

KASDD C

USED CAR SALES (UCS)

Kasbobby

Our Kasbobby



Aggyta Yosafat



M. Fatih Zain



Rif'at Fauzan



Rafi Irsyad S.

Overview

1

**Data
Description**

2

**Data
Preprocessing**

3

Exploration

4

Modeling

DATA DESCRIPTION

The background features several abstract, translucent 3D geometric shapes in shades of orange, purple, and blue, scattered around the text. These shapes include cubes, pyramids, and rectangular prisms, some of which overlap or intersect with the letters of the word 'DATA'.

Data Shape & Structure

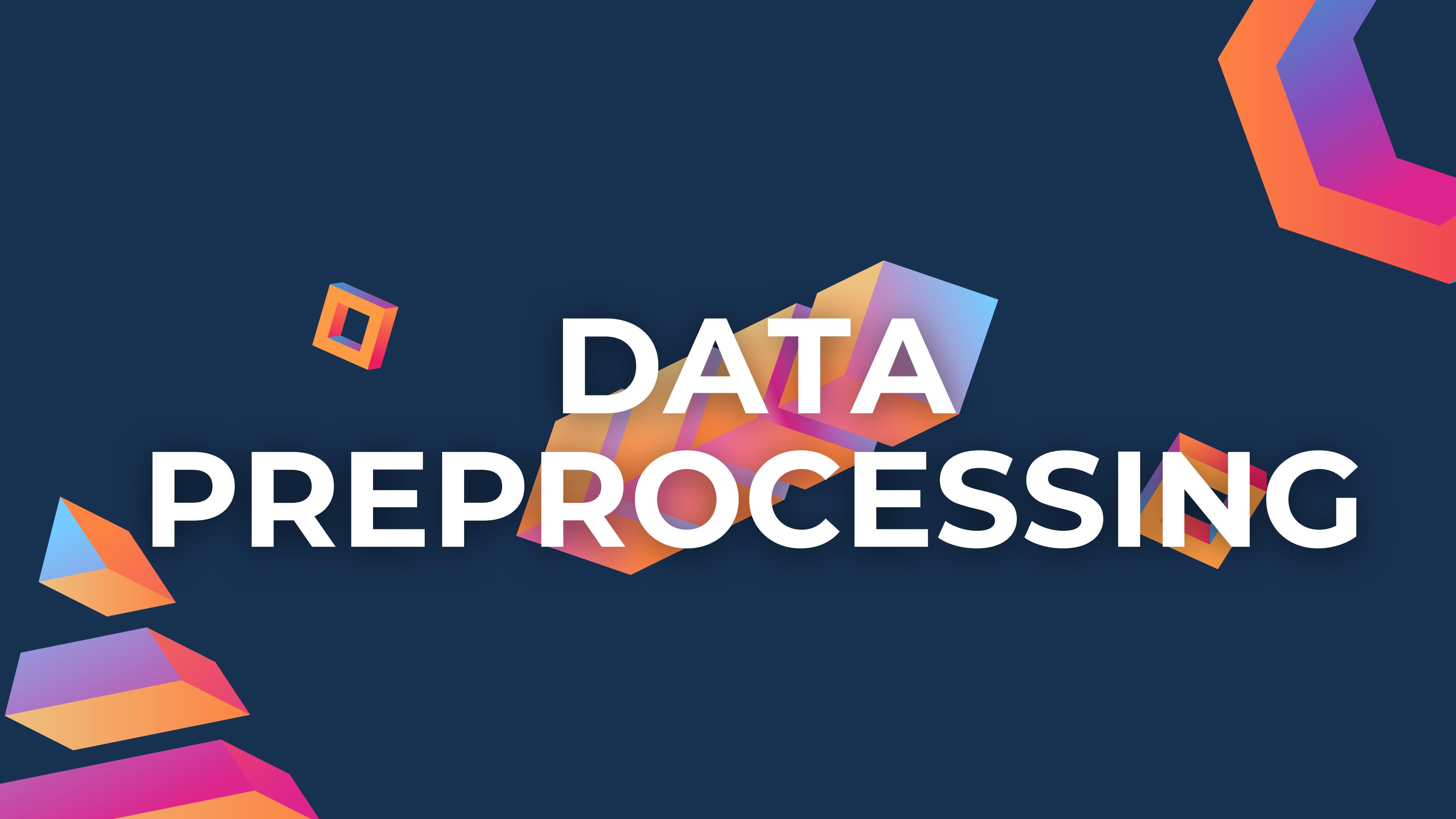
8000
BARIS

24
KOLOM

45.833%
DATA NUMERIK



DATA PREPROCESSING



Data Duplication

Duplicates

0

Missing Values

Missing
Values

0

Tidak Menangani Outliers

WHY?

Terdapat Outlier pada kolom berikut:
**Purchased Price-\$, Sold Price-\$, Sales Commission-\$,
Mileage-KM, dan Feedback**

