

# 지역 관광 활성화를 위한 관광지 개발 방향 제안

9기 김현동

10기 박태정

10기 연제명



# 목차

개요

관광지  
요소에  
따른 분석

관광지  
유형에  
따른 분석

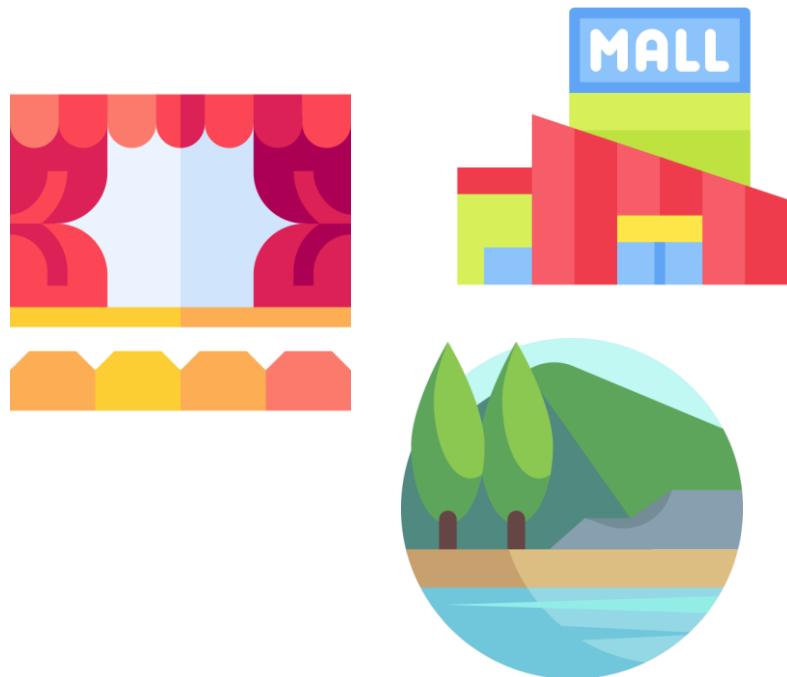
결론

느낀 점 및  
개선 방향

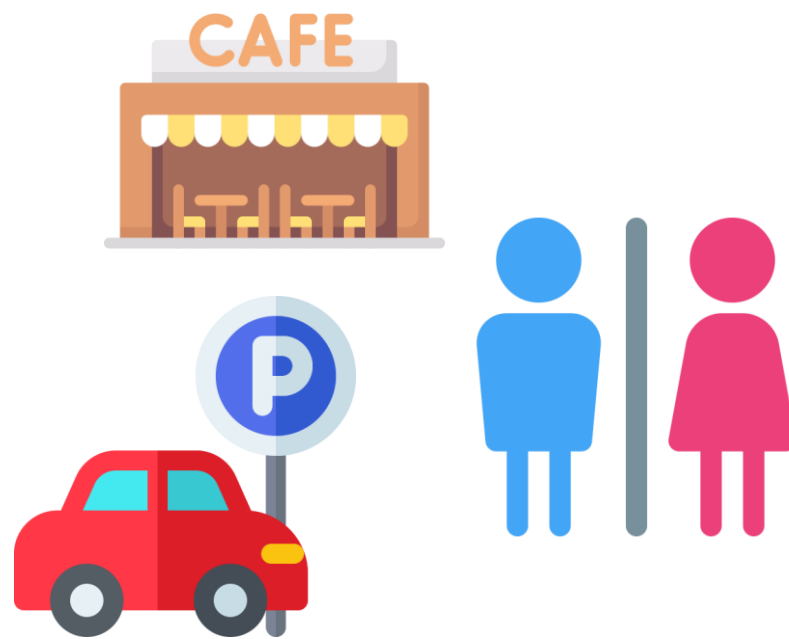


# 개요

## 관광지의 유형



## 관광지의 구성 요소





# 개요

## <Research Questions>

1. “어떤 요소”를 갖춘 관광지에 사람들이 많이 방문할까?
2. “어떤 유형”의 관광지에 사람들이 많이 방문할까?

