**BDA 공모전 인사이트 보고서**

**팀명 :** 이쿠조

**멤버들 :** 신재우, 김세원, 박희주, 남화승

**기획서 요약**

우리의 주요 목표는 5월과 6월에 네이버 및 11번가에서 제공된 데이터를 바탕으로 CJ제일제당 상품의 매출 증가를 위한 효과적인 마케팅 전략을 제안하는 것이다.

이를 위해 소비자들의 행동 패턴과 특성을 파악하는 것이 중요하다고 판단되며, 컬럼 값이 판매량에 미치는 영향을 분석한 뒤, 이를 활용하여 소비자 특성에 맞는 마케팅 전략을 제시할 계획이다. 단, 현재 제공된 데이터가 네이버와 11번가의 데이터로 한정되어 있어 마케팅 전략의 범위가 제한될 수 있다는 점을 고려해야 한다.

먼저 CJ제일제당 상품의 네이버 및 11번가 이커머스 고객 데이터에서 판매량에 유의미한 영향을 미칠 것으로 예상되는 변수들을 칼럼으로 선정하고, 데이터 전처리를 통해 데이터를 정제한 후 데이터 시각화를 활용하여 변수 간의 관계를 살펴야 한다.

각 컬럼별로 판매량에 유의미한 영향을 미치는지에 대한 가설을 설정하고, 분산분석(ANOVA), t-test 등 상황에 적합한 통계적 방법을 사용하여 가설의 타당성을 검증해야 한다. 유의미한 가설을 바탕으로 마케팅 전략과 프로모션 아이디어를 도출하고, 이를 실무에 적용하여 기업의 이익을 극대화하는데 도움을 줄 예정이다.

그러나 현재 전체 데이터가 제공되지 않아 구체적인 마케팅 전략 및 결론을 도출하기 어려우므로, 이 보고서에서는 컬럼 별 상황을 예시로 들어 각 상황에 맞는 마케팅 전략과 예상되는 기대 효과를 작성하였다.

**사용할 분석 방법들**

차후에 나올 분석 방법들에 앞서 보고서에서 사용할 분석 방법들에 대한 구체적인 세부 내용을 설명하겠다. 보고서에서 사용될 분석 방법들은 총 두가지로 ANOVA 분산 분석과 two-sample t-test 분석이다.

**Two-sample t-test :**

Two-sample t-test 분석은 데이터가 두 가지 종류 (A와B) 로 나뉘었을 때 의미 있는 차이인지 아닌지를 결정해주는 분석방법이다. 해당 분석의 가설은 아래와 같이 쓰인다.

귀무가설은 이며 대립가설은 이다. 는 모평균을 의미하며 는 그룹 A의 모평균, 는 그룹 B의 모평균을 의미한다. 파이썬 의 패키지인 SciPy 패키지를 이용해서 주어진 유의수준 하에 데이터를 A와 B로 나누었을 시에 A와 B에 차이가 있는지에 대해서 결정해주는 분석방법이다. 파이썬 에서는 하단과 같이 실행이 된다. 예시 데이터로는 그룹 A = [12, 15, 18, 24, 28] 그리고 그룹 B = [20, 25, 30, 35, 40]을 사용했다.

Text

Description automatically generated

유의수준 0.95하에 그룹 1과 그룹 2간의 차이가 있다는 분석 결과가 나왔기 때문에 귀무가설을 기각하며 그룹 1과 그룹 2로 나눌 시에 유의미한 차이가 있다는 결론을 지을 수가 있다.

**ANOVA test :**

Two-sample t-test 는 오로지 두 가지의 그룹만을 고려한 것이며 더 많은 그룹들로 나눠야 할 시에 ANOVA test 혹은 ANOVA 분산분석을 사용하면 된다. 예시로 그룹 A = [12, 15, 18, 24, 28], 그룹 B = [20, 25, 30, 35, 40] 그리고 그룹 C = [30, 36, 42, 44, 45] 를 사용하겠다. 가설은 아래와 같이 수립될 수 있다.

귀무가설은 이며 대립가설은 이다. 는 모평균을 의미하며 는 그룹 A의 모평균, 는 그룹 B의 모평균, 는 그룹C의 모평균을 의미한다. Two-sample t-test 과 동일하게 파이썬 의 패키지인 SciPy 패키지를 이용해서 주어진 유의수준 하에 분석을 실행할 것이며, 위의 예시에 적용한 코드와 결과는 다음과 같다.

Text

Description automatically generated

유의수준 0.95하에 그룹 A, B 그리고 C간의 차이가 있다는 분석 결과가 나왔기 때문에 귀무가설을 기각하며 그룹으로 나눌 시에 유의미한 차이가 있다는 결론을 지을 수가 있다.

