

# 관광 빅데이터 분석 기획(안)

---

2014년 4월 30일

장운호

## I. 분석 계획 개요

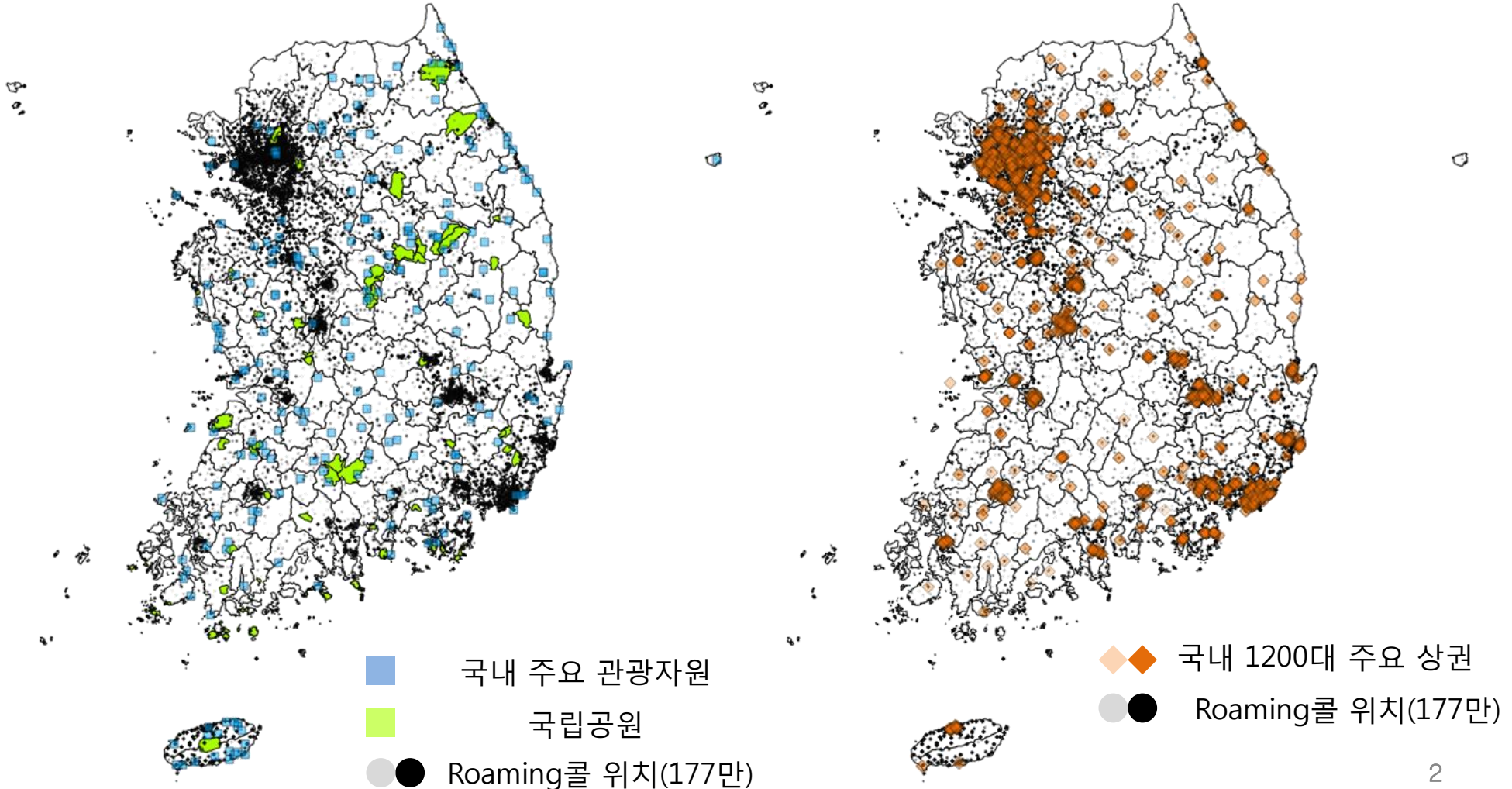
1. 데이터 개요
2. 분석 목표
3. 분석 설계

## II. 세부 분석 기획(안)

1. 검토 배경
2. 분석 목표
3. 분석 수행 계획
4. 분석 일정

177만 로밍데이터의 경위도 분포가 통념과는 달리, 국내의 주요 관광자원이나 국립공원 등보다는 유동인구가 많은 국내 1200대 상권과 외견상 유사한 분포를 보이고 있음.

