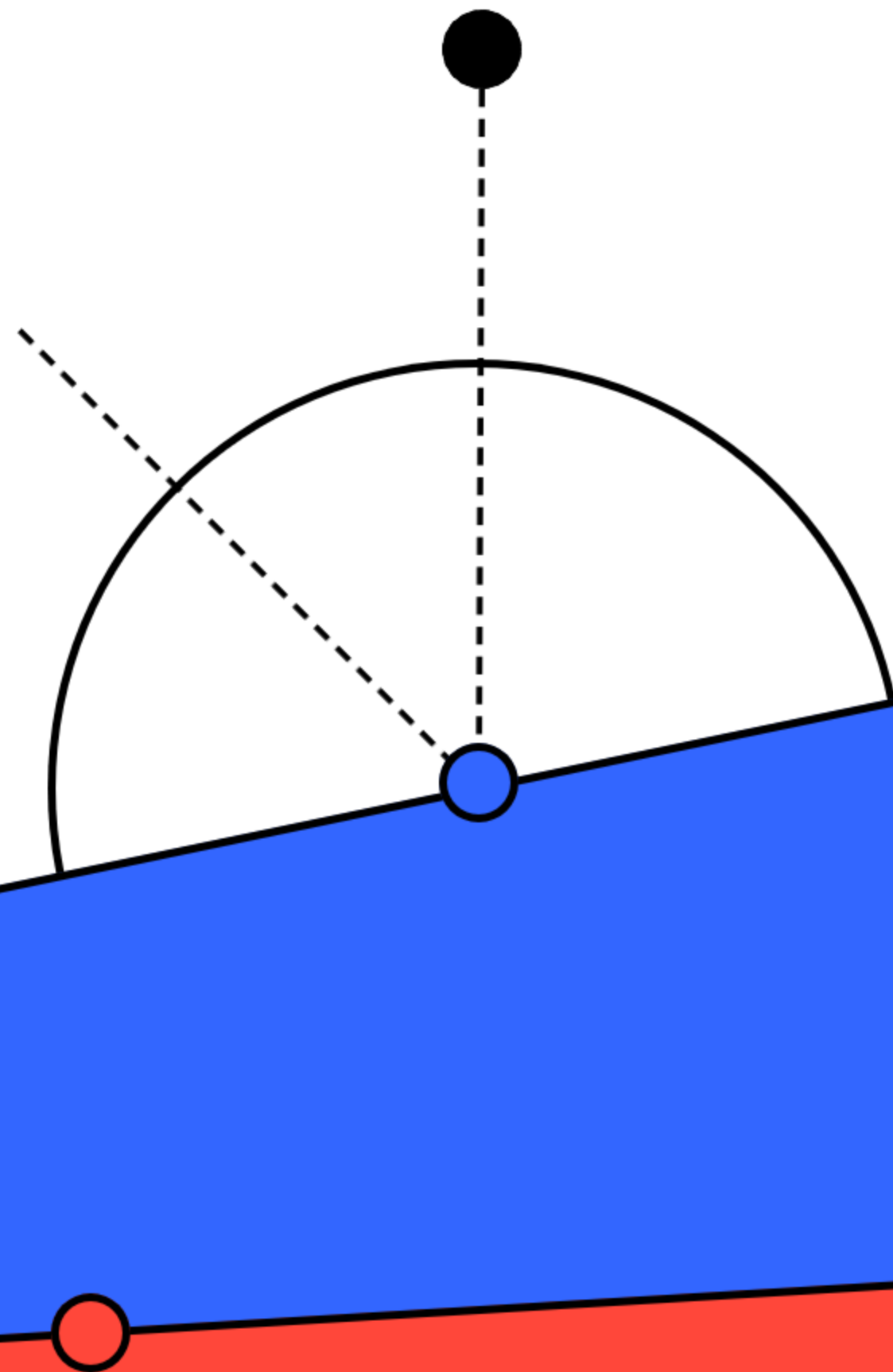


서울시 창업 위치 추천


상권 분석

32193012 이나영
32190158 공수빈



목차



- 01 진행 상황**
 - 02 전처리 과정**
 - 03 최종 데이터 생성**
- 

진행 상황

데이터 수집 (서울시 상권데이터) 100%

전처리 100%

최종 데이터 생성 (업종별 데이터) 100%

최종 데이터 생성 (상권별 데이터) 85%

◀ 진행중

클러스터링 (업종별 데이터) 100%

0% 클러스터링 (상권별 데이터)

