

Data Base

7조. 정소현 이선목 박세진 차승제

01

1

분석 목적 : 40대 50대 쇼핑 업종 소비 분석.

2

분석 방법 :

- 1) 데이터 시각화를 통해, 데이터 구조를 자세하고 직관적으로 탐색.
- 2) BoxPlot을 통해, 40대 50대 구매 수량과 금액을 분석.
- 3) WordCloud를 통해, 연령대별 소비 품목 모형화.
- 4) calendarHeat를 통해, 일자별*월별 상품 구매 빈도 모형화.

3

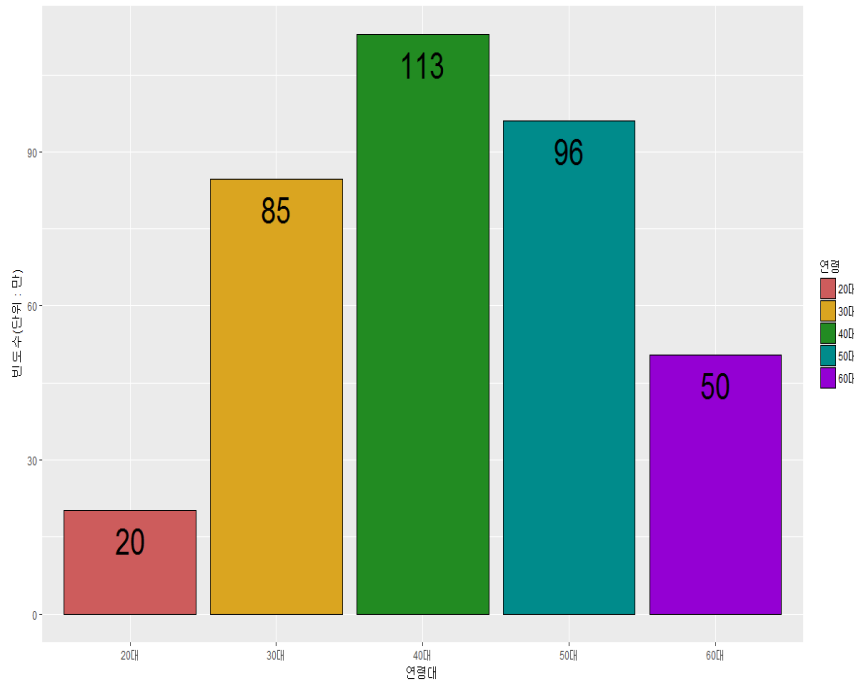
결론 도출 :

- 1) 40대 50대 총 구매 금액 차이 분석.
- 2) 40대 50대 특정 상품의 소비 분석.

