융복합 빅데이터의 지역현안 적용을 위한 분석지표 개발

-대중교통 서비스 모델 부문-

2018.3.14

최 돈 정

충남연구원 미래전략 연구단 초빙책임 연구원

Contents

01 2017년 연구회 보고

02 주요 사례 검토

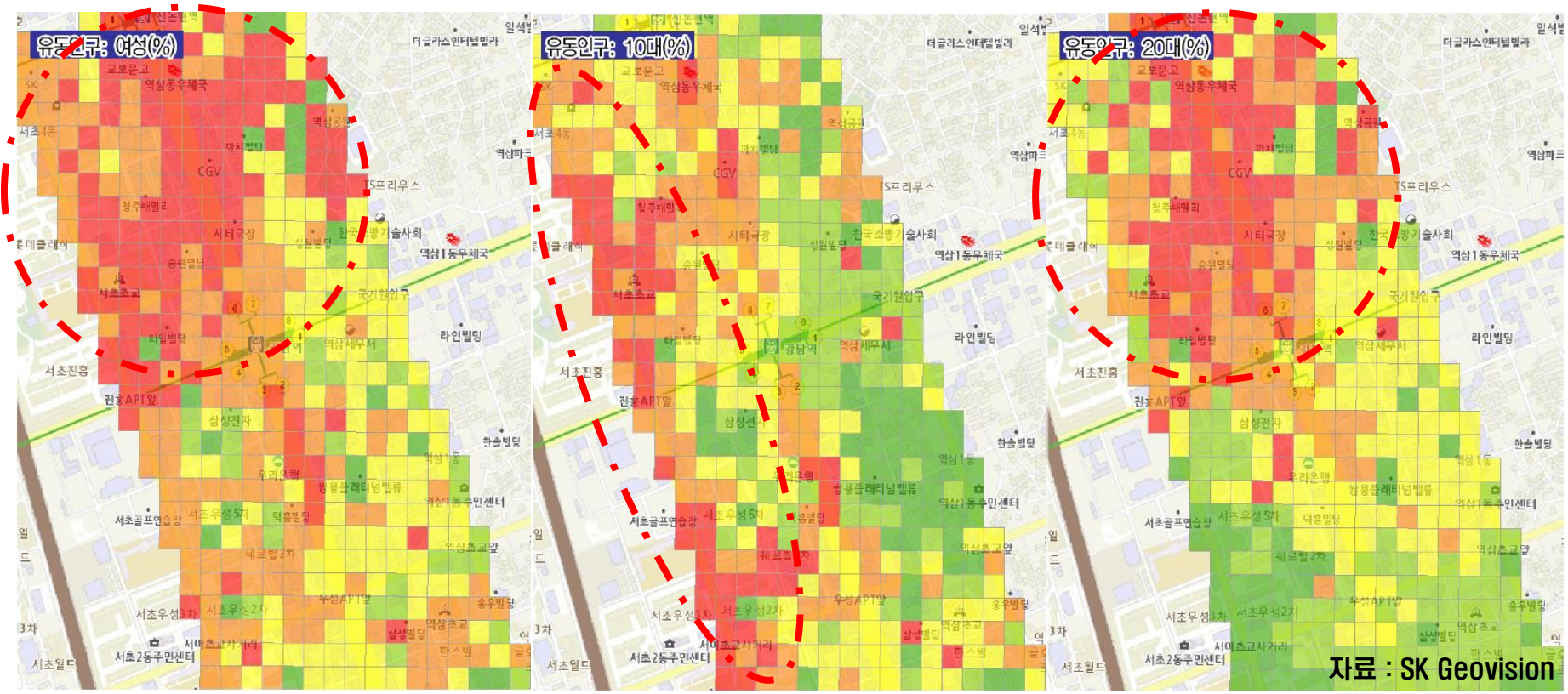
03 연구내용 및 방법

2017연구회 보고

1. 연구배경 및 목적_모바일 빅데이터와 공공데이터

➢ 모바일 빅데이터

- ✓ 대부분의 사회현상은 '공간 혹은 지리적 현상' 이며 이러한 사회현상은 개인활동의 유사성과 이질성으로 인해 발생함
- ✓ 모바일 빅데이터는 이러한 유사성과 이질성을 공간적으로 가시화 하여 개인활동의 군집에 대한 동적분포와 흐름을 파악할 수 있게 해줌
- ✓ 이러한 분포와 흐름은 공간자료와 비공간 자료를 포함한 지역의 공공데이터와 결합하여 행위를 유발하는 원인을 탐색할 수 있게 해줌
- ✓ 도시민의 행위흐름과 원인을 추론할 수 있다면 원인요소의 제어를 통해 행위의 흐름을 유도할 수 있음
- ✓ 이를 통해 '장래의 바람직한 상태를 이룩하기 위한 의도적 행위 ' 로써의 정책 의사결정이 보다 과학적인 프레임 안에서 가능해짐



- ✓ 의미 있는 정보일까?
- ✓ 어디에 쓸 수 있는 정보일까?
- ✓ 예측은 가능할까?
- ✓ 의도적으로 조절할 수 있을까?
- ✓ 가치를 부여할 수 있을까?

"왜 이런걸까?"

1. 연구배경 및 필요성_시사점

➢ **시사점**

✓ 데이터 융복합은 이미 국가적 트렌드로 발전되어 왔고 하드웨어 소프트웨어의 pool은 넘치는 상황

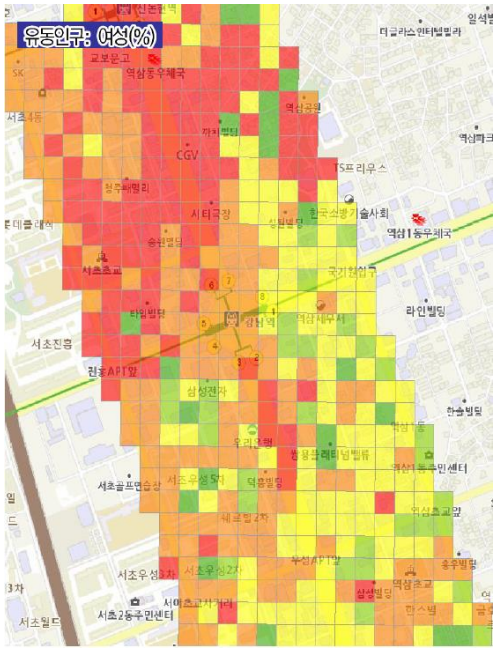
✓ 모바일 빅데이터는 인구의 행동(behavior)을 파악하는데 좋은 자료지만 그것만으로는 현상의 맥락파악이 어려움

✓ 공간데이터를 포함한 공공데이터와의 연계는 그러한 한계를 넘어 현상의 원인까지 탐색할 수 있게함(5W 1H 제공)

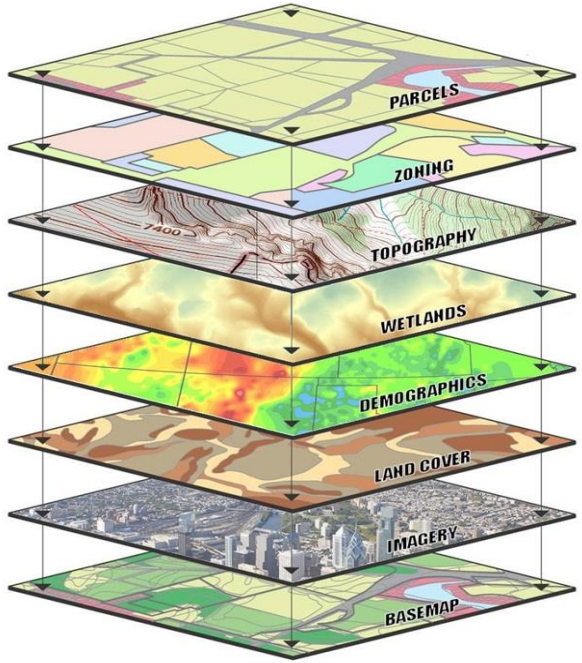
✓ 표류하는 데이터를 끌어모으고 연계하여 공간정보의 프레임 안에서 데이터간 연결고리를 찾아낼 필요가 있음

