



김태현



목차

Contents

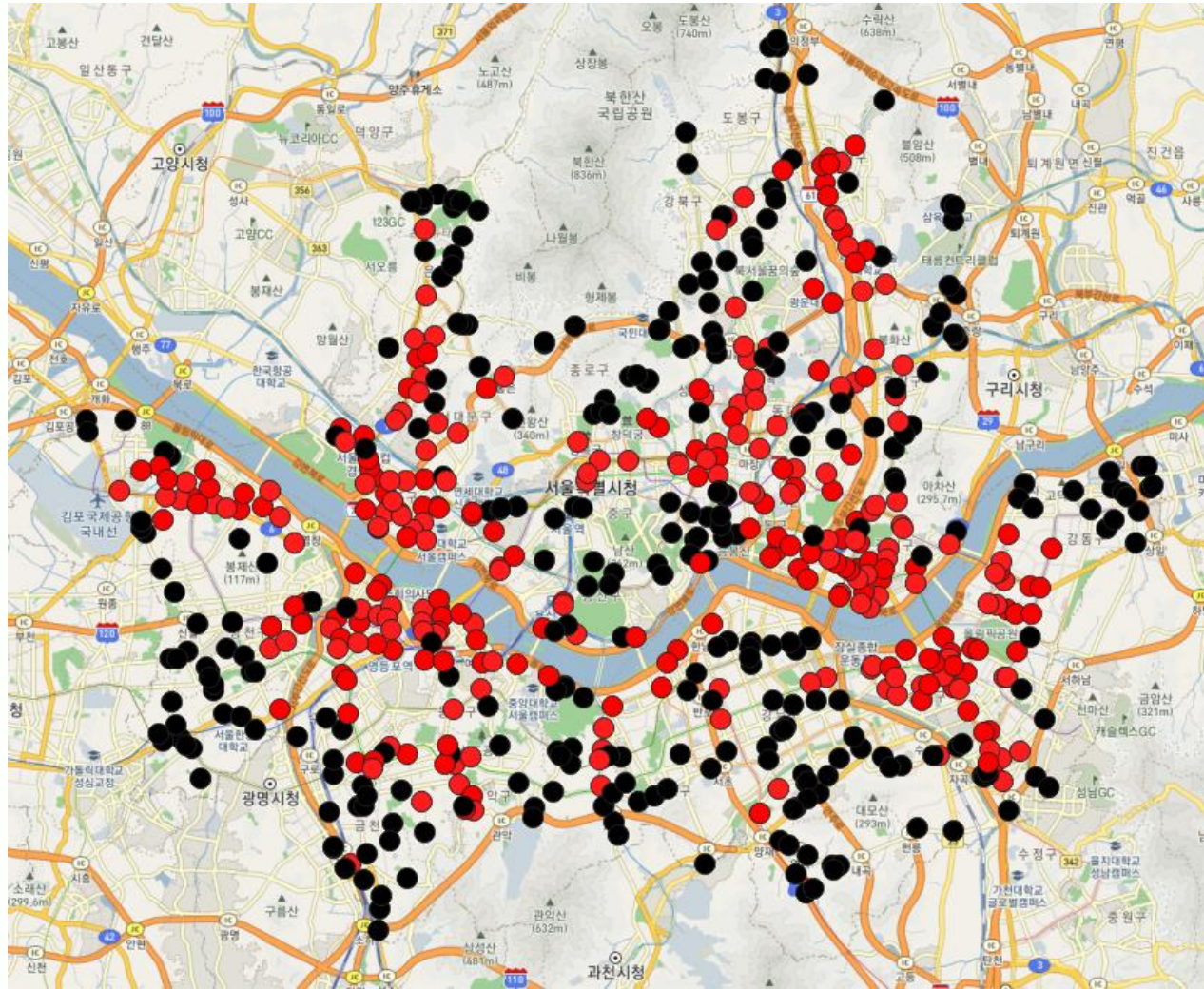
1. EDA

2. 분석 & 결과

The image features a solid blue background with a central square of a slightly lighter shade of blue. The letters 'EDA' are centered within this square in a white, bold, sans-serif font. The overall design is minimalist and modern, with sharp geometric lines and a monochromatic color scheme.

