
야구장 구글리뷰 분석

20160463 배원빈

목차



1. 프로젝트 준비



2. 과정



3. 결과 및 통계
분석



4. 결론

1. 프로젝트 준비

어떻게?

Colab을 이용한 토큰나이즈 분석기법
Voyant 툴을 통하여 시각적으로 표시
임베딩 벡터를 통한 단어 유사도 측정
실제 기사와 비교

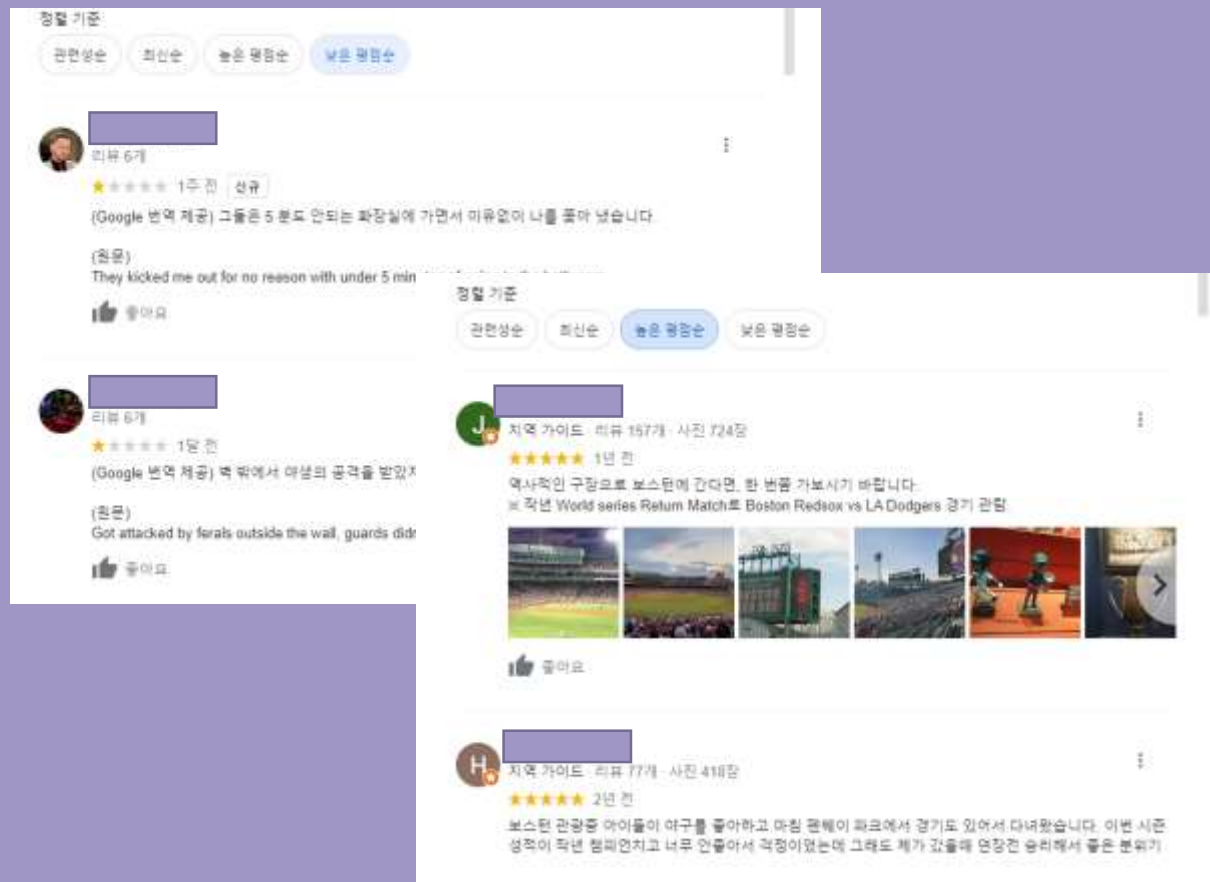
얻을 수 있는 점

구글리뷰 높은/낮은 평점순에 나오는 각각의
표현중에 무엇이 빈도수가 높을까?

다른 대중매체와 구글 리뷰와 비교했을 때, 여론

이 얼마나 공통적인지, 그리고 차이가 어느
정도인지 알 수 있을까? 등등

분석할 구글리뷰들..



분석 대상은 총 3곳

한국 사직구장

미국 텍사스 글로브 라이프 파크

미국 보스턴 펜웨이 파크

Tokenizing English Text

```
# https://www.geeksforgeeks.org/nlp-how-tokenizing-text-sentence-words-works/

# import nltk
# nltk.download('punkt')
# nltk.download('stopwords')
from nltk.tokenize import sent_tokenize
from nltk.tokenize import word_tokenize

[ ] # 내 구글 드라이브에 분석할 텍스트 파일을 업로드 함. (여기 Colab Notebooks 폴더 내)
# 왼쪽에 폴더 모양 아이콘을 눌러서 내 구글 드라이브를 Colab에 탑재함(mount).

# 텍스트 파일을 읽어옴.
with open('/content/txas_baseball_park_review_en.txt') as f:
    text = f.read()

text = text.lower()
text

# https://www.programiz.com/python-programming/methods/string/lower

'beautiful stadium but the feeling inside is more like an airport or a shopping mall. lots of
e that's what they wanted when they designed and built it. the atmosphere outside the stadium
factory or a dying mall with a baseball field in the middle. most of the concession stands are
can't open a ballpark to 100% but then keep most of the stores closed because of covid. and
initially not looking forward to it. now and now there is no seating section in the ballpark yo

[ ] # 문장 단위로 파싱.
import nltk
nltk.download('punkt')

sent_list = sent_tokenize(text)
sent_list

word_list = word_tokenize(text)
word_list
```

