



데이터 분석 분야-퓨처스리그-eco jeju

제주도 음식물양 쓰레기 예측을 통한 해결 방안 제시

팀명:감귤맛 삼다수

팀장:홍지원(ant67410@gmail.com)

팀원

권예은(yenny7013@gmail.com)

이보림(ant67410@gmail.com)

유승희(juneyo0823@naver.com)

INDEX



01.분석 배경 및 목표

- ▶ 분석 배경
- ▶ 분석 방향



02.데이터 분석

- ▶ EDA
- ▶ Feature Engineering
- ▶ 최종 DataSet
- ▶ Test Set



03.Modeling

- ▶ 모델 선정
- ▶ 모델 성능
- ▶ 예측 결과



04.해결 방안

- ▶ 해결 방안
- ▶ 의의 및 한계점



05.Reference

- ▶ 분석 도구 및 출처

01. 분석 배경 및 방향

봉개 쓰레기 소각장 포화 ‘여전’...2만7000t 야적

출처 : 제주일보

환경자원순환센터, 압축쓰레기 감안 안해

봉개동 매립장에 야적된 압축쓰레기 모습.
2019년 제주시 구좌읍 동북리에 제주환경자원순환센터(광역소각장)가 준공됐지만, 여전히 봉개동 소각장에는 압축쓰레기가 산더미처럼 쌓여있다.

3일 제주시에 따르면 지난달 기준 2만7500t의 압축쓰레기가 야적돼 있다. 높이 1m, 너비 1m인 쓰레기 더미 한 개당 무게는 900kg에 달한다. 제주시는 지난해부터는 봉개동 소각장에서 하루 평균 50t의 압축쓰레기를 소각·처리하고 있다.

제주시는 밀려드는 쓰레기를 감당하지 못하자 2015년부터 2019년까지 비닐·종이·플라스틱류를 잘게 분쇄해 압축한 압축쓰레기를 생산하기 시작했다.

제주시는 압축쓰레기를 고형연료로 만들어 도외지역으로 반출하고, 2019년에는 제주의 압축 쓰레기를 필리핀 등 다른 지역으로 반출하는 과정에서 문제가 발생해 국제적인 망신을 당하기도 했다.

제주환경자원순환센터의 최대 소각 용량은 1일 500t이지만 하루 평균 387t의 쓰레기를 소각하고 있다.

제주 명품 이호테우 해변, 실종된 시민의식에 쓰레기 천국

출처 : 국제뉴스

제주 음식물처리 시설 첫 삽... 봉개동 기한만료에 ‘발등에 불’

출처 : 국제뉴스

청정 자연의 제주도는 최근 일반 쓰레기, 음식물 쓰레기의 발생량이 포화 상태에 다다르자 해결하기 위해 각 부처에서 노력을 기하고 있음

특히, 요일 별 쓰레기 종량제, 음식물 쓰레기 종량제를 읍, 면으로 확대하는 등 제주도 지자체 내에서 노력을 지속

하지만, 대표적인 관광지로서 올 여름에도 관광지들이 몸살을 앓고 있는 모습이 문제로 대두

01. 분석 배경 및 방향

분석 배경

“분석 방향”



인구 증가의 트렌드를
반영한 인구통계적 요소



주요 산업인 관광업을 중심
으로 한 산업적 요소



두 요소의 영향으로
변화하는 경제적 요소

02.데이터 분석

02.데이터 분석

“EDA”

Feature Engineering

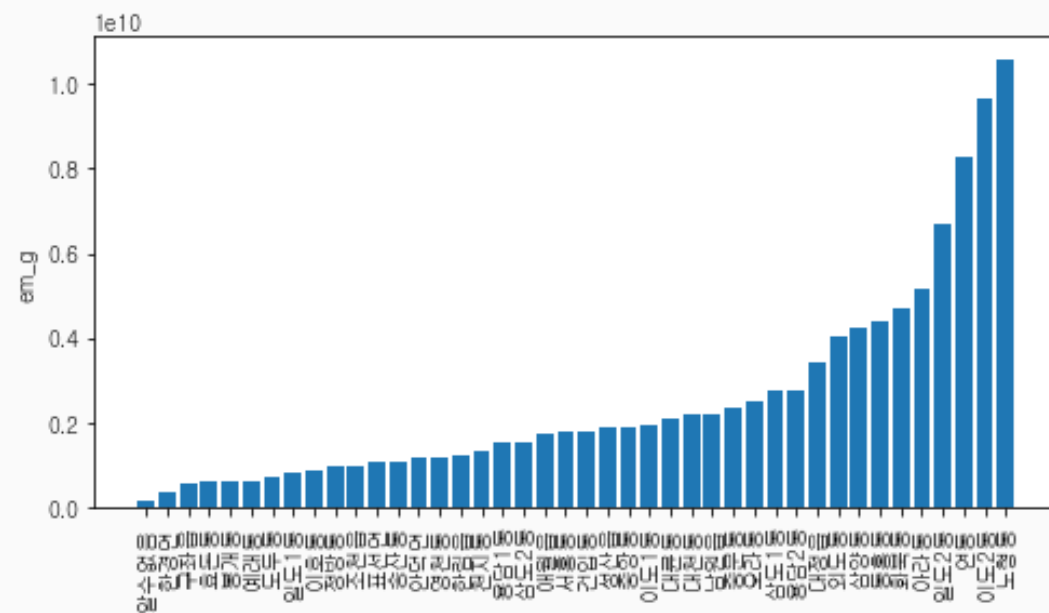
최종 Data Set

Test Set

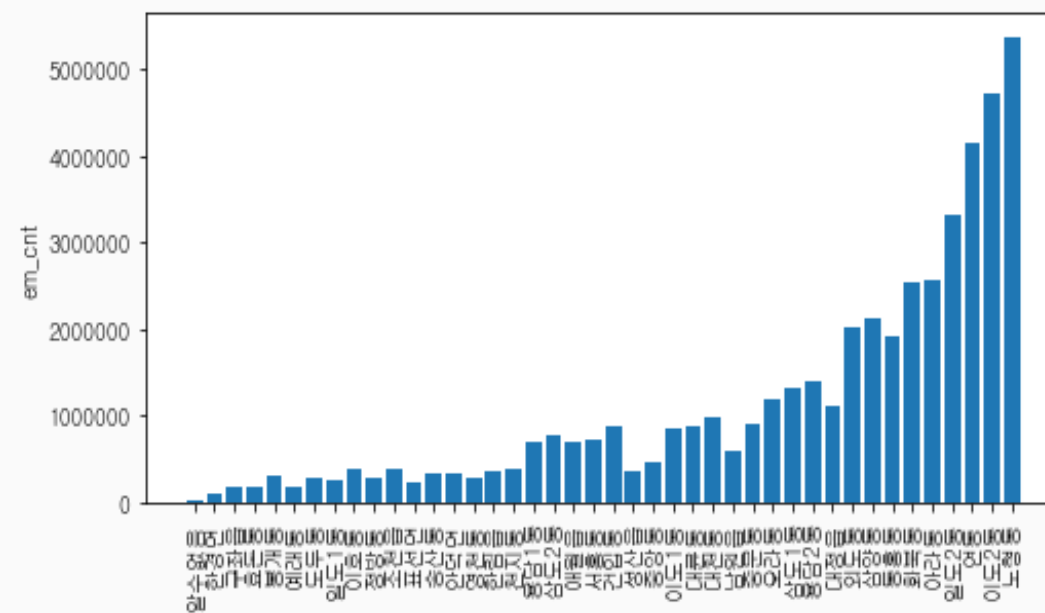
“ 01_음식물쓰레기_FOOD_WASTE_210811_update.CSV ”

<읍면동별 음식물 쓰레기>

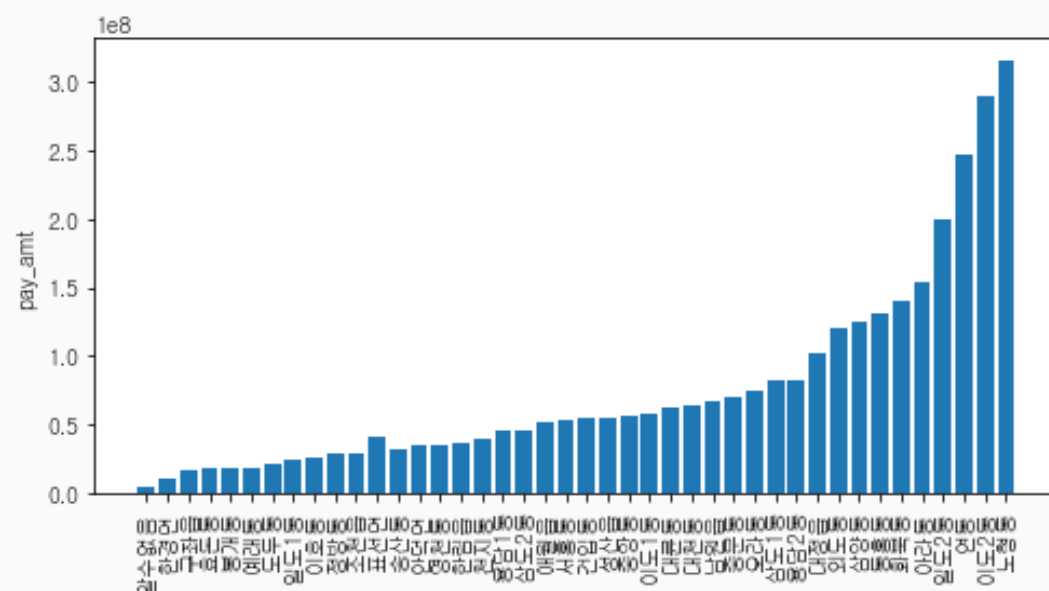
● 읍면동별 음식물 쓰레기 배출량



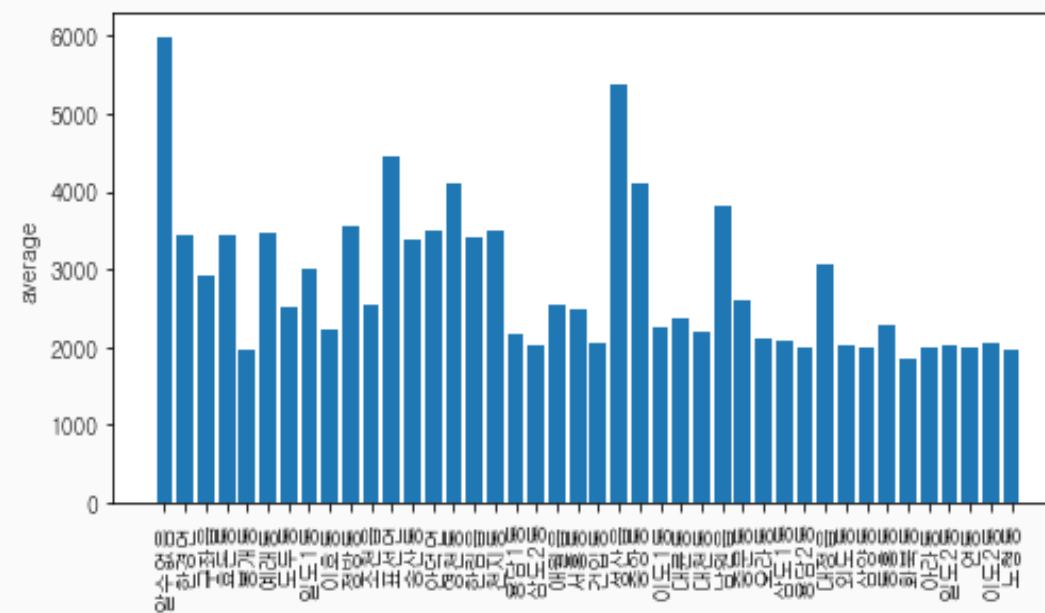
● 읍면동별 음식물 쓰레기 배출건수



● 읍면동별 총 지불금액



● 읍면동별 음식물 쓰레기 평균 배출량



음식물 쓰레기 배출량 ,
배출건수, 총 지불 금액
공동 Top 5 지역

노형동
이도2동
연동
일도2동
아라동

02.데이터 분석

“EDA”

Feature Engineering

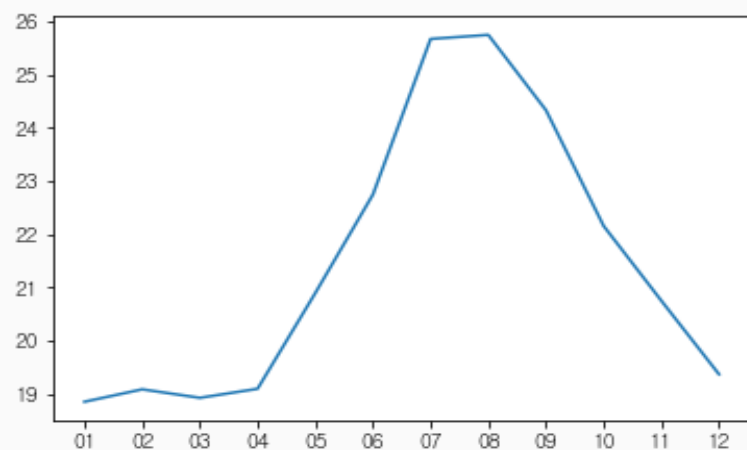
최종 Data Set

Test Set

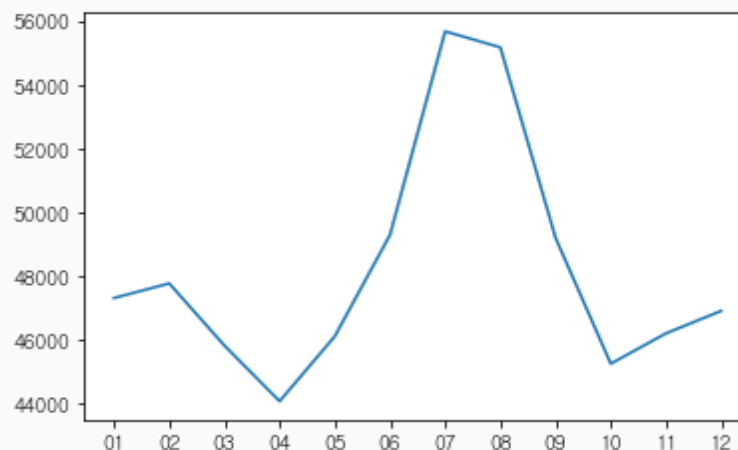
“ 01_음식물쓰레기_FOOD_WASTE_210811_update.CSV ”

<월별 음식물 쓰레기>

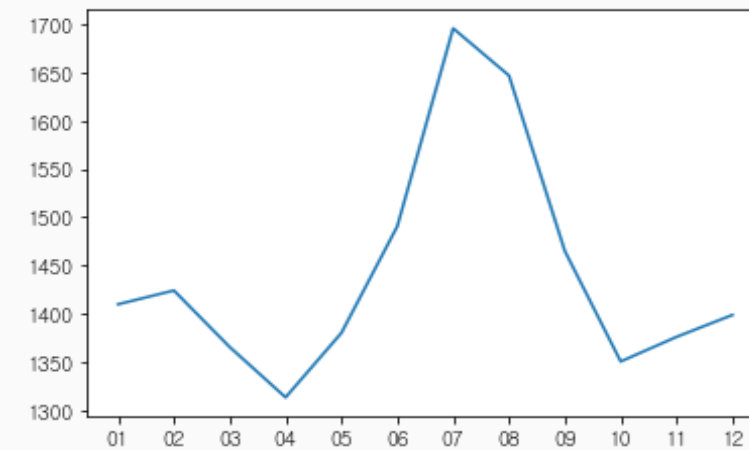
● 월별 배출건수



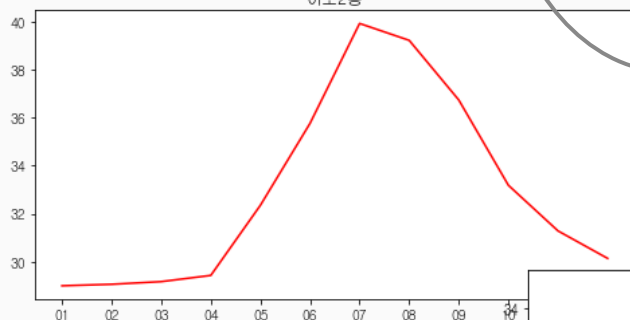
● 월별 음식물 쓰레기 배출량



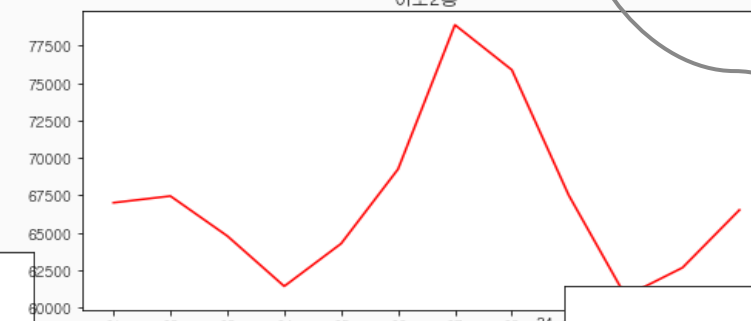
● 월별 지불금액



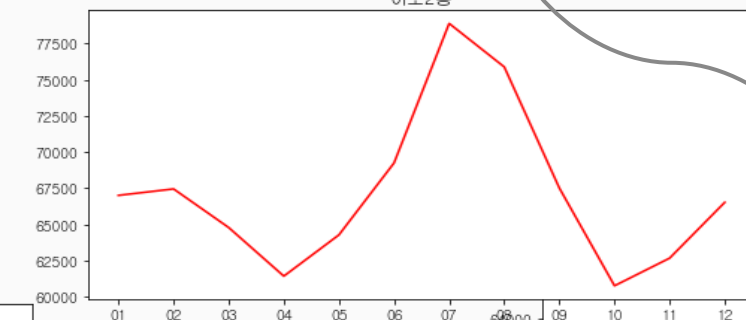
이도2동



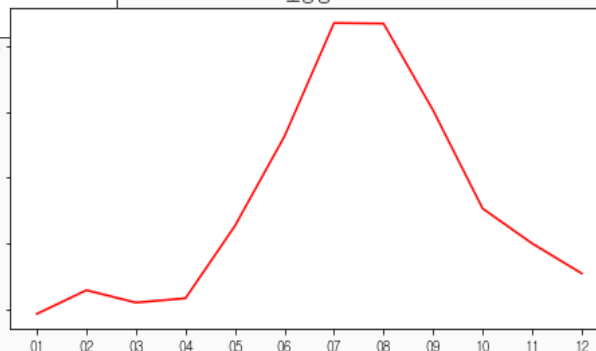
이도2동



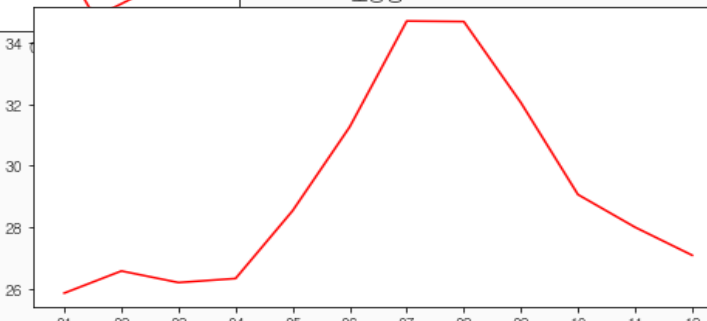
이도2동



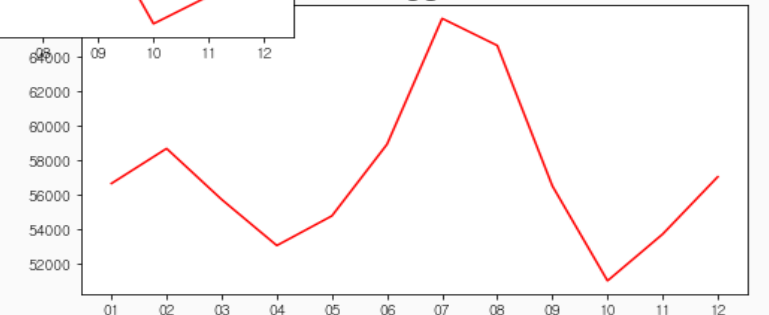
노형동



노형동



노형동



“

전체적으로 읍면동별 월별 음식물 쓰레기 배출건수, 배출량, 지불금액 추이는
제주도 전체의 월별 음식물 배출건수, 배출량, 지불금액을 따름

”

02.데이터 분석

“EDA”

Feature Engineering

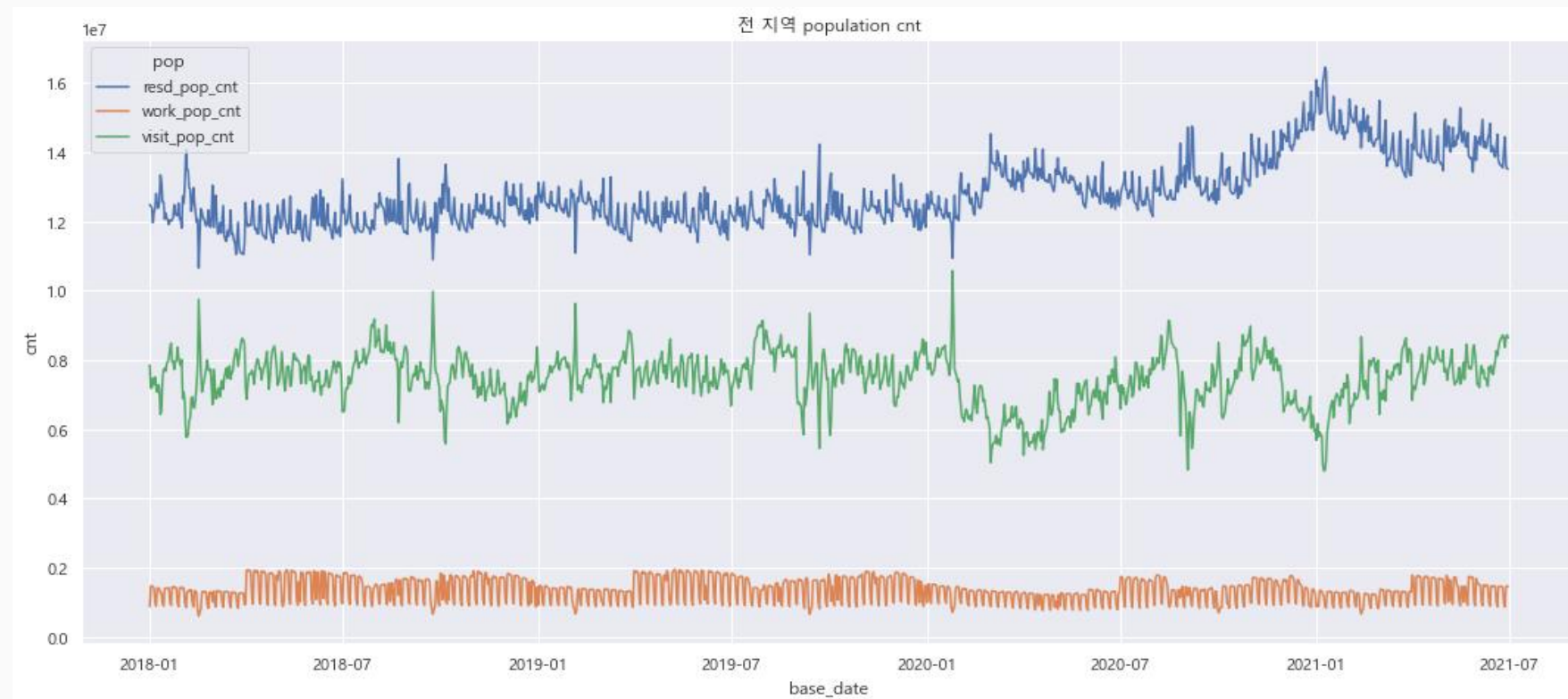
최종 Data Set

Test Set

“ 02-1_내국인유동인구_KOREAN.CSV ”

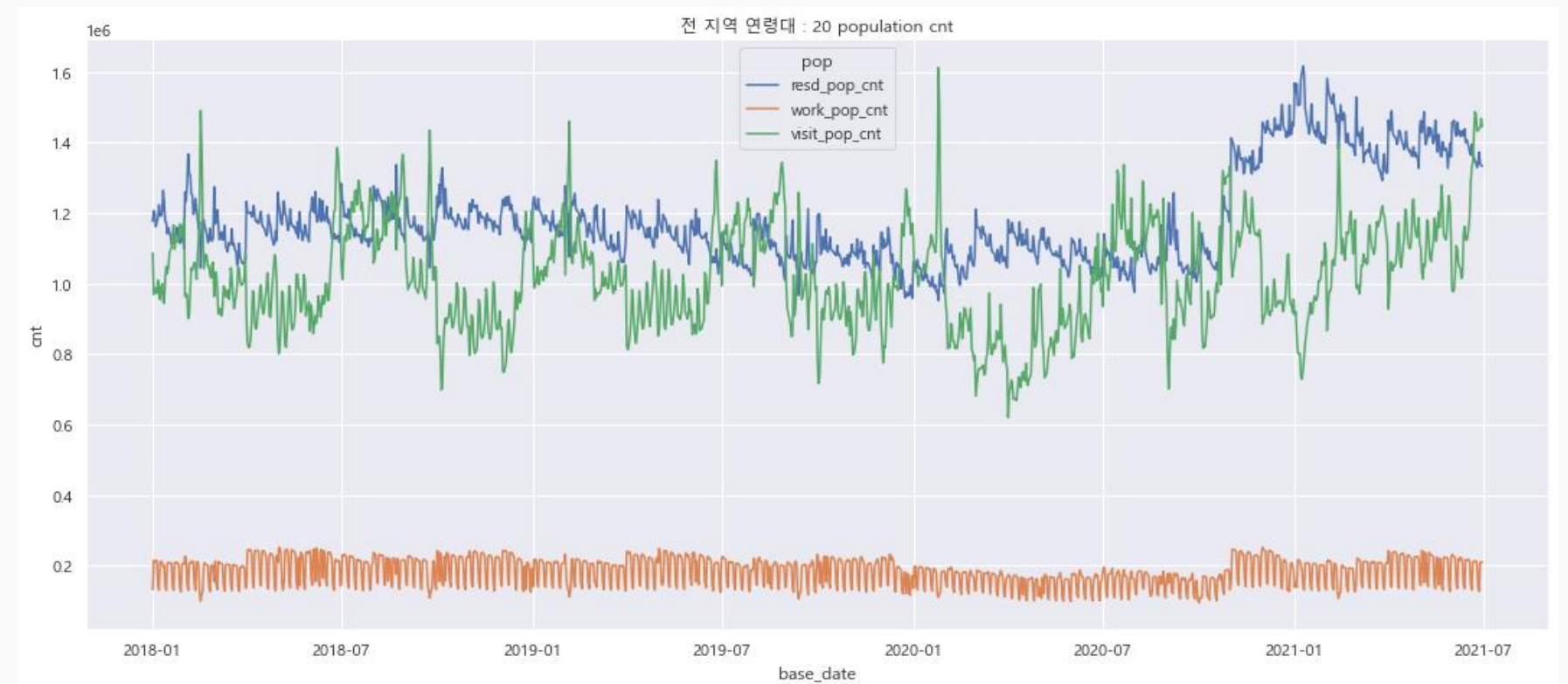
<전체 기간 기준>

● 유형별 유동인구 추이



제주도 전체지역의 인구 유형별 추이를 나타낸 그래프
거주인구, 방문인구, 근무인구 순으로 많아 나타나고 있음.

