

# 기계학습의 이해 최종발표

중고차 가격예측 모델을 이용한  
중고차 사이트 추천

2조 이승우 서준형 심현석 이동욱



# 목차

## CONTENTS

01 서론

02 Data Step

03 Analysis Step

04 결론

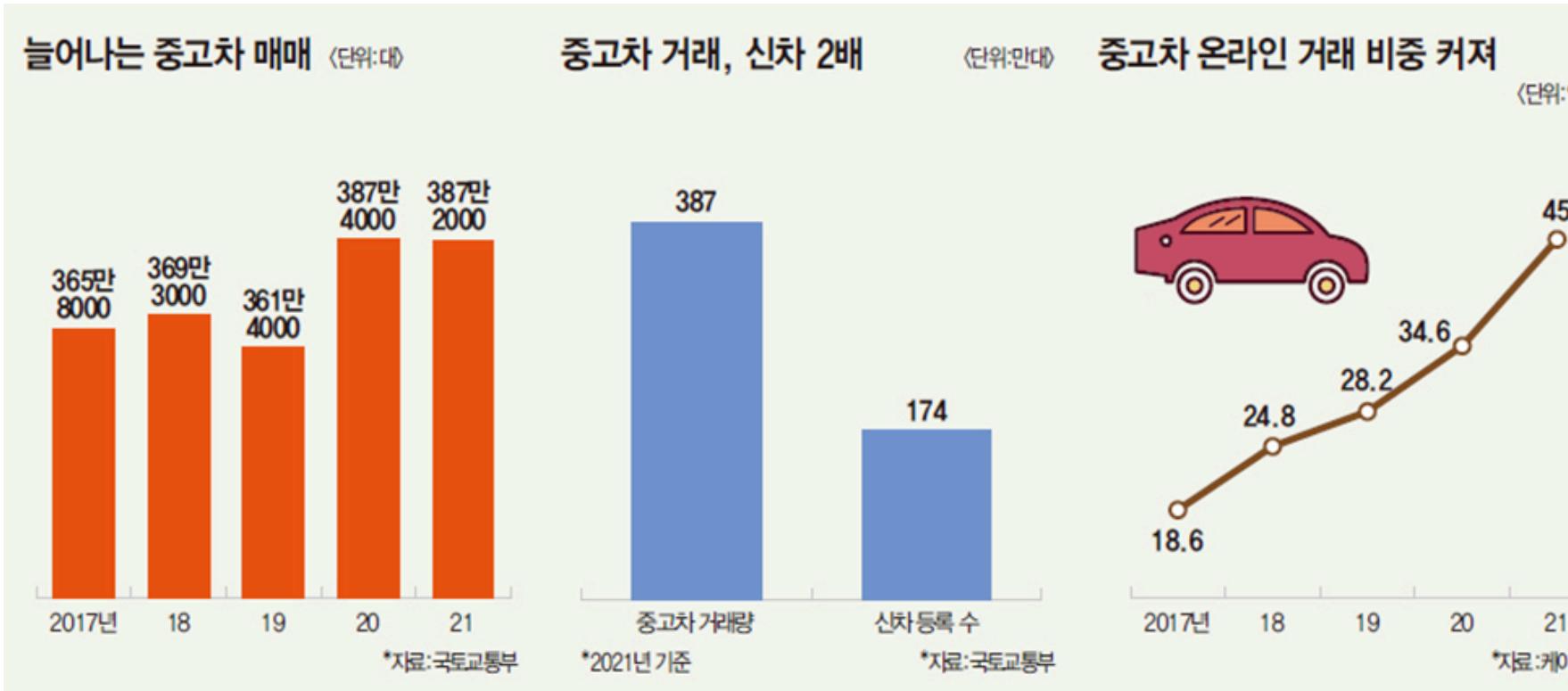


서론

01

# 분석 배경

02



중고차 시장의 확대

온라인 거래 증가

BUT

같은 차량이라도 상이한 가격

기준이 되는 가격 부재

사이트 비교 정보 부족

03

04

01

# 분석 목표

02

Mission

중고차 데이터를 기반으로 가격 예측 모델 구축

03

Vision

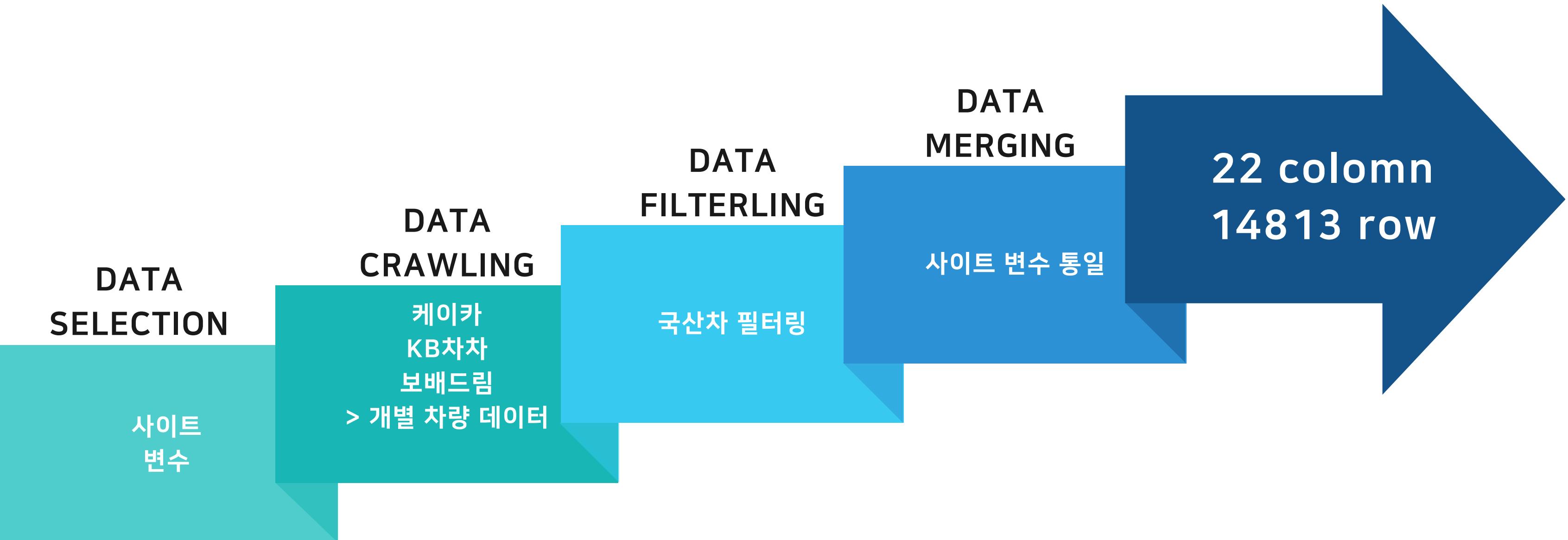
중고차 판매자가 어떤 사이트에 판매하면  
가장 이득을 볼 수 있을지 제안

04



# Data Step

# 데이터 구축 과정



01

# 데이터 설명

02

03

04

예측변수

가격

연속형 변수

주행거리

배기량

이산형 변수

연식

소유자변경횟수

범주형 변수

썬루프  
내비게이션  
스마트키  
⋮

차종  
브랜드  
색상  
연료  
차급

사이트\_KCar  
사이트\_chacha  
사이트\_bobae

옵션

# 주요 전처리 과정

01

결측치 처리

배기량, 가격

02

데이터 변환

연식

03

이상치 처리

배기량, 주행거리

04

데이터 범주화

브랜드, 연료, 색상

파생변수

차급, 차종, 전기수소

01

# 데이터 정제- 차종

02

## 차종

차이름에서부터 차종을 추출

03

### 차이름

그랜저ig 24 프리미엄



### 차종

그랜저

쏘나타 20 Ipi 스마트



쏘나타

아반떼ad 16 Ipi 스마트



아반떼

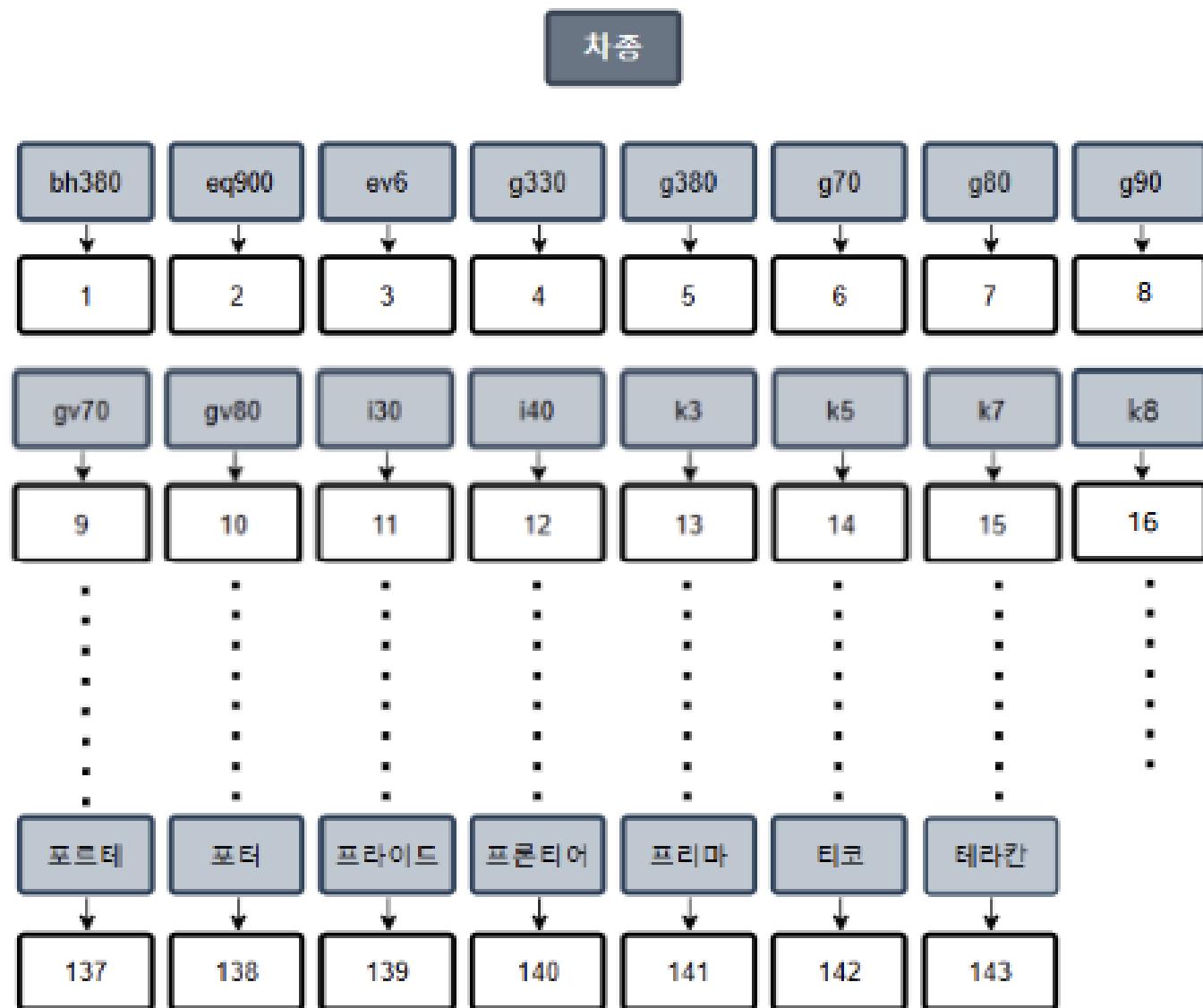
04

총 143 개의 차종 분류

01

# 차종 원핫인코딩

02



03



04

gv70 -> 9 , k3 -> 13

gv80 -> 10 , k5 -> 14

gv70 -> 2-1 , k3 -> 3-1

gv80 -> 2-2 , k5 -> 3-2

01

# 파생변수- 차급

## 차급

02

차량 크기에 대한 정보를 반영하기 위해 차종별로 차급 데이터를 수집하여 추가함

차종	차급	차종	차급	7 개 차급
모닝		쏘나타		
티코	{ 경형 }	말리부		중형
마티즈		산타페		
티볼리	{ 소형 }	임팔라		
트랙스		k9		
qm3		카니발		대형