Gensim word2vec에서 동일한 단어 임베딩 벡터 생성 방법

```
[1] import nltk
nltk.download('stopwords')

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Unzipping corpora/stopwords.zip.
True

[2] import string
from nltk.corpus import stopwords

[3] text='beautiful stadium but the feeling inside is more like an airport or a shopping mall.',
'lots of seats that are so far away from the field that i end up walking around the concession stands',
'maybe that's what they wanted when they designed and built it.',
'the atmosphere outside the stadium is super. love the bars and restaurants.',
'we're going to have some fun here over the years.',
'the place as a whole feel like a factory or a dying mall with a baseball field in the middle.',
'most of the concession stands are still closed.',
'it's hard to find the food court. or any shops.',
'i didn't see a single vendor selling merchandise anywhere.',
'you can't open a ballpark to 100% but then keep most of the stores closed because of covid, and still
'shame on you rangers.',
'no one was wearing masks and no one was enforcing masks.',
'i might go back, but definitely not looking forward to it.',
'oh and now there is no seating section in the ballpark you have to go outside.',
'i'd rather just watch the game outside on the big screen.',
'that felt more like a ballgame than being in the park.',
'we did have a good view of the field from our seats at least.',
'very nice and new.',
'great setup.',
'the 1st level seats in the 100 range are off the main concourse.',
'the upper level feels like a system of bridges with floor to ceiling windows with some great views.',
'i don't see much here that reminds of the old except the red brick archways that supports the roof tr
'the downside is the food lines and signage.',
'i lost 3 innings waiting for two things and you have to break lines constantly to get around.',
'it will be interesting to see how this plays out over time.'
```

새 시작 여수항 ●●●●●●●●●● ●●●●●●●●●●
 겨울엔 한파도 서늘도 봄기운도 조물조물
 봄에 개작2월작 작곡, 코로나여파로 100%만 업로드 박수로운 응원하였지만 이제 개본 좋았음. 시작여구들
 13대0 그 여대무신한 경기 작곡 다나왔습니다..
 부산 사직 작곡은 참 미안여구 *** 벼룩 밟은 것처럼만ㅠㅠ
 주말 가족나들이로 추천합니다
 승리의 여글스
 롯데자이언츠의 홈구장, 세계에서 가장 뜨거운 응원이 있는 곳
 자이언츠의 세번째 우승을 기대하며
 내년 봄빛이 예쁘게 파서 구경하러 갔는데 4월 1일 현재 그 예쁜 봄빛깔은 땅이 얼어있고 각각의 나무에서
 승리자이언츠 응원 땀방울
 교회 사람들이 롯데 키움여포즈 경기관람을 했다.
 최근에 부진하여 많이 없었던 롯데가
 오늘 4:1로 승리하였다.
 난 여구경기장에 자주는 안가지만
 갈때 마다 롯데가 기기는 이번이 멀어났다. 롯데에서 장으로 날 응원하여
 줄거야 아니?ㅋㅋ
 홈러온 관람 여었다.
 여 구장에도 반해 넓습니다. 부산하면 롯데 홈구장이라 응원보는 재미도 좋았어고 이랜드도 응원하. 부
 가록 만취회 합니까 좋은 곳입니다..
 구는 부산 마미고-
 마미고 부산의 여구 명칭이 시작됩니다. 롯데 자이언츠의 홈구장으로 수만명의 목소리가 하나가 되어 울려 퍼지
 캄필라야 사람이 많이 없고 완충존 선택할 수 있고 주차공간이 넓어져 생활용품리무어 용기가 구비되어 있
 어서 마미고는 언제까지도 거뜬할거야
 사직 여구장(沙池野球場)은 대한민국 부산광역시 중구에 사직동에 있는 여구장이다. 1985년부터 한국 프로야
 구인시현의 여구시현은상상 호텔
 현재 공동5위임에도 응원은 격렬할
 말을 안드려도 굳이...ㅎㅎ
 구인시현은말을 잘듣는 거 같 습니다. 지루한 일상속에서 힐링하고 싶었습니다.
 롯데 자이언츠 파이팅 !!
 라이프도 너무 좋습니여구, 그리고 중간중간 지루하지않게하는 이랜드도 재미있었여구♥
 요한 잘하고 있냐? 롯데
 자주가는 장소 미어들과 함께하기 좋은 장소

```

1 ipio install konlpy

from konlpy.tag import Hannanum
from collections import Counter

[6] with open('/content/시작 야구장 낮은말집순.txt') as f:
    text = f.read()
    text

'경기장은 낮춰지고 관중은 술먹는 위풍. 술먹구에서 술연한다. 시민의식도 왕복이런것같장. 실다
하지. ㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎㅎ
가리 구마를 달아온 해서 통고름. 그래서인지 야구장 내 매장을 음식들도 다른 야구장에 비해
격기통한 있으면 담배꽂이. 화장실 매우 좁음. 여성들 화장실 중 매우 김. 화장실청소도 붓대용
좌석도 너무 비좁음. 화장실화장품 사먹는 적인 그대로 편의시설은 대전보다 훨씬 잘 되어있음. 적
바짐. 해운이 가까 좁. 위험해술에 신세끼면야구장하자. 음식소만알고 술마시려 가는곳. 야구장
너무 재미있네요 ㅎㅎㅎ시간 전부터 표 팔아요. 화왕떡가져 마삼. 해의지가 너무 작아서 남자를 나
만 연애헤이 없음. 주변도로로 가깝게 이동가능한 연애헤음해사해한 야구 보고왔어요. 차마
'

[9] # 천다를 셀테소 분석기

hnn = Hannanum()
nouns = hnn.nouns(text)

count = Counter(nouns)
print(count)

Counter({'야구장': 11, '롯데': 11, '야구': 10, '좌석': 5, '경기': 5, '불편': 4, '사실': 4