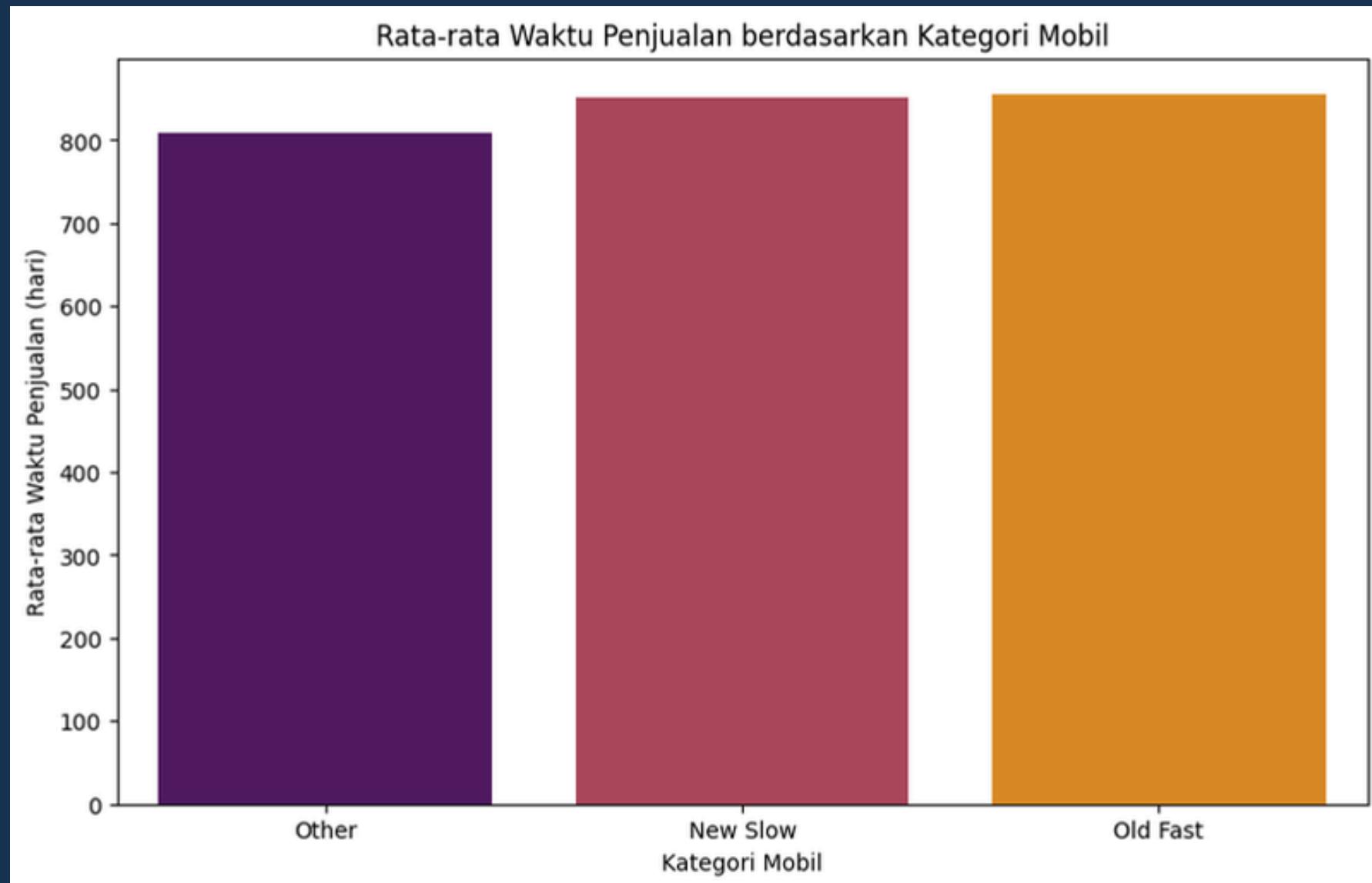
Tidak ditangani karena:

- **Nilai harga beli dan jual yang tinggi atau rendah dapat mencerminkan jenis mobil tertentu** (misalnya, mobil mewah atau murah) yang memang wajar terjadi di pasar.
- **Komisi penjualan dan Mileage-KM pengguna yang ekstrem bisa menjadi sinyal penting dalam menentukan kepuasan pelanggan.** Di beberapa model lebih memungkinkan untuk langsung drop saja.
- Selain itu, **feedback adalah target prediksi.**



EXPLORATION

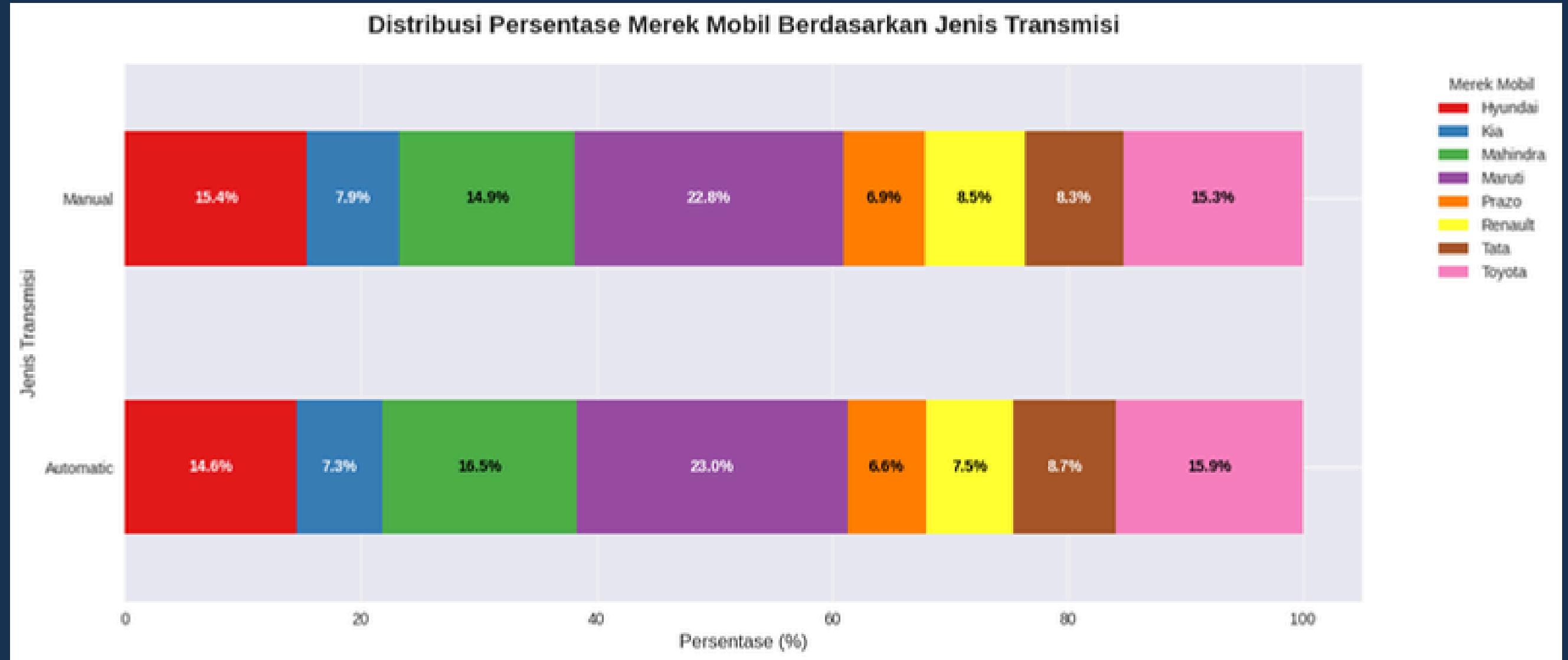
Apakah mobil yang tua tapi kencang **lebih cepat laku** dibanding mobil baru tapi tidak terlalu kencang?



