

빅데이터 기반 2019 서울시 미세먼지 원인 상관관계 분석 연구

Relationship Analysis of 2019 Particulate Matter cause in Seoul using Big Data

최예준, 이승현, 이종한

Ye-June Choi, Seung-Hyun Lee, Jong-Han Lee

요약 최근까지 미세먼지와 초미세먼지로 인해 외부 활동이 자제되었다. 미세먼지의 주원인은 화석연료가 연소하거나 자동차의 배기가스 등 대기오염물질로 알려져 있다. 이를 통해 본 논문에서는 2019 서울시 지역의 미세먼지와 초미세먼지의 평균 값을 R로 분석 및 시각화하여 대기오염물질과 미세먼지가 상관관계가 있는지 또한 다른 요인이 미세먼지에 영향을 미치는지 알아보도록 한다. 그리고 미세먼지의 영향이 다소 적은 지역을 확인하여 그 원인에 대해서도 명시하고 추후 더욱 연구가 필요한 부분에 관하여는 논하기로 한다.

주요어 미세먼지, R, 대기오염, 영향, 상관관계

Abstract Until recently, outdoor activities were restricted due to particulate matter. The main cause of particulate matter is known as air pollutants such as fossil fuel combustion or automobile exhaust gas. Through this, this thesis analyzes and visualizes the average value of particulate matter in the 2019 Seoul area as R, to find out whether there is a correlation between air pollutants and particulate matter, and whether other factors affect particulate matter. In addition, the areas where the influences of particulate matter are somewhat less will be identified, the cause of the matter will also be specified, and further studies will be discussed later.

Keyword Particulate Matter, R, Air Pollutants, Influences, Correlation

I. 서론

매해 봄철마다 돌아오는 중국 발 황사 문제와 더불어 미세먼지 및 초미세먼지 경보, 호흡기 질환 급증, 시정장애 등과 같은 문제 발생으로 미세먼지에 대한 관심과 관련 연구 필요성이 대두되고 있다. 실생활에서는 최근까지 미세먼지로 인해 외부 활동이 자제되었으며 마스크 착용이 생활화 되고 있으며 미세먼지 및 초미세먼지로 인한 질병 및 질환으로 사람들의 피해가 증가 하고 있다.

미세먼지 및 초미세먼지의 원인등의 키워드를 네이버 검색어를 기준으로 크롤링하여 기사의 제목을 수집하여 R프로그래밍을 이용하여 Word Cloud를

이용한 미세먼지 원인에 대한 반응 및 데이터를 시각화 한다.

미세먼지 농도를 조사 하기 위해 서울시 25 개의 각 자치구에서 측정, 수집된 미세먼지와 초미세먼지를 서울 열린데이터광장 데이터를 기초로 수집, 저장, 처리, 분석 단계를 거쳐 R 프로그램을 이용하여 통계 분석과 다양한 데이터 조작 및 시각화를 한다.

미세먼지 및 초미세먼지의 원인이 되는 대기오염 및 교통량의 데이터를 수집하여 해당년도의 지역별 오염정도와 교통량 등을 비교하여 지역별 차이의 존재 유무를 파악한다.

II. 관련 기술

1. Crawling

크롤링(crawling) 혹은 스크레이핑(scraping)은 웹 페이지를 그대로 가져와서 거기서 데이터를 추출해 내는 행위다. 크롤링하는 소프트웨어는 크롤러(crawler)라고 부른다.

2. R Programing

1993년 뉴질랜드에 위치한 오클랜드 대학교에서 개발된 통계 및 그래프 작업을 위한 인터프리터 프로그래밍 언어이다. R이 가지는 특이사항 중 하나로, color를 colour라고 쓰는 것 또한 인정된다. 영국식 영어를 사용하는 뉴질랜드에서 개발돼서 얻은 특징이다.

3. Data Mining

데이터베이스 속에서 의미 있으면서 유용하고 가치 있는 정보를 찾아내는 인간행동이다. 학문적으로 엄밀하게는 이 인간행동 중에서도 부분적인 과정만을 일컫지만 일상생활에서는 모든 과정을 일컫는다. 이런 과정을 지식발굴과정(KDD)라 부른다.

