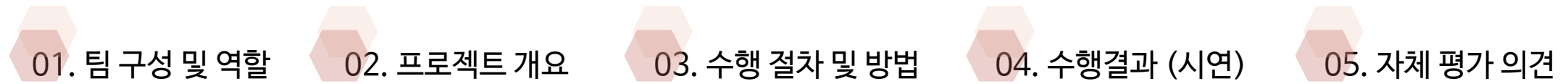


용산구 상권 매출 예측 서비스

- 
01. 팀 구성 및 역할
 02. 프로젝트 개요
 03. 수행 절차 및 방법
 04. 수행결과 (시연)
 05. 자체 평가 의견

용산구 상권 매출 예측 서비스



- 로그인/회원가입
- 지도 API 연동
- Git 관리
- 보고서 작성

- 데이터 수집
- 데이터 분석
- 프론트 참견
- PPT 제작, 발표
- 보고서 작성

데이터 수집
데이터 분석
프론트 적극 참견



웹 디자인
웹 퍼블리싱
웹 화면 구현

지도 API 연동
웹 화면 구현

기능정의서 작성

[illegible]

OS



Framework

django

Database



MariaDB

IDE



Language / Library



Chart.js

Collaboration



- 프로젝트 github
- 프로젝트 github (data분석)



Figma

- 프로젝트 Figma

Yongsan

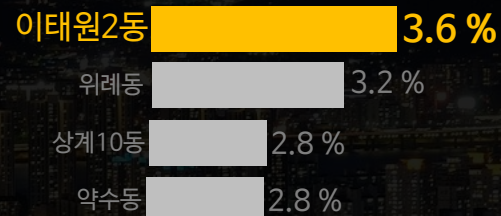
An aerial night photograph of the Yongsan District in Seoul, South Korea. The image shows a dense urban landscape with numerous high-rise apartment buildings and commercial structures, all illuminated by city lights. In the background, the dark silhouette of a mountain is visible, topped with the brightly lit N Seoul Tower. The word "Yongsan" is superimposed in a large, white, sans-serif font in the center of the image.

용산·이태원·한남동 지역은
대통령 집무실 및 미군기지 이전으로 인해
이태원을 비롯한 용산구가 주목을 받고 있다.

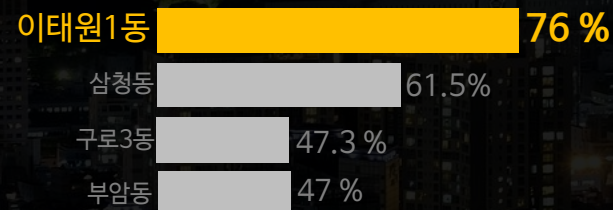
‘우리마을가게 상권분석서비스’에 따르면
‘뜨는 동네’ 카테고리에서는 TOP10에서도 상위권을 보이고 있으나,
‘뜨는상권’ 카테고리에서는 하위권을 보이고 있다.



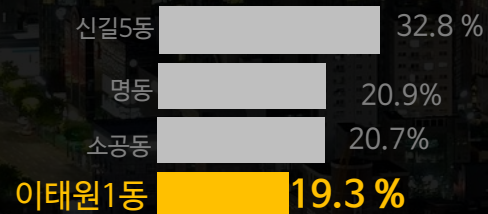
점포수 1위



매출액 1위



유동인구 4위



[출처] 우리마을가게 상권분석서비스 뜨는상권 순위TOP10

용산구 상권활성화 및 예비창업자를 위한 매출 예측 서비스

의사결정에 도움을 줄 수 있는 다양한 상권분석 정보를 제공하여, 상권활성화 및 예비창업을 지원하기 위함

용산구 지역의 예비 창업자와 자영업자에게 운영 및 의사결정에 유용한 상권정보 제공을 통해 용산구 상권 활성화 방안 모색 및 정책 수립 및 집행 과정에 활용

탐색적 데이터 분석을 수행하기 위한 데이터 정제, 시각화 방법 학습



데이터 출처

DATA

서울열린데이터광장 (클릭 시 링크로 이동됩니다)

- 우리마을가게_상권분석서비스(신_상권_추정매출)_2019년.csv
- 우리마을가게_상권분석서비스(신_상권_추정매출)_2020년.csv
- 우리마을가게_상권분석서비스(신_상권_추정매출)_2021년.csv