● 연령별 유동인구 추이



연령별 유동인구 중 20대의 유동인구 추이 그래프
특정 기간대에는 방문인구가 근무인구보다 더 많이 발생.

02.데이터 분석

“EDA”

Feature Engineering

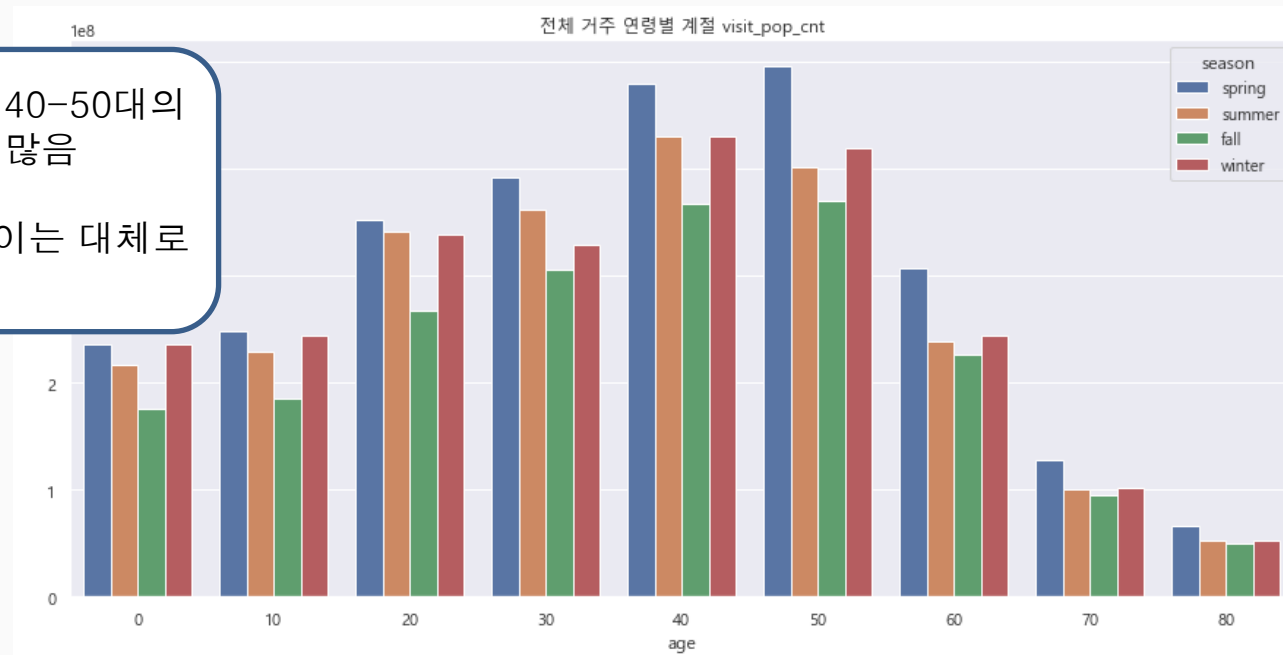
최종 Data Set

Test Set

“ 02-1_내국인유동인구_KOREAN.CSV ”

<방문 인구 기준>

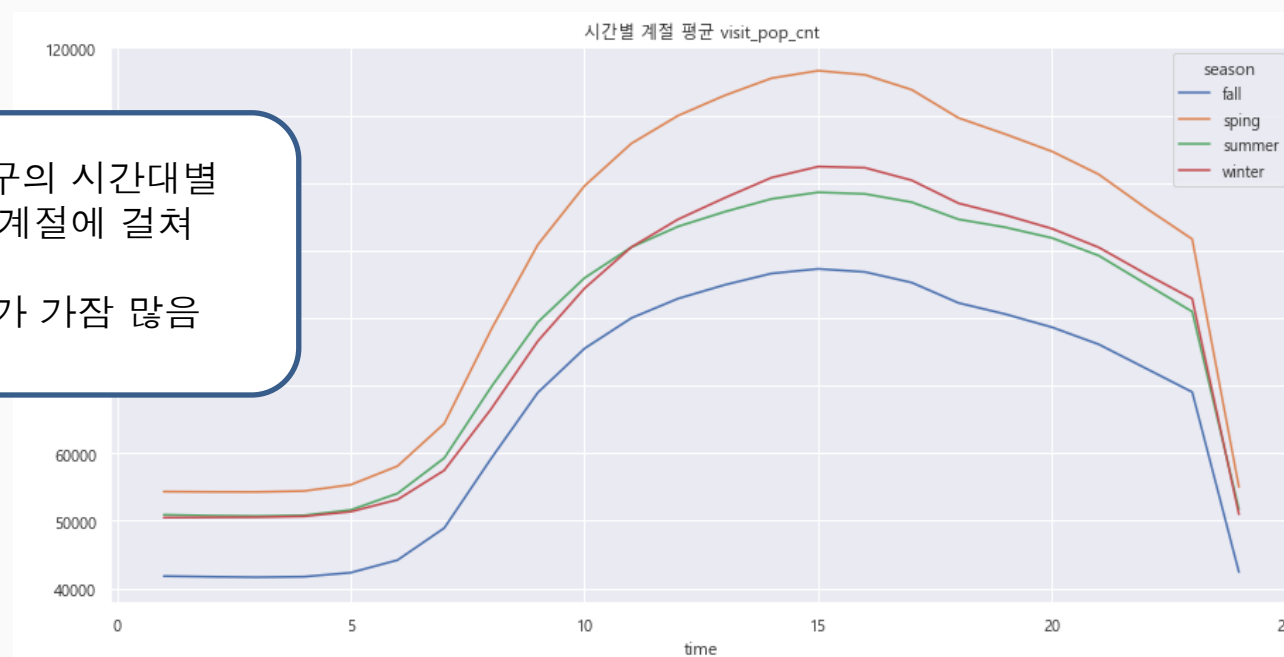
● 연령대별 계절별



내국인 유동인구 중 40-50대의 방문인구 수가 가장 많음

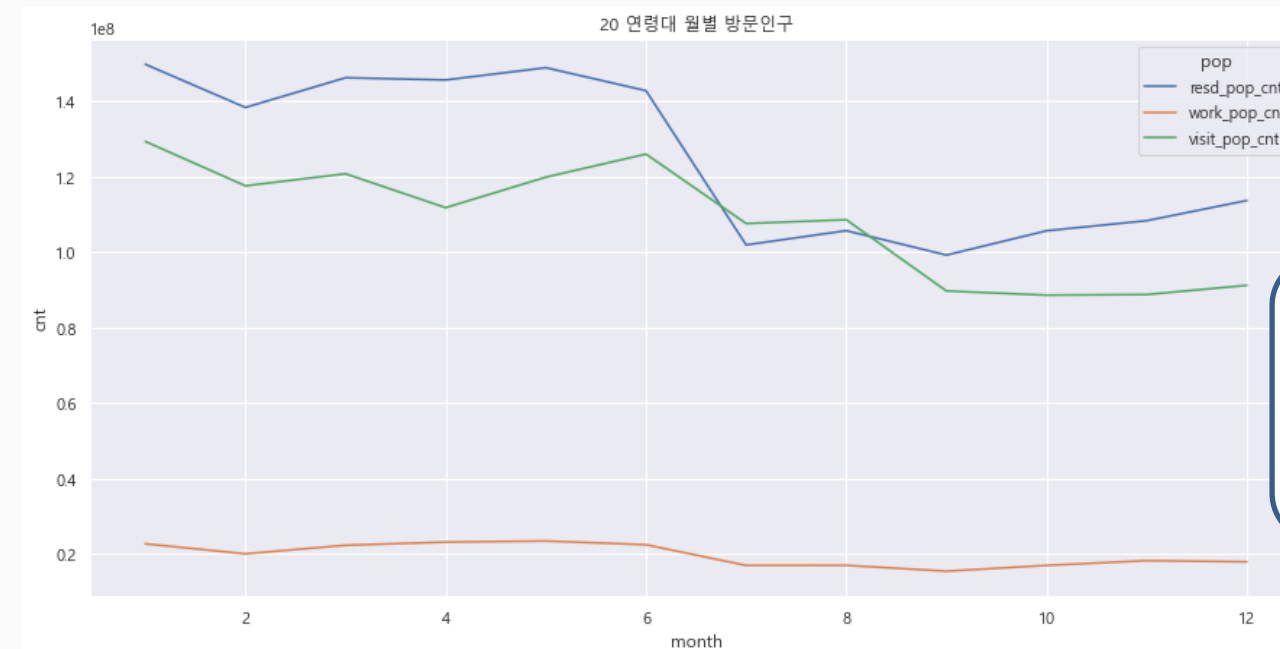
연령대별 계절별 추이는 대체로 비슷한 양상

● 시간별 계절별



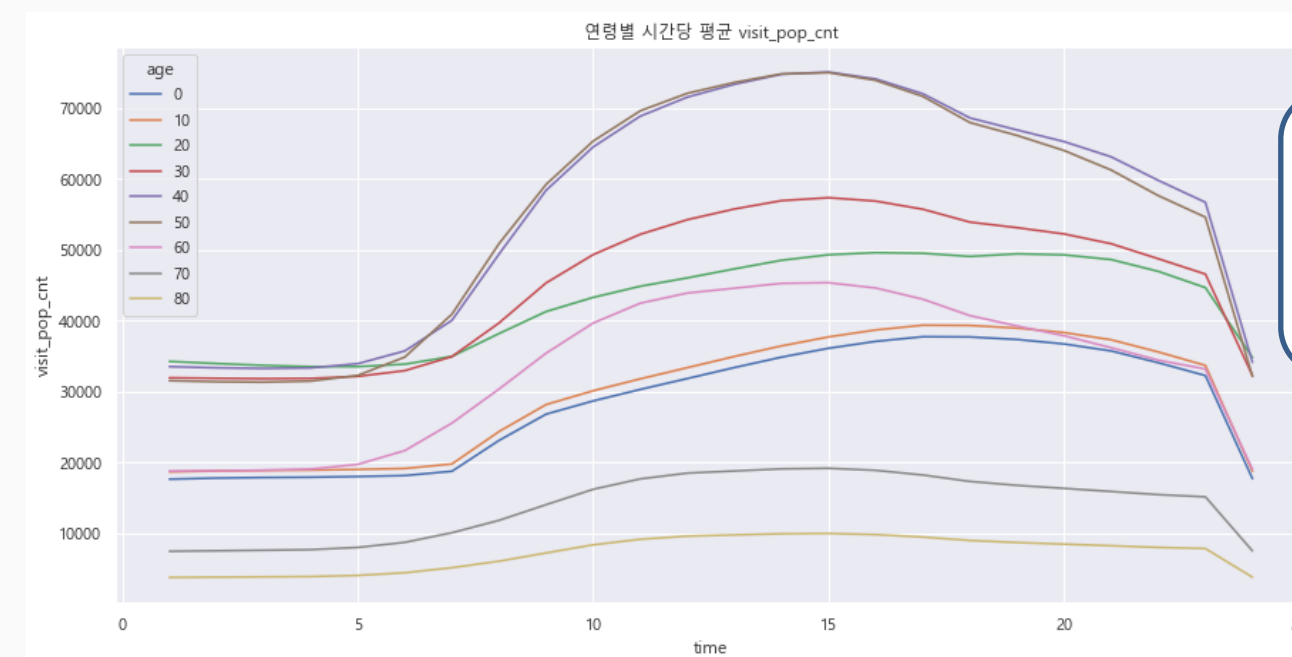
내국인 방문인구의 시간대별 인구수는 모든 계절에 걸쳐 비슷한 추세
15시에 인구수가 가장 많음

● 월별 유형별



월별 내국인 수는 거주인구, 방문인구, 근무인구 순으로 많음
7~8월은 방문인구가 거주인구수보다 높음

● 시간별 연령별



내국인 방문인구의 연령대별 인구수는 모든 연령대에 걸쳐 비슷한 추세
10시~20시 인구수가 많음

02.데이터 분석

“EDA”

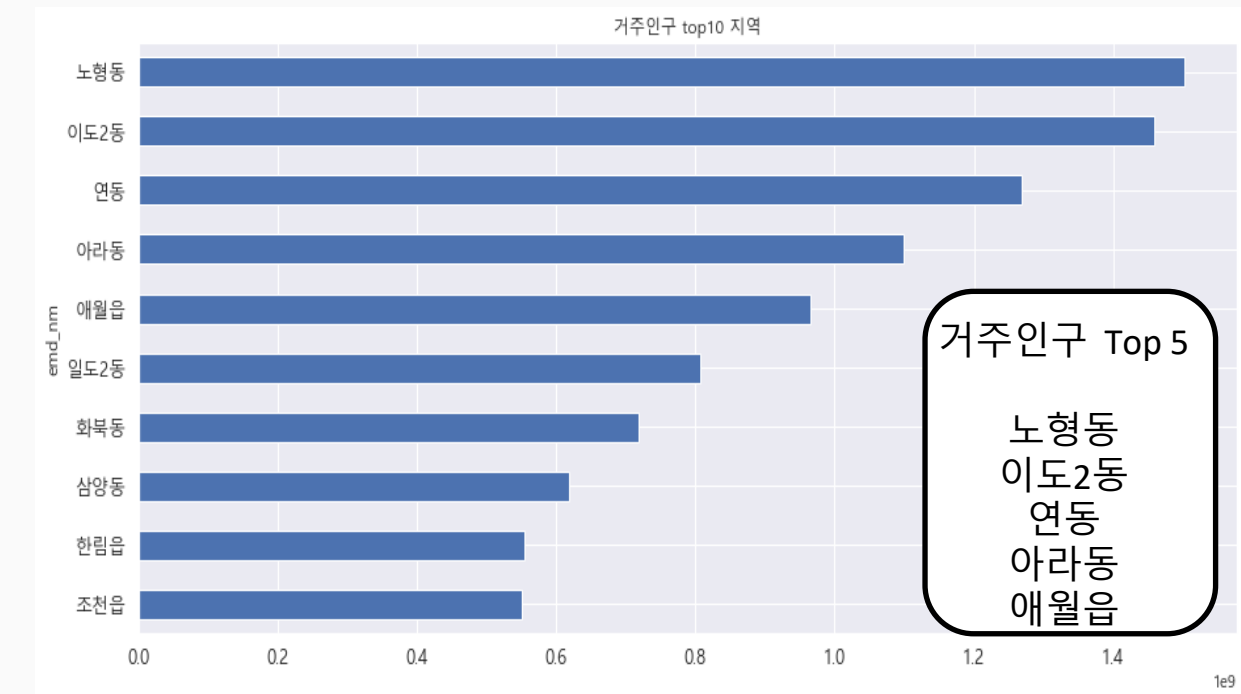
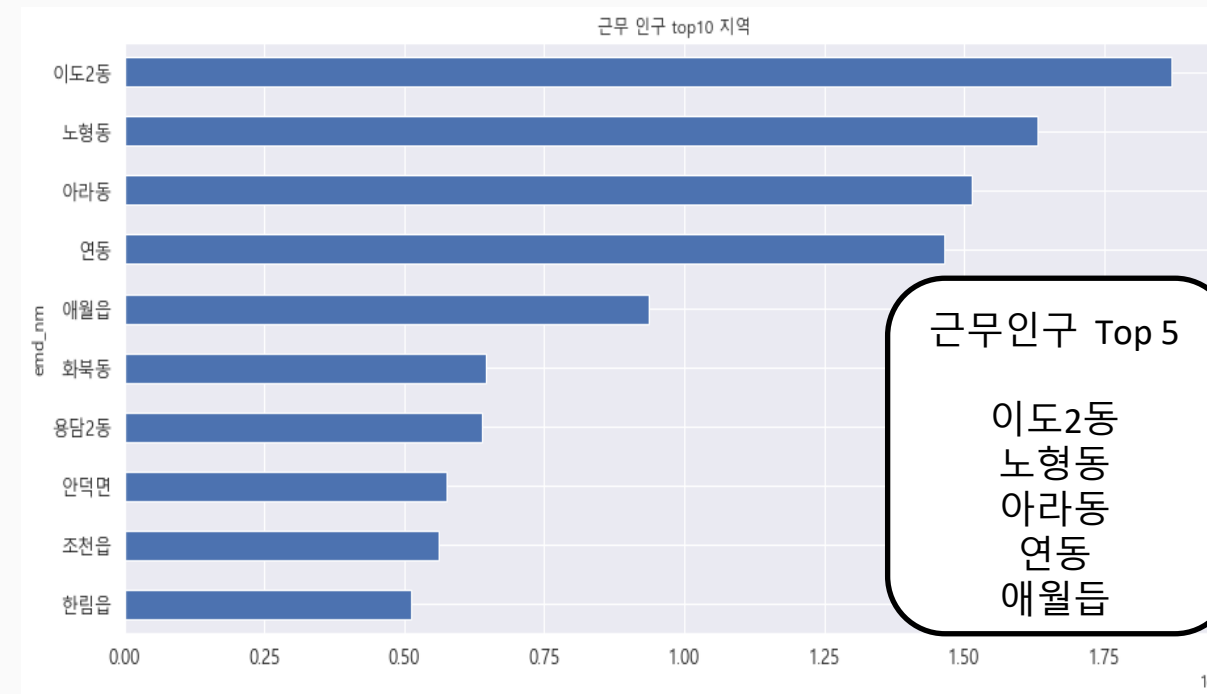
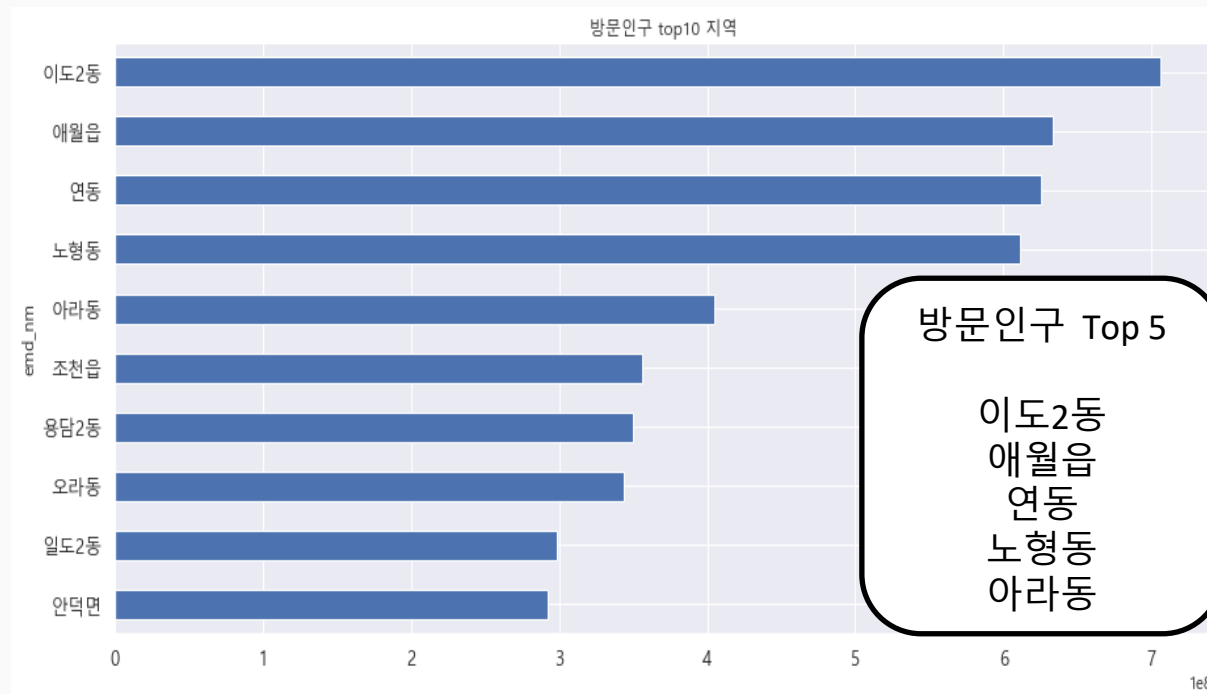
Feature Engineering

최종 Data Set

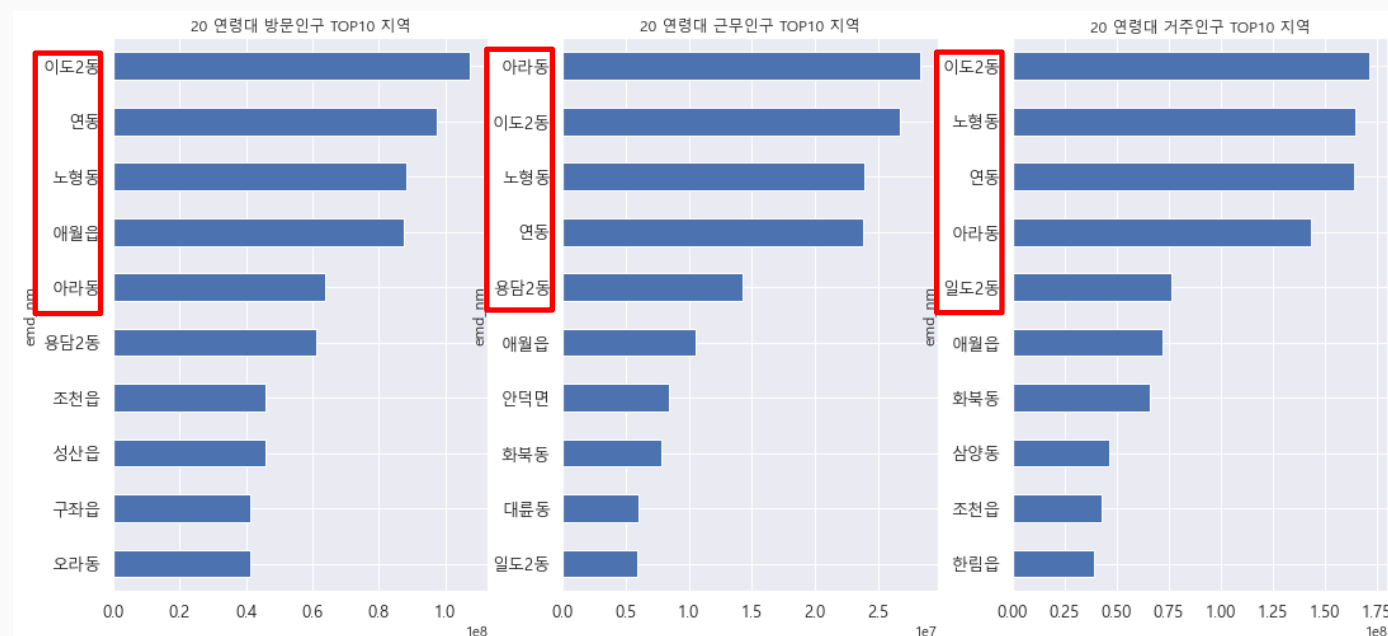
Test Set

“ 02-1_내국인유동인구_KOREAN.CSV ”

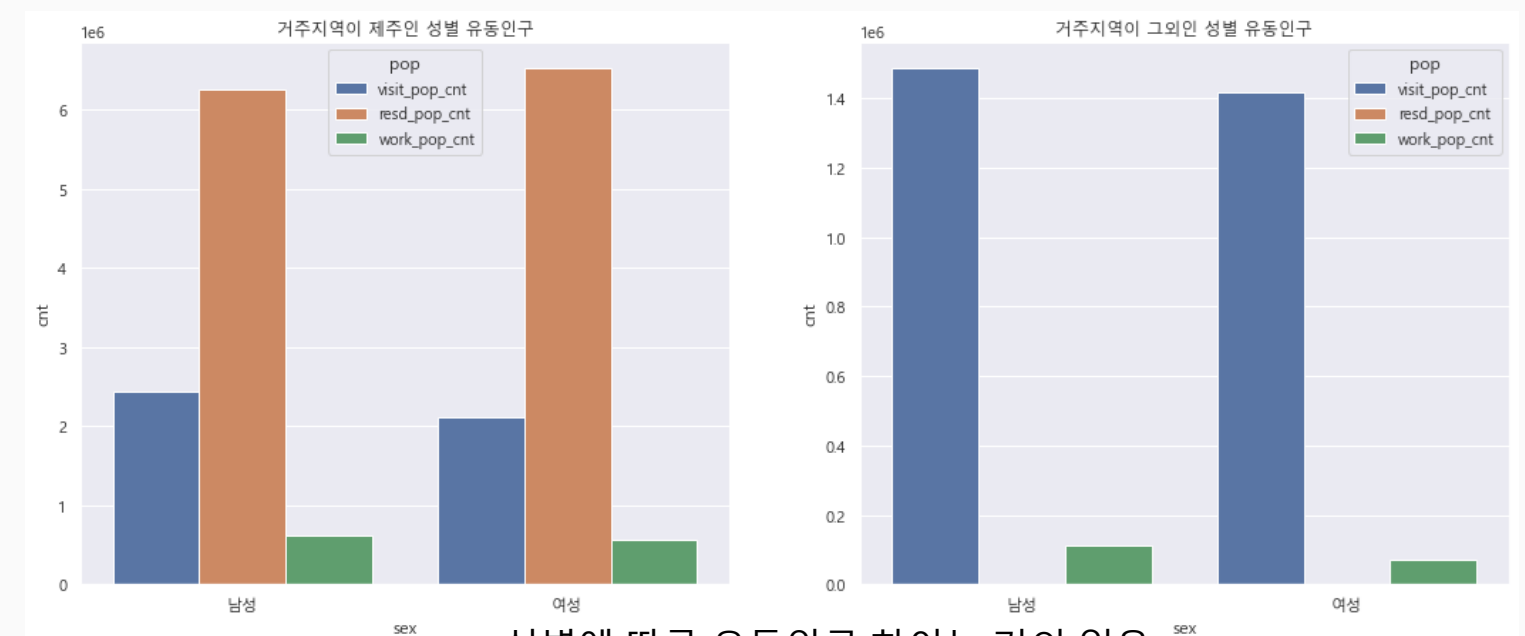
● 인구 유형별 Top 10



● 20대 인구 유형별 Top 10



● 성별에 따른 유동인구 차이



성별에 따른 유동인구 차이는 거의 없음

02.데이터 분석

“EDA”

