558,811 entries dengan 16 fitur
2. Tipe dari data int, float dan object, dan terlihat sudah sesuai dengan kolomnya. Namun, untuk kolom 'sale date' tipe data yang awalnya 'object', dirasa perlu diubah ke date time.
3. Terdapat nilai Null pada beberapa fitur
4. Dataset tidak memiliki data duplikat sehingga tidak diperlukan penanganan.

```
dfcar.isnull().sum()
```

year	0
make	10301
model	10399
trim	10651
body	13195
transmission	65353
vin	0
state	0
condition	11794
odometer	94
color	749
interior	749
seller	0
mmr	0
sellingprice	0
saledate	0
dtype:	int64

```
dfcar.duplicated().sum()
```

```
0
```

b. Data Exploration (Numerical)

	year	condition	odometer	mmr	sellingprice
count	558811.000	547017.000	558717.000	558811.000	558811.000
mean	2010.039	3.425	68323.196	13769.325	13611.262
std	3.967	0.949	53397.753	9679.875	9749.657
min	1982.000	1.000	1.000	25.000	1.000
25%	2007.000	2.700	28374.000	7100.000	6900.000
50%	2012.000	3.600	52256.000	12250.000	12100.000
75%	2013.000	4.200	99112.000	18300.000	18200.000
max	2015.000	5.000	999999.000	182000.000	230000.000

1. Condition memiliki value antara 1-5.
2. Nilai dari odometer terlihat memiliki jarak yang jauh antara min dan max dimana nilai min nya adalah 1 dan max nya hampir 1 juta. Tentu ini adalah hal yang terkesan janggal. Selain itu, nilai mean dan median nya juga terhitung sangat jauh yaitu 16000.
3. Secara keseluruhan, nilai MMR dan selling price tidak terlalu berbeda. Namun, nilai min-max dan mean-median nya sangat jauh, sehingga perlu diperhatikan lebih lanjut untuk tahap berikutnya.
4. Beberapa kolom di fitur numerical juga terdapat null value, sehingga perlu ditangani pada tahap berikutnya.

c. Data Exploration (Categorical)

```
dfcar[cat].describe()
```

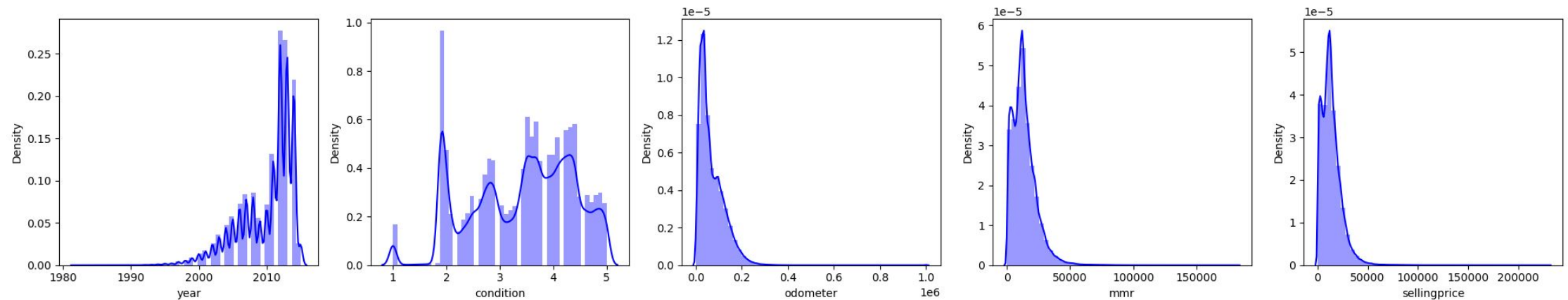
	make	model	trim	body	transmission	vin	state	color	interior	seller	saledate
count	548510	548412	548160	545616	493458	558811	558811	558062	558062	558811	558811
unique	96	973	1975	86	2	550296	38	20	17	14264	3745
top	Ford	Altima	Base	Sedan	automatic	wbanv13588cz57827	fl	black	black	nissan-infiniti It	Tue Feb 10 2015 01:30:00 GMT-0800 (PST)
freq	93554	19349	55817	199437	475914	5	82945	110970	244325	19693	5334

1. Dari data diatas, kita bisa melihat secara sekilas nilai apa yang paling teratas, sering muncul, dan berapa jumlah unique value dari masing-masing fitur.
2. Sebagai contoh, Ford merupakan brand yang paling populer di antara 96 total brand lain. Selain itu, mobil tipe sedan juga termasuk yang paling favorit.
3. Transmisi automatic lebih dicari customer dibanding manual dengan jumlah lebih dari 90%.
4. Warna hitam tetap mendominasi baik di eksterior maupun interior.
5. Untuk hasil 'first' dan 'last' hanya berlaku terhadap data date, sehingga tidak heran jika kolom yang lain dikatakan null value secara keseluruhan.

Exploratory Data Analysis

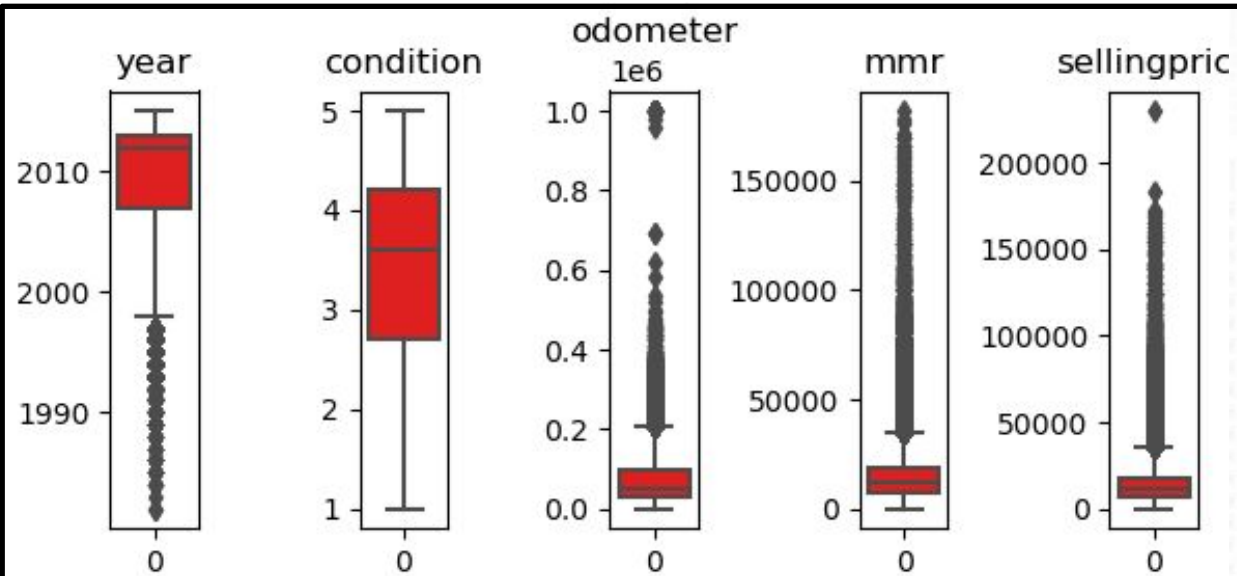
a. Univariate Analysis

- Distribusi Data



1. Dari kolom 'year' kita bisa melihat bahwa penjualan cenderung mengalami peningkatan pada setiap tahunnya dan puncak penjualan berada di antara tahun 2010 hingga 2015. Kita juga bisa melihat bahwa kolom 'year' termasuk left-skewed.
2. Pada grafik 'condition', kondisi mobil yang berada di kisaran angka 2 menjadi mayoritas dari kondisi lainnya. Hal ini menunjukkan bahwa kondisi mobil bekas bernilai kurang baik. Namun, tidak sedikit juga kondisi mobil yang bernilai 3 keatas hingga 4.
3. Untuk odometer dapat terlihat bahwa mayoritas berada di angka 0 hingga 0.2 yang artinya odometer berkisar antara 0 hingga 200k. Sedangkan untuk pemakaian mobil diatas 200k terhitung jarang.
4. Kolom 'mmr' dan 'selling price' terlihat serupa dimana harga jual mobil berkisar di angka 0-50000.
5. Dari kolom-kolom diatas, kolom 'odometer', 'mmr', dan 'selling price' termasuk right-skewed.

- Outliers



Dapat dilihat bahwa mmr dan selling price memiliki outlier paling banyak, begitu juga dengan kolom odometer dan year. Untuk kuartil dan mediannya pun juga terkesan jauh. Hal ini dapat terjadi karena banyaknya data mobil (500k++) sehingga nilai mmr dan selling price maupun odometer juga sangat bervariasi.

