

Python mini project

1조 김재훈 김진문 유효상 이예린

선호 여행 유형에 따른
국내 여행지 추천

목차

- 1 프로젝트 개요
- 2 데이터 전처리
- 3 관광소비 데이터 분석
- 4 관광 키워드 분석
- 5 GUI를 이용한 여행지 추천



Part 1

프로젝트 개요

주제 선정 & 데이터 소개

프로젝트 목표

국내 여행 지역별 트렌드 및 평균 지출액 분석을 통한 **국내 여행 가이드라인 제시**

◦ 사용 데이터 목록

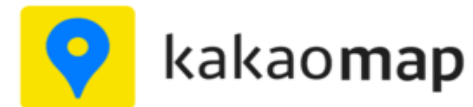
네이버 블로그 크롤링

네이버 검색광고 키워드 도구

한국관광 데이터랩 지역별 관광 지출액 데이터

카카오 맵 데이터 (GUI적용)

NAVER Developers



📋 국내 인기여행지 TOP5선정



📋 선정 이유

한국 관광 데이터랩 2022년 데이터 기준

검색건수 및 관광소비 TOP 3 + 관광중심지역 “제주”

Part 2

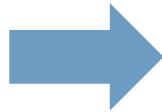
데이터 전처리

관광 소비 데이터 & 블로그 크롤링 데이터

전처리 1. 지역/중분류별 관광 소비 데이터

5개 지역의 2020년 1월~2023년 4월 소비데이터

	광역자치체	중분류	지출액	전년도 지출액	기간
0	강원도	관광총소비	98147222.0	95967578.0	2020-01
1	강원도	호텔	5005872.0	4202995.0	2020-01
2	강원도	콘도	9896940.0	9083866.0	2020-01
3	강원도	캠핑장/펜션	95516.0	62609.0	2020-01
4	강원도	기타숙박	4855888.0	5069995.0	2020-01
...
71	제주특별자치도	관광유원시설	877755.0	925673.0	2023-04
72	제주특별자치도	골프장	6549081.0	8147719.0	2023-04
73	제주특별자치도	기타레저	399040.0	449688.0	2023-04
74	제주특별자치도	문화서비스	1143212.0	1012993.0	2023-04
75	제주특별자치도	식음료	40691345.0	44580103.0	2023-04



3885 rows × 5 columns

분기/지역별 관광총소비금액 가공 후 표준화

	date	서울	경기	강원	부산	제주
20년 1분기	20-1	0.310958	0.082537	0.142471	0.239348	0.0
20년 2분기	20-2	0.313617	0.503944	0.341708	0.376353	0.131291
20년 3분기	20-3	0.222233	0.335854	0.568224	0.333959	0.431741
20년 4분기	20-4	0.132784	0.279908	0.247798	0.137114	0.411563
21년 1분기	21-1	0.0	0.0	0.0	0.0	0.151658
21년 2분기	21-2	0.315072	0.456364	0.397509	0.226808	0.70101
21년 3분기	21-3	0.058229	0.281314	0.603639	0.104429	0.50248
21년 4분기	21-4	0.479671	0.576495	0.438852	0.443207	0.740573
22년 1분기	22-1	0.330863	0.225612	0.235472	0.210573	0.571264
22년 2분기	22-2	0.980062	1.0	0.695165	0.925835	1.0
22년 3분기	22-3	0.893774	0.893673	1.0	0.962415	0.868711
22년 4분기	22-4	1.0	0.884441	0.581403	1.0	0.672229
23년 1분기	23-1	0.829663	0.552319	0.394206	0.796328	0.504932

관광 소비 데이터 중분류 항목

- 호텔
- 콘도
- 캠핑장/펜션
- 기타숙박

숙소

- 식음료

음식

- 골프장
- 관광기념품
- 레저용품쇼핑
- 대형쇼핑몰
- 관광유원시설
- 스키장

관광

전처리 2. 네이버 블로그 크롤링 & 결과 단어 빈도수 분석

* 2023 관광트렌드 전망 및 분석 보고서 참고

선호 여행 스타일 **테마 3개**를 선정하여 **지역별로** 검색

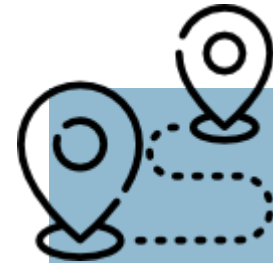
검색 결과 **상위 100개** 블로그 제목과 세부 내용 추출!



