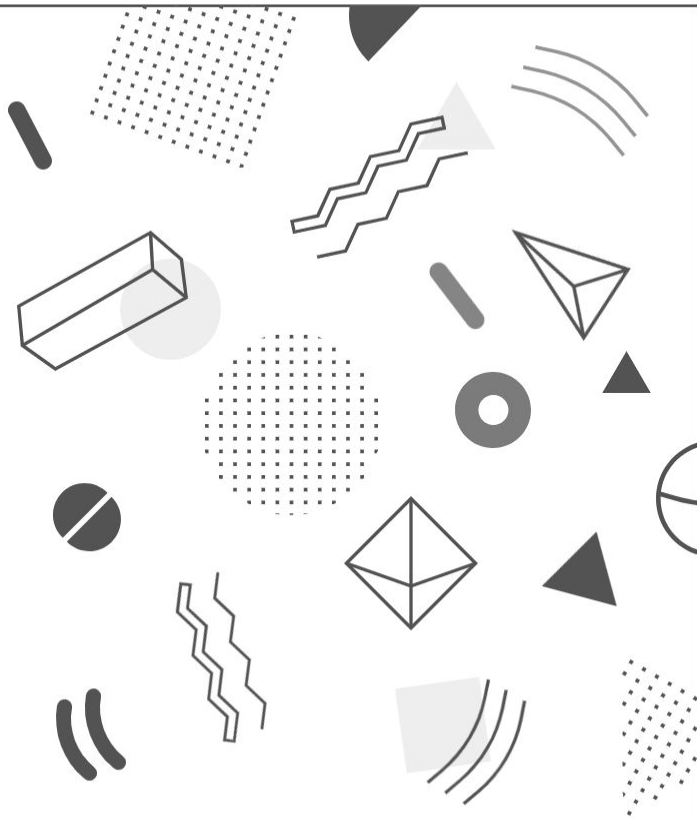


KT AIVLE School

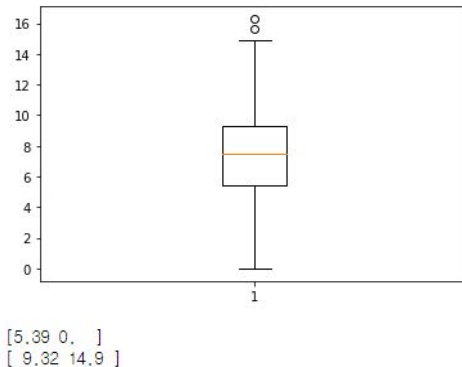
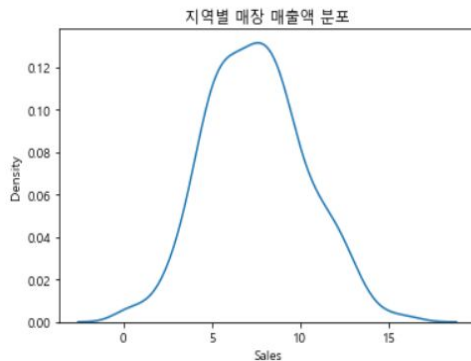
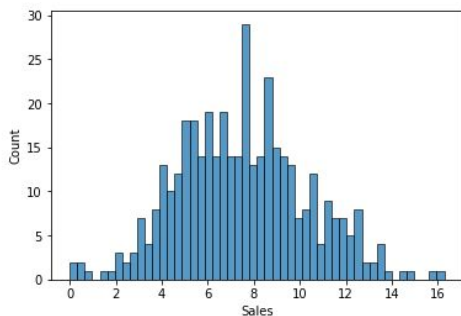
[미니 프로젝트] 카세트 매출액 분석

수도권 2반 3조(이세빈, 조혜빈, 차선희, 차혜돈, 최기호)



