# 1 문제정의

### 1.0 주제선정 배경

* 어려운 경기 속에서 조금이라도 경제적인 요소를 고려하는 사람들을 위해 유지비가 낮은 차량 선택에 도움을 주는 예측 모델을 만드는 것을 목표로 한다.

### 1.1 문제설명

경제적인 요소를 고려하는 사람들을 위해 연비가 높은 차량을 선정할 때 원하는 연비의 차량에 대해 필요한 대략적인 차량가격을 예측하고자 한다.

### 1.2 자료 수집처

* ‘https://www.data.go.kr/data/15005963/fileData.do’ 공공데이터 포털
* ‘https://www.carisyou.com/’ 카이즈유

### 1.3 변수 설명

* 모델명 : 자동차 모델명
* 제조사 : 국내외 자동차 제조사
* 유종 : 자동차 유종
* 배기량 : 자동차 배기량
* 변속형식 : 자동차 변속기어형식
* 복합연비 : 도심연비와 고속도로주행 연비에 각각 55%, 45%의 가중치를 적용하여 산출된 연비
* 등급 : 도심연비와 고속도로 주행 연비에 55, 45%의 가중치를 적용해 산출하는 연료비율당 이동 거리, 즉 연료의 효율 나타냄
* CO2배출량 : 자동차 이산화탄소 배출량
* 분류 : 배기량을 기준으로 나눈 경차, 소형차, 중형차, 대형차
* 가격 : 가격은 카이즈유 사이트를 기준으로 하여 최고값으로 한다, 카이즈유에 없는 경우는 구글 검색을 통하여 구한다.

# 2 DataFrame 생성

import pandas as pd

data = pd.read\_csv('data.csv', encoding = 'utf-16')  
data

Unnamed: 0

모델명

제조사

유종

배기량

변속형식

복합연비

등급

CO2배출량

분류

가격

0

0

모닝(JA) 1.0가솔린\_밴형 (15)

기아

휘발유

998

자동4

15.4

경형

106.0

경차

1180

1

1

모닝(JA)1.0가솔린 (15)

기아

휘발유

998

자동4

15.4

경형

106.0

경차

1195

2

2

모닝(JA) 1.0가솔린 (16인치 타이어) (15)

기아

휘발유

998

수동5

15.3

경형

106.0

경차

1234

3

3

모닝(JA) 1.0가솔린\_밴형 (16인치 타이어) (15)

기아

휘발유

998

수동5

15.3

경형

106.0

경차

1280

4

4

스파크 밴 1.0

한국지엠

휘발유

999

수동5

15.0

경형

109.0

경차

1020

…

…

…

…

…

…

…

…

…

…

…

…

110

115

“K7 2.2디젤 (PE/17”“타이어)”

기아

경유

2199

자동8

14.6

2등급

130.0

대형차

3500

111

116

“K7 2.2디젤 (17”“타이어)”

기아

경유

2199

자동8

14.6

2등급

130.0

대형차

3505

112

117

스팅어 2.2디젤 2WD\_18인치 타이어 (15)

기아

경유

2199

자동8

14.5

2등급

131.0

대형차

3870

113

118

“K7 2.2디젤 (PE/18”“타이어)”

기아

경유

2199

자동8

14.5

2등급

131.0

대형차

3515

114

119

C220 d

벤츠

경유

2143

자동9

14.5

2등급

131.0

대형차

4650

115 rows × 11 columns

del data['Unnamed: 0']

data

모델명

제조사

유종

배기량

변속형식

복합연비

등급

CO2배출량

분류

가격

0

모닝(JA) 1.0가솔린\_밴형 (15)

기아

휘발유

998

자동4

15.4

경형

106.0

경차

1180

1

모닝(JA)1.0가솔린 (15)

기아

휘발유

998

자동4

15.4

경형

106.0

경차

1195

2

모닝(JA) 1.0가솔린 (16인치 타이어) (15)

기아

휘발유

998

수동5

15.3

경형

106.0

경차

1234

3

모닝(JA) 1.0가솔린\_밴형 (16인치 타이어) (15)

기아

휘발유

998

수동5

15.3

경형

106.0

경차

1280

4

스파크 밴 1.0

한국지엠

휘발유

999

수동5

15.0

경형

109.0

경차

1020

…

…

…

…

…

…

…

…

…

…

…

110

“K7 2.2디젤 (PE/17”“타이어)”

기아

경유

2199

자동8

14.6

2등급

130.0

대형차

3500

111

“K7 2.2디젤 (17”“타이어)”

기아

경유

2199

자동8

14.6

2등급

130.0

대형차

3505

112

스팅어 2.2디젤 2WD\_18인치 타이어 (15)

기아

경유

2199

자동8

14.5

2등급

131.0

대형차

3870

113

“K7 2.2디젤 (PE/18”“타이어)”

기아

경유

2199

자동8

14.5

2등급

131.0

대형차

3515

114

C220 d

벤츠

경유

2143

자동9

14.5

2등급

131.0

대형차

4650

115 rows × 10 columns

### 데이터 앞 행 6개

data.head(6)

