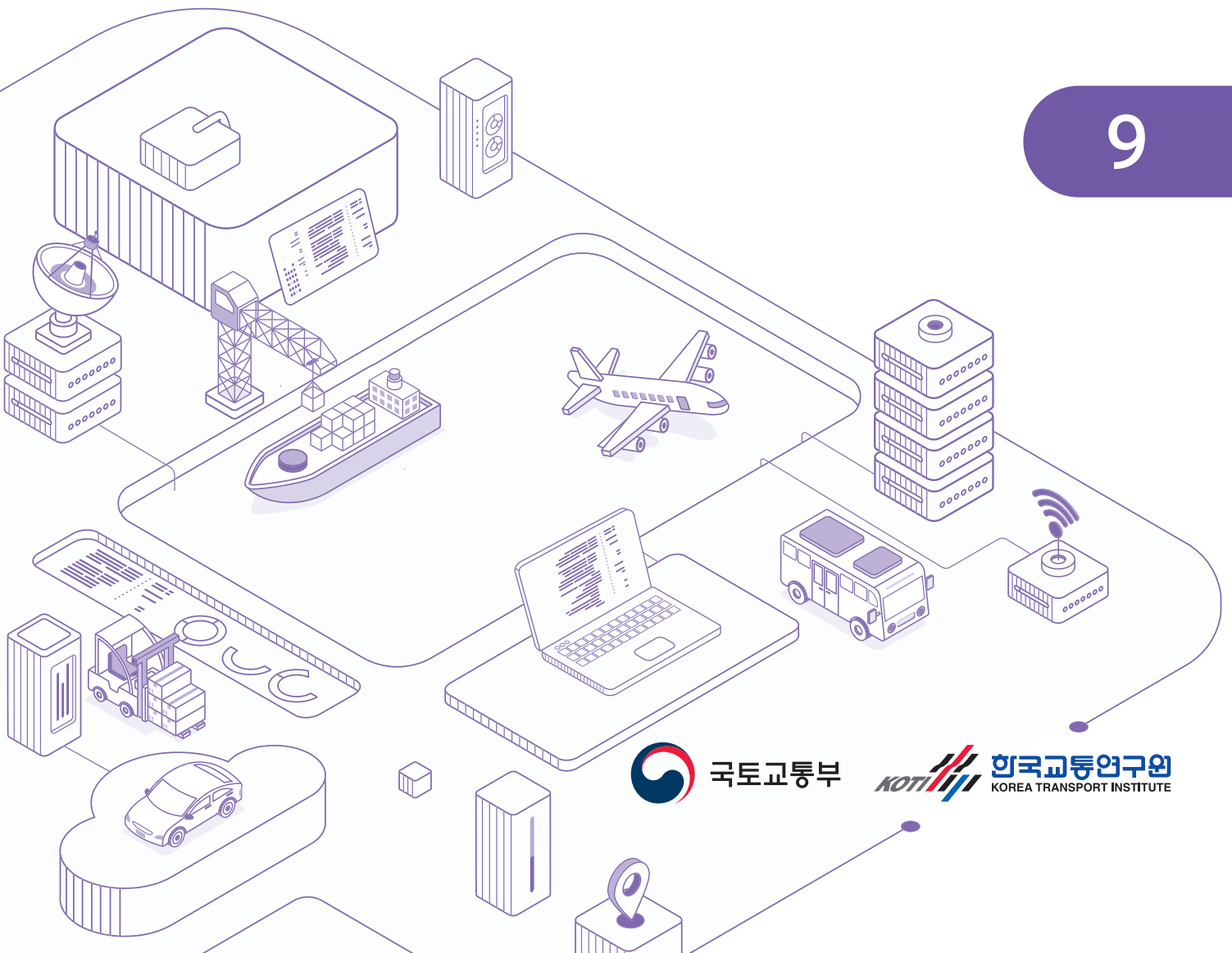


2025년 국가교통조사 및 분석 차량 GPS 빅데이터 구축

2025. 12.

9



제 출 문

국토교통부 장관 귀하

본 보고서를 “2025년 국가교통조사 및 분석”의
최종보고서로 제출합니다.

2025년 12월

한국교통연구원

원장 김 영 찬

**본 『2025년 국가교통조사 및 분석』은 다음 연구진에 의해
수행되었습니다.**

참 여 연 구 진

<한국교통연구원>	
연구책임자	<ul style="list-style-type: none"> ◦ 조종석 연구위원
연 구 진	<ul style="list-style-type: none"> ◦ 김주영, 천승훈 선임연구위원 ◦ 박용일, 원민수, 장동익 연구위원 ◦ 이송봉, 이종우, 정승환, 황순연 부연구위원 ◦ 김규진, 김동호. 신영권 책임전문원 ◦ 가보연, 이선아, 홍성표 주임연구원 ◦ 강국수, 곽명신, 권기훈, 권순옥, 김수아, 박미란, 박준호, 심지윤, 오연선, 이동엽, 이새봄, 이채영, 이해선, 정정호, 조영찬 연구원 ◦ 김다영 연구조원
<한국해양수산개발원>	
연 구 진	<ul style="list-style-type: none"> ◦ 이호춘 부연구위원 ◦ 소애립, 정희진, 허성례 전문연구원 ◦ 김기진 연구원
<한국항공협회>	
연 구 진	<ul style="list-style-type: none"> ◦ 최인영 실장 ◦ 손병열, 백준군 차장 ◦ 김진성, 김지한 과장 ◦ 김창욱 대리 ◦ 박다영 연구원

『2025년 국가교통조사 및 분석』

보고서 구성 및 담당연구진

번호	과제명	연구진
제 1권	요약보고서	조종석, 신영권, 가보연
제 2권	전국 여객 기종점통행량 예비조사	김동호, 이선아, 조영찬
제 3권	전국 여객 기종점통행량 보완갱신	김동호, 강국수, 권순욱, 박미란
제 4권	교통분석용 네트워크 구축	황순연, 이동엽, 이새봄
제 5권	항공 여객 기종점통행량 조사	한국항공협회
제 6권	전국화물 기종점통행량 보완갱신	원민수, 오연선, 정다빈
제 7권	연안화물 기종점통행량 구축	한국해양수산개발원
제 8권	KTDB 모빌리티 기반지도 구축	천승훈, 정정호
제 9권	차량 GPS 빅데이터 구축	이승봉, 이채영
제10권	모바일통신 빅데이터 구축	장동익, 이종우, 곽명신, 심지윤
제11권	국가교통통계DB구축	박용일, 권기훈, 김수아
제12권	특별교통대책기간 통행실태조사	김주영, 권순욱
제13권	교통접근성지표 구축	홍성표, 박준호

『2024년 국가교통조사 및 분석』

과제별 공동참여·위탁용역 사업자

【공동사업 참여기관】

- 전국 여객O/D 현행화 공동사업(수도권 부문)
 - 경기연구원, 인천연구원, 서울연구원
- 항공O/D 및 특성 조사
 - (사)한국항공협회

【위탁용역 사업자】

- 전국 여객 기종점통행량 예비조사
 - ㈜컨슈머인사이트
- 전국 여객 O/D 현행화 공동사업(대구광역시권)
 - 홍익대학교 산학협력단
- 전국 여객 O/D 현행화 공동사업(대전세종충청권)
 - ㈜신명이앤씨
- 전국 여객 O/D 현행화 공동사업(제주권)
 - ㈜모비크리에티브
- 통행수요 구축 및 검증 프로그램 개발
 - 계명대학교 산학협력단
- 교통부문 네트워크 갱신을 위한 GIS기반 교통망 기초자료 구축
 - 서울시립대학교 산학협력단

【위탁용역 사업자】

- 화물 기종점 통행량 보완·강신을 위한 교통물류 실증 데이터 수집·가공·전처리 용역
- ㈜노트스퀘어
- 모빌리티 빅데이터를 활용한 KTDB 기반지도 및 차량 통행DB 구축
- ㈜큐빅웨어
- 모바일 부문 ViewT 서비스 현행화 및 경로형 데이터 가공
- ㈜큐빅웨어, ㈜엘지유플러스
- 특별교통대책기간 통행실태조사
- ㈜컨슈머인사이트
- GTFS 기반 대중교통 네트워크 구축
- ㈜슈퍼무브

최종보고서 목차

- 제 1권 요약보고서
- 제 2권 전국 여객 기종점통행량 예비조사
- 제 3권 전국 여객 기종점통행량 보완갱신
- 제 4권 교통분석용 네트워크 구축
- 제 5권 항공 여객 기종점통행량 조사
- 제 6권 전국화물 기종점통행량 보완갱신
- 제 7권 연안화물 기종점통행량 구축
- 제 8권 KTDB 모빌리티 기반지도 구축
- 제 9권 차량 GPS 빅데이터 구축
- 제 10권 모바일통신 빅데이터 구축
- 제 11권 국가교통통계DB구축
- 제 12권 특별교통대책기간 통행실태조사
- 제 13권 교통접근성지표 구축

• 목차

요약

제1장 과업의 개요	1
제1절 과업의 배경 및 목적	3
1. 과업의 배경	3
2. 과업의 목적	4
제2절 과업의 범위 및 내용	5
1. 과업의 범위	5
2. 과업의 내용	5
제2장 차량 GPS 빅데이터 수집 및 분석	7
제1절 차량 GPS 빅데이터 수집 개요	9
1. 차량 GPS 빅데이터 수집	10
제2절 차량 GPS 원시데이터 분석	19
제3장 차량 GPS 데이터 가공 및 DB 구축	25
제1절 차량 DB 가공 및 구축 프로세스	27
제2절 차량 GPS 데이터 전처리 및 경로 가공	29
1. 링크 단위 티맵 내비게이션 데이터 전처리 및 경로 가공	29
2. 포인트 단위 DTG 데이터 전처리 및 경로 가공	34
3. 경로 데이터 구축 및 경로 압축(통합)	40
제3절 교통량 추정 DB 구축	45
제4절 기초교통 DB 구축	53
1. 기초교통 DB 구축 개요	53
2. 속도빈도분포 DB 구축	54
3. 기종점 통행 DB 구축	58
제5절 차량 교통망 성능평가지표 구축	64
1. 차량 교통망 성능평가지표 구축 개요	64
2. 교통환경지표 산출을 위한 입력데이터 현행화	66
3. 주요 교통망 성능평가지표 분석 결과	72
4. 차량 교통망 성능평가지표 활용	78

제6절 프로토타입 회전교통량 추정 DB 구축	79
1. 회전교통량 추정 DB 구축 개요	79
2. 회전교통량 추정 DB를 위한 교통 네트워크 구축	79
제4장 차량 GPS 데이터 구축·분석 시스템개선	103
제1절 차량 GPS 데이터 구축·분석 시스템 개요	105
1. 차량 데이터 가공 이력 항목	105
제2절 차량 GPS 데이터 가공 모니터링 기능 개발	107
제3절 데이터 분석을 위한 신규 기능 개발	112
제5장 결론 및 향후 계획	115
1. 사업결과 요약	117
2. 사업결과의 활용방향	117
3. 향후 계획	120

• 표목차

〈표 2-1〉 차량 GPS 데이터 특징	10
〈표 2-2〉 데이터 수집 목록 및 기준연도	10
〈표 2-3〉 수집 데이터의 시간적 범위	11
〈표 2-4〉 링크 기반 내비게이션 데이터 형식	13
〈표 2-5〉 통행 구성을 위한 경로 시각화 예시	14
〈표 2-6〉 포인트 기반 내비게이션 데이터의 주행경로 데이터 형식	15
〈표 2-7〉 포인트 기반 내비게이션 데이터의 사용자 정보 데이터 형식	15
〈표 2-8〉 포인트 기반 내비게이션 데이터의 주행 고유ID와 주행 탐색ID 관계 데이터 형식	15
〈표 2-9〉 내비게이션 데이터 차종 정보	16
〈표 2-10〉 내비게이션 데이터 유종 정보	16
〈표 2-11〉 DTG 데이터의 포맷	18
〈표 2-12〉 DTG 데이터의 자동차 유형 코드	18
〈표 2-13〉 데이터의 용량 및 개별 차량의 링크 발생 건수	19
〈표 2-14〉 티맵 내비게이션 데이터 차종 분석 결과	20
〈표 2-15〉 티맵 내비게이션 데이터 유종 분석 결과	20
〈표 2-16〉 링크 기반 내비게이션 데이터 현황	21
〈표 2-17〉 연도별 내비게이션 데이터 프로파일링 비교	21
〈표 2-18〉 포인트 기반 내비게이션 데이터 현황	21
〈표 2-19〉 DTG데이터 차종 분석 결과	22
〈표 2-20〉 DTG 데이터 현황	22
〈표 3-1〉 티맵 오류 이벤트 필터링 비율	32
〈표 2-2〉 티맵 필터링 및 보정 데이터 표준 DB	33
〈표 3-3〉 2024년 DTG 수집 데이터 정상/오류 여부 정리	34
〈표 3-4〉 DTG 데이터 오류 발생 필터링 비율	36
〈표 3-5〉 DTG 전처리 데이터 DB	37
〈표 3-6〉 DTG 필터링 및 보정 데이터 표준 DB	39
〈표 3-7〉 티맵 상세도로망 기반 경로 데이터 표준 DB	42
〈표 3-8〉 DTG 상세도로망 기반 경로 데이터 표준 DB	42
〈표 3-9〉 경로 데이터 압축 및 통합 경로 데이터 NoSQL DB 포맷	44
〈표 3-10〉 머신러닝 기반 추정교통량 모형 구축 프로세스	48
〈표 3-11〉 교통량 추정을 위한 입력 데이터 DB	49
〈표 3-12〉 추정교통량 DB 테이블 구성	50
〈표 3-13〉 속도빈도분포 DB 포맷	55
〈표 3-14〉 연도별 도로등급별 평균속도	56
〈표 3-15〉 연도별 시도별 평균속도	57
〈표 3-16〉 개별 차량별 기종점 통행 DB 포맷	59

〈표 3-17〉 기종점 통행량 DB 포맷	61
〈표 3-18〉 기종점 통행량 DB 포맷	63
〈표 3-19〉 교통망 성능평가지표 DB 생성 기준	64
〈표 3-20〉 차량 교통망 성능평가지표	65
〈표 3-21〉 IPCC Guideline 온실가스 추정방법 비교	66
〈표 3-22〉 차종별 유형별 규모별 전국 연간주행거리	68
〈표 3-23〉 차종별 연료별 전국 연간주행거리	69
〈표 3-24〉 이산화탄소 배출량 네트워크 점유율	69
〈표 3-25〉 이산화탄소 배출량의 배출계수	71
〈표 3-26〉 시도별 차량주행거리 비교(평일 기준)	72
〈표 2-27〉 상세도로망 단위 속도지표 DB 포맷	73
〈표 3-28〉 시도별 혼잡시평균속도 비교(평일 기준)	74
〈표 3-29〉 시도별 정상시평균속도 비교(평일 기준)	74
〈표 2-30〉 상세도로망 단위 환경지표 DB 포맷	75
〈표 3-31〉 시도별 이산화탄소배출량 비교(평일 기준)	75
〈표 2-32〉 상세도로망 단위 혼잡강도 DB 포맷	76
〈표 3-33〉 시도별 혼잡시간강도 비교(평일 기준)	76
〈표 3-34〉 시도별 혼잡빈도강도 비교(평일 기준)	77
〈표 3-35〉 교통 네트워크 대표노드 테이블 정의	81
〈표 3-36〉 교통 네트워크 대표링크 테이블 정의	81
〈표 3-37〉 대표링크 기준 회전정보 테이블 정의	83
〈표 3-38〉 회전교통량 추정을 위한 주요 학습데이터	86
〈표 3-39〉 세종특별자치시 관측 회전교통량(연 평균) 테이블 정의	87
〈표 3-40〉 세종특별자치시 관측 회전교통량(일 단위) 테이블 정의	88
〈표 3-41〉 교통 네트워크 단위 링크 경로 데이터 정의	89
〈표 3-42〉 프로브 및 속도 데이터 정의	90
〈표 3-43〉 세종특별자치시 교통량 추정 학습데이터 테이블 정의	92
〈표 3-44〉 파라미터 최적화	95
〈표 3-45〉 수집 주기별 방향별 회전교통량 추정 결과	98
〈표 4-1〉 데이터 수집 현황 로그 항목	106
〈표 4-2〉 데이터 전처리 로그 항목	106
〈표 5-1〉 데이터 분석·활용 시스템 활용 실적	118
〈표 5-2〉 데이터 제공 및 분석 지원 주요 사례	119

그림목차

〈그림 2-1〉 터맵 네트워크와 상세도로망 네트워크 상세도 비교	17
〈그림 2-2〉 시도별 연평균 일 내비게이션 프로브 샘플을	23
〈그림 3-1〉 차량 GPS 데이터 가공 및 구축 프로세스	27
〈그림 3-2〉 링크 기반 데이터 전처리 및 경로 데이터 가공 프로세스	28
〈그림 3-3〉 포인트 기반 데이터 전처리 및 경로 데이터 가공 프로세스	28
〈그림 3-4〉 차량 GPS 데이터를 활용한 개별차량 경로DB 구축 프로세스	29
〈그림 3-5〉 데이터 적재	30
〈그림 3-6〉 자료형 변환	30
〈그림 3-7〉 데이터 포맷 변환	30
〈그림 3-8〉 주행 시작일 필드와 링크 진입시간 필드 날짜 불일치 오류 예시	31
〈그림 3-9〉 링크 중복 경로 예시	31
〈그림 3-10〉 일자오류 예시	31
〈그림 3-11〉 터맵 네트워크 미참조 링크	32
〈그림 3-12〉 진입시간, 진출시간 재계산	33
〈그림 3-13〉 압축 파일에 대한 손상 확인	34
〈그림 3-14〉 DTG 데이터 오류 검토	35
〈그림 3-15〉 좌표오류 및 시간오류 필터링 방법론	36
〈그림 3-16〉 통행분리 유형 판단을 위한 데이터 분석	38
〈그림 3-17〉 지하철도 음영구간의 데이터 재생성	38
〈그림 3-18〉 터널 음영구간의 데이터 재생성	39
〈그림 3-19〉 정지 시 꼬임구간의 데이터 재생성	39
〈그림 3-20〉 맵매칭 프로세스	40
〈그림 3-21〉 링크 매칭 정보를 통한 통행속도 산출	41
〈그림 3-22〉 통행병합 예시1	41
〈그림 3-23〉 통행병합 예시2	41
〈그림 3-24〉 압축된 경로 데이터 형태	43
〈그림 3-25〉 교통량 추정 및 구축 프로세스	47
〈그림 3-26〉 도로 등급별 신뢰도 분석 결과	52
〈그림 3-27〉 교통량 수준별 신뢰도 분석 결과	52
〈그림 3-28〉 기초교통 DB 구축 프로세스	53
〈그림 3-29〉 속도빈도분포 구축 프로세스	54
〈그림 3-30〉 연도별 도로등급별 속도 빈도분포 결과 비교	55
〈그림 3-31〉 개별 차량별 기종점 통행 DB 계산 예시	59
〈그림 3-32〉 기종점 통행시간 및 통행속도 DB 계산 예시	60
〈그림 3-33〉 기종점 통행량 DB 구축 프로세스	62
〈그림 3-34〉 기점과 종점의 시군구에 따른 집계 방식 차이	62

〈그림 3-35〉 기점과 종점의 시군구에 따른 집계 방식 차이	63
〈그림 3-36〉 교통량 추정 및 교통망 성능평가지표 구축 프로세스	64
〈그림 3-37〉 속도빈도분포 형태	70
〈그림 3-38〉 통계청의 실험적 통계에 활용되는 차량 교통망 성능평가지표	78
〈그림 3-39〉 교통 네트워크 개념	80
〈그림 3-40〉 대표노드 및 대표링크 구축 예시	81
〈그림 3-41〉 대표링크별 회전정보 구축 방법	83
〈그림 3-42〉 대표링크별 회전정보 구축 예시	83
〈그림 3-43〉 네트워크별 노드-링크 체계(도식화)	84
〈그림 3-44〉 세종특별자치시 교통 네트워크	85
〈그림 3-45〉 네트워크 연결성 변수 개념	91
〈그림 3-46〉 관측 교통량과 추정 회전교통량 산점도	99
〈그림 3-47〉 시간대별 추정 회전교통량 신뢰도 분석	100
〈그림 3-48〉 방향별 추정 회전교통량 신뢰도 분석	101
〈그림 4-1〉 모니터링 프로세스	107
〈그림 4-2〉 모니터링 메뉴 구조도	108
〈그림 4-3〉 데이터 수집 현황 모니터링 화면	109
〈그림 4-4〉 데이터 전처리 모니터링 화면	109
〈그림 4-5〉 기초 데이터 모니터링 화면	110
〈그림 4-6〉 기초 데이터 조건 설정 화면	111
〈그림 4-7〉 기초 데이터 시각화 화면	111
〈그림 4-8〉 기초 데이터 상세 내용 화면	111
〈그림 4-9〉 교통지표 우선순위 분석 예시 화면	112
〈그림 4-10〉 연도별 교통지표 변화분석 예시 화면	113



후
야

1. 과업의 개요

가. 과업의 배경 및 목적

1) 과업의 배경

- 전국 단위 공공의 기초 교통데이터 수집 및 지점 데이터 분석의 한계
 - 교통 분야의 기초 데이터인 교통량, 속도 데이터는 한국건설기술연구원, 한국도로공사, 광역지자체 등에서 조사하고 있으나, 데이터의 공간적 수집 범위가 매우 낮아 전국의 교통 현황을 분석하기에 한계가 있음
 - 공공에서 수집하고 있는 데이터는 지점 정보 중심의 데이터로 단편적 정보만 확인할 수 있어 시공간적으로 연결된 교통의 흐름을 파악하기엔 한계가 있음
- 전국 단위 일관된 교통DB 및 통행지표 부재
 - 데이터 수집기관, 수집방식 등 기관별로 상이하게 수집되는 데이터 수집 체계는 전국을 일관된 기준으로 분석하기엔 한계가 있음
- 빅데이터 전처리·가공 환경 및 빅데이터 기반 차량통행 분석시스템 부재
 - 차량 GPS 빅데이터를 효율적으로 전처리·가공하기 위한 환경이 부족하며, 과거에는 차량의 시·공간적 통행패턴과 현황을 분석하기 위해 모형 중심의 프로그램을 이용하여 분석을 수행하였음
- 국가통합교통체계효율화법 시행령 제8조1)에 의거하여 교통물류활동으로 발생

하는 교통혼잡 등 교통 관련 사회적 외부비용을 국가교통물류에 관한 지표를 설정하여 고시하도록 규정하고 있음

2) 과업의 목적

- 국가통합교통체계효율화법²⁾에 의거 국가교통데이터베이스 구축 및 운영을 위한 기초교통DB 및 통행지표 생성
 - 개별 차량 기반의 교통량, 속도, 차량주행거리 등 교통분야의 핵심 기초교통 DB 구축(공공 조사교통량 약 1만 개 지점, 전국 도로 62만 개 링크)
 - 차량 이용 및 교통물류 활동으로 인한 교통혼잡, 환경오염, 온실가스 배출 등 다양한 측면의 도로 교통망 성능평가 지표 구축
 - 지자체별로 상이한 방식으로 수집되고 구축되는 교통DB·통행지표의 한계를 개선하기 위한 전국단위 일관된 교통DB·통행지표 구축 필요
 - 여객 O/D구축 및 교통수요예측을 위한 기반데이터(승용차 O/D 데이터 등) 제공
- 국가교통 데이터 제공 및 교통 분야 주요 정책 지원
 - 이용자의 요구에 대응할 수 있는 데이터 제공 및 분석지원
 - 탄소공간지도 사업, 환경소음측정망 선정 사업, 지자체 실증사업 등 주요 정책 제언을 위한 데이터 및 분석 지원(매년 요청 및 제공)

나. 과업의 범위

1) 시간적 범위

- 기준연도 : 2024년
- 데이터 수집 : 2024년, 2025년 상반기(내비게이션)
- 데이터 가공 및 DB구축 : 2024년

1) 국가교통조사에는 다음 각 호 사항이 포함되어야 한다 - 교통물류활동으로 발생하는 교통혼잡, 교통사고, 환경오염, 온실가스 배출 등 교통 관련 사회적 비용, 그 밖에 교통 관련 정책 및 계획의 수립, 교통시설 투자분석 및 평가에 필요한 사항

2) 국가통합교통체계효율화법 제12조(국가교통조사), 제14조(정보통신수단 등을 통한 교통조사) 및 동법 시행령 제8조(국가교통조사의 실시)

2) 공간적 범위

- 전국 2차로 이상 도로 및 주요 도로³⁾

3) 내용적 범위

- 차량 원시 GPS 빅데이터 전처리 및 DB가공·구축
 - 차량 GPS 데이터, DTG 데이터 수집 및 검토
 - 원시 데이터 전처리 및 경로 데이터 가공
 - 기초 교통 DB구축 및 검증
- 교통량 추정
 - 네트워크 연결성을 고려한 교통량 추정
 - 프로토타입 회전교통량 추정
 - 기초 교통DB 구축을 위한 교통량 추정 신뢰도 분석 및 연도별 패턴 분석
- 차량 통행지표 가공 알고리즘 개선 및 DB구축
 - 통행지표 가공 알고리즘 개선 및 고도화
 - 목적별 전국 통행지표 산출 및 DB 구축
- 데이터 제공 및 분석지원

3) 주요 도로 : 편도 1차로 도로 중 데이터 수집의 안정성이 확보된 도로

2. 차량 GPS 빅데이터 수집 및 분석

가. 차량 GPS 빅데이터 수집 개요

1) 데이터 수집 대상

- 본 과업의 기준연도는 2024년 기준으로 차량 GPS 빅데이터 기반 기초교통 및 교통망 성능평가지표DB를 구축하기 위해 다음과 같은 데이터를 수집함
 - 링크 기반 내비게이션 데이터는 티맵의 내비게이션 사용자의 주행 정보, 주행 정보를 도로구간(링크)별로 나타내는 도로 네트워크 데이터를 수집함
 - 포인트 기반 내비게이션 데이터는 티맵의 내비게이션 사용자의 주행 정보, 주행 정보를 GPS 단위로 데이터를 수집하므로 초 단위 데이터를 수집함
 - 포인트 기반 사업용 차량 운행기록계(DTG) 데이터는 한국교통안전공단에서 제공하며, 사업용 차량에 부착된 운행기록계로 수집되는 버스, 화물, 택시의 주행 정보 데이터를 수집함

〈표 1〉 데이터 수집 목록 및 기준연도

구분	데이터 목록	수집처	수집연도
링크 기반	내비게이션 도로 네트워크	티맵	2024년 1-12월, 2025년 상반기
	내비게이션 주행 정보		
포인트 기반	내비게이션 주행 GPS 정보	티맵	2024년 6, 7, 10, 12월
	사업용 차량 운행기록계 버스 주행 정보	한국교통안전 공단	2024년 1-12월
	사업용 차량 운행기록계 화물 주행 정보		
사업용 차량 운행기록계 택시 주행 정보			

- 시의성 있는 데이터 제공을 위해 당해연도인 2025년 상반기 데이터를 하반기 부터 수집하여 빠르게 가공 및 구축 가능한 지표를 산출함
 - 링크 기반 내비게이션 데이터는 2024년, 2025년 12주, 명절을 포함하여 월별 추세를 파악할 수 있는 기간으로 총 185일을 수집하며, 2025년 7월 데이터까지 데이터를 수집하였음
 - 포인트 기반 내비게이션 데이터는 2024년 각 4주, 여름과 겨울 추세를 파악할 수 있는 기간으로 총 31일을 수집하였으며, 수집 과정에서 데이터 오류가 확

인되어 일부 기간에 대해 재요청을 수행함

- 특히 포인트 기반 데이터는 초기 계획 대비 수집 시기가 변경되었으며, 데이터 제공기관의 서버 장애로 1~5월 데이터가 삭제됨에 따라, 대체 가능한 기간인 2024년 6월, 7월, 10월, 12월 데이터로 조정하여 수집을 완료함

〈표 2〉 수집 데이터의 시간적 범위

구분	링크 기반 내비게이션		포인트 기반 내비게이션
	2024년	2025년	2024년
1월	22일 ~ 28일 (7일)	20일 ~ 26일 (7일)	-
2월	13일 ~ 18일 (7일)	10일 ~ 16일 (7일)	-
3월	11일 ~ 17일 (7일)	10일 ~ 16일 (7일)	-
4월	15일 ~ 21일 (7일)	14일 ~ 20일 (7일)	-

6월	10일 ~ 16일 (7일)	16일 ~ 22일(7일)	17일 ~ 23일(7일)
7월	15일 ~ 21일 (7일)	21일 ~ 27일(7일)	8일 ~ 14일(7일)

9월	19일 ~ 25일 (7일)	15일 ~ 21일 (7일)	-
10월	14일 ~ 20일 (7일)	13일 ~ 19일 (7일)	21일 ~ 27일(7일)
11월	11일 ~ 17일 (7일)	10일 ~ 16일 (7일)	-
12월	16일 ~ 22일 (7일)	15일 ~ 21일 (7일)	9일 ~ 15일(7일)
명절	설날 : 2월 09일 ~ 2월 12일 (4일)	설날 : 1월 27일 ~ 1월 30일 (4일)	추석 : 9월 16일 ~ 9월 18일 (3일)
	추석 : 9월 16일 ~ 9월 18일 (3일)	추석 : 10월 03일 ~ 10월 09일 (7일)	
	합계	90일	
합계	90일	95일	31일

2) 데이터 수집 대상의 특징 및 형식

- 링크 기반 내비게이션 데이터는 링크의 진입시간, 진출 시간, 속도 등의 정보가 포함되어 있으며, 차종과 유종, 목적지 정보가 포함되어 제공받음
- 포인트 기반 내비게이션 데이터는 GPS좌적 X좌표, Y좌표에 따른 주행 시간, 속도, 고도값 등의 정보가 있으며, 주행차량의 배기량과 연식을 포함하여 제공받아 링크 기반 내비게이션 데이터와 차이가 있음
- 사업용 운행기록계(DTG) 데이터는 사업용 차량의 운행정보를 실시간으로 저장하여 변화하는 차량운행 상황을 기록하는 디지털 운행기록계를 통해 수집되는 데이터로 자동차의 순간속도, GPS, 분당 엔진회전수, 가속도, 자동차 유형 등의 정보가 제공되며, 1초 단위의 GPS 좌표를 수집함

나. 차량 GPS 원시데이터 분석

- 2024년 티맵 내비게이션 데이터는 2023년 대비 약 8.2% 증가된 총 858억 건의 개별 차량 주행 발생 건수(이벤트 수)가 수집되었음
- 2024년 DTG 데이터는 2023년 대비 약 39.2% 증가한 총 2조 8천 9백억 건의 이벤트가 수집되어, 사업용 차량 중심의 고해상도 시계열 데이터 확보가 크게 향상됨

〈표 3〉 데이터의 용량 및 개별 차량의 링크 발생 건수

데이터명	용량(TB)			이벤트 수(천 건)		
	2022년	2023년	2024년	2022년	2023년	2024년
링크 기반 내비게이션 데이터	19.4	22.3	23.9	70,349,051	79,364,038	85,844,541
DTG 버스 데이터	9.64	9.22	2.57	772,320,634	770,774,514	213,774,634
DTG 화물차 데이터	3.93	4.32	1.04	317,497,577	355,365,816	86,153,122
DTG 택시 데이터	5.91	10.24	3.85	491,244,500	953,355,667	352,145,991

- 티맵 내비게이션 데이터의 차종 및 유종 정보는 전체 이용자가 아닌 자동차보험 할인 프로그램에 참여한 “운전습관 가입자”를 통해 수집되므로, 모든 주행 데이터에 대해 해당 정보가 포함되지 않는 구조적 한계가 존재함
- 티맵 내비게이션 데이터를 통한 차종 정보 분석 결과
 - 주행 고유 ID 기준 분석 결과, 2024년 데이터의 약 82% 이상에서 차종 정보가 확인되어 전반적인 활용성은 확보된 것으로 판단됨
 - 연도별로 차종 분석 결과, 2024년 데이터는 승용차가 약 75%로 가장 많이 발생하였고, 점차적으로 증가하고 있는 것으로 분석됨

〈표 4〉 티맵 내비게이션 데이터 차종 분석 결과

차종 정보	2022년	2023년	2024년
승용차	55.2%	68.6%	74.8%
경차	2.4%	3.0%	3.3%
중형승합차	1.7%	1.8%	1.3%
대형승합차	0.5%	0.6%	0.5%
소형화물차	0.5%	0.8%	1.1%
대형화물차	0.03%	0.1%	0.8%
특수화물차	0.02%	0.3%	0.3%
정보 없음	39.6%	24.9%	17.8%
합계	100%	100%	100%

- 티맵 내비게이션 데이터를 통한 유종 정보 분석
 - 내비게이션 데이터의 주행 고유 ID별로 유종 정보가 포함된 비율을 분석한 결과, 2024년 데이터는 약 75% 이상의 데이터에서 유종 정보가 확인되었음
 - 연도별로 유종 분석 결과, 2024년 데이터는 휘발유에서 약 56%로 가장 많이 발생하였고, 점차적으로 증가하고 있는 것으로 분석됨

〈표 5〉 티맵 내비게이션 데이터 유종 분석 결과

유종 정보	2022년	2023년	2024년
휘발유	41.4%	51.3%	56.2%
경유	16.6%	20.1%	20.4%
LPG	4.7%	5.6%	5.5%
고급휘발유	1.1%	1.3%	1.4%
전기	1.1%	1.5%	2.0%
정보 없음	35.3%	20.3%	14.5%
합계	100%	100%	100%

- DTG 데이터는 자동차 유형 정보를 활용하여 차종별 분포를 분석하였으며, 총 125,506천 대의 주행 고유 ID 기준으로 구성됨
 - 분석 결과, 일반화물이 52.30%로 가장 높은 비중을 차지하였으며, 전세버스(25.89%), 시내버스(13.17%) 순으로 나타남
 - 그 외 농어촌버스(2.39%), 개인화물(2.38%), 마을버스(1.84%), 시외버스(1.36%) 등으로 구성되어, 화물 및 특정 버스 유형 중심의 데이터 구조를 보임

〈표 6〉 DTG데이터 차종 분석 결과

차종 정보	주행 고유 ID 대수 (천대)	비율
시내버스	16,532	13.17%
농어촌버스	2,994	2.39%
마을버스	2,306	1.84%
시외버스	1,713	1.36%
고속버스	841	0.67%
전세버스	32,495	25.89%
일반화물	65,638	52.30%
개인화물	2,988	2.38%
합계	125,506	100%

3. 차량 GPS 데이터 가공 및 DB구축

가. 차량 DB가공 및 구축 방법론 설정

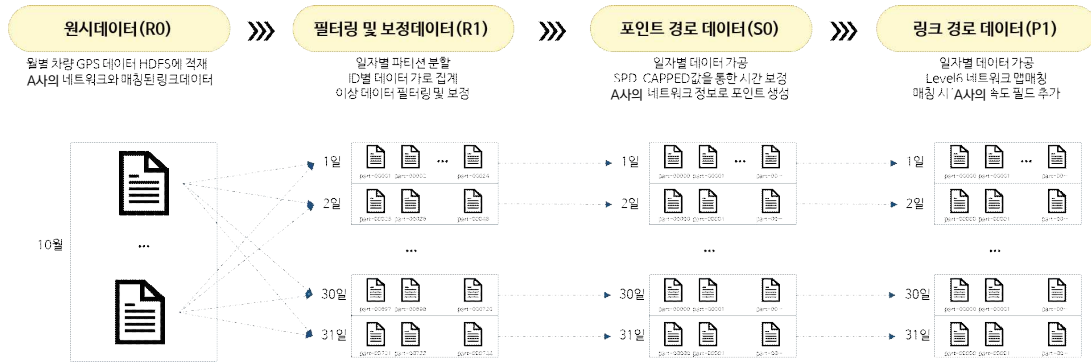
- 본 사업에서 활용하는 차량 GPS 데이터는 내비게이션 데이터와 사업용 DTG 데이터로 구성되며, 데이터 활용을 위해서는 각 데이터의 특성을 고려한 분석 및 가공·검증 과정이 필요함
- 티맵 내비게이션 데이터는 링크 기반, DTG 데이터는 포인트 기반으로 수집되어 데이터 구조와 수집 주기에서 차이가 존재하므로, 데이터 유형별 특성에 적합한 전처리 및 가공 모듈을 적용함
- 차량 GPS 데이터 가공 및 구축 프로세스는 데이터 전처리 → 개별 차량 경로 DB 구축 → 1차 가공 DB(기초교통 DB) 구축 → 2차 가공 DB(교통망 성능평가 지표 DB) 구축의 단계로 체계적으로 수행됨



〈그림 1〉 차량GPS 데이터 가공 및 구축 프로세스

- 티맵 내비게이션 데이터는 링크 기반 데이터로 제공되며, 이를 활용하기 위해 링크 단위 정보를 1초 단위 궤적(포인트) 데이터로 변환하여 표준 포인트 기반 경로 데이터를 우선 구축함
- 생성된 포인트 기반 경로 데이터는 티맵 네트워크의 세분화된 구조를 고려하여

포인트 단위 변환 및 재매칭을 수행한 후 Level6 네트워크와의 맵매칭을 통해 이기종 네트워크 간 차이를 보정하고 표준화된 링크 경로 데이터로 재구성함



〈그림 2〉 링크 기반 데이터 전처리 및 경로 데이터 가공 프로세스

- DTG 데이터는 포인트 기반 데이터로 포인트 기반 데이터 맵매칭 및 경로 가공 모듈을 이용하여 가공 및 구축함(해당 모듈은 R&D사업을 통해 개발된 것임)

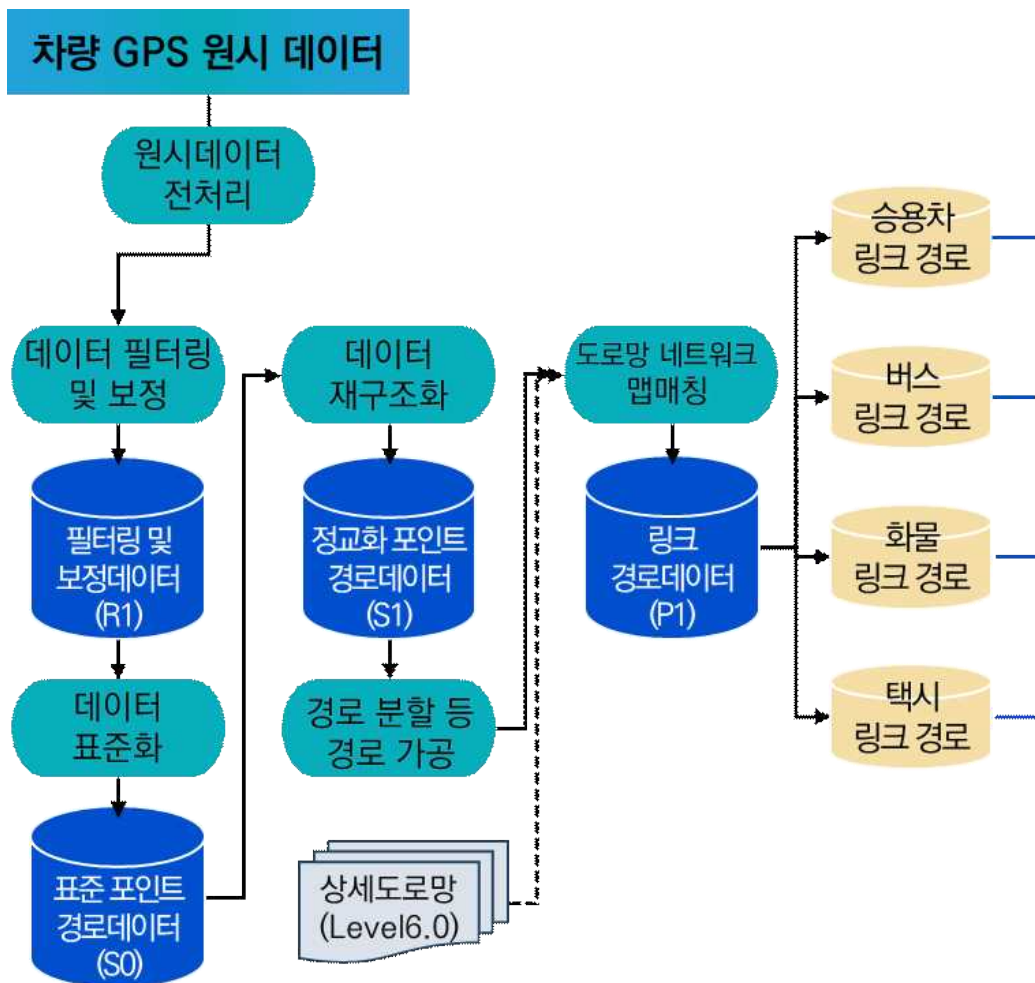


〈그림 3〉 포인트 기반 데이터 전처리 및 경로 데이터 가공 프로세스

나. 차량 GPS 데이터 전처리 및 경로 가공

1) 개별차량 경로DB 구축 프로세스

- 차량 GPS 데이터를 활용한 개별차량 경로 DB 구축은 데이터 유형별 특성을 반영하여 단계적으로 수행됨
 - 링크 기반 데이터는 이미 도로 네트워크와 맵매칭된 구조로 제공되므로, 별도의 궤적 재구조화 없이 정합성 검증 및 경로 생성 중심으로 구축함
 - 반면, 포인트 기반 데이터는 GPS 수신 오류 및 노이즈가 포함될 수 있어, 궤적 보정 및 정합화 등 재구조화 과정을 거쳐 경로 데이터를 생성함



<그림 4> 차량 GPS 데이터를 활용한 개별차량 경로DB 구축 프로세스

2) 원시데이터 전처리

○ 원시데이터 적재 및 전처리

- 대용량의 압축 파일을 빅데이터 가공 시스템으로 처리하기 위해 원시 데이터를 분할하여 HDFS에 적재함



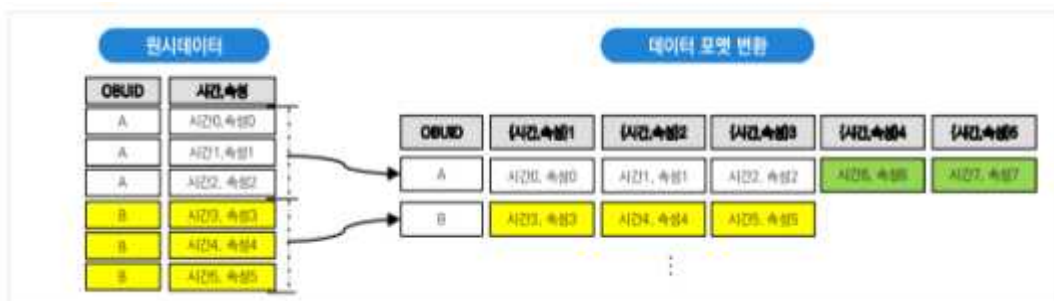
〈그림 5〉 데이터 적재

- 원시데이터를 일자별로 분리하고, 불필요한 속성정보를 제거하여 데이터 용량을 축소시킴



〈그림 6〉 자료형 변환

- 빅데이터 환경의 처리속도 향상을 위해 데이터 포맷을 변환



〈그림 7〉 데이터 포맷 변환

○ 내비게이션 데이터 필터링 및 보정

- 데이터 품질 확보를 위해 시간, 필드, 속도 등 오류 유형별 기준을 정의하고 이에 따른 필터링 및 보정 과정을 수행함
- (시간 오류) 주행 시작일 필드 오류 유형 : 주행 시작일(R_P_REQDATE)과 링크 진입 시간(ENTRY_TIME) 필드의 일자가 맞지 않는 경우 링크 진입시간을 기준으로 주행 일자에 대해 정렬하여 일자별로 분할함

DATE	TRIP_INIT_SESSION_ID	LINK_SEQ	MESH_CD	LINK_ID	LINK_DIR	LENGTH	DRIVING TIME	ENTRY_TIME	EXIT_TIME
20210209	105520210208235734644500 ***	1	4787	16033	0	63.894627	6.790106	20210209000226	20210209000232
20210209	105720210208235718506500	1	5631	114	0	121	44.565998	20210209000065	20210209000737
20210209	108120210208232709308500	1	4777	3744	1	28	2.8363636	20210209073338	20210209073339
20210209	105020210209175213265501	1	7413	28853	1	84	27.49091	20210210030545	20210210030559
20210209	105020210209210128425500	1	4775	9789	1	231.51741	11.882713	20210210084021	20210210084032

<그림 8> 주행 시작일 필드와 링크 진입시간 필드 날짜 불일치 오류 예시

- (시간 오류) 링크 시퀀스 필드 오류 유형 : 링크 진입시간(ENTRY_TIME)과 링크 진출 시간(EXIT_TIME) 필드의 시간이 링크 시퀀스(LINK_SEQ) 순서와 다를 경우 진입시간, 진출시간, 링크 시퀀스 순서로 정렬함
- (필드 오류) 링크 중복 오류 유형 : 동일 시간에서 발생한 링크의 중복 경로를 보정함



<그림 9> 링크 중복 경로 예시

- 일자오류(경로 기준일 불일치): 링크 진입시간이 기준 일자와 다르거나 당일·익일 범위를 벗어나는 경우 오류 데이터로 판단하여 제거함

20240122	108020240122090702014501	2	4787	14341	1	48	9	20240122090709	20240122090714	WN	26.5	26.5
20240122	108020240122090702014501	5	4787	4221	0	58	1	20240122090715	20240122090718	WN	26.5	26.5
20240122	1046020240122090702014505	1	6717	05360	1	106	1	20240122090721	20240122090724	WN	26	26
20240122	1041020240122090702014502	1	4785	20739	0	56	0	20240122143907	20240122143917	WN	5	5
20240122	1041020240122090702014502	1	4785	20739	1	56	0	20240122143912	20240122143912	WN	5	5