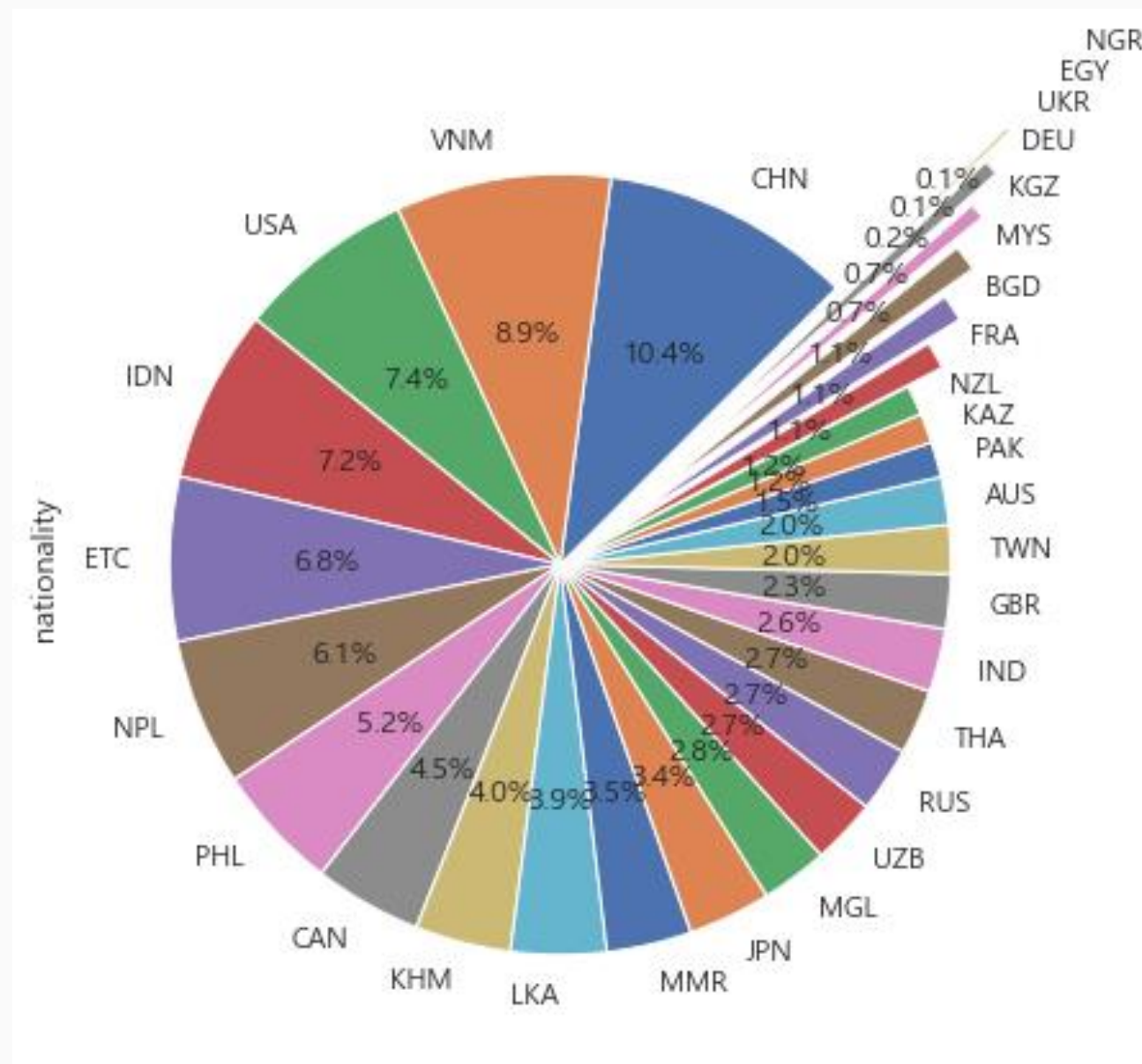
Feature Engineering

최종 Data Set

Test Set

“ 02-2_장기체류 외국인 유동인구_LONG_TERM_FRGN.CSV ”

<국적별 비중>



장기체류 외국인 유동인구 중
비중이 가장 많은 국적 Top 5

중국
베트남
미국
인도
기타

02.데이터 분석

“EDA”

Feature Engineering

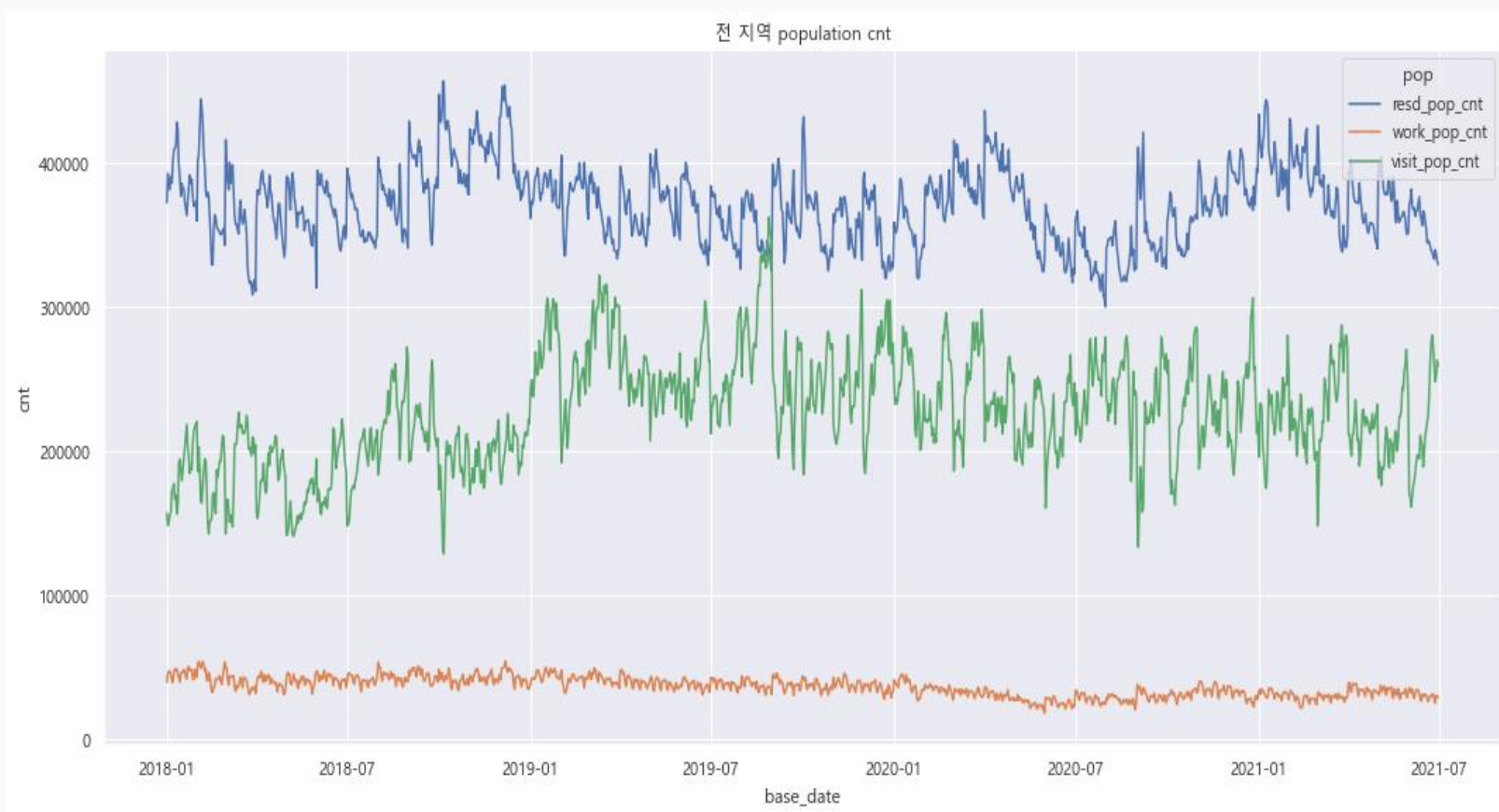
최종 Data Set

Test Set

“ 02-2_장기체류 외국인 유동인구_LONG_TERM_FRGN.CSV ”

<전체 기간 기준>

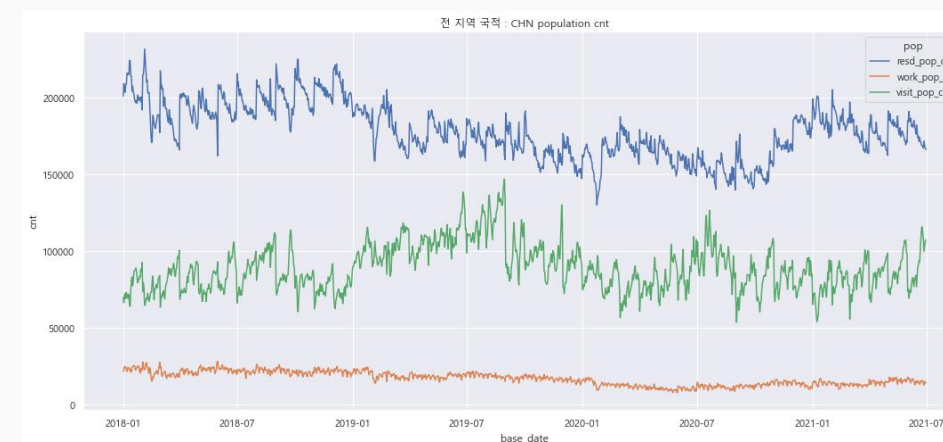
● 전체 지역 유형별 인구 추이



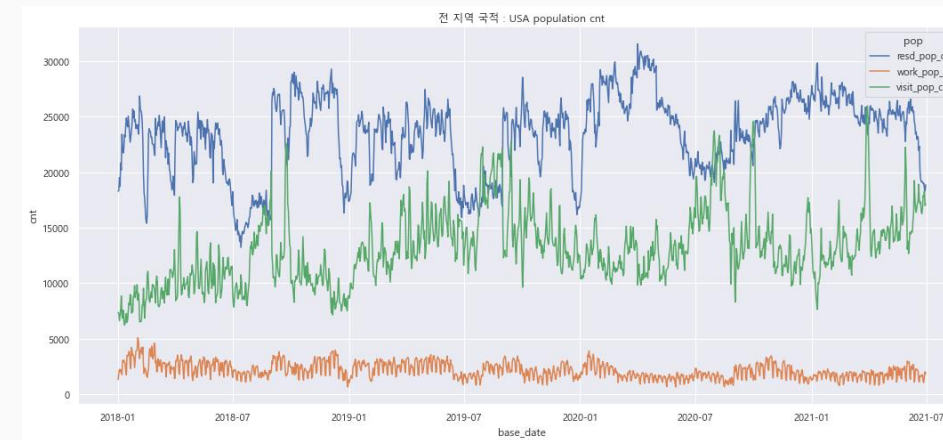
장기체류 외국인 유동인구의 제주도 전체의 유형별 인구 추이를 나타낸 그래프
내국인 유동인구와 비슷하게 인구 수가 거주인구, 방문인구, 근무인구 순으로 많음

● 전체 지역 국적별 인구 추이

CHN



USA



국적 비중을 가장 많이 차지하는
4개의 나라별로 인구 유형별 추이를
나타낸 그래프

베트남에서 방문인구가 변동폭이 크고
거주인구와 수가 비슷하거나 많음

VNM



ETC



02.데이터 분석

“EDA”

Feature Engineering

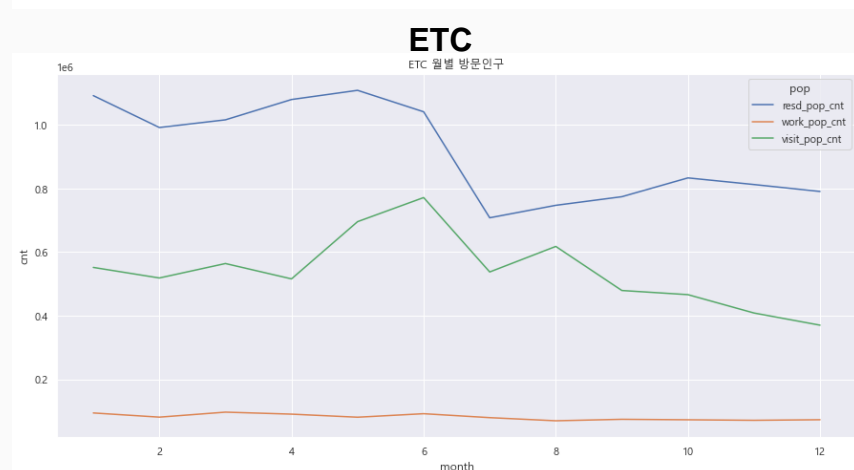
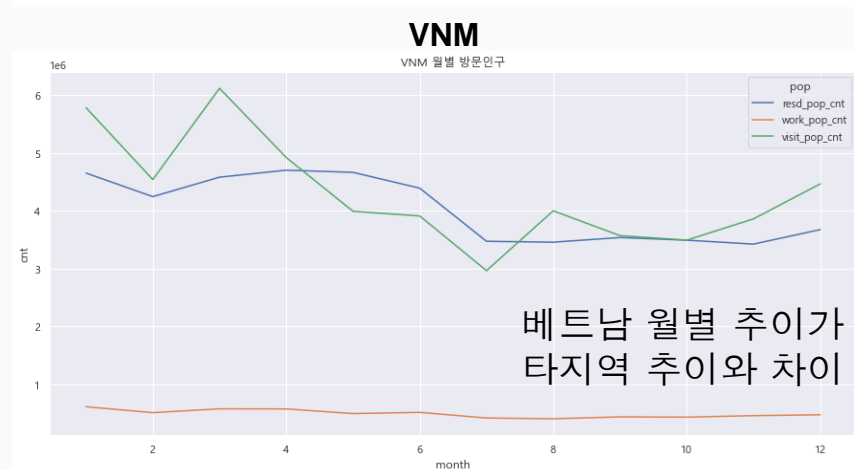
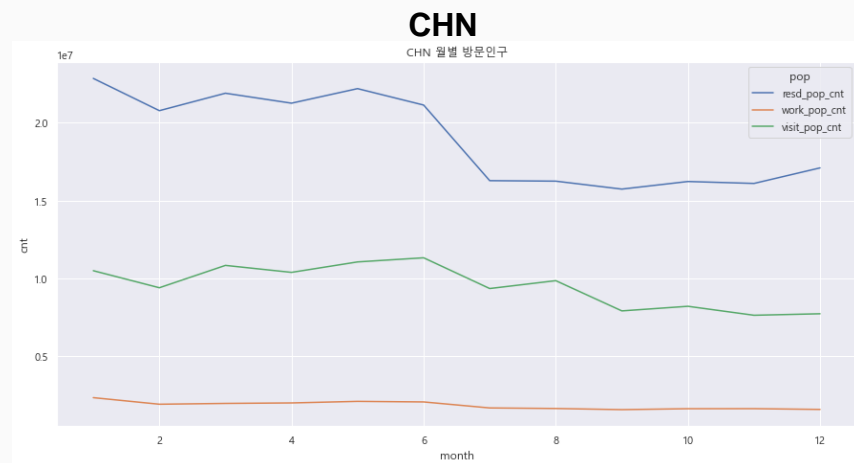
최종 Data Set

Test Set

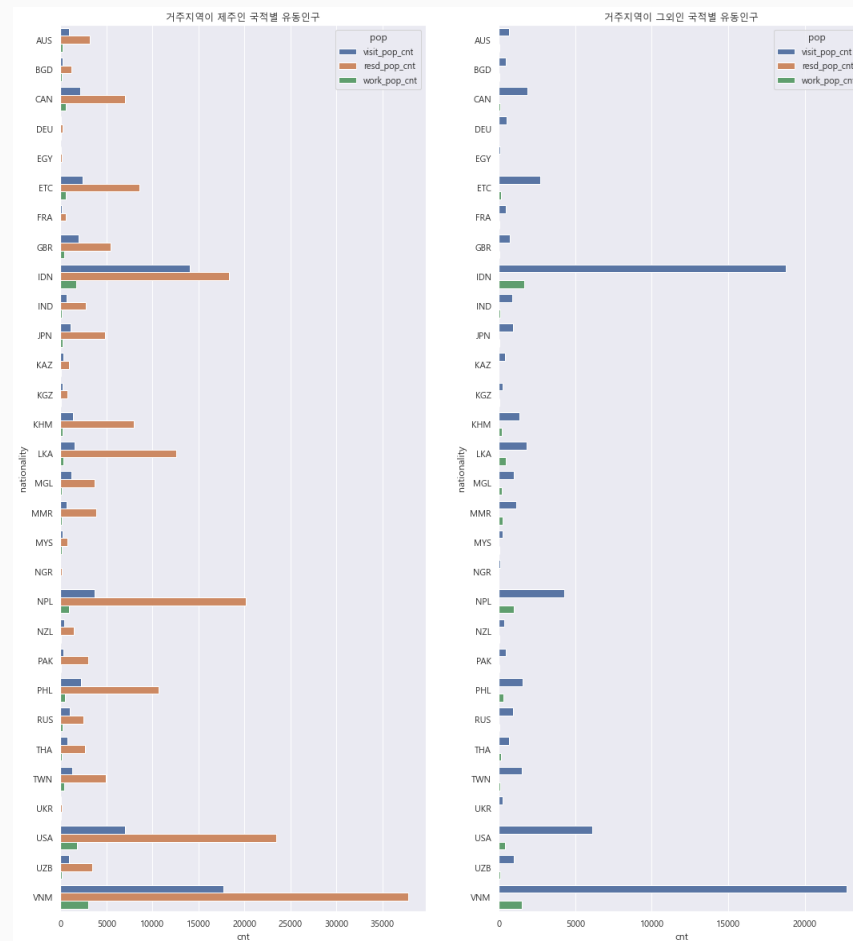
“ 02-2_장기체류 외국인 유동인구_LONG_TERM_FRGN.CSV ”

<시간 기준>

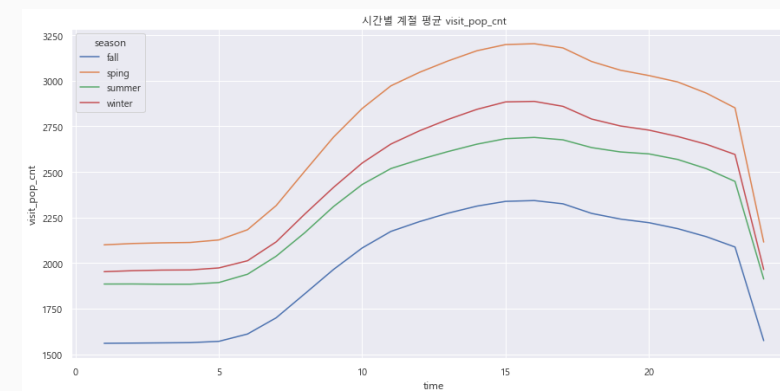
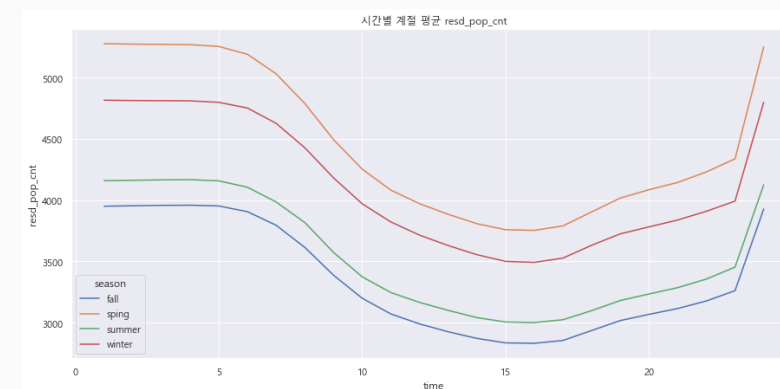
● 월별



● 일평균



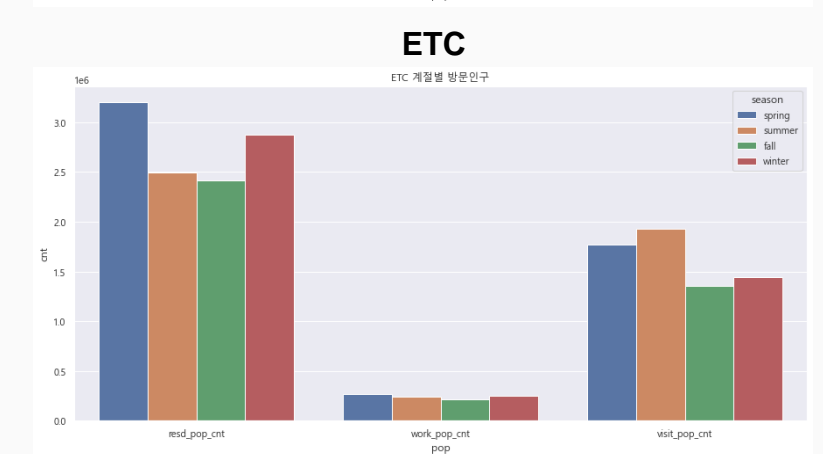
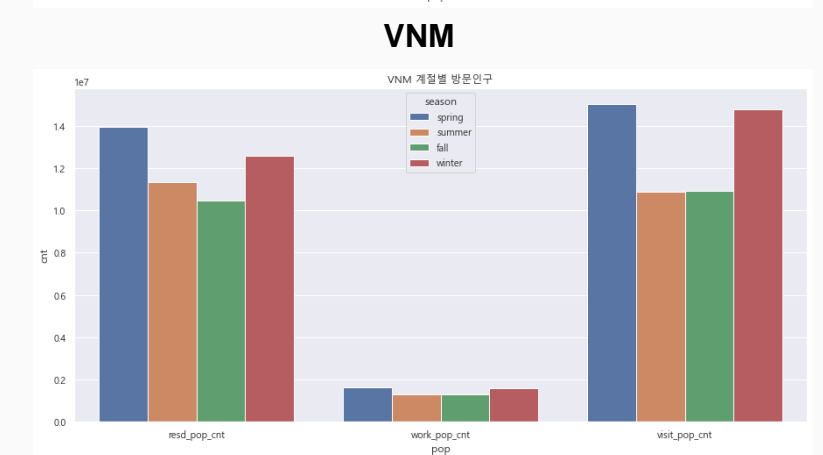
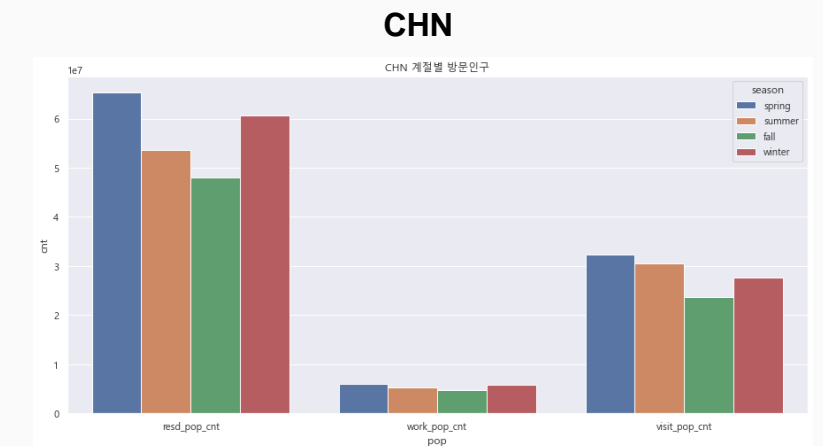
● 시간별



거주인구와 방문인구의 시간대별 추이가
거의 반대로 나타남

거주인구 : 10시 ~ 20시 낮음
방문인구 : 10시 ~20시 높음

● 계절별



02.데이터 분석

“EDA”

