

Term Project 최종 발표

- 코로나 사태와 인터넷 쇼핑 검색, 클릭량의 상관 관계 분석

컴퓨터공학부
201502755 김미연

발표 진행 순서

1. 프로젝트 목표 소개
2. 중간발표까지의 진행 상황
3. 중간발표 이후 진행 상황
4. 최종 결과 분석
5. 고찰
6. QnA

1. 프로젝트 목표 소개


세 가지 가설

1. 코로나 사태 이후, 화장품/미용 상품 중, 얼굴 위쪽을 미용할 수 있도록 해주는 상품의 검색 및 클릭량이 얼굴 아래쪽을 미용하는 상품의 검색 및 클릭량보다 상대적으로 높아졌을 것이다.
2. Pandemic 상황이 지속되면서, 인터넷 쇼핑에서 클릭량 및 검색량이 증가한 상품들이 있을 것이고, 그에 반해 하락한 상품들이 있을 것이다.
3. 코로나 확진자 수가 많았던 날 / 적었던 날에 유독 검색 및 클릭량이 증가 / 감소했던 특정 상품들이 있을 것이다.

**위 세 가지 가설을 최대한 증명해보이고, 자세히 분석하여
코로나 시대에 쇼핑 분야에서 의미 있는 insight를 제공하는 것이 목표!**

2. 중간발표까지의 진행 상황

Naver의 쇼핑인사이트 API를 사용해 데이터 수집.

쇼핑인사이트 API 개요 

쇼핑인사이트 API는 [네이버 데이터랩의 쇼핑인사이트](#)를 API로 실행할 수 있게 하는 RESTful API입니다.

네이버 통합검색의 쇼핑 영역과 [네이버쇼핑](#)에서의 검색 클릭 추이 데이터를 JSON 형식으로 반환합니다. 쇼핑 분야의 분야별 검색 클릭 추이와 특정한 쇼핑 분야에서 검색 키워드별 검색 클릭 추이를 확인할 수 있습니다.

API를 호출할 때는 쇼핑 분야와 검색 키워드, 검색 조건을 JSON 형식의 데이터로 전달합니다.

쇼핑인사이트 API의 하루 호출 한도는 1,000회입니다.

2017년 8월 1일부터 쇼핑인사이트 API 지원 시작.

-> 2017년 8월 1일부터 2020년 10월 31일까지의 데이터 수집.

중간 발표 전까지 수집한 데이터

- 1. 패션의류_남성의류
- 1. 패션의류_여성언더웨어_잠옷
- 1. 패션의류_여성의류
- 2. 패션잡화_남성가방
- 2. 패션잡화_남성신발
- 2. 패션잡화_모자
- 2. 패션잡화_벨트
- 2. 패션잡화_선글라스_안경테
- 2. 패션잡화_스글
- 2. 패션잡화_시계
- 2. 패션잡화_양말
- 2. 패션잡화_여성가방
- 2. 패션잡화_여성신발
- 2. 패션잡화_여행용가방_소품
- 2. 패션잡화_장갑
- 2. 패션잡화_주얼리
- 2. 패션잡화_지갑
- 2. 패션잡화_패션소품
- 2. 패션잡화_헤어엑세서리
- 3. 화장품_미용_남성화장품
- 3. 화장품_미용_네일케어
- 3. 화장품_미용_마스크_팩
- 3. 화장품_미용_바디케어
- 3. 화장품_미용_베이스메이크업
- 3. 화장품_미용_뷰티소품
- 3. 화장품_미용_색조메이크업
- 3. 화장품_미용_선크어
- 3. 화장품_미용_스킨케어
- 3. 화장품_미용_클렌징
- 3. 화장품_미용_향수
- 3. 화장품_미용_헤어스타일링
- 3. 화장품_미용_헤어케어
- 4. 디지털_가전
- 5. 가구_인테리어
- 6. 출산_육아
- 7. 식품
- 8. 스포츠_레저
- 9. 생활_건강
- 10. 여가_색화류의

- DVD_50000059_2017-08-01_to_2020-10-31.csv
- 건강관리용품_50000064_2017-08-01_to_2020-10-31.csv
- 건강측정용품_50000063_2017-08-01_to_2020-10-31.csv
- 공구_50000165_2017-08-01_to_2020-10-31.csv
- 관상어용품_50000057_2017-08-01_to_2020-10-31.csv
- 구강위생용품_50000072_2017-08-01_to_2020-10-31.csv
- 냉온_찜질용품_50000071_2017-08-01_to_2020-10-31.csv
- 눈건강용품_50000073_2017-08-01_to_2020-10-31.csv
- 당뇨관리용품_50000065_2017-08-01_to_2020-10-31.csv
- 문구_사무용품_50000158_2017-08-01_to_2020-10-31.csv
- 물리치료_저주파용품_50000069_2017-08-01_to_2020-10-31.csv
- 반려동물_50000155_2017-08-01_to_2020-10-31.csv
- 발건강용품_50000074_2017-08-01_to_2020-10-31.csv
- 블루레이_50000081_2017-08-01_to_2020-10-31.csv
- 생활용품_50000078_2017-08-01_to_2020-10-31.csv
- 세탁용품_50000062_2017-08-01_to_2020-10-31.csv
- 수납_정리용품_50000076_2017-08-01_to_2020-10-31.csv
- 수집품_50000056_2017-08-01_to_2020-10-31.csv
- 실버용품_50000067_2017-08-01_to_2020-10-31.csv
- 악기_50000156_2017-08-01_to_2020-10-31.csv
- 안마용품_50000075_2017-08-01_to_2020-10-31.csv
- 육실용품_50000157_2017-08-01_to_2020-10-31.csv
- 원예_식물_50000079_2017-08-01_to_2020-10-31.csv
- 음반_50000058_2017-08-01_to_2020-10-31.csv
- 의료용품_50000066_2017-08-01_to_2020-10-31.csv
- 자동차용품_50000055_2017-08-01_to_2020-10-31.csv
- 재활운동용품_50000068_2017-08-01_to_2020-10-31.csv
- 정원_원예용품_50000080_2017-08-01_to_2020-10-31.csv
- 종교_50000060_2017-08-01_to_2020-10-31.csv
- 좌욕_좌훈용품_50000070_2017-08-01_to_2020-10-31.csv
- 주방용품_50000061_2017-08-01_to_2020-10-31.csv
- 청소용품_50000077_2017-08-01_to_2020-10-31.csv
- 화방용품_50000054_2017-08-01_to_2020-10-31.csv

Date	Ratio
2017-08-01	0.55217
2017-08-02	0.55963
2017-08-03	0.54305
2017-08-04	0.49157
2017-08-05	0.41233
2017-08-06	0.51597
2017-08-07	0.61159
2017-08-08	0.61998
2017-08-09	0.63022
2017-08-10	0.61852
2017-08-11	0.54462
2017-08-12	0.44706
2017-08-13	0.49371
2017-08-14	0.59758
2017-08-15	0.57727
2017-08-16	0.70128
2017-08-17	0.65296
2017-08-18	0.59782
2017-08-19	0.49653
2017-08-20	0.5726
2017-08-21	0.70279
2017-08-22	0.71894
2017-08-23	0.69147
2017-08-24	0.68534
2017-08-25	0.60734
2017-08-26	0.51904
2017-08-27	0.57956
2017-08-28	0.66272
2017-08-29	0.87417
2017-08-30	0.77248

총 601,128 개의 클릭량 데이터 수집 완료

3. 중간발표 이후 진행 상황

기말발표까지의 프로젝트 진행 간트차트(예정)

	11월 둘째 주	11월 셋째 주	11월 넷째 주	12월 첫째 주	12월 둘째 주	12월 셋째 주
코로나 확진자 데이터 수집						
시각화 SW 학습						
시각화 SW 플랫폼 조사						
사용 플랫폼 결정						
해당 플랫폼 학습						
데이터 시각화 진행						
시각화 결과 분석						
기말발표 자료 준비 및 발표						

1) 국내 코로나 확진자 데이터 수집

2) 시각화 소프트웨어 패키지에 대해 학습

- 현재까지는 matplotlib을 통해 간단한 시각화만 해본 상태.
- csv 파일을 이용해 그래프를 그리는 Gophi 사용법에 대해 알아보는 중.

3) 사용 플랫폼 결정 후, 해당 시각화 소프트웨어 패키지를 이용한 시각화

4) 시각화 결과 분석

1-1) 국내 일별 코로나 확진자 데이터 수집

- <https://blog.itpaper.co.kr> 에서 제공하는 Open API 사용

-> 2020-02-17부터의 국내 일별 코로나 확진자 데이터를 수집.

- 통계청에서 제공하는 일별 확진자 데이터 csv 파일 사용

-> 2020-04-01부터의 공식 확진자 데이터를 제공.

-> 2020-02-17 ~ 2020-03-31의 데이터는 Open API를,
2020-04-01 ~ 2020-12-11의 데이터는 통계청 csv 파일을 이용.

1-2) 쇼핑 클릭량 데이터 추가 수집

- 11월 중순부터 일별 확진자 수 폭증.

- > 2020-11-01 ~ 2020-12-11까지의 클릭량 데이터 추가 수집

2) 18년도, 19년도, 20년도의 각 카테고리별 클릭량 분석

- 18년도, 19년도, 20년도의 동일 카테고리에 대한
연간 평균 클릭량, Pearson Correlation 분석.
 - > 각 해의 데이터는 2월 1일부터 12월 12일 까지로 범위를 정해 분석에 사용.
 - > 18년도, 19년도 대비 20년도의 클릭량 평균이 몇 % 증가, 혹은 감소했는지 분석.
 - > Pearson Correlation을 통해 다른 해와 20년의 클릭량 차이가 얼마나 다른지 분석.
- 20년도의 클릭량 데이터, 코로나 일별 확진자 수에 대한
Pearson Correlation 분석
- Graph는 참고용으로 사용! (계획 변동)

```

data = pd.read_csv(file_name)
data = data['Ratio']

data1 = data.loc[31:344]
data_2018 = []
sum_2018 = 0

for ratio in data1:
    data_2018.append(ratio)
    sum_2018 += ratio

data2 = data.loc[396:709]
data_2019 = []
sum_2019 = 0

for ratio in data2:
    data_2019.append(ratio)
    sum_2019 += ratio

data3 = data.loc[761:]
data_2020 = []
sum_2020 = 0

for ratio in data3:
    data_2020.append(ratio)
    sum_2020 += ratio

all_data = []
all_data.append(data_2018)
all_data.append(data_2019)
all_data.append(data_2020)

df = pd.DataFrame(all_data).T
df.columns = ['2018', '2019', '2020']

skp = [(sum_2018 / len(data_2018), sum_2019 / len(data_2019), sum_2020 / len(data_2020))]
avg_data = pd.DataFrame(skp, columns=df.columns, index=['Average'])

corr = df.corr(method='pearson')
corr = corr.append(avg_data)

corr_file_name = title + '_' + category + '_corr.csv'
corr.to_csv(corr_file_name, encoding='utf-8-sig')

print(corr)

