

날씨에 따른 배달 수요 예측

2020년 서울 / 경기권 데이터 중심으로

우리 같이
[분석]해볼래요?



신속 분석



정확 예측

플레이데이터 전정훈 / 최슬기 / 최창수



01 개요

-
- 1. 주제선정
 - 2. 데이터 수집
 - 3. 데이터 정제(전처리)

02 데이터 분석

-
- 1. EDA
 - 2. 모델링

03 양상분석

-
- 1. 랜덤포레스트
 - 2. 소프트맥스

04 결론

-
- 1. 결론
 - 2. 서비스 제안

01

개요

다들 식사는 잡혔어?



1. 주제선정

세상에서 [밥]에 가장 진심인 민족

아니근데 진짜 한국인은 밥에 미쳤음
혼날때 : 너 밥도 없을 줄 알어!
작업걸때 : 저랑 밥 한번 드실래요?
고마울때 : 야 진짜 고맙다 나중에 밥 한끼
찐하게 사줄게
안부물어볼때 : 밥은 먹고 지내냐?
한국인이 행복할 때 : 밥 맛있게 먹었을까?
아플때 : 밥은 꼭 먹으라고 함

먹보 한국인

팔로우

- 안부- 밥한번 먹자
- 만류- 도시락 싸들고 다니기
- 금강산- 식후경
- 격무- 먹고 살자고 하는 짓
- 고민- 먹고 살기 힘들다
- 걱정- 뭐 해먹고 사는가
- 회사- 한솥밥
- 아픔- 밥이 보약
- 불호- 밥맛 없음
- 칭찬- 맛집
- 감방- 콩밥
- 무임승차- 밥상에 숟가락 얹기
- 욕- 빌어먹을 놈

한국인은 밥심이다 ㄹㅇ임
인사: 밥먹었어??
헤어질 때: 나중에 밥 한번 살게 고맙다
감사인사: 헬 밥 한번 살게 한번만 도와줘
설득하기: 밥 한번 살게 한번만 도와줘
ㅜㅜ
풀러팅: 밥 한끼 같이 먹을까요? ㅎㅎ
혼내는 이유: 밥 안먹어서
걱정하는 이유: 밥 한끼 안먹었다고
일부 사람들의 이상형: 밥 잘먹는 사람

날씨에 어울리는 음식을
찾아 먹는 것이 자연스러움

더운 날에는 치킨에 맥주



비오는 날에는 전과 막걸리



추운 날에는 뜨끈한 국물



사장님~ 이젠 데이터로 예측해요!

Covid19으로 많은 요식업 자영업자들은 높은 임대료와 치솟는 물가에 배달 전문(위주) 업체로 변경해 배달 시장이 크게 성장했다.

이에 맞춰 자영업자들이 보다 효율적인 가게 운영을 위해 배달 날씨 데이터 분석을 통한 배달 수요를 예측해 배달플랫폼 입장 사장님들이 참고할 수 있는 단기적 운영 전략 정보를 제공한다.



2. 데이터 수집

Covid19으로 배달 시장이 활성화된
2020년 데이터 중심으로 수집

DATA1

날씨별
배달 품목차이

주요칼럼 : 시군구명, 온도, 습도,
강수량, 업종 별 배달
횟수

제공 : 경기대학교 빅데이터센터
출처 : KT 통신 빅데이터 플랫폼

DATA2

주문지역
인구 특성

주요칼럼 : 시군구코드, 시군구명,
총인구수

제공 : 경기대학교 빅데이터센터
출처 : KT 통신 빅데이터 플랫폼

DATA3

서울시
초미세먼지
경보 현황

주요칼럼 : 발표시간, 예보등급
제공 : 서울특별시 대기정책과
출처 : 서울 열린데이터 광장

3. 데이터 정제(전처리)

**수집한 데이터를 가공해 탐색적분석과 예측 분석을 위한
Sectors[업종]로 target값 설정**

	datetime	Day of the week	city	population	hour	rain	dust	humidity	precipitation	temp	windspeed	holiday	sectors
0	2020-01-01	Wednesday	41281	467673	0	0	NaN	75	0.0	-2.7	0.9	1	6
1	2020-01-01	Wednesday	41281	467673	1	0	NaN	67	0.0	-1.8	1.2	1	6
2	2020-01-01	Wednesday	41281	467673	11	0	NaN	72	0.0	0.6	1.0	1	1
3	2020-01-01	Wednesday	41281	467673	12	0	NaN	74	0.0	0.2	0.7	1	1
4	2020-01-01	Wednesday	41281	467673	13	0	NaN	73	0.0	0.0	0.9	1	1
...
100212	2020-06-14	Sunday	11290	441812	19	0	37.0	81	0.0	18.7	1.0	0	3
100213	2020-06-19	Friday	11290	441812	11	0	25.0	62	0.0	24.2	2.6	0	3
100214	2020-06-20	Saturday	11290	441812	21	0	39.0	65	0.0	22.8	1.3	0	3
100215	2020-06-25	Thursday	11290	441812	11	1	15.0	88	0.0	20.6	1.5	0	3
100216	2020-06-25	Thursday	11290	441812	13	0	15.0	90	0.5	20.3	1.3	0	3

100217 rows × 13 columns

3. 데이터 정제(전처리)

모델 성능 개선을 위해 변수조절(위도, 경도 추가)