### <목표>

다양한 통계적 검정 방법을 활용하여 각 질문에 대한 답을 찾아보고, box plot과 word cloud 등을 활용해 해당 질문에 대한 답들을 시각화한다.

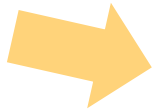


## 관광지 요소에 따른 분석

### <데이터와 방법론>

#### <데이터>

- 전국관광지정보표준데이터
- 주요관광지점 입장객 데이터
- 인기관광지\_외지인 데이터



- 총 245개 관광지 (전국관광지정보데이터 & 주요관광지 입장객 데이터 병합)
  - 가설 검정에 사용
- 총 63개 관광지 (전국관광지정보데이터 & 인기관광지 데이터 병합)
  - 워드클라우드(인기 관광지 키워드)에 사용

#### <방법론>

- Pearson 상관계수
- Mann-Whitney U검정
- T검정
- Word Cloud
- Box Plot



## <전국관광지정보표준데이터>

-



## 관광지 요소에 따른 분석

### <데이터 전처리>

#### <주요관광지점 입장객 데이터>

1. '내국인' 방문자 수 외 삭제
2. 컬럼명 변경 및 정리
3. '시도명' 값 일치
4. 문자형 데이터 정수형으로 변환
5. 결측치 제거

#### <인기관광지\_외지인 데이터>

1. '시도명' & '시군구명' 컬럼으로 변환
2. '시도명' 값 일치시키기
3. 불필요한 컬럼 삭제
4. "인기관광지\_외지인 데이터"에 "전국관광지정보표준데이터"를 병합
5. '관광지명', '시도명', '시군구명'이 모두 일치하는 행들을 추출하여 병합
  - 총 63개 관광지



## 관광지 요소에 따른 분석

### <분석 과정>

#### <Pearson 상관계수>

1. 가설 설정
  - $H_0$  : 관광지의 편의시설 유무는 총 방문자 수와 상관관계가 없다.
  - $H_1$  : 관광지의 편의시설 유무는 총 방문자 수와 상관관계가 있다.
2. 해당 편의시설과 총 방문자 수 컬럼 선택
3. 상관계수와 p-value 계산
4. 가설 검정 수행

#### <T검정>

1. 가설 설정
  - $H_0$  : 편의시설의 유무에 따른 총 방문자 수에는 차이가 없다
  - $H_1$  : 편의시설의 유무에 따른 총 방문자 수에는 차이가 있다.
2. 각 그룹의 총 방문자 수 분포 확인
3. 표본의 개수 확인
4. 등분산 검정 (Levene 검정)
5. T검정 수행
6. 가설검정 수행





## 관광지 요소에 따른 분석

### <분석 과정>

#### <Mann-Whitney U검정>

1. 가설 설정
  - $H_0$ : 편의시설의 유무에 따른 총 방문자 수에는 차이가 없다
  - $H_1$ : 편의시설의 유무에 따른 총 방문자 수에는 차이가 있다.
2. 각 그룹의 총 방문자 수 분포 확인
3. 표본의 개수 확인
4. 등분산 검정 (Levene 검정)
5. Mann-Whitney U검정 수행
6. 가설검정 수행



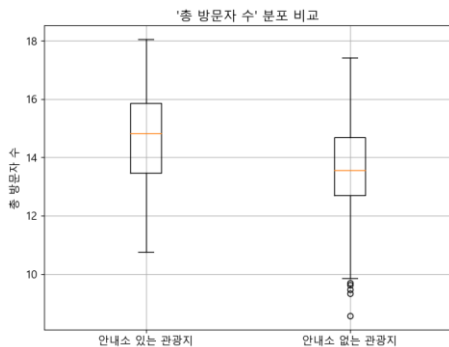


## 관광지 요소에 따른 분석

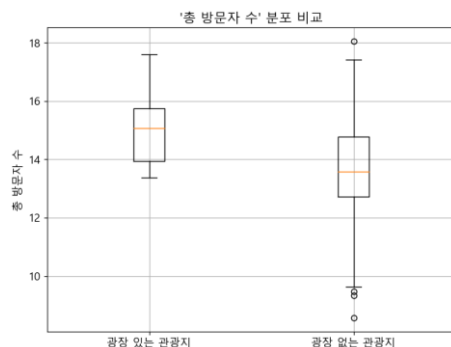
### <EDA 결과>

#### <유의미한 차이를 보이는 요소>

- 안내소
- 광장



<안내소>



<광장>

#### <유의미한 차이가 없는 요소>

- 화장실
- 주차장
- 관리사무소
- 샤워장
- 수유실
- 휴게/쉼터
- 공원
- 카페/커피
- 숙박시설
- 운동 및 오락시설
- 휴양 및 문화시설
- 접객시설
- 지원시설