<그림 10> 일자오류 예시

- 속도 오류(이상 속도값) : GPS 수신 오류 등으로 인해 비정상적으로 산출된 속도값 (200km/h 초과)은 이상치로 판단하여 필터링함

○ DTG 데이터 필터링 및 보정

- DTG 데이터는 대용량 데이터 특성상 추출 과정에서 파일 손상 및 일부 데이터

누락이 발생할 수 있어, 압축 해제 및 적재 단계에서 오류 여부를 점검하고 이상 데이터는 재요청함

- 압축 해제 과정에서 손상된 파일 및 데이터 오류를 확인하며, 단말기 ID 누락, 특정 기간(휴일 등) 데이터 누락 사례를 검출함

<그림 11> DTG 데이터 오류 검토

- DTG 데이터 검증 및 필터링은 오류 탐지, 보정, 삭제 및 중복 제거 과정을 통해 데이터 품질을 향상시키는 방식으로 수행함

- (속도 기준) 정지 상태(3km/h 이하)로 판단되는 데이터는 분석 목적에 따라 제거
- (좌표 기준) 좌표 누락 또는 오류로 위치 식별이 불가능한 데이터는 제거
- (시간 기준) 시간정보 오류를 필터링하고, Timestamp 형식으로 표준화하여 저장함
- DTG 원시데이터는 압축 해제 및 이상 여부 검증 후 HDFS에 적재하고, 사전에 정의된 오류 유형별 보정 기준을 적용하여 전처리를 수행함

- (좌표 오류) 수집된 포인트 데이터가 대한민국 육지 영역을 벗어나는 경우 오류로 판단하여 해당 좌표 데이터를 제거함
- (시간 오류) 제공 기준 일자와 불일치하거나 비정상적인 시간이 입력된 경우 오류 데이터로 판단하여 제거함
- (중복오류) 동일 시간에 여러 데이터가 발생한 경우 → 중복 데이터 제거
- 통행정보 외 불필요 컬럼을 제거하고 차중 정보를 포함한 전처리를 수행하여 데이터 용량을 효율화함

3) 경로 데이터 구축 및 경로 압축(통합)

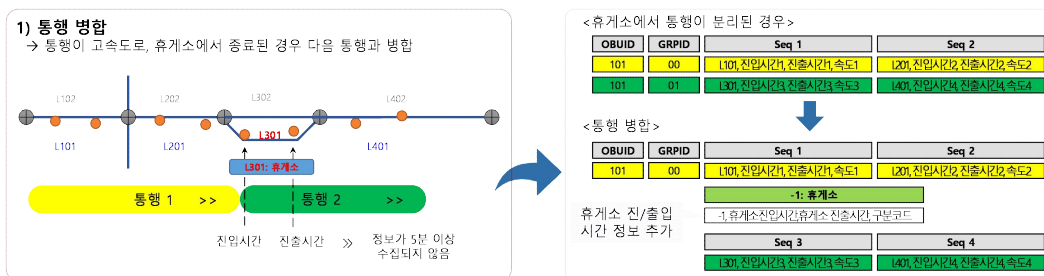
- 티맵과 DTG의 상세도로망 기반 경로 데이터 가공 및 DB 구축 프로세스는 동일함
- 본 사업에서 구축하는 기초 교통 및 통행지표 DB의 공간적 범위는 상세도로망 네트워크 기준으로 필터링 및 보정된 표준DB와 도로망 네트워크를 맵매칭 후 링크 기준의 경로 데이터 생성 필요
- 맵매칭된 링크 경로 데이터를 통해 링크 통행속도 산출
 - 링크 통과시 포인트의 좌표 정보를 통해 통행속도를 산출
 - 궤적별 포인트-링크 간 매칭된 페어 정보를 통하여 개별 차량의 링크 통행속도를 산출

순번	포인트시간	위도	경도	링크ID	링크출발거리	링크종료거리	GIS...	GIS...	GPS...	GPS...
125	2019/05/15 20:35:13	967962.028	1951976.266	57276491201	68.49	162.19	230.68	87.53	195.56	88.00
126	2019/05/15 20:35:14	967985.671	1951984.697	57276491201	93.49	137.20	230.68	87.53	195.56	88.00
127	2019/05/15 20:35:15	968009.154	1951991.78	57276491201	118.00	112.68	230.68	87.53	195.56	88.00
128	2019/05/15 20:35:16	968032.637	1952000.000	57276491201	142.98	88.98	230.68	87.53	195.56	88.00
129	2019/05/15 20:35:17	968056.120	1952008.213	57276491201	167.96	65.26	230.68	87.53	195.56	88.00
130	2019/05/15 20:35:18	968079.603	1952016.426	57276491201	189.00	41.68	230.68	87.53	195.56	88.00
131	2019/05/15 20:35:19	968102.594	1952024.639	57276491201	212.06	18.62	230.68	87.53	195.56	88.00
132	2019/05/15 20:35:20	968126.077	1952032.852	57276491001	4.73	234.83	230.68	83.95	233.24	84.00

<그림 12> 포인트 정보와 매칭된 링크의 통행속도 산출

- 휴게소 또는 고속도로에서 내비게이션 종료 또는 실행 시 통행의 시종점으로 적합하지 않기에, 해당 구간은 연속된 통행을 유지할 수 있도록 통행을 병합함
 - 통행이 병합되는 경우는 휴게소-휴게소, 고속도로-고속도로, 도시고속도로-고속도로 조합의 경우 발생
 - 통행 병합 시 시간 정보 및 구분 코드 정보를 추가하여, 데이터 분석 및 검증 시 활용할 수 있도록 함

· 통행 병합 -> 휴게소, (도시)고속도로에서의 연속된 통행 유지



<그림 13> 통행병합 예시

○ 티맵 상세도로망 기반 경로 데이터 가공 및 표준DB 형태로 변환 및 저장

〈표 7〉 티맵 상세도로망 기반 경로 데이터 표준 DB

K-V	순번	필드명	필드설명	자료형	비고
K	1	OBU_ID	주행 고유 ID	string	
	2	GRP_ID	통행 그룹 ID	integer	
	3	CAR_MODEL	차종("운전습관 가입자" 대상)	string	
	4	CAR_FUEL	유종("운전습관 가입자" 대상)	string	
	5	DEST_NAME	목적지명	string	
	6	DB_NAME	목적지 중분류 정보(POI가 아닌 주소 검색 등의 경우null)	string	
V	7	vLinkId	가상링크ID	long	
	8	inTime	링크 진입시간	long	
	9	outTime	링크 진출시간	long	
	10	lev6Len	6레벨 링크 길이	double	
	11	lev6Spd	6레벨 링크 속도	double	
	12	gpsLen	링크에 매칭된 포인트 간 길이의 합	double	
	13	tmapSpd	tmap 링크 속도	double	

○ DTG 상세도로망 기반 경로 데이터 가공 및 표준DB 형태로 변환 및 저장

〈표 8〉 DTG 상세도로망 기반 경로 데이터 표준 DB

K-V	순번	필드명	필드설명	자료형	비고
K	1	OBU_ID	주행 고유 ID	string	
	2	GRP_ID	통행 그룹 ID	integer	
	3	VEH_TYPE	차종	double	
	4	AVG_GAP	음영구간을 제외한 원본-Recon 포인트 간 거리 오차의 평균	double	
	5	RELIABILITY1	이상치 판단 후 궤적 신뢰성	double	
	6	RELIABILITY2	필터링 후 궤적 신뢰성	double	
	7	MATCHING_RATE	포인트-링크 매칭률	double	
	8	CONNECTING_RATE	링크 연결률	double	
V	9	vLinkId		long	
	10	inTime	링크 진입시간	long	
	11	outTime	링크 진출시간	long	
	12	lev6Len	6레벨 링크 길이	double	
	13	lev6Spd	6레벨 링크 속도	double	
	14	gpsLen	링크에 매칭된 포인트 간 길이의 합	double	
	15	tmapSpd	tmap 링크 속도	double	

다. 교통량 추정DB 구축

1) 교통량 추정 목적

- 전국 관측 교통량은 약 1만 개 지점 수준으로, 상세도로망(Level6) 링크 수(약 62만 개) 대비 약 1.6%에 불과하여 공간적 커버리지가 매우 제한적임
- 또한 지자체 관측교통량은 차종 미구분, 일부 시간대 미관측 등 데이터 제약이 존재하여 정책·운영에 필요한 차종별·시간대별 교통현황 파악에 한계가 있음
 - 이에 따라 제한된 관측자료를 보완·확장하여 전국 도로구간 단위의 차종별·시간대별 교통량을 추정함으로써, 교통수요 분석, 혼잡관리, 신호운영 및 교통정책 수립을 지원하기 위한 기초 데이터 확보가 필요함

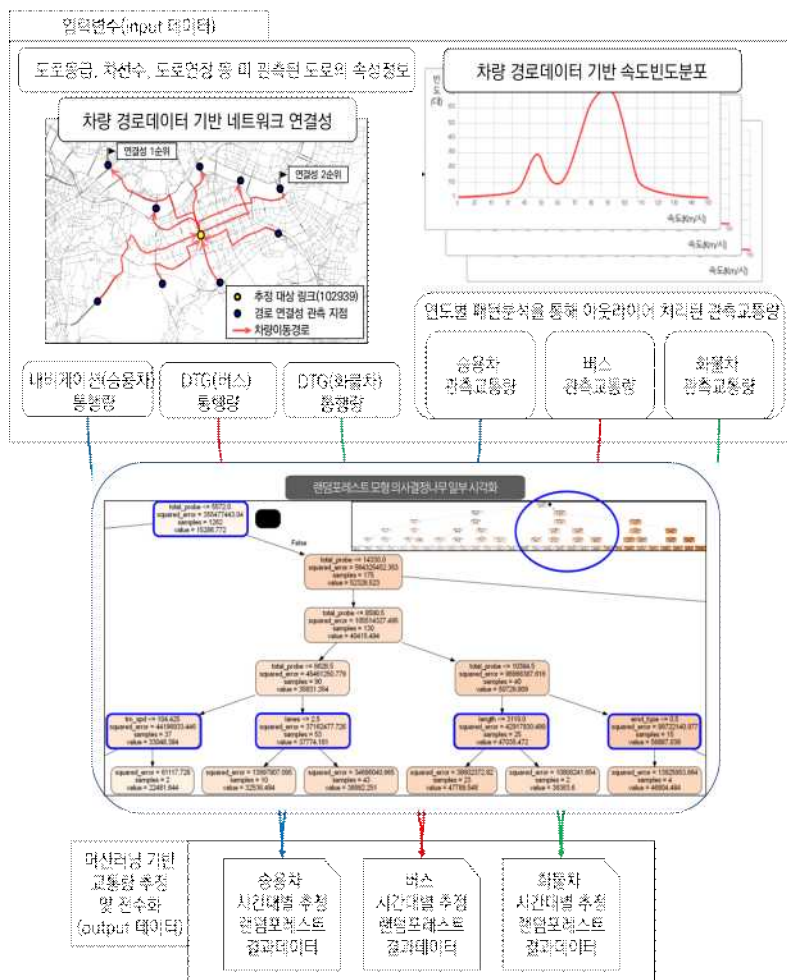
2) 교통량 추정 입력 데이터 구축 프로세스

- 교통량 추정을 위한 입력 데이터 구축
 - 상세도로망(Level6) 네트워크 기반 경로 데이터로부터 생성된 속도 프로파일 DB를 활용하여 교통량 추정 입력 데이터 DB를 구축함
 - 데이터 정합성 확보를 위해 시간대 및 링크 기준으로 데이터 구조를 통일하여 분석 가능한 형태로 재구성함
- 차종별 통행량 및 속도 데이터 생성
 - 차종별 링크 단위 속도빈도분포에서 평균속도를, 통행량 데이터에서 차종별 통행량을 추출하여 전 링크에 대한 차종별 통행량 및 속도 데이터를 생성함
- 관측 링크 및 전체 링크 대상 입력 데이터 가공
 - 관측교통량이 존재하는 링크를 대상으로 머신러닝 학습에 활용되는 관측지점 입력 데이터를 가공함
 - 학습된 모델을 기반으로 관측지점을 포함한 전체 링크의 교통량을 추정하기 위해 전 링크 대상 입력 데이터를 구축함

3) 교통량 추정 입력 데이터 구축

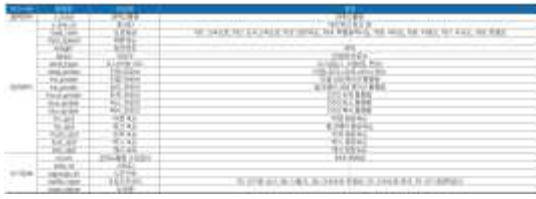
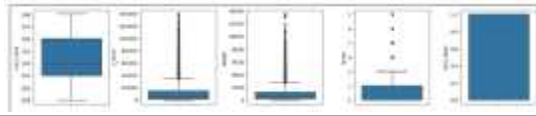
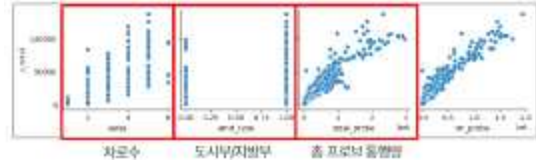
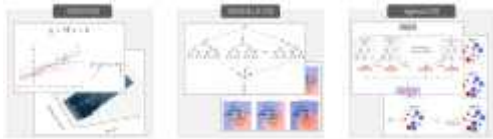
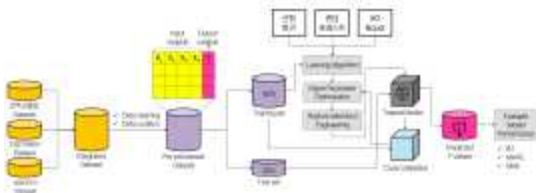
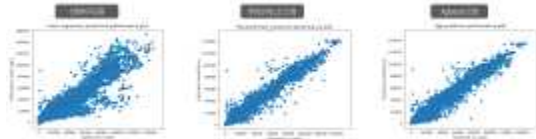


○ 학습데이터 구성

- 약 1만 개 관측지점을 학습지점과 검증지점으로 구분하고, 이 중 약 8,500개 지점을 학습데이터로 활용함
- 학습데이터는 차종별 통행량, 속도, 도로 특성(차선수, 도로등급 등)과 함께 네트워크 연계 정보를 포함하도록 구성함
- 티맵 및 DTG 경로 데이터의 연속 주행 기록을 활용하여 대상 관측지점과 연결된 인접 지점 정보를 추가 변수(네트워크 연결성 지표)로 생성함
- 개별 차량 경로를 기반으로 통행량, 차선수, 도로등급 등을 고려하여 최대 10개의 인접 후보 지점을 도출하고, 이 중 연결성이 높은 3개 지점을 선별하여 네트워크 연결성 변수로 반영함. 이를 통해 단일 링크가 아닌 인접 링크 간 상호 영향성을 반영하여 교통량 추정 정확도를 향상시킴



〈그림 14〉 교통량 추정 및 구축 프로세스

〈표 9〉 머신러닝 기반 추정교통량 모형 구축 프로세스

구분	설명	예시																																
1차가공DB를 활용한 입력변수 구축	관측교통량, 차량 프로브 통행량, 평균속도 등 교통량 추정을 위한 입력 변수 구축																																	
데이터 전처리	수집된 데이터의 이상치 검토 및 전처리																																	
탐색적 데이터분석	변수별 상관관계 분석을 통한 입력 변수 선정																																	
입력변수 선정	도로등급, 제한속도, 도로연장, 차로수, 도시부/지방부, 차량프로브 통행량, 평균속도	<table border="1"> <thead> <tr> <th>번호</th> <th>변명</th> <th>단위명</th> <th>설명</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>road_rank</td> <td>도로등급</td> <td>고속도로, SA(국도/도도, 일반국도, 특별광역시/도, RAS, 지방도, 시군도, 민공로</td> </tr> <tr> <td>2</td> <td>max_speed</td> <td>제한속도</td> <td></td> </tr> <tr> <td>3</td> <td>length</td> <td>용구연장</td> <td>미터</td> </tr> <tr> <td>4</td> <td>lanes</td> <td>차로수</td> <td>단방향 차로수</td> </tr> <tr> <td>5</td> <td>emid_type</td> <td>도시/지방 코드</td> <td>도시(0)-1, 지방(0), 미지</td> </tr> <tr> <td>6</td> <td>total_probe</td> <td>전체 프로브</td> <td>티엠 정량값+티엠웨이 정량값+트릭 정량값+버스 정량값+택시 정량값</td> </tr> <tr> <td>7</td> <td>tm_spd</td> <td>티엠 속도</td> <td>티엠 연평균 평균 속도 티엠 정량 평균 속도</td> </tr> </tbody> </table>	번호	변명	단위명	설명	1	road_rank	도로등급	고속도로, SA(국도/도도, 일반국도, 특별광역시/도, RAS, 지방도, 시군도, 민공로	2	max_speed	제한속도		3	length	용구연장	미터	4	lanes	차로수	단방향 차로수	5	emid_type	도시/지방 코드	도시(0)-1, 지방(0), 미지	6	total_probe	전체 프로브	티엠 정량값+티엠웨이 정량값+트릭 정량값+버스 정량값+택시 정량값	7	tm_spd	티엠 속도	티엠 연평균 평균 속도 티엠 정량 평균 속도
번호	변명	단위명	설명																															
1	road_rank	도로등급	고속도로, SA(국도/도도, 일반국도, 특별광역시/도, RAS, 지방도, 시군도, 민공로																															
2	max_speed	제한속도																																
3	length	용구연장	미터																															
4	lanes	차로수	단방향 차로수																															
5	emid_type	도시/지방 코드	도시(0)-1, 지방(0), 미지																															
6	total_probe	전체 프로브	티엠 정량값+티엠웨이 정량값+트릭 정량값+버스 정량값+택시 정량값																															
7	tm_spd	티엠 속도	티엠 연평균 평균 속도 티엠 정량 평균 속도																															
모형 검토	선형회귀모형, 랜덤포레스트모형, XGboost 모형 선정																																	
교통량 추정모형 구축	전처리된 데이터와 선정된 모형을 활용하여 교통량 추정모형 구축																																	
교통량 추정 및 평가	구축된 교통량 추정의 추정교통량 신뢰도 분석																																	
교차검증	최종 선정된 랜덤포레스트 모형에 대한 교차검증을 통해 일반화 성능 검토																																	
변수 분석	교통량 추정에 영향을 끼치는 변수별 영향 분석																																	

〈표 10〉 교통량 추정을 위한 입력 데이터 DB

순번	필드명	필드설명	자료형
1	v_link_id	가상링크ID	Long
2	traffic_type	조사지점 코드	integer
3	length	도로연장	string
4	lanes	차선수	string
5	road_rank	도로등급	Integer
6	time_code	시간코드	integer
7	v_total_wd	전체 관측교통량	double
8	v_auto_wd	승용 관측교통량	double
9	v_bus_wd	버스 관측교통량	double
10	v_truck_wd	트럭 관측교통량	double
11	total_probe	전체 통행량	integer
12	tm_probe	티맵 통행량	integer
13	truck_probe	버스 통행량	integer
14	bus_probe	트럭 통행량	integer
15	taxi_probe	택시 통행량	integer
16	tm_spd	티맵 속도	double
17	truck_spd	버스 속도	double
18	bus_spd	트럭 속도	double
19	taxi_spd	택시 속도	double
19	emd_type	행정구역 동 여부	Integer
21	tlfd_code	도로의 기능 구분 코드	Integer
22	length	도로길이	double
23	vkt	총도로연장	double
24	k	티맵 밀도	double
25	sido_id	시도ID	Integer
26	tm_scale	승용 스케일	double
27	bus_scale	버스 스케일	double
28	truck_scale	트럭 스케일	double
29	volume_1	연결된 관측교통량1	Double
30	probe_1	연결된 프로브통행량1	Double
31	volume_2	연결된 관측교통량2	Double
32	probe_2	연결된 프로브통행량2	Double
33	volume_3	연결된 관측교통량3	Double
34	probe_3	연결된 프로브통행량3	Double

4) 추정교통량 구축

- 전국에 조사된 교통량은 약 1만 개 지점으로 상세도로망 Level6네트워크의 링크 갯수(약 62만 개)에 비해 매우 낮은 샘플율을 가지고 있음
- 각 지자체에서 조사한 관측교통량은 차종이 구분되어 있지 않거나, 24시간 전체 시간대에 조사하지 않는 경우가 많아 차종별 시간대별 교통량 추정을 위해 상시지점의 차종을 추정함
- 본 과업을 통해 교통량 추정 알고리즘을 개선하여 과거연도(2019~2023년)데이터를 재산정 및 2024년 교통량을 추정하였음
- 기초교통 DB 기반으로 전국 2차로 이상 도로에 대한 추정교통량DB 구축함
 - 기본 구축범위는 차종(승용차, 버스, 트럭), 평일, 시간(전일, 시간대별)단위, Level6 링크 단위 구축
 - 집계단위 : 상세도로망, 주요도로망, 표준노드링크, 행정구역(시도/시군구/읍면동)
- 한국개발연구원(2008)에 따르면 추정교통량의 신뢰성을 평가하기 위해 조사교통량과의 오차율 기준으로 추정교통량에 대한 평가를 수행하고 있으며, 본 연구에서 분석하고 있는 조사지점의 교통량 수준을 고려하면 ±15% 이내의 오차율이 유의한 추정 범위라고 볼 수 있음
- 추정교통량의 신뢰도 평가지표는 평균절대비율오차(Mean Absolute Percent Error, MAPE, %)를 활용함

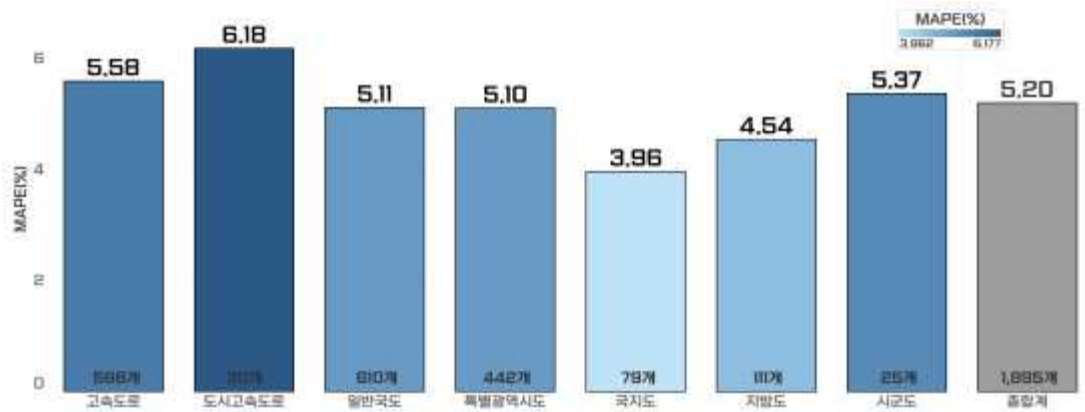
$$MAPE(\%) = \frac{\sum_{i=1}^n \left(\frac{|q_i - \hat{q}_i|}{q_i} \right) \times 100}{n},$$

여기서, q_i = 지점 i 의 조사교통량(대/일)

\hat{q} = 지점 i 의 추정교통량(대/일)

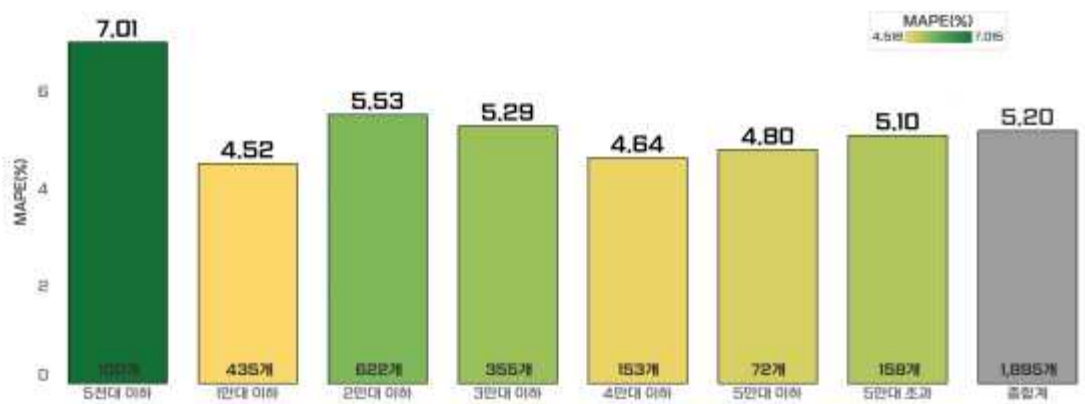
n = 전체 지점 개수(개)

- 도로등급별 오차율(MAPE) 분석 결과, 고속도로 5.6%, 일반국도 5.1%, 특별·광역시도 5.1%로 나타남
 - 전반적으로 약 5% 수준의 오차율을 보이며 ITS 성능평가 기준 상급 수준을 충족함



- ※ 참고1. 평가지점은 ITS 성능평가 기준(1시간 기준 200대 이상인 도로를 대상)으로 선정함
- 참고2. MAPE 목표치는 한국개발연구원의 직접 영향권 범위 참고함
- 참고3. 2024년 추정교통량 기준 분석결과

〈그림 15〉 도로 등급별 신뢰도 분석 결과



- ※ 참고1. 평가지점은 ITS 성능평가 기준(1시간 기준 200대 이상인 도로를 대상)으로 선정함
- 참고2. MAPE 목표치는 한국개발연구원의 직접 영향권 범위 참고함
- 참고3. 2024년 추정교통량 기준 분석결과

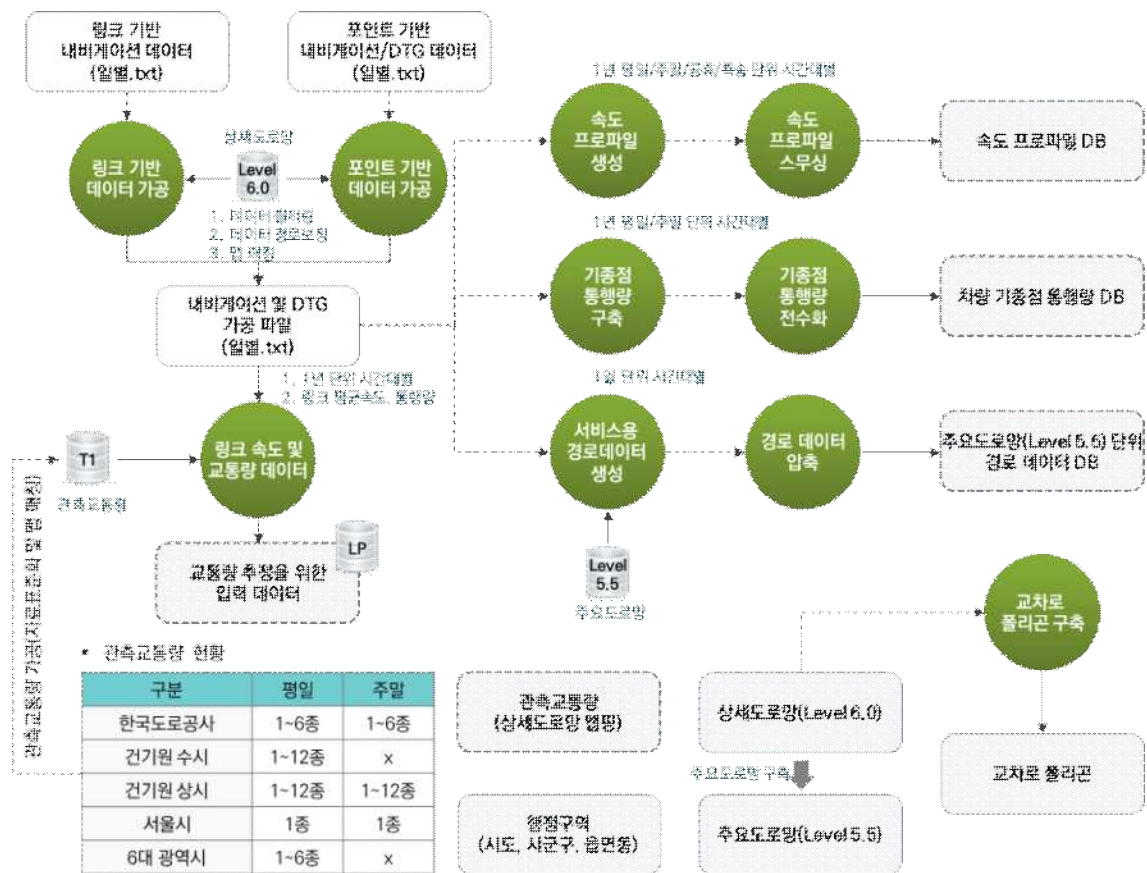
〈그림 16〉 교통량 수준별 신뢰도 분석 결과

라. 기초교통 DB 구축

1) 기초교통 DB 구축 개요

○ 전국 단위 교통량 전수화 및 통행지표 산출을 위해 경로 데이터와 관측교통량을 활용한 1차 가공 기반의 기초교통 DB를 구축함

- 경로 데이터의 속도 정보를 활용하여 시간대별 속도빈도 분포를 구축함
- 차량의 기종점(OD) 기반 통행량 DB를 구축하여 이동 특성을 반영함
- 링크별 통행량 및 속도 정보를 결합하여 교통량 추정을 위한 입력 데이터 DB를 구축함

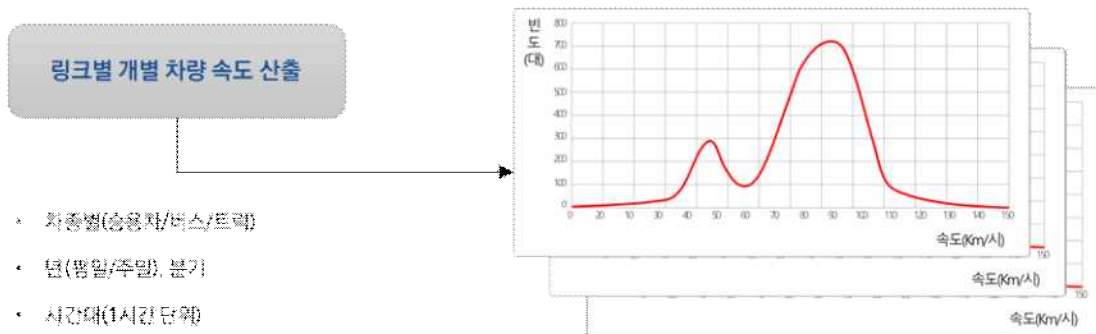


〈그림 17〉 기초교통 DB 구축 프로세스

2) 속도 빈도분포 DB 구축

○ 링크별 속도 빈도분포 집계

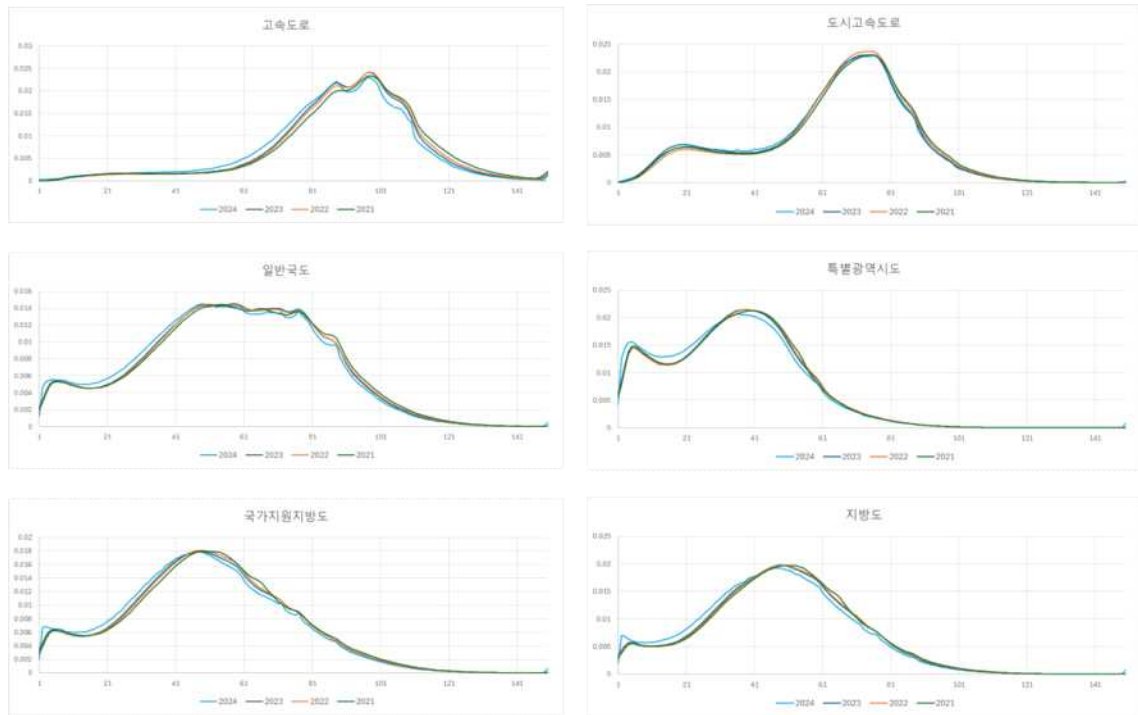
- 교통량 추정 DB와 통행지표 구축을 위하여 속도 빈도분포 DB를 구축
 - 속도지표, 혼잡지표, 환경지표 등 다양한 통행지표 등에 활용
- 링크에 매칭된 궤적 정보를 이용하여 개별차량의 속도 산출 및 산출된 속도 범위에 따른 집계를 수행
 - 링크별로 1km/h 단위의 통행속도(1km/h ~ 150km/h, 초과시 150km/h)의 통행량 집계
- 개별 링크의 연/평일주말/1시간 단위의 속도 빈도분포 구축



〈그림 18〉 속도 빈도분포 구축 프로세스

○ 연도별 도로등급별 속도 빈도분포 비교

- 2021년 ~ 2024년의 도로등급별 속도 빈도분포를 비교분석한 결과 연도별 비슷한 패턴이 유지됨을 확인함
- 2021년 ~ 2023년에 비해 2024년의 도로등급별 속도 빈도분포 데이터는 전체적인 패턴은 유사하나 저속 구간의 비율이 증가한 것으로 확인됨



〈그림 19〉 연도별 도로등급별 속도 빈도분포 결과 비교

○ 도로등급별 평균속도 비교분석

- 2023년 대비 2024년 평균속도는 모든 도로등급에서 약 3.53% 감소함
- 2023년 대비 2024년 평균속도가 가장 크게 감소한 도로등급은 고속도로로 약 8.12%의 감소폭을 보였음

〈표 11〉 연도별 도로등급별 평균속도

단위 : km/h

도로등급	평균속도				'21년 대비 '22년 증감율	'22년 대비 '23년 증감율	'23년 대비 '24년 증감율
	2021년	2022년	2023년	2024년			
고속도로	97.70	95.69	94.03	86.39	-2.07%	-1.73%	-8.12%
도시고속도로	66.34	66.49	65.20	63.33	0.23%	-1.94%	-2.87%
일반국도	60.01	59.51	58.99	57.84	-0.83%	-0.87%	-1.94%
특별광역시도	26.82	26.66	26.32	25.42	-0.59%	-1.28%	-3.41%
국가지원지방도	48.99	48.47	48.01	47.18	-1.06%	-0.94%	-1.72%
지방도	45.33	44.94	44.43	43.48	-0.87%	-1.13%	-2.14%
시군도	31.48	31.12	30.82	29.61	-1.13%	-0.98%	-3.91%
총계	35.97	35.58	35.23	33.99	-1.07%	-1.00%	-3.53%

* 데이터는 절대적으로 참값이 아니며, 지속적으로 신뢰도를 높이기 위한 정산작업 수행

○ 시도별 평균속도 비교분석

- 연도별 평균속도를 분석했을 때 2021년 35.97km/h, 2022년 35.58km/h, 2023년 35.23km/h, 2024년 33.99km/h로 매년 속도가 감소하고 있음
- 2023년 대비 2024년 평균속도가 가장 크게 감소한 지역은 세종으로 약 4.19%, 그 다음 인천에서 약 4.17% 감소하였음
- 나머지 시도에서도 2023년 대비 2024년 평균속도가 감소하여 전체 시도별 평균속도는 3.53% 감소하였음

〈표 12〉 연도별 시도별 평균속도

단위 : km/h

시도	평균속도				'21년 대비 '22년 증감율	'22년 대비 '23년 증감율	'23년 대비 '24년 증감율
	2021년	2022년	2023년	2024년			
서울	24.75	24.80	24.49	23.77	0.21%	-1.25%	-2.94%
부산	29.73	29.54	29.10	28.07	-0.62%	-1.49%	-3.54%
대구	30.44	30.48	31.26	30.22	0.13%	2.54%	-3.32%
인천	30.04	29.73	29.40	28.17	-1.04%	-1.12%	-4.17%
광주	29.29	28.99	28.75	27.86	-1.02%	-0.84%	-3.08%
대전	31.12	30.78	30.33	29.53	-1.11%	-1.44%	-2.63%
울산	33.62	33.30	32.83	31.66	-0.93%	-1.41%	-3.55%
세종	34.73	34.12	33.96	32.54	-1.76%	-0.48%	-4.19%
경기	33.18	32.80	32.47	31.22	-1.17%	-1.01%	-3.82%
강원	39.19	38.75	38.50	37.07	-1.13%	-0.64%	-3.72%
충북	39.32	38.67	38.32	37.01	-1.63%	-0.90%	-3.41%
충남	40.02	39.28	38.94	37.57	-1.85%	-0.88%	-3.52%
전북	39.29	38.88	38.54	37.20	-1.03%	-0.87%	-3.47%
전남	40.87	40.49	40.07	38.90	-0.92%	-1.04%	-2.92%
경북	40.85	40.39	39.82	38.54	-1.13%	-1.41%	-3.22%
경남	37.08	36.77	36.39	35.13	-0.83%	-1.05%	-3.45%
제주	35.13	34.81	34.56	33.42	-0.91%	-0.71%	-3.31%
평균	35.97	35.58	35.23	33.99	-1.07%	-1.00%	-3.53%

* 데이터는 절대적으로 참값이 아니며, 지속적으로 신뢰도를 높이기 위한 정산작업 수행

마. 기종점 통행 DB 구축

1) 기종점 통행DB 구축 개요

- 기종점 통행 DB는 일자 단위 개별 통행 DB 구축과 연단위 집계 DB 구축의 2 단계로 구성됨
 - 일자별 차량별 기종점 통행DB를 구축한 후, 이를 집계하여 차종별 통행시간, 통행거리, 통행속도 등을 포함한 통합 DB로 확장함
- 일자별 개별 차량 기종점 통행 DB 구축
 - 개별 차량의 링크 경로 데이터를 활용하여 차량별 기종점(Origin-Destination) 통행 정보를 생성함
 - 시간적 범위: 일자 단위
 - 공간적 범위: 전국 읍·면·동 간 통행
- 연단위 차종별 기종점 통행 DB 구축
 - 일자별 통행 DB를 기반으로 연단위 집계를 수행하여 차종별 기종점 통행 DB를 구축함
 - 시간적 범위: 연단위(평일/주말/공휴일/특송 구분, 전일 및 시간대별)
 - 공간적 범위: 전국 행정구역(시도/시군구/읍면동) 간 통행

2) 일자별 차량별 기종점 통행 DB 구축

- 경로별 출발 링크와 도착 링크의 읍면동 정보, 노드 정보, 링크종별 정보를 생성함
 - 총 통행시간 = 도착 링크의 진출 시간 - 출발 링크의 진입 시간 (링크종별 16인 휴게소 제외)
 - 총 통행거리 = obu_id 및 group_id별 경로 통행 링크 길이의 합계
 - 통행속도 = 총 통행거리 / 총 통행시간



<그림 20> 개별 차량별 기종점 통행 DB 계산 예시

<표 13> 개별 차량별 기종점 통행 DB 포맷

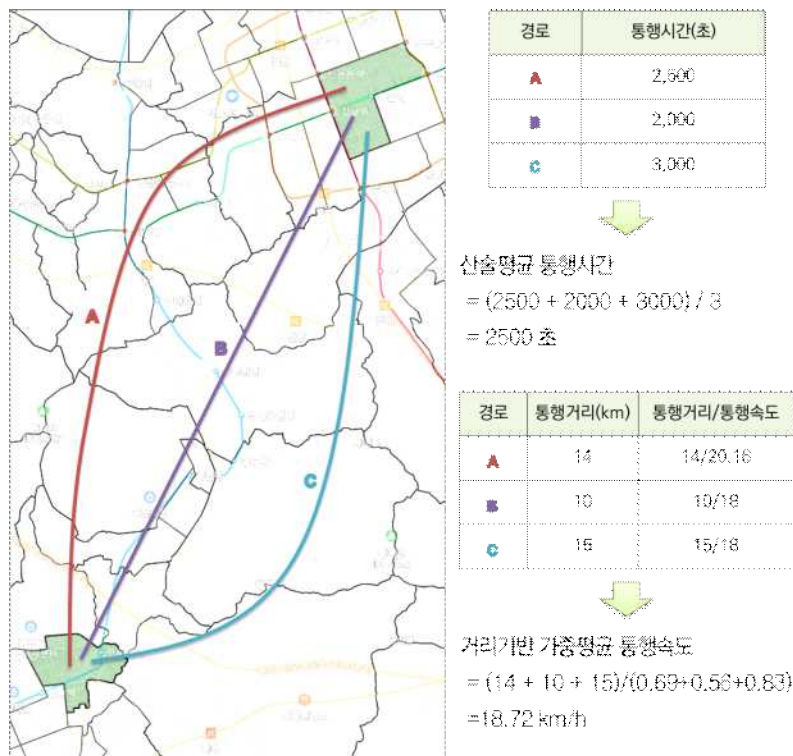
순번	필드명	필드설명	자료형	비고
1	obu_id	개별차량ID	VARCHAR	
2	group_id	통행그룹ID	INT	
3	o_link_id	기점 링크ID	BIGINT	
4	f_node	기점 노드ID	BIGINT	
5	f_node_x	기점 노드 X좌표	DOUBLE	
6	f_node_y	기점 노드 Y좌표	DOUBLE	
7	o_emd_code	기점 읍면동 코드	INT	
8	o_datetime	기점 날짜시간	DATETIME	연월일 시분초
9	o_time	기점 시간(초)	INT	0 ~ 86400
10	d_link_id	종점 링크ID	BIGINT	
11	t_node	종점 노드ID	BIGINT	
12	t_node_x	종점 노드 X좌표	DOUBLE	
13	t_node_y	종점 노드 Y좌표	DOUBLE	

순번	필드명	필드설명	자료형	비고
14	d_emd_code	중점 읍면동 코드	INT	
15	d_time	중점 시간(초)	INT	
16	total_traveltime	총 통행시간	DOUBLE	
17	total_distance	총 통행거리	DOUBLE	
18	avg_spd	평균속도	DOUBLE	

3) 차종별 기종점 통행 DB 구축

○ 경로별 출발 링크와 도착 링크의 행정구역 정보를 생성함

- 산술평균 통행시간 = 총 통행시간 합계 / 개별 차량 경로 수
- 산술평균 통행거리 = 총 통행거리 합계 / 개별 차량 경로 수
- 거리기반 가중 조화평균 통행속도 = 총 통행거리 합계 / '총 통행거리 / 통행 속도'의 합계
- 속도 계산 시, 속도가 있는 링크 정보만 사용함

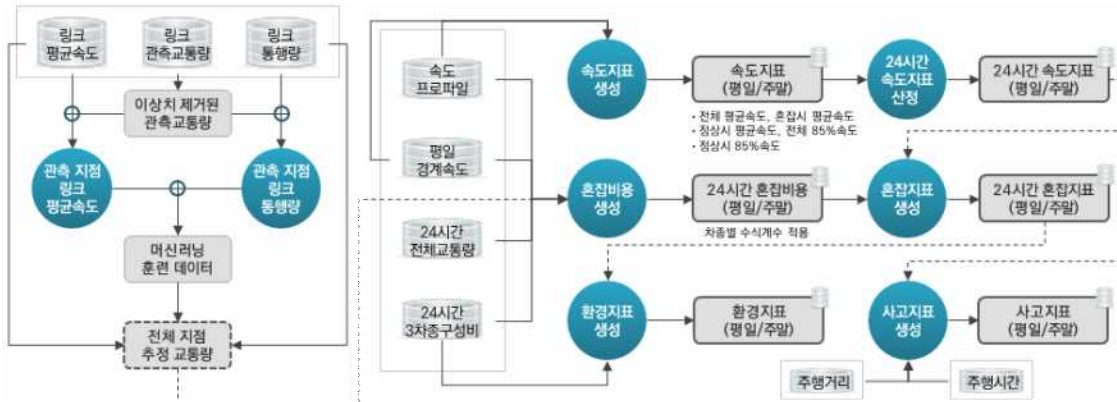


<그림 21> 기종점 통행시간 및 통행속도 DB 계산 예시

바. 차량 교통망 성능평가지표 구축

1) 차량 교통망 성능평가지표 구축 개요

- 차량 교통망 성능평가지표 DB는 1차 가공된 속도 프로파일과 추정된 교통량을 기반으로 전국 혼잡지표, 속도지표, 환경지표, 안전지표를 구축함



〈그림 22〉 교통량 추정 및 차량 통행지표 구축 프로세스

- 교통망 성능평가지표 구축을 위한 기초교통 DB는 다음과 같이 정의된 범위가 존재하여 범위에 따라 지표를 구축함

〈표 14〉 교통망 성능평가지표 DB 생성 기준

데이터명	연	분기	평일 주말	시간	차종	상세 도로망	주요 도로망	행정 구역
속도 빈도분포	○	○	○	○	○	○	○	○
교통량지표	○		○	○	○	○	○	○
속도지표	○		○	○	○	○	○	○
혼잡지표	○		○	○		○	○	○
환경지표	○		○			○	○	○
안전지표	○		○		○	○	○	○

