

중고차 가격예측 모델을 이용한 중고차 사이트 추천



2조 이승우 서준형 심현석 이동욱

목차

1. 분석목적

2. 데이터 구축

- 데이터 전처리
- 데이터 설명

3. 데이터 분석 시나리오

- Data step
- Analysis step

4. 추후 계획



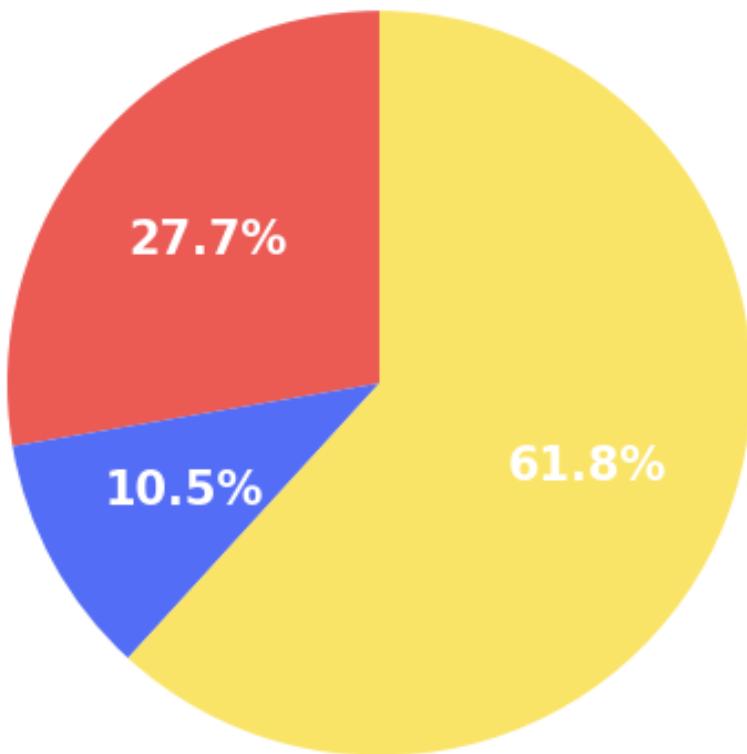
1. 분석 목적

중고차 사이트 데이터를 통해 중고차 가격 예측 모델 구축

대상 : 중고차 판매자

예측된 중고차 가격을 바탕으로
판매자가 중고차 거래 사이트 중
어떤 사이트에 판매하면 가장 이득을 볼 수 있을지 예측

2. 데이터 구축



총 수집 데이터 : 16082

예측 변수 - '가격'

