

데이터기반 행정으로 국민의 삶의 질을 개선하라!

데이턴 십 해커 톤 제4회

서울시 불법주정차 감소를 위한
유희공간 활용 및 공영주차장 추가 설치 입지 분석

분석 결과보고서

참여15조 : 파킹프렌즈(경인권)

참여자 : 이은지

임지수

지하연

한규원

목 차

1. 분석 개요	3
1.1. 분석 배경 및 개요	3
1.2. 분석 목적 및 방향	8
1.3. 분석 결과 활용 방안	9
2. 분석 데이터	11
2.1. 분석 데이터 목록.....	11
2.2. 데이터 상세 설명.....	12
2.3. 데이터 정제 방안.....	14
3. 분석 프로세스.....	15
3.1. 분석 프로세스	15
3.2. 분석 내용 및 방법	16
4. 분석결과.....	20
5. 활용 방안	46
5.1. 문제점 개선 방안	46
5.2. 업무 활용 방안	46
6. 참고자료	47
7. 부록.....	49

1. 분석 개요

1.1. 분석 배경 및 개요

1.1.0 분석 개요

우리나라의 인구 밀집도는 전 세계 3 위에 해당되며 이에 따라 자동차 등록 대수는 계속해서 증가하고 있다. 이는 다양한 사회적 문제를 불러일으킬 수 있는 충분한 조건이며, 그 중 불법주정차는 가장 큰 문제로 인식되고 있다.

현재 불법주정차를 감소시키기 위한 여러 정책들이 펼쳐지고 있지만, 근본적인 주차문제를 해결하지 못하고 있다.

이를 해결하고자 서울시의 주차공간 확보에 초점을 맞춰 단기적 측면과 장기적 측면의 해결방안을 강구하였다.

단기적 측면으로 부설주차장 개방 정책과 이용을 활성화시켜야 하는 타당한 이유를 도출하고, 장기적 측면으로 공영주차장의 주차장 확보율을 늘리기 위해 입지 분석을 통한 최적 입지 후보지를 선정이 주된 목적이다.

위 방안들을 토대로 불법주정차를 감소시킴과 동시에 주차장 확보율을 늘린다면, 시민들에게 지금보다 더 나은 편의생활을 할 수 있도록 제공할 수 있을 것으로 판단된다.

1.1.1 교통안전을 위협하는 불법주정차

- 불법주정차는 교통흐름을 방해해 도로의 혼잡을 야기하고, 운전자의 시야를 막아 보행자와 교통약자의 안전을 위협한다.
- 원활한 보행 및 차량 통행을 위해 주차 금지구역으로 지정한 곳에도 버젓이 주차를 하고 있음.

- 양방향 통행이 가능한 도로면도 한쪽면이 막힌 탓에 차량들은 일방 통행로처럼 좁아진 도로. 운전자들의 불편 호소.
- 소방차, 구급차 등 긴급차량 진입 방해로 인해 골든 타임 놓침.
- ‘민식이 법’ 이후에도 스쿨 존이나 어린이구역의 불법주정차 여전.
‘초등학교’, ‘유치원’, ‘어린이 집’ 위주로 데이터를 확보하고 조사한 결과, 매년 6 만건을 약간 밑도는 수준으로 크게 늘거나 줄지 않고 꾸준함.

※ '민식이 법' 시행에 따라 정부는 2022 년까지 전국의 모든 스쿨 존에 무인 교통 단속장비와 신호등을 100% 설치하겠다고 밝혔지만, 아직 10%밖에 설치되어 있지 않음. (2020 년 6 월 기준).

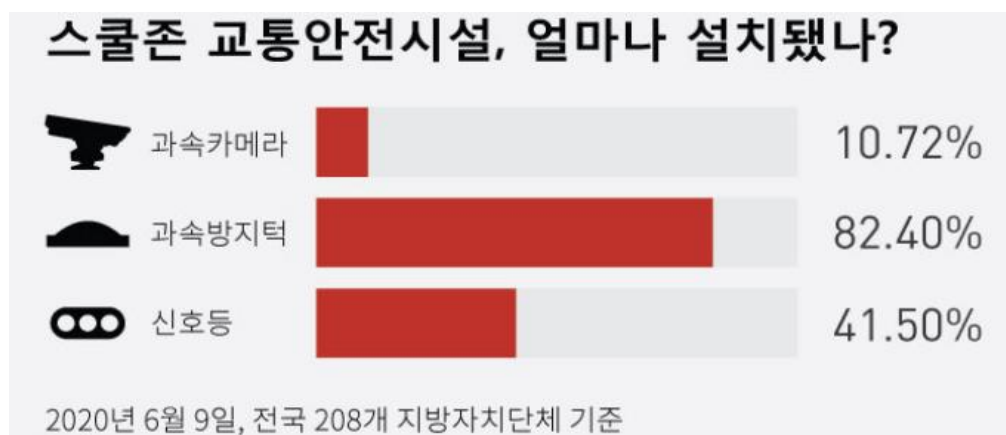
뉴스 > 경제

불법주정차 피하다 '깡' 3년간 4천7백 건...과태료는 단 4만 원

이병주 기자 | 기사입력 2021-07-25 19:30 | 최종수정 2021-07-25 20:30

0

<그림 1.출처: MBN 종합뉴스>



<그림 2. 출처: SBS NEWS>

1.1.2 자동차 등록현황

전국 자동차 등록대수 2000 만 시대, 서울특별시(이하, 서울시)도 이미 300 만을 돌파한 지 오래전이다.

구분	2015	2016	2017	2018	2019	2020
자동차등록대수	3,057	3,083	3,116	3,124	3,124	3,157
자가용 승용차	2,437	2,475	2,514	2,527	2,539	2,940

<표 1. 자동차 등록현황(출처: 서울 특별시 교통)>

(2020 년 12 월 31 일 기준, 단위: 천대)

- 고질적인 사회문제로, 자동차 수의 증가함에 따라, 도심의 불법주정차 문제는 경제적, 사회적으로 막대한 비용이 들고 있으며, 비단 서울시만의 문제가 아님.
- 주차공급확대와 이의 활용방안에 대하여 다양한 정책들을 시도하며 문제 해결에 많은 노력을 기울이고 있음.

1.1.3 불법주정차로 인한 사회적 문제점

불법주정차로 인한 다양한 사회적 문제들이 유발되었고, 이를 해결할 방안의 필요성이 점차 증대되고 있다.

- 전국 주차장 공급량은 지속적으로 증가 추세이지만, 주차시설과 주차수요 간 시공간적 불균형으로 인하여 주차수요와 공급의 편차가 계속해서 증가하는 지역 발생.
 - 주차장 공급량의 대부분은 부설주차장(전체 대비 약 92%)이며, 주거지역에서 가깝고 주차가 상대적으로 용이한 노상, 노외주차장의 공급비중은 크지 않다.

- 주택가 및 상업 밀집 지역 이면도로 등에서 불법주정차로 인해 운전자 및 보행자의 시야를 제한하여 안전 및 불편을 초래하고, 도로교통 혼잡 유발.
 - 20 년 기준 불법주정차로 인한 유발사고 건수는 569 건으로 전년 대비 약 **41.5%** 증가하였고, 올 상반기는 492 건으로 작년 같은 기간 대비 **86.5%** 증가했다. 특히 어린이 보호구역에서 **어린이 교통사고**가 증가하고 있다.
- 부족한 주차공간으로 인한 이웃과 계속된 갈등으로 인해 주민의 공동체 의식과 생활 편의성 악화.
 - 불법주정차 단속 강화를 위한 주민신고제 운영 보름만에 신고 건수가 기존의 3배를 초과했다. 불법 방지 목적의 제도가 오히려 사회적 불신과 갈등을 부추겨 또 다른 문제를 야기한다는 지적이 나온다.
- 불법주정차로 인해 발생하는 **사회적비용**(교통혼잡비용, 사고발생비용, 긴급차량 길 막음 위험비용, 배회주행 처리비용 등) 발생.
 - 서울시의 불법주정차로 인한 사회경제적 비용은 연간 4조 8970억원(삼성교통안전문화연구소, 2016). 그 중 국내 대도시에서 주차를 위해 배회하고 대기하는 차량으로 인한 상습적인 정체가 사회적 문제로 지적되고 있다.
 - 긴급 출동 시 골든 타임을 놓쳐 발생하는 인명 및 재산피해 등 이 증가하는 추세이다.
- 도심 지역의 시간에 따른 유동 인구 편차로 인해 탄력적인 조정에 어려움이 존재. 이러한 한계 극복을 위해 기존주차장의 효율적 운영과 새로운 주차장 신설에 대한 필요성이 대두.
 - 주차장 규모를 평균 주차 수요량에 맞추게 되면 차량의 도로 점유율이 높은 시간에 주차공간 부족으로 인한 문제가 발생하게 되는 반면, 점유율이 높은 시간대를 위주로 주차 수요량을 맞춰 주차공간을 확보하게 되면 점유율이 낮은 시간대에 주차공간이 과도하게 비어 있게 되어 비효율 문제가 발생한다.

1.1.4 공영주차장 사업 추진 현황

대도시 도심지역은 토지 가격이 높고 주차장으로 이용 가능한 공간적 여유가 적다. 따라서 물리적으로 주차장을 확보하기 위해서는 막대한 예산이 소요되며, 추가적으로 확보할 수 있는 주차면도 한정적이다.

또한 도심지역은 유동인구가 많고 시간에 따른 유동인구 편차 또한 크게 나타나기 때문에 주차장의 공급량을 탄력적으로 조정하기 어려운 실정이다.

- 지자체별로 불법주정차 문제를 해결하고자 여러 대안책을 강구해왔지만, 일회성 혜택에 불과.
- 근본적인 문제인 주차시설 부족과 주차수요간 시공간적 불균형을 해결하지 않는 이상 앞으로 더욱 큰 사회적 비용을 초래할 것으로 예상.
- 자동차등록수가 매년마다 증가하지만, 서울시에서 주차장 사업을 확대할 수 있는 면적은 제한적.
- 공영주차장 사업 확대.
(위 정책에는 토지매입과 공사비 등 막대한 예산이 추정)
(주차 면 1칸 당 1억. 2018년 기준.)
- 막대한 예산을 투자한들 기대 수익을 거둘 수 없는 실정.

1.1.5 불법주정차 피해 사례

- 주차된 화물차 들이받고 숨겨... 불법주차 화물차주에 ‘실형’ 선고
(2021 년 2 월)
: 충남 천안의 한 대로변에서 SUV 차량이, 세워져 있던 화물차를 들이받으며 이 사고로 차에 타고 있던 4 명 중 2 명이 숨짐. 화물차가

주차된 구간이 황색 복선(2 줄)으로 표시된 주차 금지 구간이었기 때문에, 화물차주를 업무상 과실치사 혐의로 불구속 입건.

- 불법주정차로 인한 소방활동 지연 사례
: ‘서울 홍제동 주택 화재사건’ (2001 년 3 월 4 일)
- 주차공간이 협소하여, 불법주정차로 인한 구조활동 지연으로 인하여 화재사건으로 현장에 출동했던 소방관 6 명이 순직하는 참사 발생.

1.2 분석 목적 및 방향

현재 불법주정차와 관련된 문제를 해결하기 위한 여러 방안들이 이미 시행 중이다. (주차단속 알림 시스템, 과태료 인상, 주민신고제, 단속 관련 규제 수립 및 수정 등) 하지만 감소세는 여전히 미미하므로, 근본적인 주차 문제를 해결하기 위해 새로운 주차장 확보에 초점을 맞춰 주차환경을 개선하는 방향으로 목적으로 두었다.

기존 공영주차장 및 단속 CCTV 위치 등 기존에 시행되고 있는 정책에 추가적으로 공영주차장 확충과 유휴공간을 부설주차장으로 활용한다면 시너지 효과가 극대화될 것으로 예상된다.

서울시 자치구별, 연도별, 불법주차장 단속건수, 부설주차장, 노상·노외주차장, 민원 행정기관 등 여러 요인들을 분석해 신설 공영주차장의 최적 입지를 선정해 주차장 확보가 주 목표이다.

1.2.1 분석 목적 및 방향

- 단기적 측면: 우리 주위에 아직 활용되지 않은 잠재된 장소와 유휴 공간을 활용하여 궁극적인 주차장 확보율을 높여 주차장 효율 극대화.

- 장기적 측면: 불법주정차가 많이 일어나는 지역 특성을 분석하여 새로운 공영주차장 신설이 필요한 지역들의 우선순위를 구하여 최적 입지 선정.

※ 서울시의 주차장 확보율은 130%를 상회하지만 불법주차 문제 해결되지 않고 있음. 차량 1대는 출발지와 목적지를 가지고 있으므로 1대의 차량은 최소 2개의 주차장의 수요를 발생시키므로 이상적인 최소 주차장 확보율은 200%

1.3. 분석 결과 활용 방안

1.3.1 유휴공간 활용 및 공유 사업 추진(주민센터, 경기장, 공원, 학교 등 민원행정기관 위주)

서울의 기존 공영주차장은 이미 한정되어 있으며, 공영주차장을 새로 확충하기 위한 제약사항이 많다.

- 그린벨트를 제외하고 유휴지 확보가 어려운 도시로서 주차장 확보를 위해 필요한 공간 확보에 명확한 한계가 있기 때문에, 기존의 주차장의 이용 효율을 활성화시킬 필요가 있음.
- 부설주차장개방으로 정부 지원 사업을 실시하고 있지만, 아직 부설 주차장 사업에 참여하고 있지 않은 잠재적 사업체나 장소가 많음.
- 매년 주택가 인근 부설주차장을 개방하기 위해 주차장 시설 개선 공사로 최고 25 백만원까지 지원하고 있지만 넉넉치 않은 예산의 한계로 매년 감소중인 상태.

※ 부설주차장 확대를 위해 정책 변경 제안하는 방안.

1.3.2 공영주차장 앱 사업 계획 추진

기존의 앱과 차별화된 유희공간 위치서비스와 공영주차장 위치서비스를 제공함과 동시에 서울시 주차관련 민원을 통해 토대로 추가적인 기능을 제공할 계획

● 기존 앱 한계점

- 공영 주차장의 정보를 실시간으로 누구나 볼 수 있다는 장점이 있지만, 다운로드 및 API 공개방식으로 제공하기 때문에 형식적으로 공개할 뿐, 기능을 다루지 못하는 사람에게는 제한적일 수 있음.
- 기존의 서울시 주차정보안내시스템 어플리케이션이 존재하지만, 유지보수가 늦어 사용자들에게 높은 만족감을 제공하지 못하고 있음.

● 해결방안

- 다양한 정보와 편의를 제공하기 위해 기존 서울주차정보 어플리케이션에 사람들이 필요로 하는 기능들을 추가해 새로운 서비스를 제공하는 방식으로 접근.
- 기존의 유희공간 위치서비스와 공영주차장 위치서비스를 제공함과 동시에 서울시 주차관련 민원데이터의 워드 클라우드와 빈도분석을 통해 토대로 추가적인 기능을 제공할 계획.

1.3.3 공영주차장 최적 입지 선정

장기적 관점으로 볼 때, 서울시 공영주차장 확대의 필요성은 기정 사실화. 정부에서도 지속적으로 서울시 자동차 등록 수 현황을 파악하고 있으며, 국민들의 편의를 위해 주차장 지원 사업은 절대적으로 필요한 실정이다.

- 앞서 말했듯이 현실적으로 당장 해결할 수 있는 문제는 아니지만, 주차장 설치 금지지역을 제외하고 필요한 면적 확보가 우선시 되어야함.
- 여러 제약 사항들 때문에 지금 당장 사업을 추진할 수 없지만, 사용 가능한 지역들을 미리 파악하고 선점해 놓는 것이 궁극적인 목적.