FRONT END

서울열린데이터광장 (클릭 시 링크로 이동됩니다)

서울시 우리마을가게 상권분석서비스 상권영역 4개 유형
꼭짓점 좌표 데이터(TBGIS_TRDAR_RELM.shp)



칼럼 목록

지역을 나타내는 상권_구분_코드_명, 상권_코드_명,
업종을 나타내는 서비스_업종_코드_명
매출액을 나타내는 분기별_매출, 금액_요일별, 시간대별, 성별에 따른 매출금액
결제 건수를 나타내는 분기별_매출_건수, 요일별_매출_건수 등 총 80개의 칼럼이 있다.

#	Column	Non-Null Count	Dtype
0	기준_분기_코드	14504 non-null	int64
1	상권_구분_코드_명	14504 non-null	object
2	상권_코드_명	14504 non-null	object
3	서비스_업종_코드_명	14504 non-null	object
4	분기별_매출_금액	14504 non-null	float64
5	분기별_매출_건수	14504 non-null	int64
6	주중_매출_금액	14504 non-null	float64
7	주말_매출_금액	14504 non-null	float64
8	월요일_매출_금액	14504 non-null	float64
9	월요일_매출_건수	14504 non-null	int64
10	화요일_매출_금액	14504 non-null	float64
11	화요일_매출_건수	14504 non-null	int64
12	수요일_매출_금액	14504 non-null	float64
13	수요일_매출_건수	14504 non-null	int64
14	목요일_매출_금액	14504 non-null	float64
15	목요일_매출_건수	14504 non-null	int64
16	금요일_매출_금액	14504 non-null	float64
17	금요일_매출_건수	14504 non-null	int64
18	토요일_매출_금액	14504 non-null	float64
19	토요일_매출_건수	14504 non-null	int64
20	일요일_매출_금액	14504 non-null	float64
21	일요일_매출_건수	14504 non-null	int64
22	시간대_매출_금액	14504 non-null	float64
23	시간대_매출_건수	14504 non-null	int64
24	성별_매출_금액	14504 non-null	float64
25	성별_매출_건수	14504 non-null	int64
26	연령대_10_매출_금액	14504 non-null	float64
27	연령대_10_매출_건수	14504 non-null	int64
28	연령대_20_매출_금액	14504 non-null	float64
29	연령대_20_매출_건수	14504 non-null	int64
30	연령대_30_매출_금액	14504 non-null	float64
31	연령대_30_매출_건수	14504 non-null	int64
32	연령대_40_매출_금액	14504 non-null	float64
33	연령대_40_매출_건수	14504 non-null	int64
34	연령대_50_매출_금액	14504 non-null	float64
35	연령대_50_매출_건수	14504 non-null	int64
36	연령대_60_이상_매출_금액	14504 non-null	float64
37	연령대_60_이상_매출_건수	14504 non-null	int64
38	점포수	14504 non-null	int64

dtypes: float64(47), int64(3), object(3)



데이터 정의

매출금액, 건수, 점포수는 float type으로 되어있으며 그 외 상권이나 업종을 의미하는 칼럼은 object로 되어있다.

상권_구분_코드_명	
골목상권	골목상권 중 점포 밀집도가 높은 상권으로 총 1,090개 골목상권
발달상권	유통산업발전법 제5조의 법조항에 따라 2천 제곱미터 이내 50개 이상의 상점이 분포하는 경우 “상점가”라 하고, 배후지를 고려하지 않은 도보이동이 가능한 범위내의 상가업소밀집지역을 발달상권으로 정의
전통시장	오랜 기간에 걸쳐 일정한 지역에서 자연발생적으로 형성된 상설시장이나 정기시장
관광특구	관광활동이 주로 이루어지는 지역적 공간 內 입지한 상권

