

포털사이트 검색 키워드와 실제 온라인 쇼핑 구매 품목 사이의 상관관계

경제학과 201900740 박혜인

2022년 2학기 컴퓨터공학부 빅데이터입문

지도교수: 최대진 교수님 | 발표일: 2022.12.08

Index

1

프로젝트 배경 및 목표

01. 프로젝트 배경 및 Question

02. 프로젝트 목표

2

연구대상 선정 및 가설 설정

3

데이터 수집 및 전처리

4

데이터 마이닝

01. EDA

02. 상관관계 분석

5

결론 및 제언

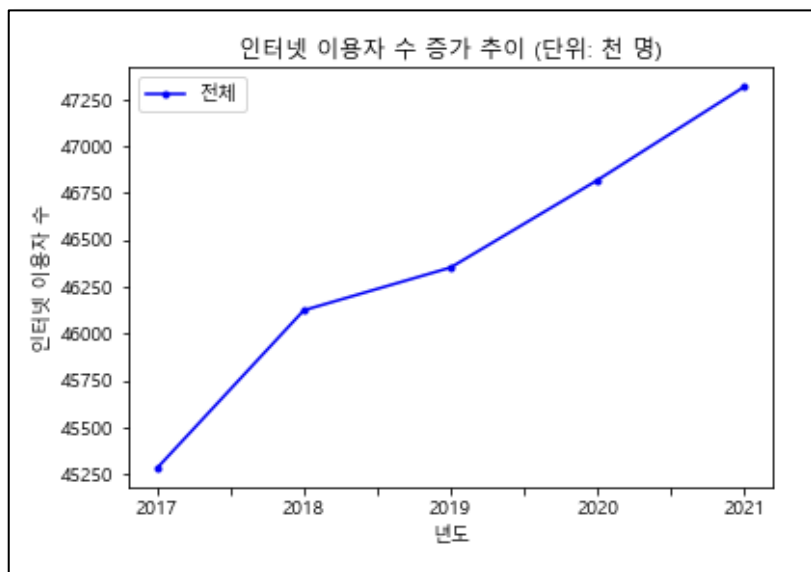
6

어려웠던 점 및 느낀점, 참고문헌

Chap 1. 프로젝트 배경 및 목표

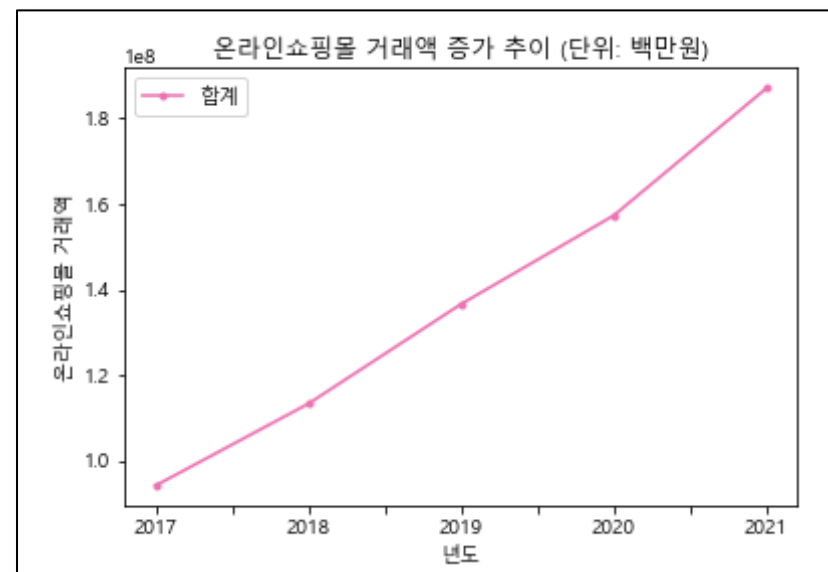
프로젝트 배경 (1)

- 인터넷 이용자 수 증가



과학기술 정보통신부의 「인터넷이용실태조사」을 살펴보면
2017~2021년 5년동안 100만명 가까이 증가함

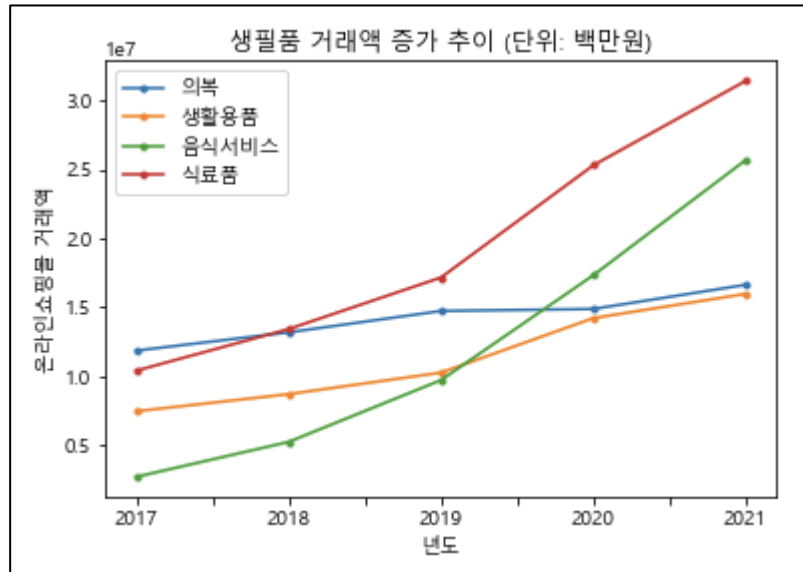
- 온라인 쇼핑 플랫폼의 활성화 및 거래액 증가



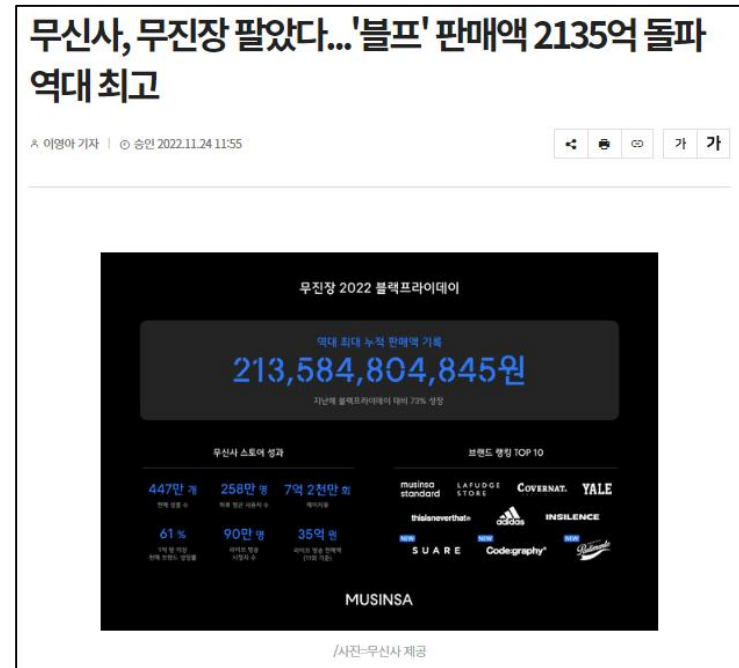
코로나 19로 인한 온라인 플랫폼 활성화와 더불어,
통계청 「온라인쇼핑동향조사」를 살펴보면 2017~2021년
5년동안 쇼핑물 거래액에 거의 2배 가까이 증가함

프로젝트 배경 (2)

- 온라인으로 생필품을 구매하는 경우 多



온라인 쇼핑이 증가함과 동시에, 마스크를 구매하거나 생필품을 온라인으로 구매하는 경우가 자연스럽게 많아짐



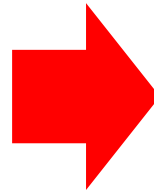
의류의 경우에도, 최근 '무진장 블랙프라이데이' 때 약 2135억을 판매해 역대 최대 수치를 기록할 만큼 온라인쇼핑 플랫폼이 활성화됨을 확인할 수 있음

Question

인터넷 이용자 수의 지속적인 증가

온라인 쇼핑 플랫폼 활성화 및 거래액 증가

온라인으로 생필품을 구매하는 경우가 많아짐



실제로 **관심을 많이 가지는 품목**과
실제로 **구매하는 품목** 사이에 **관련**이 있는가?

