**PACKUS 성장 전략 :**

**세분화된 고객 관리와 상품별 수요 예측**

**- 목차 -**

**[서론]**

패커스의 세분화된 고객 관리와 상품별 수요 예측의 필요성

**[본론]**

1. 세분화된 고객 관리  
   1) 고객군 별 맞춤형 전략  
   2) 고객등급 재개편

2. 상품별 수요 예측

**[결론]**

대시 보드를 통한 패커스의 세부적 고객 관리와 수요 예측,

**팀명 : 히스토그램**

**조원 : 김연강 신은수 이승한 정민지**

****

**[ 서론 ]**

동남아 음식, 양꼬치, 디저트. 요기요에서 발표한 2018년 신규 배달 메뉴 Best 3에 해당하는 음식들이다. 이제는 터치 몇 번으로 원하는 다양한 종류의 음식들이 집까지 오는 세상이 되었다. 심지어 빙수, 마라탕 등도 배달될 정도로 이전에 비해 훨씬 다양한 음식들이 배달 가능해졌고, 이와 함께 배달 음식 시장의 규모도 크게 성장했다. 이러한 시장 상황의 변화에 따라, 배달음식 용기 제조업체인 패커스에게도 변화가 필요해 보인다.

1. 성장세인 패커스의 “**고객 관리의 필요성”**  
 패커스는 고객 관리 제도를 전면적으로 재개편 할 필요가 있어 보인다. 패커스는 현재 active 유저 수와 매출액이 증가하고 있는 성장 기업이다.( [그림 2] 참조 ) 이러한 기업의 현 상황에 따라, CRM을 통한 세부적인 고객 관리가 그 어느 때보다 필요한 시점이다. 하지만 이러한 유저가 증가하고 있는 기업의 상황과 달리, 효율적이고 정확한 고객 관리가 이루어지고 있지는 않은 듯 하다. 지난 7월에 도입된 고객 등급 체계의 현황을 보면, 고객 등급별 구조가 매우 불균형한 것을 확인할 수 있다. 만들어진 6종류의 등급 체계에서, 2019년 1분기 기준으로 하나의 등급(BRONZE)에 약 93.1%의 고객이 몰려 있고, 최상위 등급(VVIP)에는 단 한 명의 고객도 있지 않고, 차상위 등급(VIP)에는 4명의 고객 만이 있었다. ( [표1], [그림 1] 참조 ) 이 고객 등급 제도를 수정하는 것 뿐만 아니라, 고객들의 개인 정보 및 주문 내역 데이터를 토대로 고객 맞춤형 전략을 세울 필요가 있다. 이러한 필요성에 따라, [본론 1-1]에서는 “고객 맞춤형 전략”을, [본론1-2]에서는 “고객 등급 재개편”에 관한 내용이 다루어진다.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **[표 1] 현재 패커스의 고객 등급 체계** | | | | | |
| 회원 구분 | BRONZE | SILVER | GOLD | VIP | VVIP |
| 분기별  누적 결제 | ~100만원 | 100만원~ | 300만원~ | 500만원~ | 1천만원~ |
| 2019.1분기 구매 기준 | 2017명  (93.1%) | 125명  (5.77%) | 20명  (0.9%) | 4명  (0.2%) | 0명  (0%) |

|  |
| --- |
|  |
| **[그림 1] 현 등급체계를 기준으로 한 고객 분류** |

2. 배달 음식의 다변화에 따른 세분화된 배달용기 “**수요 예측의 필요성”**

기존의 배달 음식의 대부분은 중국 음식과 치킨이 차지했다. 하지만 이제는 다양한 종류의 음식들이 배달 가능 해졌고, 배달 음식 간의 품목별 점유율에도 변화가 일어나고 있다. [그림 3]은 SKT 이용자들의 품목별 배달음식 주문 건수에, 연도 별 SKT의 이동통신사내 점유율을 나눠서 계산한 ‘총 배달음식 주문 건수’를 보여준다. 이를 통해 상위 4 종류의 배달 음식(중국음식, 족발/보쌈, 치킨, 피자)들 간의 격차도 줄어들고 있음을 확인할 수 있다.( 사용 데이터 :SKT 2013-2019 서울시 배달업종별 이용 통화량 )

이에 따라, 다양한 종류의 음식들을 담을 다양한 종류의 용기들이 필요해지고 있다. 용기의 종류가 많지 않았을 때에는 전체 배달용기의 개수를 예측하는 것도 의미가 있을 수 있었지만, 이제는 보다 세부적으로 다양한 배달 용기 종류 별로 정확한 각각의 수요 예측을 할 필요성이 있어 보인다. 이러한 필요성에 따라, [본론 2]에서는 각각의 물품 별로 수요 예측을 하는 모델에 관한 내용이 다루어진다.