- 차량 교통망 성능평가지표는 도로망의 현황과 성능을 평가하기 위한 지표를 의미하며, 약 17종의 차량 통행지표를 구축

〈표 15〉 차량 교통망 성능평가지표

지표명		지표설명
교통량 지표	관측교통량	공공에서 조사하는 교통량으로 수집기관, 조사기간 등이 상이하여 동일한 기준으로 도로망 네트워크에 매칭하여 나타낸 연평균 일단위 조사교통량
	추정 교통량	특정 시간대 동안 해당 도로구간을 통과하는 차량 대수
	차량 주행거리	추정 교통량 기준의 차량주행거리
속도지표	평균속도	전체 차량의 속도를 평균한 값
	정상시 평균속도	전체 차량 중 교통 혼잡을 경험한 차량들의 평균속도
	혼잡시 평균속도	전체 차량 중 교통 혼잡을 경험하지 않은 차량들의 평균속도
혼잡지표	혼잡시간강도	전체 차량의 총 통행시간 중 교통 혼잡을 경험한 차량들의 총 통행시간 비율
	혼잡빈도강도	전체 차량의 총 통행시간 중 교통 혼잡을 경험한 차량들의 총 통행량 비율
	교통혼잡비용	차량들이 도로상에서 교통혼잡이 발생하여 정상속도 이하로 운행하게 됨으로써 추가로 발생하게 되는 비용
환경지표	이산화탄소배출량	차량들로 주행으로 발생하는 대기오염물질 배출량
	미세먼지배출량	
	일산화탄소배출량	
	휘발성 유기화합물배출량	
	질소산화물배출량	
안전지표	과속비율	과속기준치를 초과하는 차량의 비율
	속도 표준편차	개별 차량들의 속도의 퍼짐 정도
	위험운전행동 이벤트 비율	위험운전행동기준을 개별 차량 속도기준에 따라 위험운전행동이벤트 발생 건수를 비율로 나타냄

2) 차량 교통망 성능평가지표 활용

- 차량 교통망 성능평가지표를 View-T 웹서비스를 통해 '통행지표'와 '데이터 다운로드' 등의 기능에 업로드하여 데이터를 제공하고 있음
- 교통량지표 중 차량주행거리는 통계청의 실험적 통계에서 차량 GPS 기반 도로 유형별 주행거리 통계로 활용되고 있음



〈그림 23〉 통계청의 실험적 통계에 활용되는 차량 교통망 성능평가지표

- 교통혼잡지표 중 교통혼잡비용은 통계청의 e-나라지표에서 도로교통 혼잡비용 통계로 활용되고 있음
- 교통환경지표 중 이산화탄소배출량은 국토교통부 탄소공간지도에서 수송부문 배출량 데이터로 활용되고 있음

사. 회전교통량 추정DB 구축

1) 회전교통량 DB 구축 개요

- 현재 교통량 추정 알고리즘은 회전을 고려하지 않고, 링크별 통행량과 경로 연결성을 고려하여 교통량 추정DB를 구축함
- 회전교통량을 추정 가능성을 검토하기 위해 프로토타입으로 특정 지자체를 설정하여 해당 지자체의 미관측된 교차로 구간에 대한 회전교통량을 추정함
- 스마트 교차로 데이터 수집이 가능한 지역이면서 교차로 구간에 대한 현장조사 데이터 수집가능 일자를 고려하여 세종특별자치시시를 대상으로 1일 회전교통량 추정을 수행하였고, 회전교통량 DB를 구축함
- 프로토타입의 회전교통량 추정DB 구축을 통해 지자체의 스마트 교차로 데이터를 이용한 전국 회전교통량 DB구축을 검토함

2) 회전교통량 DB 구축을 위한 교통 네트워크 활용

- 회전교통량 추정을 위해서는 교차로 단위의 회전방향 정보를 확인할 수 있는 네트워크가 필요함
- 이에 상세도로망(Level6.0)을 기반으로 회전정보를 포함한 교차로 중심의 상위 네트워크를 구축하고, 이를 교통 네트워크라 명명함
 - 교차로/IC/JC 지점에 노드를 생성하고, 생성된 노드와 노드를 연결하여 링크를 구축
 - 이렇게 구축된 노드와 링크를 각각 대표노드, 대표링크라 함
 - 대표노드는 교차로의 중심점을 기준으로 설정하여 모든 접근로가 수렴하는 기준점 역할 하고, 대표링크는 각 접근로와 연결하는 가상의 링크를 구성하여 회전방향을 명시적으로 표현함



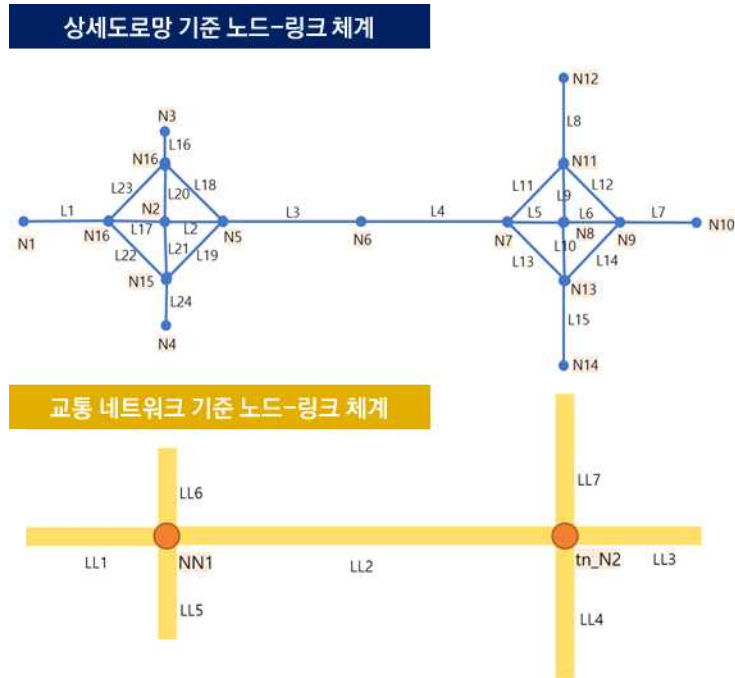
〈그림 24〉 교통 네트워크 개념

- 대표링크 기준의 회전정보는 대표 가상링크 ID, 시작 및 종료 대표노드 ID, 회전정보 코드, 시작 및 종료 상세도로망(Level6.0) 노드 ID, 상세도로망 기준의 링크 경로, 회전 후의 다음 대표 가상링크 ID를 포함함

〈표 16〉 대표링크 기준 회전정보 테이블 정의

No	Column	Type	설명	비고
1	tn_f_node	Integer	시작 대표노드 ID	
2	tn_t_node	Integer	종료 대표노드 ID	
3	tn_turn_code	Text	회전 정보(코드)	L01: 좌회전, S01: 직진 R01: 우회전, U01: 유턴
4	f_node6	Integer	시작 상세 노드 ID	
5	t_node6	Integer	종료 상세 노드 ID	
6	path	Text	상세도로망 링크 경로	
7	nxt_vlink	Text	다음 대표 가상링크 ID	진출 대표링크
8	tn_v_link_id	Integer	대표 가상링크 ID	진입 대표링크

- 연속된 2개의 교차로를 대상으로 상세도로망 기준의 노드-링크 체계와 교통 네트워크 기준의 대표노드-대표링크 체계는 다음과 같이 도식화할 수 있음



<그림 25> 네트워크별 노드-링크 체계(도식화)

- 상기 기준을 토대로 교차로 회전교통량 추정을 위한 분석 기반을 마련하기 위해 기존 상세도로망 네트워크를 고도화한 세종시 범위의 교통 네트워크를 구축함



<그림 26> 세종특별자치시 교통 네트워크

3) 회전교통량 추정을 위한 학습데이터 구축

- 교차로 회전교통량 추정모형을 구축하기 위해 관측 회전교통량, 차량 GPS 데이터, 교차로 기하구조 정보를 주요 학습데이터로 활용하고, 이를 통해 교차로 별 통행 패턴의 물리적 제약요인을 반영할 수 있도록 함
 - 관측 회전교통량
 - 세종시 스마트교차로에서 수집된 데이터
 - 추가 현장조사를 통해 교차로별 회전방향(좌회전·직진·우회전)별 교통량
 - 차량 GPS 데이터: 내비게이션 기반의 주행 궤적을 활용하여, 링크 단위의 probe 통행량, 평균속도, 교통밀도, 네트워크 연결성 변수 산출
 - 네트워크 연결성은 차량 GPS의 연속 통행 패턴을 분석하여 링크 간 관계를 수치화 함
 - 교차로 기하구조 데이터: 각 교차로의 차로 수, 도로 연장, 도로등급 등 물리적 구조 정보를 기반으로 구축
- 시간적 집계 단위는 5분, 15분, 60분으로 구분하여 학습데이터를 구축
 - 시간 세분화 수준에 따른 추정 결과의 신뢰도와 모형의 민감도를 비교·분석하기 위함
- 학습 데이터 구성 시, 교통량 수준이 높은 스마트교차로와 상대적으로 낮은 일반 교차로의 데이터를 함께 포함하여, 도시 전역을 대상으로 한 모형의 강건성과 일반화 성능을 확보함
 - 다만, 심야시간대는 교통량이 적어 신뢰도 확보가 어려워 제외하였으며, U턴은 발생 빈도가 낮고 데이터 불균형이 심해 분석대상에서 제외함

〈표 17〉 회전교통량 추정을 위한 주요 학습데이터

분류	활용 데이터	집계 단위
관측 회전교통량	스마트교차로 CCTV 수집 회전교통량	5분, 15분, 60분
	현장 조사 수집 회전교통량	
차량 GPS 데이터	내비게이션 GPS 궤적 데이터 기반 회전교통량(probe)	
	링크 단위 평균속도, 밀도	
	네트워크 연결성	
교차로 기하구조	차로 수, 도로 연장, 도로등급 등	

- 교통량 추정모형 학습을 위해 교통네트워크 가상링크 ID, 회전 ID, 일자, 시간 코드를 기준으로 관측교통량, 프로브 통행량·속도, 네트워크 연결성, 교차로 기하구조 데이터를 통합하여 학습데이터로 구축함
 - 학습데이터는 교통의 시공간적 특성과 물리적 요인을 함께 반영하도록 구성
 - 교통량 대비 프로브 대수는 변동성이 높아 학습 효율과 예측 성능 향상을 위해 민맥스 스케일링 기법으로 정규화하여 학습의 안정성과 균형을 확보함

〈표 18〉 세종특별자치시 교통량 추정 학습데이터 테이블 정의

No	Column	Type	설명	모형 사용 유무	비고
1	tn_v_link_id	Integer	교통 네트워크 가상링크ID		교통네트워크 단위
2	turn_id	Text	회전방향 ID	○	S01: 직진 R01:우회전 L01:좌회전 U01:유턴
3	date	Text	일자		yyyyMMdd
4	time_code	Integer	시간코드	○	0~287: 5분 단위 0~95: 15분 단위 0~23: 60분 단위
5	total_probe	Real	프로브		대
6	auto_spd	Real	속도	○	km/h
7	length	Real	대표 가상링크 길이	○	km
8	road_rank	Integer	대표 가상링크 도로등급	○	101 : 고속도로 102 : 도시고속도로 103 : 일반국도 104 : 특별광역시도 105 : 국가지원지방도 106 : 지방도 107: 시군도 108: 연결램프
9	lanes	Integer	대표 가상링크 차선수	○	
10	k	Real	대표 가상링크 밀도	○	프로브 / 속도
11	emd_type	Integer	대표 가상링크 읍면동 여부		0: 동 아님 1: 동 맞음
12	sido_id	Integer	대표 가상링크 시도ID		

No	Column	Type	설명	모형 사용 유무	비고
13	traffic_type	Text	조사기관		
14	total	Real	관측교통량	O(라벨)	대
15	total_scale	Real	프로브 민맥스 스케일링값	O	0~1 스케일링값
16	total_1	Real	연결된 관측교통량_1	O	대
17	total_scale_1	Real	연결된 프로브 민맥스 스케일링값_1	O	0~1 스케일링값
18	total_2	Real	연결된 관측교통량_2	O	대
19	total_scale_2	Real	연결된 프로브 민맥스 스케일링값_2	O	0~1 스케일링값
20	total_3	Real	연결된 관측교통량_3	O	대
21	total_scale_3	Real	연결된 프로브 민맥스 스케일링값_3	O	0~1 스케일링값

4) 회전교통량 추정모형 개발

- 회전교통량 관측자료가 존재하지 않는 미관측 교차로를 대상으로 직진, 좌회전, 우회전 방향별 교통량을 추정하기 위한 LightGBM 모델 기반 머신러닝 모형을 개발함
- 회전교통량 추정에 가장 적합한 모델을 선정하기 위하여 동일한 분석 환경 기준 Ridge, Elastic Net, Random Forest, Extra Trees, XGBoost, CatBoost, LightGBM 모델을 대상으로 추정 성능 평가를 진행하였고, LightGBM이 5분, 15분, 60분 모두에서 가장 우수한 추정 정확도를 보여 선정함
- 데이터 전처리 과정에서는 범주형 변수를 모형이 인식할 수 있도록 원-핫(One-Hot Encoding)을 적용하였으며, 전체 학습 파이프라인을 자동화하여 데이터 입력에서 예측 산출까지 일관된 절차로 수행되도록 구성함
- 또한 교차검증(K-Fold Cross Validation) 기법을 적용하여 예측값을 반복 학습함으로써 과적합을 방지하고 일반화 성능을 강화함
- 하이퍼 파라미터 최적화를 통해 추정 모델의 성능 및 일반화 성능을 확보함

4) 회전교통량 추정 구축결과 분석

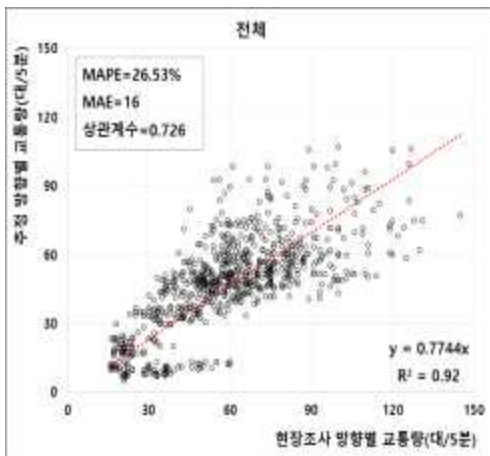
- 수집 주기별 회전교통량 추정 결과 분석을 통해 집계 단위가 확대될수록 상대 오차가 감소하고 설명력이 향상됨
 - (5분) MAPE 26.53%, MAE 16대, 상관계수 0.726, $R^2=0.92$
 - (15분) MAPE 22.57%, MAE 36대, 상관계수 0.836, $R^2=0.9424$
 - (60분) MAPE 13.76%, MAE 80대, 상관계수 0.916, $R^2=0.9772$
- 상관계수, R^2 , 산점도상 회귀선 기울기의 분석에서도 집계 단위의 확장에 따라 안정적인 추정 결과를 확인할 수 있음
 - 상관계수는 (5분) 0.726 (15분) 0.836 (60분) 0.916로 나타나 집계 단위 확대에 따라 추정치-실측치 간 선형 관계가 강화됨을 확인함
 - R^2 는 (5분) 0.92 (15분) 0.94 (60분) 0.98로 나타나 집계 단위 확대에 따른 설명력을 확보하고 특히 60분 기준 실측 변동의 약 98%를 설명하는 높은 예측 일관성을 보임
 - 회귀선 기울기는 (5분) 0.7744x (15분) 0.8595x (60분) 0.9967x로 집계 단위 확장에 따라 1에 수렴하는 경향성을 보임
- 방향별로 분석한 결과에서도 집계 단위의 확대에 따른 추정 오차율 감소의 경향성이 공통적으로 나타남
 - 좌회전은 MAPE 기준 (5분) 33.87%, (15분) 27.68%, (60분) 25.72%로 집계 단위의 확대에 따라 오차율이 개선되나 가장 높은 오차율 수준을 보임
 - 직진은 시간 집계 단위가 확장될수록 오차율 및 MAE의 개선 폭이 가장 크며, MAPE 기준 (5분) 25.13%, (15분) 21.91%, (60분) 8.94%로 나타나며, 60분 기준 MAPE 10% 미만으로 실제 활용 가능 수준의 정확도를 보임
 - 우회전은 MAPE 기준 (5분) 21.02%, (15분) 15.72%, (60분) 11.81%로 5분, 15분 단위에서 가장 낮은 오차율을 보이고, 60분 기준에서도 MAPE 11.81%로 안정적인 성능을 확보함
- 수집 주기 단위의 확대에 따라 추정의 정확도가 전반적으로 개선됨을 확인하였

으며 이후 교통량 모니터링·계획 등 실제 활용 수준의 분석 시에는 15분 또는 60분 기준의 추정 결과를 활용하여 안정적 정확도를 확보가 필요함

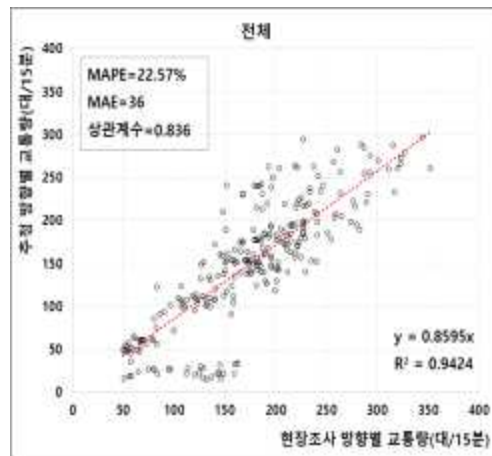
〈표 19〉 수집 주기별 방향별 회전교통량 추정 결과

수집 주기별 방향별 회전교통량 추정 결과						
구분	5분		15분		60분	
	MAPE(%)	MAE(대)	MAPE(%)	MAE(대)	MAPE(%)	MAE(대)
좌회전	33.87	16	27.68	32	25.72	112
직진	25.13	16	21.91	38	8.94	65
우회전	21.02	15	15.72	35	11.81	82
전체(교차로)	26.53	16	22.57	36	13.76	79

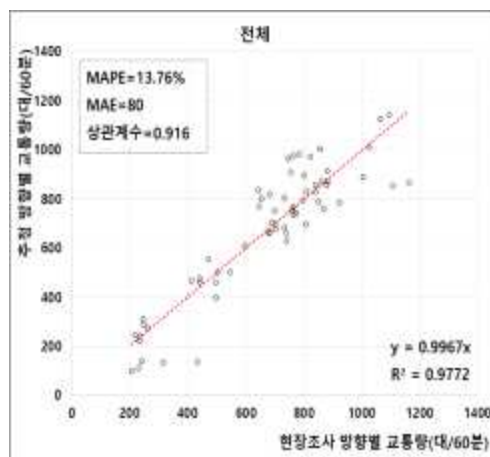
수집 주기 5분 - 분석 결과



수집 주기 15분 - 분석 결과



수집 주기 60분 - 분석 결과



4. 차량 GPS 데이터 구축·분석시스템 개선

가. 차량 데이터 구축·분석 시스템 개요

- 데이터 수집과 가공 각 단계를 모니터링하여 데이터가 정상적으로 처리되고 있는지 실시간으로 확인할 수 있는 서비스 지원
- KTDB 기반 지도 및 차량 통행 데이터셋을 체계적으로 관리하고 분석에 활용할 수 있는 시스템 구축
- 데이터셋 가공·구축 시스템 관리
 - 서버 상태 모니터링 환경 구축을 통해 실시간 상태 확인 및 장애 예방 가능
 - 데이터 가공 모듈 개선 시 서버 업데이트 및 개발 정보 지속 관리
- 분석 시스템 관리
 - 수집 현황과 데이터 전처리 상태를 제공하기 위한 대시보드 서비스 제공
 - 공간 단위별로 구축된 교통망 성능 평가 지표 및 통행 분석 DB 업로드 및 버전 관리 지원
 - 데이터 작업 현황에 따라 데이터베이스(MySQL) 저장 및 관리

나. 데이터 가공 모니터링 기능 개발

- 원시 데이터 수집 및 데이터 전처리에서 이상치 제거, 정렬, 좌표 변환, 시간 정보 변환 등의 전처리 과정을 각 단계별로 수행하며, 일자별 오류 유형, 가공 용량, OBU 등 수집된 로그 기반으로 모니터링 기능 개발

1) 기능 구성

① 데이터 수집 현황 분석기능

- 원시 데이터의 파일 유효성 검사 정보를 테이블 및 그래프로 제공

- 주요 항목 : 파일 수, 오류 파일 수, 총용량 등
- 년/월/일 별 데이터 용량, 건수를 시각적으로 표현

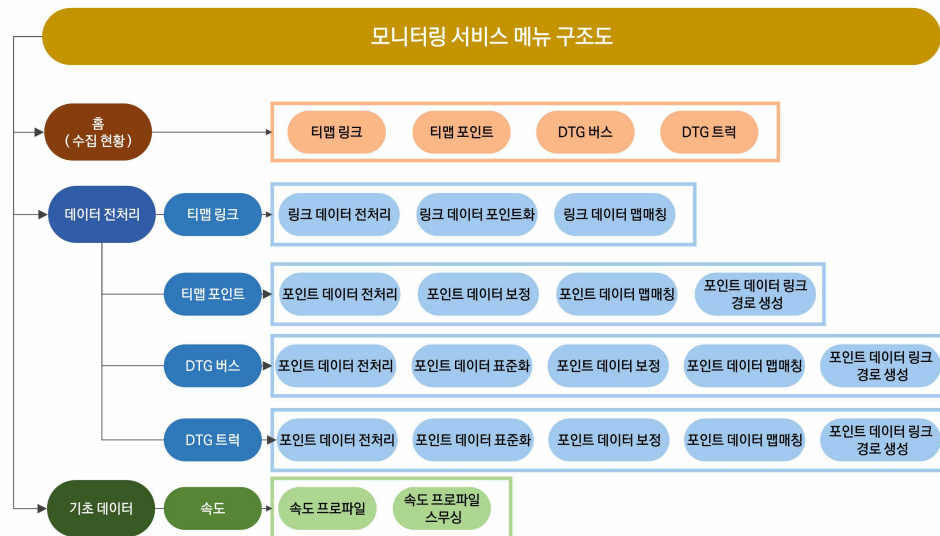
② 데이터 전처리 현황 분석기능

- 이상치 제거, 정렬, 좌표 변환, 시간 정보 변환 등의 데이터 전처리 작업에서 발생하는 로그 정보를 테이블 및 그래프로 제공

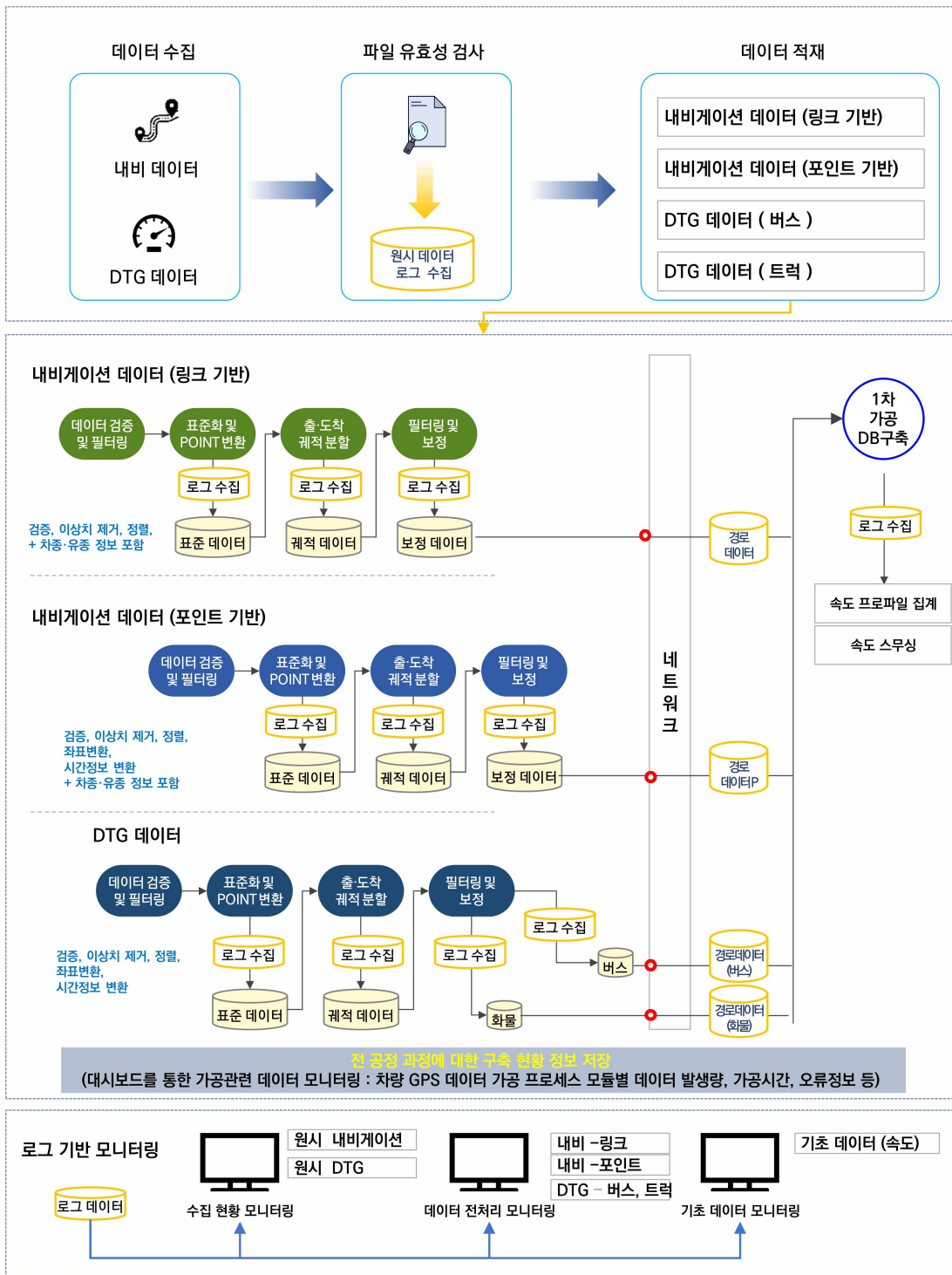
- 주요 항목 : 작업별 처리 데이터 수, 처리 성공 및 오류 수 등
- 전처리 과정에서 데이터의 상태, 정보 등 각 전처리 과정에 맞는 정보 시각적으로 표현

③ 기초 데이터 현황 분석기능

- 전처리 데이터와 속도 프로브 데이터에서 생성되는 기초 데이터의 로그 정보를 테이블 및 그래프로 제공



<그림 27> 모니터링 메뉴 구조도



〈그림 28〉 모니터링 프로세스

2) 기능 구축 결과

- 선택한 년도, 기간에 따라 이벤트 건수, 파일 수, 용량 및 데이터 현황 등을 테이블을 통해 사용자는 데이터를 구체적으로 확인할 수 있도록 함



〈그림 29〉 데이터 수집 현황 모니터링 화면

- 선택한 표출 컬럼에 따라 테이블 데이터가 구성되며, 각 전처리 단계에서 발생하는 수집된 로그 정보를 자세하게 확인할 수 있도록 함



〈그림 30〉 데이터 전처리 모니터링 화면

다. 데이터 분석을 위한 신규 기능 개발

1) 교통지표 우선순위 분석 기능 개발

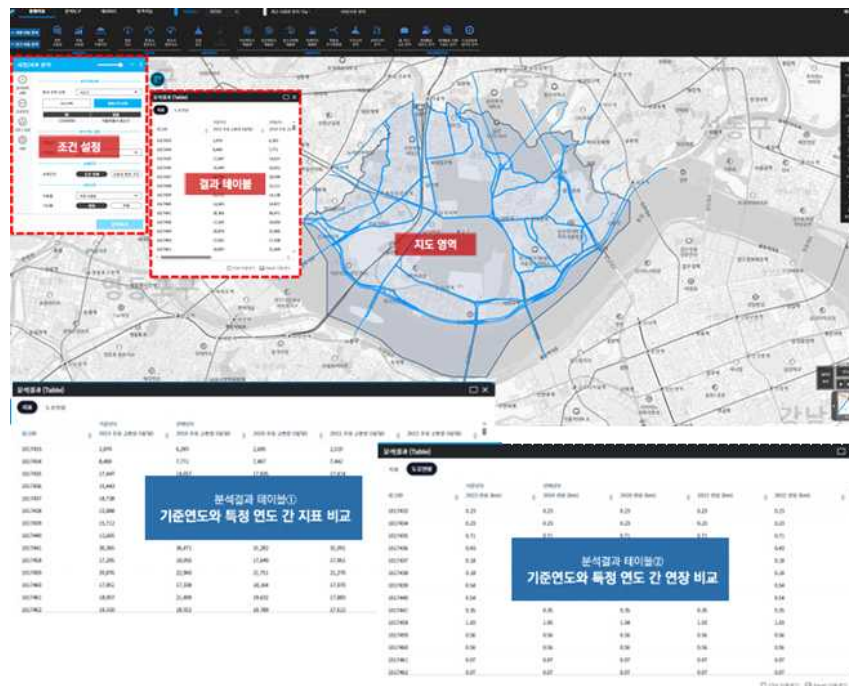
- (기능 목적) 교통SOC의 진단 및 평가를 위한 교통망 성능평가지표 DB 구축 결과를 지역별로 한눈에 순위별로 시각화하여 분석할 수 있도록 개발함
- 조건 설정
 - 분석지역 선택: 시도 또는 시군구 단위 선택 가능하도록 함
 - 기간/시간 선택: 분석 지표에 대한 기간과 시간을 선택 가능하도록 함
 - 지표 선택
 - 지표별(주요): 분석 기준이 되는 대표지표 1개를 선택 가능하도록 함
 - 지표별(보조): 주요지표와 연관되는 복수개 지표를 선택 가능하도록 함
 - 도로조건 선택: 선택한 분석지역 내 도로등급, 차선수, 링크유형 조건으로 2차 필터링 할 수 있도록 함



〈그림 31〉 교통지표 우선순위 분석 예시 화면

2) 연도별 교통지표 변화분석 기능 개발

- (기능 목적) 교통정책 효과 분석을 위해 교통망성능평가지표를 연도별로 변화된 결과를 한눈에 시각화하여 분석할 수 있도록 개발함
- 조건 설정
 - 분석영역 선택: 행정구역 단위 선택 또는 사용자 선택 영역
 - 분석연도 선택: 기준연도 선택, 기준연도를 기준으로 분석하고 싶은 연도 선택
 - 분석연도에 대한 상세조건 선택
 - 기준연도만 선택 시: '전체 도로', '신성 및 변경 구간' 중 택 1
 - 선택연도 선택 시: '전체 도로', '신설 및 변경 구간 포함' 중 택 1
 - 분석지표 선택: 분석지표 및 지표에 대한 기간(평일/주말, 시간대) 선택
 - 비교 분석(사용자 정의): 비교 분석은 기준연도에서 선택연도에 대한 증감률 정보 제공(디폴트), 사용자가 수식을 적용하고 결과를 확인할 수 있도록 함
 - 결과: 설정에 의한 최종 결과는 테이블로 표출



〈그림 32〉 연도별 교통지표 변화 분석 시각화



제1장 과업의 개요

제1절 과업의 배경 및 목적

제2절 과업의 범위 및 내용

제1장 과업의 개요

제1절 과업의 배경 및 목적

1. 과업의 배경

- 전국 단위 공공의 기초 교통데이터 수집 및 지점 데이터 분석의 한계
 - 교통 분야의 기초 데이터인 교통량, 속도 데이터는 한국건설기술연구원, 한국도로공사, 광역지자체 등에서 조사하고 있으나, 데이터의 공간적 수집 범위가 매우 낮아 전국의 교통 현황을 분석하기에 한계가 있음
 - 공공에서 수집하고 있는 데이터는 지점 정보 중심의 데이터로 단편적 정보만 확인할 수 있어 시공간적으로 연결된 교통의 흐름을 파악하기엔 한계가 있음
- 전국 단위 일관된 교통DB 및 통행지표 부재
 - 데이터 수집기관, 수집방식 등 기관별로 상이하게 수집되는 데이터 수집 체계는 전국을 일관된 기준으로 분석하기엔 한계가 있음
- 빅데이터 전처리·가공 환경 및 빅데이터 기반 차량통행 분석시스템 부재
 - 차량 GPS 빅데이터를 효율적으로 전처리·가공하기 위한 환경이 부족하며, 과거에는 차량의 시·공간적 통행패턴과 현황을 분석하기 위해 모형 중심의 프로그램을 이용하여 분석을 수행하였음
- 국가통합교통체계효율화법 시행령 제8조1)에 의거하여 교통물류활동으로 발생

1) 국가교통조사에는 다음 각 호 사항이 포함되어야 한다 - 교통물류활동으로 발생하는 교통혼잡, 교통사고, 환경오염, 온실가스 배출 등 교통 관련 사회적 비용, 그 밖에 교통 관련 정책 및 계획의 수립, 교통시설 투자분석 및 평가에 필요한 사항

하는 교통혼잡 등 교통 관련 사회적 외부비용을 국가교통물류에 관한 지표를 설정하여 고시하도록 규정하고 있음

2. 과업의 목적

- 국가통합교통체계효율화법²⁾에 의거 국가교통데이터베이스 구축 및 운영을 위한 기초교통DB 및 통행지표 생성
 - 개별 차량 기반의 교통량, 속도, 차량주행거리 등 교통분야의 핵심 기초교통 DB 구축(공공 조사교통량 약 1만 개 지점, 전국 도로 62만 개 링크)
 - 차량 이용 및 교통물류 활동으로 인한 교통혼잡, 환경오염, 온실가스 배출 등 다양한 측면의 도로 교통망 성능평가 지표 구축
 - 지자체별로 상이한 방식으로 수집되고 구축되는 교통DB·통행지표의 한계를 개선하기 위한 전국단위 일관된 교통DB·통행지표 구축 필요
 - 여객 O/D구축 및 교통수요예측을 위한 기반데이터(승용차 O/D 데이터 등) 제공
- 국가교통 데이터 제공 및 교통 분야 주요 정책 지원
 - 이용자의 요구에 대응할 수 있는 데이터 제공 및 분석지원
 - 탄소공간지도 사업, 환경소음측정망 선정 사업, 지자체 실증사업 등 주요 정책 제언을 위한 데이터 및 분석 지원(매년 요청 및 제공)

2) 국가통합교통체계효율화법 제12조(국가교통조사), 제14조(정보통신수단 등을 통한 교통조사) 및 동법 시행령 제8조(국가교통조사의 실시)

제2절 과업의 범위 및 내용

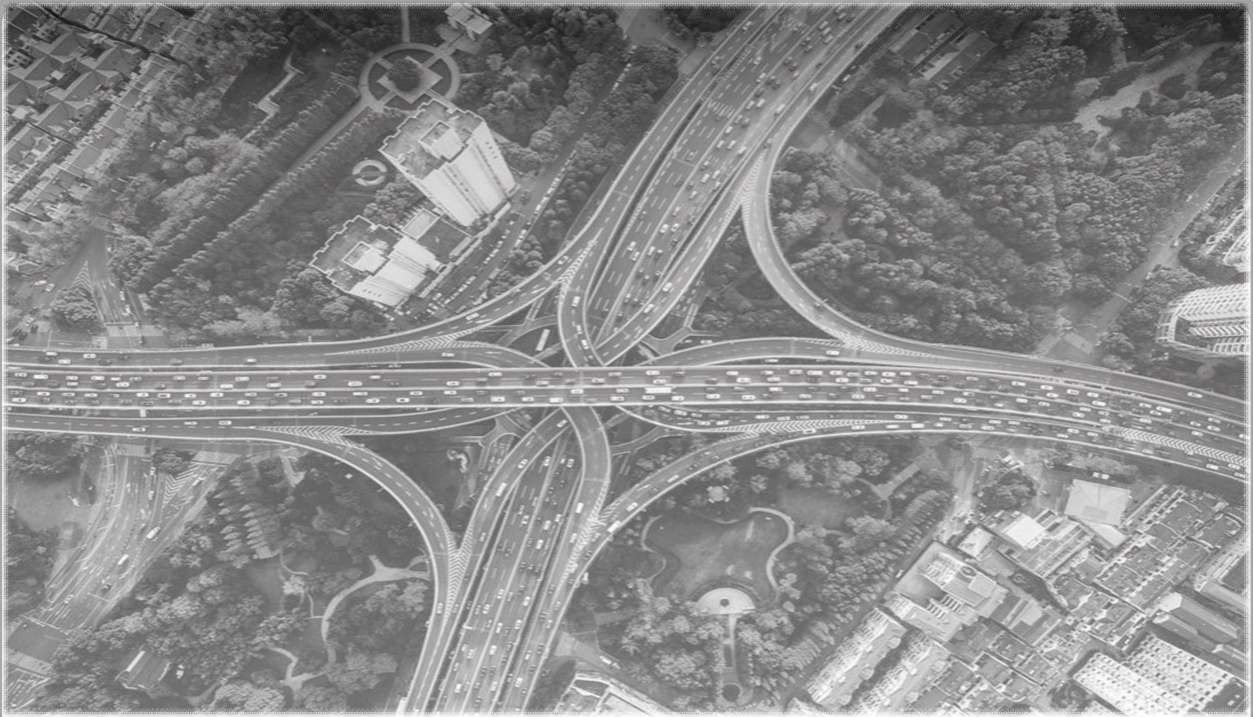
1. 과업의 범위

- 공간적 범위
 - 전국 2차로 이상 도로 및 주요 도로³⁾
- 시간적 범위
 - 데이터 수집 : 2024년
 - 데이터 가공 및 DB구축 : 2024년

2. 과업의 내용

- 차량 원시 GPS 빅데이터 전처리 및 DB가공·구축
 - 차량 GPS 데이터, DTG 데이터 수집 및 검토
 - 원시 데이터 전처리 및 경로 데이터 가공
 - 기초 교통 DB구축 및 검증
- 교통량 추정
 - 네트워크 연결성을 고려한 교통량 추정
 - 프로토타입 회전교통량 추정
 - 기초 교통DB 구축을 위한 교통량 추정 신뢰도 분석 및 연도별 패턴 분석
- 차량 통행지표 가공 알고리즘 개선 및 DB구축
 - 통행지표 가공 알고리즘 개선 및 고도화
 - 목적별 전국 통행지표 산출 및 DB 구축

3) 주요 도로 : 편도 1차로 도로 중 데이터 수집의 안정성이 확보된 도로



제2장 차량 GPS 빅데이터 수집 및 분석

제1절 차량 GPS 빅데이터 수집 개요

제2절 차량 GPS 원시데이터 분석

제2장 차량 GPS 빅데이터 수집 및 분석

제1절 차량 GPS 빅데이터 수집 개요

- 차량 GPS 데이터는 차량의 이동경로에 대해 시간 연속성과 공간 연결성이 동시에 수집되는 데이터로 차량의 위도, 경도, 고도 등 개별 차량의 속성 정보를 포함하고 있음
- 본 사업에서 수집한 차량 GPS 데이터는 티맵의 내비게이션 데이터⁴⁾와 한국교통안전공단의 DTG데이터이며, 데이터 수집 기관마다 활용되는 단말기 차이로 인해 데이터 수집 방식, 수집 주기 등 수집되는 속성 정보의 차이가 있음
 - 내비게이션 데이터는 경로탐색 시 차량의 운행정보가 생성되어 시간, 위치, 속도 정보를 포함하고 있음
 - DTG 데이터는 차량운행기록계를 통해 사업용 차량의 운행정보(시간, 위치, 속도, 가속도 등)를 기록하며, 대상은 버스·화물·택시로 구분됨
- 차량 GPS 데이터는 도로속성 정보가 연결된 링크 기반 데이터와 X,Y 위치 좌표가 수집되는 포인트 기반 데이터로 구분됨
 - 내비게이션 데이터는 링크 기반 데이터와 포인트 기반 데이터로 구분됨
 - 링크 기반 내비게이션 데이터는 포인트 기반으로 구축된 정보를 도로속성 정보를 포함하는 맵 링크ID로 매칭되어 제공되므로 위치좌표마다 생성된 가속도, 경사도 등 정보가 누락되어 수집되었음

4) 내비게이션 데이터는 데이터 신뢰도를 높이기 위해 2022년 사업부터 샘플율이 높은 티맵 내비게이션 데이터를 수집하여 2020년 데이터부터 원천 데이터가 변경됨

〈표 2-1〉 차량 GPS 데이터 특징

구분	링크 기반 내비게이션 데이터	포인트 기반 내비게이션 데이터	DTG 데이터
주행 ID 생성기준	경로 탐색 시 ID 생성	경로 탐색 시 ID 생성	차량등록번호와 주행시작일로 ID 생성
속성 정보 생성기준	내비게이션 경로 탐색 시	내비게이션 경로 탐색 시	차량 운행 시
데이터 형태	SK맵 링크 단위	1초 ~ 3초 단위 포인트	1초 단위 포인트
좌표체계	Korea 2000	Korea 2000	WGS84

1. 차량 GPS 빅데이터 수집

가. 데이터 수집 대상

- 본 과업은 2024년을 기준연도로 설정하고, 차량 GPS 빅데이터 기반 기초교통 및 교통망 성능평가지표 DB 구축을 위해 다음과 같은 데이터를 수집함
 - 링크 기반 내비게이션 데이터는 티맵 이용자의 주행 정보를 도로구간(링크) 단위로 구조화한 도로 네트워크 데이터를 수집함
 - 포인트 기반 내비게이션 데이터는 티맵 이용자의 주행 정보를 GPS 단위로 수집하며, 초 단위 시계열 데이터로 구축함
 - 금년도 사업부터 포인트 기반 티맵 데이터를 최초로 도입하였으며, 수집 과정에서 발생한 오류 데이터를 재요청하여 순차적으로 보완·수집함
 - 포인트 기반 사업용 차량 운행기록계(DTG) 데이터는 한국교통안전공단에서 제공받으며, 버스·화물·택시 등 사업용 차량의 주행 정보를 수집함

〈표 2-2〉 데이터 수집 목록 및 기준연도

구분	데이터 목록	수집처	수집연도
링크 기반	내비게이션 도로 네트워크	티맵	2024년 1-12월, 2025년 상반기
	내비게이션 주행 정보		
포인트 기반	내비게이션 주행 GPS 정보	티맵	2024년 6, 7, 10, 12월
	사업용 차량 운행기록계 버스 주행 정보	한국교통안 전공단	2024년 1-12월
	사업용 차량 운행기록계 화물 주행 정보		
사업용 차량 운행기록계 택시 주행 정보			

- 시의성 있는 데이터 제공을 위해 당해연도(2025년) 상반기 데이터를 선제적으로 수집하여, 신속한 가공·분석을 통해 활용 가능한 지표를 조기 산출함
 - 링크 기반 내비게이션 데이터는 2024년 및 2025년 12주, 명절 기간을 포함하여 월별 및 계절별 추세를 파악할 수 있도록 총 185일을 수집하였으며, 최신성 확보를 위해 2025년 7월 데이터까지 확보함
 - 포인트 기반 내비게이션 데이터는 2024년 기준 4주와 여름·겨울 계절 특성을 반영할 수 있는 총 31일을 대상으로 수집하였으며, 수집 과정에서 데이터 오류가 확인되어 일부 기간에 대해 재요청을 수행함
 - 특히 포인트 기반 데이터는 초기 계획 대비 수집 시기가 변경되었으며, 데이터 제공기관의 서버 장애로 1~5월 데이터가 삭제됨에 따라, 대체 가능한 기간인 2024년 6월, 7월, 10월, 12월 데이터로 조정하여 수집을 완료함

〈표 2-3〉 수집 데이터의 시간적 범위

구분	링크 기반 내비게이션		포인트 기반 내비게이션
	2024년	2025년	2024년
1월	22일 ~ 28일 (7일)	20일 ~ 26일 (7일)	-
2월	13일 ~ 18일 (7일)	10일 ~ 16일 (7일)	-
3월	11일 ~ 17일 (7일)	10일 ~ 16일 (7일)	-
4월	15일 ~ 21일 (7일)	14일 ~ 20일 (7일)	-

6월	10일 ~ 16일 (7일)	16일 ~ 22일(7일)	17일 ~ 23일(7일)
7월	15일 ~ 21일 (7일)	21일 ~ 27일(7일)	8일 ~ 14일(7일)

9월	19일 ~ 25일 (7일)	15일 ~ 21일 (7일)	-
10월	14일 ~ 20일 (7일)	13일 ~ 19일 (7일)	21일 ~ 27일(7일)
11월	11일 ~ 17일 (7일)	10일 ~ 16일 (7일)	-
12월	16일 ~ 22일 (7일)	15일 ~ 21일 (7일)	9일 ~ 15일(7일)
명절	설날 : 2월 09일 ~ 2월 12일 (4일) 추석 : 9월 16일 ~ 9월 18일 (3일)	설날 : 1월 27일 ~ 1월 30일 (4일) 추석 : 10월 03일 ~ 10월 09일 (7일)	추석 : 9월 16일 ~ 9월 18일 (3일)
합계	90일	95일	31일