프로젝트 목표

- ✓ 사람들이 실제로 관심 있게 검색하는 상품과 구매하는 상품 사이에 차이가 있는지 확인
(네이버 검색트렌드의 검색량, 네이버 쇼핑인사이트 클릭량, 통계청 온라인쇼핑동향조사 데이터 활용)
- ✓ 검색 키워드와 온라인 쇼핑 구매 품목이 비례할 경우,
현재 인기있는 검색 키워드로 앞으로 어떤 상품이 더 많이 팔릴지 예측할 수 있음
- ✓ 만약, 정확히 알 수 없다면, 검색어 트렌드에 영향을 많이 받는 품목이 무엇인지 알아보고,
상대적으로 트렌드에 영향을 많이 받는 품목은 향후 상품 마케팅을 진행할 때 이 점을 활용할 수 있음

Chap 2. 연구대상 선정 및 가설 설정

분석 대상(품목) 설정

- '네이버 데이터랩' 쇼핑인사이트 분야별 인기검색어 중 8가지 품목 선택

- 2022년 11월 기준으로, 인기검색어에 있는 품목 활용
- 검색트렌드에 영향을 받는 품목과 영향을 받지 않는 품목을 골고루 8가지 선정
- 가능하면, 세분화된 검색어를 활용 (예: 패딩 → 내셔널지오그래픽 패딩)

2022.11. 2022.11.01. ~ 2022.11.30.	2022.11. 2022.11.01. ~ 2022.11.30.	2022.11. 2022.11.01. ~ 2022.11.30.	2022.11. 2022.11.01. ~ 2022.11.30.	2022.11. 2022.11.01. ~ 2022.11.30.	2022.05. 2022.05.01. ~ 2022.05.31.	2022.11. 2022.11.01. ~ 2022.11.30.
1 잠옷	1 양말	1 물티슈	1 오메가3	1 덴탈마스크	1 포켓몬방	1 샴푸
2 남자팬티	2 실내화	2 담배품	2 양배추즙	2 차량용방향제	2 아르기닌	2 바디워시
3 원피스	3 어그슬리퍼	3 휴간소음매트	3 사과	3 kf94마스크새부리형	3 닭가슴살	3 탈모샴푸
4 내셔널지오그래픽패딩	4 탈슬리퍼	4 젖병소독기	4 비오틴	4 크리스마스트리	4 식용유	4 바디로션
5 맨투맨	5 페이스삭스	5 백일상대여	5 절임배추	5 차량용핸드폰거치대	5 알티지오메가3	5 트리트먼트
6 남자니트	6 크룩스	6 어린이마스크	6 콜라겐	6 월세전반	6 단백질보충제	6 마스크팩
7 숏패딩	7 나이키운동화	7 아기로션	7 과메기	7 전자담배액상	7 비오틴	7 클렌징폼
8 후드티	8 목도리	8 유아마스크	8 산양유단백질	8 싱크대수전	8 생수	8 선크림
9 남성팬티	9 캐리어	9 순산양분유	9 비타민d	9 kf94마스크	9 프로틴	9 수분크림
10 패딩조끼	10 안전화	10 포켓몬마스크	10 유산균	10 새부리형마스크	10 콜라겐	10 여성청결제
< 패션의류 >	< 패션잡화 >	< 출산, 육아 >	< 식품 >	< 생활, 건강 >	< 식품 >	< 화장품, 미용 >

분석 대상(품목) 선정 결과

- 분석 품목 선정 결과

- 실생활 필수품: 물티슈, 마스크, 샴푸, 오메가 3



- 실생활 사치품: 내셔널지오그래픽 패딩, 나이키 운동화, 포켓몬빵, 아이폰



분야별 인기 검색어에는 없지만

20대들 사이에서 인기있는 품목이므로 추가

가설 설정

필수적으로 필요한 물티슈, 마스크, 샴푸, 오메가3 품목의 경우, 네이버 검색트렌드
검색량과 관계 없이 네이버쇼핑 클릭량 및 온라인쇼핑 거래액이 일정할 것이다.



이외에 내셔널지오그래픽 패딩, 나이키 운동화, 포켓몬빵, 아이폰 품목의 경우,
네이버 검색트렌드 검색량과 네이버쇼핑 클릭량 및 온라인쇼핑 거래액이 비례할 것이다.



Chap 3. 데이터 수집 및 전처리

데이터셋 설명

- 사용 데이터 목록

출처	자료원 명	시기	형태	비고
네이버 오픈 API	데이터 랩(검색어트렌드 API)	2019.01~2022.10	JSON	월별
	데이터 랩(쇼핑인사이트 API)	2019.01~2022.10	JSON	월별
통계청	온라인 쇼핑몰 판매 매체별/상품군별 거래액	2017~2021	CSV	23개 품목, 연도별 프로젝트 배경 소개에 사용
		2019.01~2022.10	CSV	6개 품목, 월별
	인터넷 이용자 수	2017~2021	CSV	프로젝트 배경 소개에 사용

API 데이터 수집 및 전처리

- 네이버 오픈 API “검색어트렌드”

- JSON 파일 내용을 CSV 형태로 저장 후 불러와서 분석
- 날짜: 2019-01-01 ~ 2022-10-31 (월 단위, 45개)
- 주제어와 키워드
 - groupName: 물티슈, keyword: {"물티슈", "wet tissue"}
 - groupName: 마스크, keyword: {"마스크", "mask", "KF94"}
 - groupName: 샴푸, keyword: {"샴푸", "shampoo"}
 - groupName: 오메가3, keyword: {"오메가3", "omega3", "알티지오메가3"}
 - groupName: 내셔널지오그래픽패딩, keyword: {"내셔널지오그래픽", "NationalGeographic", "DownJacket"}
 - groupName: 나이키운동화, keyword: {"나이키운동화", "에어포스", "Nike", "Airforce"}
 - groupName: 포켓몬빵, keyword: {"포켓몬빵", "Pokemon Bread", "띠부싹", "띠부띠부싹"}
 - groupName: 아이폰, keyword: {"아이폰", "iPhone"}
- PC, 모바일 기기 모두 포함 / 전 연령층 / 전 성별