Kolom yang terlihat paling normal hanyalah kolom 'condition'. Hasil analisis ini merupakan bentuk visual dari statistik data yang telah kita lakukan sebelumnya pada tahap Data Exploration. Faktor ini disebabkan besarnya data yang juga meningkatkan nilai variasi pada kolom-kolom tsb.

Note terkait proses data pre-processing:

- Untuk distribusi numeric yang skewed, perlu dilakukan normalisasi
- Perlu penanganan terhadap nilai-nilai outlier, termasuk nilai-nilai janggal yang disebutkan di atas
- Untuk kolom kosong juga perlu penanganan lebih lanjut, apakah akan diisi dengan median atau dihapus sekaligus

- Check Nilai yang Tidak Wajar

Contoh



	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

Beberapa kejanggalan ditemukan pada data, diantaranya:

- Nilai minimal tidak wajar dari kolom 'mmr' (25), 'selling price' (1), dan 'odometer' (1)
- Dari data min dan max kolom 'odometer' diatas terlihat bahwa data tersebut tidak hanya mengandung nilai min dan max yang jauh tetapi pada kolom lain juga tidak normal. Hal tersebut bisa dilihat dari sampel diatas dimana baris min dan max tersebut juga banyak mengandung null value, nilai mmr dan selling price yang rendah, dan kondisi yang tidak bagus. Tidak hanya itu, baris-baris ini juga terhitung banyak yaitu lebih dari 1000 data.
- Nilai min dari selling price yang bernilai 1, terdapat 4 data dari keseluruhan dataset. Hal ini juga perlu menjadi catatan dimana hal ini terkesan tidak masuk akal.
- Dari data diatas dimana nilai mmr hanya 25, juga terdapat kejanggalan pada kolom yang lainnya. Yaitu, banyak terdapat null value, nilai 'selling price' yang tergolong kecil, kondisi yang tidak bagus.

- Data Insight

	make	count
0	Ford	93554
1	Chevrolet	60197
2	Nissan	53946
3	Toyota	39871
4	Dodge	30710

Ford adalah merek mobil paling favorit, diikuti oleh Chevrolet, Nissan, Toyota, dan Dodge. Perbedaan penjualan sangat besar, dengan mobil-mobil teratas terjual puluhan ribu unit, sedangkan yang terendah hanya satu unit.

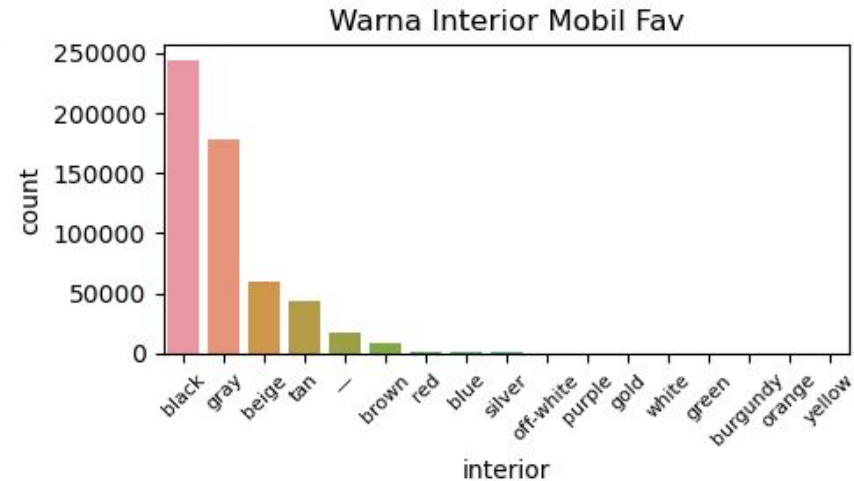
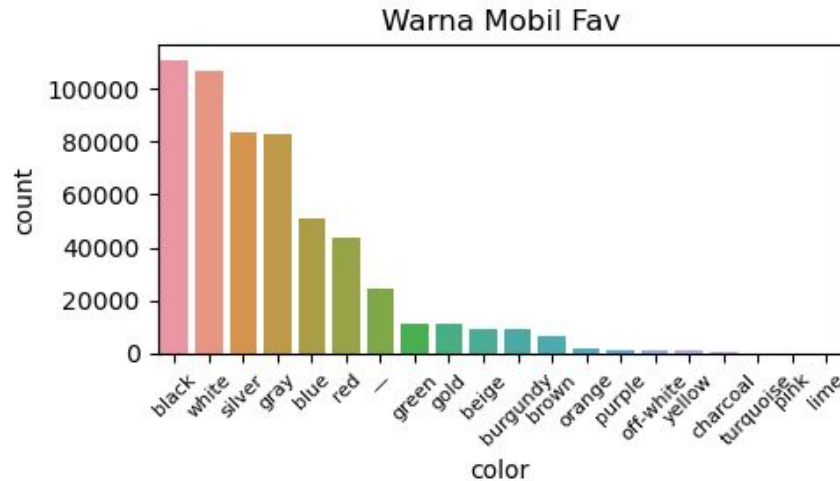
	model	count
0	Altima	19349
1	F-150	14479
2	Fusion	12946
3	Camry	12545
4	Escape	11861

Sementara itu, Altima menjadi model mobil bekas paling populer, diikuti oleh F-150 sebagai peringkat kedua. Ini mencerminkan pola yang sama dengan analisis merk mobil, di mana perbedaan antara yang paling populer dan yang kurang populer sangat signifikan.

Adapun insight lainnya adalah :

- Mayoritas mobil yang dijual menggunakan transmisi automatic.
- Pembelian mobil cenderung terjadi di akhir tahun hingga awal tahun.
- Pembelian mobil terjadi pada top hari yaitu hari selasa, rabu, dan kamis. Dengan pembelian terbanyak terjadi pada hari selasa.

- Data Insight



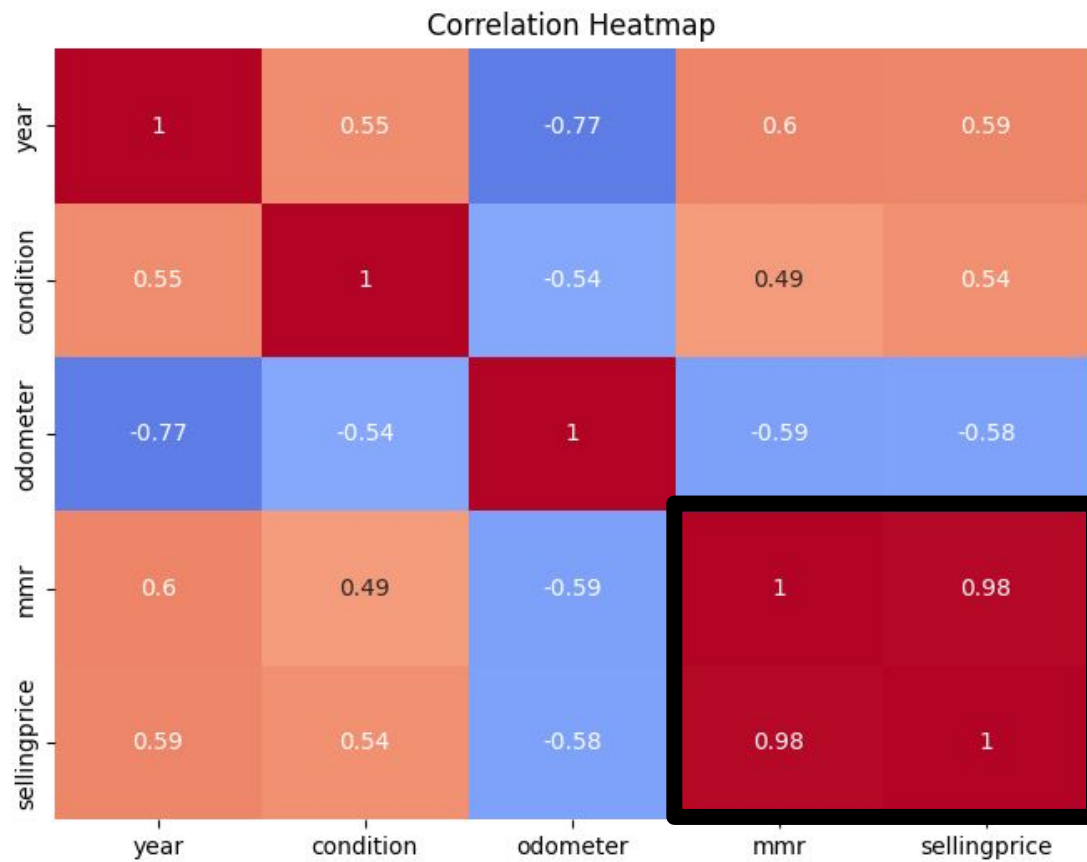
- Warna mobil dan warna interior mobil paling disukai adalah warna hitam.
- Warna interior mobil yang lazim digunakan adalah hitam, abu-abu, beige, dan tan.
- Terdapat missing value (-) pada warna dan interior mobil yang perlu dilakukan penanganan pada tahap berikutnya

