

2021년 「국가교통조사」 최종보고서

# 모 바 일 교 통 빅 데 이 터 구 축 및 활 용

9

# 제 출 문

국토교통부장관 귀하

본 보고서를 「2021년도 국가교통조사」 최종보고서로 제출합니다.

2021년 12월

한국교통연구원

원장 오 재 학



**본 『2021년도 국가교통조사』는 다음 연구진에 의해  
수행되었습니다.**

## **참 여 연 구 진**

<b>&lt;한국교통연구원&gt;</b>	
연구책임자	<ul style="list-style-type: none"> <li>◦ 김주영 연구위원</li> </ul>
연 구 진	<ul style="list-style-type: none"> <li>◦ 최정민, 조종석, 천승훈, 조범철 연구위원</li> <li>◦ 박용일, 황순연, 장동익, 원민수 부연구위원</li> <li>◦ 김동호, 신영권 책임전문원</li> <li>◦ 김규진, 김정은 주임전문원</li> <li>◦ 강국수, 곽명신, 김관용, 김성민, 김운태, 김은미, 김 현, 박미란, 박준호, 백현진, 오연선, 이선아, 이슬기, 이채영, 이해선, 조은아 채정표, 홍성표 연구원</li> <li>◦ 강아라 연구조원</li> </ul>
<b>&lt;한국해양수산개발원&gt;</b>	
연 구 진	<ul style="list-style-type: none"> <li>◦ 이호춘 부연구위원, 최건우, 황수진 전문연구원, 박일란 선임사무관</li> <li>◦ 류희영 연구원</li> </ul>
<b>&lt;한국항공협회&gt;</b>	
연 구 진	<ul style="list-style-type: none"> <li>◦ 성인영 실장</li> <li>◦ 최인영 과장, 김창욱 대리</li> </ul>

# 『2021년도 국가교통조사』

## 보고서 구성 및 담당연구진

번 호	과 제 명	연 구 진
제 1권	요약보고서	최정민, 신영권, 박준호
제 2권	전국여객O/D 조사	조종석, 조범철, 최정민, 강국수, 박미란 채정표, 이슬기, 이선아, 백현진
제 3권	항공여객O/D 조사	한국항공협회
제 4권	전국 화물O/D보완갱신	김주영, 김정은, 오연선, 김운태
제 5권	전국 화물O/D조사 예비조사	김주영, 김정은, 오연선, 김운태
제 6권	해상화물O/D 보완갱신	한국해양수산개발원
제 7권	KTDB 플랫폼 기반지도 구축	김동호, 김관용
제 8권	차량 GPS 빅데이터 구축 및 활용	천승훈, 김성민, 이채영
제 9권	모바일 교통빅데이터 구축 및 활용	원민수, 조은아
제10권	국가교통통계DB구축	박용일, 곽명신
제11권	특별교통대책기간 통행실태조사	유한솔, 김은미, 우왕희
제12권	교통유발원단위 상세분석	황순연, 김현
제13권	국가교통물류경쟁력지표 조사연구	장동익, 홍성표

**『2021년도 국가교통조사』**  
**과제별 공동참여·위탁용역 사업자**

【공동사업 참여기관】
<ul style="list-style-type: none"> <li>• 전국 여객O/D 현행화 공동사업 (수도권 부문)               <ul style="list-style-type: none"> <li>- 경기연구원, 인천연구원, 서울연구원</li> </ul> </li> <li>• 전국 여객O/D 현행화 공동사업 (대구광역시권 부문)               <ul style="list-style-type: none"> <li>- 대구경북연구원</li> </ul> </li> <li>• 항공O/D 및 특성 조사               <ul style="list-style-type: none"> <li>- (사)한국항공협회</li> </ul> </li> </ul>
【위탁용역 사업자】
<ul style="list-style-type: none"> <li>• 개인통행실태조사               <ul style="list-style-type: none"> <li>- (주)컨슈머인사이트</li> </ul> </li> <li>• 여객교통시설물 이용실태조사               <ul style="list-style-type: none"> <li>- (주)서던포스트</li> </ul> </li> <li>• 교통량조사               <ul style="list-style-type: none"> <li>- (주)코리아데이터네트워크</li> </ul> </li> <li>• 고속도로 휴게소 조사               <ul style="list-style-type: none"> <li>- (주)동해엔지니어링</li> </ul> </li> <li>• 전세버스 조사               <ul style="list-style-type: none"> <li>- (주)서던포스트</li> </ul> </li> <li>• 도로 및 철도 교통분석용 네트워크 보완갱신               <ul style="list-style-type: none"> <li>- (주)올포랜드, (주)엔토포스</li> </ul> </li> </ul>

### 【위탁용역 사업자】

- 영업용 화물차 운행기록계 자료를 이용한 화물 기종점통행량 및 운행특성 분석
  - ㈜노트스퀘어
- 전국화물O/D조사 예비조사
  - ㈜코리아데이터네트워크
- 모빌리티 빅데이터 DB구축 및 온라인 서비스 유지보수
  - ㈜엔제로, ㈜큐빅웨어
- 모바일통신 원천 DB제공 및 구축
  - ㈜오픈메이트
- 모바일통신 데이터 가공 알고리즘 최적화 및 시스템 연결
  - ㈜오픈메이트온
- 특별교통통행실태조사 및 만족도 조사
  - ㈜리서치랩
- 빅데이터 관련 위탁용역 감리
  - ㈜악티보
- 국가교통조사 효율성 및 활용도 제고 방안 연구
  - 홍익대학교 산학협력단

## **최종보고서 목차**

- 제 1권 요약보고서**
- 제 2권 전국여객 O/D조사**
- 제 3권 항공여객 O/D 조사**
- 제 4권 전국화물 O/D 보완갱신**
- 제 5권 전국화물 O/D조사 예비조사**
- 제 6권 해상화물 O/D 보완갱신**
- 제 7권 KTDB 플랫폼 기반지도 구축**
- 제 8권 차량GPS 빅데이터 구축 및 활용**
- 제 9권 모바일 교통빅데이터 구축 및 활용**
- 제 10권 국가교통통계 DB구축**
- 제 11권 특별교통대책기간 통행실태조사**
- 제 12권 교통유발원단위 상세분석**
- 제 13권 국가교통물류경쟁력지표 조사연구**

# 목 차

## 요 약

제1장	과업의 개요 .....	3
제1절	배경 및 목적 .....	3
제2절	범위 및 내용 .....	5
제2장	모바일통신 빅데이터 가공알고리즘 개발 및 고도화 .....	9
제1절	모바일통신 빅데이터 가공알고리즘 주요 개념 .....	9
제2절	전처리 알고리즘 고도화 .....	11
제3절	체류 정보 및 통행목적 추정 알고리즘 고도화 .....	16
제4절	가공알고리즘 데이터 최적화 및 속도 개선 .....	25
제3장	모바일통신 빅데이터 기반 통행 DB 구축 및 검증 .....	31
제1절	기준년도 통행DB 구축 .....	31
제2절	구축 DB 검증 .....	42
제4장	모바일통신 빅데이터 기반 교통지표 개발 및 서비스 제공 .....	63
제1절	개요 .....	63
제2절	기반 통행 DB 재구조화 .....	65
제3절	교통지표 개발 및 분석 서비스 제공 .....	89
제5장	결론 및 차년도 수행계획 .....	117
제1절	결론 .....	117
제2절	차년도 수행계획 .....	118



## 표 목 차

〈표 1〉 모바일통신 빅데이터 기반 DB 형태 .....	6
〈표 2〉 모바일통신 빅데이터 DB 테이블 정의서 .....	8
〈표 3〉 체류지 식별 기준 .....	9
〈표 4〉 지역별 통행량 비교 결과 .....	12
〈표 5〉 지역별 출퇴근 통행량 비교 결과 .....	13
〈표 6〉 지역별 출퇴근 통행시간 비교 결과 .....	14
〈표 7〉 주제별 키워드 사전 조건 .....	24
〈표 8〉 주요 통행 구간 분석기능의 연도별 표출 형태 변경 .....	27
〈표 2- 1〉 체류지 유형별 식별 기준 .....	19
〈표 2- 2〉 개선 알고리즘 단계별 코드 구성 내용 .....	26
〈표 2- 3〉 전체 프로세스별 소요시간 .....	26
〈표 2- 4〉 산출물 상세 소요시간 .....	27
〈표 3- 1〉 모바일통신 빅데이터 DB 형태 .....	31
〈표 3- 2〉 모바일통신 빅데이터 DB 테이블 정의서 .....	33
〈표 3- 3〉 체류지 유형별 식별 기준 .....	37
〈표 3- 4〉 KT 시장점유율 .....	40
〈표 3- 5〉 지역별 통행량 비교 결과 .....	43
〈표 3- 6〉 시간대별 통행량 패턴 분석 (모바일통신 빅데이터 기준) .....	46
〈표 3- 7〉 요일별 통행량 패턴 분석 (모바일통신 빅데이터 기준) .....	49
〈표 3- 8〉 연령별 통행량 패턴 분석 (모바일통신 빅데이터 기준) .....	51
〈표 3- 9〉 지역별 출·퇴근 통행량 비교 결과 .....	54
〈표 3- 10〉 지역별 출퇴근 통행시간 비교 결과 .....	56
〈표 4- 1〉 과업의 내용적 범위 .....	63
〈표 4- 2〉 모바일통신 빅데이터 구조 .....	67
〈표 4- 3〉 교통폴리곤 기준 집계 통행량 컬렉션의 구성 .....	70
〈표 4- 4〉 읍면동 출발 기준 집계 통행량의 컬렉션의 구성 .....	73
〈표 4- 5〉 읍면동 도착 기준 집계 통행량의 컬렉션의 구성 .....	74
〈표 4- 6〉 읍면동 기준의 집계 통행량 컬렉션 구성 .....	75
〈표 4- 7〉 시군구 출발 기준 집계 통행량 컬렉션 구성 .....	76
〈표 4- 8〉 시군구 도착 기준의 집계 통행량 컬렉션 구성 .....	77
〈표 4- 9〉 시군구 기준의 집계 통행량 컬렉션 구성 .....	78
〈표 4- 10〉 시도 출발 기준의 집계 통행량 컬렉션 구성 .....	79
〈표 4- 11〉 시도 도착 기준의 집계 통행량 컬렉션 구성 .....	80
〈표 4- 12〉 시도 기준의 집계 통행량 컬렉션 구성 .....	81
〈표 4- 13〉 교통폴리곤 기준 통행량 DB 검증 항목 .....	86
〈표 4- 14〉 행정구역 단위 통행량 합계 비교 검증 항목 .....	87



〈표 4- 15〉 통행목적별 집계 통행량 합계 비교 검증 항목 .....	88
〈표 4- 16〉 주제별 키워드 사전 조건 .....	107
〈표 4- 17〉 주요 통행 구간 분석기능의 연도별 표출 형태 변경 .....	110

## 그림 차례

〈그림 1〉 모바일통신 빅데이터 가공알고리즘 .....	3
〈그림 2〉 통행자 실제 위치 추정 방법 예시 .....	4
〈그림 3〉 최장체류지 식별 후, 체류시간 aggregation 예시 .....	5
〈그림 4〉 통행지표 산출물 예시 .....	5
〈그림 5〉 출발/도착 지역 선택방식 변경(좌), 도착지 하위 행정구역 전체 선택(우) .....	16
〈그림 6〉 기종점 인구 통행 구간 분석의 시도단위 전국 분석 .....	17
〈그림 7〉 고령자 통행 비중 분석의 읍면동 단위 분석 결과 표출 .....	17
〈그림 8〉 고령자 주요 통행 구간 분석의 읍면동 단위 분석 결과 표출 .....	18
〈그림 9〉 분석 조건/결과 영역의 스크롤 이동으로 통행 거리/통행시간 조건 변경 .....	18
〈그림 10〉 출·퇴근 통행시간 분석 도구 화면 .....	19
〈그림 11〉 출·퇴근 통행량 분석 도구 화면 .....	20
〈그림 12〉 근무형태에 따른 통근 통행 분석화면과 전국 시군구별 지가지수 산점도 .....	20
〈그림 13〉 고용인구 변화에 따른 통근 통행 분석화면 .....	21
〈그림 14〉 출·퇴근 통행 연결성 분석화면 .....	21
〈그림 15〉 계절별 핫플레이스의 분석 결과 .....	22
〈그림 16〉 관광 특화지역 분석 결과의 상세 정보 톨팁 고도화 .....	22
〈그림 17〉 관광지 내·외부 통행비율 분석 결과의 UI 개선 전/후 비교 .....	23
〈그림 18〉 도심공동화 심각도 분석 결과의 UI .....	23
〈그림 19〉 분석 조건 값 UI 변경 .....	24
〈그림 20〉 범례 설정 UI 변경 .....	25
〈그림 21〉 분석기능 튜토리얼 UI .....	26
〈그림 22〉 주요 통행 구간 분석의 전국 분석 결과 및 통행량 분포도 그래프 .....	27
〈그림 23〉 정보 연계분석기능 화면 .....	28
〈그림 24〉 유입 기준 폴리곤 형태의 통행 변화 시뮬레이션 .....	29
〈그림 25〉 지역 내 기준 폴리곤 및 희망선 형태의 통행 변화 시뮬레이션 .....	29
〈그림 26〉 관광 특화지역 분석기능 화면 .....	30
〈그림 27〉 도심공동화 심각도 분석화면 .....	31
〈그림 1- 1〉 과업의 수행 절차 .....	6
〈그림 2- 1〉 모바일통신 빅데이터 가공알고리즘 .....	9
〈그림 2- 2〉 선분 이력 데이터로의 변환 예시 .....	12
〈그림 2- 3〉 시간적 데이터 보정 방법 예시 .....	13
〈그림 2- 4〉 통행자 실제 위치 추정 방법 예시 .....	14
〈그림 2- 5〉 통행자 실제 위치 추정 방법 예시 .....	15
〈그림 2- 6〉 시·공간적 데이터 보정(전처리) 적용 전, 후 통행 궤적 비교 .....	15
〈그림 2- 7〉 최장체류지 식별 예시 .....	16
〈그림 2- 8〉 최장체류지 이외의 체류지에 체류한 시간 중 가장 긴 시간 식별 예시 .....	16

〈그림 2- 9〉 첫 번째 주 체류지(최장체류지) 점유시간 산정 예시 .....	17
〈그림 2- 10〉 최장체류지 이외 체류지 중 최장체류지 식별 .....	17
〈그림 2- 11〉 두 번째 주 체류지 식별 및 점유시간 산정 예시 .....	18
〈그림 2- 12〉 통행시간 계산 및 잠재체류 식별 예시 .....	20
〈그림 2- 12〉 주체류지 기준 trip type 구분 .....	21
〈그림 2- 13〉 일별 데이터에 대한 주체류지 간 통행량 산정 예시 .....	21
〈그림 2- 14〉 잠재체류 식별 예시 .....	22
〈그림 2- 15〉 Data Lagging 예시 .....	23
〈그림 2- 16〉 통행지표 산출물 예시 .....	24
〈그림 3- 1〉 선분 이력으로의 변환 예시 .....	34
〈그림 3- 2〉 체류 순서 부여 (개인별 이동궤적 형성) .....	35
〈그림 3- 3〉 이동정보와 체류 정보 구분 (예시) .....	36
〈그림 3- 4〉 체류지 유형 구분 (예시) .....	37
〈그림 3- 5〉 위치정보, 시간정보, 연령 정보 변환 (예시) .....	38
〈그림 3- 6〉 출발/도착 구분 및 통행량 집계 (예시) .....	39
〈그림 3- 7〉 기지국 기반 데이터의 통행시간 산출 방법 .....	41
〈그림 3- 8〉 지역별 통행 발생량 비교 결과 .....	44
〈그림 3- 9〉 지역별 통행 도착량 비교 결과 .....	44
〈그림 3- 10〉 시간대별 통행 발생량 패턴 분석 (모바일기지국데이터 기준) .....	47
〈그림 3- 11〉 시간대별 통행 도착량 패턴 분석 (모바일통신 빅데이터 기준) .....	47
〈그림 3- 12〉 요일별 시간대에 따른 통행 발생량 변화 (모바일통신 빅데이터 기준) .....	49
〈그림 3- 14〉 요일별 시간대에 따른 통행 도착량 변화 (모바일통신 빅데이터 기준) .....	50
〈그림 3- 14〉 연령별 통행 발생량 패턴 분석 (모바일통신 빅데이터 기준) .....	52
〈그림 3- 15〉 연령별 통행 도착량 패턴 분석 (모바일통신 빅데이터 기준) .....	52
〈그림 3- 16〉 지역별 출·퇴근 통행량 분포 비교 결과 .....	55
〈그림 3- 17〉 시도별 통근 통행시간 비교(단위 : 분) .....	57
〈그림 3- 18〉 통행시간 기준별 통근 인구 비율 비교 .....	58
〈그림 3- 19〉 출근 시간 밀도 그래프 .....	59
〈그림 3- 20〉 퇴근 시간 밀도 그래프 .....	59
〈그림 3- 21〉 출근 속도 밀도 그래프 .....	60
〈그림 3- 22〉 퇴근 속도 밀도 그래프 .....	60
〈그림 4- 1〉 모바일통신 빅데이터 기반 분석 프로세스 .....	64
〈그림 4- 2〉 분석 레이어 구축 프로세스 .....	65
〈그림 4- 3〉 모바일통신 빅데이터의 형식변환 및 거리 산출 .....	67
〈그림 4- 4〉 전처리 데이터 형식 .....	67
〈그림 4- 5〉 교통폴리곤 기준 통행량 데이터 구조 .....	69
〈그림 4- 6〉 집계 통행량 DB 구축 프로세스 .....	71
〈그림 4- 7〉 행정구역 단위의 출발/도착 기준 통행량 DB 구조 .....	72

〈그림 4- 8〉 분석 특성 집계 통행량 데이터 재구성 조건 및 DB 구조 .....	82
〈그림 4- 9〉 분석 특성(요일, 기간, 시간, 연령대) 테스트 조건 설명과 조건 설정 UI 설계 예시 .....	83
〈그림 4- 10〉 시간대 코드별 평일/주말 집계 통행량 데이터 정의서 예시 .....	83
〈그림 4- 11〉 계절별 시간대 코드별 평일/주말 집계 통행량 데이터 정의서 예시 .....	84
〈그림 4- 12〉 분석 특성(통행목적) 테스트 조건 설명과 조건 설정 UI 설계 예시 .....	84
〈그림 4- 13〉 분석 특성(통행목적) 집계 통행량 데이터 정의서 예시 .....	85
〈그림 4- 14〉 통행목적별 집계 통행량 데이터 조건 .....	85
〈그림 4- 15〉 교통폴리곤 기준 통행량 DB 검증 쿼리와 결과 예시(시도별 통행량 확인) ....	87
〈그림 4- 16〉 View-T 시스템 구성도 .....	89
〈그림 4- 17〉 심층분석 도구 화면 구성 .....	90
〈그림 4- 18〉 간편분석 도구 화면 구성 .....	91
〈그림 4- 19〉 View-T Light 화면 구성 .....	92
〈그림 4- 20〉 출발/도착 지역 선택방식 변경(좌), 도착지 하위 행정구역 전체 선택(우) ....	93
〈그림 4- 21〉 기종점 인구 통행 구간 분석의 시도단위 전국 분석 결과 표출 .....	94
〈그림 4- 22〉 고려자 통행 비중 분석의 읍면동 단위 분석 결과 표출 .....	95
〈그림 4- 23〉 고려자 주요 통행 구간 분석의 읍면동 단위 분석 결과 표출 .....	96
〈그림 4- 24〉 분석 조건/결과 영역의 스크롤 이동으로 통행 거리/통행시간 조건 변경 .....	97
〈그림 4- 25〉 출·퇴근 통행시간 분석 도구 화면 .....	98
〈그림 4- 26〉 출·퇴근 통행량 분석 도구 화면 .....	99
〈그림 4- 27〉 근무 형태에 따른 통근 통행 분석화면과 전국 시군구별 지가지수 산점도 ...	100
〈그림 4- 28〉 고용인구 변화에 따른 통근 통행 분석화면 .....	101
〈그림 4- 29〉 출·퇴근 통행 연결성 분석화면 .....	102
〈그림 4- 30〉 계절별 핫플레이스의 계절/기간 선택 .....	103
〈그림 4- 31〉 계절별 핫플레이스의 분석 결과 .....	103
〈그림 4- 32〉 관광 특화지역 분석 결과의 상세 정보 톨팁 고도화 .....	103
〈그림 4- 33〉 관광지 내·외부 통행 비율 분석 결과의 UI 개선 전/후 비교 .....	104
〈그림 4- 34〉 도심공동화 심각도 분석 결과의 UI .....	105
〈그림 4- 35〉 간편분석 도구 화면 디자인 개선 전/후 비교 .....	105
〈그림 4- 36〉 지역 선택방식 변경 .....	106
〈그림 4- 37〉 선택한 분석 지역이 분석 주제에 동적으로 반영 .....	106
〈그림 4- 38〉 분석 조건 값 UI 변경 .....	107
〈그림 4- 39〉 범례 설정 UI 변경 .....	108
〈그림 4- 40〉 분석기능 튜토리얼 UI .....	109
〈그림 4- 41〉 주요 통행 구간 분석의 전국 분석 결과 및 통행량 분포도 그래프 .....	110
〈그림 4- 42〉 정보 연계분석 기능 화면 .....	111
〈그림 4- 43〉 유입 기준 폴리곤 형태의 통행변화 시뮬레이션 .....	112
〈그림 4- 44〉 지역 내 기준 폴리곤 및 희망선 형태의 통행변화 시뮬레이션 .....	112

〈그림 4- 45〉 관광 특화지역 분석 기능 화면 .....	113
〈그림 4- 46〉 도심공동화 심각도 분석화면 .....	114

요 약

---



## 요 약

### 1. 과업의 개요

#### 가. 과업의 배경 및 목적

- 최근 교통 분야에서 각종 모빌리티 정보(예를 들어, 차량GPS 데이터, 대중교통카드 데이터, DTG 데이터, 모바일통신 빅데이터 등)를 활용하여 개별 통행정보를 추출하고 각종 통행 관련 분석이 가능해지면서 다양한 방법으로 활용되고 있으나, 각 데이터의 특성에 따라 분석 대상과 범위가 제한적임
- 그러므로, 전 국민의 95% 이상이 사용하고, 이용 교통수단에 구매받지 않고 구독이 가능한 모바일통신 빅데이터를 이용한 개별통행DB 구축 및 활용이 요구되고 있음
  - 모바일 기기와 기지국 간의 송·수신 이력인 모바일 통신 빅데이터는 대부분의 국민이 이용하고 있는 모바일 기기를 통해 축적되기 때문에 표본율이 매우 높고, 이동수단에 구매받지 않고(항공 제외), 기기의 전원을 끄지 않는 이상 개인의 이동 궤적이 상세히 기록되기 때문
  - 1년 365일 수집되는 모바일통신 빅데이터는 사회적 현안에 즉각적으로 대응할 수 있는 시의성 있는 모빌리티 정보를 제공할 수 있음
- 그러나, 기존 민간과 공공에서는 다양한 방식으로 모바일통신 빅데이터를 가공하여 유동인구 형태로 활용하고 있지만 개별통행 행태를 분석하기에는 한계점이 있으며, 이를 극복하기 위한 기술적 DB 구축 및 분석 방법이 요구되고 있음
- 모바일통신 빅데이터 중 기지국 기반의 데이터를 이용하여 시의성 있는 O/D 기반의 통행정보(통행량, 통행시간)생성 방법을 정의하고 DB 구축 및 제공
- 통행자별 이용 교통정보 정보가 포함된 기지국 기반 데이터를 이용하여 통행DB를 구축하고, 이를 기반으로 교통수요를 분석할 수 있는 지표, 기능 등을 개발하여 고도화된 교통 정책 수립 지원에 활용하고자 함



## 나. 과업의 범위

- 시간적 범위
  - 모바일통신 빅데이터 가공알고리즘 개발: 2020년 원천데이터
  - 2020년도 기준 기반 DB 구축: 2020년 1월 1일 ~ 2020년 12월 31일 (1년)
  - 웹 서비스를 위한 각종 분석도구 개발: 2019년 1월 1일 ~ 2019년 12월 31일
    - 2020년도 연구 성과물인 2019년도 기준 기반 DB 활용
- 공간적 범위: 전국

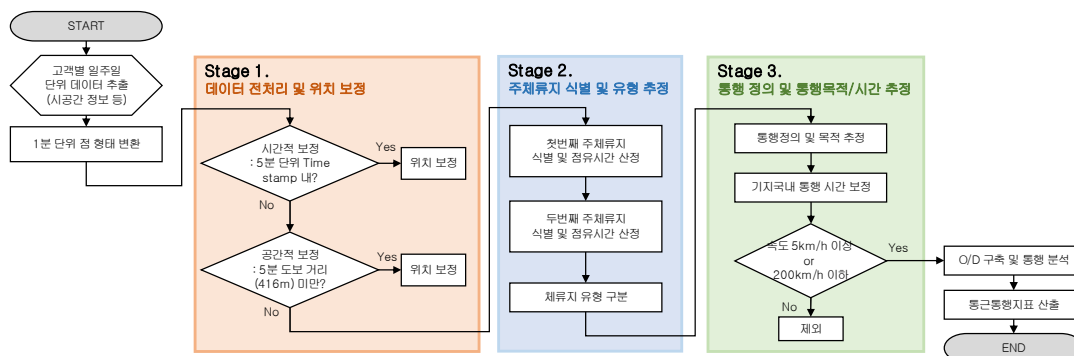
## 다. 과업의 내용

- 모바일통신 빅데이터 가공알고리즘 개발 및 고도화
  - 전처리 알고리즘 고도화
  - 체류정보 및 통행목적 추정 알고리즘 고도화
  - 가공알고리즘 데이터 최적화 및 속도 개선
- 모바일통신 빅데이터 기반 통행 DB 구축 및 검증
  - 기준년도 통행DB 구축
  - 기 구축된 통행DB 검증
- 모바일통신 빅데이터 기반 교통 지표 개발 및 서비스 제공
  - 주요 교통지표 생성 및 분석
  - 분석 서비스 개발 및 제공
  - 기반DB 현행화 및 속도 개선

## 2. 모바일통신 빅데이터 가공알고리즘 개발 및 고도화

### 가. 가공알고리즘 주요 개념

- 입력데이터(input data)로 각 통행자에 대한 7일 치 원천데이터가 사용되며 각 통행자에 대한 개별통행 특성을 기반으로 주체류지를 식별하고 통행지표를 산출
- 데이터 전처리 단계에서 신호 이상 현상 및 신호 이상치를 정형화할 필요 없이 데이터를 시·공간적으로 보정함으로써 비현실적으로 발생하는 통행자의 이동 궤적을 보정
- 기지국 기반 데이터 기록빈도 및 체류시간을 기반으로 통행자의 개별 통행 특성을 고려한 주체류지 식별을 수행하기 때문에 정기 출퇴근 통행뿐만 아니라 야간 근무자를 포함한 비정기 출퇴근 통행도 식별 가능
- 개별 통행자에 대한 주체류지를 식별한 후, 해당 위치에서의 체류시작시간과 체류종료시간을 기반으로 통근 시간 및 통행량 등 통행지표 산출

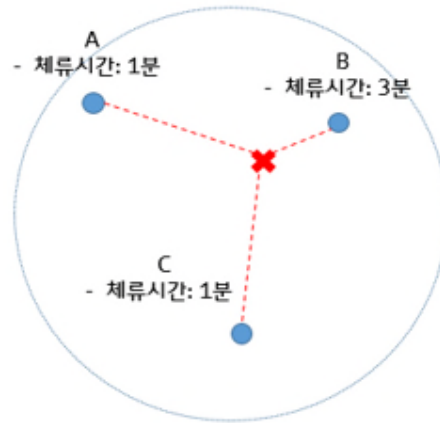


<그림 1> 모바일통신 빅데이터 가공알고리즘

### 나. 전처리 알고리즘 개발

- 적절한 공간적 해상도를 바탕으로 통행자의 통행속도를 고려한 이동 궤적을 식별하기 위해서 5분이라는 기준 시간을 정하고 해당 시간대에 기록된 기지국의 기록빈도를 기반으로 통행자의 실제 위치를 추정하고자 함
  - - 짧은 시간 동안 발생한 핑퐁 핸드오버 현상에 의한 체류지 식별률 저하 및 비현실적으로 기록된 이동궤적을 보정하기 위하여 위해 특정 시간 단위(5분)별로 데이터를 분할
- 데이터를 분할 한 후, 각 시간 단위(time slice) 내에 기록된 기지국 데이터의 기록빈도를 기반으로 통행자의 실제 위치 추정

- 통행자의 실제 위치 추정 방법: 각 기지국에서의 체류시간을 기반으로 가중치를 주어 통행자 위치 좌표(X)를 계산



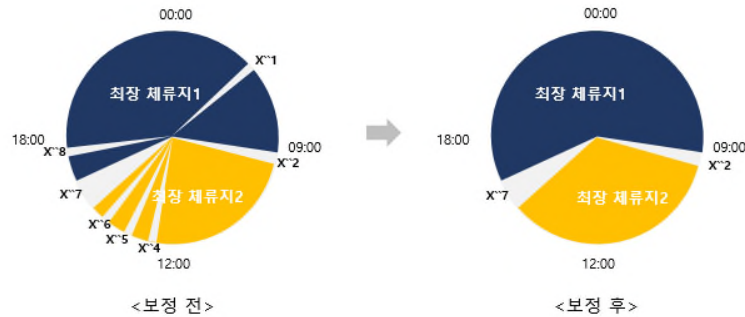
<그림 2> 통행자 실제 위치 추정 방법 예시

- 보행자 평균 도보 속도(5km/h)를 기준으로 5분(시간적 데이터 보정 시 설정한 시간 단위) 거리(약 416m) 범위 내 위치한 체류지들에 대해 각 체류지에서의 체류시간을 기반으로 해당 시간대에서의 통행자 실제 위치를 추정하고자 함
- 원천데이터에 대해 시·공간적으로 신호 이상 기록을 보정해줌으로써, 분 단위로 연속해서 변화하거나 좁은 공간적 범위 내에 산발적으로 기록되는 기지국 위치 보정 가능

#### 다. 체류정보 및 통행목적 추정 알고리즘 개발

- 데이터 전처리 단계에서 추정한 통행자의 실제 위치를 기준으로 통행자가 일주일 동안 각 위치에 몇 분 체류하였는지 계산 후, 최장체류지 식별
- 통행자가 최장체류지 이외의 체류지들에서 연속적으로 체류한 시간을 일별로 계산하여 그중에서 가장 긴 시간을 식별
- 해당 시간 이외에는 통행자가 최장체류지에 체류한 것으로 지정하며, 해당 시간에 속하는 체류지들에 대하여 일주일 기준 체류시간이 가장 긴 체류지 식별(2번째 최장 체류지 식별 시, 최장체류지로 지정된 체류지들을 후보에서 제외하고 식별)
- 일별로 2번째 최장체류지가 처음 기록된 시각(t1)과 마지막으로 기록된 시각(t2)를 식

별한 후,  $t_1$ 부터  $t_2$ 까지 통행자가 2번째 최장체류지에 체류한 것으로 보정(단,  $t_1$ 부터  $t_2$  사이에 1번째 최장체류지로 지정된 체류지가 존재하는 경우, 2번째 최장체류지가 없는 것으로 판단)



<그림 3> 최장체류지 식별 후, 체류시간 aggregation 예시

- 2개의 최장 체류지에 대해서 일주일 중 발생 빈도를 계산하여, 거주지 및 근무지 식별
  - 발생 빈도가 더 많은 것을 거주지, 적은 것을 근무지로 식별하며 발생 빈도가 같은 경우, 일주일 기준 체류시간이 더 긴 것을 거주지, 적은 것을 근무지로 식별
- 근무지로 식별된 체류지에 대해 일주일 중 근무지에 체류한 시간대를 계산하여 동일 시간대에 주 2회, 회당 3시간 미만 체류한 경우는 근무지에서 제외
- 거주지와 근무지 위치는 각 주체류지로 구분된 최장체류지의 위치로 지정
- 원천데이터에 대해 신규 알고리즘을 적용하면 <그림 4>와 같은 형태의 통행지표가 산출되며, 해당 산출물에는 개별 고객에 대한 일자별 통행지표가 포함되어 있음

날짜	2019-04-01
고객식별번호	112233
나이	30분
성별	M
출근시작시간	8:30
출근종료시간	9:00
출근소요시간	30분
출근 중 잠재체류 횟수	0회
출근 중 잠재체류 시간	0분
퇴근시작시간	18:27
퇴근종료시간	19:30
퇴근소요시간	63분
퇴근 중 잠재체류 횟수	1
퇴근 중 잠재체류 시간	35
집 위치	127.491, 36.455
회사 위치	127.485, 36.607
집-회사 간 거리	16.9 km
출근시간	30분
퇴근시간	X
출근속도	33.8 km/h
퇴근속도	X

<그림 4> 통행지표 산출물 예시

### 3. 모바일통신 빅데이터 기반 통행 DB 구축 및 검증

#### 가. DB 설계

- 「개인정보보호법」과 「위치정보의 보호 및 이용 등에 관한 법률」에 따라 개인의 위치와 이동궤적을 추적할 수 없도록 기반 DB는 다음 <표 14-1>과 같이 집계된 형태로 구축하고자 함

<표 1> 모바일통신 빅데이터 기반 DB 형태

출발						도착				
날짜	요일*	시간	폴리곤	시도	체류지 유형**	날짜	시간	폴리곤	시도	체류지 유형**
20180402	월	13	48521351	11	집	20180402	14	65421584	11	회사
20180403	화	06	54754213	21	집	20180403	08	32158421	11	회사
20180404	수	09	15486211	33	집	20180404	09	15486213	33	학교

총통행량	보정계수	성별	연령	잠재통행 고려 안함								
				통행량(비율)				시간			거리 평균	속도 평균
				통행 비율	30분 이상	60분 이상	120분 이상	평균	분산	중앙값		
10	0.34	M	20	1	0.7	0.5	0.4	35	6.98	26	8.24	29.8
25	0.57	F	50	0.8	0.9	0.8	0.2	33	6.77	24	9.45	11.9
8	0.47	M	10	1	0.9	0.6	0.1	25	4.32	19	10.5	23.5

잠재통행 30분 고려 안함									
통행량(비율)					시간			거리 평균	속도 평균
통행비율	잠재체류 횟수	30분 이상	60분 이상	120분 이상	평균	분산	중앙값		
0.8	0	0.8	0.6	0.8	35	6.98	26	5.89	36.9
0.7	3	0.7	0.5	0.2	-	-	-	9.57	10.4
0.8	2	0.8	0.5	0.1	-	-	-	10.12	20.3

잠재통행 60분 고려 안함									
통행량(비율)					시간			거리 평균	속도 평균
통행비율	잠재체류 횟수	30분 이상	60분 이상	120분 이상	평균	분산	중앙값		
0.9	0	0.7	0.6	0.5	35	6.98	26	7.4	33.6
0.8	2	0.8	0.6	0.2	-	-	-	9.9	10.9
0.9	0	0.9	0.6	0.1	25	4.32	19	10.1	22.8

주 \*) 요일: 출발 및 도착 요일 (1~7, 일~월)

주 \*\*) 유형: 개인별 출발지 및 도착지 유형 (H: 집, C: 회사, S: 학교, R: 정기 잠재체류지, X: 기타 잠재체류지)

- 개인의 통행 정보가 노출되지 않도록 기점 정보(출발 일자, 출발 시간대, 출발 위치, 출발 체류지 유형), 종점 정보(도착 일자, 도착 시간대, 도착 위치, 도착 체류지 유형), 통행자 정보(성, 연령)를 기준으로 통행 정보가 같은 인구를 집계
- 일자 정보는 1일 단위로 구축하도록 함
- 출발 시간 및 도착 시간 정보는 집계된 인구가 3명 이하가 될 가능성을 줄이면서, DB 사용자에게 가능한 섬세한 통행 정보를 제공할 수 있도록 한 시간 단위로 입력하도록 함
- 출발지와 도착지 위치는 「위치정보의 보호 및 이용 등에 관한 법률」에 저촉되지 않도록 기지국 좌표 그대로 노출하지 않고 매칭 되는 교통폴리곤(교통폴리곤은 기지국의 수신범위를 고려하여 구축한 모바일통신 빅데이터 기반 DB 분석 단위) ID로 변환하도록 함
- 출발 및 도착 체류지 유형은 평소 통행자가 해당 위치에 얼마나 자주, 오래 머무르는 지를 추정할 수 있는 정보를 제공하기 위한 것으로, 「2019년 국가교통조사」에서 설정한 체류지 식별 기준에 따라 출발지(출발 기지국 좌표)와 도착지(도착 기지국 좌표)의 특성을 확인한 후 각각 체류지 유형을 구분하여 입력하도록 함
- 통행자에 대한 정보를 확인할 수 있도록 성, 연령 필드 구성하고, 통행 정보가 지나치게 세분되지 않도록 가능한 통행 특성이 유사한 계층을 묶어 10세 단위로 연령을 구분하여 입력하도록 함
- 통행량 필드에는 전술한 기준에 의해 집계된 인구(단말기 수)를 입력하되, 집계된 통행량이 3 이하의 값을 갖는 경우, 2장에서 개발한 3통행 미만 보정계수를 적용하여 추정된 통행량을 입력하도록 함
- 새로운 지표 개발을 위해 통행시간 필드를 추가
- 통행시간 필드에는 집계된 통행 정보를 기준으로 산출한 평균통행시간 값을 입력
- 5가지 유형으로 구분된 기종점 기준, 통행 도중 인식한 기지국에서의 체류시간에 대해 고려 안함, 60분으로 나누어 통행량 및 통행시간을 산출함

&lt;표 2&gt; 모바일통신 빅데이터 DB 테이블 정의서

컬럼 ID	컬럼명	Type	비고
o_base_ymd	출발 일자	string	yyyymmdd
o_base_dow	출발 요일	string	0 ~ 6 : 일요일 ~ 토요일
o_timezn_cd	출발 시간대	string	00~23 (1시간 단위)
o_polygon	출발 폴리곤ID	string	-
o_trip_type	출발 체류지 유형	string	H: 집, C: 회사, S: 학교, R: 정기 잠재체류지, X: 기타 잠재체류지
d_base_ymd	도착 일자	string	yyyymmdd
d_timezn_cd	도착 시간대	string	00~23 (1시간 단위)
d_polygon	도착 폴리곤ID	string	-
d_trip_type	도착 체류지 유형	string	H: 집, C: 회사, S: 학교, R: 정기 잠재체류지, X: 기타 잠재체류지
age_cd	연령	integer	0~110 (예: 10: 10세 이상 20세 미만)
sex_type_cd	성	string	M: 남성, F: 여성
travel_volume	총 통행량	integer	횟수 단위
travel_time_mean	통행시간_평균	integer	1분 단위
travel_time_med	통행시간_중앙값	integer	1분 단위
travel_time_var	통행시간_분산	integer	-
travel_distance	통행거리	integer	km 단위
travel_speed	통행속도	integer	km/h 단위

## 나. DB 구축 방법

- 원천데이터 수집 및 이상치 제거
  - 2020년 1월 1일부터 2020년 12월 31일까지 기록된 로그를 수집
  - 기지국 좌표가 Null값이거나 행정동 정보가 매칭되지 않는 데이터를 제거
- 선분 이력으로 변환
  - 포인트 단위의(밀리세컨 단위) 로그를 선분 단위의(분 단위) 로그로 변환
  - 닷(dot) 형태로 기록된 로그를 기록된 시간 순서대로 연결한 후, 최초 기록된 시간(이하 ‘체류시작시간’)과 마지막으로 기록된 시간(이하 ‘체류종료시간’)정보만 추출하여 단말기가 해당 기지국에 식별된 시간(이하 ‘체류시간’)을 산출
    - 체류시간은 체류종료시간에서 체류시작시간을 뺀 값이며, 초(sec) 단위는 생략
- 통행 정보 왜곡 데이터 보완
  - 데이터 보완 범위를 한정하기 위해 로그 기록을 한 달 단위로 구분

- 통행 정보 왜곡 데이터를 탐색하고 보완하는 범위를 1개월로 한정 (예: 4월 1일부터 4월 30일까지의 로그 기록을 연결하여 탐색하고 보완)
- 통행 정보를 왜곡할 수 있는 ‘필요 이상의 핸드오버 데이터(Unnecessary Handover)’, ‘핑퐁 핸드오버(Ping-pong Handover)로 인한 데이터’를 본 과업에서 개발한 전처리 기술을 통해 보정함
- 체류순서 부여
  - 로그 기록 일자, 체류시작시간, 체류종료시간을 기준으로 로그 기록 순서에 따라 개인별 데이터를 정렬하여 체류 순서를 구분
  - 고객식별번호(단말기 구분 코드)와 데이터 생성일자를 기준으로 KEY 값을 갖는 필드를 형성한 후(‘기준일-고객식별번호’), 같은 KEY 값을 갖는 데이터 내에서 정렬
- 체류 정보 추출
  - 이동 중에 기록된 로그 기록을 제외하고 체류 중에 기록된 로그 기록만을 추출
  - 체류시간이 15분을 초과하는 로그 기록만 추출
    - 트립타임 필드를 추가하여 체류시간이 15분 이하인 경우는 ‘경로(Pass-by)’로, 15분 초과인 경우는 ‘체류(Stay)’로 구분
- 체류지 유형 구분
  - 체류지 식별 기준(〈표 14-3〉 참조)에 따라 체류지 유형을 구분하여 필드 추가

**<표 3> 체류지 식별 기준**

체류지 유형		식별 기준			
		체류 특성		통행자 연령	비고
		체류 시간	체류 빈도		
주 체류지	집	알고리즘 기준 최장 체류지 후보 1, 2 중	일주일 기준 최다빈도 체류지	-	첫 번째 주 체류지
	회사	동시간대 3시간 이상	주 2회 이상	30세 이상 60세 미만	두 번째 주 체류지 중
	학교	동시간대 3시간 이상	주 2회 이상	20세 미만	두 번째 주 체류지 중
잠재 체류지	정기적	3시간 이상	주 2회 이상	-	두 번째 주 체류지 중 회사나 학교로 구분되지 않은 체류지 중
	비정기 적	3시간 이상	주 2회 미만	-	



- 위치정보, 시간정보, 연령 정보 변환
  - 개인의 이동궤적을 추적할 수 없도록 원천데이터에서 출발지와 도착지의 위치 보를 분석 맵의 교통폴리곤 ID로 변경한 후, 시간정보(체류시작시간, 체류종료시간)에서 분 정보를 생략한 시간대 필드를 생성하고, 1세 단위의 연령 정보를 10세 단위로 변경함
- 통행량 집계
  - 앞서 부여한 체류 순서에 따라 출발과 도착을 구분한 다음, 기점 정보(출발 일자, 출발 시간대, 출발 위치, 출발 체류지 유형), 종점 정보(도착 일자, 도착 시간대, 도착 위치, 도착 체류지 유형), 통행자 정보(성, 연령)가 동일한 로그 기록을 집계함
  - 마지막 체류지가 출발 정보로 구분되고, 도착 정보가 없는 경우 이동하지 않은 것으로 보고 '무통행'으로 간주함
- 통행량 보정
  - 통행량이 3 이하(0~3통행)인 경우 모두 3통행으로 변환한 후, 3통행 미만 보정계수(2장 참고)를 적용하여 통행량을 보정
  - 개인정보보호법에 의거, 3 통행 이하의 값은 실제 값이 아닌, 추정 값을 입력
  - 실제 통행 분석에 활용할 때에는 통신업체의 시장점유율을 기준으로 전수화 하여 사용
  - 과학기술정보통신부에 고시되고 있는 무선통신서비스 가입자 통계 정보 중 LTE 가입 정보를 활용하여 KT 시장점유율 산출 가능
- 평균통행시간 산출
  - 모바일통신 빅데이터는 기지국 기반 로그 기록이므로 통행시간 산출 시, 기종점 부분에서 통행시간이 하향계산되는 경우가 발생함
  - 이와 같은 문제를 해결하고자, 기지국의 평균 수신 범위를 계산하여 시도별 기지국 내 이동 시간 산출 후, 통행 시간을 보정함
  - 각 개인의 통행정보에서 통행시간(도착시간-출발시간, 분 단위)을 산출한 다음, 기종점 정보(기지국 위·경도 좌표)와 시도(행정구역)정보를 매칭해 기종점 지점에서 시도별 기지국 내 이동 평균 시간을 합산하여 최종 통행시간 계산함
  - 개별 통행에 대해 기점 정보(출발 일자, 출발 시간대, 출발 위치, 출발 체류지 유형), 종점 정보(도착 일자, 도착 시간대, 도착 위치, 도착 체류지 유형), 통행자 정보(성, 연령)가 동일한 로그 기록의 통행시간을 집계한 후 집계한 통행시간을 산술평균하여 산출함

#### 다. DB 검증 방법

- 기지국 기반의 데이터를 이용하여 구축된 모바일통신 빅데이터 통행DB의 신뢰성 및 정확성을 검증하기 위하여 기존 가구통행실태조사자료(2016년 기준, 한국교통연구원)와 인구총조사자료(2015년 기준, 통계청)를 활용
- 기지국 기반의 데이터를 이용하여 모바일통신 빅데이터 통행DB와 기존 가구통행실태조사자료와 인구총조사자료의 통행량, 통행시간 등을 비교·분석하고, 단순 비교할 수 없는 모바일통신 빅데이터 기반 상세지표(예를 들어, 상세 시간대별, 요일별, 성별, 연령별, 희망통근시간 등)는 통상적인 기대 통행패턴을 벗어나는지 확인
  - 지역별 통행 발생량·도착량을 가구통행실태조사자료와 상대적으로 비교하여, 유사한 패턴을 보이는지 확인
  - 시간대별, 요일별 통행 발생량·도착량을 확인하여, 통상적인 첨두시간 및 평일·주말 통행패턴이 나타나는지 확인
  - 성별, 연령별 통행 발생량·도착량을 확인하여, 통상적인 인구비율 및 연령대별 통행 패턴이 나타나는지 확인
  - 지역별 출퇴근 목적 통행량·통행시간을 가구통행실태조사자료, 인구총조사자료와 상대적으로 비교하여, 유사한 패턴을 보이는지 확인
  - 출퇴근 목적 통행시간·통행속도와 희망통근시간(60분)기준 통행패턴을 확인하여, 통상적인 통행패턴이 나타나는지 확인
- 단, 데이터 가공 시간 때문에 1년 치 모든 데이터를 가공하여 과업 기간 내에 검증할 수 없으므로, 통계적으로 충분히 유의한 10만 명의 표본 자료를 이용하여 검증을 시행함
  - 표본 유의성을 분석한 결과, 약 2만 명의 표본 이후 통행시간의 평균 오차는  $\pm 0.1$  분 이하임.

