

생활폐기물 발생 구역 GIS 분석

2020.10.12 - 2020.12.31

I 분석배경 및 목적

해당 시는 폐기물과 관련된 화재사고가 발생하는 등 폐기물에 의한 문제를 갖고 있으며, 그 중 생활폐기물과 불법폐기물은 지자체의 관리가 필요함. 따라서 생활폐기물 수거데이터를 토대로 생활폐기물의 추이와 집중 관리가 필요한 장소를 분석함.

더불어 생활폐기물의 성상별 수거횟수, 평균중량 등의 정량지수를 토대로 청소관리체계에서 인적, 물적 자원의 분배를 개선할 수 있는 인사이트를 도출함.

II 분석개요

1. 활용데이터

1) 폐기물 중량데이터 2) 공간 데이터 3) 생활권 데이터

활용 데이터		구분	중요도	생성주기	지역속성
1)	동부적환장 계근량 (19.01.01-19.12.31)	정형/내부	필수	시간	구역, 행정동
	동부적환장 계근량 (20.01.01-20.08.31)	정형/내부	필수	시간	구역, 행정동
	서부자원화시설 전체계근 (19.01.01-19.12.31)	정형/내부	필수	시간	구역, 행정동
	서부자원화시설 전체계근 (20.01.01-20.08.31)	정형/내부	필수	시간	구역, 행정동
2)	거래처별 처리구역	정형/내부	필수	연간	행정동
	행정동 SHP	정형/내부	필수	연간	행정동
3)	행정동별 거주인구현황 (2019년)	정형/내부	선택	연간	행정동



[폐기물 중량데이터]



[거래처별 담당구역]

- 폐기물 중량 데이터는 입고 대상만 조회했으며, 품목, 일자, 계근량, 거래처, 운행차량번호 등이 수집됨.
- 품목, 거래처 데이터는 중복, 오기를 정제하여 표준화했으며, 화성시 재활용품 수거기준에 맞춰 새로운 카테고리를 부여함.
- 거래처별 담당 구역을 기준으로 폐기물 수거 지역을 추론했으며, 폐기물 중량데이터 중 거래처가 표기된 데이터를 1차 집계하고 행정동이 표기된 경우를 2차 집계함.

데이터 한계

- 상차시각과 공차시각데이터는 이상치가 다수로 품목별 평균 수거 시간 파악 불가능
- 폐기물의 정확한 수거 지점을 알 수 없음
- 일부 품목의 경우 특정 업체에서 전량 수거하므로 수거지역을 추론할 수 없음

생활폐기물 발생 구역 GIS 분석

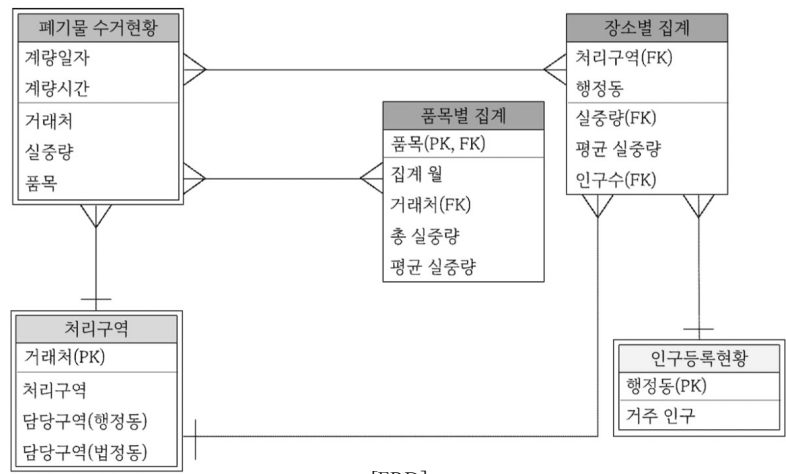
2020.10.12 – 2020.12.31

II 분석개요

2. 분석프로세스



[분석 프로세스]



[ERD]