Feature Engineering

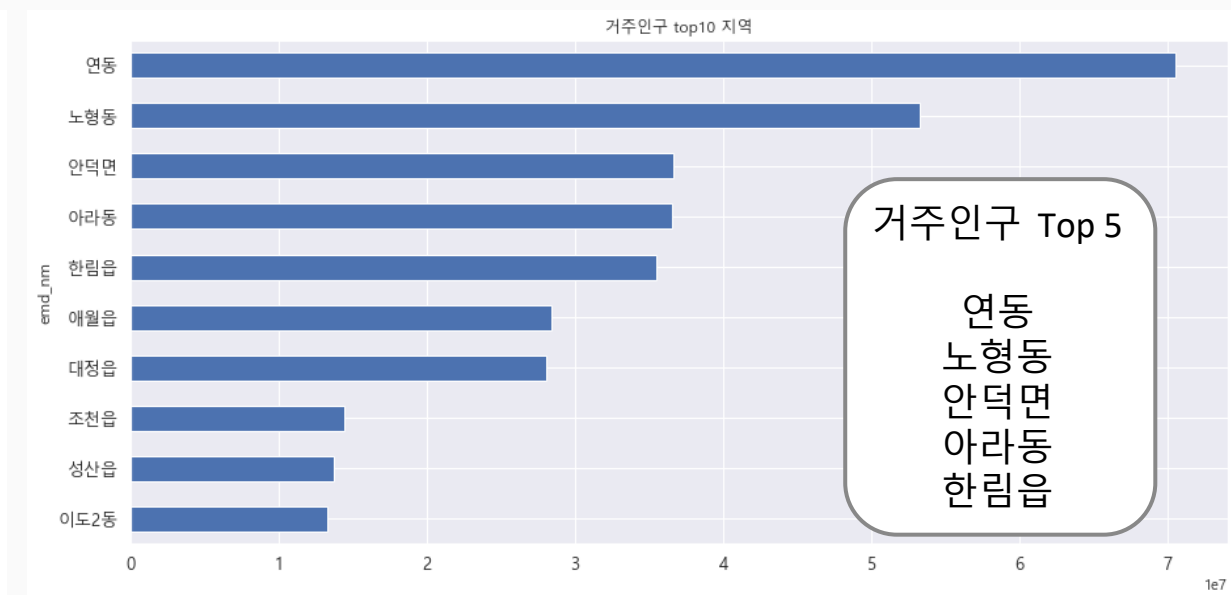
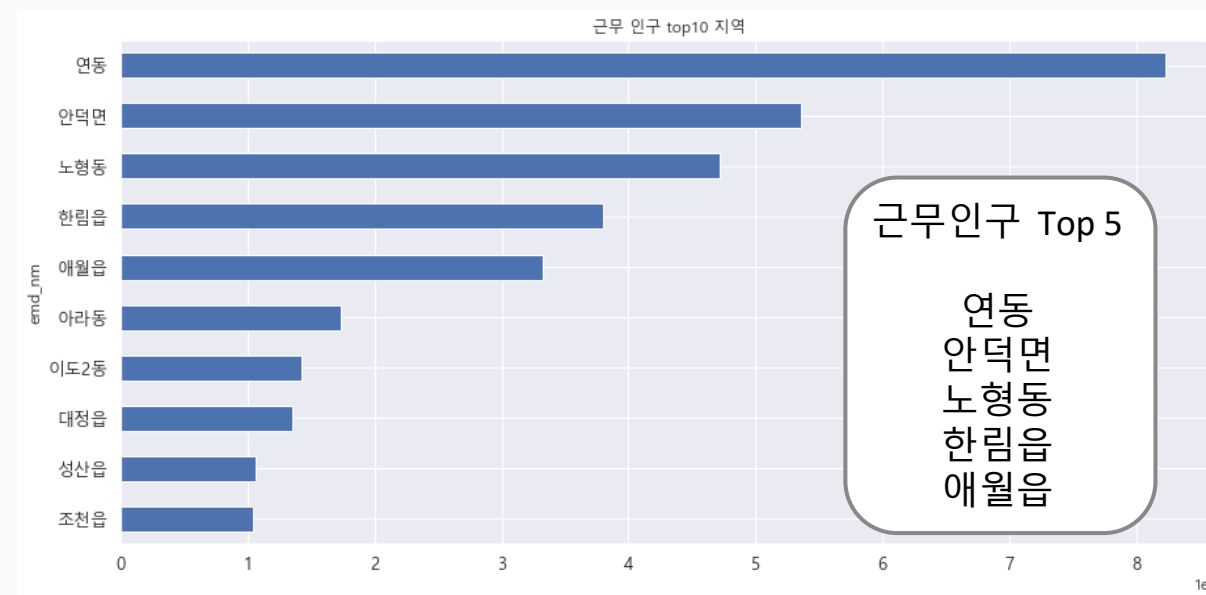
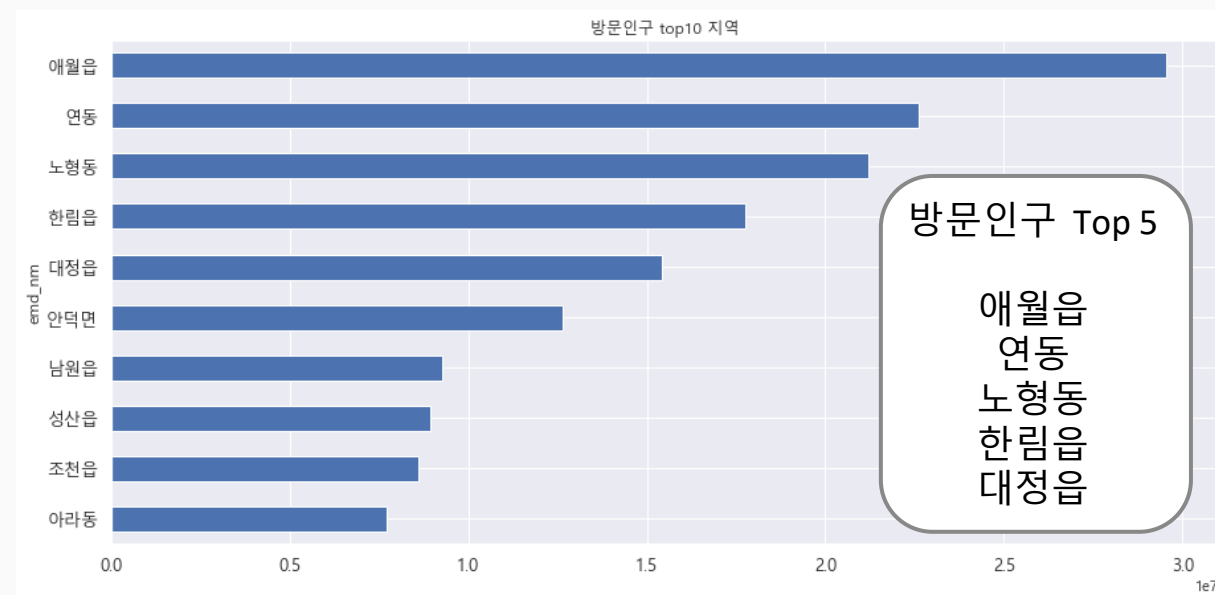
최종 Data Set

Test Set

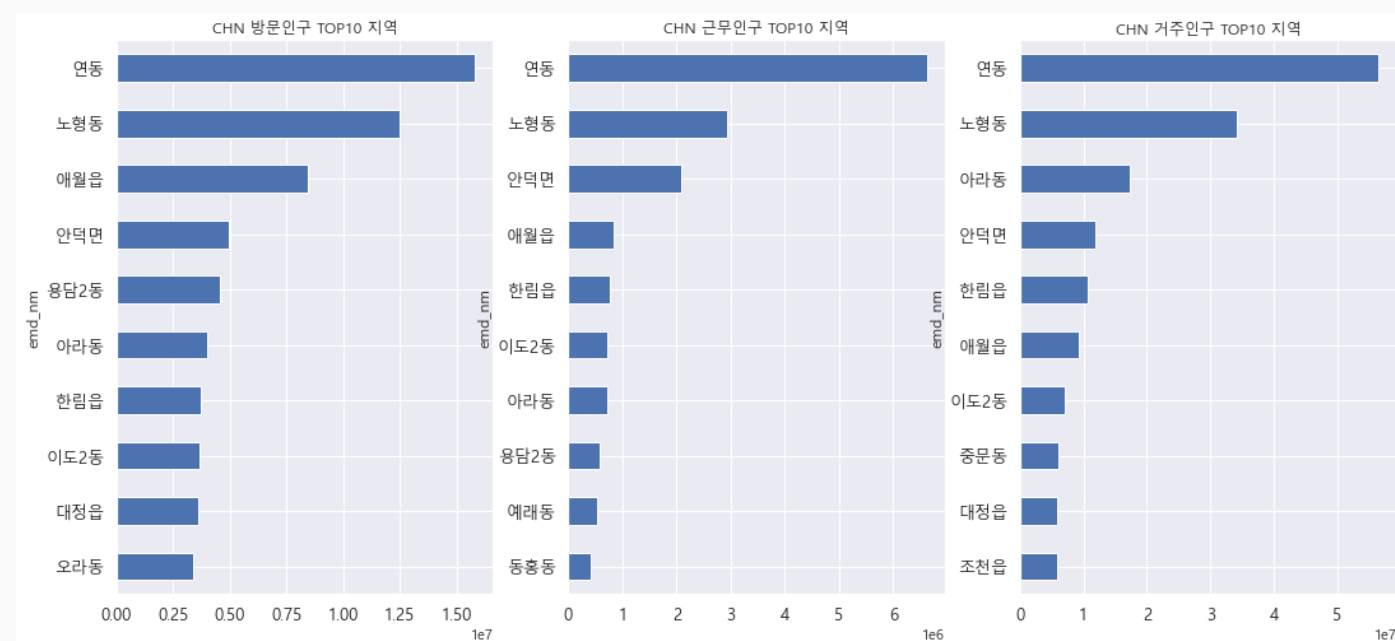
“ 02-2_장기체류 외국인 유동인구_LONG_TERM_FRGN.CSV ”

<Top 10 기준>

● 인구 유형별 Top 10



● 중국의 유형별 Top 10



방문인구 Top 5

연동
노형동
애월읍
안덕면
용담2동

근무인구 Top 5

연동
노형동
안덕면
애월읍
한림읍

거주인구 Top 5

연동
노형동
아라동
안덕면
한림읍

02.데이터 분석

“EDA”

Feature Engineering

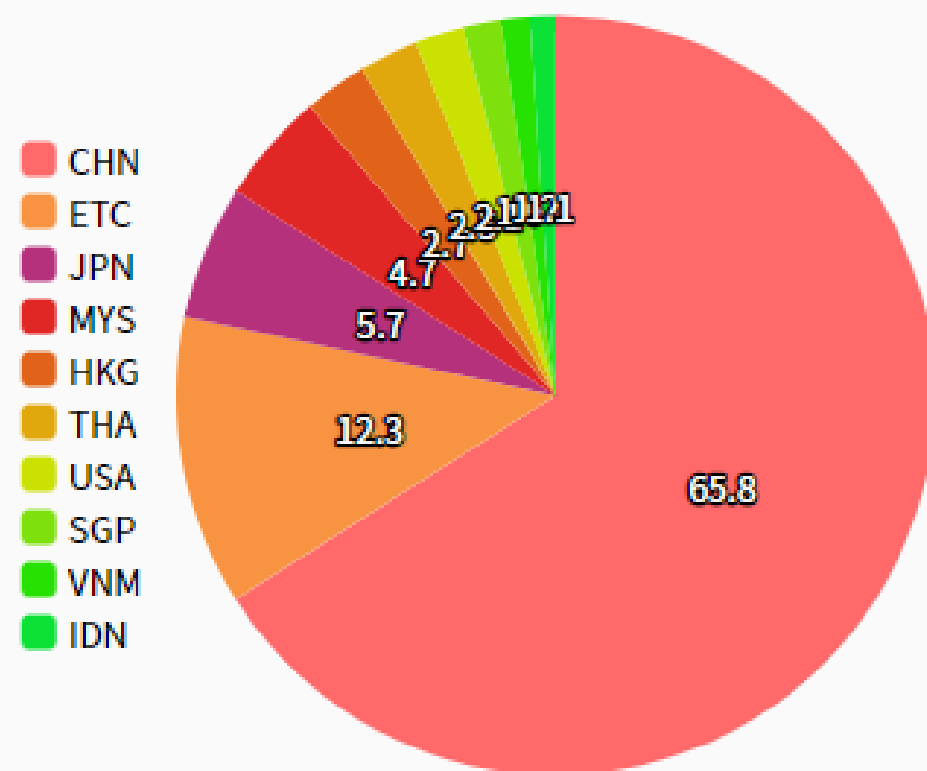
최종 Data Set

Test Set

“ 02-3_단기체류 외국인 유동인구_SHORT_TERM_FRGN.CSV ”

<국가별 방문 비율>

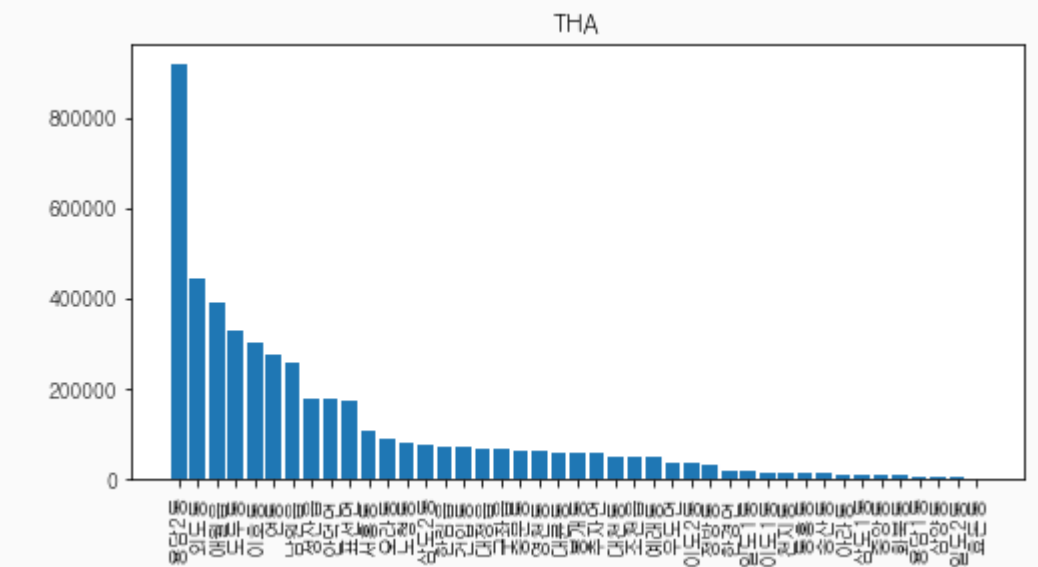
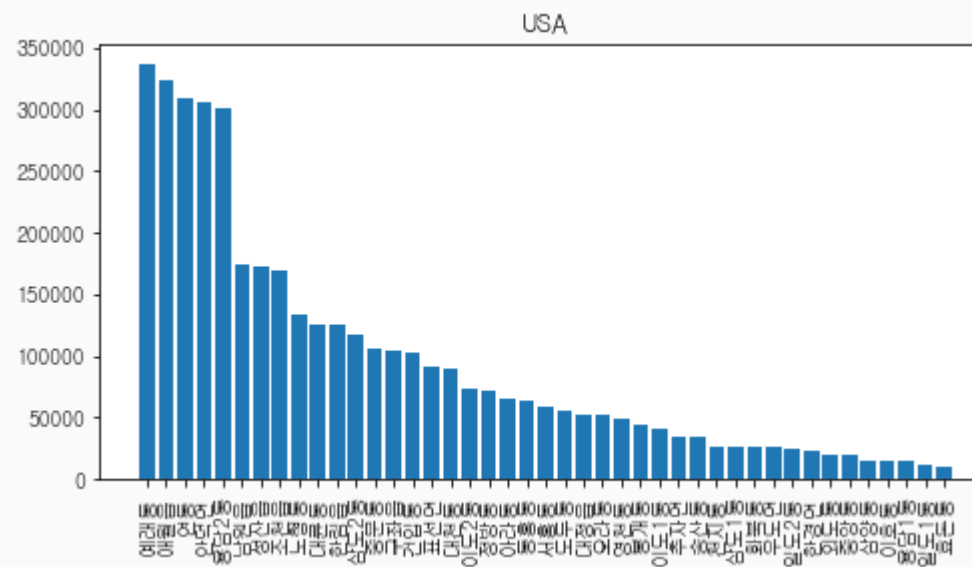
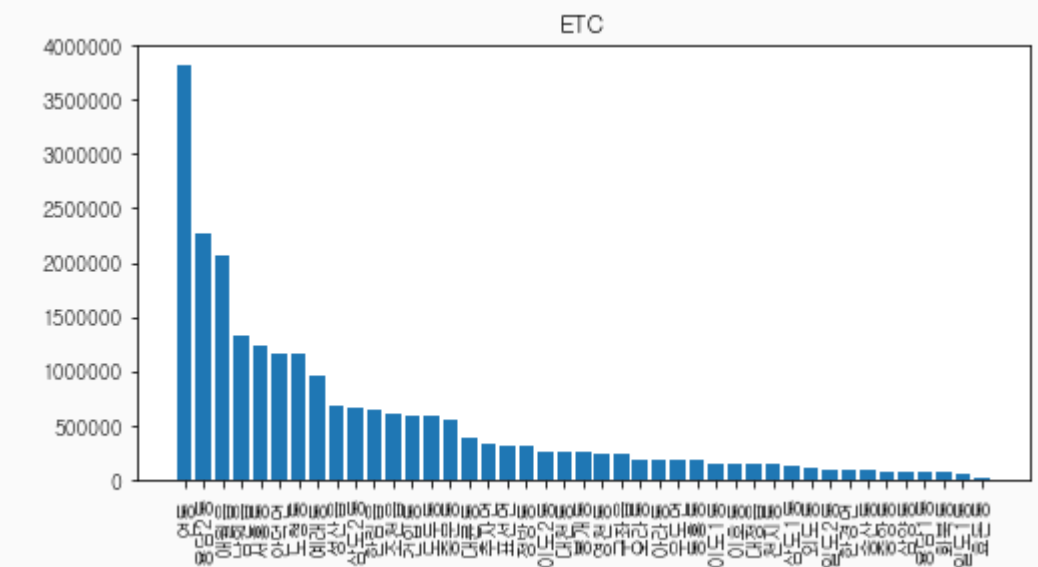
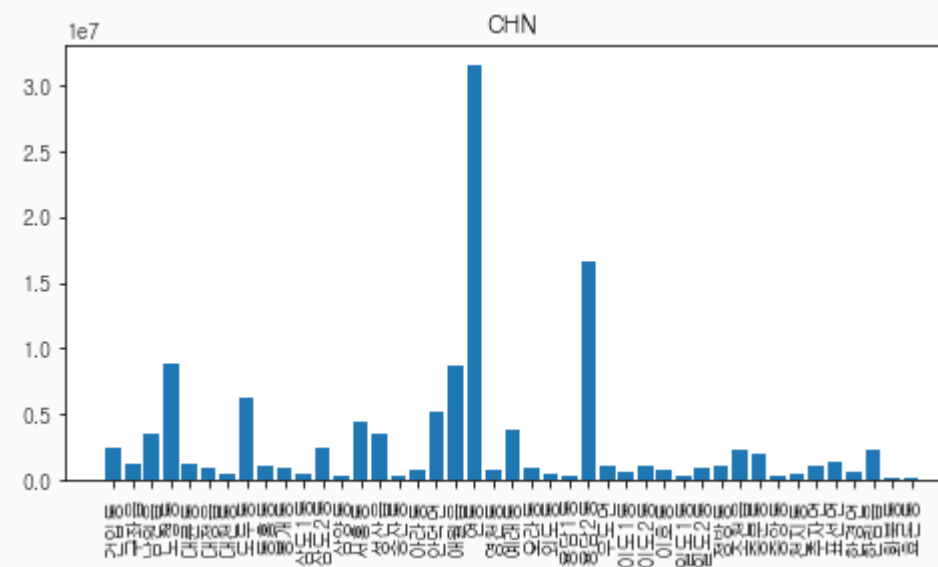
● 국가별 방문 비율



단기체류 외국인 유동인구 중
비중을 가장 많이 차지하는 국적 TOP 5

중국
기타
일본
말레이시아
홍콩

● 지역별 방문 비율



02.데이터 분석

“EDA”

Feature Engineering

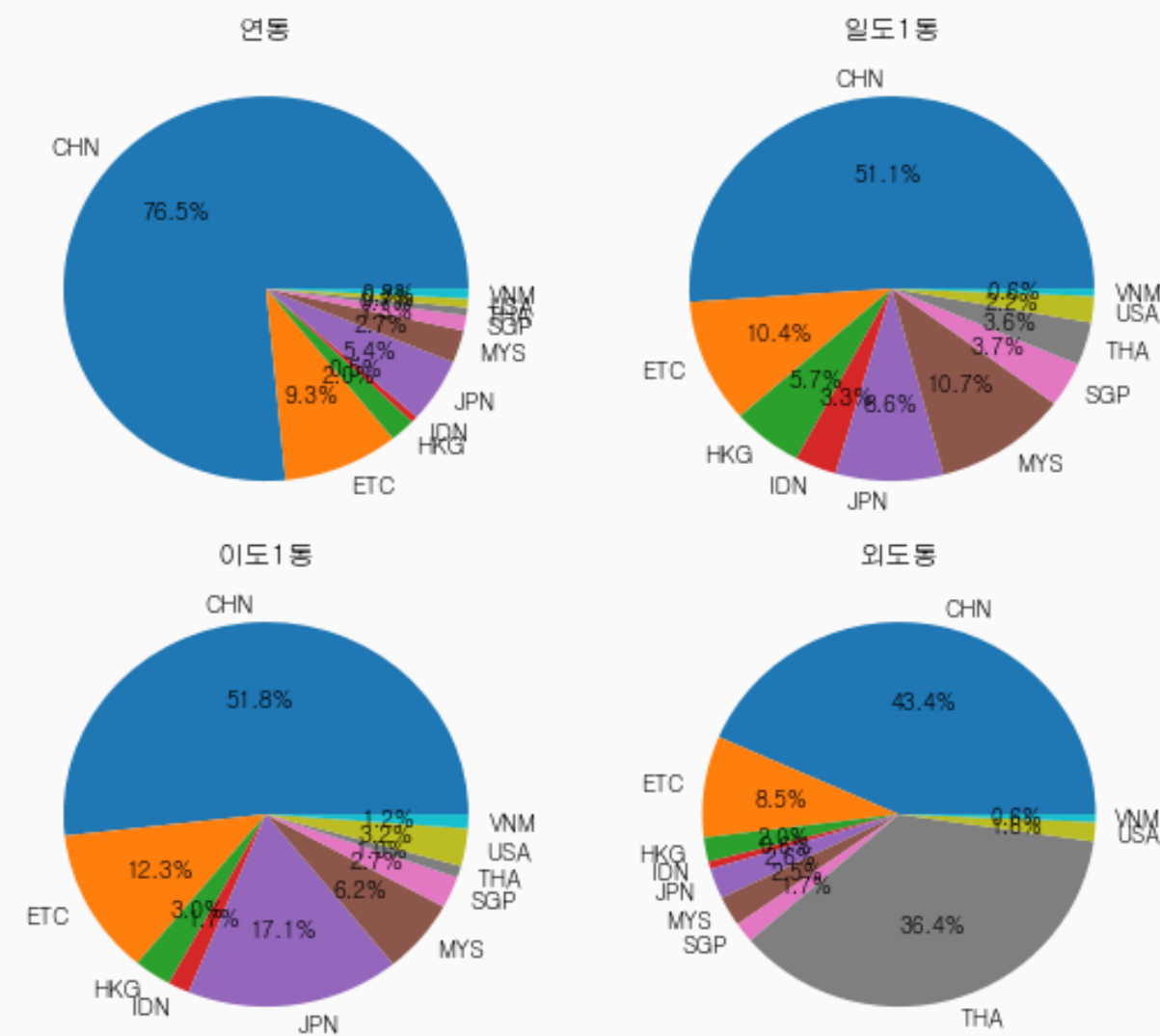
최종 Data Set

Test Set

“ 02-3_단기체류 외국인 유동인구_SHORT_TERM_FRGN.CSV ”