## 라. DB 검증 결과

### ○ 통행 발생·도착패턴 비교 및 분석 결과

<표 4> 지역별 통행량 비교 결과

(단위: 통행/일(비율%))

구분	가구통행실태조사자료 기준*		모바일통신 빅데이터 기준	
	발생 (비율)	도착 (비율)	발생 (비율)	도착 (비율)
서울	17,996,813 (20.7)	17,971,878 (20.7)	1,545,273 (19.5)	1,709,372 (21.6)
부산	6,457,538 (7.4)	6,458,293 (7.4)	513,449 (6.5)	500,869 (6.3)
대구	4,333,058 (5.0)	4,330,603 (5.0)	388,500 (4.9)	360,704 (4.6)
인천	4,370,318 (5.0)	4,433,014 (5.1)	405,912 (5.1)	364,511 (4.6)
광주	3,117,712 (3.6)	2,942,575 (3.4)	219,652 (2.8)	199,102 (2.5)
대전	3,091,862 (3.6)	3,101,487 (3.6)	233,980 (3)	223,273 (2.8)
울산	2,117,733 (2.4)	2,109,519 (2.4)	171,036 (2.2)	171,523 (2.2)
세종	389,864 (0.4)	391,571 (0.5)	49,631 (0.6)	44,047 (0.6)
경기	18,668,530 (21.5)	18,628,440 (21.4)	1,993,158 (25.2)	1,840,414 (23.3)
강원	2,714,015 (3.1)	2,715,277 (3.1)	264,098 (3.3)	275,009 (3.5)
충북	2,803,473 (3.2)	2,793,476 (3.2)	257,411 (3.3)	264,394 (3.3)
충남	3,871,861 (4.5)	3,863,535 (4.4)	322,793 (4.1)	341,216 (4.3)
전북	2,942,438 (3.4)	2,943,452 (3.4)	261,467 (3.3)	264,746 (3.3)
전남	2,606,282 (3.0)	2,783,178 (3.2)	253,020 (3.2)	274,287 (3.5)
경북	4,431,201 (5.1)	4,442,218 (5.1)	404,548 (5.1)	436,929 (5.5)
경남	5,297,159 (6.1)	5,300,768 (6.1)	489,631 (6.2)	495,935 (6.3)
제주	1,659,446 (1.9)	1,660,019 (1.9)	127,386 (1.6)	128,273 (1.6)
계	86,869,303 (100)	86,869,303 (100)	7,905,399 (100)	7,905,399 (100)

주 : \* 가구통행실태조사자료는 2016년 기준 승용차(택시), 버스, 철도(일반/고속/지하철), 항공, 해운수단에 대한 목적통행량임.

- 가구통행실태조사자료 기준 1인당 일평균 통행량(발생+도착)은 약 3.4회이며, 모바일 통신 빅데이터 기준 1인당 일평균 통행량(발생+도착)은 약 3.1회 임
- 가구통행실태조사자료 기준 전국 통행 발생·도착량은 1.74통행/일, 1.74통행/일이며, 모바일통신 빅데이터 기준 전국 통행 발생·도착량은 1.5통행/일, 1.6통행/일 임
- 모바일통신 빅데이터 기준 통행량과 가구통행실태조사자료 기준 통행량 매우 유사하게 나타남
- 하지만 가구통행실태조사자료가 더 높은 통행량을 띄는데, 이러한 차이는 기존의 가구통행실태조사는 출근, 등교, 업무, 쇼핑, 귀가 등 모든 통행목적에 따라 통행량을 집계하지만, 모바일통신 빅데이터는 체류시간 기준으로 체류지를 구분하기 때문에 기타잠재체류지에 속하지 않은 이동이 집계되지 않기 때문에 통행량차이가 과소추정되었다고 판단됨
- 지역별 통행량 비율을 살펴보면, 평균적으로는 약 0.41%의 차이로 기존 가구통행실태조사자료와 매우 유사한 패턴을 나타냄

## ○ 목적별(출퇴근) 통행량 비교 및 분석 결과

- 지역별 출퇴근 목적 통행 발생량·도착량을 가구통행실태조사자료·인구총조사자료와 상대적으로 비교하여, 유사한 패턴을 보이는지 확인함

&lt;표 5&gt; 지역별 출퇴근 통행량 비교 결과

(단위: 통행/일(비율%))

구분	가구통행실태조사자료 기준		인구총조사자료 기준	모바일통신 빅데이터 기준	
	출근 (비율)	귀가*	통근**	출근	퇴근
서울	4,516,466(20.7)	8,413,729(22.1)	4,597,322(19.6)	650,284(20.4)	547,269(19.6)
부산	1,492,300(6.8)	3,051,128(8.0)	1,500,101(6.4)	207,420(6.5)	180,879(6.5)
대구	974,423(4.5)	1,801,446(4.7)	1,064,227(4.5)	157,348(4.9)	137,939(4.9)
인천	1,213,843(5.6)	1,794,429(4.7)	1,344,311(5.7)	167,280(5.3)	145,001(5.2)
광주	838,601(3.8)	1,279,355(3.4)	639,412(2.7)	88,578(2.8)	77,006(2.8)
대전	736,843(3.4)	1,313,431(3.4)	672,856(2.9)	94,244(3.0)	82,383(3.0)
울산	514,493(2.4)	995,135(2.6)	538,654(2.3)	69,686(2.2)	59,901(2.1)
세종	89,203(0.4)	149,020(0.4)	92,403(0.4)	18,441(0.6)	16,517(0.6)
경기	5,272,517(24.1)	7,799,702(20.5)	5,745,634(24.4)	818,330(25.7)	711,400(25.5)
강원	644,578(2.9)	1,186,771(3.1)	703,266(3.0)	100,084(3.1)	90,309(3.2)
충북	637,895(2.9)	1,182,200(3.1)	742,399(3.2)	99,482(3.1)	90,888(3.3)
충남	875,177(4.0)	1,616,017(4.2)	1,032,913(4.4)	123,410(3.9)	113,776(4.1)
전북	665,597(3.0)	1,263,474(3.3)	811,899(3.5)	100,648(3.2)	91,070(3.3)
전남	706,911(3.2)	1,241,157(3.3)	864,498(3.7)	94,807(3)	86,754(3.1)
경북	1,007,670(4.6)	2,045,073(5.4)	1,280,135(5.4)	152,433(4.8)	141,449(5.1)
경남	1,372,291(6.3)	2,361,671(6.2)	1,560,582(6.6)	190,028(6)	171,842(6.2)
제주	291,636(1.3)	581,151(1.5)	309,133(1.3)	50,611(1.6)	44,769(1.6)
계	21,850,443(100)	38,074,889(100)	23,499,745(100)	3,183,114(100)	2,789,152(100)

주 : \*가구통행실태조사자료는 2016년 기준 출퇴근에 대한 목적통행량이며, 귀가통행 통행목적에 상관없이 집으로 향하는 모든 통행을 의미함

\*\*인구총조사(통계청)자료는 2015년도 기준이며, 출근과 퇴근이 구분되어 있지 않음

○ 목적별(출퇴근) 통행시간 비교 및 분석 결과

- 지역별 출퇴근 목적 통행시간을 가구통행실태조사자료·인구총조사자료와 상대적으로 비교하여, 유사한 패턴을 보이는지 확인함

<표 6> 지역별 출퇴근 통행시간 비교 결과

(단위: 분)

구분	가구통행실태조사자료 기준		인구총조사 자료 기준	모바일통신 빅데이터 기준	
	출근	귀가*	통근**	출근	퇴근
서울	34.8	30.1	45.7	32.6	35.1
부산	30.1	25.6	37.0	28.9	30.7
대구	29.5	28.2	33.4	26.6	28.2
인천	36.6	32.5	43.6	34.3	35.8
광주	20.8	18.3	30.2	24.9	26.7
대전	23.3	21.7	32.3	25.5	27.2
울산	25.6	24.5	28.6	26.3	27.9
세종	15.6	15	29.1	29.6	29.4
경기	41.7	36.3	41.5	35.3	36.9
강원	17.6	20.3	22.1	22.1	23.4
충북	17.9	17.8	25.6	24	25.6
충남	15.6	16.7	22.9	24.4	25.7
전북	16	15.4	23.2	23.4	25
전남	13.3	13.6	21.3	23.9	25.2
경북	18.8	20	22.1	23.4	24.5
경남	22.5	21.1	25.6	25.4	26.9
제주	20.8	21.5	23.6	26.6	28.9
계	23.6	22.3	35.2	29.8	31.5

주 : \*가구통행실태조사자료는 2016년 기준 출근·귀가 통행에 대한 일평균시간이며, 귀가통행 통행목적에 상관없이 집으로 향하는 모든 통행을 의미함

\*\*인구총조사자료는 2015년 기준 통근 통행에 대한 소요 시간별 자료를 일평균시간 형태로 재구성하였으며, 출근과 퇴근이 구분되어 있지 않음

## 4. 모바일통신 빅데이터 기반 교통 지표 개발 및 서비스 제공

### 가. 개요

- 2018년부터 교통 모니터링·데이터 제공·분석 플랫폼(View-T)를 통해 모바일 통신 빅 데이터를 기반으로 하는 통행 분석 서비스를 제공해왔음
- 해당 년도에는 2019년 기준 모바일통신 빅데이터 기반 통행 DB를 기반으로 통행분석 서비스를 제공하고, 사용자가 시스템에 구축된 기반 데이터를 이종 데이터와 연계하여 분석하거나 다른 방식으로 결과를 표출하는데 활용할 수 있도록 데이터 다운로드 서비스 제공함

### 나. 모바일통신 빅데이터 분석 및 구축 프로세스

- 개별 사람의 시공간적 움직임이 휴대폰 신호를 통해 인근 기지국에 송신되어 수집되는 데이터로 가상기지국 단위로 정보가 수집되고 있으나, 데이터 제공 시에는 개인정보 보호법과 위치정보의 보호 및 이용 등에 관한 법률 등으로 개인의 위치와 이동궤적이 추정되지 않는 집계 형태여야 함. 집계 형태는 개인의 위치와 이동궤적이 추정되지 않도록 집계구의 데이터와 기지국 간의 정보를 이용하여 구축한 교통폴리곤 단위로 변환 후 OD형태로 데이터가 제공되고 있음. 데이터 제공을 위한 데이터 처리 과정은 아래와 같음
  - 원시 데이터 전처리
  - 데이터 로그별 이동과 체류 정보 구분
  - 위치 정보 변환 및 데이터 집계
- 데이터 구축은 분석 맵 레이어 구축 및 모바일 데이터 전처리와 통행량 데이터 가공으로 구분함
- 분석 맵 레이어 구축은 격자형 폴리곤 데이터와 교통폴리곤 데이터 및 행정구역 데이터를 이용해 각각의 폴리곤 형상 레이어를 경량화하는 작업과 폴리곤 중심 위치 보정을 통해 포인트 레이어 구축 작업을 진행함
- 모바일 데이터 기반의 교통 분석용 데이터를 그대로 서비스 시 많은 데이터를 가져오는 데에 발생하는 처리 속도 이슈로 인하여 행정 경계별로 그룹화하여 조회 데이터 건수를 감소시키고 속도를 개선함

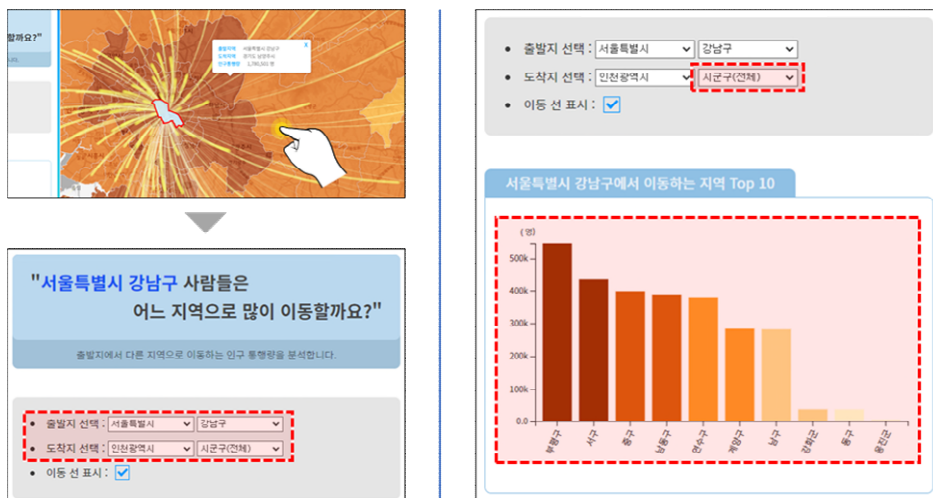
- 모바일 데이터 전처리는 누락 또는 오류 데이터를 분류하고 정제 데이터의 교통폴리곤과 행정구역을 공간적으로 결합시키고 출발 교통폴리곤과 도착 교통폴리곤의 중심 포인트 간 직선거리를 추가함
- 전처리한 모바일 통행량 데이터를 이용해 검색 속도 향상을 위한 인덱스를 생성하고 교통폴리곤 단위로 데이터를 그룹화하여 집계 통행량 DB를 구축함
- 행정구역 단위별로 서비스를 하기 때문에 구축한 교통폴리곤 단위 집계 통행량 DB를 이용하여 출발과 도착 읍면동 기준, 시군구 기준, 시도 기준으로 각각 집계 통행량 DB를 구축함
- 서비스 만족도를 향상시키기 위하여 검색 및 표출 속도를 개선하고자 모바일통신 빅데이터 속성정보를 가지고 분석 특성을 정의 후 행정구역 단위 집계 통행량 DB를 이용해 분석 특성 기준 집계 통행량 DB를 구축함

## 다. View-T Light

### 1) 주요통행분석

#### ○ 주요 통행지역 분석

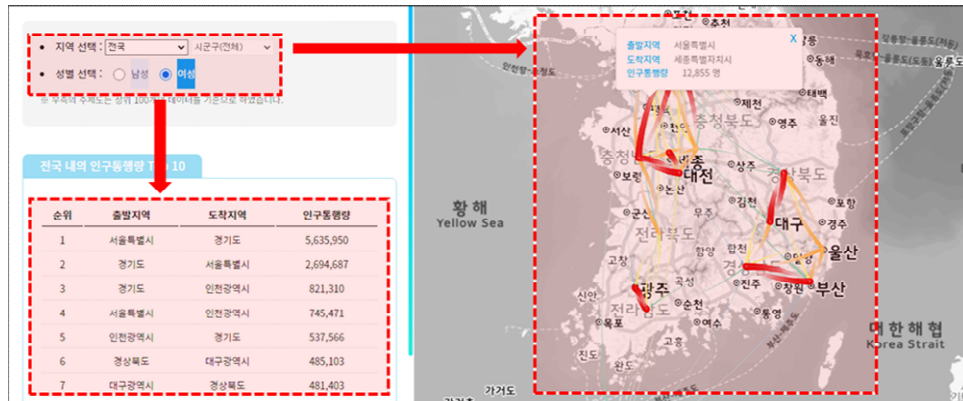
- 지도 위에서 출발지와 선택지를 각각 선택하는 복잡한 조작 방식을 개선하기 위하여 분석 결과 영역에서 콤보 박스를 통해 간편하게 출발/도착지를 선택할 수 있도록 개선
- 도착지 선택 시 시군구 지역 전체를 선택할 수 있게 하여 해당 선택한 지역들의 순위를 비교할 수 있도록 분석 결과 표시



<그림 5> 출발/도착 지역 선택방식 변경(좌), 도착지 하위 행정구역 전체 선택(우)

### ○ 주요 통행구간 분석

- 기존 지도에서 시군구 단위로 선택할 수 있던 지역 선택의 범위를 콤보박스를 이용하여 시도 단위까지 선택 가능하게 하고, 전국을 선택할 수 있도록 개선
- 기존에는 여성의 데이터로만 분석하였으나 분석 조건에 남성을 선택할 수 있는 UI를 추가하여 성별에 따라 분석할 수 있도록 개선

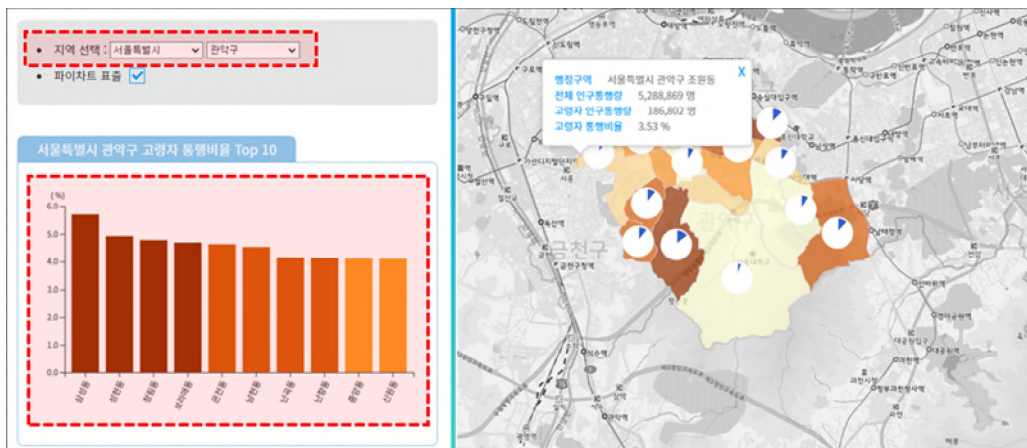


<그림 6> 기종점 인구 통행 구간 분석의 시도단위 전국 분석

## 2) 고령자 통행 분석

### ○ 고령자 통행 비중 분석

- 기존 지도를 클릭하여 시도 단위로 분석 지역을 선택할 수 있던 지역 범위를 시군구 단위까지 선택 가능하게 하고, 읍면동 단위까지의 분석 결과를 표출할 수 있도록 개선

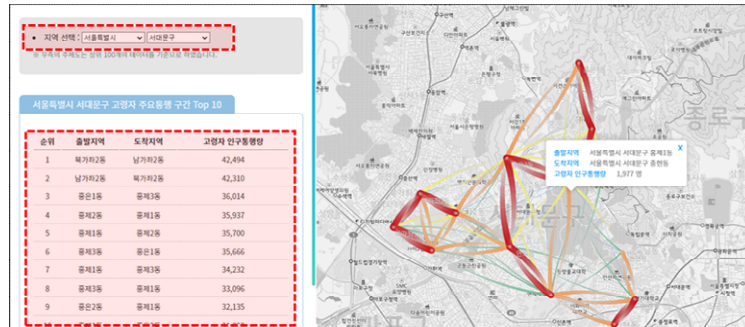


<그림 7> 고령자 통행 비중 분석의 읍면동 단위 분석 결과 표출



○ 고령자 주요 통행 구간 분석

- 기존 지도를 클릭하여 시도 단위로 분석 지역을 선택할 수 있던 지역 범위를 시군구 단위까지 선택 가능하게 하고, 읍면동 단위까지의 분석 결과를 표출할 수 있도록 개선

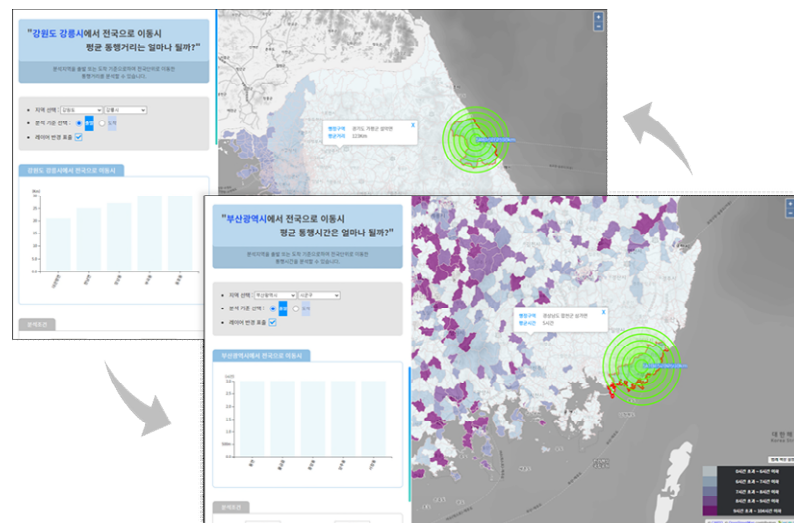


〈그림 8〉 고령자 주요 통행 구간 분석의 읍면동 단위 분석 결과 표출

3) 출·퇴근 통행분석

○ 통행시간/거리 분석

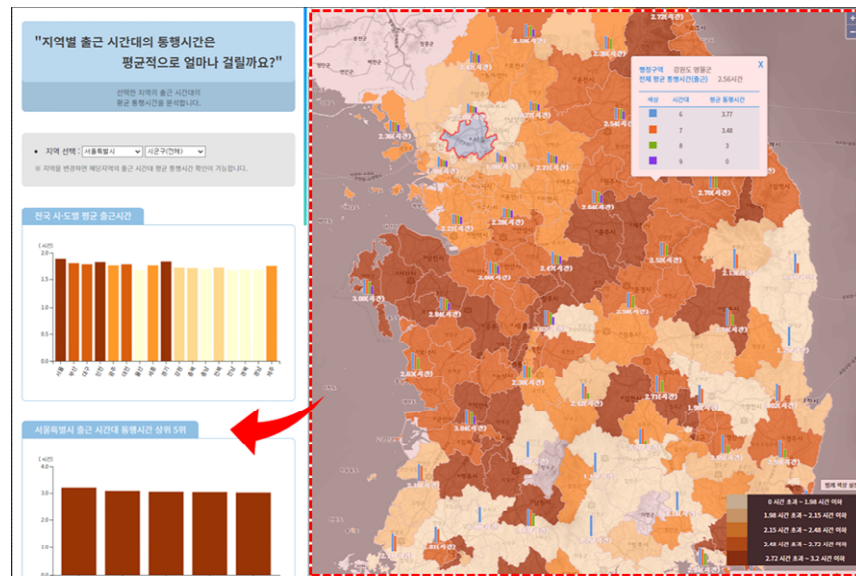
- 선택 지역을 출발/도착으로 하여 각 지역별 평균 통행시간과 거리를 분석
- 화면 좌측 분석 조건 및 결과 영역의 스크롤을 이동하여 평균 통행시간과 평균 통행 거리 조건을 전환하고, 지역 선택, 분석 기준 선택 (출발, 도착), 레이어 반경 표출의 조건을 설정 및 분석 결과의 상위 5개 지역을 그래프로 표출
- 화면 우측 주제도에서는 선택 지역을 기준으로 각 지역별 평균 통행시간과 거리를 범례 범위에 따른 색상으로 표시하고 마우스 커서에 위치하는 지역의 상세 정보를 툴팁으로 표출



〈그림 9〉 분석 조건결과 영역의 스크롤 이동으로 통행 거리/통행시간 조건 변경

○ 출·퇴근 통행시간 분석

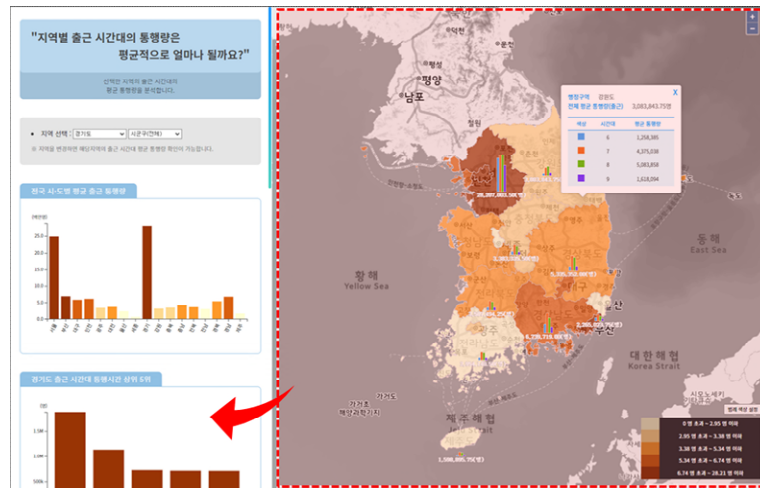
- 시도, 또는 시군구 단위의 지역을 선택하면 해당 지역을 기준으로 출근, 퇴근 시간대의 평균 통행시간을 분석하고, 평균 통행시간의 상위 5개, 하위 5개 지역을 그래프로 표출
- 주제도에서는 각 행정구역별 출근 시간대(6-9시), 퇴근 시간대(18-21시)의 각각의 시간대별 통행시간을 그래프와 툴팁을 통해 상세 표출
- 화면 좌측 분석 조건 및 결과 영역의 스크롤을 이동하여 출근 시간대와 퇴근 시간대의 조건을 각각 달리하여 분석



<그림 10> 출·퇴근 통행시간 분석 도구 화면

○ 출·퇴근 통행량 분석

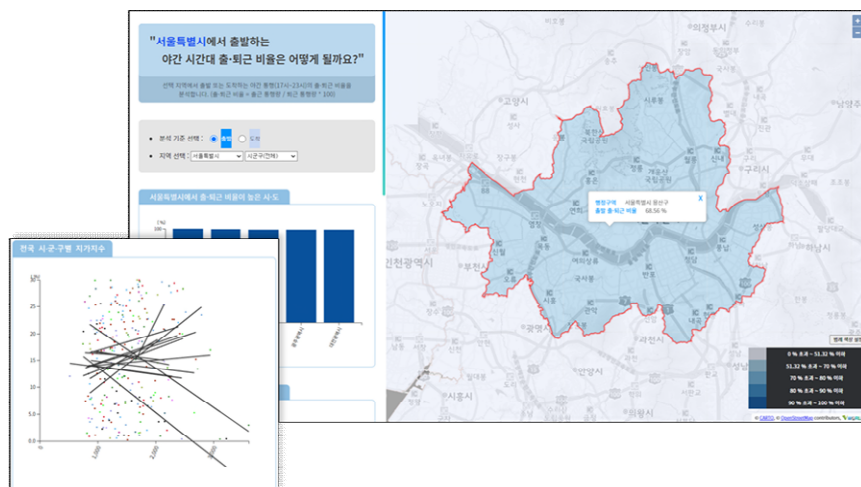
- 시도, 또는 시군구 단위의 지역을 선택하면 해당 지역을 기준으로 출근, 퇴근 시간대의 평균 통행량을 분석하고, 평균 통행량의 상위 5개, 하위 5개 지역을 그래프로 표출
- 주제도에서는 각 행정구역별 출근 시간대(6-9시), 퇴근 시간대(18-21시)의 각각의 시간대별 통행량을 그래프와 툴팁을 통해 상세 표출
- 화면 좌측 분석 조건 및 결과 영역의 스크롤을 이동하여 출근시간대와 퇴근시간대의 조건을 각각 달리하여 분석



<그림 11> 출·퇴근 통행량 분석 도구 화면

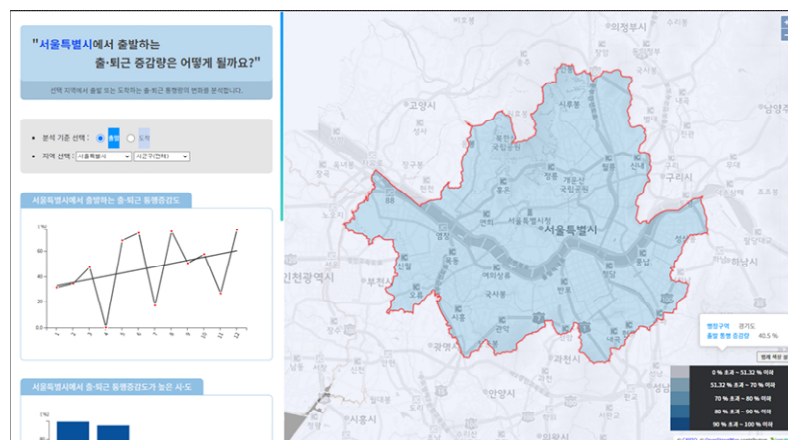
○ 근무형태에 따른 통근 통행 분석

- 야간 시간대(17-23시)에 선택 지역에서 출발 또는 선택 지역으로 도착하는 통행량에서 출퇴근의 비율을 분석하는 신규 분석 도구
- 좌측 분석 조건 및 결과 영역에서는 분석 기준을 출발 또는 도착으로 설정하고 기준이 되는 지역을 시도 또는 시군구 단위로 선택
- 조건을 설정하면 출퇴근 비율을 분석하여 좌측 영역에 출근 비율이 높은 상위 5개 지역과 퇴근 비율이 높은 상위 5개 지역을 그래프로 표출하고, 우측 주제도 영역에서는 각 지역별 범례 범위에 해당하는 색상을 표출
- 분석 결과 그래프 아래에는 전국 시도, 시군구별 지가지수 관련 산점도 그래프를 표출



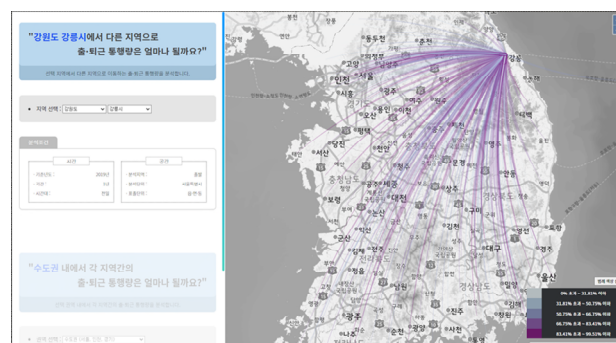
<그림 12> 근무형태에 따른 통근 통행 분석화면과 전국 시군구별 지가지수 산점도

- 고용인구 변화에 따른 통근 통행 분석
  - 출퇴근을 목적으로 이동하는 인구통행량 데이터로 선택 지역에서 출발 또는 선택 지역으로 도착하는 월별 인구통행량의 증감율을 분석하는 신규 분석도구
  - 좌측 분석 조건 및 결과 영역에서는 분석 기준을 출발 또는 도착으로 설정하고 기준이 되는 지역을 시도 또는 시군구 단위로 선택
  - 조건을 설정하면 선택 지역을 기준으로 출발 도착하는 출퇴근 목적의 통행량을 월별 증감도와 증감도가 높은 5개 지역과 증감도가 낮은 5개 지역의 순위 그래프를 좌측 영역에 표출하고, 우측 주제도 영역에서는 각 지역별 범례 범위에 해당하는 색상을 표출



<그림 13> 고용인구 변화에 따른 통근 통행 분석화면

- 출퇴근 통행 연결성 분석
  - 선택 지역을 기준으로 각 지역으로의 출퇴근 통행량을 분석하는 신규 분석도구
  - 화면 좌측 분석 조건 영역의 스크롤을 이동하여 선택 지역 기준의 전국 통행 조건과 선택 권역 내 통행 조건을 전환하고, 지역 또는 권역을 선택하는 UI 표출
  - 지역 또는 권역을 선택 시 선택된 조건으로 분석된 결과가 주제도에서 지역 연결선으로 표출



<그림 14> 출·퇴근 통행 연결성 분석화면

#### 4) 관광·연휴분석

##### ○ 계절별 핫플레이스 분석

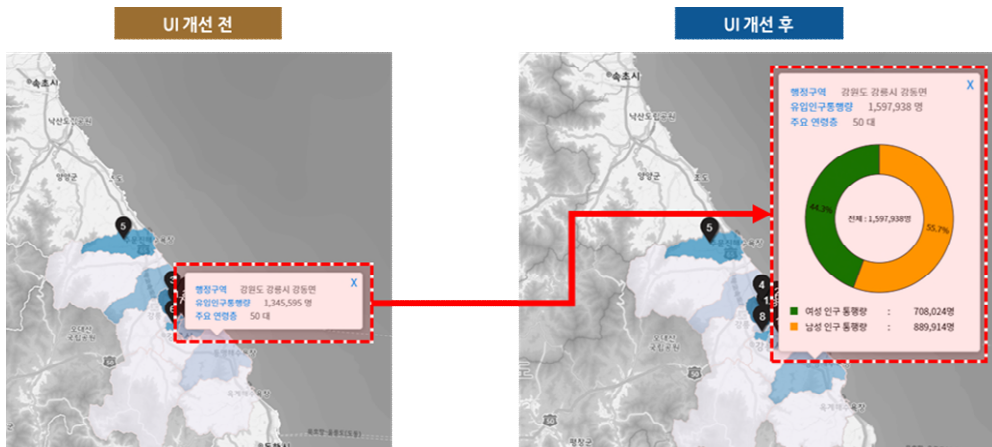
- 분석 결과 그래프에 지역명을 시도 단위부터 상세하게 표기하고 전체 이용량을 여성과 남성의 이동량으로 구분하여 표시 추가



<그림 15> 계절별 핫플레이스의 분석 결과

##### ○ 관광 특화지역 분석

- 분석 결과를 나타내는 툴팁에 남/녀 성별 통행량을 구분하여 표시하고 이를 알아보기 쉽도록 파이그래프로 표현

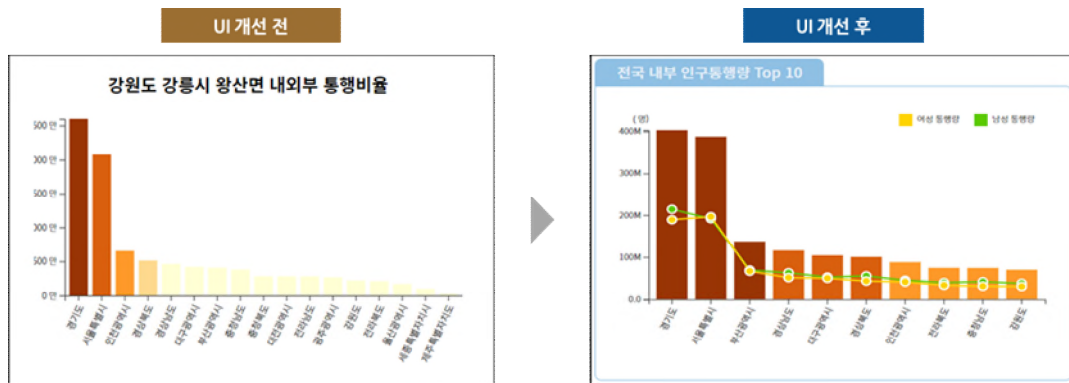


<그림 16> 관광 특화지역 분석 결과의 상세 정보 툴팁 고도화

##### ○ 관광지 내·외부 통행비율 분석

- 분석 결과 그래프에 표시되는 지역을 순위가 높은 1~10까지로 요약하여 표시하고 각 지역별 인구통행량의 남/녀 성별을 구분하여 결과 표출



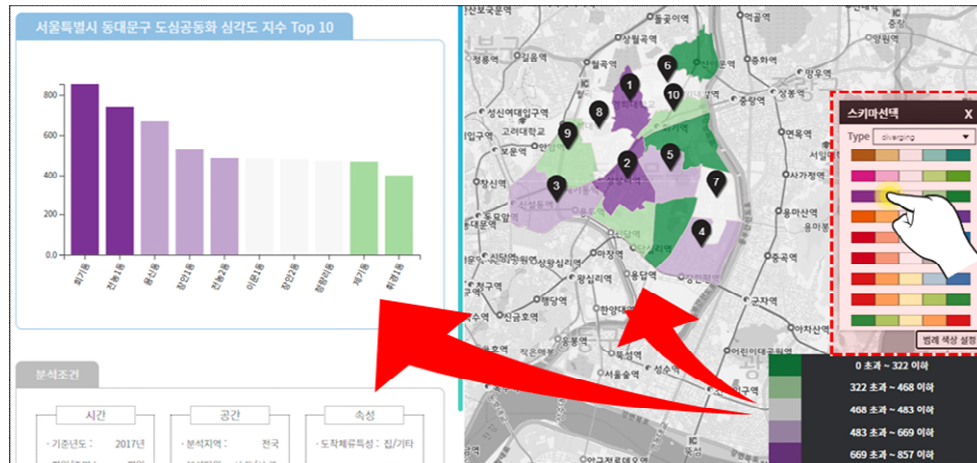


<그림 17> 관광지 내·외부 통행비율 분석 결과의 UI 개선 전/후 비교

## 5) 경제·사회활동분석

### ○ 도심공동화 심각도 분석

- 분석 결과가 색상 단계로 구분되는 표출 방식에서 결과의 관계 파악 및 확실한 의미 전달을 할 수 있도록 여러 가지 색상을 조합한 색상 스키마를 사용하여 표현 방식을 변경할 수 있는 기능 추가



<그림 18> 도심공동화 심각도 분석 결과의 UI

## 라. View-T Expert

### 1) 사용자 편의성을 위한 UI/UX 개선

- 심층적인 분석보다 간단한 분석을 선호하는 사용자를 위해 조건 설정, 범례 설정을 간소화함
- “키워드 분석” 옵션을 조건 설정 부분에 추가하여, 사용자가 상세하게 조건을 입력하지 않아도 분석을 수행할 수 있도록 함
- 사용자가 자주 사용하는 분석 조건을 기본 옵션으로 두어, 사용자의 조건 선택을 최소화



<그림 19> 분석 조건 값 UI 변경

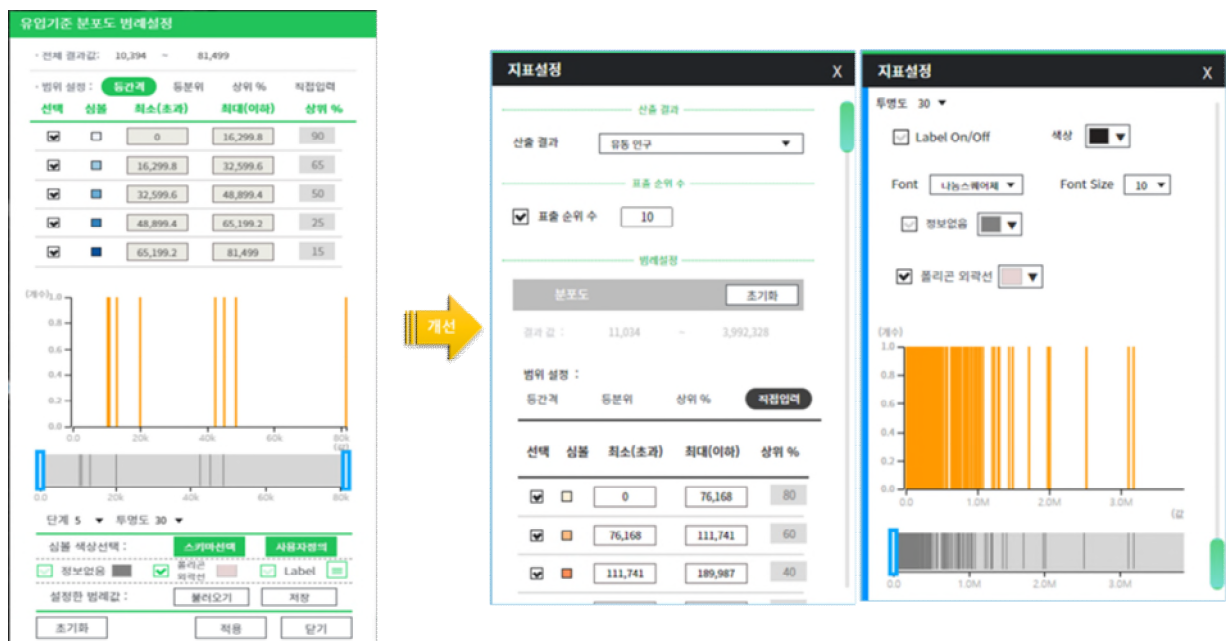
<표 7> 주제별 키워드 사전 조건

통행목적	출발 조건		도착 조건		기타 조건	비고
	시간대	트립타입	시간대	트립타입		
출근	06-09	집 (H)	07-10	회사 (C)	20-60대	필수
등교	06-09	집 (H)	07-10	학교 (S)	10-20대	필수
퇴근	18-21	회사 (C)	19-22	집 (H)	20-60대	필수
하교	15-21	학교 (S)	16-21	집 (H)	10-20대	필수
귀가	18-20	정기적 잠재체류지 (R),	19-23	집 (H)	-	필수
		비정기적 잠재체류지 (X)				
여가	-	집 (H),	-	정기적 잠재체류지 (R),	-	필수

		회사(C), 학교(S), 정기적 잠재체류지(R), 비정기적 잠재체류지(X)		비정기적잠재체류지(X)		
사적업무 후 복귀	-	정기적 잠재체류지(R), 비정기적 잠재체류지(X)	-	회사(C), 학교(S)	10~60대	보류

\* ① 도착지 타입이 ‘-’인 경우, ② 회사(C)→학교(S) 또는 학교(S)→회사(C) 타입인 경우, ③출발일자와 도착일자가 다른 경우는 제외

- 필수입력 조건이 아닌 범례 및 표출 정의를 별도의 아이콘으로 배치하여 레이아웃의 여유 공간을 확보하고, 요소 전체 UI를 개선해 사용자가 정보를 식별하는데 수월할 수 있도록 개선



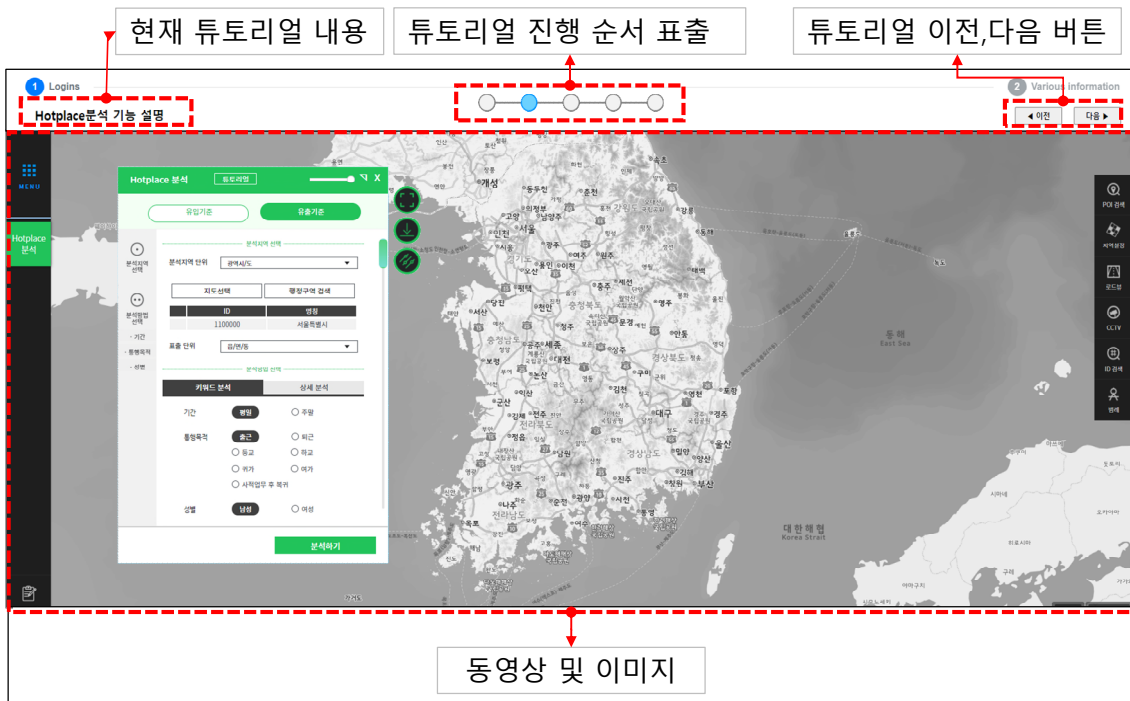
<그림 20> 범례 설정 UI 변경

## 2) 분석 기능 튜토리얼 제작

- 통신 주요 분석 기능의 기능 및 분석 결과 설명, 분석 도구의 사용 방법을 제공
  - 분석 도구 사용이 익숙하지 않은 사용자에게 동영상 또는 그림을 통하여 튜토리얼을 제공함으로써 사용 편의를 높이고, 조건 설정 등 분석 과정 및 분석결과에 대한 사용자의 이해도를 높일 수 있도록 함



- 이전, 다음 버튼을 통하여 사용자가 원하는 튜토리얼을 진행할 수 있도록 구현
- 상단 중앙에 튜토리얼 진행 순서를 표시하는 상태바 표시
- 좌측 상단에 현재 진행 중인 튜토리얼의 기능명 표시
- 중앙에 해당 분석 기능의 사용 방법의 이해를 도울 수 있는 동영상 혹은 이미지를 표시



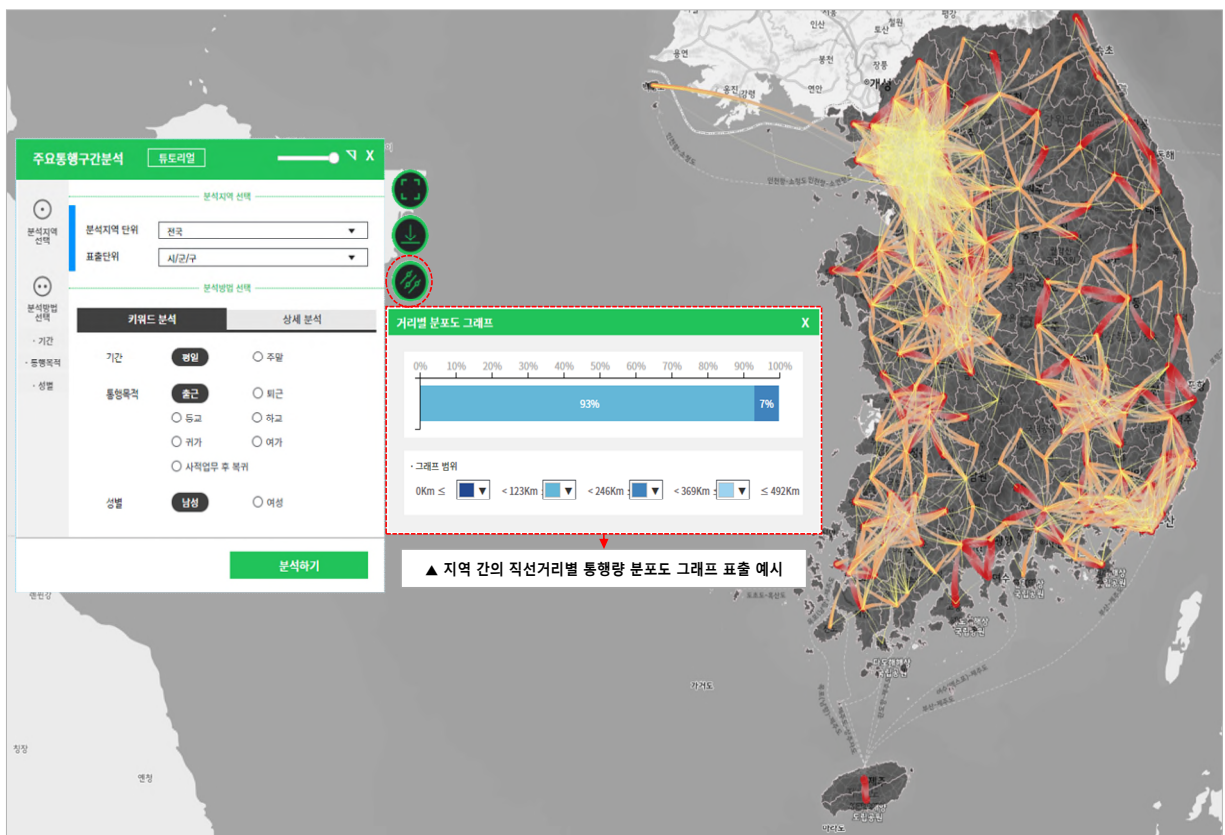
<그림 21> 분석 기능 튜토리얼 UI

### 3) 기존 분석 기능 (주요통행구간 분석) 개선

- 단거리 구간 통행과 장거리 구간 통행의 분석이 가능하도록 개선
  - 전국의 시도, 시군구 단위 간 구간 분석이 가능하도록 개선
  - 단거리 구간 통행과 장거리 구간 통행 비교를 위한 그래프 표시

&lt;표 8&gt; 주요통행구간 분석 기능의 연도별 표출형태 변경

구분	권한	변경 전	변경 후
분석 지역 선택	일반사용자	최대 5개 (시군구, 읍면동)	변화 없음
	고급사용자	일반사용자 기능에 더하여 시도 단위는 1개 선택가능	전국 단위 분석 가능 (상세 조건 선택 제한)
표출 단위 선택	일반사용자	분석 지역 하위 단위 선택 (시군구 → 읍면동)	최소 표출 단위 교통폴리곤 → 읍면동
	고급사용자		전국 단위 경우 시도, 시군구 단위 선택