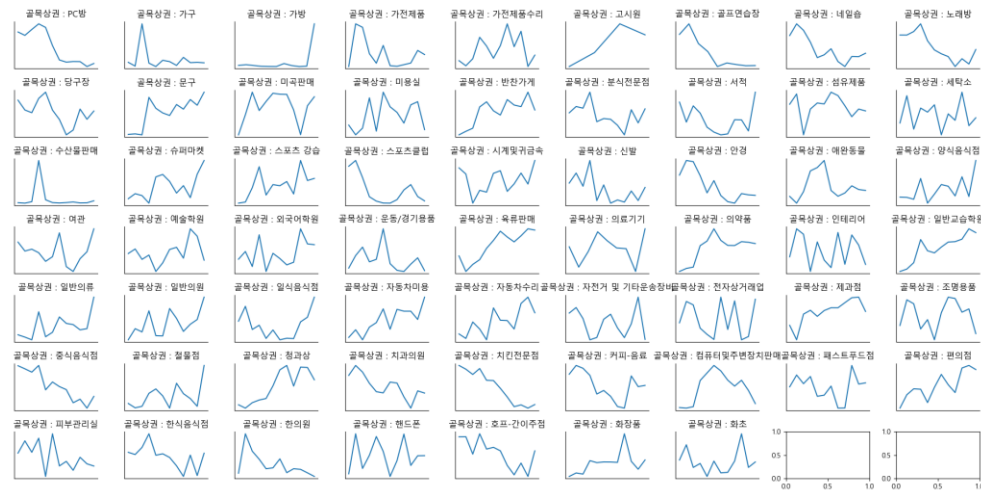
상권_코드_명	
용산세무서	실제 해당 지역을 대표하는 건물들의 명칭으로 대표 건물 주위의 상권들을 의미함
한남초등학교	
이태원(이태원역)	
⋮	

서비스_업종_코드_명	
한식음식점	사업체수가 많고, 종사자수가 5인 미만인 소규모 사업체 비중이 높으며, 창업 등 진출입이 용이한 업종으로 정의
양식음식점	
PC방	
⋮	100개 생활밀접업종정의.pdf

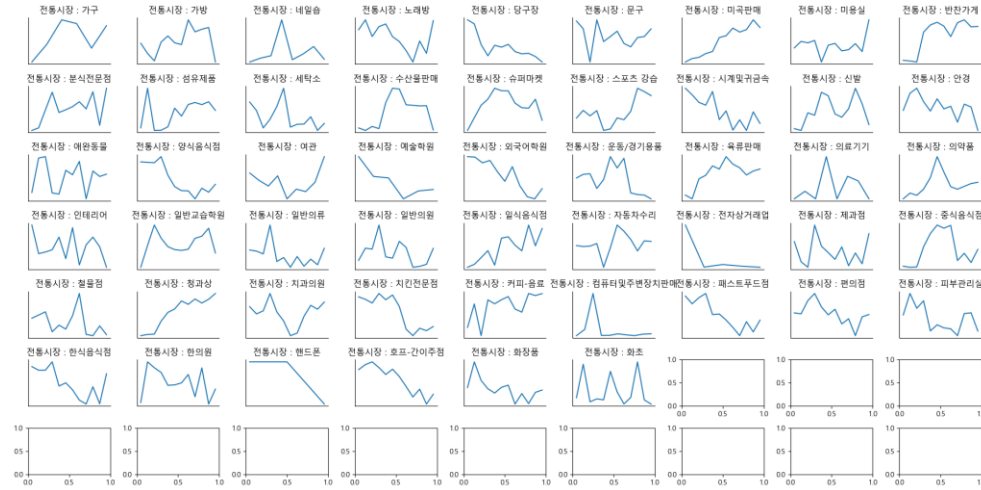
데이터 탐색

골목상권, 발달상권, 전통시장, 관광특구 네 곳 상권의 서비스업종의 매출을 확인 매출이 급감 급증하는 구간이 존재하는 것을 확인

[골목상권 | 분기별 (2019년 1분기 ~ 2021년 4분기) 매출 추이 그래프]



