

# 중고차 시장의 신뢰도 향상을 위한 ML 제안



김민규

# CONTENTS



1

## 분석 배경/목표

- 중고차 시장의 성장
- 투명하지 못한 거래 행태
- 대기업의 독과점 우려

2

## EDA

- 데이터 이해
- 데이터 전처리
- 데이터 EDA

3

## 알고리즘 개발

- 다항회귀 모델의 원리
- 산업의 현재동향
- 산업의 잠재시장 파악

4

## 프로세스 결과

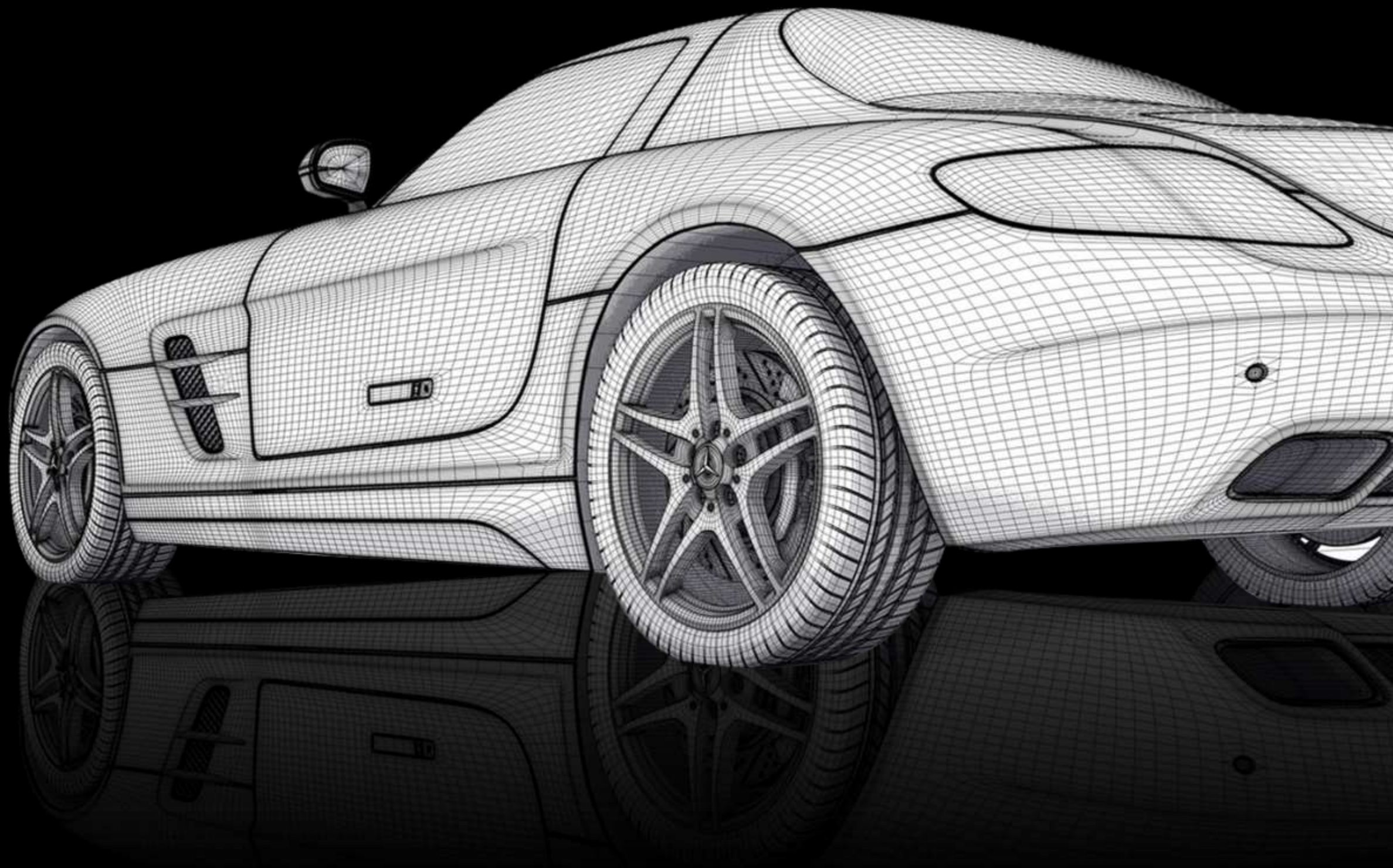
- 활용 방안 / 기대 효과
- 개선 방향



# 1

## 분석 배경/목표

- 중고차 시장의 성장
- 투명하지 못한 거래 행태
- 대기업의 독과점 우려





# 중고차 시장의 성장

## 중고차 거래, 신차 2배

〈단위:만대〉

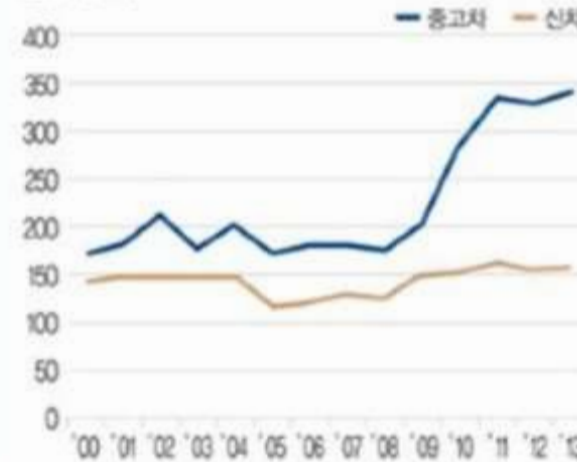


\*2021년 기준

\*자료:국토교통부

### 신차와 중고차 판매량

단위: 만 대



자료: 국토교통부·KTB투자증권

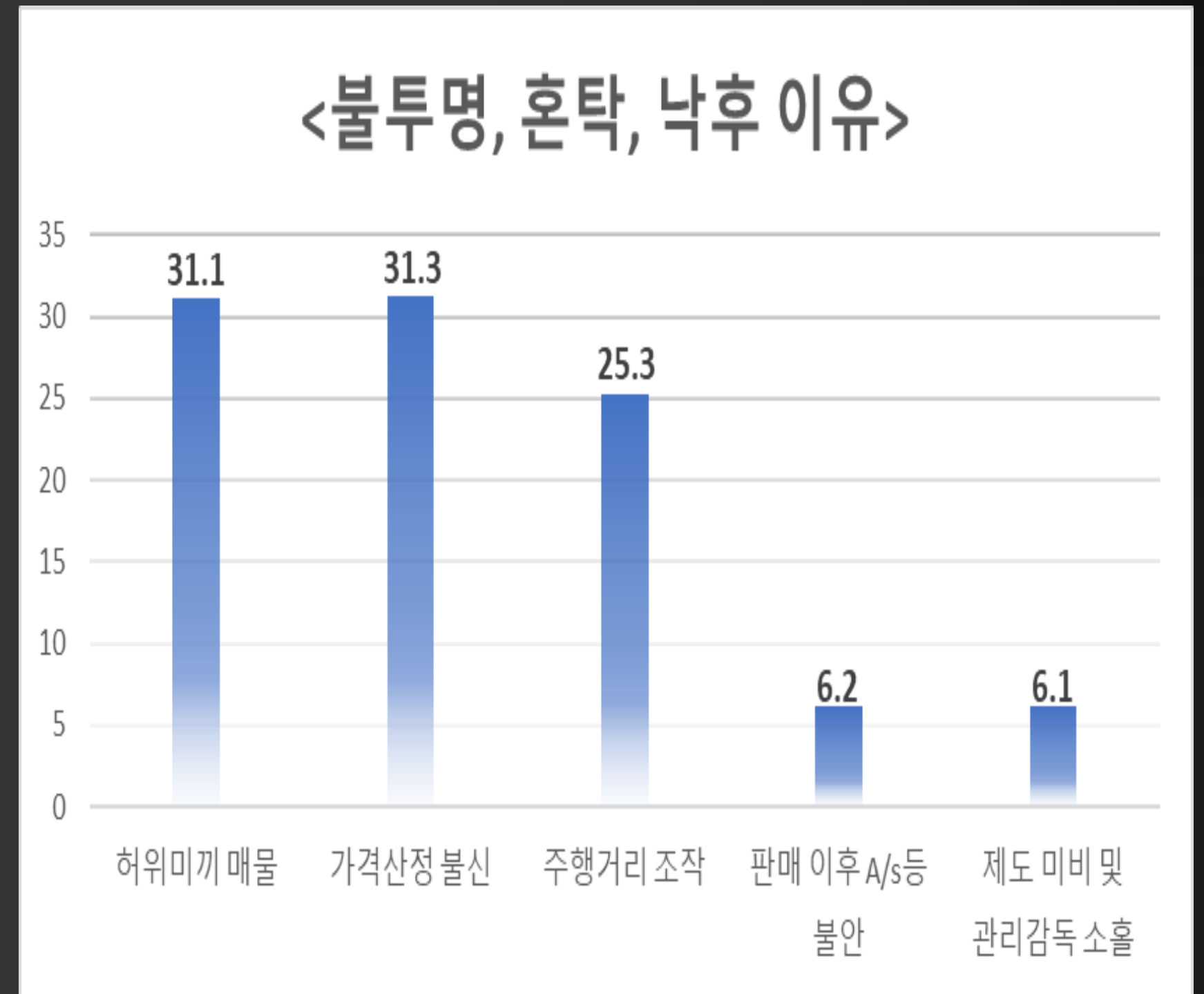
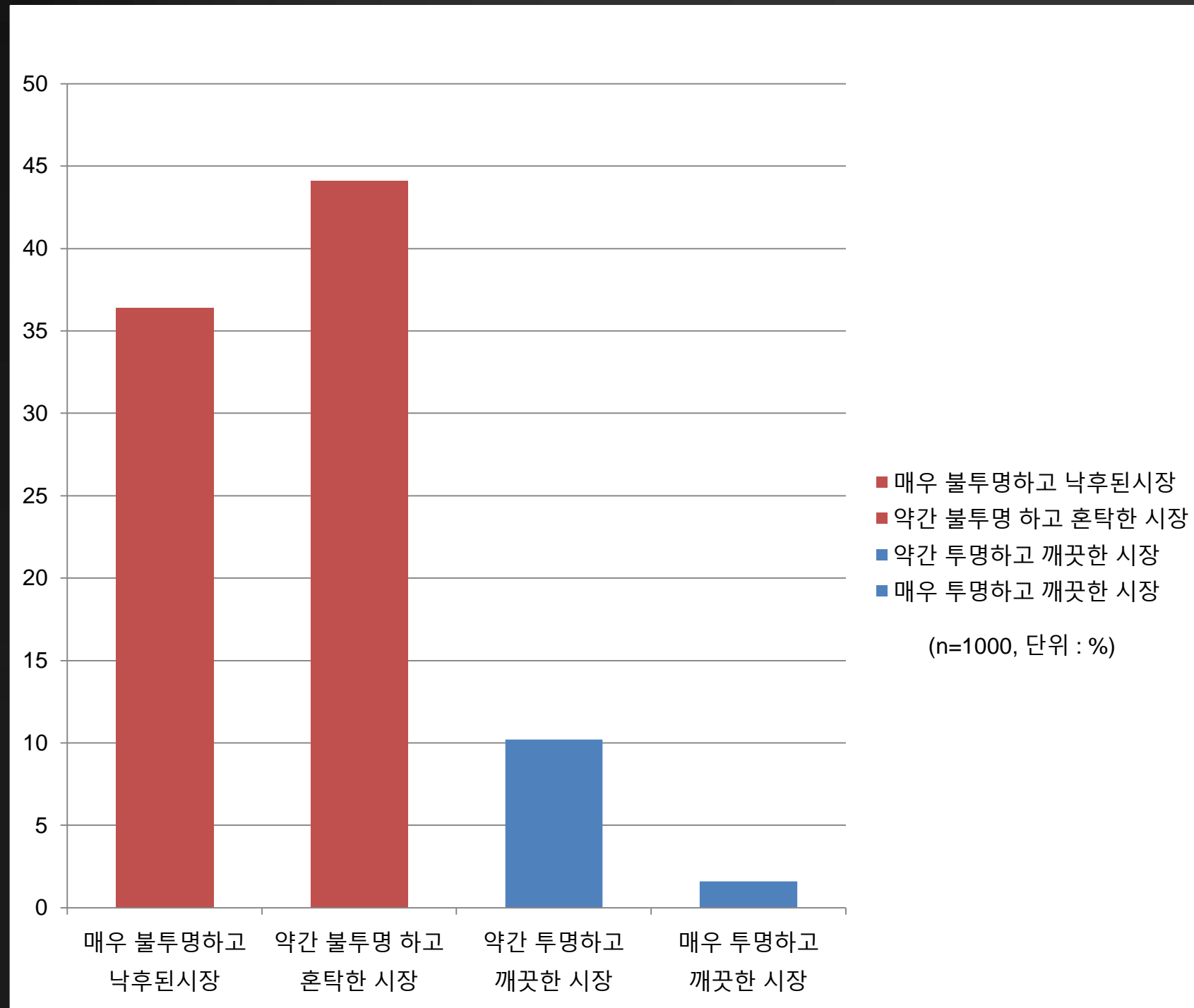


### 중고차 시장 추이



자료: 국토교통부

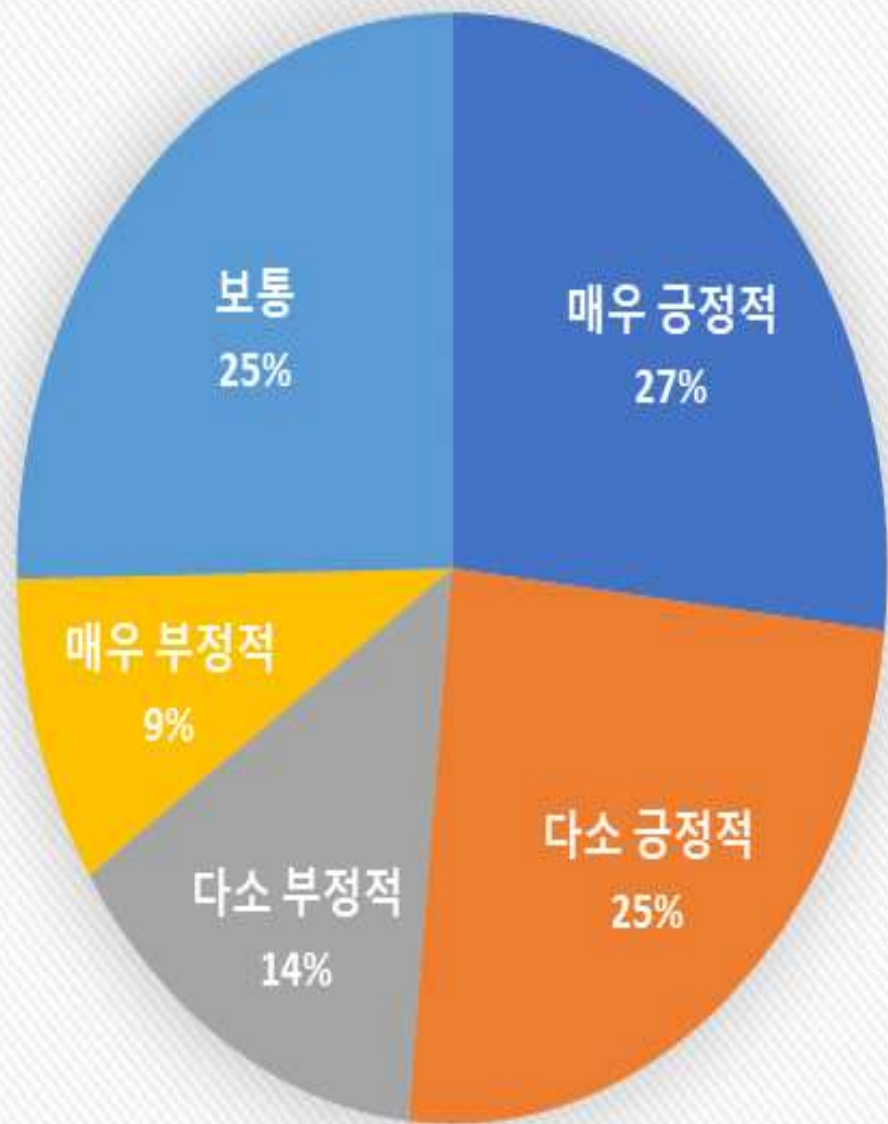
# 소비자의 낮은 신뢰도





# 대기업의 시장 진출 의견 및 독과점 우려

<대기업의 중고차 시장 진입 의견>



■ 매우 긍정적 ■ 다소 긍정적 ■ 다소 부정적 ■ 매우 부정적 ■ 보통

신차·중고차 가격 비교

현대 투싼		kia 쏘렌토	
가솔린 1.6 터보 2WD 인스퍼레이션		4세대 디젤 2.2 2WD 노블레스	
신 차	중고차	신 차	중고차
3345만원	3370만원	3953만원	3797만원
-개별소비세 3.5%	-2020년 11월	-개별소비세 3.5%	-2020년 6월
-옵션 190만원	-옵션 187만원	-옵션 350만원	-옵션 350만원

자료: 현대차·기아-엔카닷컴

시사저널e.