**주어진 데이터의 정의**

예시 데이터로부터 ‘거래처’, ‘주문일’, ‘사은품여부’, 주문/배송상태’, ‘상품명’, ‘상품수량’, ‘마감구분’ 컬럼이 있음을 확인하였다. 주어진 데이터를 기반으로 파생변수를 추가하였고, 최종적으로 사용될 데이터의 칼럼은 다음과 같다.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 상품명 | 거래처 | 주문  월 | 주문  일 | 사은품 여부 | 주문  상태 | 용량 | 묶음  수량 | 마감  구분 | 상품  수량 | 총  수량 |

**상품명 :**

해당 상품명을 의미한다. 주어진 데이터는 ‘고메치킨 후라이드 550Gx3개’, ‘고메 바삭촉촉한 깐풍기 450G/냉동’ 과 같은 형식이기 때문에, re 패키지를 이용해서 오로지 상품명만 나오게끔 만들 것이다. 예시 데이터에 이것을 적용한 파이썬 코드와 결과는 다음과 같다.

Graphical user interface, text, application

Description automatically generated

다음 과정을 거치면 모든 데이터를 상품명으로 분류 가능하다. 다수의 상품들이 있기 때문에 오로지 상품명으로 분석하는 것은 불가능하다고 판단되고, 추후 나머지 컬럼들과 관련 지어 결과를 도출하는 데에 사용할 것이다.

**거래처 :**

주어진 데이터에서 이미 어떤 거래처를 사용하는지에 대한 데이터가 있기 때문에 구분없이 해당 컬럼을 사용하면 된다. 거래처 컬럼의 정의는 바로 상품을 구매할 때 어떤 거래처에서 구매가 진행되는지 (예: 네이버, 11번가) 이다.

**주문 월 :**

주어진 예시 데이터에서는 20220506이며 해당 데이터 값은 2022년 5월 6일을 의미한다. 날짜 형태로 컬럼을 변형시킨 뒤에 해당 월만 고려해서 데이터 수집이 가능하다. 다음은 해당 내용을 파이썬에 적용시켜본 것이다.

Text

Description automatically generated with medium confidence

**주문 일:**

주문 월이 이 컬럼의 파생변수로 생각해도 되며, 주어진 예시 데이터에서 이미 주문 일까지 보여주고 있기 때문에 구분 없이 해당 컬럼을 사용하면 된다. 주문 일 컬럼의 정의는 바로 어떤 특정 날짜에 상품을 구매했는지 이며 주문 월 컬럼의 정의는 그 날짜를 월별로 나눈 것이다.

**사은품 여부 :**

예시 데이터에서는 사은품의 여부를 Y와 N으로 구분하고 있기 때문에 query 함수를 응용해서 Y와 N 두 그룹으로 데이터를 나눌 수가 있다. 다음과 같이 파이썬에서 데이터를 나누면 된다.

**Text

Description automatically generated**

**주문 상태 :**

주문 상태의 컬럼은 상품을 구매한 뒤에 구매 선택을 취소했는지 안 했는지에 대한 여부다. 보고서를 작성하면서 의논해본 결과 주문 상태를 통해서 가능한 분석은 현재 레벨에서는 없다는 결론을 내렸기 때문에 넘어가겠다.

**용량 / 묶음 수량 / 마감 구분:**

주어진 데이터에서 상품명을 살펴보면 용량, 묶음 수량 그리고 마감구분에 대한 내용까지 있다. 용량이란 해당 상품의 총량을 무게로 구분하는 것이며 묶음 수량이란 해당 상품이 몇개로 팔렸는지에 대한 여부이다. 마감 구분은 냉동, 냉장 혹은 상온을 구분해주는 컬럼이다. 위의 상품명 컬럼과 같이 여기에서도 re 패키지를 사용해서 분류를 해야 한다. 다음은 해당 내용을 파이썬에 적용시킨 것이다.

Text

Description automatically generated

**상품 수량 / 총 수량 :**

상품 수량이란 것은 소비자가 구매를 했을 때 상품을 얼마나 많이 구입했는가를 수치로 나타낸 것이며 예시 데이터에서 보이듯이 구분을 할 필요는 없다. 하지만 이 수치는 묶음 수량까지 고려를 못한 수치이며 그것까지 고려한 새로운 파생변수가 바로 총 수량이다. 총 수량 = 상품 수량 X 묶음 수량 이며 매출에 대한 정확한 데이터가 없을 시에 이 컬럼을 사용하면 된다. 매출은 총 수량 X 가격 으로 계산이 되기 때문이다. 파이썬에서 총 수량은 다음과 같이 구현해내면 된다.

Table

Description automatically generated

**컬럼 별 분석과 전략**

**1. 거래처** :

**가설** : 상품이 판매되는 거래처에 따라서 상품의 수요가 다를 것이다.

**분석 방법**  
거래처의 경우에는 요인(Factors) 밖에 없기 때문에 데이터들이 요인 1 인 네이버, 그리고 요인 2 인 11번가로 나눠진다. 우리가 궁금한 점은 거래처에 따라서 상품의 수요가 다를 것인지에 관한 것이기 때문에 귀무가설을 거래처로 나눈 평균이 똑같다, 그리고 대립가설을 평균에 차이가 있다로 정하면 된다.

