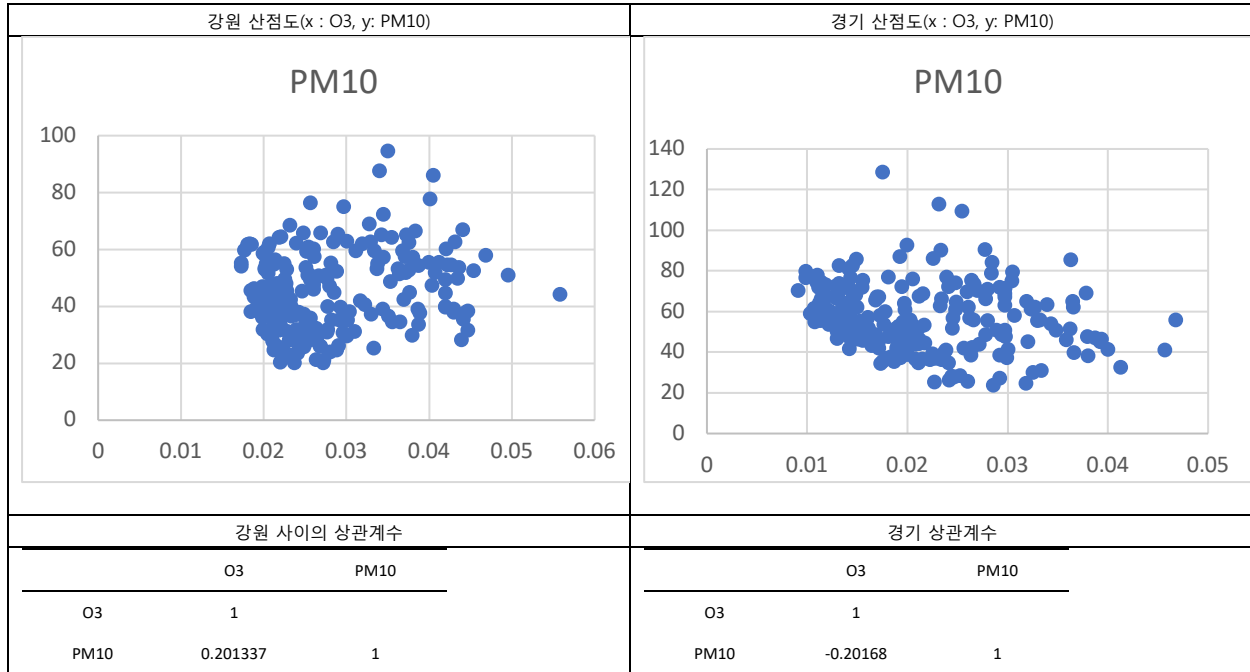


- 모든 데이터의 차트 중에 유일하게 오존(O3)만이 '상승 추세'를 가지고 있었다.
- 미세먼지나 대기중의 화학물질이 많은데 이제서야 호흡기 질환이 문제가 되는 이유가 뭘까 고민해보았더니, 그 답이 '오존'에 있을 수 도 있다는 가설을 세우게 되었다.
- 그렇다면 오존과 미세먼지간의 상관계수를 분석하고 차후에 '호흡기 질환자의 수 x 오존의 농도' / '호흡기 질환자의 수 x 대기 오염물질의 농도'를 분석한다면 실제로 인간에게 영향을 미치는 종속변수를 파악하려 하였다.

2. 문제 해결 과정

- 지역 측정소 월별 평균 데이터의 '오존', 'PM10'데이터의 상관계수를 산점도로 표현
- 산점도를 확인하니, 강원은 약한 양의 상관계수를 갖고, 경기 지역은 약한 음의 상관계수를 가짐



3. 결론

- 오존은 즉각적으로 미세먼지에 영향을 미치지 않음을 상관계수로서 확인할 수 있었다. 하지만 오존이 지구온난화를 일으켜 더욱 거시적인 영향을 미침을 연구한 여러 연구결과를 자료탐색과정에서 찾을 수 있었다.

4. 아쉬웠던 점

- 이번에 빅데이터 분석과 분석 툴인 '주피터 노트북', '판다스 라이브러리' 등을 모두 처음 접하며, 전처리에만 할애된 시간의 대부분을 사용하느라, 다른 데이터셋과 합쳐 심층적인 딥 러닝 모델을 사용하지 못하여 너무나 아쉬웠다. 하지만 데이터를 처리하는 과정에서 처음 사용해보는 여러 분석 툴들을 다룰 줄 알게 되었고, 기존에 경영학도로서 가지고 있었던 통계학적 지식을 여러 분야에 사용할 수 있다는 인사이트를 얻었다.
- 한가지 예시로는, 현재 쇼핑물을 운영 중인데, 크롤링과 시각화를 통하여 특정 아이템의 신빙성 있는 중고가 가격을 측정할 수 있는 모델을 구상해 보았다. 예를 들어, 크롤링을 하는 도중에 '파손', '부품용'과 같은 단어가 제목이나 내용에 포함되어 있으면 해당 값은 불용치 처리하여 평균으로 처리하여 평균 중고 값을 계산한다는 등의 인사이트이다.
- 오존과 미세먼지의 상관계수 분석 뿐만 아니라, 오존과 호흡기 질환 자 수와 기타 대기오염물질과 호흡기 질환 자 수의 관계도 추가적으로 분석 하였어야 했는데, 아직까지 능력의 부족으로 데이터 셋을 다룰 수 없었고, 통계학적 지식이 부족하여 더욱 더 공부해야 할 필요성을 느꼈다.₩

사용한 데이터 셋
(에어코리아 국내 대기데이터 2001~2021 년) https://www.airkorea.or.kr/web/last_amb_hour_data?pMENU_NO=123