02

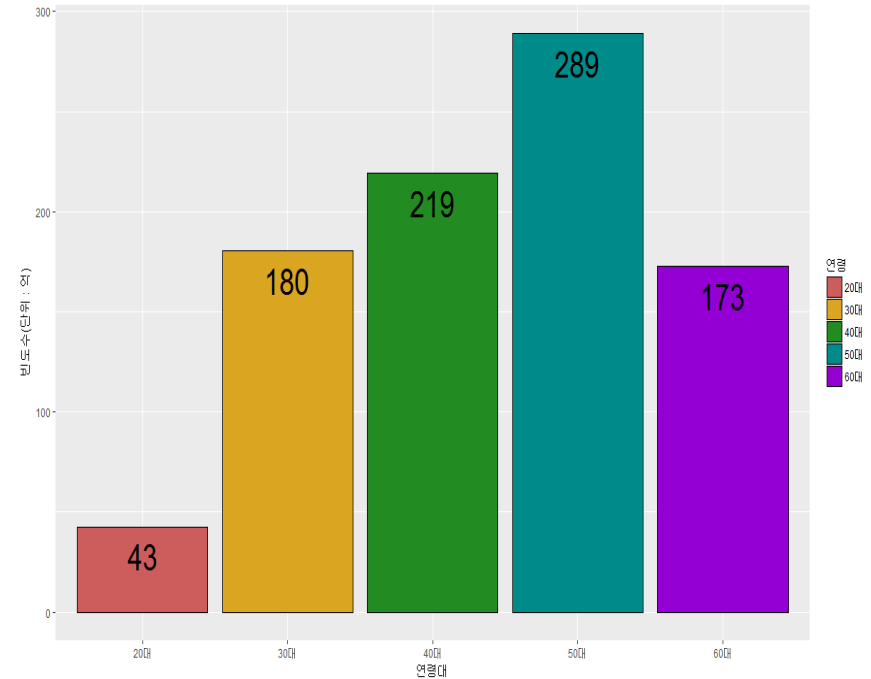
➤ 연령대별 쇼핑 이용 수

=> 40대 > 50대 > 30대 > 60대 > 20대



➤ 연령대별 쇼핑 이용 총 금액

=> 50대 > 40대 > 30대 > 60대 > 20대

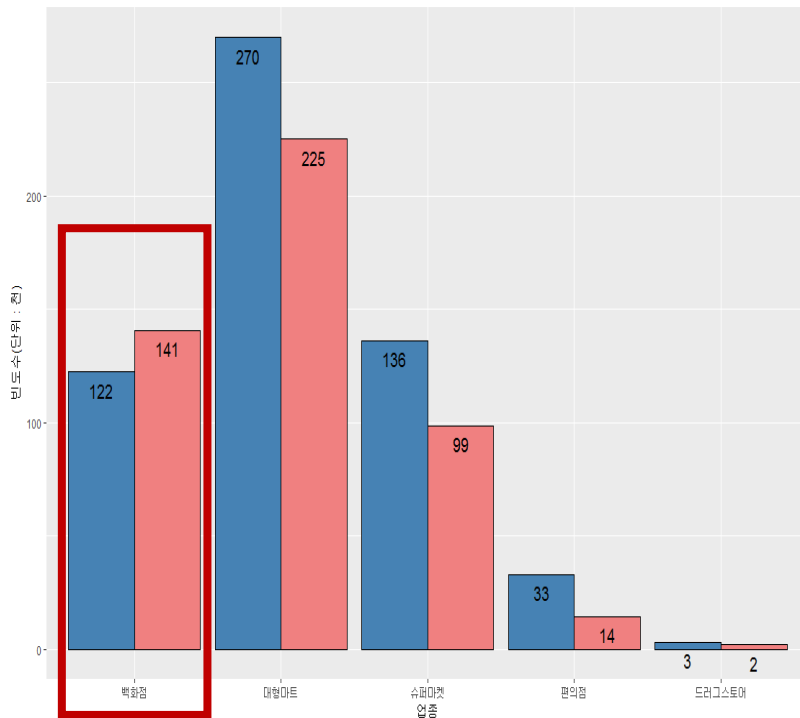


02

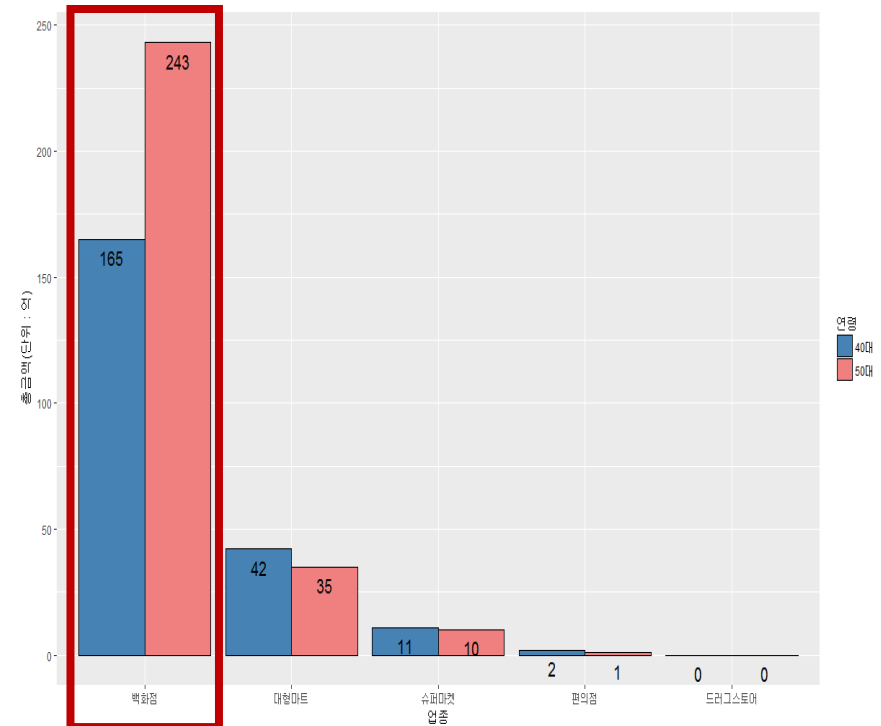
➤ 40대 50대 쇼핑 업종 이용 수

➤ 40대 50대 쇼핑 업종 이용 총 금액

=> (백화점) 50대 이용 수 > 40대 이용 수



=> 백화점 > 대형마트 > 슈퍼마켓 ...



03

빈도수 : 40대 > 50대, 총 금액 : 40대 < 50대
=> '백화점'에서 큰 금액 차이 발생.