3. 분석방법

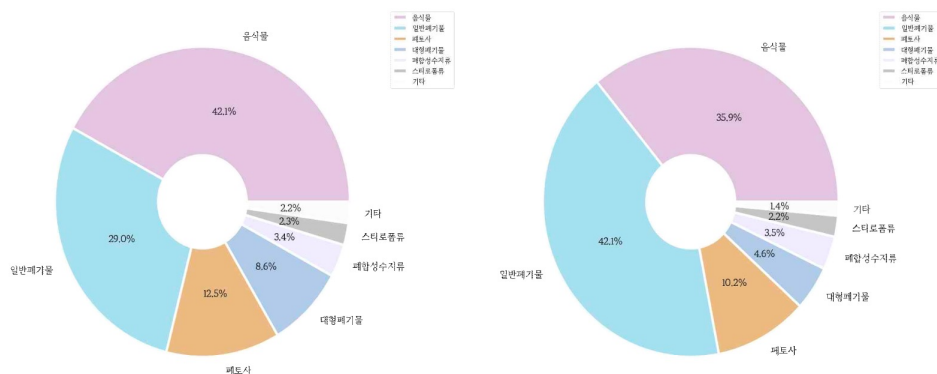
- 분석 도구 : Python, QGIS
- 원본 데이터에서 필요한 feature를 추출하여 ERD 구상 후 물리적 모델링
- 품목별, 기간별 폐기물을 집계를 통하여 폐기물 수거현황 파악 【1 품목별 집계】
- 폐기물 중량데이터의 수거장소(거래처, 행정동) 정제 후 GIS 분석 【2 GIS분석】
- 폐기물 중량데이터의 시계열 분석(Seasonal ARIMA model) 【3 시계열 분석】

생활폐기물 발생 구역 GIS 분석

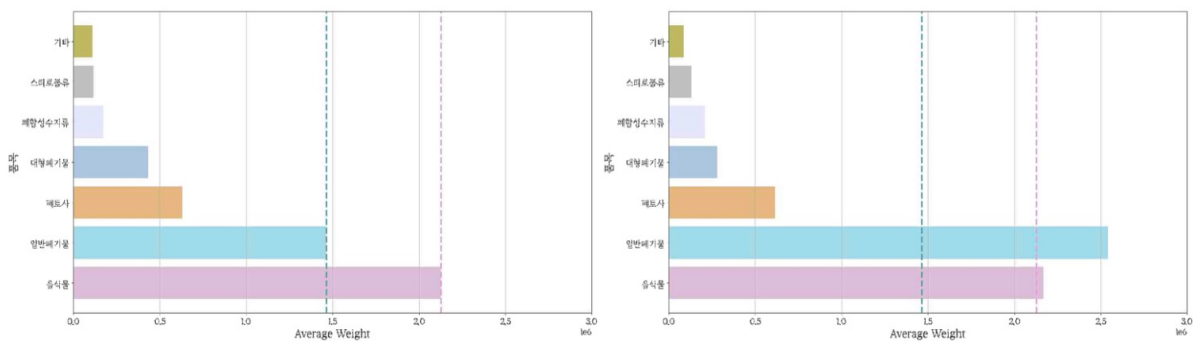
2020.10.12 - 2020.12.31

Ⅲ EDA Object

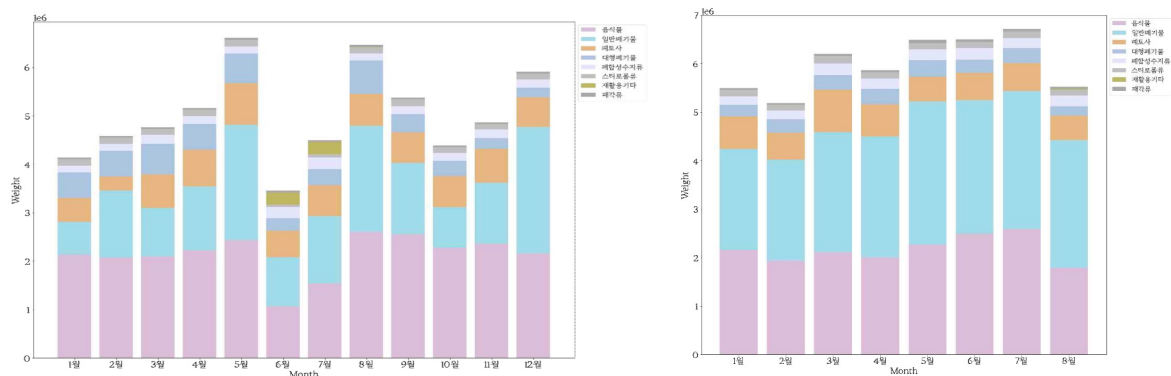
1. 품목별 집계



[생활폐기물의 구성]