8월 중고차 매입 시세 상승률

단위: %. 7월 대비 상승률.



자료: AJ 셀카, 첫차



# 분석 목표

구분	브랜드	차종	출품대수	낙찰	낙찰률(%)	희망가대비 평균상승률(%)	인기도
대형	현대차	에쿠스(VI)	168	75	44.6	↑ 2.9	★
	현대차	제네시스(DH)	165	98	59.4	↑ 4.3	★★
	현대차	그랜저(HG)	524	270	51.5	↑ 4.6	★★
	기아차	K9	43	21	48.8	↑ 3.9	★
	기아차	더뉴K7	162	76	46.9	↑ 6.2	★
중형	현대차	쏘나타(YF)	324	173	53.4	↑ 6.5	★★
	현대차	쏘나타(LF)	258	152	58.9	↑ 4.9	★★★
	기아차	더뉴K5	224	122	54.5	↑ 4.5	★★★
	현대차	i40	14	7	50.0	↑ 6.5	★★
	르노삼성	SM5(신형)	33	20	60.6	↑ 3.3	★★
준중형	현대차	아반떼(MD)	74	53	71.6	↑ 2.4	★★★
	현대차	i30(신형)	12	10	83.3	↑ 10.4	★★
	르노삼성	뉴SM3(신형)	33	19	57.6	↑ 4.3	★★
소형	현대차	엑센트(RB)	79	36	45.6	↑ 6.5	★★
경차	기아차	레이	103	73	70.9	↑ 6.3	★★★
	기아차	올뉴모닝	141	100	70.9	↑ 4.3	★★★
	한국GM	스파크	45	28	62.2	↑ 3.3	★★
SUV/RV	현대차	싼타페(DM)	126	82	65.1	↑ 2.8	★★★
	기아차	뉴쏘렌토R	17	14	82.4	↑ 4.0	★★★
	기아차	더뉴스포티지R	73	42	57.5	↑ 3.4	★★
	기아차	모하비	33	14	42.4	↑ 3.0	★★
	기아차	카니발R	66	37	56.1	↑ 4.1	★★
승합	현대차	그랜드스타렉스	159	112	70.4	↑ 6.5	★★★
화물	현대차	포터II	206	147	71.4	↑ 8.0	★★★
	기아차	봉고III(화물)	122	62	50.8	↑ 6.6	★★

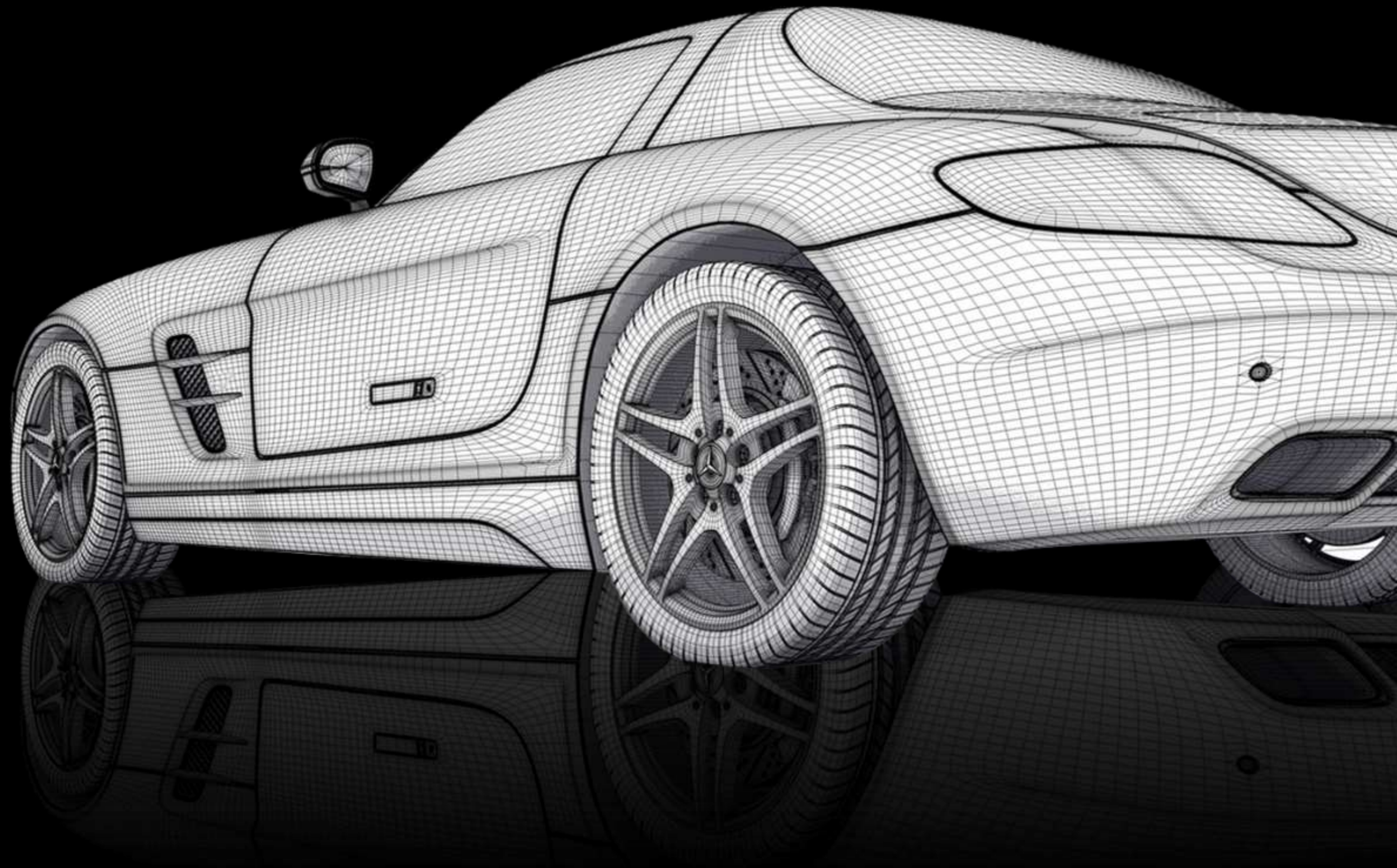




# 2

## EDA

- 데이터 이해
- 데이터 전처리
- 데이터 EDA





# 데이터 소개: 중고차 시장 데이터

## \* 범주 자료형

title: 제조사 모델명

location: 판매처(나이지리아 도시)

isimported: 현지 사용 여부

engine: 엔진 종류

transmission: 트랜스미션 종류

fuel: 연료 종류

paint: 페인트 색상

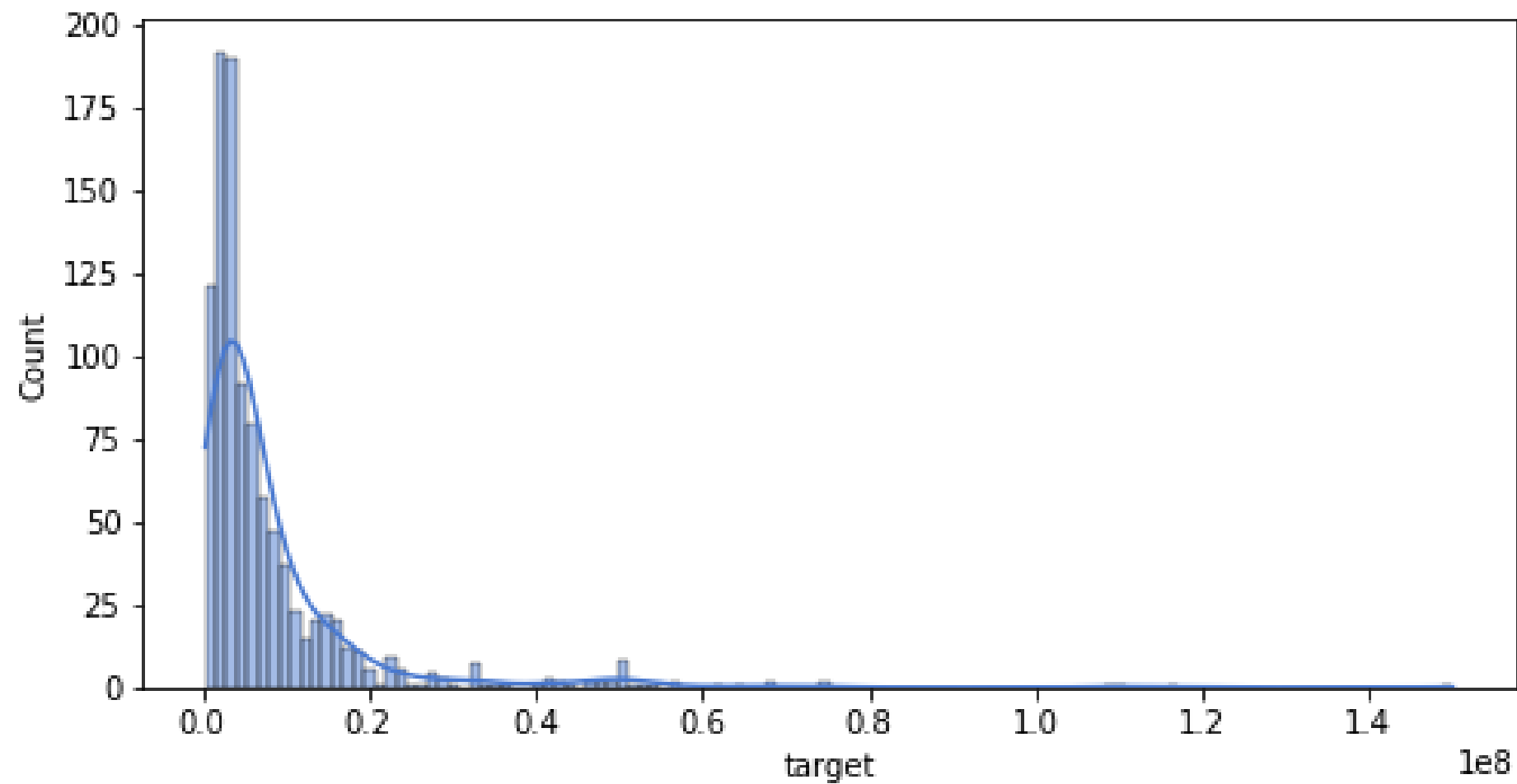
## \* 숫자/연속 자료형

odometer: 주행 거리

year: 제조년도

target(price): 자동차 가격

# 자동차 가격

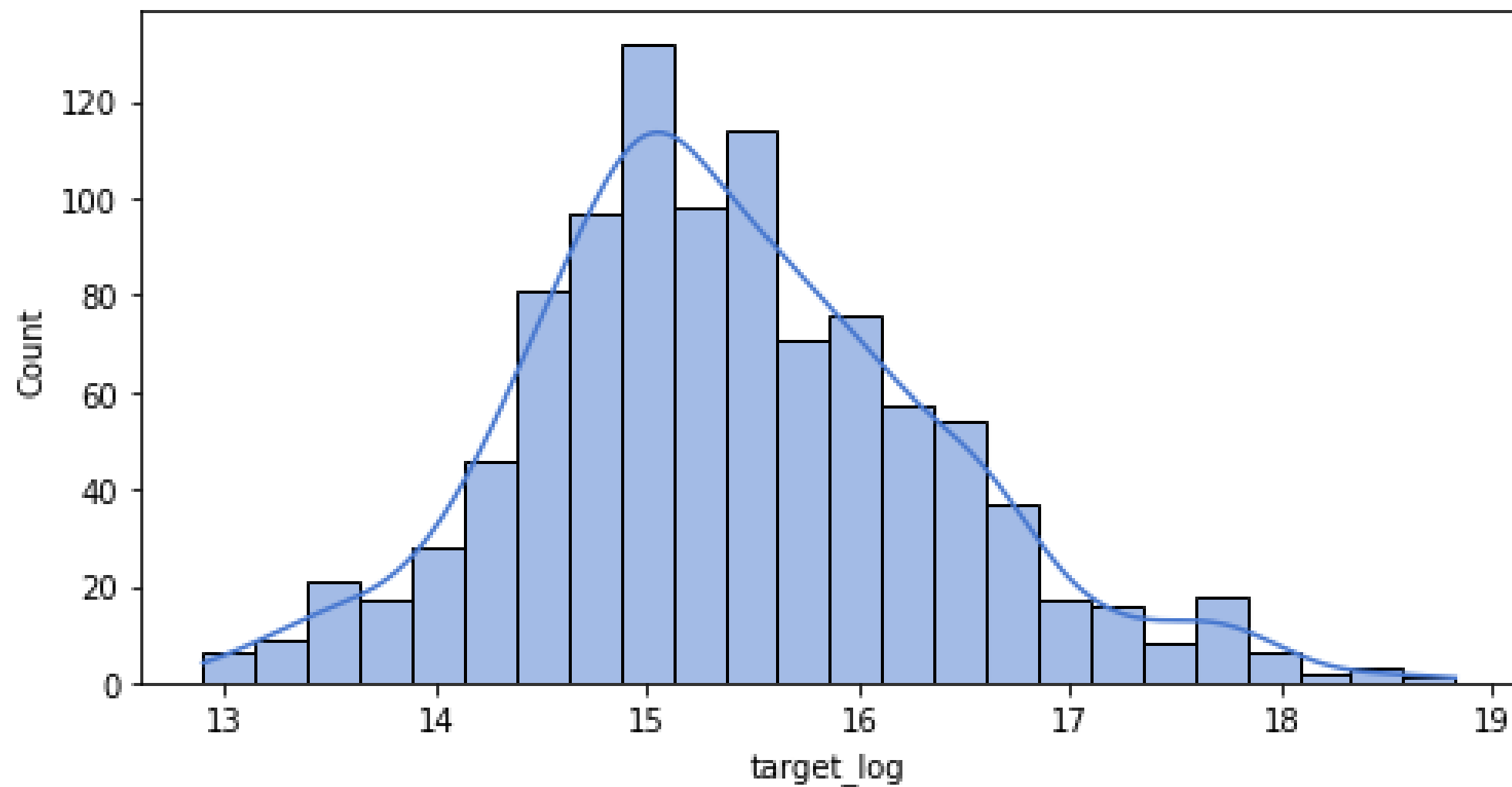


샘플 수	1,015
최소값	400,000
평균	8,243,204.45
최대값	150,015,008
비대칭도(왜도)	4.95

. 가격대가 왼쪽으로 치우쳐 높은 비대칭도를 보이고 있다.



