예비창업자들을위한 상권분석서비스

스타트라이팀 홍승표 / 서단우 / 조은진 / 차지인 / 함은정

() 팀원소개

서단우

상권 분석용 데이터 수집, 데이터 분석, 데이터 전처리, 오피니언 마 이닝 및 모델 학습, 데이터 시각화

차지인

기획 발표 준비, 오피니언 분석용 데 이터 수집, 데이터 분석, 데이터 시 각화 (매출 분석), 언어 모델 비교

홍승표

기획 발표 준비, 오피니언 분석용 데이터 수집, 데이터 분석, 데이터 시각화(업종 분석 및 시각화 전반) 웹 크롤링, 오피니언 분석용 데이 터 수집, 데이터 분석, 오피니언 마 이닝, 데이터 시각화(인구 분석)

조은진

상권 분석용 데이터 수집, 데이터 분석, 감성분석 모델 테스트, 데이 터 시각화(지역 분석), 최종 발표 자료 제작

함은정

목차

- 1
 제안배경
- 2 프로젝트소개
- 3 구현 및시행착오
- 4 시연영상
- 5 기대효과 및 활용방안

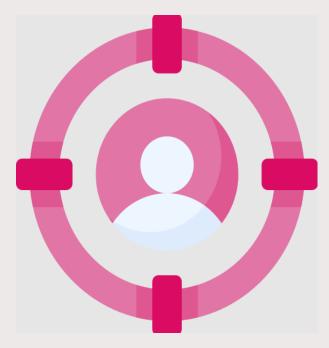
Part 1, 제안배경

1제안배경서비스소개

START STARTLIGHT LIGHT

스타트 라이트는 창업을 준비하는 예비 창업자들의 환한 앞날을 향해 나아갈 것입니다

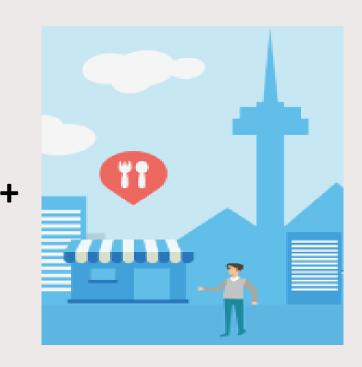
1 **제안배경** 서비스 소개



이용자가 아이템, 타겟층, 위치 선택



해당 창업 아이템의 적합도를 점수로 표현



조건에 대한 적합한 창업을 추천

1 제안배경 제안배경 및 필요성



1 제안배경 제안배경 및 필요성



폐업으로 이어지는 점포의 증가

 창업에 대한 수요 증가와 반대로, 지속되지 못하고 5년 이내 폐업하는 점포의 비율이

10곳 중 7곳

● 생존 기간

: 프랜차이즈 가맹점 > 비가맹점

: 공동 사업자 > 개인 사업자

Why? "정보의 차이"

해당 지역의 의견을 반영한 **적절한 창업 정보**를 알려준다면?

1 제안배경 경쟁서비스분석

사건 사건 카카	소상공인 상권정보
업종 범위	창업 가능한 모든 범위의 업종
장점	 대한민국전지역의업종,매출인구, 지역분석등보고서를통해 창업안정성등을수치적으로제공 입지업종분석보고서를통해, 도로변의업종에대한정보제공
단점	UI의 복잡성

마이프차	마이프차
업종 범위	브랜드 점포
장점	• 브랜드점포위치정보제공, 동단위의추천업종제공
	• 창업비용,월평균매출액, 1평당월평균매출액제공
단점	제휴 관계의 브랜드만 표시 됨

VS

1 제안배경 ^{차별성}

통계기반상권분석



SNS기반상권분석 Google Maps

Part 2, 프로젝트소개

2 프로젝트 소개 개발목표

개발 목표

빅데이터를 이용한 오피니언 마이닝 기반 창업 추천 서비스

개발 내용

내용

통계 데이터를 이용한 상권 분석





사용 Tool

Pycharm, Tableau 등을 이용하여 확인 및 분석

오피니언 마이닝을 이용한 창업 추천



Selenium, BS을 사용, 웹 크롤링을 통한 오피니언 데이 터 사용

지도 API를 이용한 위치 정보 조회



Open Street Map을 사용

프로젝트 소개 제안 배경 및 필요성

달성 목표	달성 전략							
데이터 분석 유효성 확인	TO SERVICE THAT HE SERVICE THE SERVICE THAT HE							
오피니언 마이닝 모델 학습	###							



