



## 데이터 수집 계획 및 보고서

효과적인 마케팅을 위한 온라인 쇼핑 소비자 평가 데이터 분석

송하은

23.05.16 – 23.05.26

## 목차



### 개요

프로젝트명, 수행기간, 시스템 환경



### 분석 배경 및 필요성



### 데이터 전처리



### 데이터 전처리 결과



### 데이터 분석

가설, 결과, 근거



### 한계점

어려웠거나 부족했던 점

## 개요

### 프로젝트명

효과적인 마케팅을 위한  
온라인 쇼핑 소비자 데이터 분석

### 수행 기간

2023.05.16 – 2023.05.26

### 시스템 환경

window 10  
conda 23.1.0  
python 3.8.10  
pandas  
matplotlib  
seaborn

## 분석 배경 및 필요성

급변하는 현대사회에서 온라인을 이용한 쇼핑은 점차 확대되고 있다.

지속적으로 온라인을 통한 소비자거래가 늘어남에 따라  
2019년부터 2021년도까지의 3년간 서울시 온라인 쇼핑을 이용한 소비자를 대상으로 집계된  
평가데이터를 파악하고 이를 바탕으로 현 소비자들이 선호하는 쇼핑물 유형을 분석하고,  
효과적인 마케팅을 위해 접목시키고자 한다.

## 데이터 전처리

```
In [13]: ## 필요없는 나머지 년도 제거 (2019~2021년 제외)
year = shop[shop['평가년도'] <= 2018].index

shop.drop(year, inplace=True)

year2 = shop[shop['평가년도'] > 2021].index

shop.drop(year2, inplace=True)
shop[:3]
```

Out [13]:

평가년 도	연번	쇼핑물 구분	쇼핑물 명	도메인명	소비자보호 평가	이용자만족평 가	피해발생평 가	전체평 가
50	2021	2899	종합쇼핑물	코스트코 www.costco.co.kr	49.0	27.58	10.0	86.58
51	2021	2934	의류	보리보리 www.boribori.com	48.0	28.28	10.0	86.28
52	2021	2935	의류	무신사 store.musinsa.com	48.0	30.01	8.0	86.01

가장 최근 데이터인 2022년은 월말까지 데이터가 존재하지 않기 때문에 2019 – 2021년까지로 기준년도 설정, 최근 3년간의 데이터만 필요하므로 나머지 년도의 데이터 삭제

```
In [14]: ## 필요없는 열 제거
shop.drop(columns=['연번', '도메인명', '피해발생평가'], inplace=True)

## 열 이름 변경
shop.rename(columns={'쇼핑물 구분':'쇼핑물구분', '쇼핑물 명':'쇼핑물명'}, inplace=True)
shop[:3]
```

Out [14]:

평가년도	쇼핑물구분	쇼핑물명	소비자보호평가	이용자만족평가	전체평가	
50	2021	종합쇼핑물	코스트코	49.0	27.58	86.58
51	2021	의류	보리보리	48.0	28.28	86.28
52	2021	의류	무신사	48.0	30.01	86.01

분석에 사용되는 데이터를 제외한 나머지 열 삭제 후, 띄어쓰기로 작성되어 있는 열 이름 변경

## 데이터 전처리

```
In [15]: ## 결측치 있는지 확인
shop.isnull().sum()

## 결측치 있는 행 추출
shop[shop.쇼핑물구분.isnull()]

## 결측치 있는 데이터 다른년도에 (티켓)으로 있으므로 결측치 -> 티켓으로 변경
shop.fillna('티켓', inplace=True)
shop[shop.쇼핑물구분 == '티켓']

## 결측치 있는지 재확인
shop.isnull().sum()
```

```
Out [15]: 평가년도      0
쇼핑물구분      0
쇼핑물명      0
소비자보호평가      0
이용자만족평가      0
전체평가      0
dtype: int64
```