“숙소 추천”



“맛집 추천”



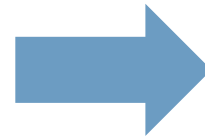
“액티비티 추천”

전처리 2. 네이버 블로그 크롤링 & 결과 단어 빈도수 분석

```
keywords = ['서울 숙소 추천', '서울 맛집 추천', '서울 액티비티 추천',
            '강원 숙소 추천', '강원 맛집 추천', '강원 액티비티 추천',
            '제주 숙소 추천', '제주 맛집 추천', '제주 액티비티 추천',
            '부산 숙소 추천', '부산 맛집 추천', '부산 액티비티 추천',
            '경기도 숙소 추천', '경기도 맛집 추천', '경기도 액티비티 추천']
```

크롤링 결과 단어 빈도수 분석

	서울숙소	서울맛집	서울관광	강원숙소	강원맛집	강원관광
0	서울 숙소 추천 포포 인츠 바이 쉐라톤 서울역 호텔	더현대서울 맛집 추천 호우섬 예약 후기	서울 프리다이빙 액티비티 남자 이색 취미 생활 운동 추천	아기랑 여행 강원정동진 숙소 추천 비치 크루즈 테라스	강원정선우정식당정선맛집추천	강원도액티비티놀이거리 강원쥬라기랜드 사육 풀잼
1	서울 을지로 숙소 추천 호텔국도 조식패키지	여의도 더현대서울 맛집 추천 호우섬 탐광카멜레온 피	서울여행 밀림에서 서울 요트 놀거리 한강 액티비티 추천	강원양양 미풍가양양 에어비앤비 강원도 독채 감성 숙소 추천	삼척맛집추천 명륜진 사갈비 강원삼척점 삼척 고기집 추천	강촌레일파크 쥘라인 강원 공기 액티비티 즐기기 추천
2	서울 근교 경기도 이천 에어비앤비 숙소 추천 알이 하우스	서울 맛집 추천 혼밥은 좋은 마요유 라탕	서울 이색데이트 실내 액티비티 추천 스케이트파크 하남 오버헤드	강원속초시 속초 주차장체크인어 메너티 숙소추천아고다	속초 강원 속초 대포항대게 여러가지 많이 주는 동맛집 추천	아이랑 강원쥬라기랜드 즐기고 옴 강원도 액티비티놀이거리