API 데이터 수집 및 전처리

- 네이버 오픈 API "검색어트렌드"

```
{
  "startDate": "2019-01-01",
  "endDate": "2022-10-31",
  "timeUnit": "month",
  "results": [
    {
      "title": "물티슈",
      "keywords": ["물티슈", "wet tissue"],
      "data": [
        {
          "period": "2019-01-01",
          "ratio": 0.56559
        },
        {
          "period": "2019-02-01",
          "ratio": 0.53429
        },
        {
          "period": "2019-03-01",
          "ratio": 0.61167
        },
        {
          "period": "2019-04-01",
          "ratio": 0.4996
        },
        {
          "period": "2019-05-01",
          "ratio": 0.54495
        },
        {
          "period": "2019-06-01",
          "ratio": 0.53549
        },
        {
          "period": "2019-07-01",
          "ratio": 0.56008
        },
        {
          "period": "2019-08-01",
          "ratio": 0.59615
        },
        {
          "period": "2019-09-01",
          "ratio": 0.59615
        },
        {
          "period": "2019-10-01",
          "ratio": 0.59615
        }
      ]
    }
  ]
}
```

```
{
  "startDate": "2019-01-01",
  "endDate": "2022-10-31",
  "timeUnit": "month",
  "results": [
    {
      "title": "내셔널지오그래픽패딩",
      "keywords": ["내셔널지오그래픽", "National Geographic", "Down Jacket"],
      "data": [
        {
          "period": "2019-01-01",
          "ratio": 4.10906
        },
        {
          "period": "2019-02-01",
          "ratio": 3.20178
        },
        {
          "period": "2019-03-01",
          "ratio": 2.8197
        },
        {
          "period": "2019-04-01",
          "ratio": 2.35521
        },
        {
          "period": "2019-05-01",
          "ratio": 2.21309
        },
        {
          "period": "2019-06-01",
          "ratio": 2.17889
        },
        {
          "period": "2019-07-01",
          "ratio": 2.5676
        },
        {
          "period": "2019-08-01",
          "ratio": 3.30746
        },
        {
          "period": "2019-09-01",
          "ratio": 5.25736
        },
        {
          "period": "2019-10-01",
          "ratio": 7.42609
        }
      ]
    }
  ]
}
```

↓ JSON 파일 CSV로 변환

	Date	WetTissue	Mask	Shampoo	Omega3	NationalGeographic	Nike	PokemonBread	iPhone
0	2019-01-01	0.56559	1.61046	0.41763	1.05838	4.10906	5.54810	0.07909	8.47009
1	2019-02-01	0.53429	0.49862	0.35894	0.91135	3.20178	6.25700	0.07069	7.84667
2	2019-03-01	0.61167	3.51814	0.38366	1.01046	2.81970	7.18418	0.07543	7.81961
3	2019-04-01	0.49960	0.48491	0.36484	0.93520	2.35521	6.40646	0.07845	7.52693
4	2019-05-01	0.54495	0.43436	0.37971	0.87952	2.21309	6.08084	0.08395	7.74526
5	2019-06-01	0.53549	0.23072	0.41051	0.87101	2.17889	5.26553	0.08441	7.50547
6	2019-07-01	0.56008	0.21801	0.45389	0.93296	2.56760	5.54546	0.08800	8.23143

반환값 ratio

구간별 검색량의 상대적인 비율.

구간별 결과에서 가장 큰 값을 100으로 설정한 상댓값

API 데이터 수집 및 전처리

- 네이버 오픈 API “쇼핑인사이트”

- 검색트렌드와 마찬가지로, JSON 파일 내용을 CSV 형태로 저장 후 불러와서 분석
- 날짜: 2019-01-01 ~ 2022-10-31 (월 단위, 45개)
- 주제어와 키워드

- name: 물티슈, param: 50000117
- name: 마스크, param: 50002057
- name: 샴푸, param: 50000297
- name: 오메가3, param: 50002447
- name: 내셔널지오그래픽패딩, param: 50000837
- name: 나이키운동화, param: 50000788
- name: 포켓몬빵, param: 50000149
- name: 아이폰, param: 50001519

name은 쇼핑 분야 이름

param은 쇼핑 분야 코드

(네이버쇼핑 카테고리 선택 시 URL에 나오는 cat_id)

- PC, 모바일 기기 모두 포함 / 전 연령층 / 전 성별

API 데이터 수집 및 전처리

- 네이버 오픈 API "쇼핑인사이드"

```
{
  "startDate": "2019-01-01",
  "endDate": "2022-10-31",
  "timeUnit": "month",
  "results": [
    {
      "title": "물티슈",
      "category": ["50000117"],
      "data": [
        {
          "period": "2019-01-01",
          "ratio": 0.66213
        },
        {
          "period": "2019-02-01",
          "ratio": 0.66222
        },
        {
          "period": "2019-03-01",
          "ratio": 0.90056
        },
        {
          "period": "2019-04-01",
          "ratio": 0.55077
        },
        {
          "period": "2019-05-01",
          "ratio": 0.71133
        },
        {
          "period": "2019-06-01",
          "ratio": 0.61536
        },
        {
          "period": "2019-07-01",
          "ratio": 0.63082
        }
      ]
    }
  ]
}
```

```
{
  "startDate": "2019-01-01",
  "endDate": "2022-10-31",
  "timeUnit": "month",
  "results": [
    {
      "title": "오메가3",
      "category": ["50002447"],
      "data": [
        {
          "period": "2019-01-01",
          "ratio": 2.53916
        },
        {
          "period": "2019-02-01",
          "ratio": 2.09052
        },
        {
          "period": "2019-03-01",
          "ratio": 3.28553
        },
        {
          "period": "2019-04-01",
          "ratio": 3.11702
        },
        {
          "period": "2019-05-01",
          "ratio": 5.12498
        },
        {
          "period": "2019-06-01",
          "ratio": 4.85277
        },
        {
          "period": "2019-07-01",
          "ratio": 6.05629
        }
      ]
    }
  ]
}
```

```
{
  "startDate": "2019-01-01",
  "endDate": "2022-10-31",
  "timeUnit": "month",
  "results": [
    {
      "title": "포켓몬빵",
      "category": ["50000149"],
      "data": [
        {
          "period": "2019-01-01",
          "ratio": 32.18244
        },
        {
          "period": "2019-02-01",
          "ratio": 32.02583
        },
        {
          "period": "2019-03-01",
          "ratio": 32.82346
        },
        {
          "period": "2019-04-01",
          "ratio": 29.76936
        },
        {
          "period": "2019-05-01",
          "ratio": 28.58545
        },
        {
          "period": "2019-06-01",
          "ratio": 29.15926
        },
        {
          "period": "2019-07-01",
          "ratio": 27.48441
        }
      ]
    }
  ]
}
```

↓ JSON 파일 CSV로 변환

	Date	WetTissue	Mask	Shampoo	Omega3	NationalGeographic	Nike	PokemonBread	iPhone
0	2019-01-01	0.66213	2.07532	1.93409	2.53916	34.85331	18.53460	32.18244	0.98396
1	2019-02-01	0.66222	0.88453	1.53928	2.09052	16.88162	21.06782	32.02583	1.57132
2	2019-03-01	0.90056	6.45820	1.64664	3.28553	13.90070	22.38283	32.82346	2.10071
3	2019-04-01	0.55077	0.71351	1.55176	3.11702	8.08272	20.02252	29.76936	1.61800
4	2019-05-01	0.71133	0.47408	1.64314	5.12498	3.32740	19.81175	28.58545	1.49712
5	2019-06-01	0.61536	0.50464	1.80212	4.85277	2.29096	16.50768	29.15926	1.54392
6	2019-07-01	0.63082	1.08087	1.94925	6.05629	3.09402	17.05898	27.48441	2.09457

반환값 ratio

구간별 검색량의 상대적인 비율.

