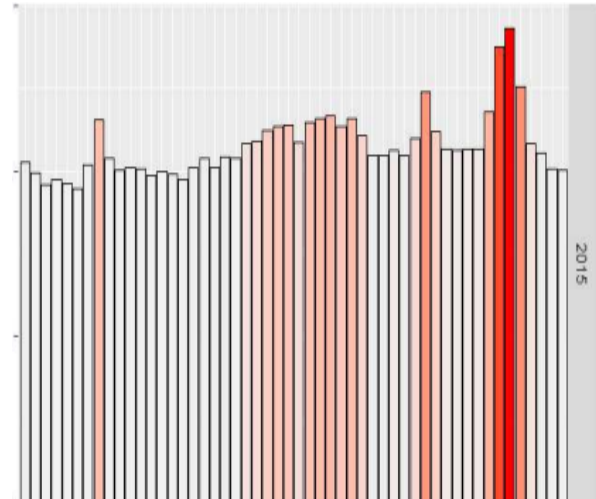
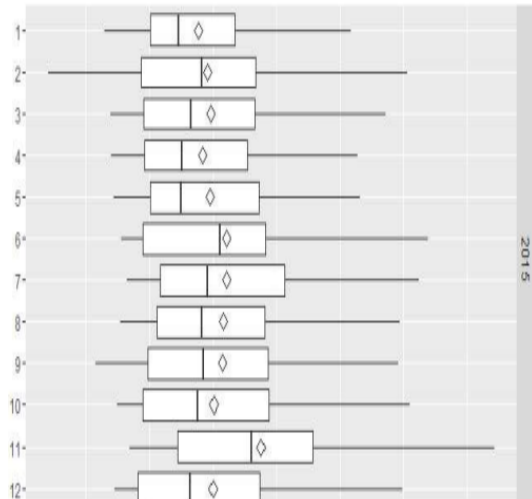


데이터를 통해 음식물쓰레기 발생량을 분석하고, 맞춤형 정책 제안

이경환

먼저 찾아본 논문을 읽어보았다. 요약해보자면 청주시의 월별 음식물 발생량은 7월달에 가장 많이



발생하였고 12월달에 가장 적은양을 보였다. 이를 주별로 분석한 그래프와 비교해보자면 주로 김장철이 있는 11월, 추석연휴 9월, 설날 연휴인 1월달에 발생량이 높다는 걸 알 수 있었다.

<월별 음식물 쓰레기 발생량>

<주별 음식물 쓰레기 발생량>

실제로 내가 해보기 위해 직접 공공데이터 포털에 들어가 음식물쓰레기 배출량 자료를 찾았다.

하지만 찾은 엑셀 자료 안에 모든 지역의 쓰레기 배출량이 다 나와있었기 때문에 데이터 가공을 하였다.

배출월	광역시도	기초지자체	배출량(톤)
7	서울특별시	강서구	193385
7	서울특별시	구로구	55209
7	서울특별시	금천구	372222
7	서울특별시	영등포구	987471
7	서울특별시	동작구	475039
7	서울특별시	서초구	399451
7	서울특별시	강동구	267455
7	서울특별시	송파구	1885677
7	부산광역시	중구	9596
7	부산광역시	서구	97104
7	부산광역시	동구	59401
7	부산광역시	영도구	187073
7	부산광역시	부산진구	708005
7	부산광역시	동래구	528949
7	부산광역시	남구	581596
7	부산광역시	북구	853434
7	부산광역시	해운대구	1248638
7	부산광역시	사하구	793504
7	부산광역시	금정구	460538
7	부산광역시	강서구	375034
7	부산광역시	연제구	438459
7	부산광역시	수영구	321783
7	부산광역시	사상구	675908
7	부산광역시	기장군	412302
7	대구광역시	중구	159911
7	대구광역시	동구	179954