|  |  |
| --- | --- |
|  |  |
| **[그림 2] 기간별 총 구매금액, 1인당 구매금액, \*Active User 수”** | **[그림 3] 종류 별 배달음식 주문 건수** |

\* Active User : 회원 가입 시 +1 , 최종 주문일이 2019.4월 이전인 경우 -1 하여 날짜 별로 계산

( 2019.6월까지의 데이터가 있으므로, 5월에 최종 주문한 사람을 이탈한 고객으로 보기 어렵다 )

**[ 본론 1-1. 고객군 별 맞춤형 전략 ]**

1) 사용 데이터

(1) 주문 데이터 : DB분석\_서버1\_2017년 고객 주문 통합 리스트.xls, DB분석\_서버1\_2018년 고객 주문 통합 리스트.xls,

2019년 상반기 고객 주문 통합 리스트.xls

(2) 유저 데이터 : (통합)DB분석\_회원관리서버.xls

2) 데이터 전처리 및 변수 선택

고객 클러스터링을 하기에 앞서서, 고객들의 특성을 대변할 수 있는 네 분류의 변수들( 고객 개인 정보, 주문 관련 정보, 혜택 관련 정보, 그 외 정보)을 선정하였다. ( 표[2] 참조 )이 변수들은, 기존 두 데이터에 있는 변수들과, 새롭게 생성한 몇 개의 변수들로 이루어진다. 회원의 현재 활동 여부를 알기 위해, (데이터 상 가장 마지막 거래 기록이 있는) 6월 19일에서 최근 접속일을 뺀 ‘최근 접속까지 일 수’, 서비스 이용기간을 파악하기 위해 최종 접속일에서 회원 가입일을 뺀 ‘이용 기간’, 어떠한 방식으로 주문을 하는지를 파악하기 위해 PC 주문 건수와 Mobile 주문 건수의 비율을 나타내는 ‘주문 경로’ 변수들을 추가하였다. 이 새로운 변수들을 포함하여 총 12개의 변수들을 선정하였다. 유의미한 차이를 보일 수도 있지만 약 30%의 값이 NA인 배송지역 변수 (‘주소(동/읍/면) ’)과, 다중 공산성을 방지하기 위해 누적 주문건수와 상관 관계(0.997)가 높은 ‘총 실주문 건수’는 이용할 변수에서 제외하였다.

|  |  |
| --- | --- |
| **[표 2] 고객의 특성을 대변할 수 있는 네 분류의 변수들 ( 총 12개 )** | |
| **1. 고객 개인 정보** | (1)회원 구분(개인/사업자), (2) 회원 등급(일반/전화주문/사업자/기타 등),  (3) 이용 기간, (4) 최근 접속까지 일 수, (5) 평생회원 여부 |
| **2. 주문 관련 정보** | (6) 누적 주문 건수, (7) 총 구매 금액 |
| **3. 혜택 관련 정보** | (8) 총 사용 적립금, (9) 쿠폰 할인 금액 |
| **4. 그 외 정보** | (10) 주문 경로(PC와 Mobile 주문 비율), (11) 추천아이디 유무,  (12) 회원 가입 경로(잡지/인쇄물/인터넷/지인 추천/기타/없음) |

|  |
| --- |
|  |
| **[그림 4] 주성분 개수에 따른 Explained Variance Ratio** |

이 12개의 변수들 중, 범주형 변수인 ‘회원 구분, 회원 등급, 회원 가입 경로’는 더미화 하여 총 22개의 변수로 된 데이터 셋으로 변형 하였다. 변수들 간의 규모 차이가 큰 것을 조정하기 위해 수치형 변수들은 전부 정규화를 하였고, 생성하게 될 고객 군집 (클러스터)의 추후 해석에 지장을 미칠 수 있는 이상치 (Z-score의 범위가 (-3,3)을 벗어나는 데이터)를 제거하여 클러스터링을 진행하였다.

3) PCA를 통한 차원 축소

위에서 생성된 22개의 변수를 가진 4583명의 데이터의

차원을 축소하기 위해 주성분 분석(PCA)를 시행하였다. [그림4]에서 볼 수 있듯, 총 5개의 변수로 약 80%의 설명력을,

8개의 변수로 약 90.4%의 설명력을 가진다. 주성분의 개수를

5개부터 8개까지 모두 이용하여 클러스터링을 해본 결과,

8개를 이용하여 군집화를 진행하는 것이 군집 간의 가장

뚜렷한 구분을 가짐을 알 수 있었다. 이에 따라, 총 8개의

변수로 차원을 축소한 이후 클러스터링을 진행하였다.