4. Word Cloud

워드 클라우드 기법(Word Cloud)을 이용한 최근 소비자학 연구 트렌드 분석은 텍스트 시각화 기법들 중 대표적인 방법으로, 텍스트와 같은 비정형 데이터를 그래프(원형)의 형태로 시각화해준다.

워드 클라우드 분석의 주요한 특징은 단어의 출현 빈도에 따라 키워드를 시각화하여 보여준다.

III. 본 론

1. 워드클라우드 이용

미세먼지의 실시간 반응과 원인과 관련된 키워드를 사하기 위해 네이버 검색어 “미세먼지 원인” 키워드를 통해 지면 뉴스 Title를 기반으로 Word Cloud를 진행하여 주요 키워드를 조사했다.

연	매	려	회	제	개	제	등	원	공	고	신	영	지	정	영	이	변
94	81	14	13	12	8	7	6	6	6	5	5	5	5	5	5	4	4
서울	세	한	영	상	채	간	경	국	대	국	대	국	대	국	대	국	대
4	4	4	4	3	3	3	3	3	3	3	3	3	3	3	3	3	3

그림 1. 뉴스 주요 키워드 조사 분석 결과

Fig 1. Outcome of news keyword analysis

R 스튜디오를 사용하여 주요 키워드 분석 결과 뉴스 미세먼지 키워드 분석 결과 중국발, 자동차, 경유차, 배출, 석탄 같은 원인에 대한 키워드가 나타난 것을 볼 수 있다.



그림 2. Word Cloud를 통해 나온 이미지

Fig 2. Image from word Cloud

```
base_url <- "https://search.naver.com/search.naver?where=news&query=%EB%A%B8%
urls<-NULL

for(x in 0:9){
  urls<-c(urls,paste0(base_url,x*10+1,seq=''))
}
urls

new_links <- NULL

for(url in urls){
  html <- read_html(url)
  new_links <- c(new_links,html %>%
    html_nodes('.news_tit') %>%
    html_text()
  )
}
txt <- new_links
```

그림 3. 크롤링 코드 분석

Fig 3. Analysis of crawling code

사람들의 관심사를 분석하기 위해 “미세먼지 원인” 관련 지면 기사들을 크롤링 하여 페이지당 10건의 뉴스가 보여지도록 url<-c(url, paste0(base_url, x*10+1, seq="")) 명령어로 최대한 많은 양의 기사 title을 가져올 수 있도록 url을 설정하고 각 기사의 html_text()를 이용하여 title 제목만을 추출했다. 그 결과는 그림 4와 같다.

[illegible]

그림 4 뉴스 타이틀 크롤링 결과

Fig 4. Outcome of news title crawling

```
library(rvest)
library(tidyverse)
library(xml2)
library(wordcloud)
library(KoNLP)
library(RColorBrewer)
useJongDnC()

base_url <- "https://search.naver.com/search.naver?where=news&query=%EC9A%FB8E%CB4%B8%EB%A8%BCE%CA7%B8%20%E0%B9%B0%EC"
urls<-NULL
for(x in 0:9){
  urls<-c(urls,paste0(base_url,x*10+1,seq=''))
}

new_links <- NULL
for(url in urls){
  html <- read_html(url)
  new_links <- c(new_links,html %>%
    html_nodes('.news_tit') %>%
    html_text())
}

txt <- new_links
txt <- str_replace_all(txt, "\\W+", " ")
txt<-gsub("\\d+", "",txt)
txt<-gsub("\\n+", "",txt)
txt<-gsub("[A-z]", "",txt)
txt<-gsub("[:cntrl:]", "",txt)
nouns <- KoNLP::extractNoun(txt)
nouns <- sapply(txt,extractNoun,USE_NAMES=F)
nouns_unlist <- unlist(nouns)
wordcount <- Filter(function(x)(nchar(x)>1),nouns_unlist)
head(wordcount, 100)
write(wordcount,"1.txt")
rev <- read.table("1.txt")
rownames(rev)
wordcount <- table(rev)
head(sort(wordcount, decreasing=T),30)
palette <- brewer.pal(9,"Set1")
wordcloud(names(wordcount),freq=wordcount,scale=c(5,0.5),rot.per=0.25,min.freq=1,
  random.order=F,random.color=T,colors=palette, max.words = 100)
```