2. 분석 데이터

2.1. 분석 데이터 목록

- 서울시 구별 위치 데이터
- 서울시 동 별 단속건수
- 서울시 동 별 주차장 통계
- 서울시 동 별 30 년 이상 노후주택 현황
- 서울시 버스 정류소 좌표 데이터
- 서울시 지하철 정류장
- 민원행정기관전자지도
- (경기도)2014 년주정차단속현황.xlsx
- (경기도)2015 년주정차단속현황.xlsx
- (경기도)2016 년주정차단속현황.xlsx
- (경기도)2017 년주정차단속현황.xlsx
- (경기도)2018 년주정차단속현황.xlsx
- (경기도)2019 년주정차단속현황.xlsx
- 21 년 1 월 기준 서울시 자동차 등록현황(동 별, 연료 별).xlsx
- 2014-2019 의왕, 파주, 포천 부설주차장 데이터.xlsx
- 민주주의 서울 <https://democracy.seoul.go.kr/front/index.do>

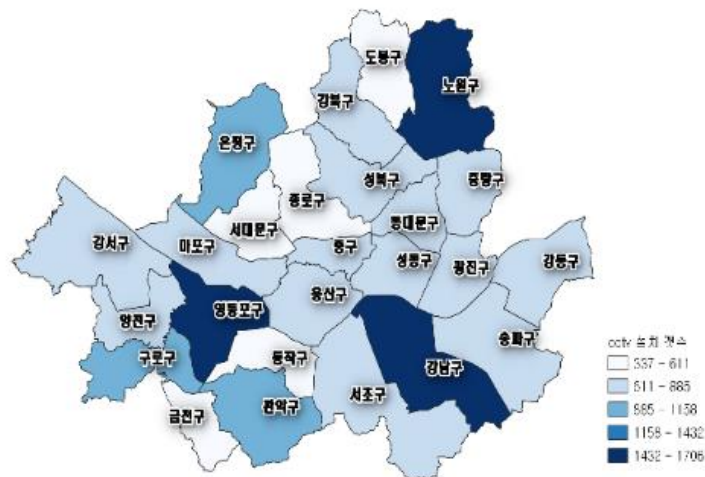
2.2 데이터 상세 설명

활용데이터	구분	중요도	생성주기	데이터소스
서울시 주차장 공간데이터	정형/내부	필수	1년	주차장명, 시간제주차_기본단위 요금, 주차요금 지불방식, X좌표,Y좌표
CCTV위치정보	정형/내부	필수	분기 1회	위도, 경도, 자치구, 단속지점명, 현장구분
도서관위치도	정형/내부	필수	1년	도서관명, 설명, 면좌표, 선좌표, 구명, 동명
문화시설 위치도	정형/내부	필수	1년	문화시설명, 설명, 면좌표, 선좌표, 구명, 동명
주민자치센터 위치도	정형/내부	필수	1년	주민자치센터명, 설명, 면좌표, 선좌표, 구명, 동명
학교 공간데이터	정형/외부	필수	1년	교명, 주소, 전화번호, 위도, 경도
공공체육시설 공간데이터	정형/내부	필수	1년	체육시설명, 설명, 면좌표, 선좌표, 구명, 동명
구별 불법주정차 단속건수	정형/내부	필수	1년	연도별 단속건수
서울시 시,군구 위치 데이터	정형/내부	필수	1회성	구이름, 인덱스(고유값)
서울 지하철 역 위치데이터	정형/내부	선택	1년	역명, 주소, X좌표, Y좌표
경기도 주,정차 단속현황	정형/내부	선택	1년	시, 군, 명, 월별 단속현황
서울시 자동차 등록현황	정형/내부	선택	정기(매년)	승용차, 승합차, 화물자동차, 특수 자동차
서울시 노후 주택수	정형/내부	선택	수시	주택수, 비율, 순위
20년 동별 자동차 면수	정형/내부	선택	1년	노상주차장 개소, 면수, 노외주차장 개소
민주주의서울 사이트	비정형/외부	선택	실시간	서울시 실시간 시민의견 정보

<표 2. 데이터 상세설명>

● 샘플분석

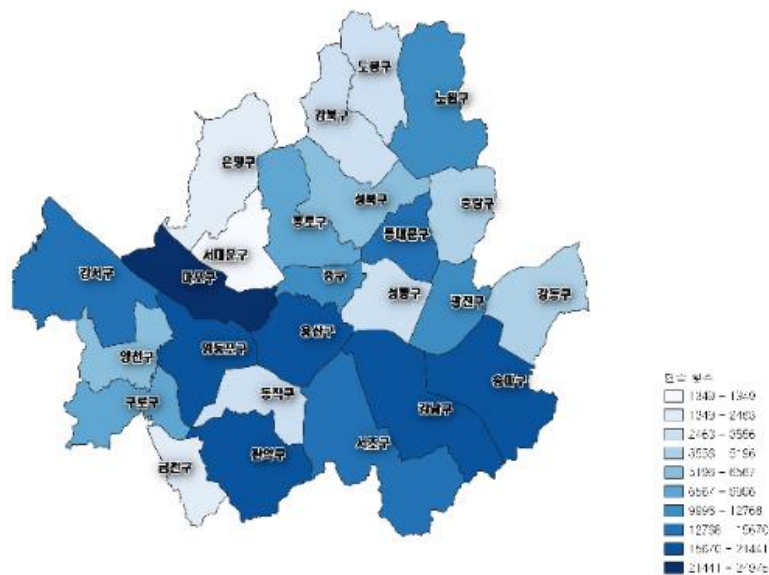
<2016 년 CCTV 설치 개수.SHP>



<그림 3.CCTV 설치 개수.SHP>

- 서울지역 CCTV 설치 개수를 파악해 향후 공영주차장 추가 증설을 위한 단속 CCTV 개수가 불법주정차에 얼마나 영향이 가는지를 파악해 최적입지선정에 중요한 변수로 사용할 예정.

<2016 년 구별 단속 횟수.SHP>



<그림 4.구별 단속 횟수.SHP>

- 단속횟수를 종속변수 y 로 두어 회귀분석을 진행하기위해 미리 구별단속 횟수를 샘플 분석.

2.3. 데이터 정제 방안..... 00

2.3.1 민주주의 서울 <https://democracy.seoul.go.kr/front/index.do>

- 크롤링하여 민원 데이터를 문장형태로 가져온 다음 데이터를 Tokenizer 를 활용하여 전처리.

[13]:	NO	tokenizer
0	안녕하세요 오세훈시장님 취임하심에 진심으로 축하드립니다저는 양천구 신정동에 거주하는...	[안녕하세요, 오세훈시장님, 취임하심에, 진심으로, 축하드립니다저는, 양천구, 신정...
0	노상흡연이 빈번한 곳에 흡연 부스 설치와 기타 설치 사항을 제안하고 싶습니다 작은 ...	[노상흡연이, 빈번한, 곳에, 흡연, 부스, 설치와, 기타, 설치, 사항을, 제안...
0	서울시내버스 시지원금지노인유료 각종버리 배불리기 텔레비에도 버리방영 시내버스 자율운...	[서울, 시지원금지노인유료, 각종버리, 배불리기, 텔레비에도, 버리방영, 시내버스,...
0	종로구에는 노인분들이 많이 사십니다그러다보니 집안의 자질구래한 일든죽 전기고장수도고...	[종로구에는, 노인분들이, 많이, 사십니다그러다보니, 집안의, 자질구래한, 일든죽,...
0	안녕하세요 코로나단계 인이상집합금지예대한 규제로 동거가족과 식당을이용하는경우현제주민...	[안녕하세요, 코로나, 인이상집합금지예대한, 규제로, 동거가족과, 식당을이용하는경우...
---	---	---
0	요즘 각 가정에서는 대부분 계란을 깨틀이를 사서 먹습니다계란을 사오면 바로 등에 담...	[요즘, 각, 가정에서는, 대부분, 계란을, 깨틀이를, 사서, 먹습니다계란을, 사오...
0	제목 격리해제자 퇴소시간 조정 요청 서울시 발전을 위한 귀 부서의 노고에 감사를 ...	[제목, 격리해제자, 퇴소시간, 조정, 요청, 서울, 발전을, 위한, 귀, 부서의,...
0	대형마트 주말 영업 제한을 풀어주세요	[대형마트, 주말, 영업, 제한을, 풀어주세요]
0	광화문광장 정비 사업의 일환으로 기존 인도도 정비를 하고 계시는데요 동화면세점 앞 ...	[광화문광장, 정비, 사업의, 일환으로, 기존, 인도도, 정비를, 하고, 계시는데요,...
0	매일 아침 강변도로를 달리면서 상쾌하고 행복하게 출근을 하는 운전자 입니다즐거운 월...	[매일, 아침, 강변도로를, 달리면서, 상쾌하고, 행복하게, 출근을, 하는, 운전자,...

<그림 5.Tokenizer 활용 전처리>

2.3.2 부설주차장 사업 추진 진위여부 회귀분석 (단기적 측면 정제 방안)

- 구별 불법주정차 단속건수, 노상주차장 수, 노외주차장 수, 부설주차장 수, CCTV 설치수, 자동차 등록대수, 등록 인구수, 노후 주택수 데이터 수집.
- 결측 값을 평균값으로 대체.
- EDA 로 데이터 분포 확인.
- 각 데이터의 년도와 자치구 순서를 통일시켜 하나의 데이터프레임 생성.

2.3.3 공영주차장 최적입지선정 과정 (장기적 측면 정제 방안)

- 서울시 노상 노외 주차장 공간데이터, 구별위치데이터, 민원행정기관전자지도 파일을 불러옴.
- 필요 데이터를 수집(서울시 노외주차장 공간데이터, 서울시 노상주차장 공간데이터, 서울시 구별 위치 데이터, 민원행정기관전자지도).
- 수집된 데이터를 분석 목적에 맞는 형태로 전처리 과정.
- Python 을 활용하여 정규화 및 합계 순위 출력.

- folium 을 활용해 최적입지 선정 확인
- QGIS 를 활용해 격자표시 시각화

3. 분석 프로세스

3.1. 분석 프로세스

3.1.1 데이터준비

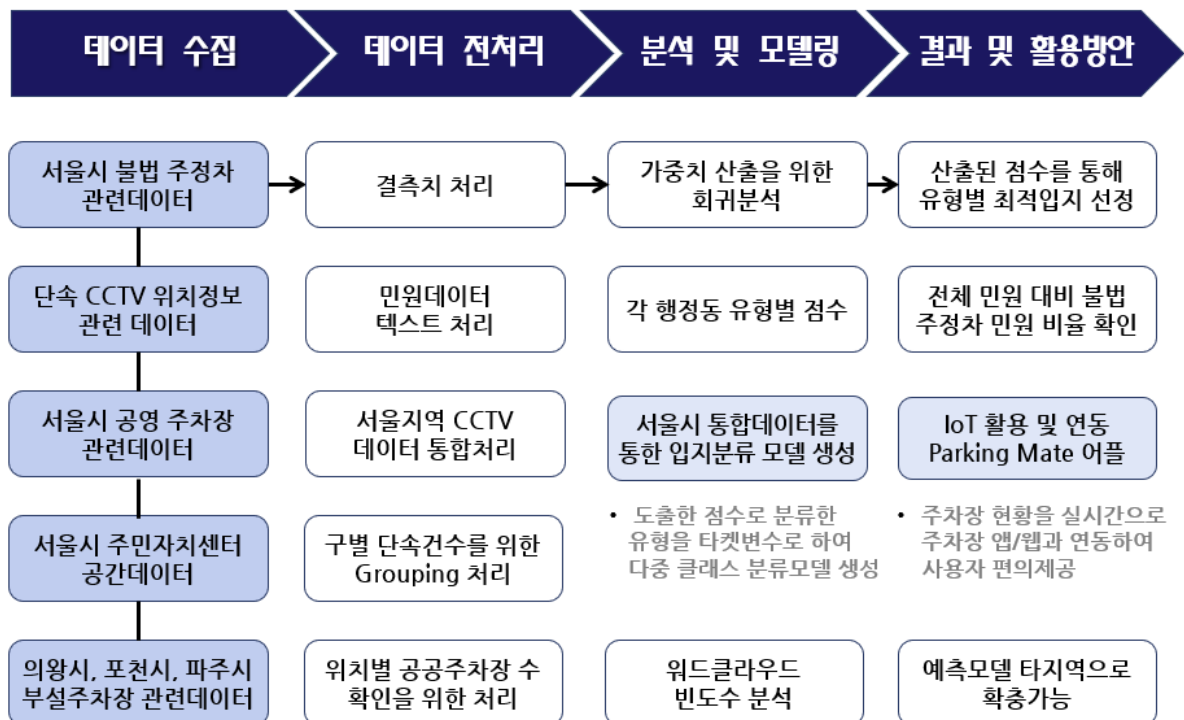
- 단기적 해결방안:

유휴공간 활용을 위한 데이터를 수집/정제하여 사용가능한 데이터 셋을 준비

- 장기적 해결방안:

공영주차장 최적입지를 위한 불법주정차에 대해 영향을 미치는 데이터를 수집 정제하여 분석에 활용

3.1.2 분석 프로세스



<그림6.데이터 분석 과정>

3.1.3 결과 도출 및 시각화

데이터 수집, 전처리 과정을 거쳐 데이터 분석 결과를 시각적으로 표현하고, 공영주차장 확대를 위한 최적입지선정 모델을 개발하여 서울시에 적용.

3.2 분석 내용 및 방법

불법주정차에 대한 심각성을 파악하고자 시민들의 민원을 워드 클라우드와 빈도분석을 실시하고 이를 통해 주차장 확보율과 주차장 위치의 중요성을 알아보고자 한다.

주차장 위치와 확보율이 불법주정차의 건수에 영향을 미침으로, 주차장 확보를 위해 지역적 특성과 여러가지 환경 요인들을 파악한 후, 단기적 측면과 장기적 측면으로 나눠 데이터 분석을 시행한다.

- **단기적 측면**으로 불법주정차단속에 영향을 미치는 요인을 알아보기 위해 회귀분석을 실행하고, 부설주차장의 필요성 여부를 파악하기 위해 주차장 수와 단속건수의 상관분석을 통하여 진위여부를 판단한다.
- **장기적 측면**으로 공영주차장의 최적 입지 선정하기위해 입지결정요인에 점수를 부여하여 우선순위별로 위치를 선정한다.

3.2.1 워드 클라우드 & 빈도수 분석 / 웹 크롤링

민주주의 서울 사이트에서 서울시 국민들의 민원을 데이터화하여 전체 민원중 불법주정차 관련 키워드 빈도분석 및 불법주정차에 대한 인식 파악

- 단기적 측면

3.2.2 주차장 수와 단속건수 관계 파악/상관분석 및 회귀분석

두 변수 간의 관련성 여부를 알기 위해 상관분석과 회귀분석을 실행하여 p-value 값으로 진위여부 판단.

3.2.3 불법주정차 단속건수에 영향을 미치는 변수 파악/회귀분석

불법주정차 단속 건수에 영향을 미치는 변수를 알기 위해 서울시 구별로 회귀분석 실행.

3.2.4 확보율이 높은 구와 낮은 구 현황 확인

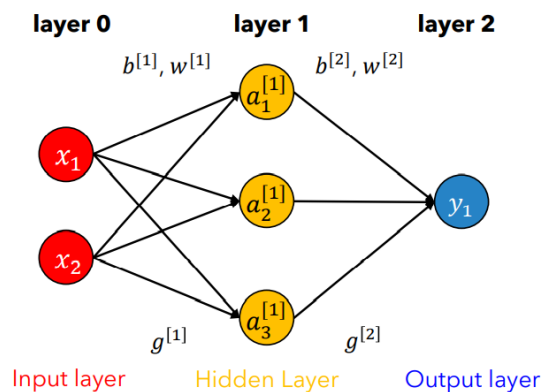
구별로 주차장 확보율을 구한 후, 상위 구 3개와 하위 구 3개를 비교하여 현황파악.

