

계획적으로

함께

끝까지

데이터 분석가에

적합한 역량

계획적으로

= 목표에 이를 수 있도록 발판을 만드는 계획성

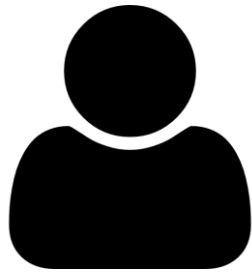
함께

= 각자 맡은 바 책임감 있게, 서로 소통하며

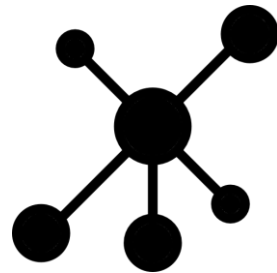
끝까지

= 원하는 결과를 얻을 때까지 포기하지 않는 끈기

Contents



**ABOUT
ME**



MY SKILL



PROJECT

ABOUT ME

PROFILE

송현근

Song Hyun Keun

1996.09.18

010-4050-5823

hksong5823@gmail.com

SCHOOL

2012.03 ~ 2015.02 안산 양지고등학교

2015.03 ~ 2021.08 송실대학교 정보통계보험수리

CERTIFICATION

2021.04.09 데이터 분석 준전문가 ADsP

2021.01.29 자동차 운전 면허 (2종 보통)

PERSONAL ACTIVITY



<https://github.com/songhk8634>

MY SKILL



PROJECT

1. 뉴스 큐레이션 서비스 (NewsCat)

2. LH 구내식당 식수 인원 예측

3. 날씨에 따른 소비패턴 분석

4. 사용자 맞춤 웹툰 추천 서비스

뉴스 큐레이션 서비스 (NewsCat)

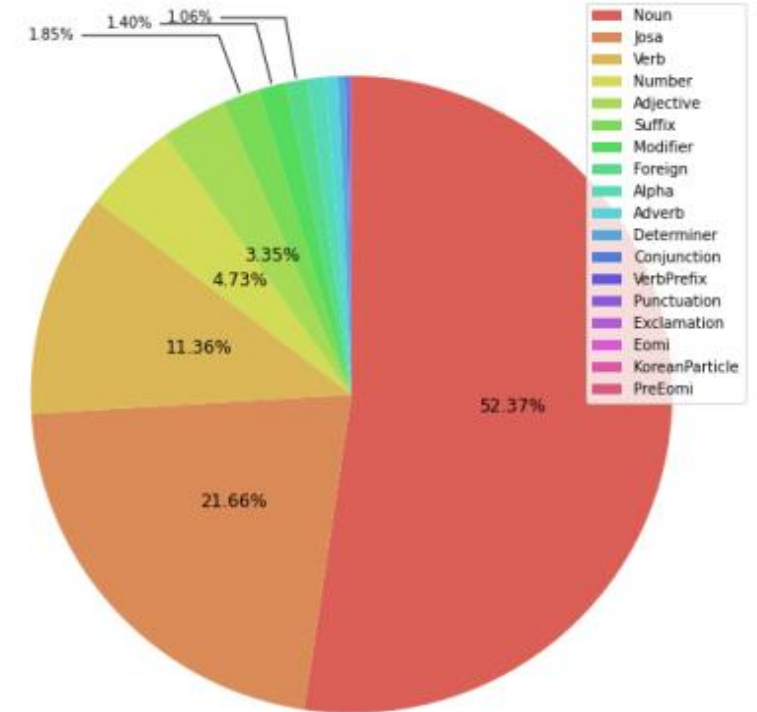


프로젝트 목적

- 포털 사이트 개입 없이 사용자의 선호도만 고려하는 뉴스 추천 서비스 제공

프로젝트 내용

- 6개 신문사(추가 가능)에서 각 5개 분야(정치, 경제, 사회, 국제, 스포츠)뉴스 기사 크롤링
- 보도자료 형식으로 배포된 중복 기사 제거
- 형태소 분석 후 Tf-idf 활용하여 각 기사의 키워드 추출
- 키워드 간 코사인 유사도 계산
- 사용자 맞춤 기사 3개, 다양한 뉴스 제공을 위한 랜덤 기사 2개 제공
- 유저 정보가 없는 '콜드 스타트' 해결 위해 처음에는 선택한 카테고리 내에서 랜덤하게 뉴스 제공



TF(d, t)

특정 문서 d에서의 특정 단어 t의 등장 횟수

DF(t)