①

단순히 백화점 이용 수가 많아서?
=> 백화점 이용 수

②

1인당 상품 구매 수가 많아서?
=> 1인당 상품 구매 수

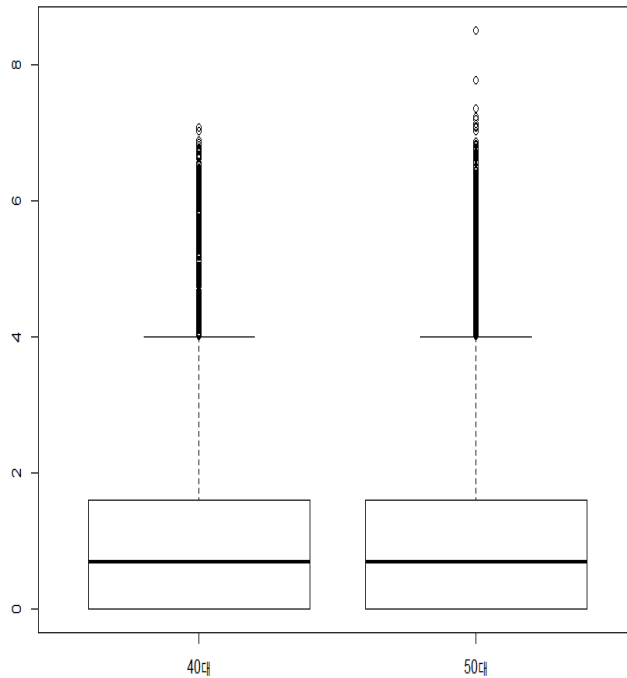
③

구매하는 상품의 금액이 비싸서?
=> 1인당 소비 금액

03

➤ 백화점 1인당 구매 수량 분포

1인당 백화점에서 하루 구매수량



	Min	1 st Qu	Median	Mean	3 rd Qu	Max
40대	1	1	2	4.759	5	953
50대	1	1	2	5.17	5	2407

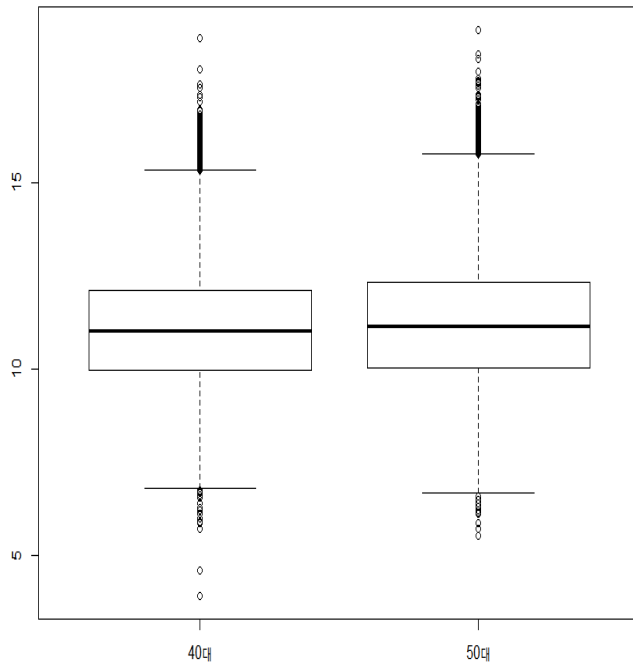
- 40대 Min == 50대 Min
- 40대 1st Qu == 50대 1st Qu
- 40대 3rd Qu == 50대 3rd Qu