01

# 파생변수- 전기수소

02

## 전기수소

전기차, 수소차의 경우 배기량이 존재하지 않아 별도의 파생변수로 분리하여 구분

03

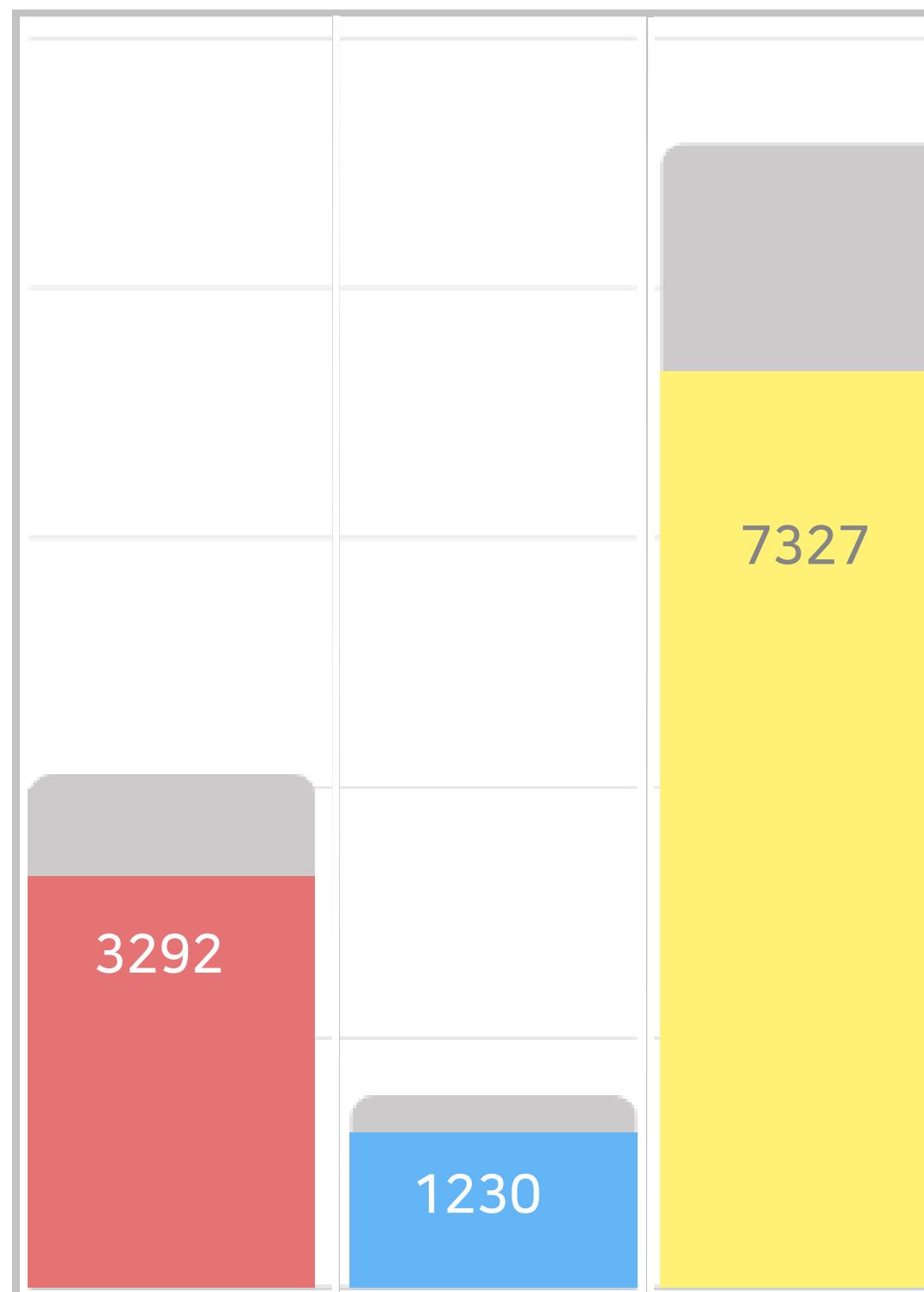
차종	배기량	전기수소
그랜저	2359	0
아반떼	1591	0
포터	0	1
아이오닉	0	1

04



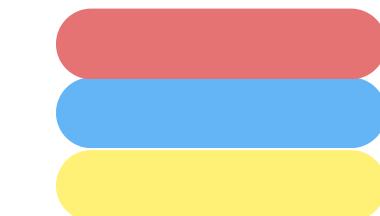
# Analysis Step

# TRAIN TEST 분리



Kcar\_train  
 Bobae\_train  
 Chacha\_train  
 test\_data

총 train data : 11,849개

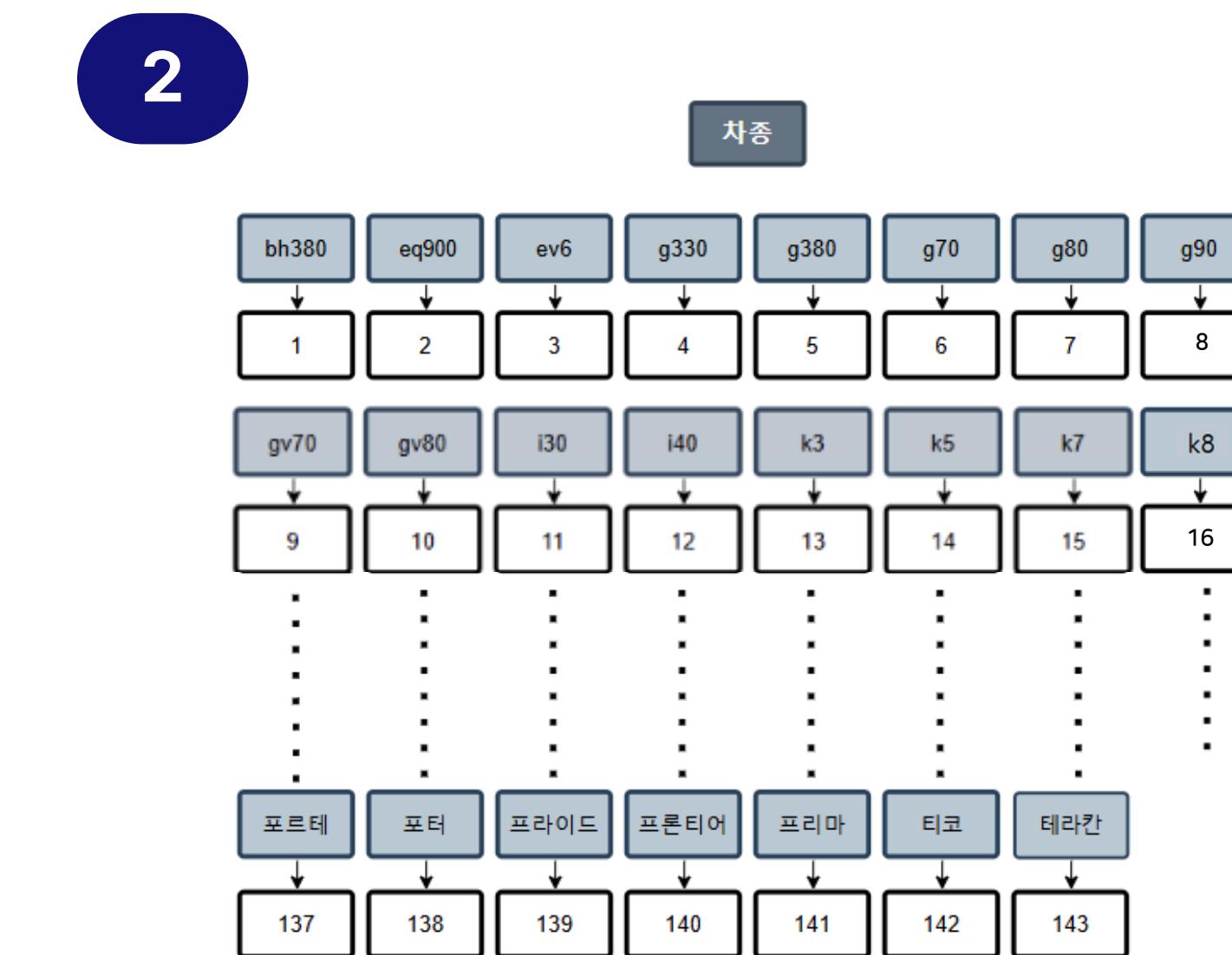


총 test data : 2,964개



# 분석 시나리오

## (1) One-Hot Encoding 방식 비교



X2

01

# 분석 시나리오

02

주행거리

배기량

03

1

MinMax Scaling

2

Standard Scaling

3

MaxAbs Scaling

4

Robust Scaling

04

X2 X4 X4

01

# 분석 시나리오

02

## (3) 사용모델

03

Linear Regression

Ridge

Lasso

SVM

Decision Tree

Random Forest

Gradient Boosting

K-Nearest Neighbors

MLP Neural Network

XGBoost

04

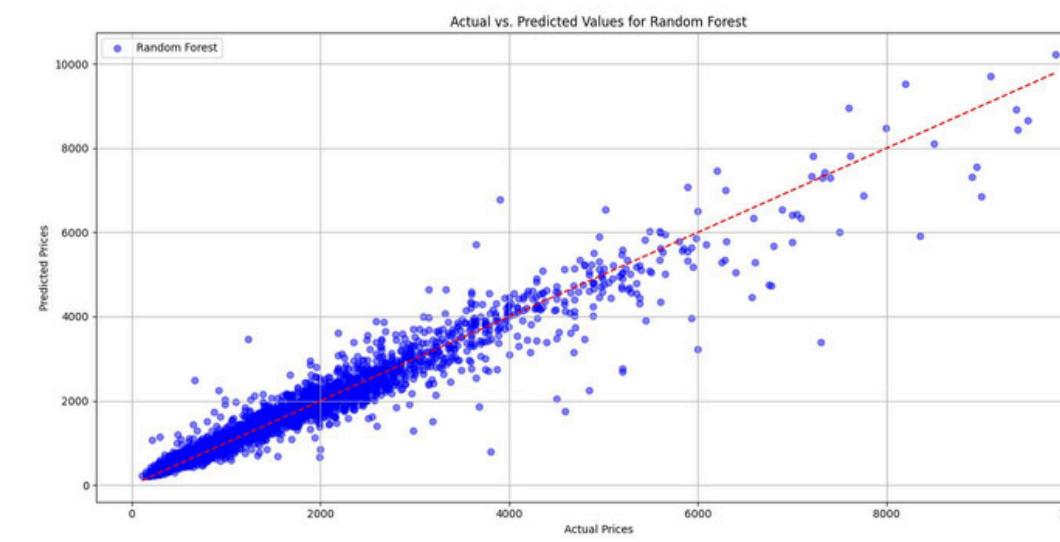
$$X2 \times 4 \times 4 \times 10 = 320$$