특정 단어 t가 등장한 문서의 수

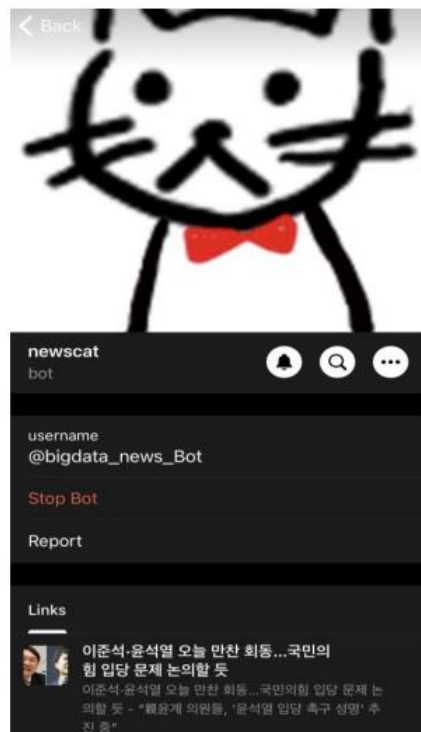
IDF(d, t)

DF(t)에 반비례하는 수

$$TF(d, t) * IDF(d, t) = TF-IDF(d, t)$$

$$idf(d, t) = \log\left(\frac{n}{1 + df(t)}\right)$$

뉴스 큐레이션 서비스 (NewsCat)



MZ세대의 성향을 반영한
B급 감성의 뉴스캐트 컨셉



뉴스 카테고리 선택



추천 뉴스 제공



서비스 피드백(만족도)

뉴스 큐레이션 서비스 (NewsCat)



앞으로 발전 방향

- 유저 증가 & 유저 피드백 log 누적
=> 협업 필터링 사용
=> 추천 정확도 향상
- 카테고리 & 해시태그 활용
=> 특정 분야 전문 뉴스 큐레이션 서비스 가능
ex) 해외 축구 이슈, 주식