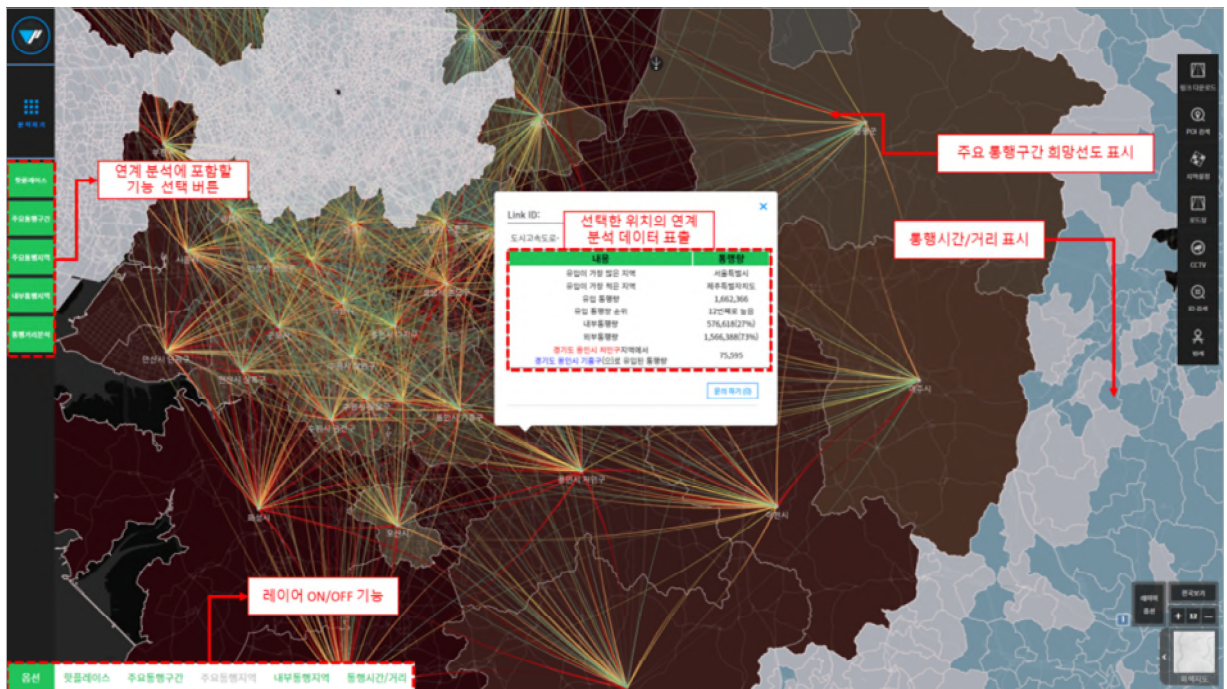
&lt;그림 22&gt; 주요 통행 구간 분석의 전국 분석 결과 및 통행량 분포도 그래프

## 5) 심층 분석을 가능케 하는 분석 기능 개발

### ① 정보 연계 분석 기능

- 여러 분석 기능을 선택하고 그에 따른 결과를 동시에 주제도로 표출하여 확인 할 수 있는 기능

- 핫플레이스, 주요통행구간, 주요통행지역, 내부통행지역, 통행거리분석을 연계분석 가능함
- 정보 연계 분석 기능 UI
  - 각 분석 기능을 선택한 후 분석하기 버튼을 클릭하여 연계분석을 실행함
  - 주제도상 클릭하여 연계 분석된 데이터를 표출함
  - 하단 옵션 버튼을 통하여 연계분석 기능들의 레이어를 끄거나 켜서 개별 주제도를 볼 수 있음
  - 희망선도를 통해서 주요 통행구간을 인지할 수 있음
  - 희망선도 혹은 지역을 선택하여 해당 위치의 데이터 개괄을 볼 수 있음



<그림 23> 정보 연계분석 기능 화면

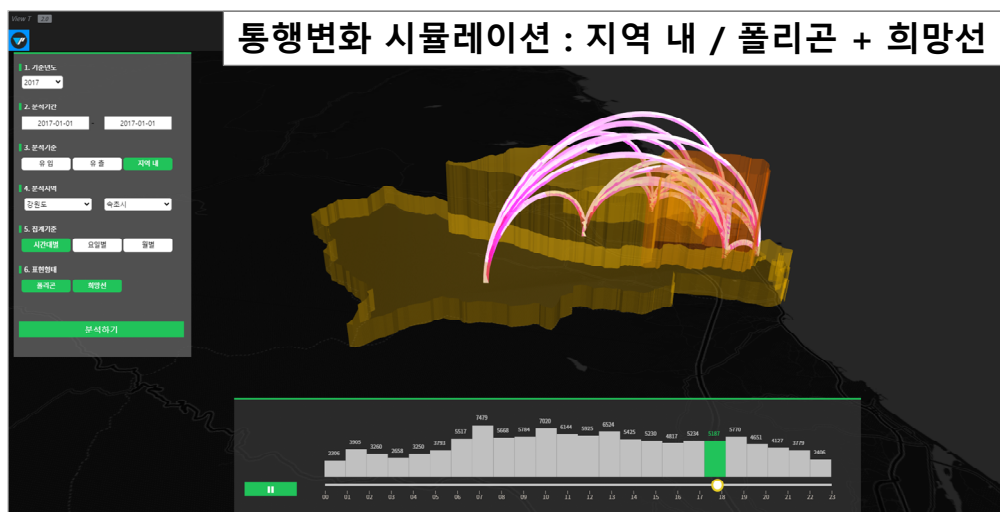
## ② 통행 변화 시뮬레이션 기능

- 통행 변화를 시계열로 확인할 수 있는 기능
- 통행 변화 시뮬레이션 분석 기능 UI
  - 분석유형(유입, 유출 및 O/D기준), 분석 지역 및 기간을 설정해 사용자 의도에 맞는 공간데이터를 시각화하여 통행 변화의 흐름 분석이 가능하도록 개발

- 시물레이션은 시간대별, 요일별, 월별을 1년치 데이터를 기준으로 집계하여 통행의 흐름을 보여줄 수 있도록 개발
- 예시) 시간대별 - 00시~24시, 요일별 - 월요일 ~ 일요일, 월별 - 1월 ~12월
- 시물레이터를 통해 시간에 따른 데이터 변화를 제어할 수 있음
- 재생, 정지, 재생트랙 제어 등



<그림 24> 유입 기준 폴리곤 형태의 통행변화 시물레이션



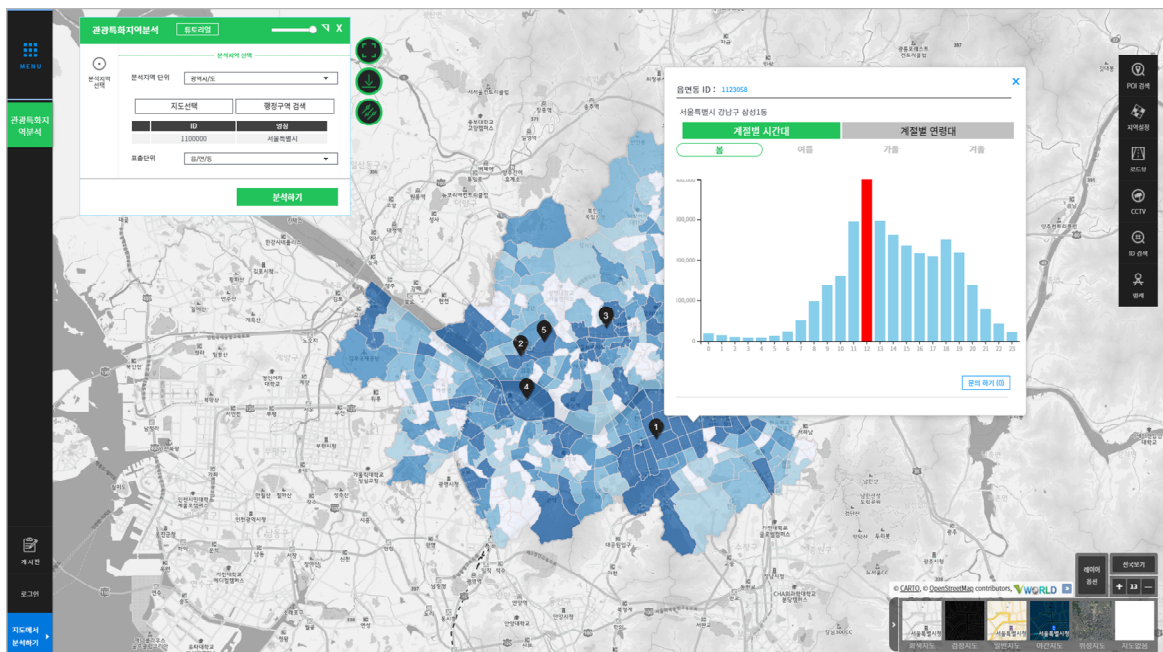
<그림 25> 지역 내 기준 폴리곤 및 희망선 형태의 통행변화 시물레이션

6) View-T Light의 분석 서비스 중 일부를 Expert에 개발

① 관광 특화지역 분석 기능



- 계절에 따라 통행량 변화가 큰 지역을 확인 할 수 있는 기능
  - 도착지 트립타입이 비정기적 잠재체류지인 통행을 지역별, 계절별로 분류 표시
- 관광 특화지역 분석 기능 UI
  - 분석 지역 단위를 광역시/도, 시/군/구에서 선택
  - 분석하길 원하는 지역을 지도선택 혹은 행정구역 검색을 통하여 선택
  - 표출 단위는 분석 지역 단위의 하위 지역 중 선택
  - 상위 5개 지역을 주제도 상에 표시
  - 해당 지역의 계절별 시간대와 계절별 연령대 통행량의 변화를 순위 그래프로 표시

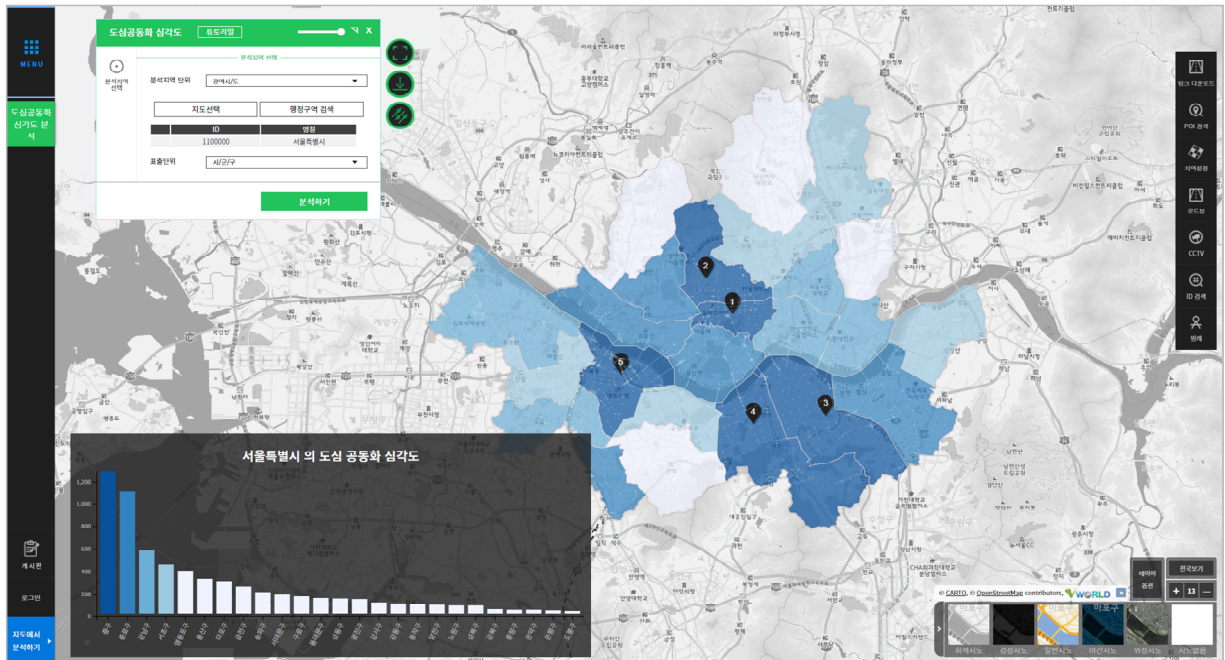


<그림 26> 관광 특화지역 분석 기능 화면

## ② 도심공동화 심각도 분석 기능

- 도시 중심부의 상주인구가 줄어드는 현상을 확인할 수 있는 기능
  - 심야 시간대 집으로의 통행량과 낮 시간대 통행량의 비율을 통한 상주인구 추정
- 도심공동화 심각도 분석 기능 UI
  - 분석 지역 단위를 광역시/도, 시/군/구에서 선택
  - 분석하길 원하는 지역을 지도선택 혹은 행정구역 검색을 통하여 선택

- 표출 단위는 분석 지역 단위의 하위 지역으로 고정됨
- 상위5개 지역을 주제도 상에 표시
- 해당 지역의 공동화 심각도를 순위 그래프로 표시



<그림 27> 도심공동화 심각도 분석화면

## 5. 결론 및 차년도 수행계획

### 가. 결론

- 본 과업에서는 기존 방법보다 정확하게 통행시간 및 통행량 등 통행지표를 산출하기 위하여 '20에 개발한 모바일 기지국 원천데이터 가공알고리즘을 보완함
  - 알고리즘은 합리적 수준의 사·공간적 데이터 군집 (Data aggregation and smoothing) 과정을 제시하여 신호 이상 현상을 효과적으로 처리하고, 개별통행 특성 및 패턴을 고려한 주요 통행목적지 추정을 통해 야간 및 비정기 근무 통행을 구분함
  - 또한, 기존 통신폴리곤 기준의 위치 추정 정보를 GPS와 같은 점단위의 위치 추정 정보로 변환하는 방법을 제안하여 모바일통신 빅데이터의 공간적 해상도를 개선하였으며, 향후 이동경로 및 수단 추정을 위한 개발 알고리즘의 적용방안을 검토함
- 2020년 1월 1일부터 12월 31일까지 생성된 로그 기록에 모바일통신 빅데이터 가공알고리즘을 적용하여 기존 DB를 구축하고 검증함
  - 각 개인별 로그 기록은 「개인정보보호」, 「위치정보의 보호 및 이용 등에 관한 법률」에 저촉되지 않도록 집계한 형태의 DB를 설계하였으며, 이를 기준으로 DB를 구축함
  - 모바일 기지국 데이터 기반으로 구축된 통행DB의 시간대별, 요일별, 성별, 연령별 통행량 및 통행시간을 기존 국가교통DB사업의 '가구통행실태조사자료 (2016년)'와 통계청의 '인구총조사자료 (2015년)' 비교하여 검증함
- 모바일통신 빅데이터 기반 교통지표를 개발하고, 분석 서비스의 성능 및 사용성을 개선하기 위하여 DB를 재구조화하여 서비스를 제공함
  - 또한, 2개 이상의 분석 수행 결과를 주제도로 동시에 확인할 수 있는 '정보 연계분석 기능' 개발하고, 통행 변화를 시계열로 확인할 수 있는 '통행 변화 시뮬레이션 분석 기능' 개발함

### 나. 차년도 수행계획

- 모바일통신 빅데이터 가공알고리즘 고도화
  - 데이터 전처리 알고리즘 개선 및 고도화

- 지역특성(도시부/비도시부), 인적특성(연령, 휴대폰 이용률 등)에 따른 로그 특성 분석 및 기 전처리 알고리즘 보완
- 통행목적 구분 알고리즘 개선 및 고도화
  - 모바일통신 빅데이터 기반 추정 가능 체류지(거주지, 근무지, 여가 등) 선정
  - 거주지 및 근무지 추정 로직 개선
- 모바일통신 빅데이터 기반 통행DB 구축 및 검증
  - 모바일통신 빅데이터 기반 통행DB 구축 (2021년도 전국)
  - 기 구축된 DB (2020년도 전국) 검증
    - 가구통행실태조사, 인구총조사 등 기존 조사기반 통계를 이용한 검증
- 모바일통신 빅데이터 기반 교통지표 현행화 및 서비스 제공
  - 교통지표 현행화
    - 1차 지표 : 통근·통학 통행량·통행시간, 기타통행 통행량 (1시간, 격자 단위의 전국 대상 1년 DB)
    - 통행지표 : 연령별·성별 주요 통행 구간, 통행 연결성, 핫플레이스, 관광/고령자/계절별 통행 특성 등
    - 생활·경제지표 : 고용인구 변화율, 야간근무 비율, 연령별 근무 형태, 도심공동화 심각도, 경제활동 의존도 및 수용도 등





## 제1장 과업의 개요

---

제1절 배경 및 목적

제2절 범위 및 내용



## 제1장 과업의 개요

### 제1절 배경 및 목적

#### 1. 배경

- 최근 교통 분야에서 각종 모빌리티 정보(예를 들어, 차량GPS 데이터, 대중교통카드 데이터, DTG 데이터, 모바일통신 빅데이터 등)를 활용하여 개별 통행정보를 추출하고 각종 통행 관련 분석이 가능해지면서 다양한 방법으로 활용되고 있으나, 각 데이터의 특성에 따라 분석 대상과 범위가 제한적임
  - 차량에 탑재된 기기(내비게이션, DTG 등)나 도로에 설치된 기기(검지기, DSRC 등)를 통해 차량 통행자에 대한 통행 정보는 취득 가능하나, 자전거·도보 등으로 이동한 통행자에 대한 통행 정보는 표본율이 낮은 가구통행실태조사 결과에만 의존하고 있는 실정
  - 대중교통 이용자에 대한 통행은 교통카드 DB(대중교통 승하차 시 태그 정보)로 대부분 추정 가능하나, 잠재적으로 대중교통을 이용할 가능성이 있는 타 교통수단 이용자(차량, 도보 등)에 대한 통행은 일부만 추정 가능
- 그러므로, 전 국민의 95% 이상이 사용하고, 이용 교통수단에 구애받지 않고 구독이 가능한 모바일통신 빅데이터를 이용한 개별통행DB 구축 및 활용이 요구되고 있음
  - 모바일 기기와 기지국 간의 송·수신 이력인 모바일통신 빅데이터는 대부분의 국민이 이용하고 있는 모바일 기기를 통해 축적되기 때문에 표본율이 매우 높고, 이동수단에 구애받지 않고(항공 제외), 기기의 전원을 끄지 않는 이상 개인의 이동 궤적이 상세히 기록되기 때문
  - 1년 365일 수집되는 모바일통신 빅데이터는 사회적 현안에 즉각적으로 대응할 수 있는 시의성 있는 모빌리티 정보를 제공할 수 있음
- 그러나, 기존 민간과 공공에서는 다양한 방식으로 모바일 통신 빅데이터를 가공하여 유동인구 형태로 활용하고 있지만, 개별통행 행태를 분석하기에는 한계점이 있으며, 이를 극복하기 위한 기술적 DB 구축 및 분석 방법이 요구되고 있음
  - 모바일 통신 데이터는 기존 설문 조사 방식의 데이터 수집 방법과 비교하여 개인 및

통행에 대한 정보가 매우 제한적이며, 시간 단위 위치 정보를 이용하여 통행의 유무, 통행목적, 통행시간 등을 추정해야 함

- 또한, 기지국 위치로 대표되는 모바일 기지국 데이터의 공간적 저해상도 문제 때문에 정확한 통행 위치와 정밀한 통행시간을 추정하는데 어려움이 있음

## 2. 목적

- 모바일 통신 빅데이터 중 기지국 기반의 데이터를 이용하여 시의성 있는 O/D 기반의 통행 정보(통행량, 통행시간)생성 방법을 정의하고 DB 구축 및 제공
  - 기존 모바일 기지국 데이터의 공간적 저해상도를 해결할 수 있는 방안을 제시
  - 모바일 기지국 데이터를 이용한 각종 체류지(집, 회사 등) 추정 방법을 제시
  - 지역별 기지국 간격 등을 고려한 정밀한 통행시간 추정 방법 제시
  - 시의성 있는 전국 교통폴리곤별 1시간 단위 연령별/성별 통행량 및 통행시간 제공
- 통행자별 이용 교통정보 정보가 포함된 모바일통신 빅데이터를 이용하여 통행DB를 구축하고, 이를 기반으로 교통수요를 분석할 수 있는 지표, 기능 등을 개발하여 고도화된 교통 정책 수립 지원에 활용하고자 함

## 제2절 범위 및 내용

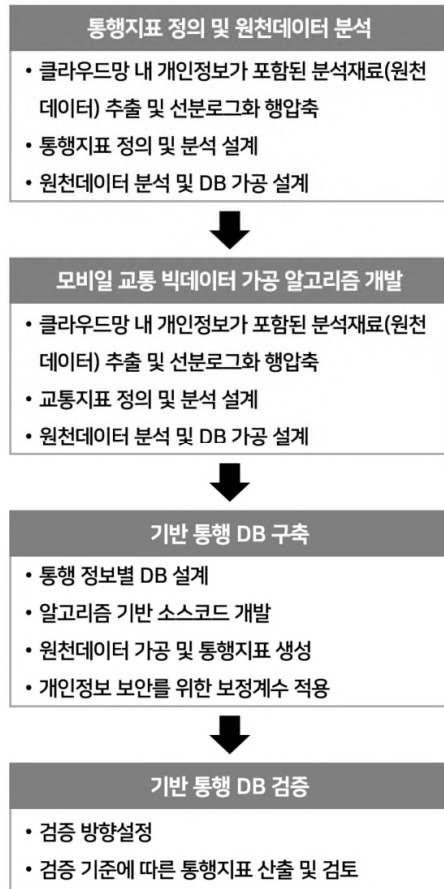
### 1. 범위

- 시간적 범위
  - 모바일통신 빅데이터 가공알고리즘 개발: 2020년 원천데이터
  - 2020년도 기준 기반 DB 구축: 2020년 1월 1일 ~ 2020년 12월 31일 (1년)
  - 웹 서비스를 위한 각종 분석 도구 개발: 2019년 1월 1일 ~ 2019년 12월 31일
    - 2020년도 연구 성과물인 2019년도 기준 기반 DB 활용
- 공간적 범위: 전국

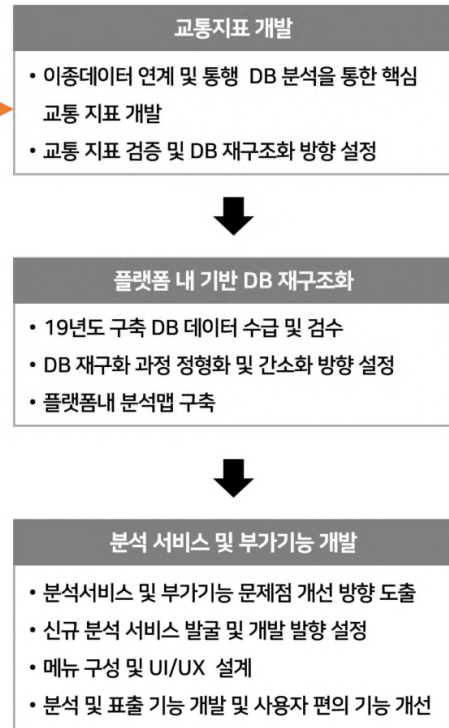
### 2. 내용

- 모바일통신 빅데이터 가공알고리즘 개발 및 고도화
  - 전처리 알고리즘 고도화
  - 체류 정보 및 통행목적 추정 알고리즘 고도화
  - 가공알고리즘 데이터 최적화 및 속도 개선
- 모바일통신 빅데이터 기반 통행 DB 구축 및 검증
  - 기준년도 통행DB 구축
  - 기 구축된 통행DB 검증
- 모바일통신 빅데이터 기반 교통지표 개발 및 서비스 제공
  - 주요 교통지표 생성 및 분석
  - 분석 서비스 개발 및 제공
  - 기반DB 현행화 및 속도 개선

### 모바일 교통빅데이터 DB 구축 및 검증 부문



### 모바일 교통빅데이터 기반 교통지표 및 서비스 개발



<그림 1- 1> 과업의 수행 절차

## 제2장 모바일통신 빅데이터 가공알고리즘 개발 및 고도화

---

제1절 모바일통신 빅데이터 가공  
알고리즘 주요 개념

제2절 전처리 알고리즘 고도화

제3절 체류정보 및 통행목적 추정  
알고리즘 고도화

제4절 가공알고리즘 데이터 최적화 및  
속도 개선



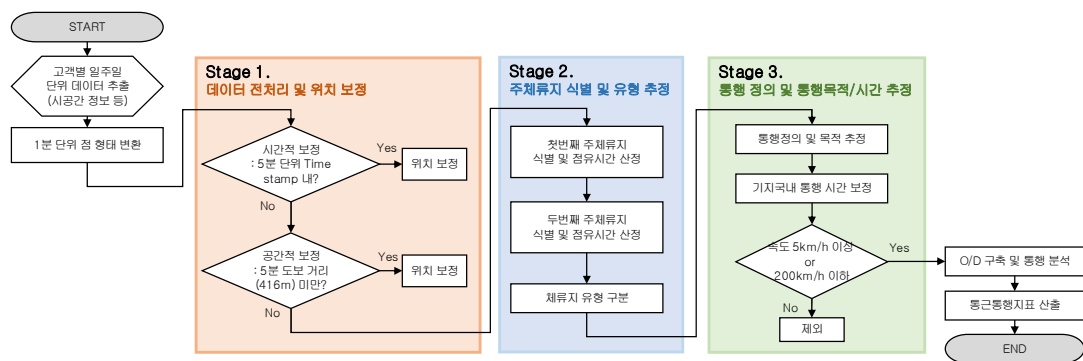


## 제2장 모바일통신 빅데이터 가공알고리즘 개발 및 고도화

### 제1절 모바일통신 빅데이터 가공알고리즘 주요 개념

#### 1. 가공알고리즘 주요 개념

- 입력데이터(input data)로 각 통행자에 대한 7일 치 원천데이터가 사용되며 각 통행자에 대한 개별 통행 특성을 기반으로 주체류지를 식별하고 통행지표를 산출
- 데이터 전처리 단계에서 신호 이상 현상 및 신호 이상치를 정형화할 필요 없이 데이터를 시·공간적으로 보정함으로써 비현실적으로 발생하는 통행자의 이동 궤적을 보정
- 기지국 데이터 기록빈도 및 체류시간을 기반으로 통행자의 개별 통행 특성을 고려한 주체류지 식별을 수행하기 때문에 정기 출퇴근 통행뿐만 아니라 야간 근무자를 포함한 비정기 출퇴근 통행도 식별 가능
- 개별 통행자에 대한 주체류지를 식별한 후, 해당 위치에서의 체류시작시간과 체류종료시간을 기반으로 통근시간 및 통행량 등 통행지표 산출



<그림 2- 1> 모바일통신 빅데이터 가공알고리즘

## 2. 가공알고리즘 단계별 수행사항

- 기지국과 단말기 간에 발생하는 불규칙적인 신호 이상 현상을 보정함
- 개별 통행 특성 및 지역 간 기지국 거리 등을 고려하여 이동/체류를 구분하고 체류지 유형을 추정함
- 통행자에 대한 개별 통행 특성을 기반으로 주체류지를 식별하고 통행지표를 산출
- 데이터 전처리 단계에서 신호 이상 및 신호 이상치를 정형화할 필요 없이, 데이터를 시·공간적으로 보정함으로써 비현실적으로 발생하는 통행자의 이동궤적을 보정
- 기지국 데이터 기록빈도 및 체류시간을 기반으로 통행자의 개별통행 특성을 고려한 주체류지 식별을 수행하며 정기 출퇴근뿐만 아니라 야간 근무자를 포함한 비정기 출퇴근 통행도 식별함
- 개별 통행자에 대한 주체류지를 식별한 후, 해당 위치에서의 체류시작시간과 체류종료시간을 기반으로 통근 시간 및 통행량 등 통행지표 산출

## 3. 기존 가공알고리즘 개선 및 고도화

- 모바일 기지국 데이터를 기반으로 한 개인의 이동 및 통행목적에 파악할 수 있는 개별 통행사슬(Trip Chain) DB를 구축하기 위한 효율적인 가공알고리즘의 개발이 필요함
- 고도화된 개별 통행사슬 DB를 구축하기 위한 알고리즘 고도화는 크게 데이터 전처리 및 이상치 처리(핸드오버 이슈 해결), 공간적 보정, 주 체류지 식별 및 유형 구분(통행목적) 세분화로 구분하여 보완하였음

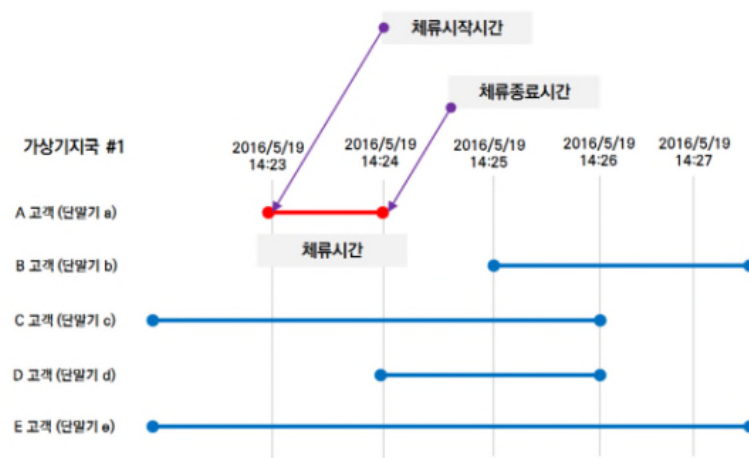
## 제2절 전처리 알고리즘 고도화

### 1. 전처리 알고리즘 주요 개념 및 단계

- 적절한 공간적 해상도를 바탕으로 통행자의 통행속도를 고려한 이동궤적을 식별하기 위해서 5분이라는 기준 시간을 정하고 해당 시간대에 기록된 기지국의 기록빈도를 기반으로 통행자의 실제 위치를 추정
- 짧은 시간 동안 발생한 핑퐁 핸드오버 현상에 의한 체류지 식별률 저하 및 비현실적으로 기록된 이동궤적을 보정하기 위하여 위해 특정 시간 단위(5분)별로 데이터를 분할
- 데이터를 분할 한 후, 각 시간 단위(time slice) 내에 기록된 기지국 데이터의 기록빈도를 기반으로 통행자의 실제 위치 추정
- 통행자의 실제 위치 추정 방법: 각 기지국에서의 체류시간을 기반으로 가중치를 주어 통행자 위치 좌표(X)를 계산
- 보행자 평균 도보 속도(5km/h)를 기준으로 5분(시간적 데이터 보정 시 설정한 시간 단위) 거리(약 416m) 범위 내 위치한 체류지들에 대해 각 체류지에서의 체류시간을 기반으로 해당 시간대에서의 통행자 실제 위치를 추정
- 원천데이터에 대해 시·공간적으로 신호 이상 기록을 보정해줌으로써, 분 단위로 연속해서 변화하거나 좁은 공간적 범위 내에 산발적으로 기록되는 기지국 위치 보정 가능

## 2. 모바일 기지국 원천데이터 전처리

- 포인트 단위의(밀리세컨 단위) 로그를 선분 단위의(분 단위) 로그로 변환
  - 닷(dot) 형태로 기록된 로그를 기록된 시간 순서대로 연결한 후, 최초 기록된 시간(이하 ‘체류시작시간’)과 마지막으로 기록된 시간(이하 ‘체류종료시간’)정보만 추출하여 단말기가 해당 기지국에 식별된 시간(이하 ‘체류시간’)을 산출함
  - 체류시간은 체류종료시간에서 체류시작시간을 뺀 값이며, 초(sec) 단위는 생략함



주: 자료를 참고하여 도식화 함

자료: 박미울·주은정(2017), LTE 시그널 정보를 이용한 위치정보가공 및 유동인구집계 방법, 한국통신학회 2017년도 동계종합학술 발표회, p.285.

<그림 2- 2> 선분 이력 데이터로의 변환 예시

### 3. 통행 정보 왜곡 데이터 보정

#### 가. 시간적 데이터 보정

- Liu T. et al.(2018)<sup>1)</sup>이 제안한 내용에 의거, 적절한 공간적 해상도를 바탕으로 통행자의 통행속도를 고려한 이동궤적을 식별하기 위해서 5분이라는 기준 시간을 정하고 해당 시간대에 기록된 기지국의 기록빈도를 기반으로 통행자의 실제 위치를 추정하고자 함
- 짧은 시간 동안 발생한 핑퐁 핸드오버 현상에 의한 체류지 식별률 저하 및 비현실적으로 기록된 이동궤적을 보정하기 위하여 위해 특정 시간 단위(5분)별로 데이터를 분할

체류시작시간	체류종료시간	기지국	체류시간
9:00	9:00	A	1
9:01	9:02	B	2
9:03	9:03	C	1
9:04	9:07	B	4
9:08	9:08	A	1
9:09	10:05	B	57
10:06	10:07	A	2
...	...	...	...



체류시간대	통행자 위치
09:00~09:04	$(1A+3B+1C) / 5 = X^1$
09:05~09:09	$(1A + 4B) / 5 = X^2$
...	...

<그림 2- 3> 시간적 데이터 보정 방법 예시

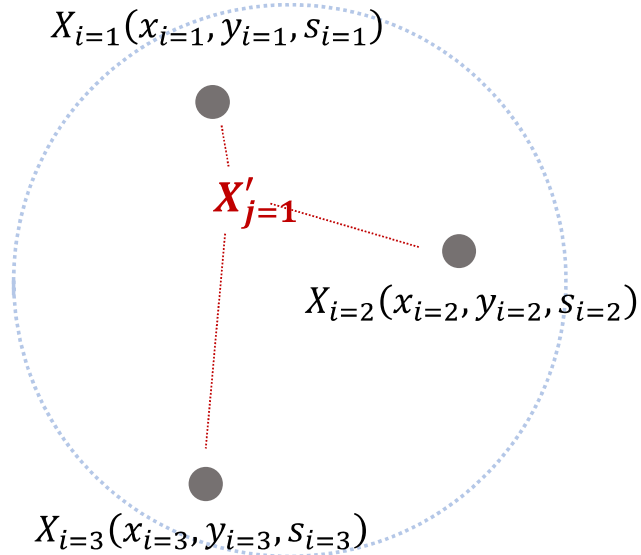
<sup>1)</sup> T. Liu, Z. Yang, Y. Zhao, C. Wu, Z. Zhou, Y. Liu, "Temporal understanding of human mobility: A multi-time scale analysis, PLOS ONE, Nov 2018

- 데이터를 분할 한 후, 각 시간 단위(time slice) 내에 기록된 기지국 데이터의 기록빈도를 기반으로 통행자의 실제 위치 추정

※ 통행자의 실제 위치 추정 방법: <그림 2-3>에서 해당 time slice 내에는 3개의 기지국 좌표(A, B, C)가 기록되어 있으며, 통행자의 실제 위치를 추정할 때는 각 기지국에서의 체류시간을 기반으로 가중치를 주어 통행자 위치 좌표(X)를 계산

- 시간 단위  $a$ 로 각 시그널 정보  $X_i(x_i, y_i, s_i)$ 를 그룹화한 후, 각 그룹별 시그널 정보의 위치( $x_i, y_i$ )와 체류시간( $s_i$ )을 가중치로 삼각 측량하여 통행자의 실제 위치  $X'_j(x'_j, y'_j, s'_j)$ 를 추정. 단,  $i$ 은 각 시간 단위  $a$ 안에 있는 시그널 로그 순서( $i = 1, 2, \dots, n$ ),  $j$ 은 보정된 로그 순서( $j = 1, 2, \dots, m$ ).

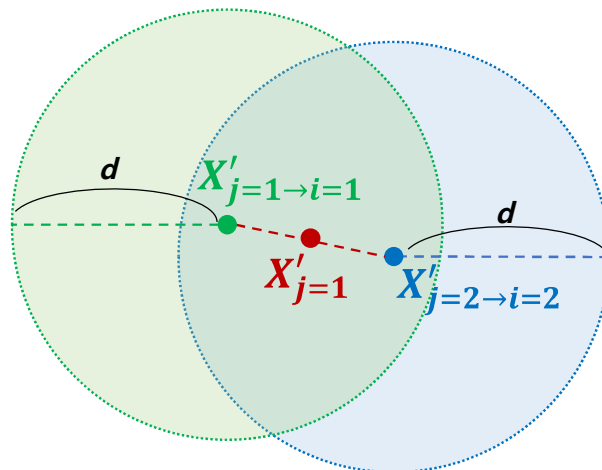
$$X'_j(x'_j, y'_j) = X'_j\left(\frac{\sum_{i=1}^n x_i \times s_i}{\sum_{i=1}^n s_i}, \frac{\sum_{i=1}^n y_i \times s_i}{\sum_{i=1}^n s_i}\right) \quad \text{for } j = 1, 2, 3, \dots, m \quad \text{식 (1)}$$



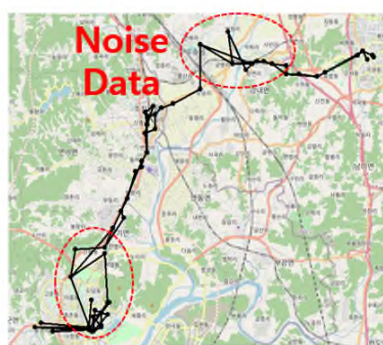
<그림 2- 4> 통행자 실제 위치 추정 방법 예시

### 나. 공간적 데이터 보정

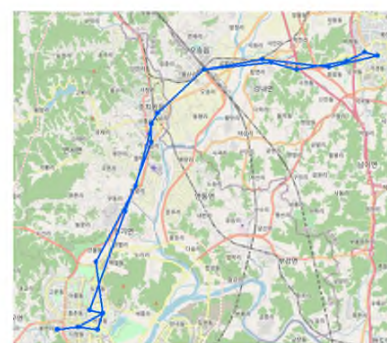
- Browning R. C., Baker E. A., Herron J. A., and Kram R.(2006)<sup>2)</sup>이 제시한 보행자 평균 도보 속도(5km/h)를 기준으로 5분(시간적 데이터 보정 시 설정한 시간 단위) 거리(약 416m) 범위 내 위치한 체류지들에 대해 각 체류지에서의 체류시간을 기반으로 해당 시간대에서의 통행자 실제 위치를 추정하고자 함
- 즉, 특정 시간 단위  $a$ 로 보정된 각 시그널 정보  $X'_j$ 의 위치가 최소 이동 속도  $v_{\min}$  기준  $a$ 시간 안에 도달할 수 있는 거리  $d$ 보다 작으면, 식 (1)을 이용하여 위치를 다시 보정함
- 원천데이터에 대해 시·공간적으로 신호 이상 기록을 보정해줌으로써, 분 단위로 연속해서 변화하거나 좁은 공간적 범위 내에 산발적으로 기록되는 기지국 위치 보정 가능



<그림 2- 5> 통행자 실제 위치 추정 방법 예시



<데이터 전처리 전 통행 궤적>



<데이터 전처리 후 통행 궤적>

<그림 2- 6> 시·공간적 데이터 보정(전처리) 적용 전, 후 통행 궤적 비교

<sup>2)</sup> R. Browning, E. A. Baker, J. A. Kram, "Effects of obesity and sex on the energetic cost and preferred speed of walking", Journal of Applied Physiology, 2006



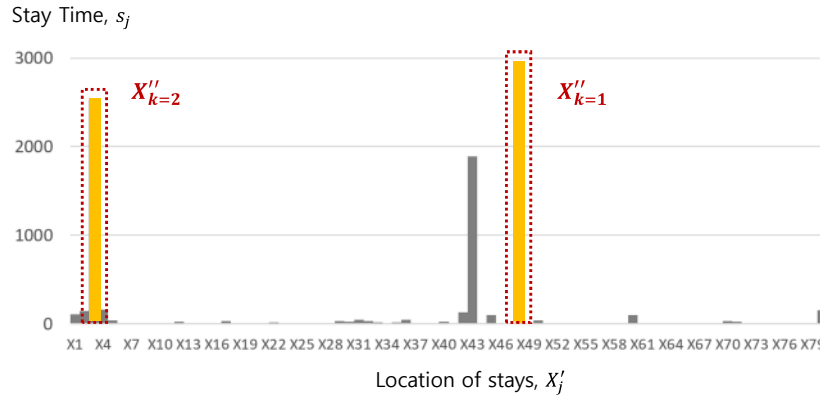
### 제3절 체류정보 및 통행목적 추정 알고리즘 고도화

#### 1. 체류지 식별 및 체류 정보 부여

##### 가. 주체류지 식별 및 점유시간 산정

- 데이터 전처리 단계에서 추정한 통행자의 실제 위치( $X'_j(x'_j, y'_j, s'_j)$ )를 기준으로 통행자가 일주일 동안 각 위치에 몇 분 체류하였는지 계산 후, 최장체류지( $X''_{k=1}(x''_{k=1}, y''_{k=1}, s''_{k=1})$ ) 식별

$$X''_{k=1}(x''_{k=1}, y''_{k=1}, s''_{k=1}) = [X'_j(x'_j, y'_j, s'_j) \mid \max(s'_j)] \quad \text{식 (2)}$$



<그림 2- 7> 최장체류지 식별 예시

- 통행자가 최장체류지 이외의 체류지들에서 연속적으로 체류한 시간을 일별로 계산하여 그중에서 가장 긴 시간을 식별

체류시작시간	체류종료시간	체류시간	통행자 위치
00:00	04:56	04:57	최장 체류지
05:01	05:02	00:02	X <sup>~</sup> 1
05:02	09:03	04:02	최장 체류지
09:04	09:06	00:03	X <sup>~</sup> 2
09:07	09:10	00:04	X <sup>~</sup> 3
09:10	09:15	00:06	X <sup>~</sup> 4
09:20	17:40	08:21	X <sup>~</sup> 5
17:40	17:45	00:06	X <sup>~</sup> 6
17:45	17:50	00:06	최장 체류지
17:50	17:54	00:05	X <sup>~</sup> 7
18:00	21:30	03:31	최장 체류지
...	...	...	...

→ 시간 1: 6분

← 시간 2가 가장 김

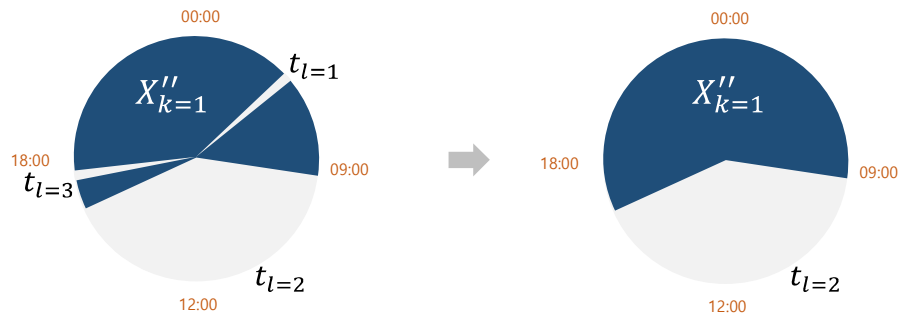
→ 시간 2: 8시간 42분

→ 시간 3: 10분

<그림 2- 8> 최장체류지 이외의 체류지에 체류한 시간 중 가장 긴 시간 식별 예시

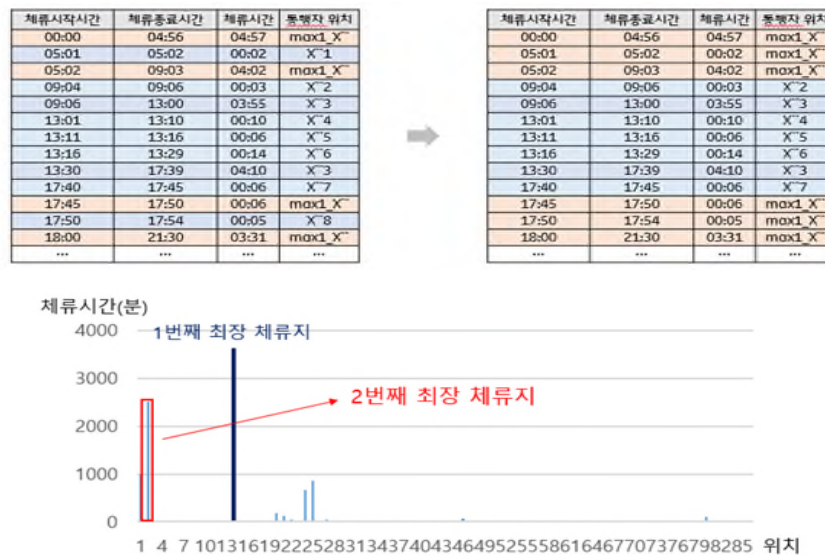
- 하루 24시간 중, 첫 번째 주 체류지  $X''_{k=1}(x''_{k=1}, y''_{k=1}, s''_{k=1})$  이외의 점유시간  $t_l$  중 최장 점유시간만 남겨놓고 나머지 시간을 첫 번째 주 체류지의 점유시간으로 산정함

$$X''_{k=1}(x''_{k=1}, y''_{k=1}, s''_{k=1}) = X_{k=1}(x''_{k=1}, y''_{k=1}, s''_{k=1} + \sum_{l=1}^r t_l - \max(t_l)) \quad \text{식 (3)}$$



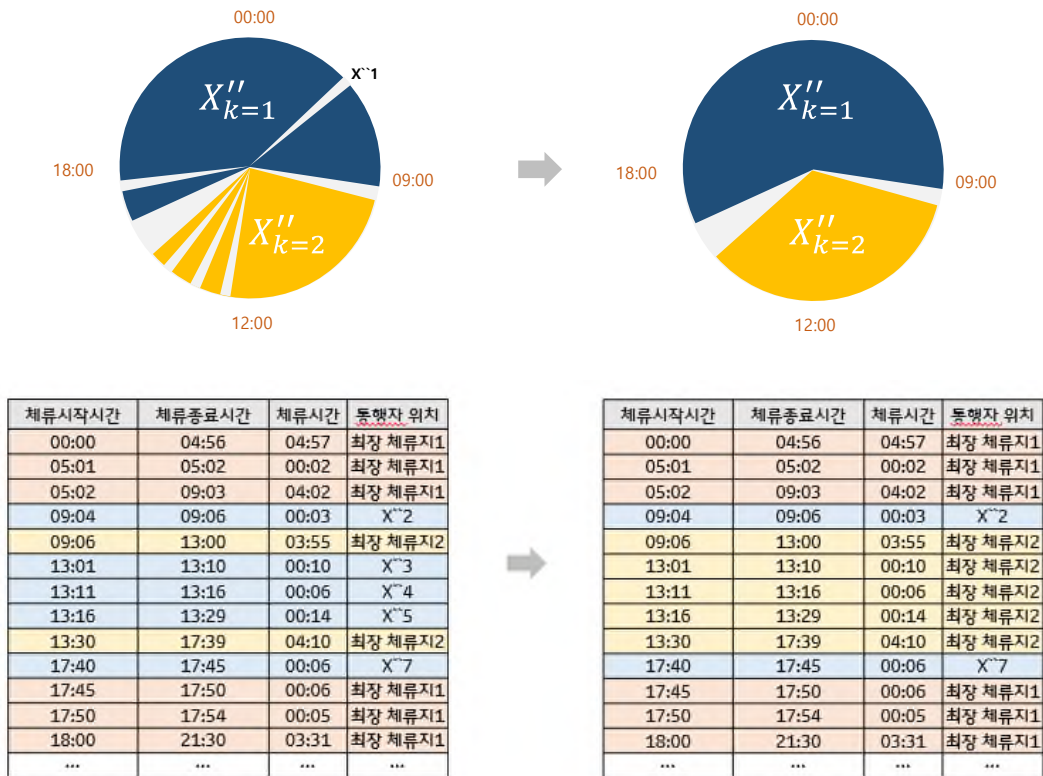
<그림 2- 9> 첫 번째 주 체류지(최장체류지) 점유시간 산정 예시

- 해당 시간 이외에는 통행자가 최장체류지에 체류한 것으로 지정하며, 해당 시간에 속하는 체류지들에 대하여 일주일 기준 체류시간이 가장 긴 체류지 식별(2번째 최장체류지 식별 시, 최장체류지로 지정된 체류지들을 후보에서 제외하고 식별)



<그림 2- 10> 최장체류지 이외 체류지 중 최장체류지 식별

- 일별로 2번째 최장체류지가 처음 기록된 시각(t1)과 마지막으로 기록된 시각(t2)를 식별한 후, t1부터 t2까지 통행자가 2번째 최장체류지에 체류한 것으로 보정(단, t1부터 t2 사이에 1번째 최장체류지로 지정된 체류지가 존재하는 경우, 2번째 최장체류지가 없는 것으로 판단)



<그림 2- 11> 두 번째 주 체류지 식별 및 점유시간 산정 예시

- 2개의 최장체류지에 대해서 일주일 중 발생 빈도를 계산하여, 거주지 및 근무지 식별
  - ※ 발생 빈도가 더 많은 것을 거주지, 적은 것을 근무지로 식별하며 발생 빈도가 같은 경우, 일주일 기준 체류시간이 더 긴 것을 거주지, 적은 것을 근무지로 식별
- 근무지로 식별된 체류지에 대해 일주일 중 근무지에 체류한 시간대를 계산하여 동일 시간대에 주 2회, 회당 3시간 미만 체류한 경우는 근무지에서 제외
- 거주지와 근무지 위치는 각 주체류지로 구분된 최장체류지의 위치로 지정

## 2. 통행 정보 추정

### 가. 통행목적 분류

- 정의된 통행에서 출발 체류지 유형과 도착 체류지 유형을 통해 해당 통행에 대한 목적을 정의함
- 예를 들어, 출발 체류지 유형이 집, 도착 체류지 유형이 회사일 경우, 통행목적은 출근이라고 구분하고, 출발 체류지 유형이 회사, 도착 체류지 유형이 집을 경우, 통행목적은 퇴근이라고 구분함

<표 2- 1> 체류지 유형별 식별 기준

체류지 유형		식별 기준			
		체류 특성		통행자 연령	비고
		체류시간	체류 빈도		
주 체류지	집	알고리즘 기준 최장 체류지 후보 1, 2 중	일주일 기준 최다빈도 체류지	-	첫 번째 주 체류지
	회사	동시간대 3시간 이상	주 2회 이상	30세 이상 60세 미만	두 번째 주 체류지 중
	학교	동시간대 3시간 이상	주 2회 이상	20세 미만	두 번째 주 체류지 중
잠재 체류지	정기적	3시간 이상	주 2회 이상	-	두 번째 주 체류지 중 회사나 학교로 구분되지 않은 체류지 중
	비정기 적	3시간 이상	주 2회 미만	-	

## 나. 통행시간/속도 추정

- 통행시간의 경우, 거주지와 근무지에서의 체류시작시간과 체류종료시간을 기반으로 출근 및 퇴근 시간을 계산

※ <그림 2-11>에서 출근 시간은 09:03부터 09:30까지인 27분으로 계산하며, 퇴근 시간은 17:39부터 18:25까지인 45분으로 계산

체류시작시간	체류종료시간	체류시간	통행자 위치
00:00	04:56	04:57	거주지
05:01	05:02	00:02	거주지
05:02	09:03	04:02	거주지
09:04	09:06	00:03	X`1
09:07	09:29	00:23	X`2
09:30	13:00	03:31	근무지
13:01	13:10	00:10	근무지
13:11	13:16	00:06	근무지
13:16	13:29	00:14	근무지
13:30	17:39	04:10	근무지
17:40	17:45	00:06	X`3
17:45	18:20	00:36	X`4
18:21	18:25	00:05	X`5
18:25	18:30	00:06	거주지
18:30	18:34	00:05	거주지
18:34	21:53	03:20	거주지
...	...	...	...