3.2.5 서울시 동 별로 세분화하여 주차장 확보율과 단속 수 관계 파악

동 별로 주차장 확보율 구하고 시간대별 단속건수와 비교.

3.2.6 서울시 동 별로 주차 수요 예측 분석을 통해 예측

주차 수요 예측하는 데에 인공지능망의 한 종류인 다층 퍼셉트론(multilayer perceptron, MLP)을 이용하여 주차 수요를 예측하고자 함.



<그림 7. 다층 퍼셉트론>

3.2.7 타 지역 경기도를 예시로 들어 부설주차장과 불법주정차 단속건수 관계 파악/상관분석

타 지역인 경기도를 예시로 들어 부설주차장이 증가함에 따라 불법주정차가 감소하는지에 대한 진위여부 판단.

3.2.8 부설주차장으로 활용할 수 있는 유휴공간 개수 파악

구 별과 동 별로 나눠 학교 문화시설, 경기장, 민원행정기관 등을 카운트하여 시각화

3.2.9 단기적 결론 도출 및 결론 바탕으로 앱 서비스 출시

유휴공간이 포함된 주차장 앱을 만들어 서비스 제공.

● 장기적 측면

3.2.10 공영주차장 최적입지선정을 위해 데이터 수집 및 전처리

- 데이터 수집

- 면 레이어: 서울시_구별
- 점 레이어: 노상, 노외주차장, 민원행정기관, 도서관, 문화시설, CCTV 단속 위치, 주민자치센터, 체육시설, 학교, 산, 한강
- 좌표계: EPSG: 4326 - WGS 84
- 격자 레이어: MMQGIS 활용하여 서울시 구별 격자 생성
 - Geometry Type: Rectangles
 - X spacing: 300
 - Y spacing: 300
 - Units: Layer Units
 - Extent: Layer Extent
 - Layer: 서울시_구별

● 전처리 과정

- 민원행정기관을 서울지역으로 벡터 -> 조서도구 -> 위치 선택
 - 다음 위치에서 객체 선택: 민원행정기관
 - 객체 위치(도형 서술자): 교차(intersect)
 - 다음과 같은 객체를 비교: 서울시_구별
- 서울 지역 산 별 위치정보를 하나의 데이터로 취합 후 서울시 격자에서 제외.

※ 최적입지선정때, 최적입지가 산중에 입지선정 되는 것을 방지.

- 격자 안에 있는 포인트 개수를 세고 하나의 데이터로 취합
 - 벡터 -> 분석도구 -> 폴리곤 안에 있는 포인트 개수 계산 활용
 - 폴리 건: 서울시_구별 격자
 - 포인트: 노상, 노외주차장, 민원행정기관, 도서관, 문화시설, CCTV 단속 위치, 주민자치센터, 체육시설, 학교
 - 개수 필드이름: 포인트이름
 - 선택된 개체를 FinalFile.csv 형태로 저장

3.2.11 공영주차장 최적입지 선정을 위한 점수계산 및 최적입지 확인

	left	bottom	right	top	lon	lat	학교	도서관	문화시설	주민자치센터	체육시설	민원행정기	cctv	주차장
2	971700	1951200	972000	1951500	127.1813	37.56108	0	0	0	0	0	1	0	0
3	971700	1951500	972000	1951800	127.1813	37.56378	0	0	0	0	0	2	0	0
4	971700	1951800	972000	1952100	127.1812	37.56648	0	0	0	0	0	0	0	0
5	971400	1949400	971700	1949700	127.1779	37.54484	0	0	0	0	0	0	0	0
6	971400	1951200	971700	1951500	127.1779	37.56107	1	0	0	0	0	0	0	0
7	971400	1951500	971700	1951800	127.1779	37.56377	1	0	0	0	0	1	1	0
8	971400	1951800	971700	1952100	127.1778	37.56647	1	0	0	0	0	1	0	0
9	971400	1952100	971700	1952400	127.1778	37.56918	1	0	0	0	0	1	0	0
10	971400	1952400	971700	1952700	127.1778	37.57188	0	0	0	0	0	0	0	0
11	971400	1952700	971700	1953000	127.1778	37.57459	0	0	0	0	0	0	0	0
12	971400	1953000	971700	1953300	127.1778	37.57799	0	0	0	0	0	0	0	0
13	971100	1949400	971400	1949700	127.1745	37.54483	0	0	0	0	0	0	0	0
14	971100	1949700	971400	1950000	127.1745	37.54754	0	0	0	0	0	1	1	0
15	971100	1950000	971400	1950300	127.1745	37.55024	0	0	0	0	0	0	2	0
16	971100	1950300	971400	1950600	127.1745	37.55295	0	0	0	0	0	0	0	0
17	971100	1952100	971400	1952400	127.1744	37.56917	0	0	0	0	0	0	0	0
18	971100	1952400	971400	1952700	127.1744	37.57187	0	0	0	0	0	0	0	0
19	971100	1952700	971400	1953000	127.1744	37.57458	0	0	0	0	0	0	0	0
20	971100	1953000	971400	1953300	127.1744	37.57728	0	0	0	0	0	0	0	0
21	971100	1953300	971400	1953600	127.1744	37.57999	0	0	0	0	0	0	0	0
22	971100	1951200	971400	1951500	127.1745	37.56106	0	0	0	0	0	0	0	9
23	971100	1951800	971400	1952100	127.1744	37.56647	0	0	0	0	0	0	2	30

<그림8. FinalFile.csv 파일>

- FinalFile.csv 파일을 불러옴.
- 불법주정차 감소영향을 주는 변수 선정
- 변수 별 가중치 부여
- null 값 0 으로 처리
- 위에서 선정된 변수 Min - Max 정규화
- 합계 컬럼을 새로 만들어 정규화 된 것들을 sum()을 활용하여 합계를 냄
- 합계 컬럼을 등수를 매겨 줌(ascending = True 활용해 오름차순으로 정렬)
- folium 을 활용해 지도에 상위 20 위 공영주차장 최적입지를 확인

3.2.12 QGIS 격자 표시로 선정된 최적 입지 시각화

- Python 에서 가공한 데이터의 등수를 활용하여 단계별 시각화
- OSM 지도를 이용하여 정확한 위치 확인
- 격자 표시 시각화를 통해 전체적으로 공영주차장이 필요한 곳이 어디인지 분포를 확인함
- OSM 지도를 이용하여 정확한 위치 확인

4. 분석결과

4.1 워드 클라우드 시각화 & 빈도수 분석



<그림9.서울 전체 민원>



<그림10.불법주정차 민원 키워드>

- 워드 클라우드를 사용하여 서울 시민들이 생각하는 특정 단어의 빈도나 중요성을 나타내기 위한 지표.

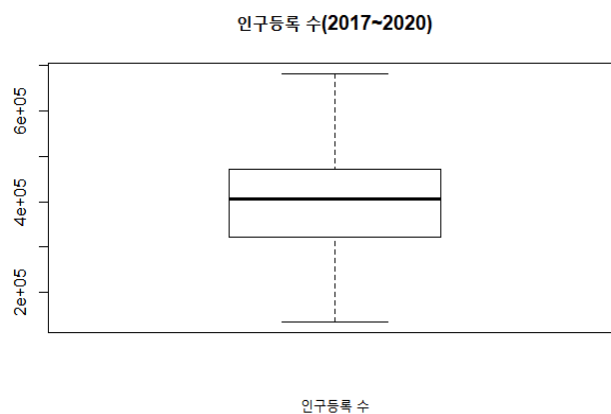
● 단기적 측면 분석 결과

4.2 서울자치구 별로 불법주정차에 영향을 미치는 요인 회귀분석

※노상 주차장: 노상주차장은 자동차의 주차를 위해 도로의 노면 또는 교통광장의 일정한 구역에 설치하는 시설로서 일반의 사용을 위해 마련된 것

※노외 주차장: 도로 위나 교통 광장에 설치된 주차 지역 이외의 장소에 마련한 주차 공간 또는 주차 시설

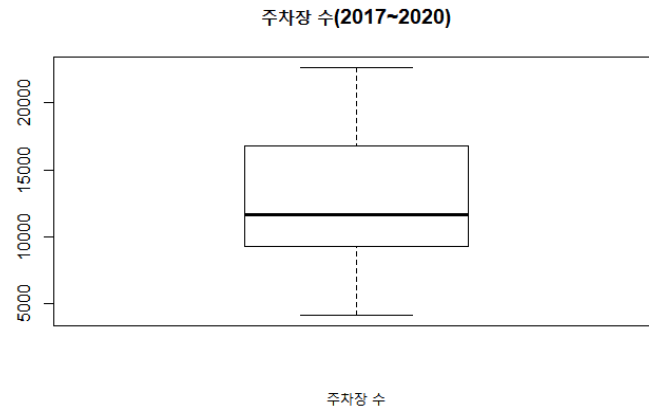
– 독립변수와 종속변수의 box plot



<그림 11. 인구 등록 수>

· 인구등록 수 기초 통계량

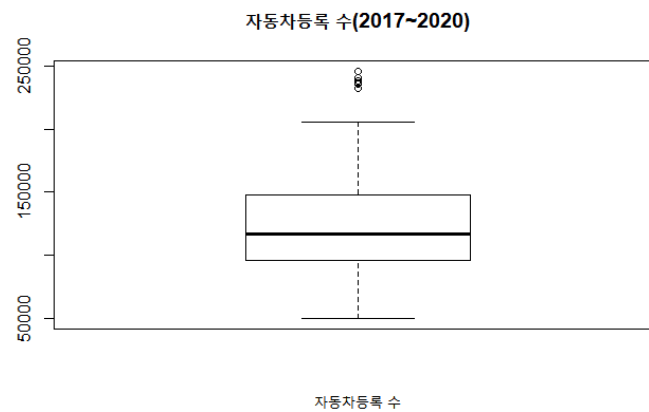
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
134600	323100	405500	401000	469900	682700



<그림 12.주차장 수>

· 주차장 수 기초 통계량

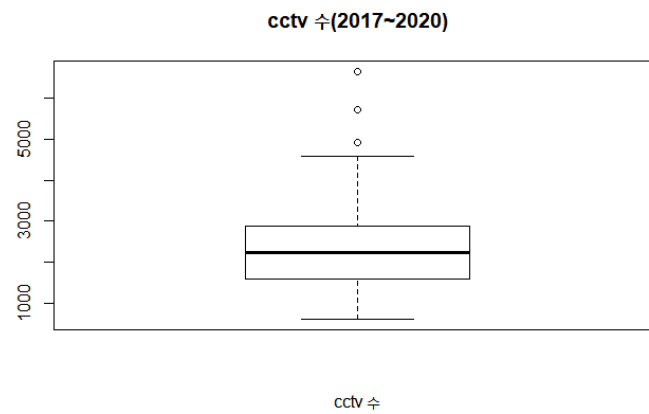
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
4152	9319	11690	12460	16590	22670



<그림 13.자동차 등록 수>

· 자동차등록 수 기초 통계량

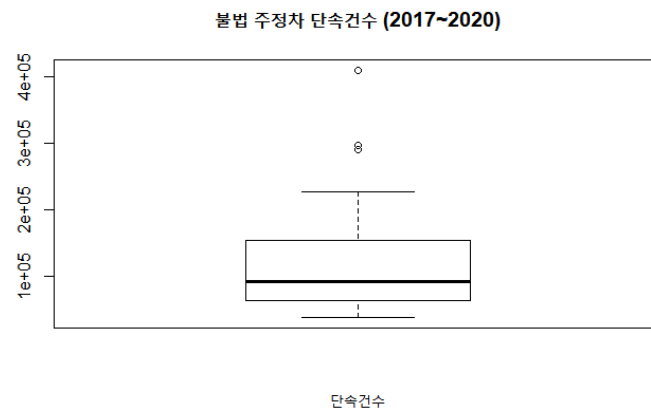
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
50050	96290	116900	125200	147000	245800



<그림 14.CCTV 수>

· CCTV 수 기초통계량

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
615	1612	2236	2380	2873	6645



<그림 15.불법 주 정차 단속건수>

· 불법 주 정차 단속건수 기초통계량

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
37110	63760	91920	108500	153200	410700

□ 독립변수들과 종속변수 상관분석

```
> cor.test(park_info$총_주차장수,park_counts$단속건수)

Pearson's product-moment correlation

data: park_info$총_주차장수 and park_counts$단속건수
t = 2.2836, df = 98, p-value = 0.02456
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.02966472 0.40338214
sample estimates:
      cor
0.2247732

> cor.test(c$pop_counts,park_counts$단속건수)

Pearson's product-moment correlation

data: c$pop_counts and park_counts$단속건수
t = 2.0115, df = 98, p-value = 0.04702
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.002812401 0.380651044
sample estimates:
      cor
0.1991205
```

<그림 16. 상관분석>

- 단속건수와 총 주차장 수의 상관계수는 0.224 이고 p-value=0.02456 으로 유의수준 5%하에서 귀무가설(‘상관계수는 0 이다.’)을 기각하므로 상관계수는 통계적으로 유의미함.
- 단속건수와 인구등록 수의 상관계수는 0.199 이고 p-value=0.04702 로 유의수준 5%하에서 귀무가설(‘상관계수는 0 이다.’)을 기각하므로 상관계수는 통계적으로 유의미함.


```
> cor.test(car_counts$소계,park_counts$단속건수)

Pearson's product-moment correlation

data: car_counts$소계 and park_counts$단속건수
t = 5.8238, df = 98, p-value = 7.309e-08
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.3449955 0.6397558
sample estimates:
cor
0.5070539

> cor.test(b$cctv_counts,park_counts$단속건수)

Pearson's product-moment correlation

data: b$cctv_counts and park_counts$단속건수
t = 3.3977, df = 98, p-value = 0.0009833
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1369483 0.4898193
sample estimates:
cor
0.324634
```

<그림 17. 상관분석>

- 단속건수와 자동차등록대수 의 상관계수는 0.507 이고 p-value=7.309e-08 로 유의수준 5%하에서 귀무가설(‘상관계수는 0 이다.’)을 기각하므로 상관계수는 통계적으로 유의미함.
- 단속건수와 CCTV 설치 개수의 상관계수는 0.324 이고 p-value=0.0009 로 유의수준 5%하에서 귀무가설(‘상관계수는 0 이다.’)을 기각하므로 상관계수는 통계적으로 유의미함.

● 불법주정차 단속건수에 영향을 미치는 요인을 알기 위한 회귀분석

※ 변수명 설명

- | | |
|-----------------|-------------------|
| - x1: 노상주차장 수 | -x5: 구별 CCTV 설치 수 |
| - x2: 노외주차장 수 | -x6: 구별 자동차등록대수 |
| - x3: 부설주차장 수 | -x7: 구별 등록 인구수 |
| - x8: 구별 노후주택 수 | |

```
> summary(fit4)

Call:
lm(formula = y ~ x1 + x2 + x3 + x5 + x6 + x7 + x8)

Residuals:
    Min       1Q   Median       3Q      Max
-87987 -23478 -5643  29707 157997

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 86757.2643 16301.9038   5.322 0.00000071995931632 ***
x1           -9.5706    9.2355  -1.036   0.30278
x2           93.4950   111.7632   0.837   0.40502
x3            1.6600    1.4639   1.134   0.25977
x5            1.8009    4.3876   0.410   0.68244
x6            2.0064    0.2176   9.223 0.00000000000000987 ***
x7           -0.5424    0.1616  -3.357   0.00115 **
x8           -0.3409    0.4672  -0.730   0.46750
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 40810 on 92 degrees of freedom
Multiple R-squared:  0.5852,    Adjusted R-squared:  0.5537
F-statistic: 18.54 on 7 and 92 DF,  p-value: 0.000000000000003387
```

<그림 18. 회귀분석>

- 구별 자동차 등록대수의 p-value=0.00000, 구별 등록인구의 p-value=0.00000. 즉 두 변수는 유의수준 5% 하에서 유의미함.
- 수정된 결정계수는 0.5537 로 약 55.37%의 설명력을 가짐. 모형의 p-value=0.00000 임으로 귀무가설('회귀계수는 0 이다.')을 기각함. 즉 유의수준 5%하에서 통계적으로 유의미함.