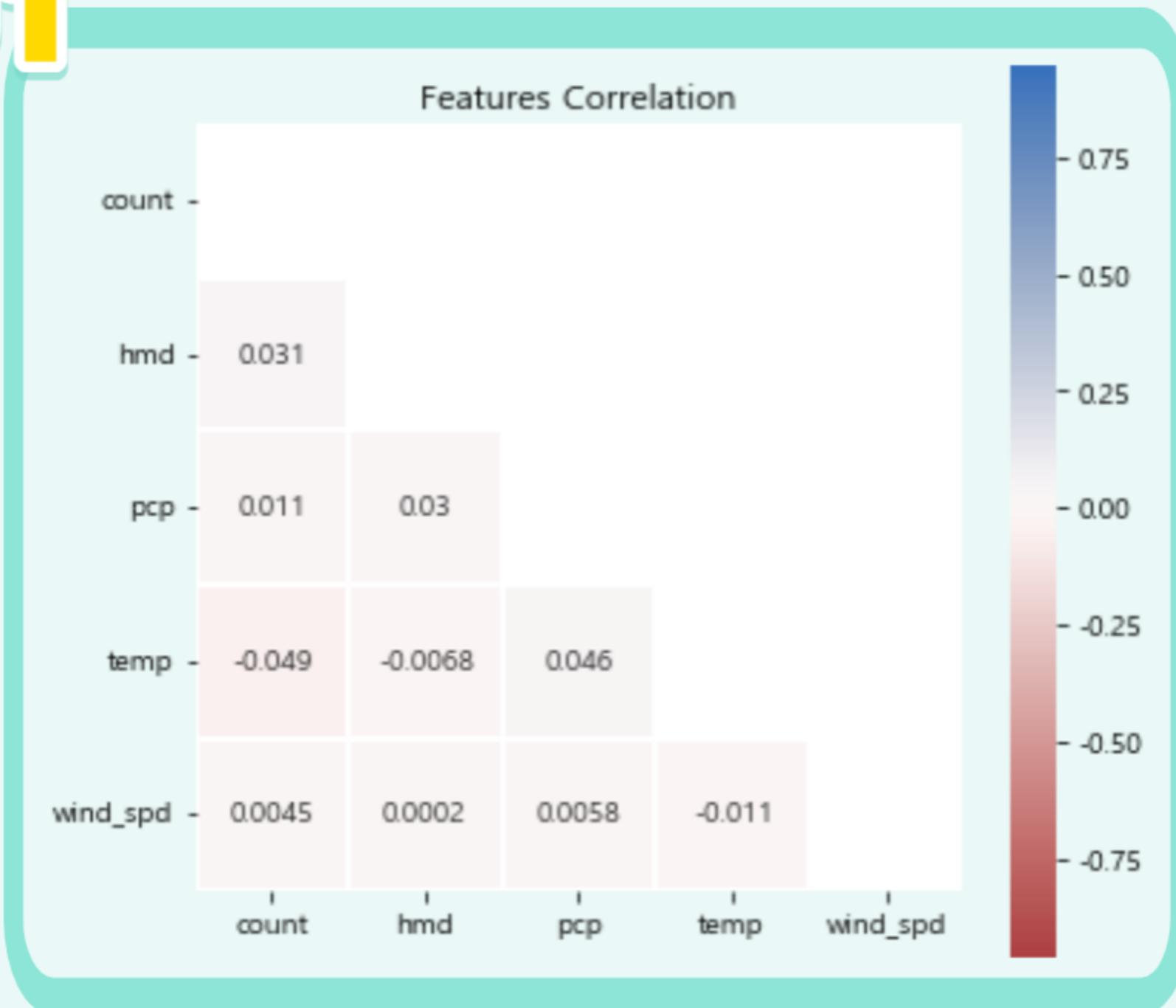
	datetime	Day of the week	city	population	hour	rain	dust	humidity	precipitation	temp	windspeed	holiday	sectors	latitude	longitude
0	2020-01-01	Wednesday	41281	467673	0	0	NaN	75	0.0	-2.7	0.9	1	6	37.646130	126.864424
1	2020-01-01	Wednesday	41281	467673	1	0	NaN	67	0.0	-1.8	1.2	1	6	37.646130	126.864424
2	2020-01-01	Wednesday	41281	467673	11	0	NaN	72	0.0	0.6	1.0	1	1	37.646130	126.864424
3	2020-01-01	Wednesday	41281	467673	12	0	NaN	74	0.0	0.2	0.7	1	1	37.646130	126.864424
4	2020-01-01	Wednesday	41281	467673	13	0	NaN	73	0.0	0.0	0.9	1	1	37.646130	126.864424
...
100212	2020-06-14	Sunday	11290	441812	19	0	37.0	81	0.0	18.7	1.0	0	3	37.591508	127.019798
100213	2020-06-19	Friday	11290	441812	11	0	25.0	62	0.0	24.2	2.6	0	3	37.591508	127.019798
100214	2020-06-20	Saturday	11290	441812	21	0	39.0	65	0.0	22.8	1.3	0	3	37.591508	127.019798
100215	2020-06-25	Thursday	11290	441812	11	1	15.0	88	0.0	20.6	1.5	0	3	37.591508	127.019798
100216	2020-06-25	Thursday	11290	441812	13	0	15.0	90	0.5	20.3	1.3	0	3	37.591508	127.019798

