

미 세 먼 지
비 즈 니 스 아 이 디 어

빅콘테스트 공모전

-Innovation분야-

데이터로 만드는 세상

박병규 김지현 김영진 김민호 김봉영

Contents

Unit 01 | 분석 배경 및 목적

Unit 02 | 분석 과정

Unit 03 | 분석 결과

Unit 04 | 결론 및 한계/시사점

미세먼지 위험에 대한 수용자의 인식과 의견 형성에 관한 연구
: 프로모션 뉴스 인지정도에 따른 분석

김영옥 이현승 이혜진 장유진

미세먼지 위험에 대한 인식과 소셜미디어 의견이
미세먼지에 대한 대응행동에 미치는 영향

연세대학교 커뮤니케이션대학원
언론학 전공
김 민 정

미세먼지 위험에 대한 수용자의 인식과 의견
형성에 관한 연구

프로모션 기사 인식정도와 관여도에 따른 분석*

김영옥 이화여자대학교 커뮤니케이션미디어학부 교수**
이현승 이화여자대학교 커뮤니케이션미디어학부 박사과정***
이혜진 이화여자대학교 커뮤니케이션미디어학부 석사과정****
장유진 이화여자대학교 커뮤니케이션미디어학부 석사과정****

미세먼지 위험에 대한 인식과
인간의 사고와 행동에 영향이 있음.

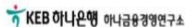
미세먼지가 바꾼 소비 행태 변화

2019년 4월



2019년 하나금융 경영 연구소에서
국내 미세먼지와
소비행태를 파악하는 연구를 진행함.

01 분석 배경



결론 및 시사점

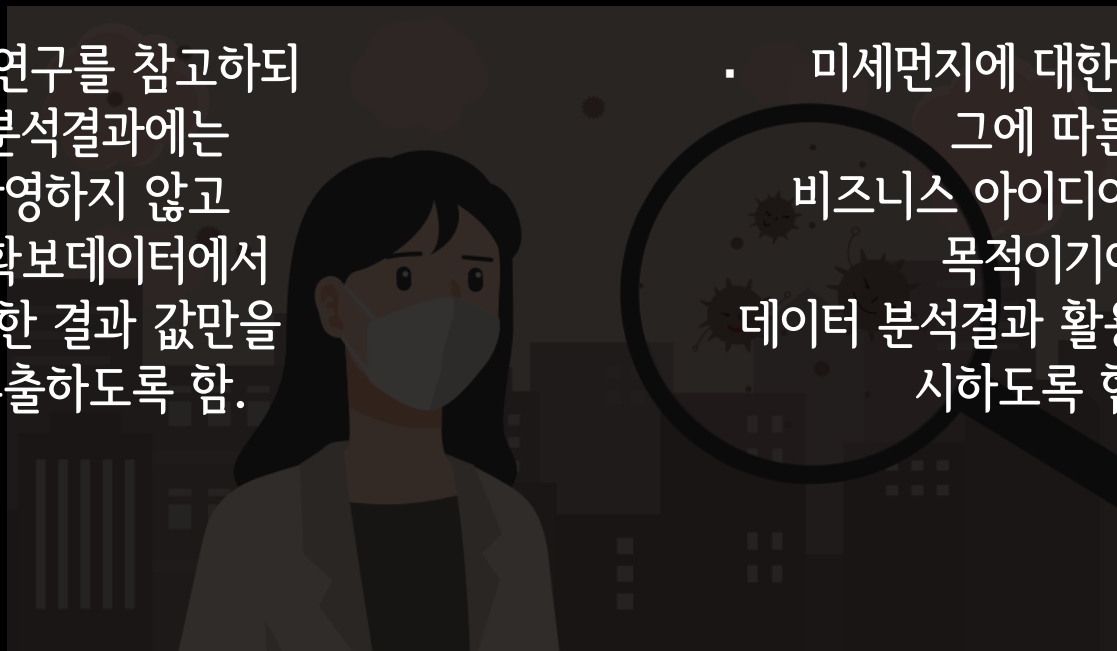
- 국내 미세먼지 농도는 세계 약 70개국 중 상위 26위이며, OECD 회원국 중 4번째로 높은 수준
- 단, 연도별 국내 미세먼지 농도는 지난 20여년 간 꾸준히 감소하고 있는 추세
- 미세먼지에 대한 한국인의 인식과 불안감은 최근 크게 증가하고 있으며, 2013년부터 미세먼지 관련 뉴스량이 급증한 것이 주요 원인 중 하나로 판단됨
- 미세먼지 관련 뉴스에서 가장 언급이 많이 된 연관 키워드는 “중국”이며, 대체로 보수 성향 언론사에서 그 정도가 더 강했음. 반면 진보 성향 언론사는 “경유차”, “석탄”이 상대적으로 많아, 미세먼지의 원인을 바라보는 시각에 다소 차이가 있는 것으로 나타남
- 실제 미세먼지 농도와 카드 소비의 상관 관계보다, 미세먼지 관련 뉴스량과 카드 소비의 상관 관계가 더 높아, 소비자는 미세먼지 관련 뉴스를 통해 미세먼지량을 인식하고, 이것이 소비 행동에 영향을 주고 있음을 확인함
- 미세먼지 관련 뉴스가 증가하였을 때, 대형마트, 수퍼마켓, 주유소, 차량 정비, 레저, 놀이공원, 호텔, 음식점 등, 외출이 필요한 분야의 소비가 감소하는 사실을 밝힘
- 반면 인터넷 쇼핑, 홈쇼핑, 신차 구매, 이비인후과, 소아과, 세탁, 화초/와인, 침구류, 사우나 등은 미세먼지 관련 뉴스가 증가함에 따라 매출이 증가하는 현상을 확인함

결론 :

1. 국내 미세먼지농도는 줄어들고 있음.
- 2.국민들의 불안감은 증가하고 있으며, 뉴스량을 주요원인 중 하나로 판단
- 3.미세먼지 농도 보다 미세먼지 관련 뉴스량이 카드소비와 상관관계가 더 높음.

분 석 목 적

- 선행 연구를 참고하되
분석결과에는
반영하지 않고
현 확보데이터에서
분석한 결과 값만을
추출하도록 함.
- 미세먼지에 대한 영향과
그에 따른
비즈니스 아이디어 창출이
목적이기에
데이터 분석결과 활용방안을 제
시하도록 함.

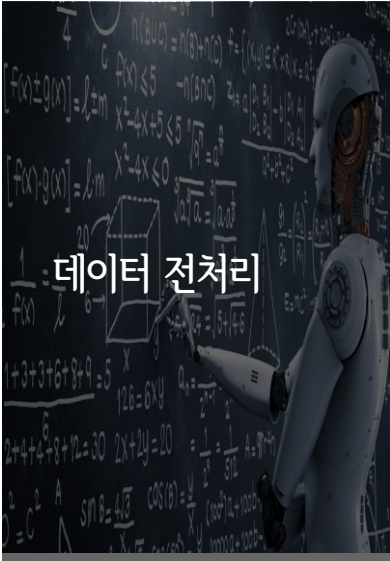


02 분석 과정




데이터 분할

A pair of hands is shown holding four blue puzzle pieces, illustrating the process of data partitioning.



데이터 전처리

A white robot is pointing at a chalkboard filled with mathematical equations and formulas, representing data preprocessing.



데이터 분석

A magnifying glass is focused on a bar chart with four bars of different colors (purple, orange, blue, red) on a grid, symbolizing data analysis.

<데이터 분할>

각각의 데이터

GS리테일(유통)

K-weather(기상)

SK 텔레콤(유동인구)

와이즈넷(SNS)

신한카드(카드)



통일된 데이터

행정동별

날짜를 기준으로

CSV파일로

변환

분석 시
데이터 끌어오기/
병합 쉬움!!

02 분석 과정

kweather_save4 > jongro

이름

- buarm
- changsin1
- changsin2
- changsin3
- cungwun
- gahee
- gyonam
- hyehwa
- jongro56
- jongro1234
- leehwa
- muack
- pungchang
- sajic
- samchung
- sungin1
- sungin2

각 데이터별 파일폴더 생성

같은 행정동 이름의
Csv파일 생성

gs_save > jongro

이름

- buarm
- changsin1
- changsin2
- changsin3
- cungwun
- gahee
- gyonam
- hyehwa
- jongro56
- jongro1234
- leehwa
- muack
- pungchang
- sajic
- samchung
- sungin1
- sungin2

