

# 감성 및 매출 요인을 고려한 상권 분석 시스템 설계

: 호프 및 간이주점을 중심으로



팀장 2019112464 남경현

팀원 2021112415 김민지

2021112397 성호정

2019112520 임형준

2019112443 허재원

---

# CONTENTS

---

01

## 연구 배경

사회적 현황  
연구 필요성

02

## 연구 목표

Persona 설정  
현행 서비스 분석  
연구 목표

03

## 연구 프레임워크

연구 프레임워크

04

## 연구 방법론

상권 분석  
상권 내 입지 분석  
상권 별 감성 분석

05

## 연구 결과

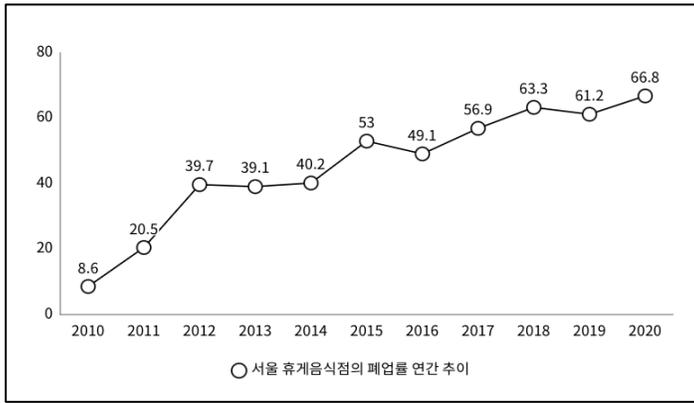
최종 사용자 시나리오  
결론 및 향후 연구

06

## Appendix

# 1. 사회적 현황 및 연구 필요성

## 서울시 음식점 폐업률 증가



<그림 1> 서울시 음식점 폐업률 연간 추이

- 1) 행정안전부는 서울시 내 휴게음식점의 **절반이 3년 이내 폐업**하고, 폐업률이 60%가 초과되어 자영업자 파산 위험이 증가하고 있음을 발표함
- 1) 통계청에 따르면, 소규모 자영업의 신규 **창업은 감소하는 추세**를 보이고 있음

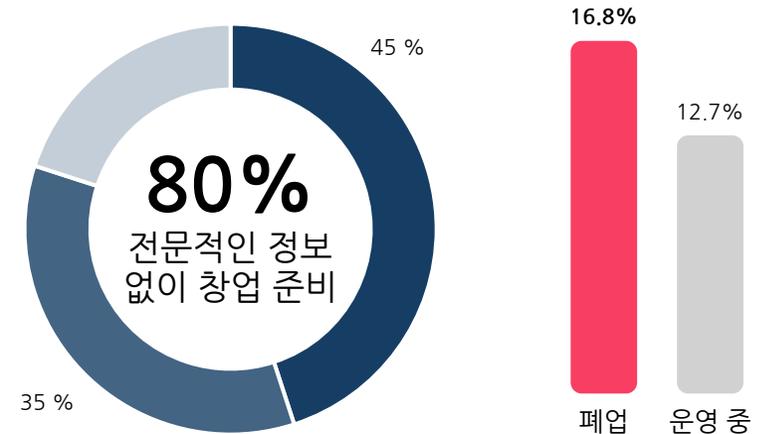
## 외식 산업에서의 컨셉의 중요성

refer	외식업의 컨셉 및 브랜드의 역할
2)서희정·박지연 (2019)	외식 기업 마케팅커뮤니케이션이 브랜드 태도와 신뢰에 유의한 영향을 미침
3)이영미 (2018)	외식기업 브랜드 태도와 신뢰는 구매의도에 유의한 영향을 미침
4)고호석 (2021)	외식기업은 타 업체와 차별화된 특성을 살려 독특한 서비스를 이용, 고객들에게 제공할 수 있도록 노력해야 함

<표 1> 외식업의 컨셉 및 브랜드의 역할

- 4) 외식 산업은 **고객의 경험**을 중요시하고, 경험에 대한 중요성은 **타 산업 대비 매우 높은 비중**을 차지함
- 4) 대체적으로 소비자들의 구매 의도는 소비자들의 구매 행동 예측을 위한 행동 지표로서, 높은 구매 의도는 구매 행동으로 연결될 가능성이 높음

## 소상공인의 정보 비대칭성



<그림 2> 소상공인 창업 준비 수준 <그림 3> 창업 시 정보 부족을 겪은 업체 현황

- 5) 중소기업청 전국 소상공인 실태조사에 따르면 소상공인 중 45%는 **아무 정보 없이**, 35%는 친인척 등 **비전문적 정보**를 통해 창업을 준비하고 있음
- 6) 폐업한 음식점 업체의 경우, **정보 부족**을 창업 과정에서 겪은 어려움으로 응답한 비중이 더 높았음

1) 서울 휴게음식점 절반 3년내 폐업...재작년부터 폐업률 60%대 <https://www.yna.co.kr/view/AKR20200429054800003>  
 2) 서희정, & 박지연. (2019). 체육관 지도자의 마케팅커뮤니케이션이 관원생의 체육관 만족과 신뢰 그리고 충성도에 미치는 영향. 대한무도학회지, 21(4), 101-112.  
 3) 이영미. (2018). 외식기업의 사회적 책임활동이 브랜드자산, 브랜드 태도 및 구매의도에 미치는 영향: 신뢰의 조절효과를 중심으로.  
 4) 고호석. (2021). 외식기업의 브랜드 증거와 마케팅커뮤니케이션이 브랜드 태도, 신뢰 및 구매의도에 미치는 영향. 지역산업연구, 44(1), 143-165.  
 5) 핀다 사용자 10명 중 9.4명 "상권분석 데이터 도움됐다" <https://www.hankyung.com/article/2024010427555>  
 6) 김상호, 이계원, 홍연아, 문동현, 허성운, 이육직, & 신성용. (2020). 포용성장 및 지속가능성 관련 식품제조업 분야 실태조사 결과. 한국농촌경제연구원 기본연구보고서, 1-207.

## 2-1. Persona 설정

### 예비 창업자 인터뷰를 통한 사용자 요구 분석

“유망한 지역에서 창업을 원하지만 정보가 부족해요”



이현수 | 27세 | 직업: 창업희망자 | 거주지: 수원

#### Background

- 대학 졸업 후 서울 자영업 창업을 준비하고 있음
- 아르바이트로 모은 돈과 정부 지원 정책을 통해 창업 자금을 마련함

#### Personality



#### Goals & Challenges

- 사업의 장기적인 성공을 위해, **향후 전망이 좋은 지역에서** 가게를 열고자 함
- **사업 실패의 부담**이 크며 신중한 의사 결정 필요
- 서울 지역에 대한 **정보가 부족**하며 여러 상권을 비교 조사하여 결정을 내리고자 함

“회사 일 때문에 직접 조사할 시간이 부족해요”



김민수 | 32세 | 직업: 회사원 | 거주지: 서울

#### Background

- 회사를 그만 두고 ‘와인 주점’ 창업을 고려 중
- 창업 준비 초기로, 회사 일과 병행하느라 사전 조사를 할 시간이 부족함

#### Personality

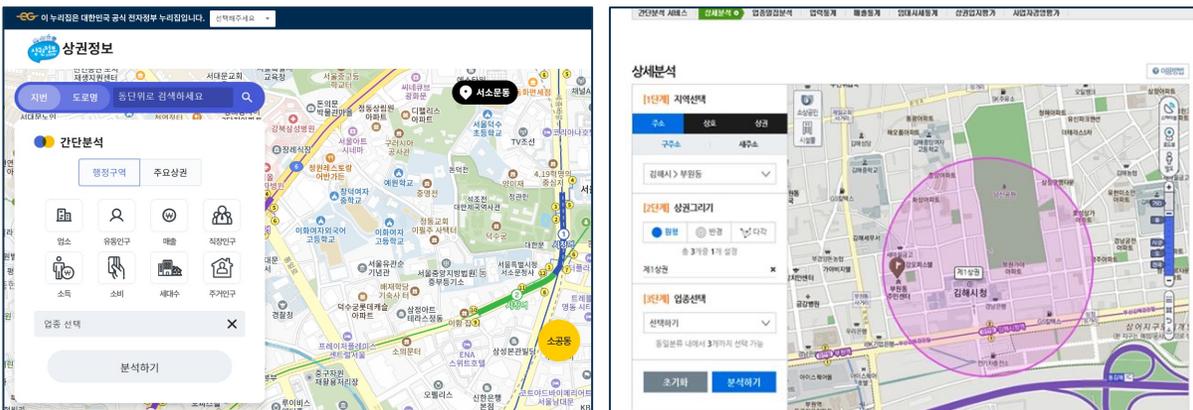


#### Goals & Challenges

- 상권 분석 서비스를 활용하여 사업을 시작할 유망한 위치를 탐색하고자 함
- **와인 주점 분위기를** 선호하는 지역 선별 필요
- **시간 부족**으로 인해 **광범위한 조사가 어려움**
- 신뢰할 수 있고 효율적인 분석 도구가 필요