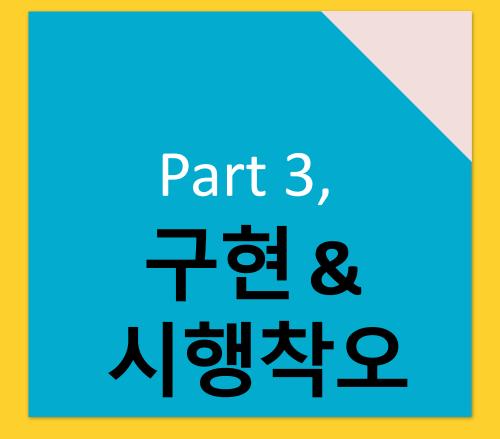
2 프로젝트소개 swor 분석

- 다양한 시각화 자료를 제공함으로써 사용자들이 직접 인사이트를 도출할 수 있게 함
- 예비 창업자들에게 창업 아이템에 대한 수치 점수 제공 S

0

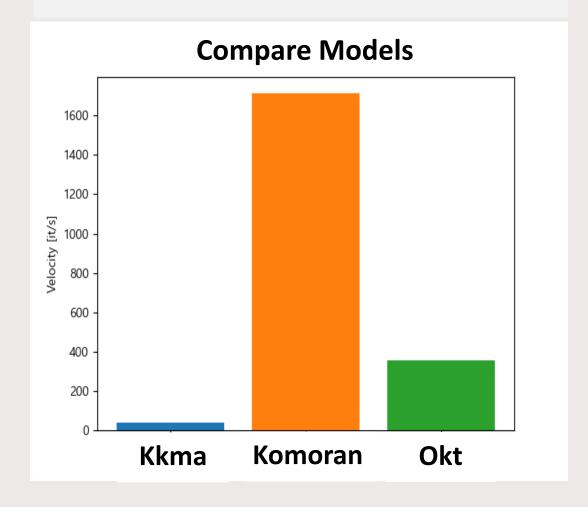
- 다양한 창업 정책과 창업 수요의 증가
- 창업 산업이 성장하면서
 사용자 수와 수요가 증가
 할 가능성이 있음

- 상권의 데이터가 부족하거나 품질이 낮을 경우 추천의 정확도가 떨어질 수 있음
- 정보 수집 과정에서 개인정보 보호 문제 발생 가능
- T '상권정보' 등 대규모 데이터를 가진 유사 서비스
 - 불확실한 경제 상황, 시장 변동 등의 외부 요인으로 시스템 운영 에 영향 미칠 가능성 있음



3 구현 및 시행착오 _{구현}

모델선택-형태소분석

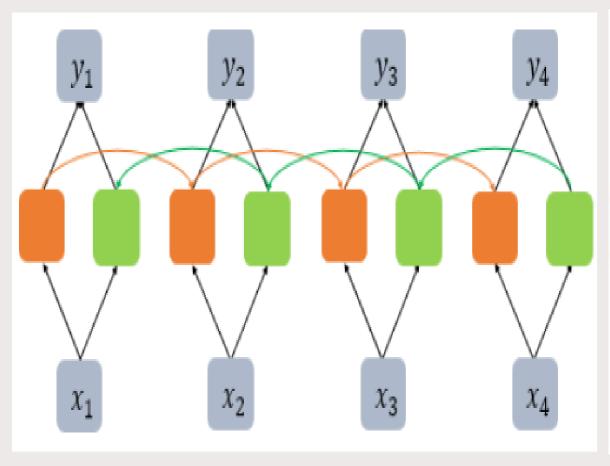


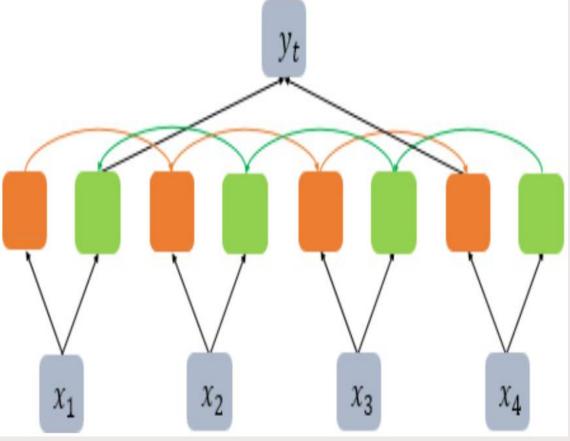
처리 결과 비교

"파이썬이 좋은감요 좋아함요 "

모델	결과
Kkma	파이/썬/이/좋/은/감/요/좋아하/ ㅁ /요
Komoran	파이썬/이/좋/은/감/요/좋아하/ ㅁ /요
Okt	파이썬/이/좋/은감/요/좋아함요

BiLSTM을 텍스트 평가에 사용





BiLSTM을 텍스트 평가에 사용

```
def sentiment_predict(new_sentence):
 print(new_sentence)
 new = new_sentence
 new_sentence = re.sub(r'[^¬-ㅎㅏ-ㅣ가-힣 ]','', new_sentence)
 new_sentence = okt.morphs(new_sentence, stem=True) # 토큰화
 new_sentence = [word for word in new_sentence if not word in stopwords] # 불용어 제거
 # 단어 등장 사전(n : 명사, a : 형용사)
 word_list = []
 for i,j in okt.pos(new):
                           # 명사 추출
   if i=='Noun':
     word_list.append(i)
   if j=='Adjective':
     word list.append(i) # 형용사 추출
  top_word = word_count(word_list) # 단어 빈도수 top 10
  encoded = tokenizer.texts_to_sequences([new_sentence]) # 정수 인코딩
 print(encoded)
 pad new = pad sequences(encoded. maxlen = maxlen) # 패팅
 score = np.round(float(loaded_model.predict(pad_new)).4) # 예측 score 소수점 넷째자리
  if(score > 0.5):
   print("{:.2f}% 확률로 긍정 리뷰입니다.₩n".format(np.round(score * 100,3)))
 else:
   print("{:.2f}% 확률로 부정 리뷰입니다.₩n".format(np.round((1 - score) * 100,因)))
  return score, top word
```