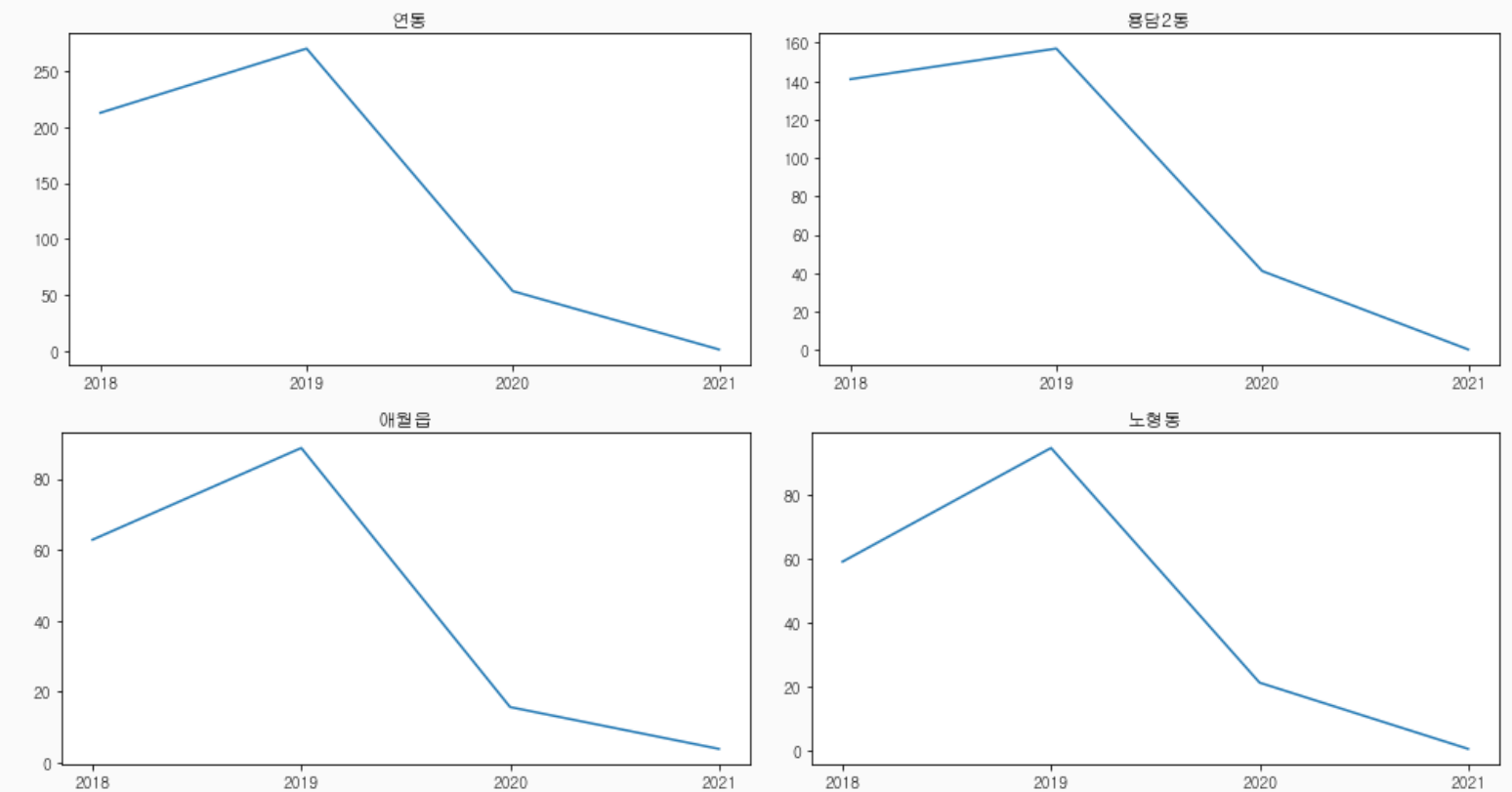
<지역별 방문 비율>

● 외국인 방문 비율

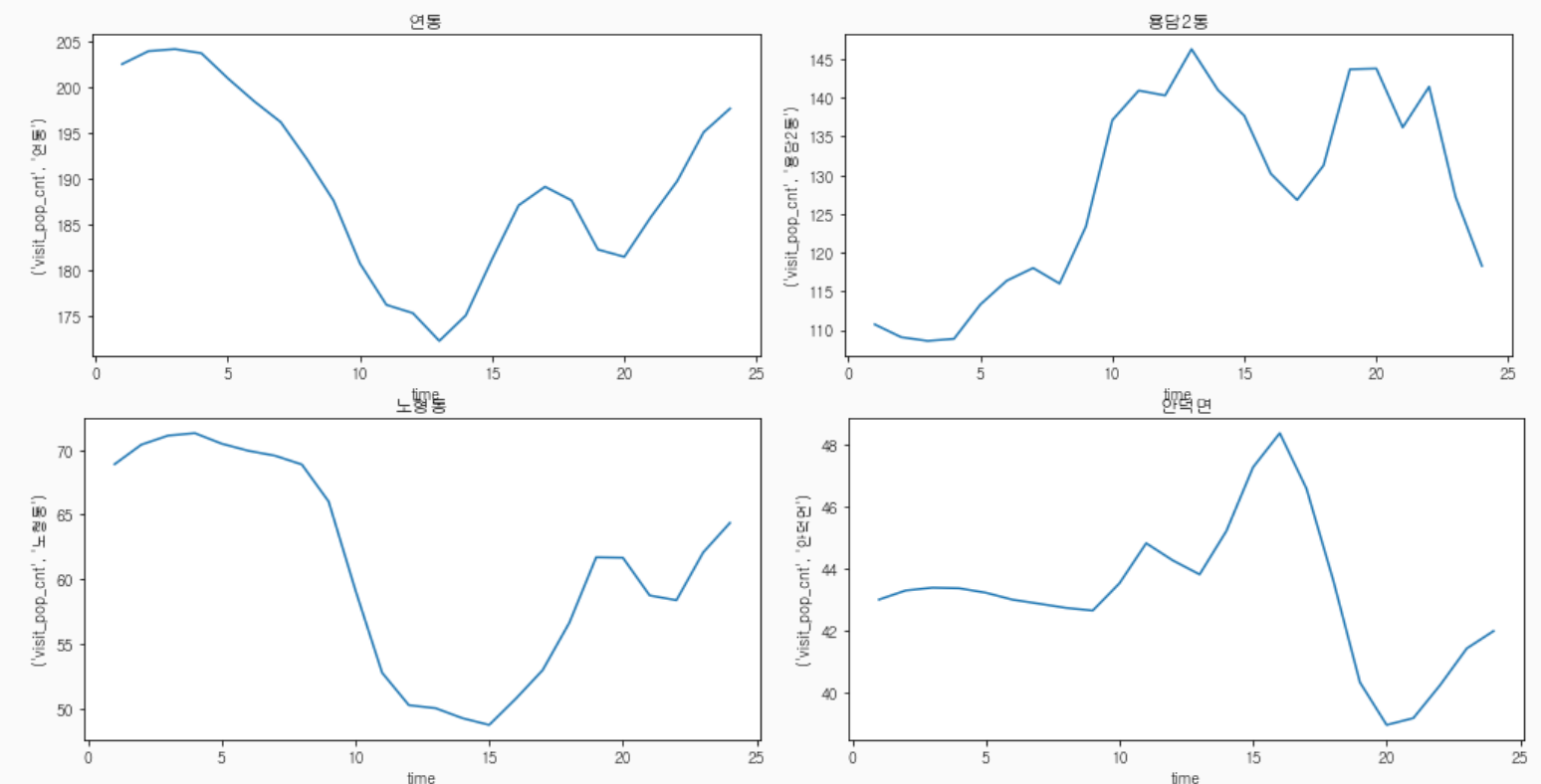


단기체류 외국인의 국적별 방문인구 비율 결과 전체적으로 중국이 가장 높고
이도 1동에는 일본, 외도동에는 태국이 비교적 높음

● 연도별 외국인 평균



● 시간대별 외국인 평균



02.데이터 분석

“EDA”

Feature Engineering

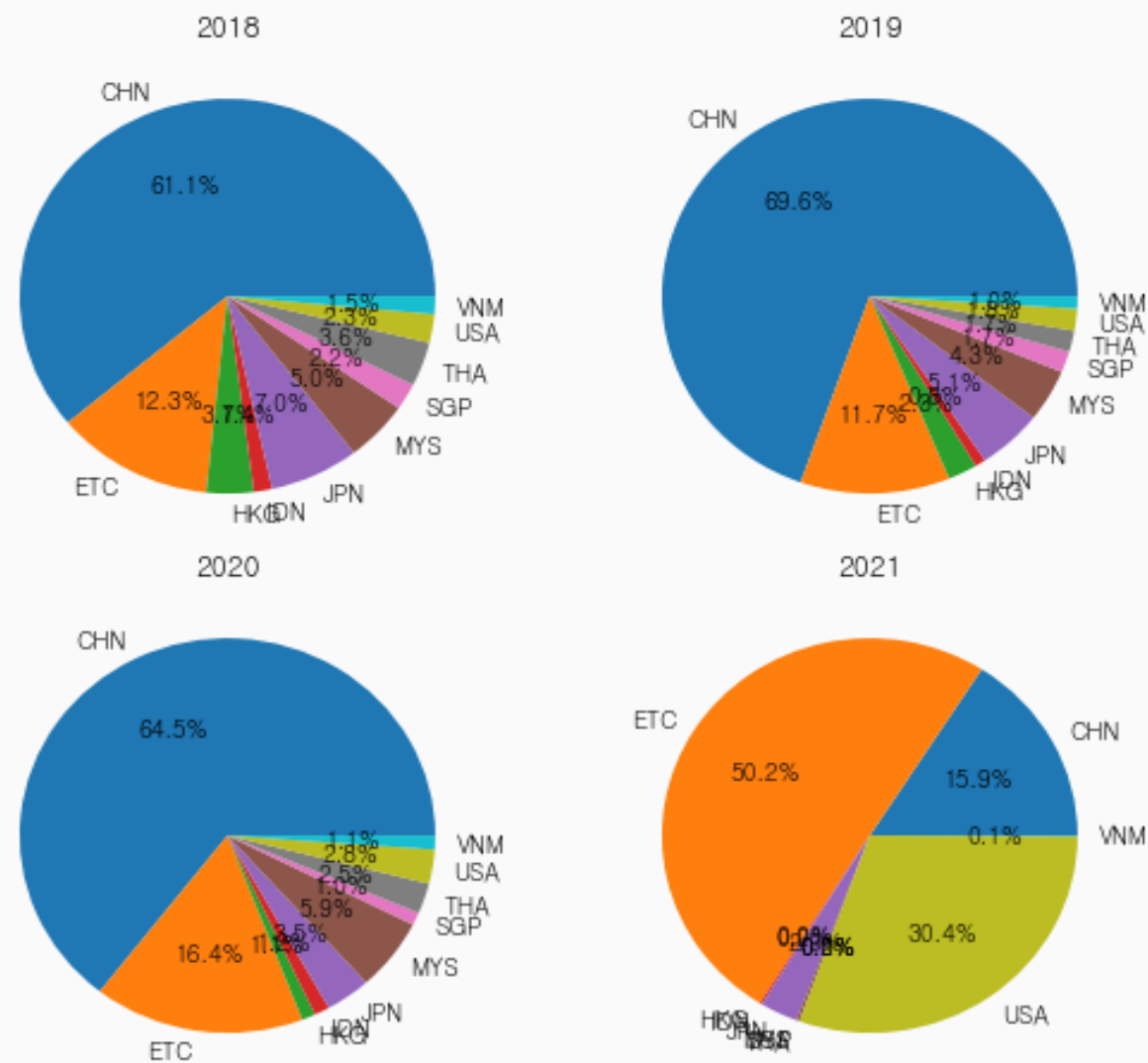
최종 Data Set

Test Set

“ 02-3_단기체류 외국인 유동인구_SHORT_TERM_FRGN.CSV ”

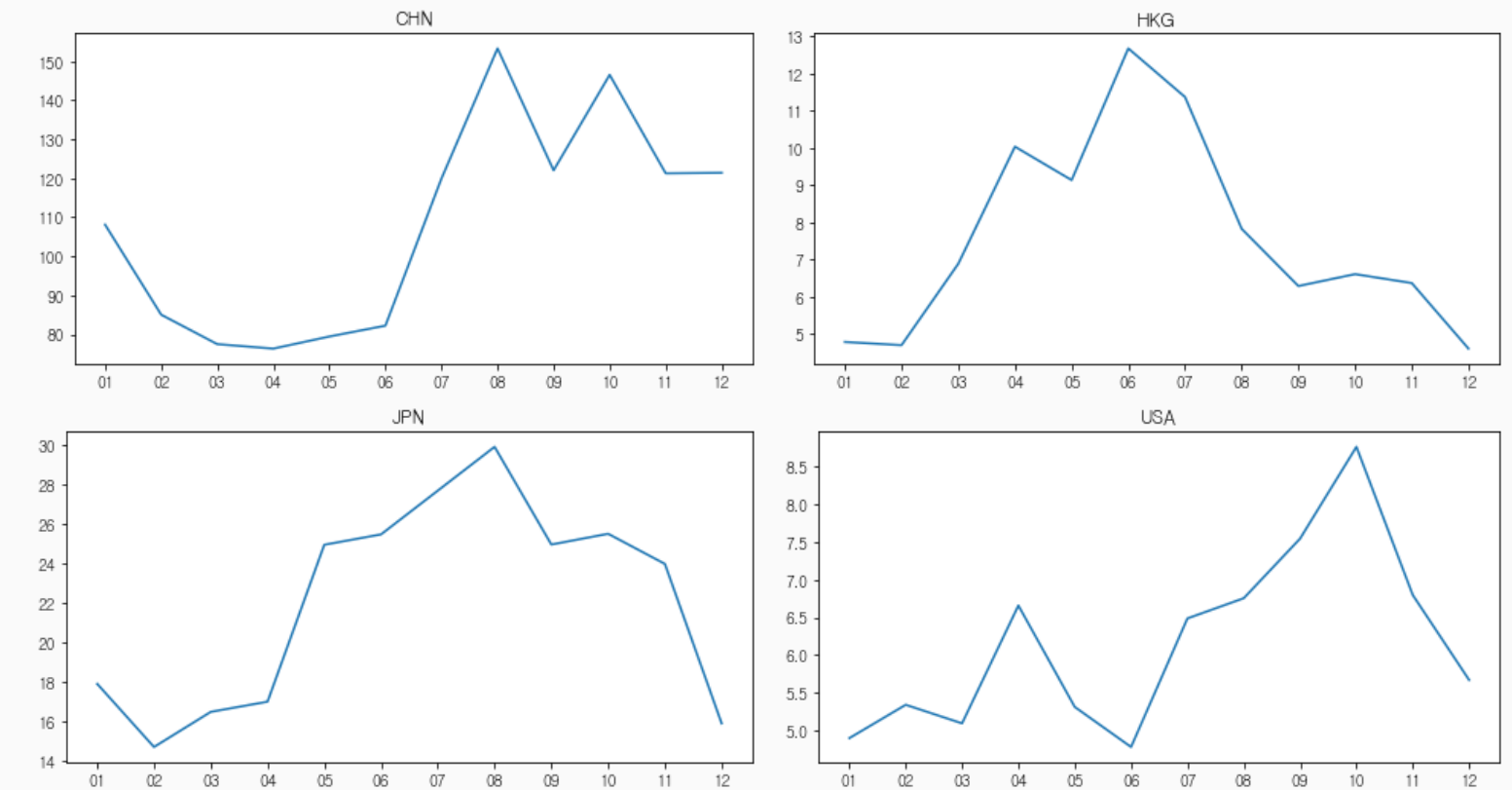
<시간대별 방문 비율>

● 연도별 외국인 방문 비율

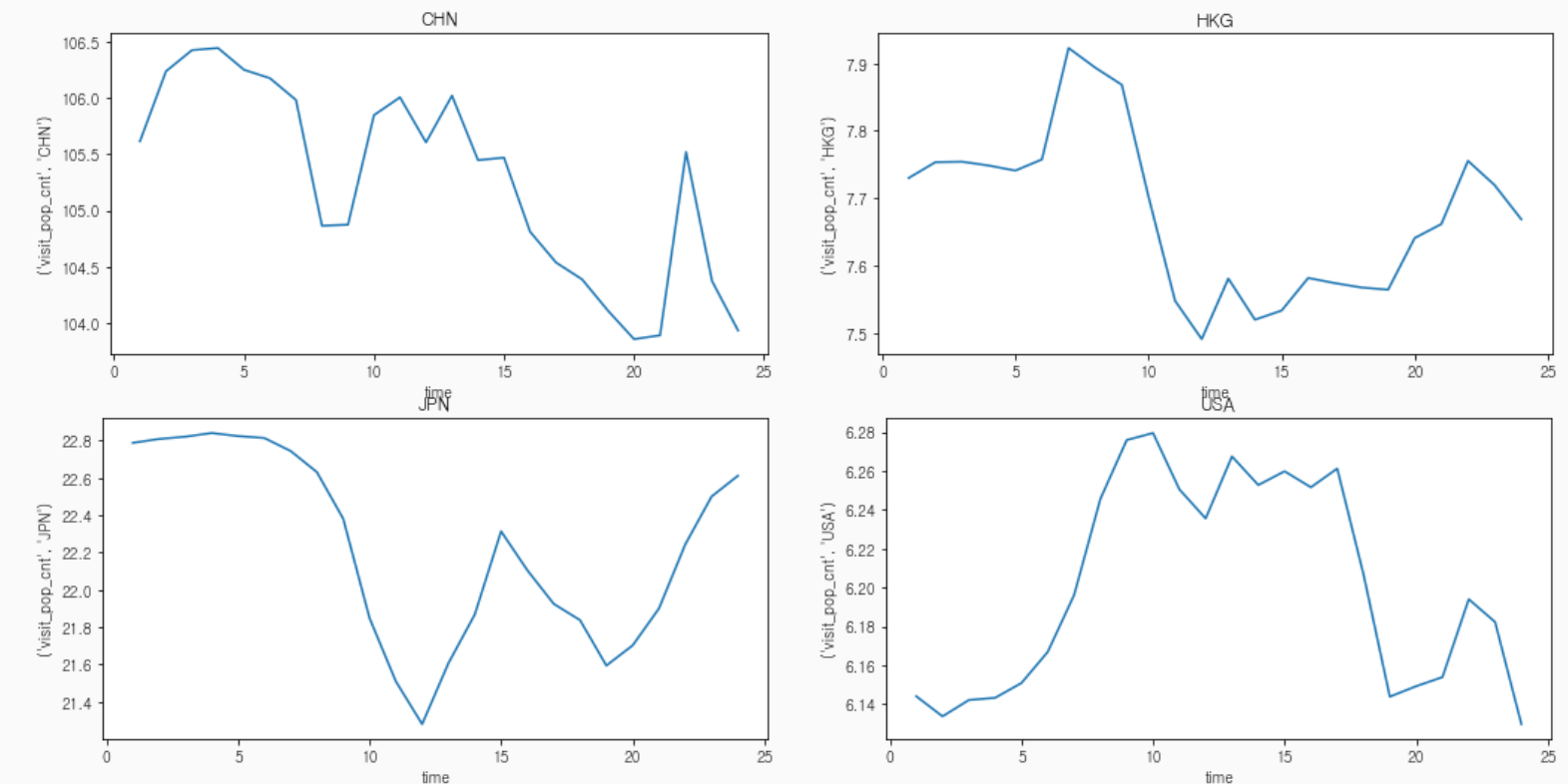


단기체류 외국인의 연도별 방문인구 비율 결과
전체적으로 중국, 일본 순으로 비율이 높음

● 월별 외국인 평균



● 시간대별 외국인 평균



02.데이터 분석

“EDA”

Feature Engineering

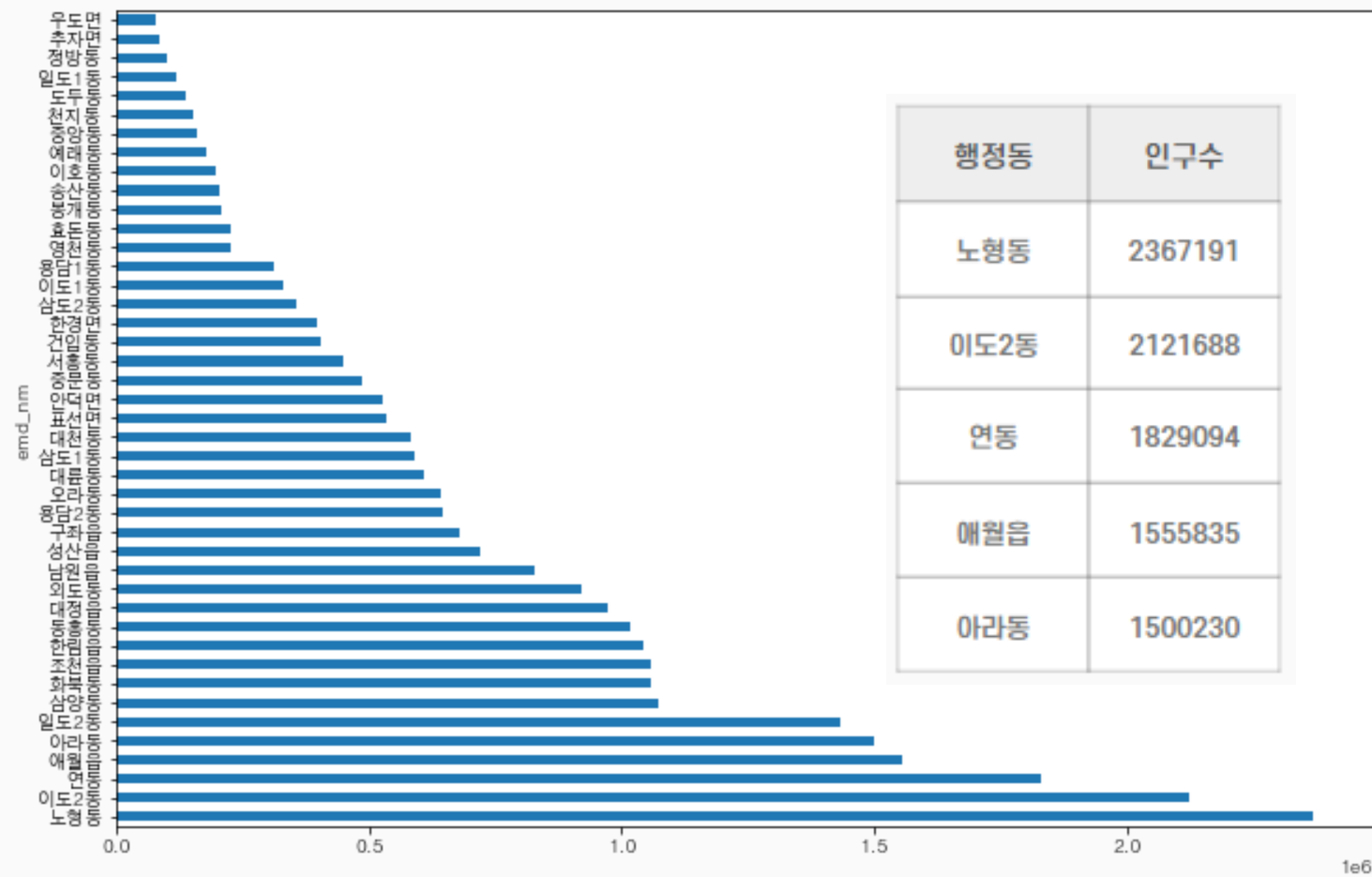
최종 Data Set

Test Set

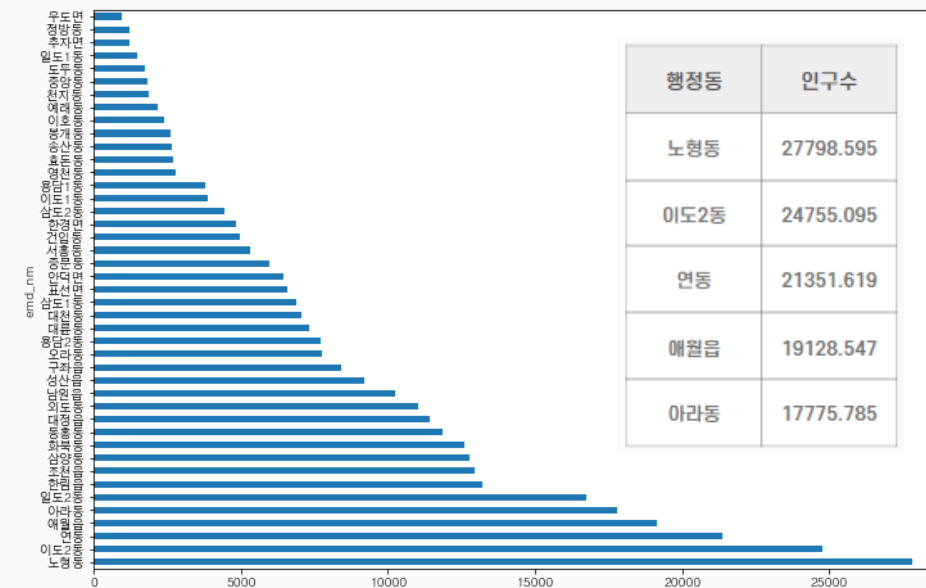
“ 03_거주인구_RESIDENT_POP.CSV ”

<읍면동별 인구특성>

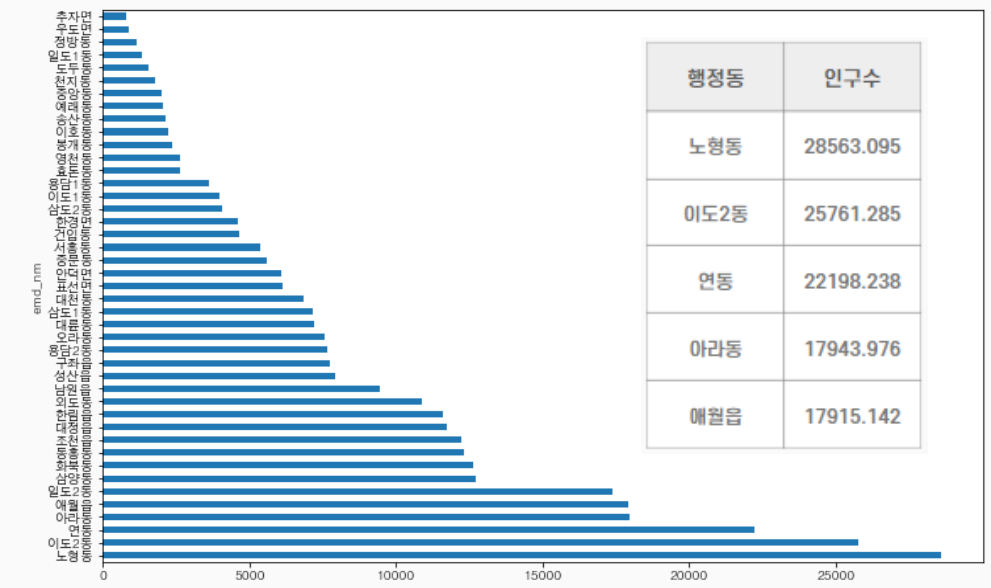
● 읍면동별 전체 인구



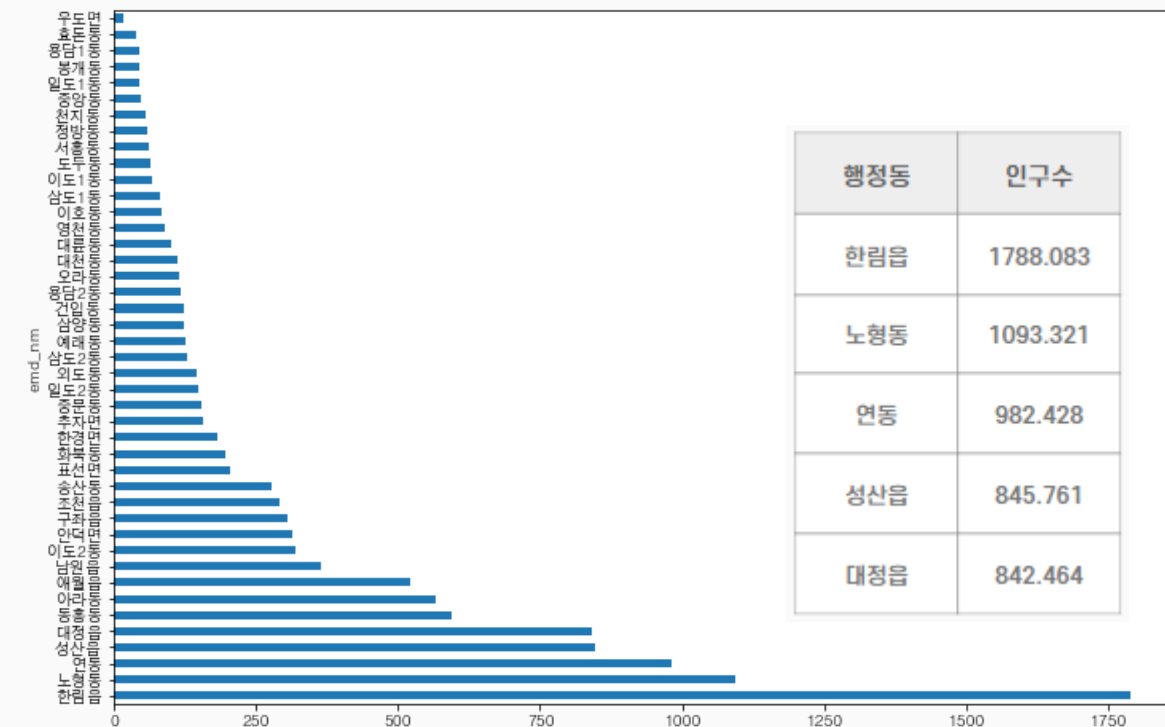
● 남성



● 여성



● 외국인



02.데이터 분석

“EDA”