# 모델 1차 선정 TOP 50

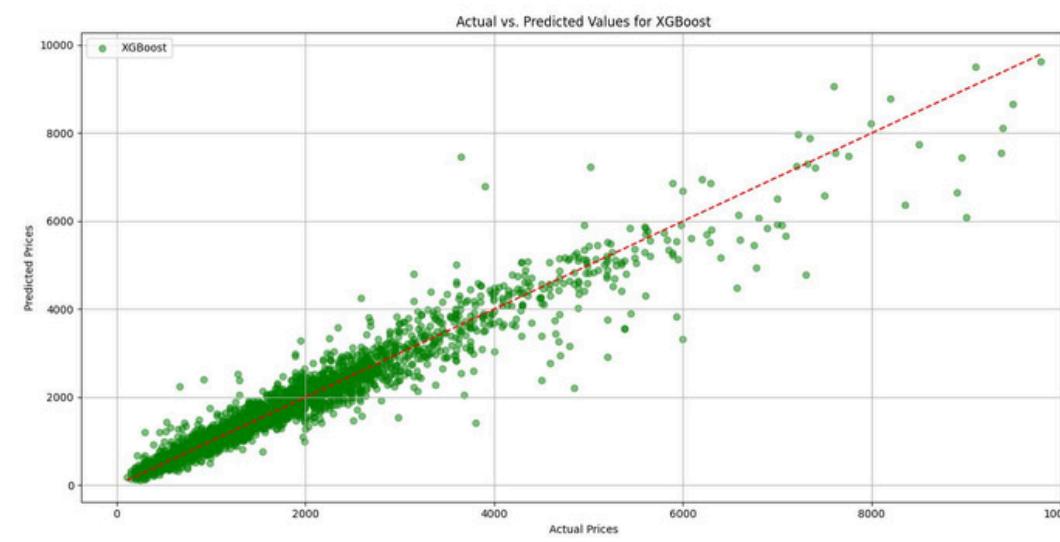
Rank	Model	Best RMSE	Best One-Hot	Best Scaling 1	Best Scaling 2
1-4	Random Forest	320.1373	1	Robust	Standard
5-20	XGBoost	322.8842	2	-	-
21-43	Random Forest	322.9068	1	Standard	Standard
44	MLP Neural Network	331.4931	2	Robust	Robust
45-48	Random Forest	332.1191	2	Standard	MaxAbs
49	MLP Neural Network	333.7729	2	MaxAbs	Standard
50	Random Forest	334.4661	2	Standard	Robust

# 1차 모델 비교

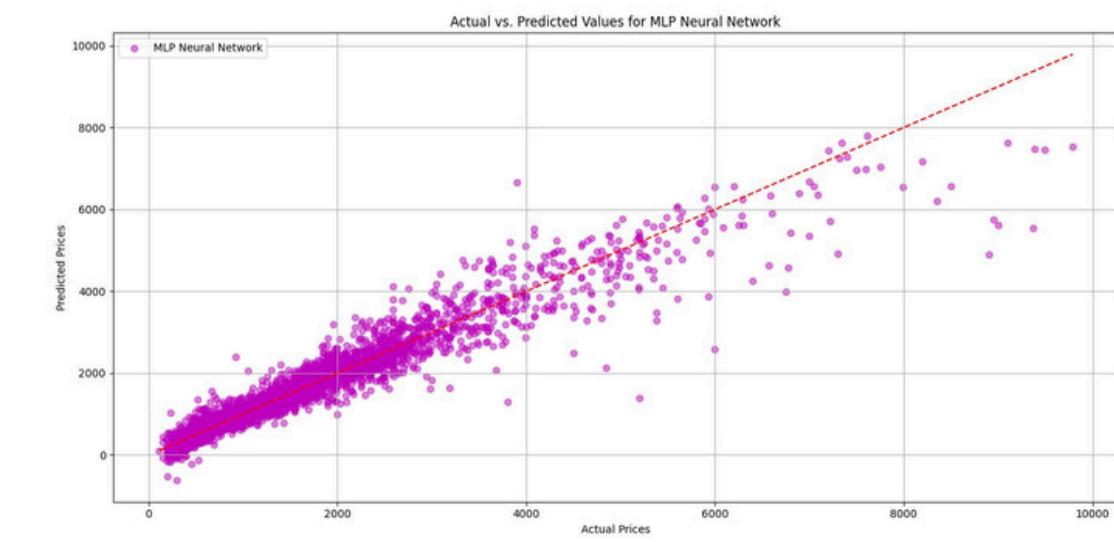
세 가지 모델의 예측 값 vs 실제 값



Random Forest  
RMSE : 320.1373



XGBoost  
RMSE : 322.8842



MLP Neural Network  
RMSE : 331.4931

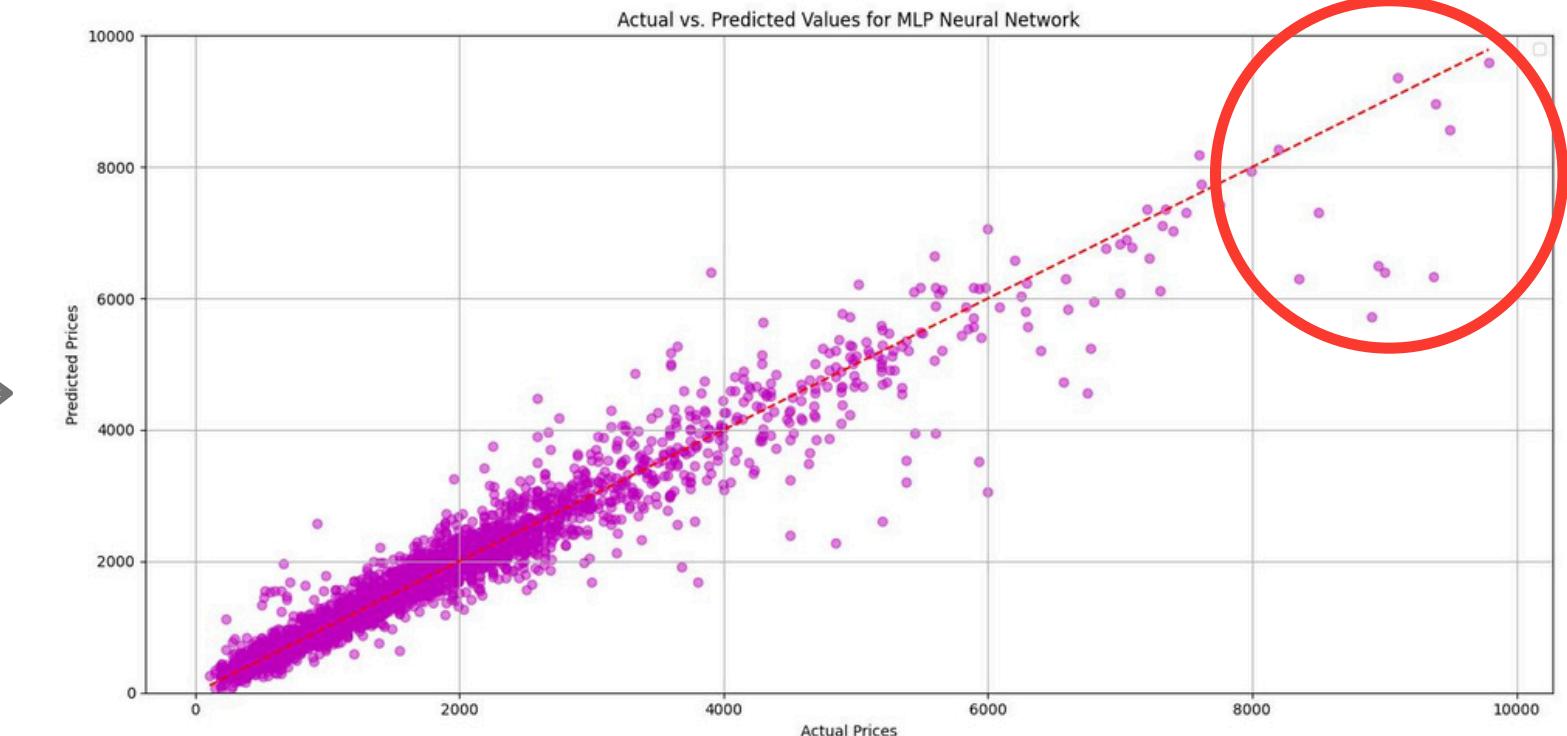
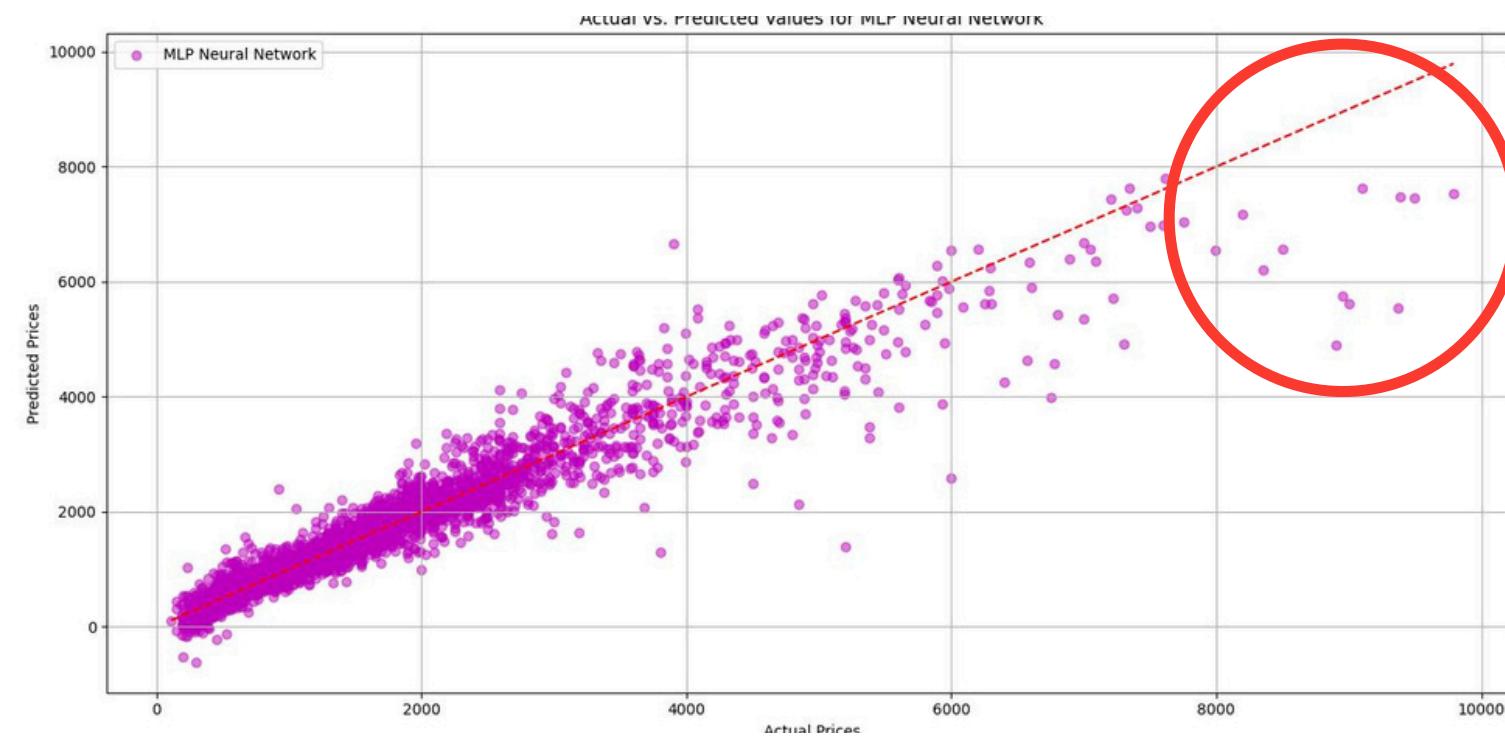
# 모델 2차 선정