✓ 그 연결고리가 의사결정의 시발점이 되는 '의제' 즉, 의제의 근거를 직관이 아닌 데이터에서 찾아내는 것이 우리의 목적

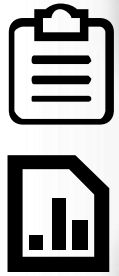
유동인구 자료



공간자료

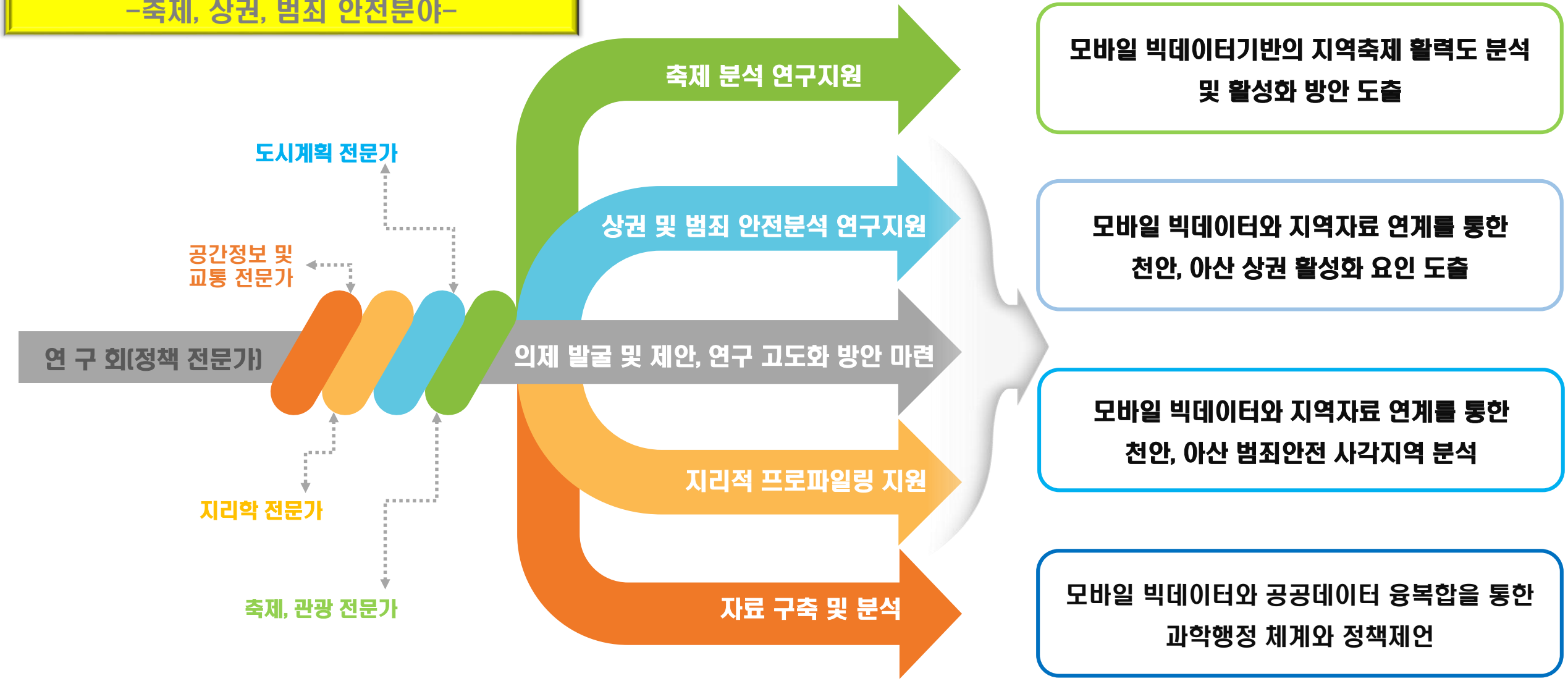


행정자료
통계자료

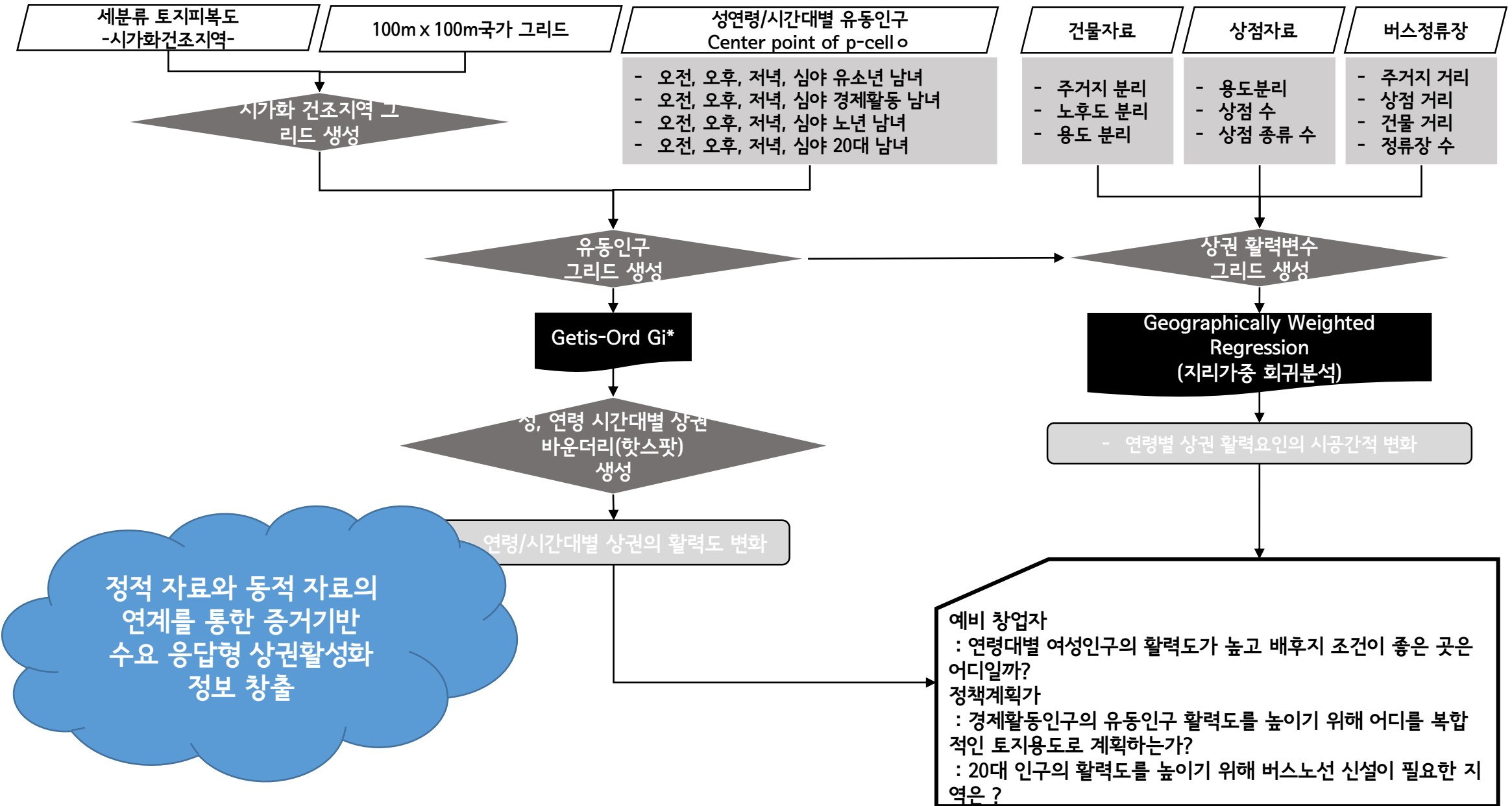


2-4. 추진체계

모바일 빅데이터와 공공데이터, 공간데이터를 연계한 공공정책 발굴
-축제, 상권, 범죄 안전분야-

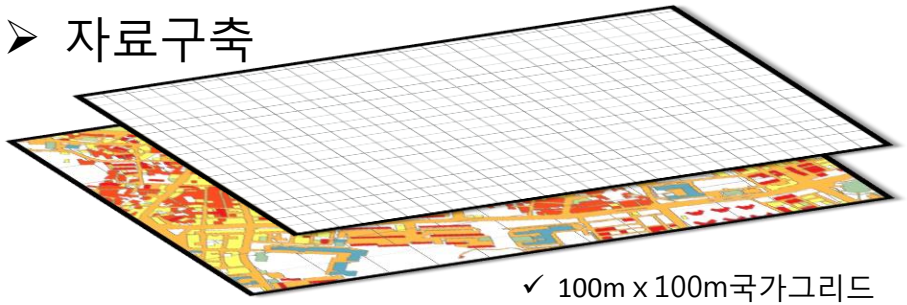


4. 연구 프레임


세분류 토지피복도
-시가화건조지역-
100m x 100m국가 그리드
성연령/시간대별 유동인구
Center point of p-cell
- 오전, 오후, 저녁, 심야 유소년 남녀
- 오전, 오후, 저녁, 심야 경제활동 남녀
- 오전, 오후, 저녁, 심야 노년 남녀
- 오전, 오후, 저녁, 심야 20대 남녀
건물자료
- 주거지 분리
- 노후도 분리
- 용도 분리
상점자료
- 용도분리
- 상점 수
- 상점 종류 수
버스정류장
- 주거지 거리
- 상점 거리
- 건물 거리
- 정류장 수
시가화 건조지역 그리드 생성
유동인구 그리드 생성
상권 활력변수 그리드 생성
Getis-Ord Gi*
Geographically Weighted Regression
(지리가중 회귀분석)
성, 연령 시간대별 상권 바운더리(핫스팟) 생성
- 연령별 상권 활력요인의 시공간적 변화
연령/시간대별 상권의 활력도 변화
정적 자료와 동적 자료의 연계를 통한 증거기반 수요 응답형 상권활성화 정보 창출
예비 창업자
: 연령대별 여성인구의 활력도가 높고 배후지 조건이 좋은 곳은 어디일까?
정책계획가
: 경제활동인구의 유동인구 활력도를 높이기 위해 어디를 복합적인 토지용도로 계획하는가?
: 20대 인구의 활력도를 높이기 위해 버스노선 신설이 필요한 지역은 ?

4. 자료구축 및 주요 분석 방법

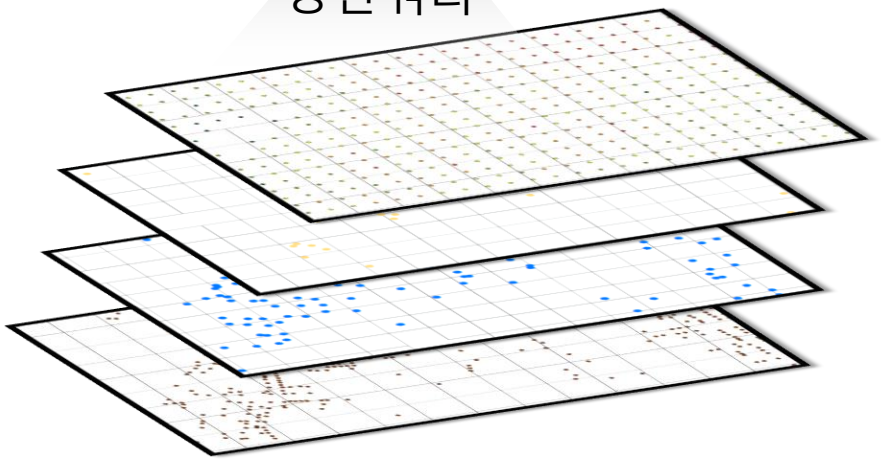
➢ 자료구축



✓ 100m x 100m국가그리드

✓ 세분류 토지피복도

공간쿼리



✓ P-cell기반의 성/연령/시간대별 인구

✓ 관광, 여가, 오락 상점

✓ 소매상점

✓ 음식점

.

.

➢ 주요 연구방법

➢ 공간 자기상관 (spatial Autocorrelation)

- 공간적 유사성과 공간적 이질성을 측정하는 공간통계학 이론
- 모든 공간은 영향을 주고받지만 가까운 공간이 먼곳보다 더 많은 영향을 주고받음
- 일반적으로 회귀모형의 y항에 공간자기상관 존재할 시 공간회귀모형의 검토가 필요

➢ Getis-Ord Gi*

- 통계적으로 유의한 수준 안에서 높은 값의 군집(hotspot)과 낮은 값의 군집(cold spot)을 가시적으로 표출, 국지적 공간자기상관 측정 도구

➢ 지리가중 회귀분석(Geographically Weighted Regression)

- 국지적 선형 회귀계 수를 추정하기 위해 이웃하고 있는 관측값들에 대하 여 거리조락에 따른 가중치를 산출하여 모델을 추정 , 특정 지역에 가까운 지역일수록 가중 치를 많이 부여하고 멀리 떨어져 있는 지역일수록 가 중치를 적게 부여하면서 해당 지역에 대한 회귀계수 를 추정하는 방법.

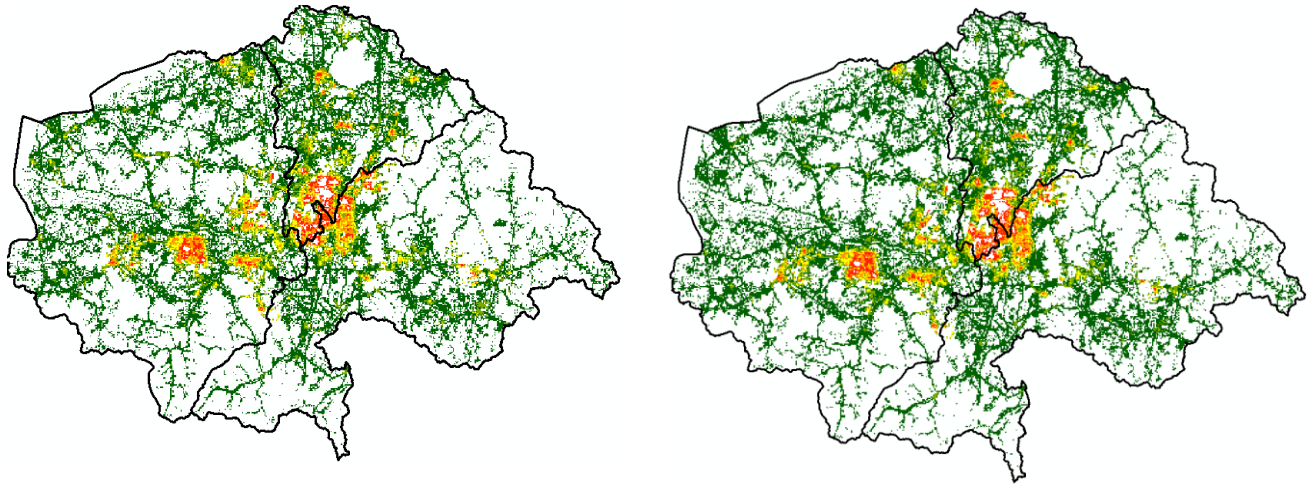
$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k (u_i, v_i)\, x_{ik} + \varepsilon_i$$

= (ui, υi)는 공간상에서 존재하는 i번째 지역이 위치한 지점의 좌표
= 이에 따라 개별 그리드에 대한 추정이 가능함

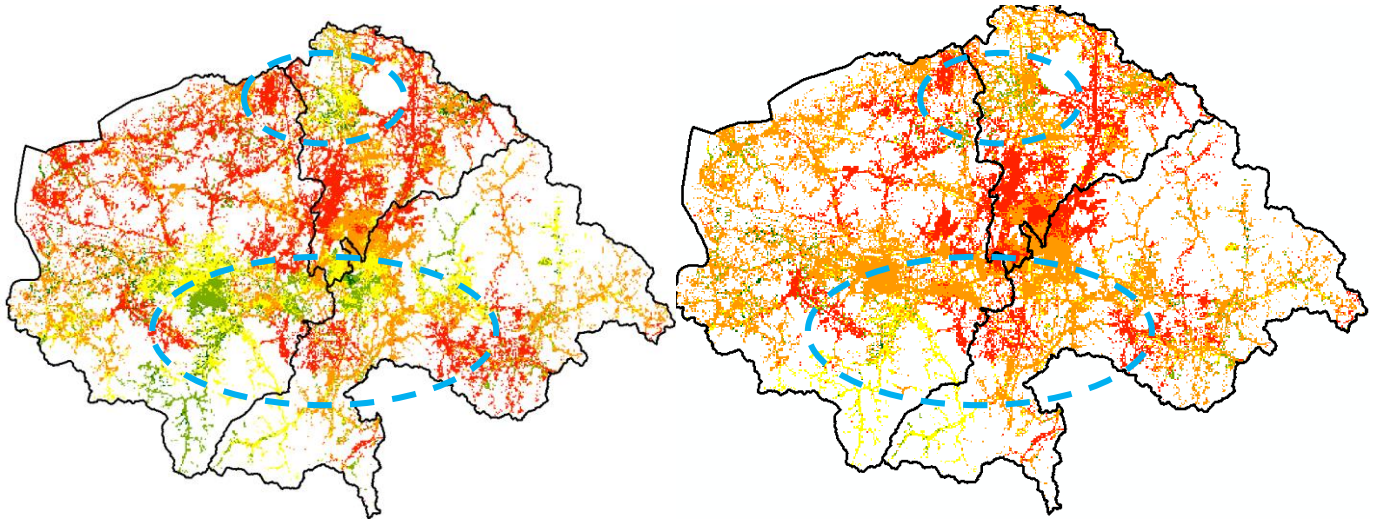
5. 분석변수 설정_동적변수(y)

- 성/연령/시간대별 일평균 유동인구의 비중
- (오전/오후/저녁/심야))

	Field	Type	설명
1	spo_no_cd	V_WString	국가 그리드 코드
2	TIME_NM	V_WString	시간대 구분
3	MAN_YOUTH	Double	남자 유소년 인구
4	WMAN_YOUTH	Double	여자 유소년 인구
5	SUM_YOUTH	Double	유소년 합계 인구
6	MAN_20	Double	남자 20대 인구
7	WMAN_20	Double	여자 20대 인구
8	SUM_20	Double	20대 합계 인구
9	MAN_30_40	Double	남자 30대 40대 인구
10	WMAN_30_40	Double	여자 30대 40대 인구
11	SUM_30_40	Double	30대 40대 합계 인구
12	MAN_EAP	Double	남자 경제활동인구
13	WMAN_EAP	Double	여자 경제활동인구
14	SUM_EAP	Double	경제활동인구 합계
15	MAN_OLD	Double	남자 노년층
16	WMAN_OLD	Double	여자 노년층
17	SUM_OLD	Double	노년층 인구 합계
18	MAN_SUM	Double	남성 인구 합계
19	WMAN_SUM	Double	여성 인구 합계
20	SUM_POP	Double	총 인구 합계
21	WEEK_END	V_WString	요일 구분 (주중, 주말, 공휴일)



✓ 오후 시간대의 경제활동 남성(좌)과 여성(우)의 유동인구 수치는 흡사한 분포를 보임



✓ 그러나 각각의 성별 총 유동인구로 보정을 해주면 해당 시간대의 유동인구 비중이 성별로 다르게 분포함을 알 수 있음, 이는 모든 연령대와 시간대에서 공통적으로 나타나는 차이

✓ 본 발표에서는 오전, 저녁 시간대의 남성과 여성 경제활동 인구에 대한 분석결과만 공개

6. 분석변수 설정_정적변수(X)

분류	내용	근거
토지이용 혼합도	단위 격자안의 토지이용 용도의 복합성 지수	Frank et al., 2005
대중교통 접근성	격자안 모든 주택에서 도로망 기준 400m이내의 평균 버스 승강장 수와 거리	Tresidder, 2005
	격자안 모든 상가에서 도로망 기준 400m이내의 평균 버스 승강장 수 와 거리	
상점 접근성	격자망 안의 모든 주택에서 400m이내 상점 까지의 평균거리	Kang, 2013
상점 수	격자망 안의 모든 주택에서 400m이내 상점 까지의 상점 수	
상점종류 다양성	격자망 안의 상점 종류 수	
배후지 주거분포	격자망 중심점으로부터 400m안의 모든 주택 수	Cardozo et al., 2012