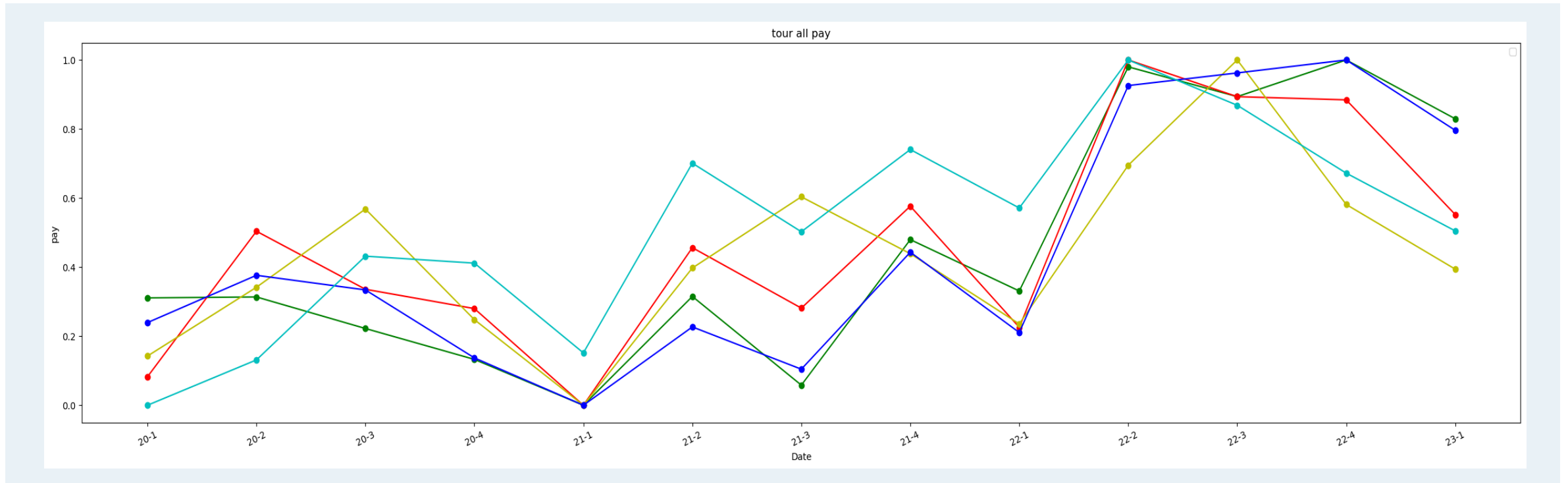
	tour	word	count
0	('액티비티', 125)	액티비티	125
1	('추천', 73)	추천	73
2	('강원', 70)	강원	70
3	('강원도', 59)	강원도	59
4	('여행', 38)	여행	38
...
1555	('죽도해변은', 1)	죽도해변은	1
1556	('이래로', 1)	이래로	1
1557	('수많은', 1)	수많은	1
1558	('스쿨들이', 1)	스쿨들이	1
1559	('잡으면서', 1)	잡으면서	1

Part 3

관광 소비 데이터 분석

관광 소비 경향 예측

관광총소비는 꾸준히 증가 추세를 보임



2020년

2021년

2022년

2023년 1분기

관광 소비 데이터 분석

	광역지자체	중분류	지출액	전년도 지출액	기간
0	강원도	관광총소비	98147222.0	95967578.0	2020-01
1	강원도	호텔	5005872.0	4202995.0	2020-01
2	강원도	콘도	9896940.0	9083866.0	2020-01
3	강원도	캠핑장/펜션	95516.0	62609.0	2020-01
4	강원도	기타숙박	4855888.0	5069995.0	2020-01
...
71	제주특별자치도	관광유원시설	877755.0	925673.0	2023-04
72	제주특별자치도	골프장	6549081.0	8147719.0	2023-04
73	제주특별자치도	기타레저	399040.0	449688.0	2023-04
74	제주특별자치도	문화서비스	1143212.0	1012993.0	2023-04
75	제주특별자치도	식음료	40691345.0	44580103.0	2023-04

3885 rows × 5 columns



ARIMA 통계 모델을 이용해 예측!

2020년 1분기

~

2023년 1분기



2023년 하반기

ARIMA model

(Auto Regressive Integrated Moving Average Model)

시계열을 예측하는 접근 방법

AR(Auto regression) + MA(Moving Average)를 합친 모형으로 시계열 데이터의 정상성을 가정

정상성이란 평균과 분산이 시간의 흐름에 따라 변하지 않음을 의미

최적의 p, D, Q값 찾기

P: AR모형의 차수 D: 시계열 함수의 미분횟수 Q: MR모형의 차수

```
p = d = q = range(0,5)
pdq = list(itertools.product(p,d,q))

maf=[]
for param in pdq:
    try:
        model_arima = SARIMAX(s2.food.values, order=param)
        model_arima_fit = model_arima.fit()
        print(param,model_arima_fit.aic)
        maf.append(model_arima_fit.aic)
    except:
        continue
min(maf)
```

```
(0, 0, 0) 437.3161655068327
(0, 0, 1) 390.1504453908173
(0, 0, 2) 347.1107278171959
(0, 0, 3) 328.68943720946277
(0, 0, 4) 321.8092143145106
(0, 1, 0) 158.9176592700121
(0, 1, 1) 153.98473297666254
(0, 1, 2) 154.34549027153463
(0, 1, 3) 155.1559048513624
(0, 1, 4) 154.10658115787277
(0, 2, 0) 192.33611686834638
(0, 2, 1) 161.54349937149686
(0, 2, 2) 157.3463442130121
(0, 2, 3) 155.66580974989668
(0, 2, 4) 156.4518190316177
(0, 3, 0) 231.67959415549154
(0, 3, 1) 193.94905850700027
```