데이터 전처리

GS리테일

데이터가 없는
값은
0으로 치환

K-weather

NAN값,
9999는
0으로
치환

SNS

TITLE에
미세먼지 단어만
크롤링해서
최대한 광고제거

카드데이터 활용한 군집형성

평일 데이터 = 0
주말 데이터 = 1

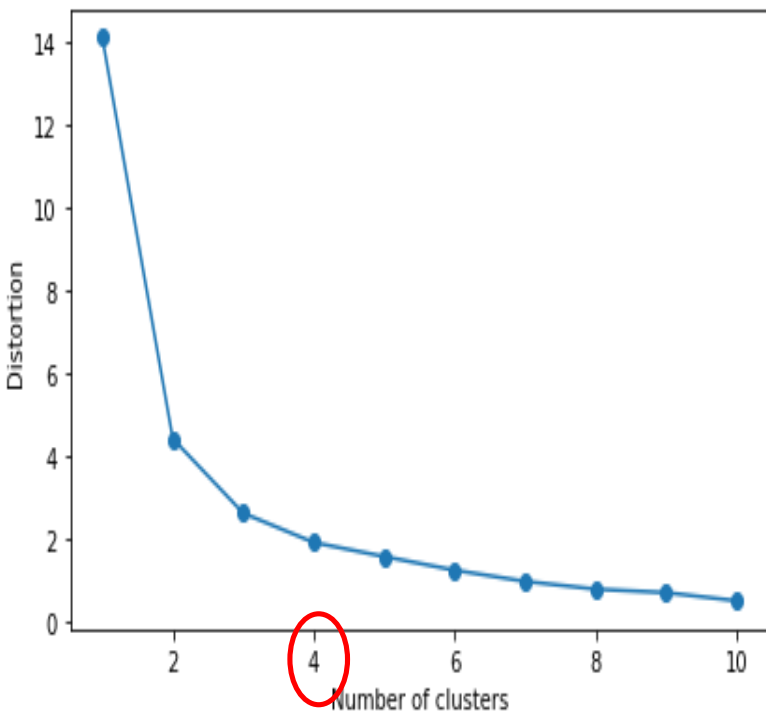
평일 \Rightarrow 주말
0 \Rightarrow 1

카드 사용 변화
카드 횟수 변화

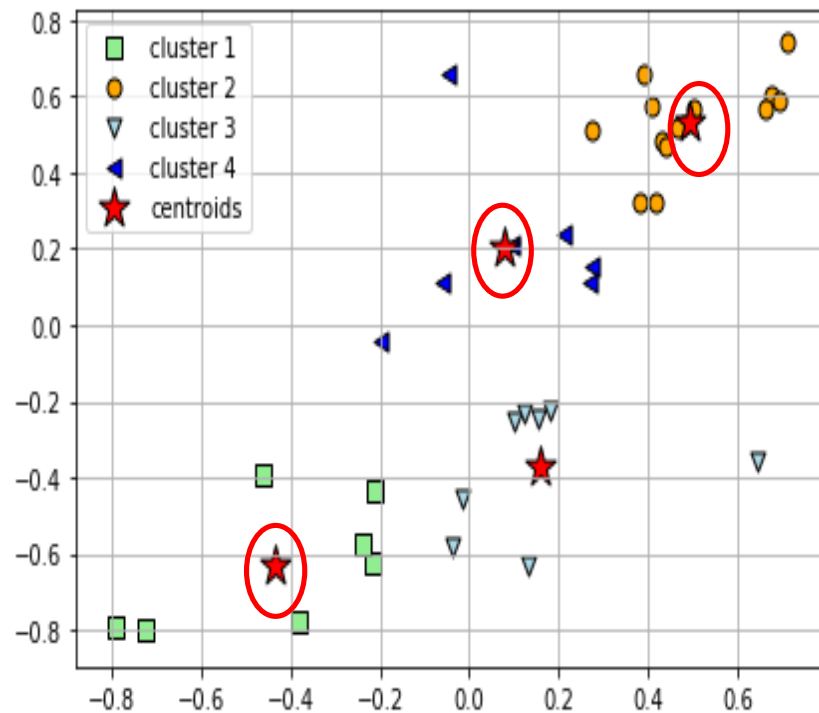
X = 주말로 변화했을 때 카드 사용 변화 상관관계
Y = 주말로 변화했을 때 카드 횟수 변화 상관관계

카드데이터 활용한 군집형성

최대 군집 개수 파악



K-means 알고리즘 활용한 군집분류



군집 별 특징

Cluster 1

평일 -> 주말

금액 ↓

빈도 ↓

Cluster 2

평일 -> 주말

금액 ↑

빈도 ↑

Cluster 3

평일 -> 주말

금액 ↑

빈도 ↓

Cluster 4

평일 -> 주말

금액 변화 없음

빈도 변화 없음

군집 별 특징

Cluster 1

병원과 관련된
매출액이
(의료부문)

가장 높은 지역

Cluster 2

요식업과
유통업의
매출금액이
높은 지역들

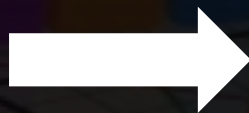
Cluster 3

요식업과
유통업의
매출금액이
높은 지역들

Cluster 4

주로
상가보다는
주거형태의
지역들

카드 데이터를 바탕으로
군집화 실시
=
소비행태가 비슷한
지역



따라서
GS리테일
소비패턴을
파악하는데
군집을 활용하기 유용함.

GS데이터 분석

GS 미세먼지 요인 PCA 차원축소

```

from sklearn.decomposition import PCA

for i in range(0, len(jongro_list)):
    pca = PCA(n_components=1)
    X = jongro_list[i][['pm10', 'temp', 'humi', 'pm25']].fillna(0).values
    W = pca.fit_transform(X)
    jongro_list[i]['PCA_기상'] = W

for i in range(0, len(jongro_list)):
    pca = PCA(n_components=1)
    X = jongro_list[i][['pm10', 'pm25']].fillna(0).values
    W = pca.fit_transform(X)
    jongro_list[i]['PCA_미세먼지'] = W

for i in range(0, len(jongro_list)):
    pca = PCA(n_components=1)
    X = jongro_list[i][['PCA_기상', 'total_counts']].fillna(0).values
    W = pca.fit_transform(X)
    jongro_list[i]['PCA_기상_SNS'] = W
  
```

1 차원으로 축소

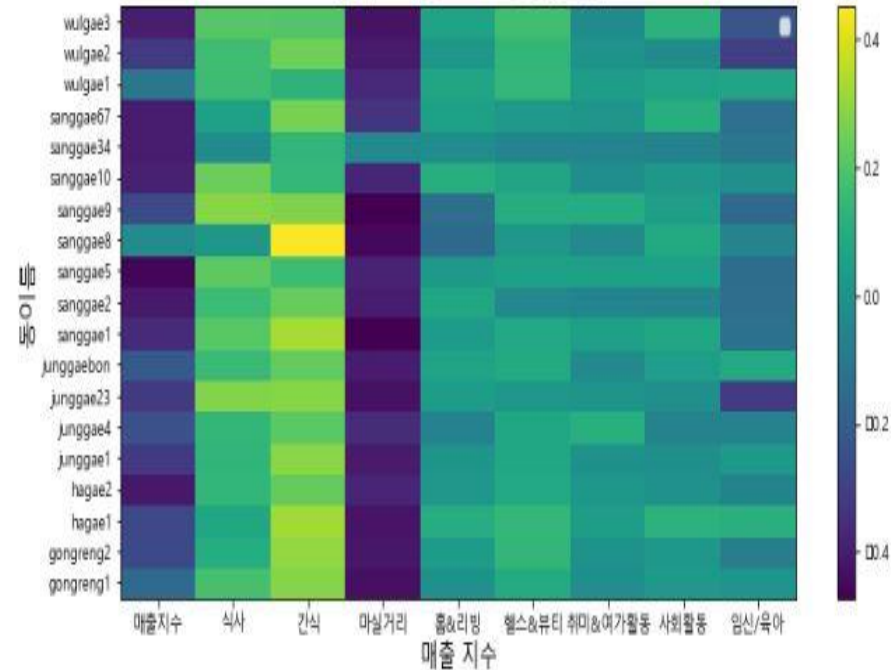
미세먼지 요인으로
GS 데이터를 2차원으로 나타내기
위해
PCA로 독립변수 생성

(온도+미세먼지 = 기상)
Only 미세먼지 = 미세먼지
기상+ SNS데이터 = 기상_SNS)

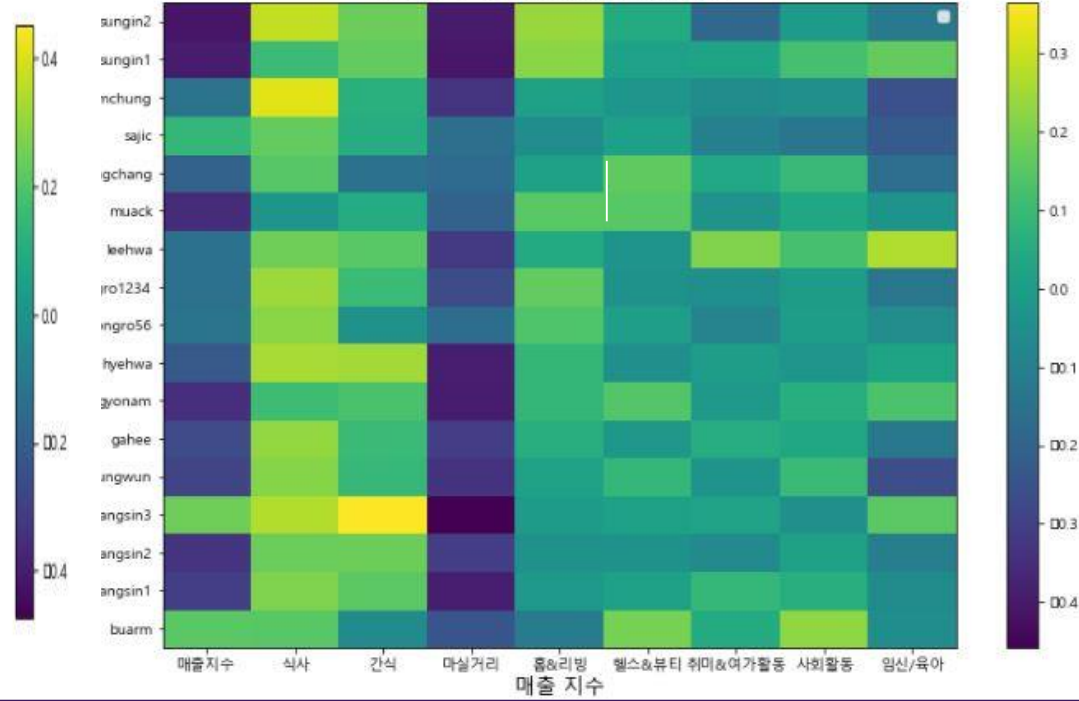
GS데이터 분석

기상+SNS로 축소한 값의 상관관계가 가장 잘 나타남.

