

CAR PRICE PREDICTION

6조짜리 자동차



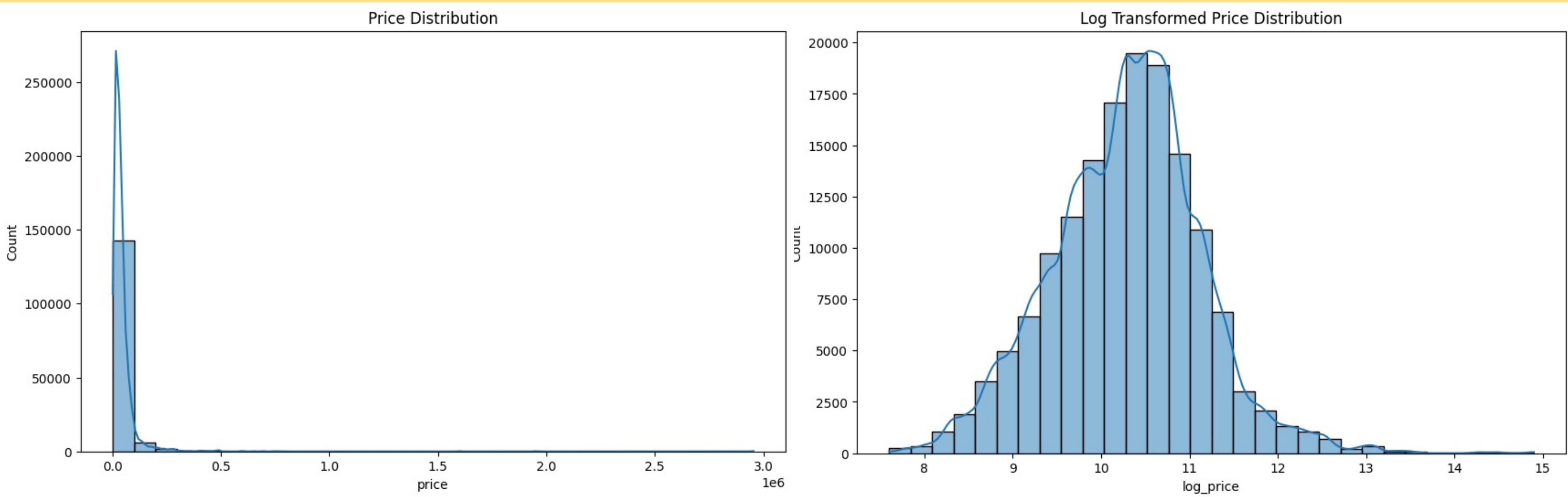
EDA



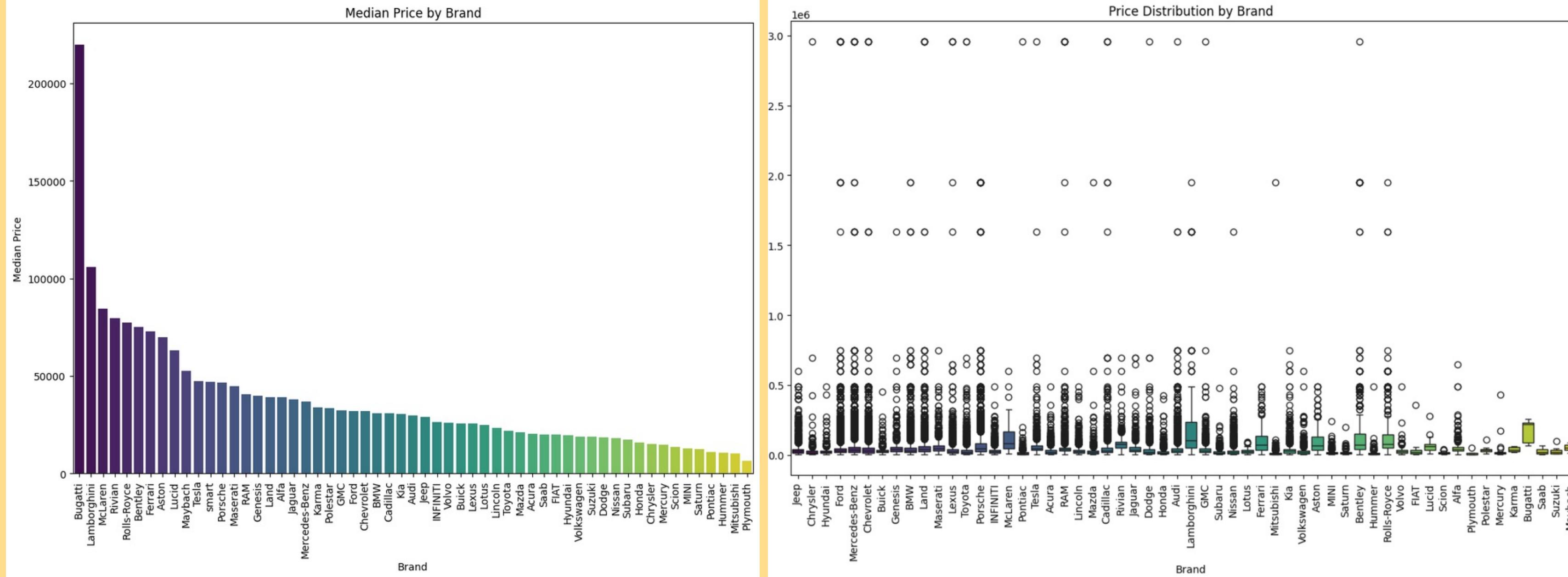
DATA

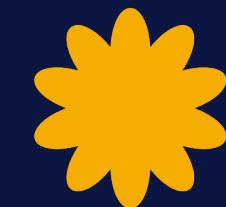
컬럼명	설명
id	매물 고유 아이디
brand	차량의 브랜드
model	차량의 모델명
release year	차량의 제조 연도
milage	차량의 주행 거리
fuel type	차량의 연료 유형
engine	차량의 엔진 타입
transmission	차량의 변속기 유형
exterior color	차량의 외장 색상
interior color	차량의 내장 색상
accident report	차량의 사고 이력
clear title	클린 타이틀 (차량 운행에 중대한 지장을 주는 사고가 없었음을 의미)
price (target)	차량 가격