02



1. EDA

01

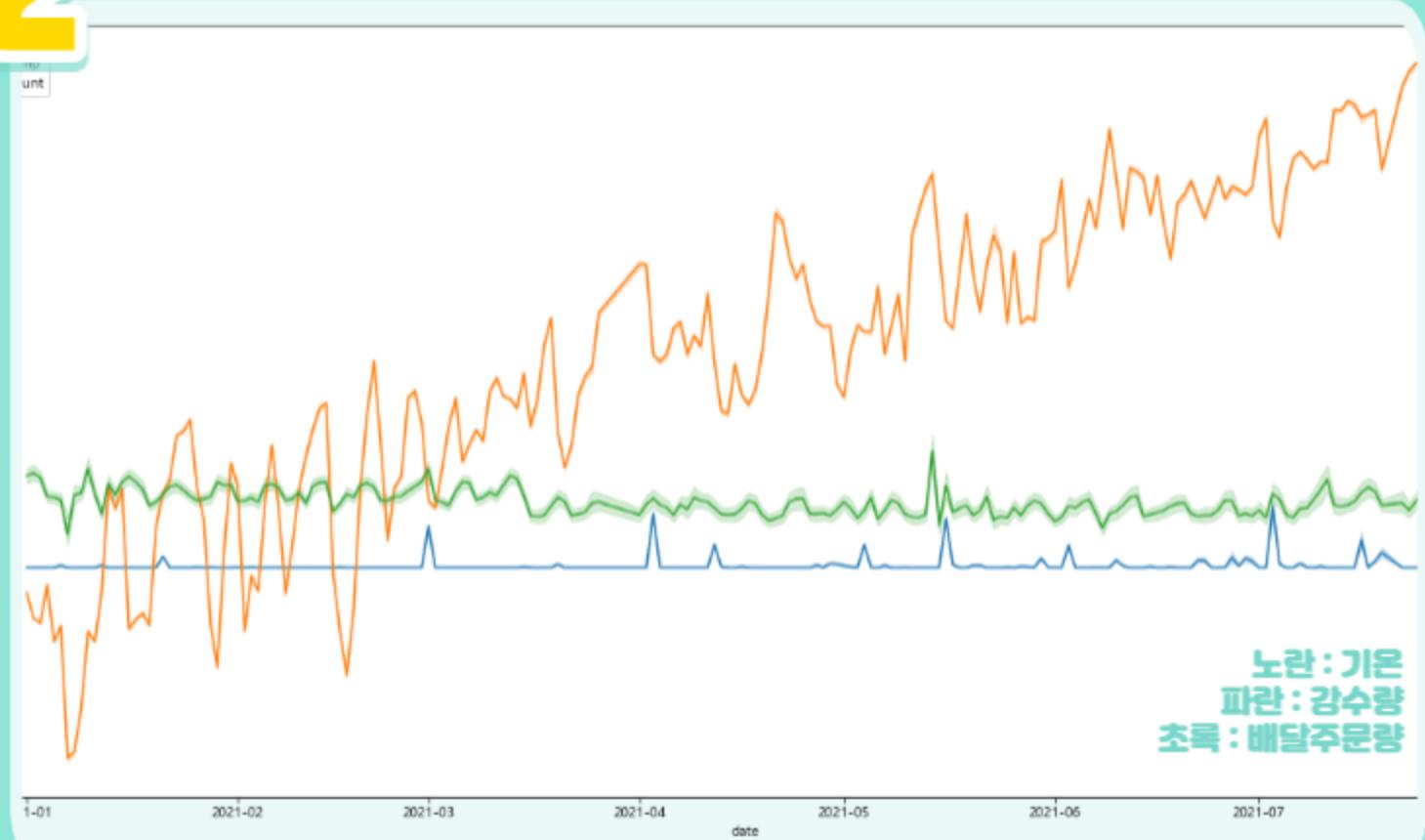


기온, 습도, 강수량, 풍속 간의
상관관계는 약하다.