Feature Engineering

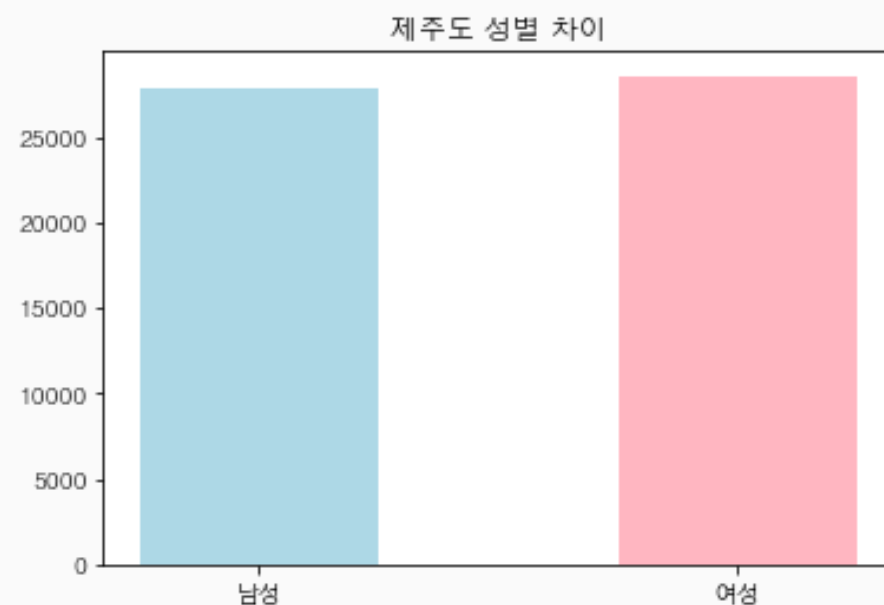
최종 Data Set

Test Set

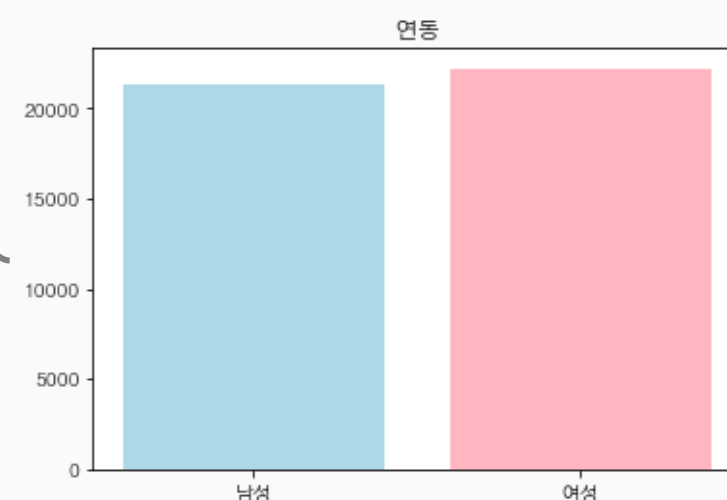
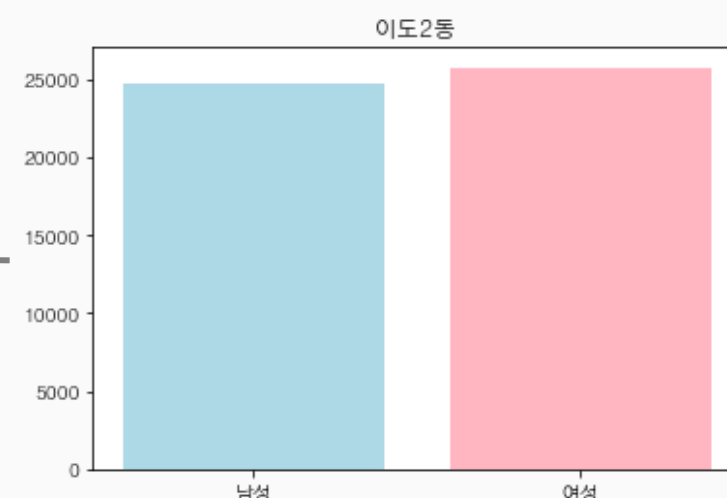
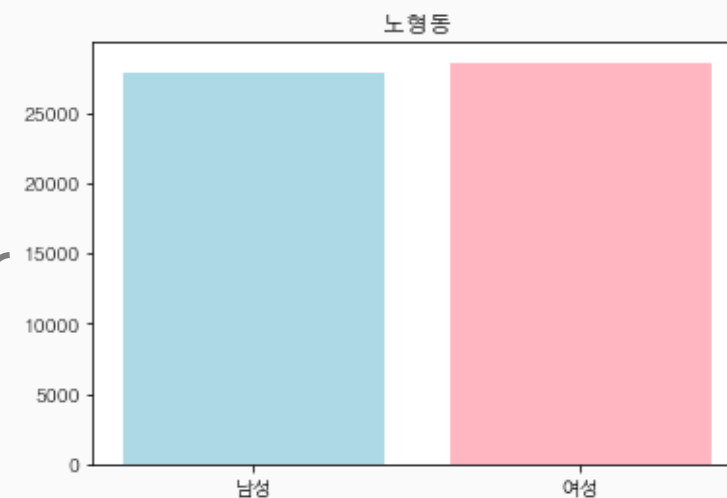
“ 03_거주민구_RESIDENT_POP.CSV ”

<인구특성별 차이>

● 성별에 따른 차이

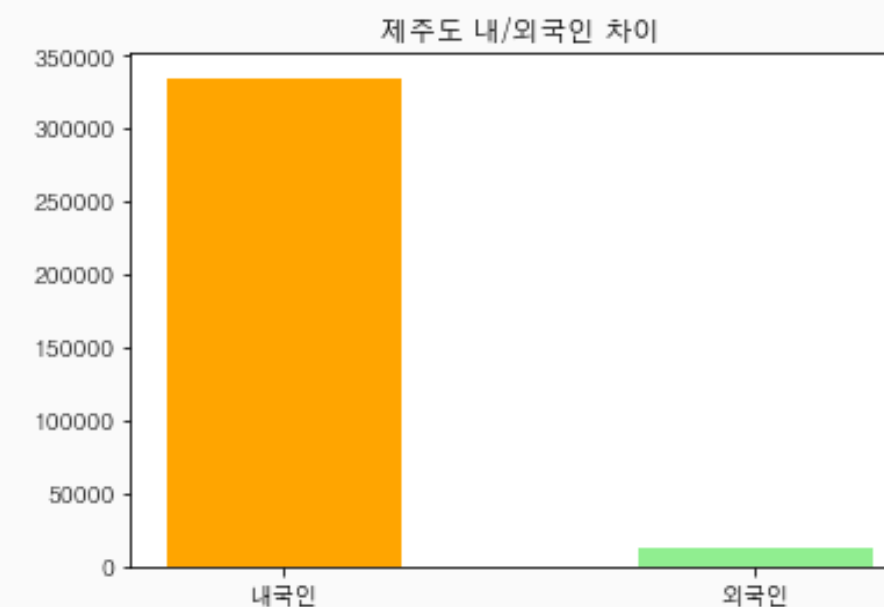


제주도에 거주하는 성별에 따른 인구 수 차이는 거의 없음

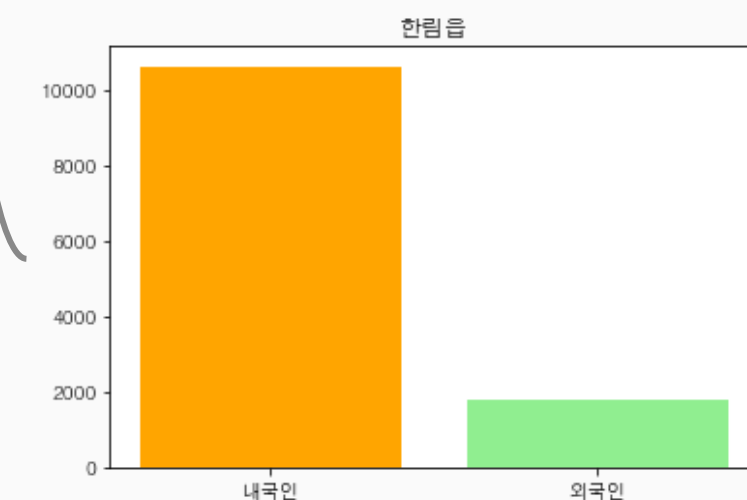
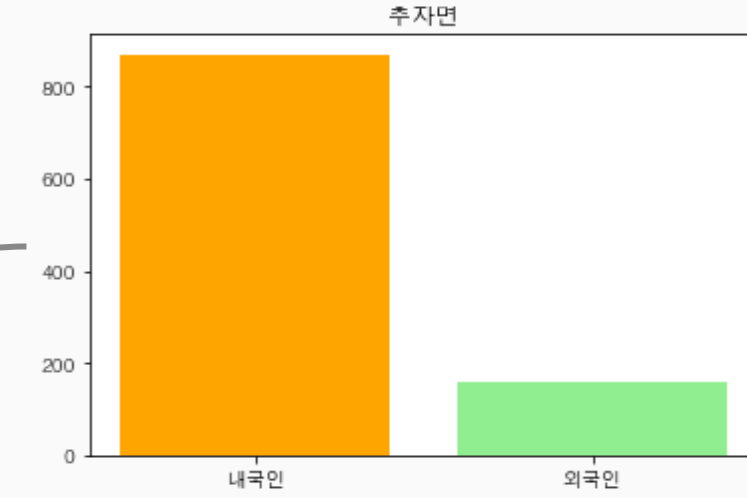
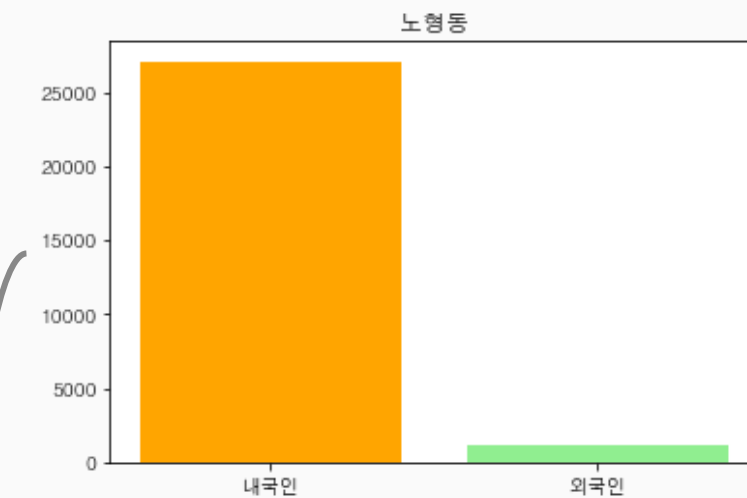


읍면동별로 거주하는 성별에 따른 인구 수 차이는 거의 없음

● 내/외국인에 따른 차이



제주도 전체 거주하는 내/외국인 차이와 읍면동별 거주하는 내/외국인 차이 양상이 비슷함



02.데이터 분석

“EDA”

Feature Engineering

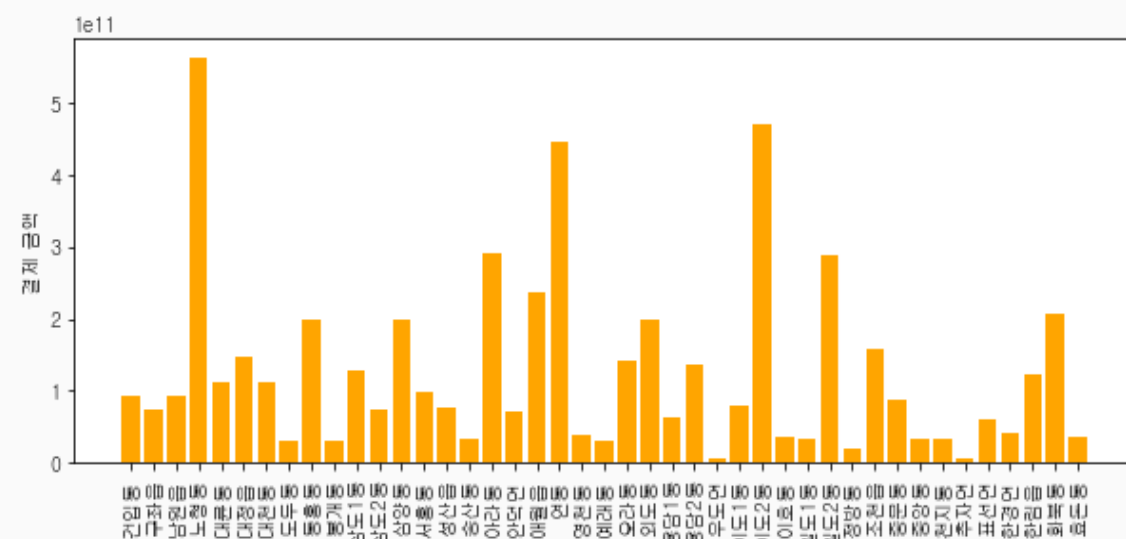
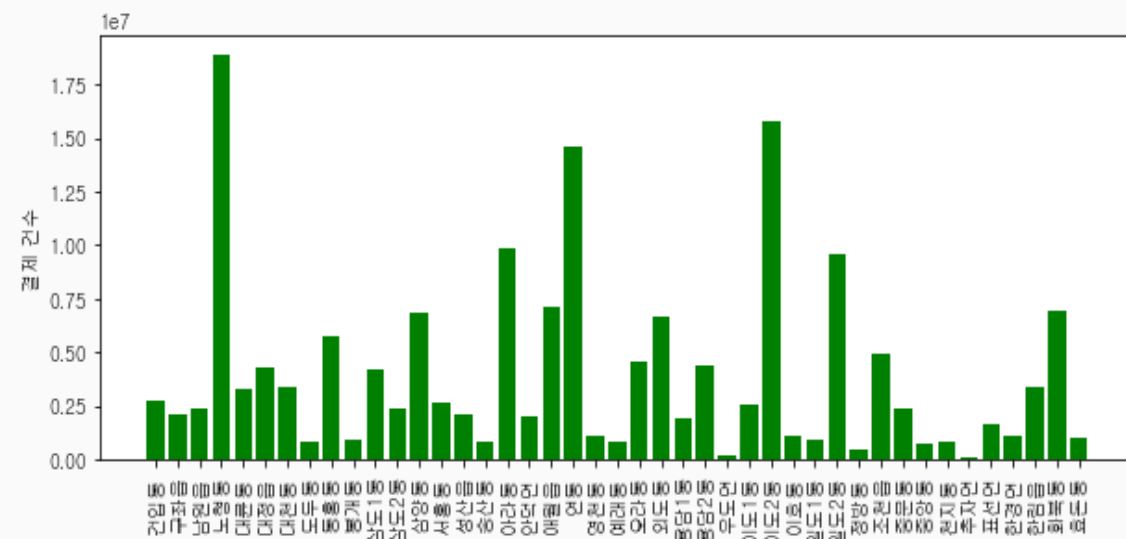
최종 Data Set

Test Set

“ 04_음식관련 카드소비_CARD_SPENDING.CSV ”

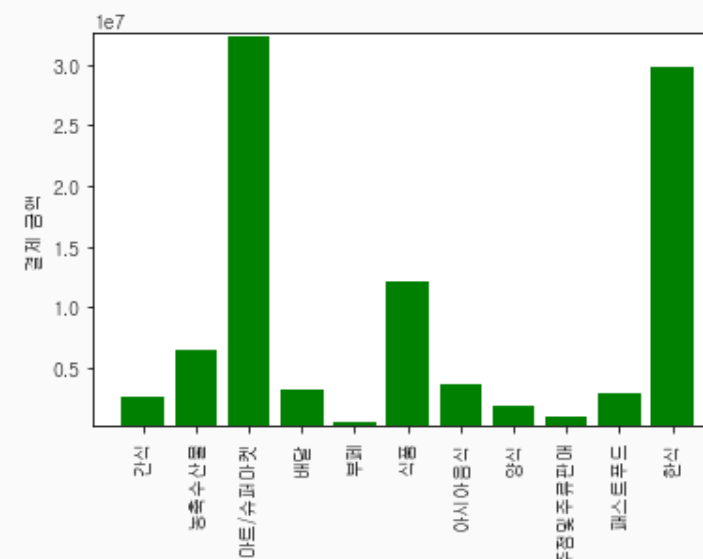
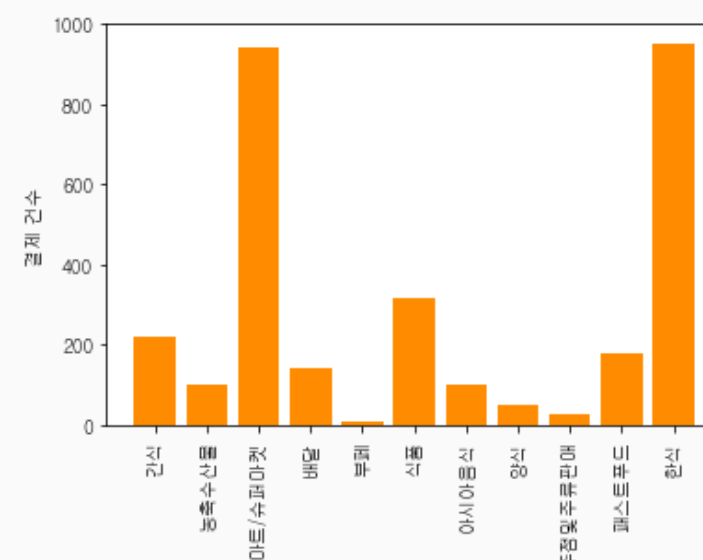
<결제건수 & 결제금액>

● 읍면동별 결제건수와 결제금액

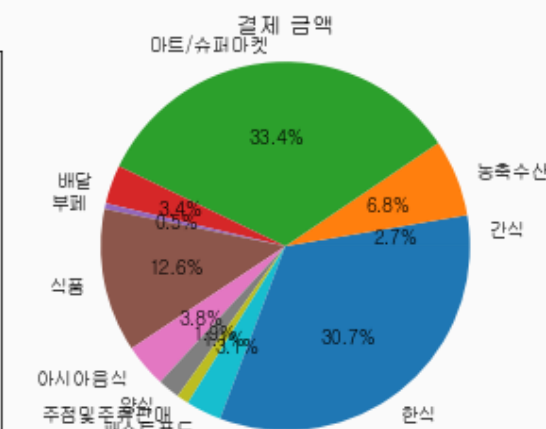
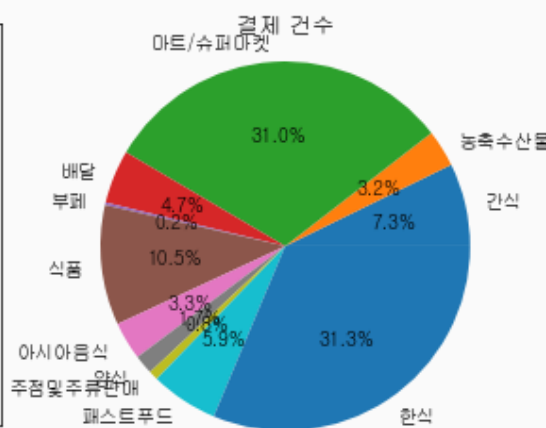


읍면동별 결제건수와 결제금액 추이가 비슷하게 나타남

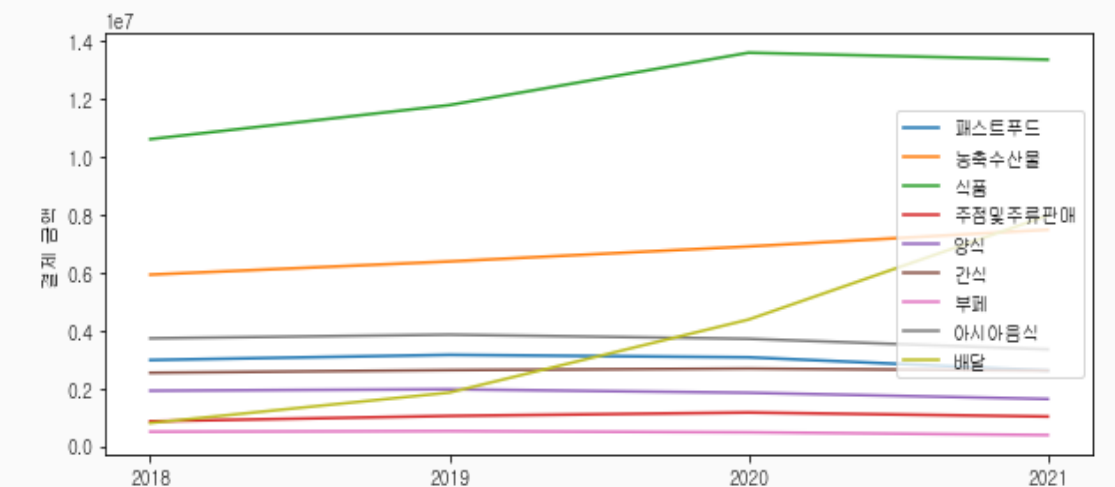
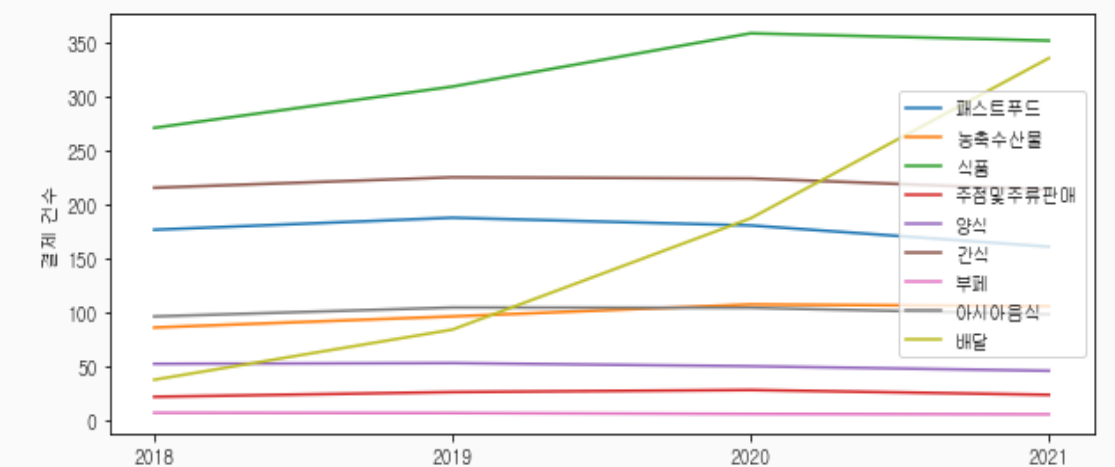
● 업종별 결제건수와 결제금액



마트/슈퍼마켓과 한식의 결제건수와 결제금액 비중이 높음



● 업종별 결제건수와 결제금액 추이



식품의 결제건수와 결제금액이 가장 높고
배달 업종의 추이가 지속적으로 증가하는 추세

02. 데이터 분석

“EDA”