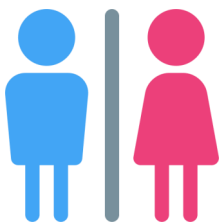


## 관광지 요소에 따른 분석

### <EDA 결과>

#### <화장실>

- Pearson 상관계수
  - 상관계수: 0.066
- T검정
  - 표본 수: 197/48
  - 등분산
  - p-value: 0.305



#### <주차장>

- Pearson 상관계수
  - 상관계수: 0.051
- T검정
  - 표본 수: 145/100
  - 등분산
  - p-value: 0.424



#### <관리사무소>

- Pearson 상관계수
  - 상관계수: 0.073
- T검정
  - 표본 수: 72/173
  - 등분산
  - p-value: 0.252



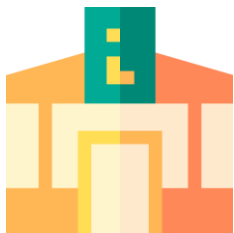


## 관광지 요소에 따른 분석

### <EDA 결과>

#### <안내소>

- Pearson 상관계수
  - 상관계수: 0.318
- T검정
  - 표본 수: 32/213
  - 이분산
  - p-value: 3.60e-07



#### <샤워장>

- Pearson 상관계수
  - 상관계수: -0.017
- Mann-Whitney U검정
  - 표본 수: 14/231
  - 등분산
  - p-value: 0.476



#### <광장>

- Pearson 상관계수
  - 상관계수: 0.165
- Mann-Whitney U검정
  - 표본 수: 9/236
  - 이분산
  - p-value: 0.007





## 관광지 요소에 따른 분석

### <EDA 결과>

#### <수유실>

- Pearson 상관계수
  - 상관계수: -0.000
- Mann-Whitney U검정
  - 표본 수: 10/235
  - 등분산
  - p-value: 0.795



#### <휴게/침터>

- Pearson 상관계수
  - 상관계수: -0.052
- Mann-Whitney U검정
  - 표본 수: 10/235
  - 등분산
  - p-value: 0.602



#### <공원>

- Pearson 상관계수
  - 상관계수: -0.045
- Mann-Whitney U검정
  - 표본 수: 7/238
  - 등분산
  - P-value: 0.987





## 관광지 요소에 따른 분석

### <EDA 결과>

#### <카페/커피>

- Pearson 상관계수
  - 상관계수: 0.002
- Mann-Whitney U검정
  - 표본 수: 4/241
  - 등분산
  - p-value: 0.430



#### <숙박시설>

- Pearson 상관계수
  - 상관계수: -0.037
- T검정
  - 표본 수: 103/142
  - 등분산
  - p-value: 0.565



#### <운동 및 오락시설>

- Pearson 상관계수
  - 상관계수: 0.038
- T검정
  - 표본 수: 92/153
  - 등분산
  - P-value: 0.550



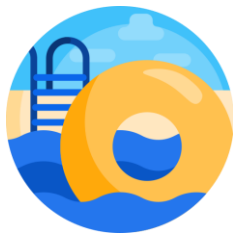


## 관광지 요소에 따른 분석

### <EDA 결과>

#### <휴양 및 문화시설>

- Pearson 상관계수
  - 상관계수: -0.001
- T검정
  - 표본 수: 122/123
  - 등분산
  - p-value: 0.993



#### <접객시설>

- Pearson 상관계수
  - 상관계수: -0.047
- T검정
  - 표본 수: 74/171
  - 등분산
  - p-value: 0.468



