

온라인 소비 집단의 프로파일링을 통한 O2O, O4O 마케팅 방안 탐색

팀명 : 뉴브이 (김아영, 박재영, 이동호)

2021 금융 데이터 경진대회 결과 제출 양식

□ 요약

해당 프로젝트는 온라인 소비 집단의 프로파일링을 통한 O2O(Offline-to-Online), O4O(Online-for-Offline) 서비스 마케팅 방안 탐색을 주요 주제로 선정하고 진행하였다. 먼저 소비자 집단을 나누고, 집단 별 유의미한 마케팅 방안을 찾는 분석을 하였다.

포스트 코로나 시대에 성장하는 O2O 서비스 사업에 대응하는 새로운 전략 수립이 필요성에 대해 제기하였다. O2O/O4O 서비스는 그 영향력을 확대해 나가고 있으나, 아직 온라인으로 뛰어들지 못한 사업자, 그리고 소비자들을 대상을 구별해 내고 그에 맞는 마케팅 전략 수립이 필요하다. 우리는 이러한 문제의식을 바탕으로 온라인과 오프라인 데이터를 비교분석하여 새로운 소비집단들을 구별해내고, 각 집단에 필요한 적절한 마케팅 전략을 통해 오프라인과 온라인의 상생의 아이디어를 제시하였다.

분석의 목적은 소비집단을 새로 프로파일링하는 것으로, 금융데이터거래소에서 제공받은 “비씨카드 - 온라인 품목별 소비데이터” / “신한카드 - 광역시도별 업종별 가맹점 데이터”를 기반으로 진행하였다. 분석의 순서는 다음과 같다.

- 유의한 파생변수 설정 > 온라인 소비 데이터 사용자 군집화 > 군집 별 특성 도출
> 오프라인 가맹점 데이터 소비 집단 예측 > 군집별 아이디어 제안

우선 온라인 소비데이터를 K-means 알고리즘을 사용하여 군집분석을 진행하였고, 4 개의 소비집단으로 분류하였다. 이를 확장시켜 LightGBM 알고리즘을 사용, 오프라인 가맹점데이터를 동일 소비집단으로 예측/분석하여 오프라인 소비 특성을 파악 할 수 있도록 하였다. 최종적으로 각 소비집단별 특성 파악, 마케팅 방안 제시 순으로 분석 프로젝트를 진행 하였다.

다음은 4 개의 군집별로 뚜렷하게 나타나는 특성들을 정리한 것이다.

- A:
남성 100%로 이루어진 집단. 30~50 대에 분포. 자녀가 있는 가구. O2O 서비스의 소비가 적음. 특정 품목의 온라인 소비가 뚜렷하지 않음.
- B:
온라인 소비금액이 압도적으로 많은 집단. O2O 서비스의 이용이 활발. 자녀가 있는 가구. 교육열이 높은 강남구, 송파구, 서초구, 노원구, 양천구, 강동구에 주로 분포. 코로나 전/후 온라인 소비 차이가 크지않음.
- C:
20 대 MZ 세대. 1 인가구 100%. 1 인가구가 많은 관악구에 주로 분포. O2O 서비스와 e 머니, 상품권 이용이 매우 활발. e 머니를 이용한 현명한 온라인 소비를 하는 집단으로 추정.

- D:

여성 100%로 이루어진 집단. 30~50 대에 분포. 자녀가 있는 가구. O2O 서비스의 소비가 적음. 생활용품, 건강식품, 신선제품 등에 주로 소비. 주부로 추정.

위의 4 개의 군집은 다시 O2O 서비스 이용량에 따라 다시 2 가지로 구분이 가능하다.

- O2O 의존도 높은 집단 : B / C 군집
- O2O 의존도 낮은 집단 : A / D 군집

30-50 대 남성으로 이루어진 A 집단의 오프라인 소비 중, '전문용역 서비스'와 '병원', '학원' 부분이 우세한 것을 볼 수 있다. '전문용역서비스'의 소분류에는 공인노무사, 번역/통역, 법무사, 변리사, 변호사, 세무사 등이 있다. '전문용역서비스'와 '병원' 부분이 우세한 것은 법률상담 분야와 헬스케어의 O2O 서비스 사용이 아직 활발하지 않기 때문이라고 추정된다. '학원' 부분이 우세한 것은 A 집단이 모두 자녀가 있는 가구로 구성 된 것을 보아 자녀들의 교육비를 오프라인으로 결제한 것으로 생각할 수 있다. 이 분야의 O2O 서비스들은 모두 오프라인 상에서 수요가 있지만 온라인 상에서 아직 활성화 되지 않은 서비스라는 공통점이 있다. 다음 O2O 서비스 사례들로 A 집단의 온라인 소비를 증가시킬 수 있을 것이라 판단된다.

30-50 대 여성으로 이루어진 D 집단의 오프라인 소비 중에는 이미 O2O 서비스가 활발하게 이루어지고 있는 '레저'와 '숙박'이 많은 부분을 차지한다. 이 집단은 숙박/레저의 O2O 서비스에 대해 인지하지 못하고 있거나 오프라인에 비해 별다른 장점을 찾지 못했을 가능성이 높다. D 집단의 온라인 소비 증가를 위해서 이 분야의 O2O 서비스 사례들은 30-50 대 여성들을 타겟으로 한 활발한 마케팅과 전략이 필요할 것으로 분석된다.

O2O 서비스의 이용이 활발한 집단 (B, C 집단)에게는 O4O 를 활용하여 새로운 사업 구축하는 방안이 있다. B,C 집단은 이미 온라인에서 소비를 활발하게 이용하고 있으며, 적극적으로 O2O 서비스를 이용하고 있다. O2O 서비스를 적극적으로 이용하는 소비자집단은 O4O 서비스로 유도하여 오프라인으로 다시 소비자들을 유도하여, 온,오프라인의 상생을 이끌수 있다.

위의 아이디어를 통해 우리는 3 가지의 측면에서 기대효과를 살펴볼 수 있는데, 소비자 특성 별 마케팅을 진행해야하는 온라인 사업자 측면, 오프라인 사업 의존도가 높은 오프라인 사업자 측면, 그리고 금융 서비스를 제공해야하는 은행측면에서 각기 다른 효과를 기대해 볼 수 있다.

온라인 사업자 측면

- 소비자 유형을 바탕으로 한 행동 패턴 분석으로 차별적인 맞춤형 체험, 구매 공간으로서의 오프라인 구축과 이를 바탕으로 한 효과 적인 옴니 채널 방안 마련의 가능성.

오프라인 사업자 측면

- O2O 서비스가 활발하지 않은 분야의 경우 온라인과 협업해서 새로운 플랫폼을 구축해 신규 유저 영입하기.

은행 측면

- O2O 서비스 관련 스타트업 창업 지원시 사업타당성 분석의 기반 제공.
- O2O 서비스 ,E 머니/상품권은 미래에 새로운 화폐가 될 가능성이 있으므로 오프라인 가맹점들은 자신의 가맹점 포인트를 현금과 교환시키는 작업이 중요. 은행 전용 O2O 포인트 앱을 만들어 저축을 유도.

1. 주제

해당 프로젝트는 온라인 소비 집단의 프로파일링을 통한 O2O(Offline-to-Online), O4O(Online-for-Offline) 서비스 마케팅 방안 탐색을 주요 주제로 선정하고 진행하였다. 먼저 소비자 집단을 나누고, 집단 별 유의미한 마케팅 방안을 찾는 분석을 하였다.

온라인 소비자 집단은 “비씨카드 - 온라인 품목별 소비데이터”를 기반으로 분류하였으며, 동일 분류 모델을 활용하여 “신한카드 - 광역시도별 업종별 가맹점 데이터”의 오프라인 소비자 집단을 구분하였다.