Diagram

Description automatically generated with low confidence

우선적으로는 시각화를 통해서 X 변수는 거래처, Y 변수는 매출을 나타내는 변수로 정하면 된다. 두 거래처의 평균에 대한 확연한 차이가 보인다면 귀무가설을 기각하면서 마케팅 전략으로 넘어가면 되지만 그것이 아니라면 ANOVA를 통한 분산분석을 이용해서 차이가 존재하는지를 살피면 된다.

**분석목적**  
위의 분석에서 거래처로 소비자들을 나눌 수가 있다는 가정하에 특정 거래처를 사용함으로써 그 거래처의 이용자들의 특징 혹은 공통점 등을 살펴서 소비자들의 성격을 파악을 할 수가 있다. 소비자들의 특징이 중요한 이유는 바로 그들 대상으로 마케팅 전략을 구성해야 하기 때문이며 그러기 위해서는 소비자들이 대체적으로 어떠한 공통점을 가졌는지를 알아야 하기 때문이다. 아래에는 주어진 거래처들인 네이버와 11번가로 나눌 수가 있다는 상황에서 분석 결과와 인용이 가능한 마케팅 전략, 그리고 그에 대한 예상 기대효과를 살펴보겠다.

**네이버 쇼핑 분석 결과** :   
2022년도 6월 ‘COSIN’ 에 작성된 글에 따르면 핵심 이용 층이 20 ∙ 30대로 비중의 50% (출처 1) 이상을 사용되는 것을 나타내고 있다. 20 ∙ 30대들은 흔히 말하는 MZ 세대에 속해 있으며 개인주의 성향이 강하며 이득이 있어야만 구매를 한다는 성격이 특징이다. 또 다른 특징으로는 이들이 SNS를 굉장히 자주 이용하기에 인플루언서들에 대한 영향이 매우 크다는 점이다. 소비자들의 성격에 대해서 파악이 완료가 되었기에 이들에게 특화된 마케팅 전략을 세우면 된다.

**마케팅 전략 예시 1** :   
하단의 컬럼들과 함께 분석을 해본 결과 네이버 쇼핑에서 여름 직전 달인 5월달 대체적으로 CJ제일제당의 다이어트 상품들인 곡물 쉐이크 프리미엄 선식 보틀의 판매량이 높았다는 예시를 들겠다. 여름 몸을 위해서 다이어트를 했다는 것을 알 수가 있으며 네이버 쇼핑에서 나온 결과이기에 20 ∙ 30대들이 핵심 소비자들이라는 것도 알 수가 있다. 묶음 판매를 진행시켜서 한번에 10개씩 구매하면 싸게 판매하는 행사를 진행시켜서 소비자 입장에서는 이득이라는 인식을 심어줄 수 있기에 소비자들의 요구를 충족시키는 마케팅 전략을 진행할 수가 있다. 예상하는 기대 효과는 곡물 쉐이크 프리미엄 선식 보틀의 수요 수량의 증가이며 이것으로 매출을 증가할 수가 있다는 결론을 내릴 수가 있다.

**마케팅 전략 예시 2** :  
위와 똑같은 예시로 네이버 쇼핑에서 5월달에 곡물 쉐이크 프리미엄 선식 보틀의 판매량이 높았다는 가정을 하며, 그에 대한 이유는 다이어트라고 들겠다. 20 ∙ 30대들이 인플루언서들에 대한 영향을 많이 받는다는 점을 이용하여서 상품 이미지에 다이어트에 대한 인식이 매우 좋은 인플루언서와 함께 상품을 판매하는 마케팅 전략을 진행시키면 된다. 20 ∙ 30대들이 해당되는 인플루언서를 믿으면서 상품을 구매하게 되기에 예상하는 기대 효과는 당연스레 곡물 쉐이크 프리미언 선식 보틀의 수요의 증가이기에 이 마케팅 전략을 통해서 매출을 올릴 수가 있다는 결론을 내릴 수가 있다.

**11번가 분석 결과** :  
네이버 자료를 참조한 똑같은 글에 의하면 11번가의 핵심 이용 층은 40 ∙ 50대들이다 (출처 1). 이들의 대부분은 직장인 혹은 은퇴를 준비하는 단계에 도달한 사람들이기에 구매력이 엄청나다는 점과 미래에 대한 투자가 우선이라는 특징을 살필 수가 있다. 그렇다는 것은 구매를 통해서 얼마나 이득을 쟁취할 수 있기보다는 상품의 품질에 더 신경을 많이 쓴다는 분석을 지을 수가 있다. 이것에 특화된 마케팅 전략들을 구성해서 판매를 하면 된다.