[생활폐기물 월평균 발생량]



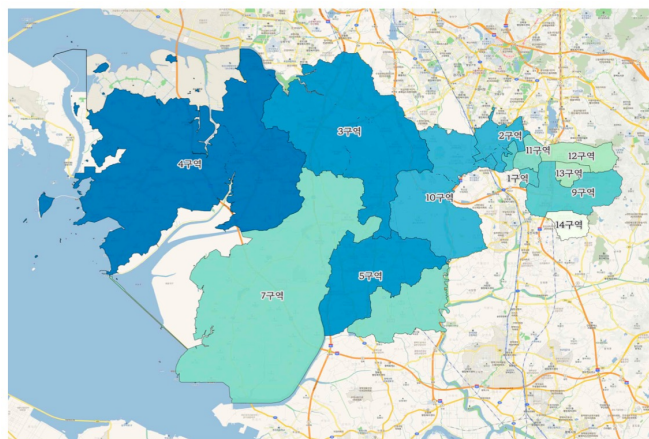
[생활폐기물 연간 발생량 추이]

생활폐기물 발생 구역 GIS 분석

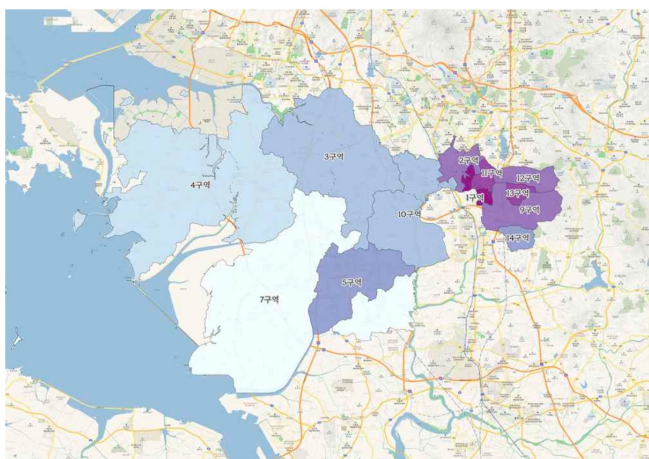
2020.10.12 - 2020.12.31

Ⅲ EDA Object

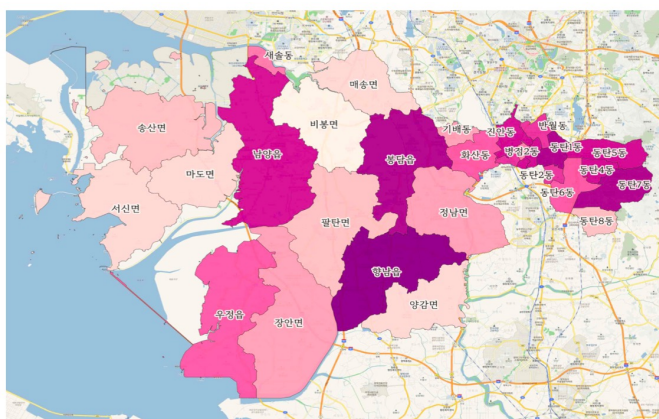
2. 장소별 집계



[일반폐기물 총 발생량]



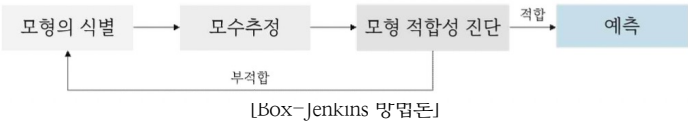
[면적당 발생비율]



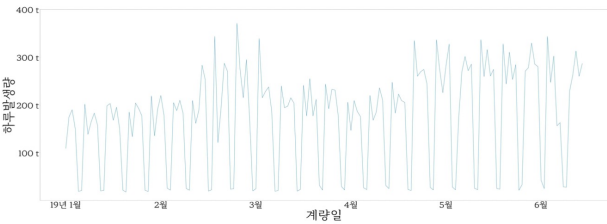
[행정동 단위 발생 추정도]