Feature Engineering

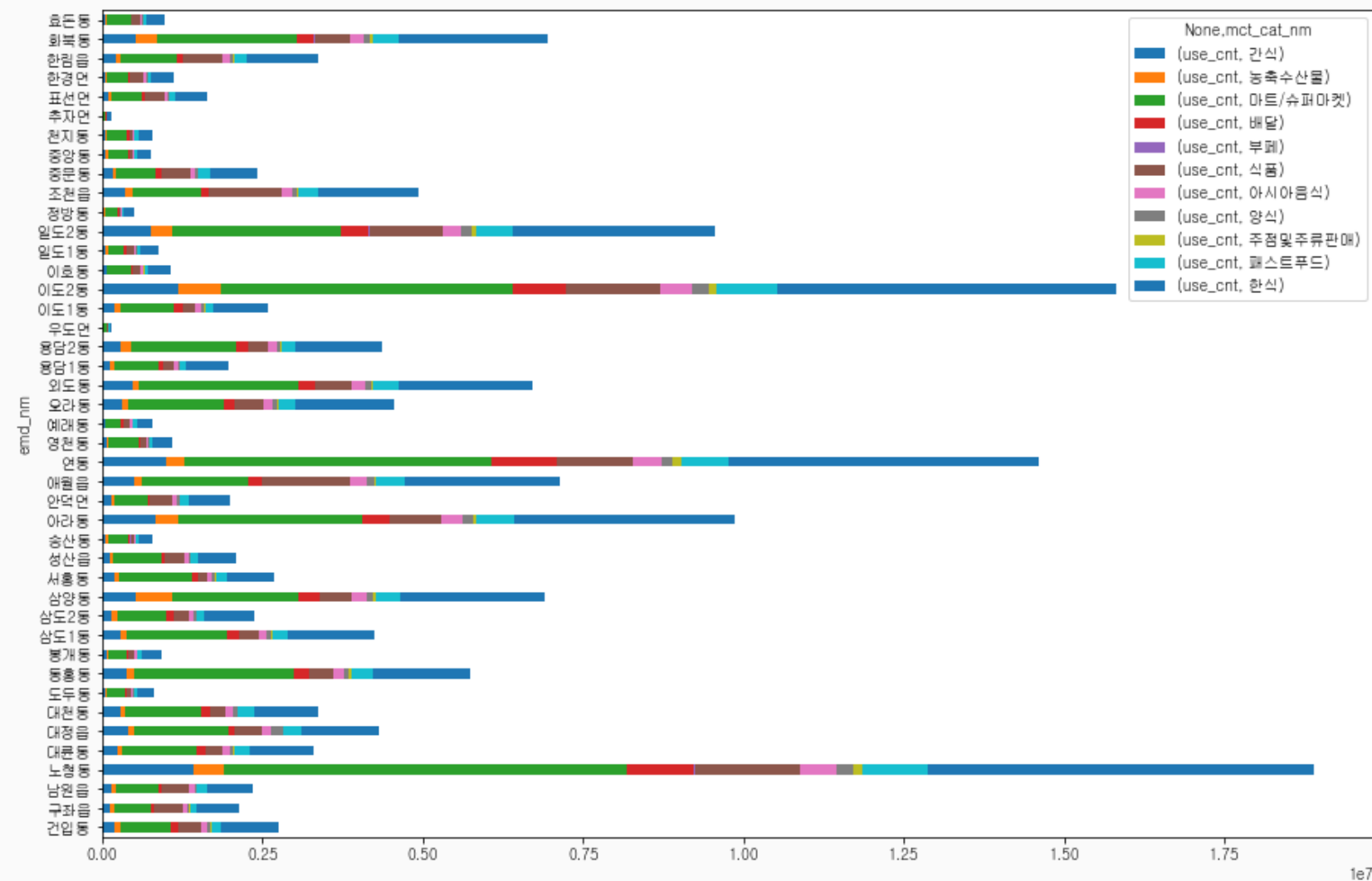
최종 Data Set

Test Set

“ 04_음식관련 카드소비_CARD_SPENDING.CSV ”

<결제건수 & 결제금액>

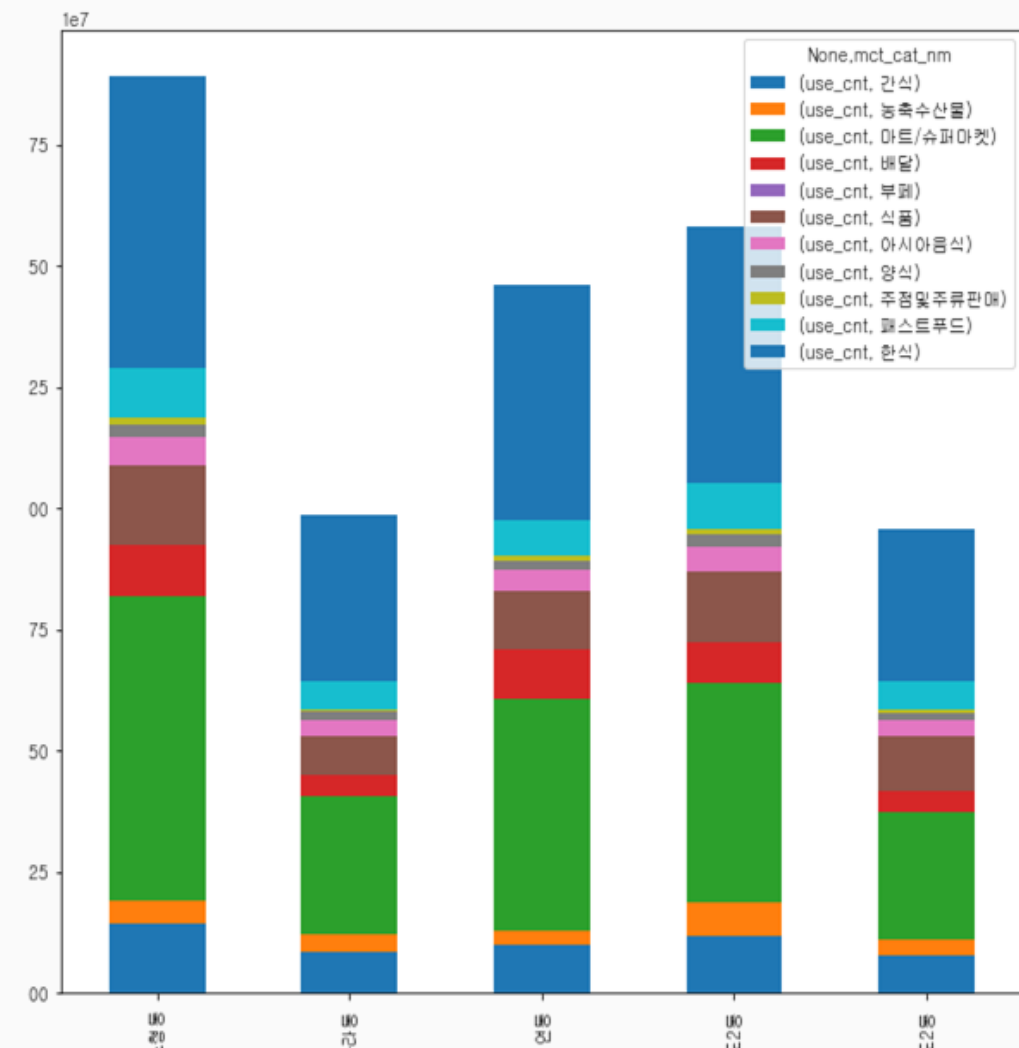
● 지역별 업종별 결제건수 총합비교



지역별로 업종별 누적 결제건수 총합을 나타낸 그래프



결제건수 Top 5 지역



총합 결제건수 Top 5

아라동
일도2동
이도 2동
연동
노형동

02.데이터 분석

EDA

“Feature Engineering”

최종 Data Set

Test Set

01

의미 있는 feature 도출

Total_pop

일별 읍면동별 전체 인구수 (연령대, 성별, 내/외국인 상관없이)

= (일별 읍면동 별 모든 연령대, 성별 인구의 합) / 시간대의 갯수
->인구는 시간대별 변화율이 큰 데이터가 아니기 때문에 일별로 구한 후 평균 처리

Visit_pop

제주에 방문하는 인구 수

= 단기 체류 외국인 인구 + 장기 체류 외국인 중 방문 인구 + 내국인 유동인구 중 방문인구

Jeju_resd_pop

제주에 거주하는 인구 수

= 거주지역이 제주인 내국인의 거주, 방문, 근무인구 + 거주지역이 제주인 외국인 거주, 인구, 방문인구

Total_frgn_pop

외국인 인구 수

= 일별 읍면동 별 장기 체류 외국인 수 + 단기체류 외국인 수

Youth_pop

제주 청년 인구 수

= 읍면동 별 2030세대의 인구 수

Delivery_amt

배달 업종의 결제 금액

= 일별 읍면동 별 배달 업종의 결제 금액의 합

02.데이터 분석

EDA

“Feature Engineering”

최종 Data Set

Test Set

02

결측치 처리

- Food waste Tab

year	2018	2019	2020	2021
emd_nm				
구좌읍	0	1629	28698	14130
한림읍	0	2441	34199	18245
한경면	0	856	19314	10431
조천읍	0	2226	38404	20105
애월읍	730	4960	45359	24734
알수없음	1090	221	255	188
중앙동	1825	1463	1306	724
성산읍	2195	14878	14776	6587
표선면	2494	12029	12200	5979
정방동	2807	2758	2551	1252
천지동	3614	3296	3231	1424
일도1동	3648	3642	3640	1622
효돈동	4670	2246	2595	2088
송산동	4899	4397	4830	2309
예래동	5397	5774	5998	3155
서홍동	5925	5219	5821	3300
안덕면	6463	18165	20502	10381
영천동	8288	5254	5155	4165
도두동	8439	8419	8033	4181
이도1동	9089	9123	9098	4189

year	em_g
2018	117482300
2019	7527000
2020	14053569
2021	12202100

<연도별 '알수없음' 배출 쓰레기양>

-연도별 쓰레기 발생 집계 읍면동 개수가 상이
2018년 37개 동 + '알수 없음'
2019년 41개 동 + '알수 없음'
2020년 41개 동 + '알수 없음'
2021년 41개 동 + '알수 없음'

-4개년도 모두 우도면, 추자면 data 없음

- '알수없음'으로 집계되는 배출거점도 일정하지 않아 '알수없음'의 음식물 쓰레기 배출량도 연도별로 큰 차이가 발생

-다른 data tab에서는 '알수없음' 이 없거나 4개년도 모두 43개동으로 일정

<연도별 읍면동별 일별 배출거점지역 개수>

Food waste 에서 특정 읍면동의 data가 없어 dataset 구성 X
일관되지 않은 '알수없음'의 처리로 예측의 성능 감소 우려

Solution

Food waste Tab의 읍면동에 맞춰 Dataset 구성
일관된 '알수 없음' 예측을 위한 새로운 row 추가
'알수없음' 행정동 제거

'전체' = 일별 모든 읍면동의 데이터 SUM

'전체' - 'Food waste 기준으로 읍면동의 SUM'
= '알수 없음' 쓰레기 배출량

	year	month	date	emd_nm	em_g	total_pop	제주 _resd_ratio	제주 _resd_pop	visit_ratio
37	2018	1	1	전체	85755000	679922	0.695323	535888.570169	36.867978
75	2018	1	2	전체	90217650	679922	0.685813	534509.814556	34.455467
113	2018	1	3	전체	80277600	679922	0.684410	532472.061474	34.561440
151	2018	1	4	전체	80138700	679922	0.669223	514678.201983	36.043898
189	2018	1	5	전체	75798600	679922	0.684943	527383.565802	35.013158
...
50721	2021	6	26	전체	93151550	696950	0.714449	598908.232477	36.447751
50762	2021	6	27	전체	103777500	696950	0.732825	615909.320383	35.265150
50803	2021	6	28	전체	96764250	696950	0.689116	581919.109130	36.322925
50844	2021	6	29	전체	93479700	696950	0.685673	577437.646506	36.696385
50885	2021	6	30	전체	91728950	696950	0.685614	576750.852147	36.525970

02.데이터 분석

EDA

“Feature Engineering”

최종 Data Set

Test Set

02 결측치 처리

- Food waste
 - 일도 1동의 경우 2021년 6월 18일 이후 food waste data 없음

	base_date	city	emd_cd	emd_nm	em_area_cd	em_cnt	em_g	pay_amt
2177717	2021-06-14	제주시	50110510	일도1동	W6X163	3	2700	80
2177719	2021-06-14	제주시	50110510	일도1동	W6X165	3	7250	217
2177720	2021-06-14	제주시	50110510	일도1동	W6X166	19	25700	768
2177721	2021-06-14	제주시	50110510	일도1동	W6X167	4	3500	104
2177722	2021-06-14	제주시	50110510	일도1동	W6X168	4	3050	90
2177723	2021-06-14	제주시	50110510	일도1동	W6X169	10	16700	497
2179686	2021-06-15	제주시	50110510	일도1동	W6X172	13	35700	1067
2182313	2021-06-16	제주시	50110510	일도1동	W6X172	7	15650	469
2184311	2021-06-17	제주시	50110510	일도1동	W6X172	12	26650	796
2186374	2021-06-18	제주시	50110510	일도1동	W6X172	8	18100	541



다른 Tab의 일도 1동 데이터도
‘2021-06-18’ 까지 추출
Test set 추출 시, 6월 예측까지 추출

- 단기 체류 외국인 유동인구
 - 단기 체류 외국인의 경우에는 일별로 데이터가 있는 읍면동의 개수가 달라짐

-> 단순히 방문이 없는 경우이므로 0으로 처리

- 음식 관련 카드 소비
 - 각 열에 대해서 행정동 별 시계열 데이터를 만든 다음 interploate로 결측치 보간
 - 2018년 1월 1일이 결측치인 경우 해당 값에만 중앙값을 채움

emd_nm	효돈동	도두동
base_date		
2018-01-01	1	1
2018-01-02	1	1
2018-01-03	1	1
2018-01-04	1	1
2018-01-05	1	1
...
2021-06-26	0	0
2021-06-27	0	1
2021-06-28	0	0
2021-06-29	1	0
2021-06-30	1	1

<일별 단기체류 외국인 방문인구 유무>



Linear Interpolation

데이터를 시계열로 학습하여 연속성 있는 패턴을 발견한다. 앞,뒤의 값을 통하여 결측치를 선형으로 보간하는 방법

03

Scaling & Encoding

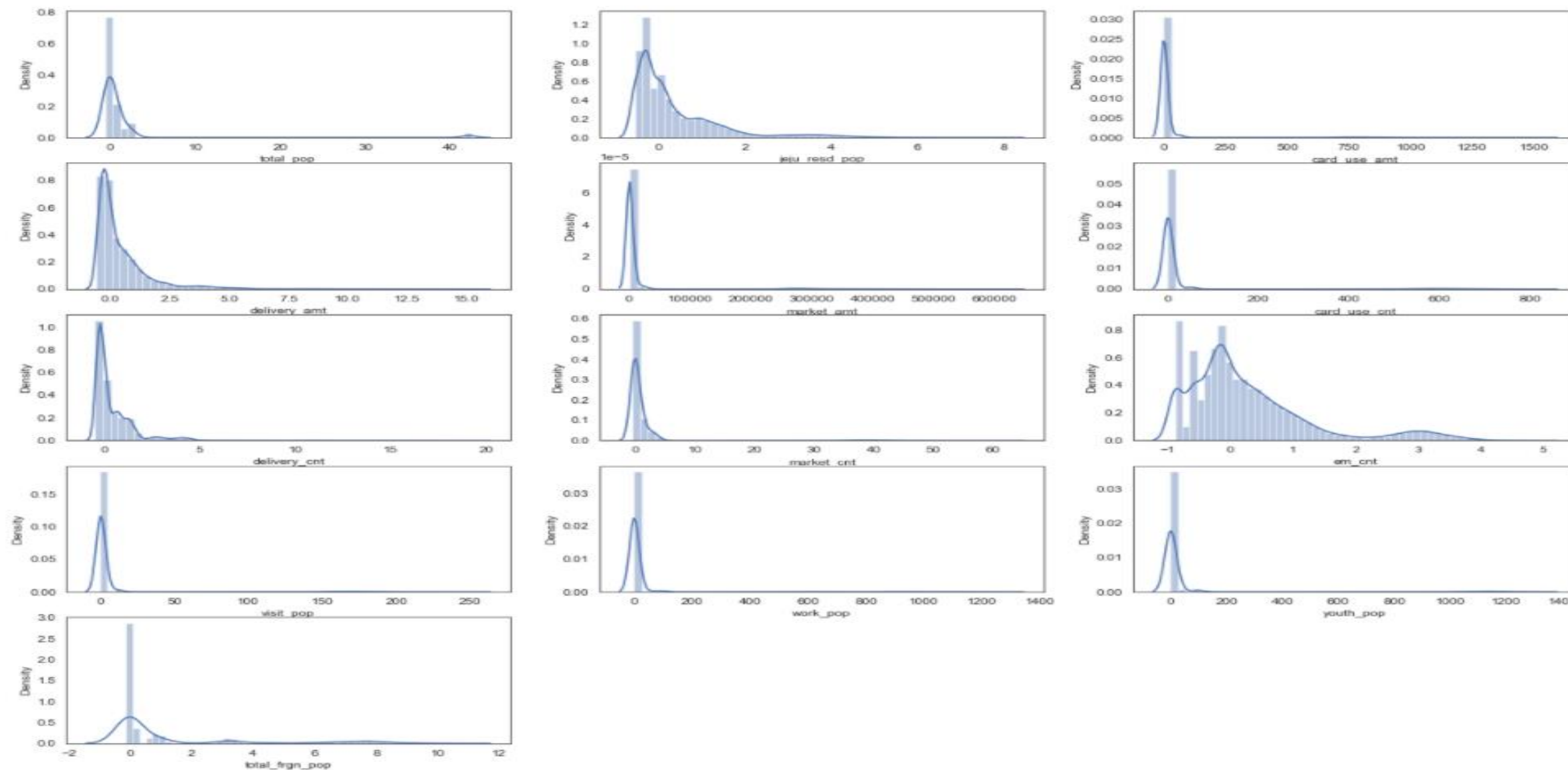
- Robust Scaler

평균과 분산 대신에 중앙값(median)과 사분위 값(interquartile range)을 사용하여 scaling 하는 기법으로, 이상치를 제거하므로 이상치의 영향을 최소화한다.

즉, 특성들이 같은 스케일을 갖게 되지만 **평균 대신 중앙값**을 사용하여, 극단값에 영향을 적게 받는 표준화 방식이라고 할 수 있다.

- Ordinal Encoder

Categorical data 중 순서에 의미를 부여하는 encode로 Base_date를 encoding 하여 날짜를 Categorical data로 변환한다

**base_date**

0	0.0
1	0.0
2	0.0
3	0.0
4	0.0
...	...
50881	1276.0
50882	1276.0
50883	1276.0
50884	1276.0
50885	1276.0

02.데이터 분석

EDA

Feature Engineering

“최종 Data Set”

Test Set

total_pop

제주도 총 인구 수

delivery_amt

배달 결제 금액

delivery_cnt

배달 결제 건수

total_frgn_pop

총 외국인 인구 수

jeju_resd_pop

제주 거주 인구 수

market_amt

마트/ 슈퍼마켓 결제 금액

market_cnt

마트/ 슈퍼마켓 결제 건수

card_use_amt

신용 카드 결제 금액

card_use_cnt

신용 카드 결제 건수

em_cnt

음식물 쓰레기 배출 건수

Visit_pop

거주 지역 그외 방문 인구

Work_pop

근무 인구

Youth_pop

청년 인구

02.데이터 분석



base_date	emd_nm	em_g	total_pop	jeju_resd_pop	card_use_amt	delivery_amt	market_amt	card_use_cnt	delivery_cnt	market_cnt	em_cnt	visit_pop	work_pop	youth_pop	total_frgn_pop
2018-01-01	전체	85755000	679922	535888.570169	3.786578e+09	20891720.0	1.602637e+09	108983.0	1057.0	40848.0	32512	7.210523e+06	8.301891e+05	54466.952499	2215.223209
2018-01-02	전체	90217650	679922	534509.814556	3.406059e+09	15597155.0	1.286966e+09	107164.0	839.0	38738.0	34664	6.761153e+06	1.368438e+06	53973.690842	2221.350506
2018-01-03	전체	80277600	679922	532472.061474	3.499232e+09	13453350.0	1.292679e+09	109765.0	728.0	38284.0	31025	6.770127e+06	1.369782e+06	53705.165769	2172.085577
2018-01-04	전체	80138700	679922	514678.201983	3.329829e+09	14592029.0	1.215080e+09	107110.0	740.0	36878.0	31540	6.993326e+06	1.351504e+06	53151.118467	2111.044522
2018-01-05	전체	75798600	679922	527383.565802	3.733755e+09	21105231.0	1.290936e+09	112064.0	1009.0	37686.0	29709	6.825653e+06	1.317354e+06	53360.869746	2055.834760
...
2021-06-30	표선면	1377450	12681	10137.380983	9.063287e+07	1855783.0	3.831039e+07	1336.0	75.0	423.0	370	1.614560e+05	2.405863e+04	914.664667	12.910869
2021-06-30	한경면	973850	9495	7257.895822	4.526361e+07	697661.0	1.767975e+07	1286.0	40.0	423.0	269	1.041218e+05	1.633182e+04	608.427552	22.165953
2021-06-30	한림읍	3298100	24462	20632.727912	1.301568e+08	5508665.0	4.207137e+07	3044.0	257.0	669.0	870	3.018026e+05	5.060204e+04	2064.041876	48.329788
2021-06-30	화북동	3749600	24630	24117.268821	1.918163e+08	13370520.0	4.846083e+07	5644.0	644.0	1645.0	2255	2.433203e+05	5.097469e+04	1810.431884	18.483512
2021-06-30	효돈동	1011050	5314	3726.154500	3.104390e+07	291740.0	1.556358e+07	714.0	15.0	322.0	313	4.439045e+04	5.386495e+03	283.896570	8.603778

ARIMA (Autoregressive Integrated Moving Average) Model

ARIMA 모델 전, 알아야 할 개념

- Autocorrelation(자기 상관)

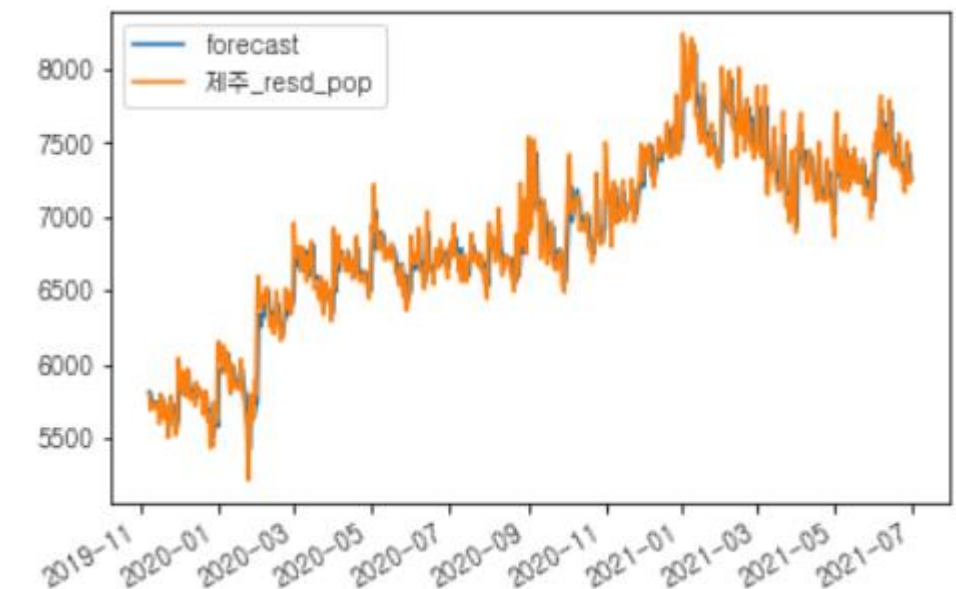
-> 한 random variable에 대해서 이전의 값이 이후의 값에 영향을 미치고 있는 있을 때, 자기상관 데이터라고 한다.

- Moving Average(이동 평균)

-> 한 random variable의 평균값이 지속적으로 증가 또는 감소하는 경향의 경우 이동 평균 데이터라고 하며, 예측 오차를 이용하여 이후의 값을 예측한다.

- Stationarity(정상성)

평균, 분산이 시간에 따라 일정한 성질로, 시계열 데이터의 특성이 시간의 흐름에 따라 변하지 않는다. 그러므로, 정상성 특징을 지닌 데이터는 추세나 계절성이 없다



<ARIMA 모델 이용하여 예측 결과>

y' = d차 차분을 구한 시계열 데이터
P = 자기 회귀 부분의 차수
q = 이동 평균 부분의 차수

최적의 (p,q,d)를 구해야 한다

$$y'_t = c + \phi_1 y'_{t-1} + \phi_2 y'_{t-2} + \dots + \phi_p y'_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t$$

ARIMA 모델은 시계열을 예측하는 대표 적인 모델

DATA에 나타나는 Autocorrelation(자기상관)을 표현하지만, 단순히 과거의 데이터를 사용하는 것이 아닌 **과거의 데이터가 지닌 추세까지** 반영한다.

또한, 시계열 데이터의 정상성을 가정하는데 그 이유는 복잡한 패턴을 모델링하여 분석하기 어렵기 때문이다. 정상성을 만족하지 않는 데이터는 로그 변환, 차분 등으로 전 처리를 시행한다.

