

발 간 등 록 번 호

B551982-2019-000014-10

2019 지방자치단체 빅데이터 분석 사례집

B I G
Bio health
DATA



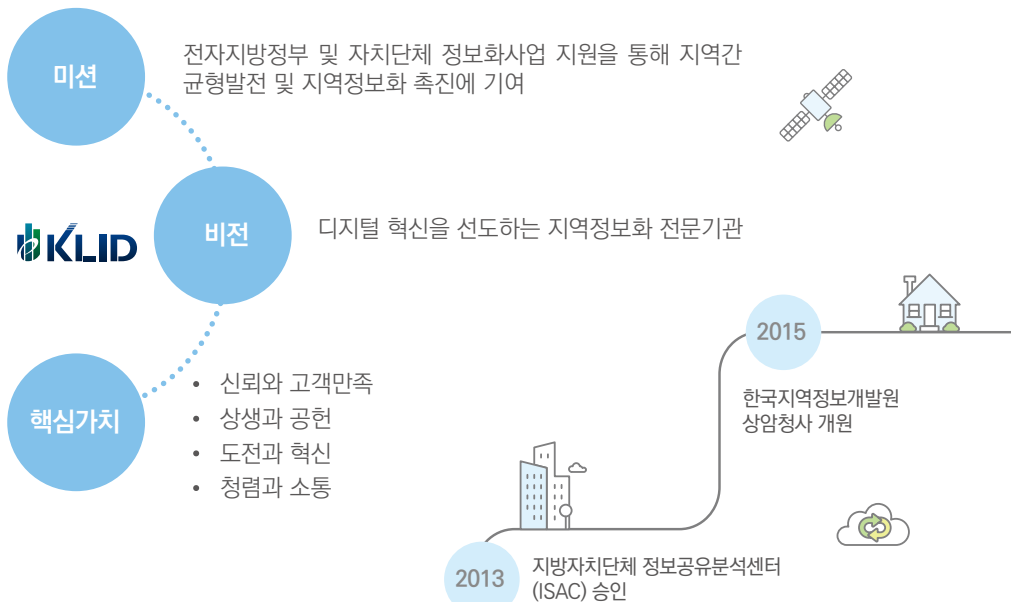
한국지역정보개발원
Korea Local Information Research & Development Institute

■ 한국지역정보개발원 소개

●● 설립근거 및 목적

한국지역정보개발원은 전자정부법 제72조에 따라 설립되어 전자지방정부 구현 및 지역정보화에 기여하고 있습니다. 안전하고 스마트한 최고의 지역정보화 전문기관으로서 정보화사업 지원, 지방자치단체 위탁사무, 연구 및 교육, 정보시스템의 개발·구축·확산, 국제교류·협력 및 국외 확산·보급을 위해 힘쓰고 있습니다.

●● 미션, 비전 및 핵심가치



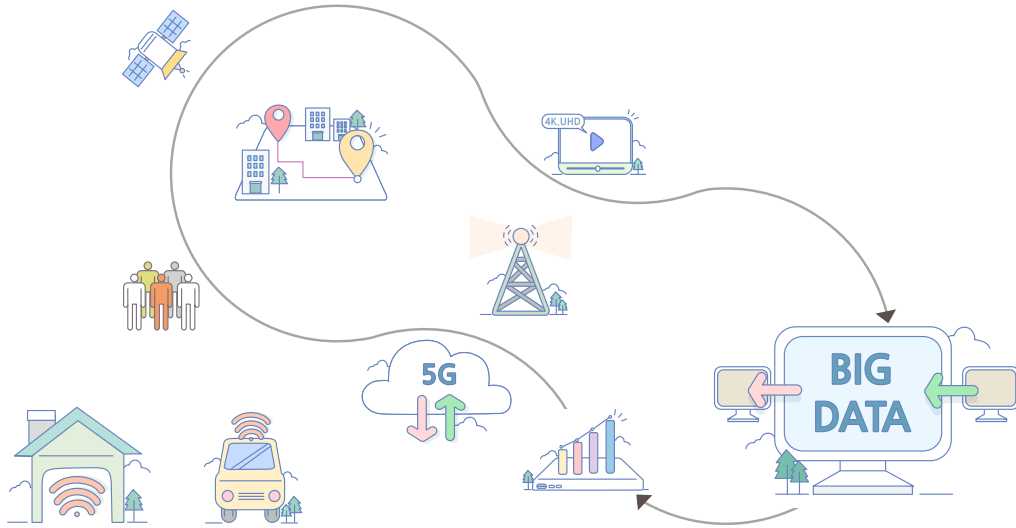
●● 주요업무

- 지방자치단체의 체계적인 정보화 발전을 위한 연구와 컨설팅 지원
- 지방자치단체 행정정보시스템 구축 및 안정적 운영지원
- 지방자치단체 정보시스템 장애와 사이버공격으로부터 보호
- 중앙 및 지방자치단체의 전자정부와 개인정보 보호를 위한 인프라 제공
- 지역경제 활성화와 정보화 글로벌 역량 강화를 위해 전자상거래 및 정보화 교육 제공

●● 주요연혁



■ 사례집 발간 배경 및 활용 방안



- 지방자치단체 빅데이터 담당자 분석 및 활용 역량 부족
- 지방자치단체 간 빅데이터 분석 및 활용에 대한 격차 완화 필요
- 사례 중심의 실무자 맞춤형 빅데이터 분석 사례 구성
- 빅데이터 분석 사례 조사 결과에 대한 추진 단계별 분석 절차
- 지방자치단체 빅데이터 분석 추진 시 활용도 높은 사례집 발간

■ 연도별 빅데이터 분석 우수 사례 발굴

2016년 우수사례	2017년 우수사례	2018년 우수사례	2019년 우수사례
"경기 남양주시" 기초 인구 분석	"제주특별자치도" 공공데이터 활성화 정책	"제주특별자치도" 대중교통 활용 이동형 bt 플랫폼 구축	"KLID" 사회 경제적 예측 분석을 통한 지자체 효율적 예산 수립
↓	↓	↓	↓
"경기도" CCTV 사각지대 분석	"부산광역시" 도시양극화 분석	"광주광역시" 도시안전 제고를 위한 취약지 분석	"KLID" 도로 미세먼지 발생량 예측 및 노면 청소차량 운행 효율화
	↓	↓	↓
	"경상남도" 응급환자 골든 타임확보를 위한 분석	"서울 성북구" 빅데이터 기반 지역경제활성화 분석	"부산광역시" 기상조건에 따른 시내버스 정류장 정상성 분석
	↓	↓	↓
	"경기 남양주시" 보건의료지표의 행정 활용	"경기 평택시" 민생경제 빅데이터 분석	"제주특별자치도" 전기차 충전 인프라 운영·관리 및 콜센터 운영 정책 수립을 위한 EV콜센터 민원 분석
	↓	↓	↓
	"경기 광주시" 당직민원 빅데이터 분석	"충북 청주시" 무인민원발급기 이용 현황 분석	"서울특별시 성동구청" 금연구역 후보지 도출 및 관리를 위한 흡연 민원지역 빅데이터 분석





한국지역정보개발원 소개	<u>02</u>
--------------	-----------

발간배경	<u>03</u>
------	-----------

01+

[KLID] 사회·경제적 예측 분석을 통한 지자체 효율적 예산 수립	<u>06</u>
---------------------------------------	-----------

02+

[KLID] 도로 미세먼지 발생량 예측 및 노면 청소차량 운행 효율화	<u>33</u>
--	-----------

03+

[부산광역시] 기상조건에 따른 시내버스 정류장 정시성 분석	<u>58</u>
----------------------------------	-----------

04+

[제주특별자치도] 전기차 충전 인프라 운영·관리 및 콜센터 운영 정책 수립을 위한 EV콜센터 민원 분석	<u>107</u>
---	------------

05+

[서울특별시 성동구] 금연구역 후보지 도출 및 관리를 위한 흡연 민원지역 빅데이터 분석	<u>152</u>
--	------------



[2019년 KLID 신규분석모델 ①]

사회·경제적 예측 분석을 통한 지자체 효율적 예산 수립

| 한국지역정보개발원, 펜타시스템테크놀로지(주)컨소시엄 |

요약

2019년 한국지역정보개발원에서는 『KLID 소관 정보시스템 보유데이터 현황조사 및 분석모델 개발』 사업을 통해 전국 지자체에서 공통적으로 적용 가능한 신규 분석모델을 발굴하였다.

그 중 첫 번째 과제인 「사회·경제적 예측 분석을 통한 지자체 효율적 예산 수립」 모델의 경우, 지자체별 지출/예산 데이터와 주민등록인구, 재정자립도 등 다양한 사회·경제적 변수를 이용하여 각 지자체의 향후 세목별 지출을 예측하는 과제로서, 지자체별로 갖고 있는 예산 수립/집행 과정에서 발생할 수 있는 애로사항을 해결하고, 과학적인 예산 수립을 지원할 목적으로 추진되었다.

분석을 위해 사용한 데이터는 지자체별 세목별 예산내역과 세목별 지출내역, 지자체 유형, 재정자립도, 주민등록연앙인구, 주민등록인구, 이동자수, 출산율, 사망률, 농가수, 임가수, 어가(해수면)수, 어가(내수면)수 등 총 18개 항목이다.

주요 분석 내용으로는 첫째, 지출내역과 각종 데이터들(18개 항목)을 결합하여, 세목별 지출예측 모델을 생성하고, 둘째, 주민등록연앙인구, 이동자수, 출산율, 사망률 데이터를 이용하여 미래 인구에 대한 추계를 실시한다. 셋째로 지출 예측 모델에 사용되는 다양한 데이터들에 대해 예측을 실시하고, 마지막으로 예측된 인구 데이터와 다양한 데이터들로 데이터마트를 생성한 뒤, 제작된 세목별 지출 예측 모델에 적용하여 미래 지출 데이터를 예측 및 각종 시각화를 실시하였다.



가 추진배경

■ 기존의 세목별 예산수립방안에서 벗어난 예산 수립의 과학적인 근거 마련

- 현재 행정안전부 자치분권 로드맵(안)에 따라 강력한 재정분권이 추진되고 있으며, 지자체별로 예산 수립 시 예산편성지침에 따라 예산을 편성하도록 되어있으나, 대부분 직전년도 예산 및 지출액에 비례하여 책정하고 있는 상황임
- 이로 인해, 지자체에서는 당초 편성된 목적에 맞게 효율적으로 예산을 활용하지 못하고 추가경정예산을 편성하는 등의 문제점 발생
 - 2017년 기준 지자체 추경 총액은 45조원으로 본예산 총액의 약 17%에 달하고 있으며, 지속적으로 증가 추세
 - 최초예산과 추경예산을 합친 총 예산 중 활용하지 못한 세계잉여금의 경우 44조원에 달하여 추경 총액인 45조원과 차이를 보이지 않음
 - 예산 수립이 효율적인 방식으로 이루어지지 않고 있으며, 해당 예산 또한 지자체별로 충분히 활용하지 못하고 있는 것으로 볼 수 있음

나 프로젝트 개요

■ 분석 개요

- 분석기간 : 2019.10. ~ 2019.12.
- 분석데이터 : 18개 항목(지자체별/세목별 예산, 지자체별/세목별 지출, 지자체별/연령별/성별 주민등록 연앙인구 및 주민등록인구, 지자체 유형분류, 지자체별 재정자립도 등)

■ 분석의 필요성

- 지자체별로 예산 수립시, 예산편성지침에 따라 예산을 편성하도록 되어있으나, 대부분 직전년도 예산 및 지출액에 비례하여 책정
- 이로 인해 예산을 효율적으로 활용하지 못하는 문제가 발생하였으며, 그 결과 지자체별 추경예산이 급증하는 등의 문제 발생



■ 주요 분석내용

- 예산/지출/인구 데이터 현황분석을 통한 주요 요인 도출
- 다양한 요인을 바탕으로 세목별 지출예측 모델 생성
- 전국 인구데이터에 대한 미래추계
- 추계된 인구를 비롯한 활용데이터와 모델을 결합하여, 분석 지자체의 연도별/세목별 지출금액 예측 및 시각화



다 적용 프로세스

■ 분석 전체 프로세스

- 기획관리(문제의식 → 사업계획 → 완료보고)
- 데이터분석(분석 설계 → 데이터 수집 및 가공 → 데이터 분석 → 데이터 시각화)
- 행정적용(시범적용)





기획·관리 단계

1. 기획·관리 단계 요약

단계	주요 내용
문제의식	<ul style="list-style-type: none"> ■ 대부분의 지자체에서 예산 편성 시 직전년도 예산 및 지출에 비례하여 차년도 예산을 책정하기 때문에 예산 활용에 있어 비효율성 발생 ■ 빅데이터를 활용한 과학적 예산책정 의사결정 도구 필요
사업계획	<ul style="list-style-type: none"> ■ 세목별 지출예측 모델 생성 <ul style="list-style-type: none"> - 지자체 유형분류, 농림수산업 가구수 등 ■ 미래인구 추계 <ul style="list-style-type: none"> - 주민등록연앙인구, 사망률, 이동자수, 출산율 ■ 기타 데이터 추계 <ul style="list-style-type: none"> - 재정자립도, 지자체 유형분류, 농림수산업 가구수 등 ■ 데이터마트 생성 및 지출데이터 예측/시각화
완료보고	<ul style="list-style-type: none"> ■ 착수보고, 중간보고, 완료보고 형태로 보고회 진행 ■ 각 보고회 추진 시, 자문위원, KLID 부서 담당자 등 의견 수렴 실시

10



2. 문제의식

과제	<ul style="list-style-type: none"> ■ 빅데이터를 활용한 과학적 예산책정 의사결정 도구 필요
추진내용	<ul style="list-style-type: none"> ■ 담당자 인터뷰를 통한 사회현안 및 문제의식 도출 <ul style="list-style-type: none"> - 원칙적으로는 예산 편성 시 예산편성지침을 따라야함 - 하지만, 대부분의 지자체에서 실제로 예산을 편성할 때, 직전년도 예산 및 지출에 비례하여 차년도 예산을 책정하여 예산 활용에 애로사항 발생 ■ 언론 등을 통한 문제의식 관련 현황 파악 <ul style="list-style-type: none"> - 현재 행정안전부 자치분권 로드맵(안)에 따라 강력한 재정분권이 추진되고 있으며, 상세 추진내역으로 자율과 책임이 담보된 지방 재정 운용 기반을 마련하도록 정책을 추진 중임 - 하지만, 2017년 기준 지자체 추경 총액은 45조원으로 본예산 총액의 약 17%에 달하고 있으며, 지속적으로 증가 추세

추진내용	<ul style="list-style-type: none"> ■ 최초예산과 추경예산을 합친 총 예산 중 활용하지 못한 세계잉여금의 경우 44조원에 달하여 추경 총액인 45조원과 차이를 보이지 않음 <ul style="list-style-type: none"> - 예산 수립이 효율적인 방식으로 이루어지지 않고 있으며, 해당 예산 또한 지자체별로 충분히 활용하지 못하고 있는 것으로 볼 수 있음
Tip	<ul style="list-style-type: none"> ■ 실무 담당자의 의견을 수렴하고, 수렴한 사회현안이 실제로 발생한 이슈인지 파악해야 함 ■ 데이터 분석을 통한 문제해결 가능 여부를 파악하는 것이 중요 ■ 유관기관에서 보유한 데이터 수집 가능 여부 및 데이터 정제 작업 소요기간 검토 필요

3. 사업계획

■ 사업에 대한 요구사항을 정의하여 계획을 수립하고 확정하는 단계

■ 사업범위를 결정하고 사업내용을 구체화함

- 사업계획 수립 시 현황 및 문제점, 사업내용 등에 대한 정의가 필요

과제	<ul style="list-style-type: none"> ■ 구체적이고 명확한 사업계획 수립 필요 ■ 과학적이고 효율적인 방식의 예산수립의 필요성을 강조하여, 전국 각 지자체별 의사 결정자에게 본 분석모델의 효과성을 인지시키는 것이 중요함 	
추진내용	<ul style="list-style-type: none"> ■ 유관기관 인터뷰 및 과제 도출 <ul style="list-style-type: none"> • ('19.09.) KLID 재정정보부 인터뷰 및 후보과제도출 • ('19.10.) 분석과제 서면자문 실시 및 분석과제 선정 <ul style="list-style-type: none"> - 자문위원단 및 지자체 실무자 대상 자문 실시 • ('19.10.) 분석대상 선정 ■ 분석 데이터 확보(~'19.11.) 	
	공공 데이터	<ul style="list-style-type: none"> • 지방재정365 : 지자체별 예산/지출 데이터, 지자체 유형분류, 재정자립도 • 국가통계포털 : 주민등록연앙인구, 주민등록인구, 사망률, 이동자수, 출산율, 농림수산업 가구수, 산림 면적, 미세먼지, 하수도 사업비, 사업체 수, 문화 및 관광산업 종사자 등

Tip

- 내부 데이터 중에도 실제로 빅데이터 분석에서 활용될 수 없는 데이터가 있을 수 있으므로, 이를 사업계획 단계에서 파악하는 것이 중요
- 초기에 발견하지 못할 경우, 사업 진행과정에서 분석에 필요한 데이터가 부족하거나, 활용 불가하여 분석을 진행할 수 없는 경우가 발생할 수 있음

4. 완료보고

과제

- 지자체 예산·지출내역 및 사회·경제 예측 분석을 통한 효율적 예산 수립

추진내용

- 분석 대상 확정(전국 기초자치단체)
- 예산/지출/인구 현황분석을 통한 분석 인사이트 도출
- 세목별 지출예측모델 생성
- 지출예측모델 과정에서 활용된 인구 데이터 추계
- 지출예측모델 과정에서 활용된 기타 데이터 추계
- 추계한 데이터를 바탕으로 데이터마트 생성
- 생성된 데이터마트를 세목별 지출 예측모델에 적용하여, 지자체별/연도별/세목별 지출금액 예측 및 시각화
- 완료보고('19.12.)
 - 사업 완료보고, 자문 및 의견수렴 등

Tip

- 분석과제에 대한 정책개발 및 적용방안을 도출하기 위해서는 유관기관 및 담당부서와의 긴밀한 협조가 필요함



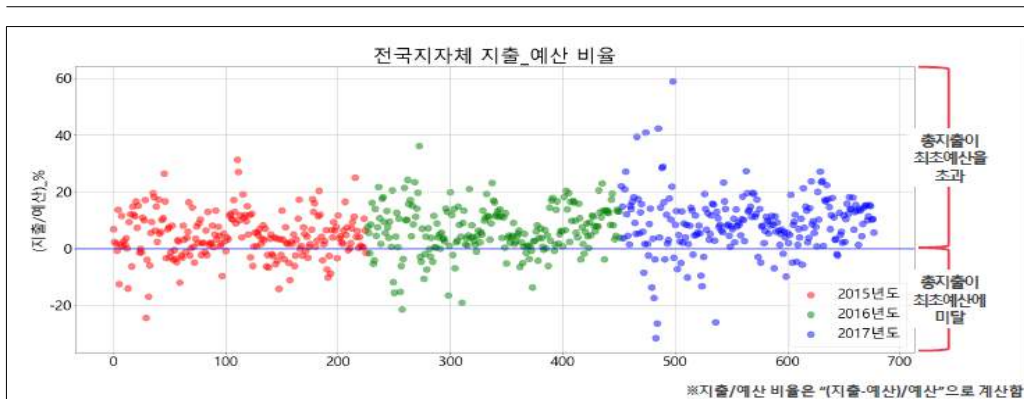
5. 분석결과



〈분석 프로세스〉

■ 예산/지출 비교분석

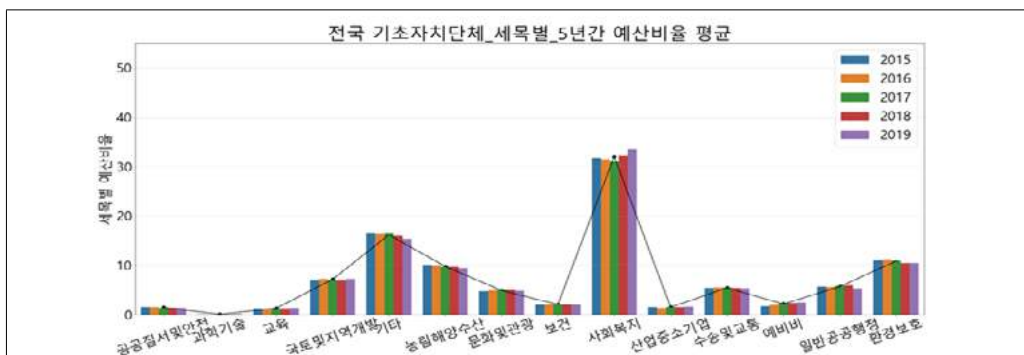
- 전국단위 기초자치단체의 2015년도 ~ 2017년도의 지출/예산 비율에 대한 현황분석을 실시하여, 지자체별로 예산 수립 및 지출이 효율적으로 이루어졌는지 확인
- 대다수의 지자체에서 총 예산 대비 총 지출이 많은 것으로 파악되었으며, 일부 지자체에서는 최초 예산대비 40% 이상의 추가 지출이 발생하였고, 최고 약 60%까지 초과 사용한 지자체가 존재하는 것으로 확인
- 해당 지자체에서는 추경예산을 편성하여, 예산 초과분을 집행한 것으로 보임
- 일부 지자체에서는 최초 예산대비 지출이 적었으며, 최초 수립예산대비 25%까지 사용하지 않은 지자체도 존재함
- 전반적으로 지자체별로 예산 수립 및 지출이 효율적으로 이루어지지 않은 것으로 파악됨



〈2015 ~ 2017년 전국지자체 수립예산대비 지출비율〉

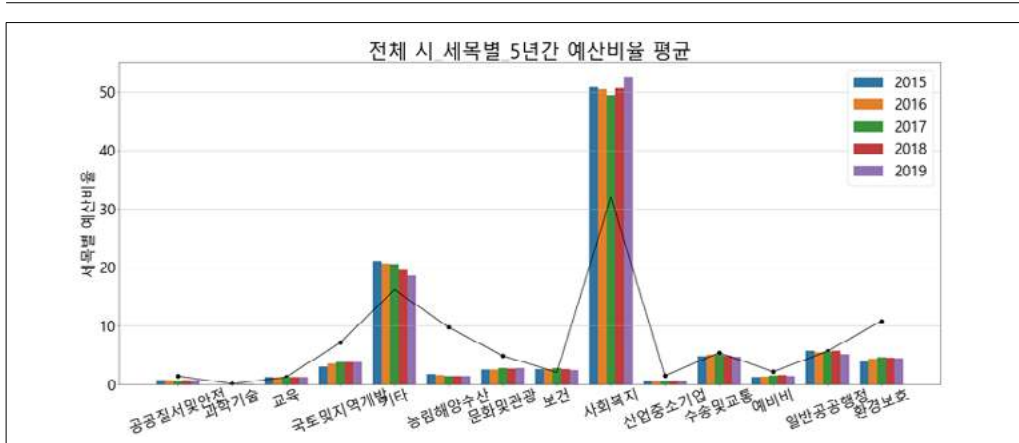
■ 예산 현황분석

- 기초자치단체의 연도별/세목별 예산 비율 평균을 구하고 시각화하여, 각 세목별 추이 확인
- 광역자치단체 중 시(특별시 및 광역시)에 속한 기초자치단체(군/구)의 연도별/세목별 예산비율 평균을 구하고 전국단위 평균 추이와 비교
- 광역자치단체 중 도에 속한 기초자치단체(시/군)의 연도별/세목별 예산비율 평균을 구하고 전국 단위 및 시 단위 평균 추이와 비교
- 광역자치단체 중 시와 도의 차이를 확인하기 위해 시/도별 예산평균 추이 분석 실시
- 시/도 내 기초자치단체별로 예산평균 추이를 확인하여 특이사항 도출
- 예산 현황분석 결과, 시/도에는 뚜렷한 패턴 차이가 존재하며, 각 광역자치단체 내 기초자치단체가 수행하는 역할이나, 해당 기초자치단체의 특성에 따라 다른 패턴을 보이는 것으로 파악됨
- 예비비에 매우 높은 비중의 금액을 편성하는 기초자치단체가 일부 존재

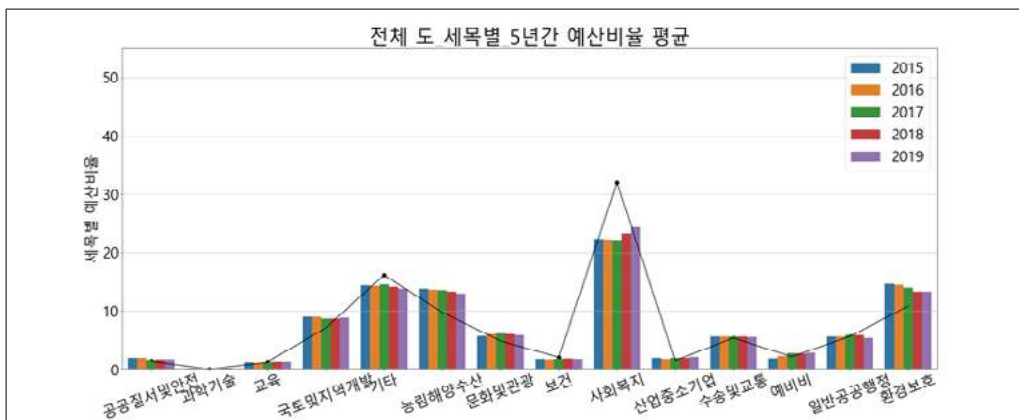


〈5년간 예산 평균 추이분석(전국)〉





〈광역자치단체별/세목별 5년간 예산 평균 추이분석(시)〉



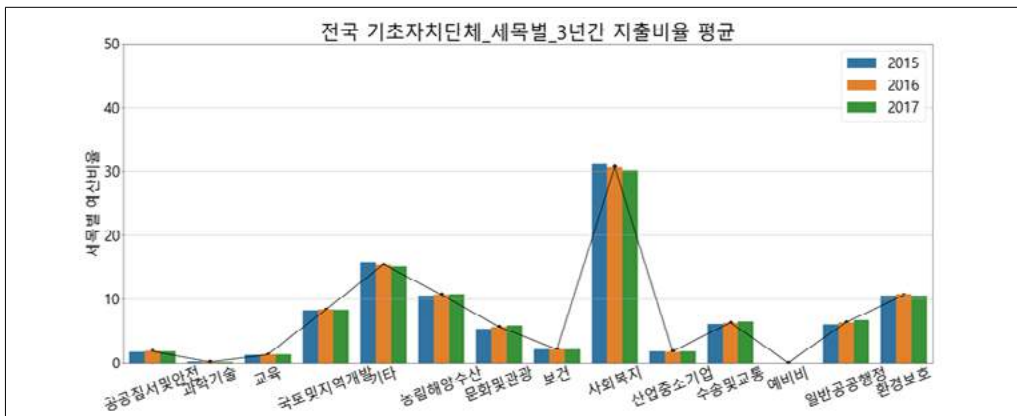
〈광역자치단체별/세목별 5년간 예산 평균 추이분석(도)〉

■ 지출 현황분석

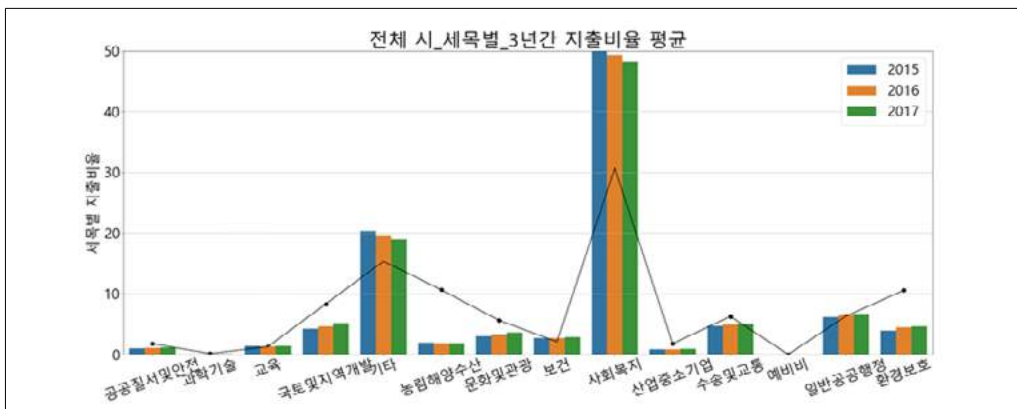
- 전국 모든 기초자치단체의 연도별/세목별 지출 비율 평균을 구하고 시각화하여, 각 세목별 추이 확인
- 광역자치단체 중 시(특별시 및 광역시)에 속한 기초자치단체(군/구)의 연도별/세목별 지출비율 평균을 구하고 전국단위 평균 추이와 비교
- 광역자치단체 중 도에 속한 기초자치단체(시/군)의 연도별/세목별 지출비율 평균을 구하고 전국단위 및 시 단위 평균 추이와 비교
- 광역자치단체 중 시와 도의 차이를 확인하기 위해 시/도별 지출평균 추이 분석 실시
- 시/도 내 기초자치단체별로 지출평균 추이를 확인하여 특이사항 도출



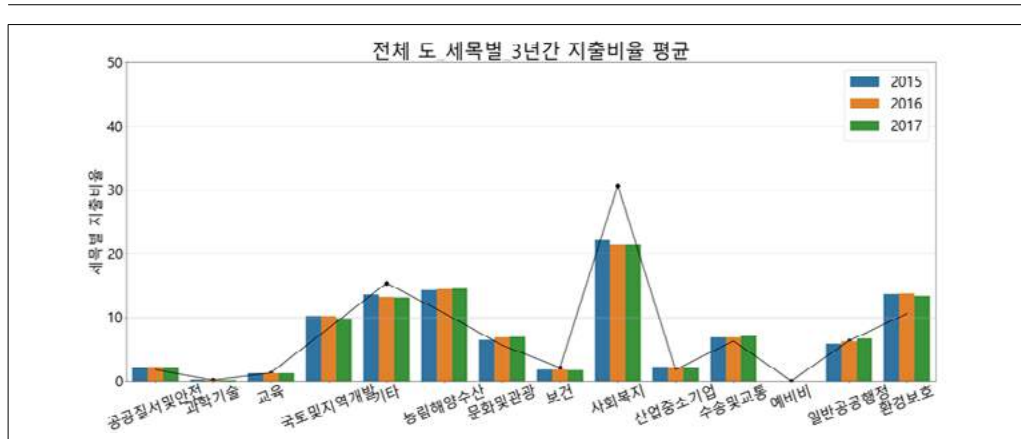
- 기초자치단체별 세목별 비율 상 특이값을 갖는 지자체를 확인하고, 해당 특이값을 갖는 기초자치단체의 데이터는 제거하여 진행
- 시/도 내 기초자치단체별로 지출평균 추이를 확인하여 특이사항 도출
- 기초자치단체별 세목별 비율 상 특이값을 갖는 지자체를 확인하고, 해당 특이값을 갖는 기초자치단체의 데이터는 제거/대치하여 진행
- 지출 현황분석 결과, 예산과 마찬가지로 시/도에는 뚜렷한 패턴 차이가 존재하며, 각 광역자치단체 내 각 기초자치단체가 수행하는 역할이나, 해당 기초자치단체의 특성에 따라 다른 패턴을 보이는 것으로 파악
- 과학기술분야는 226개 분석대상 지자체 중 10개 기초자치단체에서만 지속적으로 지출이 발생하나, 매우 낮은 비율의 지출이 발생하므로, 분석대상에서 제외
- 광역자치단체 중 “도”에서는 농림수산분야가 상대적으로 큰 비중을 차지하기 때문에, 농림수산분야를 추론할 수 있는 변수 추가 필요
- 분석과정에서 지자체 유형이나, 역할을 판단할 수 있는 데이터 추가 필요



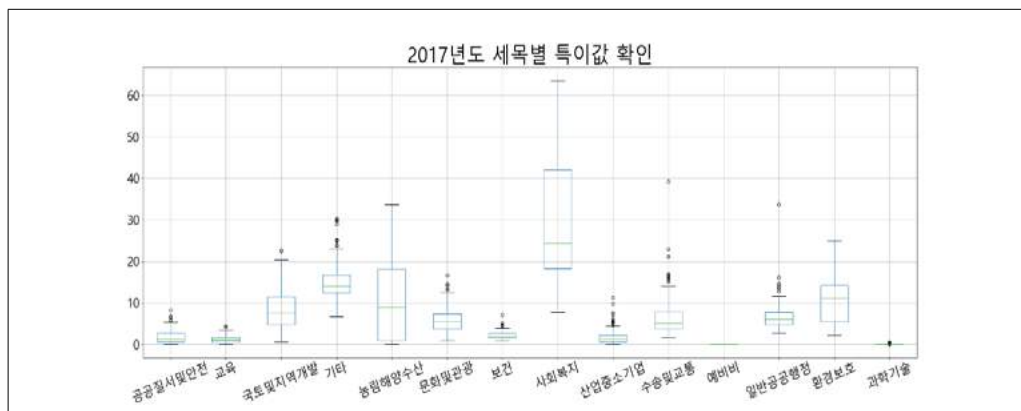
〈3년간 지출 평균 추이분석(전국)〉



〈광역자치단체별/세목별 3년간 예산 평균 추이분석(시)〉



〈광역자치단체별/세목별 3년간 예산 평균 추이분석(도)〉



〈2017년 기초자치단체별/세목별 지출 특이값 확인〉

■ 인구 현황분석

- 전국 모든 기초자치단체의 인구 변화 추이를 확인하여 분석 인사이트 도출
- 각 광역자치단체별 구도심으로 파악되는 지역에서는 대체로 인구가 지속적으로 감소 추세를 보이고 있으며, “도”의 경우 “군”의 인구감소가 눈에 띈다
- 각 광역자치단체별 신도시로 파악되는 지역에서는 대체로 인구가 지속적으로 증가하는 추세이며, “도”의 경우 “시”의 인구증가가 눈에 띈다
- 분석과정에서 지자체 유형이나, 공동주택입주일 등의 데이터 추가 필요



순위	광역자치체명	기초자치체명	평균증감률	2010 증감률	2011 증감률	2012 증감률	2013 증감률	2014 증감률	2015 증감률	2016 증감률	2017 증감률
1	부산광역시	영도구	-2.25	-0.91	-2.45	-2.4	-2.83	-2.73	-2.06	-2.34	-2.25
2	대구광역시	서구	-1.97	-0.26	-0.89	-1.08	-2.19	-2.15	-2.25	-3.17	-3.77
3	전라남도	고흥군	-1.47	-1.6	-1.78	-1.44	-1.3	-1.4	-1.49	-1.38	-1.36
4	대구광역시	남구	-1.45	-0.39	-1.23	-0.22	-1.57	-1.55	-1.61	-2.75	-2.29
5	전라남도	보성군	-1.42	-0.52	-1.84	-1.76	-1.41	-1.08	-1.16	-1.94	-1.61
6	경상북도	의성군	-1.24	-0.47	-1.85	-1.21	-1.72	-1.52	-1.33	-0.85	-1
7	전라남도	해남군	-1.21	-1.93	-1.55	-0.25	-0.6	-0.9	-1.02	-1.41	-2.02
8	충청남도	부여군	-1.11	-0.71	-1.37	-1.01	-0.97	-1.09	-0.85	-1.34	-1.57
9	전라북도	김제시	-1.08	-0.45	-1.31	-0.85	-1.19	-1.22	-1.54	-1.06	-0.98
10	충청남도	예산군	-1.01	-0.18	-0.67	-0.63	-0.78	-0.34	-1.69	-2.57	-1.23
11	경상북도	영덕군	-0.99	-0.8	-1.41	-1.32	-0.29	-1.39	-1	-0.35	-1.34
12	전라남도	함평군	-0.96	-1.04	-1.55	-0.98	-0.48	-1.46	-0.61	-1.37	-0.2
13	부산광역시	사하구	-0.95	-0.3	-0.92	-0.76	-1.03	-1.64	-1.63	-0.94	-0.39
14	전라남도	화순군	-0.92	-0.44	-0.55	-0.88	-0.8	-1.56	-1.38	-0.83	-0.95
15	서울특별시	도봉구	-0.89	-0.39	-0.74	-0.9	-1.02	-1.36	-0.7	-0.86	-1.16
16	전라남도	진도군	-0.82	-0.55	-0.71	-0.88	-0.63	-0.76	-0.82	-1.23	-0.98
17	충청북도	단양군	-0.65	-0.08	-0.7	-1.08	-0.26	-0.72	-0.6	-0.84	-0.94
18	경상북도	상주시	-0.63	-0.57	-0.87	-0.48	-0.8	-0.44	-0.5	-0.56	-0.84
19	충청북도	옥천군	-0.55	-0.17	-0.98	-0.47	-0.9	-0.56	-0.12	-0.26	-0.96
20	전라남도	완도군	-0.51	-0.06	-0.43	-0.31	-0.47	-0.83	-0.3	-0.65	-1.06

...

〈인구 지속 감소지역 상위 20개 지역〉

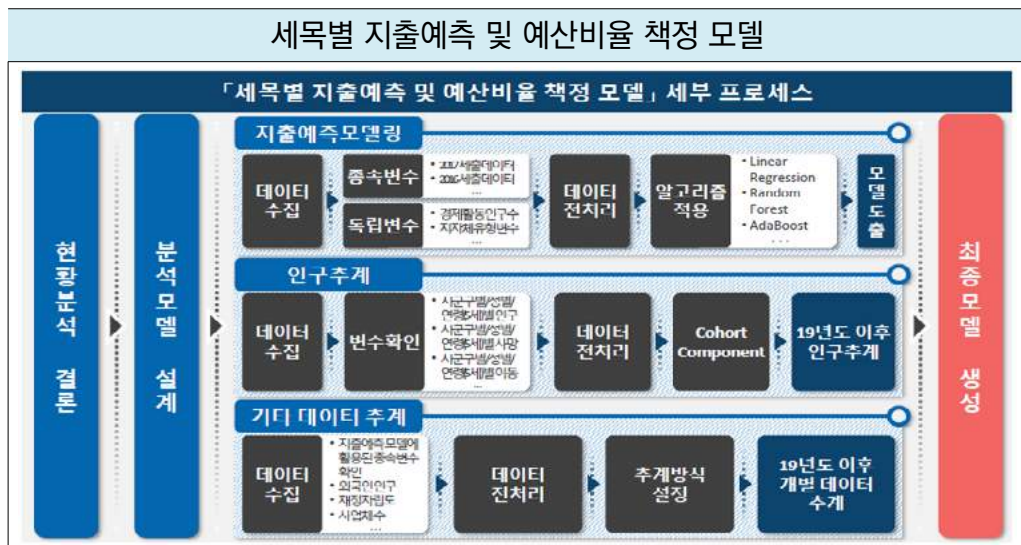
순위	광역자치체명	기초자치체명	평균증감률	2010 증감률	2011 증감률	2012 증감률	2013 증감률	2014 증감률	2015 증감률	2016 증감률	2017 증감률
1	부산광역시	강서구	8.83	3.68	3.42	1.66	6.18	13.95	18.83	15.12	7.78
2	부산광역시	기장군	7.79	14.74	5.40	5.99	15.66	9.24	5.77	3.55	1.97
3	경기도	김포시	7.18	5.55	7.83	11.84	8.65	8.97	2.84	3.84	7.88
4	경기도	광주시	4.76	4.70	6.18	3.93	4.01	4.24	4.59	4.84	5.56
5	경기도	화성시	4.40	2.91	2.16	1.69	0.91	2.00	10.29	7.44	7.83
6	경기도	파주시	3.91	10.10	6.51	4.07	1.91	2.40	2.91	1.76	1.64
7	경상남도	양산시	3.85	3.93	2.37	1.92	3.37	4.58	3.05	5.23	6.78
8	충청북도	청주시	3.58	1.99	0.91	0.74	0.91	23.57	0.05	0.39	0.05
9	인천광역시	송추	3.52	3.57	0.74	5.41	9.25	3.86	2.29	0.66	2.39
10	대전광역시	유성구	3.41	5.67	4.49	4.06	2.50	4.30	2.40	2.36	1.52
11	인천광역시	서구	3.41	3.18	6.88	7.74	4.29	1.77	1.19	1.21	1.03
12	경기도	남양주시	3.02	7.41	2.54	3.64	3.13	2.90	2.70	1.33	0.48
13	경기도	양평군	2.92	4.79	3.75	2.78	1.40	1.70	2.79	2.82	3.36
14	인천광역시	연수구	2.67	2.85	3.11	1.63	2.64	3.46	2.69	3.00	1.98
15	충청남도	아산시	2.46	3.41	3.52	2.17	2.35	2.40	1.29	1.74	2.81
16	충청북도	진천군	2.31	1.04	1.54	0.65	1.81	0.88	4.31	2.90	5.33
17	경기도	용인시	2.27	4.45	2.22	2.22	2.79	2.08	1.53	1.58	1.31
18	인천광역시	옹진군	2.14	2.84	3.98	2.84	3.29	0.02	1.25	1.86	1.04
19	광주광역시	광산구	2.09	5.92	3.38	2.55	2.02	1.12	0.87	0.57	0.27
20	경기도	평택시	2.03	2.30	1.77	1.74	1.78	1.70	2.44	2.24	2.27

...