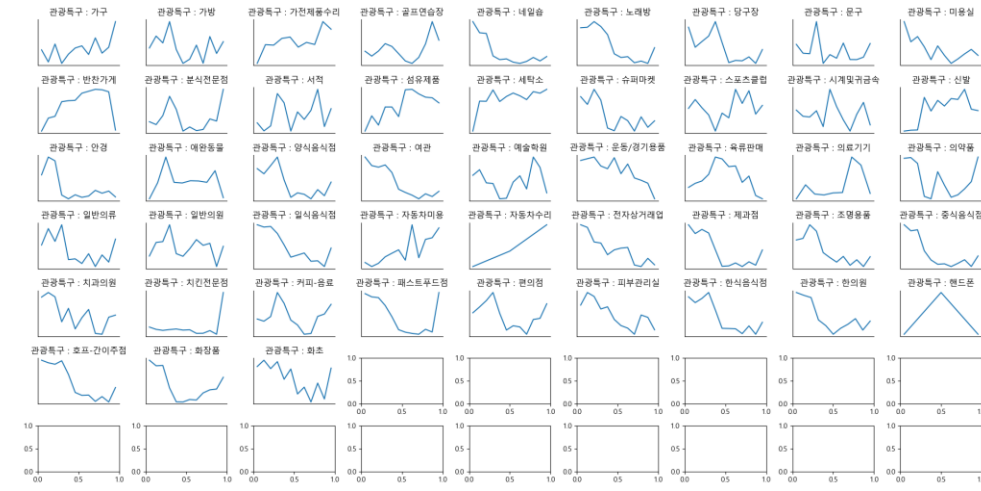
[전통시장 | 분기별 (2019년 1분기 ~ 2021년 4분기) 매출 추이 그래프]



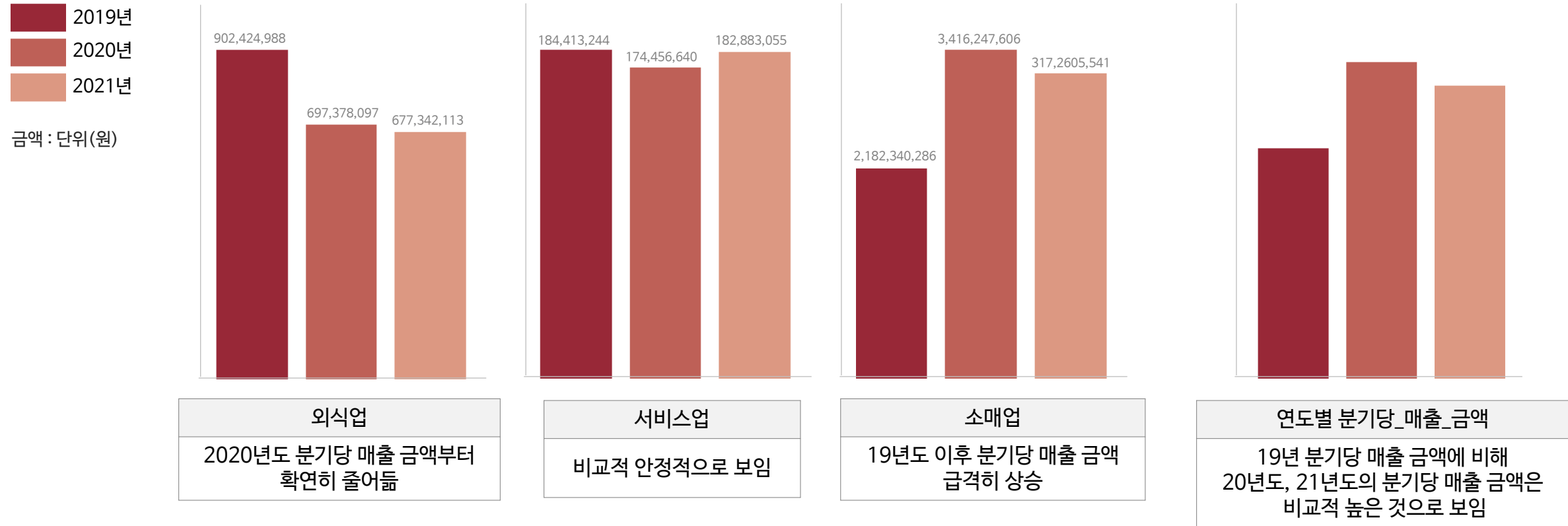
[발달상권 | 분기별 (2019년 1분기 ~ 2021년 4분기) 매출 추이 그래프]



[관광특구 | 분기별 (2019년 1분기 ~ 2021년 4분기) 매출 추이 그래프]



서비스업종(대분류)별 매출금액 변화 추이



서비스 업종별 매출금액변화 추이를 살펴보면 **코로나 거리두기 2단계 초과로 시행한 기간**
매출이 **급감하거나 급증**하는 구간이 일치하는 것을 확인할 수 있다.