EDA

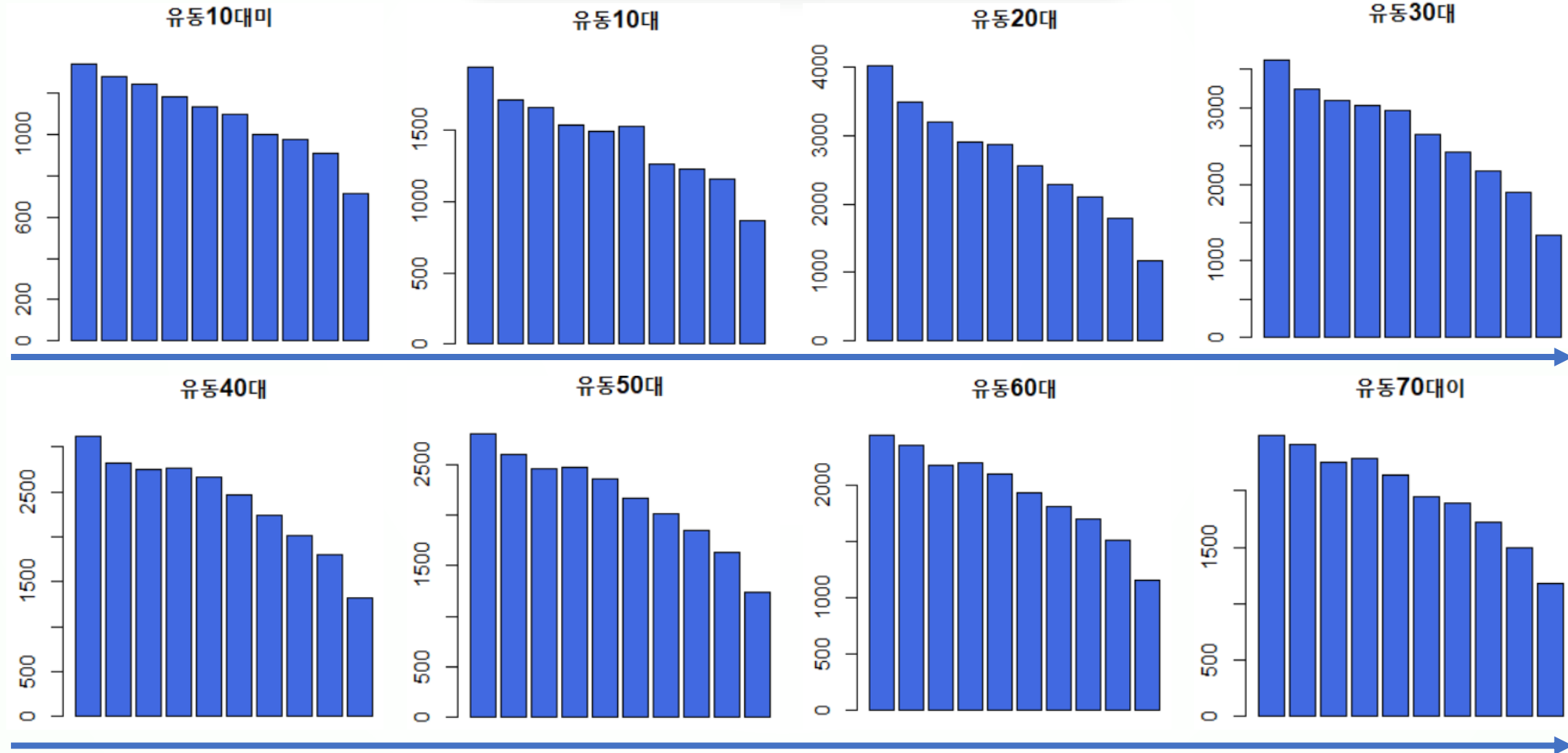
대여소 이용량 상위20%, 하위 20%



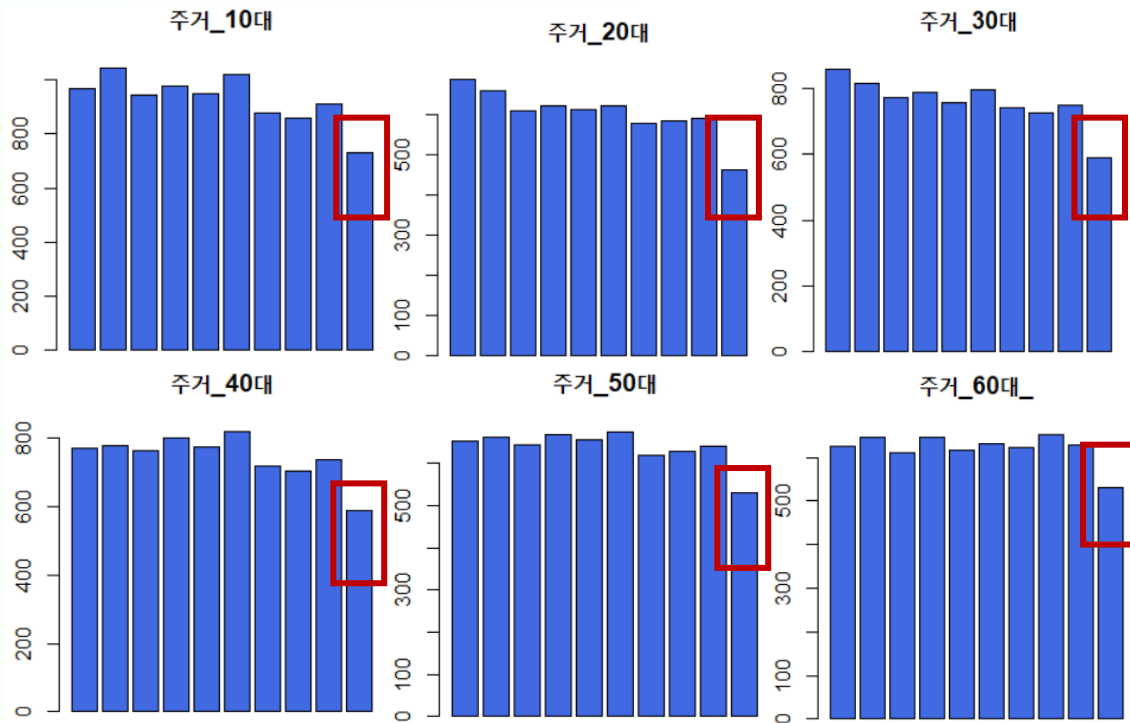