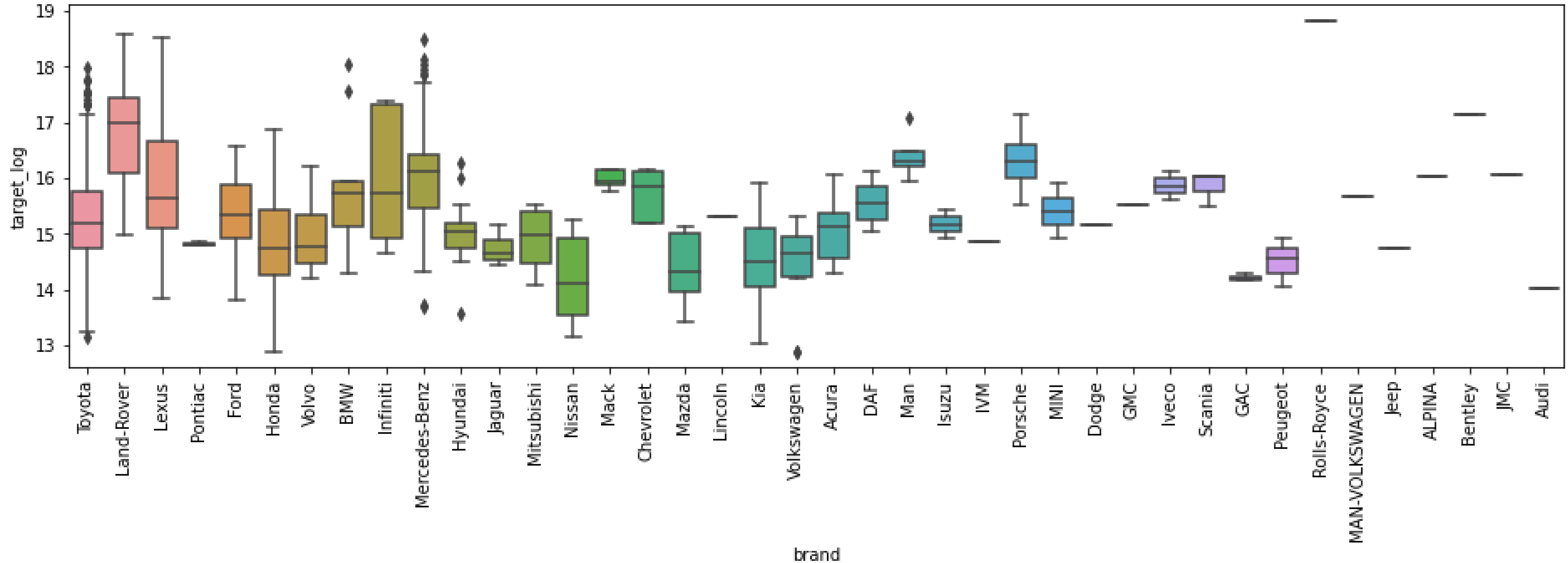
# 자동차 가격의 로그값



샘플 수	1,015
최소값	12.9
평균	15.4
최대값	18.8
비대칭도(왜도)	0.38

시각화를 더 용이하게 하기 위해서 이 후 그래프에서는 자동차 가격의 로그값을 사용

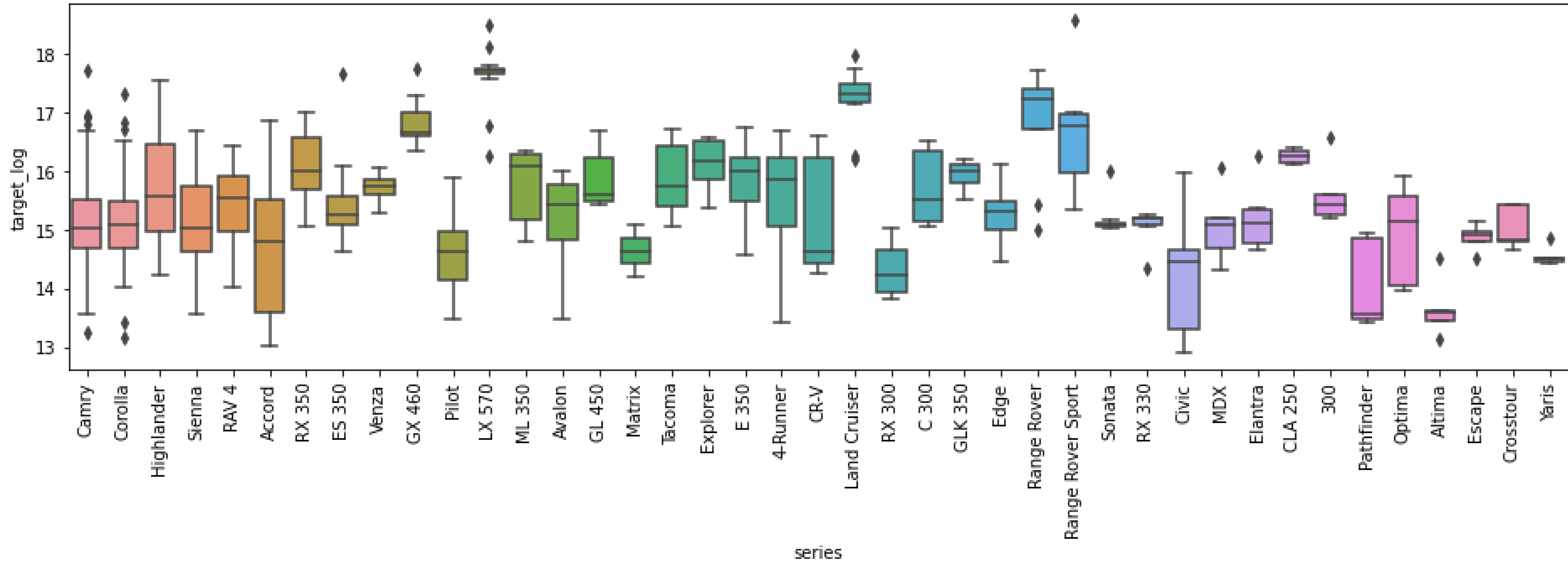
# 브랜드 vs. 가격



- 다양한 차종이 중고차 시장에 나와있는 브랜드의 경우 가격대의 분포가 더 넓게 형성되어 있다.
- 각 브랜드마다 형성된 가격대라는 것이 존재한다.

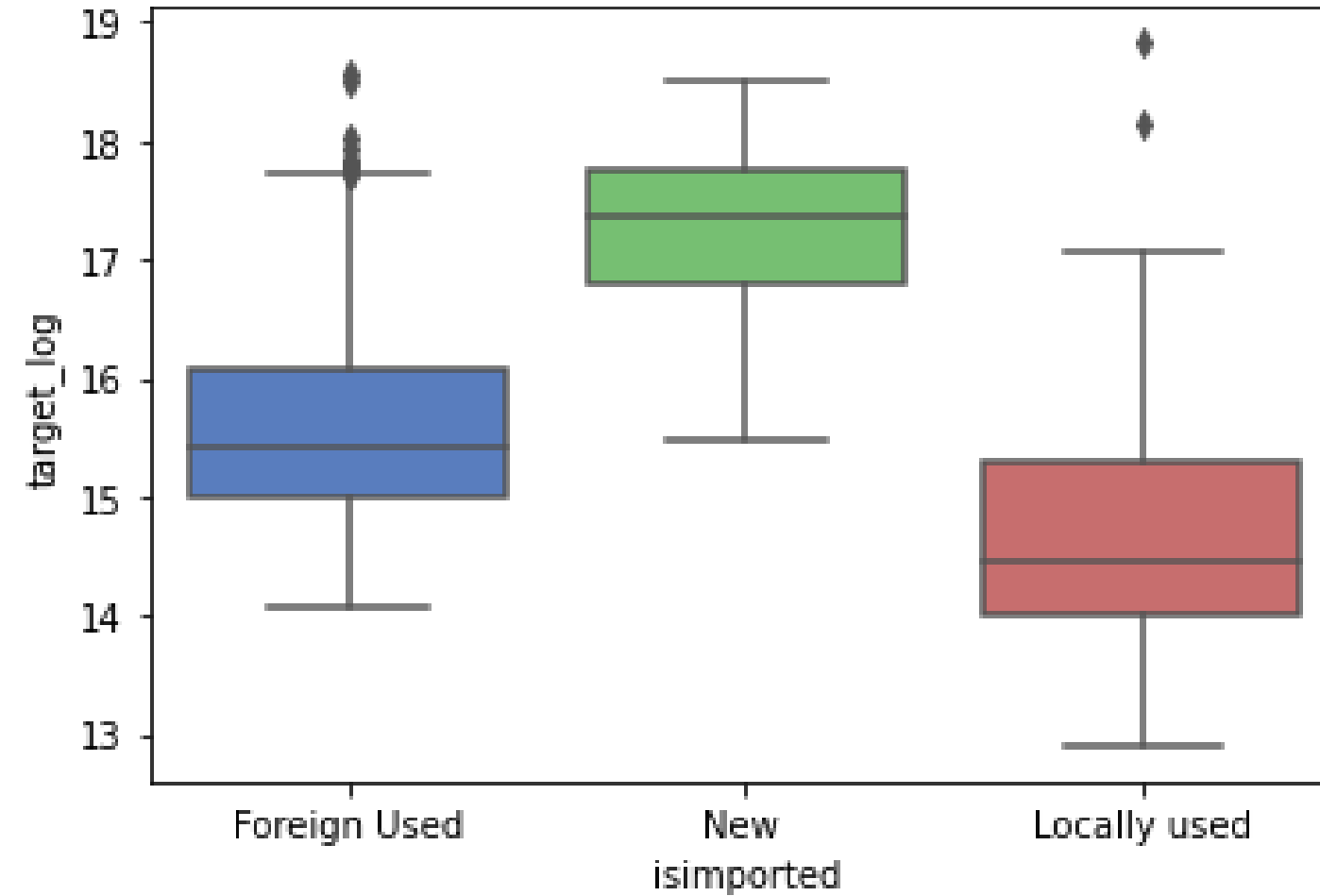
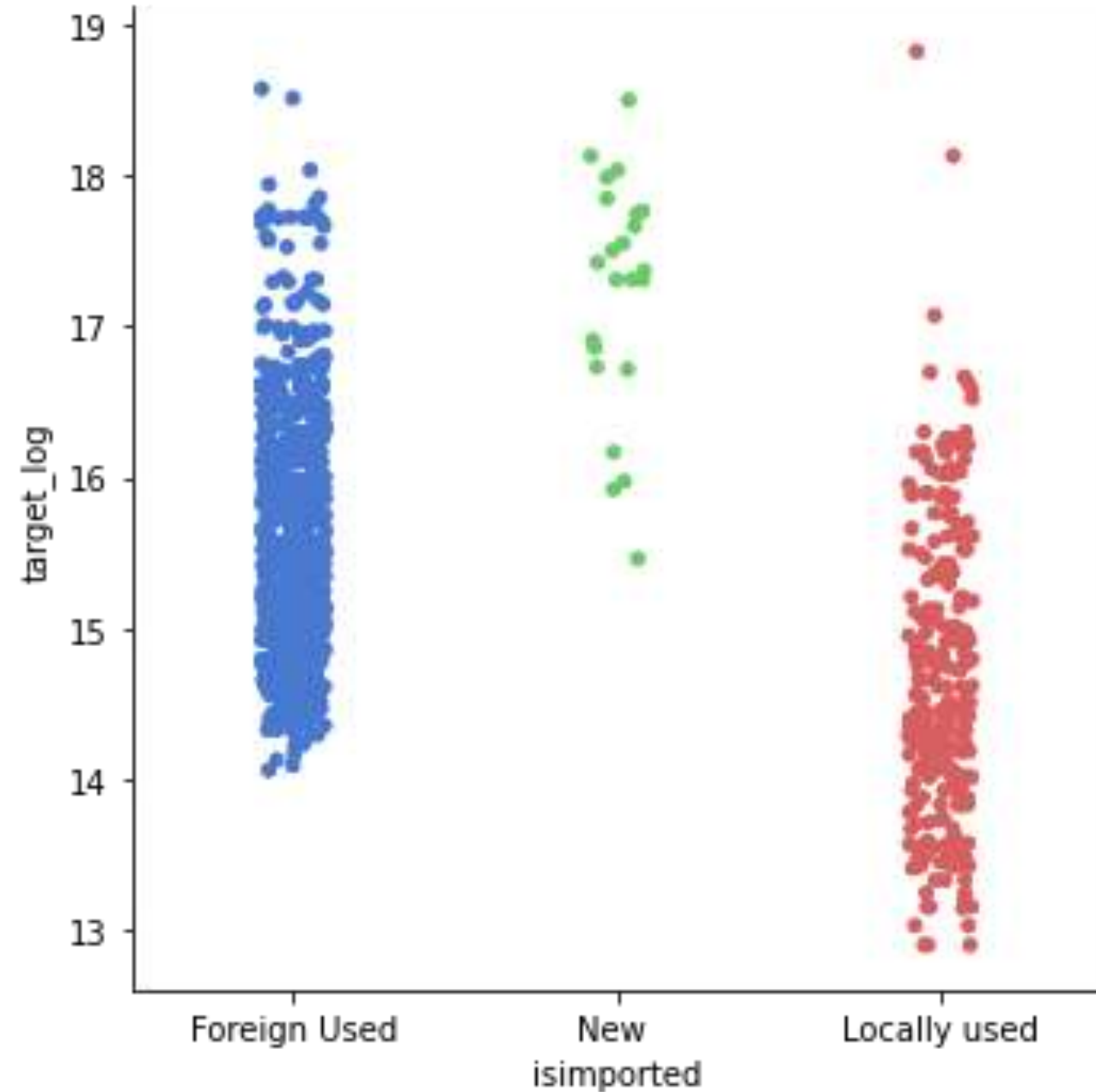


# 자동차 모델 vs. 가격 (5개 이상 데이터가 존재하는 차종의 시세)



- 중고차 시장에 많이 나와있는 모델일수록 가격대의 분포가 더 넓다.
- 모델별 형성된 시세가 있으나 다른 요소에 따라서 사분위수 범위 밖의 극단치가 존재한다.