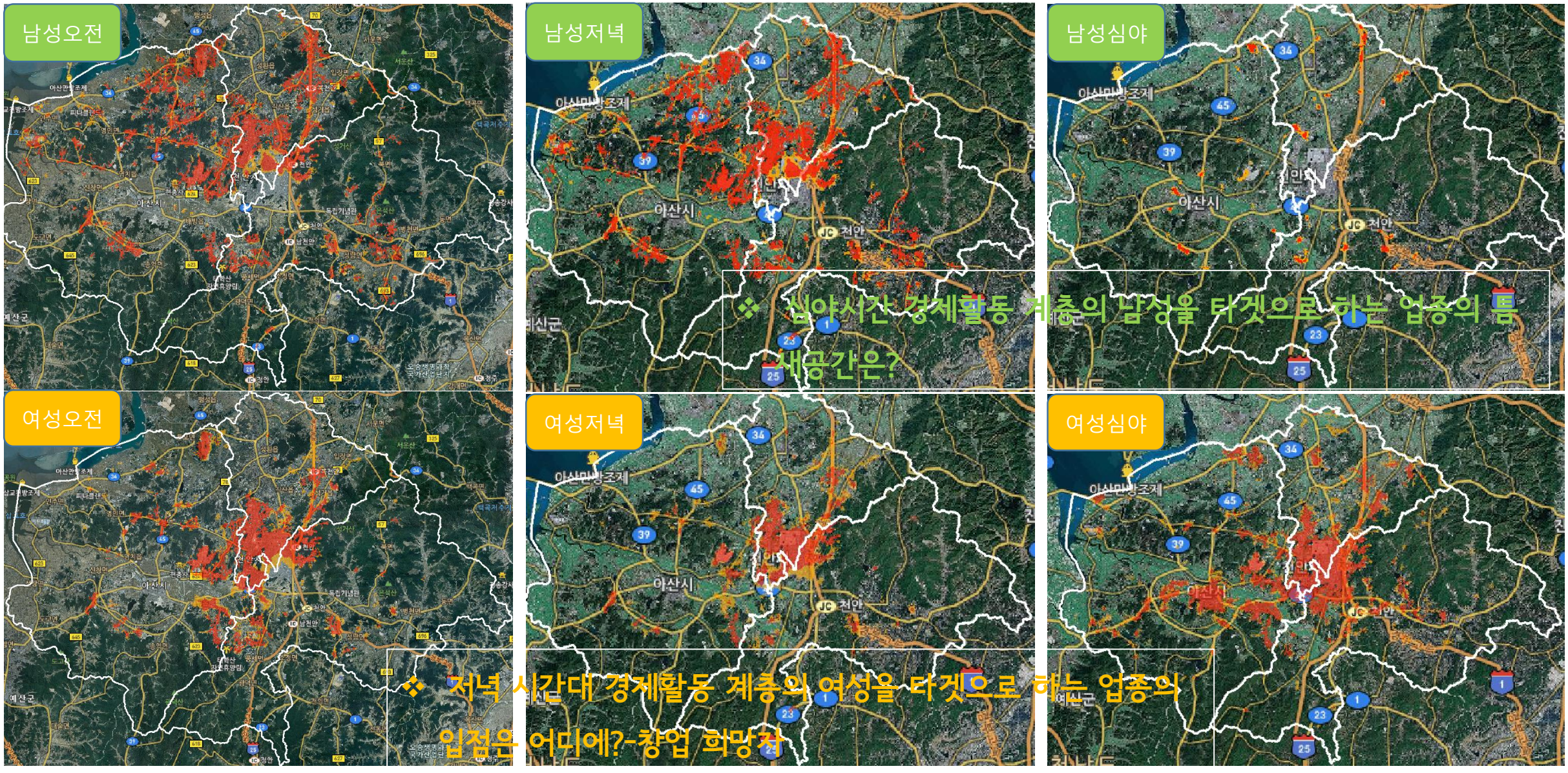
✓ 연령별 상권 활력도에 영향을 미치는 변수는 공통적으로 적용된 사례가 있는 경우만 분석에 활용

: 교통접근성, 상점 접근성, 배후지 주거, 토지이용 혼합도

✓ 추가적으로 연령에 맞춘 변수의 개발이 필요함

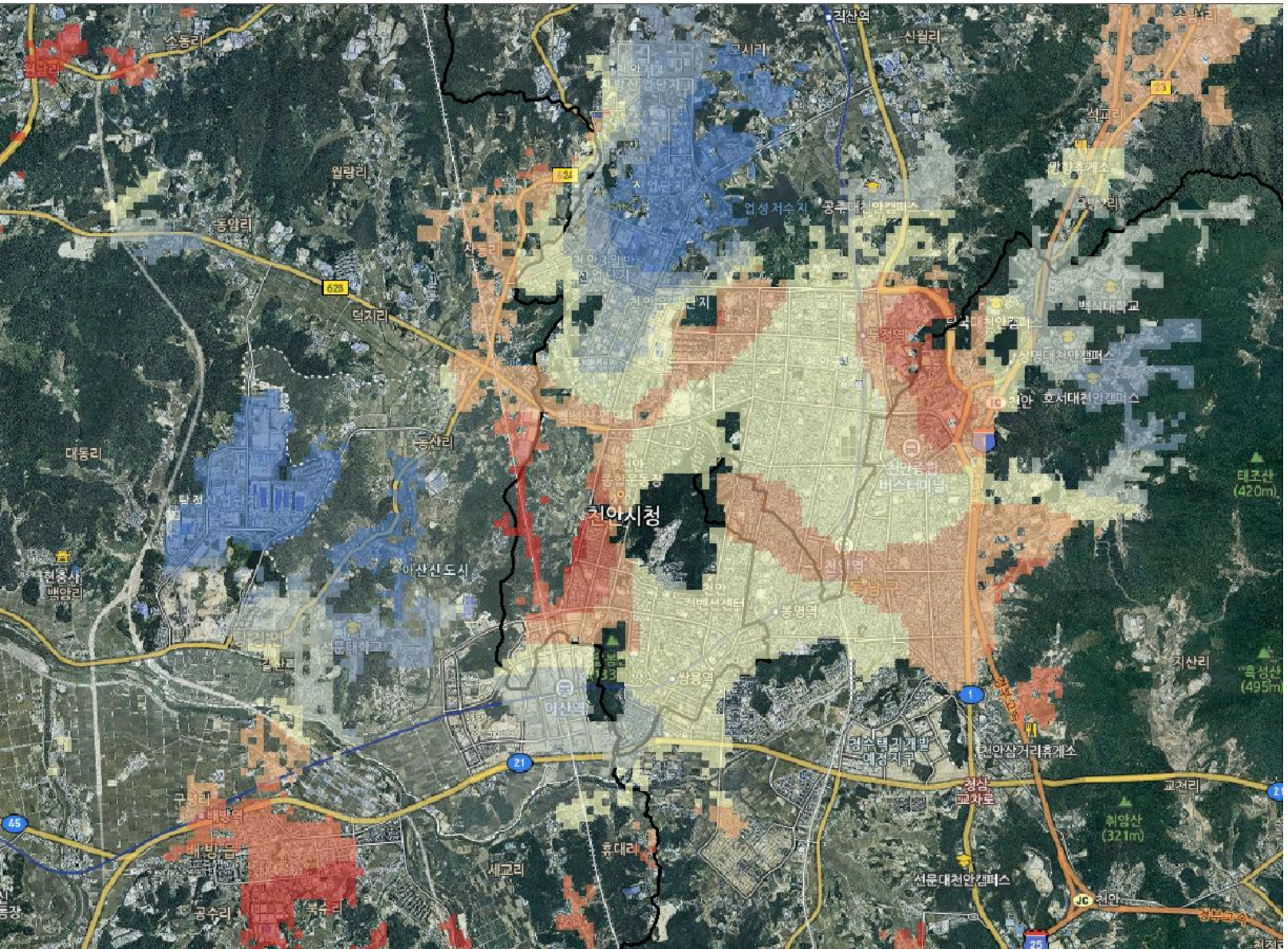
7. 주요 연구결과_성 연령별/시간대별 유동인구의 패턴변화

➢ 시간대별/성별 일평균 경제활동인구의 핫스팟



8. 주요 연구결과_유동인구 활력도의 영향요인에 관한 지리가중회귀분석결과

➢ 저녁 시간대 여성 경제활동 인구와 토지이용 혼합도



✓ 불당동, 배방읍, 두정역, 천안역인근은 토지이용 혼합도가 저녁 시간의 여성 유동인구에 (+)한 영향력을 미치는 것으로 나타남

✓ 반면 탕정 신도시, 천안 지방산업단지 주변의 토지이용 혼합도는 저녁시간대 여성 유동인구에 (-)한 영향력을 갖는것으로 나타남

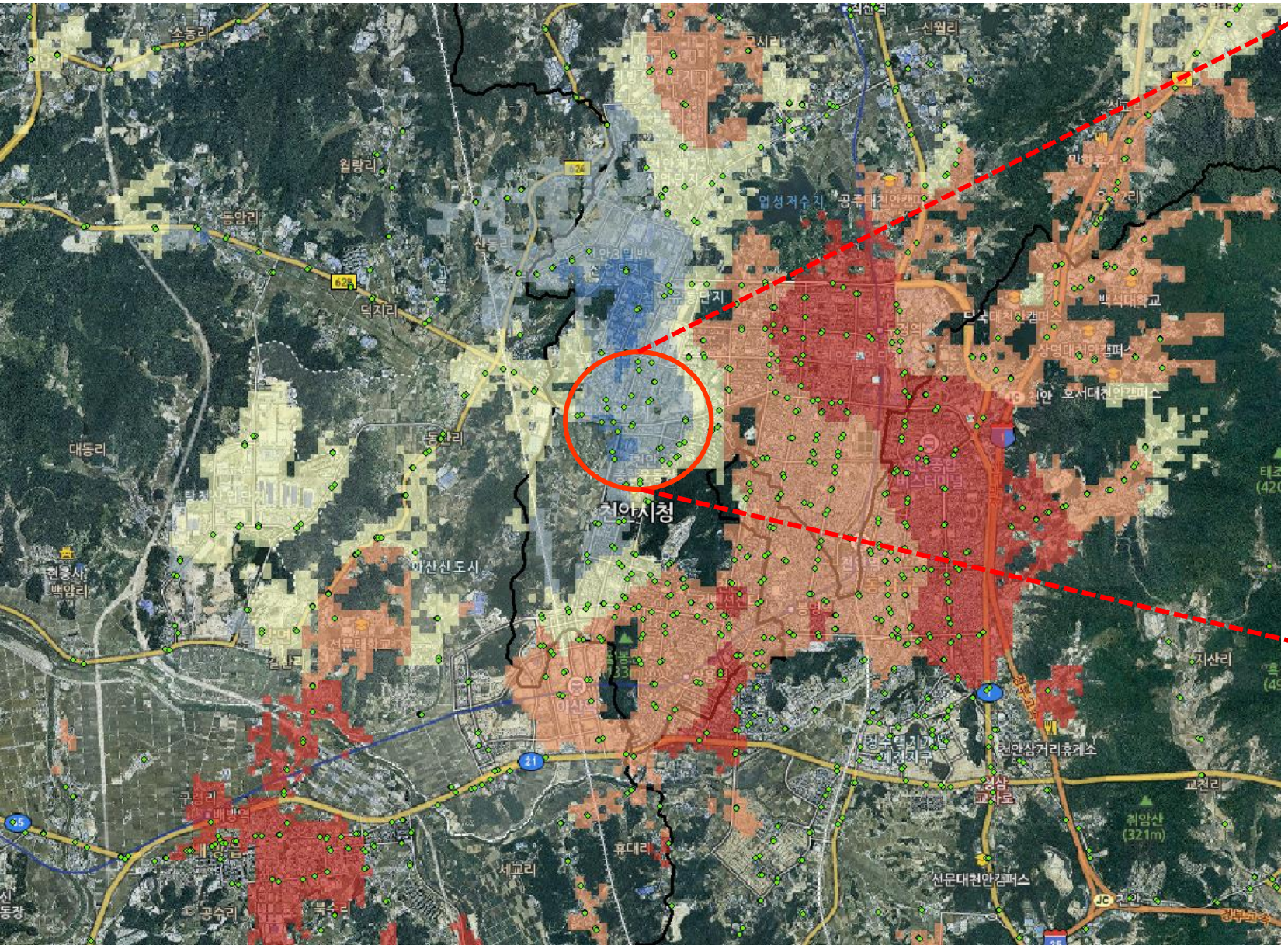
✓ 이는 저녁 시간대는 대단위 주거지를 포함한 지역에서의 토지이용 혼합도가 여성의 활동을 촉진한다는 의미

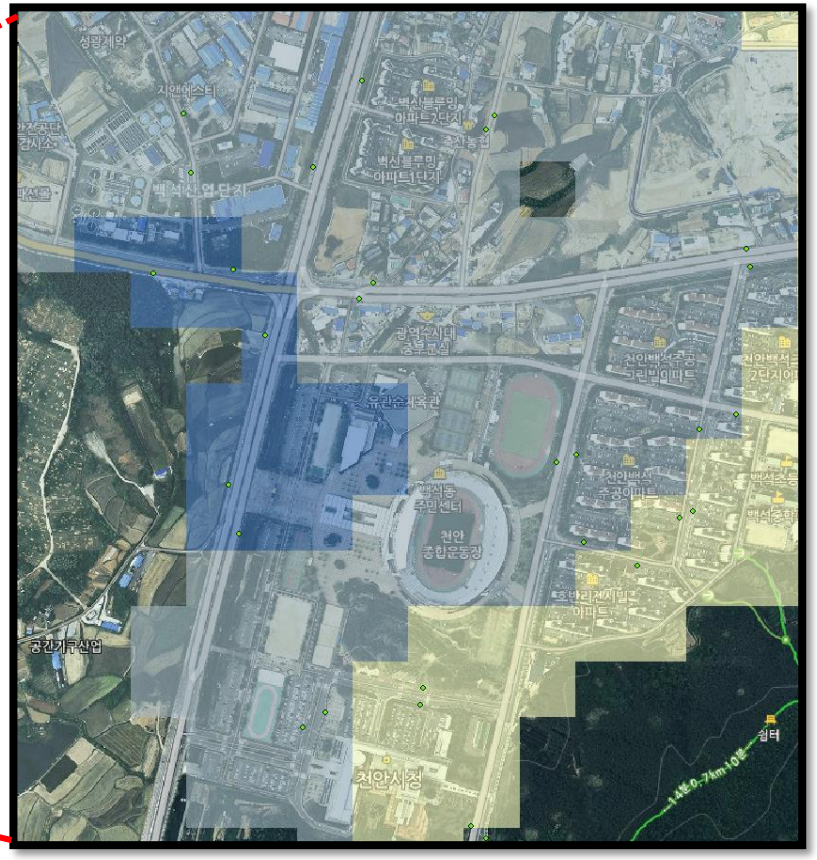
❖ 저녁 시간대 경제활동 계층의 여성을 타겟으로 하는 업종의 입점은 어디에?-창업 희망자

❖ 탕정지구의 혼합적 토지이용개발이나 천안역 인근의 도시재생 사업을 통해 상권 활성화를 도모하는것이 타당한가?- 정책 담당

9. 주요 연구결과_유동인구 활력도의 영향요인에 관한 지리가중회귀분석결과2

➢ 저녁 시간대 여성 경제활동 인구와 주거지→버스정류장 까지의 거리





✓ 유동인구가 많은곳 들의 군집지역에서 버스정류장의 거리가 멀다면 버스노선 신설을 통한 대중교통 확충을 통해 주변지역 상권과 연계할 수 있지 않을까?