〈인구 지속 증가지역 상위 20개 지역〉

■ 현황분석 결과

- 현황분석 결과를 토대로 분석 인사이트 도출
- 지자체 유형분류 등의 세목별 예측을 위한 추가 데이터 도출
- 지출/예산 세목 및 지출 지자체 특이값 도출



〈세목별 지출예측 및 예산비율 책정모델 세부 프로세스〉

■ 지출예측 모델링

- 필요한 데이터를 수집하여 전처리하고 다양한 예측 알고리즘을 적용하여, 가장 뛰어난 예측력을 보이는 모델을 선정
- Linear Regression, AdaBoost, Random Forest 등의 알고리즘을 사용하였으며, 그 결과 Random Forest 모델이 가장 좋은 것으로 파악됨

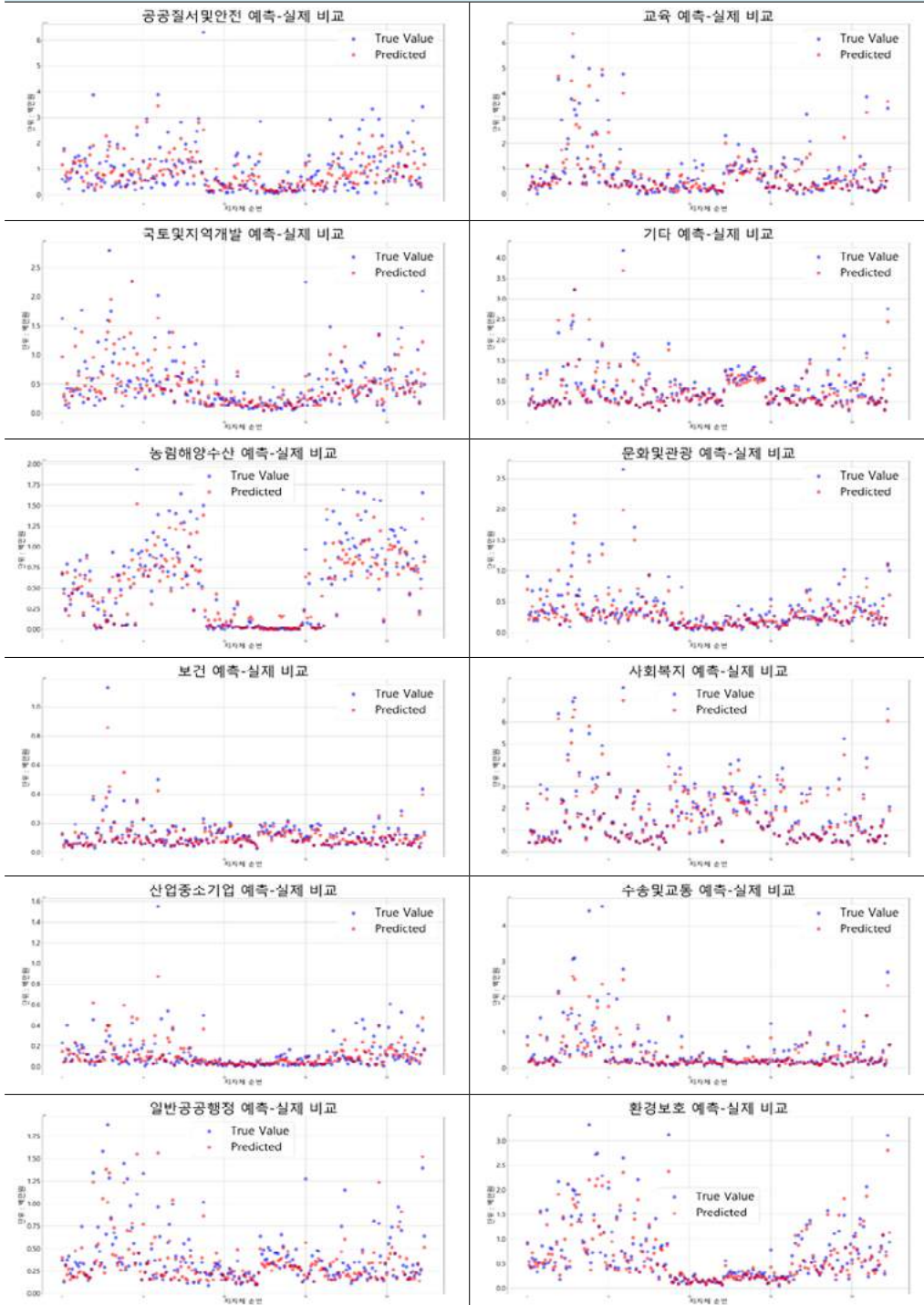


〈세목별 최적 알고리즘 정확도 검증결과〉

■ 인구추계

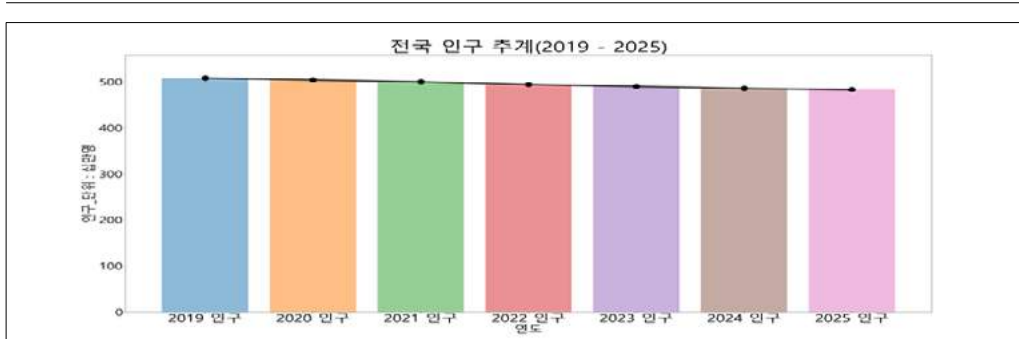
- 주민등록 연앙인구와 모의출산율 데이터로부터 지자체별/연도별/가임인구 연령별/지역별 미래출산율과 출생인구 추론
- 사망률 데이터로부터 지자체별/성별/연령별 생존율과 미래 생존인구 추론
- 순이동자수 데이터로부터 지자체별/성별/연령별 미래 순이동자수 추론

Random Forest 알고리즘 검증결과



〈세목별 최적 알고리즘 정확도 검증결과 시각화〉





〈인구추계 결과〉

■ 기타 데이터 추계

- 지출예측 모델링에 활용된 지자체 유형, 재정자립도, 농림수산 가구, 임야면적 등의 다양한 데이터에 대한 1차 년도, 3차 년도, 5차 년도 값 예측 실시
- 각 데이터별 특성 및 증감 추세를 확인하고, 해당 추세를 반영하여, 추계 실시
- 외국인 인구, 문화종사자수, 사업체수, 농림수산 가구 데이터에 대해서는 증감추세를 반영하여 데이터 추계
- 임야면적, 미세먼지 측정값, 하수도 사업비, 재정자립도, 지자체 유형은 최신수치를 이용하거나, 분석대상년도의 평균을 이용하여 데이터 추계 실시

■ 최종모델 생성

- 추계한 인구 데이터와, 기타 데이터를 이용하여 예측 분석을 위한 데이터마트 생성
- 데이터마트를 생성된 각 세목별 지출예측 모델에 대입하여 결과 도출



〈지자체별 지출예측 및 예산비율 도출 분석 세부 프로세스〉

분석결과 도출 및 시각화

■ 세목별 예측 금액 도출

- 분석대상 지자체 및 연도를 지정하여 세목별 지출예상금액을 도출하고, 해당 결과 시각화 실시

■ 연차별 증감율 변화 추이

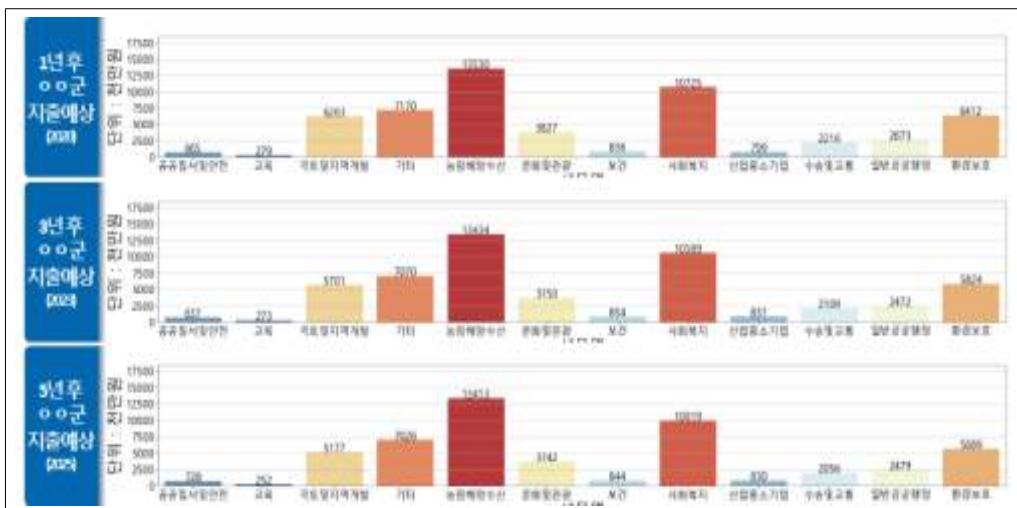
- 2019년도 세목별 예상 지출금액 대비 1차 년도, 3차 년도, 5차 년도의 예상 지출금액 증감율을 계산하고 시각화 실시

■ 세목별 예산 추천 비율

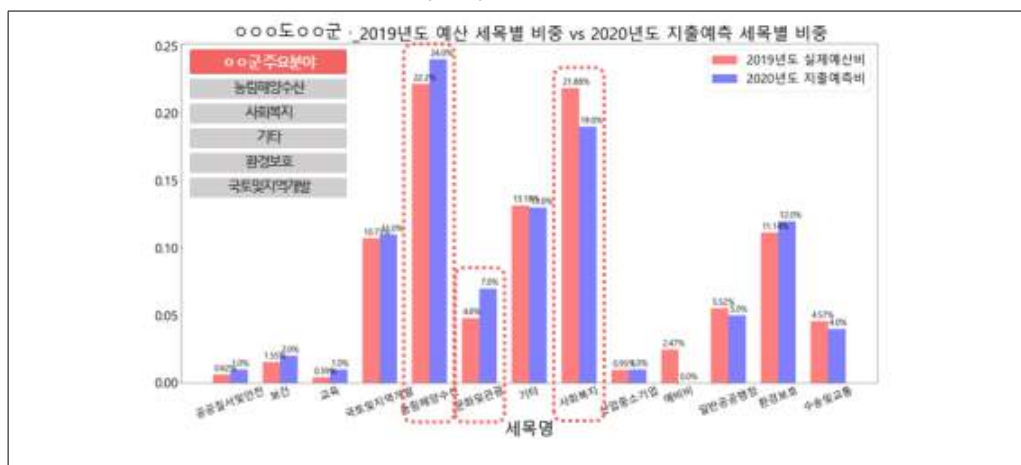
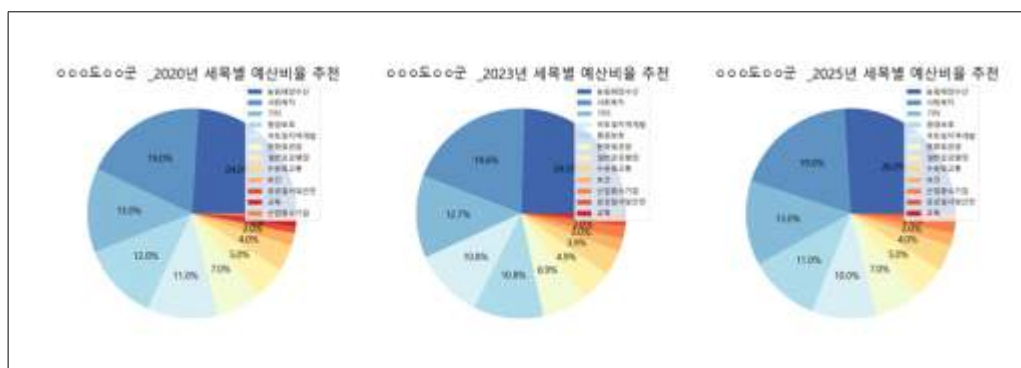
- 예측된 지출예상금액을 바탕으로 분석 대상 지자체의 1차 년도, 3차 년도, 5차 년도 세목별 예산 추천 비율 추천
- 해당 결과물로부터 분석 대상 지자체의 주요 세목 도출

■ 세목별 지출예측 비중 비교

- 2019년도 실제 예산 대비 1차 년도, 3차 년도, 5차 년도의 예산 추천 비중 비교
- 세목별 예산 추천 비중에서 도출한 주요 세목 위주로 확인



〈분석대상 지자체 1차/3차/5차 년도 세목별 지출예측금액 비교〉





데이터 분석 단계

1. 데이터 분석 단계 요약

단계	주요 내용
분석 설계	<ul style="list-style-type: none"> ■ 분석 범위 <ul style="list-style-type: none"> • 공간적 범위 : 전국 기초자치단체 • 시간적 범위 : 세목별 지출예측 모델(2015년 ~ 2017년) 인구추계 모델(2009년 ~ 2018년) • 분석 단위 : 과학기술 분야를 제외한 12개 세목 ■ 분석 모형 <ul style="list-style-type: none"> • 다양한 분석 알고리즘을 활용한 세목별 지출 예측 • 코호트 요소방법을 이용한 인구 추계
데이터 수집 및 가공	<ul style="list-style-type: none"> ■ 지방재정365를 통해 공개된 예산/지출 공공데이터 수집 ■ 지자체 유형 데이터의 경우 지방재정365에 공개된 지방재정 분석보고서의 내용 확인 ■ 국가통계포털 웹을 통하여 공공데이터 수집 ■ 현황분석을 통해 데이터 특이사항을 확인하고 공백처리 및 특이값 제거 ■ 지자체명 최신화 및 다양한 데이터를 지자체명 기준으로 결합
데이터 분석 및 검증	<ul style="list-style-type: none"> ■ 세목별 지출예측모델 <ul style="list-style-type: none"> • 2015년과 2016년 데이터를 바탕으로 다양한 알고리즘을 활용하여 분석모델을 만들고, 2017년도 데이터로 모델 검증 과정을 거쳐 최적의 모델 선정 ■ 데이터 추계 <ul style="list-style-type: none"> • 세목별 지출예측모델에 활용된 데이터들의 미래 추계 • 인구 데이터는 코호트 요소 방법을 이용하여 미래 추계 • 기타 데이터는 현재 추세를 반영하여 미래 추계 ■ 지자체별/연도별/세목별 지출 예측 <ul style="list-style-type: none"> • 추계 데이터를 이용하여 데이터마트 구성 • 데이터마트 내 데이터를 세목별 지출예측모델에 적용하여 지자체별/연도별/세목별 지출금액 예측 실시
데이터 시각화	<ul style="list-style-type: none"> ■ 세목별 지출 예측 및 세목별 추천 예산 비율 ■ '19년도 대비 세목별 비중 및 세목별/연도별 금액 증감비교



2. 분석 설계

■ 세목별 지출예측모델

- 다양한 데이터와 다양한 알고리즘을 활용하여 최적의 지출예측모델 설계

■ 활용 데이터 추계

- 지출예측모델에 활용되는 다양한 데이터의 특성을 확인하고, 미래 데이터 추계방법 설계

과제	<ul style="list-style-type: none"> ■ 세목별 지출예측모델 ■ 활용 데이터 추계 ■ 지자체별/연도별/세목별 지출 예측
추진내용	<ul style="list-style-type: none"> ■ 세목별 지출예측 모델 <ul style="list-style-type: none"> • 지출데이터와 지출예측을 위한 다양한 데이터 결합 <ul style="list-style-type: none"> - 지자체명 기준으로 전체 데이터를 결합 - 최신 지자체명을 기준으로 변경된 지자체명 현행화 • 데이터 전처리 <ul style="list-style-type: none"> - 지자체별/연령별/성별 데이터에 대해 전처리를 실시하여, 경제활동가능인구와 비경제활동인구 변수로 분리 - 지자체 유형 데이터를 예측모델에서 활용하기 위하여 dummy변수로 제작 • 훈련용 데이터와 검증용 데이터 분리 <ul style="list-style-type: none"> - 2015, 2016년 데이터를 훈련용 데이터로 이용 - 2017년 데이터를 검증용 데이터로 이용 • 세목별로 다양한 알고리즘 적용 <ul style="list-style-type: none"> - 각 세목별로 Linear Regression, Random Forest, AdaBoost 등의 다양한 예측 알고리즘 적용 - 가장 좋은 예측력을 보여준 Random Forest 모델을 최종 모델로 선정 ■ 활용 데이터 추계 <ul style="list-style-type: none"> • 세목별 지출예측 모델을 만드는 과정에서 활용한 데이터 정리 • 인구데이터 추계 <ul style="list-style-type: none"> - 코호트 요소 방법을 이용하기 위해 현재인구, 출산율, 사망률 및 순이동자수 데이터 수집 - 지자체별/연령별 가임인구와 지자체별/연령별 출산율을 지자체명 및 연령을 기준으로 결합하고 회귀분석을 통해 미래 출산율 및 미래 출생인구 추계



추진내용	<ul style="list-style-type: none"> - 지자체별/연령별/성별 사망률을 이용하여, 지자체별/연령별/성별 생존율을 계산하고 미래 지자체별/연령별/성별 생존인구 추계 - 지자체별/연령별/성별 순이동자수의 경우 최신 추세가 미래에 유지될 것이라고 가정하여, 미래 지자체별/연령별/성별 순이동자수 추계 - 추계된 데이터와 인구데이터의 연령대를 표준화하고, 지자체명/연령/성별로 맵핑을 실시하여, 코호트 요소 방법을 통해 지자체별/연령별/성별인구 추계 실시 • 기타 데이터 추계 - 인구데이터 외 세목별 지출예측모델에서 활용된 다양한 데이터에 대한 추계 실시 - 각 데이터별로 추세를 확인하고, 해당 추세를 반영하여 예측 ■ 지자체별/연도별/세목별 지출 예측 <ul style="list-style-type: none"> • 활용 데이터 추계과정에서 생성된 인구추계데이터 및 기타 추계데이터를 지자체명을 이용하여 맵핑하고, 데이터마트로 구축 • 해당 데이터마트를 세목별 지출예측 모델에 적용하고, 각 지자체별/연도별/세목별 지출금액 예측결과 도출 • 각 지자체별로 “연도별/세목별 지출예상금액”, “2019년도 예상 지출 대비 연차별 증감률 변화 추이”, “연도별/세목별 예산 추천 비중”, “2019년도 예산 세목별 비중과 미래 연도별 세목별 지출예측 비중 비교” 시각화 자료 생성
Tip	<ul style="list-style-type: none"> ■ 분석설계 과정에서 가장 중요한 것은 유관기관 및 해당 모델 활용자가 원하는 방향 및 형태로 분석 결과를 보여주는 것임 ■ 모델 활용자가 필요로 하는 결과물의 형태를 파악하고, 분석 모델에 적용해야만 실제 해당 모델의 활용도를 높일 수 있음

3. 데이터 수집 및 가공

■ 분석결과 시각화 및 유관기관, 관련부서 담당자에게 제공

과제	■ 분석에 필요한 데이터(18종)를 수집 및 가공		
추진내용	데이터명	기간	출처
	구조별 기능별 세출결산	2015~2017	지방재정365
	구조별 기능별 세출예산	2015~2019	지방재정365
	지방재정분석 지자체 유형	2019	지방재정365
	재정자립도	2015~2019	지방재정365
	시군구,성,연령(5세)별 주민등록연앙인구	2009~2018	국가통계포털
	행정구역 읍면동별(5세)별 주민등록인구	2019.07	국가통계포털
	시군구,성,연령(5세)별 순이동자수	2018	국가통계포털
	시군구 합계출산율 모의 연령별 출산율	2009~2018	국가통계포털
	시군구,성,연령(5세)별 사망률	2018	국가통계포털
	전검업별 농가	2010, 2015	국가통계포털
	전검업별 어가(해수면)	2010, 2015	국가통계포털
	전검업별 어가(내수면)	2010, 2015	국가통계포털
	전검업별 임가	2010, 2015	국가통계포털
	시군구별 산업소분류별 총괄	2010, 2015	국가통계포털
	행정구역별 면적 및 축적 산림	2015	국가통계포털
	면적 소유별, 관리기관별, 시군구별	2010	국가통계포털
	도시별 재정현황 하수도	2015~2017	국가통계포털
	미세먼지 PM2.5	2015~2019	국가통계포털
	시군구별 외국인 주민 현황	2009~2018	국가통계포털
■ 데이터 가공			
<ul style="list-style-type: none"> 구조별 기능별 세출 예산/결산 <ul style="list-style-type: none"> 변경된 지자체명을 반영하여 지자체명 최신화 연도별/세목별/광역자치단체별/기초자치단체별로 데이터 정렬 기초자치단체가 없는 세종/제주 및 광역자치단체 본청 제외 지자체 유형분류 <ul style="list-style-type: none"> 명목형 변수이므로, 분석에 활용하기 위해 dummy변수화 주민등록 연앙인구/순이동자수/사망률 <ul style="list-style-type: none"> 승격 지자체, 통합 지자체, 명칭 변경 지자체 반영 			



추진내용	<ul style="list-style-type: none"> • 전·겸업별 농가/어가(해수면)/어가(내수면)/임가 <ul style="list-style-type: none"> - 각 데이터들을 지자체명을 이용하여 결합 - 결합된 데이터를 합쳐 1차 산업가구로 변경 - 총조사 데이터이므로 5년 단위로 조사되며, 따라서 2016, 2017년 데이터는 2010년과 2015년의 추세로 예측 • 시군구별 사업소분류별 총괄 <ul style="list-style-type: none"> - 전체 항목별 사업체수에서 농림수산업, 관광 및 문화 관련 업종을 제외하여 사업체수 데이터 제작 - 관광 및 문화 관련 업종의 종사자수를 바탕으로 관련 데이터 제작 - 총조사 데이터이므로 5년 단위로 조사되며, 따라서 2016, 2017년 데이터는 2010년과 2015년의 추세로 예측 • 행정구역별 면적 및 축적(산림)/면적 소유별, 관리기관별, 시군구별 <ul style="list-style-type: none"> - 지자체명을 이용하여 맵핑을 실시하고, 추세를 바탕으로 2016, 2017년도 데이터 예측 • 도시별 재정현황 하수도 <ul style="list-style-type: none"> - 지자체명을 최신화하여 맵핑 실시 • 미세먼지 PM2.5 <ul style="list-style-type: none"> - 전국 모든 기초자치단체 데이터를 수록하고 있지 않음 - 공백 값은 동일한 지자체 유형분류를 가진 기초자치단체와, 동일한 광역자치단체 소속의 기초자치단체를 이용하여 평균 대치 • 시군구별 외국인 주민 현황 <ul style="list-style-type: none"> - 국내 거주 외국인의 현황을 파악하기 위해, 해당 항목에 해당하는 외국인 항목만 사용하고, 해당 합계를 구하여 데이터로 이용
Tip	<ul style="list-style-type: none"> ■ 보유 데이터에 맞춰 분석을 수행하는 것이 아니라, 분석모델을 설정하고 빅데이터 분석모델에 필요한 데이터를 정의해야함

4. 데이터 분석 및 검증

과제	<ul style="list-style-type: none"> • 분석 분야에 맞는 세부 분석모델을 설정하고 분석 실시 • 사업 담당자와의 지속적인 논의를 통하여 분석모델 방향 점검 및 결과물 검토 실시
추진내용	<ul style="list-style-type: none"> ■ 분석 대상 확정 <ul style="list-style-type: none"> • 분석 대상은 전국 기초자치단체이며, 기초자치단체가 없는 제주, 세종과 각 광역자치단체별 본청은 패턴이나 예측모델을 만들기에 데이터가 부족하여 분석대상에서 제외함 ■ 분석 데이터 확정 <ul style="list-style-type: none"> • 현황분석 과정에서, 세목별 특징과 각 지자체별 특징에 따른 예/결산 및 인구 패턴을 확인하고, 이를 토대로 예산/지출/인구 분석 데이터 외 재정자립도, 지자체 유형, 농림수산물가공 등의 데이터 추가/확정 ■ 세목별 지출예측 모델 <ul style="list-style-type: none"> • 지자체별/연도별/세목별 지출 데이터를 종속변수로, 인구와 기타 현황분석 결과 도출된 데이터를 독립변수로 하여 지출예측 모델링 실시 • 2015, 2016년도 데이터를 훈련 데이터로, 2017년도 데이터를 검증 데이터로 활용 • 각 연도별/세목별로 이상치를 갖는 지자체 데이터 제거 • 과학기술 세목은 226개의 분석대상 지자체 중 10개의 지자체에서만 실제로 집행되고 있고, 해당 지자체들에서도 작은 비중을 차지하므로, 분석대상에서 제외 • 다양한 알고리즘을 활용하여 분석을 실시하며, 각 결과물을 확인, 최적의 분석 모델 도출 ■ 지자체별/연도별/연령별/성별 인구추계 <ul style="list-style-type: none"> • 각 지자체별 인구데이터에서 가임여성 인구를 추출하고, 추출한 가임여성 인구 데이터와 각 지자체별/연령별 출산율 데이터를 결합하여 출생인구 계산 • 출산율 데이터에 대해 회귀분석을 실시하여, 미래 출산율 계산 • 지자체별/연도별/연령별/성별 사망률을 이용하여 지자체별/연도별/연령별/성별 생존율 도출 및 생존인구 계산 • 지자체별/연도별/연령별/성별 생존율 및 순이동자수 데이터는 최신 데이터가 유지된다고 가정 • 현재의 주민등록인구 데이터를 코호트 요소방법에 적용하여 인구추계 실시 (1차 년도, 3차 년도, 5차 년도 추계) • 코호트 요소방법 : $P_{t+1} = \sum_{i=1}^N [P_t^i - D_{t,t+1}^i + M_{t,t+1}^T] + B_{t,t+1}$

추진내용

- 기타 데이터 추계
 - 인구데이터 이외 세목별 지출예측모델에 활용하기 위한 데이터에 대해 각 데이터 특성에 맞게 추계
 - 외국인 데이터, 문화종사자 데이터, 사업체 데이터, 농림수산 가구 데이터의 경우 추세를 반영하여 미래 데이터 추계
 - 임야, 하수도, 미세먼지, 재정자립도, 지자체 유형 데이터는 분석기간 평균 수치나 최신 수치 유지
- 지자체별/연도별/세목별 지출 예측
 - 추계된 분석 데이터를 지자체별/연도별(1차 년도, 3차 년도, 5차 년도)로 결합하고, 세목별 지출모델에 적용하여 세목별 지출 예측
 - 각 지자체별/연도별/세목별 지출 예측을 실시하고 해당 결과를 각 지자체별로 “세목별 지출예측”, “세목별 추천 예산비율”, ’19년도 실제예산비중 대비 추천 세목별 비중”, ’19년도 대비 연도별 비교”등의 시각화된 자료 제공



5. 데이터 시각화

과제	<ul style="list-style-type: none"> ■ 예산/지출/인구데이터 현황분석 ■ 인구추계 모델 ■ 지자체별/연도별/세목별 예측
추진내용	<ul style="list-style-type: none"> ■ 현황분석 결과 <ul style="list-style-type: none"> • 현황분석 결과 <ul style="list-style-type: none"> - 예산-결산 비교, 예산, 결산, 인구데이터에 대해 현황분석을 실시하는 과정에서 시각화 결과물 제시 ■ 인구추계 결과 <ul style="list-style-type: none"> • 인구추계 결과 <ul style="list-style-type: none"> - 2019년도 전국인구 대비 전국인구 추계 시각화 결과물 도출 - 2019년도 전국인구/성별/연령분포 대비 추계된 전국인구/성별/연령분포 시각화 결과물 도출 ■ 지자체별/연도별/세목별 예측 결과 <ul style="list-style-type: none"> • 분석결과 시각화 <ul style="list-style-type: none"> - 지자체별/연도별/세목별 예측된 지출금액을 활용하여, 관계자의 인사이트 도출을 위한 적합한 형태의 시각화 결과물 제시 • 시각화 결과물 종류 <ul style="list-style-type: none"> - 지자체별/연도별/세목별 예측된 지출금액 시각화 - 지자체별/연도별/세목별 추천 예산비율 시각화 - 지자체별/연도별/세목별 19년도 예산대비 예측연도 예산추천비율 비교 시각화 - 지자체별/연도별/세목별 19년도 예상지출금액 대비 예측연도 지출예산금액 증감 비교 시각화 - 지자체별/연도별 총 추계인구/경제활동가능 추계인구/비경제활동 추계인구 증감 시각화 분석 결과 저장 ■ 분석 프로그램 상에 분석 결과물을 로컬 내 지정 폴더에 저장
Tip	<ul style="list-style-type: none"> ■ 빅데이터 분석 결과를 시각화하는 과정에서, 필요한 시각화 자료인지, 민감한 자료는 아닌지 등에 대해 내부 협의가 필요함 ■ 유관기관과 및 담당 부서와의 사전 협의를 통해 시각화 결과물의 형태 결정 필요



행정적용 단계

1. 행정적용 단계

- 본 분석은 공공데이터만을 이용하여 분석을 진행한 시범분석과제임
- 본 분석을 실제 분석 모델로 설계할 경우, 지자체 보유데이터(ex : 지자체별 공동주택 준공일/입주일, 소방인력 데이터, 경찰인력 데이터, 소방장비 및 시설 등) 민간데이터(ex : 소비 데이터, 교통 데이터, 유동인구 데이터)등을 확보하면 예측 정확도 상승을 기대할 수 있음
- 추후 본 분석 과정에서 상기한 데이터를 추가하여 고도화하고, 실제 지자체별 예산수립 관계자의 의사결정 지원 가능

2. 행정정책개선

- 기존의 예산 편성과 지출 내역을 확인하고, 예산상 초과분 및 부족분을 검토하여 예/결산 현황 파악
- 빅데이터 기반의 효율적 예산 수립 지원
- 효율적 방식의 예산수립을 통해 효과적인 지방재정 운영 도모





[2019년 KLID 신규분석모델 ②]

도로 미세먼지 발생량 예측 및 노면 청소차량 운행 효율화

| 한국지역정보개발원, 평택시청, 펜타시스템테크놀로지(주)컨소시엄 |

요약

2019년 한국지역정보개발원에서는 『KLID 소관 정보시스템 보유데이터 현황조사 및 분석모델 개발』 사업을 통해 전국 지자체에서 공통적으로 적용 가능한 신규 분석모델을 발굴하였다.

그 중 두 번째 과제인 『도로 미세먼지 발생량 예측 및 노면 청소차량 운행 효율화』 모델의 경우, 기존 미세먼지 분석은 공기 중에 있는 대기미세먼지 분석이 다수였으나, 본 분석은 도로 미세먼지 발생량 분석을 진행 후 노면 청소차량의 운행 효율화에 목적을 두고 추진되었다.

전국 미세먼지 배출원별 기여율의 결과를 확인한 결과, 도로 미세먼지의 기여율이 가장 높았으며, 도로 미세먼지를 감소시키기 위해 각 지자체에서는 많은 방법을 활용하여 도로 미세먼지 줄이기에 노력하고 있다.

주요 분석 내용으로는 첫째, 도로 교통량, 평균속도, 도로 길이 등의 데이터를 활용하여 현황 분석 및 미세먼지 발생량을 예측하고, 둘째, 예측한 값을 바탕으로 평균 미세먼지 발생량 추세 및 패턴을 분석하였으며, 셋째, 각 데이터를 노선 설계 프로세스에 대입하여 노면 청소차량의 노선을 설계하고 시각화하였다.



추진배경

- 전국 미세먼지 배출원별 기여율 확인 결과 도로 재비산먼지의 기여율이 가장 높으며, 도로 미세먼지를 감소시키기 위하여 각 지자체에서는 많은 방법을 활용하고 있음
- 도로 재비산먼지는 자동차 배기가스, 타이어 마모, 브레이크 패드 마모 등에 의해 도로 위 침적된 먼지가 차량의 이동에 의해 대기 중으로 재비산되는 입자상 물질을 의미함
- 대부분의 지자체에서는 미세먼지 발생량을 고려하지 않은 채 노면 청소차량을 운행하고 있음
- 현재 전국적으로 미세먼지 관련 이슈가 증가하고 있으며, 본 분석에서는 지자체 보유 데이터를 이용하여 미세먼지 발생량을 예측하고, 예측된 데이터로 효율적인 노면 청소차량의 노선 설계를 진행
- 미세먼지 발생량 분석을 통하여 설계된 노선을 노면 청소차량에 적용시켜 미세먼지 발생량을 효과적으로 감소시키고자 함

34



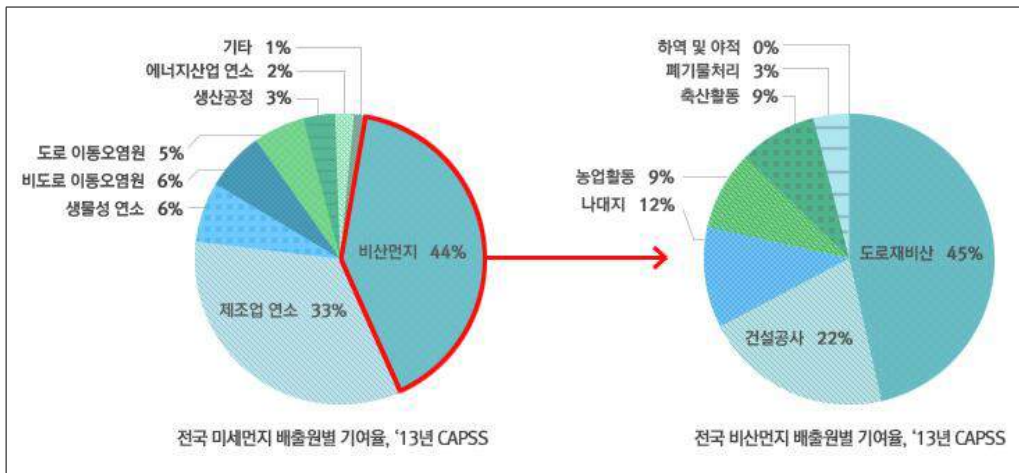
프로젝트 개요

■ 분석개요

- 분석기간 : 2019.10 ~ 12
- 분석데이터 : 평택시 교통정보센터의 평균속도 및 교통량 데이터, 전국 표준 노드 · 링크

■ 분석의 필요성

- 지자체에서는 도로 재비산먼지 측정자료를 바탕으로 효율적 도로환경 관리를 통해 도로 재비산먼지를 감소시켜야 함(환경부, ‘수도권 대기환경 개선에 관한 특별법’)
- 이를 위해, 환경부에서는 집중관리도로 지정, 청소차량 운행횟수 증가 등의 정책을 추진 중에 있음
- 효과적인 도로 미세먼지 발생량 감소를 위해서는 데이터 기반의 과학적 분석을 통한 노면 청소차량 운행 경로 설계가 필요함



〈도로 재비산먼지 배출 현황〉

- 각 지자체에서는 최근 건강과 환경에서 대두되고 있는 미세먼지를 감소시키기 위한 정책을 수립해야함
- 기존 노면 청소차량의 운행 노선은 미세먼지 발생량을 고려하지 않은 운행노선으로 운행 중임
- 본 분석에서는 미세먼지를 감소시키기 위한 방안으로, 과거 미세먼지 발생량 데이터 기반의 효과적인 노면 청소차량 운행노선 설계방안 제시함

■ 주요 분석내용

- 교통량 및 도로 길이 등의 데이터를 활용하여 미세먼지 발생량 예측
- QGIS를 활용한 도로 각 구간의 미세먼지 발생량 시각화
- 미세먼지 발생량과 도로 소통 판정을 활용하여 노면 청소차량의 노선 설계



적용 프로세스

■ 분석 전체 프로세스

- 기획관리(문제의식 → 사업계획 → 완료보고)
- 데이터분석(분석 설계 → 데이터 수집 및 가공 → 데이터 분석 → 데이터 시각화)
- 행정적용(시범적용)





라 기획·관리 단계

1. 기획·관리 단계 요약

단계	주요 내용
추진내용	<ul style="list-style-type: none"> 미세먼지 발생량을 고려하지 않은 기존 운행 노선에 대하여 데이터 기반의 효율적인 운행 노선으로 설계가 필요함
사업계획	<ul style="list-style-type: none"> 도로 각 구간의 현황 분석 <ul style="list-style-type: none"> 분석 기간의 평균 미세먼지 발생량 추세 및 패턴 분석 도로 미세먼지 발생량 예측 <ul style="list-style-type: none"> 교통량 및 도로 길이 등을 사용하여 미세먼지 발생량 예측 시간대별 미세먼지 발생량을 도로 각 구간별로 파악하여 노선 설계의 기초 자료로 활용 효율적 운행 노선 설계 <ul style="list-style-type: none"> 도로 각 구간의 현황과 미세먼지 발생량을 바탕으로 효율적 운행 노선 설계 데이터마트 생성 및 도로 미세먼지 발생량 예측/시각화
완료보고	<ul style="list-style-type: none"> 착수보고, 중간보고, 완료보고 형태로 보고회 진행 각 보고회 추진 시, 자문위원, 부서 담당자 등 의견 수렴 실시

37



2. 문제의식

■ 데이터 기반의 효율적인 운행 노선 설계가 필요

추진과제	<ul style="list-style-type: none"> 문제점 확인 문제해결을 위해 필요한 활동
추진내용	<ul style="list-style-type: none"> 문제점 확인 <ul style="list-style-type: none"> 각 지자체에서는 최근 건강 및 환경에 문제가 되는 미세먼지를 감소시키기 위하여 다양한 노력을 기울이고 있음 현재 노면 청소차량의 운행은 민원을 기반으로 운행되며, 미세먼지 발생량을 고려하지 않고 운행됨 문제해결을 위해 필요한 활동 <ul style="list-style-type: none"> 미세먼지 발생량 현황분석 및 예측 효과적인 노면 청소차량의 운행 노선 설계
Tip	<ul style="list-style-type: none"> 데이터 분석을 통한 문제해결 가능 여부를 파악하는 것이 중요함 유관기관에서 보유한 데이터 수집 가능 여부 및 데이터 정제 작업 소요기간 검토 필요

3. 사업계획

- 사업에 대한 요구사항을 정의하여 계획을 수립하고 확정하는 단계
- 사업범위를 결정하고 사업내용을 구체화함
 - 사업계획 수립 시 현황 및 문제점, 사업내용 등에 대한 정의가 필요

과제	<ul style="list-style-type: none"> ■ 구체적이고 명확한 사업계획 수립 필요 ■ 현황분석을 통한 노선설계와 도로 미세먼지 감소 	
추진내용	<ul style="list-style-type: none"> ■ 유관기관 인터뷰 및 과제 도출 <ul style="list-style-type: none"> • ('19.09.) KLID 민원정보부 인터뷰 및 후보과제 도출 • ('19.10.) 분석과제 서면자문 실시 및 분석과제 선정 <ul style="list-style-type: none"> - 자문위원단 및 지자체 실무자 대상 서면자문 실시 • ('19.10.) 분석대상 지자체 선정 • ('19.10.) 경기도 평택시 관계자 인터뷰 실시 ■ 분석 데이터 확보(~'19.11.) 	
	외부 데이터	<ul style="list-style-type: none"> · 교통량 데이터(평택시 교통행정과) · 교통 속도 데이터(평택시 교통행정과) · 살수차 운행노선(평택시 자원순환과) · 임대 살수차 운행일지(평택시 환경정책과)
	<ul style="list-style-type: none"> ■ 분석 실시(~'19.12.) 	
Tip	<ul style="list-style-type: none"> ■ 내부 데이터에도 실제로 빅데이터 분석에서 활용될 수 없는 데이터가 있을 수 있으므로, 이를 사업계획 단계에서 파악하는 것이 중요 ■ 초기에 발견하지 못할 경우, 사업 진행과정에서 분석에 필요한 데이터가 부족하거나, 활용 불가하여 분석을 진행할 수 없는 경우가 발생할 수 있음 	



4. 완료보고

■ 분석 결과물 활용에 대한 공감대 형성을 위한 의사소통의 방법으로 완료보고회를 실시

■ 도로미세먼지 발생량을 반영하여 노면 청소차량의 효율적 운행 노선 설계

과제	<ul style="list-style-type: none"> ■ 도로 미세먼지 발생량 예측 및 노면 청소차량 운행 효율화
추진내용	<ul style="list-style-type: none"> ■ 분석 대상 지자체 및 대상 도로 확정 <ul style="list-style-type: none"> • 평택시 전 도로 중 평택시청 관할 지역으로 도로 확정 ■ 속도별 배출계수 생성 ■ 도로 미세먼지 발생량 데이터 마트 생성 ■ 평택시청 관할 도로 각 구간별 현황 분석 ■ 도로 각 구간별 미세먼지 발생량 예측 <ul style="list-style-type: none"> • 교통량, 도로 길이, 배출계수를 활용한 미세먼지 발생량 예측 ■ 시간별 도로 각 구간의 미세먼지 발생량 시각화 ■ 운행 노선 설계 프로세스 적용 ■ 노선 설계 시각화 ■ 완료보고 (19.12) <ul style="list-style-type: none"> • 사업 완료보고, 자문 및 의견 수렴 등
Tip	<ul style="list-style-type: none"> ■ 분석과제에 대한 정책개발 및 적용방안을 도출하기 위해서는 유관기관 및 담당부서와의 긴밀한 협조가 필요함

5. 분석결과

추진 프로세스

■ 추진 프로세스

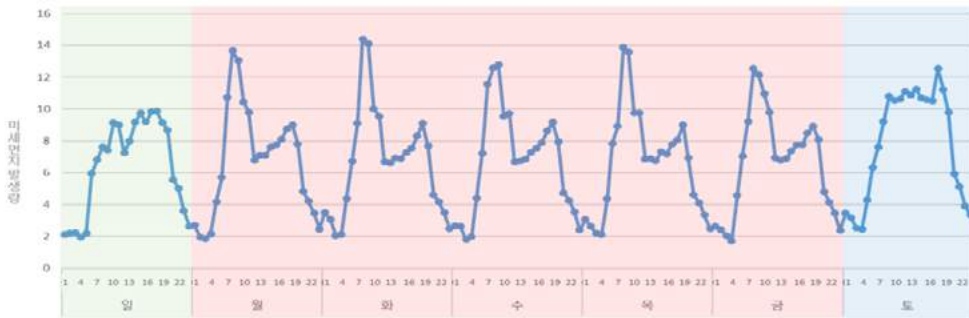


- 본 분석은 데이터수집, 데이터 전처리, 현황분석, 노선설계 순으로 이루어짐
- 데이터 수집 단계에서는 지자체로부터 분석범위내의 도로 구간별 교통량데이터와 구간별 속도데이터 수집을 진행함
- 현황분석 단계에서는 평일의 시간별, 구간별 미세먼지 발생량 분석을 진행함
- 노선설계 단계에서는 현황분석 결과를 바탕으로 평일의 시간별, 구간별 미세먼지 발생량 예측을 하고, 이를 바탕으로 시간대별 노선설계를 진행함

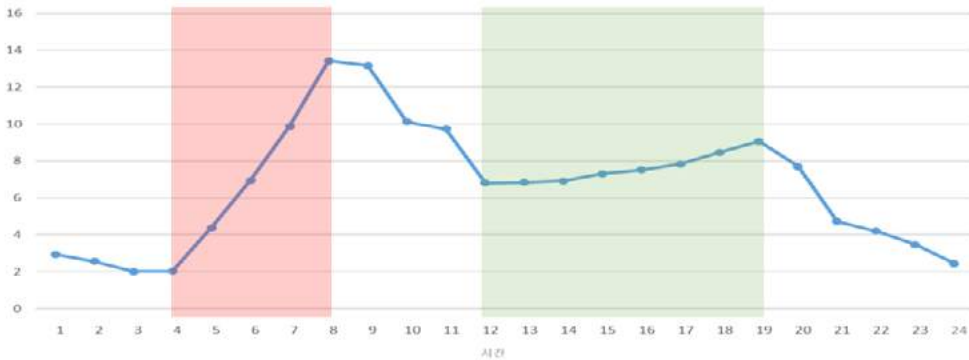


전체 미세먼지 발생 현황

■ 요일별 시간별 평균 미세먼지



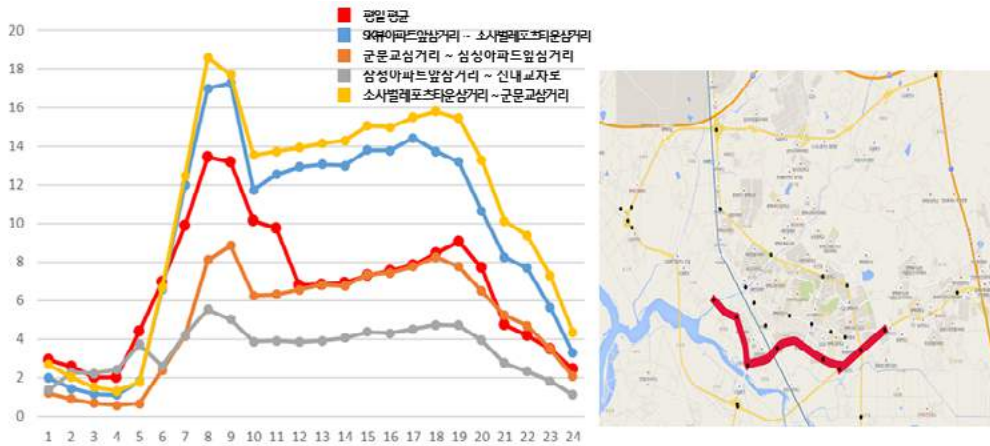
- 요일별 시간별 미세먼지 발생 현황을 나타내어 요일별 패턴을 도출함
- 월, 화, 수, 목, 금(평일)의 경우 시간에 따른 미세먼지 변화 패턴이 비슷하며, 토요일과 일요일의 경우 각각 다른 패턴을 보임
- 본 분석에서는 미세먼지 발생량을 예측하기 위해 표본수가 많은 평일을 지정하여 분석을 진행함



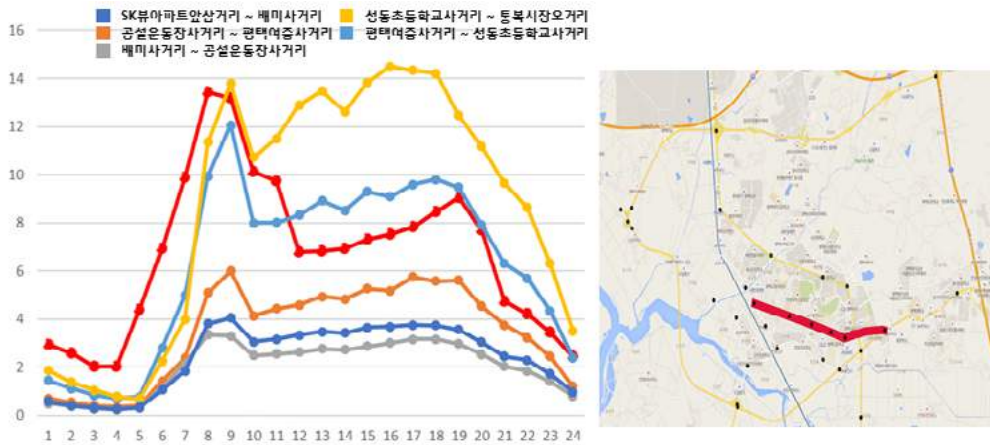
- 월요일부터 금요일까지(평일) 모든 구간의 시간별 미세먼지 발생량 평균을 나타냄
- 4시부터 8시까지 급격한 증가를 보이며, 8시에 최고 미세먼지 발생량을 나타내며, 12시까지 감소하는 추세를 보임
- 12시부터 19시까지 차츰 증가하며 19시부터 급격한 감소가 시작되어 3시에 가장 낮은 미세먼지 발생량을 보임

시간별 구간별 미세먼지 발생량 현황 분석

도로별 구간별 시간별 미세먼지



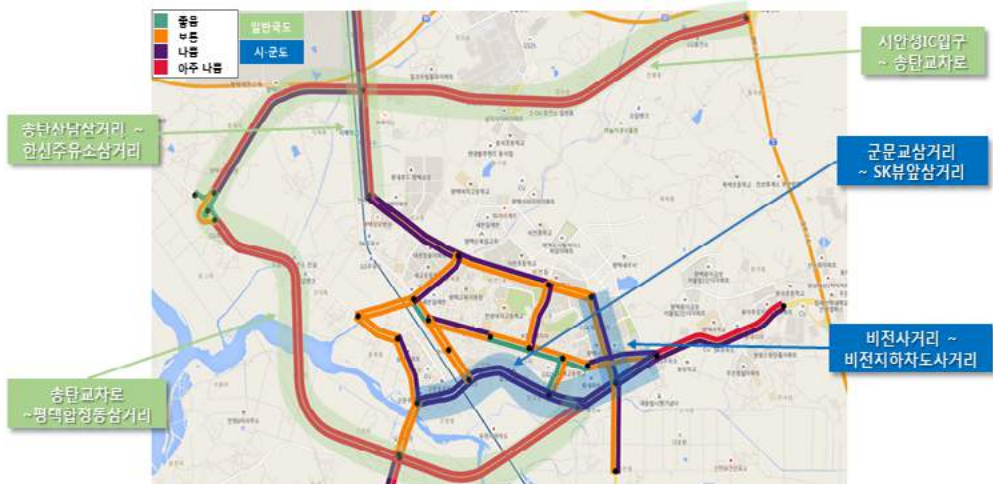
〈평남로 구간별 시간별 미세먼지〉



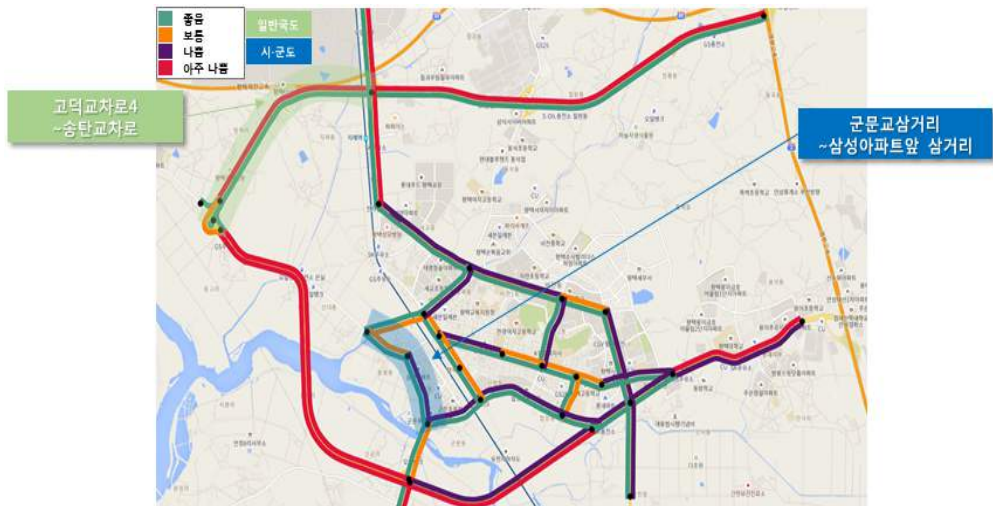
〈중양로 구간별 시간별 미세먼지〉

- 평남로와 중양로의 각 구간별 시간별 미세먼지 현황을 꺾은선 그래프로 나타냈음
- ‘일반국도 1호선’, ‘일반국도 38호선’, ‘일반국도 45호선’, ‘경기대로’ 4개의 일반 국도와 ‘평택로’, ‘평택4로’, ‘평남로’, ‘중양로’, ‘동작로’, ‘공설운동장길’, ‘팽성로’ 7개의 시·군도의 각 구간별 시간별 미세먼지 현황 분석을 진행하였음