변수들의 분포(유동인구)

*x축 : 우측으로 갈수록 이용량 ↓

EDA

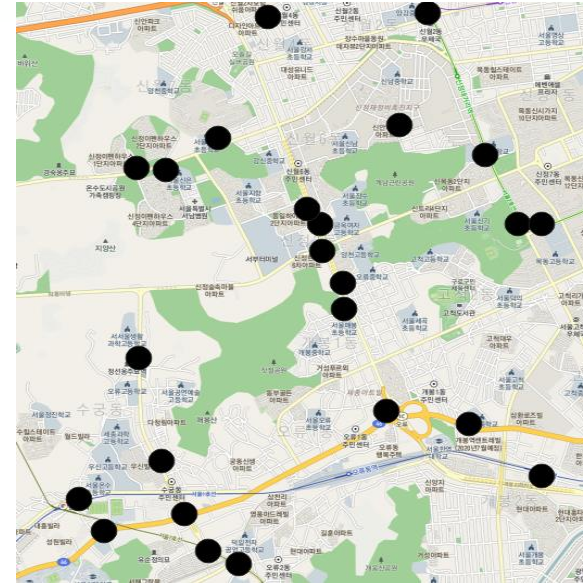
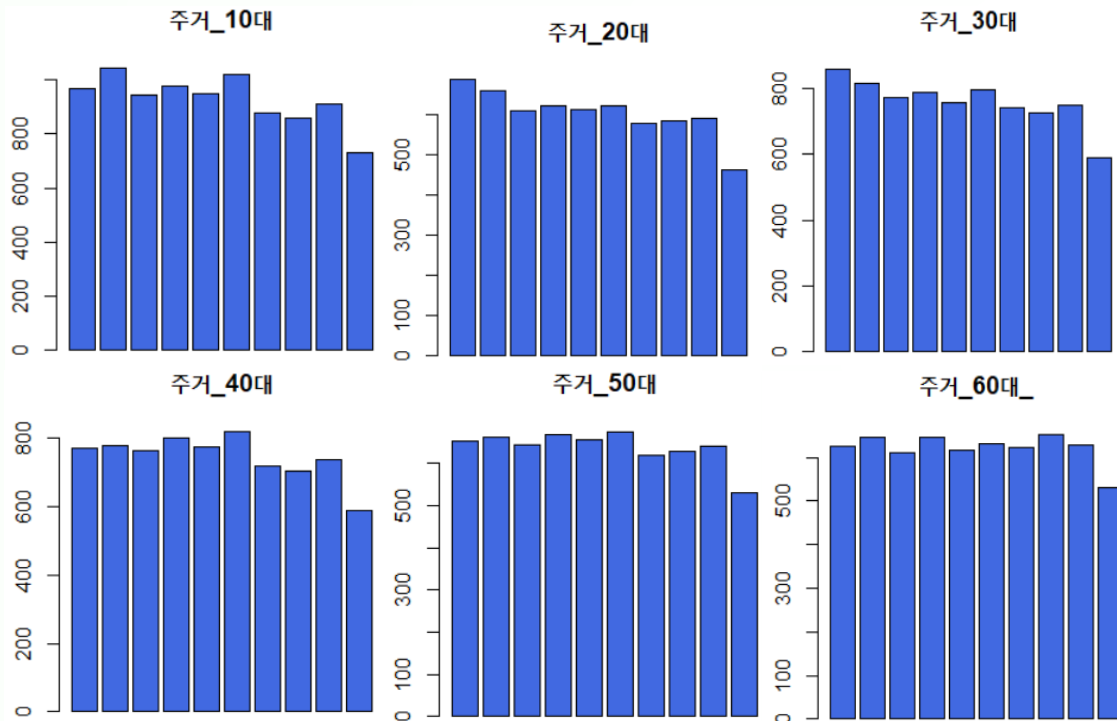