분류된 소비집단을 기반으로 각 집단 별 O2O, O4O 서비스 마케팅방안을 수립하였으며, 오프라인 소비를 선호하는 집단은 O2O 서비스로 온라인 소비를 증가시킬 수 있는 방안, 소비를 선호하는 집단을 O4O 서비스로 오프라인으로 유도하는 방안을 각각 도출하였다.

2. 배경 및 필요성

포스트 코로나 시대, e 커머스의 폭발적인 성장

e 커머스 시장은 매해 꾸준한 증가세를 보이고 있었지만, 코로나 바이러스로 인한 팬데믹 사태이후 더욱 폭발 적인 성장세를 보이고 있다. 시장조사업체 이마케터에 따르면 지난해 한국의 이커머스 매출은 1041 억달러(약 120 조 4000 억원)로 전년대비 19.5% 증가했다. 그중에서도 O2O 서비스의 성장이 더욱 돋보이고 있다. 이는 코로나 19 를 사태를 겪으며 사람들은 오프라인 공간을 대체할 수 있는 온라인 공간에서 많은 시간을 보내고 있다고 해석할 수 있다.

사회적 거리두기, 비대면 서비스 확대로 오프라인과 온라인의 경계가 허물어지고 있으며, 기존의 e 커머스 시장의 성장세로 인해 공간이라는 단어가 더 이상 오프라인 영역만을 의미하지 않게 되었다. 온라인과 오프라인 채널의 결합이 빠르게 이루어지고 있으며, 이에 따라 새로운 소비구조인 옴니채널이 등장 하였고, 최근에는 O2O 서비스 그리고 발전된 개념인 O4O 서비스가 등장하였다.

O2O? O4O?

O2O(Online to Offline) 는 온라인과 오프라인을 연결한 마케팅으로, 오프라인 서비스를 온라인의 소비자와 연결해주는 전자상거래 서비스를 의미한다. 우리 곁의 대표적인 사례로 배달의 민족, 에어비앤비가 있다. 온라인과 오프라인의 중개자 역할에 치중하는 서비스이다.

O4O(Online for Offline) 는 오프라인을 위한 온라인이라는 점에서 O2O 서비스와 차이가 있으며, 온라인과 오프라인을 중개하는 역할을 넘어 온라인의 DB 와 첨단기술을 기반으로 오프라인으로 사업 영역을 확대하는 서비스라 볼 수 있다. O4O 는 O2O 를 활발히 전개하고 있는 서비스에 도입할 수 있는 새로운 형태의 채널이 될 수 있다고 전망 되어 진다. 대표적인 사례로 아마존 고 , 무신사 테라스가 있다.

기존에는 오프라인 유통업계의 부진으로 온라인 중심으로의 진출이 활발했다면 최근에는 경쟁이 더욱 심화하는 상황에서 온·오프라인을 넘나드는 사업 기틀 구축에 더욱 박차를 가하고 있다. 온라인 중심 업체들은 성장세를 바탕으로 오프라인으로 진출하고, 기존 오프라인 중심 업체들은 온라인과의 결합을 통해 활동 범위를 확장하고 있는 것이다.

O2O 서비스의 성장, 그리고 그림자

과학기술정보통신부의 발표에 따르면, '20 년 O2O 서비스 플랫폼에서 이루어진 총 거래액은 약 126 조 원으로, 전년(약 97 조 원) 대비 29.6% 성장한 것으로 나타났다. 또한 O2O 서비스 기업은 총 678 개로 전년 대비 123 개 증가한 것으로 조사되었다.

2021 년 현재, 과거에 오프라인에서 향유했던 다양한 서비스는 스마트폰이라는 품팩터의 등장 이후 내 손안에서 대부분 이용할 수 있게 되었다. 그리고 오프라인 서비스를 전개하던 기업들에게 온라인과의 결합이 늦어지는 것은 시장에서 도태 되는 것은 당연한 수순이 되어 버렸다.



그림 1-1. 2020 년 O2O 서비스매출액 <과학기술 통신부>

어느정도 규모가 갖춰진 기업들은 이러한 당면 과제들에 대응하였지만 개인사업들은 이에 빠르게 대응하지 못하는 문제를 낳았다. 또한 이러한 추세에 따라가지 못한 소비자들은 “디지털 소외계층”이 되는 새로운 문제가 발생하기도 하였다. 갤럽조사에 따르면 2021 년 한국 성인 스마트폰 사용률이 95%에 육박하지만, 아직 이들에 의한, 그리고 위한 서비스는 아직 충분하지 못하다고 볼 수 있다.

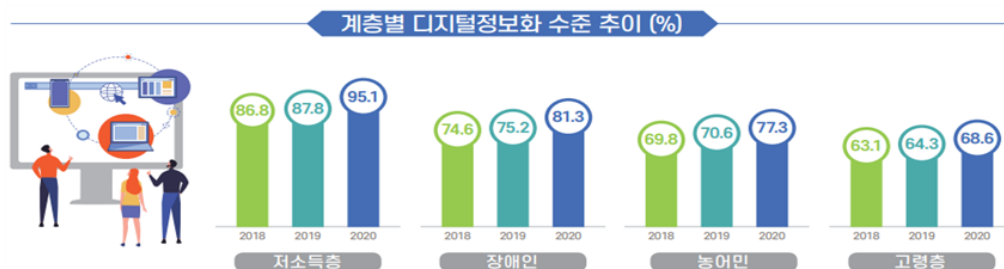


그림 1-2. 취약계층 디지털정보화 추이 <과학기술 통신부>

새로운 소비집단 분류 기준

이처럼 지속적으로 성장하는 O2O 서비스 사업에 대응하는 새로운 전략 수립이 필요하게 되었다. 분명히 O2O/O4O 서비스는 그 영향력을 확대해 나가고 있으나, 아직 온라인으로 뛰어들지 못한 사업자, 그리고 소비자들을 대상을 구별해 내고 그에 맞는 마케팅 전략 수립이 필요하다.

우리는 이러한 문제의식을 바탕으로 온라인과 오프라인 데이터를 비교분석하여 새로운 소비집단들을 구별해내고, 각 집단에 필요한 적절한 마케팅 전략을 통해 오프라인과 온라인의 상생의 아이디어를 제시하고자 한다.

3. 아이디어 제안 및 분석 결과

분석 개요

분석의 목적은 소비집단을 새로 프로파일링하는 것으로, 금융데이터거래소에서 제공받은 “비씨카드 - 온라인 품목별 소비데이터” / “신한카드 - 광역시도별 업종별 가맹점 데이터”를 기반으로 진행하였다. 분석의 순서는 다음과 같다.

- 유의한 파생변수 설정 > 온라인 소비 데이터 사용자 군집화 > 군집 별 특성 도출 > 오프라인 가맹점 데이터 소비 집단 예측 > 군집별 아이디어 제안

우선 온라인 소비데이터를 K-means 알고리즘을 사용하여 군집분석을 진행하였고, 4 개의 소비집단으로 분류하였다. 이를 확장시켜 LightGBM 알고리즘을 사용, 오프라인 가맹점데이터를 동일 소비집단으로 예측분석하여 오프라인 소비 특성을 파악 할 수 있도록 하였다. 최종적으로 각 소비집단별 특성 파악, 마케팅 방안 제시를 진행한다.

파생변수 설정

군집 분석에 앞서, 학습의 정확도를 높이기 위해 BC 카드의 온라인 소비데이터를 이용하여 3 개의 파생변수 XYZ, COVID, 지속소비등급 컬럼을 추가하였다. 각 컬럼의 구분 기준은 아래와 같다 .