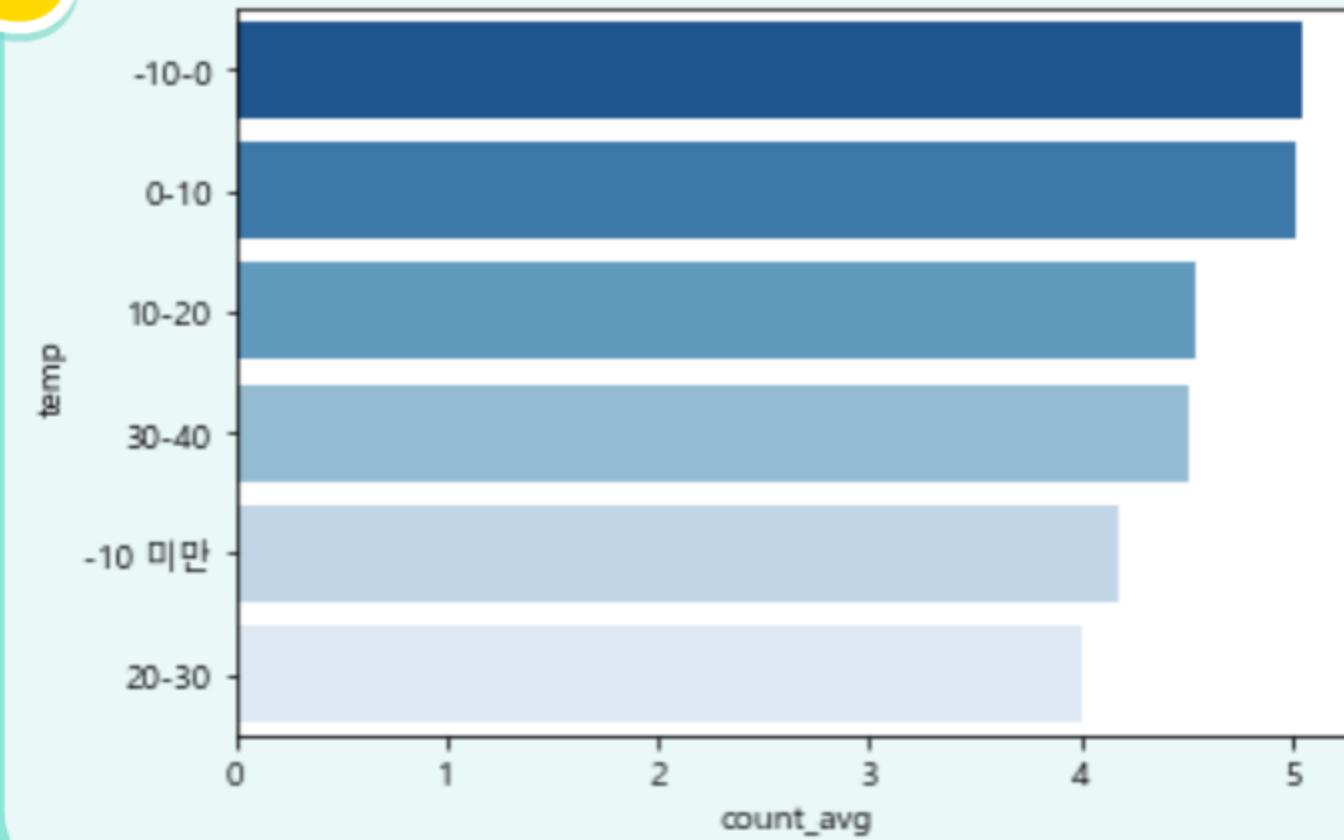
02. 데이터분석

기온이 낮은 기울 한겨울에 배달주문이 증가하고
기온이 높아질수록 감소한다.