설명 변수 - 22개



4589개의 데이터



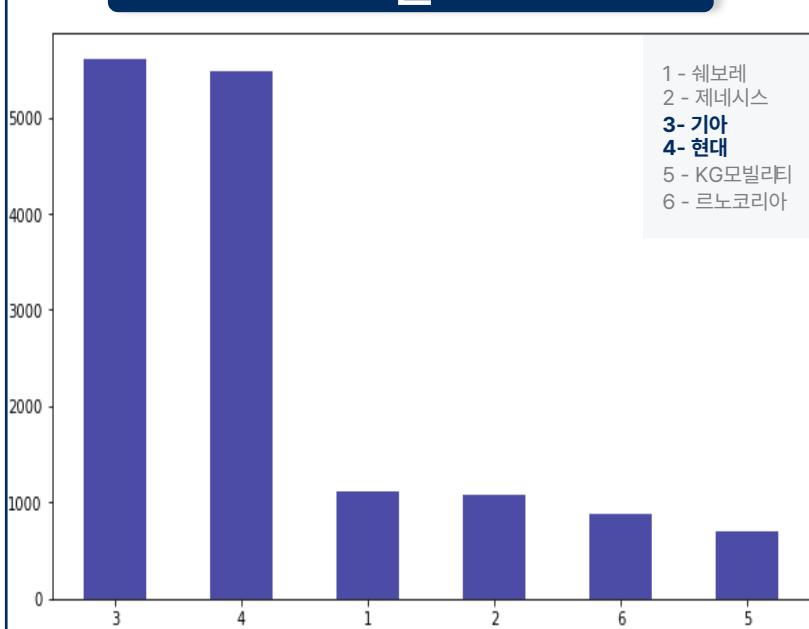
9579개의 데이터



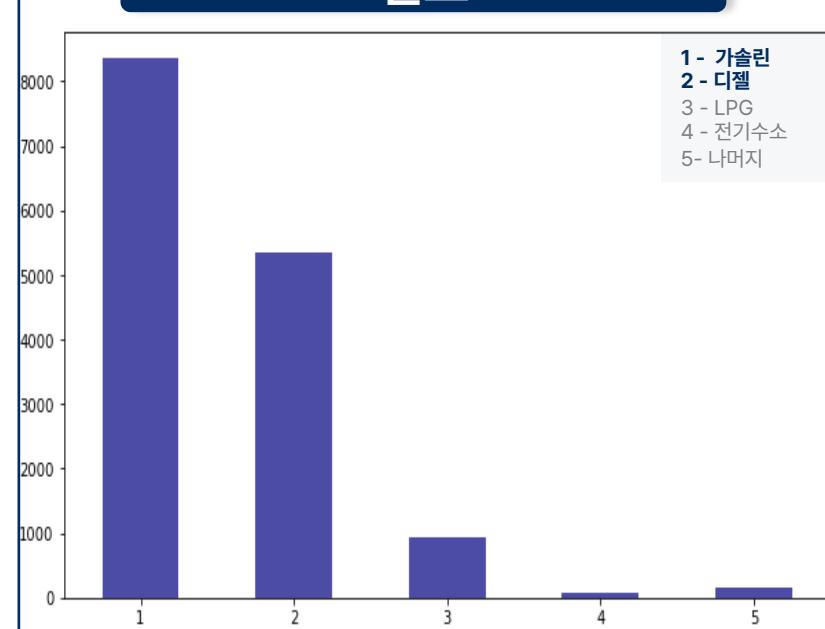
1914개의 데이터

2-1. 데이터 전처리

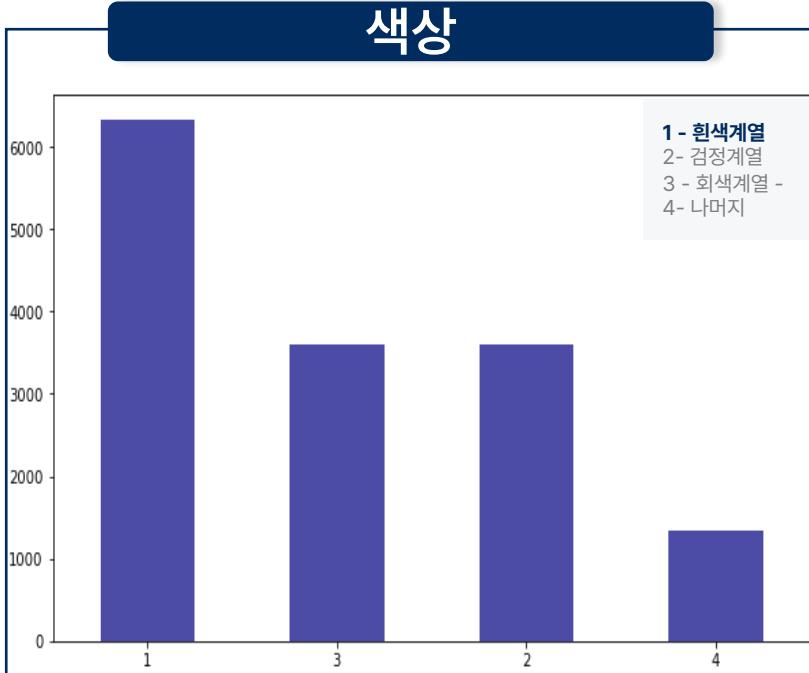
브랜드



연료



색상



연식

누적 개월 수로 변경 (24.05기준)