10.1 주요 일정진행(1차 연구모임_2017.3.27)

10.1 주요 일정진행(2차 연구모임_2017.6.1)

10.1 주요 일정진행(3차 연구모임_2017.7.25)

10.1 주요 일정진행(4차 연구모임_2017.9.27)

10.1 주요 일정진행(최종보고회_2017.11.27)

10.1 주요 일정진행(최종보고회_2017.11.27)

10.1 주요 일정진행(최종보고회_2017.11.27)

10.1 주요 일정진행(충청남도 데이터기반 행정 빅데이터 활용에 관한 조례, 2017년 12.29일 공포)

충청남도 데이터기반행정 빅데이터 활용에 관한 조례안

(이공휘 의원 대표발의)

의안번호	제 호

발의연월일 : 2017년 10 월 일

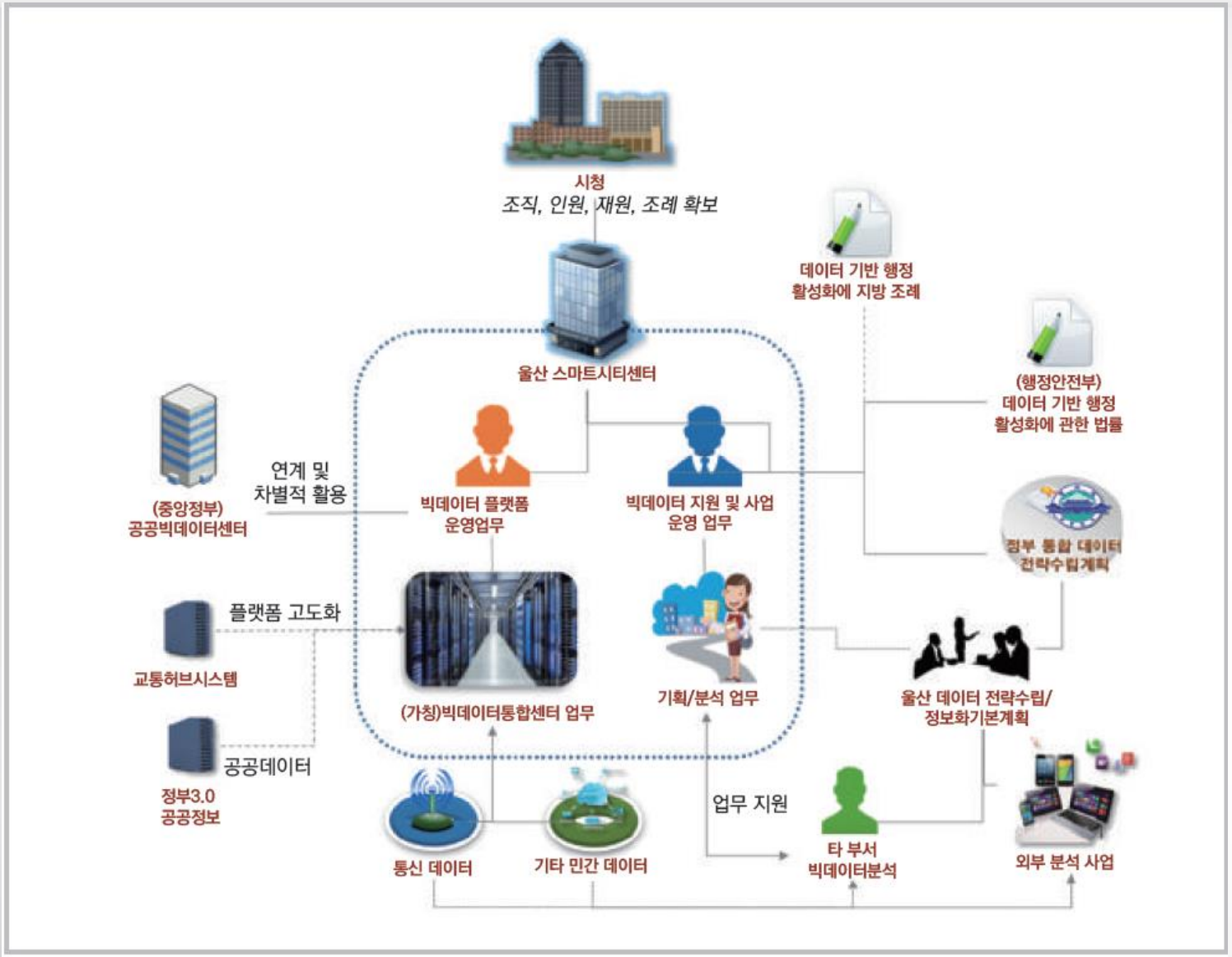
발 의 자 : 이공휘,김 연,조길행, 김동욱,김종문,유찬종, 유익환,홍재표 의원

찬 성 자 : 김원태 의원 외 11인

1. 제안이유

○ 충청남도가 데이터기반행정 빅데이터 활용을 통한 행정의 효율성을 확보하고, 도민 서비스 제공 및 민간 활용지원으로 도민의 삶의 질 향상과 지역 경제 발전을 도모하고자 함

제언1_빅데이터 플랫폼의 구축



자료: 울산발전 연구원(2017)

- **전담 조직과 인력의 확충**

- 데이터의 속성에 따른 전문성과 네트워킹이 가능한 전문성 확보

- 민간과 공공의 가교역할 수행

- **분석 주제 기획 및 공동의 연구수행을 위한 민관 거버넌스 구축**

- 분석주제기획 및 검증, 데이터 구축, 분석 알고리즘 검토 및 결과 검증 단계까지 지속적으로 함께할 수 있는 실국& 전문가 위원회 필요

- **데이터에 대한 인식 전환 및 협조체계 강화**

- 소유→활용을 위한 원료로써의 데이터 태러다임 변화 수용

❖ 현안에 관한 빅데이터 분석 프레임 개발

✓ 공간정보의 프레임 안에서 지역의 데이터를 융복합 하여 현안에 관한 정보를 생산하고 유의미한 인사이트를 제시

: 지역축제의 방문객 산정이 과대추정 되고 그로인한 부정적 이미지와 예산의 낭비를 초래한다는 근거제시

: 단일 상권의 활력을 촉진하는 요소가 공간별, 시간대별로 상이하다는 근거를 정량적, 시각적으로 제시하여 추후 상권분석 시 활용할 새로운 빅데이터 분석 프레임을 제시함

❖ 충청남도 빅데이터 관련 전문가 거버넌스의 뼈대 구축

✓ 연구회 활동을 통해 각 분야의 전문가 네트워크가 형성 되었고 올해 연구회에도 연속적으로 참여

✓ 네트워크를 통해 추후 진행될 전문 조직 및 충남도 빅데이터 분야를 주도할 거버넌스의 한 축완성

❖ 충남도 빅데이터 조례 마련

공공부문의 모바일 데이터 & 지역자료 연계 사례 검토

사업명	사업비				투입공수
	계	개발비	직접경비	기타	
관광분야 표준분석모델	128,012	65,374	51,000	11,637	93
교통(버스)분야 표준분석모델	94,922	65,292	21,000	8,629	92
민원분야 표준분석모델	56,666	50,514	1,000	5,151	63
CCTV분야 표준분석모델	70,329	42,936	21,000	6,393	59
공동주택분야 표준분석모델	60,985	54,441	1,000	5,544	82
토지 상권분석	355,218	262,926	60,000	32,292	359
차량 DTG 빅데이터 분석	226,072	175,520	30,000	20,552	232
시설입지선정 빅데이터 분석	350,514	288,649	30,000	31,864	379
누락 적발 빅데이터 분석	300,102	272,820	-	27,282	343
기초생활인프라 빅데이터 분석	334,497	244,088	60,000	30,408	319
사기 적발 빅데이터 분석	321,192	291,992	-	29,199	367

주 : 직접경비의 경우 민간데이터 구매비(유동인구, 카드매출 등) 포함
자료 : 행정안전부 빅데이터 사이트 : http://www.bigdata.go.kr

2. 사례 1

- 과제명 : 환자안전 조기 이상감지 시스템 구축
- 수행기간 : 2016년 6월 ~ 11월
- 주요 내용 : 의약품안전사용정보(DUR)의 실시간 데이터 기반 감염병 이상징후 모니터링 시스템 구축
- 활용 데이터 : 의약품 처방 데이터(2010년 이후 약 51억건), 진료기록 데이터(약 2조9천억건)
- 수행기관 : 건강보험심사평가원, 디비디스커버, 오픈메이트

※ 한국정보화진흥원 정책본부 미래전략센터의 '데이터 기반 미래전략 정책지원' 사업으로 추진됨

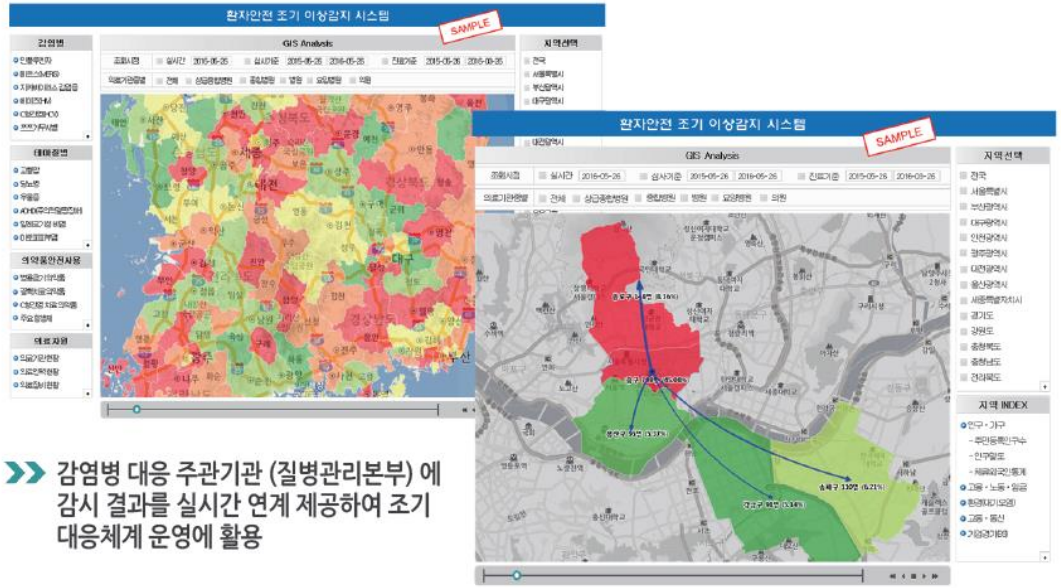
감염병 추적을 위한 데이터 분석 프로세스

- **1단계** **Target 질환 선정** : 모니터링이 필요한 우선 항목 선정
- **2단계** **처방내역 분석** : Target 질환의 최근 3~5년간 청구자료 분석
- **3단계** **의약품 처방패턴(처방약 조합SET) 구축** : 공통으로 처방된 약제 조합 도출
- **4단계** **의약품 처방패턴 타당성 검증** : 처방패턴을 조건으로 역검증(질병 역산출)
- **5단계** **DUR 시스템 연계** : 처방패턴과 일치하는 처방 발생시 감염병 의심