협업 필터링

Collaborative Filtering

① 두 사용자 모두 읽음



② 비슷한 사용자



User A



User B

③ A가 읽음



④ B에게 추천

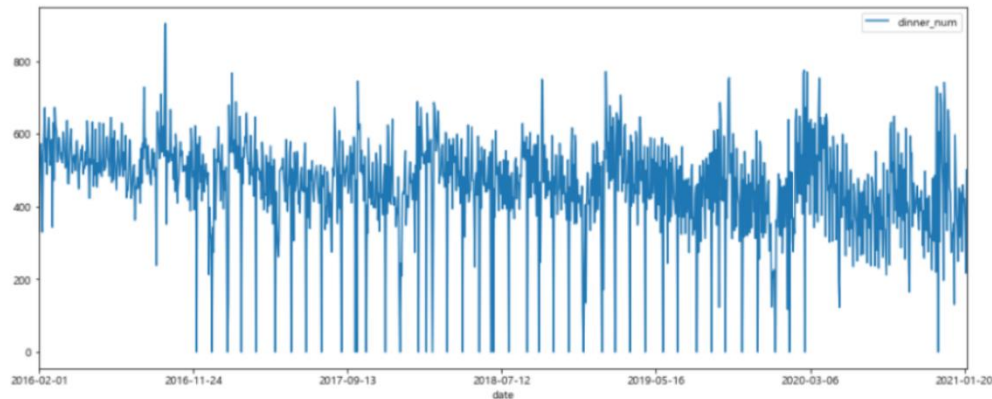
LH 구내식당 식수 인원 예측

프로젝트 목적

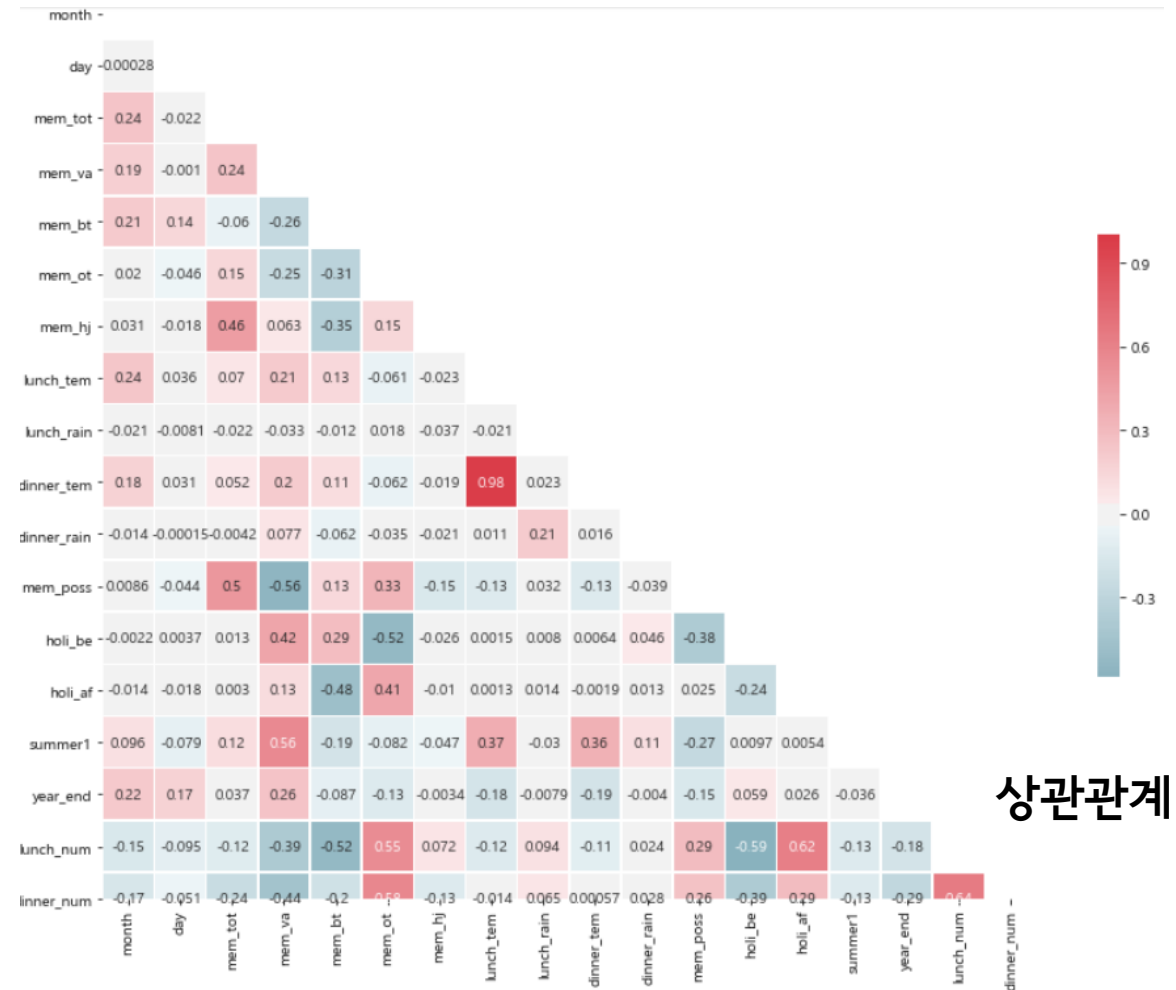
- LH 구내식당 식수 인원을 예측하여 잔반 발생량 감소
- 다른 기업들에서 정확한 식수 인원 예측을 통해 잔반 처리 비용 절감