#### <지원시설>

- Pearson 상관계수
  - 상관계수: -0.005
- T검정
  - 표본 수: 40/205
  - 등분산
  - P-value: 0.940





## 관광지 요소에 따른 분석

### <EDA 결과>

#### <인기 관광지 소개 키워드>



#### <전체 관광지 소개 키워드>





## <데이터와 방법론>

### <데이터>

- 행정안전부\_지역별(법정동) 성별 연령별 주민등록 인구수 6월(7.11)
  - 인기관광지\_외지인 데이터
- 

### <방법론>

- Pearson 상관계수
- T검정
- Box Plot

## 관광지 유형에 따른 분석

### <데이터와 방법론>

#### <데이터>

- 행정안전부\_지역별(법정동) 성별 연령별 주민등록 인구수 6월(7.11)
- 인기관광지\_외지인 데이터

#### <방법론>

- Pearson 상관계수
- T검정
- Box Plot

외지인 검색 vs 총 방문자 수

- T검정 결과 두 분포의 평균이 같다는 것 확인
- p-value: 0.136

## 관광지 유형에 따른 분석

### <데이터 전처리>

#### <인기 관광지 데이터>

1. 관광지로 분류하기 부적절한 "관광지명" 제거
  - 대형마트, 백화점, 쇼핑몰, 면세점, 레저스포츠시설(야구장), 영화관, 자동차극장, 교통시설
2. 표본수가 너무 적은 유형 삭제
  - 농/산/어촌체험, 캠핑, 수상레저스포츠, 기타숙박, 모텔, 펜션/민박
3. 인구수 데이터와 통합, 2가지 데이터로 분석
  - tour\_result
  - general\_tour\_result

#### <관광지 유형>

- |                     |                |
|---------------------|----------------|
| • 호텔 (143)          | • 콘도미니엄 (64)   |
| • 육상레저스포츠 (139)     | • 전시시설 (74)    |
| • 자연경관(하천/해양) (117) | • 테마공원 (60)    |
| • 시장 (113)          | • 종교성지 (56)    |
| • 도시공원 (85)         | • 자연관광지 (43)   |
| • 기타관광 (73)         | • 역사유적지 (41)   |
| • 복합관광지 (65)        | • 기타문화관광지 (29) |

## 관광지 유형에 따른 분석

### <분석 과정>

#### <Pearson 상관계수>

1. 가설 설정
  - $H_0$ : 관광지의 유형은 외지인 검색 수와 상관관계가 없다.
  - $H_1$ : 관광지의 유형은 외지인 검색 수와 상관관계가 있다.
2. 상관계수와 p-value 계산
3. 가설 검정 수행

#### <T검정>

1. 가설 설정
  - $H_0$ : 관광지 유형에 따라 외지인 검색 수에 차이가 없다
  - $H_1$ : 관광지 유형에 따라 외지인 검색 수에 차이가 있다.
2. 각 그룹의 총 외지인 검색 수 분포 확인
3. 표본의 개수 확인
4. 등분산 검정 (Levene 검정)
5. T검정 수행
6. 가설검정 수행

## 관광지 유형에 따른 분석

### <EDA 결과>

#### <관광지 유형과 외지인 검색의 상관관계>

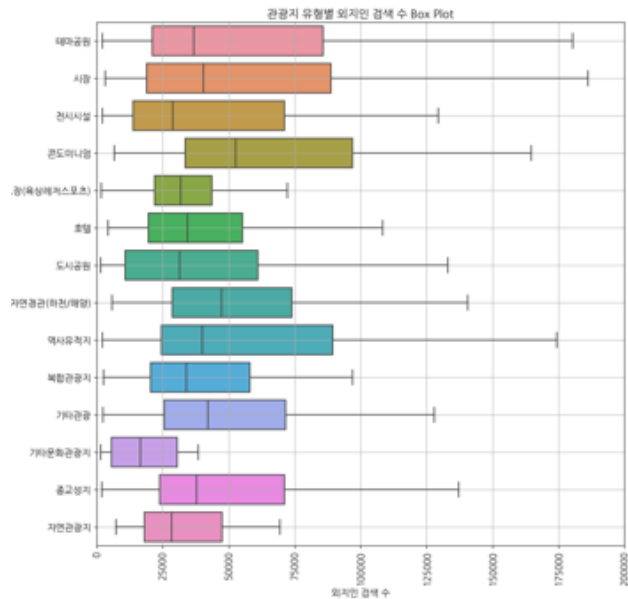
- 상관계수가 0.4 이상인 관광지 유형
  - 콘도미니엄
    - 상관계수: 0.640
  - 호텔
    - 상관계수: 0.519
  - 복합관광지
    - 상관계수: 0.451
  - 자연경관(하천/해양)
    - 상관계수: 0.427