- 로밍 데이터에 국내에 거주하면서 로밍을 이용하는 외국인 들의 수발신 내역도 포함된 것으로 판단됨.



로밍 데이터 분석 및 매쉬업을 통한 로밍 이용고객의 관광 유형 분석으로  
 고객을 세분화 하고 이들 유형에 맞는 차별화된 마케팅 수단을 패키지로 만들어 제공함으로써  
 국내 상권 활성화에 기여할 수 있는 Data Model 개발

구분	데이터	분석	국내 1200대 상권과의 연계 시스템 Ideation
주요 내용	27.4만 관광객의 117만 여건의 로밍데이터  수/발신 기지국의 경위도위치와 수/발신 시간 활용가능	발생 콜당 경위도 차이를 활용한 이동거리 계산으로 관광목적, 사업목적 등의 고객유형 구분 (Classification & Prediction).  수/발신콜 발생시간을 활용한 개인별 방문 주기 또는 패턴, 주 활동 시간대 추정 가능 (Sequential Pattern)	미국 펜실바니아주의 Foursquare와 연계한 지역관광 활성화 등의 사례 벤치마킹 등을 통한  지역상권과의 연계 방안 Ideation 및 Prototyping
기타	발생 콜당 경위도 차이를 이동거리로 환산 가능.	분석도구 R을 활용한 데이터 처리 프로그램 작성 및 분석 Package 활용 (Kmeans Classification, Logistic regression 등) .	“Pensyllivania partners with foursquare” ( <a href="http://Machable.com">Machable.com</a> 기사 참조)

