

TRIPLE S 최종 보고서

- 서울시 창업 위치 추천 -



과 목 명		실무중심산학협력프로젝트2
담당교수		김승훈
학 과		컴퓨터공학과
학 번		32193012, 32190158
이 름		이나영, 공수빈
제 출 일		2023.12.21

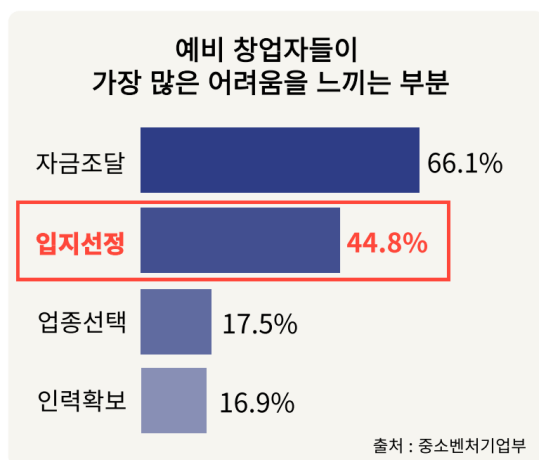
목 차

1. 프로젝트 개발 동기	3
2. 프로젝트 수행 내용 및 방법	4
2.1. 데이터 수집	4
2.2. 데이터 전처리	4
2.3. 데이터 생성	5
2.3.1. 업종 데이터	5
2.3.2. 상권 데이터	7
2.3.3 대표 데이터 및 대표 데이터 생성 알고리즘	7
2.4. 웹사이트 구현	9
3. 프로젝트 결과물	11
4. 기대 효과 및 개선 방향	19
4.1. 기대 효과	19
4.2. 개선 방향	19
5. 팀원 간 역할 분담	20
6. 참고 문헌	21

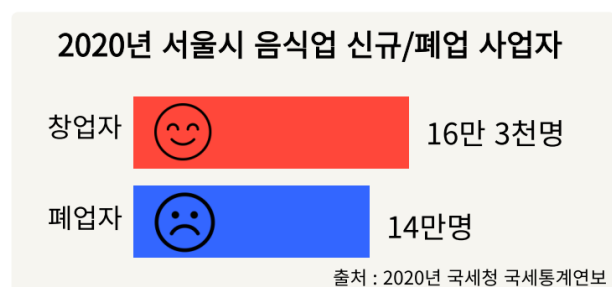
1. 프로젝트 개발 동기

자영업자들이 창업을 준비하는 기간은 6개월 미만 62%, 1년 미만 77.7%로 대부분이 1년을 넘지 않는다. 이들은 창업 준비로 주로 동일 업종 종사 경험, 시장조사, 사업계획서 작성 등을 하지만 정작 상권분석만을 따로 배우거나 체험하는 경우는 드물다. 실제로 예비 창업자들이 가장 많은 어려움을 느끼는 부분에서 입지 선정은 44.8%로 큰 부분을 차지한다. 즉, 상권 분석의 중요성을 알지만 스스로 분석하기에 너무 많은 상권과 상권 특성으로 인해 막막함을 느껴 상권 분석에 많은 투자를 하지 않는 경우가 많은 것이다. 때문인지 2020 서울시 자영업자 현황을 보면 신규 사업자가 16만 3천명인데 폐업 사업자가 14만명으로 상당히 많은 수의 자영업자가 매년 폐업하고 있음을 알 수 있다.

이러한 예비 창업자들의 상권 분석에 대한 어려움을 덜어주고자 서울시의 10가지 요식업을 대상으로 한 상권 분석 시스템을 제작하게 되었다. TRIPLE S의 주요 특징은 다음과 같다. 첫째, 상권을 비슷한 특성을 가진 상권끼리 묶어 정보를 제공한다. 서울시에는 약 1700개의 상권이 존재한다. 예비 창업자들이 겪는 가장 큰 어려움도 수많은 상권을 전부 분석하고, 수치를 종합적으로 판단해야 된다는 것이었다. 이러한 수고를 덜고자 상권들을 비슷한 특성을 가진 상권끼리 묶어 상권 그룹으로 제공한다. 둘째, 예측 데이터를 제공한다. 원천 데이터를 기반으로 상권 데이터와 업종 데이터를 생성하고, 이 두 데이터를 기반으로 사용자의 요구에 맞게 수치를 예측하여 제공한다. 따라서 현재 특정 상권에 특정 업종의 운영 중인 점포가 없더라도 데이터를 제공할 수 있다.



[사진 1]



[사진 2]

2. 프로젝트 수행 내용 및 방법

2.1. 데이터 수집

서울 열린데이터 광장에서 '상권 별 추정매출', '상권 별 집객시설', '상권 별 생활인구', '상권 영역' 데이터를 수집한다. 또한 각 상권 별 시간대별 유동인구 수 데이터를 생성하기 위해 '지하철역/버스 좌표', '지하철역 외부코드', '버스정류장 정류장 번호', '지하철 시간대별 승/하차 인원 정보', '버스노선별 정류장별 승/하차 인원 정보' 데이터와 ODSay 사이트의 '반경 내 대중교통 POI 검색 API' 를 사용해 상권 중심 좌표를 기준으로 반경 500m 내의 대중교통 데이터를 수집한다.