# 2-2. 현행 서비스 조사 및 연구 목표

## 기존 상권 분석 서비스의 한계

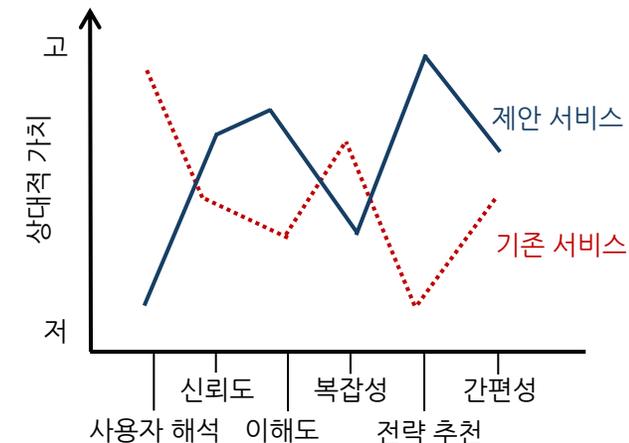


<그림 4> 서울시 상권정보 분석 사이트

- 7)(강진희, 2017)에서 기존 상권 분석 서비스에서는 상권에 대한 정보를 제공하되 **사용자가 이를 해석해야 한다는 한계점**을 말함
- 기존 상권 분석 서비스는 제공하는 정보를 활용하는 인사이트 정도의 수준에 그치고, 해석 및 **전략 개발은 창업자의 역량에 달려 있음**

## 서비스 차별점 도출: Service Value Curve

<b>Eliminate</b> - 사용자 해석	<b>Raise</b> - 간편성 - 이해도 - 신뢰도
<b>Reduce</b> - 복잡성	<b>Create</b> - 전략 추천

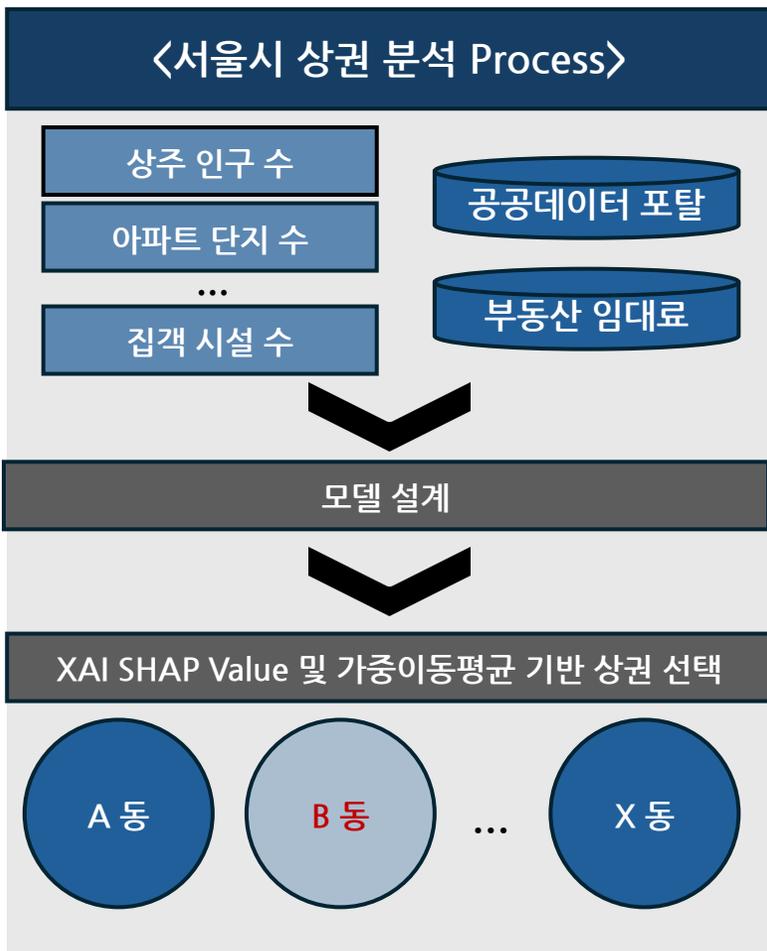


<그림 5> ERRC Grid & Value Curve

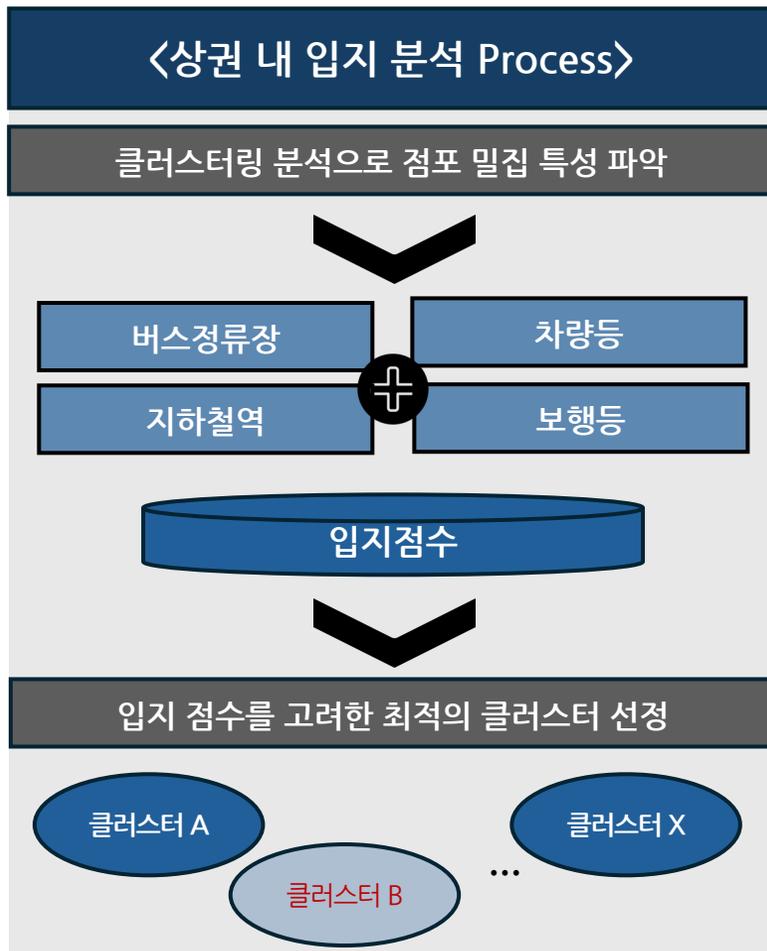
- ERRC Grid를 통한 Value Factor 선정 및 평가된 Value Factor를 기준으로 Service Value Curve 제작함
- 사용자 개인의 해석을 대신하여, 데이터 기반의 전략 추천을 추가한 신서비스를 제안함

“ 예비 창업자들의 주관적인 해석 오류를 줄이고, 데이터에 기반한 객관적이고 효율적인 창업 입지 선정 및 전략 수립을 지원 ”