1	year	month	city	town	ton
2	2017	7	GwangJu	DongGu	123848
3	2017	7	GwangJu	SeoGu	850090
4	2017	7	GwangJu	NamGu	511996
5	2017	7	GwangJu	BukGu	705765
6	2017	7	GwangJu	GwangSar	457718
7	2017	8	GwangJu	DongGu	114029
8	2017	8	GwangJu	SeoGu	766248
9	2017	8	GwangJu	NamGu	463874
10	2017	8	GwangJu	BukGu	658861
11	2017	8	GwangJu	GwangSar	412755
12	2017	9	GwangJu	DongGu	102998
13	2017	9	GwangJu	SeoGu	691837
14	2017	9	GwangJu	NamGu	414656
15	2017	9	GwangJu	BukGu	585064
16	2017	9	GwangJu	GwangSar	376982
17	2017	10	GwangJu	DongGu	118250
18	2017	10	GwangJu	SeoGu	812175
19	2017	10	GwangJu	NamGu	478270
20	2017	10	GwangJu	BukGu	685107

먼저 내가 원하는 지역만 남기고 다 삭제 시키고, 데이터 가공을 위해 한글자료를 영어로 바꾸었다.

<처음 가공 전 데이터>

<가공 후 데이터>

. 다음 파이썬을 통해 데이터분석과 추출을 진행하였다.

```

import pandas as pd
import matplotlib.pyplot as plt
from google.colab import files

df = pd.read_excel ("/content/drive/MyDrive/GwangJu.xlsx")

# 'year' 열이 2018인 데이터만 추출하기
df_2019 = df[df['year'] == 2019]

# 'town' 열이 'DongGu'인 데이터만 선택하기
df_donggu_2019 = df_2018[df_2019['town'] == 'DongGu']

# 'month'와 'ton' 데이터 선택하기
df_donggu_2018_month_ton = df_donggu_2019[['month', 'ton']]

# 막대 그래프 그리기
plt.bar(df_donggu_2018_month_ton['month'],
df_donggu_2018_month_ton['ton'])
plt.xlabel('Month')
plt.ylabel('Ton')
plt.title('DongGu Monthly Ton in 2019')
plt.ylim([50000,250000]) # y축 범위 설정
plt.ticklabel_format(axis='y', style='plain')
plt.show()

# 데이터 불러오기
df = pd.read_excel('/content/drive/MyDrive/GwangJu.xlsx')

# 2019년 9월 데이터가 빠져있는 경우
if df.loc[(df['year'] == 2019) & (df['month'] == 9)].empty:
    # 2017년과 2018년 9월 데이터 평균 구하기
    mean_2017_9 = df.loc[(df['year'] == 2017) & (df['month'] == 9),
'ton'].mean()
    mean_2018_9 = df.loc[(df['year'] == 2018) & (df['month'] == 9),
'ton'].mean()
    # 평균값으로 데이터 채우기
    df.loc[(df['year'] == 2019) & (df['month'] == 9), 'ton'] =
(mean_2017_9 + mean_2018_9) / 2

# 서구의 각 월별 막대그래프 그리기
df_seo = df.loc[df['town'] == 'SeoGu']
plt.bar(df_seo['month'], df_seo['ton'])

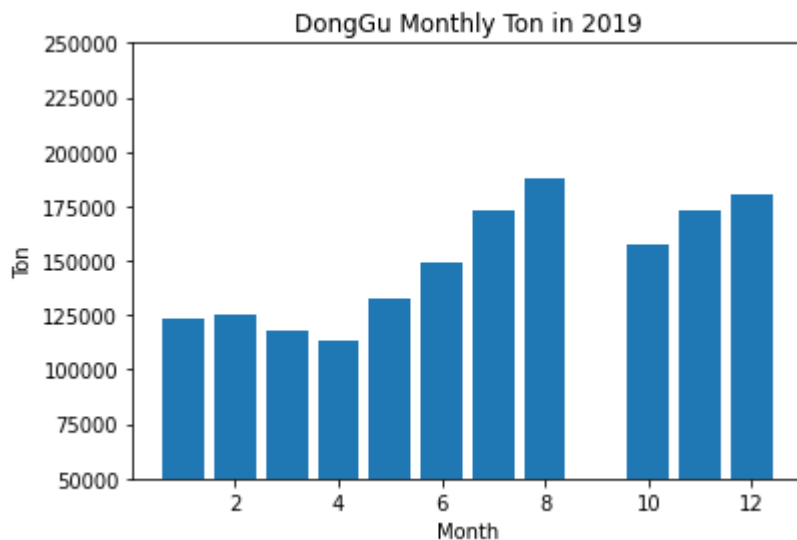
```

```
plt.xticks(range(1, 13))
plt.title('SeoGu Monthly Ton in 2019')
plt.ylim([500000,1300000]) # y축 범위 설정
plt.ticklabel_format(axis='y', style='plain')

# 그래프 출력
plt.show()
각 월별 막대그래프 그리기
df_gwangsan = df.loc[df['town'] == 'GwangSanGu']
plt.bar(df_gwangsan['month'], df_gwangsan['ton'])
plt.xticks(range(1, 13))
plt.title('GwangSanGu Monthly Ton in 2019')
plt.ylim([500000,1300000]) # y축 범위 설정
plt.ticklabel_format(axis='y', style='plain')
# 그래프 출력
plt.show()
```

