



시계열 분석 및 모델링을 통한 업종 추천

: 암구정동 및 신촌동을 중심으로

INDEX



01
Outline

Outline

1-1 팀 소개, 역할 분담



TEAM | [REDACTED]



데이터 전처리
통계적 모델링
GPT 모델링



데이터 시각화
자료수집
시계열 모델링



데이터 전처리
딥러닝 모델링
매출 예측 및 시각화



데이터 시각화
매출 관련 외부 요인 조사
논문 분석



데이터 전처리
데이터 시각화
시계열 모델링



데이터 시각화
딥러닝 모델링
발표

Outline

1-2 기획 배경 및 목표



🔑 예비창업자의 핵심역량 : 시간에 따라 변화하는 상권과 소비자의 니즈 파악

[메트로신문] 창업은 정보 전쟁이다. 창업아이템을 분석하는 것은 물론, **상권조사와 분석, 유사한 아이템을 운영하고 있는 자영업자들의 실질적 운영노하우 분석 등 다양한 정보와 점검은 필수다.** 초보 창업자들에게 적은 노력으로 다양한 창업정보를

© 메트로서울, [이상현 칼럼] 프랜차이즈 창업 필수 분석 10선

권 분석의 중요성은 잘 알고 있다. 상권 분석을 한마디로 정리하자면 '수익성 분석'이다. 우리나라처럼 자영업자 간 경쟁이 치열한 상권에서는 개별 점포의 투자 수익성을 따져보는 일이 매우 중요해졌다. **주변 상권에 어떤 업종, 어떤 콘셉트의 가게가 새로 문을 열고, 문을 닫는지, 어떤 인구집객시설이 생기느냐에 따라 내 가게의 매출 성과가 달라질 수도 있기 때문이다.** 그렇다면 상권 분석은 어떤 방법으로 진행해야 하는 것일까? 해당 상권에 처음 진입하는 소상공인

© 시사저널, '포스트 코로나'로 바뀐 상권의 속살을 살펴라

✓ 상권 변화를 이끄는 외생적 요인 : ex. 사회 기조, 유행, 임대료, 교통 시설, 생활인구 등 …

© 세계로컬타임즈

K-POP 한류 열풍 경제 영향력 상승가도?

역직구몰 매출 '수직상승'…성형외과 외국인 환자도 증가

© 서울경제

KB국민카드 "벚꽃 피니 주변 상권 매출 급증"

© 서울경제

9시 영업 제한에 7주 연속 매출 '반토막'

Outline

1-2 기획 배경 및 목표



“정성적, 정량적인 다양한 외생적인 요인에도 불구하고, 굳건히 성장하는 상권과 쇠락세에서 벗어나지 못하는 지역은 있다!”

성장과 쇠퇴를 경험하고 있는 그 지역 상권의 특징은 무엇일까?

그 지역에 창업을 준비하고 있는 이들에게 추천할 수 있는 업종은 무엇일까?

02
Design

Design

2-1 개발 과정



프로젝트 전체 로드맵



I
데이터 수집 및 전처리



II
데이터 분석 후 가설 세우기



III
다양한 모델 학습 및 시각화



IV
시각화 그래프 통해 업종 추천

- 서울 열린데이터광장에 서 데이터 수집
- 전처리 결과 : **912348행 X 22열**

- 분석 목적에 맞는 지역 선택
→ 압구정동, 신촌동
- 연분기 별 데이터로 정렬하여 '총 매출 금액'을 기준으로 경우의 수를 만들어 분석

- 통계적 모델링 이용해 지표 선정 및 업종 선정
- 시계열(**SARIMA**), 딥러닝(**LSTM**), **Chat-GPT** 활용하여 모델링 및 학습
- 2023년 매출 예측 및 시각화**

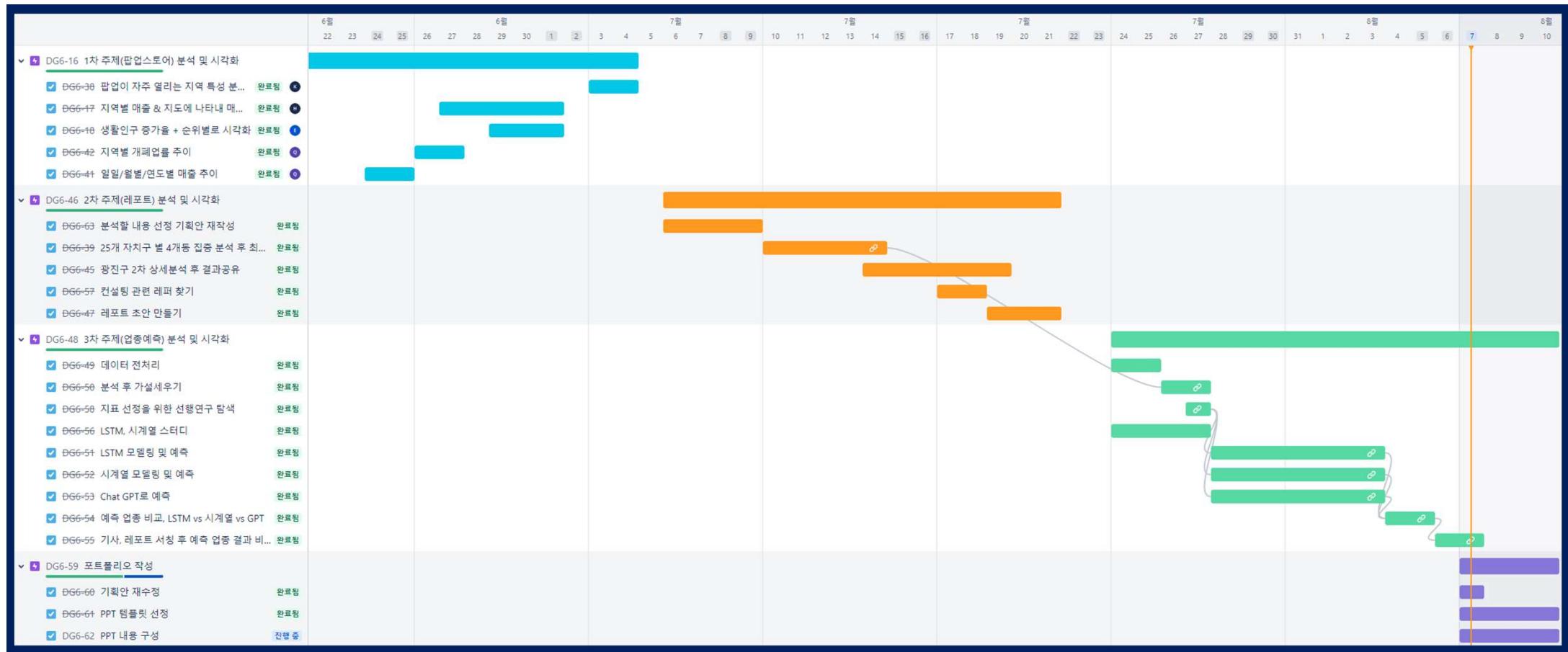
- 가설 검증
- 시각화 그래프로 추천/비추천 업종 선정
- 실제 기사/레포트와 예측한 결과 비교

Design

2-1 개발 과정



프로젝트 전체 로드맵



Design

2-2 사용한 협업 Tool



Notion

Notion Dashboard:

- Top bar: Ambulance icon, redacted area.
- Left sidebar:
 - 목차 (Table of Contents)
 - 회의록 (Meeting Log)
 - 강사님/멘토링 피드백 (Mentoring Feedback)
 - 일정 (Calendar)
 - 팀원 공유 플더 (Team Member Shared Folders)
 - 수행일지 풀더 (Performance Log Folder)
 - 제출 폴더 (Submission Folder)
- Main area:
 - 강사님/멘토링 피드백**: A table with columns 날짜 (Date), 내용 (Content), and 분류 (Category). Rows include:
 - 2023년 8월 5일: 3차 멘토링 (마지막) - Mentoring
 - 2023년 7월 26일: 삼성전자 주가 알려주신 내용 - 강사님
 - 2023년 7월 24일: 시계열 분석에 관하여 질의응답 - 강사님
 - 2023년 7월 22일: 2차 멘토링 - Mentoring
 - 2023년 7월 21일: 금요일의 8교시 (ft. 멘토링을 앞두고) - 강사님
 - 2023년 7월 14일: 서울시 전체 분석한 결과 공유하고 향후 방향성 피드백 - 강사님
 - 2023년 7월 6일: 주제 변경하고 대중적은 기획안으로 받은 피드백 - 강사님
 - 2023년 8월 11일: 1차 멘토링 질의사항 정리 - Mentoring
 - 2023년 7월 1일: 지금까지 무엇을 하였는가. - In progre..
 - 2023년 7월 24일: 시계열/딥러닝 레퍼런스 - Done
 - 2023년 7월 14일 - 2023년 7월 19일: 광진구 결과 공유 - Done
 - 2023년 7월 13일: 웹서비스든 모델링이든 금융컨설팅과 연관된 레퍼 찾기 - Done
 - 2023년 7월 10일 - 2023년 7월 14일: 행정동 13동 결과 공유 - Done
 - 2023년 7월 4일 - 2023년 7월 28일: 우리의 데이터는 어디로 가는가 ep02. - Done
 - 2023년 6월 16일: 본 프로젝트 주제 토의 - Done

Slack

Slack Channel: 데이터분석_시각화

Recent messages:

- 6일 전: 3 고정됨, 데이터구조, 수행일지풀더, 7월 4주차 기획안.docx ...
7월 5일 수요일 ~ 7월 6일 목요일
오전 9:41
[Redacted] 지는 아이디어
이에 상권분석 보고서를 제공하는 것도 궁금마케팅이라고 볼 수 있을지 궁금하네요.
bc카드에서 대박상경이라는 서비스를 제공하고 있는데 주요 기능 중에 성장성과 안정성으로 상권을 순서대로 알려주는 것도 있다고 해요.
안정성은 현재 우리의 데이터셋으로 생활인구도 높고 매출도 높은 상권이라고 할 수 있을 것 같고,
성장성은 그 상권분석지표 LL이거로 확인할 수 있을 것 같아서 던져놓습니다.
추가적으로 개인사업자 정보가 개방되있대요. json 형식이긴데 성장성을 평가하거나
나 안정성을 평가할 때 활용할 수 있지 않을까요 싶습니다. 모델링으로의 연결은 고민
이 필요합니다!
- 5일 전: KMSL: 516469393.3726969
image.png
- 16일 전: 늦은 시간인데 고생하셨어요!! 후다닥 읽어보겠습니다
- 전송됨 16일 전
[Redacted] 전송됨
건설이라는 부분이 확실히 애매한데 차라리
'본 프로젝트에서는 이러한 상권 정보를 예비 창업자, 혹은
사업을 확장하고자 하는 사업가들에게 제공하여 개
선된 환경에서 창업, 운영을 할 수 있도록 도움을 줄 수
있는 레포트를 제공하고자 한다.'
이런식은 어떨까요
- 5일 전: ETS도 경향이 비슷하긴하네요
[Redacted] LSTM이랑 더 가까워보이기도하고..
- 16일 전: 건설이라는 단어를 적으면 상권분석에 더해진 구체
성이 있어야겠네요ㅠㅠ 이번 발표는 인규님 말씀대
로 컨설팅 단어를 빼는게 좋을 것 같아요!
- 16일 전: 확실이 컨설팅은 저희가했던 방향하고는 맞지 않는 것
같네요 저희가 진행한 방향은 그냥 상권을 분석했다 정도만 될 거 같고 아래 사진에서 말하는 입지?에 좀 더
가까울 거 같아요

Design

2-3 데이터 선정 및 전처리



- 데이터 출처 : 서울 열린데이터 광장
- 활용 데이터 : 2017 ~ 2022년의 ① 상권-추정매출, ② 상권-점포 ③ 상권영역 정보, ④ 업종 정보

* 업종정보는 제공된 pdf를 바탕으로 별도 생성

Step 1.] 매출, 점포 6년치 데이터 준비

```
1 ### 매출 데이터 6개년 합치기 --> 열은 모두 똑같길래 concat 사용
2 dfs = [sales_2022, sales_2021, sales_2020, sales_2019, sales_2018, sales_2017]
3 sales_6years = pd.concat(dfs, ignore_index=True)

1 ### 점포 데이터 6개년 합치기
2 dfs = [store_2022, store_2021, store_2020, store_2019, store_2018, store_2017]
3 store_6years = pd.concat(dfs, ignore_index=True)
```

Step 2.] 매출, 점포 데이터 합치고 + 컬럼 생성 및 삭제

```
1 # 두 데이터프레임 (sales, store)의 지정 컬럼의 값이 같을 때 점포데이터의 남은 컬럼 값 가져오기
2 df_merged = pd.merge(sales_6years, store_6years, on=['기준_년_코드', '기준_분기_코드', '상권_구'])

1 # 총 매출 금액 컬럼 생성
2 df_merged['총_매출_금액'] = df_merged['주중_매출_금액'] + df_merged['주말_매출_금액']
3 df_merged.columns
```

Step 3.] 자치구, 행정동 및 업종 정보 매칭하여 컬럼 생성

```
1 ### '서비스_업종_코드_명'(업종명)에 따라 대분류, 중분류 채우기
2
3 # '대분류', '중분류' 컬럼 생성
4 df_preprocessed['대분류'] = ''
5 df_preprocessed['중분류'] = ''
6
7 # 중분류, 대분류 채우기
8 for index, row in df_preprocessed.iterrows():
9     mask = category['업종명'] == row['서비스_업종_코드_명']
10    matching_data = category[mask]
11    if not matching_data.empty:
12        df_preprocessed.at[index, '대분류'] = matching_data['대분류'].values[0]
13        df_preprocessed.at[index, '중분류'] = matching_data['중분류'].values[0]
```

```
7 # 자치구, 행정동 채우기
8 for index, row in sales_6years.iterrows():
9     mask = seoul_gudong['상권_코드'] == row['상권_코드']
10    matching_data = seoul_gudong[mask]
11    if not matching_data.empty:
12        sales_6years.at[index, '자치구'] = matching_data['자치구'].values[0]
13        sales_6years.at[index, '행정동'] = matching_data['행정동'].values[0]
```