2.2. 데이터 전처리

데이터 전처리에는 Jupyter Notebook을 사용한다. pandas, numpy 라이브러리를 이용해 DataFrame 형태로 전처리하기 편리하고, 이후 데이터 생성에 필요한 여러 클러스터링 알고리즘을 제공하는 점을 고려하였다. 첫 번째로 피처를 삭제 및 병합하여 정리한다. 추정매출 데이터에서 요식업 업종만 남기고 삭제하고, 매출 비율 관련 피처는 매출금액, 건수로 비율을 알 수 있기 때문에 중복된 열로 판단하여 삭제했다. 직장인구 데이터에서는 총직장인구 열만 사용하였고, 집객시설 데이터는 집객시설이 19개 항목으로 과하게 세부적으로 나뉘어져 있어 9개 항목으로 축소했다. 두 번째, 결측치 처리 과정에서는 결측치 피처가 7개 이상인 행, 총 매출 데이터가 없는 행, 각 업종별 중요 피처 데이터가 없는 행은 삭제하고 삭제 과정에서 삭제되는 행이 전체 행의 10%를 넘는 경우 데이터 손실이 크다고 판단하여 치환 방법을 사용했다. 치환 방법은 상관계수가 가장 높은 열의 데이터로 치환하였고, 상관계수가 모호할 경우 전년도 데이터에서 해당 업종 분기의 해당 상권 피처 데이터로 치환했다. 점포수는 개인정보보호를 위해 점포수가 2 이하인 경우 '*'로 표시되어 있어 (분기당 매출금액의 중위값)/(중위값인 행의 점포수)보다 분기당 매출 금액이 크다면 2, 작다면 1로 치환해주었다. 마지막으로, 사용자가 입력한 행정동에 속한 상권의 정보를 제공해야 하므로 각 상권에 행정동 코드를 맵핑해주었고, 클러스터링을 위해 모든 데이터를 numeric type으로 변환하고 업종별로 0~9의 숫자 라벨을 부여했다.

2.3. 데이터 생성

매출은 업종 특성과 상권 특성이 합쳐져 형성된다. 즉, 같은 업종이더라도 상권의 특성에 따라 잘 될 수도, 안 될 수도 있고 반대로 같은 상권이더라도 더 잘 되는 업종이 있을 수 있다. 따라서 두 특성을 모두 반영하기 위해 원천 데이터를 바탕으로 상권의 영향을 받지 않은 업종 자체 특성 정보를 지닌 업종 데이터와 업종의 영향을 받지 않은 상권 자체 특성 정보를 지닌 상권 데이터를 생성하고, 이 두 데이터와 대표 데이터 생성 알고리즘을 이용해 사용자가 희망하는 행정동의 업종에 대한 특성을 분석, 예측한 상권 그룹 대표 데이터를 생성한다. 사용자는 최종적으로 생성되는 상권 그룹 대표 데이터를 시각화된 형태로 제공받게 된다. 예를 들어 업종 특성은 타깃 성별, 연령층, 가격대 등이 있을 수 있고, 상권 특성은 거주 인구수, 유동인구 특성, 소득 소비 등이 있을 수 있다.

2.3.1. 업종 데이터

업종 데이터는 서울시의 약 1700개의 수많은 상권을 유사성을 띄는 상권끼리 묶고, 분류된 각 상권 그룹의 특성을 사용자에게 제공하기 위함이다. 상권을 유사성으로 묶어 분류하기 위해서는 라벨링이 되어있지 않은 원천 데이터에서 패턴을 찾아내야 하고 이를 위해 비지도 학습인 클러스터링을 사용했다. 데이터 전처리 단계 에서와 마찬가지로 Scikit-learn, Standard scaler, PCA 등 클러스터링에 유용한 여러 라이브러리를 제공하는 Jupyter Notebook을 사용했고 그 중에서도 K-means++ 알고리즘을 사용했다.

클러스터링 진행 과정은 다음과 같다.

(1) 피쳐 개수 축소

현재 원천 데이터는 피쳐의 개수가 총 89개로 클러스터링 하기에 적절하지 않은 상태이다. 피쳐의 개수가 지나치게 많으면 원활한 클러스터링이 되지 않고 시각화를 통해 클러스터링 적합도를 확인하기에도 어려워 피쳐를 89개에서 중요 피쳐 9개로 축소했다. 9개의 피쳐는 최대한 89개 피쳐의 특성을 전부 포함할 수 있도록 선정했다.

(2) 데이터 스케일링

각 피쳐 값의 범위를 맞춰주기 위해 Standard Scaler를 사용해 스케일링을 진행한다.

(3) 군집 개수 설정

군집의 응집도를 평가하는 실루엣 계수를 이용해 적절한 군집의 개수를 찾아 업종별, 분기별 6~13개의 군집으로 결정했다.

(4) 클러스터링

K-means++ 알고리즘을 이용해 클러스터링을 진행한다.

(5) 클러스터링 적합도 확인

PCA 차원 축소 알고리즘을 사용해 2차원으로 시각화하여 클러스터링 적합도를 평가한다.

(6) 클러스터링 재 진행

위와 같은 과정을 거쳐 업종별, 분기별 6~13개의 군집을 가진 데이터가 생성되었다. 그런데 각 데이터별로 최적의 클러스터링 결과를 내고자 군집 개수를 모두 다르게 설정했던 점이 오히려 군집이 너무 많아져 클러스터링 사용 목적을 잃게 되는 상황이 발생했다. 상권의 개수가 너무 많아 상권 분석에 어려움을 느꼈던 사용자에게 유사한 특성을 가진 상권끼리 묶어 편리하게 정보를 얻을 수 있도록 하고자 했기 때문이었다. 또, 같은 업종이지만 분기별로 군집의 개수가 다른 경우 시간의 흐름에 따른 상권 군집 변화를 보여주기 어려워 분기별 데이터의 목적을 잃게 되는 문제가 발생했다. 그래서 모든 업종에서 실루엣 계수가 비교적 높았던 5개의 군집으로 설정하여 클러스터링을 다시 진행하였다.