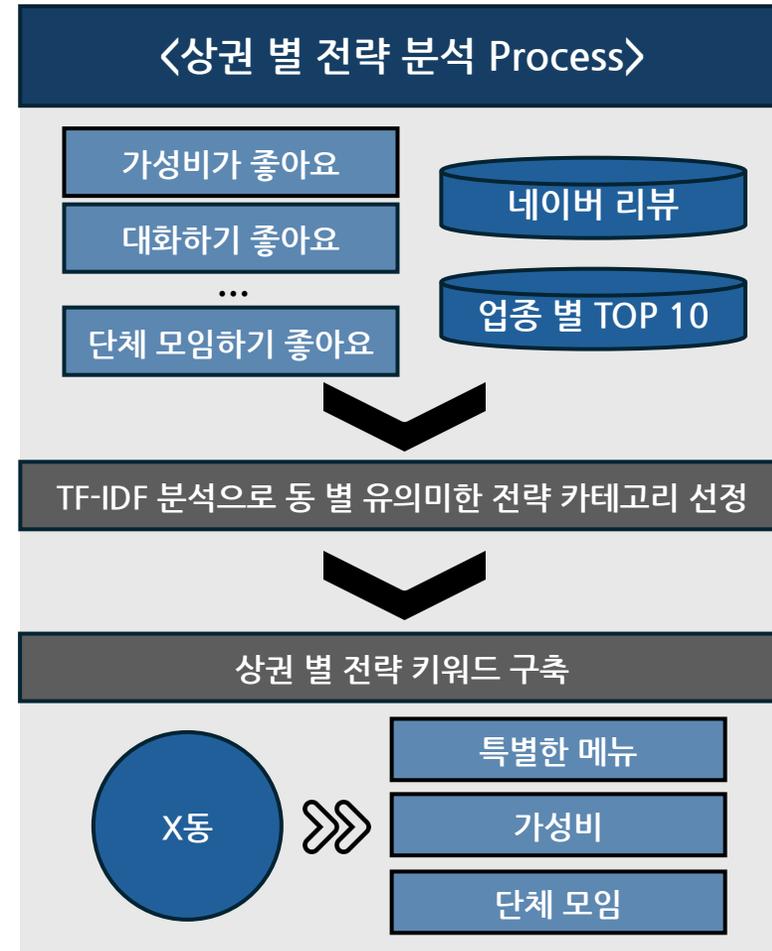
# 3. 연구 프레임워크



<업종 별 유망 상권 선정 >



<입지 점수를 통한 상권 내 입지 분석>



<전략 추천을 위한 키워드 구축 >

## 4-1. 연구 방법론 - 상권 분석

### 데이터 수집 및 모델 학습

#### 독립변수 수집 및 전처리

- 상권의 **매출에 영향을 미치는 변수**를 <sup>8)</sup>선행연구 기반으로 선정
- 서울 열린 데이터 광장 상권정보 데이터 수집
- 상관성 분석을 통해 상관계수가 <sup>9)</sup> ±0.7 이내인 변수로 최종 독립변수 선정

#### 최종 독립변수 선정

변수 명	설명	변수 타입
연령대 별 상주인구	해당 행정동의 20,30,40,50, 60대 상주 인구	연속형
연령대 별 유동인구	해당 행정동의 20,30,40,50, 60대 유동 인구	연속형
연령대 별 직장인구	해당 행정동의 20,30,40,50, 60대 직장 인구	연속형
평균 소득 금액	해당 행정동의 평균 소득 금액	연속형
아파트 수	해당 행정동 내에 존재하는 아파트의 수	연속형
집객 시설 수	해당 행정동 내에 존재하는 집객 시설 수	연속형
변화 지표	해당 행정동의 인프라 변화량	범주형

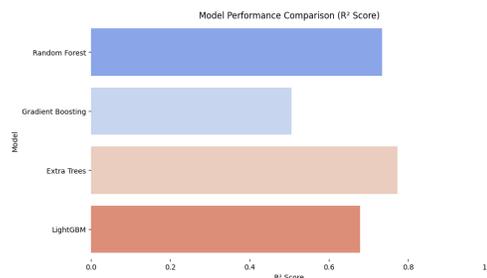
<표 2> 최종 독립변수 설명

#### 종속변수 선정 및 수집

- 상권의 **상이한 임대료 값과 점포수**를 반영하기 위해 행정동 별 부동산&점포 수 데이터를 수집
- **행정동 별 매출을 임대료와 점포수로 나눠서** 종속변수로 사용

#### 모델 학습

$$X = \text{상주인구, 유동인구} \dots \quad Y = \frac{\text{직종 별 상권의 총 매출}}{\text{해당 상권의 점포 수} * \text{상권 부동산 평균값}}$$



<그림 6> 모델 간 R-squared 성능 비교 시각화

모델	R-squared
RandomForest	0.747
GradientBoosting	0.509
ExtraTrees	0.779
LightGBM	0.679

<표 3> 모델 간 R-squared 성능 비교

- Pycaret AutoML 라이브러리를 사용하여 최적의 모델을 설계
- 하이퍼파라미터 튜닝 진행하여 회귀 모델의 설명력을 높임

# 4-1. 연구 방법론 - 상권 분석

## XAI 기반 중요도 산출

: 설계한 모델을 기반으로 각 변수가 매출 값에 어느정도 중요도를 가지는지 설명가능한 인공지능 SHAP value를 사용해 업종별로 도출

### 업종 별 중요도 해석

변수	호프 - 간이주점 Feature 중요도
20대 직장인구 수	34%
20대 유동인구 수	14%
30대 직장인구 수	14%
아파트 단지 수	11%
월 평균 소득 금액	9%
집객 시설 수	7%

<표 4> 호프 - 간이주점 Feature 별 매출에 끼치는 영향

Feature	양식 음식점 Feature 중요도
월 평균 소득 금액	29%
30대 유동인구 수	23%
30대 직장인구 수	22%
아파트 단지 수	13%
집객 시설 수	10%
50대 직장인구 수	3%

<표 5> 양식음식점 Feature 별 매출에 끼치는 영향

- 호프 간이주점 - 20대 직장인구 수가 34%로 매출에 가장 큰 영향을 끼침
- 양식 음식점 - 월 평균 소득 금액이 29%로 매출에 가장 큰 영향을 끼침
- 업종에 따라 상이한 변수 영향력을 바탕으로 <Feature 중요도> 도출

## WMA 산출 및 유망상권 도출 : 호프 및 간이주점 중심으로

$H_0$  : 유망 상권은 매출에 큰 영향을 끼치는 요인이 증가하는 추세를 보임

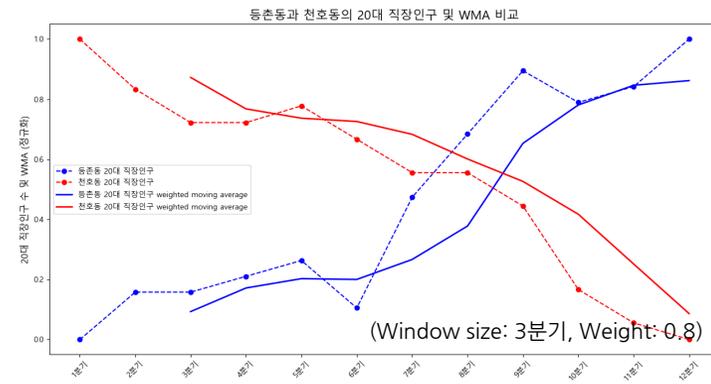
$$Target = \sum P_i \times WMA_i$$

$P_i$  : Feature 중요도  
 $WMA_i$  : Feature 변화를 가중이동평균

- 상권의 분기 별 Feature 변화율을 구하고, 변화 추세를 보기 위해 WMA를 도출함
- 상권의 <Feature 변화율 WMA>에 <Feature 중요도> 를 곱하여 Target 값 산출

유망 상권 순위	호프-간이주점
1	신사동
2	등촌동
3	중계동
...	...
379	천호동

<표 6> SHAP value, 변화율 기반으로 상권 순위 도출



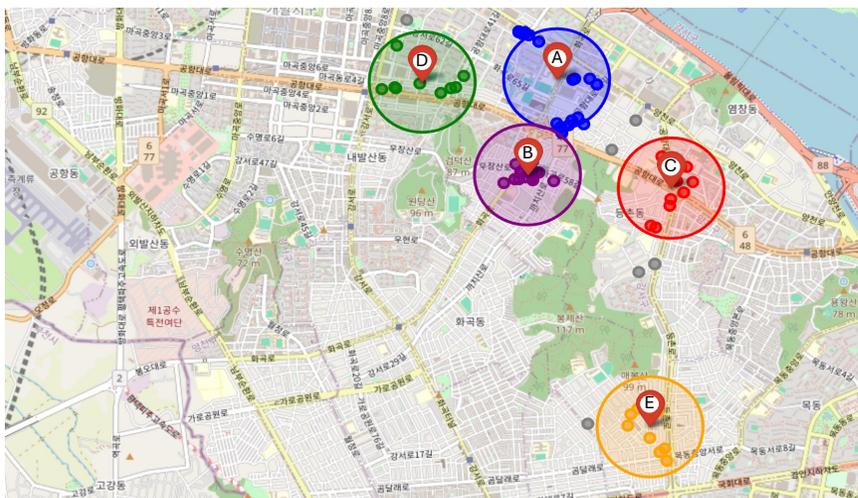
<그림 7> 등촌동, 천호동의 핵심 Feature 분기 별 가중이동평균 비교

- 등촌동** : 핵심 Feature의 WMA가 분기별로 증가하는 추세를 보이는 **유망 상권**
- 천호동** : 핵심 Feature의 WMA가 분기별로 감소하는 추세를 보이는 **쇠퇴 상권**