- XYZ 컬럼: 연령에 따른 세대 구분 (X 세대: 50 대, 60 대 이상 / Y 세대: 30 대, 40 대 / Z 세대: 20 대 미만, 20 대)
- Covid 컬럼: 2020 년을 기준으로 코로나 전 후 구분 (b_cov : 2019 년 이전, a_cov: 2020 년 이후)
- 지속소비등급 컬럼: 소비건수에 따른 등급 분류 (브론즈: 9 건 미만, 실버: 9~39 건, 골드: 40~3999 건, 다이아: 4000 건 이상)

	기준년월	품목대분류코드	품목대분류명	품목중분류코드	품목중분류명	성별	연령	가구생애주기	고객소재지_광역시도	고객소재지_시군구	고객소재지_읍면동	매출금액	매출건수	XYZ	Covid	지속소비등급
0	201903	10	의류	1010	여성의류	여성	40	초중고자녀가구	서울특별시	구로구	공동	782521	13	Y	b_cov	실버
1	201903	21	e상품/서비스	2110	o2o서비스	여성	20	신혼영유아가구	서울특별시	구로구	공동	963789	29	Z	b_cov	실버
2	201903	19	출산/육아	1913	육아용품서비스	여성	30	신혼영유아가구	서울특별시	구로구	공동	924309	46	Y	b_cov	골드
3	201903	21	e상품/서비스	2111	e머니/상품권	여성	30	신혼영유아가구	서울특별시	구로구	공동	1089360	54	Y	b_cov	골드
4	201903	17	생활	1710	생활용품	여성	30	신혼영유아가구	서울특별시	구로구	공동	1656740	58	Y	b_cov	골드

그림 2. 파생변수: XYZ, Covid, 지속소비등급

파생변수의 유의성을 검증하기 위해 각각의 파생변수 컬럼을 ‘품목중분류명’ 컬럼과 유의수준 0.05 하에서 카이제곱 독립성 검정을 진행하였다(둘다 범주형이기 때문에). 귀무가설과 대립가설을 설정하고 P-value 값이 0.05 보다 낮은 지 확인하였다.

- 귀무가설: 품목중분류명과 3 개의 파생변수는 독립적이다 (관계가 없다).
- 대립가설: 품목중분류명과 3 개의 파생변수는 독립적이지 않다 (관계가 있다).

```
result = pd.crosstab(Seoul.지속소비등급 , df.품목중분류명)
stats.chi2_contingency(observed = result)
```

```
(57949.06094223453,
0.0,
147,
```

```
result = pd.crosstab(Seoul.Covid , df.품목중분류명)
stats.chi2_contingency(observed = result)
```

```
(2032.4777815632297,
0.0,
49,
```

```
result = pd.crosstab(Seoul.XYZ , df.품목중분류명)
stats.chi2_contingency(observed = result)
```

```
(8922.191396550026,
0.0,
98,
```

그림 3. p-value 값

위의 결과와 같이 세 변수 모두 P-value 값이 0.0 으로, 유의수준 0.05 보다 작으므로 귀무가설을 기각하고 대립가설을 채택하였다. 이는 각 파생변수들의 값에 따라 품목중분류명의 값이 유의미한 차이가 있다고 해석할 수 있다.

군집화

파생변수를 추가한 데이터의 분석결과가 더 유효해지도록 데이터를 로그 변환 하는 과정을 거쳤다. 로그변환을 통해 데이터 간 편차를 줄여 왜도(skewness)와 첨도(Kurtosis)를 줄여, 데이터의 정규성을 높였다.

	품목중분류명	성별	가구생애주기	고객소재지_시군구	XYZ	Covid	지속소비등급
0	3.636719	0.693359	1.609375	1.946289	0.693359	0.693359	1.386719
1	1.098633	0.693359	1.386719	1.946289	1.098633	0.693359	1.386719
2	3.689453	0.693359	1.386719	1.946289	0.693359	0.693359	0.000000
3	0.693359	0.693359	1.386719	1.946289	0.693359	0.693359	0.000000
4	3.257812	0.693359	1.386719	1.946289	0.693359	0.693359	0.000000

그림 4. 데이터 정규화 (로그변환)

군집분석에는 K-means 클러스터링 알고리즘을 사용하였고, 해당 알고리즘은 비지도학습으로 군집의 수를 미리 정해주어야 하기 때문에 엘보우 기법과 실루엣 기법을 사용하여 군집(K)의 개수를 정하였다. 분석 결과, 군집 개수가 4 일때 실루엣 계수가 0.651 로 가장 높게 나타나, 군집 개수를(K)을 4 개로 정하여 군집을 실시하였다.

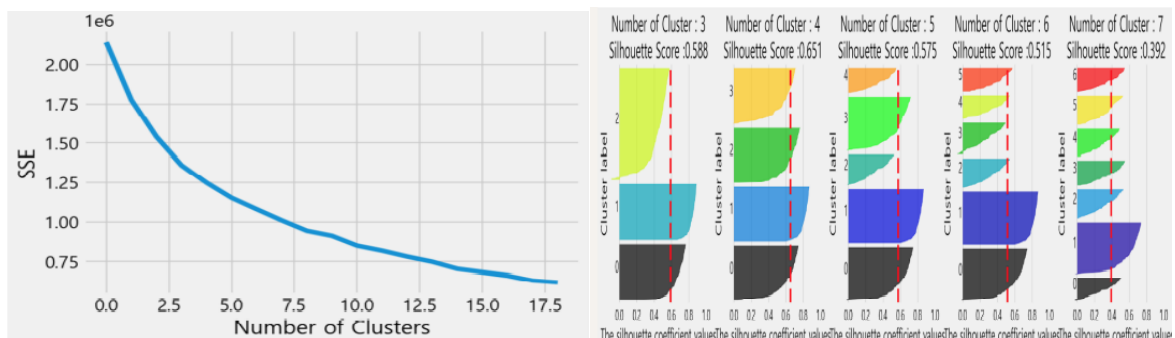


그림 5-1, 5-2. 엘보우 기법, 실루엣 기법 그래프

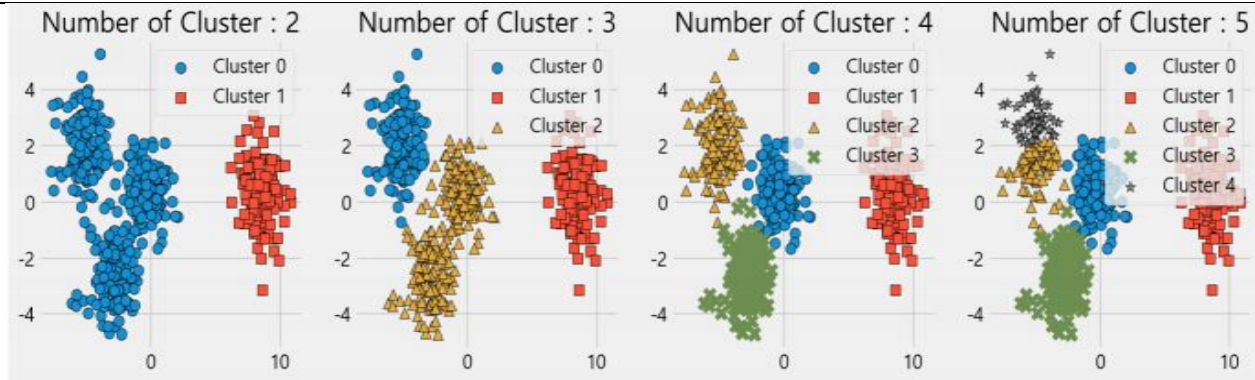


그림 6. 군집(K) 개수에 따른 시각화

군집 분석을 토대로 온라인 소비데이터에 군집화된 그룹들(A(1), B(2), C(3), D(4))을 의미하는 ‘cluster’ 컬럼을 추가하였다. 그리고 최종적으로 그림 7-1 와 같이 각 군집의 소비 건수가 동일하지 않기 때문에, 총 소비금액에 소비 건수를 나눈 ‘한번소비’ 컬럼도 추가하였다. 정제 완료된 온라인 소비데이터에서 나뉜 4 개의 군집은, 4 개의 소비자 집단으로 볼 수 있다. 우리는 4 개의 소비자 집단을 시각화 분석을 통해 여러 시각에서 탐색해보고, 각 집단 별 특성을 파악하고자 한다.

```
cus_cluster['cluster'].value_counts()
```

```
4    104563
3     74562
1     67043
2     59954
Name: cluster, dtype: int64
```