그림 5. 크롤링 후 전체 wordcloud code

Fig 5. Word Cloud code after crawling

뉴스 타이틀 크롤링 후 gsub 과 filter 사용으로
2글자이상 명사 추출 후 RColorBrewer 을 이용하여
word cloud 에 컬러감으로 시각화를 극대화 했다.

2. 서울시 구별 미세먼지 데이터 분석 및 시각화

1) 데이터 수집

본 논문에서 분석 하고자하는 서울시 지역별 미세먼지 데이터는 데이터셋(서울 데이터 열린 광장) 포털에서 조회 후 다운로드하여 수집하였다. 수집된 데이터는 2013 년에 공개된 데이터 이며 url 은 다음과 같다.

https://data.seoul.go.kr/dataList/OA-2220/F/1/datasetView.do;jsessionid=8A20A9F90045C1143FB406FD8ABAA73D.new_portal-svr-11,

<https://data.seoul.go.kr/dataList/10637/S/2/datasetView.do>

수집한 데이터는
기간별_일평균_대기환경_정보_2019 년 csv, 서울시
대기오염 통계.csv 파일을 이용하였다.

2) 데이터 처리

데이터 처리에는 R 스튜디오를 사용하였다. 2019
기간별 대기환경 데이터에는 측정일자, 권역코드, 권역명
측정소코드 및 이름, 대기환경 정보가 담겨있다.

그러나 본 논문에서 분석 할 데이터는 측정일자, 권역명, 측정소명, 미세먼지, 초미세먼지 이므로 이 외의 변수는 삭제한다.

측정일자	구역코드	구역명	측정소코드	측정소명	미세먼지(μg/m ³)	오존(ppm)	이산화질소	일산화탄소	이황화수소	조미세먼지(μg/m ³)
20191231	100	도심권	111121	종구	26	18	0.023	0.012	0.4	0.002
20191231	100	도심권	111123	종로구	27	19	0.021	0.015	0.4	0.003
20191231	100	도심권	111131	용산구	14	15	0.02	0.015	0.4	0.003
20191231	101	서북권	111181	은평구	24	13	0.024	0.01	0.4	0.003
20191231	101	서북권	111191	서대문구	23	9	0.023	0.013	0.5	0.002
20191231	101	서북권	111201	마포구	22	12	0.019	0.017	0.4	0.002
20191231	102	동북권	111141	광진구	22	11	0.021	0.015	0.4	0.002
20191231	102	동북권	111142	성동구	25	17	0.022	0.019	0.3	0.003
20191231	102	동북권	111151	중랑구	24	12	0.02	0.017	0.4	0.003
20191231	102	동북권	111152	동대문구	25	13	0.022	0.015	0.4	0.003
20191231	102	동북권	111161	성북구	26	12	0.025	0.016	0.6	0.003
20191231	102	동북권	111171	도봉구	19	12	0.029	0.01	0.3	0.002
20191231	102	동북권	111291	강북구	32	17	0.023	0.021	0.3	0.002
20191231	102	동북권	111311	노원구	22	17	0.022	0.013	0.4	0.004
20191231	103	서남권	111212	강서구	29	16	0.024	0.015	0.4	0.004
20191231	103	서남권	111221	구로구	24	14	0.027	0.014	0.2	0.003
20191231	103	서남권	111231	영등포구	22	14	0.02	0.016	0.5	0.002

그림 6. 2019 서울시 대기환경 csv 파일

Fig 6. 2019 atmosphere environment in Seoul csv file

2019 csv 파일에서 불필요한 변수는 삭제 한 후 R 스튜디오 read.csv 하여 데이터를 처리한 코드는 그림 7 과 같다.

