



우리가  
도와줄게!

리뷰 보고 갔는데 당황했니?

상혜준 안수정  
이종승 이주연



머신러닝을 이용한  
강남역 회식장소 **리뷰 분류** 프로젝트

# 목차

- 01 기획 의도
- 02 데이터 수집 및 전처리
- 03 EDA
- 04 모델링
- 05 결론 및 에필로그





1

”

## 기획 의도

# 01 기획 의도



회식하기 좋은 강남역 회식장소 아시는 분 있어요?

안녕하세요~ 전 입사한지 얼마 안된 신입사원인데요, 혹시 회식하기 좋은 강남역 회식장소 어디 없나요?  
저희 부서끼리 입사기념 회식하려고 하는데 어디로 가야될지 몰라서 글을 올리게 됐습니다. 제가 신입이고 막내이라 회식 계획 담당하게 됐어요ㅎㅎ  
첫 직장이고 첫 회식이라 회식 장소를 잘 고르고 싶거든요ㅠ 강남역에 맛집이 많지만 회식하기 좋은 장소를 찾기가 쉽지 않더라고요ㅠ<sup>1</sup>  
강남역에서 일한 직장 선배들! 회식하기 좋은 장소 추천 부탁드리겠습니다. 감사합니다. :)



50 회사에서 이번달 업무성과 좋다고 회식한다며 강남역 회식 장소 추천이요~

이번달 업무성과가 좋아서 팀장님이 기분 최고시네요  
이럴때 맛있는거 먹어야 하는데...  
괜찮은 강남역 회식 장소 없을까요?  
평상시 먹을 수 있는 그런거 말고 팀장이 쏘실때  
좋은데 가고 싶어요

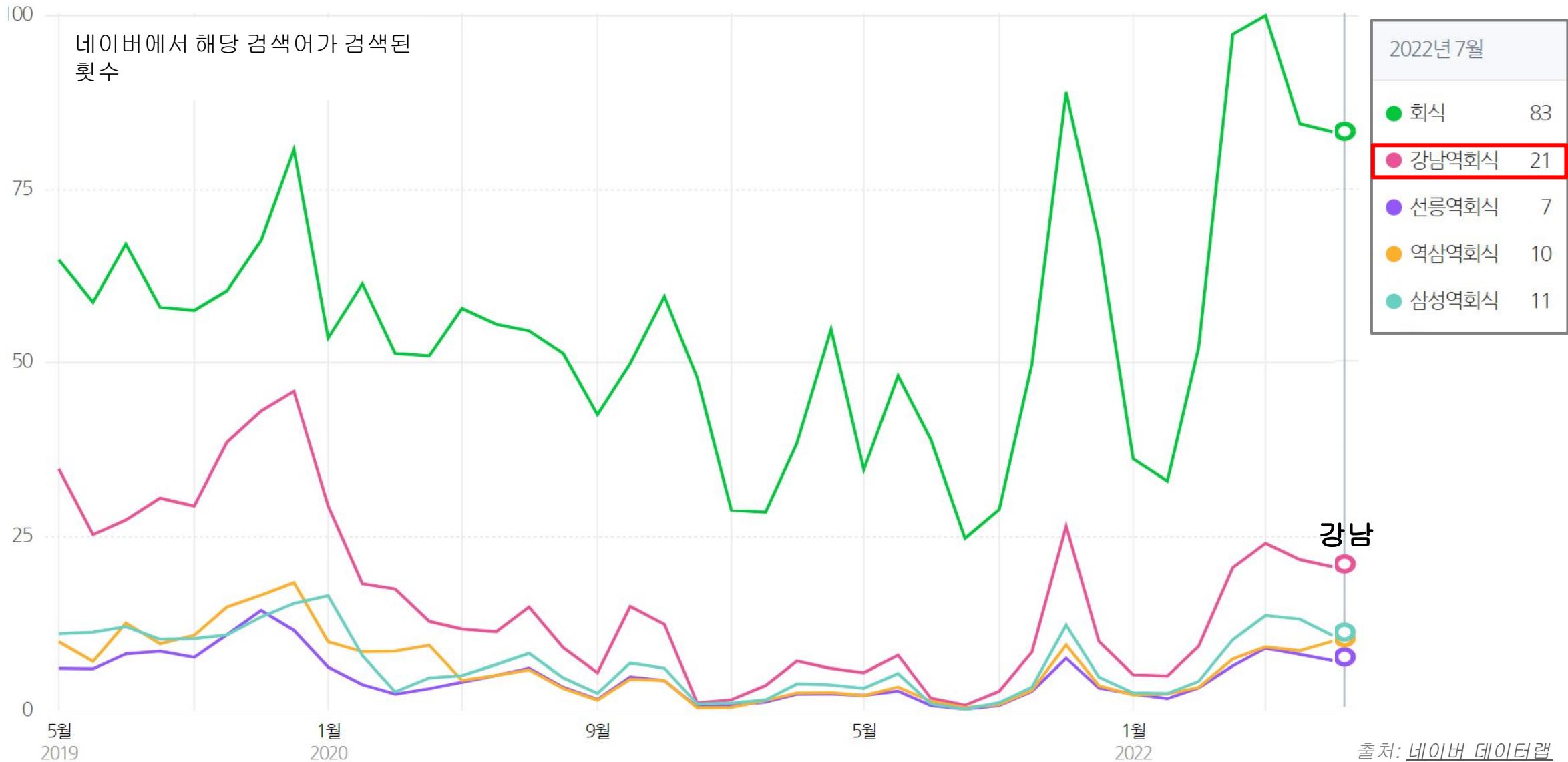


100 강남역 회식장소 있을까요

팀원 4-5명정도 회식 가능한곳이요

출처: 네이버 지식인

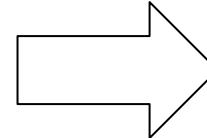
# 01 기획 의도





## 회식 장소 선정시 고려 사항

1. 신입은 아직 회사 근처 식당은 잘 모름
2. 회식장소까지의 거리 고려
3. 남녀별 선호 식당
4. 식당 리뷰는 다 읽어보아야 함



도움이 되는 리뷰 선정

# 01 기획 의도



aa\*\*\*\*

리뷰 64 · 사진 196 · 팔로워 5



매장이 넓고 품 형식이라 프라이빗하게 식사하기 좋았습니다.

한우차돌과 활전복, 관자구이의 삼합 조합도 맛있었고, 전체적으로 음식이 정갈해서 만족스러운 식사였어요~

주차도 편리해서 재방문 의사 있습니다^^\*

😊 음식이 맛있어요 +3

7.12.화 · 1번째 방문 · 영수증



aji\*\*\*\*

리뷰 6 · 사진 1

팔로우

## 백억전복삼합

차돌삼합 주문함 - 구워주시고 나감 - 고기양이 부족하여 등심추가 주문 (등심은 차돌삼합이랑 화로가 달라 버너랑 불판 치우침) - 식사 마친 후 화장실을 가다가 의자 뒤쪽 구석에 놓여있는 차돌접시 발견

직원분을 호출하여 여쭤보니, 구워드시면 된다고 안내했다고 함.

사전에 안내 받지 못했고, 보이지도 않는 곳에 있는 접시를 손님들이 어떻게 알아서 구워 먹느냐, 또 그럼 등심을 추가 주문했을 때 차돌이 한참 남은 것을 직원분은 보셨을텐데 상식적으로 불판을 치우기 전에 한번 더 여쭤보셨어야 하는거 아니냐라고 했으나, 본인들은 이런경우가 많이 있었는데도 한번도 환불해준적 없고, 손님잘못이라 함 처음에는 남은 고기를 구워먹거나 포장해서 집에 가서 구워먹으라고 함. 실랑이를 하다가 결국 차돌값만 환불받음.