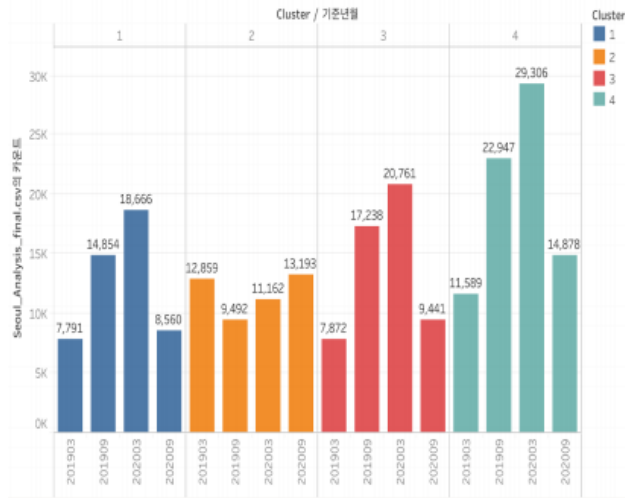
그림 7-1. 각 군집별 소비 건수

	기준년월	품목대분류코드	품목대분류명	품목중분류코드	품목중분류명	성별	연령	가구생애주기	고객소재지_광역시도	고객소재지_시군구	고객소재지_읍면동	매출금액	매출건수	XYZ	Covid	지속소비등급	cluster	한번소비
0	201903	10	의류	1010	여성의류	여성	40	초중고자녀가구	서울특별시	구로구	공동	782521	13	Y	b_cov	실버	4	60193.923077
1	201903	21	e상품/서비스	2110	o2o서비스	여성	20	신혼영유아가구	서울특별시	구로구	공동	963789	29	Z	b_cov	실버	4	33234.103448
2	201903	19	출산/육아	1913	육아용품서비스	여성	30	신혼영유아가구	서울특별시	구로구	공동	924309	46	Y	b_cov	골드	2	20093.673913
3	201903	21	e상품/서비스	2111	e머니/상품권	여성	30	신혼영유아가구	서울특별시	구로구	공동	1089360	54	Y	b_cov	골드	2	20173.333333
4	201903	17	생활	1710	생활용품	여성	30	신혼영유아가구	서울특별시	구로구	공동	1656740	58	Y	b_cov	골드	2	28564.482759