최종적으로 총 89개의 피처를 가진 업종 데이터가 생성된다.

	상권코 드	분기당매출건 수	주중매출건 수	주말매출건 수	월요일매출건 수	...	11시~14시유동인 구	14시~17시유동인 구	17시~21시유동인 구	21시~24시유동인 구	cluster
0	2130326	20796	14494	6302	2848	...	432245	501141	941946	323593	0
1	2130325	5870	4022	1848	866	...	408065	477503	744147	281506	0
2	2130324	39282	24304	14978	5121	...	325962	385854	578937	203902	0
3	2130323	51927	36280	15647	8240	...	346323	424170	668960	221300	0
4	2130322	15868	10923	4945	2067	...	58956	68358	107563	46626	0
...
1460	2110005	12423	8578	3845	869	...	171336	214601	301124	106398	0
1461	2110004	11413	7304	4109	1356	...	635222	737785	892915	315001	0
1462	2110003	5116	4288	828	944	...	198560	238837	305327	103004	0
1463	2110002	15456	12895	2561	2923	...	445539	497466	616231	219220	0
1464	2110001	26389	16834	9555	3179	...	76459	96002	109188	37317	0

[사진 3] 업종 데이터

2.3.2. 상권 데이터

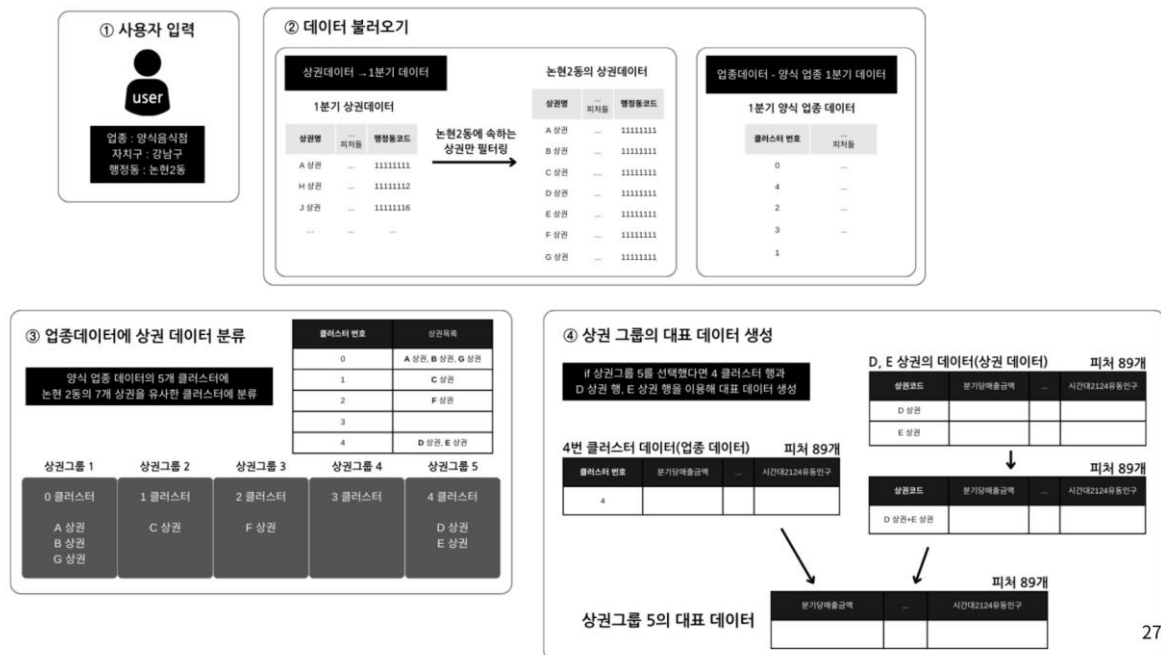
상권 데이터는 업종에 관계없이 상권만의 특성 정보를 반영한 데이터로 원천 데이터에서 같은 상권을 기준으로 묶은 형태로 가공한다. 결과, 총 89개의 피처를 가진 상권 데이터가 생성된다.

	상권 코드	분기당매출 건수	주중매출건 수	주말매출건 수	월요일매출 건수	...	시간대11~14유동 인구수	시간대14~17유동 인구수	시간대17~21유동 인구수	시간대21~24유동 인구수	행정동코 드
0	1001491	1302205	786681	515524	135625	...	427406	562386	921216	385347	11170650
1	1001492	5198622	4431089	767533	862188	...	2765755	3333748	6395073	1642232	11140520
2	1001493	1234674	937489	297185	180034	...	1994525	2254692	2899096	1031914	11140590
3	1001494	2178355	1681487	496868	322729	...	3699463	4091380	5320223	1428696	11110615
4	1001495	1904851	1236301	668550	233448	...	105810	125122	211767	75950	11710566
...
1574	2130322	40000	27128	12872	5234	...	58956	68358	107563	46626	11740620
1575	2130323	86557	56631	29926	13088	...	346323	424170	668960	221300	11740660
1576	2130324	80461	51910	28527	10381	...	325962	385854	578937	203902	11740685
1577	2130325	9135	6092	3043	1273	...	408065	477503	744147	281506	11740530
1578	2130326	128863	91544	37319	16171	...	432245	501141	941946	323593	11740560

[사진 4] 상권 데이터

2.3.3 대표 데이터 및 대표 데이터 생성 알고리즘

대표 데이터 생성 알고리즘은 앞서 생성한 업종 데이터와 상권 데이터를 모두 반영하여 특정 업종과 상권의 예측 데이터를 생성하는 단계이다. 해당 알고리즘을 통해 현재 특정 상권에 운영중인 특정 업종의 점포가 없더라도 사용자에게 추정 매출, 추정 유동인구 수 등의 예측 데이터를 제공할 수 있다. 알고리즘은 다음과 같다.