01

MLP Neural Network

02

## Gridsearch



RMSE : 331.4931

RMSE : 324.2291

03

04

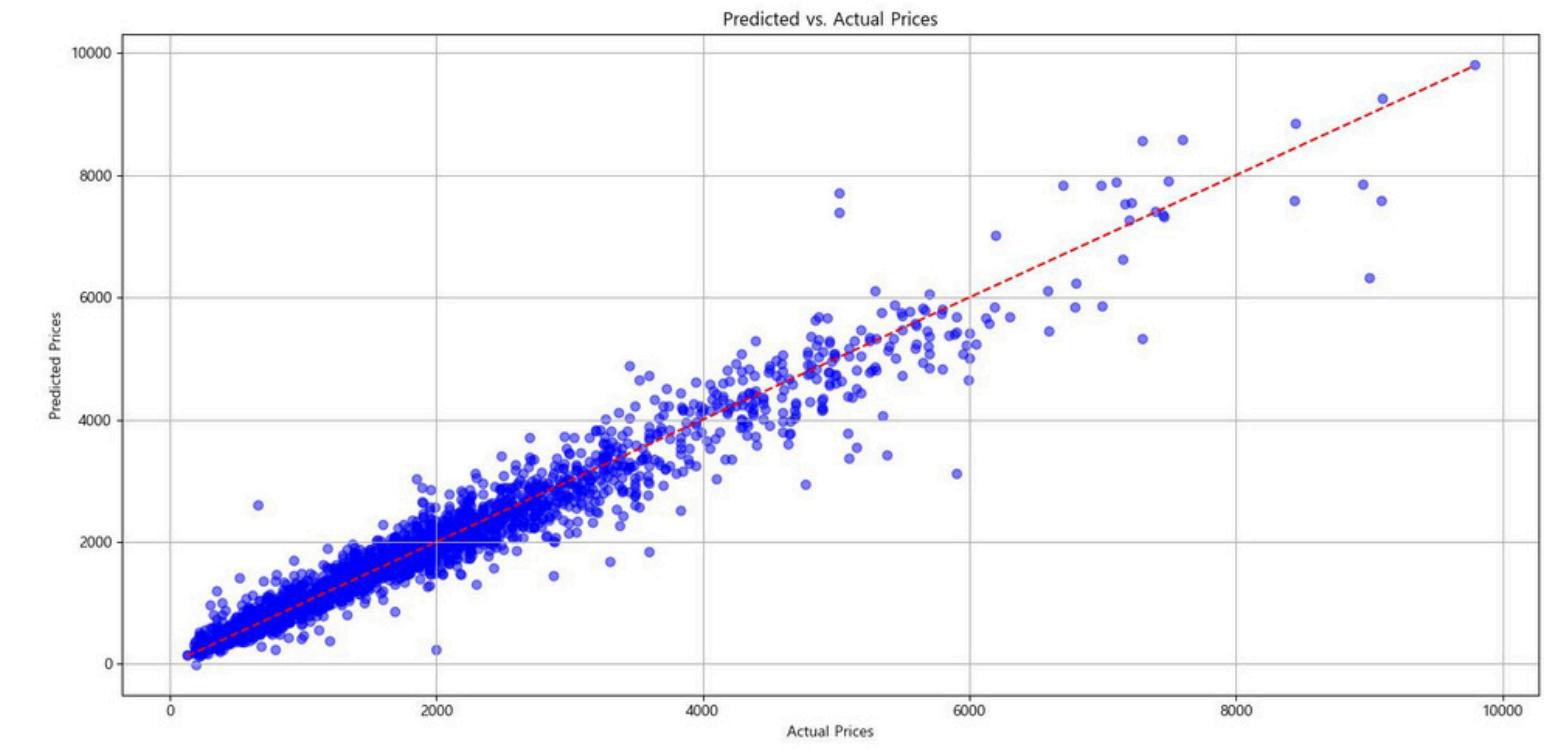
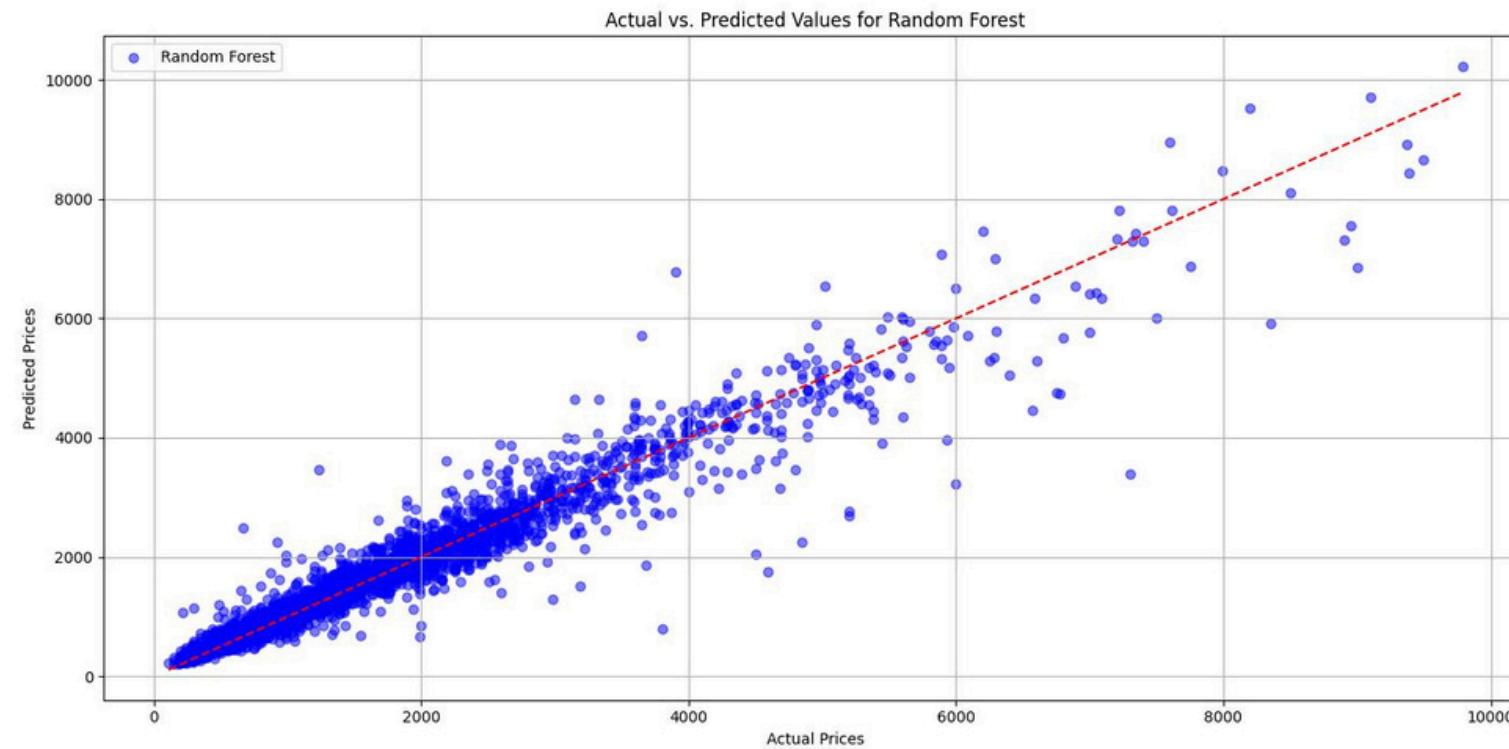
# 모델 2차 선정

01

Random Forest

02

## Gridsearch



RMSE : 320.1373

RMSE : 319.9967

03

04

# 모델 2차 선정

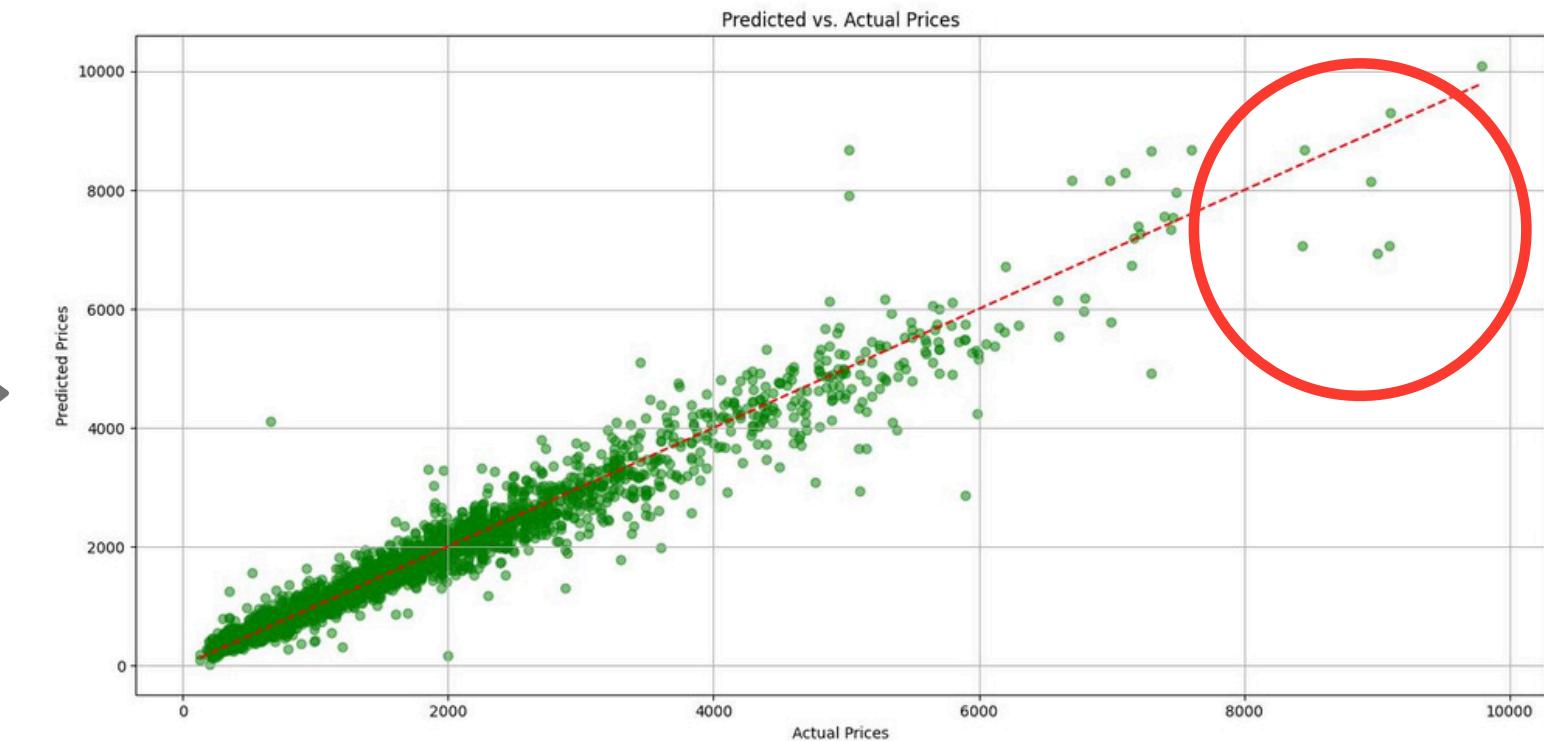
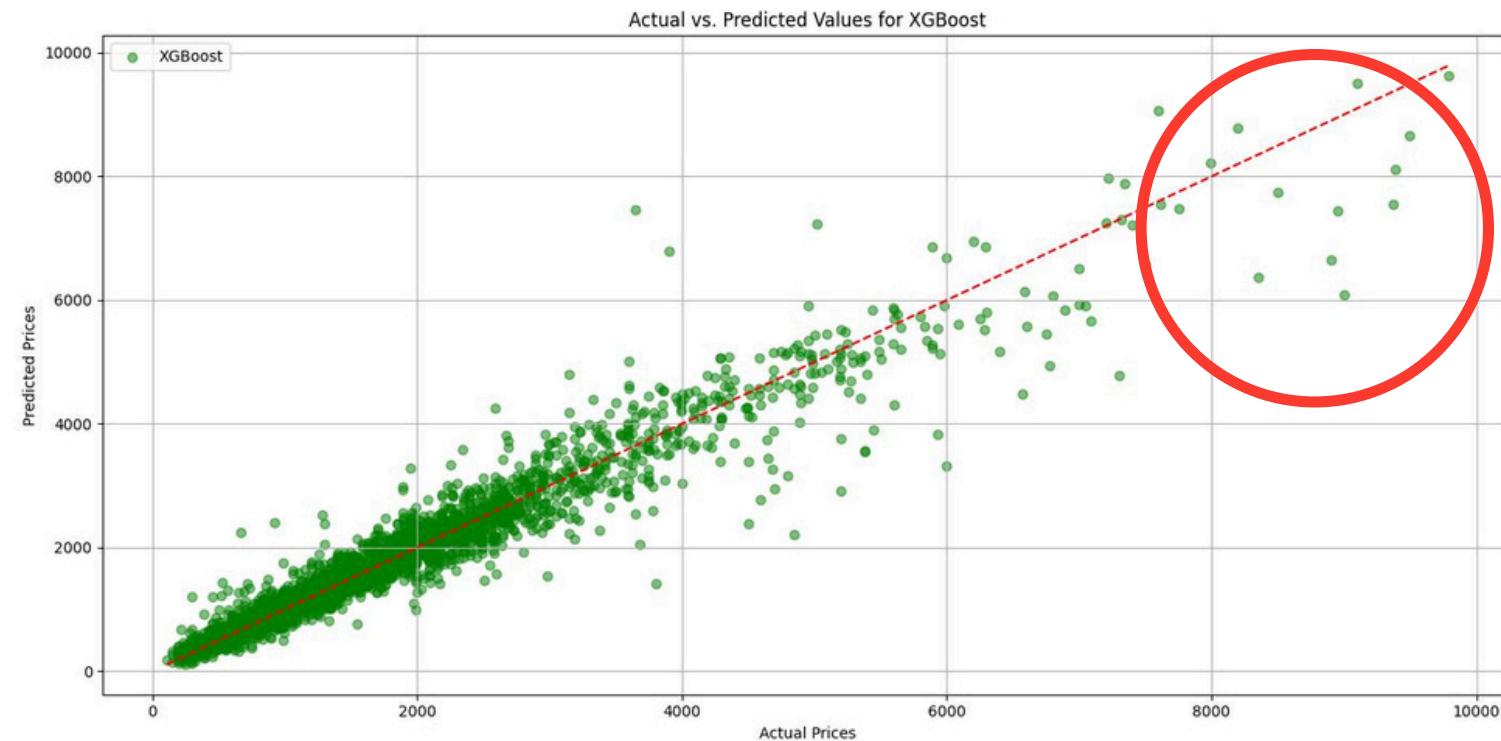
01