4) 클러스터링

고객 클러스터링을 하는 대에 있어서, SOM(Self-Organizing-Map)기법을 사용하여 총 4개의 클러스터( 2x2 격자 : [0 0], [0 1], [1 0], [1 1] )를 생성하였다. SOM 기법을 사용한 이유는 세 가지가 있다. 첫째로, [표 3]에서 확인할 수 있듯 이 기법은 K-means, DBSCAN, Mean-shift 등의 다른 클러스터링 기법들 보다 높은 silhouette score을 가져 더 나은 성능을 보였다. 둘째, SOM을 통해 생성되는 클러스터들은 다른 기법들과는 다르게 ‘격자 형식의 클러스터’로 생성 되어 인접한 클러스터 간에는 유사한 특징을 가진다는 정보를 확인할 수 있다. ( [그림 5] 참조 ) 마지막으로, SOM을 통해서 생성된 4개의 클러스터들의 3개의 주성분(PC1~PC3)을 가지고 3차원으로 표현해 보았을 때에도 클러스터들 간의 구분이 잘 이루어 진 것을 확인할 수 있다. ( [그림 6] 참조 ) 현재는 2x2 SOM

|  |
| --- |
|  |
| **[그림 6] 3차원으로 본 (주성분 1~3) 클러스터** |

이 가장 좋은 성능을 보였으나, 추후 패커스의 이용자 수가 훨씬 더 증가하게 된다면 3x3 SOM(9개의 클러스터)을 통해 보다 세부적으로 고객군을 분류할 수 있을 것이다.

|  |  |  |
| --- | --- | --- |
| **[표 3] 클러스터링 기법 간의 성능 비교** | | |
|  | 사용 기법 | Silhouette Score |
| 2 | K-Means(k=5) | 0.256 |
| 3 | DBSCAN | 0.01 |
| 4 | Mean-Shift | 0.341 |
| 5 | SOM 2x2 | 0.347 |
| 6 | SOM 3x3 | 0.335 |

|  |
| --- |
|  |
| **[그림 5] 격자 형태로 클러스터를 생성하는 SOM** |

위와 같이 클러스터링 된 네 개의 고객군을 “5)고객군 별 고객 특징”과 “6)고객군 별 구매 품목”의 차원에서 비교 분석하였다.

( 생성된 4개의 클러스터 ( [0 0], [0 1], [1 0], [1 1] )를 앞으로 각각 고객군 1, 고객군 2, 고객군 3, 고객군 4라고 표현한다 )

5-1) (고객군 별) “고객 특징”

고객군 별로 세 가지 측면 (‘구매력’, ‘이용 행태’,’혜택 활용’)에서 비교한 결과를 요약해서 나타내면 [표 4]와 같다.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **[표 4] 고객 군별(클러스터 별) 비교 분석** | | | | | |
|  | | **고객군 1** | **고객군 2** | **고객군 3** | **고객군 4** |
|  | 고객 수 | 992 명(21.6%) | 1484 명(32.4%) | 650 명(14.2%) | 1457 명  (31.8%) |
| **구매력** | 총 주문금액 | 56만 원 | 45만 원 | 22만 원 | 17만 원 |
| 건당 주문금액 | 6.4 만 원 | 6.3 만 원 | 6.2 만 원 | 5.8 만 원  **[그림 7] 고객군별 누적주문건수 비교** |
| 주문건수 | 8.2 건 | 6.7 건 | 3.3 건 | 2.6건 |
| **이용 행태** | 전화주문고객 비율 | 5.2 % | 52.3 % | 5.7 % | 49.4 %  **[표 4] 고객 군별(클러스터 별) 비교 분석**  기업 입장에서 (클러스터 간에 상대적으로) 높은 수익을 낼 수 있는 항목에 ‘파란색’을, 그렇지 않은 항목에 ‘분홍색’을 칠했다. 이 정보를 토대로, 고객 군에 따라 어떠한 측면에 맞춰서 전략을 펼쳐야 할지에 대해 예상해볼 수 있다. |
| 이용기간 | 146일 | 131일 | 27일 | 12일 |
| 최근 접속 | 55일 | 51일 | 322일 | 274일 |
| **혜택 활용** | 적립금 사용률 | 15 % | 2 % | 1 % | 0.1 % |
| 쿠폰 할인액 | 965 원 | 359원 | 950 원 | 488 원 |