결측치 있는 행을 먼저 확인하여 해당 행의 다른 년도  
데이터 유무 파악 후,  
[티켓]으로 해당 데이터가 확인되므로 결측치를 모두 [티켓]으로  
이름 변경

```
In [16]: ## 필요없는 데이터 확인
com = shop[shop['쇼핑물구분'] == '컴퓨터'].index

## 필요없는 데이터(컴퓨터) 제거
shop.drop(com, inplace=True)

## 제거(컴퓨터) 확인
shop[shop.쇼핑물구분 == '컴퓨터']

## 필요없는 데이터 확인
commerce = shop[shop['쇼핑물구분'] == '소셜커머스'].index

## 필요없는 데이터(소셜커머스) 제거
shop.drop(commerce, inplace=True)

## 제거(소셜커머스) 확인
shop[shop.쇼핑물구분 == '소셜커머스']
```

```
Out [16]: 평가년도  쇼핑물구분  쇼핑물명  소비자보호평가  이용자만족평가  전체평가
```

분석에 사용되지 않는 데이터를 확인하여  
[컴퓨터]와 [소셜커머스] 데이터를 삭제

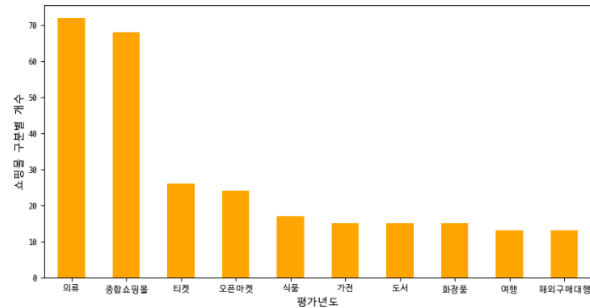
## 데이터 전처리 결과

```
In [23]: ## 시각화
plt.rc('font', family='Dcoding', size=9)
plt.figure(figsize=(10, 5))

shop_category.plot(kind='bar',
                    rot=0,
                    color='orange')

plt.xlabel('평가년도', fontsize=11)
plt.ylabel('쇼핑몰 구별별 개수', fontsize=11)
plt.title('쇼핑몰구분에 따른 이용 수', pad=15, fontsize=15)
plt.show()
```

쇼핑몰구분에 따른 이용 수



- [가설1] 가장 많이 이용하는 쇼핑몰구분은 의료일 것이다.
  - 2019년부터 2021년까지 중 가장 많이 이용한 쇼핑몰구분은 의료로 총 72개이며, 가장 낮은 쇼핑몰구분은 해외구매대행과 여행으로 총 13개이다.

### 가설1

#### : 쇼핑몰 구분에 따른 이용 수

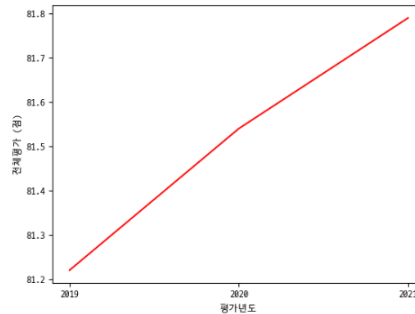
대체로 코로나 이전부터 온라인 시장에서 인기가 좋았던 쇼핑몰구분이 상위에 많이 분포되어 있고, 해외와 관련된 쇼핑몰구분은 타격을 입은 것을 볼 수 있다.

```
In [25]: ## 시각화
from matplotlib.ticker import MaxNLocator
ax = plt.figure().gca()
shop_year.plot(ylabel='전체평가 (점)',
               color='r')

ax.xaxis.set_major_locator(MaxNLocator(integer=True))
plt.title('평가년도별 전체평가 평균', pad=15, fontsize=12)

Out[25]: Text(0.5, 1.0, '평가년도별 전체평가 평균')
```

평가년도별 전체평가 평균



- [가설2] 평가년도에 따른 전체평가가 점점 상승할 것이다.
  - 온라인 쇼핑 전체평가의 평균은 2019년 81.22로 가장 낮았으며, 2021년에 81.79로 가장 높으며 점점 상승하는 추세이다.