[illegible]

그림 7. csv 파일 데이터 처리 코드

Fig 7. Processing code of csv file data

read.csv 함수를 사용하여 파일을 불러온 후 각 측정소명을 약어로 바꿔주었다.

이후 서울시 지역별 미세먼지와 초미세먼지의 년 평균을 구하기 위해 mean() 함수를 사용했으며 그래프로 시각화 하기 위해 data.frame 로 값을 저장해주었다.

3) 데이터 시각화

데이터 처리 이후 googleVis 라이브러리를 통해 2019 년도 서울시 지역별 미세먼지 평균을 height=400, weight=500 으로 막대그래프를 사용하여 시각화 하였다. 그림 8 과 같다.

가. Googlevis 를 사용한 막대그래프

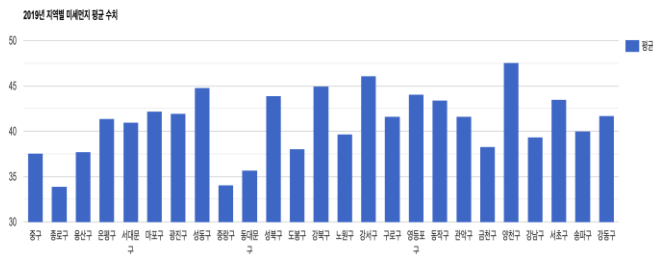


그림 8. GoogleVis 를 이용한 미세먼지 평균 그래프 시각화

Fig 8. Visualization of average using GoogleVis

막대 그리프 외에 서울시 지역을 지도로 시각화하여 알아보기 위하여 ggplot 라이브러리로 시각화하였다.

이 결과 2019 년도 미세먼지 농도 평균 양천구가 가장 높은 것을 알 수 있으며 이후 강서구 강북구 순으로 높았다. 반면 중랑구의 평균이 가장 낮았던 것을 알 수있다.

나. ggmap 을 사용한 지도 시각화

먼저 서울시 행정구역 시군구 정보를 알기위해 데이터셋에서 2014 년도에 공개된 csv 파일을 다운 받았다. 이 파일에서의 시군구코드와 시군구명들을 제외한 칼럼들은 삭제한 후 data frame 로 만들어진 평균의 값과 조인시켜 dustmean 의 파일명으로 csv 파일을 저장했다.

	측정소명	평균	id
1	중구	0.003402	11140
2	종로구	0.00323	11110
3	용산구	0.003725	11170
4	은평구	0.004376	11380
5	서대문구	0.003992	11410
6	마포구	0.003379	11440
7	광진구	0.003812	11215
8	성동구	0.003573	11200
9	중랑구	0.004941	11260
10	동대문구	0.004299	11230
11	성북구	0.003317	11290
12	도봉구	0.003812	11320
13	강북구	0.002758	11305
14	노원구	0.004489	11350
15	강서구	0.004719	11500
16	구로구	0.00459	11530
17	영등포구	0.003354	11560

그림 9. dustmean csv 파일

Fig 9. dustmean csv file

ggmap, ggplot 을 이용하여 시각화 한 코드는 다음과 같다.

```
library(ggmap)
library(ggplot2)
library(raster)
library(rgeos)
library(maptools)
library(rgdal)

P <- read.csv("dustmean.csv", header = TRUE) #시각화할 데이터셋
map <- shapefile("TL_SCCO_SIG.shp") #지리 정보 데이터셋
map <- spTransform(map, CRSobj = CRS("+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs"))
new_map2 <- fortify(map, region = "SIG_CD")
View(new_map2)
new_map2$id <- as.numeric(new_map2$id)
seoul_map <- new_map2[new_map2$id <= 11740,]
seoul_dust2 <- merge(seoul_map, P, by='id')
plot <- ggplot() + geom_polygon(data = seoul_dust2, aes(x=long, y=lat, group=group, fill = 평균))
plot + scale_fill_gradient(low = "#FFFFFF", high = "#FF9900", space = "Lab", guide = "colourbar")
+ theme_bw() + labs(title = "2019 서울시 초미세먼지 평균")
+ theme(panel.grid.major.x = element_blank(), panel.grid.minor.x = element_blank(),
        panel.grid.major.y = element_blank(), panel.grid.minor.y = element_blank(),
        plot.title = element_text(face = "bold", size = 15, hjust = 0.5))
```