위 표를 간단히 요약 하자면, 우선 크게 고객군1&2와 고객군 3&4는 ‘구매력’과 ‘이용 기간(최종 주문일-회원 가입일)’ 측면에 의해 크게 나누어 진다. 고객군 1&2는 오래 이용했고, 현재도 이용하고 있고, 구매 액 및 건수도 높은 고객군이다. 이에 반해, 고객군3&4는 장기간 접속하지 않고 이용기간도 짧은 구매력 낮은 고객군이다. 고객군 1&3과 고객군 2&4는 ‘주문방식’에 의해 구분된다. 고객군 1&3은 전화주문고객비율이 5%정도 밖에 되지 않지만, 고객군 2&4는 그 비율이 50%정도나 된다. 또한 눈에 띄는 점은, 고객군2&3&4는 적립금 사용률이 2%도 되지 않지만, 고객군1은 무려 15%나 되어서 적립금을 알뜰히 이용하는 고객군임을 알 수 있다. 그 외에 여러 세부 정보들은, 고객 관리를 효과적으로 하기 위해 제작한 대시 보드를 통해서 확인할 수 있다. 대시 보드에 들어가는 여러 정보들 중 하나의 예시로, 고객군 별 누적 주문건수 차이를 확인해보면 다음과 같다. ( [그림 7] 참고 )

5-2) (고객군 별) “고객 특징”에 따른 제언

**[고객군 1]**  구매력, 이용 행태, 혜택 활용 세 가지 측면에 있어서 모두 기업에게 가장 필요한 고객 군이라고 볼 수 있다. 다른 군집들이 이와 유사한 형태를 가지도록 유도해야 할 것이다. 하지만 이와 동시에, 이들을 꾸준히 보유하기 위해 맞춤형 마케팅을 하면 좋을 것이다. ( 맞춤형 마케팅 : 아래 6-1.에서 언급할 고객군 별 구매 물품 추천 )

**[고객군2]** 고객군1과 비교했을 때, 전화주문 고객 비율과 적립금 사용률에서 두드러진 차이를 보인다. 인터넷이 아니라 전화를 통해서 주문을 하는 고객이 많기 때문에 적립금 사용률도 함께 낮은 것으로 보인다. 전화를 통해 주문하는 경우, 모바일 가입 유도를 통해 혜택 정보들을 알림으로써 구매 방식의 전환을 유도할 수 있다. 각종 혜택 및 적립금 정보들에 대해 알게 되고 기업에게 더 충성도 높은 고객으로 남을 수 있다. 고객군 1에 못지않은 구매력을 갖춘 고객이기 때문에 이러한 노력이 고객군 3,4에 비해 더 필요할 것이다

**[고객군 3]** 구매력은 낮지만, 전화주문고객 비율이 낮고 대부분 모바일을 통해 주문을 하는 고객임을 알 수 있다. 고객군2나 4처럼 모바일 이용도가 낮은 고객들에 비해, 모바일 이용도가 높기 때문에 홍보나 마케팅을 펼치기 용이한 측면이 있다. 이러한 고객 특성을 활용하여 맞춤형 마케팅을 하면 좋을 것이다.

**[고객군 4]** 주문금액, 주문건수가 네 고객군 중 가장 낮고, 서비스 평균 이용기간이 12일도 되지 않은 1회성 고객에 가깝다고 볼 수 있다. 뿐만 아니라, 주문자의 50%가 전화를 통해서 주문을 한 고객으로서 홍보나 마케팅에 대한 접근도 용이하지 않은 고객이다. 기업의 한정된 자원을 이 고객 군에게 사용하기 보다는, 다른 세 고객 군에게 사용하는 것이 더 나은 판단일 수도 있다.

6-1) (고객군 별) “구매 품목”

고객군 별로 구매하는 물품들 간의 연관성을 파악하기 위해, 장바구니 분석을 실행하였다. 전체 장바구니에는 667개의 품목들이 있었으나, 보다 유의미한 분석을 위해 일부 상품들을 묶는 과정을 거쳤다. 실제 물품은 동일하지만 이름이 약간 다르거나, 용기의 색깔을 제외한 나머지 특징이 동일한 물품( ex. 흰색/검은색 탕 용기)등은 동일한 품목으로 묶어서, 총 266종류의 품목으로 축소시켜서 사용하였다. 이렇게 전처리 된 데이터를 바탕으로, 전체 고객을 대상으로 확인한 주요 판매 물품과, 고객군별 주요 구매 물품은 확연한 차이를 보였다.