2-4 데이터 분석 후 지역 선택



25개 자치구에서 매출 금액, 점포 밀도, 생활인구 기반으로 구별 3개의 행정동 추출해 분석한 결과 + 기사 참고하여 지역 선정

신촌동

제작: 김민수

업종수_대분류_상점수+매출 대조

업종수_증분류_상점수

업종수_업종명_상점수

(%)

(2) 면접별 나간다 매출

/4) 업종별 해당 지역의 동종업종 점포
밀집부지 대비 증가

(6) 여종별 매출 견수

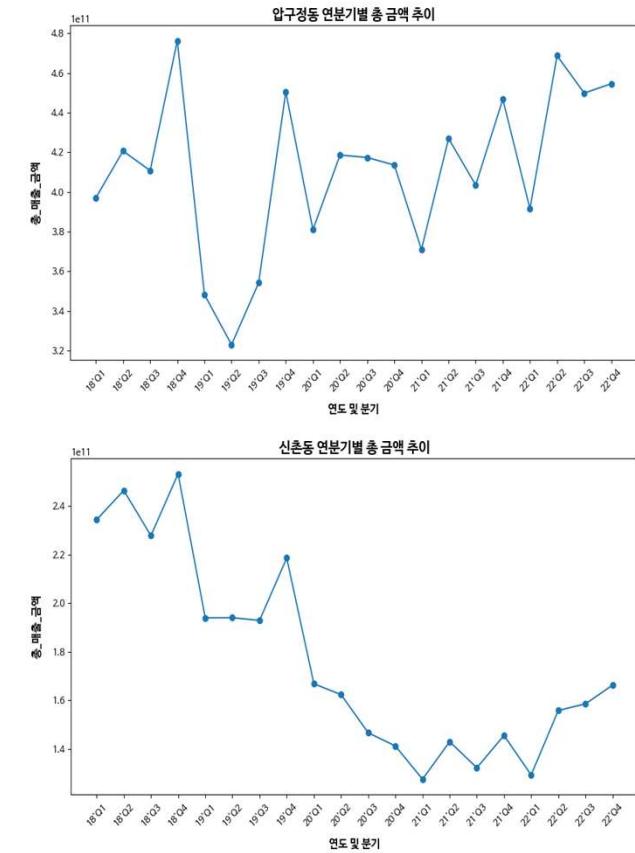
99

104

/8) 언뜻 볼 때 달 지역의 개언률 / 폐언률

/9/ 언종별 여령별 매출 거수

/10) 해당 행정동의 잘 나가는 업종별 상
과관계 분석하기



Design

2-4 선택한 지역 깊이 분석



데이터 로드

압구정동

/1/ 업종별, 요일별 매출 금액 및 비율

업종수_대분류_상점수+매출 대조

업종수_중분류_상점수

업종수_업종명_상점수

요일별 매출금액

/2/ 업종별 성별 매출 비율(ex. 남 xx% 여 yy%)

성별_매출금액

성별_매출건수

/3/ 업종별 시간당 매출

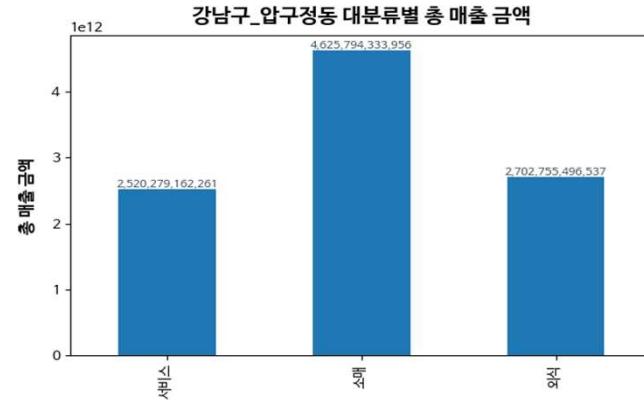
연분기별 매출 주이

/8/ 업종별 해당 지역의 개폐업률, 개폐업수

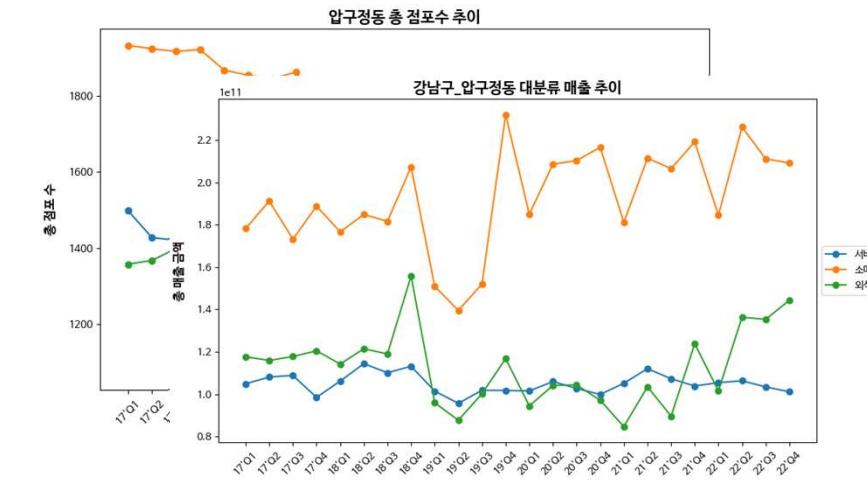
순점포증감률

총점포수 주이

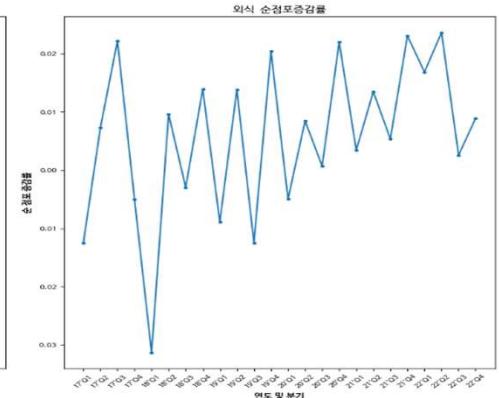
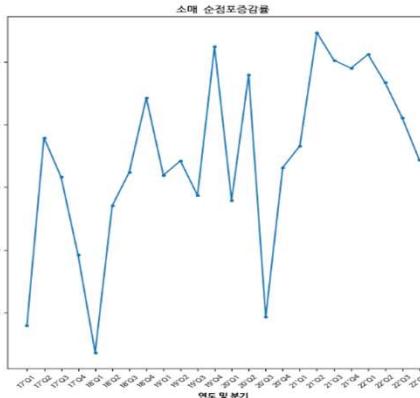
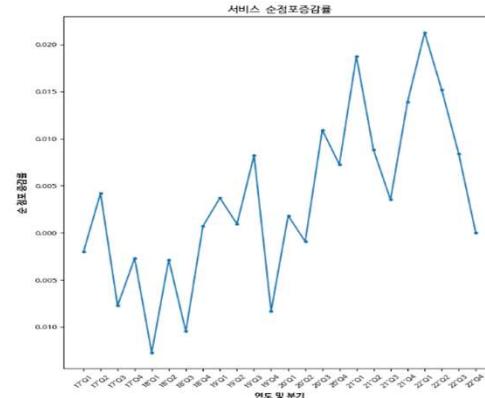
업종별 매출



