

# 제주도 음식물쓰레기양 예측 및 주요요인 도출

Covid-19와 사각지대를 중심으로



갯슈

김수현 (myksh0903@naver.com)  
고가연 (imgayoun0@gmail.com)  
이수현 (happy\_shsh@naver.com)



# 목차

## 01 문제정의

## 02 데이터 이해

## 03 EDA

- 3-1 인사이트1
- 3-2 인사이트2

## 04 전처리

- 4-1 데이터 merge
- 4-2 결측 데이터 처리
- 4-3 최종 데이터셋 및 변수
- 4-4 feature 예측 모델 선정

## 05 모델링

- 5-1 프로세스
- 5-2 7,8월 feature 예측
- 5-3 데이터 merge
- 5-4 모델 선정
- 5-5 최종 모델 적용 및 예측결과

## 06 감소방안

- 6-1 주요요인 및 해결방안
- 6-2 기대효과

# 01 문제정의



# 01 문제정의



문제정의



데이터이해



EDA



전처리



모델링



감소방안



"포화율 99.9%" 제주도는 왜 '쓰레기 섬'이 됐을까?  
제주 쓰레기매립장마다 '꽉꽉'...이미 한계점 넘어섰다  
성난 주민들 매립장 '봉쇄'...음식물쓰레기 대란 '일촉즉발'

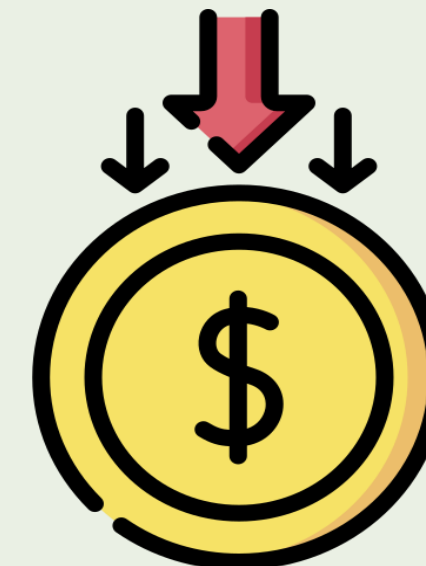
1인당 하루평균 생활쓰레기  
배출량 국내 1위



제주도 내  
매립지 부족 및 주민갈등 심화



음식물쓰레기 처리비용 과다

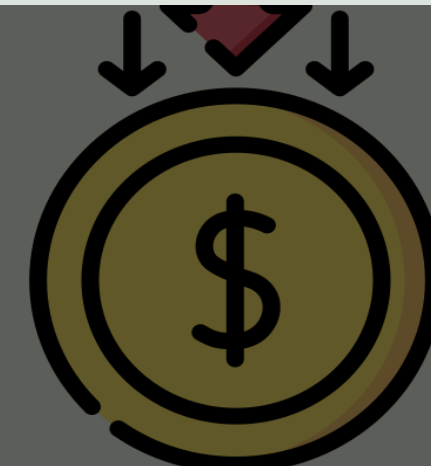




NEWS

"포화율 99.9%" 제주도는 왜 '쓰레기 섬'이 됐을까?  
제주 쓰레기매립장마다 '꽉꽉'...이미 한계점 넘어섰다

이를 해결하고자 음식물 쓰레기 배출량 예측 모델을 개발하고  
배출량 감소 방안을 도출하고자 한다.