### 가설2

#### : 평가년도별 전체평가 평균

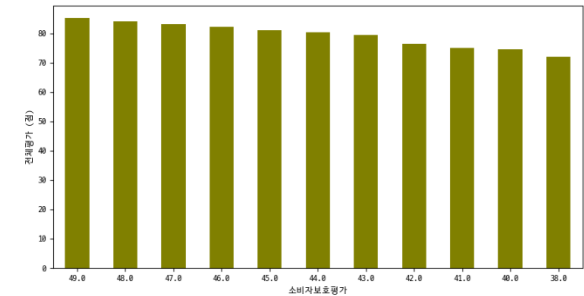
코로나로 인해 온라인 쇼핑을 이용하는 소비자가 증가하며 전체평가도 함께 높아지고 있다고 추측된다.

```
In [30]: ## 시각화
shop_all_user.plot(kind='bar',
                   figsize=(10, 5),
                   rot=0,
                   color='olive',
                   ylabel='전체평가 (점)',
                   )

plt.title('소비자보호평가에 따른 전체평가', pad=15, fontsize=12)

Out[30]: Text(0.5, 1.0, '소비자보호평가에 따른 전체평가')
```

소비자보호평가에 따른 전체평가



- [가설3] 소비자보호평가가 점수가 높을수록 전체평가가 높을 것이다.
  - 소비자보호평가 점수가 가장 높은 49점일 때, 전체평가 점수도 85.22점으로 가장 높다.
  - 소비자보호평가 점수가 가장 낮은 41점일 때, 전체평가 점수도 72.05점으로 가장 낮다.
  - 소비자보호평가 점수에 따라 전체평가 점수를 나열해 비교해봤을 때, 소비자보호평가 점수와 전체평가 점수가 비례한다는 것을 알 수 있다.

### 가설3

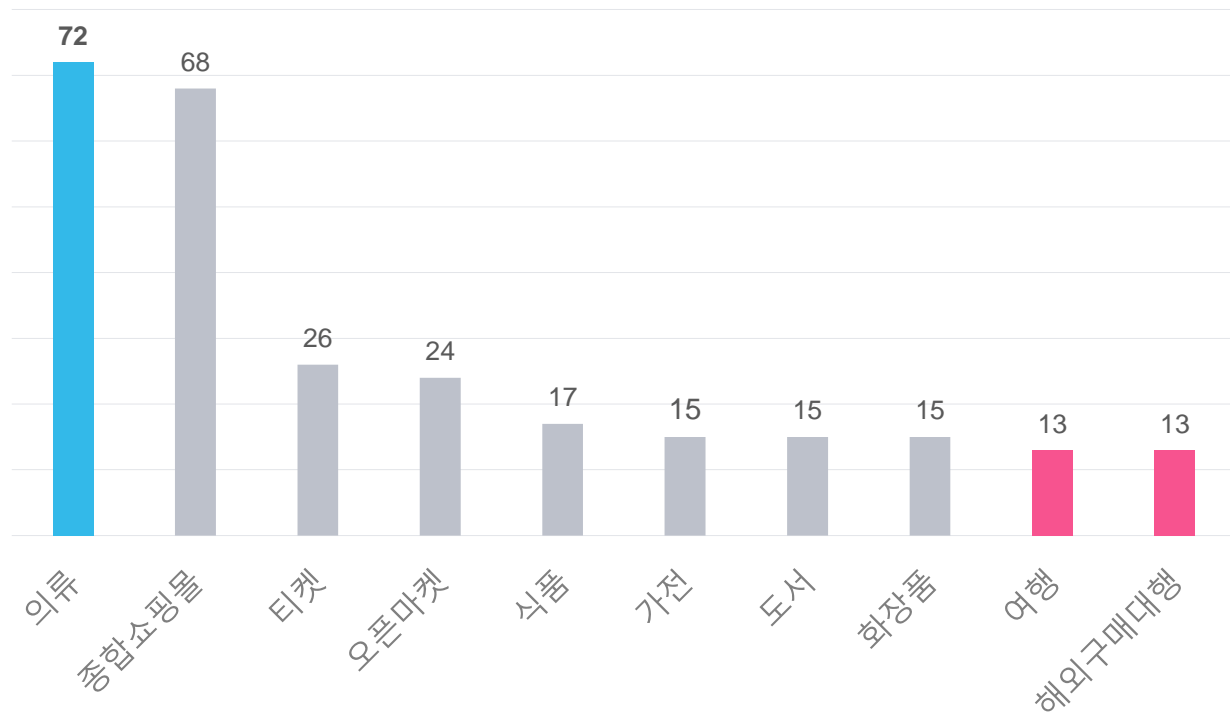
#### : 소비자보호평가에 따른 전체평가

온라인 쇼핑 이용자가 증가함에 따라 온라인 거래를 통한 사기범죄가 늘어나고 이에 따라 소비자보호평가가 중요해졌다는 것을 확인할 수 있었다.