- 노상, 노외, 부설주차장 수와 단속건수와의 회귀분석

```
> summary(fit)

Call:
lm(formula = y ~ x1 + x2 + x3)

Residuals:
    Min       1Q   Median       3Q      Max
-87567 -36079 -10526  32490 277303

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 53249.858   18433.314   2.889  0.00478 **
x1             10.786     11.573    0.932  0.35364
x2            353.687     156.297    2.263  0.02589 *
x3              1.901       1.304    1.458  0.14810
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 58740 on 96 degrees of freedom
Multiple R-squared:  0.1033,    Adjusted R-squared:  0.07525
F-statistic: 3.685 on 3 and 96 DF,  p-value: 0.01464
```

<그림19. 회귀분석>

- 노외주차장(x2)의 p-value는 0.02589로 유의수준 5%하에서 귀무가설('x2의 회귀계수는 0이다.')을 기각함. 즉 노외주차장은 통계적으로 유의미함. 수정된 결정계수는 0.07525로 약7.5%의 설명력을 가짐. 모형의 p-value=0.01464로 유의수준 5% 하에서 통계적으로 유의미함.

- 주차장 수가 불법주정차 단속건수에 미치는 영향을 알기 위한 상관분석

```
> cor.test(park_info$총_주차장수, park_counts$단속건수)

Pearson's product-moment correlation

data:  park_info$총_주차장수 and park_counts$단속건수
t = 2.2836, df = 98, p-value = 0.02456
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.02966472 0.40338214
sample estimates:
cor
0.2247732
```

<그림19. 회귀분석>

- 총 주차장 수와 단속건수의 상관계수는 0.224 이고 p-value=0.02456 으로 유의수준 5%하에서 귀무가설('상관계수는 0 이다.')을 기각함. 즉, 통계적으로 유의미함.

● 총 주차장 수와 단속건수와의 회귀분석

```
> model<-lm(park_counts$단속건수~park_info$총_주차장수) #회귀선 적합
> summary(model)
```

Call:

```
lm(formula = park_counts$단속건수 ~ park_info$총_주차장수)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-78597 -42158 -18387  34958 279497
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   72718.220  16780.718    4.333 3.56e-05 ***
park_info$총_주차장수    2.873     1.258    2.284  0.0246 *
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

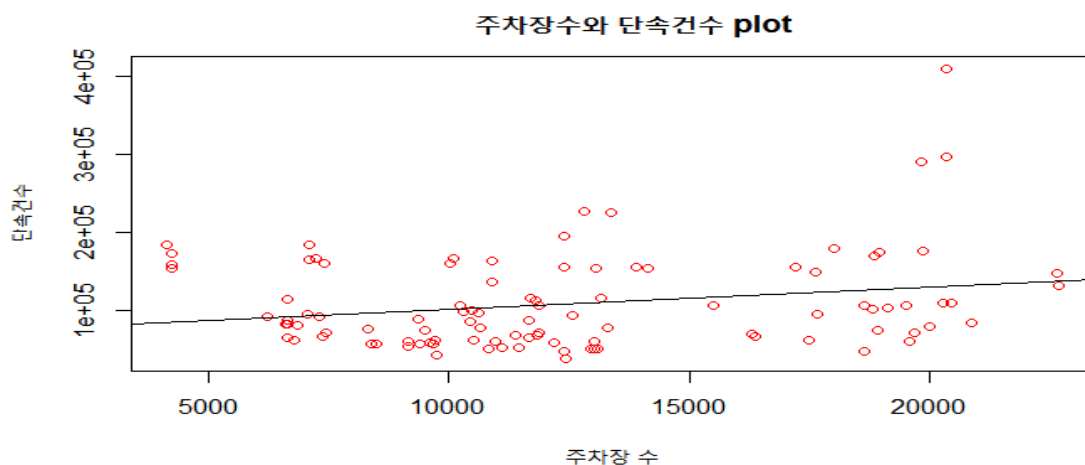
Residual standard error: 59820 on 98 degrees of freedom

Multiple R-squared: 0.05052, Adjusted R-squared: 0.04083

F-statistic: 5.215 on 1 and 98 DF, p-value: 0.02456

<그림20. 회귀분석>

- 총 주차장 수의 p-value 는 0.0246 으로 유의수준 5%하에서 귀무가설('총 주차장 수의 회귀계수는 0 이다.')을 기각함. 즉 회귀계수는 통계적으로 유의미함. 결정계수는 0.05052 로 약 5.05% 설명력을 가짐.
- p-value 가 0.02456 으로 유의수준 5%하에서 귀무가설('회귀계수는 0 이다.')을 기각. 즉 모형이 통계적으로 유의미함.



<그림21. 주차장 수와 단속건수의 회귀직선 적합>

- 회귀분석 결론

노상, 노외, 부설 주차장 수가 불법주정차 단속건수에 미치는 영향을 알기 위해 회귀분석을 시행한 결과, 결정계수가 0.075로 매우 낮게 나옴.

총 주차장 수와 단속건수와의 회귀분석에서도 결정계수가 0.0408로 매우 낮게 나옴.

- 결정계수가 낮아 불법주정차 단속건수를 노상, 노외, 부설 주차장 수, 총 주차장 수로 설명하기 부족.
- 구별로 분석하는 것은 의미가 없다고 판단. 따라서 동 별 분석이 적합.

4.3 서울시 동 별 세분화 후 주차장 확보율과 단속 수 관계

- 주택가 주차장 확보율(노상 + 노외 + 부설(일반주택, 공용주택) 주차장)

※ 일반건축물 부설주차장 제외

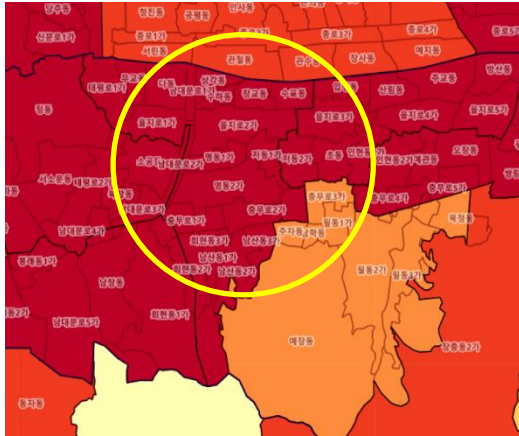
- 높은 확보율 가진 구: 강북구(118.6%), 광진구 (116.9%), 은평구(117.7%)
- 낮은 확보율 가진 구: 중구(89.3%), 영등포구(83.0%), 금천구(75.9%)

구별 주차장 확보율과 단속 수 상관계수: -0.57

→ 주차장 확보율이 높을수록 단속 수가 적다.

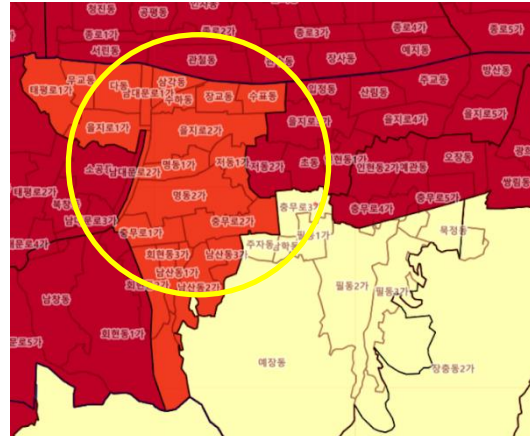
- 주차장 확보율 최저지역: 불법주정차 단속 횟수 많음.

(실제 중구 명동 19.05%, 중구는 대부분 70%가 안됨)



(평균)

<그림 22>

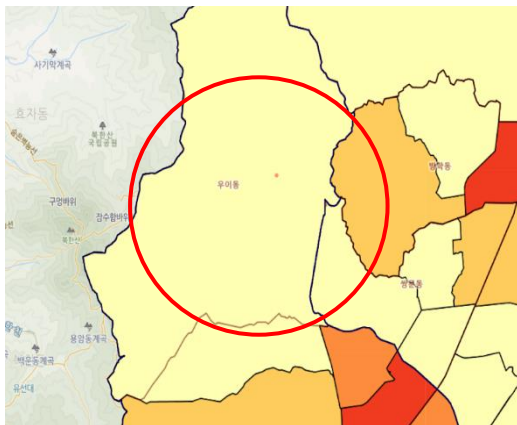


(심야)

<그림 23>

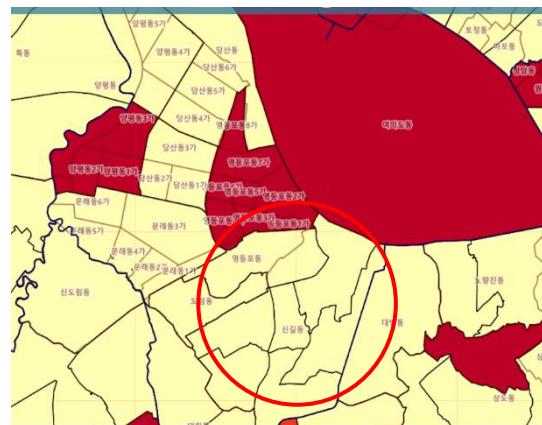
- 주차장 확보율 최고지역: 불법주정차 단속 횟수 적음

(강북구 우이동 172.89%, 영등포구 영등포본동 126.43%)



(평균)

<그림 24>



(심야)

<그림 25>

→ 접근성이 좋은 주차장 확보율 필요

(출처: https://map.seoul.go.kr/spm/gly/policy/view.do?POLICY_NO=165&THEMADO_NO=2142#modal7)

4.4 서울시 동 별 주차 수요 예측

- 동 별 단속건수, 자동차등록대수, 30 년이상 노후주택 수, 주차장 수, 버스 정류장 수, 지하철 정류장 수, 주차장 확보율 데이터를 가지고 수요 예측을 하고자 함.

※변수 구분

-종속변수: 주차장 면수

-독립변수: 단속건수, 자동차등록대수, 30 년이상 노후 주택수, 주차장 수, 버스 정류장 수, 지하철 정류장 수, 주차장 확보율

	crackdown_count	car_count	house_count30	parking_count	parking_c	bus	subway	parking_p
0	4581	22423	6246	612	10507	78	3	47
1	27407	20953	1340	2281	20950	44	5	100
2	24356	39094	11145	1150	33172	64	5	85
3	3986	22699	3192	306	19459	44	4	86
4	15740	20178	1327	1034	22233	51	5	110
...
123	6887	12214	1162	2056	12185	51	0	100
124	14756	22830	2722	1916	18317	80	0	80
125	4043	4710	1027	582	4247	13	1	90
126	10231	18032	61	95	22324	68	1	124
127	705	11839	2262	1315	9412	42	3	79

<그림 26.데이터 형태>

※변수명 설명

-crackdown_count: 단속건수

-car_count: 자동차등록대수

-house_count30: 30 년이상 노후주택 수

-parking_count: 주차장 수

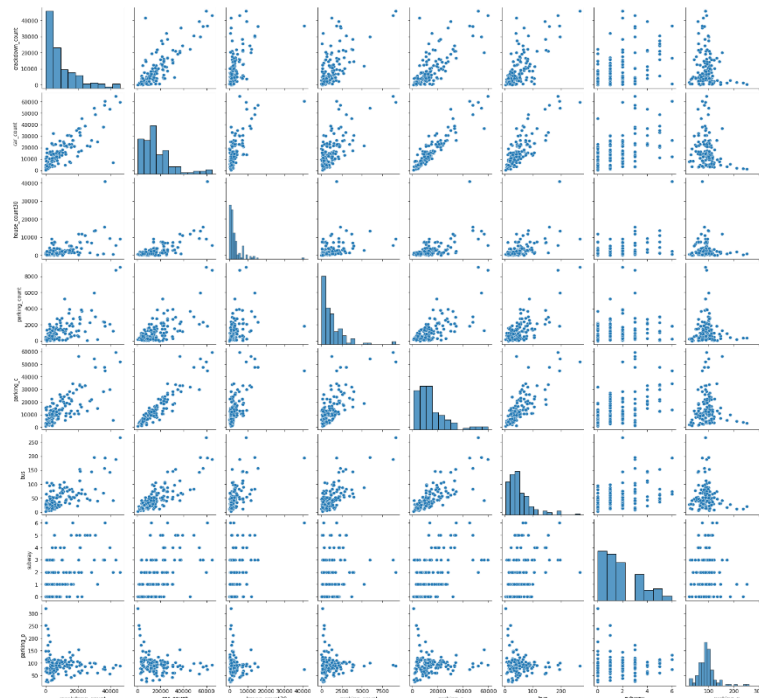
-bus: 버스 정류장 수

-subway: 지하철 정류장 수

-parking_p: 주차장 확보율

-parking_c: 주차장 면수

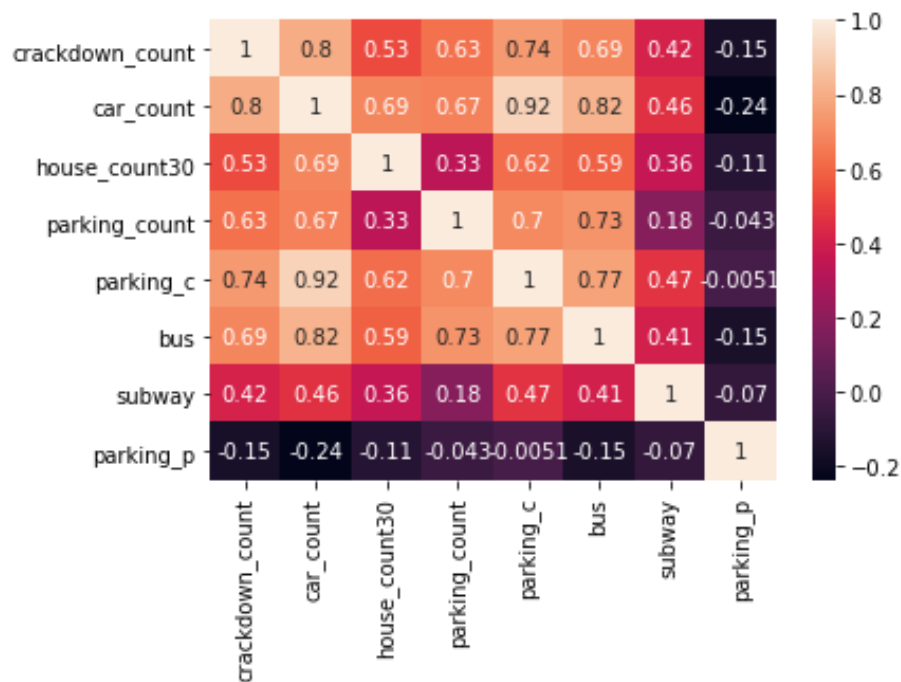
- 주차 수요 예측에 영향이 미치는 요인들을 상관분석 시행



<그림 27. Pair plot>

→ 독립변수와 종속변수가 대체적으로 선형적 관계가 있는 걸로 보임.

- 상관 계수 값 확인



<그림 28. 상관계수 확인>

→ Parking_c(주차장 면수)와 parking_p(주차장 확보율)의 상관계수가 -0.0051 로 약한 음의 상관관계가 보임.

즉, 주차 수요 예측에 포함하지 않는 것으로 결정.

● 데이터 정규화

특성(컬럼)의 스케일과 범위가 다르므로 정규화 필요

	crackdown_count	car_count	house_count30	parking_count	parking_c	bus	subway
0	4581	22423	6246	612	10507	78	3
1	27407	20953	1340	2281	20950	44	5
2	24356	39094	11145	1150	33172	64	5
3	3986	22699	3192	306	19459	44	4
4	15740	20178	1327	1034	22233	51	5
...
123	6887	12214	1162	2056	12185	51	0
124	14756	22830	2722	1916	18317	80	0
125	4043	4710	1027	582	4247	13	1
126	10231	18032	61	95	22324	68	1
127	705	11839	2262	1315	9412	42	3

<그림 29.정규화 하기 전>

	crackdown_count	car_count	house_count30	parking_count	parking_c	bus	subway
0	-0.533377	0.389242	0.667953	-0.450734	-0.363544	0.565702	0.780195
1	1.656367	0.279349	-0.404655	0.689670	0.506511	-0.219104	2.028507
2	1.363678	1.635517	1.739030	-0.083126	1.524783	0.242547	2.028507
3	-0.590456	0.409875	0.000251	-0.659819	0.382289	-0.219104	1.404351
4	0.537128	0.221412	-0.407497	-0.162387	0.613404	-0.057526	2.028507
...
123	-0.312157	-0.373953	-0.443571	0.535931	-0.223742	-0.057526	-1.092273
124	0.442731	0.419668	-0.102506	0.440271	0.287143	0.611867	-1.092273
125	-0.584988	-0.934930	-0.473087	-0.471232	-0.885094	-0.934662	-0.468117
126	0.008639	0.060984	-0.684285	-0.803993	0.620985	0.334877	-0.468117
127	-0.905209	-0.401987	-0.203076	0.029616	-0.454774	-0.265269	0.780195

<그림 30. 정규화 후>

● 모델 구축

총 128 개의 데이터 중에서 80%는 train data 로 사용하고 20%는 test data 로 사용.

train data 중 25%를 검증데이터로 사용하여 학습된 모델을 평가하고 최적의 모델을 선택.

즉, 검증데이터는 과적합 확인용으로 사용.