```

```

import matplotlib.pyplot as plt
import numpy as np

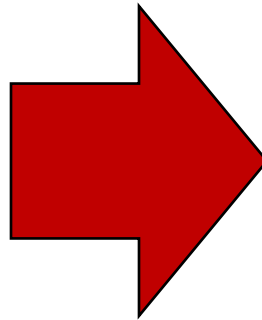
a = np.arange(0, 2, 0.2)

plt.rc('font', family = 'Malgun Gothic')
plt.rcParams['axes.unicode_minus'] = False
plt.style.use("ggplot")

plt.figure(figsize=(20,10))
plt.plot(shopping_df)
plt.title(name, fontsize=20)
plt.xticks([0, 365, 730, len(shopping_df)], labels=['2018-01-01', '2019-01-01', '2020-01-01', '2020-12-11'], fontsize=20)

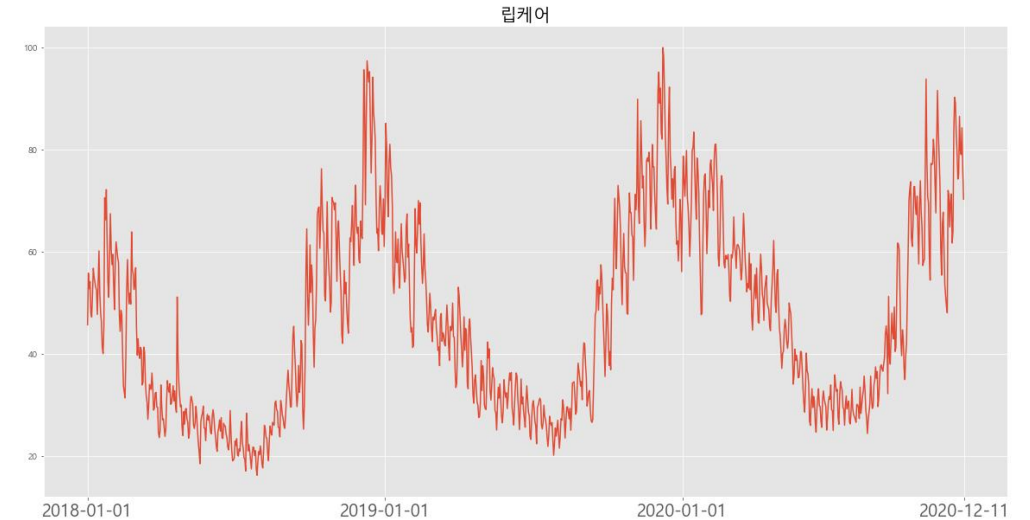
fig = plt.gcf()
graph_file_name = title + '_' + category + '_graph.png'
fig.savefig(graph_file_name, dpi=fig.dpi)

```



	2018	2019	2020
2018	1	0.418368	-0.00057
2019	0.418368	1	-0.20751
2020	-0.00057	-0.20751	1
Average	56.59876	60.76241	52.87935

연도별 Pearson Correlation 분석 & 클릭량 평균

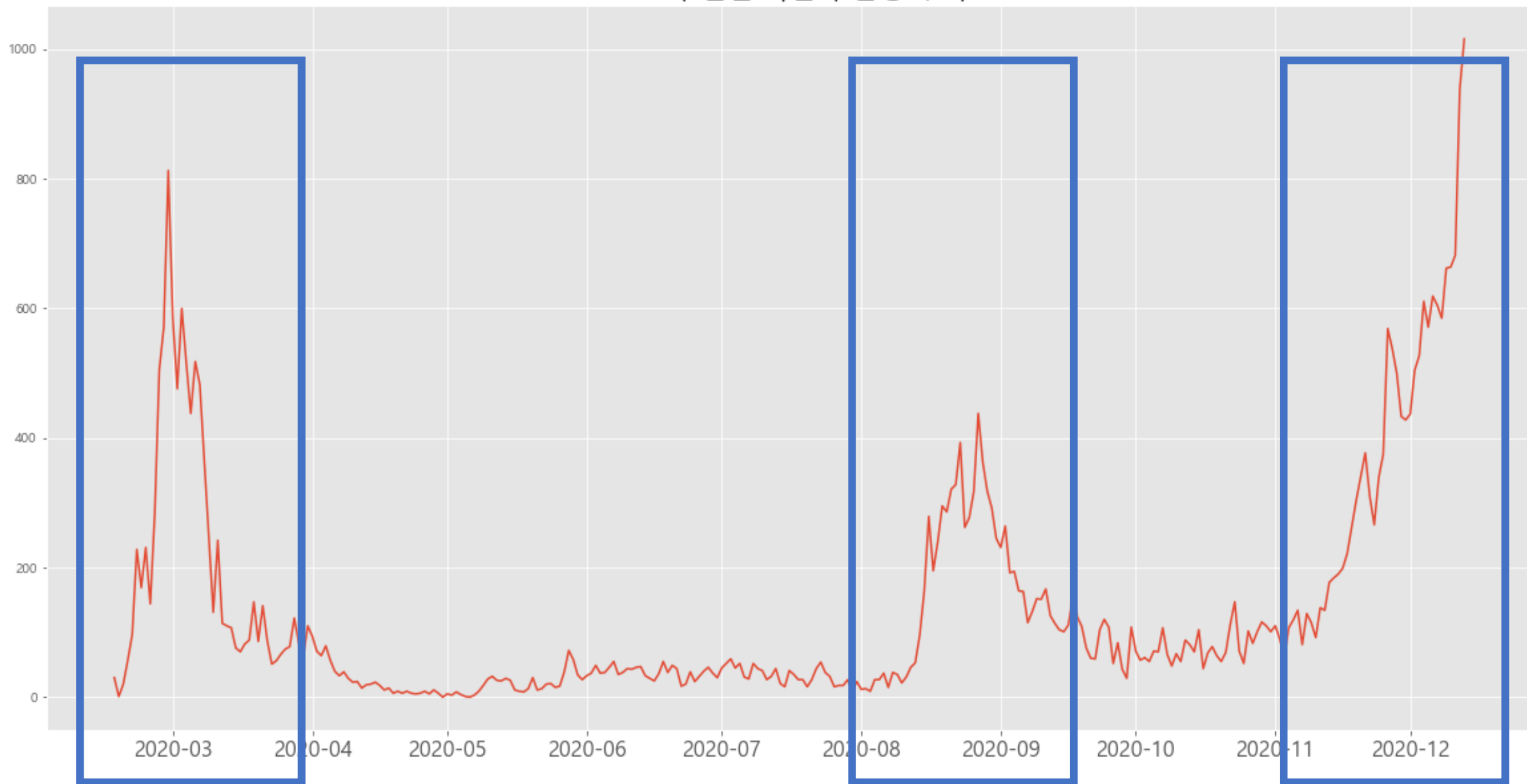


클릭량 Ratio 그래프

※ 분석 시 주의한 점 ※

- 18년도, 19년도, 20년도에 걸쳐 꾸준한 증감세를 보이면
트렌드나 유행에 의한 것이거나 다른 요소의 개입일 것이라 생각.
- 반짝 유행으로 인한 클릭량 평균의 변동인지를 살펴봤음.
ex) 게임 카테고리의 경우, 동물의 숲, 사이버펑크 2077과 같은 outlier 고려.
- 계절 유행 상품의 경우(18, 19, 20년도의 상관계수가 높은 경우),
코로나 일별 확진자와의 상관계수 분석에서 유의미하게 나와도 의심.

코로나 일별 확진자 발생 추이



4. 최종 결과 분석

1번 가설 검증

1. 코로나 사태 이후, 화장품/미용 상품 중, 얼굴 위쪽을 미용할 수 있도록 해주는 상품의 검색 및 클릭량이 얼굴 아래쪽을 미용하는 상품의 검색 및 클릭량보다 상대적으로 높아졌을 것이다.

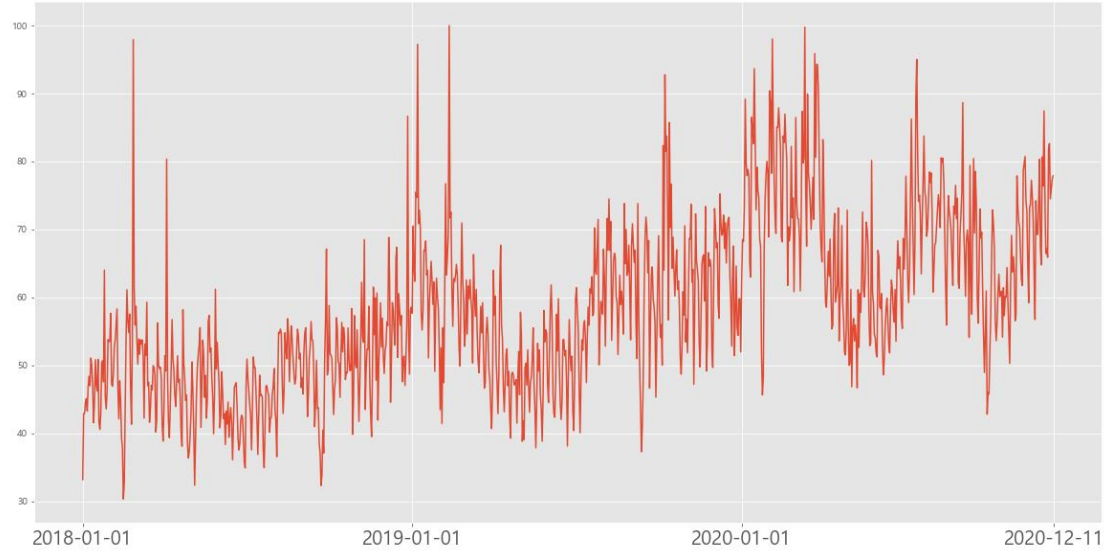
- 분석에 사용한 카테고리 :