■ 분석대상도로 시간별 미세먼지 현황



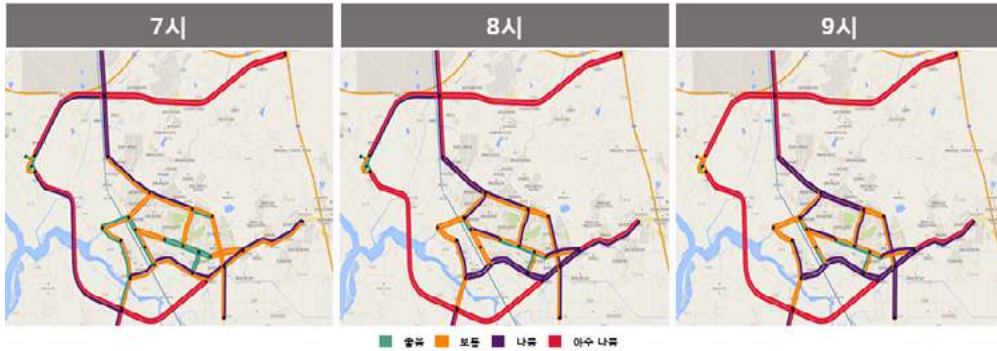
〈분석 대상도로 8시 미세먼지 현황〉



〈분석대상 도로 18시 미세먼지 현황〉

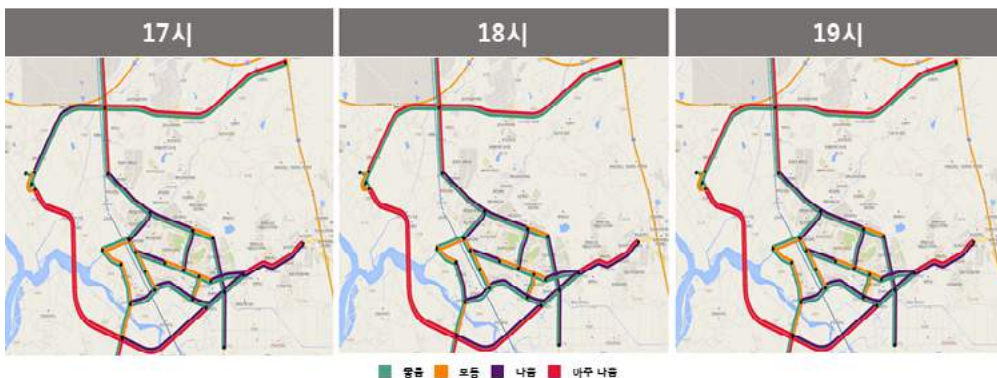
- 본 분석의 분석대상도로 각 구간별 1시부터 24시까지 미세먼지 현황을 나타냄
- 분석 대상 도로의 시간 변화에 따른 미세먼지 현황을 볼 수 있음
- 서안성IC입구 ~ 신궁입체 구간의 외곽도로 미세먼지가 계속해서 나쁜 것으로 나타나며, 노면청소차량 외 다른 대책 수립이 필요 할 것으로 보임
- 본 사업 보고서에는 1시부터 24시까지 그림을 수록하였음

■ 오전 주요시간 미세먼지 현황



- 7시 ~ 8시 : 외곽에 있는 도로의 대부분이 '나쁨'과 '아주 나쁨'으로 유지하고 있으며, 시내로 유입되는 도로의 대부분은 '나쁨' 단계로 변함
- 8시 ~ 9시 : 외곽에 있는 도로와 시내로 유입되는 도로보다는 대부분이 변화가 없으며, '보통' ~ '아주 나쁨'의 단계를 유지하고 있음

■ 오후 주요시간 미세먼지 현황



- 17시 ~ 19시 : 외곽에 있는 도로의 대부분이 '나쁨'과 '아주 나쁨'으로 유지하고 있으며, 시내로 유입되는 도로의 대부분은 '나쁨' 단계로 변함
- 외곽에 있는 도로와 시내로 유입되는 도로의 대부분이 단계 변화가 없고 '좋음'~'아주 나쁨' 단계가 고루 분포 되어 있음
- 시내로 유입되는 도로보다는 외곽에 있는 도로 미세먼지 발생량의 단계가 나쁜 것으로 확인 됨
- 일반 국도 38호선 전체 구간의 경우 17시~19시 까지 '아주 나쁨'을 계속 유지하고 있음

■ 시·군도 미세먼지 순위

[illegible]

- 시·군도의 1시부터 24시까지 미세먼지 농도 순위를 나타낸 결과, 6시 ~ 24시에 소사벌레포츠타운삼거리 ~ 군문교삼거리 구간(빨간색)의 미세먼지가 가장 높은 것으로 나타남
- SK뷰아파트앞삼거리 ~ 소사벌레포츠타운삼거리(노란색), 성동초등학교사거리 ~ 통북시장오거리(초록색)구간의 경우도 6시부터 24시까지 상위의 미세먼지 발생량을 나타냄

■ 일반국도 미세먼지 순위

[illegible]

- 일반국도의 1시부터 24시까지의 미세먼지 농도 순위를 나타낸 결과 서안성IC입구 ~ 송탄 교차로 구간이 대부분 높은 미세먼지 발생량을 보임
- 고덕교차로2 ~ 신궁입체구간의 경우 9시 이후 미세먼지 발생량이 증가함

노선 설계

■ 노선설계



운행노선 설계근거



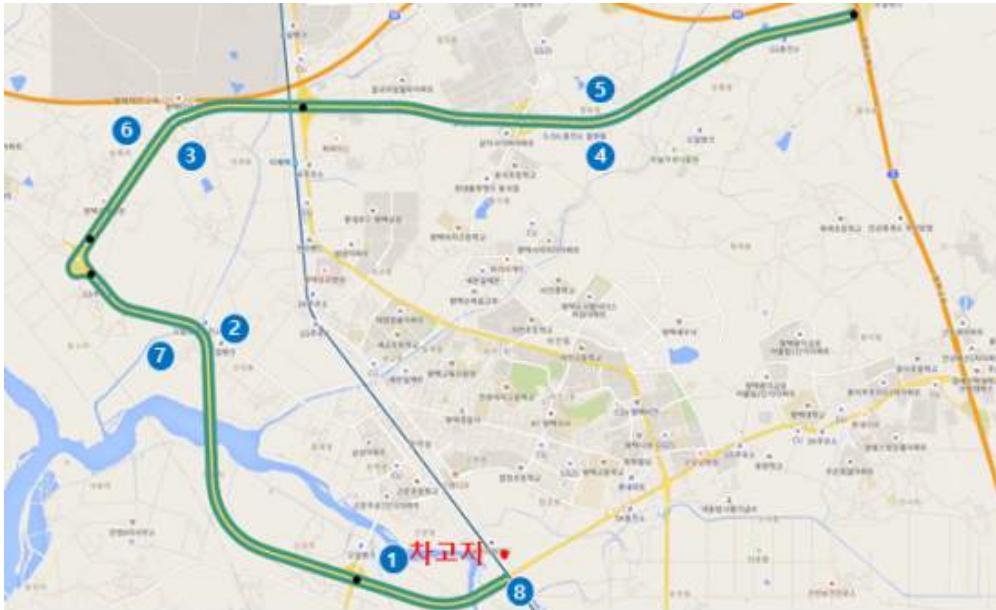
노선	운행 시간	운행 길이	소요 시간	소요 시간
운행노선 A	7시 ~ 19시	약 45km	약 3시간	[7시 ~ 19시] 전 시간대 미세먼지 발생량이 높은 외곽 구간의 노선 설계
운행노선 B	7시 ~ 9시	약 11km	약 1시간 50분	[7시 ~ 9시] 오전 주요시간대 미세먼지 발생량이 높은 시내 내부 구간의 노선 설계
운행노선 C	17시 ~ 19시	약 16km	약 2시간	[17시 ~ 19시] 오후 주요시간대 미세먼지 발생량이 높은 시내 외부 구간의 노선 설계
운행노선 D	10시 ~ 12시	약 49km	약 3시간 20분	[10시 ~ 12시] 해당 시간대 미세먼지 발생량이 높은 시내와 동쪽 외곽 구간의 노선 설계
운행노선 E	13시 ~ 16시	약 29km	약 3시간	[13시 ~ 16시] 해당 시간대 미세먼지 발생량이 높은 시내와 서쪽 외곽 구간의 노선 설계

〈운행노선 설계근거〉

- 각 시간별 구간별 미세먼지 발생량 순위와 소통판정 현황 순위를 통해 우선 진출구역 순위를 선정하였음
- 선정된 구역들 중 운행에 효율성과 타당성을 고려하여 노선 설계를 실시함
- A,B,C,D,E 5가지 노선이 설계되었으며 운행 시간과 노선 구역에 따라 나누었음



■ 운행노선-A



〈운행노선-A〉

• 경로

순서	출발	도착
1	차고지	신궁입체3
2	신궁입체3	고덕교차로2
3	고덕교차로4	송탄교차로
4	송탄교차로	서안성IC입구
5	서안성IC입구	송탄교차로
6	송탄교차로	고덕교차로4
7	고덕교차로4	고덕교차로동측
8	고덕교차로2	신궁입체3
9	신궁입체3	차고지

- 운행노선-A의 경우 평택시청 외곽국도를 중점으로 노선설계를 하였음
- 전체 시간대에서 모두 미세먼지 발생량이 높으므로, 주기적인 청소차량 운행 및 차량 운행 외 다른 미세먼지 감소를 위한 대책 수립이 필요함

■ 운행노선-B



〈운행노선-B〉

· 경로

순서	출발	도착
1	차고지	신궁입체3
2	신궁입체2	군문교삼거리
3	군문교삼거리	삼성아파트앞삼거리
4	삼성아파트앞삼거리	신대교차로
5	신대교차로	통복사거리
6	통복사거리	법원사거리
7	법원사거리	덕동사거리
8	덕동사거리	비전사거리
9	비전사거리	배미사거리
10	배미사거리	공설운동장사거리
11	공설운동장사거리	평택여중사거리
12	평택여중사거리	성동초등학교사거리
13	성동초등학교사거리	통복시장오거리
14	통복시장오거리	평택역오거리
15	평택역오거리	롯데인벤스사거리
16	롯데인벤스사거리	군문교삼거리
17	군문교삼거리	신궁입체2
18	신궁입체3	차고지

- 운행노선-B의 7시에서 9시에 운행하기 적합함
- 신궁입체로 시작하여 비전사거리에서 통복시장오거리 구간 등 시내 내부에서 발생하는 미세먼지 발생량을 기반으로 노선을 설계함



■ 운행노선-C



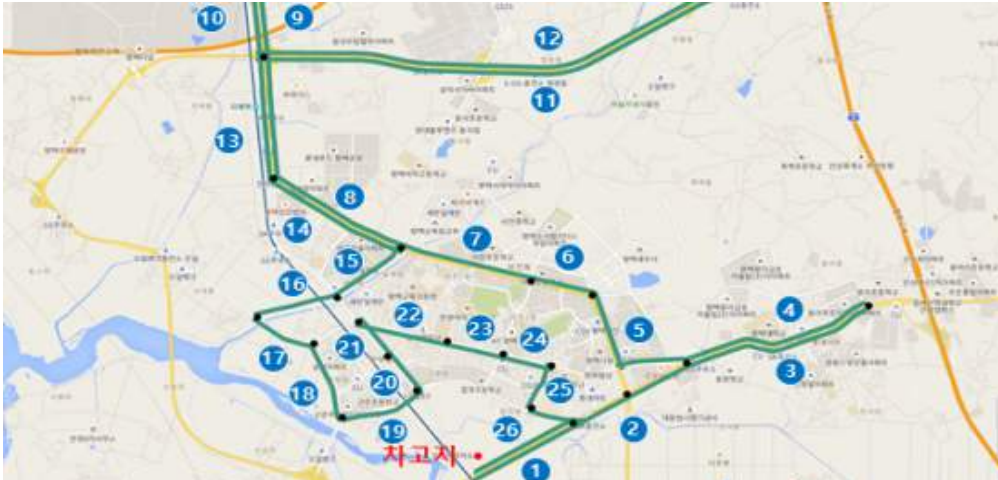
〈운행노선-C〉

• 경로

순서	출발	도착
1	차고지	평택합정동삼거리
2	평택합정동삼거리	SK뷰아파트앞삼거리
3	SK뷰아파트앞삼거리	소사벌레포츠타운삼거리
4	소사벌레포츠타운삼거리	군문교삼거리
5	군문교삼거리	삼성아파트앞삼거리
6	삼성아파트앞삼거리	신대교차로
7	신대교차로	통복사거리
8	통복사거리	법원사거리
9	법원사거리	덕동사거리
10	덕동사거리	비전사거리
11	비전사거리	비전지하차도사거리
12	비전지하차도사거리	안성천교북측
13	안성천교북측	비전지하차도사거리
14	비전지하차도사거리	차고지

- 운행노선-C의 경우 17시에서 19시에 운행하기 적합함
- 평안지하차도로 시작하여 소사벌레포츠타운 삼거리에서 안성천교북측 구간 등 시내 외부에서 발생하는 미세먼지 발생량을 기반으로 노선을 설계함

■ 운행노선-D



〈운행노선-D〉

• 경로

순서	출발	도착
1	차고지	평택합정동삼거리
2	평택합정동삼거리	SK뷰아파트앞삼거리
3	SK뷰아파트앞삼거리	평안지하차도1
4	평안지하차도1	SK뷰아파트앞삼거리
5	SK뷰아파트앞삼거리	비전사거리
6	비전사거리	덕동사거리
7	덕동사거리	법원사거리
8	법원사거리	한신주유소삼거리
9	한신주유소삼거리	송탄공단삼거리
10	송탄공단삼거리	송탄교차로
11	송탄교차로	서안성IC입구
12	서안성IC입구	송탄교차로
13	송탄교차로	한신주유소삼거리
14	한신주유소삼거리	법원사거리
15	법원사거리	통복사거리
16	통복사거리	신대교차로
17	신대교차로	삼성아파트앞삼거리
18	삼성아파트앞삼거리	군문교삼거리
19	군문교삼거리	롯데인벤스사거리
20	롯데인벤스사거리	평택역오거리
21	평택역오거리	통복시장오거리
22	통복시장오거리	성동초등학교사거리
23	성동초등학교사거리	평택여중사거리
24	평택여중사거리	공설운동장사거리
25	공설운동장사거리	소사벌레포츠타운삼거리
26	소사벌레포츠타운삼거리	차고지

- 운행노선-D의 경우 10시에서 12시에 운행하기 적합함
- 시내와 평택시 동쪽 외곽도로에서 발생하는 미세먼지 발생량을 기반으로 노선을 설계함



■ 운행노선-E



〈운행노선-E〉

• 경로

순서	출발	도착
1	차고지	신궁입체3
2	신궁입체3	고덕교차로2
3	고덕교차로동측	고덕교차로4
4	고덕교차로4	송탄교차로
5	송탄교차로	송탄공단삼거리
6	송탄공단삼거리	한신주유소삼거리
7	한신주유소삼거리	법원사거리
8	법원사거리	덕동사거리
9	덕동사거리	비전사거리
10	비전사거리	비전지하차도사거리
11	비전지하차도사거리	SK뷰아파트앞삼거리
12	SK뷰아파트앞삼거리	소사벌레포츠타운삼거리
13	소사벌레포츠타운삼거리	군문교삼거리
14	군문교삼거리	삼성아파트앞삼거리
15	삼성아파트앞삼거리	신대교차로
16	신대교차로	통복사거리
17	통복사거리	통복시장오거리
18	통복시장오거리	성동초등학교사거리
19	성동초등학교사거리	평택여중사거리
20	평택여중사거리	공설운동장사거리
21	공설운동장사거리	소사벌레포츠타운삼거리
22	소사벌레포츠타운삼거리	차고지

- 운행노선-E의 경우 13시에서 16시에 운행하기 적합함
- 시내와 평택시 서쪽 외곽도로에서 발생하는 미세먼지 발생량을 기반으로 노선을 설계함



데이터 분석 단계

1. 데이터 분석 단계 요약

단계	주요 내용
분석 설계	<ul style="list-style-type: none"> ■ 분석 범위 <ul style="list-style-type: none"> • 공간적 범위 : 평택시청 관할 도로 • 시간적 범위 : 2019.06. ~ 08. • 분석 단위 : 도로 각 구간 ■ 분석 모형 <ul style="list-style-type: none"> • 미세먼지 발생량 예측 • 효율적 운행 노선 설계
데이터 수집 및 가공	<ul style="list-style-type: none"> ■ 평택시 담당자와 인터뷰 후 데이터 수집 <ul style="list-style-type: none"> • 교통량 데이터, 평균 속도 데이터, 살수차량 운행노선, 임대살수차 운행일지 등 ■ 본 분석에 맞는 데이터 가공 <ul style="list-style-type: none"> • 데이터의 분석 기간 통일 • 데이터 시간단위 통일 • 필요 컬럼 추출 • 교통량 데이터, 평균 속도 테이블 조인
데이터 분석 및 검증	<ul style="list-style-type: none"> ■ 도로 각 구간의 미세먼지 발생량 예측 <ul style="list-style-type: none"> • 평택시청 관할 도로 분석 진행 • 3개월간 평균 미세먼지 발생량 추세 분석 • 3개월간 요일별 미세먼지 발생량 패턴 분석 • 평일 시간, 구간별 미세먼지 발생량 예측 • 도로별 미세먼지 발생량 순위 선정 ■ 노면 청소차량 운행 노선 설계 <ul style="list-style-type: none"> • 주요 시간 미세먼지 발생량 분석 • 도로 각 구간의 미세먼지 순위 선정 • 도로 각 구간의 소통판정 순위 선정 • 우선 진출구역 순위 선정 및 운행 효율성 확인
데이터 시각화	<ul style="list-style-type: none"> ■ 미세먼지 발생량 시각화 ■ 노선 설계 시각화



2. 분석 설계

■ 도로 각 구간의 미세먼지 발생량 예측

- 2019년 6월부터 8월까지의 미세먼지 발생량 예측

■ 도로 각 구간의 미세먼지 발생량 예측

- 도로 미세먼지 발생량과 도로 소통판정 데이터를 활용하여 효율적인 운행 노선 설계

과제	<ul style="list-style-type: none"> ■ 분석 범위 선정 ■ 분석 모형 설계
추진내용	<ul style="list-style-type: none"> ■ 분석 범위 <ul style="list-style-type: none"> • 2019.06. ~ 08. 평택시청 관할 도로 각 구간 ■ 분석 모형 <ul style="list-style-type: none"> • 미세먼지 발생량 예측 <ul style="list-style-type: none"> - 3개월간 미세먼지 발생량 분석 및 예측 - 도로 각 구간의 미세먼지 발생량 분석 및 예측 - 주요 시간 미세먼지 발생량 분석 - 도로별 미세먼지 발생량 순위 선정 • 노면 청소차량 운행노선 설계 <ul style="list-style-type: none"> - 도로 미세먼지 발생량 현황 확인 - 도로 소통 판정현황 확인 - 우선 진출구역 선정 및 운행 효율성 확인
Tip	<ul style="list-style-type: none"> ■ 분석설계 과정에서 가장 중요한 것은 유관기관 및 해당 모델 활용자가 원하는 방향 및 형태로 분석 결과를 보여주는 것임 ■ 모델 활용자가 필요로 하는 결과물의 형태를 파악하고, 분석 모델에 적용해야만 실제 해당 모델의 활용도를 높일 수 있음

3. 데이터 수집 및 가공

과제	<ul style="list-style-type: none"> ■ 분석에 필요한 데이터를 수집 및 가공
추진내용	<ul style="list-style-type: none"> ■ 데이터 가공 <ul style="list-style-type: none"> • 교통량 데이터 - 일별로 나뉜 데이터 통합 <ul style="list-style-type: none"> - 분 단위 형태의 데이터 시 단위 형태의 데이터로 변경 - 평택시청 관할 도로 각 구간 데이터 추출 - 분석 시 필요한 컬럼 추출 - 교통속도 데이터와 테이블 조인 • 교통 속도 데이터 <ul style="list-style-type: none"> - 2019년 6월 ~ 9월 데이터 통합 및 추출 - 평택시청 관할 도로 각 구간 데이터 추출 - 분석 시 필요한 컬럼 추출 - 교통량 데이터와 테이블 조인
Tip	<ul style="list-style-type: none"> ■ 보유 데이터에 맞춰 분석을 수행하는 것이 아니라, 분석모형을 설정하고 빅데이터 분석모형에 필요한 데이터를 정의해야함

4. 데이터 분석 및 검증

54



〈노선 설계 프로세스〉

과제	<ul style="list-style-type: none"> ■ 분석 분야에 맞는 세부 분석모델을 설정하고 분석 실시 ■ 사업 담당자와의 지속적인 논의를 통하여 분석모델 방향 점검 및 결과물 검토 실시
추진내용	<ul style="list-style-type: none"> ■ 도로 각 구간의 미세먼지 발생량 예측 <ul style="list-style-type: none"> · 미 환경부(U.S. EPA)에서 개발된 분석 도구인 MOVES(Motor Vehicle Emission)의 lookup table 방법을 사용하여 미세먼지 발생량 예측 · 미세먼지 발생량 예측 공식 $SE = EF_{ij} \times \bar{q} \times l$ <p>SE : 구간차량 미세먼지 발생량(g/section-h) EF : 배출계수(g/veh-km) i : 속도 j : 오염물질 \bar{q} : 구간별 평균 교통량(veh/h) l : 구간길이(km/section)</p> ■ 도로 각 구간의 미세먼지 발생량 분석 <ul style="list-style-type: none"> • 3개월간 평균 미세먼지 발생량 추세 분석 • 3개월간 요일별 미세먼지 발생량 패턴 분석 • 평일 시간대별 도로 각 구간의 미세먼지 발생량 분석 • 평일 시간대별 도로 각 구간간의 미세먼지 발생량 비교 분석 • 미세먼지 발생량이 높은 주요 시간대 분석 ■ 노면 청소차량 운행 노선 설계 <ul style="list-style-type: none"> • 미세먼지 발생량 현황 및 도로 소통판정 현황의 데이터 비교 분석을 통한 우선 진출 구역 순위 선정 • 우선 진출 구역의 운행 효율성 확인 • 주요 시간대의 운행노선 설계

5. 데이터 시각화

■ 분석결과 시각화 및 유관기관, 관련부서 담당자에게 제공

과제	<ul style="list-style-type: none"> ■도로 각 구간의 미세먼지 발생량 시각화 ■노면 청소차량 운행 노선 시각화
추진내용	<ul style="list-style-type: none"> ■시각화 대상 <ul style="list-style-type: none"> •평택시 지도 <ul style="list-style-type: none"> - 평택시청 관할 도로 지도 •활용 데이터 <ul style="list-style-type: none"> - 전국 표준 노드 · 링크 데이터 - 평택시청 관할 지역 도로 및 구간 - 교통량 - 평균 속도 - 도로 길이 •분석 결과 <ul style="list-style-type: none"> - 도로 각 구간의 미세먼지 발생량 - 노면 청소차량 운행노선 ■표현방법 <ul style="list-style-type: none"> •도로 각 구간의 미세먼지 발생량 <ul style="list-style-type: none"> - QGIS로 전국 표준 노드 · 링크 데이터를 평택시청 관할 지역의 도로로 추출한 뒤 미세먼지 발생량을 레이어로 표현 •노면 청소차량 운행 노선 <ul style="list-style-type: none"> - QGIS로 우선 진출구역의 노선을 레이어로 표현 ■분석 결과 저장 <ul style="list-style-type: none"> •분석 결과물을 로컬 내 지정 폴더에 저장
Tip	<ul style="list-style-type: none"> ■빅데이터 분석 결과를 시각화하는 과정에서, 필요한 시각화 자료인지, 민감한 자료는 아닌지 등에 대해 내부 협의가 필요함 ■유관기관 및 담당부서와의 사전 협의를 통해 시각화 결과물의 형태결정 필요





행정적용 단계

1. 행정적용 단계

- 본 분석은 시범으로 진행된 분석 과제로 주요시간 구간별 미세먼지 발생 변화량을 운행 노선 설계의 기본 자료로 활용하였으며, 노면 청소차량의 효율적인 운행 노선을 설계하고, 설계된 노면 청소차량의 운행 노선은 평택시청 관할 도로의 미세먼지 감소에 활용
- 추후 본 분석 시 아래의 개선사항을 추가하여 정확도를 향상 시켜 효과적인 분석으로 전 지자체에서 활용할 수 있는 분석 모델 개발
- 계측장비 추가 설치에 따른 도로 구간의 세분화를 통한 정교한 데이터 필요
 - 계측장비 추가 설치 : ATMS를 비롯한 다양한 계측 IoT장비 추가 설치를 통한 구간 세분화로 구간별 정교한 교통량과 평균속도 측정
- 구간 미세먼지 발생 특징에 따른 현실적인 방안 모색 필요
 - 구간 특성에 따른 미세먼지 해결 방안 마련 (Ex. 클린로드 설치, 공회전단속 등)
- 차종(버스, 자가용, 오토바이 등)에 따른 미세먼지 발생 예측 모델 고도화 필요
 - 차량종류에 따른 미세먼지 배출량 예측 : 버스, 자가용, 오토바이 등 차량 특성 분류를 통하여 미세먼지 배출 예측 신뢰도 향상
- 지자체 보유 관용차량 특성에 맞는 추가 노선 설계
 - 미세먼지 관련 관용차량 특성별 노선세분화 : 노면청소차량 뿐 아니라, 다른 미세먼지 관련 차량에 대한 추가적 노선 설계
- 노선설계 활용 데이터 확장
 - 지역적 특성(농어촌, 공업지대 등), 유동인구데이터, 날씨 등 추가적인 데이터 활용으로 노선 고도화

2. 행정 정책 개선

- 관내 노면 청소차량의 운영을 효율적으로 관리할 수 있음
- 노면 청소차량의 효율적 운영을 통해 효과적으로 도로 미세먼지 발생량 저감



기상조건에 따른 시내버스 정류장 정시성 분석

| 부산광역시 / 조기행 |

요약

본 사업은 기상청 공모사업의 일환으로 부산광역시가 제안한 기상과 대중교통(정시성) 분석을 통한 시민생활 개선 정책 추진사례로서, 근거 기반 정책의 근간인 데이터 분석과 행정 활용 과정을 요약하였다.

사업의 추진은 주관기관인 기상청(기상융합서비스)의 전폭적인 이해와 지원을 기초로, 전문기관들의 헌신적인 도움으로 성공할 수 있었다.(철도기술연구원의 하차지 추정, 교통안전공단의 법인택시 데이터 수급)

기존 학술연구와의 차별점은 시민의 실제 대중교통 이용 환경을 개선하려는데 주안점을 두었다는 것이며, 기존의 노선 단위 정시성 분석에서 한 걸음 나아가, 정류장 중심의 정시성에 도전한 것이다. 또한, 정시성 우선 대상 정류장 선정에 있어 탑승객수를 가중치로 사용하여 시민의 불편 정도를 고려하였다.

부산지역 날씨의 특성상 강우일(5mm이상)이 제한적이라는 점과 분석대상을 이용객 상위 40개 노선과 하위 9개 노선 등으로 제한한 점 등이 아쉽지만, 나머지 노선과 환승루트에 적용할 수 있는 분석프레임을 정립한 점은 강조할 만하다.

분석결과는 대시보드 형태로 정리되어 담당부서와 합의된 내용에 대해 공개될 예정이며, 프로젝트를 위해 수집/처리된 Data는 관내 대학(부산대)의 수업/연구에 교재로 쓰도록 제공되었다.





추진배경

1. 부산시 빅데이터 전략

■ 관련 조례

- “시민의 삶의 질을 향상시키고 지역경제를 활성화하는 데 이바지”

빅데이터 정의

정보처리능력을 가진 장치를 통하여 공개정보 및 이용내역 정보 등을 처리(수집, 생성, 저장, 조합, 분석, 그밖에 이와 유사한 행위를 말한다)하여 생성되는 정보 및 이로부터 가치를 추출하고 결과를 분석하는 기술을 말한다

- 부산광역시 빅데이터 활용 및 빅데이터산업 육성에 관한 조례(제5447호, 개정5982호)

- “통계 및 데이터를 활용하여 객관적 과학적 행정 수행(정책수립 및 시행)”

데이터 기반행정의 정의

공사·공단 및 출자·출연기관이 생성하거나 다른 기관 및 법인·단체 등으로부터 취득하여 관리하고 있는 통계 및 데이터를 수집·저장·가공·분석·표현하는 등의 방법으로 정책 수립 및 시행에 활용함으로써 데이터에 근거하여 객관적이고 과학적으로 수행하는 행정을 말한다.

- 부산광역시 데이터기반행정 활성화에 관한 조례(제5981호)

59



■ 부산시 빅데이터의 행정활용 전략

- “부산 빅데이터” 추진 전략

정책대안 (실용적 분석)	협업/공감 (협업+교육+생태계)	혁신/지속 (지속성장/변화)
<ul style="list-style-type: none"> • 시민의 편익증대 • 업무효율 개선 • 행정활용 중심 	<ul style="list-style-type: none"> • 빅데이터협의체(구군/기관/부울경/학교/단체) • 민간활용 촉진 	<ul style="list-style-type: none"> • 데이터행정책임관/평가 • 전문인력양성 • 조직/전문가/제도/인식
전략정보 정보융합 민간활용(개방/제공) 통계활용	Networking(중앙+공공) 정보공유(연구원+대학)	Measurement Training

I 부산시 중점 데이터

기본 데이터	중점 데이터
<div>인구</div> <div>교통</div> <div>건강</div> <div>고용</div> <div>여신</div> <div>주택</div>	<div>물류</div> <div>관광</div> <div>안전</div> <div>해양</div> <div>산업</div>
인구 전출입/피라미드/고령/1인가구	물류 항만/수출입/물류
교통 승하차/환승/혼잡	관광 유동인구/교통량/소비/문화
건강 건강지표/만성질환/의료기관	안전 지진/해일/화학/원자력
고용 고용/실업/임금/근속/출퇴근	해양 어업/수산업/환경
여신 부채/소비/소득(가계)/연체	산업 제조/서비스/문화/수송
주택 거래/유형	

2. 기상과 대중교통(정시성)의 관계

■ 날씨가 도시생활에 미치는 영향

- 여가, 교통, 건강 등 도시생활의 활동성에 폭넓게 영향을 미침
- 특히, 도시생활 중 대중교통 이용의 편리성에 미치는 영향이 큼

■ 기상으로 인한 대중교통 불편사항

- (가설1) 비가 오는 날 대중교통 이용에 불편이 많을 것이다
- (가설2) 비가 오는 날 대중교통 정시성이 낮아지고, 승차거부가 많을 것이다

■ 시민중심 대중교통 정책 전환

- 기존의 대중교통 정책은 공급자 중심
- 자원의 효율적인 활용과 총량기준 서비스 점검 체계 → 시민의 불편정도에 가중치 부여

■ 시민중심 대중교통 이용현황 분석이 미약

- 총괄적인 대중교통 이용현황 미흡
- 구조적인 원인, 큰 맥락에서의 불편, 기상으로 인한 문제증폭 등을 분리 필요





프로젝트 개요

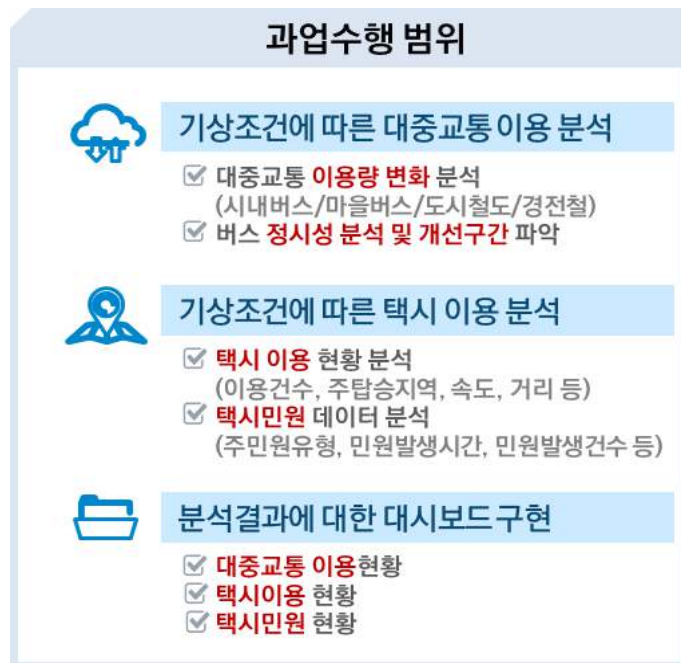
■ 개요

- 기상조건(강우)으로 인한 대중교통 이용의 불편 발생(기다림)
- 정류장 단위의 정시성 개선을 목표로 우선관리 대상을 제시

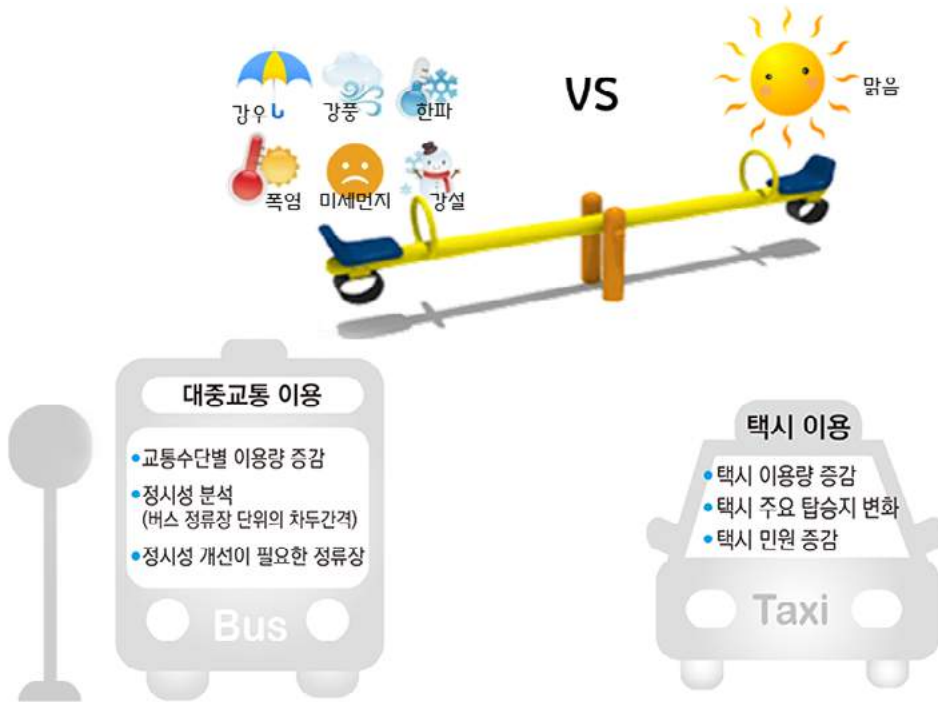
■ 분석의 필요성

- 불편의 구체적인 정의, 개선요구 구간과 시간대 실증 제시
- 시민의 불편을 고려하여 우선 개선 대상 정의
- 기존 학술 논문의 정시성 분석방법 개선(실용성 강화)

■ 분석 범위



■ 분석의 필요성



62



■ 분석의 주안점

- 현실적이고 실행 가능한 결과 도출
- 현장과 담당부서의 검증절차 확보
- 지속적으로 발전 및 확산 가능한 분석설계와 논리구조 전개
- 편리한 BI시스템으로 현장 및 담당부서가 분석결과를 쉽게 활용
- 결과의 공개(시민포함)
- 데이터의 공유(관내 대학)

■ 활용데이터

• 분석에 활용한 데이터(기상)

구분	기준	분석기간		비고	데이터출처
강우	• 5mm이상 강우일 (기상측정소 8개소 이상 에서강수가측정된날)	16일 (강수 8일/맑음 8일)	2018년 6월 ~2019년 6월	• 강우량이 5mm이하이고 기상측정소 8개소 미만 에서 측정되었을지라도 지속적으로 강수가 발생 한 경우 포함	기상자료개발포털 (기상청)
강풍	• 초속 6.5이상 강풍 분 날	8일 (강풍 4일/맑음 4일)	2018년 4월~5월	• 일별 평균 강풍 최고일 순 선정(분석기간내 강 풍주의보 없음) • 강수일 제외	
한파	• 체감온도 기준 영하 4.5도 이하인 날	8일 (한파 4일/맑음 4일)	2018년 12월 ~2019년 2월	• 체감온도 : 일별 평균기 온 x 평균 강풍	
폭염	• 체감온도기준 28도 이 상인 날	14일 (폭염 6일/맑음 8일)	2018년 7월 ~8월		
미세 먼지	• 낮시간에 미세먼지주의 보 및 경보가 발생한 날	6일 (미세먼지 3일/맑음 3일)	2018년 11월 ~2019년 4월		
강설	• 부산시 눈이 온 날 선정	2일 (폭설 1일/맑음 1일)	2019년 1월		

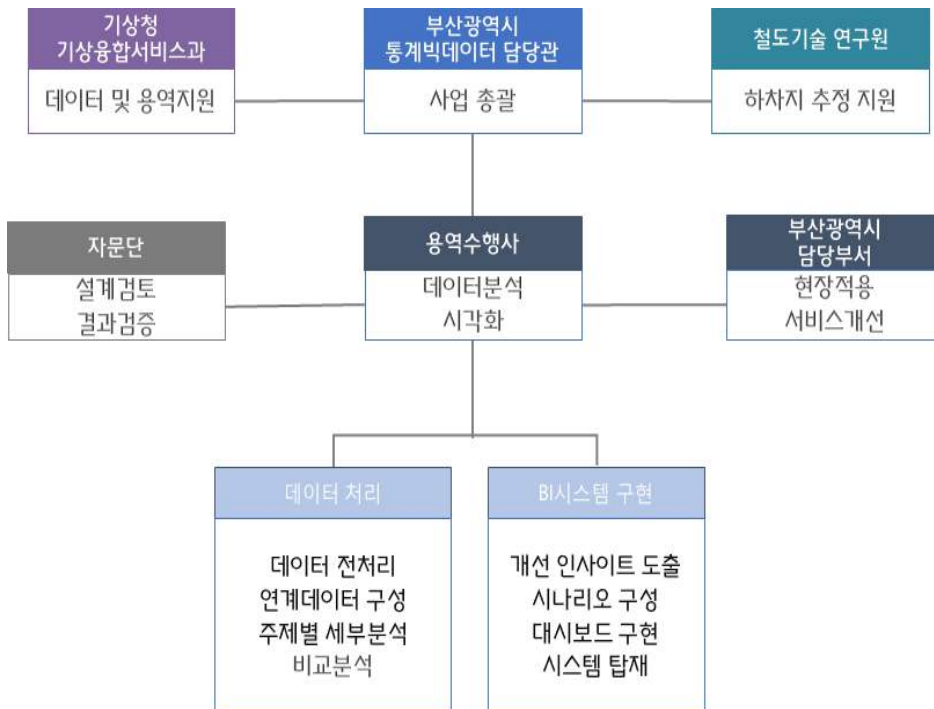
• 분석에 활용한 데이터(대중교통)

대분류	구분	데이터건수	데이터 용량 (KB)	데이터 출처	데이터 분석기간
운행정보	동해선	2,651,011	166,929	한국철도공사	2018년 1월~2019년 6월 (52일 치 활용)
	일별거래내역	168,115,091	28,433,807	부산광역시	2018년 1월~2019년 6월 (52일 치 활용)
등대콜택시	택시-등대콜	417,018	72,040	등대콜	2018년 1월~2019년 6월 (기상요건 기준 중 강설 제외 50일)
법인택시	택시-법인	5,092,148	459,862	한국교통안전공단	2018년 11월~2019년 6월 (강우일 10일치 활용)
택시민원	택시-민원	10,022	1,453	부산광역시	2014년~2019년
	택시-업체정보	481	301	부산광역시	2014년~2018년
정시성	정시성-시내버스	709,489	1,633,842	부산광역시	2018년 1월~2019년 6월
	정시성-마을버스	1,441,303	1,413,535	부산광역시	2018년 1월~2019년 6월
	정시성-시내,마을 위치 거리정보	147,870	12,472	부산광역시	2018~2019년

• 기존 논문 참고

- 이창, 고준호, 강영은, 이태경, 2011, “강우에 의한 서울시 대중교통 서비스 변화 분석 - 버스서비스의 정시성과 통행속도 변화를 중심으로”, 「국토계획 46(7)」, 대한국토도시계획학회
- 박근영, 이시복, 2012, “강우상태에 따른 대중교통 이용패턴 특성연구 - 부산광역시 버스통행을 중심으로”, 「대한토목학회논문집 D 32(1D)」, 대한토목학회
- 최상기, 이종호, 오승훈, 2013, “기상조건이 대중교통수요에 미치는 영향에 관한 연구”, 「대한토목학회논문집 33(6)」, 대한토목학회
- 이미영, 2015, “대중교통카드자료를 활용한 수도권 통행분석”, 「국토정책 Brief」, 국토연구원

■ 활용데이터

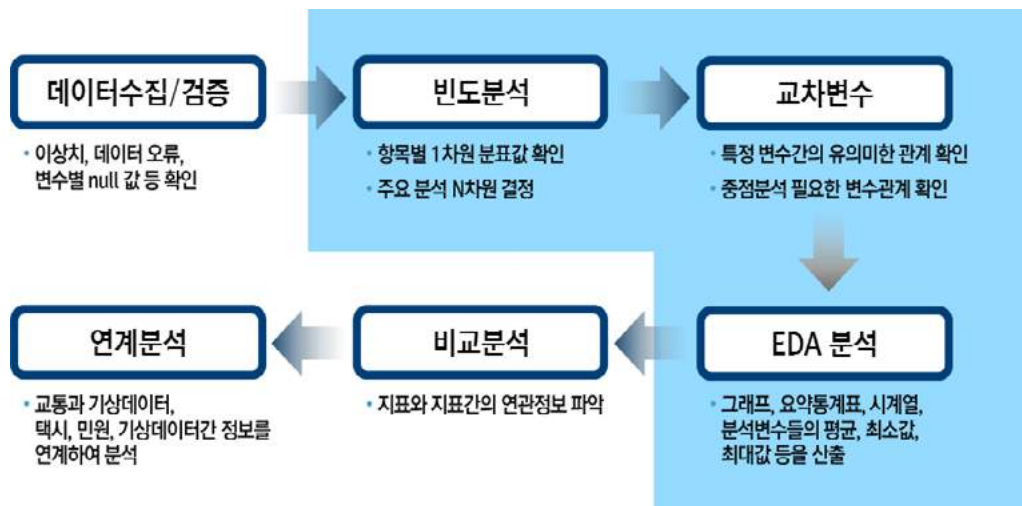


다 적용 프로세스

■ 분석 전체 프로세스



■ 데이터 분석 절차





기획·관리 단계

1. 기획·관리 단계 요약

단계	주요 내용
문제의식	<ul style="list-style-type: none"> ■ 기상 조건이 시민의 대중교통에 미치는 영향이 클 것이라고 막연히 추정해오고 있었음 ■ 비 오는 날 도심에서는 버스나 택시를 이용하기 불편하다는 민원이 있었지만, 어느 위치, 어느 구간, 어느 시간대, 어느 노선에서 민원이 많이 발생하는지에 대한 정확한 분석이 부족하였음
사업계획	<ul style="list-style-type: none"> ■ 사업의 요건 정의 <ul style="list-style-type: none"> • 현업 담당부서와 협의하고 수행팀과의 조율을 거쳐 프로젝트의 최종 결과물을 정의하고 최종적인 기대효과들을 정리(행정활용 포함) ■ 프로젝트 관리 <ul style="list-style-type: none"> • 전체적인 범위와 일정을 실천해 나가기 위해 가장 중요한 항목이 사업관리(Project Management)임 • 책임과 역할(R&R)의 정의에서부터 산출물 검수에 이르는 전반을 아우름
완료보고	<ul style="list-style-type: none"> ■ 분석의 시나리오 구성 ■ 대중교통 이용 주요현황 정리(Key Numbers) ■ 기상조건에 따른 대중교통 이용량 변화 ■ 우천시 정시성 개선요구 구간 정의 ■ BI Tool 구현(시각화)



2. 문제의식

■ 기상조건에 대응하는 대중교통 정책 필요

- 기상 조건이 시민의 대중교통에 미치는 영향이 클 것이라고 막연히 추정해오고 있었음
- 비 오는 날 도심에서는 버스나 택시를 이용하기 불편하다는 민원이 있었지만, 어느 위치, 어느 구간, 어느 시간대, 어느 노선에서 불편이 초래되는지에 대한 정확한 분석이 부족했음
- 기상청 공모가 있어 부산시(통계빅데이터담당관+대중교통부서)가 응모하여 분석을 실시함

과제	<ul style="list-style-type: none"> ■ 기상악화에 따른 대중교통 불편사항 분석 ■ 노선과 시간대에 따른 개선과제 제시(정시성)
추진내용	<ul style="list-style-type: none"> ■ 기상악화에 따른 대중교통 불편사항 분석 <ul style="list-style-type: none"> • 비가 오거나 미세먼지, 강풍 등이 있는 날 시민들의 대중교통 이용에 불편이 많다는 공감 • 늦은 시간 인구밀집 지역(서면, 해운대, 동구, 중구)은 대중교통(택시) 이용에 불편을 겪고 있다는 관점 • 시의 담당부서는 이러한 인식에 대한 정확한 데이터 분석이 필요 • 근거에 기반한 구체적이고 현실적인 대안제시가 요구됨 ■ 시민중심의 대중교통 정책 필요(정시성) <ul style="list-style-type: none"> • 기존 논의 정시성 논리 이해 • 공급자 중심의 정책에서 시민 중심으로 전환
Tip	<ul style="list-style-type: none"> ■ 익숙해진 현상에 문제의식을 갖기가 쉽지 않지만 시민 생활에 끼치는 불편 사항을 중심으로 쉽게 관점을 전환할 수 있음

3. 사업계획

■ 사업의 요건 정의

- 현업 담당부서와 협의하고 수행팀과의 조율을 거쳐 프로젝트의 최종 결과물을 정의하고 최종적인 기대효과들을 정리(행정활용 포함)