```



구글 리뷰 텍스트화
리뷰 카피 페이스트

Colab을 통한 토큰나이징 그리고 임베딩 벡터 사용

Voyant 툴을 이용
직접적인 문장 비교를 통해
임베딩 벡터값의 신뢰성 파악

Voyant와 토큰나이저를 동시에 확인해본 이유는 각 토큰마다 차이가 있음을 파악하고, 분석의 다양화를 위해서였다. 그리고 표본은 200개 정도의 문장리뷰를 대상으로 하였다. 리뷰없이 평점만 남긴 리뷰도 많았기에 직접 판별하여 문장을 선택하는 작업을 하였다.

Colab 코드

한글 및 영어 리뷰에 따른 코드들

❖ **1. gensim을 이용한 한글 구글리뷰 분석

```
[ ] # KoNLPy 설치
!pip install konlpy

[1] from konlpy.tag import Kkma

[ ] with open('/content/사직 야구장 농운평점순.txt') as f:
    text = f.read()
    text

[ ] # 문장 단위로 끊음
aa = Kkma()
sentences = aa.sentences(text)

[ ] print(sentences)

[ ] # 각 문장에서 명사들만 취함
word_list = []
for sen in sentences:
    word_list.append(aa.nouns(sen))

[ ] print(word_list)
```

❖ 2. gensim을 이용하여 비슷한 단어(context word) 발견

<https://radimrehurek.com/gensim/models/word2vec.html>

```
[ ] from gensim.models import Word2Vec

[ ] model = Word2Vec(word_list, size=10, min_count=1, sg=1, workers=1)

[ ] # 고유 단어 목록 확인
```

▶ len(vocab)

[] model.wv['최고']

▶ model.wv.most_similar('최고', topn=6)

❖ 한국어 텍스트의 토큰화 작업

한국어 텍스트 KoNLPy라는 파이썬 한국어 NLP 패키지를 사용합니다.

<https://konlpy-ko.readthedocs.io/ko/v0.4.3/>

더블클릭 또는 Enter 키를 눌러 수정

```
[ ] !pip install konlpy

from konlpy.tag import Hannanum
from collections import Counter

[ ] with open('/content/사직 야구장 농운평점순.txt') as f:
    text = f.read()
    text
```

▶ 한나눔 형태소 분석기

```
hnn = Hannanum()
nouns = hnn.nouns(text)

count = Counter(nouns)
print(count)
```

❖ Review Tokenizeing

```
[3] from nltk.tokenize import sent_tokenize
from nltk.tokenize import word_tokenize
```

▶ # 텍스트 파일을 읽어들이고,
with open('/content/texas baseball park review en.txt') as f:
 text = f.read()

text = text.lower()
text

<https://www.programiz.com/python-programming/methods/string/lower>

```
[26] # 문장 단위로 쪼갬
import nltk
nltk.download('punkt')
```

```
sent_list = sent_tokenize(text)
sent_list
```

```
[28] word_list = word_tokenize(text)
word_list
```

```
[29] # https://www.kite.com/python/answers/how-to-remove-all-punctuation-marks-with-nltk-in-python  
# 알파벳과 숫자만 추출  
  
new_words = [word for word in word_list if word.isalnum()]  
new_words
```

```
[30] # https://stackoverflow.com/questions/20510768/count-frequency-of-words-in-a-list-and-sort-by-frequency  
# 단어별 빈도를 계산.  
  
from collections import Counter  
counts = Counter(new_words)  
print(counts)
```


Colab 코드

한글 및 영어 리뷰에 따른 코드들

```
[31] # https://www.geeksforgeeks.org/removing-stop-words-nltk-python/  
# 불용어 목록을 읽어들이.  
  
nltk.download('stopwords')  
  
from nltk.corpus import stopwords  
print(stopwords.words('english'))  
  
[32] # 불용어를 제거함(filter).  
  
stop_words = set(stopwords.words('english'))  
  
# filtered_sentence = [w for w in new_words if not w in stop_words]  
  
filtered_sentence = []  
  
for w in new_words:  
    if w not in stop_words:  
        filtered_sentence.append(w)  
  
print(filtered_sentence)  
  
[33] # 불용어를 제거한 단어의 빈도 계산.  
  
filtered_counts = Counter(filtered_sentence)  
print(filtered_counts)
```