실시간 감염병 모니터링 서비스

조회 결과
: 특정 거주지역의 호나자가 이용한 의료기관 소재지를 추적하여, 환자의 이동경로를 분석할 수 있음

실시간 감염병 모니터링 서비스

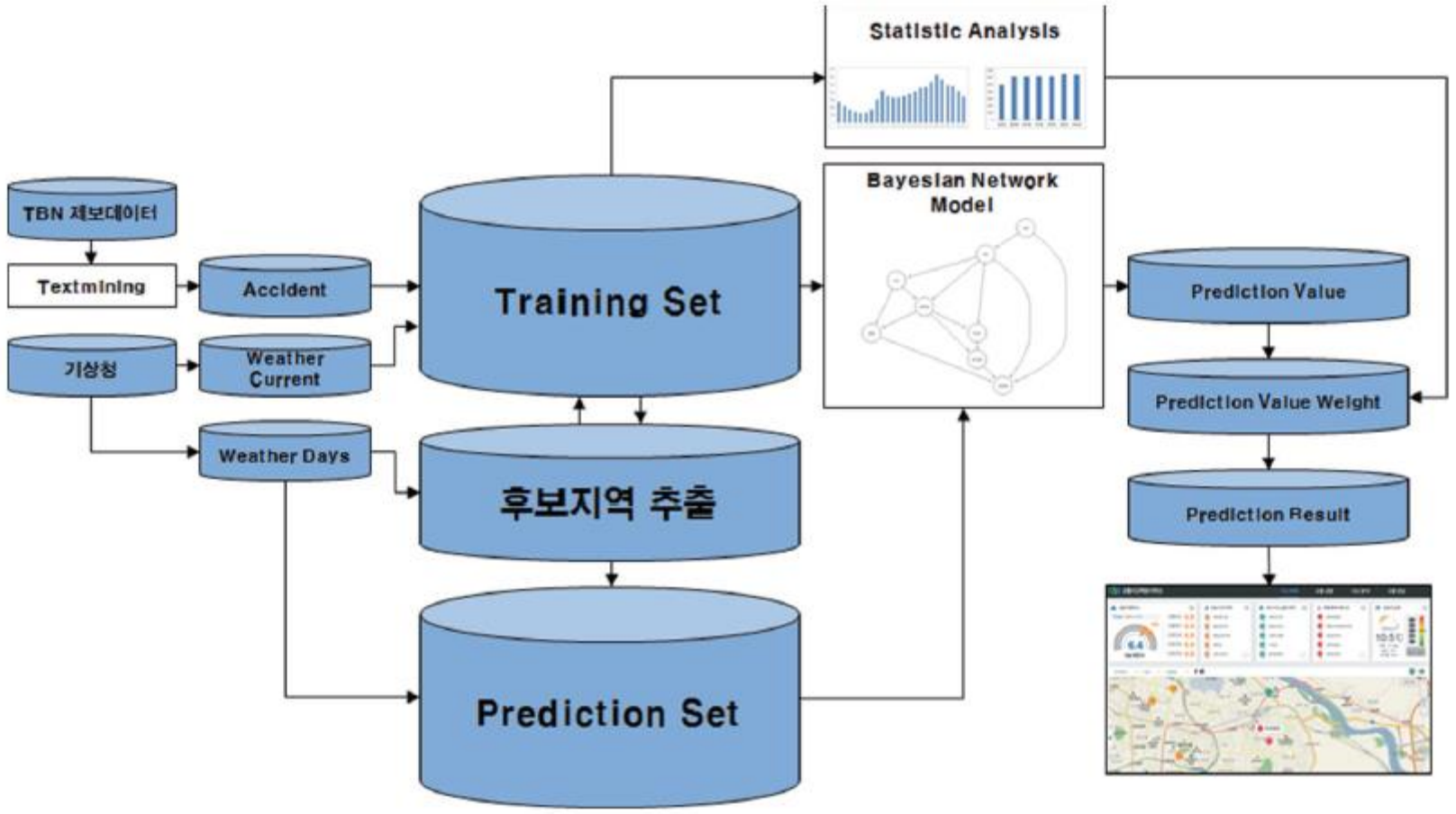


>> 감염병 대응 주관기관 (질병관리본부) 에 감시 결과를 실시간 연계 제공하여 조기 대응체계 운영에 활용

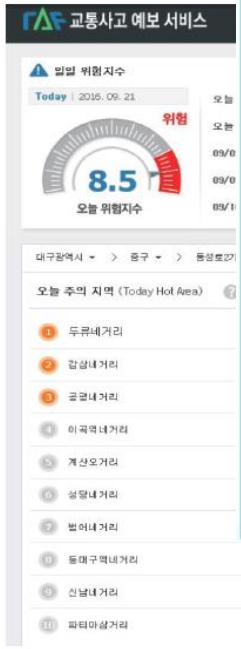
2. 사례 2

- 과제명 : 교통사고 감소를 위한 데이터 분석 기반 사고예보 서비스
- 수행기간 : 2016년 6월
- 주요 내용
  - 교통사고 제보, 사고 기록
  - 분석 결과를 교통방송에서
  (대구 · 부산)
- 활용 데이터
  - 도로교통공단 한국교통방
  - 도로교통공단 : 교통사고
  - 기상청 : 기상데이터(20
- 수행기관 : ㈜더아이엠
  ※ 한국정보화진흥원 정책분

교통사고 예보를 위한 데이터 분석 프로세스



교통사고 예보 서비스



2017-01-13 18:35:00.0
2017-01-13 18:31:00.0
2017-01-13 18:30:00.0
2017-01-13 18:27:00.0
2017-01-13 18:24:00.0

고 급증 지역
동서고가도로
광안대교
거면교차로
북양터널어귀교차로
산장

사상구

구별 교통사고 정보

강서구 2	금정구 1	기장군 0	남구 0
동구 2	동래구 2	부산진구4	북구 0

버스관련 입니다.
승용차관련 입니다.
단독 입니다.
택시관련 입니다.
동구 의 금요일 에 교통사고가 많이 발생한 유형은 승합차관련 입니다.
동래구 의 금요일 에 교통사고가 많이 발생한 유형은 승용차관련 입니다.
부산진구 의 금요일 에 교통사고가 많이 발생한 유형은 택시관련 입니다.
북구 의 금요일 에 교통사고가 많이 발생한 유형은 승용차관련 입니다.

2. 사례 3_지자체별 빅데이터 사업 현황

지자체	기존 사업	당해연도 사업
서울	• 심야 올빼미버스 배차 분석 • 지역상권 분석 등 다수	• 빅데이터 캠퍼스 운영
부산	• 도시서비스분석(2015년) • 부산관광산업 동향분석(2016년) • 공공보건사업 수행을 위한 소지역 건강지표 개발(2016년) • 소지역중심 만성질환 유병률 지표 산출 연구(2016년)	• 빅데이터 통합 플랫폼 구축 및 분석사업 • 부산관광산업 동향분석
대구	• 빅데이터 통계분석시스템 구축(1차)	• 빅데이터 통계분석시스템 구축(2차)
인천	• 내외국인 관광객 유동인구 빅데이터분석(2015년) • 인천광역시 빅데이터 정보화 전략 계획수립 및 민원데이터 분석(2016년)	• 내외국인 방문 관광지 및 축제 빅데이터 분석
광주	• 시민의소리(민원) 분석(2014년) • 시내버스 효율적 운영(2014년) • 교통사고 및 청소년 자살예방을 위한 빅데이터분석(2015년) • 빅데이터분석 플랫폼 구축 및 활용서비스 개발(2016년)	• 빅데이터 분석 플랫폼 고도화
울산	• 교통 데이터허브 시스템 구축사업 • 「태화강봄꽃대향연」 축제효과분석	• 빅데이터 기반 사회 안전취약지역 분석
세종	• 복지 자원분석 서비스	• VOC(시민의소리) 기반 민원 분석서비스
경기	• 2016년 빅데이터분석사업(7개 과제) – 공동주택 부조리 분석, CCTV 분석모델 고도화 및 확산 외 다수 • 2015년 빅데이터분석사업(11개 과제) – 공동주택 관리비·계약 부조리 분석, CCTV 사각지대분석 외 다수 • 빅데이터 활용 문화 확산 – 빅포럼 개최, 빅데이터 활용 아이디어 공모전 외	• 2017년 빅데이터분석사업(6개 과제)
강원	• 전통시장활성화(2015년)	

지자체	기존 사업	당해연도 사업
충북	• 빅데이터 활용 중장기 기본계획 수립(2016년)	• 2017년도 공공빅데이터 신규분석 과제수요조사 • 지자체 빅데이터 공통기반 공동활용 수요조사
충남	• 없음	• 빅데이터 기반 지하수 분석관리시스템 구축
전북	• 전주시 상수도누 수율 빅데이터 분석 • 전라북도 미세먼지 원인 분석	• 전라북도 빅데이터 기본계획('17년~'21년) 수립 • 빅데이터 시범분석 과제(3개) • 빅데이터 역량 강화교육
전남	• 호남선 KTX개통에 따른 관광객 행태 분석(2015년) • SNS기반 관광 키워드 빅데이터분석(2016년)	• 관광 패턴분석을 통한 관광지 거점 연계교통망 확충
경남	• 응급환자 골든타임 확보를 위한 빅데이터 분석 • 중국인 관광객유치 전략마련을 위한 빅데이터 분석	• 과제 미 확정
제주	• 공공빅데이터 표준분석 모델 구축(4개 분야) • 개별관광 증가에 따른 관광객 소비패턴 분석 • 관광산업 일자리 미스매치 해소를 위한 빅데이터 분석 • 내도 관광객 취향 분석을 위한 빅데이터 분석 • 스마트 관광 저장분석 플랫폼 구축	• 교통, 전기차 분야 빅데이터 분석과제 추진

연구내용 및 방법

1. 민간+공공데이터 융복합을 통한 대중교통 서비스 표준분석 모델 적용 및 정책제언

추진 목적

교통카드 데이터 와 민간 데이터를 융합 분석함으로써 데이터 기반의 대중교통 정책 수립 및 활용을 지원함

수행 실적

사례 및 현황분석

1. 기존 사례 분석
2. 요구사항 정의

데이터 수급

1. 제주도/포항시 데이터
   (교통카드 데이터, 버스정류장, 노선별 버스 정류장, 시청/보건소/병원/사회복지시설)
2. 주거인구 데이터(국토지리정보원)
3. 유동인구 데이터(민간 데이터)

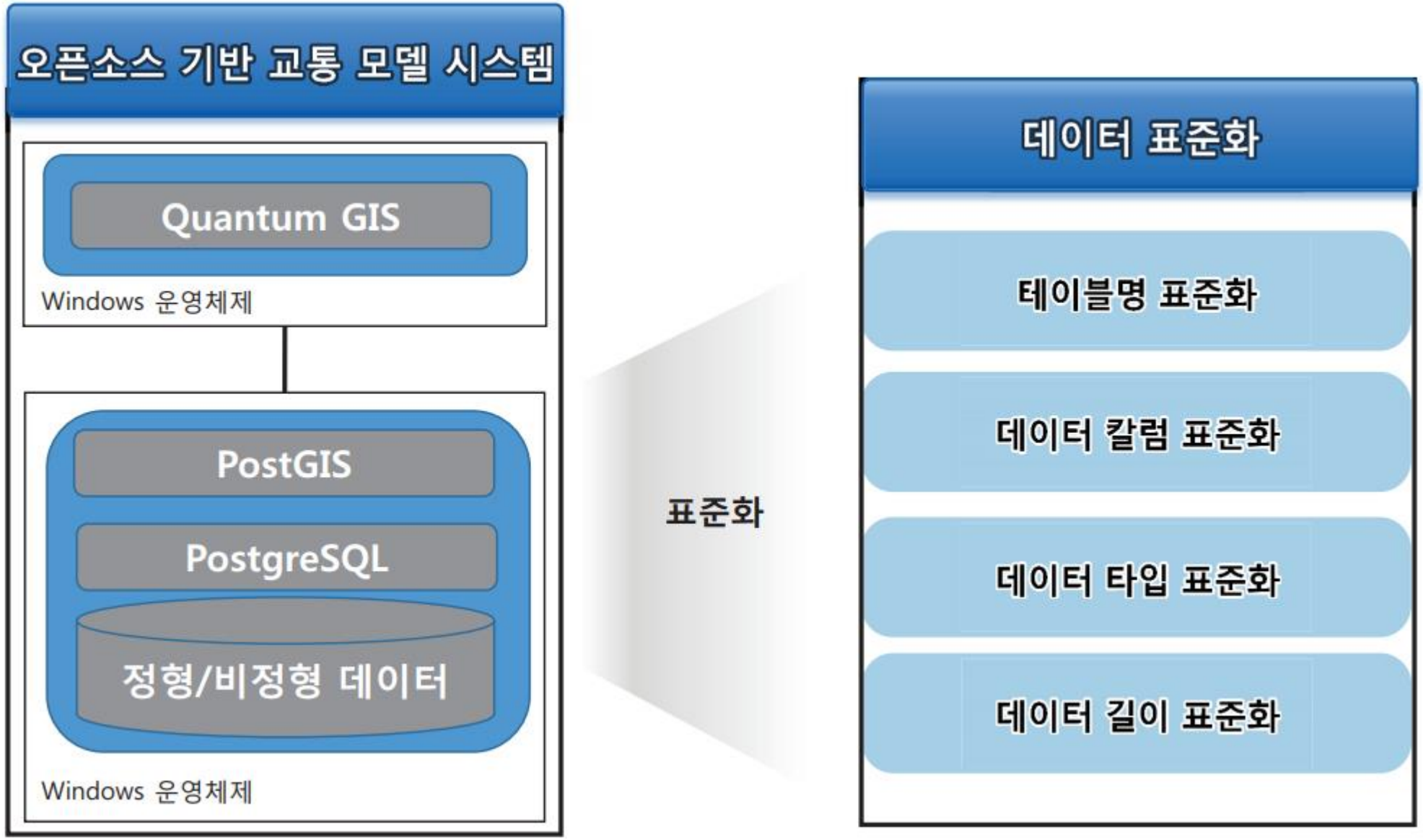
분석 대상

1. 대중교통 사각지대 노선 분석
2. 저상버스 도입 노선 분석
3. 버스 중복도 분석
4. 환승 편의시설 입지 분석
5. 탄력배차제 노선 선정
6. 통행패턴 분석

1. 민간+공공데이터 융복합을 통한 대중교통 서비스 표준분석 모델 적용 지자체

분석 모델	수집 데이터	보유 기관	수집방식	수집유형
1. 대중교통 사각지대 분석	버스정류장	제주도/포항시	Off-Line	CSV
	주거인구(전체 인구)	국토지리정보원		
	유동인구	민간기업		
2. 저상버스 도입을 위한 노선 분석	노선별 버스 정류장	제주도/포항시		
	시청/병원/보건소 /사회복지시설	제주도/포항시		
	주거인구(65세 이상인구)	국토지리정보원		
	유동인구	민간기업		
3. 탄력배차제를 위한 노선 분석	교통카드 데이터	제주도/포항시		
4. 환승 편의시설 설치를 위한 정류장 위치 분석	교통카드 데이터	제주도/포항시		
5. 통행패턴(OD) 분석	교통카드 데이터	제주도/포항시		
6. 버스 중복도 분석	노선별 버스 정류장 데이터	제주도/포항시		

1. 민간+공공데이터 융복합을 통한 대중교통 서비스 표준분석 모델 표준화



1. 민간+공공데이터 융복합을 통한 대중교통 서비스 표준분석 모델 분석 프로세스

1. Pre Processing

1.1 데이터 수집

1.2 데이터 전처리

1.3 데이터 검증

2. 표준 모델 분석

2.1 교통카드 데이터

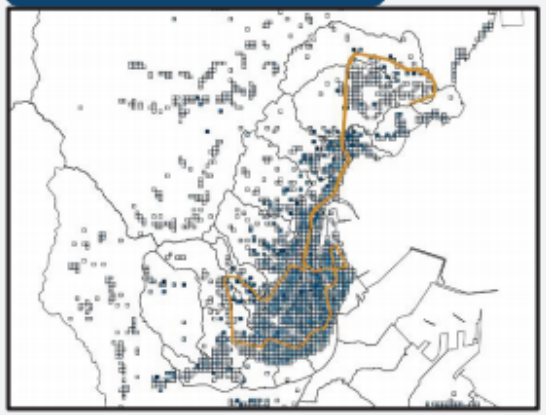
- 2.1.1 카드 현황 분석 (승차, 환승, 하차)
- 2.1.2 통행 패턴분석
- 2.1.3 탄력배차제 분석
- 2.1.4 환승 편의 시설 입지 분석

2.2 버스정류장/주거 인구 /유동 인구 데이터

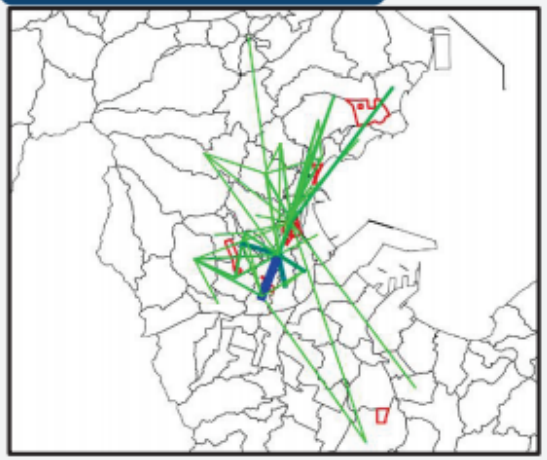
- 2.2.1 대중교통 사각지대 분석
- 2.2.2 저상버스 도입노선 분석
- 2.2.3 버스노선 중복도 분석

3. GIS 시각화

3.1 주제도 시각화



3.2 통행패턴 시각화



1. 민간+공공데이터 융복합을 통한 대중교통 서비스 표준분석 모델_하차정보 추정 예시

□ 분석 방법은 개인이 집에서 출발하여 다시 집으로 돌아온다는 가정하에서 하차 정류장을 추정함.

