



머신러닝을 이용한 제주도 음식물 쓰레기 양 분석 및 해결 방안 제시

2021 빅콘테스트 - 제주
ECO

청정푸어클린제주

- 임성준, lsj9862@gmail.com
- 권휘성, 0518hs@naver.com
- 김승종, sjkhalladay@gmail.com

목차



1. 분석 배경 및 목표

- 분석 배경
- 분석 목표

2. 데이터 전처리

- 데이터 소개
- 데이터 가공 및 병합

3. EDA

- 데이터 주기성 확인
- 특성 선택
- 클러스터링

4. 모형 설계 및 학습

- 사용 모형 설명
- 학습 과정 설명

5. 예측 결과

- 예측 결과

6. 분석결과

- 분석 결과 요약
- 분석 결과 활용 및 해결방안
- 기대효과

7. 보완점 및 후속 연구

- 보완점 및 후속 연구



1. 분석 배경 및 목적

- 분석 배경
- 분석 목적

분석 배경 및 목적



● 분석 배경

- 제주 지역 음식물 쓰레기는 다른 지역에 비해 상당히 많은 편이며 특히, 관광객이 많은 여름철에는 음식물 쓰레기 발생량이 증가하는 모습을 보이고 있다. 또한 코로나 19로 인해 해외여행 대체 여행지로 제주를 많이 찾는 경향을 보이고 있다.
- 제주도는 음식물 쓰레기 분리배출을 전도지역으로 확대하는 등 음식물 쓰레기 처리를 위한 노력을 해왔으나 그 효과를 아직 직접적으로 보지 못하였다
- 이에 많은 분야에서 활용되고 있는 빅데이터 분석 기법을 통해 음식물 쓰레기 발생에 영향을 주는 요인을 찾고 음식물 쓰레기 발생량을 예측해보는 모델을 구축하고자 한다.

분석 배경 및 목적



- 분석 목표

행정동별
음식물 쓰레기
발생량 예측

음식물 쓰레기
발생에 영향을
주는 요인 탐색

음식물 쓰레기
감소 방안 제안



2. 데이터 전처리

- 데이터 소개
- 데이터 가공 및 병합

데이터 전처리



● 데이터 소개

1. 제공데이터: 01_음식물쓰레기_FOOD_WASTE.csv

base_date	city	emd_cd	emd_nm	em_area_cd	em_cnt	em_g	pay_amt
2018-01-01	제주시	50110670	외도동	W6X062	2	15500	464
2018-01-01	제주시	50110630	아라동	W6XA97	25	59800	1787
2018-01-01	제주시	50110630	아라동	W6XA96	8	25350	758
2018-01-01	제주시	50110630	아라동	W6XA95	12	30000	898
2018-01-01	제주시	50110630	아라동	W6XA94	14	44050	1317
...
2021-06-30	제주시	50110250	한림읍	W6XEB4	12	58250	1743
2021-06-30	제주시	50110250	한림읍	W6XEB7	15	111000	3326
2021-06-30	제주시	50110250	한림읍	W6XEB8	21	48300	1444
2021-06-30	제주시	50110250	한림읍	W6XEA2	5	22050	661
2021-06-30	제주시	50110640	오라동	W6XAD5	2	2900	87

45,103,395 행, 7 변수
2018.01.01~2021.06.30 데이터 (일별)

***알수없음 : 행정동 코드가 매칭 안 되는 city컬럼**

데이터 전처리



- 데이터 소개

2. 제공데이터: 02-1_내국인유동인구_KOREAN.csv

base_date	emd_nm	sex	age	resd_pop_cnt	work_pop_cnt	visit_pop_cnt
2018-01-01	한림읍	여성	0	638.5509	15.5400	151.5149
2018-01-01	한림읍	여성	0	0.0000	3.7249	171.3458
2018-01-01	한림읍	여성	10	668.4577	17.0308	166.0500
2018-01-01	한림읍	여성	10	0.0000	4.2577	195.8539
2018-01-01	한림읍	여성	20	609.7710	5.9490	148.7246
...
2021-06-30	예래동	남성	20	0.0000	0.0000	599.1969
2021-06-30	예래동	남성	20	187.5990	41.9998	75.5996
2021-06-30	예래동	남성	10	121.8467	0.0000	4.5128
2021-06-30	예래동	남성	80	122.2190	0.0000	4.5032
2021-06-30	예래동	남성	80	0.0000	0.0000	31.5179

2,211,445 행, 7 변수
2018.01.01~2021.06.30 데이터 (일별)

데이터 전처리



● 데이터 소개

3. 제공데이터 : 02-2_장기체류 외국인 유동인구_LONG_TERM_FRGN.csv

base_date	time	nationality	city	emd_cd	emd_nm	resd_pop_cnt	work_pop_cnt	visit_pop_cnt	resd
2018-01-01	1	CAN	제주시	50110250	한림읍	6.3381	0.0000	4.5802	제주
2018-01-01	1	CHN	제주시	50110250	한림읍	442.1091	0.0000	5.7051	제주
2018-01-01	1	CHN	제주시	50110250	한림읍	0.0000	7.8695	51.3461	그외
2018-01-01	1	DEU	제주시	50110250	한림읍	0.0000	0.0000	8.2099	그외
2018-01-01	1	ETC	제주시	50110250	한림읍	15.9184	0.0000	2.8758	제주
...
2021-06-30	24	THA	서귀포시	50130620	예래동	0.0000	0.0000	3.3558	제주
2021-06-30	24	TWN	서귀포시	50130620	예래동	0.0000	0.0000	12.2655	그외
2021-06-30	24	USA	서귀포시	50130620	예래동	22.0624	0.0000	0.0000	제주
2021-06-30	24	USA	서귀포시	50130620	예래동	0.0000	0.0000	26.4749	그외
2021-06-30	24	VNM	서귀포시	50130620	예래동	3.7754	0.0000	0.0000	제주

23,837,317 행, 10 변수
2018.01.01~2021.06.30 데이터
(시간대별)

데이터 전처리



● 데이터 소개

4. 제공데이터 : 02-3_단기체류 외국인 유동인구_SHORT_TERM_FRGN.csv

base_date	time	nationality	city	emd_cd	emd_nm	visit_pop_cnt
2018-01-01	1	CHN	제주시	50110250	한림읍	66.3170
2018-01-01	1	ETC	제주시	50110250	한림읍	21.1308
2018-01-01	1	HKG	제주시	50110250	한림읍	3.6771
2018-01-01	1	JPN	제주시	50110250	한림읍	5.5962
2018-01-01	1	USA	제주시	50110250	한림읍	3.9078
...
2021-06-30	24	USA	서귀포시	50130610	중문동	9.2120
2021-06-30	24	CHN	서귀포시	50130610	중문동	0.6039
2021-06-30	24	ETC	서귀포시	50130610	중문동	11.5473
2021-06-30	24	CHN	서귀포시	50130620	예래동	0.2013
2021-06-30	24	HKG	서귀포시	50130620	예래동	0.0166

6,059,044 행, 7 변수
2018.01.01~2021.06.30 데이터
(시간대별)

데이터 전처리



● 데이터 소개

5. 제공데이터 : 03_거주인구_RESIDENT_POP.csv

base_year	base_month	city	emd_cd	emd_nm	sex	resid_reg_pop	foreign_pop	total_pop
2018	1	제주시	50110590	건입동	남성	5085	146	5231
2018	1	제주시	50110590	건입동	여성	4715	82	4797
2018	1	제주시	50110256	구좌읍	남성	7965	368	8333
2018	1	제주시	50110256	구좌읍	여성	7609	184	7793
2018	1	서귀포시	50130253	남원읍	남성	9806	428	10234
...
2021	6	제주시	50110250	한림읍	여성	10341	1140	11481
2021	6	제주시	50110600	화북동	남성	12062	169	12231
2021	6	제주시	50110600	화북동	여성	12238	161	12399
2021	6	서귀포시	50130550	효돈동	남성	2660	39	2699
2021	6	서귀포시	50130550	효돈동	여성	2597	18	2615

3,612 행, 9 변수
2018.01.01~2021.06.30 데이터
(월별)

데이터 전처리



● 데이터 소개

6. 제공데이터 : 04_음식관련 카드소비_CARD_SPENDING.csv

base_date	city	emd_cd	emd_nm	mct_cat_cd	mct_cat_nm	use_cnt	use_amt
2018-01-06	제주시	50110650	연동	C00100	한식	3643	127777300
2018-01-09	제주시	50110650	연동	C00500	패스트푸드	432	6711675
2018-01-15	제주시	50110650	연동	C01200	농축수산물	236	16089579
2018-01-15	제주시	50110650	연동	A00200	마트/슈퍼마켓	4031	121979867
2018-01-20	제주시	50110650	연동	C01000	식품	633	30410674
...
2020-10-16	제주시	XXXXXXXX	알수없음	C01400	배달	5	127250
2019-08-31	서귀포시	XXXXXXXX	알수없음	C01400	배달	5	39974
2020-10-03	서귀포시	XXXXXXXX	알수없음	C01400	배달	10	329832
2020-10-23	서귀포시	XXXXXXXX	알수없음	C01400	배달	5	68715
2020-12-21	제주시	XXXXXXXX	알수없음	C01400	배달	10	187452