⇒ **40대와 50대의 1인당 구매 수량이 비슷하다**

03

➤ 백화점 1인당 소비 금액 분포

1인당 하루 백화점에서 소비금액



	Min	1 st Qu	Median	Mean	3 rd Qu	Max
40대	50	21500	60800	214731	185000	159811920
50대	250	22800	69980	278144	226750	199648000

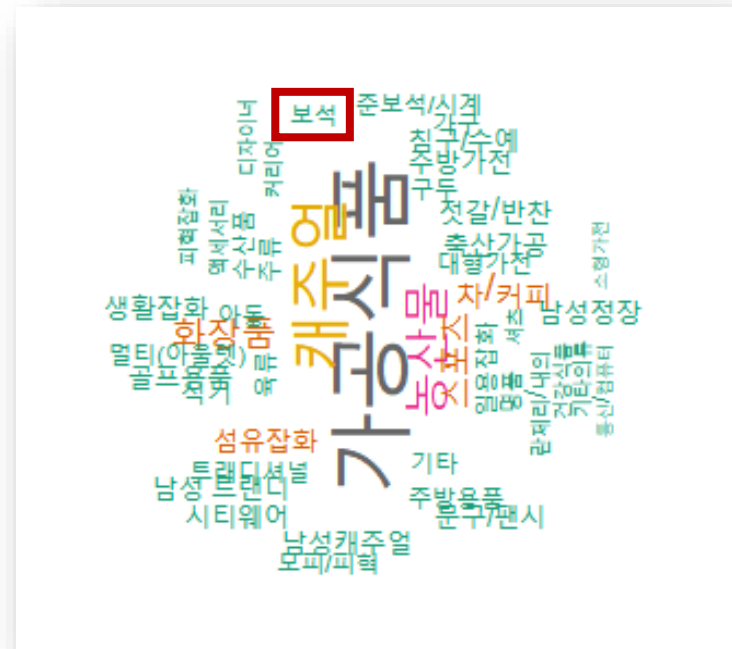
- Mean >>>> 3rd Qu
⇒ 오른쪽으로 꼬리가 매우 긴 분포, 극단 현상이 심함
- 50대 Mean >>>> 40대 Mean
⇒ **50대가 40대보다 1인당 소비 금액이 더 크다**

03

➤ 40대 백화점 소비 품목 (중분류)



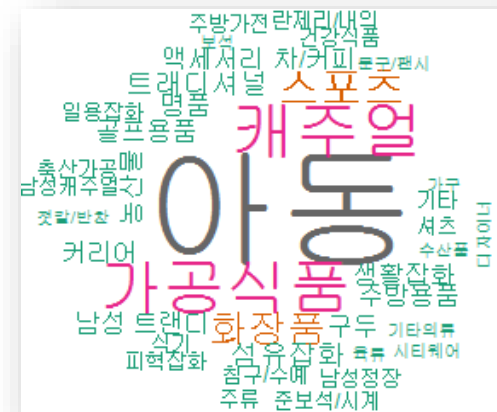
➤ 50대 백화점 소비 품목 (중분류)