# 02 데이터 이해



# 02 데이터 이해

## 사용 데이터



### 제공 데이터

FOOD\_WASTE.csv

음식물쓰레기 배출 데이터

KOREAN.csv

내국인 유동인구 데이터

LONG\_TERM\_FRGN.csv

장기체류 외국인 유동인구 데이터

SHORT\_TERM\_FRGN.csv

단기체류 외국인 유동인구 데이터

RESIDENT.csv

거주인구 데이터

CARD\_SPENDING.csv

음식 관련 카드 소비 데이터



### 외부 데이터

제주특별자치도\_편의점정보.csv

제주특별자치도상가(상권)정보.csv  
(출처 : 제주데이터허브)에서 편의점 데이터만 추출한 데이터

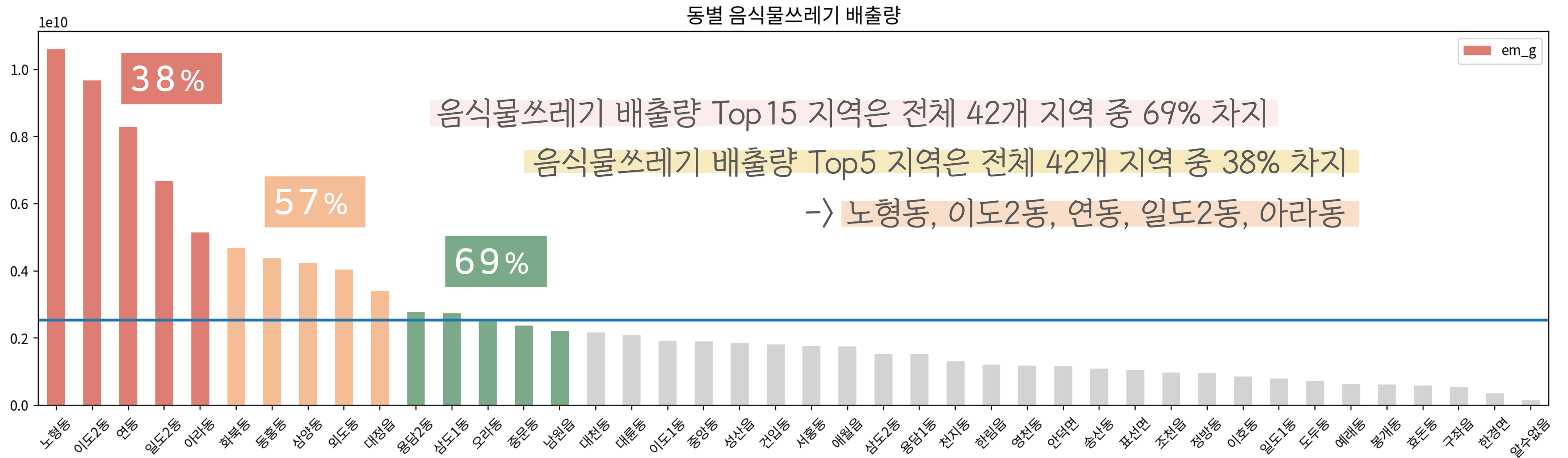
# 03 EDA







## 1) 음식물쓰레기의 약 70%는 소수의 주요 지역에서 배출된다.





배출량 상위 top5 지역은 유동인구, 거주인구, 배달소비, 마트소비가 높다.