## 4-2. 연구 방법론 - 상권 내 입지 분석

**상권 내 최적 입지 선정:** 호프 및 간이주점 중심으로

- 10) 선행 연구에 따르면 업종마다 클러스터가 형성되는 양상이 다르며
- 8) 외식업종과 같은 경우는 밀집되어 있는 것이 더 높은 매출로 이어짐



<그림 8> 등촌동 호프-간이주점 클러스터링 결과

- DBScan을 활용하여 ε 을 300m로 설정한 뒤 클러스터를 진행함
- 등촌동의 호프-간이주점의 경우 5개의 클러스터가 나타났으며, 높은 밀집도를 보임
- 클러스터를 통해 상권을 세분화한 후 다음 절차로 입지를 분석함

- 11) 입지 선정에 있어 가시성과 접근성이 중요한 요인으로, 이를 교통 요소로 파악함
- 교통시스템의 구성요소인 12) “이용자”, “운송수단”, “도로시설”에 해당하는 요인으로 입지 점수를 도출함

$$N_{S,k} = \sum_{i=1}^n I(d(S, B_i) \leq T)$$

$$TS_i = 25 \times \sum \min\left(\frac{N_{S,k}}{\mu_{S,k}}, 1\right)$$

- S : 클러스터 중심점의 위치
- k : 교통 요소 종류
- B<sub>i</sub> : 점포 i 번째 교통 요소의 위치 (i = 1, 2, ..., n)
- d(S, B<sub>i</sub>) : 점포와 교통 요소 사이의 거리
- T : 임계값; 0.5km
- I : 함수; 조건 만족 시 1 반환, 불만족 시 0 반환
- μ<sub>S,k</sub> : 모든 N<sub>S,k</sub> 의 평균

클러스터	버스정류장	지하철역	차량등	보행등	입지 점수
A	0.53719	1.0	0.645503	0.714286	72.42
B	0.713092	1.0	0.803124	0.89264968	85.22
C	0.954821	1.0	0.947303	0.962137	96.61
D	0.950413	0.0	0.539683	0.371429	46.54
E	1.0	0.0	1.0	0.985714	74.64

<표 7> 등촌동 호프-간이주점 입지점수

10) Jung, S. S., & Jang, S. S. (2019). To cluster or not to cluster?: Understanding geographic clustering by restaurant segment. International Journal of Hospitality Management, 77, 448-457.  
 11) 한주옥, & 노태욱. (2016). AHP를 활용한 주유소 입지 결정 요인에 관한 연구. 부동산연구, 26(4), 147-165.  
 12) 이동민. (2008). 교통이용자 측면 (Transportation User Perception)을 고려한 교통공학연구. 교통기술과정책, 5(1), 246-250.

## 4-3. 연구 방법론 - 상권 별 감성 분석

### 데이터 수집

- 상권 별 '네이버지도 조회수 상위 10개 업소'를 해당 상권의 고객 감성을 잘 반영하는 업소로 정의
- 상권 별로 상위 10개 업소의 '카테고리-빈도' 데이터를 수집하여 빈도 합산
- 일반적으로 자주 등장하거나, 분석에 유의하지 않은 카테고리 제거

제거한 카테고리
음식이 맛있어요, 친절해요, 술이 다양해요

리뷰 카테고리	빈도
기본 안주가 좋아요	54
특별한 메뉴가 있어요	43
가성비가 좋아요	33
대화하기 좋아요	32
오래 머무르기 좋아요	23
단체모임 하기 좋아요	20

<표 8> 수집 데이터: 네이버 리뷰 카테고리

### 리뷰 데이터 분석

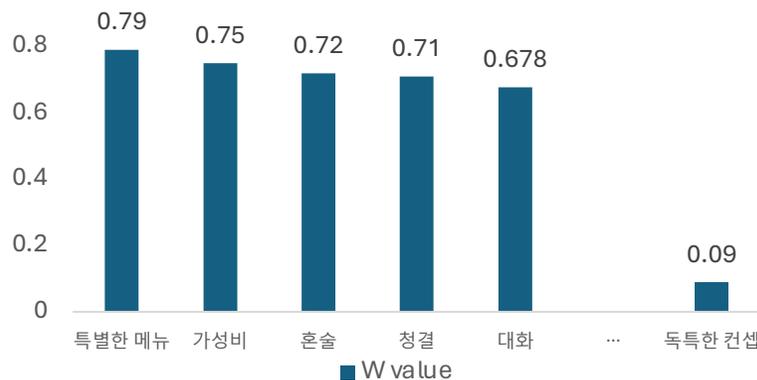
TF - IDF 방법론 선정

$$W_{i,j} = tf_{i,j} \times \log \left( 1 + \frac{\#Docs}{df_i} \right)$$

- 각 상권 별로 유의미한 카테고리를 발견하기 위함
- 여러 지역에서 일반적으로 등장하는 카테고리 보다 소수 지역에서 특별하게 발생하는 카테고리에 가중치를 둠

\* Term Frequency : 한 상권에서 해당 카테고리의 발생 빈도  
Doc Frequency : 해당 카테고리가 발생한 동의 수

카테고리 별 TF - IDF 결과 (등촌동)



<그림 9> 등촌동 TF-IDF 결과

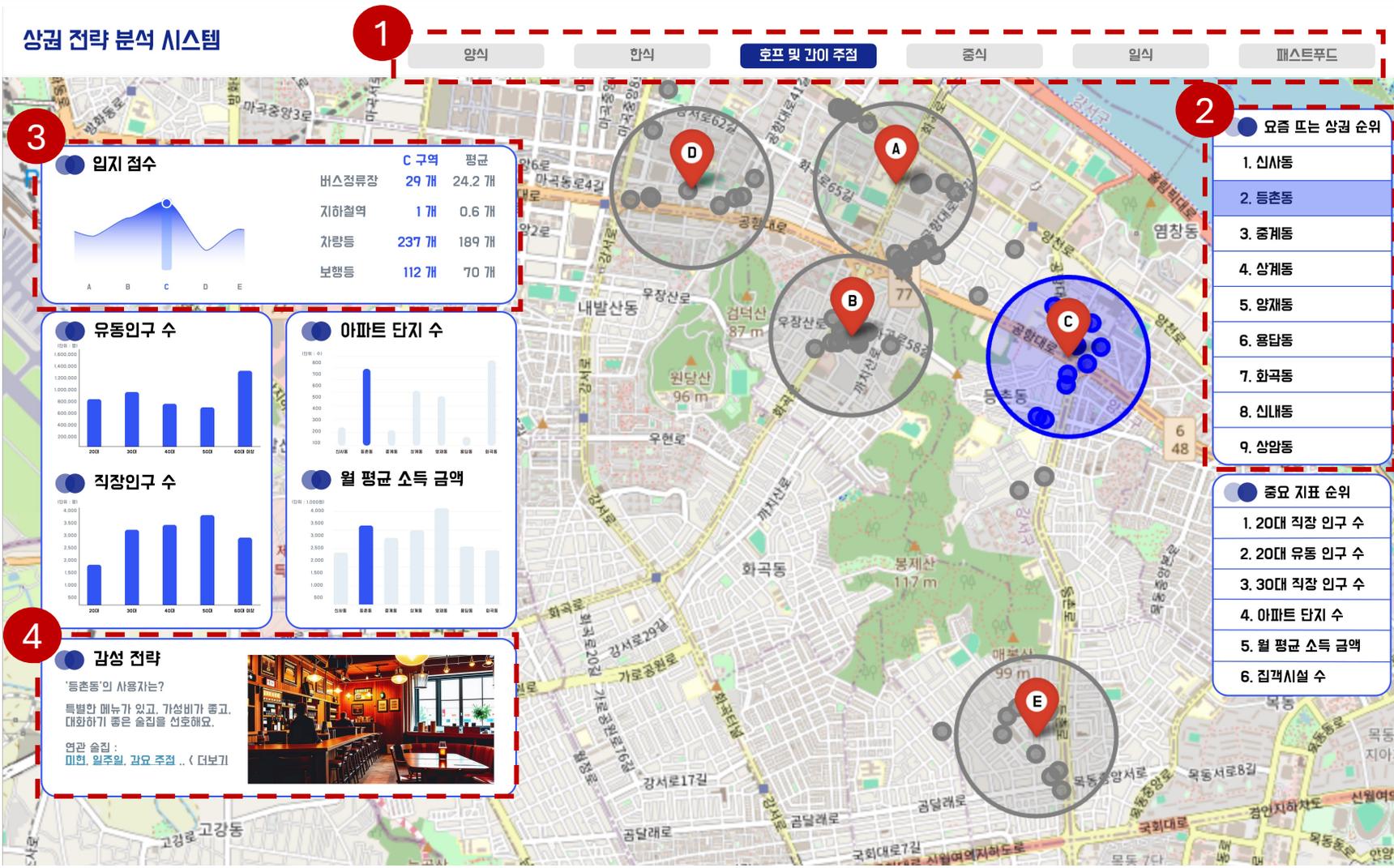
### 전략 사전 구축

동	상위 카테고리 ( W Value )
필동	대화하기 좋아요 (0.86), 기본 안주가 좋아요 (0.85), 가성비가 좋아요 (0.84),,,
을지로동	인테리어가 멋져요 (0.99), 특별한 메뉴가 있어요 (0.99), 가성비가 좋아요 (0.95),,,
명동	단체모임 하기 좋아요 (0.84), 대화하기 좋아요 (0.82), 기본 안주가 좋아요 (0.80),,,
등촌동	특별한 메뉴가 있어요 (0.79), 가성비가 좋아요 (0.75), 대화하기 좋아요 (0.72),,,
서교동	인테리어가 멋져요 (1.01), 룸이 잘 되어있어요 (1.05), 음악이 좋아요 (0.98),,,