▸ Gensim word2vec에서 동일한 단어 임베딩 벡터 생성 방법

```
[25] import nltk  
from nltk.tokenize import sent_tokenize  
nltk.download('stopwords')  
  
[26] import string  
from nltk.corpus import stopwords  
  
[27] with open('/content/tejas baseball park review en.txt') as f:  
    text = f.read()  
  
    text = text.lower()  
    text  
  
[33] import nltk  
nltk.download('punkt')  
  
sent_list = sent_tokenize(text)  
sent_list  
  
[34] def preprocessing(sent_list): #corpus  
    stop_words = set(stopwords.words('english'))  
    training_data = []  
    sentences = sent_list  
    sentences = list(filter(None, sentences))  
    for i in range(len(sentences)):  
        sentences[i] = sentences[i].strip()  
        sentence = sentences[i].split()  
        x = [word.strip(string.punctuation) for word in sentence  
              if word not in stop_words]  
        x = [word.lower() for word in x]  
        training_data.append(x)  
    return training_data  
  
#corpus.split(" ")
```

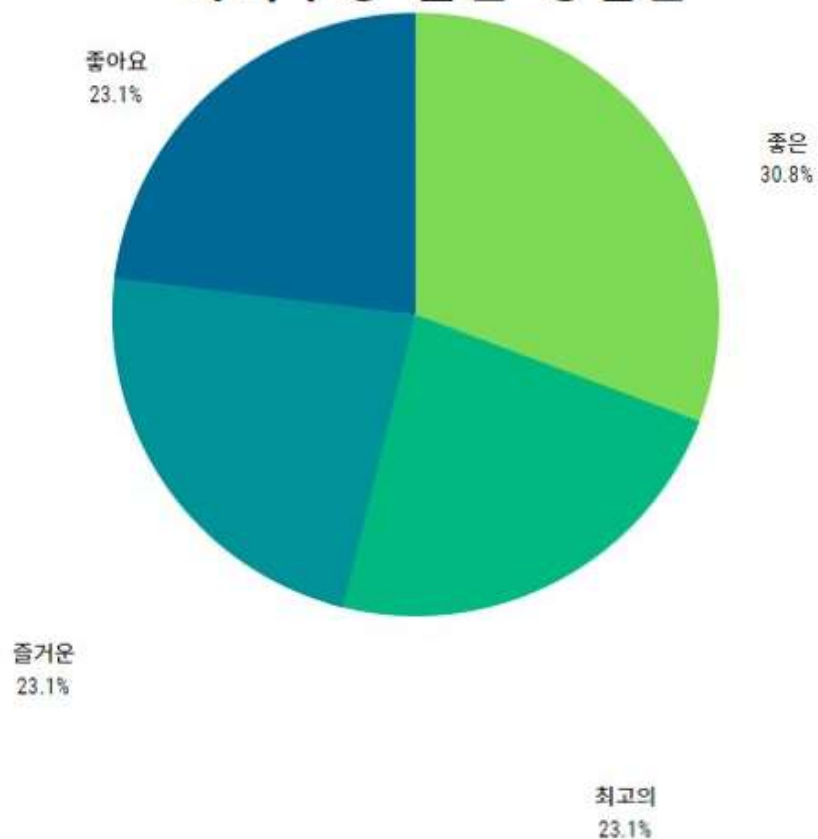
```
train_data = preprocessing(sent_list) #corpus  
  
print(train_data)  
  
[ ] # train_data  
  
[39] from gensim.models import Word2Vec  
  
[40] model = Word2Vec(train_data, size=100, min_count=1, sg=1, workers=1, seed=0)  
  
[44] model.wv.most_similar('good', topn=10)
```

Corpus가 아닌 문장 단위로 쪼개는 sent_tokenize 방식을 통해 문장을 나누고, 임베딩 값을 측정하고자 했다.