```
model=Sequential()
model.add(Dense(units=4,activation='linear',input_dim=X_df.shape[1]))
model.add(Dense(units=1,activation='linear'))
model.summary()
```

Model: "sequential_5"

Layer (type)	Output Shape	Param #
dense_10 (Dense)	(None, 4)	28
dense_11 (Dense)	(None, 1)	5
Total params: 33		
Trainable params: 33		
Non-trainable params: 0		

<그림 31. 모델 구축>

입력은 6 개, 출력은 1 개, 은닉층의 수는 1 개이고 은닉층의 노드 수는 4 개이다. Activation 은 선형 회귀 모델이므로 'linear'로 사용. 총 파라미터 수는 33 개이다.

● 파라미터 학습 결과

가중치:

$$w^{[1]} = \begin{pmatrix} 0.601779 & 0.613918 & -0.24532 & -0.05764 & 0.523053 & 0.04486 \\ 0.127833 & 0.377245 & 0.486988 & -0.90814 & -0.58592 & 0.226981 \\ 0.350024 & -1.03548 & -0.31244 & -0.0537 & 0.595538 & -0.22853 \\ 0.401068 & -0.51629 & -0.05394 & -0.38962 & 0.561033 & -0.12746 \end{pmatrix}$$

편향:

$$b^{[1]} = \begin{pmatrix} 0.011138 \\ -0.01388 \\ -0.00516 \\ -0.00056 \end{pmatrix}$$

가중치:

$$w^{[2]} = \begin{bmatrix} 0.32421 & -0.2272 & -0.71734 & 0.014549 \end{bmatrix}$$

편향:

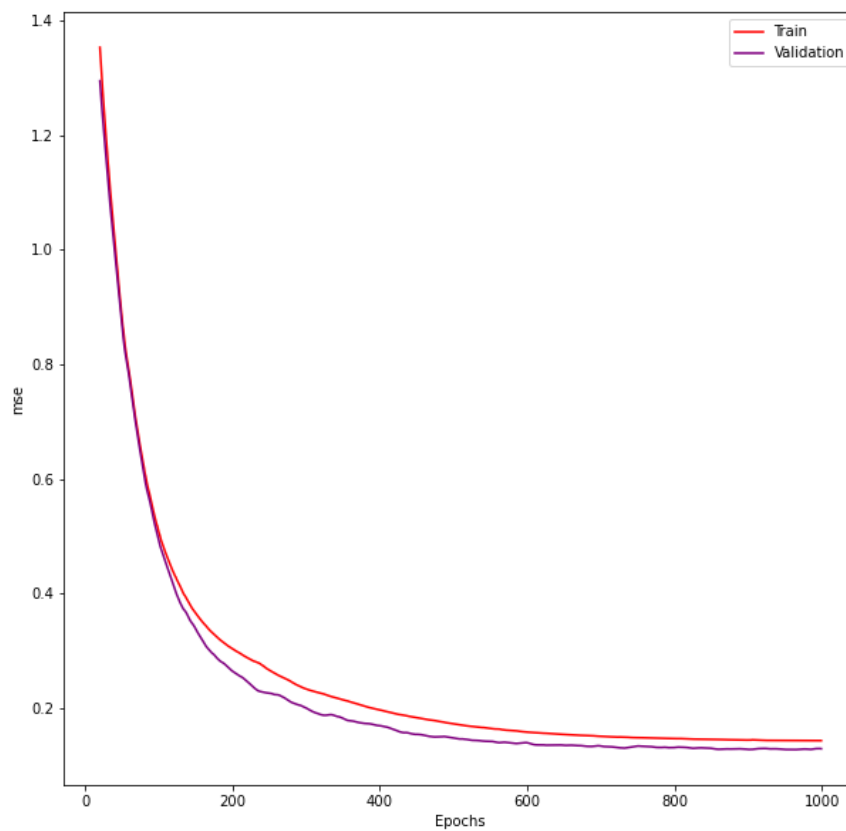
$$b^{[2]} = \begin{bmatrix} 0.006674 \end{bmatrix}$$

- 모형 평가

- MSE 값 확인

테스트 데이터를 통한 모형 평가-> MSE: 0.1098 MAE:0.2407

훈련데이터를 통한 모형 평가-> MSE: 0.1394 MAE:0.2511



<그림 32. 에폭별 비용함수(MSE)값의 변화>

※ 분석 결과: 테스트 데이터의 MSE 값과 훈련데이터의 MSE 값은 큰 차이가 나지 않고 <그림 32>을 보면 훈련데이터와 검증데이터의 비용함수(MSE)값이 비슷한 것으로 보이므로 과적합은 일어나지 않은 것으로 보임.

- 결정계수 확인

```
from sklearn.metrics import r2_score
print('설명력:', r2_score(y_test_origin, y_pred_origin))
```

설명력: 0.8317009511794283

→ 결정계수가 0.8317로 모델은 약 83.17%의 설명력을 가지는 것을 알 수 있음

● 예측

test data 를 이용하여 예측 값과 실제 관측 값 비교

(정규화 된 값을 평균과 표준편차를 이용하여 원래 값으로 계산)

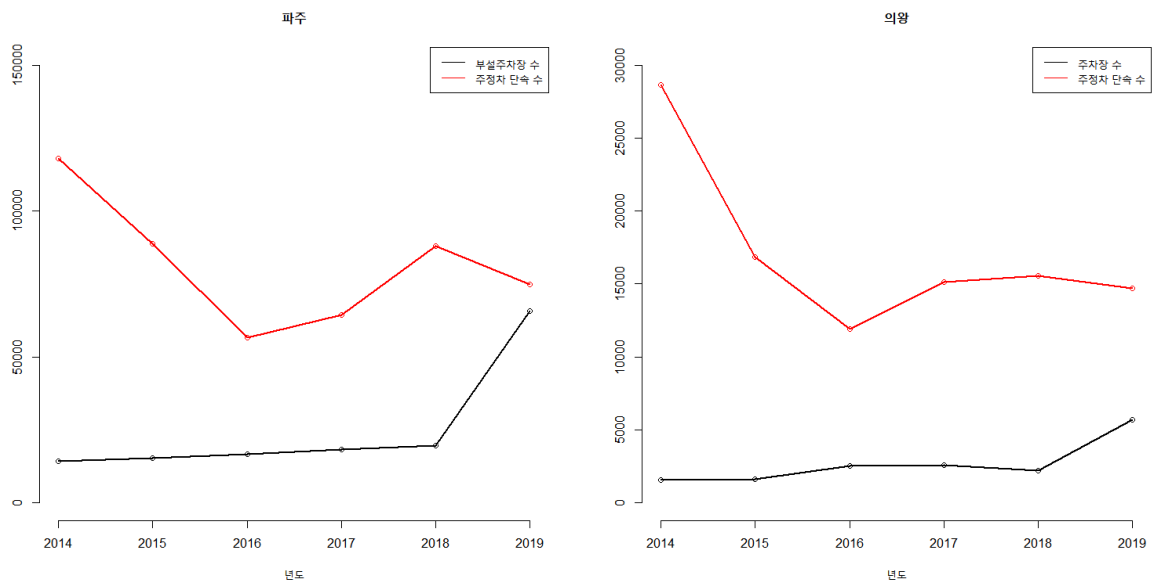
- 정규화 된 값

예측 값↵		실제 관측 값↵	
predict_정규값		parking_c	
0	-0.978843	104	-1.076635
1	0.306583	95	0.095020
2	-0.735342	41	-0.928085
3	-0.301925	123	-0.223742
4	-0.924401	115	-0.938499
5	-0.277945	59	-0.166088
6	0.586997	53	0.938081
7	-0.747928	63	-0.906589
8	2.216765	45	1.263175
9	1.269448	39	1.109626
10	1.561846	2	1.524783
11	-0.182480	47	0.128679
12	0.170618	120	0.207911
13	0.379869	52	0.627651
14	-0.664839	38	-0.932667
15	0.540831	1	0.506511 ↵
↵			

- 원래 값

예측 값↵		실제 관측 값↵	
predict_원래값		parking_c원래값	
0	3122.0	104	1948.0
1	18550.0	95	16011.0
2	6044.0	41	3731.0
3	11247.0	123	12185.0
4	3775.0	115	3606.0
5	11534.0	59	12877.0
6	21916.0	53	26130.0
7	5893.0	63	3989.0
8	41478.0	45	30032.0
9	30107.0	39	28189.0
10	33617.0	2	33172.0
11	12680.0	47	16415.0
12	16918.0	120	17366.0
13	19430.0	52	22404.0
14	6891.0	38	3676.0
15	21362.0 ↵	1	20950.0 ↵

4.5 경기도 예시: 부설주차장 증가에 따라 불법주정차 수 감소.

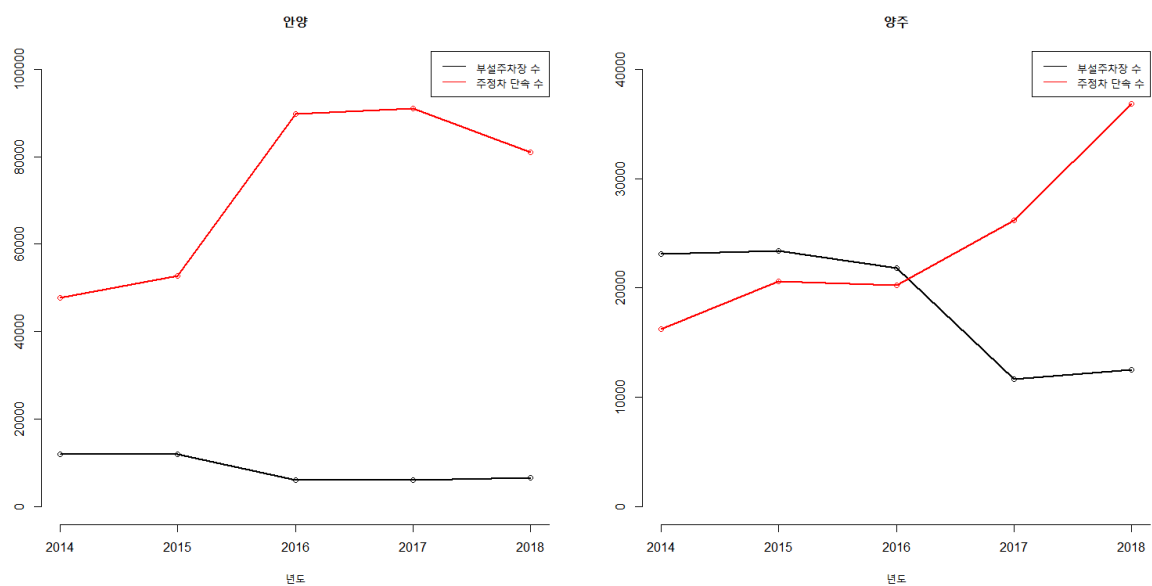


<그림 33. 파주, 의왕 주차장 수 및 단속건수 연도별 현황>

※ 부설주차장 수 단속건수 상관관계수