<표 9> 동 별 상위 감성 카테고리 도출

- TF - IDF 분석 결과를 통해, 상권 별 적합한 감성 전략 도출. 해당 결과는 본 서비스의 **창업 지역 별 전략 추천**에 사용됨
- 을지로동에 방문하는 고객은 **인테리어, 분위기** 등을 중요시 하였으며, 서교동의 고객은 **룸이 잘 되어있는** 업소에 좋은 점수를 주었음

# 5-1. 연구 결과 - 최종 사용자 시나리오



1. 관심있는 외식업 업종 선택
2. 유망 상권 순위 제공 및 창업 희망지 선택
3. 가장 높은 입지 점수를 가지는 구역 정보 제공
4. 감성 전략 추천 및 연관 업소 정보 제공

**현직자 Interview: 외식업 종사자**

감성 전략 추천의 신뢰도가 중요할 것 같으나, 서교동(홍대)의 추천 전략이 인테리어, 룸 술집, 음악인 것으로 보았을 때 타당성이 있음. 궁극적으로 창업을 위해서는 현장에서 세밀하게 조사할 필요가 있지만, 그 전에 큰 틀을 잡기 위해서 참고할 때 유용한 것 같음

<그림 10> 최종 결과물 시나리오. 창업 입지 전략 추천 대시보드

## 5-2. 결론 및 향후 연구

### 산업공학적 의의

- 정보 수집 및 분석 시간을 최소화할 수 있는 효율적인 시스템 설계
- 데이터 분석에 체계적인 방법론을 적용하였고, 의사 결정에 반영
- 데이터 분석을 통해 얻은 자료에 대한 깊이 있는 이해를 바탕으로 유의미한 결과를 도출함

### 연구 실용성



<정부>

자영업 폐업률 감소로  
경제 활성화



<창업자>

합리적인 추천 시스템을  
통한 성공적 창업



<고객>

요구사항에 맞는 서비스로  
고객 만족도 증가

### 연구 한계점

- 수집 가능한 데이터 범위가 제한적이며, 보다 신뢰성 높은 입지 선정에 위해서는 다양한 요건이 추가적으로 고려되어야 함
- ‘호프 및 간이주점’ 내부의 여러 세부 카테고리 고려하지 못함
- 세부적인 전략 추천을 위해서는 더욱 다변화된 분석 필요

### 향후연구

#### 모델 향후 연구

- 개별 점포 데이터를 통한 개인화된 추천 시스템 고도화
- 시계열 분석을 통해 사용자에게 매출 예측값 제공

#### 서비스 향후 연구

- 비즈니스 모델을 결합한 결과물의 서비스화 방안 수립
- 다양한 업종을 추가적으로 분석함으로써 사용자 선택 확장

## \* Appendix - 출처

### 참고 문헌

- 1) 서울 휴게음식점 절반 3년내 폐업...재작년부터 폐업률 60%대 <https://www.yna.co.kr/view/AKR20200429054800003>
- 2) 서희정, & 박지연. (2019). 체육관 지도자의 마케팅커뮤니케이션이 관원생의 체육관 만족과 신뢰 그리고 충성도에 미치는 영향. 대한무도학회지, 21(4), 101-112.
- 3) 이영미. (2018). 외식기업의 사회적 책임활동이 브랜드자산, 브랜드 태도 및 구매의도에 미치는 영향: 신뢰의 조절효과를 중심으로.
- 4) 고희석. (2021). 외식기업의 브랜드 증거와 마케팅커뮤니케이션이 브랜드 태도, 신뢰 및 구매의도에 미치는 영향. 지역산업연구, 44(1), 143-165.
- 5) 핀다 사용자 10명 중 9.4명 “상권분석 데이터 도움됐다” <https://www.hankyung.com/article/2024010427555>
- 6) 김상호, 이계임, 홍연아, 문동현, 허성운, 이육직, & 신성용. (2020). 포용성장 및 지속가능성 관련 식품제조업 분야 실태조사 결과. 한국농촌경제연구원 기본연구보고서, 1-207.
- 7) 강진희. (2017). Huff 의 확률적 상권분석 모형과 ESRI 모형을 활용한 외식산업 상권분석모형 개발. 관광연구, 32(8), 189-211.
- 8) 김현철, & 이승일. (2019). 서울시 골목상권 매출액에 영향을 미치는 요인에 관한 연구. 서울시연구, 20(1), 117-134.
- 9) Nettleton, D. (2014). Commercial data mining: processing, analysis and modeling for predictive analytics projects. Elsevier.
- 10) Jung, S. S., & Jang, S. S. (2019). To cluster or not to cluster?: Understanding geographic clustering by restaurant segment. International Journal of Hospitality Management, 77, 448-457.
- 11) 한주옥, & 노태욱. (2016). AHP 를 활용한 주유소 입지 결정 요인에 관한 연구. 부동산연구, 26(4), 147-165.
- 12) 이동민. (2008). 교통이용자 측면 (Transportation User Perception) 을 고려한 교통공학연구. 교통기술과정책, 5(1), 246-250.

### Data 출처

- [서울 열린 데이터 광장] 서울시 상권분석서비스(상주인구-행정동) <https://data.seoul.go.kr/dataList/OA-22183/S/1/datasetView.do>
- [서울 열린 데이터 광장] 서울시 상권분석서비스(직장인구-행정동) <https://data.seoul.go.kr/dataList/OA-22184/S/1/datasetView.do>
- [서울 열린 데이터 광장] 서울시 상권분석서비스(길단위인구-행정동) <https://data.seoul.go.kr/dataList/OA-22178/S/1/datasetView.do>
- [서울 열린 데이터 광장] 서울시 상권분석서비스(추정매출-행정동) <https://data.seoul.go.kr/dataList/OA-22175/S/1/datasetView.do>
- [서울 열린 데이터 광장] 서울시 상권분석서비스(점포-행정동) <https://data.seoul.go.kr/dataList/OA-22172/S/1/datasetView.do>
- [서울 열린 데이터 광장] 서울시 상권분석서비스(집객시설-행정동) <https://data.seoul.go.kr/dataList/OA-22169/S/1/datasetView.do>
- [서울 열린 데이터 광장] 서울시 상권분석서비스(소득소비-행정동) <https://data.seoul.go.kr/dataList/OA-22166/S/1/datasetView.do>
- [서울 열린 데이터 광장] 서울시 상권분석서비스(아파트-행정동) <https://data.seoul.go.kr/dataList/OA-22163/S/1/datasetView.do>
- [서울 열린 데이터 광장] 서울시 상권분석서비스(영역-행정동) <https://data.seoul.go.kr/dataList/OA-22160/S/1/datasetView.do>
- [서울 열린 데이터 광장] 서울시 상권분석서비스(상권변화지표-행정동) <https://data.seoul.go.kr/dataList/OA-15575/S/1/datasetView.do>
- [공공데이터 포털] 서울특별시\_보행등 위도 경로 현황 <https://www.data.go.kr/data/15124273/fileData.do>
- [서울 열린 데이터 광장] 서울시 역사마스터 정보 <https://data.seoul.go.kr/dataList/OA-21232/S/1/datasetView.do>
- [서울 열린 데이터 광장] 서울시 버스정류소 위치정보 <https://data.seoul.go.kr/dataList/OA-15067/S/1/datasetView.do>
- [공공데이터 포털] 서울특별시\_신호등 데이터 <https://www.data.go.kr/data/15124269/fileData.do>
- 네이버지도 리뷰 크롤링 - 연관도 및 인기도 순 <https://zrr.kr/b0dg>