price distribution



price distribution





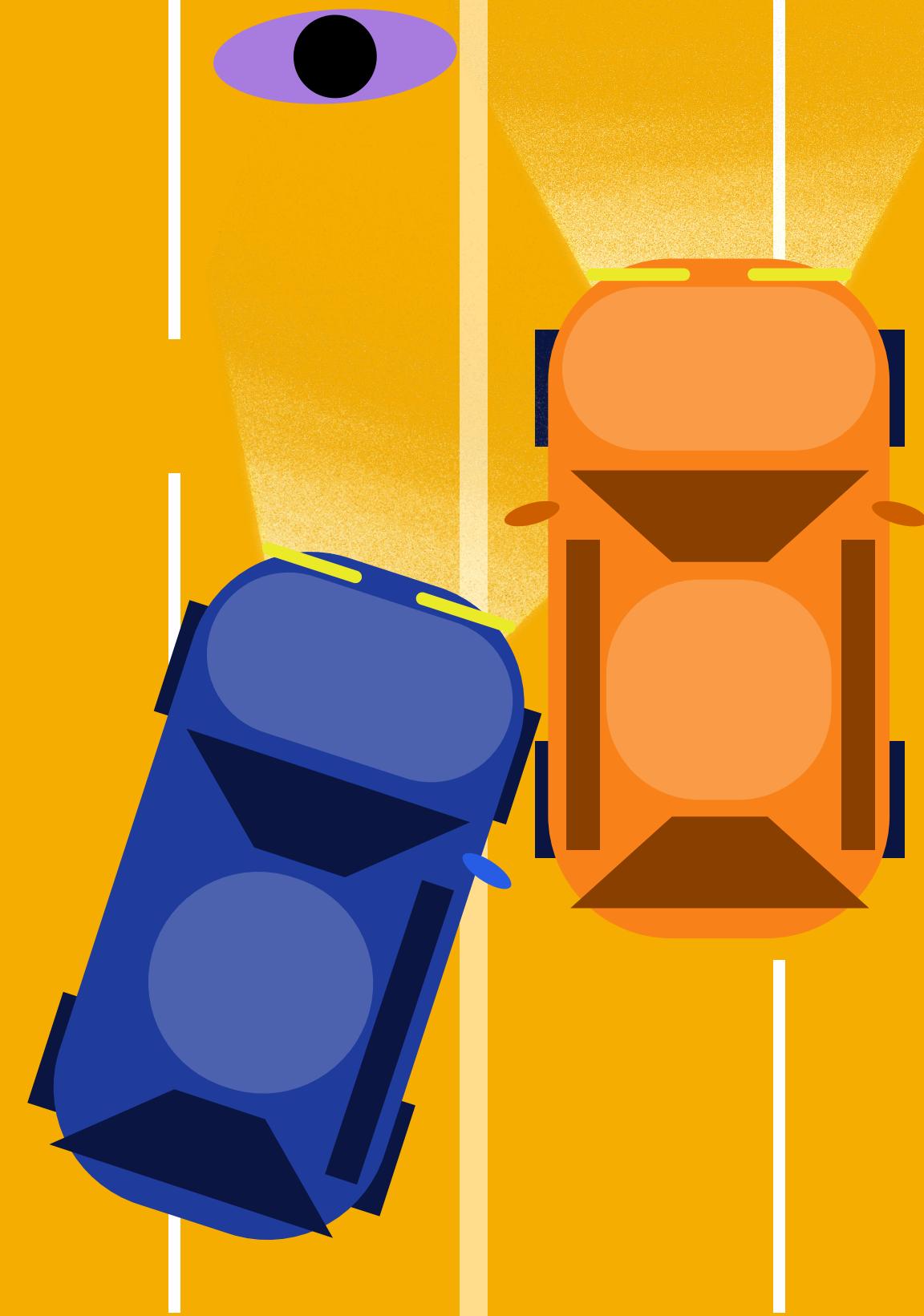
NULL 결측치 처리

fuel_type

accident_report

clean_title

engine



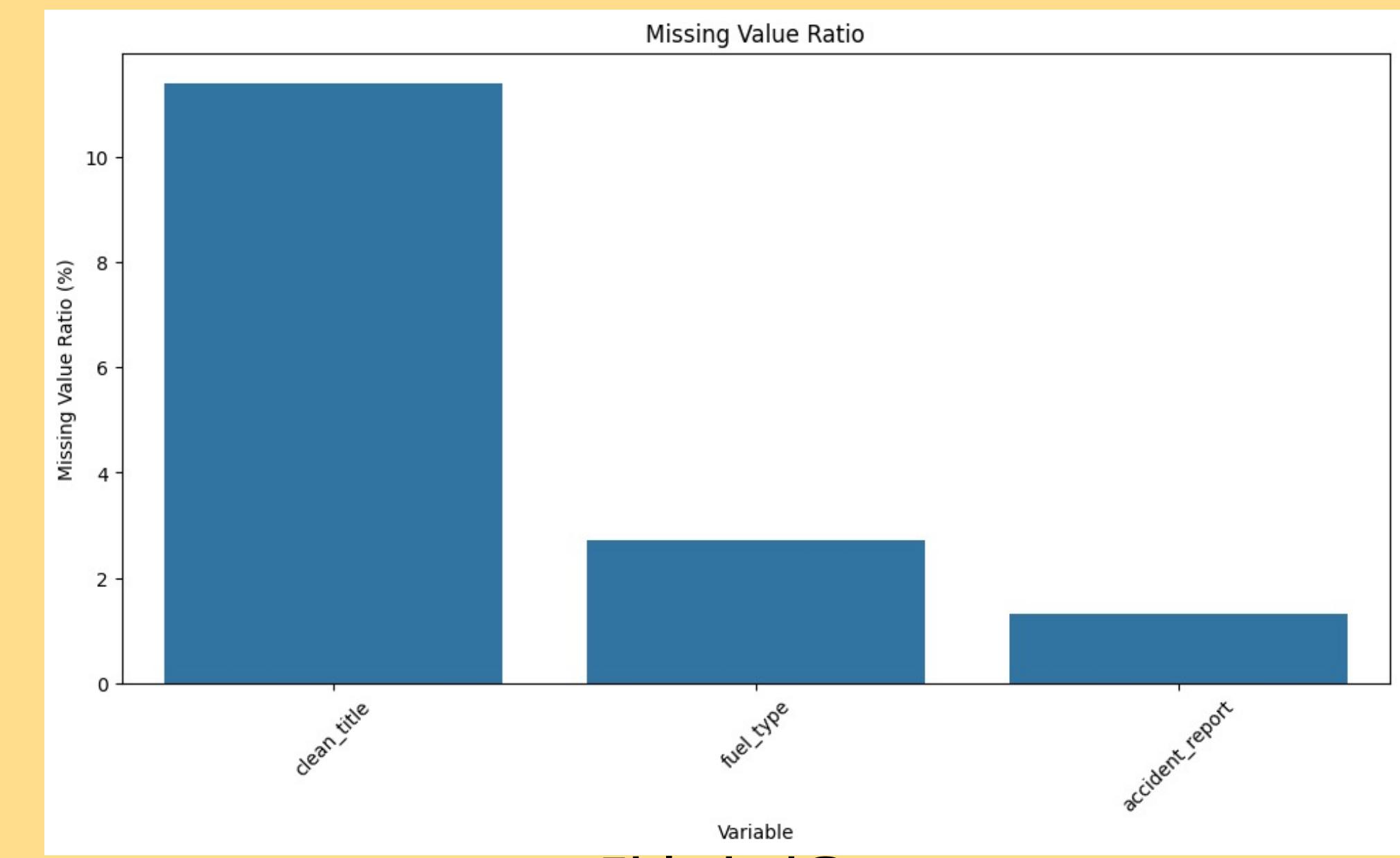
결측치

id	0
brand	0
model	0
release_year	0
milage	0
fuel_type	4085
engine	0

결측치 개수

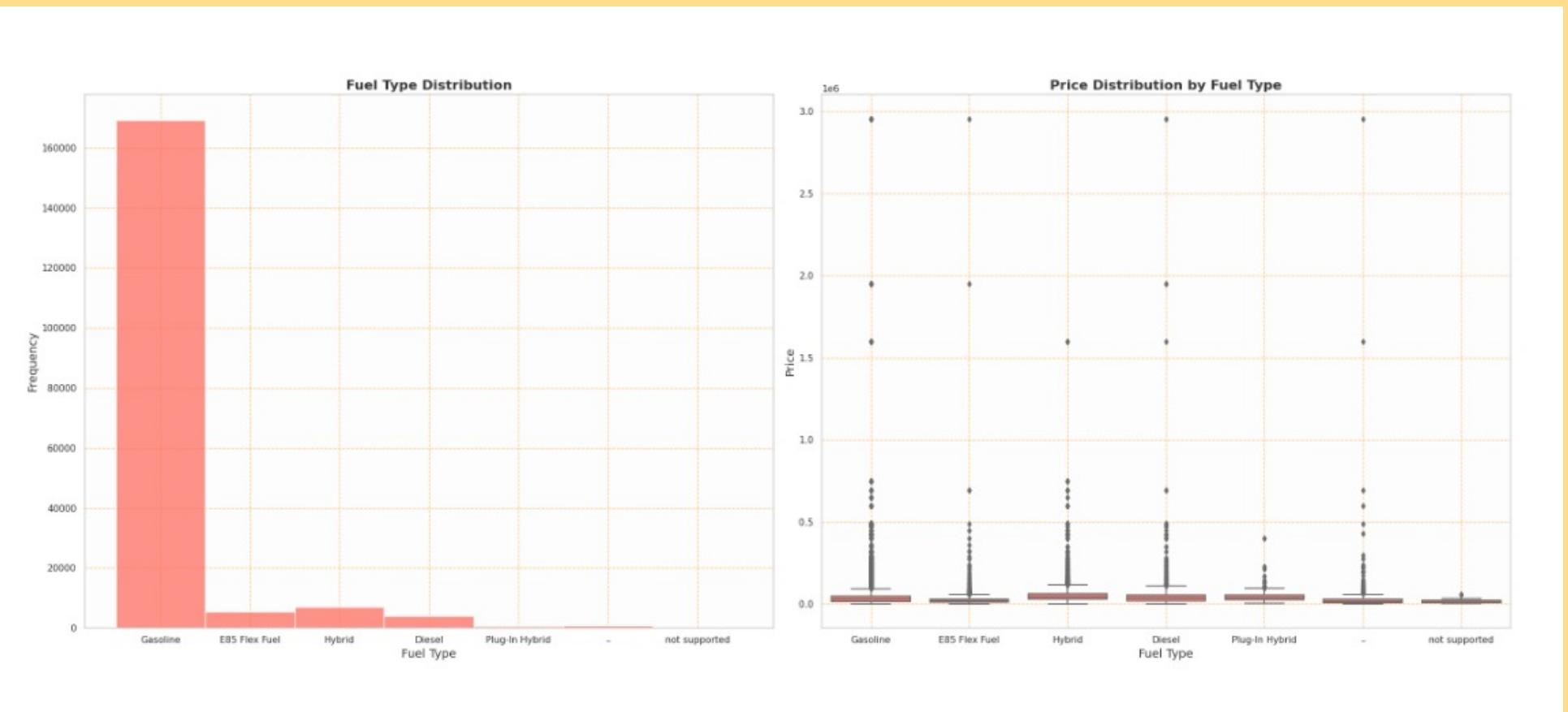
transmission	0
exterior_color	0