문제점

나는 2019년의 데이터를 알고 싶었지만 준비한 자료에 9월의 배출량이 나와있지 않았다.



<9월달이 안 보이는 자료>

그래서 2017년 자료의 9월자료와 2018년의 9월 자료의 평균을 내서 그 값을 2019년 9월 자료로

사용했다.

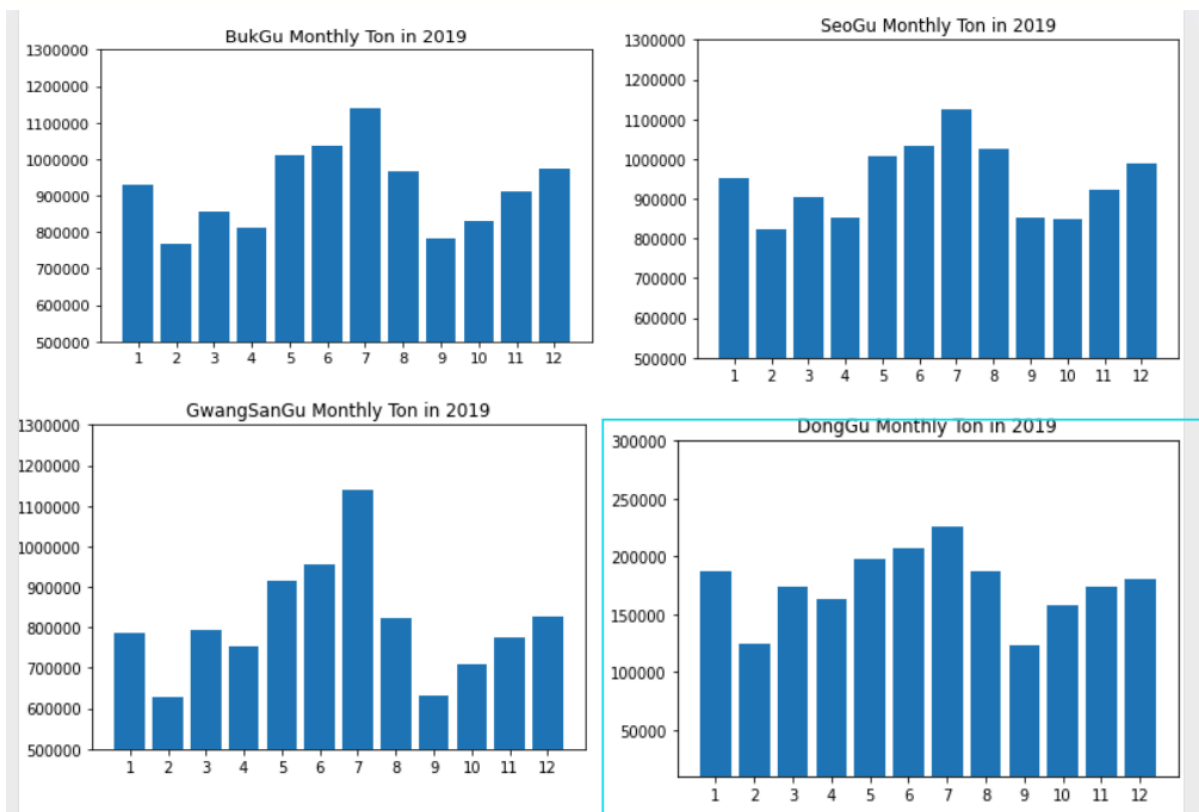
```
# 2019년 9월 데이터가 빠져있는 경우
if df.loc[(df['year'] == 2019) & (df['month'] == 9)].empty:
    # 2017년과 2018년 9월 데이터 평균 구하기
    mean_2017_9 = df.loc[(df['year'] == 2017) & (df['month'] == 9), 'ton'].mean()
    mean_2018_9 = df.loc[(df['year'] == 2018) & (df['month'] == 9), 'ton'].mean()
    # 평균값으로 데이터 채우기
    df.loc[(df['year'] == 2019) & (df['month'] == 9), 'ton'] = (mean_2017_9 + mean_2018_9)
/ 2
```

