

따릉이 대여소별 수요예측

Jingle Bike

고정* 김서* 김수* 김진* 오덕* 원용* 이은*





Why Public Bike?

*대중교통 분담률 향상

※ 사람들이 통행할 때 하루 중 이용하는 교통수단의 분포를 비율로 나타낸 것을 의미 / 친환경 교통사용량 측정 지표

건강한 도시 형성과 환경문제 완화 & 자전거 이용활성화 등을 위한 정책적 대안

2015년 시범사업 이후 서울시 공공자전거에 대한 관심과 수요는 지속적으로 증가

So, 서울시曰 “공공자전거 대여소와 자전거를 지속 확충 예정”

공공자전거 이용에 영향을 미치는 요인들을 파악할 필요





CONTENTS



01

Overview

02

Process

03

Analysis

04

Modeling

05

Insights





01 Overview

“Right Place, Right Amounts”

-목적 : 서울시 공공자전거 '따릉이'의 **사용량을 예측**, 서비스 질 제고

-Target User: 서울시 따릉이 정책 담당자 or 민간 서비스 제공자 or Share-Mobility 사업자

-기대효과: ① 유휴 수량 최소화를 통한 **예산 집행 효율화**

② 거치소간 대여, 반납 불균형 해소를 통한 따릉이 **운송 비용 절감**

③ 비화석 에너지원 사용 장려를 통한 **교통환경 및 공기질 개선 기여**



02 Process_01Variables

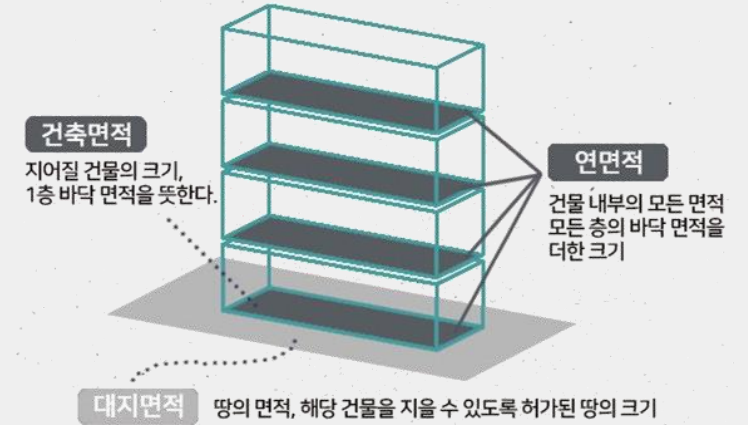
상권분석

연면적

* 건물 내부의 모든 면적

주요시설까지의 거리

* 대학교, 지하철역, 공공시설(체육센터 등)



유동인구

생활인구

* 주민등록에 등록된 거주자가 아니라 특정 시간, 특정 지역에 실제 생활하는 인구

지하철 승하차 인원

대여소 지리특성

경사도

자전거 도로 인접성

하천까지의 거리





02 Process_02Tools

Language

Python(환경: Windows10, Jupyter Notebook) / R / SQL

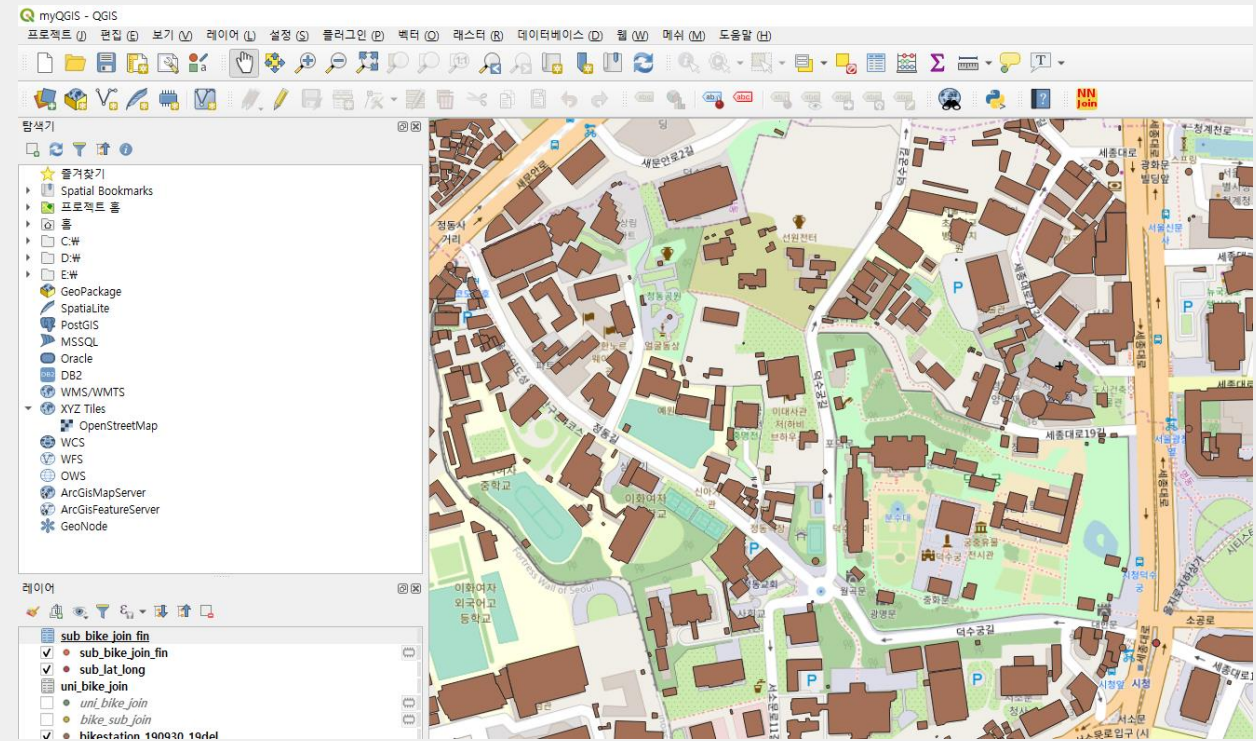
주요 Library : NumPy, Pandas, Matplotlib, Scikit-Learn, Keras