□ 하차 정류장 추정(포항시)

통행횟수	전체	하차 정류장 있음	하차 정류장 없음	하차정보 추정
1	4,436,338	304,695	4,131,643	304,695
2	6,912,302	534,718	6,377,584	6,912,302
3	1,475,721	192,124	1,283,597	1,475,721
4	469,856	69,016	400,840	469,856
5	94,510	17,305	77,205	94,510
6	21,936	4,288	17,648	21,936
7	4,809	1,121	3,688	4,809
8	1,296	337	959	1,296
9	639	140	499	639
10	150	66	84	150
11	33	0	33	33
12	36	1	35	36
13	26	23	3	26
15	15	13	2	15
합계	13,417,667	1,123,847	12,293,820	9,286,024
비율(%)		8.00	92.00	**69 .00**

□ 하차 정류장 추정(제주도)

통행횟수	전체	하차 정류장 있음	하차 정류장 없음	하차정보 추정
1	6,856,194	2,484,253	4,371,941	2,484,253
2	9,750,560	2,897,144	6,853,416	9,750,560
3	2,792,748	994,296	1,798,452	2,792,748
4	967,836	368,573	599,263	967,836
5	247,105	102,472	144,633	247,105
6	66,402	25,155	41,247	66,402
7	17,570	5,525	12,045	17,570
8	5,464	1,193	4,271	5,464
9	1,845	264	1,581	1,845
10	840	121	719	840
11	341	25	316	341
12	216	11	205	216
13	78	3	75	78
14	98	3	95	98
15	90	6	84	90
		이하 생략		
합계	20,707,506	6,879,049	13,828,457	16,335,565
비율(%)		33.22	66.78	**78.89**

□ 1번 통행에 대한 하차정보 결과에 따라 결과값이 다르게 나옴.

1. 민간+공공데이터 융복합을 통한 대중교통 서비스 표준분석 모델_Trip Chain 예시

- 분석 방법은 개인이 하루동안 대중교통을 이용한 트랜젝션 ID와 환승횟수를 기준으로 구축함

- TripChain 구성 목적
  - 승차 / 하차 / 환승 정류장 분석
  - 탄력 배차제 노선 분석
  - 환승편의 시설 입지 분석
  - 통행패턴(OD) 분석

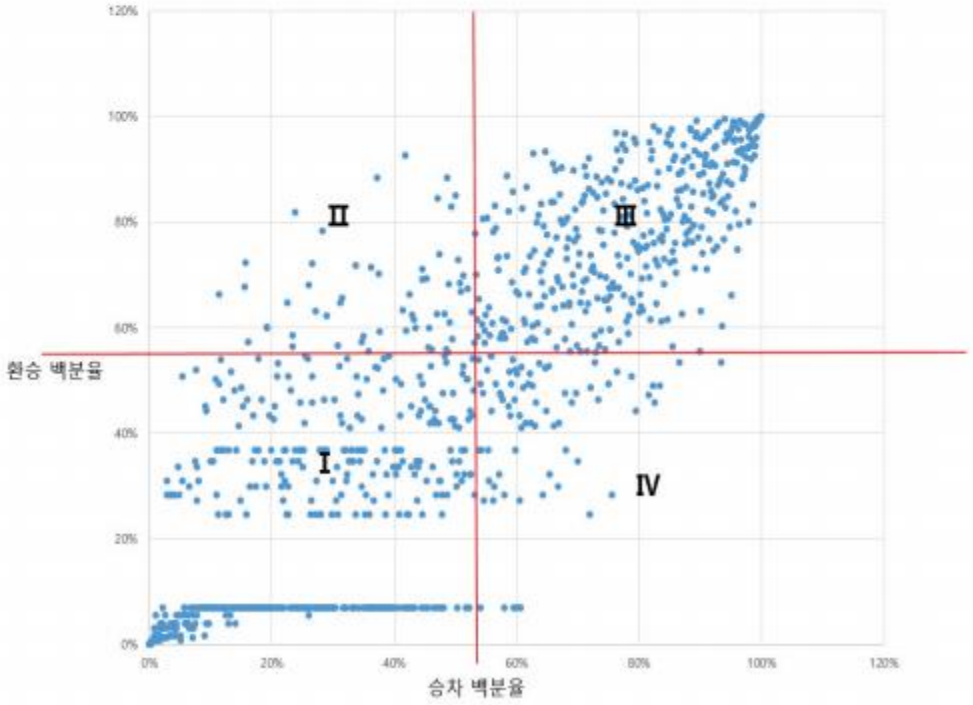
표준 원시 데이터
날짜
비식별교통카드ID
트랜젝션ID
환승횟수
교통 유형
버스노선ID
버스차량ID
승차일시
승차 정류장ID
승차 정류장 명칭
하차일시
하차 정류장ID
하차 정류장 명칭
성인 숫자
어린인 숫자
청소년 숫자
기타 숫자
거리
요금

TripChain 결과		
날짜	하차정류장명_1	하차 정류장_3
비식별카드ID	승차 시간_1	하차정류장명_3
성인 수	교통수단_1	승차 시간_3
어린이 수	버스차량_1	교통수단_3
청소년 수	버스노선_1	버스차량_3
기타 수	승차 정류장ID_1	버스노선_3
총 승객수	승차 정류장명_1	승차 정류장ID_3
총 요금	요금_2	승차 정류장명_3
총 거리	거리_2	요금_4
승차시간	하차시간_2	거리_4
교통수단	하차 정류장_2	하차시간_4
버스차량번호	하차정류장명_2	하차 정류장_4
버스노선번호	승차 시간_2	하차정류장명_4
승차정류장ID	교통수단_2	승차 시간_4
승차정류장명	버스차량_2	교통수단_4
요금	버스노선_2	버스차량_4
거리	승차 정류장ID_2	버스노선_4
요금_1	승차 정류장명_2	승차 정류장ID_4
거리_1	요금_3	승차 정류장명_4
하차시간_1	거리_3	하차시간
하차 정류장_1	하차시간_3	하차 정류장
		하차 정류장명

1. 민간+공공데이터 융복합을 통한 대중교통 서비스 표준분석 모델_환승 편의시설 위치선정

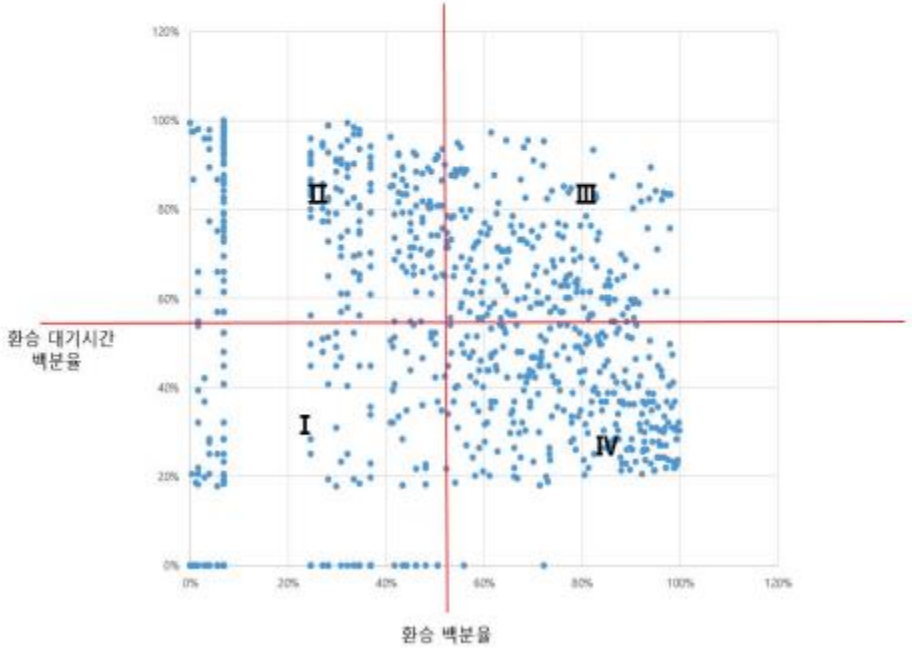
□ 승차인원과 환승 정류장 분석

- Ⅰ : 승차와 환승이 적은 경우
- Ⅱ : 승차는 적지만 환승이 많은 경우
- Ⅲ : 승차와 환승이 많은 경우
- Ⅳ : 승차는 많지만 환승이 적은 경우



□ 환승 정류장과 환승 대기 시간 분석

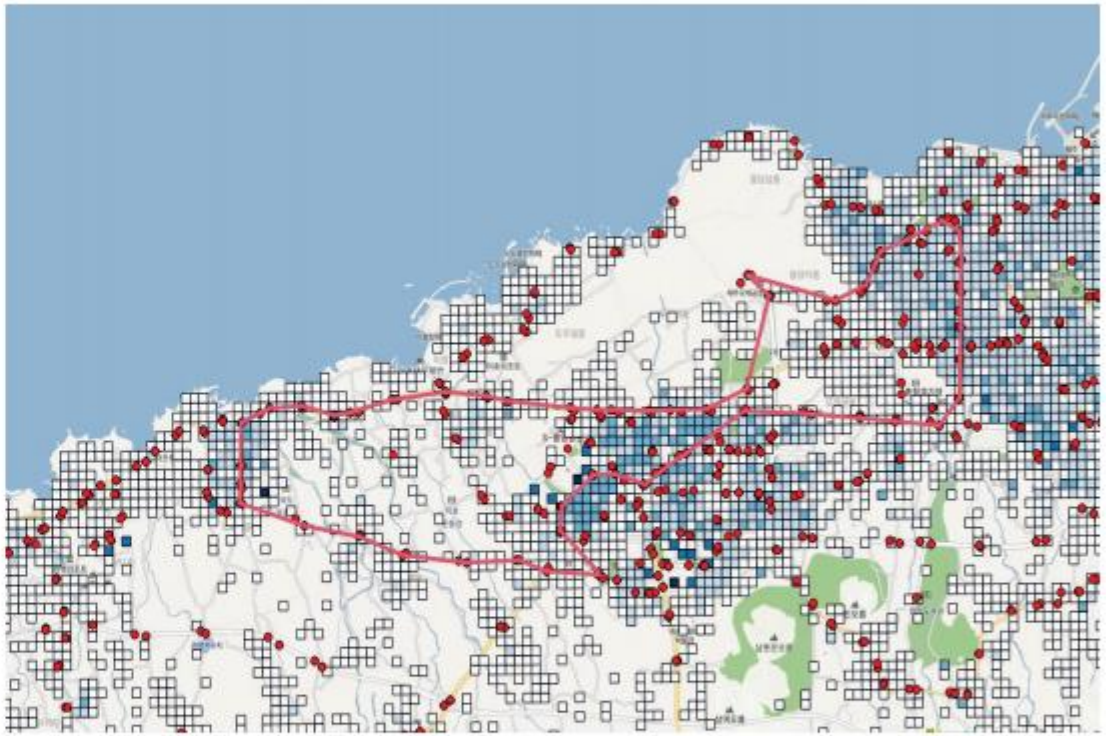
- Ⅰ : 환승 인원도 적고 환승 대기시간도 짧은 경우
- Ⅱ : 환승 인원은 적지만 환승 대기시간이 긴 경우
- Ⅲ : 환승 인원과 환승 대기시간이 긴 경우
- Ⅳ : 환승 인원은 많지만 환승 대기시간이 짧은 경우



1. 민간+공공데이터 융복합을 통한 대중교통 서비스 표준분석 모델 적용_탄력배차제 예시

□ 포항시

□ 제주도



□ 101번 요일별 승차 인원 및 주말 인원 비교

일	월	화	수	목	금	토	승차 인원 (A)	주말 인원 (B)	비율 (B/A)
44,276	81,100	82,087	83,706	82,586	85,139	66,354	525,248	110,630	21.06

□ 1009-1번 요일별 승차 인원 및 주말 인원 비교

일	월	화	수	목	금	토	승차 인원 (A)	주말 인원 (B)	비율 (B/A)
526	23	37	51	0	31	451	1,119	977	83.31

1. 민간+공공데이터 융복합을 통한 대중교통 서비스 표준분석 모델 적용_통행패턴 예시

□ 법정동 단위 OD(샘플 데이터)