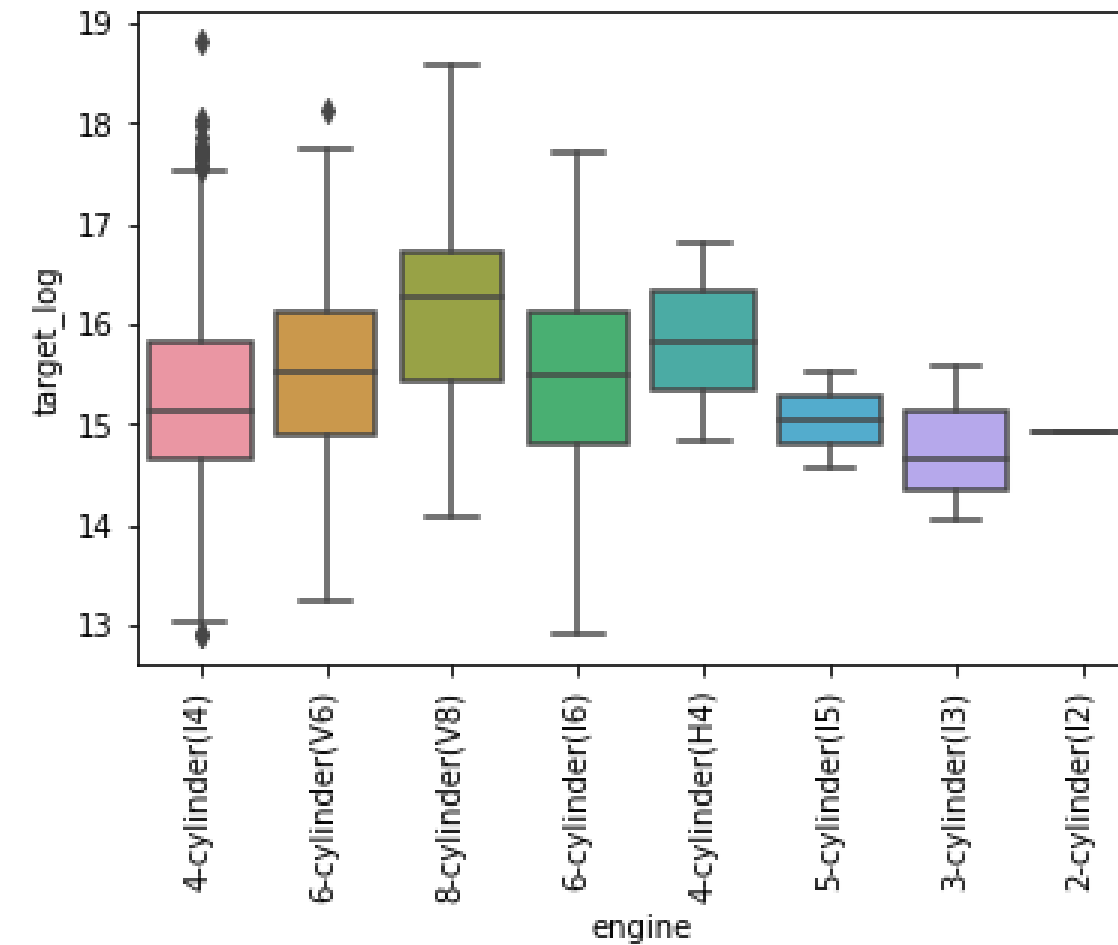
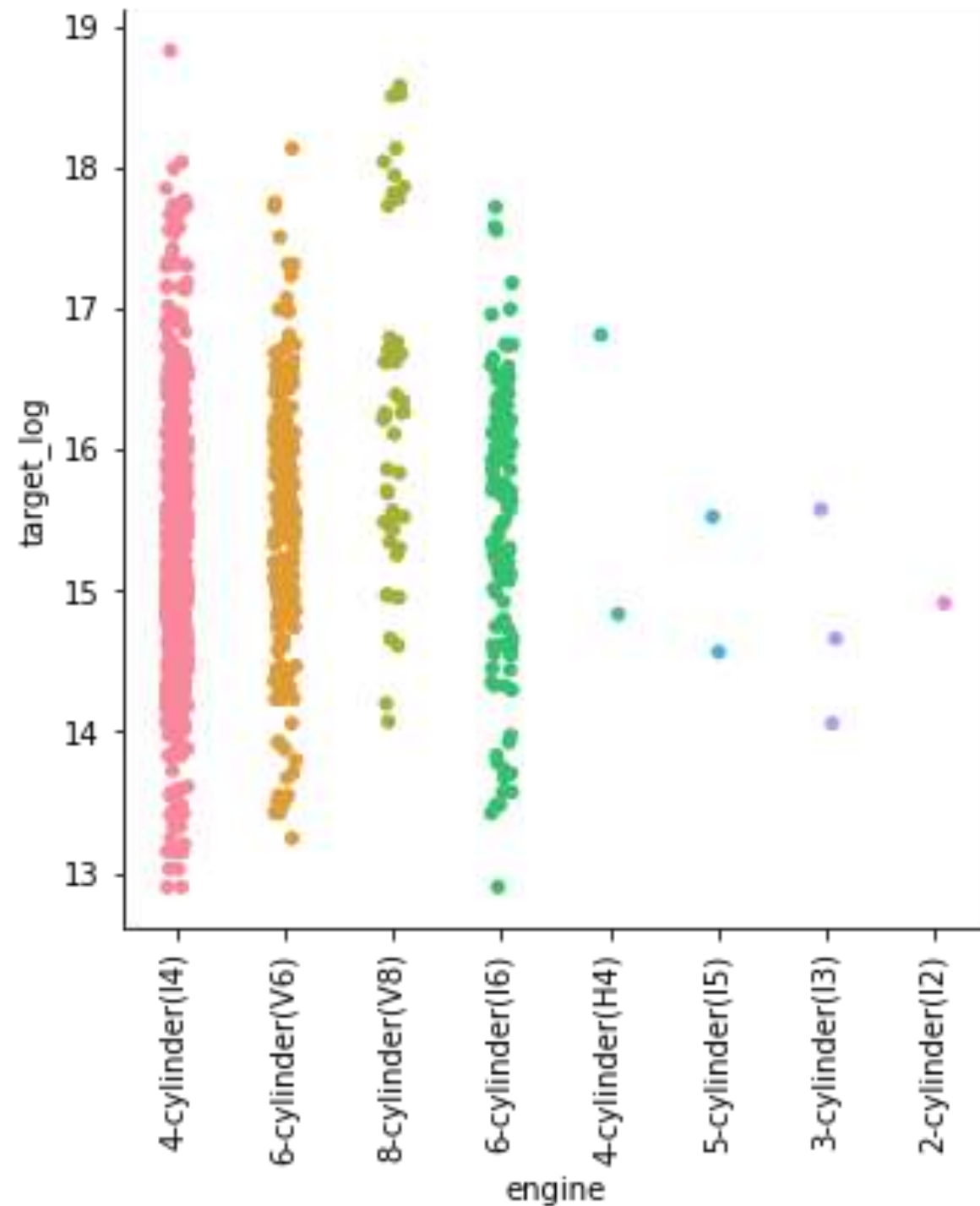
# 현지 사용 여부 vs. 가격



- 신차의 시세가 가장 높으며 그 수도 가장 적다.
- 외국에서 사용된 차량의 시세가 나이지리아 현지에서 사용된 차량의 시세보다 눈에 띄게 높다.
- 차량이 사용된 환경이 중고차 가격 결정에 큰 영향을 미친다.

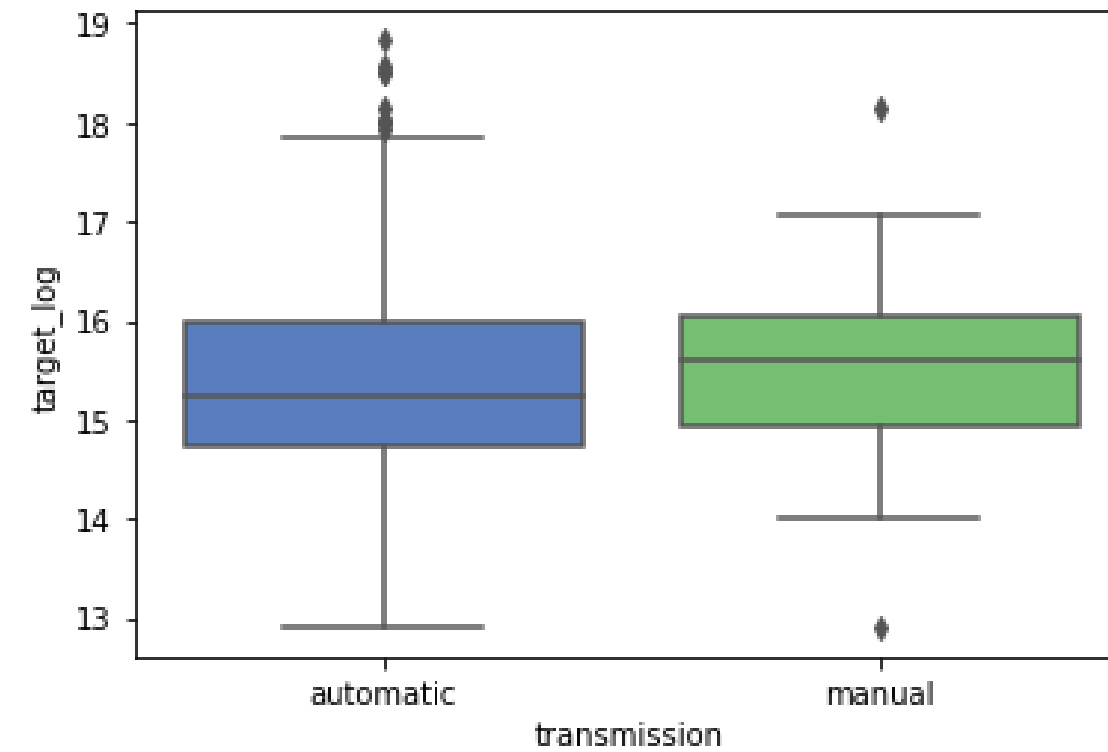
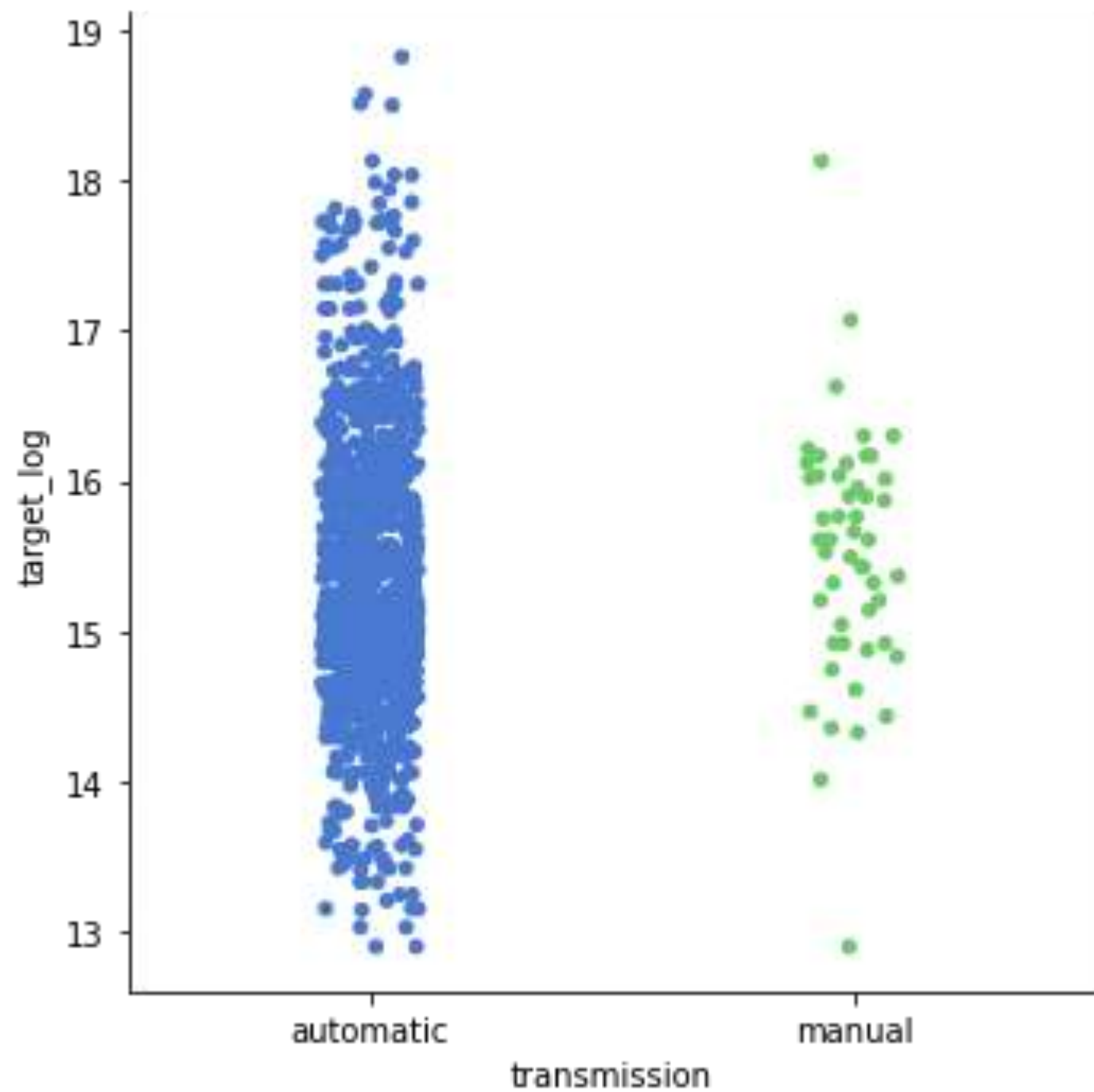


# 엔진 종류 vs. 가격



- I-4 cylinder, V-6 cylinder, I-6 cylinder 이 가장 흔하다.
- 사분위수 범위 밖에 존재하는 데이터가 많아서 엔진 종류라는 요소 하나가 가격 결정에 유의미한 영향을 준다고 단언하기 힘들다.