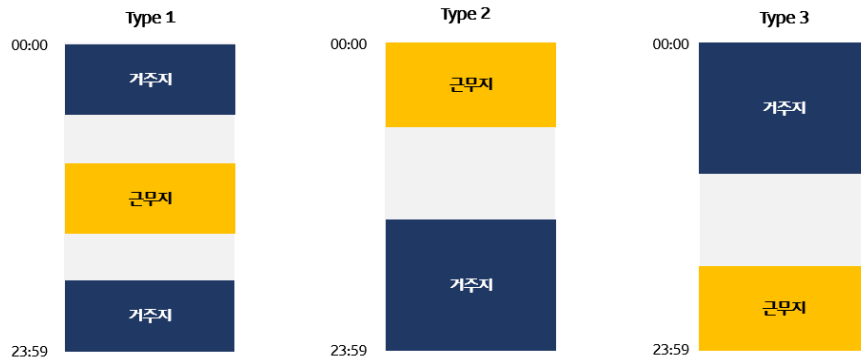
<그림 2- 12> 통행시간 계산 및 잠재체류 식별 예시

- 통행량의 경우, 주체류지(거주지 및 근무지)를 기준으로 trip type별 일별 통행량 계산

※ type 1) [거주지-근무지-거주지]인 경우, 일별 통행량=2

type 2) [근무지-거주지]인 경우, 일별 통행량=1

type 3) [근무지-거주지]인 경우, 일별 통행량=1



<그림 2- 12> 주체류지 기준 trip type 구분

	순번	고객식별번호	기준일	체류시작시간	체류종료시간	기지국	체류시간	
거주지 체류	206	164	20160525	0010	0158	I	108	
	207	164	20160525	0212	0212	M	0	
	208	164	20160525	0224	0318	I	54	
	209	164	20160525	0413	0413	M	0	
	210	164	20160525	0507	0507	I	0	
	211	164	20160525	0601	0723	M	82	
	212	164	20160525	0808	0808	I	0	
	213	164	20160525	0842	0842	M	0	
	214	164	20160525	0901	0909	I	8	
	215	164	20160525	0910	0911	M	1	
	216	164	20160525	0915	0917	I	2	
근무지 체류	217	164	20160525	0920	0924	N	4	1통행
	218	164	20160525	0925	0929	L	4	
	219	164	20160525	0932	0932	D	0	
	220	164	20160525	0935	1323	B	228	
	221	164	20160525	1329	1343	A	14	
	222	164	20160525	1345	1949	B	364	
	223	164	20160525	2006	2014	L	8	1통행
	224	164	20160525	2015	2019	I	4	
	225	164	20160525	2020	2025	N	5	
	226	164	20160525	2026	2054	I	28	
	227	164	20160525	2058	2108	M	10	
거주지 체류	228	164	20160525	2113	2114	I	1	
	229	164	20160525	2128	2128	M	0	
	230	164	20160525	2135	2143	I	8	
	231	164	20160525	2147	2158	M	11	
	232	164	20160525	2208	2337	I	89	
	233	164	20160525	2343	2343	M	0	
	234	164	20160525	2354	2354	I	0	

<그림 2- 13> 일별 데이터에 대한 주체류지 간 통행량 산정 예시

- 출근 시간과 퇴근 시간 사이에 30분 이상 체류한 기록이 있으면 잠재체류로 식별
- ※ <그림 2-14>에서 출근 시간 사이에 16:31부터 17:16까지 1회 30분 이상 체류한 기록이 있음 → 잠재체류로 판단
- 출근 또는 퇴근 중 잠재체류가 존재하는 경우 출근 시간 및 퇴근 시간 산출 대상에서 제외

	순번	고객식별번호	기준일	체류시작시간	체류종료시간	기지국	체류시간	
거주지 체류	17	9939	20160520	0004	0503	T2	300	
	18	9939	20160520	0507	0518	P2	12	
	19	9939	20160520	0519	0519	K2	1	
	20	9939	20160520	0520	0524	F2	5	
	21	9939	20160520	0525	0529	Z	5	
	22	9939	20160520	0530	0534	T	5	
	23	9939	20160520	0539	0540	J	2	
	24	9939	20160520	0541	0544	N	4	
	25	9939	20160520	0545	1554	P	610	
	26	9939	20160520	1555	1558	W	4	
	27	9939	20160520	1603	1603	R2	1	
	28	9939	20160520	1619	1628	Q2	10	
	29	9939	20160520	1631	1716	M2	46	← 잠재체류지
	30	9939	20160520	1716	1720	Q2	5	
근무지 체류	31	9939	20160520	1752	2358	T2	367	

← 잠재체류지

&lt;그림 2- 14&gt; 잠재체류 식별 예시



- 출근 속도의 경우, 거주지와 근무지 간 거리를 출근 시간으로 나누어 계산하고 퇴근 속도의 경우, 거주지와 근무지 간 거리를 퇴근 시간으로 나누어 계산
  - 출·퇴근 속도를 기준으로 이상치를 식별하고 출근 시간 및 퇴근 시간 산출 대상에서 제외
    - 출근 속도 또는 퇴근 속도가 평균 도로 속도인 5km/h 미만인 경우, ‘data lagging’ 등 데이터 수집 지연 현상에 의한 이상치로 판단
- ※ <그림 2-15>는 data lagging이 발생한 예시를 나타내며, 13:28부터 14:11에 해당하는 데이터 로그가 없음. 그림에서 파란색으로 표시된 데이터 로그가 거주지, 빨간색으로 표시된 데이터 로그가 근무지에 해당하는 로그이며 퇴근 시간이 데이터 로그가 없던 시간인 13:28부터 14:11까지인 43분으로 과대추정될 수 있음
- 출근속도 또는 퇴근 속도가 300km/h 이상인 경우, 신호 이상에 의한 체류지 식별 오류로 판단하여 출근 시간 및 퇴근 시간 산출 대상에서 제외
- ※ KTX 이동속도(300km/h)를 고려하여 임계치 설정

				체류 시작시간	체류 종료시간	경도	위도
21	20190402	387	50	2100	2318	127.0613	37.24505
22	20190403	387	50	12	1126	127.0613	37.24505
23	20190403	387	50	1220	1328	127.067	37.24596
24	20190403	387	50	1411	1730	127.0613	37.24505
25	20190403	387	50	1815	1815	127.0613	37.24505
26	20190403	387	50	1822	2334	127.0613	37.24505
27	20190404	387	50	29	719	127.0613	37.24505

<그림 2- 15> Data Lagging 예시



- 원천데이터에 대해 본 알고리즘을 적용하면 <그림 2-15>과 같은 형태의 통행지표가 산출되며, 해당 산출물에는 개별 고객에 대한 일자별 통행지표가 포함되어 있음

날짜	2020-04-01
고객식별번호	112233
나이	30분
성별	M
출근 시작시간	8 : 30
출근 종료시간	9 : 00
출근 소요시간	30분
출근 중 잠재체류 회수	0회
출근 중 잠재체류 시간	0분
퇴근 시작시간	18 : 27
퇴근 종료시간	19 : 30
퇴근 소요시간	63분
퇴근 중 잠재체류 회수	1
퇴근 중 잠재체류 시간	35
집 위치	127.491, 36.455
회사 위치	127.485, 36.607
집/회사 간 거리	16.9 km
출근시간	30분
퇴근시간	X
출근속도	33.8 km/h
퇴근속도	X

<그림 2- 16> 통행지표 산출물 예시

## 제4절 가공알고리즘 데이터 최적화 및 속도 개선

### 1. 가공알고리즘 속도 개선 방안

#### 1) 데이터베이스 최적화 방안

- 데이터베이스는 점차적으로 증가하는 대용량데이터를 고려하여, 안정적이고 효율적으로 관리 할 수 있도록 대용량 데이터베이스 설계
- 모바일통신 빅데이터의 특성을 고려하여 확장 가능한 설계, 다양한 유형별 자료를 효과적으로 구축 할 수 있는 구조로 설계
- 효율적 데이터 처리 방식을 이용하여 디스크 자원을 최소한으로 줄이며, 데이터의 액세스를 분석하여 DB성능개선 전략을 수립함
- 데이터베이스 설계 방법론에 입각한 현행 시스템 및 데이터를 분석하고 표준화 방안 등을 마련하여 시스템에서 요구하는 데이터 분석이 가능한 구조로 설계

#### 2) 알고리즘 최적화 방안

- 모바일통신 빅데이터 가공알고리즘은 대용량 원시 LTE 로그데이터를 사용하여 거주지 및 근무지 추정하고 해당 위치에서의 체류시간을 기반으로 통근 시간 및 통행량 등 통행지표 산출하는 과정으로 진행됨
- 원시데이터 전처리 과정에서는 기지국과 단말기 간에 발생하는 불규칙적인 신호 이상 현상을 보정해야 하고, 통행자의 이동궤적을 식별하는 단계에서는 비현실적인 데이터를 시·공간적으로 보정해야 함
- 따라서 효율적인 데이터 산출을 위해서 데이터의 경량화, 알고리즘의 고도화 및 분산 처리 등 데이터 생산 효율을 높이는 것이 주요 이슈임

## 2. 가공알고리즘의 단계별 세부 개선 내용

- 속도 개선을 위한 단계별 세부 알고리즘 최적화 및 수정내용은 다음과 같음

<표 2- 2> 개선 알고리즘 단계별 코드 구성 내용

구분	순번	내용
Stage 1. 데이터 전처리 및 통행자 실제 위치 추정	1	5분 단위 그룹을 만들어 좌표 가중평균화
	2	도보 속도 5km 이하 좌표, 다음 좌표와 체류시간 가중치로 보정 코드 최적화
	3	도보거리 416.7m 이하 좌표, 다음 좌표와 체류시간 가중치로 보정 코드 최적화
Stage 2. 주체류지 식별	4	일주일 기준 최장체류지 1, 2 추출(1차적으로 1을 집으로 정의) 코드 최적화
	5	최장체류지 1, 2와 좌표 간 거리 연산 후 500m 이하 좌표는 최장체류지 1, 2로 변경(DBSCAN 개념 차용)
Stage 3. 통행시간 및 통행량 산출	6	30분 이하 체류지는 제외(체류지 시간 최소기준 조작적 정의)
	7	행 압축
	8	최장체류지 1, 2 기준 OD속성 부여 코드 최적화

<표 2- 3> 전체 프로세스별 소요시간

순번	항목	내용	소요시간 (일주일 기준)	1년치 소요 시간	비고
1	KOTI_RAW_FINAL.hql	원천데이터 수집 및 저장	0.8일	41.6일	병렬처리 적용하지 않은 소요시간
2	FINAL_RUN.r	중간 산출물	2.8일	146일	병렬 수행한 결과이며 시간 계산 여유있게 한 결과
3		최종 산출물	3.2시간	7일	실제 수행 시간은 위의 중간산출물 시간에 포함 됨.
4	MASKING.r	개인 식별 위험 데이터 마스킹 처리	30분 미만	21시간 미만	병렬처리 적용하지 않은 소요시간
5	UPLOAD	데이터 반출을 위한 업로드	2.4시간	5.2일	대략적인 시간 기준

&lt;표 2- 4&gt; 산출물 상세 소요시간

순번	내용	1년치 소요 시간	비고
1	주2회 3시간 동일시간대 조건 알고리즘에 사용되는 함수	8.5시간	
2	원천데이터 raw data 필드 맞추기	9.8일	
3	CSID	7시간	
4	고객 1명 7일치 데이터 추출	1.8일	
5	원천데이터를 1분단위로 펼친 데이터 (dfs)	1.7일	
6	5분단위 time slice로 데이터를 나누고 각 time slice에서 대표위치 찾은 데이터(dfm)	6.6일	시간보정
7	이웃한 행간 속도 측정 후 5km/h 미만이면 체류시간 기준으로 위치보정	10.5일	속도보정
8	거리보정 (416m)	27.5일	공간보정
9	원천데이터에 대해 시공간적으로 보정+행압축된 데이터(a_1)	4.2일	행압축
10	a_1을 데이터테이블에서 데이터프레임으로 데이터 형식 변경	2일	
11	최장체류지1 찾기	2.5일	
12	변수 이름을 df2에서 a_1으로 다시 변경	1일	
13	max gap time을 찾기 위한 코드	11일	find_max_gap : 일별로 max gap을 찾기 위해 최장체류지가 아닌 체류지에 대한 체류시작시간, 체류종료시간이 기록됨
14	일자별 데이터에 대해 트립 타입 (집-회사-집, 집-회사, 회사-집 등)을 구분하고 max gap에 해당하는 index를 not_h라고 지정	4시간	
15	위와 같은 과정의 결과로 h_a_1과 c_a_1이 생성되며, c_a_1은 일주일동안 max gap에 해당하는 데이터들의 모음	5시간	
16	최장체류지 2를 찾기 위한 과정	3.5일	최장체류지1을 찾기 위한 과정과 동일, input 데이터만 c_a_1으로 바뀜
17	new_a_1 데이터프레임에서 최장체류지 1에 해당하는 데이터를 home='YES', 최장체류지 2에 해당하는 데이터를 home='NO'로 표시	16시간	
18	일별로 가장먼저 기록된 두번째 최장체류지 로그부터 가장 마지막에 기록된 로그까지 최장체류지 2로 지정해주는 코드	1.7일	

19	일주일에 대해 일기록빈도를 기준으로 회사와 집을 식별하는 코드	2일	기록빈도가 더 많은게 집, 기록빈도 같은경우 체류시간이 더 긴게 집
20	회사와 집 식별 후, 회사에 대해 주2회 3시간(동일시간대) 미만 체류시 회사가 없는 것으로 지정하는 코드	8.5일	
21	출퇴근 시간 및 잠재체류를 계산하기 위한 코드	36일	NEW_HC_Df 가 고객별 산출물에 해당하는 데이터프레임
22	트립타입 구분 코드	14시간	트립 타입1은 집-회사-집 트립타입2는 집-회사 트립타입3은 회사-집
23	1인에 대한 출·퇴근 시간 및 회사 집 위치 정보가 포함된 산출물	9시간	
24	모든고객(iii)에 대한 산출물	2시간	
25	회사나 학교로 이동하는 고객에 대해 출장 통행으로 추정되는 로그 식별 (출근 횟수 2회 미만 & 퇴근 횟수 2회 미만)	1.7일	
26	잠재체류 있는 경우 통행시간 산출에서 제외	3.8일	잠재체류 식별기준: 출퇴근 중 30분 이상 체류한 기록이 있음
27	회사(목적지)-집 거리 구하기	1.5일	
28	출근 속도 및 퇴근 속도 구하기	1시간 미만	시간에 30초 더해주기
29	출근 속도나 퇴근 속도가 5km/h 미만인 경우 출퇴근 시간을 NA 처리	2.5일	
30	출근 속도나 퇴근 속도가 200km/h 초과인 경우 출퇴근 시간을 NA 처리	2.6일	
31	통행량의 경우, 행별로 트립타입이 집-회사-집인 경우는 2, 집-회사 또는 회사-집인 경우는 1로 계산하면 됨	3.3일	HC개수, CH개수 count
32	잠재 통행량의 경우, 행별로 출퇴근 모두 잠재체류가 있으면 2, 출근 또는 퇴근 중 잠재체류가 하나 있으면 1, 둘 다 없으면 0	4일	

## 제3장 모바일통신 빅데이터 기반 통행 DB 구축 및 검증

---

제1절 기준년도 통행DB 구축

제2절 구축 DB 검증



## 제3장 모바일통신 빅데이터 기반 통행 DB 구축 및 검증

### 제1절 기준년도 통행DB 구축

#### 1. DB 설계

- 「개인정보보호법」과 「위치정보의 보호 및 이용 등에 관한 법률」에 따라 개인의 위치와 이동궤적을 추적할 수 없도록 기반 DB는 다음 <표 3-1>과 같이 집계된 형태로 구축하고자 함

<표 3- 1> 모바일통신 빅데이터 DB 형태

출발						도착				
날짜	요일*	시간	폴리곤	시도	체류지 유형**	날짜	시간	폴리곤	시도	체류지 유형**
20180402	월	13	48521351	11	집	20180402	14	65421584	11	회사
20180403	화	06	54754213	21	집	20180403	08	32158421	11	회사
20180404	수	09	15486211	33	집	20180404	09	15486213	33	학교

총통행량	보정계수	성별	연령	잠재통행 고려 안함								
				통행량(비율)				시간			거리 평균	속도 평균
				통행 비율	30분 이상	60분 이상	120분 이상	평균	분산	중앙 값		
10	0.34	M	20	1	0.7	0.5	0.4	35	6.98	26	8.24	29.8
25	0.57	F	50	0.8	0.9	0.8	0.2	33	6.77	24	9.45	11.9
8	0.47	M	10	1	0.9	0.6	0.1	25	4.32	19	10.5	23.5

잠재통행 30분 고려 안함									
통행량(비율)					시간			거리 평균	속도 평균
통행비율	잠재체류 횟수	30분 이상	60분 이상	120분 이상	평균	분산	중앙값		
0.8	0	0.8	0.6	0.8	35	6.98	26	5.89	36.9
0.7	3	0.7	0.5	0.2	-	-	-	9.57	10.4
0.8	2	0.8	0.5	0.1	-	-	-	10.12	20.3

잠재통행 60분 고려 안함									
통행량(비율)					시간			거리 평균	속도 평균
통행비율	잠재체류 횟수	30분 이상	60분 이상	120분 이상	평균	분산	중앙값		
0.9	0	0.7	0.6	0.5	35	6.98	26	7.4	33.6
0.8	2	0.8	0.6	0.2	-	-	-	9.9	10.9
0.9	0	0.9	0.6	0.1	25	4.32	19	10.1	22.8

주 \*) 요일: 출발 및 도착 요일 (1~7, 일~월)

주 \*\*) 유형: 개인별 출발지 및 도착지 유형 (H: 집, C: 회사, S: 학교, R: 정기 잠재체류지, X: 기타 잠재체류지)



- 개인의 통행 정보가 노출되지 않도록 기점 정보(출발 일자, 출발 시간대, 출발 위치, 출발 체류지 유형), 종점 정보(도착 일자, 도착 시간대, 도착 위치, 도착 체류지 유형), 통행자 정보(성, 연령)를 기준으로 통행 정보가 같은 인구를 집계
  - 일자 정보는 1일 단위로 구축하도록 함
  - 출발 시간 및 도착 시간 정보는 집계된 인구가 3명 이하가 될 가능성을 줄이면서, DB 사용자에게 가능한 섬세한 통행 정보를 제공할 수 있도록 한 시간 단위로 입력하도록 함
  - \* 출발 시간은 출발지에서의 체류종료시간을, 도착 시간은 도착지에서의 체류시작시간을 의미
  - 출발지와 도착지 위치는 「위치정보의 보호 및 이용 등에 관한 법률」에 저촉되지 않도록 기지국 좌표 그대로 노출하지 않고 매칭되는 교통폴리곤(교통폴리곤은 기지국의 수신범위를 고려하여 구축한 모바일 기반 DB 분석 단위) ID로 변환하도록 함
  - 출발 및 도착 체류지 유형은 평소 통행자가 해당 위치에 얼마나 자주, 오래 머무르는지를 추정할 수 있는 정보를 제공하기 위한 것으로, 「2019년 국가교통조사」에서 설정한 체류지 식별 기준에 따라 출발지(출발 기지국 좌표)와 도착지(도착 기지국 좌표)의 특성을 확인한 후 각각 체류지 유형을 구분하여 입력하도록 함
  - \* 「2019년 국가교통조사」에서 구분하고 있는 체류지 유형 중에서 집 이외 심야 시간대 주체류지는 제외하고, 총 5가지 유형으로 구분 (집, 회사, 학교, 정기적 잠재체류지, 비정기적 잠재체류지)
  - 통행자에 대한 정보를 확인할 수 있도록 성, 연령 필드 구성하고, 통행 정보가 지나치게 세분되지 않도록 가능한 통행 특성이 유사한 계층을 묶어 10세 단위로 연령을 구분하여 입력하도록 함
  - 통행량 필드에는 전술한 기준에 의해 집계된 인구(단말기 수)를 입력하되, 집계된 통행량이 3 이하의 값을 갖는 경우, 2장에서 개발한 3통행 미만 보정계수를 적용하여 추정된 통행량을 입력하도록 함
- 새로운 지표 개발을 위해 통행시간 필드를 추가
  - 통행시간 필드에는 집계된 통행 정보를 기준으로 산출한 평균 통행시간 값을 입력

&lt;표 3- 2&gt; 모바일통신 빅데이터 DB 테이블 정의서

컬럼 ID	컬럼명	Type	비고
o_base_ymd	출발 일자	string	yyyymmdd
o_base_dow	출발 요일	string	0 ~ 6 : 일요일 ~ 토요일
o_timezn_cd	출발 시간대	string	00~23 (1시간 단위)
o_polygon	출발 폴리곤ID	string	-
o_trip_type	출발 체류지 유형	string	H: 집, C: 회사, S: 학교, R: 정기 잠재체류지, X: 기타 잠재체류지
d_base_ymd	도착 일자	string	yyyymmdd
d_timezn_cd	도착 시간대	string	00~23 (1시간 단위)
d_polygon	도착 폴리곤ID	string	-
d_trip_type	도착 체류지 유형	string	H: 집, C: 회사, S: 학교, R: 정기 잠재체류지, X: 기타 잠재체류지
age_cd	연령	integer	0~110 (예: 10: 10세 이상 20세 미만)
sex_type_cd	성	string	M: 남성, F: 여성
travel_volume	총 통행량	integer	횟수 단위
travel_time_mean	통행시간_평균	integer	1분 단위
travel_time_med	통행시간_중앙값	integer	1분 단위
travel_time_var	통행시간_분산	integer	-
travel_distance	통행거리	integer	km 단위
travel_speed	통행속도	integer	km/h 단위

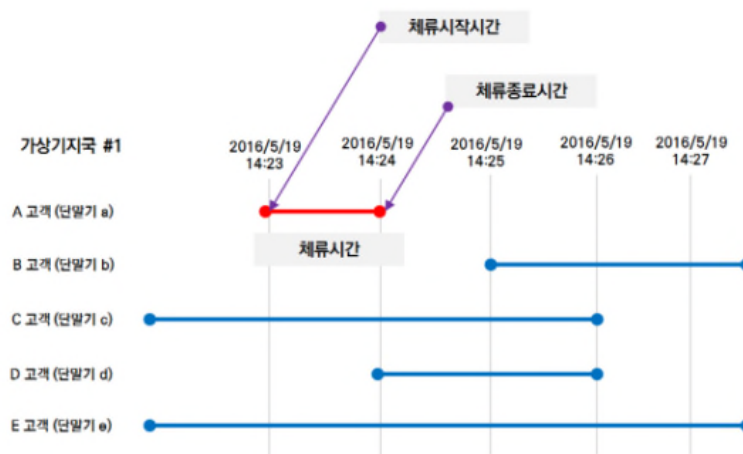
## 2. DB 구축 방법

### 1) 원천데이터 수집 및 이상치 제거

- 2020년 1월 1일부터 2020년 12월 31일까지 기록된 로그를 수집
- 기지국 좌표가 Null값이거나 행정동 정보가 매칭되지 않는 데이터를 제거

### 2) 선분 이력으로 변환

- 포인트 단위의(밀리세컨 단위) 로그를 선분 단위의(분 단위) 로그로 변환
  - 닷(dot) 형태로 기록된 로그를 기록된 시간 순서대로 연결한 후, 최초 기록된 시간(이하 ‘체류시작시간’)과 마지막으로 기록된 시간(이하 ‘체류종료시간’)정보만 추출하여 단말기가 해당 기지국에 식별된 시간(이하 ‘체류시간’)을 산출
  - 체류시간은 체류종료시간에서 체류시작시간을 뺀 값이며, 초(sec) 단위는 생략



주: 자료를 참고하여 도식화 함

자료: 박미율·주은정(2017), LTE 시그널 정보를 이용한 위치정보가공 및 유동인구집계 방법, 한국통신학회 2017년도 동계종합 학술발표회, p.285.

<그림 3- 1> 선분 이력으로의 변환 예시

### 3) 통행 정보 왜곡 데이터 보완

- 데이터 보완 범위를 한정하기 위해 로그 기록을 한 달 단위로 구분
  - 통행 정보 왜곡 데이터를 탐색하고 보완하는 범위를 1개월로 한정 (예: 4월 1일부터 4월 30일까지의 로그 기록을 연결하여 탐색하고 보완)
- 통행 정보를 왜곡할 수 있는 ‘필요 이상의 핸드오버 데이터(Unnecessary

Handover)', '핑퐁 핸드오버(Ping-pong Handover)로 인한 데이터'를 본 과업에서 개발한 전처리 기술을 통해 보정함

#### 4) 체류 순서 부여

- 로그 기록 일자, 체류시작시간, 체류종료시간을 기준으로 로그 기록 순서에 따라 개인별 데이터를 정렬하여 체류 순서를 구분
- 고객식별번호(단말기 구분 코드)와 데이터 생성일자를 기준으로 KEY 값을 갖는 필드를 형성한 후('기준일-고객식별번호'), 같은 KEY 값을 갖는 데이터 내에서 정렬

ID	그룹 번호	KEY 칼럼 (기준일- 고객식별번호)	가상기지국		체류 시작 시간	체류 종료 시간	연 령 대	체류 시간
			x좌표	y좌표				
1	1	20160520-17232311	126.914	37.545	1400	1750	33	240
2	2	20160520-17232311	126.914	37.549	1751	1752	33	1
3	3	20160520-17232311	126.916	37.548	1754	1809	33	15
4	4	20160520-17232311	126.914	37.549	1810	1813	33	3
5	5	20160520-17232311	126.915	37.551	1813	1814	33	1
6	6	20160520-17232311	126.914	37.549	1814	1815	33	1
		⋮	⋮	⋮				
18	18	20160520-17232311	126.815	37.538	1920	2359	33	279
		⋮	⋮	⋮				
453	1	20160520-1843030	126.934	37.556	0805	1041	25	156
454	2	20160520-1843030	126.936	37.556	1041	1042	25	1
455	3	20160520-1843030	126.935	37.554	1043	1044	25	1
456	4	20160520-1843030	126.933	37.552	1045	1050	25	5
457	5	20160520-1843030	126.929	37.547	1051	1052	25	1
458	6	20160520-1843030	126.923	37.547	1053	1055	25	2
459	7	20160520-1843030	126.918	37.547	1056	1059	25	3
460	8	20160520-1843030	126.914	37.545	1100	1200	25	60
		⋮	⋮	⋮				
679	227	20160522-1843030	126.934	37.556	0003	0830	25	507
680	228	20160522-1843030	126.936	37.556	0831	0833	25	2
		⋮	⋮	⋮				
711	259	20160522-1843030	126.815	37.551	0850	1100	25	130

<그림 3- 2> 체류 순서 부여 (개인별 이동궤적 형성)

#### 5) 체류 정보 추출

- 이동 중에 기록된 로그 기록을 제외하고 체류 중에 기록된 로그 기록만을 추출
  - 체류시간이 15분을 초과하는 로그 기록만 추출
    - 트립 타임 필드를 추가하여 체류시간이 15분 이하인 경우는 ‘경로 (Pass-by)’로, 15분 초과인 경우는 ‘체류(Stay)’로 구분

ID	그룹 번호	KEY 칼럼 (기준일- 고객식별번호)	가상기지국		체류 시작 시간	체류 종료 시간	연 령 대	체류 시간	트립 타임
			x좌표	y좌표					
1	1	20160520-17232311	126.914	37.545	1400	1750	33	240	체류
2	2	20160520-17232311	126.914	37.549	1751	1752	33	1	경로
3	3	20160520-17232311	126.916	37.548	1754	1809	33	15	경로
4	4	20160520-17232311	126.914	37.549	1810	1813	33	3	경로
5	5	20160520-17232311	126.915	37.551	1813	1814	33	1	경로
6	6	20160520-17232311	126.914	37.549	1814	1815	33	1	경로
		⋮	⋮	⋮					
18	18	20160520-17232311	126.815	37.538	1920	2359	33	279	체류
		⋮	⋮	⋮					
453	1	20160520-1843030	126.934	37.556	0805	1041	25	156	체류
454	2	20160520-1843030	126.936	37.556	1041	1042	25	1	경로
455	3	20160520-1843030	126.935	37.554	1043	1044	25	1	경로
456	4	20160520-1843030	126.933	37.552	1045	1050	25	5	경로
457	5	20160520-1843030	126.929	37.547	1051	1052	25	1	경로
458	6	20160520-1843030	126.923	37.547	1053	1055	25	2	경로
459	7	20160520-1843030	126.918	37.547	1056	1059	25	3	경로
460	8	20160520-1843030	126.914	37.545	1100	1200	25	60	체류
		⋮	⋮	⋮					
679	227	20160522-1843030	126.934	37.556	0003	0830	25	507	체류
680	228	20160522-1843030	126.936	37.556	0831	0833	25	2	경로
		⋮	⋮	⋮					
711	259	20160522-1843030	126.815	37.551	0850	1100	25	130	체류

<그림 3- 3> 이동정보와 체류 정보 구분 (예시)

## 6) 체류지 유형 구분

- 체류지 식별 기준(<표 3-3> 참조)에 따라 체류지 유형을 구분하여 필드 추가

&lt;표 3- 3&gt; 체류지 유형별 식별 기준

체류지 유형		식별 기준			
		체류 특성		통행자 연령	비고
		체류시간	체류 빈도		
주 체류지	집	알고리즘 기준 최장 체류지 후보 1, 2 중	일주일 기준 최다빈도 체류지	-	첫 번째 주 체류지
	회사	동시간대 3시간 이상	주 2회 이상	30세 이상 60세 미만	두 번째 주 체류지 중
	학교	동시간대 3시간 이상	주 2회 이상	20세 미만	두 번째 주 체류지 중
잠재 체류지	정기적	3시간 이상	주 2회 이상	-	두 번째 주 체류지 중 회사나 학교로 구분되지 않은 체류지 중
	비정기 적	3시간 이상	주 2회 미만	-	

ID	그룹 번호	KEY 칼럼 (기준일- 고객식별번호)	가상기지국		체류 시작 시간	체류 종료 시간	연 령 대	체류 시간	트립 타입	체류지 유형
			X좌표	Y좌표						
1	1	20160520-17232311	126,914	37,545	1400	1750	33	240	체류	집
2	2	20160520-17232311	126,914	37,549	1751	1752	33	1	경로	-
3	3	20160520-17232311	126,916	37,548	1754	1809	33	15	경로	-
4	4	20160520-17232311	126,914	37,549	1810	1813	33	3	경로	-
5	5	20160520-17232311	126,915	37,551	1813	1814	33	1	경로	-
6	6	20160520-17232311	126,914	37,549	1814	1815	33	1	경로	-
18	18	20160520-17232311	126,815	37,538	1920	2359	33	279	체류	회사
453	1	20160520-1843030	126,934	37,556	0805	1041	25	156	체류	집
454	2	20160520-1843030	126,936	37,556	1041	1042	25	1	경로	-
455	3	20160520-1843030	126,935	37,554	1043	1044	25	1	경로	-
456	4	20160520-1843030	126,933	37,552	1045	1050	25	5	경로	-
457	5	20160520-1843030	126,929	37,547	1051	1052	25	1	경로	-
458	6	20160520-1843030	126,923	37,547	1053	1055	25	2	경로	-
459	7	20160520-1843030	126,918	37,547	1056	1059	25	3	경로	-
460	8	20160520-1843030	126,914	37,545	1100	1200	25	60	체류	기타

&lt;그림 3- 4&gt; 체류지 유형 구분 (예시)

## 7) 위치정보, 시간정보, 연령 정보 변환

- 개인의 이동궤적을 추적할 수 없도록 원천데이터에서 출발지와 도착지의 위치정보를 분석 맵의 교통폴리곤 ID로 변경한 후, 시간정보(체류시작시간, 체류종료시간)에서 분 정보를 생략한 시간대 필드를 생성하고, 1세 단위의 연령 정보를 10세 단위로 변경함

ID	그룹 번호	KEY 칼럼 (기준일- 고객식별번호)	가상기지국		체류 시작 시간	체류 종료 시간	연 령 대	체류 시간	트립 타입	체류지 유형	위치 정보 변환 (교통폴리곤 ID)	체류시작 시간대	체류종료 시간대	연령대 변환
			X좌표	Y좌표										
1	1	20160520-17232311	126.914	37.545	1400	1750	33	240	체류	집	210001	14	17	30
2	2	20160520-17232311	126.914	37.549	1751	1752	33	1	경로	-	-	-	-	-
3	3	20160520-17232311	126.916	37.548	1754	1809	33	15	경로	-	-	-	-	-
4	4	20160520-17232311	126.914	37.549	1810	1813	33	3	경로	-	-	-	-	-
5	5	20160520-17232311	126.915	37.551	1813	1814	33	1	경로	-	-	-	-	-
6	6	20160520-17232311	126.914	37.549	1814	1815	33	1	경로	-	-	-	-	-
18	18	20160520-17232311	126.815	37.538	1920	2359	33	279	체류	회사	230017	19	23	30
453	1	20160520-1843030	126.934	37.556	0805	1041	25	156	체류	집	311014	08	10	20
454	2	20160520-1843030	126.936	37.556	1041	1042	25	1	경로	-	-	-	-	-
455	3	20160520-1843030	126.935	37.554	1043	1044	25	1	경로	-	-	-	-	-
456	4	20160520-1843030	126.933	37.552	1045	1050	25	5	경로	-	-	-	-	-
457	5	20160520-1843030	126.929	37.547	1051	1052	25	1	경로	-	-	-	-	-
458	6	20160520-1843030	126.923	37.547	1053	1055	25	2	경로	-	-	-	-	-
459	7	20160520-1843030	126.918	37.547	1056	1059	25	3	경로	-	-	-	-	-
460	8	20160520-1843030	126.914	37.545	1100	1200	25	60	체류	기타	312123	11	12	20

&lt;그림 3- 5&gt; 위치정보, 시간정보, 연령 정보 변환 (예시)



## 8) 통행량 집계

- 앞서 부여한 체류 순서에 따라 출발과 도착을 구분한 다음, 기점 정보(출발 일자, 출발 시간대, 출발 위치, 출발 체류지 유형), 종점 정보(도착 일자, 도착 시간대, 도착 위치, 도착 체류지 유형), 통행자 정보(성, 연령)가 동일한 로그 기록을 집계함
- 마지막 체류지가 출발 정보로 구분되고, 도착 정보가 없는 경우 이동하지 않은 것으로 보고 ‘무통행’으로 간주함

ID	그룹 번호	KEY 컬럼 (기준일- 고객식별번호)	가상기지국		체류 시작 시간	체류 종료 시간	연 령 대	체류 시간	트립 타입	체류지 유형	위치 정보 변환 (교통플리곤 ID)	체류시작 시간대	체류종료 시간대	연령대 변환	
			x좌표	y좌표											
1	1	20160520-17232311	126.914	37.545	1400	1750	33	240	체류	집	210001	14	17	30	출발
2	2	20160520-17232311	126.914	37.549	1751	1752	33	1	경로	-	-	-	-	-	1통행 ↓
3	3	20160520-17232311	126.916	37.548	1754	1809	33	15	경로	-	-	-	-	-	
4	4	20160520-17232311	126.914	37.549	1810	1813	33	3	경로	-	-	-	-	-	
5	5	20160520-17232311	126.915	37.551	1813	1814	33	1	경로	-	-	-	-	-	
6	6	20160520-17232311	126.914	37.549	1814	1815	33	1	경로	-	-	-	-	-	
18	18	20160520-17232311	126.815	37.538	1920	2369	33	279	체류	회사	230017	19	23	30	도착
453	1	20160520-1843030	126.934	37.556	0805	1041	25	156	체류	집	311014	08	10	20	출발
454	2	20160520-1843030	126.936	37.556	1041	1042	25	1	경로	-	-	-	-	-	1통행 ↓
455	3	20160520-1843030	126.935	37.554	1043	1044	25	1	경로	-	-	-	-	-	
456	4	20160520-1843030	126.933	37.552	1045	1050	25	5	경로	-	-	-	-	-	
457	5	20160520-1843030	126.929	37.547	1051	1052	25	1	경로	-	-	-	-	-	
458	6	20160520-1843030	126.923	37.547	1053	1055	25	2	경로	-	-	-	-	-	
459	7	20160520-1843030	126.918	37.547	1056	1059	25	3	경로	-	-	-	-	-	
460	8	20160520-1843030	126.914	37.545	1100	1200	25	60	체류	기타	312123	11	12	20	도착

&lt;그림 3- 6&gt; 출발/도착 구분 및 통행량 집계 (예시)



## 9) 통행량 보정

- 통행량이 3 이하(0~3통행)인 경우 모두 3통행으로 변환한 후, 3통행 미만 보정계수(2장 참고)를 적용하여 통행량을 보정
  - 개인정보보호법에 의거, 3 통행 이하의 값은 실제 값이 아닌, 추정값을 입력
- 실제 통행 분석에 활용할 때에는 통신업체의 시장점유율을 기준으로 전수화 하여 사용
  - 과학기술정보통신부에 고시되고 있는 무선통신서비스 가입자 통계 정보 중 LTE 가입 정보를 활용하여 KT 시장점유율 산출 가능

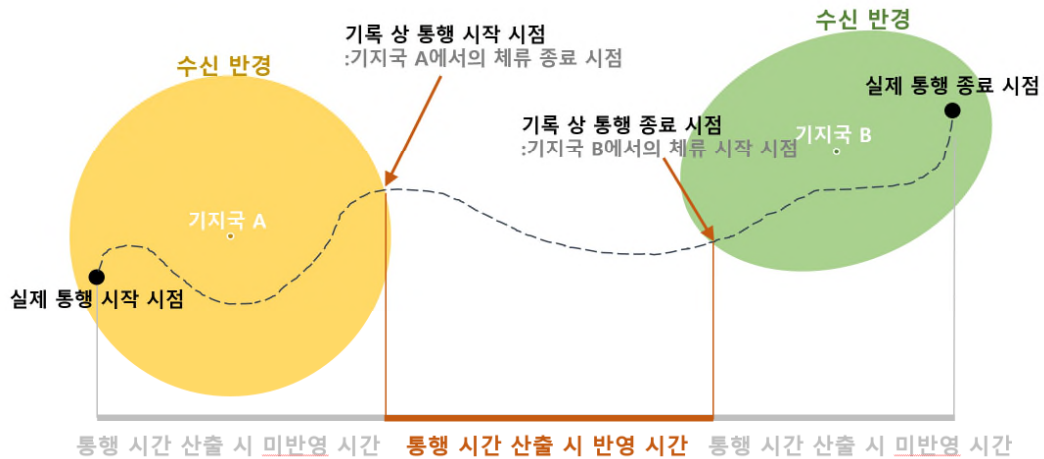
&lt;표 3- 4&gt; KT 시장점유율

구분		가입자 수	비율
LTE	SKT	24, 205, 134	43.9%
	KT	15, 259, 806	27.7%
	LGU+	12, 647, 000	22.9%
	MVNO	3, 021, 741	5.5%
	소계	55, 133, 681	100.0%

자료 : 과학기술정보통신부 “무선 통신서비스 통계 현황 (2018년 12월)”, <https://www.msit.go.kr/> (20.03.25).

## 10) 평균 통행시간 산출

- 모바일 기지국 데이터는 기지국 기반 로그 기록이므로 통행시간 산출 시, 기종점 부분에서 통행시간이 하향 계산되는 경우가 발생함



<그림 3- 7> 기지국 기반 데이터의 통행시간 산출 방법

- 위와 같은 문제를 해결하고자, 기지국의 평균 수신범위를 계산하여 시군별 기지국 내 이동 시간 산출 후, 통행시간을 보정함
  - 시군구별 면적과 기지국 수를 사용하여 시군구별 기지국 수의 밀도 계산
  - 시군구별 기지국의 평균 면적을 계산
  - 기지국의 수신범위가 원이라고 가정하고 원주율(3.14)을 통해 기지국 수신범위의 평균 반경 계산
  - 시군구 내 통행의 평균 통행속도를 구한 뒤, 해당 속도로 시군구별 기지국 수신범위의 평균 반경을 통행하는 시간 산출
- 각 개인의 통행 정보에서 통행시간(도착시간-출발시간, 분 단위)을 산출한 다음, 기종점 정보(기지국 위·경도 좌표)와 시군구(행정구역)정보를 매칭해 기종점 지점에서 시군구별 기지국 내 이동 평균 시간을 합산하여 최종 통행시간 계산함
- 개별통행에 대해 기점 정보(출발 일자, 출발 시간대, 출발 위치, 출발 체류지 유형), 종점 정보(도착 일자, 도착 시간대, 도착 위치, 도착 체류지 유형), 통행자 정보(성, 연령)가 동일한 로그 기록의 통행시간을 집계한 후 집계한 통행시간을 산술평균하여 산출함

## 제2절 구축 DB 검증

### 1. DB 검증 방법

- 기지국 기반의 데이터를 이용하여 구축된 모바일통신 빅데이터 통행DB의 신뢰성 및 정확성을 검증하기 위하여 기존 가구통행실태조사자료(2016년 기준, 한국교통연구원)와 인구총조사자료(2015년 기준, 통계청)를 활용
- 기지국 기반의 데이터를 이용하여 모바일통신 빅데이터 통행DB와 기존 가구통행실태조사자료와 인구총조사자료의 통행량, 통행시간 등을 비교·분석하고, 단순 비교할 수 없는 모바일통신 빅데이터 기반 상세지표(예를 들어, 상세 시간대별, 요일별, 성별, 연령별, 희망통근시간 등)는 통상적인 기대 통행패턴을 벗어나는지 확인
  - 지역별 통행 발생량·도착량을 가구통행실태조사자료와 상대적으로 비교하여, 유사한 패턴을 보이는지 확인
  - 시간대별, 요일별 통행 발생량·도착량을 확인하여, 통상적인 첨두시간 및 평일·주말 통행패턴이 나타나는지 확인
  - 성별, 연령별 통행 발생량·도착량을 확인하여, 통상적인 인구 비율 및 연령대별 통행 패턴이 나타나는지 확인
  - 지역별 출퇴근 목적 통행량·통행시간을 가구통행실태조사자료, 인구총조사자료와 상대적으로 비교하여, 유사한 패턴을 보이는지 확인
  - 출퇴근 목적 통행시간·통행속도와 희망통근시간(60분)기준 통행 패턴을 확인하여, 통상적인 통행패턴이 나타나는지 확인
- 단, 데이터 가공 시간 때문에 1년 치 모든 데이터를 가공하여 과업 기간 내에 검증할 수 없으므로, 통계적으로 충분히 유의한 10만 명의 표본 자료를 이용하여 검증을 시행함
  - 표본 유의성을 분석한 결과, 약 2만 명의 표본 이후 통행시간의 평균 오차는  $\pm 0.1$  분 이하임.