Destination(종점) →

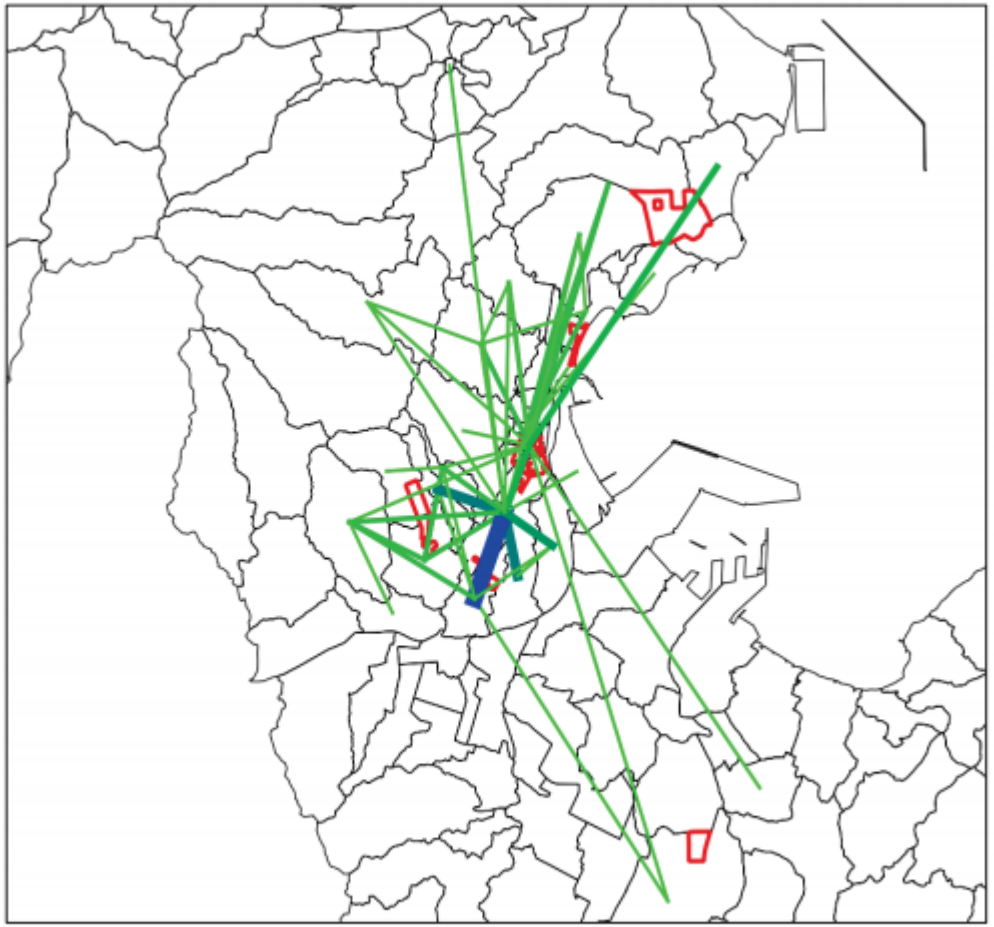
Origin(기점) ↓

법정동	괴동동	대신동	대잠동	대흥동	덕산동	덕수동
괴동동	0	8	5	1	6	3
대신동	7	0	2590	135	1035	444
대잠동	5	2844	0	2351	2010	914
대흥동	1	98	2205	0	9	22
덕산동	9	166	2534	14	0	66
덕수동	2	409	1142	35	237	0

□ 통행패턴 분석 종류

- 법정동/리 또는 행정동 단위 분석
- 격자단위(300 X 300m) 분석
- 정류장 단위 분석

□ 법정동/리 단위 통행패턴



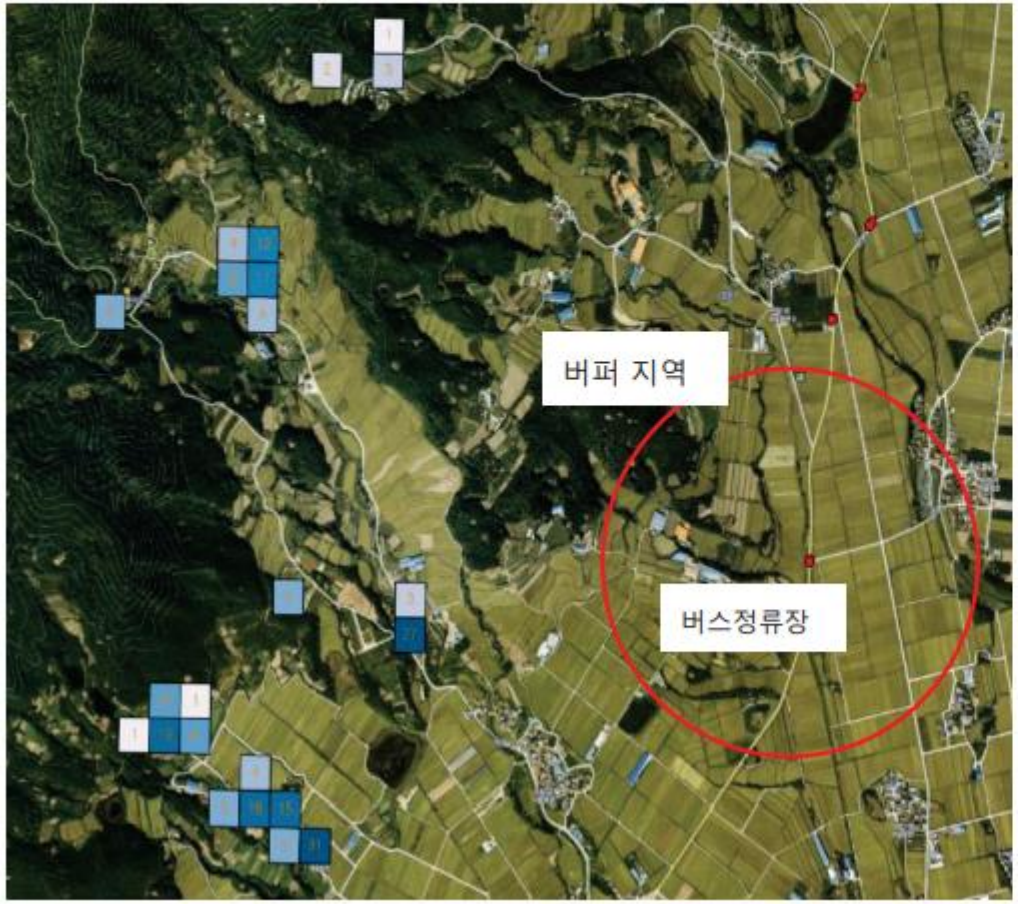
□ 선의 두께에 따라 통행량을 판단함.

1. 민간+공공데이터 융복합을 통한 대중교통 서비스 표준분석 모델_대중교통 사각지대 예시

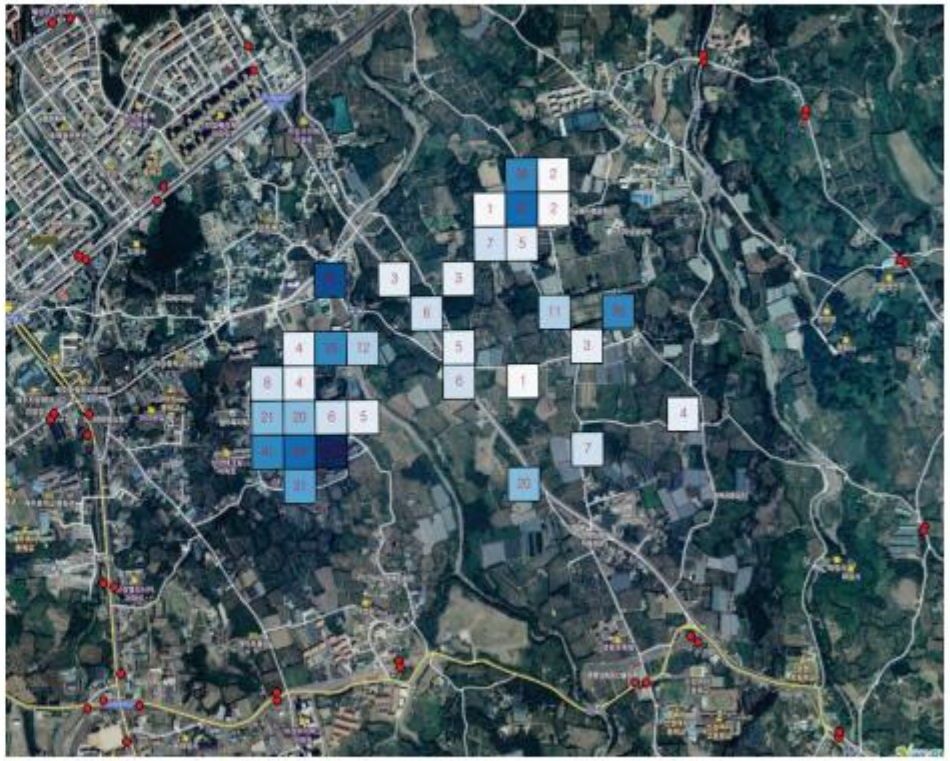
□ 분석조건

- 버퍼 영역 설정 조건은 법정동 단위은 500m, 읍면 지역은 1,000m 단위

□ 분석방법



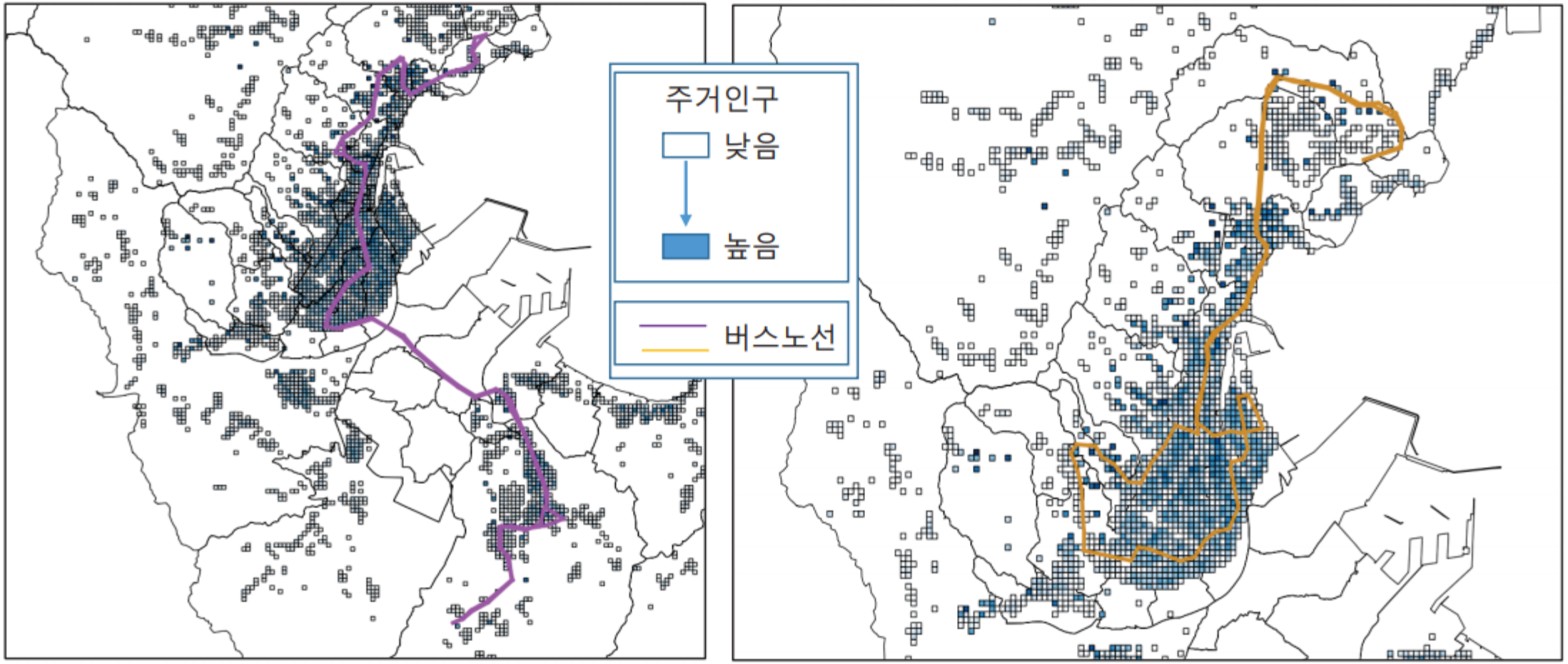
□ 시각화 방법



주거인구	정류장간 거리(m)	도로현황
555	900	연신로(편도2차선)

1. 민간+공공데이터 융복합을 통한 대중교통 서비스 표준분석 모델_저상버스 도입 예시

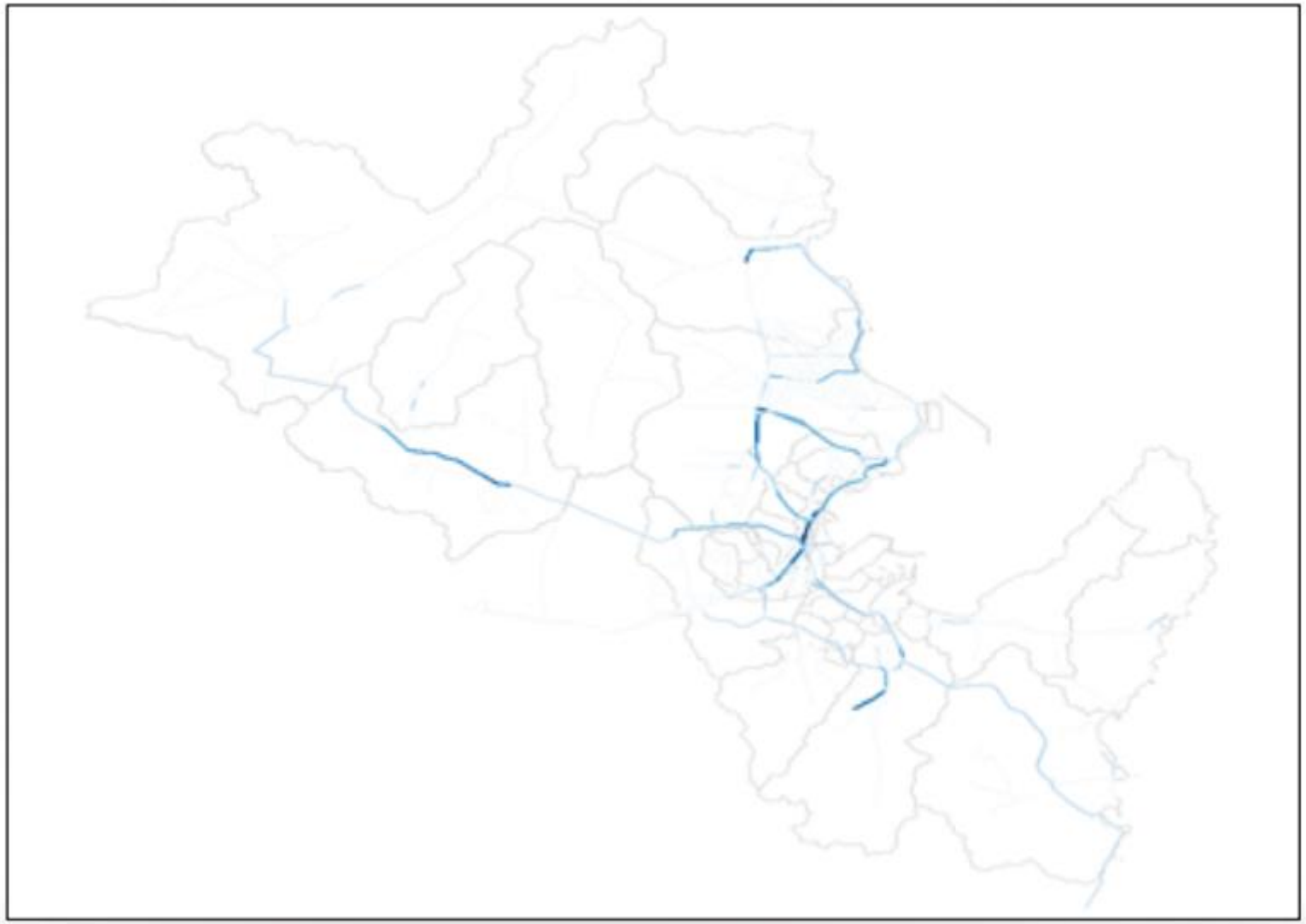
□ 저상버스 시각화 방법



1. 민간+공공데이터 융복합을 통한 대중교통 서비스 표준분석 모델_버스노선 중복도

□ 시각화 방법

- 정류장과 정류장 사이를 통과하는 버스 노선 수를 주제도로 작성함.