Dapat disimpulkan bahwa mobil tua tapi kencang (Old Fast) **tidak lebih cepat laku** dibandingkan dengan mobil baru tapi tidak kencang (New Slow).

Mobil tua tapi kencang memiliki **rerata waktu yang lebih lama untuk terjual** dibandingkan dengan mobil baru yang tidak terlalu kencang.

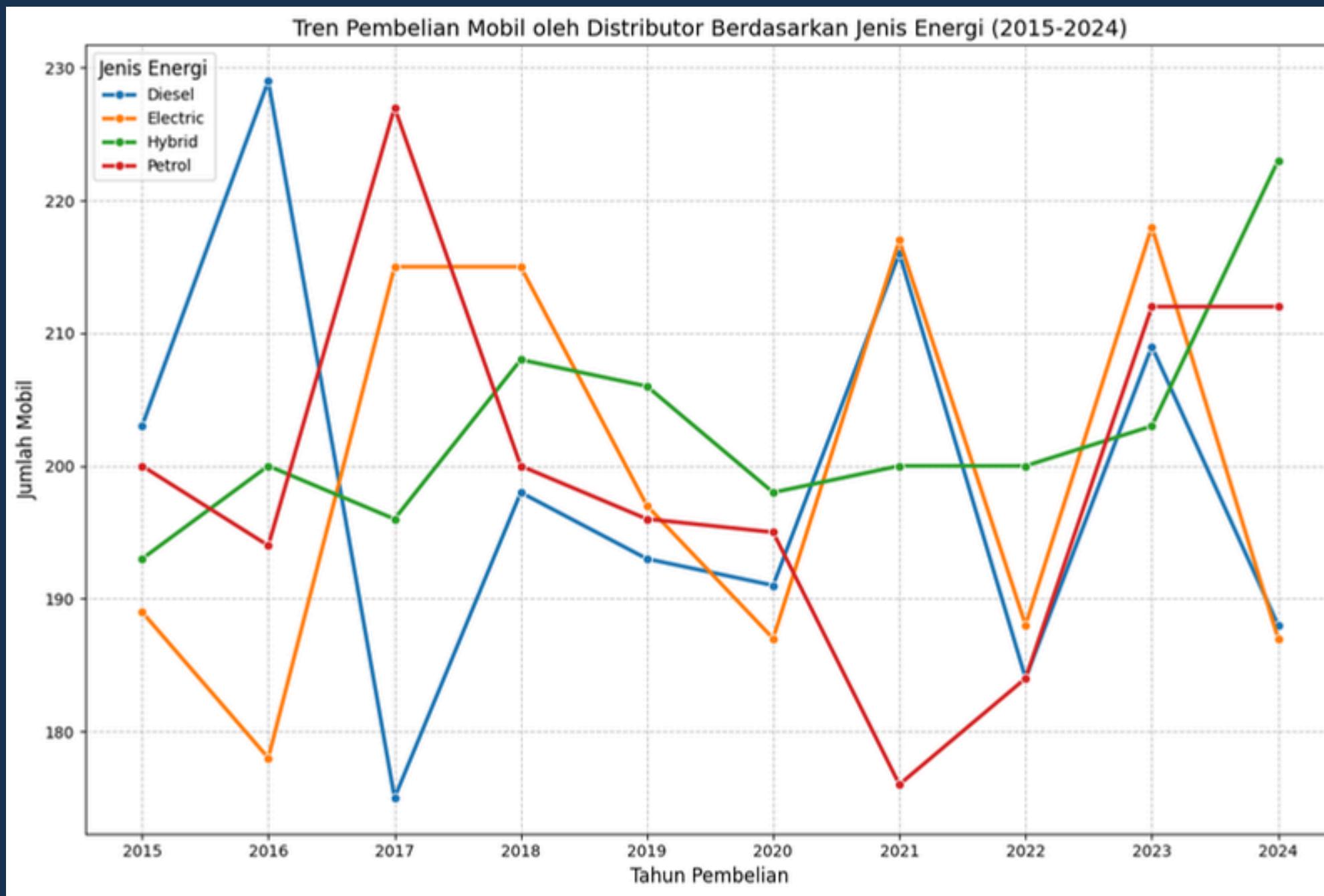
Bagaimana persentase merek mobil (Manufacturer Name) yang dibeli oleh seluruh distributor per transmisi?



Dari representasi visual terlihat bahwa merek mobil memiliki persentase jenis transmisi yang **hampir setara antara Manual dan Automatic**.