배출량 top5

노형동 이도2동 연동 일도2동 아라동

## 1 내국인 유동인구 top5

거주 -	노형동	이도2동	연동	아라동	애월읍
근무 -	이도2동	노형동	아라동	연동	애월읍
방문 -	이도2동	애월읍	연동	노형동	아라동

## 2 거주인구 top5

노형동 이도2동 연동 애월읍 아라동

## 3 배달 관련 카드 소비 top5

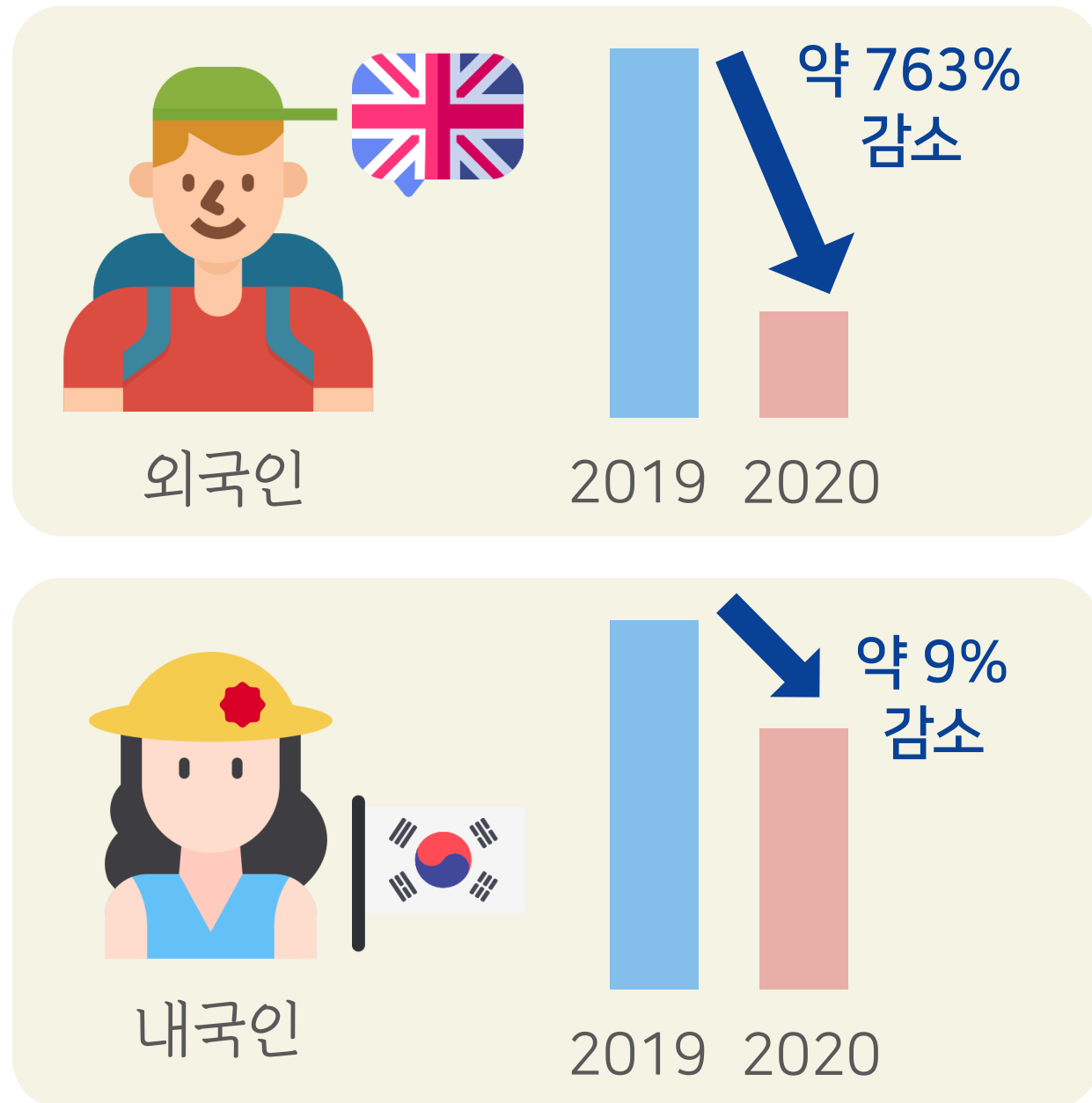
노형동 연동 이도2동 일도2동 아라동

## 4 마트 관련 카드 소비 top5

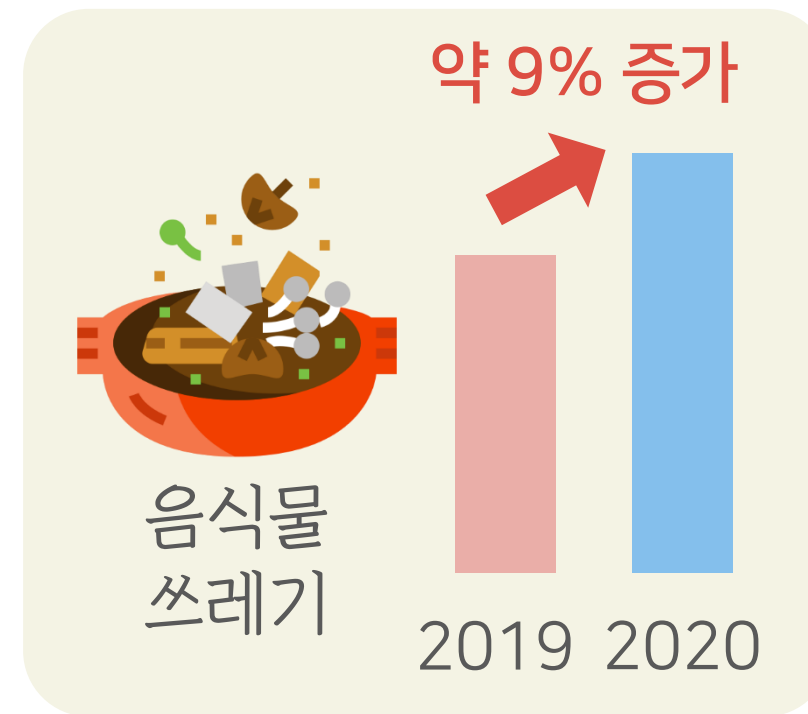
노형동 연동 이도2동 아라동 일도2동



2) 코로나19로 내/외국인 방문객 감소, but 음식물쓰레기 배출량은 증가했다.



but



why

신종 코로나바이러스  
국내 첫 확진자 발생  
☎ 승인 2020.01.20 13:38



배달관련 카드 건수

2018년 대비 2019년 : 60% 증가  
2019년 대비 2020년 : 57% 증가  
2018년 대비 2020년 : 83% 증가



# 04 전처리



# 04 전처리 ① 데이터 merge

## 년, 월, 동 별 merge

### FOOD\_WASTE

base_date
emd_cd
emd_nm
emd_area_cd
em_g

합계

merge

fin\_data1.csv

### 내국인 유동인구

base_date
emd_nm