```
> cor(park_paju,n_paju)
[1] -0.2046745
> cor(park_uiwang,n_uiwang)
[1] -0.408439
```

→ 상관관계수 -0.205, -0.408 로 음수이므로,부설주차장 수가 증가함에 따라 주 정차 단속 수가 감소



<그림 34. 안양, 양주 주차장 수 및 단속건수 연도별 현황>

※ 부설주차장 수 단속건수 상관관계

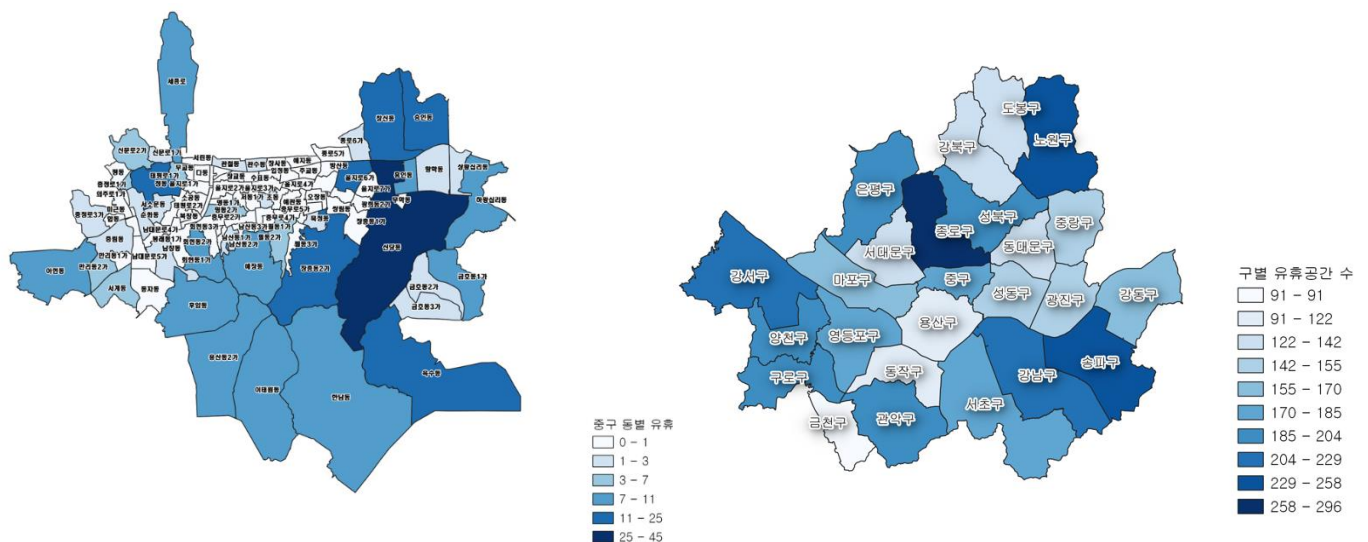
```
> cor(park_anayang,n_anayang)
[1] -0.9896681
> cor(park_yangju,n_yangju)
[1] -0.8316902
```

→ 상관계수 -0.989, -0.832 로 음수이므로,부설주차장 수가 감소함에 따라 주 정차 단속 수가 증가

4.6 부설주차장으로 활용할 수 있는 유휴공간의 개수 파악

유휴공간 (학교, '도서관', '문화시설', '주민자치센터', '체육시설', '민원행정 기관'),

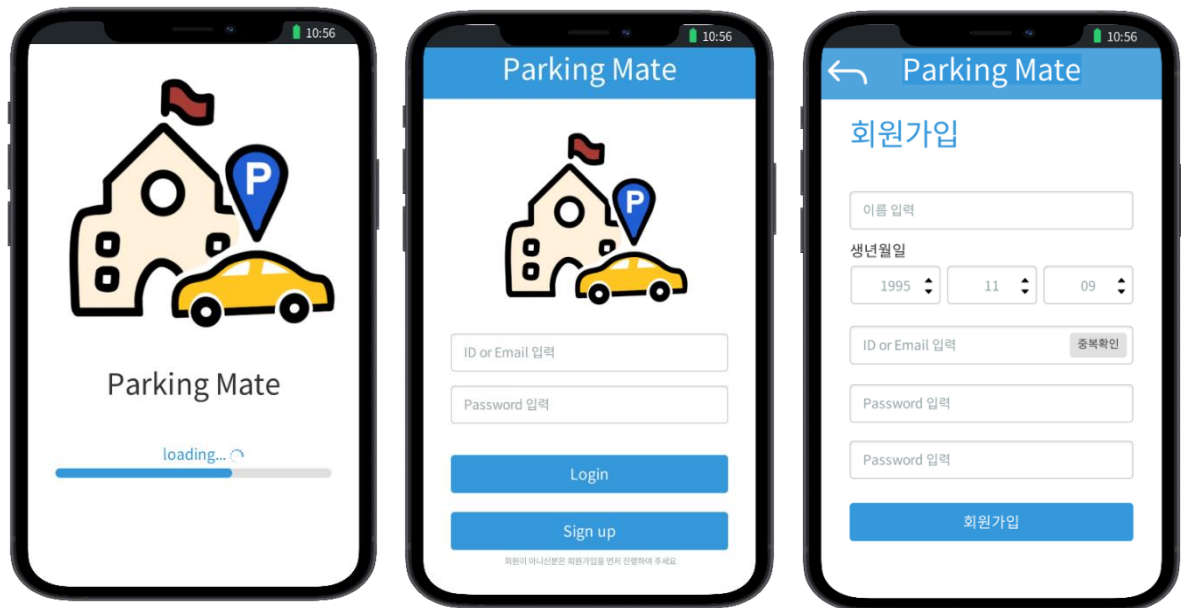
부설주차장으로 활용할 수 있는 유휴공간의 개수는 구별로 최소 91개에서 최대 296개에 달한다. 주차장 확보율이 낮은 중구의 유휴공간을 자세히 살펴보면 최대 45개의 유휴공간을 이용할 수 있음을 알 수 있다. 이를 부설주차장으로 활용한다면, 불법주정차를 줄일 수 있을 것으로 예상된다.



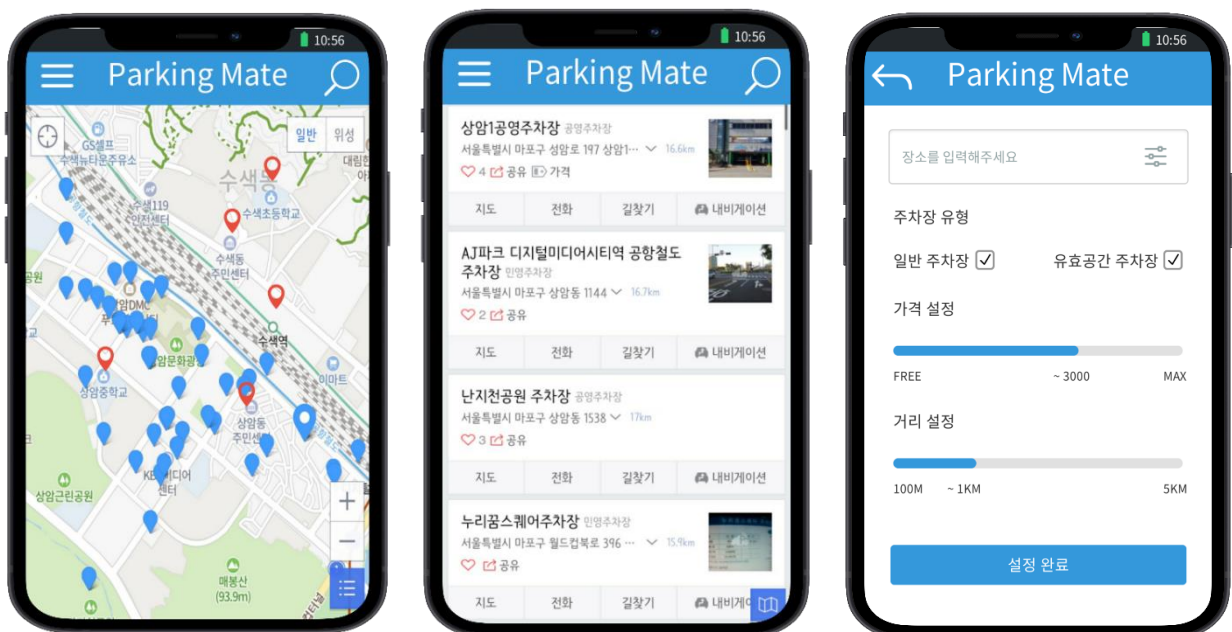
<그림 35. 구별 및 중구 동별 유휴공간 수>

4.7 부설주차장을 늘리는 방안으로 결론 도출 및 앱 서비스

단기적 측면으로 유휴공간을 활용해 접근성이 좋은 부설주차장을 이용할 수 있도록 정책 활성화와 다른 주차장 어플과 차별화된 유휴공간도 포함된 새로운 공영주차장 app 서비스 제공



<그림36 로딩 화면 →로그인 화면 → 회원가입 페이지(샘플 예시)>



<그림37. 메인 화면 지도 →메인 텍스트 → 필터 설정(샘플 예시)>



<그림38. 상세 설명 → 네비게이션 → 메인 메뉴 목록(샘플 예시)>



<그림39. 불법주정차 신고 → 유효공간 등록(샘플 예시)>

- 장기적 측면 분석 결과코드 및 시각화

4.8 공영주차장 최적 입지 시각화

2021. 8. 15.

최적입지선정

```

In [1]: #csv로 저장된 파일 불러오기

import pandas as pd
df = pd.read_csv('finalFile.csv', header=0, encoding='cp949', engine='python')

In [2]: #columns 확인

col = df.columns.tolist()
print(col)

['left', 'bottom', 'right', 'top', 'lon', 'lat', '학교', '도서관', '문화시설', '주민자치센터', '체육시설', '민원행정기관', 'cctv', '주차장']

In [3]: #columns 중에 불법주정차 감소요인에 관련된 변수만 선정
score = ['학교', '도서관', '문화시설', '주민자치센터', '체육시설', '민원행정기관', '주차장']
print(score)

#가중치변수를 score수 만큼 만들어줌.
weight = [1] * len(score)

['학교', '도서관', '문화시설', '주민자치센터', '체육시설', '민원행정기관', 'cctv', '주차장']

In [4]: ...

각각의 변수에 가중치를 줌

학교 = 2
도서관 = 1
문화시설 = 1
주민자치센터 = 1
체육시설 = 1
민원행정기관 = 1
cctv = 2
주차장 = 3
...
weight = [2,1,1,1,1,1,2,3]

```

<그림40.가중치 선정>

```

In [5]: #null값 0으로 처리

df.fillna(0, inplace=True)

In [6]: # 정규화

from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()

df[score] = scaler.fit_transform(df.loc[:,score])
df.reset_index(inplace=True)

In [7]: # 총점 계산

for i in range(0, len(score)):
    name = score[i]
    df[name] = df[name] * weight[i]

```

<그림41.정규화, 총점 계산>

2021. 8. 15.

최적입지선정

```
df.set_index(['index', 'left', 'bottom', 'right', 'top', 'lon', 'lat'], inplace=True)

df['sum'] = df.sum(1)
df.reset_index(inplace=True)
```

In [9]:

```
# 등수 계산
df['ranking'] = df['sum'].rank(method='dense', ascending=True).astype(int)

final = df.sort_values(by='ranking')
final.reset_index(drop=True, inplace=True)
```

In [10]:

```
# 상위 5개만 확인
display(final.head(5))