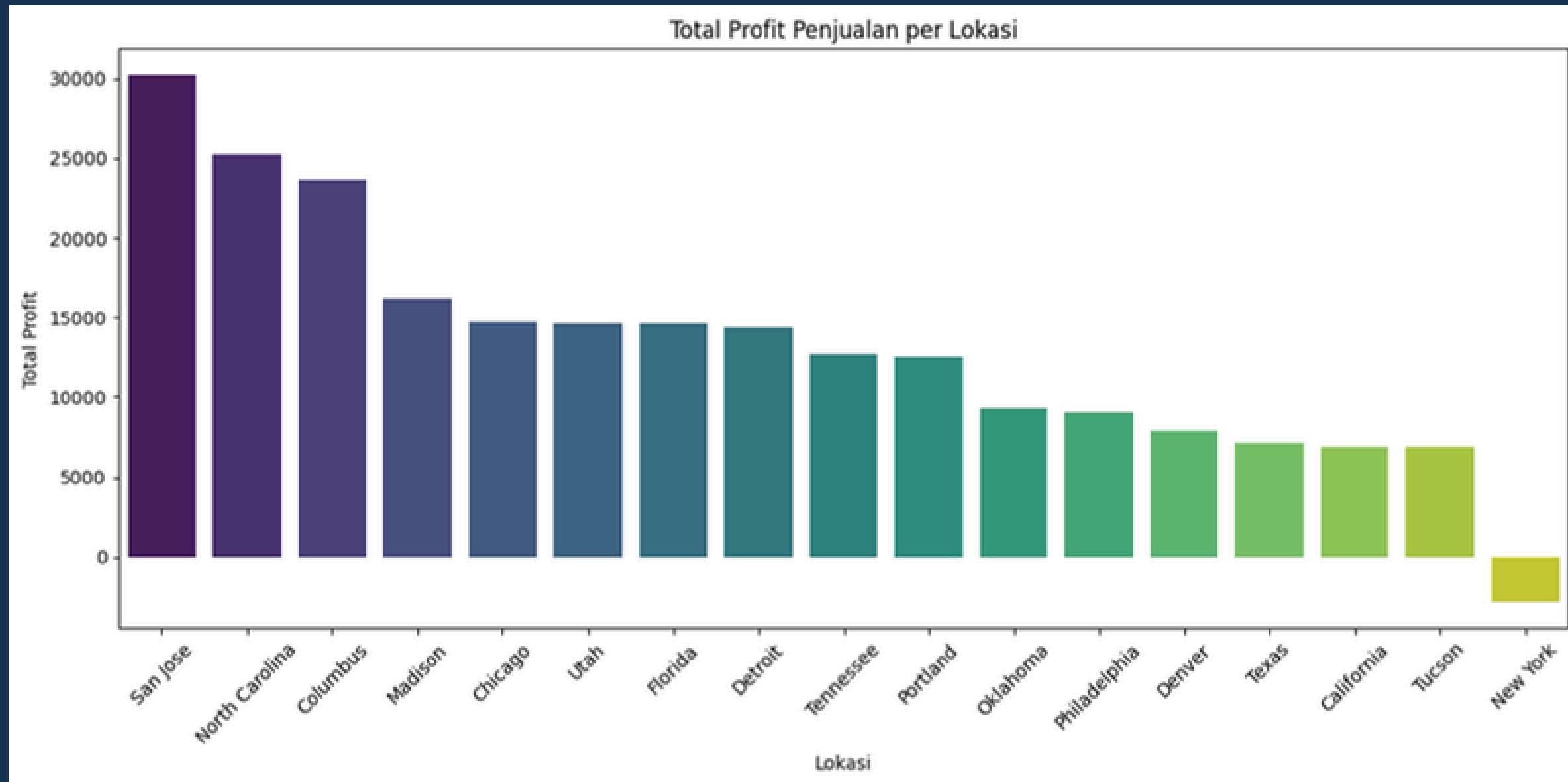
| Merek | Manual | Automatic |
|----------|--------|-----------|
| Hyundai | 15.4% | 14.6% |
| Kia | 7.9% | 7.3% |
| Mahindra | 14.9% | 16.5% |
| Maruti | 22.8% | 23.0% |
| Prazo | 6.9% | 6.6% |
| Renault | 8.5% | 7.5% |
| Tata | 8.3% | 8.7% |
| Toyota | 15.3% | 15.9% |

Bagaimana tren pembelian mobil oleh distributor per jenis energi yang digunakan per tahun?



| Year | Energy | Count |
|------|----------|-------|
| 2015 | Diesel | 203 |
| 2016 | Diesel | 229 |
| 2017 | Petrol | 227 |
| 2018 | Electric | 215 |
| 2019 | Hybrid | 206 |
| 2020 | Hybrid | 198 |
| 2021 | Electric | 217 |
| 2022 | Hybrid | 200 |
| 2023 | Electric | 218 |
| 2024 | Hybrid | 223 |