BiLSTM을 텍스트 분류에 사용

리뷰 내용	평점	감성 평가 점수
예술의 전당 맛집하면 10명의 7이상은 "백년옥" 인듯ㅎㅎ 부담없는 식사에 굿	5	0.9895
해물파전은 달걀함량이 많은 듯 그냥 그럼 18000원의 가치는 없음. 두부는 맛있음	4	0.366
이게 면이냐? 고무줄인줄 알았다	1	0.0725
난 못먹는데 매운걸좋아하는데 여기는 그냥심하게아픈맛인 데 근데 또가고싶은데??잉?	5	0.0817

3 구현 및 시행착오 구현

분석 데이터 리스트

- 🖹 상권상주인구.csv ╩
- ≧ 상권생활인구.csv ♣♣
- 🖹 상권영역.csv ╩
- 🖹 상권점포.csv ╩
- ≧ 상권직장인구.csv ♣♣
- ➡ 상권집객시설.csv ♣♣
- ≧ 상권추정매출.csv ♣♣

#	Column	Non-Null Count Dtype
0	서비스업종코드명	96282 non-null object
1	점포수	96282 non-null float64
2	유사업종점포수	96282 non-null float64
3	개업점포수	96282 non-null float64
4	폐업점포수	96282 non-null float64
5	집객시설수	96282 non-null float64

LightGBM을 이용한 상권 분석



회귀모델 평가지표

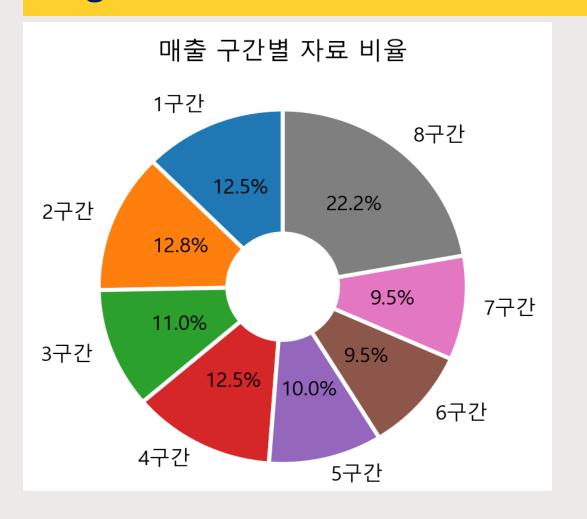
RMSE: 111.355

R2 score: 0.952

상권코드	시군구명	서비스업종명	월예상매출	점수
2110118	성동구	치킨전문점	1373.0	69.0
2110118	성동구	커피-음료	1067.0	58.0
2110118	성동구	분식전문점	1164.0	62.0
2110118	성동구	한식음식점	1288.0	66.0
2110118	성동구	제과점	1195.0	63.0
2110806	영등포구	커피-음료	831.0	50.0
2110798	영등포구	한식음식점	1223.0	64.0

3 구현 및 시행착오 _{구현}

LightGBM을 이용한 상권 분석

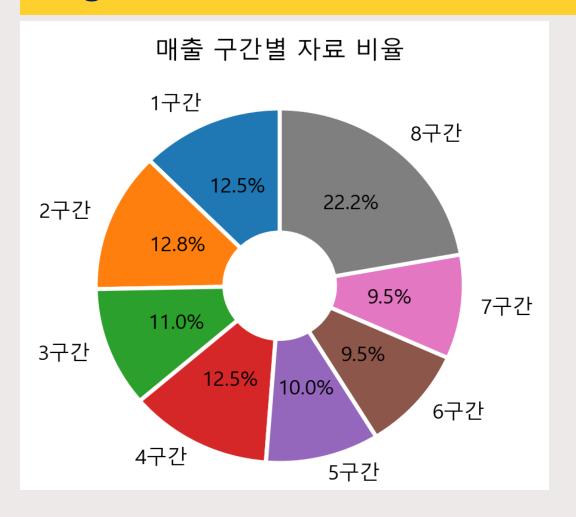


분류모델 평가

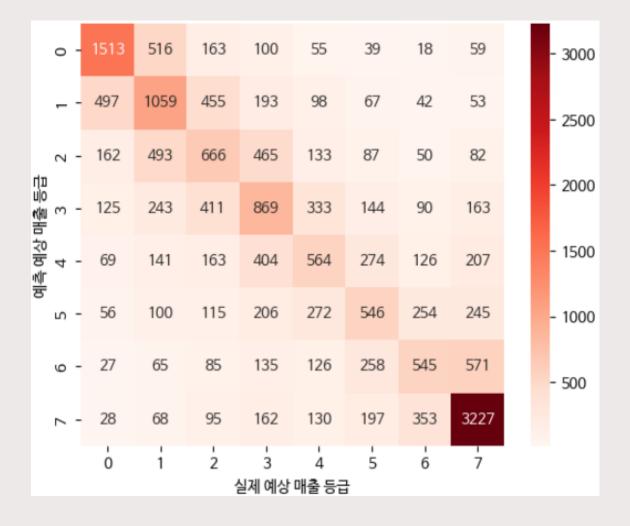
	precision	recall	f1-score
0.0	0.61	0.61	0.61
1.0	0.39	0.43	0.41
2.0	0.31	0.31	0.31
3.0	0.34	0.37	0.35
4.0	0.33	0.29	0.31
5.0	0.34	0.30	0.32
6.0	0.37	0.30	0.33
7.0	0.70	0.76	0.73