30%

클러스터 별 업종별/상권별 데이터 분석

◀ 진행중

0% 추천 서비스 설계

0% UI 구축

전처리 과정

< 추정매출(매출 금액, 매출 건수) >

- 매출 비율 관련 열은 사용하지 않음
 - 한 행에 결측치가 7개 이상이라면 행 삭제
 - 총 매출 데이터가 없는 경우 행 삭제
 - 각 업종별 중요 열의 데이터가 없는 경우 행 삭제
 - 요일별 매출은 상관계수가 가장 높은 열의 데이터로 치환
-
- 예외) 일식데이터는 2022년 null값 데이터를 2021년의 해당 행, 열 데이터로 채운 후 위 과정 수행 → 전처리 후 삭제되는 행이 전체의 10%를 넘지 않도록 했는데 일식데이터의 경우 원래대로 전처리하면 약 18%의 행이 삭제되어 데이터 손실이 컸기 때문.

전처리 과정

< 점포수 >

- 업종별로 점포수가 '*' 값인 행에 대하여 해당 행의 분기당 매출금액이
(각 업종 분기당 매출금액의 중위값)/(중위값인 행의 점포수) 보다 크다면 2, 작다면 1로 치환함
(개인정보보호를 위해 점포수가 2 이하인 경우 '*'로 표시)

< 생활인구 >

- 결측치 확인 후 분기별로 나눠서 저장

< 직장인구 >

- 원본 직장인구 데이터에서 총직장인구 열만 추출

전처리 과정

< 집객시설 >

- 관공서, 은행, 의료시설(종합병원+일반병원+약국), 교육시설(유치원+초+중+고), 대학교, 백화점, 슈퍼마켓, 여가시설(극장+숙박시설), 교통시설(공항+철도역+버스터미널+지하철역+버정) 으로 정리하였음 → 19개 항목에서 9개 항목
(주요 집객시설 외 부가적인 집객시설까지 더한 것이 총 집객시설)
- 클러스터링을 위해 모든 데이터를 numeric type으로 변환하고 불필요한 데이터(상권명 등) 열을 삭제함

최종 데이터 생성

< 유동인구 >

- 상권 중심 좌표로부터 반경 500m 내에 있는 대중교통 이용자수를 그 상권의 유동인구로 보기로 하였음
 - 상권별 데이터 특성상 시간대가 6개로 나뉘어져 있고, 1~4분기로 나뉘어져 있음
(시간대 : 00시~06시 / 06시~11시 / 11시~14시 / 14시~17시 / 17시~21시 / 21시~24시)
(1분기: 1~3월 / 2분기: 4~6월 / 3분기: 7~9월 / 4분기: 10~12월)
- 유동인구도 위와 같이 6개의 시간대와 4개의 분기로 나누었음

최종 데이터 생성

< 유동인구 >

- 제공된 상권 데이터에서 상권 중심 좌표는 EPSG 5181을 사용하고 있어서 위도, 경도 좌표로 변환해주었음
- 상권 좌표 변환 전 → 후

	상권코드	상권코드명	x	y
0	2110753	시흥동 은행나무사거리	191934	438813
1	2110427	노원 을지대학교병원	206265	459431
2	2110755	동일여자고등학교	192300	439070
3	2110739	시흥3동 우편취급국	191289	438014
4	2111017	송파나루역 3번	210222	445516
...
1666	1001493	동대문패션타운 관광특구	200974	451975
1667	1001494	종로?청계 관광특구	199768	452273
1668	1001495	잠실 관광특구	210191	446339
1669	1001496	강남 마이스 관광특구	205310	445727
1670	1001491	이태원 관광특구	199503	448326