544,805 행, 8 변수

2018.01.01~2021.06.30 데이터 (일별)
행정구역 '알수없음' 존재

데이터 전처리



● 데이터 소개

7. 외부데이터 1: 기상데이터

(<https://data.kma.go.kr/cmmn/main.do>)

일시	평균기온(° C)	평균상대습도(%)	월합강수량(00~24h만)(mm)	합계 일조시간(hr)
201801	4.9	67	58.0	73.8
201802	5.5	60	86.6	122.2
201803	11.4	72	118.0	190.9
201804	15.9	68	112.5	205.6
201805	19.1	76	98.8	159.9

- 기상 정보가 사람들의 생활에 영향을 미칠 것이라고 판단
- 종관기상관측(ASOS) 자료의 평균기온, 평균상대습도, 월합강수량, 월 합계 일조시간의 제주 지역 자료를 이용



- 데이터 소개

8. 외부데이터 2: 제주도 입도객

(<https://www.data.go.kr/>)

base_date	passengers	201806	1322085
201801	1138066	201807	1281379
201802	1003353	201808	1305745
201803	1142039	201809	1238100
201804	1342618	201810	1290110
201805	1334236	201811	1187924

- 제주 입도객수가 음식물 쓰레기 발생량에 전체적으로 영향을 줄 것이라 판단
- 제주 공항을 통해 입도한 한 달 승객수를 이용



- 데이터 소개

9. 외부데이터 3 : 소비자 심리지수

(<https://kostat.go.kr/portal/korea/index.action>)

base_date	소비자심리지수
201801	110.7
201802	109.0
201803	108.9
201804	107.9
201805	108.8
201806	106.3
201807	101.8

- 경제지표 중 하나인 소비자 심리 지수(CSI)가 소비에 영향을 미쳐 결국 쓰레기 발생량에 영향을 줄 것이라 판단
- 월별 소비자 심리지수(CSI) 자료 이용

데이터 전처리



● 데이터 소개

10. 외부데이터 4 : 코로나 신규 확진자

(<http://ncov.mohw.go.kr/>)

	일자	계(명)
0	202001	11
1	202002	2920
2	202003	6855
3	202004	979
4	202005	703
5	202006	1331
6	202007	1506

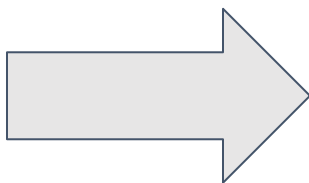
- 코로나 신규 확진자가 생활 습관에 영향을 미쳐 결국 쓰레기 발생량에 영향을 줄 것이라 판단
- 질병관리청에서 발표하는 국내 발생자, 해외 입국자 월별로 합산한 자료 이용

데이터 전처리

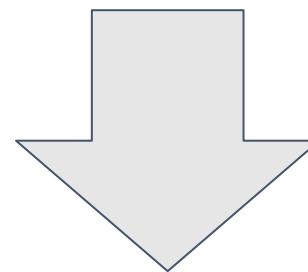


- 데이터 가공 및 병합

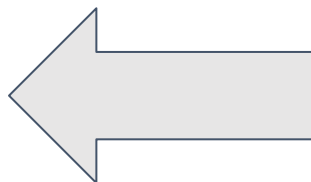
월별 합산 데이터, 일별 합산
데이터, 시간별 합산 데이터
존재



월별 합산으로 모든 데이터
가공



행정동, 월별 기준으로 병합



데이터 별 특성 고려하여
파생변수 생성
(추자면, 우도면 데이터 제거)

데이터 전처리

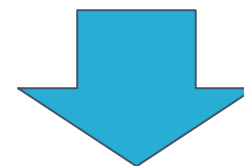


● 데이터 가공 및 병합

02-1_내국인유동인구_KOREAN.csv

base_date	emd_nm	resd_pop_cnt_m0	work_pop_cnt_m0	visit_pop_cnt_m0
201801	건입동	2.307612e+05	8627.4372	208782.6978
201801	구좌읍	3.144128e+05	32075.7221	212004.3943
201801	남원읍	3.320644e+05	12944.4662	350749.9732
201801	노형동	2.097413e+06	197420.7495	860716.8896
201801	대륜동	4.822790e+05	13800.0069	429610.7303
...
202106	표선면	3.361310e+05	27850.9888	119673.6258
202106	한경면	1.902001e+05	5892.9804	53221.7788
202106	한림읍	6.405556e+05	96647.2440	247094.1997
202106	화북동	1.310405e+06	74939.0682	548021.4722
202106	효돈동	1.292545e+05	3633.7365	35141.3668

성별, 나이,
목적(거주, 노동, 방문)
카테고리 존재



성별 나이를 묶어 집단화하여
count한 변수
(ex. 10대 여성 방문 인구, 20대
남성 노동 인구)

데이터 전처리



● 데이터 가공 및 병합

02-2_장기체류 외국인 유동인구_LONG_TERM_FRGN.csv

emd_nm	base_date	resd_pop_cnt_lf	work_pop_cnt_lf	visit_pop_cnt_lf	nationality_AUS
건입동	201801	101804.4428	16943.8851	115266.8745	33.0
건입동	201802	83297.2250	13242.1444	97014.3574	55.0
건입동	201803	94945.6814	17960.8853	108325.1290	32.0
건입동	201804	86144.4468	14824.6721	146303.3325	17.0
건입동	201805	102087.5935	16103.5984	119204.6630	44.0
...
효돈동	202102	40649.2588	1842.3114	20830.7649	61.0
효돈동	202103	34360.1109	2596.8689	21294.8326	33.0
효돈동	202104	34270.6330	2234.1941	21201.6514	81.0
효돈동	202105	27664.5582	1608.5898	30096.0542	71.0
효돈동	202106	33531.3147	1966.0956	21380.5992	32.0

국적, 주소지 여부
목적(거주, 노동, 방문)
카테고리 존재



국적 별 인구 합산 변수,
목적별 인구 합산 변수 추가

데이터 전처리



- 데이터 가공 및 병합

02-3_단기체류 외국인 유동인구_SHORT_TERM_FRGN.csv

emd_nm	base_date	visit_pop_cnt_sf	nationality_CHN	nationality_ETC	nationality_HKG
건입동	201801	78223.0796	744.0	627.0	383.0
건입동	201802	83742.4895	672.0	655.0	495.0
건입동	201803	112258.8544	742.0	743.0	679.0
건입동	201804	151971.7736	720.0	718.0	716.0
건입동	201805	148825.1216	744.0	744.0	743.0
...
효돈동	202102	81.3438	2.0	0.0	44.0
효돈동	202103	213.1029	16.0	15.0	0.0
효돈동	202104	193.8268	32.0	52.0	0.0
효돈동	202105	371.9670	236.0	106.0	24.0
효돈동	202106	204.7502	271.0	144.0	24.0

국적 변수 존재



국적 별 인구 합산
변수

데이터 전처리



● 데이터 가공 및 병합

03_거주인구_RESIDENT_POP.csv

base_date	emd_nm	resid_reg_pop_m	foreign_pop_m	total_pop_m	resid_reg_pop_w
201801	건입동	5085	146	5231	4715
201801	구좌읍	7965	368	8333	7609
201801	남원읍	9806	428	10234	9306
201801	노형동	26286	821	27107	27083
201801	대륜동	6836	83	6919	6637
...
202106	표선면	6314	211	6525	6033
202106	한경면	4627	237	4864	4531
202106	한림읍	10891	2090	12981	10341
202106	화북동	12062	169	12231	12238
202106	효돈동	2660	39	2699	2597

성별, 외국인 여부 변수
존재



성별, 외국인 여부
그룹화하여 파생 변수 생성

데이터 전처리



- 데이터 가공 및 병합

04_음식관련 카드소비_CARD_SPENDING.csv

base_date	emd_nm	use_cnt	use_amt	간식	농축수산물	마트/슈퍼마켓	배달	부페	식품
201801	건입동	57683	2024688672	31	31	31	30	9	31
201801	구좌읍	38869	1437171016	31	31	31	26	3	31
201801	남원읍	49138	1878298705	31	31	31	26	3	31
201801	노형동	397207	12301257150	31	31	31	31	23	31
201801	대륜동	65027	2414501385	31	31	31	29	5	31