Note terkait proses data pre-processing:

- Perlu dipertimbangkan mengubah lower/uppercasse terhadap nilai-nilai kategorik
- Kolom 'model' dan 'make' memiliki nilai unique yang banyak (beberapa duplicate lower/uppercasse), perlu penyederhanaan kategori
- Untuk kolom kosong juga perlu penanganan lebih lanjut, apakah akan diisi dengan modus atau dihapus sekaligus

b. Multivariate Analysis

- Korelasi Antar Fitur Numerik

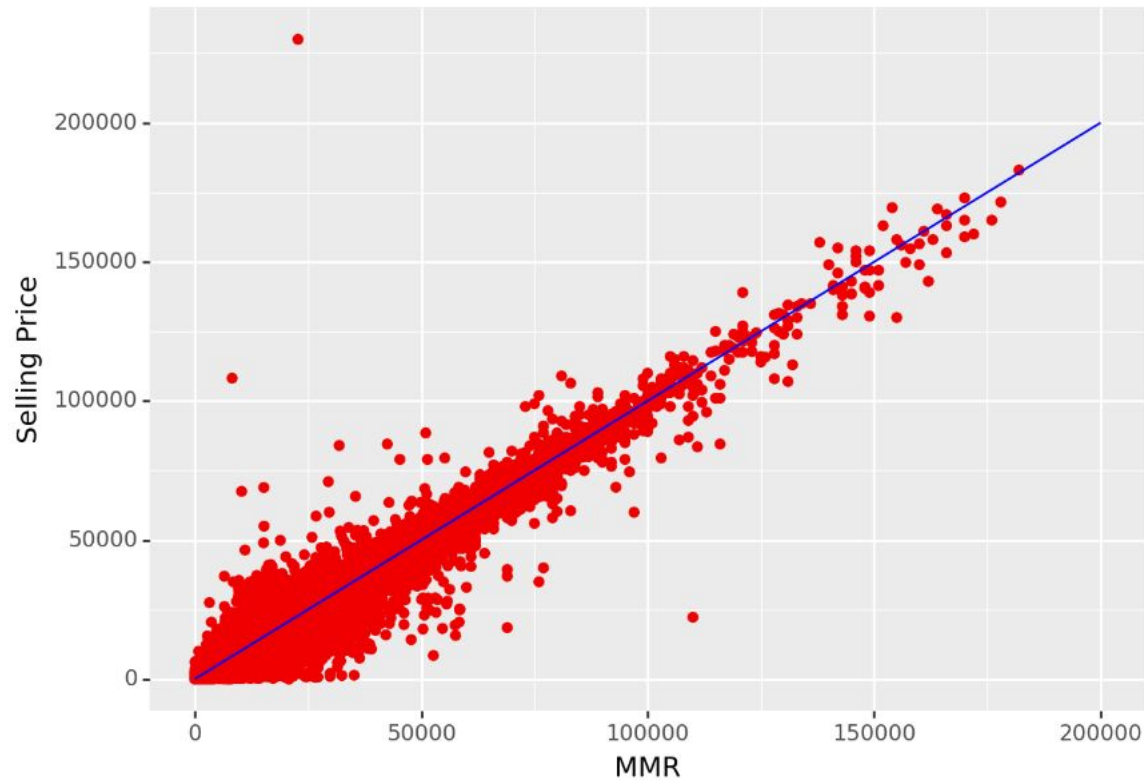


Dari grafik disamping, dapat disimpulkan:

- 'Year' berkorelasi positif dengan 'selling price', mengindikasikan mobil semakin baru, harganya cenderung lebih tinggi.
- 'Condition' berhubungan positif dengan 'selling price', artinya semakin baik kondisi mobil, semakin tinggi harganya.
- 'Odometer' berkorelasi negatif dengan 'selling price', menunjukkan mobil dengan jarak tempuh lebih rendah cenderung memiliki harga jual lebih tinggi.
- 'MMR' memiliki korelasi positif kuat dengan 'selling price', menandakan prediksi harga MMR berhubungan baik dengan harga jual aktual.
- Kolom 'Year' dan 'Odometer' juga menunjukkan korelasi negatif yang cukup tinggi, artinya mobil keluaran terbaru memiliki odometer yang cenderung kecil

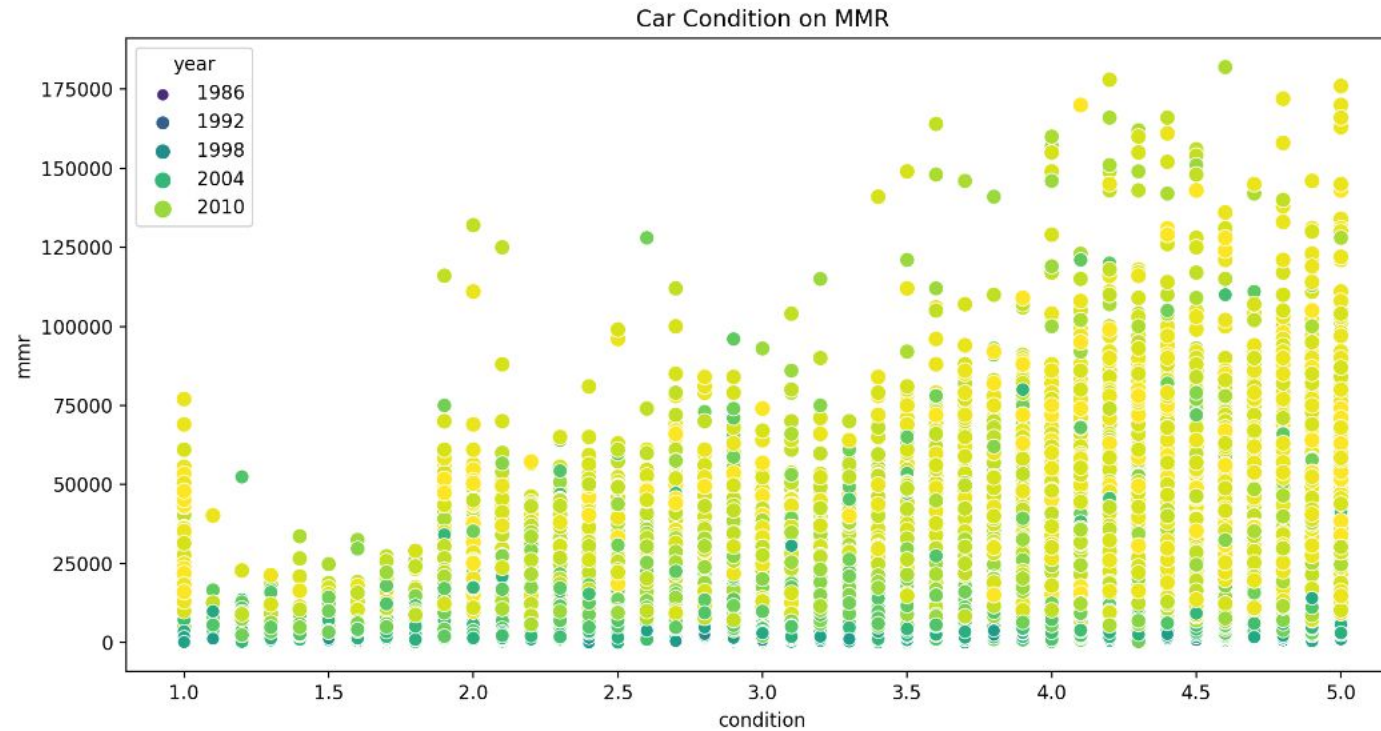
- Korelasi MMR dan Selling Price

MMR vs Selling Price



Grafik disamping menunjukkan bagaimana persebaran selling price terhadap MMR. MMR merupakan singkatan dari Manhiem Market Report yaitu sebuah spesifik report yang menjadi indikator wholesale price untuk penjualan kendaraan. Perhitungan harga didasarkan pada lebih dari 10 juta transaksi penjualan selama 13 bulan sebelumnya dengan harga yang tepat oleh penjualan kendaraan nyata di Manheim auction houses. Sehingga, bisa dikatakan bahwa MMR layak untuk dijadikan acuan harga.

