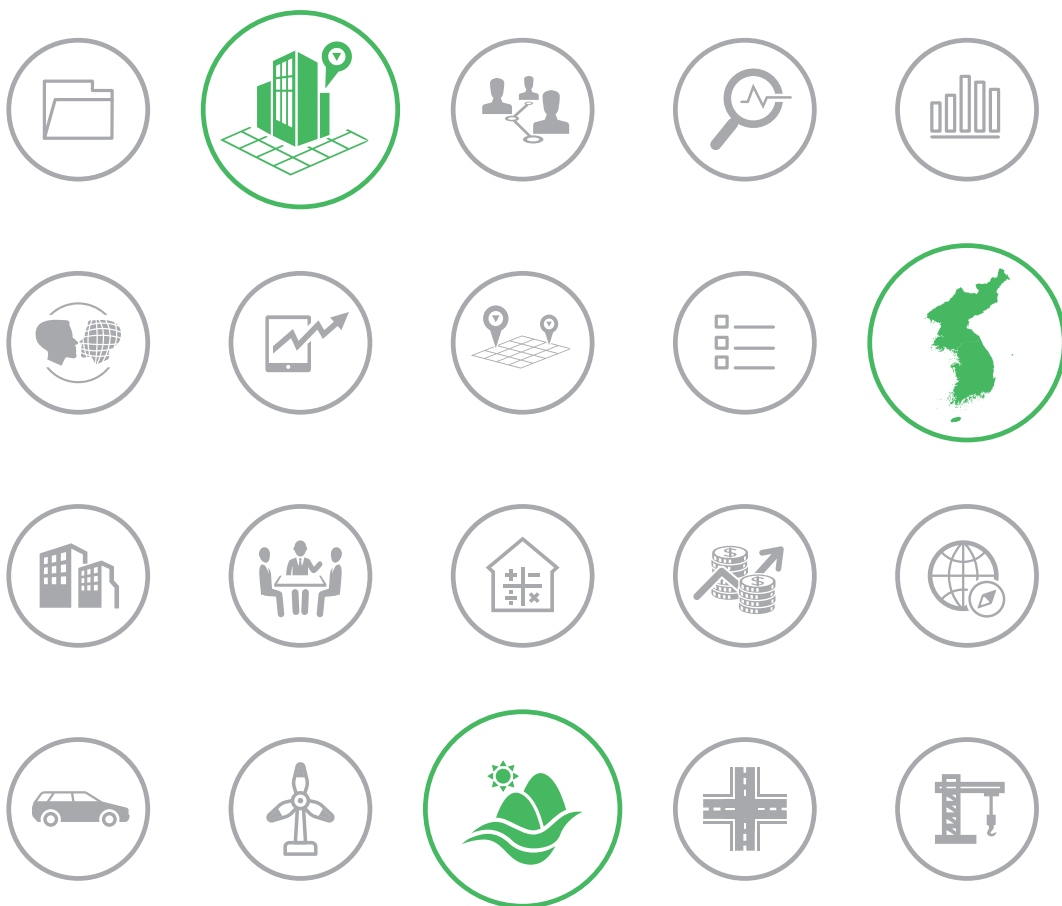




기본 | 22-28

위치기반 빅데이터를 활용한 시공간 활동패턴 분석연구

- 활동패턴의 유형 구분 및 특징을 중심으로 -

Spatio-temporal Activity Pattern Analysis by Mobile Big Data
-Activity Pattern Type and Characteristics-

김종학, 김호정, 윤서연, 연치형

기본 22-28

위치기반 빅데이터를 활용한 시공간 활동패턴 분석연구

- 활동패턴의 유형 구분 및 특징을 중심으로 -

Spatio-temporal Activity Pattern Analysis by Mobile Big Data
-Activity Pattern Type and Characteristics-

김종학, 김호정, 윤서연, 연치형

■ 저자

김종학 김호정 윤서연 연치형

■ 연구진

김종학 국토연구원 연구위원(연구책임)

김호정 국토연구원 선임연구위원

윤서연 국토연구원 연구위원

연치형 국토연구원 연구원

■ 연구심의위원

김태환 국토연구원 부원장

오성호 국토연구원 선임연구위원

이백진 국토연구원 선임연구위원

김준기 국토연구원 연구위원

오송천 국토교통부 철도건설과 과장

김형철 충남연구원 연구위원

주요 내용 및 정책 제안

FINDINGS & SUGGESTIONS



본 연구보고서의 주요 내용

- 1 스마트폰 앱으로 구축한 인간-시간-공간이 통합(joint)된 모빌리티 빅데이터로 하루 단위 개인의 활동패턴 구분기준과 특징을 데이터 가공 및 분석을 통해 파악함
- 2 개인의 하루 활동패턴 유형을 LNS(Link/Node/Stop over) 3가지 요인으로 구분하고 그 유형을 32가지로 제시
 - 활동패턴은 단순 활동패턴(SAP, Simple Activity Pattern), 복합 활동패턴(CAP, Complex Activity Pattern)으로 구분하였고 SAP가 CAP보다 약 4배 높았음
 - * 단순 활동패턴 : 하루 중 동일 목적지 1회 방문, 복합 활동패턴 : 하루 중 동일 목적지 2회 방문
 - 활동패턴을 지도상에 표출하기 위해 LNS로 구현되는 활동패턴 대시보드 개발
 - 단순 활동패턴(SAP) 81%, 복합 활동패턴(CAP)은 19%로 나타났고 SAP 유형 중 하나인 SAP 1(집-직장)이 52.8%로 32개 유형 중 가장 높았음
- 3 주중 활동시간(귀가 후 제외)은 10.3시간으로 2.1시간에 이동, 체류에 8.2시간 소비
 - 주중 활동시간이 가장 높은 곳은 울산 52.3시간, 낮은 곳은 강원 46.5시간으로 나타남
 - SAP 목적지 1회 증가 시 활동 시간은 주중 25분, 주말 63분 증가해 주말이 높았음

본 연구보고서의 시사점

- 1 (시공간 통합데이터 분석 방법) 개인별 시공간 데이터로 활동패턴 분석 방법 개발
- 2 (대시보드 개발) 개인 하루 단위 활동패턴을 지도상에 표현이 가능한 대시보드 개발
- 3 (활동패턴 구분기준 제시) 개인 활동패턴 유형을 LNS(Link/Node/Stop over)로 구분
- 4 (활용방안) 전염병, 경제둔화 등 사회변화에 따른 개인의 활동 변화 모니터링 가능



1. 연구의 개요

□ 연구의 배경 및 목적

- 인간의 활동은 물리적 공간에서 시간적 연속성을 가지고 발생하므로 인간-시간 - 공간 3개 데이터가 하나로 결합(joint) 될 때 개인의 활동패턴 파악이 가능함
- 스마트폰의 GPS 기능으로 개인의 시공간 정보가 별도의 조사 없이 빅데이터 형태로 구축 가능하게 됨
- 모빌리티 빅데이터의 공간 및 교통 분야 활용을 위해서는 원시데이터의 과학적인 가공과 분석을 통한 활동패턴 분석 방법 개발이 필요함
- 본 연구는 모빌리티 빅데이터 마이닝을 통해 하루 단위 개인 활동패턴의 유형을 구분하고 그 특징을 파악하고자 함

* 활동패턴 : 개인의 활동을 노드(물리적 목적지), 링크(이동), 경유지(시간대별 목적지)로 형상화한 것

□ 연구내용

- 본 연구는 모빌리티 앱 기반 빅데이터의 시공간 특성에 따른 데이터 가공과 시각화 과정 등을 반복하여 활동패턴 유형 구분과 그 특징을 분석
- 모빌리티 빅데이터로 활동패턴 제시를 위한 데이터 가공, 정제, 알고리즘 제시하였고 제시 방안을 토대로 모빌리티 빅데이터 시공간 특성 파악을 위한 활동패턴 대시보드 개발
- 모빌리티 빅데이터 활용 시도별 출근 및 귀가 통행 특성 및 차종별 출근 거리 등 분석
- 일 단위 개인 활동을 링크(L)/노드(N)/경유지(S)로 파악하여 활동패턴 유형을 구분하고 활동패턴 유형별 점유율, 요일별 분포 등을 분석

2. 데이터 가공 및 활동패턴 대시보드 개발

□ 원시자료 데이터 구축

- 분석데이터는 2022년도 상반기 중 휴일 영향이 적은 2022. 3. 14~4. 3 총 3주간 (21일) 데이터로 스타트업기업인 위드라이브로부터 구축

* 위드라이브(WEDRIVE) : 스타트업기업으로 2020년부터 운전자들의 이동정보 제공과 상응해 포인트(편의점, 커피숍 이용)를 주는 방식으로 데이터를 구축

- 위드라이브 가입자 1.3만 명의 3주간 자료 약 162만 개에 대한 일자별 통합, 공간조인 등의 기초가공 과정을 수행

* 원시자료 : 비 인식 개인 ID, 출발시간(월·일·시·분·초), 도착시간, 출발지/도착지 경유도 좌표 등 7개 정보로 구성

그림 1 | 원시데이터 취합 및 공간조인

Source	start	time begin	time end	origin lat	origin lng	destination lat	destination lng
1011000	44a73a7e3b3a54e73a7e	2022-03-12 15:14:00	2022-03-12 15:45:57	37.509754	127.039123	37.509752	127.040054
1011007	44a73a7e3b3a54e73a7e	2022-03-12 17:04:21	2022-03-12 17:11:15	37.509624	127.039168	37.509752	127.040746
1011008	44a73a7e3b3a54e73a7e	2022-03-12 19:50:51	2022-03-12 20:02:18	37.509769	127.039173	37.509746	127.040416
1011009	44a73a7e3b3a54e73a7e	2022-04-02 12:12:36	2022-04-02 12:45:23	37.510408	126.967125	37.510403	127.019852
1011100	44a73a7e3b3a54e73a7e	2022-04-02 12:54:02	2022-04-02 13:45:12	37.579944	127.081740	37.588124	127.044812
1011101	44a73a7e3b3a54e73a7e	2022-04-02 15:10:08	2022-04-02 15:44:16	37.587108	127.046875	37.579857	127.040339
1011102	44a73a7e3b3a54e73a7e	2022-04-02 15:56:23	2022-04-02 15:58:51	37.576713	126.944815	37.588124	127.044812
1011103	44a73a7e3b3a54e73a7e	2022-04-02 19:55:02	2022-04-02 19:59:14	37.587108	126.946763	37.588124	127.044812
1011104	44a73a7e3b3a54e73a7e	2022-04-02 17:10:53	2022-04-02 17:18:08	37.587108	126.946763	37.588124	127.044812
1011105	44a73a7e3b3a54e73a7e	2022-04-02 18:04:13	2022-04-02 19:02:17	37.588124	126.946763	37.588124	127.044812
1011106	44a73a7e3b3a54e73a7e	2022-04-02 19:02:17	2022-04-02 19:05:48	37.588124	126.946763	37.588124	127.044812
1011107	44a73a7e3b3a54e73a7e	2022-04-02 14:10:02	2022-04-02 15:08:08	37.588124	127.019852	37.588124	127.044812
1011108	44a73a7e3b3a54e73a7e	2022-03-14 10:25:46	2022-03-14 10:25:46	37.588124	127.019852	37.588124	127.044812

자료: 저자 작성

□ 데이터 가공

- 원시데이터의 7개 정보로 하루 단위 개인 활동패턴을 파악하고자 하루 중 개인의 통행 순번을 파악하는 TGN(Trip Generation Numbering) 작업을 수행함

* TGN: 일자별 개인별 통행 시·종점 위치와 시각 정보로 첫 통행, 중간통행, 마지막 통행, 일방향 통행 총 네 가지로 통행 발생 시점을 산정한 김중학 외(2021:67~71)의 연구내용을 적용함

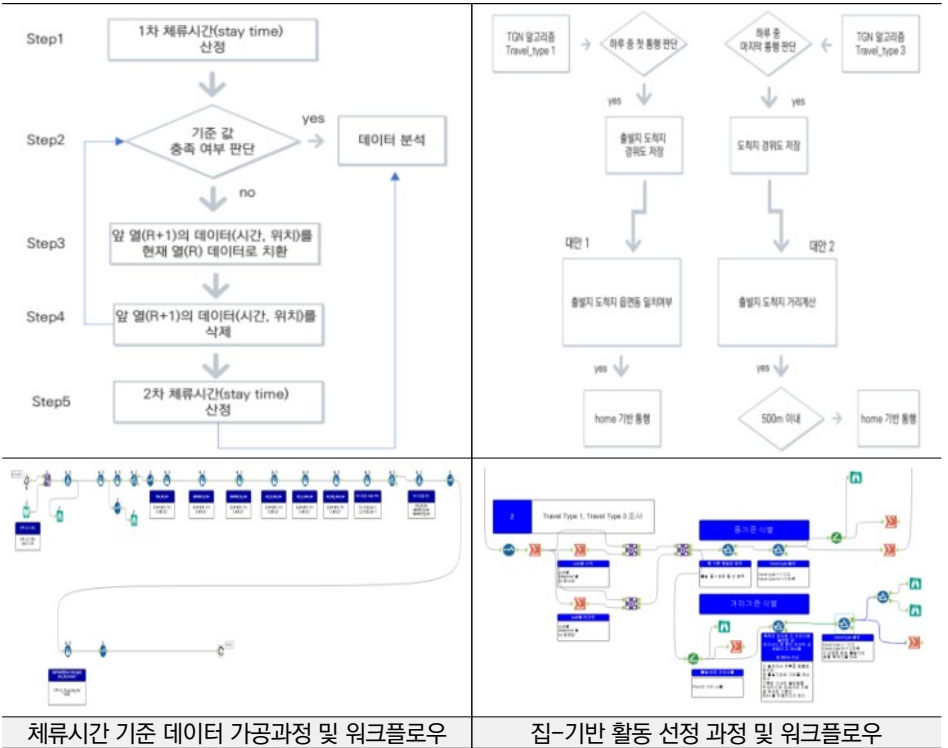
- TGN 작업과 병행하여 체류시간, 통행시간 등을 산정하였고 이후 최대 통행시간 초과, 체류시간 음수 등의 데이터를 제외하는 작업 등을 수행

- 본 연구 활동패턴은 다음의 2가지 조건을 충족하는 데이터에 대해 활동패턴 유형을 구분하고자 하였음

- 첫째, 개인의 하루 단위 활동은 시간과 공간적 연속성을 충족해야 함
- 둘째, 개인의 하루 단위 활동은 집에서 출발하여 집으로 귀가하는 활동만을 대상으로 해야 함

- 상기 2가지 조건에 동시에 부합하는 데이터를 선정하고자 다음 [그림 2]와 같은 과정을 거쳐 데이터 분석과정과 실제 수행을 위한 워크플로우를 작성
 - 시간적 공간적 연속성은 체류시간 10분 이상만 경유지로 판단하였고 공간적 연속성은 도착지와 출발지 이격거리를 500m로 가정하였음

그림 2 | 활동패턴 파악을 위한 주요 데이터 가공



자료: 저자 작성

□ 활동패턴 대시보드 개발

- 활동패턴 대시보드 개발은 원시자료의 시각화를 통한 데이터 오류 및 활동패턴 파악을 위해 필요함
 - 앞서 기술한 2가지 데이터 가공과정이 올바르게 진행되었는지를 파악하기 위해서는 데이터 시각화 과정이 필요함

- 활동패턴 대시보드는 상단 메뉴, 기본메뉴, 활동패턴 등 6개의 메뉴로 구성하였고 그 내용은 [그림 3]과 같음
 - 예로 설명하면 상단 메뉴는 활동패턴 유형, 시도, 차종, 차량 유형, 개인 ID, 일자 선택 메뉴로 구성함
 - 대시보드의 부가적 기능은 이동의 연속성을 시각적으로 검토함으로써 활동패턴별 수정 방안을 제시할 때 활용하였음

그림 3 | 활동패턴 대시보드 개발



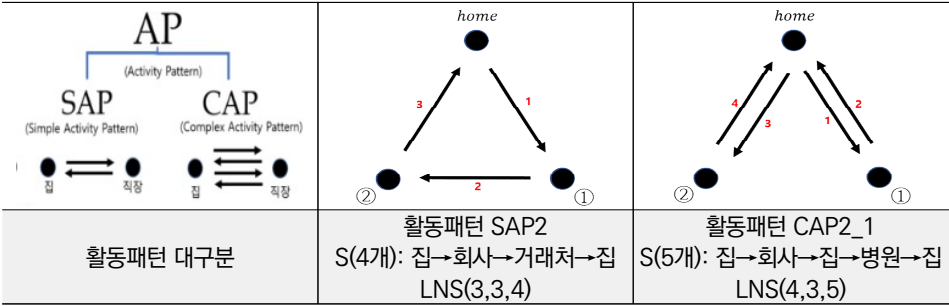
자료: 저자 작성

3. 활동패턴 유형 구분 및 특징 분석

□ 활동패턴 유형 구분

- 본 연구는 활동패턴(AP, Activity Pattern)을 하루 중 집 이외의 출발지(예: 직장)로의 회귀 통행 유무에 따라 단순 활동패턴(SAP, Simple Activity Pattern)과 복합 활동패턴(CAP, Complex Activity Pattern)으로 1차 구분함
- 활동패턴의 세부 유형은 링크(L, Link), 노드(N, Node) 이외에 경유지(S, Stop over)를 기준으로 이차적으로 소구분하였음
- 활동패턴 대시보드 개발과정을 통해 LNS(Link/Node/Stop over) 세 가지 요인으로 활동패턴 구분기준 제시, LNS 조합에 따라 활동패턴 유형을 구분
 - * 동일 목적지를 2회 이상 방문하는 복합 활동패턴을 지도상에 표현하기 위해서는 링크(L), 노드(N) 이외에 경유지(S)의 파악이 필요
 - * [그림 4]의 SAP의 세부 유형 중 하나인 L: 3개, N: 3개, S:4개로 LNS(3,3,4)인 조합은 SAP2로 구분됨

그림 4 | 활동패턴 유형 구분



자료: 저자 작성

- 단순 활동패턴(SAP)은 출발과 도착이 2회인 회귀 통행이 없어 링크(L) 개수와 노드(N) 개수가 동일한 반면 복합 활동패턴(CAP)의 경우 회귀 통행이 있어 경유지(S)의 개수가 노드, 링크 개수보다 큰 특징이 있음
- 약 6만 8천 개의 하루 단위 개인의 활동패턴을 링크(L)/노드(N)/경유지(S) 조합은 60개로 나타냈고 이 중 빈도수 낮은 것을 제외한 32개의 활동패턴 유형을 도출함

* SAP 8개, CAP 24개

* AP별 점유율 : AP1: 53.4%, AP2: 23.9% 등

□ 활동패턴 유형별 점유율

- 분석자료의 활동패턴 유형별 점유율은 단순 활동패턴(SAP) 약 81%, 복합 활동패턴(CAP)은 약 19%로 SAP가 CAP보다 약 4배 높았음
- 전체 32개 활동패턴 상위 10개 활동의 점유율은 94.4%, 이중 SAP 5개 점유율은 80.2%, CAP 5개 점유율은 14.2%로 나타남

* 활동패턴 상위 6개의 누적 점유율은 약 89%로, 10개 활동 중 약 9개를 상위 6개 패턴으로 설명할 수 있었음

그림 5 | 활동패턴 유형별 점유율

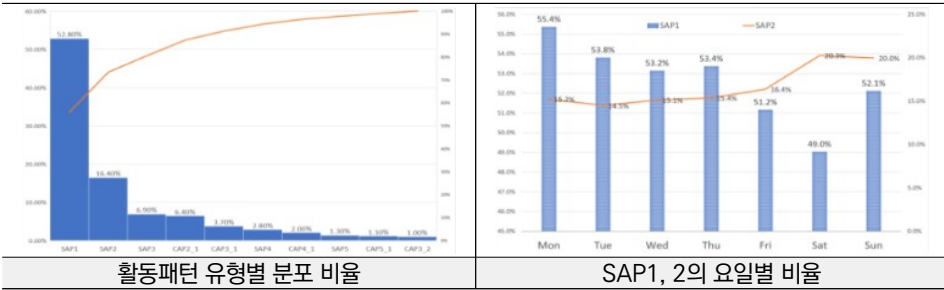
대구분	소구분	L	N	S	빈도수	비율
AP1 3개(1,2)	SAP1	2	2	3	36,024	52.8%
	CAP1_1	4	2	5	383	0.6%
	CAP1_2	6	2	7	7	0.01%
AP2 5개(1,4)	SAP2	3	3	4	11,182	16.4%
	CAP2_1	4	3	5	4,375	6.4%
	CAP2_2	5	3	6	385	0.56%
	CAP2_3	6	3	7	104	0.2%
	CAP2_기타	8	3	9	21	0.03%
AP3 5개(1,4)	SAP3	4	4	5	4704	6.9%
	CAP3_1	5	4	6	2,581	3.7%
	CAP3_2	6	4	7	684	1.0%
	CAP3_3	7	4	8	123	0.2%
	CAP3_기타	9	4	10	38	0.1%
AP4 5개(1,4)	SAP4	5	5	6	1,890	2.8%
	CAP4_1	6	5	7	1,393	2.0%
	CAP4_2	7	5	8	501	0.7%
	CAP4_3	8	5	9	102	0.2%
	CAP4_기타	10	5	11	43	0.1%
AP5 5개(1,4)	SAP5	6	6	7	868	1.3%
	CAP5_1	7	6	8	727	1.1%
	CAP5_2	8	6	9	278	0.4%
	CAP5_3	9	6	10	95	0.1%
	CAP5_기타	12	6	13	31	0.1%
AP6 4개(1,3)	SAP6	7	7	8	394	0.6%
	CAP6_1	8	7	9	324	0.5%
	CAP6_2	9	7	10	144	0.2%
	CAP6_기타	12	7	13	67	0.1%
	SAP7	8	8	9	196	0.3%
AP7 3개(1,3)	CAP7_1	9	8	10	151	0.2%
	CAP7_기타	12	8	13	116	0.2%
AP기타 2개(1,1)	SAP_기타	12	13	13	53	0.1%
	CAP_기타	12	10	13	357	0.5%
합계					68,289	100%

자료: 저자 작성

- 점유율 1, 2위인 단순 활동패턴 SAP1, SAP2의 요일별 점유율을 살펴본 결과, SAP 1은 월요일에 가장 높고 이후 금요일까지 감소하는 반면 SAP2는 주말에 증가하는 것으로 나타남

* 휴일 전일 금요일 유연근무 등으로 직장에서 일찍 나와 기타 활동을 하는 것이 이유일 수도 있음

그림 6 | 활동패턴 유형별 비율 등



자료: 저자 작성

□ 활동패턴별 활동 시간 분석

- 활동패턴 32개의 주중과 주말의 활동 시간, 체류시간, 통행시간을 검토한 결과, 일 기준 주중 활동 시간(집 귀가 후 시간 제외)은 615분(약 10.3시간), 이 중 통행에 127분(약 2.1시간), 체류에 488분(약 8.2시간)을 소비하는 것으로 나타남
- SAP1(집-직장) 활동 시간 검토 결과, 주중의 활동 시간(집 출발-집 도착)이 주말보다 3.8시간(39.5%) 높게 나타남
- SAP1의 총활동 시간은 약 9.5시간, 통행시간 1.3시간(14%), 직장 등에서의 체류 시간은 8.2시간(86%)으로 체류시간 점유율이 통행시간보다 6.3배 높았음

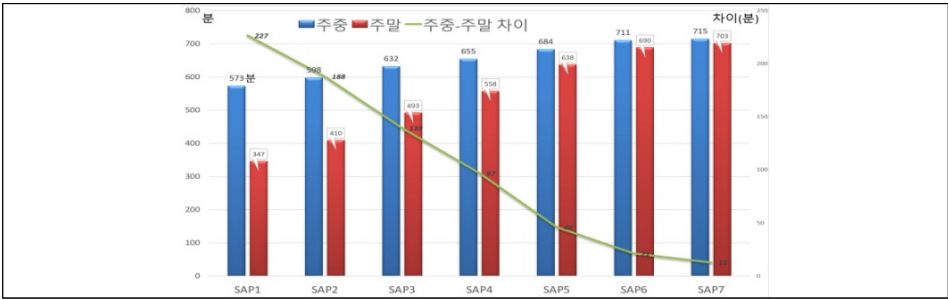
그림 7 | 주중과 주말 활동 시간(체류 및 이동) 비율



자료: 저자 작성

- SAP에서 목적지 횟수 증가에 따라 총활동 시간도 증가하는 것으로 나타나 목적지 증가로 추가적인 통행시간, 체류시간이 필요한 것으로 보여짐
- 총활동 시간의 증가추세는 목적지가 6회 이상인 SAP6 이후 둔화되었고 해당 시간 수렴치는 약 713분(11.9시간)으로 나타남
- SAP 유형 번호가 증가(목적지 증가)할수록 주중과 주말의 총활동 시간 차이는 줄어드는 것으로 나타남
 - SAP1의 주중, 주말 활동 시간 차이는 226분이었으나 SAP7의 주중과 주말의 차이는 12분으로 SAP의 총활동 시간의 수렴치는 주중과 주말 상관없이 유사한 것으로 나타남

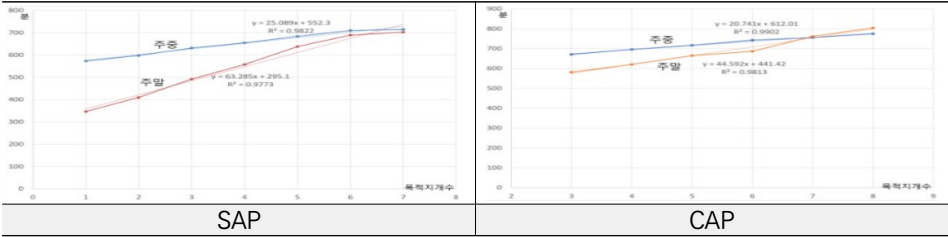
그림 8 | 주중과 주말 SAP별 총활동 시간과 차이



자료: 저자 작성

- 주중과 주말의 목적지 증가에 따른 활동 시간 변화를 살펴본 결과, SAP의 경우 주중 목적지 1회 증가 시 활동 시간이 약 25.1분, 주말 63.3분 증가해 주말의 증가치가 주중보다 높았음
- CAP의 경우 주중에는 목적지 1회 증가 시 활동 시간이 약 20.7분 증가하였고, 주말에는 44.6분 증가하는 것으로 나타남

그림 9 | 목적지 증가에 따른 SAP, CAP 활동 시간 증가율

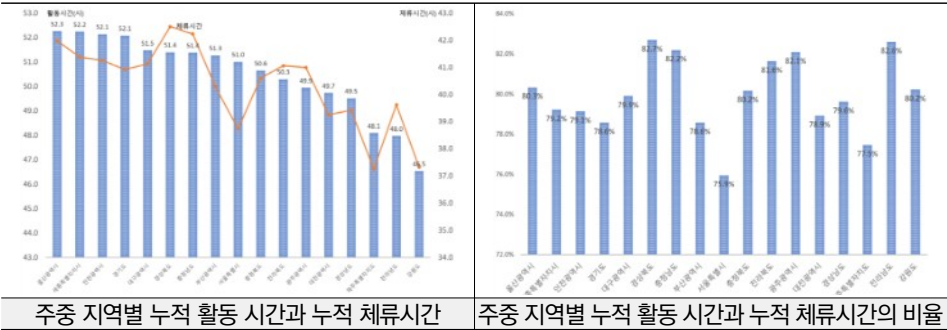


자료: 저자 작성

□ 지역별 활동 시간 분석

- 시도별 주중(월-금) 총활동 시간은 울산이 52.3시간으로 가장 높았고 강원도가 46.5시간으로 가장 낮았고 그 차이는 5.8시간 정도로 나타남
 - 주중 강원도 거주자는 울산 거주자보다 외부 활동이 5.8시간 적고, 반면 집 안에서의 내부 활동이 큰 것으로 추정됨
- 총활동 시간이 높은 시도는 울산, 세종, 인천, 경기도가 52시간, 대구, 경북, 충남, 부산, 서울 등이 51시간으로 높았음
- 활동 시간 중 체류시간이 차지하는 비율을 시도별로 살펴본 결과, 전국 평균은 약 80%로 나타남
 - 활동 시간 중 체류시간이 차지하는 비율은 서울이 75.9%로 가장 낮았고 이것은 통행에 소요하는 시간이 그만큼 타 시도보다 높다고 볼 수 있는 결과임
 - 경기도의 상기 비율은 78.6%, 인천 79.1%로 평균보다 낮았고 가장 높은 시도는 경북, 전남, 충남, 광주로 약 82% 정도였고 다음으로 높은 시도는 전북, 울산, 충북, 강원 등으로 나타남

그림 10 | 시도별 활동 시간 분석



자료: 저자 작성

- 요일별로 각 시도의 총활동 시간을 검토한 결과 수도권의 주중 활동 시간은 10.3시간, 도 지역은 9.9시간으로 나타남

5. 시사점 및 향후 과제

□ 시사점

- 본 연구는 스마트폰 앱의 개인의 위치/시간정보만으로 개인의 활동패턴을 파악하는 방법론을 노드-링크 분리 알고리즘, 시공간적 연속성 확보를 위한 데이터 가공방안 등을 제시하였다는 점에서 의의가 있음
- 가구 통행실태조사로 파악하였던 트립 체인 분석이 모빌리티 빅데이터로 이 가능하다는 것을 실제 분석을 통해 검증하였다는 점에서 의미가 있음
- 본 연구의 활동패턴 방안은 지역에서 개인의 하루 단위 활동 형상을 파악할 수 있어 개인의 지역 활동 모니터링 방안으로 활용될 수 있을 것임
- 활동패턴 분석 방법은 전염병 확산, 미세먼지 확산 등의 기후변화 등에 따른 경제활동을 간접적으로 모니터링 할 수 있는 방안이라고 사료됨
- 또한, 초광역권 설정, 광역 교통 대책 및 시설 설치를 위한 광역 교통권역 설정 등에 본 자료를 활용할 수 있을 것으로 보여짐

□ 향후 과제

- 본 연구에서 활용한 데이터(3주간)는 코로나로 인한 사회적 거리두기 시점이기 때문에 일상생활의 제약이 있었을 것으로 보이며 향후 더 많은 기간의 데이터로 활동패턴의 다양성을 분석할 필요가 있음
- 개인의 활동패턴은 개인의 속성(성별, 연령별)으로 그 차이점을 파악하지 못한 한계가 있어 추가적인 조사 등을 통한 개인 속성 정보구축이 필요함
- 기후변화, 사회현상 변화를 ICT로 파악하는 것이 우리에게 주어진 과제라고 할 때 다양한 빅데이터 가공 및 분석을 위한 노력은 지속적으로 필요함

차례

CONTENTS

주요 내용 및 정책 제언 iii

요 약 v

제1장 서론

1. 연구의 배경 및 목적 3

2. 연구의 범위 및 방법 6

3. 연구의 틀 9

4. 연구의 기대효과 11

5. 선행연구 검토 및 본 연구의 차별성 12

제2장 국내외 관련 연구

1. 개요 17

2. 모빌리티 빅데이터 18

3. 활동패턴 개념 21

4. 시사점 25

제3장 데이터 가공 및 활동패턴 대시보드 개발

1. 데이터 구축 및 검토 29

2. 활동패턴 구현을 위한 오류 데이터 필터링 방안 42

3. 집-기반 활동데이터 분석 방안	52
4. 활동패턴 대시보드 개발	66
5. 소결	76

제4장 활동패턴 유형의 구분 및 분석

1. 활동패턴의 구분 및 유형화 방안	81
2. 활동패턴 유형 분석	93
3. 활동패턴의 지역별 요일별 분석	107
4. 소결	128

제5장 시사점 및 결론

1. 시사점	133
2. 결론	142

참고문헌	145
------------	-----

SUMMARY	147
---------------	-----

표 차례

LIST OF TABLES

〈표 1-1〉 선행연구 요약 및 본 연구의 차별성	13
〈표 2-1〉 경우의 수를 고려한 이동유형 구분	19
〈표 2-2〉 정보기술의 발전에 따른 전산 환경의 변천	22
〈표 3-1〉 데이터 구축 과정	32
〈표 3-2〉 위드라이브 원시자료의 이용자 및 데이터 개수	36
〈표 3-3〉 데이터 필터링(시간기준)에 따른 데이터 개수 변화	51
〈표 3-4〉 출/도착지 위치 데이터 구축 과정	56
〈표 3-5〉 데이터 구축 과정	57
〈표 3-6〉 집 기반 분석 대안별 fs(첫 이동)와 le(마지막 이동)의 이격거리	60
〈표 3-7〉 집 기반 분석 대안별 이용자 수 도출 결과	61
〈표 3-8〉 중간 이동 포함 판단 논리구조	63
〈표 3-9〉 활동패턴 검토를 위한 대시보드 개발	68
〈표 3-10〉 활동패턴 대시보드 개발을 위한 입력 파일 가공작업	70
〈표 3-11〉 원시자료의 개인별 3일간 총활동패턴	71
〈표 3-12〉 지역별 원시데이터 표출	72
〈표 3-13〉 활동패턴 검토 사례 1: 거주지 위치, 이동 연속성	73
〈표 3-14〉 활동패턴 검토 사례 2: 거주지 위치, 이동 연속성 결측	74
〈표 4-1〉 체류시간 기준별 활동패턴 비율	83
〈표 4-2〉 링크(L)/노드(N)/경유지(S) 구분별 빈도수(예)	88
〈표 4-3〉 SAP 유형별 링크(L)/노드(N)/경유지(S) 간의 관계	90
〈표 4-4〉 CAP 유형별 링크(L)/노드(N)/경유지(S) 간의 관계	91
〈표 4-5〉 링크(L), 노드(N), 경유지(S) 기준 활동패턴 구분 예시	93
〈표 4-6〉 활동패턴 대구분 비율	96
〈표 4-7〉 상위 활동패턴 6개	99
〈표 4-8〉 활동패턴 32개 유형별 비중	100
〈표 4-9〉 AP1, SAP/CAP 활동패턴 개념도	101

〈표 4-10〉 AP2, SAP/CAP 활동패턴 개념도	102
〈표 4-11〉 AP3, SAP/CAP 활동패턴 개념도	103
〈표 4-12〉 AP4, SAP/CAP 활동패턴 개념도	104
〈표 4-13〉 AP5, SAP/CAP 활동패턴 개념도	105
〈표 4-14〉 AP6와 AP7, SAP/CAP 활동패턴 개념도	106

그림 차례

LIST OF FIGURES

〈그림 1-1〉 구글 지역사회 이동 추이 변화(2022년 5월 10일 한국 기준)	4
〈그림 1-2〉 활동패턴(예시)	5
〈그림 1-3〉 데이터 가공 및 시각화를 통한 연구 방향 설정	6
〈그림 1-4〉 모빌리티 빅데이터 가공	8
〈그림 1-5〉 활동패턴 대시보드 개발	8
〈그림 1-6〉 연구의 구성과 흐름	10
〈그림 2-1〉 서울특별시, 제천시 Multi-level 생활권 도출	18
〈그림 2-2〉 모빌리티 패턴(MP) 개념도	20
〈그림 2-3〉 시공간 경로상의 동적 활동 구분	21
〈그림 2-4〉 CDR 데이터 활용 일 단위 월 단위 활동공간 분포	23
〈그림 2-5〉 일상 활동 영역	24
〈그림 2-6〉 활동 궤적의 시공간 그룹	24
〈그림 3-1〉 위드라이브(WEDRIVE) 앱 구성	30
〈그림 3-2〉 위드라이브 제공 요일별 원시자료	30
〈그림 3-3〉 데이터 조인 및 공간조인 결과	33
〈그림 3-4〉 데이터 가공을 통한 속성정보 생성	34
〈그림 3-5〉 데이터 가공을 통한 속성정보 생성	35
〈그림 3-6〉 데이터 가공 결과	35
〈그림 3-7〉 국가교통 db의 1인당 통행발생량 변화	37
〈그림 3-8〉 서울특별시 중랑구 사례 : 000503 통행자	38
〈그림 3-9〉 경기도 오산시 사례 : 000011 통행자	39
〈그림 3-10〉 주거지 주변 배회형	40
〈그림 3-11〉 직장 주변 배회형	40
〈그림 3-12〉 서초구 고속버스 터미널(경부선) 개인 통행 시점 유형별 체류시간 분포 ...	41
〈그림 3-13〉 익일 자정 이후 도착 통행(예시)	43

〈그림 3-14〉 국내 최장 통행 거리 가정 예시	43
〈그림 3-15〉 통행시간 오류 1 : 총 통행시간 600분 이상 (예시)	44
〈그림 3-16〉 통행시간 오류 2 : 통행시간 음수	44
〈그림 3-17〉 통행시간 오류 3 : 통행시간(travel time) 음수	45
〈그림 3-18〉 체류시간 음수	47
〈그림 3-19〉 체류시간 1분 이하	48
〈그림 3-20〉 체류시간 기준 데이터 가공과정	49
〈그림 3-21〉 체류시간 기준 데이터 가공 매크로 작성	50
〈그림 3-22〉 국가교통 DB의 목적별 통행발생량(2016년)	52
〈그림 3-23〉 집 기반 활동 추출 대안별 알고리즘	54
〈그림 3-24〉 home based 활동 판단 논리구조	55
〈그림 3-25〉 집 기반 활동 판단 알고리즘	58
〈그림 3-26〉 중간 이동의 연속성 확보 구분 방안 (예시)	63
〈그림 3-27〉 노드(출/도착지) 포함 여부 논리구조	64
〈그림 3-28〉 활동패턴 미분류 데이터 복원	65
〈그림 3-29〉 대시보드 메뉴 구성항목	69
〈그림 3-30〉 활동패턴 검토 사례 3: 거주지 위치, 중간 이동 결측	75
〈그림 4-1〉 성인의 시간 사용	82
〈그림 4-2〉 체류시간 기준별 활동패턴 SAP, CAP 점유율 비율 차이	83
〈그림 4-3〉 단순 활동패턴(SAP) 개념도	84
〈그림 4-4〉 복합 활동패턴(CAP) 개념도	85
〈그림 4-5〉 활동패턴의 구분과 구성요소	86
〈그림 4-6〉 링크(L), 노드(N), 경유지(S) 기준 활동패턴 구분 예시	88
〈그림 4-7〉 링크(L)/노드(N)/경유지(S) 기준 활동패턴 유형(60개)	89
〈그림 4-8〉 단순/복합 활동패턴 비율 (단위: 건)	94

그림 차례

LIST OF FIGURES

〈그림 4-9〉 AP 유형별 AP/SAP/CAP 비율	97
〈그림 4-10〉 AP 유형별 SAP/CAP의 점유율 차이	97
〈그림 4-11〉 활동패턴 상위 10개 비율	98
〈그림 4-12〉 활동패턴별 주중과 주말 총활동 시간(분) 분포도	108
〈그림 4-13〉 활동패턴 SAP1, SAP2의 요일별 점유율	109
〈그림 4-14〉 활동패턴 SAP1, SAP2의 요일별 점유율	110
〈그림 4-15〉 전체 활동패턴의 가중 평균 주중 및 주말 체류 및 통행시간	111
〈그림 4-16〉 활동패턴별 주중과 주말 총활동 시간 분포도	112
〈그림 4-17〉 SAP1의 주중 및 주말 체류 및 통행시간	114
〈그림 4-18〉 SAP1의 주중 및 주말 총통행 거리 비교	114
〈그림 4-19〉 SAP2의 주중 및 주말 체류 및 통행시간	115
〈그림 4-20〉 SAP2의 주중 및 주말 총통행 거리 비교	116
〈그림 4-21〉 SAP2와 SAP1의 활동 시간 차이(SAP2-SAP1)	117
〈그림 4-22〉 주중과 주말 SAP별 총활동 시간과 차이	118
〈그림 4-23〉 SAP별(목적지 개수) 주중과 주말 총활동 시간	119
〈그림 4-24〉 주중과 주말 SAP별 체류시간과 통행시간 차이	120
〈그림 4-25〉 CAP2_1의 주중/주말 체류 및 통행시간	121
〈그림 4-26〉 주중과 주말 CAP별 총활동 시간과 차이	122
〈그림 4-27〉 CAP별(목적지 증가) 주중과 주말 총활동 시간 차이	123
〈그림 4-28〉 지역별 활동패턴 점유율 비교	124
〈그림 4-29〉 주중 지역별 누적 활동 시간과 누적 체류시간	125
〈그림 4-30〉 주중 지역별 누적 활동 시간과 누적 체류시간의 비율	126
〈그림 4-31〉 요일별 지역별 활동 시간	127
〈그림 5-1〉 마지막 통행 도착지의 귀가 여부 판단 알고리즘	136
〈그림 5-2〉 활동패턴 대시보드 개발	137
〈그림 5-3〉 활동패턴 대시보드 개발을 위한 데이터 구분 ID	138

〈그림 5-4〉 부산, 울산 지역 활동 연계성 검토 예시(3.14 월)	139
〈그림 5-5〉 코로나 전후 모빌리티 활동패턴 변화 (예시)	140
〈그림 5-6〉 세종, 대전시 거주자 활동패턴(SAP2) 예시	141

〈연구의 주요 용어 정리〉

구분	내용	시사점
연구 데이터 및 분석 도구	위드라이브	- 본 연구에서 활용한 모빌리티 앱 기반 빅데이터를 구축한 스타트업 기업명 (2020년부터 데이터 구축)
	알트릭스 (Alteryx)	- 빅데이터 분석 상용 프로그램으로 대용량 로딩, 연산 작업에 용이
	테블로 (Tableau)	- 활동패턴 대시보드 작성을 위해 활용한 상용 프로그램 - 활동패턴 대시보드 작성을 위한 입력자료는 알트릭스로 작성함
	워크플로우 (Workflow)	- 알트릭스에서 사용하는 용어로 분석 단위별로 생성되는 작업공간으로 알트릭스에서 제공하는 분석 툴을 연속적으로 배열한다고 해서 플로우라는 용어를 사용하는 것으로 사료됨
연구 방법	TGN (Trip Generation Numbering)	- 모빌리티 빅데이터의 하루 단위 통행의 순번을 부여하는 알고리즘 (김종학 외 2021)
	활동패턴 대시보드	- 활동패턴의 시공간적 연속성을 검토하기 위해 개발한 대시보드로 일자별, 지역별 활동패턴을 테블로의 대시보드 기능으로 구현
활동패턴	링크(L, Link)/노드(N, Node)/경유지(S, Stop over)	- 활동패턴 구현을 위한 3가지 변인으로 노드는 집을 포함한 목적지의 물리적 개수인 반면 경유지는 집을 포함한 방문지(목적지)를 시간순으로 정렬한 것(노드 : 집, 회사 / 경유지 : 집-회사-집) - 본 연구의 활동패턴 유형은 링크(L)/노드(N)/경유지(S)의 조합으로 결정
	AP(Activity Pattern) 활동패턴	- 활동패턴 유형 대구분을 의미하며 AP는 집 이외 목적지 개수와 일치함
	단순 활동패턴(SAP, Simple Activity Pattern)	- SAP는 활동패턴 중구분을 위한 용어로 순차적으로 목적지를 중복(집 제외) 없이 1일 1회만 출발하는 단순 활동패턴
	복합 활동패턴(CAP, Complex Activity Pattern)	- CAP는 2회 이상 방문하는 회귀성 활동패턴을 SAP와 구분하기 위해 제시한 활동패턴임

자료: 저자 작성



CHAPTER 1

서론

1. 연구의 배경 및 목적	3
2. 연구의 범위 및 방법	6
3. 연구의 틀	9
4. 연구의 기대효과	11
5. 선행연구 검토 및 본 연구의 차별성	12

01 서론

시공간에서 발생하는 인간의 활동은 위치로 표현되고 시간으로 그 순서를 파악할 수 있어야 한다. 개인이 휴대하는 스마트폰 GPS 기능으로 별도의 조사 없이 개인단위의 위치와 시간정보를 빅데이터 형태로 구축하고 있다. 해당 데이터로 활동패턴을 파악하기 위해서는 데이터 가공 및 분석이 필요하다. 본 연구는 스마트폰 앱을 통해 구축한 모빌리티 빅데이터로 개인 이동의 시간과 위치 정보로 하루 단위의 활동패턴 유형을 구분하고 그 특징을 분석하고자 하였다.