interior_color	0
accident_report	1991
clean_title	17205
price	0

결측치 개수



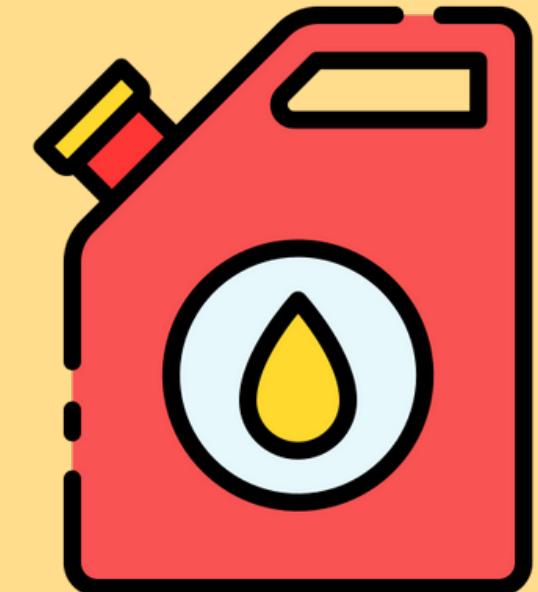
(clean_title 11.407184%/fuel_type 2.708419%/accident_report 1.320064%)

fuel_type



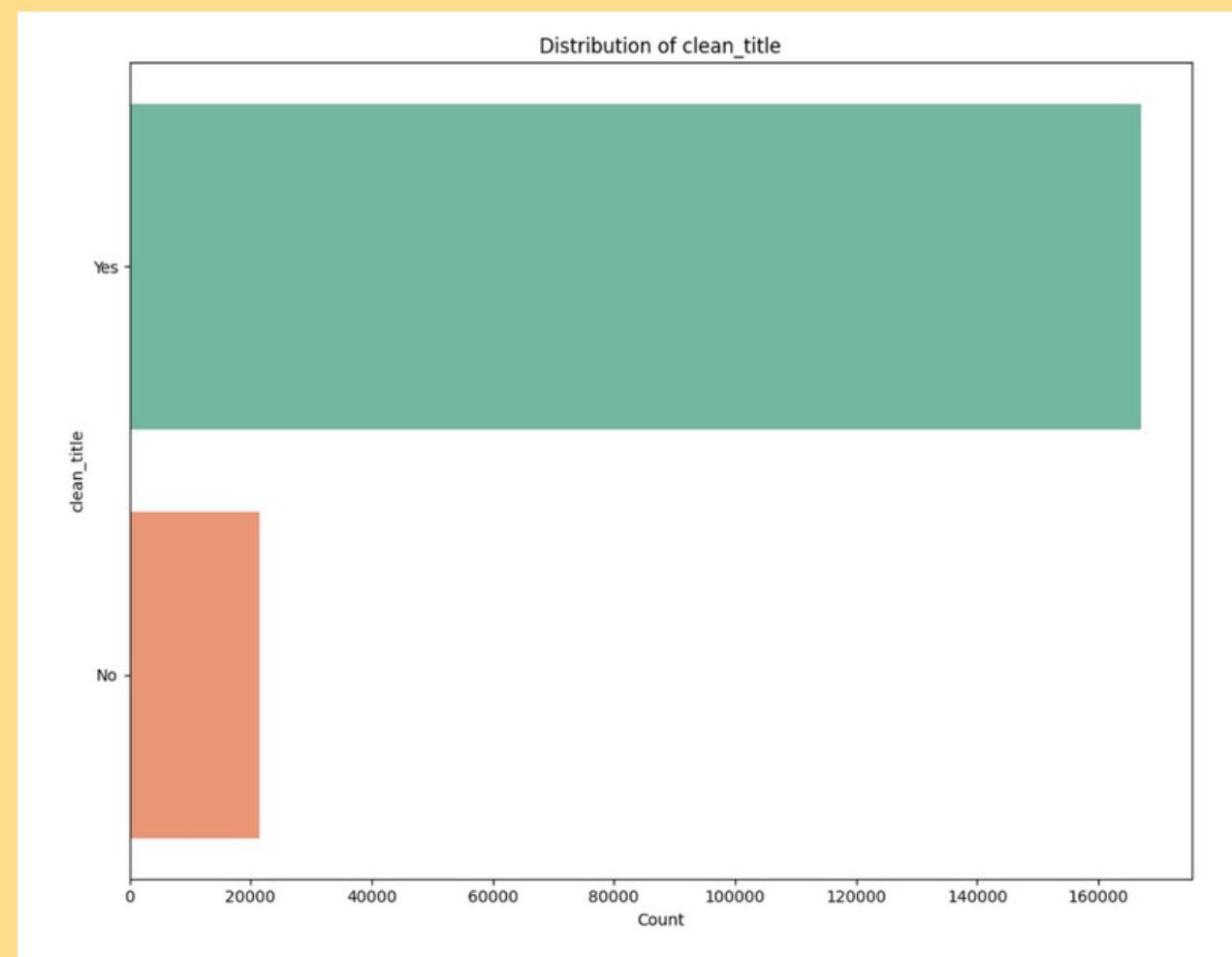
- Not supported - 삭제
- 결측치 내에 Tesla : Electric
- 결측치 x 내에 Tesla : Electric
- 나머지 결측치 : 브랜드 별 최빈값