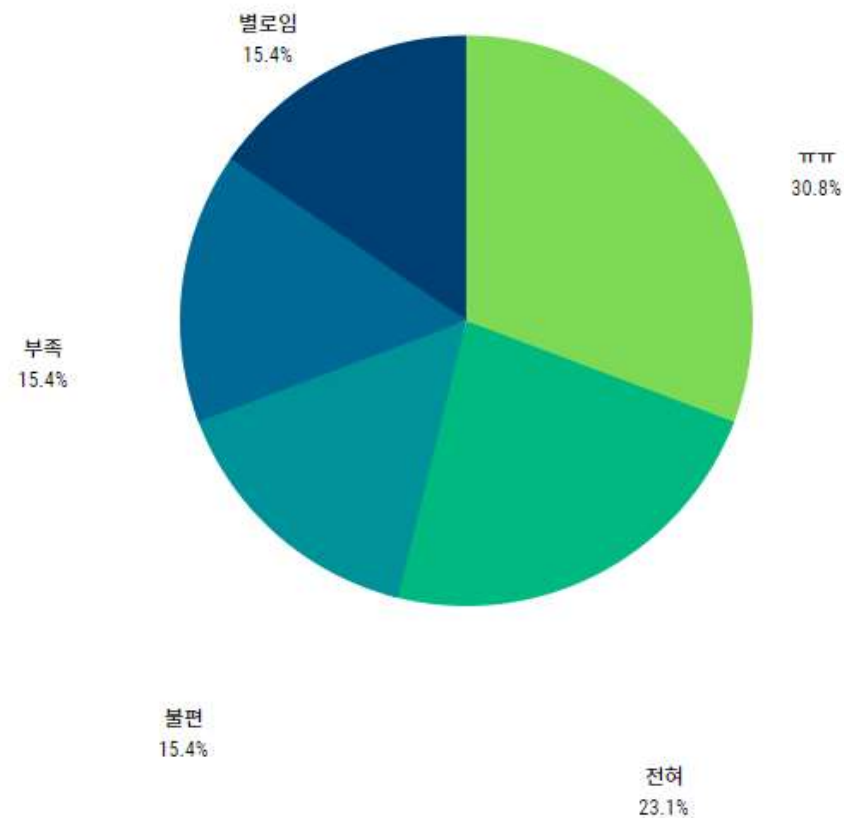
3. 결과

사직구장 구글리뷰의 높은 평점, 낮은 평점 각각에 주로 나오는 편향된 표현을 그래프로 정리

사직구장 높은 평점순

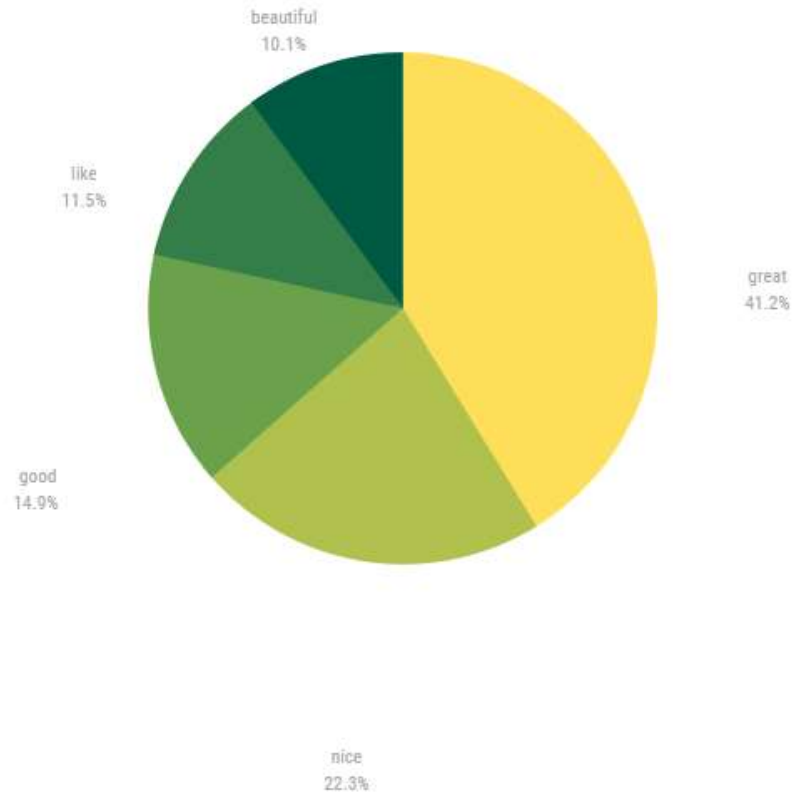


사직구장 낮은 평점순



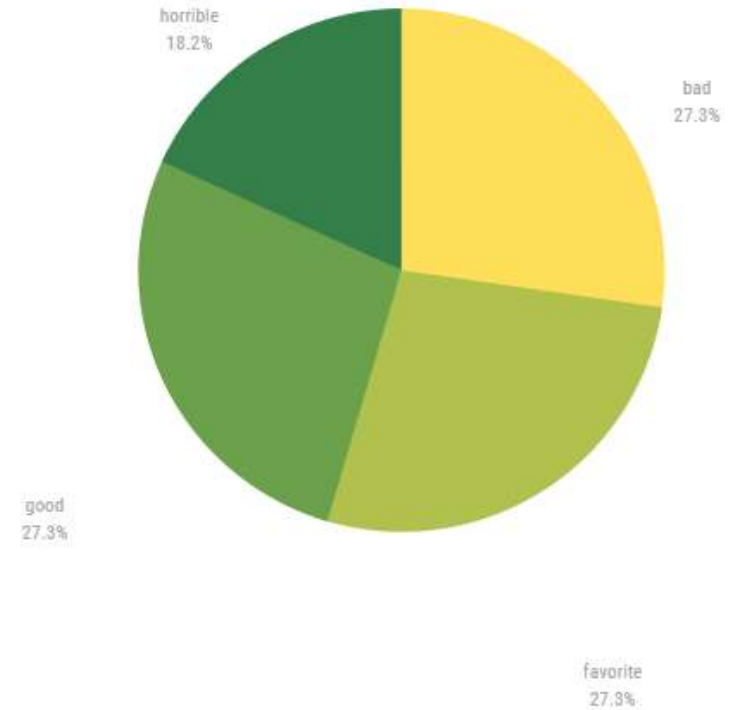
텍사스 구장 구글리뷰의 높은 평점, 낮은 평점 각각에 주로 나오는 편향된 표현을 그래프로 정리

텍사스 글로브 파크 높은 평점순



Like는 확인결과 '~와 같은'이라는 뜻으로 주로 쓰여서, 후에 분석할 때에는 빼기로 함.

