

## < 40대, 50대의 쇼핑 업종 소비 분석(7조) >

### ■ 전체적인 구성과 피드백 후 요약.

조 구성원	정소현, 이선목, 박세진, 차승제	발표자 : 박세진
중간발표 이후 수정 할 점	1. 기존에 있던 내용에서 축소. 2. 40대, 50대의 쇼핑 업종에 초점을 맞추기로 함. 3. 그래프 요소 축소 및 필요한 데이터만 사용하기로 함.	
수정 후 전체적인 내용 구성 과정	<p>&lt; 과정 &gt; ( 단계적으로 접근 )</p> <p><b>Step1. 연령대별 쇼핑 빈도수와 연령대별 쇼핑 총 금액 비교</b></p> <ul style="list-style-type: none"> <li>- 쇼핑업종 빈도수는 40대가 50대보다 더 큰 반면, 쇼핑 총 금액은 50대가 40대보다 크다는 사실을 발견.</li> <li>- 왜 이런 결과가 나왔는지를 알아보기 위해 어떤 쇼핑업종을 이용했는지를 파악.</li> </ul> <p><b>Step2. 40대, 50대의 쇼핑분야 빈도수와 쇼핑분야 총 금액 비교</b></p> <ul style="list-style-type: none"> <li>- 총 금액에서 백화점이 가장 높음.(막대그래프를 이용해서 확인)</li> <li>- 백화점에서만이 50대가 40대보다 이용 빈도수가 높다는 사실을 인지.</li> <li>-&gt; 쇼핑 총 금액이 50대가 40대보다 큰 원인을 단순하게 다음 세 가지로 추론.               <ul style="list-style-type: none"> <li>ㄱ. 50대 백화점 이용자가 많은 것</li> <li>ㄴ. 50대가 1인당 상품을 상대적으로 더 많이 구매한 것</li> <li>ㄷ. 50대가 구매한 상품 금액이 큰 것</li> </ul> </li> </ul> <p><b>Step3. 40대, 50대의 백화점 1인당 구매 상품수와 구매금액 비교</b></p> <ul style="list-style-type: none"> <li>- 데이터 확인 후 분석 저해요소 파악 : 백화점 1인당 구매 상품수량을 분석할 때, 40대와 50대의 MAX값이 각각 1202와 5000이 나왔는데 40대의 수량 MAX값인 1202은 상품을 알 수 없고 1개당 금액이 100원 이하였기에 1인당 구매수량을 분석하는데 있어 상식상 맞지 않다고 생각하여 제외하였고, 마찬가지로 50대의 수량 MAX값인 5000도 특정 행사 때 사용하기 위해 대량 구매한 빵으로, 분석에서 제외하였다.</li> <li>- 1인당 구매 상품 수 : 40대와 50대가 Min, 1st Quantile, Median, 3rd Quantile이 같음. -&gt; 두 연령대의 구매 상품 수가 거의 비슷하다고 생각. (뚜렷한 차이점을 찾기 어려움, 표본의 개수도 다르기에 상자그림만으로 정확한 판단은 어렵다. 그러나 물려있다는 것으로 판단 되어 어느 정도 근거가 된다고 생각함.)</li> <li>- 1인당 구매 금액 : 50대가 40대보다 더 크다는 사실을 통해 50대가 구매하는 상품의 가격이 크고, 백화점 이용자수가 많다는 결론 도출.</li> </ul> <p><b>Step4. 40대, 50대의 백화점 소비품목 비교</b></p> <ul style="list-style-type: none"> <li>- 40대와 50대의 소비품목 Top 50을 뽑아본 결과, 구매 품목이 거의 비슷하였다.</li> <li>- 그 중, '아동' 품목이 50대에 비해 40대에 더 구매했고, '보석' 품목이 40대에 비해 50대에 더 구매한 것을 발견 =&gt; 40대의 '아동'과 50대의 '보석'을 분석 =&gt; 같이 구매되는 상품을 확인</li> </ul>	