id	brand	model	release_year	milage	fuel_type	engine	transmission
41	Tesla	Model X Base	2023	6.0	Nan	670.0HP Electric Motor Electric Fuel System	Automatic Transmission
76	Tesla	Model S 75D	2014	119.0	Nan	518.0HP Electric Motor Electric Fuel System	Automatic Transmission
91	Tesla	Model X P100D	2019	67.0	Gasoline	Electric Motor Electric Fuel System	Automatic Transmission
159	Tesla	Model 3 Long Range	2018	36.0	Nan	Electric	Automatic
169	Tesla	Model S P100D	2021	13.0	Nan	518.0HP Electric Motor Electric Fuel System	1-Speed Automatic Transmission



accident_report & clean title

- accident_report : null값 삭제
- accident_report : none => clean title : yes
- accident_report : 1 or more accident => clean title : no



engine

- # Feature 3: Extracting horsepower (HP)
- # Feature 4: Extracting engine displacement (L)
- # Feature 5: Extracting the number of cylinders

조건 후 나온 결측치는 만든 피쳐별 최빈값으로 대체



engine
285.0HP 3.6L V6 Cylinder Engine Gasoline Fuel
250.0HP 3.5L V6 Cylinder Engine Gasoline Fuel
306.0HP 3.8L V6 Cylinder Engine Gasoline Fuel
400.0HP 3.5L V6 Cylinder Engine Gasoline Fuel
382.0HP 5.5L 8 Cylinder Engine Gasoline Fuel
3.2L V6 24V MPFI DOHC
440.0HP 6.7L 8 Cylinder Engine Diesel Fuel
302.0HP 3.5L V6 Cylinder Engine Gasoline Fuel
2.5L I4 16V GDI DOHC
355.0HP 5.3L 8 Cylinder Engine Gasoline Fuel
333.0HP 3.2L Straight 6 Cylinder Engine Gasoli...
380.0HP 3.0L V6 Cylinder Engine Gasoline Fuel
310.0HP 5.4L 8 Cylinder Engine Flex Fuel Capab...
349.0HP 3.0L V6 Cylinder Engine Gasoline Fuel
306.0HP 3.5L V6 Cylinder Engine Gasoline Fuel
453.0HP 3.0L Straight 6 Cylinder Engine Gasoli...

결측치 제거된 결과

```

1 def feature_engineering(df_cleaned):
2
3
4
5     # Feature 3: Extracting horsepower (HP) from the engine column
6     df_cleaned['horsepower'] = df_cleaned['engine'].str.extract(r'(\#\d+\#\#\d+)HP').astype(float)
7
8     # Feature 4: Extracting engine displacement (L) from the engine column
9     df_cleaned['engine_displacement'] = df_cleaned['engine'].str.extract(r'(\#\d+\#\#\d+)L').astype(float)
10
11    # Feature 5: Extracting the number of cylinders from the engine column
12    df_cleaned['number_cylinders'] = df_cleaned['engine'].str.extract(r'(\#\d+ Cylinder)').astype(str)
13    df_cleaned['number_cylinders'] = df_cleaned['number_cylinders'].replace('nan', np.nan)
14
15
16
17    df_cleaned = df_cleaned.drop('engine', axis=1)
18
19    return df_cleaned