★0.5 · 21.2.27.토 · 2번째 방문 · 예약

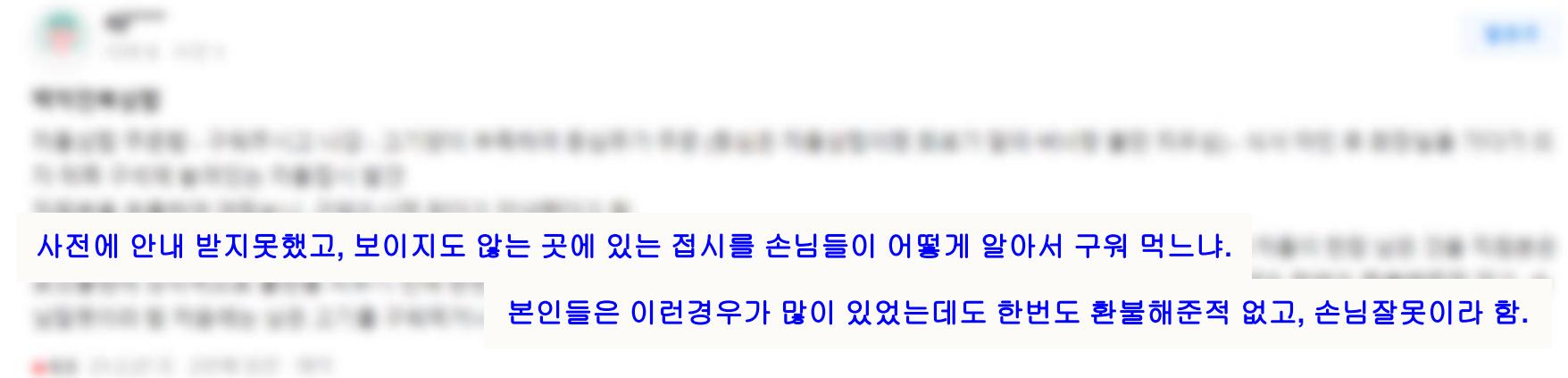
## 01 기획 의도



매장이 넓고 룸 형식이라

맛있었고

주차도 편리해서



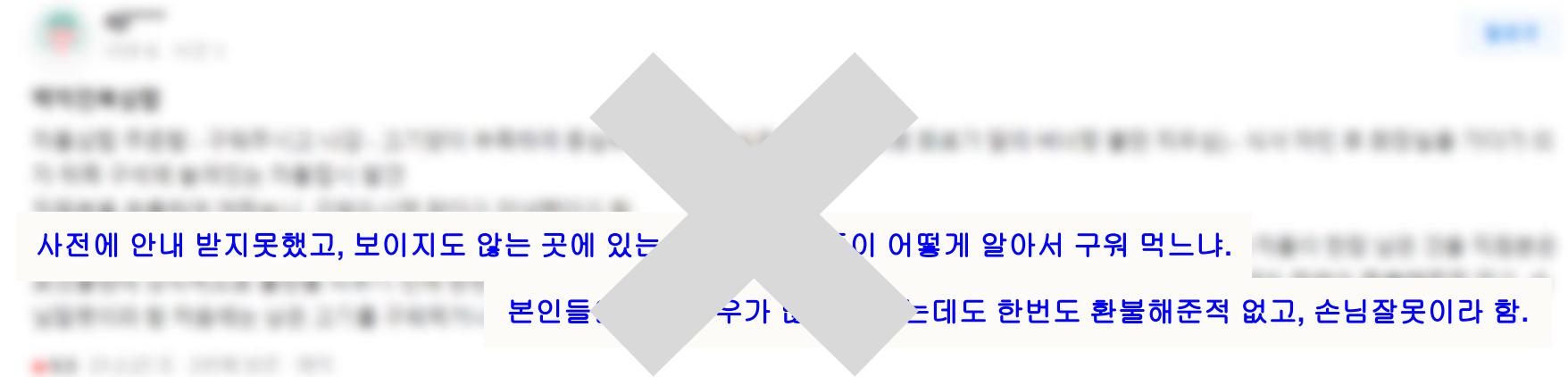
# 01 기획 의도



매장이 넓고 품 형식이라

맛있었고

주차도 편리해서



이 어떻게 알아서 구워 먹느냐.

2

”

데이터 수집 및  
전처리



# 02 데이터 수집 사이트 및 데이터 컬럼



## 데이터 수집 사이트

- 회식 장소 선정: 식신

The screenshot shows the Siksin website interface. At the top, there is a search bar with the placeholder "돈 주고 별 사는 서비스는 그만 후기들이 만드는". Below it, a dropdown menu shows "현재 맛집 지역은 서울-강남, 강남역". The main content area displays a list of dining places with their names, ratings, and brief descriptions. Two specific entries are highlighted: "시골야채된장" with a rating of 4.0 and "청류벽" with a rating of 4.2. Both entries include a "주차" (Parking) option. A red box highlights the "회식" (Dining) button in the filter section under "테마".

| 장소     | 평점  | 설명  |
|--------|-----|---|
| 시골야채된장 | 4.0 | 강남역<br>시골야채된장비빔밥, 목살김치된장(2인), 청국장(2인), 생물살삼겹살, 생삼겹살         |
| 청류벽    | 4.2 | 강남역<br>제복장반, 강자전, 냉제육, 들기름 막국수, 생대지 김치찌개식 두루치기, 계란말이, 돼지갈비찜 |

The screenshot shows the Siksin website interface with a yellow header bar. It features sections for "추천 맛집(32)", "테마(14)", and "리뷰(541)". Below this, a banner states "추천 맛집(32) 식신과 매장 방문 고객님들이 함께 엄선한 인기 맛집입니다.". Three recommended dishes are shown with "2022" badges: a spicy stew, a bowl of rice with toppings, and a large pot of soup. Below each dish is its name, rating, location, and a brief description. Red boxes highlight the "2022" badge and the "주차" (Parking) option.

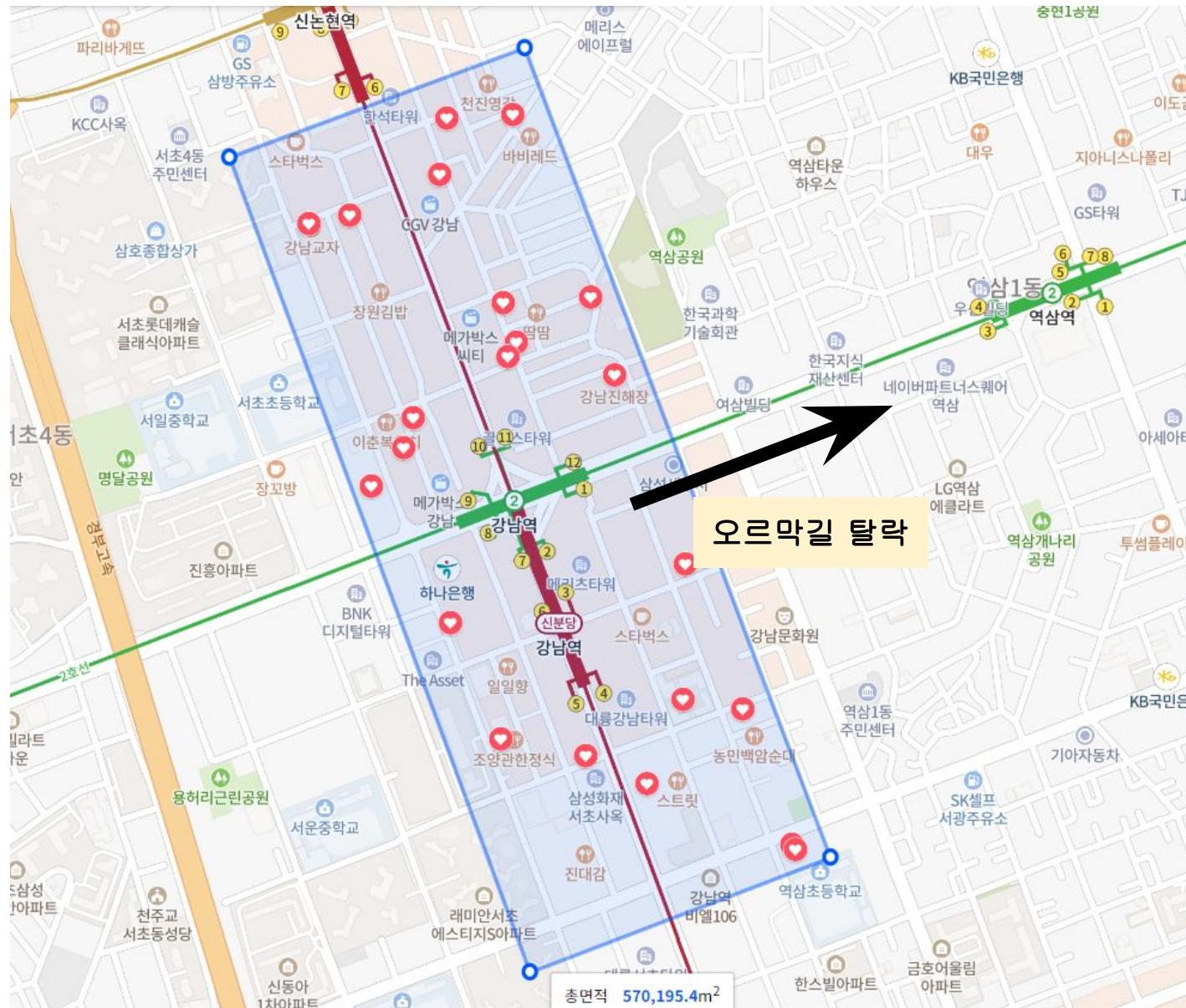
| 장소     | 평점  | 설명  |
|--------|-----|---|
| 시골야채된장 | 4.0 | 강남역<br>시골야채된장비빔밥, 목살김치된장(2인), 청국장(2인), 생물살삼겹살, 생삼겹살         |
| 청류벽    | 4.2 | 강남역<br>제복장반, 강자전, 냉제육, 들기름 막국수, 생대지 김치찌개식 두루치기, 계란말이, 돼지갈비찜 |