02

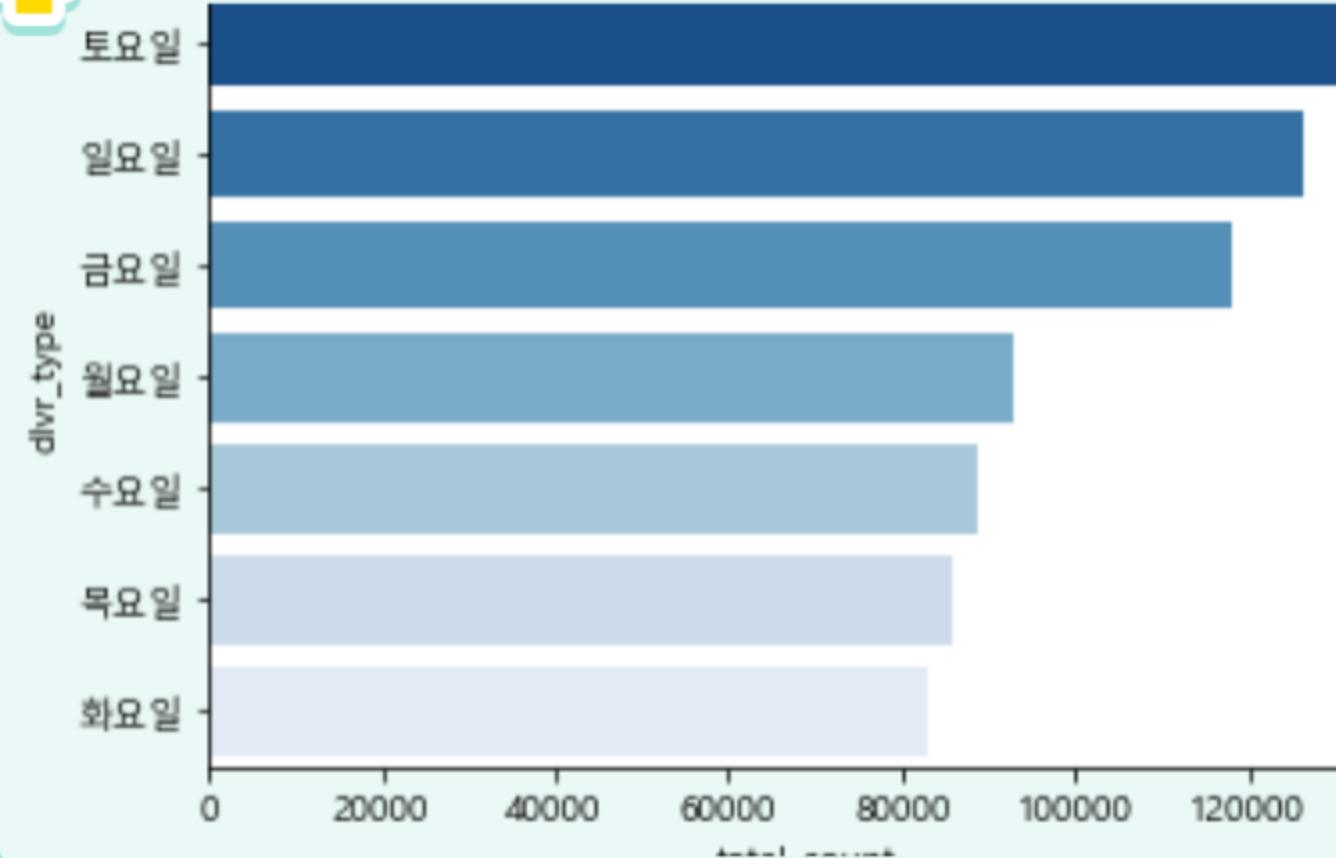


03

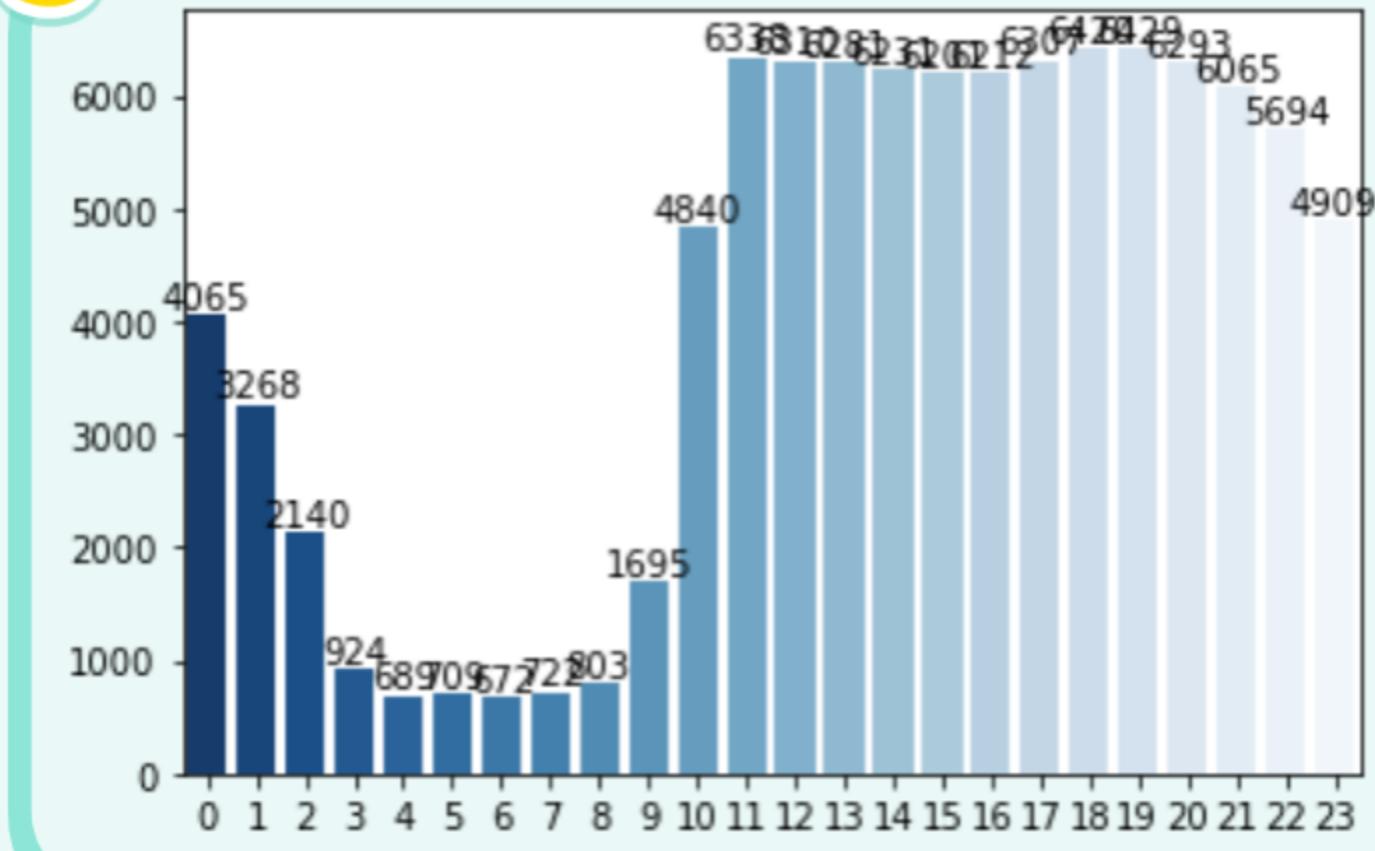


주말과 저녁~야식 시간대에 주문량이 많다.