XGBoost

02

## Gridsearch

03

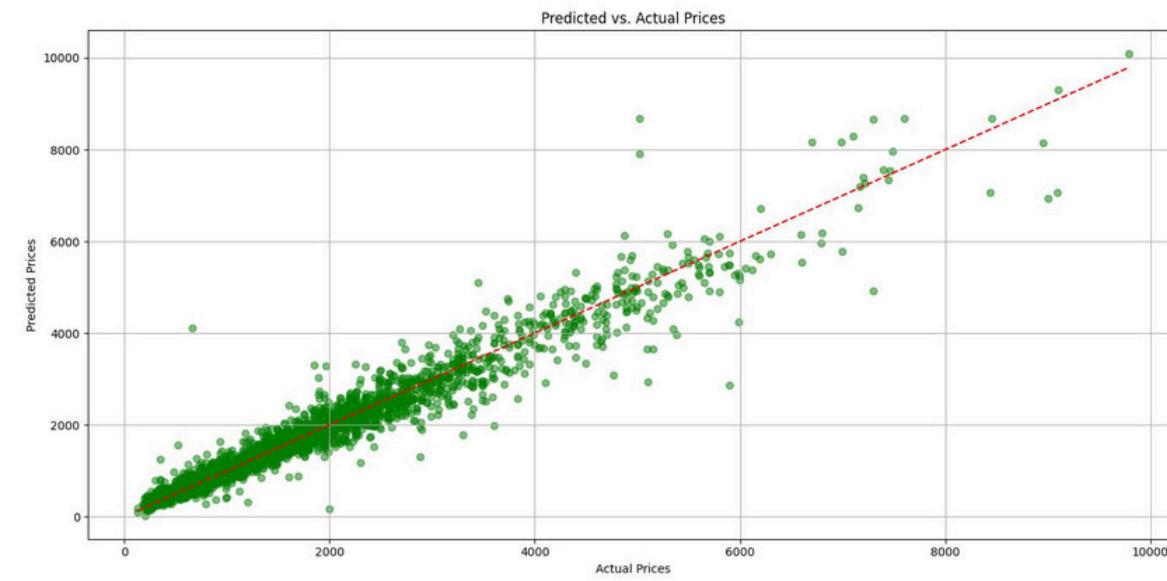


RMSE : 322.8842

RMSE : 313.3642

04

# 모델 최종 선정



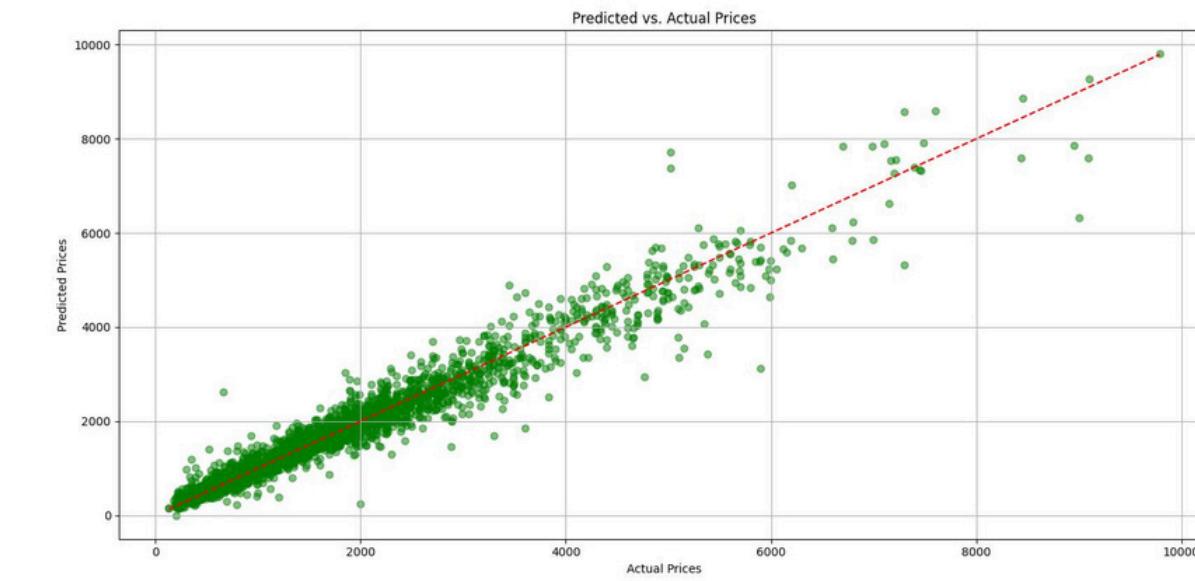
Best RMSE : 313.3642

Best Params

- 'colsample\_bytree':0.7
- 'gamma': 0
- 'learning\_rate':0.1
- 'max\_depth': 7
- 'min\_child\_weight': 1
- 'n\_estimators': 300
- 'subsample': 0.9



Optuna



Best RMSE : 300.8193

Best Params

- 'colsample\_bytree':0.661011
- 'gamma':0.142476
- 'learning\_rate':0.058869
- 'max\_depth':7
- 'min\_child\_weight':2
- 'n\_estimators':294
- 'subsample': 0.888931

The background is a grayscale aerial photograph of a city, showing a dense concentration of buildings, roads, and industrial facilities. The text is overlaid on this image.

결론

01

# 프로젝트 목표

02

Mission

중고차 데이터를 기반으로 가격 예측 모델 구축

03

Vision

중고차 판매자가 어떤 사이트에 판매하면  
가장 이득을 볼 수 있을지 제안

04

01

# SAMPLE 소형

(325구6200)기아 더 뉴 레이 밴  
1.0 가솔린 밴 2인승 기본형

판매가격  
**600만원**  
중고차시세

19년04월(19년형) | 94,951km | 가솔린 | 경기

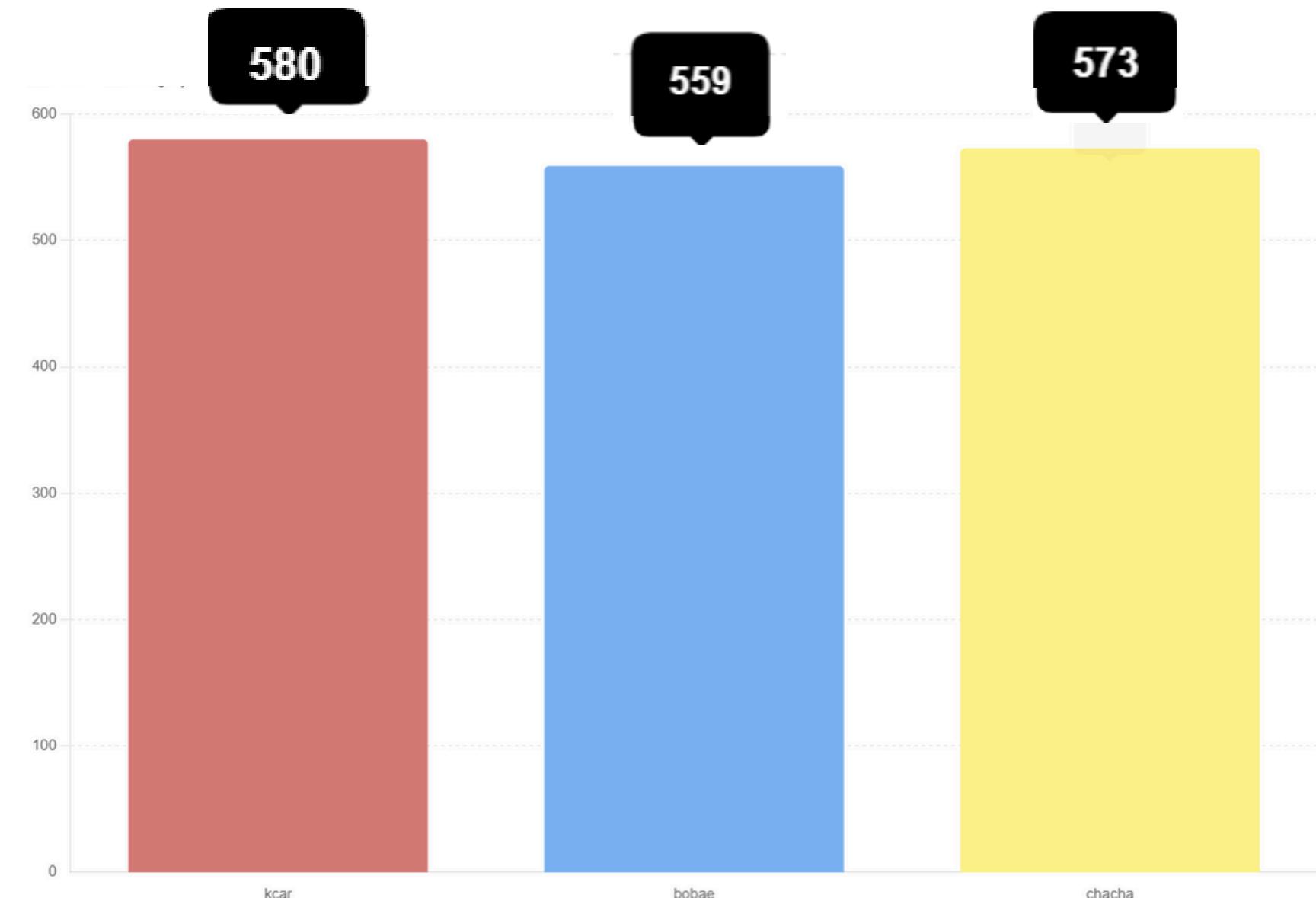


시세정보 ?