가장 낮은 값을 갖는 쌍 선택!

•
•
•

“서울 숙박”의 사용금액을 예측한 경우

최적의 p,d,q 조합으로 데이터를 ARIMA모델에 적용

```
model = SARIMAX(s1.stay.values, order=(2,1,3))
model_fit = model.fit(trend='c', full_output=True, disp=True)
print(model_fit.summary())
```

```
forecast_data = model_fit.forecast(steps=4)
bitcoin_test_df = s2
forecast_data
```

학습 데이터셋으로부터 4개월 이후를 예측

SARIMAX Results

```
=====
Dep. Variable:          y      No. Observations:          40
Model:                SARIMAX(2, 1, 3)  Log Likelihood          -89.517
Date:                 Wed, 07 Jun 2023  AIC              191.034
Time:                 09:51:24    BIC              201.015
Sample:                0      HQIC              194.615
                        - 40
Covariance Type:      opg
```

p-value <0.05

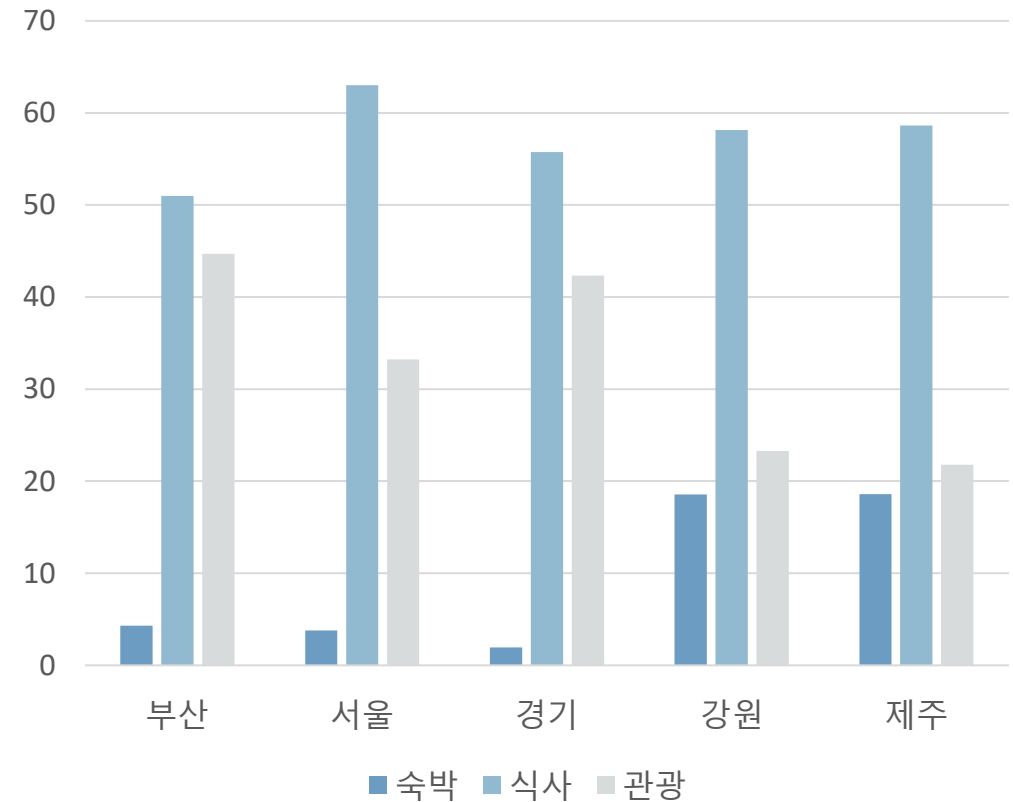
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.9862	0.052	19.069	0.000	0.885	1.088
ar.L2	-0.9852	0.025	-40.096	0.000	-1.033	-0.937
ma.L1	-1.6089	0.171	-9.425	0.000	-1.943	-1.274
ma.L2	1.4021	0.307	4.572	0.000	0.801	2.003
ma.L3	-0.6232	0.201	-3.105	0.002	-1.017	-0.230
sigma2	5.1224	1.567	3.268	0.001	2.050	8.194

```
=====
Ljung-Box (L1) (Q):          0.54  Jarque-Bera (JB):          1.09
Prob(Q):                  0.46  Prob(JB):              0.58
Heteroskedasticity (H):    0.74  Skew:                  -0.19
Prob(H) (two-sided):      0.59  Kurtosis:              2.28
=====
```