QGIS

*강력하고 편리한 지리정보체계 오픈소스 소프트웨어

연면적/최단거리 추출

연면적 : ① 벡터계산('버퍼') 최단거리 : ② NNJoin plugin 설치





02 Process_03RoadMap



적용이론 / 데이터 수집 / 전처리

변수 확정

분석 모델 선정

모델 학습 / 검증

시각화 / 결론 / 제언



03 Analysis

적용 이론

데이터 수집/전처리

변수 확정



선행연구 | 변수 선정

석사학위논문

서울시 공공자전거 이용에
영향을 미치는 물리적 환경 요인에 관한 연구
-대여소별 거리에 따른 요인의 영향력 차이를 중심으로
A Study on the Characteristics of Physical Environments
Affecting the Usage of Public Bike in Seoul, Korea
-Focusing on the Different Influences of Factors
by Distance to Bike Station

사 경 은

한양대학교 대학원



03 Analysis

데이터 수집 / 전처리

변수 확정

구분	수집 대상	
대여소 정보	대여소ID, 대여소명, 거치대 수, 위치정보(위도, 경도)	
	대여량, 반납량	
상권분석	연면적(반경 25/50/75/100m 이내 공동, 단독주택, 근린, 상업, 업무시설)	
유동인구	생활인구 (내국인 / 장기체류외국인 포함)	
	지하철역 평균 승하차 인원	
대여소 지리특성	평균경사도	
	최단거리	하천
		자전거전용도로
		주요 공원
		공공시설
		지하철역
대학교		



03 Analysis

데이터 수집 / 전처리

변수 확정

구분	수집 대상
대여소 정보	대여소ID, 대여소명, 거치대 수, 위치정보(위도, 경도)
	대여량, 반납량

대여소ID	대여소명	대여소주소	위도	경도	rent_cnt	return_cnt
101	101. (구)합정동 주민센터	서울특별시 마포구 동교로8길 58	37.549561	126.905754	114	123
102	102. 망원역 1번출구 앞	서울특별시 마포구 월드컵로 72	37.555649	126.910629	94	111
103	103. 망원역 2번출구 앞	서울특별시 마포구 월드컵로 79	37.554951	126.910835	86	86
					85	104
					113	125
					138	125
					137	157
					119	96



03 Analysis

데이터 수집 / 전처리

변수 확정

구분	수집 대상
상권분석	연면적 (반경 25/50/75/100m 이내 공동, 단독주택, 근린, 상업, 업무시설)

rackID	rackName	rackCount	p25m	s25m	p50m	s50m	p75m	s75m	p100m	s100m	g25m
101	101. (구)힐	5	1371.62	1245.51	7156.67	1829.57	16666.71	3343.56	22512.39	7114.25	2296.89



03 Analysis

데이터 수집 / 전처리

변수 확정

구분	수집 대상
유동인구	생활인구 (내국인 / 장기체류외국인 포함)
	지하철역 평균 승하차 인원

LPpopulat	LFpopulati	meanOn	meanOff
25073.43	1073.872	16163.5	17644.25
24026.86	1032.047	14833.5	15978.25
24395.18	1101.772	13336.25	14704
23212.19	1081.223	9704.5	10317.5
24450.74	1048.857	15605.6	16836.8



03 Analysis

데이터 수집 / 전처리

변수 확정

구분	수집 대상	
대여소 지리특성	최단거리	평균경사도
		하천
		자전거전용도로
		주요 공원
		공공시설
		지하철역
		대학교

slope	river_m	bikeroad_m	mainpark	genpark_n	pub_m	sub_m	univ_m
2-7%	992.7317	3.417791	889	234	1632.264	72.86659	1442.41



03 Analysis

데이터 수집 / 전처리

변수 확정

파일형식		대여소 정보	상권분석	유동인구	대여소 지리특성																																
					평균경사도	최단거리																															
						*하천	그 외																														
파일형식		CSV	CSV, Shape(.shp)	CSV, Shape(.shp)	CSV, Shape(.shp)																																
데이터 타입	범주형	ID, 대여소명	-	-	경사도 자료 → 최근접 평균경사도 (예: 0~2%, 국토부 기준)	-	-																														
	연속형	거치대수, 위·경도, 대여·반납량	토지이용면적, 연면적	생활인구, 지하철승하차인원	-	하천 변곡점별 위·경도	자전거도로변곡점, 지하철역, 공공시설 등의 위·경도																														
Tool		Jupyter Notebook	Jupyter Notebook, QGIS	MySQL, QGIS	QGIS	QGIS	QGIS																														
Key Point		· 2018년 12월 마감 기준 설치된 거치대 정보 활용 (2019년 운영 데이터 필요)	· QGIS - 버퍼 기능 사용 (영역선택 가능) · 대여소ID 기준 범위별 Group by 및 정제	· 행정동별 월별 요일별 생활인구 / daycount * GROUP BY · 월별 역별 평균승하차인원 * AVG()	· 대여소 기준 10cm 이내 최근접 경사도 NNJoin 실행	· 주요 하천에 속하는 지류, 호수, 연못 등 추가적인 요소 삭제 · 대여소 기준 NNJoin 실행	· 대여소ID 기준 NNJoin 실행																														
결과 / 예시		<table border="1"> <thead> <tr> <th>구</th> <th>ID</th> <th>명</th> <th>주소</th> <th>위도</th> <th>경도</th> <th>시작일</th> <th>사용여부</th> </tr> </thead> <tbody> <tr> <td>마포</td> <td>154</td> <td>구청역</td> <td>...</td> <td>37.5609</td> <td>126.9055</td> <td>년월일 시분</td> <td>Y/N</td> </tr> </tbody> </table>	구	ID	명	주소	위도	경도	시작일	사용여부	마포	154	구청역	...	37.5609	126.9055	년월일 시분	Y/N	<table border="1"> <thead> <tr> <th rowspan="2">ID</th> <th colspan="2">100m 면적</th> </tr> <tr> <th>공동주택</th> <th>단독주택</th> </tr> </thead> <tbody> <tr> <td>101</td> <td>22512</td> <td>7114</td> </tr> <tr> <td>102</td> <td>15177</td> <td>9170</td> </tr> <tr> <td>...</td> <td>...</td> <td>...</td> </tr> </tbody> </table>	ID	100m 면적		공동주택	단독주택	101	22512	7114	102	15177	9170	(예) 00동 1월 금요일 평균 - 내국인 : 25073.43명 - 외국인 : 1073.872명 - 평균승차: 16163.5명 - 평균하차: 17644.25명			
구	ID	명	주소	위도	경도	시작일	사용여부																														
마포	154	구청역	...	37.5609	126.9055	년월일 시분	Y/N																														
ID	100m 면적																																				
	공동주택	단독주택																																			
101	22512	7114																																			
102	15177	9170																																			
...																																			