- Korelasi Antara Condition dan MMR



Dari grafik diatas dapat disimpulkan bahwa pada setiap rentang kondisi mobil dengan mmr yang relatif rendah, terdapat mobil keluaran tahun 2004 keatas. Hal itu bisa mengindikasikan bahwa banyak mobil keluaran tahun 2004 yang dijual bekas. Selain itu, produksi mobil sejak 2010 mengalami kenaikan pada kondisi nya dan mmr yang semakin tinggi pula. Hal tersebut menunjukkan bahwa mmr memiliki pengaruh kuat terhadap bagus nya kondisi dan tahun produksi mobil.

Business Insight and Visualization

a. Persentase penjualan



	total_transaction	percentage
is_sold_below_mmr		
0	272386	48.740
1	286425	51.260

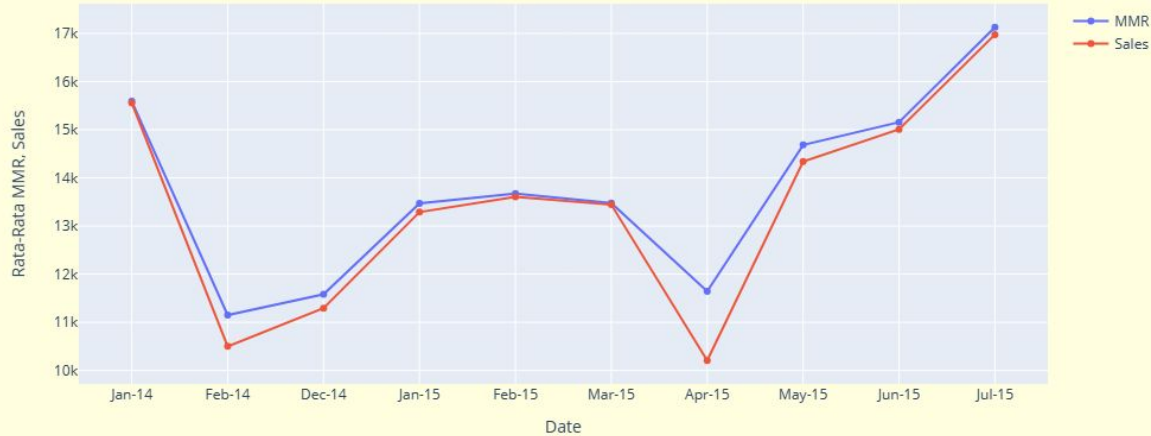
Mean of Selling Price above MMR = 14754.737728811318
Mean of Selling Price below MMR = 12523.834148555468

Perhitungan persentase penjualan dilakukan untuk mengetahui performa penjualan selama ini. Setelah dilakukan perhitungan, diperoleh informasi bahwa terdapat sekitar **51%** dari total transaksi berada dibawah harga MMR.

b. Rata-rata Sales dan MMR Dalam Periode Bulanan dan Harian

Menindak lanjuti temuan total persentase penjualan, selanjutnya dibuat plot rata-rata penjualan berdasarkan beberapa periode waktu

Rata-Rata MMR dan Sales Bulanan

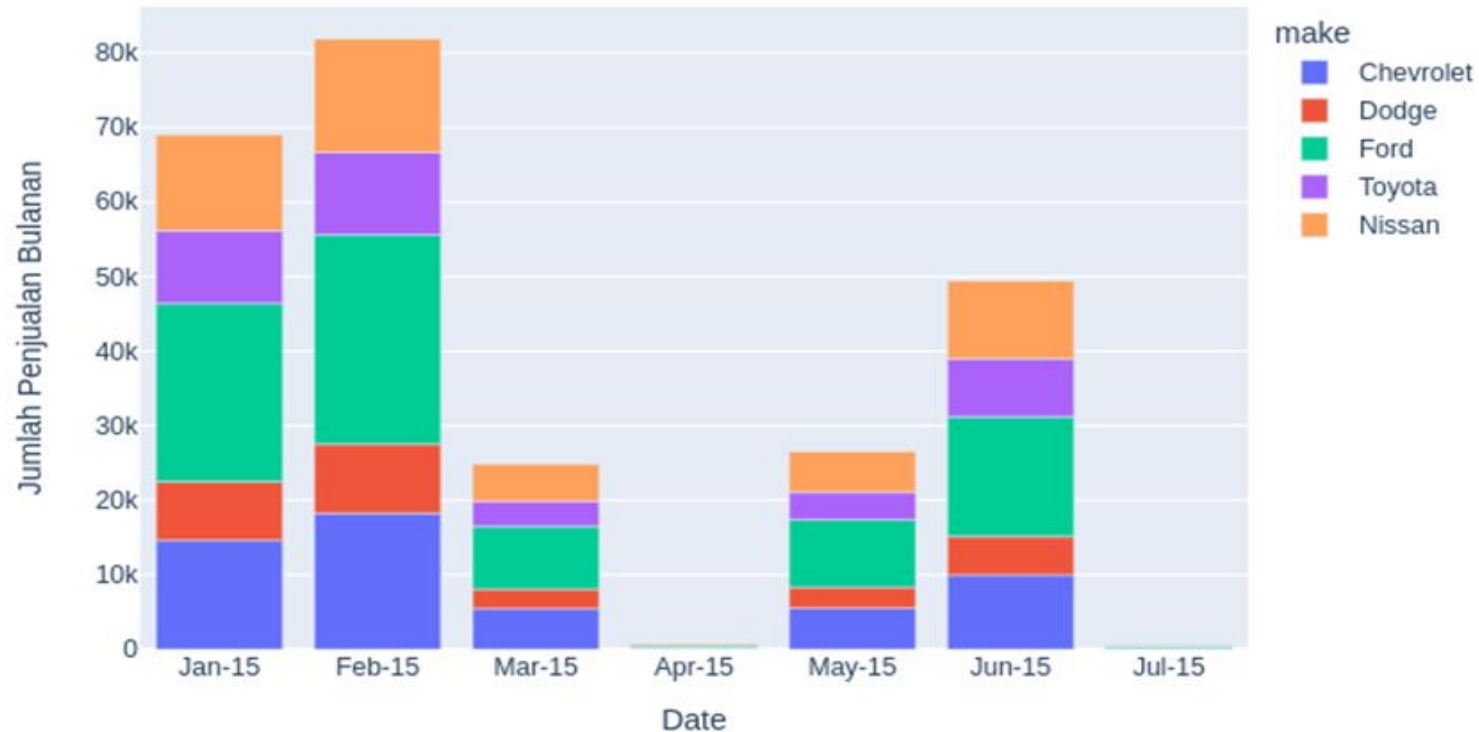


Jumlah Transaksi Bulanan



Dari grafik diatas diperoleh informasi bahwa puncak transaksi Bulanan terjadi pada Februari 2015 (sekitar 160k transaksi), setelahnya terjadi penurunan transaksi hingga April 2015 dan kembali naik hingga Juni 2015. Jumlah rata-rata sales di setiap bulannya juga terlihat selalu lebih rendah daripada nilai mmr nya.

c. 5 Brand Mobil paling Banyak Terjual Setiap Bulan di Tahun 2015



Dari Bar Plot diatas diperoleh informasi

1. Bulan Februari merupakan penjualan terbanyak dibanding bulan lainnya.
2. Mobil ford adalah merek yang paling banyak terjual terutama pada Januari dan Februari 2015. Hal ini menunjukkan bahwa Ford memiliki pangsa pasar yang kuat.
3. Bulan April dan Juli merupakan penjualan paling rendah dibandingkan dengan bulan lainnya di tahun 2015.

Data Cleansing

A. Dropping

1. Duplicated Rows

Pada tahap Data Exploration sebelumnya, terlihat bahwa tidak ada duplicate rows sehingga tidak perlu penanganan.

2. Rows with Extreme Values

Pada tahap EDA sebelumnya, kita menemukan baris-baris yang memiliki outliers ekstrim dan tidak masuk akal. Maka, di tahap data cleaning ini kami menghapus baris-baris tersebut.

3. Irrelevant Features

Beberapa fitur terlihat kurang relevan dan unique value yang dimiliki terlalu luas seperti yang sudah kami analisis sebelumnya. Sehingga diasumsikan bahwa fitur tersebut tidak memiliki pola tertentu. Fitur tersebut adalah `vin`, `selling price`, dan `saledate` yang tidak memiliki korelasi dan pola tertentu.

B. Handle Unique Value

Setelah dilakukan pembersihan data dengan menyamaratakan penulisan string, maka terlihat hasilnya dimana :

- a. ``make`` yang awalnya 96 kini hanya 61 unique values
- b. ``model`` yang awalnya 973 kini hanya 841 unique values
- c. ``trim`` yang awalnya 1975 kini hanya 1883 unique values
- d. ``body`` yang awalnya 86 kini hanya 42 unique values
- e. ``seller`` yang awalnya 14264 kini hanya 14152 unique values
- f. ``color`` dan ``interior`` berkurang 1 nilai dari unique value sebelumnya dikarenakan sudah digantikan oleh nilai modus (black).
- g. ``state`` adalah satu-satunya yang masih bernilai sama.