# \* Appendix - 일정

## Role & Responsibilities

Task	Day																														
	1주차		2주차		3주차		4주차		5주차		6주차		7주차		8주차		9주차		10주차		11주차		12주차		13주차		14주차		15주차		
	5	7	10	14	17	21	24	28	31	35	38	42	45	49	52	56	59	63	66	70	73	77	80	84	87	91	94	98	101	105	
목표설정	주제 선정	■	■	■	■																										
	문제 파악 및 수정																														
최종 제안서	구체적인 주제 선정 및 문제 정의																														
	요구사항 분석																														
	주제 관련 논문 수집																														
데이터 관리	모델 선정																														
	주제 관련 데이터 탐색																														
	데이터 정의 전 논문 조사																														
Data driven management 시스템 설계	데이터 탐색																														
	웹 데이터 크롤링																														
	개발환경 구축																														
테스트 및 검증	모델 탐색																														
	설명가능한 인공지능 구축																														
	개발환경 구축																														
실제 유저 사용 시나리오	입지분석																														
	군집 클러스터링																														
	개발환경 구축																														
실제 유저 사용 시나리오	감성 분석 근거 도출																														
	리뷰 데이터 크롤링																														
	TF-IDF 분석																														
실제 유저 사용 시나리오	감성 사전 구축																														
	감성 전략 선택의 주요 요인 도출																														
	모델 성능 평가																														
실제 유저 사용 시나리오	모델 수정 (정확도 향상)																														
	상인 인터뷰 평가																														
	ERRC Grid 생성 및 Value factor 선정																														
실제 유저 사용 시나리오	Value curve 생성																														
	UI 설계																														
	대시보드 설계																														

■	기존 일정 + 진행
■	초과 일정
■	남은 일정

## Work Breakdown Structure

Task	김민지	남경현	성호정	임형준	허재원
프로젝트 관리	일정관리	S		P	
	진행관리	P	S		
	변경관리		S		S
목표설정	주제 선정	S	P	S	S
	문제 파악 및 수정				P
최종 제안서	구체적인 주제 선정 및 문제 정의	P			
	요구사항 분석		P		
	주제 관련 논문 수집			P	
데이터 관리	모델 선정				P
	주제 관련 데이터 탐색	S			P
	데이터 정의 전 논문 조사	S	S	P	S
Data driven management 시스템 설계	데이터 탐색	P			S
	웹 데이터 크롤링				P
	개발환경 구축		P		
테스트 및 검증	모델 탐색		S		P
	설명가능한 인공지능 구축		P		
	개발환경 구축			P	S
실제 유저 사용 시나리오	입지분석			S	P
	군집 클러스터링			S	P
	개발환경 구축				P
실제 유저 사용 시나리오	감성 분석 근거 도출		P		S
	리뷰 데이터 크롤링		S		P
	TF-IDF 분석				P
실제 유저 사용 시나리오	감성 사전 구축				P
	감성 전략 선택의 주요 요인 도출	P			S
	모델 성능 평가		P		S
실제 유저 사용 시나리오	모델 수정 (정확도 향상)		S		
	상인 인터뷰 평가	S	S	S	P
	ERRC Grid 생성 및 Value factor 선정	P			
실제 유저 사용 시나리오	Value curve 생성	P		S	
	UI 설계	S	S	P	
	대시보드 설계	P		S	

P	주책임자
S	보조책임자

# \* Appendix - 머신러닝 매출 요인 분석(1/3)

## 데이터 수집

**서울시 상권분석서비스(직장인구-행정동)** 데이터

서울시 행정동 내 직장이 위치한 인구 정보를 제공합니다. 1년중 4분기에 한번 데이터를 업데이트 하...  
 수정일자: 2024-04-30    제공기관: 서울신용보증재단    제공부서: 빅데이터팀

[일반행정]    SHEET    OpenAPI

**서울시 상권분석서비스(추정매출-행정동)** 데이터

서울시 행정동별 점포들의 추정 매출 정보를 제공합니다. ※ 단위: 원 ※ 행정동 코드는 행정안전부...  
 수정일자: 2024-04-30    제공기관: 서울신용보증재단    제공부서: 빅데이터팀

[일반행정]    SHEET    OpenAPI    FILE

**서울시 상권분석서비스(길단위인구-행정동)** 데이터

서울시 행정동별 생활인구 정보를 제공합니다. ※ 행정동 코드는 행정안전부에서 고시한 "주민등록 행...  
 수정일자: 2024-04-30    제공기관: 서울신용보증재단    제공부서: 빅데이터팀

[일반행정]    SHEET    OpenAPI

**서울시 상권분석서비스(영역-행정동)** 데이터

서울시 상권분석 서비스에서 사용중인 행정동 영역 정보입니다. ※ 행정동 코드는 행정안전부에서 고시...  
 수정일자: 2024-04-30    제공기관: 서울신용보증재단    제공부서: 빅데이터팀

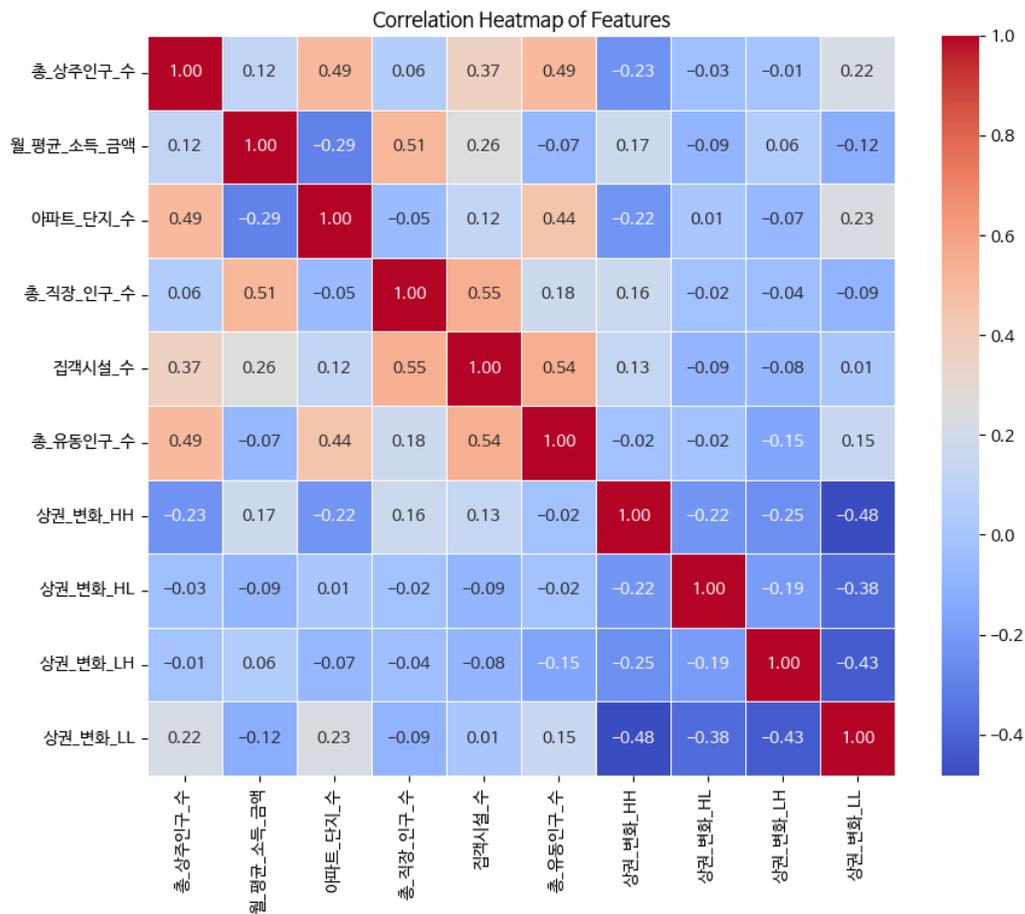
[일반행정]    SHEET    OpenAPI    FILE

**서울시 상권분석서비스(직장인구-행정동)** 데이터

서울시 행정동 내 직장이 위치한 인구 정보를 제공합니다. 1년중 4분기에 한번 데이터를 업데이트 하...  
 수정일자: 2024-04-30    제공기관: 서울신용보증재단    제공부서: 빅데이터팀