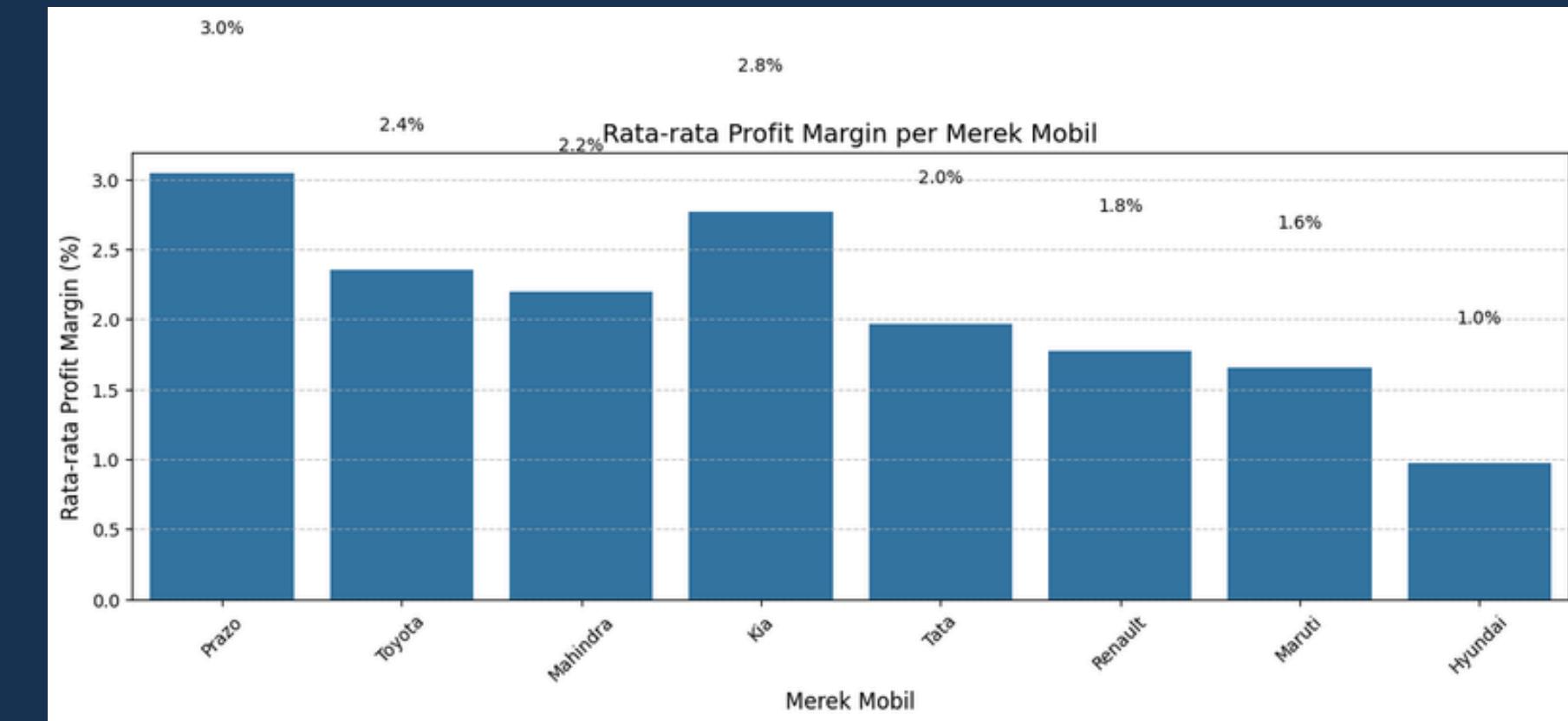
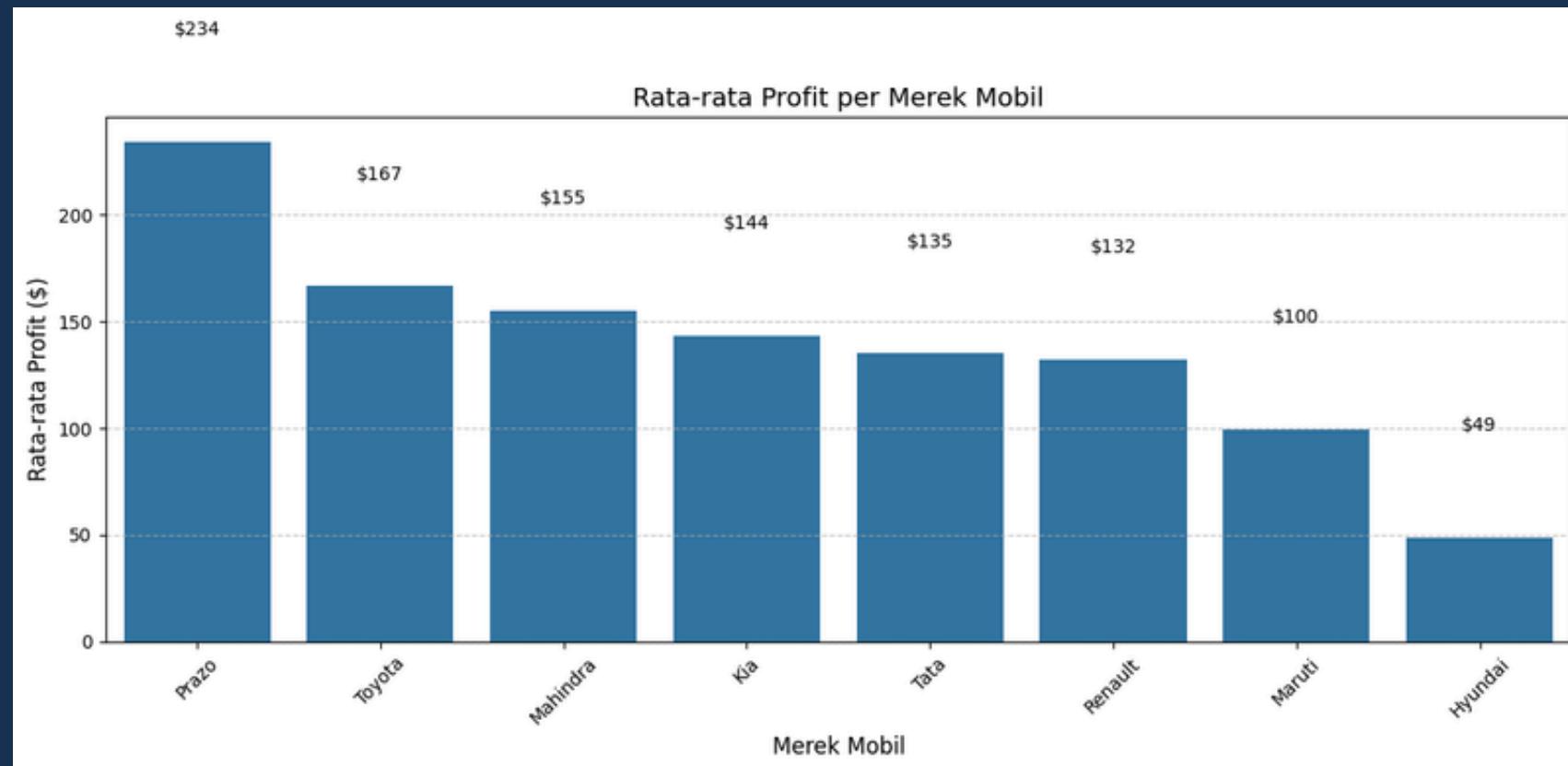
Bagaimana profit penjualan distributor di setiap lokasi? Lokasi apa yang paling menguntungkan?



