

미세먼지에 따른 업종별 신한카드 매출 분석

미세베이션 2019

이승태 (팀장) 오효은 조아란 이미연 김만덕

목차 (CONTENTS)

1. 개요
2. 미세먼지 관련 사회적 이슈
3. 데이터 전처리 및 EDA
4. 의사결정나무
5. 시계열 분석
6. 비즈니스 모델

개요 (OVERVIEW)

데이터 분석 목적

- 실제 미세먼지 농도보다는, 미세먼지에 대한 사람들의 인식이 소비 행동에 영향을 준다
- 의사결정나무를 통해, 미세먼지 실제 수치와 카드 매출액 관계를 알아본다.
- 시계열 분석을 통해, 미세먼지 SNS 언급량과 카드 매출액 관계를 알아본다.
- 미세먼지 SNS 언급량에 따라, 카드 매출이 달라지는 target 업종을 선별한다.
- 민감하게 매출이 달라지는 업종 중에서, 매출액이 높은 행정구역들을 선정한다.



분석결과를 기반으로 새로운 비즈니스 모델 제시

- 미세먼지 영향에 따른 매출 변화를 분석하여, 새로운 사업 모델에 대한 아이디어를 제시하고자 함
- 업종별, 행정동별 미세먼지 민감군에 따른 마케팅 전략 활용
- 미세먼지 민감 행정동에 소비자 참여형 미세먼지 정화 벤치를 설치하여 공유가치 창출

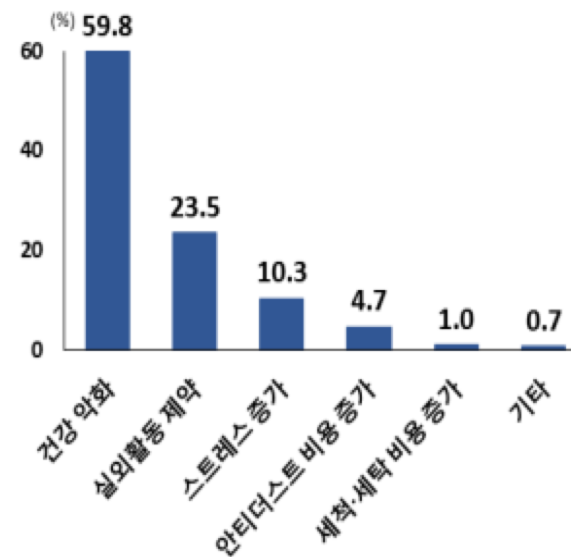
미세먼지 관련 사회적 이슈: 문제점 제시

미세먼지의 부정적인 사회 인식 및 소비 패턴에 영향

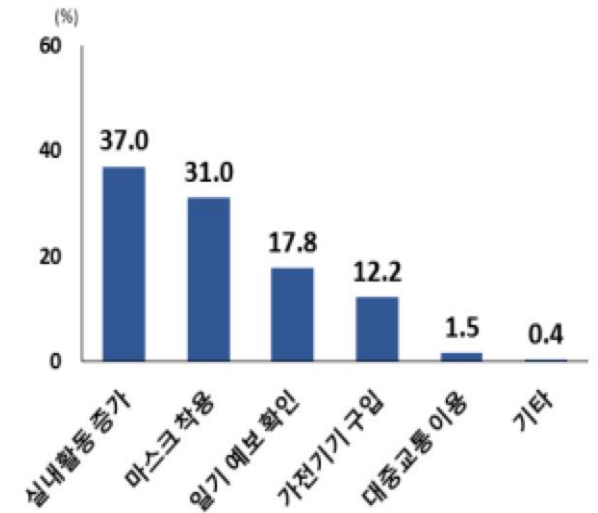
- 최근 몇 년, 미세먼지가 심해진다는 인식이 강해지면서, 소비 활동 및 일상 생활에 피해 커짐
- 미세먼지가 나빠진다는 인식이 심해질수록, 사람들의 외부활동 및 소비에 영향¹
 - 야외 활동 급감 & 실내 활동 소비 증가
 - 헬스케어 관련 건강증진 관심도 급증
- 미세먼지로 인한 소비심리 위축, 그에 따른 경제적 손실 4조원²
- 적극적으로 미세먼지를 해결하려는 사람들의 인식 변화
 - 미세먼지 해결 비용 15년 전 3,000원→ 현재 10,000원 지불 의사

미세먼지 관련 인식 설문조사²

〈미세먼지로 인한 가장 심각한 피해〉



〈미세먼지로 인한 일상생활 피해〉



1. 하나금융경영연구소 (2019).
2. 현대경제연구원 (2019).

미세먼지 관련 사회적 이슈 : 미세먼지에 대한 대중 인식

미세먼지 뉴스 많은 날, 리조트 카드매출 '울상', 세탁소 '활짝'

"소비자들이 대부분 뉴스를 통해 미세먼지 관련 정보를 인식하기 때문에, 실제 미세먼지가 측정된 농도보다는 관련 뉴스량에 소비 행동이 달라지는 현상이 나타났다"

- 전자신문 etnews -

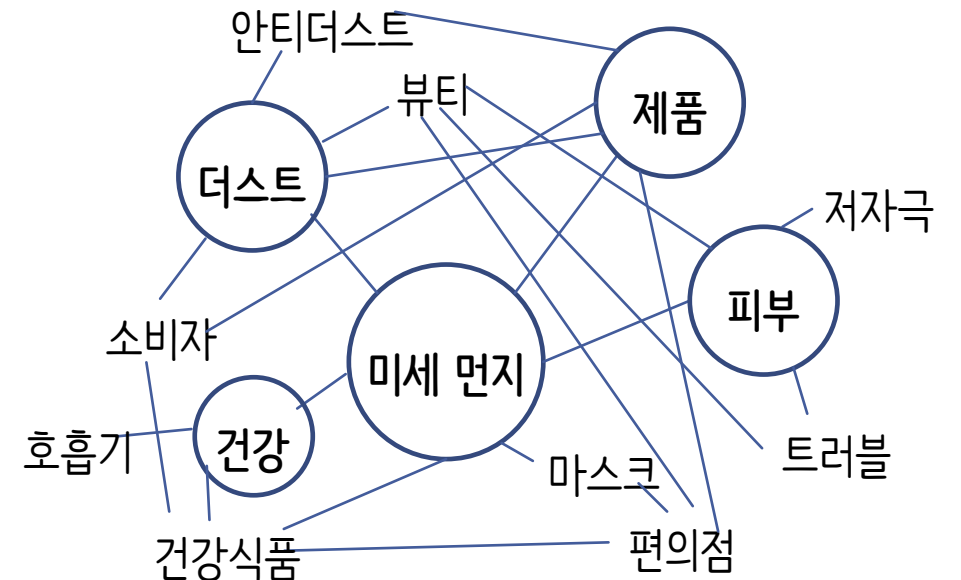
최악의 미세먼지에 "영화관·멀티플렉스 등 신용카드 사용 급증"