영업장 구분 변수 존재



영업장 구분을
dummy화하여 그룹별 합산
파생 변수 생성

데이터 전처리



- 데이터 가공 및 병합
 - 생성한 파생 변수

내외국민, 성별 정보를
무시하고
목적 별 총 인구 변수들 생성



- resd_pop_cnt
- work_pop_cnt
- visit_pop_cnt

성별, 나이별로 구분된
칼럼을 성별별, 나이별 합산



- resd_pop_cnt_w
- resd_pop_cnt_0
- total_pop_m
- 등등



파생변수 생성에 사용된
변수들 제거

데이터 전처리



- 데이터 가공 및 병합
 - 모든 데이터 병합 결과

base_date	emd_nm	em_cnt	em_g	pay_amt	use_cnt	use_amt	간식
201801	건입동	18416	45646550	1364804	57683	2024688672	31
201801	남원읍	9570	42437700	1270773	49138	1878298705	31
201801	노형동	113990	265018400	7883340	397207	12301257150	31
201801	대륜동	21666	57612600	1676850	65027	2414501385	31
201801	대정읍	10185	38885550	1164122	80495	2908778634	31
...
202106	표선면	10751	42453750	1271419	40962	1480487884	30
202106	한경면	8031	27060150	809898	30086	1096266339	30
202106	한림읍	25653	82746990	2476292	84533	3090858420	30
202106	화북동	66088	110750050	3306029	167413	4982321182	30
202106	효돈동	8720	27554100	824457	24585	864786389	30

1,666 행, 123 변수
2018.01 ~
2021.06

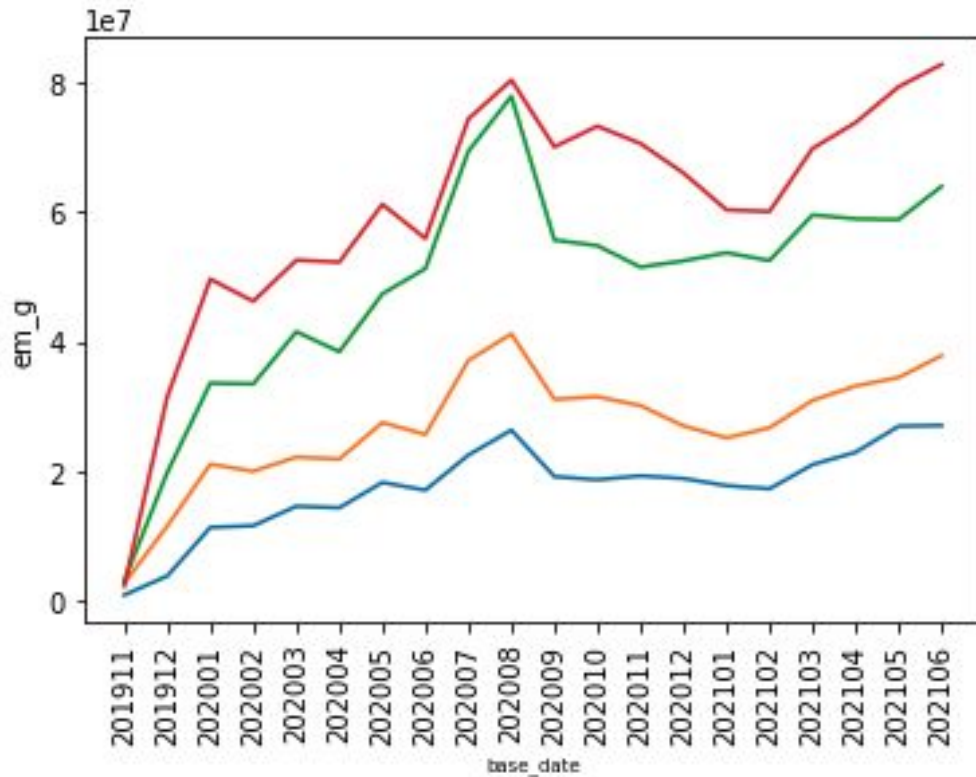


3. EDA

- 데이터 주기성 확인
 - 음식물 쓰레기 배출량의 주기성과 유사한 주기성을 가진 특성 분석 (행정동별)
- 행정동 클러스터링 (K-means Clustering)



- 데이터 주기성 확인
 - 월별 음식물 쓰레기 배출량 시각화 (랜덤 샘플)

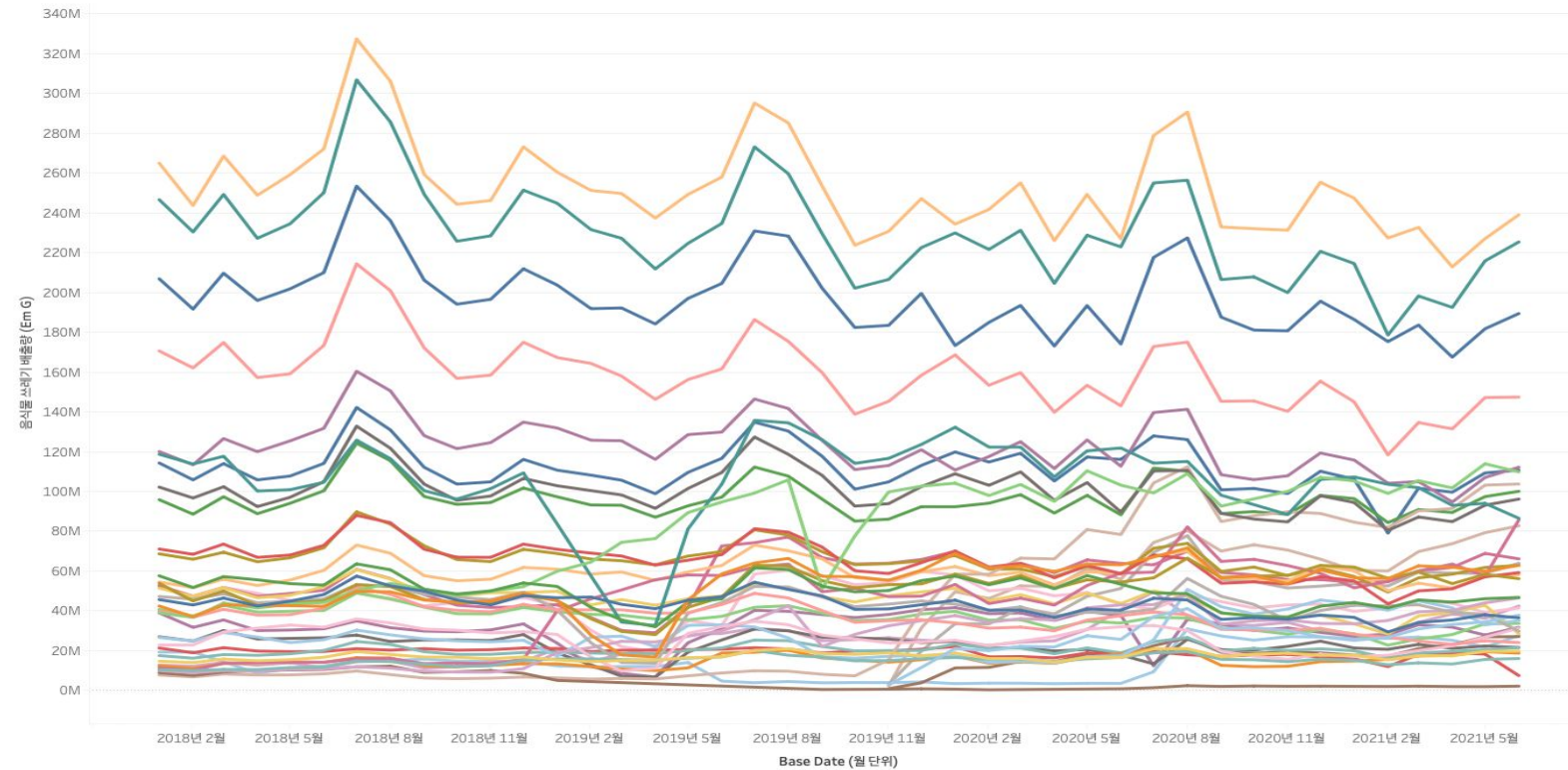


행정동 무작위 표본의 월별 음식물 쓰레기 배출량
(한경면, 구좌읍, 조천읍, 한림읍)



- 데이터 주기성 확인
 - 월별 음식물 쓰레기 배출량 시각화 (전체 행정동)