01 단변량 분석

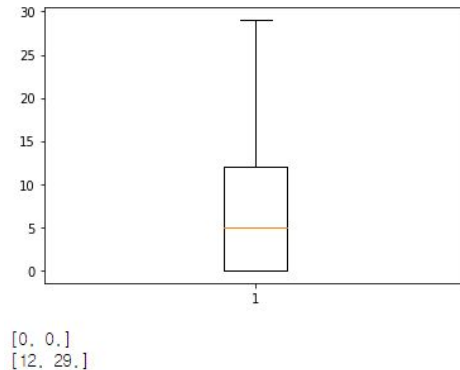
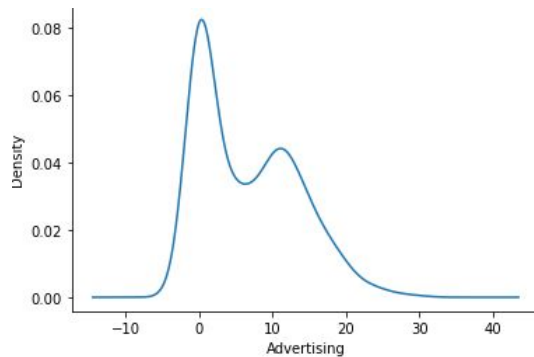
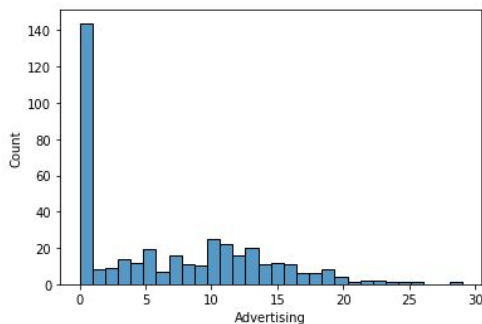
Sales



- ✓ 14,900달러 이상 판매한 매장은 이상치로 판단.
- ✓ 카시트가 하나도 판매되지 않은 매장이 존재하므로 판매하지 못한 매장을 추가 분석하여 원인을 파악해 보아야 함.
- ✓ 총 매출액이 높은 지역과 낮은 지역의 차이점을 분석하여 매장에서 수정할 수 있는 요소라면 개선사항으로 검토할 수 있음.

01 단변량 분석

Advertising

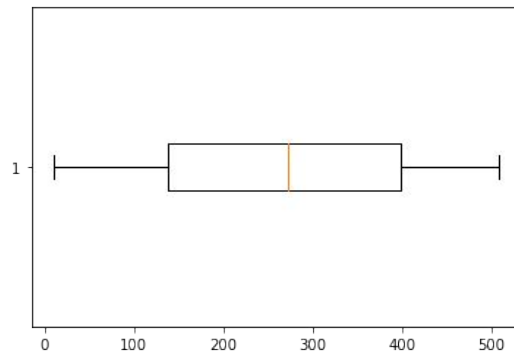
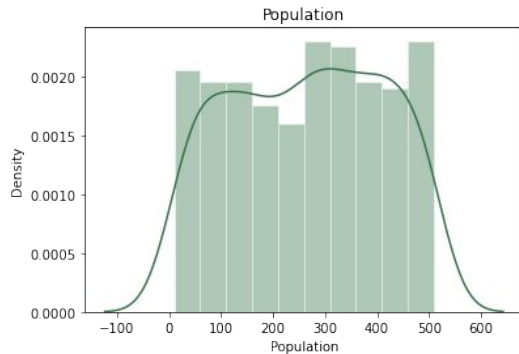


✓ 밀도함수 그래프가 왼쪽으로 치우쳐진 것으로 보아 대부분의 지역별 매장(75%)이 12,000달러 이하로 광고 비용을 사용한 것으로 보임.

✓ 25%에 해당하는 지역별 매장이 광고비를 투자하지 않아 **광고비를 사용한 매장을 따로 분석**해 볼 필요가 있음.