업종별 점포 수 및 매출 금액 변화



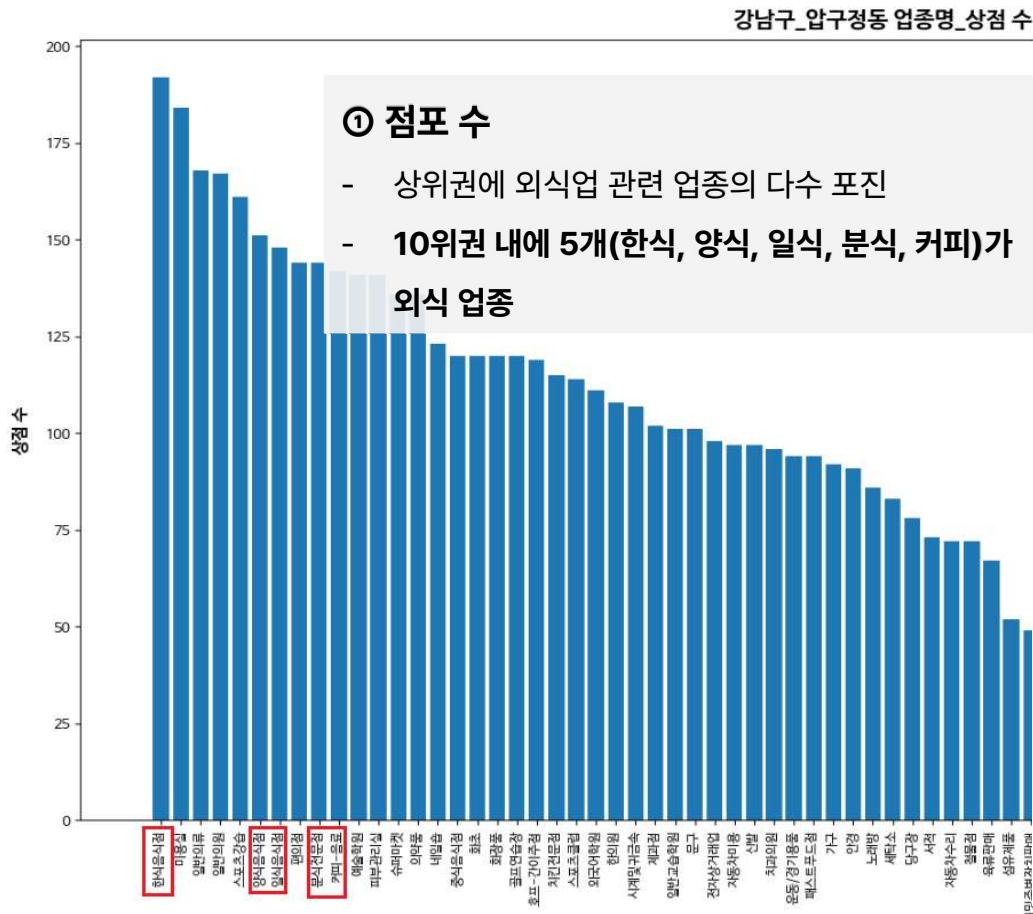
업종별 순점포증감률



2-4 깊이 분석한 후 세운 가설 - 압구정동



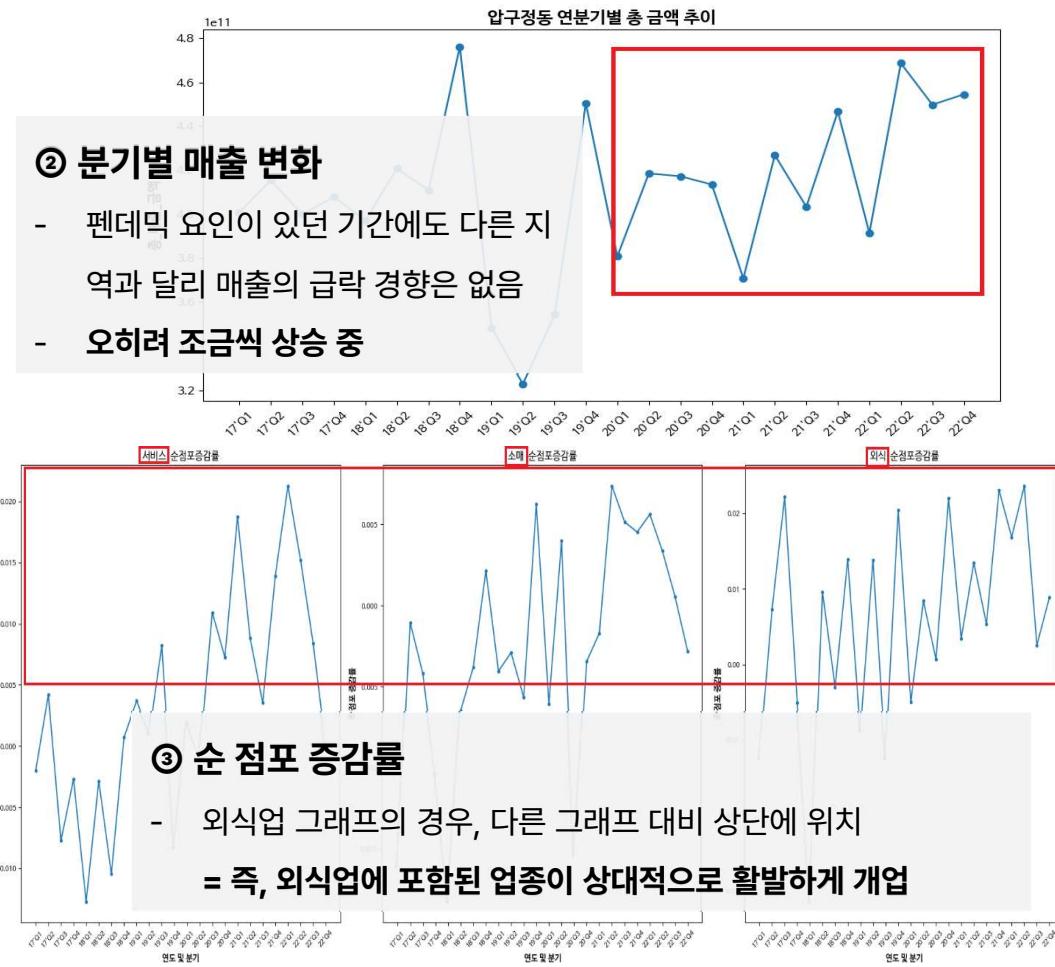
압구정동



① 점포 수

- 상위권에 외식업 관련 업종의 다수 포진
- 10위권 내에 5개(한식, 양식, 일식, 분식, 커피)가

외식 업종

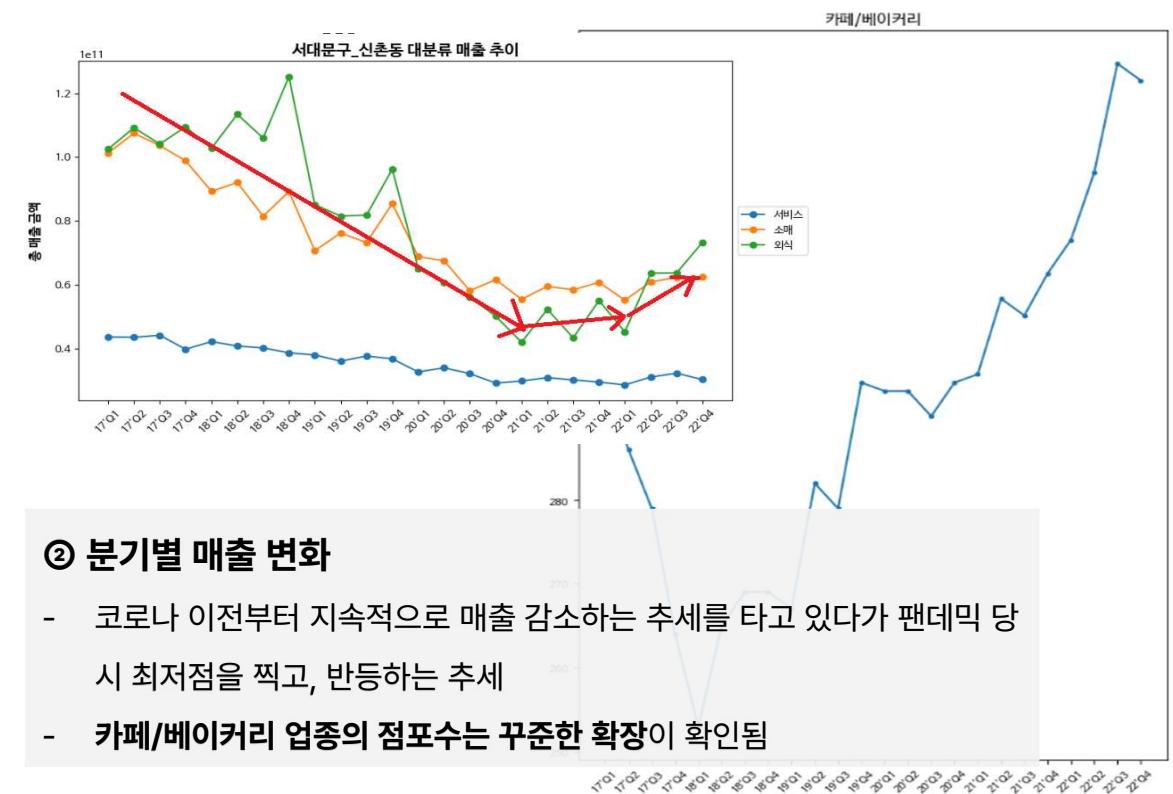
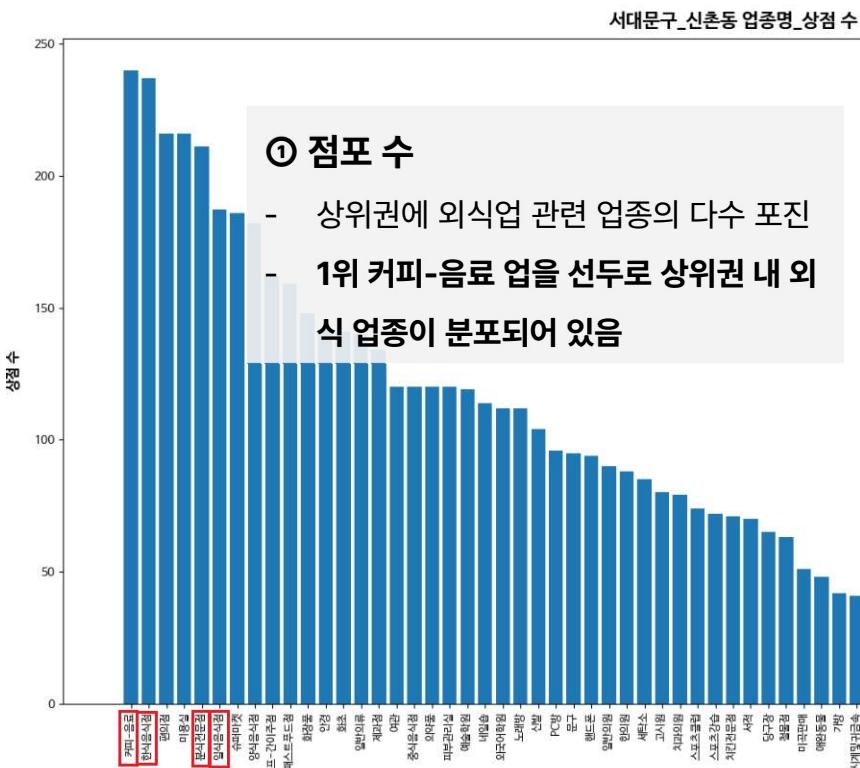


Design

2-4 깊이 분석한 후 세운 가설 - 신촌동



신촌동





압구정동

“ 압구정동의 상권은 코로나의 영향을 받지 않고,
꾸준히 성장 중이며 뜨는 업종은 주로 외식업에 있을 것이다.”

신촌동

“ 신촌동 상권은 꾸준히 하락세였으나 앞으로는 올라갈 것이며,
'카페/베이커리' 업종의 상승세가 뚜렷할 것이다.”

03
Predict

Predict

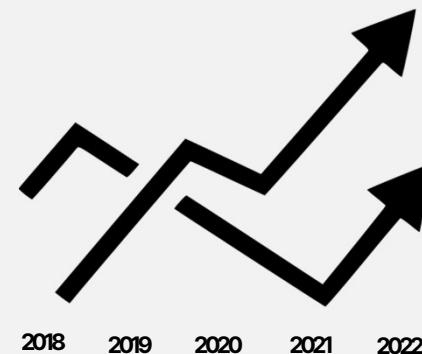
3-1 Predict



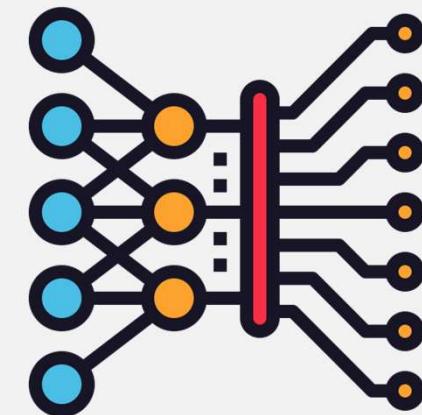
통계적 모델링



시계열 모델링



딥러닝 모델링



Predict

3-1 Predict





**Python으로 통계 모형을 분석 하기 위해,
Statsmodels 라이브러리의 OLS를 사용하여 선형 회귀 분석 진행**

⌚ 종속변수

: 총 매출 금액

⌚ 독립변수

: 총 점포수, 점포 수, 개업 점포 수, 폐업 점포 수, 프랜차이즈 점포 수, 주중, 주말 매출 금액, 남여 매출 금액, 순점포증감률, 시장성, 성장성, 업종

[종속변수와 독립변수간의 데이터들의 통계적 관계성 확인]

Predict

3-1 통계적 모델링 – 지표 선정



종속변수와 독립변수 간의
상관관계 확인

OLS Regression Results						
Dep. Variable:	총_매출_금액	R-squared:	1.000			
Model:	OLS	Adj. R-squared:	1.000			
Method:	Least Squares	F-statistic:	1.522e+26			
Date:	Fri, 04 Aug 2023	Prob (F-statistic):	9.52e-168			
Time:	04:32:35	Loo-Likelihood:	102.31			
No. Observations:	24	AIC:	-182.6			
Df Residuals:	13	BIC:	-169.7			
Df Model:	10					
Covariance Type:	nonrobust					
coef	std err	t	P> t	[0.025	0.975]	
const	0.0002	0.053	0.005	0.996	-0.114	0.114
총_점포수	9.537e-07	5.63e-05	0.017	0.987	-0.000	0.000
점포_수	-1.967e-06	6.11e-05	-0.032	0.975	-0.000	0.000
개업_점포_수	1.311e-06	5.8e-05	0.023	0.982	-0.000	0.000
폐업_점포_수	-4.768e-07	5.44e-05	-0.009	0.993	-0.000	0.000
프랜차이즈_점포_수	2.98e-06	0.000	0.026	0.980	-0.000	0.000
주중_매출_금액	1.0000	2.3e-13	4.35e+12	0.000	1.000	1.000
주말_매출_금액	1.0000	3.58e-13	2.79e+12	0.000	1.000	1.000
남성_매출_금액	-2.132e-14	5.02e-13	-0.042	0.967	-1.11e-12	1.06e-12
여성_매출_금액	-8.882e-15	2.13e-13	-0.042	0.967	-4.69e-13	4.51e-13
순점포증감률	-2.861e-06	0.000	-0.010	0.993	-0.001	0.001
시장성	-5.457e-12	6.33e-10	-0.009	0.993	-1.37e-09	1.36e-09
성장성	0	0.014	0	1.000	-0.030	0.030
Omnibus:	1.564	Durbin-Watson:	0.001			
Prob(Omnibus):	0.457	Jarque-Bera (JB):	1.386			
Skew:	-0.474	Prob(JB):	0.500			
Kurtosis:	2.301	Cond. No.	1.11e+19			

Notes:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The smallest eigenvalue is 3.21e-14. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.

R², 결정계수

: 회귀 모델의 설명력을 나타내는 지표로,
'1'에 가까울 수록 높은 성능의 회귀 모델임을 의미

Prob (F-statistic)

: 회귀식이 유의미 한지 판단할 수 있는 지표
'0.05' 이하일 경우 변수끼리 관계가 있다고 판단

AIC

: 표본의 개수와 모델 복잡성에 기반하여 모델 평가
수치가 낮을 수록 좋은 모델

BIC

: AIC보다 모델 평가 성능이 높으며,
AIC와 마찬가지로 수치가 낮을 수록 좋은 모델

1에 가까운 R²(결정계수) 값 + Prob 값이 0.05 이하 + 낮은 AIC&BIC 값 = 유의미한 통계로 판단



총 매출 금액과 통계적 유의성을 갖는 지표를 찾자 ↗ **순 점포 증감률, 시장성, 성장성 ↘**

- **순 점포 증감률** = (상점 개업 수 – 상점 폐업 수) / 전체 상점의 수
: 해당 업종의 규모 및 활성화 정도 파악 가능
- **시장성** = (행정동 내 분기 매출액 / 분기 점포 수) / (자치구 내 분기 매출액 / 분기 점포 수)
: 자치구 평균 대비 행정동의 시장성 판단 가능
- **성장성** = 이번 분기 매출액 / 지난 분기 매출액
: 지난 분기 대비 해당 분기의 매출 증감률을 비교하여 성장성 파악 가능



3-1 통계적 모델링 – 업종 선정 (압구정동)

Omnibus의 큰 값(=정규분포 따름) + 독립성이 있음 = 회귀계수들이 통계적으로 유의미

☞ coef 값이 양수인 것과 0.05 이하인 P-Value 값을 가진 업종들을 추려 시계열 분석에 사용

OLS Regression Results											
Dep. Variable:	총_매출_금액	R-squared:	0.317								
Model:	OLS	Adj. R-squared:	0.303 <th data-cs="4" data-kind="parent"></th> <th data-kind="ghost"></th> <th data-kind="ghost"></th> <th data-kind="ghost"></th>								
Method:	Least Squares	F-statistic:	23.								
Date:	Tue, 01 Aug 2023	Prob (F-statistic):	1.68e-2								
Time:	00:41:17	Log-Likelihood:	-787								
No. Observations:	3693	AIC:	1.577e+00								
Df Residuals:	3620	BIC:	1.532e+00								
Df Model:	72										
Covariance Type:	nonrobust										
	coef	std err	t	P> t	[0.025	0.975]					
Intercept	1.595e+08	7.51e+07	2.123	0.034	1.22e+07	3.07e+08					
C(임종명)[T.기구]	6.582e+08	9.42e+07	6.988	0.000	4.74e+08	8.43e+08					
C(임종명)[T.기방1]	4.192e+08	1.16e+08	3.613	0.000	1.92e+08	6.47e+08					
C(임종명)		+08	5.589	0.000	3.78e+08	7.87e+08					
C(임종명)		+08	-0.272	0.785	-1.02e+09	7.72e+08					
C(임종명)		+07	1.096	0.273	-7.58e+07	2.68e+08					
C(임종명)		+07	-0.477	0.633	-2.13e+08	1.3e+08					
C(임종명)		+07	-0.464	0.642	-2.25e+08	1.39e+08					
C(임종명)		+07	-1.462	0.144	-3.21e+08	4.68e+07					
C(임종명)[T.독서실]	-1.153e+08	1.5e+08	-0.768	0.443	-4.1e+08	1.79e+08					
C(임종명)[T.문구]	5.732e+07	8.92e+07	0.643	0.520	-1.18e+08	2.32e+08					
C(임종명)[T.미술판매]	-3.936e+07	9.98e+07	-0.394	0.693	-2.35e+08	1.56e+08					
C(임종명)[T.미술용품]	4.503e+08	8.85e+07	5.089	0.000	2.77e+08	6.24e+08					
C(임종명)[T.반찬가게]	8.673e+07	1.05e+08	0.827	0.408	-1.19e+08	2.92e+08					
C(임종명)[T.부동산중개업]	-6.674e+07	1.99e+08	-0.336	0.737	-4.56e+08	3.23e+01					
C(임종명)[T.분식점문점]	1.946e+08	1.01e+08	1.920	0.055	-4.07e+06	3.93e+08					
C(임종명)[T.사진관]	1.009e+09	1.61e+08	6.262	0.000	6.93e+08	1.33e+09					
C(임종명)[T.서적]	1.894e+08	9.37e+07	2.022	0.043	5.77e+06	3.73e+08					
C(임종명)[T.섬유제품]	8.015e+07	1.01e+08	0.791	0.429	-1.18e+08	2.79e+08					
C(임종명)[T.세무사사무소]	1.323e+08	4.57e+08	0.290	0.772	-7.64e+08	1.03e+09					
C(임종명)[T.세탁소]	-1.442e+07	9.08e+07	-0.159	0.874	-1.93e+08	1.64e+08					
C(임종명)[T.수선물판매]	-2.901e+07	1.08e+08	-0.269	0.788	-2.4e+08	1.82e+08					
C(임종명)[T.슈퍼마켓]	4.184e+08	8.95e+07	4.676	0.000	2.43e+08	5.94e+08					
C(임종명)[T.스포츠강습]	2.093e+07	8.62e+07	0.243	0.808	-1.48e+08	1.9e+08					
C(임종명)[T.스포츠클럽]	3.299e+08	8.92e+07	3.699	0.000	1.55e+08	5.05e+08					
C(임종명)[T.시계및귀금속]	1.068e+09	9.7e+07	11.009	0.000	8.78e+08	1.26e+09					
C(임종명)[T.신발]	3.517e+08	9.29e+07	3.785	0.000	1.7e+08	5.34e+08					
C(임종명)[T.의자]	-1.1e+08	1.76e+08	-0.624	0.533	-4.55e+08	2.36e+08					
C(임종명)[T.안경]	5.033e+08	9.08e+07	5.541	0.000	3.25e+08	6.81e+08					
C(임종명)[T.아동운동]	-1.102e+08	1.09e+08	-1.015	0.310	-3.23e+08	1.03e+08					
C(임종명)[T.양식식품점]	6.183e+08	1.01e+08	6.103	0.000	4.2e+08	8.17e+08					
C(임종명)[T.여관]	1.152e+08	1.04e+08	1.112	0.266	-8.79e+07	3.18e+08					
C(임종명)[T.여행사]	1.474e+08	1.24e+08	1.191	0.234	-9.53e+07	3.9e+08					
C(임종명)[T.예술품]	-8.601e+07	1.61e+08	-0.534	0.594	-4.02e+08	2.3e+08					
C(임종명)[T.예술학원]	1.814e+08	8.66e+07	2.095	0.036	1.16e+07	3.51e+08					