1. 연구의 배경 및 목적

1) 연구의 배경 및 필요성

스마트폰의 GPS 기능으로 개인의 시공간 정보가 별도의 조사 없이 빅데이터 형태로 구축 가능하게 되었다. 기존에는 가구 통행실태조사¹⁾ 또는 연구과제별로 별도의 설문 조사를 거쳐 제한된 기간과 제한된 인원의 이동현황을 조사로 파악할 수밖에 없었지만 2010년 초반부터 무선 통신사 기지국 위치 정보로 개인의 이동정보를 파악하기 시작하였고 지역별로 집계하여 사용하고 있다. 하지만, 개인단위로는 자료를 제공하고 있지 않으며 시간적 연속성 불명확한 한계가 있다. 하지만, 스마트폰의 운영체제를 개발 및 운영하는 구글(안드로이드), 애플(ios) 등은 자체 모바일 운영체제를 가지고 있어 개인 동의하에 위치 정보를 활용하고 있고, 카카오톡, 티맵 등의 개별 앱 등에서도 개인 위치 정보를 동의

1) 가구 통행실태조사: 국가교통 DB의 전국단위 여객 통행실태 조사를 위해 5년마다 샘플 조사하여 전수화

하에 수집할 수 있게 되었다.

코로나 사태 이후 구글²⁾은 자체 모빌리티 빅데이터로 전염병 확산 방지에 기여하고자 지역사회의 이동변화 추이 보고서를 국가별로 일일 리포트를 작성하고 있으며, 이 보고서는 일 단위 소매점과 공원 등의 여가 시설, 식료품점/약국, 공원, 대중교통 정거장, 직장, 주거지 등으로 위치를 구분하여 시간별, 지역별 이동 추이를 차트로 기술하고 있다. 예를 들어 [그림 1-1]과 같이 2022년 5월 10일 한국의 소매점 및 여가시설 방문객은 보고서 작성일 기준 6주간 평균치 보다 9%, 공원은 33% 증가했다고 기술하고 있다.

그림 1-1 | 구글 지역사회 이동 추이 변화(2022년 5월 10일 한국 기준)



주: 1) 구글은 한시적으로 코로나19로 인한 지역사회의 이동 추이 변화를 국가별로 분석하여 제시

2) 기준값은 보고서 작성일 6주 전 평균치

자료: 구글 코로나19 지역사회 이동성 보고서

https://www.gstatic.com/covid19/mobility/2022-05-10_KR_Mobility_Report_ko.pdf (p.1)

(2022년 5월 10일 접속)

모바일 폰 운영체계에 내장된 GPS 기능을 이용해 다양한 스타트업기업들이 LBS³⁾ 사업에 참여하고 있다. 국내에서도 2020년부터 스타트업기업(예: 위드라이브) 중심으로 개인 동의하에 개인 트래킹 데이터를 수집·가공하여 다양한 비즈니스 모델에 적용하고 있어 연구목적의 데이터 구득이 가능하게 되었다. 이는 기존 트립 체인, 모빌리티 패턴 연구가 시공간적 연속성에 대해 개념도 형태로는 제시하고 있지만 해당 패턴을 시각화

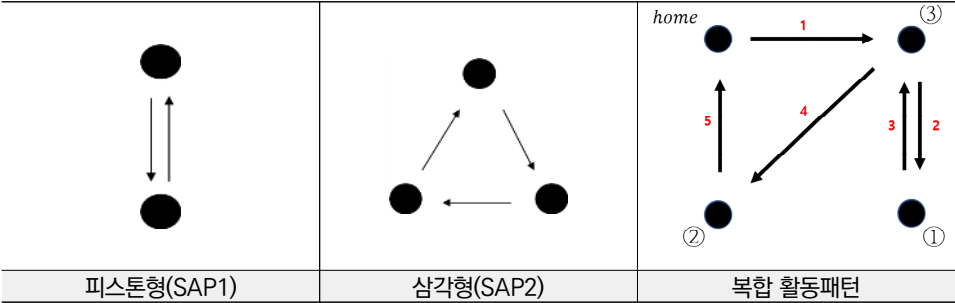
2) 구글 코로나19 지역사회 이동성 보고서

<https://www.google.com/covid19/mobility/index.html?hl=ko>

3) LBS: Location Based Service, GRAB, UBER, WEDRIVE 등

하지는 못하는 한계를 극복하는 계기가 되고 있다. 인간 활동은 물리적 공간에서 시간적 연속성을 가지고 발생하므로 발생 순서를 고려한 활동패턴 파악이 필요하고, 공간에서 이루어지는 인간의 활동은 위치(예: 경위도 좌표)로 표현되고 시간으로 활동 순서를 파악할 수 있어야 한다. 결국 모바일 폰으로 자동 수집되는 모빌리티 빅데이터로 활동패턴을 파악하기 위해서는 해당 데이터의 구축 방법과 함께 공간과 시간 연속성 여부를 판단할 수 있어야 한다. [그림 1-2]와 같은 활동패턴을 지도상에 표출하기 위해서는 도착지가 다음 통행의 출발지로 이어지는 것에 대한 시공간적 연속성을 지리 정보화하여 살펴볼 필요가 있으며, 본 연구에서 사용하는 스타트업기업의 모빌리티 빅데이터 구축 목적은 해당 기업의 비즈니스 모델로 데이터 마이닝을 통해 활동패턴 분석을 수행할 필요가 있다. 개인 이동의 시공간적 연속성을 파악해야만 구현되는 활동패턴을 개인-시간-공간 3개 정보가 통합(joint)된 모빌리티 빅데이터로 파악하려는 연구가 필요한 실정이다. 특히, 활동패턴 구성 요인으로 해당유형을 구분하는 기준이 없어 개인의 다양한 활동패턴을 파악하는 데 한계가 있었다. 활동패턴 구분기준이 없었던 가장 큰 이유는 활동패턴을 개념도로만 표현하고 실제 지도상에 프로그램화하여 활동패턴 구성 요인을 과학화하려는 노력은 적었다.

그림 1-2 | 활동패턴(예시)



주: 원 자료의 MP(Mobility pattern)를 SAP(Simple Activity Pattern), CAP(Complex Activity Pattern) 등으로 변경
 자료: 김종학 외(2021, 6) 일부 인용

2) 연구의 목적

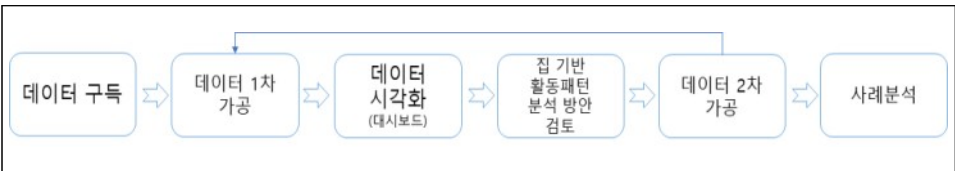
본 연구의 목적은 모빌리티 빅데이터 마이닝을 통해 활동패턴 유형 구분기준을 제시하고 활동패턴 구분 및 그 특징을 분석하는 데 있다. 본 연구는 모빌리티 빅데이터의 특징 및 한계 검토, 데이터 가공 및 분석을 통한 활동패턴 대시보드 개발, 출근/귀가 통행 분석, 활동 유형별 특징 분석(시도별, 일자별 등)을 통한 시사점을 제시하였다.

본 연구의 활동패턴 개념은 “하루 단위 개인의 집 기반 주요 활동지의 시공간 연속성을 노드(물리적 목적지), 링크(이동 방향/거리), 경유지(시간적 목적지)로 형상화한⁴⁾ 것”으로 정의하였으며, 이를 분석하기 위해 모빌리티 빅데이터의 특징과 한계를 선행적으로 파악 후 데이터 가공과 분석으로 상기 활동패턴의 형상화 작업을 위해 활동패턴 대시보드 개발을 통하여 구현하였다.

2. 연구의 범위 및 방법

본 연구는 모빌리티 앱 기반 빅데이터의 시공간 특성에 따른 데이터 가공과 시각화 과정의 피드백 등으로 활동패턴 유형을 구분하고 그 특징을 분석하였다.

그림 1-3 | 데이터 가공 및 시각화를 통한 연구 방향 설정



자료: 저자 작성

4) 형상(形象): 사물의 생긴 모양이나 형태

1) 연구의 범위 및 정의

(1) 공간적 범위

본 연구의 대상은 일별 모빌리티 빅데이터를 구축한 위드라이브 앱 데이터로 시도별 활동패턴을 분석하였다.⁵⁾

(2) 시간적 범위

연구의 시간적 범위는 활동패턴의 분석을 위해 2022년 3월부터 2022년 4월까지 3주간 데이터를 위드라이브 측으로부터 구독하여 활용하였다. 상기 기간은 연구 기간을 고려해 설정하였으며 해당 데이터는 일자, 시, 분, 초 데이터로 구성되어 있어 요일별(주중, 주말 포함), 시간대별 분석을 수행하였다.

(3) 내용적 범위

본 연구는 시공간 활동 개념, 모빌리티 빅데이터 활용 현황과 활동패턴을 다룬 국내외 연구를 검토하고, 빅데이터 1차 가공 형식 변경, 조인(join), 분리, 좌표체계 설정, 위치기반 활동 요인 파악 등을 위한 데이터 가공 알고리즘 등 시공간적 활동패턴 파악을 위한 빅데이터 기초가공을 진행하였다. 이러한 데이터 가공을 기반으로 활동패턴 대시보드 작성을 통한 모빌리티 빅데이터 시공간 특성을 파악하고자 하였으며, 이 과정에서 활동패턴 대시보드 메뉴 구성 및 디자인, 대시보드를 통한 데이터 특성 파악, 시공간 연속성 만족을 위한 데이터 가공 및 정제 수행, 집 기반 활동패턴 구현 기준 및 방안을 제시하였다.

5) 해당 데이터의 샘플 수 문제로 시, 도 이하의 공간 단위를 분석하는 데는 한계가 있었음

2) 연구 방법

(1) 국내외 활동패턴 분석연구 검토

트립 체인, 모빌리티 빅데이터 등 관련 선행연구를 검토하였으며, 특히 최근 모바일 빅데이터를 이용한 활동패턴을 다룬 연구들을 검토하였다.

(2) 모빌리티 빅데이터 가공 및 분석

출발·도착 포인트 및 이동 라인 생성, 집 기반 통행을 추출하기 위해 워크플로우로 작성하였고 이 결과를 대시보드의 입력자료로 활용하였다.

그림 1-4 | 모빌리티 빅데이터 가공



자료: 저자 작성

(3) 활동패턴 대시보드 개발

활동패턴 대시보드 개발을 위해 개발 대시보드에 대한 요구조건을 정의하였고, 대시보드 설계 및 입력자료 가공, 태블로를 활용해 대시보드를 개발하였다.

그림 1-5 | 활동패턴 대시보드 개발

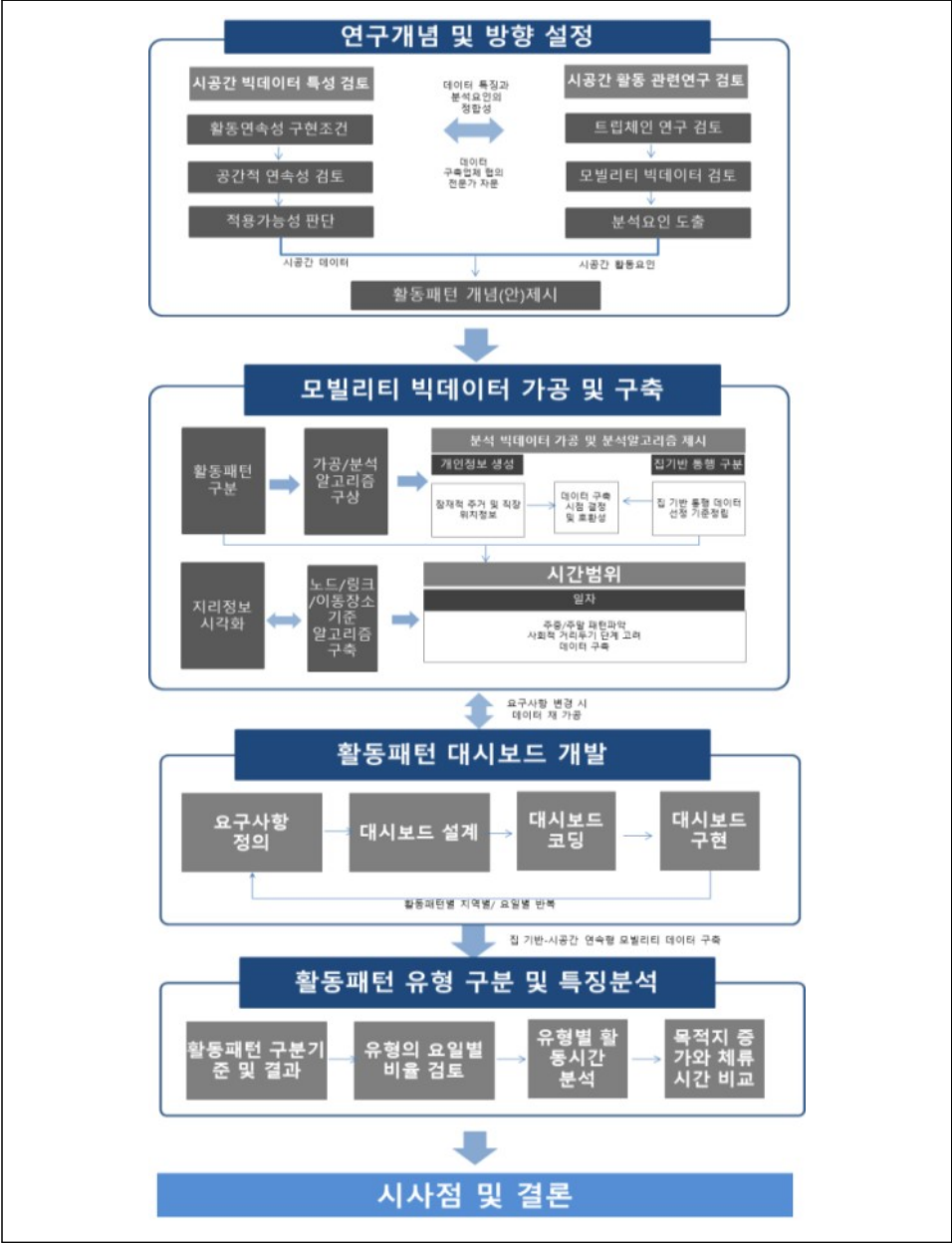


자료: 저자 작성

3. 연구의 틀

본 연구는 연구개념 정의 및 연구 방향 설정, 데이터 가공, 활동패턴 대시보드 개발, 사례 분석을 통한 정책제언 및 결론 순서로 진행하였다. 먼저, 본 연구의 개념과 방향을 설정하기 위해 모빌리티 빅데이터와 활동패턴과의 연관성에 대해 관련 문헌 검토와 데이터 시공간 특성 검토를 수행하였고, 빅데이터 구축 및 가공으로 활동패턴의 시간적 연속성과 공간적 연속성을 활용데이터로 설명 가능한지를 검토하였다. 그 중 특히 활동패턴은 일 단위 개인 활동의 시공간적 연속성 확보 가능성을 주로 검토하였으며, 주어진 데이터의 지리 정보화를 위한 출발, 도착 노드 생성, 링크 생성 등에 대한 알고리즘 개발 및 전산화(워크플로우) 작업을 함께 진행하였다. 그 이후에는 앞 단계의 모빌리티 빅데이터 구축 및 가공의 산출물을 활용한 활동패턴 대시보드 개발을 진행하였다. 모빌리티 빅데이터는 노드와 링크로 구성된 실질적인 활동패턴 시각화를 통해 데이터 시간적, 공간적 연속성을 파악할 수 있기 때문에 활동패턴의 시각화를 위해 활동패턴 대시보드를 개발하였고, 요구 사항 정의, 대시보드 메뉴 설계, 데이터 가공 등의 과정을 거쳐 알트릭스로 가공한 데이터를 기반으로 태블로에서 대시보드를 구현하였다. 마지막으로 앞서 개발한 활동패턴 대시보드 작업을 통한 활동패턴 유형을 구분하고 활동패턴 유형을 노드, 링크, 경유지 3개 변수의 조합 수로 나누어 각 유형을 지역별, 요일별로 그 특징을 분석하였으며 사례별 분석 결과에 기반한 학술적, 정책적 시사점을 도출하였다.

그림 1-6 | 연구의 구성과 흐름



자료: 저자 작성

4. 연구의 기대효과

본 연구는 모빌리티 빅데이터로 개인 활동패턴을 파악하기 위한 가공 알고리즘, 분석 방안을 선도적으로 제시하였다. 개인정보 보호법 등의 한계로 파악하기 어려운 집, 직장, 통행목적(출근, 귀가 등)을 데이터 가공 및 분석 등을 통해 파악하였다. 향후, 빅데이터 가공을 통한 개인정보 생성 방안은 향후 모빌리티 빅데이터 분석 방법 등에서 활용도가 높을 것으로 사료되며, 본 연구의 시공간적 연속성 확보를 위한 데이터 가공 알고리즘은 향후 자율주행 모빌리티 빅데이터 가공에도 활용이 가능하다. 특히, 본 연구는 모빌리티 빅데이터로 활동패턴 유형 구분기준을 LNS(노드, 링크, 경유지) 3가지 정보의 조합으로 제시하였다는 점에서 그 의미가 크다. 기존 활동패턴은 설문 조사로 진행되고 활동패턴 다양성을 파악하지 못했지만, 빅데이터는 활동 샘플 수 자체가 많아 다양한 활동패턴이 있음을 밝힐 수 있었고 그 다양성을 구분하는 기준을 제시할 수 있었다.

본 연구는 활동패턴 대시보드 개발을 통한 정책 활용도를 제고하고자 하였다. 본 연구의 태블로 기반 대시보드는 향후 추가적인 개발을 통해 지역 거주자들의 활동패턴을 지역 정책수립자들이 시각적으로 볼 수 있다는 장점이 있다. 또한, 개인 활동량, 활동패턴 변화를 시각적으로 파악할 수 있어 사회변화에 따른 국민 생활 영향 등을 파악하는 데 도움이 될 수 있다. 예를 들어 전염병 확산 전후 개인의 활동패턴 변화를 파악하거나, 신도시 초기 입주민들의 통행패턴 파악을 통한 임시 버스노선 발굴 등의 사전 광역교통개선대책 등에 활용할 수 있다.

5. 선행연구 검토 및 본 연구의 차별성

소셜 네트워크 데이터를 활용하여 개인 활동패턴을 분석한 구자용(2017)은 소셜 네트워크인 트위터 데이터로 사용자의 정보와 시간, 위치 정보를 추출하여 개인 활동패턴을 분석하였으며 시간에 따른 공간적 이동 경로를 지도화하여 활동 경로를 파악하고 개인 활동패턴을 시공간적으로 표현하여 도심 밀집도 파악 가능성을 제시하였다. 모빌리티 빅데이터를 이용한 개인의 모빌리티 패턴을 분석한 김종학 외(2021)는 연속형 데이터 분석 기법을 제시하여 모빌리티 패턴을 파악하였으며 코로나 전후 모빌리티 패턴 변화를 제시함으로써 동적 사회현상에 대한 동적 데이터 활용 가능성을 제시하였다. 미디어 이용의 시공간적 매핑에 대해 분석한 이재현(2014)은 크로노토프 개념과 시간 지리학을 원용하여 새로운 개념을 고안하고 시간과 공간을 통합적으로 접근할 수 있는 틀을 미디어 분야 활용방안으로 제시하였고 크로노토프 관점에서 미디어 이용을 지수화하는 복합 시간성과 같은 새로운 개념화를 통해 미디어 이용 자료 분석의 가능성을 제시하였다. 휴대폰 통화내역 기록 수집을 통해 월별 개인의 공간 경로와 활동패턴의 변동 및 요인을 분석한 Järv et al(2014)는 휴대 전화 기반 CDR 데이터 수집을 통해 개인의 월별/계절별 공간 행동 등을 분석하였다.

표 1-1 | 선행연구 요약 및 본 연구의 차별성

구 분		선행연구와의 차별성		
		연구목적	연구 방법	주요 연구내용
주요 선행 연구	1	• 과제명: 공간 빅데이터를 이용한 개인 활동패턴의 시공간적 탐색: 서울시 지역의 트윗 데이터를 사례로 • 연구자: 구자용(2017) • 연구목적: 소셜 네트워크 데이터를 활용한 개인 활동패턴 분석	• 기존 문헌 검토 • 빅데이터 수집 및 분석 • 데이터 마이닝 기법을 통한 공간 빅데이터 활용	• 알고리즘을 이용한 트윗 데이터의 수집 및 가공 • 시공간 지리학을 이용한 개인 활동패턴의 파악 및 분석
	2	• 과제명: 빅데이터 기반 도시 모빌리티 패턴 분석 방법론 연구 • 연구자: 김종학 외 (2021) • 연구목적: 모빌리티 빅데이터를 이용한 개인 모빌리티 패턴 분석	• 문헌 검토 • 빅데이터 가공 및 분석 • 알고리즘 제시 • 모빌리티 패턴 유형 분석	• 모빌리티 빅데이터 가공 및 연속형 데이터 분석방안 제시 • 코로나 전후 지역별 날짜별 모빌리티 패턴 변화 분석
	3	• 과제명: 미디어 이용의 시공간적 매핑 • 연구자: 이재현 (2014) • 연구목적: 크로노토프 개념과 시간 지리학을 원용하여 미디어 이용의 시공간적 분석 및 적용	• 통계조사 자료 이용 분석 • 미디어 이용 장소의 분류 • 크로노토프 접근 적용 및 개념화	• 시간 지리학 원용의 크로노토프 접근 이론화 및 신개념 고안 • 시공간적 미디어 이용 현상 분석 및 적용 가능성 시사
	4	• 과제명: Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records • 연구자: Järv et al(2014) • 연구목적: 월별 개인의 공간 경로와 활동패턴의 변동 및 요인 분석	• 휴대폰 통화 내역 기록 수집 및 표본조사 분류 • 통계자료 분석	• 휴대 전화 기반 CDR 데이터 활용 및 개인별 공간 행동 연구 • 개인 월별/계절별 공간 행동 평가 및 대인 간 변동 요인 분석
본 연구		• 과제명: 위치기반 빅데이터를 활용한 시공간 활동패턴 분석연구 • 활동패턴 구분기준 제시 및 특성 분석	• 기존 문헌 검토 • 모빌리티 빅데이터 가공 알고리즘 워크플로우 작성 • 활동패턴 대시보드 개발	• 모빌리티 빅데이터 가공 및 분석 • 출근 및 귀가 통행 분석 • 집 기반 활동패턴 구현방안 • 활동패턴 유형 구분 및 특징 분석

자료: 저자 작성



CHAPTER 2

국내외 관련 연구

1. 개요	17
2. 모빌리티 빅데이터	18
3. 활동패턴 개념	21
4. 시사점	25

02 국내외 관련 연구

국내외 관련 연구 동향은 본 연구 주제와 관련성이 높은 모빌리티 빅데이터, 활동 유형 개념 2개 주제를 선정하여 살펴보았다. 인간의 활동을 시간과 공간을 통합하여 살펴보려는 노력은 1970년 Hägerstrand부터 현재까지 지리학, 교통 등 다양한 분야에서 이루어지고 있었다. 특히, IT 기술 발달로 개인의 시공간 정보구축이 가능해지면서 더욱 활발해지고 있다.

1. 개요

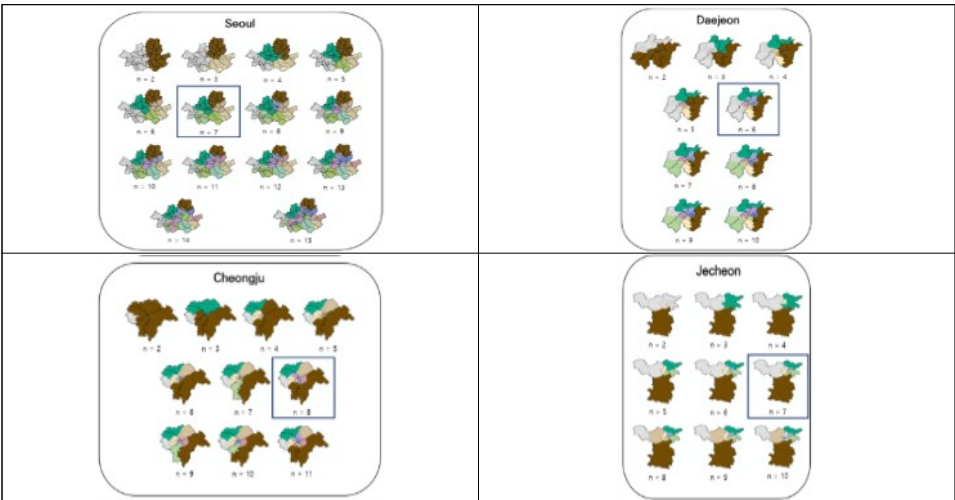
국내외 관련 연구 동향에서는 본 연구의 주제와 관련성이 높다고 판단되는 모빌리티 빅데이터와 활동 유형 개념으로 구분하여 관련 연구를 살펴보았다. 모빌리티 빅데이터 연구는 모바일 기지국 데이터의 읍면동 단위 기종점 통행량 자료로 생활권을 구분하는 연구(김규혁 외, 2021), 앱 기반 빅데이터로 모빌리티 패턴을 연구한 김종학 외 (2021)등의 연구를 살펴보았다. Shaw et al(2009)은 물리적 공간과 사이버 공간에서의 인간 활동을 분석하려는 노력도 있었다. 활동패턴 개념관련 연구는 인간, 시간, 공간이 합해진 삼간 데이터의 중요성을 강조한 김영표 외(2001) 연구와 에스토니아의 툴린 지역을 대상으로 12개월 CDR(Call Detail Records, 세부 통화기록) 자료로 계절별, 일자별 개인 활동량과 변화, 공간분포 등을 살펴본 Järv et al(2014) 연구를 살펴보았다. 또한, 시간 지리학 개념을 제안한 것으로 유명한 Hägerstrand(1970)의 ‘지역 학에서의 사람은 무엇인가’라는 연구를 살펴보았다. 그는 위치 중심의 연구에서 사람 중심의 연구가 필요하다고 주장하였다.

2. 모빌리티 빅데이터

1) 모바일 생활통행데이터 기반 도시 인구 규모별 생활권 분류 및 특성 파악¹⁾

모바일 기지국 중심의 위치 데이터로 도시별 생활권 구분을 시도한 김규혁 외(2021)의 연구를 살펴보았다. 해당 연구는 모바일 기지국 데이터 기반의 읍면동 간 기종점 통행량 자료로 Community Detection 기법으로 생활권을 구분하고자 하였다. 그는 서울특별시 2-15개, 대전광역시 2-10개, 중도시인 청주시의 경우 2-11개, 제천시의 경우 2-10개의 생활권으로 구분하였다. 이 연구는 생활권 구분을 읍면동 기준으로 하였다. 따라서, 대도시의 동 면적 대비 면 지역의 면적이 상대적으로 커 도시 규모가 작을 수록 생활권 개수가 적고 상세 구분이 어려웠을 것으로 사료된다. 향후 기지국 단위 위치 정보보다 더 상세한 위치 정보를 활용한다면 현실성이 높은 생활권 구분이 될 수 있을 것으로 보인다.

그림 2-1 | 서울특별시, 제천시 Multi-level 생활권 도출



자료: 김규혁 외(2021:671)

1) 김규혁 외(2021) 자료를 토대로 저자 작성

2) 빅데이터 기반 도시 모빌리티 패턴 분석 방법론 연구2)

김종학 외(2021)는 앱 기반 모빌리티 빅데이터로 모빌리티 패턴(MP, Mobility Pattern)을 파악하는 방법을 제시하였다. 해당 연구는 개인의 하루 단위 통행 발생 시점을 구분하기 위한 방법으로 TGN(Trip Generation Numbering) 알고리즘을 제시하였다. TGN 파악을 위해 현재 열의 개인 ID, 일자와 전·후열의 개인 ID, 일자의 일치 여부(○, ×)를 비교했다(김종학 외. 2021:67-71). 해당 연구의 TGN 알고리즘 구현을 위해 비교해야 할 경우의 수는 총 16개로 각 경우에서 판단할 조건문 내용은 [표 2-1]과 같다.

표 2-1 | 경우의 수를 고려한 이동유형 구분

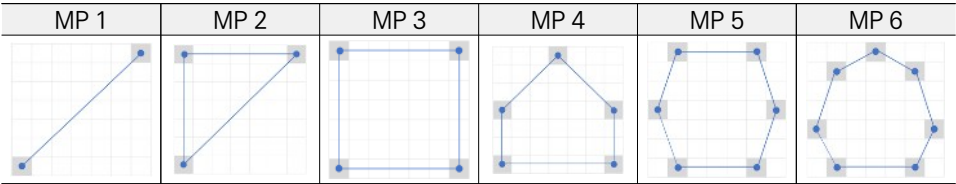
Case	ID(통행자) 2개		Date(날짜) 2개		통행 발생 시점
	전열(Row-1)	후열(Row+1)	전열(Date-1)	후열(Date+1)	
1	○	○	×	○	유형1: 첫 통행
2	×	○	×	×	
3	×	○	○	○	
4	○	○	○	○	유형2: 중간통행
5	○	○	○	×	유형3: 마지막 통행
6	○	×	○	×	
7	○	×	○	○	
8	×	○	○	×	단방향 통행
9	○	×	×	○	
10	○	×	×	×	
11	×	×	○	○	
12	×	×	×	○	
13	×	×	×	×	
14	×	○	×	×	
15	○	○	×	×	
16	×	×	○	×	

자료: 김종학 외 (2021:70)

2) 김종학 외(2021) 자료를 토대로 저자 작성

해당 연구는 집과 직장 등 집 이외의 목적지가 한곳인 모빌리티 패턴을 MP1(mobility pattern 1)으로 명명하고 추가 목적지가 증가할수록 모빌리티 패턴을 삼각형, 사각형, 오각형 등으로 그 패턴을 구분하였다.

그림 2-2 | 모빌리티 패턴(MP) 개념도



자료: 김종학 외(2021:요약 10)

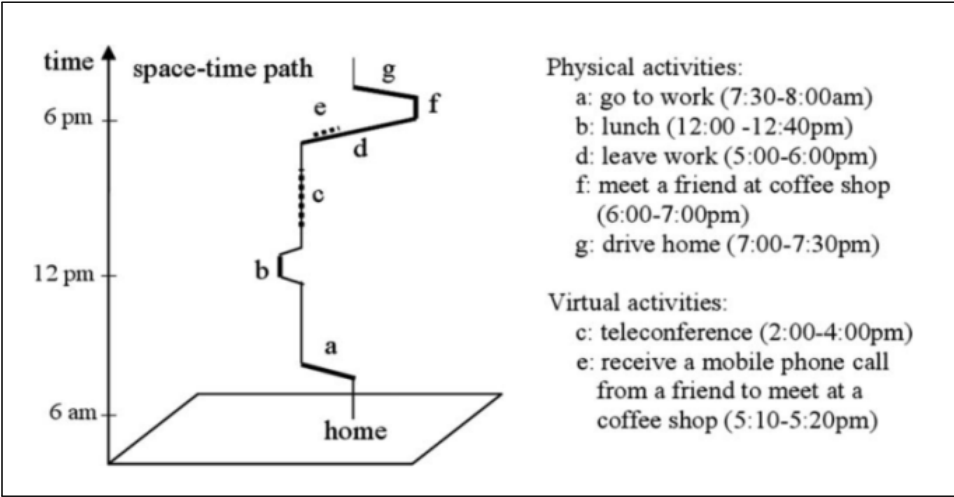
3) A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical-virtual space³⁾

Shaw et al(2009)는 ICT(정보통신기술) 발달로 시공간을 초월한 인간의 활동이 늘어남에 따라 물리적 공간과 함께 가상공간의 활동도 시공간적으로 파악하였다. 그는 시간 지리학(Time Geography)의 시공간 맥락(Space Time context)⁴⁾을 기반으로 물리적 활동과 가상적 활동을 파악하기 위해 [그림 2-3]과 같이 시공간 개인의 활동을 물리적 활동과 가상적 활동으로 구분하여 3차원으로 표현하였다. 그림에서 수직이동은 체류시간, 대각선 이동은 타 장소로의 이동으로 해당 기울기는 이동속도로 볼 수 있다. 이처럼 개인의 이동은 시공간적으로 연속성을 가지므로 빅데이터 기반의 모빌리티는 해당 데이터의 시공간적 연속성 파악이 중요한 것으로 사료된다.

3) Shaw et al(2009) 자료를 토대로 저자 작성

4) 원문은 공간과 시간이지만 본 연구의 국문 표현은 시공간으로 표기함

그림 2-3 | 시공간 경로상의 동적 활동 구분



자료: Shaw et al(2009:144)

3. 활동패턴 개념

1) 디지털시대에 대비한 사이버 국토 구축전략 연구⁵⁾

김영표 외(2001)는 우주라는 근원적 시스템이 그러하듯 모든 시스템은 시간 공간 인간이라는 세 요소가 두루 갖추어질 때 완전해진다고 주장하였다. 또한, 지리정보시스템(GIS, Geographic Information System)이 공간 위주로 자료를 관리·분석 처리하고 있어 활용성 제고를 위해 삼간(인간/시간/공간) 통합을 지향해야 한다고 하였다. 이러한 측면에서 본 연구에서 활용하고자 하는 모빌리티 빅데이터는 인간, 시간, 공간이 통합된 삼간 통합데이터라고 할 수도 있다. 해당 연구는 약 20년 전 연구이며 현시점에서 삼간 통합데이터는 구현 가능한 데이터가 되었다. 삼간 데이터의 통합은 가능하게 되었지만 모빌리티 패턴 분석과 같이 실제 활용을 위해서는 데이터의 가공 및 분

5) 김영표 외(2001) 자료를 토대로 저자 작성

석이 필요하다. 특히, 시공간 상에서의 개인의 활동은 다양하기 때문에 시간과 공간을 공유하는 개인들의 활동을 구분하기 위한 작업등이 필요하며 이러한 작업들은 지리정보시스템 이외의 타 분석 도구를 필요로 할 것이다.

표 2-2 | 정보기술의 발전에 따른 전산 환경의 변천

구분	비 공간적	공간적
정태적	[과거] 전산 자료처리의 중심	[현재] 전통적 GIS
동태적	[현재] 디지털경제의 중심	[미래] 삼간(三間) 통합

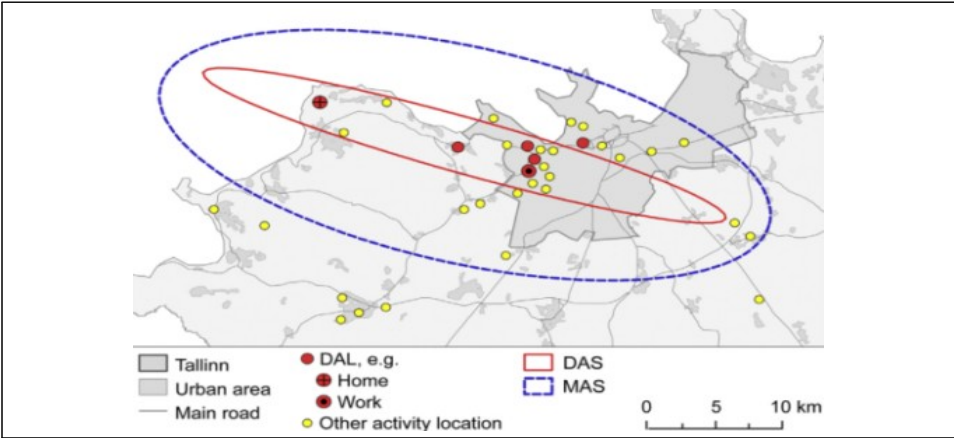
자료: 김영표 외(2001:21)

2) Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records⁶⁾

Järv et al(2014)은 에스토니아의 모바일 통신사인 EMT(Estonian Mobile Network)사의 12개월 CDR(Call Detail Records, 세부 통화기록) 자료로 개인의 ATB(Activity Travel Behaviour, 활동 이동행태)를 에스토니아의 툴린(tallinn) 지역을 대상으로 파악하였다. 그는 GIS의 SDE(Standard Deviation Ellipse, 표준타원체) 분석을 통해 하루 단위 활동공간과 월 단위 활동공간의 공간분포를 분석하여 월 단위 활동공간이 큰 것을 확인하였다. 이 연구는 집과 직장을 중심으로 개인의 활동 분포 변화를 살펴보았다. 개인의 활동패턴 파악을 위해서는 집과 직장의 위치 파악이 중요한 것으로 보인다.

6) Järv et al(2009) 자료를 토대로 저자 작성

그림 2-4 | CDR 데이터 활용 일 단위 월 단위 활동공간 분포



- 주: 1) MAL-monthly activity location,
2) DAL-daily activity location
3) NAL-new activity location
4) Mtop10-10 most frequency location
5) MAS-monthly activity space
6) DAS-daily activity space

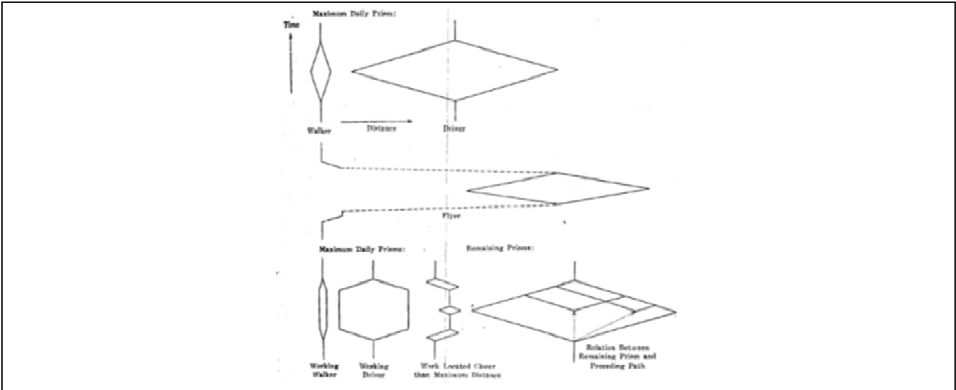
자료: Järv et al.(2014:129)

3) WHAT ABOUT PEOPLE IN REGIONAL SCIENCE⁷⁾

Hägerstrand(1970)는 상기 제목처럼 ‘지역학에서의 사람은 무엇인가?’라는 질문에 대한 답을 구하려고 노력하였고 그 노력의 결과로 시공간 개념을 제안했다고 사료된다. 지역학에서 위치에 중점을 두지 말고 사람에 중점을 두어야 한다고 주장하였다(Hägerstrand, 1970:7). 또한, 일 단위 개인의 활동 한계는 이용하는 교통수단별로 다르게 표현될 수 있다고 하면서 daily prisms(일상 활동 영역) 개념을 [그림 2-5]처럼 제시하였다. 그림에서는 도보 직장인과 운전자 직장인의 활동 영역은 차이를 설명하고 있다. 결국, 그는 이동 수단의 속도에 따라 개인의 활동 영역(prisms)의 시공간 폭은 확대/축소되는 개념을 제시했다고 볼 수 있다.

7) Hägerstrand(1970) 자료를 토대로 저자 작성

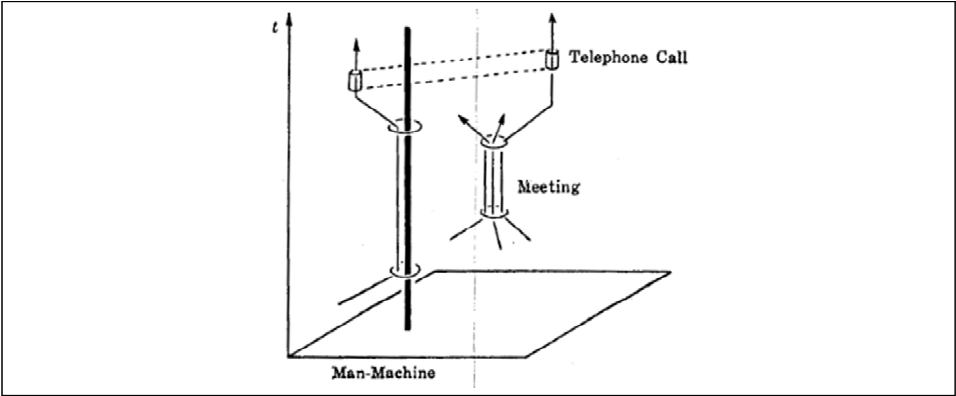
그림 2-5 | 일상 활동 영역



자료: Hägerstrand(1970:13)

Hägerstrand(1970)는 활동 궤적의 시공간 그루핑 개념을 제시하면서 전화 (machine) 등을 이용해 공간적 거리를 극복할 수 있고 이질적 공간의 개인 활동이 공유할 수 있다고 하였다. [그림 2-6]은 시간과 공간을 공유하는 개념으로 대면 회의 (meeting), 전화 통화 등을 제시한 것으로 시공간적 공유를 통한 개인 활동 개념을 선제적으로 제시한 것으로 볼 수 있는 개념도이다. 코로나 사태로 화상회의가 일상화 되고 메타버스 등 가상공간 이용이 구체화 되면서 그의 이러한 연구개념은 적용 범위가 더욱 넓어질 수 있다고 사료된다.

그림 2-6 | 활동 궤적의 시공간 그루핑



자료: Hägerstrand(1970:14)

4. 시사점

기존 연구 검토 결과, 본 연구 주제와 유사한 활동 형상, 특징 연구들은 교통, 지리학 등에서 다양하게 수행된 것으로 나타났고 이 연구들은 인간의 시공간 정보를 활용했다는 공통점이 있었다. 특히, 본 연구의 모빌리티 빅데이터는 김영표 외(2001)에서 언급한 삼간(三間, 인간/공간/시간)의 정보라고도 할 수 있다. 다만 활동패턴 분석을 위한 데이터 수집 체계가 과거에는 설문조사 등에 기반하여 수행하였다는 점이 본 연구와 다른 것으로 나타났으며, 관련 연구들은 설문 조사 등으로 연구를 수행하여 활동의 시간적 공간적 연속성 확보를 전제로 답변을 유도하였기 때문에 데이터상에서의 별도의 분석은 필요하지 않았다. 하지만 본 연구에서 수행해야 하는 활동의 시간적, 공간적 연속성 확보는 데이터 가공을 통해 확보해야 하는 문제점이 있었으나 관련 연구는 이에 대한 자세한 언급은 미흡하였다. 기지국 위치 정보로 활동패턴을 연구한 기존 연구에서는 기지국 간 핸드오버 현상에 따른 위치 오류, 기지국 단위의 궤적정보 등이 한계인 것으로 보인다. 반면 앱 기반 위치 정보는 국내에서 2020년부터 활용 가능하게 되어 비교적 최근에서야 연구되기 시작한 분야로 파악되었다. 또한 기존 연구 검토 결과 모빌리티 빅데이터로 활동패턴을 파악하기 위해서는 개인 이동의 시공간적 연속성을 확보하는 것이 중요하고 집-기반 활동 파악이 필요한 것으로 나타났다.

CHAPTER 3

데이터 가공 및 활동패턴 대시보드 개발

- 1. 데이터 구축 및 검토 29
- 2. 활동패턴 구현을 위한 오류 데이터 필터링 방안 42
- 3. 집-기반 활동데이터 분석 방안 52
- 4. 활동패턴 대시보드 개발 66
- 5. 소결 76

03 데이터 가공 및 활동패턴 대시보드 개발

모빌리티 빅데이터¹⁾ 가공을 통해 개인정보를 생성하고 활동과 시공간 연속성 검토를 위한 데이터 정제과정을 수행하였다. 또한, 활동패턴 자료의 공간적 형상 등을 검토하고 분석을 위한 활동패턴 대시보드를 개발하였다. 위드라이브로부터 구득한 데이터의 오류를 총 통행시간, 체류시간 점검, 중간 트립 점검 방안 등을 제시하여 활동패턴의 시공간적 연속성 확보, 신뢰도 높은 체류시간 등을 산정하고자 노력하였다. 또한, 집-기반 활동패턴을 구현하기 위해 최초출발지와 최종 도착지 간의 이격거리 기준 등을 설정하였다.