모델명

제조사

유종

배기량

변속형식

복합연비

등급

CO2배출량

분류

가격

0

모닝(JA) 1.0가솔린\_밴형 (15)

기아

휘발유

998

자동4

15.4

경형

106.0

경차

1180

1

모닝(JA)1.0가솔린 (15)

기아

휘발유

998

자동4

15.4

경형

106.0

경차

1195

2

모닝(JA) 1.0가솔린 (16인치 타이어) (15)

기아

휘발유

998

수동5

15.3

경형

106.0

경차

1234

3

모닝(JA) 1.0가솔린\_밴형 (16인치 타이어) (15)

기아

휘발유

998

수동5

15.3

경형

106.0

경차

1280

4

스파크 밴 1.0

한국지엠

휘발유

999

수동5

15.0

경형

109.0

경차

1020

5

스파크 밴 1.0

한국지엠

휘발유

999

무단변속

15.0

경형

109.0

경차

1060

### 데이터 뒤 행 6개

data.tail(6)

모델명

제조사

유종

배기량

변속형식

복합연비

등급

CO2배출량

분류

가격

109

C220 d 4Matic

벤츠

경유

2143

자동9

14.7

2등급

129.0

대형차

4690

110

“K7 2.2디젤 (PE/17”“타이어)”

기아

경유

2199

자동8

14.6

2등급

130.0

대형차

3500

111

“K7 2.2디젤 (17”“타이어)”

기아

경유

2199

자동8

14.6

2등급

130.0

대형차

3505

112

스팅어 2.2디젤 2WD\_18인치 타이어 (15)

기아

경유

2199

자동8

14.5

2등급

131.0

대형차

3870

113

“K7 2.2디젤 (PE/18”“타이어)”

기아

경유

2199

자동8

14.5

2등급

131.0

대형차

3515

114

C220 d

벤츠

경유

2143

자동9

14.5

2등급

131.0

대형차

4650

### 데이터 행 / 열 개수

data.shape

(115, 10)

### 데이터 요약 정보

data.info()

<class ‘pandas.core.frame.DataFrame’> RangeIndex: 115 entries, 0 to 114 Data columns (total 10 columns): # Column Non-Null Count Dtype  
— —— ————– —–  
0 모델명 115 non-null object 1 제조사 114 non-null object 2 유종 114 non-null object 3 배기량 115 non-null int64  
4 변속형식 115 non-null object 5 복합연비 115 non-null float64 6 등급 115 non-null object 7 CO2배출량 115 non-null float64 8 분류 115 non-null object 9 가격 115 non-null int64  
dtypes: float64(2), int64(2), object(6) memory usage: 9.1+ KB

data.dtypes

모델명 object 제조사 object 유종 object 배기량 int64 변속형식 object 복합연비 float64 등급 object CO2배출량 float64 분류 object 가격 int64 dtype: object

### 데이터 변수들의 값

data['모델명'].unique()

array([‘모닝(JA) 1.0가솔린\_밴형 (15)’, ‘모닝(JA)1.0가솔린 (15)’, ‘모닝(JA) 1.0가솔린 (16인치 타이어) (15)’, ‘모닝(JA) 1.0가솔린\_밴형 (16인치 타이어) (15)’, ‘스파크 밴 1.0’, ‘스파크 1.0’, ‘“스토닉 1.0 T-GDI 7DCT (15”“타이어)”’, ‘“스토닉 1.0 T-GDI 7DCT (17”“타이어)”’, ‘레이 1.0가솔린 (2015년형)’, ‘레이 1.0가솔린 (PE, 14인치 타이어)’, ‘“레이 1.0가솔린 밴형 (PE, 14인치 타이어)”’, ‘레이 1.0가솔린 밴형 (PE, 14인치 타이어)’, ‘모닝(JA) 1.0T-GDI\_14인치 (15)’, ‘레이 1.0가솔린\_밴형 (15)’, ‘다마스 5인승’, ‘다마스밴’, ‘라보롱카고’, ‘라보5MT’, ‘TWIZY Urban(2인승)’, ‘TWIZY Urban(1인승)’, ‘아이오닉 1.6GDI 플러그인하이브리드 개조차\_6DCT’, "아이오닉 1.6GDI 하이브리드 6DCT\_15’‘타이어“, ‘아이오닉 1.6GDI 플러그인하이브리드\_6DCT (15)’,”아이오닉 1.6GDI 하이브리드 6DCT\_17’‘타이어“,”니로 1.6GDI 하이브리드\_16’‘타이어 (15)“, ‘코나 1.6GDI 하이브리드 (16" 타이어)’, ‘토요타 Prius C’, ‘니로(DE) 1.6GDI 플러그인 하이브리드\_6DCT (15)’, ‘K3 1.6 디젤(4DR) 7DCT\_ISG (15)’,”아반떼(AD) 1.6디젤\_7DCT ISG (15’‘/16’‘타이어) (15)", ’엑센트 1.6디젤 (4DR) (15)’, ‘아반떼(AD) 1.6디젤 (15)’, ’CHEVROLET