## 02 데이터 수집 사이트 및 데이터 컬럼



## 데이터 수집 사이트

- 회식 장소 선정: 식신



# 02 데이터 수집 사이트 및 데이터 컬럼



## 데이터 수집 사이트

- 리뷰 데이터: 네이버 MyPlace

검색 장소, 버스, 지하철, 도로 검색

### < 시골야채된장

홈 메뉴 리뷰 사진

리뷰 299 사진 리뷰만

찌개 31 찌개 31 청국장 18 계란찜 16 비빔 만족도 71 서비스 17 가격 7 청결도 5

추천순 Beta 최신순

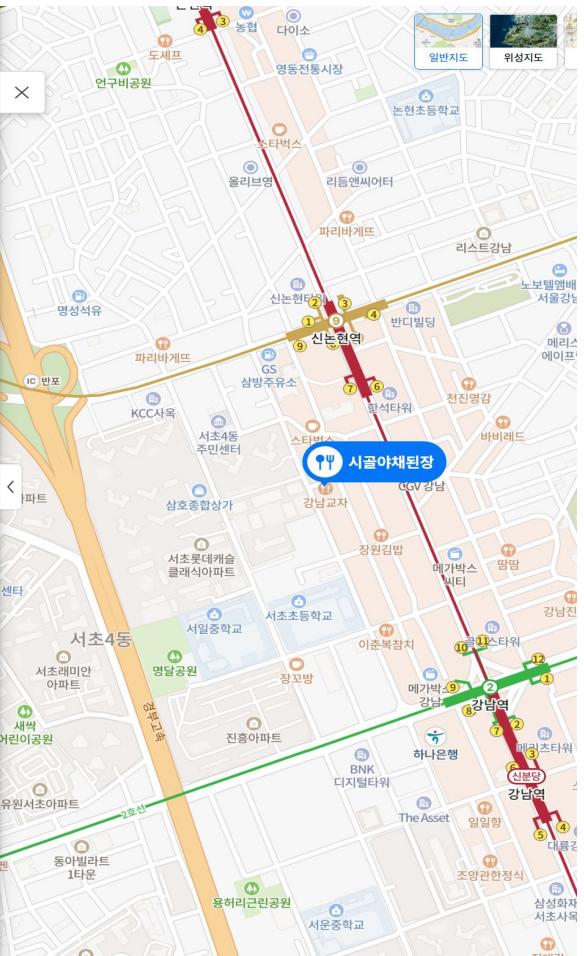
캠블러즈 리뷰 115 사진 362 팔로워 16

야채 넣고 된장 한 국자 넣고 속속 비벼먹으면 완전 맛 있음 👍 생각나는 맛, 건강해지는 맛

음식이 맛있어요 +2

8.26.금 1번째 방문 영수증

김제제 리뷰 80 사진 234 팔로워 24



추천순 Beta 최신순

캠블러즈 리뷰 115 사진 362 팔로워 16



야채 넣고 된장 한 국자 넣고 속속 비벼먹으면 완전 맛 있음 👍 생각나는 맛, 건강해지는 맛

음식이 맛있어요 +2

8.26.금 1번째 방문 영수증

행복하잖니 리뷰 636 사진 56 팔로워 3



맛은 변하고 가격은 올랐을 노맛

8.20.토 1번째 방문 영수증

## 02 데이터 수집 사이트 및 데이터 컬럼



### 데이터 기간

- 2019년 5월 26일 ~ 2022년 8월 19일 → **약 8000개**의 데이터 사용

### 데이터 컬럼 설명

- 가게 이름
  - 남/녀 비율
  - 연령별 1위, 2위
- 리뷰 날짜
- 방문자 리뷰
- helpful(label)
  - 카테고리(menu, mood, service, price, clean, room, parking, photo)에서 3개 이상 해당이 되면 helpful = 1 or helpful = 0
- 키워드 관련 항목(맛, 맛집, 메뉴 + 음식, 친절 + 직원, 때, 예약, 가격, 분위기, 추천)

## 02 데이터 및 전처리



```
text='테이블 서비스해 주시는 직원분도 친절하시고 차돌 삼합 너무 맛있어요👍👍\n전복식감0\ntext
```

'테이블 서비스해 주시는 직원분도 친절하시고 차돌 삼합 너무 맛있어요👍👍\n전복식감이 훌륭합니다 마무리 누룽지, 된찌, 물냉면까지 음식이 다 정갈하고 맛있어서 반하고 왔답니다❤️음식이 맛있어요~~ 재료가 신선해요 ㅎㅎ'

```
text_re=text.replace('👍', '굿')\ntext_re=text_re.replace('❤️', '하트')\ntext_re
```

'테이블 서비스해 주시는 직원분도 친절하시고 차돌 삼합 너무 맛있어요 굿 굿\n전복식감이 훌륭합니다 마무리 누룽지, 된찌, 물냉면까지 음식이 다 정갈하고 맛있어서 반하고 왔답니다 하트\n음식이 맛있어요~~ 재료가 신선해요 ㅎㅎ'

```
text_re=text_re.replace('\n', ' ')  
text_re=text_re.replace('\r', ' ')  
text_re=cleanText(text_re)  
text_re=cleanText2(text_re)  
text_re
```

'테이블 서비스해 주시는 직원분도 친절하시고 차돌 삼합 너무 맛있어요 굿 굿 전복식감이 훌륭합니다 마무리 누룽지 된찌 물냉면까지 음식이 다 정갈하고 맛있어서 반하고 왔답니다 하트 음식이 맛있어요 재료가 신선해요 '

이모티콘 텍스트로 대체해주는 방식

적용

의미없는 자음, 모음, 특수문자, 개행문자 제거

## 02 데이터 및 전처리



### [맞춤법 검사 전 리뷰]

```
text1 = '백억전복삼합 먹었는데 전복식감이 훌륭합니다'  
text2 = '코시국에 륨이 있어 좋습니다'
```

### [1차 시도: 맞춤법]

```
spelled_sent1 = spell_checker.check(text1)  
checked_sent1 = spelled_sent1.checked  
spelled_sent2 = spell_checker.check(text2)  
checked_sent2 = spelled_sent2.checked  
print("text1:", checked_sent1), print("text2:", checked_sent2)
```

```
text1: 백억 전복 삼합 먹었는데 전복 식감이 훌륭합니다  
text2: 고 시국에 륨이 있어 좋습니다
```

### [2차 시도: 띄어쓰기 + 맞춤법]

```
spacing = Spacing()  
new_sent1 = text1.replace(" ", "")  
kospacing_sent1 = spacing(new_sent1)  
spelled_sent1 = spell_checker.check(kospacing_sent1)  
checked_sent1 = spelled_sent1.checked  
new_sent2 = text2.replace(" ", "")  
kospacing_sent2 = spacing(new_sent2)  
spelled_sent2 = spell_checker.check(kospacing_sent2)  
checked_sent2 = spelled_sent2.checked  
print("text1:", checked_sent1), print("text2:", checked_sent2)
```

```
text1: 백억 전복 삼합 먹었는데 전복 식감이 훌륭합니다  
text2: 고 시국에 륨이 있어 좋습니다
```