[일반행정]    SHEET    OpenAPI

## 상관성 분석



수집한 변수를 기반으로  
상관성 분석을 진행

상관성 분석을 통해  
유의미한 변수만 남긴 후

해당 변수를 연령대로  
나눠서 진행

## \* Appendix - 머신러닝 매출 요인 분석(2/3)

### 최종 데이터 프레임

	연령대_20_상 주인구_수	연령대_30_상 주인구_수	연령대_40_상 주인구_수	연령대_50_상 주인구_수	연령대_60_이상 상주인구_수	연령대_20_직 장_인구_수	연령대_30_직장 _인구_수	연령대_40_직 장_인구_수	연령대_50_직 장_인구_수	연령대_60_이상 _직장_인 구_수	...	연령대_30_유 동인구_수	연령대_40_유 동인구_수	연령대_50_유 동인구_수	연령대_60_이 상_유동인 구_수	월_평균_소 득_금액	아파트_단 지_수	집객_시설_수	상권_변화_지 표	당월_매출_금 액	점 포_수
0	1594	1697	2148	1979	3023	1929	3064	2954	1874	742	...	571066	663192	539268	747777	3780222	342	93	LH	137990626	10
1	1238	1443	1476	1562	2538	6303	13536	14415	9470	4045	...	792707	756825	593110	774465	5007768	98	206	HH	1641273647	57
2	320	360	369	413	878	726	1549	1817	1381	386	...	168946	164441	137586	183076	3442534	11	49	LH	117797848	5
3	1340	1236	1628	1723	2609	695	1179	1099	798	336	...	152579	218548	207388	348752	3647449	183	84	LH	28529570	5
4	2450	2233	2767	3266	4859	2878	6197	7579	5016	1464	...	114667	167476	178642	293519	4504006	297	162	HH	127152245	15
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3338	2716	3393	3322	3043	4423	2020	3567	3667	4407	3369	...	996514	1030838	924847	1388981	3438376	359	98	HL	296573135	18
3339	3995	4327	3504	3832	6328	391	474	511	467	454	...	1538475	1440208	1361854	2266805	2824592	325	113	LL	524863259	19
3340	3628	3584	3615	3975	5565	658	1034	1162	962	810	...	1049992	1133700	1091806	1540116	2897346	456	136	HL	574120008	42
3341	7189	7606	6859	7605	11733	1369	1886	2128	2298	1762	...	2502960	2526266	2543632	4186942	2840603	624	242	LL	767759791	68
3342	3746	3861	3901	4390	6438	946	1293	1058	1026	634	...	1043582	1119880	1141680	1838115	3414406	315	129	HL	128822259	15

3343 rows x 21 columns

### 모델 학습

```

y = df['당월_매출_금액'] / df['점포_수']
X = df.drop(columns=['당월_매출_금액', '점포_수'])
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
models = {
    "Random Forest": RandomForestRegressor(n_estimators=100, random_state=42),
    "Gradient Boosting": GradientBoostingRegressor(n_estimators=100, random_state=42),
    "Extra Trees": ExtraTreesRegressor(n_estimators=100, random_state=42),
    "LightGBM": LGBMRegressor(n_estimators=100, random_state=42)
}
r2_scores = {}

```

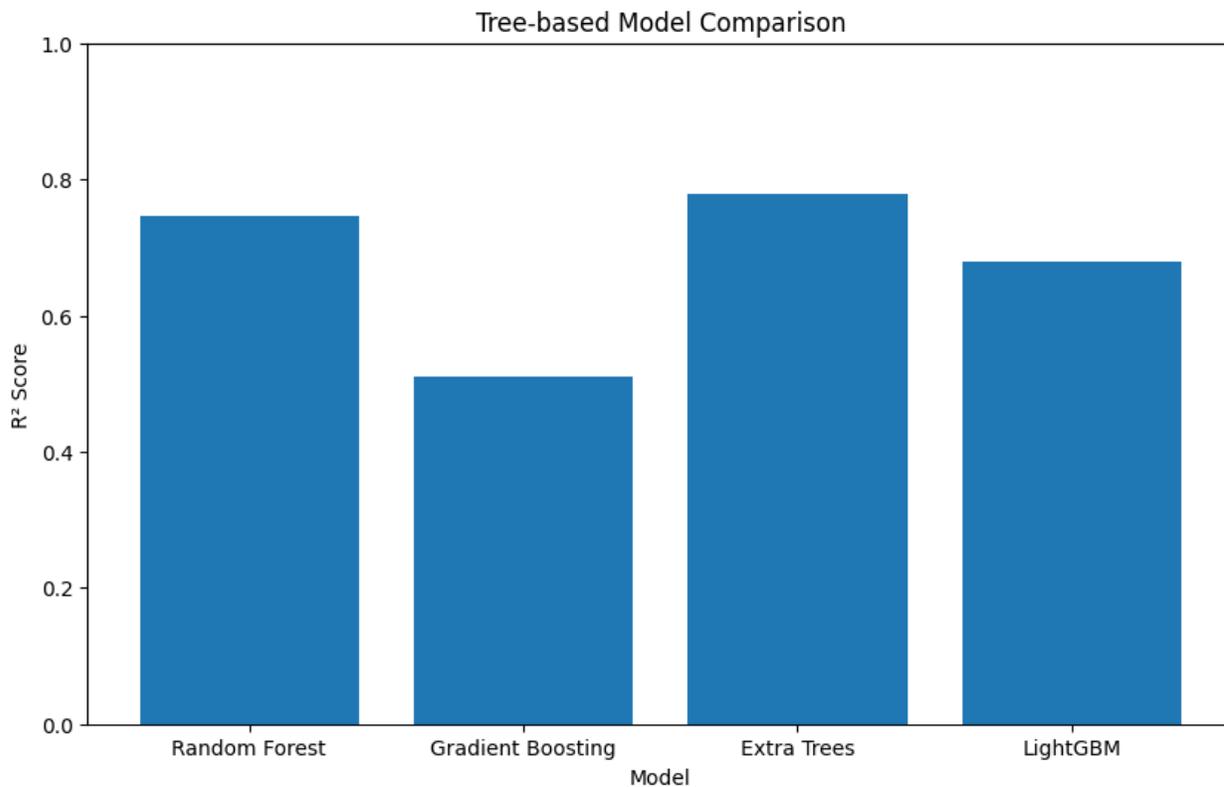
```

for name, model in models.items():
    model.fit(X_train, y_train)
    y_pred = model.predict(X_test)
    r2_scores[name] = r2_score(y_test, y_pred)
plt.figure(figsize=(10, 6))
plt.bar(r2_scores.keys(), r2_scores.values())
plt.xlabel('Model')
plt.ylabel('R^2 Score')
plt.ylim(0, 1)
plt.title('Tree-based Model Comparison')
plt.show()

```

## \* Appendix - 머신러닝 매출 요인 분석(3/3)

### 모델 학습 결과



### Feature 중요도 도출

```
importance = {
  '연령대_20_직장_인구_수': 28.102994,
  '연령대_20_유동인구_수': 8.572773,
  '연령대_30_직장_인구_수': 8.165808,
  '아파트_단지_수': 6.938153,
  '월_평균_소득_금액': 6.676604,
  '연령대_60_이상_유동인구_수': 6.416639,
  '연령대_20_상주인구_수': 4.831853,
  '연령대_30_상주인구_수': 4.374774,
  '연령대_40_유동인구_수': 2.911915,
  '연령대_50_직장_인구_수': 2.798844,
  '연령대_30_유동인구_수': 2.780336,
  '연령대_50_상주인구_수': 2.735831,
  '연령대_60_이상_직장_인구_수': 2.550702,
  '연령대_60_이상_상주인구_수': 2.163323,
  '집객시설_수': 2.117845,
  '연령대_40_직장_인구_수': 1.894948,
  '연령대_50_유동인구_수': 1.862210,
  '연령대_40_상주인구_수': 1.417412,
}
```

## \* Appendix - 클러스터링 및 입지 점수 도출

### 등촌동 술집의 좌표

	술집	위도	경도
0	인생맥주 등촌역점	37.550379	126.863713
1	미현	37.548926	126.862022
2	쌍백리	37.552036	126.865623
3	감요 주점	37.552837	126.863859
4	일주일	37.555589	126.854404
...	...	...	...
74	란(이자카야)	37.552098	126.853720
75	가야	37.552234	126.850723
76	홀릭빠	37.551984	126.849464
77	꾼노리	37.552807	126.850930
78	M2BAR	37.553358	126.850476