그림 10. 지도 시각화 코드

Fig 10. map visualization code

Ggmap, ggplot2 등 맵 제공 환경에 맞는 라이브러리들을 사용한다. 시각화할 데이터셋인 dustmean 파일을 불러 온 후 지리 정보 데이터셋을 map 에 저장시킨다.

map 을 spTransform() 함수를 이용하여 좌표계 변환을 진행한 후 fortify() 함수를 이용하여 map 을 data frame 으로 변환한다. 서울시 구는 모두 코드 id 가 11740 이하 이므로 new_map2\$id <= 11740, 코드를 사용하여 id 수를 11740 이하만 추출하여 seoul_map 변수를 생성한다.

이후 merge()함수를 사용하여 id 를 key 조인으로 하여 merge 시켜 seoul_dust2 로 저장한다.

ggplot 함수를 사용하여 경도와 위도를 설정하여 구별 경계를 나타낸 후 ggplot 을 활용하여 평균에 따라 가장 높은 구는 진한 색상 낮은 구는 연한 색상을 사용하였다. 그 결과는 그림 11 과 같다.

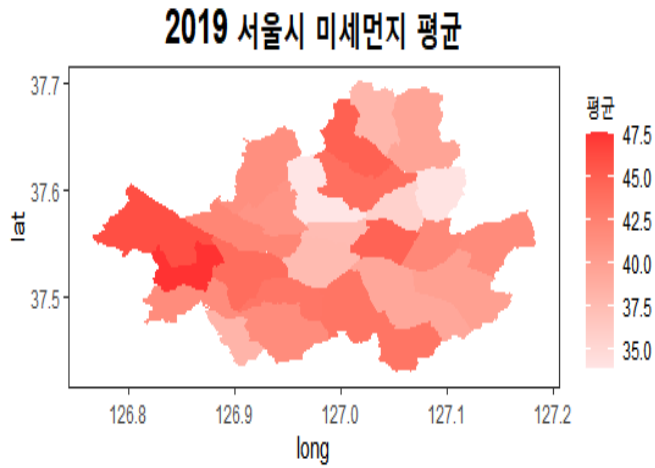


그림 11. 2019 지역별 서울시 미세먼지 평균

Fig 11. Average of particulate matter in seoul by regin in 2019

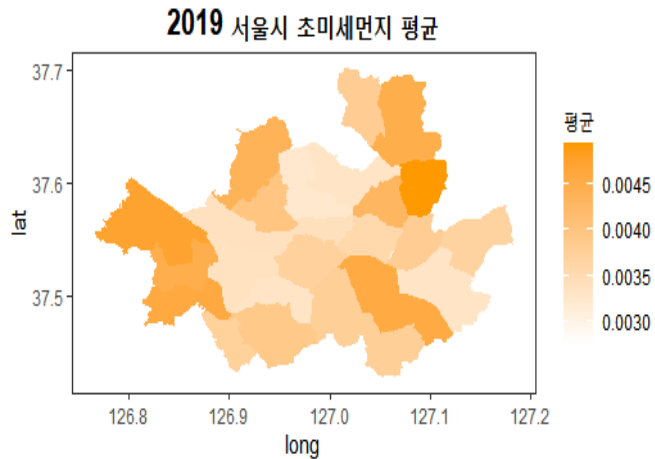


그림 12. 2019 지역별 서울시 초미세먼지 평균

Fig 12. Average of micro dust in seoul by regin in 2019

그림 11 과 12 를 보면 각 지역 마다 미세먼지와 초미세먼지의 평균 차이가 존재하는 것을 확인 할 수 있다. 그림 8 과 같이 양천구가 가장 높으며 중랑구가 낮음을 지도를 통해 시각적으로 알아볼 수 있다.