1차. 맞춤법 및 띄어쓰기 검사



2차. 추가 오류 검은 Human performance 이용



## 모인 데이터의 데이터프레임

|   | store_name | date      | review  | length | review_check                                       | check_length |
|---|------------|-----------|---|--------|--|--------------|
| 0 | 화기애애 강남 역점 | 2018.12.1 | 4명이서 5인분 이라고 되어있는 모듬 대를 시켰습니다 고기는 괜찮았고 양은 조금 ...  | 254    | 4명이서 5인분이라고 되어 있는 모둠대를 시켰습니다 고기는 괜찮았고 양은 조금 적었...  | 251          |
| 1 | 화기애애 강남 역점 | 2019.3.19 | 가격대비 고기도 맛있고 직원도 친절하고 우리가 인원이 많아 시끄러울까봐 룸을 부탁드... | 122    | 가격 대비 고기도 맛있고 직원도 친절하고 우리가 인원이 많아 시끄러울까 봄을 부탁드...  | 130          |
| 2 | 백억하누 강남 본점 | 2020.1.3  | 네이버로 예약하고 방문했는데 사정이 있어 조용한 구석자리를 부탁한다고 썼는데도 구석... | 358    | 네이버로 예약하고 방문했는데 사정이 있어 조용한 구석 자리를 부탁한다고 썼는데도 구석... | 370          |
| 3 | 강남진해장      | 2022.4.5  | 일반적인 양평해장국과 다르게 기본 간이 되어있고 베이스가 육개장과 뼈해장국같은 느...  | 343    | 일반적인 양평 해장국과 다르게 기본 간이 되어 있고 베이스가 육개장과 뼈해장국 같은...  | 348          |
| 4 | 백억하누 강남 본점 | 2022.5.7  | 개인룸이 있고 룸안에 화장실이 있어 너무 편하고 좋았는데 옆에 개인 서랍이 있는지...  | 163    | 개인 룸이 있고 룸 안에 화장실이 있어 너무 편하고 좋았는데 옆에 개인 서랍이 있는지... | 165          |



## 도움이 되는 카테고리 분류

| menu    | mood    | service       | price | clean | room   | parking | photo |
|---------|---------|---------------|-------|-------|--------|---------|-------|
| 양이 많아요  | 위치가 좋았다 | 서빙이 빠르다 / 느리다 | 가성비   | 깨끗/더럽 | 네이버 예약 | 주차      | 사진 유무 |
| 양이 적어요  | 인테리어    | 안내            | 가격 대비 | 거리두기  | 룸      | 발렛      |       |
| 맛이 쏘쏘   | 사람이 많아요 | 불친절           | 비싸    | 화장실   | 모임     |         |       |
| 잘 먹고갑니다 | 자리가 좋아요 | 진동벨 / 호출      | 돈(금액) | 환기    | 단체     |         |       |
| 메뉴명     | 춥다/덥다   | 종업원이 원가 해줌    | 포인트   | 흡연장소  | 직장동료 술 |         |       |



## helpful 포함 데이터프레임

|   | store_name | date      | review  | length | review_check                                       | check_length | helpful |
|---|------------|-----------|---|--------|--|--------------|---------|
| 0 | 화기애애 강남 역점 | 2018.12.1 | 4명이서 5인분 이라고 되어있는 모듬 대를 시켰습니다 고기는 괜찮았고 양은 조금 ...  | 254    | 4명이서 5인분이라고 되어 있는 모둠대를 시켰습니다 고기는 괜찮았고 양은 조금 적었...  | 251          | 1       |
| 1 | 화기애애 강남 역점 | 2019.3.19 | 가격대비 고기도 맛있고 직원도 친절하고 우리가 인원이 많아 시끄러울까봐 룸을 부탁드... | 122    | 가격 대비 고기도 맛있고 직원도 친절하고 우리가 인원이 많아 시끄러울까 봄을 부...    | 130          | 1       |
| 2 | 백억하누 강남 본점 | 2020.1.3  | 네이버로 예약하고 방문했는데 사정이 있어 조용한 구석자리를 부탁한다고 썼는데도 구석... | 358    | 네이버로 예약하고 방문했는데 사정이 있어 조용한 구석 자리를 부탁한다고 썼는데도 구...  | 370          | 1       |
| 3 | 강남진해장      | 2022.4.5  | 일반적인 양평해장국과 다르게 기본 간이 되어 있고 베이스가 육개장과 뼈해장국같은 느... | 343    | 일반적인 양평 해장국과 다르게 기본 간이 되어 있고 베이스가 육개장과 뼈해장국 같은...  | 348          | 1       |
| 4 | 백억하누 강남 본점 | 2022.5.7  | 개인룸이 있고 룸안에 화장실이 있어 너무 편하고 좋았는데 옆에 개인 서랍이 있는지...  | 163    | 개인 룸이 있고 룸 안에 화장실이 있어 너무 편하고 좋았는데 옆에 개인 서랍이 있는지... | 165          | 1       |

\* helpful = 카테고리가 3개 이상이면 1, 아니면 0



EC : 연결어미  
EF : 종결어미  
EP : 선어말어미  
ETM : 관형형 전성 어미  
ETN : 명사형 전성 어미  
IC : 감탄사  
JC : 접속 조사  
JKB : 부사격 조사  
JKC : 보격 조사  
JKG : 관형격 조사  
JKO : 목적격 조사  
JKQ : 인용격 조사  
JKS : 주격 조사  
JKV : 호격 조사  
JX : 보조사  
MAG : 일반 부사  
MAJ : 접속 부사  
MM : 관형사  
NNB : 의존 명사  
NNBC : 단위를 나타내는 명사  
NNG : 일반 명사  
NNP : 고유 명사

NP : 대명사  
NR : 수사  
SC : 구분자, /, :  
SE : 줄임표  
SF : 마침표, 물음표, 느낌표  
SH : 한자  
SL : 외국어  
SN : 숫자  
SSC : 닫는 괄호  
SSO : 여는 괄호  
SY : 기타 기호  
VA : 형용사  
VCN : 부정 지정사  
VCP : 긍정 지정사  
VV : 동사  
VX : 보조 용언  
XPN : 체언 접두사  
XR : 어근  
XSA : 형용사 파생 접미사  
XSN : 명사파생 접미사  
XSV : 동사 파생 접미사

### 토큰화(tag)

- tag
  - 'NNG', 'NNP', 'SN', 'NR', 'VA', 'VV', 'VCN', 'VCP', 'XPN', 'XR', 'NNBC'



### 토큰화(불용어 처리 & 원핫인코딩)

- 불용어

- 한우, 하, 포장, 평양냉면, 탕, 타이, 전복, 전, 재, 우동, 오, 들, 되, 느끼, 네이버, 가 ...

-> 총 **43개**의 불용어 제거

- One-Hot Encoding

- 딘타이펑 강남점 : 1, 백억하누 강남본점 : 2, 느린마을 양조장: 3 화기애애 강남역점 : 4 ...