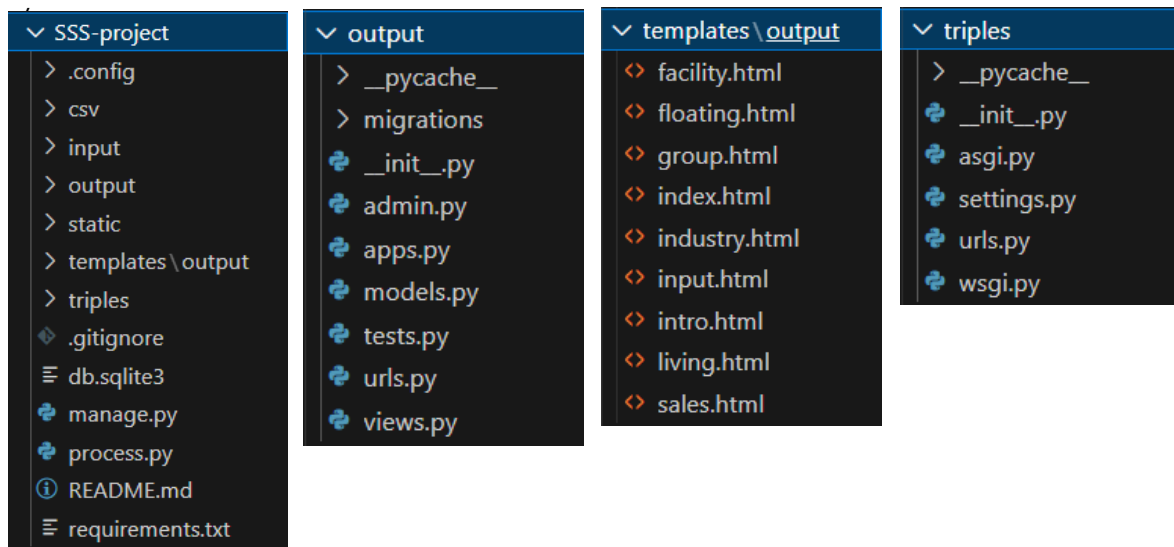


[그림 1] 대표 데이터 생성 알고리즘

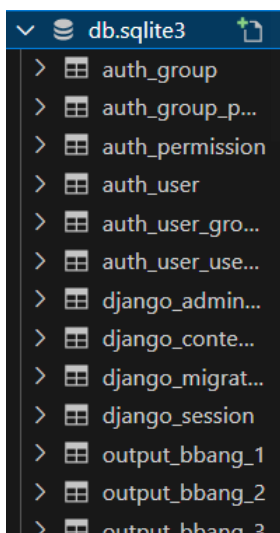
- ① 사용자로부터 업종과 행정동을 입력 받는다. 행정동에 속하는 상권들에 대해 분석하여 결과를 제공한다.
- ② 해당하는 상권 데이터와 업종 데이터를 불러온다. 상권 데이터에서 행정동의 행정동 코드를 기준으로 필터링하여 행정동에 속하는 상권 데이터만 가져오고, 사용자의 희망 업종에 대한 업종 데이터만 가져온다.
- ③ 업종 데이터에 상권 데이터를 분류한다. 분류 결과 유사한 특성을 가진 업종 클러스터 데이터와 상권 데이터가 한 그룹을 이루게 된다. 특정 클러스터와 비슷한 상권이 없다면 해당 클러스터에는 상권이 분류되지 않는다.
- ④ 각 그룹끼리 연산을 통해 상권 그룹의 대표 데이터를 생성한다. 이렇게 생성된 최종 데이터를 그래프 등으로 시각화해 사용자에게 제공한다.

2.4. 웹사이트 구현


데이터 전처리부터 생성까지 Python을 사용한 점과 Jupyter Notebook에서 사용하던 여러 라이브러리를 사용할 수 있다는 점을 고려해 Django Framework를 사용했다. Django의 기본 데이터베이스인 SQLite에 생성한 업종 데이터와 상권 데이터를 업로드해 관리하고 데이터 시각화에는 Chart.js를 사용했다. 협업에는 Git을 사용했다.



[사진 5] Dnjango




[사진 6] SQLite


SSS-project
Public














Edit Pins
Watch 0
Fork 0
Star 0

main
1 Branch
0 Tags

Add file
Code







naanayoung
uwsgi 수정

f1934a3 · last week
95 Commits

 .config	uwsgi 수정	last week
 csv	url mapping error 해결	last month
 input	ny	last month
 output	설명 수정	2 weeks ago
 static	아이콘, 글자 색 변경	2 weeks ago
 templates/output	add detail description	2 weeks ago
 triples	.config 추가, 서버 작업 중	2 weeks ago
 .gitignore	Add .gitignore file	last month
 README.md	Initial commit	last month
 db.sqlite3	url mapping error 해결	last month
 manage.py	장고 프로젝트 생성	last month
 process.py	클러스터별 중요 피쳐 모델 추가	last month
 requirements.txt	add requirements.txt	2 weeks ago

About

서울시 내에 있는 상권들을 업종별로 분석하여 예비창업자에게 조언해줄 수 있는 웹사이트

-  Readme
-  Activity
-  0 stars
-  0 watching
-  0 forks

Report repository

Releases

No releases published


Create a new release


Packages

No packages published

Publish your first package

Contributors 2


Janewary
SoobinKong


naanayoung

[사진 7] TRIPLE S의 Github repository

3. 프로젝트 결과물

TRIPLE S의 최초 화면이다. 사이트에 대해 간단히 소개하고, 상권 분석 서비스를 이용할 수 있도록 했다.