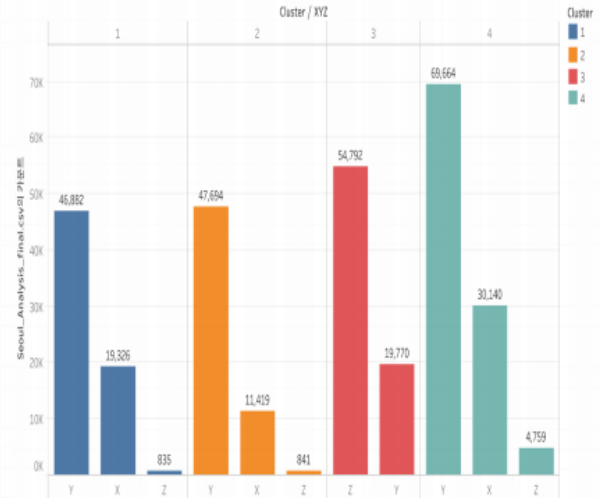
그림 7-2. 시각화 분석에 사용할 데이터

소비자 집단 시각화 분석

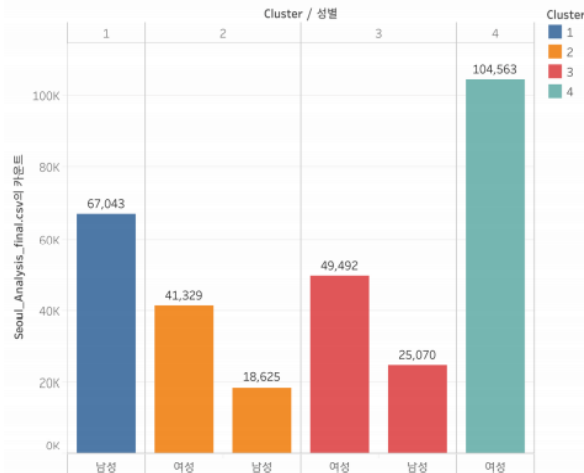
<군집분석_코로나전후>



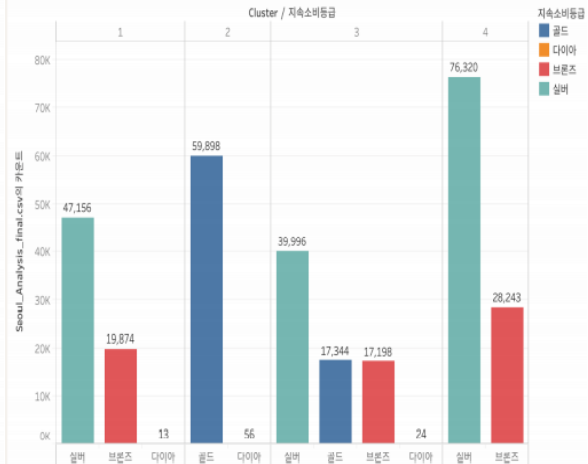
<군집분석_XYZ세대분포>



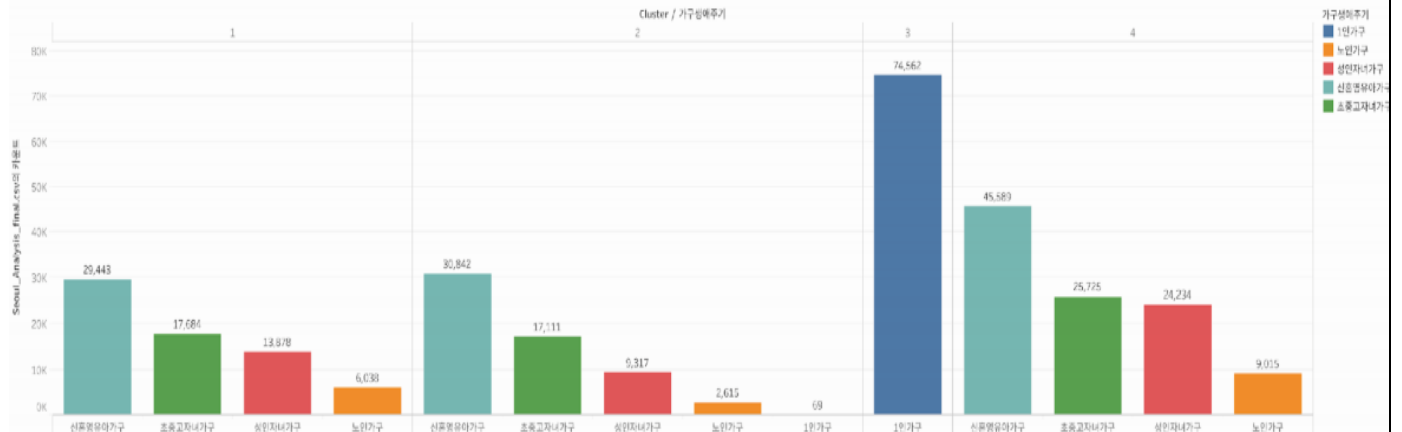
<군집분석_성별>



<군집분석_지속소비등급>



<군집분석_가구생애주기>



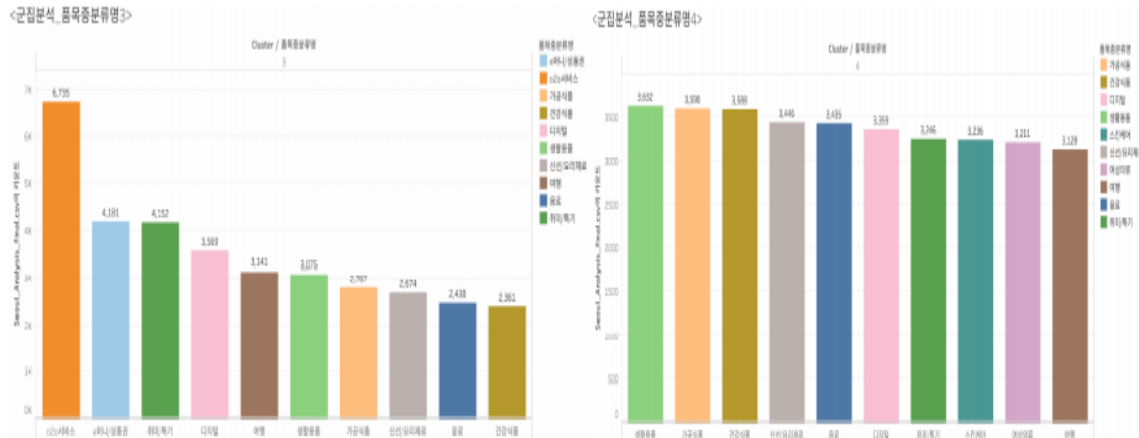
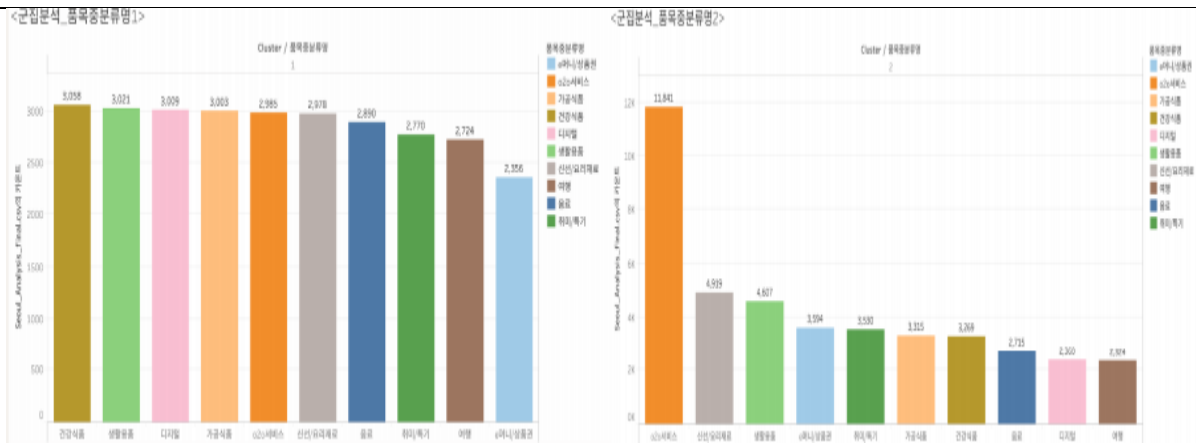