행정동별 월별 음식물 쓰레기 배출량

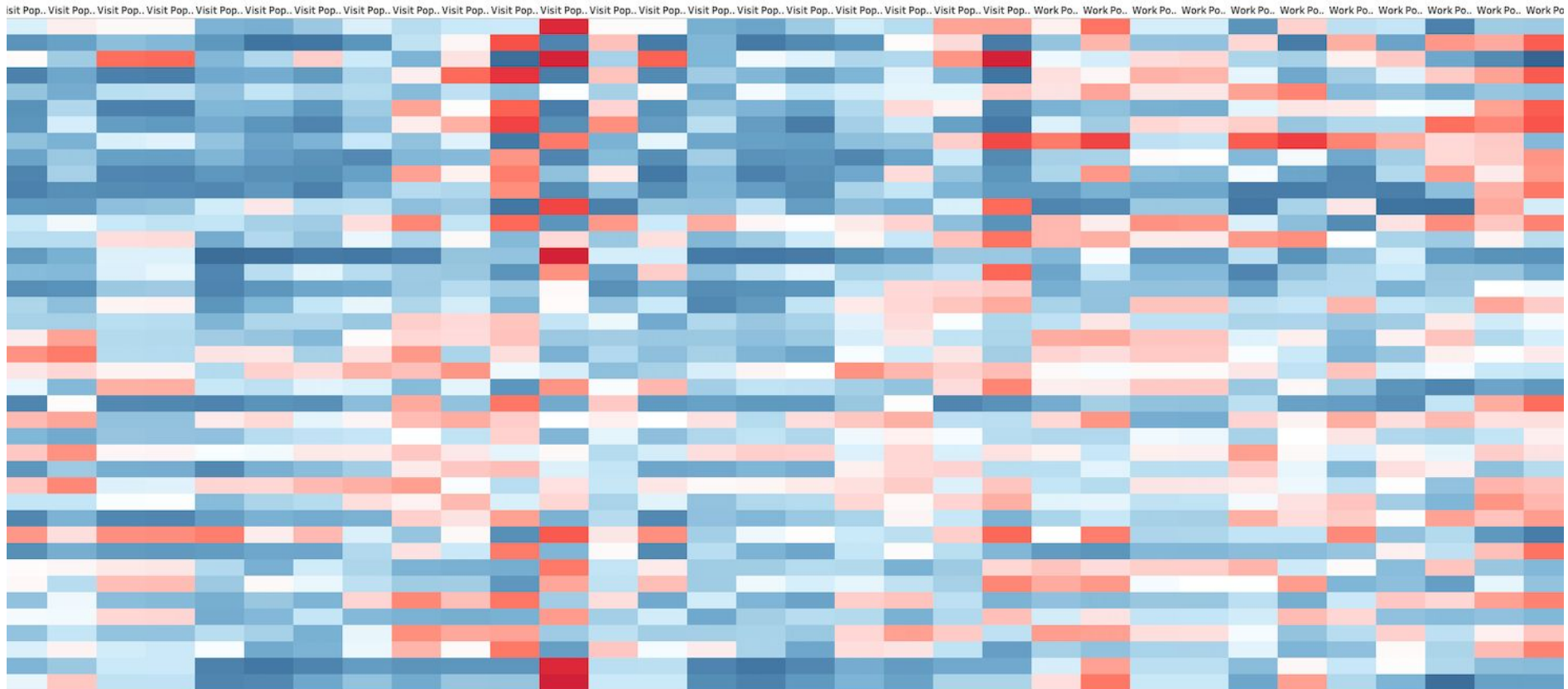


→ 음식물 쓰레기 배출량에서 주기성 시각적으로 확인

EDA



- 음식물 쓰레기 배출량의 주기성과 유사한 주기성을 가진 특성 분석 (행정동별)

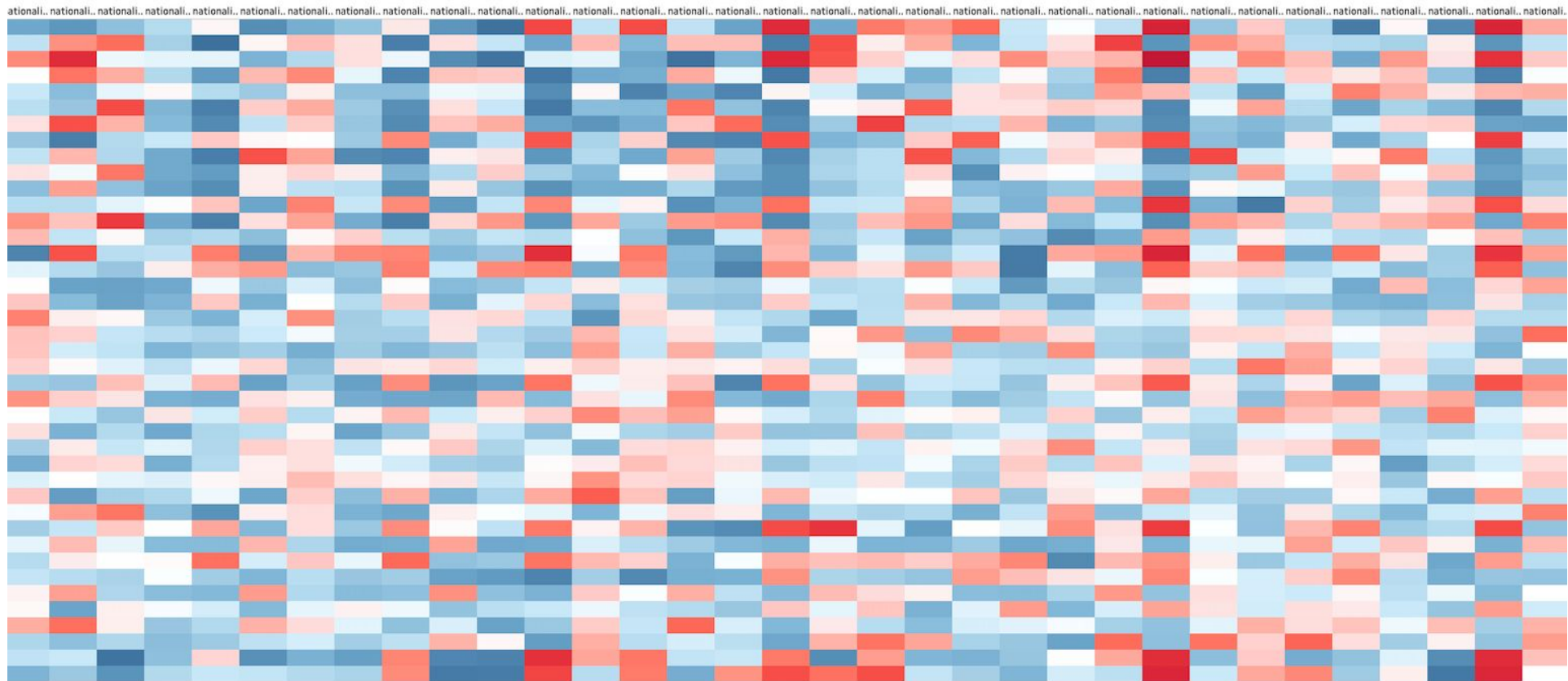


$-1 < \text{Corr} < 1$, 1에 가까울수록 파란색, -1에 가까울수록 빨간색

EDA



- 음식물 쓰레기 배출량의 주기성과 유사한 주기성을 가진 특성 분석 (행정동별)

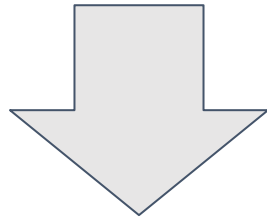


$-1 < \text{Corr} < 1$, 1에 가까울수록 파란색, -1에 가까울수록 빨간색



- 음식물 쓰레기 배출량의 주기성과 유사한 주기성을 가진 특성 분석 (행정동별)

주성분 분석(PCA)을 통해
데이터를 대표하는 특성 파악



주성분 특성과 쓰레기
배출량의 주기 유사성 확인

데이터 전처리



- 데이터 가공 및 병합
 - 행정동 '알수없음' 처리

base_date emd_nm em_cnt em_g pay_amt use_cnt use_amt

201801	알수없음	1107	8880900	266153	7933	290431278
201802	알수없음	950	7543950	226090	4922	195870617
201803	알수없음	1169	9287150	278325	5107	177446688
201804	알수없음	1305	9853350	295266	5060	168054124
201805	알수없음	1585	10290950	308329	5073	187931451

'알수없음' 데이터는 행정동이 표시된 데이터들과 성격이 다르다고 판단하여 분리

32 행, 123 변수
2018.01 ~
2021.06
(빠진 기간 존재)



- 음식물 쓰레기 배출량의 주기성과 유사한 주기성을 가진 특성 분석 (행정동별)

use_amt	0.46556179952862836
resid_reg_pop_m	0.07948236315341631
total_pop_m	0.06334587529741013
resid_reg_pop_w	0.06139580437718961
total_pop_w	0.029393560860414034
resd_pop_cnt	0.024484304666084773
work_pop_cnt	0.019732100378197337
resd_pop_cnt_0	0.018082926046373224
resd_pop_cnt_10	0.015353841626054387
resd_pop_cnt_20	0.013599582307237248
resd_pop_cnt_30	0.01283839691577956
resd_pop_cnt_40	0.011601015300918057
resd_pop_cnt_50	0.010914424381421298
resd_pop_cnt_60	0.009585282078907634
resd_pop_cnt_70	0.008982084794414213
work_pop_cnt_20	0.007960683286993427
work_pop_cnt_30	0.007901229965532297
work_pop_cnt_40	0.007121137374650828
work_pop_cnt_50	0.0070702950774300406
work_pop_cnt_60	0.006429253152553331
work_pop_cnt_80	0.006300065837637898
visit_pop_cnt_0	0.00608675094662277
visit_pop_cnt_10	0.0055165776421081765
visit_pop_cnt_20	0.005385701369642823

데이터의 전체 105개 특성 중 다음 24개의 특성이 전체 분산의 90%를 설명

- use_amt : 결제 금액
- resid_reg_pop_m : 주민등록상 거주 인구 (남성)
- total_pop_m : 총 거주 인구(남성)
- resid_reg_pop_w : 주민등록상 거주 인구 (여성)
- total_pop_w : 총 거주 인구(여성)
- etc ..