프로젝트 내용

- LH에서 제공한 식수 인원 데이터와 해당 기간 지역 기상 데이터를 수집, 병합
- => 데이터 전처리, EDA, 상관분석, 가설검정, 예측



EDA



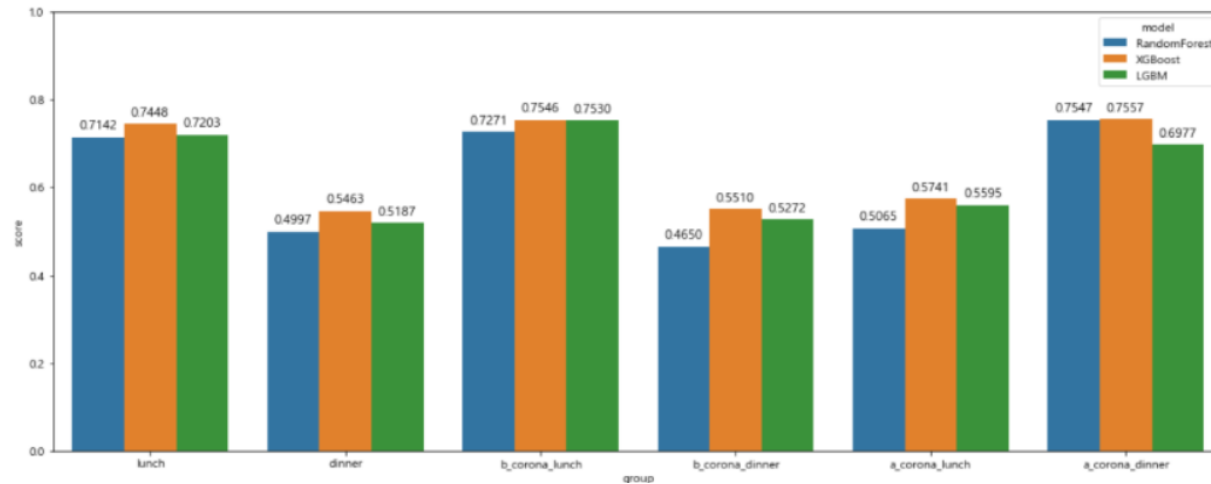
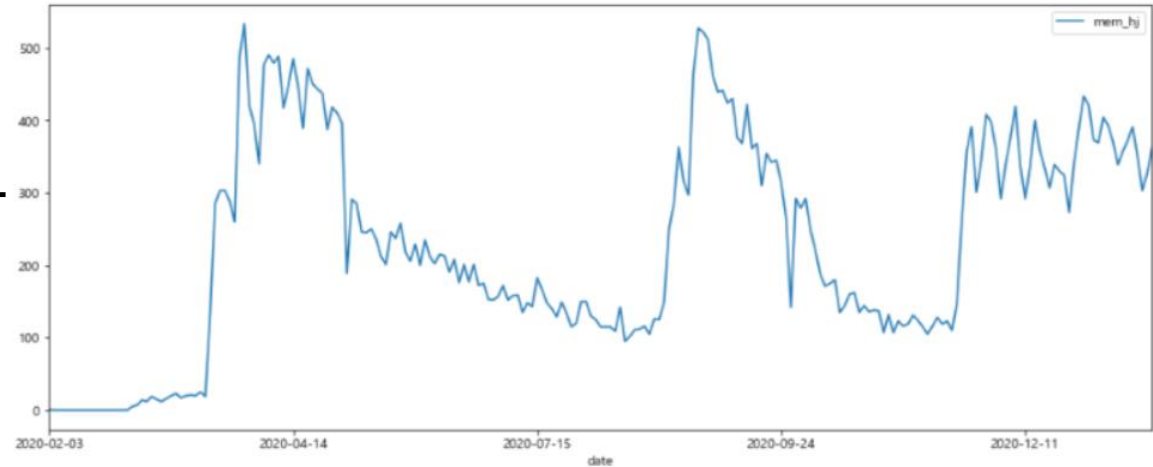
상관관계

내 구내식당 식수 인원 예측

```

=====
train_mon
1.0      vs train_mon
7.460344657903165e-47      vs train_tue
3.138241722033116e-46      vs train_wed
1.360769834405051e-82      vs train_thu
6.201291395979666e-139      vs train_fri
=====
train_tue
7.460344657903165e-47      vs train_mon
1.0      vs train_tue
0.13853103300217712      vs train_wed
1.1689702494479013e-18      vs train_thu
8.540662591390006e-86      vs train_fri
=====
train_wed
3.138241722033116e-46      vs train_mon
0.13853103300217712      vs train_tue
1.0      vs train_wed
4.7957244870189515e-12      vs train_thu
2.1986296403397977e-69      vs train_fri
=====
train_thu
1.360769834405051e-82      vs train_mon
1.1689702494479013e-18      vs train_tue
4.7957244870189515e-12      vs train_wed
1.0      vs train_thu
1.5760907265634135e-41      vs train_fri
=====
train_fri
6.201291395979666e-139      vs train_mon
8.540662591390006e-86      vs train_tue
2.1986296403397977e-69      vs train_wed
1.5760907265634135e-41      vs train_thu
1.0      vs train_fri
=====
    
```