* 코로나 거리두기 2단계 초과 시행 기간: 2020.08 ~
[출처] 질병관리청_사회적 거리두기 시행연혁_20220901.hwp

그러나,

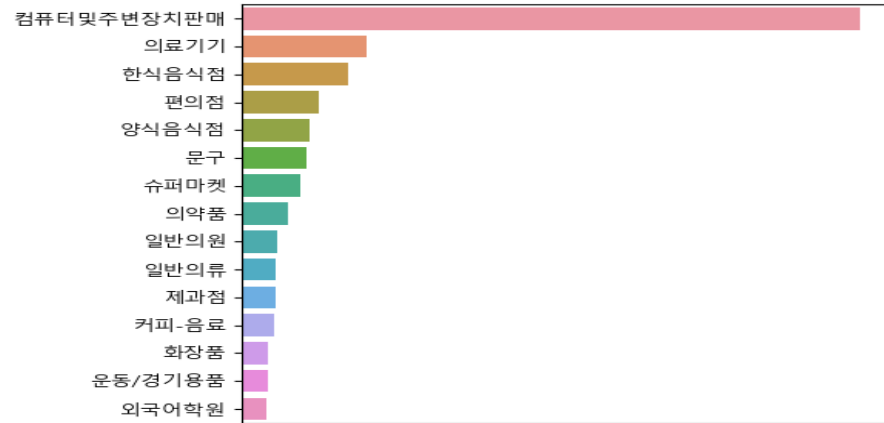
각 시기의 거리두기 단계를 새로운 칼럼을 추가하여
분기당 매출 금액과의 상관관계를 확인해보았으나
관계성이 낮은 관계로 독립변수에서 다시 제외

관련자료 링크

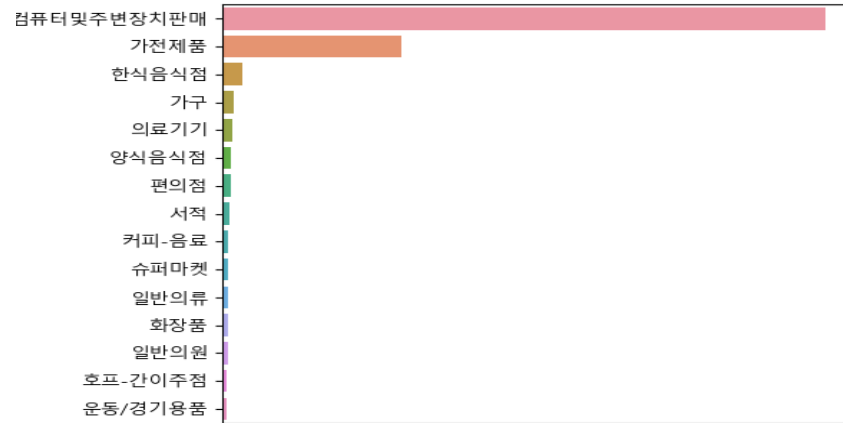
‘신용산역’, ‘용산 전자상가’가 속해 있는 골목상권, 발달상권에는 컴퓨터 및 주변장치판매의 매출비율이 월등히 높은것을 볼 수 있다.

이후 분석에서 이상치를 꼼꼼하게 확인해 보아야할 것으로 보인다.

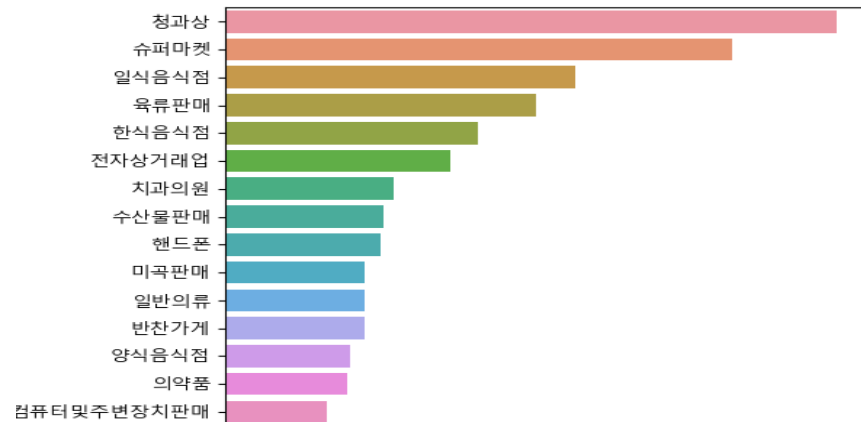
[골목상권 서비스업종별 매출]