구간별 결과에서 가장 큰 값을 100으로 설정한 상대값

통계청 데이터 수집 및 전처리

• 통계청 KOSIS 데이터

- 2019년 1월부터 2022년 10월까지 월별 데이터 수집 (단위: 백만원)
- <온라인쇼핑동향조사 상품군 분류 및 예시> 자료를 참고하여, 품목으로 설정한 제품이 어느 조사범위에 속하는지 살펴보고, 해당하는 상품분류만 남기고 나머지는 삭제

- 멀티슈 → 생활용품
- 마스크 → 기타(의료기구)
- 샴푸 → 생활용품
- 오메가3 → 기타(의료기구)
- 내셔널지오그래픽 패딩 → 의류
- 나이키 운동화 → 신발
- 포켓몬빵 → 음식료품
- 아이폰 → 가전, 전자, 통신기기

8가지 품목이 해당하는
상품분류 6가지만 사용하도록
데이터 전처리 진행

	시점	가전/전자/통신기기	의복	신발	음식료품	생활용품	기타(의료)
0	2019.01	1063081	1107321	159646	1193353	831769	223939
1	2019.02	982059	959743	156431	867434	735651	221598
2	2019.03	1324884	1238235	206810	1027271	869192	327742
3	2019.04	1134227	1204083	204171	1058853	821882	195957
4	2019.05	1296165	1278281	217438	1111609	862291	195838

<온라인쇼핑동향조사 상품군 분류 및 예시>

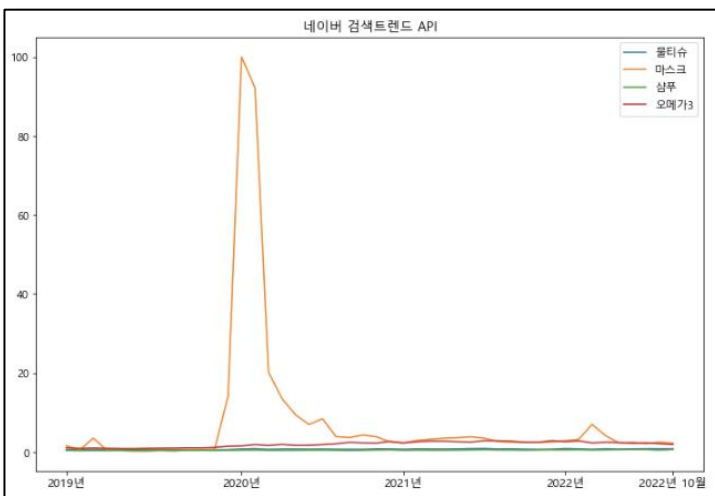
상품분류	조 사 범 위
컴퓨터 및 주변기기	PC, 노트북, 프린터, 스피커, CD형태 등 유형의 소프트웨어
가전·전자·통신기기	TV, 냉장고, 세탁기, 디지털카메라, 휴대폰 등
서적	각종 도서 (e-Book은 콘텐츠에 해당하여 조사에서 제외)
사무·문구	사무용품, 문구류, 다이어리/앨범, 종이류/복사지, 필기구 등
의복	의복류 (남성복, 여성복, 스포츠웨어, 아동·유아복 등)
신발	신발 (구두, 운동화, 샌들, 실내화, 아동화 등)
가방	가방 (핸드백, 가방, 여행용 등)
패션용품 및 액세서리	모자, 장갑, 스카프, 시계, 금반지, 각종 액세서리 등
스포츠·레저용품	운동용품, 레저용품, 등산화, 등산배낭 등
화장품	화장품, 향수, 화장품관련 소품 등
아동·유아용품	기저귀, 유모차, 그네, 아기침대, 보행기, 카시트, 인형, 완구 등
음·식·료·품	공산품류(커피, 차, 음료, 생수, 설탕, 식용유, 분유 등), 김치, 장류 및 장아찌류 등
농축수산물	곡물, 육류, 어류, 채소, 과일, 신선식품류 등
생활용품	주방용품, 침구, 비누, 샴푸, 세제, 화장지, 꽃, 화분 등
자동차 및 자동차용품	자동차, 오토바이, 튜닝/선팅용품, 내비게이션, 블랙박스, 엔진오일, 워셔액 등 자동차 관련용품
가구	가구 (장롱, 화장대, 신발장, 책상, 의자 등)
예완용품	예완용품 (사료, 장난감, 장신구 등)
여행 및 교통서비스	항공권, 교통티켓(버스, 기차), 렌터카, 숙박시설 등
문화 및 레저서비스	영화, 공연 등의 예약서비스
e쿠폰서비스	해당금액에 상응하는 서비스나 상품을 제공받을 수 있는 상품권
음식서비스	온라인 주문 후 조리되어 배달되는 음식 (피자, 치킨 등 배달서비스)
기타서비스	인화 등 주문제작, 이사, 청소 등 용역서비스, 각종 렌탈서비스
기타	문화상품권, 의료기구(안마의자제외), 골동품, 종교용품, 성인용품, 음반·비디오·악기 등

Chap 4. 데이터 마이닝

네이버 검색어트렌드 EDA

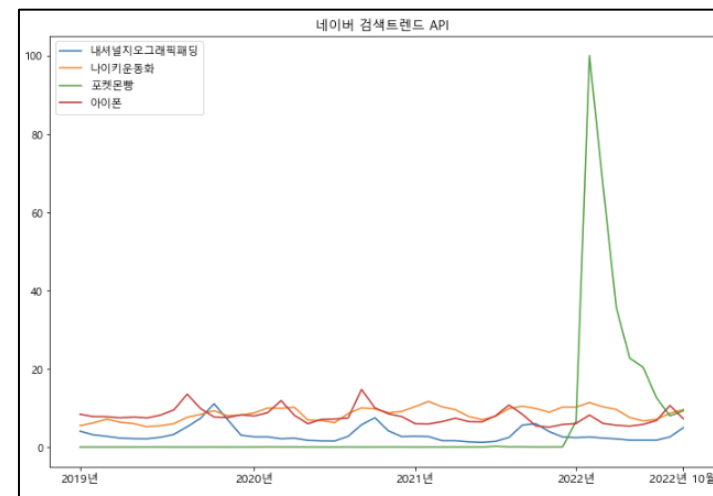
- 각 품목 검색량에 대한 시계열 그래프

< 실생활 필수품 >



- 마스크만 2020년 코로나19가 발생한 때 엄청난 검색량 차지
- 물티슈, 샴푸, 오메가3는 전반적으로 검색량이 많지 않고,
- 마스크의 경우에도 이후에는 검색량이 적은 모습을 확인할 수 있음

< 사치 품목 >

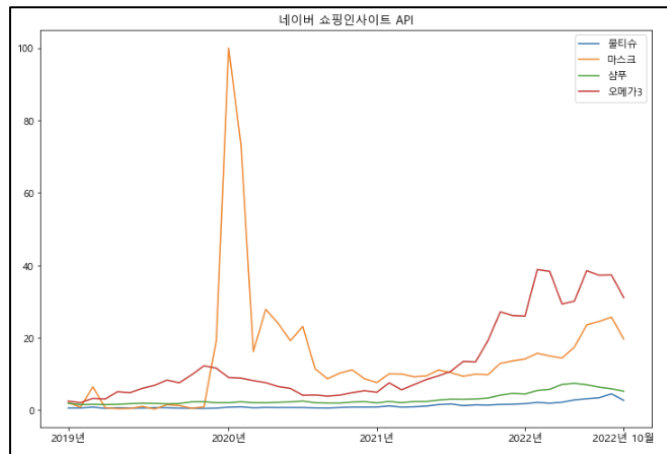


- 내셔널지오그래픽 패딩은 매년 겨울마다 검색량이 상승
- 나이키 운동화는 꾸준한 검색량을 가지고 있음
- 아이폰은 매년 새로운 모델이 나오는 하반기 초에 검색량 상승
- 포켓몬빵 신드롬을 일으킨 2022년 초에 엄청난 검색량, 현재는 낮은 수치