# 최종 결과 CSV 파일쓰기
# final.to_csv('결과_총정.csv', index=False, sep=',', encoding='cp949')
```

	index	left	bottom	right	top	lon	lat	학교	도서관	문화시설	주민자치센터	체육시설	민원행정기관	cct
0	0	937200	1953600	937500	1953900	126.790447	37.581006	0.0	0.0	0.0	0.0	0.0	0.0	0.
1	2089	959700	1950900	960000	1951200	127.045413	37.557927	0.0	0.0	0.0	0.0	0.0	0.0	0.
2	2094	959700	1952400	960000	1952700	127.045331	37.571447	0.0	0.0	0.0	0.0	0.0	0.0	0.
3	2117	959700	1965900	960000	1966200	127.044589	37.693125	0.0	0.0	0.0	0.0	0.0	0.0	0.
4	2118	959700	1966200	960000	1966500	127.044573	37.695829	0.0	0.0	0.0	0.0	0.0	0.0	0.

<그림41.등수 계산, 순위화>

In [11]:

#folium 을 활용해 지도에 상위 20위 공영주차장 최적입지를 확인

cnt = 20

import folium

for i in range(cnt):

map = folium.Map(location=[final.loc[i, 'lat'], final.loc[i, 'lon']], zoom_start = 1)

folium.Marker([final.loc[i, 'lat'], final.loc[i, 'lon']], add_to(map)

print (final.loc[i, 'lat'], final.loc[i, 'lon'])

j = i+1

display(f'상위 {j}번째 격자의 센터 위치 보기', map)

37.58100598 126.79044740000002

'상위 1번째 격자의 센터 위치 보기'

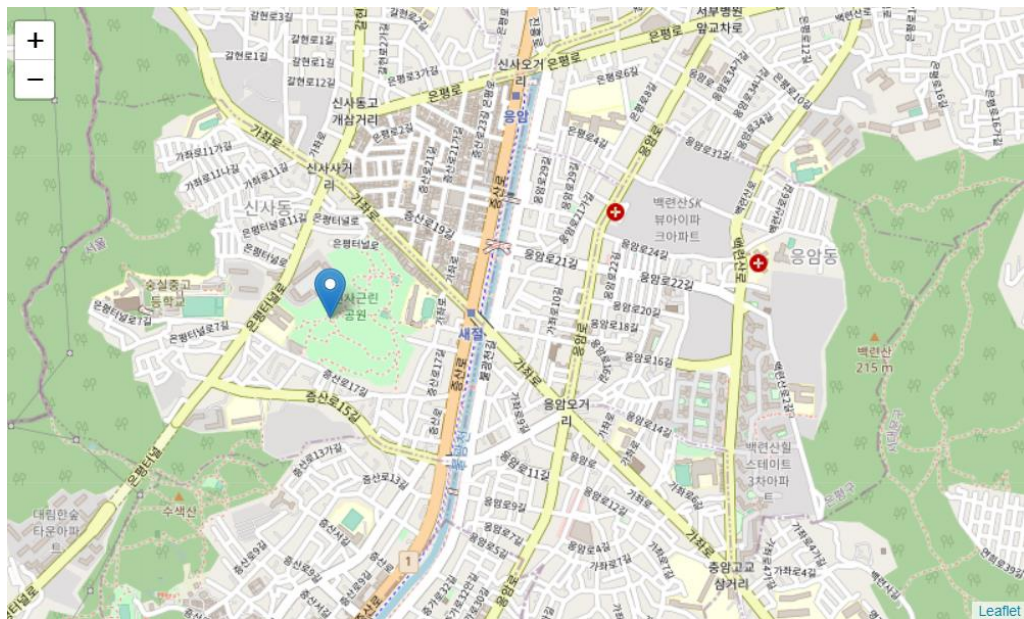
Make this Notebook Trusted to load map: File -> Trust Notebook

<그림42.상위 최적 입지 확인>

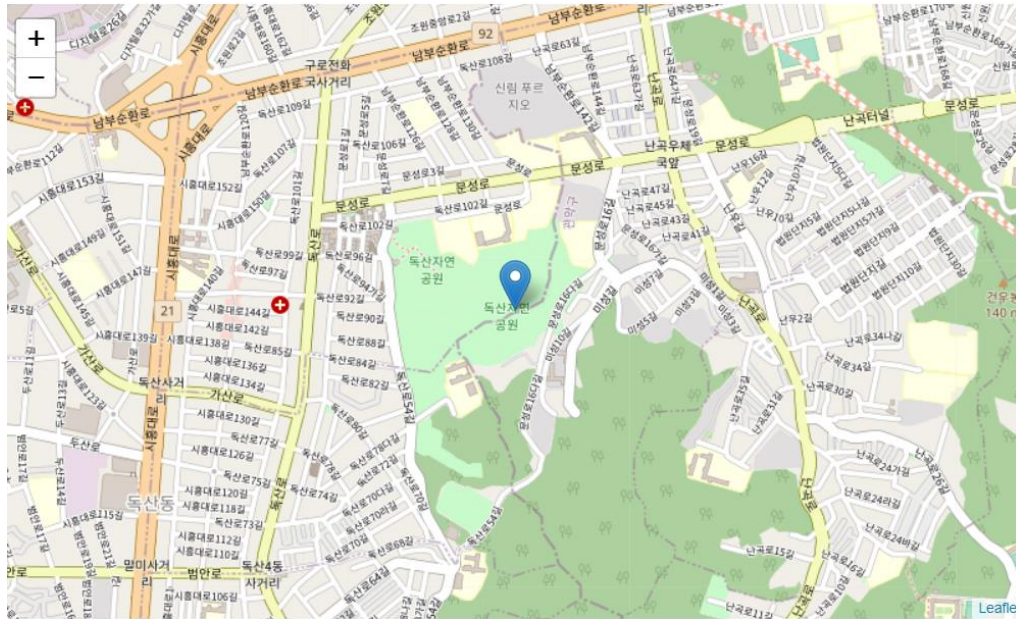
2021 년 공공 빅데이터 분석 청년 인재양성 데이터 분석 전문교육과정

left	bottom	right	top	lon	lat	학교	도서관	문화시설	주민자치선체육시설	민원행정7	cctv	주차장	sum	ranking
941700	1941600	942000	1941900	126.8424	37.47314	0	0	0	0	0	0	0	0	1
939600	1950000	939900	1950300	126.8179	37.54872	0	0	0	0	0	0	0	0	1
941700	1941900	942000	1942200	126.8423	37.47585	0	0	0	0	0	0	0	0	1
939300	1942200	939600	1942500	126.8152	37.4784	0	0	0	0	0	0	0	0	1
939900	1951800	940200	1952100	126.8212	37.56496	0	0	0	0	0	0	0	0	1
949200	1944000	949500	1944300	126.927	37.49522	0	0	0	0	0	0	0.00838	0.00838	2
951300	1946100	951600	1946400	126.9506	37.51426	0	0	0	0	0	0	0.00838	0.00838	2
938700	1950900	939000	1951200	126.8077	37.55677	0	0	0	0	0	0	0.00838	0.00838	2
947400	1943400	947700	1943700	126.9067	37.48971	0	0	0	0	0	0	0.00838	0.00838	2
946200	1945500	946500	1945800	126.893	37.50857	0	0	0	0	0	0	0.00838	0.00838	2
946800	1953300	947100	1953600	126.8992	37.57891	0	0	0	0	0	0	0.01676	0.01676	3
966300	1947300	966600	1947600	127.1203	37.52574	0	0	0	0	0	0	0.01676	0.01676	3
951000	1948500	951300	1948800	126.9471	37.53587	0	0	0	0	0	0	0.02514	0.02514	4
951600	1944600	951900	1944900	126.9541	37.50075	0	0	0	0	0	0	0.02514	0.02514	4
945300	1954800	945600	1955100	126.8821	37.59234	0	0	0	0	0	0	0.02514	0.02514	4
963300	1948200	963600	1948500	127.0863	37.53374	0	0	0	0	0	0	0.02514	0.02514	4
950400	1945500	950700	1945800	126.9405	37.5088	0	0	0	0	0	0	0.03352	0.03352	5
954600	1941600	954900	1941900	126.9882	37.47387	0	0	0	0	0	0	0.03352	0.03352	5
962400	1950000	962700	1950300	127.076	37.54993	0	0	0	0	0	0	0.03352	0.03352	5
957300	1942500	957600	1942800	127.0187	37.48211	0	0	0	0	0	0	0.03352	0.03352	5
943500	1950300	943800	1950600	126.862	37.55167	0	0	0	0	0	0	0.03352	0.03352	5
948300	1943400	948600	1943700	126.9169	37.48976	0	0	0	0	0	0	0.03352	0.03352	5

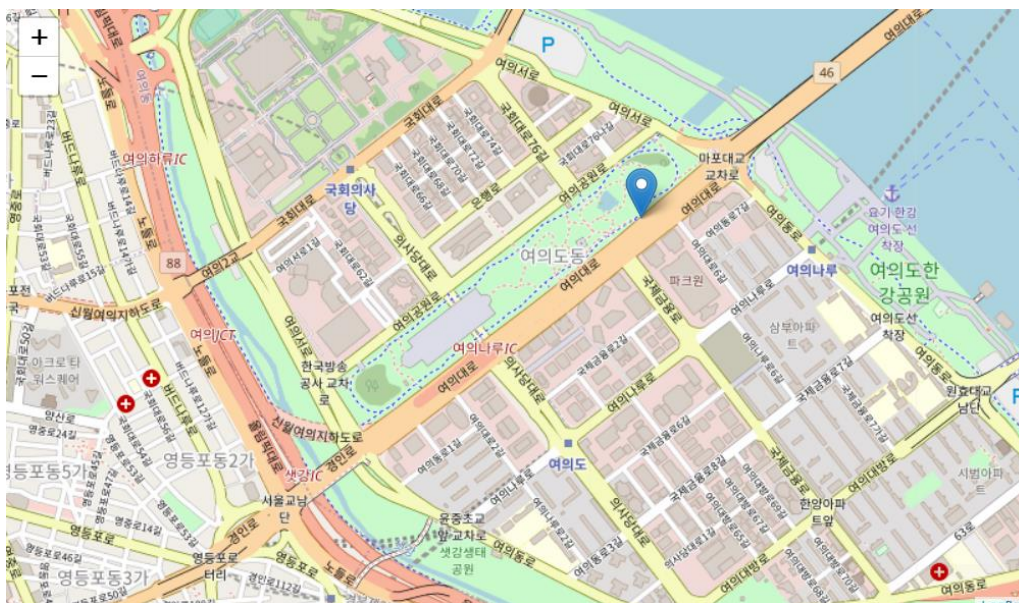
<그림43 : 서울 특별시 공영주차장 최적입지 순위화>



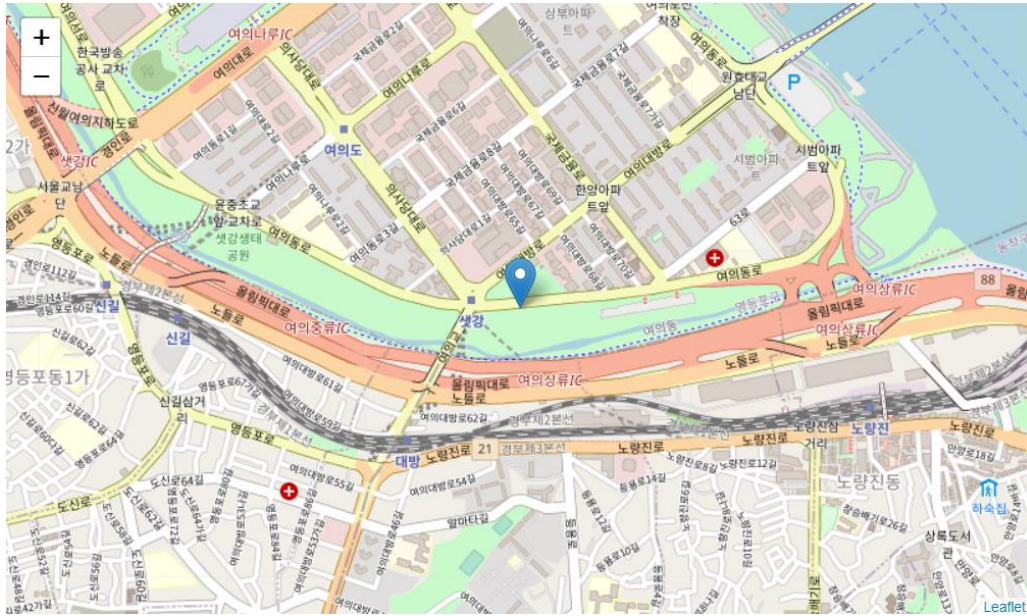
<그림44 : 최적 입지선정 후보지(1)>



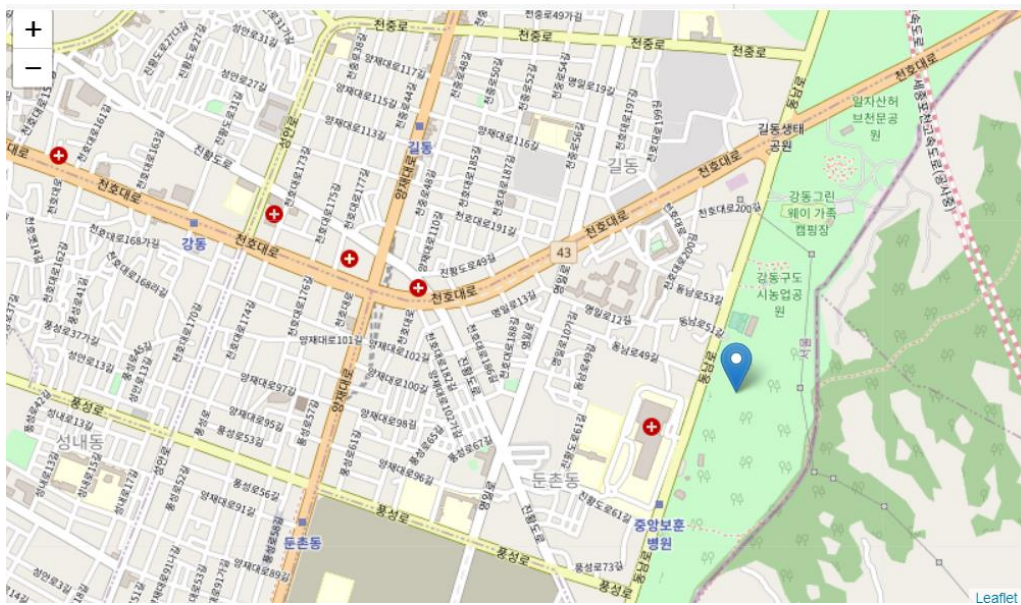
<그림45 : 최적 입지선정 후보지(2)>



<그림46 : 최적 입지선정 후보지(3)>



<그림47 : 최적 입지선정 후보지(4)>



<그림48 : 최적 입지선정 후보지(5)>

5. 활용 방안

5.1 문제점 개선 방안

- 불법주정차로 인한 불필요한 공회전과 차량 이동으로 발생하는 교통흐름 방해 등 다양한 문제로 인해 발생하는 사회적 비용 감소.
- 보행안전과 밀접한 장소(어린이보호구역, 횡단보도, 버스정류장 등)의 시야 확보로 보행자 교통사고 감소.
- 소방 및 경찰 등 긴급 출동이 요구되는 응급 차량의 통행로 확보.
- 불법주정차 감소로 인한 주민의 주거 환경 및 삶의 질 개선.
- 서울지역 불법주정차 감소 요인분석을 통한 예측 모델을 활용해 타 지역에도 확대 가능.

5.2. 업무 활용 방안

5.2.1 새로운 앱 서비스 출시

민원분석을 통해 시민들이 원하는 편의 기능(유희공간 추천, 적립포인트, 즐겨찾기 등)을 더 첨가하고,
기존 서울시에서 관리하는 “주차정보안내시스템” 지도를 새롭게 공영주차장 앱 서비스를 출시한다.

5.2.2 잠재된 부설주차장 사업 활성화 및 수정 방안 모색

기존의 부설주차장 사업이 있지만, 아직 참여하지 않은 사업체나 민원행정기관이 많다. 다양한 홍보와 정책 시도를 통해 새로운 협업체를 발굴하는 작업을 추진하도록 한다.

5.2.3 새로운 공영주차장 입지 사업 계획 추진

서울시에서 “운영하는 주택가 및 공공시설 공영주차장 건설 지원 사업”을 고려하여 유희공간, 주차장, CCTV 위주로 불법주정차 감소요인이 없는 지역 위주로 입지 후보지를 선정하였다.

6. 참고자료

참고자료

- 기사

<https://www.gidream.com/news/articleView.html?idxno=607668>

“과태료 강화” ‘스쿨 존’ 불법 주 정차 막을까

https://news.sbs.co.kr/news/endPage.do?news_id=N1005829445&plink=ORI&cooper=NAVER

스쿨 존 불법 주 정차 30만 건…꾸준한 위반이 문제

<https://www.yonhapnewstv.co.kr/news/MYH20210713025200641?did=1825m>

스쿨 존 노상주차장 폐지 첫날…변함없는 학교 앞

https://imnews.imbc.com/replay/2021/nwtoday/article/6280577_34943.html

단속 강화됐는데도…스쿨 존 불법주정차 여전

<http://news.heraldcorp.com/view.php?ud=20200522000113>

서울시민 55% “불법 주 정차 차량 단속 불만”

https://newsis.com/view/?id=NISX20210108_0001298948&cID=14001&pID=14000

시민신고제 정착되니…작년 서울 불법주정차 230만건, 5년내 최소

<http://www.shinailbo.co.kr/news/articleView.html?idxno=1387511>

성 동 구 , 불 법 주 정 차 단 속 C C T V 1 3 5 대 위 ' 한 눈 에 '

<https://news.kbs.co.kr/news/view.do?ncd=5243484&ref=A>

운전자 시야 막는 불법 주 정차…사고 위험 급증 ‘골머리’

<https://www.interview365.com/news/articleView.html?idxno=90707>

서울시, 어 린 이보 호 구 역 과 속 단 속 · 불 법 주 정 차 C C T V 설 치 확 대

<http://www.nbnnews.co.kr/news/articleView.html?idxno=523711>

의정부시, 주 정차 과태료 정리 및 주차장 확충 노력

<http://www.namdonews.com/news/articleView.html?idxno=655240>

[기동취재]광주역 일대 불법주정차 ‘몸살’

● 참고문헌

- 권 용훈, 이 세구(2013) 「공동주차장 타당성분석을 위한 수요 및 편익추정 연구」 서울공공투자 관리센터
- 임 채훈(2015) 「불법 주 정차 단속 실효성 제고 방안」 삼성교통안전연구소.
- 장재민, 김태형(2017) 「불법주정차의 사회적 비용에 기반한 주차정책방향에 대한 연구」
- 한상진, 우승국, 심 재익, 박 경욱, 엄 기종, 이동윤(2019)
「지자체교통안전정책 개선 지원사업」 한국교통연구원.
- 신 우재, 김 건우, 김 정민(2020) 「서울시 주차문제 해결을 위한 주차장 이용 효율 향상 방안 연구」
- 김 우혁, 최상민(2021) 불법주정차 실태 분석: 인천시 사례를 중심으로. 대한 전자공학회

7. 부록

● 워드 클라우드 코드

```

In [ ]:
from selenium.webdriver import Chrome
import time
import sqlite3
from pandas.io import sql
import os
import pandas as pd
import matplotlib as mpl
import matplotlib.pyplot as plt

In [ ]:
from selenium import webdriver

options = webdriver.ChromeOptions()
options.add_argument("--start-maximized");

browser = webdriver.Chrome("chromedriver", options=options)

In [ ]:
#서울민원 데이터 수집

browser.get("https://democracy.seoul.go.kr/front/allSuggest/list.do?sn2=&searchCondition=")
browser.implicitly_wait(5)

In [ ]:
articleList = []
dfList = []

for y in range(10):
    for x in range(2,11):
        for i in range(1,13):
            review = browser.find_elements_by_xpath(f'//*[@id="suggList_{i}"]')

            for tmp in review:
                df = pd.DataFrame({
                    "NO": [tmp.find_element_by_xpath('p[2]').text],

                })
                dfList.append(df)

            articleList = pd.concat(dfList)

            browser.find_element_by_xpath(f'//*[@id="content"]/div[3]/div[2]/span/a[{x}]')
            browser.implicitly_wait(10)
            browser.find_element_by_xpath(f'//*[@id="content"]/div[3]/div[2]/a[3]').click()
            browser.implicitly_wait(10)

In [ ]:
#데이터 확인
articleList = pd.concat(dfList)

In [ ]:
articleList

In [ ]:
#데이터 처리(한글, 공백 제외)
import re

articleList["NO"] = articleList["NO"].apply(lambda x: re.sub("[^가-힣Ws]", "", x))
articleList

```

```

In [ ]: #공백 row 삭제

articleList = articleList[articleList["NO"].apply(lambda x: re.sub("[^가-힣]*", "", x))
articleList

```

```

In [ ]: #WordExtractor로 단어 점수 학습

from soynlp.word import WordExtractor

word_extractor = WordExtractor(min_frequency=100,
                               min_cohesion_forward=0.05,
                               min_right_branching_entropy=0.0
                              )
word_extractor.train(articleList["NO"].values)
words = word_extractor.extract()
words

```

```

In [ ]: #Tokenizer (L-R 분리)

from soynlp.tokenizer import LTokenizer
from soynlp.word import WordExtractor
from soynlp.utils import DoublespaceLineCorpus

cohesion_score = {word:score.cohesion_forward for word, score in words.items()}
tokenizer = LTokenizer(scores=cohesion_score)

```

```

In [ ]: articleList["tokenizer"] = articleList["NO"].apply(lambda x: tokenizer.tokenize(x, re
articleList

```

```

In [ ]: #각각의 단어들을 words변수에 append 작업

words = []
for i in articleList["tokenizer"].values:
    for k in i:
        words.append(k)
words

```

```

In [ ]: #각각의 단어들을 counter를 사용하여 빈도수를 확인

from collections import Counter
cnt = Counter(words)
words = dict(cnt)
words

```

```

In [ ]: #내림차순으로 정렬

rank = sorted(words.items(),reverse=True,key=lambda item: item[1])

```

```

In [ ]: #불법주정차 관련 빈도수 분석

for i in range(len(rank)):
    print(i+1,"위: ",rank[i])

```

```
In [ ]: #wordcloud 시각화 진행

from wordcloud import WordCloud
wordcloud = WordCloud(height = 500, width=500,background_color="white")W
               .generate_from_frequencies(words)

In [ ]: wordcloud = WordCloud(font_path = 'C:/Windows/Fonts/malgun.ttf', background_color='wh

In [ ]: #wordcloud 시각화

plt.figure(figsize=(10,10))
plt.axis("off")
plt.imshow(wordcloud)
plt.show()
```

● 최적입지 선정 코드

```

In [ ]: #csv로 저장된 파일 불러오기

import pandas as pd
df = pd.read_csv('finalFile.csv', header=0, encoding='cp949', engine='python')

In [ ]: #columns 확인

col = df.columns.tolist()
print(col)