# 관광지 유형에 따른 분석

## <EDA 결과>

### <관광지 유형에 따른 외지인 검색>



- 시장, 콘도미니엄, 자연경관(하천/해양), 테마공원, 역사유적지, 전시시설 간에는 유의한 차이 없음
- 하지만 콘도미니엄, 시장과 자연경관(하천/해양), 역사유적지, 테마공원과 전시시설 순으로 더 많은 관광지 유형들과 분포에 유의한 차이를 보임
- 어느 관광지 유형이 낮다고 해석하기에는 무리

## 관광지 유형에 따른 분석

### <EDA 결과>

#### <콘도미니엄>

- 콘도미니엄 > 호텔
  - p-value: 0.015
- 콘도미니엄 > 육상레저스포츠
  - p-value: 0.004
- 콘도미니엄 > 기타문화관광지
  - p-value: 0.004
- 콘도미니엄 > 종교성지
  - p-value: 0.042
- 콘도미니엄 > 기타관광
  - p-value: 0.022
- 콘도미니엄 > 복합관광지
  - p-value: 0.014
- 콘도미니엄 > 자연관광지
  - p-value: 0.001
- 콘도미니엄 > 도시공원
  - p-value: 0.044

#### <시장>

- 시장 > 호텔
  - p-value: 0.016
- 시장 > 육상레저스포츠
  - p-value: 0.004
- 시장 > 기타문화관광지
  - p-value: 0.019
- 시장 > 기타관광
  - p-value: 0.034
- 시장 > 복합관광지
  - p-value: 0.016
- 시장 > 자연관광지
  - p-value: 0.001

#### <자연경관(하천/해양)>

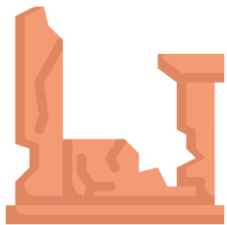
- 자연경관(하천/해양) > 육상레저스포츠
  - p-value: 0.004
- 자연경관(하천/해양) > 기타문화관광지
  - p-value: 0.001
- 자연경관(하천/해양) > 자연관광지
  - p-value: 0.000

## 관광지 유형에 따른 분석

### <EDA 결과>

#### <역사유적지>

- 역사유적지 > 육상레저스포츠
  - p-value: 0.047
- 역사유적지 > 기타문화관광지
  - p-value: 0.001
- 역사유적지 > 자연관광지
  - 0.010