단순히 고객군 별로 많이 팔린 물품의 차이를 확인하는 것을 넘어서서, 특정 물품을 구매 할 때 어떤 물품과 함께 구매가 이루어지는지 파악 하기 위해, Gephi 툴을 활용하여 고객군 들의 구매 물품들간의 연관성 분석을 해보았다. ( 다음 장 [그림 8] 참조 ) 이 분석을 통해 고객군별로 주로 운영하고 있는 음식점이 어떤 음식을 취급하는지, 함께 구매하는 물품에는 어떠한 차이가 있는지 유의미한 결과를 얻어낼 수 있었다. 이를 토대로, 고객군 별로 맞춤형 전략을 세울 수 있었다.

|  |  |
| --- | --- |
| **[ 그림 8 ] (Gephi를 활용한) 고객군별 구매 물품 분석** | |
| 아래의 그림에서, 각각의 node는 하나의 물품을 나타내고, node가 서로 연결되어 있다는 것은, 두 물품을 함께 구매했다는 것을 의미한다. ( 위의 그림의 node와 edge에 대해 간략히 설명하자면 다음과 같다. )  1) node의 크기 : degree(자신과 연결된 node의 개수)를 의미한다.   ( 시각화 단계에서, 연관성이 적은 물품들을 필터링 하기 위해 degree가 20(고객군 2의 경우 30)이하인 node는 배제하였다 )  2) node의 색깔 : modularity( 물품들 간의 cluster )를 나타낸다. ( 같은 색의 node = 같은 cluster의 물품 )  3) edge의 두께 : 연결된 두 품목간의 유사한 정도를 의미한다 ( 유사하다 = 같이 구매가 이루어지는 물품들이 유사하다 ) | |
|  |  |
| **고객군 1** | **고객군 2** |
|  |  |
| **고객군 3** | **고객군 4** |

6-2) (고객군 별 ) “구매 품목”에 따른 제언

각 고객군 만의 특징을 바탕으로, 다음과 같은 제언을 할 수 있다.4개의 고객군에서 모두 공통적으로 음식 용기로는 “도시락”용기가 가장 높은 degree를 가졌기 때문에, 이를 제외한 고객군별 main 음식에 초점을 맞춰서 제언을 했다.

**[고객군 1] 주요 제언 물품 : “중화용기”**

중화 용기와 관련된 물품이 그 어느 고객군 보다 비중이 높은 군집이다. 나무 젓가락이나, 단무지 등을 끼워 팔기를 통해 구매를 유도할 수 있을 것이다. 또한, ‘도시락 용기’ 구매자에게는 사은품(행주, 수세미, 위생장갑)이 지나칠 정도로 많이 제공되고 있는 반면, ‘중화 용기’ 구매자에게는 사은품을 거의 주지 않는 것을 확인할 수 있다. 중화용기에도 이에 알맞은 사은품을 제공하여 더욱 더 고객을 끌어올 필요가 있다.

**[고객군 2] 주요 제언 물품 : “탕 용기”**

3개(도시락,짬짜면,중화면 용기)를 제외하고, 전부 “탕 용기 (탕 용기, 감자탕 용기, 원형용기 등)”만으로 구성된 장바구니이다. 이 고객군은 탕 용기를 구매할 때, 매우 높은 연관성으로 ‘다용도 컵’(70,95,105 파이)을 함께 구매하는 것을 확인할 수 있다. 이 고객군이 탕 용기를 구매 할 때, 추천 구매 물품으로 다용도 컵을 노출시키는 것도 하나의 구매를 유도할 수 있는 방안이다.

**[고객군 3] 주요 제언 물품 : “도시락 용기”**

어느 집단에서나 모두 도시락 용기가 많이 등장한다. 하지만 고객군 3만큼 도시락 용기가 다른 상품과의 연관성이 높은 군집은 없다. 이들은 [고객군 2]가 탕 용기와 다용도 컵이 함께 구매된 것처럼, 도시락 용기와 다용도 컵이 함께 구매가 이루어진다. 마찬가지로 추천 구매 물품에 노출시키면 좋을 것이다.

**[고객군 4] 주요 제언 물품 : 사은품**

고객군 4는 ‘5-1) 고객군 별 고객특징 분석’에서 확인 했듯이, 기업에게 수입 가져오는 주요한 세 지표( 건당 주문금액, 총 주문금액, 총 주문 건수 )에서 모두 제일 낮은 금액 및 건수를 기록한 고객군이다. 하지만 이들의 장바구니를 분석한 결과, 생성된 3개의 물품 들의 클러스터(modularity)에서 모두 사은품이 가장 높은 degree를 보임을 확인할 수 있었다. 기업의 한정된 자원을 고려하면, 이들에게 마케팅이나 사은품을 제공하는 것 보다는 다른 고객군을 더 관리하고 혜택을 제공하는 것이 더 나은 선택일 수도 있다.