-> 총 **22개**의 store\_name 변경



3

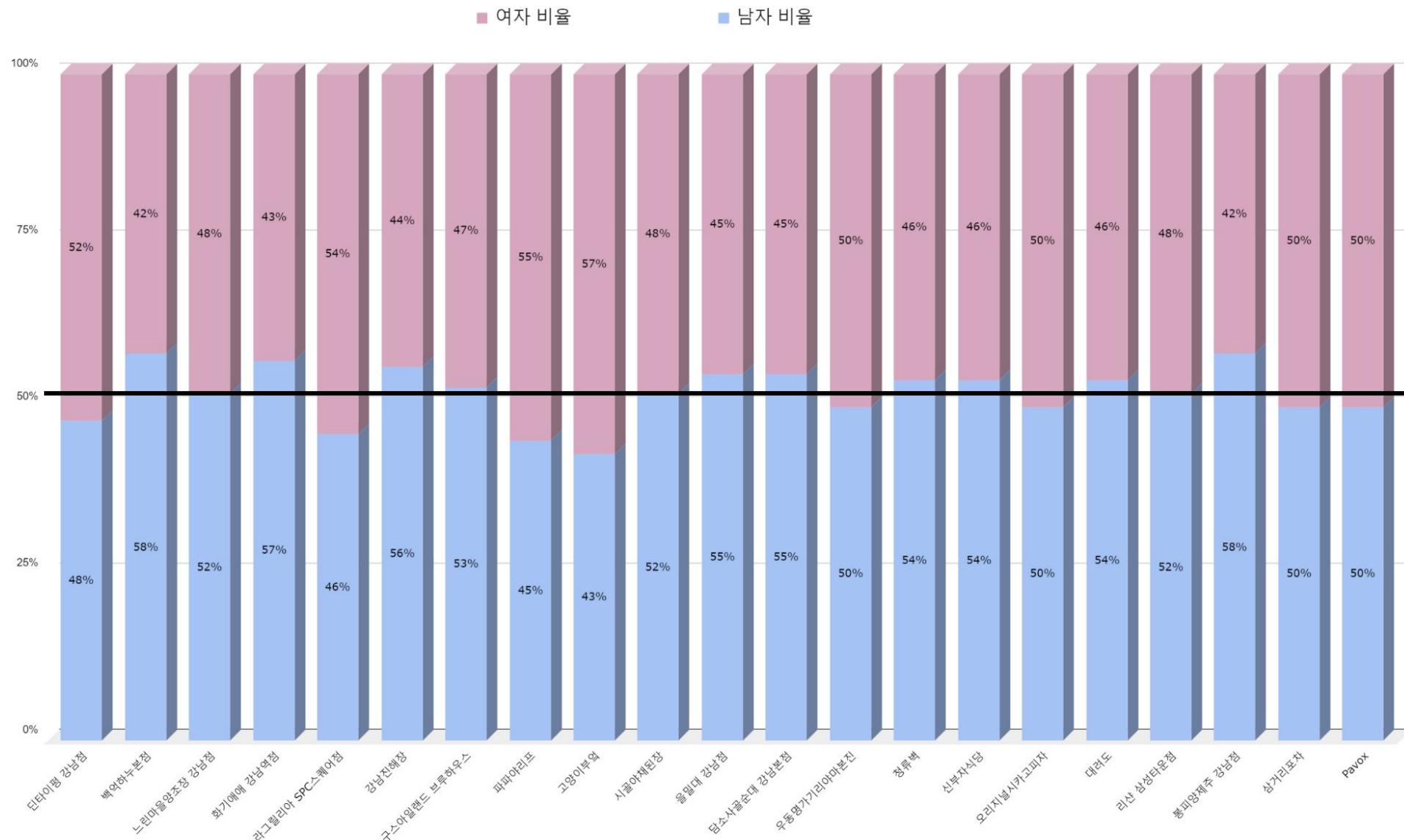
EDA

”

# 03 EDA



가게별 남녀 비율 분포도 그래프



## 03 EDA



가게별 리뷰 갯수

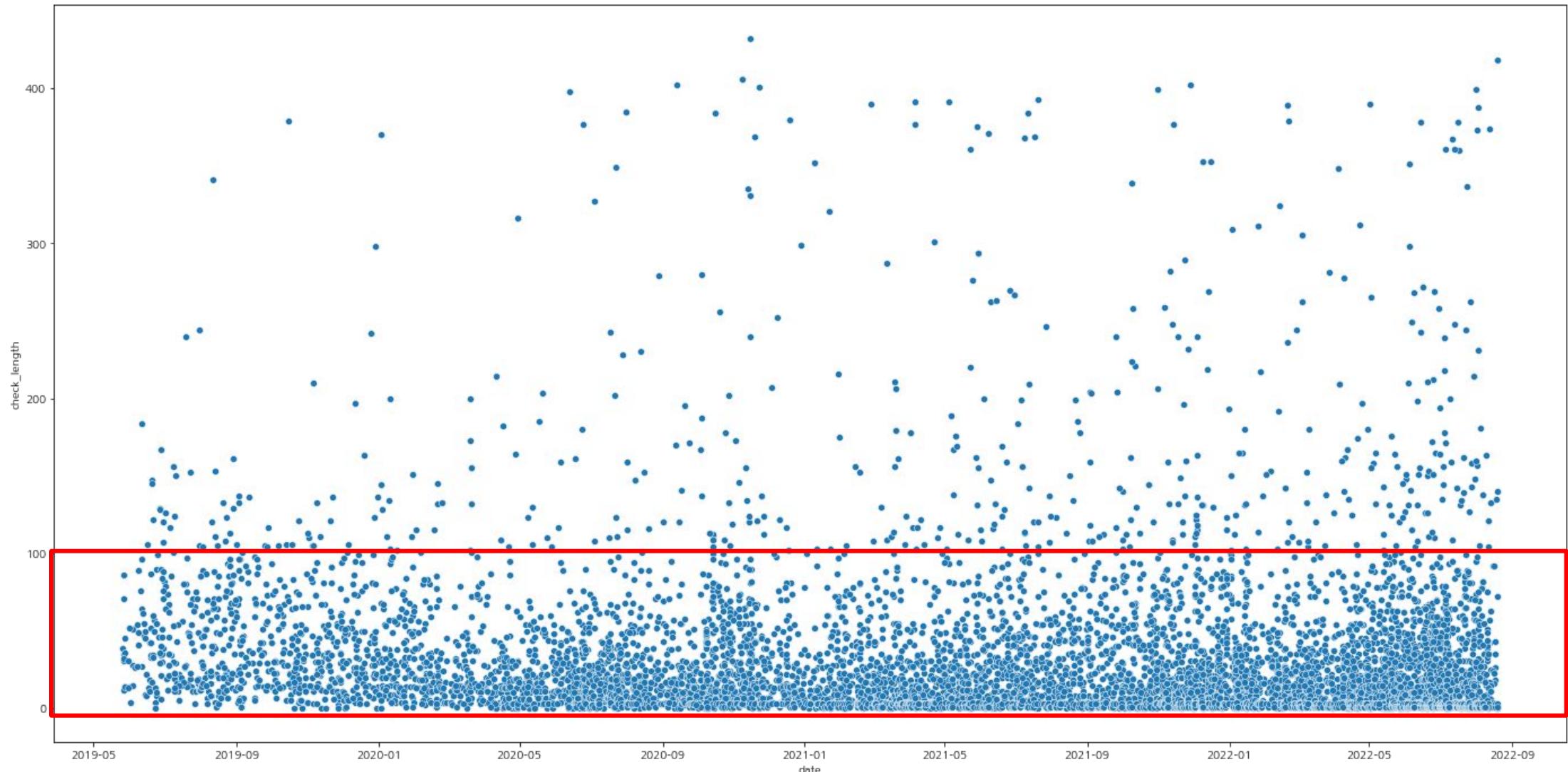
```
df_re['store_name'].value_counts()
```

|               |      |
|---------------|------|
| 딘타이펑 강남점      | 1296 |
| 백억하누 강남본점     | 1111 |
| 느린마을양조장 강남점   | 624  |
| 화기애애 강남역점     | 504  |
| 라그릴리아 SPC스퀘어점 | 435  |
| 강남진해장         | 424  |
| 구스아일랜드 브루하우스  | 404  |
| 파파야리프         | 353  |
| 고양이부엌         | 331  |
| 시골야채된장        | 286  |
| 을밀대 강남점       | 283  |
| 담소사골순대 강남본점   | 278  |
| 청류벽           | 269  |
| 우동명가기리야마본진    | 268  |
| 신부자식당         | 247  |
| 오리지널시카고피자     | 183  |
| 대려도           | 157  |
| 봉추찜닭 강남주공공이점  | 131  |
| 리샨 삼성타운점      | 112  |
| 봉피양제주 강남점     | 71   |
| Pavox         | 29   |
| 삼거리포차         | 26   |