IV Time Series Analysis

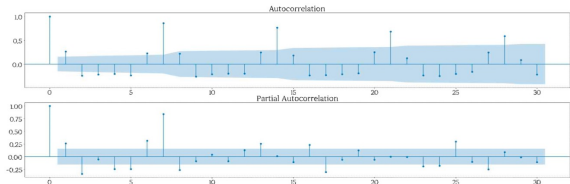
- 전체 기간 중 2019년 6월 15일 - 7월 14일에 등록된 폐기물 수거데이터가 측정되지 않음. 따라서 이전 구간의 폐기물 발생량으로 해당 구간의 폐기물 발생량을 보완하는 시계열 분석을 도입함.
- Box-Jenkins 방법론으로 접근하여 예측에 적합한 시계열분석모델을 파악함



모형의 식별과 모수의 추정



[2019년 1월 - 6월 14일 일별 폐기물 발생량]



[[위] ACF (아래) PACF]

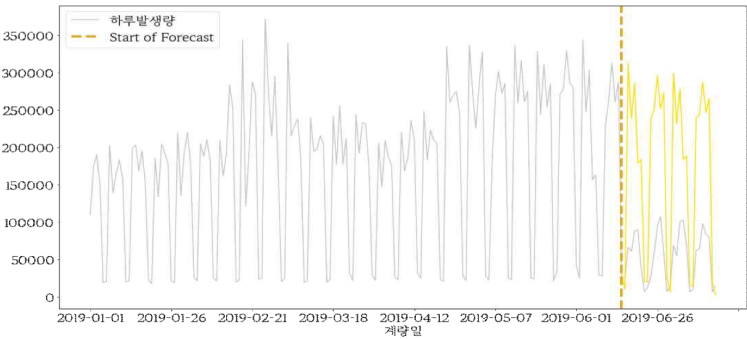
No	p	d	q	P	D	Q	R	r ² score
1	2	1	1	1	1	1	7	0.75
2	1	0	2	1	1	1	14	0.71
3	2	1	1	1	1	1	14	0.76
4	2	0	2	0	1	1	28	0.73
5	2	1	3	0	1	0	28	0.68

[r² score]

그래프에서 계절성이 확인되므로 SARIMA Model을 선정했으며, $(p,d,q)(P,D,Q,R)$ 을 추정함. ACF / PACF를 참조하여 $m=7, d=1, D=1$ 로 가정함. Auto-ARIMA package를 활용하여 $1 \leq p \leq 4, 1 \leq q \leq 4$ 의 범위 내에서 AIC가 낮은 모수를 선정하고 r^2score 로 모델 적합성을 진단함.

예측 결과

SARIMA Prediction Results (r2_score : 0.76)



[2019년 6월 15일 - 7월 14일 생활폐기물 발생량 예측 결과]

날짜	기록	추정	날짜	기록	추정	날짜	기록	추정
6/15	25,350	31212.12	6/25	56,630	249844.40	7/5	69,520	188353.90
6/16	25,710	10170.86	6/26	94,750	295710.60	7/6	7,070	15869.05
6/17	66,630	311950.10	6/27	107,502	251955.00	7/7	10,780	14914.55
6/18	61,230	238474.20	6/28	57,360	272519.40	7/8	61,940	239257.80
6/19	88,060	285954.60	6/29	7,700	24331.06	7/9	64,490	243005.70
6/20	89,770	178465.00	6/30	14,290	6706.01	7/10	98,050	287397.60
6/21	40,500	184171.10	7/1	69,320	298975.60	7/11	83,740	246575.50
6/22	6,800	21205.61	7/2	55,170	231672.60	7/12	80,570	265095.60
6/23	12,950	20187.17	7/3	100,900	277810.70	7/13	6,630	19418.42
6/24	28,330	238822.60	7/4	102,650	183401.30	7/14	14,810	2915.68

월	기록	추정값	19년 평균	2020년	추정증가율
6월	3,473,672	5,312,550	5,055,178	6,551,340	23.31%
7월	4,524,230	6,513,254		6,744,520	3.55%