01 단변량 분석

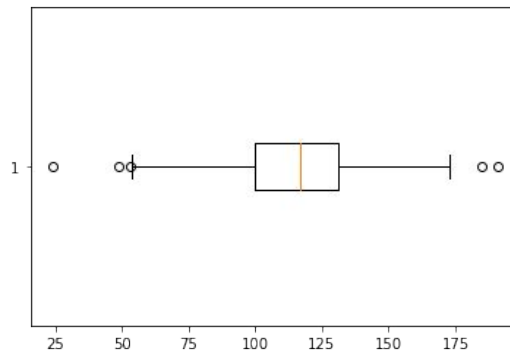
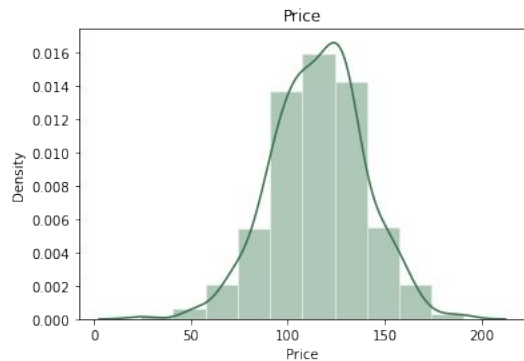
Population



- ✓ 지역 인구수는 유의미한 차이를 보이지 않음.
- ✓ 타 변수와의 관계를 파악할 필요가 있음.

01 단변량 분석

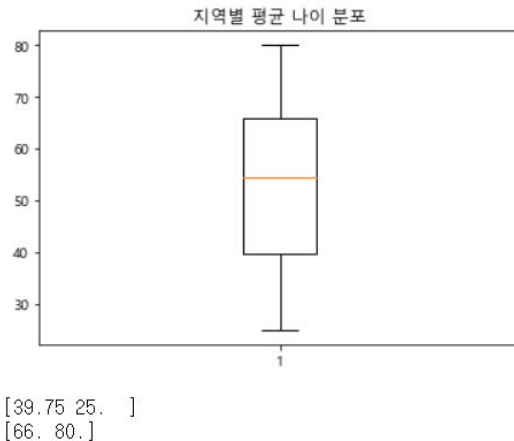
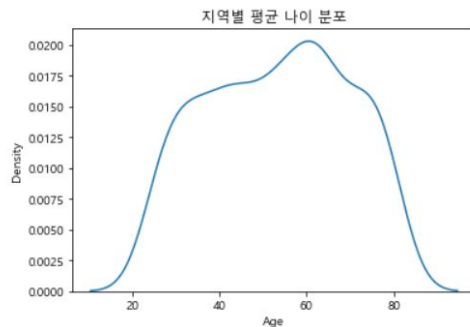
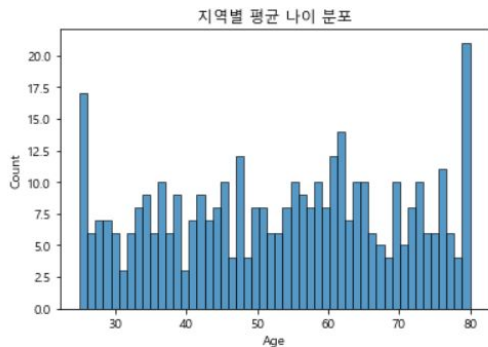
Price



- ✓ 지역별 매장 제품 가격의 최댓값과 최솟값의 차이가 약 170달러 정도임.
- ✓ 약 5만 달러 이하와 17만 3천 달러 이상에서 이상치를 보임.
- ✓ 가격이 높은 매장과 가격이 낮은 매장의 **요인 (예: 빈부격차)**을 추가 분석할 필요가 있음.