-> 얼굴 위쪽 : 마스크라, 아이라이너, 아이브로, 아이섀도

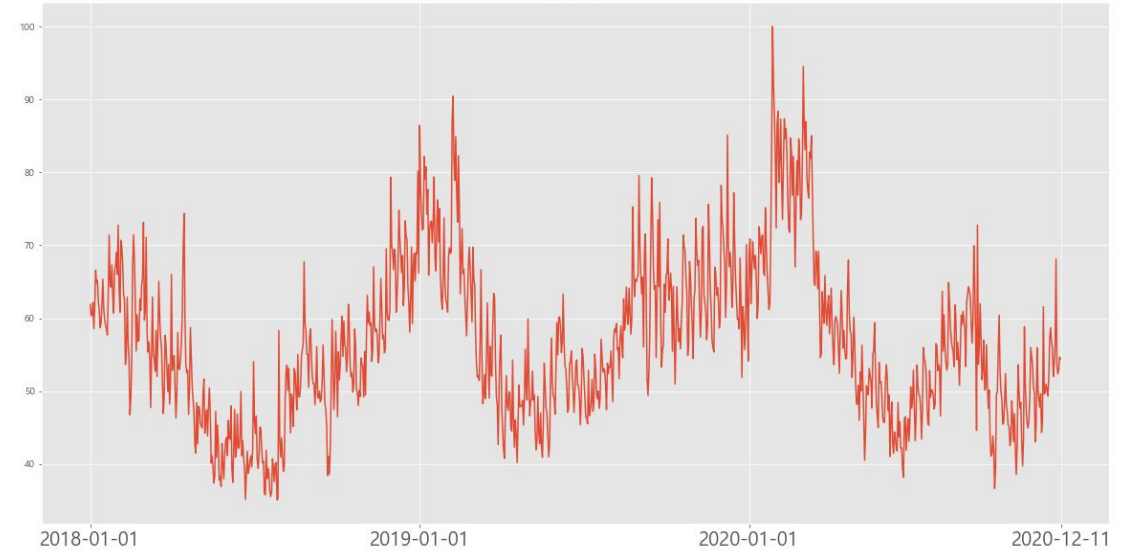
-> 얼굴 아래쪽 : 립스틱, 립라이너, 립케어, 립틴트

1) 얼굴 위쪽 미용 제품 결과 분석

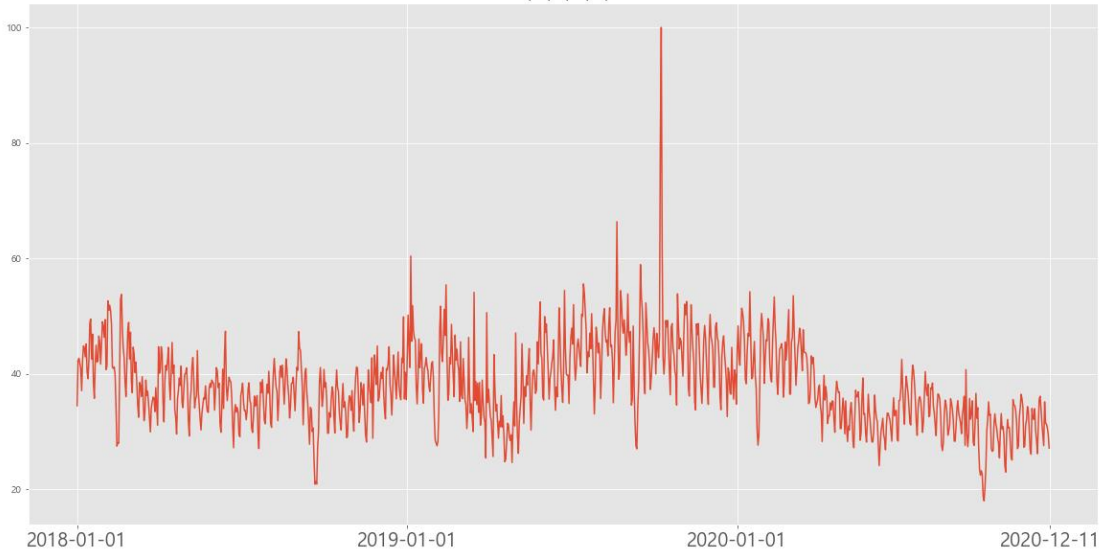
마스카라



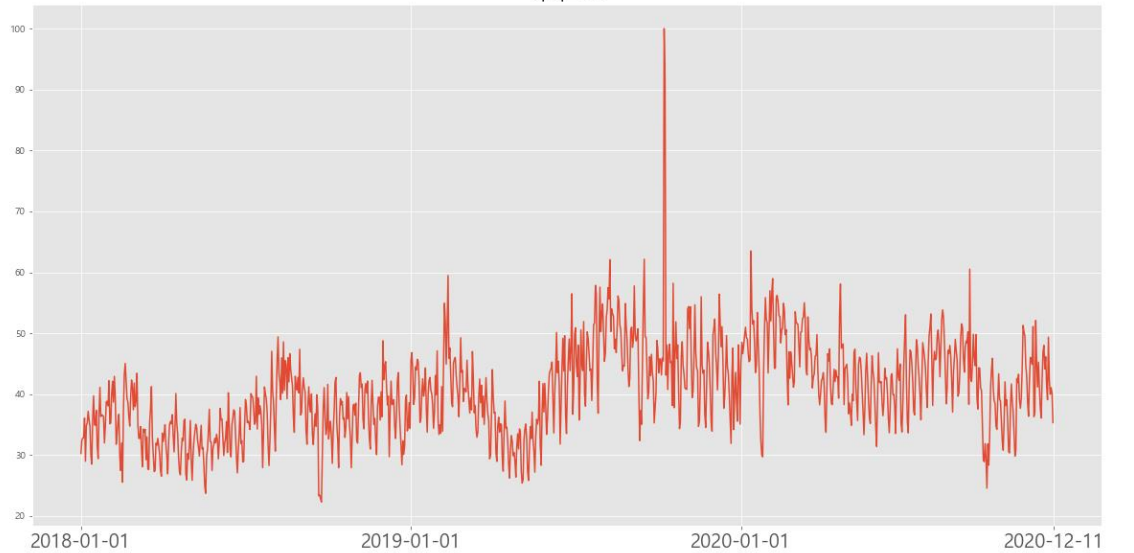
아이섀도



아이라이너

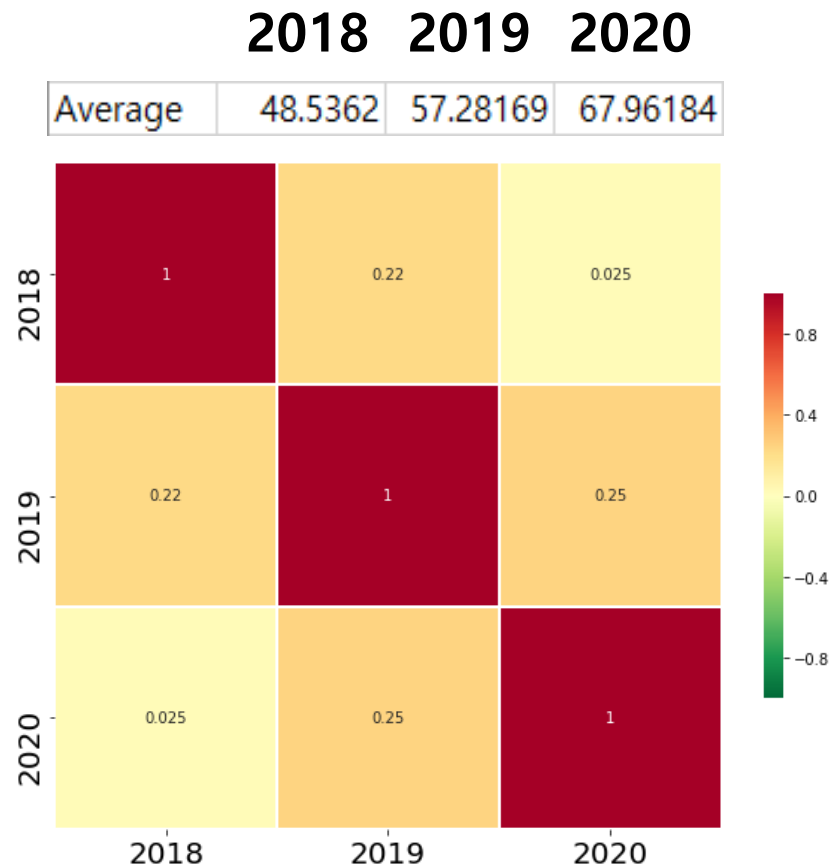


아이브로



1) 얼굴 위쪽 미용 제품 결과 분석

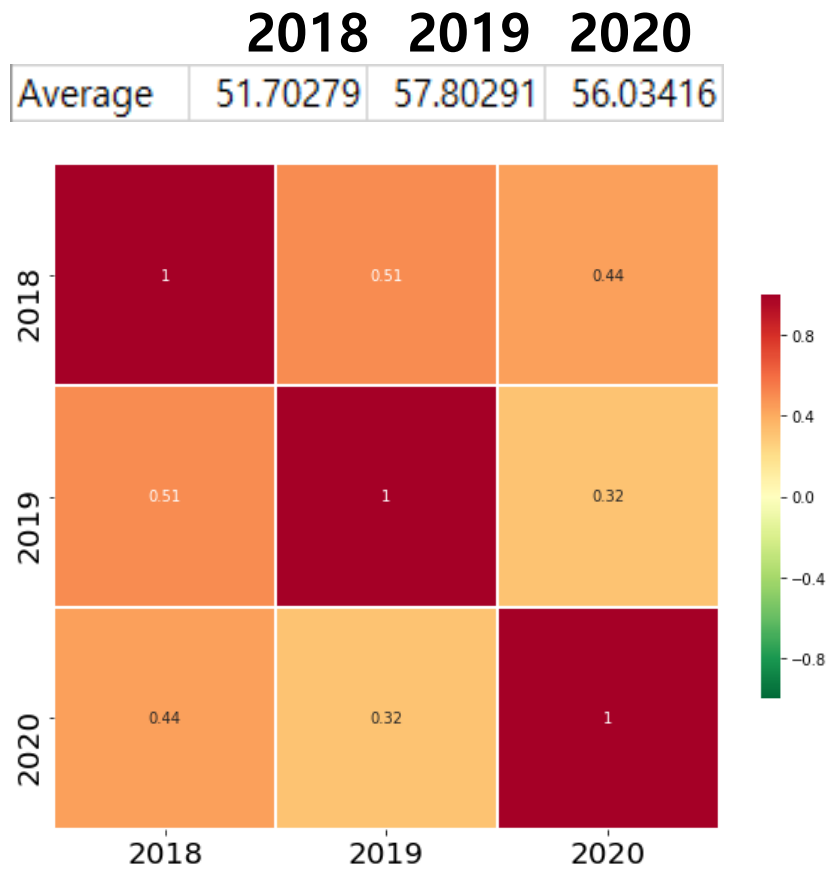
마스카라 카테고리의 연도별 Pearson Correlation과 평균 클릭량



- 꾸준한 클릭량의 상승(48% -> 57% -> 67%)
- 18년도와 20년도의 상관관계가 거의 없음.
- 하지만 18년도와 19년도의 상관관계도 낮은 편.

1) 얼굴 위쪽 미용 제품 결과 분석

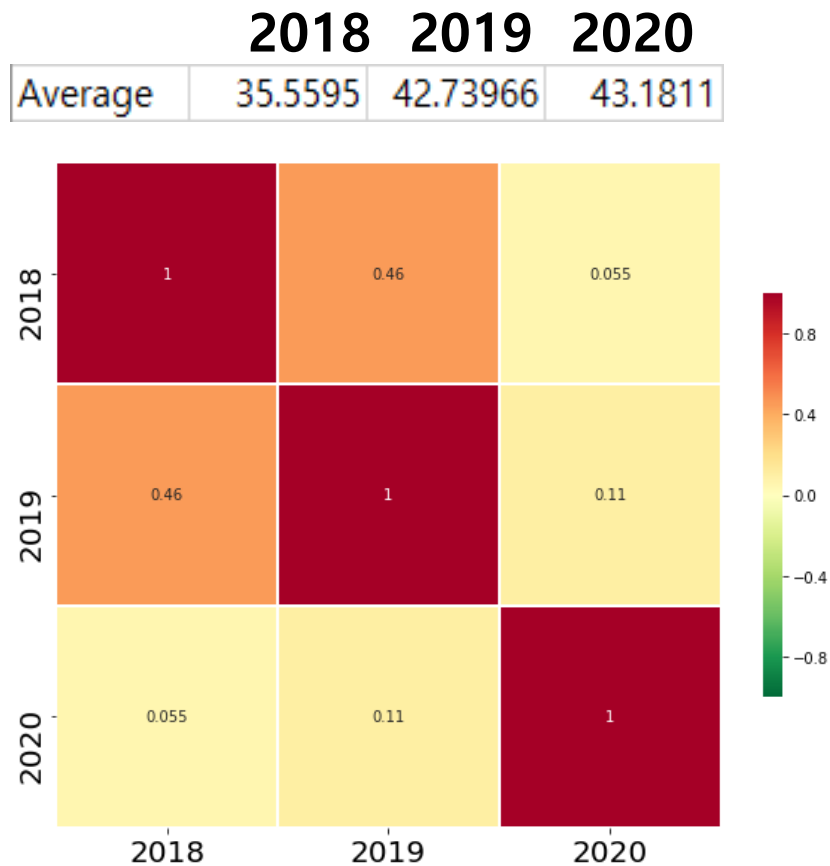
아이새도 카테고리의 연도별 Pearson Correlation과 평균 클릭량



- 클릭량 거의 변화 없음(51% -> 57% -> 56%)
- 18, 19, 20년도 모두 상관관계 뚜렷한 편.