03 Analysis

데이터 수집 / 전처리_QGIS 화면 예시

변수 확정



< 따릉이 대여소 기준 반경별 연면적 정보 추출(QGIS) >

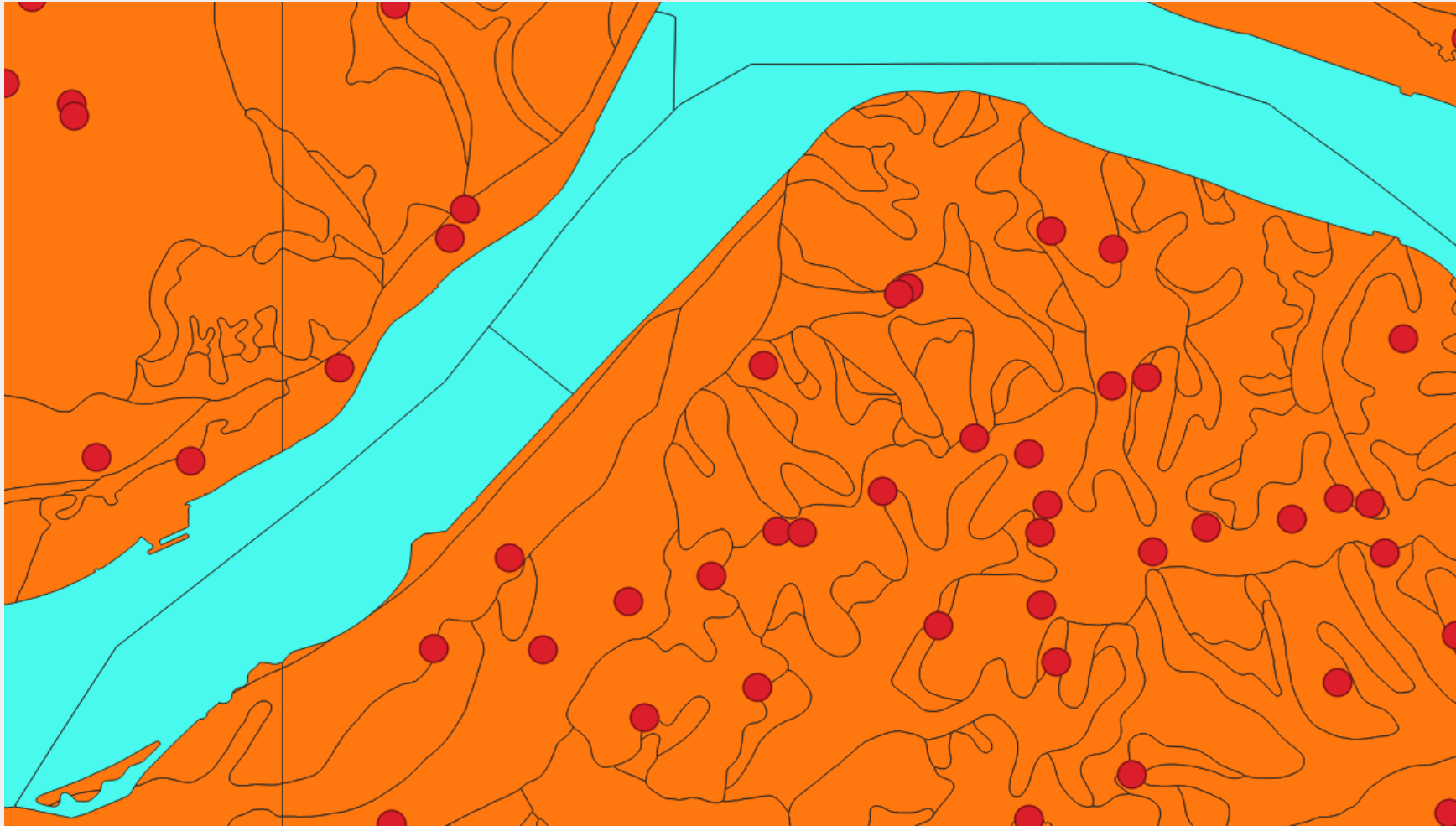
벡터 - 면적 추출



03 Analysis

데이터 수집 / 전처리_QGIS 화면 예시

변수 확정



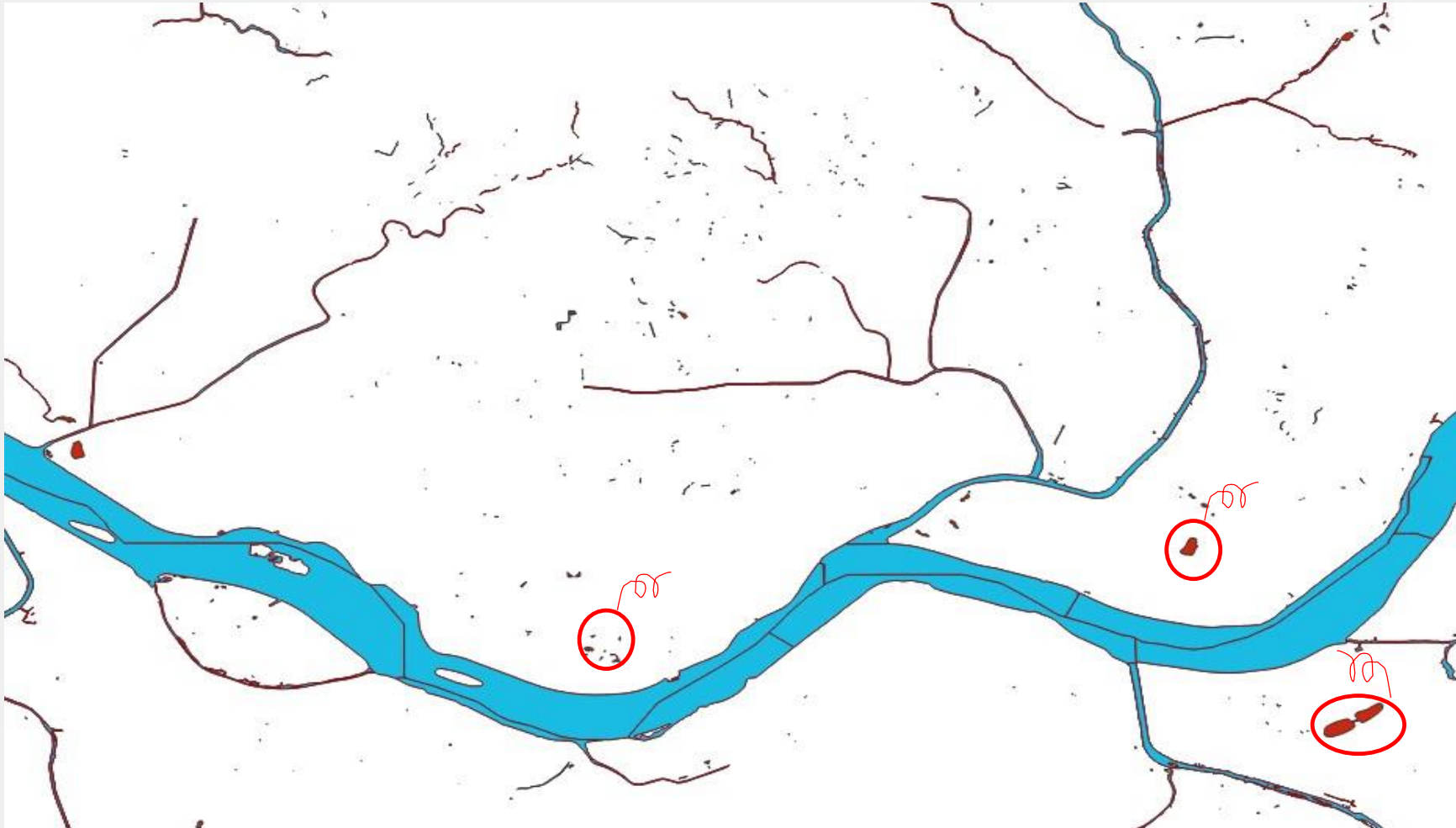
< 따릉이 대여소 기준 경사도(QGIS) >



03 Analysis

데이터 수집 / 전처리_QGIS 화면 예시

변수 확정



< 하천 위치정보와 따릉이 대여소 위치정보 지도(QGIS) >

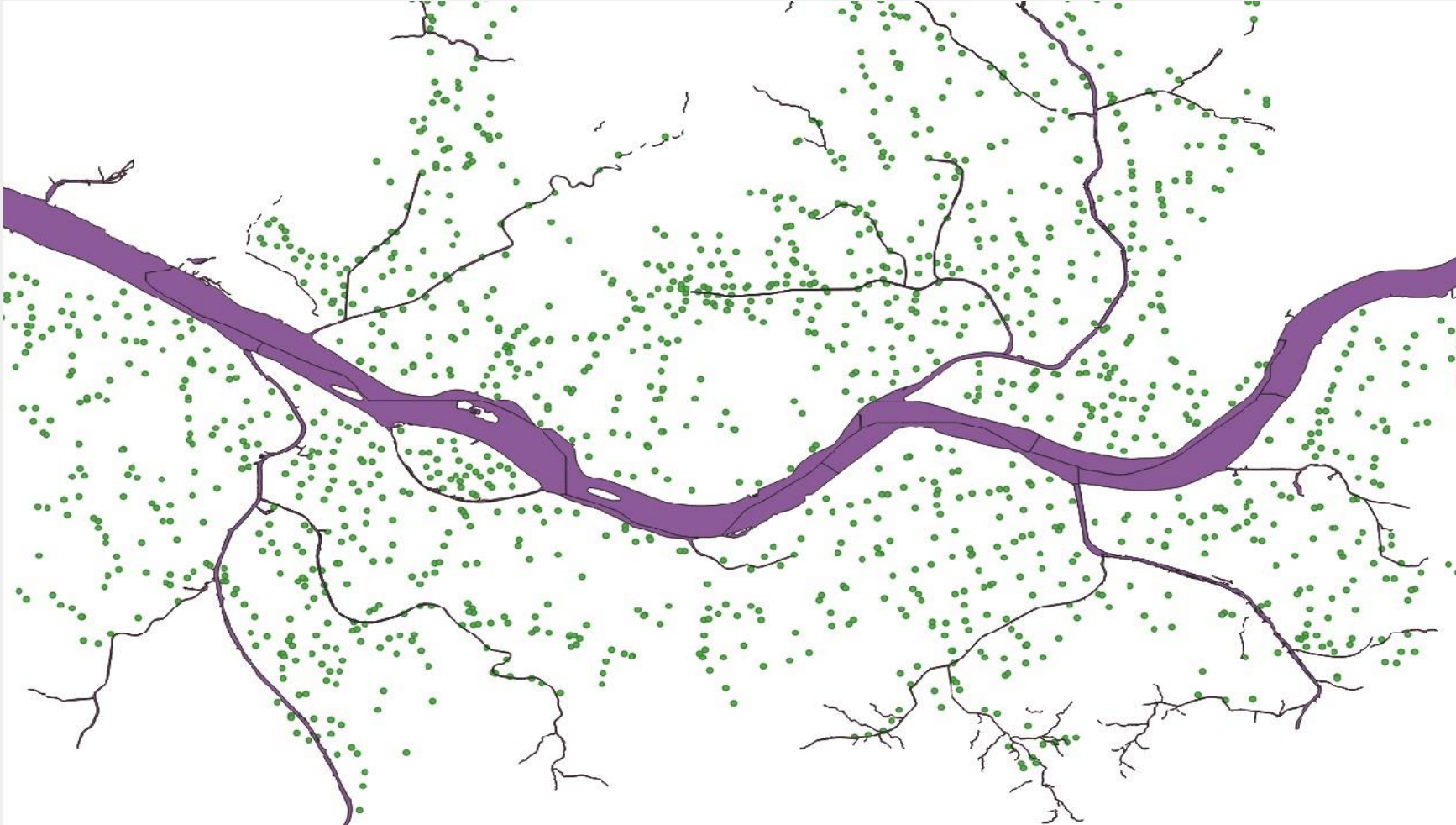
NNJoin - 최단거리



03 Analysis

데이터 수집 / 전처리_QGIS 화면 예시

변수 확정



< 하천 위치정보와 따릉이 대여소 위치정보 지도(QGIS) >

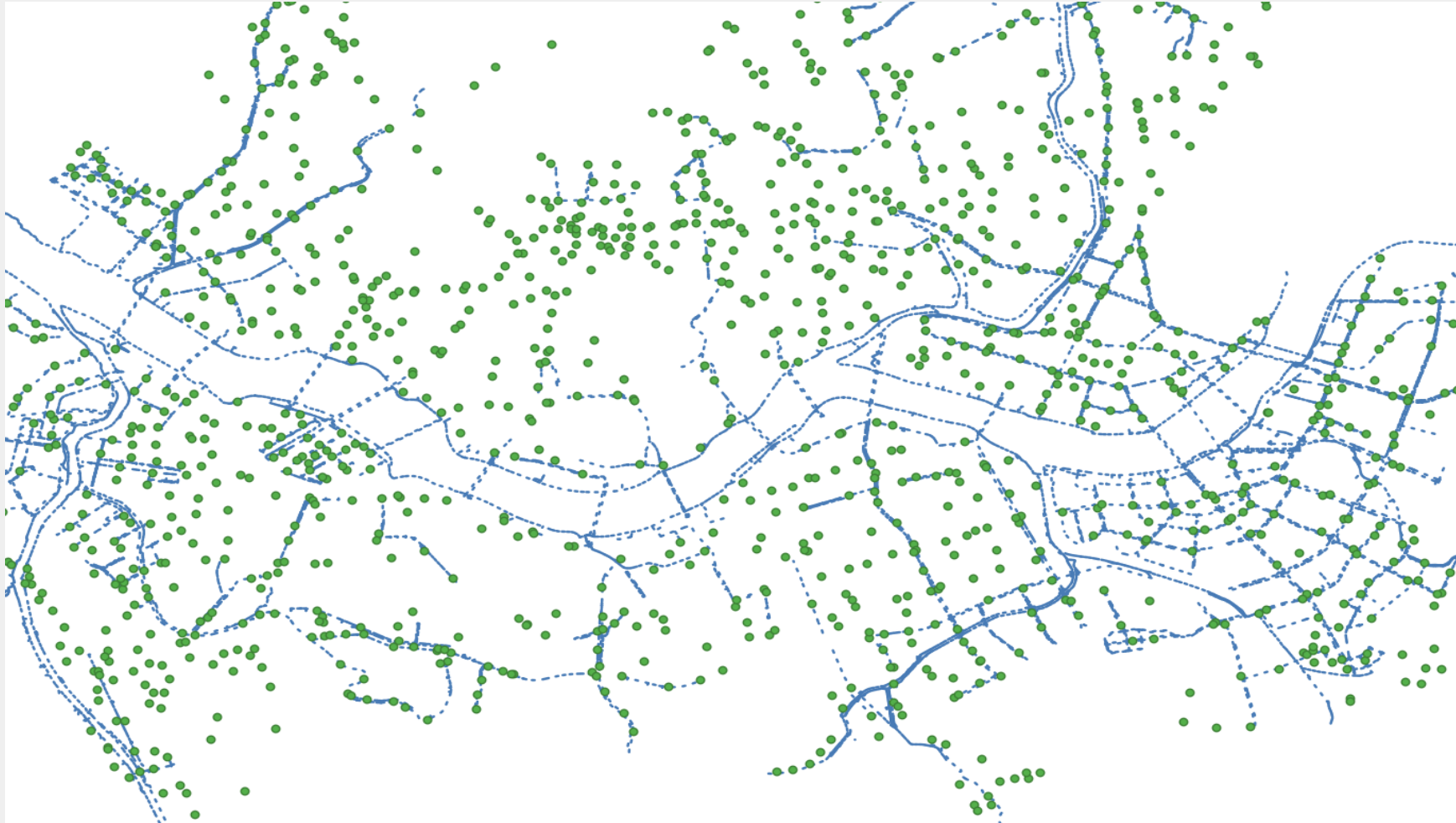
NNJoin - 최단거리



03 Analysis

데이터 수집 / 전처리_QGIS 화면 예시

변수 확정



< 자전거도로 위치정보와 따릉이 대여소 위치정보 지도(QGIS) >

NNJoin - 최단거리



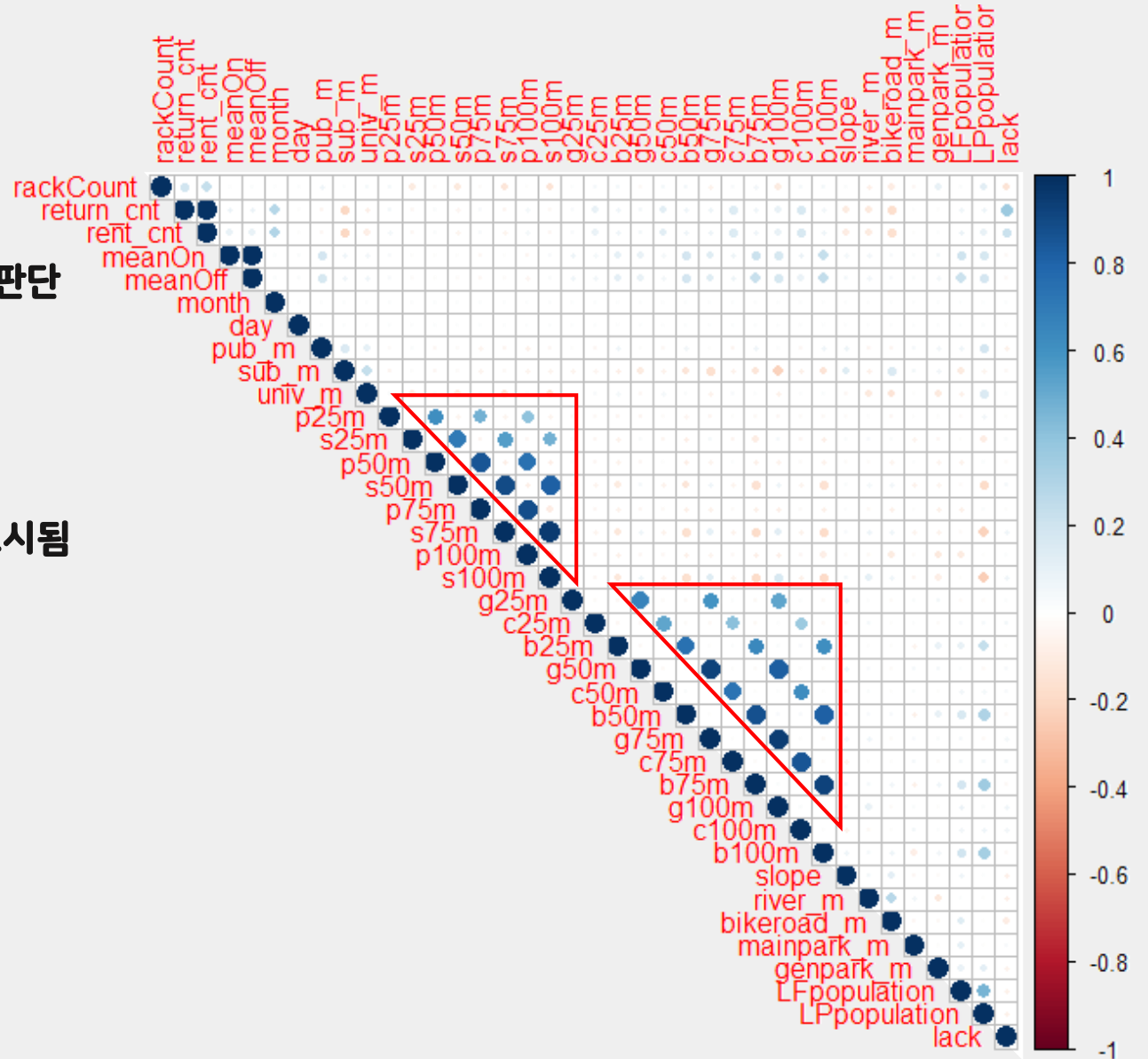
03 Analysis