과제	■프로젝트의 최종 산출물을 구체적으로 정의 ■산출물의 현장 활용 방안 제시		
추진내용	■산출물 정의		
	대중교통 이용현황	정시성 분석	현장활용 방안
	• 대중교통 승차인원 현황분석 • 노선별, 정류장별, OD별	• 정류장별, 기상조건별, 노선별, 시간대별	• 개선대상 정류장 List Up • 개선 우선순위 제시
	■현장 활용방안		
	정시성 개선	택시 이용개선	
	노선조정/교통지도/유연배차	행정지도/민원예고	
	■WBS작성 : 단위업무별로 상세 추진 일정을 합의하여 공유		
Tip	■시작하기 전이지만, 사업이 끝났을 때의 모습을 구체화하는 것이 중요		



■ 프로젝트 관리

- 전체적인 범위와 일정을 실천해 나가기 위해 가장 중요한 것은 사업관리(Project Management)임
- 책임과 역할(R&R)의 정의에서부터 산출물 검수에 이르는 전반을 아우름

과제	<ul style="list-style-type: none"> ■ 협업 기관/부서의 R&R(워크샵+정기협의) ■ 주단위 산출물 점검 ■ 자문위원 및 검증 방법 ■ 현업의 확인 및 System Delivery
추진내용	<ul style="list-style-type: none"> ■ 협업기관/부서의 R&R 정의 <ul style="list-style-type: none"> • 기상청의 공모사업에 선정됨 → 소요경비와 데이터를 지원받음 • 철도기술연구원의 하차지 추정 알고리즘을 지원 받음 (통행실적기반 대중교통운영 지원시스템(TRIPS)) • 한국교통안전공단의 법인택시 승하차 데이터 및 전처리를 지원받음 • 부산광역시 대중교통부서(버스운영과, 택시운영과)가 운영하는 시스템에서 승하차 데이터를 지원받음 ■ 주단위 산출물 점검 <ul style="list-style-type: none"> • 사업수행계획서 및 WBS(단위업무별 추진일정표) 협의 및 공유 • 주간보고서 산출물 및 추진현황 점검 • 중간보고 등을 통하여 목표일정의 달성도 점검 및 Issue해결 • 수집된 Data와 분석된 결과의 정합성 검증 ■ 현업확인 및 System Delivery <ul style="list-style-type: none"> • 분석결과의 현실성 확인(현실 대표) • 제시된 현장적용 대안의 실현 가능성 검토 • 분석결과의 BI System 구현 (분석시나리오 → 대시보드) • 다차원, 시계열, 시각화, 유관관점 등
Tip	<ul style="list-style-type: none"> ■ 계획된 일정 준수 ■ 사업수행계획서와 제안요청서의 산출물 요구사항을 엄격히 점검

4. 완료보고

- 프로젝트의 전 과정을 마치면서 주요 내용을 알기 쉽고 효율적으로 전달하는 것이 중요
- 완료보고의 구성을 의미하는 시나리오 수립과 개선효과 및 중요 구현내용 등이 중심이 됨

■ 분석의 시나리오 구성

- 다른 어떤 절차에 앞서 완료보고의 시나리오를 구성하는 것이 핵심
- 주요 내용을 효과적으로 전달하기 위한 결과의 재구성

과제	<ul style="list-style-type: none"> ■프로젝트의 개괄 ■분석결과의 요약 ■현장 활용 방안 ■시스템 구현 결과 ■제안사항
추진내용	<ul style="list-style-type: none"> ■프로젝트의 개괄 <ul style="list-style-type: none"> • 프로젝트의 목적, 기대하는 산출물 • 기간/수행업체/예산/데이터 수급 • 협업기관, R&R, 자문, 검증 • 추진일정 • 주요결과, 의미(특징) ■분석결과의 요약 <ul style="list-style-type: none"> • 대중교통 이용현황 • 정시성 분석결과(평시 vs 우천시) • 정시성 개선대상 정류장 제시 • 기대효과 ■현장 활용 <ul style="list-style-type: none"> • 시민중심 대중교통 정책 반영 • 우선 개선대상 제시 • 단계적 적용방안 ■시스템 구현 결과 <ul style="list-style-type: none"> • BI시스템 구성도 • 수록결과 요약/분류 • 시연 ■제안사항 <ul style="list-style-type: none"> • 분석결과의 의미/한계 • 향후 확대발전방안 • 보완계획
Tip	<ul style="list-style-type: none"> ■완료보고를 하기 전에 현장 사정에 맞춰 보고의 구성을 먼저 협의하는 것이 중요



■ 대중교통 이용 주요 현황 정리(Key Number)

- 분석된 대중교통 데이터 분석결과 중 대표적인 수치를 뽑아 전체를 대표하게 하는 절차로서 현황과 결과물 혹은 기대효과 등이 잘 드러나게 함

과제	<ul style="list-style-type: none"> ■ 전체적인 현황의 주요수치 요약 ■ 주안점(정시성)에 관한 기준선 제시
추진내용	<ul style="list-style-type: none"> ■ 전체적인 현황의 주요 수치 요약 <ul style="list-style-type: none"> • 대중교통 이용현황의 주요지표 파악 • 분석의 대상을 선택하는 논리 전개 • 개선이 필요한 영역의 지수현황 대조 ■ 주안점(정시성)에 관한 기준선 제시 <ul style="list-style-type: none"> • 지역, 노선, 정류장, 시간대 등에 대한 정시성 현황 제시 • 시민중심의 지표 필요성 제기 • 분석대상 선정(노선, 시간대, 기상조건, 일수, Data)
Tip	<ul style="list-style-type: none"> ■ 분석결과와 수치를 대조할 수 있는 Parameter들을 간결하게 제시 ■ 가급적 Graphic을 활용하여 즉시 파악 가능하도록 함

■ 기상 조건에 따른 대중교통 이용량 변화 제시

- 기상조건에 따른 대중교통 이용량의 변화와 민원 등의 변화를 체계적으로 비교할 수 있도록 구조적으로 제시

과제	<ul style="list-style-type: none"> ■ 대중교통 이용량 변화 ■ 민원의 변화
추진내용	<ul style="list-style-type: none"> ■ 대중교통 이용량 변화 <ul style="list-style-type: none"> • 기상조건의 정의 • 기상조건에 따른 대중교통 이용량 비교(맑은 날 대비) • 정시성의 변화 비교 • 구간/노선/시간대별 비교분석 • 정시성 지표의 변화 ■ 민원의 변화 <ul style="list-style-type: none"> • 택시이용의 현황(OD, 횟수, 금액) • 택시 민원의 분포 • 기상 조건에 따른 택시이용 변화 • 기상 조건에 따른 민원의 변화
Tip	<ul style="list-style-type: none"> ■ 비교를 위한 분석의 틀이 객관적인지 유의

■ 우천시 정시성 개선요구 구간 정의

- 프로젝트의 핵심요소중 하나인 정시성(Punctuality)에 대하여 정의를 내리고 그 결과를 구체적으로 업무개선에 적용하는 과정을 제시

과제	<ul style="list-style-type: none"> ■정시성의 관리기준 정의 ■우선 관리 대상 지정의 개선
추진내용	<ul style="list-style-type: none"> ■정시성의 관리기준 정의 <ul style="list-style-type: none"> •시물레이션 : 정시성 기준별 관리대상 정류장 • 기존 논문의 제시기준 검토 •현업의 검토와 자문위의 검증에 따른 기준 제시 ■우선 관리대상 지정의 개선 <ul style="list-style-type: none"> •이용 시민의 수를 가중치에 부여한 우선관리대상 정류장 제시 •노선별, 시간대별, 지역별 우선관리대상 정류장 List Up •지도 위에 표현(가시성 개선)
Tip	<ul style="list-style-type: none"> ■기존 논문의 관점과 본 프로젝트와 차이를 선명하게 제시 ■현장과 담당부서의 검토의견 ■실행 가능한 ‘정시성’ 정의



■ BI Tool 구현(시각화)

- 대개의 분석결과는 수치로 나타나며 때로는 단일차원이 아닌 다차원의 수치로 복잡하게 나타나기도 함
- 복잡한 결과를 논리적이고 구조적으로 재구성하여 분석결과에서 도출한 Insight를 선명하게 전달

과제	<ul style="list-style-type: none"> ■ 분석결과의 대시보드 구현 결과 ■ 대시보드의 주요 범위
추진내용	<ul style="list-style-type: none"> ■ 분석결과의 대시보드 구현 <ul style="list-style-type: none"> • 기본현황(기본통계량) • 정시성 분석(정류장 단위) • 민원 분석(택시) • 정시성 우선관리 대상 정류장 • Filter(Slide Bar) 구분 : 정시성 기준의 변경에 따른 우선대상 모의시험 • 연계 데이터와 Dimension : 주요한 변수들끼리 Filter나 Selection을 연계하여 조건분석에 활용 • 시연 : 현실의 상황에 빗대어 구현결과 대입(지도 위에 표현, Graph, List Up) ■ 대시보드의 주요범위 <ul style="list-style-type: none"> • BI System 구성도 : Diagram으로 대시보드의 구조/범위를 제시 • 수록결과 요약/분류 : Category, 배치된 Work Sheet, 분석논리/주제 등으로 요약/분류 • 대시보드구성의 시나리오(Hierarchy) : 분석주제(Topic)와 구성도(Diagram)를 유기적으로 엮어 맥락(Context)을 제시
Tip	<ul style="list-style-type: none"> ■ 대시보드의 주요 구조를 간결하게 제시 ■ 주요 결과를 슬라이드 환경에 맞도록 조정하여 표현

5. 분석결과

분석내용

■ 기본통계량 파악(Key Numbers)

- 분석의 결과를 비교할 기본 현황집계
- 교통수단별 이용건수 현황(승차/하차)

대중교통

(평일, 맑은 날 76일 평균, 단위:건)

• 교통수단별 이용건수(평일, 일평균)

이용건수	승차	하차
시내버스	656,633	656,272
마을버스	169,156	172,699
도시철도	984,607	837,381
경전철	50,948	48,638

• 분석대상 버스노선

구분	분석대상	비고
시내버스	49개 노선 (전체 144개)	일별 대당 이용객수 상위 40개, 하위 9개 노선 선정(2019.03 기준)
마을버스	44개 노선 (전체 152개)	월별 매출기준 상위 40개, 하위 4개 노선 선정
분석대상일	기상조건일: 26일 맑은 날: 28일	2018년 6월~2019년 6월

택시

(맑은 날 5일 평균, 단위:건, 원, km, km/h)

• 택시이용 현황(평일, 일평균)

이용횟수	이용건수	이용금액 (1회평균)	이동거리 (1회평균)	이용속도 (평균)
법인택시	153,489 (8,024)	6,759	5.1	26.2

() : 개인택시 이용횟수
개인택시 이용현황은 등대플 데이터 기준, 하차데이터 없음

• 분석대상 택시

구분	분석대상	비고
법인택시	13,883대	90개 회사(2018년 기준)
개인택시	10,843대	2018년 기준 (이용현황: 등대플 데이터 민원현황: 개인택시 전체)
분석대상일	기상조건일: 26일 맑은 날: 28일	2018년 6월~2019년 6월 법인택시: 강수일 16일만 적용

*하차추정: 한국철도연구원 하차 정규장 추정 알고리즘 적용, *동행실적기반 대중교통운영 지원시스템(TRIPS)
52일치(2018.11.1~2019.06.31)의 일별교통카드 이용 데이터 160,115,091건에 대해 하차 추정

■ 분석 대상 정의

- 분석에 사용되는 데이터의 구체적인 내용을 정의
- 버스: 노선수, 날짜수, 기간
- 택시: 택지구분, 차량수, 기간
- 시내버스의 이용건수 비교일: 26일(맑은 날, 비오는 날 각각)



분석내용

■ 기상조건별 대중교통 이용건수의 증감분석

- 기상조건별: 강우, 강풍, 폭염, 한파, 미세먼지, 강설
- 교통 수단별: 시내버스, 마을버스, 도시철도, 경전철
- 요일별 : 평일, 주말

평일					주말				
이용건수증감					이용건수 증감				
(단위: 맑은날 대비%, 승차기준)					(단위: 맑은날 대비%, 승차기준)				
구분	시내버스	마을버스	도시철도	경전철	구분	시내버스	마을버스	도시철도	경전철
강우	-16.2	-16.1	-13.2	-12.4	강우	-15.3	-13.7	-11.5	-12.9
강풍	-7.0	-2.8	-5.2	-3.3	강풍	9.4	8.0	8.6	7.1
폭염	-0.4	1.9	2.1	1.1	폭염	-12.9	-10.0	-8.9	-12.6
한파	-9.4	-11.0	-10.3	-8.1	한파	-5.5	-5.3	-29.0	-29.5
미세먼지	-0.6	-2.1	-5.4	-0.5	미세먼지	9.8	10.6	13.7	16.3
강설	-9.9	-12.2	-8.7	-5.9					

■ 기상 조건별 분석 결과

- 대중교통의 이용건수는 맑은 날 대비 12.4%~16.2%까지 감소
- 특히, 시내버스의 경우 비오는 날 이용건수는 맑은 날 대비 16.2%가 감소
- 주말의 경우에도 맑은 날 대비 비오는 날의 대중교통 이용건수가 감소 (11.5%~15.3%)
- 다만, 한파/미세먼지/강설의 수치는 사례수 자체가 적어서 신뢰도가 떨어짐

분석내용

■ 기상조건별 환승의 증감 분석

- 기상조건별: 맑은 날, 비오는 날
- 환승형태별: 버스↔버스, 버스↔도시철도, 도시철도↔도시철도

	1 버스 ↔ 버스	2 버스 ↔ 도시철도	3 도시철도 ↔ 도시철도
 맑은 날	83,146명	38,601명	43,234명
 강우일	68,573명	36,659명	41,192명
 증 감	-17.5%	-5.0%	-4.7%

■ 기상 조건별 환승의 증감 분석 결과

- 버스↔버스 환승의 감소폭이 가장 높았음(17.5% 감소)
- 버스↔도시철도, 도시철도↔도시철도 등은 버스만으로 이루어진 환승에 비해 감소폭이 낮음(-4.7% ~ -5.0%)
- 옥외에서 환승하는 경우, 강우가 영향을 미치는 것으로 추정



분석내용

■ 시간대별 정시성 증감 분석

- 기상조건별: 비오는 날, 맑은 날
- 시간대별: 출근시간, 낮시간, 퇴근시간

시내버스					마을버스				
분석대상 정류장 : 4,066개 (49개 노선) (평일 7시~9시)					분석대상 정류장 : 2,334개 (44개 노선) (평일 7시~9시)				
구분	맑음		기상조건		구분	맑음		기상조건	
	정류장수	비율	정류장수	비율		정류장수	비율	정류장수	비율
출근	579개	14.2%	840	20.7%	출근	269개	11.5%	280개	12.0%
낮	374개	9.2%	485	11.9%	낮	160개	6.9%	206개	8.8%
퇴근	138개	3.4%	342	8.4%	퇴근	236개	10.1%	290개	12.4%

※ 건수 : 정시성에서 벗어난 정류장 수

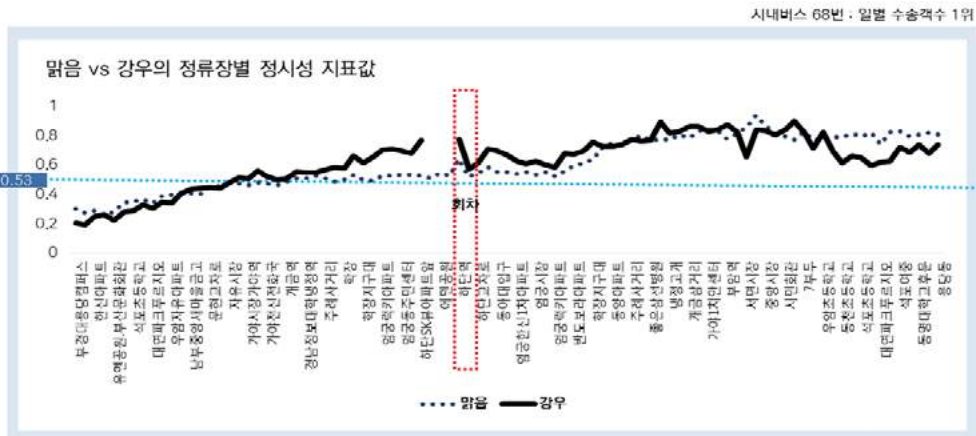
■ 기상 조건별 정시성 증감 분석 결과

- 비오는 날 정시성이 맑은 날 대비 8.4%이상 증가
- 시내버스 : 출근시간 20.7% 증가, 낮시간 11.9%증가, 퇴근시간 8.4%증가
- 마을버스의 정시성도 8.8% ~ 12.4% 증가

분석내용

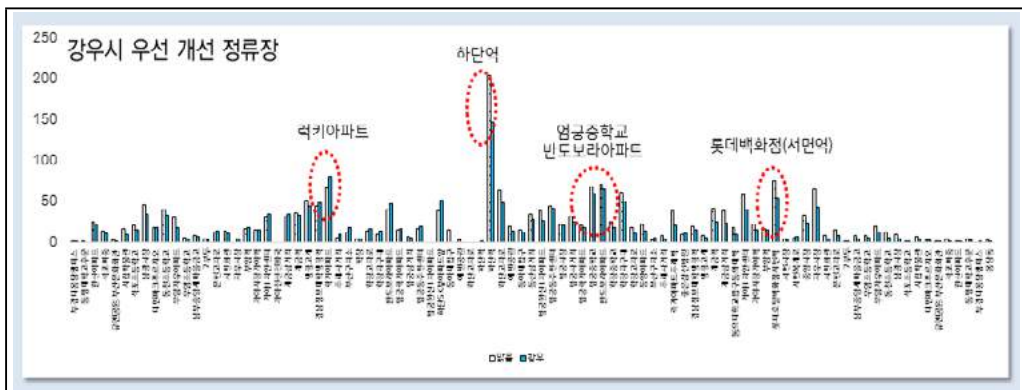
■ 기상조건별 정시성 증감 분석(예시 68번 버스, 정류장별)

- 기상조건: 맑은 날, 비오는 날
- 정시성 분석단위: 각 정류장



■ 기상조건별 정시성 증감 분석 결과(예시 68번 버스, 정류장별)

- 회차지점에 이를수록 맑은 날 대비 비가 오는 날의 정시성이 증가함
- 회차지점을 지나 돌아오는 코스도 증가하다가 종점에 가까워 오면서 정시성이 개선됨
- 대학가와 공단을 지나는 특성이 정시성에 반영되는 것으로 추정
- 각 정류장별 승차인원을 가중치로 고려했을 때 우선적으로 개선이 필요한 정류장을 선택함



분석내용

■ 기상 조건별 정시성 증감 분석결과와 시각화 (예시 68번 버스, 정류장별)

- 전체 노선 및 정류장을 지도위에 표현
- 해당 노선의 Key Number를 참조제시
- 우선관리 대상 정류장을 강조해서 표현



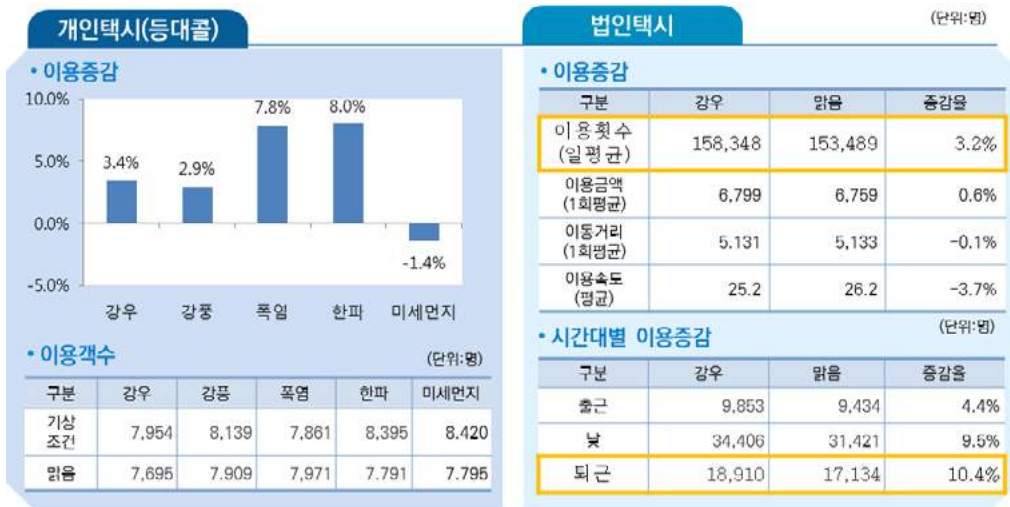
■ 기상조건별 정시성 증감 분석결과와 행정정책 적용 제안(예시 68번 버스, 정류장별)

- 정시성 개선 우선대상 정류장에 대한 행정정책을 제시
- 현업 담당자와의 협의를 거쳐 대안의 후보를 검토
- 현실성 있는 대안으로 제시

분석내용

■ 기상조건별 택시 이용횟수 분석

- 기상조건별: 강우, 강풍, 폭염, 한파, 미세먼지
- 택시 종류별: 법인택시, 개인택시
- 시간대별: 출근시간, 낮시간, 퇴근시간
- Measure(법인): 이용횟수, 이용금액, 이동거리, 이용속도



■ 기상조건별 택시 이용횟수 분석 결과

- 강우, 강풍, 폭염, 한파 등의 기상조건에 택시의 승차횟수 증가
- 특히, 폭염과 한파 등의 기상조건에서 이용횟수 증가폭이 큼(7.8%~8.0%)
- 세부분석은 법인택시 데이터를 중심으로 진행: 비오는 날 증가폭이 10.4% 큼(퇴근 시간대)



분석내용

■ 기상 조건별 택시 이용금액 분석

- 기상조건: 맑은 날, 비오는 날
- 요금대별: 500원단위(3,500원~14,000원이상)



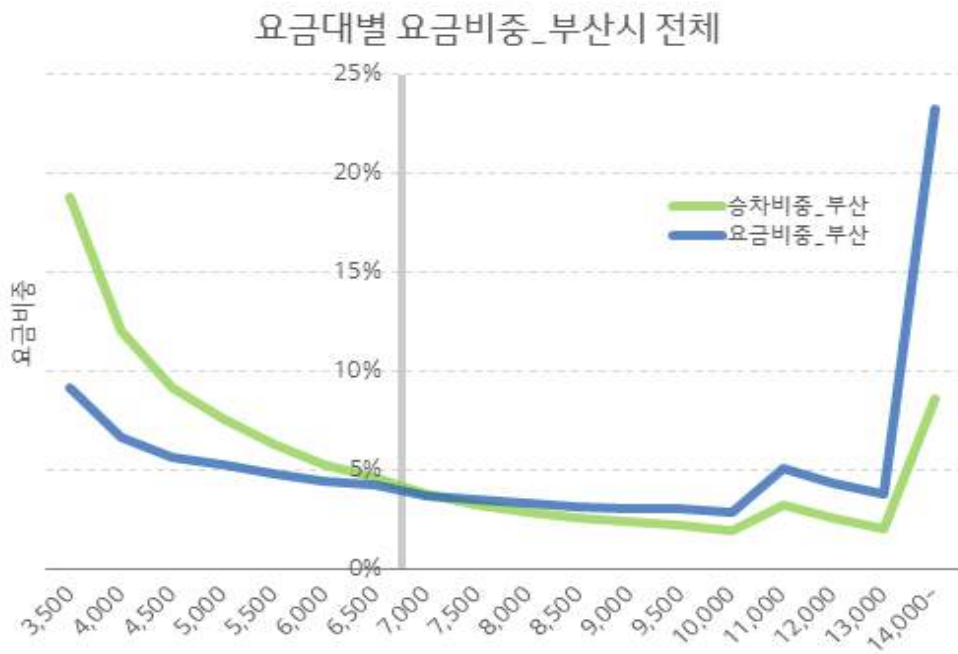
■ 기상 조건별 택시 이용금액 분석결과

- 평균 금액은 6,700원대
- 이용금액대별 분포에서는 3,500원대 승차인원이 가장 많은 비중을 차지
- ※ 택시 요금의 산술평균을 단순 적용하면 결과 해석의 오해가 우려
- 기상조건별 승차횟수의 차이는 크게 나타나지 않음 : 비오는 날이 맑은 날에 비해 일관적으로 높음

분석내용

■ 택시 요금 비중 분석

- 구분: 승차비중 vs 요금비중
- 요금대별: 500원단위(3,500원~14,000원이상)



■ 택시 요금 비중 분석 결과

- 14,000원 이상의 요금대역이 차지하는 비중이 가장 높음
- 그러나, 횡수로는 3,500원 요금대역의 비중이 가장 높음
- 많은 시민은 3,500원대 요금으로 택시를 이용하지만, 택시 기사 수입에 차지하는 비중은 14,000원 이상의 요금대역이 큼



데이터 분석 단계

1. 데이터 분석 단계 요약

단계	주요 내용
분석 설계	<ul style="list-style-type: none"> ■ 기본개념 정의(Terms and Condition) ■ 분석설계 ■ 이동데이터 연계 ■ 제한점
데이터 수집 및 가공	<ul style="list-style-type: none"> ■ 필요 데이터 정의/수집/연계 <ul style="list-style-type: none"> • 필요데이터 정의 • 기상데이터, 시내버스 승하차데이터 등 수집 • 하차 정류장 추정 알고리즘 협업 • 택시 데이터 수집 ■ 데이터 품질 검증 <ul style="list-style-type: none"> • 데이터 확인 • 신뢰성, 일관성 확인 • 연계 가능성 검토
데이터 분석 및 검증	<ul style="list-style-type: none"> ■ 탐색적 분석(EDA) <ul style="list-style-type: none"> • 기초통계량 확인 • Data 전처리 ■ 교차분석 <ul style="list-style-type: none"> • 주요 요인간의 교차 • 단일차원 분석의 검증 • 인사이트 도출 ■ 추세/시계열 분석 <ul style="list-style-type: none"> • 동일 조건, 현상과 맥락 ■ 예측 분석 <ul style="list-style-type: none"> • 패턴/알고리즘 도출, 현장 적용 ■ 분석결과의 검증 <ul style="list-style-type: none"> • 기존 보고서, 현장 담당자 등에게 검토
데이터 시각화	<ul style="list-style-type: none"> ■ 대시보드 설계 ■ 분석결과의 연계 ■ 스토리 라인 ■ 시스템에 탑재

2. 분석 설계

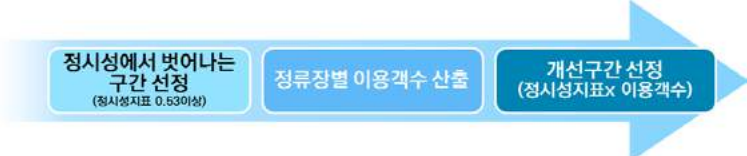
- 분석 프로젝트의 필수단계중 하나로 분석할 범주와 측정값의 범위 및 형식을 정의함
- 최종 산출물의 정의에 따른 분석 단계의 연계를 설계함

■ 기본 개념 정의 (Terms and Condition)

- 데이터를 통하여 분석할 수치의 조작적 정의 단계
- 분석의 범위와 한계 등의 조건을 명시

과제	<ul style="list-style-type: none"> ■정시성의 정의 ■대조군과 실험군 정의 ■기상조건의 정의 ■분석대상 노선과 택시의 정의
추진내용	<p>■정시성의 정의 : 버스의 차고지 출발간격이 각 정류장별 버스도착 간격으로 지켜지는 정도</p> <div data-bbox="432 938 1189 1250"> </div> <p>※ 참조 논문 : "버스 운행 신뢰성 평가를 위한 정시성지표의 개발 및 적용" 고승영외 2인(서울대학교 지구환경시스템공학부 교수), 대한교통학회지 제23권 제2호, 2005년 4월</p> <ul style="list-style-type: none"> ■대조군과 실험군의 정의 <ul style="list-style-type: none"> • 대조군 : 맑은 날(28일) • 실험군 : 비오는 날(26일) ■기상조건 : 강우, 강풍, 한파, 폭염, 미세먼지, 강설 ■분석대상 <ul style="list-style-type: none"> • 시내버스 : 상위 40개 노선, 하위 9개 노선 • 마을버스 : 상위 40개 노선, 하위 4개 노선 • 택시 : 법인택시 13,883대
Tip	<ul style="list-style-type: none"> ■분석설계에서 가장 중요한 것은 조작적 정의(Operational Definition)임



과제	<ul style="list-style-type: none"> 정시성 우선개선 정류장 선정 방법
추진내용	<ul style="list-style-type: none"> 정시성에서 벗어나는 구간 선정 <ul style="list-style-type: none"> 기존 논문의 정의를 존중하여 정시성 위배가 0.53이상인 구간을 먼저 추출 BI Tool등을 활용하여 0.53이라는 기준을 다른 수치로 자유롭게 변경하며 적용 가능  <ul style="list-style-type: none"> 정류장별 이용객수 산출 <ul style="list-style-type: none"> 각 정류장에서 버스에 탑승한 이용객수를 산출 우선 개선 정류장 선정에 가중치(Weight)로 사용함 개선 구간 선정 <ul style="list-style-type: none"> 기존 방식을 이용하여 List Up된 정시성 개선 후보 정류장들에 이용객수라는 가중치(Weight)를 적용 우선순위를 고려한 순위 재산정
Tip	<ul style="list-style-type: none"> 기존 공급자 중심의 관점을 시민 불편 관점으로 재조명하는 데 의의가 있음

※ 기존 논문이 제시하는 정시성 지표구간의 정의

• 차두간격의 균등성의 서비스 수준 기준

차두간격의 분산계수	이용자/운영자 관점
0.00 ~ 0.21	정확한 서비스 제공
0.22 ~ 0.30	차두시간에서 종종 벗어남
0.31 ~ 0.39	차두시간에서 가끔 벗어남
0.40 ~ 0.52	불규칙, 가끔 차량 몰림 현상
0.53 ~ 0.74	잦은 몰림 현상
≥ 0.75	대부분 차량이 지체를 경험

※ 기존 논문의 정시성 지표 정의

$$\text{정시성 지표}(P_2) = \frac{S_2^2}{h_t^2}$$

- P_2 = 운행 계획표 상의 배차 간격과 실제 운행시격과의 차이
- ※ 사전에 작성된 운행 계획표는 있으나, 실제 운행 횟수가 정해진 운행 횟수와 다를 경우 사용

$$S_2^2 = \frac{1}{I-1} \sum_{i=2}^I (h_i - h_t)^2$$

- t_i : i 번째 차량의 실제 정류장 도착 시간
- I : 운행 횟수
- h_t : 운행계획표상의 운행시격
- $h_i = t_i - t_{i-1}$: i 번째 차량의 운행시격



■ 분석 설계

- 분석을 진행하기 위해 필요한 단일 통계치와 다변량 통계, 그리고 목표에 이르기 위한 연계 분석 등의 구조를 사전에 설계하고 정의하는 과정

과제	<ul style="list-style-type: none"> ■ 기본통계 ■ 비교분석 방법 ■ 하차지 추정 ■ 분석의 순서/Hierarchy 정의
추진내용	<ul style="list-style-type: none"> ■ 기본통계 <ul style="list-style-type: none"> • 분석목표에 따라 정시성을 계산하는 구조 확정 • 탐색적 분석(Exploratory Analysis)을 통해 정시성의 분석에 투입되는 주요 Dimension(정류장, 노선, 날짜, 기상조건, 시간대)별 Measure(이용량, 정시성) 산출 ■ 비교분석 방법(이용량, 정시성) <ul style="list-style-type: none"> • 주요 Dimension을 서로 교차하여 분석함으로써 정류장별 정시성의 차이를 규명 • 우선관리대상 정류장 발체의 기본 비교수치 생성 ■ 하차지 추정 <ul style="list-style-type: none"> • 서울/경기와 달리 부산은 하차 시 버스카드를 Tag하지 않음 • 철도기술연구원의 지원을 받아 분석에 쓰인 모든 승차정보에 하차지점 정보(정류장)를 추가함 ■ 분석의 순서/Hierarchy 정의 <ul style="list-style-type: none"> • 목표하는 결과물에 이르기 위해 가장 우선시되는 분석단계를 정의 • 수반되는 분석주제들의 순서와 산출물/구조 등을 정의 • 절차마다 생성되는 결과물의 상호관계를 명시
Tip	<ul style="list-style-type: none"> ■ 전체 프로젝트의 성패를 좌우할 수 있는 가장 중요한 단계 중 하나가 분석 및 설계임 ■ 충분히 논의하고 결과물의 모습을 합의해야 결과에 대해 수용할 수 있음

■ 이종 데이터 연계

- 일반적인 데이터 분석의 경우에는 단일차원의 포맷으로 구성되어 있으나 대중교통 데이터 분석의 경우에는 위치정보 등이 추가로 등장함

과제	<ul style="list-style-type: none"> ■ 이종 데이터의 연계기준 정의 ■ 검증방법 ■ 데이터 보완
추진내용	<ul style="list-style-type: none"> ■ 이종 Data의 연계기준 정의 <ul style="list-style-type: none"> • 대중교통 이용 및 정시성 분석을 위해 사용되는 Data는 그 속성이 서로 다른 시스템에서 추출 • 정류장 위치 정보(번호, ID, 좌표, 지번) • 운행하는 버스정보(차량번호, 노선번호, 차고지 출발시간, 식별번호, 정류장 통과시간) • 승하차 Tag정보(식별번호, 승차지점 정보-정류장, 하차정류장-일부, 환승순서) • 정류장을 중심으로 각종 속성 정보를 연계 ■ 검증방법 <ul style="list-style-type: none"> • 이종 Data를 연결해서 구한 Measure가 UniDimension의 Measure와 동일한지 검증 • 교차분석의 결과가 UniDimension과 일관적인지 검증 • Data Set(연계 데이터 모음)이 제대로 통합되었는지 탐색적 분석(총 사례수, Measure, %, Dimension단위 탑승/정시성 등) ■ 데이터 보완 <ul style="list-style-type: none"> • 택시민원 분석의 경우, 지번이 아닌 인지적 지점(예: 롯데백화점 앞)으로 되어 있음 • 일일이 분류하여 POI로 매칭(좌표, 지번)
Tip	<ul style="list-style-type: none"> ■ 대중교통의 경우 목적과 포맷이 각각 다른 정보의 연계가 매우 중요 ■ 정류장 정보의 Update가 자주 발생하므로 주의



■ 제한점

- 짧은 프로젝트 기간과 강우일 수의 제한 등이 존재
- 유의미한 분석을 위해 시내버스 노선 및 일자에 제한을 가함

과제	<ul style="list-style-type: none"> ■ 데이터 수집 기간 ■ 분석 대상 교통수단 ■ 분석 대상 기상조건
추진내용	<ul style="list-style-type: none"> ■ 데이터 수집기간 <ul style="list-style-type: none"> • 2018년 6월 ~ 2019년 6월(1년) • 대상지역: 부산 전지역 ■ 분석대상 교통수단 <ul style="list-style-type: none"> • 대중교통 종류: 시내버스, 마을버스, 도시철도, 경전철 • 정시성 분석대상 버스 노선번호와 일자 선정(제한) ■ 분석대상 기상 조건 <ul style="list-style-type: none"> • 분석대상 기간 중 비 온 날: 20일 • 비교집단(맑은 날)은 비 온 날에 매칭 • 기타 기상조건 해당날짜와 사례수가 부족
Tip	<ul style="list-style-type: none"> ■ 날짜 수가 제한을 받게 되었지만, 노선의 선택은 객관성을 유지

3. 데이터 수집 및 가공

■ 필요 데이터 정의/수집/연계

- 프로젝트의 목표 결과물을 얻어 내는 데 필요한 데이터를 정의하고 확보하는 단계
- 여러 가지 Table단위의 개별 데이터를 수집한 후 관계에 맞게 연계하는 작업이 중요

과제	<ul style="list-style-type: none"> ■ 필요데이터 정의 ■ 기상데이터, 시내버스 승하차데이터 등 수집 ■ 하차정류장 추정 알고리즘 협업
추진내용	<ul style="list-style-type: none"> ■ 필요데이터 정의 <ul style="list-style-type: none"> • 분석결과의 목표 산출물에 필요할 것으로 판단되는 데이터 Source(Table)와 Field를 정의 • 분석의 결과를 도출하기에 충분한 기간 혹은 구간을 정의 • 1년을 전체구간으로 하되, 기상조건에 해당되는 날을 우선으로 하고 같은 조건의 맑은 날을 1:1로 매칭 ■ 기상데이터, 시내버스 승하차데이터 등 수집 <ul style="list-style-type: none"> • 데이터 정의에 맞춰 기상데이터, 시내버스 승하차 데이터 등을 수집 • 이종 데이터간(승하차데이터, 정류장 정보, 운행정보)의 연계를 위해 좌표 정보/지번 정보/행정 구역 정보 등을 보완 ■ 하차 정류장 추정 알고리즘 협업 <ul style="list-style-type: none"> • 철도기술연구원(KRRI)에 분석을 위해 최종 추출된 승차데이터를 보냄 • KRRI에서 개발한 알고리즘을 적용하여 모든 승차데이터에 추정 하차 지점을 추가하여 자체 검증 • 추정 하차지점을 추가한 데이터를 수령하여 기존 데이터 셋과 교체 ■ 택시 데이터 수집 <ul style="list-style-type: none"> • 승하차 데이터(법인택시, 개인택시) • 법인택시 승하차 데이터는 한국교통안전공단의 협조를 얻어서 수집 • 민원데이터 수집(해당 부서)
Tip	<ul style="list-style-type: none"> ■ 데이터 처리능력보다 협의/소통/협업 능력이 수집단계 성공의 관건임



■ 데이터 품질 검증

- 수집되는 데이터의 상태가 분석에 적합한 지를 검토하는 단계로서 분석 단계에 들어가기 위한 필수 과정

과제	<ul style="list-style-type: none"> ■ 데이터 확인 ■ 신뢰성, 일관성 확인 ■ 연계가능성 검토
추진내용	<ul style="list-style-type: none"> ■ 요구데이터 여부 확인 <ul style="list-style-type: none"> • 수집된 Data의 내용(Field)이 분석에 필요한 요청 Data와 일치하는 지 확인 • 확보된 Field들이 요구한 Field의 속성을 충족하는 지 확인 ■ 신뢰성 검증 <ul style="list-style-type: none"> • 확보된 Field의 결측치 여부 • 각 Field의 Value가 기대한 것을 충족하는 지 확인 • 각 Field의 Value에 Outlier가 있는 지 확인 ■ 포맷의 일관성 <ul style="list-style-type: none"> • 문자, 숫자, 날짜 등 각 Field의 Value가 일관적으로 들어 있는 지 확인 • 각 Field의 Value가 원하는 형식으로 변환 가능한지 판단 • 연관된 Table의 동일 Field와 공통된 속성의 Format인지 확인 ■ 연계성 검토 <ul style="list-style-type: none"> • 서로 다른 Table을 연결하기 위한 공통 Field를 지정가능한 지 검토 • Table을 서로 연결했을 때 기대하는 만큼의 Matching이 일어나는지 확인(두 Table의 합집합 또는 교집합)
Tip	<ul style="list-style-type: none"> ■ 이 단계를 통해 수집된 데이터의 건강상태를 따져보게 되므로 Table이름과 Field의미까지 철저히 살펴봐야 함

4. 데이터 분석 및 검증

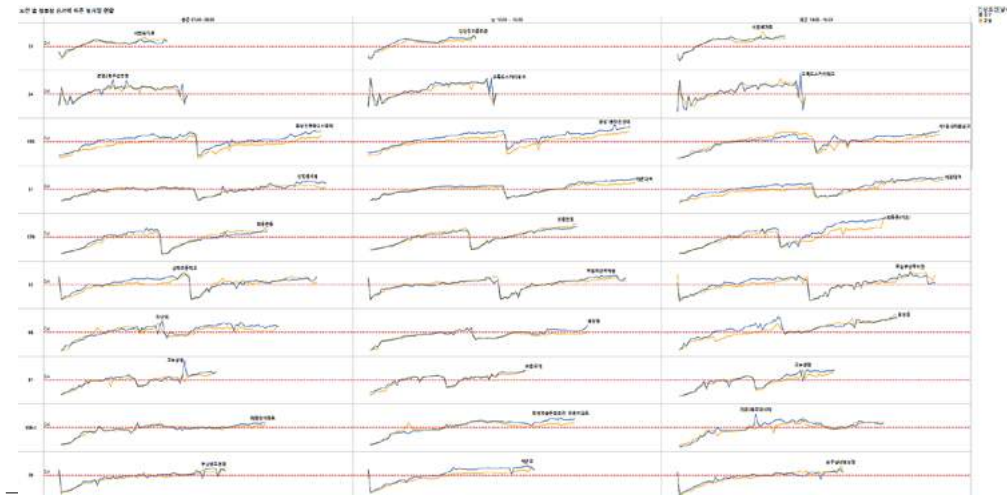
■ 탐색적 분석(EDA)

- 본격적인 분석에 들어가는 과정의 첫 단계로서 전체적인 윤곽과 주요한 관점을 파악

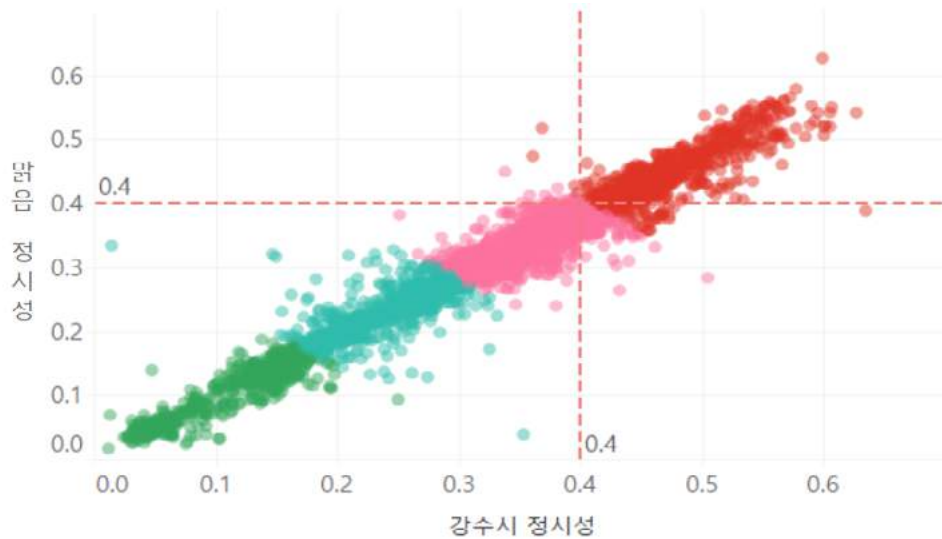
과제	<ul style="list-style-type: none"> ■ 주요 Dimension의 각각의 기초통계량(Sum, Frequency, %) 확인
추진내용	<ul style="list-style-type: none"> ■ 기초통계량 확인 <ul style="list-style-type: none"> • 주요 차원별로 합계, 평균, 개수, 비율, 결측치 등을 계산 (Parameters 포함) • 전체를 대표하고 있는 지 여부를 확인(신뢰도, 타당도) • 주요 Dimension을 상호교차분석하여 Measure가 의미있게 분할되는지 확인 (인구수 vs. 남녀) • 다차원 분석표의 비율, 개수, % 등에서 논리적 오류가 없는 지 확인 • 검증된 Dimension별로 전체를 대표하는 값(Measure)을 정의 (예) 전체 대중교통 이용량(#, %): 일별, 시간대별, 노선별, 정류장별 ■ Data 전처리 <ul style="list-style-type: none"> • 탐색적 분석 및 기초통계분석을 위하여 원시자료를 가공하는 작업 (예) 연월일시가 한꺼번에 연결되어 있는 Field를 년, 월, 일, 시로 분리: 201906091329 → 2019년 06월 09일 13시 29분 • 분석에 필요한 Data형태로 가공하는 과정에서 Outlier를 제거하거나 변경하고 논리적 산술과정을 거쳐 새로운 Field를 생성하기도 함 • 각기 다른 Raw Data Set을 가공한 후 원하는 통합 Data Mart를 만드는 것이 일반적
Tip	<ul style="list-style-type: none"> ■ 데이터 값의 편포, Value의 특성(문자/숫자/Code)등을 일목 요연하게 정리하는 절차 ■ 대체로 단일 차원의 데이터 탐색이 주를 이루지만, 일부 주요한 변수끼리의 조합을 통한 검증도 이루어짐



※ EDA분석의 예: 노선별 정시성 튀는 정류장 검토



※ 교차분석의 예: 정시성 교차분석(맑음 vs. 강수)



■ 교차분석

- 탐색적 분석에서 선발된 주요 변수들을 서로 교차하여 연관성이나 규칙을 찾아 내는 단계로 분석결과와 초보적인 인사이트를 도출

과제	<ul style="list-style-type: none"> ■ 주요 요인간의 교차 ■ 단일차원 분석의 검증 ■ 인사이트 도출
추진내용	<ul style="list-style-type: none"> ■ 주요 요인간의 교차 <ul style="list-style-type: none"> • 주요 요인(비가 오는 날 vs 맑은 날)을 중심으로 승객수(#)와 비율(%) 등을 비교분석 ■ 단일차원 분석의 검증 <ul style="list-style-type: none"> • 교차/비교 분석한 결과와 단일차원 집계값의 비교를 통해 신뢰성과 타당성 검증(분석에 사용할 의미가 있는 지 여부 판단) <p>(예) 부산지역은 강설과 영하의 기상조건에 부합하는 교통량 분석이 불가 (← 사례수가 너무 적음)</p> ■ 인사이트 도출 <ul style="list-style-type: none"> • 주요 요인간의 교차분석을 통해 관련되어 있는 변수/요인간의 상호관계를 통합 분석하는 Insight를 얻게 됨 <p>(예) 시간대별 정시성 분석, 강우가 정시성에 미치는 영향, 승객수와 정시성의 관계, 정시성 개선 우선 지역</p> <p>(예) 노선별*정류장별: 승객수(#, %)</p> <p>비오는 날*비오지 않는날: 승객수(#, %)</p>
Tip	<ul style="list-style-type: none"> ■ 교차분석은 선택적 조합을 통한 분석으로까지 발전하면서 결론의 논리를 펼치게 됨 ■ 한 가지 주제를 너무 여러 가지로 분석하기보다는, 여러 번의 교차분석과 다차원 연계분석을 반복함으로써 윤곽을 좁혀 가는 것이 바람직함