## 데이터 분석

01

[가설1]  
가장 많은 쇼핑물구분은 의류일 것이다.



### 가장 높은 쇼핑물구분

: 의류

2019년부터 2021년까지 중 가장 많이 이용한  
쇼핑물구분은 의류로 총 72개이다.

### 가장 적은 쇼핑물구분

: 여행, 해외구매대행

가장 낮은 쇼핑물구분은 해외구매대행과 여행으로  
총 13개이다.



## 데이터 분석 결과 [가설1]

01

[가설1]  
가장 많은 쇼핑물구분은 의류일 것이다.

● 3년간의 쇼핑물구분별 개수를 파악한 결과,  
가장 많은 쇼핑물 구분이 의류일 것이라고 예상한 것과 같이  
의류가 72개로 가장 많았다.  
가장 적은 쇼핑물구분은 해외구매대행과 여행으로 총 13개이다.

- 코로나 발생으로 인해 해외무역이 어려워지면서 해외여행 관련 또는 해외구매대행 쇼핑물에 타격을 입은 것으로 보인다.
- 의류는 기존에도 온라인산업이 확장된 산업 중 하나이고, 코로나로 인해 오프라인 매장 운영이 어려워지며 온라인 시장이 더욱 확대된 것으로 확인된다.

[코로나19] "신상품 주기 앞당기고 당일 배송까지" ...출구 찾는 패션업계

머니투데이방송 [이유나](#) 기자

입력 2020-03-20 11:31:37



'돌발악재 막아라' 경제팀 초긴장

신종코로나 장기화되면 회복세 보이던 경제 악영향 ... 연일 대응회의

2020-01-29 11:01:47 게재



1월 수출도 '마이너스'...14개월 연속 내리막

출판표 기자 | ⓒ 최초승인 2020.02.02 12:12:48 | ⓒ 최종수정 2020.02.02 17:12

[앵커멘트]

'코로나19' 여파로 소비자들의 쇼핑패턴이 변화하고 있습니다. 오프라인 매장 대신 온라인 쇼핑을 찾는 소비자들이 늘어나고 있는건데요. 패션업체들도 추세에 맞춰 온라인 전용 제품을 출시하는가 하면 의류 당일배송에 나서며 소비자들을 공략하고 있습니다. 이유나 기자의 보도입니다.