[사진 8] TRIPLE S 메인 페이지

TRIPLE S는 상권을 그룹으로 묶어서 그룹의 대표데이터를 제공하고 있기 때문에 이에 대해 사용자의 이해를 돕기 위한 글을 띄우고 있다. 상권 그룹에 대한 설명과 어떤 정보를 제공받아 어떻게 사용할 수 있을지 참고가 될만한 설명을 제공한다. 또한 10가지 요식업종을 분석한 결과를 매출 팁으로 제공하고 있다.



💡 상권 분석에 앞서 우리 서비스에 대해 알아보세요!

그룹이란? 비슷한 특성을 가지는 상권끼리 묶어놓은 것

서울시 상권 분석 서비스(TRIPLE S)는 업종 데이터를 분석한 뒤, 매출에 주 요인이 되는 지표들이 비슷한 경향을 띄는 상권들을 묶어 그룹으로 나누었습니다.

사용자가 선택한 희망 업종과 행동동에 따라 해당 지역의 상권들이 여러 그룹 중 하나의 그룹으로 분류되고, 그 결과 각 그룹에는 비슷한 특성을 가진 상권끼리 모이게 됩니다. 그룹마다 두드러지게 나타나는 특성을 키워드로 나타내었으니 참고해주세요!

추천 서비스 알고리즘을 통해 각 그룹에 있는 상권에서 창업을 하게 되었을 때의 가능성을 예측하실 수 있습니다. 그룹 상세 페이지에서는 매출금액, 생활인구, 유동인구, 집객시설 카테고리의 데이터를 시각화하여 보여드립니다. 매출을 발생시키거나 감소시키는 요인이 무엇인지 참고하시고 영업 시간, 휴무일, 가게 유형(저가형/고가형) 등의 경영 전략을 수립하세요!

[→ 상권 분석 보러 가기](#)

💡 10가지 요식업종의 매출 TIP을 알아보세요!

상권 분석 서비스에서 10가지 업종을 분석하여 매출이 높은 상권에서 공통적으로 보이는 특징, 업종별로 두드러지는 특성에 대해 정리하여 분기별 매출 평균 금액, 중위값을 나타낸 그래프와 함께 보여드립니다. 희망하는 업종에서의 매출에 중요한 요인들이 어떤 것들이 있는지 살펴보세요!

[→ 업종별 분석 보러 가기](#)

[사진 9] TRIPLE S의 두 가지 기능

업종별 분석 보러 가기를 눌러 페이지를 이동하면 다음과 같이 업종별 매출 평균과 순위 값을 그래프로 시각화 하여 나타내고, 각 업종에서 보이는 특성에 대해 분석한 내용을 함께 볼 수 있다. 우측의 네비게이션 바로 원하는 업종의 매출 팁으로 이동할 수 있다.

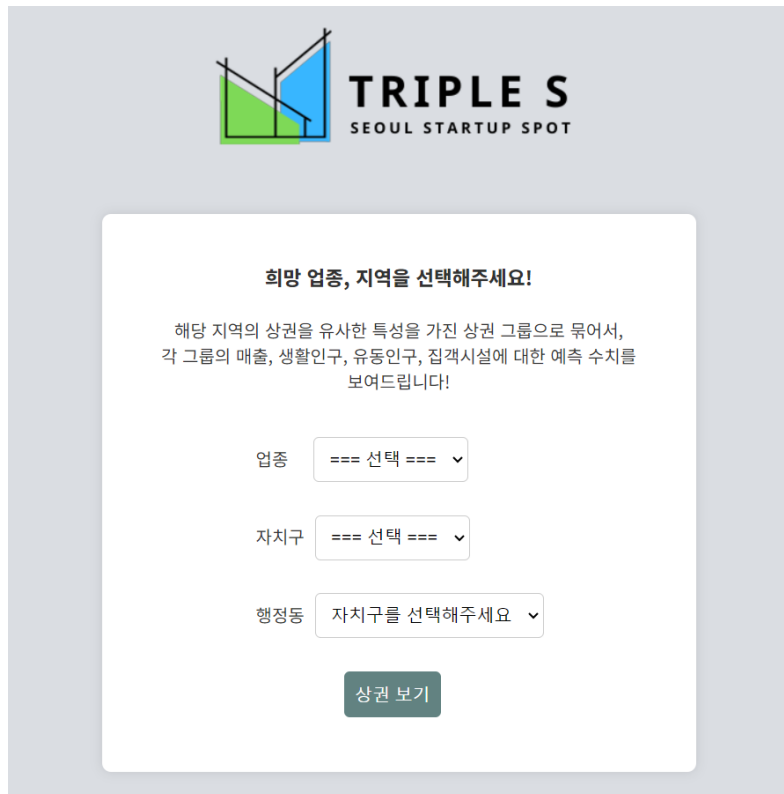


[사진 10] 한식음식점 매출 팁



[사진 11] 제과점 매출 팁


상권 분석 보러 가기를 눌러 페이지를 이동하면 희망 업종, 지역을 선택할 수 있는 사용자 입력 화면이 뜬다. 업종, 자치구, 행정동을 선택하고 상권 보기를 눌러 해당 행정동에서 해당 업종을 창업할 경우에 대한 데이터를 볼 수 있다.