**[ 본론 1-2. 고객 등급 재개편 ]**

지난 7월에 처음 도입한 패커스의 고객등급 체계는 매우 불균형적인 구조를 가지고 있다. Bronze부터 VVIP까지 총 5개의 등급에서, Bronze등급이 2017명(93.1%)을 차지하고, Silver, Gold, VIP, VVIP 각각 125,20,4,0명을 차지하고 있다. 등급을 나누는 기준 (1. 누적 구매 금액, 2.단위 기간)에 수정이 불가피해 보인다. 다음 등급으로 올라가기 위한 ‘단위 기간’당 ‘누적 구매 금액’을 수정하고자, 우선 평균 구매 주기부터 확인해 보았다. ‘2건’ 이상의 거래가 있는 사람들을 대상으로, 가장 최근 구매한 두 건의 구매일자의 차이의 평균을 구해 유저들의 ‘평균 구매 주기’를 계산한 결과, 약 22일이라는 기간이 나왔다. 현재 등급 체계에서는 단위 기간을 ‘한 분기’로 잡고 있는, 이는 최근 4~5회의 누적 구매 금액을 기준으로 삼은 꼴이다. 1회 구매 시 구매하는 금액에 따라 고객의 등급이 수시로 바뀌는 것은 고객의 입장에서 불안정성을 가져올 수 있기 때문에 4~5회의 누적구매 금액으로 기준으로 삼는 것 (즉, 현재 제도와 같이 단위 기간을 1분기로 유지하는 것)은 나쁘지 않아 보인다. 이는 결국 ‘누적 구매 금액’을 수정해서 등급 컷을 정해야 함을 의미하지만, 얼마의 금액이 최적의 등급 컷이 될지는 파악하기 어렵다. 하지만, 다른 고객등급체계를 운영하는 서비스들을 벤치 마킹하여 다음과 같이 등급 체계를 재조정해 보았다.( [그림 9] 참조 )

각각의 등급에 60%,15%,15%,7.5%,2.5%의 고객들이 들어가게 끔 조정을 해보았다. 이 비율로 정하게 될 경우, 다음 등급으로 넘어가기 위한 ‘1분기당 누적 구매금액’은 다음과 같이 수정되어야 한다 ( [표 5] 참조 )

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | Bronze | Silver | Gold | VIP | VVIP | | 기존 | ~100만 | 100만~ | 300만~ | 500만~ | 1000만~ | | 개편 | ~25만 | 25만~ | 40만~ | 80만~ | 180만~ |   \* 각 등급별 비율을 위와 같이 유지하려면, 실제 정확한 등급 컷은 차례대로 248,840원(25만), 391,250원(40만), 771,900원(80만), 1,743,620원(180만) 이다. |
| **[그림 9] 개편된 고객 등급 체계** | **[표 5] 등급 상승을 위해 필요한 1분기당 누적구매금액** |

|  |
| --- |
|  |
| **[그림 10] 등급 별 구매금액 대비 적립금 사용 비율** |

현재 등급 체계에서는, Bronze,Silver,Gold,VIP,VVIP에게 각각 0.2%,0.3%,0.4%,0.5%,0.6%의 마일리지 적립률을 제공한다. 하지만, Bronze, 구매금액 대비 사용하는 적립금의 비율을 확인하면 ( [그림 10] 참고 ) Bronze,Silver, Gold등급의 사람들은 절반 이상의 고객들이 적립금을 이용하지 않음을 알 수 있다. (각각 16,39,43% 만이 이용) 이들은 대신에, 쿠폰 이용률이 VIP(91%)보다 높음(Bronze 97%, Silver 97%, Gold 93%)에 따라, 쿠폰 혜택이 보다 적절해 보인다. 그러므로 Bronze, Silver, Gold등급에는 ‘쿠폰’혜택을, VIP와VVIP에는 ‘적립율’ 혜택을 차등적으로 제공하는 것이 더 나을 것으로 예상된다.

**[ 본론 2. 상품 수요예측 ]**

1) 목적 : 각 상품의 과거 데이터를 토대로 다음 달 판매 량(수요)을 예측

배달 업계의 성장과 간편식, HMR 수요 증가에 따른 플라스틱 용기 사용처 다양화로 인해 배달용기의 수요가 점점 증가하고 있다. 앞으로의 수요변화에 대응한 생산계획을 수립하고, 생산자원을 적기에 확보하여 과잉 재고나 재고 부족 등에 의한 손실을 줄이기 위해 각 상품의 다음달 판매수량을 예측하는 모델을 만드는 것을 목표로 하였다.