1) 얼굴 위쪽 미용 제품 결과 분석

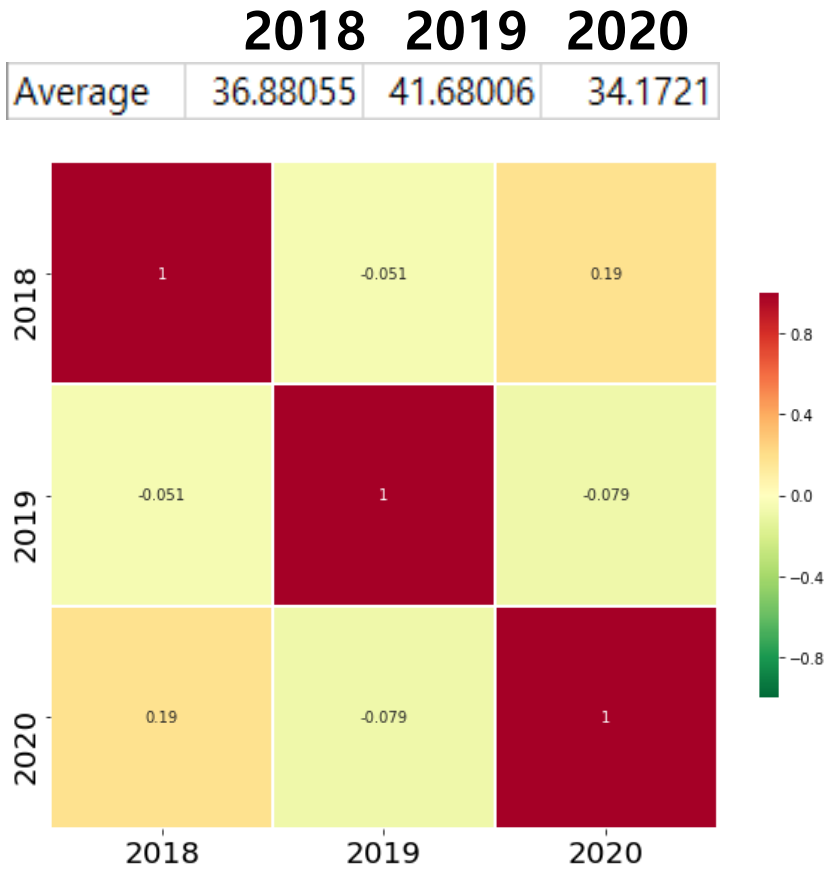
아이브로 카테고리의 연도별 Pearson Correlation과 평균 클릭량



- 18, 19년도에 비해 20년도는 상관관계가 약한 편.
- 그러나 클릭량 큰 변화 없음(35% -> 42% -> 43%)

1) 얼굴 위쪽 미용 제품 결과 분석

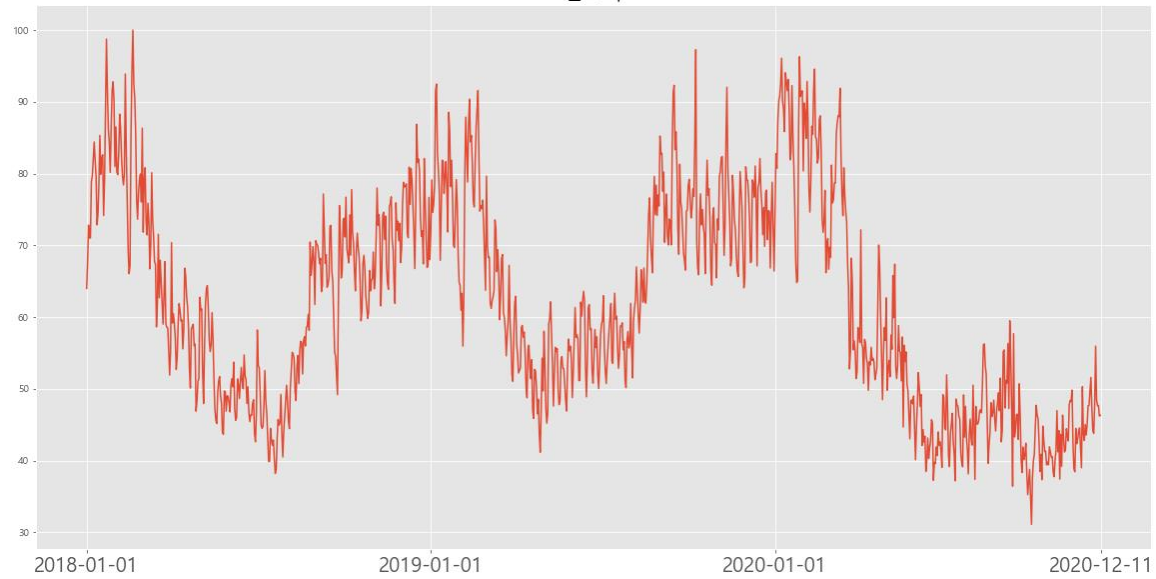
아이라이너 카테고리의 연도별 Pearson Correlation과 평균 클릭량



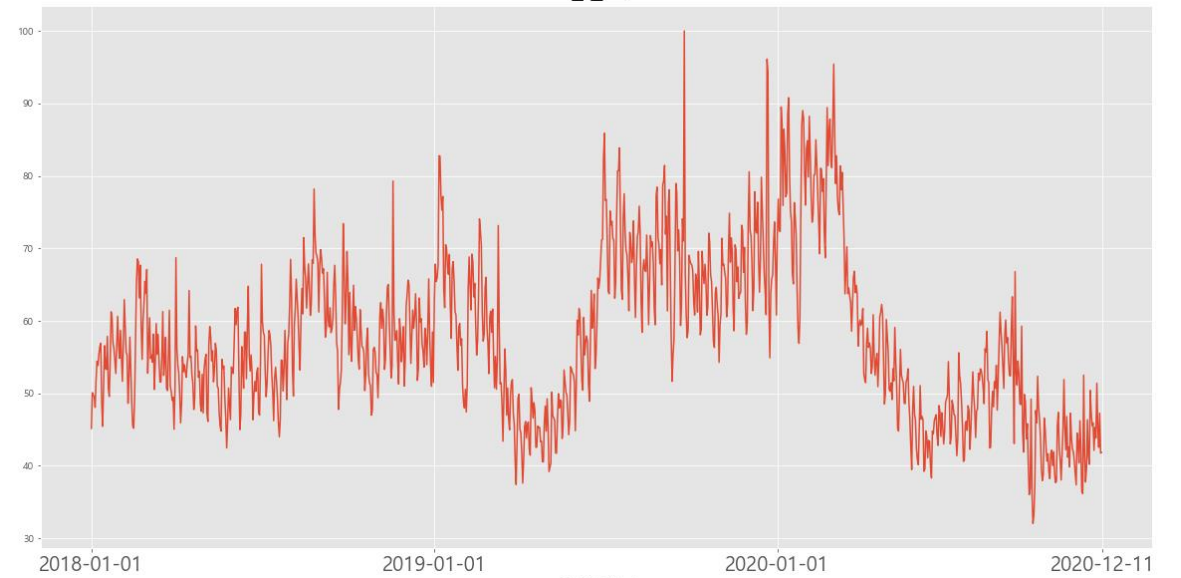
- 클릭량 오히려 감소(36% -> 41% -> 34%)
- 18, 19, 20년도 모두 상관관계 거의 없는 편.