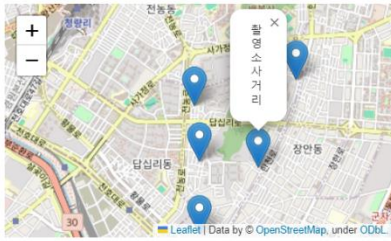
The screenshot shows a web application interface for 'TRIPLE S SEOUL STARTUP SPOT'. At the top, there is a logo consisting of a stylized 'S' made of green, blue, and yellow geometric shapes. Below the logo, the text 'TRIPLE S' is prominently displayed in a bold, black, sans-serif font, with 'SEOUL STARTUP SPOT' in a smaller font underneath. The main content area is a white rectangular box with a light gray border. Inside this box, the heading '희망 업종, 지역을 선택해주세요!' (Please select your desired business type and region!) is centered. Below the heading, a paragraph explains that the system will show estimated data for sales, population, and accommodation based on the selected business type and region. There are three dropdown menus for selection: '업종' (Business Type) with a placeholder '=== 선택 ===', '자치구' (District) with a placeholder '=== 선택 ===', and '행정동' (Administrative District) with a placeholder '자치구를 선택해주세요'. A dark green button labeled '상권 보기' (View Business Area) is positioned at the bottom of the selection area.

[사진 12] 희망 업종, 지역 선택

다음은 한식 업종, 답십리2동을 선택하고 상권 보기를 누른 경우에 대한 화면이다. 답십리2동에 있는 상권들의 위치를 지도로 표시하고 있고, 지도 우측에 현재 영업중인 한식 현황에 대한 매출금액, 점포수를 표로 띄우고 있다. 하단에는 유사한 상권끼리 묶인 상권 그룹목록이 있으며 그룹에 속하는 상권이 무엇인지 나타내고, 각 그룹에 대한 특성을 키워드로 표시하였다.



◀ 동대문구 답십리2동 한식



● 목록에 있는 상권은 현재 답십리2동에서 영업 중인 한식의 현황입니다.

상권명	월별매출금액	점포수
자동차부품상가	31,547,765	9
활영소사거리	253,492,025	21
답십리사거리	1,062,999,498	30
답십리공원	88,615,665	8
답십리시장	51,603,361	8

● 아래 상권 그룹은 답십리2동에서 한식을 창업했을 때의 예측 데이터입니다. 따라서 현재 영업 중인 점포가 없는 상권에 대한 예측 데이터도 제공됩니다.

상권 그룹 (총 1개)

→ 자동차부품상가

#남성 #30대

[상세보기](#)

상권 그룹 (총 1개)

→ 답십리시장

#남성 #30대 #의료시설

[상세보기](#)

상권 그룹 (총 1개)