01 단변량 분석

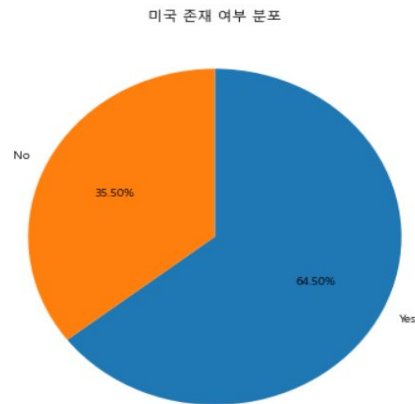
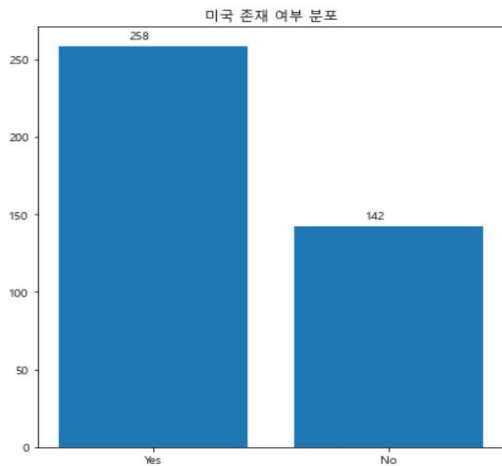
Age



- ✓ 지역 내 연령분포가 특정 연령층으로 치우치지 않은 고르게 분포된 것으로 볼 수 있음.
- ✓ 연령별 자녀유무 혹은 그 자녀의 연령대 또한 조사해 볼 필요가 있음.
- ✓ **0~5세, 80대 이상의 인구분포**가 다른 나이대에 비해 높은 경향을 보이므로 따로 분석해 볼 필요가 있음.

01 단변량 분석

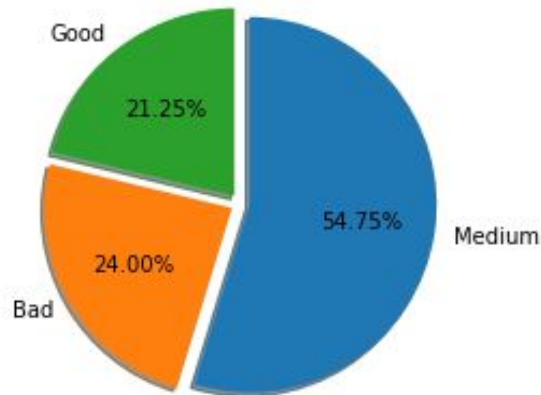
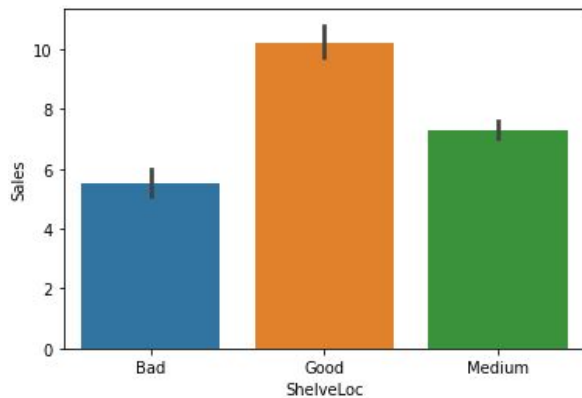
US



- ✓ 64.5% 가량의 매장이 미국 지역에 존재.
- ✓ 미국에 있는 매장과 없는 매장 차이점을 비교 분석.

01 단변량 분석

ShelveLoc

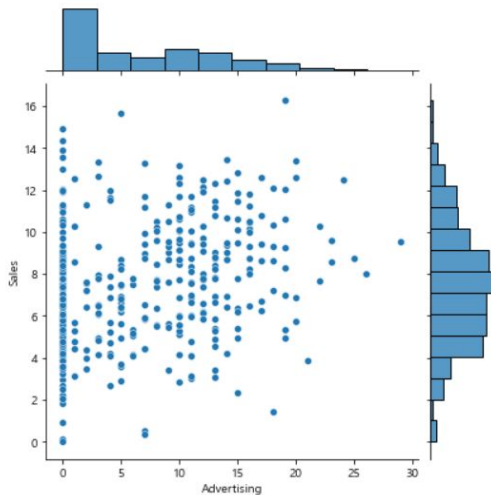


✓ 76%의 매장에서 진열상태가 양호한 편으로 분석되고 그 외 24%의 매장에서는 진열상태가 '나쁨'으로 분석됨.