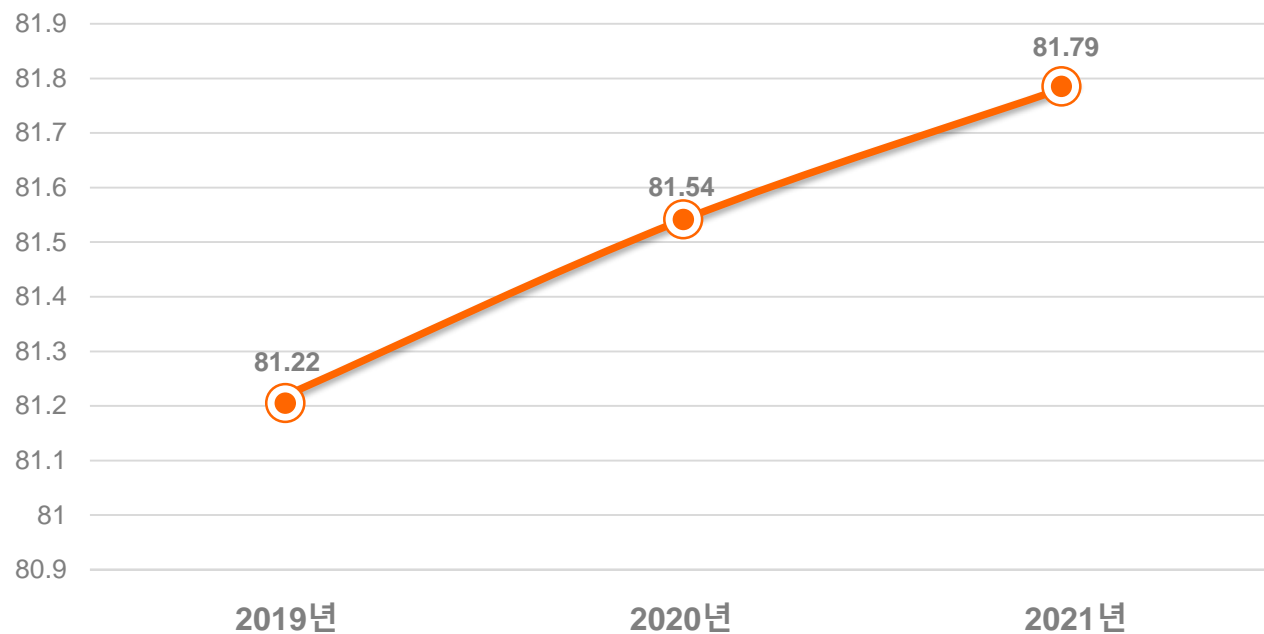
| 전년대비 6.1% 감소 433억5000만달러로 집계



올해 1월 수출이 마이너스를 기록하며 14개월 연속 마이너스 행진을 이어갔다.

산업통상자원부는 1월 수출이 지난해 같은 달보다 6.1% 감소한 433억5000만달러로 집계됐다고 1일 밝혔다.

## 데이터 분석



02

[가설2]  
평가 연도에 따른 전체평가가 점점 상승할 것이다.

가장 높은 평가 연도

: 2021년

2021년에 81.79로 가장 높으며 연도에 따라  
점점 상승하는 추세이다.

가장 낮은 평가 연도

: 2019년

온라인 쇼핑 전체평가의 평균은 2019년 81.22로  
가장 낮다.

## 데이터 분석 결과 [가설2]

02

[가설2]  
평가 연도에 따른 전체평가가 점점 상승할 것이다.

● 연도별 전체평가의 평균을 확인한 결과,  
연도가 지날수록 전체평가가 점점 높아질 것이라고 예상한 것과 같이  
2021년이 81.79점으로 가장 높았다.  
가장 낮은 평가 연도는 2019년으로 총 81.22점이다.

- 코로나 발생으로 인해 집콕생활이 늘어나면서 비대면 쇼핑이 늘어나고,  
이에 따라 오프라인을 즐겨 이용하던 50-60세대도 온라인 쇼핑이 늘어난  
것을 알 수 있다.
- 다양한 연령층이 온라인을 이용하면서 온라인 쇼핑에 대한 선호도가  
높아지고 이를 바탕으로 평가 점수도 함께 올라갔다고 추측할 수 있다.

〈연령대별 온 오프라인 카드 소비 증감액(%)〉



이러한 온라인 소비 증가 양상에서 한 가지 주목할 만한 점이 있다. 바로 디지털 기기에 익숙지 않은 5060세대의 온라인 소비가 큰 폭으로 증가했다는 점이다. 앞선 서울시 통계에 따르면 작년 한 해 동안 모든 세대의 온라인 소비가 전 년 대비 증가했지만, 그 중에서도 50대의 온라인 소비 증가율이 22.3%로 가장 높았으며 60, 70대 이상의 온라인 소비는 1% 이상 증가했다.

### 쇼핑비용 62% 온라인에서 썼다

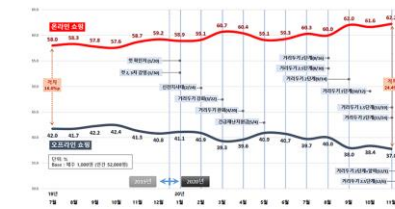
관리자 | 2020.12.23

#### 쇼핑비용 62% 온라인에서 썼다

- 작년 말~올 연초 59%에서 9월 이후 62%로 사상최대
- 코로나19로 수년 걸릴 변화 반년 만에 이뤄져
- 거리두기 영향 점점 덜 받아-증가세 계속될 듯

그렇다면 5060세대는 어떤 온라인 쇼핑 플랫폼을 사용하고 있을까? 먼저, 신선식품 배송 서비스 '마켓컬리'다. 마켓컬리는 2020년 1월 20일부터 5월 18일까지 50대, 60대 이상 회원이 전년 동기 대비 각각 112%, 122% 증가했으며 이에 따라 5060세대 매출 역시 크게 증가했다고 발표했다. 두 번째로 '카카오뱅크'의 구매 데이터 분석 결과에 따르면 2020년 추석 시즌 시은 50대 거래액과 60대 거래액은 전년 동기 각각 115%, 122% 증가하였다. 특히 60대 거래액 증가율은 전 연령대에서 가장 높은 수치이다. 이러한 추세를 발맞추어 카카오페이도 5060세대를 겨냥한 '살' 선물 기획전을 선보였다.