시세안전구간

**521~775만원**

- Kcar
- Bobae
- Chacha



02

03

04

01

# SAMPLE 중형

(152구9621)현대 더 뉴 그랜저  
3.3 캘리그래피

판매가격

**2,399만원**

중고차시세

20년04월(20년형) | 118,909km | 가솔린 | 인천

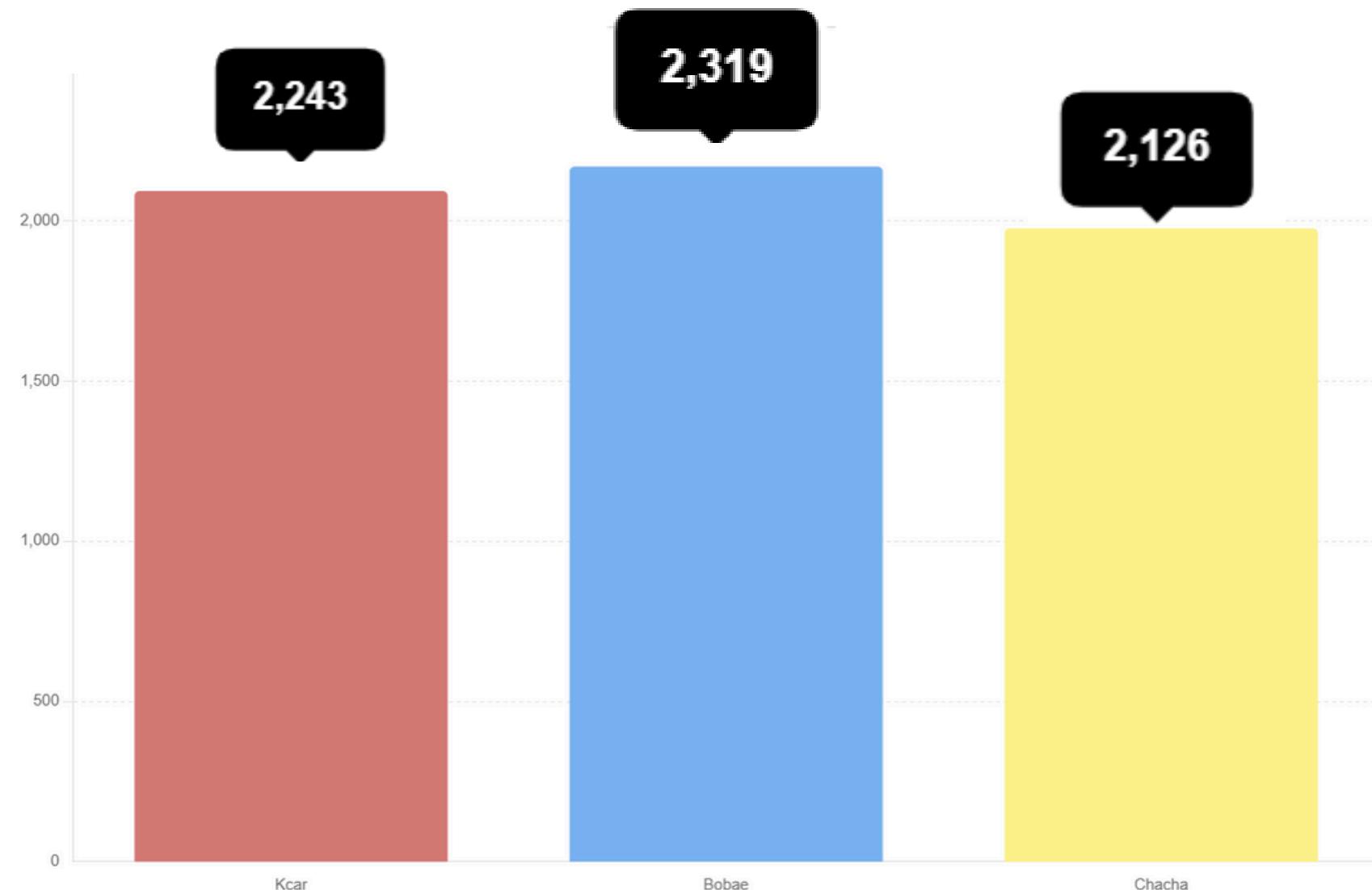


시세정보 ?

시세안전구간

**2,254~2,745만원**

- Kcar
- Bobae
- Chacha



02

03

04

01

# SAMPLE 대형

02

## 기아 더 뉴 카니발 9인승 노블레스 스페셜

174라7070 | 무사고 · 19년 12월식(20년형) · 60,243km · 디젤 · 검정색 · 오토 · 2,430만원

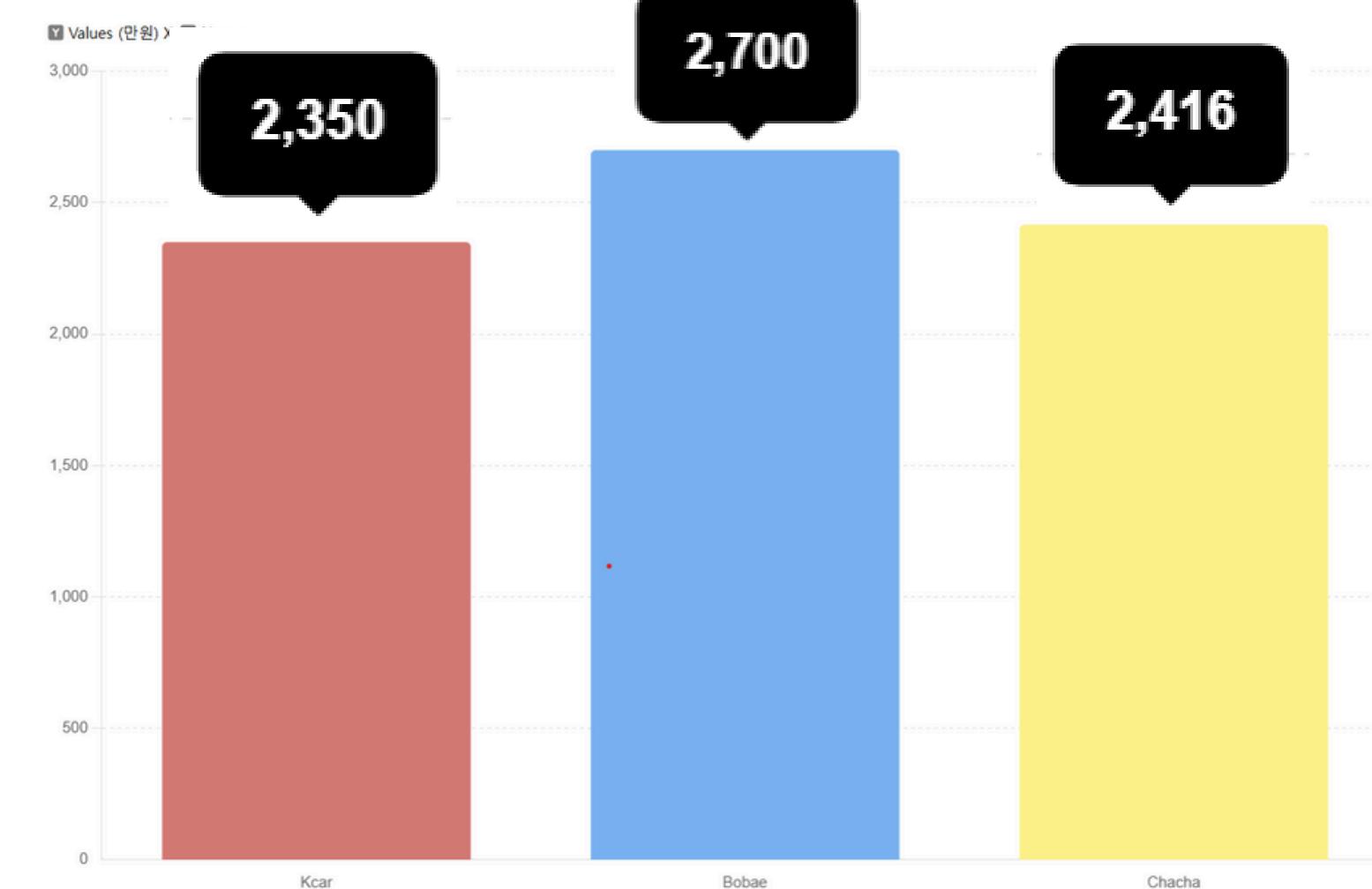


시세정보 ?