**마케팅 전략 예시 3** :  
네이버 쇼핑에 비해서 11번가에서 CJ 제일제당의 건강 상품인 ‘byocore 피부 유산균’의 매출이 높았다는 예시에서 마케팅 전략을 세우겠다. 위의 분석 결과를 통해서 상품의 품질에 대해서 더 중요시하기에 상품의 인식을 높여주는 마케팅 전략이 필요하며, 이를 토대로 광고 마케팅 전략을 추구할 수가 있다. 예시 2 와 똑같이 40대로 소비자들과 비슷한 나이 대이지만 건강한 이미지를 유지하는 공유와 같은 연예인을 상품 이미지에 상품과 함께 표시하면 예상되는 효과는 소비자들이 연예인이 가진 이미지 때문에 다른 경쟁사 상품들 보다는 해당 상품을 선택하게 될 것이며 매출은 증가하게 된다는 결론을 지을 수가 있다.

**거래처 컬럼에 관한 결론** :  
데이터가 주어지지 않았기 때문에 섣불리 결론 짓기가 어렵다는 것은 사실이다. 하지만 오로지 거래처를 컬럼으로 사용할 시에 생기는 문제는 바로 상품들의 범위가 매우 넓다는 단점이다. 위의 예시로 다이어트 식품과 건강 식품을 예시로 줬지만 20 ∙ 30대, 40 ∙ 50대 상관없이 다이어트 식품과 건강 식품들의 수요가 높아질 수도 있다는 것이다. 즉 예측은 가능하지만 그에 대한 오차가 매우 높을 것이다. 그렇기 때문에 다른 컬럼들도 함께 고려를 해서 분석을 해야 된다.

**2. 주문일 :**

**가설** : “주문일에 따라 주문량에 차이가 있을 것이다.”

**분석 방법**  
주문일의 경우 2022년 5월 1일부터 2022년 6월 30일까지, 총 61일의 데이터로 나눌 수 있다. 위에 작성한 가설에 따른 귀무가설과 대립가설은 다음과 같다.

우리가 궁금한 점은 주문일에 따라서 상품의 수요가 다른 지이기 때문에 귀무가설은 “주문일별 판매량 평균이 똑같다”, 대립가설은 “주문일별 판매량 평균이 적어도 하루는 다르다” 로 설정한다.

이를 검증하기 위해서는 1차적으로 시각화를 활용하면 된다. 주문일 컬럼을 기준으로 그룹화하여 판매량의 합계를 구한 후, X 변수로는 주문일, Y 변수로는 총판매량으로 설정한다. 주문일별로 판매량 평균에 대한 확연한 차이가 보인다면 귀무가설을 기각하면서 하단에 작성한 마케팅 전략을 활용하면 된다. 그러나 만약 시각화를 통해 눈에 띄는 차이를 보이지 않았다면 ANOVA를 통한 분산분석을 활용하여 차이가 존재하는지를 살피는 단계로 넘어간다.

**분석 목적**  
앞서 설정한 가설을 바탕으로, 우리는 주문일에 따라 주문량에 차이가 있다는 점을 활용하여 각 주문일의 특징을 파악할 수 있다. 주문량이 눈에 띄게 차이나는 주문일은, 날짜를 통해 그 특징을 파악하여 그에 알맞은 프로모션을 진행하면 된다. 예를 들어, 주문일5의 판매량이 눈에 띄게 차이 나는 경우, 우리는 이것이 어린이날로 인한 것임을 알 수 있다. 따라서 해당 일자에는 어린이날 플랫폼에 방문하는 소비자층의 특성에 맞는 마케팅 전략을 수립하면 된다.

**마케팅 전략** **1 예시 1, 어린이 날? 어른이 날! :**  
어린이 날에 이커머스 플랫폼에 방문하는 고객이 어린이가 아닌 부모라는 것을 고려하면, “어른이 날”이라는 문구로 방문 고객의 구매 욕구를 자극할 수 있다. 어린이날 상품과 어른들이 좋아할 상품을 결합한 기획전을 펼침으로써, 어린이날 상품과 더불어 추가적인 매출을 끌어올릴 수 있다. 구체적인 예시를 통해 살펴보자.

**상품A = 어린이날 상품 :** 어린이 날인 5월 5일에 상품 A 에 관한 매출이 높다는 데이터가 나왔으면 상품 A의 소비자들은 대부분 5월 5일 어린이날과 관련이 있다는 분석을 할 수가 있다. 더 정확하게는 상품 A의 소비자들은 어린이를 가진 가정이라고 좁힐 수가 있다.   
**상품B = 안주세트 :** 대개 어린 자녀를 둔 부모의 나이대는 3~40대이므로 이들이 주로 찾는 네이버 쇼핑에서 판매량이 높은 간편 냉동 식품을 상품B로 설정한다.   
⇒ 어린이날에 상품 A를 사려는 부모에게 “어른이 날”이라는 문구와 함께 상품A와 상품B를 결합한 상품C 노출시킴으로써, 상품A로 그쳤을 판매에 상품B에 판매까지 유도할 수 있다.  
  