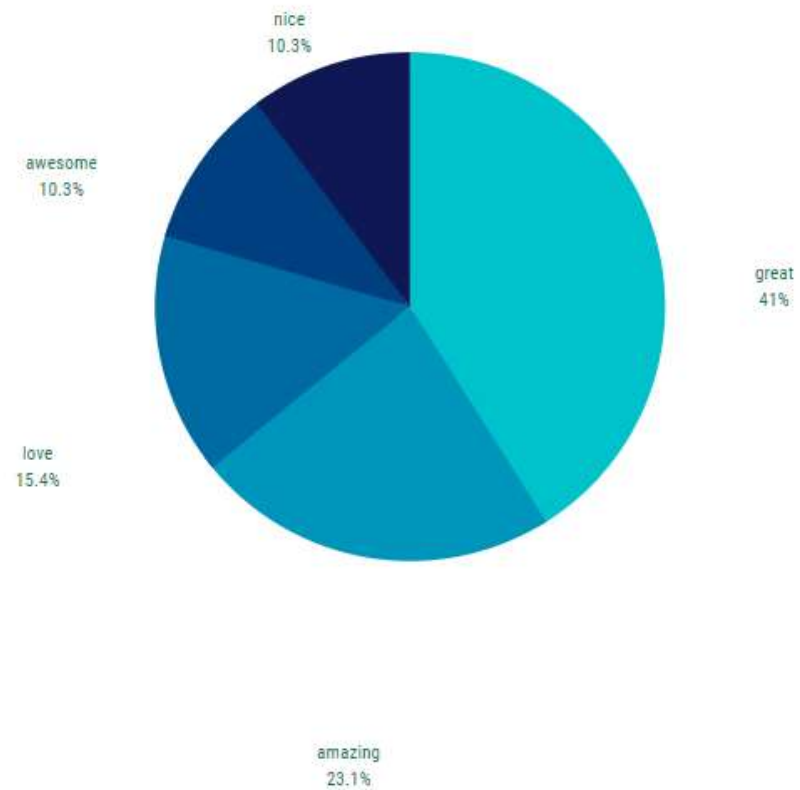
텍사스 글로브 파크 낮은 평점순



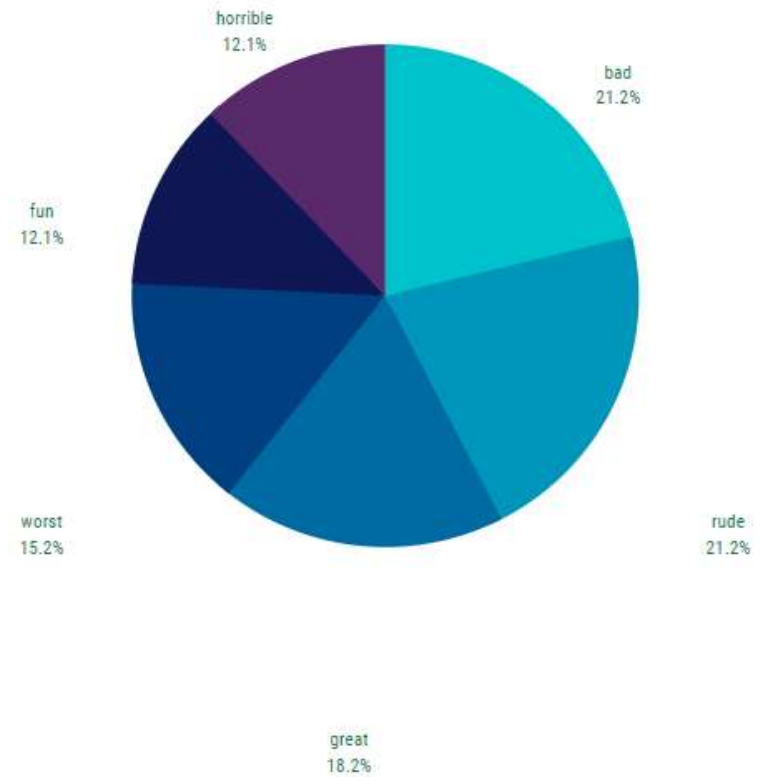
여기에 나온 good은 not good이 대부분. 그러나 몇몇 good은 긍정의 표현. Favorite은 least와 함께 주로 쓰임.

펜웨이 파크 구글리뷰의 높은 평점, 낮은 평점 각각에 주로 나오는 편향된 표현을 그래프로 정리

펜웨이 파크 높은 평점순



펜웨이 파크 낮은 평점순



Voyant를 통한 표현의 쓰임새 알아보기

Ex) 가족 단위로 관람하기 **좋은** 곳입니다.

그래서인지 야구장 내 매장들 음식들도 다른 야구장에 비해 **별로임**.