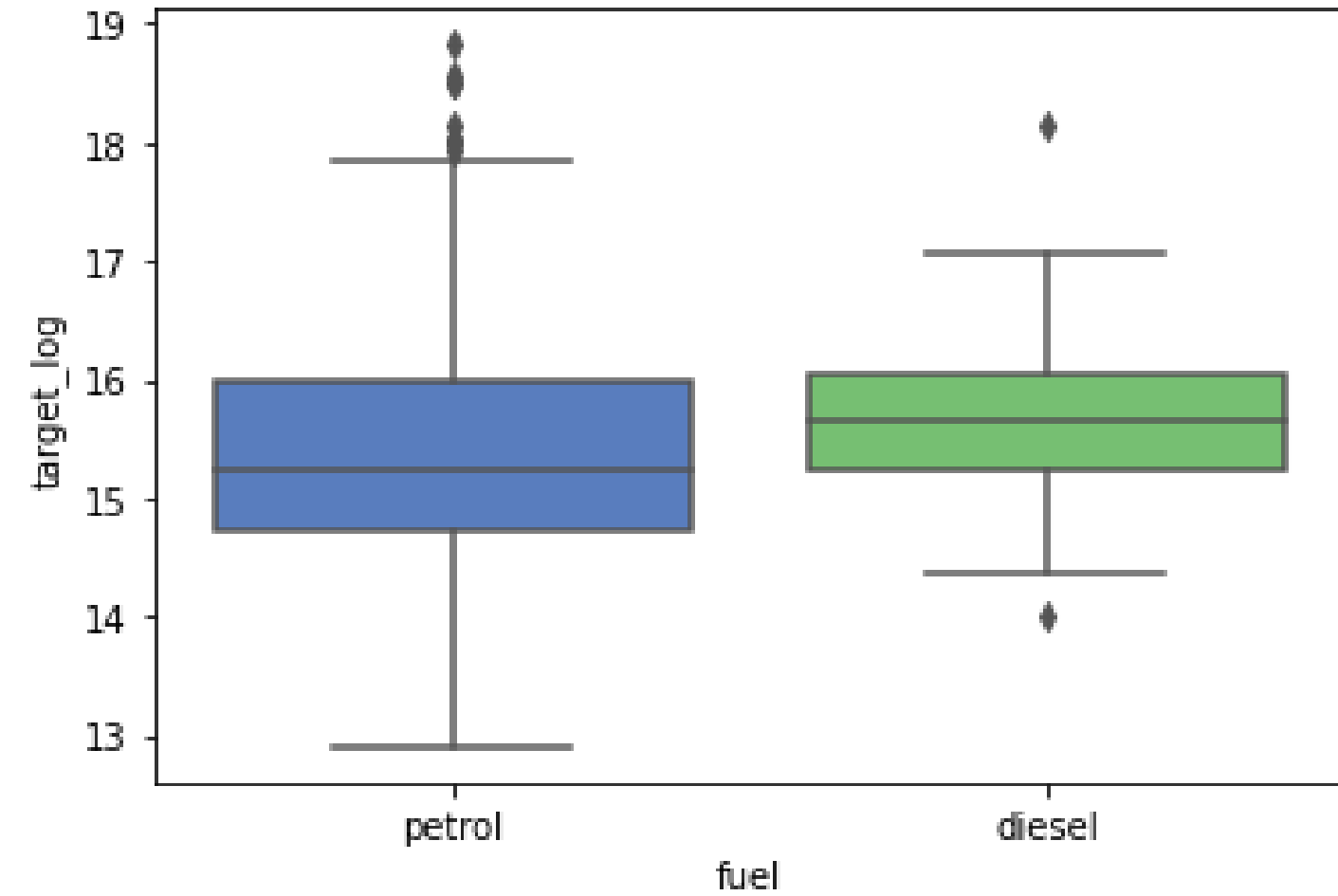
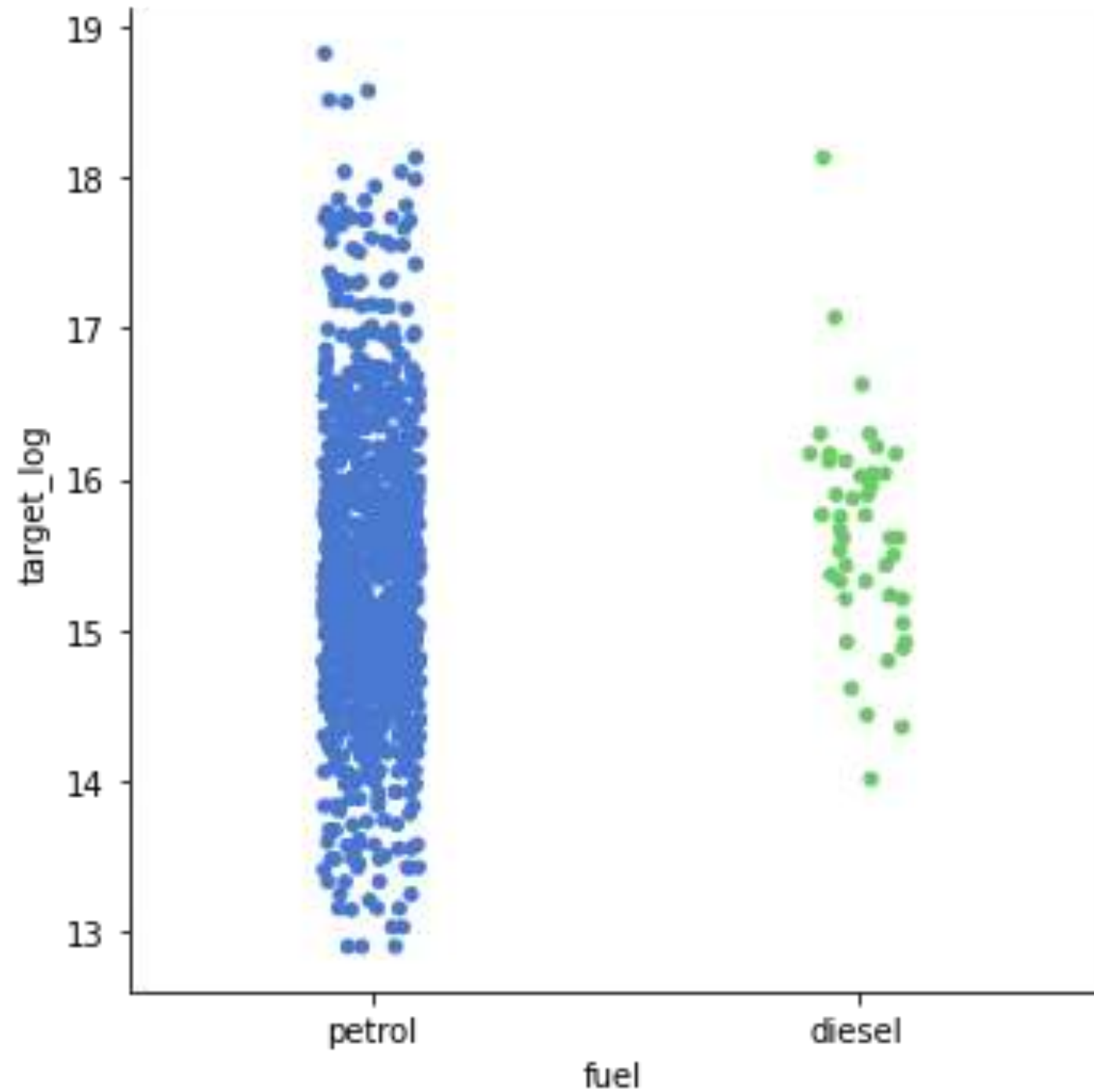
# 트랜스미션 vs. 가격



- 자동변속기 차량의 매물이 수동변속기 차량의 매물 수 보다 훨씬 더 많다.
- 변속기별 가격 분포가 크게 차이 나지 않아 변속기 종류만으로 가격 결정에 유의미한 영향을 주었다고 단언하기 힘들다.

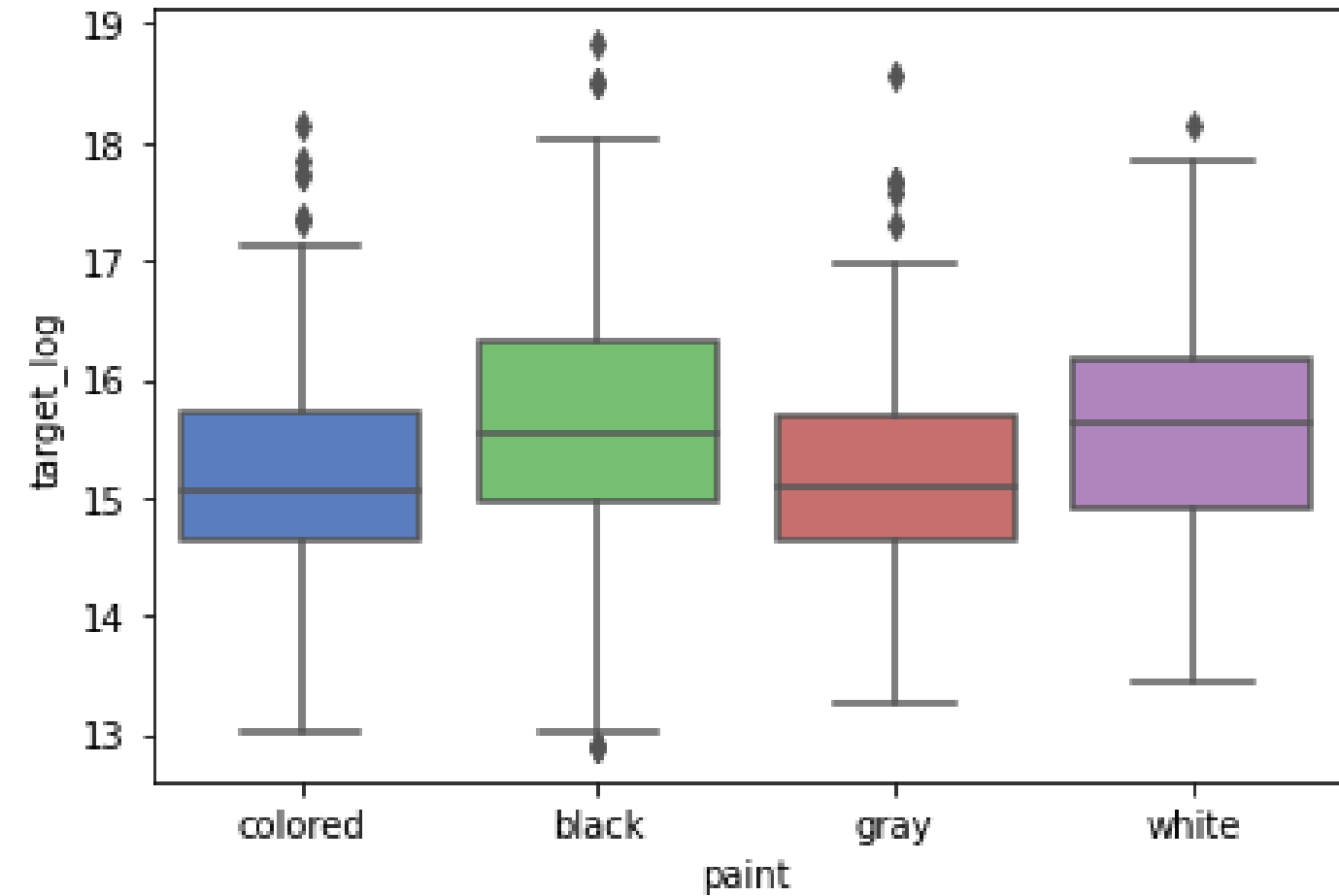
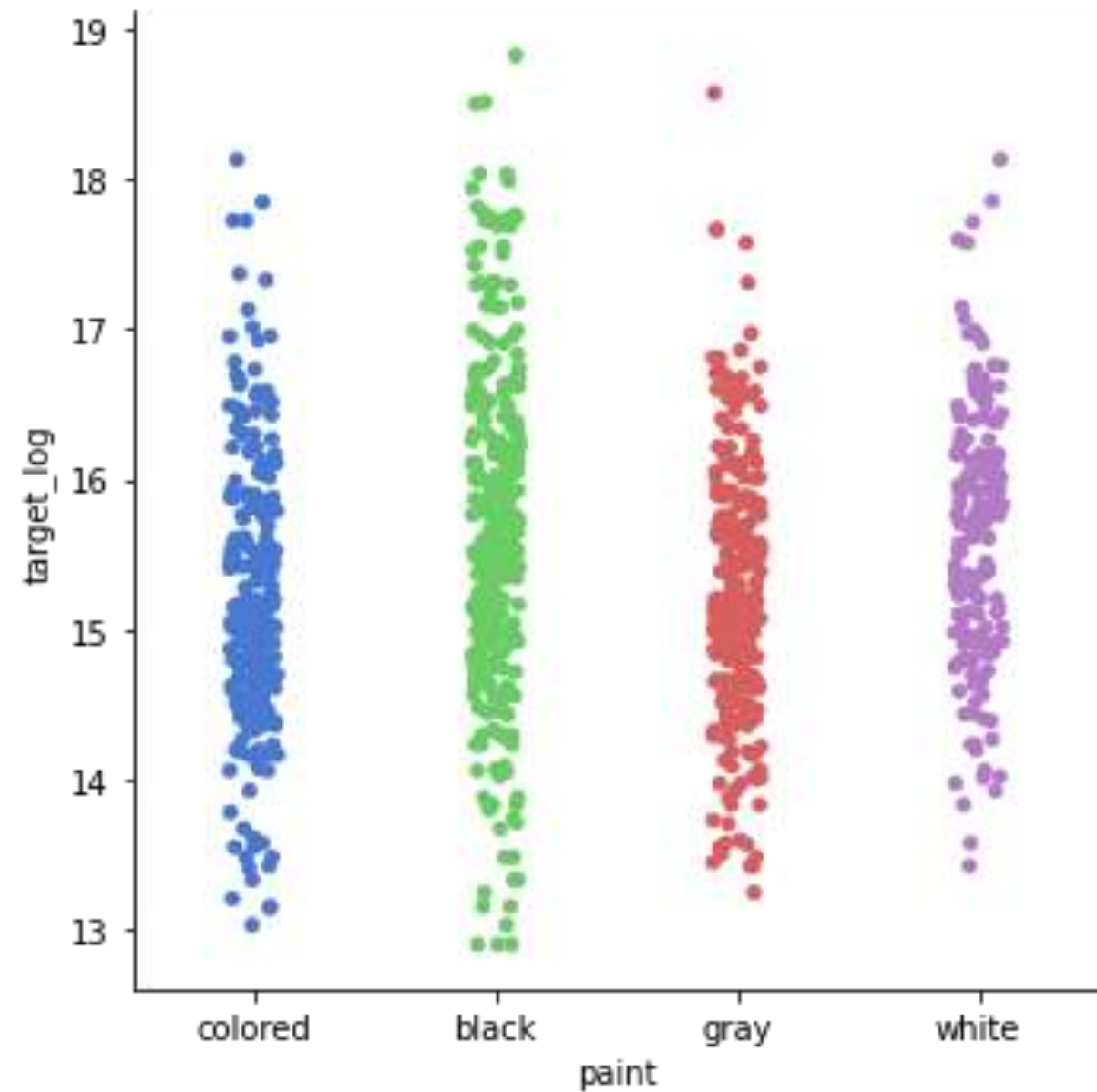


# 연료 종류 vs. 가격



- 가솔린 차량의 매물이 디젤 차량의 매물 수 보다 훨씬 더 많다.
- 연료별 가격 분포가 크게 차이 나지 않아 연료 종류만으로 가격 결정에 유의미한 영향을 주었다고 단언하기 힘들다.