다. 2015~2019 서울시 지역 미세먼지 시각화

2015,2019					
장소	2015	2016	2017	2018	2019
종로구	50	46	40	35	38
중구	44	47	40	36	38
용산구	42	42	39	34	34
성동구	46	52	46	42	42
광진구	45	48	42	40	41
동대문구	44	49	44	36	48
중랑구	45	50	45	35	45
성북구	42	49	48	41	43
강북구	45	42	37	35	39
도봉구	43	46	44	41	38
노원구	46	41	42	39	44
은평구	45	47	44	40	45
서대문구	42	50	46	41	38
마포구	46	46	41	44	40
양천구	47	47	44	37	48
강서구	48	51	47	39	47
구로구	46	51	45	43	42
금천구	47	45	41	39	39
영등포구	48	55	49	47	45
동작구	44	43	42	40	44
관악구	51	48	43	43	49
서초구	47	50	47	42	44
강남구	46	51	44	35	40
송파구	44	46	45	42	40
강동구	44	53	48	44	42

그림 12. 2015~2019 년 지역별 서울시 미세먼지평균

Fig 12. Average of particulate matter in Seoul by region from 2015 to 2019

지역별 차이의 존재 유무를 보다 정확하게 확인하기 위해 2019 년 데이터 뿐만 아니라 2015~2019 서울시 지역별 년 평균 데이터를 이용했다.

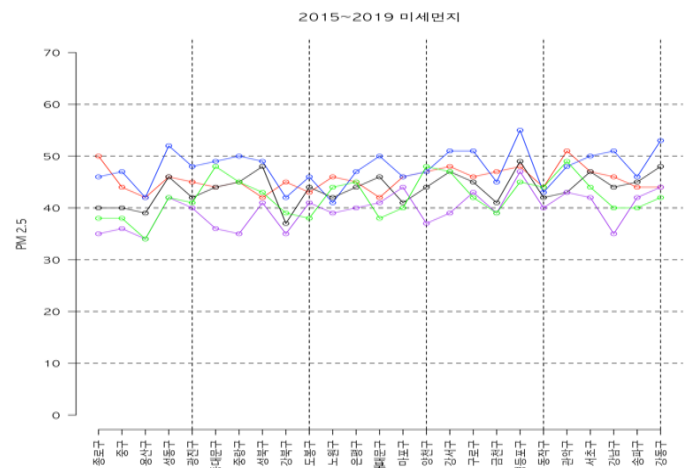


그림 13. 2015~2019 년 평균 그래프

Fig 13. 2015~2019 average graph

그림 13 를 보면 그래프의 형태가 굉장히 불규칙적으로 보이지만 특정 부분에서 유사한 부분을 발견 할 수 있다. 년도 별 특정지역의 수치가 매년 상위권인 것을 확인 할 수 있다.


```

R > for(x in range(1,3)){
+   bestbad1 <- subset(a,subset=(X2015==a2015[x]))
+   print(bestbad1$장소)
+ }
[1] "관악구"
[1] "강서구" "영등포구"
R > for(x in range(1,3)){
+   bestbad2 <- subset(a,subset=(X2016==a2016[x]))
+   print(bestbad2$장소)
+ }
[1] "영등포구"
[1] "성동구"
R > for(x in range(1,3)){
+   bestbad3 <- subset(a,subset=(X2017==a2017[x]))
+   print(bestbad3$장소)
+ }
[1] "영등포구"
[1] "성북구" "강동구"
R > for(x in range(1,3)){
+   bestbad4 <- subset(a,subset=(X2018==a2018[x]))
+   print(bestbad4$장소)
+ }
[1] "영등포구"
[1] "마포구" "강동구"
R > for(x in range(1,3)){
+   bestbad5 <- subset(a,subset=(X2019==a2019[x]))
+   print(bestbad5$장소)
+ }
[1] "관악구"
[1] "동대문구" "양천구"