1. 데이터 구축 및 검토

1) 원시자료

(1) 데이터 소스

본 연구의 모빌리티 빅데이터는 위드라이브(WEDRIVE) 라는 스타트업기업의 데이터로 해당 기업은 2020년부터 운전자들의 이동정보 제공과 상응해 포인트(편의점, 커피숍 이용)를 주는 방식으로 데이터를 구축하고 있다. 위드라이브는 1.3만 명의 해당 앱 이용자를 대상으로 이동 경로를 일자별로 구축하고 있으며 스마트 시티, 광고, 정보 제공 등으로

1) 위드라이브(2020)

수익을 얻는 구조이며, 해당 앱 이용자가 생성한 하트는 포인트로 전환되고 제품을 구매하거나 서비스를 제공 받을 수 있다.

그림 3-1 | 위드라이브(WEDRIVE) 앱 구성



주: 위드라이브 (모바일 앱 스타트업기업)

자료: 위드라이브(2020)

(2) 활용데이터 시간적 범위

분석데이터는 2022년도 상반기 중 휴일 영향이 적은 2022.3.14~4.3 총 3주간(21일) 데이터로 위드라이브로부터 구축하였으며, 위드라이브 가입자의 3주간 자료는 약 162만 개다.

그림 3-2 | 위드라이브 제공 요일별 원시자료

22.3.14(월) (약 7.4만 개)										22.3.15(화) (약 7.9만 개)									
id	id	id	id	id	id	id	id	id	id	id	id	id	id	id	id	id	id	id	id
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140
141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160
161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180
181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200
201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220
221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240
241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260
261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280
281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300
301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320
321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340
341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360
361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380
381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400
401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420
421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440
441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460
461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480
481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500
501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520
521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540
541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560
561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580
581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600
601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620
621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640
641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660
661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680
681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700
701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720
721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740
741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760
761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780
781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800
801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820
821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840
841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860
861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880
881	882	883	884	885	886	887	888	889	890	891	892	893	894	895	896	897	898	899	900
901	902	903	904	905	906	907	908	909	910	911	912	913	914	915	916	917	918	919	920
921	922	923	924	925	926	927	928	929	930	931	932	933	934	935	936	937	938	939	940
941	942	943	944	945	946	947	948	949	950	951	952	953	954	955	956	957	958	959	960
961	962	963	964	965	966	967	968	969	970	971	972	973	974	975	976	977	978	979	980
981	982	983	984	985	986	987	988	989	990	991	992	993	994	995	996	997	998	999	1000

22.3.20(일) (약 5.5만 개) 취합 (약 162만 개)

자료: 저자 작성

(3) 위드라이브의 자료수집 알고리즘

활용데이터는 기본적으로 차량 이동을 기록하는 궤적 데이터로 궤적 데이터와 트래킹 데이터 2가지 형태로 제공하고 있다. 활용데이터는 트래킹 데이터로 도보 이동은 기록에서 제외하고 있으며, 도보 이용의 판단은 스마트폰 OS에 내장되어있는 health kit(아이폰), 건강 앱(안드로이드)에서 자체 판단하는 기능을 이용하고, 추가적 판단 기준으로 geo-fencing(반경 300m)을 이동할 경우 차량 주행으로 인식하고 이동 경로를 기록하고 있다. 예로 인근 주차장에서 도보로 이동할 때 목적지 건물 출입 시의 이동궤적은 기록하지 않고 있어 이러한 데이터의 한계를 고려할 필요가 있다. 주행데이터는 도로구간의 속도 파악은 가능하나 본 연구의 활동패턴 파악과는 상이해 본 연구는 출발지, 도착지 구분이 가능한 트래킹 데이터를 활용하였다. 또한, 지하 공간 이동 시에도 위치 정보가 오류가 발생하고 있어 이러한 경우 기록에서 제외하고 있다. 통행의 연속성이 일부 통행에서 비연속적으로 발생할 수 있기 때문에 이러한 데이터의 한계를 극복할 수 있는 방안 마련이 필요하다. 현재 데이터는 수치상으로만 제시하고 있어서 지리 정보화하여 목적지, 도착지를 노드로, 이동을 링크로 표현하는 활동데이터의 지리 정보화 작업이 필요하였다.

(4) 원시자료 정보

위드라이브의 자료는 출발지 위치와 시간, 도착지 위치와 시간(연·월·일·시·분·초)으로 구성되었으며, 위드라이브 이용자의 차량 정보는 제조사, 차종(모델별) 등 비교적 상세하게 기입된 정보로 이용자의 차종을 파악할 수 있다.

2) 데이터 취합 및 공간조인

제공된 데이터의 속성정보는 7개 필드로 구성되며 그 내용은 개인 ID (비 인식 정보), 출발 시간, 출발지 경도/위도, 도착시간, 도착지 경도, 위도 좌표이다. 데이터는 먼저 일자별 분석을 위해 제공된 일자별 트래킹 데이터를 단일 파일로 합치는 데이터 조인(data join) 과정과 출발지와 도착지에 대한 행정구역 위치를 파악하기 위한 공간조인 작업을 Alteryx의 워크플로우 작성을 통해 수행하였다. 그 중 data union은 필드 구조가 동일한 여러 파일(csv)들을 동일 (레코드) 형태로 열(record)의 개수를 증가해 한 개 파일로 합치는 작업이고, 공간조인에 대해서는 행정 경계(시군구 읍면동)의 위치 정보와 위드라이브 위치 정보의 일치 여부를 판단해 해당 행정구역 위치를 위드라이브 각 통행 출발지, 도착지에 입력하는 작업을 수행하였으며 약 324만 개의 위치 정보를 공간조인 하는데 약 30분 정도 소요되었다.

표 3-1 | 데이터 구축 과정

단계	내용	워크플로우
Step 1 : Data Join	일자별 데이터를 한 개 데이터 파일로 결합	
Step 2 : Spatial Join	개별 통행의 출발지와 도착지 경위도 좌표와 행정동 지리정보를 공간조인하여 출/도착지 행정 위치(시도, 시군구, 읍면동) 속성정보 생성	

자료: 저자 작성

그림 3-3 | 데이터 조인 및 공간조인 결과

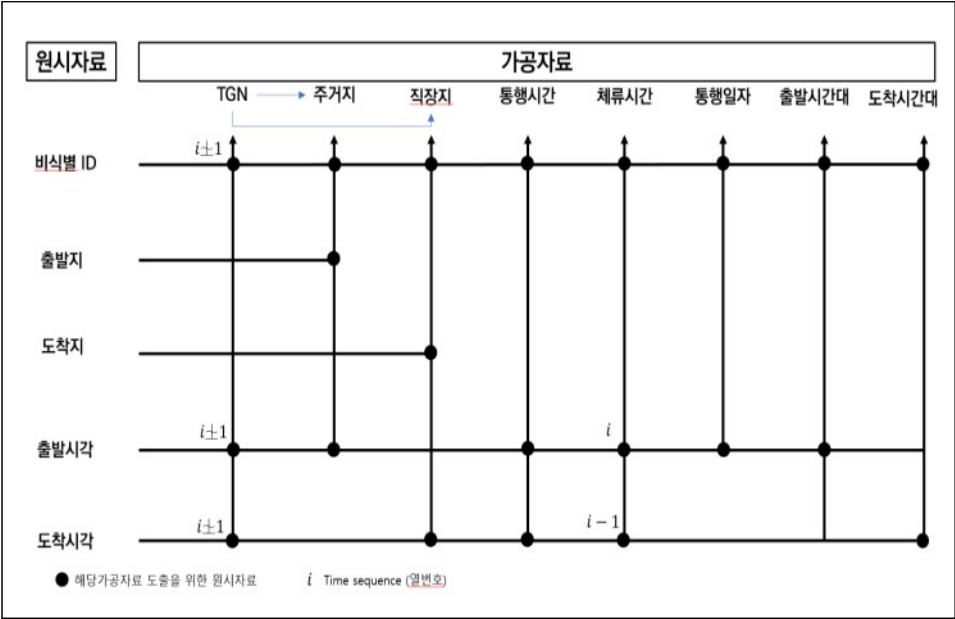
자료: 저자 작성

3) 추가속성 정보 생성

본 연구의 데이터 가공은 앞서 제시한 데이터 속성(7개 필드, 비식별 ID, 출도착 경위도, 시간)정보를 가공해 통행순서, 체류시간 등의 8개 속성정보를 생성하였다. 먼저, 김종학 외(2021)가 제안한 통행 발생 순서 도출 알고리즘인 TGN(Trip Generation Numbering)을 적용해 하루 단위의 통행 발생 순서를 파악하였다. 그 중, 통행시간은 특정 개인의 해당 일자 내 통행의 출발과 도착시간의 차로 원시자료에서 일자와 시간을 구분(split)하는 과정이 필요하다.

체류시간은 동일 ID의 동일 통행 중 전 통행의 목적지 도착시간과 현 위치의 출발시간 차이로 산정하였다. [그림 3-4]에서 검은 원은 해당 가공자료를 산정하기 위해 필요한 원시자료를 의미한다.

그림 3-4 | 데이터 가공을 통한 속성정보 생성



자료: 김종학 외(2021)를 바탕으로 저자 재작성

4) 모빌리티 빅데이터 원시자료 특성 검토

(1) 원시자료의 요일별 데이터 개수

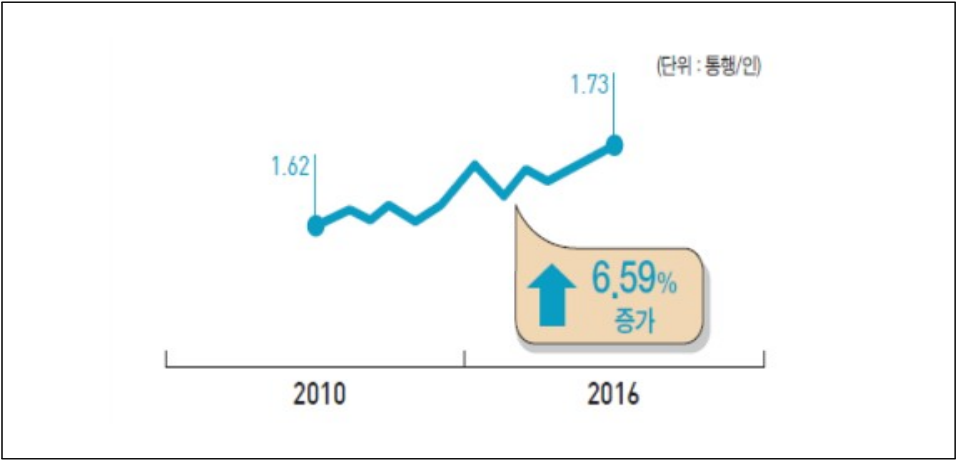
요일별 데이터(이동) 개수는 주중이 주말보다 16.1% 정도 많아 일반적인 주중과 주말 차이를 설명한다고 사료되나 그 절대량은 살펴볼 필요가 있어 제공된 자료(위드라이브 데이터)의 활용 가능성을 요일별 통행자 수, 통행 빈도수 분석을 통해 검토하였다. 결과적으로 [표 3-2]와 같이 가공 전 제공된 위드라이브 이용자는 주중 약 1만 5천 주말 약 1만 2천 명 정도로 나타났으며, 인당 통행 발생 수는 한국교통연구원의 정책자료집 (2018)의 2016년 1.73(통행/일)과 차이가 커 제공데이터의 상세 검토가 필요한 것으로 나타났다. 이에 원시자료의 통행자 수가 높은 것은 교차로 정지, 도로혼잡 등의 일시적인 정지상태도 목적지로 판단해 위드라이브 측에서 기록하였기 때문으로 판단된다.

표 3-2 | 위드라이브 원시자료의 이용자 및 데이터 개수

요일	이용자(인)	데이터 개수
월	14,537	77,123
화	14,905	81,155
수	14,942	80,887
목	14,873	80,320
금	15,230	84,233
토	13,210	67,673
일	11,786	56,162
주중 평균	14,897	80,743
주말 평균	12,498	61,918

주: 위의 데이터 날짜는 위드라이브에서 구축한 2022년 3/14~4/3 3주간 요일별 평균자료임
자료: 저자 작성

그림 3-7 | 국가교통 db의 1인당 통행발생량 변화

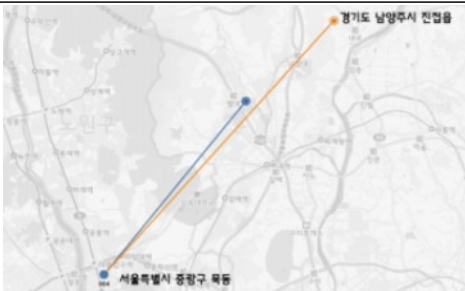
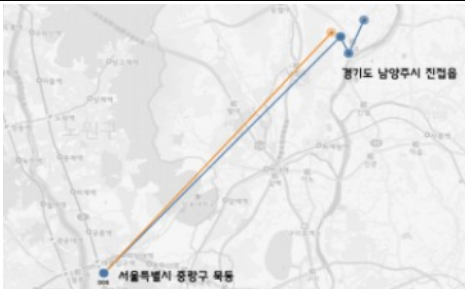
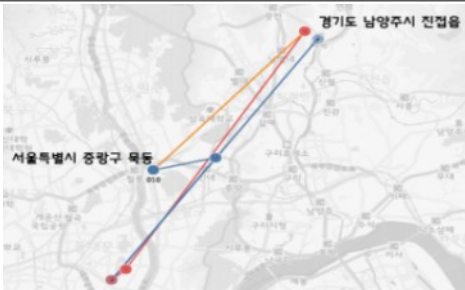


자료: 한국교통연구원(2018:16)

(2) 주거지와 직장 위치 추출 가능성 검토

위드라이브의 데이터로 주거지와 직장 추정을 위해 개인의 주중 첫 통행 사례를 지역별로 살펴보았다. 그 중 000503 가명 정보로 명명된 통행자는 2021년 3.8(월)~3.10(수) 통행 기록을 통해 집과 직장의 위치를 추정하면 집은 서울시 중랑구 묵동이고 직장은 약 24.5km 떨어진 남양주시 진접읍으로 추정 가능하다.

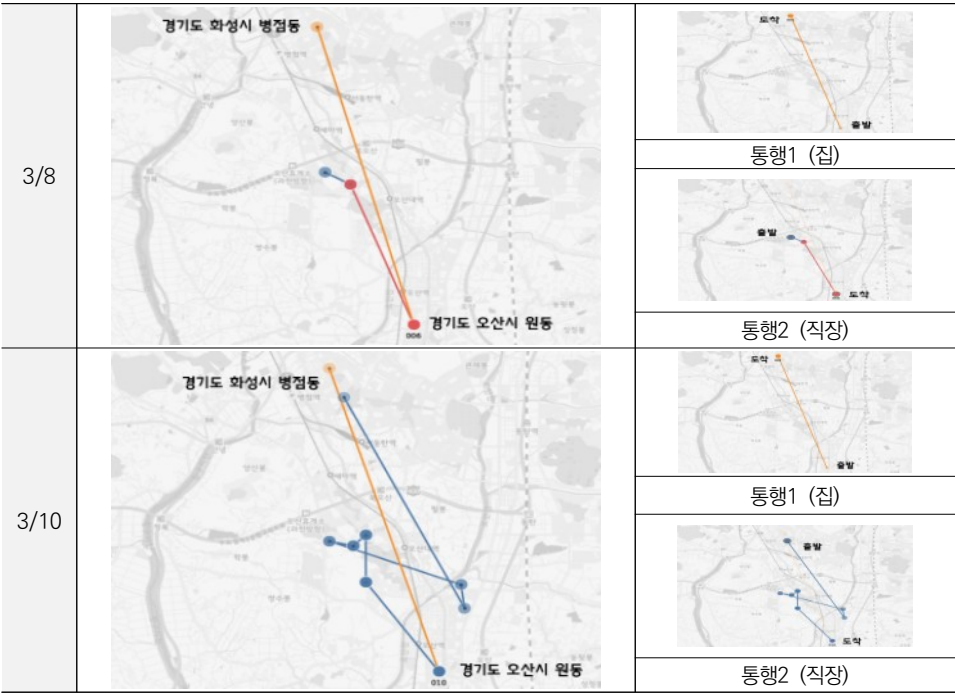
그림 3-8 | 서울특별시 중랑구 사례 : 000503 통행자

3/8		
		통행1 (집)
		
3/9		
		통행1 (집)
		
3/10		
		통행1 (집)
		
		통행2 (직장)

자료: 저자 작성

000011로 명명된 통행자는 2021년 3.8(월), 3.10(수) 통행 기록을 통해 집과 직장의 위치를 추정하면 집은 오산시 원동, 직장은 약 10.8km 떨어진 화성시 병점동으로 추정 가능하다.

그림 3-9 | 경기도 오산시 사례 : 000011 통행자



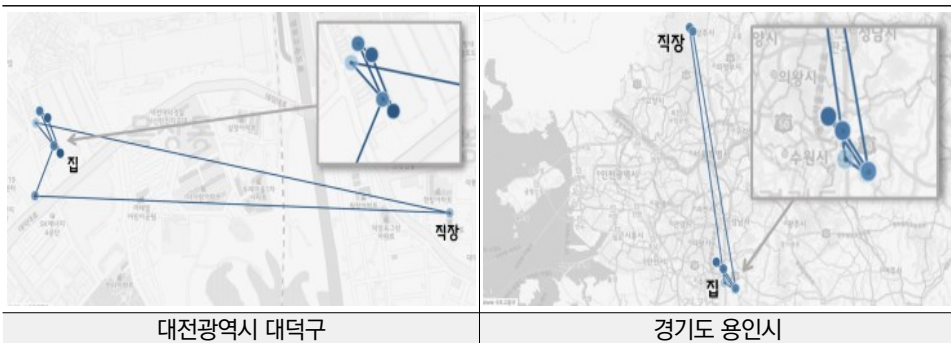
자료: 저자 작성

상기 기술한 바와 같이 주중 빈번히 오가는 하루 첫 통행의 기종점을 잠재적 주거지와 직장으로 가정하여 주거지와 직장 위치 데이터 구축이 가능하며 시공간적 연속성 측면에서 현재 통행의 출발지가 전 통행의 목적지가 아닌 공간적으로 비연속적인 데이터가 발생하는 것으로 나타나 이에 대한 필터링 기준이 필요한 것으로 보인다.

(3) 주요 활동지 주변 이동데이터 검토

원시자료 중 일부를 시각화하여 집, 직장을 중심으로 이동 형상을 검토한 결과 최종 목적지까지의 경로가 직선이 아닌 경우(유티, P턴) 집과 직장 등을 중심으로 단기간 이동하는 데이터가 많이 발생하는 것으로 나타났다. 이러한 이유는 GPS 데이터 오류 또는 원시자료의 궤적 데이터가 필터링되지 않고 트래킹 데이터로 기록되었기 때문으로 보인다. 따라서 이러한 자료는 체류시간 기준 설정 등을 통해 일원화하는 방안 마련이 필요하다.

그림 3-10 | 주거지 주변 배회형



자료: 저자 작성

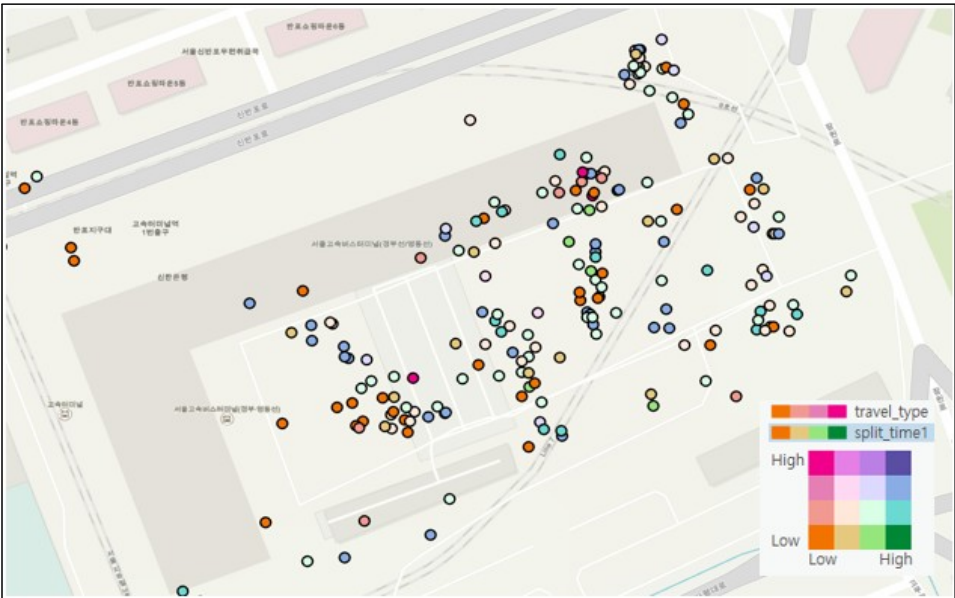
그림 3-11 | 직장 주변 배회형



자료: 저자 작성

해당 앱 데이터를 이용해 개인별 이동패턴(첫, 중간, 마지막 통행)별 도착 시간대 분포도 파악할 수 있었다. [그림 3-12]에서 Low/Low(오렌지색)의 의미는 하루 중 첫 통행이면서 오전 시간에 해당 터미널에 도착한 사람을 의미하며 High/High(보라색)는 하루 중 마지막 통행이면서 오후 늦은 시간에 도착한 사람을 의미한다. 이러한 위드 라이브 이용자의 도착 통행유형과 도착 시간대는 교차표(tabulated intersection) 분석을 통해 [그림 3-12]처럼 살펴볼 수 있다.

그림 3-12 | 서초구 고속버스 터미널(경부선) 개인 통행 시점 유형별 체류시간 분포



자료: 저자 작성

2. 활동패턴 구현을 위한 오류 데이터 필터링 방안

첫 통행(예: 출근)은 개인의 하루 중 첫 통행의 출발 시점만을 검토한 내용으로 통행 시간과 체류시간, 다음 통행의 연속성 판단 없이 분석할 수 있는 내용이다. 하지만, 활동의 연속성 검토로 파악할 수 있는 퇴근 활동, 활동패턴 유형 판단은 통행시간, 체류시간 등의 산정을 통해 분석의 기본가정이 되는 통행의 연속성을 판단할 필요가 있으며, 현재 통행 도착 시점이 다음 통행의 출발시간보다 늦거나, 국토 면적을 고려할 때 10시간 이상의 단일 통행자료는 제외하였다.

1) 통행시간 기준 오류 데이터 필터링

(1) 개요

분석데이터는 스마트폰 App(앱) 기반 데이터로 개인의 스마트폰 사용 방법, 통신오류, 데이터 수집 시 다양한 오류가 발생할 수 있다. 본 연구는 개인의 하루 단위 개별 트립을 분석하는 것으로 하루 이상의 과도한 통행시간을 분석에서 제외하는 것이 적정한지 살펴볼 필요가 있다.

위드라이브에서 제공 받은 원시자료 약 162만 개 중에서 출발일과 도착일이 다른 데이터는 약 2.2만 개로 나타났다. 이 데이터의 대부분은 22시 또는 23시에 출발해서 익일 00시 또는 01시에 도착하는 통행으로 나타났지만, 일부 데이터의 경우 16시에 출발하여 익일 10시에 도착하는 비정상적인 통행들도 포함되어 있어 분석에서 제외하였다.

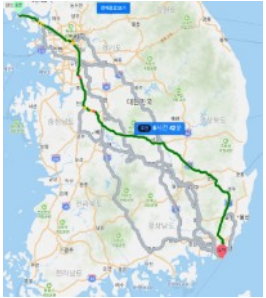
그림 3-13 | 익일 자정 이후 도착 통행(예시)

Record	time_begin	time_end	Record	time_begin	time_end
275	2022-03-17 21:12:40	2022-03-18 10:05:45	11942	2022-03-27 23:53:09	2022-03-28 00:06:59
276	2022-03-24 22:25:21	2022-03-25 00:09:05	11943	2022-03-30 23:50:15	2022-03-31 00:13:47
277	2022-03-26 22:56:04	2022-03-27 00:15:51	11944	2022-03-16 23:46:46	2022-03-17 02:53:28
278	2022-03-27 23:55:12	2022-03-28 07:32:46	11945	2022-03-19 16:28:57	2022-03-20 10:25:47
279	2022-04-03 23:15:59	2022-04-04 07:48:05	11946	2022-03-18 20:17:44	2022-03-19 14:46:48
280	2022-03-31 23:44:59	2022-04-01 00:15:59	11947	2022-03-26 19:03:18	2022-03-27 08:19:46
281	2022-03-28 21:36:00	2022-03-29 07:17:59	11948	2022-03-27 13:29:00	2022-03-28 06:31:56
282	2022-04-02 22:55:12	2022-04-03 00:14:56	11949	2022-03-28 22:03:06	2022-03-29 06:32:18
283	2022-03-24 17:48:42	2022-03-25 08:10:31	11950	2022-03-31 22:08:51	2022-04-01 06:01:12
284	2022-03-23 07:41:05	2022-03-24 07:33:47	11951	2022-03-20 23:56:22	2022-03-21 00:17:30
285	2022-04-02 22:36:02	2022-04-03 00:23:56	11952	2022-03-16 21:12:55	2022-03-17 00:40:49
286	2022-03-22 18:47:56	2022-03-23 07:41:50	11953	2022-03-22 21:02:31	2022-03-23 00:16:22
287	2022-04-02 21:14:02	2022-04-03 23:08:57	11954	2022-03-28 22:46:21	2022-03-29 01:01:51
288	2022-03-19 16:59:22	2022-03-20 10:07:17	11955	2022-03-18 14:22:39	2022-03-19 17:16:02
289	2022-03-26 16:38:00	2022-03-27 11:11:26	11956	2022-03-21 05:28:28	2022-03-22 13:04:20
290	2022-03-29 20:37:26	2022-03-30 07:14:06	11957	2022-03-26 23:42:12	2022-03-27 00:05:45
291	2022-03-29 23:25:19	2022-03-30 00:07:28	11958	2022-03-19 23:53:22	2022-03-20 00:12:25
292	2022-04-02 23:40:27	2022-04-03 00:02:24	11959	2022-03-15 22:52:54	2022-03-16 00:22:19

자료: 저자 작성

우리나라의 국토 형상은 직사각형으로 대각통행의 통행시간과 길이가 길어 최대 통행 시간 가정을 위해 강원 고성군-전남 진도군, 인천 강화-부산 영도 간 통행을 네이버의 길 찾기 기능으로 파악한 결과 북동-남서 대각통행은 692km, 8시간 50분, 북서-남동 대각통행은 501km, 6시간 42분으로 나타났다. 중간 휴식 시간을 고려한다면 국내의 1일 최대 통행시간은 약 10시간(600분)으로 설정해도 될 것으로 판단하였다.

그림 3-14 | 국내 최장 통행 거리 가정 예시

	
출발: 고성군 제진역, 도착: 진도군 진도항 통행시간 8:50분, 통행 거리 692km	출발: 강화군 교동도 지석초, 도착: 부산시 영도구 신선바위 통행시간 6:42분, 통행 거리 501km

자료: 네이버 길 찾기 화면 캡처

(2) 총 통행시간 기준 검증

분석데이터는 스마트폰 App(앱) 기반 데이터로 데이터 수집 시 다양한 오류가 발생할 수 있다. 일례로 [그림 3-15] ID가 32199인 통행자는 3월 29일 18시 출발, 3월 30일 7시 11분에 도착해 총 통행시간이 770분인 통행자로 출발지와 도착지는 경기도 구리시로 되어 있어 비정상적인 통행으로 나타났다. 이러한 통행시간 600분 이상인 통행은 약 1.1만 개로 나타났으며 익일 도착인 경우가 많았다.

그림 3-15 | 통행시간 오류 1 : 총 통행시간 600분 이상 (예시)

예1	RecordID		time_begin	time_end	travel_time_2
	32199		2022-03-29 18:20:41	2022-03-30 07:11:28	770
	33025		2022-03-17 23:28:17	2022-03-18 13:55:50	867
예2	Record	RecordID	time_begin	time_end	travel_time_2
	217	41826	2022-04-01 23:58:12	2022-04-02 17:58:11	1079
	218	41827	2022-04-01 23:59:37	2022-04-02 18:00:04	1080
600분 이상 데이터 개수	4 of 4 Fields ▾ Cell Viewer ▾ 10,707 records displayed, 259 KB ↑ ↓				
	Record	RecordID	time_begin	time_end	travel_time_2

자료: 저자 작성

(3) 통행시간 음수 수정

[그림 3-16] 예시는 레코드 271번 통행자로 4월 2일 23시 30분에 출발하여 익일인 4월 3일 19시 52분에 목적지에 도착하였다. 이 통행은 출발지와 도착지가 경기도 광주 시임에도 불구하고 통행시간은 약 1,200분으로 국토 면적을 고려할 때 통행시간이 커 통행데이터로 쓰기에는 부적절하다고 판단하였다.

그림 3-16 | 통행시간 오류 2 : 통행시간 음수

출발일시	Record		time begin			time end				
	271		2022-04-02 23:30:53			2022-04-03 19:52:53				
출도착 장소	Record	R u...	time...	loni	start_do_name	start_gu_name	start_dong_name	end_do_name	end_gu_name	end_dong_name
	271	2	04...	2022...	12	경기도	광주시	군지암읍	경기도	광주시

자료: 저자 작성

또한, 일상생활에서 친목 모임, 야근 등으로 익일 자정 이후 귀가 시 분석 프로그램에서는 통행시간이 음수로 나타나 수정이 필요하며, [그림 3-17]은 3월 28일 23시 56분에 출발하여 3월 29일 00시 12분에 도착한 경우로 기존 함수를 그대로 사용할 경우 통행시간이 -16분으로 산출되는 경우이다. 이 경우는 도출된 통행시간에 절댓값 함수를 사용하여 양수로 변환하는 작업이 필요하다.

그림 3-17 | 통행시간 오류 3 : 통행시간(travel time) 음수

출발 일시	Record		R	time_begin	time_end			
	1	1	2022-03-28 23:56:42	2022-03-29 00:12:48				
통행시간 (음수)	Record	start_do_name	start_gu_name	start_dong_name	end_do_name	end_gu_name	end_dong_name	trave time
	1	경상남도	진주시	초장동	경상남도	진주시	이현동	-16
▼								
전처리 (절댓값)	abs(DateTimeDiff([time_begin], [time_end], "minutes"))							
▼								
출발일시	Record		R	time_begin	time_end			
	1	1	2022-03-28 23:56:42	2022-03-29 00:12:48				
도착시간 (양수)	Record	start_do_name	start_gu_name	start_dong_name	end_do_name	end_gu_name	end_dong_name	trav time
	1	경상남도	진주시	초장동	경상남도	진주시	이현동	16

주: 1) alteryx의 tool 중 formula(포물러)의 시간 차이 함수(DateTimeDiff) 결과물을 절댓값(abs)으로 도출
 2) 통행시간(travel time)은 □참고
 자료: 저자 작성

2) 체류시간(stay time) 기준 검증

(1) 개요

체류시간은 현재 통행의 도착시간과 다음 통행의 출발시간과의 차이로 계산할 수 있다. 구득 원시데이터로 위 방법의 체류시간을 산정한 결과 연속적 통행이라고 보기 어려운 데이터가 발견되었다. 활동패턴의 분석 전제는 현재 통행이 다음 통행보다 시간적으로 앞서 발생했다는 전제에 있으며 이를 확보하기 위해 출발시간 기준으로 데이터를 정렬하여 사용하였으나, 데이터 검증 결과 출발시간으로는 정렬되어 있었지만 다음 통행의 출발시간이 현재 통행의 도착시간보다 빠른 경우가 있었다. 체류시간 오류 데이터는 ID 필드명이 동일한 것으로 미루어 볼 때 특정인에 한정되어 주로 발생하는 것으로 추정되며, 상기와 같은 체류시간 음수인 경우는 65명의 통행데이터 약 5천 5백 개로 나타났다. 음의 값을 가지는 체류시간은 트립 체인을 구성하는 데 있어 기본가정을 위배하는 자료로 분석상에서 제외해야 할 데이터이다. 특히 이러한 데이터는 하루 중 통행의 순번을 파악하는 데 있어 오류를 범할 수 있는 데이터이기 때문에 체류시간이 음수인 경우 음수가 나온 통행의 도착시간을 현재 통행의 도착으로 치환하는 작업이 필요하며 이러한 작업은 반복적으로 수행될 필요가 있다. 이러한 체류시간 오류 검증을 통해 치환되어야 할 데이터는 도착시간, 도착지 좌표, 도착지 등이다.

그림 3-18 | 체류시간 음수

예1	RecordID	uuid	time_begin	time_end	travel_time_2	stay_time
	41743	0708308b02124a59b17bf340847cfc29	2022-04-01 22:10:40	2022-04-02 16:11:07	1080	-1079
	41744	0708308b02124a59b17bf340847cfc29	2022-04-01 22:11:21	2022-04-02 16:11:20	1079	-1077
	41745	0708308b02124a59b17bf340847cfc29	2022-04-01 22:14:11	2022-04-02 16:14:38	1080	-1079
	41746	0708308b02124a59b17bf340847cfc29	2022-04-01 22:15:32	2022-04-02 16:15:30	1079	-1078
	41747	0708308b02124a59b17bf340847cfc29	2022-04-01 22:17:07	2022-04-02 16:17:34	1080	-1079
	41748	0708308b02124a59b17bf340847cfc29	2022-04-01 22:17:55	2022-04-02 16:17:54	1079	-1076
	41749	0708308b02124a59b17bf340847cfc29	2022-04-01 22:21:02	2022-04-02 16:21:00	1079	-1079
	41750	0708308b02124a59b17bf340847cfc29	2022-04-01 22:21:48	2022-04-02 16:22:15	1080	-1079
	41751	0708308b02124a59b17bf340847cfc29	2022-04-01 22:22:32	2022-04-02 16:22:30	1079	-1077
	41752	0708308b02124a59b17bf340847cfc29	2022-04-01 22:24:44	2022-04-02 16:25:11	1080	-1079
	41753	0708308b02124a59b17bf340847cfc29	2022-04-01 22:25:54	2022-04-02 16:26:18	1080	-1079
	41754	0708308b02124a59b17bf340847cfc29	2022-04-01 22:26:19	2022-04-02 16:26:38	1080	-1079
	RecordID	uuid	time_begin	time_end	travel_time_2	stay_time
예2	48711	081c7908ea214de99e161e0a78083124	2022-03-29 19:06:43	2022-03-30 13:07:36	1080	-1080
	48712	081c7908ea214de99e161e0a78083124	2022-03-29 19:07:23	2022-03-30 13:08:11	1080	-1080
	48713	081c7908ea214de99e161e0a78083124	2022-03-29 19:07:48	2022-03-30 13:08:42	1080	-1080
	48714	081c7908ea214de99e161e0a78083124	2022-03-29 19:08:22	2022-03-30 13:09:16	1080	-1080
	48715	081c7908ea214de99e161e0a78083124	2022-03-29 19:08:53	2022-03-30 13:09:47	1080	-1079
	48716	081c7908ea214de99e161e0a78083124	2022-03-29 19:09:53	2022-03-30 13:10:22	1080	-1079
	48717	081c7908ea214de99e161e0a78083124	2022-03-29 19:10:33	2022-03-30 13:11:27	1080	-1080
	48718	081c7908ea214de99e161e0a78083124	2022-03-29 19:11:11	2022-03-30 13:11:57	1080	-1080
	48719	081c7908ea214de99e161e0a78083124	2022-03-29 19:11:38	2022-03-30 13:12:32	1080	-1080
	48720	081c7908ea214de99e161e0a78083124	2022-03-29 19:12:09	2022-03-30 13:13:02	1080	-1079
	48721	081c7908ea214de99e161e0a78083124	2022-03-29 19:13:14	2022-03-30 13:14:08	1080	-1080
	2 of 2 Fields Cell Viewer 65 records displayed, 3790 bytes					
	Record	uuid	Count			
체류시간 음수 데이터	1	7fff704147f24a859e8fc4e5af1d8f26	1073			
	2	58d0ba29b1c44803a182577919691592	485			
	3	c3c29636a7704a728a35ecc7caba4611	419			
	4	e272e27f5ec44d2c870c0a3de6fc0e6e	418			
	5	081c7908ea214de99e161e0a78083124	395			
	6	33c083771a754ba78737f9e3f7c31495	392			
	7	25616bdad4d54120a9449b09ff71dc92	371			
	8	8766d437c80a4581b1247f0fcc09482f	335			
	9	16aaf48b39ee431fb09a9204d17033f6	332			
	10	54de2610038049f688f47f4238dbe77	245			
	11	dfa9ef0416da409a894e66d6d6ec5112	229			
	12	0708308b02124a59b17bf340847cfc29	140			
	13	1137ba277a954ebd9721b6da8f4e4536	120			
	14	ddfc5083589246f2991f40c7ae419018	111			
	15	78ffe9a749ef4e5ca86be07f1a714a9d	98			
	16	984f2c77d5de4ecda00db2ec4b320bf8	73			
	17	9c0f5b44fb7042a79decdd86cc27bccdd6	57			
	18	81ee825eff8a40ed9cebdd53ef78ce5e	49			
	19	bb6efe9d5f714e31a201f6fd7a9d07bf	39			
	20	9d57c2cc6d2c426ca54ecd01cf5e7ef0	37			

주: 상기 uuid는 개인 고유 ID이며 count는 해당 통행자의 체류시간 음수 데이터 개수임
 자료: 저자 작성

(2) 데이터 치환을 통한 데이터 정제

체류시간은 현재 통행의 도착시간과 다음 통행의 출발시간과의 차이로 계산할 수 있다. 이 방법으로 체류시간을 산정한 결과 목적지에서 특정 활동을 수행하였다고 보기 어려운 데이터들이 있었다. [그림 3-19]는 목적지에서의 체류시간이 1분 미만인 예로 목적지에 도착했다기보다는 교차로 등에서의 정차로 볼 수 있는 자료이다. [그림 3-19] 예2의 레코드 1239의 도착시간은 13:52:55이고 다음 통행인 레코드 1240의 출발시간은 13:52:57로 2초간 시차로 활동 장소에 도착했다고 보기 어려운 자료이며, 이러한 데이터가 발생하는 원인은 위드라이브 자료가 차량 궤적 데이터를 활용하기 때문으로 사료된다.

그림 3-19 | 체류시간 1분 이하

예1	Record	RecordID	uuid	time_begin	time_end	stay_time
	54	731	00090245f88441d4962e625c3954568c	2022-04-02 07:30:12	2022-04-02 07:40:10	0
	55	732	00090245f88441d4962e625c3954568c	2022-04-02 07:40:10	2022-04-02 09:00:00	0
예2	Record	RecordID	uuid	time_begin	time_end	stay_time
	88	1239	001ed0a4a9bc46e0add449d0ee86e91a	2022-03-22 09:37:37	2022-03-22 13:52:55	0
	89	1240	001ed0a4a9bc46e0add449d0ee86e91a	2022-03-22 13:52:57	2022-03-22 19:47:57	0

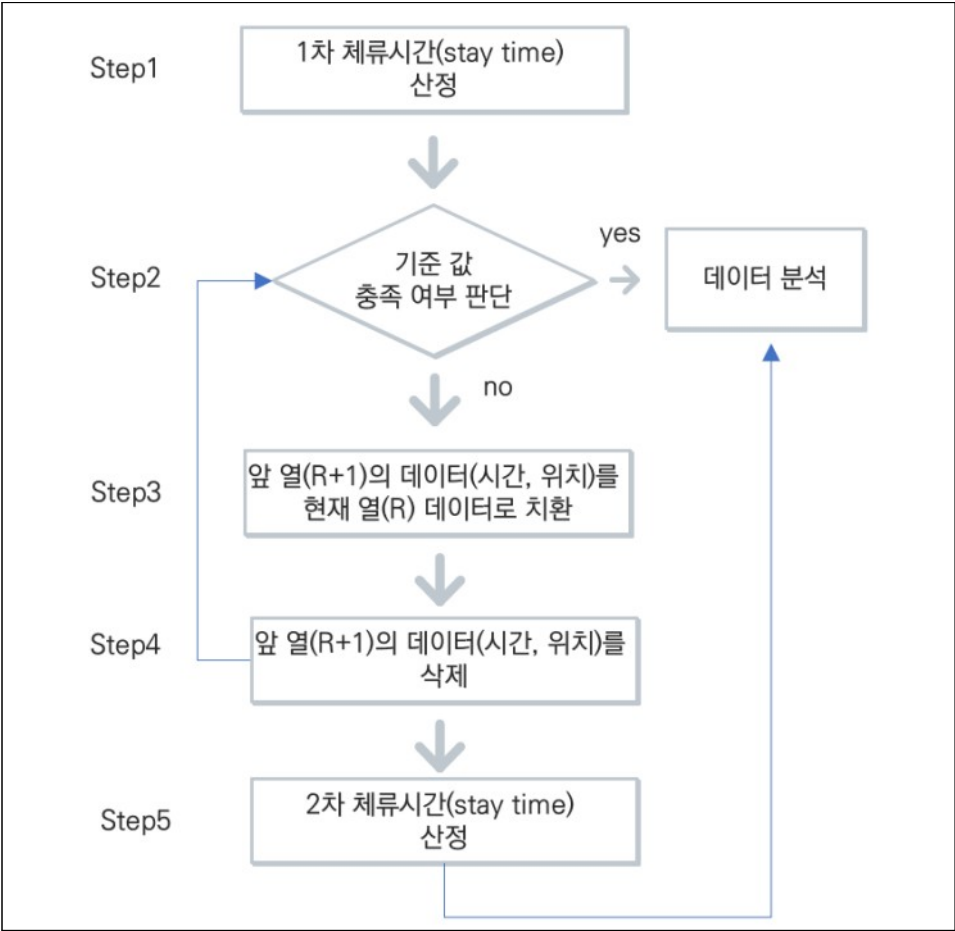
주: 상기 uuid는 개인 고유 ID임

자료: 저자 작성

여기서 주의해야 할 것은 체류시간이 특정 값(예: 1분) 이하일 경우 해당 데이터를 무조건 삭제하는 것은 통행의 연속성 확보를 위해 적절치 않다. 왜냐하면 다음 통행의 목적지 도착시간, 도착 위치 정보 등이 현재 통행의 목적지 정보로 치환되어야 하기 때문이다.

따라서 체류시간 데이터 가공과정은 5단계로 진행하였으며 체류시간 기준값이 충족 될 때까지 데이터 가공은 반복되는 작업을 거쳤다. [그림 3-20]에서 반복 작업은 Step2-Step3-Step4 과정이다.

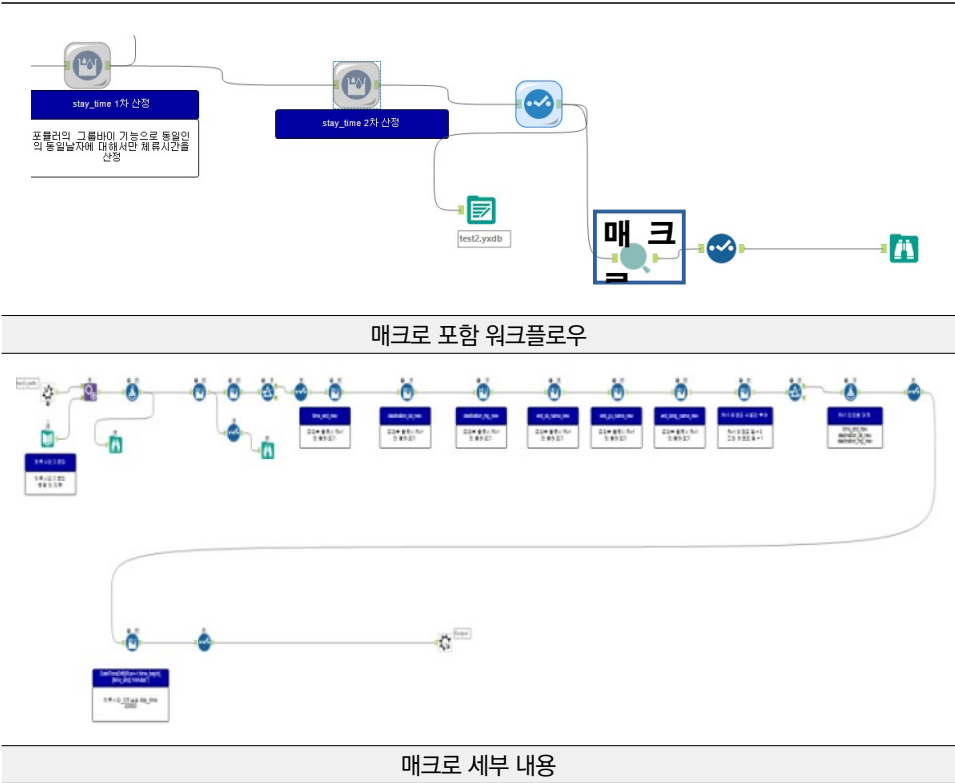
그림 3-20 | 체류시간 기준 데이터 가공과정



자료: 저자 작성

상기 플로우 차트의 내용을 실제로 구현하기 위해 Alteryx 프로그램을 활용하여 워크플로우와 기준값 변경을 용이하게 반영하기 위해 매크로를 작성하였다. [그림 3-21]의 돋보기 모양 툴이 알트릭스 프로그램의 매크로 툴이고 2열의 내용이 매크로 세부 내용이며, 체류시간 기준값은 데이터 가공단계별로 적용되기 때문에 일괄된 값의 변경을 용이하게 하기 위해 매크로 실행 전 입력값으로 설정하였다. 매크로 내부에서는 체류시간 기준값을 미충족하는 연속데이터를 판단하는 조건문과 미충족 데이터의 도착시간, 위치 정보 등을 치환하는 작업 등을 수행한다.