2) 얼굴 아래쪽 미용 제품 결과 분석

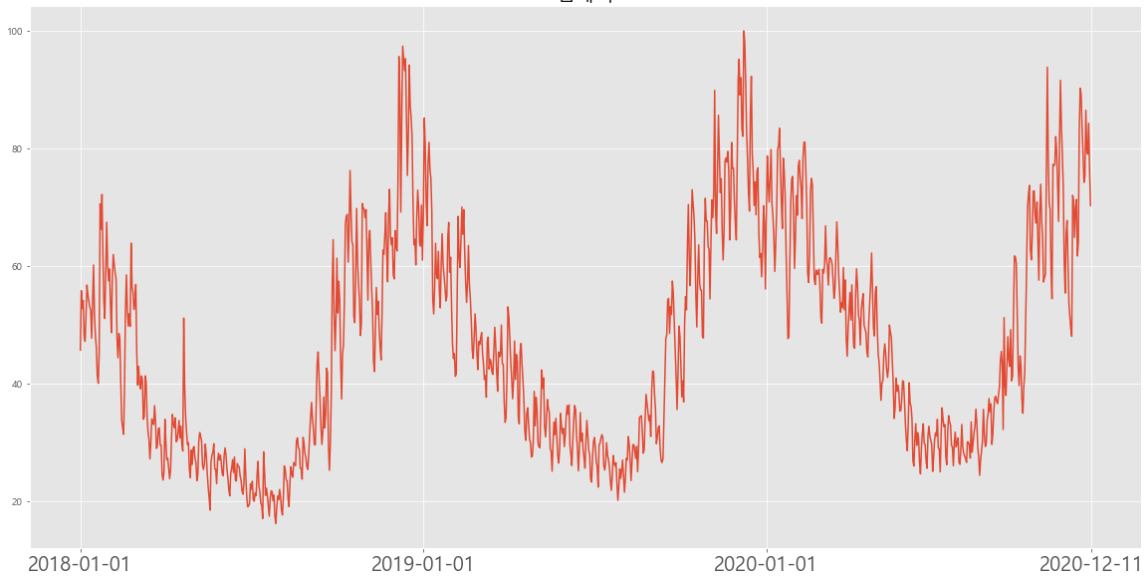
립스틱



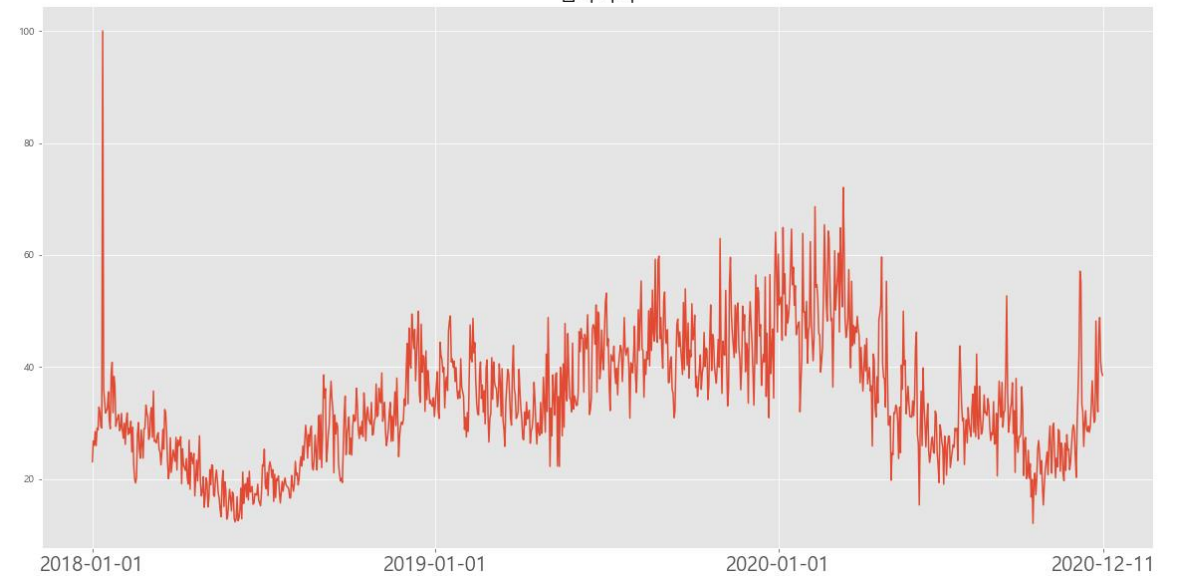
립틴트



립케어

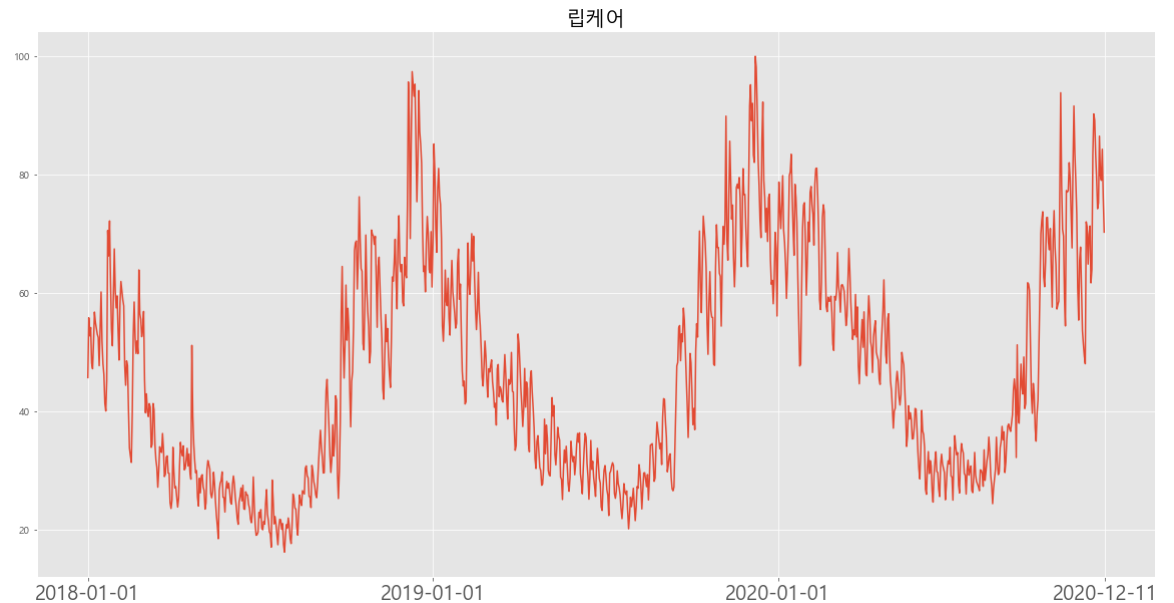
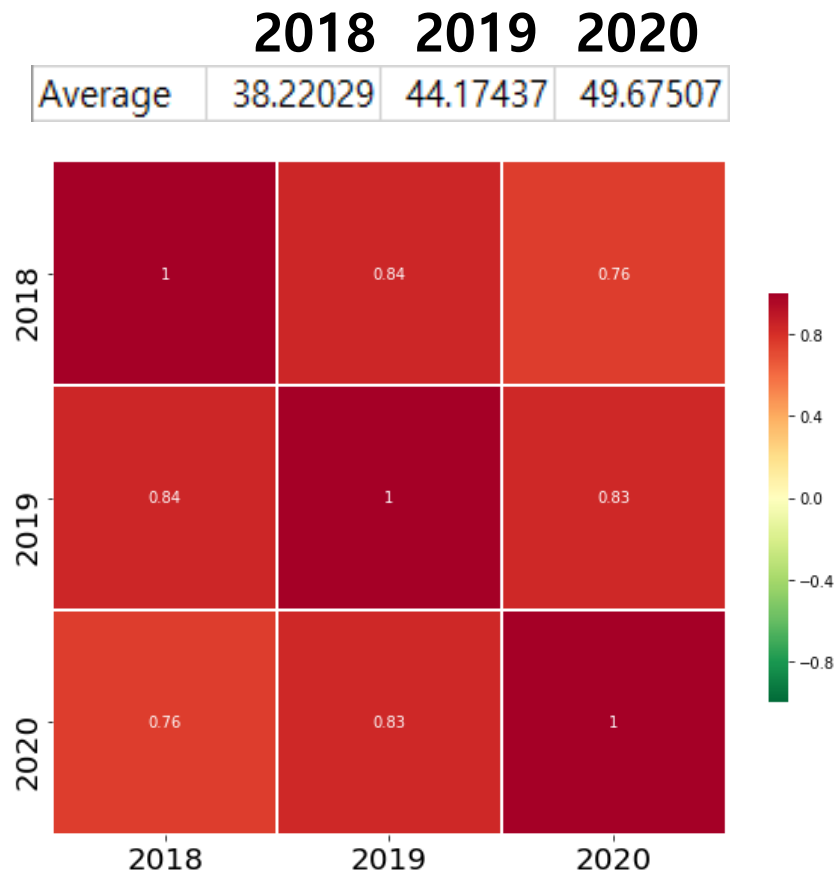


립라이너



2) 얼굴 아래쪽 미용 제품 결과 분석

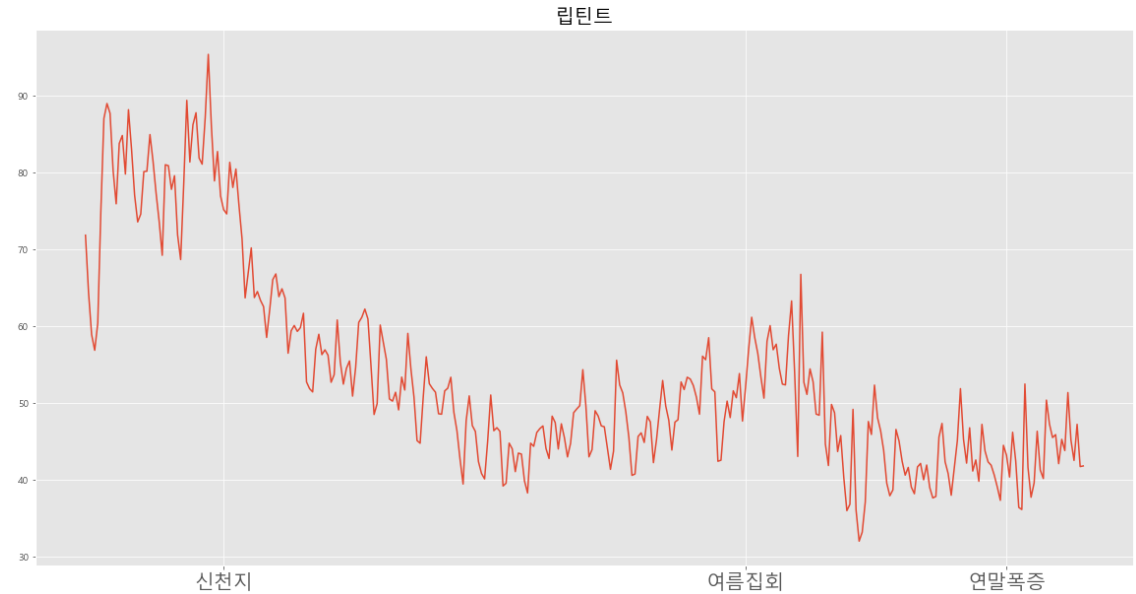
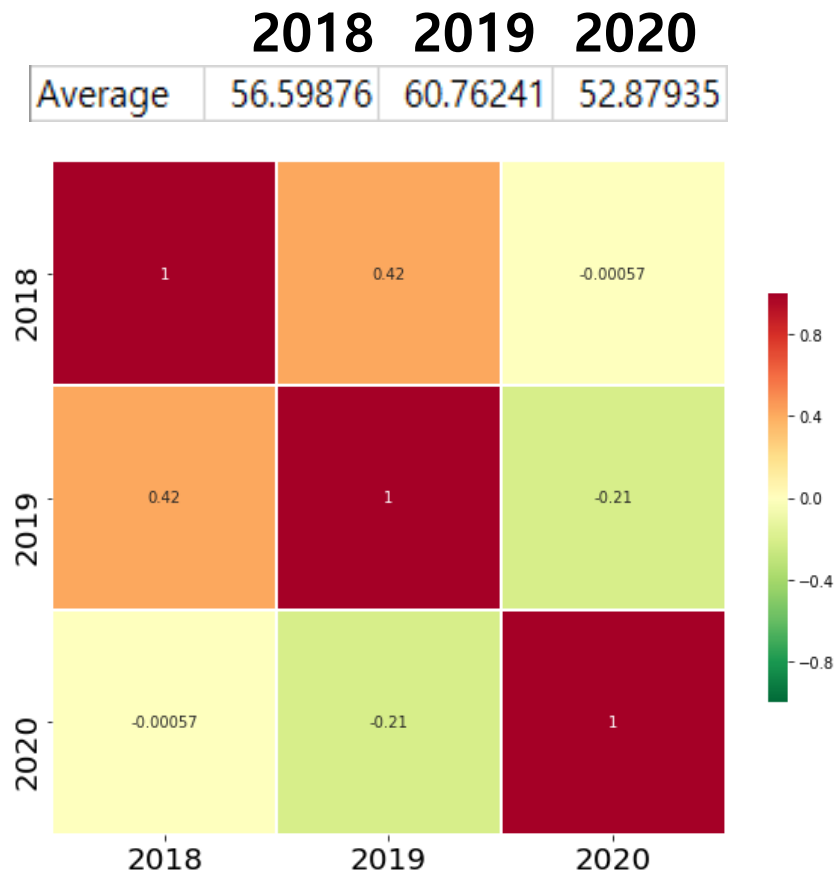
립케어 카테고리의 연도별 Pearson Correlation과 평균 클릭량



- 클릭량 꾸준히 증가(38% -> 44% -> 49%)
- 강한 상관관계(계절 영향), 매년 비슷한 추이

2) 얼굴 아래쪽 미용 제품 결과 분석

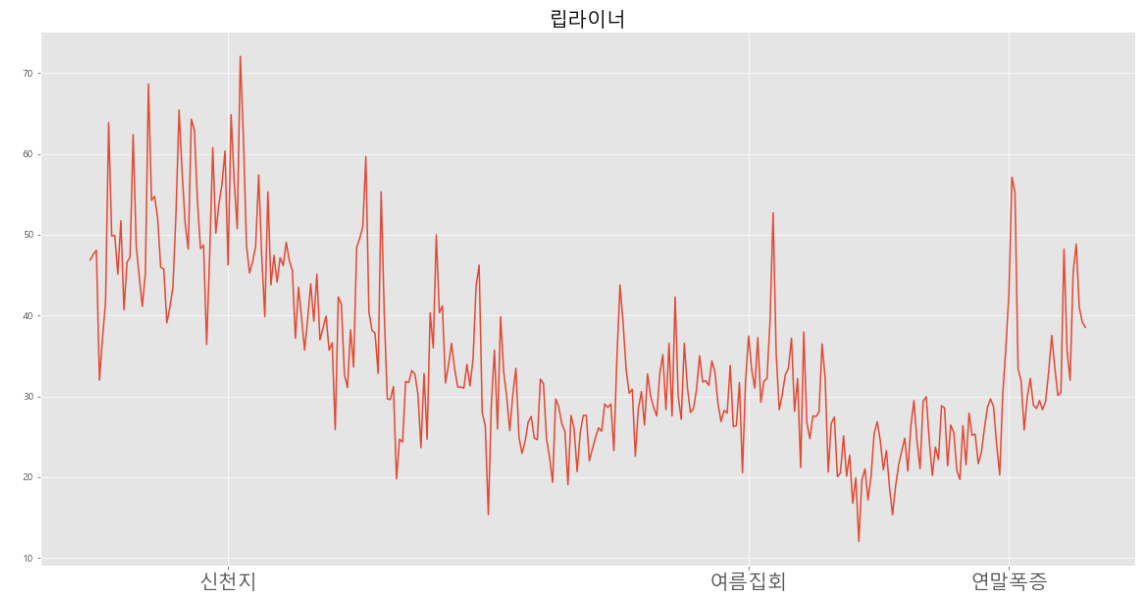
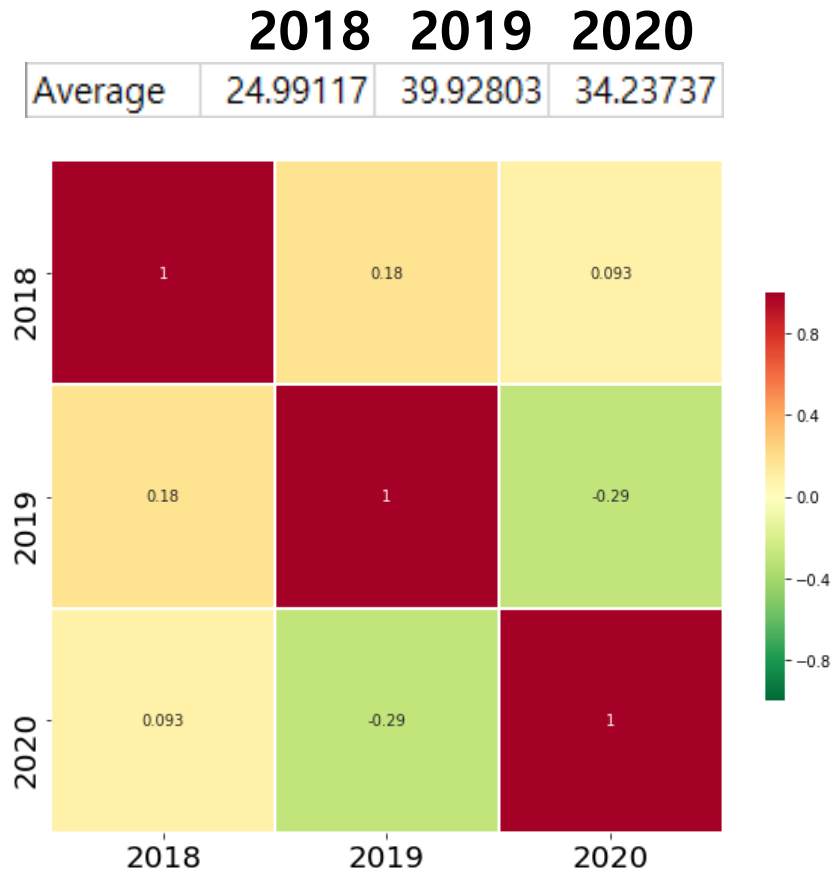
립틴트 카테고리의 연도별 Pearson Correlation과 평균 클릭량