## 03 EDA



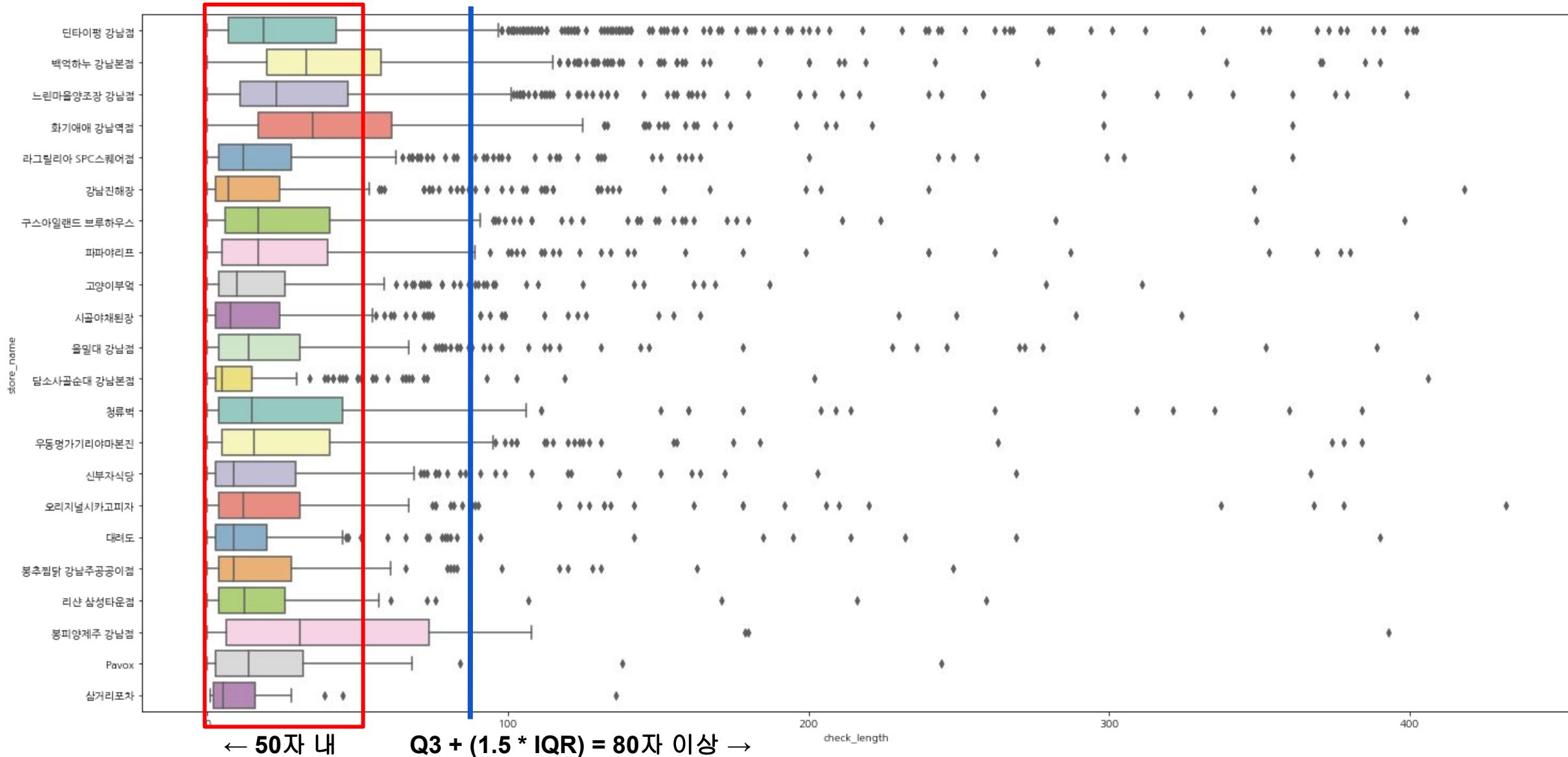
리뷰길이 비교



## 03 EDA



식당별 리뷰길이 비교



## 03 EDA



식당별 리뷰길이 이상치 확인

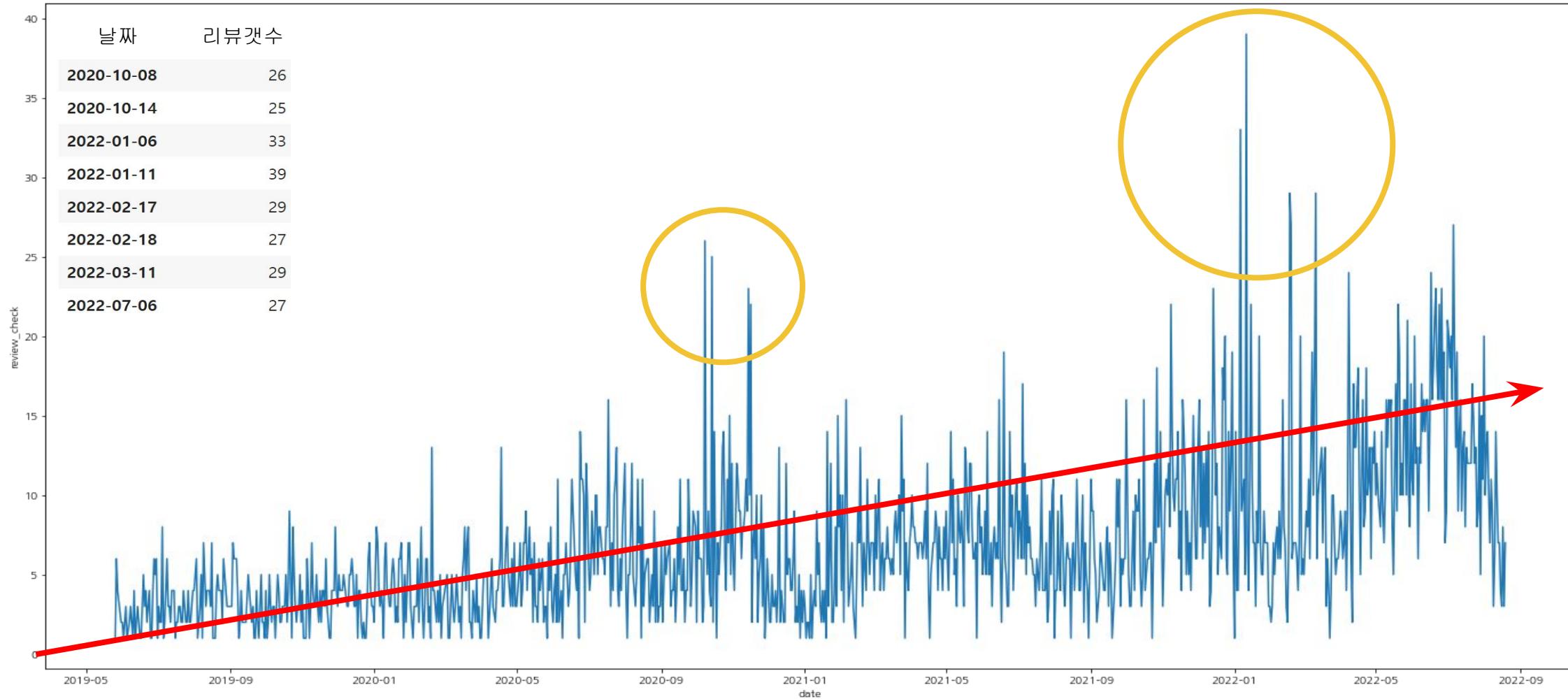
```
df2019_80=df2019.copy()
df2019_80=df2019_80[['store_name', 'date', 'review_check', 'check_length', 'helpful']]
df2019_80=df2019_80[df2019_80['check_length']>=80]
df2019_80['helpful'].value_counts()
```

```
1    427
0    345
Name: helpful, dtype: int64
```

# 03 EDA



날짜별 리뷰갯수 추세 확인



# 03 EDA



## 날짜별 리뷰 빈도수 비교



df\_re['store\_name'].value\_counts()

|               |      |
|---------------|------|
| 민타이핑 강남점      | 1296 |
| 백억하누 강남본점     | 1111 |
| 느린마을양조장 강남점   | 624  |
| 화기애애 강남역점     | 504  |
| 라그릴리아 SPC스퀘어점 | 435  |
| 강남진해장         | 424  |
| 구스아일랜드 브루하우스  | 404  |
| 파파야리프         | 353  |
| 고양이부엌         | 331  |
| 시골야채된장        | 286  |
| 을밀대 강남점       | 283  |
| 담소사골순대 강남본점   | 278  |
| 청류벽           | 269  |
| 우동명가기리야마본진    | 268  |
| 신부자식당         | 247  |
| 오리지널시카고피자     | 183  |
| 대려도           | 157  |
| 봉추찜닭 강남주공공이점  | 131  |
| 리산 삼성타운점      | 112  |
| 봉피양제주 강남점     | 71   |
| Pavox         | 29   |
| 삼거리포차         | 26   |