C. Handle Missing Value

Terdapat 9 fitur yang memiliki Null value, yaitu ``make``, ``model``, ``trim``, ``body``, ``transmission``, ``condition``, ``odometer``, ``color``, ``interior``.

Namun, dikarenakan fitur color dan interior telah dihapus, maka kita hanya perlu meng-handle fitur lainnya.

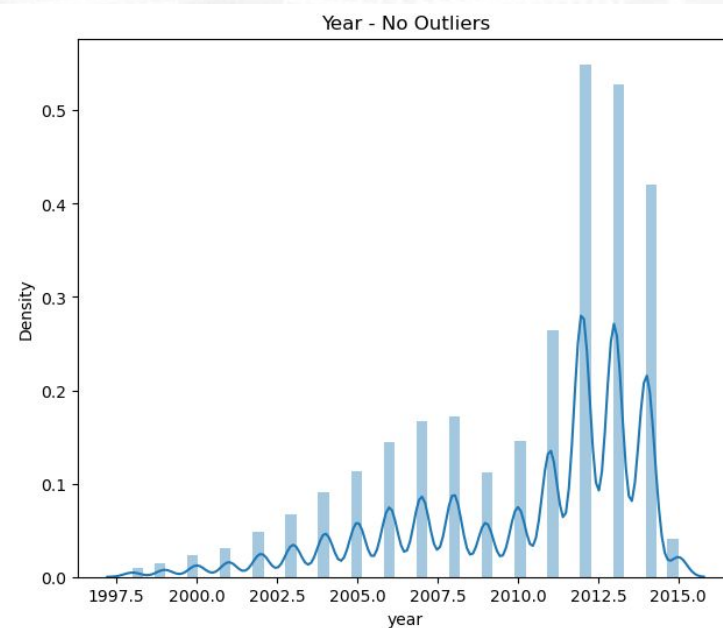
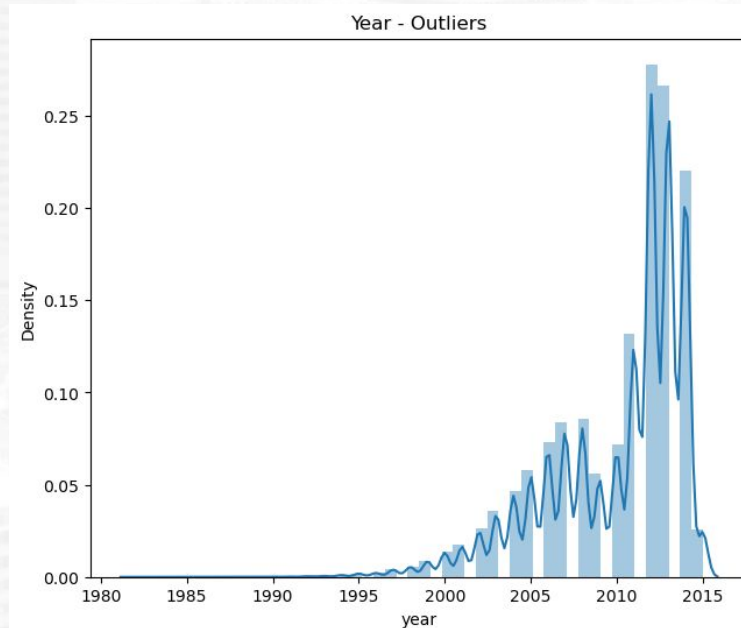
1. Seluruh data kategorikal diisi dengan nilai modus.
2. ``condition`` yang memiliki persebaran data yang lumayan normal, maka diisi dengan nilai mean.
3. ``odometer`` hanya memiliki 94 null values, maka dari itu kami menghapusnya.

D. Handle Outlier (IQR)

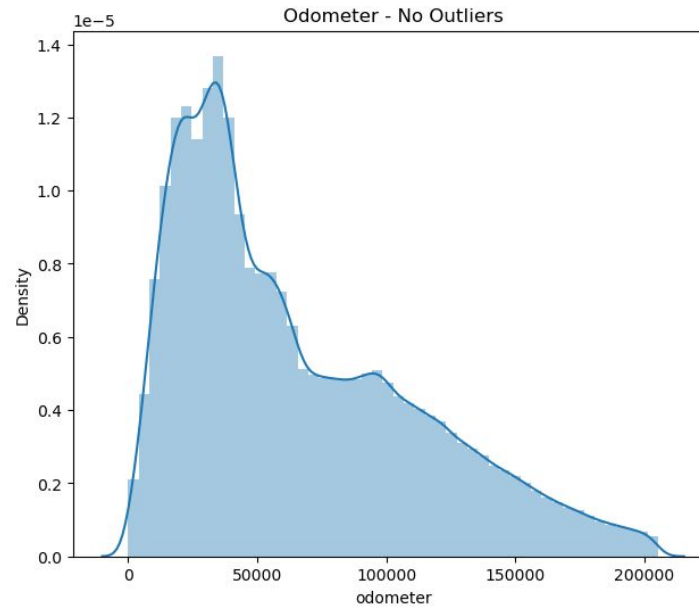
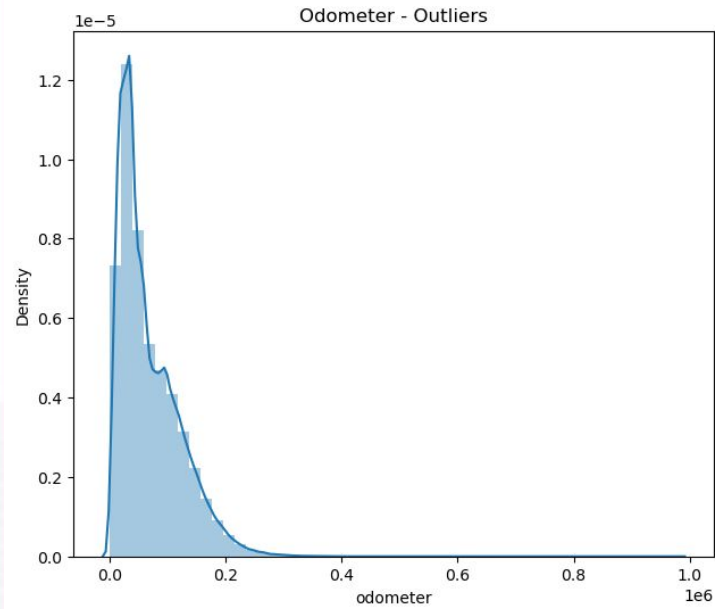
Jumlah Outliers pada kolom year : 6033 dan persentase outliers: 1.08 %

Jumlah Outliers pada kolom odometer : 10289 dan persentase outliers: 1.85 %

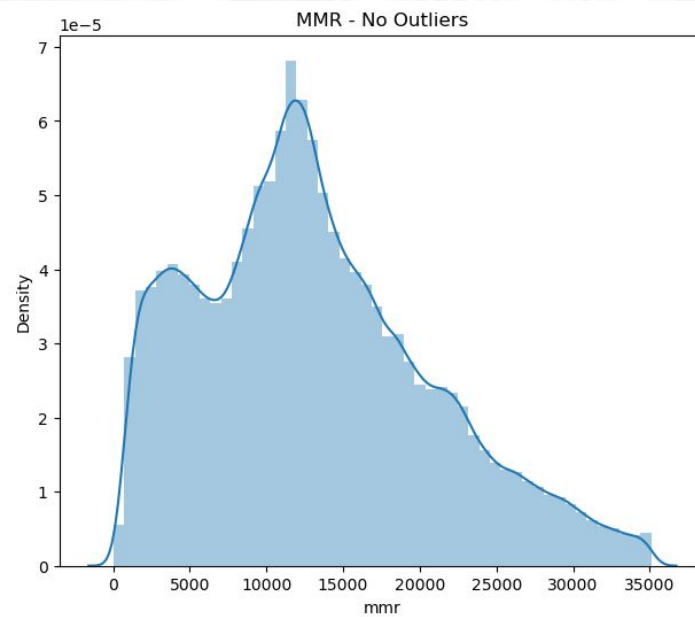
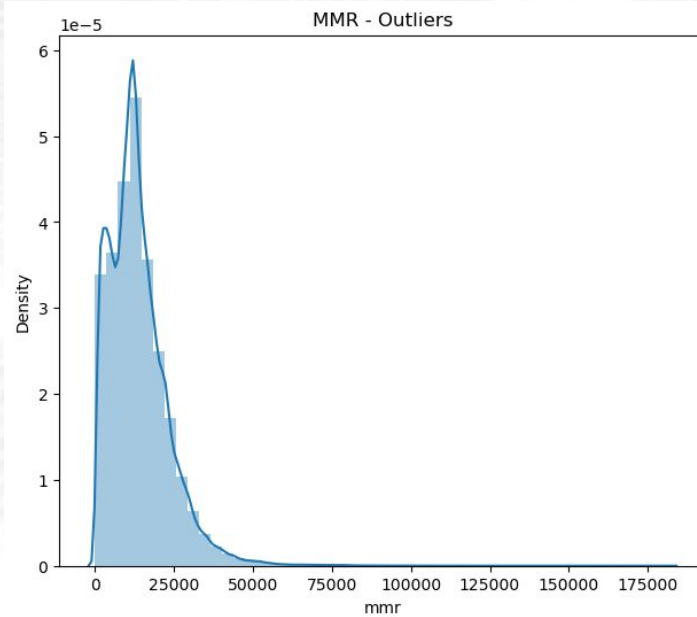
Jumlah Outliers pada kolom mmr : 16307 dan persentase outliers: 2.93 %