코로나 확산시기에 따라
증가하는 재택근무인원

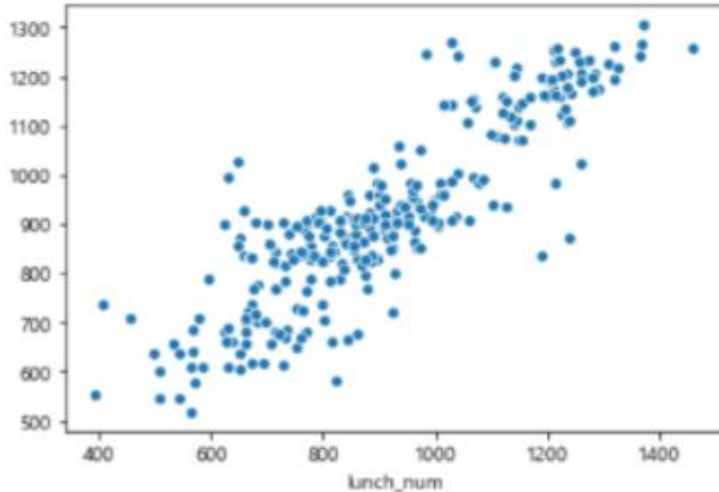


예측 모델 비교

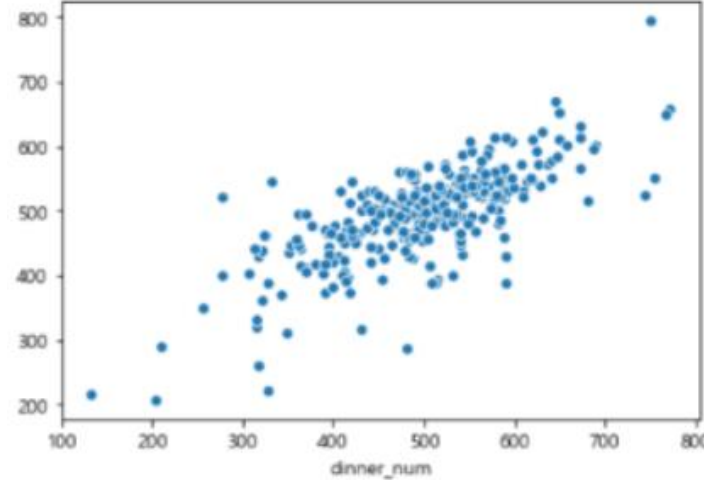
요일간 식수인원에 차이가 있는지 가설검정

내 구내식당 식수 인원 예측

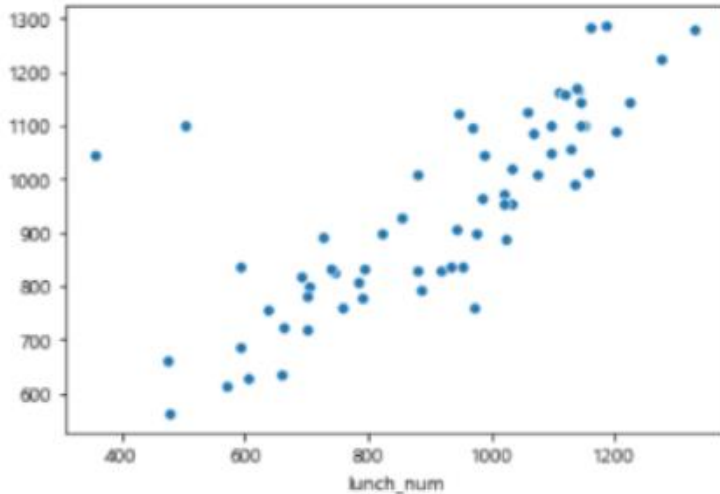
코로나 이전 점심 식수인원 예측 score : 0.7545933443756944



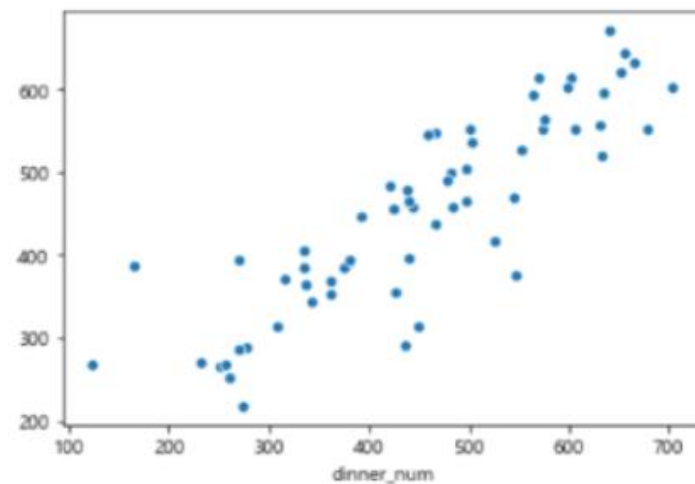
코로나 이전 저녁 식수인원 예측 score : 0.5509644717971985



코로나 이후 점심 식수인원 예측 score : 0.5740874145466672



코로나 이후 저녁 식수인원 예측 score : 0.7557081191168687



앞으로 발전 방향

- parameter tuning을 통해 더 **정확**한 예측 기대
- 내 근처 식당 데이터 **자연어 처리**를 통해 메뉴가 겹칠 경우 인원 분석