	<p>=&gt; 각 상품이 일자별, 월별로 언제 잘 팔리는지를 분석.</p> <p><b>Step5. 분석 결과 확인, 결론 도출</b></p> <ul style="list-style-type: none"> <li>- 40대의 '아동'과 50대의 '보석'이 같이 구매되는 상품은 비슷하였다.(캐주얼, 가공식품)</li> <li>- 40대의 '아동'은 어린이날이 있는 5월과 대체로 주말에 상품이 잘 팔렸다.</li> <li>- 50대의 '보석'은 7월 초 중순에 잘 팔리는 것을 확인할 수 있었다.</li> </ul> <p>==&gt; 50대가 40대보다 쇼핑금액이 큰 것은 일부분인 백화점 업종 부분에서 40대보다 빈도수가 높았고 그에 따라 제품 가격이 크기 때문.</p> <p>* 더 나아가 보석과 아동 물품에 대해 같이 구입하는 것이 캐주얼과 가공식품이란 부분에서 착안 + '아동'관련 제품은 5월, '보석'제품은 7월 초에 많이 팔림</p> <p>= &gt; 보석 중 루비 - 7월 탄생석으로써 다이아몬드와 매출을 나란히 한다는 기사를 확인.</p> <p>= &gt; 어린이날 선물로 아동관련 상품 아동 캐주얼 함께 행사 추천</p>
--	---

## ■ 발표자료 설명(PPT자료 참고)

- 주제 및 개요(슬라이드 2)

첫째, 저희 조의 주제는 40대와 50대의 쇼핑 업종 소비를 분석하는 것입니다.

둘째, 저희 조는 데이터를 시각화하고, BoxPlot, WordCloud, calenderHeat를 사용해 데이터를 탐색하고, 모형화 하여 분석하였습니다.

셋째, 그래서 저희는 40대와 50대의 총 구매 금액 차이를 분석하고, 40대와 50대의 특정 상품의 소비 패턴을 분석하였습니다.
- 기본적인 데이터 분석 및 주제 선정 이유 (슬라이드 3)

먼저, 데이터에서 연령대별로 쇼핑 업종을 분석했습니다.

연령대별로 쇼핑 이용 수와 이용 총 금액을 시각화 해본 결과,

쇼핑 이용 수는 40대 50대 30대 순이고, 쇼핑 이용 총 금액은 50대 40대 30대 순으로

이용 수는 40대가 50대 보다 더 큰데, 이용 총 금액은 50대가 40대 보다 더 크다는 사실을 발견했습니다.

그래서 40대와 50대에 초점을 맞춰서 분석하도록 하겠습니다.
- 데이터 시각화 (슬라이드 4)

40대와 50대에 초점을 맞춘 후,

이용 총 금액이 어느 쇼핑 업종에서 크게 차이 나는지 알아보기 위해,

40대와 50대의 쇼핑 업종 이용 수와 이용 총 금액을 시각화하였습니다.

그래프는 단위가 '억/원'으로, 백화점에서 40대와 50대의 이용 금액이 두드러지게 차이가 난다는 사실을 발견했습니다.

또한, 40대와 50대의 백화점 이용 빈도수를 보았을 때, 50대가 40대보다 더 많다는 것을 알 수 있습니다.
- 데이터 분석 1 (슬라이드 5)

그래서 빈도수는 40대가 더 많은데, 총 금액은 50대가 더 많은 이유가 백화점에서 큰 금액 차이가 발생했기 때문임을 알았습니다.

저희 조는 더 나아가, 단순히 50대가 40대보다 백화점 이용 수가 많아서 인지,

1인당 상품 구매 수량이 많아서 인지, 구매하는 상품의 금액이 비싼건지 더 자세하게 알아보도록 하겠습니다.
- 데이터 분석 2 (슬라이드 6)

우선 백화점에서의 1인당 상품 구매 수량을 알아보도록 하겠습니다.

근데, 여기서 하나 말씀 드리자면, 40대의 원데이터에서는 수량을 1202까지 사는 사람이 있었습니다.

확인해 본 결과 상품을 알 수 없었고, 1개당 금액이 100원 이하였기에 1인당 구매 수량을 분석하는데 불필요하다고 생각하여 제외하였습니다.

또한, 50대의 원데이터에서는 수량을 5000개 까지 사는 사람이 있었습니다.

확인해 본 결과 특정 행사(빵 나눔잔치, 기업 행사 대량 구매등) 때 사용하기 위해 빵을 구입한 것으로, 1인당 구매 수량을 분석하는데 있어서 맞지 않다고 생각하여 제외하였습니다.