네이버 쇼핑인사이드 EDA

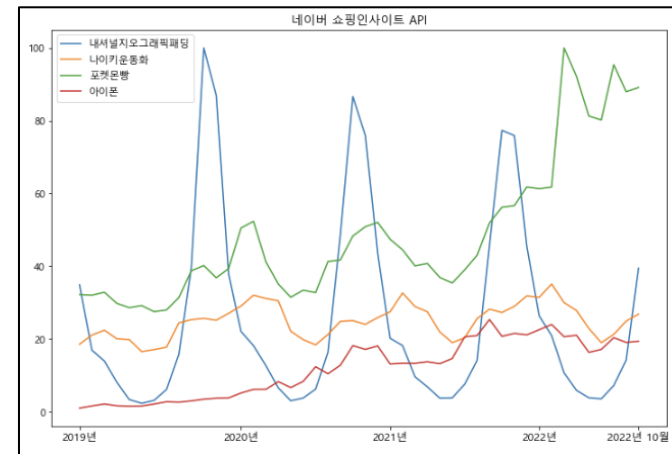
- 네이버쇼핑 각 품목 클릭량에 대한 시계열 그래프

< 실생활 필수품 >



- 마스크만 2020년 코로나19가 발생한 때부터 많은 클릭량 보유
이후에도 마스크를 꾸준히 구매함으로써, 검색량은 일정수준 유지
- 최근 건강을 챙기는 MZ세대가 늘어나, 오메가3에 대한 쇼핑 클릭량이 증가
- 물티슈와 샴푸는 주기적으로 구매율이 일정하므로, 그래프도 일정

< 사치 품목 >

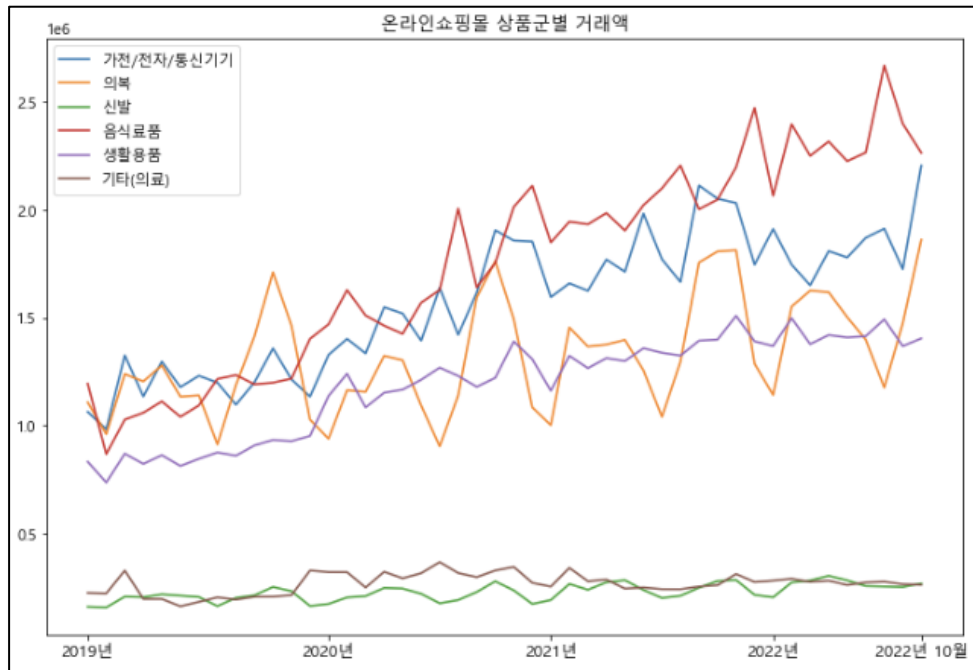


- 내셔널지오그래픽 패딩은 매년 겨울마다 클릭량이 상승
- 나이키 운동화는 매년 블랙프라이데이가 열리는 연말에 클릭량이 조금 상승
- 포켓몬빵은 신드롬을 일으킨 2022년 초 클릭량이 급격하게 상승.
하지만 온라인쇼핑몰에 품절되는 경우가 많아 다른 시기와 큰 차이가 나지 않음
- 아이폰을 사용하는 사람이 많아지면서 품목 클릭량도 증가

온라인쇼핑 거래액 EDA

- 온라인쇼핑몰 상품군별 거래액 시계열 그래프

- 통계청에서 조사한, 6가지 품목(가전/전자/통신기기, 의복, 신발, 음식료품, 생활용품, 기타(의류))에 대한 그래프



코로나 19 이후, 음식료품의 온라인쇼핑 거래액이 증가함

의복의 경우, 상대적으로 단가가 비싼 겨울 의상을 구매할 때
온라인쇼핑 거래액 증가

이외에도 가전/전자/통신기기, 생활용품도 온라인으로 구매하는
경우가 점차 증가하고 있음

적용할 데이터 마이닝 기법

- 데이터 마이닝 기법: **상관관계 분석** 사용

- 대표적인 상관관계 구하는 방법: Pearson(피어슨), Kendall(켄달), Spearman(스피어만)
→ 앞의 EDA에서 확인했듯이 변수가 모두 연속형 변수이므로, **Pearson(피어슨) 상관관계를 사용**

- Pearson : 상관관계에서 기본으로 사용되는 상관관계로, **연속형 변수의 상관관계** 측정 (신장, 몸무게) / 모수 검정
- Kendall : 변수값 대신 순위로 바뀌어서 사용하는 상관관계로, **샘플사이즈가 작거나, 데이터의 동률이 많을 때 유용** / 비모수 검정
- Spearman : 변수값 대신 순위로 바뀌어서 사용하는 상관관계로, 데이터 내 편차와 에러에 민감 / 켄달보다 높은 값 / 비모수 검정

- 구한 상관관계들이 유의한지, 유의수준 0.01, 0.05에서 유의한지 확인
- 시각화 방법: 산포도

검색량과 클릭량 사이의 상관관계 분석

- 피어슨 상관계수 검정

	피어슨 상관계수	P-Value
물티슈	0.5646152719559807	4.363724940588112e-05
마스크	0.9234368201110952	6.439861723605964e-20
삼푸	0.7764446951037397	2.2815095649046045e-10
오메가3	0.40985393283434746	0.004673740203139446
내셔널지오그래픽 패딩	0.8362009097650471	4.704985852316415e-13
나이키 운동화	0.9133191744129401	8.889876902226972e-19
포켓몬빵	0.5465609361415198	8.504496564405955e-05
아이폰	-0.2666534776053485	0.07323706689093089