## 2. DB 검증 결과

### 가. 통행 발생·도착패턴 비교 및 분석 결과

#### 1) 가구통행실태조사자료 기준 지역별 통행량 비교 결과

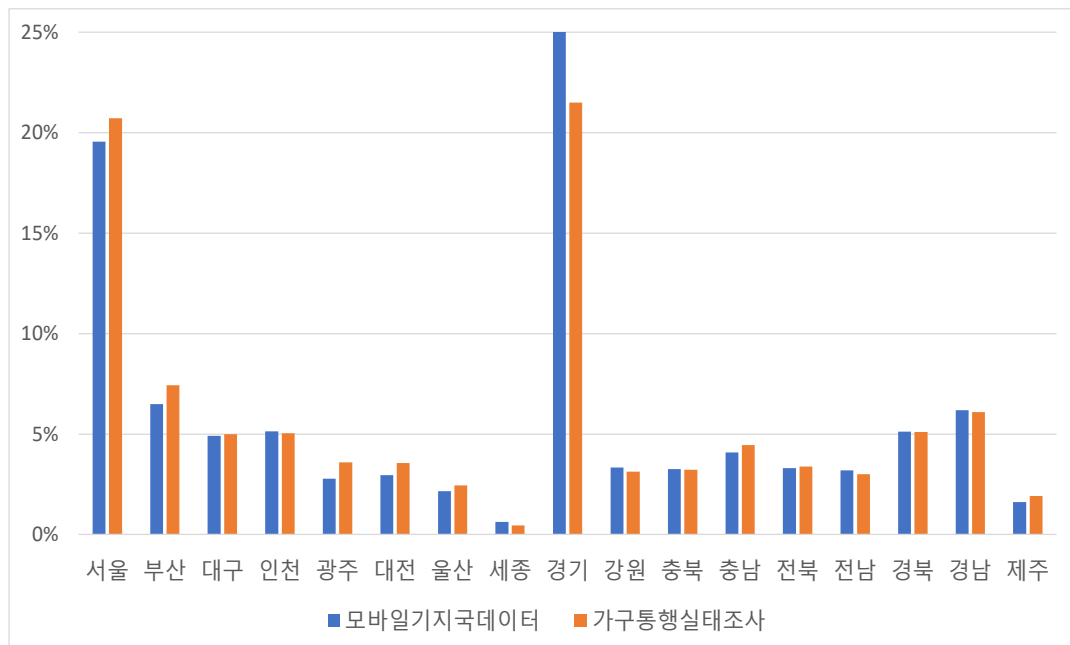
- 지역별 통행 발생량·도착량을 가구통행실태조사자료와 상대적으로 비교하여, 유사한 패턴을 보이는지 확인

<표 3- 5> 지역별 통행량 비교 결과

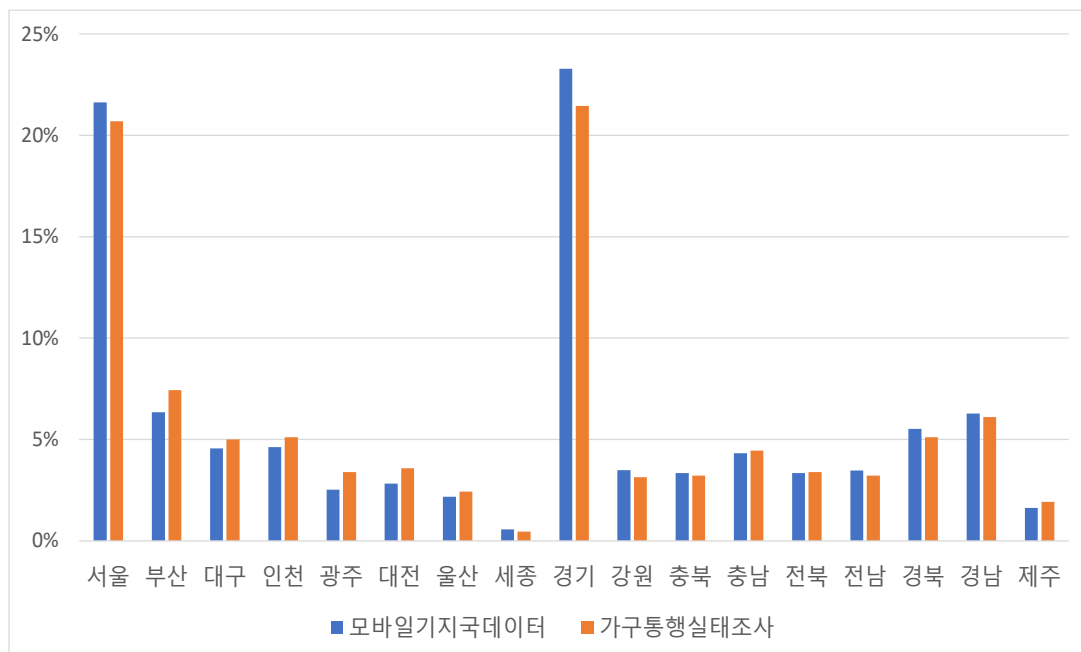
(단위: 통행/일(비율%))

구분	가구통행실태조사자료 기준*		모바일통신 빅데이터 기준	
	발생 (비율)	도착 (비율)	발생 (비율)	도착 (비율)
서울	17,996,813 (20.7)	17,971,878 (20.7)	1,545,273 (19.5)	1,709,372 (21.6)
부산	6,457,538 (7.4)	6,458,293 (7.4)	513,449 (6.5)	500,869 (6.3)
대구	4,333,058 (5.0)	4,330,603 (5.0)	388,500 (4.9)	360,704 (4.6)
인천	4,370,318 (5.0)	4,433,014 (5.1)	405,912 (5.1)	364,511 (4.6)
광주	3,117,712 (3.6)	2,942,575 (3.4)	219,652 (2.8)	199,102 (2.5)
대전	3,091,862 (3.6)	3,101,487 (3.6)	233,980 (3.0)	223,273 (2.8)
울산	2,117,733 (2.4)	2,109,519 (2.4)	171,036 (2.2)	171,523 (2.2)
세종	389,864 (0.4)	391,571 (0.5)	49,631 (0.6)	44,047 (0.6)
경기	18,668,530 (21.5)	18,628,440 (21.4)	1,993,158 (25.2)	1,840,414 (23.3)
강원	2,714,015 (3.1)	2,715,277 (3.1)	264,098 (3.3)	275,009 (3.5)
충북	2,803,473 (3.2)	2,793,476 (3.2)	257,411 (3.3)	264,394 (3.3)
충남	3,871,861 (4.5)	3,863,535 (4.4)	322,793 (4.1)	341,216 (4.3)
전북	2,942,438 (3.4)	2,943,452 (3.4)	261,467 (3.3)	264,746 (3.3)
전남	2,606,282 (3.0)	2,783,178 (3.2)	253,020 (3.2)	274,287 (3.5)
경북	4,431,201 (5.1)	4,442,218 (5.1)	404,548 (5.1)	436,929 (5.5)
경남	5,297,159 (6.1)	5,300,768 (6.1)	489,631 (6.2)	495,935 (6.3)
제주	1,659,446 (1.9)	1,660,019 (1.9)	127,386 (1.6)	128,273 (1.6)
계	86,869,303 (100)	86,869,303 (100)	7,905,399 (100)	7,905,399 (100)

주 : \* 가구통행실태조사자료는 2016년 기준 승용차(택시), 버스, 철도(일반/고속/지하철), 항공, 해운 수단에 대한 목적통행량임.



<그림 3- 8> 지역별 통행 발생량 비교 결과



<그림 3- 9> 지역별 통행 도착량 비교 결과

- 가구통행실태조사자료 기준 1인당 일평균 통행량(발생+도착)은 약 3.4회이며, 모바일 통신 빅데이터 기준 1인당 일평균 통행량(발생+도착)은 약 3.1회 임
  - 가구통행실태조사자료 기준 전국 통행 발생·도착량은 1.74통행/일, 1.74통행/일이며, 모바일통신 빅데이터 기준 전국 통행 발생·도착량은 1.5통행/일, 1.6통행/일 임
  - 모바일통신 빅데이터 기준 통행량과 가구통행실태조사자료 기준 통행량 매우 유사하게 나타남
  - 하지만 가구통행실태조사자료가 더 높은 통행량을 띄는데, 이러한 차이는 기존의 가구통행실태조사는 출근, 등교, 업무, 쇼핑, 귀가 등 모든 통행목적에 따라 통행량을 집계하지만, 모바일통신 빅데이터는 체류시간 기준으로 체류지를 구분하기 때문에 기타잠재 체류지에 속하지 않은 이동이 집계되지 않기 때문에 통행량 차이가 과소 추정되었다고 판단됨
- 지역별 통행량 비율을 살펴보면, 평균적으로는 약 0.41%의 차이로 기존 가구통행실태 조사자료와 매우 유사한 패턴을 나타냄
  - 경기지역 통행 발생량이 약 +2.2%p로 가장 큰 폭으로 과대 추정되었으며, 부산지역 통행 발생량이 약 -1.1%p로 가장 큰 폭으로 과소 추정되었으나, 평균적으로 약  $\pm 1\%$ 로 매우 신뢰성 높은 결과를 도출함

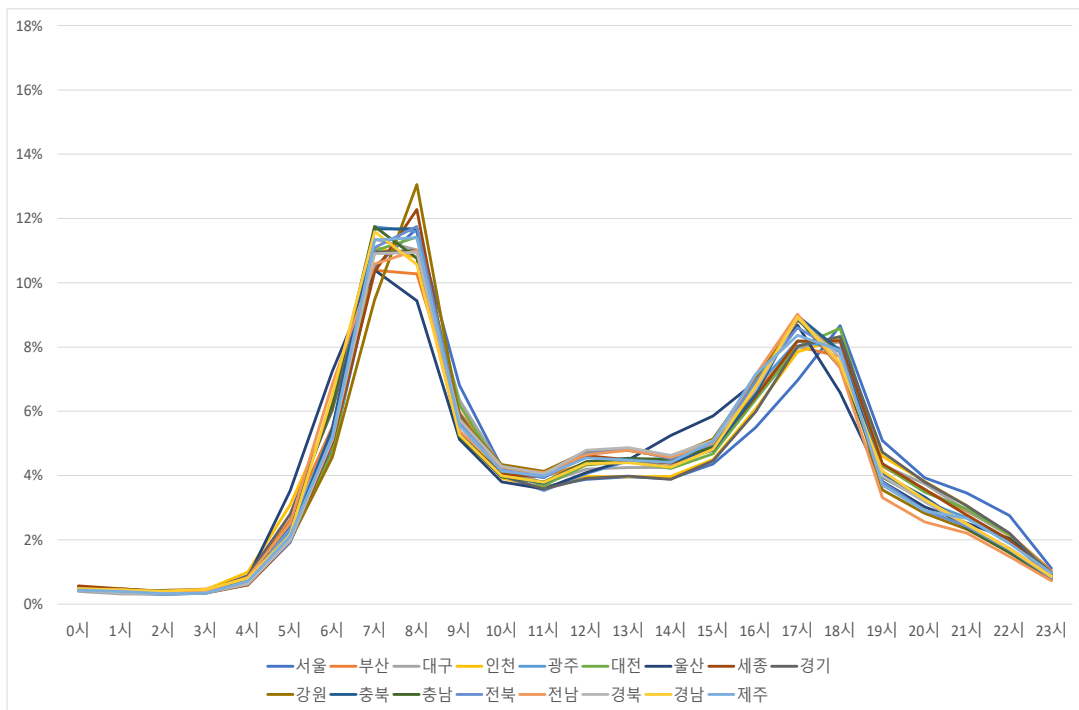
## 4) 시간대별 통행량 패턴 분석 결과

- 시간대별 통행 발생량·도착량을 확인하여, 통상적인 첨두시간 통행패턴이 나타나는지 확인함

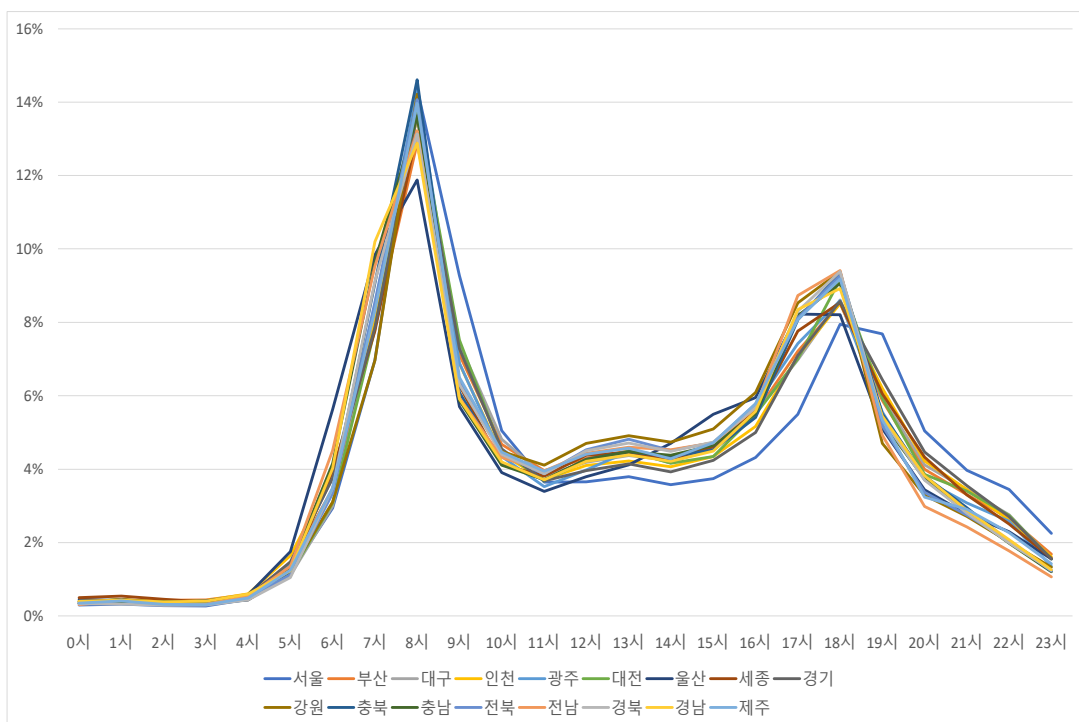
&lt;표 3- 6&gt; 시간대별 통행량 패턴 분석 (모바일통신 빅데이터 기준)

(단위: 통행/일)

구분		새벽		오전		오후		저녁	
		0시~	3시~	6시~	9시~	12시~	15시~	18시~	21시~
발 생	서울	16,651	56,604	416,252	227,475	181,636	259,954	273,283	113,237
	부산	6,873	20,220	134,047	72,833	66,993	100,115	79,618	31,262
	대구	3,989	11,648	106,341	55,795	49,330	75,169	63,061	23,121
	인천	4,936	18,403	113,148	53,433	48,231	74,638	67,641	24,843
	광주	2,453	7,396	63,107	29,279	28,057	43,590	33,557	12,129
	대전	2,901	7,333	63,417	32,770	30,378	44,586	38,335	14,235
	울산	2,324	8,168	46,246	21,374	23,608	36,651	22,931	9,433
	세종	711	1,423	13,713	6,596	6,697	9,666	8,014	2,804
	경기	23,566	80,871	558,073	268,712	234,608	367,950	336,020	122,531
	강원	3,473	8,280	71,445	37,686	37,182	55,348	37,541	12,798
	충북	2,999	7,900	74,177	33,596	34,458	52,233	39,331	12,694
	충남	4,120	12,095	92,463	41,729	43,340	65,890	46,807	15,206
	전북	3,108	7,747	73,307	35,979	36,670	53,603	38,041	12,849
	전남	3,086	9,481	71,687	34,599	35,293	53,365	33,326	11,146
	경북	4,431	11,992	109,014	56,996	57,754	83,488	60,058	20,561
	경남	6,437	16,596	139,315	63,395	63,338	99,748	72,408	24,668
	제주	1,475	4,007	35,495	17,388	17,076	26,137	18,431	7,022
도 착	서울	15,411	32,793	415,017	307,244	188,487	231,768	353,323	165,265
	부산	6,359	12,082	121,109	78,773	65,004	86,369	91,920	38,252
	대구	3,674	6,774	88,933	57,936	46,153	61,935	67,348	27,946
	인천	4,457	9,680	92,922	54,636	45,134	60,410	68,977	27,945
	광주	2,281	4,544	52,765	29,398	25,158	35,429	35,551	13,967
	대전	2,710	4,435	55,174	35,073	28,790	37,589	42,319	17,182
	울산	2,150	4,691	46,882	22,287	21,633	33,724	28,705	11,349
	세종	659	885	10,975	6,263	5,775	7,909	8,332	3,249
	경기	21,627	43,436	464,894	285,600	221,409	300,710	358,618	143,884
	강원	3,251	5,700	66,710	41,122	39,451	54,169	47,757	16,649
	충북	2,799	5,007	71,704	36,925	34,603	47,616	49,364	16,371
	충남	3,799	7,498	93,503	46,698	44,771	62,568	61,700	20,490
	전북	2,908	5,107	68,734	38,819	36,685	49,294	47,167	15,963
	전남	2,857	6,069	74,643	38,928	37,113	52,385	47,419	14,413
	경북	4,125	7,799	112,439	63,769	59,869	81,737	80,668	26,399
	경남	6,014	11,034	134,138	68,410	63,404	91,130	90,020	30,746
	제주	1,382	2,613	32,946	19,047	16,944	23,736	22,710	8,436



<그림 3- 10> 시간대별 통행 발생량 패턴 분석 (모바일기지국데이터 기준)



<그림 3- 11> 시간대별 통행 도착량 패턴 분석 (모바일통신 빅데이터 기준)



- 모바일통신 빅데이터 기준 시간대별 통행패턴 분석 결과, 모든 지역에서 전형적인 첨두시간 통행패턴을 보임
  - 오전 5시를 시작으로 통행량이 점점 증가하여 오전 9시에 감소하는 전형적인 오전 첨두시간 통행패턴이 나타나며, 오후 3시부터 점차 증가하여 오후 6시를 정점으로 오후 10시까지 이어지는 전형적인 오후 첨두시간 통행패턴을 나타냄
  - 퇴근 시간보다 출근 시간에 더 집중되는 통행패턴과 첨두시간 대비 약 1/3에 해당하는 비첨두 통행패턴 등 매우 신뢰성 있는 통행패턴 결과를 도출함
  - 지역별로도 매우 유사한 패턴을 보여, 지역별 편차 없이 신뢰성 있는 결과가 도출됨

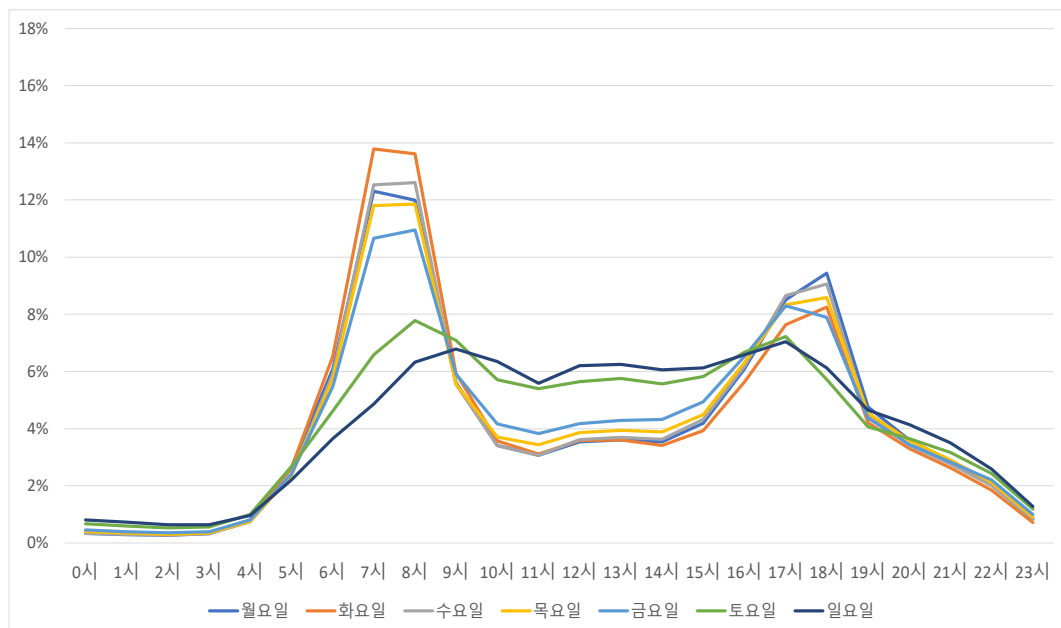
## 5) 요일별 시간대에 따른 통행량 패턴 분석 결과

- 요일별 통행 발생량·도착량을 확인하여, 통상적인 평일·주말 통행패턴이 나타나는지 확인함

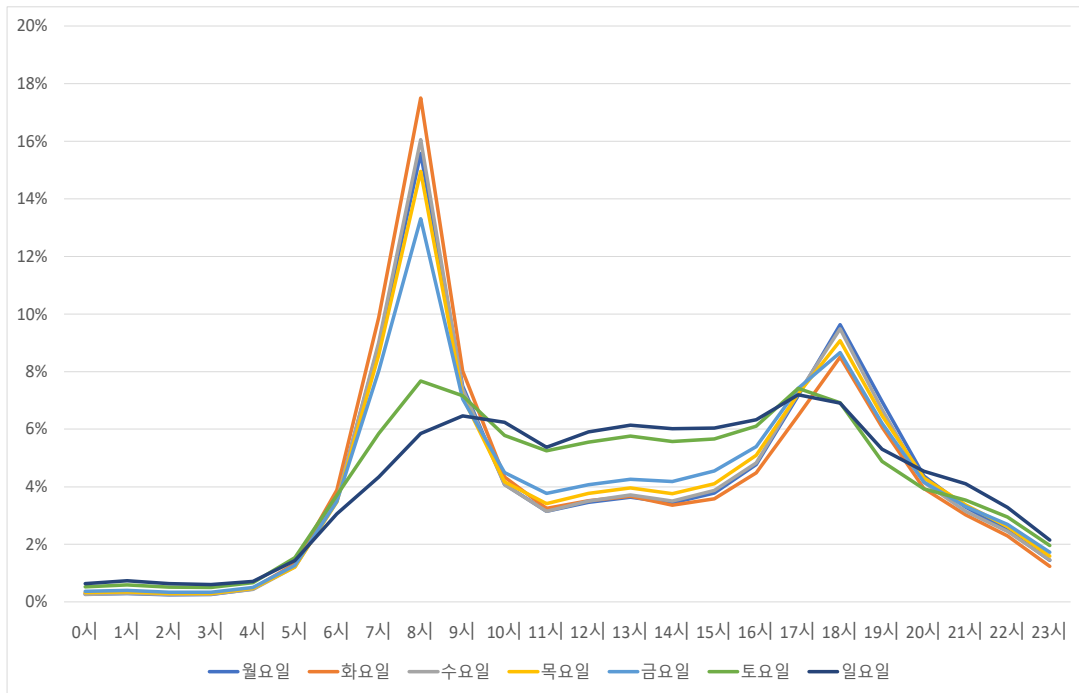
&lt;표 3- 7&gt; 요일별 통행량 패턴 분석 (모바일통신 빅데이터 기준)

(단위: 통행/일)

구분		새벽		오전		오후		저녁	
		0시~	3시~	6시~	9시~	12시~	15시~	18시~	21시~
발생	월	10,635	45,528	368,977	146,916	129,658	228,041	216,069	69,325
	화	10,537	40,580	366,144	136,010	114,353	185,400	170,077	55,939
	수	11,499	43,900	398,381	154,360	140,219	245,910	217,448	72,066
	목	14,822	49,022	417,938	180,504	166,100	272,044	237,491	84,037
	금	15,857	47,617	357,916	183,649	169,040	261,496	207,751	79,546
	토	16,911	40,336	181,173	173,684	161,968	188,355	128,314	64,612
	일	13,273	23,177	90,717	114,514	113,311	120,885	91,251	45,013
도착	월	9,768	24,078	340,926	178,810	127,945	190,910	254,805	87,905
	화	9,748	22,534	337,215	168,307	113,927	156,813	199,731	70,767
	수	10,621	24,716	367,080	187,601	137,683	204,546	260,798	90,739
	목	13,728	28,077	384,144	209,116	163,425	233,565	282,154	107,748
	금	14,654	27,995	328,452	202,646	165,561	229,563	251,654	102,345
	토	15,587	25,911	164,541	173,846	161,267	183,352	150,203	80,646
	일	12,357	16,839	81,130	110,604	110,576	119,727	102,552	58,356



&lt;그림 3- 12&gt; 요일별 시간대에 따른 통행 발생량 변화 (모바일통신 빅데이터 기준)



<그림 3- 14> 요일별 시간대에 따른 통행 도착량 변화 (모바일통신 빅데이터 기준)

- 모바일통신 빅데이터 기준 요일별 통행패턴 분석 결과, 평일과 주말 기준 전형적인 통행패턴을 보임
  - 평일인 월~금요일은 매우 뚜렷한 평일 첨두시간 통행패턴을 보이며, 주말인 토요일~일요일은 낮 시간대 비교적 완만한 전형적인 주말 통행패턴을 보임으로써, 매우 신뢰성 있는 결과가 도출되었다고 판단됨
  - 평일 요일별 통행패턴은 큰 차이가 없으며, 일요일은 토요일에 비해 완만한 낮 시간대 통행패턴을 보임

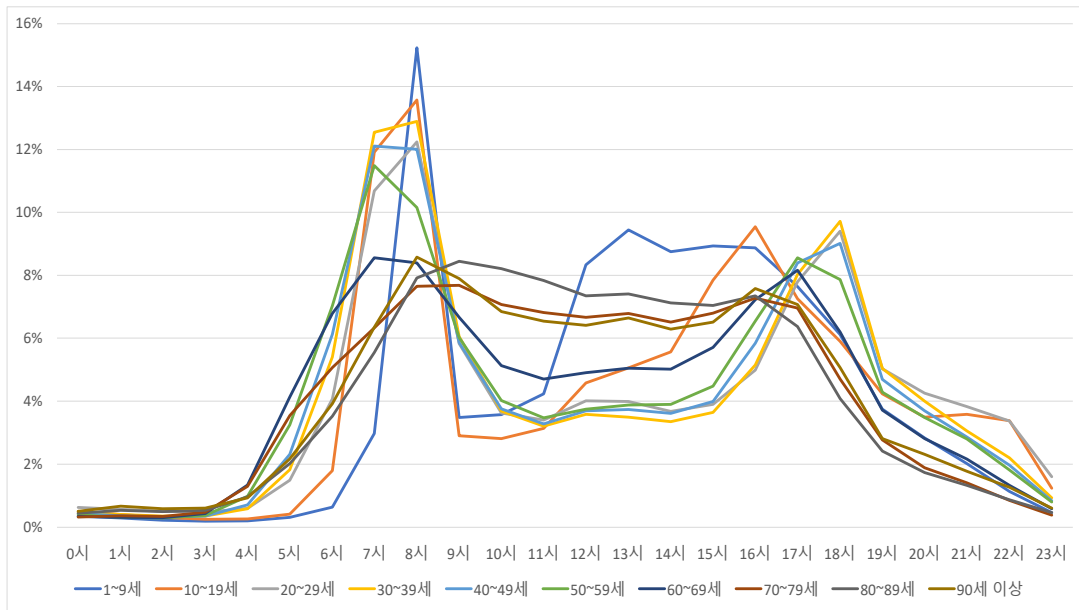
## 6) 연령별 통행량 패턴 분석 결과

- 연령별 통행 발생량·도착량을 확인하여, 통상적인 인구비율 및 연령대별 통행패턴이 나타나는지 확인

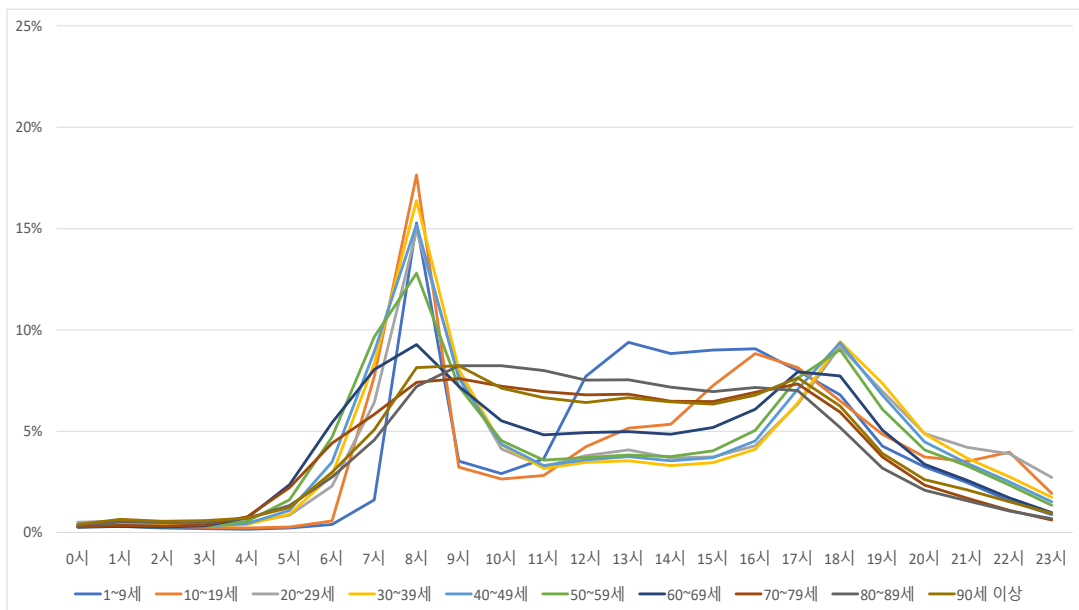
&lt;표 3- 8&gt; 연령별 통행량 패턴 분석 (모바일통신 빅데이터 기준)

(단위: 통행/일)

구분		새벽		오전		오후		저녁	
		0시~	3시~	6시~	9시~	12시~	15시~	18시~	21시~
발 생	0대	507	411	11,058	6,624	15,570	14,931	7,452	2,121
	10대	5,849	4,474	131,675	42,773	73,421	118,916	65,751	39,588
	20대	21,894	32,965	347,875	165,693	150,548	215,042	241,024	113,413
	30대	16,232	36,417	407,259	171,012	137,681	222,179	247,618	81,814
	40대	16,520	50,108	449,137	191,599	164,178	270,709	258,662	84,410
	50대	17,080	79,033	491,699	231,832	197,766	335,744	268,109	92,940
	60대	10,312	63,716	256,203	178,100	161,724	227,662	137,403	43,954
	70대	3,879	20,046	71,829	81,342	75,288	79,300	35,307	9,996
	80대	1,185	2,829	13,685	19,728	17,626	16,723	6,629	2,142
	90대	77	163	826	933	847	927	446	160
도 착	0대	477	341	10,150	5,904	15,223	15,288	8,399	2,892
	10대	5,605	3,631	125,041	41,912	71,139	116,942	72,798	45,377
	20대	20,643	23,180	304,189	195,876	148,954	184,962	271,025	139,625
	30대	15,016	21,562	364,076	205,353	136,223	184,346	285,864	107,772
	40대	15,102	27,269	410,264	226,833	161,617	227,337	306,358	110,544
	50대	15,477	42,320	465,411	262,777	193,268	286,287	328,966	119,696
	60대	9,361	37,001	245,427	189,530	159,437	207,239	174,196	56,884
	70대	3,599	12,698	66,534	82,077	75,763	78,155	45,305	12,857
	80대	1,112	2,037	11,687	19,702	17,905	17,014	8,430	2,658
	90대	71	112	709	964	854	909	558	201



<그림 3- 14> 연령별 통행 발생량 패턴 분석 (모바일통신 빅데이터 기준)



<그림 3- 15> 연령별 통행 도착량 패턴 분석 (모바일통신 빅데이터 기준)

- 모바일통신 빅데이터 기준 연령별 통행패턴 분석 결과, 학업 및 경제활동 연령(0~59세)은 뚜렷한 첨두시간 통행패턴을 보이며, 비경제활동 연령(60세 이상)은 시간대별 다소 완만한 통행패턴을 보임
  - 특히, 0~19세에 해당하는 학업활동 연령대는 다소 이른 오후 1시부터 통행량이 증가하는 패턴을 보여, 실제 상대적으로 이른 하교 시간을 정확하게 반영하는 것으로 판단됨
  - 주로 비경제활동 인구에 해당하는 60세 이상의 통행패턴은 경제활동 인구에 비해 비교적 완만한 첨두시간 통행패턴을 보여, 실제 연령별 통행패턴과 매우 유사한 결과를 도출한 것으로 판단됨

## 나. 목적별(출퇴근) 통행패턴 비교 및 분석 결과

### 1) 가구통행실태조사자료·인구총조사자료 기준 출·퇴근 통행량 비교 결과

- 지역별 출퇴근 목적 통행 발생량·도착량을 가구통행실태조사자료·인구총조사자료와 상대적으로 비교하여, 유사한 패턴을 보이는지 확인함

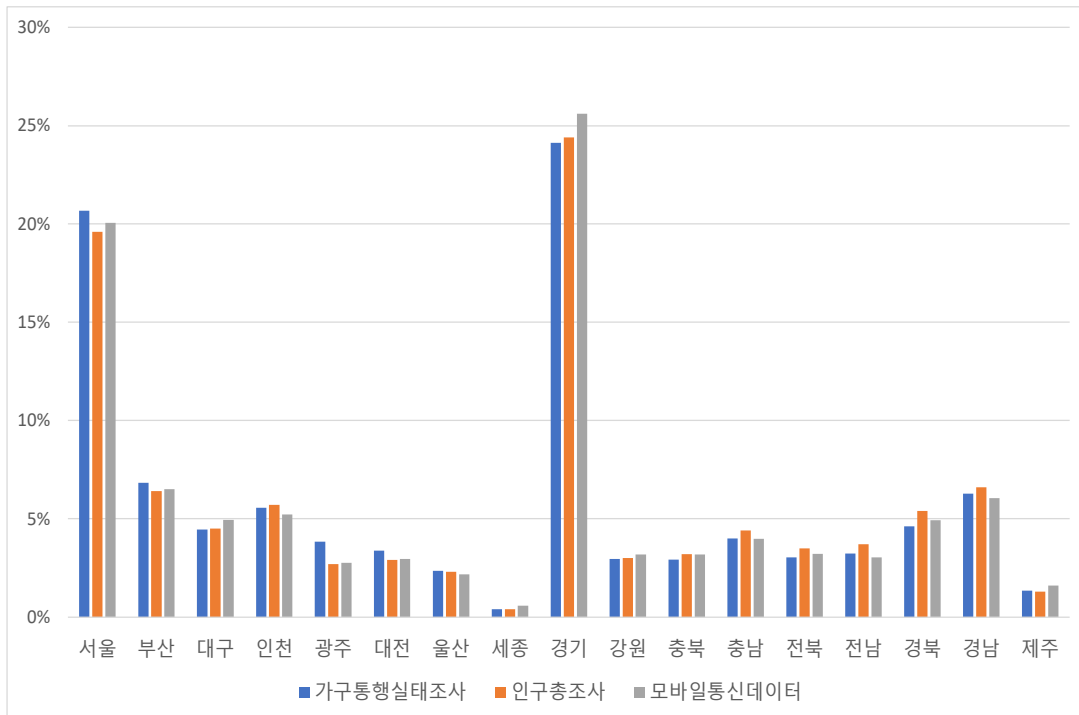
<표 3- 9> 지역별 출·퇴근 통행량 비교 결과

(단위: 통행/일(비율%))

구 분	가구통행실태조사자료 기준		인구총조사자료 기준	모바일통신 빅데이터 기준	
	출근 (비율)	귀가*	통근**	출근	퇴근
서울	4,516,466(20.7)	8,413,729(22.1)	4,597,322(19.6)	650,284(20.4)	547,269(19.6)
부산	1,492,300(6.8)	3,051,128(8.0)	1,500,101(6.4)	207,420(6.5)	180,879(6.5)
대구	974,423(4.5)	1,801,446(4.7)	1,064,227(4.5)	157,348(4.9)	137,939(4.9)
인천	1,213,843(5.6)	1,794,429(4.7)	1,344,311(5.7)	167,280(5.3)	145,001(5.2)
광주	838,601(3.8)	1,279,355(3.4)	639,412(2.7)	88,578(2.8)	77,006(2.8)
대전	736,843(3.4)	1,313,431(3.4)	672,856(2.9)	94,244(3.0)	82,383(3.0)
울산	514,493(2.4)	995,135(2.6)	538,654(2.3)	69,686(2.2)	59,901(2.1)
세종	89,203(0.4)	149,020(0.4)	92,403(0.4)	18,441(0.6)	16,517(0.6)
경기	5,272,517(24.1)	7,799,702(20.5)	5,745,634(24.4)	818,330(25.7)	711,400(25.5)
강원	644,578(2.9)	1,186,771(3.1)	703,266(3.0)	100,084(3.1)	90,309(3.2)
충북	637,895(2.9)	1,182,200(3.1)	742,399(3.2)	99,482(3.1)	90,888(3.3)
충남	875,177(4.0)	1,616,017(4.2)	1,032,913(4.4)	123,410(3.9)	113,776(4.1)
전북	665,597(3.0)	1,263,474(3.3)	811,899(3.5)	100,648(3.2)	91,070(3.3)
전남	706,911(3.2)	1,241,157(3.3)	864,498(3.7)	94,807(3)	86,754(3.1)
경북	1,007,670(4.6)	2,045,073(5.4)	1,280,135(5.4)	152,433(4.8)	141,449(5.1)
경남	1,372,291(6.3)	2,361,671(6.2)	1,560,582(6.6)	190,028(6)	171,842(6.2)
제주	291,636(1.3)	581,151(1.5)	309,133(1.3)	50,611(1.6)	44,769(1.6)
계	21,850,443(100)	38,074,889(100)	23,499,745(100)	3,183,114(100)	2,789,152(100)

주 : \*가구통행실태조사자료는 2016년 기준 출·퇴근에 대한 목적통행량이며, 귀가통행 통행목적에 상관 없이 집으로 향하는 모든 통행을 의미함

\*\*인구총조사(통계청)자료는 2015년도 기준이며, 출근과 퇴근이 구분되어 있지 않음



<그림 3- 16> 지역별 출·퇴근 통행량 분포 비교 결과

- 지역별 출퇴근 통행량 비율을 살펴보면, 평균적으로 가구통행실태조사자료 기준 약 0.64%p, 인구총조사자료 기준 약 0.3%p의 차이를 보임
  - 가구통행실태조사자료와 비교하였을 때, 부산지역의 출·퇴근 통행량이 약 2.4%p로 가장 큰 폭으로 과대 추정되었으며, 경기지역의 출퇴근 통행량이 약 3.3%로 가장 큰 폭으로 과소 추정되었으나, 평균적으로 약 0.64%p로 모바일통신 빅데이터 기준 출퇴근 통행량이 신뢰성 있는 것으로 판단됨
  - 인구총조사자료 기준 대구지역의 출퇴근 통행량이 약 0.4%p로 가장 큰 폭으로 과대 추정되었으며, 전남지역의 출퇴근 통행량이 약 0.6%p로 가장 큰 폭으로 과소 추정되었으나, 평균적으로 약 0.3%p로 모바일통신 빅데이터 기준 출퇴근 통행량이 신뢰성 있는 것으로 판단됨



## 2) 가구통행실태조사자료·인구총조사자료 기준 출퇴근 통행시간 비교 결과

- 지역별 출퇴근 목적 통행시간을 가구통행실태조사자료·인구총조사자료와 상대적으로 비교하여, 유사한 패턴을 보이는지 확인함

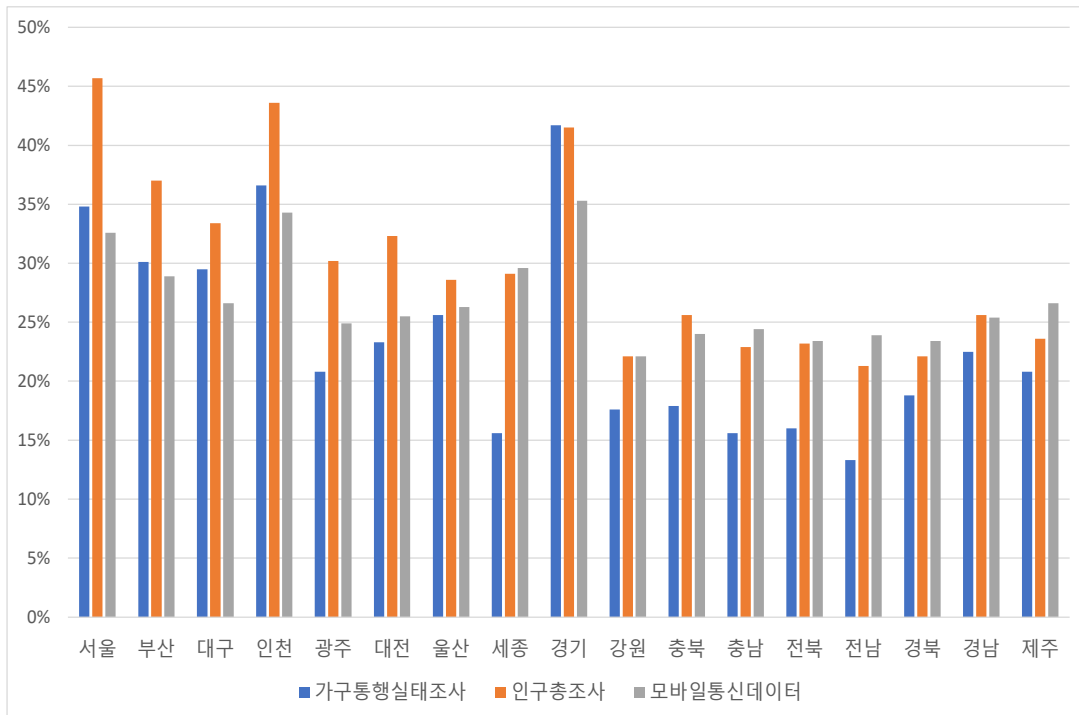
&lt;표 3- 10&gt; 지역별 출퇴근 통행시간 비교 결과

(단위: 분)

구분	가구통행실태조사자료 기준		인구총조사 자료 기준	모바일통신 빅데이터 기준	
	출근	귀가*	통근**	출근	퇴근
서울	34.8	30.1	45.7	32.6	35.1
부산	30.1	25.6	37.0	28.9	30.7
대구	29.5	28.2	33.4	26.6	28.2
인천	36.6	32.5	43.6	34.3	35.8
광주	20.8	18.3	30.2	24.9	26.7
대전	23.3	21.7	32.3	25.5	27.2
울산	25.6	24.5	28.6	26.3	27.9
세종	15.6	15	29.1	29.6	29.4
경기	41.7	36.3	41.5	35.3	36.9
강원	17.6	20.3	22.1	22.1	23.4
충북	17.9	17.8	25.6	24	25.6
충남	15.6	16.7	22.9	24.4	25.7
전북	16	15.4	23.2	23.4	25
전남	13.3	13.6	21.3	23.9	25.2
경북	18.8	20	22.1	23.4	24.5
경남	22.5	21.1	25.6	25.4	26.9
제주	20.8	21.5	23.6	26.6	28.9
계	23.6	22.3	35.2	29.8	31.5

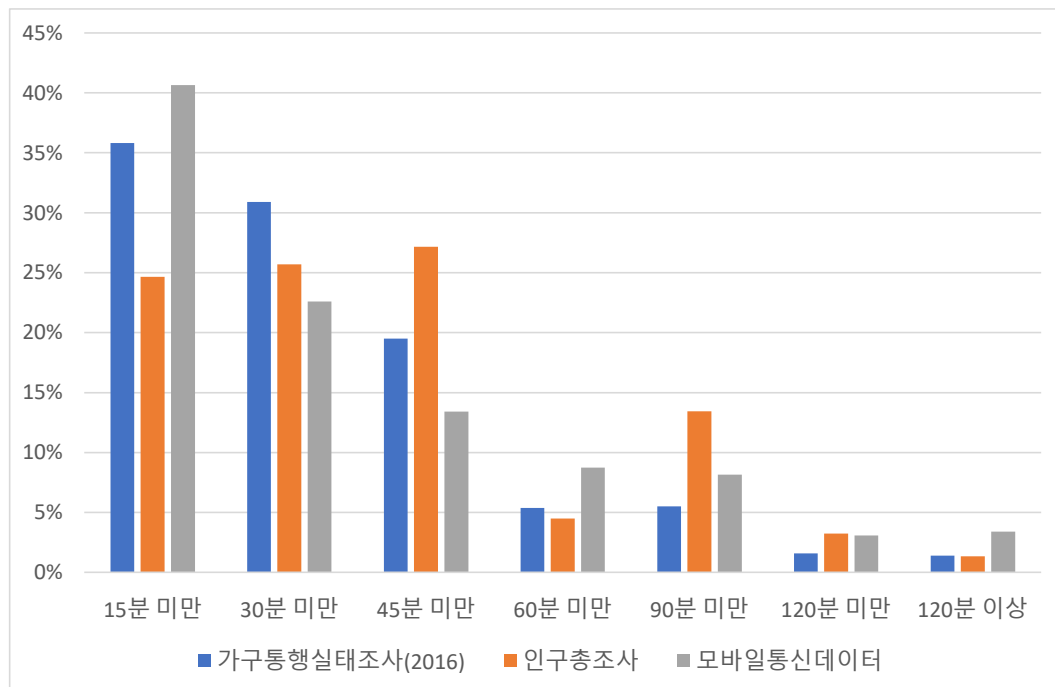
주 : \*가구통행실태조사자료는 2016년 기준 출근·귀가 통행에 대한 일평균시간이며, 귀가통행 통행목적에 상관없이 집으로 향하는 모든 통행을 의미함

\*\*인구총조사자료는 2015년 기준 통근 통행에 대한 소요 시간별 자료를 일평균시간 형태로 재구성하였으며, 출근과 퇴근이 구분되어 있지 않음



<그림 3- 17> 시도별 통근 통행시간 비교(단위 : 분)

- 가구통행실태조사자료와 인구총조사자료 기준 1인당 평균 출근시간 및 통근시간은 각각 약 23.6분, 35.2분이며, 모바일통신 빅데이터 기준은 약 29.8분임
- 각 자료별 평균 출퇴근 통행시간을 비교해보면, 인구총조사자료 기준이 가장 길고, 가구통행실태조사자료 기준이 가장 짧은 것으로 나타남
- 지역별 분석결과를 살펴보면, 모바일통신 빅데이터는 특별시·광역시 지역에서 가구통행실태조사자료와 유사한 출퇴근 통행시간 분포를 띄며, 시도에서는 인구총조사자료와 유사한 출퇴근 통행시간 분포를 띄
- 각각 지역마다 출퇴근 통행시간을 비교해보면 다소 차이가 있다고 해석할 수 있지만, 전체적인 지역적 패턴을 기준으로 보았을 때 매우 유사한 분포를 띄

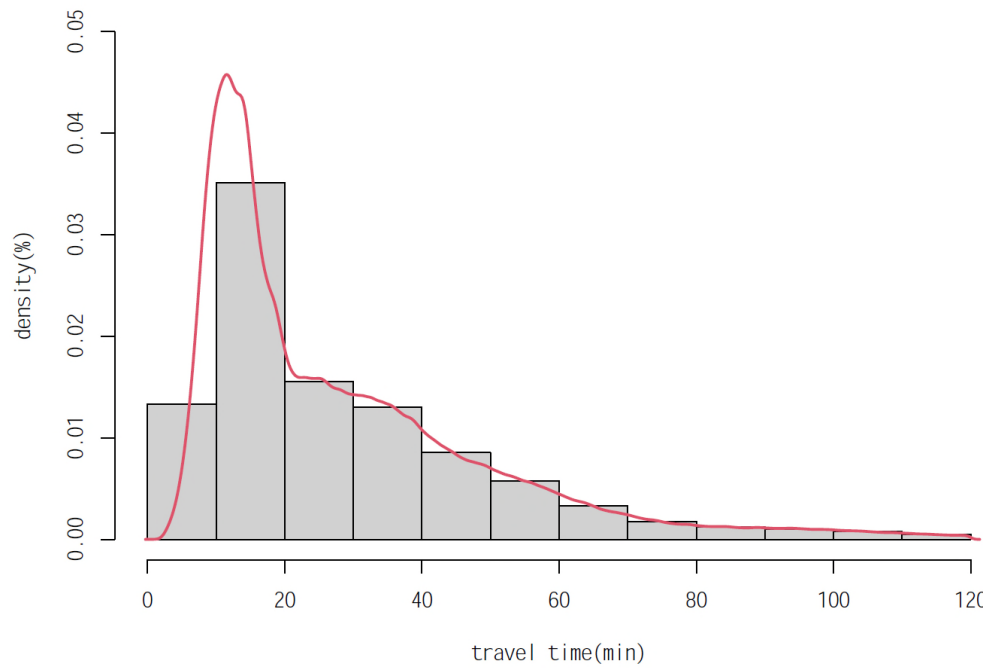


<그림 3- 18> 통행시간 기준별 통근 인구 비율 비교

- 각 자료의 통행시간 분포별 통행량을 살펴보면, 특정 시간 분포에서 다소 과대·과소 추정되는 경향이 관찰되나, 전체적으로는 기존자료 대비 매우 유사한 패턴을 보임
  - 15분 미만 출퇴근 통행량은 인구총조사자료 대비 다소 과대 추정되었으나, 가구통행조사자료 기준과는 매우 유사한 분포를 띄는 것으로 나타남
  - 30분 이상 45분 미만에서는 다소 과대추정됨
  - 45분 이상 60분 미만에서 출퇴근 통행량은 가구통행실태조사 기준과 매우 유사한 분포를 띄며, 인구총조사자료와 비교하였을때는 과소 추정된 것으로 나타남
  - 통행 시간대별로 모바일통신 빅데이터가 과소 또는 과대 추정되어 나타났지만, 출·퇴근 시간, 출퇴근 속도 밀도분포를 확인해본 결과, 매우 신뢰성 있는 분포를 보임<sup>3)</sup>
    - <그림 3-20>, <그림 3-21>은 출퇴근 시간 밀도 그래프임. 모바일통신 빅데이터 기준 출퇴근 통행시간은 10분을 정점으로 로그 노말(log-normal) 분포에 가깝게 나타남
    - <그림 3-22>와 <그림 3-23>은 출퇴근 속도 밀도 그래프임. 출퇴근 통행속도는 12km/h를 정점으로 로그 노말(log-normal) 분포에 가깝게 나타남

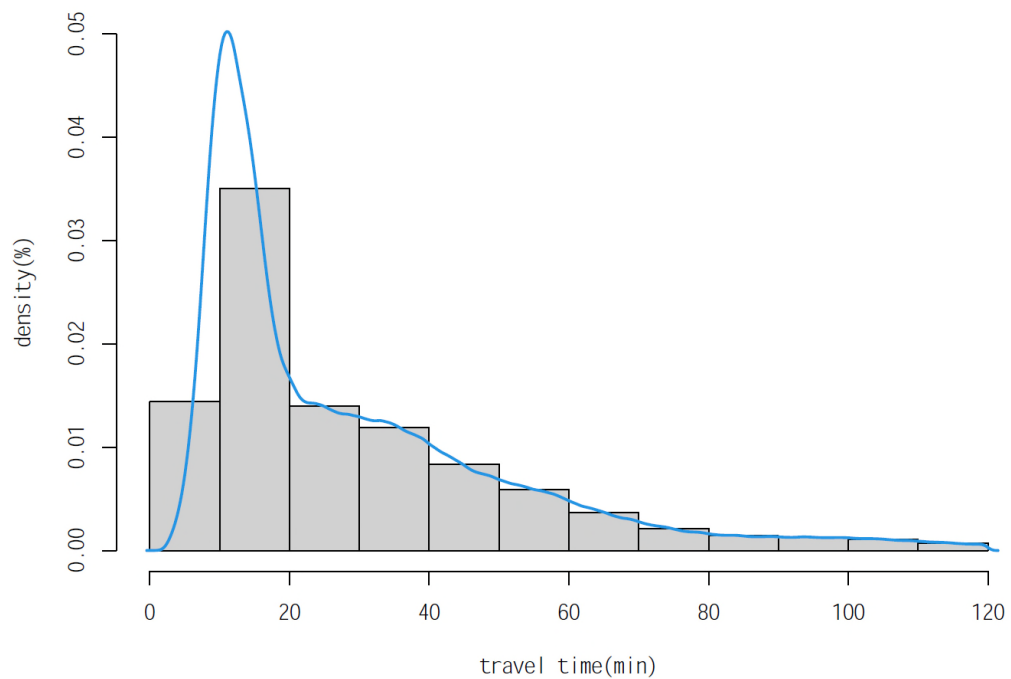
<sup>3)</sup> 노정현, ‘교통계획: 통행수요이론과 모형’, 개정판(2012), 나남.