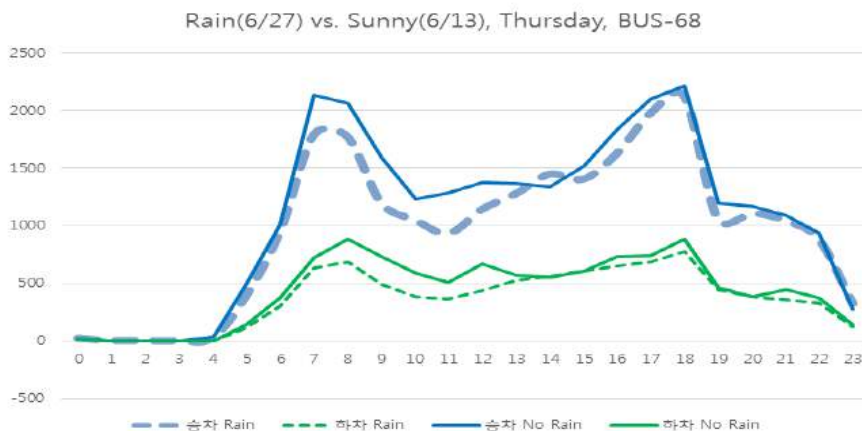


■ 추세/시계열 분석

- 시간의 흐름을 따라 분석하고자 하는 변수의 값을 비교 분석하는 기법으로 흔히 이전 동기 대비 일정한 효과가 반복되는 지를 검증하는 데 쓰임

과제	<ul style="list-style-type: none"> ■ 동일 조건 ■ 현상과 맥락
추진내용	<ul style="list-style-type: none"> ■ 동일 조건하에서 비교 <ul style="list-style-type: none"> • 전체를 대표하는 값(평균, 합계, 횟수, %)을 시간의 순서로 나열해보면, 전체적인 맥락 파악 가능 (예) 날짜/요일 순서별 대중교통 이용량 분석, 하루의 시간대별 정시성 변화 • 추세/시계열 분석에서는 시간대별 파라미터 비교 조건을 동일하게 유지하도록 유의해야 함 (예) 비오는 날의 시간대별 정시성 vs. 맑은 날의 시간대별 정시성 (같은 노선, 같은 요일, 같은 날) ■ 현상과 맥락을 연계해서 분석 <ul style="list-style-type: none"> • 현상은 특정시점을 지정하거나 특정기간을 대변함. 그 현상의 앞이나 뒤에 위치한 또 다른 현상들을 같은 조건으로 연결하여 비교하면 맥락 (Context)을 파악할 수 있으며 방향성을 예상할 수 있음 (예) 월별, 시간대별, 요일별 분석
Tip	<ul style="list-style-type: none"> ■ 동일 조건을 최대한 통제하는 것이 추세/시계열 분석의 고려사항임

※ 시계열 분석의 예(비 오는 날의 이용객수 차이, 68번 버스)



■ 예측 분석

- 과거의 이벤트 발생에 대한 주요 요인간의 기여와 구조를 이해하면 같은 조건의 다음 이벤트 발생에 대한 확률을 계산할 수 있는데, 이와 같은 접근을 기초로 예측에 활용

과제	<ul style="list-style-type: none"> ■ 알고리즘/모델 도출 ■ 패턴/조건의 적용 ■ 알고리즘/모델의 현행화 ■ 지수 vs 확률
추진내용	<ul style="list-style-type: none"> • 예측분석의 시작점은 어떤 현상의 발생에 관한 관련변수/요인 간의 변화 규칙을 찾아내는 것(Algorithm, Model) • 보통 가장 많이 사용하는 기법은 통계의 Regression(회귀) 방정식임 • 사전 단계의 탐색적 분석을 통해 회귀분석에 적절한 구간과 조건을 정의함 • 해당 조건 안에서 선형관계를 도출하여 관심의 대상인 현상에 영향을 미치는 input변수의 중요도를 파악(Model 도출) (예) 정류장별 정시성에 영향을 미치는 변수의 비중: 강우 > 노선 > 시간대 • Model에서 선정된 주요 변수의 값을 조절해 가며 목표로 하는 변수의 Parameter가 어떻게 변하는 지 확인 • 단, 과거 관측치 범주에서 살펴보고 있는 한계가 있으므로, 과거 조건을 넘는 예측은 계수 조정이 필요함(강우량 변화 → 정시성 악화) • 무엇보다 중요한 것은 Model의 예측력(Performance Power)을 유지하는 것 → 지속적인 Data Update • 지수(Indicator, Index)는 현재 상태 또는 정도를 하나의 수치로 대변할 때 쓰이지만, 확률은 목표로 하는 상황이 일어날 가능성(Probability)을 의미함 (예) 조건별 확률, 조건별 지수 (예) indicator : a thing, especially a trend or fact, that indicates the state or level of something (예) probability : the extent to which something is probable; the likelihood of something happening or being the case
Tip	<ul style="list-style-type: none"> ■ 인공지능(AI)을 적용한다 해도 기본적인 알고리즘의 정립과 환경 및 적용 조건이 중요함



■ 분석결과의 검증

- 분석한 결과를 확정하기 전에 거치는 절차로, 일종의 현장성 검토 작업임
- 사후, 심도 있는 검토를 위해 꼼꼼한 자료작성 및 기록이 필요

과제	<ul style="list-style-type: none"> ■ 기존 보고서와 비교 ■ 현장 및 담당부서 인터뷰 ■ 분석결과 수치의 정합성 검토 ■ 분석방법론의 Documentation
추진내용	<ul style="list-style-type: none"> ■ 기존 보고서와 비교 <ul style="list-style-type: none"> • 분석한 결과의 신뢰도와 타당도를 검증하기 위한 방법 중 대표적인 것은 기존분석/보고서와 일관적인 값을 나타내는 가를 검토하는 것임 • 같은 조건의 분석이어야 완벽한 검증이 가능하지만, 일부 통제된 조건에 국한한 비교도 검증차원에서는 유용 • 학술논문, 해안 등의 유사사례 등을 수집하여 방법론, 사용 데이터, 분석조건 등을 따져 이번 분석결과와 비교 ■ 현장 및 담당부서 인터뷰 <ul style="list-style-type: none"> • 분석된 대중교통 분석결과(정류장별 정시성)를 현장 담당자의 경험에 비추어 검증함 (예) 정시성이 나쁘게 평가된 정류장이 실제로 그런지 아니면, 분석의 오류인지 검증 (예) 제시한 우선순위가 현실성을 충분히 담보하는 지 검토 ■ 분석결과 수치의 정합성 검토 <ul style="list-style-type: none"> • 분석된 결과의 2차 가공 혹은 지수화 과정에서 원래 값과의 모순관계가 발생할 수 있음 → Model을 수립하는 과정에서 주요변수의 이상치를 발견하게 됨 → 주요변수의 결측치 처리 방침을 정하게 됨(보간법, Missing) • 분석과정에 대한 Documentation은 매우 중요 → 변수 생성 과정, 근거, Data 출처 등 기록
Tip	<ul style="list-style-type: none"> ■ 프로젝트의 결과와 과정을 논리적으로 축적하는 일이므로 검증절차와 꼼꼼한 기록이 중요함

5. 데이터 시각화

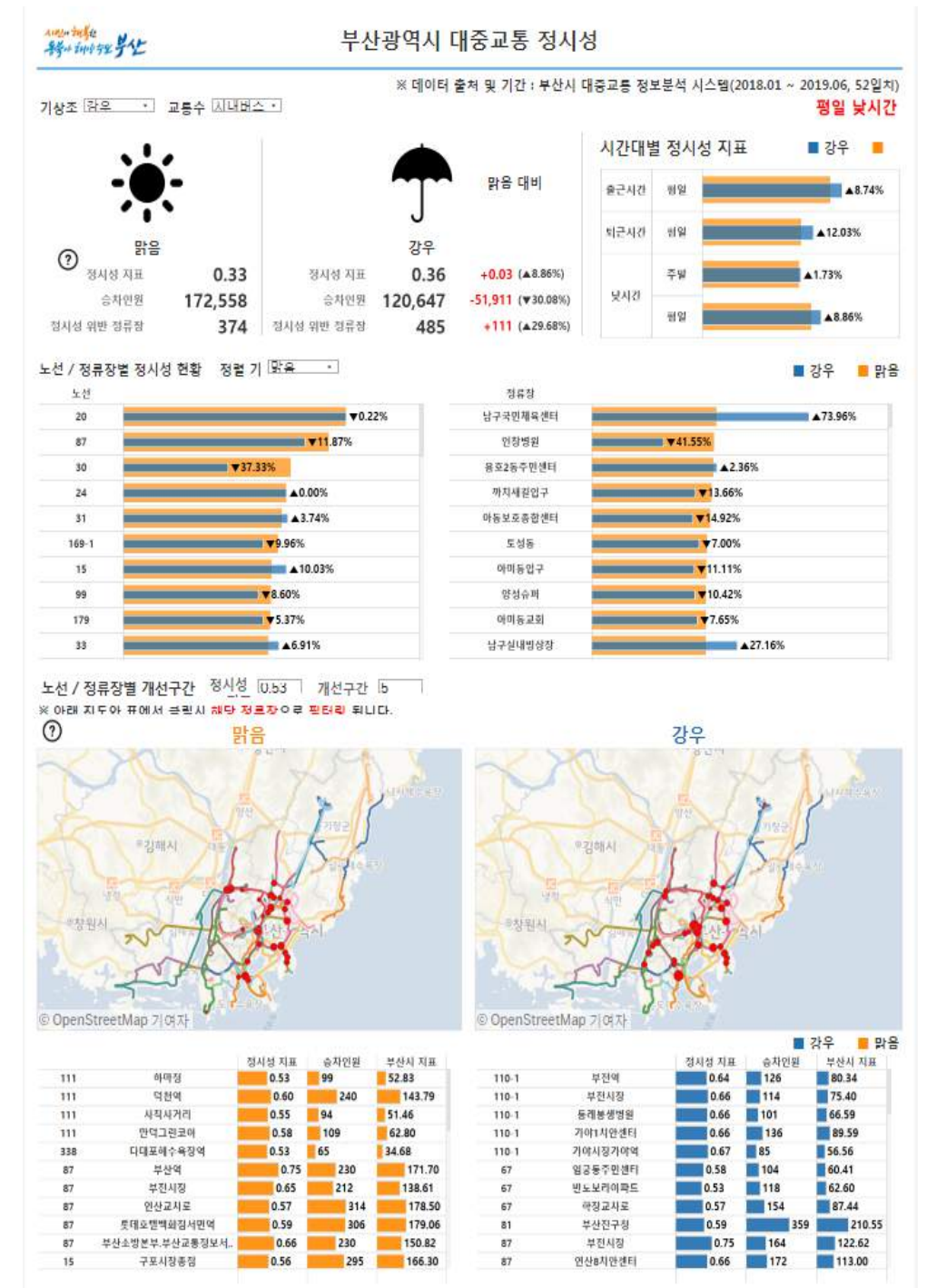
■ 대시보드 설계

- 분석결과의 핵심내용을 논리적으로 대시보드(대시보드)에 담아내는 작업은 전체 작업의 결과를 표현하는 중요한 과정임

과제	<ul style="list-style-type: none"> ■ 분석결과의 범주화(주제화) ■ 전개에 따른 우선순위 ■ 연관성을 고려한 분석 요약과 배치
추진내용	<ul style="list-style-type: none"> ■ 분석결과의 범주화 <ul style="list-style-type: none"> • 분석 대주제별로 하위/세부 분석 소주제들이 어떻게 편성되는 지 설계 • 도면(Diagram)으로 주요 결과치의 Display와 선후관계(Process)를 정의(담당자 지정, 검수일정, 검증방법, Data 구조화 등도 함께 검토) ■ 전개에 따른 우선순위 <ul style="list-style-type: none"> • 각각의 대주제와 대표적인 Worksheet들을 선택 • 대시보드를 구성하기 위한 구성요소와 논리전개에 맞게 상하위 개념으로 정리(분석의 목표에 맞는 Story Line 고려 필요) ■ 연관성(요약/배치) <ul style="list-style-type: none"> • 연관되어 있는 개념의 분석을 서로 연결하여 제시하는 전략구성 • 분석주제를 대시보드의 형태로 재구성, 인지구조에 맞게 배치함
Tip	<ul style="list-style-type: none"> ■ 단순한 대시보드 구현을 위한 전산적인 설계에 초점을 두기보다는 분석결과를 알기 쉽고 간결하게 전달하는 흐름을 고려하여야 함 ■ 과도한 정보를 주기보다는 핵심적인 내용의 연계에 주력하는 방향으로 꾸미는 것이 효과적임



※ 대시보드 구현 화면 예시(대중교통 정시성)



■ 분석결과와의 연계

- 단일 차원의 분석에서 Parameter들을 얻는다면, 교차분석 또는 연계분석(조건문 투입)을 통하여 분석결과와의 논리를 유도하기 때문에 대시보드 상에서 이러한 전개가 가능하도록 설계함

과제	<ul style="list-style-type: none"> ■ 주제별(범주별) 논리의 구성 ■ 분석 논리에 맞게 결과를 연계 ■ 조건을 달리하여 쉽게 비교해 볼 수 있도록 구성
추진내용	<ul style="list-style-type: none"> ■ 분석결과와의 구성 <ul style="list-style-type: none"> • 분석결과를 목표에 맞는 큰 주제를 기준으로 재 정돈 • 분석결과를 관점에 맞춰 재 정돈 ■ 분석 논리에 맞춰 연계 <ul style="list-style-type: none"> • 기본적인 분석결과를 바탕으로 심도있는 분석으로 이어지는 분석결과를 다양한 기법으로 연계(Filter, Select) (예) 전체 교통량 → 기상조건별 교통량 → 노선별 교통량 → 정류장별 교통량 ■ 쉽게 비교 <ul style="list-style-type: none"> • Simulation : 기준치를 Slide Bar형태로 구성하여 좌우로 조절하며 조건에 따라 달라지는 분석결과를 쉽게 비교할 수 있게 함 • 비교분석 : Combo버튼이나 Check Box등을 이용하여 단일한 조건 외에 조건들의 융합이나 배제 등을 구현, 조건에 따른 분석결과와의 변화를 비교
Tip	<ul style="list-style-type: none"> ■ 다양한 조건의 조합을 쉬운 방법으로 변경하면서 분석결과를 비교하고 인사이트를 제시하는 것이 관건 ■ 모든 경우의 수를 Filtering할 수 있게 구성하기 보다는 사전에 분석한 논리를 반영하여 핵심적인 조건의 조합을 제시하는 것이 효과적임 ■ 여건이 가능하다면, 간단한 분석 Widget을 제시하여 사용자가 쉽게 응용할 수 있도록 하는 것이 바람직함



■ Story Line

- 분석결과와 인사이트를 상대방에게 전달하고자 할 때 어떤 순서와 맥락으로 전달할 것인지를 정하는 절차로 일종의 시나리오 구성 단계임

과제	■ Audience의 눈높이 고려
추진내용	<ul style="list-style-type: none"> • 이야기하기 좋은 방식으로 설계 • 일정한 수준의 위젯 혹은 표준분석 시나리오 제시 • 대시보드의 설계구현에 있어서 무엇보다 중요한 것은 사용자의 기대 수준에 부합하는 것임(낮은 기대수준 → 높은 기대수준) <ul style="list-style-type: none"> → 실무자의 업무를 대신 또는 경감하는 수준 → 현황을 지수로 요약 → 개선점이나 우선순위를 제시 → 전략적 고려사항을 판단하게 지원 (예) 정시성 개선 우선순위 제시 • 분석 전문가가 아닌 사람이 분석 시나리오를 수립하기 쉽지 않음 <ul style="list-style-type: none"> → 현장의 주요근거 수치 혹은 논리를 중심으로 Insight(문제점/해결책)를 도출하도록 순서를 배치 • 가장 보편적인 주제를 소재로 보거나 예제를 구성하여 제시 <ul style="list-style-type: none"> → 사용자/현업의 응용분석을 유도 → 추가 개선사항을 유도
Tip	<ul style="list-style-type: none"> ■ 무엇보다 중요한 것은 Audience가 듣고자 하는 주제와 쉽게 이해할 수 있는 순서로 전달하는 것임 ■ 기술적 용어나 분석팀의 주장이 앞서기 보다는 현장 담당자가 가장 힘들어 하는 부분을 선명하게 해결해 주는 전개가 바람직 함 ■ 가급적 현실적인 예와 상황을 활용하여 분석결과와의 적용을 제시하는 방법이 효과적임

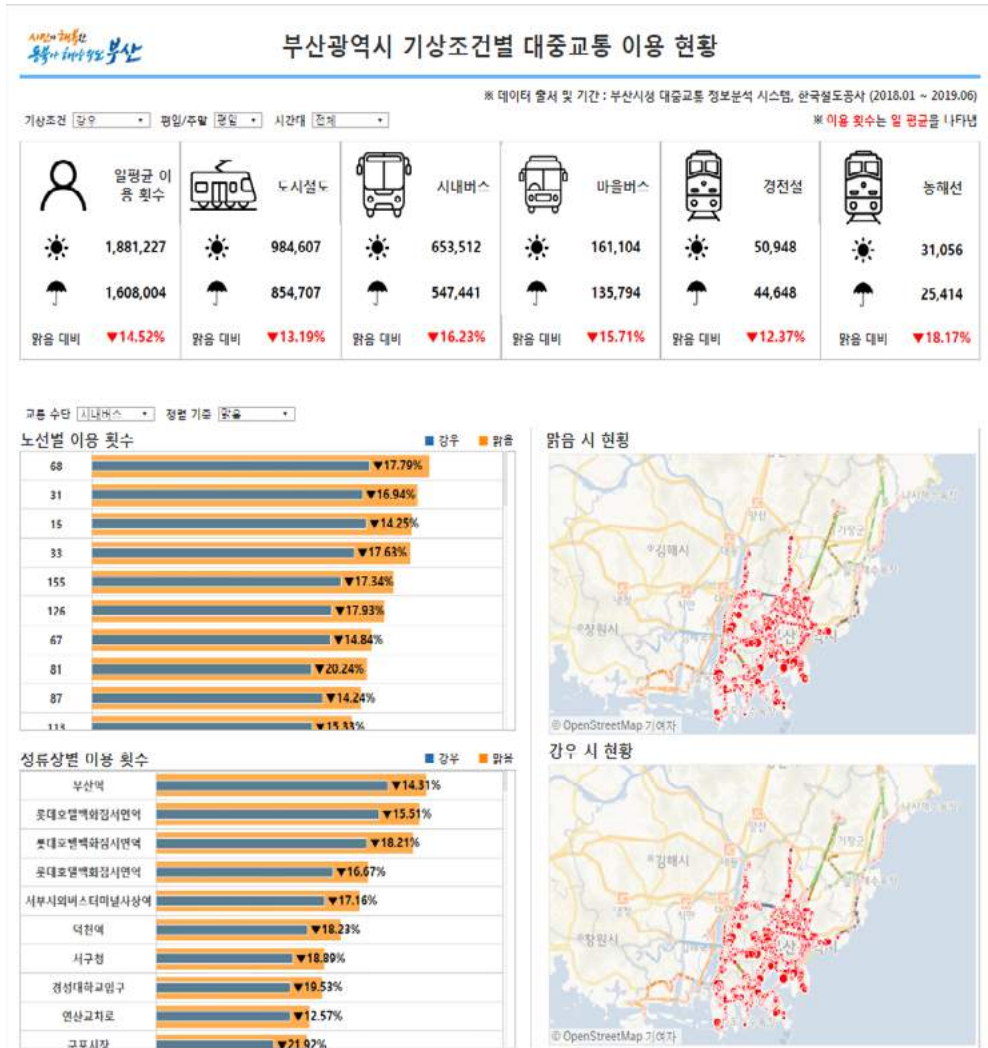
■ System Delivery

- 분석결과 및 대시보드를 시스템에 올려 많은 사용자에게 동질의 서비스를 제공하는 단계

과제	<ul style="list-style-type: none"> ■ 내부망 공유 ■ 대시민 공유
추진내용	<ul style="list-style-type: none"> ■ 내부망 공유 <ul style="list-style-type: none"> • 부서별 조회권한 부여 • 원본 변경 혹은 추가 Data 연계분석 가능 • 단순 조회(시나리오 분석) ■ 대시민 공유 <ul style="list-style-type: none"> • 부산은 빅데이터 포털이 망 외부에 위치 • 분석결과의 공유 & Download 가능 • 자기 Data의 연계 가능
Tip	<ul style="list-style-type: none"> ■ 일반적인 지식 수준을 가진 사람들이 손쉽게 접근하고 이용할 수 있도록 배려 ■ 시민들의 참여와 체험을 보장할 수 있도록 시스템 설계와 배치를 기획 ■ 매뉴얼을 마련하여 배우고자 하는 계층의 욕구에 체계적 대응



※ 대시보드 구현 화면 예시(대중교통 이용현황)





행정적용 단계

1. 행정적용 단계 요약

■ 분석요건의 정의에 담당부서의 의견 반영

- 현황과 개선목표에 대한 뚜렷한 구분
- 개선의 명분을 위한 구체적인 분석결과의 내용을 기술
- 담당부서가 목표로 하는 가시적인 개선결과 및 기대효과 정의

■ 분석의 중간결과를 공유

- 분석을 진행하면서 의미있는 경과가 있을 때마다 현장 및 담당부서와 공유하여 그 결과의 실용성을 검토
- 담당부서의 시정 또는 보완의견에 대해 즉각적인 재검토 및 피드백이 필요

■ 최종 분석결과를 현장에 적용하기 전에 현실성과 기대효과를 사전 검토

- 현장 및 담당부서와 많은 협의를 거쳐 도달한 결과물이라도 실제 단계적 적용의 현실성을 검토해야 함
- 현장적용에 따른 정량적인 개선 기대효과를 실제 적용 전에 추정(Estimating)해야 함

■ 적용 후 개선효과 검증

- 단계적 적용계획을 실제로 실천할 때는 기저수치(Baseline)를 사전에 정의해야 함
- 적용에 따른 순개선효과(Additional Growth)를 정확히 계산하여 다른 추가적인 외적 요인을 통제된 상태의 수치를 획득하여 개선효과를 검증해야 함



※ 현장적용 방안 예시(68번 버스 정시성 우선 관리대상 정류장)



2. 행정정책개선

■ 시민의 편의 증대

- 빅데이터의 행정서비스 활용 목적 중 가장 큰 항목은 시민의 편의 증대
- 대개 한정된 재원으로 욕구와 수요의 충족에 대응하기 위해 사용하는 빅데이터 분석 방법은 우선순위를 제시하는 것
- 사각지대 혹은 상대적 소외계층을 제시하는 형태로 쓰임
- 이 프로젝트에서는 각 정류장의 승하차 인원을 정시성 지수에 가중치로 부여하여 관리 대상 정류장 우선순위에 반영함

■ 행정의 효율 개선

- 이 프로젝트에서는 대중교통 담당자가 일일이 확인하거나 추적할 수 없는 분량의 데이터 및 검토항목을 빅데이터 분석 기법을 활용하여 알기 쉽게 조건에 따른 결과를 찾도록 구성함
- 이로 인해 대중교통 정시성 관리에 필요한 시간과 인력의 효율이 획기적으로 개선됨

■ 향후과제

- **전 노선 확대**(145개 노선) : 확립된 방법으로 전 노선에 걸쳐 정시성을 분석하고 관리 대상 정류장의 우선순위를 부여
- **도시철도·버스·마을버스 등 환승루트로 확대** : 버스 탑승 시민에 국한하여 분석한 결과 이므로, 환승승객에까지 확대하여 모델을 확장
- **기상조건에 따른 시민생활의 불편개선 추가 발굴**(예 : 통증)
- **상시분석** : 확보된 데이터와 가공방법을 자동화하고 DB화하여 시간의 흐름에 따라 누적된 결과와 변화를 쉽게 알 수 있도록 구현
- **사전예고제** : 정시성의 일정한 패턴과 주요 요인간의 영향요인을 모델로 구성하여 기상 조건 예보시점에 예상되는 정류장별 정시성 지수를 시민들에게 예고하고, 담당자는 필요한 대책을 사전에 수립하여 대응





전기차 충전 인프라 운영·관리 및 콜센터 운영 정책 수립을 위한 EV콜센터 민원 분석

| 제주특별자치도 / 노희섭·김기홍·박기범, 한국지역정보개발원 / 강민선 |

요약

제주특별자치도는 한국지역정보개발원과의 협업으로 제주 전기차 EV콜센터로 접수된 전기차 및 전기차 충전소 관련 민원 빅데이터 분석을 실시하였다.

제주지역 전기차는 2013년 302대, 2014년 674대에 불과하였으나, 2018년 3월 기준 10,006대의 전기차가 제주지역에 등록된 것으로 나타나, 전국 지자체 최초 전기차 운영 1만대 시대에 진입하였다.

또한 제주지역에서는 탄소제로섬 실현을 위한 전기차 보급 확대 및 연관 산업 육성을 하기 위한 다양한 시도를 하고 있으며, 개방형 충전기 제도 보완, 다중 집합장소·공동주택 멀티 충전시스템 등 민간 충전기 서비스 인프라 확충으로 이용자의 편의를 증진시키고 있다.

그러나, 전기차 보급 확대 및 전기차 렌터카 이용객 급증에 따라 발생하는 민원 또한 증가 추세에 있어, 도청에서는 전기차를 사용하는 도민과 관광객들이 전기차 이용 및 충전과 관련해 겪는 어려움을 해결하고자 ‘제주 EV콜센터’를 운영하고 있다.

콜센터에서는 365일 24시간 상담 업무를 진행하고 있으나, 가까운 충전소 위치나 충전기 사용법 등 단순 사용 미숙 민원 또한 상당수를 차지하고 있어 콜센터에 접수된 민원 내용을 분석하여 민원에 대한 재분류를 통해 효율적 안내 방법을 모색하고, 주요 고장 발생지 및 고장원인 분석을 통한 사전 예방 등 데이터에 기반한 콜센터 운영 정책 수립을 지원하고자 한다.

- 제주 전기차 보급률은 전국 기준 29%, 제주지역 전체 차량의 4% 점유
 - 전국 지자체 최초 전기차 1만대 시대 진입('18년 3월 기준 10,006대)
 - '19년 6월까지 등록된 자동차 대수는 386,690대, 전기차 대수는 17,079대로 전체 등록차량 대비 4.42% 차지, 전체 등록대수 대비 전기차 비율 증가 추세
- 전기차 용도별 등록현황은 자가용 13,121대(76.8%) > 영업용 3,523대(20.6%) > 관용차 435대(2.5%)순으로 자가용 비율이 현저히 높음
 - 영업용 전기차 중 렌터카는 2,817대(16.5%), 택시 601대(3.5%), 버스 103대(0.6%), 화물차 2대(0.0%)로 용도별로 다양하게 운행
- 전기차 충전기는 총 12,235기('19.6월 기준)로 전기차 증가에 따라 지속적으로 충전기가 추가·확대 중임
 - (관리주체별) <개방형> 2,398기(20%), <개인용> 9,837기(80%)
 - 개방형 충전기 중 공공기관 : 585기(5%), 민간사업자 : 1,813기(15%)
 - ※ 공공기관 : 지자체 및 정부(제주도청, 환경부 등)에서 설치한 충전기
 - ※ 민간사업자 : 한국전력공사, 한국전기차충전서비스, 한국전기차서비스 등
 - ※ 개인용 : 민간보급 전기차용 충전기, 공용으로 사용되지 않는 민간사업자 충전기
 - (충전방식별) <완속 충전기> 11,276기(92%), <급속 충전기> 959기(8%)
- 전기차 보급 확대 및 렌터카 이용객 급증으로 제주 EV콜센터에 접수되는 민원량 증가
 - (민원 접수량) 2015년 6월 ~ 2019년 5월(5개년)까지 총 53,378건
 - '15년 : 2,455건, '16년 : 4,110건, '17년 : 22,979건, '18년 : 14,997건, '19년 : 8,837건
 - (렌터카 이용자) 전체민원 중 렌터카 이용자가 접수한 민원은 31,557건으로 전체민원 대비 절반 이상(59.1%)을 차지
- 콜센터 접수 민원 분석을 통해 개방형 충전기 서비스 개선 사항 도출 필요
 - 주요 고장 발생지, 반복 민원사항 분석을 통한 민원량 감소 및 사전 예방책 마련
 - 단순 사용미숙에 대한 민원사항 사전 예방을 위한 효율적 안내방법 모색
 - 민원 분류 카테고리(대분류, 중분류, 소분류) 중복 확인 및 재정비를 위한 민원 내용 분석 및 민원 재분류 등



프로젝트 개요

■ 분석범위

- 시간적 범위 : 2015. 6. 1. ~ 2019. 5. 31. (총 5개년)
- 공간적·내용적 범위 : 제주 EV콜센터에 접수된 전기차/충전기 관련 민원

■ 주요 분석내용

- 반복민원 파악을 통한 행정 불편사항 해소와 담당자 및 이용자의 편의 증진 도모
- 단순 사용미숙에 대한 민원사항 사전 예방 안내방법 마련
- 주요 고장 발생지 및 고장 원인을 파악을 통한 현업 담당자의 업무 효율성 제고
- 도 구축 충전기 유료화에 따른 민원 유형 변동 분석 및 내·외부 개선점 모색
- 민원 분류 카테고리 재정비를 위한 민원 내용 분석 필요

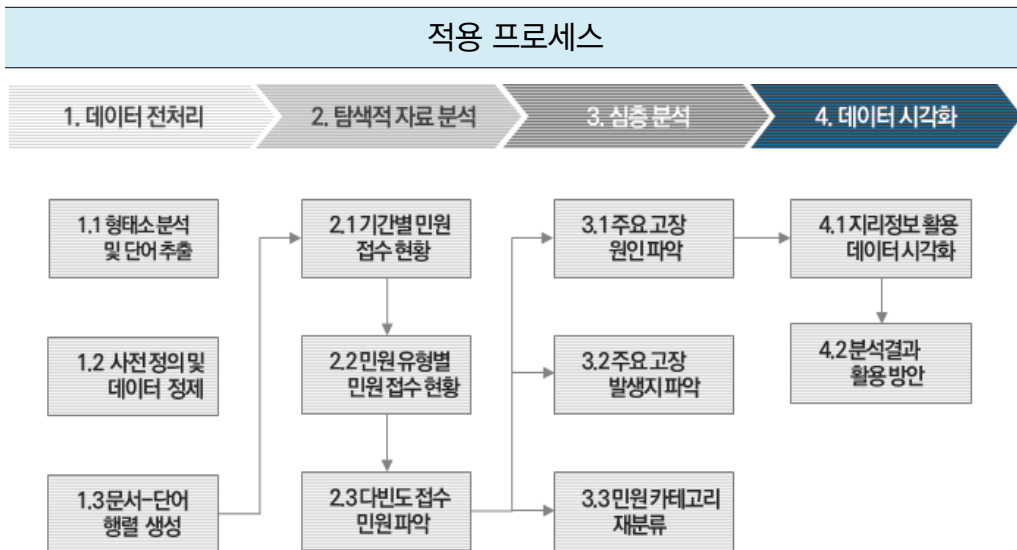
■ 활용데이터

No.	데이터명	컬럼명	비고
1	EV 콜센터 민원데이터	일자, 시간, 대분류, 중분류, 소분류, 상담 문의, 상담내용, 충전소명, 고객구분, 성별, 연령대 등	연도별 일부 상이
2	개방형 충전기 설치 현황	지역구분, 장소명, 주소, 충전기 설치 현황 (공공/민간) 등	-
3	충전기 위치 지리정보	위도, 경도 좌표값, 행정동 경계 등	-

다 적용 프로세스

■ 분석 전체 프로세스

- 텍스트 형태로 저장된 데이터 분석을 위하여 데이터 전처리 진행
- 탐색적 자료 분석(EDA)을 통해 연도별/유형별 민원 기초 통계 현황 파악
- 텍스트마이닝 기법 적용을 통한 심층 분석으로 세부 민원 분석 수행
- 분석 결과에 대한 GIS 기반 데이터 시각화를 통해 민원 콜센터 운영에 대한 정책적 시사점 도출



- 1. 데이터 전처리** : 특수문자 및 개인정보 처리 불용어, 동의어 처리, 형태소 분석, 사전(Dictionary) 정의, 문서-단어행렬 생성(민원 내 키워드 출현 빈도)
- 2. 탐색적 자료 분석** : 시기별 민원 발생빈도 추이분석, 민원 유형 분류, 다빈도 접수 민원 도출, 상위 키워드 추출
- 3. 심층 분석** : 주요 고장 원인 및 고장 발생지 파악, 동일 민원 중복 카테고리에 대한 카테고리 재분류 등
- 4. 데이터 시각화** : GIS 정보를 활용한 데이터 시각화 및 분석 결과 활용방안 모색





라 기획·관리 단계

1. 기획·관리 단계 요약

단계	주요 내용
문제의식	<ul style="list-style-type: none"> ■ 2018.3.2 제주 전기차가 전국 지자체 최초로 1만대 시대(10,006대)로 진입 하면서, 이후 관련 민원의 효율적 처리에 대한 요구 증대 <ul style="list-style-type: none"> - 제주 전기차 보급률은 전국 기준 29%, 제주 지역 전체 차량의 4% 점유 - 전체 민원은 '15년 : 2,455건, '16년 : 4,110건, '17년 : 22,979건, '18년 : 14,997건, '19년 : 8,837건으로 증가 추세 ■ 전기차 보급 확대 및 렌터카 이용객 급증에 따라 제주 EV콜센터에 접수되는 민원 증가 <ul style="list-style-type: none"> - 전체 이용자 중 렌터카 이용자(21.91%)의 민원이 총 31,557건으로 전체민원 53,378건 중 과반수(59.12%)를 차지 ■ 제주 전기차 관련 사업 활성화에 따른 수요 급증으로 관광지 주변 충전소의 이용량 과밀 해소 및 추가 배치 등 정책 수립의 객관적 근거자료 필요 ■ 콜센터 접수 민원 분석을 통해 개방형 충전기 서비스 개선 사항을 도출하고, 단순 사용미숙 등의 반복되는 민원 사항에 대한 근본적 해소 방안 마련 필요 ■ 기존 민원 분류 카테고리의 재정비 및 추가의 필요성
사업추진	<ul style="list-style-type: none"> ■ 2015년 6월 ~ 2019년 5월(5개년) 데이터를 활용한 콜센터 민원 접수에 대해 현황 분석 실시, 주요 고장 발생지 및 원인 분석 ■ 다년간의 민원 데이터 축적에 따른 연도별 키워드 분석 및 도 구축 충전기 유료화 이후 민원 유형 변동 분석 추진 ■ 효율적 충전기 설치 및 유지관리 방안 마련
사업완료	<ul style="list-style-type: none"> ■ 콜센터 및 개방형 충전기 서비스 효율적 운영 방안 및 정책 수립 <ul style="list-style-type: none"> - 빅데이터 분석 결과 공유를 위한 담당자 회의 등 개최 - 분석결과를 활용한 카테고리 재정비, 사전 안내방법 건의 등의 운영 효율화 방안 마련 - 공용충전인프라 추가 구축을 위한 최적지 도출 및 유지관리방안 마련 - 사업 추진을 위한 예산 투입의 당위성 확보

2. 문제의식

- 제주 전기차 1만대 시대 진입에 따라 관련 민원의 효율적 처리 및 개방형 충전기 최적 설치에 대한 요구 증대
- 2018.3.2 기준 제주 전기차 10,006대 / 2019.8.16. 기준 전기차 등록대수는 17,424대로 전체 자동차 등록대수(390,258대) 대비 4.46% 차지
 - EV콜센터로 접수된 전기차 관련 민원은 '15년 : 2,455건, '16년 : 4,110건, '17년 : 22,979건, '18년 : 14,997건, '19년 : 8,837건
 - 개방형 충전기 설치대수 : 제주시(573대), 서귀포시(426대)
- 제주 EV콜센터 민원 접수 현황
 - (민원인 유형별) 렌터카 이용자(31,557건), 일반 이용자(18,294건), 일반시민(1,586건) 순으로 렌터카 이용자 접수 민원이 대다수
 - (연도별/월별)
 - 2015년 - 8월(377건) > 7월(374건) > 9월(371건)
 - 2016년 - 12월(409건) > 4월(403건) > 7월(397건)
 - 2017년 - 8월(3,245건) > 7월(2,791건) > 5월(2,693건)
 - 2018년 - 8월(1,698건) > 9월(1,595건) > 12월(1,460건)
 - 2019년 - 3월(1,952건) > 5월(1,820건) > 4월(1,814건)로봄철 나들이객이 증가하는 3~5월 그리고 여름휴가가 집중된 7~8월 사이 민원 접수량이 급증하였으며, 겨울철 민원 접수량은 상대적으로 낮음
- 제주 전기차 관련 사업 활성화에 따른 개방형 충전기 수요 증가
 - 관광지 주변 이용량 과밀 지역 분산 효율화 필요
 - 추가 설치 건의에 따른 최적 입지 선정 고려 및 유지관리 강화
- 충전기 서비스 개선 사항 도출 및 반복 민원에 대한 근본적 해소 방안 마련 필요
 - 결제 및 커넥터 조작 등 단순 사용법 문의에 대한 효율적 안내방법 모색
 - 민원 분류 카테고리 재정비 필요



과제	<ul style="list-style-type: none"> ■ 문제점 도출 및 확인 ■ 문제해결을 위해 필요한 활동(대안)을 나열 ■ 필요한 활동(대안)을 통해 얻을 수 있는 효과를 확인
추진내용	<ul style="list-style-type: none"> ■ 문제점 확인 <ul style="list-style-type: none"> • 전기차 이용자 수는 렌터카가 일반인보다 낮으나, 민원 건수는 렌터카 이용자가 다수를 차지하는 것을 확인 • 제주 전기차 1만대 시대, '17년도 이후 민원 급증에 따른 민원접수 업무량 증가 • 휴가철, 공휴일, 여행적기 등의 일자 변수를 고려할 때, 특정 월에 민원이 가중되어 업무 효율성이 저하 • 결제, 단순 사용미숙 등의 반복민원에 따른 불필요한 행정력 낭비 초래 • 도 구축 충전기 유료화 이후 민원 유형 변동이 예상 • 특정 지역 과밀 해소를 위한 충전기 최적 설치 및 유지관리 방안의 필요성 ■ 문제해결을 위해 필요한 활동 나열 <ul style="list-style-type: none"> • 행정 구역별/사업자별 개방형 충전기 설치 대수 빈도 분석 • 민원인 유형별/월별 민원 발생 건수 현황 분석 • 주요 고장 발생지 및 원인 분석 • 일별 민원 데이터 시계열 분석 및 예측 • 도 구축 충전기 유료화 시행에 따른 민원 유형 변동 분석 • 연도별 민원 키워드 분석 ■ 필요한 활동을 통해 얻을 수 있는 효과를 확인 <ul style="list-style-type: none"> • 민원 유형, 가중 시간대 파악으로 불필요한 행정처리 최소화 및 업무 효율성 증진 • 지도 매핑으로 읍면동별 민원 접수 건수와 고장 발생지 및 원인 파악 • 반복 및 단순 민원 조기 식별로 인한 민원 대응력 향상 • 유료화 시행에 따른 신규 민원 유형 대비책 강구 • 키워드 분석을 통한 민원 분류 카테고리 재정비 • 개방형 충전기 최적 배치를 위한 데이터 기반 객관적 타당성 검증
Tip	<ul style="list-style-type: none"> ■ 주요 고장 원인 및 위치를 정확히 파악하기 위해 기존에 수집하던 민원 자료에 결측값을 최소화하는 수집 방식을 제안

3. 사업추진

- 자체분석 추진 시 빅데이터 전문지식 및 분석 경험 부족으로 인한 다양한 변수를 미적용한 단순 분석에 그침
 - 한국지역정보개발원에 빅데이터 분석 협업 요청
- 민원 키워드 분석모델의 경우 향후 콜센터를 운영하고 있는 지자체로의 확산을 고려, 한국지역정보개발원과 빅데이터 분석 협업 추진
 - 한국지역정보개발원과의 협업으로 민원 처리에 전문 분석 기법을 대입하여 자체 분석에서의 부족한 점을 보완하고, 경험 부족에 대한 부담을 경감
 - 프로젝트 추진에 따른 한국지역정보개발원의 전문가 분석 가이드를 통해 빅데이터 분석에 대한 추진 동력 확보

과제	<ul style="list-style-type: none"> ■ 자체분석 시 빅데이터 분석 전문지식 및 경험 부족에 따른 어려움 해결 필요 ■ 프로젝트 추진에 대한 동기 부여 및 빅데이터 활용 문제 해결을 위한 추진 동력 확보 방안 마련 필요 ■ 협업 기관(한국지역정보개발원)과 제주시 EV콜센터 간 효율적 협조 도출 필요
추진내용	<ul style="list-style-type: none"> ■ 디지털융합과-4713(2019.05.16) 빅데이터 연계 분석 협조 요청 ■ 디지털융합과-5437(2019.06.07) 빅데이터 연계 분석 업무에 따른 서약서 제출 요청 ■ 클라우드기획부-732(2019.06.07) 빅데이터 연계 분석 업무에 따른 보안 서약서 회신 ■ 디지털융합과-5521(2019.06.10) 빅데이터 연계 분석 업무 소요 데이터 제공 ■ 클라우드기획부-784(2019.06.14) 전기차 충전 인프라 운영·관리 및 콜센터 운영 정책 수립을 위한 협업 빅데이터 분석 추진 계획 수립 ■ (2019.06.21) 협업 빅데이터 분석 업무 협의회 개최 (장소 : 제주시청) ■ (2019.10.18) 협업 분석 결과 공유 회의 개최 (장소 : 한국지역정보개발원) ■ (2019.11.04) 최종 분석결과보고서 제출
Tip	<ul style="list-style-type: none"> ■ 민원 유형 및 키워드의 반복성을 고려한 전략적 접근 ■ 자체분석의 어려움을 각 부처의 빅데이터 교육 및 세미나 등에 적극적인 참석을 통해 마인드 함양과 빅데이터 분석에 대한 접근법 모색 가능



4. 완료보고

- 정책 결정자에 대해 분석결과를 보고하는 한편 업무담당자와 공유하여 분석결과물 활용에 대한 공감대 형성
- 직원들의 관심도와 공감대 형성을 위해 분석결과를 내부시스템 자료실에 게시하고, 담당자 교체 시에도 지속 활용 가능하도록 지원
- 행정안전부 국가정보자원관리원 해안(빅데이터 공통기반 시스템)에 분석사례 결과 공유·전파를 통한 타 행정기관들의 활용도 제고

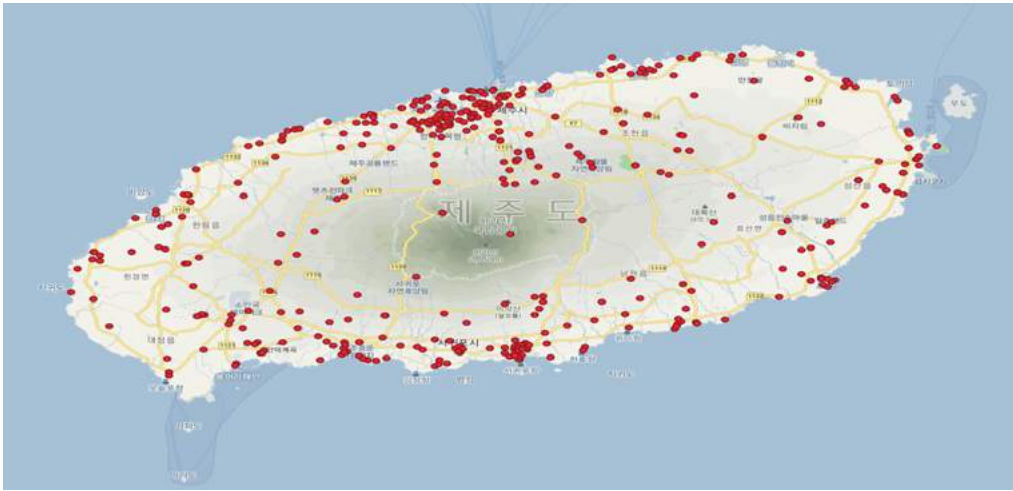
과제	<ul style="list-style-type: none"> ■ 정책결정자에 대한 분석 결과 보고를 통한 원활한 정책 수립 결정 지원 ■ 업무관련 담당자들의 분석 결과물 활용에 대한 공감대 형성 방안 모색 ■ 분석 결과물을 정책 활용 주제에 맞게 적용할 수 있는 업무 중심의 활용 방안 제시 및 정책 적용 ■ 분석결과 사례 공유 및 공동 활용 방안 모색
추진내용	<ul style="list-style-type: none"> ■ 정책결정권자에 대한 분석 결과 보고서 작성 및 결과 보고 추진 ■ 빅데이터 분석 결과의 정책 활용을 위한 관련 부서와의 유기적인 협조 체계 마련 <ul style="list-style-type: none"> • 분석 결과 자료 공유 및 관련 부서 활용 방안 제시 • 내부 종합포털사이트 자료실에 해당 결과물 게시로 빅데이터에 대한 직원 관심도와 공감대 형성에 기여 • 분석 결과 공유 이후 부족한 사항 도출 및 개선점 논의 ■ (해안) 분석 결과 공유 및 사례 전파 <ul style="list-style-type: none"> • 행정안전부 국가정보자원관리원 해안(빅데이터 공통기반 시스템) 자료 공유 • 타 지자체 공무원에게 결과 공유를 통한 새로운 행정서비스 발굴 기회 부여
Tip	<ul style="list-style-type: none"> ■ 분석 결과 활용 정책 적용 부분에 있어서 관련 부서와 충분한 사전협의를 거친 후 추진

5. 분석결과

제주 전기차 개방형 충전기 설치현황 분석

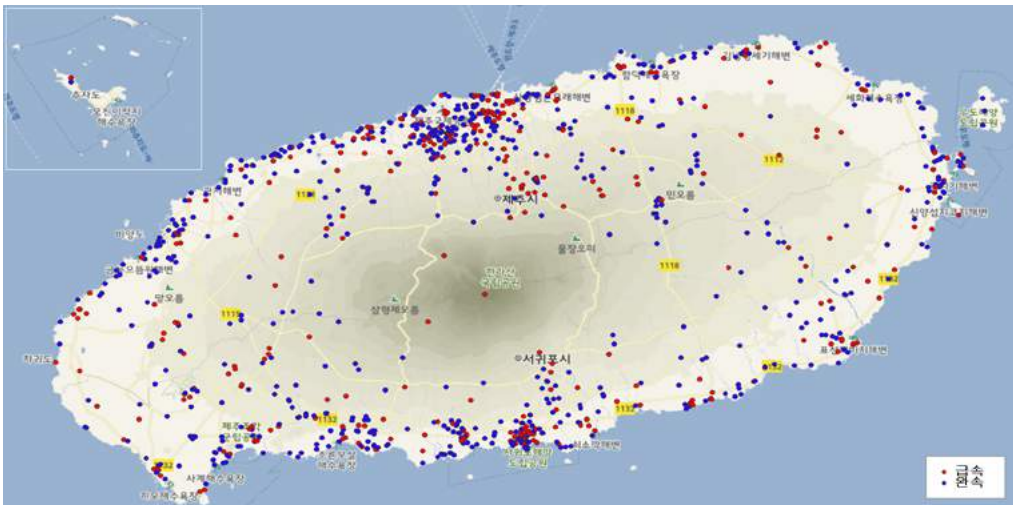
■ 제주도 전기차 충전소 분포도 (전체 : 2,158기)