그림 3-21 | 체류시간 기준 데이터 가공 매크로 작성



자료: 저자 작성

3) 데이터 정제 단계별 데이터 개수 변화

데이터 가공은 시계열적으로 통행의 순서가 맞는지 검토하기 위해 통행시간과 체류시간 변수를 이용하여 검토하였다. 본 가공에서는 시간기준(통행시간, 체류시간)만을 고려하였고 [표 3-3]과 같이 3단계 과정을 거쳐 데이터를 가공하였다.

표 3-3 | 데이터 필터링(시간기준)에 따른 데이터 개수 변화

단계	데이터 가공		데이터(통행) 개수 (단위: 건)		앱 이용자 수 (단위: 인)
			가공단계별 데이터	제외 데이터	
0	원본 자료		1,619,709	-	335,496
1	통행시간 600분 제외		1,609,902	▼ 9,807	334,896
2	체류시간 음수 제외		1,605,942	▼ 3,960	334,125
3	체류시간 기준	5분 이하 제외	1,355,183	▼ 250,759	334,740
		10분	1,232,738	▼ 122,445	334,731
		20분	1,082,552	▼ 150,186	334,722
		30분	984,420	▼ 98,132	334,708

주: 1) 데이터 개수는 각 가공단계 기준에 부합한 위드라이브 이용자의 3주간 총통행 수

2) 이용자 수는 각 가공단계에서 3주간 위드라이브 이용자 수를 의미하며 일자별 이용자는 중복됨

자료: 저자 작성

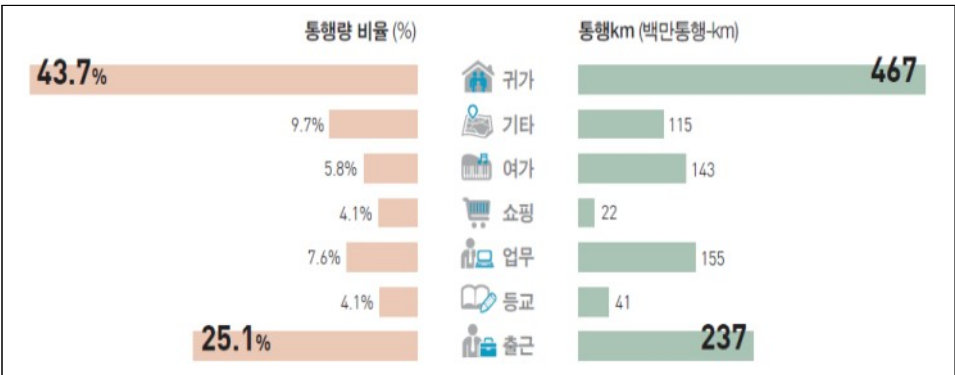
본 연구 분석 원시자료(3주간)의 총 데이터 개수는 약 162만 개로, 시간기준 가공 1단계 통행시간 600분 이상 데이터는 분석에서 제외하였다. 가공 2단계에서 체류시간 음수 데이터를 필터링 한 결과 데이터 약 4천 개 정도가 제외되는 것으로 나타났다. 3단계 체류시간 기준값으로 데이터 개수 감소분을 살펴본 결과, 체류시간 기준 5분에서는 약 25만 개, 10분은 약 12만 개, 20분은 약 15만 개, 30분은 약 10만 개 데이터 필터링이 필요하였다. 결국, 3단계를 통해 과다 통행시간과 비활동 체류시간 등으로 약 26만~64만 개가량의 비활동 데이터를 제외할 수 있었다. 가공단계별 3주간 위드라이브 총 이용자 수의 변화를 살펴보면 데이터 개수 감소율보다 낮았고 이것은 일자별 개인 전체 데이터를 필터링 한 것이 아니라 주어진 조건에 부합한 데이터만을 핀셋 필터링한 결과라고 볼 수 있는 결과로, 이러한 데이터 가공은 활동패턴의 시간적 연속성과 비논리적 활동을 선제적으로 제외시킴으로써 활동패턴의 신뢰성을 제고시킬 수 있을 것으로 사료된다.

3. 집-기반 활동데이터 분석 방안

1) 분석개요

한국교통연구원(2018)에 따르면 전체 목적통행 중 귀가 통행과 출근 통행이 차지하는 비율은 68.8%로 10 통행 중 약 7 통행으로 나타났으며, 해당 자료는 35만 가구를 대상으로 한 2016년(제4차 조사) 가구 통행실태 조사를 통해 작성된 자료로 5년마다 샘플링 자료로 작성하고 있다.

그림 3-22 | 국가교통 DB의 목적별 통행발생량(2016년)



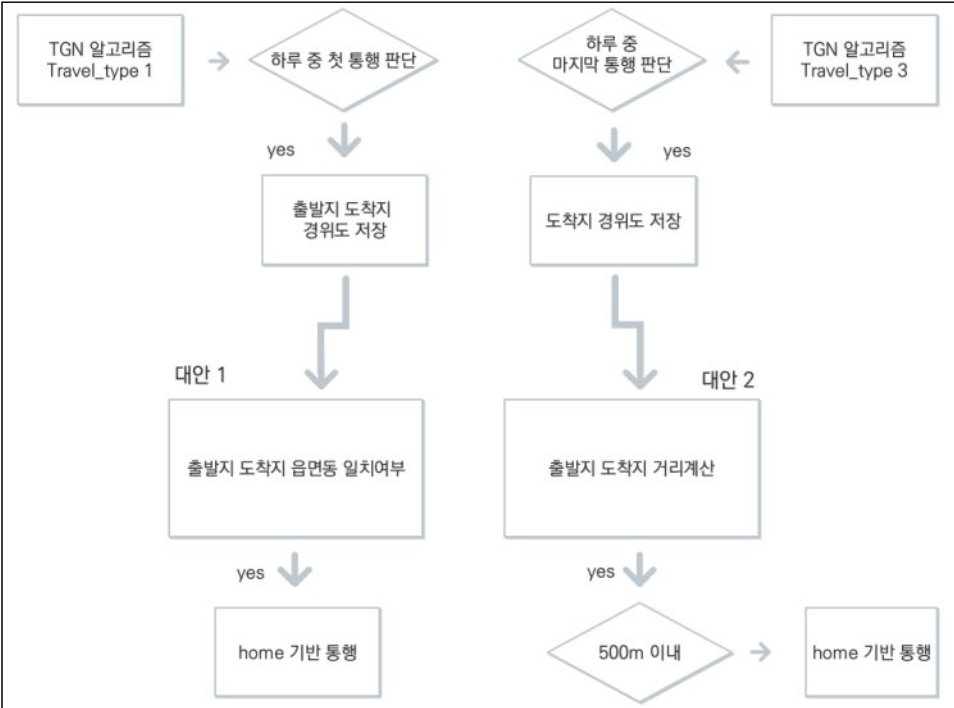
자료: 한국교통연구원 (2018:18)

위드라이브의 일일 궤적 데이터로 활동패턴을 파악하는 데 있어서 가장 먼저 작업해야 할 사항이 집 기반(home-based) 활동 여부의 판단이다. 활동의 시작과 끝이 집(home)에서 이루어지는 등 개인 일상의 중심은 집에 기반하기 때문이다. 출장/여행 등으로 당일 귀가하지 못한 것으로 파악되는 데이터는 예외적인 데이터로 간주하여 본 분석에서 제외하였다. 가구 통행실태 설문 조사 등의 조사에서는 주중과 주말의 대표적인 통행을 개인의 기억에 기반한 집 기반 통행을 조사하고 있어 별도의 집 기반 통행을 파악하려는 노력이 불필요하다. 이러한 개별 설문 조사로 가정 기반 통행을 파악할 경우 집에서 출발하여 집으로 귀가하는 통행을 전제로 조사하기 때문에 집 기반 통행이 전제로 되어 있다.

하지만, 모빌리티 빅데이터는 개인의 이동기록을 모바일 폰의 GPS 기능으로 수집하고 있어 집 기반 통행을 구분하기 위해서는 별도의 가공작업이 필요하다. 이러한 경우 먼저 개인이 하루 중 첫 통행을 확인한 후에 첫 통행의 출발지와 하루 중 마지막 통행의 목적지와 유사 정도를 판단할 필요가 있다. 여기서, 해당 판단은 분석데이터의 한계와 특징을 고려해야 하며 본 연구의 위드라이브 데이터는 도보 통행은 기록하지 않고 있어 도보 이동 시의 위치 정보는 저장되어 있지 않다. 또한 본 연구가 활용하는 데이터 구축이 활동패턴 분석 목적이 아니므로 연구 진행은 모빌리티 빅데이터 데이터 특징을 고려하여 수행이 필요하다.

집 기반 활동 여부 판단은 김종학 외(2021:67)의 TGN(Trip Generation Numbering) 알고리즘을 활용하였다. 이러한 집 기반 활동의 기본가정은 하루 중 첫 통행의 출발지와 마지막 통행의 목적지가 유사해야 한다는 가정에 있으며, 유사 여부를 판단하기 위해 2가지 대안을 가지고 분석하였다. 대안 1은 하루 중 첫 통행(fs: first start)과 하루 중 마지막 통행(le: last end)의 읍면동이 일치할 경우 집 기반 통행으로 판단하는 방안이며, 하루 중 첫 통행 출발지와 하루 중 마지막 통행 도착지 간 읍면동 일치 여부는 앞서 공간조인에서 파악한 출발지와 도착지 일치 여부로 판단한다. 대안 2는 하루 중 첫 통행(fs: first start)과 하루 중 마지막 통행(le: last end)의 이격거리를 기준으로 산정하는 방안이며, 대안 1에 비해 대안 2는 여러 가지 작업 과정이 필요하며 먼저 출발지와 목적지 간 포인트를 생성하고 생성된 포인트의 위·경도 좌표를 이용하여 거리를 계산하는 과정과 이격거리 기준값(예: 500m)에 대한 설정이 필요하다.

그림 3-23 | 집 기반 활동 추출 대안별 알고리즘

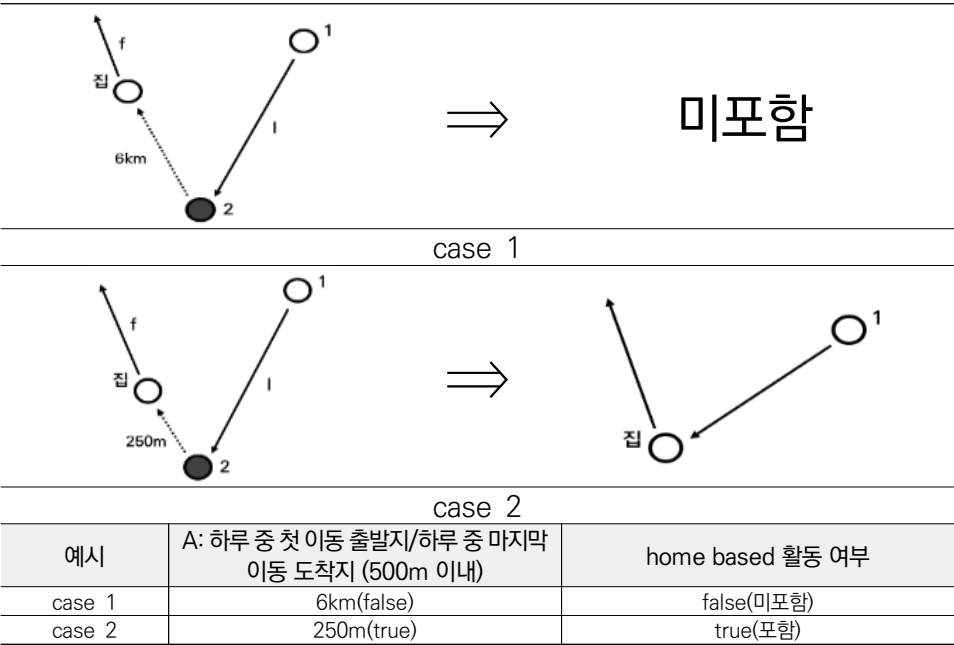


자료: 저자 작성

집 기반 통행 판단을 위해 하루 중 첫 이동의 출발지와 마지막 이동의 도착지점 간 거리는 500m로 가정하였다. 도보 이동 개인은 위드라이브 데이터에서 수집하지 않고 있어 해당 데이터의 도착지와 목적지 간 이격거리가 도보권(500m) 이내일 경우 하루 중 마지막 이동의 도착지는 집 주변의 단순경유지(주차장, 동네 마트 등)로 볼 수 있다고 가정하였다. 도보권, 역세권 등에 대한 거리는 본 연구에서 다루기는 한계가 있어 상식적으로 무리가 없다고 사료되는 500m를 고려하였다. 하루 중 첫 이동 출발지(집)와 하루 중 마지막 이동 도착지가 500m 이상일 경우 해당 이동을 집 기반 통행이라고 판단하지 않은 이유는 특정 거리 기준 없이 집 기반 통행을 판단하기 어렵고, 하루 중 마지막으로 이동한 데이터로 이후 데이터가 없어 체류시간에 따른 중간통행 여부 판단이 어렵기 때문이다. 마지막 통행의 도착지가 동일 시군, 읍면동에 있을 경우에는 예외적으로 해당 이동을 마지막이 아닌 중간 이동으로 간주할 수 있는 방안이 있으나 이 경우 집 귀가 출발시간과 도착시간

등을 추론하기 어려운 문제점이 있어 본 연구에서는 집 기반 통행만을 분석 대상으로 하기 위해 귀가 통행 여부 판단기준을 엄격히 적용하였다. 이에 대해 [그림 3-24]를 예로 설명하면 case 1은 하루 중 마지막 이동의 목적지가 집에서 6km 이상 떨어진 곳에 도착한 통행으로 이후 이동기록이 없는 자료이며, 이러한 자료는 해당일 이후 이동기록이 없어 체류시간 귀가 여부 판단이 어려워 집 기반 통행 분석데이터에서 제외하였다. 하루 중 중간 이동의 경우는 해당 이동 이후 기록으로 연속 이동 여부를 추론할 수 있으나 case 1과 같은 데이터는 추론이 어려운 점이 있어 분석에서 제외하였고, case 2의 경우는 마지막 이동 도착지와 집의 위치가 500m 이내로 도보로 귀가했다고 추론이 가능해 분석데이터에 포함하였다.

그림 3-24 | home based 활동 판단 논리구조



주: 1) f: first, 첫 이동 / l: last, 마지막 이동
 2) 1 하루 중 마지막 이동의 출발지, 2 하루 중 마지막 이동의 도착지
 자료: 저자 작성

2) 출발·도착 분리를 통한 노드, 링크, 경유지(Link, Node, Stop over) 산정

연속 이동의 출발지와 목적지 간의 관계는 현재 도착지가 이후 이동의 출발지와 유사해야 한다는 조건을 만족해야 한다. 또한 출발지와 도착지 간의 연계성 판단을 위해 위·경도 좌표를 지리 정보화하여 표출하여야 하며 표출된 각 포인트 간 거리 계산이 필요하다. 본 연구는 앞서 가공한 데이터를 이용하여 각 이동의 출발지와 도착지를 구분하여 자료를 구축하고 데이터 ID와 날짜 등으로 재조인하는 방법으로 이동 출발지와 목적지 지리정보를 생성하였다. [표 3-4]는 출발지, 도착지, 위·경도 좌표를 지도상에 표시한 것으로 약 324만 개의 포인트 데이터가 생성되었으며, 위치 정보 베이스맵은 알트릭스 제공 street map을 활용하므로 전 세계 이동표시가 가능하며 일부 미국과 유럽 이동도 나타났다.

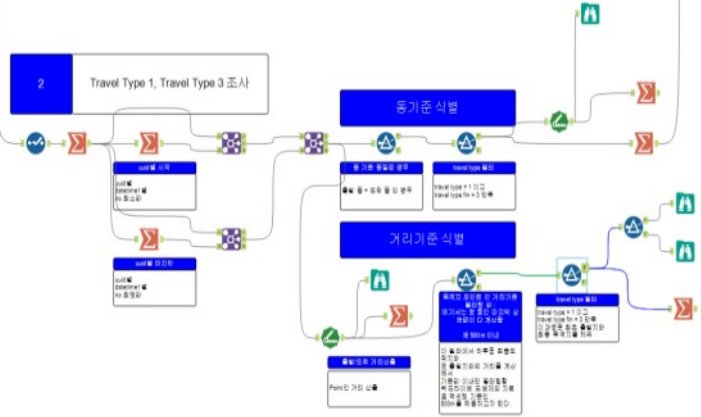
표 3-4 | 출/도착지 위치 데이터 구축 과정

내용	워크플로우
출발과 도착으로 각 열(레코드)을 분리 후 조인한 후 포인트 데이터 생성	
출발지, 도착지 위치 좌: 국토 기준 우: 세계기준	

자료: 저자 작성

집 기반 통행은 앞서 기술한 행정 명칭 일치 기준(대안 1)과 거리 기준(대안 2)으로 구분하여 분석하였으며, 분석 방법은 위·경도 좌표로 생성한 포인트 속성정보에 출발, 도착 정보를 입력하고 각 포인트 간 거리를 산정하여 판단하였다.

표 3-5 | 데이터 구축 과정

내용	워크플로우
대안 1과 대안 2에 대한 첫 이동 출발지와 마지막 이동 도착지 비교 후 두 포인트 간 거리 계산 및 시도별 거리 출력	
대안별 출발지 도착지 분석 결과	 읍면동 기준  거리 기준(1000m)

자료: 저자 작성

[그림 3-25]는 첫 이동의 출발지와 마지막 이동 도착지의 유사 유무를 두 포인트 간의 거리로 판단하는 로직을 코딩화한 것이다. 첫 번째 코딩문은 거리 기준을 500m로 설정하고 이 보다 크면 다음 이동 노드에 반영하지 않고(N), 500m와 같거나 작으면 포함(C)하는 내용이다.

그림 3-25 | 집 기반 활동 판단 알고리즘

Formula (65) - Configuration

Output Column

std1

Data Preview

IF

!ISNULL([distance])

THEN

IF [distance]*500 > [var1] THEN 'N'

ELSEIF [distance]*500 <= [var1] THEN 'C'

ELSE NULL() ENDIF

ELSEIF

[travel_type] = '1' && [category] = '01'

THEN

'S'

ELSE

NULL()

ENDIF

(활동의 연속성 판단기준)

std0

IF

[ID2] = '901' && [std1] = 'C'

THEN

'T'

ELSE

'F'

ENDIF

Data type:

V_WString

Size:

1073741823

(마지막 이동의 도착지가 첫 번째 출발지인지를 판단)

자료: 저자 작성

3) 이격거리 기준 도착지/출발지 판단방안

집 기반 통행을 기준으로 활동패턴을 분석해야 하는 이유는 앞서 기술한 바와 같이 특정 통행이 아닌 일반적인 통행을 대상으로 활동패턴을 분석하기 위해서이다. 본 연구는 김종학 외(2021)의 TGN(Trip Generation Numbering) 알고리즘을 이용해 파악한 하루 중 첫 이동(fs)의 출발지와 마지막 이동(le)의 목적지 간 행정 명칭(대안 1)과 이격거리(대안 2) 기준에 부합하는 원시자료를 필터링하는 작업을 수행하였다. 읍면동 기준(대안 1)으로 fs와 le의 이격거리를 시도별 평균 거리 기준으로 산정한 결과, 0.19~0.55km 정도로 양호한 것으로 나타났으나 최대이격거리는 3.8~18.2km까지 시도별로 차이가 큰 것으로 나타났다. 즉, 평균값은 양호하나 개별 데이터상으로는 마지막 도착지가 집이라고 판단하기에 그 거리가 큰 것으로 나타났다. 또한, 읍면동으로 할 경우 실제 두 지점 간 거리는 작지만 행정 경계가 달라 집 기반 통행에서 제외되는 경우도 있었다. 대안 2(이격거리 500m) 기준으로 할 경우 시도별 평균 이격거리는 0.17km로, 읍면동 기준보다 해당 거리가 56% 정도 감소하는 것으로 나타났다. 거리 기준으로 할 경우 시도별 차이도 동 기준에 비해 줄어드는 것으로 나타났다. 집 기반 통행 선별 기준에 따른 원시자료 유효 샘플의 감소 정도를 분석한 결과 읍·면·동으로 할 경우 일 평균 14,212인의 샘플이 확보되었으나 1,000m 기준으로 하였을 때는 8,052인, 500m 기준에서는 6,472명으로 샘플 수가 줄어드는 것으로 나타났다. 이에 지역별 활동패턴 분석을 위해서는 가능한 많은 수의 샘플 수 확보가 필요하나 집 기반 활동패턴이라는 조건을 충족하기 위해서는 거리 기준의 활동패턴 샘플 확보가 필요하다고 사료되었다.

표 3-6 | 집 기반 분석 대안별 fs(첫 이동)와 le(마지막 이동)의 이격거리

도구분	동 기준 이격거리		거리 기준(500m)	
	평균(km)	최대(km)	평균(km)	최대(km)
강원도	0.52	18.23	0.17	0.50
경기도	0.35	15.22	0.16	0.50
경상남도	0.42	8.73	0.17	0.50
경상북도	0.46	8.67	0.17	0.50
광주광역시	0.27	5.93	0.17	0.50
대구광역시	0.28	6.57	0.16	0.50
대전광역시	0.27	5.22	0.16	0.50
부산광역시	0.29	8.62	0.16	0.50
서울특별시	0.19	4.06	0.15	0.50
세종특별자치시	0.40	3.76	0.17	0.50
울산광역시	0.39	6.29	0.17	0.50
인천광역시	0.29	6.75	0.16	0.50
전라남도	0.55	14.19	0.17	0.50
전라북도	0.40	9.43	0.17	0.50
제주특별자치도	0.53	14.23	0.18	0.50
충청남도	0.50	10.59	0.17	0.50
충청북도	0.46	8.88	0.17	0.50
평균	0.39	9.14	0.17	0.50

자료: 저자 작성

표 3-7 | 집 기반 분석 대안별 이용자 수 도출 결과

(단위 : 인)

일자	읍면동 기준	최초 출발지와 최종 도착지 이격거리		
		1000m	600m	500m
2022-03-14	13,691	7,925	6,775	6,324
2022-03-15	14,111	8,475	7,335	6,856
2022-03-16	14,280	8,569	7,408	6,930
2022-03-17	14,229	8,391	7,218	6,770
2022-03-18	14,339	7,947	6,822	6,368
2022-03-19	12,183	6,329	5,378	4,988
2022-03-20	11,094	5,811	4,995	4,645
2022-03-21	14,841	8,741	7,509	7,037
2022-03-22	15,235	9,163	7,898	7,438
2022-03-23	15,403	9,262	8,004	7,506
2022-03-24	15,668	9,393	8,155	7,667
2022-03-25	15,960	8,881	7,699	7,206
2022-03-26	13,392	6,865	5,800	5,392
2022-03-27	11,844	6,061	5,168	4,801
2022-03-28	15,078	8,758	7,501	7,005
2022-03-29	15,368	9,271	8,032	7,545
2022-03-30	15,143	9,032	7,786	7,283
2022-03-31	14,721	8,544	7,305	6,797
2022-04-01	15,390	8,339	7,158	6,686
2022-04-02	14,054	7,014	5,972	5,571
2022-04-03	12,421	6,326	5,450	5,097
총합	298,445	169,097	145,368	135,912
일 평균 이용자	14,212	8,052	6,922	6,472

자료: 저자 작성

4) 활동의 시공간적 연속성 판단방안

본 자료는 차량 이동 데이터로 이동과 목적지 도착에 대한 판단작업이 중요하며, 해당 데이터는 자동차 이동궤적 데이터로 신호등에서의 2~4분 정도의 정차도 목적지로 기록하고 있어 이에 대한 필터링이 필요하다.

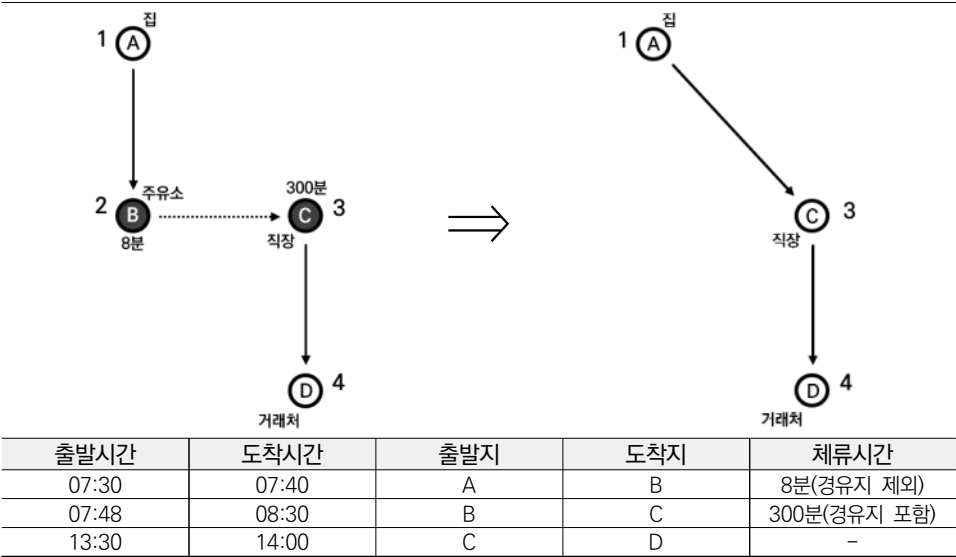
활동의 연속성을 판단하기 위해서는 첫 이동의 목적지가 두 번째 활동(중간 활동 1)의 출발지의 연속성 여부를 판단할 필요가 있고, 해당 앱의 오작동, GPS 오류, 도보 이동 등 다양한 이유로 해당 데이터의 공간적·시간적 연속성을 판단하는 기준이 필요하다. 또한 중간 이동의 경우 앞서 언급한 귀가와는 달리 이후 활동 여부를 판단할 수 있어 중간목적지를 연계하여 활동패턴을 파악하는 것도 필요하다. 본 연구는 중간 이동의 연속성을 확보하고자 현재 이동(n 열)의 목적지와 다음 이동($n+1$)의 출발지 간 거리(500m 이내)와 체류시간 10분 기준으로 판단하였다. 이에 집 기반 활동은 이격거리만을 고려한 반면, 중간 이동은 체류시간을 추가하여 중간 이동의 도착지 유무를 판단하고자 하였다. [그림 3-26]은 중간 이동이 미포함된 경우를 설명한 것으로 첫 이동의 목적지에서 8분간 체류, 중간 이동 1의 목적지에서 300분을 체류하였기 때문에 노드 2는 목적지에서 제외하고 노드 3을 목적지로 분류하였다. 즉, 집에서 출발하여 회사로 가는 도중 8분은 경유지에서 제외하였으며 데이터상으로 3(직장)의 체류시간은 300분으로 경유지에 포함한 사례이다. 분석상에서 주의해야 할 사항은 체류시간은 원 시데이터에서 계산해야 한다는 점이며, 분석데이터에서 중간 이동 포함 여부는 현재 이동(n 열) 목적지와 이후 이동($n+1$ 열)의 이격거리와 체류시간 두 가지 조건의 논리합($\wedge$)이 참일 경우에 포함시켰다.

표 3-8 | 중간 이동 포함 판단 논리구조

구분	논리 A	논리 B	논리합(A∧B)
조건	이격거리 500m 이내	체류시간 10분 이상	(500m 이내) ∧ (10분 이상)
판단 값	True	True	True

자료: 저자 작성

그림 3-26 | 중간 이동의 연속성 확보 구분 방안 (예시)

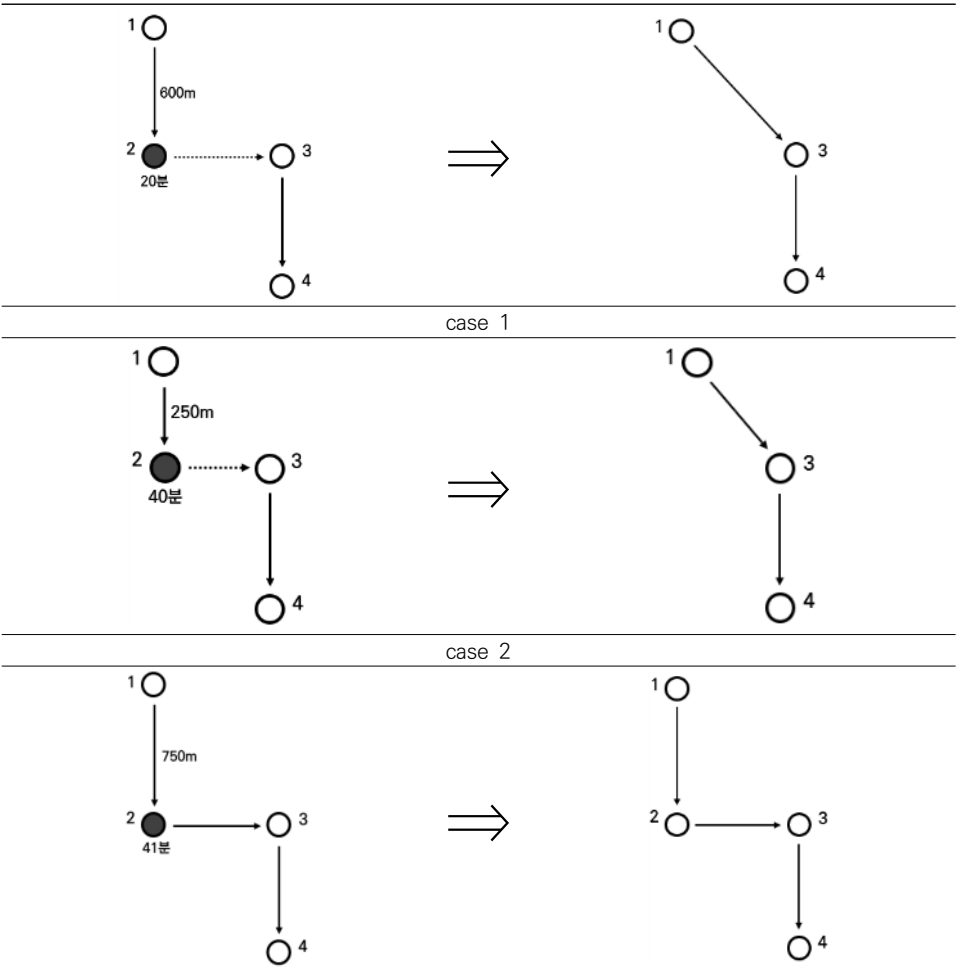


주: 체류시간은 현 통행 도착지에서의 체류시간임

자료: 저자 작성

[그림 3-27]은 노드 연속성을 체류시간과 이격거리에 따라 판단하는 사례로 case 2는 첫 이동 중 집 주변 출발 이동의 출발지를 집과 인접한(500m 이내) 경우로 출발지를 집으로 일원화한 사례이다. 마지막으로 case 3은 중간 이동의 목적지가 집과 떨어져 있고 체류시간도 10분 이상으로 중간 이동 노드(출/목적지)로 포함된 사례이다.

그림 3-27 | 노드(출/도착지) 포함 여부 논리구조



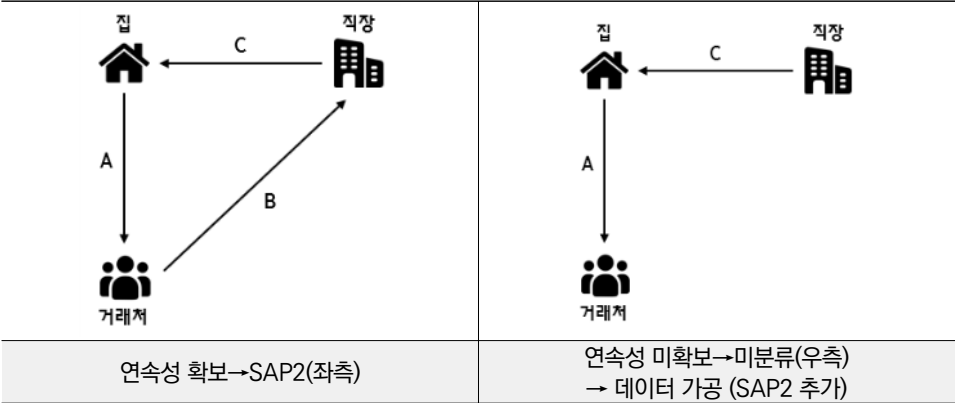
예시	A: 도착지(n)/출발지(n+1) 간 거리 (500m 이상)	B:체류시간(1분 이상) (n+1 출발시간-n 도착시간)	A∧B(논리합) (true, false)
case 1	600m (true)	8분 (false)	false
case 2	250m (false)	25분 (true)	false
case 3	750m (true)	41분 (true)	true

자료: 저자 작성

5) 미싱 데이터 복원방안

활동패턴의 연속성이 확보되지 않는 [그림 3-28]과 같은 미분류 데이터를 중간 경유지 출발지, 도착지 이격거리 파악 등의 알고리즘을 통해 활동패턴으로 구분하는 작업을 진행하였다. [그림3-28] 우측은 하루 중 첫 이동과 마지막 이동데이터는 있지만 거래처에서 직장으로의 통행 기록이 누락된 경우이다. 상기 데이터는 거래처와 직장의 이격거리와 거래처 도착시간, 직장→집 출발시간 차이 등을 고려하여 추가통행 유무를 추출하였다.

그림 3-28 | 활동패턴 미분류 데이터 복원



자료: 저자 작성

4. 활동패턴 대시보드 개발

1) 활동패턴 대시보드 개발 필요성 및 요구사항 정의

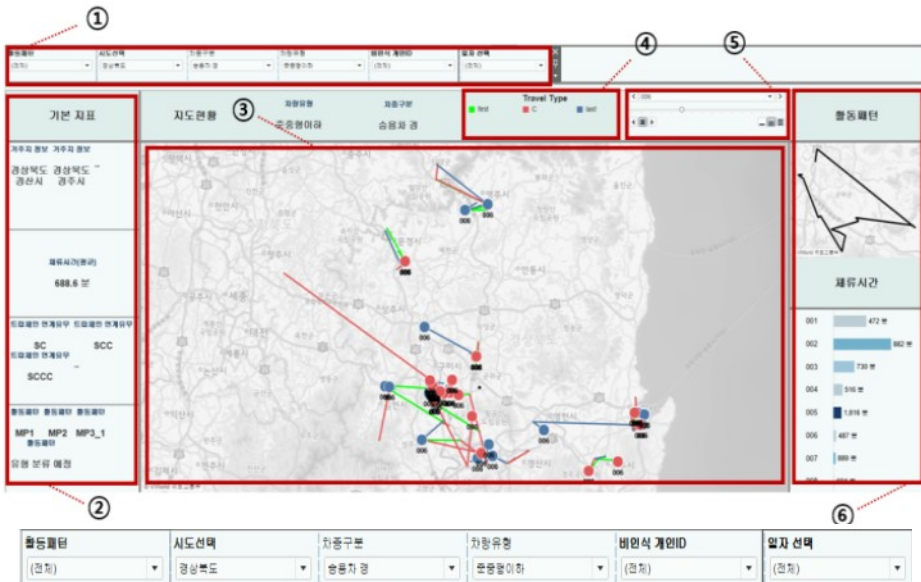
원시자료의 시각화를 통한 활동패턴을 파악하기 위해서는 주어진 기종점 위치와 시간정보를 지도상에 표출하는 시각화가 필요하다. 하지만 3주간 원시자료 약 160만 개의 개별 이동정보를 단일 지리 정보화할 경우 데이터 중복 및 겹침으로 개개인의 이동현황을 파악하는 데 한계가 있었다. 이에 본 연구에서 파악하려고 하는 집-기반의 활동패턴을 검토하기 위해서는 개인별, 일자별 이동행태를 시간순으로 구분할 필요가 있다. 활동패턴 연구에서 주의 깊게 살펴보아야 할 이동의 연속성은 주어진 데이터의 공간적 연속성을 판단할 수 있어야 하며 이를 위해서는 현재 이동의 목적지가 다음 이동의 출발지여야 한다. 이 경우 위도·경도 좌표의 일치성만으로 출발·도착지 일치성을 평가하기에는 GPS 기능의 오차로 정확한 위치로 일치하는 어려워 출발도착지 위치 이격 거리에 대한 기준설정이 필요하며 본 연구는 도보권이 대략 반경 500m인 점을 고려하였다. 활동패턴 대시보드 개발은 첫째, 개인의 일자별 이동을 포인트(출발지, 도착지)로 나타내고 출발지와 목적지는 링크(선)로 지도상에 표시할 수 있어야 하며, 둘째, 개인의 이동 경로를 선으로 요약하여 표현할 수 있어야 한다. 셋째로 개인의 하루 중 이동 시점을 첫, 중간, 마지막으로 구분할 수 있게 포인트와 링크의 색깔을 달리하여 표현할 수 있어야 하고, 넷째로 지역/일자/차종별로 구분하여 활동패턴을 지도상에 표시할 수 있어야 하며 다섯째로 이동순서별로 이동궤적을 표시할 수 있어야 한다. 여섯째로는 목적지 체류시간과 통행시간을 표시할 수 있어야 하며 일곱째로 줌 기능으로 출발지, 도착지 세부 위치정보 파악이 가능해야 한다. 마지막으로 활동패턴을 표시할 수 있어야 한다.

2) 개발과정

활동패턴 대시보드는 상기 요구 사항 정의에 기반해 메뉴 구성을 상단 메뉴, 기본메뉴, 활동패턴 등 6개의 메뉴로 구성하였다. 상단 메뉴에는 활동패턴 유형, 시·도, 차종, 차량 유형, 개인 ID, 일자 선택 메뉴로 구성하였는데, 대시보드의 장점인 필터링 툴바 기능으로 구성하여 분석가가 원하는 시·도, 차량 유형, 활동패턴 등 원하는 기준을 다양하게 설정할 수 있게 구성하였다. 기본메뉴에는 거주지, 체류시간, 트립 체인 연계 유무, 활동패턴 정보 등을 기술하였고, 이동유형에는 하루 이동을 첫, 중간, 최종으로 구분하는 출발지와 목적지를 색상으로 구분하고자 하였다. 상단 메뉴는 활동패턴, 지역(시도), 차종, 일자 등에 대한 툴바 기능으로 분석자가 원하는 시도, 일자, 패턴별로 활동패턴을 시각화하여 볼 수 있는 기능으로 개인의 일자별 활동의 연속성을 검토할 수 있다. 개발단계에서 대시보드의 주요 기능은 이동의 연속성을 시각적으로 검토함으로써 활동패턴별 수정알고리즘을 개발하는 데 있으며, 개발 완료 단계에서는 지역별 활동패턴을 일자별, 차종별로 파악하는 데 활용할 수 있을 것으로 기대된다.

표 3-9 | 활동패턴 검토를 위한 대시보드 개발

구분	내용
① 상단 메뉴	활동패턴, 시도 선택, 차종 구분, 차량 유형, 비 인식 개인 ID, 일자 선택
② 기본메뉴	거주지 정보, 체류시간(평균), 트립 체인 연계 유무, 활동패턴
③ 활동패턴	지도현황
④ 이동유형	초록-첫 이동, 빨강-중간 이동, 파랑-마지막 이동
⑤ 이동순서	이동순서 (1-2-3 등)
⑥ 이동 경로	경로 형태, 목적지별 체류시간



자료: 저자 작성

그림 3-29 | 대시보드 메뉴 구성항목

활동패턴 (전체)	시도선택 경상북도	차종구분 승용차 경	차량유형 운용중이하	비인식 개인ID (전체)	일자 선택 (전체)
--------------	--------------	---------------	---------------	------------------	---------------

활동패턴 MP2 검색 텍스트 입력 (전체) MP1 MP2 MP3_1 MP4_1 MP5_1 유형 분류 예정 활동패턴	시도선택 경상북도 검색 텍스트 입력 (전체) 강원도 경기도 경상남도 경상북도 광주광역시 대구광역시 대전광역시 부산광역시 서울특별시 세종특별자치시 충청남도 충청북도 충청북도 충청북도 시도 선택	차종구분 승용차 경 검색 텍스트 입력 (전체) 승용차 경 승용차 대 승용차 대 승용차 대 승용차 중 승합차 경 승합차 대 승합차 소 승합차 중 이륜차 대 이륜차 소 이륜차 중 해당없음 화물차 대 화물차 대형 화물차 소 화물차 중 차종 구분	비인식 개인ID (전체) 검색 텍스트 입력 (전체) 000263 000464 001196 002792 003275 003707 004205 004421 005554 006050 007764 008616 009399 012496 014563 014993 017056 017581 018345 020549 021465 비인식 개인 ID	일자 선택 (전체) 검색 텍스트 입력 (전체) 2022-03-14 2022-03-15 2022-03-16 2022-03-17 2022-03-18 2022-03-19 2022-03-20 2022-03-21 2022-03-22 2022-03-23 2022-03-24 2022-03-25 2022-03-26 2022-03-27 2022-03-28 2022-03-29 2022-03-30 2022-03-31 2022-04-01 2022-04-02 2022-04-03 일자 선택
---	--	---	---	--

자료: 저자 작성

상기 기술한 요구 사항과 메뉴 구성을 충족시키기 위해 알트릭스(Alteryx)를 이용한 [표 3-10]의 워크플로우를 작성해 원시자료를 가공하였다. 활동패턴 대시보드 개발을 위해 먼저 원시자료 개별 레코드(출발, 도착) 정보를 분리하여 개별적으로 포인트 데이터를 생성하고 다시 결합하는 작업이 필요하며, 포인트와 링크의 일치 여부를 판단하는 연결 값(key 값)을 이용하여 출발지와 도착지, 이동(선)을 연계하는 작업 또한 필요하다. 마지막으로 일자별로 개인의 이동순서 필드 생성을 통해 이동궤적을 파악할 수 있는 작업 등이 필요하다.

활동패턴 대시보드는 알트릭스(Alteryx) 가공작업을 입력자료로 하고 태블로의 대시보드 각 시트 작업을 진행하여 구현하였다.

표 3-10 | 활동패턴 대시보드 개발을 위한 입력 파일 가공작업

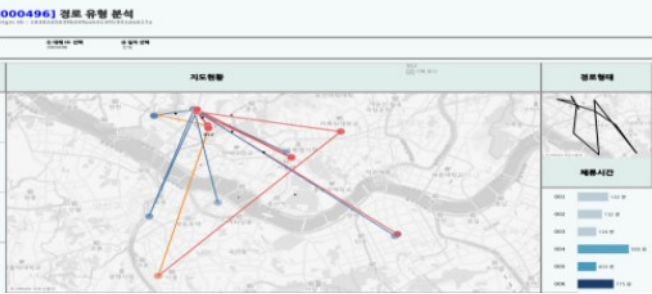
구분	워크플로우
step 1: 기종점 이동 분리	
step 2: travel type 매칭	
step 3: 대시보드 생성 입력자료 도출	

자료: 저자 작성

3) 대시보드 개발 결과

[표 3-11]은 활동패턴 대시보드를 이용하여 지역별 특정인의 3일 동안 기종점 이동 궤적을 표시한 자료이다.