2) 데이터 전처리 및 Feature Engineering

[ 사용 데이터 ]

1. 주문 데이터 : DB분석\_서버1\_2017/2018년 고객 주문 통합 리스트.xls, DB분석\_서버1\_2019년 상반기 고객 주문 통합 리스트.xls ( 이용컬럼 : ‘결제일시(입금확인일)’, ‘상품번호’, ‘주문상품명’, ‘상품구매금액’, ‘수량’, ‘취소구분’ )
2. 상품 카테고리 : 패커스 홈페이지에서 크롤링
3. 삼륭물산 주가 : 네이버 금융 사이트에서 크롤링

[ 전처리 ]

|  |  |
| --- | --- |
| 총 거래 기간은 ‘2017.01.01 ~ 2019.06.19’이지만 2019년 6월은 한 달의 데이터가 온전히 있지 않기 때문에 2017년 1월부터 2019년 5월까지의 데이터만을 사용하였다.  또한 패커스 상품 총 1231개 중 실질적으로 판매된 상품은 628개로, 그 중 매출의 90%를 차지하는 상품 135개에 대해 판매 수요를 예측함과 동시에 최근까지도 잘 팔리는 상품을 예측하고자 마지막 판매일이 2019년 4,5,6월인 상품만을 선택하여 데이터를 정리하였다. 최종적으로 130개의 상품에 대하여 수요 예측을 진행했다. |  |

|  |
| --- |
| **예시) PP 5칸 도시락용기 DS 200개 1박스(돈까스 용기)의 일별 판매추이** |
|  |

패커스 회사 설립 초기 단계에는 제품별 판매량이 수요예측을 진행하기에 충분하지 않아서 월단위로 묶은 이후 분석을 진행하였다.

[ Feature Engineering ]

변수 선택 과정에 있어서, 물가 및 배달 음식 시장 의 현황을 대표 할 수 있는 지수들도 함께 넣고 모델을 만들어서 예측을 해봤다. 하지만 패커스의 기업 성장 단계 과정 상, 성장 초기단계에 있어서인지 거시적인 요인(기온, 4종 대표적 배달 음식의 일별 배달량 등)에 의한 영향 보다, 기업 과 직접적으로 연관 되어있는 회사의 자체적인 변수들에 의해 더 영향을 많이 받는 것을 보였다. 이러한 변수들로 상품수요를 예측하는 것이 더 좋은 성능을 거두는 것을 확인할 수 있었다. 사용한 변수들은 아래와 같다.

1. Target lags : 시계열 데이터의 자기상관성을 반영하기 위해 n시차전의 판매량을 변수로 추가하였다. 최근 경향과 장기 경향을 반영하기 위해 각각 (1, 2, 3시차), (6, 12시차)를 선택하였다.(=’item\_cnt\_lag\_#’)
2. 3개월 이동 평균 및 3개월 이동 표준편차 (=’rolling\_mean\_3’, ‘rolling\_std\_3’)
3. Trend feature : 각 상품의 판매가 변화율을 반영하기 위한 변수

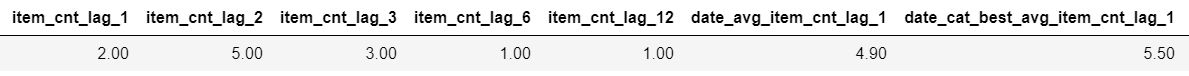
- 각 상품의 평균판매가에서 한 시차 전의 판매가가 얼마나 변했는지(=’delta\_price\_lag’)

- 한 시차 전의 판매가로부터 현재 판매가가 얼마나 변했는지(=’delta\_price’)

1. 주가 : 패커스의 모회사인 삼륭 물산의 한 시차전의 주가. (=’stock\_lag\_1’)
2. 카테고리 별 평균판매량 : 해당 카테고리의 #시차 전 평균판매량(=’date\_cat\_(카테고리명)\_lag\_#’)
3. 처음 판매가 된 기준년월 (=’item\_first\_sale’)

12시차까지를 변수로 사용하였으므로 앞의 12개월 데이터를 삭제한 후 최종 데이터를 사용하였다.

[ 최종 데이터셋 ]



최종데이터의 예시는 위와 같고, 2017년 1월 ~ 2019년 4월의 데이터를 train data로, 2019년 5월의 데이터를 우리가 최종적으로 예측할 test data로 하여 모델링을 진행하였다.

3) 모델링

(1) Tree-based model 사용 이유 : 처음에는 각각의 상품에 대해 단일변량 ARIMA모델을 사용하고자 하였으나 월별 통합 데이터는 observation의 수가 너무 적고, 시계열의 정상성 가정에 민감한 ARIMA 모델을 사용하기에 어려움이 있었다. 따라서 주가, 카테고리, 이동평균 등의 변수를 더 추가하여 분석 할 수 있고, 하나의 모델로 여러 개 상품에 대한 결과를 보여주는 tree-based model을 사용하는 것이 합리적이라고 판단하였다.

