

# Data Analysis

차량 데이터에 따른 연비 분석



이 한 길

# STACKS

분석 언어  
python



분석 도구  
scikit



분석환경  
Google colab



1

## 분석 목표 설정

분석 목적 / 활용 데이터

2

## 데이터 준비

데이터 수집 및 데이터 전처리

3

## 데이터 분석

자동차 연비 예측을 위한 선형  
회귀 분석 모델 구축

4

## 분석 결과 시각화

산점도 , 선형 회귀  
그래프로 시각화

# 1. 분석 목표 설정



## 분석 목표 설정 단계

- 차량별 스펙 데이터 활용
- 머신러닝 기반의 회귀 분석을 수행
- 항목별 연비에 미치는 영향을 확인



**최종 분석결과로 차량 연비 예측**

# 분석 데이터 조사와 선정

## 차량의 관성 주행 영향력

- ✓ 연비 산출시 배기량과 가속에 의한 엔진 회전수가 연관
- ✓ 차량 가속시 퓨얼 컷 오프(Fuel Cut Off) 구역 에 돌입하고 연비에 큰 영향

# Fuel Cut Off : 가속에 의한 관성 주행 구간

출처 : 한국에너지공단 수송에너지 자료



## 최종 데이터 선정

Model\_year(연식) , Acceleration(가속도 성능) , Displacement(배기량) ,  
weight(무게) , cylinders(실린더)

## 2. 데이터 준비

	A	B	C	D	E	F	G	H	I
1	mpg	cylinders	displacement	horsepower	weight	acceleration	model_year	origin	car_name
2	18	8	307	130	3504	12	70	1	chevrolet chevelle malibu
3	15	8	350	165	3693	11.5	70	1	buick skylark 320
4	18	8	318	150	3436	11	70	1	plymouth satellite
5	16	8	304	150	3433	12	70	1	amc rebel sst
6	17	8	302	140	3449	10.5	70	1	ford torino
7	15	8	429	198	4341	10	70	1	ford galaxie 500
8	14	8	454	220	4354	9	70	1	chevrolet impala
9	14	8	440	215	4312	8.5	70	1	plymouth fury iii
10	14	8	455	225	4425	10	70	1	pontiac catalina
11	15	8	390	190	3850	8.5	70	1	amc ambassador dpl
12	15	8	383	170	3563	10	70	1	dodge challenger se
13	14	8	340	160	3609	8	70	1	plymouth 'cuda 340
14	15	8	400	150	3761	9.5	70	1	chevrolet monte carlo
15	14	8	455	225	3086	10	70	1	buick estate wagon (sw)
16	24	4	113	95	2372	15	70	3	toyota corona mark ii
17	22	6	198	95	2833	15.5	70	1	plymouth duster
18	18	6	199	97	2774	15.5	70	1	amc hornet



- 선정 데이터 확인
- 데이터 전처리

### Data Preprocessing

```
import numpy as np
import pandas as pd

data_df = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/Data_GasMileage/auto-mpg.csv', header = 0, engine='python')
print('데이터셋 크기 : ', data_df.shape)
data_df.head()
# csv 데이터 확인
# 데이터셋은 398개이며 컬럼은 9개
```

데이터셋 크기 : (398, 9)

	mpg	cylinders	displacement	horsepower	weight	acceleration	model_year	origin	car_name
0	18.0	8	307.0	130	3504	12.0	70	1	chevrolet chevelle malibu
1	15.0	8	350.0	165	3693	11.5	70	1	buick skylark 320
2	18.0	8	318.0	150	3436	11.0	70	1	plymouth satellite
3	16.0	8	304.0	150	3433	12.0	70	1	amc rebel sst
4	17.0	8	302.0	140	3449	10.5	70	1	ford torino

### 데이터 선별 후 조회

```
data_df = data_df.drop(['car_name', 'origin', 'horsepower'], axis=1, inplace = False)
# drop 메서드를 axis=1(열을 기준으로 수행) // axis=0(행을 기준으로 수행)
# inplace = False로 설정하여 메서드 수행 후 원본은 유지하며 결과값만 산출

print('데이터셋 크기 : ', data_df.shape)
data_df.head()
```

### 3. 데이터 분석



#### ▼ Data Analysis and Machine Training

```
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score

# X, Y 분할
y = data_df['mpg']
x = data_df.drop(['mpg'], axis = 1, inplace=False)

# 훈련용 데이터와 평가용 데이터 분할
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.3, random_state = 0)

# 선형 회귀 분석 모델 생성
lr = LinearRegression()

# 분석 모델 훈련
lr.fit(x_train, y_train)

# 선형 회귀 분석 : 평가 데이터에 대한 예측 수행 -> 예측 결과 y_predict 산출
y_predict = lr.predict(x_test)
```