그림 8. 군집에 따른 특성 시각화

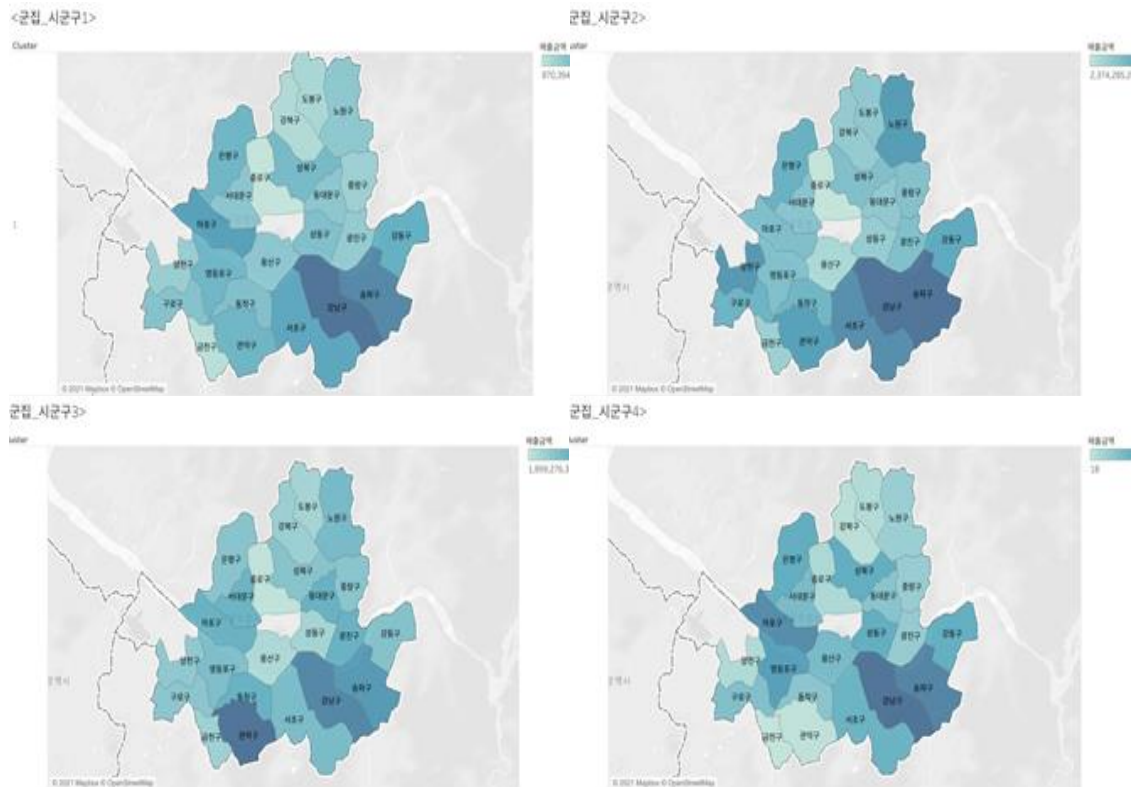


그림 9. 군집 별 지역 소비

<군집-나이대>

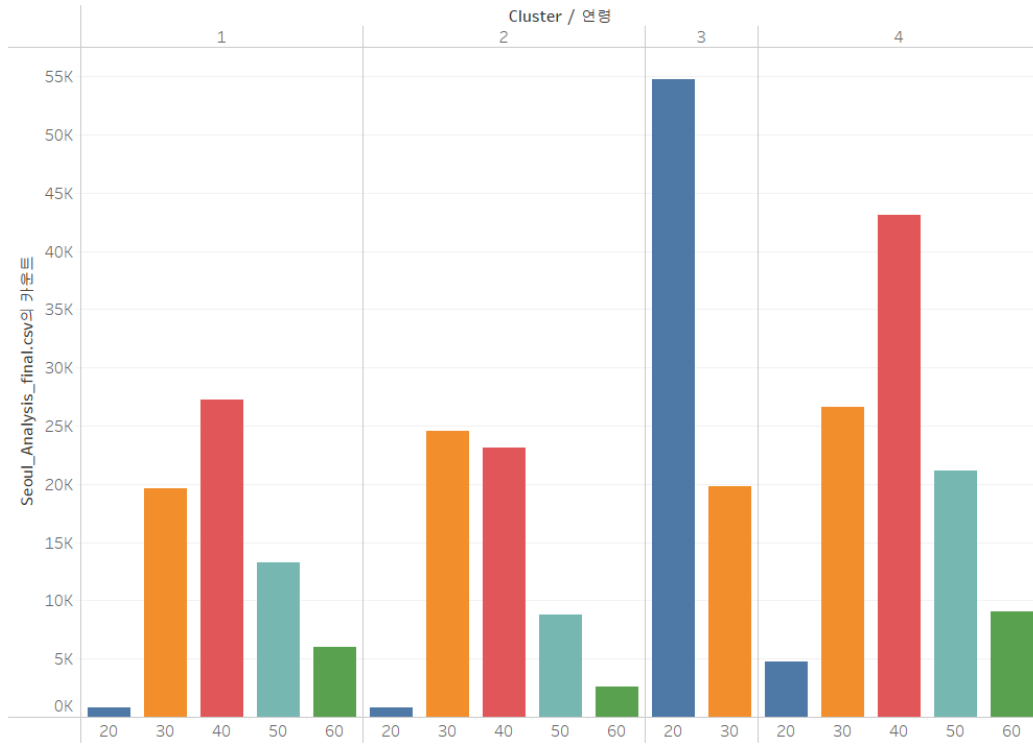


그림 10. 군집 별 나이대

<군집분석개수_매출금액>

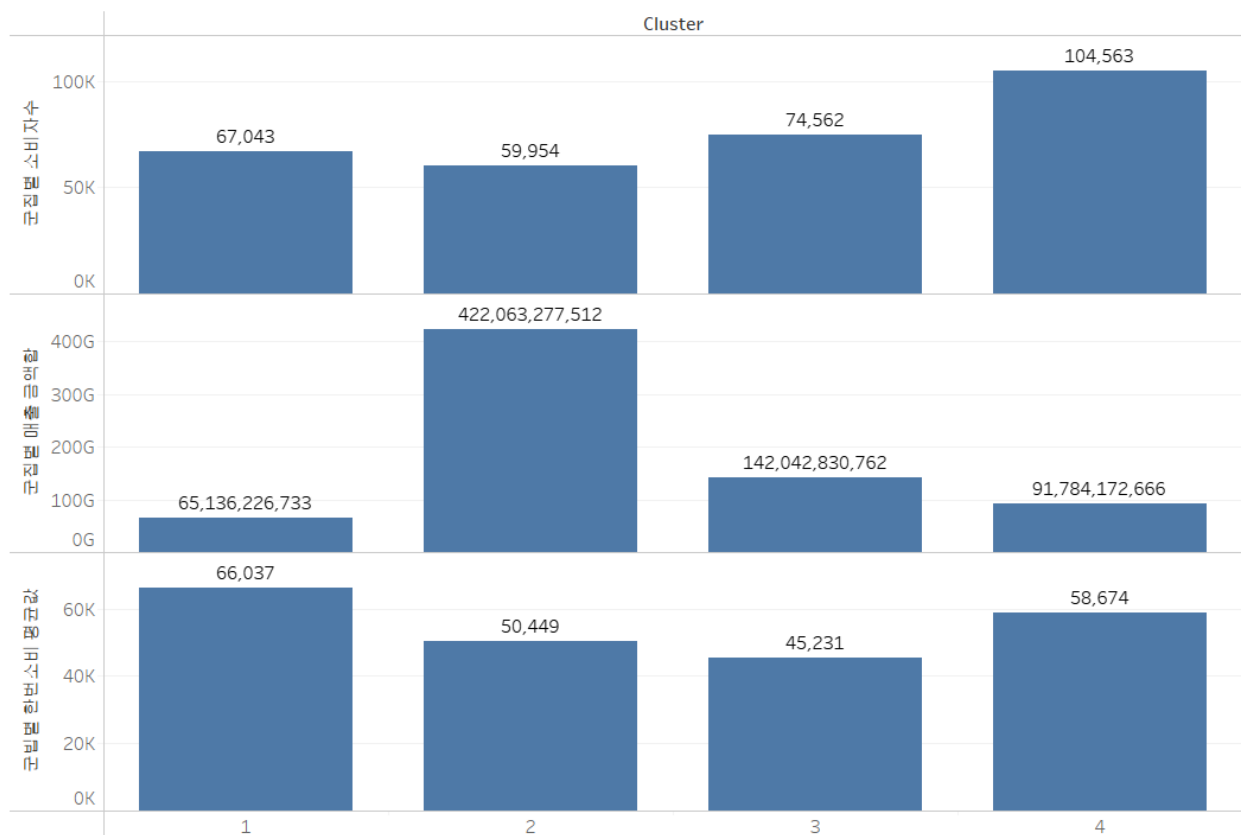


그림 11. 온라인 군집 매출금액 비교

- 카운트: 군집 별 사람의 수
- 매출금액: 군집별 매출금액의 총합
- 한번소비: 매출금액/매출건수

한번 결제한 금액은 A, D 집단이 평균적으로 더 높았지만, 총 매출금액은 B 집단이 압도적으로 많았다. B, C 집단이 한번에 더 작은 금액을 결제하지만, 더 자주 소비한다는 것을 알 수 있다.

군집화된 집단별 코로나 19 전후의 온라인 소비 금액, XYZ 세대의 분포, 성별, 지속소비등급, 가구생애주기, 품목별 소비금액을 시각화 해보았다. 연령과 세대, 구매한 품목 등에 군집별로 확실한 차이를 볼 수 있었다.

군집 별 특성

다음은 4 개의 군집별로 뚜렷하게 나타나는 특성들을 정리한 것이다.

- A :
남성 100%로 이루어진 집단. 30~50 대에 분포. 자녀가 있는 가구. O2O 서비스의 소비가 적음. 특정 품목의 온라인 소비가 뚜렷하지 않음.
- B:
온라인 소비금액이 압도적으로 많은 집단. O2O 서비스의 이용이 활발. 자녀가 있는 가구. 교육열이 높은 강남구, 송파구, 서초구, 노원구, 양천구, 강동구에 주로 분포. 코로나 전/후 온라인 소비 차이가 크지않음.
- C:
20 대 MZ 세대. 1 인가구 100%. 1 인가구가 많은 관악구에 주로 분포. O2O 서비스와 e 머니, 상품권 이용이 매우 활발. e 머니를 이용한 현명한 온라인 소비를 하는 집단으로 추정.
- D:
여성 100%로 이루어진 집단. 30~50 대에 분포. 자녀가 있는 가구. O2O 서비스의 소비가 적음. 생활용품, 건강식품, 신선제품 등에 주로 소비. 주부로 추정.

군집 별 특성 결론:

- O2O 의존도 높은 집단 : B 번 C 번 군집
- O2O 의존도 낮은 집단 : A 번 D 번 군집

오프라인 데이터 군집 예측

각 군집별 ‘온라인’ 소비 패턴을 알아보았지만, 각 군집별 ‘오프라인’ 소비의 특성은 무엇일까? 온라인 소비가 많은 집단은 오프라인으로 유도하고, 오프라인 소비가 많은 집단은 온라인으로 유도하기 위해 각 군집별 오프라인 소비의 트렌드도 함께 파악하는 것이 중요하다.

그러므로, 비씨카드의 온라인 소비 데이터로 군집을 학습시켜 신한카드의 오프라인 소비 데이터로 군집을 분류하는 모델을 만든 뒤에 각 군집별 오프라인 소비에 대한 트렌드도 분석해보았다. 머신러닝 기법으로는 light GBM 을 사용하였다.