다음 구별 그래프를 그린다

```
# ?구의 각 월별 막대그래프 그리기
df_seo = df.loc[df['town'] == 'SeoGu']
plt.bar(df_seo['month'], df_seo['ton'])
plt.xticks(range(1, 13))
plt.title('SeoGu Monthly Ton in 2019')
plt.ylim([500000, 1300000]) # y축 범위 설정
plt.ticklabel_format(axis='y', style='plain')
```

그래프 출력

```
plt.show()
```



분석

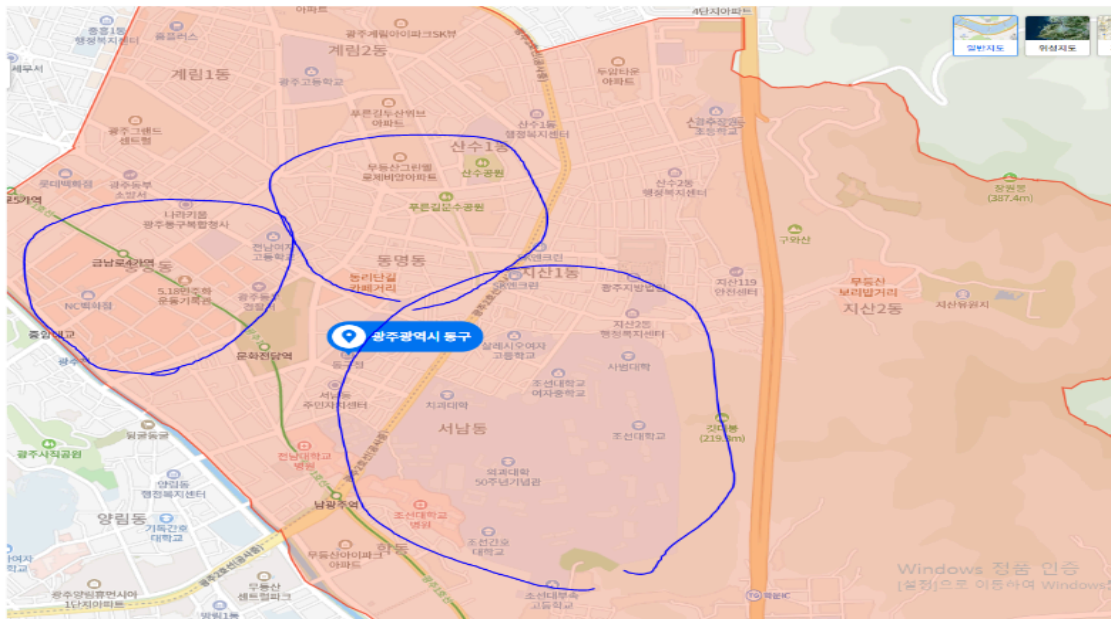
청주시는 추석과 설날이 있는 2월 9월에 음식물 쓰레기 배출량이 늘었지만, 광주는 2월 9월에 배출량이 줄었다. 나는 이것을 보고 유의미한 차이라고 생각해서 원인을 찾아보았다.

광주 광산구가 5개 “구”중에서 가장 큰 변화를 보였다.