노원구 기상+sns 상관관계



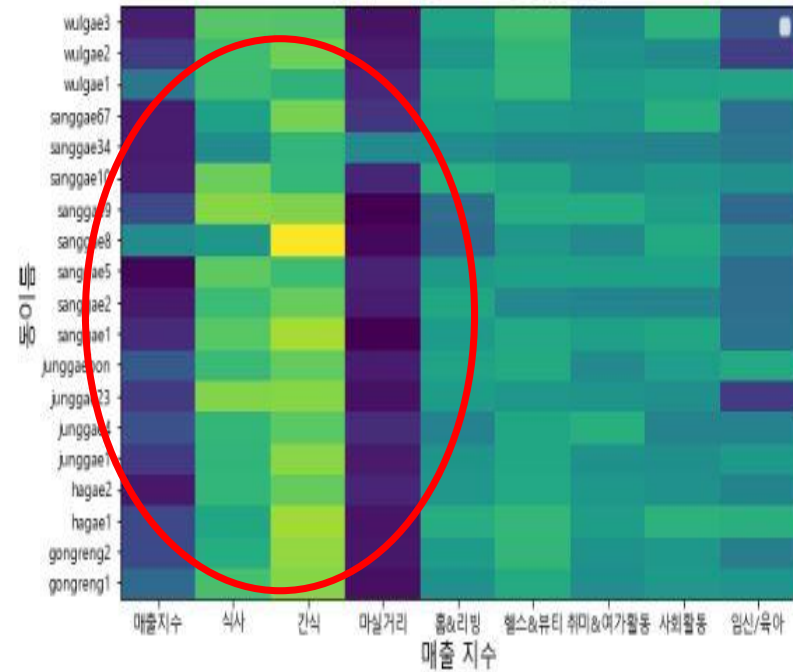
종로구 기상+sns 상관관계



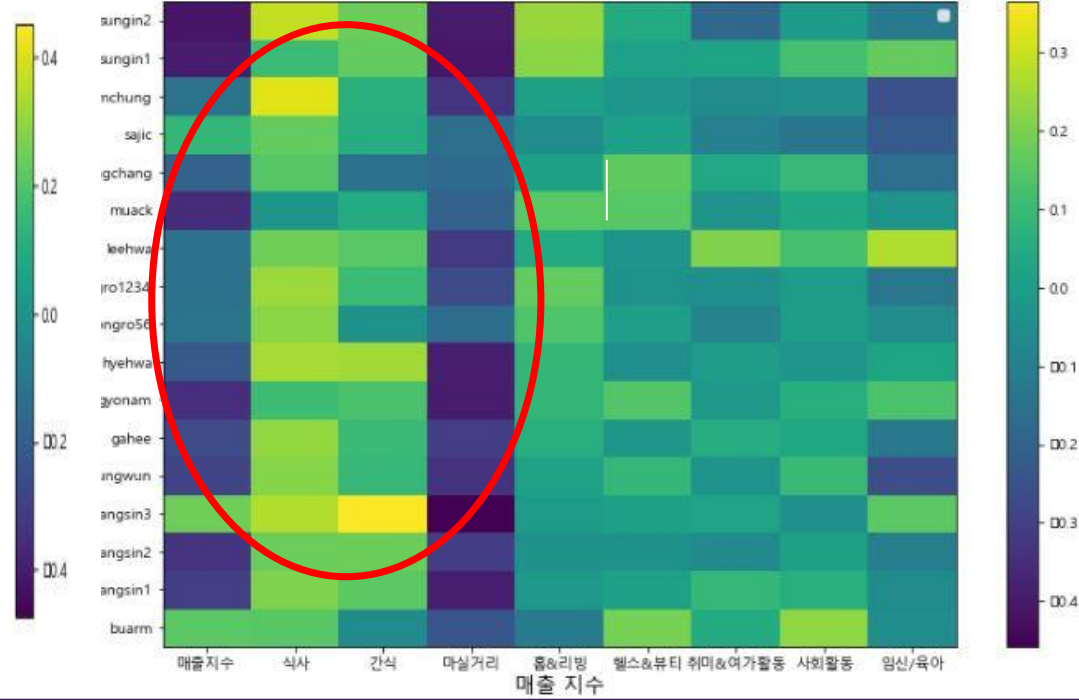
GS데이터 분석

매출/식사/간식/마실거리에서 강한 상관관계

노원구 기상+sns 상관관계



종로구 기상+sns 상관관계



비교분석을 위한 설계

미세먼지(실제 기상 + SNS)
의 변수가

GS 매출에 영향을 줄을 확인.

따라서
미세먼지 영향이 많은 날 데이터
(PCA 값 상위 20%)

-

미세먼지 영향이 적은 날 데이터
(PCA 값 하위 20%)
비교

미세먼지 영향이 많은 날 데이터
(PCA 값 상위 20%)

-

미세먼지 영향이 적은 날 데이터
(PCA 값 하위 20%)

=

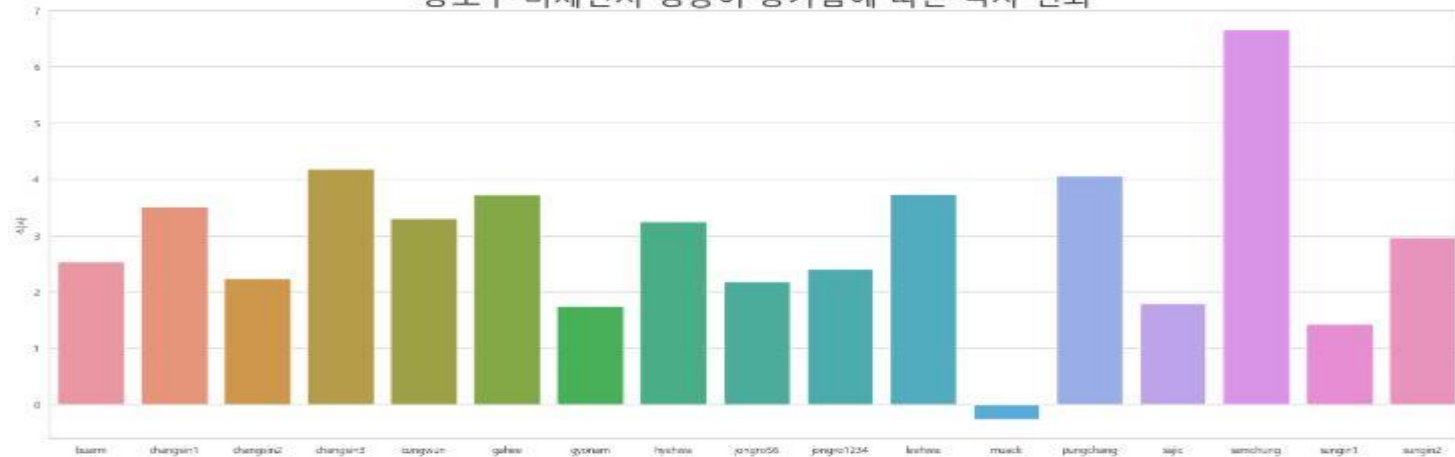
- 값이 나온 경우:
매출이 하락했음.

+ 값이 나온 경우:
매출이 상승했음.