```

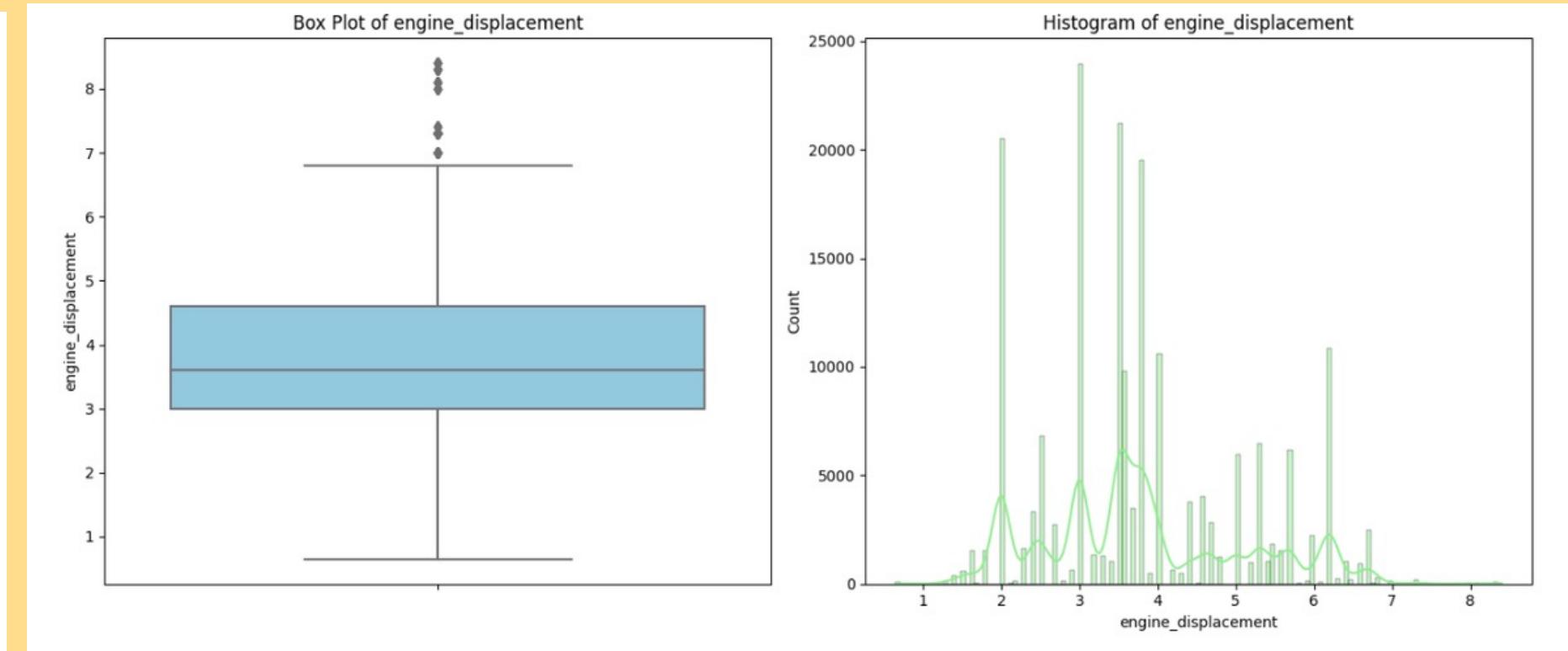
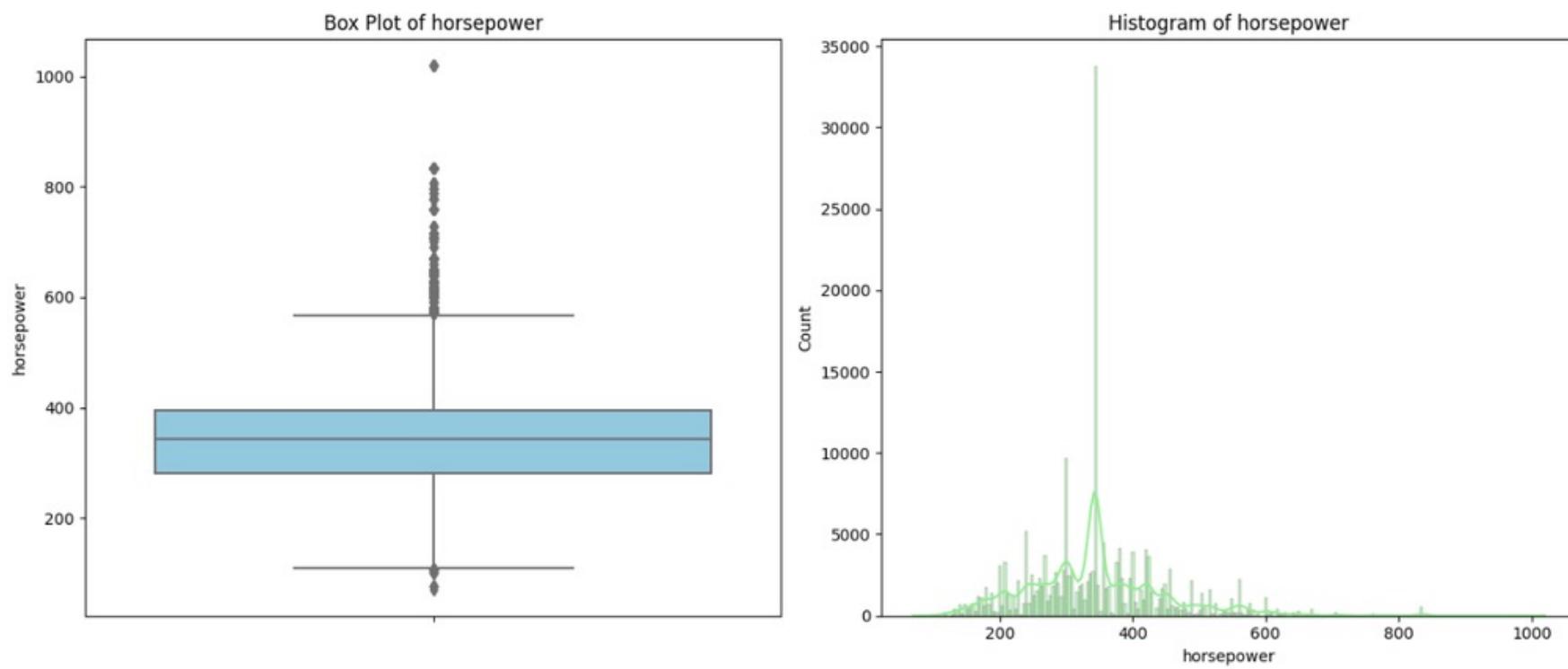
```

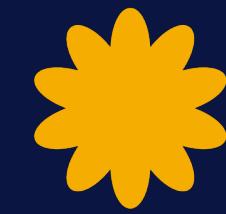
[43] 1 def null_handling(df):
2
3     df['horsepower'] = df['horsepower'].fillna(df['horsepower'].mean())
4     df['engine_displacement'] = df['engine_displacement'].fillna(df['engine_displacement'].mean())
5     df['number_cylinders'] = df['number_cylinders'].fillna(df['number_cylinders'].mode()[0])
6
7     return df

```



	#NULLS	%NULLS	#Unique_Values	Unique_Values/Range
id	0	0.0	148822	150825
brand	0	0.0	57	{Hummer, Hyundai, BMW, McLaren, Aston, Saturn,...
model	0	0.0	1896	{Suburban 1500 LT, Crossfire Limited, Tundra T...
release_year	0	0.0	34	50
milage	0	0.0	264	405.0
fuel_type	0	0.0	7	{-, Gasoline, E85 Flex Fuel, electric, Diesel,...
transmission	0	0.0	52	{6-Speed Automatic with Auto-Shift, 6-Speed, 9...
exterior_color	0	0.0	319	{Sangria Red, Ruby Flare Pearl, Volcanic Orang...
interior_color	0	0.0	156	{Navy Pier, Deep Chestnut, Ebony Black, Sandst...
accident_report	0	0.0	2	{1 or more accident reported, None reported}
clean_title	0	0.0	2	{No, Yes}
price	0	0.0	1569	2952083
horsepower	0	0.0	348	950.0
engine_displacement	0	0.0	62	7.75
number_cylinders	0	0.0	7	{5 Cylinder, 6 Cylinder, 12 Cylinder, 10 Cylin...