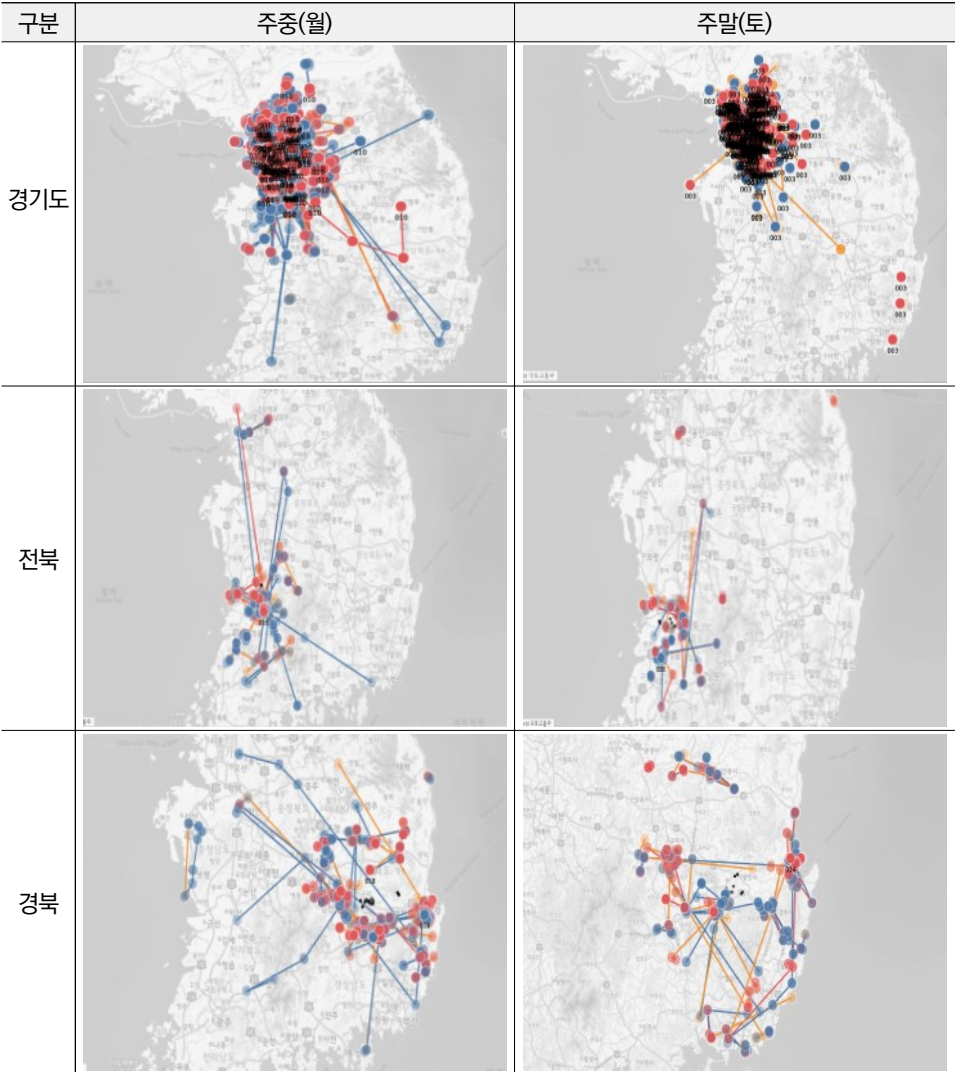
표 3-11 | 원시자료의 개인별 3일간 총활동패턴

구분	가공 전 활동패턴
서울특별시 (ID:000496)	
경기도 (ID:000011)	
대구광역시 (ID:000111)	

자료: 저자 작성

[표 3-12]는 활동패턴 대시보드를 이용하여 지역별 월요일과 토요일의 기종점 이동 궤적을 표시한 자료로 주중과 주말 자료가 상이한 것으로 나타났다.

표 3-12 | 지역별 원시데이터 표출



자료: 저자 작성

[표 3-13]은 서울시 영등포구에 거주하는 것으로 추정되는 개인의 활동 기종점 간을 나타낸 그림으로 첫 출발지(집/녹색)가 연속적으로 영등포구 대방역 인근에서 출발하는 것으로 나타났다. 첫 이동의 출발지는 일정하나 첫 이동의 도착지는 논현, 행당 등 상이한 것으로 나타났으며 중간 이동인 적색 이동의 목적지가 집과 겹치는 것으로 볼 때 출발 후 집 귀가 후 다른 목적지로 재이동을 하는 활동패턴을 가지는 것으로 사료된다.

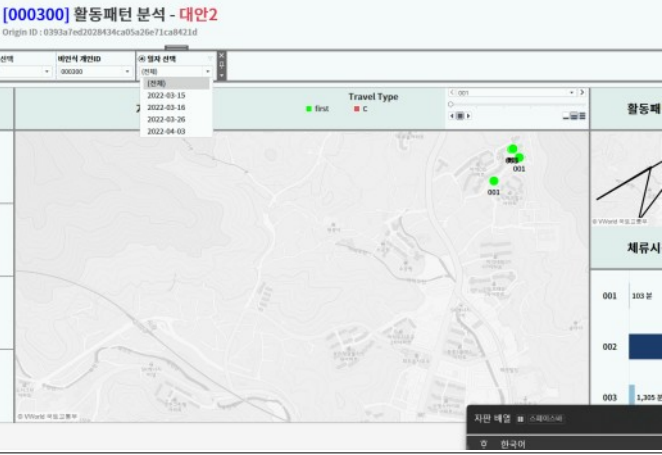
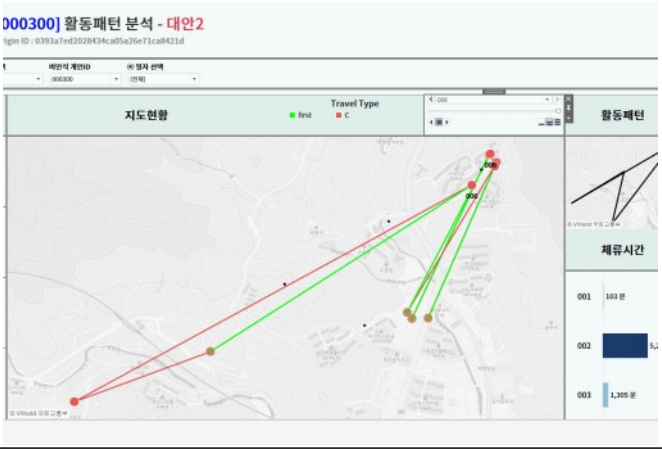
표 3-13 | 활동패턴 검토 사례 1: 거주지 위치, 이동 연속성



자료: 저자 작성

[표 3-14]는 남양주시 마도 다산 신도시의 3차 신도브래뉴 아파트가 거주지로 추정되는 이용자의 기종점 간 3.15, 3.16, 3.26, 4.3일의 4일간 이동패턴을 나타내고 있으며, 마지막 이동 목적지가 출발지와 상이해 중간 이동인 적색 포인트와 링크로 표시되고 있어 추가적인 검토가 필요하다.

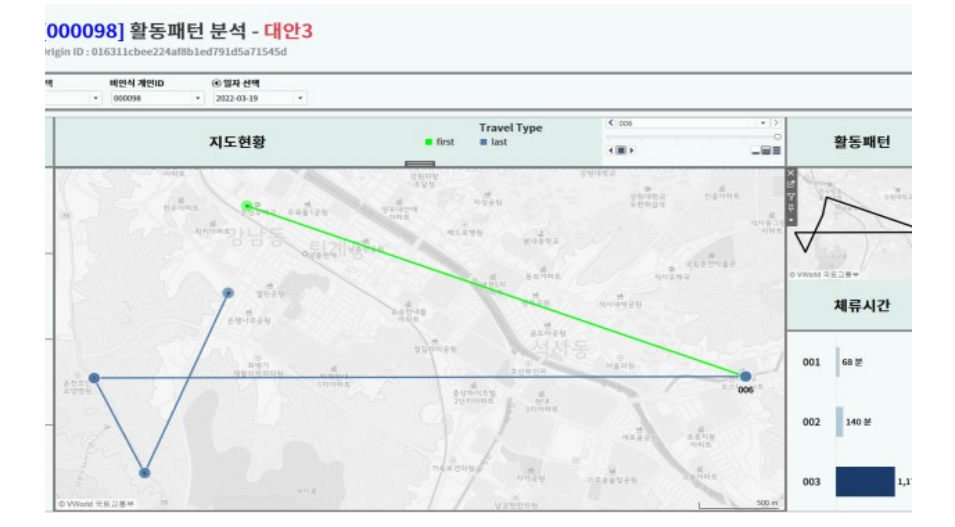
표 3-14 | 활동패턴 검토 사례 2: 거주지 위치, 이동 연속성 결측

구분	활동패턴
출발지(집)	
전체	

자료: 저자 작성

[그림 3-30]은 춘천시 춘천 현대 성우 오스타 아파트에 거주하는 이용자의 기종점 간 3.19일 이동으로 집에서 춘천우체국으로 이동 후 춘주로 92 근처로 이동한 것으로 추정되나 원시자료에서 이 이동이 미싱된 것으로 보여진다. 아래 사례의 미싱된 이동을 생성한다고 가정하면 집에서 출발하여 해당 일자에 집으로 귀가한 집-기반 이동이며, 중간목적지는 3곳인 활동패턴으로 구분될 수 있는 사례이다. 이 경우 [그림 3-30] 우측 상단의 활동패턴처럼 표현 가능하다.

그림 3-30 | 활동패턴 검토 사례 3: 거주지 위치, 중간 이동 결측



자료: 저자 작성

5. 소결

3장에서는 원시자료 가공과 특성에 대한 검토를 통해 활동패턴 분석을 위한 주거지, 직장, 출발·도착시간, 일자 등 추가속성정보를 생성하였다. 빅데이터는 개인정보 보호 등으로 개인정보가 거의 없어 하루 중 첫 통행의 출발지와 목적지의 빈도를 통해 주거지와 직장 등을 파악하였으며 이러한 노력은 특정 데이터 가공을 통해 구현될 수 있다. 본 연구의 데이터처럼 개인정보는 없지만 개인 구분 ID, 위치, 시간정보로 모빌리티, 활동 정보를 파악하기 위해서는 하루 중 통행 발생의 순서 파악이 중요하며, 향후 자율주행차량, 소형 대중교통, PM 등의 빅데이터 분석에서도 그 활용도가 높을 것으로 보여진다. 빅데이터 분석에서 중요한 것은 분석자가 원하는 결과물을 얻기 위한 데이터 요구도를 정의하고 이와 불일치 하는 데이터를 정제하는 과정이다. 본 연구는 활동패턴 분석과 함께 개인의 하루 단위 활동의 특징을 파악할 수 있는 목적지에서의 체류시간, 통행시간 등을 산정할 필요가 있었고 이 데이터를 산정하는 과정에서 총활동 시간이 하루 이상인 자료, 데이터 수집 오류로 음수인 체류시간 등이 산정되는 데이터 등을 정제하는 과정을 수행하였다. 특히, 활동패턴은 시공간적인 연속성 파악이 중요하고 이 속성조건 만족을 여러 가지 기준 등을 설정하고 그 구현 과정 등을 기술하였고, 시공간 활동패턴을 위해서는 하루 중 활동이 집에서 시작해 집에서 끝나는 가정을 만족하는 데이터 필터링 과정이 필요한 것을 확인하였다. 또한, 현재 통행의 목적지가 다음 통행의 출발지라는 전제에 부합하는 데이터만을 선정해야 하는 작업이 필요하며 이때 GPS 오류 등의 한계치를 어디까지 둘 것인가를 몇 개의 대안을 설정하여 검토하였으며 이러한 과정은 주어진 데이터의 한계를 극복하면서 공간적 연결성을 제고하는 과정이라고 사료된다. 데이터 가공과 정제과정은 분석자가 요구하는 정의에 만족하는 데이터를 찾는 과정으로 모든 분석에서 동일한 과정을 거칠 필요는 없다. 예를 들어 개별 통행을 분석한다면 본 연구처럼 통행의 시공간적 연속성을 만족시킬 필요는 없다고 사료된다.

본 연구에서는 활동패턴의 공간적 형상을 파악하고 그 특징을 좀 더 과학적으로 파

악하고자 활동패턴 대시보드를 개발하였다. 활동패턴을 설문 조사 등으로 조사한다고 하면 응답자는 현재 통행의 도착지는 다음 통행의 출발지라고 자연스럽게 기입하고 통행시간 등을 기입할 것이지만, 데이터로 이러한 활동을 파악하기 위해서는 주어진 데이터(통행)의 목적지가 다음 통행의 출발지 인지를 파악해야 하고 이 과정은 귀가 전까지 연속적으로 판단해야 한다. 본 연구는 활동패턴 대시보드 개발을 통해 하루 중 첫 통행을 출근 통행, 하루 중 마지막 통행을 귀가 통행으로 구분하여 파악할 수 있었으며, 활동패턴을 지도상에 구현하는데 필요한 데이터가 노드(Node, 출/도착지), 링크(Link, 출발지-도착지) 그리고 경유지(Stop over) 개수 등 3가지 자료가 필요함을 개발과정을 통해 파악할 수 있었다. 이러한 과정에서 교통계획, 길 찾기 등에서는 노드와 링크 데이터가 필요하지만 활동패턴 파악을 위해서는 경유지(Stop over) 개수가 패턴의 유형을 구분하는 데이터로 노드, 링크와 함께 필수적이다. 3장에서 수행한 데이터 가공, 필터링, 정제 그리고 활동패턴 대시보드 개발 등은 4장에서 분석할 내용의 기초 자료 생성과 분석 결과의 신뢰성을 제고시키기 위해서 반드시 필요한 과정이며 많은 피드백 과정을 통해 고도화하였으며, 3장에서 개발한 활동패턴 대시보드의 입력자료 등은 4장에서 제시할 활동패턴을 단순 활동패턴(SAP, Simple Activity Pattern), 복합 활동패턴(CAP, Complex Activity Pattern)으로 구분한 기초자료로 활용하였다.



CHAPTER 4

활동패턴 유형의 구분 및 분석

- 1. 활동패턴의 구분 및 유형화 방안 81
- 2. 활동패턴 유형 분석 93
- 3. 활동패턴의 지역별 요일별 분석 107
- 4. 소결 128

04 활동패턴 유형의 구분 및 분석

4장은 3장에서 개발한 활동패턴 대시보드의 노드/링크/경유지(L/N/S) 정보를 활용하여 개인의 하루 단위 활동패턴을 목적지로의 회귀성 여부에 따라 단순 활동패턴(SAP, Simple Activity Pattern)과 복합 활동패턴(CAP, Complex Activity Pattern)으로 구분하였다. 분석을 위해 목적지에서의 최소 체류시간을 선정하였다. 3장에서 제시한 분석 방법 적용 결과, 활동패턴의 시공간적 연속성을 만족하는 활동패턴 분석 대상은 약 6만 8천 개로 나타났다. 또한 L/N/S로 구분한 활동패턴 총 60개를 점유율을 기준으로 32개로 집합화하여 활동패턴 개념도와 함께 제시하였다. 하루를 기준으로 집 밖에서 활동하는 시간을 총활동 시간으로 명명하고 이를 체류시간과 동행시간으로 구분하여 지역별, 요일별로 비교하였다.

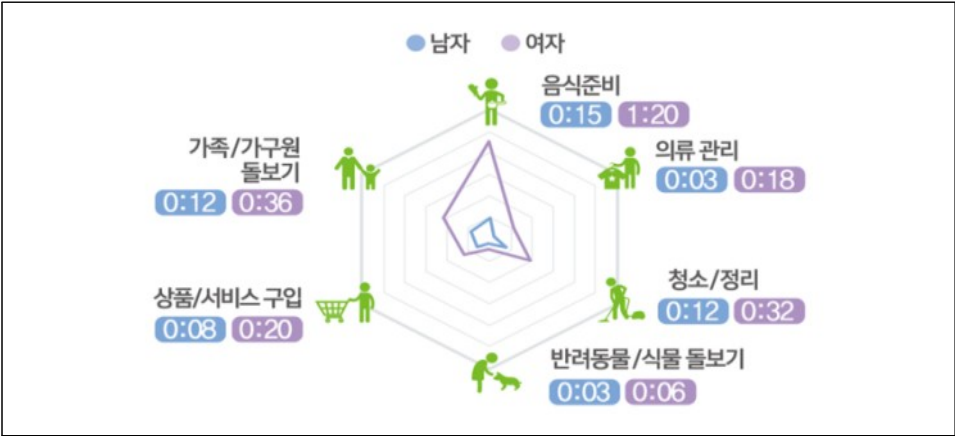
1. 활동패턴의 구분 및 유형화 방안

1) 활동패턴 파악을 위한 체류시간 선정

모빌리티 빅데이터로 개인의 일상 활동을 파악을 위해 결정해야 할 사항은 도착지점에서의 체류시간이다. 본 연구 데이터는 대부분 자동차 이용자들이 이용하는 앱이기 때문에 특정지점에서의 체류는 다양한 원인에 의해 발생할 수 있다. 일례로 주유소의 주유, 자녀들의 학원 또는 학교에서 픽업, 편의점 물품 구매 등을 위한 체류시간은 작지만 해당 활동이 빈번하고 차량 경로가 변경될 수 있고, 자녀의 어린이집, 학교로 데려다주는 활동(픽업), 앱 주문 음식 받으러 가는 활동 등은 목적지에서의 체류시간이 짧지만 해당 활동의 중요성은 크다고 사료된다.

또한, [그림 4-1]에서 알 수 있듯이 상품/서비스 구입에 여자는 20분, 남자는 8분 정도로 10분 이하의 짧은 체류시간에도 특정 활동이 이루어지고 있는 것으로 나타났다.

그림 4-1 | 성인의 시간 사용



자료: 통계청. 2019년 생활시간 조사결과. (2020:요약2)

본 연구는 활동패턴에 대한 분석연구로 활동을 위한 체류시간 기준 여부에 따라 샘플링(목적지) 수가 달라질 수 있기 때문에 체류시간을 10분, 30분 기준으로 구분하여 뒤에 기술할 활동패턴별 샘플 수의 점유율을 비교하였다.

이후 본 연구는 활동패턴을 하루 중 집 이외의 출발지(예: 직장)로의 회귀 통행 유무에 따라 활동패턴의 대구분을 단순 활동패턴(SAP, Simple Activity Pattern)과 복합 활동패턴(CAP, Complex Activity Pattern)으로 구분하였고, 각 대구분별 세 구분은 목적지 개수에 따라 구분하였다. 분석자료의 체류시간 기준별 활동패턴 비율을 살펴보면 체류시간 10분 기준의 SAP 비율은 80.23%, CAP 비율은 19.77%였고 체류시간 30분 기준에서는 SAP 87.21%, CAP 12.79%로 나타났으며, 체류시간 기준이 길수록 통행 수가 줄어 체류시간 30분 기준에서 SAP(단순 활동패턴)의 비율이 늘고 CAP(복합 활동패턴) 비율이 줄어드는 것으로 나타났다.

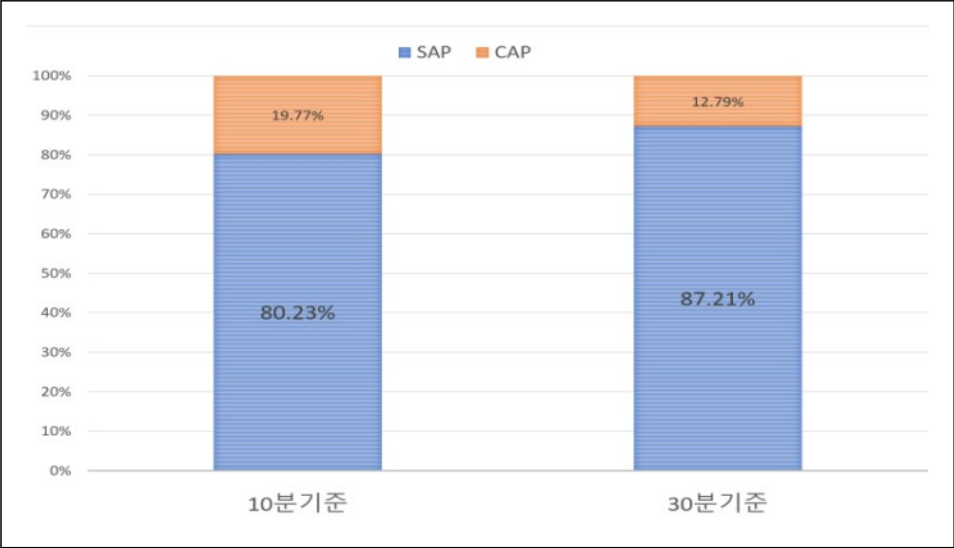
본 연구는 비교적 짧은 시간에도 특정 활동이 빈번하게 발생할 수 있다는 점 등을 고려해 체류시간 10분을 활동의 최소시간으로 정하여 활동패턴을 분석하였다.

표 4-1 | 체류시간 기준별 활동패턴 비율

활동패턴		체류시간 10분 기준		체류시간 30분 기준	
대구분	세구분	유형 내 비율	전체비율	유형 내 비율	전체비율
SAP (단순 활동패턴)	SAP1	65.75%	52.75%	72.70%	63.40%
	SAP2	20.22%	16.22%	18.85%	16.44%
	SAP3 이상	14.03%	11.25%	8.45%	7.37%
	소계	100.00%	80.23%	100.00%	87.21%
CAP (복합 활동패턴)	CAP1	2.89%	0.57%	4.92%	0.63%
	CAP2	36.95%	7.31%	52.97%	6.77%
	CAP3 이상	60.16%	11.90%	42.11%	5.38%
	소계	100.00%	19.77%	100.00%	12.79%
합계		-	100.00%	-	100.00%

자료: 저자 작성

그림 4-2 | 체류시간 기준별 활동패턴 SAP, CAP 점유율 비율 차이

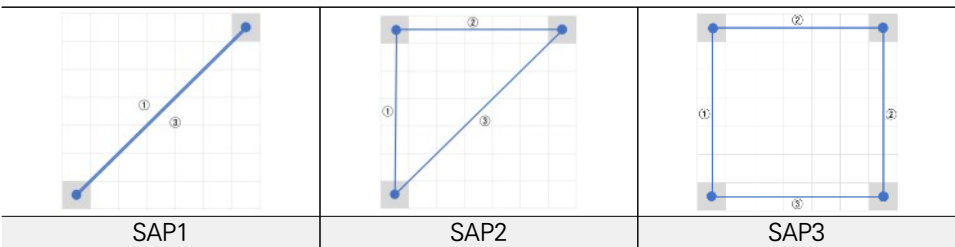


자료: 저자 작성

2) 활동패턴 구분 및 특징

김종학 외(2021)가 제시한 활동패턴은 위치 정보를 반영하지 못한 단순 활동패턴으로 이동과정에서 집을 제외한 나머지 출발지와 도착지의 중복을 고려하지 못했다. 단순 활동패턴은 삼각형, 사각형 등 다각형 형태로 설명할 수 있으며 이러한 형태를 가지기 위해서는 각 이동의 출발지와 도착지의 중복이 없어야 한다. 즉, 직장에서 거래처 방문 후 다시 직장으로의 복귀에 대한 고려는 없었다. 단순 활동패턴에서도 하루 첫, 마지막 이동이 집에서 시작해서 끝나야 된다는 전제는 충족되어야 한다.

그림 4-3 | 단순 활동패턴(SAP) 개념도



주: 1) ①-첫 이동, ②-중간 이동, ③-마지막 이동을 나타냄

2) 원자료의 그림 제목은 모빌리티 패턴(MP) 구분(개념도)이었고 본 연구에서는 복합 활동패턴과 구분하기 위해 단순 활동패턴으로 원자료의 그림 제목을 변경하였음

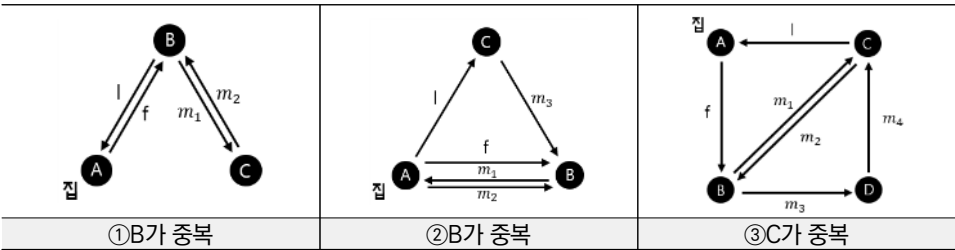
자료: 김종학 외(2021:92) 일부 편집

본 연구에서 개발한 활동패턴(AP, Activity Pattern) 대시보드로 활동패턴을 조사하여 해당 패턴을 구분하였다. 활동패턴 유형은 하루 중 집 이외의 출발지(예: 직장)로의 회귀 통행 유무에 따라 단순 활동패턴(SAP, Simple Activity Pattern)과 복합 활동패턴(CAP, Complex Activity Pattern)으로 구분하였다. 복합 활동패턴은 경유지 중 출발지로 회귀하는 통행이 있을 경우 해당되며 그 예는 [그림 4-4]와 같다. 복합 활동패턴은 집을 제외한 중간 이동 과정에서 출발지와 도착지가 중복되는 것으로 이에 대한 일치 여부를 판단하는 과정이 필요하며, 이 과정은 집-기반 활동과 동일하게 반경 500m를 기준으로 경유지의 중복 여부를 판단하였다.

[그림 4-4] ①의 활동패턴은 링크 4개, 노드 3개이며 집 이외에 장소 B로 회귀하는

복합 활동패턴으로 경유지는 집을 포함해 총 3개이며, ③의 활동패턴은 링크 6개, 노드 4개이고 C로 회귀하는 복합 활동패턴이며 집을 포함한 경유지는 5개이다.

그림 4-4 | 복합 활동패턴(CAP) 개념도



주: f-첫 이동, m-중간 이동, l-마지막 이동을 나타냄

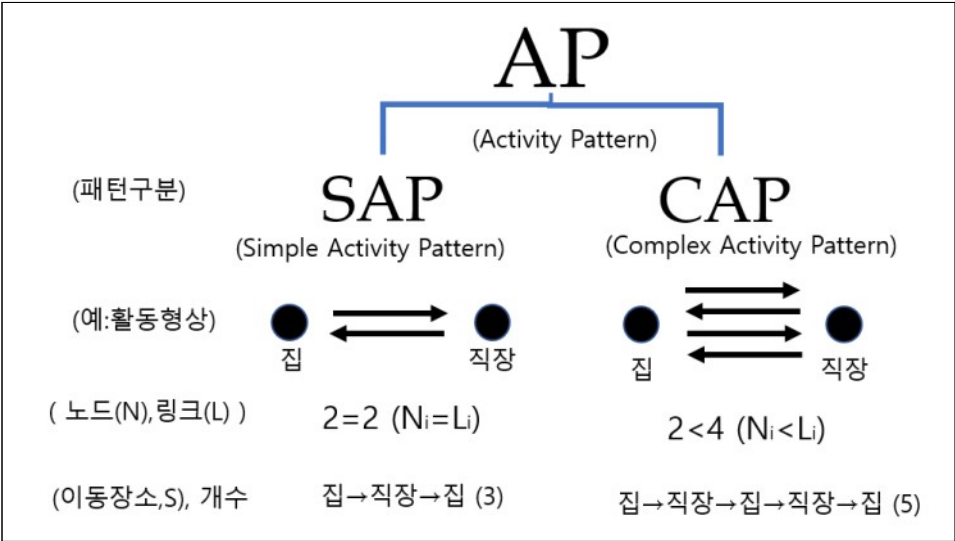
자료: 저자 작성

CAP는 하루 단위 활동패턴에서 동일 목적지를 2회 이상 방문한 활동패턴을 복합 활동패턴이라고 명명하고 그 특징을 활동의 회귀성으로 설명하고자 하였다. 예로 직장 출근 후 거래처 방문 후 직장으로 복귀할 경우 직장은 도착지(출근), 출발지(업무), 도착지(업무), 출발지(귀가)로 총 4회의 출/도착 선택지가 되며, 상기 내용을 경유지로 기술하면 집→직장→거래처→직장→집으로 표현된다. 이러한 통행은 4개의 링크(통행)와 3개의 노드(출/도착지)로 구성되며 여기서 직장은 도착지 2회, 출발지 2회를 차지하므로 복합 활동패턴(CAP)에 해당한다. 4개의 통행으로 만약 집 이외의 목적지가 서로 다를 경우 4개의 링크에 4개의 목적지가 되어야 하나 직장으로 회귀하는 통행으로 목적지(노드)는 3개가 된다. 단순 활동패턴(SAP)은 링크(출발지와 도착지 연계 화살표, Link)와 노드(출발·도착지, Node)로 링크(L) 개수와 노드(N) 개수가 일치하는 특징을 가지고 있다. 다만 복합 활동패턴(CAP)은 앞서 기술한 바와 같이 출발지로 회귀하기 때문에 링크 개수가 노드 개수보다 많은 특징이 있다.

활동패턴의 최초출발지(집)를 포함한 목적지를 경유지(Stop over)라고 명명하고 줄여서 S라고 표현하면 각 활동패턴의 S 개수는 링크보다 1이 많으며 그 이유는 경유지 개수 산정 과정에서 출발지인 집은 최초 통행 발생 장소로 관련 링크가 1개이기 때문이

다. 또한, 본 연구의 활동패턴은 집 기반 활동으로 최초출발지와 최종출발지는 집이기 때문에 모든 활동패턴의 경유지 개수는 홀수라는 특징이 있다. 가장 단순한 활동은 집-직장-집이고, 이 경우 경유지는 3으로 중간 경유지 개수와 중복 여부에 따라 활동패턴이 구분되는 특징이 있다.

그림 4-5 | 활동패턴의 구분과 구성요소



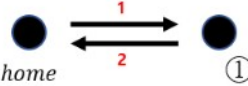
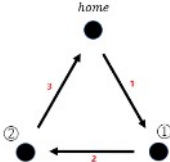
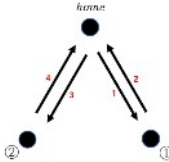
주: N : 노드, L: 링크, I: 노드 개수, k: 추가 링크 개수

자료: 저자 작성

본 연구는 활동패턴의 대시보드 자료 검토를 통해 단순 활동패턴과 복합 활동패턴의 형상을 링크(L), 노드(N), 경유지(S)로 표현해야 한다. 교통계획에서 교통수요 배정 시 활용하는 네트워크는 링크와 노드로만 구성되어 있는 반면 활동패턴을 표현하는 데는 경유지(S)가 추가되어야 하는데, 교통수요 분석에서는 출발지점과 도착지점의 최단 경로를 탐색하기 때문에 한 번 지나간 노드를 다시 돌아가는 것은 최소비용 경로 탐색의 가정에 위배 되기 때문에 경유지(S)는 불필요하다. 하지만 활동패턴에서는 최단 경로를 파악하는 것이 아니라 경유지(S) 위주로 활동패턴을 표시하고 이러한 활동에서는

집, 직장, 특정 장소 등으로의 회귀성 활동패턴이 발생하기 때문에 링크(L), 노드(N) 이외에 경유지(S)를 추가하여 분석해야 한다. 회귀성 활동은 집-직장-집-직장-집, 집-직장-거래처-직장-집 등 특정 장소로 다시 돌아가는 것으로 본 연구에서 구분한 활동패턴의 CAP가 여기에 해당한다. 본 연구는 활동패턴의 유형 구분을 위해 체류시간 10분 기준 집 기반 활동데이터를 추출한 후 하루 단위 개인의 활동에 대해 링크(L), 노드(N), 이외에 경유지(S) 개수를 분석하였다. 이에 본 연구 분석기준에 부합한 하루 단위 개인 활동의 총개수는 약 6만 8천 개로 나타났으며, 약 6만 8천 개의 활동패턴을 링크(L)/노드(N)/경유지(S) 개수 조합을 산정한 결과 총 60개로 나타났다. 60개의 의미는 주어진 빅데이터로 개인의 하루 단위 활동패턴을 유형의 최대 개수라고 볼 수 있으며, 본 연구 빅데이터 기반 활동패턴 유형의 장점은 위치기반 분석으로 회귀성 활동패턴(CAP)의 파악과 그 다양성을 파악할 수 있다는 점이다. 다만, 60개 유형의 각 빈도수(예 : 1)를 고려한 해당 패턴의 집합화 작업은 필요하다. 활동패턴 유형을 [표 4-2]의 예시로 설명하면 전체 약 6만 8천 개 활동 중 가장 많은 빈도수(약 3만 6천 개)를 점유하는 활동패턴은 L(2)/N(2)/S(3)로 노드와 링크 개수가 2개, 집-직장-집으로 경유지가 구성되는 출퇴근 활동패턴이며, 빈도수가 약 1만 1천 개로 두 번째 많은 L(3)/N(3)/S(4)는 링크 3, 노드 3개이면서 경유지가 집-직장-출장지-집으로 이루어지는 삼각형 모양이 활동패턴이다. 마지막으로 빈도수가 약 4천 개로 4번째 많은 L(4)/N(3)/S(5)는 특정 장소(예: 집)를 기준으로 2개의 다른 목적지가 있는 복합 활동패턴이었다.

그림 4-6 | 링크(L), 노드(N), 경유지(S) 기준 활동패턴 구분 예시

L/N/S 구분	L(2) N(2) S(3)	L(3) N(3) S(4)	L(4) N(3) S(5)
활동 패턴 형상(예)			
경유지(예)	집-직장-집 (3)	집-직장-출장지-집 (4)	집-병원-집-마트-집 (5)
빈도수 순위	1	2	4

주: 총활동패턴 분석개수 : 68,289 활동/인 일
자료: 저자 작성

표 4-2 | 링크(L)/노드(N)/경유지(S) 구분별 빈도수(예)

구분	링크(L)	링크(N)	경유지(S)	빈도수(단위: 건)
1	2	2	3	36,024
2	3	3	4	11,182
3	4	3	5	4,704
4	4	4	5	4,375
5	5	4	6	2,531
6	5	5	6	1,890
7	6	5	7	1,393
8	5	3	6	868
9	6	4	7	727
10	6	6	7	684
11	7	6	8	501

자료: 저자 작성

그림 4-7 | 링크(L)/노드(N)/경유지(S) 기준 활동패턴 유형(60개)

구분	링크(L)	링크(N)	이동장소(S)	빈도수
1	2	2	3	36024
2	3	3	4	11182
3	4	4	5	4704
4	4	3	5	4375
5	5	4	6	2531
6	5	5	6	1890
7	6	5	7	1393
8	6	6	7	868
9	7	6	8	727
10	6	4	7	684
11	7	5	8	501
12	7	7	8	394
13	5	3	6	385
14	4	2	5	383
15	8	7	9	324
16	8	6	9	278
17	8	8	9	196
18	9	8	10	151
19	9	7	10	144
20	7	4	8	121

구분	링크(L)	링크(N)	이동장소(S)	빈도수
22	8	5	9	102
23	9	6	10	95
24	9	9	10	87
25	10	8	11	78
26	10	9	11	71
27	10	10	11	61
28	10	7	11	51
29	11	9	12	41
30	11	10	12	38
31	9	5	10	33
32	8	4	9	30
33	11	8	12	29
34	10	6	11	24
35	12	10	13	23
36	7	3	8	20
37	12	11	13	19
38	11	11	12	14
39	11	7	12	13

구분	링크(L)	링크(N)	이동장소(S)	빈도수
40	13	10	14	12
41	12	9	13	9
42	12	12	13	9
43	10	5	11	8
44	13	11	14	8
45	6	2	7	7
46	9	4	10	7
47	13	13	14	7
48	11	6	12	5
49	12	8	13	5

구분	링크(L)	링크(N)	이동장소(S)	빈도수
50	13	8	14	4
51	13	12	14	4
52	12	7	13	3
53	13	15	14	3
54	11	5	12	2
55	13	6	14	2
56	13	9	14	2
57	9	3	10	1
58	10	4	11	1
59	13	14	14	1
60	13	17	14	1

주: 상기 60개는 6만 8천 개 데이터를 L/N/S 3개 변수로 표현 시 도출될 수 있는 조합개수임

자료: 저자 작성

상기 분석에서 활용한 링크(L)/노드(N)/경유지(S) 유형별 데이터의 규칙성을 파악하여 SAP, CAP 구분과 각 세부 유형을 구분하는 방안을 API(Activity Pattern Identification) 분석 알고리즘으로 명명하여 도출하였다. SAP는 링크(L) 개수와 노드(N) 개수가 동일한 특징이 있으며 그 이유는 집을 제외한 나머지 경유지로 회귀하는 통행이 없기 때문이다. 본 연구의 분석 대상 활동은 집 기반으로 모든 활동은 집에서 끝나는 데이터만을 추출하였기에 각 SAP의 경유지(S)는 링크 또는 노드 개수에 1을 더해야 하며, 활동패턴 SAP를 구성하는 링크(L), 노드(N), 경유지(S) 각 개수 간의 관계를 수식으로 표현하면 아래 수식 1과 같다.

$$SAP_i^S + 1 = SAP_i^N = SAP_i^L \dots\dots\dots \text{수식 1}$$

여기서,
 SAP_i^S : 단순 활동패턴(SAP) 유형 i 경유지(S)의 개수
 SAP_i^N : 단순 활동패턴(SAP) 유형 i 노드(N)의 개수
 SAP_i^L : 단순 활동패턴(SAP) 유형 i 링크(L)의 개수

표 4-3 | SAP 유형별 링크(L)/노드(N)/경유지(S) 간의 관계

구분	링크(L)	노드(N)	경유지(S)	관계 S=N+1 또는 L+1
SAP ₁	2	2	3	3=2+1
SAP ₂	3	3	4	4=3+1
SAP ₃	4	4	5	5=4+1
SAP ₄	5	5	6	6=5+1
SAP ₅	6	6	7	7=6+1

자료: 저자 작성

앞서 기술한 바와 같이 출발지로 돌아가는 회귀성 통행이 발생하는 CAP의 L/N/S 개수 간의 관계는 SAP와는 다른 특징이 있었다. CAP 특징은 경유지의 중복이 있어 경유지(S)의 개수가 노드, 링크 개수보다 큰 특징이 있는데, 예를 들어 집-직장-거래처-직장-거래처-직장-집으로 경유지를 달리하는 경우 노드는 집, 직장, 거래처로 3 곳이지만 경유지의 중복이 있기 때문에 경유지는 8곳, 링크는 6개가 된다. 본 연구는 앞서 기술한 L/N/S 유형 중 CAP의 사례를 검토하여 다음과 같은 CAP 내 L/N/S 개수 간의 관계를 도출할 수 있으며, 활동패턴 CAP를 구성하는 링크(L), 노드(N), 경유지(S)의 관계를 수식으로 표현하면 아래 수식 2와 같다.

$$\begin{aligned}
 CAP_i^S &= CAP_i^N + CAP_i^L - AP_i && \dots\dots\dots \text{수식 2} \\
 &= CAP_i^N + CAP_i^L - (CAP_i^N - 1) \\
 &= CAP_i^L + 1
 \end{aligned}$$

여기서,

- CAP_i^S : 복합 활동패턴(CAP) 유형 i 경유지(S)의 개수
- CAP_i^N : 복합 활동패턴(CAP) 유형 i 노드(N)의 개수
- CAP_i^L : 복합 활동패턴(CAP) 유형 i 링크(L)의 개수
- AP_i : 활동패턴(AP) 유형 i의 i 값

표 4-4 | CAP 유형별 링크(L)/노드(N)/경유지(S) 간의 관계

구분	링크(L)	노드(N)	경유지(S)	S=N+L-AP
CAP _{1_1}	4	2	5	5=4+2-1, 5=4+1
CAP _{2_2}	5	3	6	6=5+3-2, 6=5+1
CAP _{3_3}	6	4	7	7=6+4-3, 7=6+1
CAP _{4_3}	8	5	9	9=8+5-4, 9=8+1
CAP _{5_3}	9	6	10	10=9+6-5, 10=9+1

자료: 저자 작성

모빌리티 빅데이터 가공을 통해 개인의 하루 단위 활동패턴을 구분을 위해서는 [표 4-5]와 같이 3단계 분석과정이 필요하며 각 분석 과정마다 데이터 선별을 위한 기준설정이 필요하다. 첫 번째 단계는 김종학 외(2021)가 제시한 TGN(Trip Generation Numbering) 작업이다. 이 작업은 원시데이터(비 인식 개인 ID, 출발/도착시간, 위치 정보)로 하루 중 개인 통행의 순번을 파악하는 과정이며, 이 과정은 활동패턴을 구분하기 위한 가장 기초 작업이면서 활동패턴 구현이 불가능한 1일 1통행 자료들을 구분하는 작업이기도 하다. 이 과정에서 개인의 하루 단위 활동은 최소 첫 통행(1), 마지막 통행(3)으로 구성되어야 한다는 조건을 구현하는 작업이다. 두 번째 단계는 체류시간과 트립 체인 구성을 위한 출발·도착 위치 기준의 충족성을 평가를 통해 하루 단위 개인 활동을 LNS(Link/Node/Spot)로 표현하는 작업이다. 이때 활동은 현재 통행의 목적지가 다음 통행의 출발지가 되어야 하는 위치 기준을 충족해야 하며 앞서 기술한 것처럼 본 연구는 차량 이동데이터라는 특성을 고려해 목적지/출발지의 일치성을 도보권 500m 기준으로 판단하였다. 체류시간은 단시간 활동(주유, 자녀 픽업 등)을 고려해 10분을 기준으로 하였으며 체류시간 산정을 통해 출/도착 연속성 가정이 결여되는 데이터를 제외하는 작업을 수행하였다. 세 번째 단계는 2단계에서 산정한 각 개인의 하루 단위 LNS 정보와 빈도수를 이용하여 세부 활동패턴을 구분하는 작업을 수행하였다. 이를 통해 개인의 하루 단위 활동 유형을 LNS 정보로 유형화할 수 있으며 데이터 상에서 약 6만 8천 개 활동의 LNS 유형은 60가지로 구분되며 이중 빈도수가 높은(약 0.2% 이상) LNS를 세부 활동패턴으로 구분하였다.