예시로 2022.05.05에 고메 크리스피 핫도그의 주문량이 급증했다는 가정을 들겠다.  
상품 A : 고메 크리스피 핫도그  
상품 B : 고메 치킨 세트 (고메 순살크리스피 치킨 고추크런치, 고메치킨 후라이드)  
→ 상품 C : 고메 어린이 날? 어른이 날! 세트 (=상품A+상품B)

**마케팅 전략 예시 2, 11번가 월간(月刊) 십일절 활용 :**  
11번가에서 매달 11일에 진행하는 11번가 대표 프로모션 데이로, 2019년부터 누적 거래액 3조 이상을 기록할 만큼 많은 이들의 관심을 모으는 행사이다. 즉, 매달 11일은 판매량이 급격하게 늘어난다는 뜻이다. 2022년 5월 11일, 6월 11일에 많이 팔린 상품을 내림차순 정렬한 후, 많이 팔린 상품의 브랜드를 분석한 후, 2023년 5월 11일, 6월 11일에는 동일 브랜드 내 판매량이 저조했던 상품을 기획상품으로 묶어 판매한다.

**3. 사은품 여부 :**

**가설 :** 두가지 상품이 같은 상품인 경우, 사은품이 있는 상품의 수요가 더 높을 것이다.

**분석방법**  
사은품의 컬럼의 경우 사은품이 있다 없다 로 두가지 데이터 형태로 주어진다. 이런 경우 t-test 를 사용해서 사은품을 포함시키면 의미 있는 차이가 있는지를 찾을 수가 있다. 귀무가설은 사은품 포함 시킨 것과 아닌 것의 모평균들이 똑같다는 것이며 대립가설은 두 모평균들이 똑같지 않다는 것이다.  
적절한 유의수준을 고른 후 사은품으로 나눠도 되는지 안되는지에 대한 결정을 내릴 수가 있으며, 이를 통해서 소비자들의 성격을 결정할 수가 있다.  
  
**분석 목적**  
사은품이랑 같이 판매한 상품의 매출이 올랐다는 가설, 즉 위의 귀무가설이 기각되었다는 가정하에 해당 상품은 사은품을 포함시키면 의미 있는 차이를 만들 수가 있다. ­­해당 상품의 소비자들은 이 상품을 사은품의 가격까지 더 내면서까지 필수로 필요하다는 분석을 할 수가 있다. 이 점을 이용해서 매출이 적은 다른 상품과 함께 사은품으로 판매도 가능하며, 아예 상품 자체의 가격을 올려서 매출을 올리는 전략도 시행할 수가 있다.

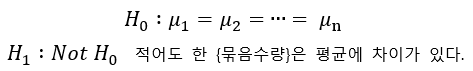
**마케팅 전략 예시 1 :**  
CJ제일제당에서 판매하는 상품인 컨디션 스틱 컨디션 맛이 사은품과 함께 판매 시 매출이 올랐다는 상황을 예시로 두겠다. 이것을 통해서 우리는 컨디션 스틱 컨디션 맛 상품의 소비자들이 손해를 보더라도 그들에게는 필요한 제품이라는 분석을 할 수가 있다. 여기에서 제시할 수 있는 마케팅 전략은 바로 가격을 올리는 것이다. 수요의 가격 탄력성이 비탄력적이라는 분석을 할 수가 있다. 그러므로 예상하는 기대효과는 가격을 올려도 매출이 올라간다는 것이다.

**마케팅 전략 예시 2 :**  
위의 예시와 같이 컨디션 스틱 컨디션 맛이 사은품과 함께 판매 시 매출이 올랐다는 가정과 컨디션 스틱 애플 스틱 맛 상품의 매출이 적다는 예시를 두겠다. 비슷한 상품이지만 애플 스틱 맛의 매출이 적다는 것은 소비자들에게 노출이 많이 안됐다는 분석을 할 수가 있으며, 이를 기반으로 마케팅 전략을 세울 수가 있다. 컨디션 맛의 사은품으로 애플 스틱 맛 상품을 마케팅 전략으로 세울 수가 있으며, 이에 대한 예상 기대 효과는 바로 애플 스틱 맛에 대한 노출도의 증가이며 이것은 애플 스틱 맛의 매출의 증가로 이어지게 된다.

**4. 묶음 수량 , 용량, 상품명 (3가지 컬럼을 활용) :**

**가설 :** 각 상품을 용량으로 그룹화하였을 때, 묶음 수량에 따라 판매량이 다를 것이다.

**분석 방법**  
각 상품별 묶음 수량은 1개부터 n개까지 분류한다고 할 때(n : 각 브랜드에서 설정하는 최대 묶음 개수), 위 가설을 통해 귀무가설과 대립가설은 다음과 같이 설정할 수 있다. 우리가 궁금한 것은 묶음 수량에 따라서 판매량이 다른 지이기 때문에 귀무가설은 “묶음 수량으로 분류하였을 때, 판매량 평균이 똑같다”, 대립가설은 “묶음 수량으로 분류하였을 때, 적어도 한 분류의 묶음 수량의 판매량 평균이 다르다” 로 설정한다.



