

데이터 시각화 연습문제 (3) - seaborn 막대, 빈도 그래프

다음의 데이터는 1999년과 2008년에 미국 EPA에서 조사하여 발표한 자동차 주요 모델별 연비 데이터의 일부이다.

id	manufacturer	model	displ	year	cyl	trans	drv	cty	hwy	fl	class
1	audi	a4	1.8	1999	4	auto(l5)	f	18	29	p	compact
2	audi	a4	1.8	1999	4	manual(m5)	f	21	29	p	compact
3	audi	a4	2	2008	4	manual(m6)	f	20	31	p	compact
4	audi	a4	2	2008	4	auto(av)	f	21	30	p	compact
5	audi	a4	2.8	1999	6	auto(l5)	f	16	26	p	compact

전체 데이터는 아래의 URL을 통해 확인할 수 있다.

<https://data.hossam.kr/D01/mpg.xlsx>

데이터는 234 개의 행이 있으며, 각 행은 다음과 같은 변수로 구성되어 있다.

변수	설명
manufacturer	자동차 제조사
model	자동차 모델명
displ	자동차 배기량
year	제조년도
cyl	엔진 실린더 수
trans	자동차 트랜스미션 종류
drv	자동차 구동 방식. f=전륜구동, r=후륜구동, 4=사륜구동
cty	도심 연비 (마일/갤론)
hwy	고속도로 연비 (마일/갤론)
fl	연료 종류
class	자동차 분류

문제1

자동차 제조사별로 출시한 자동차 수를 하나의 그래프로 시각화 하시오.

문제2

자동차 제조사에 따른 자동차 수를 제조년도를 범주로 하여 하나의 그래프로 시각화 하시오.

문제3

구동방식에 따른 자동차 수를 하나의 그래프로 시각화 하시오.

문제4

자동차 구동 방식에 따른 도심 연비 평균을 하나의 그래프로 시각화 하시오.

문제5

자동차 제조사에 따른 도심 연비 평균과 고속도로 연비 평균을 서브플롯 형태로 시각화 하시오.

문제6

배기량과 도심연비, 배기량과 고속도로 연비의 상관관계를 확인하고자 한다.

서브플롯에 각 그래프를 시각화 하고 추세선을 포함시키시오.

문제7

철수는 새로 자동차를 구입하고자 한다.

좋은 자동차를 구입하는데 있어서 가장 중요시 되는 조건은 평균연비이다.

데이터 전처리 요구사항

1. 데이터 프레임에 평균연비에 대한 파생변수를 생성하시오.
2. 전체 평균연비에 대한 평균값을 구하고 이 평균값보다 평균연비가 높은 경우 'Pass', 그렇지 않을 경우 'Fail'로 표시할 수 있는 파생변수를 평가 라는 이름으로 추가하시오.

3. 평균연비에 대한 백분율을 구한 후 90점 이상일 경우 A, 80점 이상일 경우 B, 70점 이상일 경우 C, 그 외의 경우는 D로 하여 등급이라는 컬럼을 추가하시오.

다음의 네가지 지표를 확인할 수 있는 서브플롯을 제시하라

- 자동차 구동방식별 평가에 대한 빈도
- 자동차 구동방식별 등급에 대한 빈도
- 자동차 연료종류 평가에 대한 빈도
- 자동차 연료종류 등급에 대한 빈도
- 자동차 분류별 평가에 대한 빈도
- 자동차 분류별 등급에 대한 빈도
- 평균연비는 고속도로연비와 도심연비에 대한 평균이다.