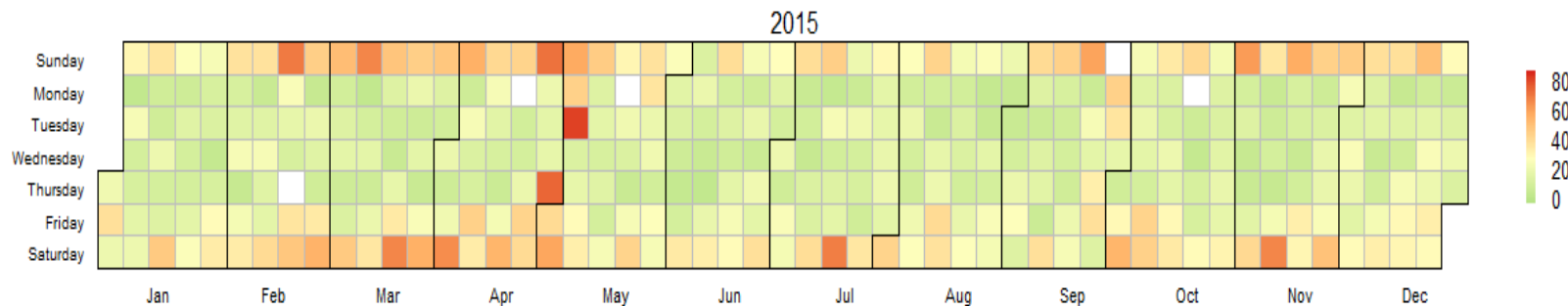
03

➤ 40대 백화점 “아동” 소비 품목 분석

같이 구매되는 상품	빈도수
캐주얼	413
가공식품	407
스포츠	233



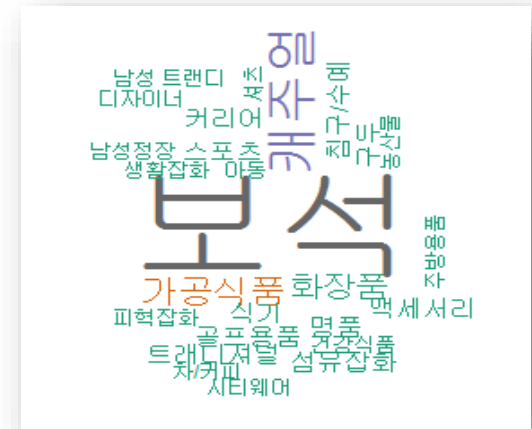
Calendar Heat Map of 40대 백화점 아동 날짜별 구매 빈도



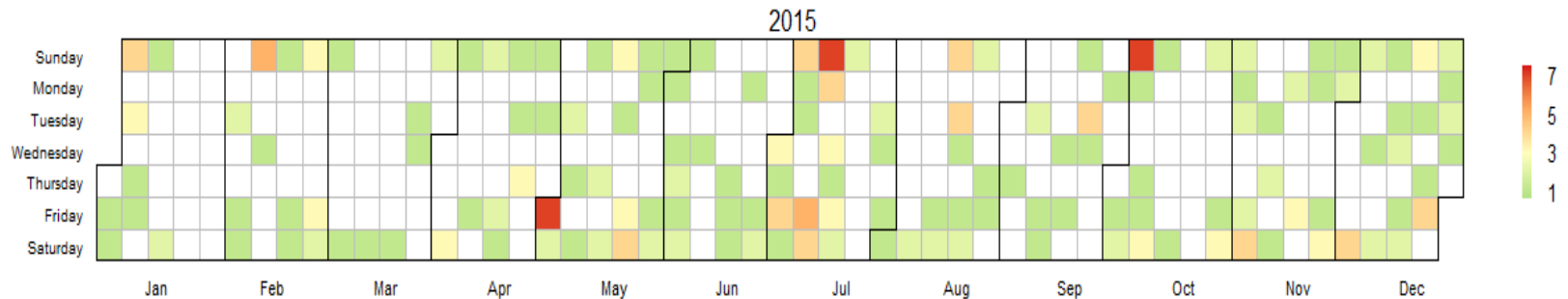
03

➤ 50대 백화점 “보석” 소비 품목 분석

같이 구매되는 상품	빈도수
캐주얼	52
가공식품	25
화장품	19



Calendar Heat Map of 50대 백화점 보석 날짜별 구매 빈도



04



- 40대 50대의 쇼핑 총 금액 차이에 있어서
백화점 이용 금액 차이가 큰 영향을 미친다.



- 40대보다 50대의 **백화점 이용 수가 많고,**
구매 상품 1개의 가격 또한 50대가 더 크다.



- 40대의 '아동'은 어린이날이 있는 5월과 주말에,
50대의 '보석'은 7월 초중순에 잘 팔린다.

Q & A

**THANK
YOU**