1. 문제설정 및 목적

최근 미세먼지로 인해서 야외활동에 많은 제한이 되고 있고 이와 관련한 사회문제가 대두되어지고 있다. 이와 관련하여서 사람들의 소비패턴에도 영향을 미칠 것이라고 예측되어진다. 이를 마크로밀 엠브레인의 리서치 데이터를 바탕으로 파악해보고자 한다.

첫번째로 외식문화에 영향을 미칠 것으로 예상된다. 고객들의 야외활동이 제한됨에 따라서 고객들이 외식을 하기보다는 배달음식을 많이 선호할 것이라고 예측되고, 이를 고객들의 결제정보를 통해서 확인해 보겠다.

두번째로 고객들의 쇼핑행태에서 오프라인매장을 통한 쇼핑보다는 온라인 쇼핑몰의 결재량이 늘어날 것으로 예측된다.

2. 전처리

먼저 미세먼지와 관련된 데이터를 기상청등에서 공개된 공공정보를 활용하여 지역별 일자별 미세먼지의 영향을 크롤링등을 활용하여 수집하겠다.

결재정보 테이블을 활용하여 고객들의 결재량과 결재금액을 사용하여 고객들의 소비패턴을 파악하겠다. 먼저 배달음식결재와 외식업체 결재금액혹은 결재 횟수를 비교하고, 두번째로 오프라인 쇼핑몰과 인터넷 쇼핑몰에 있어서도 같은 방식으로 비교를 진행하겠다.

3. 모델링

우선 고객 개개인에 대한 모델을 세우는 데는 한계가 존재하므로 비슷한 행동양식을 갖는 고객들을 클러스터링 하겠다. 이를 바탕으로 각 클러스터별로 미세먼지 수준에 따른 고객들의 소비패턴에 영향이 있는지 분산분석등의 방법을 통해 검증해 보겠다.

추가적으로 일정 미세먼지수준 하에서 고객들의 소비량을 예측하고 이를 바탕으로 온라인/오프라인 매장에서 고객의 수요를 예측하는데 활용 할 수 있는 모델을 tree기반의 random forest, xgboost등의 모델을 활용하여 설립하도록 하겠다.

4. 기대효과

미세먼지로 인한 고객 소비패턴의 유의미한 차이가 발생하는 것을 확인 할 수 있다면, 현재 미세먼지의 효과가 단순한 국민들의 건강문제 뿐만이 아닌 국가경제에도 영향을 줄 수 있다는 사실을 제안 할 수 있고 이를 바탕으로 국가정책 설립에 있어서 영향을 줄 수 있을 것이다.

또한 미세먼지 수준별 고객들의 소비패턴을 예측가능 하다면 업체들의 영업패턴에 적용을 통해 수익성을 극대화 시킬 수 있을 것으로 예상된다. 예를 들어 미세먼지가 적은 날 외식업체들이 평소보다 많은 고객들이 방문할 것을 예측하고 영업준비를 하는 등의 효과를 줄 수 있을 것이다.