

[0110]_EDA_방법론_이원준 과제

경기도 지역화폐 발행/이용 내역과 골목상권의 시장규모 및 매출 추이의 상관관계 분석.

1. 데이터 수집

- 지역화폐 발행 지역선정.

서울시에서 발행하는 서울사랑상품권과 경기도 지역별 지역화폐 데이터를 비교해봤을 때 서울시에서 발행하는 화폐에 대한 발행액, 이용금액 데이터셋이 없으므로 경기도 지역으로 선정했습니다.

- 지역화폐 데이터

경기데이터드림에 있는 '지역화폐 발행액, 이용금액' 데이터 사용

- 시장규모 및 매출 추이 데이터

경기 지역의 상권별로 최근 21-4분기의 매출 데이터는 있지만 과거 데이터가 없어서 매출 추이를 파악하기 어려움. 시장규모의 추이를 파악하기 위해서 경기도시장상권진흥원에서 제공하는 지역/상권별 점포수 현황을 크롤링해서 사용하기로 했습니다.

2. 데이터 전처리

- 지역화폐 데이터

카드형, 모바일형, 지류상품권형 등 3종류의 지역화폐 발행액이 있어서 합산했습니다. 또한, 점포수의 변화 데이터가 분기별로 나뉘어져있어서 월별로 나타난 데이터를 분기별로 합산했습니다.

- 지역별 골목상권 점포수 현황 데이터

시군구 중 구 단위로 나뉘어져있는 행은 drop 했습니다.

지역화폐 데이터와 merge 할 때 분기별/지역별로 merge하기 위해서 column으로 들어가있는 분기를 stack 후 reset_index를 통해 하나의 column으로 만들었습니다. 그 후 merge 시켜 최종적인 분기별/지역별 지역화폐 발행액, 점포수 데이터셋을 만들었습니다.

3. 데이터 Scaling

점포수와 발행액 단위에 차이가 있어 minmaxscaling을 진행했습니다.

4. 데이터 시각화 및 상관관계 분석

Scatterplot을 활용해 지역별로 점포수 와 지역화폐 발행액의 추이를 시각화했습니다. 그 후 pearson 상관계수를 분석하였습니다.

5. 한계점 및 개선방안

데이터 분석 전 예상했던 방향은 상관관계가 없다라고 예상을 하고 시작했으나 많은 지역에서 높은 상관관계를 나타내고 있었습니다. 다만, 안양시, 광명시, 과천시, 의왕시에서

만 음의 상관관계를 보이는 것을 알 수 있었습니다. 개선방안은

- 시장규모 및 매출의 추이를 점포수의 추이로 분석한 것이 타당한지 검토
 - 특정 지역에서만 상관관계가 없거나 음의 상관관계를 보이는 이유에 대한 검토
- 등이 있을 것 같습니다.