경제안정지수 생각

모든 기준은 다 동 별로

지수에 필요한 변수들: 동 별 카드소비 금액 증감률, 동 별 유동인구 수 증감률, 동 별 배달 전환 가능 여부, 사회적 거리두기 단계

변수생성

* **동 별** 카드 소비 금액 증감률, 물류 배송건수 증감률, 유통 비중 증감률

*비슷한 수식으로 물류, 유통 업종 별 점수 산출*

*유통 데이터 => 전체에서 해당 산업군에서의 위치/비중을 나타냄 이것을 card 점수에 고려-> 산업 대비의 의미 -> 업종이 별로 안 나와있음 -> 코로나 이후에도 비율은 거의 다 똑같음, 시장 규모가 변한 것은 아님 0.23*

* 논의 이슈
* 1. **월 별**로 지수를 제시해줄 것인지**, 주차별**로 제시할 것인지, **매일** 제시할 것인지 고려 필요 (또한, 평균값으로 산출할지 중앙값으로 산출할지 통계량 정해야 함)  
  주차 별이면 올해 3월 첫째 주 해당업종 카드소비액 … 이런 식  
  요일도 영향을 미칠 수 있음!  
  -> 최종 지수를 월별, **주차 별**, 매일, 요일 별 제공할지 결정/바이러스가 있을 경우에 어떻게 될 것인지  
  -> 해당 식의 *대조군*이 주차 별, +(주중/주말) 별 결정 -> 생각!!
* 2. 증감률을 그대로 지수에 반영할 것인지, ~~구간화를 할 것인지~~  
  ~~구간화 예시: {-50이하: 매우 불황, -50~-30: 불황… 1%이상: 호황}~~
* 단순 카드소비액이 아닌 업종별 선호도를
* 동 별 유동인구 *증감률로 할지 그냥 인구수로 할지 고민 필요* -> 유동인구가 골목 상권에 크게 좌지우지할 것이므로
* 논의 이슈
* 1. 위의 1번과 비슷한 문제
* 2. 나이 대/성별 유동인구, 시간대 별 유동인구 -> 어떻게 합칠 것인지(평균 또는 중앙값으로 합치긴 해야 함) 나이 & 시간대 둘 다 구해서 수식 합치는 방법도 있음(예: float\_age, float\_time을 수식적으로 합함. 단순 2로 나눠서 더하거나…) 그러나 나이대 별, 시간 별로 준 의미가 잘 드러나지 않을 수 있음 -> 단순 유동인구  
  나이/시간대 고려안하고 싹 더해서 그 주의 유동인구 구할 수도 있음.  
  아니면 합산 유동인구로 회귀분석을 수행해 영향력을 파악한 후 시간대별/성별 실시간 골목 상권 지수를 제안할 수도 있음(기대효과 부분과 밀접)
* 3. 업종 별 유동인구 – 불가능
* 동 별 배달 전환 가능 여부
* 업종 별 배달 전환이 가능한지 1,0으로 표시. 그리고 해당 동의 비율을 평균 구하면 0~1사이의 확률값이 나옴.
* 물류 데이터의 경우, 골목 상권과는 좀 거리가 있어 보임.
* 사회적 거리두기 단계
* 해당 시기의 거리두기 단계 시행을 숫자로 표시
* 3/22일부터 본격적인 거리두기가 시행된 것을 알 수 있음
* ***사회적 거리두기를 통해서 1,2,3단계 간 매출 및 유동인구 수에 차이가 있는지 검정 필요(anova 분석)***

경제 지수 (아주 러프) -> 더 많이 고민해 봐야 함(가중치를 두거나 로그를 씌우거나…)

가중치들은 eda나 분석을 통해 도출!

<예상 테이블> (동, 32개 업종)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 날짜 | 구 | 동 | 업종 | 작년 동일 대비 카드소비 매출액 증감률 | 물류 증감률 | ~~유통비중 증감률~~ | 유동인구 수 | 배달 전환여부 | 사회적 거리두기 단계 |
| 2020.04.01 | 중구 | 혜화동 | 10 | -22.3% | +1.03% | +3.5% | -14% | 0 | 2 |
| 2020.04.01 | 중구 | 혜화동 | 20 | -2.3% |  | -18% | -14% | 0 | 1 |
| 2020.04.01 | 중구 | 혜화동 | 21 | +1.5% | -20% |  | -14% | 1 | 0 |