1671 rows × 4 columns

	상권코드	상권코드명	x	y
0	2110753	시흥동 은행나무사거리	126.908844	37.448687
1	2110427	노원 을지대학교병원	127.070978	37.634469
2	2110755	동일여자고등학교	126.912978	37.451006
3	2110739	시흥3동 우편취급국	126.901564	37.441482
4	2111017	송파나루역 3번	127.115614	37.509060
...
1666	1001493	동대문패션타운 관광특구	127.011025	37.567312
1667	1001494	종로?청계 관광특구	126.997374	37.569998
1668	1001495	잠실 관광특구	127.115275	37.516476
1669	1001496	강남 마이스 관광특구	127.060059	37.511003
1670	1001491	이태원 관광특구	126.994377	37.534435

1671 rows × 4 columns

최종 데이터 생성

< 유동인구 >

- 반경내 대중교통 POI 검색 API를 이용해 상권 중심 좌표 반경 500m 내 지하철역의 이름과 정류장 ARS 코드를 가져와서 저장하였음

ODSay

LAB

API소개

Document

Application

Community

데이터바우처

구매문의

로그인/회원가입

Document

Guide

API Release Note

API Reference

Console

반경내 대중교통 POI 검색 (Public Transit POI within Radius) Release v1

지정된 포인트와 반경을 가지고 대중교통 POI정보를 검색합니다.

* 로그인 하시면 콘솔에서 API를 호출 해보실 수 있습니다.

호출 URL

https://api.odsay.com/v1/api/pointSearch?lang=&x=&y=

GET 파라미터 (GET Parameter)

lang*	결과 언어 선택 (국문:0, 영문:1, 일문:2, 중문(간체):3, 중문(번체):4, 베트남어:5)	ex) 1
x*	x좌표(경위도)	ex) 126.933361407195
y*	y좌표(경위도)	ex) 37.3643392278118
radius	반경(m) (default:250)	ex) 250
stationClass	정류장 종류	ex) 1:2

실행

	상권코드	상권코드명	x	y	버스	지하철
0	2110753	시흥동 은행나무사거리	126.908844	37.448687	[18864, 18141, 18982, 18981, 18968, 18969, 189...	[]
1	2110427	노원 을지대학교병원	127.070978	37.634469	[11466, 11467, 11417, 11835, 11448, 11470, 114...	['하계']
2	2110755	등일여자고등학교	126.912978	37.451006	[18786, 18959, 18573, 18219, 18979, 18978, 188...	[]
3	2110739	시흥3동 우편취급국	126.901564	37.441482	[18540, 18166, 18014, 18013, 18169, 18534, 181...	[]
4	2111017	송파나루역 3번	127.115614	37.509060	[24202, 24391, 24220, 24225, 24219, 24199, 242...	['송파나루']
...
1666	1001493	동대문패션타운 관광특구	127.011025	37.567312	[2280, 2711, 2920, 2174, 2908, 2173, 2709, 223...	['신당', '동대문역사문화공원', '동대문']
1667	1001494	종로7청계 관광특구	126.997374	37.569998	[1247, 1769, 1019, 1176, 2226, 1018, 1020, 152...	['을지로4가', '종로3가', '종로5가']
1668	1001495	잠실 관광특구	127.115275	37.516476	[24903, 24241, 24396, 24211, 24210, 24729, 241...	['한성백제', '용촌토성']
1669	1001496	강남 마이스 관광특구	127.060059	37.511003	[23198, 23818, 23905, 23199, 23801, 23197, 239...	['봉은사', '삼성']
1670	1001491	이태원 관광특구	126.994377	37.534435	[3284, 3283, 3191, 3190, 3572, 3569, 3192, 319...	['이태원']