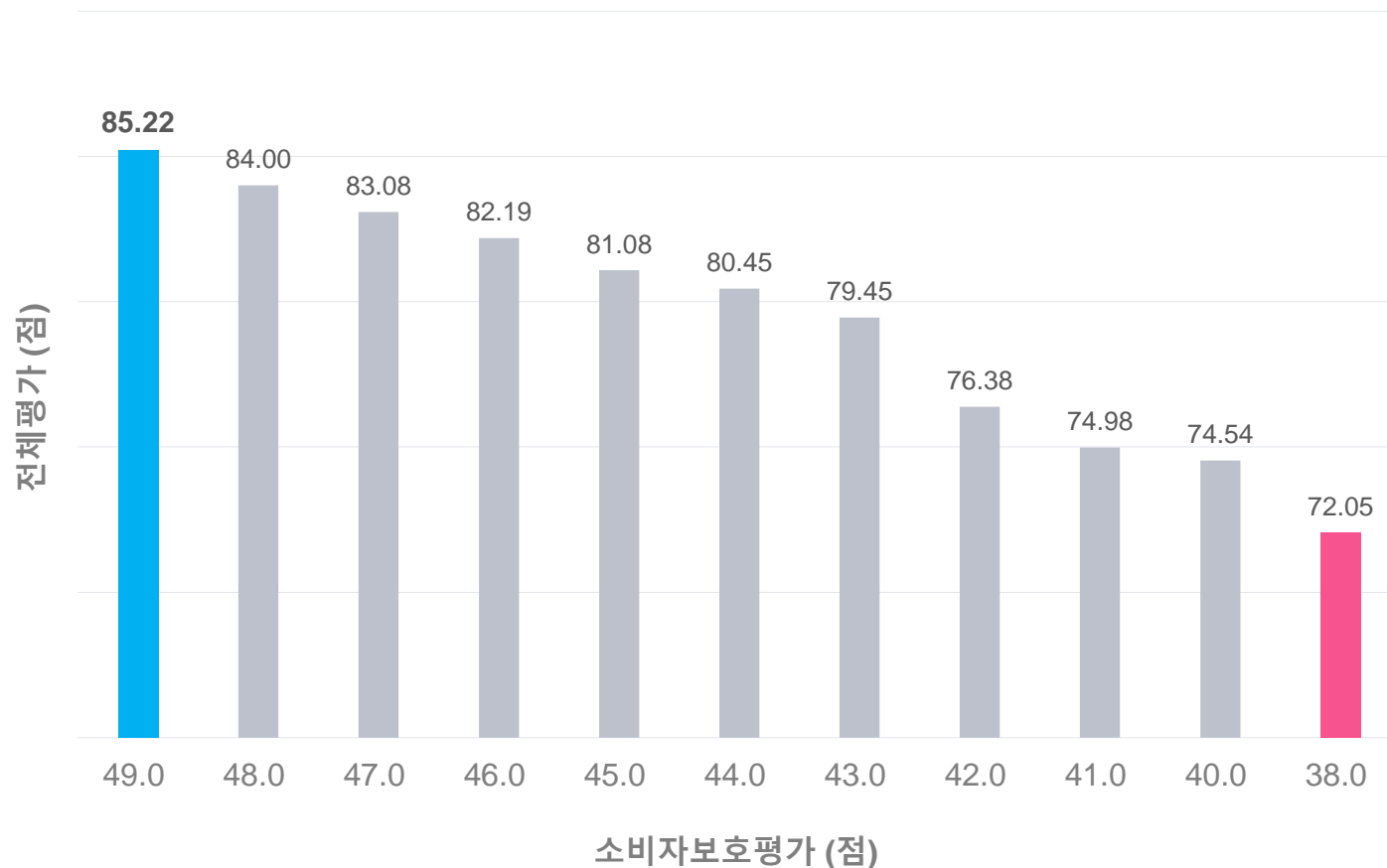
〔그림1〕 쇼핑비 온-오프라인 지출 비중 추이 (2019.7-2020.11)



A1. 지난 한달 간 쇼핑비를 100%라고 했을 때 오프라인 지출 비중이 차지 했는지 (%)이다.  
〔그림1〕 쇼핑비 온-오프라인 지출 비중 추이

작년 하반기부터 올해 초까지 58~59%선을 유지하던 쇼핑 지출 중 온라인 비중이 연초부터 급등해 지난 9월 이후 62%선까지 올라섰다. 수년 걸릴 쇼핑 형태 변화가 약 반년 만에 이뤄졌으며 이를 촉발한 것은 코로나19 확산을 막기 위한 사회적 거리두기와 이에 따른 연례도 줄었다.

