

10주차

피드백

상권 단위에 대한 데이터를 봤음 좋겠다?-> 서울시에서 어떤 기준으로 나뉘든 다음주에 발표하면 좋을듯.

서울시의 상권 개수 -> 정확히 몇개인지는 기억나지 않지만 꽤 많았음.

상권 범위? -> 포함되지 않는 부분은 없는 것으로 확인됐다.

상권과 지하철역을 어떻게 묶을지..?

음식점 카테고리 8개가 다인가? -> 아니잖아!!!!!! 10개인데 청과물 이런거 포함할지 고민중이다.

유동인구와 매출의 연관성으로 무엇을 이야기할 수 있겠냐?

데이터를 학습시켜서 클러스터링 할 수 있을지....? -> 생각만해도 쯔니 어려워보임;

좀 더 구체적인 데이터가 필요할듯.....

유동인구에 대한 지수를 뽑아내고 음식점 종류를 같이 봐서 한 곳이 아니라 여러 개를 보기...

상세하게... 해볼 필요가 있을 것.....

머신러닝....? 이거는 절대 못할거같은데 어떡하지

유동인구와 업종별 상권 분석을 먼저 집중적으로 해봐라....

해야할 일

1. 상권에 포함되는 지하철역/버스정류장 데이터 만들기 - mapping
2. 방향성? 생각해오기. -> 최종 목표는 이런 쪽으로 가면 좋을 것 같다 최소 1가지 이상
3. 기존 카테고리(10가지)에 소상공인시장진흥공단의 소분류 카테고리 이용할 수 있을지 생각해오기.

사용자가 분석하고 싶은 업종을 입력해
그 업종이 잘 되는 대표 상권을 보여주고
그 상권들의 특성 알려주기?

업종 선택 - 10가지 중에
상권 선택 . 희망위치를 선택

<https://www.data.go.kr/data/15067631/fileData.do#layer-api-guide>

소상공인시장진흥공단

우리가 지금 하려고 하는 분석

EDA (Exploratory Data Analysis: 탐색적 데이터 분석)

유동인구와 업종별 상관 분석을 먼저 집중적으로 할 것

아무래도.. 우리 진짜 아무것도 모르고 분석하겠다고 한거같아서 현타옴..

인프런에서 듣고있는 강의도 너무 바탕지식이 없는 상태에서 들어봤자 소용이 없을거같은
어쩌면 1학기 + 방학동안 계속 이 분석을 진행해야할거같은

참고 자료들

<https://dacon.io/codeshare/4899>

EDA (탐색적 데이터 분석) 사용 설명서 (1) - EDA & 통계치 분석

1. 문제 정의	2. 데이터 수집	3. 데이터 전처리	4. 데이터 모델링	5. 시각화 및 탐색
<ul style="list-style-type: none">- 분석 대상의 이해- 객관적이고 구체적으로 분석 대상 정의	<ul style="list-style-type: none">- 필요한 데이터 요건 정의- 데이터 소재 파악 및 확보	<ul style="list-style-type: none">- 오류 사항 점검 및 조치- 데이터 구조 및 특성 변경	<ul style="list-style-type: none">- 다양한 관점을 반영한 데이터 설계- 관련 테이블간 관계 설정	<ul style="list-style-type: none">- 다양한 유형의 데이터 시각화- 문제 해결을 위한 인사이트 도출

<https://wikidocs.net/book/3558>

개발자가 데이터 분석 준전문가 되기

http://contents.kocw.or.kr/document/RISS/Silla/ppt1_eda.pdf

EDA의 4가지 주제

<https://koreascience.kr/article/JAKO201831960581440.pdf>

탐색적 자료 분석(EDA) 기법을 활용한 국내 11개 대표 온라인 쇼핑몰 BEST 100 비교

우리 주제가 뭔가 작품이 나오기보다 이것처럼 논문을 써야할거같은 주제라는 생각이 듬..