02.데이터 분석



각 feature별로, 읍면동 별로 ARIMA 모델을 적용하여 2021-07-01 ~ 2021-08-31 data를 예측하여 7,8월 음식물 쓰레기 배출량을 예측할 Test dataset을 구성



base_date	emd_nm	card_use_cnt	visit_pop	youth_pop	total_pop	jeju_resd_pop	market_amt	card_use_amt	delivery_cnt	market_cnt	em_cnt	work_pop	total_frgn_pop	delivery_amt
2021-07-01	건입동	2149.200947	186013.345208	918.426171	9110.000000	6373.946295	2.124696e+07	5.949093e+07	243.573062	621.829694	691.101412	22279.504950	18.726354	6.051195e+06
2021-07-01	구좌읍	1888.344788	248254.331973	1581.126294	15955.000000	13322.572504	1.787525e+07	6.669524e+07	95.135022	460.313851	498.108840	39094.694098	21.372934	2.296941e+06
2021-07-01	남원읍	2033.618443	197437.511454	1083.864525	19353.564276	13976.727235	3.102192e+07	8.008490e+07	114.509864	572.678068	591.539913	31182.018084	32.357490	2.344106e+06
2021-07-01	노형동	15891.179785	560200.270413	5120.128184	57473.000000	52538.355479	1.515518e+08	4.561962e+08	2232.411807	4692.282906	4541.465743	134960.666900	68.696882	5.190421e+07
2021-07-01	대륜동	2945.429814	229885.394412	1620.484395	15277.000000	13743.968561	4.030110e+07	9.897430e+07	324.000062	886.679954	768.790514	34735.201229	18.243095	7.627131e+06
...
2021-08-31	표선면	1405.848020	169165.168058	951.892810	12681.000000	10389.363506	1.781009e+07	5.198330e+07	91.211040	404.954847	359.987455	22057.011743	13.355359	2.271676e+06
2021-08-31	한경면	1074.256764	122062.153919	656.824383	9495.000000	7287.302869	1.251964e+07	3.822899e+07	37.481373	306.674145	280.516475	15558.745952	24.812021	8.405571e+05
2021-08-31	한림읍	2876.701236	315808.323921	2137.575902	24462.000000	20938.154470	2.693904e+07	1.055959e+08	243.206314	672.739217	889.049337	46252.758838	50.141742	6.211514e+06
2021-08-31	화북동	5688.235168	252468.251400	1793.119429	24630.000000	24524.363963	4.707253e+07	1.678146e+08	560.305073	1526.696734	2243.415085	45586.976305	17.741194	1.284811e+07
2021-08-31	효돈동	831.940062	42317.331897	293.381777	5314.000000	3804.823261	1.319068e+07	2.959266e+07	33.580763	296.205064	300.601780	4968.595897	8.255073	6.830147e+05

03. Modeling

03. Modeling

Model 선정

“Model 성능”

예측 결과

AutoML(Automated Machine Learning)

AutoML이란 아래와 같은 반복적인 기계 학습 모델 개발 작업을 자동화하는 프로세스

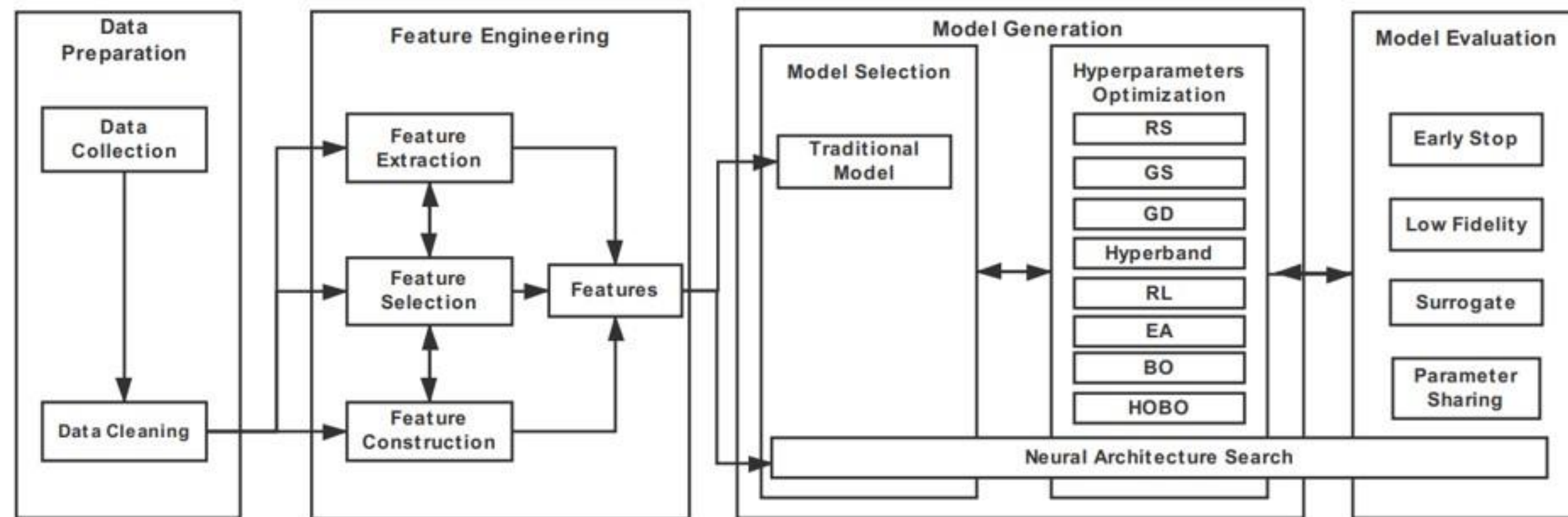
AutoML 라이브러리 중 **PyCaret** 사용

- scikit-learn 패키지를 기반으로 하고 있음
- 다양한 함수를 지원하여 유용하게 사용 가능

많은 알고리즘에 대한 결과를
한 번에 볼 수 있어서 시간 단축

+

더 효율적인 방향으로 학습



03. Modeling

평가 지표 RMSE

잔차 제곱합에 루트를 씌웠기 때문에 큰 오류값 차이에 대해서 크게 패널티를 주게 되어 이상치의 영향을 많이 받지 않으므로 평가 지표로 선택

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

STEP 1 파라미터 튜닝이 되지 않은 채로 가장 잘 나온 top3 모델 선정

```
best_3 = compare_models(sort = 'RMSE', n_select = 3, fold=5)
print(best_3)
```

	Model	MAE	MSE	RMSE	R2
catboost	CatBoost Regressor	2.782356e+05	1.569324e+11	3.903804e+05	0.8947
gbr	Gradient Boosting Regressor	3.132330e+05	1.943525e+11	4.372704e+05	0.8678
lightgbm	Light Gradient Boosting Machine	3.212560e+05	2.076384e+11	4.527010e+05	0.8596
et	Extra Trees Regressor	3.223967e+05	2.140643e+11	4.578839e+05	0.8556
rf	Random Forest Regressor	3.448495e+05	2.400723e+11	4.864237e+05	0.8371

STEP 2 3개의 단일 모델에 대해서 각각 파라미터 튜닝

```
model_01 = create_model('catboost')
model_01 = tune_model(model_01, fold= 5, n_iter = 100, optimize = 'RMSE',
                      verbose = True, choose_better = True)
tuned_models.append(model_01)
```

	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	258108.3528	1.107263e+11	332755.6009	0.9167	0.0410	0.0321
1	230700.3106	9.870029e+10	314166.0285	0.9424	0.0414	0.0301
2	273178.6622	1.281180e+11	357935.7050	0.8955	0.0460	0.0360
3	318676.6563	2.550656e+11	505040.1617	0.8630	0.1005	0.0494
4	304430.5911	1.851872e+11	430333.8540	0.8625	0.0578	0.0406
Mean	277018.9146	1.555595e+11	388046.2700	0.8960	0.0574	0.0376
SD	31637.1219	5.792285e+10	70566.0042	0.0310	0.0224	0.0069

03. Modeling

STEP 3 3개의 단일 모델을 기반으로 한 배깅, 부스팅, 스택킹, 보팅 진행

```
stacking = stack_models(estimator_list=[model_01, model_02, model_03],  
                        meta_model = model_02, fold=5, optimize='RMSE')  
prediction_models.append(stacking)
```

	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	280235.2220	1.321729e+11	363555.9039	0.9006	0.0458	0.0354
1	263128.9188	1.369748e+11	370101.0600	0.9201	0.0537	0.0356
2	274306.5705	1.299725e+11	360516.9942	0.8940	0.0464	0.0363
3	353464.8691	2.638925e+11	513704.7294	0.8583	0.0954	0.0525

Stacking

```
boo_01 = ensemble_model(model_01, method = 'Boosting', fold=5, optimize = 'RMSE')  
prediction_models.append(boo_01)
```

	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	279751.6766	1.264038e+11	355533.1618	0.9049	0.0437	0.0348
1	262790.3706	1.242489e+11	352489.5276	0.9275	0.0470	0.0348
2	301851.4021	1.631766e+11	403951.2415	0.8669	0.0519	0.0397
3	333516.0366	2.941370e+11	542343.9980	0.8420	0.1053	0.0517

Boosting

```
bag_01_15 = ensemble_model(model_01, n_estimators = 15, fold=5, optimize = 'RMSE')  
prediction_models.append(bag_01_15)
```

	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	276371.3220	1.311472e+11	362142.4532	0.9014	0.0442	0.0343
1	260609.4268	1.311285e+11	362116.6513	0.9235	0.0465	0.0339
2	303391.4984	1.643563e+11	405408.8248	0.8660	0.0528	0.0403
3	334443.0102	3.019139e+11	549466.9055	0.8378	0.1057	0.0516

Bagging

```
blended = blend_models(estimator_list = best_3, fold = 5)  
prediction_models.append(blended)
```

	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	266655.7152	1.236710e+11	351668.8456	0.9070	0.0432	0.0333
1	252203.9604	1.137174e+11	337220.1372	0.9336	0.0431	0.0328
2	292255.8048	1.674774e+11	409240.0704	0.8634	0.0521	0.0385
3	316069.6835	2.583469e+11	508278.3560	0.8612	0.1015	0.0494

Blending

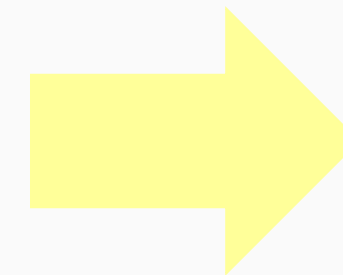
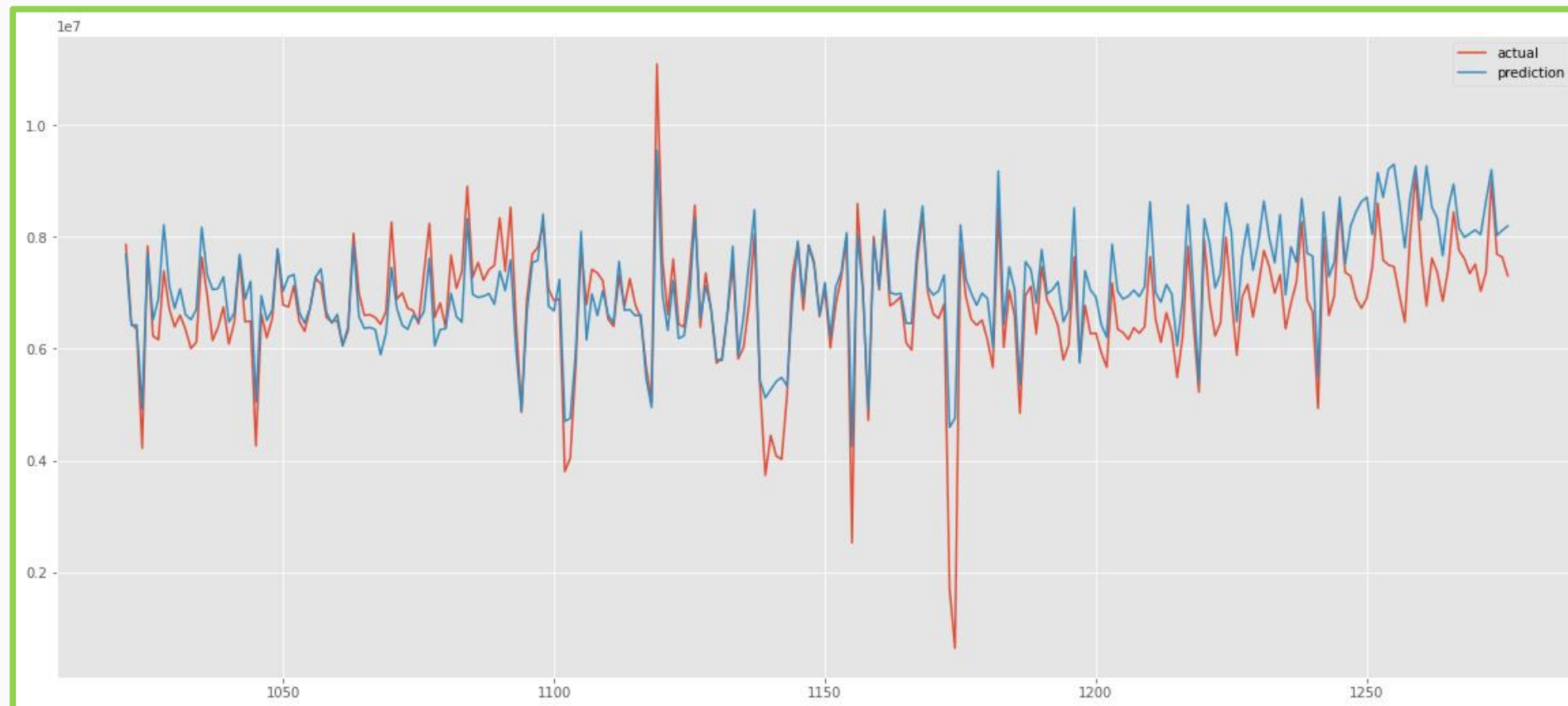
03. Modeling

STEP 4 fold 평균 성능이 좋은 모델 3개를 선정하여 finalize 진행

->이전 과정에서는 입력한 데이터셋에서 train과 valid dataset을 나누어 학습했다면, 해당 단계에서는 전체 훈련 데이터셋을 이전에 진행한 파라미터를 토대로 한 번 학습한다.

```
final_01 = finalize_model(model_02)
final_02 = finalize_model(boo_02)
final_03 = finalize_model(blended)
```

STEP 5 최종 모델의 예측값을 실제 검증 데이터셋의 결과와 비교



제주도 7,8월
음식물 쓰레기
배출량 예측

03. Modeling

7,8월 예측 결과

NO	행정동명	7월 배출량(g)	8월 배출량(g)
1	한림읍	43166465.6	70881406.3
2	애월읍	6570723.0	9912937.6
3	구좌읍	30396282.5	29812952.3
4	조천읍	47070125.0	52000691.5
5	한경면	8586445.7	19413513.1
6	일도1동	15924583.9	19672648.9
7	일도2동	157815735.5	141508193.2
8	이도1동	45378874.7	47796486.6
9	이도2동	227837543.9	229562759.8
10	삼도1동	68764904.9	61934906.0
11	삼도2동	34935141.5	35148724.3
12	용담1동	37655924.0	35808802.7
13	용담2동	62324810.4	64386435.4
14	건입동	47448317.7	42484157.6
15	화북동	118758880.4	106707166.9
16	삼양동	102487757.0	94408609.7
17	봉개동	15526977.1	16808132.6
18	아라동	118899403.4	113206377.6
19	오라동	59421070.9	55851654.3
20	연동	211733744.0	180073550.9

21	노형동	274821929.2	222764816.4
22	외도동	92212826.1	88199621.9
23	이호동	20973347.3	20375198.8
24	도두동	16061999.0	17756910.6
25	대정읍	63075533.3	91954297.9
26	남원읍	67707033.7	56766859.2
27	성산읍	46685498.7	51914128.9
28	안덕면	17047415.2	31846452.7
29	표선면	23863518.2	24174961.7
30	송산동	24064330.9	26233258.7
31	정방동	24937079.1	24840363.4
32	중앙동	45873937.2	49550066.1
33	천지동	28736728.6	35215799.2
34	효돈동	16289410.8	13626148.2
35	영천동	29420358.1	26781497.9
36	동홍동	98721244.2	108973923.9
37	서홍동	41042724.6	48201876.8
38	대륜동	38279861.0	51181886.8
39	대천동	52331056.7	49863987.1
40	중문동	41477750.8	60397178.8
41	예래동	15527432.1	16325509.2
42	알수없음	16756616.1	0.0

행정동명 : 전체 인 배출량



나머지 행정동 배출량 합



‘알수없음’ 행정동 배출량

04. 해결 방안

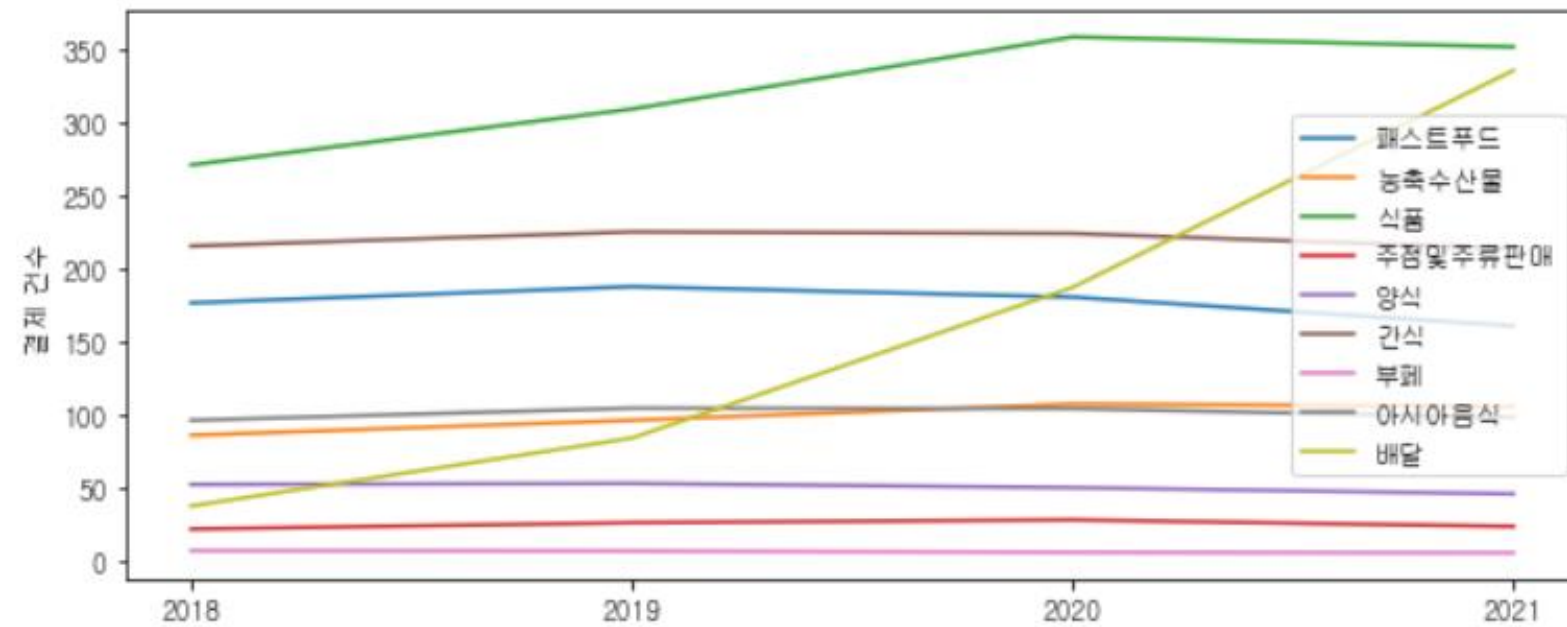
04. 해결 방안

“해결방안”

의의 및 한계점

제시 1

제주도에서도 용기를 내세요!



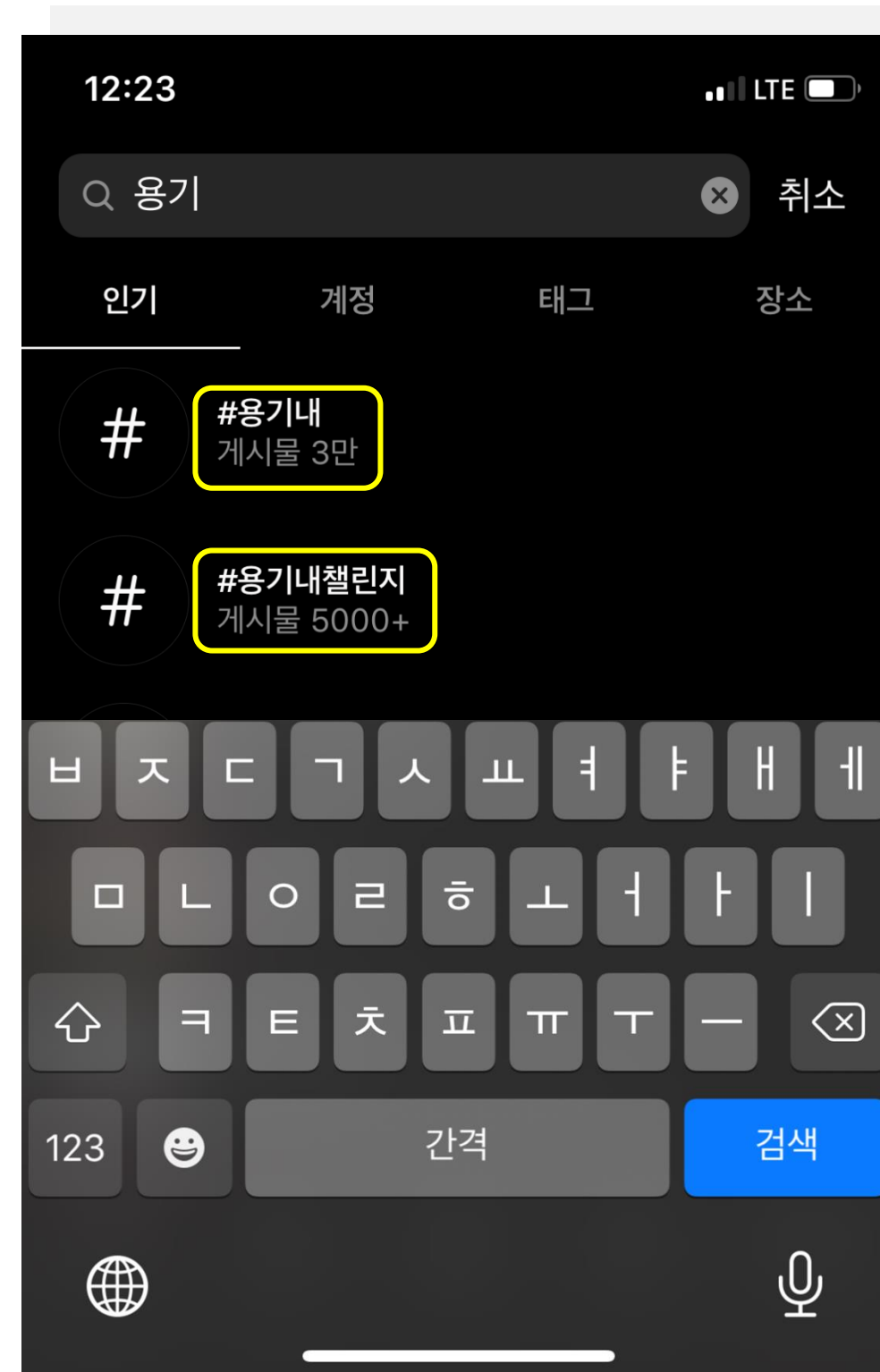
음식 소비의 문화가 변화

직접 방문 -> Take Out , Delivery로 집에서 즐기기

‘용기 내’ 챌린지란?

배달 대신 포장

일회 용기 대신 다회 용기를 내는 챌린지의 트렌드
환경을 생각하는 소비를 지향하는 소비자의 증가



“ 제주도에서도 용기내! ”

04.해결 방안

“해결방안”

의의 및 한계점

제시 1

제주도에서도 용기를 내세요!

제주도에서 용기는 특별한 용기로!



상업적 효과



남은 음식 포장 또는
배달 음식 포장으로
쓰레기 절감 효과

제시 2

I'm perfect!

버려진 폐감귤..양심도 환경도 썩고 있다

출처 : 제주일보

감귤철마다 반복되는 부패 감귤 무단투기 행위로 인해 제주 청정 환경이 위협 받고 있다.
21일 오전 서귀포시 중문동 ‘하원동 탐라 왕자묘’ 인근 도순천에 감귤이 무더기로 버려져 있었다.

매립장 포화...비상품 감귤 처리 비상

출처 : 제주MBC

봉개동 쓰레기 매립장이 포화 상태에 이르면서, 감귤 부패과 처리에 비상이 걸렸는데요.

가뜩이나 올해는 잦은 비에 부패 발생이 많아 출하를 앞둔 농가의 근심이 커지고 있습니다.

폐감귤이란?

- 기후나 재배 과정 문제로 부패가 발생하여 판매할 수 없는 감귤
 - 상품성이 떨어져 판매 가치가 떨어지는 감귤

1년에 4600톤!

이런 감귤들을 활용할 수 있는 방법이 없을까?
폐감귤을 처리하는 올바른 방법은 없을까?
폐 감귤도 어딘가에서는 가치 있을 수 있다!

I'm perfect!

제시 2

I'm perfect!

Perfect 1

부패로 인해 버려야 한다면, 제대로 올바르게 버리자

농가의 저장고에서 발생한 부패 감귤 처리 방법에 대해 농가를 교육 실시, 책임 부여
가정에서 발생한 부패 감귤은 음식물 쓰레기 종량제의 규정을 적용함을 알린다,

Perfect 2

자연으로 다시 깨끗하게 돌려주기

발효 퇴비로 제조하여 농업에서 재사용

Perfect 3

못생겨도 맛있는 B급 상품 감귤 판매 촉진

다양한 상품 개발 ex) 감귤잼

농가에서
불법으로 폐기처리하지
않도록, 발효 퇴비로 만들
어지는 기간 동안 저장할
수 있도록

퇴비화 시설 확충 필요

의의 01

- 제주도에서의 소비 트렌드 변화 발견 -> 배달 업종의 급부상

한계점 01

- 코로나로 인해 데이터의 흐름이 변화한 현실을 반영한 모델링을 하지 못함

의의 02

- 대표적인 관광지로 음식물 쓰레기 또한, 관광업으로 인한 배출이 많을 것이라고 생각했으나, 제주도 거주 인구의 영향이 더 큰 것으로 나타남

한계점 02

- 음식물 쓰레기 배출에 다양한 사회적 요인들이 영향을 끼치다 보니 많은 데이터 속에서 의미 있는 feature을 도출하고, feature의 특성에 맞게 engineering하는 과정이 부족했음.

05.Reference

분석도구



참고문헌

‘부패감굴 처리대책, 어째서 미온적인가’,제주일보,2019년 10월 13일
‘버려진 폐감굴..양심도 환경도 썩고 있다’,제주일보,2021년 3월 22일
‘매립장 포화... 비상품 감굴 처리 비상’,제주MBC,2019년 10월 11일
‘봉개 쓰레기 소각장 포화 '여전'...2만7000t 야적’,제주일보,2021년 3월 3일
‘제주 명품 이호테우해변, 실종된 시민의식에 쓰레기 천국’,국제 뉴스,2020년 5월 25일
‘제주 음식물처리시설 첫삽...봉개동 기한만료 '발등에 불',국제 뉴스,2021년 6월 8일

THANK YOU



팀명:감귤맛 삼다수
팀장:홍지원(ant67410@gmail.com)
팀원
권예은(yenny7013@gmail.com)
이보림(ant67410@gmail.com)
유승희(juneyo0823@naver.com)