나. 데이터 수집 대상의 특징 및 형식

1) 차량 내비게이션 데이터

- 링크 기반 내비게이션 데이터는 링크 단위의 진입·진출 시간, 속도 등 주행 정보와 함께 차종, 유종, 목적지 정보가 포함되어 제공됨
- 해당 데이터는 티맵에서 제공하는 도로 네트워크와의 맵매칭(Map-Matching)을 통해 실제 주행 경로 및 위치를 해석할 수 있으며, 도로 네트워크의 링크 ID는 특정 시점마다 갱신되므로 데이터 수집 기간별로 상이한 네트워크 체계를 관리·적용해야 함
- 주행 ID는 사용자가 목적지를 설정하고 경로 탐색을 수행할 때 생성되는 단위 식별자로, 동일 사용자라 하더라도 주행 단위마다 별도의 ID가 부여됨에 따라 개인 단위의 연속적인 이동경로(Trip Chain) 추적에는 한계가 존재함
- 주행 중 휴게소 이용 또는 신호대기 등 정차 상태는 연속된 주행 데이터 내에서 속도 정보로 식별 가능하나, 내비게이션을 종료한 경우 이후 구간에 대한 경로는 생성되지 않으며, 별도의 보정 또는 추정 처리는 수행되지 않음
- 출발지명(origin)이 NULL로 수집되는 경우는 GPS 좌표 정보가 존재함에도 불구하고 해당 위치가 법정동 경계와 매칭되지 않는 사례로, 사용자 단말기에서 서버로 데이터 전송 시 행정구역 코드 매핑 과정에서 발생하는 오류로 판단됨
- 차종 정보는 '운전습관 가입자' 기반으로 수집되며, 이는 사용자 설정에 의존하는 속성으로 전체 주행 데이터 중 약 70~80% 수준에서만 제공되는 것으로 확인됨
- 목적지 정보(destination)는 사용자가 경로 탐색 시 입력한 텍스트 기반 정보로 구성되며, 건물명, 상업시설, 공공기관, 교통시설 등 다양한 형태로 존재함. 주소 기반 검색 시에는 목적지 중분류 정보가 NULL로 수집되는 경우가 일부 존재함
- 목적지 중분류 정보는 티맵 POI(Point of Interest) DB와 연계되어 제공되며, 공공기관, 교육시설, 의료시설, 교통시설 등으로 세분화된 분류체계를 포함하고 있어 주행 목적(통근, 업무, 여가 등)을 간접적으로 추정하는 데 활용 가능함
- 다만, 장거리 이동 과정에서 휴게소 정차, 중간 경유지 추가, 내비게이션 재탐색 등으로 주행이 분절될 경우 새로운 주행 ID가 생성되며, 이로 인해 동일 목적지에 대한 정보가 중복 기록될 수 있음. 따라서 목적지 기반 통행 목적 분류 시에는 주행 단위 분절 특성을 고려한 추가적인 데이터 정제 및 통합 과정이 요구됨

〈표 2-4〉 링크 기반 내비게이션 데이터 형식



구분	내용	예시
P_REQDATE	주행시작일(YYYYMMDD)	20230209
TRIP_INIT_SESSION_ID	주행 고유 ID	108820220209171330540S00
LINK_SEQ	주행 내 링크 시퀀스	1
MESH_CD	MESH 코드	5716
LINK_ID	LINK 코드	512
LINK_DIR	LINK 방향 (0 : 정방향 / 1 :역방향)	1
LENGTH	LINK 길이	49
DRIVING_TIME	LINK 주행시간 (초)	41
ENTRY_TIME	링크 진입시간(연월일시분초)	20240209171332
EXIT_TIME	링크 진출시간(연월일시분초)	20240209171413
SPEED_CAPPED	CAP이 씌워진 상태로 계산된 속도값	/N
SPEED_GPS	단말기에서 올려준 속도값의 평균	4.9333
SPEED_REFINED	보완된 최종 속도값	4.9333
CAPPED_YN	CAP 씌워짐 여부	N
DEPART_NAME	출발지명(읍면동 기준, GPS 문제시 NULL)	서울특별시 강남구 역삼1동
DEPART_X	출발지 X좌표(BESSEL)	127.039416
DEPART_Y	출발지 Y좌표(BESSEL)	37.6938611
DEST_NAME	목적지명	성남시청 음악분수
DEST_POIID	목적지 POI ID	1000670
DEST_X	목적지 X좌표(BESSEL)	127.1278333
DEST_Y	목적지 Y좌표(BESSEL)	37.4176111
CB_NAME	목적지 중분류 정보(주소 검색 등의 경우 NULL)	구역내시설물
CAR_MODEL	차종("운전습관 가입자" 대상)	CT_NORMAL
CAR_FUEL	유종("운전습관 가입자" 대상)	FT_LPG

- 포인트 기반 내비게이션 데이터는 GPS 궤적(X, Y 좌표)을 기반으로 주행 시간, 속도, 고도 등의 시계열 정보를 포함하며, 차량의 배기량 및 연식 정보까지 제공된다는 점에서 링크 기반 데이터와 차별성이 있음
 - GPS 데이터는 기본적으로 1초 단위로 수집되나, 경로 이탈, 수신 불량, 맵매칭 오류 등의 영향으로 일부 구간에서는 1초 이상의 불규칙한 수집 주기가 발생함
 - GPS 속도 및 고도값이 음수로 나타나는 경우는 위성 신호 수신 불량에 따른 오류로 판단되며, 본 연구에서는 임의 보정을 수행하지 않고 분석 목적에 따라 필터링하는 방식으로 처리함
 - 배기량 및 연식 정보는 교통환경지표 산출을 위한 잠재 변수로 활용 가능하나, CP사 기반 수집 특성상 차종 및 유종 대비 표본율이 낮아 활용성에는 제한이 있음
- 포인트 기반 데이터 수집 과정에서 일부 좌표가 맵매칭을 통해 보정된 상태로

제공된 것이 확인됨에 따라, 분석의 일관성과 정확성 확보를 위해 원시 GPS 데이터 기준으로 재추출을 수행함

- 데이터 수집·분석, 오류 검증, 재요청 및 재추출 과정이 총 3회 반복 수행되었으며, 이 과정에서 2024년 상반기 원시 GPS 데이터는 데이터 제공기관 서버에 미존재함이 확인됨에 따라, 최종적으로 2024년 6월, 7월, 10월, 12월 데이터를 대상으로 수집 기간을 조정함
- 원시 GPS 데이터는 개별 포인트 단위로 구성되어 통행 단위로 직접 연결되지 않으므로, Trip_init_session_id와 session_id 간 매핑이 가능한 관계 데이터를 활용하여 단일 통행(Trip) 단위로 재구성함(<표 2-5 참조>)

<표 2-5> 통행 구성을 위한 경로 시각화 예시

trip_init_session_id 기준 시각화	session_id(85개) 기준 시각화
	

- 포인트 기반 내비게이션 데이터는 향후 안정적인 데이터 수집 및 품질 검증이 완료된 이후, 단계적으로 데이터 분석 및 DB 구축에 활용함

〈표 2-6〉 포인트 기반 내비게이션 데이터의 주행경로 데이터 형식

구분	내용	예시
P_REQDATE	주행시작일(YYYYMMDD)	20241016
TRIP_INIT_SESSION_ID	주행 고유 ID	105020220113081947967S00
GPS_X	GPS 궤적 X좌표	127.1137778
GPS_Y	GPS 궤적 Y좌표	36.79916111
GPS_TM	해당 GPS좌표 시 주행시간(연월일시분초)	2.02201E+13
SPEED_GPS	GPS 속도(GPS 오류 일 경우 음수값존재,m/s)	0
ALTITUDE	단말기로부터 수집된 상대 고도값(단위°)	0
ANGLE	단말기로부터 수집된 차량 진행 방향각(0~359°)	273

〈표 2-7〉 포인트 기반 내비게이션 데이터의 사용자 정보 데이터 형식

구분	내용	예시
P_REQDATE	주행시작일(YYYYMMDD)	20241016
TRIP_INIT_SESSION_ID	주행 고유 ID	105020220113081947967S00
DEPART_NAME	출발지명(읍면동기준, GPS 문제시 NULL)	경상북도 문경시 마성면
DEST_NAME	목적지명	양천구보건소
DEST_POIID	목적지 POI ID	459268
DEST_X	목적지 X좌표(BESSEL)	126.8680833
DEST_Y	목적지 Y좌표(BESSEL)	37.51502777
CB_NAME	목적지 중분류 정보(주소 검색 등의 경우 NULL)	의료시설
CAR_MODEL	차종("운전습관 가입자" 대상)	CT_NORMAL
CAR_FUEL	유종("운전습관 가입자" 대상)	FT_GAS
CAR_DISPLACEMENT	주행 차량 배기량("운전습관 가입자" 대상)	2199
CAR_MODELYEAR	주행 차량 연식("운전습관 가입자" 대상)	2016

〈표 2-8〉 포인트 기반 내비게이션 데이터의 주행 고유ID와 주행 탐색ID 관계 데이터 형식

구분	내용	예시
P_REQDATE	주행시작일(YYYYMMDD)	20241016
SESSION_ID	세션 ID	108620231016174806052S00
TRIP_INIT_SESSION_ID	주행 고유 ID	108320231016181319634S01

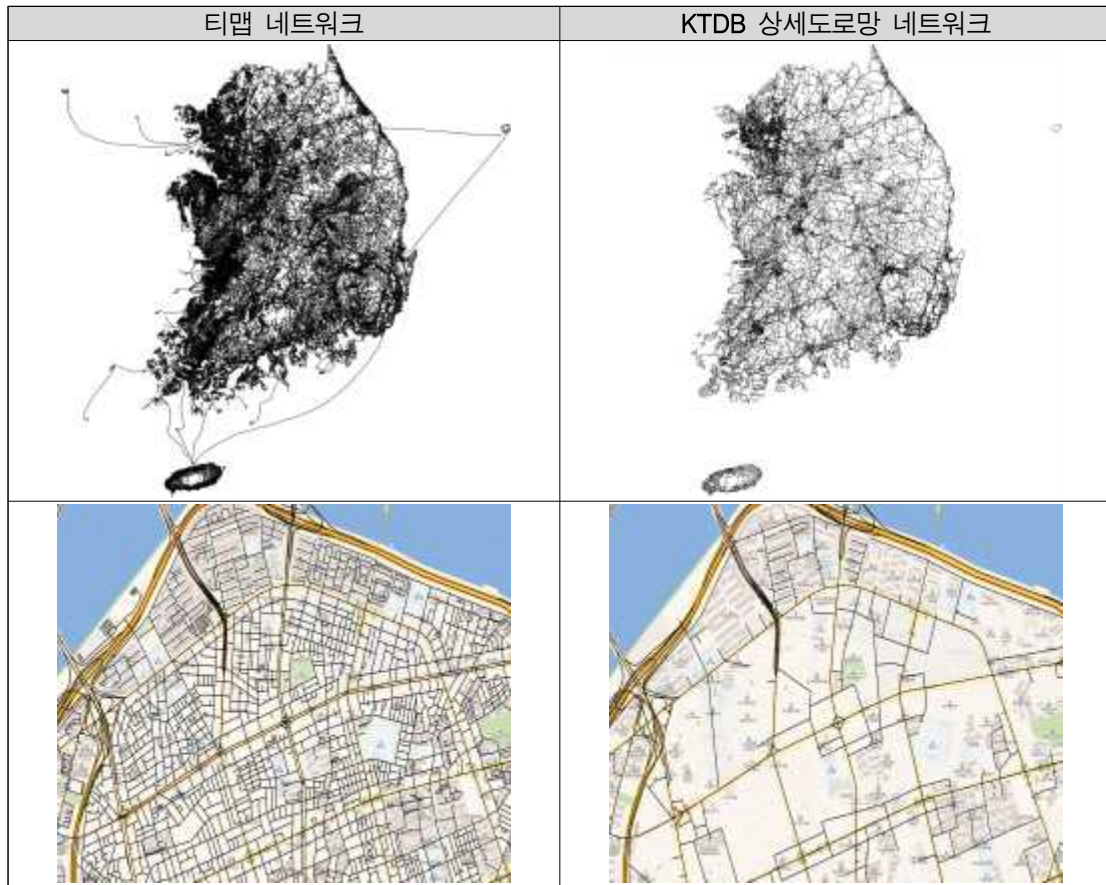
〈표 2-9〉 내비게이션 데이터 차종 정보

구분	차종 정보	차종 구분 기준
CT_COMPACT	경차	경차
CT_NOMAL	승용차	16인승 이하 승합차
CT_MIDDLE	중형승합차	17인승 이상 32인승 이하 승합차, 2.5톤 이상 5.5톤 이하 화물차
CT_LARGE	대형승합차	33인승 이상 승합차, 5.5톤 초과 10톤 미만 화물차
CT_TRUCK	소형화물차	2.5톤 미만 화물차
CT_SPECIAL	대형화물차	10톤 이상 20톤 미만 화물차, 3축 차량
CT_SAMLL_TRUCK	특수화물차	20톤 이상 화물차, 4축 이상 차량

〈표 2-10〉 내비게이션 데이터 유종 정보

구분	유종 정보
FT_GAS	휘발유
FT_DIESEL	경유
FT_LPG	LPG
FT_GASPM	고급휘발유
FT_EV	전기

- 티맵 내비게이션 데이터는 티맵 자체 도로 네트워크의 링크 단위를 기준으로 차량의 위치 및 이동경로 정보를 제공함
- 본 과업은 KTDB 상세도로망 네트워크를 기준으로 기초교통 및 통행지표 DB를 구축하는 것을 목표로 하므로, 티맵 링크 체계에서 생성된 경로를 KTDB 상세도로망 링크 체계로 변환·재구성하는 과정이 필수적으로 요구됨
 - 이를 위해 티맵 링크와 KTDB 링크 간 공간적 정합성 확보(좌표 기반 매칭), 링크 분할·병합 처리, 방향성 일치 검증 등 맵매칭 및 경로 재생성 절차를 수행함
 - 특히, 티맵 네트워크는 교차로 내부 링크 및 세부 차로 단위까지 반영된 고해상도 구조를 가지는 반면, KTDB 상세도로망은 교통분석 목적에 적합한 표준화된 링크 체계를 적용하고 있어, 두 네트워크 간 구조적 차이를 고려한 변환 로직이 필요함
 - 2024년 12월 기준 티맵 도로 네트워크의 양방향 링크 연장은 859,682km로, KTDB 상세도로망 네트워크의 227,630km 대비 약 3.7배 수준으로 더 세분화된 것으로 분석됨
 - 이에 따라, 티맵 경로는 단순 매칭이 아닌 세분 링크 통합 및 구간 분할을 통해 KTDB 기준으로 재구성하여 분석 일관성과 지표 정합성을 확보함



<그림 2-1> 티맵 네트워크와 상세도로망 네트워크 상세도 비교

2) 사업용 운행기록계 데이터

- 사업용 운행기록계(DTG) 데이터는 사업용 차량에 장착된 디지털 운행기록계를 통해 차량의 운행 상태를 실시간으로 기록·저장한 데이터로, 한국교통안전공단으로부터 제공받음
 - 해당 데이터는 GPS 좌표를 기반으로 1초 단위로 수집되며, 순간속도, 가속도, 엔진회전수(RPM), 자동차 유형 등 차량의 동적 운행 정보를 포함함
 - 특히 자동차 유형 정보는 버스, 화물, 택시 등 차종 구분이 가능하여, 차종별 이동경로 및 통행 특성 분석 등 다양한 교통 분석에 활용 가능함

〈표 2-11〉 DTG 데이터의 포맷

필드명	설명	자료형
TRIP_KEY	등록번호 & 정보발생 일시	string
DTG_MDL_NM	단말기 모델명	string
CHASSIS_NO	차량 고유번호 1	string
CAR_TP_CD	차량 유형 구분	string
CAR_REG_NO	차량 고유번호 2	string
BIZ_REG_NO	운송사업자 등록번호	string
DRIVER_CODE	운전자 코드	string
DLY_DRIV_DIST	일일 주행거리	double
ACCM_DRIV_DIST	누적 주행거리	double
OPT_SPD	차량속도	double
RPM	분당 엔진 회전수	double
BREAK_SIG	1: 유, 2: 무	integer
GPS_X	차량위치 X (WGS84 경위도)	double
GPS_Y	차량위치 Y (WGS84 경위도)	double
GPS_AGL	지점별 방위각	double
ACCEL_VX	횡가속	double
ACCEL_VY	종가속	double
CONT_CODE	통신상태코드	integer
AREA_CODE	행정기관코드, 대존코드	integer
OPT_Time	YYMMDDHHMMSSSS	string

〈표 2-12〉 DTG 데이터의 자동차 유형 코드

구분 코드	자동차 유형	구분 코드	자동차 유형
00	미분류	17	장외차량
11	시내버스	21	일반택시
12	농어촌버스	22	개인택시
13	마을버스	31	일반화물
14	시외버스	32	개인화물
15	고속버스	41	비사업용차량
16	전세버스	-	-

제2절 차량 GPS 원시데이터 분석

가. 데이터 프로파일링

- 2024년 티맵 내비게이션 데이터는 2023년 대비 약 8.2% 증가한 총 858억 건의 개별 차량 주행 이벤트가 수집되어, 데이터 규모 및 활용 가능성이 지속적으로 확대되고 있음
- 2024년 DTG 데이터는 2023년 대비 약 39.2% 증가한 총 2조 8천 9백억 건의 이벤트가 수집되어, 사업용 차량 중심의 고해상도 시계열 데이터 확보가 크게 향상됨

〈표 2-13〉 데이터의 용량 및 개별 차량의 링크 발생 건수

데이터명	용량(TB)			이벤트 수(천 건)			
	2022년	2023년	2024년	2022년	2023년	2024년	
링크 기반 내비게이션 데이터	19.4	22.3	26.3	70,349,051	79,364,038	85,844,541	
DTG 데이터	버스	9.64	9.22	10.56	772,320,634	770,774,514	893,255,731
	화물차	3.93	4.32	5.01	317,497,577	355,365,816	421,405,927
	택시	5.91	10.24	17.1	491,244,500	953,355,667	1,579,449,885

- 티맵 내비게이션 데이터의 차종 및 유종 정보는 전체 이용자가 아닌 자동차보험 할인 프로그램에 참여한 “운전습관 가입자”를 통해 수집되므로, 모든 주행 데이터에 대해 해당 정보가 포함되지 않는 구조적 한계가 존재함
- 티맵 내비게이션 데이터를 통한 차종 정보 분석 결과
 - 주행 고유 ID 기준 분석 결과, 2024년 데이터의 약 82% 이상에서 차종 정보가 확인되어 전반적인 활용성은 확보된 것으로 판단됨
 - 다만 일부 데이터는 차종 정보가 결측되어 있어 분석 시 표본 대표성에 대한 보완 검토가 필요하며, 2024년 기준 승용차 비중이 약 75%로 가장 높고 증가 추세를 보이며, 개인 이동 중심 특성을 반영하므로 차종별 분석 시 편중을 고려해야 함

〈표 2-14〉 티맵 내비게이션 데이터 차종 분석 결과

차종 정보	2022년	2023년	2024년
승용차	55.2%	68.6%	74.8%
경차	2.4%	3.0%	3.3%
중형승합차	1.7%	1.8%	1.3%
대형승합차	0.5%	0.6%	0.5%
소형화물차	0.5%	0.8%	1.1%
대형화물차	0.03%	0.1%	0.8%
특수화물차	0.02%	0.3%	0.3%
정보 없음	39.6%	24.9%	17.8%
합계	100%	100%	100%

- 티맵 내비게이션 데이터를 통한 유종 정보 분석
 - 내비게이션 데이터의 주행 고유 ID별로 유종 정보가 포함된 비율을 분석한 결과, 2024년 데이터는 약 75% 이상의 데이터에서 유종 정보가 확인되었음
 - 연도별로 유종 분석 결과, 2024년 데이터는 휘발유에서 약 56%로 가장 많이 발생하였고, 점차적으로 증가하고 있는 것으로 분석됨

〈표 2-15〉 티맵 내비게이션 데이터 유종 분석 결과

유종 정보	2022년	2023년	2024년
휘발유	41.4%	51.3%	56.2%
경유	16.6%	20.1%	20.4%
LPG	4.7%	5.6%	5.5%
고급휘발유	1.1%	1.3%	1.4%
전기	1.1%	1.5%	2.0%
정보 없음	35.3%	20.3%	14.5%
합계	100%	100%	100%

- 내비게이션 데이터의 월별 데이터 발생 현황
 - 2024년 링크 기반 내비게이션 데이터의 총 용량은 26.27TB이며, 총 858억 건의 이벤트 수가 수집되어 개별 차량 경로 수는 총 8.4억 건의 경로가 수집되었음
 - 2023년 데이터와 비교하여 데이터 용량은 약 6TB 증가, 이벤트 수는 약 122억 건 증가, 경로 수는 약 2.2억 건 증가한 것으로 분석됨

〈표 2-16〉 링크 기반 내비게이션 데이터 현황

구분	용량	이벤트 수(개)	경로 수(개)
1월	2.02 TB	6,564,525,898	62,041,360
2월	3.18 TB	10,362,522,198	90,126,152
3월	2.18 TB	7,092,691,866	66,974,548
4월	2.22 TB	7,218,826,009	68,827,938
5월	2.39 TB	7,820,738,090	79,946,461
6월	2.28 TB	7,471,586,755	73,231,616
7월	2.10 TB	6,883,426,048	67,573,651
8월	2.23 TB	7,308,722,534	70,281,512
9월	3.00 TB	9,808,730,771	92,639,314
10월	1.59 TB	5,221,534,825	58,828,837
11월	1.61 TB	5,261,601,837	58,657,668
12월	1.48 TB	4,829,634,827	54,088,201
합계	26.27 TB	85,844,541,658	843,217,258

- 2022년 데이터부터 용량이 증가한 것은 차종, 유종 등 속성정보 추가로 인한 것임

〈표 2-17〉 연도별 내비게이션 데이터 프로파일링 비교

구분	2021년	2022년	2023년	2024년	'21년 대비 '22년 증감율	'22년 대비 '23년 증감율	'23년 대비 '24년 증감율
데이터 용량	7.2TB	19.4TB	20.28TB	26.27TB	145.1%	4.5%	29.5%
이벤트 건수	586억 건	703억 건	738억 건	858억 건	35.0%	4.9%	16.3%
경로 수	4.5억 건	5.4억 건	6.2억 건	8.4억 건	15.1%	3.8%	30.6%

- 포인트 기반 내비게이션 데이터는 수집이 완료되었으나, 현재 데이터를 가공할 수 있는 물리적 환경 및 리소스 부족으로 인해 한달 치 데이터에 대해서만 가공을 수행하였음

- 추후 링크 기반 내비게이션 데이터 가공이 완료된 이후 리소스 확보를 한 뒤 포인트 기반 내비게이션 데이터 가공을 수행할 예정임

〈표 2-18〉 포인트 기반 내비게이션 데이터 현황

구분	용량	이벤트 수(개)	경로탐색 ID수(개)	경로 수(개)
6월	2.91 TB	121,737,726,456	888,055,973	78,295,622
7월	2.17 TB	91,054,459,590	663,344,684	65,034,960
10월	2.20 TB	92,543,622,111	727,372,064	63,136,476
12월	1.58 TB	65,233,314,016	561,747,934	32,077,915

- DTG 데이터는 자동차 유형 정보를 활용하여 차종별 분포를 분석하였으며, 총 125,506천 대의 주행 고유 ID 기준으로 구성됨
- 분석 결과, 일반화물이 52.30%로 가장 높은 비중을 차지하였으며, 전세버스(25.89%), 시내버스(13.17%) 순으로 나타남
- 그 외 농어촌버스(2.39%), 개인화물(2.38%), 마을버스(1.84%), 시외버스(1.36%) 등으로 구성되어, 화물 및 특정 버스 유형 중심의 데이터 구조를 보임

〈표 2-19〉 DTG데이터 차종 분석 결과

차종 정보	주행 고유 ID 대수 (천대)	비율
시내버스	16,532	13.17%
농어촌버스	2,994	2.39%
마을버스	2,306	1.84%
시외버스	1,713	1.36%
고속버스	841	0.67%
전세버스	32,495	25.89%
일반화물	65,638	52.30%
개인화물	2,988	2.38%
합계	125,506	100%

- DTG데이터의 용량 및 이벤트 건수 현황
 - DTG 데이터는 대용량 특성으로 인해 기존 압축 해제 후 가공 과정에서 처리 시간이 과다 소요되는 문제가 발생함에 따라, 2023년부터는 압축 해제 없이 처리 가능한 다수의 분할된 gz 파일 형태로 데이터 추출 방식을 개선함
 - 전체 압축 데이터 용량은 약 33.5TB이며, 약 2조 9천억 건의 이벤트 수집됨

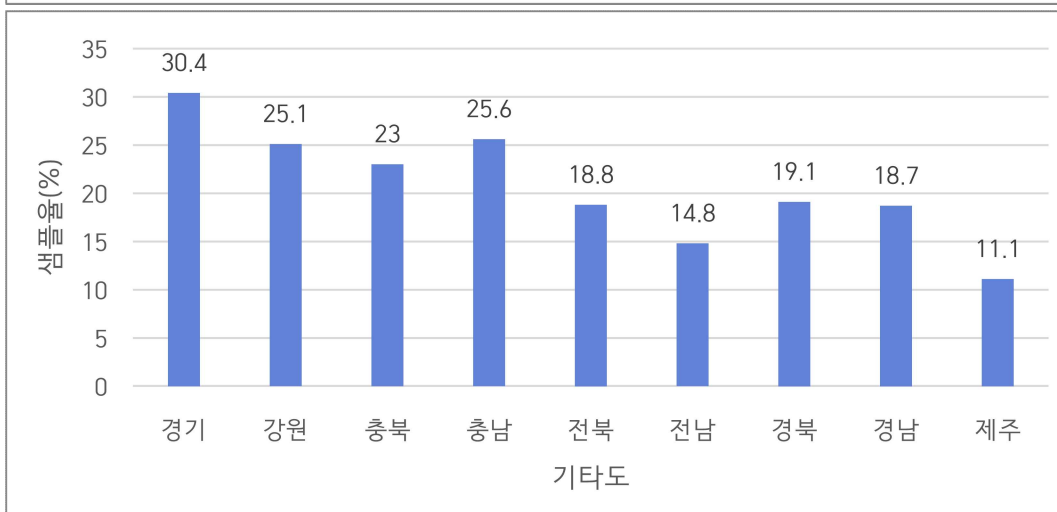
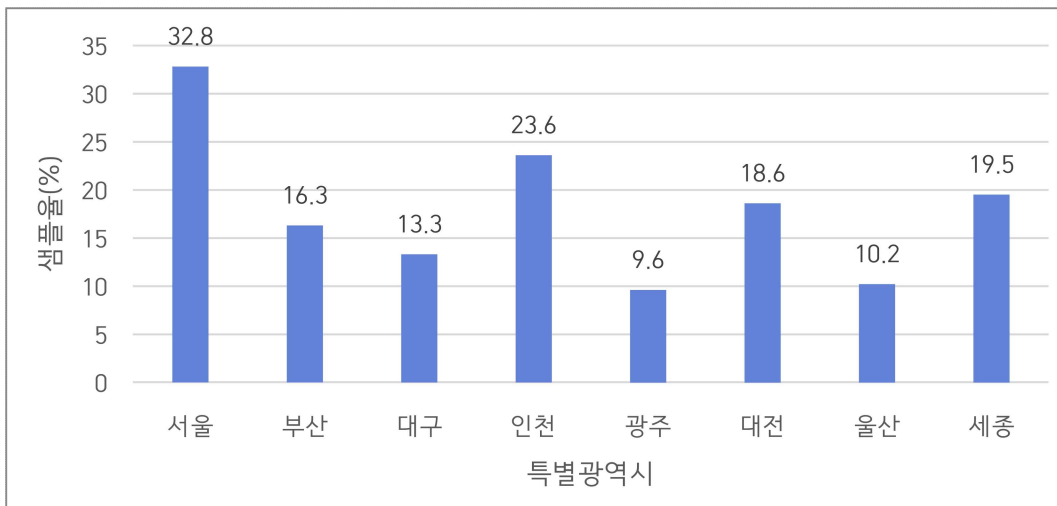
〈표 2-20〉 DTG 데이터 현황

구분	버스		화물		택시	
	용량	이벤트 수(십억 건)	용량	이벤트 수(십억 건)	용량	이벤트 수(십억 건)
1월	831.6	68.8	390.24	32	1,239.39	117.5
2월	811.27	63.9	326.7	26.5	1,257.68	120.3
3월	910.08	71.8	363.42	29.4	1,420.27	128.7
4월	918.34	77.4	357.64	28.5	1,485.3	128.7
5월	913.64	71.8	390.56	31.3	1,397.86	120.9
6월	889.84	73.2	396.51	32.1	1,303.76	113.7
7월	892.34	72.4	398.21	32.9	1,570.13	148.5
8월	863.87	71.1	412.23	34.4	1,567.93	137.2
9월	862.28	74.8	373.18	30.7	1,499.71	134.3
10월	1006.64	83.1	549.62	44.9	1,590.91	139.4
11월	963.85	80.3	586.22	49.2	1,573.5	140.1
12월	953.24	84.7	587.3	49.6	1,608.68	150.2
합계	10,816.99	893.3	5,131.83	421.5	17,515.12	1579.5

나. 데이터 샘플율 분석

○ 티맵 내비게이션 데이터를 통한 관측지점 샘플율 분석 결과

- 내비게이션 데이터를 활용하여 시도별 연평균 일 프로브 샘플율을 분석함
 - 공공 관측교통량 데이터 및 연평균 일 통행량 데이터를 기반으로 샘플율을 산출하였으며, 2024년 상세도로망 네트워크 기반 관측교통량 DB를 활용하여 분석함
 - 시도별 샘플율 분석 결과, 수도권(서울, 경기, 인천)과 충청권, 세종시에서 상대적으로 높은 샘플율이 나타나는 것으로 분석되며, 이는 해당 지역에서 내비게이션 이용률이 높고 데이터 수집 밀도가 높은 특성을 반영한 결과로 판단됨
 - 반면 일부 비수도권 지역은 상대적으로 낮은 샘플율을 보이며, 이에 따라 교통량 추정 및 지표 산출 시 추가적인 보정 필요성이 존재함



〈그림 2-2〉 시도별 연평균 일 내비게이션 프로브 샘플율



제3장 차량 GPS 데이터 가공 및 DB 구축

제1절 차량 DB 가공 및 구축 프로세스

제2절 차량 GPS 데이터 전처리 및 경로 가공

제3절 교통량 추정 DB 구축

제4절 기초교통 DB 구축

제5절 차량 교통망 성능평가지표 구축

제6절 프로토타입 회전교통량 추정DB 구축

제3장 차량 GPS 데이터 가공 및 DB 구축

제1절 차량 DB 가공 및 구축 프로세스

- 본 사업에서 활용하는 차량 GPS 데이터는 내비게이션 데이터와 사업용 DTG 데이터로 구성되며, 데이터 활용을 위해서는 각 데이터의 특성을 고려한 분석 및 가공·검증 과정이 필요함
- 티맵 내비게이션 데이터는 링크 기반, DTG 데이터는 포인트 기반으로 수집되어 데이터 구조와 수집 주기에서 차이가 존재하므로, 데이터 유형별 특성에 적합한 전처리 및 가공 모듈을 적용함
- 차량 GPS 데이터 가공 및 구축 프로세스는 데이터 전처리 → 개별 차량 경로 DB 구축 → 1차 가공DB(기초교통 DB) 구축 → 2차 가공DB(교통망 성능평가 지표 DB) 구축의 단계로 체계적으로 수행됨



〈그림 3-1〉 차량 GPS 데이터 가공 및 구축 프로세스

- 티맵 내비게이션 데이터는 링크 기반 데이터로 제공되며, 이를 활용하기 위해 링크 단위 정보를 1초 단위 궤적(포인트) 데이터로 변환하여 표준 포인트 기반 경로 데이터를 우선 구축함
- 생성된 포인트 기반 경로 데이터는 티맵 네트워크의 세분화된 구조를 고려하여 포인트 단위 변환 및 재매칭을 수행한 후 Level6 네트워크와의 맵매칭을 통해 이기종 네트워크 간 차이를 보정하고 표준화된 링크 경로 데이터로 재구성함



<그림 3-2> 링크 기반 데이터 전처리 및 경로 데이터 가공 프로세스

- DTG 데이터는 포인트 기반 데이터로 포인트 기반 데이터 맵매칭 및 경로 가공 모듈을 이용하여 가공 및 구축함(해당 모듈은 R&D사업을 통해 개발된 것임)



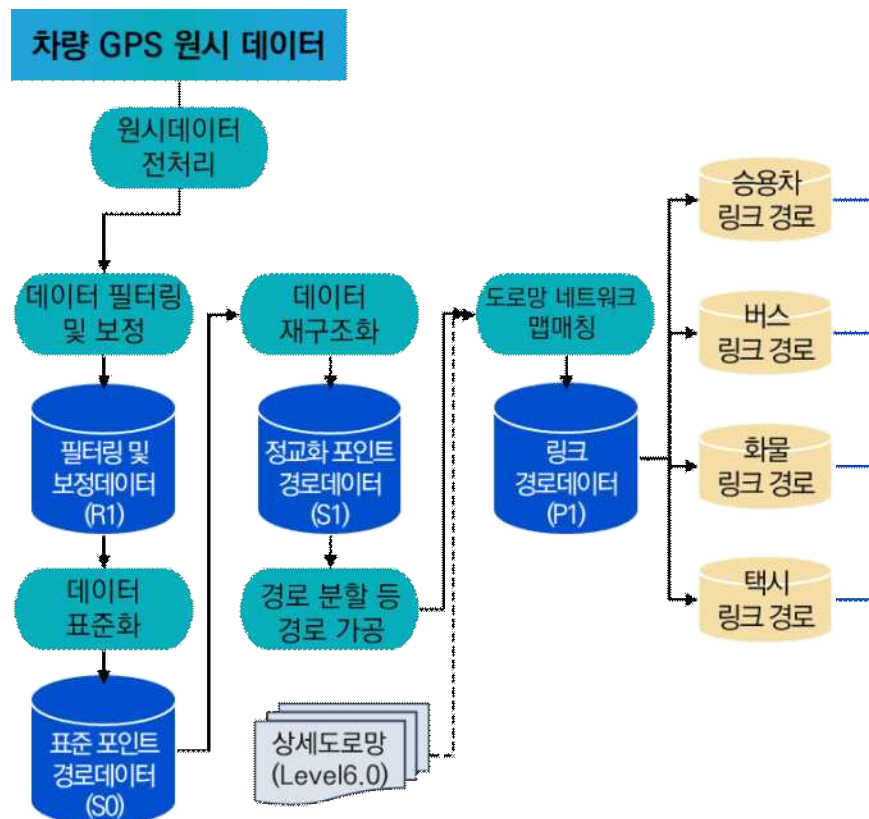
<그림 3-3> 포인트 기반 데이터 전처리 및 경로 데이터 가공 프로세스

제2절 차량 GPS 데이터 전처리 및 경로 가공

1. 링크 단위 티맵 내비게이션 데이터 전처리 및 경로 가공

가. 개별차량 경로DB 구축 프로세스

- 차량 GPS 데이터를 활용한 개별차량 경로 DB 구축은 데이터 유형별 특성을 반영하여 단계적으로 수행됨
 - 링크 기반 데이터는 이미 도로 네트워크와 맵매칭된 구조로 제공되므로, 별도의 궤적 재구조화 없이 정합성 검증 및 경로 생성 중심으로 구축함
 - 반면, 포인트 기반 데이터는 GPS 수신 오류 및 노이즈가 포함될 수 있어, 궤적 보정 및 정합화 등 재구조화 과정을 거쳐 경로 데이터를 생성함



〈그림 3-4〉 차량 GPS 데이터를 활용한 개별차량 경로DB 구축 프로세스

나. 원시데이터 전처리

○ 원시데이터 적재 및 전처리

- 대용량의 압축 파일을 빅데이터 가공 시스템으로 처리하기 위해 원시 데이터를 분할하여 HDFS에 적재함



<그림 3-5> 데이터 적재

- 원시데이터를 일자별로 분리하고, 불필요한 속성정보를 제거하여 데이터 용량을 축소시킴



<그림 3-6> 자료형 변환

- 빅데이터 환경의 처리속도 향상을 위해 데이터 포맷을 변환



<그림 3-7> 데이터 포맷 변환

다. 데이터 필터링 및 보정

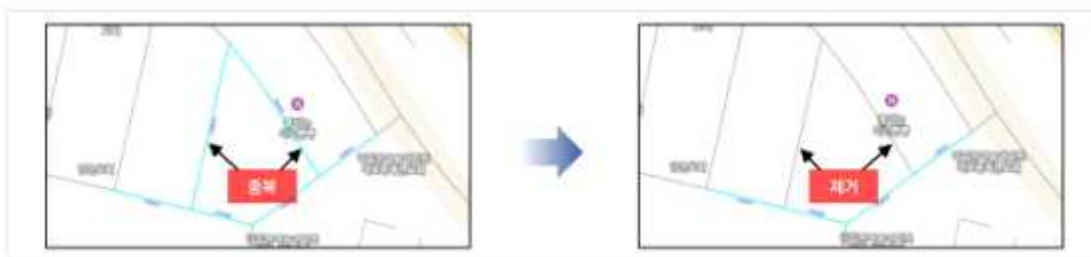
○ 내비게이션 데이터 필터링 및 보정

- 데이터 품질 확보를 위해 시간, 필드, 속도 등 오류 유형별 기준을 정의하고 이에 따른 필터링 및 보정 과정을 수행함
- 주행 시작일 필드 오류 유형 : 주행 시작일(R_P_REQDATE)과 링크 진입시간(ENTRY_TIME) 필드의 일자가 맞지 않는 경우 링크 진입시간을 기준으로 주행 일자에 대해 정렬하여 일자별로 분할함

DATE	TRIP_INIT_SESSION_ID	LINK_SEQ	MESH_CD	LINK_ID	LINK_DIR	LENGTH	DRIVING TIME	ENTRY_TIME	EXIT_TIME
20210209	105520210208235734644500	1	4787	16033	0	63.894627	6.790106	20210209000226	20210209000232
20210209	105720210208235718508500	1	5631	114	0	121	44.565998	20210209000653	20210209000737
20210208	108120210208232709308500	1	4777	3744	1	25	2.8363636	20210209073336	20210209073339
20210209	105020210208175213265501	1	7418	28853	1	84	27.49091	20210210030545	20210210030559
20210209	105020210208210128425500	1	4775	9789	1	231.51741	11.862713	20210210084021	20210210084032

〈그림 3-8〉 주행 시작일 필드와 링크 진입시간 필드 날짜 불일치 오류 예시

- 링크 시퀀스 필드 오류 유형 : 링크 진입시간(ENTRY_TIME)과 링크 진출시간(EXIT_TIME) 필드의 시간이 링크 시퀀스(LINK_SEQ) 순서와 다를 경우 진입시간, 진출시간, 링크 시퀀스 순서로 정렬함
- 링크 중복 오류 유형 : 동일 시간에서 발생한 링크의 중복 경로를 보정함



〈그림 3-9〉 링크 중복 경로 예시

- 경로 기준일 불일치 오류 유형 : 링크 진입시간이 기준 일자와 다르거나 당일·익일 범위를 벗어나는 경우 오류 데이터로 판단하여 제거함

20240122	102020240122090707014901	2	4787	14361	1	82	5	20240122090709	20240122090714	WH	20.0	20.0
20240122	102020240122090707014901	5	4787	4221	0	56	1	20240122090715	20240122090715	WH	20.5	20.5
20240122	102020240122090707014901	1	5217	40840	1	104	1	20240122090724	20240122090740	WH	25	25
20240122	104120240122114203035502	1	0780	20738	0	56	0	20240122114307	20240122114307	WH	3	3
20240122	104120240122114203035502	1	4788	10714	1	56	0	20240122114312	20240122114313	WH	4	4

〈그림 3-10〉 일자오류 예시

- 이상 속도값 발생 오류 유형 : GPS 수신 오류 등으로 인해 비정상적으로 산출된 속도값(200km/h 초과)은 이상치로 판단하여 필터링함

- 2024년 티맵 원시데이터의 전체 이벤트 수 대비 오류 이벤트 수의 비율은 약 2.4%로 확인됨
 - 2024년에 일자를 기준으로 오류 데이터를 판단하는 로직이 추가되었기에 기존에 비해 오류 이벤트 건수가 증가하였음

<표 3-1> 티맵 오류 이벤트 필터링 비율

연도	전체 이벤트 수	오류 이벤트 수	필터링 비율
2022년	70,349,051,290	224,456,055	0.32%
2023년	73,819,850,279	85,733,051	0.12%
2024년	85,844,541,658	2,087,872,587	2.43%

- 통행정보 외 불필요 컬럼을 제거하고 차종 정보를 포함한 전처리를 수행하여 데이터 용량을 효율화함
 - 2024년 티맵 원시데이터 전처리 소요기간은 약 2일이 소요되었으며, 전처리 과정을 통해 전체 데이터 용량은 약 4% 수준으로 축소됨

라. 포인트 변환

- 궤적별 연속된 링크 데이터에 대해 시간을 기준으로 통행을 분리함
 - 차량의 궤적 데이터는 1일 기준으로 데이터가 연속 수집되기 때문에 출·도착 기준의 통행 분리가 필요함
 - 궤적별 데이터가 5분 이상 미수집될 시 해당 지점을 기준으로 통행을 분리
- 링크 기준 데이터를 1초 단위 포인트 데이터 변환하여 생성 및 저장
 - 티맵 원시 내비게이션 데이터에 나타난 링크 길이 값이 존재하지 않는 경우가 있기 때문에 티맵 네트워크의 링크 길이를 참조하여 링크의 속도를 기준으로 주행시간 재계산
 - 티맵 네트워크와 매칭되지 않는 레코드 제거

```

20211014,105820211014235834660500,1,4787,27500,1,20,0,2,1010102,20211014235900,20211014235905,32,99999725,31,0,31,0,y
20211014,105820211014235834660500,2,4787,27504,0,11,0,1,2,20211014235900,20211014235906,33,00000000000001,14,32,00000000000007,y
20211014,105820211014235834660500,3,4787,27505,1,11,0,1,2,20211014235900,20211014235907,33,00000000000001,35,0,35,0,y
20211014,105820211014235834660500,4,4787,27499,0,20,0,2,1010102,20211014235907,20211014235909,32,99999725,10,32,99999725,y
20211014,105820211014235834660500,5,4787,27499,1,11,0,1,2,20211014235900,20211014235910,33,00000000000001,35,0,35,0,y
20211014,105820211014235834660500,6,4787,31131,0,15,0,1,6363636,20211014235910,20211014235911,33,00000079333333,14,32,00000073333333,y
20211014,105820211014235834660500,7,4787,31130,1,51,0,6,47927,20211014235911,20211014235918,26,712176300363504,10,26,712176300363504,n
20211014,105820211014235834660500,8,4787,30633,0,79,0,0,610102,20211014235918,20211014235925,32,9999930079749,34,0,34,0,y
20211014,105820211014235834660500,9,4787,30634,1,12,0,1,3090909,20211014235925,20211014235929,33,0000002916666,10,33,0000002916666,y
20211014,105820211014235834660500,10,4787,31700,0,11,0,1,2,20211014235929,20211014235927,32,00000000000001,14,32,00000000000007,y
20211014,105820211014235834660500,11,4787,31703,1,26,0,2,8263636,20211014235927,20211014235929,33,00000042307693,41,0,41,0,y
20211014,105820211014235834660500,12,4787,27494,1,20,0,2,0363636,20211014235929,20211014235932,33,00000042307693,42,0,42,0,y
20211014,105820211014235834660500,13,4787,27494,1,20,0,2,1010102,20211014235932,20211014235933,32,99999725,10,32,99999725,y
20211014,105820211014235834660500,14,4787,27485,1,20,0,2,1010102,20211014235933,20211014235935,32,99999725,40,0,40,0,y
20211014,105820211014235834660500,15,4787,49150,0,13,0,1,4101010,20211014235935,20211014235937,33,00000042307693,33,0,33,0,y
20211014,105820211014235834660500,16,4787,49140,1,34,0,5,606215,20211014235937,20211014235942,21,8329122230901,10,21,8329122230901,n
    
```



<그림 3-11> 티맵 네트워크 미참조 링크

- 궤적별 첫 링크의 시작 시간을 기준으로 재계산된 주행시간을 누적하여 링크의 진입과 진출 시간을 조정

시간 보정 전						시간 보정 후				
순번	진입시간 (sec)	진출시간 (sec)	주행시간 (sec)	링크길이 (m)	캡 속도 (km/h)	계산된 주행시간 (millisec)	순번	진입시간 (millisec)	진출시간 (millisec)	주행시간 (millisec)
1	1633657281	1633657281	10	89	22.9620576	10808.427	1	1633657281000	1633657281308	10808
2	1633657281	1633657281	20	87	12.52166762	19162.61	2	1633657281808	1633657281971	19253
3	1633657281	1633657285	4	21	15.08886947	4688.9007	3	1633657281971	1633657285770	4698
4	1633657285	1633657307	22	78	13.08373128	21481.784	4	1633657285770	1633657307232	21462
5	1633657307	1633657316	9	38	14.21919766	9610.796	5	1633657307232	1633657316953	9621
6	1633657316	1633657325	9	57	23.79010544	8647.2438	6	1633657316953	1633657325800	8647
7	1633657325	1633657327	2	15	28.98712773	1821.9745	7	1633657325800	1633657327863	1802
8	1633657327	1633657333	5	57	35.53301577	5774.0584	8	1633657327863	1633657333075	5774
9	1633657333	1633657337	4	45	34.58832941	4682.368	9	1633657333075	1633657337758	4682
10	1633657337	1633657344	7	35	18.58051359	6807.9854	10	1633657337758	1633657344741	6808
11	1633657344	1633657354	10	52	19.72996913	9485.104	11	1633657344741	1633657354283	9488
12	1633657354	1633657377	23	283	46.28127458	22791.075	12	1633657354283	1633657377320	22781
13	1633657377	1633657388	9	137	51.58130664	9306.368	13	1633657377320	1633657388325	9306
14	1633657388	1633657394	5	104	45.49400564	8229.656	14	1633657388325	1633657394756	8230
15	1633657394	1633657401	107	148	4.552305042	13131.854	15	1633657394756	1633657406069	13132
16	1633657407	1633657580	13	107	31.50289292	12727.454	16	1633657406069	1633657583115	12727
17	1633657580	1633657579	19	47	8.598289818	19682.91	17	1633657583115	1633657603968	19683