"미세먼지가 심한 날일수록 신용카드 사용이 증가하는 등 소비자들의 소비 경향에도 영향을 주는 것으로 나타나...야외 활동은 줄고 실내에서 주로 머물면서 멀티플렉스나 영화관, 복합쇼핑몰 등의 신용카드 매출이 크게 증가"

- 경향신문 -

미세먼지에 관한 기사가 대중들의 인식에 주는 영향력

미세먼지 관련 키워드 크롤링: 미세먼지, "제품", "피부", "공기", "건강"



데이터 전처리 &
EDA

시계열 분석



의사결정나무

비즈니스 모델

데이터 사용 변수

서울시 종로구와 노원구의 36개 행정동에서 수집된 데이터들의 설명

| 데이터 이름 | 사용 변수 | 활용 분석 | 출처 및 데이터 수집 기간 |
|----------|--|---|----------------------------------|
| 1. 카드 | 날짜, 행정동, 행정구, 연령, 성별, 업종, 매출액 | <ul style="list-style-type: none">• EDA• 의사결정나무• 시계열 분석 | 신한 카드 (2018/04/01~2019/03/31) |
| 2. 환경 기상 | 날짜, 시간, 행정동, 행정구, 습도, 소음, 미세먼지(pm10), 초미세먼지(pm2.5), 온도 | <ul style="list-style-type: none">• EDA• 의사결정나무 | 케이웨더 (2018/04/01~2019/03/31) |
| 3. SNS | 날짜, 미세먼지 언급, 그 외 단어 언급 | <ul style="list-style-type: none">• EDA• 시계열 분석 | 와이즈넷 (2018/04/01~2019/03/31) |

데이터 전처리 및 EDA:

(1) SNS 데이터

Konlpy twitter를 이용한 분석 과정

수집

- SNS 1~8에서 무작위로 데이터 수집

전처리

- Konlpy twitter로 명사만 추출 후, 빈도 카운트
- (실행시간 이득위해 빈도수를 비율로 바꿔, 새로 txt 작성)

분석

- 피부, 마스크, 아이, 케어 등 미세먼지관련해서, 대부분의 사람들이 건강/뷰티에 관심있음을 알 수 있음

워드클라우드 이미지



데이터 전처리 및 EDA:

(2) 환경 데이터

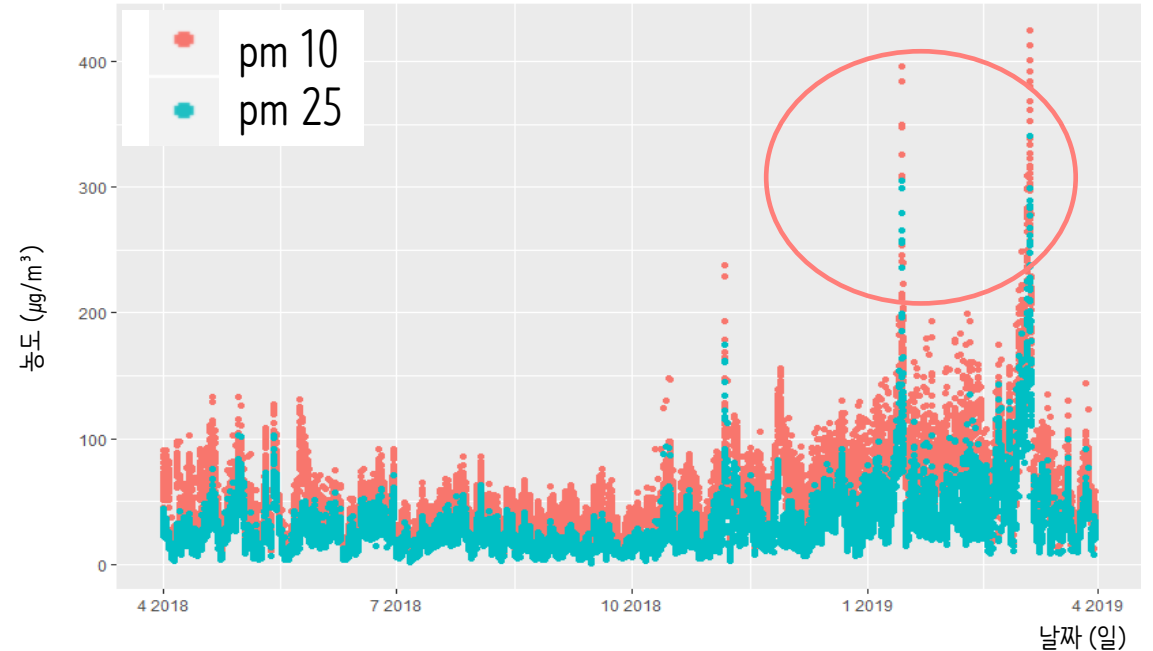
미세먼지 데이터 전처리 과정

- 결측치 처리
 - 1) 인접 측정소들 값들의 평균으로 대체
 - 2) 월별, 일별, 시간별 행정동 미세먼지 계산 시, 결측치 제거됨
- 극단치 처리
 - (특정 관측치 앞 5분 관측들의 최대값 $\times 10$)가 (특정 관측치) 보다 작을 경우, 해당 관측치를 결측값 처리
- 파생 변수 생성
 - 미세먼지 수치를 환경부 기준으로 등급화:
“0” 좋음, “1” 보통, “2” 나쁨, “3” 매우 나쁨

| 날짜 | 행정동 | 행정구 | humi | noise | pm10 | pm25 | temp | pm25등급 | pm10등급 |
|------------|-----------|-----|----------|----------|----------|----------|----------|--------|--------|
| 2018-04-01 | 종로1,2,3,4 | 0 | 60.6081 | 51.90208 | 64.66134 | 32.31477 | 16.67361 | 1 | 1 |
| 2018-04-02 | 종로1,2,3,4 | 0 | 58.21557 | 52.86427 | 62.81318 | 25.43179 | 19.59567 | 1 | 1 |
| 2018-04-03 | 종로1,2,3,4 | 0 | 62.77242 | 53.24526 | 61.51412 | 18.09337 | 19.30075 | 1 | 1 |