● 변수 확정_수집된 변수들의 상관관계 분석(R Studio)

√ 전체 변수들의 상관관계는 -0.4~+0.4 사이로 낮다고 판단

√ 25m, **50m**, 75m, 100m 연면적 변수

- 이전 범위의 특성을 포함하기 때문에 상관관계 높게 표시됨





03 Analysis

● 변수 확정_One-hot Encoding & Normalization

범주형 자료
월, 요일, 경사도
one-hot encoding

rackID	slope_0-2%	slope_2-7%	slope_7-15%	slope_15-30%	slope_30-60%	slope_60-100%	slope_경고?	day_Monday	day_Tuesday
101	0	0	0	0	0	0	1	0	0
101	0	0	0	0	0	0	1	1	0
101	0	0	0	0	0	0	1	0	0
101	0	0	0	0	0	0	1	0	0
101	0	0	0	0	0	0	1	0	0
...
3542	0	0	1	0	0	0	0	0	0
3542	0	0	1	0	0	0	0	0	0
3542	0	0	1	0	0	0	0	0	0
3542	0	0	1	0	0	0	0	0	1
3542	0	0	1	0	0	0	0	0	0

연속형 자료
연면적, 거리, 생활인구, 지하철 이용 등
Normalization

	count	mean	std	min	25%	50%	75%	max