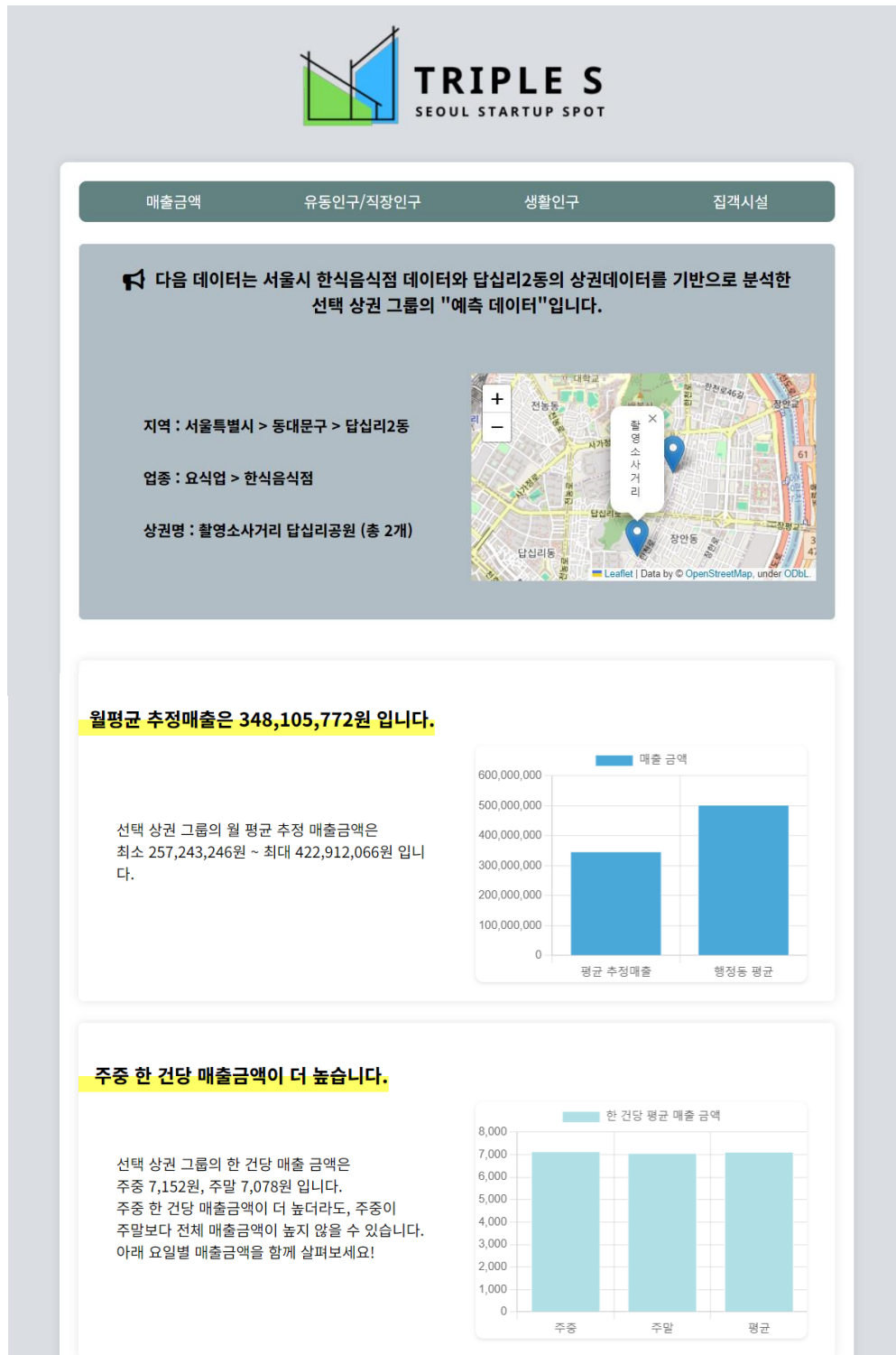
→ 답십리사거리

#남성 #30대 #평일 #직장인구 #관광서 #은행

[상세보기](#)

[사진 13] 상권 그룹 페이지

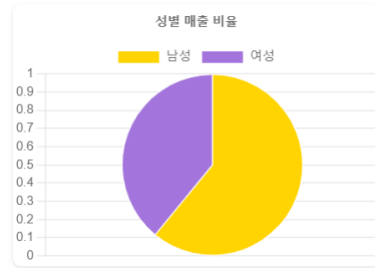
상권 특성을 매출금액, 유동인구/직장인구, 생활인구, 집객시설 네 가지 카테고리로 나누어 데이터를 제공한다. 항목별로 시간대별, 연령대별, 성별 데이터 수치를 시각화 하여 그래프로 나타내고 있다.



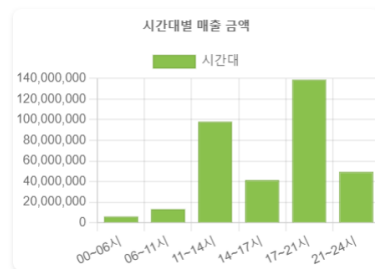
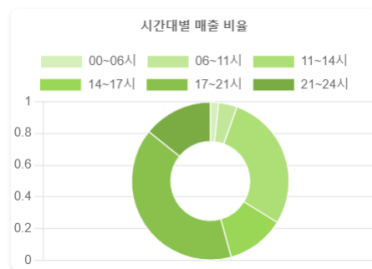
[사진 14] 매출금액 카테고리

남성 매출금액이 여성 매출금액 보다 많습니다.

여성 매출금액은 전체 매출의 39.1%,
남성 매출금액은 전체 매출의 60.9%를 차지합니다.
남성 매출금액이 여성 매출금액보다 월등히 높아,
남성이 여성보다 훨씬 소비가 많은 것으로 보입니
다.



17시~21시 매출금액이 가장 높습니다.

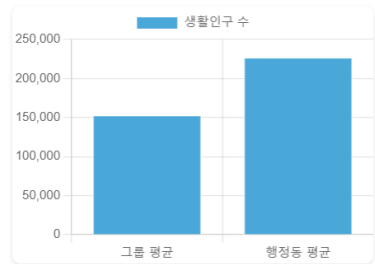


[사진 15] 매출금액 카테고리

① 생활인구란 특정 지역, 특정 시점에 존재하는 모든 인구를 말합니다. 해당 지역에 주소지를 둔 사람은 물론 일시적
으로 서울을 찾은 인구가 모두 해당됩니다. 생활 인구가 소비자로 이어질 수 있도록 마케팅 전략을 세워야합니다.

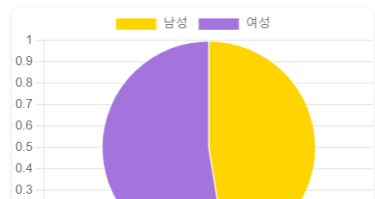
월평균 생활인구수는 152,976명 입니다.

선택 상권 그룹의 월 평균 추정 생활인구수는
최소 105,885명 ~ 최대 184,017명 입니다.



여성 생활인구가 남성 생활인구 보다 많습니다.

여성 생활인구수는 전체 생활인구수의 52.6%,
남성 생활인구수는 전체 생활인구수의 47.4%를 차
지합니다.

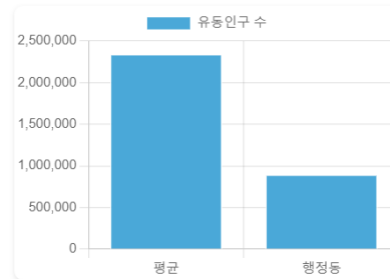


[사진 16] 생활인구 카테고리

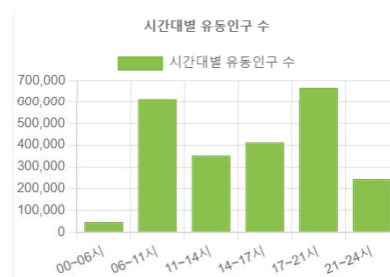
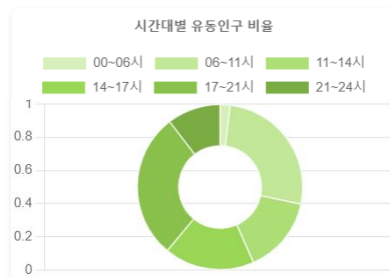
① 유동인구란 일정 기간 동안 한 지역을 오고 간 사람의 수를 말합니다. 단순히 많은 유동인구가 큰 매출로 이어지는 것이 아니기 때문에 매출과 연관이어 흐르는 유동인구가 많은 지역인지, 소비자로 이어지는 유동인구인지, 판단하는 것이 중요합니다.