〈그림 3-12〉 진입시간, 진출시간 재계산

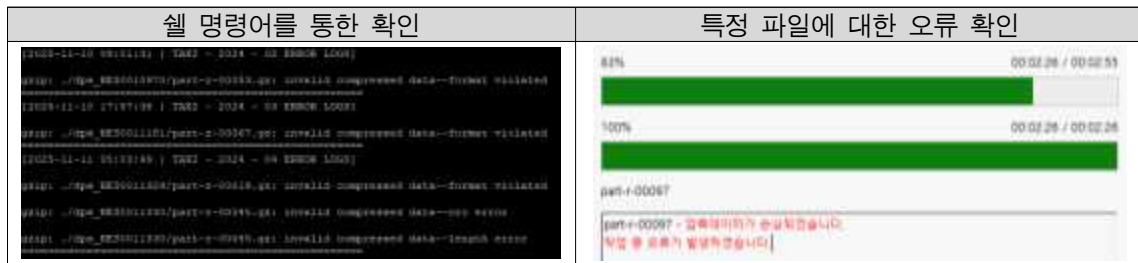
〈표 2-2〉 티맵 필터링 및 보정 데이터 표준 DB

K-V	순번	필드명	필드설명	자료형	비고
K	1	OBU_ID	주행 고유 ID	string	
	2	CAR_MODEL	차종("운전습관 가입자" 대상)	string	
	3	CAR_FUEL	유종("운전습관 가입자" 대상)	string	
	4	DEST_NAME	목적지명	string	
	5	DB_NAME	목적지 중분류 정보(POI가 아닌 주소 검색 등의 경우null)	string	
V	6	TIME	데이터 발생 시각	long	Unixtime
	7	X	X 좌표	double	UTM-K
	8	Y	Y 좌표	double	UTM-K
	9	TMAPID	tmap 링크ID	long	
	10	SPD_CAPPED	CAP이 씌워진 상태로 계산된 속도값	double	

2. 포인트 단위 DTG 데이터 전처리 및 경로 가공

가. DTG 데이터 오류 검토 및 재요청

- DTG 데이터는 추출 및 수집 과정에서 파일 손상, 압축 오류 등 원시 데이터 수준의 오류가 발생할 가능성이 있어, 데이터 적재 전 사전 검증 프로세스를 수행함
 - HDFS 적재 이전 단계에서 각 gzip 압축 파일에 대해 손상 여부를 점검하여 오류 데이터를 사전에 식별함
 - 파일 단위 무결성 검증을 수행하고, 특정 파일에 대한 오류 발생 여부를 확인하고, 오류가 확인된 파일은 재요청 및 재수집을 통해 데이터 품질을 확보함

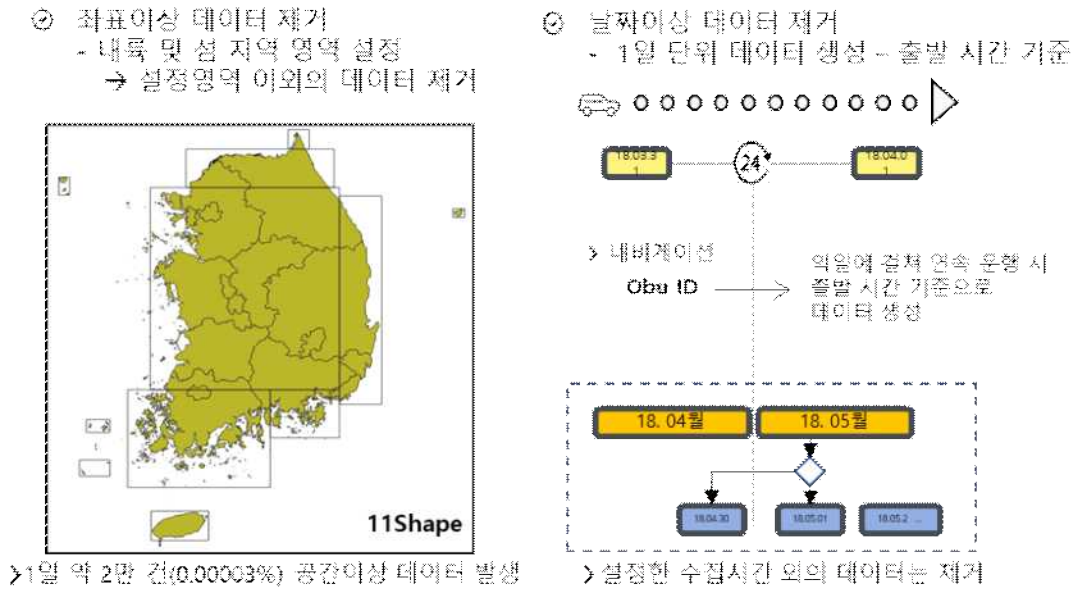


〈그림 3-13〉 압축 파일에 대한 손상 확인

- DTG 데이터는 추출 및 수집 과정에서 일부 파일에 대해 압축 오류 및 손상 문제가 발생하는 것으로 확인됨에 따라, 월별·차종별 데이터 단위로 정상 여부를 점검함

〈표 3-3〉 2024년 DTG 수집 데이터 정상/오류 여부 정리

월	순번	버스		택시		화물	
		구분	오류여부	구분	오류여부	구분	오류여부
1월	1	dps_RES0011186		dps_RES0011158		dps_RES0010972	
	2	dps_RES0011187		dps_RES0011159		dps_RES0010973	
	3	dps_RES0011188		dps_RES0011180		dps_RES0010974	
2월	1	dps_RES0011857	오류	dps_RES0010970	오류	dps_RES0010975	
	2	dps_RES0011858	오류	dps_RES0010945		dps_RES0010949	
	3	dps_RES0011859		dps_RES0012342		dps_RES0010976	
3월	1	dps_RES0011493		dps_RES0011181	오류	dps_RES0010977	
	2	dps_RES0011861	오류	dps_RES0012343		dps_RES0010950	
	3	dps_RES0011862	오류	dps_RES0011183		dps_RES0010951	
4월	1	dps_RES0011863		dps_RES0012344		dps_RES0011279	
	2	dps_RES0011073		dps_RES0011329	오류	dps_RES0011280	
	3	dps_RES0011074		dps_RES0011330	오류	dps_RES0011281	
5월	1	dps_RES0011524		dps_RES0011331		dps_RES0011282	
	2	dps_RES0011076		dps_RES0011332		dps_RES0011283	
	3	dps_RES0011077		dps_RES0011333		dps_RES0011284	오류
...							
9월	1	dps_RES0011499		dps_RES0011569		dps_RES0011606	
	2	dps_RES0011580		dps_RES0011584		dps_RES0011607	
	3	dps_RES0011581	오류	dps_RES0011585		dps_RES0011608	
...							
12월	1	dps_RES0012451		dps_RES0012460		dps_RES0012469	
	2	dps_RES0012452		dps_RES0012461		dps_RES0012470	
	3	dps_RES0012453		dps_RES0012462		dps_RES0012471	



〈그림 3-15〉 좌표오류 및 시간오류 필터링 방법론

- (중복오류) 동일 시간에 여러 데이터가 발생한 경우 → 중복 데이터 제거
- 2024년 DTG 데이터의 전체 이벤트 대비 오류 이벤트 비율은 약 2.11%로 확인됨
 - 연도별로 전체 이벤트 수는 지속적으로 증가하는 반면, 오류 이벤트 수는 감소하여 데이터 품질이 전반적으로 개선된 것으로 분석됨
 - 링크 단위 내비게이션 데이터 대비 오류 비율이 높은 것은 포인트 기반 GPS 좌표 수집 과정에서 오류 발생 가능성이 상대적으로 높기 때문임

〈표 3-4〉 DTG 데이터 오류 발생 필터링 비율

연도	전체 이벤트 수	오류 이벤트 수	필터링 비율
2022년	1,581,062,712,184	73,642,723,727	4.66%
2023년	2,079,495,998,909	61,697,555,110	2.97%
2024년	2,894,111,544,402	60,974,458,298	2.11%

- 통행정보 외 불필요 컬럼을 제거하고 차종 정보를 포함한 전처리를 수행하여 데이터 용량을 효율화함
 - 2024년 전처리 소요기간은 약 34일로, 2023년(약 28일) 대비 데이터 규모 증가에 따라 처리 기간이 증가한 것으로 분석됨
 - 데이터 적재에는 별도로 약 72일이 소요되었으며, 전처리 과정을 통해 전체

데이터 용량은 약 67% 수준으로 축소됨

- 개인정보 보호를 위해 개인식별 정보를 제거하고, 전처리 결과를 DB 형태로 변환·저장함

〈표 3-5〉 DTG 전처리 데이터 DB

K-V	순번	필드명	필드설명	자료형	비고
K	1	OBU_ID	주행 고유 ID	string	
	2	VEH_TYPE	차종	string	
V	3	TIME	포인트 시간	long	Unix Time
	4	X	X좌표	double	UTM-K
	5	Y	Y좌표	double	UTM-K

다. DTG데이터 필터링 및 보정 프로세스

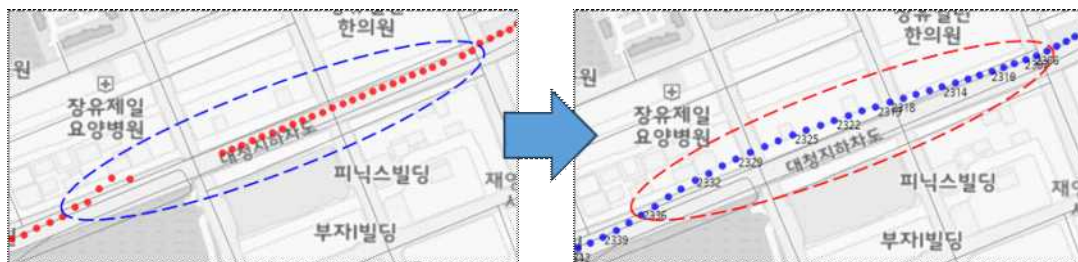
- 전처리 결과로 생성된 DB로부터 궤적별 차량의 출발/도착을 구분하여 분할
 - 차량의 궤적 데이터는 1일 기준으로 데이터가 연속 수집되기 때문에 출·도착 기준의 통행분리가 필요함
 - 포인트 단위인 DTG 데이터의 경우 링크 단위 데이터에 비해 좀 더 상세하게 통행분리 유형을 판별할 수 있음
- 사업용 차량의 출발·경유·도착에 대한 판단 및 통행분리를 위해 차량 유형별로 데이터를 분석했을 때 다음과 같이 구분됨
 - 목적지까지 정차없이 도착하는 통행
 - 장거리 운행 중 휴게소에 정차 및 주유하기 위한 정차나 배송업무 등으로 인해 반복 정차하는 경우 등 경유지를 거쳐 목적지에 도착하는 통행
 - 차고지에서 시동을 끄지않고 정차 및 차고지를 계속 주행하는 경우는 운반한 물건을 적재하는 과정의 통행
 - 회차지에서 정차하지 않고 주행하거나 몇 시간 이상 정차하는 등의 통행



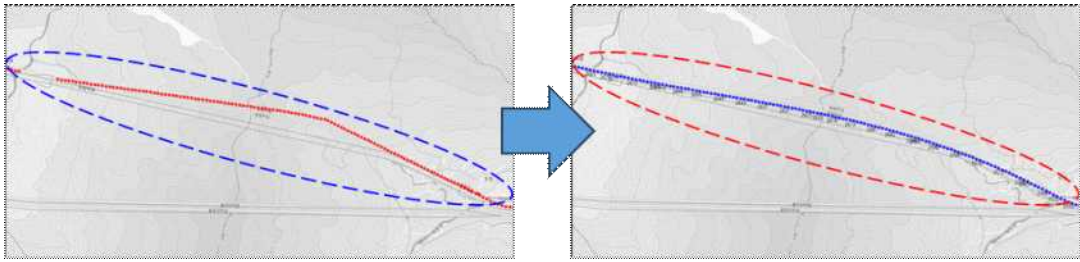
〈그림 3-16〉 통행분리 유형 판단을 위한 데이터 분석

- 사업용 차량 통행은 운행 목적과 패턴이 다양하므로 일정 기준을 적용하여 통행을 분리함
 - 연속 포인트 간 좌표가 동일한 경우, 반경 50m 이내 포인트를 집계하여 300개 이상(약 5분 이상 정지)일 경우 해당 지점을 기준으로 통행을 분리함
 - 버스·화물차는 노선 반복 및 회귀 특성을 고려하여 차고지·회차지 후보를 시간·공간 조건으로 도출하고, 이를 기반으로 통행 분리 지점을 판단함

- GPS 포인트의 오차 범위 및 주변 환경에 따라 발생하는 오류를 줄여 통행 정보, 속도 등의 지표 산출 시 오차를 줄이기 위해 보정 작업을 진행
 - (지하차도, 터널, 고가 밑) 음영 구간에서는 GPS 데이터가 수신되지 않을 수 있으므로, 이 경우 내비게이션은 인위적인 더미 데이터(왼쪽)를 생성함. 이는 실제 차량의 움직임과 다른 데이터이므로 더미 데이터를 제거 후 음영 구간의 데이터를 재생성(오른쪽)하여야 함

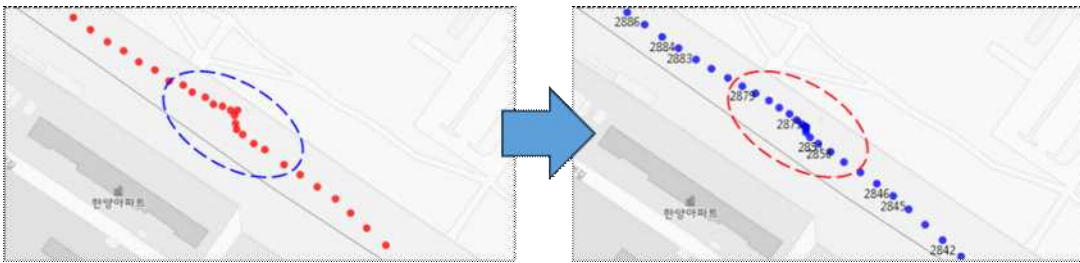


〈그림 3-17〉 지하차도 음영구간의 데이터 재생성



〈그림 3-18〉 터널 음영구간의 데이터 재생성

- (정지구간) 차량이 신호대기 등으로 일시 정차하는 경우 GPS 좌표가 불안정하게 변동(데이터 누락, 반복 순환 패턴 등)하는 오류가 발생하며, 이는 실제 주행과 불일치하므로 감속 - 정지 - 가속 단계로 구간을 재구성하여 정지 구간 데이터를 보정함



〈그림 3-19〉 정지 시 꼬임구간의 데이터 재생성

- DTG 데이터 필터링 및 보정에 약 9일 소요되었으며, 약 12배 용량이 증가함
- 필터링 및 보정 데이터를 표준DB 형태로 변환 및 저장

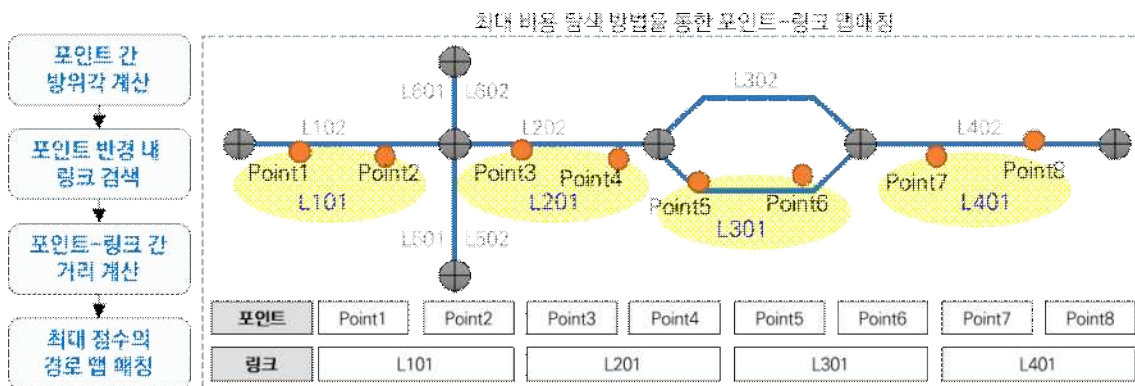
〈표 3-6〉 DTG 필터링 및 보정 데이터 표준 DB

K-V	순번	필드명	필드설명	자료형	비고
K	1	OBUID	통행 ID	string	
	2	VEH_TYPE	차종	string	
	3	AVG_GAP	음영구간을 제외한 원본-Recon 포인트 간 거리 오차의 평균	double	
	4	RELIABILITY1	이상치 판단 후 궤적 신뢰성	double	
	5	RELIABILITY2	필터링 후 궤적 신뢰성	double	
V	6	TIME	포인트 시간	long	Unix Time
	7	X	경도	double	UTM-K
	8	Y	위도	double	UTM-K
	9	SPD	포인트 속도	double	
	10	ACC	포인트 가속도	double	
	11	TYPE	포인트 유형	double	

3. 경로 데이터 구축 및 경로 압축(통합)

가. 상세도로망 기반 경로 데이터 가공 및 DB 구축

- 티맵과 DTG의 상세도로망 기반 경로 데이터 가공 및 DB 구축 프로세스는 동일함
- 본 사업에서 구축하는 기초 교통 및 통행지표 DB의 공간적 범위는 상세도로망 네트워크 기준으로 필터링 및 보정된 표준DB와 도로망 네트워크를 맵매칭 후 링크 기준의 경로 데이터 생성 필요
 - 차량 GPS 데이터의 위치 정보를 기준으로 각 포인트의 진행 방향각 정보 계산. 방향각은 진북을 기준으로 시계방향인 우측 방향각도를 계산
 - 포인트 자료 주변 도로 네트워크(링크) 검색. 검색된 링크를 대상으로 포인트와 링크와의 최단 거리, 링크에서의 방위각 차이, 링크를 따라서 이동한 거리 정보를 계산하여 점수를 부여
 - 모든 포인트는 주변에 검색된 링크를 대상으로 거리 비용, 방위각 비용, 전환 비용을 통해 모든 점수가 최대 비용을 나타내는 경로를 선정 및 적재



〈그림 3-20〉 맵매칭 프로세스

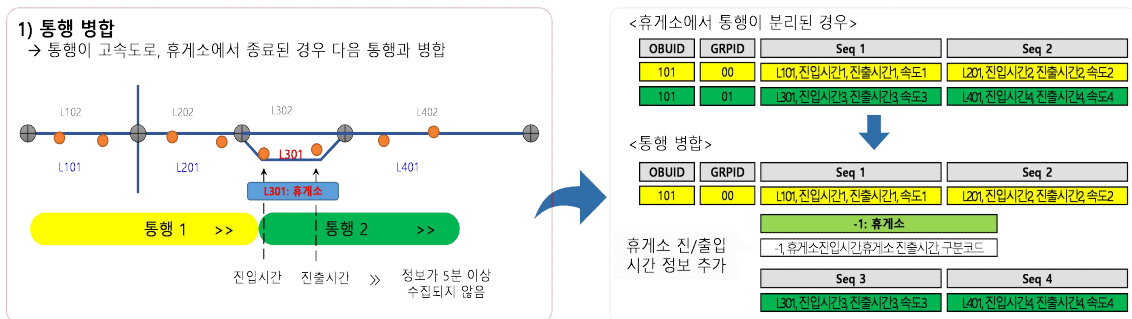
- 맵매칭된 링크 경로 데이터를 통해 링크 통행속도 산출
 - 링크 통과시 포인트의 좌표 정보를 통해 통행속도를 산출
 - 궤적별 포인트-링크 간 매칭된 페어 정보를 통하여 개별 차량의 링크 통행속도를 산출

선택	순번	포인트시간	위도	경도	...	링크ID	링크운행거리	링크남은거리	GIS...	GIS...	GPS...	GPS...
☑	125	2019/05/15 20:35:13	967962.028	1951976.266	2..	57276491201	68.49	162.19	230.68	87.53	195.56	88.00
☑	126	2019/05/15 20:35:14	967985.671	1951984.697	2..	57276491201	93.49	137.20	230.68	87.53	195.56	88.00
☑	127	2019/05/15	968009.154	1951991.78	2..	5727649120		112.68	230.68			00
☑	128	2019/05/15						88.98	230.68			00
☑	129	2019/05/15	968055.856	1952001.929	2..	5727649120		55.26	230.68			00
☑	130	2019/05/15 20:35:18	968079.209	1952005.186	2..	57276491201	189.00	41.68	230.68	87.53	195.56	88.00
☑	131	2019/05/15 20:35:19	968102.594	1952007.398	2..	57276491201	212.06	18.62	230.68	87.53	195.56	88.00
☑	132	2019/05/15 20:35:20	968126.022	1952008.604	2..	57276491001	4.73	234.83	239.56	83.95	233.34	84.00

〈그림 3-21〉 링크 매칭 정보를 통한 통행속도 산출

- 휴게소 또는 고속도로에서 내비게이션 종료 또는 실행 시 통행의 시종점으로 적합하지 않기에, 해당 구간은 연속된 통행을 유지할 수 있도록 통행을 병합함
 - 통행이 병합되는 경우는 휴게소-휴게소, 고속도로-고속도로, 도시고속도로-고속도로 조합의 경우 발생
 - 통행 병합 시 시간 정보 및 구분 코드 정보를 추가하여, 데이터 분석 및 검증 시 활용할 수 있도록 함

• 통행 병합 → 휴게소, (도시)고속도로에서의 연속된 통행 유지



〈그림 3-22〉 통행병합 예시1



〈그림 3-23〉 통행병합 예시2

- 티맵 상세도로망 기반 경로 데이터 가공 및 표준DB 형태로 변환 및 저장

〈표 3-7〉 티맵 상세도로망 기반 경로 데이터 표준 DB

K-V	순번	필드명	필드설명	자료형	비고
K	1	OBU_ID	주행 고유 ID	string	
	2	GRP_ID	통행 그룹 ID	integer	
	3	CAR_MODEL	차종("운전습관 가입자" 대상)	string	
	4	CAR_FUEL	유종("운전습관 가입자" 대상)	string	
	5	DEST_NAME	목적지명	string	
	6	DB_NAME	목적지 중분류 정보(POI가 아닌 주소 검색 등의 경우null)	string	
V	7	vLinkId	가상링크ID	long	
	8	inTime	링크 진입시간	long	
	9	outTime	링크 진출시간	long	
	10	lev6Len	6레벨 링크 길이	double	
	11	lev6Spd	6레벨 링크 속도	double	
	12	gpsLen	링크에 매칭된 포인트 간 길이의 합	double	
	13	tmapSpd	tmap 링크 속도	double	

- DTG 상세도로망 기반 경로 데이터 가공 및 표준DB 형태로 변환 및 저장

〈표 3-8〉 DTG 상세도로망 기반 경로 데이터 표준 DB

K-V	순번	필드명	필드설명	자료형	비고
K	1	OBU_ID	주행 고유 ID	string	
	2	GRP_ID	통행 그룹 ID	integer	
	3	VEH_TYPE	차종	double	
	4	AVG_GAP	음영구간을 제외한 원본-Recon 포인트 간 거리 오차의 평균	double	
	5	RELIABILITY1	이상치 판단 후 궤적 신뢰성	double	
	6	RELIABILITY2	필터링 후 궤적 신뢰성	double	
	7	MATCHING_RATE	포인트-링크 매칭률	double	
	8	CONNECTING_RATE	링크 연결률	double	
V	9	vLinkId		long	
	10	inTime	링크 진입시간	long	
	11	outTime	링크 진출시간	long	
	12	lev6Len	6레벨 링크 길이	double	
	13	lev6Spd	6레벨 링크 속도	double	
	14	gpsLen	링크에 매칭된 포인트 간 길이의 합	double	
	15	tmapSpd	tmap 링크 속도	double	

나. 경로 데이터 압축 및 통합 경로 데이터 구축

○ 개요

- 티맵 내비게이션 데이터와 DTG 데이터는 동일한 기준과 절차로 경로 데이터를 압축·통합 구축하며, 대용량 데이터 처리 효율성과 온라인 서비스 성능 최적화를 목적으로 수행함

○ 구축 방법

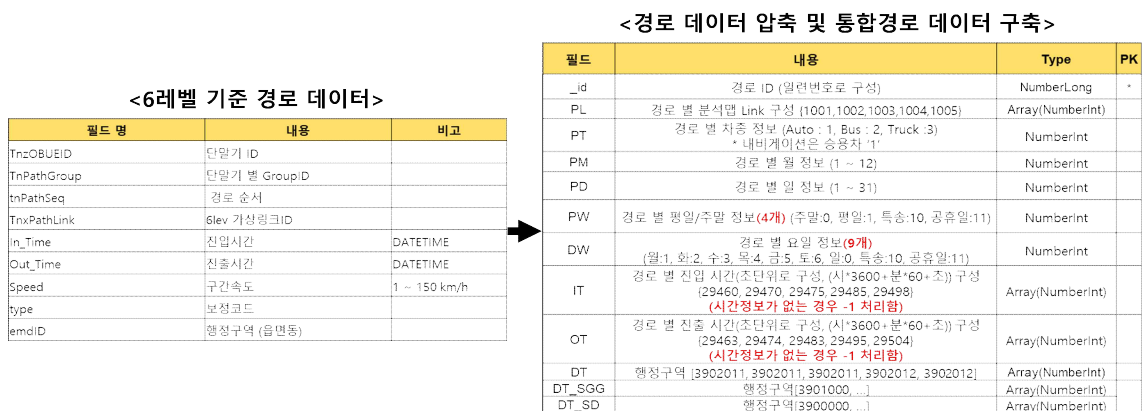
- 상세도로망(Level6) 단위 경로 데이터는 대용량으로 인해 검색 및 조회 성능에 한계가 있어, 주요도로망(Level5.5) 단위로 집계·압축하여 서비스용 통합 경로 데이터로 재구성함
- 이 과정에서 공간적 연속성과 경로 특성은 유지하되, 불필요한 세분 정보를 축약하여 데이터 활용성과 처리 효율성을 확보함

○ 성능 및 용량 개선 효과

- 상세도로망 기준 약 610GB 데이터를 150GB 수준으로 압축하여 약 75% 이상의 용량을 절감하고, 데이터 경량화를 통해 온라인 서비스의 검색 속도 및 응답 성능 개선

○ 저장 및 운영 체계

- 압축·통합된 경로 데이터는 대용량 분산처리와 확장성이 용이한 NoSQL 구조로 저장하여 안정적인 데이터 조회 및 서비스 제공이 가능한 운영 환경을 구축함



《그림 3-24》 압축된 경로 데이터 형태

〈표 3-9〉 경로 데이터 압축 및 통합 경로 데이터 NoSQL DB 포맷

순번	필드명	필드설명	자료형	비고
1	_id	경로ID	NumberLong	
2	PL	경로별 분석맵 링크 구성	Array(NumberInt)	
3	PT	경로별 차종 정보	NumberInt	
4	PM	경로별 월 정보	NumberInt	
5	PD	경로별 일 정보	NumberInt	
6	PW	경로별 평일/주말 정보	NumberInt	
7	DW	경로별 요일 정보	NumberInt	
8	IT	경로별 진입 시간	Array(NumberInt)	
9	OT	경로별 진출 시간	Array(NumberInt)	
10	DT	행정구역_읍면동	Array(NumberInt)	
11	DT_SGG	행정구역_시군구	Array(NumberInt)	
12	DT_SD	행정구역_시도	Array(NumberInt)	

제3절 교통량 추정 DB 구축

가. 교통량 추정 목적

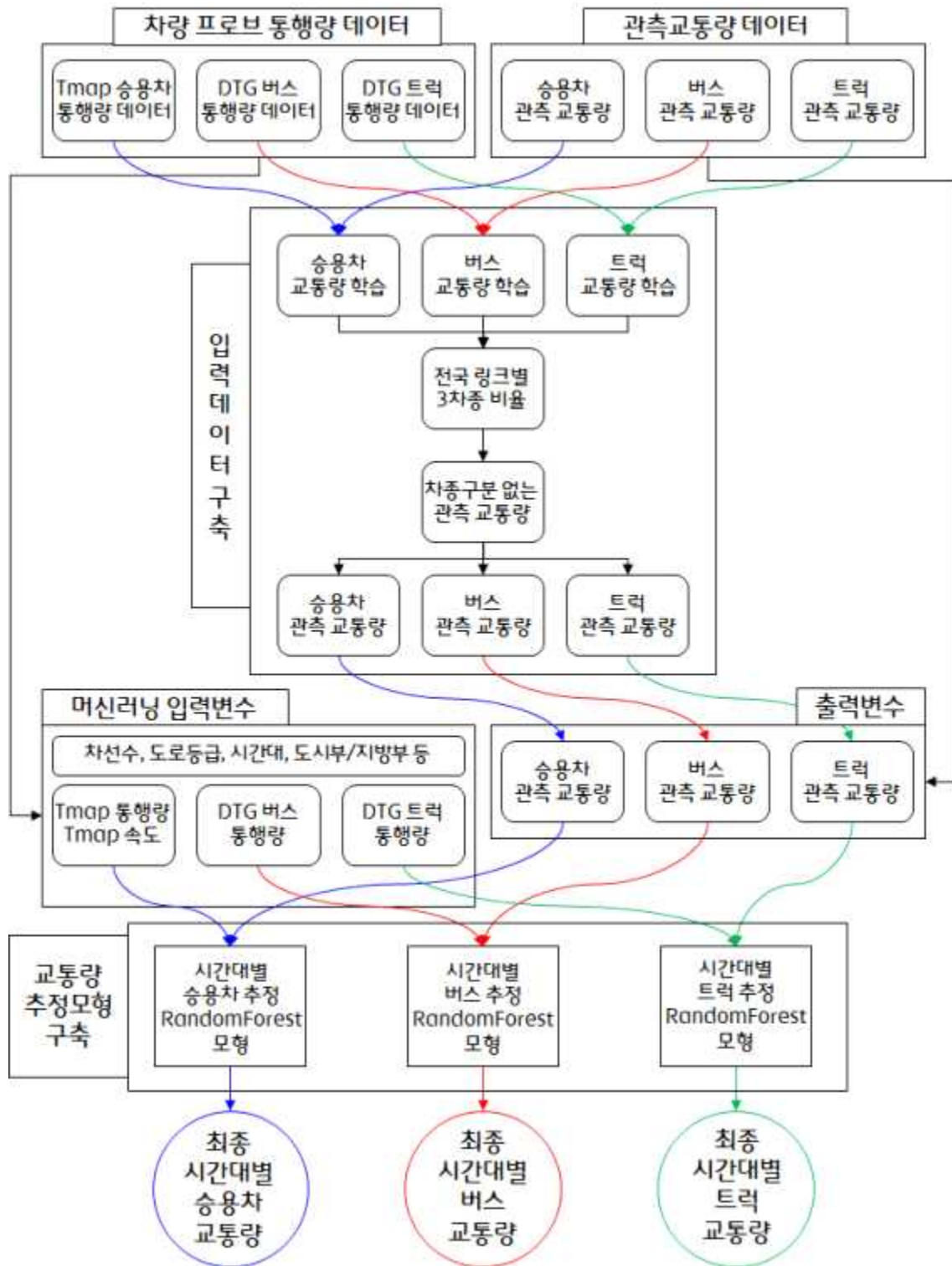
- 전국 관측 교통량은 약 1만 개 지점 수준으로, 상세도로망(Level6) 링크 수(약 62만 개) 대비 약 1.6%에 불과하여 공간적 커버리지가 매우 제한적임
- 또한 지자체 관측교통량은 차종 미구분, 일부 시간대 미관측 등 데이터 제약이 존재하여 정책·운영에 필요한 차종별·시간대별 교통현황 파악에 한계가 있음
 - 이에 따라 제한된 관측자료를 보완·확장하여 전국 도로구간 단위의 차종별·시간대별 교통량을 추정함으로써, 교통수요 분석, 혼잡관리, 신호운영 및 교통정책 수립을 지원하기 위한 기초 데이터 확보가 필요함

나. 교통량 추정 프로세스

- 교통량 추정은 차량 프로브 데이터와 관측교통량 데이터를 결합한 단계적 분석 절차를 통해 수행됨
 - Step ① (데이터 구축 및 정합화) 프로브 통행량과 관측교통량 데이터를 통합·정합화하여 분석 기반 데이터를 구축함
 - Step ② (차종 정보 보완) 차종 정보가 없는 관측지점 및 링크를 대상으로 차종을 추정하여 데이터의 완전성을 확보함
 - Step ③ (입력변수 구축) 교통량, 속도, 네트워크 특성 등 교통량 추정을 위한 머신러닝 입력변수를 구축함
 - Step ④ (추정 모형 개발) 머신러닝 기반 차종별·시간대별 교통량 추정 모형을 개발 및 학습함
 - Step ⑤ (검증 및 보정) 연도별 비교 및 관측값 대비 추정값 검증을 통해 모델의 신뢰도를 평가하고 보정함
 - Step ⑥ (최종 산출) 전국 도로구간 단위의 시간대별·차종별 교통량을 산출함


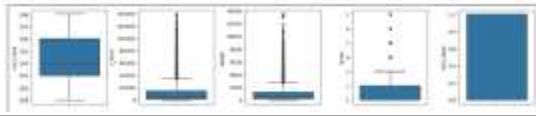
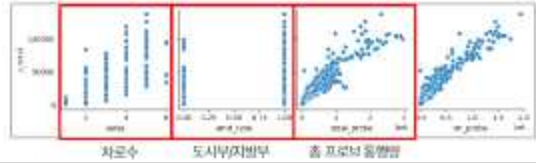
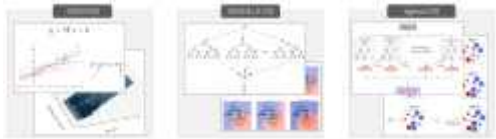
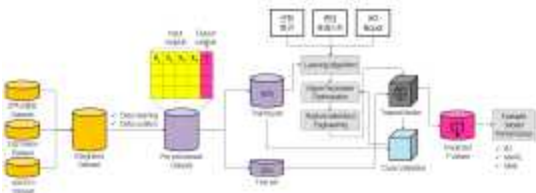
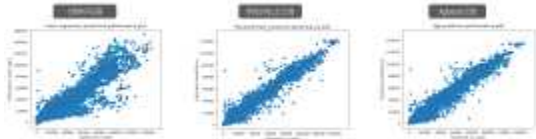


다. 교통량 추정 입력 데이터 구축 프로세스

- 교통량 추정을 위한 입력 데이터 구축
 - 상세도로망(Level6) 네트워크 기반 경로 데이터로부터 생성된 속도 프로파일 DB를 활용하여 교통량 추정 입력 데이터 DB를 구축함
 - 데이터 정합성 확보를 위해 시간대 및 링크 기준으로 데이터 구조를 통일하여 분석 가능한 형태로 재구성함
- 차종별 통행량 및 속도 데이터 생성
 - 차종별 링크 단위 속도빈도분포에서 평균속도를, 통행량 데이터에서 차종별 통행량을 추출하여 전 링크에 대한 차종별 통행량 및 속도 데이터를 생성함
- 관측 링크 및 전체 링크 대상 입력 데이터 가공
 - 관측교통량이 존재하는 링크를 대상으로 머신러닝 학습에 활용되는 관측지점 입력 데이터를 가공함
 - 학습된 모델을 기반으로 관측지점을 포함한 전체 링크의 교통량을 추정하기 위해 전 링크 대상 입력 데이터를 구축함
- 학습데이터 구성
 - 약 1만 개 관측지점을 학습지점과 검증지점으로 구분하고, 이 중 약 8,500개 지점을 학습데이터로 활용함
 - 학습데이터는 차종별 통행량, 속도, 도로 특성(차선수, 도로등급 등)과 함께 네트워크 연계 정보를 포함하도록 구성함
 - 티맵 및 DTG 경로 데이터의 연속 주행 기록을 활용하여 대상 관측지점과 연결된 인접 지점 정보를 추가 변수(네트워크 연결성 지표)로 생성함
 - 개별 차량 경로를 기반으로 통행량, 차선수, 도로등급 등을 고려하여 최대 10개의 인접 후보 지점을 도출하고, 이 중 연결성이 높은 3개 지점을 선별하여 네트워크 연결성 변수로 반영함
 - 이를 통해 단일 링크가 아닌 인접 링크 간 상호 영향성을 반영하여 교통량 추정 정확도를 향상시킴



〈그림 3-25〉 교통량 추정 및 구축 프로세스

〈표 3-10〉 머신러닝 기반 추정교통량 모형 구축 프로세스

구분	설명	예시																																
1차가공DB를 활용한 입력변수 구축	관측교통량, 차량 프로브 통행량, 평균속도 등 교통량 추정을 위한 입력 변수 구축																																	
데이터 전처리	수집된 데이터의 이상치 검토 및 전처리																																	
탐색적 데이터분석	변수별 상관관계 분석을 통한 입력 변수 선정																																	
입력변수 선정	도로등급, 제한속도, 도로연장, 차로수, 도시부/지방부, 차량프로브 통행량, 평균속도	<table border="1"> <thead> <tr> <th>번호</th> <th>변수명</th> <th>유형명</th> <th>설명</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>road_rank</td> <td>도로등급</td> <td>고속도로, SA(국도/도도), 일반국도, 특별광역시/도, R(시도), 지방도, 시군도, 민중로</td> </tr> <tr> <td>2</td> <td>max_speed</td> <td>제한속도</td> <td></td> </tr> <tr> <td>3</td> <td>length</td> <td>도로연장</td> <td>미터</td> </tr> <tr> <td>4</td> <td>lanes</td> <td>차로수</td> <td>단방향 차로수</td> </tr> <tr> <td>5</td> <td>emid_type</td> <td>도시/지방 코드</td> <td>도(시/군) 1, 지방(시, 군) 0</td> </tr> <tr> <td>6</td> <td>total_probe</td> <td>전체 프로브</td> <td>티엠 정량량+티엠크웨이 정량량+트릭 정량량+버스 정량량+택시 정량량</td> </tr> <tr> <td>7</td> <td>tm_spd</td> <td>티엠 속도</td> <td>티엠 연평균 평균 속도</td> </tr> </tbody> </table>	번호	변수명	유형명	설명	1	road_rank	도로등급	고속도로, SA(국도/도도), 일반국도, 특별광역시/도, R(시도), 지방도, 시군도, 민중로	2	max_speed	제한속도		3	length	도로연장	미터	4	lanes	차로수	단방향 차로수	5	emid_type	도시/지방 코드	도(시/군) 1, 지방(시, 군) 0	6	total_probe	전체 프로브	티엠 정량량+티엠크웨이 정량량+트릭 정량량+버스 정량량+택시 정량량	7	tm_spd	티엠 속도	티엠 연평균 평균 속도
번호	변수명	유형명	설명																															
1	road_rank	도로등급	고속도로, SA(국도/도도), 일반국도, 특별광역시/도, R(시도), 지방도, 시군도, 민중로																															
2	max_speed	제한속도																																
3	length	도로연장	미터																															
4	lanes	차로수	단방향 차로수																															
5	emid_type	도시/지방 코드	도(시/군) 1, 지방(시, 군) 0																															
6	total_probe	전체 프로브	티엠 정량량+티엠크웨이 정량량+트릭 정량량+버스 정량량+택시 정량량																															
7	tm_spd	티엠 속도	티엠 연평균 평균 속도																															
모형 검토	선형회귀모형, 랜덤포레스트모형, XGboost 모형 선정																																	
교통량 추정모형 구축	전처리된 데이터와 선정된 모형을 활용하여 교통량 추정모형 구축																																	
교통량 추정 및 평가	구축된 교통량 추정의 추정교통량 신뢰도 분석																																	
교차검증	최종 선정된 랜덤포레스트 모형에 대한 교차검증을 통해 일반화 성능 검토																																	
변수 분석	교통량 추정에 영향을 끼치는 변수별 영향 분석																																	

라. 교통량 추정을 위한 입력 데이터 구축 결과

- 관측링크와 전체링크의 입력 데이터는 동일한 구조로 구축하여 모델 학습과 적용 간 일관성을 확보함
- 입력 데이터는 교통량 추정 알고리즘 개발 과정에서 수행한 탐색적 분석을 통해 선정된 핵심 변수로 구성함

〈표 3-11〉 교통량 추정을 위한 입력 데이터 DB

순번	필드명	필드설명	자료형
1	v_link_id	가상링크ID	Long
2	traffic_type	조사지점 코드	integer
3	length	도로연장	string
4	lanes	차선수	string
5	road_rank	도로등급	Integer
6	time_code	시간코드	integer
7	v_total_wd	전체 관측교통량	double
8	v_auto_wd	승용 관측교통량	double
9	v_bus_wd	버스 관측교통량	double
10	v_truck_wd	트럭 관측교통량	double
11	total_probe	전체 통행량	integer
12	tm_probe	티맵 통행량	integer
13	truck_probe	버스 통행량	integer
14	bus_probe	트럭 통행량	integer
15	taxi_probe	택시 통행량	integer
16	tm_spd	티맵 속도	double
17	truck_spd	버스 속도	double
18	bus_spd	트럭 속도	double
19	taxi_spd	택시 속도	double
19	emd_type	행정구역 동 여부	Integer
21	tlfd_code	도로의 기능 구분 코드	Integer
22	length	도로길이	double
23	vkt	총도로연장	double
24	k	티맵 밀도	double
25	sido_id	시도ID	Integer
26	tm_scale	승용 스케일	double
27	bus_scale	버스 스케일	double
28	truck_scale	트럭 스케일	double
29	volume_1	연결된 관측교통량1	Double
30	probe_1	연결된 프로브통행량1	Double
31	volume_2	연결된 관측교통량2	Double
32	probe_2	연결된 프로브통행량2	Double
33	volume_3	연결된 관측교통량3	Double
34	probe_3	연결된 프로브통행량3	Double

마. 추정교통량 구축

○ 추정교통량 구축 범위는 다음과 같음

- 기초 교통 DB를 기반으로 전국 2차로 이상 도로를 대상으로 추정교통량 DB를 구축함
 - 구축 범위는 차종(승용차, 버스, 트럭), 요일패턴(평일, 주말), 시간(전일 및 시간대별), Level6 링크 단위로 설정함
 - 집계 단위는 상세도로망, 주요도로망, 표준 노드·링크, 행정구역(시도/시군구/읍면동) 등으로 확장 가능하도록 구성함
- 관측지점을 대상으로 추정 결과와 실측 교통량을 비교하고, 연도별 추이 분석을 통해 추정교통량의 신뢰성을 검증함

〈표 3-12〉 추정교통량 DB 테이블 구성

No	Column	Type	설명	코드	코드정보
1	V_link_id	Integer	가상링크ID	-	-
2	Week_code	Integer	요일코드	0, 1	주말, 평일
3	time_code	Integer	시간코드	1-24	24시간
4	total_veh	Double	교통량(전체)	-	-
5	auto	Double	교통량(승용차)	-	-
6	bus	Double	교통량(버스)	-	-
7	truck	Double	교통량(화물차)	-	-

바. 교통량 추정 검증 결과

- (목적) 관측교통량이 존재하는 지점을 기준으로 추정교통량과 실측값을 비교하여, 추정 결과의 현실 재현 수준과 신뢰도를 평가함
- 평가 기준 및 방법
 - 조사교통량 대비 오차율을 기반으로 신뢰도를 평가하며, 한국개발연구원 (2008) 기준을 적용하여 ±15% 이내를 유의한 추정 범위로 설정함
 - 추가적으로 ITS 성능평가 기준(VDS검지기)을 적용하여 교통정보의 정확도 및 신뢰도 확보 여부를 함께 검증함
 - 분석 대상은 1시간 기준 200대 이상이며, 관측교통량과 추정교통량이 모두 존재하는 도로구간으로 설정함
 - 추정교통량의 신뢰도 평가지표는 평균절대비율오차(Mean Absolute Percent Error, MAPE, %)를 활용함

$$MAPE(\%) = \frac{\sum_{i=1}^n \left(\frac{|q_i - \hat{q}_i|}{q_i} \right) \times 100}{n},$$

여기서, q_i = 지점 i 의 조사교통량(대/일)

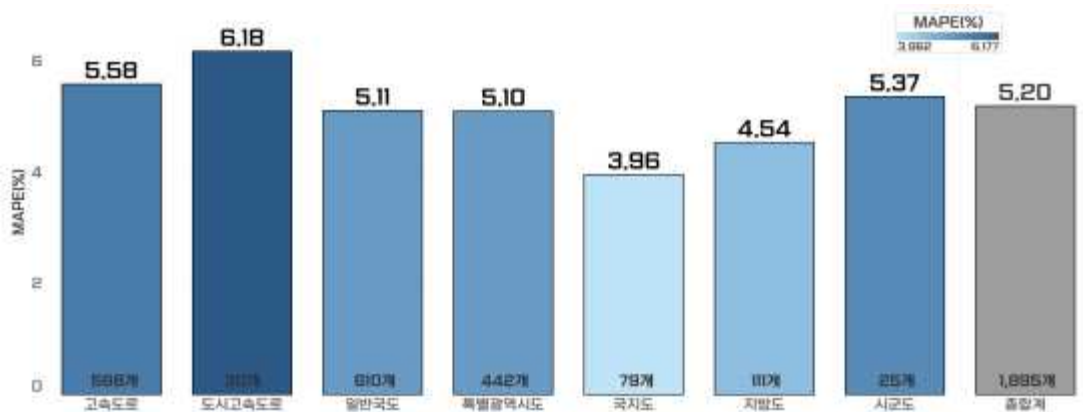
\hat{q} = 지점 i 의 추정교통량(대/일)

n = 전체 지점 개수(개)

- 검증 결과
 - 도로등급별 오차율(MAPE) 분석 결과, 고속도로 5.6%, 일반국도 5.1%, 특별·광역시도 5.1%로 나타남
 - 전반적으로 약 5% 수준의 오차율을 보이며 ITS 성능평가 기준 상급 수준을 충족함
 - 또한 교통량이 증가할수록 오차율이 감소하는 경향이 확인됨

○ 시사점

- 관측지점 기반 교통량을 전국 도로망으로 확장·추정하는 데 있어 높은 신뢰성과 적용 가능성을 확보함
- 특히 주요 간선도로 및 교통량 수준이 높은 구간에서는 정책 수립 및 운영의 사결정에 활용 가능한 수준의 정확도를 확보함
- 향후 저수준 교통량 및 미관측 구간에 대한 추가 보정 및 학습데이터 확대를 통해 추정 정확도와 활용 범위를 지속적으로 확대할 필요가 있음