- 클릭량 증가하다 감소(56% -> 60% -> 52%)
- 18, 19의 상관관계에 비해 약한 상관관계
- 그래프 상으로 코로나 사건 이후로 감소하는 추이

2) 얼굴 아래쪽 미용 제품 결과 분석

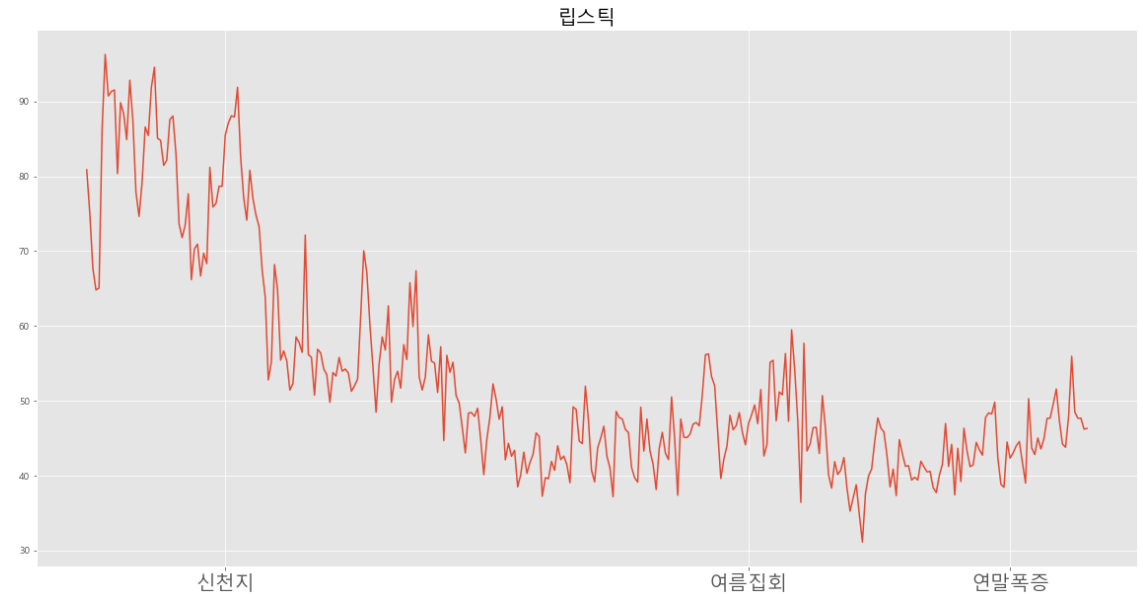
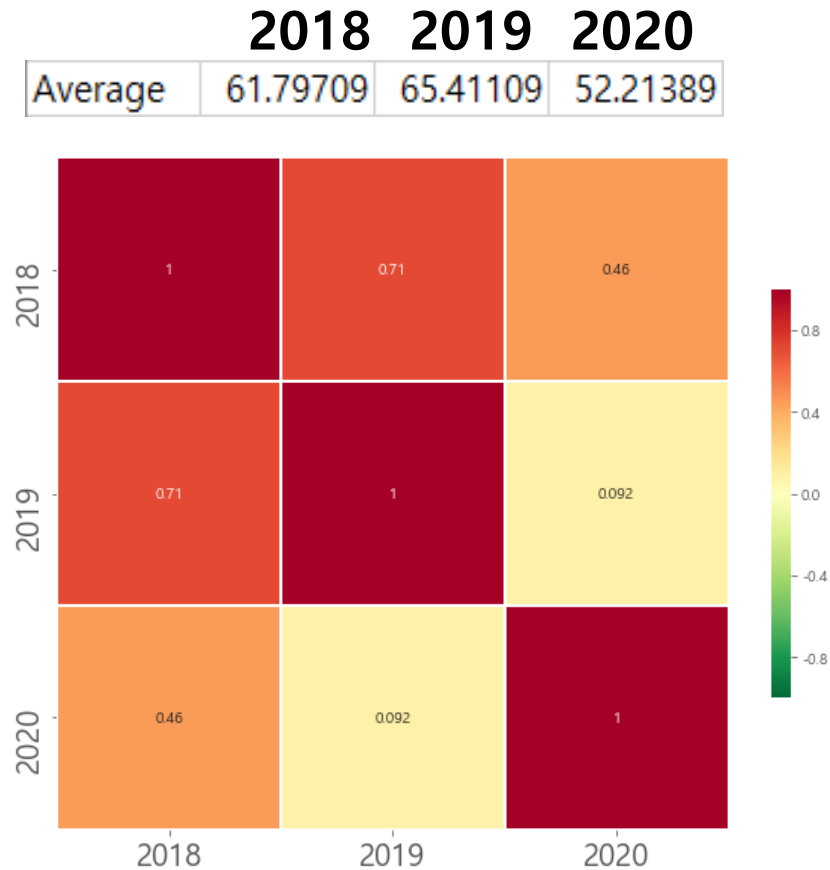
립라이너 카테고리의 연도별 Pearson Correlation과 평균 클릭량



- 클릭량 증가하다 감소(56% -> 60% -> 52%)
- 그래프 상으로 코로나 사건 이후로 감소하는 추이

2) 얼굴 아래쪽 미용 제품 결과 분석

립스틱 카테고리의 연도별 Pearson Correlation과 평균 클릭량



- 클릭량 증가하다 감소(61% -> 65% -> 52%)
- 13% 감소, 19년도와 상관관계 거의 없음.
- 그래프 상으로 코로나 사건 이후로 감소하는 추이

1번 가설 검증

결론 1(얼굴 위쪽 미용 제품)

- 마스크라는 클릭량이 상승했으나, 꾸준한 상승이므로 코로나 영향이라 보기 힘들.
- 다른 위쪽 미용 물품들은 클릭량 변화가 거의 없거나 오히려 감소.
- 눈에 띄는 상관관계 변화도 없음.

코로나 사태 이후, 얼굴 위쪽을 미용할 수 있도록 해주는 상품의
검색 및 클릭량이 증가했을 것이라는 가정은 들어맞지 않았음.

1번 가설 검증

결론 2(얼굴 아래쪽 미용 제품)

- 립케어는 사람들이 여전히 많이 찾는 겨울에 인기있는 상품.
- 다른 아래쪽 미용 물품들은 클릭량이 감소. 코로나 사태 발발과 어느 정도 맞는 추이.
- 립틴트, 립라이너는 감소폭이 비교적 적은 편(8%, 5%)이나, 립스틱은 13% 감소

얼굴 아래쪽을 미용할 수 있도록 해주는 상품의 검색 및 클릭량 감소에
코로나 사태가 영향을 어느 정도 줬을 가능성이 있다고 판단이 가능.

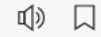
국제 >

불황때 뛰는 '립스틱 지수'... 이번엔 마스크에 갇혔네

여성 90% "마스크 쓰면 안 발라"

강다운 기자

입력 2020.07.30 03:00



경기가 불황에 빠졌는지를 판단하는 풍향계로 활용되던 '립스틱 지수'가 코로나발(發) 경기 침체 국면에선 들어맞지 않고 있다.

립스틱 지수는 경기가 어려울수록 립스틱 판매량이 증가하는 현상을 지수로 표현한 것이다. 1930년대 대공황 시기에 소비가 침체한 와중에도 립스틱 판매량은 오히려 늘어난 데서 고안됐다. 여성들이 경제적으로 어려울 때 드레스·신발 같은 값비싼 제품 대신 상대적으로 저렴한 립스틱을 구매하며 '작은 사치'를 하기 때문이다. 2001년 9·11테러 때도 립스틱 판매가 증가하자 미국 유명 화장품

회사 에스티로더가 이런 현상을 지수화해 '립스틱 지수'를 발표했다. 이후 립스틱 판매량은 경기 상황을 판단하는 지표가 됐다.

립스틱 지수가 코로나 불황에는 들어맞지 않았다는 사실을 수치적으로 증명해보았다!

2번 가설 검증

2. Pandemic 상황이 지속되면서, 인터넷 쇼핑에서 클릭량 및 검색량이 증가한 상품들이 있을 것이고, 그에 반해 하락한 상품들이 있을 것이다.

- 약 **430여 개** 카테고리의 클릭량을 분석

- 1) 건강관리용품 카테고리의 변화
- 2) 패션의류/잡화 + 베이스메이크업 카테고리의 변화
- 3) 재택 근무, 원격 수업 등으로 인한 변화
- 4) 주방 물품들에 대한 변화
- 5) 취미와 관련된 상품들 - 3번 가설과 함께 설명

1) 건강관리용품 카테고리

마스크

전년도 대비 클릭량 **24배**,
9.2% 증가 (0.4% -> 9.6%)



체온계

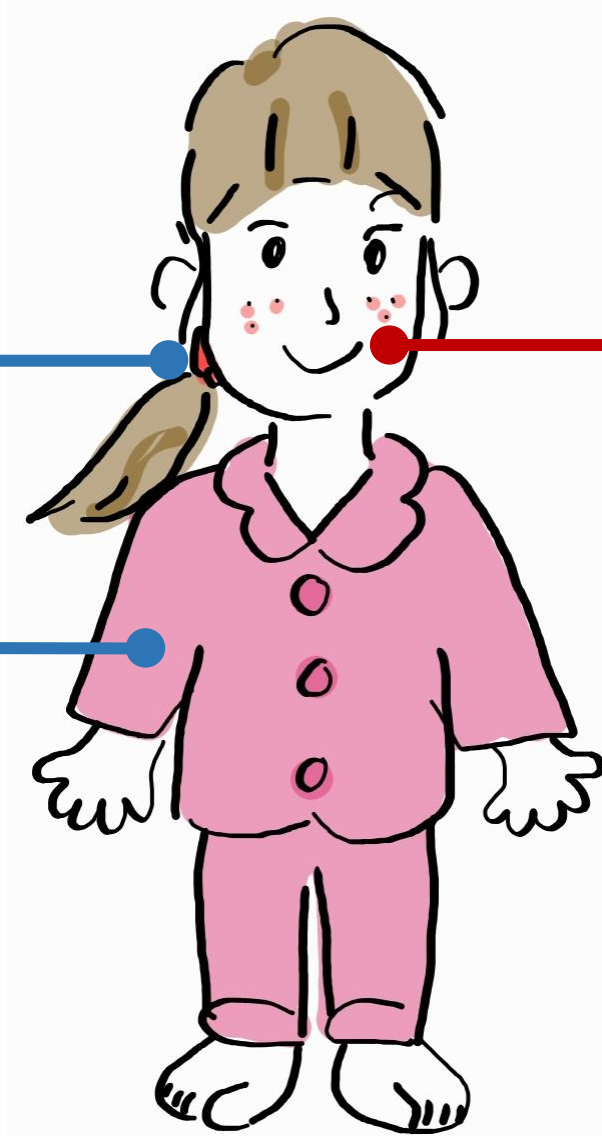
전년도 대비 클릭량 **9.6배**,
9.2% 증가 (1.2% -> 11.5%)