유의확률
경우 유의

디아고스티노검정

: 정규성 판단 지표로 값이 클수록 정규분포로 판단

C(업종별)[T.인테리어]	3.472e+08	1.42e+08	2.445	0.015	6.88e+07	6.26e+08
C(업종별)[T.일반교습학원]	4.554e+08	9.39e+07	4.848	0.000	2.71e+08	6.39e+08
C(업종별)[T.일반외익점]	4.178e+08	1.e+08	3.815	0.000	2.03e+08	6.33e+08
일반외익점]	3.926e+08	9.6e+07	4.091	0.000	2.04e+08	5.81e+08
일식음식점]	4.867e+08	9.57e+07	5.088	0.000	2.99e+08	6.74e+08
자동차마음]	9.163e+07	8.86e+07	1.034	0.301	-8.21e+07	2.65e+08
자동차수리]	1.632e+08	9.6e+07	1.701	0.089	-2.49e+07	3.51e+08
자전거 및 기타운송장비]	2.213e+08	1.01e+08	2.185	0.029	2.27e+07	4.2e+08
전자상거래업]	5.304e+08	1e+08	5.287	0.000	3.34e+08	7.27e+08
C(업종별)[T.제과점]	2.367e+08	9.6e+07	2.467	0.014	4.86e+07	4.25e+08
C(업종별)[T.조명용품]	6.008e+08	1.08e+08	5.573	0.000	3.89e+08	8.12e+08
C(업종별)[T.주유소]	1.948e+08	1.1e+08	1.764	0.078	-2.17e+07	4.11e+08
C(업종별)[T.중식음식점]	3.737e+08	9.08e+07	4.115	0.000	1.96e+08	5.52e+08
C(업종별)[T.철물점]	1.297e+08	9.32e+07	1.392	0.164	-5.3e+07	3.12e+08
C(업종별)[T.청과상]	-9.445e+06	1.26e+08	-0.075	0.940	-2.56e+08	2.37e+08
C(업종별)[T.치과의원]	9.049e+08	9.22e+07	9.811	0.000	7.24e+08	1.09e+09
C(업종별)[T.침기전문점]	1.949e+06	8.92e+07	0.022	0.983	-1.73e+08	1.77e+08
C(업종별)[T.커피~음료]	2.357e+08	1.08e+08	2.186	0.029	2.43e+07	4.47e+08
C(업종별)[T.컴퓨터및주변장치판매]	2.106e+08	1.01e+08	2.079	0.038	1.2e+07	4.09e+08
C(업종별)[T.통신기기수리]	-1.293e+08	1.46e+08	-0.887	0.375	-4.15e+08	1.57e+08
C(업종별)[T.패스트푸드점]	1.936e+08	9.63e+07	2.011	0.044	4.83e+06	3.82e+08
C(업종별)[T.편의점]	7.653e+08	9.32e+07	8.215	0.000	5.83e+08	9.48e+08
C(업종별)[T.피부과의사]	1.923e+08	9.5e+07	1.919	0.	0.	0.
디아고스티노검정						
: 정규성 판단 지표로 값이 클수록 정규분포로 판단						
독립성 판단 지표						
: 1.5 ~ 2.5 사이일 경우 독립적으로 판단						
C(업종별)[T.화장품]	2.034e+09	1.26e+08	16.179	0.0	0.	0.
C(업종별)[T.화초]	5.707e+07	9.01e+07	0.633	0.527	-1.2e+08	2.34e+08
C(업종별)[T.회계사사무소]	-1.03e+08	2.71e+08	-0.180	0.704	-6.34e+08	4.28e+08
Omnibus:	1076.131	Jarbin-Watson:	1.878			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	3457.837			
Skew:	1.470	Prob(JB):	0.00			
Kurtosis:	6.718	Cond. No.	89.1			

.4
0
독립성 판단 지표

: 1.5 ~ 2.5 사이일 경우
독립적으로 판단



3-1 통계적 모델링 – 업종 선정 (신촌동)

신촌동의 회귀분석 결과 역시, Omnibus의 큰 값(=정규분포 따름)을 띠며 독립성이 있음

☞ 결정계수는 낮아 설명력은 낮으나 회귀 계수들이 통계적으로 유의미하다고 판단됨

OLS Regression Results													
Dep. Variable:	총 매출	R-squared:											
Model:	OLS	0.322											
Method:	Least Squares												
Date:	Fri, 28 Jul 2023	Prob (F-statistic):											
Time:	04:14:39	Log-Likelihood:											
No. Observations:	3593	AIC:											
Df Residuals:	3525	BIC:											
Df Model:	67												
Covariance Type:	nonrobust												
coef	std err	t	P> t	[0.025	0.975]								
Intercept	1.443e+08	2.91e+07	4.952	0.000	8.72e+07	2.01e+08							
C(업종명)[T. 가구]	-8.275e+07	1.85e+08	-0.446	0.655	-4.46e+08	2.81e+08	C(업종명)[T. 우도/경기용품]	4.524e+08	5.62e+07	8.046	0.000	3.42e+08	5.63e+08
C(업종명)[T. 가방]	-3.07e+07	5.77e+07	-0.532	0.595	-1.44e+08	8.25e+07	C(업종명)[T. 의약품]	2.533e+08	6.24e+07	4.058	0.000	1.31e+08	3.76e+08
C(업종명)[T. 가전제품]	-1.235e+08	7.1e+07	-1.739	0.082	-2.63e+08	1.57e+07	C(업종명)[T. 인테리어]	-5.734e+06	5.94e+07	-0.096	0.923	-1.22e+08	1.11e+08
C(업종명)[T. 고시원]	-4.3e+07	4.68e+07	-0.919	0.358	-1.35e+08	4.87e+07	C(업종명)[T. 일반교습학원]	3.755e+08	4.19e+07	8.964	0.000	2.93e+08	4.58e+08
C(업종명)[T. 골프연습장]	1.065e+08	6.24e+07	1.706	0.088	-1.59e+07	2.29e+08	C(업종명)[T. 일반의류]	-1.334e+08	6.76e+07	-1.972	0.049	-2.66e+08	-7.57e+05
C(업종명)[T. 네일숍]	-7.52e+07	3.95e+07	-1.903	0.057	-1.53e+08	2.28e+06	C(업종명)[T. 조명용품]	6.274e+08	8.02e+07	7.821	0.000	4.7e+08	7.85e+08
C(업종명)[T. 노래방]	-1.008e+08	4.16e+07	-2.422	0.015	-1.82e+08	-1.92e+07	C(업종명)[T. 주유소]	1.724e+08	5.49e+07	3.142	0.002	6.49e+07	2.8e+08
C(업종명)[T. 당구장]	-8.062e+07	4.91e+07	-1.643	0.101	-1.77e+08	1.56e+07	C(업종명)[T. 일반의원]	5.366e+08	4.62e+07	11.605	0.000	4.46e+08	6.27e+08
C(업종명)[T. 독서실]	-0.191e+08	6.92e+07	-1.471	0.141	-2.38e+08	3.39e+07	C(업종명)[T. 일식음식점]	6.826e+07	3.68e+07	1.856	0.064	-3.85e+06	1.4e+08
C(업종명)[T. 문구]	4.052e+08	4.06e+07	9.984	0.000	3.26e+08	4.85e+08	C(업종명)[T. 자전거 및 기타운송장비]	-4.721e+07	9.61e+07	-0.491	0.623	-2.36e+08	1.41e+08
C(업종명)[T. 미국판매]	-1.179e+07	4.68e+07	-0.252	0.801	-1.04e+08	8e+07	C(업종명)[T. 전자상거래업]	-1.438e+08	2.61e+08	-0.552	0.581	-6.55e+08	3.67e+08
C(업종명)[T. 미용실]	-5.285e+07	3.76e+07	-1.404	0.160	-1.27e+08	2.1e+07	C(업종명)[T. 제과점]	-6.502e+07	6.62e+07	-0.983	0.326	-1.95e+08	6.47e+07
C(업종명)[T. 반찬가게]	1.771e+08	6.62e+07	2.677	0.007	4.74e+07	3.07e+08	C(업종명)[T. 철물점]	1.41e+08	4.01e+07	3.514	0.000	6.23e+07	2.2e+08
C(업종명)[T. 불광점]	8.377e+07	7.29e+07	1.149	0.251	-5.92e+07	2.27e+08	C(업종명)[T. 청과상]	-1.92e+07	7.75e+07	8.659	0.000	5.19e+08	8.23e+08
C(업종명)[T. 분식전문점]	-3.714e+07	4.01e+07	-0.925	0.355	-1.16e+08	4.15e+07	C(업종명)[T. 치과의원]	1.077e+09	9.61e+07	11.206	0.000	8.88e+08	1.26e+09
C(업종명)[T. 사진관]	-4.728e+07	6.14e+07	-0.771	0.441	-1.68e+08	7.3e+07	C(업종명)[T. 치킨전문점]	6.783e+07	4.19e+07	1.619	0.105	-1.43e+07	1.5e+08
C(업종명)[T. 서적]	4.733e+07	4.39e+07	1.077	0.281	-3.88e+07	1.33e+08	C(업종명)[T. 커피-음료]	3.919e+07	3.78e+07	1.036	0.300	-3.5e+07	1.13e+08
C(업종명)[T. 섬유제품]	-1.411e+08	6.14e+07	-2.299	0.022	-2.61e+08	-2.08e+08	C(업종명)[T. 컴퓨터및주변장치판매]	3.714e+08	6.48e+07	5.730	0.000	2.44e+08	4.98e+08
C(업종명)[T. 세무사사무소]	3.697e+08	2.61e+08	1.419	0.156	-1.41e+08	8.81e+07	C(업종명)[T. 통신기기수리]	-9.141e+07	1.1e+08	-0.834	0.405	-3.06e+08	1.24e+08
C(업종명)[T. 세탁소]	-1.298e+08	4.12e+07	-3.150	0.002	-2.11e+08	-4.9e+07	C(업종명)[T. 패션푸드점]	1.313e+08	3.91e+07	3.360	0.001	5.47e+07	2.08e+08
C(업종명)[T. 슈퍼마켓]	2.579e+07	3.91e+07	0.660	0.509	-5.08e+07	1.02e+08	C(업종명)[T. 편의점]	3.467e+08	3.64e+07	9.527	0.000	2.75e+08	4.18e+08
C(업종명)[T. 스포츠강습]	4.774e+07	4.77e+07	1.001	0.317	-4.58e+07	1.41e+08	C(업종명)[T. 피부관리실]	1.535e+07	4.32e+07	0.356	0.722	-6.93e+07	1e+08
C(업종명)[T. 스포츠클럽]	-2.813e+07	4.32e+07	-0.651	0.515	-1.13e+08	5.65e+07	C(업종명)[T. 한복점]	-1.42e+08	9.61e+07	-1.478	0.140	-3.3e+08	4.64e+07
C(업종명)[T. 스포츠클럽]	1.198e+08	5.94e+07	2.016	0.044	3.31e+06	2.36e+08	C(업종명)[T. 한식음식점]	3.386e+08	3.92e+07	8.648	0.000	2.62e+08	4.15e+08
C(업종명)[T. 신발]	5.442e+07	4.25e+07	1.280	0.200	-2.89e+07	1.38e+08	C(업종명)[T. 한의원]	3.095e+08	4.16e+07	7.439	0.000	2.28e+08	3.91e+08
C(업종명)[T. 약기]	-1.258e+08	8.69e+07	-1.417	0.331	-3.75e+07	2.4e+08	C(업종명)[T. 핸드폰]	-1.503e+07	4.19e+07	-0.359	0.720	-9.72e+07	6.71e+07
C(업종명)[T. 애완동물]	9.331e+07	3.75e+07	2.4	0.000	-1.05e+08	-1.01e+08	C(업종명)[T. 호프-간이주점]	-1.06e+07	3.8e+07	-0.279	0.780	-8.51e+07	6.39e+07
C(업종명)[T. 약식음식점]	-7.467e+07	5.43e+07	-1.370	0.103	-1.01e+08	3.11e+07	C(업종명)[T. 회장품]	5.904e+07	4.62e+07	1.277	0.202	-3.16e+07	1.5e+08
C(업종명)[T. 여관]	2.102e+08	3.86e+07	5.451	0.000	1.35e+08	2.86e+08	C(업종명)[T. 회초]	-7.807e+07	3.76e+07	-2.074	0.038	-1.52e+08	-4.27e+06
C(업종명)[T. 여행사]	-1.539e+07	4.16e+07	-0.370	0.711	-9.7e+07	6.62e+07	Skew:	1.697	Prob(JB):	0.00			
C(업종명)[T. 예술학원]	1.173e+07	1.1e+08	0.107	0.915	-2.03e+08	2.27e+08	Kurtosis:	7.736	Cond. No.	64.0			
C(업종명)[T. 한구]	6.672e+07	3.95e+07	1.688	0.091	-1.08e+07	1.44e+08	Notes:						
C(업종명)[T. 외국어학원]	-8.264e+07	8.69e+07	-0.951	0.342	-2.53e+08	8.78e+07	[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.						