resd_pop_cnt
work_pop_cnt
visit_pop_cnt

평균

### 단기/장기체류 유동인구

base_date
emd_nm
resd_pop_cnt
work_pop_cnt
visit_pop_cnt

평균

### 거주인구

base_year/month
emd_nm
resid_reg_pop
foreign_pop
total_pop

평균

### 음식 관련 카드 소비

base_date
emd_nm
mct_cat_nm
use_cnt

평균

### 파생 변수

계절
공휴일 수
편의점 수

### 1 2018.01~2019.10 기간 내 구좌읍, 조천읍, 한경면, 한림읍의 부재

- 구좌읍, 조천읍, 한경면, 한림읍 4개동의 데이터는 2019년 11월부터 존재함
- 즉, 4개동의 2020년은 1~12월이 모두 존재함
- 다른 동의 2020년 대비 2018, 2019년 피쳐 증감률에 대해 평균을 계산하여 4개동에 적용하여 결측 데이터를 채워넣음

### 2 '알수없음'의 부재 및 피쳐 NaN

- '알수없음'은 2019년 2~8월, 2020년 3~5월에 존재하지 않음
- '알수없음'에 대한 피쳐는 모두 NaN으로 존재함
- 결측치 NaN을 0으로 변환하고 '알수없음'이 존재하지 않는 년월에도 0으로 채워넣음

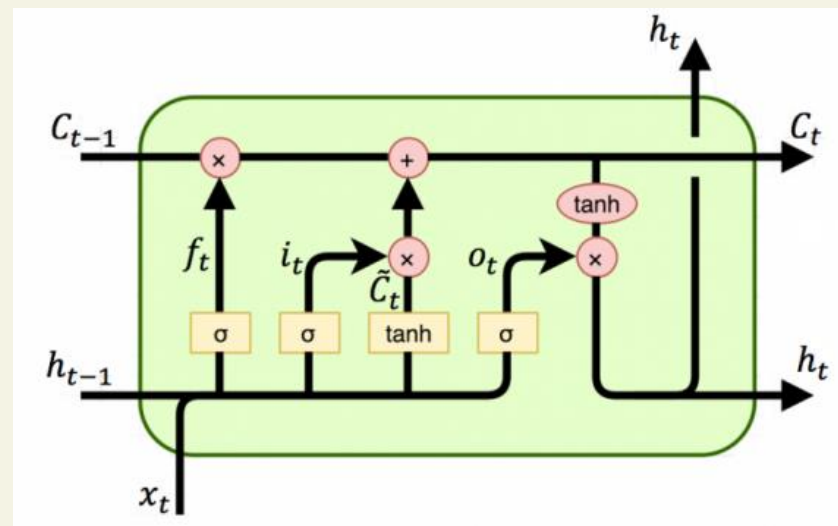
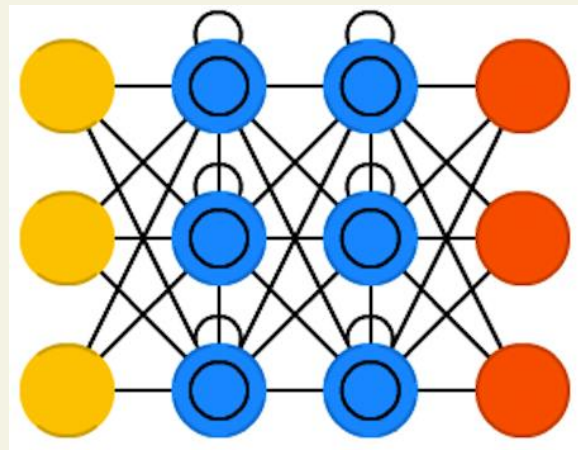