Dari representasi visual didapatkan informasi **bahwa lokasi San Jose merupakan distributor yang paling menguntungkan.**

Selain itu dapat dilihat pada bar chart **total profit ditampilkan dalam \$.** Dapat dilihat juga **New York mengalami kerugian.**

Merk mobil apa yang memberikan profit tertinggi?

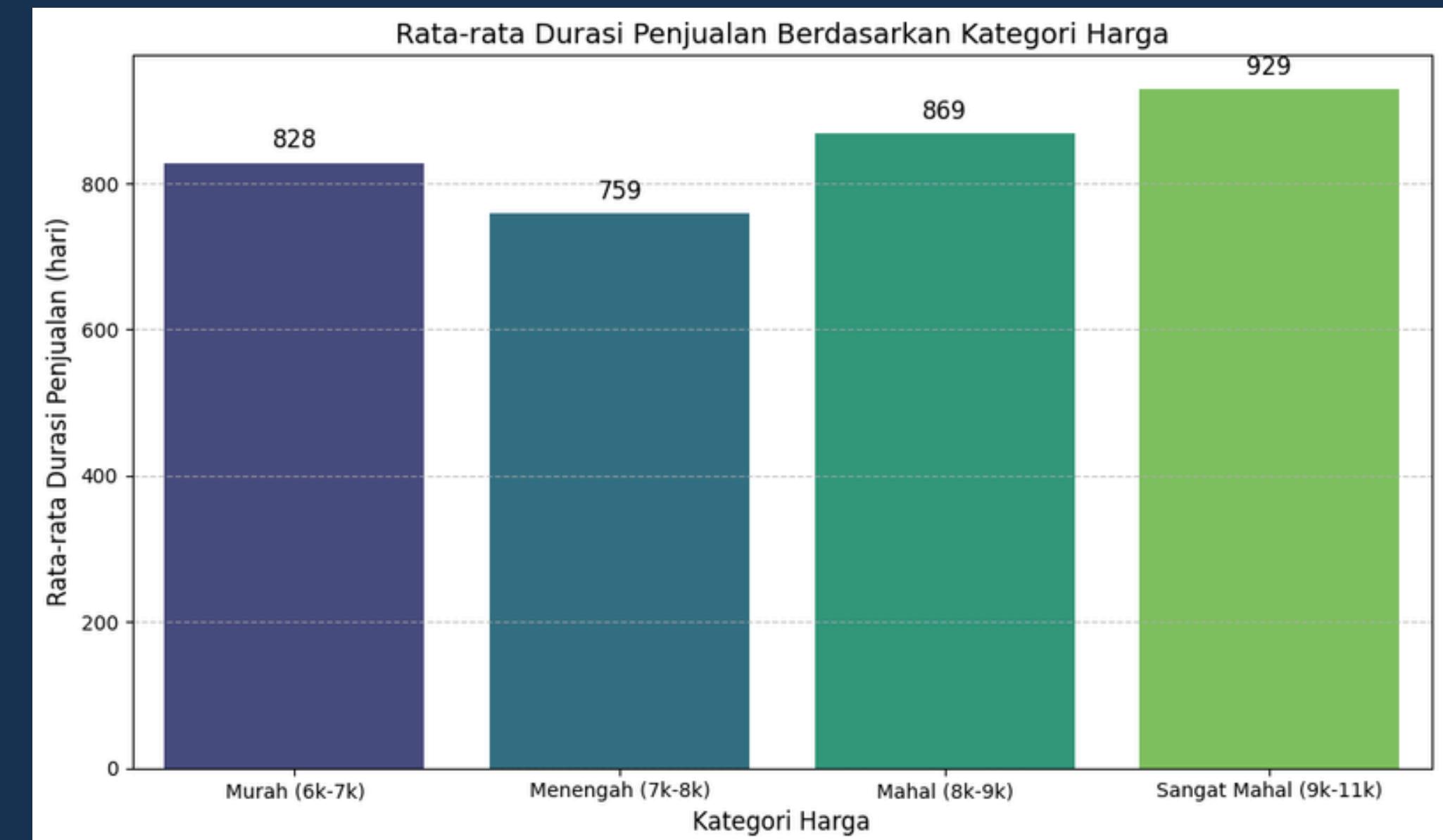


Merk Prazo memiliki profit tertinggi, diikuti Toyora dan Mahindra. Merek seperti Toyota dan Mahindra memberikan profit tinggi tetapi tidak dengan margin terbaik.

Merek dengan volume penjualan tinggi dapat memberikan kontribusi total profit besar meskipun margin per unit lebih rendah

Apakah harga mobil berpengaruh terhadap waktu penjualan?