이를 검증하기 좋은 방법은 시각화이다. 주어진 데이터에서 상품명 컬럼을 기준으로 그룹화하고, 용량 컬럼을 기준으로 또 한번 그룹화 한 후,  X 변수로는 묶음수량, Y 변수로는 총판매량을 설정한다. 이렇게 하면, 같은 상품명의 상품들에 대하여 용량 별 그래프를 시각화 할 수 있다. 다시 말해, 전체 상품의 종류가 r개이고, 각 상품의 용량 종류가 i개( i = 1, ```, p) 라고 하면, r \* i 개의 그래프를 시각화 할 수 있다.   
  
이렇게 시각화한 그래프를 통해, 각 상품의 용량별로 판매량이 줄어들기 시작하는 묶음 수량, 즉 변곡점이 존재함을 알 수 있다.(변곡점 : 판매량이 줄어들기 시작하기 직전의 묶음 수량) 그래프가 꺾이는 지점을 해당 브랜드의 변곡점으로 정의하고 하단의 마케팅 전략을 진행한다.  
  
우리가 예상하는 분석 결과의 예시로 비비고 왕교자 용량 종류 :350g,  455g, 1.05kg, 1.5kg, 총 4가지 용량의 상품에 대하여, 각각의 묶음 수량 변곡점은 다음과 같다.   
350g : 5개 / 455g : 5개 / 1.05kg : 3개 / 1.5kg : 2개  
⇒ 해당 값을 활용하여 우리는 여러가지 세일즈 및 프로모션 전략을 펼 칠 수 있다.

**분석 목적**  
상품의 용량별로 그룹화하여 묶음 개수 별 판매량의 차이를 분석함으로써 우리는 판매량이 줄어들기 시작하는 특정 시점, ‘변곡점’을 찾아낼 수 있다. 변곡점은 사람들이 어느 정도의 묶음수량까지 구매 의사가 있는지 확인할 수 있으므로, 상품별로 변곡점을 적극적으로 활용하면, 소비자로부터 최대한의 소비를 이끌어낼 수 있다. 우리는 최대한의 소비를 유도하기 위한 마케팅 전략을 다음과 같이 수립해보았다.

**마케팅 전략 예시, 시각화 자료를 기반으로, 그래프의 기울기에 따라 변곡점의 묶음수량을 팔거나 변곡점+1 지점의 묶음수량을 판매한다 :**  
비비고 왕교자 1.05kg 상품의 묶음개수 변곡점이 3개라고 가정할 때,  다음과 같은 세일즈 및 마케팅 전략을 펼칠 수 있다.

1. 그래프의 기울기가 급한 경우 : 그래프의 기울기가 급하다는 것은 소비자가 가격 탄력적이라는 것이다. 묶음 수량이 소비자의 결정에 미치는 영향이 크다는 뜻으로, 아무리 프로모션을 많이 진행한다고 하더라도 변곡점 이상의 묶음 수량을 판매할 때 매출 증대를 유도할 수 없다. 따라서 이 경우는 정확히 변곡점에 해당하는 상품을 이커머스 검색 결과 상단에 노출시키는 마케팅 전략을 활용한다.

2. 그래프의 기울기가 완만한 경우 : 그래프의 기울기가 완만하다는 것은 소비자가 가격 비탄력적이라는 것이다. 즉, 약간의 혜택이 있다면 소비자는 가격을 더 지불하더라도 손쉽게 변곡점 이상의 수량을 구매하기 때문에, 매출 증대를 유도할 수 있다. 이 때 우리는 묶음수량이 {변곡점+1}개인 상품의 단위 가격을 낮춤으로써 소비자의 구매를 유도할 수 있다. 다만, 해당 소비자는 가격 비탄력적이므로 단위가격을 큰 폭 낮추는 것보다는 소폭만 낮춘 후 검색 결과 상단에 노출시킴으로써 제품 클릭 수를 높이는 것을 목표로 한다.

**5. 마감 구분 :**

**가설 :** 상품의 마감구분에 따라서 상품의 수요가 다를 것이다.

**분석 방법**마감구분의 경우에는 데이터가 냉장, 냉동, 상온으로 구분된다고 볼 수 있다. 상품의 마감구분에 따라 상품의 수요가 다를 것인지가 조사 목적이기 때문에 '마감구분으로 나누었을 때 판매량 평균이 모두 같다'를 귀무가설로, '마감구분으로 나누었을 때 적어도 하나의 판매량 평균이 다르다'를 대립가설로 둔다.

**Text, letter

Description automatically generated**

우선 x축은 마감구분, y축은 매출을 나타내는 변수로 시각화 한다. 시각화한 자료를 통해 대략적인 판매량 차이를 확인한 후 ANOVA를 통한 분산분석을 이용해 귀무가설을 채택할 것인지, 기각할 것인지를 확인한다.

**분석 목적**

ANOVA 분산분석을 통해 마감구분에 따라 상품의 수요가 차이가 있을 경우, 냉장/냉동/상온 제품을 주로 소비하는 고객들의 특징 또는 공통점을 예상해 볼 수 있다. 이를 통해 마감구분에 따른 소비자의 특성을 이용해 마케팅 전략을 세울 수 있게 되고 이는 판매량의 증대로 이어질 것이라 볼 수 있다.