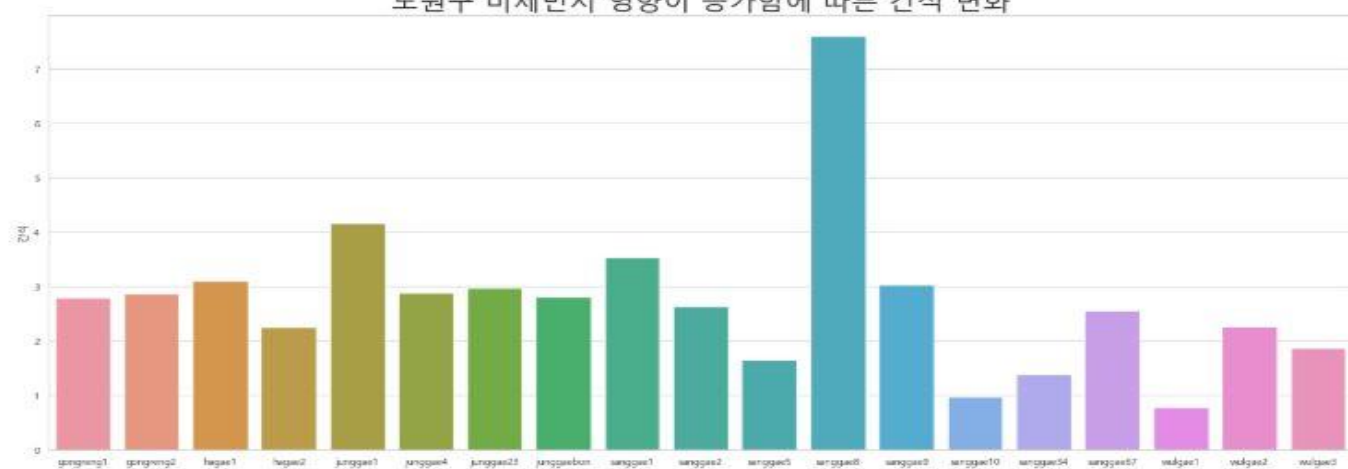
종로구 미세먼지 영향이 증가함에 따른 간식 변화



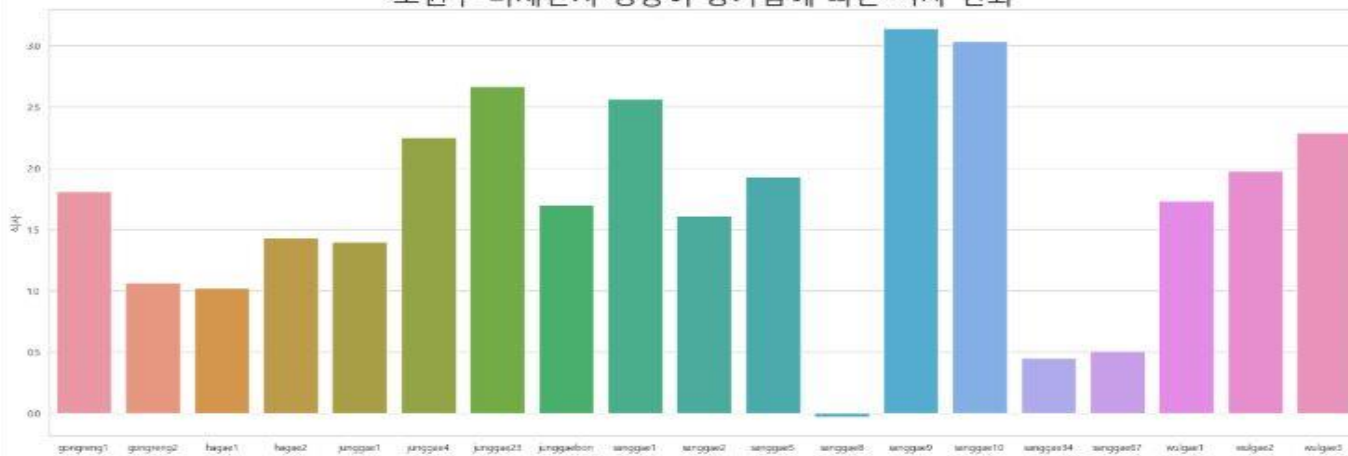
종로구 미세먼지 영향이 증가함에 따른 식사 변화

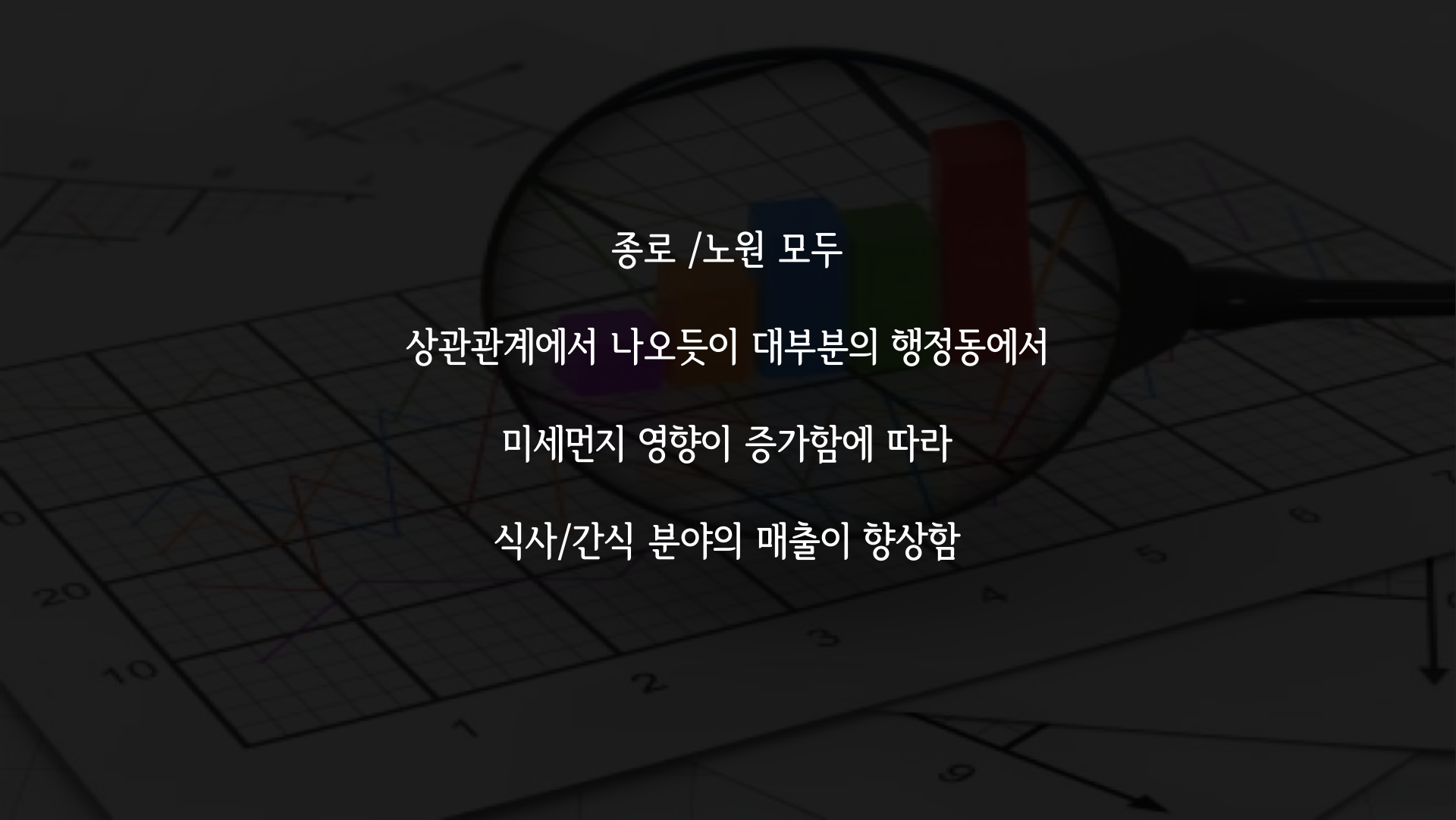


노원구 미세먼지 영향이 증가함에 따른 간식 변화



노원구 미세먼지 영향이 증가함에 따른 식사 변화





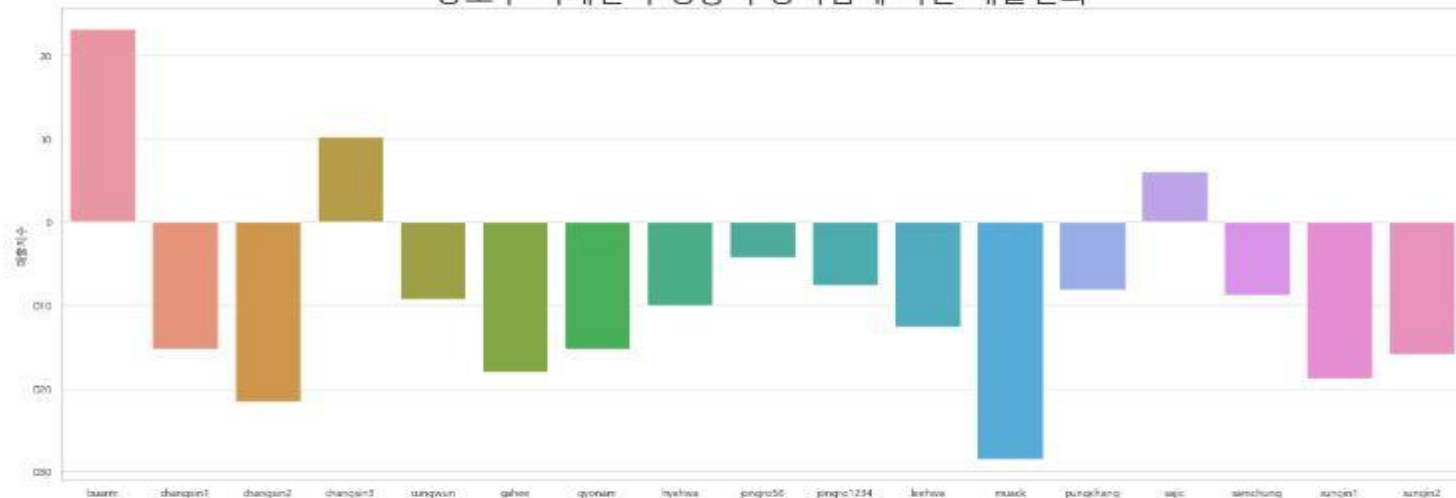