Koefisien Skewness Awal year : -1.186
Koefisien Skewness Baru year : -1.017



Koefisien Skewness Awal odometer : 1.239
Koefisien Skewness Baru odometer : 0.841



Koefisien Skewness Awal mmr : 1.998
Koefisien Skewness Baru mmr : 0.544

Data Preparation

A. Feature Extraction

Pada tahapan ini dilakukan pembuatan Feature – Feature baru dari Feature – Feature yang telah ada sebelumnya. Beberapa Feature baru yang dibuat diantaranya :

1. **Overall** (Ekstraksi fitur `condition` dan `odometer`)

Fitur yang berisikan nilai dimana (Poor, Fair, Good, Very Good) dihasilkan melalui ekstraksi fitur kondisi dan odometer berdasarkan nilai mean mmr nya.

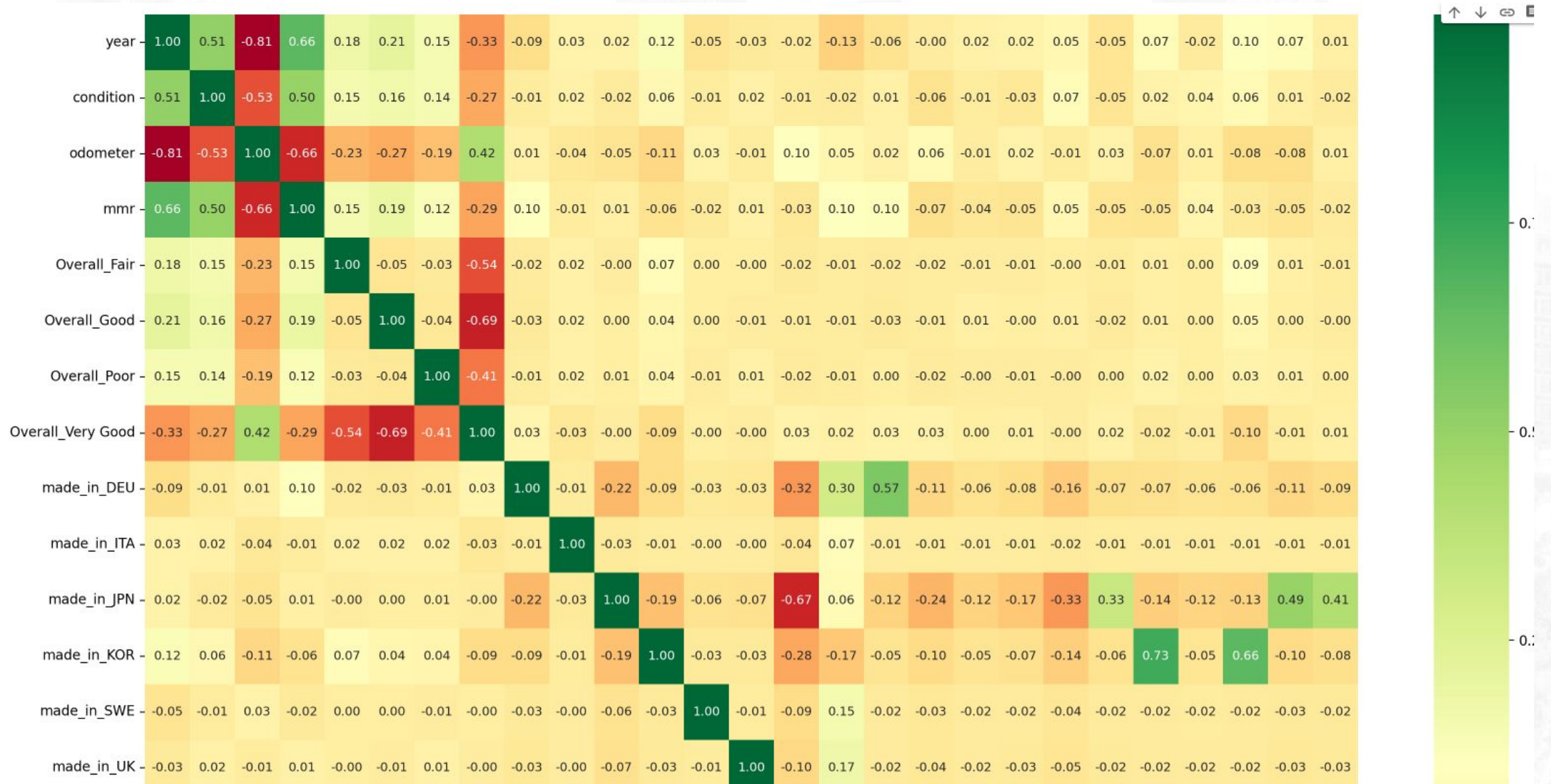
2. **made_in** (Ekstraksi fitur `make`)

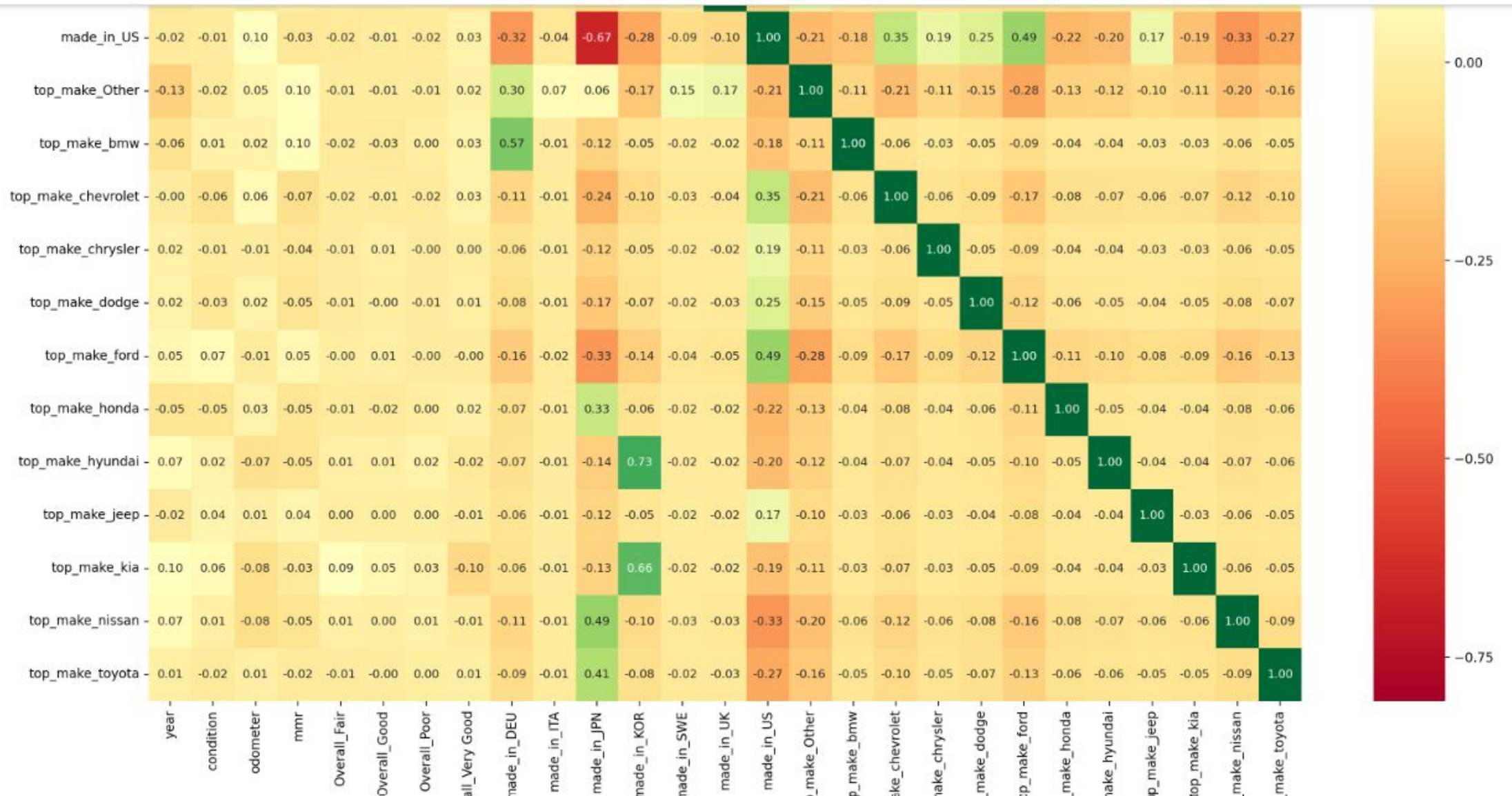
Feature yang berisikan kode negara dimana perusahaan pembuat mobil berasal. Feature ini diperoleh setelah dilakukan feature extraction pada kolom `make` yang dikaitkan dengan list negara asal yang telah dibuat. Dengan pengelompokkan ini diharapkan dapat membantu penggambaran kolom `make` pada model dan membantu meningkatkan akurasi model