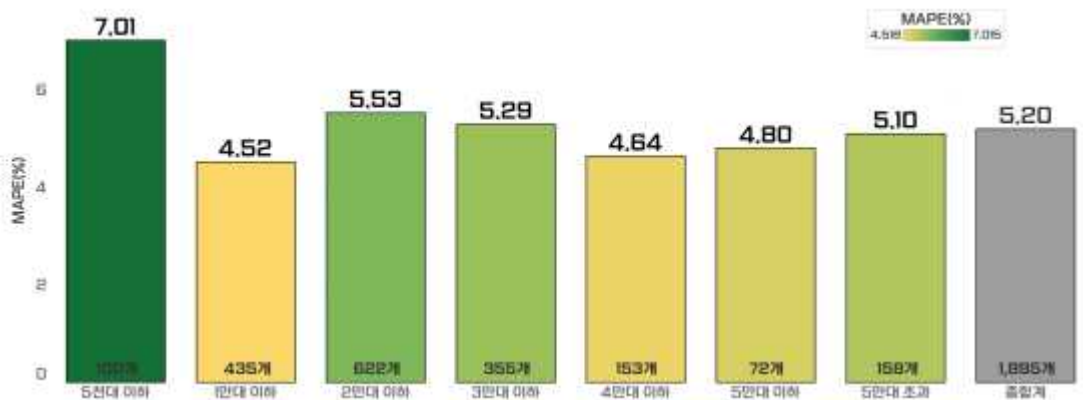


※ 참고1. 평가지점은 ITS 성능평가 기준(1시간 기준 200대 이상인 도로를 대상)으로 선정함

참고2. MAPE 목표치는 한국개발연구원의 직접 영향권 범위 참고함

참고3. 2024년 추정교통량 기준 분석결과

〈그림 3-26〉 도로 등급별 신뢰도 분석 결과



※ 참고1. 평가지점은 ITS 성능평가 기준(1시간 기준 200대 이상인 도로를 대상)으로 선정함

참고2. MAPE 목표치는 한국개발연구원의 직접 영향권 범위 참고함

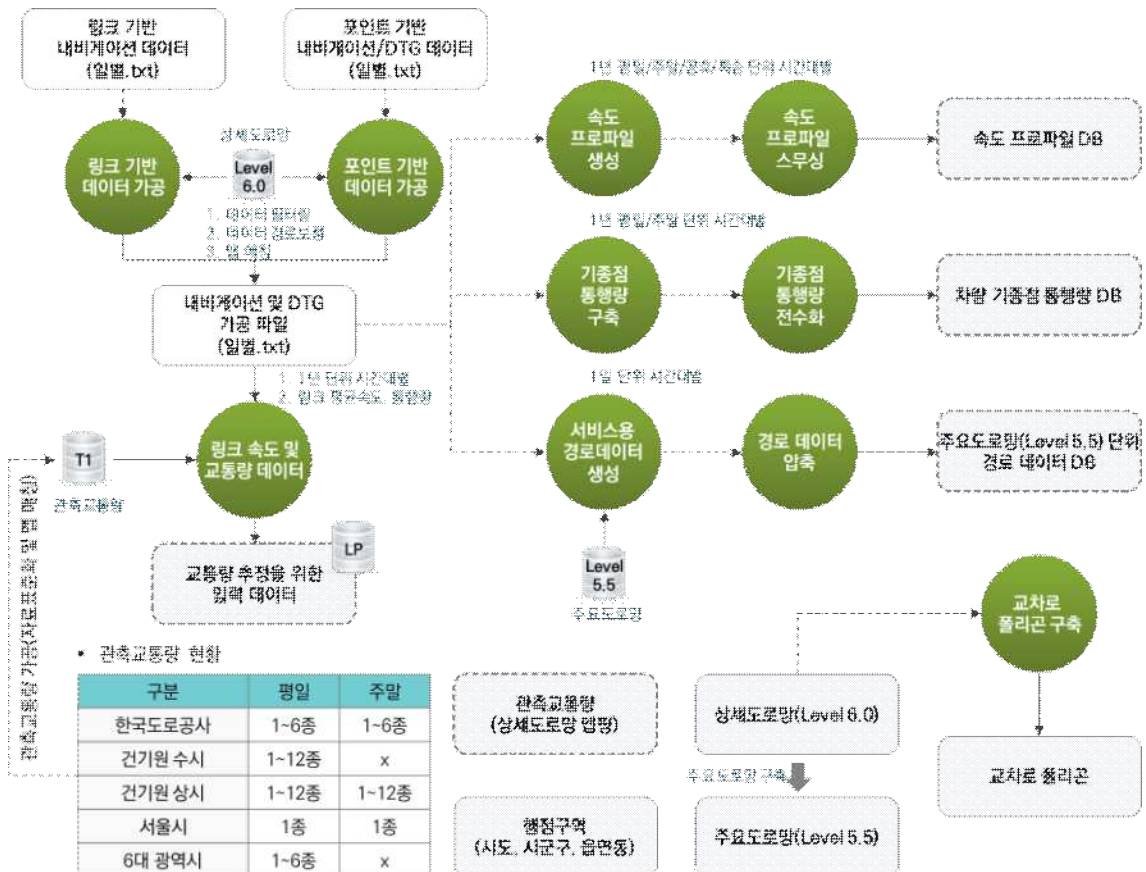
참고3. 2024년 추정교통량 기준 분석결과

〈그림 3-27〉 교통량 수준별 신뢰도 분석 결과

제4절 기초교통 DB 구축

1. 기초교통 DB 구축 개요

- 전국 단위 교통량 전수화 및 통행지표 산출을 위해 경로 데이터와 관측교통량을 활용한 1차 가공 기반의 기초교통 DB를 구축함
 - 경로 데이터의 속도 정보를 활용하여 시간대별 속도빈도 분포를 구축함
 - 차량의 기종점(OD) 기반 통행량 DB를 구축하여 이동 특성을 반영함
 - 링크별 통행량 및 속도 정보를 결합하여 교통량 추정을 위한 입력 데이터 DB를 구축함

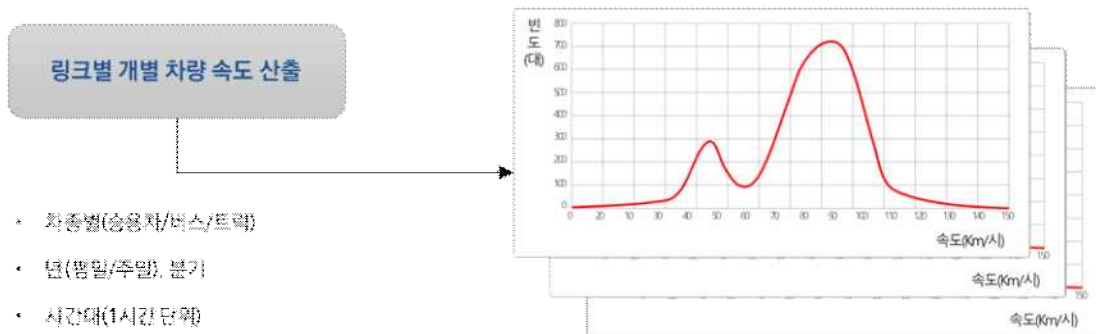


〈그림 3-28〉 기초교통 DB 구축 프로세스

2. 속도빈도분포 DB 구축

가. 속도빈도분포 구축 프로세스

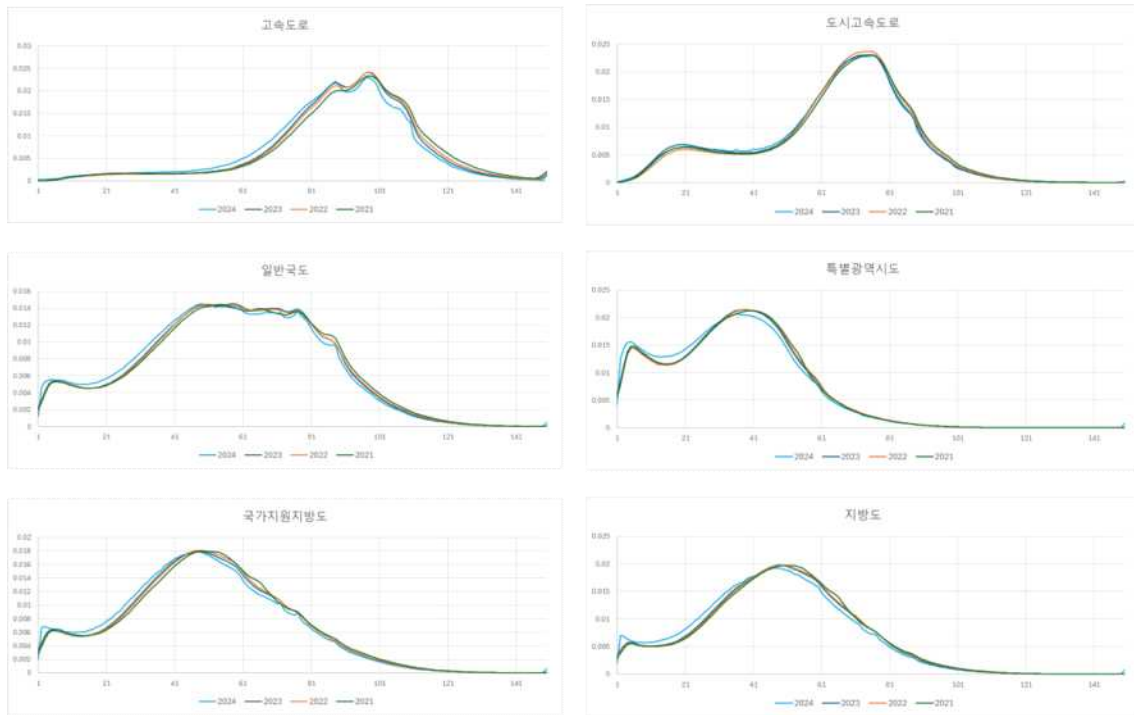
- 링크별 속도빈도분포 집계
 - 교통량 추정 DB와 통행지표 구축을 위하여 속도빈도분포 DB를 구축
 - 속도지표, 혼잡지표, 환경지표 등 다양한 통행지표 등에 활용
 - 링크에 매칭된 궤적 정보를 이용하여 개별차량의 속도 산출 및 산출된 속도 범위에 따른 집계를 수행
 - 링크별로 1km/h 단위의 통행속도(1km/h ~ 150km/h, 초과시 150km/h)의 통행량 집계
 - 개별 링크의 연/평일주말/1시간 단위의 속도빈도분포 구축



〈그림 3-29〉 속도빈도분포 구축 프로세스

나. 속도빈도분포 구축 결과

- 연도별 도로등급별 속도빈도분포 비교
 - 2021년 ~ 2024년의 도로등급별 속도빈도분포를 비교분석한 결과 연도별 비슷한 패턴이 유지됨을 확인함
 - 2021년 ~ 2023년에 비해 2024년의 도로등급별 속도빈도분포 데이터는 전체적인 패턴은 유사하나 저속 구간의 비율이 증가한 것으로 확인됨



〈그림 3-30〉 연도별 도로등급별 속도 빈도분포 결과 비교

○ 속도빈도분포 DB 형태

〈표 3-13〉 속도빈도분포 DB 포맷

순번	필드명	필드설명	자료형	비고
1	v_link_id	상세도로망 가상링크ID	long	
2	quarter_code	분기코드	int	1 ~4
3	week_code	주중코드	int	1: 평일 2: 주말
4	time_code	시간코드	int	0: 전일 1~24: 시간대
5	road_name	도로명	string	
6	road_rank	도로등급	int	
7	lanes	차선수	int	
8	length	도로길이	double	
9	emd_id	읍면동ID	int	
10	avg_spd	평균속도	double	
11	med_spd	중위속도	double	
12	probe	통행량	int	
13	spd_0kmh	0km/h 속도 통행량	int	
14	spd_1kmh	1km/h 이하 속도 통행량	int	
...	
163	spd_150kmh	149km/h 속도 초과 통행량	int	

○ 도로등급별 평균속도 비교분석

- 2023년 대비 2024년 평균속도는 모든 도로등급에서 약 3.53% 감소함
- 2023년 대비 2024년 평균속도가 가장 크게 감소한 도로등급은 고속도로로 약 8.12%의 감소폭을 보였음

<표 3-14> 연도별 도로등급별 평균속도

단위 : km/h

도로등급	평균속도				'21년 대비 '22년 증감율	'22년 대비 '23년 증감율	'23년 대비 '24년 증감율
	2021년	2022년	2023년	2024년			
고속도로	97.70	95.69	94.03	86.39	-2.07%	-1.73%	-8.12%
도시고속도로	66.34	66.49	65.20	63.33	0.23%	-1.94%	-2.87%
일반국도	60.01	59.51	58.99	57.84	-0.83%	-0.87%	-1.94%
특별광역시도	26.82	26.66	26.32	25.42	-0.59%	-1.28%	-3.41%
국가지원지방도	48.99	48.47	48.01	47.18	-1.06%	-0.94%	-1.72%
지방도	45.33	44.94	44.43	43.48	-0.87%	-1.13%	-2.14%
시군도	31.48	31.12	30.82	29.61	-1.13%	-0.98%	-3.91%
총계	35.97	35.58	35.23	33.99	-1.07%	-1.00%	-3.53%

* 데이터는 절대적으로 참값이 아니며, 지속적으로 신뢰도를 높이기 위한 정산작업 수행

○ 시도별 평균속도 비교분석

- 연도별 평균속도를 분석했을 때 2021년 35.97km/h, 2022년 35.58km/h, 2023년 35.23km/h, 2024년 33.99km/h로 매년 속도가 감소하고 있음
- 2023년 대비 2024년 평균속도가 가장 크게 감소한 지역은 세종으로 약 4.19%, 그 다음 인천에서 약 4.17% 감소하였음
- 나머지 시도에서도 2023년 대비 2024년 평균속도가 감소하여 전체 시도별 평균속도는 3.53% 감소하였음

〈표 3-15〉 연도별 시도별 평균속도

단위 : km/h

시도	평균속도				'21년 대비 '22년 증감율	'22년 대비 '23년 증감율	'23년 대비 '24년 증감율
	2021년	2022년	2023년	2024년			
서울	24.75	24.80	24.49	23.77	0.21%	-1.25%	-2.94%
부산	29.73	29.54	29.10	28.07	-0.62%	-1.49%	-3.54%
대구	30.44	30.48	31.26	30.22	0.13%	2.54%	-3.32%
인천	30.04	29.73	29.40	28.17	-1.04%	-1.12%	-4.17%
광주	29.29	28.99	28.75	27.86	-1.02%	-0.84%	-3.08%
대전	31.12	30.78	30.33	29.53	-1.11%	-1.44%	-2.63%
울산	33.62	33.30	32.83	31.66	-0.93%	-1.41%	-3.55%
세종	34.73	34.12	33.96	32.54	-1.76%	-0.48%	-4.19%
경기	33.18	32.80	32.47	31.22	-1.17%	-1.01%	-3.82%
강원	39.19	38.75	38.50	37.07	-1.13%	-0.64%	-3.72%
충북	39.32	38.67	38.32	37.01	-1.63%	-0.90%	-3.41%
충남	40.02	39.28	38.94	37.57	-1.85%	-0.88%	-3.52%
전북	39.29	38.88	38.54	37.20	-1.03%	-0.87%	-3.47%
전남	40.87	40.49	40.07	38.90	-0.92%	-1.04%	-2.92%
경북	40.85	40.39	39.82	38.54	-1.13%	-1.41%	-3.22%
경남	37.08	36.77	36.39	35.13	-0.83%	-1.05%	-3.45%
제주	35.13	34.81	34.56	33.42	-0.91%	-0.71%	-3.31%
총계	35.97	35.58	35.23	33.99	-1.07%	-1.00%	-3.53%

* 데이터는 절대적으로 참값이 아니며, 지속적으로 신뢰도를 높이기 위한 정산작업 수행

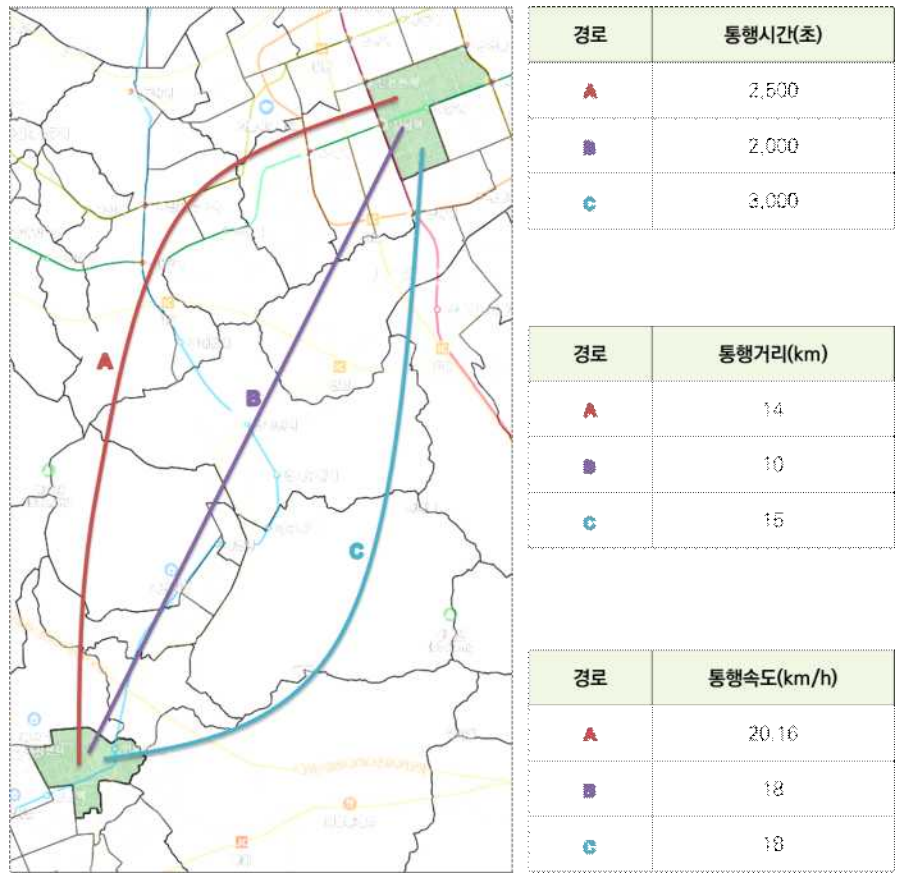
3. 기종점 통행 DB 구축

가. 기종점 통행 DB 구축 프로세스

- 기종점 통행 DB는 일자 단위 개별 통행 DB 구축과 연단위 집계 DB 구축의 2 단계로 구성됨
 - 일자별 차량별 기종점 통행DB를 구축한 후, 이를 집계하여 차종별 통행시간, 통행거리, 통행속도 등을 포함한 통합 DB로 확장함
- 일자별 개별 차량 기종점 통행 DB 구축
 - 개별 차량의 링크 경로 데이터를 활용하여 차량별 기종점(Origin-Destination) 통행 정보를 생성함
 - 시간적 범위: 일자 단위
 - 공간적 범위: 전국 읍·면·동 간 통행
- 연단위 차종별 기종점 통행 DB 구축
 - 일자별 통행 DB를 기반으로 연단위 집계를 수행하여 차종별 기종점 통행 DB를 구축함
 - 시간적 범위: 연단위(평일/주말/공휴일/특송 구분, 전일 및 시간대별)
 - 공간적 범위: 전국 행정구역(시도/시군구/읍면동) 간 통행

나. 일자별 차종별 기종점 통행 DB 구축

- 경로별 출발 링크와 도착 링크의 읍면동 정보, 노드 정보, 링크종별 정보를 생성함
 - 총 통행시간 = 도착 링크의 진출 시간 - 출발 링크의 진입 시간 (링크종별 16인 휴게소 제외)
 - 총 통행거리 = obu_id 및 group_id별 경로 통행 링크 길이의 합계
 - 통행속도 = 총 통행거리 / 총 통행시간



〈그림 3-31〉 개별 차량별 기종점 통행 DB 계산 예시

〈표 3-16〉 개별 차량별 기종점 통행 DB 포맷

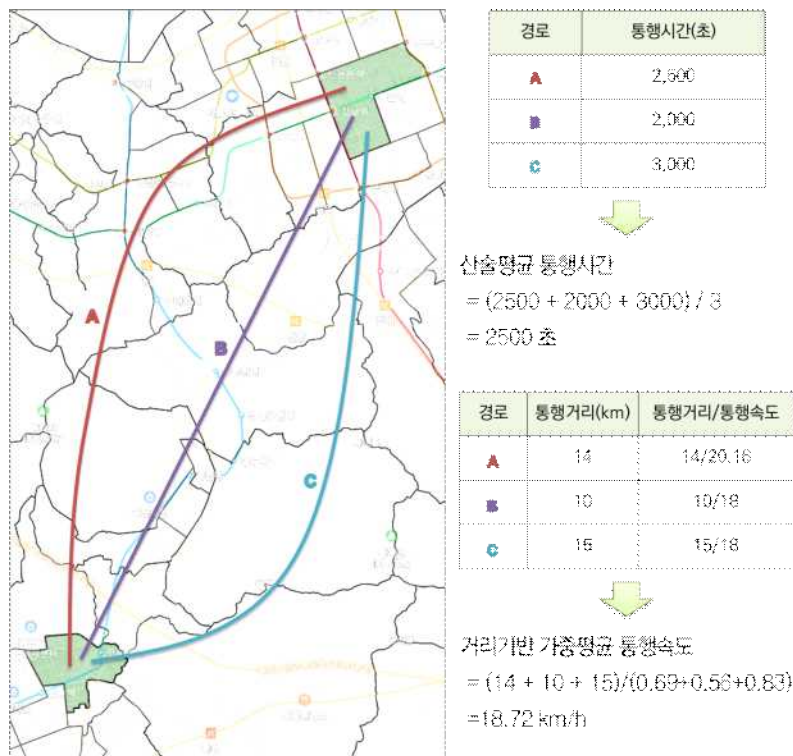
순번	필드명	필드설명	자료형	비고
1	obu_id	개별차량ID	VARCHAR	
2	group_id	통행그룹ID	INT	
3	o_link_id	기점 링크ID	BIGINT	
4	f_node	기점 노드ID	BIGINT	
5	f_node_x	기점 노드 X좌표	DOUBLE	
6	f_node_y	기점 노드 Y좌표	DOUBLE	
7	o_emd_code	기점 읍면동 코드	INT	
8	o_datetime	기점 날짜시간	DATETIME	연월일 시분초
9	o_time	기점 시간(초)	INT	0 ~ 86400
10	d_link_id	종점 링크ID	BIGINT	
11	t_node	종점 노드ID	BIGINT	
12	t_node_x	종점 노드 X좌표	DOUBLE	
13	t_node_y	종점 노드 Y좌표	DOUBLE	

순번	필드명	필드설명	자료형	비고
14	d_emd_code	중점 읍면동 코드	INT	
15	d_time	중점 시간(초)	INT	
16	total_traveltime	총 통행시간	DOUBLE	
17	total_distance	총 통행거리	DOUBLE	
18	avg_spd	평균속도	DOUBLE	

다. 차종별 기종점 통행 DB 구축

○ 경로별 출발 링크와 도착 링크의 행정구역 정보를 생성함

- 산술평균 통행시간 = 총 통행시간 합계 / 개별 차량 경로 수
- 산술평균 통행거리 = 총 통행거리 합계 / 개별 차량 경로 수
- 거리기반 가중 조화평균 통행속도 = 총 통행거리 합계 / '총 통행거리 / 통행 속도'의 합계
- 속도 계산 시, 속도가 있는 링크 정보만 사용함



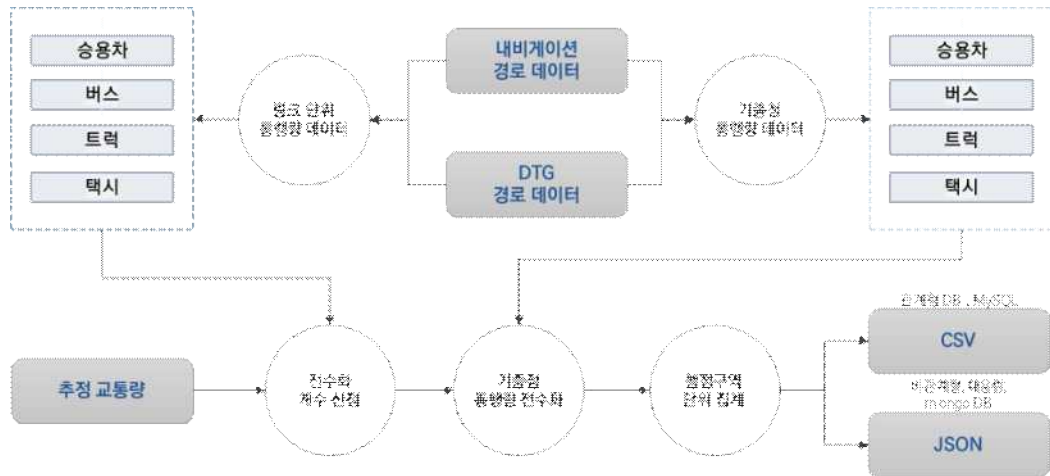
〈그림 3-32〉 기종점 통행시간 및 통행속도 DB 계산 예시

〈표 3-17〉 기종점 통행량 DB 포맷

순번	필드명	필드설명	자료형	비고
1	ST_ZONE_CD	출발 행정구역 코드	INT	
2	ED_ZONE_CD	도착 행정구역 코드	INT	
3	t1	0~1시의 통행거리	INT	km
4	t2	1~2시의 통행거리	INT	
5	t3	2~3시의 통행거리	INT	
6	t4	3~4시의 통행거리	INT	
7	t5	4~5시의 통행거리	INT	
8	t6	5~6시의 통행거리	INT	
9	t7	6~7시의 통행거리	INT	
10	t8	7~8시의 통행거리	INT	
11	t9	8~9시의 통행거리	INT	
12	t10	9~10시의 통행거리	INT	
13	t11	10~11시의 통행거리	INT	
14	t12	11~12시의 통행거리	INT	
15	t13	12~13시의 통행거리	INT	
16	t14	13~14시의 통행거리	INT	
17	t15	14~15시의 통행거리	INT	
18	t16	15~16시의 통행거리	INT	
19	t17	16~17시의 통행거리	INT	
20	t18	17~18시의 통행거리	INT	
21	t19	18~19시의 통행거리	INT	
22	t20	19~20시의 통행거리	INT	
23	t21	20~21시의 통행거리	INT	
24	t22	21~22시의 통행거리	INT	
25	t23	22~23시의 통행거리	INT	
26	t24	23~24시의 통행거리	INT	
27	tall	전일 통행량	INT	

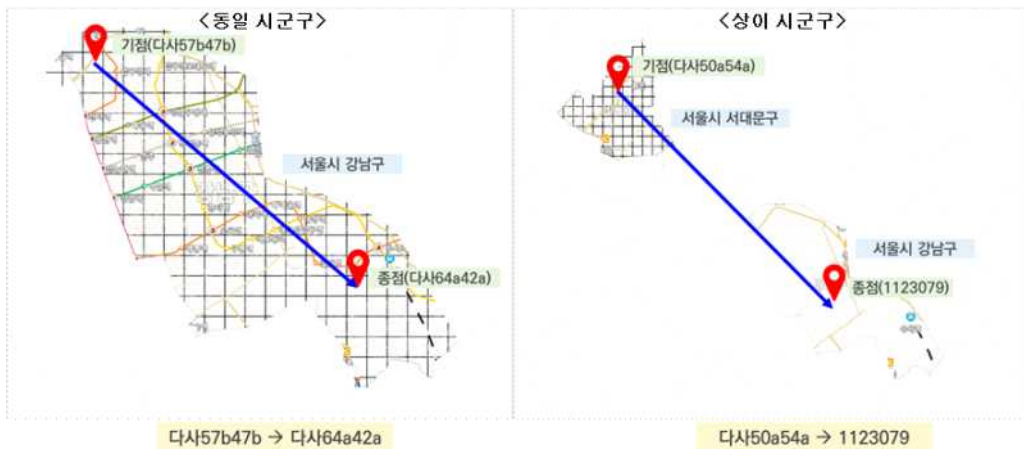
○ 일자별 개별 차량별 기종점 통행DB를 이용하여 연단위 차종별 기종점 통행량 DB를 구축

- 티맵, DTG 경로 데이터를 활용해 차종별 기종점 통행량 DB를 구축



〈그림 3-33〉 기종점 통행량 DB 구축 프로세스

- 기종점의 시군구가 동일하면 격자-격자 통행량, 상이하면 격자-읍면동 통행량을 집계



〈그림 3-34〉 기점과 종점의 시군구에 따른 집계 방식 차이

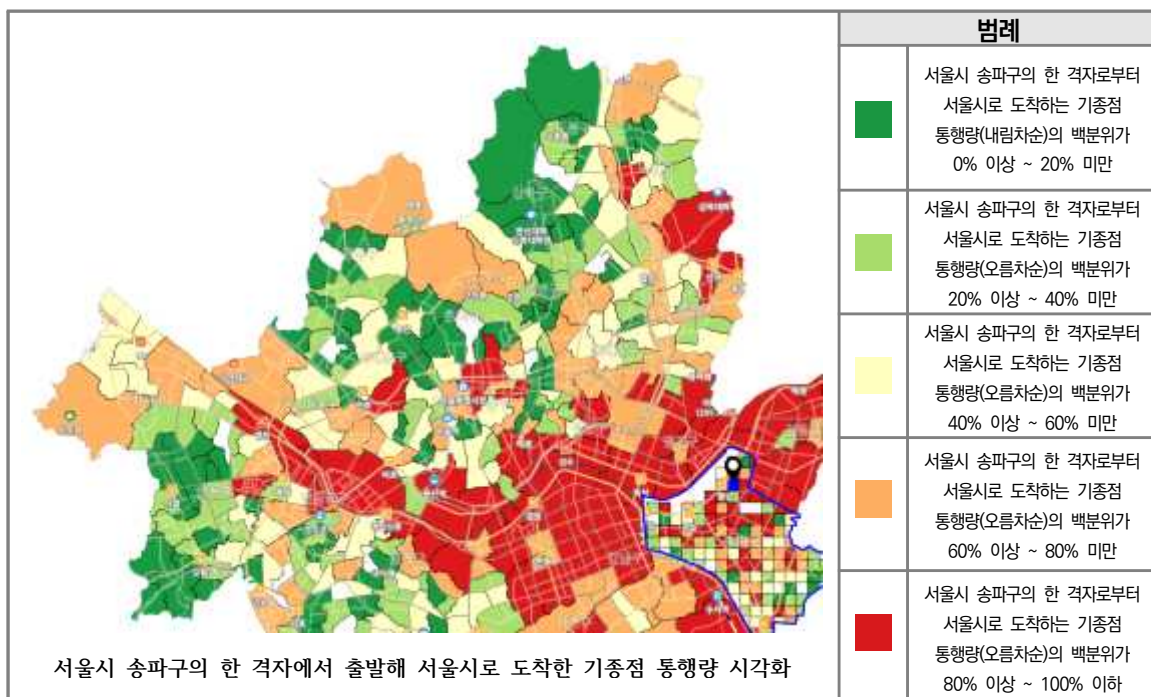
- 기종점 단위 통행량 데이터 집계
 - 티맵, DTG 경로 데이터를 통해 기점과 종점의 시군구가 동일할 경우 격자-격자 통행량을, 상이할 경우 격자-읍면동 통행량을 집계
- 기종점 통행량 DB구축 결과
 - 링크 통행량 집계 시 읍면동 코드 라인에 대한 판별 및 집계 이뤄지기에 다른 과정에 비해 상대적으로 시간이 오래 소요

〈표 3-18〉 기종점 통행량 DB 포맷

순번	필드명	필드설명	자료형	비고
1	DATA_STRD_YM	기준년월	string	
2	DPTR_LTTC_ID	출발 격자ID	string	
3	DPTR_LEGALDONG_CD	출발 법정동코드	int	
4	DPTR_SIGNGU_CD	출발 시군구코드	int	
5	DPTR_CTPR_CD	출발 시도코드	int	
6	ARVL_PYN_ID	도착 폴리곤ID	string	
7	ARVL_LEGALDONG_CD	도착 법정동코드	int	
8	ARVL_SIGNGU_CD	도착 시군구코드	int	
9	ARVL_CTPR_CD	도착 시도 코드	int	
10	WEEK_CD	주 코드	int	
11	TM_CD	시간 코드	int	
12	OD_TYPE_CD	OD 유형 코드	int	
13	DLMN_TRFC_QTY	연단위 통행량	double	

○ 기종점 통행량 결과 시각화 예시

- 동일 시군구 내에서는 비교적 골고루 통행량이 분포함을 확인
- 상이 시군구의 가까운 읍면동으로 통행량이 많은 것으로 확인

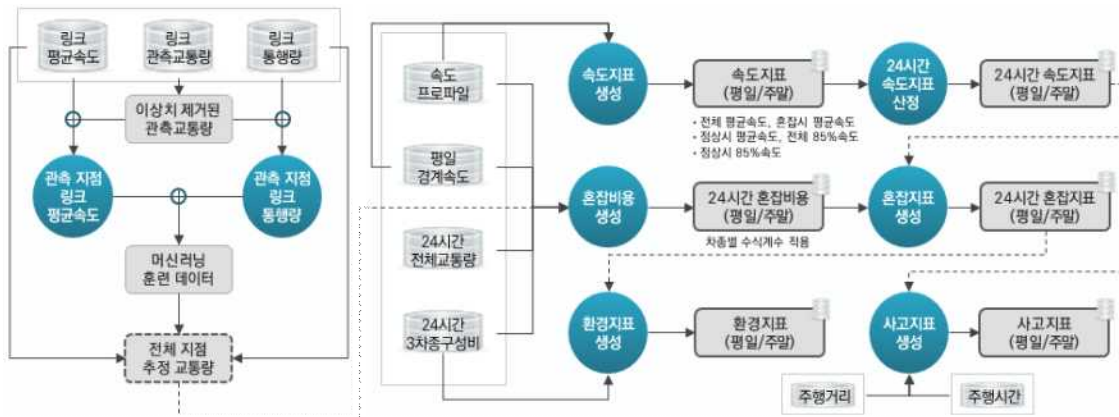


〈그림 3-35〉 기점과 종점의 시군구에 따른 집계 방식 차이

제5절 차량 교통망 성능평가지표 구축

1. 차량 교통망 성능평가지표 구축 개요

- 차량 교통망 성능평가지표 DB는 1차 가공된 속도 프로파일과 추정된 교통량을 기반으로 전국 혼잡지표, 속도지표, 환경지표, 안전지표를 구축함



〈그림 3-36〉 교통량 추정 및 교통망 성능평가지표 구축 프로세스

- 차량 교통망 성능평가지표 구축을 위한 기초교통 DB는 다음과 같이 정의된 범위가 존재하여 범위에 따라 지표를 구축함

〈표 3-19〉 교통망 성능평가지표 DB 생성 기준

데이터명	연	분기	평일 주말	시간	차종	상세 도로망	주요 도로망	행정 구역
속도 빈도분포	○	○	○	○	○	○	○	○
교통량지표	○		○	○	○	○	○	○
속도지표	○		○	○	○	○	○	○
혼잡지표	○		○	○		○	○	○
환경지표	○		○			○	○	○
안전지표	○		○		○	○	○	○

- 차량 교통망 성능평가지표는 도로망의 현황과 성능을 평가하기 위한 지표를 의미하며, 아래 표와 같이 교통량, 속도지표, 혼잡지표, 환경지표, 안전지표로 구분

〈표 3-20〉 차량 교통망 성능평가지표

지표명		지표설명
교통량	관측교통량	공공에서 조사하는 교통량으로 수집기관, 조사기간 등이 상이하여 동일한 기준으로 도로망 네트워크에 매칭하여 나타낸 연평균 일단위 조사교통량
	추정교통량	특정 시간대 동안 해당 도로구간을 통과하는 차량 대수
	차량주행거리	추정 교통량 기준의 차량주행거리
속도지표	평균속도	전체 차량의 속도를 평균한 값
	정상시 평균속도	전체 차량 중 교통 혼잡을 경험한 차량들의 평균속도
	혼잡시 평균속도	전체 차량 중 교통 혼잡을 경험하지 않은 차량들의 평균속도
혼잡지표	혼잡시간강도	전체 차량의 총 통행시간 중 교통 혼잡을 경험한 차량들의 총 통행시간 비율
	혼잡빈도강도	전체 차량의 총 통행시간 중 교통 혼잡을 경험한 차량들의 총 통행량 비율
	교통혼잡비용	차량들이 도로상에서 교통혼잡이 발생하여 정상속도 이하로 운행하게 됨으로써 추가로 발생하게 되는 비용
환경지표	이산화탄소배출량	차량들로 주행으로 발생하는 도로이동오염물질 배출량
	미세먼지배출량	
	일산화탄소배출량	
	휘발성 유기화합물배출량	
	질소산화물배출량	
안전지표	과속비율	과속기준치를 초과하는 차량의 비율
	속도 표준편차	개별 차량들의 속도의 퍼짐 정도
	위험운전행동 이벤트 비율	위험운전행동기준을 개별 차량 속도기준에 따라 위험운전행동이벤트 발생 건수를 비율로 나타냄

2. 교통환경지표 산출을 위한 입력데이터 현황화

가. 교통환경지표 산출 방법론 정의

- 교통환경지표는 도로이동오염원 배출량 중 이산화탄소배출량, 일산화탄소배출량, 질소산화물배출량, 미세먼지배출량, 휘발성유기화합물배출량으로 구성되어 있으며, 도로이동오염원 배출량 산정을 위한 방법은 기본적으로 IPCC (Intergovernmental Panel on Climate Change) Guideline⁵⁾에서 제시하고 있음
- IPCC Guideline 보고서에서는 배출량 산정을 위한 방법으로 Tier 1, Tier 2, Tier 3을 제시하고 있으며, 본 과업에서는 가장 정확한 산정을 할 수 있는 Tier 3 방법론을 기반으로 도로구간별(링크별) 차량의 배출량을 산정하였음

〈표 3-21〉 IPCC Guideline 온실가스 추정방법 비교

방법	산정식	특징
Tier 1	$Emissions = \sum(Fuel_a \times EF_a)$ <p><i>Emissions</i> : 배출량(kg) <i>Fuel_a</i> : 연료 a 소비량(TJ), <i>EF_a</i> : 배출계수(kg/TJ) <i>a</i> : 연료 종류(예, 휘발유, 경유, LPG 등)</p>	Top-down 방식 연료별 IPCC 배출계수 적용, 차량주행거리는 고려되지 않음
Tier 2	$Emissions = \sum(Fuel_{a,b,c} \times EF_{a,b,c})$ <p><i>Emissions</i> : 배출량(kg) <i>Fuel_{a,b,c}</i> : 이동 배출원 활동도에 대한 연료 소비량(TJ) <i>EF_{a,b,c}</i> : 배출계수(kg/TJ) <i>a</i> : 연료 종류(예, 휘발유, 경유, LPG 등), <i>b</i> : 차종 <i>c</i> : 배출제어기술(제어장치의 미장착, 촉매변환장치 등)</p>	Bottom-up 방식 차종별, 배출제어기술에 따른 배출계수 적용, 차종별 에너지 소비량 자료 필요, 차량주행거리는 고려되지 않음
Tier 3	$Emissions = \sum(Distance_{a,b,c,d} \times EF_{a,b,c,d}) + \sum C_{a,b,c,d}$ <p><i>Emissions</i> : 배출량(kg) <i>EF_{a,b,c,d}</i> : 배출계수(kg/km) <i>Distance_{a,b,c,d}</i> : 주어진 이동 배출원 활동도에 대해 열적으로 안정된 엔진 운전 단계 동안의 차량주행거리(VKT)(km) <i>C_{a,b,c,d}</i> : 예열 중 배출량(cold start)(kg) <i>a</i> : 연료 종류(예, 휘발유, 경유, LPG 등), <i>b</i> : 차종 <i>c</i> : 배출제어기술(제어장치의 미장착, 촉매변환장치 등) <i>d</i> : 운전 조건(예, 도심 또는 교외의 도로종류, 기타 환경적 요인)</p>	Bottom-up 방식 차종별, 배출제어기술에 따른 배출계수 적용, 배출량을 가장 정확히 산정하는 수준이기 때문에 많은 자료가 요구됨

자료 : 국가 온실가스 인벤토리 작성을 위한 2006 IPCC 가이드라인, 환경부(2008)

5) 국가 온실가스 인벤토리 작성을 위한 2006 IPCC 가이드라인, 환경부(2008)

- Tier3 방법론에서 배출제어기술, 운전 조건에 대한 자료는 차량 GPS 데이터를 통해 확인하기 어려우므로 제외하였으며, 차종별 연료별에 대한 배출계수 및 차량주행거리 자료를 이용하여 배출량을 산정함
 - 여기서 차량주행거리 자료는 교통량 기반 차량주행거리이며, KTDB 상세도로망 level6 네트워크⁶⁾의 링크별 교통량을 통해 산정된 링크별 차량주행거리 데이터를 이용함
 - 차종별 규모별 연료별 네트워크 점유율은 링크별 도로를 주행하는 차량들의 차종, 규모, 연료별로 구성하는 비율을 의미하며, 교통안전공단에서 제공하는 자동차 주행거리 자료를 바탕으로 도로상에 점유하고 있는 차량의 비율을 산정함

$$Emissions = \sum (Distance_{a,b} \times EF_{a,b} \times NOR_{a,b})$$

$$Distance_{a,b} = Volume_a \times RoadLength$$

Emissions : 배출량 (kg)

EF_{a,b} : 배출계수 (kg/km)

Distance_{a,b} : 차량주행거리 (VKT) (km)

NOR_{a,b} : 차종별 규모별 연료별 네트워크 점유율 (%)

a : 차종 (승용차, 버스, 화물차)

b : 연료 (휘발유, 경유, LPG, HEV)

- 한국교통안전공단의 자동차주행거리는 자동차검사 시 차량 등록지와 실제 운행지가 다를 수 있기 때문에 지역별 차량주행거리가 실제 차량주행거리와 차이가 있으므로 교통량 기반 차량주행거리 데이터를 이용함
- 기존 도로이동오염원 배출량 산정을 위한 배출계수는 아래의 문헌에서 제공하고 있는 배출계수를 이용하였으나, 최신 배출계수를 반영하고 있지않기 때문에 현재 제시되고 있는 배출계수와 차이가 있음
 - 수송부문 온실가스 기후변화대응 시스템 구축(II) - 자동차 온실가스 Bottom-up 배출계수 개발의 온실가스(CO₂) 배출계수와 국가 대기오염물질 배출량 산정방법 편람(III)의 배출계수를 이용하였음

6) 도로망 네트워크는 2023 도로업무편람의 2022년 111,420km 도로연장을 기준으로 약 97% 도로를 커버하는 상세도로망 level6 네트워크 데이터를 활용함

- 본 사업에서 사용된 최신 배출계수는 환경부의 국가온실가스종합센터와 국가미세먼지정보센터에서 제공하는 배출계수를 이용하였음
 - 이산화탄소배출량의 배출계수는 “국가온실가스종합정보센터”에서 제공하고 있는 자료이며, 도로수송 부문에서 가장 최신 자료인 “2021년 승인 국가 온실가스 배출·흡수계수”의 1A3b도로수송의 배출계수를 이용함
 - 일산화탄소배출량, 질소산화물배출량, 미세먼지배출량, 휘발성유기화합물배출량의 배출계수는 “국가미세먼지정보센터”에서 제공하고 있는 “국가 대기오염물질 배출량 산정방법 편람(VI)”의 도로이동오염원의 차종별 연료별 배출계수를 이용함
- 기존 활용된 배출계수는 '06년 - '08년 연식에 적용된 계수값으로 최근 배출계수를 반영이 필요하며, 차종별 규모별 연료별 네트워크 점유율도 각 연도에 맞는 값으로 반영해야하므로 교통환경지표 모듈 개선이 필요함
 - 최신화된 배출계수와 차종별 규모별 연료별 네트워크 점유율을 산정하여 교통환경지표 모듈을 개선함
- 차종별 규모별 연료별 네트워크 점유율은 도로를 주행하는 차량의 공간적 해상도를 할당하기 위해 적용되는 것으로 한국교통안전공단의 차량의 차종/규모/연료별 연간주행거리와 국토교통 통계누리의 자동차등록대수를 통해 산정함

〈표 3-22〉 차종별 유형별 규모별 전국 연간주행거리

차종		등록대수(천대)	연간주행거리(백만km)	비율	
승용차 (기타형 제외)	경형	2,025	16,586	6.4%	100%
	소형	223	1,443	0.6%	
	중형	12,873	159,942	61.4%	
	대형	6,469	82,486	31.7%	
버스	시내버스	41,193	-	39.6%	100%
	시외버스	24,201	-	60.4%	
	고속버스	18	-		
	전세버스	38,561	-		
화물차 (특수용도형 제외)	소형	2,524	32,975	66.8%	100%
	중형	329	6,075	12.3%	
	대형	220	10,345	20.9%	

자료1 : 2024년 자동차 주행거리통계, 교통안전정보관리시스템, 교통안전공단

자료2 : 2024년 12월 자동차 등록자료 통계, 국토교통 통계누리, 국토교통부

〈표 3-23〉 차종별 연료별 전국 연간주행거리

차종		등록대수(천대)	연간주행거리(백만km)	비율	
승용차	휘발유	12,353	126,131	48.4%	100%
	경유	5,363	73,890	28.4%	
	LPG	1,603	26,663	10.2%	
	기타	2,271	33,772	13.0%	
승합차	휘발유	7	92	0.7%	100%
	경유	556	9,811	71.4%	
	LPG	64	898	6.5%	
화물차	기타	51	2,942	21.4%	100%
	휘발유	30	351	0.5%	
	경유	3,255	57,271	88.7%	