- 음식물 쓰레기 배출량의 주기성과 유사한 주기성을 가진 특성 분석 (행정동별)

```
use_amt  
resid_reg_pop_m  
total_pop_m  
resid_reg_pop_w  
total_pop_w  
resd_pop_cnt  
work_pop_cnt  
resd_pop_cnt_0  
resd_pop_cnt_10  
resd_pop_cnt_20  
resd_pop_cnt_30  
resd_pop_cnt_40  
resd_pop_cnt_50  
resd_pop_cnt_60  
resd_pop_cnt_70  
work_pop_cnt_20  
work_pop_cnt_30  
work_pop_cnt_40  
work_pop_cnt_50  
work_pop_cnt_60  
work_pop_cnt_80  
visit_pop_cnt_0  
visit_pop_cnt_10  
visit_pop_cnt_20
```

- 외부 기상데이터, 외부 입도객, 소비자 심리지수(CSI), 코로나 신규 확진자
- 외부 데이터를 원본 데이터에 추가하여 주성분 분석 재실시
=> 기존에 주성분으로 선택된 특성에 영향을 주지 못함



- 음식물 쓰레기 배출량의 주기성과 유사한 주기성을 가진 특성 분석 (행정동별)

부패
주점및주류판매

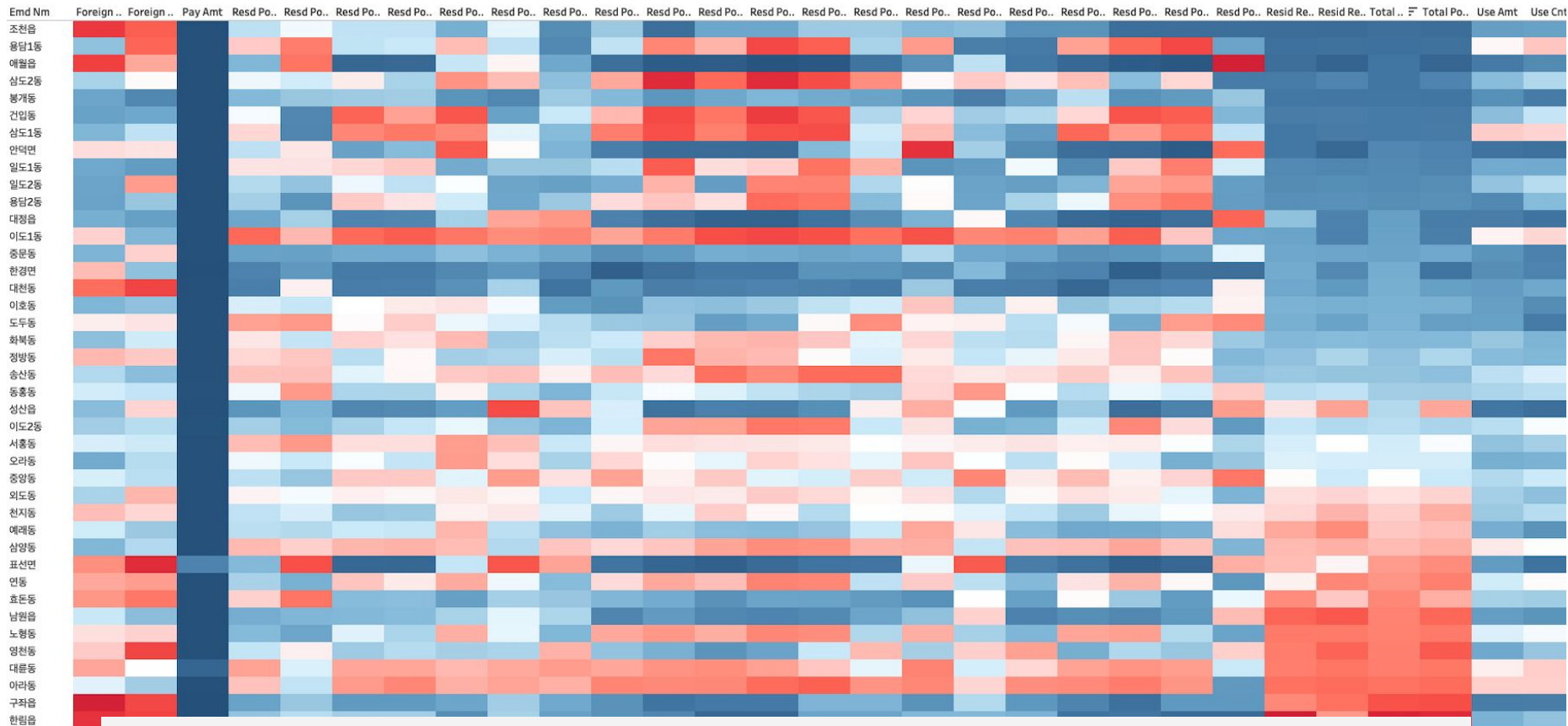
0.2752960683198296
0.11695028804264733

앞선 PCA 결과에서,
결제 금액(use_amt)과 인구수와 관련한 특성을
제거한 후 분산에 가장 많은 영향을 끼치는
특성 2개를 다시 계산한 결과는 다음과 같다.

EDA



- 음식물 쓰레기 배출량의 주기성과 유사한 주기성을 가진 특성 분석 (행정동별)

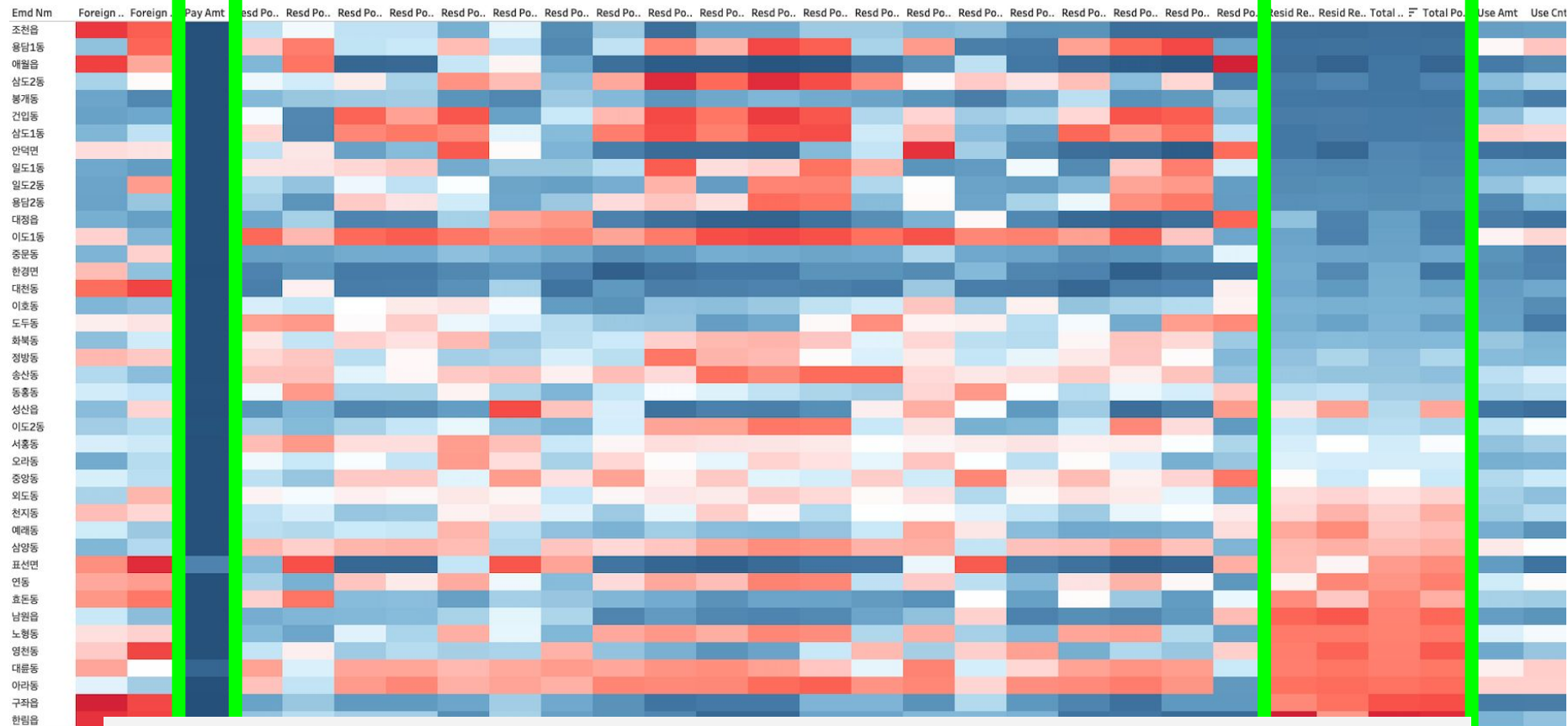


$-1 < \text{Corr} < 1$, 1에 가까울수록 파란색, -1에 가까울수록 빨간색

EDA



- 음식물 쓰레기 배출량의 주기성과 유사한 주기성을 가진 특성 분석 (행정동별)