**2,430** 만원

시세안전구간

**2,185~2,661**만원



03

04

01

# Summary

가격 예측 모형을 통해 중고차 판매자는 최대 판매수익 기대 가능

02

STEP 1  
차량 정보 입력



주행거리  
배기량  
차종  
차급

STEP 2  
판매가 예측



사이트	예측가
KCar	2,350만원
보배드림	2,700만원
KB차차차	2,416만원

STEP 3  
사이트 추천



보배드림  
추천

03

04

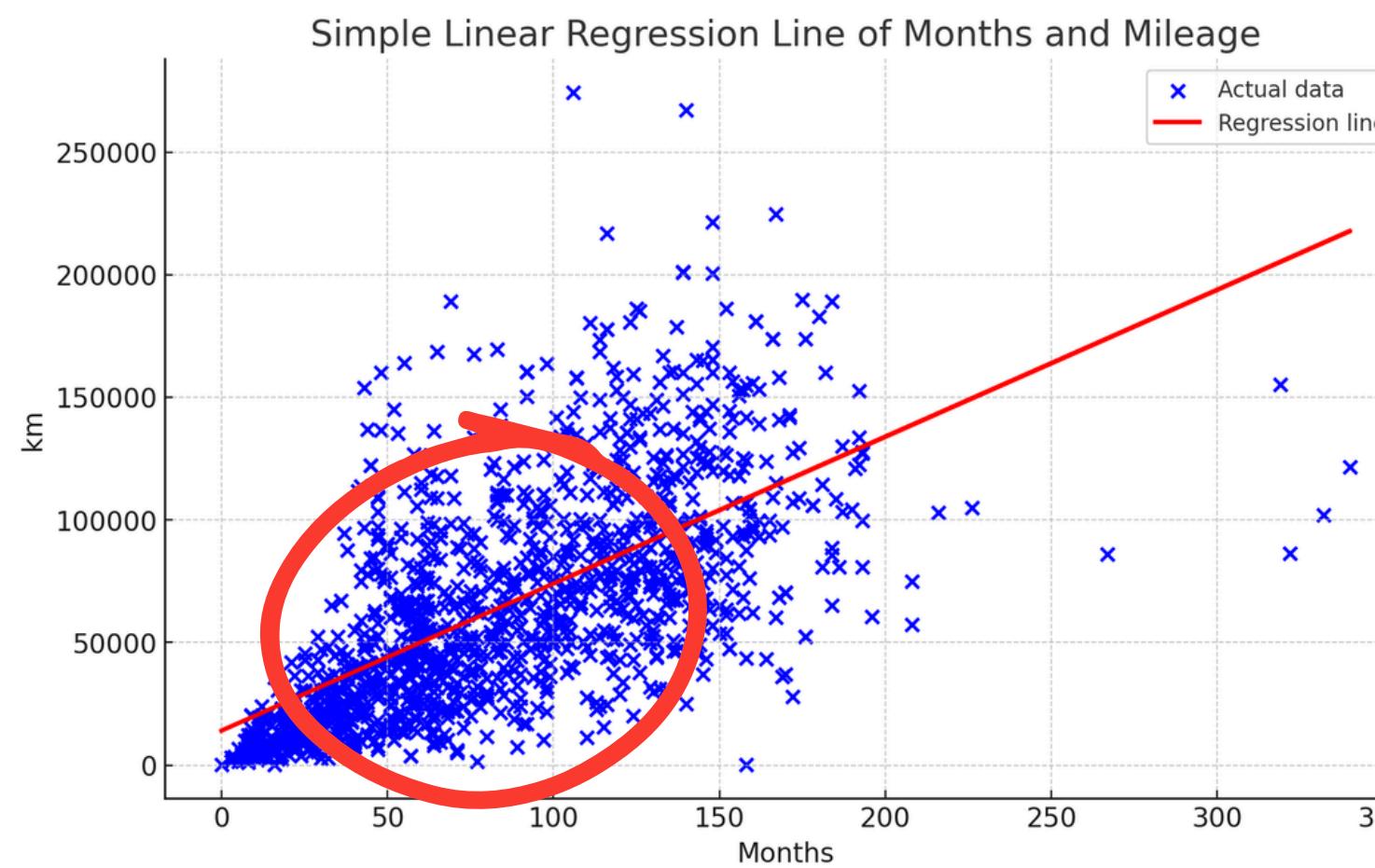


01

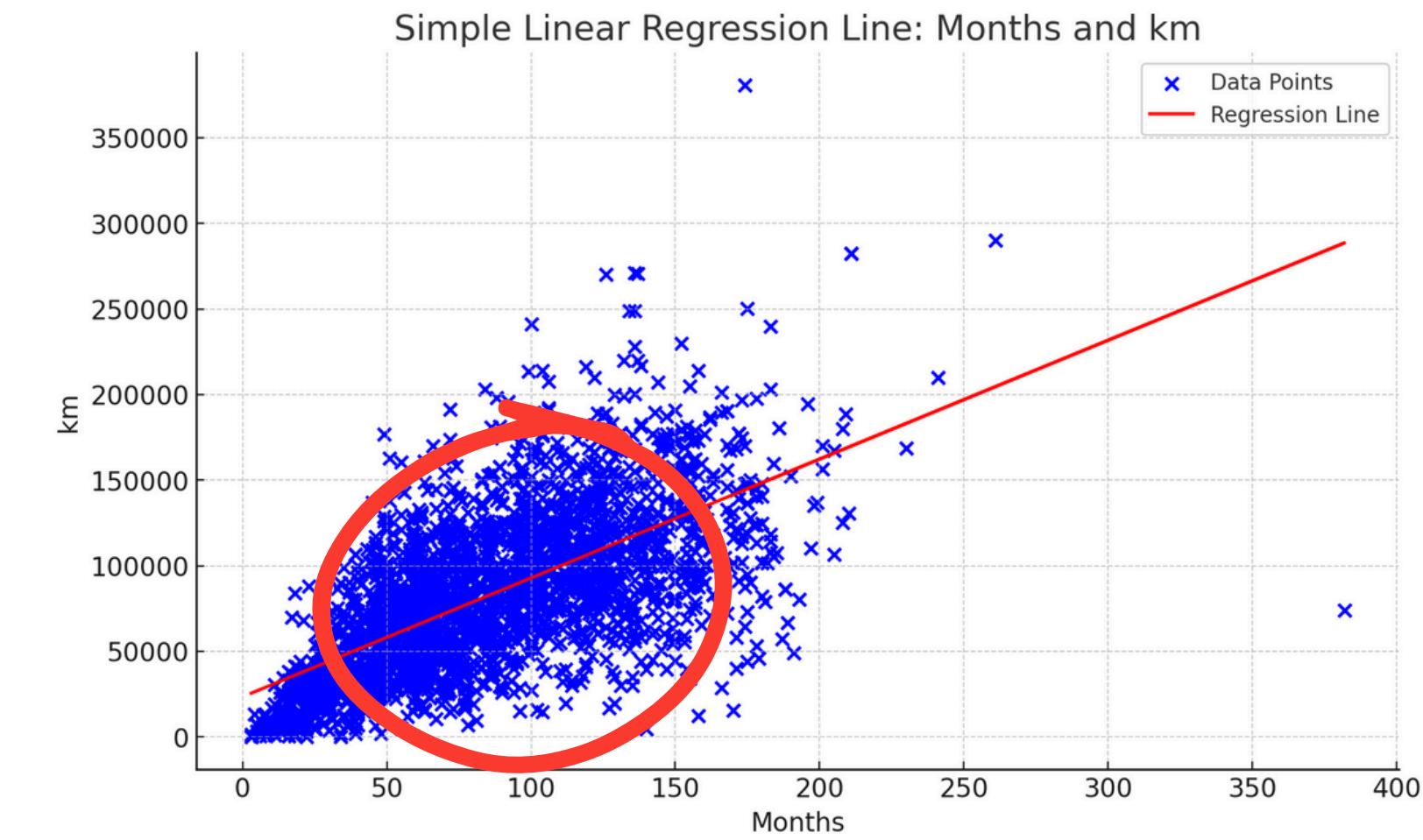
# 시간의 경과에 따른 중고차 주행거리 변화 추이 확인

- 각 차급을 대표하는 차량 데이터 선정
- 데이터 분포 확인 후 연식에 따른 주행거리의 회귀직선 추정

02



소형차

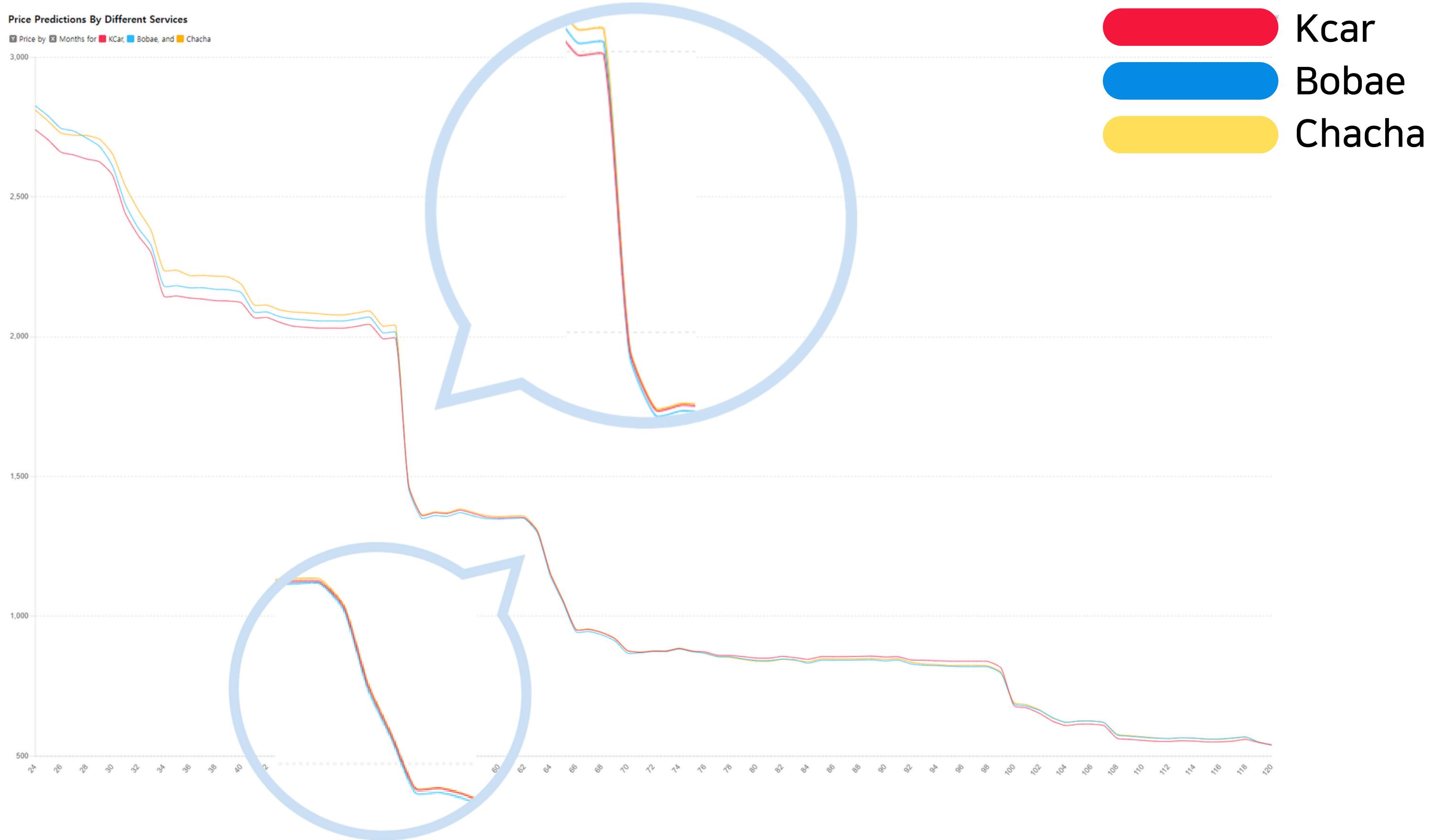


중형차

03

04

# 소형 중고차 가격의 변화 추이



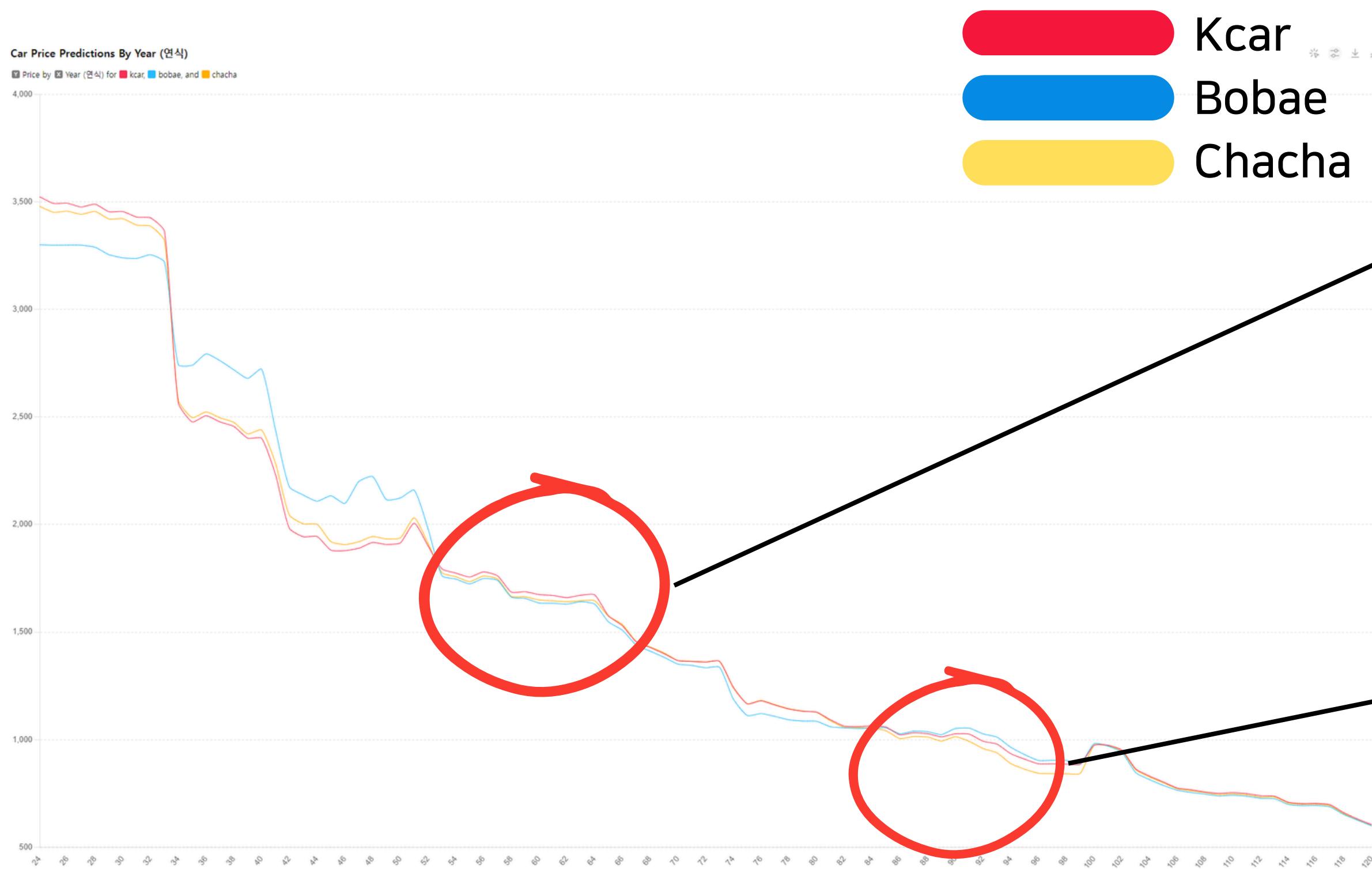
# 01 기대효과

예측 모형을 통한 개인별 중고차 적정 판매 시기 제안



01