자료1 : 2024년 자동차 주행거리통계, 교통안전정보관리시스템, 교통안전공단

자료2 : 2024년 12월 자동차 등록자료 통계, 국토교통 통계누리, 국토교통부

○ 이산화탄소 배출량 산출을 위한 네트워크 점유율은 다음과 같음

- '24년 이산화탄소 배출량 점유율은 대부분 크게 차이 나지 않으나, 하이브리드 차량이 전년대비 약 1%p증가되어 하이브리드 차량이 점점 증가하고 있는 것으로 나타남

〈표 3-24〉 이산화탄소 배출량 네트워크 점유율

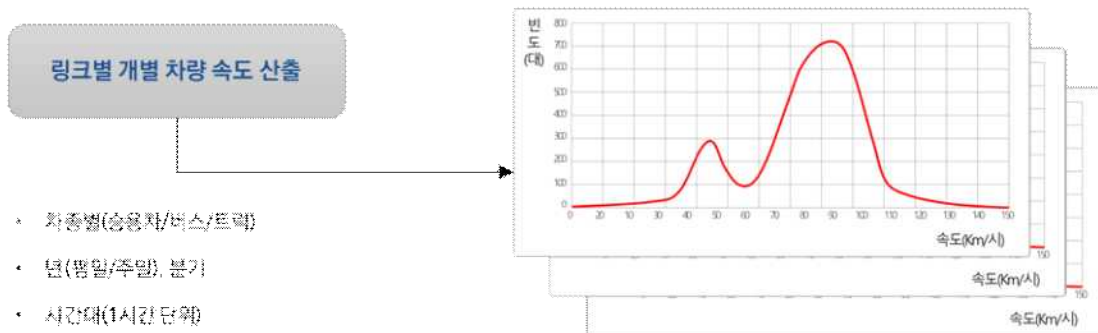
차종	연료	네트워크 점유율					
		20년	21년	22년	23년	24년	
승용차	경형	휘발유	5.7%	5.6%	5.4%	5.3%	5.3%
		LPG	1.4%	1.3%	1.2%	1.2%	1.1%
	소형	휘발유	0.5%	0.5%	0.4%	0.3%	0.3%
		경유	0.4%	0.4%	0.3%	0.2%	0.2%
		LPG	0.1%	0.1%	0.1%	0.1%	0.1%
	중형	휘발유	29.7%	29.5%	29.4%	29.7%	29.7%
		경유	22.1%	20.7%	20.2%	18.6%	17.4%
		LPG	7.4%	6.8%	6.5%	6.6%	6.3%
		HEV	2.6%	3.3%	4.7%	6.3%	8.0%
	대형	휘발유	14.4%	15.6%	15.4%	15.4%	15.3%
		경유	10.8%	10.9%	10.6%	9.7%	9.0%
		LPG	3.6%	3.6%	3.4%	3.4%	3.2%
HEV		1.3%	1.7%	2.4%	3.2%	4.1%	
버스	대형(시내버스)	CNG	39.5%	39.5%	39.9%	39.7%	39.6%
	대형(고속버스)	경유	49.9%	48.6%	48.2%	47.6%	46.5%
		CNG	10.6%	11.9%	11.9%	12.7%	13.9%
화물차	소형	경유	66.8%	67.5%	68.3%	66.9%	66.8%
	중형	경유	14.2%	13.0%	12.5%	12.6%	12.3%
	대형	경유	19.0%	19.5%	19.2%	20.5%	20.9%

○ 이산화탄소배출량의 배출계수

- 2021년 승인 국가 온실가스 배출·흡수계수의 1A3b도로수송의 배출계수에서 차종 버스를 제공하지 않고, 승합 대형을 전세광역과 시내로 구분하여 제공하므로 해당 항목을 버스의 배출계수*로 이용함

* 제시되지 않는 계수 중 차종·연료가 동일한 자동차의 규모별 항목이 없는 경우에 상위 규모의 동일 차종·연료 차량을 적용

- 배출계수의 산출식에 나타난 차속은 개별차량 경로데이터를 이용하여 구축한 링크별 차량의 속도빈도분포 데이터를 이용함



<그림 3-37> 속도빈도분포 형태

《표 3-25》 이산화탄소 배출량의 배출계수

차종	연료	배출계수		
		차속구분	산출식 (y: 배출량(g/km), x:차속(km/h))	
승용차	경형	휘발유	65.4km/h 미만	$y = 1065.1722x^{-0.5889}$
			65.4km/h 이상	$y = 0.0225x^2 - 3.3075x + 212.8460$
		LPG	65.4km/h 미만	$y = 989.9413x^{-0.5937}$
			65.4km/h 이상	$y = 0.0172x^2 - 2.3601x + 167.3842$
	소형	휘발유	65.4km/h 미만	$y = 1256.0382x^{-0.5914}$
			65.4km/h 이상	$y = 0.0252x^2 - 3.7270x + 245.9051$
		경유	65.4km/h 미만	$y = 1037.3974x^{-0.5800}$
			65.4km/h 이상	$y = 0.0133x^2 - 1.3612x + 129.4859$
		LPG	65.4km/h 미만	$y = 1223.8670x^{-0.6046}$
			65.4km/h 이상	$y = 0.0188x^2 - 2.7902x + 203.7804$
	중형	휘발유	65.4km/h 미만	$y = 1446.3728x^{-0.5793}$
			65.4km/h 이상	$y = 0.0343x^2 - 5.4212x + 339.8479$
		경유	65.4km/h 미만	$y = 1153.5685x^{-0.5507}$
			65.4km/h 이상	$y = 0.0226x^2 - 3.0857x + 225.8804$
		LPG	65.4km/h 미만	$y = 1513.8104x^{-0.6075}$
			65.4km/h 이상	$y = 0.0245x^2 - 3.6654x + 257.7428$
		HEV	65.4km/h 미만	$y = 211.9807x^{-0.1884}$
			65.4km/h 이상	$y = 0.0205x^2 - 2.8635x + 190.4598$
	대형	휘발유	65.4km/h 미만	$y = 2022.6604x^{-0.6183}$
			65.4km/h 이상	$y = 0.0374x^2 - 5.9783x + 385.8791$
		경유	65.4km/h 미만	$y = 1149.2206x^{-0.5313}$
			65.4km/h 이상	$y = 0.0246x^2 - 3.3168x + 239.5643$
		LPG	65.4km/h 미만	$y = 1967.2719x^{-0.6616}$
			65.4km/h 이상	$y = 0.0295x^2 - 4.6079x + 301.8248$
HEV		65.4km/h 미만	$y = 522.2199x^{-0.3855}$	
		65.4km/h 이상	$y = 0.0205x^2 - 2.8635x + 190.4598$	
버스	시내	CNG	$y = 5054.5880x^{-0.4910}$	
	전세 광역	경유	65.4km/h 미만	$y = 4317.2386x^{-0.5049}$
			65.4km/h 이상	$y = 0.1829x^2 - 29.8145x + 1670.8962$
		CNG	65.4km/h 미만	$y = 5011.6276x^{-0.5579}$
65.4km/h 이상			$y = 0.1122x^2 - 17.5798x + 1141.5327$	
화물차	소형	경유	65.4km/h 미만	$y = 1250.4831x^{-0.4630}$
			65.4km/h 이상	$y = 0.0292x^2 - 2.9530x + 258.3205$
	중형	경유	65.4km/h 미만	$y = 1385.8860x^{-0.4184}$
			65.4km/h 이상	$y = 1.6720x + 141.2224$
	대형	경유	$y = 3351.2892x^{-0.4407}$	

3. 주요 교통망 성능평가지표 분석 결과

가. 교통량 지표

- 교통 분야의 핵심적인 기초 자료로 활용되는 교통량을 도로망 네트워크 기반으로 구축된 지표들을 의미하며, 도로를 주행하는 모든 차량들의 교통량과 교통량기반 주행거리를 의미함
 - 링크별, 평일/주말별, 시간대별/일별 산정
 - 관측교통량 : 현장조사 및 검지기를 통해 조사한 교통량으로 한국건설기술연구원, 한국도로공사, 8대 특광역시에서 제공한 자료를 도로망 네트워크에 매칭하여 구축
 - 추정교통량: 혼잡을 경험하지 않은 차량들의 평균속도
 - 차량주행거리: 추정교통량을 이용하여 도로를 주행하는 모든 차량들의 이동거리의 합
- 평일 기준의 시도별 차량주행거리는 다음과 같음
 - 차량주행거리는 전년대비 약 5.6% 증가하였으며, 충북에서 7.6%로 가장 크게 증가, 전북에서 0.8%로 가장 적게 증가한 것으로 분석됨

〈표 3-26〉 시도별 차량주행거리 비교(평일 기준)

단위 : km*대/일

구분	2023년	2024년	증감률
서울특별시	100,830,532	108,397,403	7.5%
부산광역시	47,367,297	50,394,522	6.4%
대구광역시	40,718,021	43,531,788	6.9%
인천광역시	50,905,698	53,738,051	5.6%
광주광역시	27,345,615	28,157,731	3.0%
대전광역시	25,548,690	25,874,733	1.3%
울산광역시	26,488,086	27,665,855	4.4%
세종특별자치시	8,081,393	8,652,021	7.1%
경기도	327,798,467	350,797,670	7.0%
강원특별자치도	47,432,367	48,526,686	2.3%
충청북도	55,661,865	59,908,464	7.6%
충청남도	76,333,856	79,984,263	4.8%

구분	2023년	2024년	증감률
전북특별자치도	51,672,187	52,077,799	0.8%
전라남도	58,648,111	61,040,340	4.1%
경상북도	90,463,169	93,533,040	3.4%
경상남도	90,106,255	96,448,208	7.0%
제주특별자치도	18,080,419	19,228,298	6.3%
합계	1,143,482,026	1,207,956,874	5.6%

* 데이터는 절대적으로 참값이 아니며, 지속적으로 신뢰도를 높이기 위한 정산작업 수행

나. 속도지표

○ 도로를 주행하는 모든 차량들의 속도를 평균하거나 도로등급별 혼잡경계속도 기준에 따라 차량들의 속도를 평균한 것을 의미함

- 링크별, 분기별, 평일/주말별, 시간대별/일별 산정

- 평균속도: 모든 차량들의 평균속도
- 정상시 평균속도: 혼잡을 경험하지 않은 차량들의 평균속도
- 혼잡시 평균속도: 혼잡을 경험한 차량들의 평균속도

〈표 2-27〉 상세도로망 단위 속도지표 DB 포맷

순번	필드명	필드설명	자료형	비고
1	v_link_id	가상링크ID	long	
2	qrtr_cd	분기코드	integer	
3	week_cd	평일/주말코드	integer	
4	time_cd	시간코드	integer	
5	avg_spd	평균속도	double	
6	noml_avg_spd	정상시 평균속도	double	
7	cgst_avg_spd	혼잡시 평균속도	double	
8	total_85_avg_spd	전체 85% 평균속도	double	
9	noml_85_avg_spd	상위 85% 평균속도	double	

○ 평일 기준의 시도별 평균속도는 다음과 같음

〈표 3-28〉 시도별 혼잡시평균속도 비교(평일 기준)

단위 : km/h

구분	2023년	2024년	증감률
서울특별시	16.56	15.58	-5.95%
부산광역시	18.91	17.82	-5.76%
대구광역시	20.54	19.24	-6.31%
인천광역시	17.61	16.40	-6.88%
광주광역시	19.14	17.80	-6.99%
대전광역시	19.66	18.51	-5.82%
울산광역시	20.89	19.58	-6.29%
세종특별자치시	22.78	21.21	-6.88%
경기도	22.22	21.02	-5.42%
강원특별자치도	26.86	25.20	-6.18%
충청북도	26.35	24.93	-5.38%
충청남도	27.08	25.68	-5.17%
전북특별자치도	26.54	25.01	-5.75%
전라남도	27.18	25.72	-5.36%
경상북도	27.65	26.10	-5.58%
경상남도	25.58	24.13	-5.67%
제주특별자치도	23.88	22.64	-5.19%
합계	24.07	22.69	-5.75%

* 데이터는 절대적으로 참값이 아니며, 지속적으로 신뢰도를 높이기 위한 정산작업 수행

〈표 3-29〉 시도별 정상시평균속도 비교(평일 기준)

단위 : km/h

구분	2023년	2024년	증감률
서울특별시	39.76	44.52	11.97%
부산광역시	41.46	45.93	10.77%
대구광역시	43.83	48.39	10.39%
인천광역시	41.58	46.00	10.64%
광주광역시	42.00	47.65	13.43%
대전광역시	43.23	47.75	10.46%
울산광역시	43.76	48.01	9.70%
세종특별자치시	44.47	48.16	8.29%
경기도	46.30	50.42	8.91%
강원특별자치도	49.37	52.51	6.35%
충청북도	50.01	53.64	7.25%
충청남도	50.44	53.86	6.78%
전북특별자치도	49.97	53.60	7.27%
전라남도	50.83	54.08	6.41%
경상북도	51.10	54.36	6.38%
경상남도	48.47	52.56	8.45%
제주특별자치도	46.35	48.87	5.45%
합계	47.44	51.26	8.05%

* 데이터는 절대적으로 참값이 아니며, 지속적으로 신뢰도를 높이기 위한 정산작업 수행

다. 환경지표

- 차량들로 인해 발생하는 도로이동오염원의 배출량을 의미하며, 도로 특성에 따른 속도를 적용한 차종별 도로이동오염원의 배출량을 구축함으로써 전국 상세 도로망 기반의 환경지표를 구축함

- 링크별, 분기별, 평일/주말별, 시간별 산정

〈표 2-30〉 상세도로망 단위 환경지표 DB 포맷

순번	필드명	필드설명	자료형	비고
1	v_link_id	가상링크ID	long	
2	week_cd	평일/주말코드	integer	
3	co_emissions	일산화탄소 배출량	double	
4	co2_emissions	이산화탄소 배출량	double	
5	nox_emissions	질소산화물 배출량	double	
6	pm_emissions	미세먼지 배출량	double	
7	vocs_emissions	휘발성유기화합물 배출량	double	

- 평일 기준 시도별 이산화탄소 배출량은 다음과 같음

〈표 3-31〉 시도별 이산화탄소배출량 비교(평일 기준)

단위 : g/일

구분	2023년	2024년	증감률
서울특별시	23,614,848,433	23,226,111,034	-1.65%
부산광역시	10,564,014,157	10,379,856,033	-1.74%
대구광역시	9,338,968,036	9,055,264,761	-3.04%
인천광역시	11,544,929,280	11,079,311,657	-4.03%
광주광역시	6,219,970,568	5,892,415,911	-5.27%
대전광역시	5,914,438,398	5,410,770,856	-8.52%
울산광역시	6,005,137,224	5,803,387,631	-3.36%
세종특별자치시	1,721,694,374	1,706,532,054	-0.88%
경기도	69,623,866,714	68,454,977,098	-1.68%
강원특별자치도	9,515,653,778	9,060,992,885	-4.78%
충청북도	11,709,054,585	11,515,457,043	-1.65%
충청남도	15,800,660,876	15,261,490,919	-3.41%
전북특별자치도	11,049,266,928	10,339,283,329	-6.43%
전라남도	12,337,000,758	12,084,667,369	-2.05%
경상북도	19,079,281,256	18,190,599,405	-4.66%
경상남도	19,120,245,818	19,035,835,703	-0.44%
제주특별자치도	4,077,653,016	4,049,843,948	-0.68%
합계	247,236,684,199	240,546,797,637	-2.71%

* 데이터는 절대적으로 참값이 아니며, 지속적으로 신뢰도를 높이기 위한 정산작업 수행

라. 혼잡지표

- 도로를 주행하는 모든 차량들이 혼잡을 경험했을 때의 지표를 의미함
 - 혼잡시간강도: 도로를 주행하는 모든 차량이 경험한 총 통행시간 대비 혼잡을 경험한 차량의 총 통행시간의 비율
 - 혼잡빈도강도: 도로를 주행하는 모든 차량 중 혼잡을 경험한 차량의 비율

〈표 2-32〉 상세도로망 단위 혼잡강도 DB 포맷

순번	필드명	필드설명	자료형	비고
1	v_link_id	가상링크ID	long	
2	week_cd	평일/주말코드	integer	
3	cgst_time_intnsty	혼잡시간강도	double	
4	cgst_frqcy_intnsty	혼잡빈도강도	double	

- 평일 기준 시도별 혼잡시간강도는 다음과 같음

〈표 3-33〉 시도별 혼잡시간강도 비교(평일 기준)

단위 : %

구분	2023년	2024년	증감률
서울특별시	77.40	80.88	4.49%
부산광역시	67.60	74.24	9.81%
대구광역시	66.43	72.41	9.00%
인천광역시	65.85	72.14	9.55%
광주광역시	71.52	76.38	6.80%
대전광역시	66.71	72.08	8.05%
울산광역시	62.06	68.86	10.95%
세종특별자치시	60.04	67.83	12.97%
경기도	68.28	73.90	8.23%
강원특별자치도	58.48	65.56	12.11%
충청북도	59.88	66.80	11.54%
충청남도	59.54	66.41	11.54%
전북특별자치도	60.06	66.63	10.94%
전라남도	57.29	63.82	11.40%
경상북도	58.02	64.89	11.85%
경상남도	62.82	69.72	10.98%
제주특별자치도	61.79	67.18	8.73%
합계	63.29	69.63	10.01%

* 데이터는 절대적으로 참값이 아니며, 지속적으로 신뢰도를 높이기 위한 정산작업 수행

○ 평일 기준 시도별 혼잡빈도강도는 다음과 같음

〈표 3-34〉 시도별 혼잡빈도강도 비교(평일 기준)

단위 : %

구분	2023년	2024년	증감률
서울특별시	63.65	67.26	5.67%
부산광역시	54.07	59.93	10.83%
대구광역시	53.53	58.53	9.34%
인천광역시	51.81	57.46	10.90%
광주광역시	59.44	63.29	6.48%
대전광역시	53.88	58.21	8.02%
울산광역시	49.81	55.29	11.01%
세종특별자치시	49.16	55.38	12.66%
경기도	56.02	60.99	8.87%
강원특별자치도	48.44	53.55	10.57%
충청북도	49.18	54.30	10.39%
충청남도	49.13	54.30	10.52%
전북특별자치도	49.58	54.23	9.37%
전라남도	47.02	51.53	9.59%
경상북도	47.86	52.77	10.25%
경상남도	52.16	57.35	9.96%
제주특별자치도	51.83	55.67	7.41%
합계	51.97	56.92	9.53%

* 데이터는 절대적으로 참값이 아니며, 지속적으로 신뢰도를 높이기 위한 정산작업 수행

4. 차량 교통망 성능평가지표 활용

- 차량 교통망 성능평가지표를 View-T 웹서비스를 통해 '통행지표'와 '데이터 다운로드' 등의 기능에 업로드하여 데이터를 제공하고 있음
- 교통량지표 중 차량주행거리는 통계청의 실험적 통계에서 차량 GPS 기반 도로 유형별 주행거리 통계로 활용되고 있음



〈그림 3-38〉 통계청의 실험적 통계에 활용되는 차량 교통망 성능평가지표

- 교통혼잡지표 중 교통혼잡비용은 통계청의 e-나라지표에서 도로교통 혼잡비용 통계로 활용되고 있음
- 교통환경지표 중 이산화탄소배출량은 국토교통부 탄소공간지도에서 수송부문 배출량 데이터로 활용되고 있음

제6절 프로토타입 회전교통량 추정 DB 구축

1. 회전교통량 추정 DB 구축 개요

가. 회전교통량 추정DB 구축배경

- 현재 교통량 추정 알고리즘은 회전을 고려하지 않고, 링크별 통행량과 경로 연결성을 고려하여 교통량 추정DB를 구축함
- 회전교통량을 추정 가능성을 검토하기 위해 프로토타입으로 특정 지자체를 설정하여 해당 지자체의 미관측된 교차로 구간에 대한 회전교통량을 추정함

나. 회전교통량 추정DB 구축목적

- 차량 교통망 성능평가지표는 전국 연평균 평일 단위의 시간적 범위로 DB구축하고 있으나, 스마트 교차로 데이터 수집이 가능한 지역이면서 교차로 구간에 대한 현장조사 데이터 수집가능 일자를 고려하여 세종시를 대상으로 1일 회전교통량 추정을 수행하였고, 회전교통량 DB를 구축함
- 프로토타입의 회전교통량 추정DB 구축을 통해 지자체의 스마트 교차로 데이터를 이용한 전국 회전교통량 DB구축을 검토함

2. 회전교통량 추정 DB를 위한 교통 네트워크 구축

가. 회전교통량 DB 구축을 위한 교통 네트워크 활용

1) 교통 네트워크 개요

- 회전교통량 추정을 위해서는 교차로 단위의 회전방향 정보를 확인할 수 있는 네트워크가 필요함

- 이에 상세도로망(Level6.0)을 기반으로 회전정보를 포함한 교차로 중심의 상위 네트워크를 구축하고, 이를 교통 네트워크라 명명함
 - 교차로/IC/JC 지점에 노드를 생성하고, 생성된 노드와 노드를 연결하여 링크를 구축
 - 이렇게 구축된 노드와 링크를 각각 대표노드, 대표링크로 정의함



<그림 3-39> 교통 네트워크 개념

2) 교통 네트워크

① 대표노드 및 대표링크

- 대표노드는 실제 도로와 도로가 교차하는 지점에 생성하고, 생성한 대표노드와 상세도로망(Level6.0) 노드와의 관계 생성을 위해 대표노드 속성정보에 상세도로망의 노드정보를 입력함
- 도로와 도로가 만나는 교차지점, IC, JC에 각각 대표노드를 생성하고, 통행이 가능한 구간에 대해 대표노드와 대표노드를 이어 대표링크를 생성함
 - 대표노드는 교차로의 중심점을 기준으로 설정하여 모든 접근로가 수렴하는 기준점 역할 하고, 대표링크는 각 접근로와 연결하는 가상의 링크를 구성하여 회전방향을 명시적으로 표현함



〈그림 3-40〉 대표노드 및 대표링크 구축 예시

○ 대표노드 및 대표링크의 테이블 정의는 다음과 같음

- 대표노드는 대표노드 ID, 상세도로망 노드 ID, 위치 정보(x, y 좌표)와 연결된 상세도로망의 노드 리스트를 포함하여 노드 간 공간적 관계를 표현함

〈표 3-35〉 교통 네트워크 대표노드 테이블 정의

No	Column	Type	설명	비고
1	tn_node_id	Integer	교통 네트워크 대표노드 ID	
2	node_id_l6	Integer	상세도로망(Level6.0) 노드 ID	
3	x	Real	x 좌표	
4	y	Real	y 좌표	
5	node_type	Integer	노드유형	
6	node_name	Text	노드 명칭	
7	node_list	Text	연결된 상세도로망(Level6.0) 노드 정보	

- 대표링크는 대표링크 ID, 상·하행 기준 대표노드 ID, 차로 수, 행정구역, 링크 길이 등 주요 속성을 정의함
- 또한 상세도로망과 동일하게 대표링크를 양방향 하나의 링크 ID로 관리하되, 상·하행을 논리적으로 구분하기 위해 가상링크 ID를 부여하여 방향성 정보 관리하도록 함

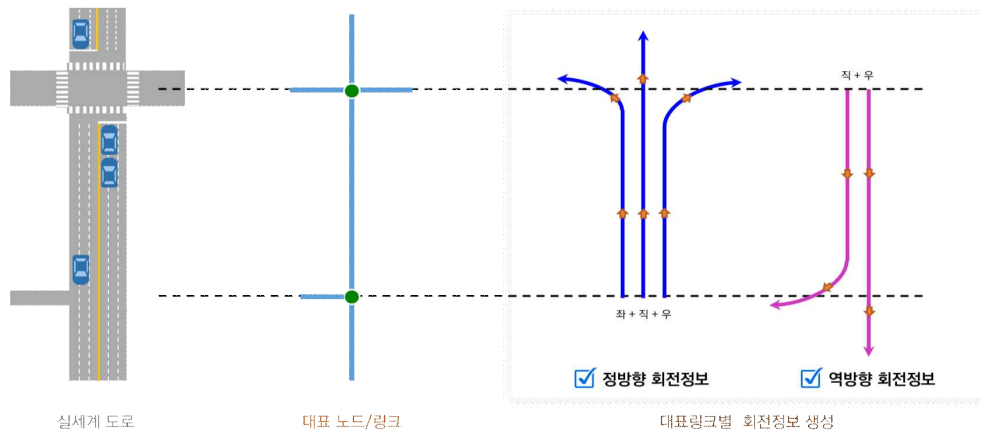
〈표 3-36〉 교통 네트워크 대표링크 테이블 정의

No	Column	Type	설명	비고
1	tn_link_id	Integer	교통 네트워크 대표링크 ID	

No	Column	Type	설명	비고
2	tn_up_f_no	Integer	상행 시작 대표노드 ID	
3	tn_up_t_no	Integer	상행 종료 대표노드 ID	
4	tn_dw_f_no	Integer	하행 시작 대표노드 ID	
5	tn_dw_t_no	Integer	하행 종료 대표노드 ID	
6	road_rank	Text	도로등급	
7	road_name	Text	도로명칭	
8	oneway	Integer	일반통행	
9	road_no	Text	도로번호	
10	link_type	Text	링크종별	
11	up_lanes	Integer	상행 차로 수	
12	dw_lanes	Integer	하행 차로 수	
13	lane	Text	전체 차로 수	
14	tn_up_v_li	Integer	상행 대표 가상링크 ID	
15	tn_dw_v_li	Integer	하행 대표 가상링크 ID	
16	sido_id	Text	시도 행정구역 ID	
17	sigungu_id	Text	시군구 행정구역 ID	
18	emd_id	Text	읍면동 행정구역 ID	
19	tn_length	Real	대표링크 길이	

② 회전정보 구축

- 동일한 대표링크를 기준으로 차량 통행 방향별 회전정보를 생성함
 - 양방향 통행이 가능한 도로의 경우 정방향과 역방향을 구분하여 모든 통행 가능 방향을 고려하여 회전정보를 생성함
 - 대표링크는 양방향 기준 하나의 ID를 통해 관리하되, 상하행을 명확히 구분하기 위해 가상 링크ID를 부여하여 방향성 정보를 체계적으로 관리함



〈그림 3-41〉 대표링크별 회전정보 구축 방법

- 회전정보는 대표링크와 상세도로망(Level6.0) 링크 간 관계 테이블로, 대표링크 기준으로 회전 방향별 좌회전, 직진, 우회전, U턴 4가지로 회전을 구분함

대표링크별 회전정보 구축



〈그림 3-42〉 대표링크별 회전정보 구축 예시

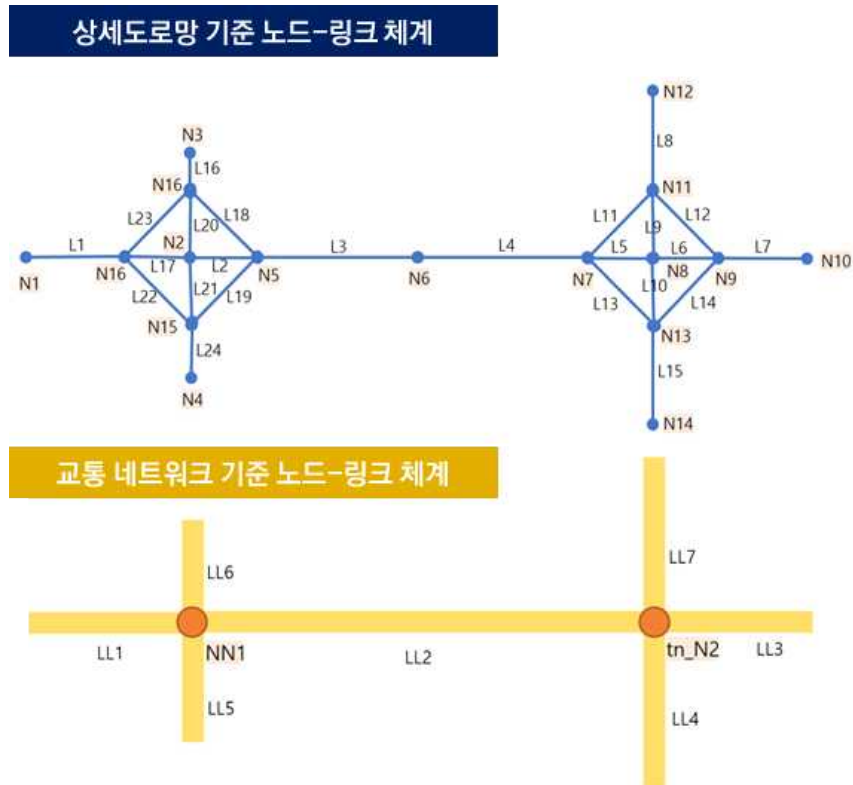
- 대표링크 기준의 회전정보는 대표 가상링크 ID, 시작 및 종료 대표노드 ID, 회전정보 코드, 시작 및 종료 상세도로망(Level6.0) 노드 ID, 상세도로망 기준의 링크 경로, 회전 후의 다음 대표 가상링크 ID를 포함함

〈표 3-37〉 대표링크 기준 회전정보 테이블 정의

No	Column	Type	설명	비고
1	tn_f_node	Integer	시작 대표노드 ID	
2	tn_t_node	Integer	종료 대표노드 ID	

No	Column	Type	설명	비고
3	tn_turn_code	Text	회전 정보(코드)	L01: 좌회전, S01: 직진 R01: 우회전, U01: 유턴
4	f_node6	Integer	시작 상세 노드 ID	
5	t_node6	Integer	종료 상세 노드 ID	
6	path	Text	상세도로망 링크 경로	
7	nxt_vlink	Text	다음 대표 가상링크 ID	진출 대표링크
8	tn_v_link_id	Integer	대표 가상링크 ID	진입 대표링크

- 연속된 2개의 교차로를 대상으로 상세도로망 기준의 노드-링크 체계와 교통 네트워크 기준의 대표노드-대표링크 체계는 다음과 같이 도식화할 수 있음



<그림 3-43> 네트워크별 노드-링크 체계(도식화)

- 상기 기준을 토대로 교차로 회전교통량 추정을 위한 분석 기반을 마련하기 위해 기존 상세도로망 네트워크를 고도화한 세종시 범위의 교통 네트워크를 구축함



〈그림 3-44〉 세종특별자치시 교통 네트워크

다. 회전교통량 추정을 위한 학습데이터 구축

- 교차로 회전교통량 추정모형을 구축하기 위해 관측 회전교통량, 차량 GPS 데이터, 교차로 기하구조 정보를 주요 학습데이터로 활용하고, 이를 통해 교차로별 통행 패턴의 물리적 제약요인을 반영할 수 있도록 함
 - 관측 회전교통량
 - 세종시 스마트교차로에서 수집된 데이터
 - 추가 현장조사를 통해 교차로별 회전방향(좌회전·직진·우회전)별 교통량
 - 차량 GPS 데이터: 내비게이션 기반의 주행 궤적을 활용하여, 링크 단위의 probe 통행량, 평균속도, 교통밀도, 네트워크 연결성 변수 산출
 - 네트워크 연결성은 차량 GPS의 연속 통행 패턴을 분석하여 링크 간 관계를 수치화 함
 - 교차로 기하구조 데이터: 각 교차로의 차로 수, 도로 연장, 도로등급 등 물리적 구조 정보를 기반으로 구축
- 시간적 집계 단위는 5분, 15분, 60분으로 구분하여 학습데이터를 구축

- 시간 세분화 수준에 따른 추정 결과의 신뢰도와 모형의 민감도를 비교·분석하기 위함
- 학습 데이터 구성 시, 교통량 수준이 높은 스마트교차로와 상대적으로 낮은 일반 교차로의 데이터를 함께 포함하여, 도시 전역을 대상으로 한 모형의 강건성과 일반화 성능을 확보함
- 다만, 심야시간대는 교통량이 적어 신뢰도 확보가 어려워 제외하였으며, U턴은 발생 빈도가 낮고 데이터 불균형이 심해 분석대상에서 제외함

〈표 3-38〉 회전교통량 추정을 위한 주요 학습데이터

분류	활용 데이터	집계 단위
관측 회전교통량	스마트교차로 CCTV 수집 회전교통량	5분, 15분, 60분
	현장 조사 수집 회전교통량	
차량 GPS 데이터	내비게이션 GPS 궤적 데이터 기반 회전교통량(probe)	
	링크 단위 평균속도, 밀도	
	네트워크 연결성	
교차로 기하구조	차로 수, 도로 연장, 도로등급 등	-

1) 관측 회전교통량 기반 데이터 구축

- 세종특별자치시에서 현재 설치·운영 중인 31개 지점 개별 스마트교차로에 대한 2025년 7월 24일의 CCTV 기반 수집된 회전교통량 데이터를 수집함
 - 차량 GPS 데이터의 수집 일자는 관측 회전교통량 수집 기간과 동일한 일자 및 시간 범위로 정합성을 맞추어 학습데이터로 구축함
- 또한, 스마트교차로 외의 일반 교차로를 포함한 세종시 전역의 회전교통량을 확보하기 위해 2025년 7월 24일 현장조사를 실시함
 - 조사 대상은 총 63개 교차로(스마트교차로 31개소, 일반교차로 32개소)이며, 오전 첨두(07시~09시), 오후 비첨두(12시~14시), 오후 첨두(17시~19시)의 6시간 동안 회전교통량을 직접 계측하여 추가 학습데이터로 활용함
- 이렇게 구축된 세종시의 관측 회전교통량 데이터는 교통 네트워크 가상링크 ID, 관측 지점 정보, 회전 방향, 시간 정보 및 교통량 항목으로 구성되며, 각 교차로별 회전패턴을 정량적으로 표현할 수 있도록 정의함

〈표 3-39〉 세종특별자치시 관측 회전교통량(연 평균) 테이블 정의

No	Column	Type	설명	비고
1	tn_v_link_id	Integer	교통 네트워크 가상링크ID	
2	road_rank	Text	도로등급	
3	sido_id	Text	시도ID	
4	traffic_type	Text	조사기관	
5	spot_id	Text	관측지점 ID	
6	spot_name	Text	관측지점 명칭	
7	camera_dir	Text	카메라 방향정보	
8	dir	Integer	방향 정보	
9	move	Text	회전 정보	L01 : 좌회전 S01 : 직진 R01 : 우회전 U01 : 유티
10	week_code	Integer	주말/평일정보	0: 평일 1 :주말
11	day_code	Integer	요일 정보	
12	hour_code	Integer	시간단위 코드	
13	min_code	Integer	분단위 코드	
14	time_info	Integer	수집시간 정보	
15	tn_v_link_N	Integer	회전을 통해 도착한 링크 ID	
16	lanes	Integer	세종시에서 제공하는 차선수	
17	auto	Real	회전한 승용차 교통량	
18	bus	Real	회전한 버스 교통량	
19	truck	Real	회전한 화물차 교통량	

〈표 3-40〉 세종특별자치시 관측 회전교통량(일 단위) 테이블 정의

No	Column	Type	설명	비고
1	tn_v_link_id	Integer	교통 네트워크 가상링크ID	
2	road_rank	Text	도로등급	
3	sido_id	Text	시도ID	
4	traffic_type	Text	조사기관	
5	spot_id	Text	관측지점 ID	
6	spot_name	Text	관측지점 명칭	
7	camera_dir	Text	카메라 방향정보	
8	dir	Integer	방향 정보	
9	move	Text	회전 정보	L01 : 좌회전 S01 : 직진 R01 : 우회전 U01 : 유턴
10	date	Text	수집 일자 정보	yyyy-MM-dd
11	week_code	Integer	주말/평일정보	0: 평일 1 :주말
12	day_code	Integer	요일 정보	
13	hour_code	Integer	시간단위 코드	
14	min_code	Integer	분단위 코드	
15	time_info	Integer	수집시간 정보	
16	tn_v_link_N	Integer	회전을 통해 도착한 링크 ID	
17	lanes	Integer	세종시에서 제공하는 차선수	
18	auto	Real	회전한 승용차 교통량	
19	bus	Real	회전한 버스 교통량	
20	truck	Real	회전한 화물차 교통량	

2) 내비게이션 기반 차량 GPS 데이터 기반 데이터 구축

- 회전교통량 추정을 위한 학습데이터를 구축하기 위해서는 내비게이션 기반 차량 GPS 데이터를 교통 네트워크 단위로 정제·구축하는 과정이 필수적임

① 교통 네트워크 단위 링크 경로 데이터 구축

○ 상세도로망(Level6.0)을 기반으로 한 링크 경로 데이터를 구축하고, 대표링크 기준의 회전정보 테이블을 활용하여 교통 네트워크 단위의 링크 경로 데이터를 구축함

- 2025년 7월 24일 하루 동안 세종특별자치시를 경유한 차량 GPS 데이터 추출하여, 교통 네트워크 단위의 링크 경로 데이터 구축

○ 교통 네트워크 단위의 링크 경로 데이터는 다음과 같이 정의함

〈표 3-41〉 교통 네트워크 단위 링크 경로 데이터 정의

No	Column	Type	설명	비고
1	OBUID	Text	차량용단말기식별부호	1일 단위 가상키
2	GRPID	Integer	통행그룹ID	통행분리 키
3	CVT_CD	Text	차종구분 코드	CT_COMPACT CT_NORMAL CT_MIDDLE CT_LARGE CT_TRUCK CT_SPECIAL CT_SMALL_TRUCK
4	CFT_CD	Text	유종구분 코드	FT_GAS FT_DIESEL FT_LPG FT_GASPM FT_EV
5	DST_NM	Text	목적지명	
6	DST_CTGRY	Text	목적지 카테고리	중분류 정보
7	TN_V_LINK_ID	Integer	교통네트워크 가상링크 ID	
8	TURN_ID	Text	회전방향 ID	S01:직진 R01:우회전 L01:좌회전 U01:유턴 Z00:연결링크없음 -1,-9:회전정보없음

No	Column	Type	설명	비고
9	FLAG	Integer	회전방향 사용 플래그	0: 정상 -1:회전방향미사용
10	LINK_ENT_TM	Integer	링크 진입 시간	Unixtime
11	LINK_EXT_TM	Integer	링크 진출 시간	Unixtime
12	GIS_LINK_LNTH	Real	GIS 네트워크의 링크 길이	m
13	GIS_LINK_KPH	Real	GIS 네트워크의 링크 길이로 계산된 속도	km/h
14	GPS_LINK_LNT H	Real	동일 링크에 매칭된 포인트의 거리	m
15	GPS_LINK_KPH	Real	포인트의 거리로 계산된 속도	km/h

② 프로브 통행량 및 속도 데이터 구축

- 앞서 구축된 교통 네트워크 단위 링크 경로 데이터를 통해 개별차량의 경로 데이터를 회전별 링크 단위의 일별 5분, 15분, 60분 단위로 집계하여 각각의 프로브 통행량과 평균속도를 구축함
- 구축한 프로브 통행량과 평균속도를 바탕으로 대표 링크별 밀도(k) 값을 산출하여 회전교통량 추정을 위한 학습데이터로 활용함

〈표 3-42〉 프로브 및 속도 데이터 정의

No	Column	Type	설명	비고
1	tn_v_link_id	Integer	교통네트워크 가상링크 ID	교통 네트워크
2	turn_id	Text	회전방향 ID	S01: 직진 R01: 우회전 L01: 좌회전 U01: 유턴
3	date	Text	일자	yyyyMMdd
4	time_code	Integer	시간코드	5분 단위: 0~287 15분 단위: 0~95 60분 단위: 0~23
5	probe	Integer	프로브 통행량	대
6	spd	Real	평균속도	km/h

③ 네트워크 연결성 변수 구축

- 네트워크 연결성 변수는 추정 대상 링크와 공간적으로 인접한 주변 관측지점의 교통량 특성을 통계적으로 연계하는 개념으로, 다음과 같은 절차로 구축함
 - 내비게이션 기반 차량 GPS 데이터를 이용해 미관측 링크를 통과하는 모든 차량의 이동 경로 추출
 - 이후 개별 차량의 주행 경로를 분석하여, 추정 대상 링크와의 물리적 거리, 통행량 수준, 차로 수 등의 요소를 종합적으로 고려
 - 이를 기반으로 연결성이 높은 최대 10개 지점을 후보군으로 설정하고, 이 중 공간적·통행 특성이 가장 유사한 상위 3개 지점을 최종적으로 선정하여 네트워크 연결성 변수로 구축함
- 연결성 변수는 추정 대상 링크와 동일한 도로등급 그룹 내에서 최단 거리상에 존재하는 관측지점의 정보만을 활용함으로써 공간적 연계성을 극대화함
- 모든 연결성 데이터는 일평균 60분 단위(ADT) 기준으로 작성되었으며, 동일한 변수를 5분 및 15분 단위 학습데이터에도 동일하게 적용함



〈그림 3-45〉 네트워크 연결성 변수 개념

3) 교차로 기하구조 기반 데이터 구축

- 기하구조 변수는 교차로 및 링크의 물리적 특성과 용량 제약을 정량적으로 반

영하여, 구조적 차이에 따른 교통량 변동을 모형에 고려할 수 있도록 함

- 주요 변수는 차로 수, 도로 연장, 도로등급 등으로, 교차로의 교통 처리 능력과 도로 기능 위계를 함께 반영하였으며, 별도의 테이블로 분리하지 않고 차량 GPS 기반 입력 데이터에 통합함

4) 교통량 추정 학습데이터 구축 및 전처리

- 교통량 추정모형 학습을 위해 교통네트워크 가상링크 ID, 회전 ID, 일자, 시간 코드를 기준으로 관측교통량, 프로브 통행량·속도, 네트워크 연결성, 교차로 기하구조 데이터를 통합하여 학습데이터로 구축함
 - 학습데이터는 교통의 시공간적 특성과 물리적 요인을 함께 반영하도록 구성
 - 교통량 대비 프로브 대수는 변동성이 높아 학습 효율과 예측 성능 향상을 위해 민맥스 스케일링 기법으로 정규화하여 학습의 안정성과 균형을 확보함

〈표 3-43〉 세종특별자치시 교통량 추정 학습데이터 테이블 정의

No	Column	Type	설명	모형 사용 유무	비고
1	tn_v_link_id	Integer	교통 네트워크 가상링크ID		교통네트워크 단위
2	turn_id	Text	회전방향 ID	○	S01: 직진 R01:우회전 L01:좌회전 U01:유턴
3	date	Text	일자		yyyyMMdd
4	time_code	Integer	시간코드	○	0~287: 5분 단위 0~95: 15분 단위 0~23: 60분 단위
5	total_probe	Real	프로브		대
6	auto_spd	Real	속도	○	km/h
7	length	Real	대표 가상링크 길이	○	km
8	road_rank	Integer	대표 가상링크 도로등급	○	101 : 고속도로 102 : 도시고속도로 103 : 일반국도 104 : 특별광역시도 105 :

No	Column	Type	설명	모형 사용 유무	비고
					국가지원지방도 106 : 지방도 107: 시군도 108: 연결램프
9	lanes	Integer	대표 가상링크 차선수	O	
10	k	Real	대표 가상링크 밀도	O	프로브 / 속도
11	emd_type	Integer	대표 가상링크 읍면동 여부		0: 동 아님 1: 동 맞음
12	sido_id	Integer	대표 가상링크 시도ID		
13	traffic_type	Text	조사기관		
14	total	Real	관측교통량	O(라벨)	대
15	total_scale	Real	프로브 민맥스 스케일링값	O	0~1 스케일링값
16	total_1	Real	연결된 관측교통량_1	O	대
17	total_scale_1	Real	연결된 프로브 민맥스 스케일링값_1	O	0~1 스케일링값
18	total_2	Real	연결된 관측교통량_2	O	대
19	total_scale_2	Real	연결된 프로브 민맥스 스케일링값_2	O	0~1 스케일링값
			...		
54	total_20	Real	연결된 관측교통량_20	O	대
55	total_scale_20	Real	연결된 프로브 민맥스 스케일링값_20	O	0~1 스케일링값

다. 교차로 회전교통량 추정모형 개발

- 회전교통량 관측자료가 존재하지 않는 미관측 교차로를 대상으로 직진, 좌회전, 우회전 방향별 교통량을 추정하기 위한 LightGBM 기반 머신러닝 모형을 개발함
- 회전교통량 추정에 가장 적합한 모델을 선정하기 위하여 동일한 분석 환경 기준 Ridge, Elastic Net, Random Forest, Extra Trees, XGBoost, CatBoost, LightGBM 모델을 대상으로 추정 성능 평가를 진행하였고, LightGBM이 5분, 15분, 60분 모두에서 가장 우수한 추정 정확도를 보여 선정함

- 데이터 전처리 과정에서는 범주형 변수를 모형이 인식할 수 있도록 원-핫(One-Hot Encoding)을 적용하였으며, 전체 학습 파이프라인을 자동화하여 데이터 입력에서 예측 산출까지 일관된 절차로 수행되도록 구성함
- 또한 교차검증(K-Fold Cross Validation) 기법을 적용하여 예측값을 반복 학습함으로써 과적합을 방지하고 일반화 성능을 강화함

1) 결측치 및 이상치 처리

① 결측치 처리

- 학습데이터 특성상 일부 링크에서 특정 피쳐 값이 생성되지 않아 결측치 발생할 수 있음
- 결측값이 포함될 경우 모형 학습의 안정성과 계산 일관성이 저하될 수 있으므로 모든 입력 피쳐에 대해 결측값이 존재하는 항목은 값을 '0'으로 대체함
 - 입력 피쳐 대부분이 교통량, 속도, 도로길이 등 연속형 수치형 변수로 구성되어 있어, 결측값을 0으로 처리하더라도 데이터의 왜곡이 크지 않음

② 이상치 처리

- 학습데이터에는 프로브 데이터 수집 오류나 관측교통량 입력 오류 등으로 발생한 이상치(outlier)가 존재할 수 있음
- 이상치는 모델 학습 시 예측 오차를 증가시키고 회귀계수의 안정성을 저하시키므로 전처리 과정에서 제거함
 - 관측교통량 대비 프로브 교통량의 비율이 0.01 미만 또는 0.4 이상인 경우 이상치로 간주하여 제외
 - 관측교통량이 프로브 교통량보다 작은 경우 데이터 신뢰도가 낮다고 판단하여 학습대상에서 제외