데이터 정의 (개인 단위 인구 데이터)

SKT 모바일 빅데이터의 개인 단위 인구 데이터는 비식별화를 완료한 데이터로서 개인을 특정할 수는 없지만 설정 조건에 따라 휴대폰 사용자의 시공간적 패턴을 추출할 수 있는 데이터 입니다.

원본
- 개인 단위 인구 데이터
- 전국 1일 약 100억건

추출범위
- 주요 상권 지역 및 인구 밀집 지역
- 지정 기간 (약 1주이내)

추출조건
- 분석 지역의 IN/OUT Time DATA 적용
- 체류 시간 기준에 따른 인구 이동 특성 분석

공간조건
- 분석 지역의 기지국 단위 DATA 적용
- 분석 지역 기지국 전파 영역 적용

적용사례
- O/D 동선 분석 적용
- 체류시간, 체류인구 등 분석 적용

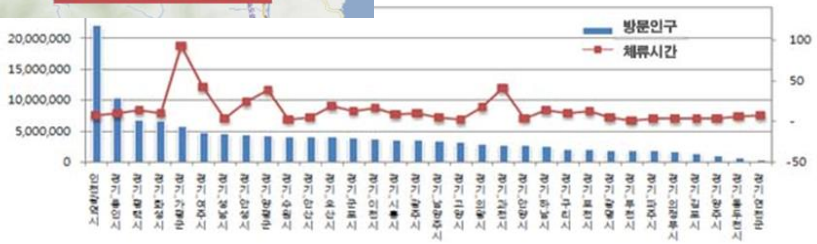
분석결과
- 1) 인구 이동에 따른 인구 데이터 추출
- 2) 인구 특성 데이터 추출 (성/연령, 체류시간 등)

- 개인 단위 인구 데이터는 모바일 빅데이터로 생산 가능한 가장 정밀한 인구 데이터로서 비식별화를 완료한 데이터이나 개인을 유추 할 수 있는 개연성이 있어 분석 완료된 통계성 집계 데이터만을 제공함.

거점별 체류 시간

단위 : 명

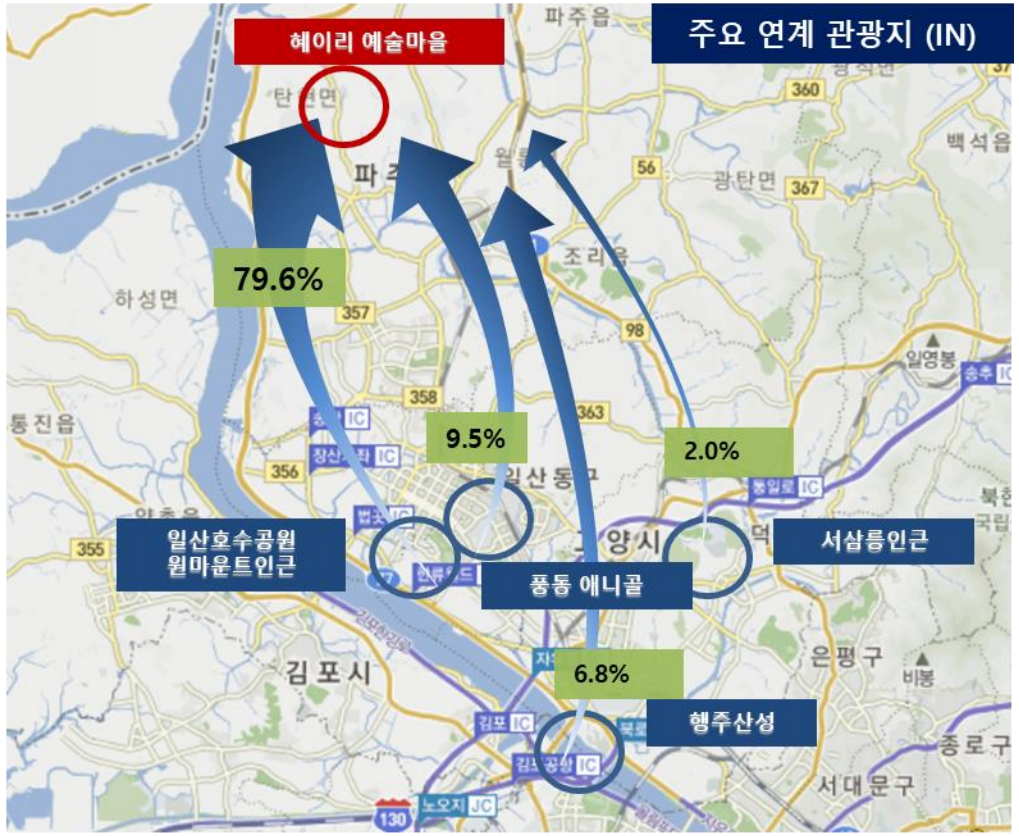
거점명	시군명	평균 체류일수
A거점	담양군	1.4
	함평군	1.5
	여수시	2.1
거점 평균 체류기간	1.8	
B거점	부산광역시	2.1
	양산시	1.5
	거제시	1.8
거점 평균 체류기간	1.9	



연계 상권 지역 이동 패턴 분석 (예시)

경기 서북부 관내/외 연계 상권 지역에 대한 인구 이동 패턴 분석을 통하여 주요 연계 지역을 파악하고, 이에 대한 관내 유입 방안과 정책 발굴을 위한 분석 데이터 제공

순위	연계 관광지(IN)	인구수	점유율
1	호수공원인근→헤이리 예술마을	3,033	79.6%
2	풍동애니골→헤이리 예술마을	360	9.5%
3	행주산성→헤이리 예술마을	258	6.8%
4	서삼릉_인근골프장→헤이리 예술마을	78	2.0%
5	서오릉→헤이리 예술마을	54	1.4%
6	고양향교인근→헤이리 예술마을	27	0.7%
7	배다골테마파크→헤이리 예술마을	2	0.1%
합 계		3,812	



계룡산 국립공원 탐방객 이동 동선 분석 (예시)

국립공원 탐방센터 인근 탐방객의 이동 동선으로 파악하고, 이에 따른 탐방로 주변의 시설물 계획 입안 및 탐방객 관련 정책 입안 자료로 활용하기 위한 분석 진행

이동동선	인구수	비율
동학사	72,219	66.3%
동학사>계룡산_정상>동학사	**29,268**	**26.9%**
동학사>계룡산_정상>갑사	**3,520**	**3.2%**
동학사>갑사	1,129	1.0%
동학사>상신	750	0.7%
동학사>수통골	487	0.4%
동학사>계룡산_정상>상신	292	0.3%
동학사>계룡산_정상>갑사>신원사	282	0.3%
동학사>갑사>신원사	263	0.2%
동학사>계룡산_정상>수통골	161	0.1%
동학사>계룡산_정상>상신>갑사	117	0.1%

< 분석 한계점 >

기지국 위치 기준의 개인 동선 분석은 이동 동선상의 기지국 존재 여부에 따른 그 정밀도가 결정되며, 비도심 지역의 기지국 조밀도가 낮은 경우 그 오차가 크게 발생할 수 있음.

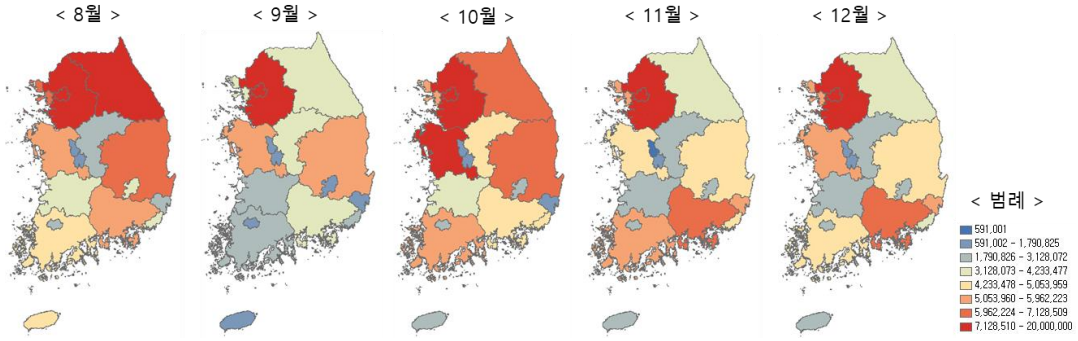
지역간 이동량 분석 (예시)

행정구역간, 특정 지역간 인구 이동량 분석을 통하여 교통 정책이나 지역간 연계성을 평가할 수 있는 기초 자료로 활용 가능
(ex. 교통 OD 분석, 관광지 연계성 분석, 상업지역간 연계성 분석 등에 활용)

O/D	서울	부산	대구	인천	광주	대전	울산	세종	.....	제주	합계
서울		5,328,217	2,170,382	11,908,960	2,114,398	1,901,924	1,656,665	682,012	.....	7,119,955	177,606,153
부산	5,963,757		2,548,914	1,902,165	543,622	713,188	8,392,583	134,237	.....	3,271,225	76,837,594
대구	2,563,784	2,885,894		933,356	94,717	444,068	1,013,840	111,452	.....	2,708,926	47,806,839
인천	12,096,658	1,123,126	434,541		381,826	520,424	306,477	142,094	.....	817,335	46,121,610
광주	3,394,796	631,143	110,831	1,367,002		475,436	141,603	165,718	.....	1,865,134	50,459,764
대전	5,158,885	784,816	441,051	1,740,164	350,073		384,101	1,859,544	.....	942,652	36,099,130
울산	1,854,895	3,415,210	1,016,684	656,203	102,883	323,633		73,226	.....	691,489	18,470,233
세종	2,630,564	390,887	158,651	725,405	282,562	1,467,544	203,668		.....	406,925	14,222,543
경기	104,403,721	5,280,472	2,422,391	22,975,752	1,900,931	2,149,079	1,534,719	883,055	.....	5,777,543	210,443,600
강원	11,941,463	1,076,810	729,665	4,253,263	294,028	530,499	414,856	89,309	.....	791,579	38,275,997
충북	6,300,012	896,235	2,023,515	2,973,344	517,090	1,147,703	289,499	978,755	.....	920,294	38,030,016
충남	8,340,717	1,160,397	764,366	4,826,216	666,823	3,974,016	419,547	1,524,217	.....	864,782	51,914,066
전북	8,158,867	888,174	553,754	3,787,768	3,124,112	2,127,752	279,331	364,037	.....	1,911,291	51,458,272
전남	5,614,965	1,589,201	312,424	1,874,658	9,752,878	664,007	297,950	161,887	.....	2,047,747	37,497,885
경북	6,872,288	4,573,350	12,458,765	3,821,633	615,895	1,252,075	1,947,200	147,598	.....	1,463,148	54,421,078
경남	4,214,109	15,326,456	3,474,415	1,068,116	313,792	528,373	23,945,352	82,171	.....	2,505,702	67,104,277
제주	2,816,472	710,735	467,340	338,417	629,356	88,553	170,347	23,959	.....		8,534,696
합계	192,325,953	46,061,123	30,087,689	65,152,422	21,684,986	18,308,274	41,397,738	7,423,271	.....	34,105,727	1,025,303,753

< 행정구역별 인구 이동량 분석 >

- 시도별, 읍면동별 인구 이동량 분석
- 개인 단위 인구 데이터 적용
- 특정 지역에 존재하는 인구가 이후 특정 시점에 어디로 이동했는지에 대한 분석을 통하여 공간별, 시간별 인구 이동의 특성을 분석한 데이터



< 분석 한계점 >

- 1일 기준 약 3Pb 이상의 빅데이터 적용
- 물리적 처리 시간이 많이 소요됨
- 개인을 추정할 수 있는 데이터로서 개인정보보호법에 의거 통계성 데이터만 제공 가능

관광객 이동 동선 분석 (예시)

개별 관광지를 방문하는 관광객의 이동 동선 분석을 통하여 관광객의 주요 경로를 분석하고, 관광지 개발로 인한 수요를 관내 관광지로 확산하는 정책의 기초 자료로 활용됨

남해군(연령층 주요 방문 관광지)
젊은층 : 남해편백자연휴양림, 상주은모래비치
장년층 : 해오름예술촌, 양모리학교
노년층 : 금산과 보리암, 이순신순국공원

관광지명	10대 점유율	20대 점유율	30대 점유율	40대 점유율	50대 점유율	60대 점유율	70대 점유율
가천다랭이마을	7%	9%	22%	26%	22%	11%	4%
금산과 보리암	3%	7%	14%	26%	33%	15%	3%
남해편백자연휴양림	10%	3%	27%	31%	16%	11%	2%
독일마을	6%	10%	20%	25%	25%	12%	4%
상주은모래비치	6%	11%	26%	24%	22%	9%	2%
양모리학교	6%	9%	22%	32%	20%	8%	2%
이순신순국공원	4%	7%	19%	27%	25%	11%	7%
해오름예술촌	9%	5%	28%	29%	14%	9%	6%

감사합니다.