1671 rows × 6 columns

최종 데이터 생성

< 유동인구 >

버스 이용자수

- 월마다 갱신되는 서울시 버스노선별 정류장별 시간대별 승하차 인원 정보 데이터 가져오기
(2022년 3, 4분기와 2023년 1, 2분기 데이터를 가져왔음)
- 버스 승하차 승객수 데이터 시간대별로 합치기
- 노선이 겹쳐서 생기는 중복 정류장 데이터 합치기
- 분기별로 합치기
- 서울시에 있지 않은 정류장 데이터가 섞여있어서 서울시 버스 정류장만 남기기

지하철 이용자수

- 월마다 갱신되는 서울시 지하철 호선별 역별 시간대별 승하차 인원 정보 데이터 가져오기
(2022년 3, 4분기와 2023년 1, 2분기 데이터를 가져왔음)
- 지하철 승하차 승객수 데이터 시간대별로 합치기
- 호선이 겹쳐서 생기는 중복 지하철역 데이터 합치기
- 분기별로 합치기

최종 데이터 생성

< 유동인구 >

총 유동인구

- 앞에서 API 호출을 통해서 저장한 반경 내 지하철역, 버스정류장 ARS 코드 정보를 이용해서 각 상권별 시간대별 유동인구 데이터 만들기

	상권코드	상권코드명	00시~06시유동인구	06시~11시유동인구	11시~14시유동인구	14시~17시유동인구	17시~21시유동인구	21시~24시유동인구
0	2110753	시흥동 은행나무사거리	50653	451376	250560	300015	427015	158611
1	2110427	노원 을지대학교병원	108798	1459341	723562	827117	1271399	488187
2	2110755	동일여자고등학교	38419	362511	197829	234344	313151	116109
3	2110739	시흥3동 우편취급국	25340	237076	89731	119042	212037	74120
4	2111017	송파나루역 3번	42842	596696	316202	357721	642543	259258
...
1666	1001493	동대문패션타운 관광특구	302585	2780313	1994525	2254692	2899096	1031914
1667	1001494	종로?청계 관광특구	167720	4334021	3699463	4091380	5320223	1428696
1668	1001495	잠실 관광특구	13858	226257	105810	125122	211767	75950
1669	1001496	강남 마이스 관광특구	94171	1831723	848659	1026131	1815123	516101
1670	1001491	이태원 관광특구	73043	633325	427406	562386	921216	385347