- 내륙보다는 주로 해안가 도로 및 제주공항과 청사, 관광지 인근을 따라 분포



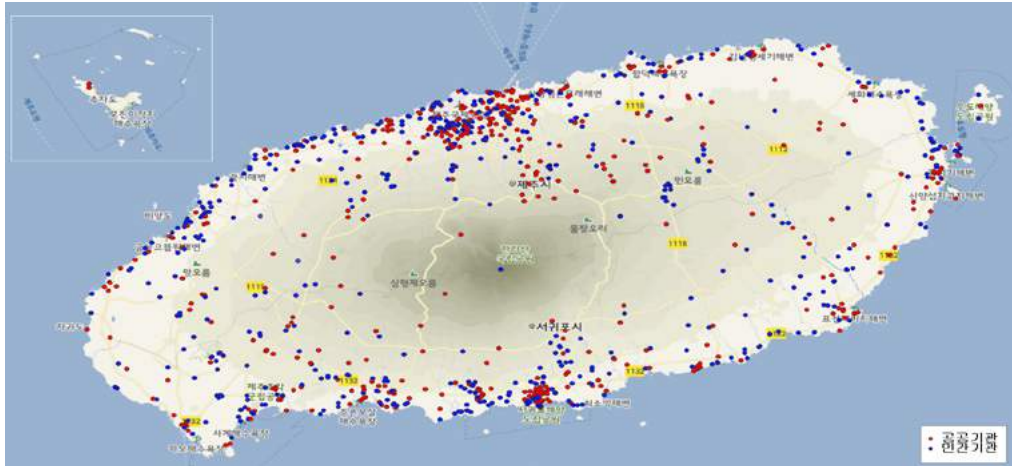
■ 제주도 전기차 충전소 분포도 - 충전방식별(급속/완속)

- 전기차 충전기 총 2,158대 중 급속 충전기는 582기, 완속 충전기 1,576기 설치
- 급속, 완속 충전기 모두 구비한 충전소는 급속 충전소로 분류



■ 제주도 전기차 충전소 분포도 - 사업자별 [공공(689기), 민간(1,469기)]

- 공공기관에 비해 민간기관이 충전 인프라 다수 보유



(단위: 기)

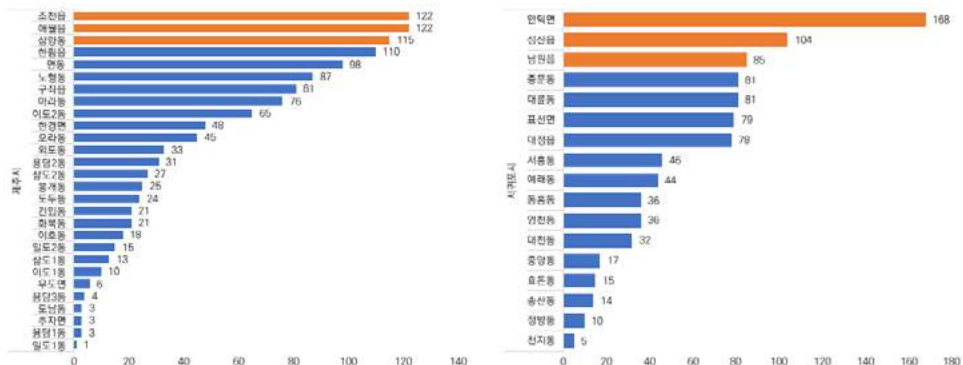
구 분		공공기관	민간기관
기관 수		4	13
충전기 종류	급속	382	148
	완속	255	1,321
	기타	52	-

※ 공공기관: 도청, 환경부, 한국전력 등 4개, 민간기관: KT, 포스코, 지엔텔, 클린일렉스 등 13개

※ 충전기(기타): 도청 교통약자용 충전기, 급속으로 분류

■ 행정구역별 개방형 충전기 설치 대수 - 총 2,158기 [급속(582기), 완속(1,576기)]

- (제주시) 조천읍(122기), 애월읍(122기), 삼양동(115기)
- (서귀포시) 안덕면(168기), 성산읍(104기), 남원읍(85기)



EV콜센터 민원 접수 현황 기초 통계분석

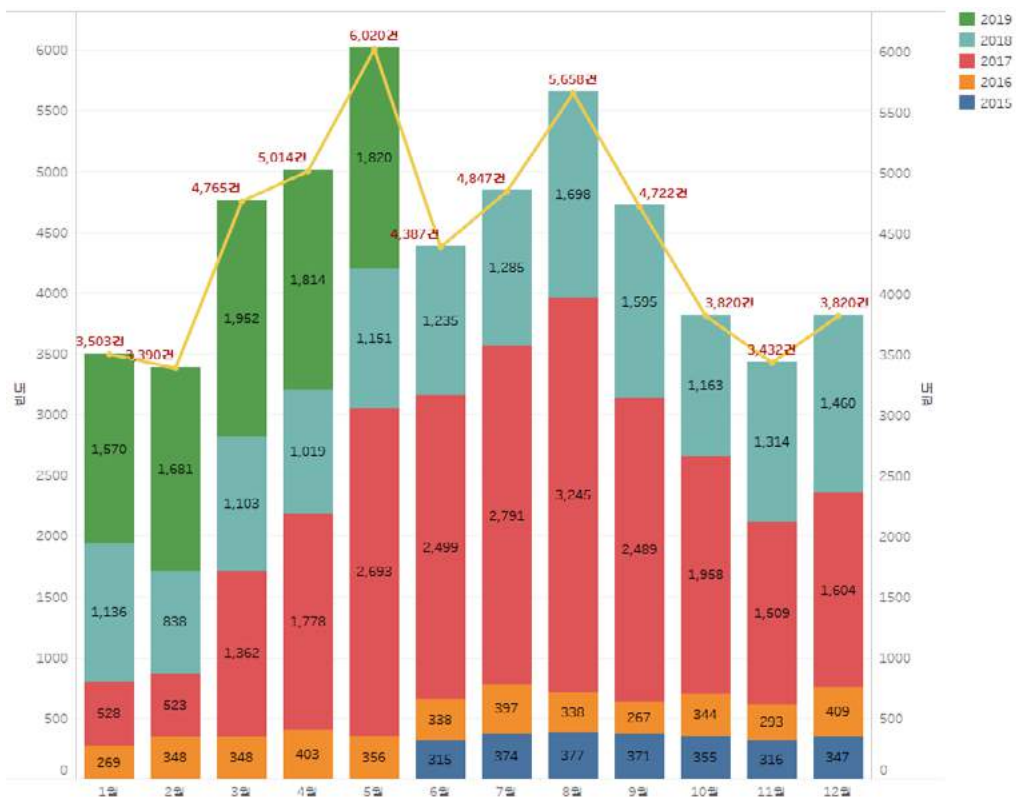
연도별 민원 접수 현황

(단위 : 건)

구분	2015년 9~12월	2016년	2017년	2018년	2019년 1~5월	합계
민원 수	2,455	4,110	22,979	14,997	8,837	53,378

- 최근 5년간 민원 수는 2017년 가장 높았으며, 향후 연간 민원 건수는 전년대비 증가 추세를 지속할 것으로 판단됨

연도별 · 월별 민원 접수 현황



- 민원 발생 건수는 봄 > 여름 > 가을 > 겨울 순으로 집계됨
- 향후 '19년도 하반기 자료를 고려한다면 5월 이후 민원 접수량이 소폭 증가할 것으로 예상됨



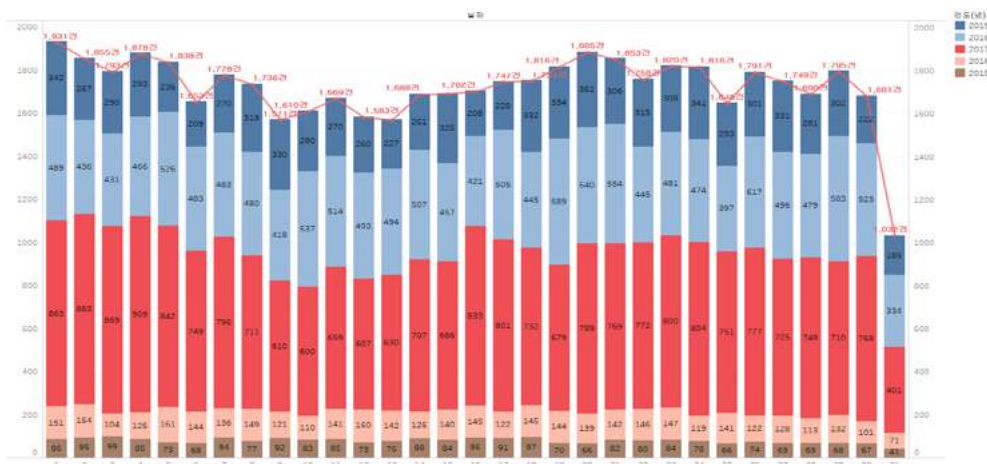
연도별·월별 민원 접수 빈도

(단위: 건)

구 분	2015년	2016년	2017년	2018년	2019년	월평균	합계
1월	-	269	528	1,136	1,570	875.8	3,503
2월	-	348	523	838	1,681	847.5	3,390
3월	-	348	1,362	1,103	1,952	1,191.3	4,765
4월	-	403	1,778	1,019	1,814	1,253.5	5,014
5월	-	356	2,693	1,151	1,820	1,505.0	6,020
6월	315	338	2,499	1,235	-	1,096.8	4,387
7월	374	397	2,791	1,285	-	1,211.8	4,847
8월	377	338	3,245	1,698	-	1,414.5	5,658
9월	371	267	2,489	1,595	-	1,180.5	4,722
10월	355	344	1,958	1,163	-	955.0	3,820
11월	316	293	1,509	1,314	-	858.0	3,432
12월	347	409	1,604	1,460	-	955.0	3,820

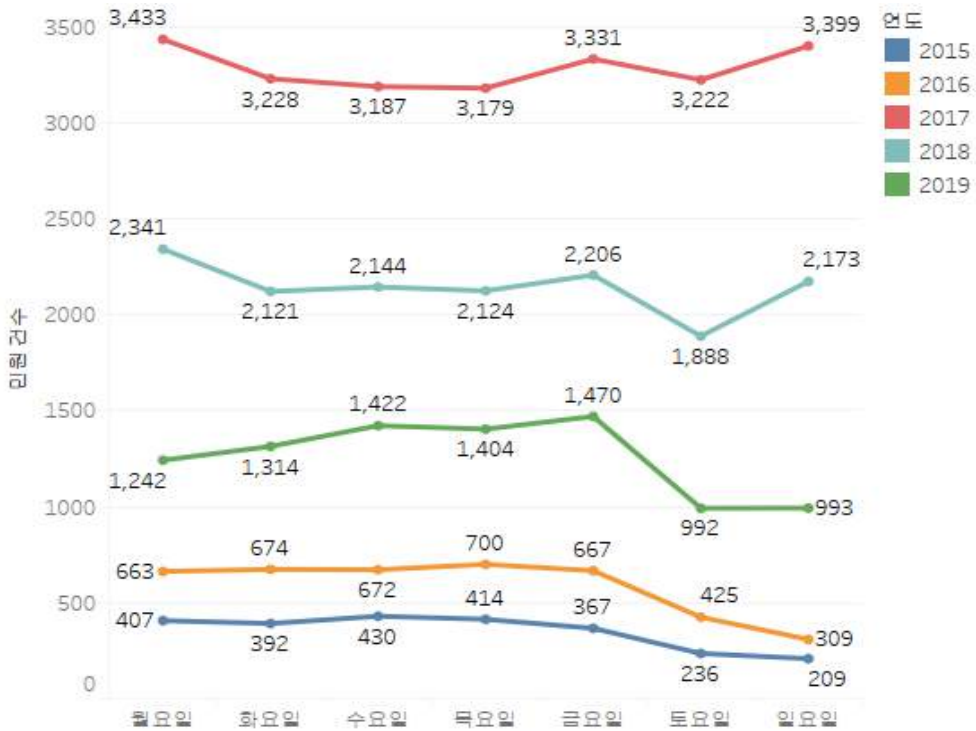
- 월별 평균 민원 접수 건수를 살펴보면 봄철 날들이객이 증가하는 4월, 5월 그리고 여름휴가가 집중된 7월, 8월 민원이 급증하는 것으로 나타남
- 민원 접수 상위 월(4~5월, 7~8월)의 접수 건수는 21,539건으로 전체 접수 민원(53,378건)의 40.4%를 차지

연도별·일자별 민원 접수 현황



- 매월 초(1~5일), 중순 말(19~21일)등이 특히 높으며 대체적으로 고르게 나타남

■ 전체 연도 요일별 민원 수 통계

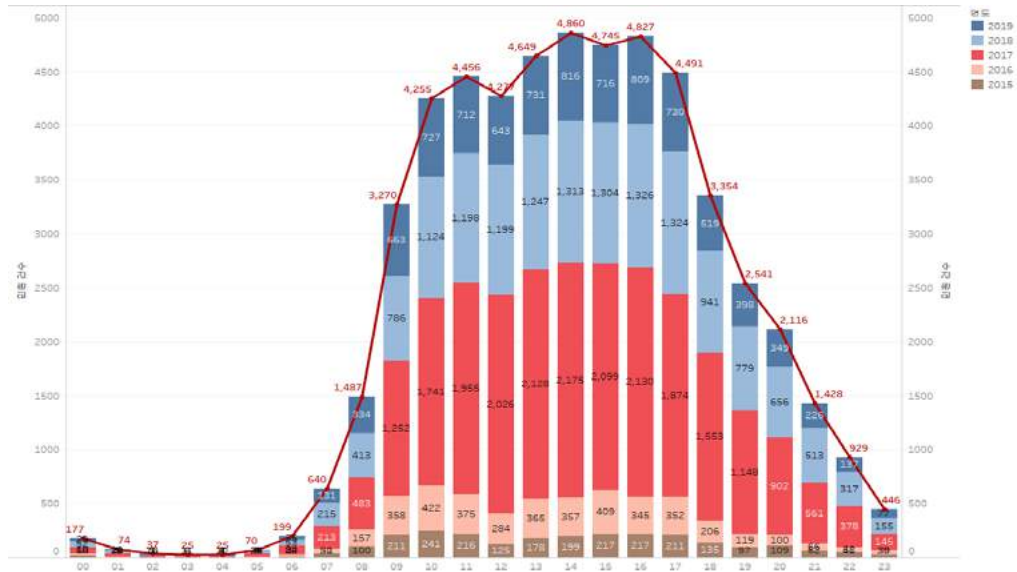


- 월요일(8,086건) > 금요일(8,041건) > 수요일(7,855건) > 목요일(7,821건) > 화요일(7,729건) > 일요일(7,083건) > 토요일(6,763건) 순으로 민원 발생
- 주말대비 주중 민원 접수 건수가 더 높은 것으로 나타남

	월요일	화요일	수요일	목요일	금요일	토요일	일요일	합계
2015년	407	392	430	414	367	236	209	2,455
2016년	663	674	672	700	667	425	309	4,110
2017년	3,433	3,228	3,187	3,179	3,331	3,222	3,399	22,979
2018년	2,341	2,121	2,144	2,124	2,206	1,888	2,173	14,997
2019년	1,242	1,314	1,422	1,404	1,470	992	993	8,837
합계	8,086	7,729	7,855	7,821	8,041	6,763	7,083	53,378



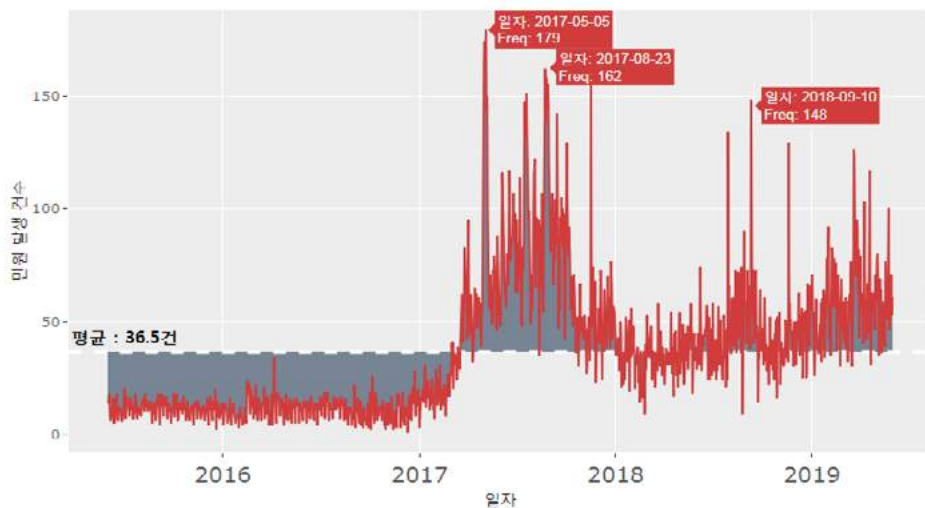
연도별/시간대별 민원 접수 현황



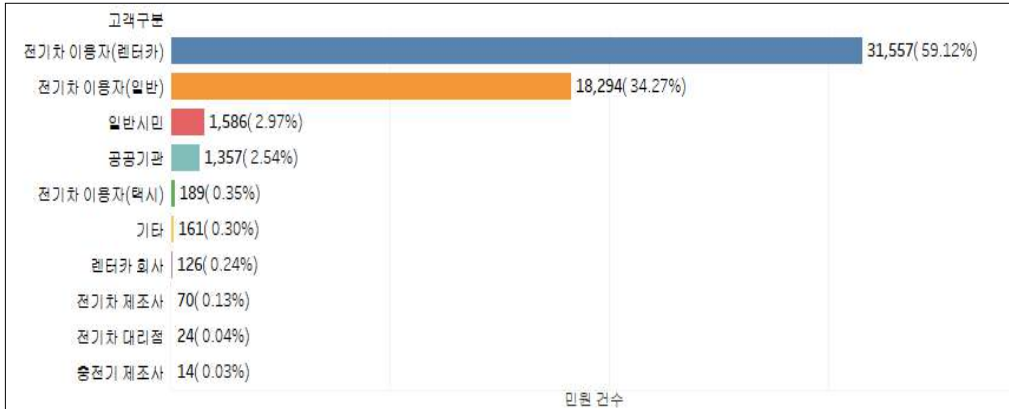
- 14시가 가장 높으며, 10~17시 사이에 민원 접수 건수가 집중되어 있음

특정 일자 접수 건수 통계

- 민원 평균 접수 건수는 36.5건, '17년 3월 이후 평균 상한
- 어린이날, 추석 등 매년 공휴일 및 여름휴가철 전후에 높은 수치를 기록함



■ 민원인 유형별 통계



구분	빈도(건)	비율(%)
전기차 이용자(렌터카)	31,557	59.12
전기차 이용자(일반)	18,294	34.27
일반시민	1,586	2.97
공공기관	1,357	2.54
전기차 이용자(택시)	189	0.35
기타	161	0.30
렌터카 회사	126	0.24
전기차 제조사	70	0.13
전기차 대리점	24	0.04
충전기 제조사	14	0.03
합계	53,378	100.00

■ 연도별 민원인 유형 구분체계가 상이하여 최근 기준(2019년)으로 재분류 함

※ 공공기관 : 지자체 관리자, 공공기관 등

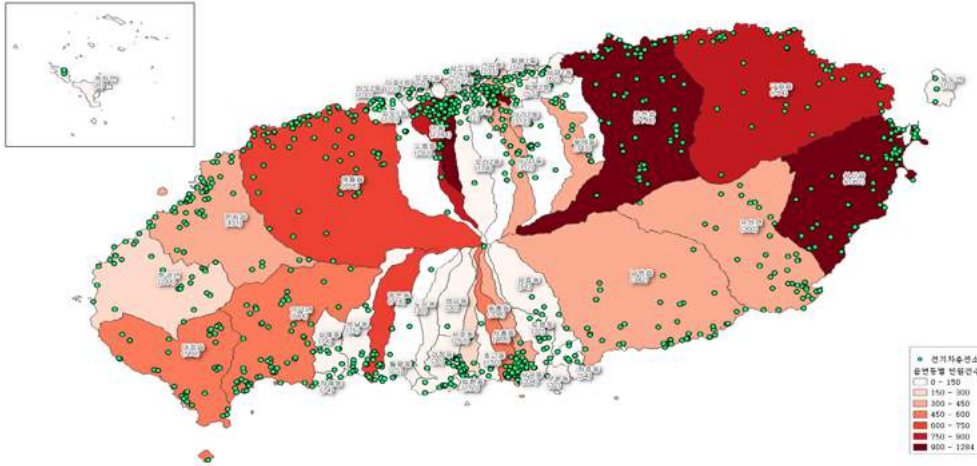
※ 일반시민 : 일반인, 아파트 관계자, 일반시민 등

- 전기차 이용자(렌터카) > 전기차 이용자(일반) > 일반시민 순으로 민원 건수 추이를 보임
- 관광객 급증에 따라 일반 이용자에 비해 렌터카 이용자의 민원 접수 건수가 높은 것으로 판단



주요 고장 발생지 및 원인 분석

읍면동별 민원 접수 건수 및 충전소 지도

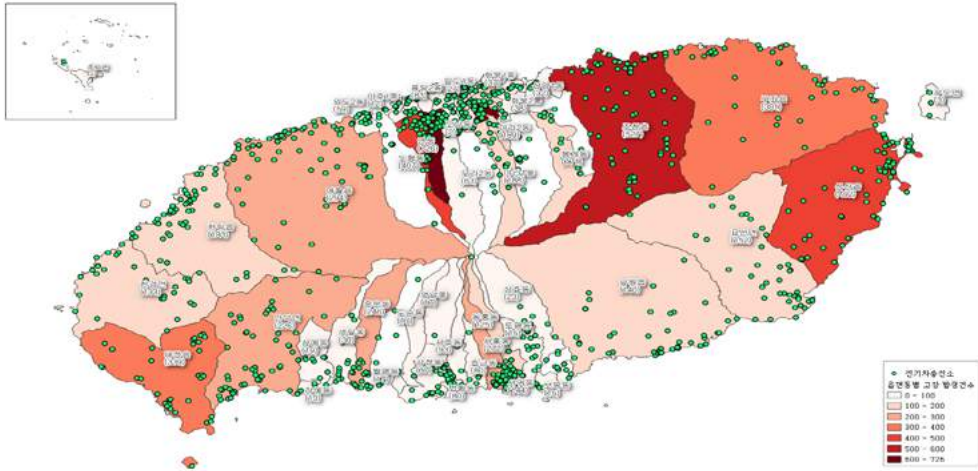


(단위 : 건)

읍면동	민원 접수 건수	읍면동	민원 접수 건수	읍면동	민원 접수 건수
건입동	233	서홍동	468	이도2동	1,117
구좌읍	784	성산읍	1,062	이호동	47
남원읍	342	송산동	97	일도1동	37
노형동	797	아라동	624	일도2동	317
대륜동	553	안덕면	584	정방동	22
대정읍	569	애월읍	658	조천읍	1,148
대천동	115	연동	1,284	중문동	716
도남동	2	영천동	73	중앙동	109
도두동	137	예래동	223	천지동	170
동홍동	338	오라동	247	추자면	6
봉개동	347	외도동	200	표선면	360
삼도1동	26	용담1동	14	한경면	244
삼도2동	226	용담2동	170	한림읍	431
삼양2동	2	우도면	10	화북동	113
삼양동	125	이도1동	53	효돈동	108
합계 (주소 표기)		15,308	합계 (주소 누락)		38,070
전체 합계		53,378			

- 주소 누락 38,070건을 제외한 15,308건을 대상으로 주요 민원 접수 건수 조사
- 연동, 이도2동, 조천읍, 성산읍 등 관광지과 공항 및 청사 인근에 민원 발생 건수 다

■ 읍면동별 고장 발생 건수 및 충전소 지도



(단위 : 건)

읍면동	고장 발생 건수	읍면동	고장 발생 건수	읍면동	고장 발생 건수
건입동	153	서흥동	272	이도2동	662
구좌읍	371	성산읍	492	이호동	25
남원읍	140	송산동	45	일도1동	27
노형동	407	아라동	377	일도2동	147
대륜동	253	안덕면	226	정방동	14
대정읍	339	애월읍	294	조천읍	527
대천동	68	연동	726	중문동	249
도남동	2	영천동	38	중앙동	34
도두동	62	예래동	76	천지동	81
동홍동	125	오라동	135	추자면	3
봉개동	171	외도동	118	표선면	152
삼도1동	17	용담1동	6	한경면	133
삼도2동	74	용담2동	83	한림읍	193
삼양2동	2	우도면	3	화북동	75
삼양동	65	이도1동	30	효돈동	46
고장 민원 (합계)		7,548	기타 민원 (합계)		7,760
전체 합계		15,308			

- 주소가 명시된 15,308건 중 단순 사용자 미숙을 제외한 7,548건을 대상으로 조사
- 전체 민원 53,378건 중 7,548건(14.14%)이 고장 발생 민원에 해당
- 제2청사, 제주시청 등이 위치한 연동, 이도2동 / 관광지 인근에 위치한 조천읍, 성산읍이 가장 다

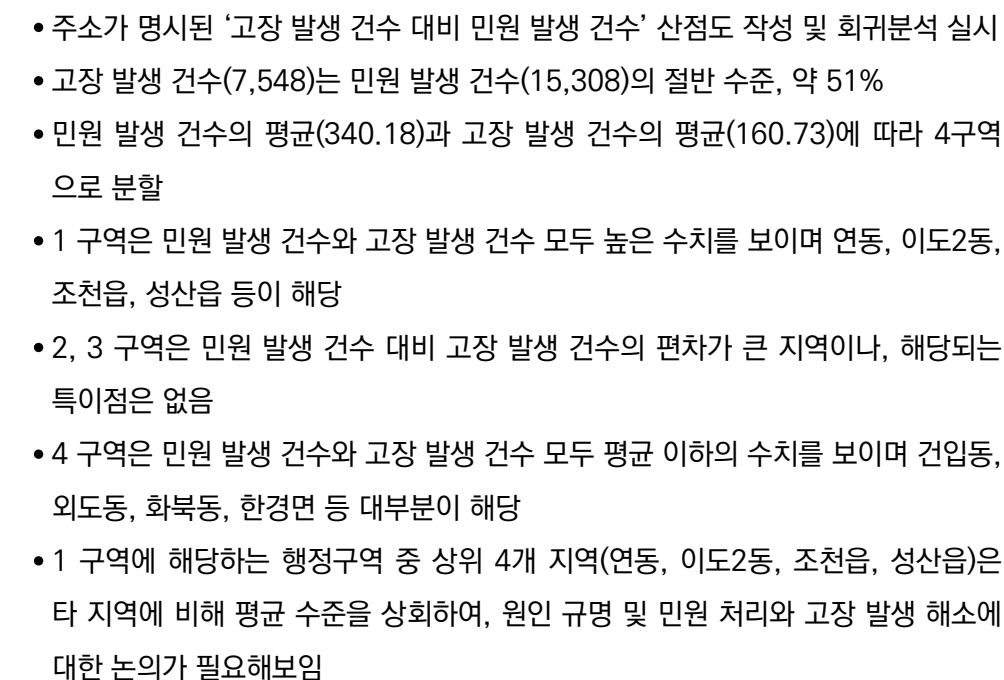
■ 읍면동별 충전기 1대당 평균 민원 및 고장 발생 감당 건수

(단위 : %)

읍면동	충전기 1대 당 민원 건수	충전기 1대 당 고장 건수	읍면동	충전기 1대 당 민원 건수	충전기 1대 당 고장 건수
건입동	11.10	7.29	예래동	5.07	1.73
구좌읍	9.68	4.58	오라동	5.49	3.00
남원읍	4.02	1.65	외도동	6.06	3.58
노형동	9.16	4.68	용담1동	4.67	2.00
대륜동	6.83	3.12	용담2동	5.48	2.68
대정읍	7.29	4.35	우도면	1.67	0.50
대천동	3.59	2.13	이도1동	5.30	3.00
도남동	0.67	0.67	이도2동	17.18	10.18
도두동	5.71	2.58	이호동	2.61	1.39
동홍동	9.39	3.47	일도1동	37.00	27.00
봉개동	13.88	6.84	일도2동	21.13	9.80
삼도1동	2.00	1.31	정방동	2.20	1.40
삼도2동	8.37	2.74	조천읍	9.41	4.32
삼양2동	0.50	0.50	중문동	8.84	3.07
삼양동	1.09	0.57	중앙동	6.41	2.00
서홍동	10.17	5.91	천지동	34.00	16.20
성산읍	10.21	4.73	추자면	2.00	1.00
송산동	6.93	3.21	표선면	4.56	1.92
아라동	8.21	4.96	한경면	5.08	2.77
안덕면	3.48	1.35	한림읍	3.92	1.75
애월읍	5.39	2.41	화북동	5.38	3.57
연동	13.10	7.41	효돈동	7.20	3.07
영천동	2.03	1.06			

- 충전기 설치 대수 대비 민원 및 고장 발생이 많이 접수되는 지역은 우선 관리 대상으로 분류
- 충전기 설치 대수가 많은 지역은 대체로 민원 및 고장 발생 접수도 많으나, 예외 지역도 있음
- 천지동, 일도2동, 이도2동, 봉개동, 연동 등은 충전기 설치 대수 대비 민원 발생량이 많음
- 천지동, 이도2동, 일도2동, 연동, 건입동 등은 충전기 설치 대수 대비 고장 발생량이 많음

$$\text{민원건수} = 0.51 \cdot \text{고장건수} - 5.61$$

$$(R^2=0.8917)$$


■ 주요 고장 발생 충전소

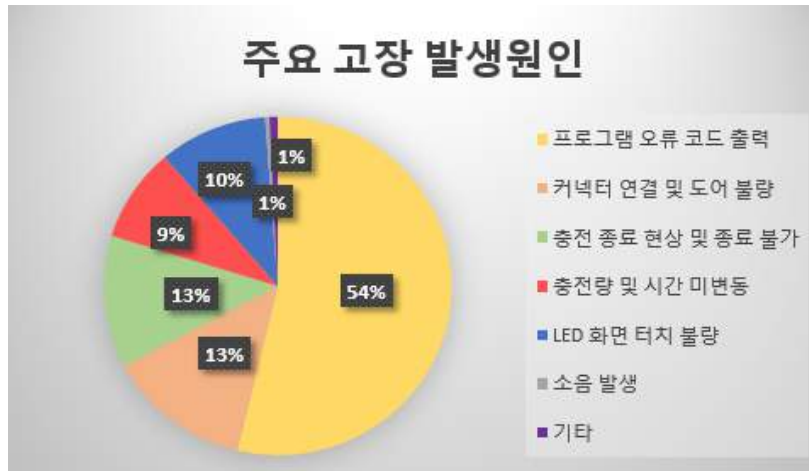


(단위 : 건, %)

충전소명	읍면동	고장 발생 건수	읍면동별 전체 고장 발생 건수	비율(%)
제주도청 1청사	연동	350	726	48.21
제주시청	이도2동	290	662	43.81
성산읍사무소	성산읍	179	492	36.38
서귀포시청 1청사	서홍동	143	272	52.57
아라동주민센터	아라동	138	377	36.6
성산일출봉	성산읍	131	492	26.63
조천읍사무소	조천읍	107	527	20.3
노형동주민센터	노형동	97	407	23.83
애월읍사무소	애월읍	90	294	30.61
안덕면사무소	안덕면	89	226	39.38
합계		1,614	4,475	36.07

- 수급 자료 중, 구체적인 충전소 위치가 명시된 주요 고장 발생 상위 10개 지역 추출
- 면/동사무소, 시/도청 등 관공서와 성산일출봉 등의 관광지 인근 충전소에서 주로 고장 발생 (‘제주도청 1청사’〉‘제주시청’〉‘성산읍사무소’ 등)
- 해당 읍면동의 전체 고장 발생 건수 대비 충전소별 고장 발생 건수의 비율은 ‘서귀포시청 1청사’〉‘제주도청 1청사’〉‘제주시청’ 등의 순으로, 해당 읍면동에서 우선 관리 지역으로 분류
- 전체 고장 발생 건수 7,548건 중 상위 10개의 충전소에서 1,614건(약 21.4%)이 발생함에 따라, 해당 지역을 우선적으로 관리한다면 20%가량의 고장 발생 민원 경감 효과를 기대할 수 있음
- 전체 민원 발생 건수 53,378건 중 상위 10개의 충전소에서 1,614건(약 3.0%)이 발생함에 따라, 해당 지역을 우선적으로 관리한다면 3%가량의 전체 민원 경감 효과를 기대할 수 있음

■ 주요 고장 발생원인



(단위: 건)

주요 고장 발생원인	건수
프로그램 오류 코드 출력	4,058
커넥터 연결 및 도어 불량	1,014
충전 종료 현상 및 종료 불가	952
LED 화면 터치 불량	761
충전량 및 충전 시간 미변동	675
소음 발생	30
기타	58
합계	7,548

(단위: 건)

오류코드	건수	오류코드	건수
42	133	50	11
30	66	22	11
41	59	60	10
26	56	72	9
14	47	54	9
109	29	24	9
21	26	52	7
49	25	47	7
151	24	109	7
34	22	45	5
16	18	110	5
6	18	19	5
37	17	76	4
32	17	2	4
64	13	61	3
55	13	40	3
36	12	기타	31
4	12	합계	747

- 수급 자료 중, 구체적인 충전소 위치가 명시된 7,548건을 대상
- 인증 및 충전 과정의 프로그램 오류 코드 출력, 화면 오류 표시가 4,058건으로 주요 고장 발생 원인 1순위
- 문의 내용 중, 구체적인 오류코드가 명시된 747건의 주요 발생 코드는 42, 30, 41, 26, 14번
- 기타 고장 원인으로는 사용 미숙, 연결 시간 초과 등의 단순 고장으로 분류
- 추가적인 자료 수급이 이루어진다면 주요 고장 발생원인의 더 정확한 식별이 가능할 것으로 기대



도 구축 충전기 유료화에 따른 민원 변동 분석

■ 민원 기술통계량 변동 확인

- 2019년 5월 이후 도 구축 충전기 유료화에 따른 전후 자료 비교 검토 시행
- 유료화 전(47개월 분 자료), 유료화 후(1개월 분 자료)

(단위 : 건)

구분	민원 건수 총계	민원 건수 월 평균
유료화 전	51,558	1,096
유료화 후	1,820	1,820

(단위 : 건)

민원인 유형	구분	
	유료화 전	유료화 후
전기차 이용자(일반인)	17,407 (33.76%)	887 (48.74%) ↑
전기차 이용자(렌터카)	30,774 (59.69%)	783 (43.02%) ↓
전기차 이용자(택시)	178	11
전기차관련사업자	187	10
지자체관리자	1,252	105
일반인	1,545	16
기타	215	8
총계	51,558	1,820

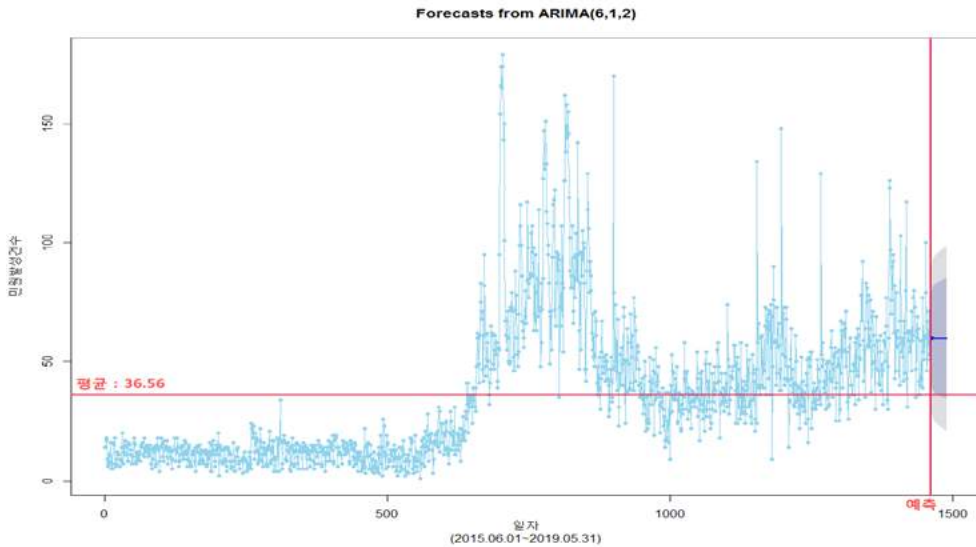
- 비율 수치를 고려할 때, 유료화 전은 전기차 이용자(렌터카) 민원인이 다수였으나, 유료화 후 전기차 이용자(일반인) 민원인이 다수
- 다른 민원인 유형은 대체로 소폭의 변동이 있으나, 유의한 차이는 없다고 판단됨

(단위 : 건)

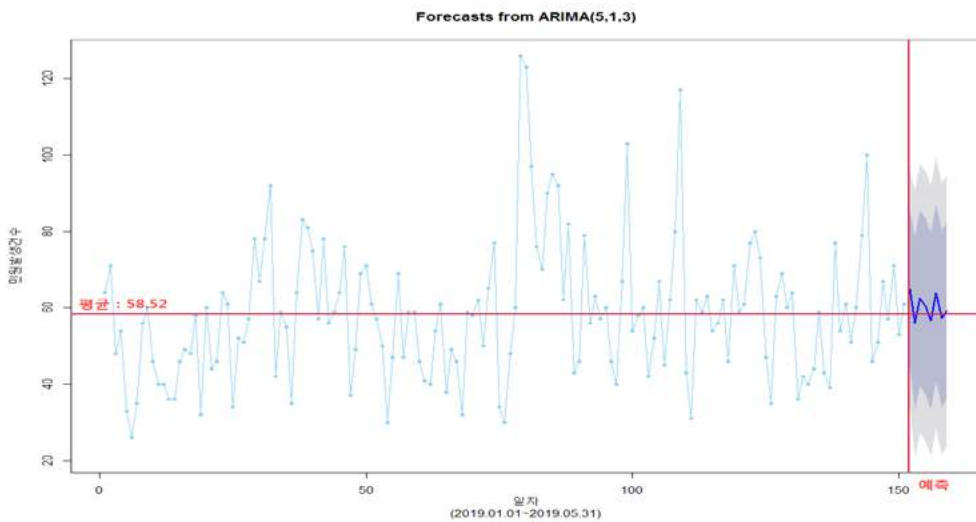
주요 민원 내용	구분	
	유료화 전	유료화 후
공용충전인프라	38,351 (74.38%)	1,582 (86.92%) ↑
일반	3,101	-
사용법	1,060	-
전기차	3,335	145
민간보급충전기	1,010	-
앱	371	7
전기차 및 충전기 민간보급	743	51
충전불가	2,932	-
기타	655	35
총계	51,558	1,820

- 유료화 전, 후 모두 공용충전인프라 관련 민원이 다
- 비율 수치를 고려할 때, 다른 민원 내용은 대체로 유의한 차이가 없다고 판단됨

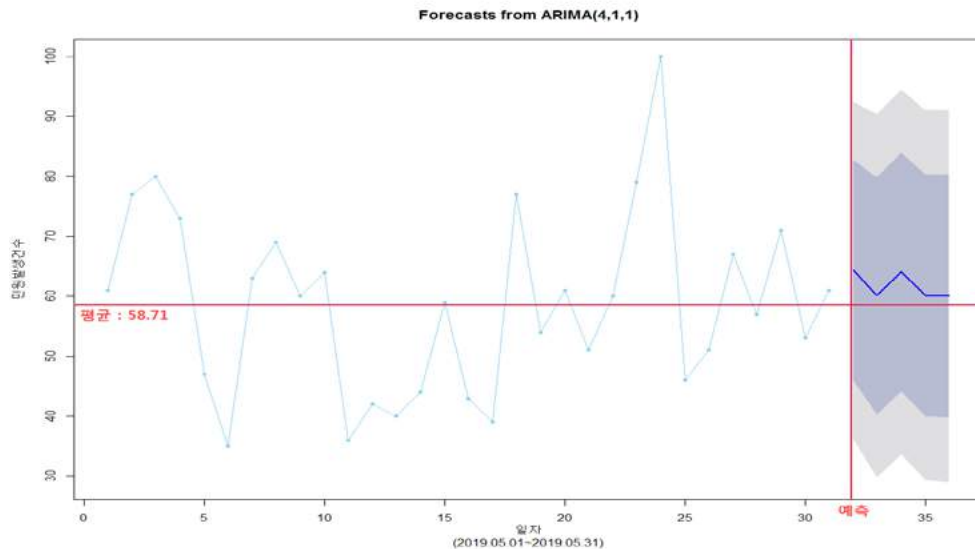
■ 민원 접수 건수 변동 분석



- '15~'19년도(1,460일) 기준 도 구축 충전기 유료화 이후 민원 발생 건수 예측
- 예측 구간 평균은 59.79로 전체 평균 36.56보다 높은 수치
- 향후 예측 30일 추계 및 신뢰구간(CI)



- '19년도(151일) 기준 도 구축 충전기 유료화 이후 민원 발생 건수 예측
- 예측 구간 평균은 60.08로 19년도 평균 58.52보다 조금 높은 수치
- 향후 예측 8일 추계 및 신뢰구간(CI)



- '19년도 5월(31일) 기준 도 구축 충전기 유료화 이후 민원 발생 건수 예측
- 예측 구간 평균은 61.75로 19년도 5월 평균 58.71보다 조금 높은 수치
- 향후 예측 5일 추계 및 신뢰구간(CI)

(단위 : 건)

활용 자료 기준	예측 개수	예측 평균	신뢰 하한	신뢰 상한
'15~19년도(1,460일)	30일	59.79	24.13	95.47
'19년도(151일)	8일	60.08	25.30	94.87
'19년도 5월(31일)	5일	61.75	31.62	91.87

131



※ 신뢰 수준 95%, CI는 개별 예측 자료의 평균값을 산출

※ 예측 자료 개수는 활용 자료 개수를 기준으로 각각의 시계열 모델 실효성에 위배되지 않는 최적 개수를 선정하였음

- 3가지 활용 자료 모두 원 자료 평균보다 예측 자료 평균이 높은 것을 확인, 유료화 이후 평균 이상 수준의 민원 발생 건수를 유지할 것으로 예측
- 활용 자료 기준을 구체화할수록 예측 평균 민원 건수는 소폭 증가, 신뢰구간 범위는 감소
- 유료화 이후 민원 접수 자료의 추가 수급('19.06~)이 이루어진다면 더욱 예측도 높은 시계열 모델을 제안할 수 있음

■ 민원 유형 변동 분석(상위 20개 키워드 추출)

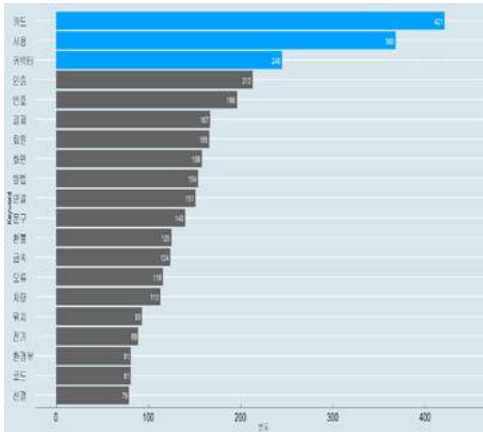
(단위 : 건)

순위	유료화 이전				유료화 이후	
	15~19년도 총괄		19년도		19년도(5월)	
	키워드	빈도	키워드	빈도	키워드	빈도
1	충전	48,500	충전	7,123	충전	1,352
2	위치	8,358	카드	1872	카드	421
3	카드	7,490	사용	1696	사용	368
4	사용	7,087	커넥터	1266	커넥터	245
5	급속	5,473	급속	841	인증	213
6	커넥터	4,828	화면	805	번호	196
7	방법	3,584	인증	761	결제	167
8	차량	3,304	방법	732	회원	166
9	인증	3,229	결제	702	화면	158
10	오류	2,698	번호	663	방법	154
11	이용	2,572	차량	628	연결	151
12	연결	2,449	위치	625	문구	140
13	화면	2,411	연결	609	환불	125
14	완속	2,231	문구	515	급속	124
15	설치	1,929	오류	487	오류	116
16	코드	1,911	회원	439	차량	113
17	번호	1,826	환경부	438	위치	93
18	서귀포	1,702	환불	395	전기	89
19	결제	1,544	이용	351	환경부	81
20	종료	1,427	전기	345	코드	81

- 유료화 이후 두드러지는 민원 변동 특이 사항은 없으나 회원(167건), 환불(125건) 등의 결제 관련 문의 소폭 증가
- 유료화 이후 민원 자료 추가 확보 시, 키워드 빈도 변화 추이가 두드러질 것으로 예상



■ ('19년도 5월 상위 빈도 키워드 Top 20 & 워드클라우드)



※ 공통적 1위인 '충전' 키워드는 제외하고 시행

- 주요 키워드 : 카드(421), 사용(368), 커넥터(245) 등
- 결제 관련 신규 키워드 식별



연도별 반복민원 키워드 분석

■ 연도별 상위 빈도 키워드 추출(unigram 모델)

(단위 : 건)