일별 미세먼지 농도 (2018년 4월~ 2019년 3월)

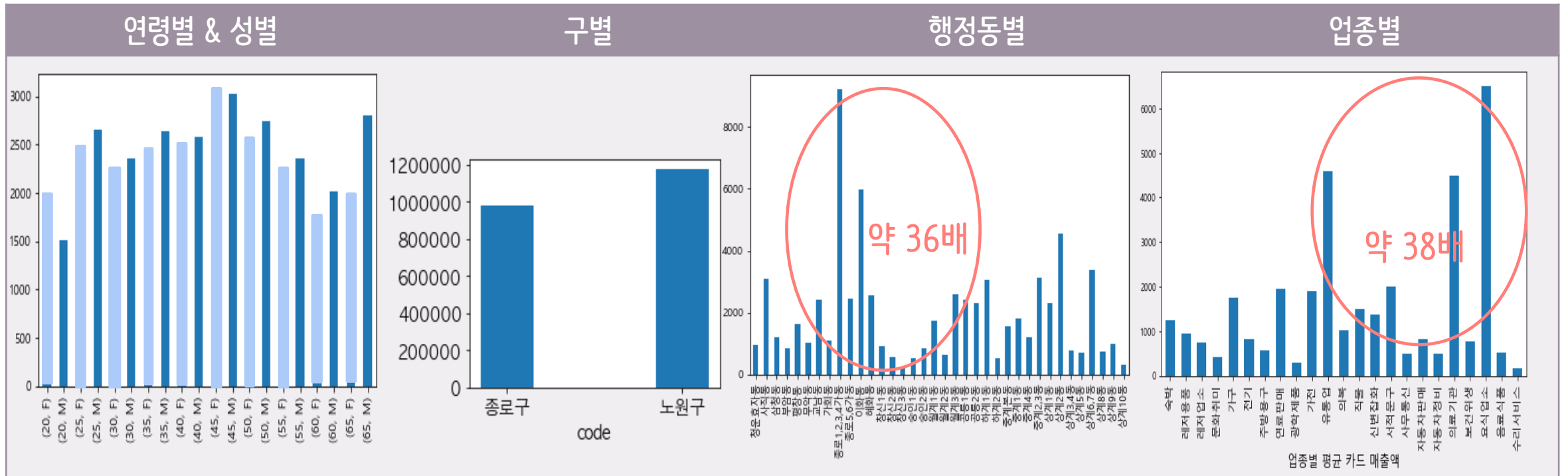
- 겨울, 봄철 (11월~3월) 미세먼지 농도 높은 추세



데이터 전처리 및 EDA:

(3) 신한 카드 데이터

- 특성별 카드 총 매출액 비교: 연령, 성별, 구별 내에서의 매출액보다, 행정동과 업종 내에서의 매출액 차이가 비교적 더 큼
- 총 매출액 가장 많은 동: 종로1,2,3,4가동 (9,128,000원), 가장 적은 동: 창신1동 (256,000원)
- 총 매출액 가장 많은 업종: 요식업 (6,504,000원), 가장 적은 업종: 수리 서비스 (172,000원)