```

그림 14. 2015~2019 년 년도별 평균 미세먼지 수치 상위권 지역

Fig 14. Top regions in 2015~2019 yearly average particulate matter

그림 13 의 시각화된 그래프 자료로 정확한 수치를 확인하기 어려워 보다 직관적인 장소를 확인하기 위해 재분석을 했다. 각 년도별 미세먼지 수치를 오름차순으로 정렬 후 상위권 3 개의 수치를 a201X[] 배열에 저장 A201X[] 값을 바탕으로 for 문을 이용하여 년도별 상위권 지역을 보다 명확하게 확인 했다.

2015 년 관악구,강서구,영등포구

2016 년 영등포구,성동구

2017 년 영등포구,성북구,강동구

2018 년 영등포구,마포구,강동구

2019 년 관악구,동대문구,양천구

영등포구 4 회, 관악구 2 회, 강동구 2 회,동대문구 1 회, 양천구 1 회, 마포구 1 회, 성동구 1 회, 강서구 1 회

해당 자료를 통해 영등포구가 2015~2019 년 동안 항상 상위권을 차지함을 알 수 있다.

3. 서울시 미세먼지 요인 데이터 분석

기간	자치구	구분	사업체수
2018	서울시	합계	823,385
2018	종로구	합계	39,952
2018	중구	합계	60,957
2018	용산구	합계	20,813
2018	성동구	합계	27,868
2018	관진구	합계	24,535
2018	동대문구	합계	31,878
2018	충랑구	합계	28,024
2018	성북구	합계	23,287
2018	강북구	합계	18,918
2018	도봉구	합계	18,894
2018	노원구	합계	26,616
2018	은평구	합계	24,873
2018	서대문구	합계	19,796
2018	마포구	합계	37,232
2018	양천구	합계	26,126
2018	강서구	합계	38,524
2018	구로구	합계	38,553
2018	금천구	합계	32,472
2018	영등포구	합계	42,589
2018	동작구	합계	19,916
2018	관악구	합계	26,105
2018	서초구	합계	46,540
2018	강남구	합계	71,373
2018	송파구	합계	47,772
2018	강동구	합계	29,772

그림 15. 2018 년 서울시 사업체 현황

Fig 15. Businesses in Seoul in 2018

미세먼지는 가스,석탄,석유 등 연소 시 발생하는 대기오염 물질이다. 그림 15 는 공장 및 가스, 석탄, 석유 등 연소가 필요한 사업체의 지역별 사업체 수를 가져온 것이다.

그림 14 의 내용을 바탕으로 미세먼지의 수치가 영등포구가 항상 상위권인 것을 고려하여 보면 주원인이 되는 사업체의 수는 서울시의 약 5%에 불과하며 사업체가 가장 높은 곳은 강남에 해당한다.



그림 16 영등포구 먹자골목 지도

Fig 16. Yeongdeungpo-gu Food Alley Map

2019 년 11 월 까지 영등포구의 미세먼지 측정 장소는 당산 1 동 주민센터 옥상이며 당산 1 동에는 먹자 골목이 형성 되어 있어 먹자 골목에서 나오는 각종 연기들이 미세먼지 측정의 정확성을 떨어트린다.

그로 인해 작년 11 월 우리은행 영등포구청점 건물 옥상으로 측정 위치를 이전하였다. 그 후 2020 년

1 분기 총 6 월까지의 미세먼지 수치가 하위권으로 조정되었다.