순위	2015년		2016년		2017년		2018년		2019년		전체기간	
	키워드	빈도	키워드	빈도	키워드	빈도	키워드	빈도	키워드	빈도	키워드	빈도
1	충전기	1,942	충전기	3,075	충전	10,590	충전기	6,045	충전	3,317	충전기	21,914
2	충전	813	충전	1473	충전기	7,808	충전	4,926	충전기	3,044	충전	21,119
3	카드	630	카드	988	근처	6,461	커넥터	1,980	카드	1,847	근처	9,248
4	사용	432	사용	831	충전소	3,909	사용	1,861	사용	1,687	카드	7,409
5	설치	430	급속	605	급속	2,310	카드	1,854	커넥터	1,237	사용	7,042
6	급속	277	전기	422	사용	2,231	근처	1,721	가능	875	충전소	6,110
7	차량	213	완속	387	카드	2,090	급속	1,429	급속	823	급속	5,444
8	커넥터	199	설치	378	가능	1,729	충전소	1,392	화면	793	커넥터	4,711
9	가능	195	커넥터	323	인증	1,174	가능	1,351	근처	730	가능	4,405
10	화면	195	화면	300	부근	1,073	이용	1,230	인증	721	인증	3,137
11	연결	169	차량	284	방향	1,034	인증	1,079	결제	680	차량	3,120
12	완속	144	가능	255	오류	1,003	차량	1,034	충전소	663	전기	2,537
13	전기	143	종료	233	차량	986	오류	933	번호	646	오류	2,494
14	종료	140	근처	227	커넥터	972	연결	878	차량	603	이용	2,430
15	발급	135	신청	217	전기	897	전기	735	연결	602	연결	2,400
16	수리	120	수리	209	서귀포	781	코드	727	문구	515	화면	2,216
17	근처	109	시작	188	코드	670	해피	613	오류	461	완속	1,879
18	하려	100	환경	183	인근	658	차저	612	환경부	431	설치	1,869
19	위치	98	진행	179	위치	639	완속	505	회원	422	코드	1,685
20	언제	92	이용	176	이용	604	화면	490	발생	364	번호	1,499

- 충전 관련 카드 사용, 충전기 위치 문의, 커넥터 연결 및 분리 관련 반복 민원 多
- 급속이 완속 충전기보다 설치 대수는 적으나 민원 발생량은 多
- 오류, 화면, 문구, 인증, 결제 등 시스템 관련 민원도 多



연도별 상위 빈도 키워드 추출(Top100, unigram 모델)



〈2015년〉



〈2015년(고빈도 키워드 제거-충전기, 충전)〉

※ 최다빈출단어 ‘충전기’, ‘충전’ 키워드는 제외하고 워드클라우드 작성

- 주요키워드 : 충전기(1,942), 충전(813), 카드(630), 사용(432), 설치(430) 등
- 충전기 설치 현황 문의, 충전 카드, 급속 충전기 사용법, 충전기 설치 관련 민원 다수 접수



〈2016년〉



〈2016년(고빈도 키워드 제거-충전기, 충전)〉

- 주요키워드 : 충전기(3,075), 충전(1,473), 카드(988), 사용(831), 급속(605) 등
- 급속·완속 충전기 사용법, 커넥터 관련 민원 다수 접수



〈2017년(고빈도 키워드 제거-충전기, 충전)〉

- 주요키워드 : 충전(10,590), 충전기(7,808), 근처(6,461), 충전소(3,909), 급속(2,310) 등
- 민원 접수량 폭발적 증가, 근처 충전기 위치 문의 민원 급증



〈2018년(고빈도 키워드 제거-충전기, 충전)〉

- 주요키워드 : 충전기(6,045), 충전(4,926), 커넥터(1,980), 사용(1,861), 카드(1,854) 등
- 고빈도어 제거 후 다양한 다빈도어 출현, 커넥터 연결 관련 민원, 렌트카 민원 증가

연도별 상위 빈도 키워드 추출(bigram 모델)

(단위: 건)

순위	2015년		2016년		2017년		2018년		2019년		전체기간	
	키워드	빈도	키워드	빈도	키워드	빈도	키워드	빈도	키워드	빈도	키워드	빈도
1	충전기 설치	285	충전기 사용	512	근처 충전	2,371	충전기 사용	1,066	충전기 사용	758	충전기 사용	3,846
2	충전기 사용	281	급속 충전기	510	근처 충전소	1,845	급속 충전기	810	사용 가능	379	급속 충전기	2,959
3	급속 충전기	236	완속 충전기	334	충전기 사용	1,229	오류 코드	611	급속 충전기	338	근처 충전	2,838
4	완속 충전기	129	충전기 설치	216	급속 충전기	1,065	해피 차저	605	급속 충전	296	근처 충전소	2,413
5	카드 발급	120	충전기 카드	186	근처 충전기	1,040	충전기 이용	515	커넥터 연결	283	근처 급속	1,670
6	커넥터 연결	118	환경 공단	167	급속 충전	908	커넥터 연결	512	카드 인증	282	급속 충전	1,657
7	충전기 카드	115	충전기 화면	145	근처 급속	851	차저 카드	497	근처 급속	245	카드 인증	1,641
8	민간 보급	87	충전 시작	144	카드 인증	764	카드 인증	463	카드 번호	241	충전 가능	1,445
9	충전기 화면	86	충전기 이용	142	충전 가능	718	근처 급속	448	환경부 카드	239	완속 충전기	1,427
10	사용 가능	84	민간 보급	134	방향 충전	668	충전 가능	421	해피 차저	209	사용 가능	1,425
11	충전기 커넥터	82	충전 종료	132	오류 코드	602	근처 충전소	416	해당 충전기	204	오류 코드	1,417
12	환경 공단	80	카드 인증	108	사용 가능	518	급속 충전	405	충전 불가	202	커넥터 연결	1,334
13	공단 카드	75	충전기 충전	107	부근 충전기	490	완속 충전기	381	완속 충전기	200	근처 충전기	1,296
14	가정용 충전기	73	커넥터 연결	106	인근 충전	477	안전 교육	357	오류 코드	198	해피 차저	1,294
15	충전기 위치	72	공단 카드	105	해피 차저	462	근처 충전	354	카드 사용	187	충전기 이용	1,160
16	충전 카드	71	카드 발급	105	완속 충전기	383	사용 가능	350	발생 충전	181	차저 카드	1,027
17	충전 가능	69	충전 카드	102	동시 충전	382	이용 가능	332	충전 가능	176	충전 시작	979
18	충전기 충전	66	가정용 충전기	100	주변 충전	374	충전 시작	332	커넥터 분리	172	충전기 설치	967
19	충전 종료	65	사용 가능	94	부근 충전	372	충전기 커넥터	296	차저 카드	169	충전기 카드	762
20	충전기 이용	62	충전기 위치	94	차저 카드	346	dc콤보	256	충전 시작	160	충전 종료	703



연도별 상위 빈도 워드클라우드(Top100, bigram 모델)



〈2015년〉



〈2016년〉

- [전기차 도입기] 민원 접수량 자체는 많지 않지만, 충전기 설치 위치, 충전기 사용법 등 단순 사용법 관련 민원이 다수 접수됨
- 주요키워드(2015년) : 충전기 설치(285), 충전기 사용(281), 급속 충전기(236), 완속 충전기(129) 등
- 주요키워드(2016년) : 충전기 사용(512), 급속 충전기(510), 완속 충전기(334), 충전기 설치(216) 등



〈2017년〉



〈2018년〉

- [전기차 성장기] 민원 접수량 대폭 증가, 근처 충전소 위치 문의, 급속·완속 충전기 사용법, 카드 인증 오류, 충전 시작 후 바로 종료되는 현상 등 단순 사용법 문의 외에도 충전기 동작 중 발생하는 오류를 문의하는 민원이 많이 접수됨
- 2017년 민원 접수 최대치 갱신 후 2018년 민원 접수량은 소폭 감소 추세를 보임
- 주요키워드(2017년) : 근처 충전(2,371), 근처 충전소(1,845), 충전기 사용(1,229), 급속 충전기(1,065), 근처 충전기(1,040) 등
- 주요키워드(2018년) : 충전기 사용(1,066), 급속 충전기(810), 오류 코드(611), 해피 차저(605), 충전기 이용(515) 등



〈2019년〉

〈전체기간〉

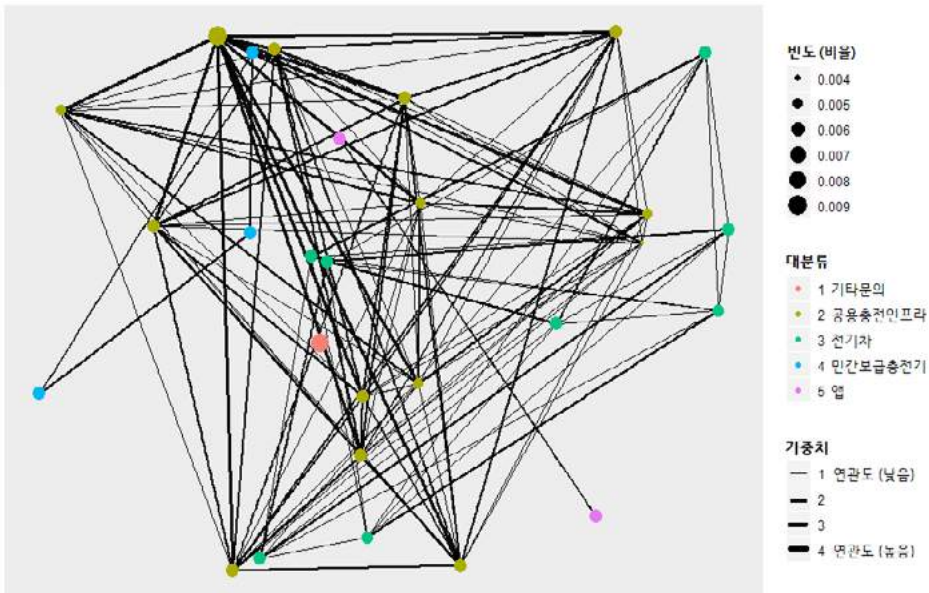
- 주요키워드(2019년) : 충전기 설치(285), 충전기 사용(281), 급속 충전기(236), 완속 충전기(129) 등
- [전기차 성숙기] 해피 차저 카드 사용 문의, 오류코드 발생 관련 문의, 환경부 카드 인증, 커넥터 연결·분리 문제 등 전기차를 이용하면서 발생하는 각종 문제점 문의
- 전기차 도입기 및 성장기에는 근처 충전소 위치 등에 대한 단순 문의가 대부분이었다면, 2018년 이후에는 다빈도 키워드의 종류도 다양화



민원 분류 카테고리 재정비

■ 분류 간 연관도 및 군집 분석

- 대분류와 중분류간의 연관성 및 빈도를 고려하여 가중치 산정
- 최적의 대분류 및 중분류 항목 개수 선정 타당성 검증



- 색 구분 군집 수를 고려하여 대분류 5항목 선정
- ‘공용충전인프라’ 대분류는 가중치가 높은 연결선을 다수 보유하여 중복 및 유사 항목 제거를 검토한 후 중분류를 선정하는 것이 타당함
- ‘전기차’ 대분류 역시 가중치가 높은 연결선을 다수 보유하여 중복 및 유사 항목 제거를 검토한 후 중분류를 선정하는 것이 타당함
- ‘기타문의’ 대분류는 특별한 연결선이 없으므로 기존 체계를 유지(1~2항목의 중분류 선정)
- ‘앱’ 대분류는 연결선 및 노드 수를 고려하여 1~3항목의 중분류 선정
- ‘전기차 및 충전기 민간보급(민간보급충전기)’ 대분류는 연결선 및 노드 수를 고려하여 2~4항목의 중분류 선정

■ 대/중분류별 피벗테이블

- 기존 수급 자료('18.01 자료 수급 양식 변경 이후)를 기준으로 빈도별 피벗테이블 작성
- (기존) 대분류 5항목, 중분류 23항목
- (개선) 대분류 5항목, 중분류 17항목

행 레이블	개수	색 구분	신설	이동	병합
공용충전인프라	19739				
공용충전기 고장	5336				
공용충전기 기타	736				○
공용충전기 사용	9459				
기타	49				○
위치 공용	4159				
기타문의	497				
건의사항	11				
기타	486				
앱	155				
기타	2				○
앱 고장	11				
앱 기타	29				○
앱 사용	113				
전기차	1947				
전기차 고장	51				○
전기차 기타	173				○
전기차 사용	371				○
전기차고장	160				○
전기차기타	645				○
전기차사용	547				○
전기차 및 충전기 민간보급	794				
기타 민간보급	115		○		○
기타 민간보급	185		○		○
민간보급충전기 고장	301		○	○	
민간보급충전기 사용	78		○	○	
전기차 구매	26			○	○
전기차구매	89			○	○
총합계	23132	-	4건	4건	7건

행 레이블	개수
공용충전인프라	19739
고장	5336
사용	9459
위치	4159
기타	785
민간보급충전기	679
고장	301
사용	78
위치	0
기타	300
앱	155
고장	11
사용	113
기타	31
전기차	2062
고장	211
사용	918
구매	115
기타	818
기타문의	497
건의사항	11
기타	486
총합계	23132

■ 카테고리 개선안 세부사항

- (대분류)
 - 기존 '전기차 및 충전기 민간보급' 항목은 카테고리 중복 여부를 고려하여 '민간보급 충전기'로 신설 및 변경
 - 중복 제거한 중분류 항목 수를 고려하여 대분류 5항목 제안
- (중분류)
 - '앱 기타', '기타(앱)' / '전기차 구매', '전기차구매' 등 동일 민원 건에 대해 통일되지 않은 항목 7건 병합
 - 중복 제거 후 중분류 17항목 편성 및 기존 민간보급충전기 관련 중분류 4건 이동
- (제언사항)
 - 복합민원에 대해서는 카테고리 선정이 어려워 결측값(기타)으로 분류, 자료 수급의 개선점 모색이 필요
 - 민간보급충전기의 위치 문의는 과거 발생 이력을 고려하여 중분류 항목으로 편성



효율적 충전기 설치 및 유지관리 방안

■ 설치 관련 민원 건수 통계량

(단위: 건)

구분	2015년	2016년	2017년	2018년	2019년	합계
전체 민원	2,455	4,110	22,979	14,997	8,837	53,378
설치 관련 민원	381	345	420	349	220	1,715
비율	15.52%	8.39%	1.83%	2.33%	2.49%	-

- '12년도 환경부 전기차 보급계획에 따른 전기차 장려정책 및 충전 인프라 구축 확대로 설치 관련 민원이 주요 요인이었으나 해를 거듭하며 감소하는 추세
- '15년도 이후 설치 요구사항보다 유지관리에 초점을 맞춘 민원 발생('17년도 이후 증가 추세)

■ 민원 주요 내용 분석

(단위: 건)

구분	주요 내용
설치	직접적인 설치 요청(관광지, 개인 사업장, 가정집, 아파트 등)
설치	설치 일정, 요금, 기준, 차량 종류, 소요 기간 문의
설치	설치 이전 문의(방법 및 양도와 절차)
유지관리	계량기 설치 문의
유지관리	여러 차량 동시충전 가능 여부 문의
유지관리	차광막 설치 및 방수 보호덮개 건의
기타	설치된 충전기의 사용자 기준 문의
기타	다양한 차종의 충전기 및 케이블 구비 요청
기타	완속 충전기의 충전 속도 개선, 급속으로 대체 여부 문의
기타	충전케이블 위치 변경 요구

- 워드클라우드 분석 결과를 바탕으로 고빈도 키워드에 대해 설치 및 유지관리 관련 문의사항 세부 내용 검토
- 설치 일정에 대한 단순 문의가 가장 많았으며, 민원 빈발 구역을 기준으로 설치 수요량 측정 제안
- 계량기, 충전 케이블 등의 부속품 편의성 개선 필요
- 바닷가 주변 지역 및 강수량을 고려한 방수 덮개 보급, 일조량에 따른 차광막 설치로 충전 시간의 효율성 증대
- 동시충전 및 다양한 차종의 충전 가능 여부 검토, 완속의 충전 속도 향상을 위한 방안 마련
- 개인 설치 및 사용 기준 관련 매뉴얼 제공에 따른 사용자 접근성 향상



데이터 분석 단계

1. 데이터 분석 단계 요약

단계	주요 내용
분석 설계	<ul style="list-style-type: none"> ■ 데이터를 분석하기 위한 계획을 세우는 것으로 분석모델을 결정 ■ 분석목적과 활용목적을 명확히 하고, 그에 적합한 방법론을 정함
데이터 수집 및 가공	<ul style="list-style-type: none"> ■ 분석에 필요한 내부 보유데이터(raw data)수집 및 외부 데이터 보유 기관 확인 후 데이터 수집, 필요에 따라 기관 간 협업 <ul style="list-style-type: none"> • (내부) 민원 데이터, 개방형 충전기 설치 현황 데이터 - 제주 EV 콜센터를 통해 쿼리를 제공받아 5개년도 데이터 일괄 추출 • (외부) GIS 관련 데이터, 제주도 전기차 충전소 데이터 활용(공공데이터 포털) ■ 유사 카테고리별 통합, '18년 이후 변동 자료 양식 통일 <ul style="list-style-type: none"> • 카테고리 특수문자 구분 삭제, 분석에 필요한 칼럼 추출 ■ 결측값(missing value) 및 이상치(outlier) 처리, 데이터 가공 등 <ul style="list-style-type: none"> • 분석이 용이한 형태로 데이터 전처리, 최종 분석용 데이터 셋 생성
데이터 분석 및 검증	<ul style="list-style-type: none"> ■ 1단계 : 개방형 충전기 설치 현황 분석 ■ 2단계 : 민원 기초 통계 및 현황 분석 ■ 3단계 : 주요 고장 발생지 및 원인 분석 ■ 4단계 : 충전기 유료화에 따른 민원 유형 변동 분석 ■ 5단계 : 연도별 반복민원 키워드 분석 ■ 6단계 : 최적 입지 선정 및 효율적 유지관리 방안 마련
데이터 시각화	<ul style="list-style-type: none"> ■ 분석결과를 바탕으로 R 활용, 빅데이터 시각화 ■ GIS 활용 최적 입지 선정, 지도 시각화



2. 분석설계

■ 분석 설계란, 결과를 도출하기 위해 필요한 전략을 세우는 일로, 분석을 계획하는 단계

- 분석 설계에서는 분석을 통해 알아내고자 하는 대상(분석의 목적, 내용, 결과를 통해 도출되는 기대효과)을 명확하게 규정
- 분석 목적에 적합한 분석모형을 결정하여야 함
- 분석설계는 관련 주제에 대한 데이터 분석모형을 정하는 것임

■ 통계적 분석 방법을 쓸 것인가를 고려

- 분석 단위 결정
 - 시간적 분석 단위 : 시간대별, 일별, 요일별, 월별, 연도별
 - 공간적 분석 단위 : 행정구역별
- 분석을 위한 적절한 통계 SW의 사용 여부를 함께 고려함
 - 오픈소스 기반의 R 프로그램 및 시각화 솔루션 QGIS, Tableau 등을 활용

과제	<ul style="list-style-type: none"> ■ 분석결과와 활용 목적과 기대효과를 구체적으로 정의 ■ 분석목적과 기대효과에 적합한 분석 방법 선정 ■ 분석 시나리오 작성 및 이에 근거하여 필요한 DATA의 종류, 유형 등 결정
추진내용	<ul style="list-style-type: none"> ■ 분석 결과와 활용목적과 기대효과를 구체적으로 정의 <ul style="list-style-type: none"> • 활용목적 : 시간별 현황 분석을 통한 민원 급증구간 파악, 민원인 유형별 민원 서비스 이용자 파악, 키워드 분석을 통한 반복 민원 처리, 충전기 유료화 이후 유형 변동 여부 확인, 주요 고장 발생지 및 원인에 대한 구체적인 통계 자료를 제공함으로써 데이터 기반 행정 정책 결정에 활용하고자 함 • 기대효과 : 전기차 콜센터 운영과 업무 효율을 증진하고, 충전기 최적 입지 선정 및 유지관리 체계 확립, 데이터에 기반한 과학적 행정 정책 수립과 맞춤형 서비스 제공 실현 ■ 적합한 분석 방법 선정 <ul style="list-style-type: none"> • 상위 빈도 키워드 식별, 충전기 설치현황, 시간대별/일별/요일별/월별/연도별/민원인 유형별 민원 건수 통계를 도출하고, 분석목적별 유의미한 특성 파악 • 1차적으로 파악한 특성을 시각적으로 표현하기 위해 도표, 그래프, 지도 등으로 분석 결과 산출 ■ 분석 시나리오 및 분석에 필요한 데이터 항목, 유형 등을 결정 <ul style="list-style-type: none"> • 분석 시나리오 설정 : 연구내용을 명확하게 하고 분석에 불필요한 것들을 배제하는 등 원활한 분석 활동을 위해 시나리오를 설정



추진내용	<p>■ 분석 시나리오</p> <p>(1단계) 개방형 충전기 설치 현황 기초 통계 분석</p> <ul style="list-style-type: none"> • 기존 전기차 충전소 현황 지도 시각화 • 사업자별/행정구역별 충전기 설치 현황 파악 <p>(2단계) 민원 건수 기초 통계 분석</p> <ul style="list-style-type: none"> • 시간대별/일별/요일별/월별/연도별 민원 건수 추이 • 민원인 유형별(일반인, 렌터카 등) 민원 접수 통계 산출 <p>(3단계) 주요 고장 발생지 및 원인 분석</p> <ul style="list-style-type: none"> • 읍면동별 민원 건수 및 충전기 설치 현황 추이 • 읍면동별 고장 발생 건수 지도 시각화 및 주요 원인 파악 <p>(4단계) 충전기 유료화 이후 민원 유형 변동 분석</p> <ul style="list-style-type: none"> • 민원 건수 시계열 자료/문의사항 기술 통계량 확인 • 유료화 전후 데이터 분할 및 키워드 비교 분석 <p>(5단계) 반복민원 파악을 위한 텍스트마이닝</p> <ul style="list-style-type: none"> • 연도(2015~2019년)에 따른 Top 20 키워드 추출 • 주요/기타 키워드 파악을 위한 워드클라우드 시각화 • 반복 및 단순 민원 현황 파악, 분류 카테고리 재선정 건의 <p>(6단계) 효율적 충전기 설치 및 유지관리 방안 마련</p> <ul style="list-style-type: none"> • 설치 및 유지관리 관련 민원 건수 통계량 도출 • 세부 내용 추출 및 개선 사항 모색 <p>■ 데이터 분석에 필요한 항목과 수치 정의</p> <ul style="list-style-type: none"> • 측정항목 및 분석범주 : 시간대별/일별/요일별/월별/연도별/민원인 유형별 민원 건수, 제주 전기차 충전소 현황, 행정동별 충전기 설치 현황 통계 등
Tip	<p>■ 분석 설계는 본격적인 분석을 시작하기 위한 전반적인 과정을 계획하는 단계로 명확하고 구체적인 설계를 통해 원래 분석 목적대로 진행을 원칙으로 함</p>

3. 데이터 수집 및 가공

■ 분석에 필요한 원 자료(RAW DATA)를 보유한 기관 등을 확인하여 적절한 방법으로 수집 및 가공

과제	<ul style="list-style-type: none"> ■ 분석에 필요한 데이터 리스트 및 보유기관 확인, 데이터 수집 ■ 수집 후 분석이 용이한 형태로의 데이터 가공
추진내용	<ul style="list-style-type: none"> ■ 데이터 보유 주체(기관) 확인 및 수집 <ul style="list-style-type: none"> • 분석에 필요한 내부 보유데이터(raw data)수집 및 외부 데이터 보유 기관 확인 후 데이터 수집, 필요에 따라 기관 간 협업 • (내부) 민원 데이터, 개방형 충전기 설치 현황 데이터 - 제주 EV 콜센터를 통해 쿼리를 제공받아 5개 년도 데이터 일괄 추출 • (외부) GIS 관련 데이터, 제주도 전기차 충전소 데이터 활용(공공데이터포털) ■ 유사 카테고리별 통합, 18년 이후 변동 자료 양식 통일 <ul style="list-style-type: none"> • 카테고리 특수문자 구분 삭제, 분석에 필요한 칼럼 추출 ■ 결측값(missing value) 및 이상치(outlier) 처리 등 데이터 전처리 <ul style="list-style-type: none"> • 분석이 용이한 형태로 데이터 전처리(preprocessing) ■ 수집한 데이터 가공 및 최종 분석용 데이터 셋 생성 <ul style="list-style-type: none"> • 수집한 데이터를 전처리, 재가공하여 분석 목적에 맞는 최종 분석용 데이터 셋을 구축하여 구체적인 분석을 수행하기 위한 준비
Tip	<ul style="list-style-type: none"> ■ 연도별 통계 자료에 대하여 데이터 접수 양식 (예) 기존 월/일 구분 → 일자 일괄 통합)이 상이함 <ul style="list-style-type: none"> ※ 집계방식, 기준시점에 따라 확인 필요 ■ 충전소 현황 데이터 수집 시 자료 변동의 소지가 있을 수 있으므로 최신의 파일 확보 필요

4. 데이터 분석 및 검증

- 수집하여 가공한 정보(DATA)를 적절한 BI(Business Intelligence) 소프트웨어를 이용하여 분석함
- 수집된 정보(DATA)의 품질을 확인하고, 예상되는 분석 산출물에 대한 정의 필요
- 분석결과가 도출될 때마다, 협업담당자와의 의사소통을 통해 검증 필요

과제	<ul style="list-style-type: none"> ■ 분석 산출물에 대한 정의 ■ 수집 정보(DATA)에 대한 품질 검증 실시 ■ 분석 설계 시 작성했던 분석 시나리오에 따라 도출하고자한 주제별 분석 실시 ■ 분석 결과 확인을 통한 분석방법 검증
추진내용	<ul style="list-style-type: none"> ■ 분석 산출물에 대한 정의 <ul style="list-style-type: none"> • 산출물 정의 <ul style="list-style-type: none"> - 일자별 민원 특성 및 충전기 현황을 분석하여 분석 결과 산출 - 분석 결과를 바탕으로 반복 및 단순 민원 처리 업무 효율성 증진, 민원 급증 구간 운영 유연화, 유료화 이후 유형 변동 여부 파악, 주요 고장 원인에 대한 대처, 카테고리 재정비 및 유지관리 방안 마련 • 산출물의 시각화 여부 <ul style="list-style-type: none"> - 그래프, 표, 지도 등을 사용하여 전달력이 높고, 현업 업무활용도가 높은 시각화 결과물 도출 ■ 수집 정보(DATA)에 대한 품질 검증 실시 <ul style="list-style-type: none"> • 수집 정보는 해당 민원 콜센터에서 직접 제공한 데이터로 별도의 품질 검증이 필요하지는 않지만, 연도별 발생 건수를 비교해보았을 때, 변동의 폭이 너무 큰 연도는 없는지에 대한 판단 후, 현업 담당자의 사실 여부 확인 ■ 분석시나리오에 따른 주제별 분석 실시 <ul style="list-style-type: none"> • 민원 현황 및 충전소와 충전기 설치 현황 분석에 대한 시나리오를 기반으로 작성 • 민원 접수 건수 현황 시간대별/일별/요일별/월별/연도별/민원인 유형별 기초 통계 분석 실시 • 읍면동별 고장 발생 건수로 분류하여 주요 고장 발생지 및 원인 파악과 지도 시각화 • 충전기 유료화 이후 민원 건수/문의사항 등의 변동 여부 기술 통계량 분석 • 유료화 전후 데이터 분할하여 시계열 자료 검토, 경향성 확인 및 예측값 산출 • 연도별 민원 문의사항 상위 빈도 키워드 추출 및 워드클라우드 기반 시각화 • 중복 카테고리 제거 및 빈도표 작성을 통한 재정비 • 민원 세부 내용 분석을 통한 최적 설치 입지 검토 및 효율적인 유지관리 방안 마련



추진내용	<ul style="list-style-type: none"> ■ 분석 결과 확인을 통한 분석방법 검증 <ul style="list-style-type: none"> • 각 주제별 산출된 분석 결과를 바탕으로 현업 담당자와 내용을 검토하여 확인 과정을 거침 • 분석 방법 및 산출물 검증을 통해 오류 내용을 수정하여 재분석하는 과정 반복, 유의한 결과가 도출될 수 있도록 함
Tip	<ul style="list-style-type: none"> ■ 어떤 데이터 분석방법론을 통해 분석 결과를 도출하느냐도 중요하지만, 수집 정보 (DATA)의 품질 검증은 해당 업무 담당자와의 협업을 통해 오류 파악을 하는 것이 훨씬 효과적임 ■ 관련 업무 담당자와의 논의 시, 관련 업무의 문제점이나 기존에 실시 정책에 대한 오류를 지적하고자 하는 것이 아님을 밝히고, 본 분석의 목적을 명확히 설명하고 도출되는 결과에 대한 시사점을 분명하게 제시하여야 함

5. 데이터 시각화

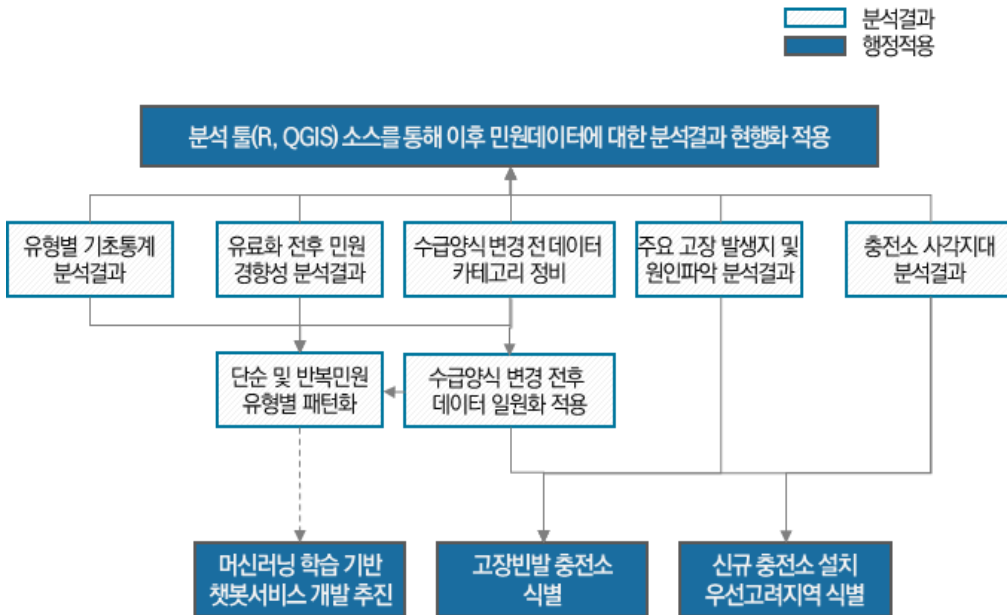
■ 분석결과를 바탕으로 R, QGIS를 활용, 빅데이터 및 지도 시각화

과제	■ 분석결과를 바탕으로 R, QGIS 활용, 빅데이터 및 지도 시각화
추진내용	<ul style="list-style-type: none"> • 민원 현황 분석결과를 현업 담당자들이 쉽게 접근하고 활용할 수 있도록 하기 위해서는 R 기반 통계량 시각화 대시보드 구현 필요 • 전기차 충전소 설치 현황 GIS 기반 지도 시각화 • 민원 시간대별/일별/요일별/월별/연도별/민원인 유형별 접수 건수 파악 • 행정동별/사업자별 충전기 설치 현황 분석 • 주요 고장 발생지 및 원인 분석 • 유료화 이후 민원 유형 변동 여부 시계열 그래프 표현 및 기술통계량, 변동 유형 비교 분석 • 연도별 민원 상위 빈도 키워드 추출 및 텍스트 분석 • 중복 제거 및 피벗테이블 작성을 통한 카테고리 재정비 • 설치 및 유지관리 관련 민원 세부 내용 분석
Tip	<ul style="list-style-type: none"> ■ 본 분석에서 적용한 분석기간(15.6~19.5)이후 데이터의 분석 및 시각화를 위해서는 추가적으로 형식에 맞는 데이터 수집 필요



행정적용 단계

1. 행정적용 단계 요약



2. 행정정책개선

■ 제주특별자치도 구축 개방형 충전기 운영관리체계 구축 및 개선에 반영

- 충전기 플랫폼 운영 추진('20년 예산 350,000천원 책정)
 - 충전기 유료 서비스 과금 · 정산 관리 및 고장 · 수리 이력의 체계적 관리 목적
 - 전기차 및 충전기 통합 콜센터 운영
 - 데이터 관리를 통한 민원패턴 분석 및 충전인프라 운영관리 정책 수립
- ※ 빅데이터 분석을 통해 '15.6 ~ '17.12 데이터 카테고리를 자료수집 양식이 변경된 '18 이후 수급 양식에 맞게 일원화 적용 가능해짐
 - 전기차 이용자의 관련 민원(불편)사항 즉시 해결 및 통합관리로 이용자 만족도 제고
- 개방형 충전인프라 구축 및 기능 개선('20년 예산 300,000천원 책정)



- 개방형 충전인프라 기능개선 및 노후화 시설(캐노피 등)정비 등을 통한 전기차 이용자의 안전과 충전 편의 증진으로 전기차 이용환경 최적화
- 개방형 충전인프라 구축을 통한 전기차 이용자들의 불편 최소화
- ※ 빅데이터 분석을 통해 민원 빈발 충전소 및 충전소 설치 사각지대의 직관적 현황 파악 및 이를 바탕으로 고장빈발 충전소 식별 및 신규 충전소 최우선 설치 추진 가능해짐

■ 반복 및 단순민원에 대한 유형별 패턴화를 통해 유관부서의 민원대응 리소스 절감 방안 추진

- 빅데이터 분석을 바탕으로 한 민원유형별 패턴화를 통한 머신러닝 학습 반영 가능해짐
- 머신러닝 학습을 바탕으로 챗봇 서비스 개발 및 이로 인한 민원대응 리소스의 최적화 및 절감 가능해짐

※ 실효성 있는 민원 챗봇 개발은 빅데이터 분석을 통한 데이터 패턴화 없이 구현이 어려움

- 챗봇은 앱처럼 따로 설치하거나 사용법을 배울 필요 없이 메시지만 보내면 원하는 서비스를 실시간으로 받을 수 있음
- 챗봇은 민원대응부서에는 리소스를 최적화시킬 수 있고 민원인에게는 시간절약 및 편의성(24시간 민원 응대 가능)을 줄 수 있는 수단임



금연구역 후보지 도출 및 관리를 위한 흡연 민원지역 빅데이터 분석

| 서울특별시 성동구 / 강성호·원준호·이현정 |

요약

국민건강 증진과 쾌적한 환경조성을 위해 매년 금연구역을 확대하고 있음에도 불구하고 간접흡연 피해에 대한 민원이 지속적으로 증가하고 있는 현실이다. 또한 기존의 금연구역 조례 지정은 뚜렷한 기준 없이 지자체별로 개별 운영하고 있어 과학적인 근거 기준 마련이 필요하다.

이에 본 분석은 비흡연자와 흡연자의 갈등을 최소화하고 흡연 민원 스마트 관리체계를 통한 금연구역의 효율적인 관리를 위해 빅데이터를 활용하였다. 이를 위해, 흡연 민원 및 흡연 과태료 단속 데이터를 분석하여 흡연 핫스팟(hotspot)¹⁾ 지역을 파악하고 시간대별 집중지역을 도출하였다. 또한 흡연유발지수 산출 모델을 개발하여 국토교통부 표준격자(100m*100m) 단위별 흡연 발생 가능성을 정량화하고 흡연유발지수 상위지역을 흡연집중지역으로 분류하여 금연구역 후보지를 선정하는 기준을 제시하였다.

이를 통해 우리구 흡연지도를 제작하여 관내 흡연 관련 현황을 파악하는 정책지도로 활용하고 시간대별 흡연 단속원 배치계획을 수립하여 현재 민원지역 위주 단속의 한계를 극복하는 인적자원 활용으로 단속 효율성을 높이며 금연구역 조례 지정을 위한 과학적 기준을 마련하여 비흡연자와 흡연자가 상생할 수 있는 도시 공간 생태계의 기반을 마련하고자 한다.

1) 핫스팟 분석(Hotspot Analysis)은 공간 현상의 입지나 분포 패턴을 분석하는 기법 중 하나로, 핫스팟 지역은 특정 장소의 값의 높고 높은 값을 갖는 인접 지역들로 둘러싸여 있는 지역이다



가 추진배경

■ 흡연 단속 인력 한계로 민원발생 지역 위주 지도 점검 실시

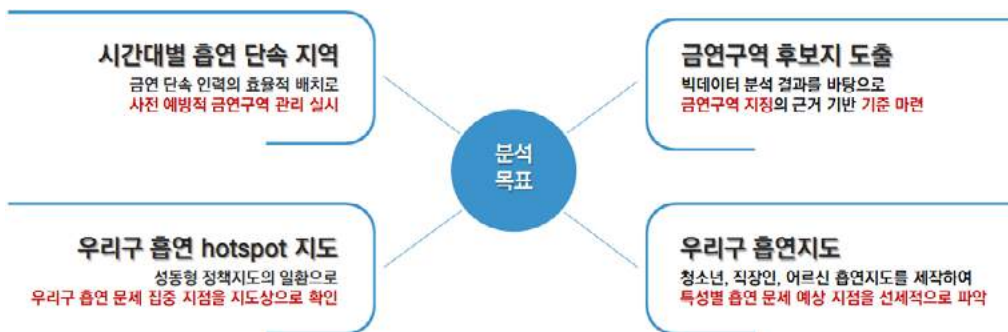
- 간접흡연 피해민원 2015년 128건에서 2019년 8월 기준 370건으로 2.9배 증가
- 단속인력 1인당 금연시설 약 2,000개소 관리
- 단속 인력의 효율적 배치로 사전 예방적 금연구역 관리 필요

■ 비금연구역의 간접흡연 피해 지속 증가로 금연구역 확대 필요

- 금연구역 사각지대인 주택가, 공개공지, 이면도로 등 공공장소 내 간접흡연 피해민원 지속 증가(전체 민원의 52%)

■ 금연정책 추진시 비흡연자와 흡연자의 갈등 심각

- 실외 금연구역 확대 [비흡연자] ↔ 공공 흡연시설 확충 [흡연자]
- 과학적 근거 기반 금연구역 지정 필요
- 빅데이터 분석자료를 근거로 금연구역 지정의 우선순위 결정





프로젝트 개요

■ 분석개요

- 분석기간: 2019년 8월 ~ 10월(3개월)
- 분석공간범위: 서울 성동구 관내, 국토교통부 표준격자(100m*100m)
- 분석시간범위: 2016년 ~ 2018년(3개년)

■ 주요 분석내용

- 흡연민원 및 단속 과태료 현황 파악
 - 흡연 민원 통계 분석: 연도별/행정동별/월별/요일별/민원 유형별 분석
 - 흡연 단속 과태료 통계 분석: 연도별/행정동별/월별/요일별/시간대별 분석
- 흡연유발지수 개발 및 연령층별 유형화
 - 청소년, 직장인, 어르신 연령층별 유형화 및 흡연유발지수 개발
 - 흡연유발지수 분포를 통해 연령층별 흡연집중지역 도출
- 금연구역 후보지 도출
 - 전체 및 연령층별 흡연집중지역을 확인
 - 흡연집중지역 중 현재 금연구역이 아닌 지역을 금연구역 후보지로 도출

■ 활용데이터

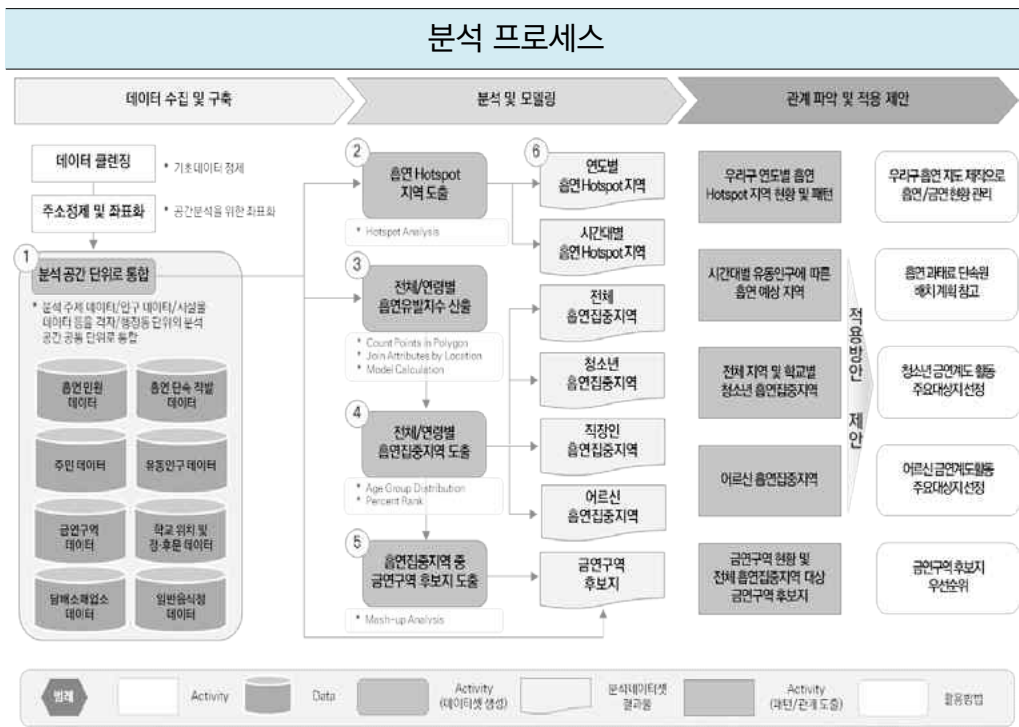
순번	데이터명	시간범위	출처
1	응답소 흡연 관련 민원신고	2016 - 2018	서울시
2	보건소 흡연 민원 접수 대장	2016 - 2018	성동구 보건소
3	새울 흡연 민원 신고	2016 - 2018	성동구
4	흡연 단속 과태료 대장	2016 - 2018	성동구 보건소
5	금연구역 지정 고시 현황	2019	성동구 보건소
6	성동구 학교 위치(정문, 후문)	2019	성동구 빅데이터센터
7	성동구 지하철 출입구 위치	2019	성동구 빅데이터센터
8	성동구 버스정류소 위치	2019	서울시
9	길단위 추정 유동인구	2018	서울시
10	담배소매업소	2019	행정안전부
11	일반음식점	2019	행정안전부
12	PC방(인터넷컴퓨터게임시설제공업소)	2019	행정안전부
13	국토교통부 표준격자(100m*100m)	2019	국토교통부



다 적용 프로세스

■ 분석 전체 프로세스

- 사업기획: 문제의식 → 현업부서 요청 → 분석 → 행정적용
 - 데이터분석: 분석모델설계 → 데이터 수집 및 구축 → 분석 및 검증 → 결과 도출 → 적용 제안
 - 시스템 연계: 분석 기초 데이터 → 분석 결과 데이터 → 분석 결과
- (* 성동구 빅데이터 분석 과제는 '성동구 빅데이터 분석플랫폼'에 플러그인 방식으로 탑재되어 데이터 및 분석 결과 공유)





라 기획·관리 단계

1. 기획·관리 요약

단계	주요 내용
문제의식	<ul style="list-style-type: none"> ■ 흡연 단속 인력 한계로 민원발생 지역 위주 지도 점검 실시 ■ 비금연구역의 간접흡연 피해 지속 증가로 금연구역 확대 필요 ■ 금연정책 추진 시 비흡연자와 흡연자의 갈등 심각
사업추진	<ul style="list-style-type: none"> ■ 요구사항 정의 ■ 분석 계획 수립 및 확정 ■ 사업범위, 사업기간, 사업 담당자 문서화
사업완료	<ul style="list-style-type: none"> ■ 흡연민원 및 단속 과태료 현황 파악 ■ 흡연유발지수 개발 및 연령층별 유형화 ■ 금연구역 후보지 도출

2. 문제의식

- 흡연 단속 인력 한계로 민원발생 지역 위주 지도 점검 실시
 - 간접흡연 피해민원 2015년 128건에서 2019년 8월 기준 370건으로 2.9배 증가
 - 단속인력 1인당 금연시설 약 2,000개소 관리
 - 단속 인력의 효율적 배치로 사전 예방적 금연구역 관리 필요
- 비금연구역의 간접흡연 피해 지속 증가로 금연구역 확대 필요
 - 금연구역 사각지대인 주택가, 공개공지, 이면도로 등 공공장소 내 간접흡연 피해민원 지속 증가(전체 민원의 52%)
- 금연정책 추진 시 비흡연자와 흡연자의 갈등 심각
 - 실외 금연구역 확대 [비흡연자] ↔ 공공 흡연시설 확충 [흡연자]
 - 과학적 근거 기반 금연구역 지정 필요
 - 빅데이터 분석 자료를 근거로 금연구역 지정의 우선순위 결정



과제	<ul style="list-style-type: none"> ■ 문제점 확인 ■ 데이터 분석 방안 도출 ■ 데이터 분석 결과 예측 및 구체화
추진내용	<ul style="list-style-type: none"> ■ 문제점 확인 <ul style="list-style-type: none"> • 현업 부서 이슈로 부각되고 있는 주요 항목 정리 ■ 데이터 분석 방안 도출 <ul style="list-style-type: none"> • 주요 이슈 및 문제점에 대한 데이터 분석 가능성 검토 • 중앙부처 및 타 지자체 빅데이터 분석 사례 검토 • 데이터 분석을 위한 데이터 확보 방안 검토 ■ 데이터 분석 결과 예측 및 구체화 <ul style="list-style-type: none"> • 주요 항목별 데이터 분석 결과 예측 • 주요 항목별 빅데이터 분석 방법 및 모델 검토 • 빅데이터 분석 결과 구체화
Tip	<ul style="list-style-type: none"> ■ 데이터를 통한 문제해결 가능 여부를 파악하는 것이 중요 ■ 현업 부서의 경우 빅데이터 분석을 통한 문제해결 가능 여부를 판단하기 어려운 측면이 있으므로 이에 대한 지원 필요



3. 사업추진

■ 분석을 위한 요구사항 정의

■ 분석 계획 수립 및 확정

■ 사업범위, 사업기간, 사업 담당자 문서화

과제	<ul style="list-style-type: none"> ■요구사항 정의 ■분석 계획 수립 및 확정 ■사업범위, 사업기간, 사업 담당자 문서화
추진내용	<ul style="list-style-type: none"> ■요구사항 정의 <ul style="list-style-type: none"> • 현업부서 담당자와의 협의를 통한 세부 분석 주제 검토 • 주요 분석 주제별 분석 가능 여부 검토 및 확정 ■보유 데이터 검토 및 데이터 확보 가능 여부 검토 <ul style="list-style-type: none"> • 각 부서별 보유데이터 검토 후 데이터 제공 가능 여부 검토 • 성동구 자체 수급 가능 데이터와 외부 공공기관 수급 가능 데이터 분류 및 수급 • 외부 공공기관 및 민간데이터의 경우 데이터 보유기관 제공 가능 여부 파악을 통해 확보 가능 데이터와 불가능 데이터 분류 ■사업 범위 및 소요 기간 확정 <ul style="list-style-type: none"> • 데이터 수급 가능 여부를 기준으로 전체 요구사항 중 데이터 확보가 가능한 범위로 사업 범위 확정 • 분석 과제별 소요 기간 검토 및 확정
Tip	<ul style="list-style-type: none"> ■내부 데이터 중 실제로 분석 활용이 불가능한 데이터가 있으므로 사업계획 단계 시 파악하는 것이 중요 ■사업 초기에 데이터 수급 여부를 잘 검토하지 않으면 분석 진행 중 분석이 불가능 하거나 재분석이 필요한 경우 발생