종로 /노원 모두

상관관계에서 나오듯이 대부분의 행정동에서

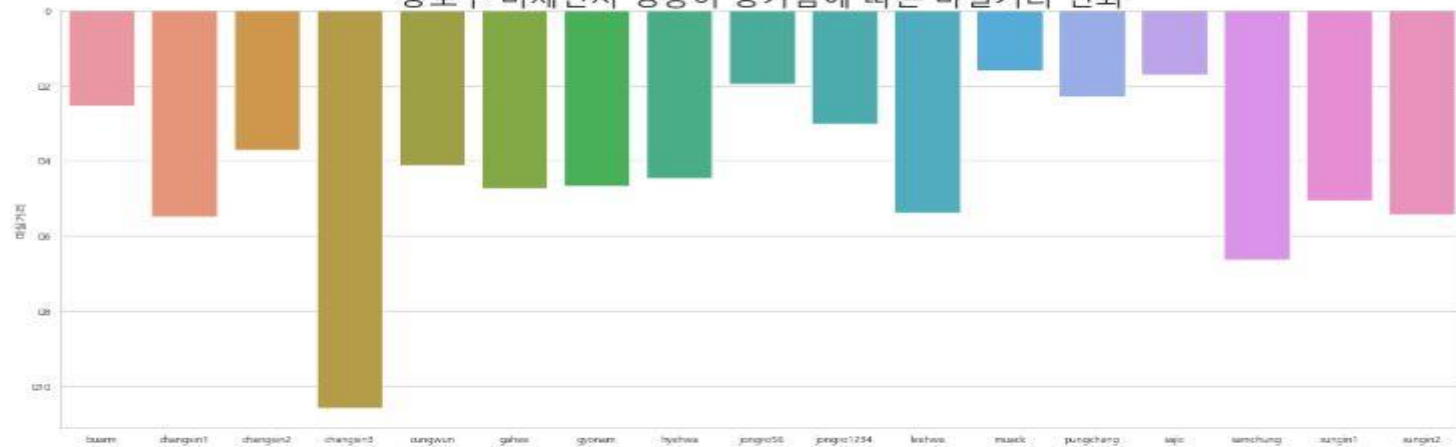
미세먼지 영향이 증가함에 따라

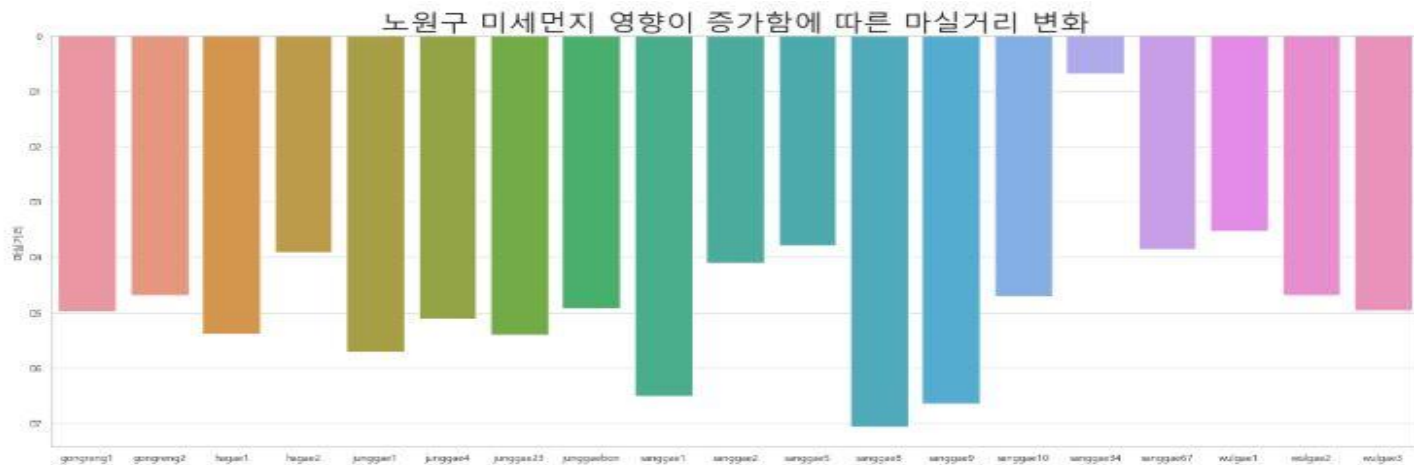
식사/간식 분야의 매출이 향상함

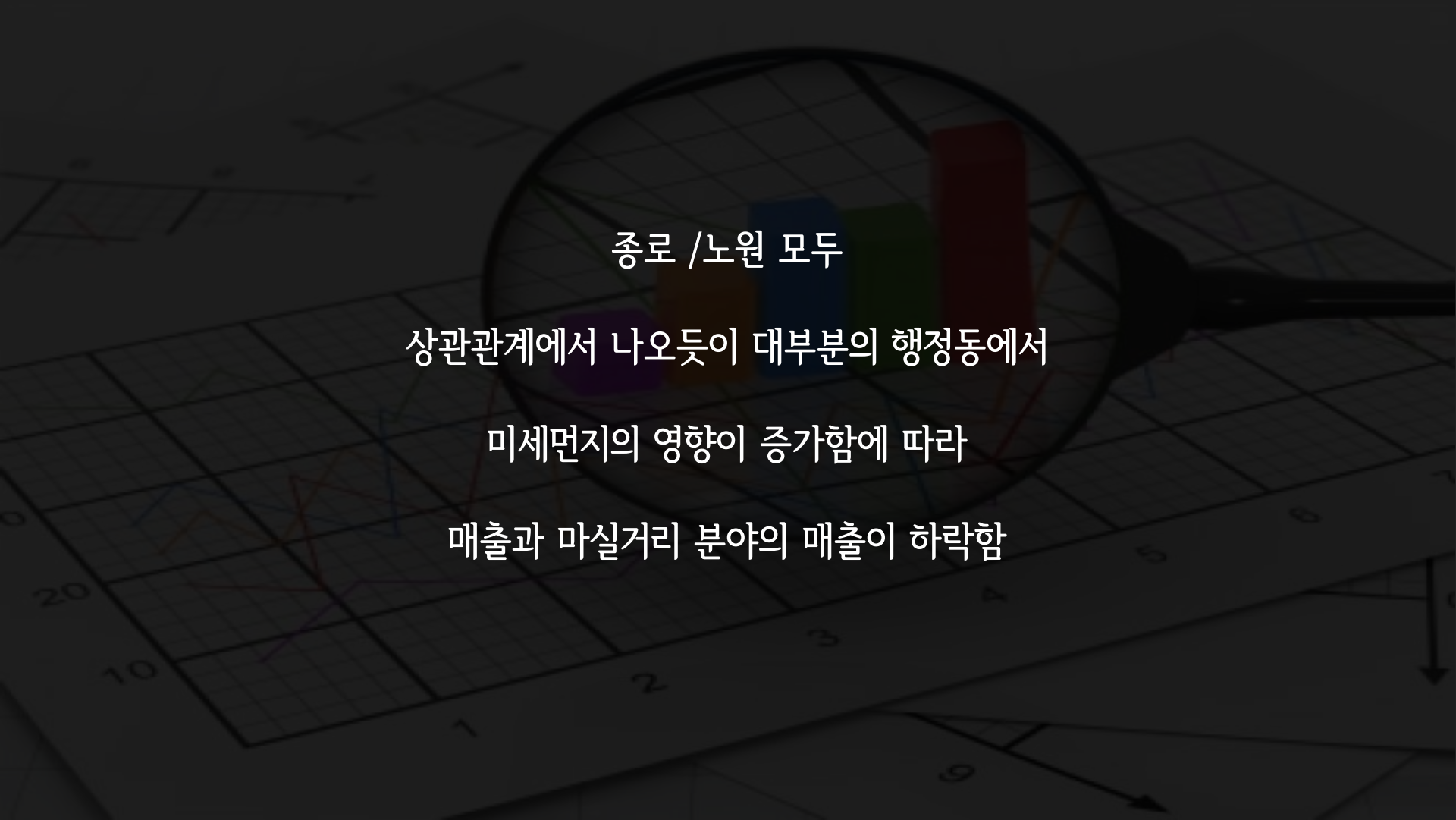
종로구 미세먼지 영향이 증가함에 따른 매출변화



종로구 미세먼지 영향이 증가함에 따른 마실거리 변화