피어슨 상관계수를 구한 결과, 마스크와 나이키 운동화가 1에 가까운 양의 상관관계를 가지고 있음을 확인

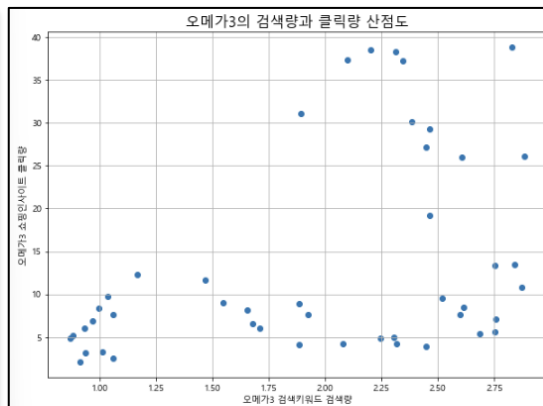
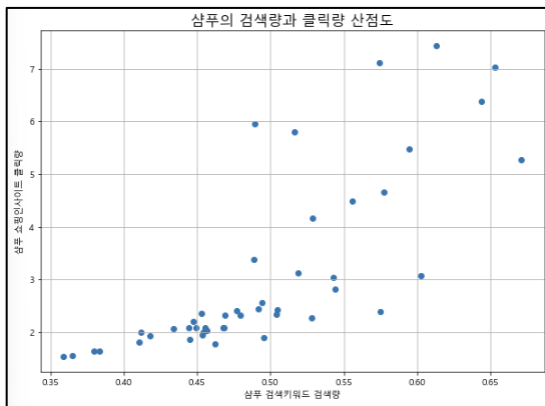
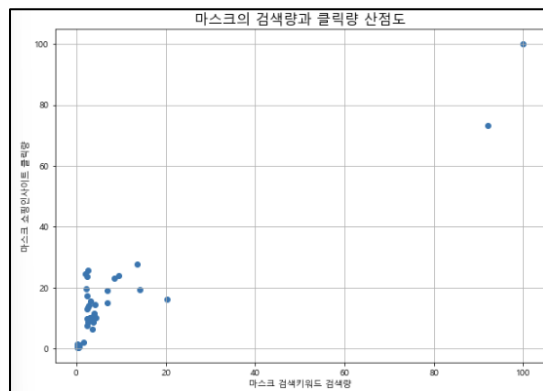
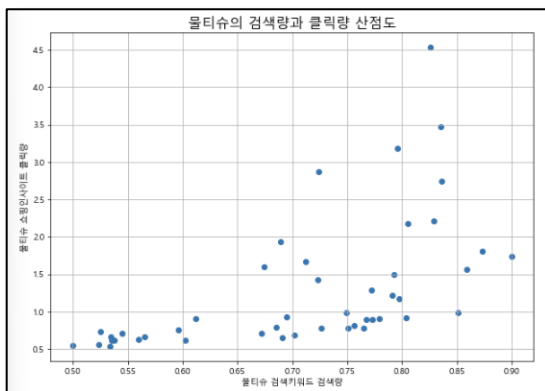


P-value의 경우, 아이폰을 제외한 7가지 품목에서 0.01, 0.05의 유의수준에서 유의하다는 것을 확인

따라서, 유의수준 0.05에서도 유의하지 않은 아이폰 품목은 검색량과 클릭량 사이의 상관관계 분석에서 제외

검색량과 클릭량 사이의 상관관계 분석

- 물티슈, 마스크, 샴푸, 오메가3 산포도 (실생활 필수품)



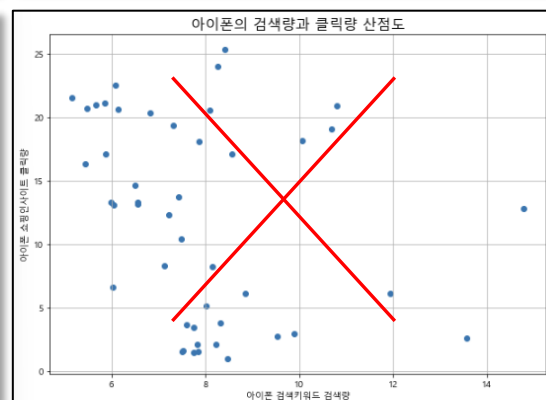
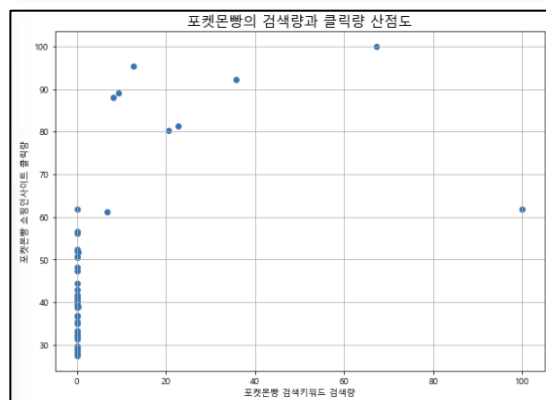
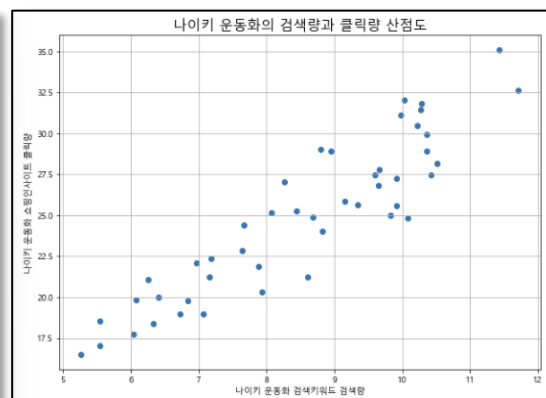
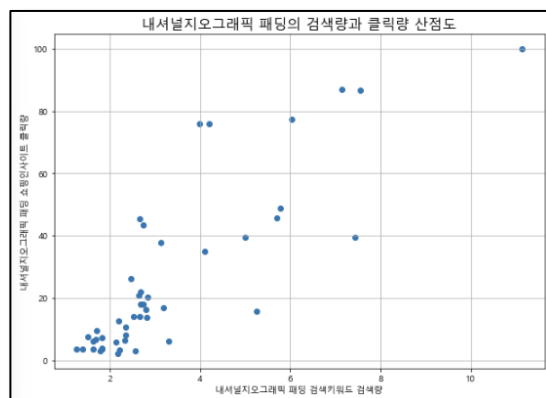
물티슈와 오메가3는 각각 상관계수가 0.5, 0.4으로
검색량과 쇼핑 품목 클릭량 사이에 적당한 상관관계 존재

마스크의 경우, 상관계수가 0.92로 검색량과 클릭량
사이에 강한 상관관계가 있음. 이는 코로나 19로 인해,
마스크를 구매하고자 하는 사람들의 영향도 배제할 수 X

샴푸도 상관계수가 0.77로 상관관계가 높은 편.
주로, 개인이 사용하는 샴푸 브랜드가 있기 때문에,
검색량과 클릭량 사이에 양의 상관관계가 있다고 볼 수 있음

검색량과 클릭량 사이의 상관관계 분석

- 내셔널지오그래픽 패딩, 나이키 운동화, 포켓몬빵, 아이폰 산포도 (사치품)



내셔널지오그래픽은 상관계수가 0.84로 검색량과 클릭량 사이에 강한 양의 상관관계를 가지고 있음. 겨울 시즌에 유행하는 상품인만큼, 분포는 아래에 쏠려 있음.