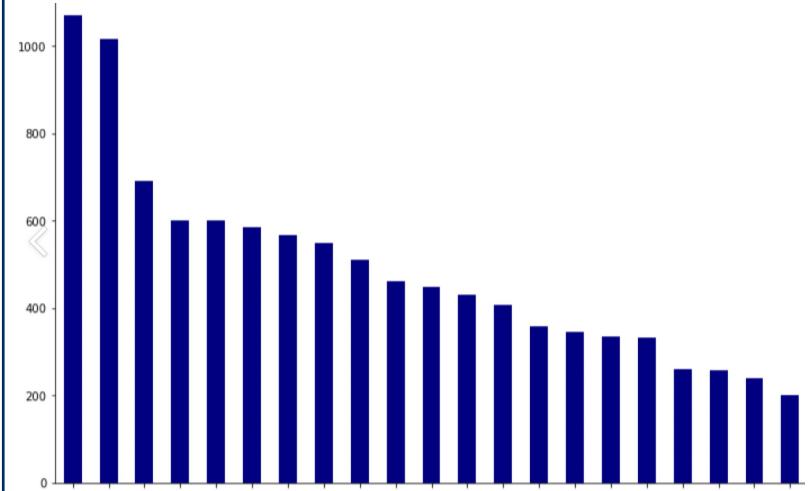
24.01 ➡ 4

23년 12월 ➡ 5

2-1. 데이터 전처리

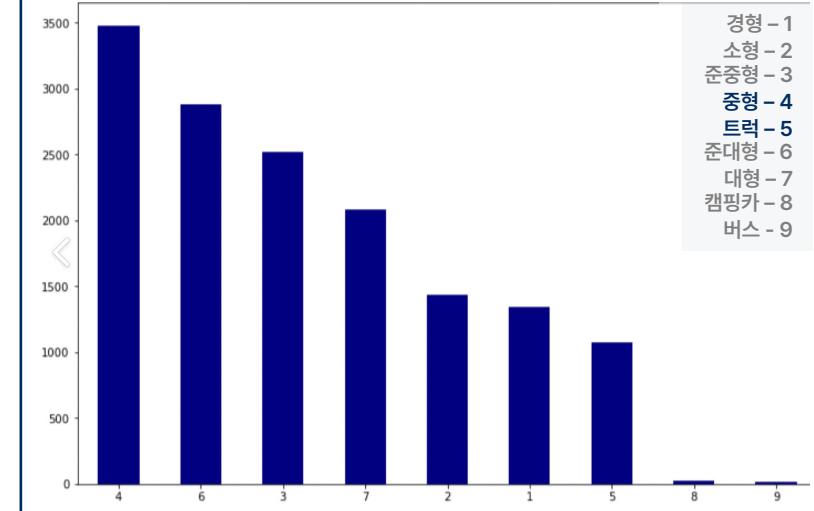
차종

차이름 \Rightarrow 차종



차급

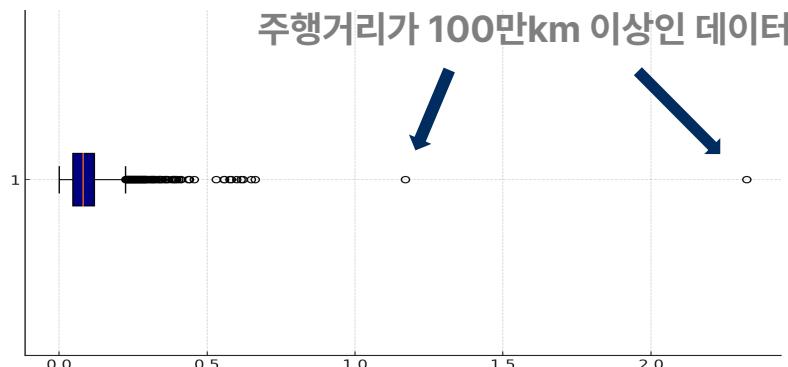
차크기 \Rightarrow 차급



이상치

배기량, 주행거리 이상치 제거

주행거리가 100만km 이상인 데이터



결측치

배기량

같은 차종 평균값

가격

- [계약] \Rightarrow 제거
- 0 \Rightarrow 중고차 시세 중앙값

2-2. 데이터설명

예측변수

가격 = Y

설명변수

수치형 변수

연식 = X_15	주행거리 = X_16	배기량 = X_17	소유자변경횟수 = X_18
-----------	-------------	------------	----------------

범주형 변수

Binary

썬루프 = X_1	내비게이션 = X_2	스마트키 = X_3	오토라이트 = X_4	주차감지센서 = X_5
열선시트 = X_6	통풍시트 = X_7	가죽시트 = X_8	사고이력 = X_9	변속기 = X_10

Label Encoding

차종 = X_11	브랜드 = B	색상 = C	연료 = F	차급 = G
-----------	---------	--------	--------	--------

One-Hot Encoding

사이트_Kcar = X_12	사이트_Bobae = X_13	사이트_Chacha = X_14
-----------------	------------------	-------------------