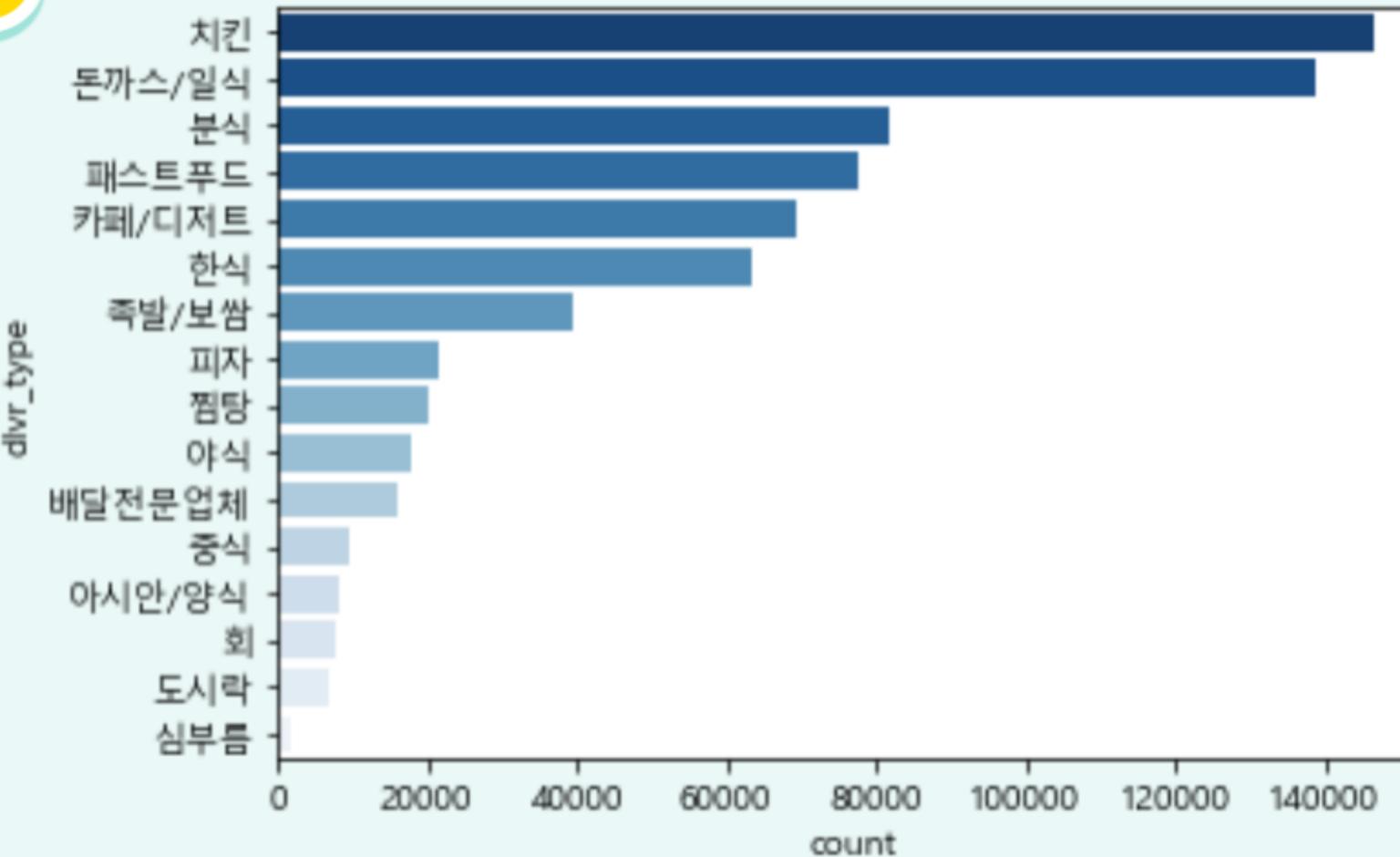
04



05

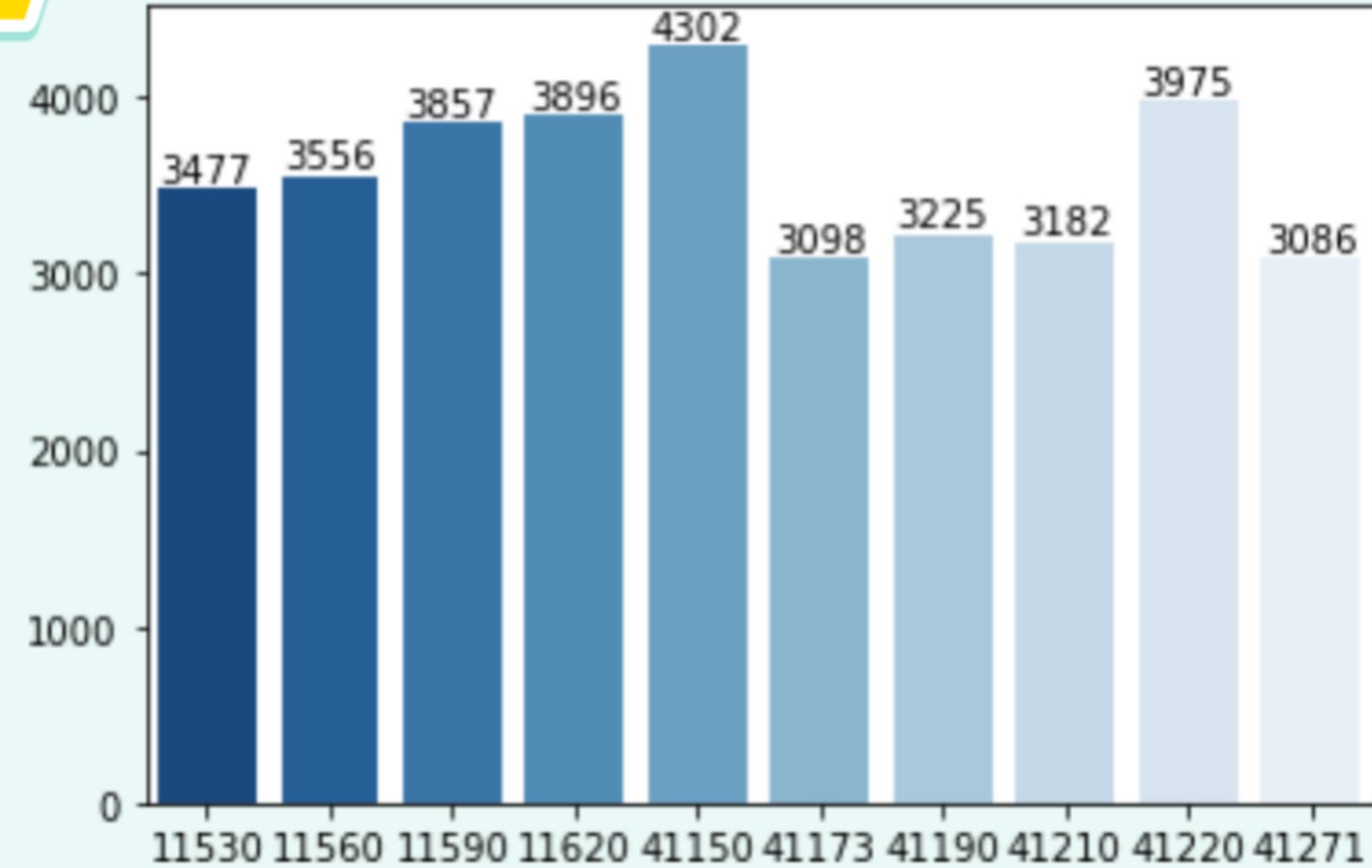


06



치킨, 돈까스, 분식 등
튀김 메뉴가 있는 업종들의
배달주문이 많다.

07



경기도 의정부시가 가장 많은 배달주문을 했고 그 뒤는 영등포구, 동작구, 관악구 등 서울시 주요 생활지구의 배달주문량이 많다.

2. 모델링

로지스틱 회귀분석

목적

새로운 변수가 주어졌을 때, 종속
변수 어디에 속하게 되는지 확률
계산해 수요예측

모델링

X = 강수, 습도, 강수량, 기온, 풍속
y = 업종
train,test 7:3 분리
sklearn LogisticRegression

성능평가

test 데이터에 target 예측
실제와 일치하는 비율 정확도로 측정
accuracy : 0.320

2. 모델링

소프트맥스

목적

3개 이상으로 분류하는 다중분류
문제에 적합

모델링

기존 로지스틱 회귀분석에서
`multi_class = multinomial` 추가

성능평가

`accuracy`는 0.01 소폭 상승했으나
`f1_weighted`로 계산한 값은 0.177
로 성능이 매우 낮음

2. 모델링

의사결정나무

목적

의사결정 규칙을 시각화해 의사결정이 이뤄지는 시점과 성과를 한 눈에 파악해 간편한 해석 가능

모델링

sklearn DecissionTreeRegressor
Depth 조절하며 모델링

성능평가

불순도 척도인 entropy가 2점대로 성능이 좋지 못함

2. 모델링

SVM

목적

매우 강력하고 선형이나 비선형 분류, 회귀 이상치 탐색에 유용하고 복잡한 분류 문제에 적합

모델링

풍속값 log 변환 후
sklearn LinearSVC(선형)
sklearn PolynomialFeatures
(비선형)
두 가지 방식 모두 진행