날씨에 따른 소비패턴 분석

프로젝트 목적

- 날씨에 따라 연령별/성별 소비자들의 소비 트렌드를 분석하여 마케팅적 요소에 사용 가능

프로젝트 내용

- 지역, 연령층, 성별, 상품 분류, 판매량 등으로 이루어진 구매 데이터 (2018.01.01 ~ 2019.12.31)
- 기상 데이터 (지역별 기온, 강수량, 습도, 적설량, 미세먼지)
- 지역별 가중치 위한 주민등록 인구 데이터
- 물품별 수요예측

기준

9°C 이상 9°C 미만

습도

80% 이상 60% 미만

메세인지

나쁨

나이

40

성별

F

상품유형

●식품●냉난방가전●뷰티

날씨/연령/성별에 따른 상품 수요

상품선택

스킨케어

날씨에 따른 제품의 소분류별 판매 비율

날씨에 따른 연령별 소비트렌드

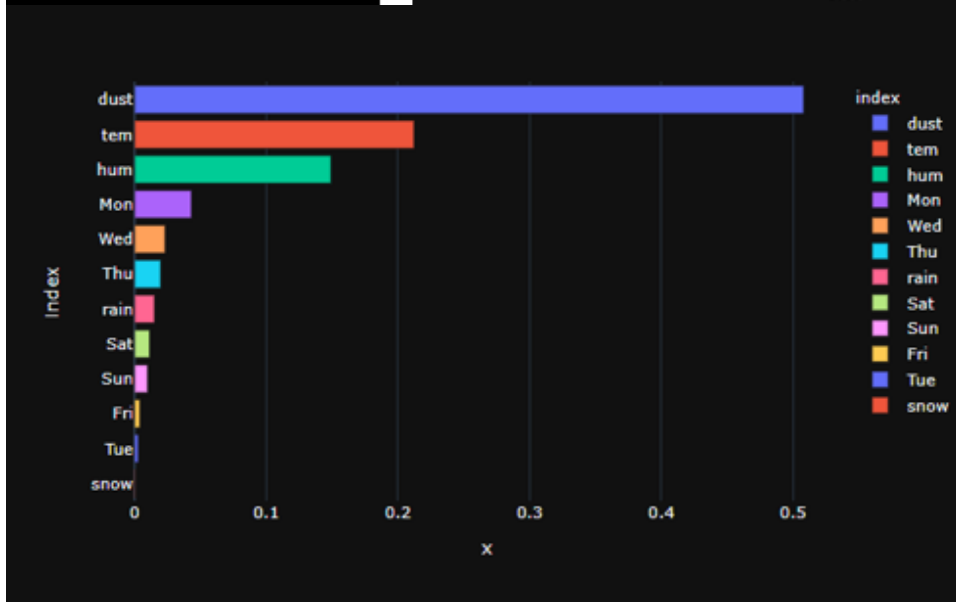
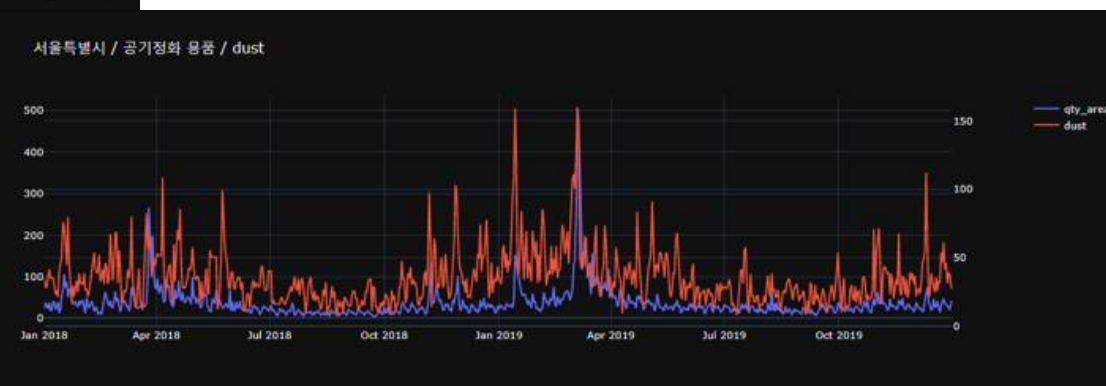
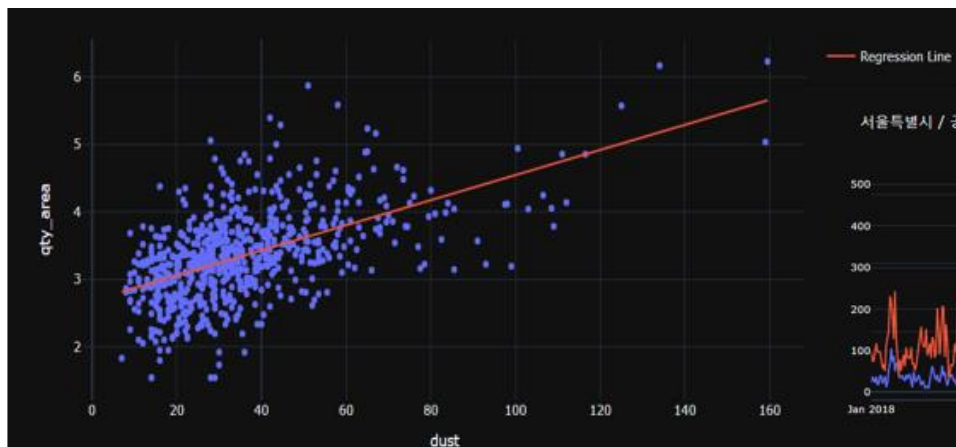
날씨에 따른 성별 소비트렌드