2023년 8월 소비 예측 결과(%)

(%)	숙박	식사	관광
부산	4.315	50.995	44.691
서울	3.774	63.019	33.207
경기	1.930	55.748	42.322
강원	18.568	58.138	23.295
제주	18.596	58.615	21.787

23년 8월 소비예측 결과 그래프



Part 4

관광 키워드 분석

블로그 크롤링 & 네이버 검색광고 도구

관 광 키워드 분석

	연관키워드	click_cnt
142	홍대맛집	1464.2
136	다이어트식단	1216.5
4	광화문맛집	1116.7
5	서울가볼만한곳	1064.6
214	도시락배달	935.5
127	찜질방	911.6
20	인사동맛집	887
186	강남역맛집	877.8
15	명동맛집	846.2
13	서울역맛집	823.3
270	국내기차여행	745.1
259	한정식맛집	738.6
131	여의도맛집	737.6
123	서울전시회	725.3
158	잠실맛집	698.3
218	왕십리맛집	694.2
154	국내여행	686.4
3	종로맛집	654.4
8	강남맛집	646.4
258	영등포맛집	634.1
1	서울맛집	633.6
159	국내여행사	632.2
128	냉면	627.3



키워드 광고 검색 광고 도구

(단어,빈도수)

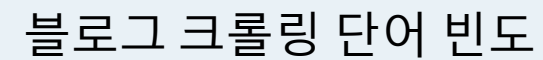
('부산', 202)
 ('숙소', 122)
 ('추천', 116)
 ('호텔', 64)
 ('부산숙소추천', 60)
 ('부산숙소', 36)
 ('해운대', 34)
 ('부산여행', 30)
 ('오션뷰', 23)
 ('광안리', 21)
 ('영도', 19)
 ('기장', 17)
 ('여행', 16)
 ('남포동', 16)
 ('펜션', 15)
 ('호캉스', 12)
 ('엘시티', 12)
 ('수', 12)
 ('부산호텔추천', 12)
 ('스테이화가', 11)

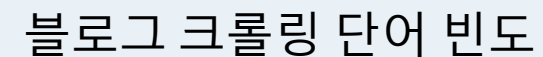
크롤링 단어 빈도 분석

#Word Cloud



©Saebyeol Yu. Saebyeol's PowerPoint







키워드 광고 검색 Top 50



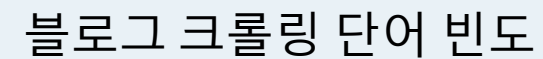
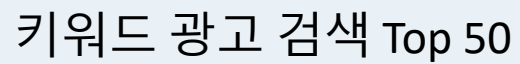
블로그 크롤링 단어 빈도