✓ 진열상태가 ' 좋음'으로 분석되는 매장이 '나쁨'으로 분석되는 매장보다 적게 나올 만큼 진열상태가 좋지 않은 매장이 많기 때문에 본사 차원에서 **매장관리에 더 신경** 써야 할 것.

02 이변량 분석

Advertising -> Sales



상관계수 : 0.2695067813769019
p-value : 4.377677110302732e-08

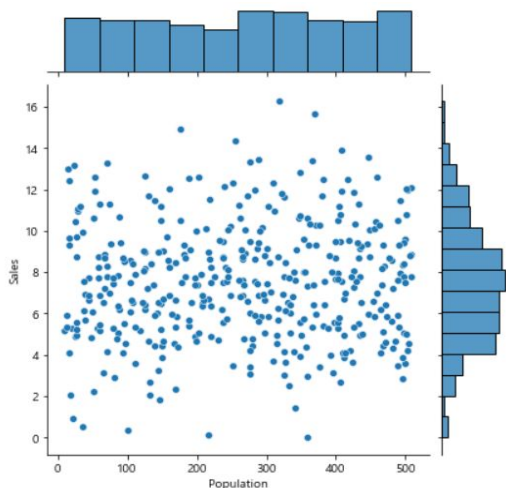
✓ 광고비가 0인 지역 매장을 제외하고는 매출액과의 관계 분포가 **약한 양의 상관관계**를 갖는 것으로 보임.

✓ 양의 상관관계를 가지고 있으나 광고비 0인 지역을 고려한 추가 분석이 필요.

✓ p-value 값이 일반적인 유의수준인 0.05보다 작으므로 두 변수 간의 상관관계가 유의미하다고 볼 수 있음.

02 이변량 분석

Population -> Sales



상관계수 : 0.05047098447203919
p-value : 0.31398160931903

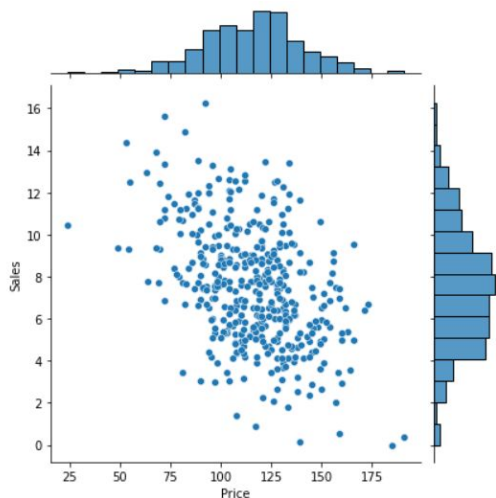
✓ 상관관계는 존재하지 않음.

✓ 대부분 고르게 분포.

✓ p-value 값이 일반적인 유의수준인 0.05보다 크므로
인구와 판매량에는 상관관계가 있다고 보기 어려움.