**마케팅 전략 예시 1, 냉동 닭가슴살 → 헬스인 타겟팅 :**

냉동 닭 가슴살의 주요 소비자 중 하나를 헬스 하는 사람으로 특정 할 수 있다. 따라서 이를 활용하여 헬스 관련 인플루언서 마케팅을 한다면, 헬스에 관심이 있는 사람들에게 확실한 광고 효과를가진다고 판단할 수 있다. 타 플랫폼을 통한 광고 시, 네이버 쇼핑 또는 11번가 판매 링크를 첨부함으로써 소비자가 네이버 쇼핑, 11번가로 유입하도록 유도한다.

**마케팅 전략 예시 2, 1~2인 가구, 냉동 HMR 하루 끼니 기획전 :**

2020년 닐슨코리아의 조사에 따르면, CJ 제일제당이 냉동 HMR 시장을 가장 많이 점유하고 있다. HMR 주 소비자층이 1인가구와 2인가구라는 점을 생각해보았을 때, 이들을 타겟으로 세일즈 전략을 펼치는 것이 효과적이다. 마크로밀엠브레인의 2021년 소비자 인식 조사에 따르면 HMR을 주 1회 이상 섭취한다는 조사 결과에 따라, 하루 끼니와 간식을 해결할 수 있는 “하루 끼니 기획전”을 제안한다. CJ의 냉동 HMR로는 고메 중화와 같은 식사 상품도 있지만, 고메 핫도그 / 치킨과 같이 간식 또는 안주거리로 요긴한 상품 또한 있다. 하루 끼니 기획전은 마트에서 “10개 골라 담으면 8750원”과 같은 행사를 하는 것과 같은 기획전처럼 진행된다. 위의 분석 결과, 냉동에서 많이 팔린 상품 중 비슷한 가격대의 상품을 제시하여 소비자들이 n개를 꽉 채워 구매하도록 유도함으로써 매출 증대를 꾀할 수 있다.

**제언 1 (구체적인 시간까지 고려한 주문 시간 컬럼) :**

위의 예시 데이터에 따르면 구체적인 주문 시간을 추출할 수 있는 데이터는 안 주어졌다. 그렇기 때문에 구체적인 구매 시간까지 주어졌다는 가정하에 제언을 진행하겠다.

**가설** : 주문시간에 따라 주문량에 차이가 있을 것이다

주문시간은 하루 24시간을 30분 단위로 나누어, 총 48개의 데이터로 분류하여 분석한다. 위에 작성한 가설에 따른 귀무가설과 대립가설은 다음과 같다.

Text, logo

Description automatically generated with medium confidence

(주문시간대 : 00:00:00~00:29:59 → 주문시간1 / 00:30:00~00:59:59 → 주문시간 2 등)

우리가 궁금한 점은 주문시간에 따라서 상품의 수요가 다른 지이기 때문에 귀무가설은 “주문시간별 판매량 평균이 똑같다”, 대립가설은 “적어도 하나의 주문시간대의 판매량 평균은 다르다” 로 설정한다.

이를 검증하기 위해서는 1차적으로 시각화를 활용하면 된다. 주문시간 컬럼을 기준으로 그룹화하여 판매량의 합계를 구한 후, X 변수로는 주문시간, Y 변수로는 총판매량으로 설정한다. 주문시간별로 판매량 평균에 대한 확연한 차이가 보인다면 귀무가설을 기각하면서 하단에 작성한 마케팅 전략을 활용하면 된다. 그러나 만약 시각화를 통해 눈에 띄는 차이를 보이지 않았다면 ANOVA를 통한 분산분석을 활용하여 차이가 존재하는지를 살피는 단계로 넘어간다.

**분석목적**

앞서 설정한 가설을 바탕으로, 우리는 주문시간에 따라 주문량에 차이가 있다는 점을 활용하여 각 주문시간의 특징을 파악할 수 있다. 주문량이 눈에 띄게 차이나는 주문시간대는, 해당 시간대에 플랫폼에 방문할 고객의 특징을 파악하여 그에 알맞은 프로모션을 진행하면 된다.

**마케팅 전략 예시, 시간대별 타겟 배너 및 팝업 광고 :**

오전 9시 ~ 오후6시까지는 학생과 직장인은 각각 학교와 직장에서 공부하고 일하는 시간이기 때문에 해당 시간대에 방문하는 고객은 대개 주부라는 것을 유추해낼 수 있다. 다시 말해, 해당 시간대에 구매하는 제품들은 주부들이 자주 구매하거나 애용하는 제품이라고 할 수 있다.