	기준년월	품목대분류코드	품목대분류명	품목중분류코드	품목중분류명	성별	연령	가구생애주기	고객소재지_광역시도	고객소재지_시군구	고객소재지_읍면동	매출금액	매출건수	XYZ	Covid	지속소비등급	cluster	한번소비
0	201903	10	의료	1010	여성의료	여성	40	초중고자녀가구	서울특별시	구로구	공동	782521	13	Y	b_cov	실버	4	60193.923077
1	201903	21	e상품/서비스	2110	2o서비스	여성	20	신혼영유아가구	서울특별시	구로구	공동	963789	29	Z	b_cov	실버	4	33234.103448
2	201903	19	출산/육아	1913	육아용품서비스	여성	30	신혼영유아가구	서울특별시	구로구	공동	924309	46	Y	b_cov	골드	2	20093.673913
3	201903	21	e상품/서비스	2111	e상/상품권	여성	30	신혼영유아가구	서울특별시	구로구	공동	1089360	54	Y	b_cov	골드	2	20173.333333
4	201903	17	생활	1710	생활용품	여성	30	신혼영유아가구	서울특별시	구로구	공동	1656740	58	Y	b_cov	골드	2	28564.482759

그림 9-1. 트레인 시킬 비씨카드의 온라인 소비 데이터

	기준년월	광역시도명	업종대분류	업종중분류	신규가입건수	해지가맹건수	매출가맹건수	카드매출금액	카드매출건수	현당매출금액	현당매출금액
0	201903	강원	문화레저	동행	3	1	12	16036000	500	1336292	32071
1	201903	강원	문화레저	동행	5	1	190	283445000	7826	1491816	36218
2	201903	강원	문화레저	동행명연	0	1	88	341305000	4910	3878461	66512
3	201903	강원	문화레저	레저	0	1	80	3575639000	25016	44695491	142934
4	201903	강원	문화레저	레저	1	1	3	748000	9	249333	83111
***	***	***	***	***	***	***	***	***	***	***	***
23867	202009	충북	문화레저	전용매장	0	0	5	2199000	415	439677	5300
23868	202009	충북	문화레저	전용매장	0	0	10	82163000	3971	8216290	20691
23869	202009	충북	문화레저	전용매장	0	0	0	0	0	0	0
23890	202009	충북	문화레저	전용매장	31	23	1726	10572495000	1365915	6125432	7740
23891	202009	충북	NaN	NaN	0	0	0	0	0	0	0

그림 9-2. 테스트할 신한카드의 오프라인 소비 데이터

- Train X: (비씨카드) 기준년월, 품목대분류명, 품목중분류명, 고객소재지_광역시도, 한번소비
- Train Y: cluster
- Test X: (신한카드) 기준년월, 업종대분류, 업종중분류, 광역시도명, 건당매출금액

온라인 데이터(그림 9-1)를 training set 으로 이용하여 학습을 시킨 뒤, 같은 변수의 오프라인 데이터(그림 9-2)를 test set 으로 이용하여 오프라인 소비 패턴을 군집별로 분류해보았다.

```
x_test['cluster'] = w_pred
```

```
x_test['cluster'].value_counts()
```

```
4    831
1    319
3    216
2     25
Name: cluster, dtype: int64
```

그림 10. 오프라인 데이터로 학습된 군집의 count

오프라인 소비가 A, D 집단보다 현저하게 적은 B, C 집단의 카드 건수가 작게 나오는 것을 확인할 수 있다.

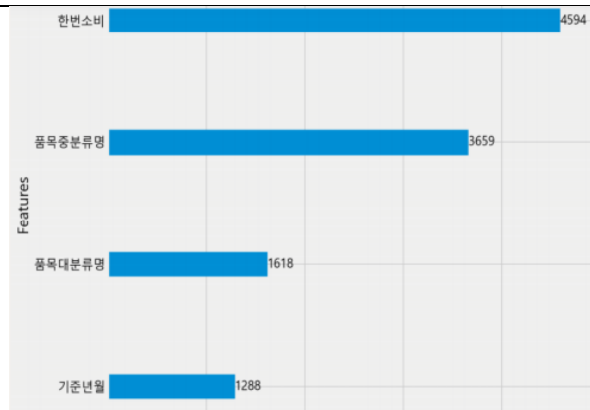


그림 11. 학습에 중요한 변수들

군집분류 시 가장 중요하게 나타난 변수는 ‘한번소비’와 ‘품목중분류명’이었다. 건당 얼마만큼의 금액을 소비하는지, 어떤 품목을 소비하는지가 군집을 분류하는 데에 가장 중요하게 작용했다고 볼 수 있다.

군집별 아이디어 제안

위의 군집분석을 통한 소비집단 분류로, 크게 온라인 서비스를 활발히 이용하는 소비집단 (B,C)과 온라인보다는 오프라인 서비스를 주로 이용하는 소비집단(A,D)로 나눌 수 있다. 이를 통해 A,D 집단은 온라인 O2O 서비스로 유도하고, B,C 집단은 오프라인 O4O 서비스로 유도하는 아이디어를 제안하고자 한다.

오프라인에서 온라인으로 유도 / (A,D) 집단

신한카드의 오프라인 가맹점 데이터로 O2O 서비스의 소비가 적은 집단 (A, D 집단)의 오프라인 소비패턴을 분석하여 오프라인 소비를 온라인으로 끌고와 더욱 간편한 소비를 유도할 수 있다.

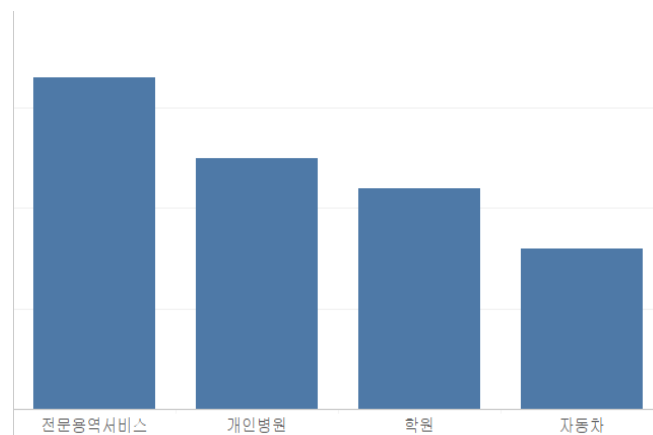


그림 12. A 집단의 품목별 오프라인 소비패턴

그림 12. 에 따르면 30-50 대 남성으로 이루어진 A 집단의 오프라인 소비 중, ‘전문용역 서비스’와 ‘병원’, ‘학원’ 부분이 우세한 것을 볼 수 있다.

‘전문용역서비스’의 소분류에는 공인노무사, 번역/통역, 법무사, 변리사, 변호사, 세무사 등이 있다. ‘전문용역서비스’와 ‘병원’부분이 우세한 것은 법률상담 분야와 헬스케어의 O2O 서비스 사용이 아직 활발하지 않기 때문이라고 추정된다. ‘학원’ 부분이 우세한 것은 A 집단이 모두 자녀가 있는 가구로 구성 된 것을 보아 자녀들의 교육비를 오프라인으로 결제한 것으로 생각할 수 있다.

이 분야의 O2O 서비스들은 모두 오프라인 상에서 수요가 있지만 온라인 상에서 아직 활성화되지 않은 서비스라는 공통점이 있다. 다음 O2O 서비스 사례들로 A 집단의 온라인 소비를 증가시킬 수 있을 것이라 판단된다.