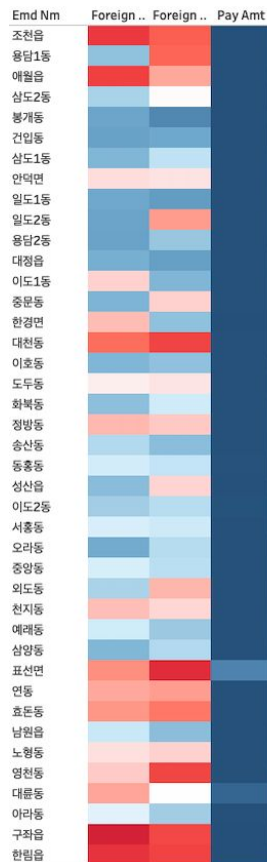


주성분으로 선택된 특성들과 쓰레기 배출량의 주기 유사성

EDA



- 음식물 쓰레기 배출량의 주기성과 유사한 주기성을 가진 특성 분석 (행정동별)

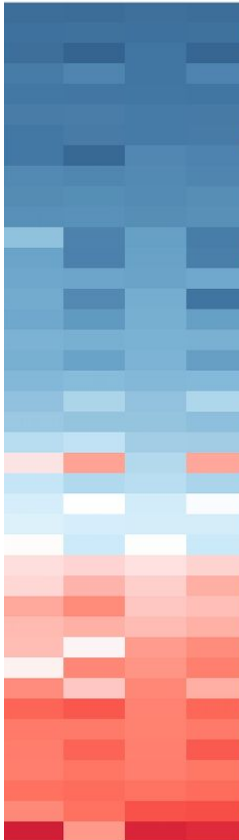


결제 금액 (Pay_Amt) 특성값은 모든 행정동에서
쓰레기 배출량과 매우 유사한 주기성을 보유



- 음식물 쓰레기 배출량의 주기성과 유사한 주기성을 가진 특성 분석 (행정동별)

Resid Re.. Resid Re.. Total .. F² Total Po.



주민등록상 거주 인구,
총 거주 인구의 경우는 쓰레기 배출량과
주기성이 행정동 사이에서 일관적이지 않다

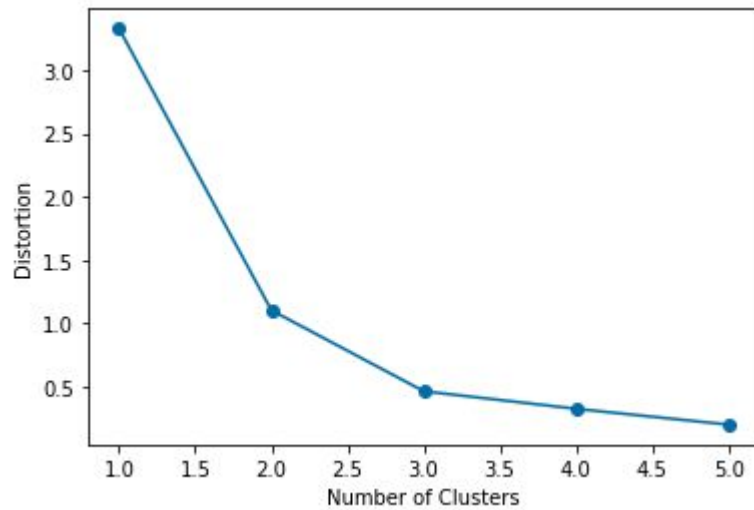


전체 행정동을 해당 특성값을 이용해
유사한 행정동으로 클러스터화

EDA



- 행정동 클러스터링 (K-means Clustering)



특성값으로 주민등록상 거주인구
(resid_pop_w + resid_pop_m)
사용

EDA



- 행정동 클러스터링 (K-means Clustering)

Cluster0

emd_nm
동흥동
서흥동
송산동
오리동
이도2동
이호동
점방동
해양동
화북동

Cluster

1 emd_nm
건입동
대정동
대천동
도두동
봉개동
삼도1동
삼도2동
안덕면
애월읍
용담1동
용담2동
이도1동
일도1동
일도2동
조천읍
중령동
한림면

Cluster

2 emd_nm
구좌읍
남원읍
노량동
대천동
삼양동
성산읍
아라동
연동
영천동
예래동
외도동
천지동
표선면
한림읍
홍동동

```
corr(resd_pop_cnt, em_g) for cluster0 : 0.856575256482039
corr(resd_pop_cnt, em_g) for cluster1 : 0.8286772330836815
corr(resd_pop_cnt, em_g) for cluster2 : 0.8864824039716626
```




4. 모형 설계 및 학습

- 사용 모형에 대한 설명
- 학습 과정에 대한 설명

모형 설계 및 학습



- 사용 모형에 대한 설명

- 분석에 사용된 모형

1. ARIMA 모형

- 2021년 7~8월의 'em_g'를 제외한 특성치(X) 예측
- 2012년 7~8월의 '알 수 없음'의 'em_g' 예측을 위해 사용

2. XGBoost 모형

- 앞서 예측한 2021년 7~8월 features(X)를 이용해 'em_g'를 예측하기 위해 사용

모형 설계 및 학습



- 사용 모형에 대한 설명

- 1. ARIMA

- AR(AutoRegression) 모형과 MA(Moving Average) 모형을 합친 모형
 - 시계열 데이터의 정상성(Stationary)를 가정

※ 정상성 : 평균, 분산이 시간에 따라 일정한 성질

(1) AR 모형 : 자기상관성을 시계열 모형으로 구성, 이전 자신의 관측값이 이후 자신의 관측값에 영향을 준다는 아이디어 기반

(2) MA 모형 : 예측 오차를 이용하여 미래를 예측하는 모형

∴ $ARIMA(p, d, q)$: d 차 차분한 데이터에 $AR(p)$, $MA(q)$ 모형을 합친 모형



- 사용 모형에 대한 설명

2. XGBoost

- XGBoost 모형의 정의

- (1) 여러 개의 Decision Tree를 조합해 사용하는 Ensemble 알고리즘
- (2) Gradient Boosting 알고리즘을 분산환경에서도 실행할 수 있도록 구현해놓은 라이브러리

- XGBoost 모형의 특징

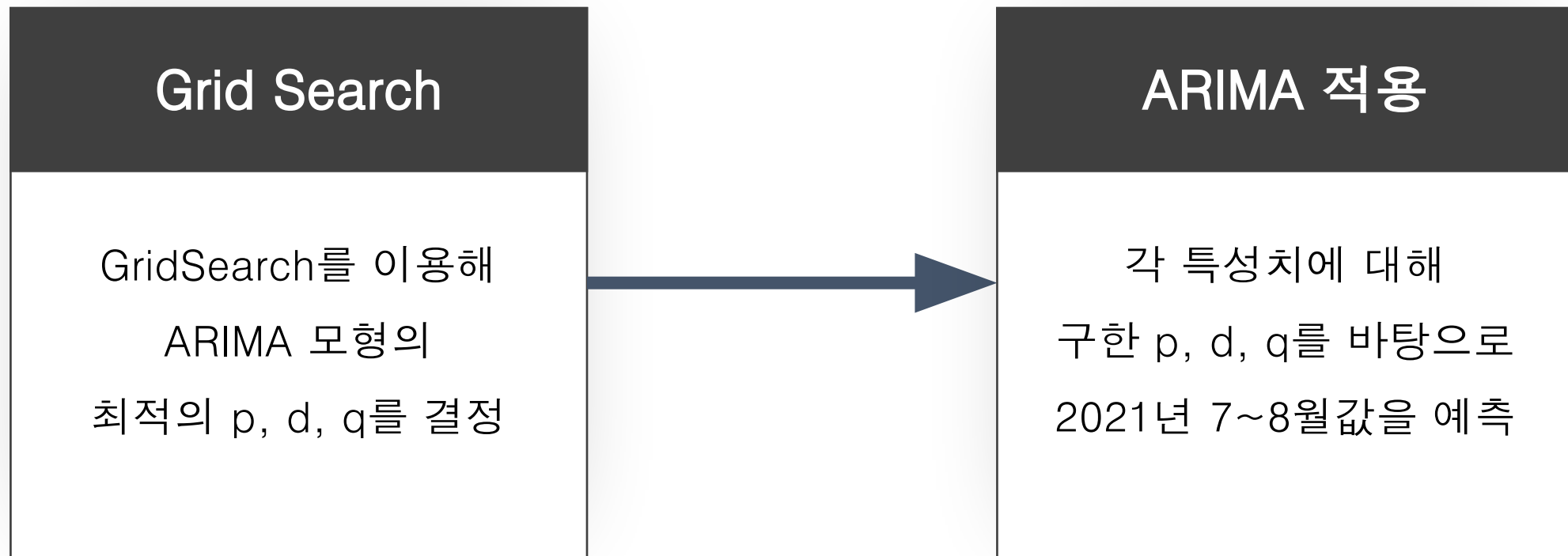
- (1) 성능이 좋고, 컴퓨팅 자원 활용률이 좋음
- (2) Overfitting을 방지할 수 있음
- (3) Cross Validation을 지원