전국(출근)

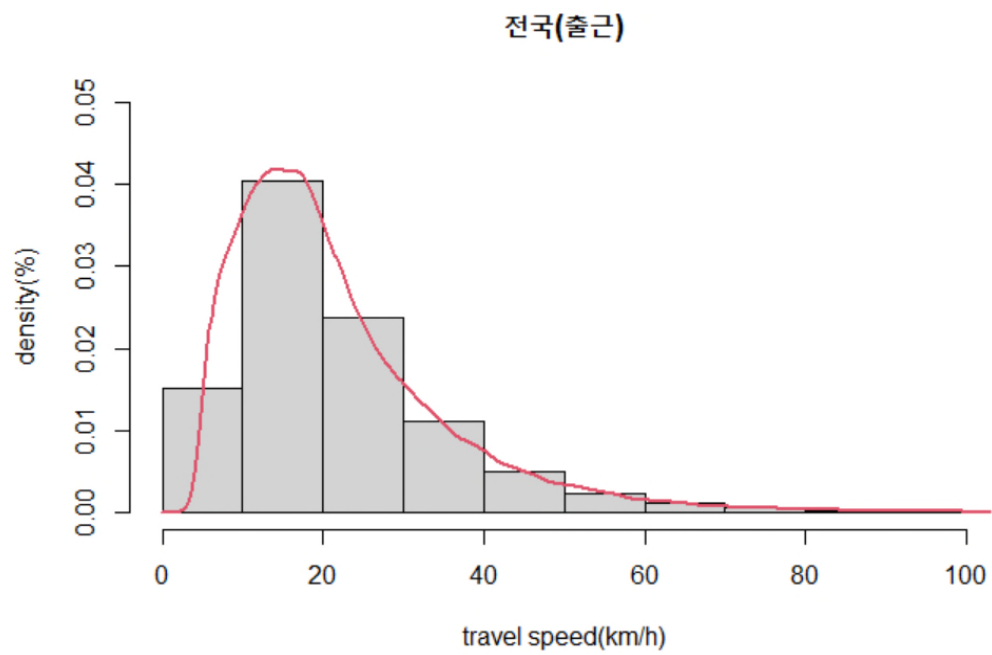


&lt;그림 3- 19&gt; 출근 시간 밀도 그래프

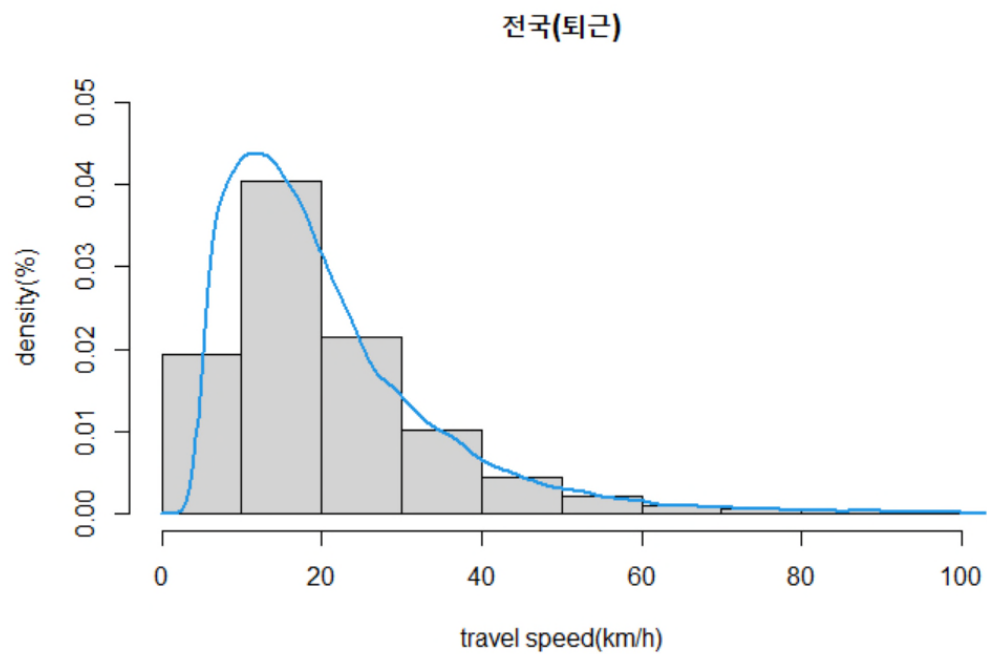
전국(퇴근)



&lt;그림 3- 20&gt; 퇴근 시간 밀도 그래프



<그림 3- 21> 출근 속도 밀도 그래프



<그림 3- 22> 퇴근 속도 밀도 그래프

## 제4장 모바일통신 빅데이터 기반 교통지표 개발 및 서비스 제공

---

### 제1절 개요

### 제2절 기반 통행 DB 재구조화

### 제3절 교통지표 개발 및 분석 서비스 제공



## 제4장 모바일통신 빅데이터 기반 교통지표 개발 및 서비스 제공

### 제1절 개요

#### 1. 개발 방향 및 범위

- 2018년부터 교통 모니터링·데이터 제공·분석 플랫폼(View-T)를 통해 모바일 통신 빅데이터를 기반으로 하는 통행 분석 서비스를 제공해왔음
- 해당연도에는 2019년 기준 모바일통신 빅데이터 기반 통행 DB를 기반으로 통행 분석 서비스를 제공함
- 사용자가 시스템에 구축된 기반 데이터를 이종 데이터와 연계하여 분석하거나 다른 방식으로 결과를 표출하는데 활용할 수 있도록 데이터 다운로드 서비스 제공

<표 4-1> 과업의 내용적 범위

구분		주요내용
데이터 재구조화 및 교통지표 개발		모바일통신 빅데이터 가공 및 전처리 모바일통신 빅데이터 행정구역 및 분석 특성별 DB 구축 모바일통신 빅데이터 이종 데이터 간 결합 및 분석 목적별 교통지표 개발 모바일 기지국 수신반경에 기반한 사람 모빌리티 기반지도 구축
서비스	분석도구 고도화	주요 통행지역 / 통행 구간 분석 도구 고도화 내부 통행지역 및 통행시간/거리 분석 도구 고도화 관광특화지역/고령자 주요통행 구간 분석/코로나19 위험지역 분석
	이용자 편의성 개선 및	GIS 링크 표출 방식 개선 및 링크 다운로드 기능 개선
	유지보수	홈페이지 운영 및 DB 유지보수
	신규 개발	대용량 빅데이터 처리를 위한 최적화 시스템 구성 모바일 환경 View-T 서비스 개발

#### 2. 모바일통신 빅데이터 기반 통행 DB 분석 프로세스

##### 가. 분석 프로세스 주요 개념

- 개별 사람의 시공간적 움직임이 휴대폰 신호를 통해 인근 기지국에 송신되어 수집되

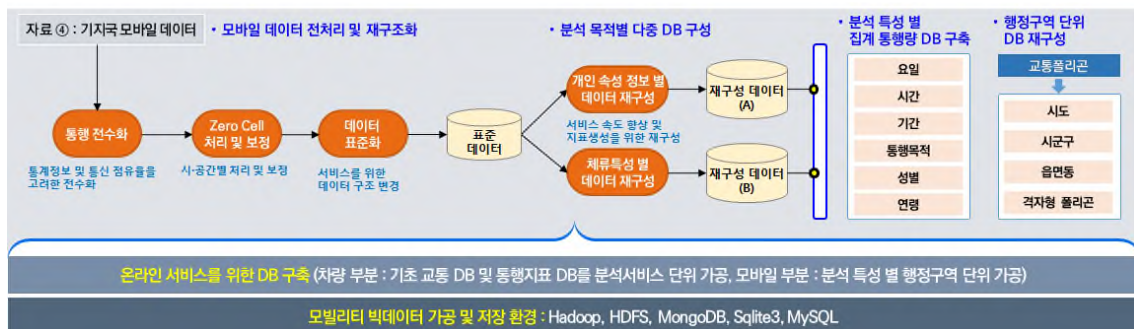


는 데이터로 가상기지국 단위로 정보가 수집되고 있으나, 데이터 제공 시에는 개인정보보호법과 위치정보의 보호 및 이용 등에 관한 법률 등으로 개인의 위치와 이동궤적이 추정되지 않는 집계 형태여야 함. 집계 형태는 개인의 위치와 이동궤적이 추정되지 않도록 집계구의 데이터와 기지국 간의 정보를 이용하여 구축한 교통폴리곤 단위로 변환 후 OD형태로 데이터가 제공되고 있음. 데이터 제공을 위한 데이터 처리 과정은 아래와 같음

- 원시데이터 전처리
  - 원시 데이터는 고객 식별번호, 가상기지국, 체류시작시간, 체류종료시간, 연령대, 성별 등의 정보로 구성
  - 고객 식별번호와 시간정보를 이용하여, 선을 그린 후 데이터 전환 및 이상치 제거
- 데이터 로그별 이동과 체류 정보 구분
  - 체류시간 및 위치정보를 이용하여 사람의 이동인지 체류인지 판단
  - 체류의 판단 후 연령대별로 주 몇 회 이상 체류와 체류 시간대에 대한 기준을 두어 통행목적(집/학교/직장/여가 등) 구분
- 위치정보 변환 및 데이터 집계
  - 가상기지국의 위치정보와 교통폴리곤의 영역과 공간조인 하여 정보 매칭
  - 매칭 정보를 이용하여 교통폴리곤 기준으로 데이터 가공
  - 기종점 정보(일자, 시간대, 위치, 체류지 유형), 통행자 정보(성별, 연령), 통행목적 정보가 동일한 인구를 기준으로 집계. 집계된 값은 최소 3이상되도록 보정
  - 집계된 통행량(3통행)을 사용 시 실제 통행보다 과대하게 집계되는 문제로 인하여 보정계수를 적용하여 1통행, 2통행에 대한 데이터를 생성

#### 나. 분석 프로세스 과정

- 분석 프로세스는 크게 ① 데이터 재구조화 → ② 서비스 기반 DB 생성 → ③ 지표 생성 → ④ 온라인 서비스를 위한 데이터 구축 순으로 진행



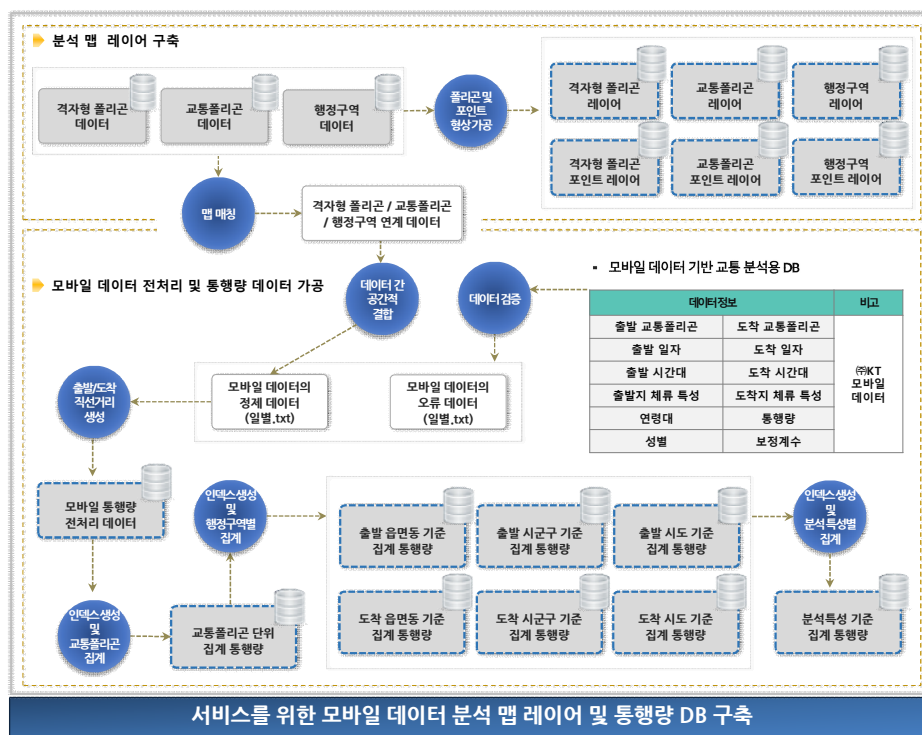
<그림 4- 1> 모바일통신 빅데이터 기반 분석 프로세스

## 제2절 기반 통행 DB 재구조화

### 1. 기반 통행 DB 재구조화

#### 가. 모바일통신 빅데이터 재구조화를 위한 분석 레이어

- 데이터 구축은 분석 맵 레이어 구축 및 모바일 교통빅데이터 전처리와 통행량 데이터 가공으로 구분함



<그림 4-2> 분석 레이어 구축 프로세스

- 분석 맵 레이어 구축은 격자형 폴리곤 데이터와 교통폴리곤 데이터 및 행정구역 데이터를 이용해 각각의 폴리곤 형상 레이어를 경량화하는 작업과 폴리곤 중심 위치 보정을 통해 포인트 레이어 구축 작업을 진행함
- 모바일통신 빅데이터 기반의 교통 분석용 데이터를 그대로 서비스 시 많은 데이터를 가져오는 데에 발생하는 처리 속도 이슈로 인하여 행정경계별로 그룹화하여 조회 데이터 건수를 감소시키고 속도를 개선함
- 모바일통신 빅데이터 전처리는 누락 또는 오류 데이터를 분류하고 정제 데이터의 교

통폴리곤과 행정구역을 공간적으로 결합시키고 출발 교통폴리곤과 도착 교통폴리곤의 중심 포인트 간 직선거리를 추가함

- 전처리한 모바일 통행량 데이터를 이용해 검색 속도 향상을 위한 인덱스를 생성하고 교통폴리곤 단위로 데이터를 그룹화하여 집계 통행량 DB를 구축함
- 행정구역 단위별로 서비스를 하기 때문에 구축한 교통폴리곤 단위 집계 통행량 DB를 이용하여 출발과 도착 읍면동 기준, 시군구 기준, 시도 기준으로 각각 집계 통행량 DB를 구축함
- 서비스 만족도를 향상시키기 위하여 검색 및 표출 속도를 개선하고자 모바일통신 빅데이터 속성정보를 가지고 분석 특성을 정의 후 행정구역 단위 집계 통행량 DB를 이용해 분석 특성 기준 집계 통행량 DB를 구축함

## 나. 모바일통신 빅데이터 재구조화

### 1) 모바일통신 빅데이터 재구조화 필요성

- 수급한 데이터의 분석과 누락된 데이터 또는 이상치데이터 여부에 대한 검증으로 데이터 파악
- View-T 서비스는 분석 지역과 표출 단위에 대한 선택이 필수적으로 GIS 분석 최소 단위인 교통폴리곤과 교통폴리곤이 속하는 행정구역의 매칭 정보가 필요함
- 검색 쿼리와 DB 적재 용량 부분을 고려하여 문자형 데이터를 숫자 형식으로 변환 작업이 필요함
- 과대하게 집계된 3 통행에 대해서 보정계수를 적용하여 1 통행과 2 통행 데이터를 생성
- 분석 서비스에 활용하기 위해 출발 교통폴리곤과 도착 교통폴리곤 간의 직선거리를 생성

## 2) 모바일통신 빅데이터 구조 예시

&lt;표 4- 2&gt; 모바일통신 빅데이터 구조

칼럼 명	비고	칼럼 명	비고
출발 교통폴리곤 ID	-	도착 교통폴리곤 ID	-
출발 일자	yyyymmdd	도착 일자	yyyymmdd
출발 시간대	0 - 23	도착 시간대	0 - 23
출발 체류지 유형	H, C, S, X	도착 체류지 유형	H, C, S, X
연령대	0 - 110	성별	M, F
통행량(인구수)	-	보정계수	-

## 3) 모바일통신 빅데이터 검증

- JAVA 기반의 검증시스템을 개발하여 설계 DB를 기준으로 수집된 통신 데이터를 전수 검증하고, 정보가 누락된 통신 데이터에 대하여 오류 데이터를 생성 후 확인

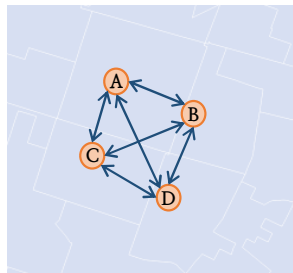
## 4) 모바일통신 빅데이터의 문자형 데이터 형식변환 및 거리 데이터 생성

- 전처리 작업 시간을 줄이기 위해 모바일통신 빅데이터 검증과 동시에 작업 진행함

■ 데이터 형식 변환 예시

기존	변환후	비고
H	11	집
C	12	회사
S	13	학교
R	14	종교집회장소
X	15	기타 잠재체류지
F	1	여자
M	2	남자

■ 거리 데이터 산출 예시



출발 폴리곤	도착 폴리곤	거리(m)
A	B	100
A	C	150
A	D	200
B	C	200
B	D	120
C	D	130

&lt;그림 4- 3&gt; 모바일통신 빅데이터의 형식변환 및 거리 산출

## 5) 전처리 데이터 DB 예시

- 전처리 과정을 거친 데이터는 MongoDB에 아래와 같은 형식으로 적재

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
출발 교통폴리곤ID	출발 읍면동ID	출발 시군구ID	출발 시도ID	출발 일자	출발 시간대	출발 체류지 유형	도착 교통폴리곤ID	도착 읍면동ID	도착 시군구ID	도착 시도ID	도착 일자	도착 시간대	도착 체류지 유형	연령대	성별	통행량	거리
110012	1101057	1101000	1100000	20180406	17	12	110013	1101058	1101000	1100000	20180406	18	11	20	1	15	100

&lt;그림 4- 4&gt; 전처리 데이터 형식

## 다. 모바일통신 빅데이터의 이종 데이터 간 결합 및 분석 목적별 다중 DB 재구조화

### 1) 모바일통신 빅데이터와 이종 데이터의 결합

- 모바일통신 빅데이터를 활용하여 각종 통행 관련 분석이 가능해졌지만, 데이터 특성에 따라 대상과 범위가 제한적임
- 다른 분야에서 발생한 서로 다른 데이터인 이종 데이터를 모바일통신 빅데이터와 시간공간적으로 매칭하고자 함
- 모바일통신 빅데이터와 결합을 위한 이종 데이터의 예로는 기초 통계자료, POI, 부동산 지가지수 등이 있음

## 라. 모바일통신 빅데이터 행정구역 및 분석 특성별 DB 재구조화

### 1) 모바일통신 빅데이터를 이용한 교통폴리곤 단위 집계 통행량 DB 재구조화

- 모바일 전처리 통행량 데이터를 그대로 서비스 하게 될 시 1시간 이상의 처리 시간이 소요되는 문제가 발생함
- 이는 서비스 시 문제로 판단되어 신속한 서비스를 위하여 경량화한 집계 통행량 DB 구축이 필요함
- 통행 유출과 유입 데이터 분석에 용이하도록 출발 기준, 도착 기준 행정구역 단위별로 집계 통행량 DB 구축함
- 집계 기준은 출발 지역과 도착 지역, 성별, 연령대, 출발 체류 특성, 도착 체류 특성의 조합으로 키를 생성하고 집계 기준 키와 일치하는 데이터들은 출발 시간대와 도착 시간대 기준으로 묶어 데이터를 배열 형태로 구성함

필드	_id	CG	A1	E1	A2	E2	A3	E3
Type	object	object 배열	Integer	Integer	Integer	Integer	Integer	Integer
컬럼 구성	"_id":{ "A" : 출발 분석구ID, "E" : 도착 분석구ID, "J" : 성별, "I" : 연령대, "D" : 출발 체류특성, "H" : 도착 체류특성 }	"CG":[ { "C" : 출발 시간대, "G" : 도착 시간대, "K" : [ { "L" : 기준일, "V" : 통행량, }, { "L" : 기준일, "V" : 통행량, } ]} ]	"A1": 출발 읍면동ID	"E1": 도착 읍면동ID	"A2": 출발 시군구ID	"E2": 도착 시군구ID	"A3": 출발 시도ID	"E3": 도착 시도ID

&lt;그림 4- 5&gt; 교통폴리곤 기준 통행량 데이터 구조

- 출발 교통폴리곤 ID 도착 교통폴리곤 ID, 체류특성, 연령대, 성별 필드를 키(필드명:\_id)로 생성하여 중복을 제거함
- 키의 조건과 일치하는 데이터들의 시간대를 배열화하여 집계함
- 배열화 된 시간대 그룹(필드명:CG)에 일자별 통행량을 그룹화하여 데이터를 생성함
- 해당 데이터를 생성함과 동시에 해당 교통폴리곤에 해당하는 A1~E3의 출발 도착 행정구역 ID 필드를 삽입하여 차후 행정구역 단위의 출/도착 데이터 생성 시 참조할 수 있도록 구성함
- 분석 조건에 맞는 데이터 그룹을 이용하여 약 145억 건의 데이터를 9억 건의 데이터로 줄일 수 있음
- View-T의 교통폴리곤 단위 서비스를 통신 데이터로 분석 시 1시간 이상이 걸렸지만 교통폴리곤 기준 통행량 DB 구축 후에는 5분 이내로 서비스됨
- 출발/도착의 읍면동, 시군구, 시도를 추가함으로써 특정 행정구역에 포함된 교통폴리곤 단위의 통행량 집계 서비스 시 처리 속도가 빠르며 읍면동 단위의 출발/도착 기준 통행량 구축 시 별도로 교통폴리곤과 매칭되는 행정구역들을 찾아서 추가하지 않아도 되기 때문에 구축 시간이 단축됨

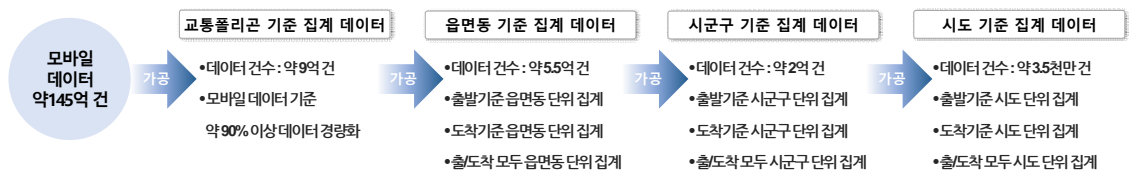
&lt;표 4-3&gt; 교통폴리곤 기준 집계 통행량 컬렉션의 구성

컬렉션명	TMOBILE_2017			
컬렉션 설명	교통폴리곤 기준의 집계 통행량			
데이터 수	889,081,736		데이터 크기	약 700GB
KEY Name	KEY	VALUE TYPE	데이터 구성	비고
식별자	_id	Object	_id : { “A” : 출발분석구코드, “E” : 도착분석구코드, “J” : 성별, “I” : 연령대, “D” : 출발체류특성, “H” : 도착체류특성 }	- A, E, D, H, J, I : integer - 체류특성 코드 ·H:11, N:12, C:13, S:14, R:15, X: 16 - 성별 코드 ·F:1, M:2
출도착시간대 인구수	CG	Object 배열	CG : [ { “C” : 출발시간대, “G” : 도착시간대, “K” : [{ “B”:출발날짜, “F”:도착날짜, “V”:인구수, }, { “B”:출발날짜, “F”:도착날짜, “V”:인구수, }] ]	- C, G, B, F, V : integer - K : object 배열 - V : _id, C, G, B, F에 해당하는 원시데이터의 인구수
출발 읍면동	A1	INT	“A1” : 출발 읍면동 코드	
도착 읍면동	E1	INT	“E1” : 도착 읍면동 코드	
출발 시군구	A2	INT	“A2” : 출발 시군구 코드	
도착 시군구	E2	INT	“E2” : 도착 시군구 코드	
출발 시도	A3	INT	“A3” : 출발 시도 코드	
도착 시도	E3	INT	“E3” : 도착 시도 코드	
분석구 간 거리	S	DOUBL E	“S” : 분석구 간 거리	

## 2) 모바일통신 빅데이터를 이용한 행정구역 단위 집계 통행량 DB 구축(추후 갱신 예정)

- 교통 폴리곤 기준 통행량 DB를 이용하여 행정구역 단위의 서비스를 하게 될 시 검색 및 처리 속도가 느려 원활한 서비스에 어려움이 있음. 처리 속도를 향상하고, 서비스를 원활하게 할 수 있도록 행정구역 단위의 DB 가공이 필요함
- 행정구역 단위 집계 통행량 DB 구축은 다음과 같이 진행함
  - 교통폴리곤 기준 통행량 데이터를 이용하여 각 행정구역 단위의 데이터를 생성함
  - 행정구역 단위의 통행량 DB는 GIS 분석 서비스 및 데이터 다운로드 시에 적합하도록 행정구역 단위별로 출발 행정구역 기준 집계 데이터, 도착 행정구역 기준 집계 데이터, 출발과 도착이 모두 행정구역 단위인 집계 데이터를 구축함
  - 출발 행정구역 기준 집계 데이터와 도착 행정구역 기준 집계 데이터는 크게 유출과 유입 데이터 분석에 사용되며 출발과 도착이 모두 행정구역 단위인 데이터는 지역 간, 지역 내 matrix 데이터 분석에 사용됨
  - 구축 순서는 읍면동 -> 시군구 -> 시도로 진행되며 상위 행정구역 단위 통행량 생성 시, 구축 시간을 최소화하기 위하여 집계한 하위 행정구역 단위 통행량을 이용하여 구축함  
(읍면동 단위 : 교통폴리곤 기준 통행량, 시군구 단위 : 읍면동 단위 통행량, 시도 단위 : 시군구 단위 통행량)
  - 교통폴리곤 기준 통행량과 마찬가지로 MongoDB의 배열 필드를 이용하여 집계함

## ■ 교통폴리곤 / 행정구역 집계 통행량 DB 구축 과정



## ■ 집계 통행량 DB 구조 - 출발 읍면동 기준 예시

필드	Object	Array	int	int	int	int	int
컬럼 구성	<pre> {   "id": {     "A1": 출발 읍면동 코드     "E": 도착 교통폴리곤 코드     "J": 성별     "I": 연령대     "D": 출발 체류특성     "H": 도착 체류특성   } } </pre>	<pre> CG: [ {   "C": 출발 시간대   "G": 도착 시간대   "K": [ {     "SD": 출발날짜     "ED": 도착날짜     "V": 통행량   }, ..... ] } ] </pre>	도착 읍면동 코드	출발 시군구 코드	도착 시군구 코드	출발 시도 코드	도착 시도 코드

&lt;그림 4- 6&gt; 집계 통행량 DB 구축 프로세스



- 행정구역 단위의 통행량에 대한 데이터 구조는 다음과 같음
  - 각 행정구역 DB에 맞게 출발 도착에 대한 행정구역 ID로 키를 구성

필드	_id					
기준	출발 읍면동	도착 읍면동	출발 시군구	도착 시군구	출발 시도	도착 시도
컬럼 구성	"_id":{ "L": 기준일 "A1": 출발 읍면동ID, "E": 도착 분석구ID "J": 성별, "I": 연령대, "D": 출발 체류특성, "H": 도착 체류특성 }	"_id":{ "L": 기준일 "A": 출발 분석구ID, "E1": 도착 읍면동ID "J": 성별, "I": 연령대, "D": 출발 체류특성, "H": 도착 체류특성 }	"_id":{ "L": 기준일 "A2": 출발 시군구ID, "E": 도착 분석구ID "J": 성별, "I": 연령대, "D": 출발 체류특성, "H": 도착 체류특성 }	"_id":{ "L": 기준일 "A": 출발 분석구ID, "E2": 도착 시군구ID "J": 성별, "I": 연령대, "D": 출발 체류특성, "H": 도착 체류특성 }	"_id":{ "L": 기준일 "A3": 출발 시도ID, "E": 도착 분석구ID "J": 성별, "I": 연령대, "D": 출발 체류특성, "H": 도착 체류특성 }	"_id":{ "L": 기준일 "A": 출발 분석구ID, "E3": 도착 시도ID "J": 성별, "I": 연령대, "D": 출발 체류특성, "H": 도착 체류특성 }

<그림 4- 7> 행정구역 단위의 출발/도착 기준 통행량 DB 구조

- 구축하고자 하는 행정구역 단위의 출발/도착 별로 기준 키(필드명: \_id)의 출발/도착 행정구역을 변경하여 집계함
- 교통폴리곤 기준 통행량과 동일하게 출발지 행정구역인 A1(읍면동), A2(시군구), A3(시도)와 도착지 행정구역인 E1(읍면동), E2(시군구), E3(시도)가 존재함
- 행정구역 단위로 통행량을 집계하기 때문에 교통폴리곤 기준 통행량보다 데이터 건수가 줄어들어 교통 폴리곤 이외 행정구역 단위 서비스 시 처리 속도가 향상됨
- 행정구역 단위별로 출발, 도착, 출/도착이 모두 행정구역인 데이터를 사용자 분석 조건에 맞게 데이터를 검색 할 수 있도록 하여 하나의 DB 사용보다 조금 더 안정된 서비스 제공 가능

&lt;표 4- 4&gt;읍면동 출발 기준 집계 통행량의 컬렉션의 구성

컬렉션명	TMOBILE_UMD_POLY_2017			
컬렉션 설명	읍면동 출발 기준의 집계 통행량			
데이터 수	531,880,964		데이터 크기	약 580GB
KEY Name	KEY	VALUE TYPE	데이터 구성	비고
식별자	_id	Object	_id : { “A1” : 출발 읍면동코드, “E” : 도착 분석구코드, “J” : 성별, “I” : 연령대, “D” : 출발 체류 특성, “H” : 도착 체류 특성 }	- A1, E, D, H, J, I : integer - 체류특성 코드 ·H:11, N:12, C:13, S:14, R:15, X: 16 - 성별 코드 ·F:1, M:2
출·도착 시간대 인구수	CG	Object 배열	CG : [ { “C” : 출발 시간대, “G” : 도착 시간대, “K” : [{ “B”:출발 날짜, “F”:도착 날짜, “V”:인구수, }, { “B”:출발 날짜, “F”:도착 날짜, “V”:인구수, }] ]	- C, G, B, F, V : integer - K : object 배열 - V : _id, C, G, B, F에 해당하는 원시데이터의 인구수
도착 읍면동	E1	INT	“E1” : 도착 읍면동 코드	
출발 시군구	A2	INT	“A2” : 출발 시군구 코드	
도착 시군구	E2	INT	“E2” : 도착 시군구 코드	
출발 시도	A3	INT	“A3” : 출발 시도 코드	
도착 시도	E3	INT	“E3” : 도착 시도 코드	

&lt;표 4- 5&gt; 읍면동 도착 기준 집계 통행량의 컬렉션의 구성

컬렉션명	TMOBILE_POLY_UMD_2017			
컬렉션 설명	읍면동 도착 기준의 집계 통행량			
데이터 수	530, 646, 371		데이터 크기	약 580GB
KEY Name	KEY	VALUE TYPE	데이터 구성	비고
식별자	_id	Object	_id : { “A” : 출발분석구코드, “E1” : 도착읍면동코드, “J” : 성별, “T” : 연령대, “D” : 출발체류특성, “H” : 도착체류특성 }	- A, E1, D, H, J, I : integer - 체류특성 코드 ·H:11, N:12, C:13, S:14, R:15, X: 16 - 성별 코드 ·F:1, M:2
출도착시간대 인구수	CG	Object 배열	CG : { { “C” : 출발시간대, “G” : 도착시간대, “K” : [{ “B”:출발날짜, “F”:도착날짜, “V”:인구수, }, { “B”:출발날짜, “F”:도착날짜, “V”:인구수, }] } }	- C, G, B, F, V : integer - K : object 배열 - V : _id, C, G, B, F에 해당하는 원시데이터의 인구수
출발읍면동	A1	INT	“A1” : 도착읍면동코드	
출발시군구	A2	INT	“A2” : 출발시군구코드	
도착시군구	E2	INT	“E2” : 도착시군구코드	
출발시도	A3	INT	“A3” : 출발시도코드	
도착시도	E3	INT	“E3” : 도착시도코드	

&lt;표 4- 6&gt; 읍면동 기준의 집계 통행량 컬렉션 구성

컬렉션명	TMOBILE_UMD_2017			
컬렉션 설명	기준 출도착 행정구역이 읍면동인 집계 통행량			
데이터 수	293, 991, 951		데이터 크기	약 480GB
KEY Name	KEY	VALUE TYPE	데이터 구성	비고
식별자	_id	Object	_id : { “A1” : 출발읍면동코드, “E1” : 도착읍면동코드, “J” : 성별, “I” : 연령대, “D” : 출발체류특성, “H” : 도착체류특성 }	- A1, E1, D, H, J, I : integer - 체류특성 코드 ·H:11, N:12, C:13, S:14, R:15, X: 16 - 성별 코드 ·F:1, M:2
출도착시간대 인구수	CG	Object 배열	CG : [ { “C” : 출발시간대, “G” : 도착시간대, “K” : [{ “B”:출발날짜, “F”:도착날짜, “V”:인구수, }, { “B”:출발날짜, “F”:도착날짜, “V”:인구수, }] ]	- C, G, B, F, V : integer - K : object 배열 - V : _id, C, G, B, F에 해당하는 원시데이터의 인 구수
출발시군구	A2	INT	“A2” : 출발시군구코드	
도착시군구	E2	INT	“E2” : 도착시군구코드	
출발시도	A3	INT	“A3” : 출발시도코드	
도착시도	E3	INT	“E3” : 도착시도코드	

&lt;표 4- 7&gt; 시군구 출발 기준 집계 통행량 컬렉션 구성

컬렉션명	TMOBILE_SGG_POLY_2017			
컬렉션 설명	시군구 출발 기준의 집계 통행량			
데이터 수	164, 414, 067		데이터 크기	약 450GB
KEY Name	KEY	VALUE TYPE	데이터 구성	비고
식별자	_id	Object	<pre> _id : {   "A2" : 출발시군구코드,   "E" : 도착분석구코드,   "J" : 성별,   "T" : 연령대,   "D" : 출발체류특성,   "H" : 도착체류특성 } </pre>	<ul style="list-style-type: none"> <li>- A2, E, D, H, J, I : integer</li> <li>- 체류특성 코드 <ul style="list-style-type: none"> <li>·H:11, N:12, C:13, S:14, R:15, X: 16</li> </ul> </li> <li>- 성별 코드 <ul style="list-style-type: none"> <li>·F:1, M:2</li> </ul> </li> </ul>
출도착시간대 인구수	CG	Object 배열	<pre> CG : [ {   "C" : 출발시간대,   "G" : 도착시간대,   "K" : [{     "B":출발날짜,     "F":도착날짜,     "V":인구수,   },   {     "B":출발날짜,     "F":도착날짜,     "V":인구수,   } ] } ] </pre>	<ul style="list-style-type: none"> <li>- C, G, B, F, V : integer</li> <li>- K : object 배열</li> <li>- V : _id, C, G, B, F에 해당하는 원시데이터의 인구수</li> </ul>
도착읍면동	E1	INT	"E1" : 도착읍면동코드	
도착시군구	E2	INT	"E2" : 도착시군구코드	
출발시도	A3	INT	"A3" : 출발시도코드	
도착시도	E3	INT	"E3" : 도착시도코드	

&lt;표 4- 8&gt; 시군구 도착 기준의 집계 통행량 컬렉션 구성

컬렉션명	TMOBILE_POLY_SGG_2017			
컬렉션 설명	시군구 도착 기준의 집계 통행량			
데이터 수	162,303,592	데이터 크기	약 450GB	
KEY Name	KEY	VALUE TYPE	데이터 구성	비고
식별자	_id	Object	_id : { “A” : 출발분석구코드, “E2” : 도착시군구코드, “J” : 성별, “T” : 연령대, “D” : 출발체류특성, “H” : 도착체류특성 }	- A, E2, D, H, J, I : integer - 체류특성 코드 ·H:11, N:12, C:13, S:14, R:15, X: 16 - 성별 코드 ·F:1, M:2
출도착시간대 인구수	CG	Object 배열	CG : { { “C” : 출발시간대, “G” : 도착시간대, “K” : [{ “B”:출발날짜, “F”:도착날짜, “V”:인구수, }, { “B”:출발날짜, “F”:도착날짜, “V”:인구수, }] } }	- C, G, B, F, V : integer - K : object 배열 - V : _id, C, G, B, F에 해당하는 원시데이터의 인구수
출발읍면동	A1	INT	“A1” : 도착읍면동코드	
출발시군구	A2	INT	“A2” : 출발시군구코드	
출발시도	A3	INT	“A3” : 출발시도코드	
도착시도	E3	INT	“E3” : 도착시도코드	

&lt;표 4- 9&gt; 시군구 기준의 집계 통행량 컬렉션 구성

컬렉션명	TMOBILE_SGG_2017			
컬렉션 설명	기준 출도착 행정구역이 시군구인 집계 통행량			
데이터 수	12, 690, 584		데이터 크기	약 150GB
KEY Name	KEY	VALUE TYPE	데이터 구성	비고
식별자	_id	Object	_id : { “A2” : 출발시군구코드, “E2” : 도착시군구코드, “J” : 성별, “T” : 연령대, “D” : 출발체류특성, “H” : 도착체류특성 }	- A2, E2, D, H, J, I : integer - 체류특성 코드 ·H:11, N:12, C:13, S:14, R:15, X: 16 - 성별 코드 ·F:1, M:2
출도착시간대 인구수	CG	Object 배열	CG : [ { “C” : 출발시간대, “G” : 도착시간대, “K” : [{ “B”:출발날짜, “F”:도착날짜, “V”:인구수, }, { “B”:출발날짜, “F”:도착날짜, “V”:인구수, }] ]	- C, G, B, F, V : integer - K : object 배열 - V : _id, C, G, B, F에 해당하는 원시데이터의 인구수
출발시도	A3	INT	“A3” : 출발시도코드	
도착시도	E3	INT	“E3” : 도착시도코드	

&lt;표 4- 10&gt; 시도 출발 기준의 집계 통행량 컬렉션 구성

컬렉션명	TMOBILE_SIDO_POLY_2017			
컬렉션 설명	시도 출발 기준의 집계 통행량			
데이터 수	33, 441, 520		데이터 크기	약 270GB
KEY Name	KEY	VALUE TYPE	데이터 구성	비고
식별자	_id	Object	_id : { “A3” : 출발시도코드, “E” : 도착분석구코드, “J” : 성별, “T” : 연령대, “D” : 출발체류특성, “H” : 도착체류특성 }	- A3, E, D, H, J, I : integer - 체류특성 코드 ·H:11, N:12, C:13, S:14, R:15, X: 16 - 성별 코드 ·F:1, M:2
출도착시간대 인구수	CG	Object 배열	CG : [ { “C” : 출발시간대, “G” : 도착시간대, “K” : [{ “B”:출발날짜, “F”:도착날짜, “V”:인구수, }, { “B”:출발날짜, “F”:도착날짜, “V”:인구수, }] ]	- C, G, B, F, V : integer - K : object 배열 - V : _id, C, G, B, F에 해당하는 원시데이터의 인구수
도착읍면동	E1	INT	“E1” : 도착읍면동코드	
도착시군구	E2	INT	“E2” : 도착시군구코드	
도착시도	E3	INT	“E3” : 도착시도코드	



&lt;표 4- 11&gt; 시도 도착 기준의 집계 통행량 컬렉션 구성

컬렉션명	TMOBILE_POLY_SIDO_2017			
컬렉션 설명	시도 도착 기준의 집계 통행량			
데이터 수	32,907,812		데이터 크기	약 270GB
KEY Name	KEY	VALUE TYPE	데이터 구성	비고
식별자	_id	Object	_id : { “A” : 출발분석구코드, “E3” : 도착시도코드, “J” : 성별, “I” : 연령대, “D” : 출발체류특성, “H” : 도착체류특성 }	- A, E3, D, H, J, I : integer - 체류특성 코드 ·H:11, N:12, C:13, S:14, R:15, X: 16 - 성별 코드 ·F:1, M:2
출도착시간대 인구수	CG	Object 배열	CG : [ { “C” : 출발시간대, “G” : 도착시간대, “K” : [{ “B”:출발날짜, “F”:도착날짜, “V”:인구수, }, { “B”:출발날짜, “F”:도착날짜, “V”:인구수, }] ]	- C, G, B, F, V : integer - K : object 배열 - V : _id, C, G, B, F에 해당하는 원시데이터의 인구수
출발읍면동	A1	INT	“A1” : 도착읍면동코드	
출발시군구	A2	INT	“A2” : 출발시군구코드	
출발시도	A3	INT	“A3” : 출발시도코드	

&lt;표 4- 12&gt; 시도 기준의 집계 통행량 컬렉션 구성

컬렉션명	TMOBILE_SIDO_2017			
컬렉션 설명	기준 출도착 행정구역이 시도인 집계 통행량			
데이터 수	116, 070		데이터 크기	약 18GB
KEY Name	KEY	VALUE TYPE	데이터 구성	비고
식별자	_id	Object	_id : { “A3” : 출발시도코드, “E3” : 도착시도코드, “J” : 성별, “I” : 연령대, “D” : 출발체류특성, “H” : 도착체류특성 }	- A3, E3, D, H, J, I : integer - 체류특성 코드 ·H:11, N:12, C:13, S:14, R:15, X: 16 - 성별 코드 ·F:1, M:2
출도착시간대 인구수	CG	Object 배열	CG : [ { “C” : 출발시간대, “G” : 도착시간대, “K” : [{ “B”:출발날짜, “F”:도착날짜, “V”:인구수, }, { “B”:출발날짜, “F”:도착날짜, “V”:인구수, }] }] }	- C, G, B, F, V : integer - K : object 배열 - V : _id, C, G, B, F에 해당하는 원시데이터의 인구수

### 3) 분석 특성에 따른 집계 통행량 DB 재구조화

- View-T 서비스 만족도를 높이기 위하여 검색, 표출 속도를 개선하기 위한 작업이 필요함
- 서비스 속도를 개선하기 위하여 모바일통신 빅데이터 속성정보 및 통행목적 정보를 기준으로 데이터를 재구성함

■ 분석 특성 집계 통행량 조건 구분

분류	옵션	조건			
분석 기간	주중	월, 화, 수, 목, 금			
	주말	토, 일			
이동 목적	출근	집 - 회사	06시~09시 - 07시~10시	20대 ~ 60대	
	등교	집 - 학교	06시~09시 - 07시~10시	10대 ~ 20대	
	퇴근	회사 - 집	18시~21시 - 19시~22시	20대 ~ 60대	
	하교	학교 - 집	15시~21시 - 16시~21시	10대 ~ 20대	
	귀가	정기적, 비정기적 잠재체류지 - 집	18시~20시 - 19시~23시	전체 연령대	
	여가	집, 회사, 학교, 정기적, 비정기적 잠재체류지 - 정기적, 비정기적 잠재체류지	전체 시간대	전체 연령대	
	사적업무 후 복귀	정기적, 비정기적 잠재체류지 - 회사, 학교	전체 시간대	10대 ~ 60대	
	* ① 도착지 타임이 -인 경우, ② 회사(C)→학교(S) 또는 학교(S)→회사(C) 타임인 경우, ③ 출발일자와 도착일자가 다른 경우는 제외				
성별	남자	-			
	여자	-			

■ 분석 특성 통행량 DB 구조

필드	Object	int	int	Double
컬럼 구성	_id : { "A1": 출발 읍면동 "E1": 도착 읍면동 "J": 성별 "P": 통행목적 코드 }	명 일 집계 통행량	주말 집계 통행량	읍면동 직선 거리
필드	int	int	int	int
컬럼 구성	출발 시군구	도착 시군구	출발 시도	도착 시도