맛집+회식 관련된 키워드(빈도수)

- 맛
- 맛집
- 메뉴+음식
- 친절+직원
- 때
- 예약
- 가격
- 분위기
- 주전



## 모델링에 사용할 DataFrame

```
RангIndex: 7822 entries, 0 to 7821
Data columns (total 34 columns):
 #   Column      Non-Null Count Dtype  
--- 
 0   review_check    7822 non-null   object  
 1   check_length     7822 non-null   int64  
 2   단타이핑        7822 non-null   int64  
 3   백억하누         7822 non-null   int64  
 4   느린마을         7822 non-null   int64  
 5   화기애애         7822 non-null   int64  
 6   라그릴리아      7822 non-null   int64  
 7   강남진해장       7822 non-null   int64  
 8   구스             7822 non-null   int64  
 9   파파야리프      7822 non-null   int64  
 10  고양이부엌        7822 non-null   int64  
 11  시골야채된장      7822 non-null   int64  
 12  을밀대           7822 non-null   int64  
 13  담소             7822 non-null   int64  
 14  정류벽          7822 non-null   int64  
 15  기리야마         7822 non-null   int64  
 16  시카고           7822 non-null   int64  
 17  대려도           7822 non-null   int64  
 18  신부자           7822 non-null   int64  
 19  봉추찜닭         7822 non-null   int64  
 20  리산             7822 non-null   int64  
 21  봉피양          7822 non-null   int64  
 22  Pavox            7822 non-null   int64  
 23  삼거리           7822 non-null   int64  
 24  맛               7822 non-null   int64  
 25  맛집             7822 non-null   int64  
 26  메뉴_음식         7822 non-null   int64  
 27  친절_직원         7822 non-null   int64  
 28  때               7822 non-null   int64  
 29  예약             7822 non-null   int64  
 30  가격             7822 non-null   int64  
 31  분위기           7822 non-null   int64  
 32  추천             7822 non-null   int64  
 33  helpful          7822 non-null   int64  
dtypes: int64(33), object(1)
```

|   | review_check  | check_length | 단타이평 | 백억하누 | 느린마을 | 화기애애 | 라그릴리아 | 강남진해장 | 구스 | 파파야리프 | ... | 맛 | 맛집 | 메뉴_음식 | 친절_직원 | 때 | 예약 | 가격 | 분위기 | 추천 | helpful |
|---|---|--------------|------|------|------|------|-------|-------|----|-------|-----|---|----|-------|-------|---|----|----|-----|----|---------|
| 0 | 네이버로 예약하고 방문했는데<br>사정이 있어 조용한 구석 자리<br>를 부탁한다고 썼는데도 구...  | 370          | 0    | 1    | 0    | 0    | 0     | 0     | 0  | 0     | ... | 0 | 0  | 1     | 1     | 0 | 1  | 0  | 0   | 0  | 1       |
| 1 | 일반적인 양평 해장국과 다르게<br>기본 간이 되어 있고 베이스가<br>육개장과 뼈해장국 같은...   | 348          | 0    | 0    | 0    | 0    | 0     | 1     | 0  | 0     | ... | 1 | 0  | 1     | 0     | 0 | 0  | 1  | 0   | 0  | 1       |
| 2 | 개인 룸이 있고 룸 안에 화장실<br>이 있어 너무 편하고 좋았는데<br>옆에 개인 서랍이 있는...  | 165          | 0    | 1    | 0    | 0    | 0     | 0     | 0  | 0     | ... | 1 | 0  | 1     | 1     | 0 | 0  | 0  | 0   | 0  | 1       |
| 3 | 분위기도 괜찮고 맛있어요 네이<br>버예약 덕분에 에이드도 서비스<br>로 나왔네요 가격은 비싸지... | 59           | 1    | 0    | 0    | 0    | 0     | 0     | 0  | 0     | ... | 1 | 0  | 0     | 0     | 0 | 1  | 1  | 1   | 0  | 1       |
| 4 | 룸으로 이용했는데 화장실도 딸<br>려 있어서 너무 편했어요 삼합<br>재료들이 신선해서 생각보...  | 276          | 0    | 1    | 0    | 0    | 0     | 0     | 0  | 0     | ... | 1 | 0  | 1     | 0     | 0 | 0  | 0  | 0   | 0  | 1       |

4

”

모델링



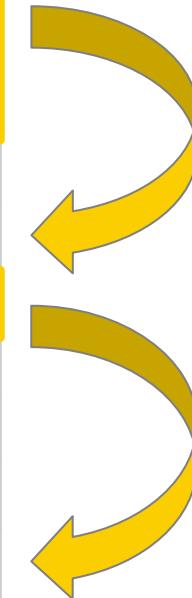
## 04 모델링



문장 1 : 서비스가 좋아서 또 가고 싶어요  
문장 2 : 이 식당 진짜 맛집이에요 좋아요 좋아  
문장 3 : 친절하고 서비스도 좋아요  
문장 4 : 이 식당 강추합니다.

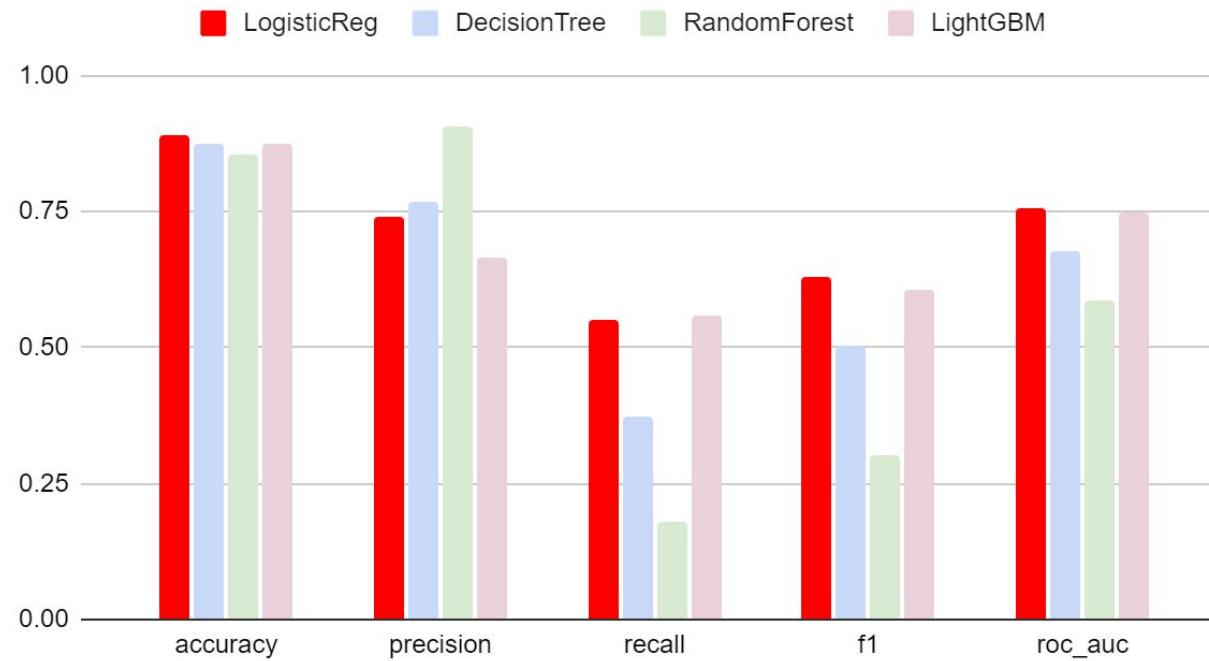
TF-IDF?