표 4-5 | 링크(L), 노드(N), 경유지(S) 기준 활동패턴 구분 예시

구분	1단계	2단계	3단계
분석 알고리즘	TGN (Trip Generation Numbering)	LNS (Link/Node/Spot)	API (Activity Pattern Identification)
	하루 단위 개인 통행의 순번을 첫(1), 중간(2, 중복 가능), 마지막(3)으로 구분	빅데이터의 출발·도착 위치, 시간정보로 하루 단위 개인별 노드, 링크, 경유지 정보를 산정	2단계 산정 노드/링크/장소 정보를 활용한 활동패턴 구분
선정 기준	하루 중 개인별 1건 발생 데이터 등 제외	목적지 체류시간 기준 (본 연구 10분) 출/도착지 공간거리 기준 (본 연구 500M) 체류시간 음수 데이터 제외 및 데이터 치환	점유율 10건 이상 패턴을 활동패턴으로 구분
구현 방법	워크플로우 작성	워크플로우 작성	워크플로우 작성

주: 1) 2단계 기준설정의 데이터 치환은 제외 데이터의 목적지 도착시간 치환을 의미함

2) 워크플로우는 알트릭스 프로그램을 이용하여 작성함

3) TGN : 김종학 외(2021:67-71) 적용

자료: 저자 작성

2. 활동패턴 유형 분석

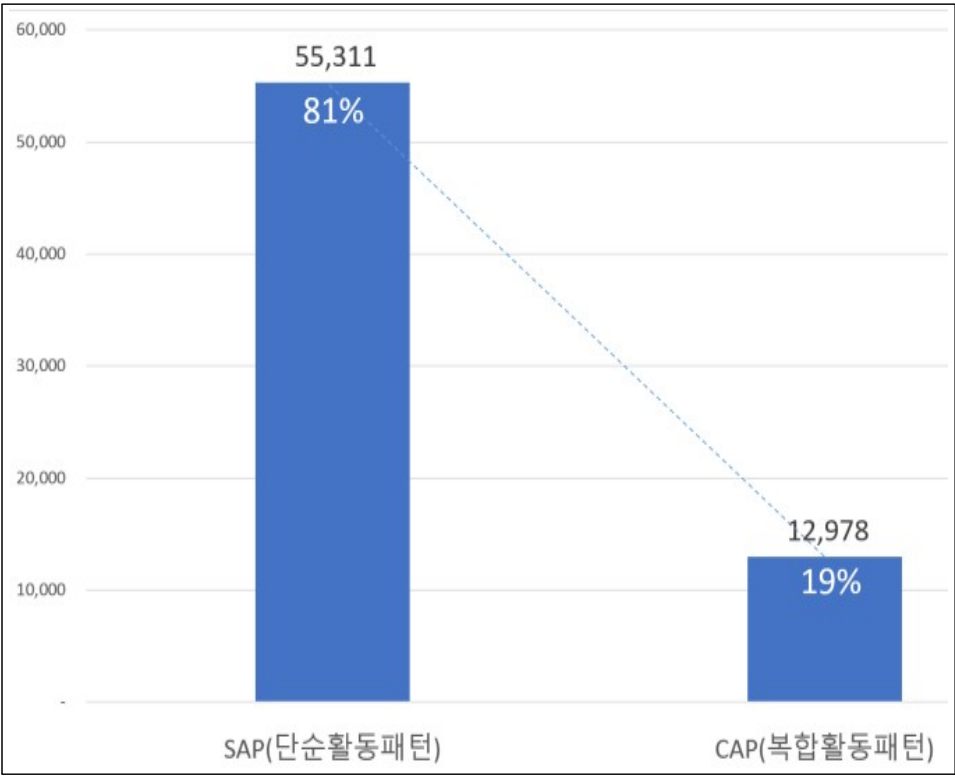
1) 활동패턴 유형 분석 결과

앞서 기술한 방법대로 AP(활동패턴)을 단순 활동패턴(SAP, Simple Activity Pattern)과 복합 활동패턴(CAP, Complex Activity Pattern)의 빈도수로 각 패턴 비율을 산정하였다. 산정 결과, 단순 활동패턴(SAP) 비율은 81%였고 복합 활동패턴(CAP) 비율은 19%로 SAP 비율이 CAP보다 약 4배 높았다.

본 연구에서 복합 활동패턴(CAP)의 존재와 그 유형을 구분할 수 있었던 것은 활동

패턴 대시보드로 활동의 위치 정보를 파악하였기 때문이며, 향후 연구에서도 활동패턴은 이동의 연속성과 활동 위치 파악이 전제되어야 할 것이다. 이 분석의 의미는 기존 가구 통행실태조사 등에서 파악하기 어려웠던 집 이외의 동일 목적지 2회 이상 방문인 복합 활동의 존재와 그 비율을 파악했다는 점이다. 통행실태조사는 대표성이 있는 통행만을 조사해 빈도수가 낮은 통행 파악에는 한계가 있으나 빅데이터는 다양한 활동자료 분석이 가능하다고 볼 수 있는 결과이다.

그림 4-8 | 단순/복합 활동패턴 비율 (단위: 건)



자료: 저자 작성

활동패턴은 집 이외의 목적지 개수를 기준으로 AP1, AP2, AP3 등 AP(Activity Pattern) 7개로 구분하였으며 그 이상의 노드로 구성된 것은 AP 기타로 구분해 총 8개로 구분하였다. 이때 각 AP 유형의 노드 개수는 집을 포함하기 때문에 유형 번호 +1의 노드 개수를 가지는 특징이 있다. 앞서 기술한 바와 같이 L/N/S 정보로 구분 가능한 활동패턴은 총 60개였고 각 L/N/S 정보를 이용하여 SAP, CAP를 구분할 수 있었으며, 활동패턴 대구분 내용을 살펴보면 노드가 2개인 AP1의 비율이 전체 데이터 중 53.4%로 가장 높은 것으로 나타났다. AP1 활동패턴은 하루 중 집과 특정 장소 두 곳을 오가는 활동으로 특정 장소가 직장이라면 분석 활동 2개 중 1개 이상이 집과 직장을 오가는 활동이라는 의미이며, AP1 데이터 내의 단순 활동패턴과 복합 활동패턴 비율은 99:1로 집과 특정 장소(예: 직장)만을 오가는 활동패턴이 절대적으로 높게 나타났다. 다만, 집과 특정 장소(예: 직장)를 하루 2회 이상 오고 가는 활동패턴(집-직장-점심(집)-직장-집)도 발생하는 것으로 나타났다. 대구분 AP는 집 이외의 목적지 개수에 따른 구분으로 AP 번호가 증가할수록 해당유형의 비율은 감소하는 것으로 나타났다. 특히 목적지 횟수가 증가함에 따라 해당 AP 유형의 비율은 감소하였으며 AP1(53.4%)과 AP2(23.6%)의 비율 차이는 약 30%p, AP2(23.6%)와 AP3(11.9%)의 차이는 약 12%p로 나타났다. 이러한 결과는 AP1/2 유형에서 목적지 1회 증가 시 해당 활동의 비율은 약 절반씩 감소하는 경향이 있다고도 볼 수 있으며 추후 개인별로 복합 활동과 단순 활동의 비율을 분석한다면 개인의 SAP, CAP 비율 비교를 통해 개인의 일상 활동과 특정 활동이 구분될 수 있음을 검토할 수 있을 것이다. 즉, 최종적으로 집과 직장을 오가는 개인이 특정일에는 출장을 가거나, 친지 방문, 여행 등의 활동도 구분하여 파악할 수 있을 것이다.

표 4-6 | 활동패턴 대구분 비율

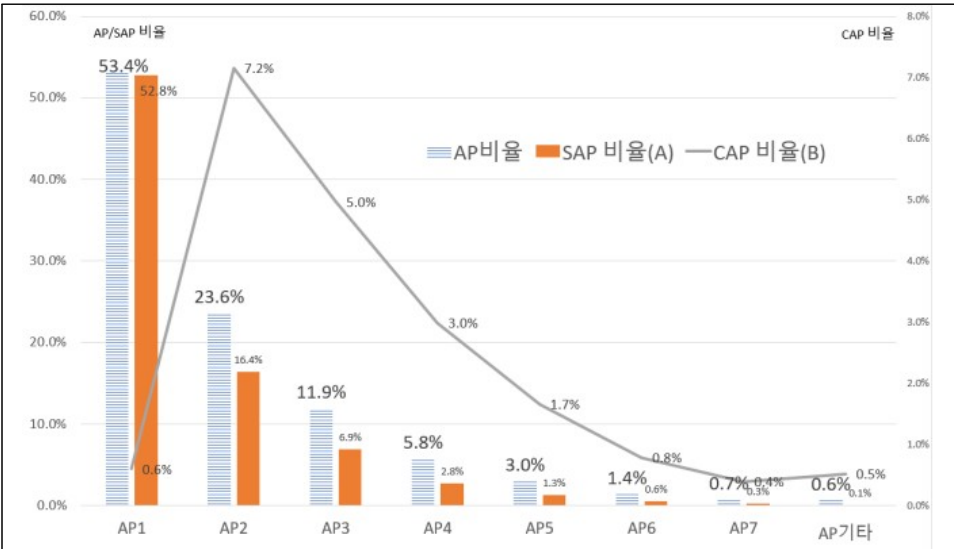
대구분	비율(A+B)	SAP 비율(A)	CAP 비율(B)	AP 유형 내 상대 비율	
				SAP : A/(A+B)	CAP: B/(A+B)
AP1	53.4%	52.8%	0.6%	99%	1%
AP2	23.6%	16.4%	7.2%	70%	30%
AP3	11.9%	6.9%	5.0%	58%	42%
AP4	5.8%	2.8%	3.0%	48%	52%
AP5	3.0%	1.3%	1.7%	43%	56%
AP6	1.4%	0.6%	0.8%	43%	57%
AP7	0.7%	0.3%	0.4%	43%	57%
AP 기타	0.6%	0.1%	0.5%	13%	87%
합	100%	81%	19%	-	-

주: 합은 소수점 첫째 자리에서 반올림한 수치임

자료: 저자 작성

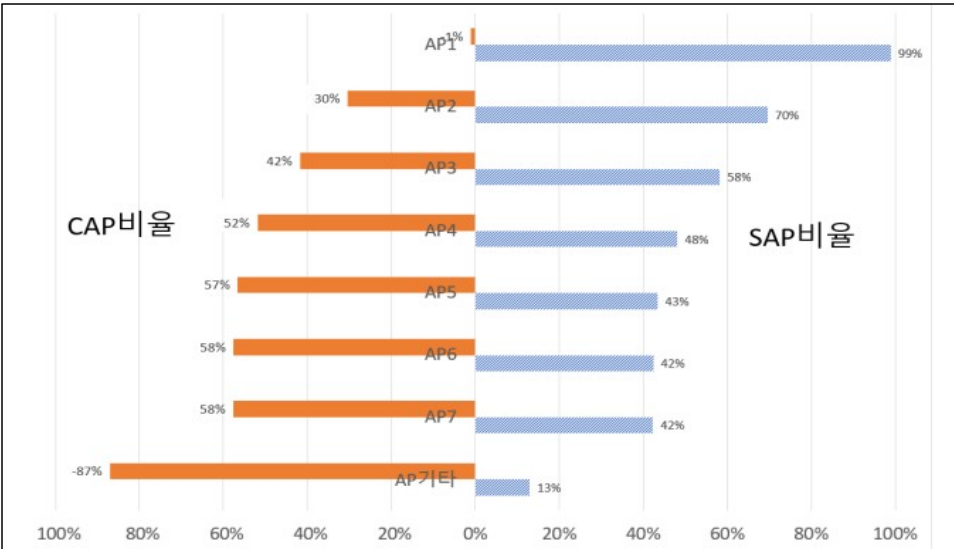
[표 4-6]을 활용하여 AP 번호 증가와 SAP, CAP 비율의 변화를 차트를 작성하여 검토하였다. 결과적으로 SAP는 목적지가 증가할수록 해당유형의 점유율이 줄어들었고, CAP 비율은 목적지 회수가 2곳인 경우 7.2%로 가장 높았고 이후 감소하는 것으로 나타났다. 이 중 CAP 2의 활동패턴은 여러 가지가 있을 수 있으며 예를 들어 집→직장→거래처→직장→집 등의 직장 중심의 활동이 있을 수 있으며, 집→마트→집→병원→집으로 집 중심의 활동이다. 즉, 복합 활동패턴은 단일 거점 중심의 활동패턴 비율이 가장 많은 것으로 나타났다. 동일 AP 유형 내에서 SAP, CAP 비율 차이 검토 결과, AP1, AP2, AP3까지는 해당유형 내의 SAP 점유율이 높았으나 AP4부터는 복합 활동패턴인 CAP의 점유율이 높은 것으로 나타났으며, 목적지가 4개 이상인 활동패턴의 경우 복합 활동패턴의 비율이 단순 활동패턴 비율보다 높다고 볼 수 있다. 복합 활동 비율은 AP2(집 이외 목적지 2곳)부터 비율이 증가하기 시작하였으며 목적지 횟수가 증가해 거점 중심의 활동이 증가한다고 볼 수 있다. 특히 개인의 하루 활동패턴을 분석한 결과 새로운 목적지를 방문하기보다는 특정지점으로의 회귀성 통행이 많다는 것을 보여주는 결과이다.

그림 4-9 | AP 유형별 AP/SAP/CAP 비율



자료: 저자 작성

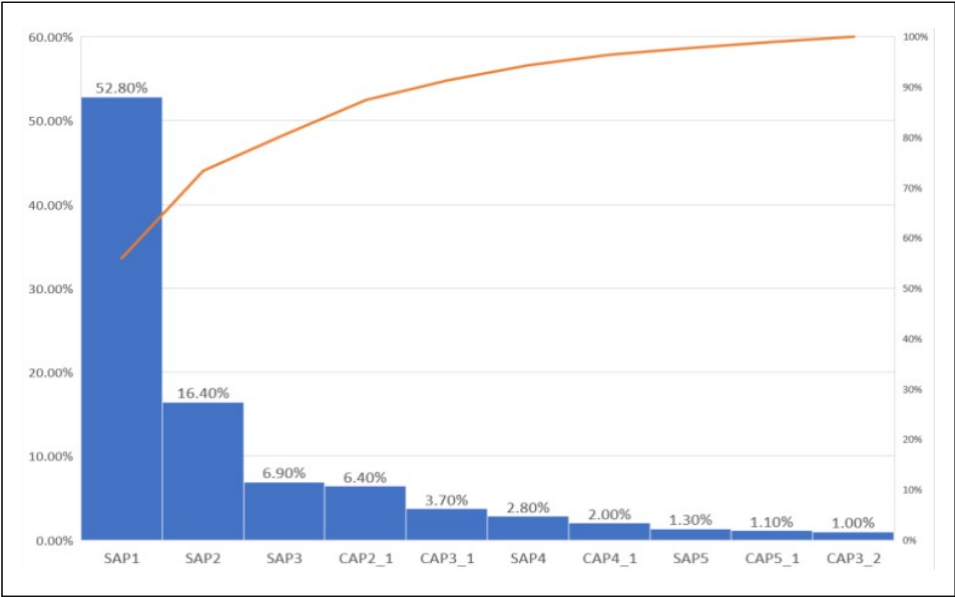
그림 4-10 | AP 유형별 SAP/CAP의 점유율 차이



자료: 저자 작성

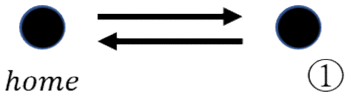
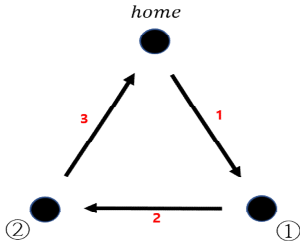
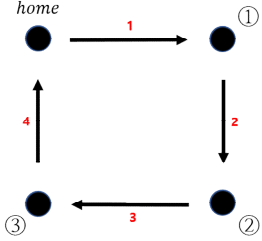
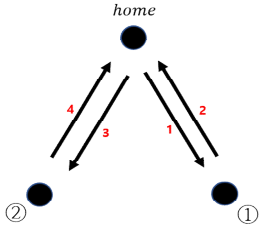
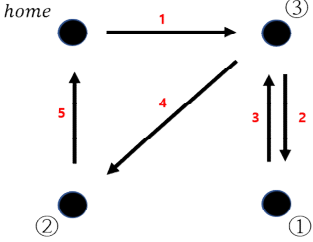
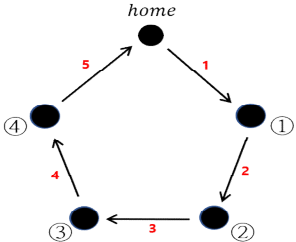
앞서 기술한 바와 같이 L/N/S 정보로 구분 가능한 활동패턴은 총 60개였으며 이 중 빈도수가 10 이하 등으로 낮은 일부 활동패턴을 노드 수를 기준으로 각 AP의 기타로 구분하고 AP 7개 유형에 포함되지 않은 활동은 기타로 구분하여 총 32개의 활동패턴을 도출하였다. 단, CAP 1_2의 경우 빈도수는 낮지만 집-직장 왕복 횟수가 높고 집-직장 기반 활동패턴임을 고려하여 별도로 구분하였다. 32개 활동패턴 중 상위 10개 활동패턴의 유형과 점유율 검토 결과, 상위 10개 활동의 점유율은 94.4%였고 이중 SAP는 5개, 점유율 80.2%, CAP 5개 점유율은 14.2%로 나타났다. 전체 32개 활동패턴 중 점유율이 높은 상위 6개의 누적 점유율은 89%로 6개 패턴이 10개 활동 중 약 9개를 설명할 수 있음을 의미한다. 즉, 개인 활동패턴은 다양하지만 6개 패턴으로 압축될 수 있음을 의미하는 결과이며, 상기 6개 중 CAP는 CAP2_1, CAP3_1 등 2개가 해당되어 활동패턴 파악을 위해서는 위치 정보를 이용한 CAP 검토도 중요하다고 볼 수 있다.

그림 4-11 | 활동패턴 상위 10개 비율



자료: 저자 작성

표 4-7 | 상위 활동패턴 6개

	
1순위(52.8%) : SAP_1	2순위(16.4%) : SAP_2
	
3순위(6.9%) : SAP_3	4순위(6.4%) : CAP_{2_1}
	
5순위(3.7%) : CAP_{3_1}	6순위(2.8%) : SAP_4

주: ①, ②는 집 이외의 목적지를 통행순서대로 기술한 것임
 자료: 저자 작성

활동패턴 유형 SAP 8개, CAP는 24개이며 각 유형별 빈도수와 점유율은 아래와 같다. 전체 유형 중 약 19%가 CAP라는 것은 활동패턴의 다양성을 설명하기 위해서는 비율은 낮지만 CAP 유형 구분이 중요한 의미를 가진다고 사료된다.

표 4-8 | 활동패턴 32개 유형별 비중

대구분	소구분	L	N	S	빈도수	비율
AP1 3개(1, 2)	SAP1	2	2	3	36,024	52.8%
	CAP1_1	4	2	5	383	0.6%
	CAP1_2	6	2	7	7	0.01%
AP2 5개(1, 4)	SAP2	3	3	4	11,182	16.4%
	CAP2_1	4	3	5	4,375	6.4%
	CAP2_2	5	3	6	385	0.59%
	CAP2_3	6	3	7	104	0.2%
	CAP2_기타	8	3	9	21	0.03%
AP3 5개(1, 4)	SAP3	4	4	5	4704	6.9%
	CAP3_1	5	4	6	2,531	3.7%
	CAP3_2	6	4	7	684	1.0%
	CAP3_3	7	4	8	121	0.2%
	CAP3_기타	9	4	10	38	0.1%
AP4 5개(1, 4)	SAP4	5	5	6	1,890	2.8%
	CAP4_1	6	5	7	1,393	2.0%
	CAP4_2	7	5	8	501	0.7%
	CAP4_3	8	5	9	102	0.2%
	CAP4_기타	10	5	11	43	0.1%
AP5 5개(1, 4)	SAP5	6	6	7	868	1.3%
	CAP5_1	7	6	8	727	1.1%
	CAP5_2	8	6	9	278	0.4%
	CAP5_3	9	6	10	95	0.1%
	CAP5_기타	12	6	13	31	0.1%
AP6 4개(1, 3)	SAP6	7	7	8	394	0.6%
	CAP6_1	8	7	9	324	0.5%
	CAP6_2	9	7	10	144	0.2%
	CAP6_기타	12	7	13	67	0.1%
AP7 3개(1, 3)	SAP7	8	8	9	196	0.3%
	CAP7_1	9	8	10	151	0.2%
	CAP7_기타	12	8	13	116	0.2%
AP 기타 2개(1, 1)	SAP_기타	12	13	13	53	0.1%
	CAP_기타	12	10	13	357	0.5%
합계					68,289	100%

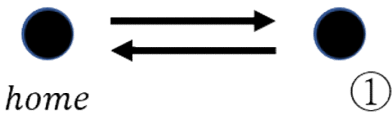
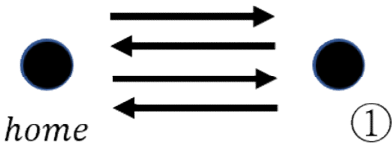
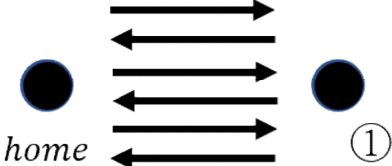
주: 1) 위 유형 중 기타 L/N/S는 해당 활동패턴 중 소구분으로 구분되지 않은 유형들의 각 L/N/S 평균치를 의미

2) 비율의 합은 소수점 첫째 자리에서 반올림한 수치임

자료: 저자 작성

본 연구는 [표 4-8]에서 파악된 L/N/S 정보를 이용하여 추정 가능한 활동패턴 개념도를 작성하였다. [표 4-9]는 SAP, CAP 활동패턴별 활동 형상을 개념도 형식으로 표현한 것으로, 그림에서 복합 활동패턴은 특정지점을 2회 이상 방문한 것임을 알 수 있다. 일례로, [표 4-9]에서 CAP_{1_2} 는 링크(L) 6개(왕복 3회), 노드(2개, 예: 집, 직장) 그리고 최초출발지 (집)을 포함한 도착지는 총 7개인 활동패턴이다.

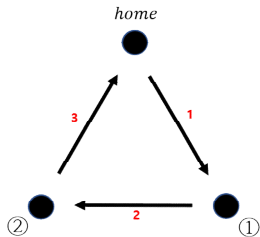
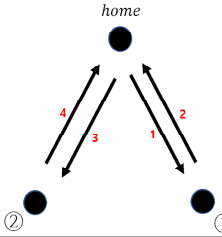
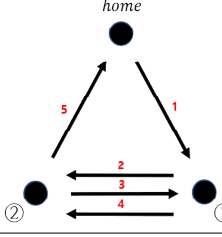
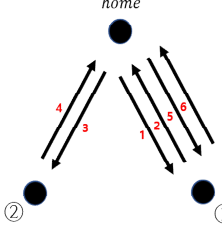
표 4-9 | AP1, SAP/CAP 활동패턴 개념도

대구분	소구분	활동패턴 그림	활동패턴 설명
AP ₁	SAP ₁		집-①-집 L:2 N:2 S:3
	CAP _{1_1}		집-①-집-①-집 L:4 N:2 S:5
	CAP _{1_2}		집-①-집-①-집-①-집 L:6 N:2 S:7

자료: 저자 작성

[표 4-10]에서 $CAP_{2,2}$ 는 링크(L) 5개, 노드 3개 (예: 집, 직장, 거래처) 그리고 하루 중 최초출발지인 집을 포함한 S가 6개인 활동패턴이다.

표 4-10 | AP2, SAP/CAP 활동패턴 개념도

대구분	소구분	활동패턴 그림	활동패턴 설명
AP_2	SAP_2		집-①-②-집 L:3 N:3 S:4
	$CAP_{2,1}$		집-①-집-②-집 L:4 N:3 S:5
	$CAP_{2,2}$		집-①-②-①-②-집 L:5 N:3 S:6
	$CAP_{2,3}$		집-①-집-②-집-①-집 L:6 N:3 S:7

주: 적색 번호는 통행순서/원 번호는 목적지
자료: 저자 작성

[표 4-11]에서 CAP_{3_1} 는 링크(L) 5개, 노드 4개(예: 집, 거래처, 직장, 마트), 최초출발지 집을 포함한 목적지가 6개인 활동패턴으로 집에서 거래를 경유하여 직장에 갔다가 다시 거래처 방문 후 마트 이용 후 귀가한 활동패턴이라고 추정할 수 있다. 이처럼 AP 증가 시 하나의 활동패턴에는 여러 가지 목적통행이 포함되며 복잡한 활동패턴이 발생한다.

표 4-11 | AP3, SAP/CAP 활동패턴 개념도

대구분	소구분	활동패턴 그림	활동패턴 설명
AP_3	SAP_3		집-①-②-③-집 L:4 N:4 S:5
	CAP_{3_1}		집-③-①-③-②-집 L:5 N:4 S:6
	CAP_{3_2}		집-③-①-③-②-③-집 L:6 N:4 S:7

주: 적색 번호는 통행순서/원 번호는 목적지
자료: 저자 작성

표 4-12 | AP4, SAP/CAP 활동패턴 개념도

대구분	소구분	활동패턴 그림	활동패턴 설명
AP ₄	SAP ₄		집-①-②-③-④-집 L:5 N:5 S:6
	CAP _{4_1}		집-①-②-①-③-④-집 L:6 N:5 S:7
	CAP _{4_2}		집-①-②-①- ②-③-④-집 L:7 N:5 S:8
	CAP _{4_3}		집-①-집-①-집- ②-③-④-집 L:8 N:5 S:9

자료: 저자 작성

표 4-13 | AP5, SAP/CAP 활동패턴 개념도

대구분	소구분	활동패턴 그림	활동패턴 설명
AP ₅	SAP ₅		집-①-②-③-④-⑤-집 L:6 N:6 S:7
	CAP _{5_1}		집-①-②-①- ③-④-⑤-집 L:7 N:6 S:8
	CAP _{5_2}		집-①-②-①-②- ③-④-⑤-집 L:8 N:6 S:9
	CAP _{5_3}		집-①-②-①-②- ③-④-⑤-④-집 L:9 N:6 S:10

자료: 저자 작성

표 4-14 | AP6와 AP7, SAP/CAP 활동패턴 개념도

대구분	소구분	활동패턴 그림	활동패턴 설명
AP_6	SAP_6		집-①-②-③- ④-⑤-⑥-집 L:7 N:7 S:8
	CAP_{6_1}		집-①-②-①-③- ④-⑤-⑥-집 L:8 N:7 S:9
	CAP_{6_2}		집-①-②-①-③- ④-⑥-⑤-⑥-집 L:9 N:7 S:10
AP_7	SAP_7		집-①-②-③- ④-⑤-⑥-⑦-집 L:8 N:8 S:9
	CAP_{7_1}		집-①-②-①-②-③- ④-⑤-⑥-⑦-집 L:9 N:8 S:10

자료: 저자 작성

3. 활동패턴의 지역별 요일별 분석

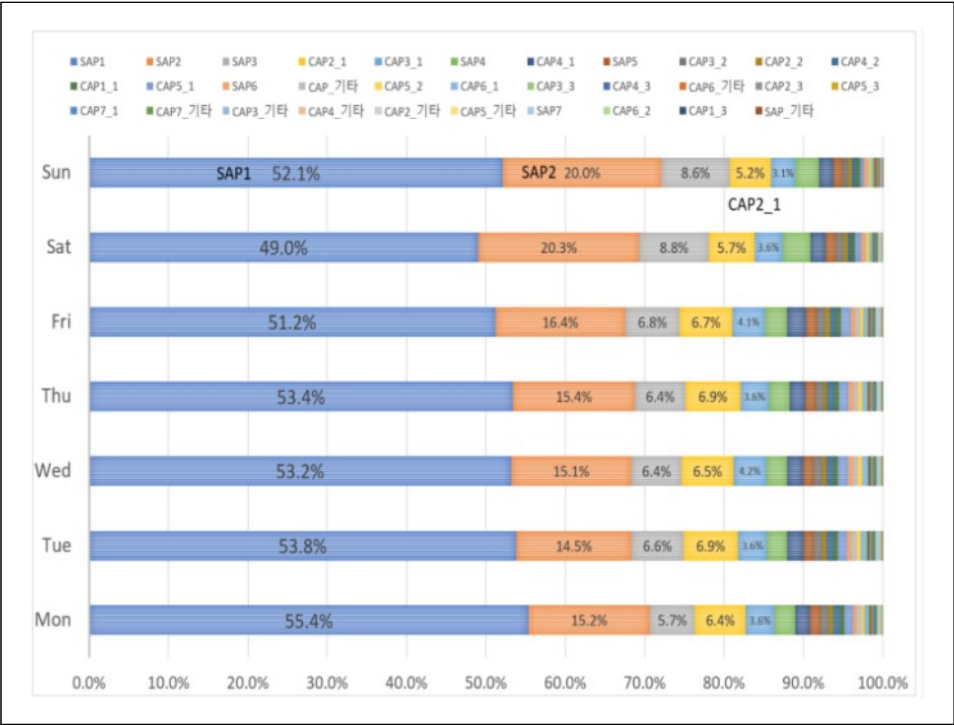
본 연구는 활동패턴 대시보드 가공작업에서 도출된 L/N/S 정보로 개인의 하루 단위 활동패턴을 상기에서 기술한 32개 유형과 매칭하는 작업을 수행하였으며, 개인의 활동패턴 자료와 활동패턴 세부 통행자료를 이용하여 활동 시간, 체류시간, 통행시간을 산정하였다. 활동패턴 구분 전 자료의 각 통행 단위의 이동시간과 체류시간 데이터를 이용하여 집에서 출발한 시점부터 귀가 시간까지를 총활동 시간으로 정의하여 산정하였다. 또한, 각 출발과 도착시간으로 통행시간을, 현 도착지 도착시간과 다음 통행의 출발시간을 이용하여 체류시간을 산정하였다. 여기서 기술하는 체류시간은 하루 중 집으로 최종 귀가 전 방문한 장소에서 머무른 시간의 합이며 귀가 후 집에서 머무는 시간은 포함되지 않고, 통행시간은 하루 중 집 귀가를 포함한 목적지 이동에 걸리는 시간의 합이다. 이러한 개인의 활동은 다양하며 그 다양한 활동을 대표하는 활동 시간을 산정하는 작업은 여러 가지 방법이 있을 수 있다. 본 연구는 앞서 구분한 활동패턴 유형 32개의 총 데이터에서 차지하는 비율을 가중치로 활동 시간에 대한 가중평균값을 산정하였다.

집과 직장만을 오가는 활동패턴의 총활동에서 차지하는 비율을 고려하여 활동 시간을 산정함으로써 좀 더 현실적인 활동 시간을 산정할 수 있었으며, 일반 조사에서는 대부분 빈번한 활동의 활동 시간 등만을 조사하기 때문에 본 연구에서 제시한 복합 활동의 존재 여부는 물론 해당 활동의 활동 시간에 대한 데이터 수집이 어려울 수 있는 부분이 있다. 활동 시간 분석은 먼저 활동패턴 유형 32개의 가중평균값을 산정한 후 SAP, CAP 활동 시간을 검토하였다.

1) 활동패턴의 요일별 점유율

활동패턴 32개의 각 점유율을 요일별로 살펴본 결과 하루 전체 활동패턴 중 SAP1이 차지하는 비율은 49%~55.4%로 나타났고 SAP2의 점유율은 14.5%~20.3% 사이에 분포하는 것으로 나타났다. 요일별로 높은 활동패턴은 전체 순위와 유사하게 SAP1, SAP2가 모든 요일에서 가장 점유율이 높았고 SAP1, SAP2, SAP3인 주요 단순 활동패턴의 주중 점유율은 75% 이상이었고 주말은 이보다 약 4%p 높은 79%로 나타났다.

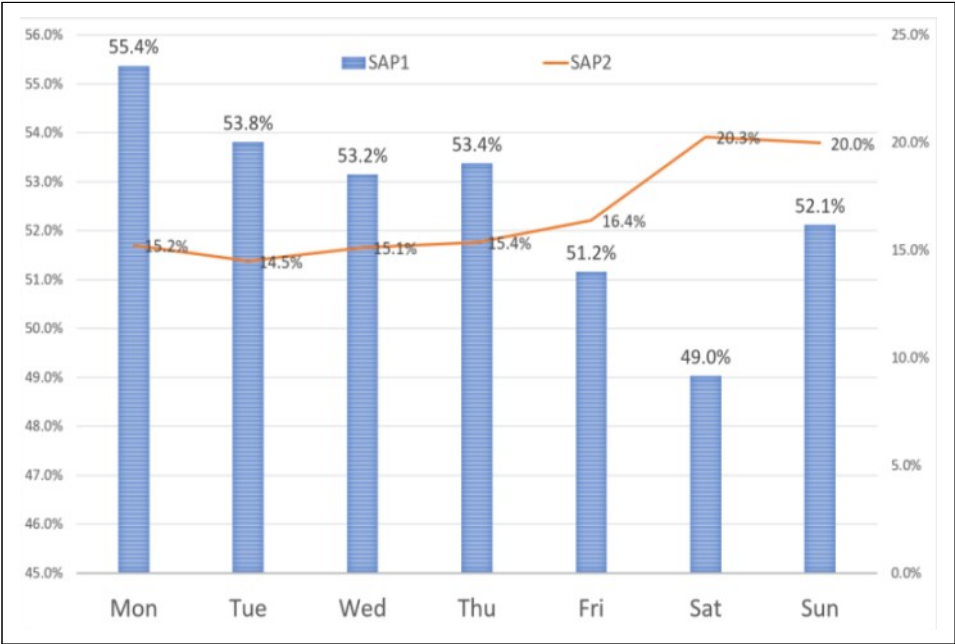
그림 4-12 | 활동패턴별 주중과 주말 총활동 시간(분) 분포도



자료: 저자 작성

점유율이 1, 2위인 단순 활동패턴인 SAP1, SAP2의 요일별 점유율을 살펴본 결과, SAP1은 월요일이 55.4% 가장 높고 이후 감소하였고 금요일은 주중 가장 낮은 51.2%를 차지하는 것으로 나타났다. 그 중 사무직의 경우 월요일에 내부 회의 일정이 많아 외부 출장보다는 내부 업무에 집중하기 때문에 이러한 현상이 나타나는 것이라고 보여졌다. 금요일의 경우 SAP1의 비율은 51.2%로 주중 가장 낮았지만 SAP2의 비율은 주중 가장 높게 나타났으며, 금요일은 휴일 전일로 유연근무 등의 경우 직장에서 일찍 나와 기타 활동을 하는 것으로 보인다. SAP2의 점유율은 월요일 15.2%에서 목요일까지 크게 변화가 없었으나 금요일 16.4%, 토요일 20.3%, 일요일 20.0%로 주말이 주중보다 높았다. SAP2 활동패턴이 주중보다 주말에 더 많이 발생하는 활동패턴이라고 볼 수 있으며 집에서 나와 쇼핑, 외식 등의 연계 활동이 많은 주말 활동 특성도 하나의 이유가 될 것으로 보여진다.

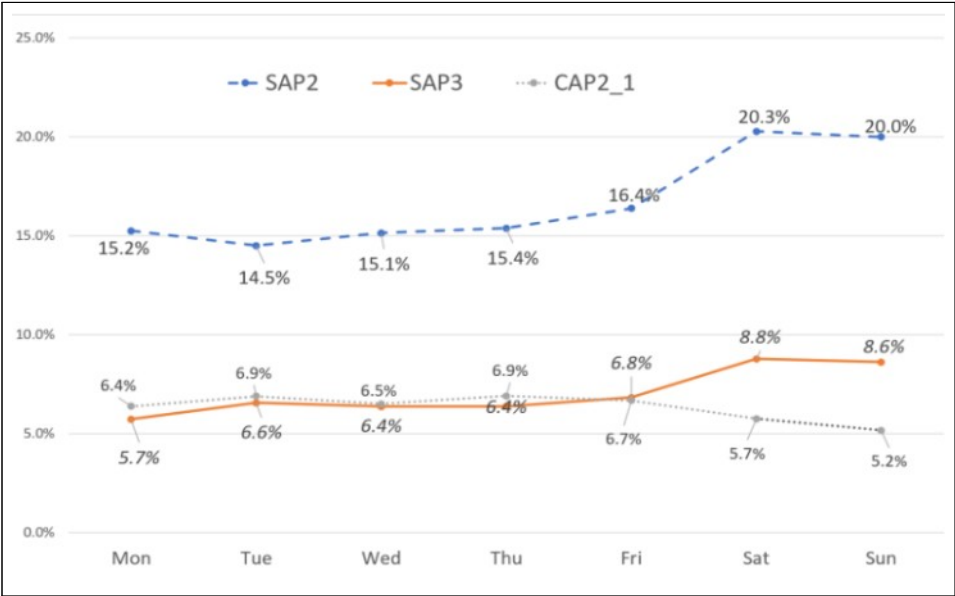
그림 4-13 | 활동패턴 SAP1, SAP2의 요일별 점유율



자료: 저자 작성

앞서 검토한 요일별 활동패턴의 점유율 검토 결과, SAP2/SAP3/CAP2_1의 주중과 주말 점유율 경향이 상이해 추가비교 하였다. SAP3 점유율은 화요일 6.6%에서 금요일 6.8%로 변화가 적었으나 토요일 8.8%, 일요일 8.6%로 주말 점유율이 주중 점유율보다 높았으며, SAP2의 점유율은 주중 약 15.3%, 주말 약 20.2%로 주말이 약 4.9%p나 높았다. 반면, CAP2_1의 주중 점유율은 약 6.7%, 주말 점유율은 약 5.5%로 주중이 주말보다 1.2%p 높았다. 일반적으로 주말의 활동량이 주중보다 적은 것을 고려하면 SAP2, SAP3 활동패턴은 주중 발생량이 높은 활동패턴이라고 할 수 있으며, SAP2, 3의 주말 활동패턴의 경우 외식만을 하거나, 쇼핑만을 하기보다는 2개 이상의 활동을 연속적으로 한다고 볼 수 있는 결과이다. 반면, 복합 활동패턴인 CAP2_1(예: 집-직장-집-직장-집)가 주말에 적다는 것은 주말에는 중간에 귀가했다가 다시 나가는 패턴은 낮고 회귀성이 없는 활동패턴이 증가한다고도 볼 수 있는 결과이다.

그림 4-14 | 활동패턴 SAP1, SAP2의 요일별 점유율

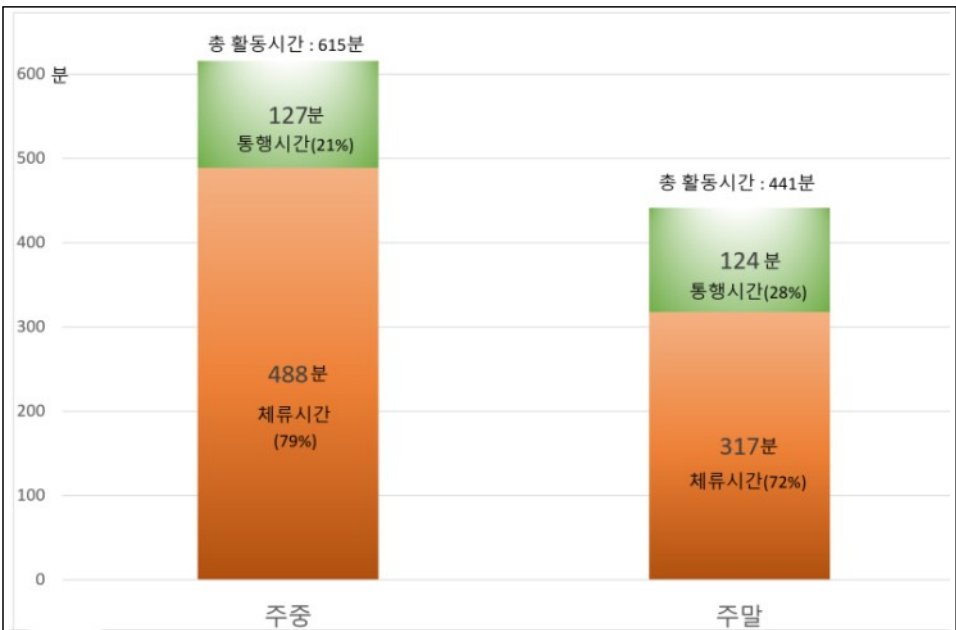


자료: 저자 작성

2) 전체 활동패턴의 활동 시간 분석

활동패턴 32개의 각 점유율을 가중치로 주중과 주말의 활동 시간, 체류시간, 통행시간을 검토하였다. 검토 결과, 하루를 기준으로 주중 활동 시간(집 귀가 후 시간 제외)은 615분(약 10.3시간)으로 나타났고 이중 통행시간에 127분(약 2.1시간), 체류에 488분(약 8.2시간)을 소비하는 것으로 나타났다. 주말의 총활동 시간은 441분(약 7.4시간)으로 나타났고 이중 통행시간에 124분(약 2.1시간), 체류에 317분(약 5.3시간)을 소비하는 것으로 나타났으며, 주중과 주말의 총활동 시간 중 체류시간과 통행시간이 차지하는 비율은 주중이 79:21, 주말이 72:28로 나타났다. 이것은 주중 집 밖에서의 100분 활동을 위해 평균 21분을 이동에 투자하고 있는 것으로도 해석할 수 있다.

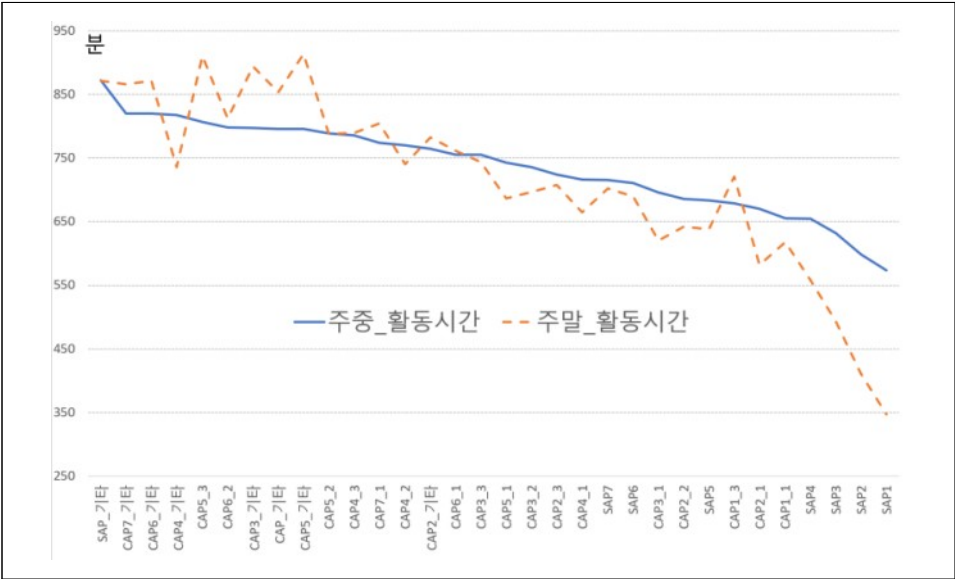
그림 4-15 | 전체 활동패턴의 가중 평균 주중 및 주말 체류 및 통행시간



자료: 저자 작성

활동패턴별 활동 시간의 분포는 주중 573분~873분 사이에 분포하고 있었으며 주말은 347분~872분 사이에 분포하고 있었다. 복합 활동패턴 중 목적지 회수가 많은 복합 활동패턴에서 주말의 활동 시간이 주중보다 높은 것으로 나타났지만, 복합 활동패턴 중 목적지 회수가 낮은 해당 활동패턴에서는 주중이 높아지는 역전 현상을 보이고 있었다.