데이터 전처리 &
EDA

시계열 분석



의사결정나무

비즈니스 모델

의사결정나무: 미세먼지+ 카드 데이터 결합

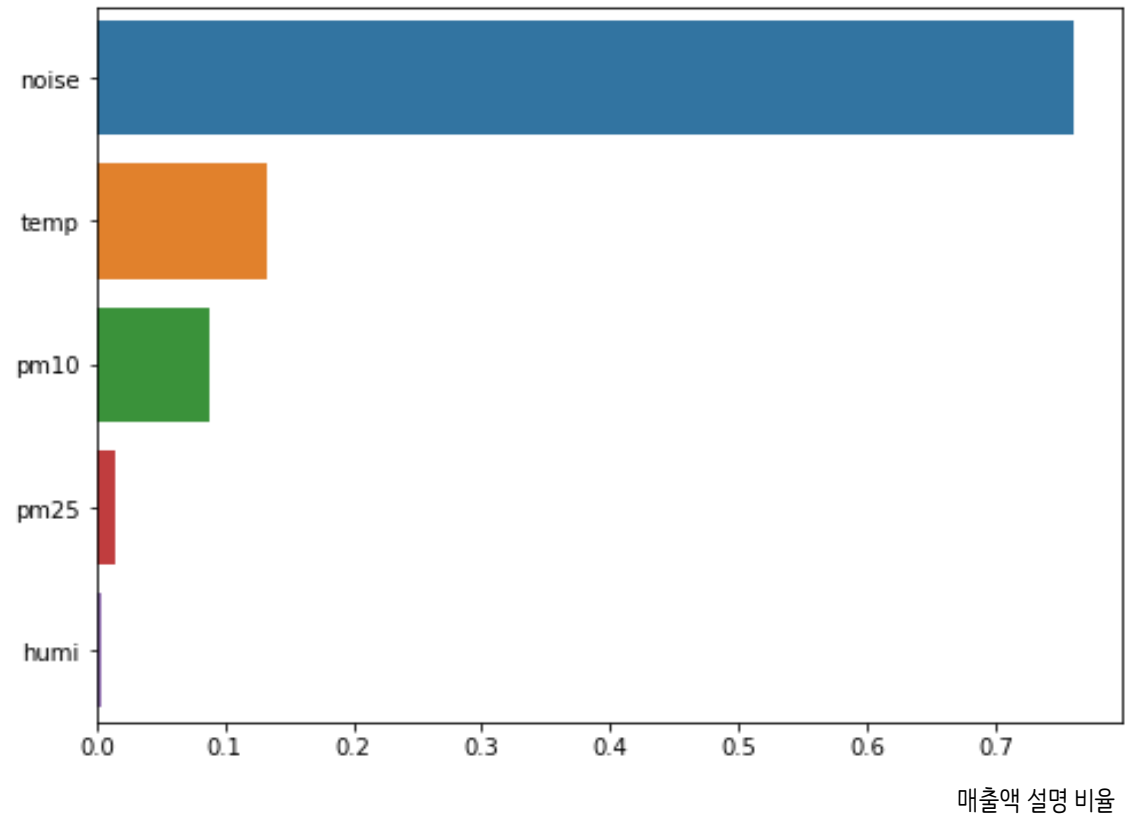
미세먼지 농도가 카드 매출에 주는 영향력을 의사결정 나무로 측정

- 의사결정나무를 통해, 기후변수들과 매출액의 관계에서 유의미한 변수를 파악
- 독립변수를 소음, 기온, 미세먼지 농도, 습도로 하고, 종속변수를 매출액으로 놓고 분석
- 모델 분석 결과 (R-squared)

train accuracy : 0.011

test accuracy : 0.015

- 분석 결과: 실제 미세먼지 농도는 개개인 카드 소비에 직접적인 영향력이 없는 것으로 나타난다



〈의사결정나무〉

데이터 전처리 &
EDA

시계열 분석



의사결정나무

비즈니스 모델

시계열 분석: 실제 미세먼지 농도 VS 일일 SNS 미세먼지 글 수

1. 의사결정 나무를 통해서 미세먼지 농도의 매출액에 대한 영향력이 유의미하지 않음을 확인



2. 보다 정밀한 비교를 위해 시계열 모형 SARIMAX를 이용해, 미세먼지 농도와 일일 SNS미세먼지 글 수를 각각 독립 변수로 설정하여 영향력 비교분석



3. 분석 결과, pm10 미세먼지 농도보다 일일 SNS 미세먼지 글 수가 영향력이 있는 것으로 판단



4. 일일 SNS미세먼지 글 수를 독립변수로 한 최종 모형 결정

〈SARIMAX 미세먼지 계수 비교분석 결과〉

| 업종 코드 | 업종 | 실제 미세먼지 계수 (pm10) | 미세먼지 인식 계수 (일일 미세먼지 글 수) |
|-------|-------|-------------------|--------------------------|
| 10 | 숙박 | 0 | -0.187 |
| 20 | 레저용품 | 0 | 0.1749 |
| 22 | 문화취미 | 0 | -0.162 |
| 62 | 자동차정비 | 0 | 0.2346 |
| 70 | 의료기관 | 0 | 0.471 |



유의하다고 판단

시계열 분석: SNS + 카드 데이터 결합, 모형 소개

시계열 분석을 위해, SARIMAX 모형 사용

- 데이터 결합 과정
 - SNS 데이터와 카드 매출 데이터를 날짜 기준으로 병합 후, 모든 행정구역에서 업종별 일일 매출액의 총 합을 계산
- SNS 글 수와 카드 매출은 시간 첨자가 존재하는 시계열 데이터
 - 매출액은 일주일 단위 주기가 존재하므로, seasonal effect 고려할 필요성 있음
- 시간 변화에서 SNS 미세먼지 글 수가 매출액에 주는 유의미한 관계를 분석하기 위해, SARIMA모형에 독립변수를 추가한 SARIMAX 모형 사용

SARIMAX(p,d,q) (P,D,Q)_S 모형

$$\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D\left(Y_t - \sum_{i=1}^k \beta_i X_{it}\right) = \theta_q(B)\Theta_Q(B^S)\epsilon_t$$