날씨에 따른 수요예측

지역
서울특별시

물품
공기정화 용품

날씨요인
dust



수요예측

기온 20
강수량 0
습도 40
적설량 0
미세먼지 20

월요일 1
화요일 0
수요일 0
목요일 0
금요일 0
토요일 0
일요일 0

결과

22.52 개입니다

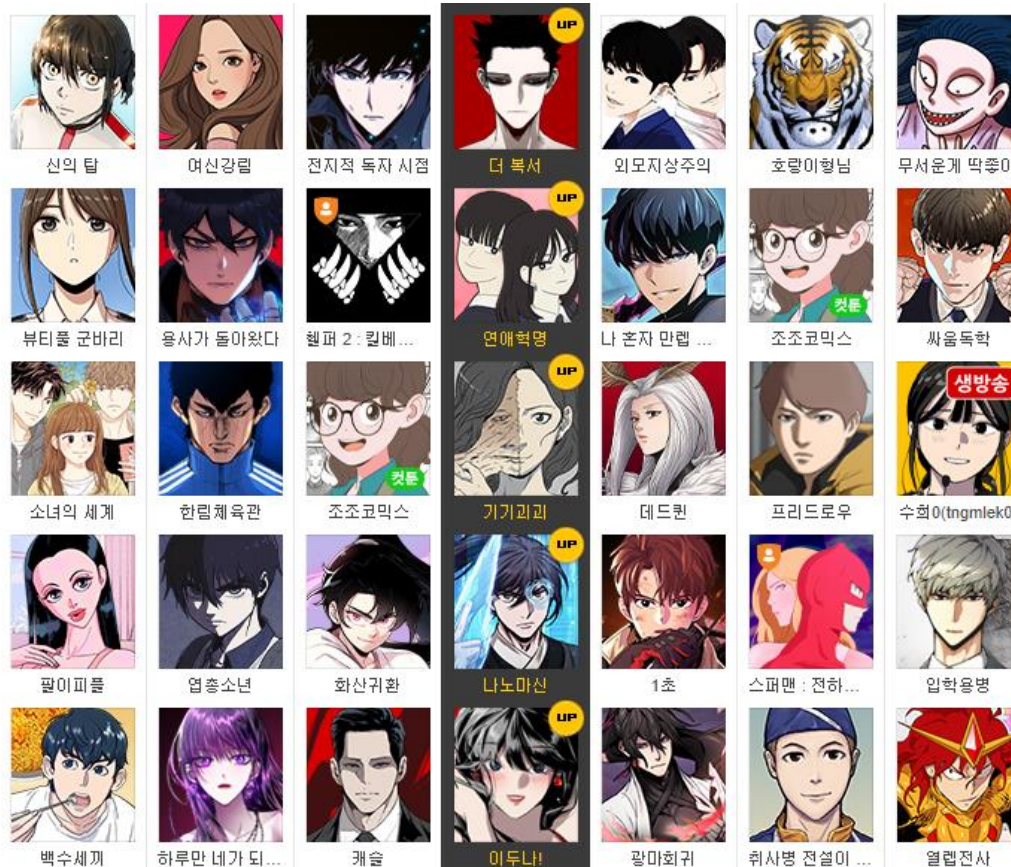
사용자 맞춤 웹툰 추천 서비스

프로젝트 목적

- 사용자가 인상깊게 봤던 웹툰을 입력하면
그와 비슷한 장르, 그림체를 갖고 있는 웹툰 추천

프로젝트 내용

- 네이버 웹툰 대표 이미지, 소개글, 댓글, 제목, 장르 크롤링
- 대표 이미지 분석 후 분류
- 소개글, 댓글 자연어 처리 후 분류
- 서비스 제공하는 웹페이지 구현