79 rows x 3 columns

### 클러스터링 코드

```

import folium
import numpy as np

# 서울시청을 중심으로 하는 지도 생성
map = folium.Map(location=[37.5665, 126.9780], zoom_start=11)

# 각 클러스터에 다른 색상을 할당
colors = ['blue', 'gray', 'gray', 'gray', 'gray', 'gray'] # 클러스터 수에 따라 색상을 추가할 수 있습니다.

# 각 점을 지도에 추가
for _, row in stores_df1.iterrows():
    if row['cluster'] != -2: # 노이즈 포인트 제외
        folium.CircleMarker([row['위도'], row['경도']],
                             radius=5,
                             color=colors[row['cluster'] % len(colors)],
                             fill=True,
                             fill_color=colors[row['cluster'] % len(colors)],
                             fill_opacity=0.7).add_to(map)

# 클러스터 중심 및 반경 300m의 원 시각화
for cluster in np.unique(stores_df1['cluster']):
    if cluster != -1: # 노이즈 포인트는 제외
        cluster_points = stores_df1[stores_df1['cluster'] == cluster][['위도', '경도']].values
        if len(cluster_points) > 0:
            cluster_center = np.mean(cluster_points, axis=0)

            # 클러스터 중심에 마커 추가
            folium.Marker(cluster_center,
                           icon=folium.Icon(color='red', icon='info-sign'),
                           popup=f'Cluster {cluster} Center').add_to(map)

            # 반경 300m의 원 추가
            folium.Circle(cluster_center,
                           radius=400, # 반경 300m
                           color=colors[cluster % len(colors)],
                           fill=True,
                           fill_color=colors[cluster % len(colors)],
                           fill_opacity=0.2).add_to(map)

# 지도를 출력
map

```

### 입지 점수 도출 코드

```

# 각 가게마다 주변 보행등 개수 계산 함수
def count_nearby_traffic_lights(store_lat, store_lon, threshold=0.5):
    """주어진 가게 위치에서 threshold km 이내에 있는 보행등의 개수를 반환합니다."""
    count = 0
    for _, traffic_light in traffic_lights_df.iterrows():
        traffic_light_location = (traffic_light['위도'], traffic_light['경도'])
        store_location = (store_lat, store_lon)
        distance = great_circle(traffic_light_location, store_location).km
        if distance <= threshold:
            count += 1
    return count

# 가게마다 주변 보행등 개수 계산 및 결과 저장
stores_df['보행등'] = stores_df.apply(lambda row: count_nearby_traffic_lights(row['위도'], row['경도']), axis=1)

# 결과 출력
stores_df

```

cluster	위도	경도	버스정류장	지하철역	차량등	보행등	
0	0	37.551624	126.863806	13	1	122	50
1	1	37.559461	126.853967	25	1	231	93
2	2	37.559245	126.842427	29	1	237	112
3	3	37.534487	126.861982	23	0	102	26
4	4	37.552586	126.851461	31	0	253	69

# \* Appendix - 리뷰 분석

## 수집 리뷰 데이터 프레임

😊 "음식이 맛있어요"	544
👉 "술이 다양해요"	311
💬 "인테리어가 멋져요"	289
👉 "특별한 메뉴가 있어요"	238
👉 "대화하기 좋아요"	156
💖 "친절해요"	141
👉 "매장이 청결해요"	113
👉 "기본 안주가 좋아요"	111
🎵 "음악이 좋아요"	98

```

driver = webdriver.Chrome()
df = pd.read_excel("점포_리뷰_리뷰.xlsx")
print(df)

for idx, row in df.iterrows():
    if not pd.isnull(row['리뷰']): continue

    url = row['주소']
    driver.get(url)

    iframe = WebDriverWait(driver, 10).until(EC.presence_of_element_located((By.CSS_SELECTOR, "#entryIframe")))
    driver.switch_to.frame(iframe)
    time.sleep(2)

    while True:
        try:
            bt_spread = WebDriverWait(driver, 2).until(EC.presence_of_element_located((By.CLASS_NAME, "dP0sq")))
            bt_spread.click()
        except:
            break

    review = ""
    elements1 = driver.find_elements(By.CLASS_NAME, "t315F")
    elements2 = driver.find_elements(By.CLASS_NAME, "Cl0Ly")

    for e1, e2 in zip(elements1, elements2):
        word = e1.text.replace("\n", "")
        score = e2.text.split("\n")[1]
        review += word + "-" + score + "/"
    
```

장소	점포명	순위	위도	경도	주소	리뷰
필동	보스호프	1	37.5615	126.9968	https://ma	음식이 맛있어요-361/친절해요-229/특별한
필동	하안집 3호	2	37.5616	126.9967	https://ma	가성비가 좋아요-144/음식이 맛있어요-139
필동	하안집 2호	4	37.56131	126.9966	https://ma	가성비가 좋아요-91/음식이 맛있어요-78/친
필동	한남북옛국	3	37.55816	126.9959	https://ma	음식이 맛있어요-304/술이 다양해요-141/친
필동	브아브아	5	37.56143	126.9939	https://ma	음식이 맛있어요-181/술이 다양해요-95/친
필동	시오	6	37.56131	126.9968	https://ma	음식이 맛있어요-607/인테리어가 멋져요-3
필동	술무로	7	37.56122	126.9967	https://ma	음식이 맛있어요-100/기본 안주가 좋아요-
필동	오매가매	8	37.56096	126.9925	https://ma	음식이 맛있어요-329/친절해요-142/단체도
필동	미식싸롱	9	37.56279	126.9919	https://ma	음식이 맛있어요-26/술이 다양해요-20/친
필동	보금보	10	37.56216	126.9906	https://ma	음식이 맛있어요-109/인테리어가 멋져요-1
을지로동	주마등	1	37.56535	126.9913	https://ma	음식이 맛있어요-230/가성비가 좋아요-99/
을지로동	을지로 골	2	37.56556	126.99	https://ma	음식이 맛있어요-1002/특별한 메뉴가 있어

## TF-IDF 분석

```

# 장소 별 카테고리 발생 횟수 합산
for row in df.iloc():
    review = dic[row["주소"]]
    rev_text = row["리뷰"].split("/")[:-1]

    for r in rev_text:
        term = r.split("-")[0]
        cnt = int(r.split("-")[1])
        if term in ["음식이 맛있어요", "친절해요", "술이 다양해요"]: continue

        if term not in dic_terms: dic_terms[term] = 0
        if term in review: review[term] += cnt
        else: review[term] = cnt

docs = len(dic)
TF_IDF = dic

# 해당 term이 있는 doc의 수 count(IDF)
for doc in TF_IDF.keys():
    for term in dic_terms.keys():
        if term in TF_IDF[doc]: dic_terms[term] += 1
print(dic_terms)

# TF-IDF
for doc in TF_IDF.keys():
    d = TF_IDF[doc]
    for term in d.keys():
        #TF-IDF 계산
        cnt = d[term]
        idf = math.log10(1+ docs/dic_terms[term])
        d[term] = math.log10(1+cnt) * idf
    
```

```

을지로동
[('인테리어가 멋져요', 0.9924611193760158), ('특별한 메뉴가 있어요', 0.9985331239393501), ('가성비가 좋아요', 0.9493010984055524), ('기본 안주가 좋아요', 0.9465174091519896), ('음악이 좋아요', 0.9279027759826597), ('대화하기 좋아요', 0.8940234533395463), ('단체모임 하기 좋아요', 0.847972550553187), ('매장이 청결해요', 0.808488239260127), ('오래 머무르기 좋아요', 0.7289754196055052), ('재료가 신선해요', 0.6939799127484703), ('양이 많아요', 0.6907031581125757), ('매장이 넓어요', 0.6859731866141117), ('메뉴 구성이 알차요', 0.680349280125435), ('혼술 하기 좋아요', 0.6721988094290369), ('화장실이 깨끗해요', 0.6626865667888862), ('특별한 날 가기 좋아요', 0.6460489132745685), ('컨셉이 독특해요', 0.6203318619464232), ('뷰가 좋아요', 0.5959396577395639), ('음식이 빨리 나와요', 0.5591127933472761), ('아늑해요', 0.5457412998893724), ('야외 공간이 멋져요', 0.5442398126445194), ('사진이 잘 나와요', 0.5437143497367392), ('혼밥하기 좋아요', 0.52670863655223941), ('주차하기 편해요', 0.43088342771835975), ('음료가 맛있어요', 0.37040203346725215), ('건강한 맛이에요', 0.3624762331578262), ('비싼 만큼 가치있어요', 0.3353303625148477), ('항신료가 강하지 않아요', 0.3353303625148477), ('좌석이 편해요', 0.2718571748683696), ('차분한 분위기에요', 0.270714608138923), ('파티하기 좋아요', 0.23958375817013625), ('잡내가 적어요', 0.21041093737452468), ('라이브공연이 훌륭해요', 0.18986563624075592), ('코스요리가 알차요', 0.1812381165789131), ('집중하기 좋아요', 0.14362780923945326), ('롤이 잘 되어있어요', 0.14362780923945326), ('아이와 가기 좋아요', 0.14362780923945326), ('반려동물과 가기 좋아요', 0.1305433316289997)]
    
```

## \* Appendix - 인터뷰

### 인터뷰 현장 사진 자료