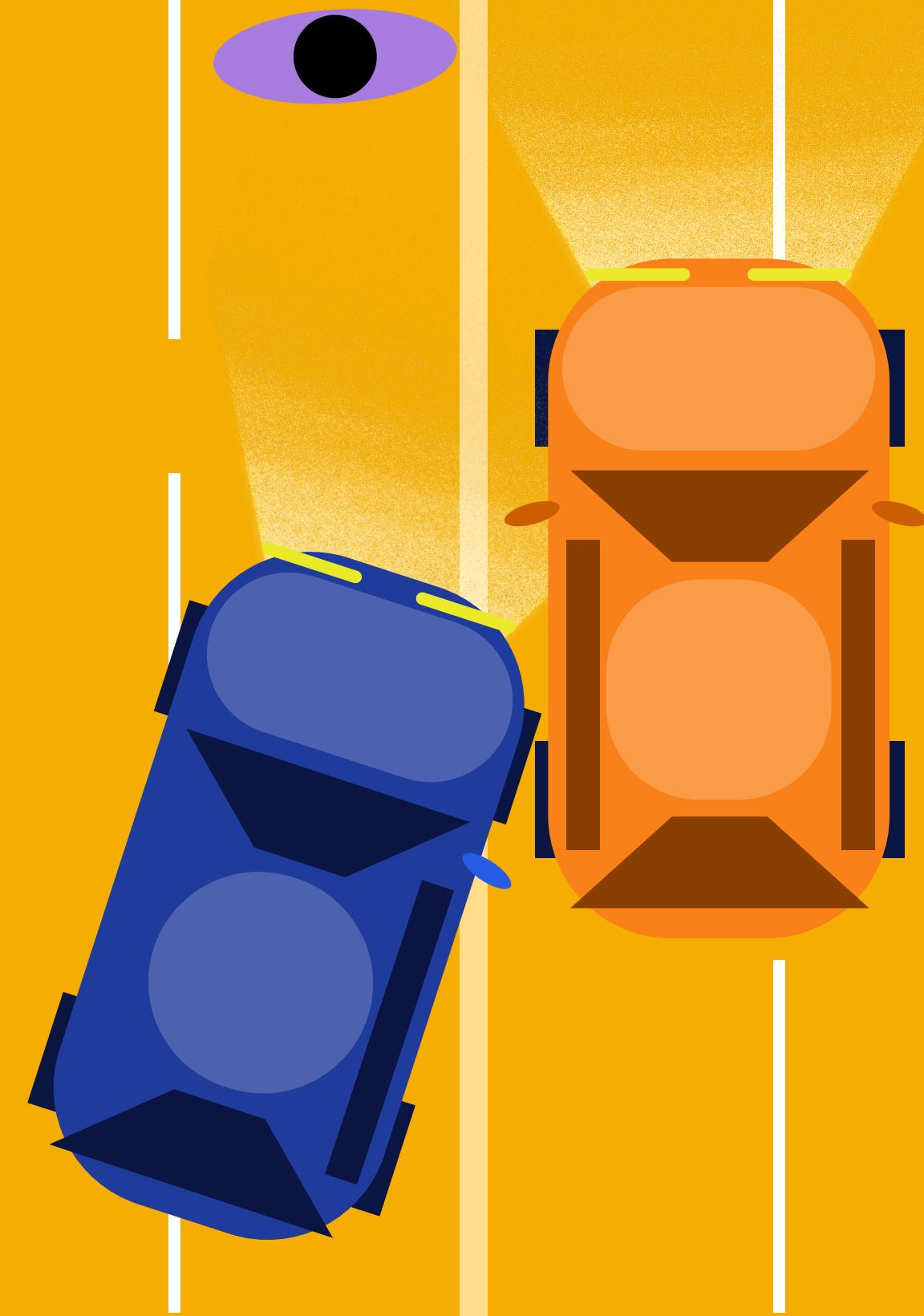




인코딩

transmission

interior_color/exterior_color



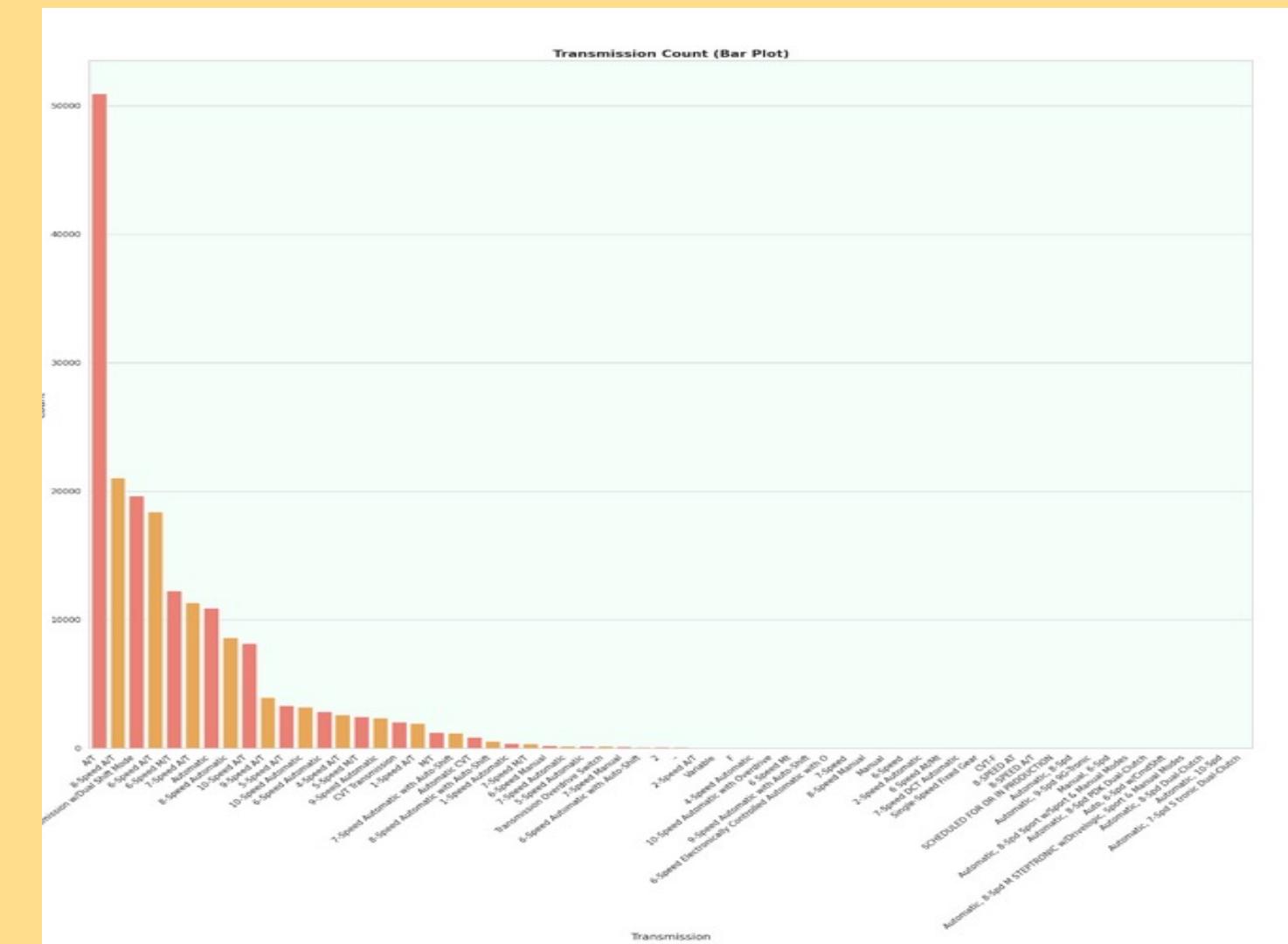
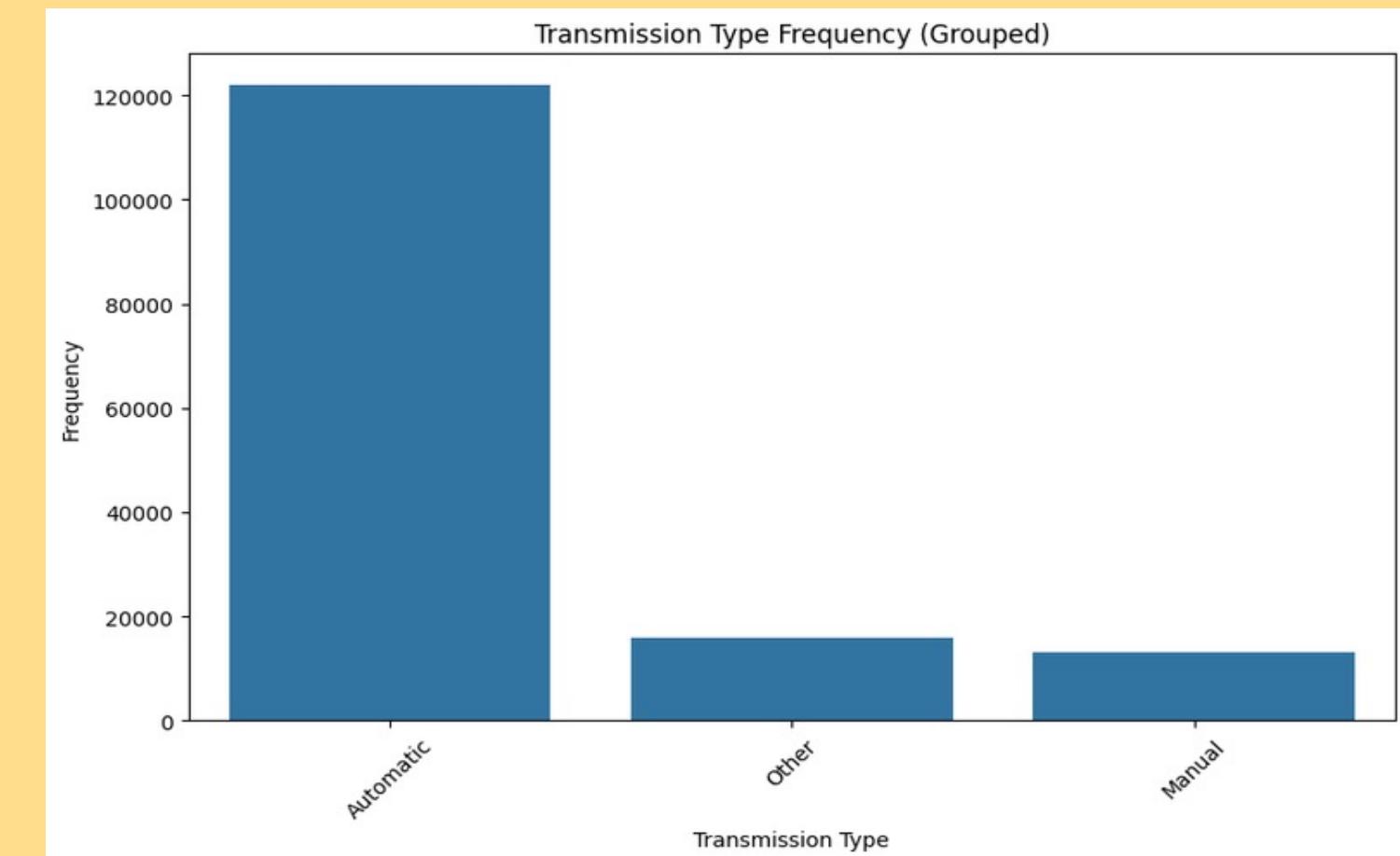
Transmission