$$\text{where } \theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q,$$

$$\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p,$$

$$\Theta_Q(B^S) = 1 - \Theta_1 B^S - \dots - \Theta_Q B^Q S,$$

$$\Phi_P(B^S) = 1 - \Phi_1 B^S - \dots - \Phi_P B^P S,$$

Y : 매출액
X : 일일 미세먼지 글 수
S : 계절차분 차수
D : 계절차분 횟수
P : 계절 AR 모형의 차수

Q : 계절 MA 모형의 차수
d : 비계절 차분 횟수
p : 비계절 AR모형의 차수
q : 비계절 MA모형의 차수
B : 후진 연산자

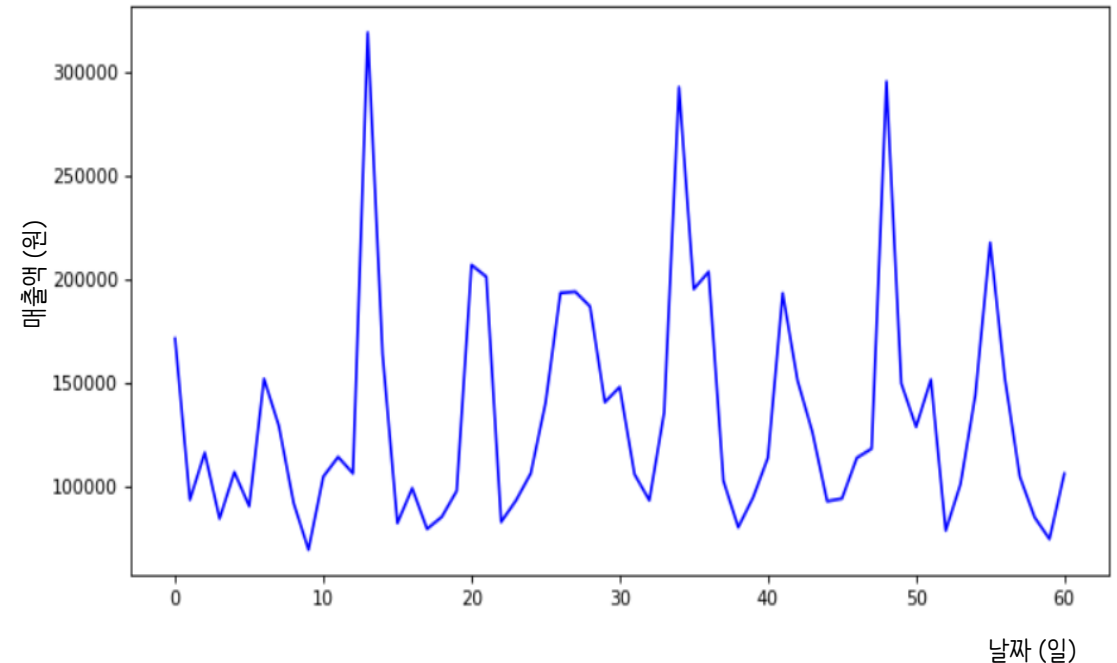
시계열 분석: 분석 데이터 예시 -숙박

예시 : 숙박 업종의 분석 데이터

| 날짜 | 업종 | 매출액 | 미세먼지 인식 계수 (일일 미세먼지 글 수) |
|----------|---------|--------|--------------------------------|
| 18.04.01 | 10 (숙박) | 171130 | 2419 |
| 18.04.02 | 10 (숙박) | 93369 | 2811 |
| 18.04.03 | 10 (숙박) | 116075 | 2762 |
| 18.04.04 | 10 (숙박) | 84254 | 2799 |
| 18.04.05 | 10 (숙박) | 106643 | 1911 |

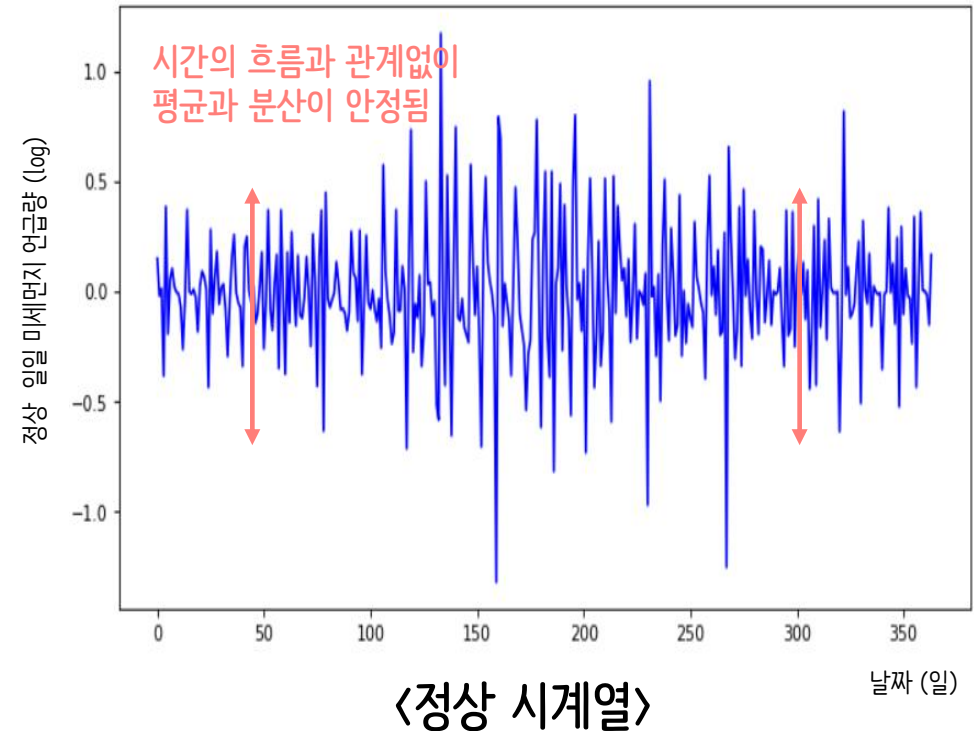
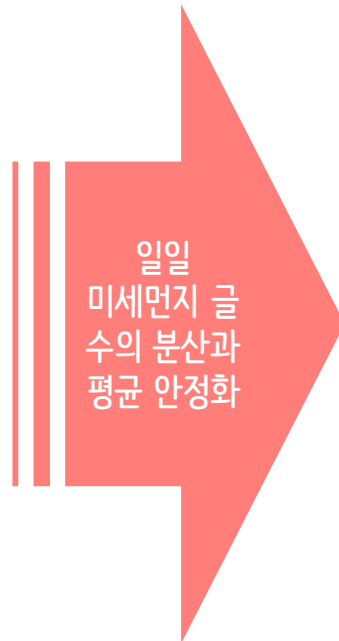
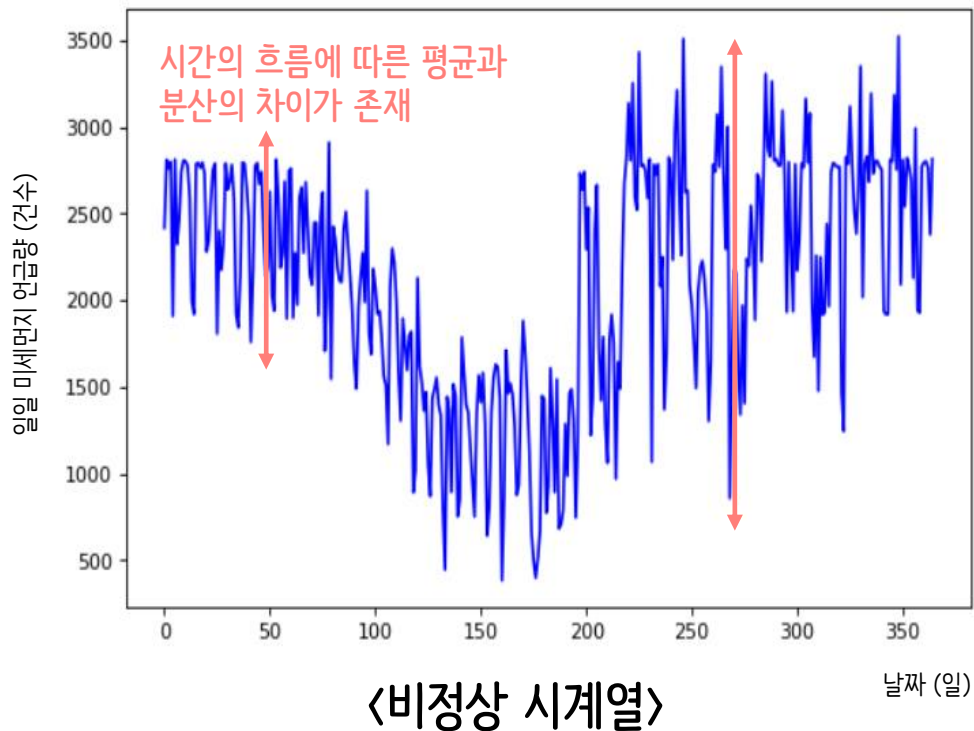
예시 : 숙박 업종의 단기 매출액 그래프