오후 6시 ~ 오후 12시까지는 학교와 직장에서 집으로 돌아온 후의 시간이기 때문에 해당 시간대에 방문하는 고객은 대개 학생과 직장인이라는 것을 유추해낼 수 있다. 다시 말해, 해당 시간대에 구매하는 제품들은 학생과 직장인들이 자주 구매하거나 애용하는 제품이라고 할 수 있다

따라서 각각의 주문시간대의 데이터만 sort한 후, 브랜드별로 그룹화하여, 해당 시간대에 가장 구매 량이 높은 브랜드를 쇼핑 플랫폼 배너에 광고한다. “주부를 위한”, “직장인들을 위한”과 같은 문구로 타겟층을 언급함으로써 클릭을 유도하는 것도 좋은 방법이다.

**제언 2 (노출 순위를 컬럼으로 사용한 분석) :**

상품 페이지를 웹 크롤링 (Web Crawling) 을 통해서 노출 순위에 관한 분석도 가능하지만 해당 내용은 주어진 데이터로는 추출할 수가 없는 데이터이기 때문에 제언으로 두겠다.

**가설**  
귀무가설 : μ\_1페이지 = μ\_2페이지 = μ\_3페이지 = … 판매량이 모두 같을 것이다.  
대립가설 : μ\_1페이지, μ\_2페이지, μ\_3페이지, … 중 적어도 한 가지는 판매량이 다를 것이다.

**분석 방법**

노출 순위를 페이지 별로 범주화 한다. 각 범주에 속한 개별 품목들의 판매량을 조사한 후, 해당 범주의 판매량 평균값을 구한다. 그 후 각 범주들을 기준으로 ANOVA test를 진행하여 상품의 노출 순위 페이지에 따라 판매량을 통계적으로 모두 같다고 말할 수 있는지 검정을 진행한다.

**분석 목적**

실제로 진행해보아야 정밀한 결론을 내릴 수 있지만, 통념 적인 기준에서 상품의 노출 순위는 판매량에 강한 영향을 미칠 것이라고 예상할 수 있다. 네이버 쇼핑에서는 상품 품목을 검색하면 네이버 랭킹순에 따른 정렬이 가장 먼저 검색된다. 이는 네이버의 내부 알고리즘에 따라 적합도 지수, 인기도 지수, 신뢰도 지수를 점수화 하여 정렬하게 되는 방식인데, 광고상품은 별도기준에 따라 상단에 정렬되는 방식이다. 여러 회사에서 비용을 지불하면서까지 네이버 쇼핑에서의 상단에 정렬하려고 하는 것을 보아 노출순위와 판매량은 강한 상관관계가 있을 것으로 예측된다. 11번가 또한 추천상품, 포커스클릭, 파워상품, 플러스상품 등 노출순위를 높일 수 있는 많은 광고를 운영하고 있다.

그러나 상관관계가 존재하더라도, 얼마나 유의미한 상관관계를 나타내는지는 마케팅에서 강하게 고려해야 할 요소이다. 상관관계가 약하면 비싼 비용을 지불하면서 광고를 하는 것이 손해일수도 있기 때문이다. 따라서 -1 ~ 1 사이의 값을 가지며 정규화 된 척도인 상관계수(correlation coefficient)를 계산하여 광고가 어느 정도의 비용을 지불할 가치가 있는지 알아볼 수 있다.  
  
**마케팅 전략 예시 1**  
11번가와 네이버 중 평균적인 상관계수가 더 높게 나타나는 집단에서 더 많은 광고를 진행할 수 있다. 현재까지의 판매량이 유사하다는 가정 하에 상관계수가 높은 집단이 11번가로 나타난다면, 추천상품 / 포커스클릭 / 파워상품 / 플러스상품의 광고를 모두 게시하거나, 더욱 높은 비용 지불을 통해 최상단에 광고를 진행하는 방식을 고려해볼 수 있다.

**마케팅 전략 예시 2**  
마진, 판매량, 상관계수 3가지를 모두 고려하면 더욱 경제학적인 비용 계산 책정이 가능하다. 따라서 상관계수가 유사하다면 (마진\*판매량)이 높은 상품의 노출순위를 높여 더 경제적인 마케팅을 진행할 수 있다.

**맺음말**

컬럼 별 가설 설정을 통한 데이터 분석으로 마케팅 전략을 수립하는 것은 효과적인 방식이다. 그러나 지도 학습(Supervised Learning)을 활용한 모델링을 진행하고, 판매 성공 여부를 예측함으로써 더욱 정밀한 통찰을 얻을 수 있다. 파이썬의 사이킷런 라이브러리를 사용하여 의사결정나무, 로지스틱 회귀, 랜덤 포레스트, 서포트 벡터 머신, 신경망 등 다양한 방법을 추가적으로 고려해볼 수 있다. 이 중 이진 분류를 통해 판매 성공 여부를 예측하는 의사결정나무를 활용하여 정보 획득을 최대화하는 칼럼을 확인하고 우선순위를 책정함으로써 더 효과적인 마케팅 전략을 수립할 수 있다. 이 접근 방식은 공모전의 Track 2에 해당하는 모델링 파트에 적합하며, 추후 연구를 통해 더 나은 방법을 제안할 수 있다.