(2) 모델링 및 모델 평가

LightGBM, XGBoost, Random Forest, Extratree 총 4개의 모델을 사용하였다. 모델의 과적합 방지를 위해 교차검증을 하고자 데이터가 시계열로 의미가 있을 때 사용하는 시계열 cross-validation을 사용하였다. 각각 최근 3개월(2019년 2,3,4월)을 valid set으로 하고 지난 날을 train set으로 하여 평가지표(R^2, RMSE)를 계산한 뒤 평균 내서 비교하였다. 모델링 이후, SHAP 지수를 이용하여 만들어진 모델을 분석해보았다. 수요량 예측에 중요한 변수들을 파악해본 결과, 기업 입장에서 통제 가능한 변수들 보다, 통제 가능하지 않은 변수들(한 시차 전, 두 시차 전의 판매량)이 높은 중요도를 가짐을 확인할 수 있었다. 이를 통해, 패커스는 다른 변수들의 변화를 유도하여 판매량을 늘리는 전략을 세우기 보다, 각각 품목의 변화 추이를 파악하여 이러한 정보를 바탕으로 추후 판매량을 예측하는 전략이 더 낫다는 결론을 내릴 수 있었다.

|  |  |
| --- | --- |
|  | cross-validation 결과, Random Forest는 86%의 설명력을 가지고  RMSE 또한 네 개의 모델 중 가장 낮다. |
| **[표 6] Cross-Validation 결과** |
|  | 최종 예측결과는 LightGBM과 XGBoost 등 부스팅 모델이 더 잘 예측한 것 같아 보이지만 오버피팅이 되기 쉬운 부스팅 모델의 단점으로 인한 결과로 생각된다. 따라서 앞의 cross-validation 결과와도 큰 차이가 나지 않고 최종 예측 성능도 안정적인 Random Forest를 최종적인 수요예측 모델로 사용하는 것이 타당하다 |
| **[표 7] 최종 예측 결과** |

|  |
| --- |
|  |
| **[ 그림 11 ] 입금 배송일 차이** |

4) 활용방안

[그림 11]을 보면, 입금일과 배송일 사이의 차이가 많이 나는 경우가 보인다. 이는 곧 고객 입장에서 배송이 지연되는 경우가 종종 있음을 의미한다. 패커스 상품 주문 수 38561건 중 입금과 배송시작일이 하루 이상 차이 나는 경우는 총 6100건, 주말을 감안했을 때 3일 이상 차이 나는 경우는 621건이었다. 심지어 일주일 이상 차이 나는 건은 344건이었다. 너도나도 총알배송을 외치는 시대에 이러한 배송지연은 곧 소비자이탈로 이어질 수 있다. 정확한 수요예측을 통한 적절한 재고 확보로 입금 확인일과 배송 시작일의 차이를 줄임으로써 이전보다 빠른 배송이 가능해지고 소비자에게 만족스러운 서비스를 제공할 수 있다. 이를 통해 소비자의 신뢰를 확보하여 패커스의 경쟁력 향상이 가능하다.

|  |
| --- |
|  |
| **[그림 12] 세부적인 맞춤형 전략을 위한‘고객 관리 대시 보드’**  ( 고객군 별로 기본적인 수치들에 대한 상세비교 뿐만 아니라, 구매 특성에 따른 ‘추천 판매 물품’등에 대한 정보들도 함께 담겨 있다 ) |

**[ 결론 ]**

패커스는 지금 그 어느 때보다 중요한 시점에 있다. 배달 음식의 품목들이 다양해지고 이에 따라 시장규모도 커지고 있을 뿐만 아니라, 패커스 내에서도 고객 수가 꾸준히 증가하는 추세를 보이고 있다. 이러한 상황에서, 패커스는 다양해지는 품목들에 알맞게 (1) 정확한 품목별 수요예측과, 늘어나고 있는 고객들을 잘 관리하기 위한 (2) 고객군 별 맞춤형 전략을 펼칠 필요가 있다. 단순히 거래금액만이 아닌, 다양한 특징들로 보다 종합적으로 고객군을 세분화 한 뒤, 각 군집 별로 (장바구니 분석을 통해) 구매 특성을 파악하여 맞춤형 전략을 세워야 한다. 이러한 고객관리를 효과적으로 하기 위해, 다음과 같은 ‘고객 관리 대시 보드’를 제작하였다.( [그림 12] 참조 ) 이 대시 보드에는, 고객군 별로 기본적인 정보(유저 수, 매출 기여도, 이용 기간, 적립금 및 쿠폰에 따른 반응율)뿐만 아니라, 이들의 구매 내역들을 장바구니 분석하여 얻어낸 ‘추천 판매 물품’등에 대한 정보들도 담겨 있다. 이를 통해, 패커스가 고객군별로 세부적인 맞춤형 전략을 펼쳐서 늘어나고 있는 고객들에 대한 충성도를 유지시킬 뿐만 아니라, 새로운 고객들도 유입할 수 있을 것으로 기대한다. 이와 함께, 상품별 정교한 수요예측을 통해 재고관리 비용 감소 및 빠른 배송으로 인한 만족도 증가도 이끌어낼 수 있을 것이다.