문장단위로 분리해 본 결과, 사직구장의 높은 평점순에서 나온 긍정적인 표현들은 공통적으로 **장소, 응원** 등에 쓰이며, 이를 통해 높이 평가하고 있다.

반면, 낮은 평점순에서 나온 부정적인 표현들은 **낙후된 시설, 청결하지 못한 위생상태**등에 쓰이며, 이를 통해 낮게 평가하고 있다.

단, 높은 평점순에서 나온 리뷰가 무조건 긍정적인 표현만 있지는 않았다.

그리고 Voyant에서 직접 확인한 문장들에 쓰인 표현과 임베딩하여 얻은 각 표현의 유사 표현을 비교해보면 어느정도 비슷하다는 것을 알 수 있으니, 이는 임베딩한 값이 어느정도 신뢰받을 수 있다고 할 수 있겠다.

```
model.vw.most_similar('최고', topn=6)
```

```
[('정', 0.8358801603317261),  
 ('문구', 0.8034607172012329),  
 ('1패', 0.8001145720481873),  
 ('공간', 0.6879863142967224),  
 ('음악', 0.6763362884521484),  
 ('구도', 0.6357241868972778)]
```

```
model.vw.most_similar('불편', topn=6)
```

```
[('바퀴벌레', 0.7883530259132385),
```

```
model.vw.most_similar('부족', topn=6)
```

```
[('미', 0.7913756966590881),  
 ('개선', 0.7707507014274597),  
 ('정체', 0.768353283405304),
```

Voyant를 통한 표현의 쓰임새 알아보기

Ex) Great food. And what a great park to see a game.

Least favorite is the dome itself.

Paid \$55.00 all you can eat food was trouble very sorry we were so disappointed

텍사스 글로브 파크의 경우, 높은 평점순에서 나온 편향된 표현들은 주로 야구를 보기 좋은 **경관**, **좌석**등을 높이 평가할 때 쓰였고,

낮은 평점순에서 나온 편향된 표현들은 주로 **음식**이나 **티켓 가격대**를 향한 **불만**, **불편한 시설** 등을 낮게 평가할 때 쓰였다.

임베딩 벡터값에도 어느정도 반영되어 있다. 그리고 높은 평점순에는 주로 레인저스와 유사하다는 값이 많은 것으로 보아, 텍사스 경기장이다 보니 텍사스 팬들이 좋은 리뷰를 남기는 경우가 있다고 느꼈다.

```
model.wv.most_similar('nice', topn=10)
```

```
[('hang', 0.8516708016395569),  
 ('king', 0.8347517848014832),  
 ('park', 0.8217498064041138),
```

```
model.wv.most_similar('great', topn=10)
```

```
[('us', 0.7823994159698486),  
 ('won', 0.7715296745300293),  
 ('rangers', 0.7648227214813232),  
 ('open', 0.7597079277038574),  
 ('even', 0.7587273120880127),  
 ('86', 0.7521730661392212),  
 ('seat', 0.7415736317634583),
```

```
model.wv.most_similar('favorite', topn=10)
```

```
[('behind', 0.7939848303794861),  
 ('lower', 0.7689269781112671),  
 ('district', 0.753911018371582),  
 ('price', 0.7399929761886597),  
 ('cream', 0.7156827449798584),
```

```
model.wv.most_similar('bad', topn=10)
```

```
[('one', 0.8462221622467041),  
 ('cozy', 0.8389073610305786),  
 ('10pm', 0.8318966627120972),  
 ('glove', 0.7831375002861023),  
 ('access', 0.7758855223655701),
```

```
model.wv.most_similar('horrible', topn=10)
```

```
[('porch', 0.8024390339851379),  
 ('put', 0.7958182692527771),  
 ('pepper', 0.7919301986694336),  
 ('play', 0.7641103267669678),  
 ('empty', 0.7473374009132385),  
 ('there's', 0.7364282011985779),  
 ('center', 0.7266366481781006),  
 ('miss', 0.7215284109115601),  
 ('sound', 0.7098920345306396),
```

Voyant를 통한 표현의 쓰임새 알아보기

Ex) Sam was a **great** guide with a lot of knowledge regarding the park and team.
Great guide and loved the grounds, with all that history.
Security guards and employees extremely **rude**.

가장 오래된 야구역사를 가진 펜웨이 파크 리뷰는 높은 평점순에서는 편향된 표현이 주로 **역사 가이드나 역사, 그리고 경기장 자체**에 쓰였으며 이를 높이 평가하고 있었고,

낮은 평점순에서는 편향된 표현이 주로 **경기장에 들어서는 팬분들이나 스태브를 향한 불만, 낙후된 시설**에는 낮게 평가하며 쓰였고, 경기장 자체에는 의외로 높게 평가하는데에 쓰였다.