나이키 운동화는 상관계수가 0.913으로 검색한 사람이 쇼핑 품목을 클릭했다고 볼 수 있을 정도로 강한 상관관계

포켓몬빵은 검색한 만큼 쇼핑 품목을 클릭하지는 않음. 적당한 양의 상관관계 0.55

아이폰은 유의하지 않으므로 분석에 포함하지 않음

검색량과 클릭량 사이의 상관관계 분석

- 정리해보면

- 실생활 필수품인 **마스크**와 **샴푸**: 검색량과 쇼핑 품목 클릭량이 비례 (상관계수 각각 0.92, 0.77)
 - 마스크: 코로나 19가 발생한 시점에 검색량이 급격하게 많아졌을 수도 있음을 고려
 - 샴푸: 여성의 경우, 개인이 사용하는 샴푸 브랜드가 있을 수 있음. 이러한 관점에서 검색량과 클릭량은 비례한다고 볼 수 0
 - 사치품으로 분류한 **내셔널지오그래픽 패딩**: 겨울 시즌 상품임에도 불구하고, 검색량과 클릭량이 비례 (상관계수 0.84)
나이키 운동화: 검색량과 쇼핑 품목 클릭량이 0.913으로, 강한 상관관계를 나타냄
 - 하지만, **포켓몬빵**의 경우, 온라인에서 구매하기 힘들고 주로 오프라인에서 구매하기 때문에 강한 상관관계를 나타내지 못함
- ▶ 유의: 검색량과 클릭량이 비례한다고, 실제 포털 사이트 검색 키워드와 온라인쇼핑 구매 품목이 동일하다고 볼 수 없음

검색량과 거래액 사이의 상관관계 분석

- 피어슨 상관계수 검정

	피어슨 상관계수	P-Value
물티슈	0.8618726303464963	1.4593005580658092e-14
마스크	0.3246466412393815	0.0277170142552584
삼푸	0.7780914827937033	1.975036476943985e-10
오메가3	0.4691100584124052	0.00100684427971092
내셔널지오그래픽 패딩	0.5035487694183518	0.00036068026898323553
나이키 운동화	0.41713918764699986	0.003925797870409121
포켓몬빵	0.41200804264725727	0.004440604899880114
아이폰	-0.3387119207913757	0.021303602828846305

피어슨 상관계수를 구한 결과, 검색량과 거래액 사이에 아주 강한 양의 상관관계를 가지고 있는 것은 물티슈와 삼푸 1에 가까운 양의 상관관계를 가지고 있음을 확인

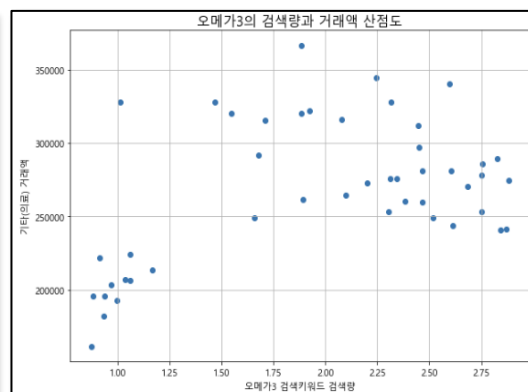
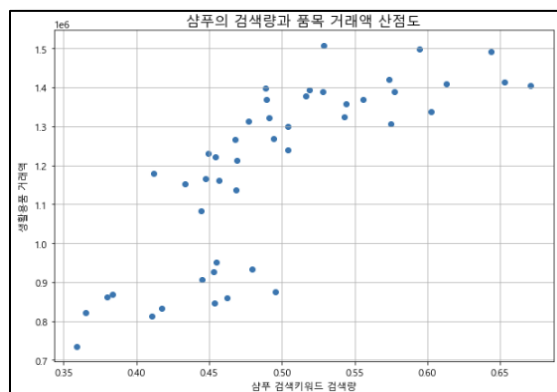
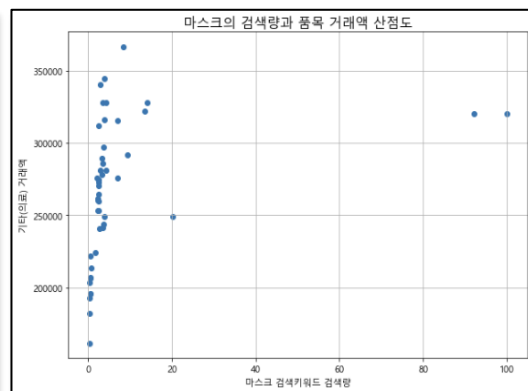
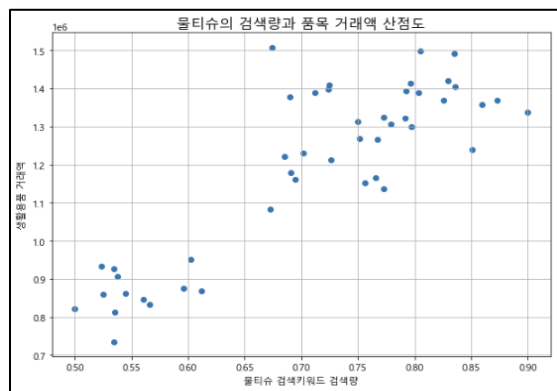


P-value의 경우, 마스크와 아이폰은 0.05 수준에서 유의 나머지 품목은 0.01의 유의수준에서도 유의함

따라서, 이 분석에서는 유의수준 0.05로 두고 분석 (제외하는 값 없이 진행)

검색량과 거래액 사이의 상관관계 분석

- 물티슈, 마스크, 샴푸, 오메가3 산포도 (실생활 필수품)



물티슈와 샴푸의 생활용품 거래액과의 상관계수는 각각 0.86, 0.77로 강한 양의 상관관계를 이루고 있음

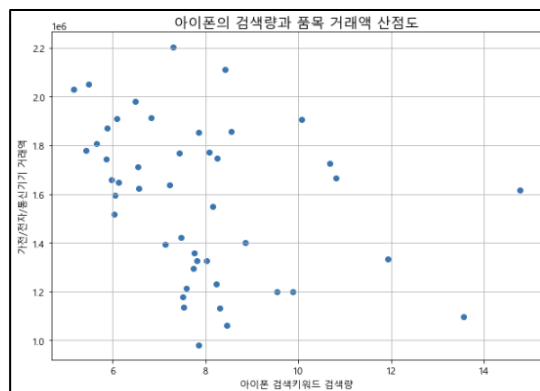
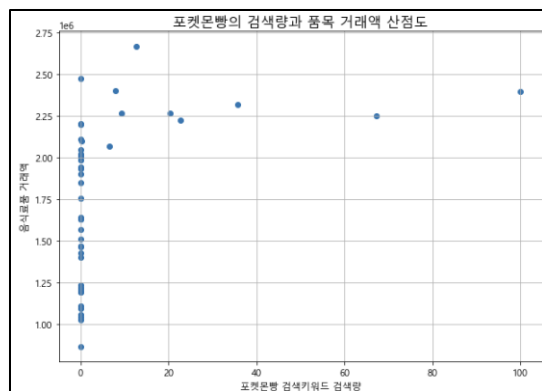
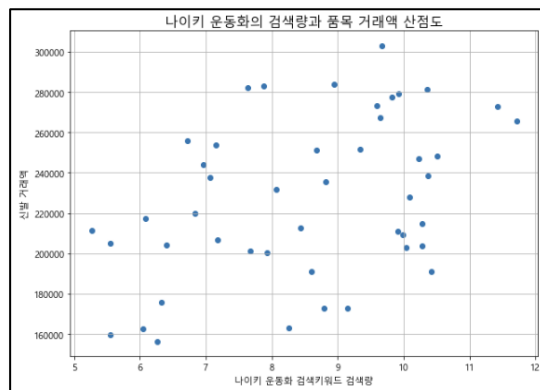
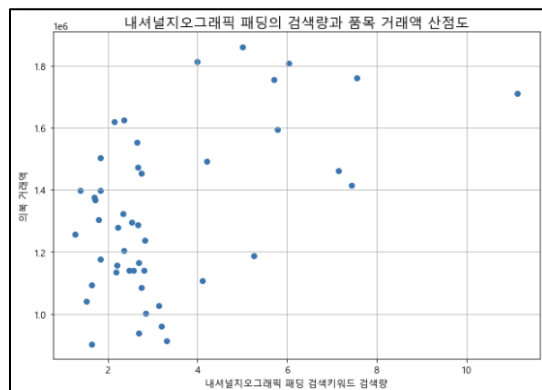
물티슈와 샴푸의 산포도도 우상향하는 모습을 보이기에 검색 키워드를 검색한만큼, 거래액이 증가했다고 볼 수 있음