fin\_data1.csv


year, month	범주형	년, 월
emd_nm	범주형	제주특별자치도 42개 행정동명 + 알수없음
long_resd, long_work, long_visit	연속형	장기체류 외국인 유동인구 (거주인구, 근무인구, 방문인구 월별 평균)
short_visit_pop_cnt	연속형	단기체류 외국인 유동인구 (방문인구 월별 평균)
season	범주형	계절 (봄:0, 여름:1, 가을:2, 겨울:3)
holiday_cnt	연속형	해당 년월 공휴일 수
배달_cnt, 마트_cnt	연속형	음식관련 카드 사용 수 (배달, 마트)
내국resd, 내국work, 내국visit	연속형	내국인 유동인구 (거주인구, 근무인구, 방문인구 월별 평균)
거주foreign, 거주total, 거주resid	연속형	거주인구 (내국인, 외국인, total 월별 평균)
편의점 수	연속형	각 행정동 편의점 수
em_g	연속형	음식물 쓰레기 배출량 (월별 합계, Label)

# 04 전처리 ④ feature 예측 모델 선정

## 1 후보1 - LSTM



셉 호크라이터

- Long Short-Term Memory
- 장단기 메모리망
- 은닉층의 각 뉴런에  Memory Cell 추가  
-> 내부에 저장공간을 가짐
- 7,8월 모든 feature('em\_g'(Label) 포함) 동시 예측
- window=5로 설정, 5개월로 다음 1달을 예측하는 시계열 프로세스