선정된 업종들

● 압구정동

: 편의점 / 서적 / 육류판매 / 미용실 / 호프-간이주점 / 한식음식점 / 중식음식점 / 양식음식점 / 화장품
외국어학원 / 커피-음료 / 전자상거래업 / 일식음식점 / 의약품 / 일반의류 / 운동-경기용품 / 스포츠클럽
한의원 / 신발 / 안경 / 치과의원 / 일반의원 / 제과점 / 가방 / 시계및귀금속 / 컴퓨터및주변장치판매
인테리어 / 슈퍼마켓 / 일반교습학원 / 패스트푸드점 / 예술학원 / 가구  **총 32개 업종**

● 신촌동

: 육류판매 / 호프-간이주점 / 예술학원 / 슈퍼마켓 / 일식음식점 / 화장품 / 운동-경기용품 / 커피-음료
조명용품 / 한식음식점 / 일반의원 / 외국어학원 / 중식음식점 / 편의점 / 분식전문점 / 패스트푸드점
일반의류 / 의약품 / 가방 / 신발 / 서적 / 안경 / 미용실 / 양식음식점 / 한의원 / 제과점 / 치과의원
스포츠클럽 / 시계및귀금속  **총 29개 업종**

Predict

3-2 Predict





SARIMA

- 시계열 데이터의 자기 회귀, 이동 평균 성분을 활용하여 예측하는 **ARIMA 모델**에 + 계절성까지 고려하여 예측할 수 있는 모델
- 단점 : 자동적으로 최적의 파라미터 설정이 불가해 사용자의 설정에 따라 모델 성능에 영향을 미침

ETS

- 시계열 데이터의 추세(Trend), 계절성(Seasonality), 잔차(Error)를 분해하고 패턴을 파악하여 미래 값을 예측할 수 있는 모델
- 단점 : 외생적인 요인을 고려하지 않음

외생적인 요인이 작용하는 매출 데이터 특성을 고려해 **SARIMA로 최종 결정**

Predict



3-2 예측 모델 선정 : 시계열 데이터 모델링

[Step 1.] 최적 파라미터 선정

방법 ①. 매출로 파라미터 산출

```
1 import itertools
2 p = range(0, 3)
3 d = range(0, 1)
4 q = range(0, 3)
5 pdq = list(itertools.product(p, d, q))
6 seasonal_pdq = [(x[0], x[1], x[2], 4) for x in pdq]
7
8 AIC = []
9 params = []
10 for i in pdq :
11     for j in seasonal_pdq :
12         try :
13             model = SARIMAX(df_yang['총매출'].values, order=(i), seasonal_order = (j))
14             model_fit = model.fit()
15             print(f'SARIMA : {i},{j} >> AIC : {round(model_fit.aic, 2)}')
16             AIC.append(round(model_fit.aic, 2))
17             params.append((i, j))
18
19         except Exception as e:
20             print(e)
21             continue
22
23
24 SARIMA : (0, 0, 0),(0, 0, 0, 4) >> AIC : 2.557955635328519e+16
25 SARIMA : (0, 0, 0),(0, 0, 1, 4) >> AIC : 1138.33
26 SARIMA : (0, 0, 0),(0, 0, 2, 4) >> AIC : 1127.11
27 SARIMA : (0, 0, 0),(1, 0, 0, 4) >> AIC : 1106.3
28 SARIMA : (0, 0, 0),(1, 0, 1, 4) >> AIC : 1100.20
29
30 optim = [(params[i], j) for i, j in enumerate(AIC) if j == min(AIC)]
31 print(optim)
32
33 [((1, 0, 0), (0, 0, 1, 4)), 1082.85]
```

방법②. auto_arima 모듈 활용

```
1 auto_arima_model = auto_arima(df_yang['총매출'], start_p=1, start_q=1,
2                                 max_p=3, max_q=3, m=4, seasonal=True,
3                                 d=1, D=1,
4                                 max_P=3, max_Q=3,
5                                 trace=True,
6                                 error_action='ignore',
7                                 suppress_warnings=True,
8                                 stepwise=False)
```

```
10 auto_arima_model.summary()
```

ARIMA(2,1,0)(1,1,0)[4]	: AIC=857.958, Time=0.04 sec
ARIMA(2,1,0)(1,1,1)[4]	: AIC=857.831, Time=0.09 sec
ARIMA(2,1,0)(1,1,2)[4]	: AIC=857.207, Time=0.14 sec
ARIMA(2,1,0)(2,1,0)[4]	
ARIMA(2,1,0)(2,1,1)[4]	Best model: ARIMA(0,1,1)(3,1,0)[4]
ARIMA(2,1,0)(3,1,0)[4]	Total fit time: 19.472 seconds

SARIMAX Results	
Dep. Variable:	y
No. Observations:	24
Model:	SARIMAX(0, 1, 1)x(3, 1, 0, 4)
Log Likelihood	-418.675
Date:	Fri 04 Aug 2023
AIC	847.349

☞ 두 가지 방법에서 나온 파라미터 값을 적용해보고,

오차 값(RMSE)가 낮은 파라미터로 선정하여 모델링

Predict

3-2 예측 모델 선정 : 시계열 데이터 모델링



[Step 2.] 모델 학습

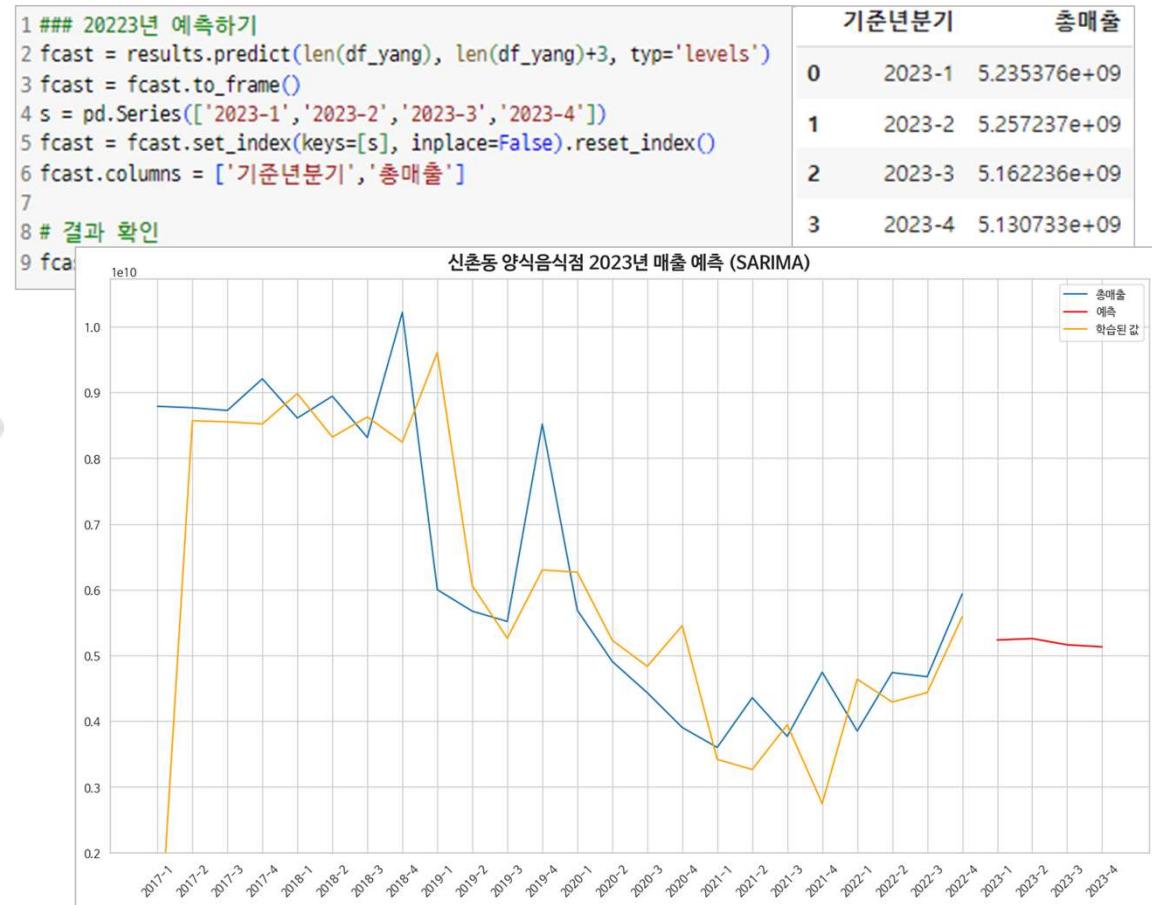
```
1 # 트레이닝/테스트 데이터 나누기
2 train = df_yang.iloc[:20]
3 test = df_yang.iloc[20:]

1 # 최적 파라미터 값으로 모델링
2 model = SARIMAX(train['총매출'], order=(1,0,0), seasonal_order=(0,0,1,4))
3
4 # 모델링 값 학습
5 results = model.fit()
6
7 # 결과 확인
8 results.summary()

SARIMAX Results
Dep. Variable: 총매출 No. Observations: 20
Model: SARIMAX(1, 0, 0)x(0, 0, [1], 4) Log Likelihood -450.430
Date: Fri, 04 Aug 2023 AIC 906.860
Time: 02:22:19 BIC 909.848
Sample: 0 HQIC 907.444
- 20

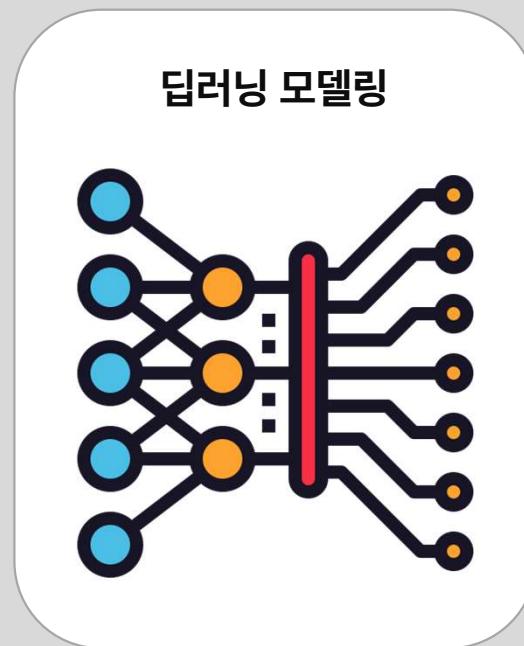
Covariance Type: opg
      coef    std err      z   P>|z| [0.025  0.975]
ar.L1  0.9577    0.061  15.829     0.000  0.839  1.076
ma.S.L4 0.4757    0.314   1.516     0.130 -0.139  1.091
sigma2 2.118e+18 1.36e-19 1.56e+37 0.000 2.12e+18 2.12e+18
Ljung-Box (L1) (Q): 1.20 Jarque-Bera (JB): 3.63
Prob(Q): 0.27 Prob(JB): 0.16
Heteroskedasticity (H): 1.66 Skew: -0.74
Prob(H) (two-sided): 0.52 Kurtosis: 4.47
```

[Step 3.] 2023년 매출 예측하여 시각화



Predict

3-2 Predict

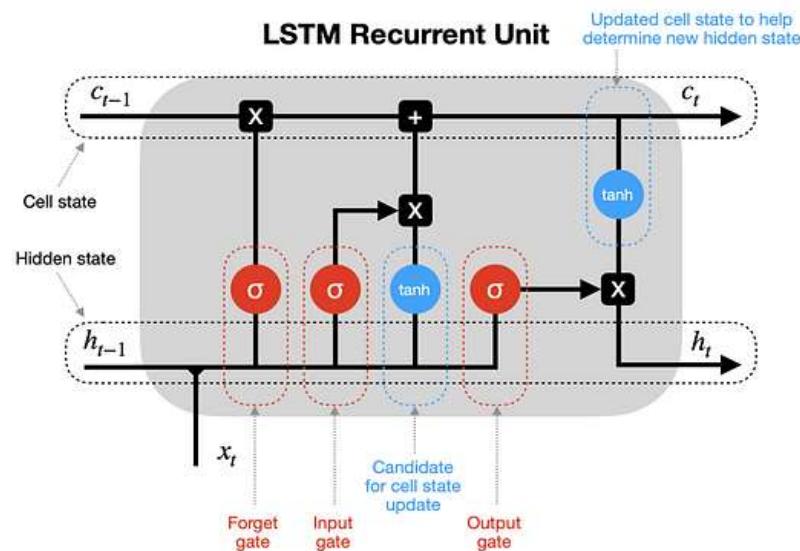




3-2 예측 모델 선정 : 딥러닝 모델링

LSTM

- RNN이 가지고 있던, 입력 데이터의 길이가 길어지는 경우에 생기는 기울기 소실(Gradient Vanishing) 문제가 개선된 모델
- 시간에 따른 주가 변동, 기온 변화, 특정 질병의 환자 수 변화 등 **시계열 데이터 예측**에 주로 쓰인다.