모형 설계 및 학습



- 학습 과정에 대한 설명

1. 'em_g'를 제외한 모든 특성치 예측



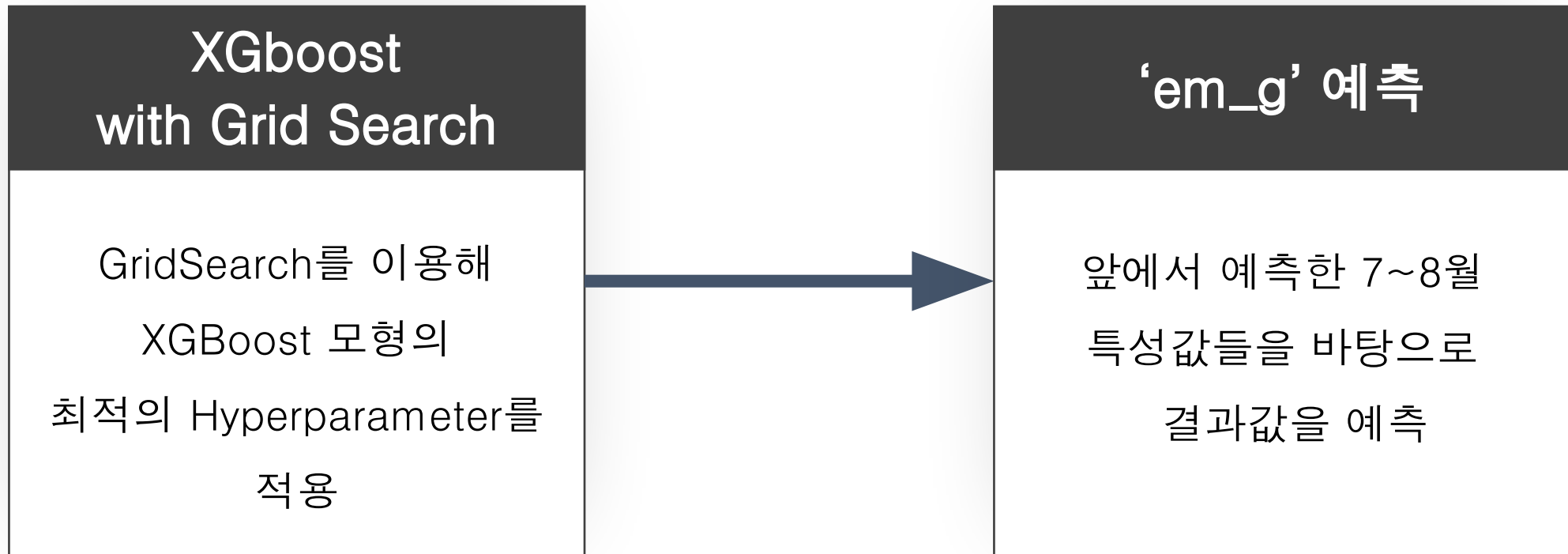
모형 설계 및 학습



- 학습 과정에 대한 설명

- 2. 'em_g' 예측

※ cluster별로 학습을
진행

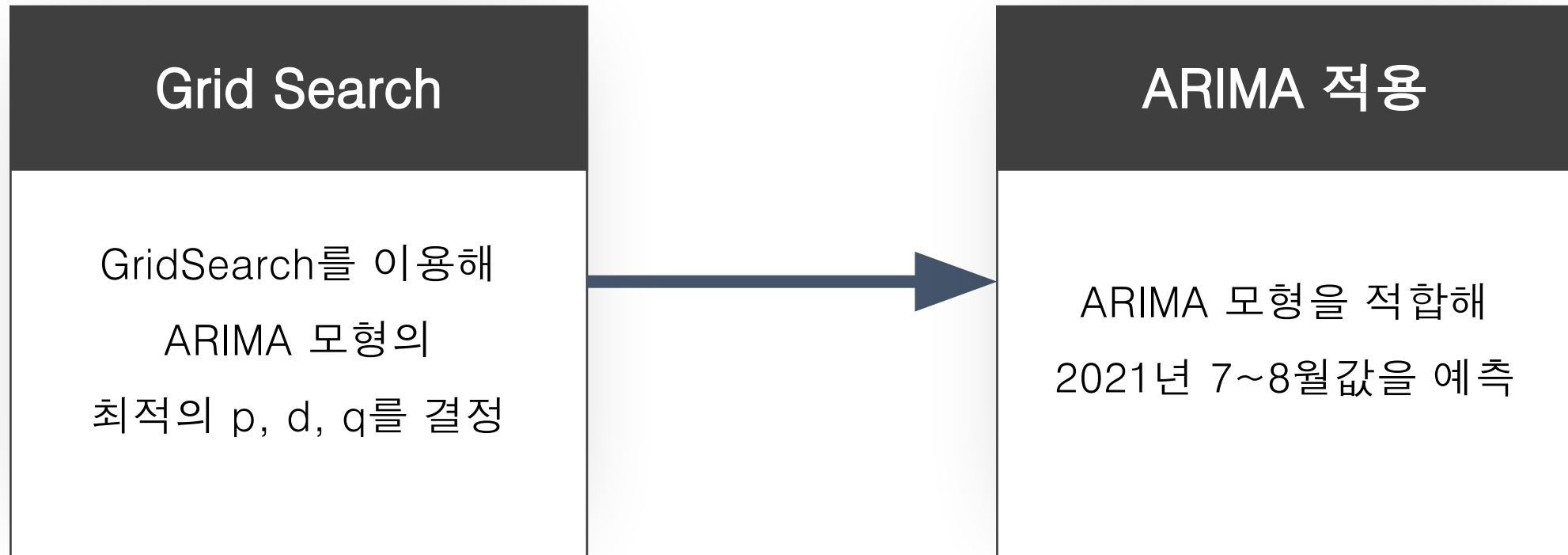


모형 설계 및 학습



- 학습 과정에 대한 설명

- 3. '알 수 없음'의 'em_g' 예측





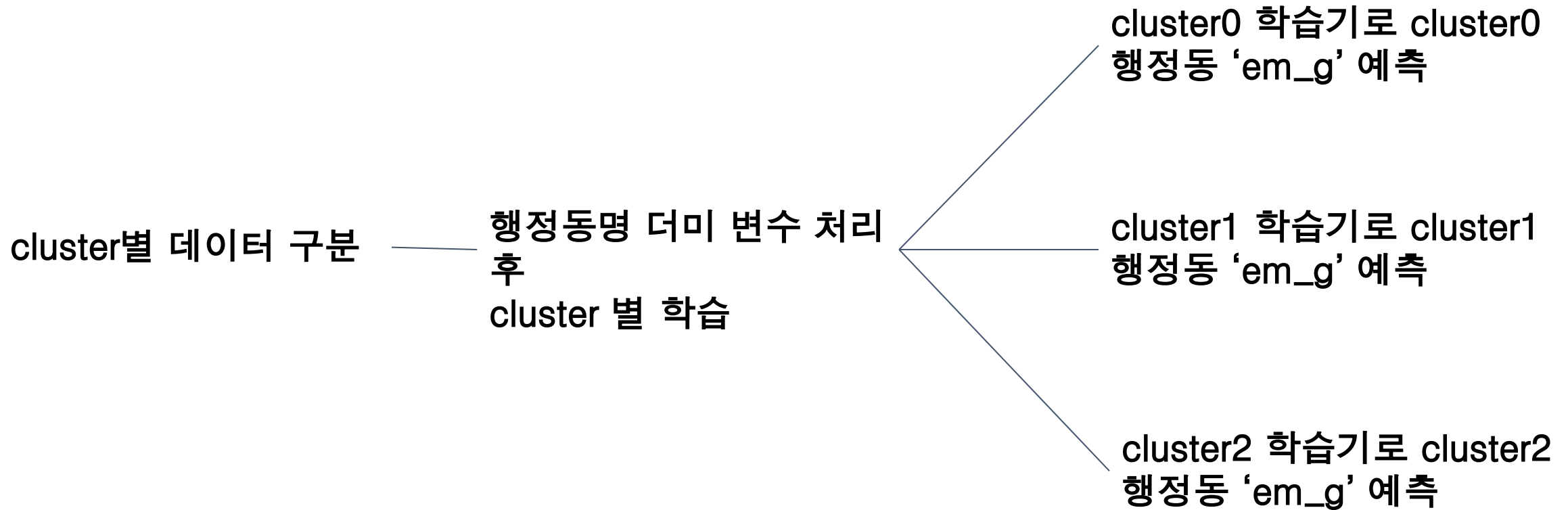
5. 예측 결과

- 예측 결과

예측 결과



- 행정동별 'em_g' 예측 과정



예측 결과



● 행정동별 'em_g' 예측 결과

3	NO	행정동명	7월 배출량(g)	8월 배출량(g)
4	1	한림읍	76778448	76256576
5	2	애월읍	108957280	112246832
6	3	구좌읍	39795844	34764300
7	4	조천읍	61386704	58378620
8	5	한경면	29644448	31131196
9	6	일도1동	14373044	21500950
10	7	일도2동	147700064	145234720
11	8	이도1동	33096408	37137744
12	9	이도2동	224458592	225256352
13	10	삼도1동	58883848	58818372
14	11	삼도2동	22848922	22142278
15	12	용담1동	33627776	33511354
16	13	용담2동	62890780	61851712
17	14	건입동	35695204	35511996
18	15	화북동	109995856	109876696
19	16	삼양동	94409216	93803336
20	17	봉개동	16025016	16015494
21	18	아라동	114257904	115207664
22	19	오라동	58733140	56874364
23	20	연동	185771776	185866384
24	21	노형동	236255888	234016224