키워드 광고 검색 Top 50




블로그 크롤링 단어 빈도



Part 5

데이터 활용 WITH GUI



**블로그 크롤링
& 빈도분석**

**국내 여행지 선택
가이드라인 제시**

**관광 소비경향
예측 결과**

활용 예시 1



나는 여행할 때 숙소3, 음식 2, 관광 5 비율 정도로 중요하게 생각해

X 3 X 2 X 5

(%)	숙박	식사	관광
부산	4.315	50.995	44.691
서울	3.774	63.019	33.207
경기	1.930	55.748	42.322
강원	18.568	58.138	23.295
제주	18.596	58.615	21.787

숙박 중요도를 입력하시오(숙박, 음식, 관광 총합은 10) : 3
 음식 중요도를 입력하시오(숙박, 음식, 관광 총합은 10) : 2
 관광 중요도를 입력하시오(숙박, 음식, 관광 총합은 10) : 5

	지역	숙박	음식	관광	계산값
0	부산	4.315	50.995	44.691	338.390
1	경기	1.930	55.748	42.322	328.896
4	서울	3.774	63.019	33.207	303.395
2	강원	18.567	58.138	23.295	288.452
3	제주	19.596	58.615	21.787	284.953

→ 부산 여행 추천 !



활용 예시 2



나는 여행할 때 숙소5, 음식 4, 관광 1 비율 정도로 중요하게 생각해

X 5 X 4 X 1

(%)	숙박	식사	관광
부산	4.315	50.995	44.691
서울	3.774	63.019	33.207
경기	1.930	55.748	42.322
강원	18.568	58.138	23.295
제주	18.596	58.615	21.787

숙박 중요도를 입력하시오(숙박, 음식, 관광 총합은 10) : 5
 음식 중요도를 입력하시오(숙박, 음식, 관광 총합은 10) : 4
 관광 중요도를 입력하시오(숙박, 음식, 관광 총합은 10) : 1

	지역	숙박	음식	관광	계산값
3	제주	19.596	58.615	21.787	354.227
2	강원	18.567	58.138	23.295	348.682
4	서울	3.774	63.019	33.207	304.153
1	경기	1.930	55.748	42.322	274.964
0	부산	4.315	50.995	44.691	270.246

→ 제주도 여행 추천 !





나는 여행할 때 음식이 가장 중요해! 숙소나 관광은 상관없어

(%)	숙박	식사	관광
부산	4.315	50.995	44.691
서울	3.774	63.019	33.207
경기	1.930	55.748	42.322
강원	18.568	58.138	23.295
제주	18.596	58.615	21.787

	지역	숙박	음식	관광	계산값
4	서울	3.774	63.019	33.207	407.5475
3	제주	19.596	58.615	21.787	396.5325
2	강원	18.567	58.138	23.295	395.3450
1	경기	1.930	55.748	42.322	389.3700
0	부산	4.315	50.995	44.691	377.4900

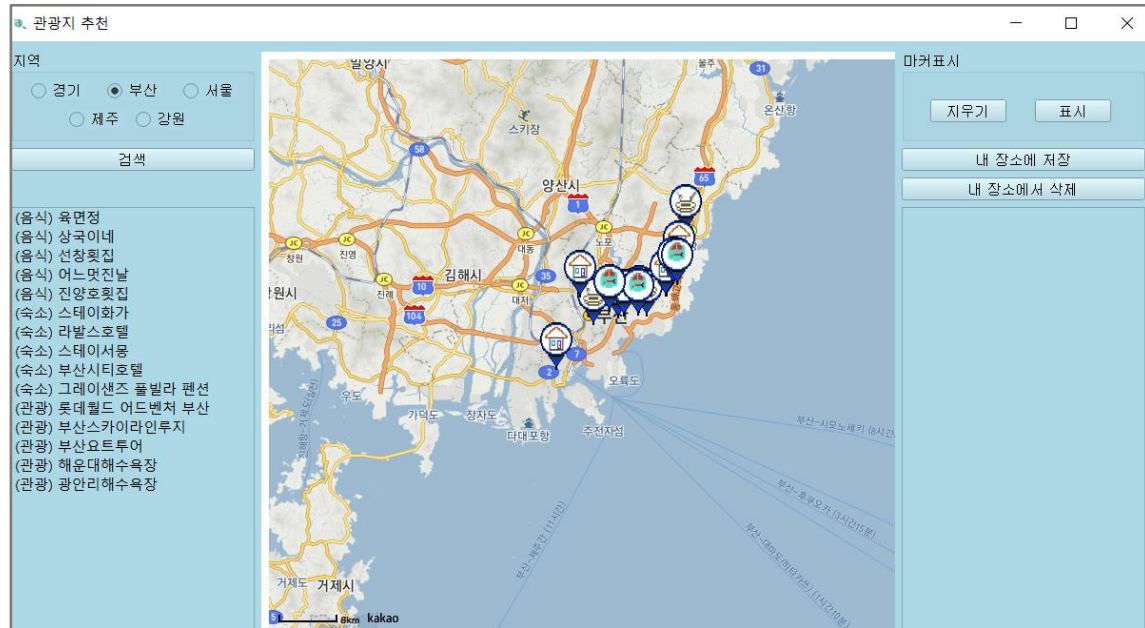
	지역	숙박	음식	관광	계산값
1	경기	1.930	55.748	42.322	445.9302
4	서울	3.774	63.019	33.207	442.7534
0	부산	4.315	50.995	44.691	436.1628
2	강원	18.567	58.138	23.295	391.9841
3	제주	19.596	58.615	21.787	388.7706