Mobil kategori menengah terjual paling cepat dengan rata-rata 759 hari, sementara mobil sangat mahal membutuhkan waktu terlama dengan rata-rata 929 hari. Mobil kategori murah justru memiliki durasi penjualan lebih lama yaitu 828 hari dibandingkan kategori menengah, menunjukkan bahwa **faktor selain harga juga berpengaruh terhadap kecepatan penjualan.**





The background features a dark navy blue gradient. Several 3D cubes are scattered across the scene, some floating and others appearing to be part of a larger structure. These cubes are rendered in a translucent style with a color palette of orange, pink, purple, and blue. In the center, the word "MODELING" is written in a large, bold, white sans-serif font. The letters are partially obscured by the overlapping 3D cubes, creating a sense of depth and integration.

MODELING

KLASIFIKASI



Model Klasifikasi Preprocessing

| | |
|----------------------------|--|
| Feature Engineering | <ul style="list-style-type: none">• Membuat fitur ‘Car_Age’• Membuat fitur ‘Price_Difference’ |
| Encoding | Menggunakan LabelEncoding |
| Handle Imbalance | Synthetic Minority Oversampling Technique (SMOTE) |
| Data Splitting | Rasio 80:20 |

Model Klasifikasi

Hyperparameter Tuning

Decision Tree

```
DecisionTreeClassifier
```

```
DecisionTreeClassifier(class_weight='balanced', criterion='entropy',
random_state=42)
```

Random Forest

```
RandomForestClassifier
```

```
RandomForestClassifier(bootstrap=False, class_weight='balanced', max_depth=10,
min_samples_leaf=4, min_samples_split=10,
random_state=42)
```

Naive Bayes

```
GaussianNB
```

```
GaussianNB(var_smoothing=np.float64(0.012915496650148827))
```

KNN

```
KNeighborsClassifier
```

```
KNeighborsClassifier(metric='manhattan', n_neighbors=3)
```

Model Klasifikasi

F1 Score

Decision
Tree

0.54524

Random
Forest

0.51572

Naive
Bayes

0.52492

KNN

0.55288

REGRESI



Model Regresi Preprocessing

| | |
|---------------------|--|
| Feature Engineering | Membuat fitur ‘Car_Age’ |
| Pipeline | <ul style="list-style-type: none">• Menggunakan LabelEncoding• Menggunakan StandardScaler |
| Data Splitting | Rasio 80:20 |

Model Regresi Hyperparameter Tuning

Random Forest

```
RandomForestRegressor
RandomForestRegressor(max_depth=15, max_features=None, min_samples_leaf=4,
                     min_samples_split=5, n_estimators=200, n_jobs=-1,
                     random_state=42)
```

XGBoost

```
XGBRegressor
XGBRegressor(base_score=None, booster=None, callbacks=None,
             colsample_bylevel=None, colsample_bynode=None,
             colsample_bytree=0.9, device=None, early_stopping_rounds=50,
             enable_categorical=False, eval_metric='rmse', feature_types=None,
             feature_weights=None, gamma=None, grow_policy=None,
             importance_type=None, interaction_constraints=None,
             learning_rate=0.1, max_bin=None, max_cat_threshold=None,
             max_cat_to_onehot=None, max_delta_step=None, max_depth=5,
             max_leaves=None, min_child_weight=None, missing=nan,
             monotone_constraints=None, multi_strategy=None, n_estimators=100,
             n_jobs=None, num_parallel_tree=None, ...)
```

Gradient Boosting

```
best_estimator_: GradientBoostingRegressor
GradientBoostingRegressor(learning_rate=0.15, n_estimators=200, random_state=42)
GradientBoostingRegressor
GradientBoostingRegressor(learning_rate=0.15, n_estimators=200, random_state=42)
```

Model Regresi

Linear
Regression

0.86365

Lasso
Regression

0.86365

Ridge
Regression

0.86365

Random
Forest

0.85419

XGBoost

0.86891

Gradient
Boosting

0.87161

CLUSTERING



Model Clustering Preprocessing

Selected Features

Cluster menggunakan fitur :

- 'Price-\$'
- 'Engine Power-HP'
- 'Manufactured Year'
- 'Number of Doors'
- 'Number of Seats'
- 'Car Type'

Pipeline

- Menggunakan MinMaxScaler
- Menggunakan OneHotEncoder

Model Clustering K-Means

Cluster 1 (23%)

- **Car Type : Truck (1304), Convertible (542)**
- Price: Premium
- Number of Seats: Standard capacity
- Number of Doors: Less doors
- Engine Power: High-performance
- Manufactured Year: New

Cluster 2 (31%)

- **Car Type : Hatchback (2478)**
- Price: Budget
- Number of Seats: Standard capacity
- Number of Doors: Standard doors
- Engine Power: Standard-performance
- Manufactured Year: Older

Cluster 3 (31.5%)

- **Car Type : SUV (2519)**
- Price: Premium
- Number of Seats: High capacity
- Number of Doors: Standard doors
- Engine Power: High-performance
- Manufactured Year: Older

Cluster 4 (14.5%)

- **Car Type : Sedan (1157)**
- Price: Mid-range
- Number of Seats: High capacity
- Number of Doors: Standard doors
- Engine Power: Low-performance
- Manufactured Year: New

Model Clustering

Agglomerative Single Link

Cluster 1 (47.8%)

- **Car Type : SUV (2519), Truck (1304)**
- Price: Premium
- Number of Seats: High capacity
- Number of Doors: Less door
- Engine Power: High-performance
- Manufactured Year: Recent

Cluster 2 (6.8%)

- **Car Type : Convertible (542)**
- Price: Premium
- Number of Seats: Standard capacity
- Number of Doors: Less door
- Engine Power: High-performance
- Manufactured Year: Older

Cluster 3 (14.5%)

- **Car Type : Sedan (1157)**
- Price: Mid-range
- Number of Seats: High capacity
- Number of Doors: Standard door
- Engine Power: Low-performance
- Manufactured Year: New

Cluster 4 (31.0%)