p25m	117194.0	2430.282956	11095.588000	0.000000	0.000000	0.000000	0.000000	243460.860000
s25m	117194.0	84.097510	266.851180	0.000000	0.000000	0.000000	0.000000	4159.160000
p50m	117194.0	7365.194095	20088.864110	0.000000	0.000000	64.010000	7970.400000	385944.250000
s50m	117194.0	514.930785	984.874653	0.000000	0.000000	0.000000	615.920000	6396.680000
p75m	117194.0	14819.848372	28595.705104	0.000000	0.000000	4155.935000	19994.200000	559498.380000
s75m	117194.0	1557.417668	2420.535435	0.000000	0.000000	315.605000	2278.360000	14910.050000
p100m	117194.0	25418.836528	39443.404693	0.000000	779.950000	12463.831750	36151.020000	699974.860000

	p25m_normed	s25m_normed	p50m_normed	s50m_normed	p75m_normed	s75m_normed	p100m_normed
0	-0.095413	4.352285	-0.01038	1.334829	0.064585	0.737912	-0.073687
1	-0.095413	4.352285	-0.01038	1.334829	0.064585	0.737912	-0.073687





04 Modeling

분석모델 선정

모델학습 / 검증

Step1. Scikit Learn 탐색

√ (다중회귀) 모델 탐색

LinearRegression

Ridge

Lasso

Elastic

LassoLars

SGDRegressor

XGBRegressor

Step2. Least Error

√ MAE(Mean Absolute Error) Best값 찾기

lr 0.53

ridge 0.53

lasso 0.64

elastic 0.62

LassoLars 0.64

SGDRegressor 0.53

XGBRegressor 0.28

결정계수 $R^2 = 0.79$

Step3. Keras 탐색

√ 더 정교화된 모델 필요

√ 다층으로 구성된 Keras 모델 개발

◆ 1차 Test 모델 spec

Activation Func.		Relu
Layer 갯수		In/Out : 각 1 Hidden : 3
compiler	Loss	mse
	Optimizer	adam
	Metrics	mae
Batch_size		100