[여러 지표를 넣어서 LSTM 모델 생성]

Predict

3-2 예측 모델 선정 : 딥러닝 모델링



[Step 1.] 데이터 생성

```
### X / y 데이터 생성  
X = sin_wfood[['성장성', '시장성', '순점포증감률']]  
y = sin_wfood.loc[:, ['분기당_매출_금액']]
```

```
# 결과 확인하기  
print(X)  
print('-'*80)  
print(y)
```

	총_매출_금액	순점포증감률	시장성	성장성
0	25499817180	0.330033	2.799161e+06	1.000000
1	26698024549	-0.664452	5.674356e+06	1.046989
2	27541994924	1.633987	2.401260e+06	1.031612
3	31308246219	-0.327869	3.585769e+06	1.136746
4	29115910548	-6.271777	4.238515e+06	0.929976

[Step 2.] 정규화

```
### Normalization 진행  
  
# MinMaxScaler import  
from sklearn.preprocessing import MinMaxScaler  
  
# MinMaxScaler 객체 생성  
mmsx = MinMaxScaler()  
mmsy = MinMaxScaler()
```

```
# 스케일링  
X_mms = mmsx.fit_transform(X)  
y_mms = mmsy.fit_transform(y)  
  
# 결과 확인  
print(f'X_mms 모양 : {X_mms.shape}')  
print('-'*80)  
print(f'y_mms의 모양 : {y_mms.shape}')
```

```
X_mms 모양 : (28, 3)  
-----  
y_mms의 모양 : (28, 1)
```

[Step 3.] train, test 데이터 생성

```
# window_size=2인 X, y 데이터 생성  
X_data, y_data = make_dataset(X_mms, y_mms, 2)  
  
# 결과 확인하기  
print(f'X_data 데이터의 모양 : {X_data.shape}')  
print('-'*80)  
print(f'y_data 데이터의 모양 : {y_data.shape}')  
  
X_data 데이터의 모양 : (26, 2, 3)  
-----  
y_data 데이터의 모양 : (26, 1)
```

```
# 평가용 데이터 크기 설정  
TEST_SIZE = 4
```

```
# 학습용 데이터 생성  
X_train = X_data[:-TEST_SIZE, :, :]  
y_train = y_data[:-TEST_SIZE, :]
```

```
# 평가용 데이터 생성  
X_test = X_data[-TEST_SIZE:, :, :]  
y_test = y_data[-TEST_SIZE:, :]
```

Predict



3-2 예측 모델 설정 : 딥러닝 모델링

[Step 4.] LSTM 모델 생성

```
# 랜덤 시드 설정
random.seed(0)
tf.random.set_seed(0)
np.random.seed(0)
initializer = tf.keras.initializers.GlorotUniform(seed=0)

# 모델 정의
model = tf.keras.Sequential()
model.add(tf.keras.layers.LSTM(units=64,
                               return_sequences=True,
                               kernel_initializer=initializer,
                               input_shape=X_train[0].shape))
model.add(tf.keras.layers.LSTM(units=32, kernel_initializer=initializer))
model.add(tf.keras.layers.Dense(1, kernel_initializer=initializer))

# 모델 컴파일
model.compile(loss='mean_squared_error', optimizer = Adam(learning_rate = 0.001))

# 모델 확인
model.summary()

Model: "sequential_16"
-----  
Layer (type)          Output Shape         Param #
-----  
lstm_32 (LSTM)        (None, 2, 64)       17408  
lstm_33 (LSTM)        (None, 32)          12416  
dense_16 (Dense)      (None, 1)           33  
-----  
Total params: 29,857  
Trainable params: 29,857  
Non-trainable params: 0
```

[Step 5.] LSTM 모델 학습

```
# early stop 설정
early_stop = tf.keras.callbacks.EarlyStopping(monitor='val_loss',
                                              verbose=1,
                                              patience=2)

# ModelCheckpoint 설정
save_file_path = 'stock_model.h5'
checkpoint = tf.keras.callbacks.ModelCheckpoint(save_file_path,
                                                monitor='val_loss',
                                                verbose=1,
                                                save_best_only=True)

# 학습
history = model.fit(X_train,
                      y_train,
                      epochs=100,
                      validation_split=0.4,
                      callbacks=[early_stop, checkpoint])

Epoch 1/100
1/1 [=====] - ETA: 0s - loss: 0.4097
Epoch 1: val_loss improved from inf to 0.02842, saving model to stock_model.h5
1/1 [=====] - 6s 6s/step - loss: 0.4097 - val_loss: 0.0284
Epoch 2/100
1/1 [=====] - ETA: 0s - loss: 0.3970
Epoch 2: val_loss improved from 0.02842 to 0.02545, saving model to stock_model.h5
1/1 [=====] - 0s 138ms/step - loss: 0.3970 - val_loss: 0.0255
```

Predict

3-2 예측 모델 선정 : 딥러닝 모델링



[Step 5.] 예측과 시각화

```
# 테스트 데이터를 이용하여 예측하기  
pred = loaded_model.predict(X_test)  
print(pred)
```

```
1/1 [=====] - 1s 841ms/step  
[[0.13471709]  
[0.13036254]  
[0.12495787]  
[0.12391124]]
```

```
### 평가용 데이터에 대한 예측의 결과 --> 데이터프레임 생성  
df_test = pd.DataFrame(y_test, columns=['y_test'])  
df_test['predict'] = pred  
date = df_retail.iloc[-4:, 0]  
df_test.index = date
```

```
# 결과 확인하기  
print(df_test)
```

	y_test	predict
분기		
2023-1	0.394868	0.492979
2023-2	0.431277	0.513105
2023-3	0.528249	0.514893
2023-4	0.519651	0.507244

```
### 결과 시각화  
plt.figure(figsize=(12, 6))  
plt.plot(df_retail['분기'], df_retail['총_매출_금액'], color='green', label = 'Original')  
plt.plot(df_test['분기'], df_test['test_origin'], color='blue', label = 'test_origin')  
plt.plot(df_test['분기'], df_test['predict_origin'], color='red', label = 'predict_origin')  
plt.xlabel('time')  
plt.ylabel('price')  
plt.xticks(rotation=45)  
plt.yticks(rotation=45)  
plt.legend()  
plt.show()
```



Predict

3-2 Predict



- ① 모델을 사용한 분석 결과와 GPT에 넣어서 얻은 결과를 비교
- ② 트렌드인 GPT를 이용해보는 새로운 경험

Predict

3-2 Chat-GPT로 예측하기



test size = 4로 해서 train, test 데이터 분리해줘.

다음 4분기 예측을 진행해줘!

김은성의 어시스턴트 GPT4

주어진 데이터프레임을 train, test 데이터로 분리하겠습니다. 이를 위해 pandas의 'iloc' 메서드를 사용합니다.

LSTM 모델을 생성하고 학습시켜줘.

김은성의 어시스턴트 GPT4

시계열 데이터를 다루기 위해 LSTM 모델을 생성하고 학습시키겠습니다. 먼저, 필요한 라이브러리를 임포트하고 데이터를 전처리한 다음, 모델을 정의하고 학습시키겠습니다.

```
python
import numpy as np
from keras.models import Sequential
from keras.layers import LSTM, Dense
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import mean_squared_error

# 데이터 전처리
train_values = train_data['총매출_정규화'].values.reshape(-1, 1)
test_values = test_data['총매출_정규화'].values.reshape(-1, 1)
```

김은성의 어시스턴트 GPT4

다음 4분기의 예측을 진행하겠습니다. 먼저, 학습한 LSTM 모델을 사용하여 다음 4개의 예측값을 구해볼 수 있습니다. 다음 코드는 학습된 LSTM 모델을 사용하여 4분기 동안의 예측을 수행하며 결과를 출력합니다.

python

0에서 3까지 반복
for i in range(4):
 # 예상되는 데이터를 입력합니다.
 prediction = model.predict(last)
 # 예상되는 데이터를 출력합니다.
 print(prediction)

마지막 예상되는 데이터를 출력합니다.
last = prediction

다음 4분기 예측 결과(총매출):

분기 1: 10223016852.57
분기 2: 10081244524.17
분기 3: 9957476986.89
분기 4: 9858654461.02

이렇게 학습된 LSTM 모델을 사용하여 다음 4분기 동안의 총매출을 예측할 수 있습니다. 그러나, 이 예측 결과는 현재 LSTM 모델의 성능과 데이터셋에 따라 변동될 수 있으며, 실제 결과와 다를 수 있습니다. 특히 LSTM이 갖는 한계와 시계열 도메인의 특성에 따라 더 성능이 좋은 예측 방법을 찾아 적용해볼 필요가 있습니다.



3-2 Chat-GPT로 예측하기



Chat GPT 를 활용, 업종별 총 매출 금액 데이터에 기반하여 1년 뒤 매출 상승이 기대되는 업종을 예측

업 구 정 동

1년 뒤의 업종 매출 순위 예측:
 화장품: 106.12%
 양식음식점: 97.64%
 운동/경기용품: 8.28%
 커피-음료: 7.49%
 한식음식점: 5.41%
 슈퍼마켓: 3.95%
 미용실: 3.80%
 패스트푸드점: 3.65%
 일반교습학원: 3.63%
 서적: 2.95%
 컴퓨터및주변장치판매: 0.93%
 인테리어: 0.85%
 예술학원: 0.39%
 일반의류: -0.12%
 가구: -0.67%
 육류판매: -1.59%
 중식음식점: -2.20%
 일반의원: -2.25%
 전자상거래업: -2.38%
 외국어학원: -3.56%
 치과의원: -4.07%
 조명용품: -4.71%
 의약품: -6.72%
 스포츠클럽: -11.97%
 제과점: -18.72%
 편의점: -20.09%
 일식음식점: -23.73%
 한의원: -32.49%
 호프-간이주점: -41.00%
 시계및귀금속: -43.87%
 안경: -57.32%
 자전거 및 기타운송장비: -60.80%
 신발: -67.66%
 가방: -86.20%

신 촌 동

1년 뒤의 업종 매출 순위 예측:
 화장품: 49.54%
 신발: 15.79%
 커피-음료: 14.61%
 예술학원: 9.37%
 시계및귀금속: 6.90%
 호프-간이주점: 3.17%
 육류판매: 1.82%
 치과의원: 1.38%
 의약품: 0.53%
 서적: 0.40%
 한의원: 0.18%
 가전제품: -0.15%
 한식음식점: -0.16%
 편의점: -0.68%
 안경: -2.71%
 양식음식점: -2.98%
 슈퍼마켓: -5.56%
 일식음식점: -7.64%
 일반의원: -9.56%
 인테리어: -9.73%
 중식음식점: -10.39%
 제과점: -13.51%
 스포츠클럽: -13.91%
 컴퓨터및주변장치판매: -14.60%
 일반교습학원: -16.90%
 조명용품: -16.98%
 일반의류: -17.10%
 외국어학원: -26.22%
 패스트푸드점: -28.45%
 운동/경기용품: -34.44%
 분식전문점: -44.47%
 가방: -47.73%
 미용실: -48.70%

04

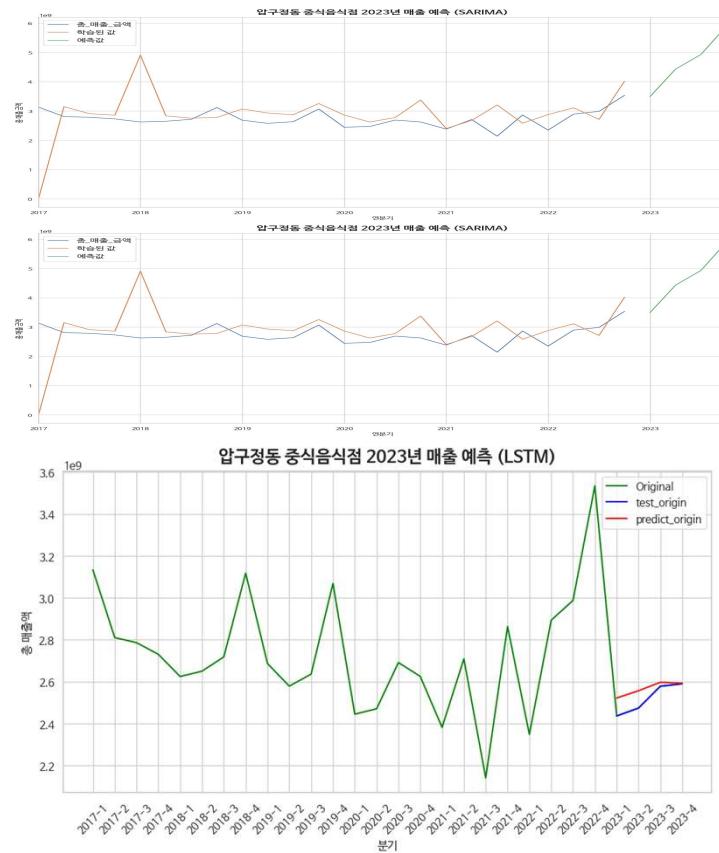
Conclusion

Conclusion

4-1 결론 – 추천/비추천 업종



시계열(SARIMA)와 딥러닝(LSTM)의 예측 그래프가 모두 상승이면 추천, 하락이면 비추천 업종으로 분류



Conclusion

4-1 결론 – 추천/비추천 업종



시계열(SARIMA)와 딥러닝(LSTM)의 예측 그래프가 모두 상승이면 추천, 하락이면 비추천 업종으로 분류

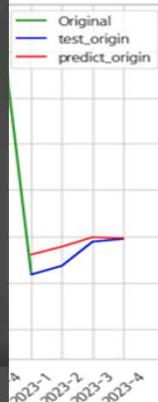
▷ 압구정동 추천

한식음식점(18.2%) / 중식음식점(20.8%)
/ 양식음식점(13.2%) / 화장품(12.3%)
/ 외국어학원(5.5%) / 커피-음료(4.7%)
/ 전자상거래업(15.4%) / 일식음식점(13.1%)
/ 의약품(29.1%) / 일반의류(2.9%)



▷ 신촌동 추천

중식음식점(21.7%)
/ 편의점(5.1%)



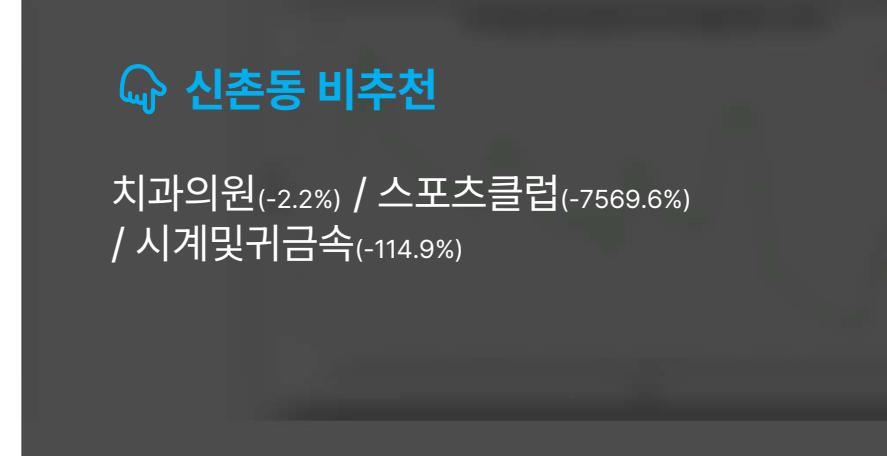
▷ 압구정동 비추천

일반교습학원(-8.3%) / 패스트푸드점(-6.1%)
/ 예술학원(-12.3%) / 가구(-55.7%)