# 페인트 색상 vs. 가격



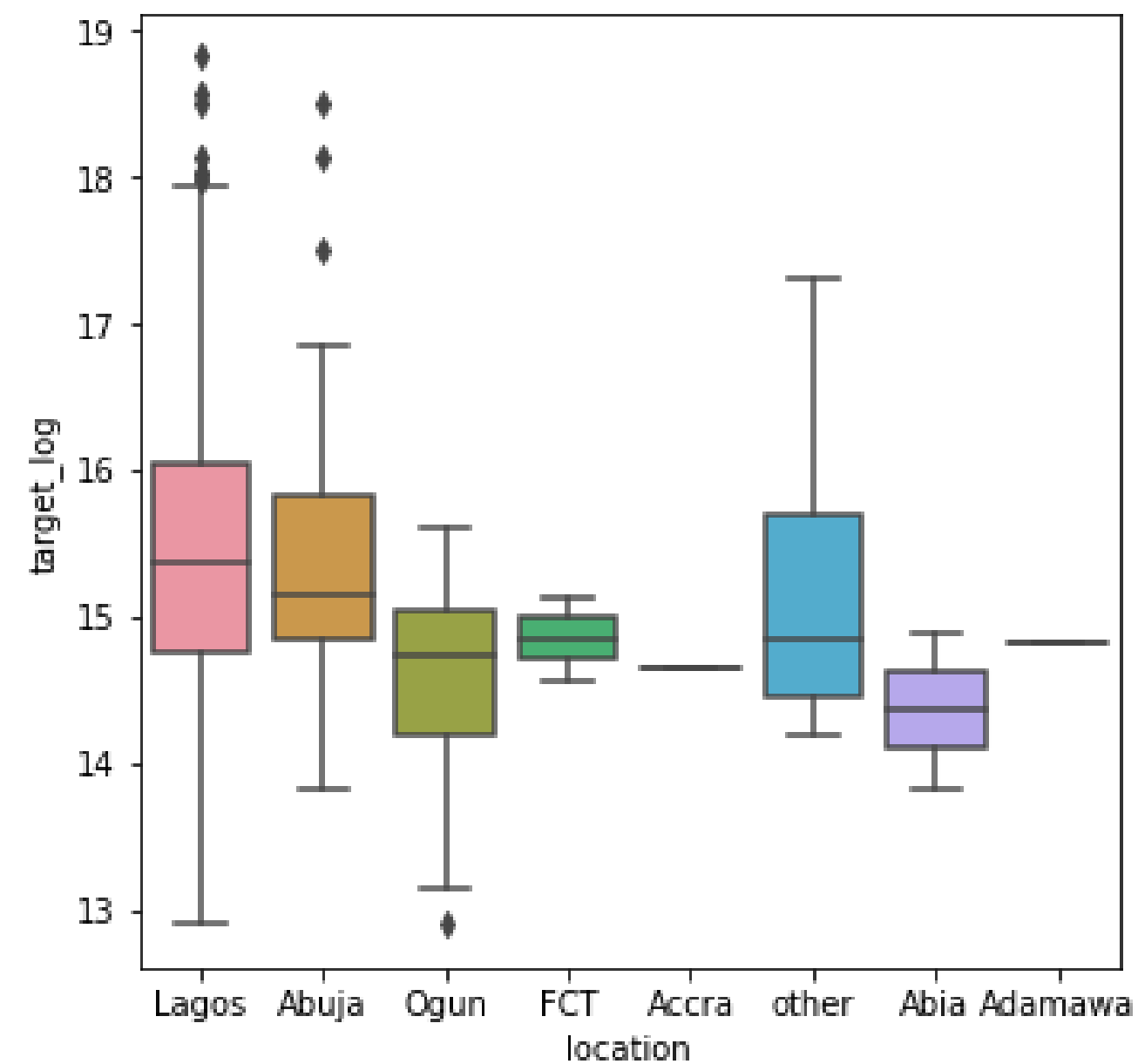
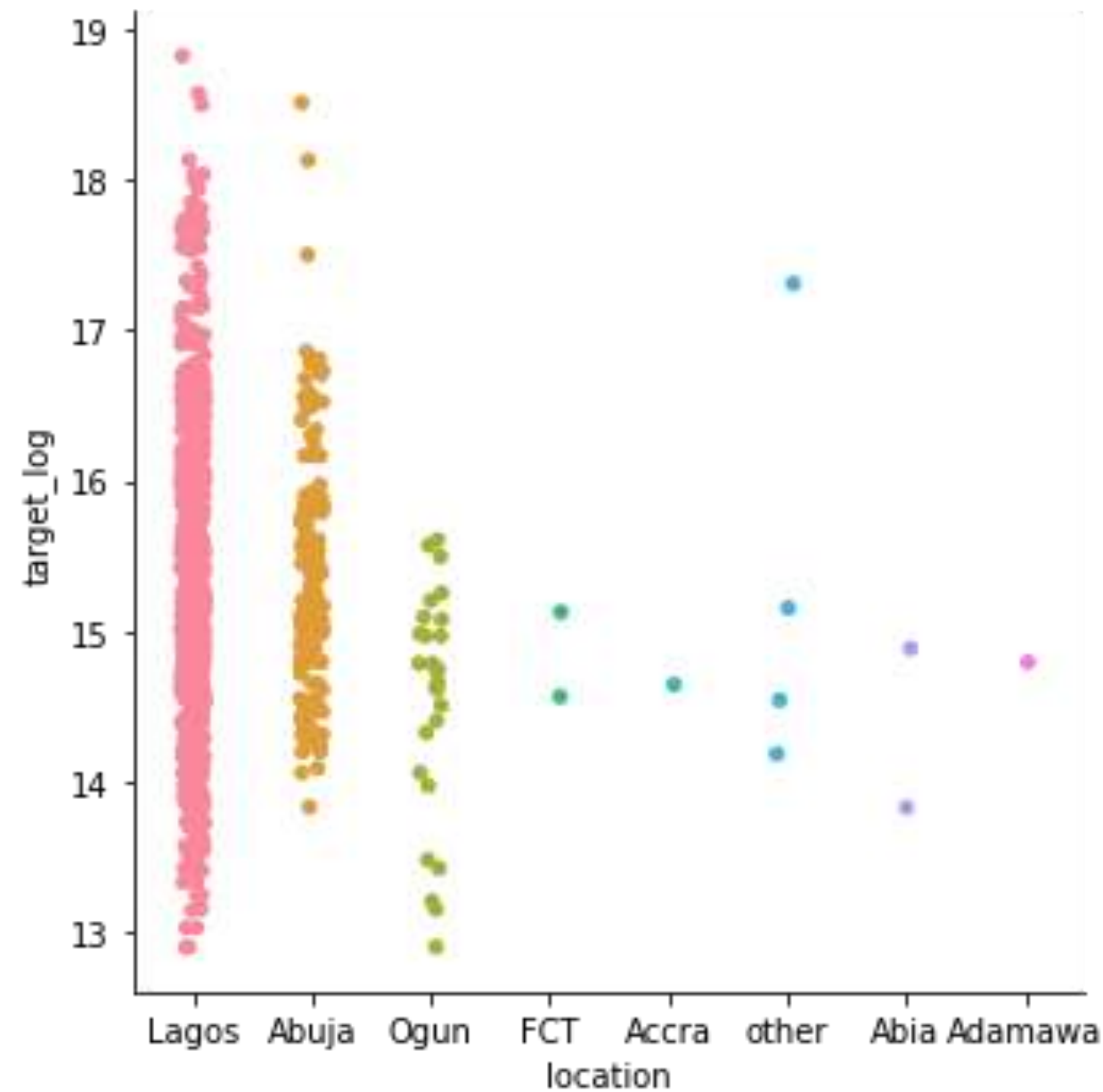
평균 값 : White  $\geq$  Black  $>$  gray  $\geq$  colored

이상치 개수 : 범위 밖의 데이터가 많다.

차량 색상이 가격 결정 요소에 유의미한 영향을 주었다고 말하기 힘들다.

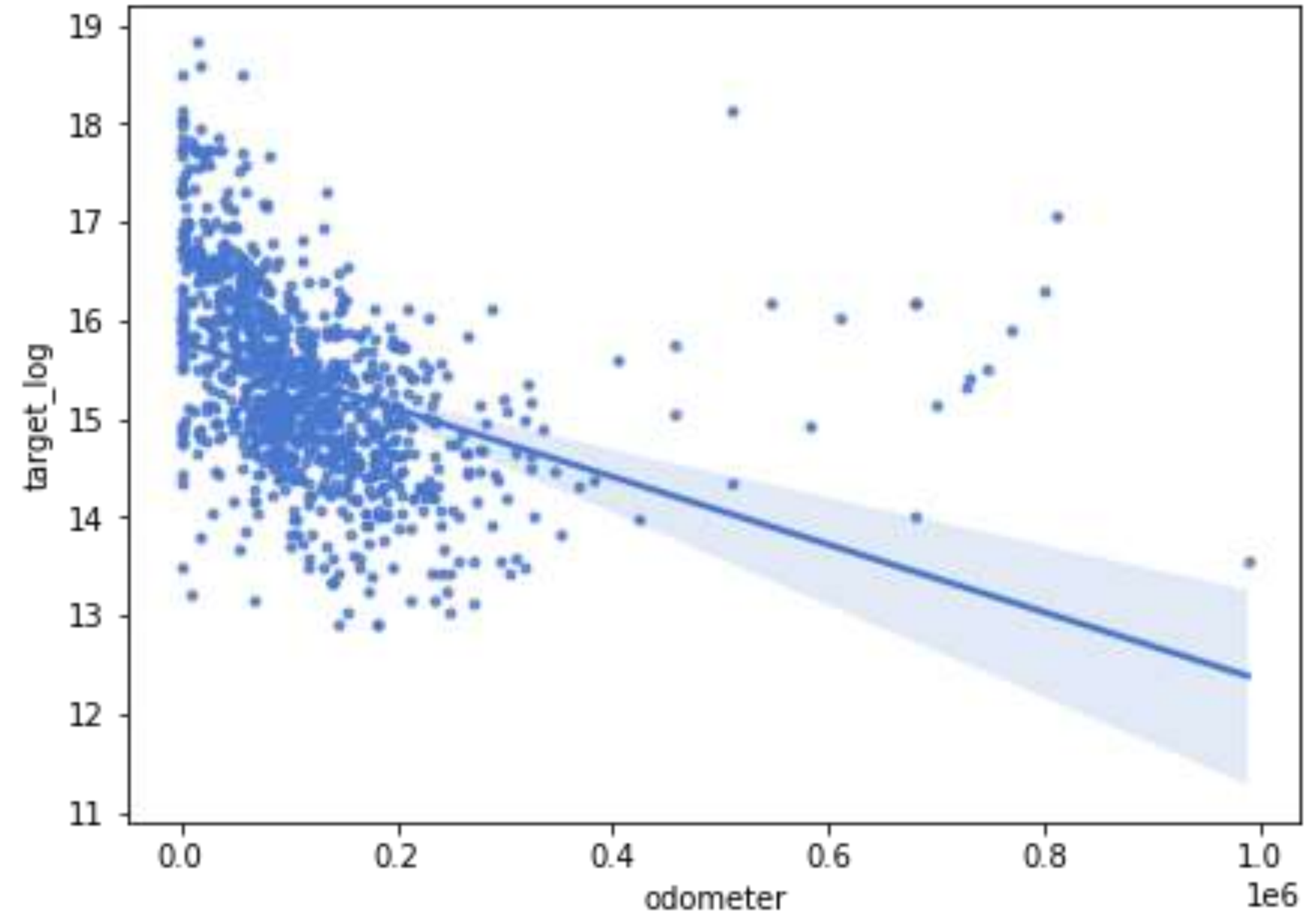
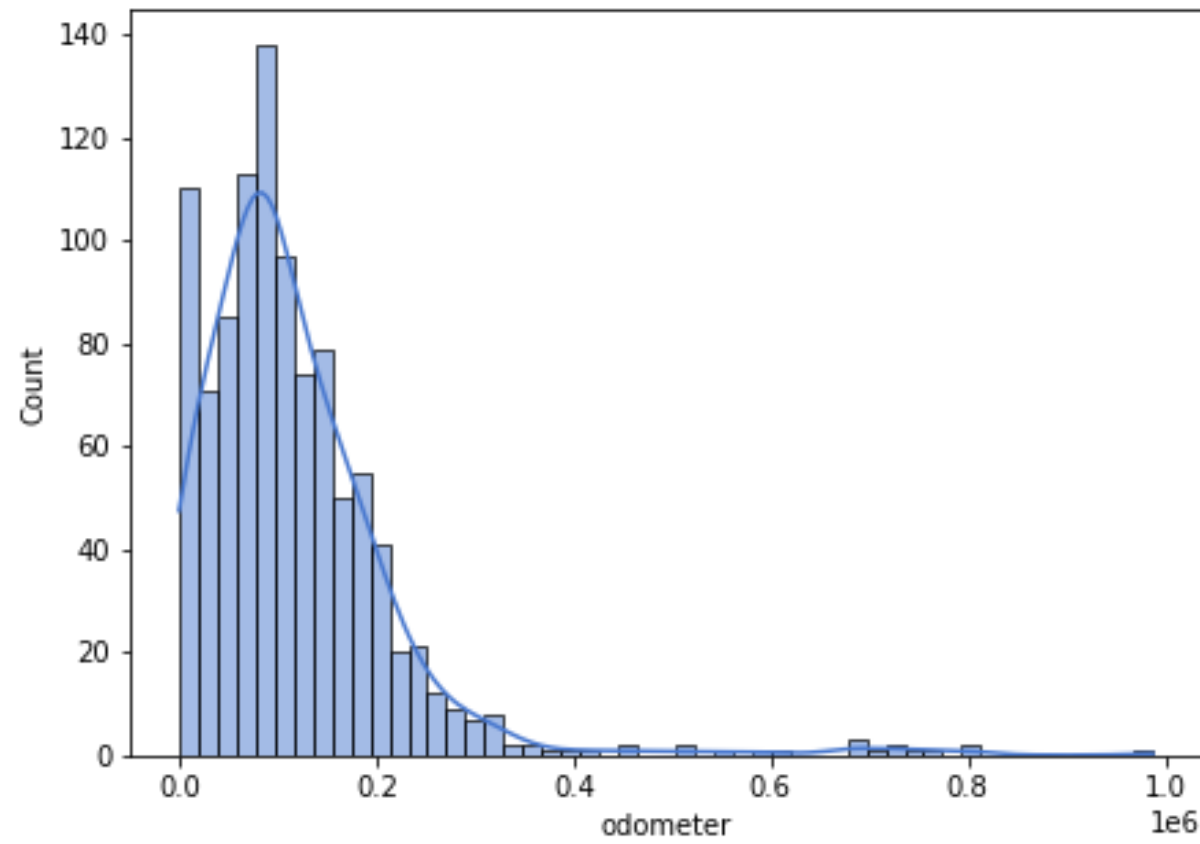