분석한 결과는 40대와 50대의 Min, 1st quantile, 3rd quantile, Median이 서로 같았습니다.

평균이 미세하게 다르긴 하지만 Min과 median, 1\*3 quantile이 같아 몰려있다고 생각하였고, 1인당 구매 수량이 거의 같다고 해도 된다는 판단하였습니다.

### ● 데이터 분석 3 (슬라이드 7)

그 다음으로 백화점에서의 1인당 소비 금액을 알아보도록 하겠습니다.

앞서 제외한 데이터는 뺀 결과로, 50대의 평균이 40대의 평균보다 65000원 정도 더 높은 것을 알 수 있었습니다.

즉, 1인당 구매 수량은 거의 같으나, 소비 금액은 50대가 더 큰 것을 보아,

50대가 구매하는 상품 가격이 40대가 구매하는 상품 가격보다 크다는 사실을 분석 할 수 있었습니다.

백화점에서 50대가 이용수가 많고, 50대가 구매하는 상품 가격이 크기에,

백화점에서 금액 차이가 크게 났음을 알 수 있었습니다.

여기까지가 저희 조의 궁극적인 문제 분석이었고, 부가적으로 40대와 50대의 백화점 소비 품목을 보아 특정 상품에 따른 소비 패턴을 알아보도록 하겠습니다.

### ● 데이터 분석 3 (슬라이드 8)

40대와 50대의 백화점 소비 품목 top 50을 본 결과, 40대와 50대의 크게 달라지는 품목들이 존재 하지 않아,

그 중 '아동' 품목이 50대에 비해 40대에 더 구매했고, '보석' 품목이 40대에 비해 50대에 더 구매한 것을 발견해,

40대의 '아동'과 50대의 '보석'을 중점으로 분석했습니다.

### ● 데이터 분석 4 (슬라이드 9)

뭐 거창한 분석은 아니지만 그 특정 품목과 함께 구매되는 상품과 그 특정 품목이 일자별,월별로 언제 잘 팔리는 지 분석해 보도록 하겠습니다.

40대의 '아동'을 먼저 보시면, 같이 구매되는 상품들을 wordcloud로 모형화 한 뒤, top3의 빈도수를 봤고, calenderheat를 통해 언제 잘 팔렸는지 알아봤습니다.

분석해 본 결과, 유독 5월1일과 5월6일에 많이 판매됨을 알 수 있습니다.

그래서, 더 많은 수익을 위해, 5월 행사인 어린이날 선물로 아동 상품과 아동 캐주얼을 함께 행사하는게 어떨까 추천하는 바입니다.

### ● 데이터 분석 5 (슬라이드 10)

다음으로 50대의 '보석'을 보시면, 같이 구매되는 상품들을 wordcloud로 모형화 한 뒤, top3의 빈도수를 봤습니다.

앞에서 보신 같이 구매되는 상품과 캐주얼, 가공식품이란 부분이 비슷하다는 사실을 알 수 있습니다.

그리고, calenderheat을 통해 언제 잘 팔렸는지 분석해 본 결과, 유독 7월 초에 많이 판매됨을 알 수 있습니다.

보석 중 루비가 7월의 탄생석이라고 합니다.

그래서, 7월이 오기 전에, 루비로 여러 악세사리를 만든다면 더 많은 수익을 볼 수 있지 않을까 합니다.

### ● 결과 도출 (슬라이드 11)

저희 조가 분석한 결론을 정리해보겠습니다.

40대와 50대를 분석한 결과, 빈도수에 비해 쇼핑 총 금액이 차이가 컸고,

백화점 이용 금액 차이가 큰 영향을 미쳤다는 사실을 알아냈습니다.

또한 백화점 이용 금액 차이는 50대가 백화점 이용 수가 많고, 구매 상품 1개의 가격 또한 컸기에 크게 나타났음을 알았습니다.

추가로, 40대와 50대의 특정 상품을 잡아

40대의 아동은 어린이날이 있는 5월과 주말에, 50대의 보석은 7월 초중순에 잘 팔린다는 사실을 알았습니다.

쇼핑 업종은 백화점에서 큰 수입을 얻는다고 볼 수 있는데, 이렇게 40대와 50대의 소비 패턴을 파악해, 그에 알맞은 행사나 제품들을 만들어낸다면, 큰 효과를 얻을 수 있다고 생각하는 바입니다.