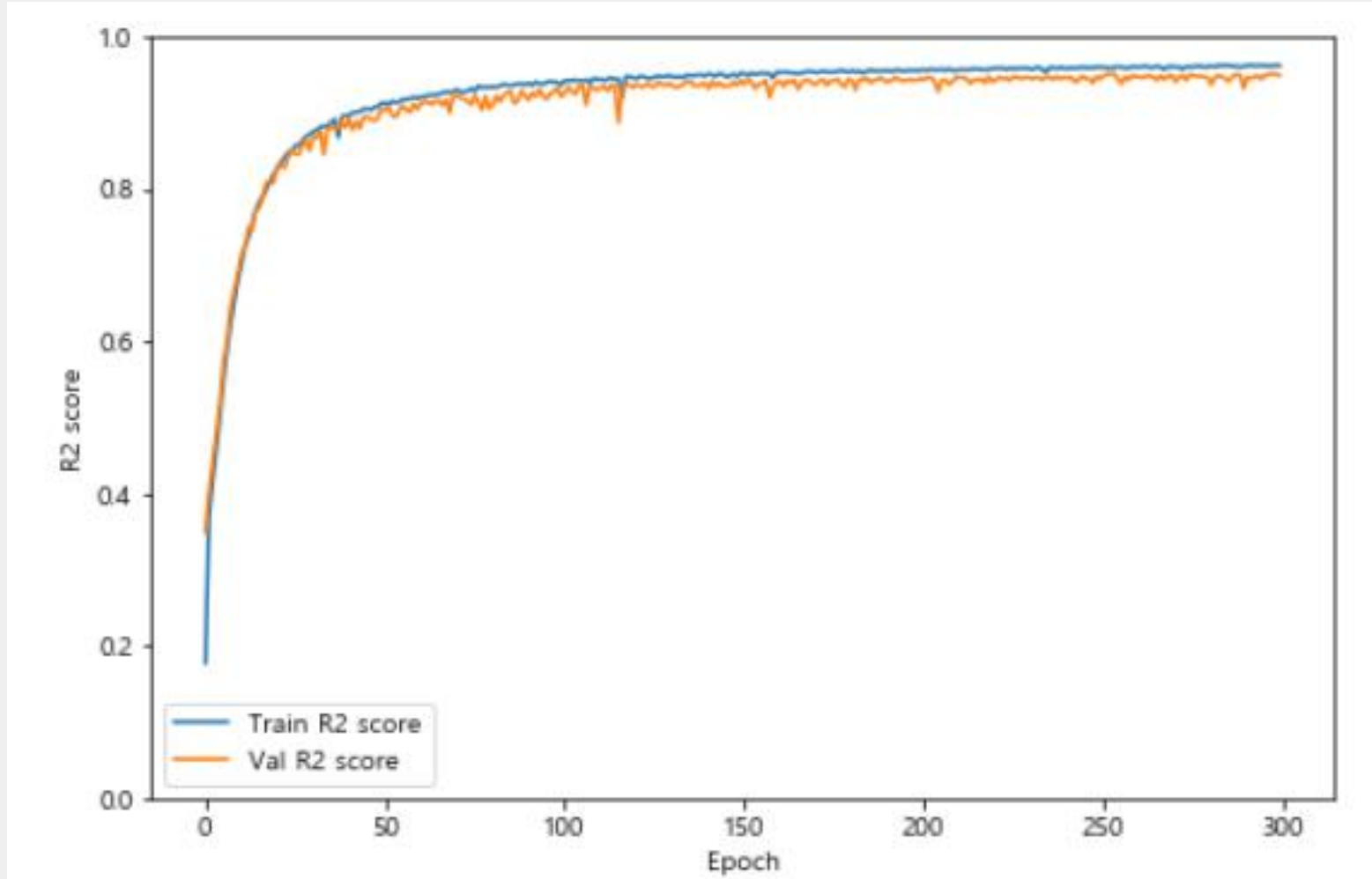


05 Insights

● 시각화

● 결론

● 제언



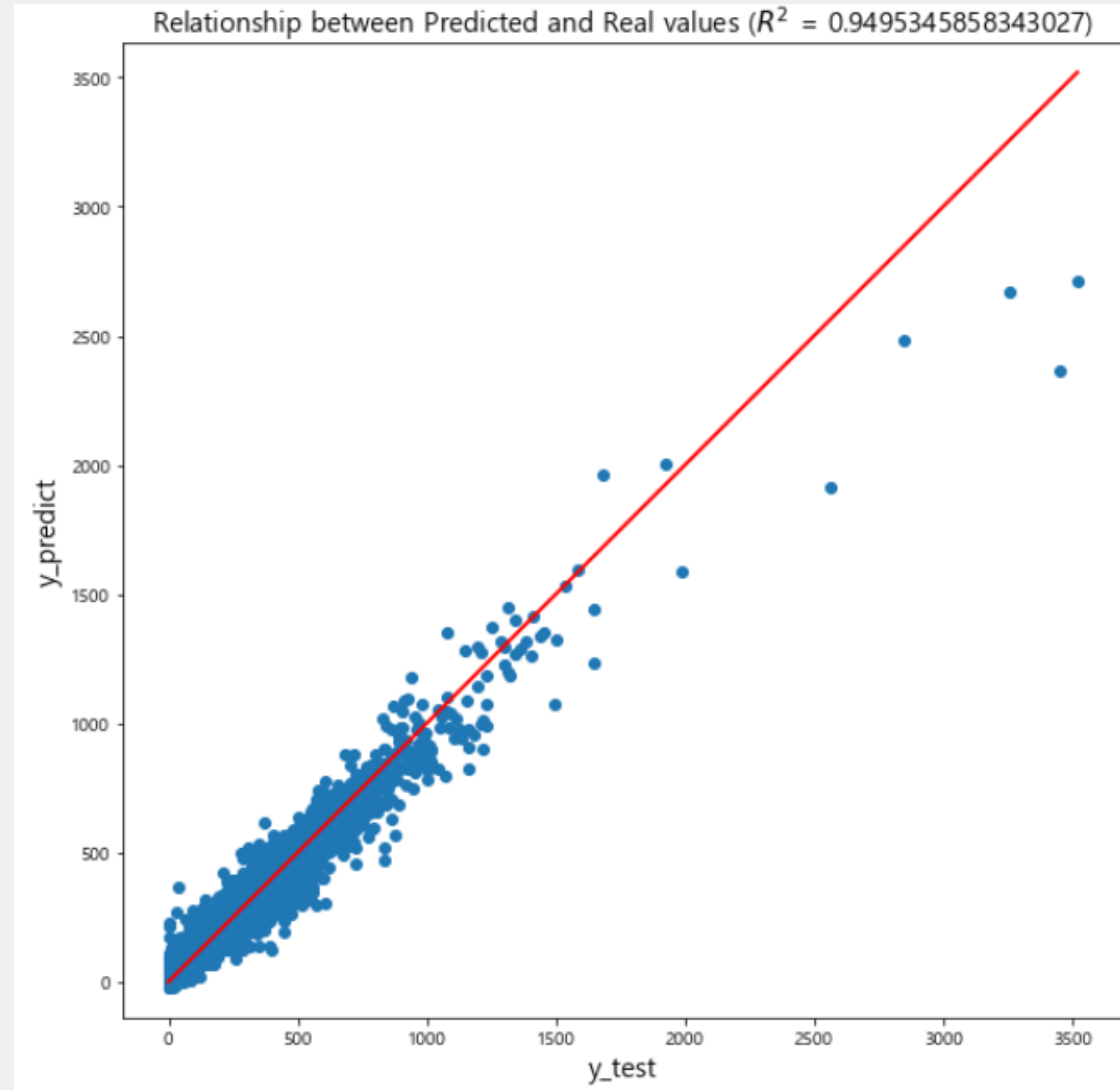


05 Insights

시각화

결론

제언



$R^2 = 0.95$



05 Insights

결론

제언

Parameter		값
Activation Func.		Relu
Layer 갯수		In/Out : 각 1 Hidden : 1
Node 갯수		128
Optimizer		adam
compiler	Loss	mse
	Metrics	mae
		R^2 score (custom)
epoch		300
Validation_split		0.2



05 Insights

제언_ "따름이 예보"

**대여소 후보지역의 독립변수를 측정 및 데이터 정제 후 입력할 경우
미래의 사용량이 예측되어 Right Place, Right Amounts 따름이 행정 달성!**

따름이 대여소별 수요예측

대여소 선택

1702 . 녹전역 1번출구 앞

대여소번호	구분	실제 대여량	실제 반납량	예측 대여량	예측 반납량	대여 오차	반납 오차
1702	1월	133	167	268	188	-135	-21
	2월	189	195	271	211	-82	-16
	3월	364	411	445	438	-81	-27