ㄹㅇ 의식의 흐름이니까 무시해

최종적으로 뭘 할지.. 목표를 정해보자

원래 하려고 했던 것

- 사용자에게 희망 창업 업종, 지역을 입력 받아 그 지역에서 해당 업종의 Top 10 가게 주변을 분석해 이를 토대로 위치를 추천해주는 것

새로 생각해본 것

- 사용자가 업종, 지역을 입력하면 그 지역에서 해당 업종의 Top 10 가게 주변을 분석해서 공통적인 특성을 알려주는 것. 서울시 상권분석 사이트를 참조. 차별점은 Top10 가게를 위주로 분석했다는 것. 원래 하려고 했던 것에서 위치 추천을 제외한것.
- 사용자가 업종, 상권을 선택하게 하여서(그러면 업종의 수가 좀 부족하더라도 업종&상권이기 때문에 선택지가 매우 많아지는것이 아닐까) 해당 상권에서 해당 업종이 갖는 유리한 부분, 불리한 부분을 나눠서 알려주고 점수를 매겨서 보여주기
=> 이 방법이 우리가 분석한 걸 더 잘 사용할 수 있을 거 같긴 함

상권에 포함되는 지하철역/버스정류장 데이터 만들기 - mapping

이거 너무 어려운데

상권 안에 있는 지하철역, 버스정류장만 할 것인가??

우선 상권 좌표가 어디쯤 찍히는지 알아야 할듯

1. -사용자가 알고 싶은 업종을 입력하면(창업하고 싶은) 이 업종의 서울시 top 10을 뽑아서 그 top 10의 상권 분석? -> top 10이 한 상권에 전부 있는 경우 그~~렇게까지 유의미한 데이터는 아닐듯.
2. 사용자가 업종과 위치 입력 -> 그 업종의 특성과 위치(상권)의 특성 분석 -> 둘이 잘 맞는지? 그 상권에서 그 업종을 창업하는 게 적합한지?

<https://theorydb.github.io/dev/2019/08/25/dev-ml-insight/#%EA%B7%B8%EB%A6%AC%EA%B3%A0-%ED%94%84%EB%A1%9C%EA%B7%B8%EB%9E%98%EB%A8%B8-vs-%EC%97%B0%EA%B5%AC%EC%9E%90>

-> 한번 읽어보면 좋을

먼저 인사이트를 도출해야하는데...

강의를 듣기보다 데이터분석 관련 책을 읽는 게 낫다는 생각이 들었음.

scikit-learn

- 분류 : 주어진 데이터가 어느 클래스에 속하는지 판별
- 회귀 : 전달된 데이터를 바탕으로 값을 예상.
- 클러스터링 : 전달된 데이터를 어떤 규칙에 따라 나누는 것.

< 전체적인 틀 재설정 >

- 수빈: 사용자가 업종, 상권을 선택하게 하여서(그러면 업종의 수가 좀 부족하더라도 업종&상권이기 때문에 선택지가 매우 많아지는것이 아닐까) 해당 상권에서 해당 업종이 갖는 유리한 부분, 불리한 부분을 나눠서 알려주고 점수를 매겨서 보여주기
- 나영: 사용자가 업종과 위치 입력 -> 그 업종의 특성과 위치(상권)의 특성 분석 -> 둘이 잘 맞는지? 그 상권에서 그 업종을 창업하는 게 적합한지?

1. 사용자가 업종, 상권 두 가지 데이터를 선택.

업종 탭

ex) 한식음식점

-> 어떤 데이터 필요? 음식분류 카테고리

지역 탭

ex) 사직동

사용자가 두 가지를 선택한 후 => 분석

그 지역에 포함되는 모든 상권을 분석해서 각 상권의 특성, 정보 같은 것들을 다 보여주고 (시간대별 유동인구, 집객시설 수, 주변 상주인구, 전체 매출에서 한식음식점이 차지하는 비율, 전체 가게(10개 업종)중에서 한식음식점 개수가 차지하는 비율, 한식음식점의 추정 매출 평균&중위값,

한식음식점의 적합도에 대해서는 상권마다 점수로 매겨서 보여주기

시간대별 유동인구 & 시간대별 매출

ex) 유동인구가 아침과 밤에 집중되어있는데, 한식음식점은 점심 저녁의 유동인구가 중요하다면 점수가 깎임

-> 어떤 데이터 필요? 상권데이터(행정동코드 있는),

2-1 업종의 특성을 분석 -> 업종이 있는 상권의 특성 분석

2-2 상권을 분석 -> 상권 특성 분석

3. 2-1과 2-2를 합쳐서 이 업종이 상권(2-2)에서 창업되는 게 적절한지? 도출

- 상권 단위 분석이기 때문에 지역에 갇혀있는 것은 아님!