광주광역시 검색 결과

#	행정구역	평균	남자	여자
1	광주광역시 동구	45.4	44.2	46.7
2	광주광역시 남구	44	42.7	45.2
3	광주광역시 북구	43.4	42.1	44.7
4	광주광역시 서구	43.1	41.8	44.4
5	광주광역시 광산구	39.7	38.8	40.5

다음은 광주 “구”별 평균 연령이다. 광산구가 가장 낮은 평균 연령을 보여줬기 때문에 명절에 부모님을 찾으러 가는게 일반적이기 때문에 2월과9월 사람이 다른 지역으로 빠져나가 음식물 쓰레기양이 주는데 영향을 줬을 수 있다고 생각한다. 동구는 평균연령이 가장 높은데도 큰 변화를 보여주었다. 그래서 나는 동구 지도를 찾아보았다.



동구에는 시내, 동명동, 조선대학교가 절반 이상을 차지했다. 이 주변에는 상업시설이 많이 있기때문에 사람들이 다른 지역으로 이동하면 상대적으로 장사가 잘 안 될 것이고, 그러면 음식물 쓰레기 배출량도 줄어들게 될거라고 생각했다.

충청북도 청주시 검색 결과

#	행정구역	평균	남자	여자
1	충청북도 청주시 상당구 낭성면	60.1	58.1	62.3
2	충청북도 청주시 상당구 미원면	60.1	57.8	62.4
3	충청북도 청주시 서원구 현도면	60.1	57.7	62.6
4	충청북도 청주시 청원구 북이면	58.7	55.6	62.2
5	충청북도 청주시 상당구 문의면	58.6	56	61.5
6	충청북도 청주시 상당구 가덕면	58.3	56.5	60.2
7	충청북도 청주시 상당구 성안동	51.4	49.7	53.2
8	충청북도 청주시 상당구 남일면	51	49.2	52.8
9	충청북도 청주시 청원구 내덕1동	50.9	48.9	52.9
10	충청북도 청주시 서원구 사직1동	50	49	51.1
11	충청북도 청주시 서원구 남이면	49.2	48.3	50.1
12	충청북도 청주시 서원구 수곡1동	48.6	46.6	50.8
13	충청북도 청주시 상당구 영운동	48	46.4	49.6
14	충청북도 청주시 청원구 우암동	47.7	46.2	49.3
15	충청북도 청주시 서원구 수곡2동	47.5	45.6	49.2
16	충청북도 청주시 상당구 탑대성동	47.4	45.7	49.1
17	충청북도 청주시 상당구 중앙동	46.7	45.1	48.3

다음으로 청주시와 비교해보았다. 청주시는 광주광역시에 비해 평균연령이 더 높았다. 그러므로 부모님들을 보러 가는 가족들이 다른 지역에서 많이 유입되면 음식물 쓰레기 발생량도 늘어날 수 있다고 생각하였다.

결론

2월과 9월에 광주광역시와 청주시가 같은 내용의 음식물쓰레기 배출 관련 홍보를 한다면

홍보인한 효과가 차이가 날 수 있다고 생각한다. 청주시는 쓰레기가 많이 배출 될 수 있는 2월

9월에 홍보를 한다면 큰 효과를 누릴 수 있지만, 광주광역시는 그러지 않을 것이다. 따라서

각 지자체마다 쓰레기가 많이 나오는 기간을 분석해 자신들에게 맞는 정책을 만들어야할 것이다.

느낀점

내가 원하는 자료를 찾는데 시간이 많이 걸렸다. 또 찾아서 이제 거의 다 된 줄 알았지만, 자료에 누락된

부분이 있어서, 그 부분을 어떤 방식으로 채워넣을지 많은 고민을 했다. 나는 예전 자료들을 보고 평균을 구해 채워 넣었지만, 좀 더 시간이 있었다면 각 년도 마다 배출량이 증가했는지 감소했는지 알아내서 추세에 맞게 값을 넣을 수 있었을 것이다.

직접 우리 지역이 언제 음식물 쓰레기가 많이 나오는지 직접 파이썬으로 그래프를 그리면서 확인 해보니 좋았다. 그래프를 보고 직접 무엇이 원인이 되어 많이 쓰레기가 배출되는지 찾는 과정에서 재미있었다.

.