2) 패션의류/잡화 + 베이스메이크업 카테고리

헤어끈

전년도 대비 클릭량 2.2배,
27% 증가 (22% -> 49%)



여성잠옷/홈웨어

전년도 대비 클릭량 1.61배,
19% 증가 (31% -> 50%)

베이스메이크업
전년도 대비 클릭량 0.7배,
23% 감소 (80% -> 57%)

3) 재택 근무, 원격 수업 등으로 인한 변화

모니터 주변기기

전년도 대비 클릭량 3.39배,
33% 증가 (13% -> 46%)



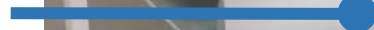
태블릿PC액세서리

전년도 대비 클릭량 2.2배,
33% 증가 (29% -> 62%)



의자

전년도 대비 클릭량 1.5배,
19% 증가 (37% -> 56%)



4) 주방 물품들에 대한 변화

식기

전년도 대비 클릭량 1.9배,
26% 증가 (13% -> 46%)



그릇장/컵보드

전년도 대비 클릭량 1.9배,
28% 증가 (32% -> 60%)

카테고리명	연도별 클릭량 평균 변화	비고
패션의류, 패션잡화 카테고리		
잠옷/홈웨어	31% -> 31% -> 50%(+19%)	18, 19년도와 강한 상관관계
신발끈	33% -> 44% -> 63%(+19%)	19년도와 약한 상관관계
기타신발용품	14% -> 22% -> 47%(+25%)	18, 19년도와 강한 상관관계
여행용가방/소품	52% -> 61% -> 41%(+20%)	18, 19년도와 약한 양의 상관관계
헤어핀	29% -> 40% -> 56%(+16%)	전년도 대비 약 1.39배
헤어끈	15% -> 22% -> 49%(+27%)	전년도 대비 약 2.2배
야구모자	47% -> 50% -> 61%(+11%)	18년도와 상관관계 거의 없음, 19년도와 약한 양의 상관관계
주얼리	51% -> 59% -> 72%(+13%)	19년도와 상관관계 거의 없음
안경소품	14% -> 21% -> 37%(+16%)	18, 19년도와 상관관계 거의 없음.
안경줄(마스크줄)	9% -> 17% -> 36%(+19%)	전년도 대비 클릭량 약 2.1배
화장품/미용 카테고리		
스킨케어	51% -> 64% -> 74%(+10%)	꾸준한 클릭량 증가
클렌징	49% -> 61% -> 69%(+8%)	꾸준한 클릭량 증가
남성화장품	49% -> 57% -> 63%(+6%)	꾸준한 클릭량 증가
베이스메이크업	69% -> 80% -> 57%(-23%)	18, 19년도와 상관관계 거의 없음, 전년도 대비 약 0.7배
BB크림	52% -> 57% -> 38%(-19%)	18년도와 상관관계 거의 없음, 전년도 대비 약 0.6배
CC크림	53% -> 51% -> 25%(-26%)	18년도와 약한 음의 상관관계, 전년도 대비 약 0.49배
프라이머	47% -> 53% -> 37%(-16%)	18년도와 상관관계 거의 없음, 전년도 대비 약 0.69배
파우더	61% -> 56% -> 34%(-22%)	전년도 대비 약 0.6배

카테고리명	연도별 클릭량 평균 변화	비고
베이스메이크업세트	18% -> 33% -> 3%(-30%)	18, 19년도와 상관관계 거의 없음, 전년도 대비 0.09배
립스틱	61% -> 65% -> 52%(-13%)	19년도와 상관관계 거의 없음, 전년도 대비 0.79배
립케어	38% -> 44% -> 49%(+5%)	꾸준한 상승폭
립글로스	36% -> 42% -> 36%(-6%)	19년도와 상관관계 거의 없음.
립틴트	56% -> 60% -> 52%(-8%)	18년도와 상관관계 거의 없음.
아이섀도	51% -> 57% -> 56%(-1%)	
아이라이너	36% -> 41% -> 34%(-7%)	18년도와 약한 상관관계, 19년도와 상관관계 거의 없음.
마스카라	48% -> 57% -> 67%(+10%)	18년도와 상관관계 거의 없음, 꾸준한 상승폭
아이브로	35% -> 42% -> 43%(+1%)	18, 19년도와 상관관계 거의 없음.
블러셔	55% -> 50% -> 46%(-4%)	꾸준한 감소폭
하이라이터	41% -> 53% -> 39%(-14%)	19년도와 약한 양의 상관관계, 전년도 대비 0.74배
색조메이크업세트	36% -> 44% -> 33%(-11%)	전년도 대비 0.75배
마스크팩	52% -> 61% -> 66%(+5%)	19년도와 상관관계 거의 없음.
네일케어	42% -> 50% -> 56%(+6%)	꾸준한 상승폭
매니큐어	36% -> 44% -> 67%(+23%)	전년도 대비 1.5배
향수	49% -> 62% -> 70%(+8%)	19년도와 상관관계 거의 없음, 꾸준한 상승폭
바디케어	27% -> 36% -> 46%(+10%)	19년도와 약한 양의 상관관계
헤어스타일링	40% -> 55% -> 67%(+12%)	18년도와 약한 양의 상관관계, 19년도와 상관관계 거의 없음
헤어스프레이	18% -> 26% -> 50%(+24%)	18년도와 약한 양의 상관관계, 전년도 대비 1.92배
파마약	25% -> 26% -> 43%(+17%)	전년도 대비 1.7배

카테고리명	연도별 클릭량 평균 변화	비고
디지털/가전 카테고리		
영상가전	36% -> 41% -> 53%(+12%)	18년도와 약한 음의 상관관계, 19년도와 상관관계 거의 없음
영상가전악세사리	27% -> 36% -> 67%(+31%)	18, 19년도와 약한 상관관계, 전년도 대비 1.86배
프로젝터주변기기	44% -> 46% -> 64%(+18%)	19년도와 약한 양의 상관관계, 전년도 대비 1.4배
생활가전	39% -> 45% -> 64%(+19%)	
청소기	35% -> 41% -> 59%(+18%)	19년도와 상관관계 거의 없음, 전년도 대비 1.4배
다리미	21% -> 26% -> 45%(+19%)	18, 19년도와 상관관계 거의 없음, 전년도 대비 1.7배
스탠드	32% -> 38% -> 64%(+26%)	18년도와 상관관계 거의 없음, 전년도 대비 1.7배
구강청정기	33% -> 51% -> 64%(+13%)	18년도와 약한 음의 상관관계, 꾸준한 상승폭
자외선소독기	0.7% -> 1.1% -> 9.5%(+8%)	18, 19년도와 상관관계 거의 없음. 전년도 대비 8.5배
손소독기	1% -> 6% -> 23%(+17%)	18, 19년도와 상관관계 거의 없음. 전년도 대비 3.8배
자동차기기	62% -> 61% -> 72%(+11%)	19년도와 상관관계 거의 없음.
PC	48% -> 41% -> 58%(+17%)	18, 19년도와 약한 양의 상관관계, 전년도 대비 1.4배
주방가전	41% -> 50% -> 69%(+19%)	19년도와 상관관계 거의 없음, 전년도 대비 1.4배
가스레인지후드	38% -> 43% -> 66%(+23%)	18년도와 약한 상관관계, 전년도 대비 1.53배
오븐	36% -> 37% -> 56%(+19%)	18년도와 상관관계 거의 없음, 전년도 대비 1.48배
식기세척/건조기	20% -> 33% -> 68%(+35%)	18, 19년도와 상관관계 거의 없음, 전년도 대비 2.1배
커피머신	28% -> 37% -> 62%(+25%)	전년도 대비 1.65배
커피메이커	26% -> 35% -> 57%(+22%)	전년도 대비 1.6배
제빵기	13% -> 12% -> 23%(+11%)	18, 19년도와 상관관계 거의 없음, 전년도 대비 1.85배

카테고리명	연도별 클릭량 평균 변화	비고
거품/반죽기	18% -> 22% -> 44%(+22%)	18년도와 상관관계 거의 없음, 전년도 대비 2배
전기그릴	14% -> 15% -> 32%(+17%)	18, 19년도와 약한 상관관계, 전년도 대비 2.1배
믹서기	37% -> 42% -> 59%(+17%)	19년도와 약한 상관관계, 전년도 대비 1.4배
분쇄기	11% -> 21% -> 49%(+27%)	18, 19년도와 약한 상관관계, 전년도 대비 2.26배
모니터주변기기	10% -> 13% -> 46%(+33%)	18년도와 약한 양의 상관관계, 전년도 대비 3.39배
모니터받침대	7.7% -> 8.9% -> 45%(+36%)	19년도와 상관관계 거의 없음, 전년도 대비 5.1배
태블릿PC액세서리	16% -> 29% -> 63%(+34%)	19년도와 약한 양의 상관관계, 전년도 대비 2.2배
PC마이크	5.7% -> 6.8% -> 17%(+10%)	18, 19년도와 약한 상관관계, 전년도 대비 2.5배
가구/인테리어 카테고리		
침실가구	49% -> 53% -> 68%(+15%)	18, 19년도와 약한 상관관계
거실가구	46% -> 48% -> 62%(+14%)	18년도와 약한 음의 상관관계, 19년도와 상관관계 거의 없음
주방가구	47% -> 50% -> 67%(+17%)	18, 19년도와 약한 음의 상관관계, 전년도 대비 1.32배
왜건/카트	26% -> 37% -> 64%(+27%)	18, 19년도와 약한 음의 상관관계, 전년도 대비 1.74배
주방수납장	35% -> 40% -> 64%(+24%)	18, 19년도와 상관관계 거의 없음, 전년도 대비 1.6배
그릇장/컵보드	30% -> 32% -> 60%(+28%)	18, 19년도와 약한 상관관계, 전년도 대비 1.9배
수납가구	47% -> 49% -> 64%(+15%)	18, 19년도와 상관관계 거의 없음, 전년도 대비 1.3배
수납장	44% -> 39% -> 62%(+23%)	18, 19년도와 상관관계 거의 없음, 전년도 대비 1.6배
CD/DVD장	39% -> 42% -> 63%(+21%)	18, 19년도와 상관관계 거의 없음, 전년도 대비 1.5배
서재/사무용가구	36% -> 40% -> 58%(+18%)	18, 19년도와 상관관계 거의 없음, 전년도 대비 1.46배
의자	32% -> 37% -> 56%(+19%)	18, 19년도와 약한 상관관계, 전년도 대비 1.5배