## 데이터 분석



03

[가설3]  
소비자보호평가 점수가 높을수록 전체평가가 높을 것이다.



소비자보호평가 점수에 따라 전체평가 점수를 나열해 비교해봤을 때, 소비자보호평가 점수와 전체평가 점수가 비례하다는 것을 알 수 있다.

소비자보호평가 대비 전체평가

가장 높은 점수 : 85.22점

소비자보호평가 점수가 가장 높은 49점일 때,  
전체평가 점수도 85.22점으로 가장 높다.

소비자보호평가 대비 전체평가

가장 낮은 점수 : 72.05점

소비자보호평가 점수가 가장 낮은 38점일 때,  
전체평가 점수도 72.05점으로 가장 낮다.

## 데이터 분석 결과 [가설3]

03

[가설3]  
소비자보호평가 점수가 높을수록 전체평가가  
높을 것이다.

- 소비자보호평가별 전체평가의 평균을 확인한 결과, 소비자보호평가가 높을수록 전체평가가 높을 것이라고 예상한 것과 같이 소비자보호평가 49점일 때, 전체평가가 85.22점으로 가장 높았다. 가장 낮은 평가 년도는 소비자보호평가가 38점일 때, 전체평가는 72.05점이다.
- 코로나 발생으로 인해 온라인 쇼핑이 급증하면서 그로 인한 쇼핑물 피해 사례도 함께 늘어났다.
- 소비자들이 방역물품을 다량으로 구매하면서 이를 빌미로 사기피해가 늘어나며 소비자보호법 개정을 요구하는 등 소비자보호에 대한 인식이 강해졌다고 볼 수 있다. 이에 따라, 소비자보호법에서 좋은 평가를 받을수록 전체평가 역시 좋은 점수를 받았다고 추측할 수 있다.

전자상거래 소비자 피해를 막기 위해  
서울시와 한국소비자연맹이 설립해 운영하는  
[서울시전자상거래센터]



## 한계점 및 보완할 점

### 한계점

- ① 서울시의 상위100개 쇼핑몰로 집계된 데이터이기 때문에 결과를 일반화하기 어려움.
- ② [가설1]의 경우, 쇼핑몰구분별 매출액의 차이가 있기 때문에 가장 많이 집계된 쇼핑몰구분은 의류이지만, 수익률도 이와 비례한 지 확인 불가능.
- ③ [가설3]의 경우, 소비자보호평가의 세부적인 기준을 알 수 없어서 이 점수가 전체평가와 어떤 상관관계를 이루고 있는지 정확한 파악이 어려움.

### 보완할 점

- ① 특정 쇼핑몰구분에 대한 지역별 통계 수치를 함께 비교해도 좋을 것이다.
- ② 평가를 해준 소비자의 연령대, 성별 등을 바탕으로 통계 수치를 비교해도 좋을 것이다.
- ③ 최근 연도인 2022년, 2023년에 대한 결과가 없으므로 최근 연도에 따른 통계 수치를 비교해도 좋을 것이다.

[ 평가 데이터 기준 관련 기사 링크 ]

<http://www.discoverynews.kr/news/articleView.html?idxno=759930>

소감

문서형식으로 작성된 데이터를 필요한 부분만 가공하고, 시각화한 데이터를 토대로 분석하여 결과를 만드는 과정을 직접 경험할 수 있었습니다.

순차적으로 코드 작업을 하고 직접 설정한 가설대로 결론을 추출하면서 계획적이고 꼼꼼하게 일을 수행하는 것이 중요하다는 것을 느끼는 과정이었습니다.

작성한 코드대로 진행 되가는 과정을 보며 신기하기도 하고 새로운 경험이라 어렵기도 했지만 이번 과정으로 코드작업의 다양성을 배울 수 있었습니다.

THANK YOU !