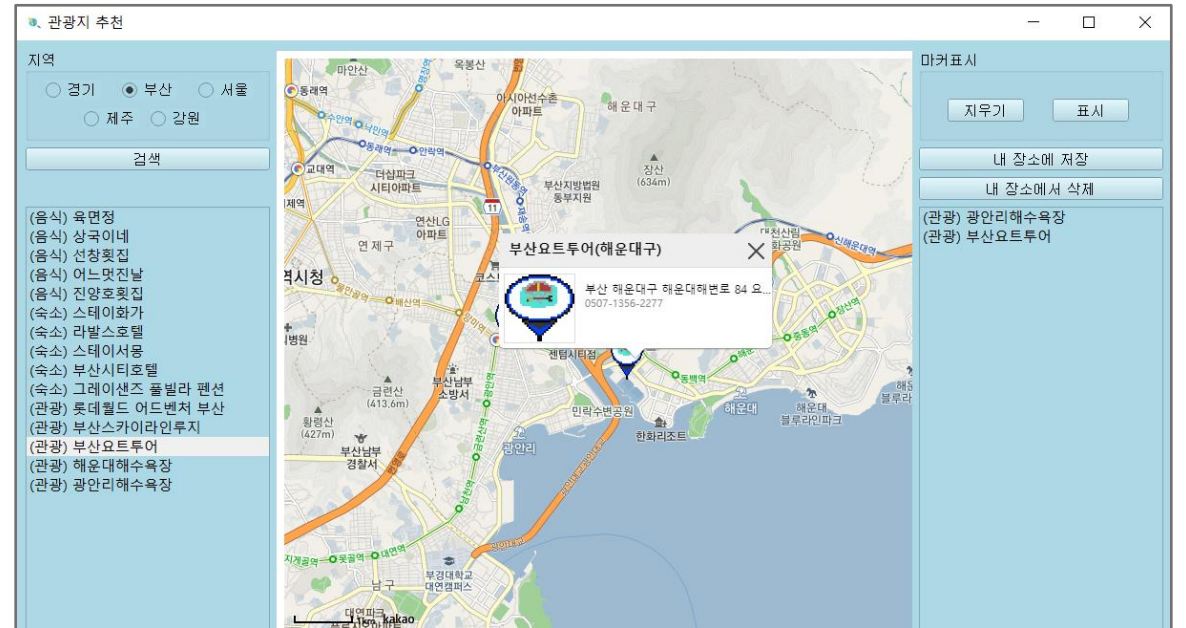
음식이 최우선 순위일 때
두가지 지역 추천 가능



지역별 인기 장소 검색



원하는 장소 추가/삭제 기능



Q&A



THANK YOU