반면, 마스크와 오메가3의 기타(의료)와의 상관계수는 각각 0.32, 0.47로 중간 정도의 양의 상관관계를 지님

하지만, 산포도 상으로 분포가 흩어져있거나, 한 줄로 쏠려 있어 상관관계가 있다고 말하기에 명확하지 않음

검색량과 거래액 사이의 상관관계 분석

- 내셔널지오그래픽 패딩, 나이키 운동화, 포켓몬빵, 아이폰 산포도 (사치품)



내셔널지오그래픽 패딩의 경우, 검색량과 거래액 사이의 상관관계수가 0.50이지만, 겨울에 주로 판매되는 의류임을 고려하면, 다른 변수들과 약간 차이가 있음

나이키 운동화의 경우, 분포가 조금 흩어져 완벽하게 상관관계를 가지고 있다고 말하기에 명확하지 않음

포켓몬빵의 경우, 갑자기 신드롬을 일으켰기에 이전까지 상관관계가 없다가, 갑자기 상관이 있어졌을 수도 있음 (0.50)

아이폰은 -0.339로 음의 상관관계를 지님
즉, 아이폰을 검색한다고해서 사람들이 구매하는 것은 아님

검색량과 거래액 사이의 상관관계 분석

- 정리해보면

- 실생활 필수품인 물티슈와 샴푸: 검색량과 온라인쇼핑 품목 거래액이 비례 (상관계수 각각 0.86, 0.77)
→ 산포도 상으로도 우상향하는 모습이 뚜렷하게 보이기 때문에, 검색량만큼 거래액이 증가했다고 볼 수 있음
 - 사치품으로 분류한 내셔널지오그래픽 패딩: 겨울에 주로 판매되는 의류임에도 불구하고, 중간 정도의 상관관계를 가짐 (0.50)
포켓몬빵: 2022년 갑작스러운 판매 증가가 있을 수 있기 때문에, 급격하게 수치가 변했을 가능성 있음
아이폰: 상관계수가 -0.339로 음의 상관관계를 지님.
 - 일반적으로, 검색을 한다고 해서 핸드폰을 그 자리에서 구매하지 않듯이, 사람들의 일반적인 소비를 반영한 점도 확인할 수 있음
- ▶ 유의: 검색량과 거래액이 비례한다고, 실제 포털 사이트 검색 키워드와 온라인쇼핑 구매 품목이 동일하다고 볼 수 없음

Chap 5. 결론 및 제언

가설 검증 확인 (1)

필수적으로 필요한 물티슈, 마스크, 샴푸, 오메가3 품목의 경우, 네이버 검색트렌드
검색량과 관계 없이 네이버쇼핑 클릭량 및 온라인쇼핑 거래액이 일정할 것이다.



마스크의 경우, 검색량과 네이버쇼핑 클릭량이 높은 상관관계가 있다.

물티슈의 경우, 검색량과 온라인쇼핑 거래액이 높은 상관관계가 있다.

샴푸의 경우, 검색량과 네이버쇼핑 클릭량, 온라인쇼핑 거래액 모두 높은 상관관계가 있다.

오메가3를 제외한, 나머지 3가지 상품은 검색량이 클릭량, 온라인쇼핑 거래액과 상관관계가 있다.

가설 검증 확인 (2)

이외에 내셔널지오그래픽 패딩, 나이키 운동화, 포켓몬빵, 아이폰 품목의 경우,
네이버 검색트렌드 **검색량**과 네이버쇼핑 **클릭량** 및 **온라인쇼핑 거래액**이 **비례**할 것이다.



내셔널지오그래픽 패딩의 경우, 검색량과 클릭량, 온라인쇼핑 거래액에 상관관계가 있다.

나이키 운동화의 경우, 검색량과 클릭량은 높은 상관관계가 있다.

아이폰은 검색량과 클릭량은 비교할 수 없고, 온라인쇼핑 거래액과는 음의 상관관계를 지닌다.

포켓몬빵을 제외한, 나머지 3가지 상품은 검색량이 클릭량, 온라인쇼핑 거래액과 상관관계가 있다.

최종 결론 및 제언

• 최종 결론 및 제언

사람들은 물건을 구매하기 전에, 구매하고자 하는 상품에 대해 검색하며 온라인쇼핑 거래로 이어지는 경우가 많음

하지만, 포켓몬빵과 같이 주로 오프라인에서 구매가 이루어지는 상품의 경우에는 상관관계가 없음
또한, 아이폰과 같이 고가의 상품의 경우에는 검색량과 클릭량, 거래액이 비례하지 않는 경우가 많음
(일반적인 사람들은 고가의 상품을 사기 전에 여러 가지 상품을 비교해보고 찾아보기 때문)

또, 클릭량과 온라인쇼핑 거래액에 해당 상품의 판매 수치를 완벽하게 대신할 수 없음 (광고를 통해 유입 되었을지, 아이쇼핑을 했을지 아무도 모름)
검색을 하는 이유가 물건을 구매하려는 목적이 아니라, 단순히 유행을 하거나 주위 친구들과 이야기를 하다가 검색을 했을지 알 수 없음

검색량과 네이버쇼핑 클릭 수, 그리고 온라인쇼핑 거래액이 완벽하게 관련이 있다고 볼 수는 어렵지만
대체적으로, 상품을 구매하기 전에 위와 같은 행동을 한다는 것을 염두해두면, 앞으로의 상품 판매도 충분히 예측할 수 있다

Chap 6. 어려웠던 점 및 느낀점, 참고문헌

TroubleShooting & Impression

- 프로젝트를 진행하면서 어려웠던 점

- 네이버 오픈 API를 처음 다뤄봐서 데이터를 수집하는 데에 생각했던 것보다 시간이 오래 걸렸음
 - 결과를 보면 왜 오래 걸렸는지 모르겠지만, 거의 10일 가까이 API를 활용하는 공부만 진행하여 다른 것들을 할 시간이 부족했음
- 적용한 데이터 마이닝 기법: 상관관계 분석
 - 다양한 기법을 사용해보고 싶어서 폭넓게 찾아봤지만, 상관관계 분석에 그쳤던 점이 아쉬웠음

- 느낀점

- 프로젝트의 처음부터 끝까지 내 손으로 다 끝내 본 경험이 처음이라 굉장히 의미 있었음
- 정말 단순한 개념이었지만, 오랜 시간 이해하지 못하는 스스로를 보면서 좌절도 많이 했지만, 결국 해내고 난 후에 느껴지는 뿌듯함에 빅데이터 분석에 또 다시 흥미를 느끼게 된 계기가 됨
- 기회가 된다면, 방학 때 Side Project로써, 전공지식도 살려서 데이터 마이닝을 해보고 싶음

참고문헌

- 노호영, 김성용, 김태영. "소비자의 인터넷 검색활동이 농축산물 구매에 미치는 영향" 농업경제연구60, no.2 (2019): 23-42.
- 김영식, 이승호. "네이버 데이터랩 검색어 트렌드 서비스를 이용한 온라인 포털에서의 한약재 검색 트렌드와 의미에 대한 고찰" 대한본초학회지(본초분과학회지)36, no.5 (2021): 1-14.
- 과학기술정보통신부, 「인터넷이용실태조사」, 2021, 2022.11.21, 인터넷 이용자수
- 통계청, 「온라인쇼핑동향조사」, 2022.10, 2022.11.21, 온라인쇼핑몰 판매매체별/상품군별거래액
- 이영아, "무신사, 무진장 팔았다... '블프' 판매액 2135억 돌파 역대 최고", Tech M, 2022년 11월 24일자

Q & A