4. 완료보고

■ 흡연민원 및 단속 과태료 현황 파악

■ 흡연유발지수 개발 및 연령층별 유형화

■ 금연구역 후보지 도출

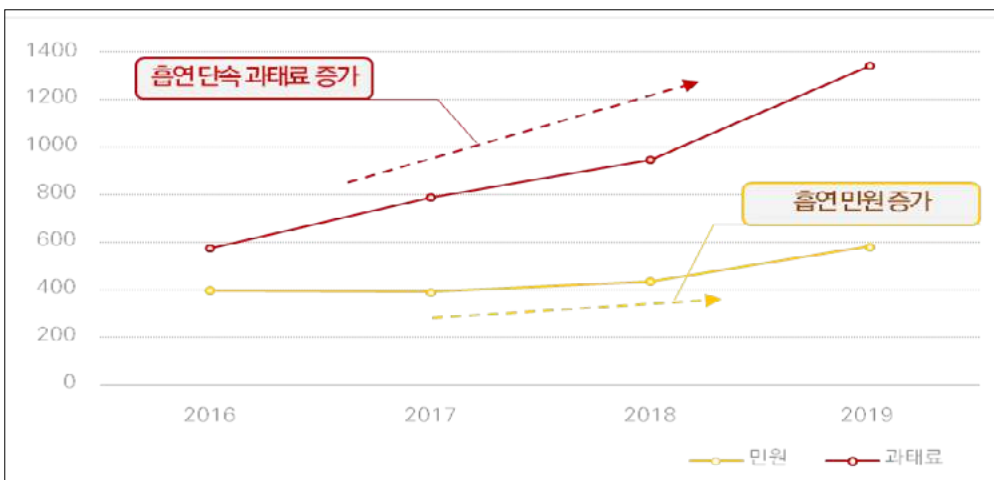
과제	<ul style="list-style-type: none"> ■ 흡연민원 및 단속 과태료 현황 파악 ■ 흡연유발지수 개발 및 연령층별 유형화 ■ 금연구역 후보지 도출
추진내용	<ul style="list-style-type: none"> ■ 흡연민원 및 단속 과태료 현황 파악 <ul style="list-style-type: none"> • 흡연민원 통계분석 : 연도별/행정동별/월별/요일별/민원 유형별 분석 • 흡연단속 과태료 통계분석 : 연도별/행정동별/월별/요일별/시간대별 분석 ■ 흡연유발지수 개발 및 연령층별 유형화 <ul style="list-style-type: none"> • 청소년, 직장인, 어르신 연령층별 유형화 및 흡연유발지수 개발 • 흡연유발지수 분포를 통해 연령층별 흡연집중지역 도출 ■ 금연구역 후보지 도출 <ul style="list-style-type: none"> • 전체 및 연령층별 흡연집중지역을 확인 • 흡연집중지역 중 현재 금연구역이 아닌 지역을 금연구역 후보지로 도출
Tip	<ul style="list-style-type: none"> ■ 분석과제에 대한 정책개발 및 적용방안 도출 시 담당부서와 협조 필요

5. 분석결과

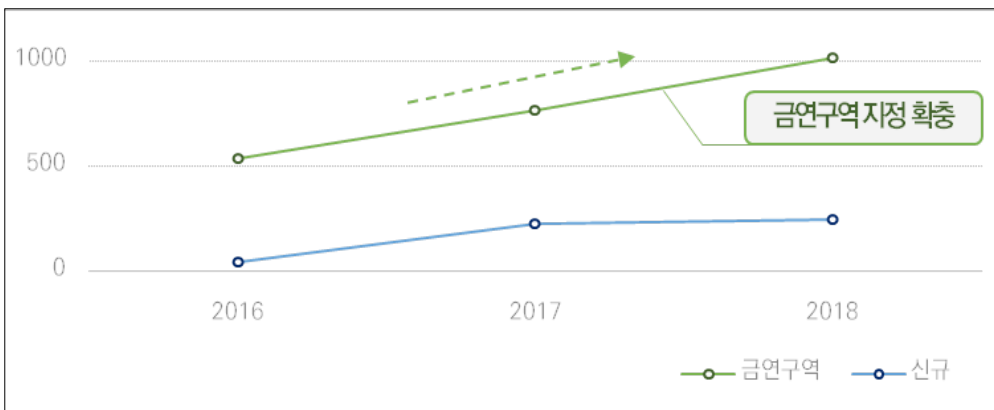
1. 흡연 민원 및 단속 통계분석

■ 연도별 흡연 민원 및 단속 발생 건수

- 흡연 민원은 2016년 283건에서 2019년 584건으로 지속적 증가 추세
- 흡연 단속 과태료는 2016년 576건에서 2019년 1,346건으로 대폭 증가 추세
- 흡연관련 민원 증가로 단속, 계도 강화 및 금연구역 확충
- 보다 효율적인 금연구역 지정 및 단속활동 계획을 통해 민원 감소 강구
- 지속적 민원증가로 흡연 과태료 단속 강화 및 금연구역 확충



〈연도별 흡연 민원 및 단속 발생 건수〉

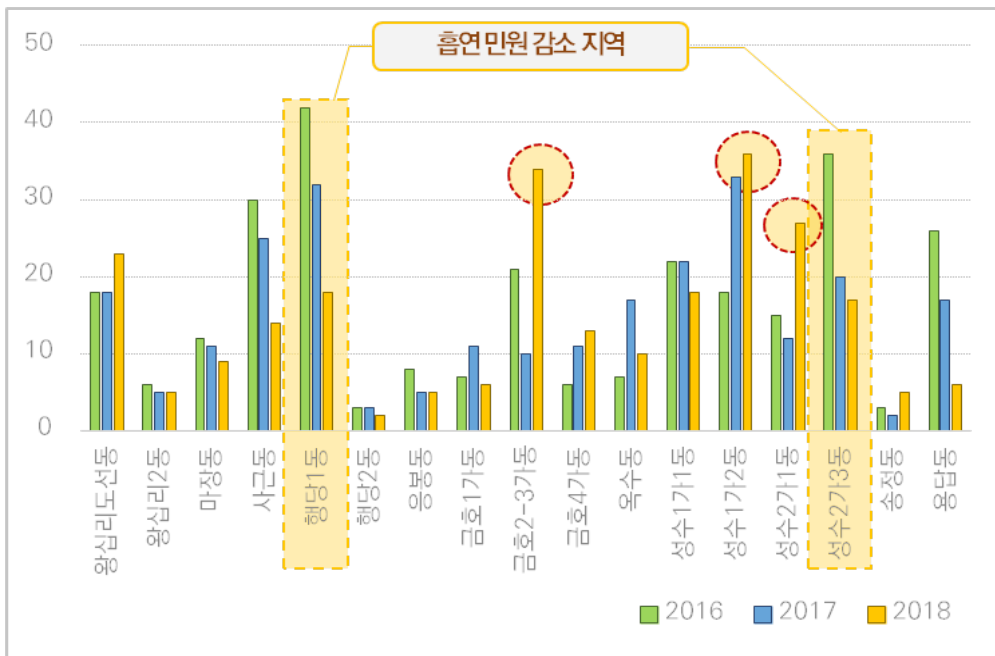


〈연도별 금연구역 지정 현황〉



■ 행정동별 흡연 민원 발생 건수

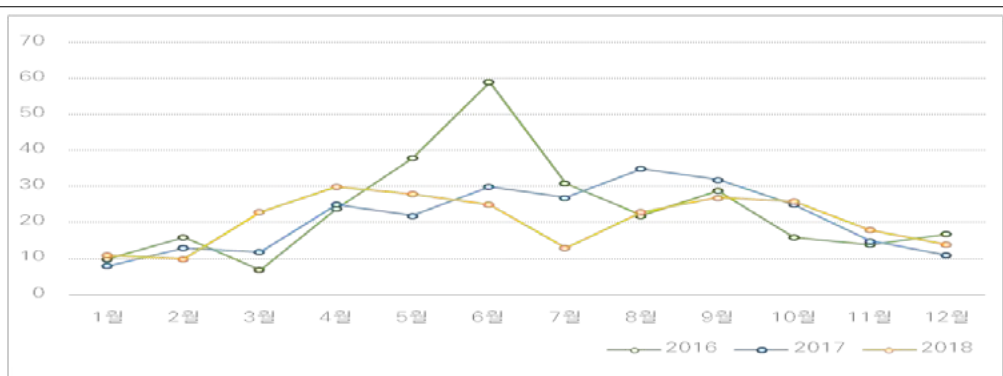
- 행당1동과 성수2가3동은 2016년도에는 민원이 많았으나 지속적으로 감소 추세
- 행당1동의 민원감소는 압도적인 단속 건수의 영향이 있을 것으로 판단됨
- 성수1가2동이 14.52%, 금호2-3가동이 13.71%, 성수2가1동이 10.89%로 2018년 흡연민원이 많은 상위권 지역
- 금호2-3가동의 경우 지역 수준이 높아짐에 따라 주변 환경에 따른 불편 신고가 증가했을 것으로 예상됨
- 성수1가2동과 성수2가1동의 경우 지역의 상권이 개발됨에 따라 유동인구 증가로 인한 흡연 불편 피해가 증가했을 것으로 예상됨



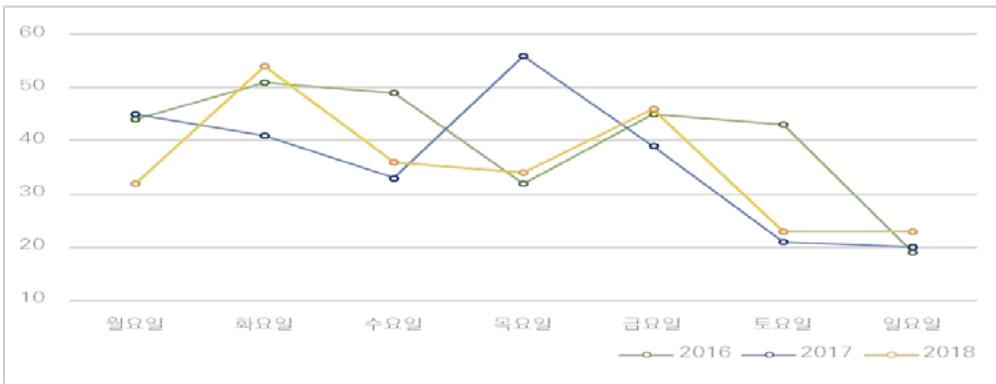
〈행정동별 흡연 민원 및 단속 발생 건수〉

■ 월별/요일별 흡연 민원 발생 건수

- 월별, 요일별 분석에서는 부각되는 신고 패턴 없음
- 2018년 민원은 4월, 9월과 화요일, 금요일에 집중적
- 7월과 주말에는 민원 신고가 적게 나타남



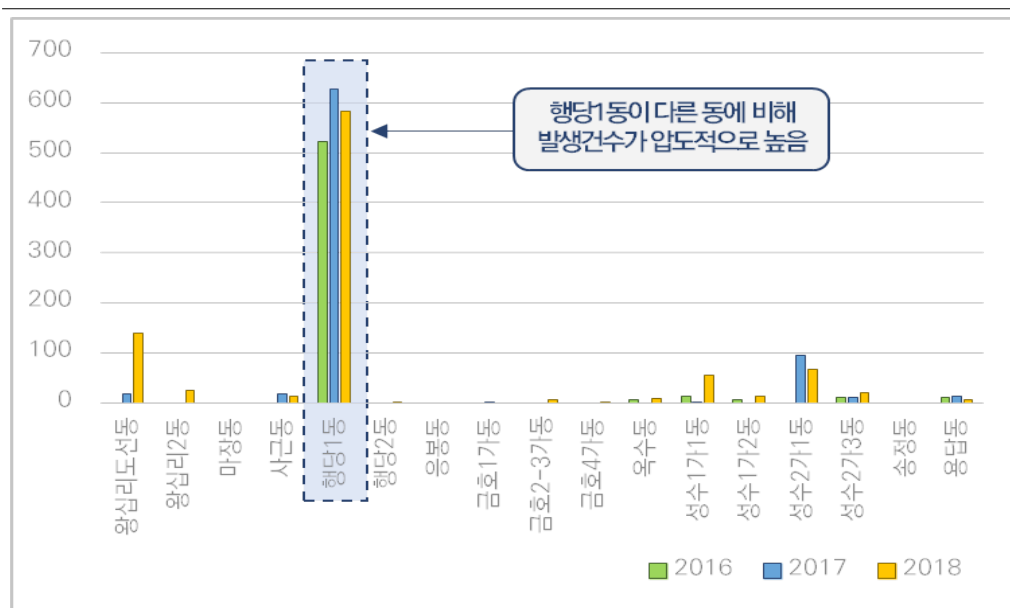
〈월별 흡연 민원 발생 건수〉



〈요일별 흡연 민원 발생 건수〉

■ 흡연 민원 워드 클라우드

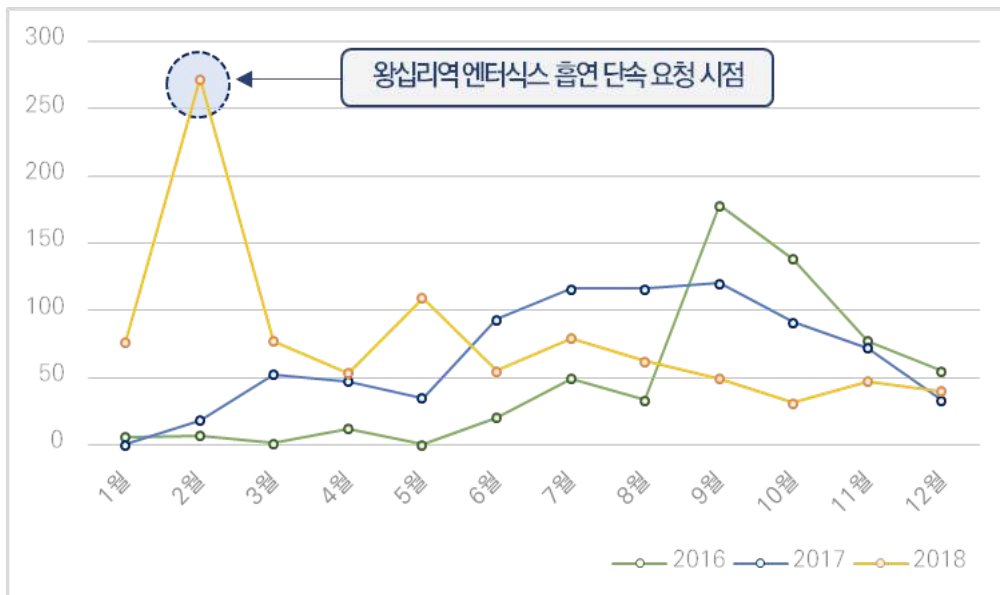
- 텍스트 마이닝 결과 총 5,339개의 단어 중 성동구(395개), 단속(248개), 담배꽂초(216개), 쓰레기(202개), 금연구역(178개) 이 집중적으로 도출됨
- 편의점, 정류장, 지하철 출구, 놀이터 등 특정 장소를 지칭하는 단어들이 많이 도출됨
- 어린이, 아이들, 유모차, 유치원 등 어린이와 관련된 단어들이 도출되었으며, 간접 흡연으로 인한 어린이들의 피해가 막대하다는 점을 유추 가능
- 청소년, 고등학교 등 청소년과 관련된 단어들이 도출되었으며, 청소년 흡연으로 인한 피해가 있다는 점을 확인
- 그 외에도 CCTV, 과태료, 시정조치, 스티커 등 요청, 수거, 처리를 요청하는 단어들이 도출되었으며, 결과 확인을 요청하는 단어들이 다음으로 도출됨



〈행정동별 흡연 단속 발생 건수〉

■ 월별 흡연 단속 발생 건수

- 2018년 과태료 단속은 2월에 가장 많이 나타남
- 왕십리역 엔터식스에서 흡연 단속 요청을 반영하여 단속을 집중한 결과



〈월별 흡연 단속 발생 건수〉



■ 요일별 흡연 단속 발생 건수

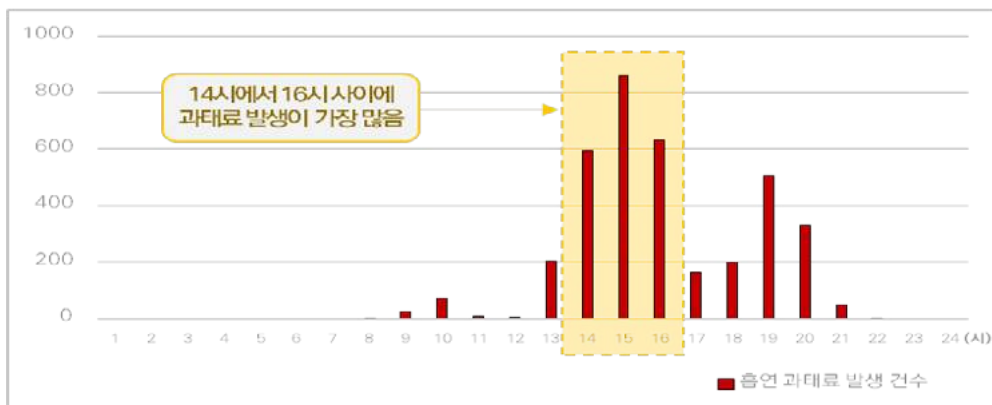
- 2018년 흡연 단속은 목요일에 31.58%로 가장 많으며, 주말 단속은 없음
- 목요일은 월3회 야간 단속을 추가적으로 운영하고 있어 단속 건수 많음



〈요일별 흡연 단속 발생 건수〉

■ 시간대별 흡연 단속 발생 건수

- 현재 단속원은 '오전9시 - 오후5시' 와 '오후1시 - 5시' 두 개 조로 나누어 운영
- 단속이 집중되는 '오후2시 - 4시'에 가장 많은 단속 건수가 나타남

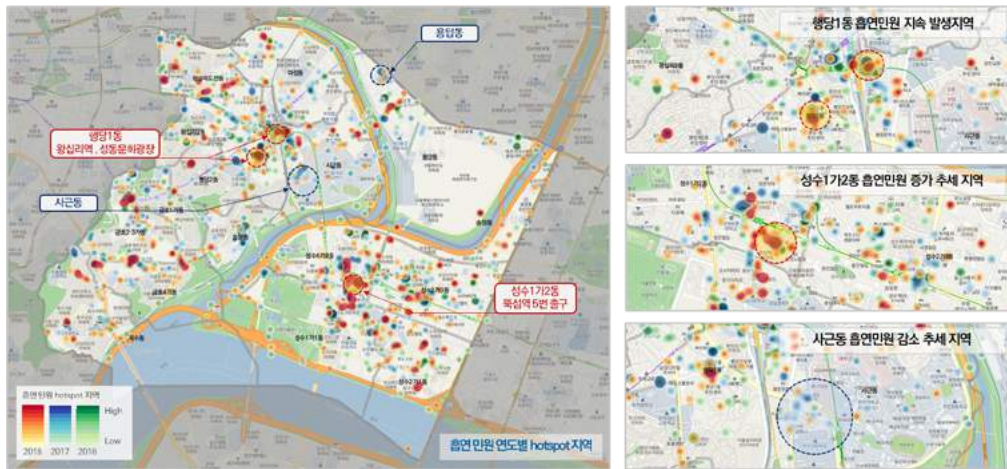


〈시간대별 흡연 단속 발생 건수〉

2. 흡연 Hotspot 분석

■ 흡연 민원 연도별 hotspot 분석

- 흡연 민원은 2016년도 대비 2018년에 108.97% 증가
- 용담동, 행당1동, 사근동은 민원 감소 추세인 지역이나 행당1동은 국지적으로 민원이 지속적으로 발생하고 있는 왕십리역과 성동문화광장이 확인됨
- 성수1가2동은 2년 동안 민원이 2배 증가한 지역으로 독섬역 주변을 중심으로 민원이 집중
- 지속적 민원증가로 흡연 과태료 단속 강화 및 금연구역 확충



〈흡연 민원 연도별 hotspot 지역〉

■ 흡연 단속 연도별 hotspot 분석

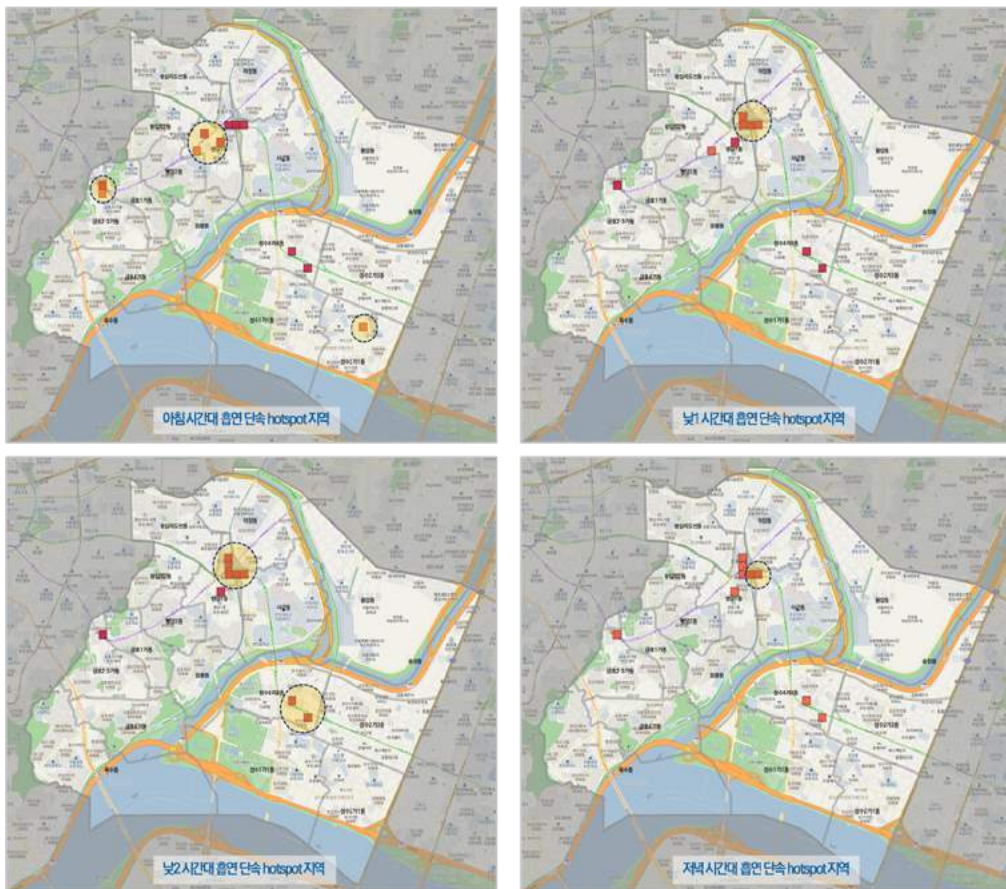
- 흡연 단속은 2016년도 대비 2018년에 164.93% 증가 추세
- 행당1동 왕십리역사와 왕십리광장이 전체 과태료의 68.55%를 차지
- 성수2가3동 성수역 인근 지역은 지속적인 과태료 부과 지역으로 확인됨



〈흡연 단속 연도별 hotspot 지역〉

■ 시간대별 흡연 민원 hotspot 분석

- 아침(06:00~10:59), 낮1(11:00~13:59), 낮2(14:00~16:59), 저녁(17:00~20:59)
- 시간대별 유동인구와 민원 분포를 활용한 단속 필요지역 분석
- 전체 유동인구가 많은 왕십리역 주변과 직장인 집중지역인 성수동 독섬역 주변은 전체 시간 집중 지역
- 아침에는 왕십리역교차로, 신금호역교차로, 경수초등학교 인근에서 집중되었다가 정오부터는 왕십리광장으로 이동
- 늦은 낮 시간대는 왕십리광장과 독섬역 인근에서 집중되었다가 저녁 시간대는 사근동 음식점 밀집지역으로 이동



〈시간대별 흡연 단속 hotspot 지역〉

3. 흡연집중지역

■ 흡연유발지수 모델

- 흡연유발지수는 흡연 민원 및 단속 데이터와 주거인구, 유동인구, 시설물 데이터를 활용하여 흡연 발생 가능성을 정량적으로 산출한 지수
- 흡연유발지수 산출 모델을 개발·적용함으로써 수치화된 정량지표를 기반으로 흡연집중 지역을 도출하여 연령대별로 분류된 각 유형별로 맞춤형 정책을 적용 가능

$$\begin{aligned} \text{흡연 발생건수} &= \text{흡연 발생건수} = \text{흡연 민원 건수} + \text{흡연 단속 건수} \\ \text{전체 흡연유발지수} &= +(\text{주거인구} \times \text{가중치}) + (\text{유동인구} \times \text{가중치}) + (\text{담배판매업소 점포수} \times \text{가중치}) + (\text{음식점 점포수} \times \text{가중치}) \\ &+ (\text{버스정류소 수} \times \text{가중치}) + (\text{지하철 출입구 수} \times \text{가중치}) \end{aligned}$$



* 청소년 (10대), 직장인(20~50대), 어르신(60대 이상) 연령을 대상

〈흡연유발지수 산출 모델〉

■ 전체 흡연지도 분석

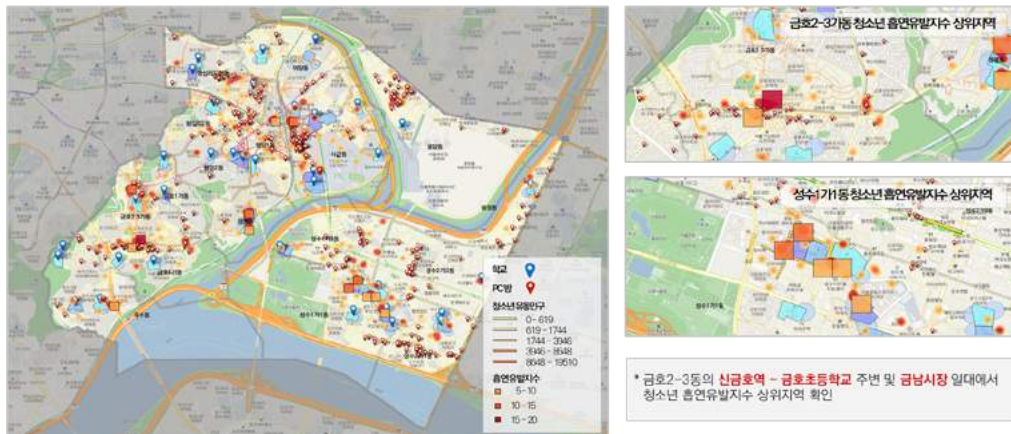
- 행당1동과 사근동에서 흡연유발지수 상위 0.1% 지역 확인 - 왕십리역사 및 주변 상권을 중심으로 발생
- 행당2동, 옥수동에서 상위 0.5% 지역 확인 - 행당역, 옥수역 등 지하철역 주변에 발생
- 흡연유발지수 상위지역은 주로 지하철역을 중심으로 도출되며 주변 상권(담배판매업소, 음식점) 및 유동인구가 영향을 끼쳤을 것으로 판단됨



〈전체 흡연집중지역〉

■ 청소년 흡연지도 분석

- 청소년 흡연은 단속 과태료(50%) 부과 대상이기도 하지만 연령대를 감안하여 교육 및 계도를 위한 자료로 활용 필요
- 청소년 흡연유발자수 상위지역은 금호2-3가동, 성수1가1동, 응봉동, 사근동에서 확인
- 금남시장 일대에 청소년 흡연 근절을 위한 계도 활동 집중 필요

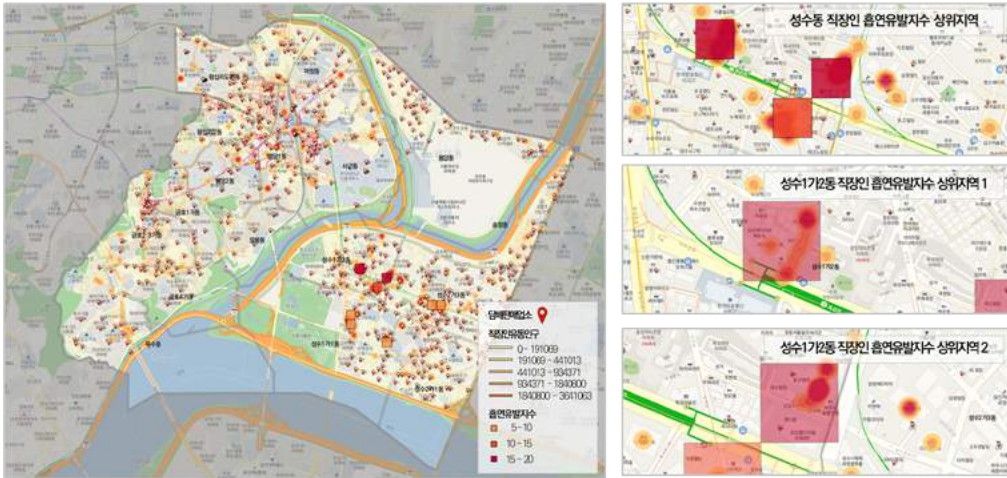


〈청소년 흡연집중지역〉

■ 직장인 흡연지도 분석

- 직장인 흡연집중지역은 성수동 일대에서 확인
- 성수1가2동은 독섬역을 중심으로 서울숲엘타워와 서울숲IT밸리 인근에서 흡연유발 지수가 높게 도출

- 성수1가1동은 서울숲역 인근 서울지식산업센터 주변에서 흡연유발지수가 높게 도출



〈직장인 흡연집중지역〉

■ 어르신 흡연지도 분석

- 어르신 흡연집중지역은 광장 및 공원의 휴게 공간이 조성되어 있는 지역에서 발생
- 행당1동의 성동문화광장은 소월아트홀 앞의 휴게 공간으로 왕십리역 주변에 위치해 유동인구가 많아 특히 집중지역으로 확인됨
- 금호2-3가동은 인근에 소공원이 조성되어 있는 금호2-3가주민센터와 금호자이2차 아파트에서 확인됨

171

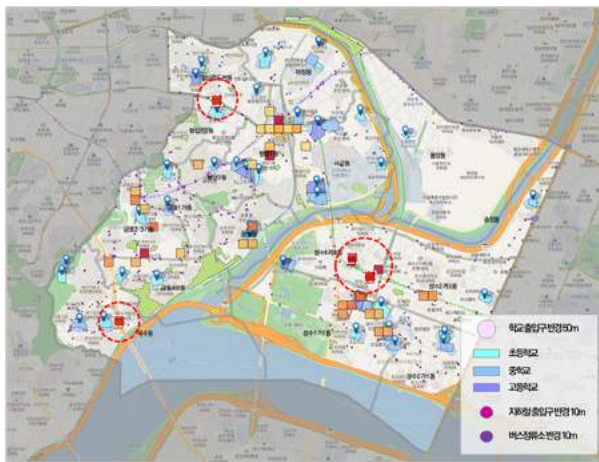


〈어르신 흡연집중지역〉

4. 금연구역 후보지 도출

■ 금연구역 후보지 분석

- 흡연유발지수 상위지역 중 현재 금연구역이 아닌 지역 도출
- 왕십리2동의 상왕십리역 ~ 무학초등학교 일대
- 성수동 서울숲엘타워와 서울숲IT밸리 주변
- 옥수역 / 옥수역교차로 주변



〈금연구역 후보지〉



구분	내용
분류	• 유동인구 밀집 지역
현황	• 상왕십리역 1~11번 출구 샌드박스 인근 • 지하철도구 및 샌드박스 주변 유계공간 연결된 지역 중 연 예상
시사점	• 지하철도구가 있으며 유동인구 밀집 지역임을 고려한 금연구역 지정 필요 • 흡연 단속 공백 최소화를 위해 단속반 투입 필요 및 단속장 종연자에 대한 과태료 부과 • 과태료 부과 내용을 포함한 흡연 단속 경고문 부착

〈금연구역 후보지 상세분석 예시〉

■ 흡연집중지역

- 흡연유발지수 상위지역인 흡연집중지역의 각 유형별 TOP 10
- 흡연유발지수 동점 지역 발생시 TOP 10에서 일부 증감
- 일부 다른 흡연지도와 중복 지역 존재

유형	순번	지역	행정동
전체 흡연지도	1	왕십리 광장~성동구청 사이 일대	행당1동
	2	왕십리역 광장 일대	행당1동
	3	상왕십리역 일대	왕십리2동
	4	옥수역 일대	옥수동
	5	왕십리역 6번 출구~동마중학교 일대	사근동
	6	왕십리 문화공원~왕십리역 이스타빌 일대	행당1동
	7	왕십리 민자역사 일대	행당1동
	8	무학여고 앞 정류장~ 무학여고 사거리 일대	행당2동
	9	한양시장 일대	사근동
	10	왕십리역 성동 경찰서 앞 3,7번 출구 일대	행당1동
청소년 흡연지도	1	금호 푸르지오 아파트~금남시장 일대	금호2,3가동
	2	동마중학교~왕십리역 6번출구 일대	사근동
	3	경일고 뒷편서울숲M타워~장미아파트 일대	성수1가1동
	4	신금호역3번 출구~금호초등학교 일대	금호2,3가동
	5	응봉역1번 출구~광희중학교 일대	응봉동
	6	경일고등학교~동부공원 일대	성수1가1동
	7	왕십리역 광장 일대	행당1동
	8	응봉역~응봉초등학교 일대	응봉동
	9	경동초등학교 ~ 경일중학교 사이	성수1가1동
	10	금호역남지빌딩 삼거리~구립 금강경로당 일대	금호4가동
	11	옥수역 일대	옥수동
	12	경동초등학교 일대	성수1가1동
	13	성수공업고등학교 일대	성수2가1동
직장인 흡연지도	1	코오롱디지털 타워 3차~서울숲IT밸리 일대	성수1가2동
	2	뚝섬역2번 출구~서울숲엘타워 일대	성수1가2동
	3	뚝섬역5번 출구~서울숲A타워 일대	성수1가2동
	4	서울숲역~장미아파트 사이 왕십리로 2길	성수1가1동
	5	서울숲역 지식산업센터 일대	성수1가1동
	6	서울 뚝섬 우체국~현대그린아파트 일대	성수1가1동
	7	성수역 2,3번 출구 일대	성수2가3동
	8	성수역 제일빌딩~영동 테크노 타워 일대	성수1가2동
어르신 흡연지도	1	성동 문화광장 일대	행당1동
	2	금호자이2차아파트~금호동 성당 일대	금호2,3가동
	3	금호역2,3번 출구 일대	금호4가동
	4	서울 응봉초등학교~구립 가람어린이집 일대	응봉동
	5	금호초등학교~금호 2·3가동 주민센터 일대	금호2가동
	6	소월아트홀~대영빌딩 일대	행당1동
	7	행당 대림 아파트 일대	행당2동
	8	금동 어린이 공원~금호 사거리 사이 무수막길 일대	금호2,3가동
	9	신금호역1번 출구~신금호파크자이아파트 일대	금호2,3가동



데이터 분석 단계

1. 데이터 분석 단계 요약

단계	주요 내용
분석 설계	<ul style="list-style-type: none"> ■ 분석용어, 분석공간단위, 분석시간단위 정의 ■ 분석시나리오 구성과 분석주제 설계 ■ 분석주제별 분석방법, 시각화 방안, 활용방안 구성
데이터 수집 및 가공	<ul style="list-style-type: none"> ■ 분석 모델별 필요 데이터 정의 ■ 데이터 보유기관 확인 및 데이터 제공 가능여부 협의 ■ 분석 모델별 데이터 수집 및 가공 ■ 개인정보 해당 필드 삭제
데이터 분석 및 검증	<ul style="list-style-type: none"> ■ 흡연민원 및 단속 과태료 현황 파악 <ul style="list-style-type: none"> • 흡연민원 통계분석: 연도별/행정동별/월별/요일별/민원 유형별 분석 • 흡연단속 과태료 통계분석: 연도별/행정동별/월별/요일별/시간대별 분석 ■ 흡연유발지수 개발 및 연령층별 유형화 <ul style="list-style-type: none"> • 청소년, 직장인, 어르신 연령층별 유형화 및 흡연유발지수 개발 • 흡연유발지수 분포를 통해 연령층별 흡연집중지역 도출 ■ 금연구역 후보지 도출 <ul style="list-style-type: none"> • 전체 및 연령층별 흡연집중지역을 확인 • 흡연집중지역 중 현재 금연구역이 아닌 지역을 금연구역 후보지로 도출
데이터 시각화	<ul style="list-style-type: none"> ■ 통계 분석 차트 시각화 ■ GIS기반 분석결과 시각화



2. 분석설계

■ 분석목적에 맞는 결과를 도출하기 위해서 분석목적, 분석시나리오, 분석주제, 분석 방법을 기획

과제	<ul style="list-style-type: none"> ■ 분석용어, 분석공간단위, 분석시간단위 정의 ■ 분석시나리오 구성과 분석주제 설계 ■ 분석주제별 분석방법, 시각화 방안, 활용방안 구성
추진내용	<ul style="list-style-type: none"> ■ 분석용어 정의 <ul style="list-style-type: none"> • 흡연지도: 흡연유발지수가 높은 지역을 연령층 별로 분류하여 분석한 지도 • 흡연유발지수: 흡연 민원 및 단속 과태료 현황 데이터와 주거인구, 유동 인구, 시설 데이터를 활용하여 흡연 발생 가능성을 산출 • 흡연집중지역: 흡연유발지수가 높게 나타난 지역 • 연령층: 청소년(10대), 직장인(20대~50대), 어르신(60대 이상) ■ 분석공간단위 정의 <ul style="list-style-type: none"> • 성동구 관내 및 국토교통부 표준격자(100m*100m) ■ 분석시간단위 정의 <ul style="list-style-type: none"> • 2016년 ~ 2018년(3개년)
Tip	<ul style="list-style-type: none"> ■ 분석 설계는 분석을 시작하기 전에 전반적인 과정을 계획하는 단계로 구체적인 분석 설계를 통해 분석목적에 맞는 세부 분석 주제를 분석 시나리오에 맞게 설계 ■ 분석 설계 과정에서 현업부서 및 활용 담당자가 필요한 형태로 분석결과를 도출할 수 있도록 충분한 협의 필요

3. 데이터 수집 및 가공

■ 분석주제별 필요한 원시데이터(Raw Data)를 보유 기관에 공급요청 및 수급

■ 수급한 원시데이터(Raw Data)는 분석단위에 맞는 형태로 정제 및 가공

과제	<ul style="list-style-type: none"> ■ 분석 모델별 필요 데이터 정의 ■ 데이터 보유기관 확인 및 데이터 제공 가능여부 협의 ■ 분석 모델별 데이터 수집 및 가공 ■ 개인정보 해당 필드 삭제
추진내용	<ul style="list-style-type: none"> ■ 분석 모델별 필요 데이터 정의 ■ 데이터 보유기관 확인 및 데이터 제공 가능여부 협의 <ul style="list-style-type: none"> • 내부 수급 가능 데이터는 담당 부서와 협의하여 보유 데이터를 확인하고 공문을 통해 수급 • 공공데이터는 각 포털을 통해 수집 ■ 분석 모델별 데이터 수집 및 가공 <ul style="list-style-type: none"> • 수집한 데이터는 분석 단위 및 목적에 맞는 데이터가 존재하는지 검증 작업을 수행한 후 분석 모델별 데이터 구성 • 주소형태로 된 데이터 필드는 POINT 좌표 형태로 지오코딩 • 본 분석은 국토교통부 표준격자(100m*100m)로 분석공간단위를 설정하여 모든 데이터는 국토교통부 표준격자를 기준으로 가공 ■ 개인정보 해당 필드 삭제 <ul style="list-style-type: none"> • 민원 정보나 과태료 정보는 담당 부서에 데이터 요청 시 긴밀한 협조를 통해 개인정보 필드 삭제 후 공급 및 수급
Tip	<ul style="list-style-type: none"> ■ 데이터 수집 전 필요 데이터와 해당 필드를 정확히 파악한 후 요청하는 것이 필요 ■ 개인정보는 개인을 식별할 수 없도록 개인정보 필드를 삭제하여 공급 및 수급이 될 수 있도록 담당부서와 협의 필수



4. 데이터 분석 및 검증

■ 분석단위에 맞는 형태로 가공된 데이터를 통계분석 혹은 공간분석도구를 활용하여 분석하고 검증

과제	<ul style="list-style-type: none"> ■ 분석설계 단계에서 수립한 분석주제별로 분석 수행 ■ 분석 결과에 맞는 검증 실시 ■ 분석 산출물에 대한 정의
추진내용	<ul style="list-style-type: none"> ■ 흡연 민원지역 통계분석 <ul style="list-style-type: none"> • 흡연 민원 통계분석: 연도별/동별/월별/요일별/민원유형별 흡연민원 발생 건수 • 흡연 단속 통계분석: 연도별/동별/월별/요일별/시간별 흡연 단속 발생 건수 ■ 흡연 민원지역 hotspot 분석 <ul style="list-style-type: none"> • 연도별 흡연 민원 hotspot 분석 • 시간대별 흡연 민원 hotspot 분석 • 연도별 흡연 과태료 hotspot 분석 • 흡연 지도 분석 • 전체 흡연 지도 분석 • 청소년 흡연 지도 분석 • 직장인 흡연 지도 분석 • 어르신 흡연 지도 분석 • 금연구역 및 흡연지역 현황 분석 • 금연구역 후보지 분석
Tip	<ul style="list-style-type: none"> ■ 수집 데이터의 품질검증은 해당 데이터 담당자와의 협업을 통해 오류파악을 하는 것이 효과적 ■ 분석 결과 검증 시 협업 부서 담당자와 협업을 통해 통찰(insight)을 도출 하는 것이 효과적

5. 데이터 시각화

■ 분석결과를 바탕으로 분석주제를 가장 효과적으로 표현할 수 있는 시각화 방안 구성

과제	<ul style="list-style-type: none"> ■ 통계 분석 차트 시각화 ■ GIS기반 분석결과 시각화
추진내용	<ul style="list-style-type: none"> ■ 흡연 민원지역 통계분석 차트 시각화 <ul style="list-style-type: none"> • 흡연 민원 통계분석: 연도별/동별/월별/요일별/민원유형별 흡연민원 발생 건수 • 흡연 단속 통계분석: 연도별/동별/월별/요일별/시간별 흡연 단속 발생 건수 ■ 흡연 민원지역 hotspot 분석 GIS 시각화(shp) <ul style="list-style-type: none"> • 연도별 흡연 민원 hotspot 분석 • 시간대별 흡연 민원 hotspot 분석 • 연도별 흡연 과태료 hotspot 분석 ■ 흡연 지도 분석 GIS 시각화(shp) – 국토교통부 표준격자 (100m*100m) 단위 <ul style="list-style-type: none"> • 전체 흡연 지도 분석 • 청소년 흡연 지도 분석 • 직장인 흡연 지도 분석 • 어르신 흡연 지도 분석 • 금연구역 및 흡연지역 현황 분석 • 금연구역 후보지 분석
Tip	<ul style="list-style-type: none"> ■ 연도별 연계분석을 위해서는 오픈소스 GIS 소프트웨어를 사용하는 것을 권장



바 행정적용 단계

1. 행정적용 단계 요약

업무 적용	<ul style="list-style-type: none"> ■ 관내 금연구역 및 흡연집중지역 현황 관리 ■ 연도별 흡연 관련 민원 및 단속 현황 관리 ■ 시간대별 흡연 단속원 배치 ■ 청소년 금연 계도 활동 경로 참고 ■ 어르신 금연 계도 활동 지역 참고 ■ 금연구역 후보지 우선순위 목록 참고 ■ 우리구 흡연지도 정책지도화
단계적 확대	<ul style="list-style-type: none"> ■ 연도별 관내 금연구역 및 흡연집중지역 현황 관리 ■ 연도별 흡연 관련 민원 및 단속 현황 관리 ■ 연도별 금연구역 후보지 우선순위 목록 관리 ■ 연도별 우리구 흡연지도 정책지도화 ■ 간접흡연 위험지역에 대한 예측정보 활용으로 금연관리 효율성 제고 ■ 비흡연자와 흡연자가 공존할 수 있는 도시공간 생태계 기반 마련
성과 측정	<ul style="list-style-type: none"> ■ 정책 개발에 활용되어 주요 정책 개발 및 개선 내용 건수 ■ 흡연 관련 민원 증감 성과 측정 ■ 흡연 과태료 단속 증감 성과 측정



사례집을 만든 사람들

한국지역정보개발원

클라우드기획부 김유석 부장·예광호 수석·강민선 선임

지방자치단체 및 외부 전문가

부산광역시

조기행 통계빅데이터담당관

제주특별자치도

노희섭 미래전략국장·김기홍 디지털융합과장·박기범 빅데이터팀장

서울특별시 성동구

강성호 빅데이터센터장·원준호 주무관·이현정 주무관

펜타시스템테크놀로지 (주)컨소시엄

김동찬 책임·윤서현 사원·정대진 주임·최재혁 주임

❖ 본 사례집의 내용 중 빅데이터 분석 추진 공무원의 의견 등은 개발원의 공식적인 입장이 아님을 밝힙니다.

2019 지방자치단체 빅데이터 분석 사례집

발 행 2020년 1월
발행처 한국지역정보개발원
발행인 지 대 범

사례집의 무단전재를 금하며,
가공·인용할 때에는 반드시 「한국지역정보개발원,
2019 지방자치단체 빅데이터 분석 사례집」이라고 밝혀주시기 바랍니다.

내용에 관한 문의나 제안은 아래의 연락처로 해주시기 바랍니다.

- 주소 : 03923, 서울특별시 마포구 성암로 301, 한국지역정보개발원
- 전화 : 02-2031-9100
- 팩스 : 02-2031-9360

사례집은 한국지역정보개발원 홈페이지(www.klid.or.kr) 정보화자료실 -
정보화연구자료에서 보실 수 있습니다.

본 사례집의 내용 중 빅데이터 분석 추진 공무원의 의견 등은
한국지역정보개발원의 공식견해와 다를 수 있습니다.

간행물 등록번호
B551982-2019-000014-10