카테고리명	연도별 클릭량 평균 변화	비고
사무/교구용가구	34% -> 39% -> 59%(+20%)	18, 19년도와 약한 상관관계, 전년도 대비 1.5배
홈데코	43% -> 47% -> 63%(+16%)	19년도와 상관관계 거의 없음, 전년도 대비 1.3배
주방데코	44% -> 50% -> 68%(+18%)	19년도와 상관관계 거의 없음, 전년도 대비 1.36배
수예	24% -> 29% -> 55%(+26%)	18, 19년도와 상관관계 거의 없음, 전년도 대비 1.9배
보석십자수	3.7% -> 6.5% -> 41%(+35%)	전년도 대비 6.4배
인테리어소품	41% -> 45% -> 65%(+20%)	18, 19년도와 약한 양의 상관관계
출산/육아 카테고리		
유모차	60% -> 54% -> 40%(-14%)	18, 19년도와 상관관계 강함
카시트	46% -> 47% -> 43%(-3%)	18, 19년도와 상관관계 거의 없음
구강청결용품	27% -> 30% -> 31%(+1%)	18, 19년도와 상관관계 강함, 거의 일정하게 유지
외출용품	66% -> 64% -> 48%(-18%)	18, 19년도와 상관관계 거의 없음
유아가구	65% -> 67% -> 75%(+8%)	18년도와 상관관계 거의 없음, 19년도와 약한 음의 상관관계
출산/돌기념품	46% -> 47% -> 62%(+15%)	18년도와 약한 음의 상관관계
수영복/용품	22% -> 24% -> 12%(-12%)	강한 양의 상관관계
유아발육용품	51% -> 51% -> 62%(+11%)	19년도와 약한 음의 상관관계
교재/서적	35% -> 37% -> 58%(+21%)	18, 19년도와 강한 상관관계
인형	27% -> 25% -> 36%(+11%)	18, 19년도와 강한 상관관계
위생/건강용품	1% -> 1% -> 12%(+11%)	18년도와 상관관계 거의 없음, 전년도 대비 12배
완구/매트	33% -> 33% -> 45%(+12%)	18, 19년도와 강한 상관관계

카테고리명	연도별 클릭량 평균 변화	비고
식품 카테고리		
축산	21% -> 29% -> 48%(+19%)	18년도와 상관관계 거의 없음, 19년도와 약한 양의 상관관계
수산	28% -> 34% -> 56%(+22%)	18년도와 약한 양의 상관관계
농산물	30% -> 35% -> 56%(+21%)	18년도와 상관관계 거의 없음
반찬	28% -> 33% -> 48%(+15%)	18년도와 상관관계 거의 없음, 19년도와 약한 양의 상관관계
김치	14% -> 20% -> 39%(+19%)	18, 19년도와 강한 상관관계
음료	44% -> 50% -> 69%(+19%)	18년도와 약한 음의 상관관계, 19년도와 약한 양의 상관관계
다이어트식품	41% -> 43% -> 59%(+16%)	18년도와 약한 양의 상관관계
쿠키박스	14% -> 13% -> 35%(+22%)	18년도와 상관관계 거의 없음, 19년도와 약한 상관관계
냉동/간편조리식품	27% -> 38% -> 59%(+21%)	18, 19년도와 상관관계 거의 없음.
가공식품	30% -> 36% -> 58%(+22%)	18년도와 상관관계 거의 없음, 19년도와 약한 상관관계
스포츠/레저 카테고리		
등산화	37% -> 34% -> 51%(+17%)	
캠핑	20% -> 23% -> 53%(+30%)	18, 19년도와 강한 상관관계
스키/보드	11% -> 21% -> 10%(-11%)	
헬스	31% -> 33% -> 54%(+21%)	18, 19년도와 약한 양의 상관관계
러닝머신	23% -> 33% -> 59%(+26%)	18, 19년도와 상관관계 거의 없음, 전년도 대비 1.78배
아령/덤벨	11% -> 12% -> 28%(+16%)	18, 19년도와 상관관계 거의 없음, 전년도 대비 2.34배
배드민턴	56% -> 72% -> 40%(-32%)	18, 19년도와 약한 음의 상관관계

카테고리명	연도별 클릭량 평균 변화	비고
생활/건강 카테고리		
공구	31% -> 36% -> 56%(+20%)	18년도와 약한 음의 상관관계, 19년도와
화방용품	35% -> 44% -> 57%(+13%)	18년도와 약한 양의 상관관계
관상어용품	26% -> 35% -> 58%(+23%)	18, 19년도와 상관관계 없음
수집품	45% -> 48% -> 64%(+16%)	
자동차용품	44% -> 53% -> 71%(+18%)	18년도와 약한 음의 상관관계, 19년도와 약한 양의 상관관계
반려동물	33% -> 45% -> 65%(+20%)	18년도와 상관관계 거의 없음
음반	34% -> 24% -> 52%(+28%)	18년도와 상관관계 거의 없음
악기	40% -> 41% -> 64%(+23%)	18, 19년도와 상관관계 거의 없음, 전년도 대비 1.53배
기타	26% -> 21% -> 48%(+27%)	18, 19년도와 약한 상관관계, 전년도 대비 2.3배
피아노	31% -> 36% -> 52%(16%)	전년도 대비 1.4배
건강관리용품	0.9% -> 1% -> 10%(+9%)	18년도와 상관관계 거의 없음, 전년도 대비 10배
먼지차단마스크	0.3% -> 0.4% -> 9.6%(+9%)	18년도와 상관관계 거의 없음, 전년도 대비 24배
주방용품	32% -> 37% -> 59%(+22%)	18, 19년도와 상관관계 거의 없음, 전년도 대비 1.6배
식기	25% -> 29% -> 55%(+26%)	18, 19년도와 상관관계 거의 없음, 전년도 대비 1.9배
조리기구	28% -> 32% -> 55%(+23%)	18, 19년도와 약한 상관관계, 전년도 대비 1.7배
커피용품	26% -> 29% -> 46%(+17%)	18, 19년도와 약한 상관관계, 전년도 대비 1.55배
제과/제빵용품	19% -> 23% -> 42%(+19%)	18년도와 약한 상관관계, 전년도 대비 1.8배
의료용품	7% -> 8% -> 14%(+6%)	18년도와 상관관계 거의 없음, 19년과 약한 음의 상관관계
눈건강용품	35% -> 41% -> 57%(+16%)	18년도와 상관관계 거의 없음, 19년도와 약한 양의 상관관계

카테고리명	연도별 클릭량 평균 변화	비고
실비용품	29% -> 40% -> 55%(+15%)	18년도와 상관관계 거의 없음, 꾸준한 상승폭
구강위생용품	47% -> 60% -> 75%(+15%)	18, 19년도와 상관관계 거의 없음.
발건강용품	26% -> 31% -> 49%(+18%)	18년도와 약한 음의 상관관계
수납/정리용품	33% -> 38% -> 58%(+20%)	18, 19년도와 약한 상관관계, 전년도 대비 1.53배
생활용품	12% -> 16% -> 27%(+11%)	18, 19년도와 상관관계 거의 없음, 전년도 대비 1.7배
청소용품	30% -> 38% -> 55%(+17%)	18년도와 약한 음의 상관관계, 19년과 약한 양의 상관관계
원예/식물	25% -> 29% -> 43%(+14%)	18, 19년도와 강한 상관관계
정원/원예용품	20% -> 27% -> 51%(+24%)	18, 19년도와 강한 상관관계, 전년도 대비 1.9배

3번 가설 검증

3. 코로나 확진자 수가 많았던 날 / 적었던 날에 유독 검색 및 클릭량이 증가 / 감소했던 특정 상품들이 있을 것이다.

```
c_data = pd.read_csv('corona_date.csv')
c_data = c_data['확진자']
c_data = c_data.loc[:298]

corona_data = []

for count in c_data:
    corona_data.append(count)

r_data = pd.read_csv(file_name)
r_data = r_data['Ratio']

r_data = r_data.loc[777:]

ratio_data = []

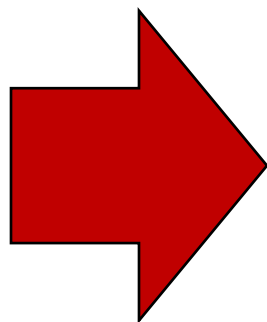
for ratio in r_data:
    ratio_data.append(ratio)

all_data = []
all_data.append(corona_data)
all_data.append(ratio_data)

print(corona_data[20])
print(ratio_data[20])

df = pd.DataFrame(all_data).T
df.columns = ['일별 확진자', 'Ratio']

corr = df.corr(method='pearson')
corr_file_name = 'Pearson_' + title + '_' + category + '_corr.csv'
corr.to_csv(corr_file_name, encoding='utf-8-sig')
```



	일별 확진자	Ratio
일별 확진자	1	-0.1087618
Ratio	-0.1087618	1

5) 취미와 관련된 상품들



수예(십자수 등) 용품

상관관계 **0.661411**로, 강한 양의 상관관계
전년 대비 클릭량 **1.9배**(26% 상승)



쿠킹박스

상관관계 **0.548608**로, 강한 양의 상관관계
전년 대비 클릭량 **2.7배**(22% 상승)

4) 취미와 관련된 상품들



러닝머신/덤벨(헬스 용품)
상관관계 **0.594342**로, 강한 양의 상관관계
전년 대비 클릭량 **1.78배**(26% 상승)



악기
상관관계 **0.353114**로, 양의 상관관계
전년 대비 클릭량 **1.53배**(23% 상승)

카테고리명	일별 코로나 확진자와의 상관계수	비고
여행용가방/소품	-0.357218	18, 19년도와 약한 양의 상관관계
신발용품	-0.393454	
여성가방	-0.329915	18년도와 약한 양의 상관관계
야구모자	-0.417597	18, 19년도와 상관관계 거의 없음
벨트	-0.349038	18, 19년도와 상관관계 거의 없음
아이새도	0.359245	18, 19년도와 강한 양의 상관관계
향수세트	0.574186	19년도와 상관관계 거의 없음
바디케어	0.354484	18, 19년도와 약한 양의 상관관계
영상가전액세서리	0.362089	18, 19년도와 약한 상관관계
스탠드	0.50639	19년도와 상관관계 거의 없음
손소독기	0.497521	18, 19년도와 상관관계 거의 없음
자외선소독기	0.361515	18, 19년도와 상관관계 거의 없음
믹서기	-0.339226	19년도와 상관관계 거의 없음
모니터 주변기기	0.42412	18년도와 약한 상관관계
책상	0.360126	18, 19년도와 약한 상관관계, 18, 19년도의 상관관계는 강함
수예	0.661411	18, 19년도와 약한 상관관계, 18, 19년도의 상관관계는 강함
자수	0.437544	19년도와 상관관계 거의 없음
퀼트/펠트	0.556045	19년도와 약한 음의 상관관계
원단	0.40943	18, 19년도와 약한 상관관계

카테고리명	일별 코로나 확진자와의 상관계수	비고
인테리어소품	0.412286	18, 19년도와 약한 상관관계
공구	0.366854	18년도와 약한 음의 상관관계, 19년도와 약한 양의 상관관계
쿠킹박스	0.548608	18년도와 상관관계 거의 없음
밥/죽 쿠킹박스	0.453824	18년도와 약한 음의 상관관계, 19년도와 상관관계 거의 없음
찌개/국 쿠킹박스	0.508873	18, 19년도와 약한 양의 상관관계
주방용품	0.346462	18년도와 약한 음의 상관관계, 19년과 상관관계 거의 없음
식기	0.352204	18, 19년도와 상관관계 거의 없음
잔/컵	0.415323	18년도와 상관관계 거의 없음, 19년과 약한 양의 상관관계
커피용품	0.384233	18, 19년도와 약한 상관관계
헬스	0.295841	18, 19년도와 약한 양의 상관관계
러닝머신	0.594342	18년도와 상관관계 거의 없음, 19년도와 약한 상관관계
아령/덤벨	0.422964	18, 19년도와 상관관계 거의 없음
배드민턴	-0.38783	18, 19년도와 약한 음의 상관관계
악기	0.353114	18년도와 상관관계 거의 없음, 19년과 약한 양의 상관관계

5. 고찰

- **클릭량 ≠ 판매량**

- 본 프로젝트의 최대 한계점.
- Pandemic 시대에 아이쇼핑만 했을 가능성도 충분.
- 클릭량 데이터와 판매량 데이터 사이의 상관관계를 알아볼 수 있으면 좋겠음.

- **증가했다, 감소했다 -> 코로나 때문이다?**

- 성급한 일반화의 오류가 될 수 있음. 매우 위험!!
- 클릭량에 영향을 주는 요인은 굉장히 많을 것(유행, 경제 등등)
- 단정짓지는 않았다.

- **클릭량 추이 분석을 이용한 RecSys 구축** 해볼 수 있으면 참 좋겠다

6. QnA