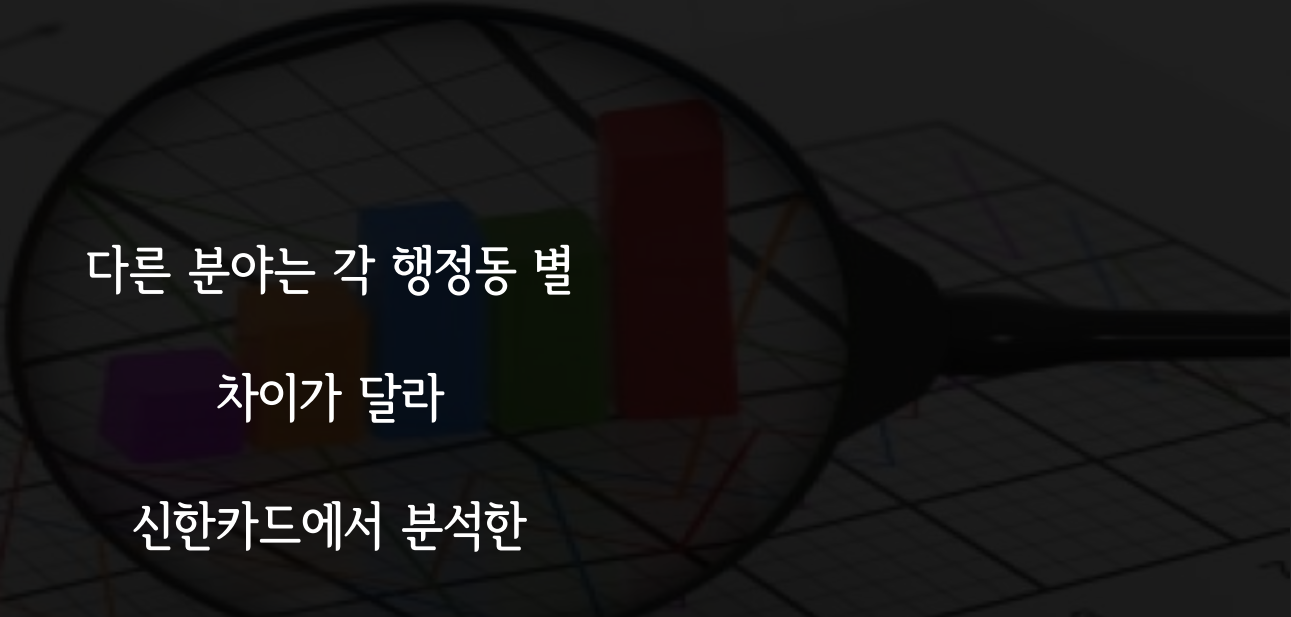


종로 /노원 모두

상관관계에서 나오듯이 대부분의 행정동에서

미세먼지의 영향이 증가함에 따라

매출과 마실거리 분야의 매출이 하락함



다른 분야는 각 행정동 별

차이가 달라

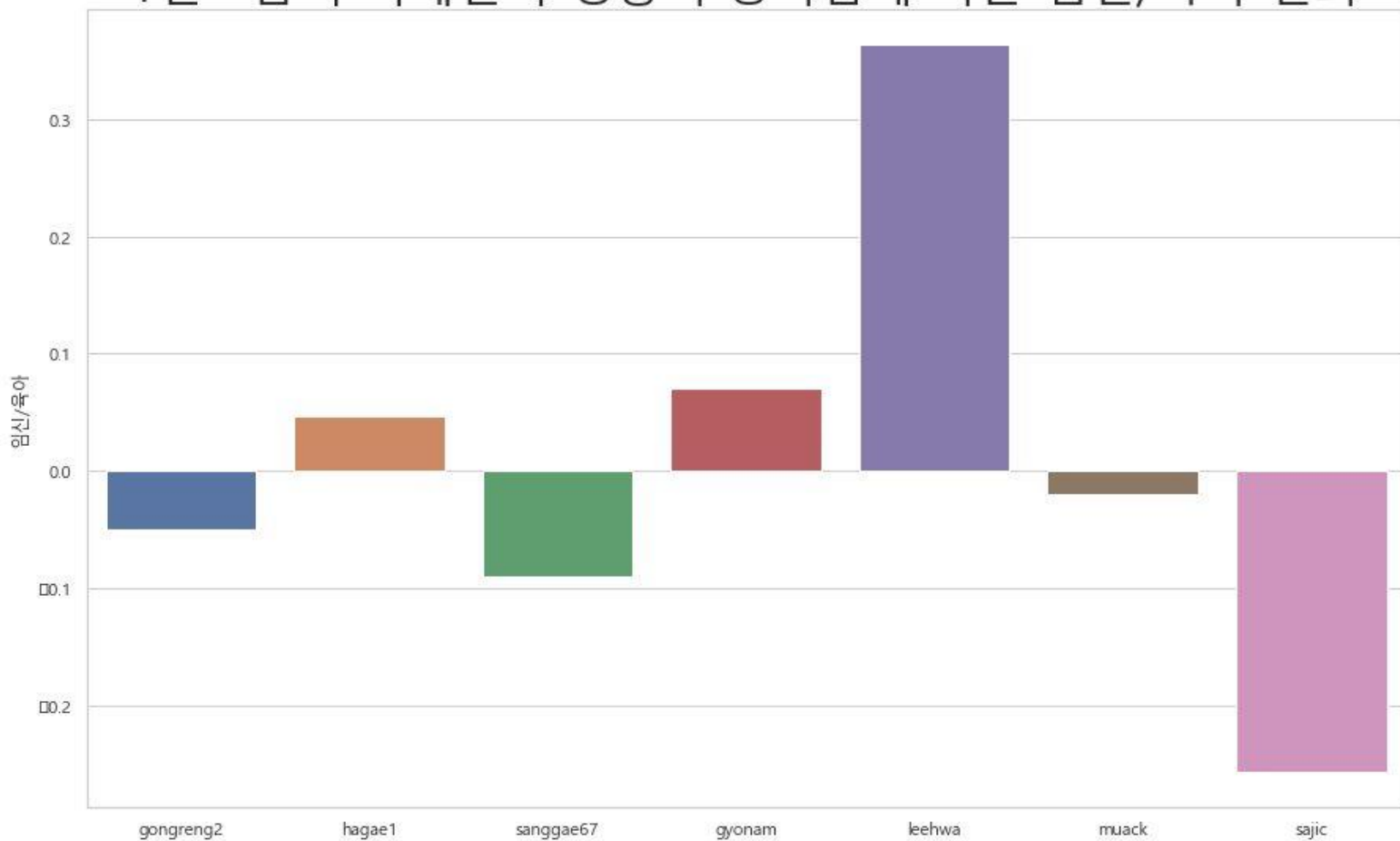
신한카드에서 분석한

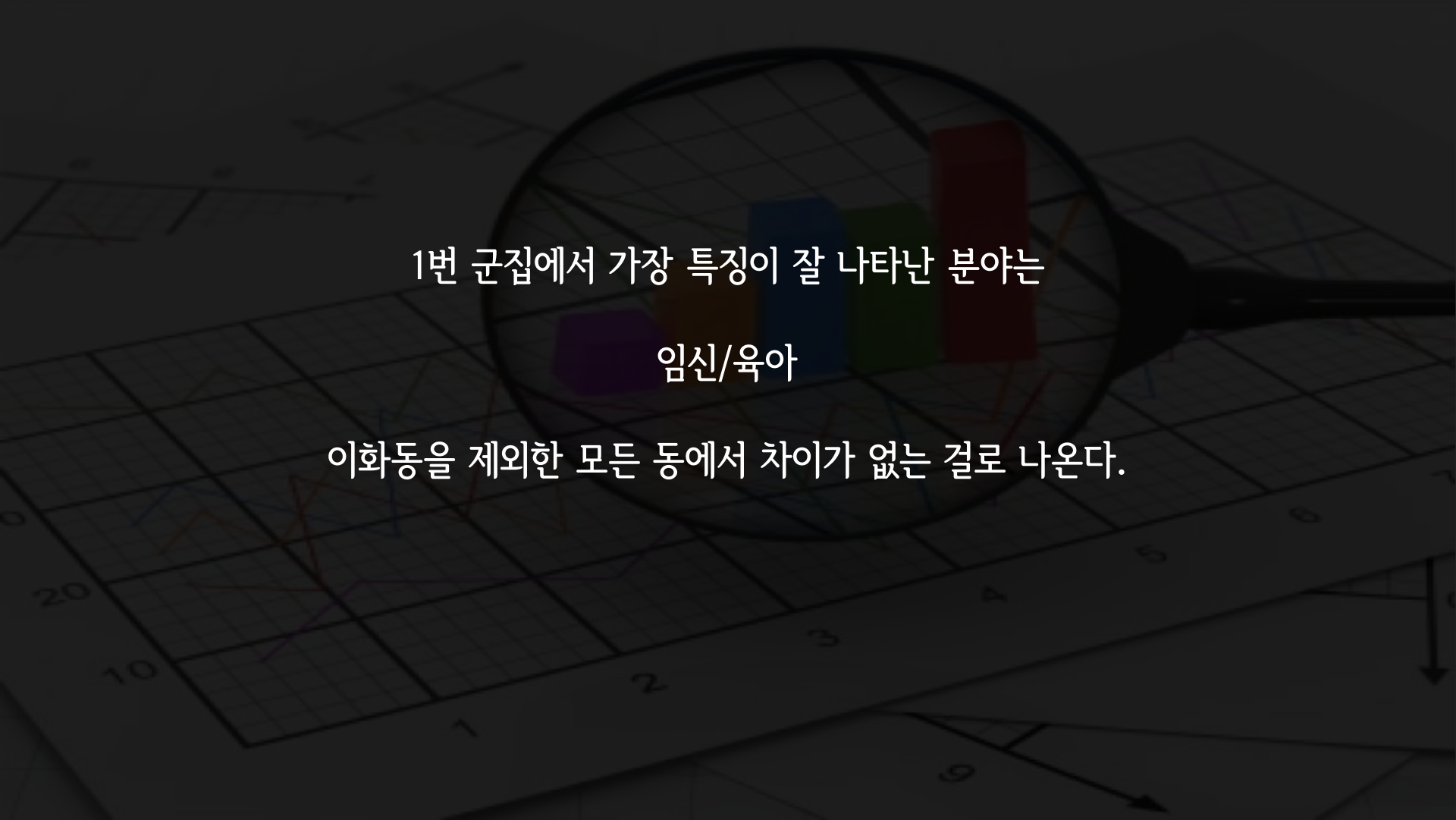
군집을 기준으로 데이터를 확인해 봄.