3. **top_make** (Ekstraksi fitur `year` dan `make`)

Mengekstraksi fitur make dengan menjadikan 10 top merk mobil dengan penjualan mobil terbanyak dari tahun pembuatan mobil.

Correlation Matrix 2





B. Feature Encoding

Selanjutnya adalah melakukan encoding terhadap fitur-fitur yang masih bertipe object, yaitu :

1. One-Hot Encoding : Overall, made in, top make

...	top_make_chevrolet	top_make_chrysler	top_make_dodge	top_make_ford	top_make_honda	top_make_hyundai	top_make_jEEP	top_make_kia
...	0	0	0	0	0	0	0	1
...	0	0	0	0	0	0	0	1
...	0	0	0	0	0	0	0	0

2. Target Encoding : Fitur model, trim, body, dan state
3. Count Encoding : Fitur seller, color, interior

C. Feature Selection

Memilih atau menghapus fitur-fitur yang tidak lagi diperlukan untuk masuk ke tahap pemodelan.

Modelling and Evaluation

Untuk metode awal kami menggunakan Linear Regression serta Regularization Lasso dan Ridge.

A. Linear Regression

MAE: 2468.738886308443
RMSE: 3340.305301190921
R2 score: 0.8068971312419558

Dikarenakan skor yang kurang memuaskan, kami akan mencoba beberapa metode lain yaitu Random Forest dan Boosting.

B. Lasso

MAE: 2468.790584640927
RMSE: 3340.33560087233
R2 score: 0.8068936279801807

C. Ridge

MAE: 2468.751038282878
RMSE: 3340.3150323384825
R2 score: 0.8068960061260888

D. Random Forest Regressor

```
MAE train: 249.24368197061415
MAE test: 675.3143389666736
RMSE train: 451.59455435411223
RMSE test: 1223.7941125480165
R2 score train: 0.9964442452180161
R2 score test: 0.9740801120792703
```

E. Catboost Regressor

```
MAE train: 918.5333764968341
MAE test: 930.3384890568701
RMSE train: 1439.3781961134798
RMSE test: 1470.3113808830235
R2 score train: 0.9638769635308235
R2 score test: 0.9625859228072585
```

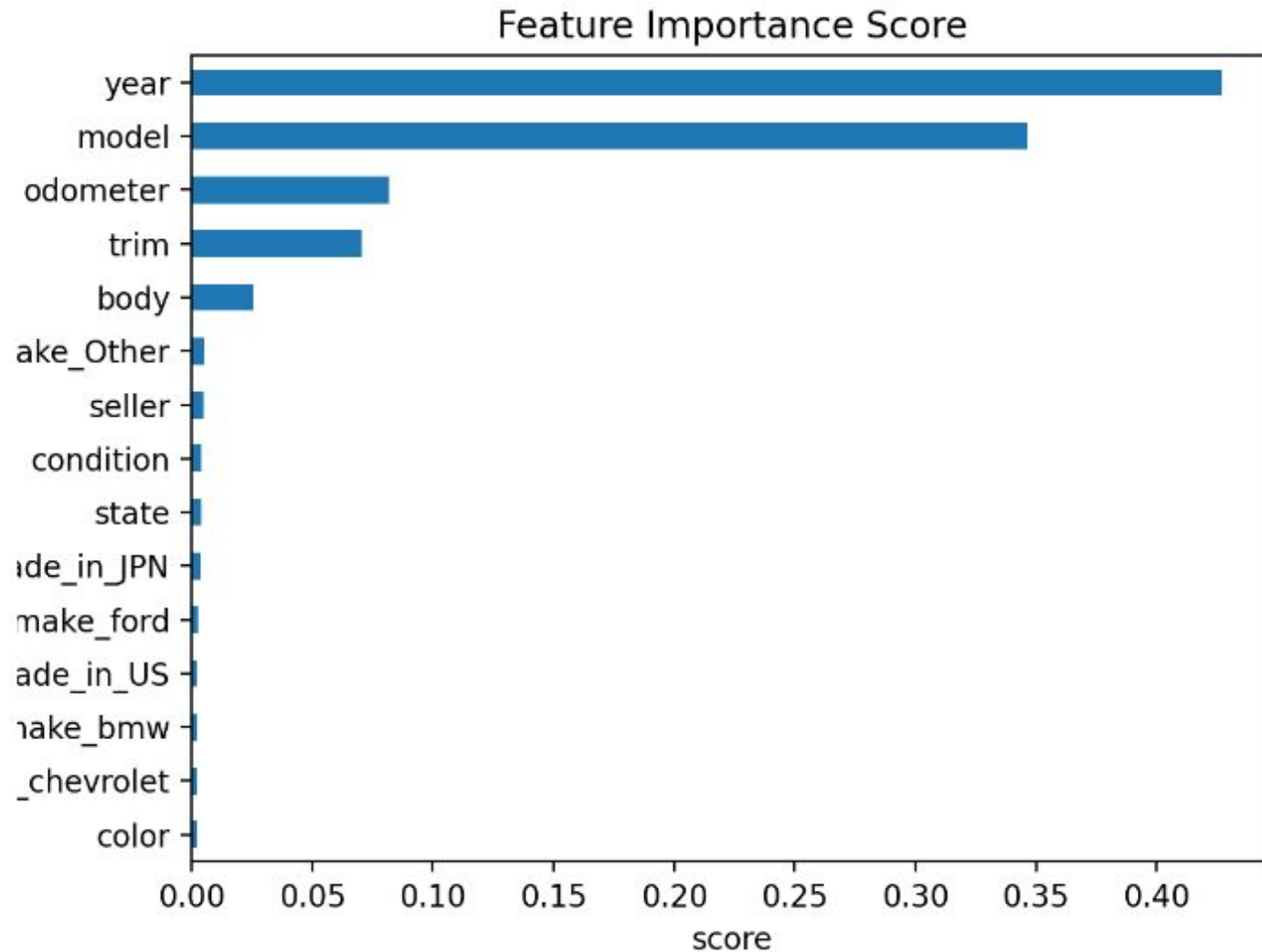
F. XGBoost Regressor

```
MAE train: 976.6041523871259
MAE test: 1000.2500079120172
RMSE train: 1500.1049592424351
RMSE test: 1552.3379414399467
R2 score train: 0.9607646343907252
R2 score test: 0.9582949213059464
```

Dari beberapa algoritma pemodelan yang sudah kami coba, kami memilih Random Forest sebagai model terbaik karena memiliki R2 yang paling tinggi serta MAE dan RMSE paling kecil. Model juga tidak overfit maupun underfit. Lalu, selanjutnya kami melakukan cross validation dengan 5 k-fold. Berikut nilai CV dengan mean dan std nya.