▷ 신촌동 비추천

치과의원(-2.2%) / 스포츠클럽(-7569.6%)
/ 시계및귀금속(-114.9%)



Conclusion

4-1 결론 – GPT vs 시계열과 딥러닝



압구정동

화장품: 106.12%
양식음식점: 97.64%
운동/경기용품: 8.28%
커피-음료: 7.49%
한식음식점: 5.41%
슈퍼마켓: 3.95%
미용실: 3.80%
패스트푸드점: 3.65%
일반교습학원: 3.63%
서적: 2.95%
컴퓨터및주변장치판매: 0.93%
인테리어: 0.85%
예술학원: 0.39%
일반의류: -0.12%
기구: -0.67%
육류판매: -1.59%
중식음식점: -2.20%
일반의원: -2.25%
전자상거래업: -2.38%
외국어학원: -3.56%
치과의원: -4.07%
조명용품: -4.71%
의약품: -6.72%
스포츠클럽: -11.97%
제과점: -18.72%
편의점: -20.09%
일식음식점: -23.73%
한의원: -32.49%
호프-간이주점: -41.00%
시계및귀금속: -43.87%
안경: -57.32%
자전거 및 기타운송장비: -60.80%
신발: -67.66%
가방: -86.20%

추천 업종

비추천 업종

신촌동

화장품: 49.54%
신발: 15.79%
커피-음료: 14.61%
예술학원: 9.37%
시계및귀금속: 6.90%
호프-간이주점: 3.17%
육류판매: 1.82%
치과의원: 1.38%
의약품: 0.53%
서적: 0.40%
한의원: 0.18%
가전제품: -0.15%
한식음식점: -0.16%
편의점: -0.68%
안경: -2.71%
양식음식점: -2.98%
슈퍼마켓: -5.56%
일식음식점: -7.64%
일반의원: -9.56%
인테리어: -9.73%
중식음식점: -10.39%
제과점: -13.51%
스포츠클럽: -13.91%
컴퓨터및주변장치판매: -14.60%
일반교습학원: -16.90%
조명용품: -16.98%
일반의류: -17.10%
외국어학원: -26.22%
패스트푸드점: -28.45%
운동/경기용품: -34.44%
분식전문점: -44.47%
가방: -47.73%
미용실: -48.70%

[시계열과 딥러닝을 활용하여 예측한 결과와 Chat-GPT가 예측한 결과를 비교했을 때, 예측도가 많이 떨어진다는 것을 알 수 있다.]

Conclusion



4-1 결론 – 가설검증 (압구정동)

“ 압구정동의 상권은 코로나의 영향을 받지 않고, 꾸준히 성장 중이며 또는 업종은 주로 외식업에 있을 것이다.”

압구정동

경제

‘가논강압’ 상권, 잘나가는 비결(2) ‘국밥’ 잘 나가는 강남?...압구정은 ‘신구조화’

나건웅 기자 wasabi@mk.co.kr
조동현 기자 cho.donghyun@mk.co.kr
입력 : 2023-04-10 09:51:01 수정 : 2023-04-10 09:51:31

가

입구정로데오: 다양한 포트폴리오

주말 매출 증가 유일…명실상부 ‘핫플’

압구정로데오는 다양한 포트폴리오에 힘입어 선전한 케이스다. 비교적 크지 않은 상권임에도 불구하고 외식을 비롯해 카페, 주점, 의료, 패션, 전시 등 다양한 업태의 매장이 오밀조밀 끌고루 자리 잡고 있다.

의료 매출이 두드러진 타 지역 상권과 달리 외식 업종도 힘을 냈다. 특히 최근 MZ세대 사이에서 인기가 많은 **‘바 카페’**가 무려 8000%가 넘는 성장을 보였다. 2019년 3억원이 채 안 됐지만 2022년에는 232억원에 달했다. 이 밖에도 양식(131억원), 한식-백반(102억원), 커피 전문점(65억원) 등이 고른 성장을 보였다.

 **압구정동 추천**

- 한식음식점(18.2%) / 중식음식점(20.8%)
- / 양식음식점(13.2%) / 화장품(12.3%)
- / 외국어학원(5.5%) / 커피-음료(4.7%)
- / 전자상거래업(15.4%) / 일식음식점(13.1%)
- / 의약품(29.1%) / 일반의류(2.9%)

도이가 물려 있어 깊은 흥사이에 더
많이 찾는 핫플레이스가 하니들씨 늘며 유동 인구가 늘었고 그 결과 인근에 음식점 및 카페들이 또다시 생겨나
한구조가 직동하고 있다.

상권의 부활

시울 전통 관광명소인 명동 동대문홍대에 이어 입구성과 경기 대형 스크린을 갖춘 Z
가이드북 대신 인스타그램과 유튜브 등 SNS로 관광지를 찾거나, K팝 스타가 언급한 상소를 찾는 Z
세대(1990년대 중반~2000년대 초반 출생) 외국인 관광객의 영향으로 풀이된다.
CI 물리브영이 외국인 매출 집중 상권을 분석한 결과, 명동과 동대문, 홍대 등 전통 관광 상권에는
한국 초행자나 가족 단위 관광객이, 입구점과 성수 상권에는 20대 젊은 층이 주로 방문한 것으로 나
타난다.

압구정에는 젠틀몬스터의 디저트 가게인 '뉴데이크 하우스'와 '런던 베이글 뮤지엄', '노티드', 카멜 커피' 등 소셜 미디어에서 인기 있는 맛집들이 집중해 있다. 성수는 패션 및 뷰티 브랜드의 팝업스토어가 몇 개 있어 젊은 층 사이에서 트렌디한 곳으로 손꼽히는 곳이다.

앞으로도 성장할 상권으로 예상되며, 외식업이 더욱 성행하고 있음을 확인

Conclusion

4-1 결론 – 가설검증 (신촌동)



신촌동

“신촌동 상권은 꾸준히 하락세였으나 앞으로는 올라갈 것이며, ‘카페/베이커리’ 업종의 상승세가 뚜렷할 것이다.”

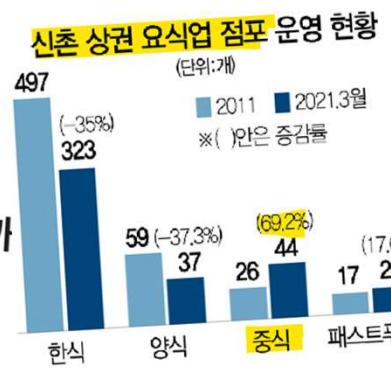
마라탕의 인기는 어디까지?

[르포] “신촌 상권 매출 상승 중”...
등록일 2022.11.01 19:16 | 게재일 2022.11.01
조문한 기자 | 입력 2023.06.27 18:38 | 댓글 0

마라탕의 유행은 계속되고 있다. 네이버 기준
량 조회수는 총 40만건으로 나타났다. 연령별 ◇ “대중교통전용지구 해제 효과 있어”... 반대하던 식당 ‘점심시간 손님’
이런 가운데 최근 신촌상권 매출이 상승했다는 조사결과가 나와 관심이 집중되고 있다. 대중교통전용지구 해제 효과
로 해석할 수 있어서다. ‘나이스지니데이터’에 따르면 2023년 1분기 신촌상권의 월평균 매출액은 501억원이며 2019
년 1분기 480억원 대비 21억원이 증가했다. 16일 한국경제신문 보도에 따르면 지난 3월부터 5월 기간 BC카드 이용
은 2019년 동기 대비 카드 이용액이 12.1%, 결제건수는 14.3%가 늘었다.

신촌동 추천
중식음식점(21.7%)
/ 편의점(5.1%)

해제될까



물론 이들 업종도 코로나19의 타격을 피하지는 못했지만 전체 평균이나 한식·양식 매장
간 감소 폭에 비해 충격이 덜했다. 해외여행이 보편화되고 주변 국가에서 조리법을 체득해 실
제 창업으로 이어진 결과가 늘어났다는 분석이다. 실제로 초밥이나 라멘 같은 일식 메뉴나
훠궈·양꼬치·마라탕 등을 파는 중식당을 주변에서 쉽게 찾아볼 수 있게 됐다. 패스트푸드
메뉴 역시 다양화되는 추세다.

대중교통 전용지구 해제 + 업종 제한 폐지 수순을 밟고 있어 신촌 상권의 매출 상승이 기대됨
카페/베이커리의 경우 1년 생존률이 낮아지고 있어 쇠퇴하는 업종이 될 듯
마라에 대한 식지 않는 인기 + 대학가 특성상 회전률이 높은 중식음식점이 성행할 것으로 확인됨

Conclusion

4-1 결론 깊이 보기

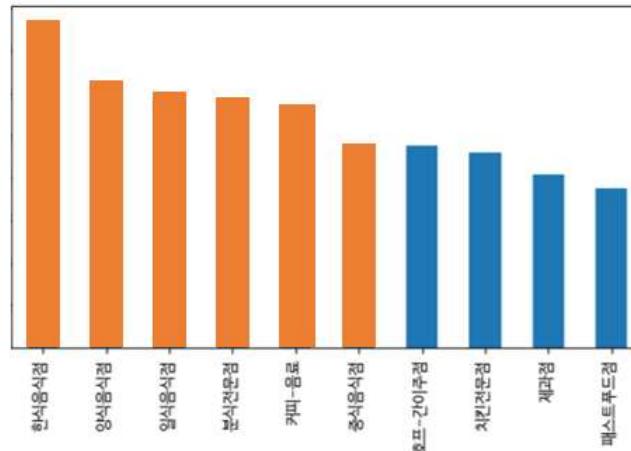


압구정동 추천 업종 ① 외식업

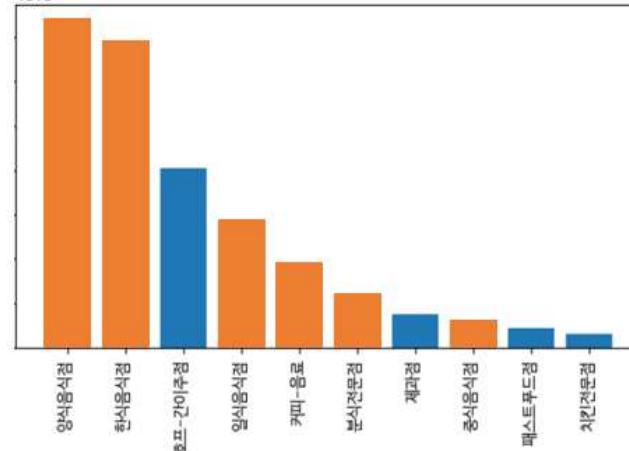
양식음식점, 한식음식점, 일식음식점, 중식음식점

팬데믹 기간에 발생한 많은 공실에 대응하기 위해 임대료를 낮추며 음식점 유입 성공!

강남구_압구정동 외식 상점수



강남구_압구정동 외식 총 매출 금액



- 한식, 양식음식점 : 매출과 상점수에서 모두 TOP3
- 특히 양식음식점 분기 평균 매출은 2020년 4분기 대비 2022년 4분기 **55.39% 증가**
- 199개였던 양식음식점 매장수 → 2년 새 241개로 증가