# 판매처 vs. 가격



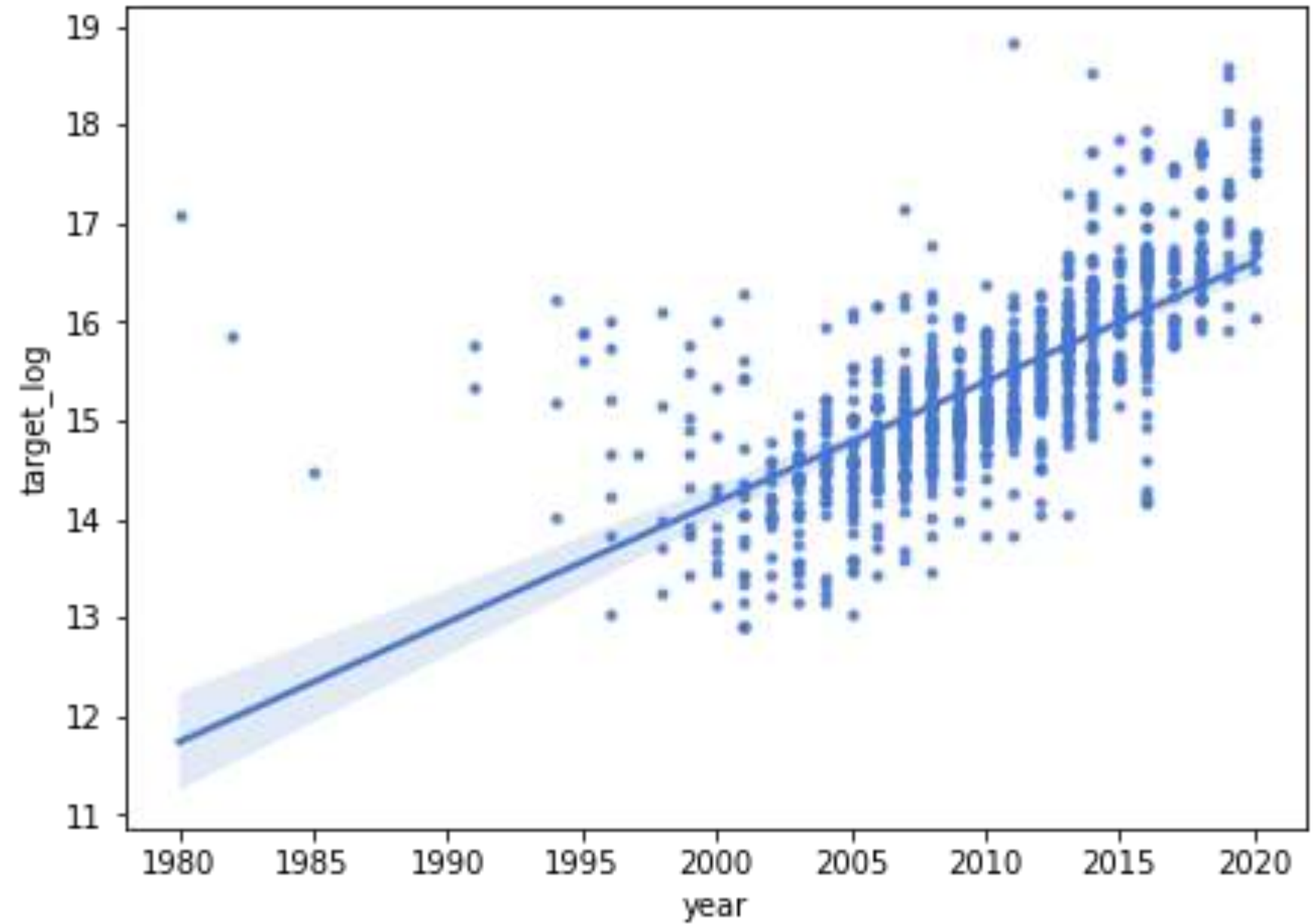
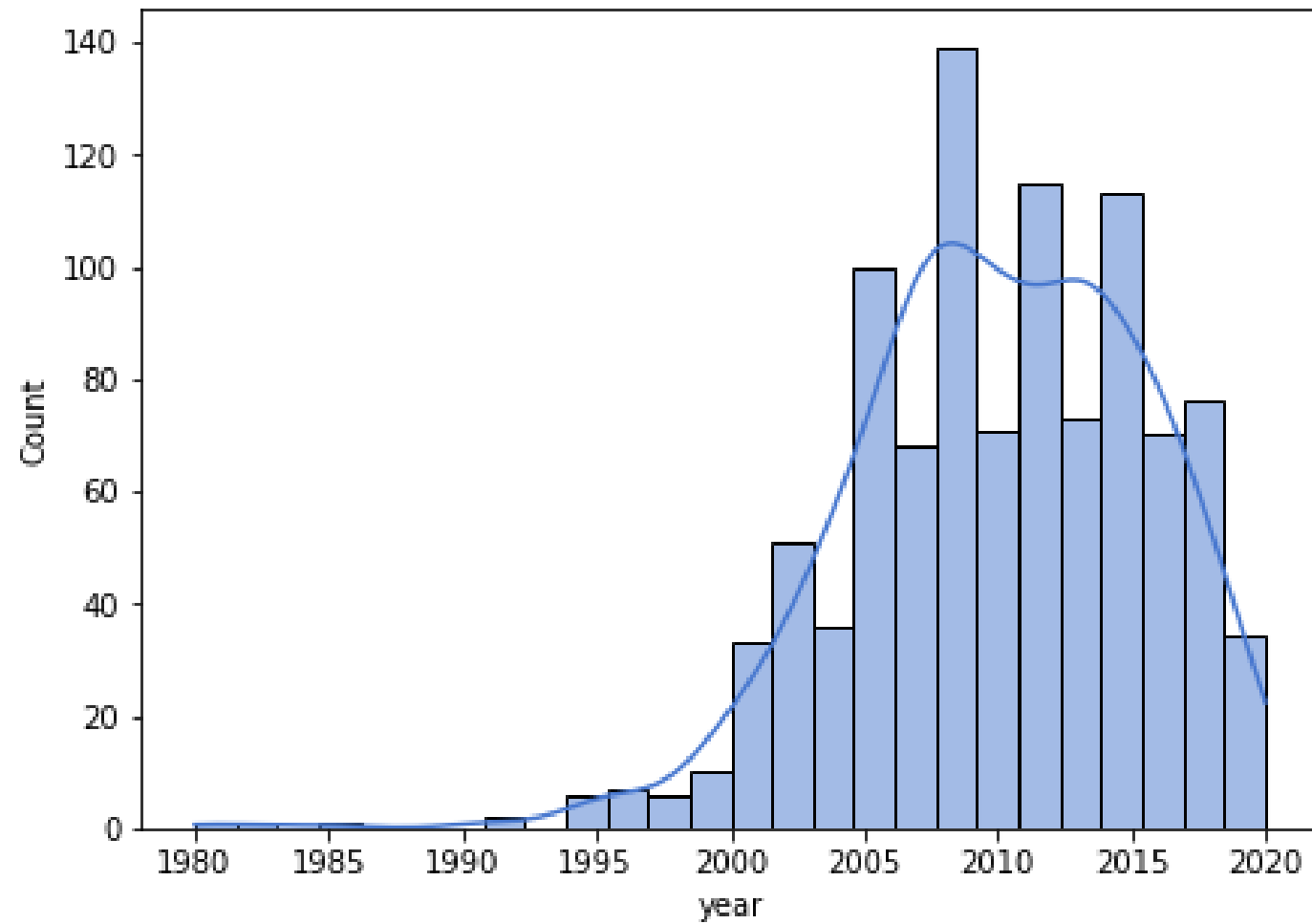
- 주로 Lagos에서 수입해오며 Abuja와 Ogun이 그 뒤를 따르며 다른 세 곳의 경우, 데이터가 부족하다.
- 수입량이 많은 Lagos의 가격대 분포가 가장 넓다.
- Ogun에서 수입된 차량의 경우 가격대가 더 낮게 형성이 되어 있다.

# 주행 거리



. 주행거리가 커질 수록 시세가 하락한다.

# 제조년도



- 제조년도가 현재에 더 가까울 수록 시세가 높아진다.



# INSIGHT

- ✓ 제조사와 모델별로 형성된 시세가 있음을 확인하였다.
- ✓ 엔진, 변속기, 연료, 색 종류 등은 각각의 요소들이 시세 형성에 유의미한 영향을 미쳤다가보다는 다른 요소와 함께 유기적으로 영향을 미쳤을 확률이 높다.
- ✓ 특정한 판매처의 눈에 띄는 낮은 시세와, 현지 사용 여부에 따른 시세의 큰 차이는, 차량의 사용 여부와 차량이 사용된 환경이 시세에 유의미한 영향을 미쳤음을 보여준다.
- ✓ 주행 거리와 시세는 반비례하며 제조 연도와 시세는 정비례한다.

## ➤ 적용 가능한 서비스 영역 :

제조사와 모델, 엔진, 연료, 색 등은 구매자가 쉽게 확인할 수 있는 외적 요소들이며 이 중 시세 형성에 큰 영향을 미치는 요소는 제조사와 모델이다.

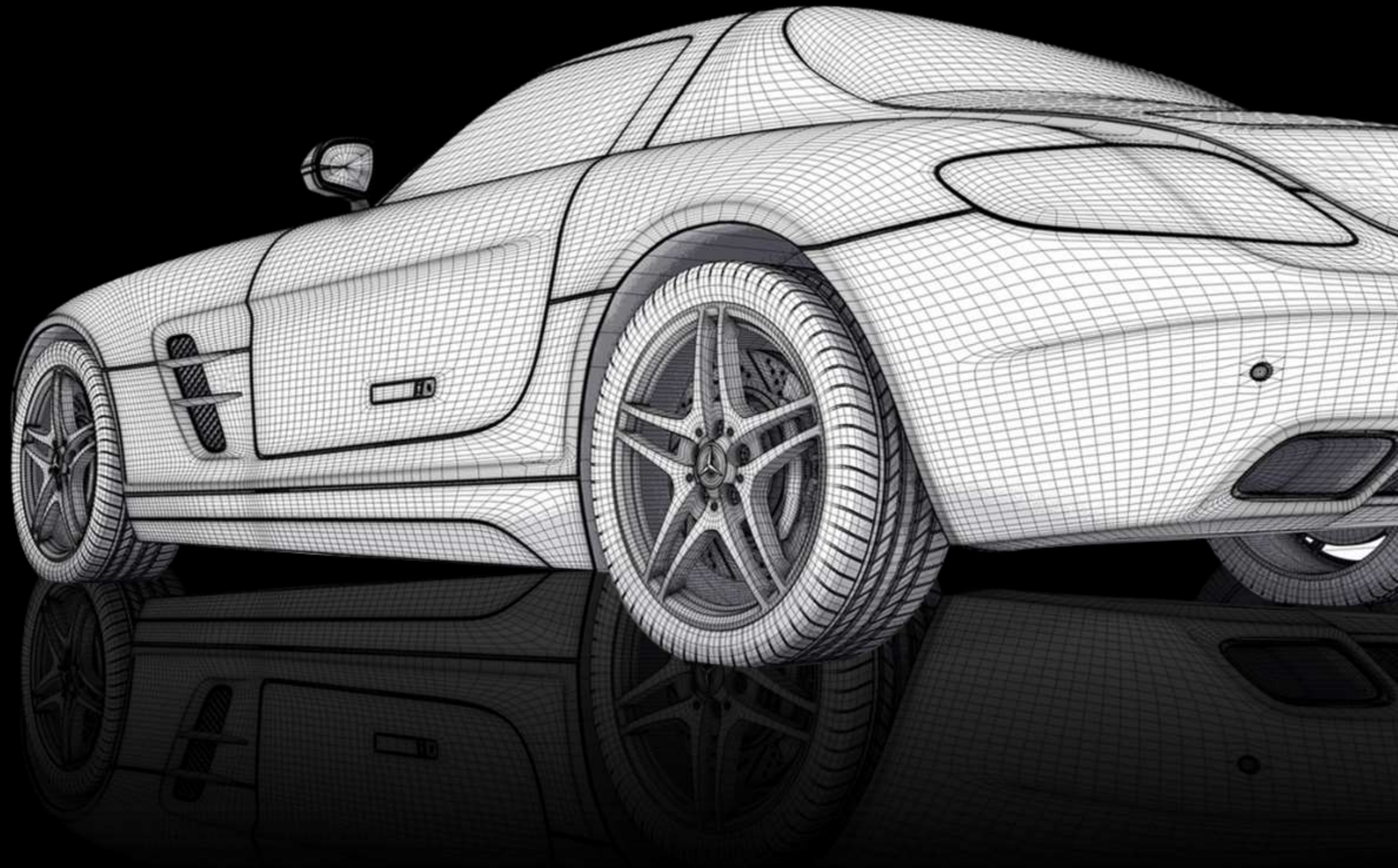
하지만, 차량이 사용된 환경과 조작이 가능한 주행 거리 등처럼 눈으로 쉽게 확인할 수 없는 요소들 또한 가격에 유의미한 영향을 미침을 확인했다.

이러한 부분의 믿을 만한 정보를 제공하는 중고차 딜러에게서 소비자들이 중고차를 믿고 구매할 수 있을 것이다.

3

## 알고리즘 개발

- 다항회귀 모델의 원리
- 산업의 현재동향
- 산업의 잠재시장 파악



# NMAE 함수

```
import numpy as np
from sklearn.metrics import make_scorer

def nmae(true, pred):
    mae = np.mean(np.abs( ((true)) - ((pred)) ))
    score = mae / np.mean(np.abs( ((true)) ))
    return score

NMAE = make_scorer(nmae, greater_is_better=False)

add_metric('nmae2', 'NMAE', nmae, greater_is_better = False )

# best= compare_models(sort='mae', n_select=4)
```

## NMAE

1. 값이 작을수록 잘 예측
2. NMAE는 scale이 다른 데이터셋의 MAE 비교에 용이
3. NMAE는 scale이 다른 데이터셋의 MAE 비교에 용이



	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	NMAE	TT (Sec)
et	Extra Trees Regressor	0.2303	0.1236	0.3446	0.8676	0.0210	0.0150	0.0150	0.909
gbr	Gradient Boosting Regressor	0.2429	0.1221	0.3452	0.8689	0.0209	0.0158	0.0158	0.240
rf	Random Forest Regressor	0.2477	0.1303	0.3559	0.8607	0.0216	0.0161	0.0161	0.834
br	Bayesian Ridge	0.2599	0.1756	0.4108	0.8056	0.0254	0.0169	0.0169	0.053
ridge	Ridge Regression	0.2613	0.1677	0.4034	0.8146	0.0248	0.0170	0.0170	0.019
lightgbm	Light Gradient Boosting Machine	0.2761	0.1513	0.3846	0.8378	0.0233	0.0179	0.0180	0.099
lr	Linear Regression	0.2952	0.2075	0.4498	0.7701	0.0279	0.0192	0.0192	0.340
omp	Orthogonal Matching Pursuit	0.2957	0.2123	0.4566	0.7651	0.0282	0.0192	0.0192	0.022
dt	Decision Tree Regressor	0.3195	0.2249	0.4709	0.7560	0.0286	0.0208	0.0208	0.026
ada	AdaBoost Regressor	0.3652	0.2225	0.4695	0.7599	0.0284	0.0238	0.0237	0.194
en	Elastic Net	0.4772	0.4035	0.6315	0.5517	0.0389	0.0310	0.0310	0.021
lasso	Lasso Regression	0.5045	0.4341	0.6558	0.5206	0.0402	0.0328	0.0328	0.022
huber	Huber Regressor	0.5952	0.6099	0.7771	0.3281	0.0479	0.0387	0.0387	0.074
knn	K Neighbors Regressor	0.6315	0.6678	0.8134	0.2787	0.0496	0.0412	0.0411	0.076
llar	Lasso Least Angle Regression	0.7582	0.9340	0.9641	-0.0097	0.0586	0.0493	0.0493	0.316
dummy	Dummy Regressor	0.7582	0.9340	0.9641	-0.0097	0.0586	0.0493	0.0493	0.016
par	Passive Aggressive Regressor	1.4113	7.8812	2.1024	-7.0254	0.1669	0.0931	0.0916	0.024

```

1 lgb = create_model('lightgbm')
2 lgb = create_model('rf')
3
4 # top5 = [rank for rank in best]
5 # top5_tune = [tune_model(i) for i in top5]

```

	MAE	MSE	RMSE	R2	RMSLE	MAPE	NMAE
Fold							
0	0.2711	0.1617	0.4021	0.8184	0.0245	0.0177	0.0177
1	0.2366	0.1078	0.3284	0.9005	0.0202	0.0154	0.0154
2	0.2320	0.0868	0.2946	0.8814	0.0178	0.0150	0.0151
3	0.2624	0.1162	0.3409	0.8584	0.0214	0.0174	0.0172
4	0.2603	0.1266	0.3559	0.8686	0.0222	0.0172	0.0170
5	0.2076	0.0956	0.3091	0.8920	0.0186	0.0134	0.0134
6	0.2164	0.1006	0.3172	0.8978	0.0189	0.0138	0.0140
7	0.2963	0.2002	0.4474	0.7920	0.0267	0.0189	0.0192
8	0.2778	0.2225	0.4717	0.8104	0.0276	0.0178	0.0180
9	0.2166	0.0853	0.2920	0.8874	0.0180	0.0142	0.0141
Mean	0.2477	0.1303	0.3559	0.8607	0.0216	0.0161	0.0161
Std	0.0285	0.0459	0.0604	0.0377	0.0034	0.0018	0.0019

```

1  # 앙상블
2  blender = blend_models(estimator_list=compare_models(n_select=5,
3                                                         sort='nmae2'))
4
5  # blended = blend_models(estimator_list = best,
6  #                          fold = 10,
7  #                          #method = 'soft',
8  #                          optimize='mae',
9  #                          )

```

	MAE	MSE	RMSE	R2	RMSLE	MAPE	NMAE	
Fold								
0	0.2409	0.1309	0.3618	0.8529	0.0220	0.0157	0.0157	
1	0.1902	0.0603	0.2455	0.9444	0.0149	0.0124	0.0123	
2	0.2124	0.0764	0.2764	0.8956	0.0167	0.0138	0.0138	
3	0.2545	0.1159	0.3405	0.8588	0.0212	0.0169	0.0167	
4	0.2404	0.1195	0.3457	0.8760	0.0220	0.0158	0.0157	
5	0.1841	0.1096	0.3310	0.8762	0.0200	0.0119	0.0119	
6	0.1731	0.0774	0.2782	0.9214	0.0169	0.0112	0.0112	
7	0.2658	0.1615	0.4018	0.8322	0.0239	0.0171	0.0172	
8	0.2471	0.1978	0.4448	0.8315	0.0258	0.0158	0.0160	
9	0.2105	0.0880	0.2967	0.8838	0.0184	0.0138	0.0137	
Mean	0.2219	0.1137	0.3322	0.8773	0.0202	0.0144	0.0144	
Std	0.0306	0.0399	0.0579	0.0345	0.0033	0.0020	0.0020	

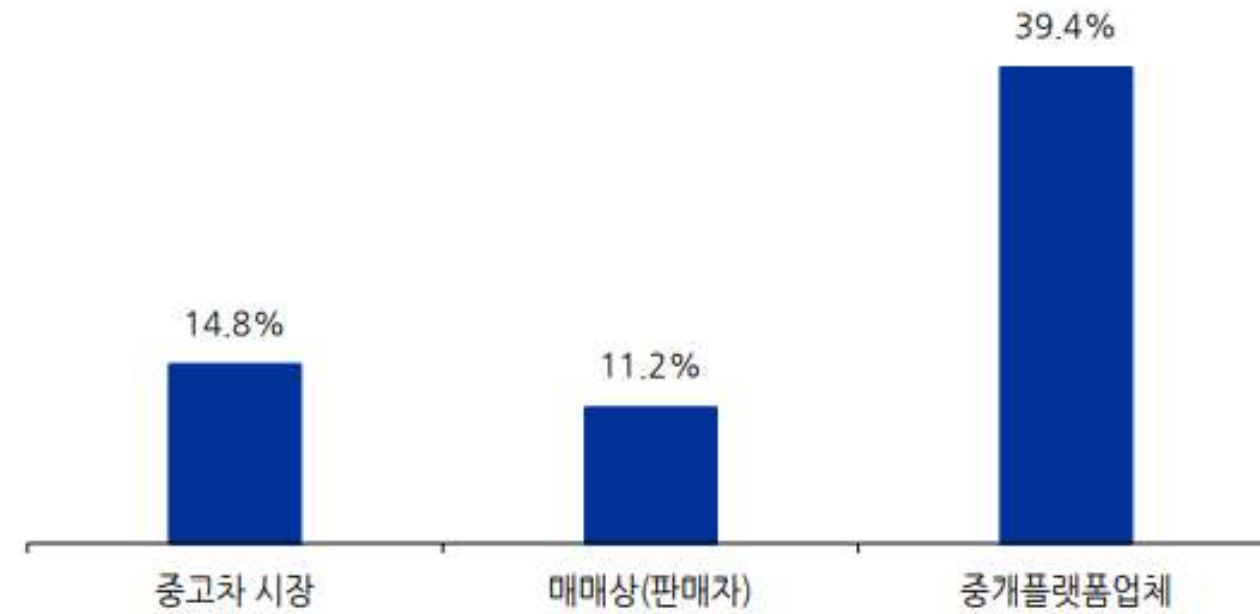
# 앙상블 학습(Ensaemble Learning)  
# 여러 개의 분류기를 생성하고,  
#그 예측을 결합함으로써 보다  
정확한 예측을 도출하는  
기법입니다.

n\_select : Accuracy 기준으로 성능이  
좋은  
5개의 모델만 top5 변수에  
입력합니다.