3 <mark>구현 및 시행착오</mark> _{구현}

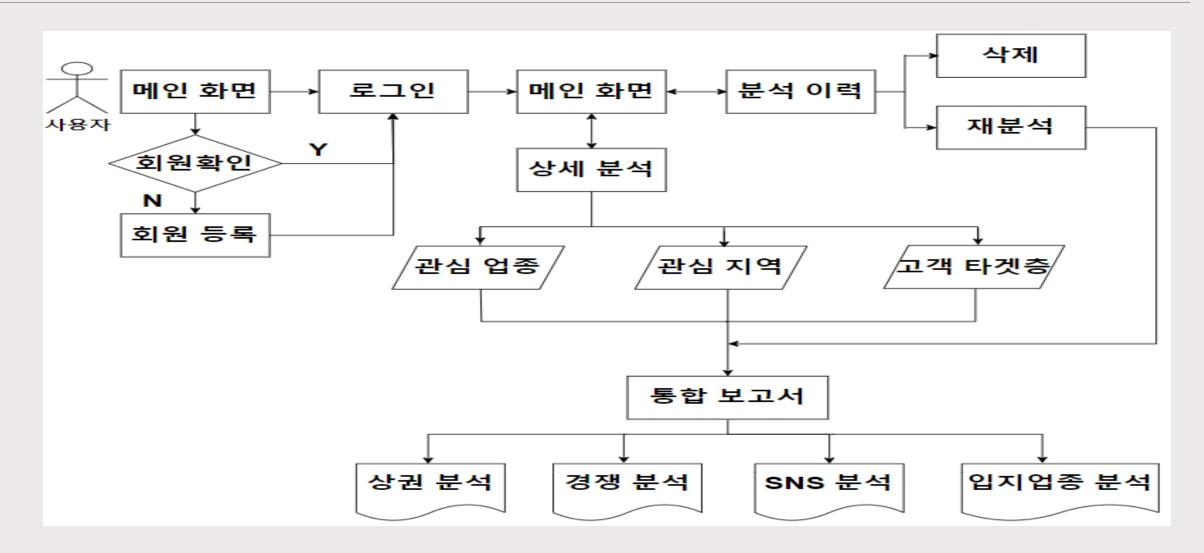
LightGBM을 이용한 상권 분석



분류모델 평가

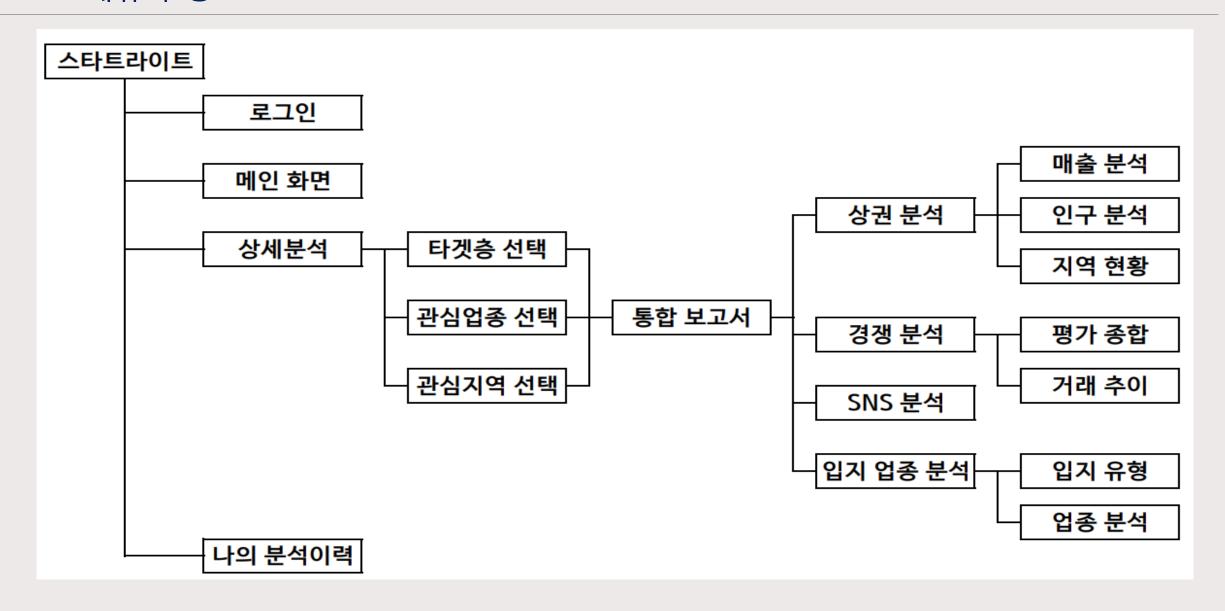


3구현및시행착오

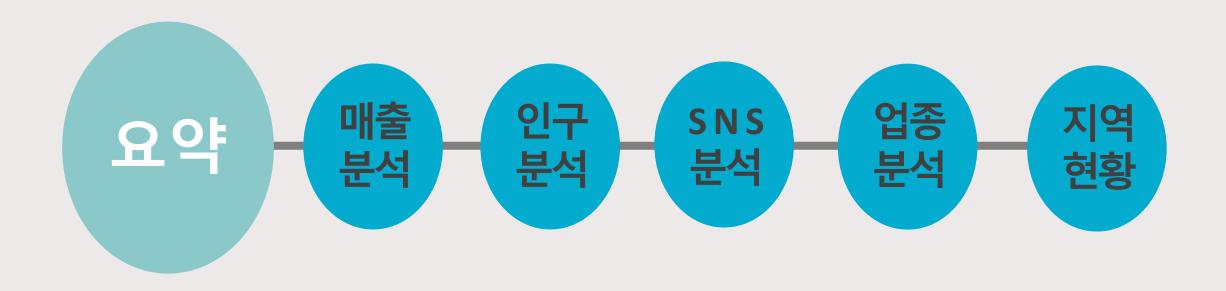


©	아이디	ID		varchar2(15)	
	이름	NAME		varchar2(20)	
	성별	GENDER		char(1)	
	나이	AGE		number(3)	
	연락처	PHONE		char(15)	+
	주소	ADDRESS		varchar2(300)	
	최근접속일	ACCESS_DATE		char(20)	
	회원탈퇴여부	DEL_ID	"Y","N"	char(1)	
	회원가입일	REGISTE_DATE		char(20)	
	비밀번호	PASSWORD		varchar2(300)	
	회원유형	MEMBER_TYPE		varchar2(10)	