이로서 알게 된 것은 평점순에 따라 대체로 편향된 표현들이 극단적으로 갈리는 경우가 있었지만, 그럼에도 높게 평가하고자 하는 건 높게, 낮게 평가하고자 하는 건 낮게 평가하려는 리뷰도 존재하다는 점이였다. 그리고 영어리뷰는 직접적으로 불만을 언급한다기 보다는, 돌려서 말하는 느낌이 있었다.

```
model.wv.most_similar('great', topn=10)
```

```
[('hard', 0.8150391578674316),  
 ('ruin', 0.7711430788040161),  
 ('honestly', 0.6944494247436523),  
 ('super-friendly', 0.6895115375518799),
```

```
model.wv.most_similar('amazing', topn=10)
```

```
[('blocks', 0.8462756872177124),  
 ('clear', 0.8130497932434082),  
 ('magical', 0.7554185390472412),  
 ('', 0.739719808101654),  
 ('pre', 0.6937789916992188),  
 ('looked', 0.6924610733985901),  
 ('park', 0.6816368699073792),
```

```
model.wv.most_similar('bad', topn=10)
```

```
[('one', 0.8373358249664307),  
 ('access', 0.786113440990448),  
 ('to', 0.7752424478530884),  
 ('toward', 0.7706998586654663),  
 ('human-size', 0.7364516258239746),  
 ('life', 0.7336831092834473),  
 ('throw', 0.7268449664115906),  
 ('bleachers', 0.7006937861442566),
```

```
model.wv.most_similar('rude', topn=10)
```

```
[('longest', 0.8618735074996948),  
 ('concession', 0.8150543570518494),  
 ('bucks', 0.8142864108085632),  
 ('lost', 0.8129188418388367),  
 ('stadium', 0.7971088290214539),  
 ('money', 0.7626392841339111),  
 ('racist', 0.7621638178825378),
```

언론 기사와 비교해보면...

부산시, 사직야구장 재건축 등 장기표류사업 신속해결 나선다

파이낸셜뉴스 입력 2021.06.07 14:33 수정 2021.06.07 14:33

스포츠 >

[부산현장]한용덕 경기감독관 "사직 외야에 물 철벽철벽, 부상 우려"(직격인터뷰)

낙후된 시설에 우려를 표현하고 있으며, 기관에서는 이를 개선시키
고자 하고있고, 야구팬들은 응원을 하러 다니는 모습이 기사에 나오
는 것을 통해, 구글리뷰에서 사람들이 생각했던 것이 미디어에도 비
슷하게 나오는 것으로 보아 이는 구글 리뷰가 다른 매체와도 연결되
는 여론이라고 볼 수 있다는 점을 시사.



포토출

사직구장 찾아 응원하는 야구팬

언론기사와 비교해보면...

TEXAS RANGERS

Outside Food, Drinks Allowed at Globe Life Field Soon

The changes go into effect Friday

Published June 2, 2021 • Updated on June 2, 2021 at 3:11 pm



텍사스 구장은 음식을 향한 사람들의 불만 때문인지 6월 2일경에 외부 음식을 반입해도 된다는 기사를 내어 불만 해소를 꾀하고자 했다.

Fans will be allowed to attend NLCS, World Series at Globe Life Field; tickets on sale next week

MLB will make approximately 11,500 tickets available for each game.

그리고 티켓 가격과 관련된 불만은 아무래도 작년에 유일하게 열렸던 신 구장이었기에, 그리고 월드 시리즈나 NLCS 경기가 열리는 경기장이었기에, 희소가치가 올랐기 때문인지 티켓팅에 실패했거나, 비싸서 가지 못한 사람들의 불만이 대부분이었다고 본다.

언론 기사와 비교해보면...

Fan escorted from Fenway Park for unfurling 'Trump won' banner on outfield wall

Share



WCVB 5 abc

Updated: 9:12 PM EDT Jun 7, 2021

최근 기사에 따르면 "트럼프 승"이라는 정치와 관련된 문구가 펜웨이 파크에 걸린적이 있었는데, 펜웨이 파크에 입장하는 부분적인 팬들에게 향한 불만이 왜 있는가를 조금이나마 이해할 수 있었다.

Green Monster ladder: Fenway quirk,

December 5th, 2020



Ian Browne
@ianMBrowne

Share



펜웨이 파크의 크나큰 특징인 외야 벽 '그린몬스터'. 이에 팬들은 리뷰에서 정말 크다, 어메이징 하다는 이야기를 주로 했었는데, 기사 제목에 나와있듯이 '별났다'고 할 정도로 호불호가 갈린다. 구장의 역사와 구장 자체는 인정하는 분위기이지만, 이 거대한 벽에는 수많은 사연이 있다보니, 여론도 애매한 평가를 내린다고 본다.


4. 결론

1. 경기장 리뷰에서의 높은 평점순에는 긍정적인 표현이, 낮은 평점순에는 부정적인 표현이 주로 쓰였다는 것을 알 수 있었다. 그리고 구글 리뷰는 자유로운 의견 표출 공간이었다.

2. 임베딩 벡터로 분석한 값과 실제 텍스트를 비교해 보았을 때, 유사도 측정이 세밀하게는 차이가 났지만, 의미가 어느정도 일치하는 부분이 있었다.

3. 구글 리뷰와 언론 기사와의 비교를 통하여, 대중 매체의 여론이 구글 리뷰와 어느정도 일치할 수 있다는 점을 알 수 있었다.

4. 야구장을 높이 평가하는 이유는 야구장 시설 자체를 평가하는 것 뿐만이 아닌, 야구장에 가는 것 자체에(가족과 함께 가거나 애인과 가는 등) 의미를 높게 부여할 수도 있다는 점을 알 수 있었다.



감사합니다!