-> 주말마다 매출액이 급격하게 증가함

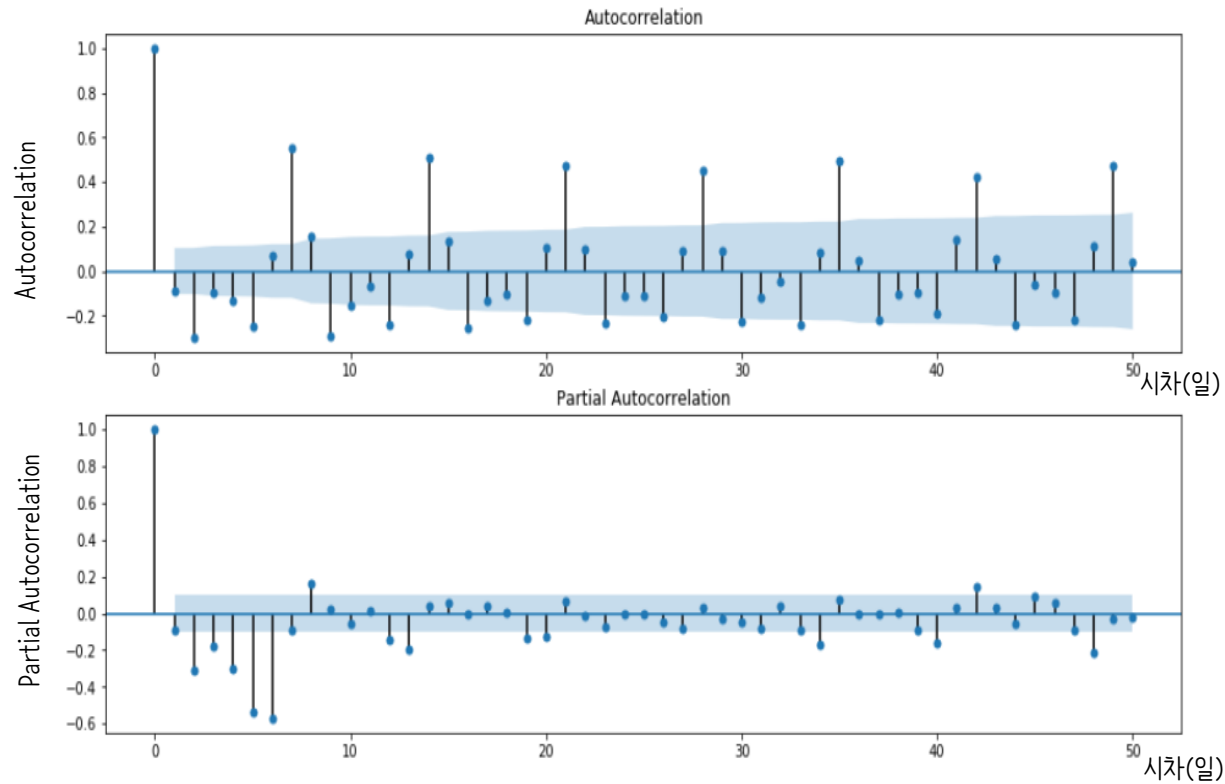


시계열 분석: SARIMAX 모형 사용

- 모든 변수가 정상 시계열이라는 조건들을 (평균 & 분산 일정) 만족해야, SARIMAX 모형을 시계열 분석으로 적용 가능
- 시간의 흐름에 따라서, <비정상 시계열> 그래프에서 허구적 회귀가 발생 가능성 ↑
- 불안정 시계열의 정상화를 위해서, 분산 안정화 변환(로그)과 평균 안정화 변환(차분) 실시



시계열 분석: 모형 차수 탐색



〈숙박 매출액 acf, pacf 결과〉

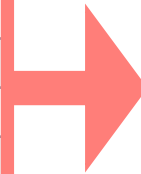
SARIMA 모형 차수 탐색 과정

- 업종별 매출액 변수도 같은 과정을 통해 정상화 시킨 후 모형의 차수 탐색
- 모형의 차수는 acf, pacf 도표를 참고하여 AIC, BIC를 기준으로 선택
- acf 가 7을 주기로 높은 값을 가짐
→ 주기 7의 계절 차분 실시(D=7)
- acf는 첫번째 시차와 두번째 시차에서 유의한 값을 갖고, pacf는 6번째 시차까지 유의한 값을 가지므로 비계절적 차수도 ARMA(6,2)까지 고려해 볼 수 있음

시계열 분석: 모형 적합, 미세먼지 민감 업종 도출

미세먼지 계수 절대값이 높은 기준으로 민감 업종 선별

| 업종 코드 | 대분류 업종 | order | seasonal order | 미세먼지 계수 |
|-------|--------|---------|----------------|---------|
| 92 | 수리서비스 | (1,1,1) | (0,1,1,7) | 0.5091 |
| 70 | 의료기관 | (1,1,1) | (0,1,1,7) | 0.471 |
| 62 | 자동차정비 | (1,1,1) | (0,1,1,7) | 0.2346 |
| 10 | 숙박 | (1,1,1) | (3,1,0,7) | -0.187 |
| 20 | 레저용품 | (1,1,1) | (0,1,1,7) | 0.1749 |
| 22 | 문화취미 | (1,1,2) | (1,1,0,7) | -0.162 |
| 71 | 보건위생 | (1,1,1) | (0,1,1,7) | 0.1513 |
| 81 | 음료식품 | (0,1,1) | (0,1,1,7) | 0.1457 |
| 21 | 레저업소 | (1,1,1) | (1,1,2,7) | 0.0791 |
| 33 | 연료판매 | (1,1,1) | (1,1,1,7) | 0.0328 |
| 40 | 유통업 | (1,1,1) | (1,1,1,7) | 0 |
| 42 | 의복 | (1,1,1) | (0,1,1,7) | 0 |
| 50 | 서적문구 | (1,1,1) | (1,1,1,7) | 0 |
| 80 | 요식업소 | (1,1,1) | (1,1,1,7) | 0 |