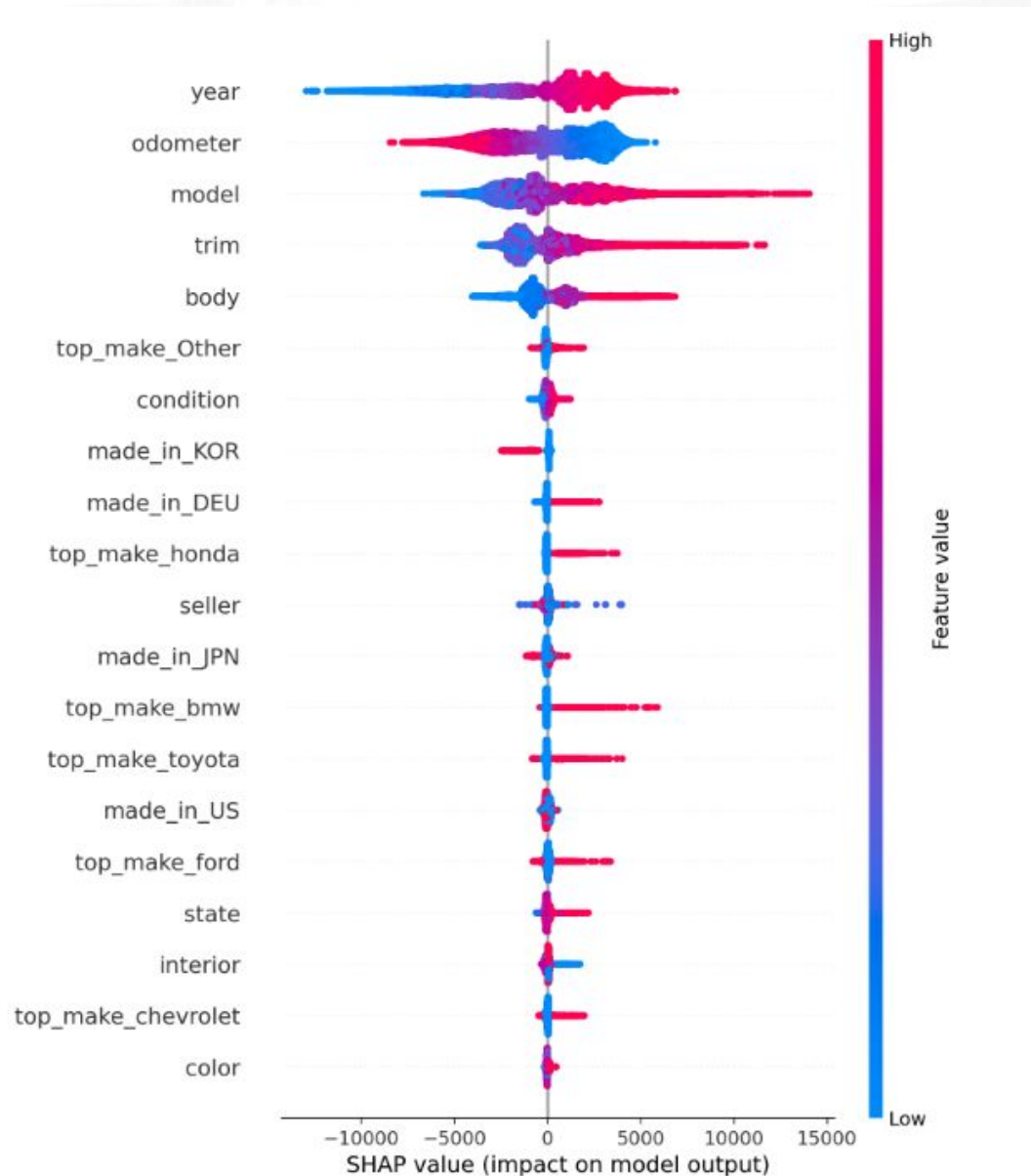
```
Accuracy with cross validation, 5 = [97.22%, 97.42%]
```

Feature Importance Graph



Pada bar chart feature importance di samping terlihat bahwa kesenjangan fitur yang paling relevan dengan yang lainnya sangatlah jauh. Hal ini mengindikasikan bahwa dalam penentuan harga jual mobil bekas, fitur-fitur yang dominan lah yang pada dasarnya cukup untuk menjadi faktor penentu harga jual mobil, sedangkan faktor lainnya hanyalah faktor pendukung.

SHAP Value Graph



Jika kita lihat menggunakan SHAP graph, terlihat bahwa semakin baru tahun keluaran mobil, semakin tinggi pula harganya, begitu pula dengan model dan trim. Sebaliknya untuk odometer, semakin rendah nilainya justru harga semakin tinggi.

Sedangkan untuk fitur-fitur lainnya seperti asal negara mobil, tipe body, kondisi mobil, dll menunjukkan nilai yang tidak terlalu besar. Hal ini menunjukkan bahwa faktor tersebut hanyalah faktor tambahan dalam menentukan harga jual mobil.

RESULT

New Selling 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

Perbandingan dengan menggunakan sampel data yang sama

		Persentase	Total Transaksi	Rata-rata Harga Mobil
Sebelum Modelling	Di atas MMR	48.44%	51136	\$14,816
	Di bawah MMR	51.56%	54426	\$12,506
Sesudah Modelling	Di atas MMR	51.88%	54761	\$12,775
	Di bawah MMR	48.12%	50801	\$13,422

Dengan sampel sebesar 20%, (105.562 data) prediksi harga yang dibuat menunjukkan bahwa persentase harga dibawah mmr adalah 48.12% dimana angka ini menunjukkan penurunan sekitar 3% dari data awal dan harga diatas mmr juga 3% mengalami kenaikan dibanding data awal.

Recommendation

1. PRICING

Perbandingan dengan menggunakan sampel data yang sama

Sebelum Modelling

Rata-rata Harga Di atas MMR = \$14,816	48.44%
Rata-rata Harga Di bawah MMR = \$12,506	51.56%

Sesudah Modelling

Rata-rata Harga Di atas MMR = \$12,775	51.88%
Rata-rata Harga Di bawah MMR = \$13,422	48.12%

Setelah dilakukan pemodelan, rata-rata harga dibawah mmr adalah sebesar \$13,422. Jika dibandingkan dengan rata-rata harga mobil yang dijual dibawah mmr sebelumnya yaitu \$12,506, kita bisa menarik kesimpulan bahwa harga mobil direkomendasikan dijual \$916 (6.8%) lebih tinggi, sehingga seller bisa mendapatkan keuntungan yang lebih banyak.

Begitu pula dengan rata-rata mobil yang dijual diatas mmr yaitu \$12,775 jika dibandingkan dengan yang sebelumnya yaitu \$14,816. Kita bisa merekomendasikan bahwa harga mobil bisa dijual \$2,041 (13.7%) lebih rendah agar harga tidak terlalu terlampau mahal.

2. RECOMMENDATION SYSTEM

Membuat sistem rekomendasi yang bisa memberikan rekomendasi mobil kepada customer sehingga lebih personal dan tepat sasaran. Hal ini juga dimaksudkan agar kombinasi tepat sasaran dan pricing seller yang sesuai bisa mendulang tingkat penjualan.

3. QUALITY INVENTORY

Berdasarkan insight yang didapat pada kondisi dan odometer mobil, maka kita bisa melakukan Quality Control (QC) untuk screening mobil yang akan dijual di platform, misalnya dengan kualifikasi kondisi mobil diatas 2 atau odometer yang tidak terlalu besar. Hal tersebut bertujuan untuk meningkatkan kepercayaan customer dalam membeli mobil bekas yang kualitasnya tetap bagus.

PEMBAGIAN TUGAS

Peran dan Tanggungjawab	
1.Setiap Stage dikerjakan oleh keseluruhan anggota kelompok 2.Setiap Stage terdiri dari 2 tim, yaitu Technical Team dan Theoretical Team 3.Ketua tim (Dwi Ayu Nouvalina/Yuna) sebagai check point reviewer untuk setiap topik pada Stage	
Technical	Theoretical
Bertugas melakukan coding	Bertugas mengumpulkan link referensi dan merangkum materi tertentu untuk mempermudah tim technical dan anggota lain dalam memahami materi
Mencicil Laporan Final Project sesuai stage dan bagiannya masing-masing	Membuat laporan jawaban Homework dan Notulensi mentoring sesuai stage masing-masing

Stage	Topik	PJ	
		Technical	Theoretical
0	Menentukan permasalahan, penyelesaian, peran anggota tim, goals dan objectives, dan business metrics	Semua anggota kelompok	
1	a. Data Exploration	Donny, Yuna	Danta, Yuna
	b. EDA : Univariate Analysis	Nanda, Yuna	Ghulam, Yuna
	Multivariate Analysis	Aby, Yuna	
	c. Insights and Visualization	Irham, Yuna	Alif, Yuna
2	a. Data Cleansing	Alif, Nanda, Yuna	Danta, Yuna
	b. Data Pre-processing	Aby, Donny, Ghulam, Yuna	Irham, Yuna
3	a. Modelling	Aby, Ghulam, Yuna, Irham	Nanda, Yuna
	b. Evaluation	Danta, Alif, Yuna	Donny, Yuna
4	Slide Presentasi Final Project	Yuna	

Thank You