In [ ]: #columns 중에 불법주정차 감소요인에 관련된 변수만 선정
score = ['학교', '도서관', '문화시설', '주민자치센터', '체육시설', '민원행정기관', 'cctv', '주차장']
print(score)

#가중치변수를 score수 만큼 만들어줌.
weight = [1] * len(score)

In [ ]: ...
각각의 변수에 가중치를 줌

학교 = 2
도서관 = 1
문화시설 = 1
주민자치센터 = 1
체육시설 = 1
민원행정기관 = 1
cctv = 2
주차장 = 3
...
weight = [2,1,1,1,1,1,2,3]

In [ ]: #null값 0으로 처리

df.fillna(0, inplace=True)

In [ ]: # 정규화

from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()

df[score] = scaler.fit_transform(df.loc[:,score])
df.reset_index(inplace=True)

In [ ]: # 총점 계산

for i in range(0, len(score)):
    name = score[i]
    df[name] = df[name] * weight[i]

df.set_index(['index', 'left', 'bottom', 'right', 'top', 'lon', 'lat'], inplace=True)

df['sum'] = df.sum(1)
df.reset_index(inplace=True)

```

```

In [ ]: # 등수 계산
df['ranking'] = df['sum'].rank(method='dense',ascending=True).astype(int)

final = df.sort_values(by='ranking')
final.reset_index(drop=True, inplace=True)

In [ ]: # 상위 5개만 확인
display(final.head(5))

In [ ]: #folium 을 활용해 지도에 상위 20위 공영주차장 최적입지를 확인

cnt = 20

import folium

for i in range(cnt):
    map = folium.Map(location=[final.loc[i,'lat'],final.loc[i,'lon']], zoom_start = 1)
    folium.Marker([final.loc[i,'lat'],final.loc[i,'lon']]).add_to(map)
    print (final.loc[i,'lat'],final.loc[i,'lon'])
    i = i+1
    display(f'상위 {i}번째 객체의 센터 위치 보기', map)
    
```

● 단속 수 회귀분석 코드

```
options(scipen=999)
setwd("C:\\Users\\user\\Documents\\링 프로젝트")

#y: 불법주정차 단속건수
dat <- read.table('서울시(2017-2020)불법주정차 단속건수.txt', sep='\\t', encoding='UTF-8')
colnames(dat)<-c("행정구", 2017,2018,2019,2020)

y <- c( dat[,2], dat[,3], dat[,4], dat[,5])

#x1: 노상주차장
dat1 <- read.table('서울시(2017-2020)노상주차장.txt', sep='\\t', encoding='UTF-8')
colnames(dat1)<-c("행정구", 2017,2018,2019,2020)

x1 <- c( dat1[,2], dat1[,3], dat1[,4], dat1[,5])

#x2: 노외주차장
dat2 <- read.table('서울시(2017-2020)노외주차장.txt', sep='\\t', encoding='UTF-8')
colnames(dat2)<-c("행정구", 2017,2018,2019,2020)

x2 <- c( dat2[,2], dat2[,3], dat2[,4], dat2[,5])

#x3: 부설주차장
dat3 <- read.table('서울시(2017-2020)부설주차장.txt', sep='\\t', encoding='UTF-8')
colnames(dat3)<-c("행정구", 2017,2018,2019,2020)

x3 <- c( dat3[,2], dat3[,3], dat3[,4], dat3[,5])

#x5: 구별 cctv 설치수
dat5 <- read.table('서울시(2017-2020)cctv 설치수.txt', sep='\\t', encoding='UTF-8')
colnames(dat5)<-c("행정구", 2017,2018,2019,2020)

x5 <- c( dat5[,2], dat5[,3], dat5[,4], dat5[,5])

#x6: 구별 자동차등록대수
dat6 <- read.table('서울시(2017-2020)자동차등록대수.txt', sep='\\t', encoding='UTF-8')
colnames(dat6)<-c("년도", "행정구", "자동차등록대수")

x6<-c( dat6[,3])

#x7: 구별 등록인구수
dat7 <- read.table('서울시(2017-2020)구별등록인구수.txt', sep='\\t', encoding='UTF-8')
colnames(dat7)<-c("행정구", 2017,2018,2019,2020)

x7 <- c( dat7[,2], dat7[,3], dat7[,4], dat7[,5])

#x8: 구별 노후주택 수
dat8 <- read.table('서울시(2017-2020)구별노후주택수.txt', sep='\\t', encoding='UTF-8')
colnames(dat8)<-c("행정구", 2017,2018,2019,2020)

x8 <- c( dat8[,2], dat8[,3], dat8[,4], dat8[,5])

#단속건수에 영향을 미치는 요인 회귀분석 (귀무가설: 모든 회귀계수가 0이다)
fit4<- lm(y~x1+x2+x3+x5+x6+x7+x8)
summary(fit4)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x5 + x6 + x7 + x8)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -87987 -23478 -5643  29707 157997
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 86757.2643 16301.9038   5.322 0.00000071995931632 ***
## x1           -9.5708     9.2355  -1.036   0.30278
## x2            93.4950    111.7632   0.837   0.40502
## x3             1.6600     1.4639   1.134   0.25977
## x5             1.8009     4.3878   0.410   0.68244
## x6             2.0064     0.2178   9.223 0.000000000000000987 ***
## x7            -0.5424     0.1616  -3.357   0.00115 **
## x8            -0.3409     0.4672  -0.730   0.46750
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 40810 on 92 degrees of freedom
## Multiple R-squared:  0.5852, Adjusted R-squared:  0.5537
## F-statistic: 18.54 on 7 and 92 DF, p-value: 0.000000000000003387
```

#구별등록인구수, 구별자동차등록대수가 유의해 보인다.

```
#주차장수와 단속건수 회귀분석 (귀무가설: 모든 회귀계수가 0이다)
fit<- lm(y~x1+x2+x3)
summary(fit)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -87567 -36079 -10526  32490 277303
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 53249.858 18433.314   2.889 0.00478 **
## x1           10.786     11.573   0.932 0.35364
## x2          353.687    156.297   2.263 0.02589 *
## x3           1.901     1.304   1.458 0.14810
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 58740 on 96 degrees of freedom
## Multiple R-squared:  0.1033, Adjusted R-squared:  0.07525
## F-statistic: 3.685 on 3 and 96 DF, p-value: 0.01464
```

```
## 귀무가설 기각, 최소한 하나는 0이 아님
##결정계수가 너무 낮음 -> 설명력 부족
####구별로 정제된 수치만 놓고 봤을 때는 서울시의 주차문제를 세밀하게 판단하기 어렵다.
#####동별로 살펴보자
```

- 부설주차장 단속 수 관계 시각화

부설주차장_단속수_관계시각화.R

user

2021-08-16

```
rm(list=ls())

par(mfrow=c(1,2))

x11()
par(mfrow=c(1,2))

park_paju<-c(14315, 15358, 16770, 18270,19633,65630) #파주 부설주차장 수
n_paju<-c(117887, 88774, 56670, 64323, 88028,74803) #파주 주정차 단속 수

#관계 시각화
plot(park_paju,ylim=c(0,150000),main="파주",axes=F,xlab="년도",ylab=" ")
legend("topright",c("부설주차장 수","주정차 단속 수"),col=c('black','red') ,lty=1)
axis(1, 1:6, 2014:2019)
axis(2)
lines(park_paju,lwd=2)
lines(n_paju,col='red',lwd=2)
points(n_paju,col='red')

cor(park_paju,n_paju) #파주 부설주차장 수와 주정차 단속 수의 상관관계
```

```
## [1] -0.2046745
```

```
park_uiwang<-c(1558,1627,2530,2581,2226,5679) #의왕 부설주차장 수
n_uiwang<-c(28652,16840,11904,15109,15554,14676 ) #의왕 주정차 단속 수

#관계 시각화
plot(park_uiwang,ylim=c(0,30000),main="의왕",axes=F,xlab="년도",ylab=" ")
legend("topright",c("주차장 수","주정차 단속 수"),col=c('black','red') ,lty=1)
axis(1, 1:6, 2014:2019)
axis(2)
lines(park_uiwang,lwd=2)
lines(n_uiwang,col='red',lwd=2)
points(n_uiwang,col='red')
```



```
cor(park_uiwang,n_uiwang) #의왕 부설주차장 수와 주정차 단속 수의 상관관계
```

```
## [1] -0.408439
```

```
#####
```

```
x11()
par(mfrow=c(1,2))
options(scipen=999)

park_anayang<-c(12024,12024,6076,6076,6639) #안양 부설주차장 수
n_anayang<-c(47750,52748,89737,90992,80995) #안양 주정차 단속 수

#관계 시각화
plot(park_anayang,ylim=c(0,100000),main="안양",axes=F,xlab="년도",ylab=" ")
legend("topright",c("부설주차장 수","주정차 단속 수"),col=c('black','red'),lty=1)
axis(1, 1:6, 2014:2019)
axis(2)
lines(park_anayang,lwd=2)
lines(n_anayang,col='red',lwd=2)
points(n_anayang,col='red')

cor(park_anayang,n_anayang) #안양 부설주차장 수와 주정차 단속 수의 상관관계
```

```
## [1] -0.9896681
```

● 서울시 주차수요 예측

주차장확보율 뺐

In [100]:

```
import numpy as np
import tensorflow as tf
from tensorflow.keras import Sequential
from tensorflow.keras.layers import Dense
from sklearn import datasets
import random
from sklearn.model_selection import train_test_split
import pandas as pd
import matplotlib.pyplot as plt

SEED = 10
random.seed(SEED)
np.random.seed(SEED)
tf.random.set_seed(SEED)
print("재현을 위한 시드 고정:", SEED)

park_info=pd.read_csv('C://Users//user//Desktop//서울특별시 동별_데이터들.csv',encoding='cp949')
park_info=park_info.iloc[:,2:-1]
park_info
```

...

In [101]:

```
park_info_stats=park_info.describe() #기초통계량
park_info_stats=park_info_stats.transpose()
park_info_stats
```

...

정규화

In [102]:

```
def norm(x):
    return (x-park_info_stats['mean'])/park_info_stats['std']
normed_park_info=norm(park_info)

normed_park_info
```

...

In [103]:

```
X_df=normed_park_info.iloc[:, [0,1,2,3,5,6]] #독립변수: 단속건수, 자동차등록대수, 노후주택30년이상,
y_df = normed_park_info.iloc[:,[4]] #종속변수: 주차장면수
```

In [104]:

```
X_df
```

...

In [105]:

y_df

...

train, test로 나눔

In [106]:

```
X_train, X_test, y_train, y_test = train_test_split(X_df, y_df, test_size=0.2, shuffle=True, random
print("훈련데이터 X와 y의 차원", X_train.shape, y_train.shape)
print("테스트데이터 X와 y의 차원", X_test.shape, y_test.shape)
```

훈련데이터 X와 y의 차원 (102, 6) (102, 1)
 테스트데이터 X와 y의 차원 (26, 6) (26, 1)

In [107]:

```
model=Sequential()
model.add(Dense(units=4,activation='linear',input_dim=X_df.shape[1]))
model.add(Dense(units=1,activation='linear'))
model.summary()
```

Model: "sequential_3"

Layer (type)	Output Shape	Param #
dense_6 (Dense)	(None, 4)	28
dense_7 (Dense)	(None, 1)	5
Total params: 33		
Trainable params: 33		
Non-trainable params: 0		

In [108]:

```
model.compile(optimizer='adam', loss='mse', metrics=['mae', 'mse'])

history = model.fit(X_train, y_train, epochs=1000, batch_size=64, validation_split=0.25) # 25%가 경
```

```
Epoch 1/1000
2/2 [=====] - 0s 104ms/step - loss: 1.9537 - mae: 0.9734
- mse: 1.9537 - val_loss: 1.7753 - val_mae: 1.0460 - val_mse: 1.7753
Epoch 2/1000
2/2 [=====] - 0s 25ms/step - loss: 1.9120 - mae: 0.9613
- mse: 1.9120 - val_loss: 1.7430 - val_mae: 1.0365 - val_mse: 1.7430
Epoch 3/1000
2/2 [=====] - 0s 24ms/step - loss: 1.8724 - mae: 0.9500
- mse: 1.8724 - val_loss: 1.7117 - val_mae: 1.0274 - val_mse: 1.7117
Epoch 4/1000
2/2 [=====] - 0s 35ms/step - loss: 1.8339 - mae: 0.9389
- mse: 1.8339 - val_loss: 1.6817 - val_mae: 1.0187 - val_mse: 1.6817
Epoch 5/1000
2/2 [=====] - 0s 20ms/step - loss: 1.7896 - mae: 0.9277
- mse: 1.7896 - val_loss: 1.6541 - val_mae: 1.0107 - val_mse: 1.6541
Epoch 6/1000
2/2 [=====] - 0s 24ms/step - loss: 1.7629 - mae: 0.9191
- mse: 1.7629 - val_loss: 1.6279 - val_mae: 1.0037 - val_mse: 1.6279
Epoch 7/1000
2/2 [=====] - 0s 20ms/step - loss: 1.7368 - mae: 0.9101
```

In [109]:

```
model.weights
```

...

In [110]:

```
model.evaluate(X_test, y_test)
```

```
1/1 [=====] - 0s 16ms/step - loss: 0.1098 - mae: 0.2407 - m
se: 0.1098
```

Out[110]:

```
[0.10983861237764359, 0.24072010815143585, 0.10983861237764359]
```

In [111]:

```
model.evaluate(X_train, y_train)
```

```
4/4 [=====] - 0s 4ms/step - loss: 0.1394 - mae: 0.2511 - ms
e: 0.1394
```

Out[111]:

```
[0.13940195739269257, 0.2511116564273834, 0.13940195739269257]
```

In [112]:

```
#에폭별 비용함수 변화
plt.figure(figsize=(10,10))
plt.plot(range(20,len(history.history['loss'])), history.history['loss'][20:], label='Train', color=
plt.plot(range(20,len(history.history['val_loss'])), history.history['val_loss'][20:], label='Valida
plt.xlabel('Epochs')
plt.ylabel('mse')
plt.legend()
plt.show()
```

...

In [113]:

```
model.predict(X_test) #정규값(예측값)
predict=pd.DataFrame(model.predict(X_test),columns=['predict_정규값'])
predict
```

...

In [114]:

```
y_pred=(predict*12002.685134)+14870.507812 #평균:14870.507812 표준편차:12002.685134 #정규화 한 것을
y_pred=y_pred.rename(columns={'predict_정규값':'predict_원래값'})
y_pred
#np.round(y_pred)

y_pred_origin=y_pred.applymap(np.round)
y_pred_origin
```

...

In [115]:

```
y_test #정규값(관측값) #평균:14870.507812 표준편차:12002.685134
```

...

In [116]:

```
y_test_1=(y_test*12002.685134)+14870.507812 #평균:14870.507812 표준편차:12002.685134 #정규화 한 것
y_test_1=pd.DataFrame(y_test_1)
y_test_1=y_test_1.rename(columns={'parking_c':'parking_c원래값'})
y_test_1_1=y_test_1.applymap(np.round)
y_test_1_1
```

...

In [118]:

```
from sklearn.metrics import r2_score
print('설명력:',r2_score(y_test_1_1,y_pred_1_1))
```

설명력: 0.8317009511794283

● 주차 수요에 영향을 미치는 변수 상관분석 코드

In [1]:

```
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns

import tensorflow as tf
from tensorflow import keras
from tensorflow.keras import layers
```

In [2]:

```
park_info=pd.read_csv('C://Users//user//Desktop//서울특별시 동별_데이터들.csv',encoding='cp949')
park_info=park_info.iloc[:,[2,3,4,5,6,7,8,9]]
```

In [3]:

```
sns.pairplot(data=park_info)
```

In [4]:

```
corr=park_info.corr()
corr
```

In [5]:

```
sns.heatmap(corr,annot=True)
```