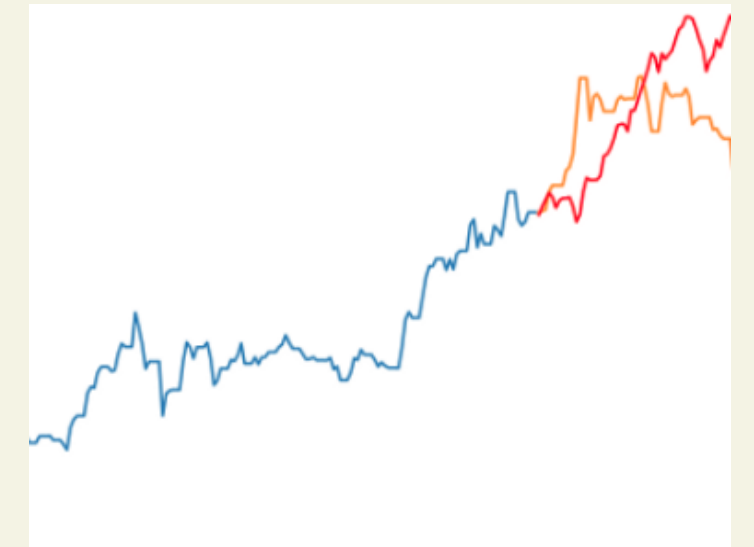
## 2 후보2 - ARIMA

$ARIMA(p, d, q)$

AR 모형 차수

차분

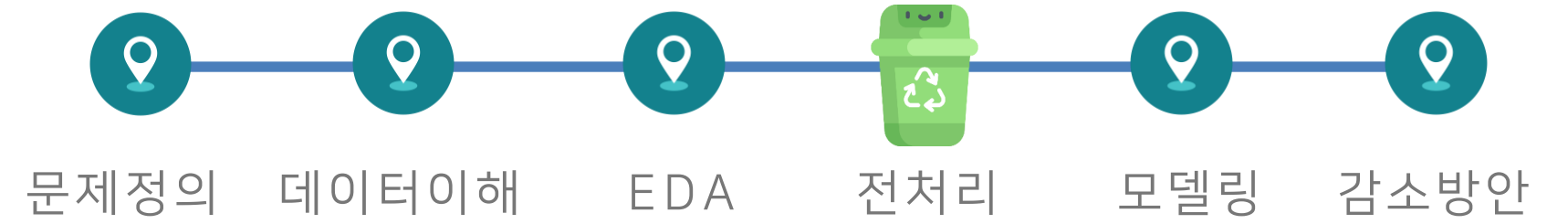
MA 모형 차수



- 시계열 예측을 위한 통계적 모델
- AutoRegressive Integrated Moving Average
- 자기회귀 누적 이동평균 모델
  - MA 최적 차수 - 자기 상관 함수 (ACF)
  - AR 최적 차수 - 부분(편) 자기 상관 함수 (PACF)
- 7,8월 'em\_g'(Label)를 제외한 피쳐 예측 후 모델 적용

# 04 전처리

## ④ feature 예측 모델 선정



### Feature 예측 결과 비교

	MSE	MAE	RMSE	R2
LSTM	0.6185	0.6625	0.7865	-0.1530
ARIMA	0.0140	0.0992	0.1181	0.9782



Feature를 한 번에 모두 예측하다 보니  
전체적인 평가 지표에서 좋지 않은 점수를 나타냄



LSTM에 비해 예측력이 우수한 점수를 나타내고 있음

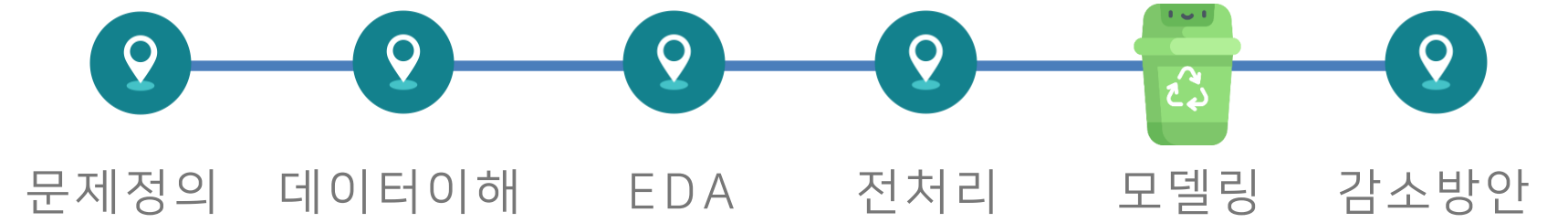
7, 8월 피쳐들을 예측하여 모델을 적용하는 방향으로 모델링 진행



# 05 모델링



# 05 모델링 ① 프로세스



# 05 모델링

② 7,8월 feature 예측



문제정의



데이터이해



EDA



전처리



모델링



감소방안

## 모델 1

대표적인 시계열 모델인 ARIMA 사용

각 동별 각 피쳐들을 auto\_arima를 사용하여 최적의 p,d,q를 적용하여 7,8월 데이터를 예측  
각 동별 피쳐들을 각각 예측을 수행해야하기 때문에 함수를 생성하여 실행

```
from pmdarima.arima import auto_arima
def for_arima(dong, feature):
    df1 = data[['year', 'month', 'emd_nm', feature]]
    # i = df1[(df1['year']==2021)&(df1['month']==5)].index[0]
```

```
    train = df1.set_index(['year', 'month'])
    test = df1.set_index(['year', 'month'])
```

```
    train = train[train['emd_nm']==dong]
    train = train.drop(['emd_nm'], axis=1)
```

```
    cols = ['long_resd', 'long_work', 'long_visit', 'short_visit_pop_cnt', '배달_cnt',
            '마트_cnt', '내국resd', '내국work', '내국visit', '거주foreign', '거주total', '거주resid']
```

```
    l1=[]
    for i in data['emd_nm'].unique():
        for j in cols:
            l1.append(for_arima(i, j))
```