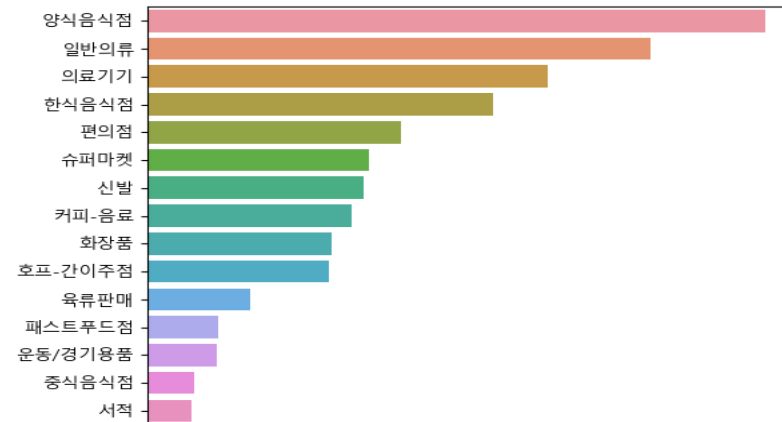
[발달상권 서비스업종별 매출]



[전통시장 서비스업종별 매출]



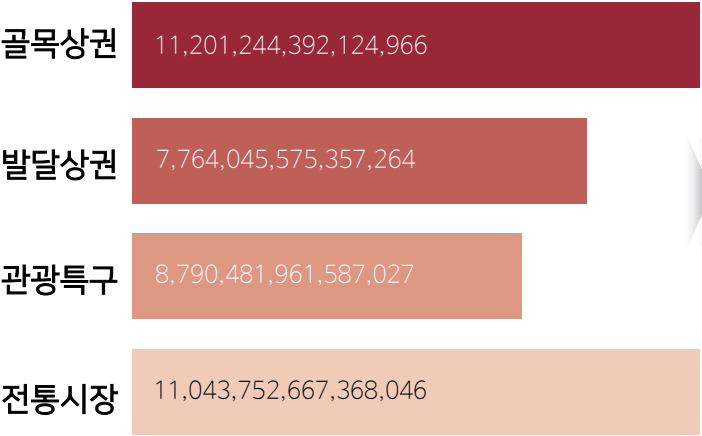
[관광특구 서비스업종별 매출]



03. 수행 절차 및 방법

정규성 확인 > 등분산성 확인 > ANOVA > 사후검정

[분기당_매출_금액(평균)]



[shapiro]

상권 구분 명	P 값	정규성	
골목상권	0.0	< 0. 05	만족
발달상권	2.687035992140221e-37	< 0. 05	만족
관광특구	5.306251271708362e-11	< 0. 05	만족
전통시장	3.635869046624615e-33	< 0. 05	만족

[levene]

등분산성		
1.007750333420473e-133	< 0.05	불만족

[welch_anova]

가설검증		
0.000	< 0.05	만족

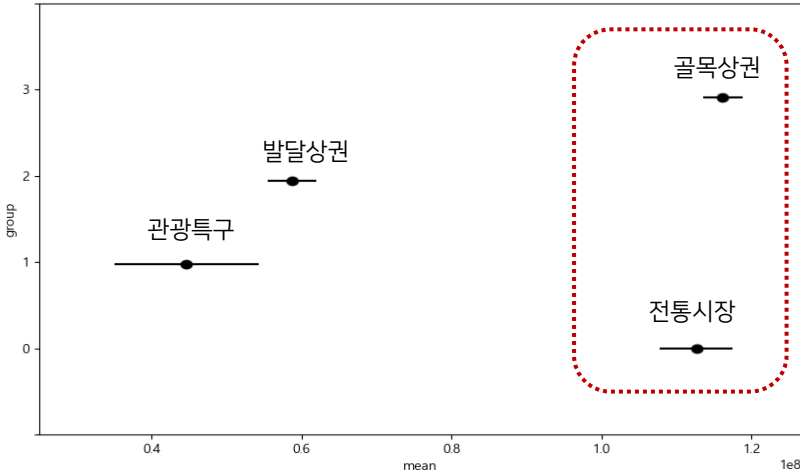
귀무기각! 사후검정 진행

[pairwise_tukeyhsd]