1번 군집에서 가장 특징이 잘 나타난 분류

1 번 군 집 = 매 출 과 빈 도 하 락

1번그룹의 미세먼지 영향이 증가함에 따른 임신/육아 변화





1번 군집에서 가장 특징이 잘 나타난 분야는

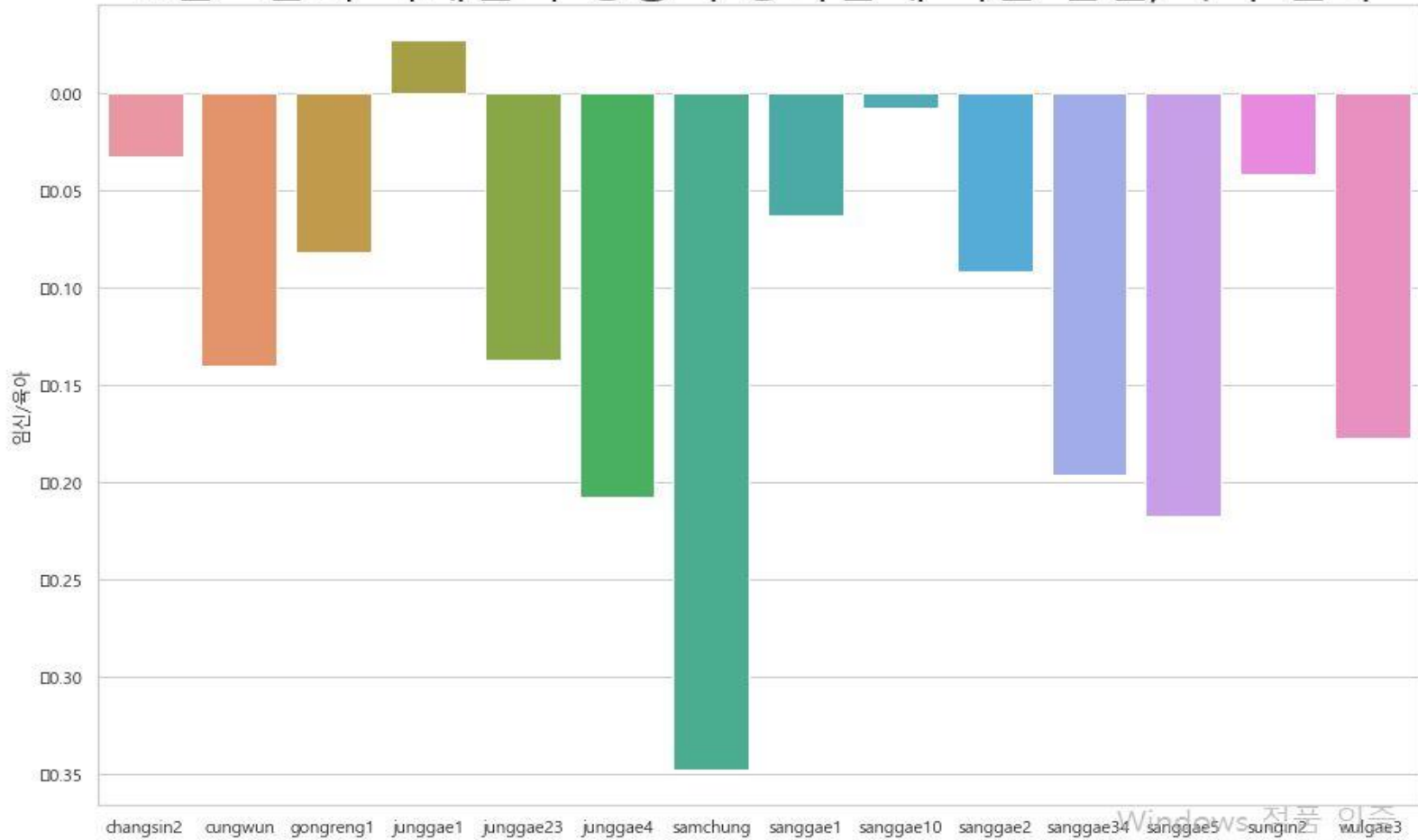
임신/육아

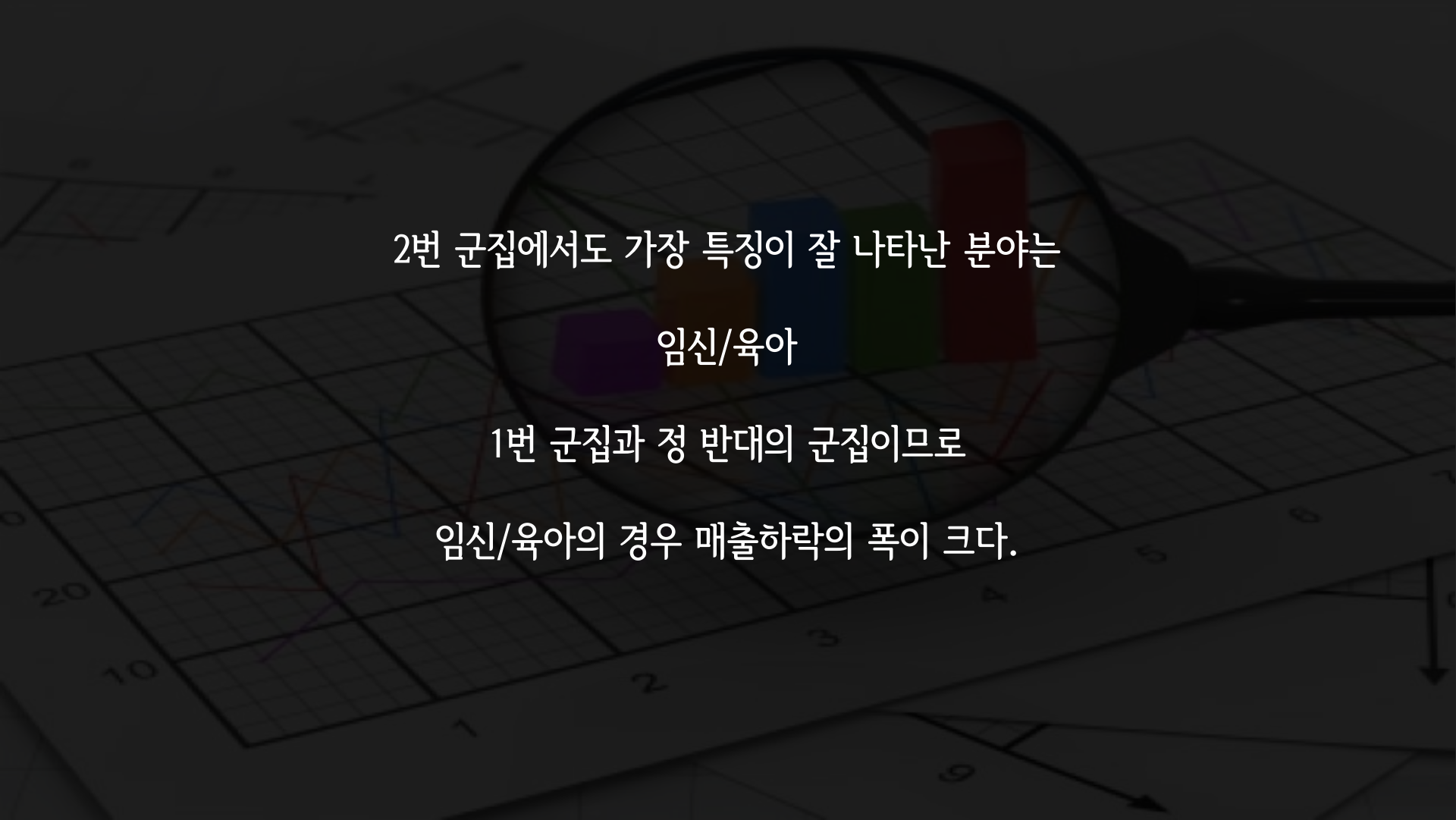
이화동을 제외한 모든 동에서 차이가 없는 걸로 나온다.

2번 군집에서 가장 특징이 잘 나타난 분류

2 번 군 집 = 매 출 과 빈 도 상 승

2번그룹의 미세먼지 영향이 증가함에 따른 임신/육아 변화





2번 군집에서도 가장 특징이 잘 나타난 분야는

임신/육아

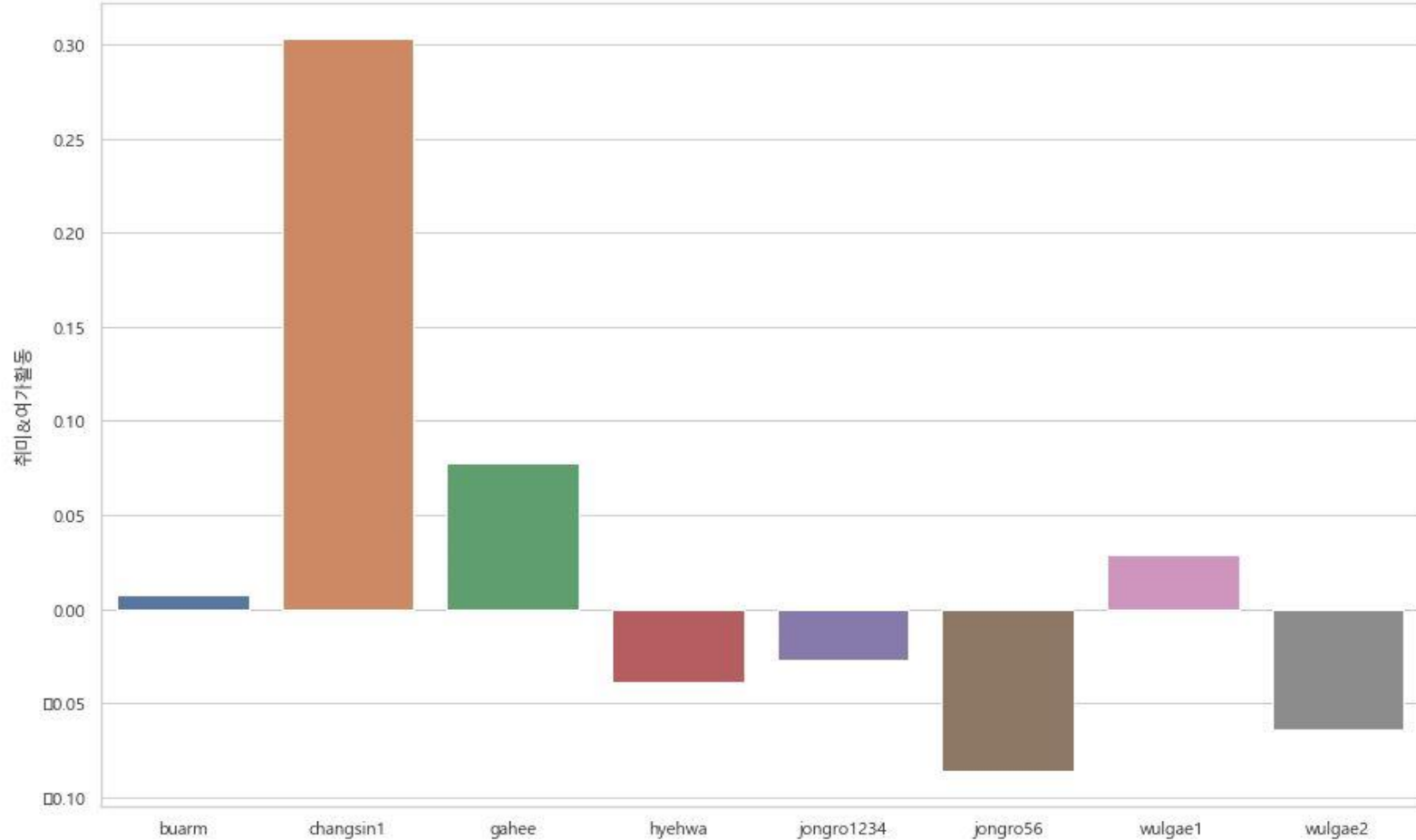
1번 군집과 정 반대의 군집이므로

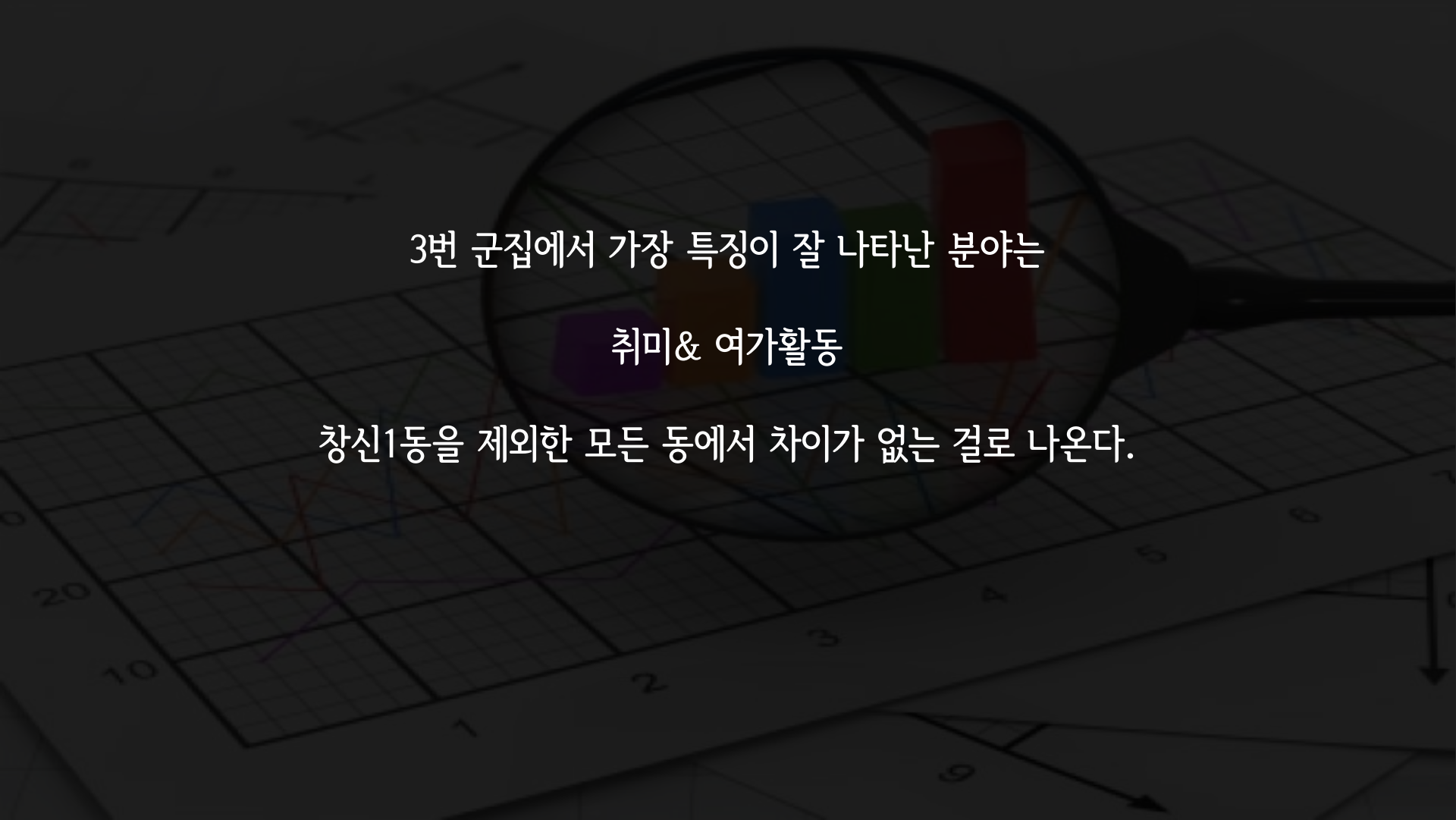
임신/육아의 경우 매출하락의 폭이 크다.