특성값으로 주민등록상 거주인구
(resid_pop_w + resid_pop_m)
사용

예측 결과



● 행정동별 'em_g' 예측 결과

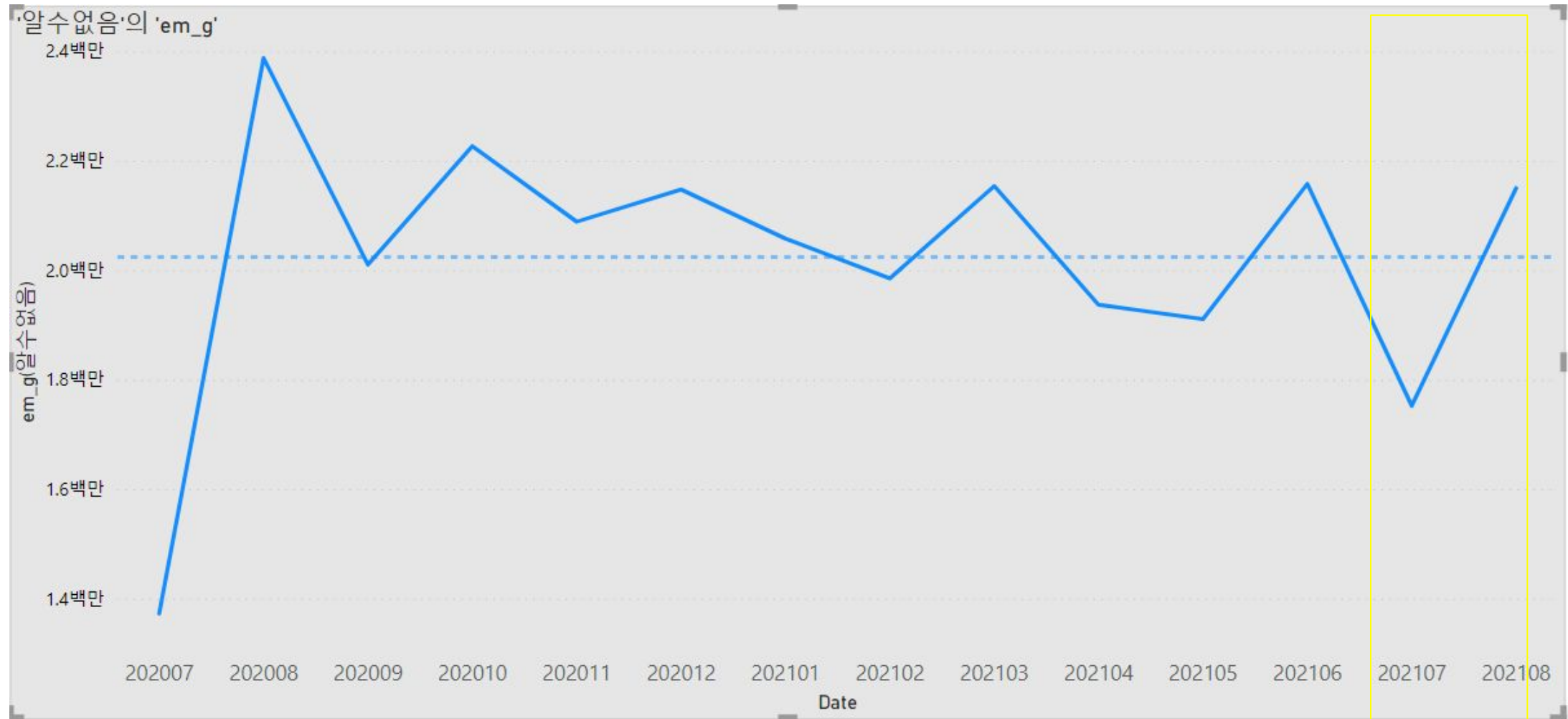
▲ 3	NO	행정동명	7월 배출량(g)	8월 배출량(g)
▼ 25	22	외도동	97666608	103085152
26	23	이호동	21121750	21078784
27	24	도두동	19816956	20113046
28	25	대정읍	112220072	112106424
29	26	남원읍	63969024	64113980
30	27	성산읍	74959552	74000608
31	28	안덕면	47442136	47865424
32	29	표선면	38173860	39983416
33	30	송산동	34142488	34203384
34	31	정방동	21618860	22297038
35	32	중앙동	41485476	41446584
36	33	천지동	31318898	32446716
37	34	효돈동	28322548	29853240
38	35	영천동	38613440	40164232
39	36	동흥동	84087000	84301392
40	37	서흥동	28846006	28837862
41	38	대륜동	46407076	53401300
42	39	대천동	57303364	57127100
43	40	중문동	66802720	67634288
44	41	예래동	19326560	19311850
45	42	알수없음	1752308	2149913

특성값으로 주민등록상 거주인구
(resid_pop_w + resid_pop_m)
사용

예측 결과



- 알수없음 'em_g' 예측 결과





6. 분석 결과

- 분석 결과 요약
- 분석 결과 활용 및 해결 방안
 - 기대효과

분석 결과



● 분석 결과 요약

- 코로나 신규 확진자 수, 제주도 입도객 수는 음식물 쓰레기 배출량에 많은 영향을 끼치지 않는다.
- 결제 금액과 음식물 쓰레기 양이 높은 상관관계를 보이는 것을 통해 많이 소비할수록 많은 양의 음식물 쓰레기가 나온다는 것을 확인할 수 있다.
- 음식물 쓰레기 배출량과 배출량에 높은 영향을 주는 특성들이 주기적으로 겨울철보다 여름철에 더 큰 값을 가지는 것을 확인할 수 있다.
- 음식물 쓰레기 배출량과 배출량에 높은 영향을 주는 특성들이 주기적으로 겨울철보다 여름철에 더 큰 값을 가지는 것을 확인할 수 있다.

분석 결과



- 분석 결과 활용 및 해결 방안

- 소비량, 인구 수에 따라 음식물 쓰레기 양이 늘어난다는 것을 확인할 수 있다.

따라서 총 음식물 쓰레기 배출량을 줄이기 위해서는 인당 음식물 쓰레기 배출량을 줄이는 것이 가장 효과적이라는 결론을 얻을 수 있다.

- 앞서 나눈 3개의 행정동 클러스터는 주민등록상 거주 인구, 총 거주 인구와 쓰레기 배출량의 주기성이 유사한 행정동의 집합이다.

따라서 클러스터별 각 행정동의 특징에 맞는 음식물 쓰레기 배출량 감소 방안을 적용하는 것이 좋을 것이라는 결론을 얻을 수 있다.

분석 결과



- 분석 결과 활용 및 해결 방안

- 해결 방안으로 음식물 쓰레기 배출량이 많을 것이라고 예상되는 기간에 인당 음식물 쓰레기 배출량이 감소되도록 유도하는 캠페인을 실시하는 방안을 생각할 수 있다.
- 또한 부페 특성이 인구와 관련한 특성을 제외하고 음식물 쓰레기 소비량에 가장 많은 영향을 주는 것을 확인할 수 있다.

각 부페에 설치된 POS기기 등으로부터 사용자와 음식 소비량등의 시계열 데이터를 얻고, 이 데이터를 이용해 향후 사용자와 음식 소비량 등을 정량적으로 예측하여 남은 음식물 양을 최소화하는 방법이 가능하다.



7. 보완점 및 후속 연구

보완점 및 후속 연구



● 보완점 및 후속 연구

- 가져온 외부 데이터 특성들이 특성 선택에 큰 영향을 끼치지 못하였는데, 실제 특성 선택에 큰 영향을 발휘할 수 있는 외부 데이터를 찾는 작업을 추가로 실시할 수 있다.
- 행정동 클러스터링 방법에 대한 후속 연구로, 특성 선택에 영향을 주는 외부 데이터가 존재한다면 클러스터별로 더 많은 변수를 추가해 모델링을 시도할 수 있다.
- 제주도의 행정구역별 지도를 시각화하기 위한 라이브러리 등의 개발 도구가 부족했고, 행정동 별 도움이 될 수 있는 사전 정보를 구하는 것이 용이하지 않았다.
행정동 별 사전 정보가 더 많이 주어진다면 이를 분석에 반영하여 추가 연구를 진행할 수 있다.
- ‘알 수 없음’ 레이블로 제공된 데이터를 하나의 행정구역으로 여기고 분석을 진행했기 때문에 ‘알 수 없음’ 데이터 예측이 부정확할 가능성이 있다.



Q&A

감사합니다.

2021 빅콘테스트 - 제주
ECO

청정퓨어클린제주

- 임성준, lsj9862@gmail.com
- 권휘성, 0518hs@naver.com
- 김승종, sjkhalladay@gmail.com