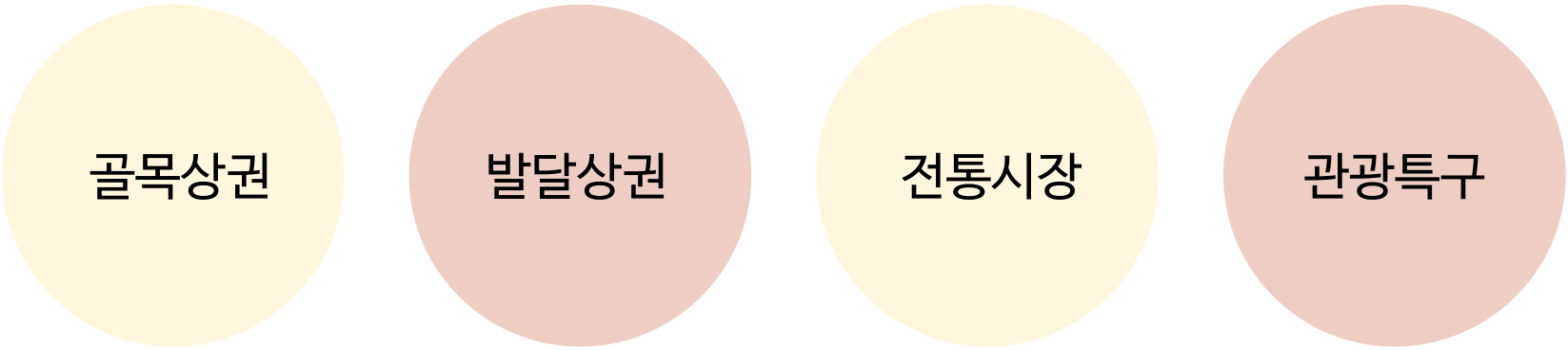
Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
0	1	-3678035766611086.0	-0.0	-4436542900335454.5	-2919528632886717.5	TRUE
0	2	-2917180537855892.0	-0.0	-3373295454359584.0	-2461065621352200.0	TRUE
0	3	190369891416402.0	0.6385	-223558632207758.3	604298415040562.2	FALSE
1	2	760855228755194.0	0.0274	59029919962619.125	1462680537547769.0	TRUE
1	3	3868405658027488.0	-0.0	3193235479721085.0	4543575836333891.0	TRUE
2	3	3107550429272294.0	-0.0	2809955881572386.0	3405144976972202.0	TRUE

Multiple Comparisons Between All Pairs (Tukey)



상권별로 매출의 차이가 있는 것으로 확인. 상권별로 서비스업종의 비중이 큰 차이가 있는 것으로 볼 수 있음
이러한 특징은 상권별 매출에 영향을 주는 독립변수도 차이가 있을 것이라 여겨진다.



골목상권

발달상권

전통시장

관광특구

```
model=smf.ols(formula, data).fit()
```

OLS Regression Results
회귀분석모형의 적절성을 위한 조건 확인해보자

- 사분위수를 이용하여 이상치 제거
- 각 상권별로 상관계수 확인 후 독립변수 설정
- 적절성을 확인하면서 변수를 바꿔 주도록 한다.

다중공선성	골목상권	발달상권	전통시장	관광특구
	3.453	-0.082	5.735	-0.271
	5.629	0.042	2.685	-2.910
	0.831	0.179	0.145	-0.003
	4.157	-0.118	5.531	1.478
	4.550	1.125	5.811	-0.273

	골목상권	발달상권	전통시장	관광특구
Adj. R-squared	0.955	0.938	0.968	0.969
Prob (F-statistic)	0.00	0.00	0.00	0.00
Durbin-Watson (독립성)	1.816	1.789	1.693	1.887
P> t	0.000	0.000	0.000	0.003
Skew (왜도)	0.298	0.756	0.072	1.577
Kurtosis (첨도)	24.371	18.690	19.290	15.342
Shapiro (정규성)	0.0	0.0	6.353347107402288e-41	2.0196184957740435e-20
선형성				
등분산성				