1671 rows × 8 columns

최종 데이터 생성

< 업종별 데이터 >

ex) 분식전문점 최종 데이터

상권코 드	분기당 매출건 수	주중매 출건수	주말매 출건수	월요일매 출건수	화요일매 출건수	수요일매 출건수	목요일매 출건수	금요일매 출건수	토요일매 출건수	...	의료 시설 수	교육 시설 수	여가 시설 수	교통 시설 수	00시 ~06시 유동인 구	06시~11 시유동 인구	11시~14 시유동 인구	14시~17 시유동 인구	17시~21 시유동 인구	21시~24 시유동 인구
1001496	184370	137220	47150	25567	24440	28586	29207	29420	25433	...	0.0	0.0	0.0	0.0	94171	1831723	848659	1026131	1815123	516101
1001495	277977	182811	95166	36138	36250	37051	35815	37557	51003	...	0.0	0.0	0.0	0.0	13858	226257	105810	125122	211767	75950
1001494	332000	248152	83848	46130	49040	49020	52979	50983	48967	...	0.0	0.0	0.0	0.0	167720	4334021	3699463	4091380	5320223	1428696
1001493	302924	224109	78815	44474	43405	43504	46437	46289	45297	...	17.0	0.0	0.0	0.0	302585	2780313	1994525	2254692	2899096	1031914
1001492	579306	501645	77661	97969	99078	98537	109702	96359	44878	...	0.0	0.0	0.0	0.0	234578	6157760	2765755	3333748	6395073	1642232
...
2110010	18487	15479	3008	3308	3212	2602	3200	3157	1504	...	0.0	0.0	0.0	0.0	9935	147900	85620	105926	127250	45263
2110008	10216	7993	2223	1236	1732	1595	1938	1492	1212	...	0.0	0.0	0.0	0.0	9047	175314	111147	136166	170999	52697
2110005	6879	4911	1968	926	854	811	1210	1110	984	...	0.0	0.0	0.0	0.0	23135	290332	171336	214601	301124	106398
2110003	4208	3330	878	735	734	498	605	758	439	...	0.0	0.0	0.0	0.0	19692	315895	198560	238837	305327	103004
2110001	10197	5733	4464	1027	1100	1350	1024	1232	2311	...	0.0	0.0	0.0	0.0	7196	120567	76459	96002	109188	37317

```
Index(['상권코드', '분기당매출건수', '주중매출건수', '주말매출건수', '월요일매출건수', '화요일매출건수', '수요일매출건수', '목요일매출건수', '금요일매출건수', '토요일매출건수', '일요일매출건수', '시간대00~06매출건수', '시간대06~11매출건수', '시간대11~14매출건수', '시간대14~17매출건수', '시간대17~21매출건수', '시간대21~24매출건수', '남성매출건수', '여성매출건수', '연령대10매출건수', '연령대20매출건수', '연령대30매출건수', '연령대40매출건수', '연령대50매출건수', '연령대60이상매출건수', '분기당매출금액', '주중매출금액', '주말매출금액', '월요일매출금액', '화요일매출금액', '수요일매출금액', '목요일매출금액', '금요일매출금액', '토요일매출금액', '일요일매출금액', '시간대00~06매출금액', '시간대06~11매출금액', '시간대11~14매출금액', '시간대14~17매출금액', '시간대17~21매출금액', '시간대21~24매출금액', '남성매출금액', '여성매출금액', '연령대10매출금액', '연령대20매출금액', '연령대30매출금액', '연령대40매출금액', '연령대50매출금액', '연령대60이상매출금액', '점포수', '총생활인구수', '남성생활인구수', '여성생활인구수', '연령대10생활인구수', '연령대20생활인구수', '연령대30생활인구수', '연령대40생활인구수', '연령대50생활인구수', '연령대60이상생활인구수', '시간대1생활인구수', '시간대2생활인구수', '시간대3생활인구수', '시간대4생활인구수', '시간대5생활인구수', '시간대6생활인구수', '월요일생활인구수', '화요일생활인구수', '수요일생활인구수', '목요일생활인구수', '금요일생활인구수', '토요일생활인구수', '일요일생활인구수', '총직장인구수', '집객시설수', '관공서수', '은행수', '대학교수', '백화점수', '슈퍼마켓수', '의료시설수', '교육시설수', '여가시설수', '교통시설수', '00시~06시유동인구', '06시~11시유동인구', '11시~14시유동인구', '14시~17시유동인구', '17시~21시유동인구', '21시~24시유동인구'], dtype=object')
```

최종 데이터프레임 칼럼 수: 89개

최종 데이터 생성

< 상권별 데이터 >

최종 업종별 데이터는 업종별, 분기별로 나누어진 상태이기 때문에 업종, 분기를 나타내는 열이 없어서 각 데이터에 업종, 분기별로 번호를 부여하여 열을 생성하였음.

한식 : 0 / 양식 : 1 / 일식 : 2 / 중식 : 3 / 분식 : 4 / 제과점 : 5 /

카페 : 6 / 치킨 : 7 / 호프 : 8 / 패스트푸드 : 9

+ 행정동 코드 매핑

상권영역 데이터를 바탕으로 각 상권코드에 해당하는 행정동 코드를 매핑

감사합니다