- **Car Type : Hatchback (2478)**
- Price: Budget
- Number of Seats: Standard capacity
- Number of Doors: Standard door
- Engine Power: Standard-performance
- Manufactured Year: Older

Model Clustering Agglomerative Complete Link

Cluster 1 (45.4%)

- **Tipe Mobil : Hatchback (2478), Sedan (1157)**
- Price: Budget
- Number of Seats: Standard capacity
- Number of Doors: Standard door
- Engine Power: Standard-performance
- Manufactured Year: New

Cluster 2 (16.3%)

- **Car Type : Truck (1304)**
- Price: Premium
- Number of Seats: Standard capacity
- Number of Doors: Less door
- Engine Power: Standard-performance
- Manufactured Year: New

Cluster 3 (31.5%)

- **Tipe Mobil : SUV (2519)**
- Price: Premium
- Number of Seats: High capacity
- Number of Doors: Standard door
- Engine Power: High-performance
- Manufactured Year: Older

Cluster 4 (6.8%)

- **Tipe Mobil : Convertible (542)**
- Price: Premium
- Number of Seats: Standard capacity
- Number of Doors: Less door
- Engine Power: High-performance
- Manufactured Year: Older

Model Clustering

Agglomerative Average Link

Cluster 1 (23.1%)

- **Car Type : Truck (1304), Convertible (542)**
- Price: Premium
- Number of Seats: Standard capacity
- Number of Doors: Less doors
- Engine Power: High-performance
- Manufactured Year: New

Cluster 2 (14.5%)

- **Car Type : Sedan (1157)**
- Price: Mid-range
- Number of Seats: High capacity
- Number of Doors: Standard doors
- Engine Power: Low-performance
- Manufactured Year: New

Cluster 3 (31.5%)

- **Car Type : SUV (2519)**
- Price: Premium
- Number of Seats: High capacity
- Number of Doors: Standard doors
- Engine Power: High-performance
- Manufactured Year: Older

Cluster 4 (31%)

- **Car Type : Hatchback (2478)**
- Price: Budget
- Number of Seats: Standard capacity
- Number of Doors: Standard doors
- Engine Power: Standard-performance
- Manufactured Year: Older

Cluster Summary

K-Means

| Cluster | Size | Avg_Price | Avg_HP | Avg_Seats | Car_Age | Car_Types | Cluster_Type |
|-----------|------|-----------|--------|-----------|---------|--------------------|----------------------------|
| Cluster 1 | 1846 | 8547.5 | 177.7 | 3.0 | 5.5 | Truck, Convertible | Mobil kerja (Truck) |
| Cluster 2 | 2478 | 7106.0 | 121.1 | 4.5 | 5.5 | Hatchback | Mobil ekonomis (Hatchback) |
| Cluster 3 | 2519 | 8601.7 | 156.3 | 6.3 | 5.6 | SUV | Mobil keluarga (SUV) |
| Cluster 4 | 1157 | 7586.3 | 107.5 | 5.0 | 5.4 | Sedan | Mobil sehari-hari (Sedan) |

Agglomerative Single Link

| Cluster | Size | Avg_Price | Avg_HP | Avg_Seats | Car_Age | Car_Types | Cluster_Type |
|-----------|------|-----------|--------|-----------|---------|-------------|-----------------------------|
| Cluster 1 | 3823 | 8431.1 | 153.3 | 5.3 | 5.5 | SUV, Truck | Mobil keluarga (SUV) |
| Cluster 2 | 542 | 9620.5 | 250.0 | 2.0 | 5.6 | Convertible | Mobil premium (Convertible) |
| Cluster 3 | 1157 | 7586.3 | 107.5 | 5.0 | 5.4 | Sedan | Mobil sehari-hari (Sedan) |
| Cluster 4 | 2478 | 7106.0 | 121.1 | 4.5 | 5.5 | Hatchback | Mobil ekonomis (Hatchback) |

Agglomerative Complete Link

| Cluster | Size | Avg_Price | Avg_HP | Avg_Seats | Car_Age | Car_Types | Cluster_Type |
|-----------|------|-----------|--------|-----------|---------|------------------|-----------------------------|
| Cluster 1 | 3635 | 7258.8 | 116.7 | 4.6 | 5.5 | Hatchback, Sedan | Mobil sehari-hari (Sedan) |
| Cluster 2 | 1304 | 8101.5 | 147.7 | 3.5 | 5.4 | Truck | Mobil kerja (Truck) |
| Cluster 3 | 2519 | 8601.7 | 156.3 | 6.3 | 5.6 | SUV | Mobil keluarga (SUV) |
| Cluster 4 | 542 | 9620.5 | 250.0 | 2.0 | 5.6 | Convertible | Mobil premium (Convertible) |

Agglomerative Average Link

| Cluster | Size | Avg_Price | Avg_HP | Avg_Seats | Car_Age | Car_Types | Cluster_Type |
|-----------|------|-----------|--------|-----------|---------|--------------------|----------------------------|
| Cluster 1 | 1846 | 8547.5 | 177.7 | 3.0 | 5.5 | Truck, Convertible | Mobil kerja (Truck) |
| Cluster 2 | 1157 | 7586.3 | 107.5 | 5.0 | 5.4 | Sedan | Mobil sehari-hari (Sedan) |
| Cluster 3 | 2519 | 8601.7 | 156.3 | 6.3 | 5.6 | SUV | Mobil keluarga (SUV) |
| Cluster 4 | 2478 | 7106.0 | 121.1 | 4.5 | 5.5 | Hatchback | Mobil ekonomis (Hatchback) |

The background features a dark navy blue gradient. On the left side, there's a large, semi-transparent shape composed of overlapping red, purple, and pink triangles. On the right side, there's a large, semi-transparent shape composed of overlapping orange and blue triangles.

THANK
YOU!