그림 4-16 | 활동패턴별 주중과 주말 총활동 시간 분포도

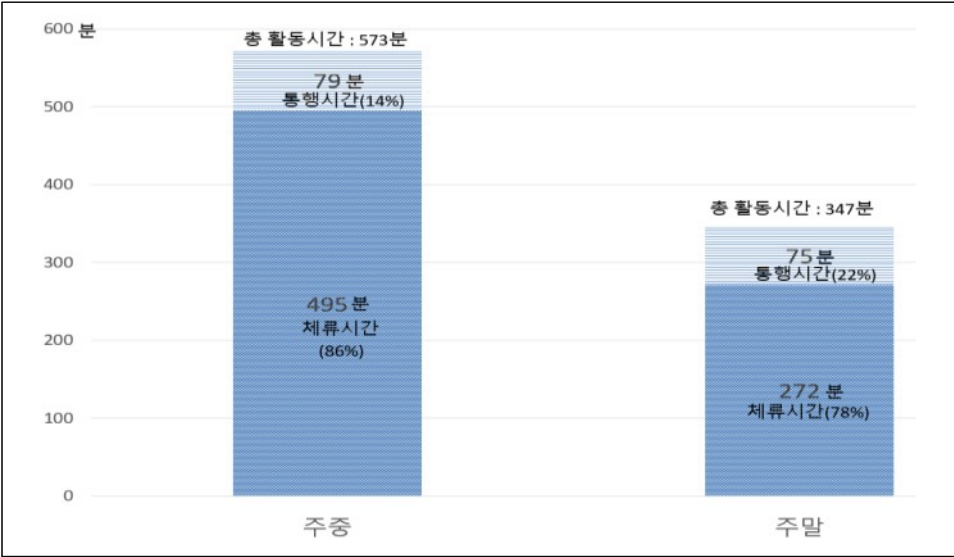


자료: 저자 작성

3) SAP별 활동 시간 등 분석

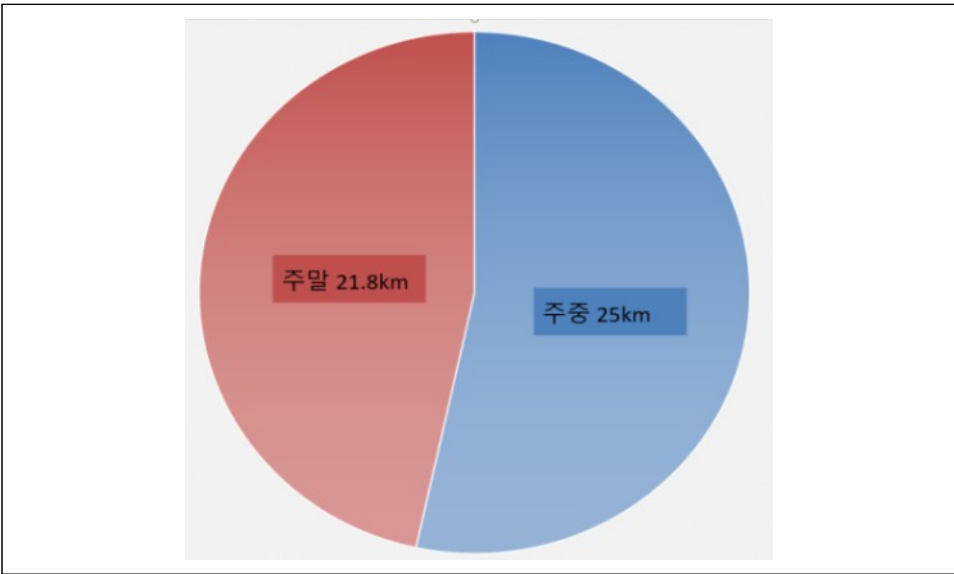
활동패턴 32개 유형 중 점유율이 가장 높은 SAP1(목적지 1곳)을 대상으로 총활동 시간, 체류시간, 통행시간의 평균을 주중과 주말로 구분하여 검토하였다. SAP1 특성상 목적지는 1개로 주중에는 집에서 직장으로의 출퇴근, 주말에는 쇼핑, 친지 방문 등이 주 활동 목적으로 추정되었다. 주중과 주말의 SAP1 평균 총활동 시간 검토 결과, 주중의 활동 시간(집 출발-집 도착)이 주말보다 3.8시간(39.5%) 높게 나타났으며, SAP1 주중 체류시간은 주말보다 223분 높았고 통행시간은 4분 정도 높아 주중과 주말 활동 시간의 주된 차이는 체류시간에 기인하는 것으로 나타났다. SAP1의 주중 총활동 시간(집 출발-집 귀가)은 약 9.5시간, 통행시간(양방향) 1.3시간, 직장 등에서의 체류시간은 8.2시간으로 체류시간 점유율이 통행시간보다 6.3배 높았다. 이에 비해 주말의 총활동 시간(집 출발-집 귀가)은 약 5.8시간, 통행시간(양방향) 1.3시간 (22%), 친지, 놀이공원, 쇼핑 등의 체류시간은 4.5시간(78%)으로 나타났다. 이는 본 연구의 체류시간은 집에서 머무는 시간을 제외한 시간으로, 하루 중 주말에 집에서 머무는 시간이 SAP1의 경우 주중보다 3.8시간 더 길다고도 볼 수 있는 결과이다. 또한 주중과 주말의 SAP1의 총이동 거리는 주중 25km, 주말 21.8km로 주중이 주말보다 약 3.2km 통행 거리가 길었다. 이는 집과 직장을 오가는 활동이 많은 SAP1 활동패턴에 서는 주중 여가 통행 거리보다 출퇴근 이동 거리가 길다고도 볼 수 있는 결과이다.

그림 4-17 | SAP1의 주중 및 주말 체류 및 통행시간



자료: 저자 작성

그림 4-18 | SAP1의 주중 및 주말 총통행 거리 비교

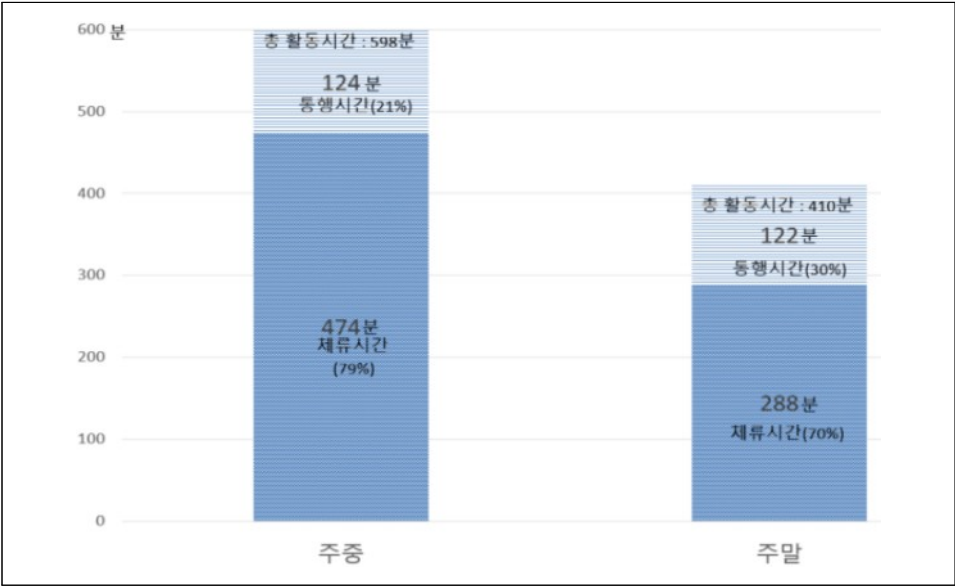


주: 상기 거리는 출발자-도착지 직선거리로 실제 거리보다 적음

자료: 저자 작성

활동패턴 유형 중 점유율이 두 번째로 높은 SAP2(목적지 2곳)를 대상으로 총활동 시간, 체류시간, 통행시간을 주중과 주말로 구분하여 검토하였다. SAP2의 목적지는 주중 직장/거래처, 주말에는 쇼핑 후 친지 방문, 관광 후 식사 등이 주목적이 될 것으로 추정되었다. 주중과 주말의 SAP2 활동 시간은 주중이 주말보다 활동 시간이 3.1시간(31.5%) 높았으며 SAP2 주중 체류시간은 주말보다 188분 높았고 통행시간은 2분 정도 높아 SAP1과 유사하게 주중과 주말의 활동 시간 차이는 체류시간에 기인하였다. SAP2의 주중 총활동 시간은 10시간, 통행시간(양방향) 2.1시간(21%), 체류시간은 7.9시간(79%)으로 체류시간 점유율이 통행시간보다 3.8배 높았으며 주말의 총활동 시간은 6.8시간, 통행시간(양방향) 2.0시간 (30%), 체류시간은 3.1시간(70%)으로 나타났다. SAP2 총활동 시간의 체류/이동시간 배분은 주중 8:2, 주말 7:3으로 주말의 통행시간 비중이 조금 높았다.

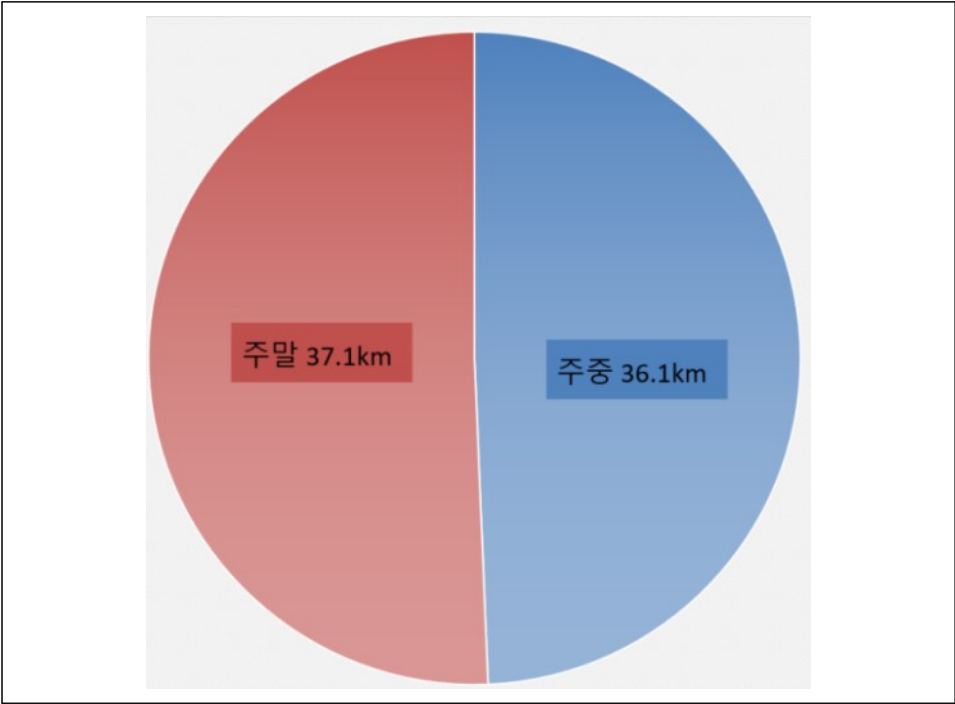
그림 4-19 | SAP2의 주중 및 주말 체류 및 통행시간



자료: 저자 작성

주중과 주말의 SAP2의 총이동 거리는 주중 36.1km, 주말 37.1km로 유사한 것으로 나타났다. SAP2의 활동 거리가 긴 것은 하루 중 방문지가 1곳에서 2곳으로 증가하였기 때문으로 목적지 횟수 증가에 비례해 총통행 거리도 증가한 것으로 나타났으며, 주말 통행의 경우 목적지 1회 증가로 총통행 거리는 15.3km(21.8—37.1) 증가하였고 주중은 11km 증가하였다. SAP1과 비교해 SAP2 주말 통행 거리 증가량이 주중에 비해 높았고 그 이유는 주말의 여가 통행 시 맛집, 관광명소 등의 방문에 기인한다고 추정되며 이는 거리에 구애받지 않고 근처 명소를 찾아가는 관광행태 등도 일부 원인이 될 수 있을 것이다.

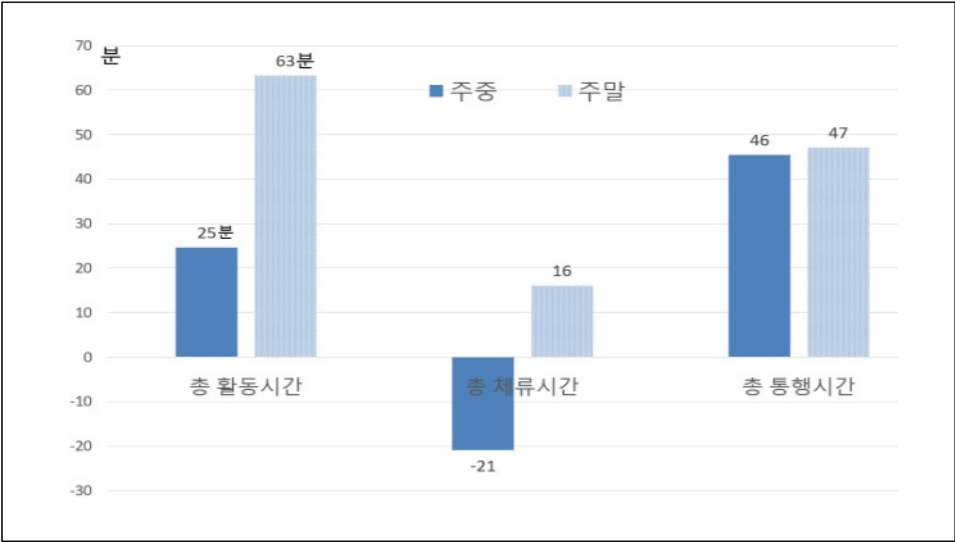
그림 4-20 | SAP2의 주중 및 주말 총통행 거리 비교



주: 상기 거리는 출발지-도착지 직선거리로 실제 거리보다 적음
자료: 저자 작성

주중과 주말의 SAP2와 SAP1 활동 시간의 차이(SAP2-SAP1)를 총활동 시간, 총 체류시간, 총 통행시간으로 구분하여 검토하였다. 검토 결과, 총활동 시간은 주중과 주말 모두 SAP2가 SAP1보다 길었고 그 정도는 주중 25분, 주말 63분이었으며, SAP2의 목적지가 SAP1보다 1개 많기 때문이며 그 차이는 주말이 주중에 비해 약 2.5 배 높았다. SAP2의 주중과 주말 총활동 시간 차이의 주된 이유는 체류시간 차이로 SAP2의 주중 체류시간은 SAP1보다 21분 적었으나 주말은 오히려 16분 길었고, SAP2의 주중 2곳에 머무는 시간은 SAP1의 주중 1곳 체류시간보다 적었으며 이는 주 중에는 정해진 근무 시간에 활동을 끝마치기 위한 노력의 결과라고도 사료된다. 주중에는 대개 업무적 체류, 주말에는 여가를 위한 체류가 대부분이라고 한다면 가보고 싶 었던 관광지, 쇼핑 등으로 체류시간이 증가한다고 볼 수 있으며, SAP2의 주중 및 주말 통행시간 증가치는 유사하나 체류시간에 차이가 있는 것은 해당 주말 활동이 대부분 여가, 관광지에서의 활동이기 때문으로 사료된다.

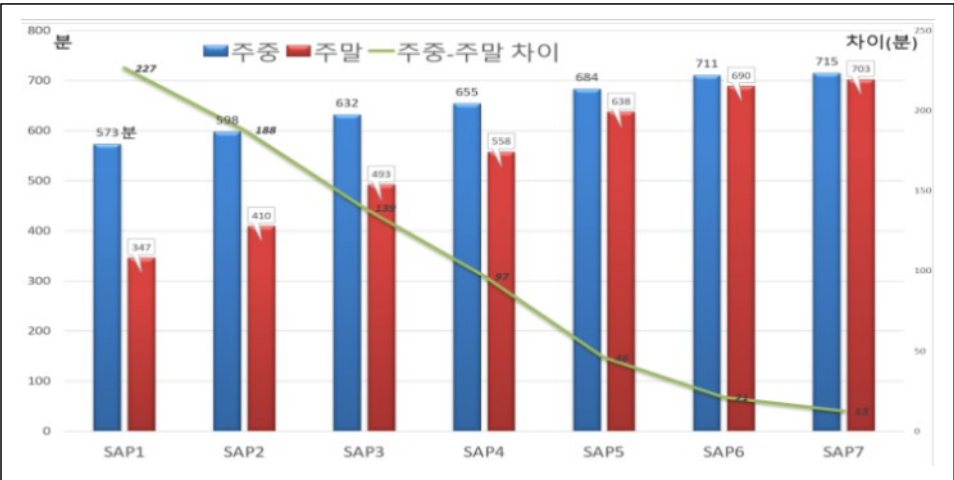
그림 4-21 | SAP2와 SAP1의 활동 시간 차이(SAP2-SAP1)



주: 상기 시간은 SAP2에서 SAP1의 시간을 차감한 결과임
 자료: 저자 작성

또한 여기서는 목적지 횟수가 증가할 경우 총활동 시간이 증가하는 경향이 있는지를 검토하였다. 총활동 시간은 주중과 주말 모두 SAP 유형 번호가 증가할수록 총활동 시간은 모두 증가하는 것으로 나타났다. SAP 유형 번호가 증가한다는 것은 앞서 기술한 것처럼 목적지 회수가 각 1회씩 증가한다는 의미로 추가적인 통행시간, 체류시간이 필요하기 때문에 총활동 시간이 증가하는 경향이 있었다. 총활동 시간의 증가추세는 목적지가 6회 이상인 SAP6 이후 둔화되었으며 주중과 주말의 SAP6/SAP7의 총활동 시간 평균은 약 705분(11.8시간)으로 나타났다. 증가추세가 둔화된다는 것은 주중 총활동 시간이 해당유형 내에서 특정 시간에 수렴한다 볼 수 있으며 본 연구에서는 약 713분(11.9시간)으로 나타났으며, 위 수렴치의 주말 값은 약 697분(11.6시간)으로 주중보다 약 16분이 낮았다. 특이한 점은 SAP 유형 번호가 증가(목적지 증가)할수록 해당유형 내 주중과 주말의 총활동 시간 차이는 줄어든다는 점이며 SAP1 주중, 주말의 총활동 시간 차이는 227분이었으나 SAP7의 해당 시간 차이는 13분으로 급격히 그 차이가 줄어들었다. 이러한 결과를 통해 SAP의 총활동 시간의 수렴치는 주중과 주말에 큰 차이가 없었다.

그림 4-22 | 주중과 주말 SAP별 총활동 시간과 차이

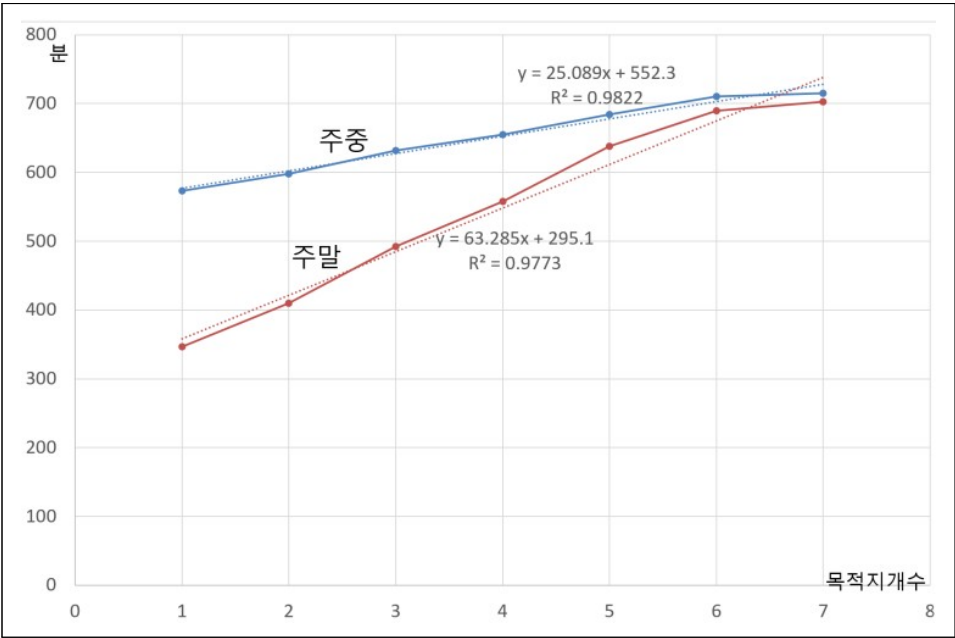


주: 상기 시간 차이는 주중에서 주말을 차감한 결과임

자료: 저자 작성

주중과 주말의 SAP 유형별 총활동 시간 변화를 살펴보기 위하여 목적지 개수별로 총활동 시간을 검토하였다. SAP 유형 번호는 목적지 개수이므로 유형 번호를 목적지 개수(X) 축에 표시하고 Y축에 해당 목적지 개수에 따른 총활동 시간을 산포도로 그리고 단순 추세식을 도출하였다. 여기서 단순 추세식을 이용한 것은 목적지 개수 증가에 따른 평균 활동 시간의 증가치를 파악하기 위해서이다. 검토 결과, 주중에는 목적지 1회 증가 시 활동 시간이 약 25.1분 증가하였다. 주말에는 약 63.3분 증가해 주말의 증가치가 주중보다 높았으며, 이러한 증가치는 목적지 6회 이상에서는 급격히 감소하는 것으로 나타났다 주중, 주말 모두 특정 값에 수렴하는 현상을 보여주었다. 이러한 결과는 주말 활동은 관광지, 쇼핑, 친지 방문, 외식 등 다양한 목적이 있고 해당 목적 달성을 위해 개인들은 주중보다 더 많은 시간을 투자한다는 것을 의미한다고 볼 수 있다.

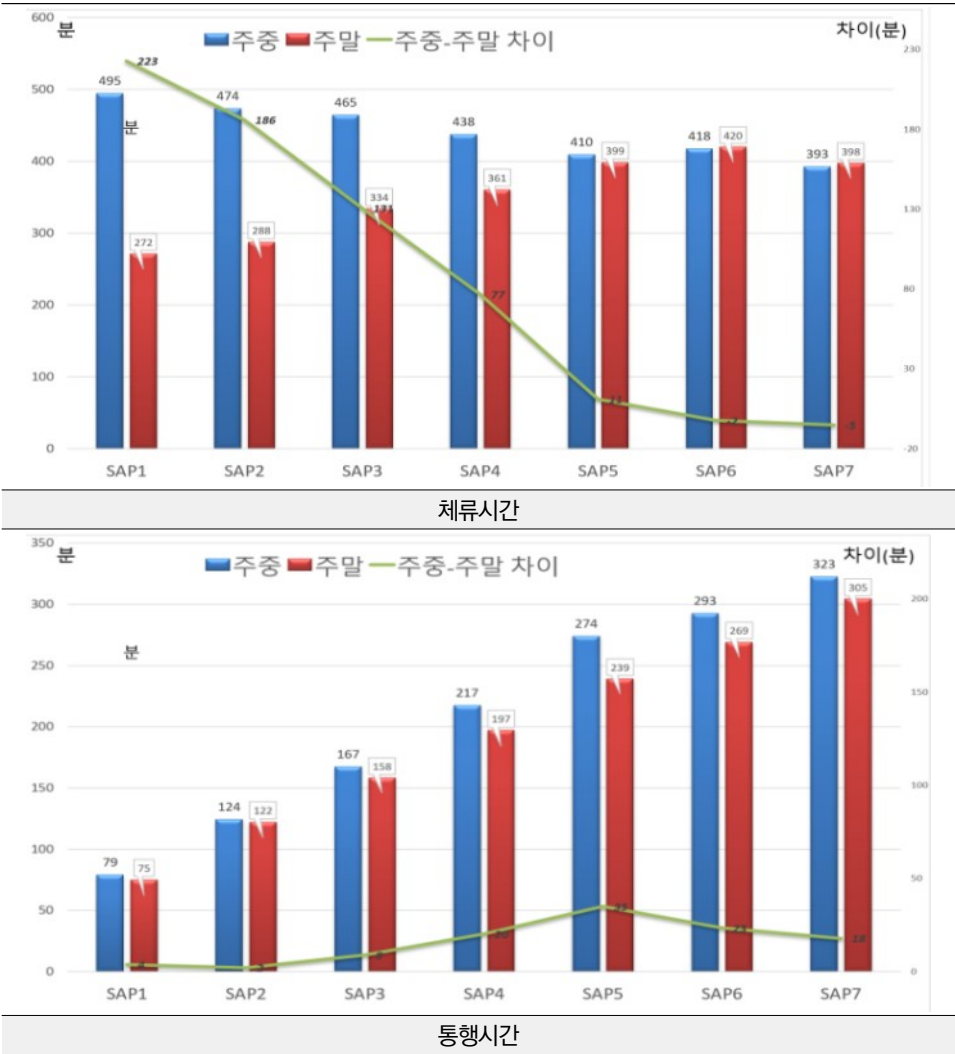
그림 4-23 | SAP별(목적지 개수) 주중과 주말 총활동 시간



주: 상기 시간 차이는 주중에서 주말을 차감한 결과임
 자료: 저자 작성

SAP 유형별 총활동 시간 차이는 체류시간에 기인하는 것으로 나타났으며, [그림 4-24]처럼 SAP 유형별 체류시간의 최대 차이는 223분, 통행시간 차이는 35분으로 나타났다.

그림 4-24 | 주중과 주말 SAP별 체류시간과 통행시간 차이



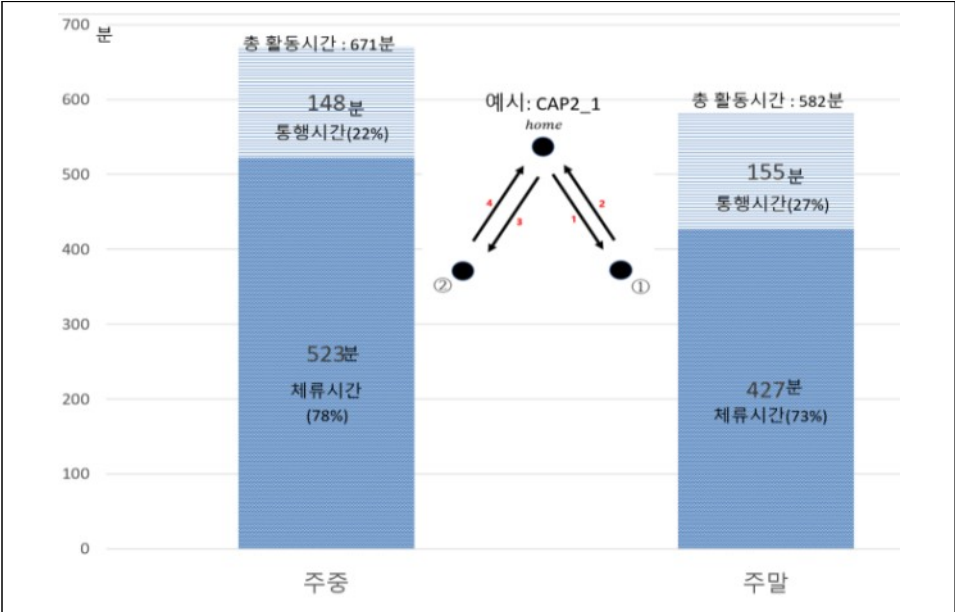
자료: 저자 작성

4) CAP별 활동 시간 등 분석

여기서는 앞서 기술한 SAP(단순 활동패턴)처럼 CAP(복합 활동패턴)에 대해서도 총활동 시간, 체류시간, 통행시간을 산정하여 CAP 패턴별로 검토하였다.

먼저, CAP(복합 활동패턴) 중 점유율이 높은 CAP2_1에 대해 검토하였는데, CAP2_1은 AP2 형태로 삼각형 활동패턴으로 집 포함 목적지가 3개이며 이중 특정지점으로 2회 회귀하는 활동패턴이다. 특히 집-직장-집-병원-집, 집-직장-거래처-직장-집 등의 활동패턴이 하나의 사례가 될 수 있다. CAP2_1의 주중과 주말 평균 총활동 시간 검토 결과, 주중이 주말보다 15.2%(89분) 높았고 그 차이는 SAP2(31.5%)보다는 낮았다, CAP2_1의 주중 체류시간은 주말보다 96분 높았고 통행시간은 7분 정도 낮았으며, CAP2_1의 총활동 시간 중 체류시간과 통행시간의 점유율은 주중 8:2, 주말 7:3 정도로 나타났다.

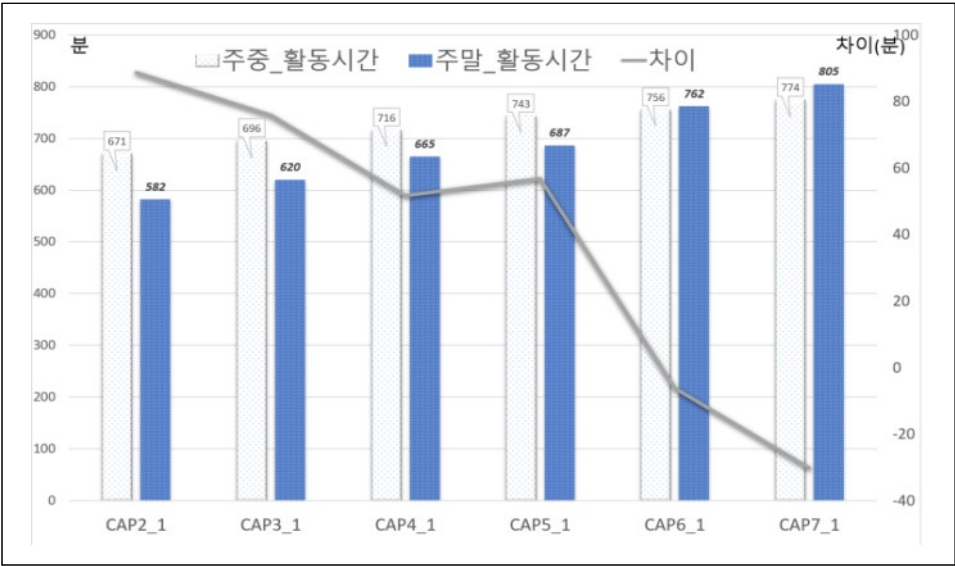
그림 4-25 | CAP2_1의 주중/주말 체류 및 통행시간



자료: 저자 작성

CAP(복합 활동패턴)의 목적지가 증가 시 활동 시간의 변화를 살펴본 결과 [그림 4-26]처럼 목적지 개수가 상대적으로 낮은 CAP2_1, CAP3_1, CAP4_1, CAP5_1에서는 주중 활동 시간이 높았으나 목적지 개수가 높은 CAP6_1, CAP7_1에서는 주말의 활동 시간이 주중보다 높은 것으로 나타났다. 특히 CAP(복합 활동패턴)의 번호가 증가할수록(목적지 증가) 주중과 주말의 차이는 감소하는 것으로 나타났다.

그림 4-26 | 주중과 주말 CAP별 총활동 시간과 차이

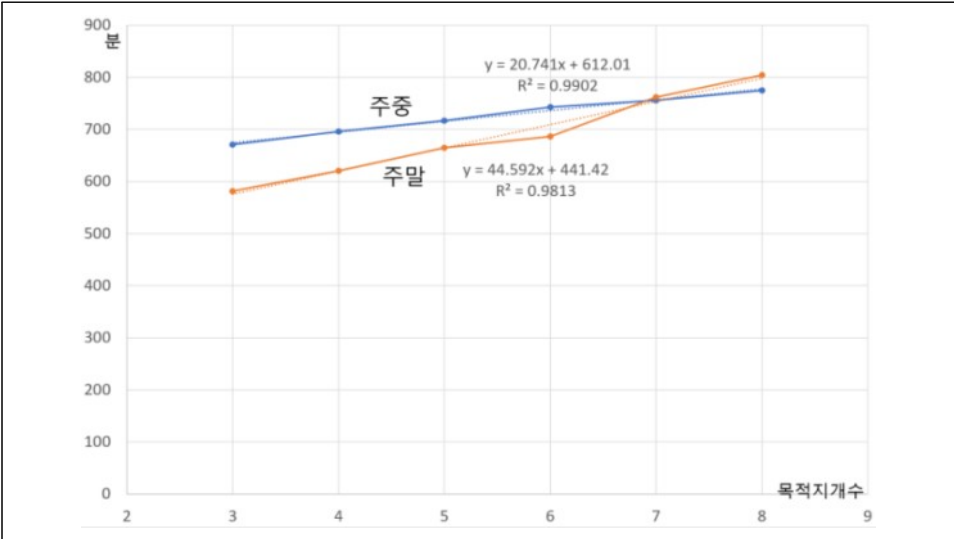


주: 1) 상기 시간 차이는 주중에서 주말을 차감한 결과임
 2) 차트의 우측 축의 범위는 -40분-80분, 좌측 축의 범위는 0분-900분
 자료: 저자 작성

또한 목적지 증가에 따른 주중과 주말의 CAP 활동패턴의 총활동 시간을 검토하였다. CAP 유형의 목적지 개수(X) 축에 표시하고 Y축에 해당 목적지 개수에 따른 총활동 시간을 산포도로 그리고 단순 추세식을 도출하였다.

검토 결과, 주중에는 목적지 1회 증가 시 활동 시간이 약 20.7분 증가하였고, 주말에는 44.6분 증가해 주말의 증가치가 주중보다 높았으며, 해당 증가치는 목적지 7회 이상에서는 주말의 활동 시간이 주중보다 높은 것으로 나타났다.

그림 4-27 | CAP별(목적지 증가) 주중과 주말 총활동 시간 차이



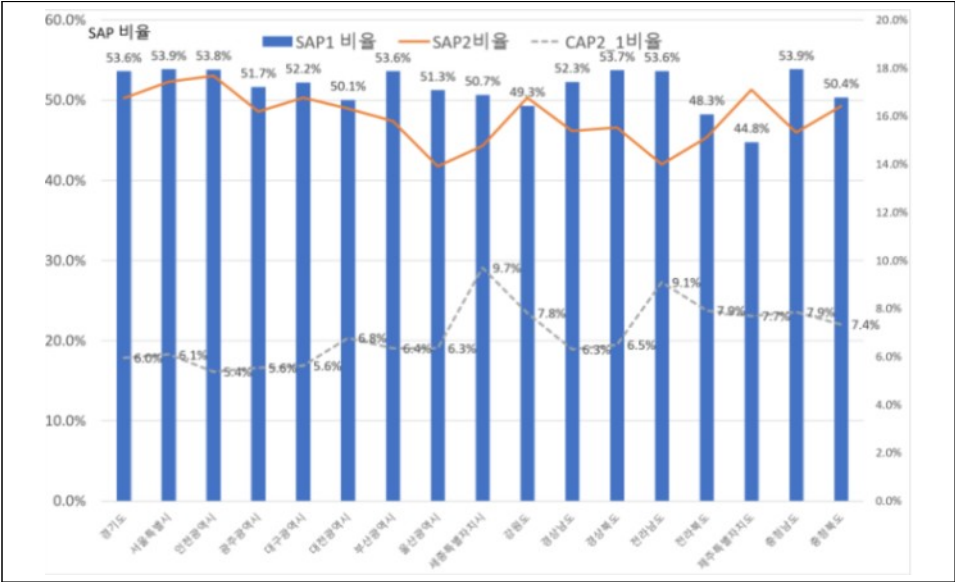
주: 1) 상기 시간 차이는 주중에서 주말을 차감한 결과임
2) 차트의 우측 축의 범위는 -40분-80분, 좌측 축의 범위는 0분-900분
자료: 저자 작성

목적지 증가에 따른 활동 시간 변화를 앞서 검토한 SAP와 연계하여 기술하면 주중 목적지 1회 증가 시 활동 시간 증가는 SAP가 CAP보다 주중과 주말 모두 더 큰 것으로 나타났다. 목적지 1회 증가 시 주중 SAP 활동 시간은 25분씩 증가한 반면 CAP는 21분 증가하였으며, 주말은 목적지 1회 증가에 따른 SAP 활동 시간은 63분씩 증가한 반면 CAP는 45분씩 증가하였다. 목적지 증가에 따른 CAP의 활동 시간 증가가 SAP보다 적은 것은 동일 목적지를 2회 이상 방문하기 때문에 체류시간이 새로운 장소를 방문할 때보다 적기 때문일 수 있다.

5) 지역별 활동 시간 등 분석

지역별 활동패턴 유형의 점유율을 살펴보고자 SAP1, SAP2, CAP2_1을 대상으로 시도별 점유율을 살펴보았는데, 전체적으로 각 활동패턴의 지역별 점유율은 대표치와 크게 다르지 않은 것으로 나타났다. SAP1의 평균 점유율 약 52%를 기준으로 지역별 분포 비율을 살펴보면 수도권 약 54%, 대전 50%, 울산 51%였고, 강원도 49%, 제주도가 45% 순으로 낮았다. 이외에 SAP2의 경우 수도권의 해당 활동패턴 점유율은 17.3%로 전국 평균 16.0%보다 조금 높았고 지방 광역시와 도 지역은 평균보다 적은 각각 15.8%, 15.6%로 나타났으며, CAP2_1의 경우는 호남권, 충청권, 제주도 등에서 해당 패턴의 점유율이 약 8%로 타 시도보다 높았고 수도권은 약 5.8%로 나타났다. 지역별 활동패턴 유형별 점유율을 살펴본 결과 지역별로 차이가 뚜렷하게 나타나지는 않았다.

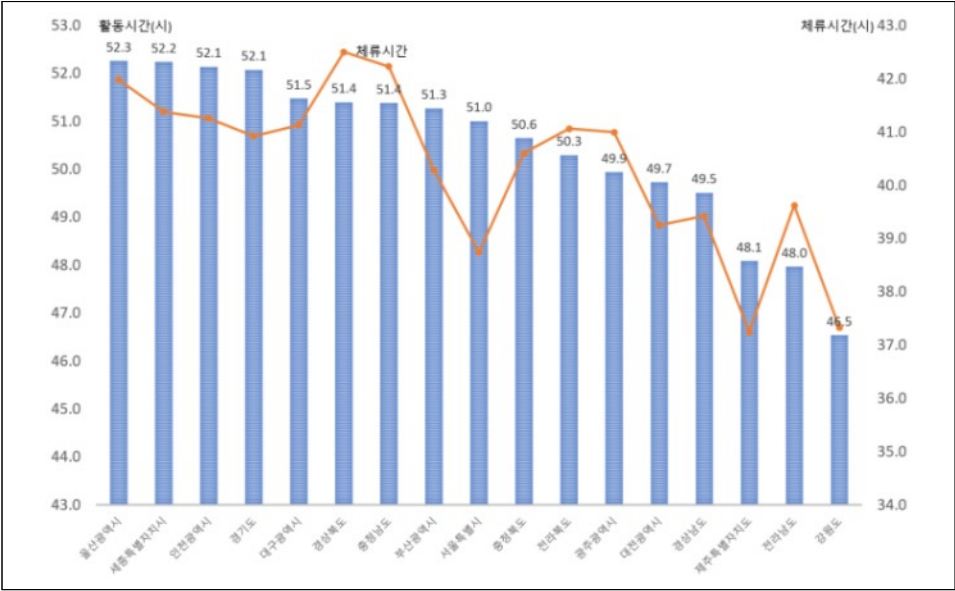
그림 4-28 | 지역별 활동패턴 점유율 비교



주: 차트의 우측 축은 SAP1 비율, 좌측 축은 SAP1, CAP2_1의 비율임
 자료: 저자 작성

요일별로 각 시도의 주중(월-금) 총활동 시간 검토 결과, 울산이 52.3시간으로 가장 높았고 강원도가 46.5시간으로 가장 낮았고 그 차이는 5.8시간이다. 본 연구의 총활동 시간은 집에서 출발해 귀가할 때까지의 시간으로 정의하였으므로 본 데이터상으로 주중 강원도 거주자는 울산 거주자보다 외부 활동이 5.8시간 적고, 반면 집 내부 활동이 그만큼 크다고 볼 수 있는 결과이다. 총활동 시간이 높은 시도를 나열하면 울산, 세종, 인천, 경기도가 52시간 이상이었으며, 대구, 경북, 충남, 부산, 서울 순으로 51시간 이상 높았다. 반대로 주 단위 활동 시간이 48시간 이하인 시도는 제주, 전남, 강원도로 나타났다. 수도권은 주 단위 활동 시간은 약 51.5시간으로 전국 평균 약 50.5시간보다 1시간 높았으며 서울의 해당 시간은 51시간으로 전국 평균보다 0.5시간 높았지만 인천, 경기도에 비해 낮았다.

그림 4-29 | 주중 지역별 누적 활동 시간과 누적 체류시간



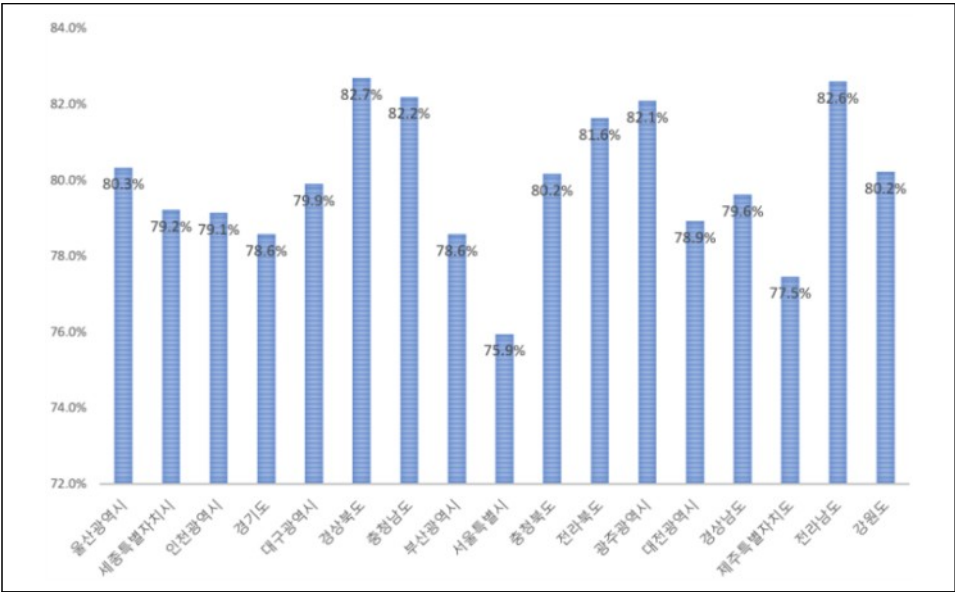
주: 1) 차트의 우측 축은 활동 시간(시), 좌측 축은 체류시간(시)임

2) 막대그래프(활동 시간), 꺾은선 그래프(체류시간)

자료: 저자 작성

요일별로 각 시도의 주중(월-금) 체류시간(집 이외 장소) 검토 결과, 경북과 충남이 42시간으로 가장 높았고 강원, 제주가 37시간으로 가장 낮았으며, 체류시간 정도는 활동 시간 감소에 비례하여 줄어드는 경향이 있었지만, 일부 시도에서는 반대되는 현상도 있었다. 활동 시간 중 체류시간이 차지하는 비율을 시도별로 살펴본 결과, 전국 평균은 약 80%로 나타났으며, 활동 시간 중 체류시간이 차지하는 비율은 서울이 75.9%로 가장 낮았고 이것은 통행으로 소요하는 시간이 그만큼 타 시도보다 높다고 볼 수 있는 결과이다. 이외에도 수도권 중 경기도의 상기 비율은 78.6%, 인천 79.1%로 평균보다 낮았다. 상기 비율이 가장 높은 시도는 경북, 전남, 충남, 광주로 약 82% 정도였고, 다음으로 높은 시도는 전북, 울산, 충북, 강원 등으로 나타났다.

그림 4-30 | 주중 지역별 누적 활동 시간과 누적 체류시간의 비율

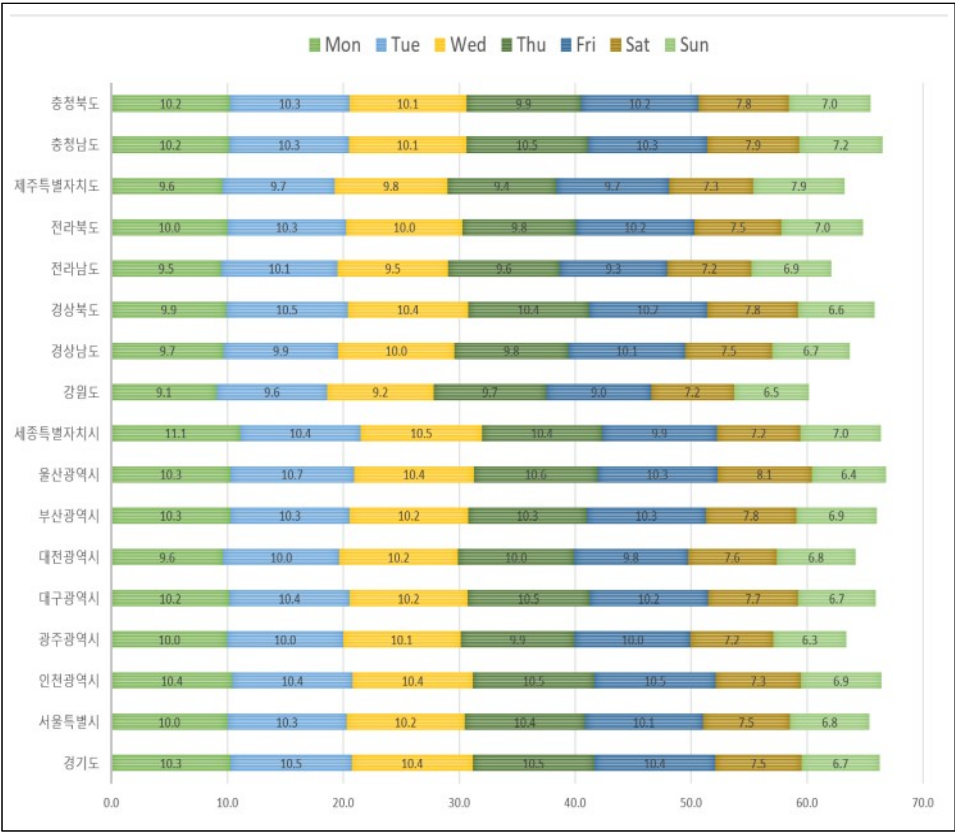


주: 체류시간/총활동 시간 비율

자료: 저자 작성

요일별로 각 시도의 총활동 시간을 검토한 결과 수도권의 주중 활동 시간은 10.4시간, 도 지역은 9.9시간으로 나타났다. 총활동 시간은 집에서 출발해 귀가할 때까지의 시간으로 정의하였으므로 수도권이 집 밖에서의 체류시간이 도 지역보다 크다고 볼 수 있었다. [그림 4-31]은 시도별 요일별 활동 시간을 검토한 결과로, 제주도, 강원도, 경상남도 등의 활동 시간이 적게 나타났다.