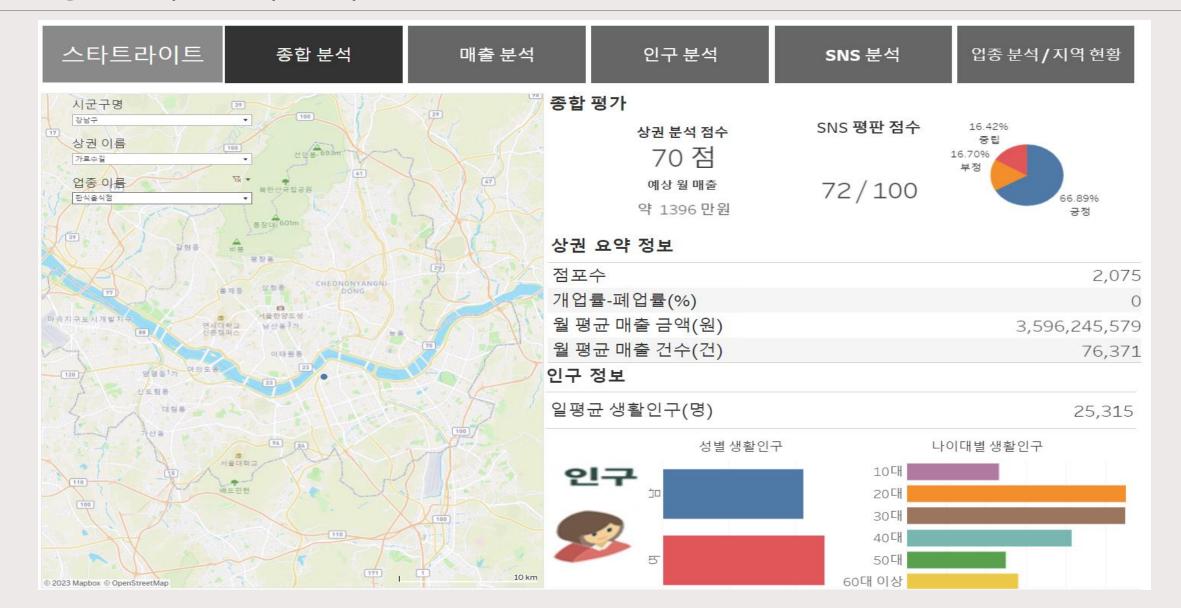
_								
뷴	분석	내	격					ANALYSIS
(⊙ ⊽	분	석번호		analysisCo	de	Domain	number(10)
п		분	석날짜		aDate			char(20)
		상	권분석보고/	Ч	reportCom	mertial	Domain	varchar(300)
×		경	맹분석보고/	Ч	reportCom	petition		varchar(300)
		SN	S분석보고시	4	reportSns			varchar(300)
		입	지업종분석!	보고서	reportLoca	ition		varchar(300)
(© 7	아	미디		ID			varchar2(15)
(⊙ ⊽	상	권번호		commertia	lCode		number(10)
					¥			
	Ş	상권	정보			C	OMMER'	ΓIAL
	(₽ ₹	상권번호	commo	ertialCode		number(10)
			관심업종	Catego	ory		char(20)	
			타겟층	Target		Domain	varchar(2	20)