성능평가

선형 f1score = 0.017
비선형 f1score = 0.467
이전 모델들 보다 성능이 많이 향상됨

03

양상블

3. 암상블

STEP
01

단순진행

accuracy =
0.987

STEP
02

교차검증 및
그리드서치

accuracy =
0.663

STEP
03

하이퍼파라미터
조정

accuracy =
0.670

STEP
04

데이터 정제

accuracy =
0.675

랜덤포레스트

각 Step 설명

Step1. 교차 검증을 진행하지 않아 제대로 된 값이 아님

Step2. 교차검증 및 그리드서치를 통한 하이퍼파라미터 조정

Step3. n_estimators 값을 줄여 하이퍼파라미터 조정했으나
train데이터 정확도가 낮아 데이터 정제 필요

Step4. 습도를 제거하고 하이퍼파라미터 조정, 0.05소폭 상승

03. 암상블

STEP

05

원핫인코딩

accuracy =
0.540

STEP

06

데이터 정제

f1_weighted =
0.728

STEP

07

보팅을 위한
모델 별 점수확인

STEP

08

보팅

랜덤포레스트

각 Step 설명

Step5. 원핫인코딩으로 그리드서치 결과 반복 확인

Step6. stratify 적용, 특성 조정, f1_weighted로 성능평가지표 변경해
보다 정확한 평가 진행

Step7. 소프트맥스를 활용했을 때 score가 낮아 악영향이 우려되어
시군구코드를 위도, 경도로 변경

Step8. 보팅 후 재진행

04



최종 TEST

로지스틱 회귀분석
&
소프트맥스

f1_weighted
0.163

의사결정나무

f1_weighted
0.712

랜덤포레스트

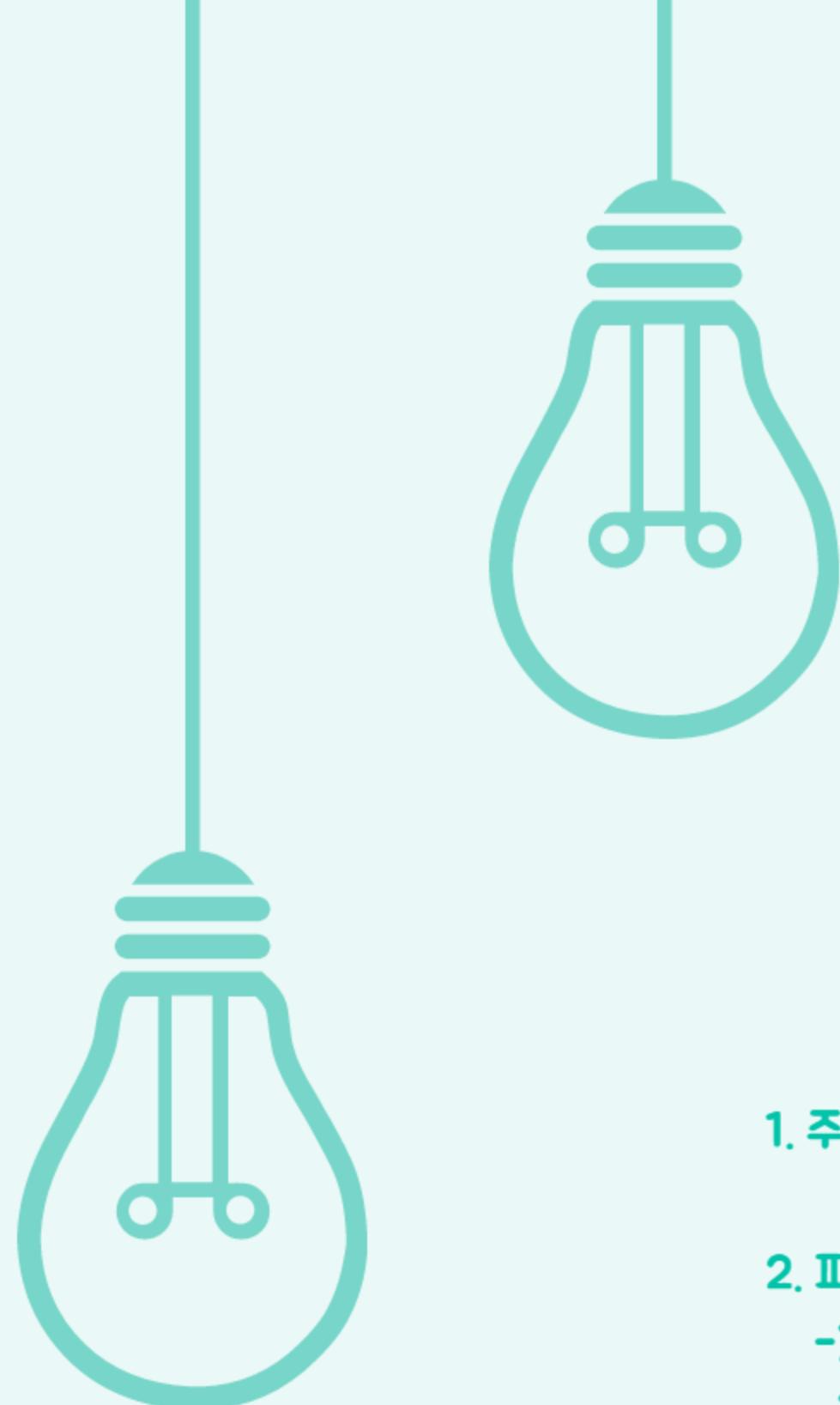
f1_weighted
0.724

보팅

f1_weighted
0.706

추후 분석

한정된 데이터로 이런 날씨면 이런 업종들이 배달량이 많아지는 정도의 예측만 진행했다.
업종 별 인기매장 및 인기메뉴, 주문자 특성 등 다양한 특성 데이터를 추가해 딥러닝 기법에 활용하면
보다 정확한 예측을 하리라 기대된다.



서비스제안

1. 주간 단위로 날씨유형 별 대응전략 알람(원재료 수급 조절, 인력 보강 등)
2. 파트타임 수시 매칭 서비스 개발
 - > 일상회복이 시작되면서 구인난이 심해짐
 - > 날씨에 따른 배달 수요가 많아질 때, 조건에 맞춰 인력 보강
(part-time으로 일하는 개념으로 접근)

APPENDIX

일정

9/14(수)	주제선정	데이터수집	데이터분석 방향 설정	역할분담
9/15(목)	데이터 정제	데이터 추가수집	데이터 전처리	EDA
9/16(금)	로지스틱 회귀분석	소프트맥스	의사결정나무	SVM
9/19(월)	양상블	시각화	취합 및 정리	PT제작
9/20(화)	최종 마무리	발표		

github

데이터 정제	https://github.com/wjswjdgns/miniproject02/blob/main/deliverydata_Notebook.ipynb
로지스틱회귀분석	https://github.com/xotmfrlox/delivery_weather/blob/main/Machine%20Learning/SVM(Support%20Vector%20Machine).ipynb
소프트맥스	https://github.com/xotmfrlox/delivery_weather/blob/f1_weighted/Machine%20Learning/Softmax.ipynb
의사결정나무	https://github.com/cuicharles17/personal_git/blob/master/decisiontree-checkpoint.ipynb
SVM	https://github.com/xotmfrlox/delivery_weather/blob/main/Machine%20Learning/SVM(Support%20Vector%20Machine).ipynb
랜덤포레스트	https://github.com/wjswjdgns/miniproject02/blob/main/deliverydata_data.ipynb

**THANK
YOU**