# 산업의 현재 동향

중고차 시장 소비자 신뢰도 조사



자료: 한국소비자연맹, 언론 보도,

현대 자동차, 기아 자동차, 현대 글로비스 등 자동차 대기업들과 그룹사들의 중고차 시장 진출이 예정되어 있고, 중개 플랫폼 업체의 신뢰도가 매우 높은 편이다.

현대차 중고차 소매 시장 진출 계획 세부 내역

시장 점유율	<ul style="list-style-type: none"> <li>3년간 단계적 진입</li> <li>22년 2.5%, 23년 3.6%, 24년 5.1%</li> </ul>
대상 차량	<ul style="list-style-type: none"> <li>5년 이하, 10km 이하 중고차, 인증 중고차 방식</li> <li>200개 이상 항목에 대한 품질검사 통과한 차량에 한정</li> </ul>
점유율 기준	<ul style="list-style-type: none"> <li>중고차 총거래대수, 사업자거래대수 산술평균</li> </ul>
매입 방식	<ul style="list-style-type: none"> <li>매입 대상 차량 제한 없음</li> <li>인증 중고차 이외 차량은 경매를 통해 판매</li> <li>신차 구입 시 보상판매(트레이드 인) 제도 도입</li> </ul>
판매 방식	<ul style="list-style-type: none"> <li>가상공간에 전시장 구축</li> <li>가상현실(VR) 및 초고화질 이미지를 활용한 중고차 상태 확인</li> <li>모바일/온라인 중심의 판매 플랫폼 구축</li> </ul>
중고차 정보 포털	<ul style="list-style-type: none"> <li>중고차 성능/상태 정보, 가격 산정, 중고차 가격 지수, 실거래 통계, 시세 통계, 모델별 판매 순위 등 제공</li> </ul>

자료: 현대차, 언론 보도

모바일 앱 기반의 중고차 거래 플랫폼 활성화



자료: 각 사, 유진투자증권



# 산업의 잠재 시장 파악



## ■ 국내 중고차 수출산업 육성전략

구분	내용
현재 산업 규모	연간 35만대
육성전략	1 관세 담당 사무소 마련
	2 질 좋은 중고차 매입
	3 해외 빅 바이어 유치
	4 인센티브·성능평가·품질보증
	5 백화점식 현대화 시설
	6 전문 소프트웨어 구축
잠재가치	연간 70만대 규모, 중고부품 산업 성장, 일자리 창출

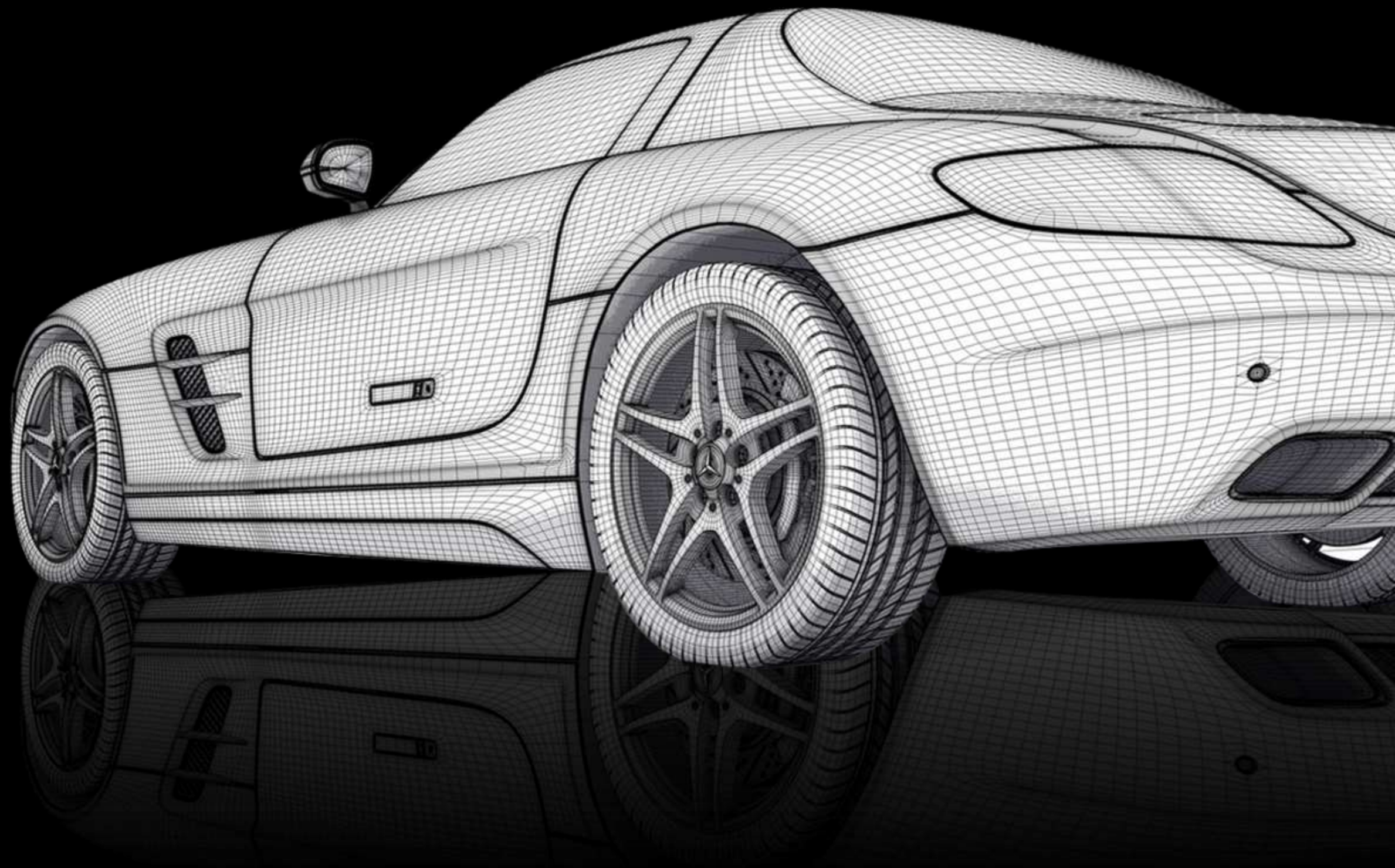




4

## 프로세스 결과

- 활용 방안 / 기대 효과
- 개선 방향





# 활용 방안 / 기대 효과



## 국산 인기모델 잔존가치 분석



캐롯과 함께하는

**Jump Warranty**  
중고차 보증서비스  
지금 만나보세요!



재제조

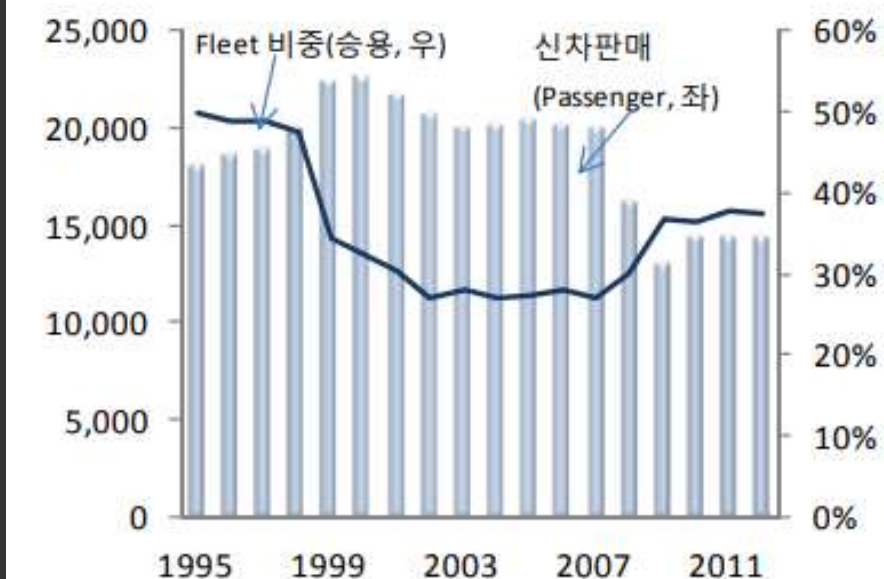
트랜스미션(변속기) (재제조)  
쌍용자동차 코란도 스포츠 6단, 2륜



재제조

트랜스미션(변속기) (재제조)  
쌍용자동차 코란도C

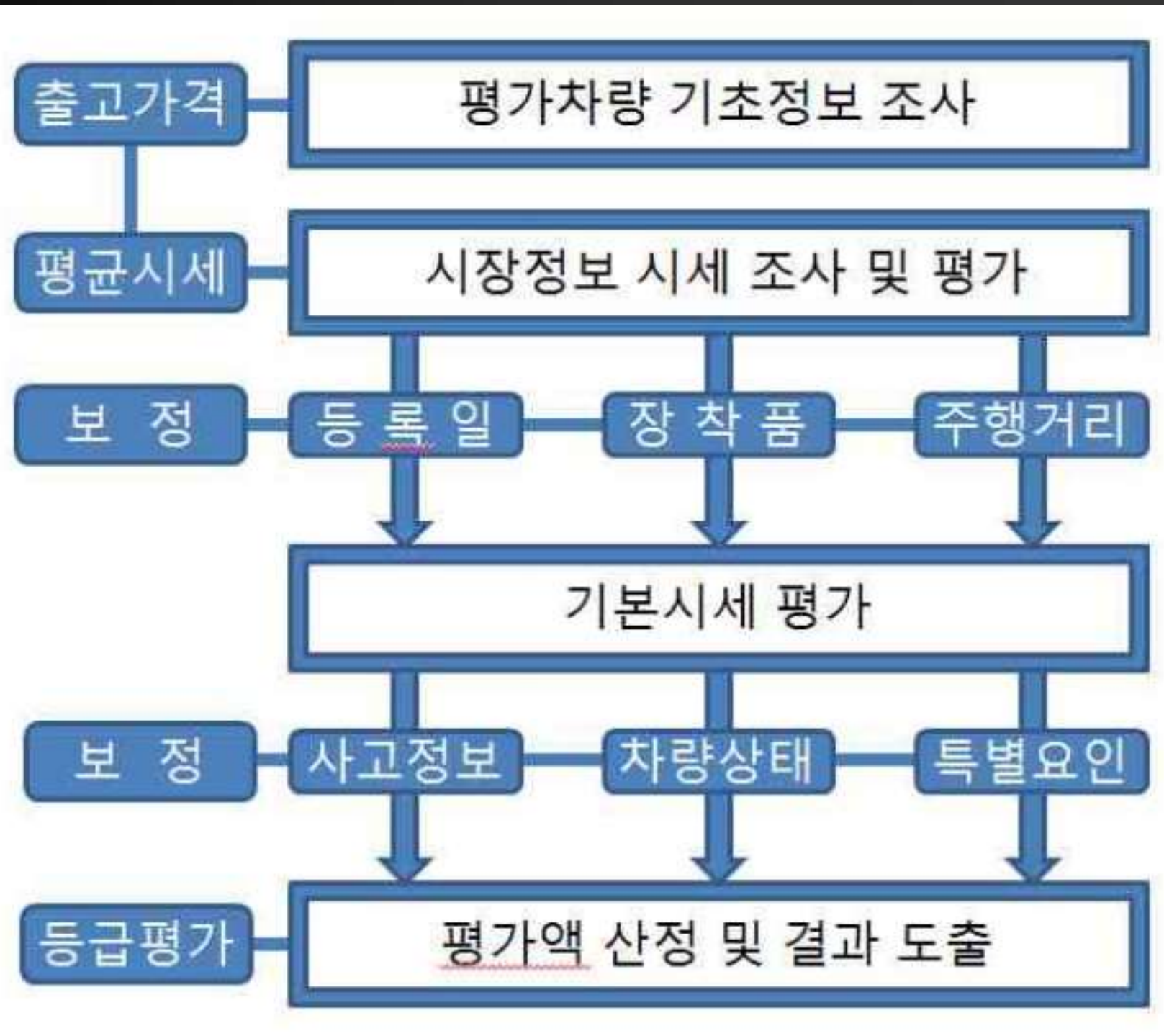
<그림 10> 미국, 신차에서 판매에서 fleet 비중 (단위:천대)



자료: NTS, BS 투자증권



# 개선 방향



평가 시스템의 신뢰성 향상을 위해서는 중고차의 거래실태와 가격결정에 영향을 미치는 요인과 특성에 대한 체계적인 분석과 이를 반영한 개선 노력이 지속되어야 할 것으로 판단된다.

# 출처

윤대권, 이해택, 김용현,  
남일우, 김흥곤, 김주동,  
윤재곤, 하성용  
/ 중고자동차 가격산정 평가  
시스템 개발

유진투자증권 / 국내 중고차  
시장의 잠재력 알아보기

[https://www.eugenefn.com/common/files/amil//20220406\\_B2510\\_leejaeil\\_877.pdf](https://www.eugenefn.com/common/files/amil//20220406_B2510_leejaeil_877.pdf)

BS 투자증권 / 중고차

[http://www.bnkfn.co.kr/uploads/11911/1/%ED%95%A9%EB%B3%B8\\_1%EB%B2%88\\_%EC%82%B0%EC%97%85.pdf](http://www.bnkfn.co.kr/uploads/11911/1/%ED%95%A9%EB%B3%B8_1%EB%B2%88_%EC%82%B0%EC%97%85.pdf)

뉴스 기사

<https://www.sedaily.com/NewsView/263HNPY50D>

<https://www.famtimes.co.kr/news/articleView.html?idxno=62157>