#### <테마공원>

- 테마공원 > 기타문화관광지
  - p-value: 0.041
- 테마공원 > 자연관광지
  - p-value: 0.028



#### <종교성지>

- 종교성지 > 기타문화관광지
  - p-value: 0.005
- 종교성지 > 자연관광지
  - p-value: 0.045



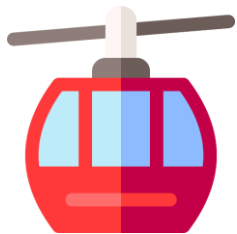


## 관광지 유형에 따른 분석

### <EDA 결과>

#### <기타관광>

- 기타관광 > 기타문화관광지
  - p-value: 0.002
- 기타관광 > 자연관광지
  - p-value: 0.029



#### <전시시설>

- 전시시설 > 자연관광지
  - p-value: 0.045
- 전시시설 > 기타문화관광지
  - p-value: 0.040



#### <도시공원>

- 도시공원 > 자연관광지
  - p-value: 0.047
- 도시공원 > 기타문화관광지
  - p-value: 0.030

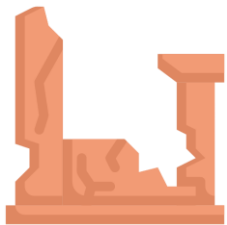


## 관광지 유형에 따른 분석

### <EDA 결과>

#### <호텔>

- 호텔 > 기타문화관광지
  - P-value: 0.015



#### <복합관광지>

- 복합관광지 > 기타문화관광지
  - P-value: 0.018



#### <육상레저스포츠>

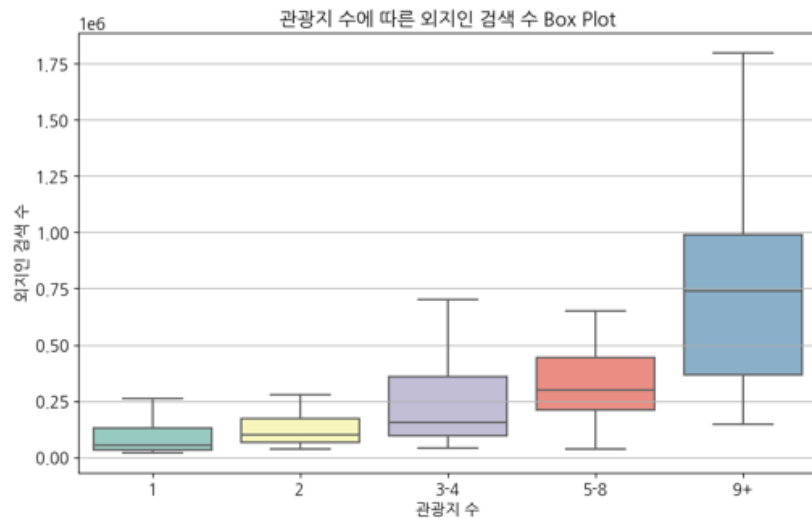
- 육상레저스포츠 > 기타문화관광지
  - p-value: 0.041



## 관광지 유형에 따른 분석

### <EDA 결과>

#### <관광지 수에 따른 외지인 검색>



- 관광지 수가 1개일 때와 2개일 때는 유의한 차이 없음
- 유형과 관계 없이 관광지 수의 증가는 외지인 검색 수의 증가로 이어짐
- T검정을 통해 관광지 수가 많을수록 외지인 검색 수도 유의미하게 증가하는 것 확인

# 결론



- 관광객 방문과 가장 큰 상관관계를 갖는 요소는 **안내소**
- 관광지의 유형보다 **관광지의 수**가 외지인 검색에 유의한 효과 제공
- 관광지 수가 2개에서 3개가 될 때 유의미한 차이 존재
  - 관광지 수가 2개인 지역은 하나 더 지으면 외지인이 더 많이 찾아올 것으로 예상
- 관광지를 설치 시 좋은 관광지 유형
  - 콘도미니엄 > 시장 > 자연경관(하천/해양) > 역사유적지 > 테마공원, 종교성지, 기타관광, 전시시설, 도시공원 > 호텔, 복합관광지 > 기타문화관광지, 자연관광지
  - 자연경관과 역사유적지: 지형과 역사의 선행 필수
  - 테마공원과 콘도미니엄: 투자 비용이 커 손실 위험 존재
  - 시장: 지역 상인들이 모여야 설치 가능
  - **기타관광, 도시공원, 전시시설**을 짓는 것이 최선의 선택으로 사료

지역 관광객 유치를 위해 안내소를 갖춘 도시공원, 전시시설 등의 다양한 관광지를 개발한다!



## <어려웠던 점>

- 원하는 형태의 데이터 부재로 전처리 과정에서의 어려움 존재
- 적은 표본 수로 가설 검정의 신뢰성 확보에 어려움 존재

## <개선 방향>

- 보다 요구사항에 부합하면서도 많은 표본 수의 데이터 활용 또는 비정형 데이터 가공을 통한 분석
- 통계학 학습을 통한 더 체계적인 가설 검정 진행
- 인구밀집지역과의 거리도 하나의 고려대상이 될 수 있을 것



**감사합니다!**