딥러닝으로 모델을 만들어 보았으나 역시나 제대로 나오지 않는다.
 결과가 나오는 값들마저 예측값이 실제값과 너무 차이 나는것을 확인 (실제 데이터의 값을 넣어서 예측한 결과값과 비교)

```
model=tf.keras.Sequential([
    tf.keras.layers.Dense(units=64, input_shape=(5,), activation='relu'),
    tf.keras.layers.Dense(units=32, activation='relu'),
    tf.keras.layers.Dense(units=1)
])

model.compile(optimizer='adam', loss='mse')
print(model.summary())

model.fit(test_x, test_y, epochs=100)
print(model.predict(test_x).flatten())
```



loss: 678439550976.0000	loss: 1177366396871376896.0000	loss: nan
loss: 918490578944.0000	loss: 1177365572237656064.0000	loss: nan
loss: 952932499456.0000	loss: 1177364472726028288.0000	loss: nan
loss: 835954868224.0000	loss: 1177363372724400512.0000	loss: nan
loss: 801391116288.0000	loss: 1177362272722222272.0000	loss: nan
loss: 748062441472.0000	loss: 1177361172720000000.0000	loss: nan
loss: 810733076480.0000	loss: 1177360072717777776.0000	loss: nan
loss: 842712154112.0000	loss: 1177358972715555552.0000	loss: nan
loss: 989801676800.0000	loss: 1177357872713333328.0000	loss: nan
loss: 912419586048.0000	loss: 1177356772711111104.0000	loss: nan
loss: 1294505607168.0000	loss: 1177355672708888880.0000	loss: nan

하지만, 현재 데이터로써는 한계가 있다고 판단

정규성을 만족하지 않는다... 하지만 프로젝트를 위해 진행한다...

골목상권	
정규성	✗
독립성	✓
선형성	✓
등분산성	✓
다중공선성	✓

발달상권	
정규성	✗
독립성	✓
선형성	✓
등분산성	✓
다중공선성	✓

관광특구	
정규성	✗
독립성	✓
선형성	✓
등분산성	✓
다중공선성	✓

전통시장	
정규성	✗
독립성	✓
선형성	✓
등분산성	✓
다중공선성	✓

골목상권 **관광특구** 상권 전통시장

상권, 업종 선택하기

상권 이름: 이태원 관광특구

대분류 업종:
외식업

양식음식점

분석하기

사용자가 입력한 값들은
파라미터로 넘어와
새로운 입력값(데이터)을 가져올 때 사용된다.



관광특구 상권의 이태원관광특구의
양식음식점의
매출 평균 금액을 가져온다.



가져온 **평균의 값**을
학습된 모델에 **새로운 값으로** 입력해준다.

```
model.predict(new_x)
```

[웹 화면 출력 내용]

- 연간 예상 매출액 **319,768,630 원**
 - 분기별 예상 매출액
 - 선택한 서비스업종의에 대한 그래프 출력
- * 분석 report는 용산구 상권의 추이와
현황 정보 등을 제공하기 위함임. 예측 X

분석코드

용산구 상권 매출 예측 서비스

데이터 추가확보

2019년 전 데이터를 추가적으로 확보하고, 대중교통 승하차인원과 유동인구 등의 매출금액과 다른 성격의 다양한 변수를 통해 상관관계를 파악한다면 더욱 정교한 모델을 구축할 수 있을 것으로 기대한다.

Model 저장

Nodel 학습 후 static에 저장하여 모델을 불러올 예정이었으나 `NameError: name 'settings' is not defined` 가 계속해서 발생하였다. Django app 초기 생성 시 문제가 있는 것으로 보인다. 추후 다시 보완하여 진행해보도록 한다.

AWS 클라우드 서버 출력

ssh 명령어를 통해 클라우드 서버에 있는 아파치를 통해서 우분투 리눅스 환경에 접속하였고, 제작 중이었던 장고 프로젝트를 실행시켜 보았다. 하지만 가용한 리소스가 충분하지 않아서 페이지 로딩이 너무 느려지는 이슈를 만나게 되었다. 따라서 현재는 로컬에서 장고 서버를 통해 웹 사이트를 출력하였으며, 이후 보완 후 다시 클라우드 서버를 통한 출력을 시도해볼 계획이다.

마이페이지 스크랩 저장 기능

출력된 분석 결과 보고서를 이미지로 DB에 저장하고 마이페이지에서 게시판 형태로 조회 및 관리가 가능하도록 구성한다. 그러면 로그인 기능과 조합하여 활용할 수 있을 것으로 기대한다.

Q & A

감사합니다.