- 법률상담 분야 O2O 서비스
 - 법률서비스 문턱이 낮아지면서 수요가 크게 늘어나는 계기로 삼을 수 있음
 - 참고 사례: 로팜, 세무통
- 상황 별 병원 찾기, 의사 상담, 건강 상식 추천 등 헬스케어 O2O 서비스
 - 포스트 코로나 시대에 최적화된 새로운 진료 모델의 출현은 필연적
 - 온라인과 오프라인 진료를 적절히 융합하여 비용 절감, 효율적인 접근성 보장 등의 문제를 해결할 수 있음
 - 참고 사례: 똑닥, 굿닥
- 학원 매칭 O2O 서비스
 - 학원과 학생/학부모를 연결하는 학원 O2O 서비스
 - 교육시장의 정보 비대칭 문제 해결
 - 학원 랭킹과 리뷰 기능 등

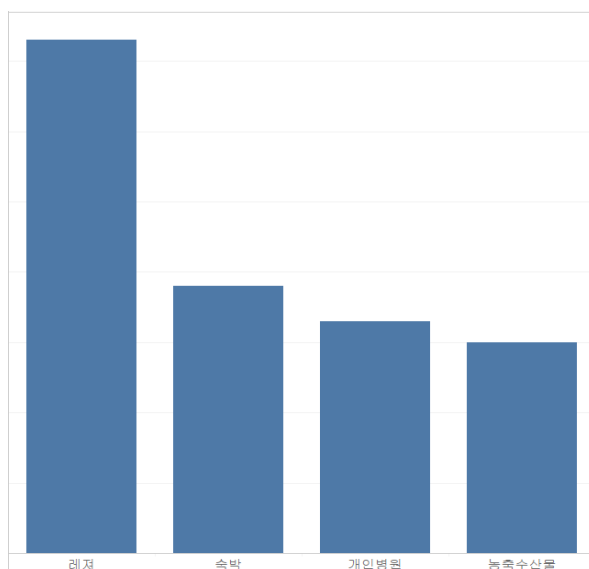


그림 13. D 집단의 품목별 오프라인 소비패턴

반면에, 그림 13에 따르면 30-50 대 여성으로 이루어진 D 집단의 오프라인 소비 중에는 이미 O2O 서비스가 활발하게 이루어지고 있는 ‘레저’와 ‘숙박’이 많은 부분을 차지한다. 이 집단은 숙박/레저의 O2O 서비스에 대해 인지하지 못하고 있거나 오프라인에 비해 별다른 장점을 찾지 못했을 가능성이 높다. D 집단의 온라인 소비 증가를 위해서 이 분야의 O2O 서비스 사례들은 30-50 대 여성들을 타겟으로 한 활발한 마케팅과 전략이 필요할 것으로 분석된다.

- 숙박/레저 O2O 서비스
 - 고객의 취향에 맞춰 숙소를 기반으로 개인화된 레저 서비스 제공 가능
 - 참고사례: 야놀자, 여기어때

온라인에서 오프라인으로 유도 / (B,C) 집단

O2O 서비스의 이용이 활발한 집단 (B, C 집단)에게는 O4O 를 활용하여 새로운 사업 구축하는 방안이 있다. B,C 집단은 이미 온라인에서 소비를 활발하게 이용하고 있으며, 적극적으로 O2O 서비스를 이용하고 있다. O2O 서비스를 적극적으로 이용하는 소비자집단은 O4O 서비스로 유도하여 오프라인으로 다시 소비자들을 유도하여, 온,오프라인의 상생을 이끌수 있다.

아마존 고, 무신사 테라스의 참고사례로 보면 기존에 온라인에서 주로 사업을 전개했던 회사지만, 온라인 거래로 쌓은 막대한 크기의 DB 와 첨단기술을 기반하여 오프라인으로 진출한 대표적인 O4O 서비스이다.

앞서 말했듯, 코로나 19 로 인해 비대면 소비가 폭발적으로 증가함에 따라 온라인 거래액이 역대 최고치를 기록하고 있지만, 반대로 오프라인 점포들은 빠르게 감소하는 매출액과 잦은 폐점으로 죽어가고 있다. 온라인 채널이 시장을 독식하는 것은 옳지 않고, 오프라인과 온라인 시장의 상생이 이루어지는 것이 바람직하다. 이에 따라, 온라인 소비만을 추구하는 집단을 오프라인으로 끌고와 오프라인 시장을 활성화하는 것이 필요하다. 이에 온라인 중심 업체들은 O4O 플랫폼을 적극적으로 활용, 오프라인으로 진출하여 활동 범위를 확장할 필요가 있다.

아래는 O2O 서비스와 온라인을 적극적으로 활용하는 집단이 주로 이용하는 서비스들의 오프라인 진출 아이디어를 정하였다.

- 음식 배달 서비스의 오프라인 진출
 - 오프라인 시식 매장을 운영
 - 지역의 배달 음식점들을 선정하여 고개들이 시식할 수 있는 오프라인 공간을 제공함으로써, 리뷰에 의존하던 음식 주문하던 약점을 극복 가능.
- 중고거래 서비스의 오프라인 진출
 - 당근마켓으로 되는 지역 중고거래앱의 오프라인 매장 마련
 - 지역 커뮤니티 오프라인 공간을 마련을 마련 하는 효과 기대
- 온라인 상거래 서비스 오프라인 진출
 - 온라인에서 많이 팔리는 물품들의 유사성을 고려해서 품목별로 체험용 공간 마련
 - 쿠팡 + 코스트코 => 쿠팡코스트

4. 기대효과

우리는 소비자 집단을 군집분석을 통해 나누고, 각 집단별 특성을 파악함으로써 각 집단 별 마케팅 아이디어를 도출 할 수 있었다. 위의 아이디어를 통해 우리는 3 가지의 측면에서 기대효과를 살펴볼 수 있는데, 소비자 특성 별 마케팅을 진행해야 하는 온라인 사업자 측면, 오프라인 사업 의존도가 높은 오프라인 사업자 측면, 그리고 금융 서비스를 제공해야 하는 은행측면에서 각기 다른 효과를 기대해 볼 수 있다.

온라인 사업자 측면

- 소비자 유형을 바탕으로 한 행동 패턴 분석으로 차별적인 맞춤형 체험, 구매 공간으로서의 오프라인 구축과 이를 바탕으로 한 효과 적인 옴니 채널 방안 마련의 가능성.

오프라인 사업자 측면

- O2O 서비스가 활발하지 않은 분야의 경우 온라인 과 협업해서 새로운 플랫폼을 구축해 신규 유저 영입하기.

은행 측면

- O2O 서비스 관련 스타트업 창업 지원시 사업타당성 분석의 기반 제공.
- O2O 서비스, E 머니/상품권은 미래에 새로운 화폐가 될 가능성이 있으므로 오프라인 가맹점들은 자신의 가맹점 포인트를 현금과 교환시키는 작업이 중요. 은행 전용 O2O 포인트 앱을 만들어 저축을 유도.

5. 활용 데이터

- [2021 금융데이터경진대회] 신한카드 - 광역시도별 업종별 가맹점 데이터
- [2021 금융데이터경진대회] 비씨카드 - 온라인 품목별 소비데이터


6. 참고자료

- 과기정통부 보도자료, “O2O 서비스, 코로나 19 비대면 시대 쾌속 성장 과기정통부, 「2020 년 O2O 서비스 산업조사」 결과 발표”, 2021. 4. 8.
- 한국갤럽조사연구소, “2012-2021 스마트폰 사용률 & 브랜드, 스마트워치, 무선이어폰에 대한 조사”, 2021.06.03
- 정보통신산업진흥원, “O2O 서비스 시장조사”, 2019.12

※ 작성 시 유의사항

- 분량 제한 없음 /글자 폰트 크기 11 포인트(한글 및 워드로 작성)
- 도표, 이미지 등 활용 가능
- 설명을 위한 추가자료 첨부 가능
- 제출 시 표지를 함께 제출하되, 식별이 가능하도록 참가자 명(팀의 경우 팀장 및 팀원명)을 작성하여 제출

2021 년 09월 06일

참가자(대표자) (인  서명)

금융보안원 귀중