- **automatic**
- **manual**
- **other**

Other에 들어간 변속기 종류: ['Transmission with Dual Shift Mode'

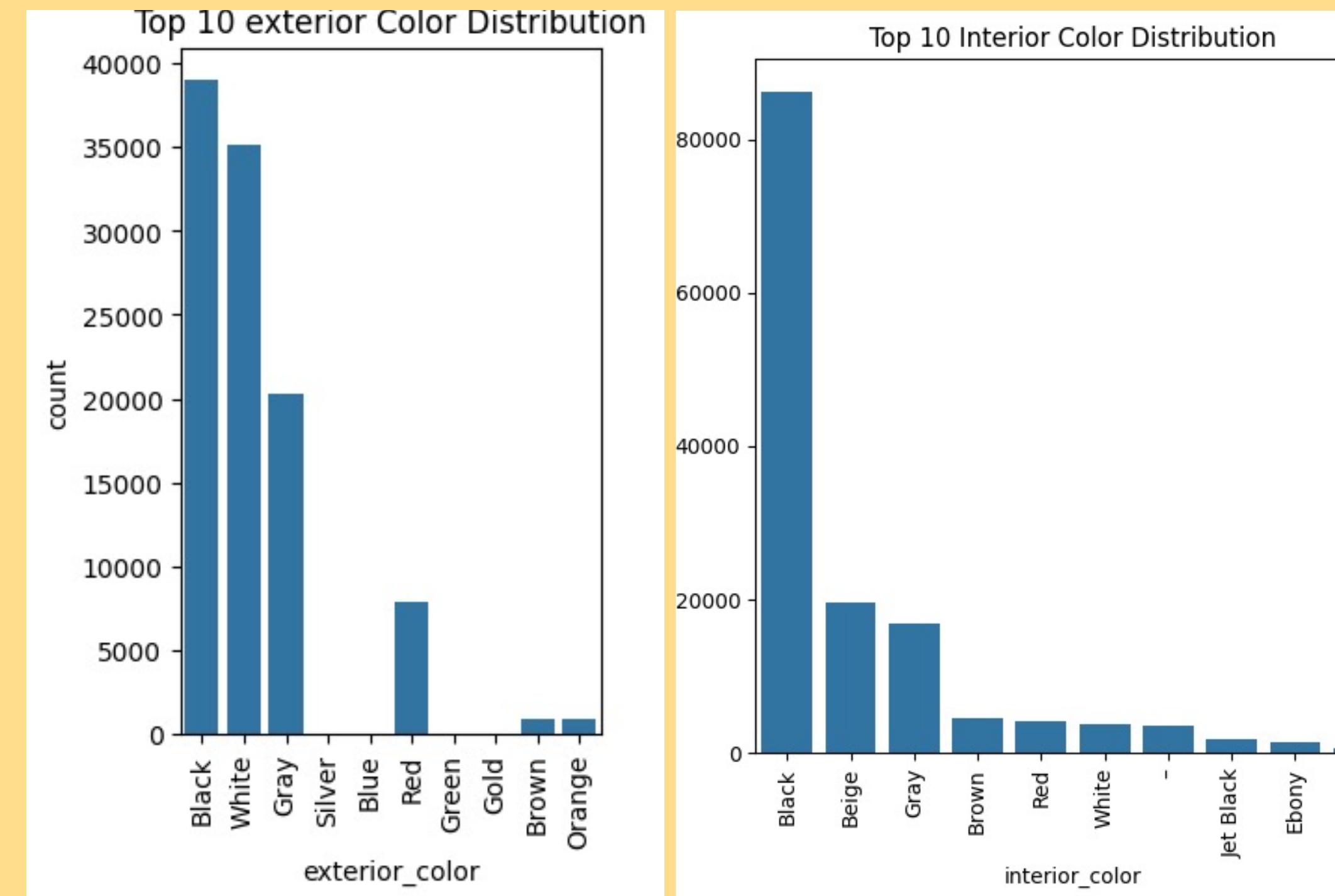
'Transmission Overdrive Switch''7-Speed''F''6-Speed''2''-'Single-