|      | 식당             | 맛집             | 진짜             | 친절             | 서비스            | 강추             | 좋아             |
|------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 문장 1 | 0              | 0              | 0              | 0              | 1              | 0              | 1              |
| 문장 2 | 1              | 1              | 1              | 0              | 0              | 0              | 2              |
| 문장 3 | 0              | 0              | 0              | 1              | 1              | 0              | 1              |
| 문장 4 | 1              | 0              | 0              | 0              | 0              | 1              | 0              |
| DF   | 2              | 1              | 1              | 1              | 2              | 1              | 4              |
| IDF  | $\ln(4/(2+1))$ | $\ln(4/(1+1))$ | $\ln(4/(1+1))$ | $\ln(4/(1+1))$ | $\ln(4/(2+1))$ | $\ln(4/(1+1))$ | $\ln(4/(4+1))$ |
|      | 0.2877         | 0.6931         | 0.6931         | 0.6931         | 0.2877         | 0.6931         | -0.2231        |
| 문장 1 | 0              | 0              | 0              | 0              | 0.2877         | 0              | -0.2231        |
| 문장 2 | 0.2877         | 0.6931         | 0.6931         | 0              | 0              | 0              | -0.4462        |
| 문장 3 | 0              | 0              | 0              | 0.6931         | 0.2877         | 0              | -0.2231        |
| 문장 4 | 0.2877         | 0              | 0              | 0              | 0              | 0.6931         | 0              |





모델 비교



|              | accuracy | precision | recall   | f1       | roc_auc  |
|--------------|----------|-----------|----------|----------|----------|
| LogisticReg  | 0.890096 | 0.738693  | 0.550562 | 0.630901 | 0.75525  |
| DecisionTree | 0.874121 | 0.769231  | 0.374532 | 0.503778 | 0.67571  |
| RandomForest | 0.856869 | 0.90566   | 0.179775 | 0.3      | 0.587962 |
| LightGBM     | 0.876677 | 0.665179  | 0.558052 | 0.606925 | 0.750136 |



## Logistic Regression

|           | Acc    | 리뷰 Vectorizer     | 리뷰 길이 | 가게명 | 키워드 |
|-----------|--------|-------------------|-------|-----|-----|
| Feature 1 | 85.88% | Tf-idf Vectorizer | O     | O   | X   |
| Feature 2 | 88.18% | CountVectorizer   | O     | O   | X   |
| Feature 3 | 88.24% | CountVectorizer   | X     | O   | X   |
| Feature 4 | 88.31% | CountVectorizer   | O     | O   | O   |
| Feature 5 | 88.56% | CountVectorizer   | X     | O   | O   |

5

”

결론 및  
에필로그



## 05 결론



|     | store_name | date       | review_check                                      | helpful | result | proba                                      |
|-----|------------|------------|---|---------|--------|--|
| 143 | 와인코르크 강남점  | 2021.03.12 | 솔직히 왜 별점이 높은지 모르겠음 샹그리아도 다른데보다 양과 맛양에 비해 가격이 높... | 1       | 0      | [0.5582672907549633, 0.44173270924503666]  |
| 144 | 와인코르크 강남점  | 2021.03.03 |   | 0       | 0      | [0.9852858614110556, 0.014714138588944412] |
| 145 | 와인코르크 강남점  | 2020.12.19 | 언제나 좋아요 4번째 방문이에요                                 | 0       | 0      | [0.963505961298203, 0.036494038701797016]  |
| 146 | 와인코르크 강남점  | 2020.11.17 | 넓직한 내부에 와인을 골라서 마실 수 있단 게 너무 좋았습니다 가격대도 다양해서 선... | 1       | 0      | [0.9138010473728951, 0.08619895262710488]  |
| 147 | 와인코르크 강남점  | 2020.11.14 | 가성비 좋은 와인집에요                                      | 1       | 0      | [0.9395363700780536, 0.060463629921946434] |
| 148 | 와인코르크 강남점  | 2020.11.03 | 분위기 좋은 곳이에요 많이 시끄럽지 않아서 좋았어요 오늘의 와인도 맛있었습니다       | 1       | 1      | [0.4807821794859657, 0.5192178205140343]   |
| 149 | 와인코르크 강남점  | 2020.10.24 | 좋아하는 곳 포인트 소멸이 빨리되는 게 아쉽다                         | 1       | 0      | [0.9700330933946226, 0.029966906605377402] |
| 150 | 와인코르크 강남점  | 2020.10.05 | 분위기 굿   | 0       | 0      | [0.9274190033238585, 0.07258099667614147]  |
| 151 | 와인코르크 강남점  | 2020.09.21 | 굿   | 0       | 0      | [0.9852858614110556, 0.014714138588944412] |
| 152 | 와인코르크 강남점  | 2020.08.23 | 친절해요  | 0       | 0      | [0.9315581698587309, 0.06844183014126916]  |
| 153 | 와인코르크 강남점  | 2020.08.14 | 가격도 저렴하고 좋아요 웃음 재방문 의사 있습니다                       | 0       | 0      | [0.8415267259437946, 0.15847327405620543]  |
| 154 | 와인코르크 강남점  | 2020.07.29 | 강남역 분위기도 좋고 저렴한 와인 맛집 와인 종류도 많고 원하는 와인 골라서 가져다... | 1       | 1      | [0.07686250520249405, 0.923137494797506]   |
| 155 | 와인코르크 강남점  | 2020.05.04 | 샹그리아 맛집 웃음  | 0       | 0      | [0.9538641206188714, 0.046135879381128585] |
| 156 | 와인코르크 강남점  | 2020.05.01 | 강남에서 제일 좋은 와인 집                                   | 0       | 0      | [0.9759838014281261, 0.024016198571873915] |
| 157 | 와인코르크 강남점  | 2019.12.31 | 와인 가볍게 즐기기 좋은 곳                                   | 0       | 0      | [0.973488916545909, 0.026511083454091082]  |

```

print('전체:', len(final)),
print('맞음:', len(final[final['helpful']==final['result']])),
print('틀림:', len(final[final['helpful']!=final['result']])))
print('확률:', len(final[final['helpful']==final['result']]))

전체: 158
맞음: 135
틀림: 23
확률: 85.44303797468355

```



## [Human performance]

- hanspell로 맞춤법 검사 -> pycospacing + hanspell 맞춤법 검사

### [맞춤법 검사 전 리뷰]

```
text1 = '백얼저복상합 먹었는데 전복식감이 출룡합니다'
text2 = '코시국에 름이 있어 좋습니다'
```

### [1차 시도: 맞춤법]

```
spelled_sent1 = spell_checker.check(text1)
checked_sent1 = spelled_sent1.checked
spelled_sent2 = spell_checker.check(text2)
checked_sent2 = spelled_sent2.checked
print("text1:", checked_sent1), print("text2:", checked_sent2)
```

```
text1: 백얼 저복 상합 먹었는데 전복 식감이 출룡합니다
text2: 고 시국에 름이 있어 좋습니다
```

### [2차 시도: 띄어쓰기 + 맞춤법]

```
spacing = Spacing()
new_sent1 = text1.replace(" ", " ")
kospacing_sent1 = spacing(new_sent1)
spelled_sent1 = spell_checker.check(kospacing_sent1)
checked_sent1 = spelled_sent1.checked
new_sent2 = text2.replace(" ", " ")
kospacing_sent2 = spacing(new_sent2)
spelled_sent2 = spell_checker.check(kospacing_sent2)
checked_sent2 = spelled_sent2.checked
print("text1:", checked_sent1), print("text2:", checked_sent2)
```

```
text1: 백얼 전복 상합 먹었는데 전복 식감이 출룡합니다
text2: 고 시국에 름이 있어 좋습니다
```

- 맞춤법 확인 결과, 추가 오류를 발견하였습니다.
- 따라서 저희가 직접 눈으로 하나씩 보면서 교정하여 문제를 해결하였습니다.



### [Machine Learning]

- 데이터 양을 봤을 때 딥러닝을 먼저 사용하기 보다 머신러닝을 사용하여 결과를 확인하였고, 성능을 높이고자 노력했습니다.
- 결과는 특정 피처 값을 사용하였을 때 88% 이상의 성능이 나와 딥러닝이 아닌 머신러닝 만으로도 충분히 높은 결과를 얻을 수 있었습니다.
- 추후에 데이터를 더 많이 모아 딥러닝을 시도해보는 것도 좋을 것 같습니다.



## Reference

- [\[NLP\] LDA 토픽 모델링을 활용한 앱 리뷰 분석 프로젝트](#)
- [\[ML 공부\] 맛집 리뷰를 통한 감성분석 \(긍정, 부정 언어 판별\) - 210521](#)
- [\[프로젝트\] 호텔 리뷰 텍스트 분석 & 분류 모델 제작](#)
- [PyKoSpacing github](#)
- [Py-hanspell github](#)
- [한국어 맞춤법/문법 검사기](#)
- [위키독스 - 딥러닝을 이용한 자연어처리 입문](#)

감사합니다