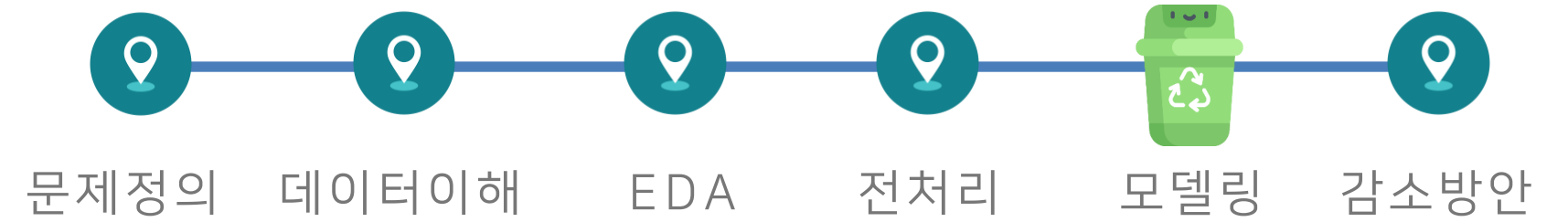
```
model_arima= auto_arima(train, trace=True, error_action='ignore', start_p=1, start_q=1, max_p=3, max_q=3,
                        suppress_warnings=True, stepwise=False, seasonal=True)
```

```
model = model_arima.fit(train[feature].values, trend='c', full_output=True, disp=True)
fore = model.predict(steps=2)
return fore[:2]
```



# 05 모델링

## ③ 데이터 merge



### 예측한 7, 8월 피쳐 데이터

	emd_nm	year	month	long_resd	long_work	long_visit
0	건입동	2021	7	5.252561	1.189391	11.031765
1	구좌읍	2021	7	12.333491	0.787917	7.123437
2	남원읍	2021	7	13.862747	0.734602	10.328282
3	노형동	2021	7	45.847444	3.932875	18.635641
4	대륜동	2021	7	6.765012	1.303481	6.697616
...	...	...	...	...	...	...
79	표선면	2021	8	6.396931	0.836936	5.901348
80	한경면	2021	8	13.813848	0.428379	9.650899
81	한림읍	2021	8	34.832140	1.781228	17.995943
82	화북동	2021	8	14.203290	1.428152	3.920178
83	효돈동	2021	8	5.175617	0.280931	3.305364

84 rows × 7 columns

### 최종 합친 데이터(2018.01~2021.08) : concat78.csv

	year	month	emd_nm	long_resd	long_work	long_visit
0	2018	1	건입동	8.583848	1.428658	9.718961
1	2018	1	구좌읍	17.971788	1.469885	7.029032
2	2018	1	남원읍	15.207991	2.274018	11.477026
3	2018	1	노형동	68.001225	5.834689	20.299049
4	2018	1	대륜동	7.739642	0.731385	9.043291
...	...	...	...	...	...	...
1843	2021	8	표선면	6.396931	0.836936	5.901348
1844	2021	8	한경면	13.813848	0.428379	9.650899
1845	2021	8	한림읍	34.832140	1.781228	17.995943
1846	2021	8	화북동	14.203290	1.428152	3.920178
1847	2021	8	효돈동	5.175617	0.280931	3.305364

1848 rows × 7 columns

### 모델 2 Regressor 모델 적용

	MSE	MAE	RMSE	R2
CatboostRegressor	0.0218	0.1157	0.1477	0.9659
ExtraTreesRegressor	0.0519	0.1774	0.2278	0.9188
XGBoostRegressor	0.0199	0.1079	0.1413	0.9688

최종 모델 선택 : XGBoostRegressor

# 05 모델링 ⑤ 최종 모델 적용 및 예측결과

사용 모델

XGBoostRegressor

하이퍼 파라미터

learning\_rate = 0.05  
max\_depth = 5  
min\_child\_weight = 4  
reg\_alpha = 0.05  
reg\_lambda = 0.2