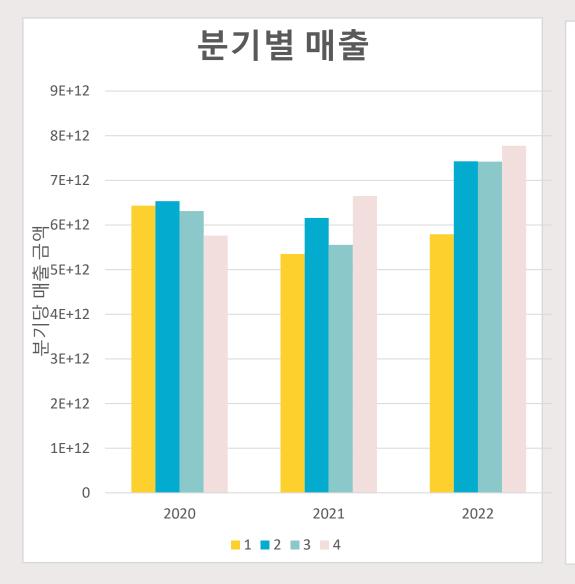


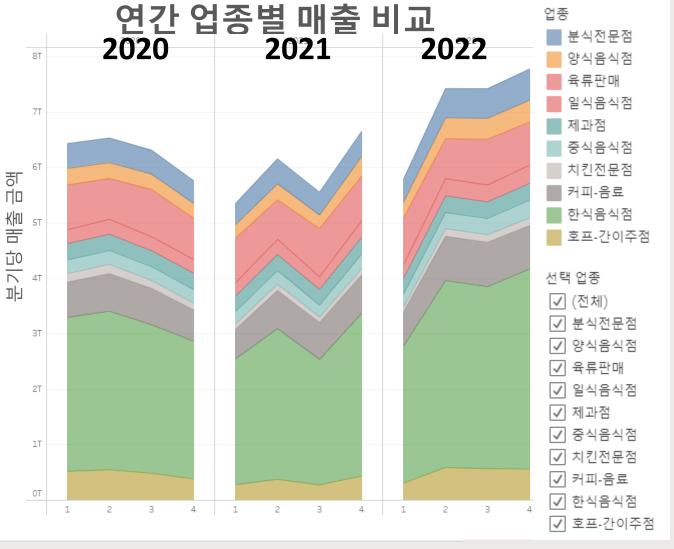


3 구현 및 시행착오 상권분석보고서 - 요약



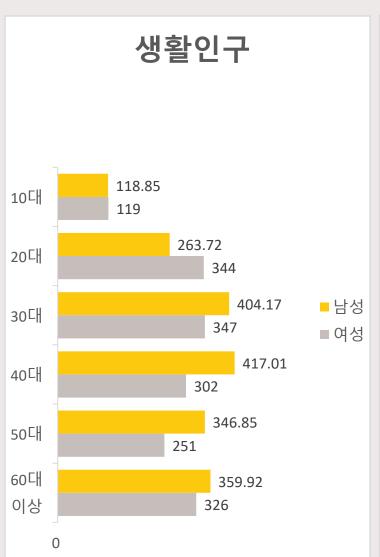
3 구현 및 시행착오 상권분석보고서-매출분석

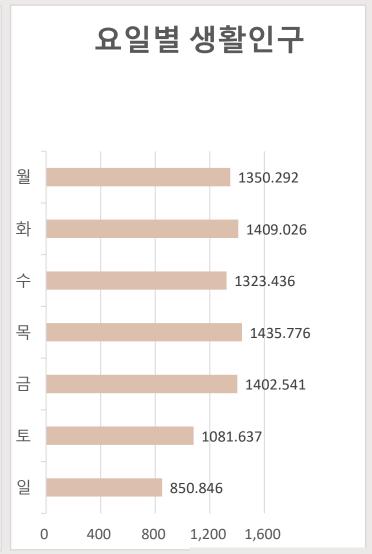




3 구현 및 시행착오 상권분석보고서-인구분석



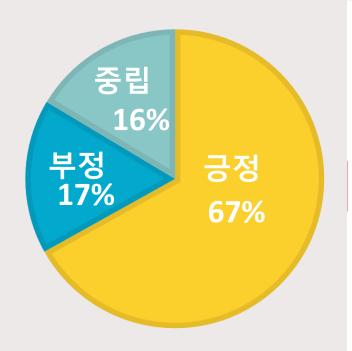




3 구현 및 시행착오 상권분석보고서-SNS분석

키-E 고리 간이주점 분식 아시안음식 양식 육류판매 음료 일식 제과점 중식 치킨 패스트푸드 **한식**

긍정 · 부정 · 중립



텍스트 마이닝

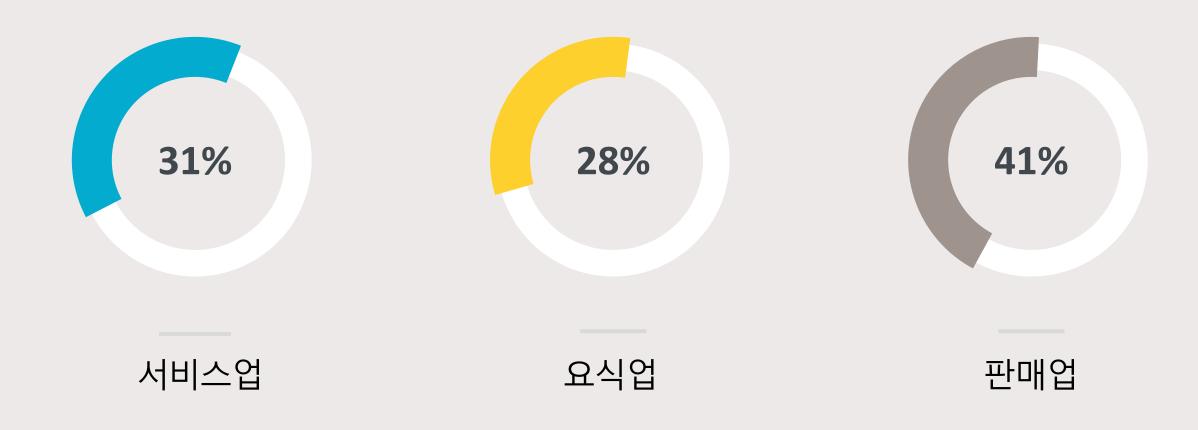
리뷰

제가 좋아하는 관자와 전복 너무 신선하니 맛있었어요

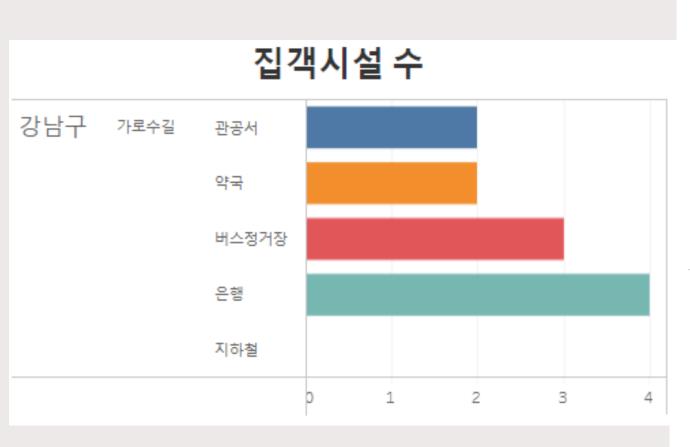
먹을 건 하나 없고 값만 비싼 집

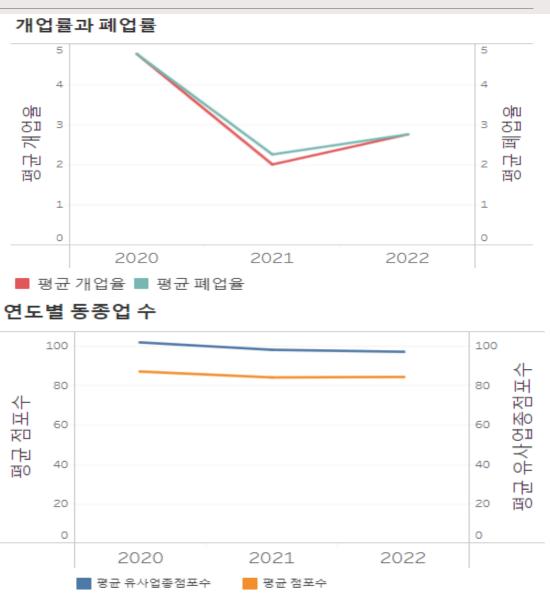
본관 꽉 차서 별관으로 옴