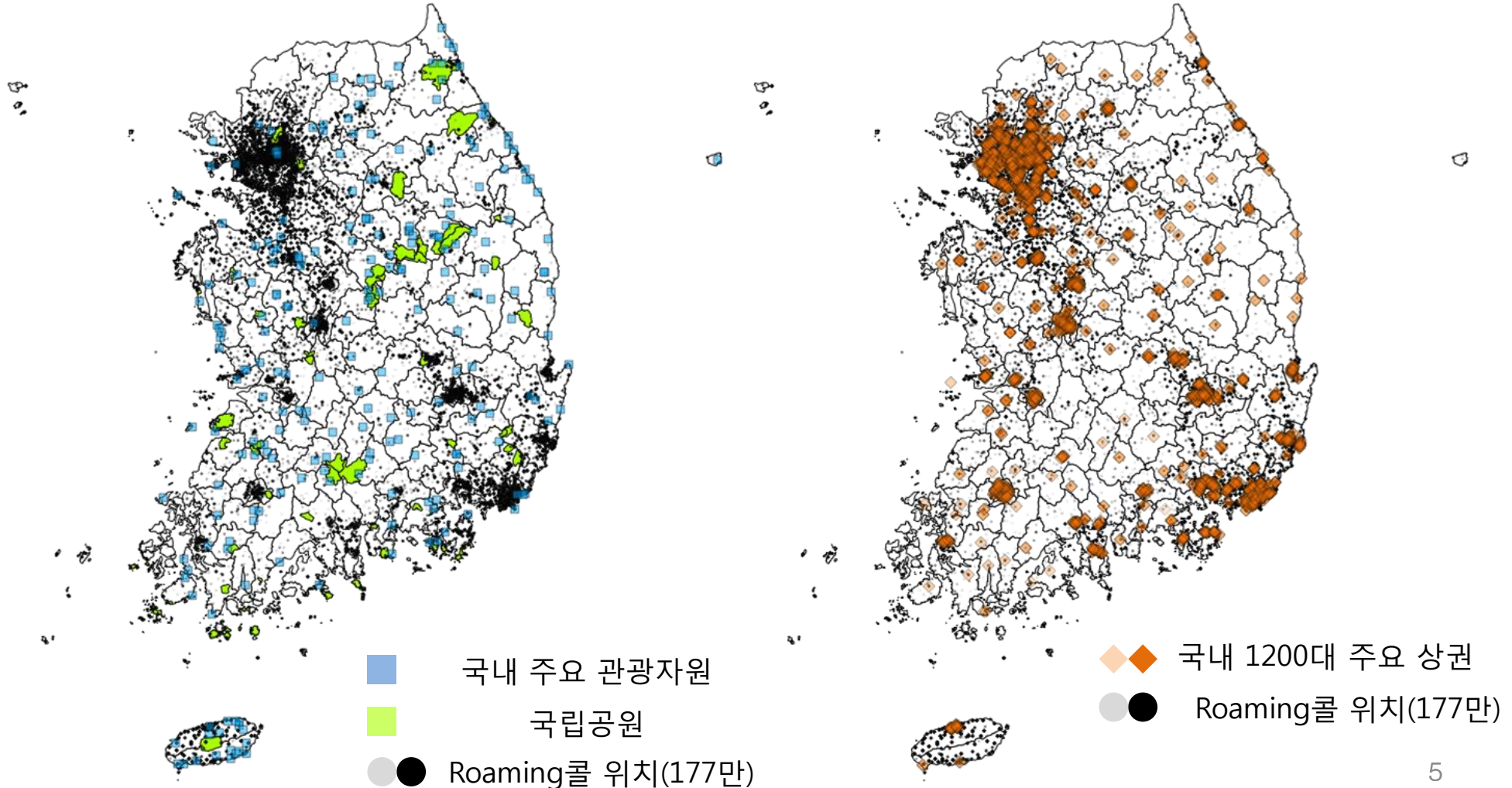
### 로밍데이터를 기반으로 콜간의 경/위도차이를 활용, 이동거리를 계산, 이를 기반으로 분류 및 Recommendation 모델 개발

- 관광객의 경우는 장거리를 이동하며, 콜발생 시점이 주로 저녁무렵일 것으로 예상됨.

분석 기법	활용 변수	적용 기법	비고
Clustering	개인별 주당 이동거리 개인별 주당 수/발신건수 수/발신건당 평균 이동거리	Hierarchical classification K-means clustering  Logistic Regression Support Vector Machine	R을 이용한 위 경도 거리계산 함수개발 필요
Buffer 검출	특정상권과의 일정거리 이내 위치여부 검출	GeoSpatial Buffer 분석	연동 불가시 R을 이용한 함수개발로 대체

177만 로밍데이터의 경위도 분포가 통념과는 달리, 국내의 주요 관광자원이나 국립공원 등보다는 유동인구가 많은 국내 1200대 상권과 외견상 유사한 분포를 보이고 있음.