Conclusion

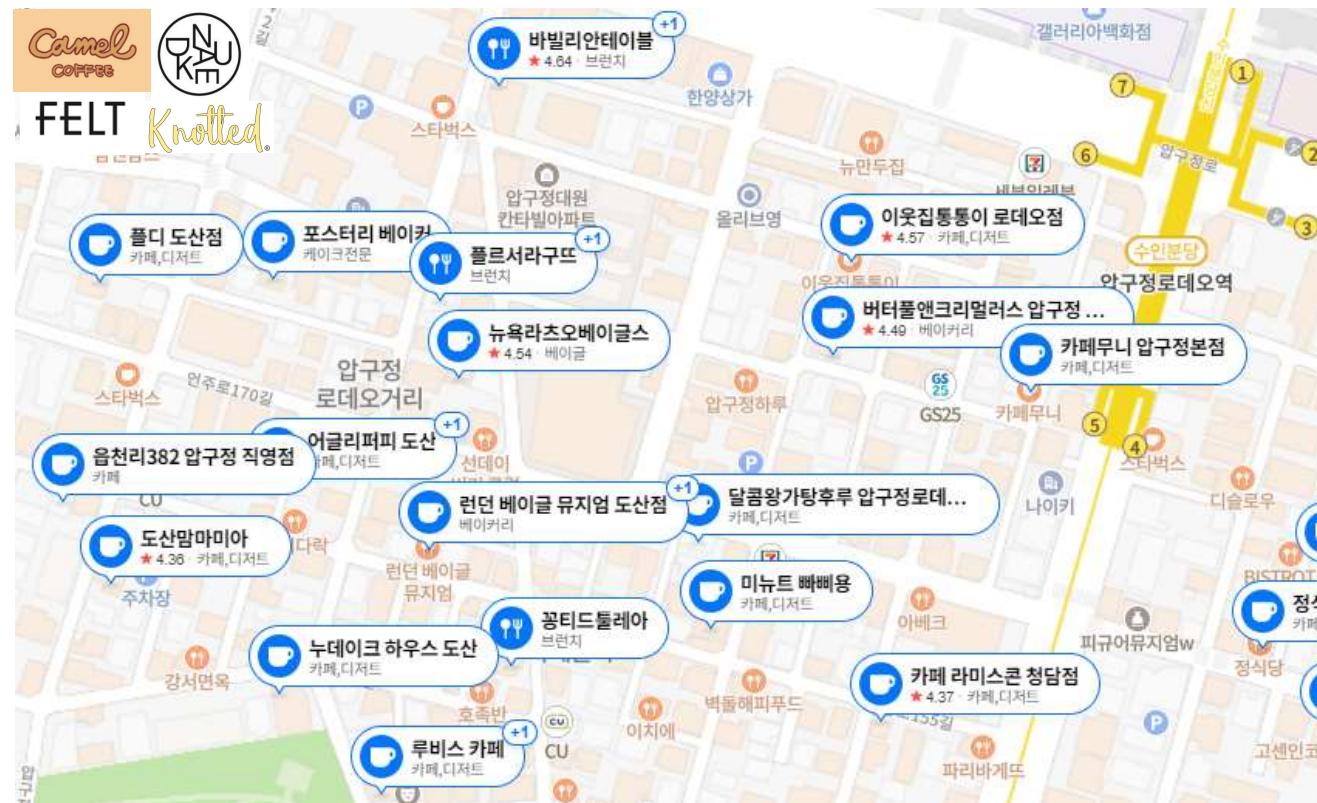


4-1 결론 깊이 보기

압구정동 추천 업종 ② 카페

특정 상권을 대표하거나 대형 상가의 핵심이 되는 61명 접포
Ex. 노터드 청담, 카멜 커피, 펄트 커피, 누데이크 하우스, 꽁티드 풀레아 등

앵커스토어가 발달하며 보다 소규모인 카페의 유입을 도와 압구정 카페거리 형성



Conclusion

4-1 결론 깊이 보기



압구정동 추천 업종 ③ 의류, 전자상거래업

☞ 앵커스토어

- 프리미엄 요가브랜드 '룰루레몬' 아시아에서 가장 큰 매장 위치
- 언더아머 플래그십 스토어



☞ 트렌디한 편집숍

- 압구정 로데오거리의 정체성이라고도 할 수 있는 편집숍
- 편집숍의 온라인 매장 운영 여파로 전자상거래업 또한 추천업종

☞ 백화점

- 2018~2022년 기준 강남구의 중심관광지 1,3위:
- 현대백화점 압구정본점, 갤러리아 백화점/명품관EAST

인기 관광지			
순위	이름	주소	구분
1	현대백화점압구정본점	서울 강남구 압구정로 165-0	백화점

Conclusion

4-1 결론 깊이 보기

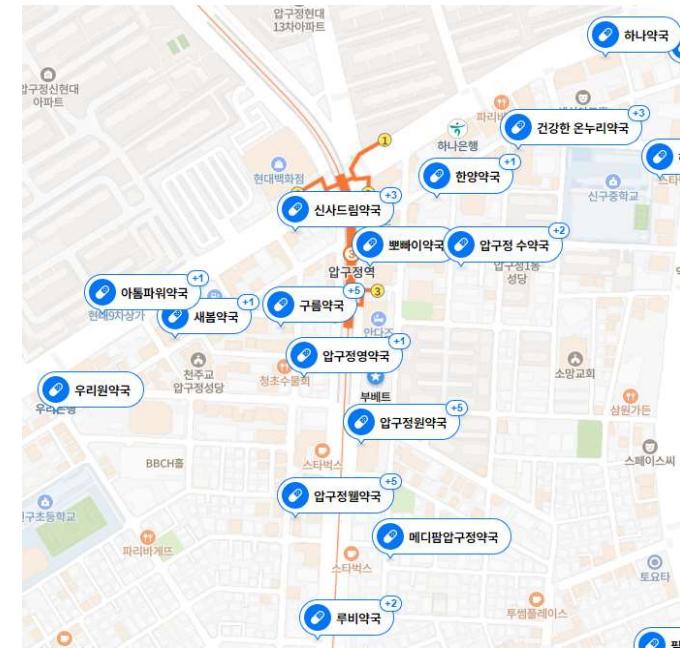


압구정동 추천 업종 ④ 화장품, 의약품



올리브영 압구정지점은 화장품 마니아, 일명 '코덕'을 위한 제품군 판매량 많음
K-콘텐츠 사업의 성행으로 많은 외국인 관광객이 한국 화장품에 대한 관심 증가

ex. 색조 제품, 기능성 화장품, 이너뷰티, 그리고 미용 소도구 등



일반의원의 매출은 서비스 부문에서 가장 높지만,
점포수는 219 > 216으로 감소 추세에 있음

반면, 공생관계인 약국은 45 > 49로 증가 추세
(2020-4분기 22년-4분기 비교)

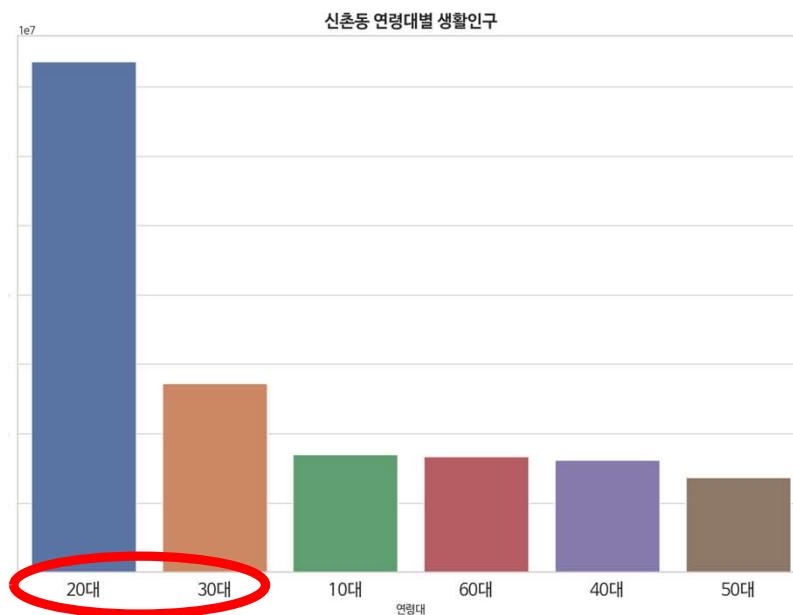
Conclusion

4-1 결론 깊이 보기



신촌동 추천 업종 ① 편의점

[신촌동 내 편의점 특수는 지속될 것으로 예측]



20대 생활인구수가 압도적으로 높음



서울시 내에서 인구대비 1인 자취가구 수가 많은 행정동 3위

Conclusion

4-1 결론 깊이 보기



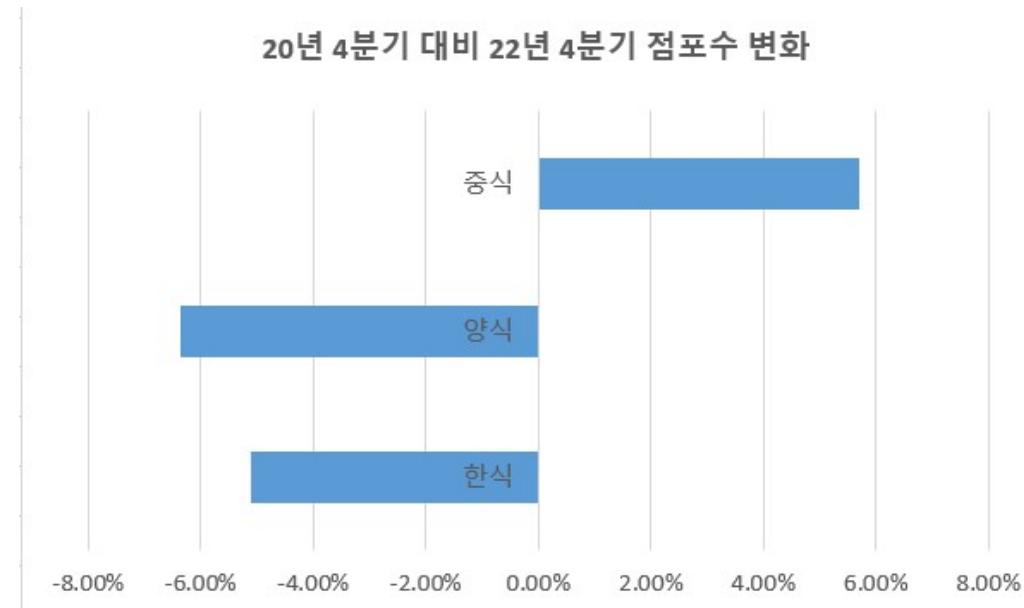
신촌동 추천 업종 ② 중식음식점



MZ세대를 중심으로 중국음식 인기

코로나19의 여파에도 중식의 인기로 식생활 트렌드를 빠르게 파악해
변화한 업체만 살아남은 것으로 분석

20년 4분기 대비 22년 4분기 점포수 변화



신촌에서 한식과 양식의 변동성이 각각 -5.11%, -6.35%로 감소하였지만
중식은 5.71%로 증가

Conclusion

4-2 기대효과



예비창업자가 얻을 수 있는 것

압구정동과 신촌동의 지역 상권의 특징을 각각 파악하고, 창업 시 업종 선택에 참고할 수 있다.

이로써 **예비 창업자의 핵심 역량인 시간에 따라 변화한 상권과, 소비자 니즈 파악 능력 향상이 기대된다.**

Conclusion

4-3 3L 회고



😊 Liked, 좋았던 것

- 협업, 의사소통을 통해 어려운 주제들을 팀원들과 함께 정하고 풀어나갔던 것 (즐거운 회의 시간♪)
- '할 일'의 우선순위를 정하고 확정된 건마다 '왜' 해야 하는지 이유를 검증했던 것
- 진행 프로세스를 모르는 사람이 갖게 될 의문, 의심을 지속해서 고민하여 방향성을 잡았던 것
- 처음 접하는 각종 툴을 기꺼이 활용했던 것
- 전반적으로 세미프로젝트보다 나아졌고, 모델링에 대한 아쉬움을 해소할 수 있었다.

😢 Lacked, 아쉽거나 부족했던 것

- 6년치의 데이터였지만, 더 많은 양의 데이터를 확보할 수 있었다면 오차를 줄일 수 있었을 텐데.. (실질적으로 분석한 행은 24행에 불과했다는 사실 😞)
- 도메인에 대한 '충분한' 지식과 조사가 부족해 주제가 2번 변경되었다.
- 두 번의 주제 변경으로 최종 주제의 분석 고도화를 할 수 있는 시간을 확보하지 못해 아쉬웠다.

☺ Learned, 배운 것

- 기획이 반이다! 구체화의 중요성, 탄탄한 토대가 좋은 프로젝트를 만든다.
- 아이디어란 한 사람의 머리에서 나오는 게 아닌 의견 교환 속에서 다듬어진다!
- 같은 결과라도 목적과 사람에 따라 다르게 해석할 수 있다.
- 하나의 단어, 조사에 따라 문장의 의도는 다르게 읽히는 구나.

Conclusion

4-4 참고문헌



저자(연도)	증속변수	독립변수	
<상권 배후지 소득에 따른 골목상권 매출변화 분석(COVID - 19 발생을 중심으로) - 링크			
황재석(2021)	골목상권 매출	상권 배후지 소득(고소득, 중소득, 저소득)	
<소매업 업종별 매출에 영향을 미치는 소득과 물리적 입지특성 분석 : 서울시 외식업, 수퍼마켓, 휴대폰 판매점을 중심으로>			
성은영(2018)	가계의 소비지출항목 (외식비, 식료품비, 자동차구입, 의류비, 의료 서비스비, 이미용 서비스, 학원비, 가계 지출비)	가구경제특성 가구인구통계 학적 특성	가처분소득, 비경상소득, 월평균 부채 상환액 가구원수, 가구주 나이, 가구주 나이 제곱, 가구주 학력(기준 : 고등학교 졸업 이하), 보유 자동차 대수, 주거면적

<코로나 19 시기에 밀레니얼세대가 상권 매출에 미치는 영향 분석 : 서울시 골목상권과 발달상권을 중심으로>		
정우영(2022)	면적 당 매출액	밀레니얼세대의 인구 비율, 유동인구 수의 평균, 상주인구 수의 평균, 다양성 지수(업종 다양성을 나타냄), 지하철역까지의 거리, 대학교까지의 거리, 도시공원까지의 거리, 버스정류장 개수

<서울시 골목상권 매출액에 영향을 미치는 요인에 관한 연구>			
김현철, 이승일(2019)	매출액	상권특성	연령별 매출비율, 시간대별 매출비율, 업종별 매출비율, 업종별 밀집도, 업종 수, 상권면적, 상가밀도, In 유동인구
		도시공간 구조 특성	도심·부도심, 용도지역, 발달상권
		배후지역 특성	In배후지상주인구, In배후지주정소득, 대형유통시설
		통제변수	건물노후도, 건폐율, 용적률, 주거용도 건물 수, 비주거용도 건물 수, In공시지가, In지하철 승하차 인원, In버스 승하차 인원

<골목상권 매출변화에 영향을 미치는 상권 특성 연구>		
김지원(2018)	2년 매출액 변화	매출 수준, 업종다양성 수준, 점포 수 수준, 상주인구 수준, 유동인구 수준, 아파트시세 수준, 점포밀도 수준, 업종다양성 변화, 유동인구 변화, 아파트시세 변화, 임대시설 변화, 상주인구 변화, 점포밀도 변화 제곱, 유동인구 변화 제곱, 업종다양성 변화 제곱

<공간 빅데이터를 활용한 소지역 상권매출에 영향을 미치는 요인분석에 관한 연구>			
이명호(2017)	카드매출	인구	주거인구수, 유동인구
		고용	직장인구수
		소득 수준	지역의 평균소득
		집적	지역상권 상점의 수

저자(연도)	분석 방법(모형)	연구(논문)
강민경(2022)	LSTM	서울시 상권변화 유형 도출 및 예측 모델을 활용한 변화예측
이동현(2022)		LSTM을 기반으로 한 서울시 성장상권과 쇠퇴상권 내 외식산업의 생존율 예측
김미성(2020)		서울시 상권 데이터의 시각화에 기반한 매출액 예측

저자(연도)	분석 방법(모형)	연구(논문)
노윤승, 도명식(2015)	SARIMA	SARIMA 모형을 이용한 철도 여객 단기 수송 수요 예측
김재호, 김장영(2022)		SARIMA 모형을 이용한 코로나19 확진자 수 예측

Q & A