3 구현 및 시행착오 상권분석보고서 - 업종·지역 분석

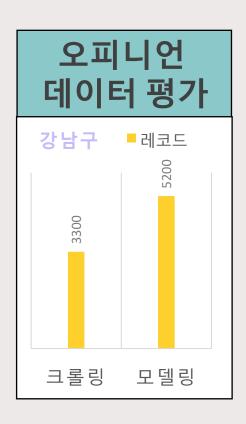


3 구현 및 시행착오 상권분석보고서-업종·지역 분석

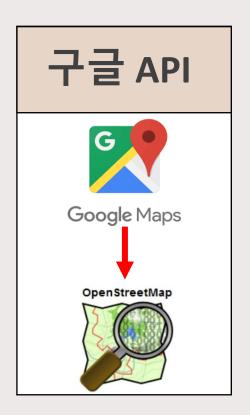




3구현및시행착오









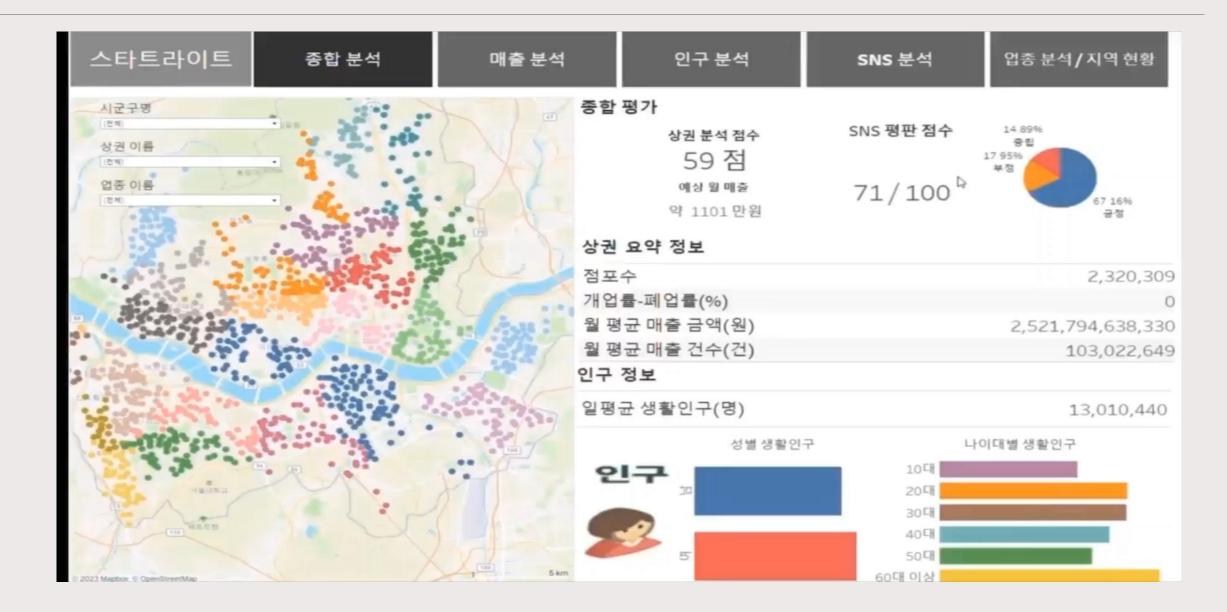
3 구현 및 시행착오 _{달성 목표}

리뷰 데이터 정확도

데이터 분석 유효성 확인 모델 학습 모피니언 마이닝 모델 학습 모피니언 데이터 평가 모데이터 수집

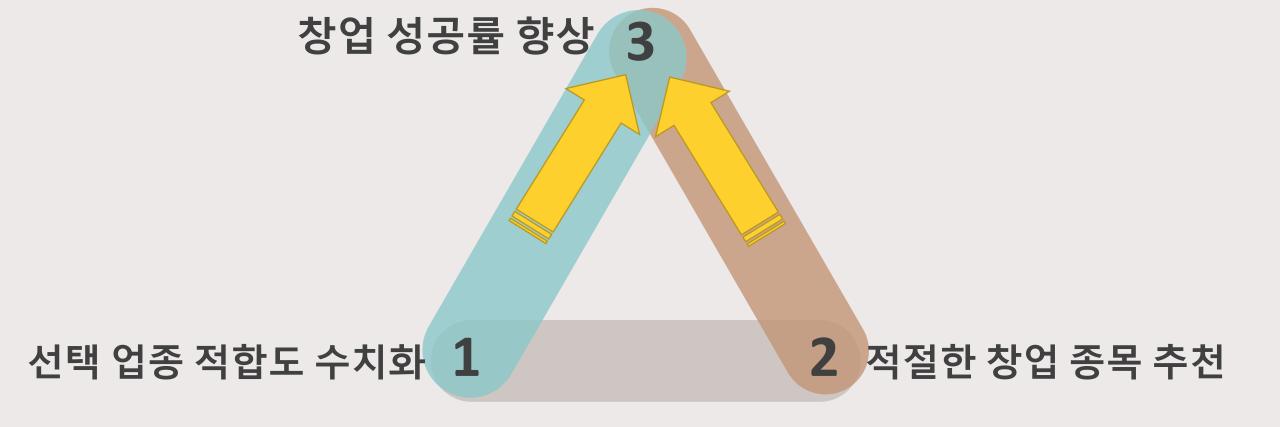
Part 4, 시연 영상

4시연영상

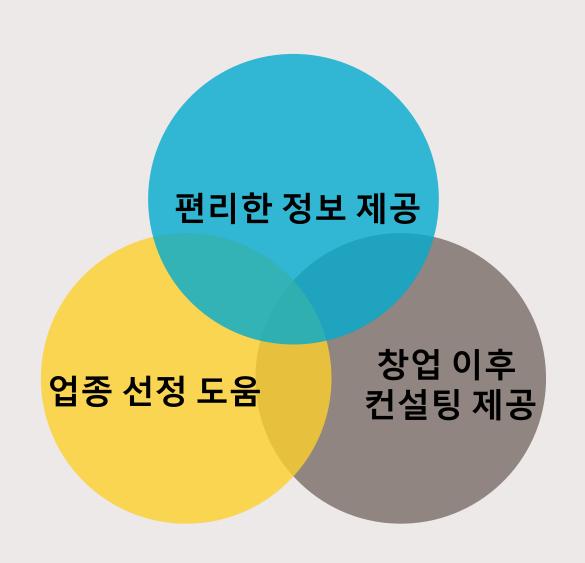


Part 5, 기대효과 & 활용방안

5 기대효과



5 활용 방안



5데이터출처

데이터명	데이터 출처	데이터내용					
직장인구, 상주인구		성별, 연령대					
생활인구		요일,연령대					
점포		상권코드명,유사업종수					
집객 시설	서울 열린 데이터 광장	집객시설수					
상권변화지표		상권변화지표명, 영업 개월 평균, 폐업 영업 개월 평균					
추정매출		서비스 업종					
리뷰 데이터	구글맵, 식신, 카카오맵	리뷰 데이터					
그외	빅데이터캠퍼스, 스마트치안 빅데이터 플랫폼, envbigdata, 공공데이터 포털, kt 통신빅데이터, 서울 연구데이터 서비스, kdx 한국데이터 거래소, 카드매출						

5계획일정

		추진 내용	4.11	4.12	4.13	4.14	4.17	4.18	4.19	4.20	4.21	4.24	4.25	4.26	4.27	4.28	5.1	5.2	5.3	5.4	5.5	5.8
Γ	기획	프로젝트 기획 발표																				
	기복	화면 설계 발표																				
		상권 분석용 빅데이터 수집																				
	분석	오피니언 데이터 수집																				
	E-7	데이터 전처리																				
		데이터 분석																				
	설계	모델링 및 데이터 학습																				
	구현	서비스 구현 및 시연																				
	발표	프로젝트 최종발표 준비																				
	24	프로젝트 발표																				

Q&A