# 중형 중고차 가격의 변화 추이



40개월 ~ 50개월

50개월 ~ 60개월

01

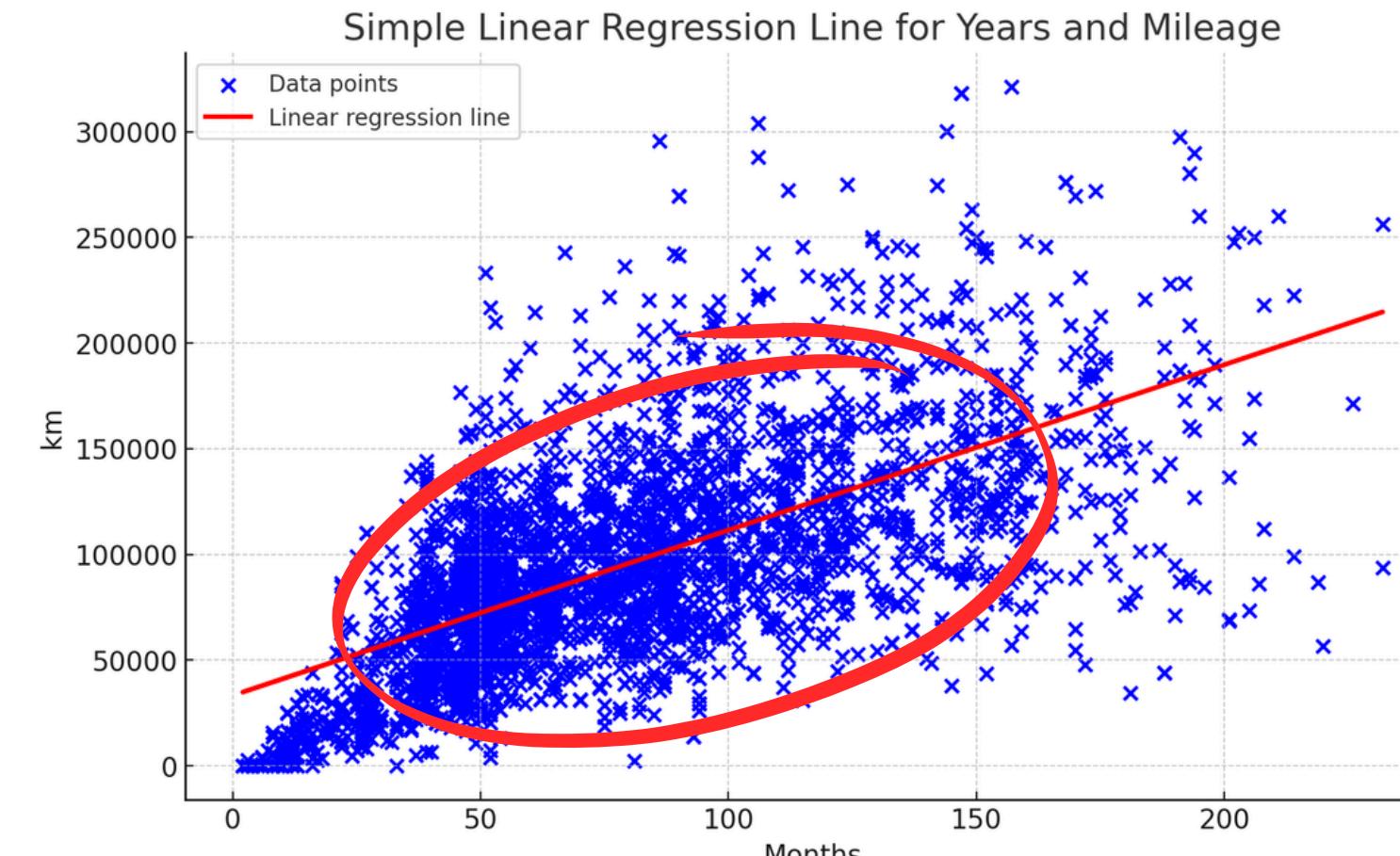
# 시간의 경과에 따른 중고차 주행거리 변화 추이 확인

- 대형 차 데이터 분포 확인 후 연식에 따른 주행거리의 회귀직선 추정

02

03

04

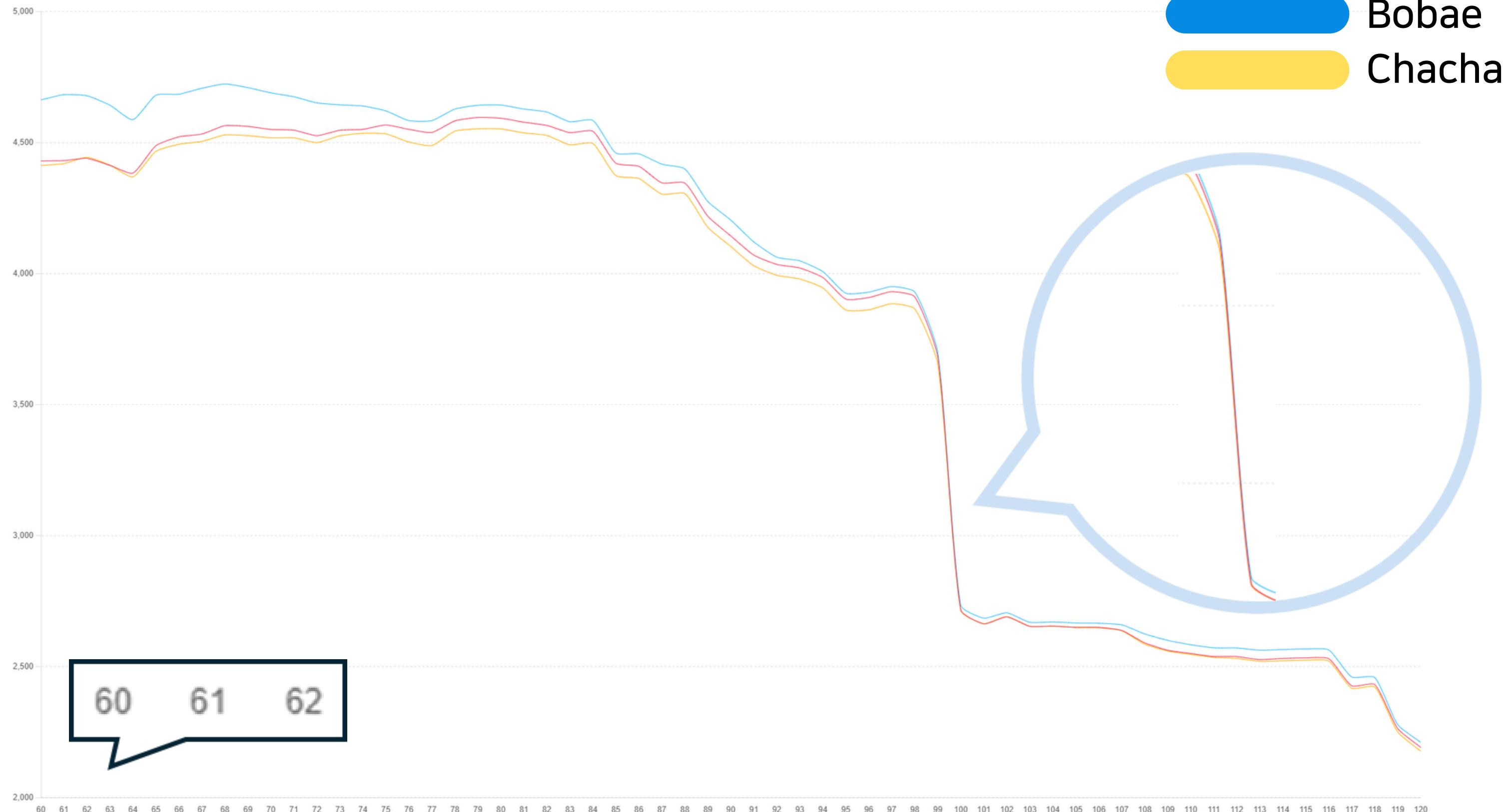


대형차

# 01 대형 중고차 가격의 변화 추이

Price Predictions By Kcar, Bobae, And Chacha (Years >= 60)

☒ Price by ☒ Years for ■ kcar, ■ bobae, and ■ chacha



01

## 프로젝트 결과

02

01

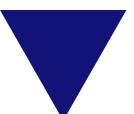
구축한 모델로 중고차 판매가 예측



03

02

예측된 가격을 비교하여 사이트 추천



04

03

중고차 적정 판매 시기 제안

# 기계학습의 이해

## 최종발표

중고차 가격예측 모델을 이용한  
중고차 사이트 추천



심현석

수학과

이승우

동물자원과학과

서준형

스마트ICT융합공학과

이동욱

스마트ICT융합공학과