#### 사이킷런 선형회귀 분석

- 성능 측정을 위한 평가 지표인 mean\_squared\_error, r2\_score 임포트
- test\_size=0.3, train\_test\_split() → 데이터를 7:3 비율로 분할하여 학습 데이터와 평가 데이터로 설정

### 3. 데이터 분석 – 회귀식 확인



```
[ ] coef = pd.Series(data = np.round(lr.coef_,2), index = x.columns)
coef.sort_values(ascending = False)
```

```
model_year      0.76
acceleration    0.20
displacement     0.01
weight          -0.01
cylinders        -0.14
dtype: float64
```

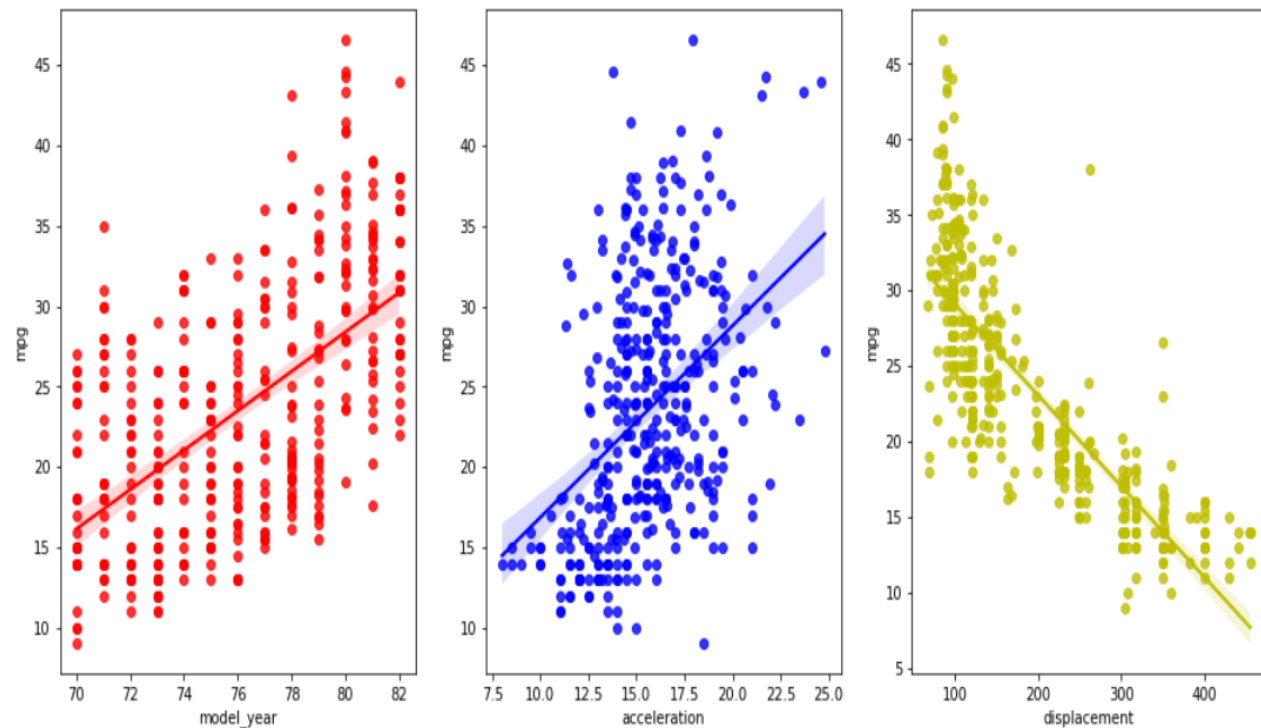
사이킷런 최종 회귀식 산출

$Y_{mpg} =$

$-0.14X_{cylinders} + 0.01X_{displacement} - 0.01X_{weight} + 0.20X_{acceleration} + 0.76X_{model\_year} - 17.55$



## 4. 데이터 시각화



산점도 와 선형 그래프로 데이터 시각화

→ 독립변수 시각화 데이터로 개별 피처의  
영향력 확인

# 분석 결론 , 연비 예측



- 시각화된 결과로 독립 변수5개에 대한 개별 연비 영향력 확인  
→ 연식이 낮을 수록, 가속능력이 좋을수록 연비가 높으며, 이와 반대로 배기량, 무게, 실린더(기통)이 낮을수록 연비가 높다

```
↳ 연비를 예측할 차량의 정보를 입력해 주세요!  
cylinders : 8  
displacement : 350  
weight : 3200  
acceleration : 22  
model_year : 99  
이 자동차의 예상 연비(MPG)는 41.32입니다.
```

- 분석 모델을 통한 연비 예측