		아황산가스(ppm)	일산화탄소(ppm)	이산화질소(ppm)	오존(ppm)
2019	종로구	0.004	0.6	0.03	0.023
2019	중구	0.003	0.6	0.032	0.027
2019	용산구	0.003	0.5	0.032	0.023
2019	성동구	0.005	0.6	0.024	0.03
2019	광진구	0.004	0.8	0.022	0.027
2019	동대문구	0.004	0.5	0.027	0.025
2019	중랑구	0.004	0.5	0.032	0.027
2019	성북구	0.004	0.7	0.027	0.022
2019	강북구	0.005	0.5	0.029	0.027
2019	도봉구	0.006	0.5	0.024	0.025
2019	노원구	0.003	0.7	0.032	0.023
2019	은평구	0.003	0.5	0.021	0.028
2019	서대문구	0.004	0.5	0.019	0.029
2019	마포구	0.004	0.5	0.026	0.022
2019	양천구	0.004	0.5	0.032	0.024
2019	강서구	0.005	0.5	0.03	0.027
2019	구로구	0.005	0.4	0.025	0.028
2019	금천구	0.004	0.5	0.031	0.022
2019	영등포구	0.003	0.5	0.031	0.022
2019	동작구	0.003	0.5	0.03	0.024
2019	관악구	0.004	0.5	0.031	0.026
2019	서초구	0.004	0.4	0.03	0.027
2019	강남구	0.005	0.5	0.028	0.022
2019	송파구	0.003	0.5	0.033	0.023
2019	강동구	0.004	0.5	0.03	0.022

그림 17. 2019 년 서울시 대기오염 물질 수치

Fig17. 2019 air pollutant levels in Seoul

2019 년 미세먼지 수치가 높았던 양천구의 대기오염 물질을 살펴보면 아황산가스 0.005 일산화탄소 0.5 이산화질소 0.032 오존 0.022 으로 다른 지역에 비해 대기오염 물질의 수치가 높지 않은 것을 확인 할 수 있다.

또한 각 대기오염 물질 별 가장 높은 수치를 가지는 지역 또한 미세먼지 수치가 높지 않다.

IV. 결 론

미세먼지 지역의 차이를 확인하기 위해 R 프로그래밍을 이용하여 위와 같이 시각화, 분석을 진행하여 지역적 차이의 분명함 시각화 하였고 WordCloud 를 통한 키워드를 분석하여 미세먼지의 주요원인이 무엇인지 확인 할 수 있었다.

미세먼지의 주 원인이 되는 석유,석탄,가스 연소 시 발생하는 대기오염 물질, 그 물질을 발생하는 사업체를 통해 각 지역별 사업체 수를 조사하여 미세먼지 수치가 높은 지역을 비교하였지만 미세먼지 수치와는 큰 영향을 작용하지 않는 것으로 나타나고 다.

대기오염 물질의 수치를 파악하여 미세먼지와 상호 연관성을 조사하였고 그 결과 또한 미세먼지 수치와 크게 상관 없는 것으로 나타나고 있다.

미세먼지를 측정하는 측정소의 위치에 따라 미세먼지 수치의 정확성을 떨어트리는 외부 요인이 작용하며 그 외 미세먼지의 원인이 되는 교통량, 국외 요인 등 고려해야 될 사항이 많은 것으로 판단된다.

결론으로 서울시 내의 자치구의 차이점은 미세먼지 측정소의 위치로 인한 외부 영향으로 미세먼지 측정의 정확성을 떨어트리는 점과 국내에서 생성되는 미세먼지의 원인이 되는 정확한 요소를 확인하기 어렵다. 국내에서 발생하는 미세먼지의 요인보다 국외의 미세먼지의 발생 요인에 대해 더 고려해 볼 필요가 있다.

References

1. 데이터안심구역,
<https://dsz.kdata.or.kr/svc/data/search.do>
2. 데이터셋(서울열린데이터광장),
https://data.seoul.go.kr/dataList/OA-2220/F/1/datasetView.do;jsessionid=8A20A9F90045C1143FB406FD8ABAA73D.new_portal-svr-11
3. 서울특별시 대기환경정보,
<https://cleanair.seoul.go.kr/2020/>
4. 서울시 행정구역 시군구 정보
<https://data.seoul.go.kr/dataList/OA-11677/S/1/datasetView.do>
5. 대한민국 최신 행정구역 파일 다운로드
<http://www.gisdeveloper.co.kr/?p=2332>
6. 에어코리아
https://www.airkorea.or.kr/web/detailViewDown?pMENU_NO=12