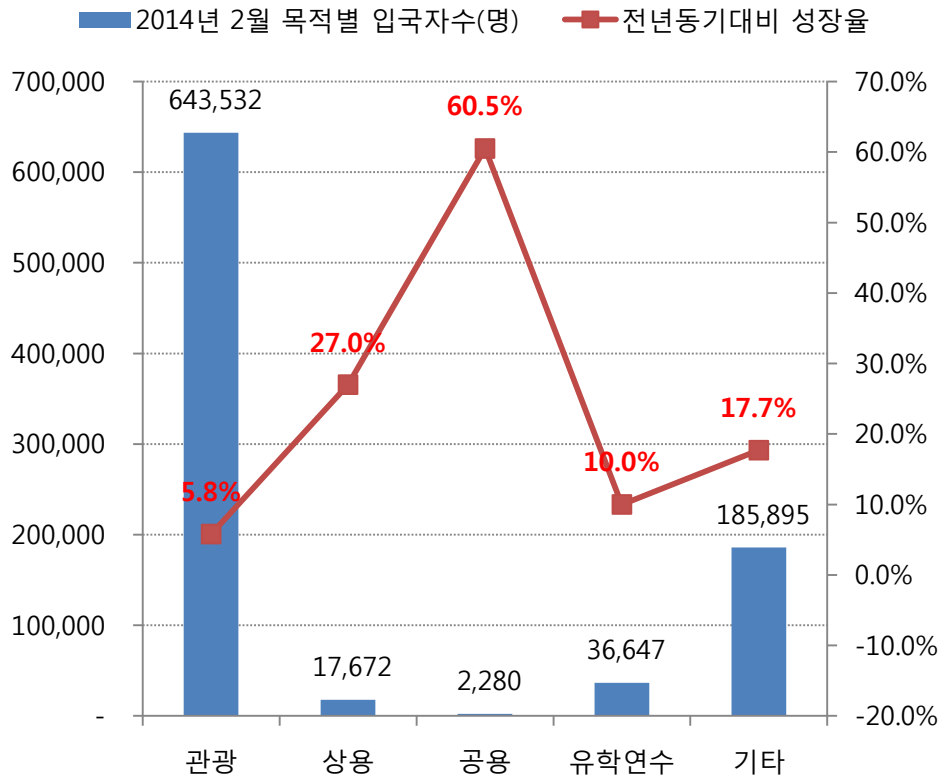
- 로밍 데이터에 국내에 거주하면서 로밍을 이용하는 외국인 들의 수발신 내역도 포함된 것으로 판단됨.



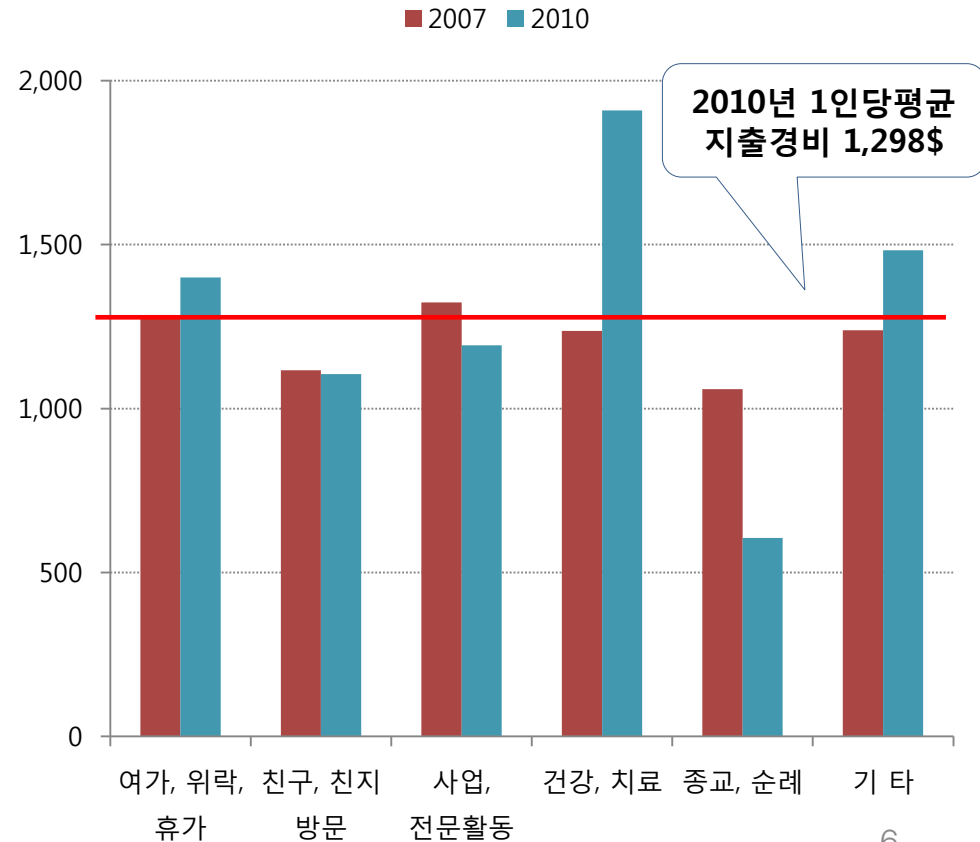
특히 관광객의 유입규모가 계속 늘고 있고 이들의 평균 지출 경비 역시 증가추세이며, 규모로는 1인당 약 154만원 수준으로 계획된 접근을 활용한 관광 소비 증대유도로 지역경제 활성화 측면에서는 신규수요 증대와 동일한 수준의 기여가 가능할 것으로 예상됨.