02 이변량 분석

Price -> Sales

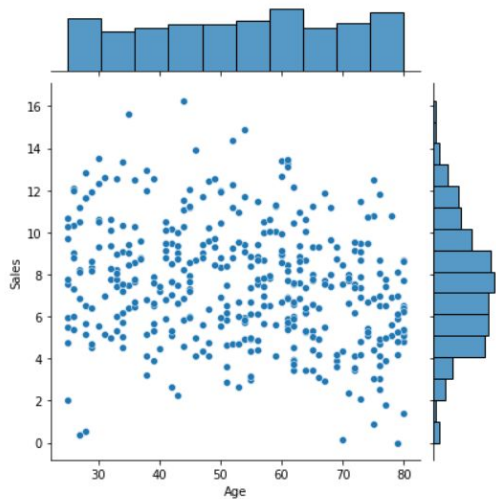


상관계수 : -0.4449507278465726
p-value : 7.61818701191294e-21

- ✓ 산점도에서 음의 관계가 형성되는 것을 발견.
- ✓ 상관계수가 -0.44495로 **비교적 강한 음의 상관관계**를 형성.
- ✓ 지역별 가격이 판매액에 영향을 미친다고 볼 수 있음.

02 이변량 분석

Age -> Sales

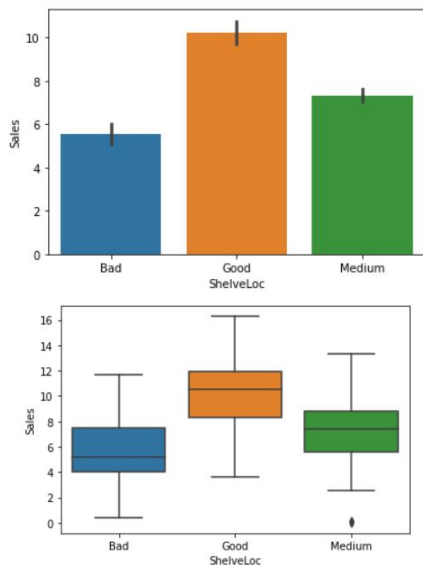


상관계수 : -0.2318154396045765
p-value : 2.7889498393711916e-06

- ✓ 산점도에서 전체적인 특징을 찾아보기 힘들.
- ✓ 구매하는 가격대에 따라 나이대별 차이가 있는 것으로 보임.
- ✓ 상관계수가 -0.23182 로 **약한 음의 상관관계**.
- ✓ 나이대가 판매액에 약한 영향을 미친다 할 수 있음.

02 이변량 분석

ShelveLoc-> Sales

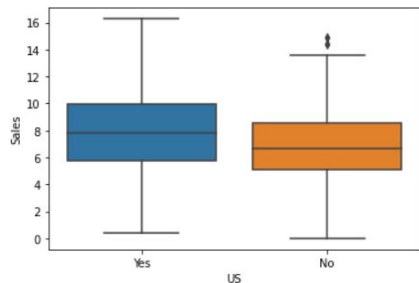
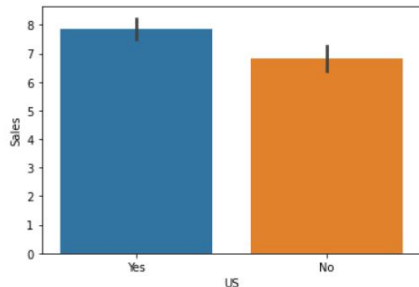


검정통계량 : 92.22990509910346
p-value : 1.26693609015938e-33

- ✓ 평균비교 그래프의 신뢰구간이 모두 겹치지 않는 것으로 판단.
- ✓ Medium의 낮은 이상치가 존재.
- ✓ p-value 값이 0.05보다 작음.
- ✓ **진열상태에 따라 판매액의 차이가 있다**고 할 수 있음.

02 이변량 분석

US-> Sales



검정통계량 : 3.589738747802499
p-value : 0.00037233958701471517

- ✓ 그래프상 신뢰구간이 겹치지 않고 미국에 존재할 때, 매출액이 약간 더 높은것으로 나타남.
- ✓ No 에 높은 이상치가 존재.
- ✓ p-value 가 0.00037로 작은 편에 속함.
- ✓ **미국에 매장 존재 여부에 따라 판매액 차이가 있다**고 할 수 있음.

03 관계정리

✓ 강한 관계의 변수

- Price, ShelfLoc

✓ 약한 관계의 변수

- Advertising, Age, US

✓ 관계 없음

- Population