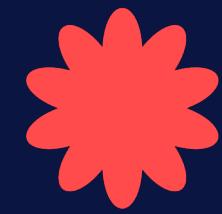
Speed Fixed Gear''SCHEDULED FOR OR IN PRODUCTION']



interior_color/exterior_color

- **exterior : black ,**
- **white, silver or not**
- **interior: black or not**





모델링

CATBOOST

XG-BOOST

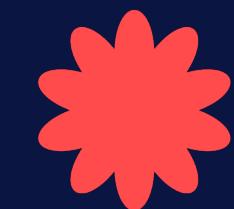


결과

모델	데이터	RMSE
CATBOOST	결측치만 제거한 데이터	70450
XG- BOOST	범주형 데이터 처리 후 (인코딩) 데이터	45360
CATBOOST	범주형 데이터 처리 후 (인코딩) 데이터	41581

FEATURE IMPORTANCE

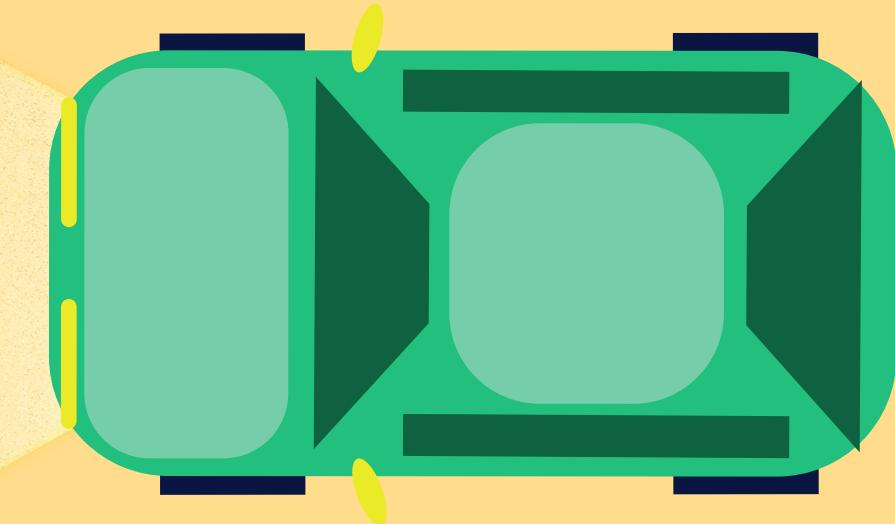
	Feature	Importance
1	model	49.635729
0	brand	21.891580
7	combined_component	7.660893
9	engine_displacement	5.825411
3	milage	4.914982
2	release_year	4.418190
8	horsepower	2.693735
15	exterior_color_BWG	0.942813
14	interior_color_black	0.620038
4	fuel_type	0.602945
10	number_cylinders	0.592758
13	transmission_Other	0.077577
5	accident_report	0.052353
11	transmission_Automatic	0.048543
12	transmission_Manual	0.022287
6	clean_title	0.000166



insight 도출



Consumer Perspective



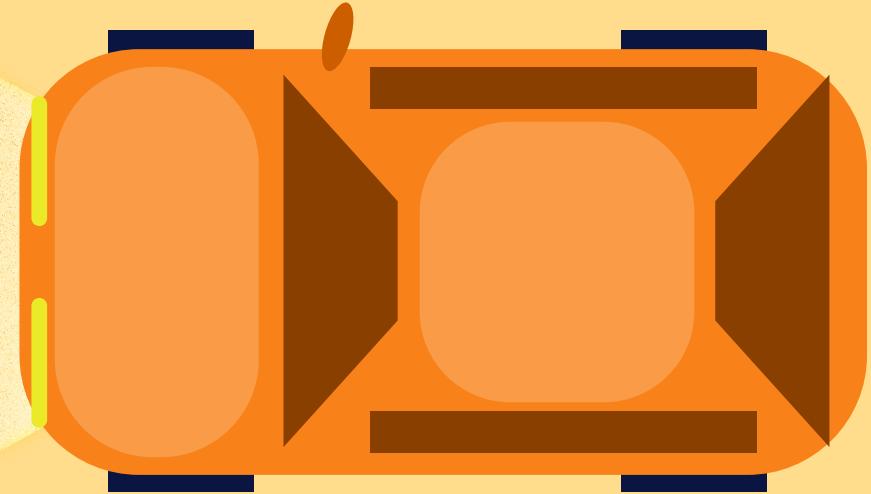
1. 투자 가치가 높은 모델과 브랜드:

- 구매자는 감가상각이 낮은 브랜드와 모델을 선택할 경우, 나중에 되팔 때도 높은 가치를 유지할 가능성이 큼, 프리미엄 브랜드와 인기 모델에 대한 잔존가치 정보를 제공하여, 소비자가 더 나은 투자 가치를 제공하는 차량 선택이 현명.

2. 친환경 차량 및 연료 효율 고려

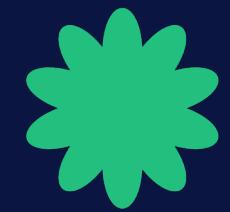
- 연료 유형(fuel_type)과 엔진 효율이 가격형성에 미치는 중요도가 크게 높지 않기에 장기적으로 효율이 좋은 전기차와 하이브리드 차량은 연비가 높고 유지비가 절감 가능성이 높음. 특히, 장기적으로 연료비를 절약하고자 하는 소비자에게 유리할 수 있음.

Seller Perspective



1) 가격 정책

- 모델(model)과 브랜드(brand)에 따른 프리미엄 가격 설정: 인기 모델 및 프리미엄 브랜드 차량의 경우 더 높은 가격을 책정하는 전략을 세우기
- 상대적으로 수요가 낮은 브랜드는 가격을 낮추어 더 빠르게 판매를 유도하거나, 할부와 같은 옵션을 제공해 구매 장벽을 낮출 수 있음



감사합니다