<그림 4- 8> 분석 특성 집계 통행량 데이터 재구성 조건 및 DB 구조

- 분석 특성별 집계 통행량 DB 구축은 교통폴리곤 단위 통행량 DB와 행정구역 단위의 통행량 DB의 키 조건에 해당하는 시간대 배열과 일자 배열을 특성에 맞게 데이터 압축 형태로 구축하고자 함

#### ① 분석 특성(요일, 기간, 시간, 연령대)에 해당하는 집계 통행량 조건 테스트

- 통신 데이터의 특성 중 사용자가 자주 바꾸는 조건을 고려하여 요일, 기간, 시간, 연령대 속성정보를 그룹화한 집계 데이터를 가공함
- 요일은 주중/주말, 분석 기간은 계절별(봄/여름/가을/겨울), 분석 시간은 오전/오후 첨두시간대인 출근 시간대와 퇴근 시간대 및 점심 시간대와 심야 시간대, 성별(남자/여자), 연령대는 청소년(학생)/경제활동 연령/고령으로 속성별로 그룹화를 함

- 분석 특성(요일, 기간, 시간, 연령대)으로 설계한 조건은 아래와 같음

❖ 조건 설명(1)

분류	옵션	조건
분석 요일	주중	월, 화, 수, 목, 금
	주말	토, 일
분석 기간	봄	3월, 4월, 5월
	여름	6월, 7월, 8월
	가을	9월, 10월, 11월
	겨울	12월, 1월, 2월
분석 시간	출근 시간대	6시 ~ 9시
	퇴근 시간대	15시 ~ 21시
	점심 시간대	11시 ~ 14시
	심야 시간대	0, 1, 2, 23시
연령대	청소년 (학생)	0대, 10대, 20대
	경제활동연령	20대, 30대, 40대, 50대, 60대
	고령	70대 ~

❖ 조건 설정 부분 UI 설계 예시(1)

1. 분석 요일  
☒ 전체 ☐ 주중 ☐ 주말

2. 분석 기간  
☒ 봄 ☐ 여름 ☐ 가을 ☐ 겨울 ☐ 휴가철  
 3월, 4월, 5월을 기준으로 함

3. 분석 시간  
☒ 출근 ☐ 퇴근 ☐ 점심 ☐ 심야  
 오전 6시부터 9시까지 기준으로 함

4. 성별  
☒ 전체 ☐ 남자 ☐ 여자

5. 연령대  
☒ 전체 ☐ 청소년  
☐ 경제활동연령 ☐ 노인

<그림 4-9> 분석 특성(요일, 기간, 시간, 연령대) 테스트 조건 설명과 조건 설정 UI 설계 예시

- 분석 특성(요일, 기간, 시간, 연령대) 조건에 대한 시간대 코드별 평일과 주말을 집계한 통행량 데이터 정의서 예시는 다음과 같음

분석 특성을 고려한 교통플리곤 단위 집계 통행량 DB (1)									
컬렉션			교통플리곤 단위 시간과 기간을 그룹화한 모바일 통행량 데이터						
컬럼 ID	_id	CG	A1	E1	A2	E2	A3	E3	S
타입	object	array	int	int	int	int	int	int	double
	_id : { "A": 출발 교통플리곤 코드 "E": 도착 교통플리곤 코드 "J": 성별 "I": 연령대 "D": 출발 제류특성 "H": 도착 제류특성 } 	CG : [ { "T": 시간 그룹 코드 "WD": 주중 통행량 "WE": 주말 통행량 } ] 	출발 음면동 코드	도착 음면동 코드	출발 시군구 코드	도착 시군구 코드	출발 시도 코드	도착 시도 코드	출/도착 교통플리곤 간 거리
비고	A, E, D, H, J, I : int	- T, WD, WE : int T : 1 (출근 시간대) - 6시 ~ 9시 T : 2 (퇴근 시간대) - 18시 ~ 21시 T : 3 (점심 시간대) - 11시 ~ 14시 T : 4 (심야 시간대) - 0, 1, 2, 23시 	7자리	7자리	7자리	7자리	7자리	7자리	Km단위

<그림 4-10> 시간대 코드별 평일/주말 집계 통행량 데이터 정의서 예시

- 분석 특성(요일, 기간, 시간, 연령대) 조건에 대한 계절별 시간대 코드별 평일과 주말을 집계한 통행량 데이터 정의서 예시는 다음과 같음

분석 특성을 고려한 교통플리곤 단위 집계 통행량 DB (2)									
컬렉션			교통플리곤 단위 계절, 시간과 기간을 그룹화한 모바일 통행량 데이터						
컬럼 ID	_id	CG	A1	E1	A2	E2	A3	E3	S
타입	object	array	int	int	int	int	int	int	double
	_id : { "A": 출발 교통플리곤 코드 "E": 도착 교통플리곤 코드 "J": 성별 "I": 연령대 "D": 출발 제류특성 "H": 도착 제류특성 }	CG : { { "N": 계절 코드 "T": 시간 그룹 코드 "WD": 주중 통행량 "WE": 주말 통행량 } }	출발 음면종 코드	도착 음면종 코드	출발 시군구 코드	도착 시군구 코드	출발 시도 코드	도착 시도 코드	출/도착 교통플리곤 간 거리
비고	A, E, D, H, J, I : int	- T, WD, WE : int N : 1 (봄) - 3월 ~ 5월 N : 2 (여름) - 6월 ~ 8월 N : 3 (가을) - 9월 ~ 11월 N : 4 (겨울) - 12, 1, 2월 T : 1 (출근 시간대) - 6시 ~ 9시 T : 2 (퇴근 시간대) - 18시 ~ 21시 T : 3 (점심 시간대) - 11시 ~ 14시 T : 4 (심야 시간대) - 0, 1, 2, 23시	7자리	7자리	7자리	7자리	7자리	7자리	Km단위

<그림 4- 11> 계절별 시간대 코드별 평일/주말 집계 통행량 데이터 정의서 예시

## ② 분석 특성(통행목적)에 해당하는 집계 통행량 조건 테스트

- 통신 데이터의 특성을 고려하여 체류 특성, 시간대, 연령대 속성정보의 조합으로 통행 목적에 대한 집계 데이터를 가공함
- 분석 기간은 주중/주말, 통행목적은 출근/등교/퇴근/하교/귀가/여가/식사, 성별은 남자/여자로 그룹화를 함
- 분석 특성(통행목적)으로 설계한 조건은 아래와 같음

### ❖ 조건 설명(2)

분류	옵션	조건		
분석 기간	주중	월, 화, 수, 목, 금		
	주말	토, 일		
통행 목적	출근	집 - 회사	06시 ~ 09시	20대 ~ 60대
	등교	집 - 학교	06시 ~ 09시	0대 ~ 20대
	퇴근	회사 - 집	18시 ~ 21시	20대 ~ 60대
	하교	학교 - 집	15시~16시 - 17시~23시	0대 ~ 20대
	귀가	비정기적 잠재체류지 - 집	19시~21시 - 21시~23시	전체 연령대
	여가	모든 트립타입 - 비정기적 잠재체류지	전체 시간대	전체 연령대
	식사	회사, 학교 - 비정기적 잠재체류지	11시 ~ 14시	20대 ~ 60대
성별	남자	-		
	여자	-		

### ❖ 조건 설정 부분 UI 설계 예시(2)

#### 1. 기간

☒ 전체 ☐ 주중 ☐ 주말

#### 2. 통행목적

☒ 출근
 ☐ 퇴근
 ☐ 등교
 ☐ 하교  
☐ 귀가
 ☐ 여가
 ☐ 식사

오전 6시 ~ 오전 9시 사이에 출발지(집)에서 목적지(회사)로 이동하는 것을 통행 목적으로 봄

#### 3. 성별

☒ 전체 ☐ 남자 ☐ 여자

<그림 4- 12> 분석 특성(통행목적) 테스트 조건 설명과 조건 설정 UI 설계 예시

- 분석 특성(통행목적) 조건에 대한 집계 통행량 데이터 정의서 예시는 다음과 같음

분석 특성을 고려한 교통플리곤 단위 집계 통행량 DB (3)										
컬렉션		교통플리곤 단위 통행목적(체류특성, 시간, 연령조합) 그룹화한 모바일 통행량 데이터								
컬럼 ID	_id	WD	WE	A1	E1	A2	E2	A3	E3	S
타 입	object	int	int	int	int	int	int	int	int	double
	_id: {             "A": 출발 교통플리곤 코드 "E": 도착 교통플리곤 코드 "J": 성별 "P": 통행목적           }	주중 통행량	주말 통행량	출발 읍면동 코드	도착 읍면동 코드	출발 시군구 코드	도착 시군구 코드	출발 시도 코드	도착 시도 코드	출/도착 교통플리곤 간 거리
비고	A, E, J, P: int	-	-	7자리	7자리	7자리	7자리	7자리	7자리	Km단위

<그림 4- 13> 분석 특성(통행목적) 집계 통행량 데이터 정의서 예시

### ③ 통행목적별 집계 통행량 DB 구축

- 분석 특성별 집계 통행량 DB 구축 테스트 결과를 기반으로 조건을 재구성하여 DB 구축
- 통행목적 중 ‘식사’는 ‘사적 업무 후 복귀’로 변경하고, 이 외의 통행목적의 조건은 출발/도착 시각, 체류 특성, 연령대가 조금씩 변경됨
- 통행목적별 집계 통행량 데이터 조건은 아래와 같음

분류	옵션	조건		
분석 기간	주중	월, 화, 수, 목, 금		
	주말	토, 일		
통행 목적	출근	집 - 회사	06시~09시 - 07시~10시	20대 ~ 60대
	등교	집 - 학교	06시~09시 - 07시~10시	10대 ~ 20대
	퇴근	회사 - 집	18시~21시 - 19시~22시	20대 ~ 60대
	하교	학교 - 집	15시~21시 - 16시~21시	10대 ~ 20대
	귀가	정기적, 비정기적 잠재체류지 - 집	18시~20시 - 19시~23시	전체 연령대
	여가	집, 회사, 학교, 정기적, 비정기적 잠재체류지 - 정기적, 비정기적 잠재체류지	전체 시간대	전체 연령대
	사적업무 후 복귀	정기적, 비정기적 잠재체류지 - 회사, 학교	전체 시간대	10대 ~ 60대
* ① 도착지 타입이 ‘-’인 경우, ② 회사(C)→학교(S) 또는 학교(S)→회사(C) 타입인 경우, ③ 출발일자와 도착일자가 다른 경우는 제외				
성별	남자	-		
	여자	-		

<그림 4- 14> 통행목적별 집계 통행량 데이터 조건

## 마. 모바일통신 빅데이터 가공 검증

### 1) 재구조화된 DB 검증 목적

- 데이터 가공 과정에서의 데이터 손실을 최소화하기 위하여 데이터 검증 항목을 도출하고 이를 기반으로 데이터를 검증할 수 있도록 함

### 2) 재구조화된 DB별 데이터 검증

- 교통폴리곤 기준 통행량 및 행정구역 단위로 출발 기준, 도착 기준, 출발/도착 기준 통행량 데이터에 대한 검증 항목을 도출 후, 이를 기반으로 검증 진행
- 검증방식은 모바일통신 빅데이터 기반 통행DB와 행정구역이 매칭된 모바일통신 빅데이터(전처리 작업 후 데이터)를 비교 검증하여 데이터 손실 여부를 검증함
- 교통폴리곤 기준 통행량 DB 검증 방법
  - 교통폴리곤 기준 통행량 DB를 가공하기 위해 참조한 모바일통신 빅데이터 전처리 DB와 항목별 통행량 합계를 비교하여 차이 발생 시 데이터 재가공 및 검증함
  - 모든 항목별 통행량 합계가 동일시 검증 완료
- 교통폴리곤 기준 통행량 DB 검증 항목은 아래와 같음

<표 4- 13> 교통폴리곤 기준 통행량 DB 검증 항목

검 증 항 목	비 고
원시 전처리 데이터와 가공 결과 데이터에 대한 출발/도착 교통폴리곤 ID별 통행량 합계 비교	출발/도착 각각의 값으로 비교 검증
원시 전처리 데이터와 가공 결과 데이터에 대한 출발/도착 시간대별 통행량 합계 비교	
원시 전처리 데이터와 가공 결과 데이터에 대한 출발/도착의 체류 특성별 통행량 합계 비교	
원시 전처리 데이터와 가공 결과 데이터에 대한 연령대별 통행량 합계 비교	-
원시 전처리 데이터와 가공 결과 데이터에 대한 성별 통행량 합계 비교	
원시 전처리 데이터와 가공 결과 데이터에 대한 교통폴리곤별 통행량 합계 비교	구축되어진 DB별 출발 또는 도착의 행정구역 ID로 비교함
원시 전처리 데이터와 가공 결과 데이터에 대한 읍면동별 통행량 합계 비교	
원시 전처리 데이터와 가공 결과 데이터에 대한 시군구별 통행량 합계 비교	
원시 전처리 데이터와 가공 결과 데이터에 대한 시도별 통행량 합계 비교	

## ❖ 교통폴리곤 기준 통행량 DB 검증 쿼리 예시

```

1 db.TMOBILE_2017.aggregate( [
2   {
3     $match:{
4       $and : [
5         {"A3":{"$gt" : 2900000} },
6         {"_id.D": 11 },
7         {"_id.H": 13 },
8         {"_id.I":{"$in" : [20,30,40,50,60]} },
9         {"$expr":{"$ne" : ["$_id.A", "$_id.E"]}} },
10    ],
11  },
12  {
13    $unwind: "$CG"
14  },
15  {
16    $unwind: "$CG.K"
17  },
18  {
19    $match:{
20      $and : [
21        {"CG.C" : {"$in" : [6,7,8,9]} },
22        {"CG.G" : {"$in" : [7,8,9,10]} },
23        {"$expr":{"$eq" : ["$CG.K.SD", "$CG.K.ED"]}} },
24    ],
25  },
26  {
27    $group : {
28      "_id" : "$E3",
29      "SUM" : {"$sum":"$CG.K.V"}
30    },
31  },
32  {
33    $limit : 100
34  },
35  ],{allowDiskUse : true})

```

## ❖ 쿼리 결과 예시

```

1 /* 1 */
2 {
3   "_id" : 1100000,
4   "SUM" : 25832885
5 },
6 /* 2 */
7 {
8   "_id" : 3700000,
9   "SUM" : 19676116
10 },
11 /* 3 */
12 {
13   "_id" : 3300000,
14   "SUM" : 12848713
15 },
16 /* 4 */
17 {
18   "_id" : 3100000,
19   "SUM" : 86253471
20 },
21 /* 5 */

```

&lt;그림 4-15&gt; 교통폴리곤 기준 통행량 DB 검증 쿼리와 결과 예시(시도별 통행량 확인)

- 행정구역 단위 통행량 DB 검증 방법
  - 읍면동 단위 DB를 가공하기 위해 참조한 교통폴리곤 단위 통행량 DB와 항목별 통행량 합계를 비교하여 차이가 발생했는지 검증함
  - 시군구 단위 DB를 가공하기 위해 참조한 읍면동 단위 통행량 DB와 항목별 통행량 합계를 비교하여 차이가 발생했는지 검증함
  - 시도 단위 DB를 가공하기 위해 참조한 시군구 단위 통행량 DB와 항목별 통행량 합계를 비교하여 차이가 발생했는지 검증함
  - 차이가 발생했을 경우, 집계 코드를 수정하여 데이터 재생성 후 다시 검증 진행
  - 모든 항목별 통행량 합계가 동일할 때까지 진행
- 행정구역 단위 통행량 DB 검증 항목은 아래와 같음

&lt;표 4-14&gt; 행정구역 단위 통행량 합계 비교 검증 항목

검 증 항 목	비 고
행정구역이 매칭된 모바일통신 빅데이터의 합계와 시도 단위 통행량 합계	출발/도착 각각의 값으로 비교 검증
행정구역이 매칭된 모바일통신 빅데이터의 합계와 시군구 단위 통행량 합계	
행정구역이 매칭된 모바일통신 빅데이터의 합계와 읍면동 단위 통행량 합계	



- 통행목적별 집계 통행량 DB 검증 방법
  - 통행목적별 집계 통행량은 교통폴리곤 기준 통행량 DB와 행정구역 단위 통행량 DB 항목별 통행량 합계를 비교하여 차이가 발생했는지 검증함
  - 차이가 발생했을 경우, 집계 코드를 수정하여 데이터 재생성 후 다시 검증 진행
  - 모든 항목별 통행량 합계가 동일할 때까지 진행
- 통행목적별 집계 통행량 DB 검증 항목은 아래와 같음

<표 4- 15> 통행목적별 집계 통행량 합계 비교 검증 항목

비교 대상	비교 조건	비교
교통폴리곤 기준 통행량 데이터	- 주중/주말 조건 통행량 합계 - 통행목적 조건 통행량 합계 성별 통행량 합계	출발/도착 각각의 값으로 비교 검증
행정구역 단위 출발/도착 통행량 데이터		

### 제3절 교통지표 개발 및 분석 서비스 제공

#### 1. View-T 시스템 개요

##### 가. View-T 시스템의 구성

- View-T 시스템은 다양한 분석 도구와 지표, 및 정보를 이용자에게 제공하는 서비스 기능과 서비스 운영, 모니터링 등을 위한 관리자 기능으로 구성



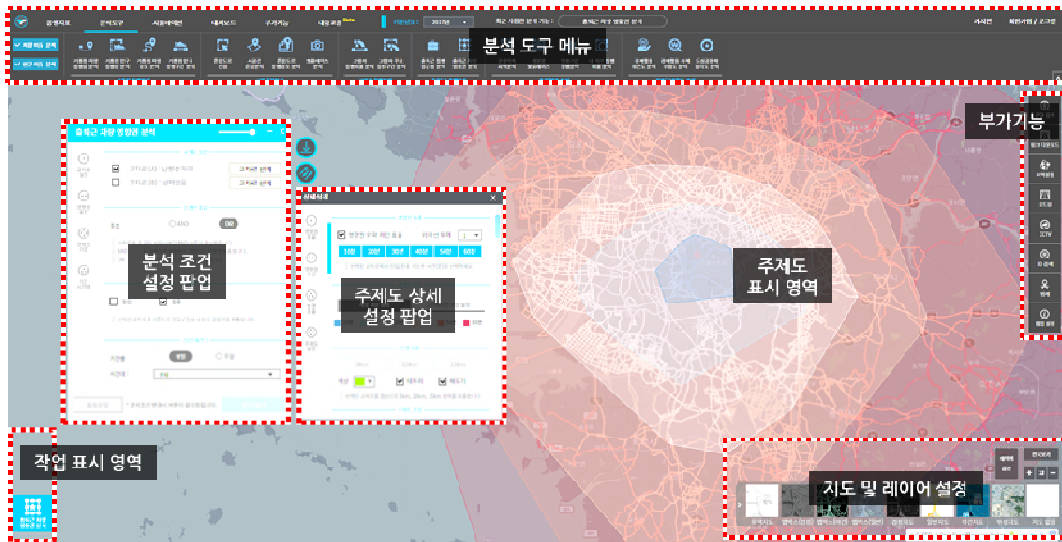
<그림 4-16> View-T 시스템 구성도

- 서비스 기능 : 교통 및 통신 데이터를 시각적으로 분석할 수 있는 Expert & Light 분석 도구와 분석 도구와 관련된 다양한 정보가 제공되며, 등록된 회원에 한하여 분석 결과와 교통지표 다운로드 서비스 등을 제공
- 관리자 기능 : 회원의 서비스 이용 권한을 관리하며, 회원의 서비스 이용 현황 및 다운로드 데이터 통계, 그리고 회원의 요청이나 문의에 대응하는 시스템을 제공

## 나. 분석 도구의 화면 구성

### 1) 심층분석 도구 (Expert) 화면 구성

- 심층분석 도구는 전문 이용자가 다양한 조건을 세밀하게 설정하여 보다 전문적이고 자세한 분석 결과를 얻을 수 있는 화면을 제공

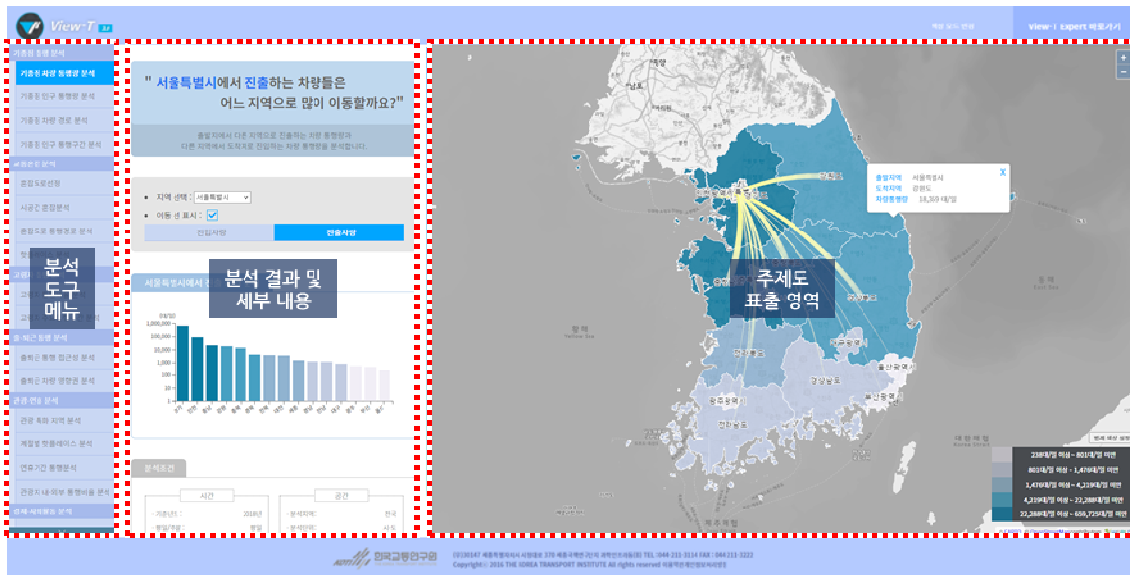


<그림 4- 17> 심층분석 도구 화면 구성

- 화면 전체를 차지하는 주제도 영역 위에 분석 도구 메뉴바가 있고, 오른쪽에 부가 기능과 지도 및 레이어 설정 UI를 배치
- 메뉴바의 메뉴 클릭 시 해당 분석 도구의 조건 설정 팝업과 작업 표시 영역에 분석 도구 아이콘 표시 (분석 조건 설정 팝업은 필요에 따라 드래그하여 위치 이동 가능)
- 분석 조건 설정 팝업에서 조건 설정 후 분석하기 버튼 클릭 시 주제도에 결과 표시 (또는 별도의 지정된 영역에서 분석 결과 표시)
- 분석 조건 설정 팝업에서 분석 결과 표시 후 상세 설정 팝업을 추가로 열어 표시된 결과의 표현 방식을 보기 쉽도록 재설정 가능

### 2) 간편분석 도구 (Light) 화면 구성

- 간편분석 도구는 일반 이용자가 쉬운 주제로 빠르고 간편하게 분석 결과를 확인하여 접근성을 높이기 위한 화면을 제공



<그림 4-18> 간편분석 도구 화면 구성

- 간편분석 도구의 화면 구성은 복잡한 심층분석 도구 화면을 단순화하여 크게 메뉴, 결과 및 세부 내용, 주제도의 영역으로 구분
- 메뉴 선택 시 해당 분석 도구의 사용 방법 및 목적 등의 자세한 설명 내용이 결과 및 세부 내용 영역에 표시하고 미리 설정된 조건에 해당하는 분석 결과를 주제도에 표시
- 분석 결과 및 세부 내용 영역에서는 미리 설정된 조건 이외에도 간단한 옵션 변경을 통해 다양하게 변화되는 분석 결과를 확인할 수 있음

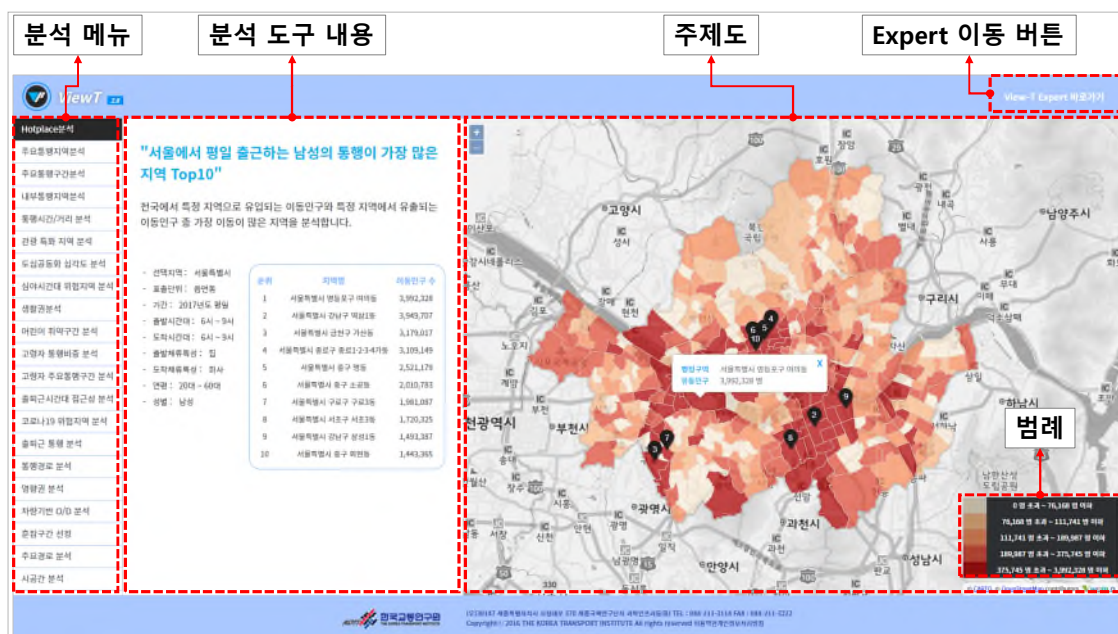
## 2. View-T Light

### 가. 개요

- View-T Light는 좀 더 간단하게, 빠르게, 이해하기 쉽게 View-T의 분석 도구들을 사용하자는 의도에서 개발된 분석 서비스로, 기존 View-T 주요 분석 도구들의 목적에 최적화된 간소화 버전을 제공하여 사용자가 몇 번의 클릭만으로 의도한 분석 결과가 시각화된 모습을 볼 수 있어 기능성을 강조하고, 사용자가 View-T 전반에 대해 직관적으로 이해할 수 있도록 도움
- 기존 View-T는 방대한 교통 데이터 위에서 복잡한 UI, 상세한 분석 조건 설정을 통하여 사용자가 원하는 차량 및 통신 이동 통행을 분석 및 시각화하고 있었으나, 이는 View-T에 대한 이해도가 낮은 사용자에게는 시스템 전반의

기능을 이해하기 어렵고, 친숙도가 떨어지는 애로사항이 있었음

- View-T Light는 각 주요 분석 도구 메뉴, 각 도구마다 간단한 설명과 기본 분석 설정값, 주제도 및 범례로 구성
  - View-T Light에서 사용할 수 있는 기능은 기존 View-T의 분석 도구 중 차량량과 통신 각 파트별 주요 분석 도구들만을 제공
  - 메뉴에서 각 분석 도구를 선택하면 이미 설정되어있는 기본 설정값과 해당 설정값에 최적화된 정제 데이터를 이용하여 우측 주제도 상에 분석 결과를 표출
  - 메뉴와 주제도 사이에는 해당 분석 도구의 대표적인 캐치프레이즈, 분석 내용에 대한 간단한 설명과 함께 분석에 사용된 기본 설정값을 표출
  - 분석 도구에 따라 이미지, 테이블 차트, 그래프 등을 함께 시각화하여 제공



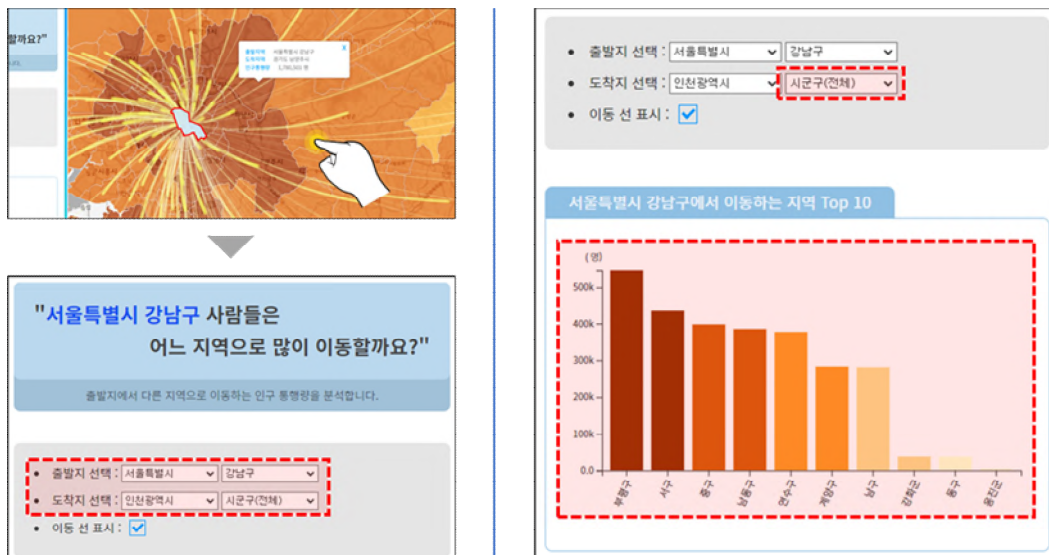
<그림 4-19> View-T Light 화면 구성

## 나. View-T Light 분석 기능

### 1) 주요 통행 분석

#### ○ 주요 통행지역 분석

- 특정 출발지에서 특정 도착지의 지역별 통행량을 간단하게 분석
- 지도 위에서 출발지와 선택지를 각각 선택하는 복잡한 조작 방식을 개선하기 위하여 분석 결과 영역에서 콤보박스를 통해 간편하게 출발/도착지를 선택할 수 있도록 개선
- 도착지 선택 시 시군구 지역 전체를 선택할 수 있게 하여 해당 선택한 지역들의 순위를 비교할 수 있도록 분석 결과 표시
- 마우스 클릭만으로 출발지와 도착지를 선택하여 유동 인구를 파악할 수 있는 기능
- 출발지를 제외한 다른 시/군/구를 선택하면 기존 선택 시/군/구에서 새로 선택한 시군구로의 유동 인구를 표출
- 선택된 2개의 시/군/구중 하나를 클릭하면 해당 지역이 선택 해제되며 나머지 선택 지역에서 전국으로의 유동 인구를 표출
- 기본 주제도 출력은 강남구를 출발지로 하는 전국 시군구로의 유동 인구를 표출
- 우측 하단에는 범례를 표시

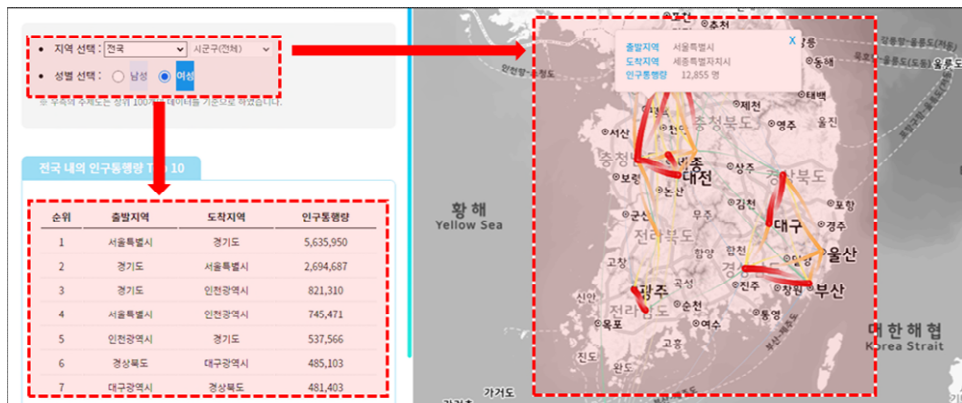


<그림 4- 20> 출발/도착 지역 선택방식 변경(좌), 도착지 하위 행정구역 전체 선택(우)



○ 주요 통행 구간 분석

- 통행이 많이 일어나는 구간을 간단하게 분석
- 기존 지도에서 시군구 단위로 선택할 수 있던 지역 선택의 범위를 콤보박스를 이용하여 시도 단위까지 선택 가능하게 하고, 전국을 선택할 수 있도록 개선
- 기존에는 여성의 데이터로만 분석하였으나 분석 조건에 남성을 선택할 수 있는 UI를 추가하여 성별에 따라 분석할 수 있도록 개선
- 시도 간 통행이 많이 일어나는 구간을 분석한 내용을 시각화 함
- 기본 주제도 출력은 전국의 시도단위에서 각각 시도에서의 출발 및 도착의 이동인구 수를 희망 선도로 표현함
- 주제도 상에 마우스 오버 이벤트로 해당 희망 선도의 출발지역, 도착지역, 통행자 수를 표현
- 좌측에는 우측 주제도의 조건 및 상위 10개 희망 선도의 내용을 표현
- 우측 하단에는 범례를 표시

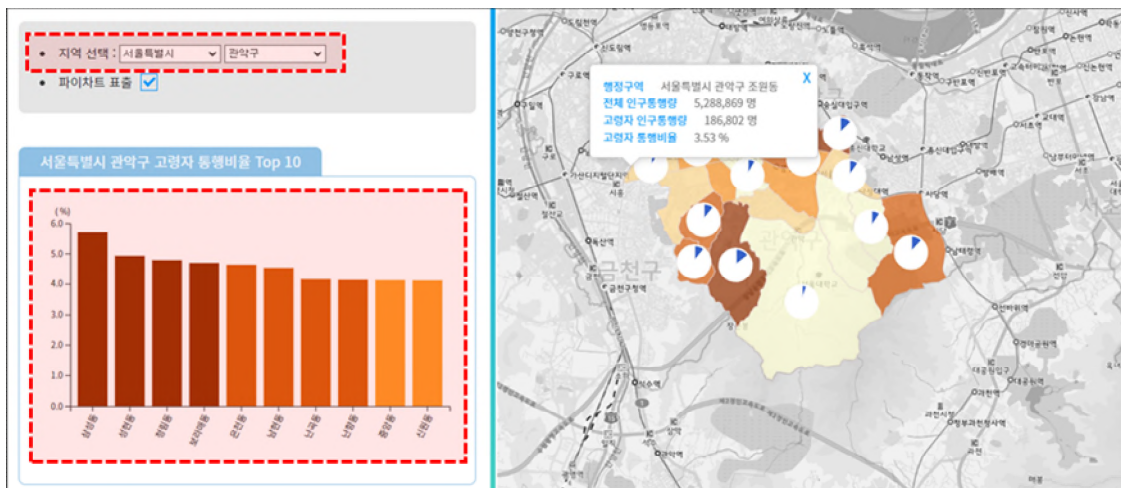


<그림 4- 21> 기종점 인구 통행 구간 분석의 시도단위 전국 분석 결과 표출

## 2) 고령자 통행 분석

### ○ 고령자 통행 비중 분석

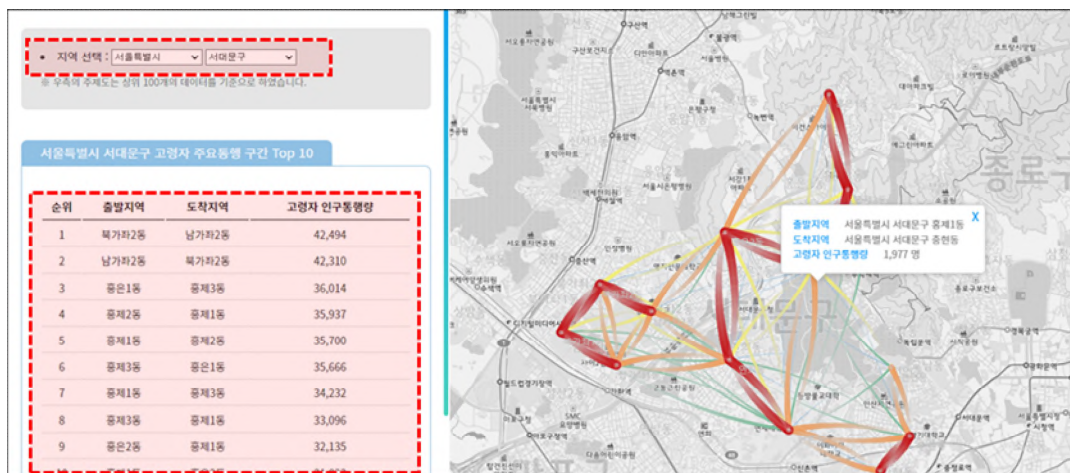
- 고령자의 통행 비중을 분석하기 위한 기능
- 기존 지도를 클릭하여 시도 단위로 분석 지역을 선택할 수 있던 지역 범위를 시군구 단위까지 선택 가능하게 하고, 읍면동 단위까지의 분석 결과를 표출할 수 있도록 개선
- 모바일통신 빅데이터에서 고령자(연령대:70~110대)의 통행 발생량을 지역별로 추출
- 데이터 가공 방법
  - . ㉠ 모바일통신 빅데이터에서 고령자(연령대: 70~110대)의 통행 발생량을 지역별로 추출
  - . ㉡ 모바일통신 빅데이터에서 전체 통행 발생량을 지역별로 추출
  - . 고령자 통행 비중 계산 ((㉠ / ㉡) \* 100)
- 주제도 상의 고령자 비중을 파이 차트로 표시
- 지도의 시도 단위의 행정구역을 클릭하면 사용자가 선택한 시도의 고령자 통행 비중에 대한 결과를 표출해주며, 마우스 오버 시 고령자 통행량과 전체 통행량의 속성정보를 표출
- 우측 하단에는 범례를 표시



<그림 4- 22> 고령자 통행 비중 분석의 읍면동 단위 분석 결과 표출



- 고령자 주요 통행 구간 분석
  - 고령자의 주요 통행 구간을 분석하기 위한 기능
  - 데이터 가공 방법
    - 모바일통신 빅데이터에서 고령자(연령대: 70~110대)의 통행만 추출하여 기종점 통행량 추출
  - 희망 선도로 고령자 주요 통행 구간을 표출함
  - 지도의 시도 단위의 행정구역을 클릭하면 사용자가 선택한 시도의 고령자 주요 통행 구간에 대한 결과를 표출해주며, 마우스 오버 시 출발지역과 도착지역, 통행량의 속성정보를 표출
  - 주제도 상에 마우스 오버 이벤트로 해당 지역의 고령자 통행량을 표현
  - 좌측에는 주요 통행 구간 상위 10개 순위를 표시
  - 우측 하단에는 범례를 표시
  - 기존 지도를 클릭하여 시도 단위로 분석 지역을 선택할 수 있던 지역 범위를 시군구 단위까지 선택 가능하게 하고, 읍면동 단위까지의 분석 결과를 표출할 수 있도록 개선

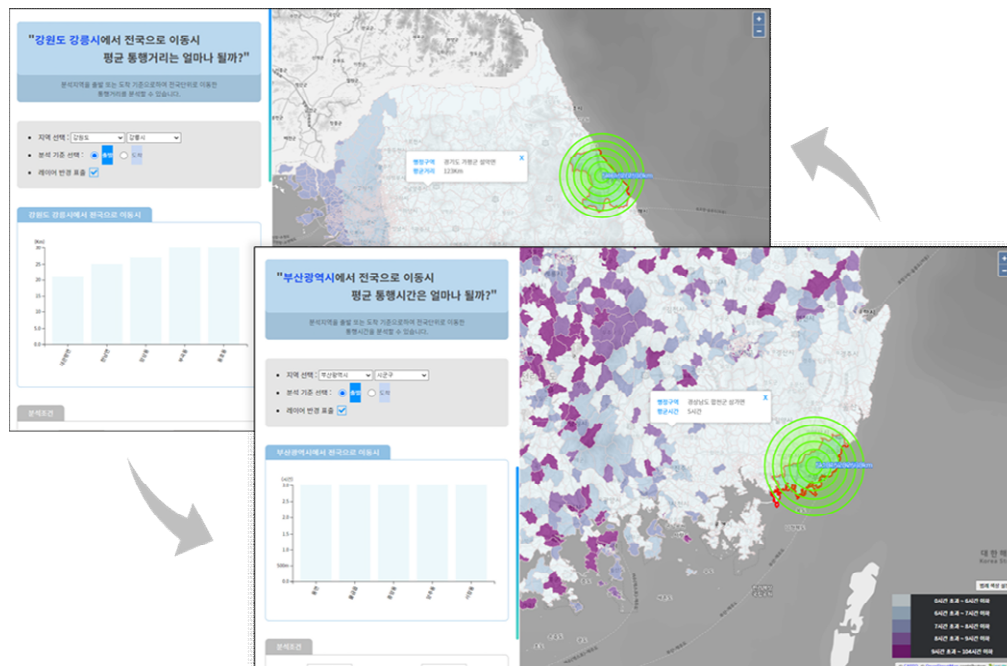


<그림 4- 23> 고령자 주요 통행 구간 분석의 읍면동 단위 분석 결과 표출

### 3) 출·퇴근 통행분석

#### ○ 통행시간/거리 분석

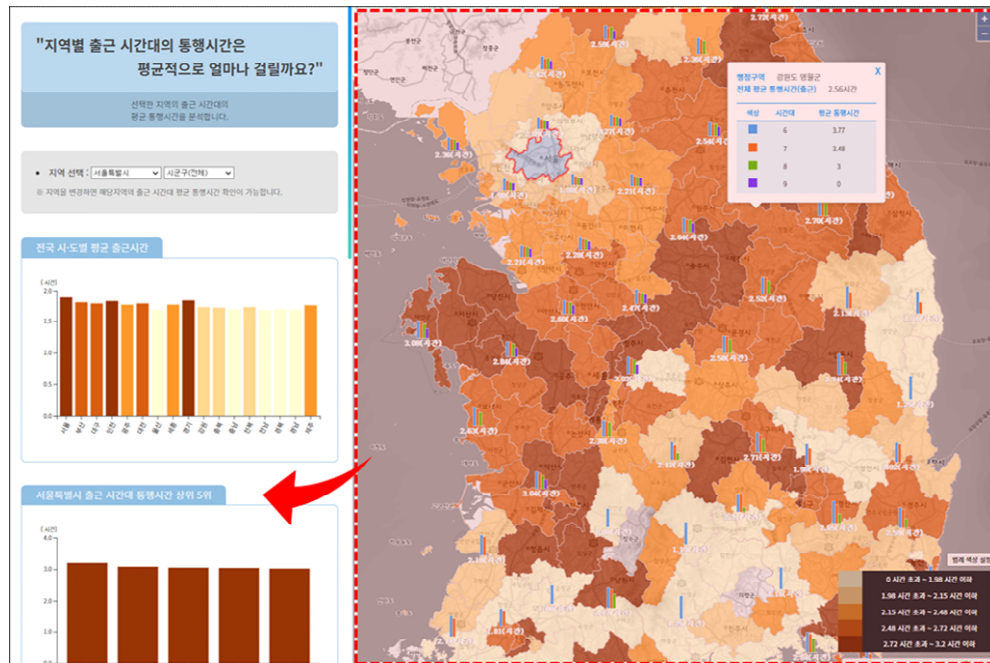
- 선택 지역을 출발/도착으로 하여 각 지역별 평균 통행시간과 거리를 분석
- 화면 좌측 분석 조건 및 결과 영역의 스크롤을 이동하여 평균 통행시간과 평균 통행거리 조건을 전환하고, 지역 선택, 분석 기준 선택 (출발, 도착), 레이어 반경 표시의 조건을 설정 및 분석 결과의 상위 5개 지역을 그래프로 표출
- 화면 우측 주제도에서는 선택 지역을 기준으로 각 지역별 평균 통행시간과 거리를 범례 범위에 따른 색상으로 표시하고 마우스 커서에 위치하는 지역의 상세 정보를 툴팁으로 표출



<그림 4-24> 분석 조건/결과 영역의 스크롤 이동으로 통행 거리/통행시간 조건 변경

○ 출·퇴근 통행시간 분석

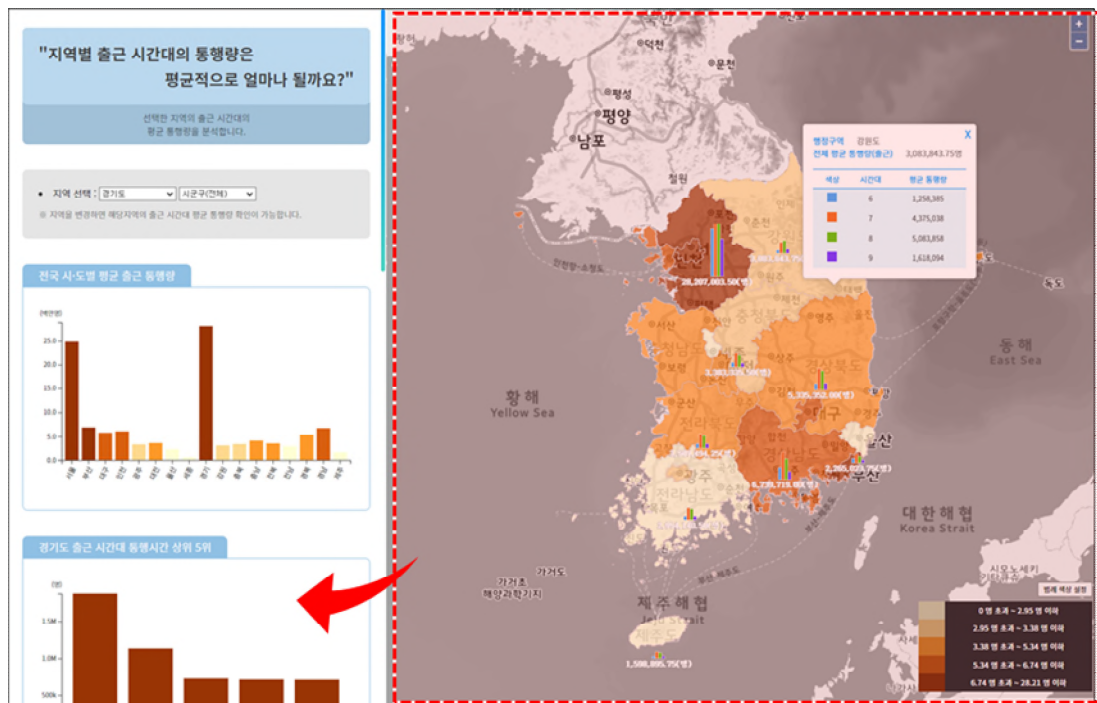
- 시도, 또는 시군구 단위의 지역을 선택하면 해당 지역을 기준으로 출근, 퇴근 시간대의 평균 통행시간을 분석하고, 평균 통행시간의 상위 5개, 하위 5개 지역을 그래프로 표출
- 주제도에서는 각 행정구역별 출근 시간대(6-9시), 퇴근 시간대(18-21시)의 각각의 시간대별 통행시간을 그래프와 톱팁을 통해 상세 표출
- 화면 좌측 분석 조건 및 결과 영역의 스크롤을 이동하여 출근 시간대와 퇴근 시간대의 조건을 각각 달리하여 분석