사용자 맞춤 웹툰 추천 서비스

형태소 분석기 사용하여 명사만 추출

- ⇒ Labeling이 되어있지 않아 Unsupervised Learning 사용 (Word2Vec)
- ⇒ 문장의 벡터를 구하기 위해 각 단어의 벡터들의 평균을 구함
- ⇒ K-Means Clustering 사용하여 가까운 벡터끼리 분류
- ⇒ 벡터의 차원이 굉장히 높아 시각화 하기 위해 2차원으로 축소

TSNE Webtoon Summary Embeddings



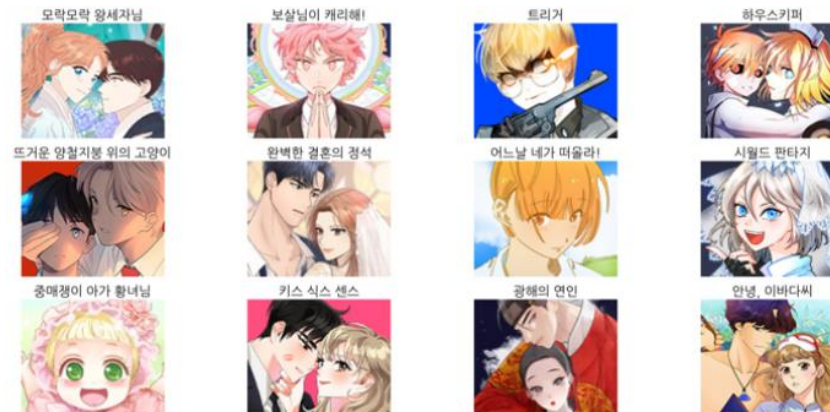
사용자 맞춤 웹툰 추천 서비스

대표이미지 분류 기법 Augnet model 사용

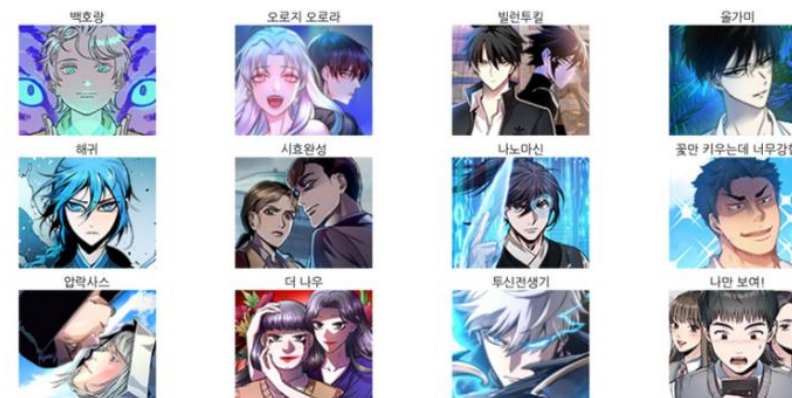
⇒ Input 전체 썸네일



전체적으로 어둡고 절망적인 느낌



풍부한 파스텔톤을 사용하여 화사한 느낌



푸른색 계열이 주를 이루고 액션장르의 느낌

사용자 맞춤 웹툰 추천 서비스



The image shows a web form for Naver's comic recommendation service. It is divided into four horizontal sections with different background colors: light blue, light green, light orange, and light red. The first section has the title 'NAVER 웹툰 추천'. The second section, '요일을 선택하세요', contains seven checkboxes for days of the week. The third section, '웹툰을 선택하세요', has a dropdown menu with 'Volvo' selected. The fourth section, '장르를 선택하세요', has a dropdown menu with '열로' selected. At the bottom, the word '추천' is centered and underlined with a pink line.

NAVER 웹툰 추천

요일을 선택하세요

☐ 월요일 ☐ 화요일 ☐ 수요일 ☐ 목요일 ☐ 금요일 ☐ 토요일 ☐ 일요일

웹툰을 선택하세요

웹툰 선택: Volvo ▼

장르를 선택하세요

장르 선택: 열로 ▼

추천

추천 서비스 제공하는 웹페이지 초안

아직 프로젝트 미완성

추후 추천 알고리즘 구축 후 웹페이지 디자인

사용자 연령/성별 정보를 입력 받아

협업 필터링에 사용 가능

VISION

무한한 데이터가 끊임 없이 생산되는
빅데이터 시대를 살아가고 있는 지금

데이터 분석능력 뿐만 아니라 다양한
능력이 요구되는 데이터 분석가가 되기
위하여 끝없이 도전하겠습니다