3번 군집에서 가장 특징이 잘 나타난 분류

3 번 군 집 = 매 출 상 승 빈 도 하 락

3번그룹의 미세먼지 영향이 증가함에 따른 취미&여가활동 변화



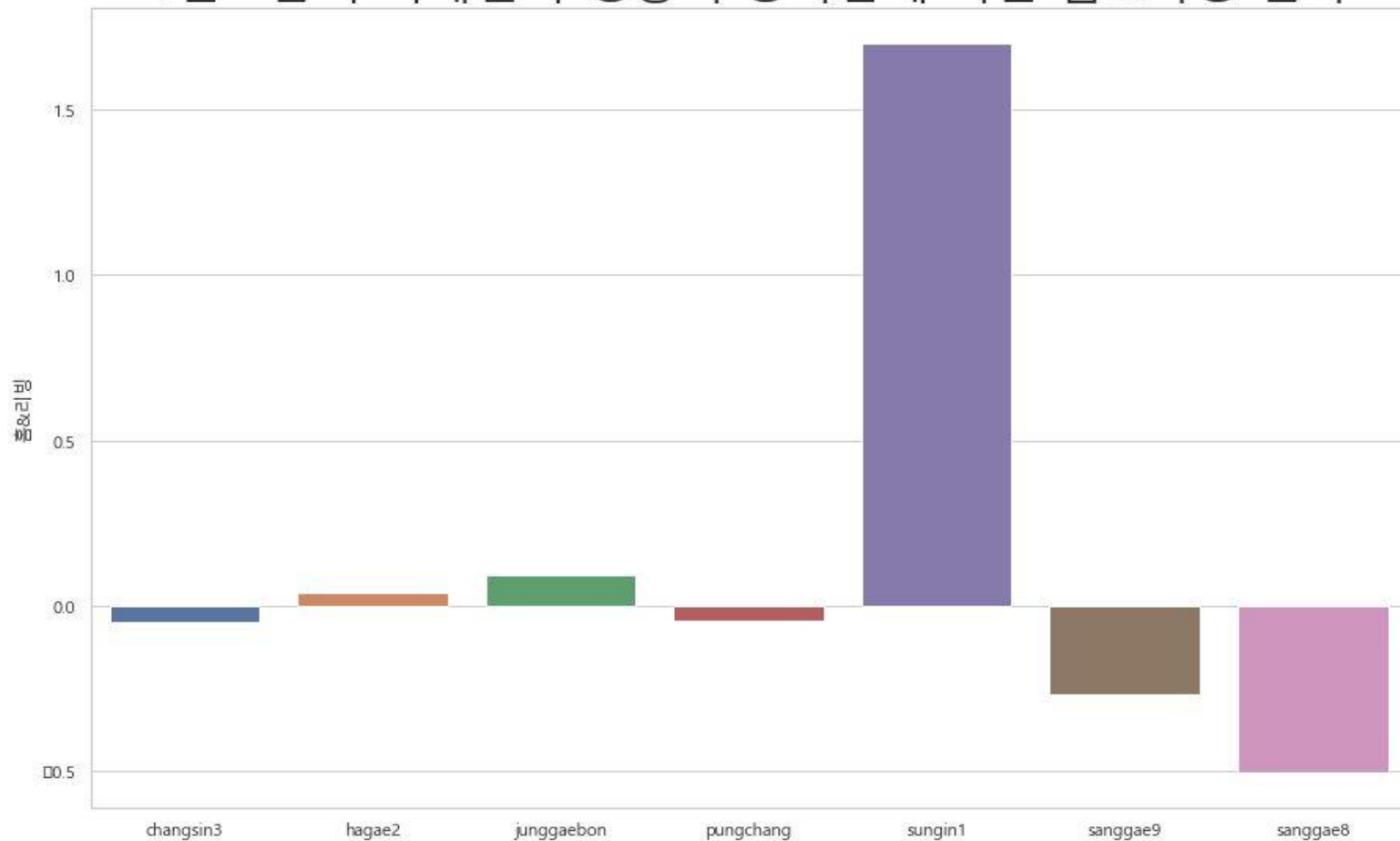


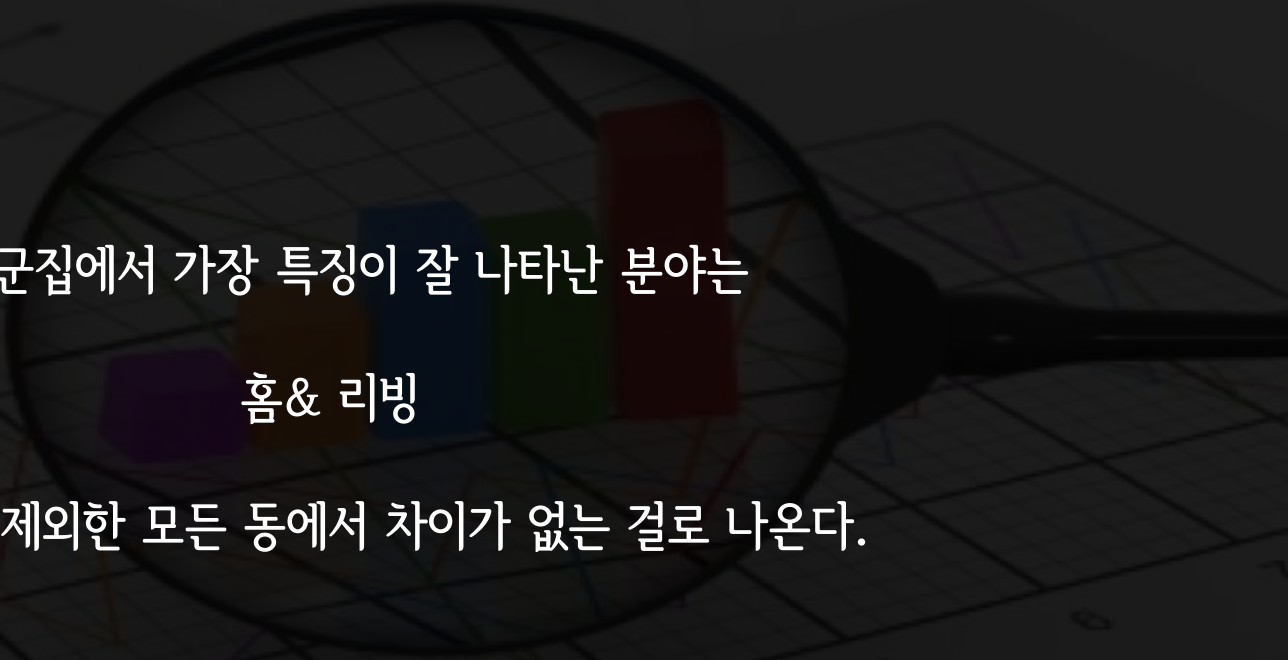
3번 군집에서 가장 특징이 잘 나타난 분야는

취미& 여가활동

창신1동을 제외한 모든 동에서 차이가 없는 걸로 나온다.

4번그룹의 미세먼지 영향이 증가함에 따른 홈&리빙 변화





4번 군집에서 가장 특징이 잘 나타난 분야는

홈& 리빙

송인1동을 제외한 모든 동에서 차이가 없는 걸로 나온다.

분석 결과

1. 미세먼지의 영향이 강해짐에 따라 종로구와 노원구의 매출이 하락함.
2. 미세먼지의 영향이 강해짐에 따라 매출/마실거리는 하락, 식사/간식은 상승함
3. 신한카드 데이터로 나눈 소비형태의 군집을 4개로 분류할 수 있음.
4. 군집별로 GS데이터 미세먼지영향에 따른 상품매출을 확인할 수 있음.

활용 방안

1. 미세먼지의 영향을 볼때 농도만이 아닌 온도와 SNS, 뉴스 등 언론의 영향도 함께 고려해야함.
2. 미세먼지에 대비한 식사/간식 품목에 홍보 및 할인 마케팅을 적용해 볼 가치가 있음.
3. 미세먼지 영향에 따른 품목의 영향을 확인하여 행정동 별 매장의 차별화된 마케팅을 적용할 수 있음.

한 계 점

1. 1년간의 데이터

-1년의 시계열 데이터만 주어져, 계절, 시간 등 매개변수들이 많아 분석에 대한 신뢰성 확보가 어려움. -

2. SNS 데이터질의 문제

-광고성 데이터가 많아 실제 미세먼지 영향력을 추출하기 어려움
또한 용량문제로 google colab활용을 함에도
분석시간이 오래 걸림. -

3. 기상 데이터의 결측 값

-가장 중요한 미세먼지 데이터의 결측 값들이 많음.-

Q & A

감사합니다