월평균 유동인구 수는 2,339,465명입니다.

선택 상권 그룹의 월평균 유동인구는
최소 2,222,789명 ~ 최대 2,491,078명 입니다.

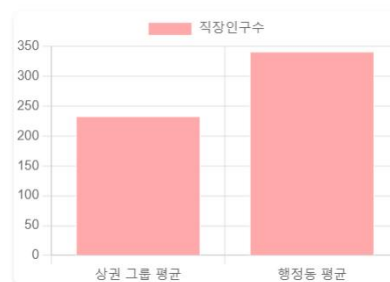


17시~21시 유동인구가 가장 많습니다.



직장인구는 평균 234명 입니다.

선택 상권 그룹은 직장인구수가 답십리2동 평균에
비해 적은 편입니다.



Triple S

[사진 17] 유동인구/직장인구 카테고리

4. 기대 효과 및 개선 방향

4.1. 기대 효과

TRIPLE S는 방대한 양의 상권 분석에 막막함을 느끼고, 상권 분석 컨설턴트를 이용하려 하니 막대한 비용이 부담되는 예비 창업자들에게 창업의 첫 시작, 입지 선정에서 도움이 될 것을 목적으로 한다. 상권분석은 실패를 줄이고 성공을 키우는 과정이다. 이들이 TRIPLE S로 여러 가정을 들어 시뮬레이션 해보고 해답을 찾는 발판이 되며, 점포의 위치가 창업 실패 요인으로 작용할 가능성이 줄어들 것을 기대한다. 나아가 상권 선정에 성공하여 지역 상권의 발전에도 기여할 것을 기대한다.

4.2. 개선 방향

(1) 창업성

TRIPLE S 프로젝트의 본래 목적은 공공데이터를 사용하였고, 예비 창업자들에게 도움을 주기 위함 이기 때문에 무료 서비스로 계획하였으나 사용자들이 증가한다면 광고 수익 모델, 또는 프리미엄 기능으로 수익을 얻을 수 있다. 창업 위치 추천 서비스이기 때문에 지역 기업이나 기타 창업 관련 광고주들로부터 광고를 받아 표시할 수 있고, 무료로 제공되는 데이터 외에 더 상세한 분석 리포트, 실시간 업데이트 등 프리미엄 기능을 유료로 제공하는 방법을 고려해볼 수 있다.

(2) 부족한 데이터

① 다양한 연도의 데이터 추가

TRIPLE S는 2022년 데이터를 기반으로 제작된 프로그램이다. 때문에 분기별 시간의 흐름에 따른 상권 변화는 알 수 있어도 연도별 시간의 흐름에 따른 데이터는 제공할 수 없었는데 2021년 데이터와 2023년 데이터를 추가하면 보완될 것이다.

② 소득, 소비 데이터 추가

소비자의 소득과 소비는 매출의 중요 특성으로 무시할 수 없는 데이터이다. 하지만 원천 데이터를 수집했던 2023년 4월 당시에는 상권 별 소득 소비 데이터가 공공데이터로 제공되지 않아 사용하지 못한 아쉬움이 있었는데, 2023년 11월부터 서울시 열린

데이터 광장에서 해당 데이터를 제공하기 시작한 것을 확인했다. 데이터의 전처리 후 보완될 가능성이 있을 것으로 보인다.

③ 프로그램 확장성

현재 TRIPLE S는 서울시의 10개 요식업만을 대상으로 데이터를 제공한다. 서울시 특성과 타 도시의 특성, 요식업 특성과 타 업종의 특성이 다르기 때문이었는데, 현재 요식업에 대한 분석과 기본 프로그램이 완성된 상태이기 때문에 서울시 외 나머지 도시의 요식업 업종 데이터도 서울시와 마찬가지로 원천 데이터가 제공된다면 보충될 가능성이 있다. 하지만 요식업을 제외한 타 업종에 대해서는 요식업 알고리즘을 그대로 적용해도 될 지의 우려가 있어, 각 업종의 특성을 분석한 뒤 알고리즘의 수정과 추가 작업을 거쳐야 할 것으로 본다.

5. 팀원 간 역할 분담

- 성명 : 이나영

■ 학과 : 컴퓨터공학과

■ 학번 : 32193012

■ 역할 : 데이터 수집, 데이터 전처리(5개 업종), 데이터 생성(상권 데이터, 업종 데이터), 추천 서비스 알고리즘 설계 및 구현, 웹사이트 UI 일부 작업, 프로그램 서버 배포

- 성명 : 공수빈

■ 학과 : 컴퓨터공학과

■ 학번 : 32190158

■ 역할 : 데이터 수집, 데이터 전처리(5개 업종), 데이터 생성(유동인구 데이터, 업종 데이터), 추천 서비스 알고리즘 일부 설계, 웹사이트 UI 작업

6. 참고 문헌

- 「상권은 매출이다」, 송규봉 지음, 북스톤, 2020.02.12.
- KOSIS 국가통계포털, 국세청 국세통보 사업자 현황, 자료갱신일:2023.08.11.
- 서울 열린데이터광장
<https://data.seoul.go.kr/>
- Odsay
<https://lab.odsay.com/>