Train : Valid = 8 : 2

train

valid

2018.01~ 2021.06

2021.07~08

X\_test

예측 결과

데이터분석분야\_퓨처스리그\_ECO제주\_갯슈\_평가데이터.csv

행정동명			
7월 배출량(g)			
8월 배출량(g)			
NO			
1	한림읍	7.578923e+07	7.592872e+07
2	애월읍	9.555955e+07	9.006181e+07
3	구좌읍	2.989358e+07	3.068265e+07
4	조천읍	6.815165e+07	6.851959e+07
5	한경면	1.913213e+07	1.944487e+07
6	일도1동	1.212487e+07	1.166209e+07
7	일도2동	1.467571e+08	1.496582e+08
8	이도1동	3.404465e+07	3.250201e+07
9	이도2동	2.251089e+08	2.201762e+08
10	삼도1동	5.509460e+07	5.379201e+07



# 06 감소방안



# 06 감소방안



문제정의



데이터이해



EDA



전처리



모델링



감소방안

## 주요요인



### 사각지대

클린하우스에 편의점, 숙박시설, 단체급식 등에서 배출되는 음식물 쓰레기도 포함될 것으로 가정

가정 외 배출 근원지를 사각지대라고 명명하고 이를 중점으로 음식물쓰레기 배출량 감소방안을 도출하고자 한다.



클린하우스 : 생활폐기물을 버리는 곳



생활폐기물 : 가전폐기물, 음식물 쓰레기 등 사업장 폐기물 외의 생활에서 발생하는 폐기물

## 해결방안



편의점

- 소비기한이 남은 음식 저장 및 공유하는 공유 냉장고 배치  
-> 취약 계층 및 재난 피해 구호물품 지원
- 음식물쓰레기 처리기 비용 일정 지원



숙박공유

- 입도객에 체류기간에 따라 환경부담금 부과  
-> 거둔 비용으로 음식물쓰레기 처리기 구입 및 숙박공유시설 주변에 설치
- 퇴실 시 음식물 배출량에 따른 환경부담금 부과



집단급식

- 누적 배출량에 따른 패널티 부과  
-> 한달 누적 기준치 제공
- 배식 완료 후 남은 음식을 다회용기에 가져가도록 장려



문제정의



데이터이해



EDA



전처리

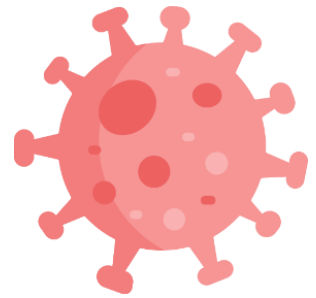


모델링



감소방안

## 주요요인



### 코로나19

#### “확진자 줄었지만” 22일까지 제주 ‘4단계’ 유지

코로나 19 유행 이후 거리두기로 인해 가정 내 식사가 증가하게 되었고 이는 곧 가정에서의 음식물쓰레기 배출 증가로 이어짐



가정 내 식사 : 가정식, 포장, 배달음식

가정 내 식사 증가를 중점으로 가정 내 음식물쓰레기 배출량 감소방안을 도출하고 그 외 다양한 음식물쓰레기 감소방안을 공유하고자 한다.

## 해결방안



가정식

-> 식재료 구입, 식사 준비, 식사로 단계를 나누어 고민

#### STEP1 구입 단계

- 마트 및 시장에 소포장, 날개 구매 장려

#### STEP2 식사 준비 단계

- 가족구성원의 정확한 식사량 파악 -> 먹을 만큼만 준비

#### STEP3 식사 단계

- 먹을 만큼만 덜어 먹기
- 실질적인 어린이 식사 교육 실시



배달/포장

- 배달 시 최소 주문 금액 완화
- 식사량 및 반찬의 선택지 제공

## 기대효과

- 가정뿐 아니라 사각지대의 음식물 쓰레기 배출량을 감소시키는 **맞춤형 방안**을 시행함으로써 클린하우스 매립지 포화 문제, 도민 간 갈등 완화 등이 기대된다.
- 또한 이러한 방안들을 클린하우스 수, 배출량이 집중된 지역 **Top6에 시범 적용**하면서 적용 구역을 늘리며 장기적으로 배출량을 감소시키는 현상을 기대한다.



감사합니다