<그림 4- 25> 출·퇴근 통행시간 분석 도구 화면

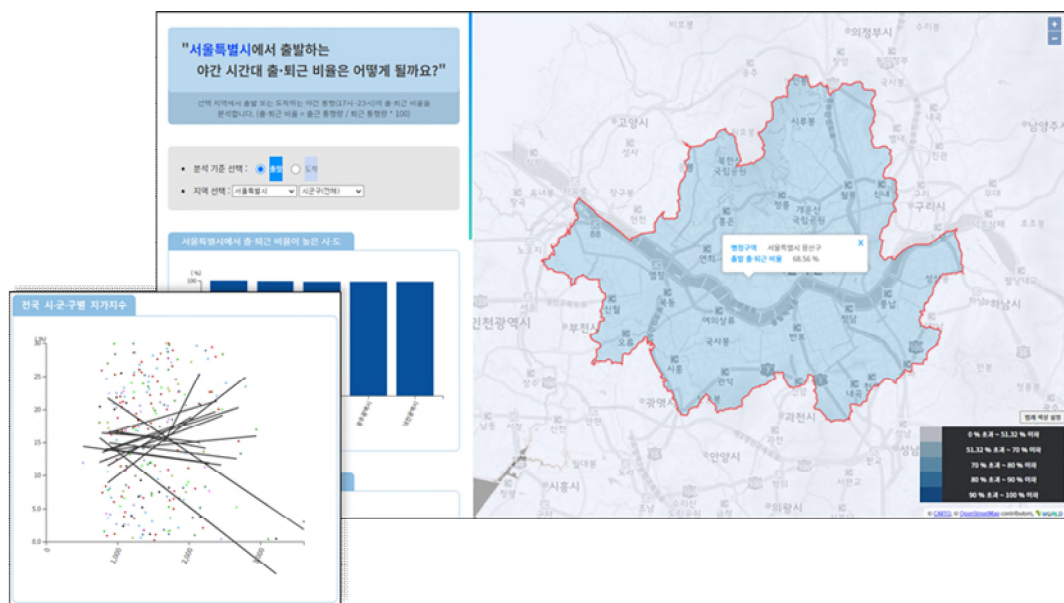
○ 출·퇴근 통행량 분석

- 시도, 또는 시군구 단위의 지역을 선택하면 해당 지역을 기준으로 출근, 퇴근 시간대의 평균 통행량을 분석하고, 평균 통행량의 상위 5개, 하위 5개 지역을 그래프로 표출
- 주제도에서는 각 행정구역별 출근 시간대(6-9시), 퇴근 시간대(18-21시)의 각각의 시간대별 통행량을 그래프와 톨팁을 통해 상세 표출
- 화면 좌측 분석 조건 및 결과 영역의 스크롤을 이동하여 출근 시간대와 퇴근 시간대의 조건을 각각 달리하여 분석



<그림 4- 26> 출·퇴근 통행량 분석 도구 화면

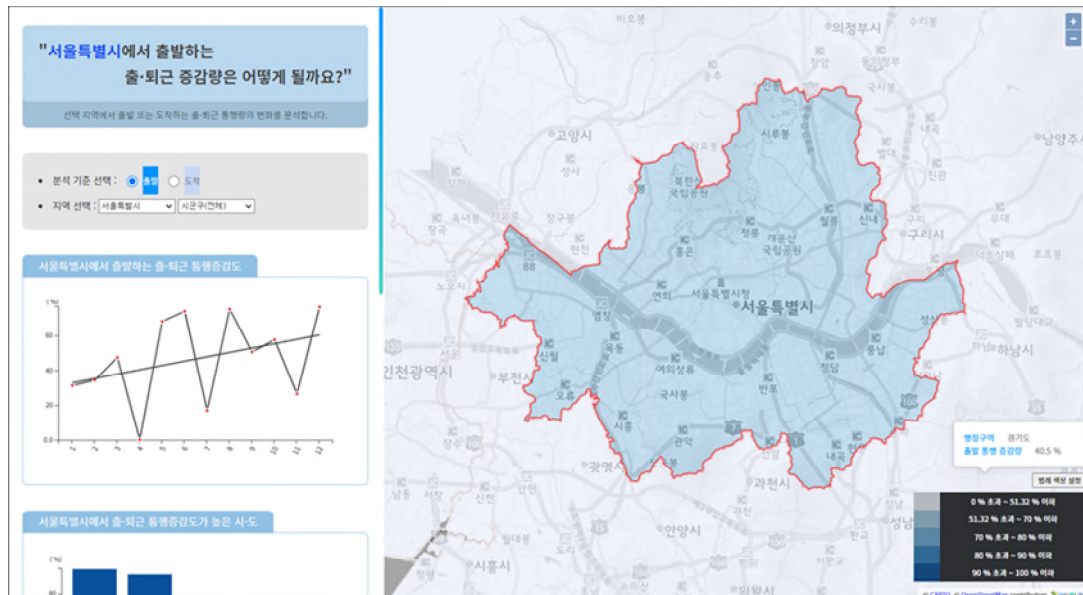
- 근무 형태에 따른 통근 통행 분석
  - 야간 시간대(17-23시)에 선택 지역에서 출발 또는 선택 지역으로 도착하는 통행량에서 출퇴근의 비율을 분석하는 신규 분석 도구
  - 좌측 분석 조건 및 결과 영역에서는 분석 기준을 출발 또는 도착으로 설정하고 기준이 되는 지역을 시도 또는 시군구 단위로 선택
  - 조건을 설정하면 출퇴근 비율을 분석하여 좌측 영역에 출근 비율이 높은 상위 5개 지역과 퇴근 비율이 높은 상위 5개 지역을 그래프로 표출하고, 우측 주제도 영역에서는 각 지역별 범례 범위에 해당하는 색상을 표출
  - 분석 결과 그래프 아래에는 전국 시도, 시군구별 지가지수 관련 산점도 그래프를 참고용으로 표출



<그림 4-27> 근무 형태에 따른 통근 통행 분석화면과 전국 시군구별 지가지수 산점도



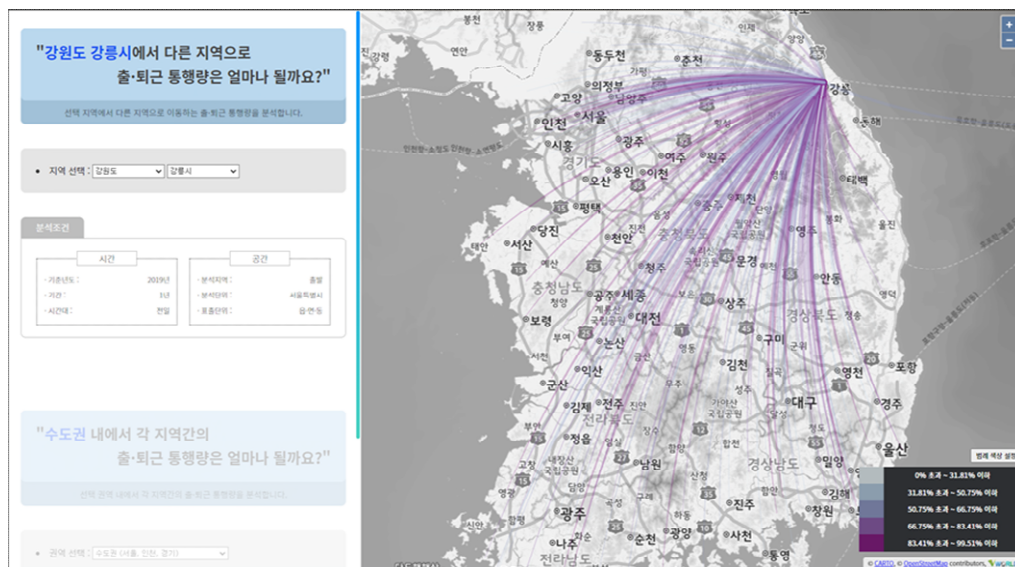
- 고용인구 변화에 따른 통근 통행 분석
  - 출퇴근을 목적으로 이동하는 인구통행량 데이터로 선택 지역에서 출발 또는 선택 지역으로 도착하는 월별 인구통행량의 증감률을 분석하는 신규 분석 도구
  - 좌측 분석 조건 및 결과 영역에서는 분석 기준을 출발 또는 도착으로 설정하고 기준이 되는 지역을 시도 또는 시군구 단위로 선택
  - 조건을 설정하면 선택 지역을 기준으로 출발 도착하는 출퇴근 목적의 통행량을 월별 증감도와 증감도가 높은 5개 지역과 증감도가 낮은 5개 지역의 순위 그래프를 좌측 영역에 표출하고, 우측 주제도 영역에서는 각 지역별 범례 범위에 해당하는 색상을 표출



<그림 4- 28> 고용인구 변화에 따른 통근 통행 분석화면

### ○ 출퇴근 통행 연결성 분석

- 선택 지역을 기준으로 각 지역으로의 출퇴근 통행량을 분석하는 신규분석 도구
- 화면 좌측 분석 조건 영역의 스크롤을 이동하여 선택 지역 기준의 전국 통행 조건과 선택 권역 내 통행 조건을 전환하고, 지역 또는 권역을 선택하는 UI 표출
- 지역 또는 권역을 선택 시 선택된 조건으로 분석된 결과가 주제도에서 지역 연결선으로 표출

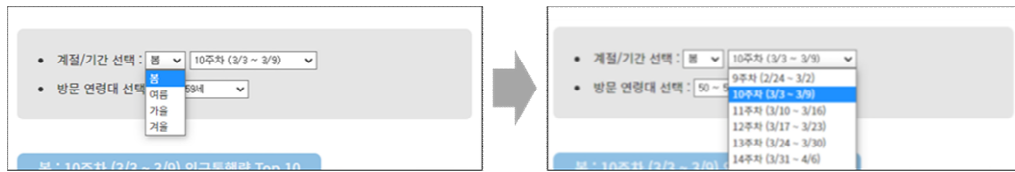


<그림 4- 29> 출·퇴근 통행 연결성 분석화면

## 4) 관광·연휴 분석

### ○ 계절별 핫플레이스 분석

- 전국에서 특정 지역으로 유입되는 이동인구와 특정 지역에서 유출되는 이동인구 중 가장 이동이 많은 지역을 계절별 간단하게 분석
- 분석 결과 그래프에 지역명을 시도단위부터 상세하게 표기하고 전체 이용량을 여성과 남성의 이동량으로 구분하여 표시 추가
- 우측에는 주제도를 표현하고, 좌측에는 우측 주제도의 조건 및 내용을 설명
- 우측 하단에는 범례를 표시
- 주제도상에 마우스 오버 이벤트로 해당 지역의 행정구역명과 유동 인구 수를 표현



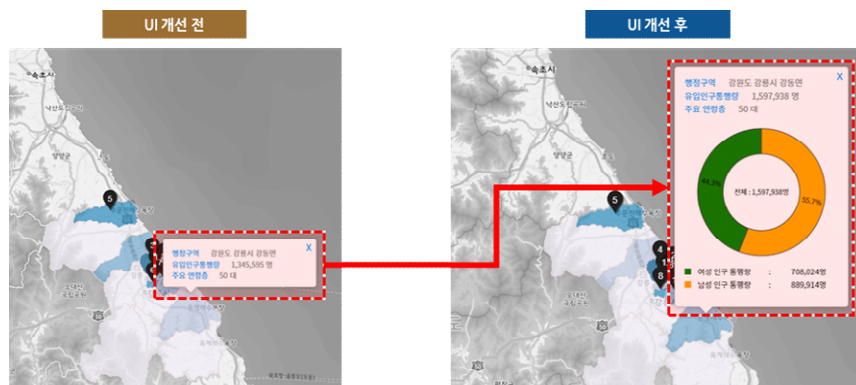
<그림 4- 30> 계절별 핫플레이스의 계절/기간 선택



<그림 4- 31> 계절별 핫플레이스의 분석 결과

○ 관광 특화지역 분석

- 관광 목적의 통행량 변화 큰 지역을 확인할 수 있는 기능
- 데이터 가공 방법 : 도착지 타입이 비정기적 체류지인 통행을 지역별, 계절별로 분류
- 기본 주제도는 강원도 강릉시의 유입량 지표를 색상으로 표출
- 유입량 지표가 높은 상위 10개 지역의 순위를 표시
- 주제도상에 마우스 오버 이벤트로 해당 지역의 유입 인구 수를 표현
- 좌측에는 우측 주제도의 조건과 지역명, 주요 방문 나이대, 계절별 통행량 차트를 표출
- 우측 하단에는 범례를 표시
- 분석 결과를 나타내는 툴팁에 남/녀 성별 통행량을 구분하여 표시하고 이를 알아보기 쉽도록 파이 그래프로 표현

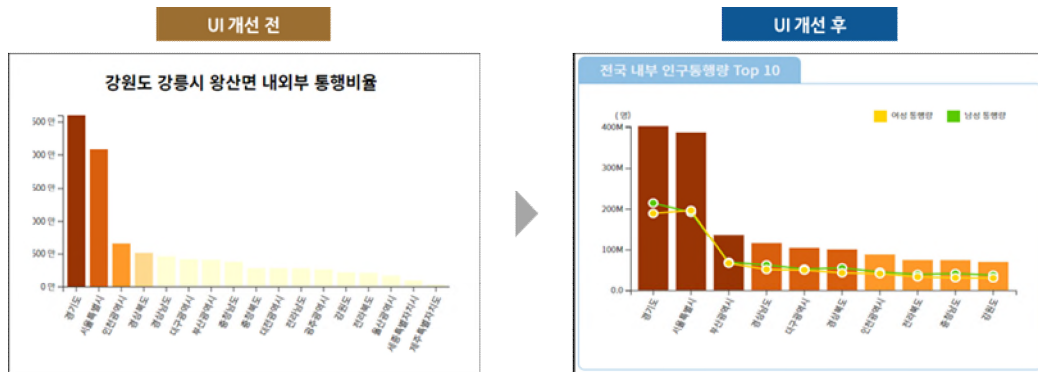


<그림 4- 32> 관광 특화지역 분석 결과의 상세 정보 툴팁 고도화



○ 관광지 내·외부 통행비율 분석

- 분석 결과 그래프에 표시되는 지역을 순위가 높은 1~10까지로 요약하여 표시하고 각 지역별 인구통행량의 남/녀 성별을 구분하여 결과 표출

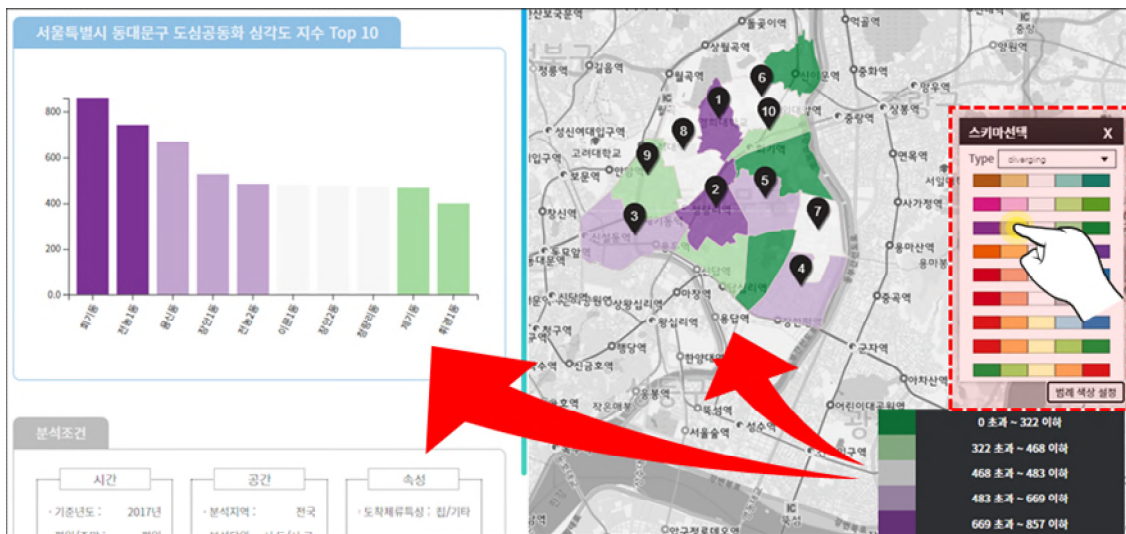


<그림 4-33> 관광지 내·외부 통행 비율 분석 결과의 UI 개선 전/후 비교

5) 경제·사회활동 분석

○ 도심공동화 심각도 분석

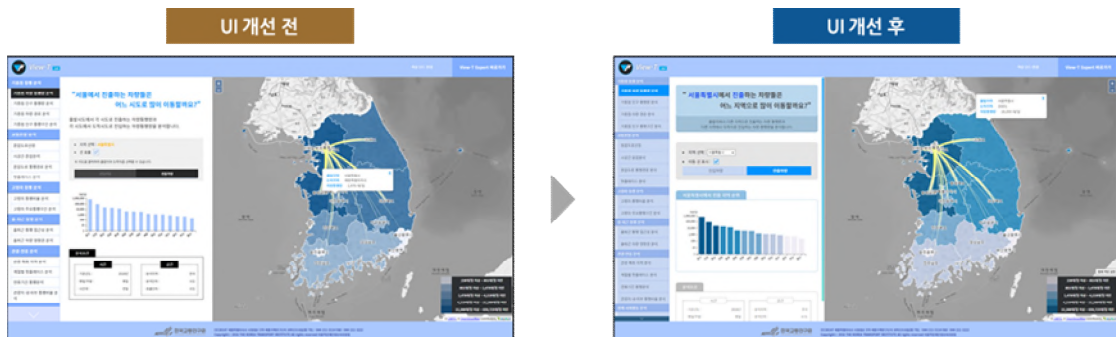
- 도심공동화가 나타나는 지역을 분석하기 위한 기능
- 데이터 가공 방법
  - . ④ 지역별 심야 시간대(도착시간 기준: 22~06시) 도착지 트립타입이 집(H)인 인구 추출
  - . ⑥ 지역별 낮시간대(도착시간 기준: 10~18시) 도착지 트립타입 전체(또는 잠재 체류지)인 인구 추출
  - . 도심공동화 지표 산출 ((⑥ / ④) \* 100)
- 기본 주제도는 서울시의 도심공동화 지표를 색상으로 표출
- 주제도 상에 마우스 오버 이벤트로 해당 지역의 도심공동화 심각도 지수를 표현
- 좌측에는 도심공동화 심각도에 대한 차트를 표출
- 우측 하단에는 범례를 표시
- 분석 결과가 색상 단계로 구분되는 표출 방식에서 결과의 관계 파악 및 확실한 의미 전달을 할 수 있도록 여러 가지 색상을 조합한 색상 스키마를 사용하여 표현 방식을 변경할 수 있는 기능 추가



<그림 4- 34> 도심공동화 심각도 분석 결과의 UI

#### 다. View-T Light 시스템 개선사항

- 메뉴와 해설, 분석 조건과 결과를 보다 명확하게 구분할 수 있도록 화면 디자인 개선



<그림 4- 35> 간편분석 도구 화면 디자인 개선 전/후 비교

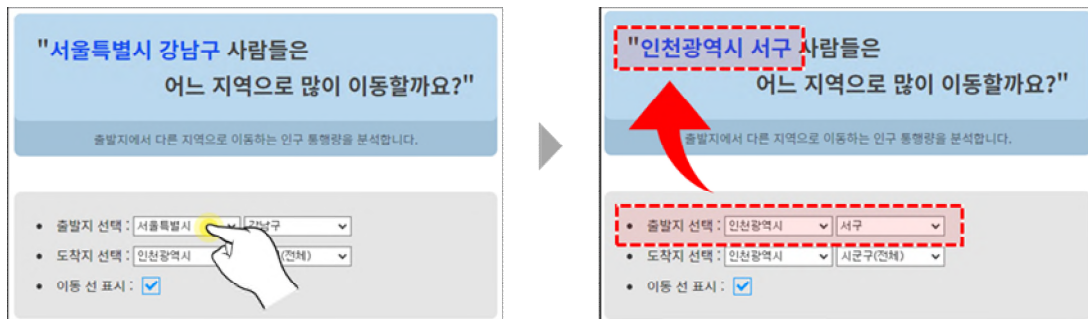
- 메뉴 부분의 크기와 형태를 조절하여 분석 결과 영역과 구분되도록 처리
- 분석 주제와 설명 문구를 이해하기 쉽도록 보완
- 분석 결과 영역의 각 구성 요소들의 (분석 주제, 해설, 조건, 결과 등) 구분이 명확해지도록 디자인 개선
- 분석 결과 그래프에 타이틀을 붙여서 분석의 목적부터 결론까지 쉽게 이해하도록 유도
- 텍스트 및 UI 요소들의 크기와 색상을 성격에 맞게 정리
- 기존에는 주제도에서 행정구역을 직접 클릭하여 선택하였으나 분석 결과와 각종 상세

정보가 중첩되어 표시되어 있고, 정확한 위치를 알고 있지 않으면 원하는 지역을 선택하기 어려운 문제가 있어서, 이를 해결하기 위하여 분석 결과 영역에 콤보박스를 통해 보다 간편하게 지역을 선택할 수 있도록 기능 개선



<그림 4- 36> 지역 선택방식 변경

- 기존의 분석 지역을 변경하여도 고정된 분석 주제 문구로 분석 도구의 이해도가 떨어지는 문제를 해결하기 위해 분석 지역 변경 시 변경된 지역명이 분석 주제에 동적으로 반영되도록 기능 개선



<그림 4- 37> 선택한 분석 지역이 분석 주제에 동적으로 반영

### 3. View-T Expert

#### 가. 사용자 편의성을 위한 UI/UX 개선

- 심층적인 분석보다 간단한 분석을 선호하는 사용자를 위해 조건 설정, 범례 설정을 간소화함
- “키워드 분석” 옵션을 조건 설정 부분에 추가하여, 사용자가 상세하게 조건을 입력하지 않아도 분석을 수행할 수 있도록 함
- 사용자가 자주 사용하는 분석 조건을 기본 옵션으로 두어, 사용자의 조건 선택을 최소화



<그림 4- 38> 분석 조건 값 UI 변경

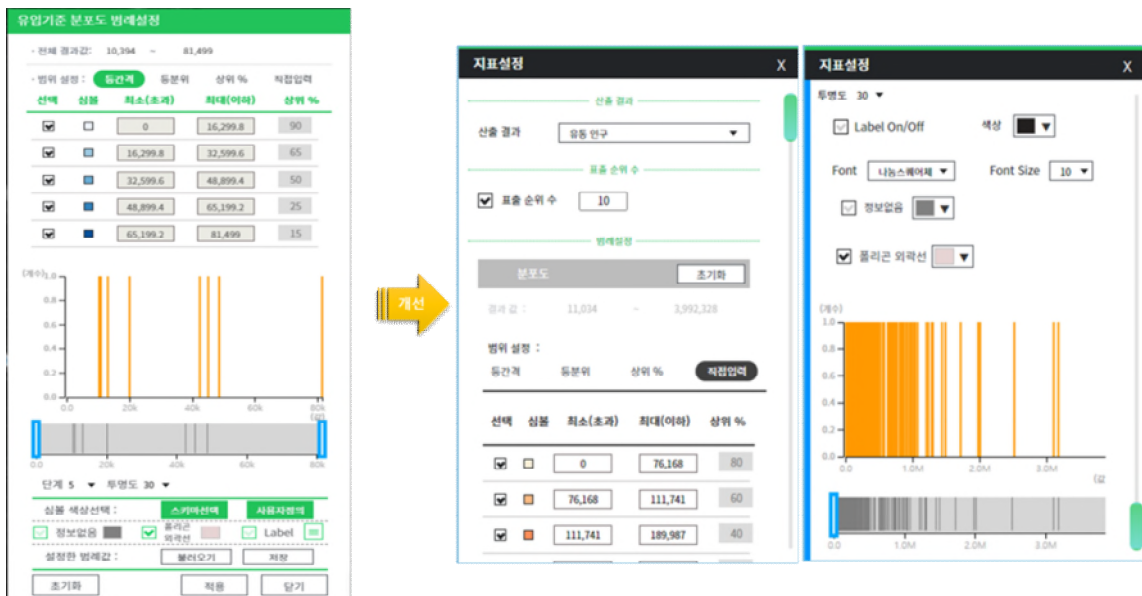
<표 4- 16> 주제별 키워드 사전 조건

통행목적	출발 조건	도착 조건	기타 조건	비고
	트립타입	트립타입		
출근	집 (H)	회사(C)	20~60대	필수
등교	집 (H)	학교 (S)	10~20대	필수
퇴근	회사(C)	집 (H)	20~60대	필수
하교	학교 (S)	집 (H)	10~20대	필수
귀가	정기적 잠재체류지 (R), 비정기적 잠재체류지 (X)	집 (H)	-	필수
여가		정기적 잠재체류지 (R),	-	필수

	회사(C),	비정기적잠재체류지(X)		
	학교(S),			
	정기적 잠재체류지(R),			
	비정기적 잠재체류지(X)			
사적업무 후 복귀	정기적 잠재체류지(R),	회사(C), 학교(S)	10~60대	보류
	비정기적 잠재체류지(X)			

\* ① 도착지 타입이 '-'인 경우, ② 회사(C)→학교(S) 또는 학교(S)→회사(C) 타입인 경우, ③출발일자와 도착일자가 다른 경우는 제외

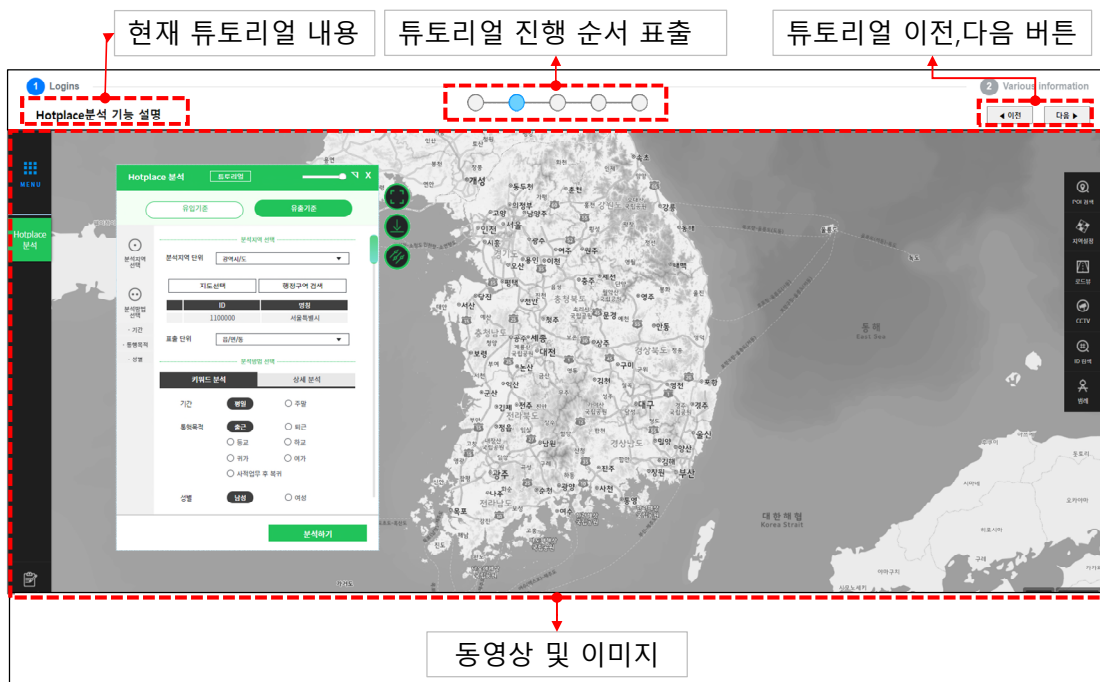
- 필수 입력 조건이 아닌 범례 및 표출 정의를 별도의 아이콘으로 배치하여 레이아웃의 여유 공간을 확보하고, 요소 전체 UI를 개선해 사용자가 정보를 식별하는데 수월할 수 있도록 개선



<그림 4- 39> 범례 설정 UI 변경

## 나. 분석 기능 튜토리얼 제작

- 통신 주요 분석기능의 기능 및 분석 결과 설명, 분석 도구의 사용 방법을 제공
  - 분석 도구 사용이 익숙하지 않은 사용자에게 동영상 또는 그림을 통하여 튜토리얼을 제공함으로써 사용 편의를 높이고, 조건 설정 등 분석 과정 및 분석 결과에 대한 사용자의 이해도를 높일 수 있도록 함
  - 이전, 다음 버튼을 통하여 사용자가 원하는 튜토리얼을 진행할 수 있도록 구현
  - 상단 중앙에 튜토리얼 진행 순서를 표시하는 상태 바 표시
  - 좌측 상단에 현재 진행 중인 튜토리얼의 기능 명 표시
  - 중앙에 해당 분석기능의 사용 방법의 이해를 도울 수 있는 동영상 혹은 이미지를 표시



<그림 4- 40> 분석기능 튜토리얼 UI

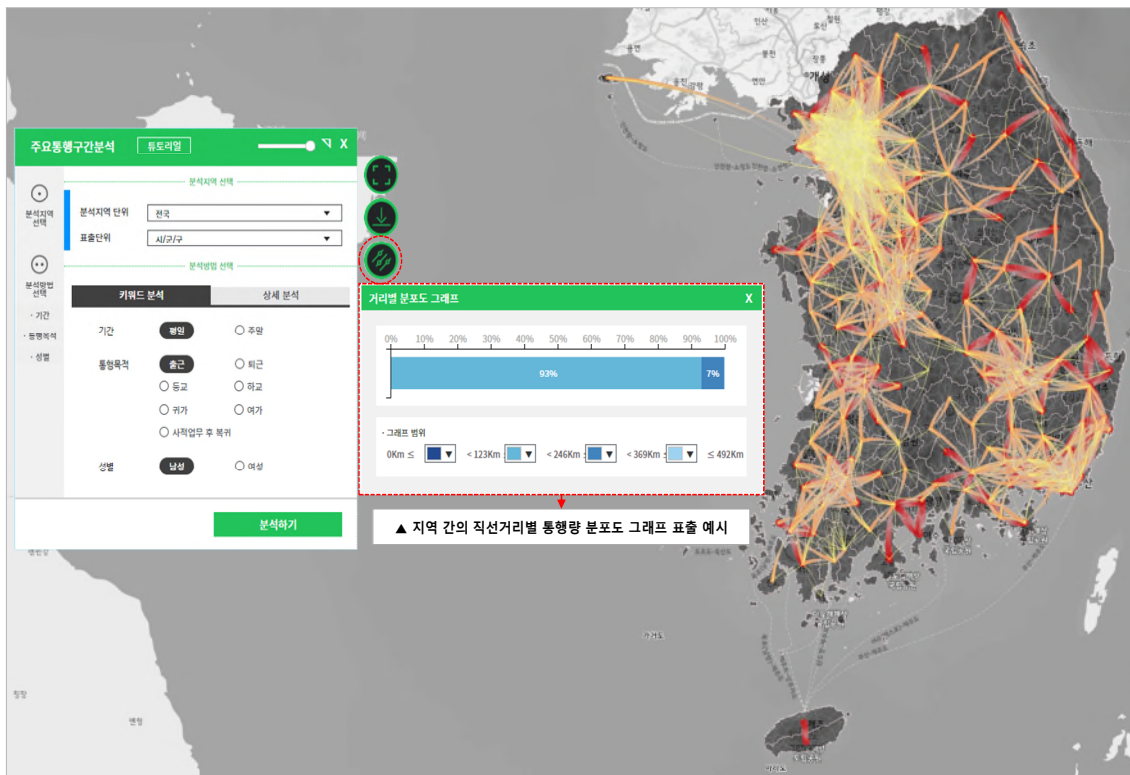


#### 다. 기존 분석 기능 (주요 통행 구간 분석) 개선

- 단거리 구간 통행과 장거리 구간 통행의 분석이 가능하도록 개선
  - 전국의 시도, 시군구 단위 간 구간 분석이 가능하도록 개선
  - 단거리 구간 통행과 장거리 구간 통행 비교를 위한 그래프 표출

<표 4- 17> 주요 통행 구간 분석기능의 연도별 표출 형태 변경

구분	권한	변경 전	변경 후
분석지역 선택	일반사용자	최대 5개 (시군구, 읍면동)	변화 없음
	고급사용자	일반사용자 기능에 더하여 시도 단위는 1개 선택가능	전국 단위 분석 가능 (상세 조건 선택 제한)
표출단위 선택	일반사용자	분석지역 하위 단위 선택 (시군구 → 읍면동)	최소 표출단위 교통폴리곤 → 읍면동
	고급사용자		전국 단위 경우 시도, 시군구 단위선택



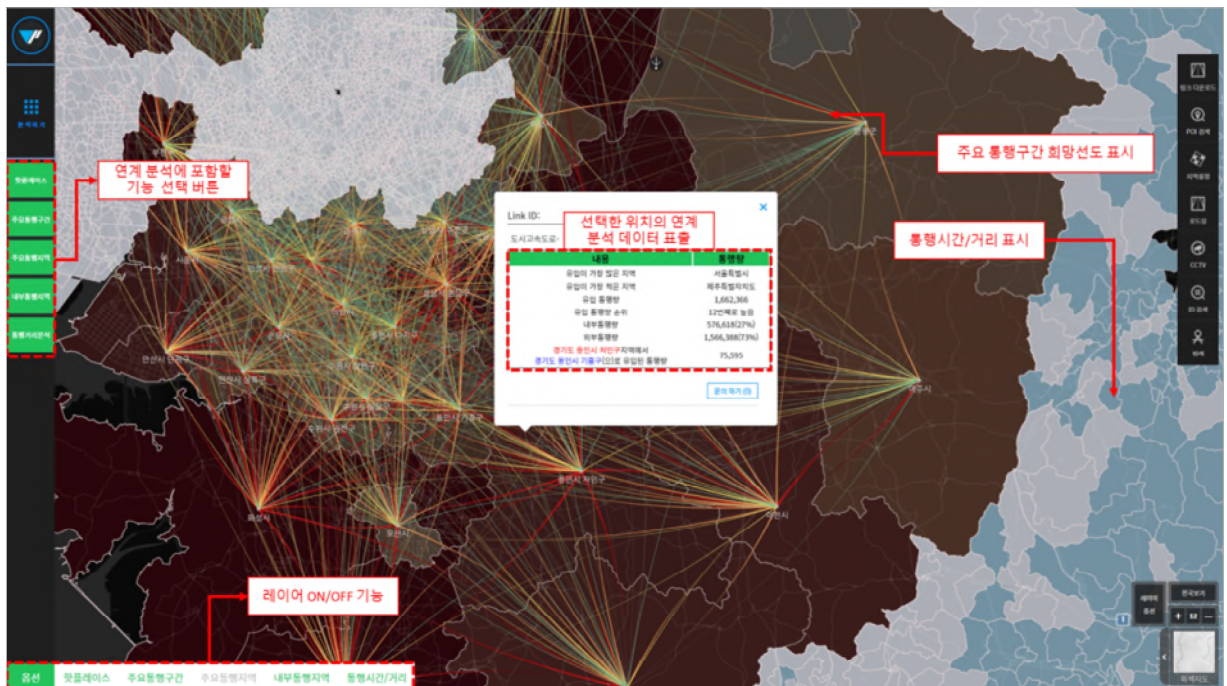
<그림 4- 41> 주요 통행 구간 분석의 전국 분석 결과 및 통행량 분포도 그래프

## 라. 신규분석 기능 개발

### 1) 심층분석을 가능케 하는 분석 기능 개발

#### ① 정보 연계분석 기능

- 여러 분석 기능을 선택하고 그에 따른 결과를 동시에 주제도로 표출하여 확인할 수 있는 기능
  - 핫플레이스, 주요 통행 구간, 주요 통행지역, 내부통행지역, 통행 거리 분석을 연계분석 가능함
- 정보 연계분석 기능 UI
  - 각 분석기능을 선택한 후 분석하기 버튼을 클릭하여 연계분석을 실행함
  - 주제도 상 클릭하여 연계 분석된 데이터를 표출함
  - 하단 옵션 버튼을 통하여 연계분석 기능들의 레이어를 끄거나 켜서 개별 주제도를 볼 수 있음
  - 희망 선도를 통해서 주요 통행 구간을 인지할 수 있음
  - 희망 선도 혹은 지역을 선택하여 해당 위치의 데이터 개괄을 볼 수 있음



<그림 4-42> 정보 연계분석 기능 화면

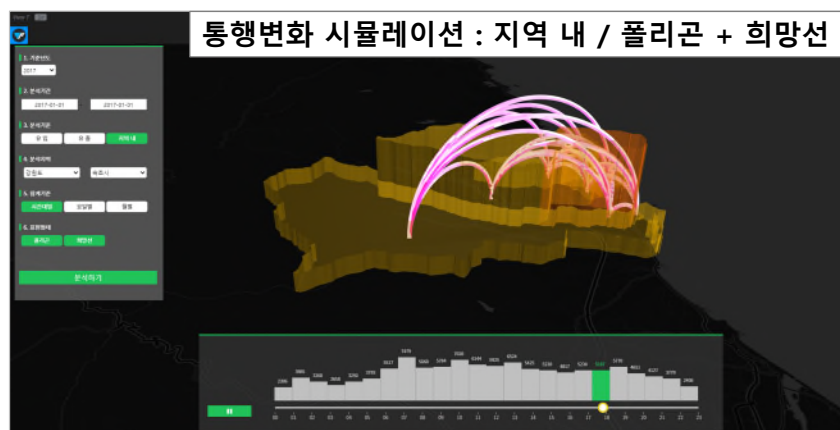


## ② 통행 변화 시뮬레이션 기능

- 통행 변화를 시계열로 확인할 수 있는 기능
- 통행 변화 시뮬레이션 분석 기능 UI
  - 분석유형(유입, 유출 및 O/D기준), 분석 지역 및 분석 기간을 설정해 사용자 의도에 맞는 공간 데이터를 시각화하여 통행 변화의 흐름 분석이 가능하도록 개발
  - 시뮬레이션은 시간대별, 요일별, 월별을 1년치 데이터를 기준으로 집계하여 통행의 흐름을 보여줄 수 있도록 개발
  - 예시) 시간대별 - 00시~24시, 요일별 - 월요일~일요일, 월별 - 1월~12월
  - 시뮬레이터를 통해 시간에 따른 데이터 변화를 제어할 수 있음
  - 재생, 정지, 재생트랙 제어 등



<그림 4- 43> 유입 기준 폴리곤 형태의 통행변화 시뮬레이션

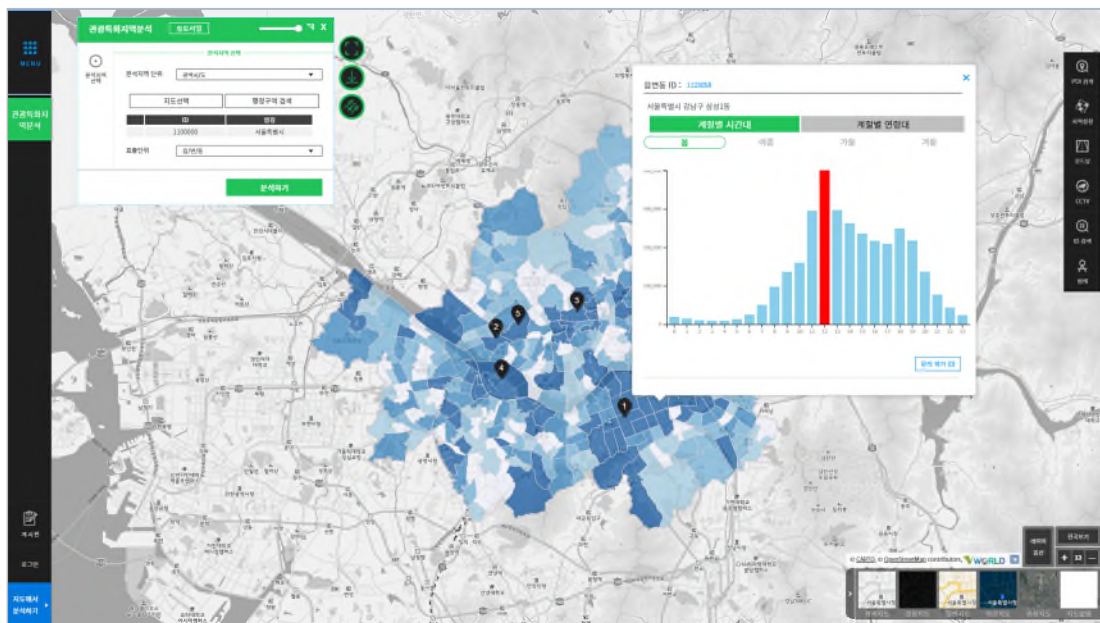


<그림 4- 44> 지역 내 기준 폴리곤 및 희망선 형태의 통행변화 시뮬레이션

## 2) View-T Light의 분석 서비스 중 일부를 Expert에 개발

## ① 관광 특화지역 분석 기능

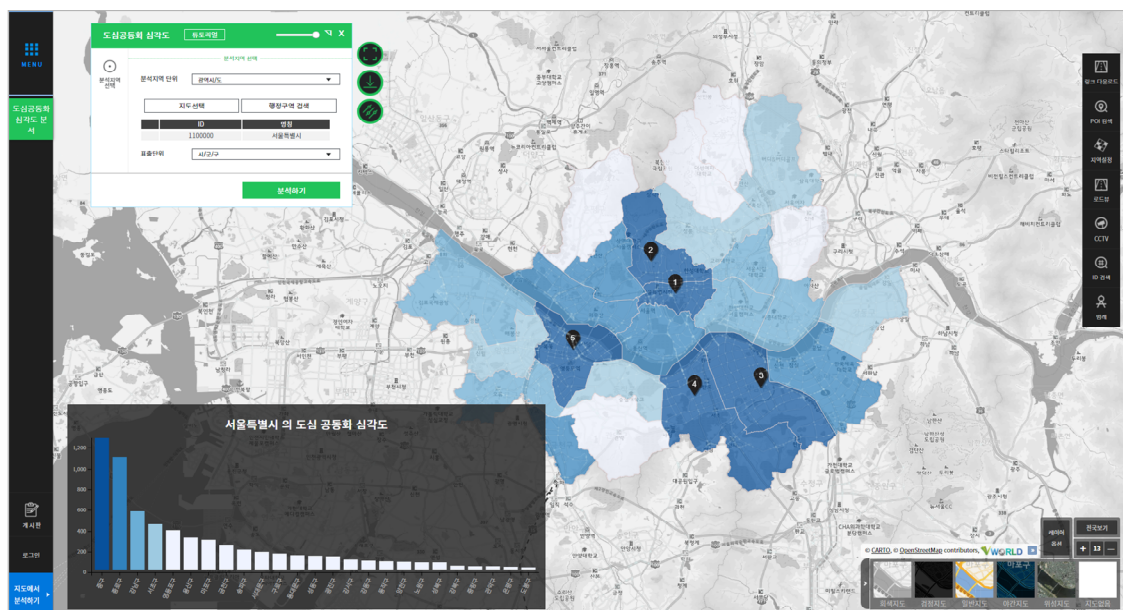
- 계절에 따라 통행량 변화가 큰 지역을 확인 할 수 있는 기능
  - 도착지 트립 타입이 비정기적 잠재체류지인 통행을 지역별, 계절별로 분류 표시
- 관광 특화지역 분석 기능 UI
  - 분석 지역 단위를 광역시/도, 시/군/구에서 선택
  - 분석하길 원하는 지역을 지도선택 혹은 행정구역 검색을 통하여 선택
  - 표출 단위는 분석 지역 단위의 하위 지역 중 선택
  - 상위 5개 지역을 주제도 상에 표시
  - 해당 지역의 계절별 시간대와 계절별 연령대 통행량의 변화를 순위 그래프로 표시



&lt;그림 4- 45&gt; 관광 특화지역 분석 기능 화면

## ② 도심공동화 심각도 분석 기능

- 도시 중심부의 상주인구가 줄어드는 현상을 확인할 수 있는 기능
  - 심야 시간대 집으로의 통행량과 낮 시간대 통행량의 비율을 통한 상주인구 추정
- 도심공동화 심각도 분석 기능 UI
  - 분석 지역 단위를 광역시/도, 시/군/구에서 선택
  - 분석하길 원하는 지역을 지도선택 혹은 행정구역 검색을 통하여 선택
  - 표출 단위는 분석 지역 단위의 하위 지역으로 고정됨
  - 상위 5개 지역을 주제도 상에 표시
  - 해당 지역의 공동화 심각도를 순위 그래프로 표시



<그림 4- 46> 도심공동화 심각도 분석화면

## 제5장 결론 및 차년도 수행계획

---

### 제1절 결론

### 제2절 차년도 수행계획



## 제5장 결론 및 차년도 수행계획

### 제1절 결론

- 본 과업에서는 기존 방법보다 정확하게 통행시간 및 통행량 등 통행지표를 산출하기 위하여 '20에 개발한 모바일통신 원천데이터 가공알고리즘을 보완함
  - 알고리즘은 합리적 수준의 시·공간적 데이터 군집 (Data aggregation and smoothing) 과정을 제시하여 신호 이상 현상을 효과적으로 처리하고, 개별통행 특성 및 패턴을 고려한 주요 통행목적지 추정을 통해 야간 및 비정기 근무 통행을 구분함
  - 또한, 기존 통신폴리곤 기준의 위치 추정 정보를 GPS와 같은 점 단위의 위치 추정 정보로 변환하는 방법을 제안하여 모바일 기지국 빅데이터의 공간적 해상도를 개선하였으며, 향후 이동 경로 및 수단 추정을 위한 개발 알고리즘의 적용방안을 검토함
- 2020년 1월 1일부터 12월 31일까지 생성된 로그 기록에 모바일통신 빅데이터 가공알고리즘을 적용하여 기존 DB를 구축하고 검증함
  - 각 개인별 로그 기록은 「개인정보보호」, 「위치정보의 보호 및 이용 등에 관한 법률」에 저촉되지 않도록 집계한 형태의 DB를 설계하였으며, 이를 기준으로 DB를 구축함
  - 기지국 기반의 모바일통신 빅데이터 기반으로 구축된 통행DB의 시간대별, 요일별, 성별, 연령별 통행량 및 통행시간을 기존 국가교통DB사업의 '가구통행실태조사자료 (2016년)'와 통계청의 '인구총조사자료 (2015년)' 비교하여 검증함
- 모바일통신 빅데이터 기반 교통지표를 개발하고, 분석 서비스의 성능 및 사용성을 개선하기 위하여 DB를 재구조화하여 서비스를 제공함
  - 또한, 2개 이상의 분석 수행 결과를 주제도로 동시에 확인할 수 있는 '정보 연계 분석 기능' 개발하고, 통행 변화를 시계열로 확인할 수 있는 '통행 변화 시뮬레이션 분석기능' 개발함

## 제2절 차년도 수행계획

- 모바일통신 빅데이터 가공알고리즘 고도화
  - 전처리 알고리즘 개선 및 고도화
    - 지역특성(도시부/비도시부), 인적특성(연령, 휴대폰 이용률 등)에 따른 로그 특성 분석 및 기 전처리 알고리즘 보완
  - 통행목적 구분 알고리즘 개선 및 고도화
    - 모바일통신 빅데이터 기반 추정 가능 체류지(거주지, 근무지, 여가 등) 선정
    - 거주지 및 근무지 추정 로직 개선
- 모바일통신 빅데이터 DB 구축 및 검증
  - 모바일통신 빅데이터 DB 구축 (2021년도 전국)
  - 기 구축된 DB (2020년도 전국) 검증
    - 가구통행실태조사, 인구총조사 등 기존 조사기반 통계를 이용한 검증
- 모바일통신 빅데이터 기반 교통지표 현행화 및 서비스 제공
  - 교통지표 현행화
    - 1차 지표 : 통근·통학 통행량·통행시간, 기타통행 통행량 (1시간, 격자 단위의 전국 대상 1년 DB)
    - 통행지표 : 연령별·성별 주요 통행 구간, 통행 연결성, 핫플레이스, 관광/고령자/계절별 통행 특성 등
    - 생활·경제지표 : 고용인구 변화율, 야간근무 비율, 연령별 근무 형태, 도심공동화 심각도, 경제활동 의존도 및 수용도 등