< 용어 정리 >

<https://golmokpolicy.seoul.go.kr/introduce.do>

상권 : 고객들을 대상으로 점포(상품 및 서비스)의 세력이 미치는 범위.

상권 배후지 : 어떤 상업지역의 상업시설(판매시설)을 주로 이용하는 고객들이 거주하는 지역

- '상권배후지'라는 개념은 골목상권에만 존재하는 개념으로, 골목상권으로부터 반경 200m의 영역입니다. 가운데가 뚫린 도넛 모양을 연상하시면 이해가 쉬우리라 생각됩니다.(가운데 뚫린 부분은 골목상권) 골목 상권보다 상권배후지의 영역 자체가 더 넓기때문에 집계되어 열린데이터 광장에서 보여지는 데이터 양이 많을 수 있습니다.

상주인구 : 한 지역에 주소를 두고 늘 거주하는 인구

생활인구 : 상주인구의 한계를 극복하기 위해 등장한 개념 - 주소만 두는 사람을 제외함

유동인구 : 일정한 기간 동안 한 지역을 오가는 사람

골목점포의 사전적 정의

1. 생활밀접업종을 포함한 점포
2. 발달상권에 포함되지 않는 점포
3. 배후지가 주거밀집 지역에 포함되는 점포
4. 전통시장에 포함되지 않는 점포
5. 길에 위치한 점포

골목상권의 사전적 정의

1. 일정 점포 수 이상의 상권
2. 골목 점포의 밀집도가 높은 상권

골목상권의 개념적 정의

1. 음식점업, 소매업, 서비스업을 영위
2. 도보를 통해 이동할 수 있는 거리
3. 주거지역이 밀집된 곳에 형성된 상권
4. 대형 유통시설을 상권 영역에 포함된 점포 제외
5. 대로, 로를 제외한 길 단위로 구성
6. 30개 이상 점포수를 포함한 골목상권 지역

발달상권의 정의

유통산업발전법 제5조의 법조항에 따라 2천 제곱미터 이내 50개 이상의 상점이 분포하는 경우 “상점가”라 하고, 배후지를 고려하지 않은 도보이동이 가능한 범위내의 상가업소밀집지역을 발달상권으로 정의

(도매·소매·음식·숙박·생활서비스·금융·부동산·학문·교육·의료복지·문화예술종교·관광여가오락 등의 8개 업종대분류 점포가 밀집한 지구로 정의)

전통시장 상권의 정의

오랜 기간에 걸쳐 일정한 지역에서 자연발생적으로 형성된 상설시장이나 정기시장

관광특구 상권의 정의

관광활동이 주로 이루어지는 지역적 공간 內 입지한 상권

<https://data.seoul.go.kr/dataList/OA-15580/S/1/datasetView.do>

서울시 우리마을가게 상권분석서비스(상권-집객시설)

<https://data.seoul.go.kr/dataList/OA-15568/S/1/datasetView.do>

서울시 우리마을가게 상권분석서비스(상권-생활인구)

<https://data.seoul.go.kr/dataList/OA-15572/S/1/datasetView.do#AXexec>

- 상권-추정매출

<https://data.seoul.go.kr/dataList/OA-15584/S/1/datasetView.do>

서울시 우리마을가게 상권분석서비스(상권-상주인구)

<https://data.seoul.go.kr/dataList/OA-15576/S/1/datasetView.do>

서울시 우리마을가게 상권분석서비스(상권-상권변화지표) -> 단순 정보 제공용

<https://data.seoul.go.kr/dataList/OA-12252/S/1/datasetView.do>

- 지하철 호선별 역별 시간대별 승하차 인원 정보 (시간대별)

<https://data.seoul.go.kr/dataList/OA-12914/S/1/datasetView.do>

- 지하철 호선별 역별 승하차 인원 정보 (시간대별x, 일별O)

<https://data.seoul.go.kr/dataList/OA-12913/S/1/datasetView.do>

- 버스노선별 정류장별 시간대별 승하차 인원 정보

<https://data.seoul.go.kr/dataList/OA-15560/S/1/datasetView.do>

- 상권영역

수빈

집객시설

생활인구

추정매출

추정매출

상주인구

유동인구(지하철,버스)