- 2012년 인당평균 지출경비는 약 1,469\$(154만원) : 2014년 4월 29일 15시 현재 와환은행 고시 매매기준율 기준

[2014년 2월 방문 목적별 입국자수 현황]



[방한 목적별 지출 경비 현황]



자료: 1) 한국관광공사 "관광통계월보" (2014.2월)  
 2) 통계청 "외래관광객 실태조사" (2006~2012)

로밍 데이터 분석 및 매쉬업을 통한 로밍 이용고객의 관광 유형 분석으로  
 고객을 세분화 하고 이들 유형에 맞는 차별화된 마케팅 수단을 패키지로 만들어 제공함으로써  
 국내 상권 활성화에 기여할 수 있는 Data Model 개발

구분	데이터	분석	국내 1200대 상권과의 연계 시스템 Ideation
주요 내용	27.4만 관광객의 117만 여건의 로밍데이터  수/발신 기지국의 경위도위치와 수/발신 시간 활용가능	발생 콜당 경위도 차이를 활용한 이동거리 계산으로 관광목적, 사업목적 등의 고객유형 구분 (Classification & Prediction).  수/발신콜 발생시간을 활용한 개인별 방문 주기 또는 패턴, 주 활동 시간대 추정 가능 (Sequential Pattern)	미국 펜실바니아주의 Foursquare와 연계한 지역관광 활성화 등의 사례 벤치마킹 등을 통한  지역상권과의 연계 방안 Ideation 및 Prototyping
기타	발생 콜당 경위도 차이를 이동거리로 환산 가능.	분석도구 R을 활용한 데이터 처리 프로그램 작성 및 분석 Package 활용 (Kmeans Classification, Logistic regression 등) .	“Pensyllivania partners with foursquare” ( <a href="http://Machable.com">Machable.com</a> 기사 참조)



관광객들의 행동 특성을 전문화 단계별로 구분하고, 고객별 관광 Needs에 맞는 정보 제시가 효과적이라는 다양한 연구결과들이 있는 바, 이를 데이터 분석과 연계하는 모델 구성에 활용할 필요가 있음.

- 로객 로밍 데이터를 활용한 데이터 매쉬업으로 추출한 고객 특성을 ICT 기술을 사용하여 적절한 마케팅 Tool과 연동시킴으로써 국내 주요 상권과의 연계를 강화하여 지역경제 활성화에 기여하도록 하는 큰 그림이 필요함.

연구 내용	영화제 참가자의 태도와 행동특성 분석	문화유산 관람객의 행동특성 분석	관람유형에 따른 관광박람회 관람객의 관람특성 분석
행동특성 구분을 위해 사용한 변수	동반인 교통수단 참가횟수 및 일수 관람시기	역사 유적지 여행경험 횟수 여행 거리 여행기간	동기 및 체험 형태 선호도 및 삶의 중심성 지식과 기술 경험 및 관여도 인구통계학적 특성
세분화 결과 유형 예시	일상적 단계 활동적 단계 진지한 단계	소극적 추구자 레저 추구자 전문성 추구자	기분 전환형 체험 활동형 전문 탐구형

자료) 관람자 유형에 따른 관광박람회 관람특성 분석내용을 표 형태로 재정리

#### 로밍데이터를 기반으로 콜간의 경/위도차이를 활용, 이동거리를 계산, 이를 기반으로 분류 및 Recommendation 모델 개발

- 관광객의 경우는 장거리를 이동하며, 콜발생 시점이 주로 저녁무렵일 것으로 예상됨.