- 미세먼지 변화에 따라 민감하게 매출 변화를 보이는 대분류 업종 8개 그리고 업종 별 매출액이 큰 행정동을 순위화 하여 14개 행정동 선별

| 업종 | 선정된 미세먼지 민감 행정동 |
|--------|--|
| 수리 서비스 | <ul style="list-style-type: none"> 종로1.2.3.4가동 |
| 의료기관 | <ul style="list-style-type: none"> 중계1동 |
| 자동차정비 | <ul style="list-style-type: none"> 공릉1동 |
| 숙박 | <ul style="list-style-type: none"> 상계1동 상계2동 |
| 레저용품 | <ul style="list-style-type: none"> 이화동 상계6.7동 |
| 문화취미 | <ul style="list-style-type: none"> 교남동 |
| 보건위생 | <ul style="list-style-type: none"> 종로5.6가동 사직동 가회동 혜화동 |
| 음료식품 | <ul style="list-style-type: none"> 중계2.3동 공릉2동 |

데이터 전처리 &
EDA

시계열 분석

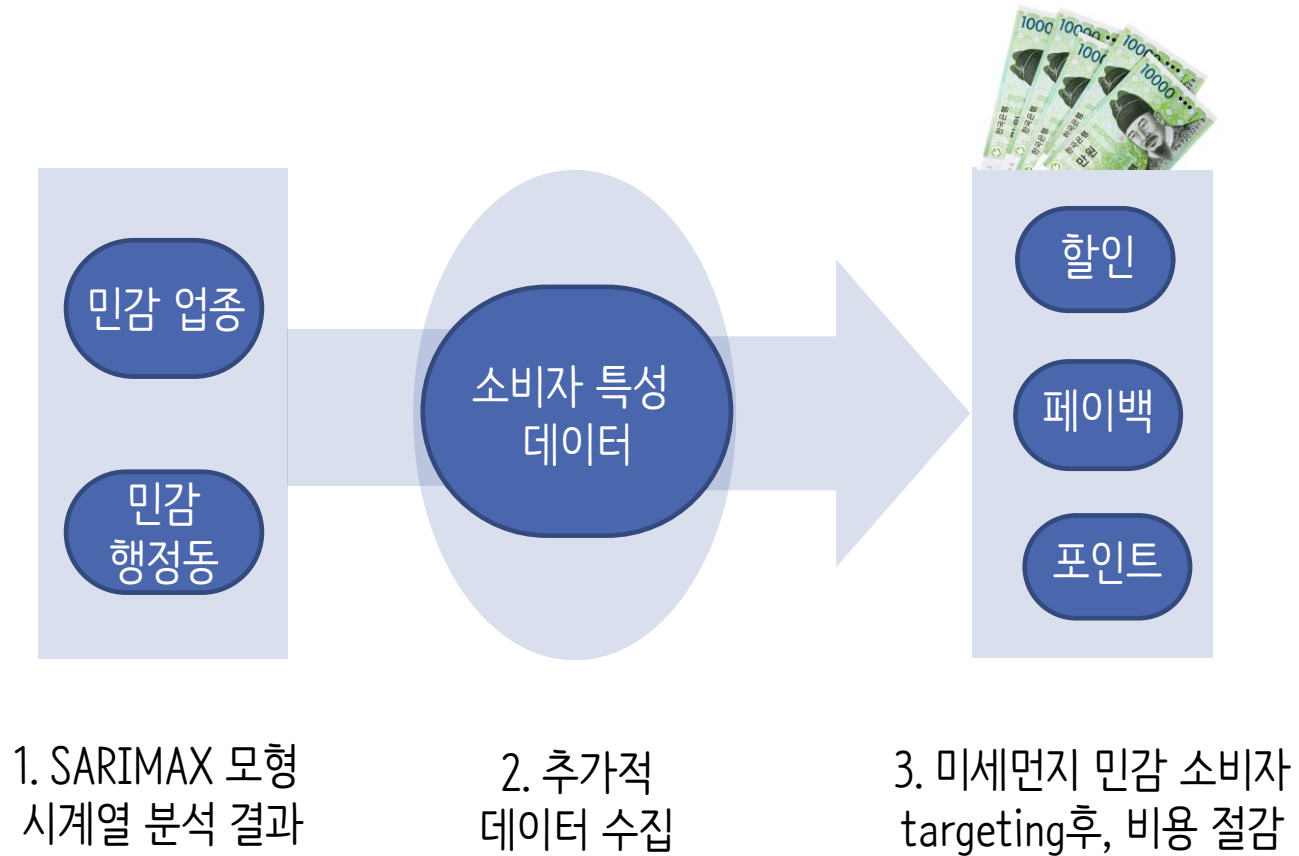


의사결정 나무

비즈니스 모델

비즈니스 모델:

(1) 신한 카드 기존 고객 데이터 활용



기존 고객 데이터를 활용해 초개인화 마케팅 전략 실행

1. 신한 카드 데이터 분석 결과
 - 실제 미세먼지 수치보다 미세먼지 인식이 매출에 중요
 - 매출에 유의한 미세먼지 민감 업종과 행정동 선정
2. 신한 카드 마케팅 현황
 - 카드수수료 인하 정책으로 마케팅 비용을 줄이는 추세
 - 초개인화 카드 출시 등 개인화 마케팅 주력
3. 미세먼지 민감 소비자 targeting
 - 미세먼지 민감 반응 고객 데이터 (민감군 vs 둔감군) 분류
 - 미세먼지에 따라 카드 혜택 할인/ 페이백/ 포인트 적립

비즈니스 모델:

(1) 미세먼지 민감 소비자 TARGETING

고비용 마케팅 구조 개선 및 비용 절감

- 업종별 매출 반응이 중요한 요소
- 미세먼지 민감군 대상 관련 마케팅 메시지 전달
- 미세먼지에 둔감하게 반응하는 고객은 평소와 비슷하게 마케팅함으로써 기회비용 절감
- 예시) 민감업종 수리서비스의 세탁소의 경우, 미세먼지 언급량이 많은 기간에 민감군에 일정 금액 할인하거나 쿠폰 지급.