3. 데이터 분석 시나리오

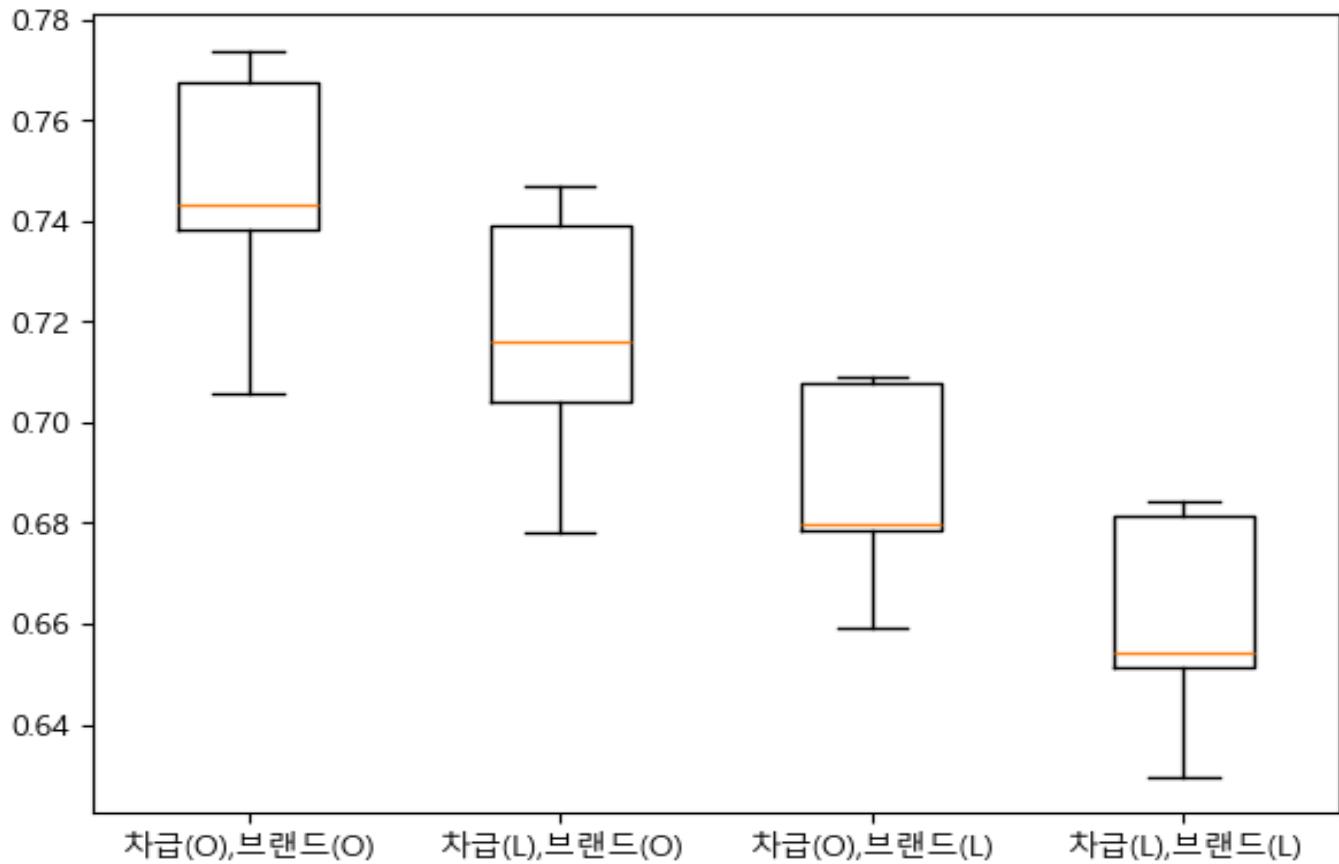
Data Step 베이스 모델 $Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_{18} X_{18i} + \gamma_1 B + \gamma_2 C + \gamma_3 F + \gamma_4 G + \epsilon_i$

시나리오 분류 기준

Label Encoding vs One-Hot Encoding

Case	차급	브랜드	연료	색상	Case	차급	브랜드	연료	색상
I	L	L	L	L	IX	O	L	L	L
II	L	L	L	O	X	O	L	L	O
III	L	L	O	L	XI	O	L	O	L
IV	L	L	O	O	XII	O	L	O	O
V	L	O	L	L	XIII	O	O	L	L
VI	L	O	L	O	XIV	O	O	L	O
VII	L	O	O	L	XV	O	O	O	L
VII	L	O	O	O	XVI	O	O	O	O

Analysis Step



4. 추후 계획

단순 선형회귀 모델 X

추가 회귀모델

Ridge

Lasso

ElasticNet

데이터 스케일링

MinMax

Normalizer

Standard

Robust Scaler

추가 모델 탐색

RandomForest

XGBoost

LSTM

etc



Q & A

2조 이승우 서준형 심현석 이동욱