그림 4-31 | 요일별 지역별 활동 시간



주: 1) 차트의 우측 축은 활동 시간(시), 좌측 축은 체류시간(시)임

2) 막대그래프(활동 시간), 꺾은선 그래프(체류시간)

자료: 저자 작성

4. 소결

4장에서는 3장에서 정립한 데이터 가공, 정제 및 집 기반 활동패턴 구현방안을 실제 적용하여 위드라이브의 모빌리티 빅데이터로 활동패턴을 도출하고 그 특징을 파악하였다. 활동패턴의 구분은 본 연구에서 활동패턴 대시보드를 작성하는데 적용한 노드, 링크, 경유지 3가지 요인을 기준으로 약 6만 8천 개의 샘플 활동패턴을 60개로 유형화하였고 이중 빈도수가 낮은 활동패턴 등을 기타로 분류하여 총 32개의 활동패턴을 제시하였다. 이로 인해 일상에서 일어나는 활동패턴은 다양하지만 노드, 링크, 경유지 3가지 요인으로 구분하면 32개로 구분할 수 있다는 것으로 향후 활동패턴 연구에서도 활용도가 높을 것으로 보여진다. 본 연구는 활동의 시공간적 연속성이 확보된 활동패턴을 구현했다는 점에서 기존 시공간적 연속성을 설문지로 파악한 가구 통행실태조사, 설문 조사 등의 연구와는 차별적이며 좀 더 현실적인 활동패턴을 파악한 연구라고 사료된다. 전체 활동패턴 중 단순 활동패턴(SAP), 복합 활동패턴(CAP)의 비율은 8:2 정도로 SAP가 CAP보다 4배 이상 높게 나타났으며 SAP1의 비중이 52.8%로 가장 높았다. 또한 32개 활동패턴 중 SAP는 8개, CAP는 24개로 빈도수는 SAP가 높지만 활동패턴을 설명하기 위한 역할은 CAP가 더 크다고 할 수 있다. SAP와 CAP의 요일별 점유율을 검토한 결과 월-토 SAP는 감소, SAP2는 주말 증가 등 일상생활 패턴을 설명하기에 각 패턴별 증감률이 적절한 것으로 나타났으며 CAP2_1의 주중 점유율은 6.7%, 주말 점유율은 5.5%로 주중이 주말보다 1.2%p 높아 주말의 회귀 통행(예: 집 귀가 후 외출)이 주중보다 적다는 결과로 일상생활 패턴을 적절히 설명하는 결과라고 사료된다.

활동패턴 각 점유율을 가중치로 활동 시간을 산정한 결과 주중 활동 시간(집 귀가 후 시간 제외)은 615분(약 10.3시간)으로 나타났고 이중 통행시간에 127분(약 2.1시간), 체류에 488분(약 8.2시간)을 소비하는 것으로 나타났다.

이러한 결과는 통계청에서 실시하는 생활시간 조사의 보완 및 검증자료로 본 연구의 결과물을 활용할 수 있는 가능성을 보여주며, 활동 시간을 체류시간과 이동시간으로

구분할 수 있고 주중과 주말에 대해 각 활동패턴 유형별로 산정할 수 있다는 것은 활동의 다양성을 시간 관점에서 기술할 수 있다는 가능성을 보여준 결과라고 사료된다. 또한, SAP, CAP 목적지 1회 추가당 체류시간을 단순회귀식을 통해 산정해 목적지 추가에 따른 체류시간의 증가치를 과학적으로 제시하였다는 점에서도 의의가 있다.

CHAPTER 5

시사점 및 결론

1. 시사점	133
2. 결론	142

05 시사점 및 결론

본 연구의 시사점은 3장~4장에서 기술한 연구 방법과 분석 결과에 근거하여 학술적 시사점과 정책적 시사점을 제시하였다. 본 연구의 모빌리티 빅데이터의 가공 및 정제방법, 집-기반 활동 분석방법론 등은 모빌리티 빅데이터 구축이 활성화될수록 학술적 활용도가 더 높아질 것으로 기대된다. 또한 정책적으로는 지역 활동 모니터링, 사회적 이벤트 전후 활동 변화 파악, 생활권 설정, 교통 DB 조사 개선 등에서 활용이 가능할 것이다

1. 시사점

1) 학술적 시사점

본 연구의 학술적 시사점은 다음과 같이 4가지로 제시하고자 한다.

(1) 개인단위 모빌리티 빅데이터 분석 방안 제시

단순 통행과 활동의 차이는 통행은 그 자체만으로 완결성을 가지는 반면 활동은 통행의 집합으로 연속성이 시간과 공간적으로 연결될 때 그 의미가 있다는 점이다. 하루 단위 활동은 작게는 2개 통행에서 크게는 10개 이상의 통행 집합으로 표현 가능하다. 개인단위의 스마트폰 앱(App)으로 수집한 위치와 시간정보를 가공 및 분석하였다는 점에서 학술적으로 의미가 있다고 사료된다. 빅데이터의 특성상 일자별 개인별 혼재된 데이터상에서 개인의 하루 단위 통행의 순서를 파악하는 것이 중요하기 때문이다.

하루 단위 활동의 시간적 공간적 연속성이 확보하기 위한 시간적 연속성 확보방안과 출발·도착지 연속성 확보방안을 제시하였다. 통행의 시간적 연속성을 판단 시 해당 장소에서의 체류시간 산정을 통한 논리적 오류 해결방안을 제시하였다. 이 과정에서 음수로 산정되는 통행시간, 최대 통행시간 등 필터링하여 활동 시간의 신뢰성을 제고해야만 한다. 일상 활동은 대개 출근, 쇼핑, 외식, 친지 방문, 여가 등 다양한 목적을 이루기 위한 통행을 하며 그 이동들이 모여 하루의 활동패턴이 형성된다고 볼 수 있다. 개인의 개별적인 통행이 시간적 연속성이 확보될 때 하루 단위의 활동패턴이 형성될 수 있기 때문이다.

개인정보 보호법에 따라 휴대폰을 포함한 모바일 장비의 개인별 이동 위치와 시간정보 취득이 불가능한 상황에서도, 이용자 자발적 정보 제공에 의해 수집된 위드라이브 회사의 자료로 개인별 시공간적 활동 위치 정보 자료를 활용한 개별 활동패턴을 분석하는 방법을 제시한 의미 있는 연구라고 사료된다. 본 연구는 위드라이브 회사의 자료를 활용하였지만, 장래에 법적 제도적 장치가 개선되어 개인정보 공개에 따른 피해를 극소화하며 공익적 자료 활용이 가능한 상황이 도래하였을 때 어떠한 형태의 모바일 위치 정보 자료들을 활용하여 분석 방법과 정책 활용이 가능한 기초적 연구로써 바탕을 마련한 연구로써 의미가 있다.

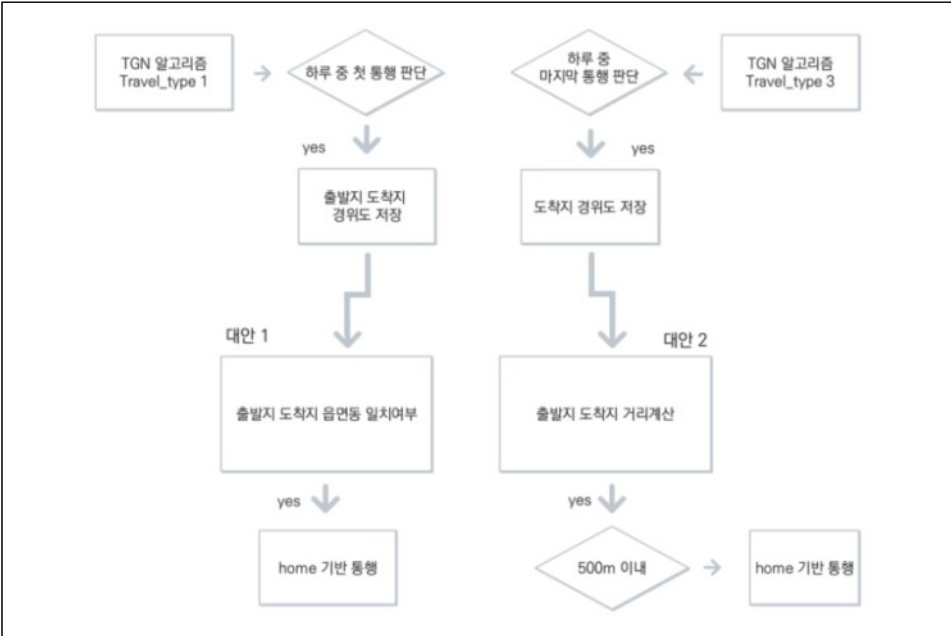
시간/공간/사람과 관련된 엄청난 규모의 data 자체는 너무 커서 오히려 의미성이 없어질 수가 있다. 따라서 이러한 Big data에서 정책에 활용이 가능한 유용한 정보를 규격화 및 표준화된 format에 따라 추출하여 축약된 형태의 2차 자료로써 저장하는 것도 중요하며, 또한 이러한 표준화된 자료가 시계열적으로 동일한 정보로써 지속적으로 축적되는 것도 의미가 있다. 본 연구에서 제시한 포맷을 토지이용 및 교통 정책을 위한 Big data의 2차 자료 가공의 표준화의 한 방편으로 제시하는 것도 필요하다.

(2) 모빌리티 빅데이터 기반의 집-기반 활동 분석 방안제시

집 기반 활동 여부 판단을 위해서는 2가지 작업이 순차적으로 필요하며 이 두 과정은 타 모빌리티 데이터로 집 기반 활동을 분석함에 있어 공통적으로 적용될 수 있는 분석 기법이라고 사료된다. 먼저 김종학 외(2021)가 제시한 통행의 순번을 파악하는 TGN(Trip Generation Numbering) 작업이 필요하다. 이 작업은 모빌리티 빅데이터 중 집-기반 통행을 파악하는 데 있어 최소 2회 이상의 통행을 선정하고 1회 통행을 제거하는 데 유용하다. 예를 들어 하루에 한 개의 통행만을 했다고 하면 이러한 데이터는 집 기반 활동 분석에서 제외되어야 한다.

두 번째 분석은 하루 중 첫 통행의 출발지와 마지막 통행의 목적지가 유사해야 한다는 가정을 만족할 수 있어야 한다. 본 연구는 귀가 여부와 유사 여부를 판단하기 위해 2가지 대안을 가지고 분석하였다. 대안 1은 하루 중 첫 통행(fs: first start)과 하루 중 마지막 통행(le: last end)의 읍면동 일치 시 집 기반 통행으로 판단하는 방안이다. 대안 2는 하루 중 첫 통행(fs: first start)과 하루 중 마지막 통행(le: last end)의 이격 거리를 기준으로 산정하는 방안이다. 대안 2는 여러 가지 작업 과정이 필요하며 먼저 출발지와 목적지 간 포인트를 생성하고 생성된 포인트의 위·경도 좌표를 이용하여 거리를 계산하는 과정과 이격거리 기준값(예: 500m)에 대한 설정이 필요하다. 본 연구는 도보권 등을 고려해 500m로 설정하였지만 충분한 데이터가 확보되고 위치 정보 오류가 적다면 그 이하로 설정하는 방안 마련도 향후 필요할 것이다. 집-기반 활동 분석을 위해 하루 중 집에서 첫 통행이 출발지와 마지막 통행의 도착지가 설정한 공간 범위에 포함되는지를 워크플로우로 구현하였다. 워크플로우는 하루 단위 개인의 통행 순번을 통해 첫 통행의 출발지와 마지막 통행의 목적지 데이터 선정 후 두 포인트의 이격거리가 설정 기준에 포함 여부로 집-기반 활동데이터를 추출하였다.

그림 5-1 | 마지막 통행 도착지의 귀가 여부 판단 알고리즘

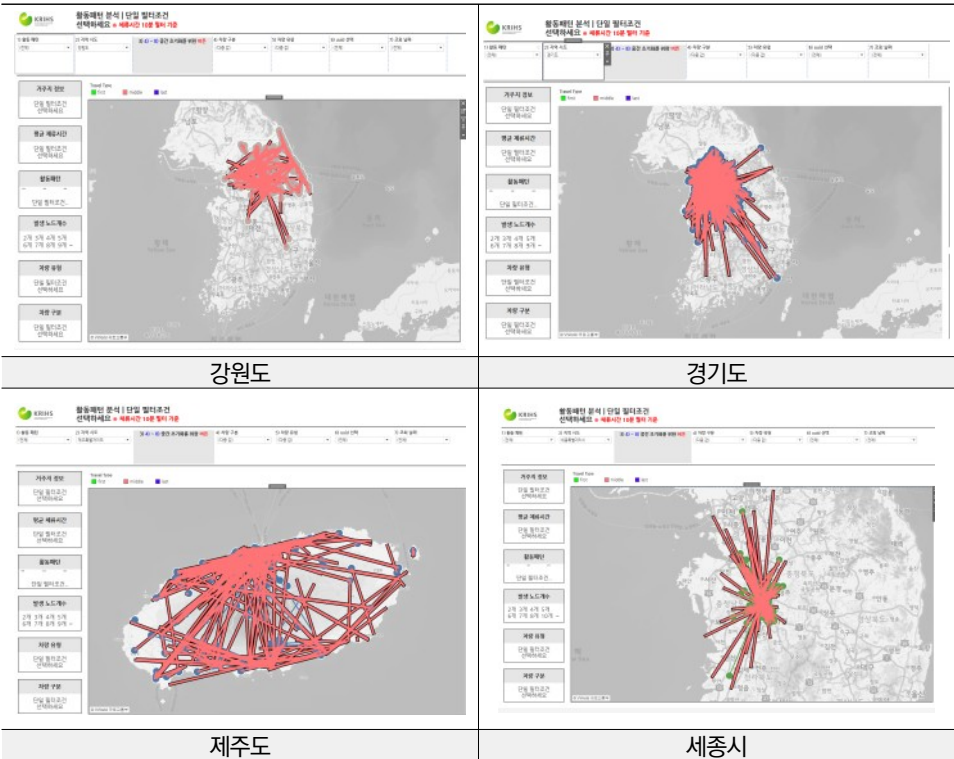


자료: 저자 작성

(3) 모빌리티 빅데이터 가치 제고를 위한 활동패턴 대시보드 개발

스마트폰의 개인위치정보, PM(Personal Mobility) 위치 정보, 택시/트럭 위치 정보, 수요응답형 교통수단 위치 정보와 시간정보의 가치 제고를 위한 대시보드 개발이 필요하다. 눈으로 보는 이동데이터는 지도상에 시각화될 때 그 가치가 제고될 수 있다. 수치상의 이동데이터는 지리정보와 연계될 때 그 활용도와 가치가 제고될 수 있으며 본 연구에서 개발한 활동패턴 대시보드 개발 방법은 유사한 모빌리티 빅데이터 분석 방안으로 활용 가능성이 높을 것으로 사료된다. [그림 5-2]는 본 연구에서 개발한 활동패턴 대시보드를 이용하여 시도별 활동 분포를 표현한 것으로 기존 교통계획의 희망선도와 다른 점은 활동패턴을 형성하기 때문에 통행과 통행 사이가 연계된 활동 분포도 작성이 가능하다.

그림 5-2 | 활동패턴 대시보드 개발



자료: 저자 작성

(4) LNS(노드, 링크, 경유지) 정보 기반 활동패턴 유형 구분기준 제시

본 연구 결과 개별 통행을 시간과 공간적으로 연속성을 가지니 형태로 표현하기 위해서는 출발지, 도착지를 표시하는 노드와 출발·도착 간을 연결하는 링크 그리고 활동의 연속성을 의미하는 경유지 3개 요인이 필요하다.

기존 교통계획 등의 통행을 표시하기 위한 노드와 링크 외 경유지를 추가한 것으로 출발지로 다시 돌아가는 통행의 회귀성 등도 파악할 수 있는 분석 방안이다. 활동패턴 유형 개수는 위 3가지 변인의 조합 가능 개수로 약 29.3만 개의 데이터로 약 6.8만 개의 활동패턴 데이터를 작성하고 이를 32개로 유형화할 수 있었다. 하루 단위 개인

통행데이터로 LNS(노드, 링크, 경유지) 정보 생성을 위해서는 [그림 5-3]과 같이 각 포인트 데이터에 3가지 ID를 부여해야 한다. ID 1은 일자별 개인별 활동패턴에 대한 고유 ID¹⁾이며, ID 2는 일자별 경유지 ID로 경유지 수로 링크 개수이다. ID 3은 위치 정보 유사성 판단을 통해 부여하는 ID 2의 번호로 SAP의 활동패턴은 그 번호가 ID 2와 같고, 복합 활동패턴(CAP)의 경우에는 회귀성 활동이기 때문에 서로 다른 번호가 있을 수 있다.

그림 5-3 | 활동패턴 대시보드 개발을 위한 데이터 구분 ID

ID1	ID2	ID3	uuid	date	ID1	ID2	ID3	uuid	date
140	000128	002	000ba264fc6048f39095b70546e60648	2022-03-17	041312	004	002	26baa6797998475c80c94f16154666d7	2022-03-28
141	000128	003	000ba264fc6048f39095b70546e60648	2022-03-17	091481	004	002	520ac8b67ca043dcbf5b3337da7a3b0c	2022-03-28
142	000129	001	000ba264fc6048f39095b70546e60648	2022-03-18	013285	002	002	0c9150658e414435ba0a274988856db2	2022-03-29
143	000129	002	000ba264fc6048f39095b70546e60648	2022-03-18	040608	002	002	261e7b045c46c6da278a243a3fc5a9	2022-03-29
144	000129	003	000ba264fc6048f39095b70546e60648	2022-03-18	013285	004	002	0c9150658e414435ba0a274988856db2	2022-03-29
145	000131	001	000ba264fc6048f39095b70546e60648	2022-03-21	040608	004	002	261e7b045c46c6da278a243a3fc5a9	2022-03-29
146	000131	002	000ba264fc6048f39095b70546e60648	2022-03-21	018517	004	002	6aa09493803d4ea8ee9f7e645727e0	2022-03-29
147	000131	003	000ba264fc6048f39095b70546e60648	2022-03-21	025135	002	002	177f9a5254411c3c8ee3c702e2f1e	2022-03-30
148	000131	004	000ba264fc6048f39095b70546e60648	2022-03-21	099478	002	002	593a14e0221e48c58d4819089643bae	2022-03-30
149	000132	001	000ba264fc6048f39095b70546e60648	2022-03-23	025135	004	002	177f9a5254411c3c8ee3c702e2f1e	2022-03-30
150	000132	002	000ba264fc6048f39095b70546e60648	2022-03-23	013066	004	002	65841cca61794183a770cc8aefef3	2022-03-30
151	000132	003	000ba264fc6048f39095b70546e60648	2022-03-23	013270	002	002	0c8cf0d087a1428082ae89690478801	2022-04-01
152	000133	001	000ba264fc6048f39095b70546e60648	2022-03-24	040248	002	002	25818c1a7b46494eb4a7a48262ca9981	2022-04-01
153	000133	002	000ba264fc6048f39095b70546e60648	2022-03-24	013270	004	002	0c8cf0d087a1428082ae89690478801	2022-04-01
154	000133	003	000ba264fc6048f39095b70546e60648	2022-03-24	011362	004	002	0a8689937e40828968ec49515b566	2022-04-02
155	000133	004	000ba264fc6048f39095b70546e60648	2022-03-24	011290	002	002	0c9150658e414435ba0a274988856db2	2022-04-03
156	000133	005	000ba264fc6048f39095b70546e60648	2022-03-24	011287	002	002	67aea8a74cd744a18a9e7cb259d05526	2022-04-03
157	000133	006	000ba264fc6048f39095b70546e60648	2022-03-24	069911	002	002	4e040ca244904ae8977f154d5c8be371	2022-04-03
158	000133	007	000ba264fc6048f39095b70546e60648	2022-03-24	011290	004	002	0c9150658e414435ba0a274988856db2	2022-04-03
159	000133	008	000ba264fc6048f39095b70546e60648	2022-03-24	011287	004	002	67aea8a74cd744a18a9e7cb259d05526	2022-04-03
160	000133	009	000ba264fc6048f39095b70546e60648	2022-03-24	091836	002	002	5251db0640444c9caf4ab4b53c7b65	2022-04-02
161	000134	001	000ba264fc6048f39095b70546e60648	2022-03-25	091836	004	002	5251db0640444c9caf4ab4b53c7b65	2022-04-02
162	000134	002	000ba264fc6048f39095b70546e60648	2022-03-25	082859	004	002	460028a77954541b535a1b7446e682	2022-03-21
163	000134	003	000ba264fc6048f39095b70546e60648	2022-03-25	109439	004	002	6299d3261564e4d082c94365ef8c1	2022-04-01
164	000134	004	000ba264fc6048f39095b70546e60648	2022-03-25	002232	004	002	03048180dc0248b9409c188ae51f1a	2022-03-22
165	000134	005	000ba264fc6048f39095b70546e60648	2022-03-25	080004	004	002	489aa51502754430b345c8e495dc086	2022-03-23
					025925	002	002	1884f738bc194c48ace91846457546da	2022-04-03
					025925	004	002	1884f738bc194c48ace91846457546da	2022-04-03
					018371	002	002	118e578a6846e41b8ea373099213cb4	2022-03-14

자료: 저자 작성

2) 정책적 시사점

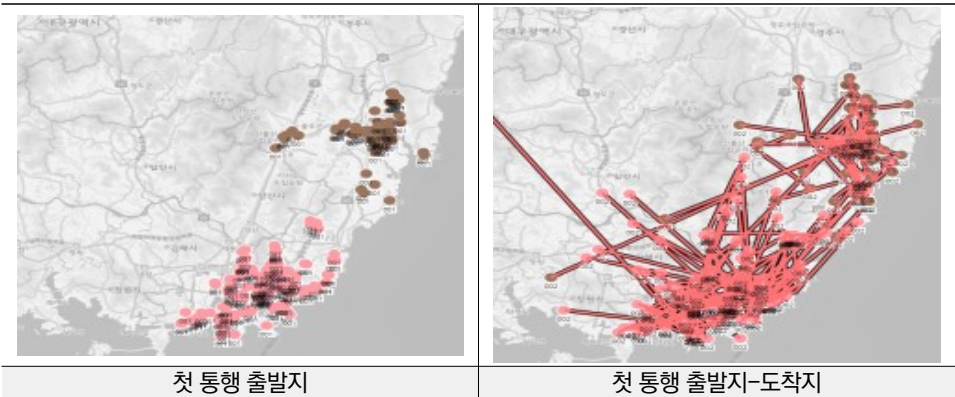
(1) 지역 활동 모니터링 활용방안

본 연구에서 제안한 활동패턴 방안은 지역에서 개인의 하루 단위 활동 형상을 파악할 수 있어 개인의 지역 활동 모니터링 방안으로 활용될 수 있을 것으로 보여진다. 개인의 활동은 시간 연속성을 가지고 공간에서 이루어지므로 공간 단위를 어떻게 구성하

1) ID1에 3330이 표시되었다면 이것은 A라는 사람이 3.14일 3번 이동했다는 의미임

느냐에 따라서 그 의미는 다양해질 수 있지만 본 연구의 데이터는 미시 데이터로 시도, 시군, 읍면동, 통계구 등 다양한 공간 단위에 적용이 가능할 것이다. 본 연구의 데이터 상에는 파악하지 못한 개인정보(연령, 성별)를 별도의 앱 설문조사 등으로 파악한다면 개인 군집별로 지역 활동패턴을 파악할 수 있을 것이다. 아래 예는 본 연구 분석 결과로 부산과 울산 거주자들의 2022. 3. 14(월)의 첫 통행(예: 출근)부터 순차적으로 이루어지는 활동패턴을 표현한 것으로 첫 통행의 출발지는 부산과 울산 거주지에 있던 사람들이 출근 등의 사유로 목적지로 이동하는 공간적 분포를 기술하고 있다. 이러한 활동행태 파악을 통해 시각화 기반의 지역 활동 모니터링이 가능할 것으로 보여진다.

그림 5-4 | 부산, 울산 지역 활동 연계성 검토 예시(3.14 월)



자료: 저자 작성

(2) 사회적 이벤트 전후 활동 특성 변화

전염병 확산, 미세먼지 확산 등의 기후변화 등에 따른 경제활동을 간접적으로 모니터링 할 수 있는 방안이라고 사료된다. 전염병 확산 시 본 연구에서 제시한 체류시간 등은 해당 지역이 전염병 위험도를 파악하는 자료로도 활용이 가능할 것으로 보여진다. 빅데이터는 미시 데이터에서 거시 데이터까지 파악이 가능해 사회적 이벤트 전후 기존 통행량 변화와 차별된 활동패턴의 변화를 파악할 수 있는 장점이 있다. 예로 경기 침체로 인한 퇴근 후 활동 변화, 주말 여가 통행 변화 등의 파악이 가능하다.

그림 5-5 | 코로나 전후 모빌리티 활동패턴 변화 (예시)

구분		코로나 확산 전	코로나 확산 후	
			case1	case2
활동 패턴 (사각)	노드4 링크4			
	활동면적	대	중	소
	체류	5곳(집귀가 가정)	좌동	4곳(집귀가 가정)
	이동	4회	좌동	좌동
	체류시간	4hr(거주지 제외 평균)	3hr(거주지 제외 평균)	5hr(거주지 제외 평균)
	패턴변화	-	×	○

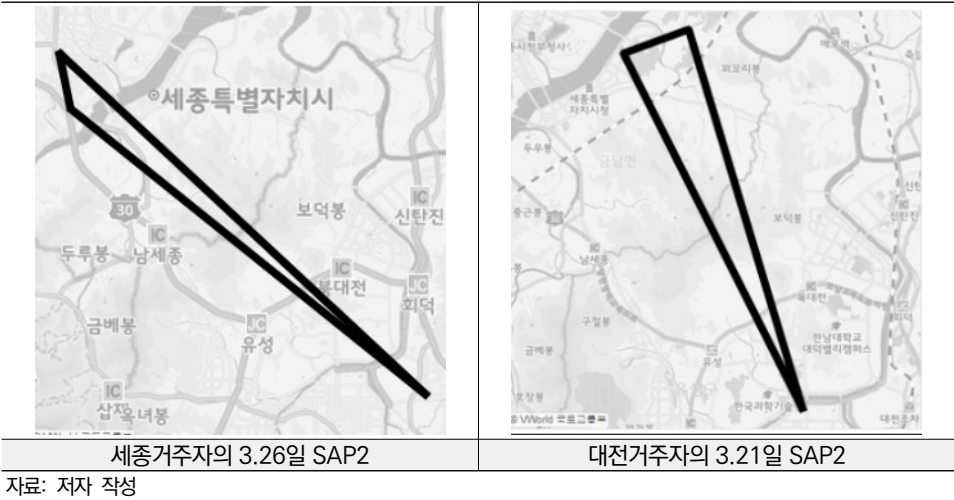
주: 상기 자료는 개념도임
자료: 저자 작성

(3) 초광역권, 광역 교통권역 설정 등의 기초자료로 활용

인구감소 대응, 초광역권 설정, 광역 교통 대책 및 시설 설치를 위한 광역 교통권역 설정 등에 본 자료를 활용할 수 있을 것으로 보여진다. 개인 활동을 본 연구에서 개발한 방법으로 파악한다면 어느 도시와의 연계가 강한 면을 나타내는지 파악이 가능할 것이다. 현행 대도시권 권역 설정은 합리적 권역 설정 기준이 부재한 상태에서 행정구역 중심으로 관습적으로 권역을 설정하는 한계가 있다.

광역권 포함 여부에 따라 교통인프라 공급과 지역개발사업 사업에서의 공공부문 재정지원 분담 비율에 차이를 두고 있기 때문에 (예: 광역 전철 지정 시 중앙정부 재원 지원 비율 증가) 광역권 설정에 따른 지자체 간 갈등 발생으로 사회적 비용이 증가하는 문제점을 보이고 있다. 유사한 통행을 보이는 동질지역을 하나의 광역권으로 설정하기 위해서 통상 권역 내 통행을 최대로, 권역 간 통행을 최소화하는 합리적인 광역권 설정 기준 마련이 필요하다. 본 연구에서 검토된 앱 기반 모빌리티 빅데이터를 이용한 출근/귀가 통행 분석, 활동패턴 유형 구분 분석 결과는 합리적인 광역권 설정에 유용하게 활용이 가능하다.

그림 5-6 | 세종, 대전시 거주자 활동패턴(SAP2) 예시



(4) 지역 간 공유 교통 인프라 선정

본 연구에서 제시한 활동패턴은 출근뿐만 아니라 출근 후 귀가 전까지의 활동을 연속적으로 표현할 수 있어 지자체가 버스노선, 차고지, 터미널 등 공유인프라를 발굴 필요성을 검토하는 자료로 활용 가능하다. 즉, 지방 분권 시대 연계 협력형 교통부문 공유인프라뿐만 아니라 보건, 의료, 복지 등 다양한 공유인프라 선정 및 평가 방법으로 활용 가능하다. 본 연구 제안한 사업발굴 과정 및 평가과정은 향후 지방 분권 시대 연계 협력형 교통부문 공유인프라뿐만 아니라 보건, 의료, 복지 등 다양한 공유인프라 선정 및 평가 방법으로 활용 가능하다.

특히, 여가 활동의 경우 광역적인 관광자원을 이용하는 활동패턴이 많을 것으로 보여지기 때문에 본 연구에서 제안한 주말 활동패턴으로 해당 관광자원이 시군단위의 시설인지, 도 단위의 시설인지, 국가 단위의 시설인지 파악이 가능해 관광자원의 등급 결정 및 지원 등의 차별화가 가능할 것으로 보인다.

(5) 사전 광역교통개선 대책 수립 및 사후 모니터링

신도시 입주 초기 노선버스 운행 전 지역민의 활동패턴 파악을 통한 신도시 내, 광역 지역 간 임시버스 마련을 위한 기초자료로 활용 가능하다. 또한, 광역교통시설 확충 효과 모니터링 평가 방법으로 적용할 수 있다. GTX 개통 전후 자료로 본 연구 제시 활동패턴을 파악한다면 실제 이용객의 동선 추정을 통한 확충 효과도 단기간에 파악할 수 있을 것으로 기대된다. 신도시, 혁신도시 등 기초조사가 미흡하거나 어려운 지역 등에서 본 연구 방법으로 데이터를 분석한다면 출퇴근 교통 문제 원인 파악이 가능하다고 판단된다. 지역별, 시간대별 해당 통행의 빈도수를 시간과 공간별로 파악 가능해 지역별로 차등적인 광역 교통 문제 발생원인 파악이 가능하기 때문이다.

2. 결론

1) 연구 종합

본 연구는 모빌리티 빅데이터의 가공, 정제, 분석을 통해 다양한 활동패턴을 특정 패턴으로 정의하여 구분하고 그 특징을 분석하고자 하였다. 스마트폰 자체의 GPS 기능과 시간정보로 다양한 앱(App)에서 스마트폰 이용자의 위치와 시간정보를 개인의 동의하에 파악하고 있으며 이 정보를 연구에 활용하였다. 기존 가구 통행실태조사, 생활시간 조사 등에서는 출근 후 거래처 방문 등의 연속적인 활동을 이용자가 기입하여 시간과 공간적인 연속성을 기본 전제로 하고 있다. 하지만, 모빌리티 빅데이터는 개인의 스마트폰 배터리 상태, 통신오류, 위치 오류 등의 다양한 문제로 활동의 연속성이 확보되지 않는 사례가 다수 발생한다. 본 연구를 수행함에 있어 개인의 하루 단위 활동의 시간적 공간적 연속성이 확보되는 데이터를 확보하는데 가장 많은 시간과 노력을 기울였다. 데이터의 형태와 특징은 제공 주체 측에서도 파악하지 못하는 여러 가지 오류가 있을 수 있으며 이러한 오류는 분석가가 데이터를 직접 보고 그 오류의 특징을

파악할 때 해결책을 모색할 수 있다. 예를 들어 출발시간 기준으로 데이터 정렬 시 다음 통행의 출발시간은 현재 통행의 도착시간보다 더 늦어야 한다는 기본조건을 충족시키지 못하는 데이터가 존재는 체류시간 산정 등을 통해 그 오류를 발견할 수 있었다.

일상 활동은 대개 출근, 쇼핑, 외식, 친지 방문, 여가 등 다양한 목적을 이루기 위한 통행을 하며 그 이동 들이 모여 하루의 활동패턴이 형성된다고 볼 수 있다. 개인의 개별적인 통행이 시간적 연속성과 공간적 연속성이 확보될 때 하루 단위 활동패턴이 파악 가능하다. 본 연구는 약 160만 개의 개별 통행데이터의 가공을 통해 시공간적 연속성이 본 연구 기준에 부합한 약 6만 8천 개의 활동데이터를 확보할 수 있었다. 상기 활동데이터를 활동패턴 대시보드 작업을 통해 제시한 각 활동의 링크, 노드, 경유지 개수의 조합으로 구분한 결과 32개의 활동패턴을 제시할 수 있었다.

활동패턴의 특징으로는 총활동 시간, 체류시간, 통행시간 등으로 구분하여 개인의 하루 단위 집 밖에서의 활동 특성을 시도별, 요일별로 분석할 수 있었다. 활동패턴 분석을 위해 본 연구에서 제시한 가공/정제/분석 방법 등은 다양한 기기로 수집되는 모빌리티 빅데이터의 기초 가정과 분석 방법으로 활용될 수 있을 것으로 기대된다.

2) 연구의 의의

스마트폰 앱 기반의 모빌리티 빅데이터로 개인의 위치 및 시간 이동정보만으로 개인의 활동패턴을 파악하였다는 점에서 의의가 있다. 최근까지 활동패턴 분석연구 트위터 활동, 설문 조사 등의 데이터는 사용한 적이 있지만 스마트폰의 앱 기반 위치 정보를 활용한 사례는 없었다. 본 연구는 앱 기반 위치 정보로 활동패턴의 시각화를 통해 다양한 개인의 활동패턴을 유형화하고 그 특징을 파악하였다는 점에서 의의가 있다. 하루 중 집 밖에서의 활동 시간을 단순히 집 출발, 귀가 시간의 차이만으로 산정한 것이 아니라 집 출발 이후 방문지를 연속적으로 파악하여 최종 귀가 여부를 시간과 위치 정보로 파악하였다는 점에서 의의가 있다. 본 연구는 하루 단위 개인의 활동패턴을 유형화하고 총활동 시간을 체류시간과 이동시간으로 구분하여 개인의 특정 장소에서의 활동

성을 판단할 수 있는 자료로 활용 가능할 것으로 보여진다. 스마트폰 앱으로 수집되는 정보 중 가장 기초적인 정보는 위치와 시간정보이며 이러한 정보를 가공 및 정제하는 방법을 제시하였다는 점에서 의의가 있다. 따라서, 본 연구에서 제시한 시공간 정보 가공방안은 관광, 전염병 예방, 교통 db, 경제활동, 지역 활동 등에서 다양하게 활용될 수 있을 것으로 기대된다. 또한, 전염병, 경제위축 등 사회변화 시 개인의 활동 변화를 파악하는데 적용 가능하다. 본 연구에서 개발한 활동패턴 대시보드 활용 시 특정 이벤트 전후의 개인 활동패턴을 적은 노력으로 분석할 수 있을 것이다.

3) 한계 및 향후 과제

본 연구에 활용한 데이터가 2022.3-2022.4, 3주간이라는 점은 데이터상으로 분명 한 한계가 있다. 특히, 해당 시기가 코로나로 인한 사회적 거리두기 시점이기 때문에 일상생활의 제약이 있었을 것으로 보여지며 이러한 현상이 일반적인 활동패턴을 설명하는 데 한계가 있었을 것으로 사료된다. 위드라이브의 모빌리티 빅데이터는 해당 앱 사용자를 기반으로 샘플 수가 적어 전국 시도 이하의 시군구, 읍면동 단위까지의 활동을 설명하는 데 한계가 있었다. 또한, 해당 앱은 자동차 이용자 대상 앱으로 대중교통, 자전거 등의 이용자들을 포함하지 못하는 한계가 있었다. 개인의 활동패턴은 개인의 속성(성별, 연령별)으로 그 차이점을 파악하지 못한 한계가 있었다고 판단된다. 개인 속성정보의 부재로 연령, 성별로 차별적인 개인의 활동패턴을 구분하지 못한 것은 아쉬운 점이었다. 향후 과제에서는 충분한 일자 또는 연 단위별 데이터를 확보해 분석 시기별 활동패턴의 다양성을 파악할 수 있어야 할 것이다. 기후변화, 경기 변화, 사회적 현상을 ICT로 파악하는 것이 우리에게 주어진 과제라고 할 때 다양한 빅데이터 분석의 실패와 시행착오에 대한 사회적 포용은 진일보한 미래 혁신으로 나아가기 위한 값진 경험이라고 사료된다.

참고문헌

REFERENCE



【 인용문헌 】

- 구자용. 2017. 공간 빅데이터를 이용한 개인 활동패턴의 시공간적 탐색: 서울시 지역의 트윗 데이터를 사례로. 국토지리학회지 51(3):259-270.
- 김규혁, 이동엽, 김동호, 원민수, 홍성민, 송태진. 2021. 모바일 생활통행데이터 기반 도시 인구 규모별 생활권 분류 및 특성 파악. 대한교통학회지 39(5):662-679.
- 김종학, 임현섭, 윤서연, 장요한. 2021. 빅데이터 기반 도시 모빌리티 패턴 분석 방법론 연구. 세종:국토연구원
- 김영표, 한선희, 김미정, 문린곤. 2001. 디지털시대에 대비한 사이버국토 구축전략 연구. 세종:국토연구원
- 위드라이브. 2020. 제품소개자료. 인천: 위드라이브
- 이재현. 2014. 미디어 이용의 시공간적 매핑: 크로노토프 접근과 그 적용. 언론정보연구 51(1):214-250.
- 통계청. 2020. 2019년 생활시간 조사결과. 7월 30일. 보도자료.
- 한국교통연구원, 2018. KTDB 정책자료집 여객 통행실태 INDEX BOOK.
- Järv, O, Ahas, R, and Witlox, F. 2014. Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records. Transportation Research Part C: Emerging Technologies no.38:122-135.
- Shaw, S. L. and Yu, H. 2009. A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical-virtual space. Journal of Transport Geography no.17:141-149
- Hägerstrand, T. 1970. WHAT ABOUT PEOPLE IN REGIONAL SCIENCE? . Ninth european congress of the regional science association, no.24:7-21.

구글 코로나19 지역사회 이동성 보고서

https://www.gstatic.com/covid19/mobility/2022-05-10_KR_Mobility_Report_ko.pdf

(2022년 5월 10일 접속)

네이버 지도 길 찾기

<https://map.naver.com/v5/directions> (2022년 6월 1일 접속)

SUMMARY



Spatio-temporal Activity Pattern Analysis by Mobile Big Data -Activity Pattern Type and Characteristics-

Kim Jonghak , Kim Hojung, Yoon Seoyeon, Yeon Chihyung

Key words: Activity Pattern, Spatio-temporal, Mobile Big Data, Activity Pattern Type

This study aims to develop a method for analyzing daily activity patterns using mobility big data mining. By combining human-time-space data, collected through GPS functionality on smart phones, individuals' activity patterns can be identified without separate surveys. The study proposes data processing and algorithm development for presenting activity patterns, and develops a dashboard for analyzing mobility big data based on its spatio-temporal characteristics. The analysis includes commuting and travel characteristics by attempt, distance traveled by vehicle type, and identification of activity pattern types based on daily activities as Link, Node, Stop over(LNS). The study contributes to understanding individual mobility patterns and provides insights for utilizing mobility big data in the transportation sector. The analysis data for the dashboard is obtained from a start-up company, WEDRIVE, which collects data by providing points to drivers who use their service. The data covers a three-week period with low holiday impact. The data is made up of seven variables and is processed to determine daily individual activity patterns.

The TGN (Trip Generation Numbering) method is used to determine the sequence of a person's travel during the day. This is done alongside determining their residence time and travel time, which is used to exclude data that exceeds the maximum travel time and negative residence time. The data is analyzed using a workflow, which includes determining that the continuity condition is met if the residence time at a location is more than 10 minutes, and the spatial continuity is met if the distance between the starting and ending point is less than 500m. The activity pattern dashboard is divided into six menus: the main menu, basic menu, activity pattern menu, and three other menus. The top menu allows the user to select the activity pattern, district, vehicle type, and individual ID, and date. The additional function of the dashboard was used to visually review the continuity of movement and to present a correction plan for each activity pattern. Activity pattern type classification and characteristic analysis are the results of analysis based on the activity pattern dashboard developed by the researcher. In this study, activity patterns were first classified into simple activity patterns (SAP) and complex activity patterns (CAP) according to the presence or absence of return traffic to places other than home during the day, and detailed types were secondarily classified based on links (L), nodes (N), and stops (S). Simple Activity Patterns (SAPs) have the same number of links (L) and nodes (N) because there is no regression passage with two departures and arrivals, whereas Composite Activity Patterns (CAPs) have regression passage, so the number of stops (S) is larger than the number of nodes and links. In this study, 68,000 daily individual activity patterns were analyzed, and 60 link (L)/node (N)/stop (S) combinations were found, of which a total of 32 activity pattern types were derived, excluding those with less than 10 frequencies. It is divided into 8 SAPs and 24 CAPs, and SAP accounts for 81% of the overall activity pattern. In addition, we analyzed the share by activity pattern type, and

the analysis showed that SAP was about four times higher than CAP. The share of the top 10 activity patterns was 94.4%, of which 5 SAP shares were 80.2% and 5 CAP shares were 14.2%. The top six cumulative share activities accounted for about 89% of the total. Finally, the activity time analysis by activity pattern was performed. In this analysis, the average activity time for each activity pattern was compared by weekday, weekend, and time zone, and the characteristics of each activity pattern were identified through this. This study proposes a methodology to identify individual activity patterns using only personal location and time information obtained from a smart phone app, utilizing node-link separation algorithms and data processing methods to ensure spatio-temporal continuity. The study also validates the use of mobility big data to analyze trip chains, which were previously only possible to obtain through household travel surveys. The proposed activity pattern analysis can be used to monitor individual local activities and indirectly monitor economic activity related to climate change and other phenomena. The data can also be used for setting up metropolitan transportation policies and facilities. Limitations include the need for more diverse data over a longer period and the lack of analysis on individual characteristics. Future efforts should continue to focus on processing and analyzing big data to understand climate and societal changes.

기본 22-28

위기 기반 빅데이터를 활용한 시공간 활동패턴 분석연구

- 활동패턴의 유형 구분 및 특징을 중심으로 -

저 자 김종학, 김호정, 윤서연, 연치형

발 행 인 강현수

발 행 처 국토연구원

출판등록 제2017-9호

발 행 2022년 12월 31일

주 소 세종특별자치시 국책연구원로 5

전 화 044-960-0114

팩 스 044-211-4760

가 격 7,000원

I S B N 979-11-5898-813-5

홈페이지 <http://www.krihs.re.kr>

© 2022, 국토연구원

이 연구보고서를 인용하실 때는 다음과 같은 사항을 기재해주시시오.

김종학, 김호정, 윤서연, 연치형. 2022. 위기 기반 빅데이터를 활용한 시공간 활동패턴 분석연구. 세종: 국토연구원.

이 연구보고서의 내용은 국토연구원의 자체 연구물로서 정부의 정책이나 견해와는 상관없습니다.

이 연구보고서는 한국출판인협회에서 제공한 KoPub 서체와 대한인쇄문화협회가 제공한 바른바탕체 등이 적용되어 있습니다.

위치기반 빅데이터를 활용한 시공간 활동패턴 분석연구

- 활동패턴의 유형 구분 및 특징을 중심으로 -

Spatio-temporal Activity Pattern Analysis by Mobile Big Data
-Activity Pattern Type and Characteristics-



제1장 서론

제2장 국내외 관련 연구

제3장 데이터 가공 및 활동패턴 대시보드 개발

제4장 활동패턴 유형의 구분 및 분석

제5장 시사점 및 결론



KRIHS 국토연구원

(30147) 세종특별자치시 국책연구원로 5 (반곡동)
TEL (044) 960-0114 FAX (044) 211-4760