새로운 비즈니스 모델 제시

- SNS 및 기사에 미세먼지 언급량이 많은 기간에 미세먼지에 매출이 민감한 상품들 프로모션
- 미세먼지 마케팅 반응에 따라 민감 고객군을 수정할 수 있는 시스템을 구축하여 지속 가능 마케팅 실현



비즈니스 모델:

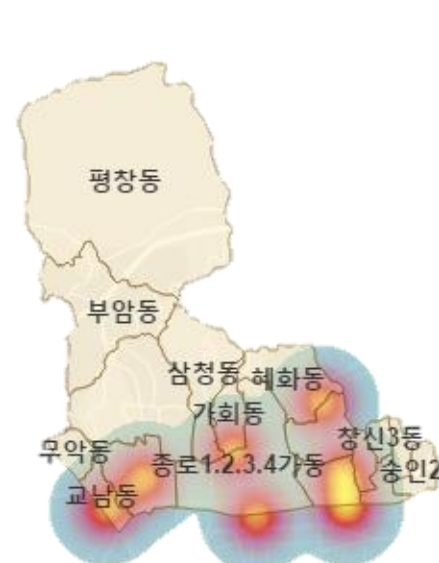
(2) 신한 카드의 미세먼지 정화 벤치

미세먼지 정화 벤치로 공유가치 창출

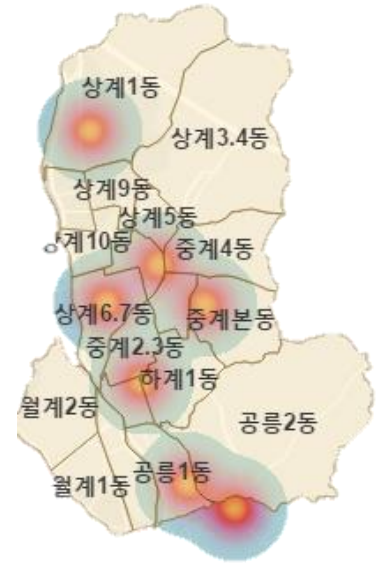


- 미세먼지에 민감한 소비자층을 공략하기 위해 “미세먼지 민감 행정동” 지역 위주로 미세먼지를 정화해 줄 수 있는 벤치 설치
- 공유가치창출(creating shared value; CSV) 의의: 신한카드의 환경 지킴이 역할을 확장하여 새로운 환경 관련 사회공헌활동 시행
- 미세먼지 관련 CSV 활동을 선도함으로써, 자사 홍보 효과 극대화
- 미세먼지 정화 벤치 활용 방안:
 1. 시민 참여형 정화 벤치: 신한 앱 결제 및 카드를 통해 시민들이 직접 기부해, 미세먼지 민감 행정구역에 새로운 미세 벤치 설치
 2. 광고판 QR코드를 운영하여 주변 미세먼지 관련 제품들을 정화 벤치를 통해서만 특별할인 쿠폰을 발급할 수 있도록 제휴

미세먼지 민감 행정구역에 정화 벤치 설치



〈종로구 민감 행정동〉



〈노원구 민감 행정동〉

추가적 의의 및 한계점

추가적 의의

- 실제 미세먼지 수치가 높은 것보다, 미세먼지가 더 많이 언급될수록, 사람들의 인식이 강해지고 더 민감하게 반응한다
- 미세먼지 수치 자체를 줄이는 것도 중요하지만, 미세먼지에 대한 시민들의 불안감을 줄이는 것도 매우 중요
 - 미세먼지 관련 언론 보도 및 사회적 언급 영향력 크다
- 카드 매출 업종별로 미세먼지 언급량에 대해 민감하게 반응하는 정도가 상이하다
 - 미세먼지에 민감한 업종을 판별해 적절한 마케팅 가능

한계점

- 카드 데이터의 내생적 한계점
 - 종로구와 노원구만의 분석 결과를 일반화하기 어려움
 - 업종별 대분류 데이터만 제공되어, 세부 업종에 대한 미세먼지의 직접적인 영향은 파악할 수 없음
 - 인구 통계학적 데이터만 제시. 소비 지수 같은 미시적 변수가 제공되지 않아서, 소비자들과 연계해서 비즈니스모델을 설계하는 데 있어 한계점 존재

참고 문헌

- 현대경제연구원 2019년 3월
- 통계청,〈데이터마이닝 의사결정나무의 응용〉, 최종후, 통계분석연구 제4권 제1호(99.봄),1999, pp61-83
- "퇴근 후 유흥소비.. 광화문,삼성역 초저녁 유출인구 많아",매일경제,
- <https://www.mk.co.kr/news/special-edition/view/2012/03/175468/>
- 한국정보화진흥원, 〈2015 서울 유동인구 조사보고서〉, 2015.12
- '미세먼지가 바꾼 소비행태 변화 KEB' - 2019년 4월 하나은행 하나금융경영연구소
- 'Solar radiation forecasting by time series models' -2018년 12월 Yu Min Suha · Heung-goo Sonb · Sahm Kim
- “'사랑의 온도'100도 넘었다”,충북일보

<https://www.inews365.com/news/article.html?no=216000>

참고 문헌

- “신한카드, 고객 위한 맞춤형 초개인화 카드 선보여“, 백세시대
<http://www.100ssd.co.kr/news/articleView.html?idxno=62658>
- “허리띠 졸라 겨우 선방한 카드사, 하반기도 ' 첩첩산중'”, 조선비즈,
<https://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=101&oid=366&aid=0000439616>
- 최악의 미세먼지에 “영화관·멀티플렉스 등 신용카드 사용 급증”, 한겨레
http://news.khan.co.kr/kh_news/khan_art_view.html?art_id=201903071041001
- 미세먼지 뉴스 많은 날 , 리조트 카드매출 ‘울상’, 세탁소 ‘활짝’
<http://www.etnews.com/20190417000170>
- konlpy api
<https://konlpy-ko.readthedocs.io/ko/v0.4.3/>