	4월	508	603	639	688	-131	-85
	5월	719	1012	834	891	-115	121
	6월	767	1024	857	869	-90	155
	7월	838	1028	774	855	64	173
	8월	744	918	819	815	-75	103
	9월	717	833	702	847	15	-14
	10월	982	1064	799	1032	183	32
	11월	704	814	638	682	66	132
	월요일	941	1131	1081	1091	-140	40
	화요일	1054	1266	1126	1084	-72	182
	수요일	1009	1193	915	1071	94	122
	목요일	921	1183	964	1073	-43	110
	금요일	921	1194	1040	1102	-119	92
	토요일	1031	1132	1019	1086	12	46
일요일	788	970	900	1007	-112	-37	



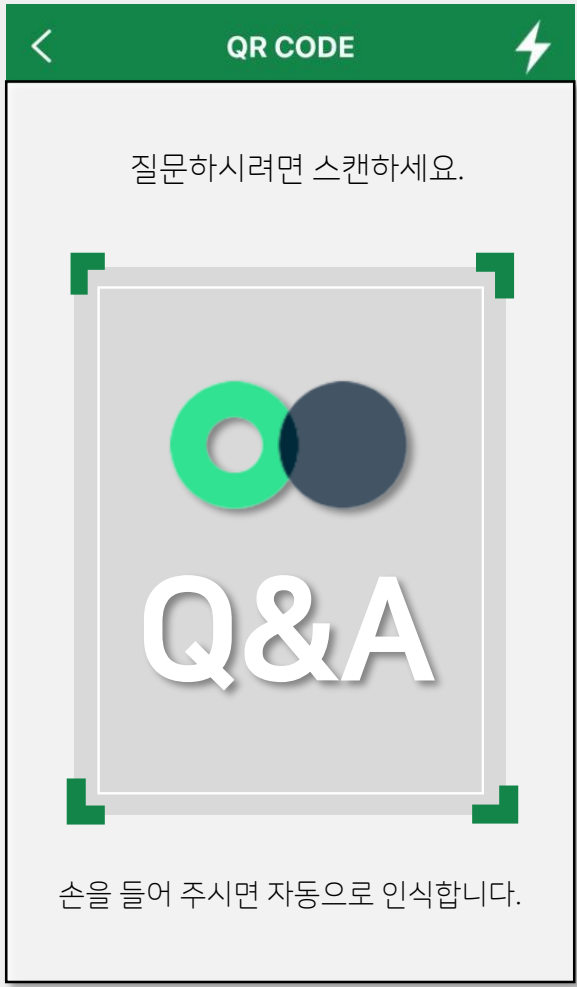
개선사항 및 향후 계획

Issue 1 : 날씨 변수 미반영 (대여소간 날씨 정보 동일) ⇨ 월 단위 변수로 대체

Issue 2 연도별 대여량 증가율 반영 미흡 (시계열 정보 누락)

Plan 1 Python 과 QGIS 연동 통한 원스톱 예측 시스템 구현

Plan 2 킥보드 등 다양한 라스트 마일 모빌리티 산업 적용





THANK YOU