분석 기법	활용변수	적용기법	비고
Clustering	개인별 주당 이동거리 개인별 주당 수/발신건수 수/발신건당 평균 이동거리	Hierarchical classification K-means clustering  Logistic Regression Support Vector Machine	R을 이용한 위 경도 거리계산 함수개발 필요
Buffer 검출	특정상권과의 일정거리 이내 위치여부 검출	GeoSpatial Buffer 분석	연동 불가시 R을 이용한 함수개발로 대체

고객별 이동거리 추출 및 수발신 로밍 건수를 기준으로 고객군 분류 및 Prediction 모델 구성 및 이를 활용한 지역 상권과의 연계방안 모색.

주요 실행 Item	5월 3일	5월 10일	5월 17일	5월 24일	5월 31일	6월 7일	6월 10일	비고
로밍 데이터 전처리								Date Class 적용(POSIXIt)
고객별 이동거리 및 수발신 건수 정리								계산 프로그램 개발
고객군 분류 (관광목적, 사업목적 등)								다양한 Clustering 분석 Algorithm 적용 비교 분석
고객군 Prediction 모델링 (Logistic Regression)								Logistic Regression
고객군별 지역상권 연계 방안 Ideation								필요시 주요 상권 우선 적용 검토
분석 보고서 작성 및 제출								

## 주요 참조자료(Reference) 리스트

---

### [각종 통계]

1. 관광 통계월보 (2013.12월 ~ 2014년 2월) : [한국관광공사](#)
2. 성별별/연령/방한목적/방한횟수별 개별여행객 1인 지출 경비(2011~2012) : [통계청](#)
3. 성별/연령/방한목적/방한횟수별 쇼핑 장소(2001~2012) : [통계청\\_주제별\\_교육문화과학\\_관광](#)
4. 연령별/방한목적별/방한횟수별 총지출경비(2006~2012) : [통계청\\_주제별\\_교육문화과학\\_관광](#)
5. 월별 한국여행시 불편사항(2004~2012) : [통계청\\_주제별\\_교육문화과학\\_관광](#)

### [관련 연구보고서]

1. 관람자 특성에 따른 관광박람회 관람특성 분석(2010) : [한양대학교 관광연구소](#)
2. 한국 관광정책 추진 사례와 향후 정책 방향(2011) : [한국 문화관광 연구원](#)
3. 관광서비스 발전을 위한 ICT 활용현황과 시사점(2013) : [정보통신정책연구원\(KISDI\)](#)
4. 관광산업고도화를 위한 생태계 구축방안 : [한국문화관광정책연구원](#)
5. 지역관광발전 추진체계 개선방안 연구 : [한국문화관광정책연구원](#)
6. 스마트관광의 추진현황 및 향후과제(2013) : [한국문화관광정책연구원](#)
7. 관광분야 사회적 협동조합 육성방안(2013) : [한국문화관광정책연구원](#)

### [관련 Link URL]

1. 미국 펜실바니아주 공식 관광 웹사이트 [www.visitpa.com](http://www.visitpa.com)
2. "Pensylvania partners with foursquare" : [Machable.com](http://Machable.com)
3. 국내주요관광자원, 국립공원 GIS-DB(TM중부원점 좌표) : [www.biz-gis.com](http://www.biz-gis.com) 자연과학분야 GIS-DB
4. 국내 주요 상권별 유동인구 정보(Katec 좌표) : [서울 열린 데이터 광장](#)