2) 원-핫 인코딩

- 학습데이터 내 범주형(categorical) 변수에 대해 이를 모델이 인식하고 계산할 수 있도록 원-핫 인코딩(One-Hot Encoding) 방식을 적용하여 수치형으로 변환함
 - 원-핫 인코딩은 각 범주를 독립된 이진 벡터로 변환하여, 변수 간에 존재하지 않는 순서나 크기 관계가 인위적으로 반영되는 문제를 방지함
 - 예를 들어 '좌회전', '직진', '우회전' 범주는 각각 0과 1의 조합으로 표현되어 알고리즘이 이를 수치 크기로 오해하지 않도록 함
- ColumnTransformer를 활용해 수치형과 범주형 변수를 동시에 처리하는 전처리 파이프라인을 구축하였으며, handle_unknown='ignore' 옵션을 적용해 학습 시 존재하지 않았던 새로운 범주형 데이터 값이 등장하더라도 오류 없이 처리되도록 함

3) 하이퍼 파라미터 최적화

- LightGBM 기반 회전교통량 추정 모형의 일반화 성능을 극대화하기 위해, 트리 복잡도, 학습률, 노드 최소 표본, 무작위성, 정규화와 관련한 파라미터에 대한 최적화 과정을 수행함
 - 하이퍼 파라미터는 모형의 복잡도와 예측 정확도를 동시에 좌우하는 요소로, 적절한 조합을 찾는 것이 모형의 신뢰성과 효율성을 확보하는 데 필수적임
- <표 3-46>과 같이 파라미터가 최적 조합으로 선정(MAPE 성능, 계산 효율성, 과적합 위험을 종합적으로 고려한 결과)되었으며, 최종적으로 동일한 설정을 수집 주기별 모든 모델에 일괄 적용함

<표 3-44> 파라미터 최적화

구분	파라미터	내용	탐색 영역 및 선택
트리 복잡도	num_leaves	트리 최대 노드 수	15, 31 , 47
	max_depth	트리 최대 깊이	6, 8 , 10
학습률	learning_rate	부스팅 스텝에서 새로운 트리의 영향	0.02, 0.05 , 0.08
노드 최소 표본	min_child_samples	노드 최소 샘플 수	5 , 10, 20

구분	파라미터	내용	탐색 영역 및 선택
무작위성	subsample	트리 생성 시 사용 데이터 비율(행)	0.6, 0.8, 1.0
	colsample_bytree	트리 생성 시 사용 feature 비율(열)	0.6, 0.8, 1.0
정규화 (규제)	reg_lambda	L2 규제	0.5, 1.0, 2.0
	reg_alpha	L1 규제	0.0, 0.5 , 1.0

4) LightGBM 모델

- 교통량 추정모형은 Gradient Boosting Decision Tree 기반의 머신러닝 알고리즘인 LightGBM 기반 회귀 구조로 설계됨
 - 해당 모형은 대용량 데이터와 고차원 특성을 가진 데이터에 대한 학습 속도와 메모리 효율성에 강점이 있음
 - 히스토그램 기반 알고리즘으로 연속 데이터를 이산화하여 처리하며, 리프 중심 트리 분할(Leaf-wise) 방식을 통해 예측 오류 손실을 최소화 함
 - GOSS(Gradient-based One-side Sampling) 기법을 통해 기울기가 큰(학습이 덜 된) 데이터에 집중하고, 기울기가 작은 데이터는 일부만 샘플링하여 계산량을 절감하면서 정확도를 유지함
 - EFB(Exclusive Feature Bundling) 기법을 통해 원-핫 인코딩 등 희소한 고차원 특성을 묶어 차원을 축소함으로써 학습 속도를 개선함

다. 교차로 회전교통량 추정 결과 분석

1) 회전교통량 추정결과

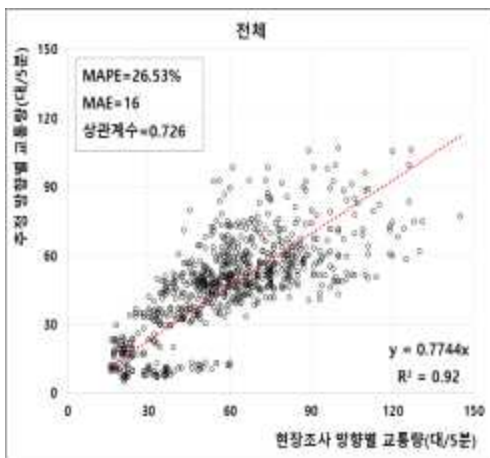
- 수집 주기별 회전교통량 추정 결과 분석을 통해 집계 단위가 확대될수록 상대 오차가 감소하고 설명력이 향상됨
 - (5분) MAPE 26.53%, MAE 16대, 상관계수 0.726, $R^2=0.92$
 - (15분) MAPE 22.57%, MAE 36대, 상관계수 0.836, $R^2=0.9424$
 - (60분) MAPE 13.76%, MAE 80대, 상관계수 0.916, $R^2=0.9772$

- 상관계수, R^2 , 산점도상 회귀선 기울기의 분석에서도 집계 단위의 확장에 따라 안정적인 추정 결과를 확인할 수 있음
 - 상관계수는 (5분) 0.726 (15분) 0.836 (60분) 0.916로 나타나 집계 단위 확대에 따라 추정치-실측치 간 선형 관계가 강화됨을 확인함
 - R^2 는 (5분) 0.92 (15분) 0.94 (60분) 0.98로 나타나 집계 단위 확대에 따른 설명력을 확보하고 특히 60분 기준 실측 변동의 약 98%를 설명하는 높은 예측 일관성을 보임
 - 회귀선 기울기는 (5분) 0.7744x (15분) 0.8595x (60분) 0.9967x로 집계 단위 확장에 따라 1에 수렴하는 경향성을 보임
- 방향별로 분석한 결과에서도 집계 단위의 확대에 따른 추정 오차율 감소의 경향성이 공통적으로 나타남
 - 좌회전은 MAPE 기준 (5분) 33.87%, (15분) 27.68%, (60분) 25.72%로 집계 단위의 확대에 따라 오차율이 개선되거나 가장 높은 오차율 수준을 보임
 - 직진은 시간 집계 단위가 확장될수록 오차율 및 MAE의 개선 폭이 가장 크며, MAPE 기준 (5분) 25.13%, (15분) 21.91%, (60분) 8.94%로 나타나며, 60분 기준 MAPE 10% 미만으로 실제 활용 가능 수준의 정확도를 보임
 - 우회전은 MAPE 기준 (5분) 21.02%, (15분) 15.72%, (60분) 11.81%로 5분, 15분 단위에서 가장 낮은 오차율을 보이고, 60분 기준에서도 MAPE 11.81%로 안정적인 성능을 확보함
- 수집 주기 단위의 확대에 따라 추정의 정확도가 전반적으로 개선됨을 확인하였으며 이후 교통량 모니터링·계획 등 실제 활용 수준의 분석 시에는 15분 또는 60분 기준의 추정 결과를 활용하여 안정적 정확도를 확보가 필요함

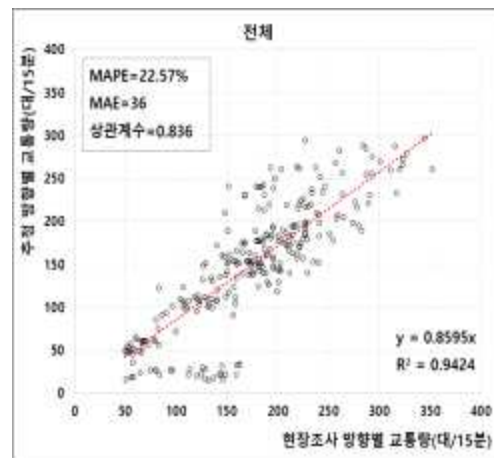
〈표 3-45〉 수집 주기별 방향별 회전교통량 추정 결과

수집 주기별 방향별 회전교통량 추정 결과						
구분	5분		15분		60분	
	MAPE(%)	MAE(대)	MAPE(%)	MAE(대)	MAPE(%)	MAE(대)
좌회전	33.87	16	27.68	32	25.72	112
직진	25.13	16	21.91	38	8.94	65
우회전	21.02	15	15.72	35	11.81	82
전체(교차로)	26.53	16	22.57	36	13.76	79

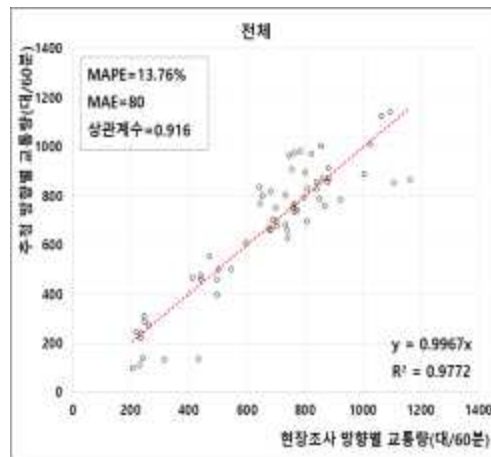
수집 주기 5분 - 분석 결과



수집 주기 15분 - 분석 결과



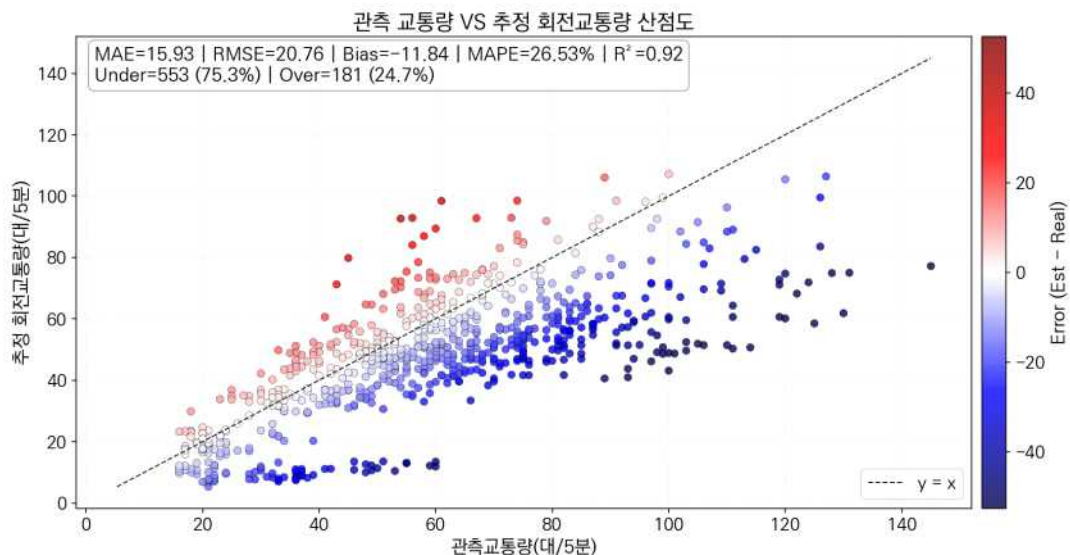
수집 주기 60분 - 분석 결과



2) 회전교통량 추정결과 산점도 분석 - 5분 기준

- 산점도는 관측 교통량(대/5분)과 추정 회전교통량(대/5분)의 대응 관계를 시각화한 것으로, 대각선($y=x$)에 가까울수록 예측 정확도가 높음

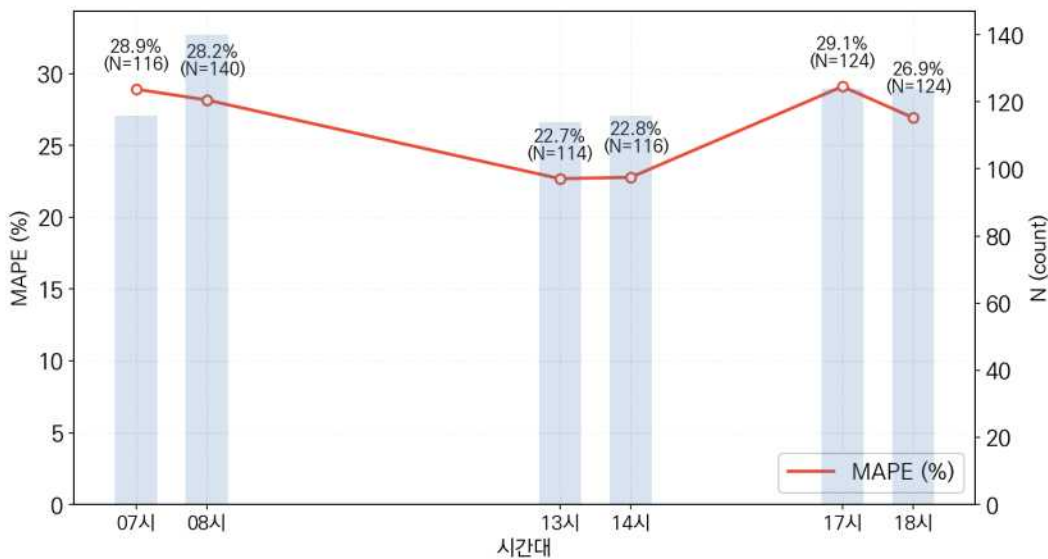
- 점의 색상은 오차의 부호와 크기를 표현하며 붉은색은 과대추정, 푸른색은 과소추정을 의미함
 - 동일 관측량에서 색이 혼재하면 오차 분산이 크고, 특정 색이 우세하면 편향을 시사함
- 전체 성능 지표는 MAE 15.93대/5분, RMSE 20.76대/5분, Bias -11.84대/5분, MAPE 26.53%, 결정계수 R^2 0.92로 나타남
- 5분 기준 평균 약 16대의 절대오차, 상대오차 약 27% 수준이며, $R^2 \approx 0.92$ 로 실측 교통량 변동의 약 92%를 설명함
 - RMSE가 MAE보다 약 5대 높게 나타나 일부 큰 오차 사례가 전체 성능을 저하시키고 있으며, 해당 케이스에 대해 데이터 품질 검증 및 사후보정의 우선 대상으로 검토할 필요가 있음
- 편향(Bias)이 -11.84대/5분이고, 과소추정 553건(75.3%)으로, 전반적으로 약 12개 가량 과소추정하는 경향이 존재함
- 구간별로 분석한 결과, 저·중간 교통량대(10~40대/5분)에서 점들이 $y=x$ 부근에 밀집하여 안정적 성능을 보이고 있음
- 반면 고교통량대(60대 이상/5분)에서는 과소추정 경향이 뚜렷해짐을 확인함



〈그림 3-46〉 관측 교통량과 추정 회전교통량 산점도

3) 시간대별 추정 회전교통량 신뢰도 분석 - 5분 기준

- 시간대별 추정 회전교통량의 신뢰도를 MAPE(%) 기준으로 비교한 결과, 전체 22.7~29.1%로 모든 시간대에서 20% 이상의 오차율을 보임
 - 오전 첨두(7 - 9시)에는 MAPE 28.2%~28.9%의 수준으로 나타나, 출근 시간대의 교통량 변동과 혼잡 등으로 인한 예측 정확도가 저하된 것으로 보임
 - 오후 비첨두(13 - 15시)에는 MAPE가 22.7~22.8% 수준으로 오전 첨두 대비 약 6%p 개선되었으며, 상대적으로 교통량 수준이 낮아 안정적인 추정 결과가 나타난 것으로 보임
 - 오후 첨두(17 - 19시)에는 26.9~29.1%의 MAPE를 보이며 17~18시의 추정 결과가 전체 시간대 중 가장 높은 오차율로 나타났으며, 퇴근 시간대 교통량 집중으로 인해 오차가 확대한 것으로 보임

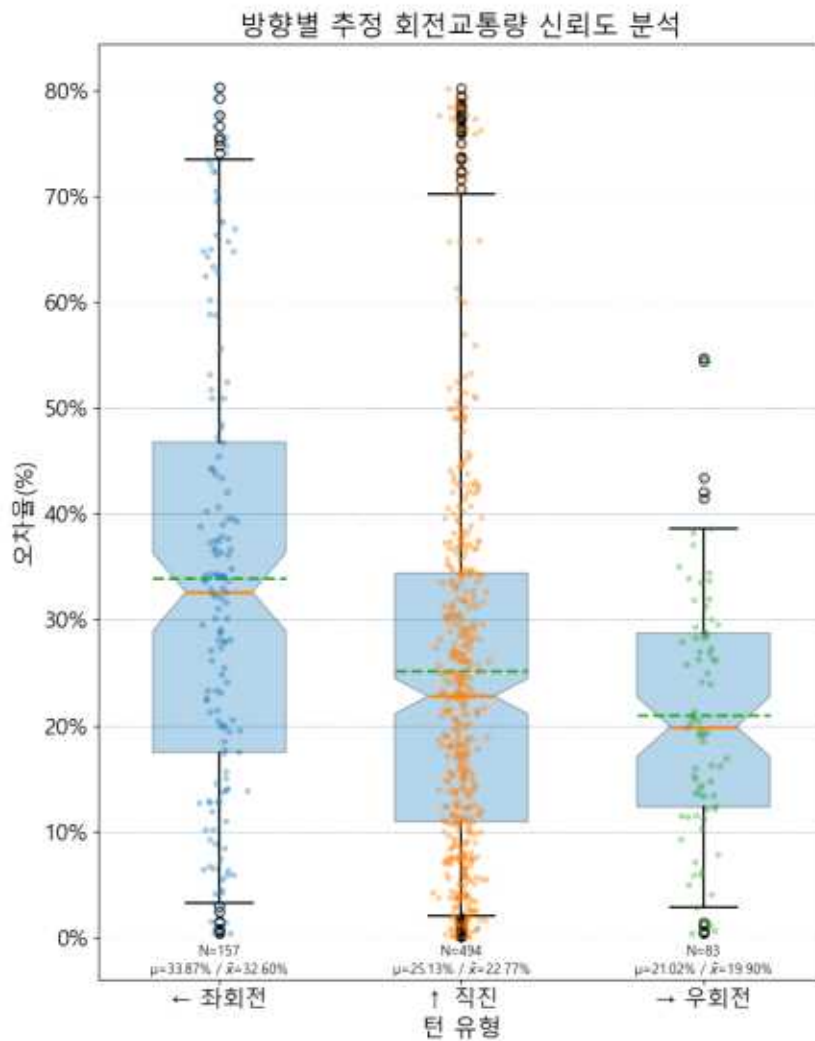


〈그림 3-47〉 시간대별 추정 회전교통량 신뢰도 분석

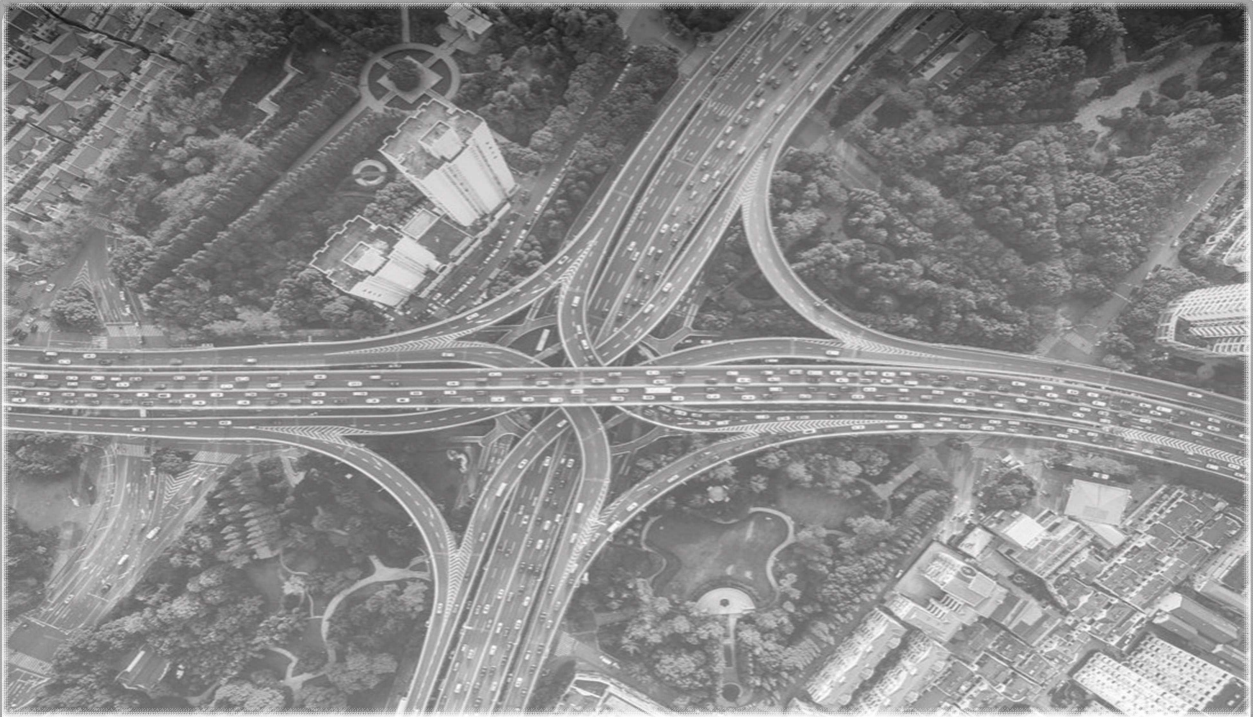
4) 방향별 추정 회전교통량 신뢰도 분석 - 5분 기준

- 좌회전·직진·우회전에 대한 5분 기준 방향으로 회전교통량 추정 오차율(%)을 분석한 결과, 평균 오차율은 좌회전 33.87%, 직진 25.13%, 우회전 21.02% 순으로 나타남
 - 중앙값은 좌회전 32.60%, 직진 22.77%, 우회전 19.90%로 분석됨

- 표본 규모는 직진 N=494, 좌회전 N=157, 우회전 N=83 순으로 나타남
 - 평균과 중앙값의 차이가 좌회전 약 1.3%p, 직진 약 2.4%p, 우회전 약 1.1%p로 비교적 근접한 것으로 나타남
- 사분위 범위와 산점도의 분포의 분석 결과는 다음과 같음
- 좌회전은 사분위 상단이 약 47%, 위스커 상한이 약 73%에 달하며, 이상치가 다수 관찰되어 세 방향 중 가장 큰 분산을 보임
 - 직진은 사분위 범위가 좌회전보다 좁게 나타나 안정적으로 확인되었으나, 많은 표본으로 인해 상하단 꼬리가 길게 분포하는 것을 보임
 - 우회전은 중앙값과 상자 범위가 가장 작으나 일부 이상치가 존재함



〈그림 3-48〉 방향별 추정 회전교통량 신뢰도 분석



제4장 차량 GPS 데이터 구축·분석 시스템 개선

제1절 차량 GPS 데이터 구축·분석 시스템 개요

제2절 차량 GPS 데이터 가공 모니터링 기능 개발

제3절 데이터 분석을 위한 신규 기능 개발

제4장 차량 GPS 데이터 구축·분석 시스템개선

제1절 차량 GPS 데이터 구축·분석 시스템 개요

- 데이터 수집과 가공 각 단계를 모니터링하여 데이터가 정상적으로 처리되고 있는지 실시간으로 확인할 수 있는 서비스 지원
- KTDB 기반 지도 및 차량 통행 데이터셋을 체계적으로 관리하고 분석에 활용할 수 있는 시스템 구축
- 데이터셋 가공·구축 시스템 관리
 - 서버 상태 모니터링 환경 구축을 통해 실시간 상태 확인 및 장애 예방 가능
 - 데이터 가공 모듈 개선 시 서버 업데이트 및 개발 정보 지속 관리
- 분석 시스템 관리
 - 수집 현황과 데이터 전처리 상태를 제공하기 위한 대시보드 서비스 제공
 - 공간 단위별로 구축된 교통망 성능 평가 지표 및 통행 분석 DB 업로드 및 버전 관리 지원
 - 데이터 작업 현황에 따라 데이터베이스(MySQL) 저장 및 관리

1. 차량 데이터 가공 이력 항목

가. 데이터 수집 현황

- 전처리 단계 이전에 원시 데이터의 파일 정보, 용량, 오류 등 기초적인 상태를 사전에 확인할 수 있는 정보 수집

〈표 4-1〉 데이터 수집 현황 로그 항목

구분		설명
원시 데이터	티맵 링크	원시 데이터의 이벤트 건수, 파일 건수, 용량 등 기초 정보의 오류 정보 수집
	티맵 포인트	
	DTG 버스, 트럭	

나. 데이터 전처리 정보

- 원시 데이터의 이상치 제거, 정렬, 좌표 변환, 시간 정보 변환 등의 전처리 과정을 각 단계별로 수행하며 일자별 오류 유형, 가공 용량, OBU 정보 등 전처리 관련 정보를 수집

〈표 4-2〉 데이터 전처리 로그 항목

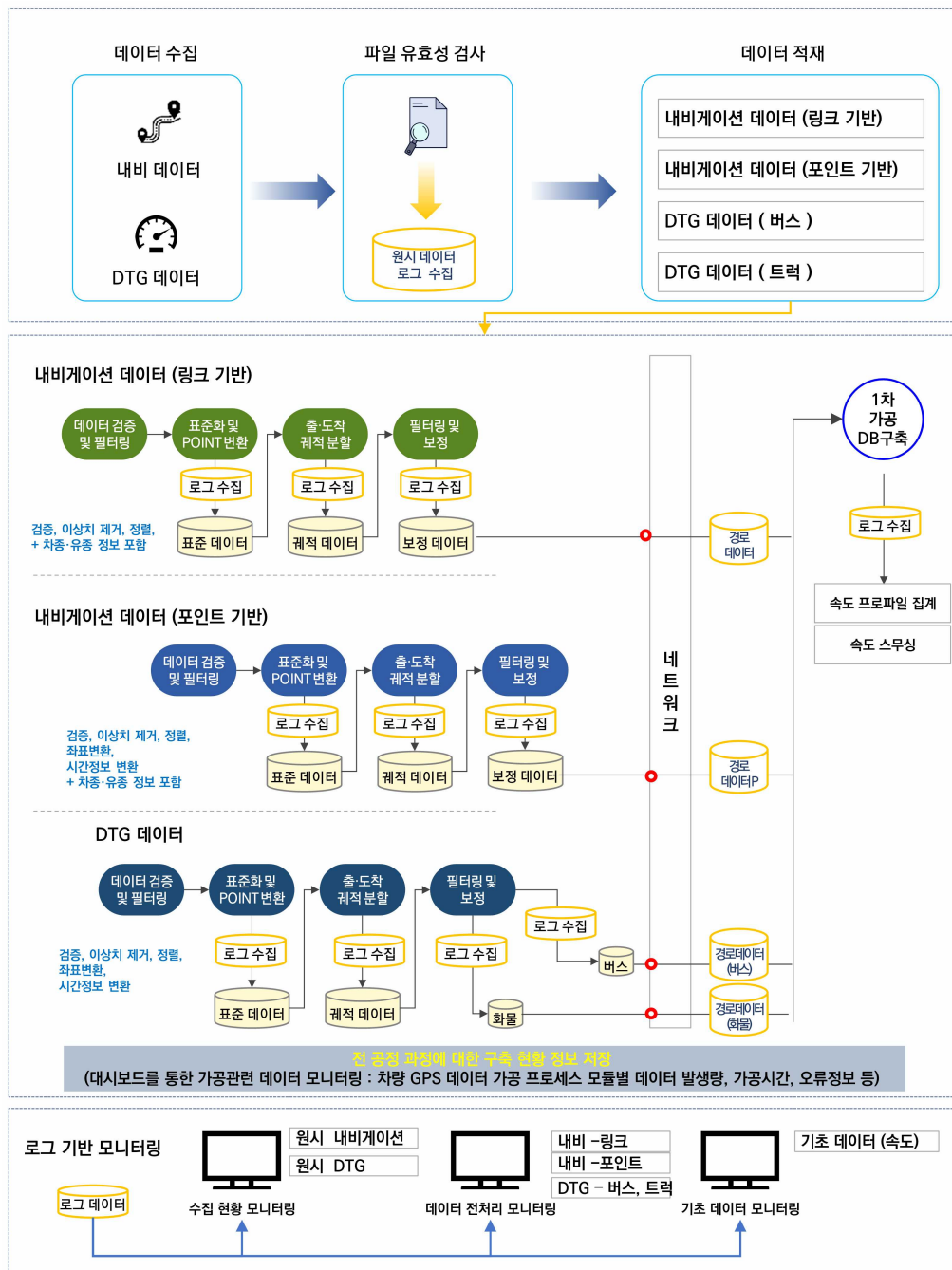
구분		설명
티맵 링크	링크 데이터 전처리	데이터 검증 및 필터링 정보
	링크 데이터 포인트화	표준화 및 Point 변환 정보
	링크 데이터 맵매칭	데이터 맵매칭 정보
티맵 포인트	포인트 데이터 전처리	데이터 검증 및 필터링 정보
	포인트 데이터 보정	표준화 및 Point 변환
	포인트 데이터 맵매칭	데이터 맵매칭 정보
	포인트 데이터 링크 경로 생성	데이터 링크 경로 생성 정보
DTG 버스, 트럭	포인트 데이터 전처리	포인트 데이터 전처리 정보
	포인트 데이터 표준화	포인트 데이터 표준화 정보
	포인트 데이터 보정	포인트 데이터 보정 정보
	포인트 데이터 맵매칭	포인트 데이터 맵매칭 정보
	포인트 데이터 링크 경로 생성	포인트 데이터 링크 경로 생성 정보

다. 기초 교통DB 구축 정보

- 데이터 전처리를 통해 가공된 데이터와 속도 프로파일이 집계, 스무싱 처리된 정보를 수집
 - 속도프로파일 집계 정보 및 이상치 정규화 분포 정보

제2절 차량 GPS 데이터 가공 모니터링 기능 개발

- 원시 데이터 수집 및 데이터 전처리에서 이상치 제거, 정렬, 좌표 변환, 시간 정보 변환 등의 전처리 과정을 각 단계별로 수행하며, 일자별 오류 유형, 가공 용량, OBU 등 수집된 로그 기반으로 모니터링 기능 개발



<그림 4-1> 모니터링 프로세스

가. 기능 구성

1) 데이터 수집 현황

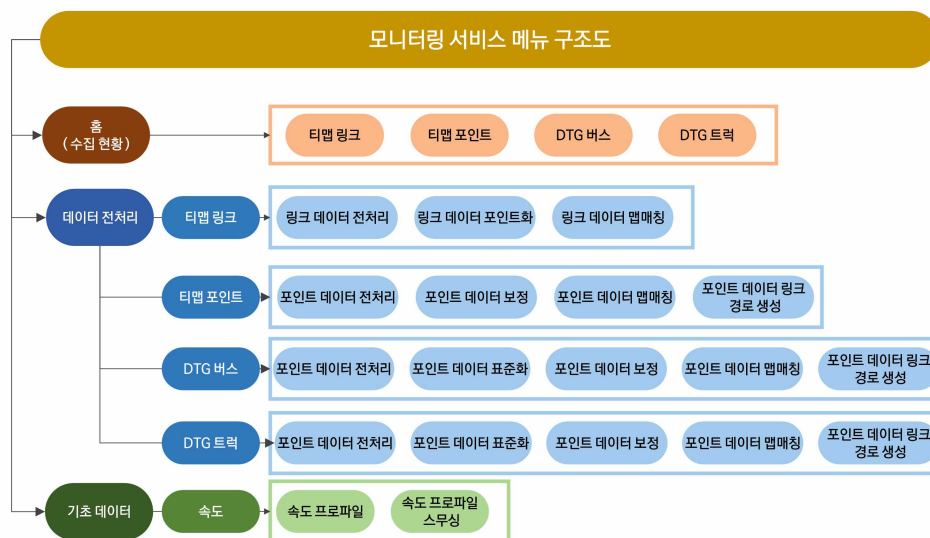
- 원시 데이터의 파일 유효성 검사 정보를 테이블 및 그래프로 제공
 - 주요 항목 : 파일 수, 오류 파일 수, 총용량 등
 - 년/월/일 별 데이터 용량, 건수를 시각적으로 표현

2) 데이터 전처리

- 이상치 제거, 정렬, 좌표 변환, 시간 정보 변환 등의 데이터 전처리 작업에서 발생하는 로그 정보를 테이블 및 그래프로 제공
 - 주요 항목 : 작업별 처리 데이터 수, 처리 성공 및 오류 수 등
 - 전처리 과정에서 데이터의 상태, 정보 등 각 전처리 과정에 맞는 정보 시각적으로 표현

3) 기초 데이터

- 전처리 데이터와 속도 프로브 데이터에서 생성되는 기초 데이터의 로그 정보를 테이블 및 그래프로 제공



〈그림 4-2〉 모니터링 메뉴 구조도

나. 기능 구축 결과

- 선택한 년도, 기간에 따라 이벤트 건수, 파일 수, 용량 및 데이터 현황 등을 테이블을 통해 사용자는 데이터를 구체적으로 확인할 수 있도록 함



<그림 4-3> 데이터 수집 현황 모니터링 화면

- 선택한 표출 컬럼에 따라 테이블 데이터가 구성되며, 각 전처리 단계에서 발생하는 수집된 로그 정보를 자세하게 확인할 수 있도록 함



<그림 4-4> 데이터 전처리 모니터링 화면

- 조건 설정에 따라 각 데이터 기초에서 발생하는 수집된 로그 정보를 자세하게 확인할 수 있도록 함



<그림 4-5> 기초 데이터 모니터링 화면

- 조건 설정
 - 데이터 선택 : 각 데이터 기초 데이터 항목을 표출 및 해당 항목을 선택하여 모니터링할 수 있도록 함
 - 네트워크 선택 : 데이터 기초에 사용되는 네트워크를 선택하여 모니터링할 수 있도록 함
 - 년도 선택 : 각 데이터 기초에 수집된 로그 정보를 년도별로 모니터링할 수 있도록 함
 - 기간 설정 : 각 데이터 기초에 수집된 로그의 기간을 설정하여 모니터링할 수 있도록 함
 - 표출 컬럼 선택 : 수집된 기초 데이터의 세부 정보를 선택하여 사용자 맞춤형 모니터링 제공



<그림 4-6> 기초 데이터 조건 설정 화면

○ 시각화

- 조건 설정에 따라 각 데이터 기초에서 생성되는 용량, 가공 시간, 입력 OBU, 출력 OBU 등 차트를 통해 분석할 수 있도록 함



<그림 4-7> 기초 데이터 시각화 화면

○ 상세 내용

- 조건 설정에 따라 각 데이터 기초에서 발생하는 수집된 로그 정보를 자세하게 확인할 수 있도록 함

날짜	프로세스 코드	데이터 유형	가공 데이터의 기준네트워크	프로세스의 잡 진행 시간	입력 데이터 용량	출력 데이터 용량
2023	속도 프로파일 스무싱	DTG_TAXI	7레벨	600.0분	54.74MB	38.77MB
입력 데이터 이벤트 개수		스무싱 실패 개수	입력 데이터 오류 개수	출력 데이터 오류 개수	크기 오류 개수	출력 데이터 이벤트 개수
4,901,729,815		5,273,096,119	4,810,668,951	33,850,115	50,470,175	35,103,855

<그림 4-8> 기초 데이터 상세 내용 화면

제3절 데이터 분석을 위한 신규 기능 개발

가. 교통지표 우선순위 분석 기능 개발

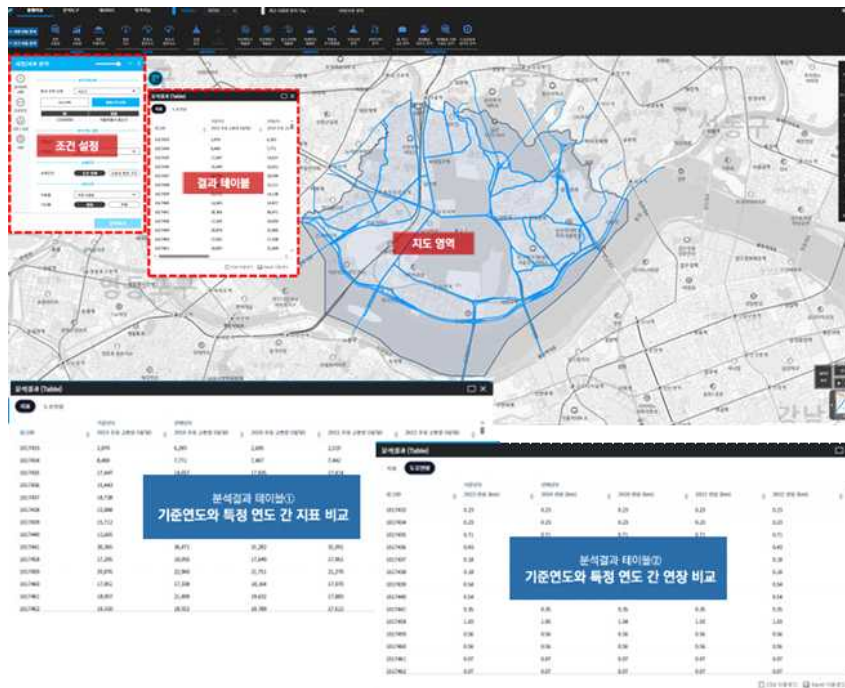
- (기능 목적) 교통SOC의 진단 및 평가를 위한 교통망 성능평가지표 DB 구축 결과를 지역별로 한눈에 순위별로 시각화하여 분석할 수 있도록 개발함
- 조건 설정
 - 분석지역 선택: 시도 또는 시군구 단위 선택 가능하도록 함
 - 기간/시간 선택: 분석 지표에 대한 기간과 시간을 선택 가능하도록 함
 - 지표 선택
 - 지표별(주요): 분석 기준이 되는 대표지표 1개를 선택 가능하도록 함
 - 지표별(보조): 주요지표와 연관되는 복수개 지표를 선택 가능하도록 함
 - 도로조건 선택: 선택한 분석지역 내 도로등급, 차선수, 링크유형 조건으로 2차 필터링 할 수 있도록 함



〈그림 4-9〉 교통지표 우선순위 분석 예시 화면

나. 연도별 교통지표 변화분석 기능 개발

- (기능 목적) 교통정책 효과 분석을 위해 교통망성능평가지표를 연도별로 변화된 결과를 한눈에 시각화하여 분석할 수 있도록 개발함
- 조건 설정
 - 분석영역 선택: 행정구역 단위 선택 또는 사용자 선택 영역
 - 분석연도 선택: 기준연도 선택, 기준연도를 기준으로 분석하고 싶은 연도 선택
 - 분석연도에 대한 상세조건 선택
 - 기준연도만 선택 시: '전체 도로', '신성 및 변경 구간' 중 택 1
 - 선택연도 선택 시: '전체 도로', '신설 및 변경 구간 포함' 중 택 1
 - 분석지표 선택: 분석지표 및 지표에 대한 기간(평일/주말, 시간대) 선택
 - 비교 분석(사용자 정의): 비교 분석은 기준연도에서 선택연도에 대한 증감률 정보 제공(디폴트), 사용자가 수식을 적용하고 결과를 확인할 수 있도록 함
 - 결과: 설정에 의한 최종 결과는 테이블로 표출



<그림 4-10> 연도별 교통지표 변화분석 예시 화면



제5장 결론 및 향후 계획

제5장 결론 및 향후 계획

1. 사업결과 요약

- 본 과업은 전국 상세도로망 네트워크 기준 교통망 성능평가지표 구축 및 지표 구축을 위한 기초교통DB를 가공·구축하였음
- 2024년 기준 전국 개별차량 경로DB 구축, 기초교통 DB 구축, 교통량 추정, 교통망 성능평가지표에 대해 구축하였음
- 2024년 기준 교통망 성능평가지표를 산정하기 위한 교통량 추정 입력데이터를 구축하여 추정교통량DB를 구축함
 - 교통량 추정 알고리즘은 학습 데이터 누적방식의 교통량 추정모형으로 입력데이터의 품질이 중요하므로 관측교통량과 프로브 통행량의 이상치를 제거하였음
- 회전교통량 추정DB 구축가능성 검토를 위해 특정 지역을 대상으로 프로토타입 회전교통량 추정을 수행하였으며, 특정 1일 회전교통량을 추정DB 구축함
- 2024년 기준 교통망 성능평가지표를 활용하여 통계청의 “차량 GPS 기반 도로 유형별 주행거리”를 실험적 통계로 공표

2. 사업결과의 활용방향

- 데이터 분석·활용시스템을 이용하여 이용자 문의나 공문을 통해 이용자 맞춤형 데이터를 제공함
 - 특히, 타 기관에서 운영하는 시스템이나 사업의 기반데이터로 이용되어 매년

제공하고 있음

- 주로 중앙부처 및 지자체의 교통운영 및 계획을 위한 정책 수립 근거자료로 사용되었으며, 최근 빅데이터의 관심이 커지면서 학계·개인의 연구 목적으로도 폭넓게 활용되고 있음
- 서울대, 한양대, 인하대 등 대학교에서 개인연구를 위한 데이터 요청 건이 많음
 - 데이터 분석·활용 시스템인 View-T를 통해 통행지표 데이터 및 경로데이터 (분석도구기능)를 제공하였으며, 데이터 제공 실적이 3,744에서 17,220건으로 전년 대비 약 25% 증가함

<표 5-1> 데이터 분석·활용 시스템 활용 실적

기능	설명	데이터 건수
링크 다운로드	도로 구간(링크)별로 통행지표 데이터를 다운받을 수 있음	6,621
월 데이터 다운받기	월별로 구성된 데이터를 다운받을 수 있음	4,865
통행지표 데이터 다운로드	통행지표를 행정구역별, 도로별, 시간대별 등을 선택하여 데이터를 다운받을 수 있음	4,725
분석도구 결과 다운로드	분석도구를 이용하여 분석한 결과를 다운받을 수 있음	1,009
총합		17,220

〈표 5-2〉 데이터 제공 및 분석 지원 주요 사례

1차 가공DB (속도분포, 경로데이터 등)	▣ 법정계획 및 정책 수립 지원 · 국가-지자체 도로정책 수립 지원 연구	수시
	▣ 데이터 관련 사업 지원 · 부동산 빅데이터 플랫폼 구축사업 · 서울시 개인차량 이동행태 분석 사업	상시 수시
	▣ 민간 및 학술적 연구 지원 · 온실가스 저감을 위한 국토도시공간 계획 및 관리기술 연구 · 영업용 화물자동차의 충전인프라 구축 연구 · 전기차 배터리 수명개선용 BMS 알고리즘 연구 · 국도30호선 차량 과속관련 교통안전 연구	수시
	▣ 법정계획 및 정책 수립 지원 · 통계청의 차량 GPS 기반 도로유형별 주행거리 실험적 통계 지원 · 도로건설계획(2026-2030) 수립 연구 · 지방 교통시설 국고보조금 관리 방안 수립 연구	상시 수시
기초지표 (교통량, 속도, 차량주행거리)	▣ 데이터 관련 사업 지원 · 교통 접근성 지표 산출을 위한 기초데이터 지원 · 교통 빅데이터 플랫폼 구축사업 · 환경소음측정망 적정 측정 지점 선정 사업 · 탄소공간지도 기반 계획지원 기술개발 사업 · COMPAS 국토도시데이터 분석과제 공모전	상시
	· 머신러닝을 이용한 소음지도 작성 사업 · 교통 빅데이터 활성화를 위한 지자체 지원사업 · 경기도 공공투자사업 투자심사 사전검토	수시
	▣ 민간 및 학술적 연구 지원 · 국토교통부 주관 데이터 공모전 · 도심항공 UAM 운용을 위한 연구	상시
	· 대도시권 광역교통 주관 경진대회 · 울릉도 교통현안 분석 · 탄소중립 수송부문 감축 고도화 기술 개발 연구	수시
성능평가지표 (혼잡지표, 환경지표, 안전지표 등)	▣ 법정계획 및 정책 수립 지원 · e-나라지표의 도로교통 혼잡비용 지원 · 행정중심복합도시 장래 도로계획 수립	상시 수시
	▣ 데이터 관련 사업 지원 · 대도시권 광역교통 통계 분석 · 교통 빅데이터 플랫폼 구축사업 · COMPAS 국토도시데이터 분석과제 공모전	상시
	· 교통 빅데이터 활성화를 위한 지자체 지원사업	수시
	▣ 민간 및 학술적 연구 지원 · 국토교통부 주관 데이터 공모전 · 대도시권 광역교통 주관 경진대회 · 인천광역시 보통교부세 진단 및 보정수요발굴 연구 · 해운대구 탄소중립 녹색성장 기본계획 수립연구	상시 수시

* 혼잡지표(교통혼잡비용, 혼잡시간강도, 혼잡빈도강도, 혼잡시평균속도), 환경지표(이산화탄소배출량, 일산화탄소배출량, 미세먼지배출량, 질소산화물배출량, 휘발성유기화합물배출량), 안전지표(운전자피로도, 과속비율, 속도편차)

3. 향후 계획

- 전국 2024년 교통망 성능평가지표 산정결과과 경로DB를 View-T 웹서비스를 통해 공표
- 2024년 기준 교통망 성능평가지표를 활용하여 통계청의 “차량 GPS 기반 도로 유형별 주행거리”를 실험적 통계로 공표
- 민간 모빌리티 데이터 제공 기관의 서버 환경과 수집 방식에 따라 동일한 컬럼을 가지고 있는 데이터라도 데이터 형태가 일부 다르게 나타나고 있으므로 이에 이전과 다른 이상치를 판별할 수 있는 원시데이터 이력관리, 메타데이터 생성 등 데이터의 품질을 향상시킬 연구가 필요함
- 민간 모빌리티 데이터 제공 기관의 데이터 판매 정책으로 인한 원시데이터 수집의 한계점이 존재하며, 가공데이터를 수집함에 따라 달라지는 데이터의 형태, 일관성을 보완할 수 있는 방법을 검토할 필요가 있음

9 2025년 국가교통조사 및 분석 차량 GPS 빅데이터 구축