변수들의 분포(주거인구)

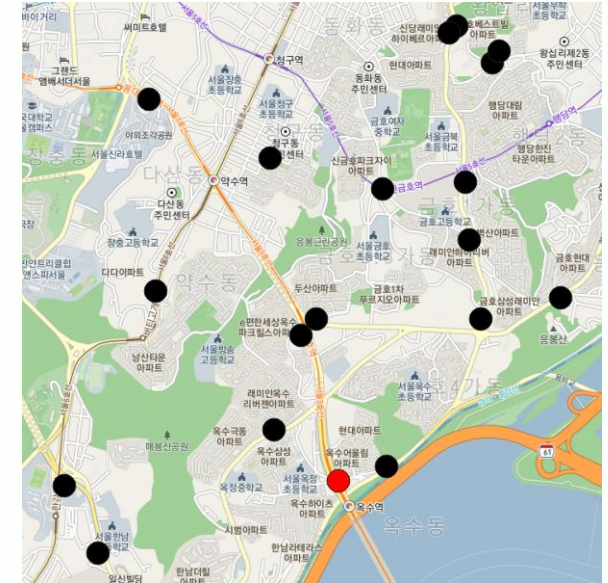


양재IC부근(서울외곽)

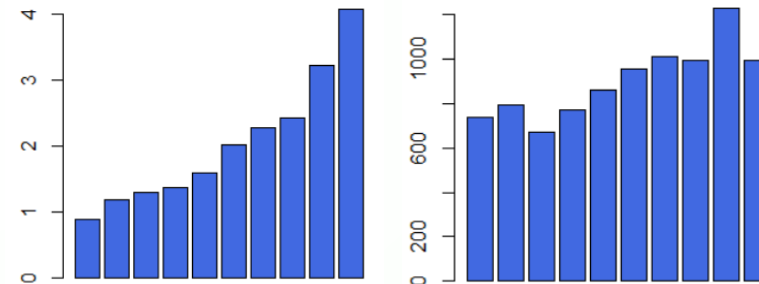
변수들의 분포(주거인구)



신정동, 개봉동(서울외곽)
평균_경사



이태원동, 옥수동(서울중심)
거리_하천



유동인구 & 주거인구

EDA

주거_30대	-0.04	-0.02	0.04	0.22	0.16	0.25	0.39	0.25
주거_20대	0.05	0.02	0.07	0.27	0.21	0.31	0.41	0.32
주거_60대	-0.05	-0.04	0.03	0.28	0.17	0.27	0.38	0.24
주거_50대	-0.06	-0.04	0.04	0.25	0.18	0.28	0.41	0.28
주거_40대	-0.09	-0.07	0.03	0.2	0.15	0.23	0.37	0.3
주거_10대	-0.1	-0.08	0.02	0.17	0.13	0.2	0.36	0.3

유동_20대
유동_30대
유동_40대
유동_70대
유동_50대
유동_60대
유동_10대
유동_10대

20/30/40대
: 주거공간과 생활공간이 나뉨

10대미만/10/50/60/70대
: 주거공간과 생활공간이 비슷함

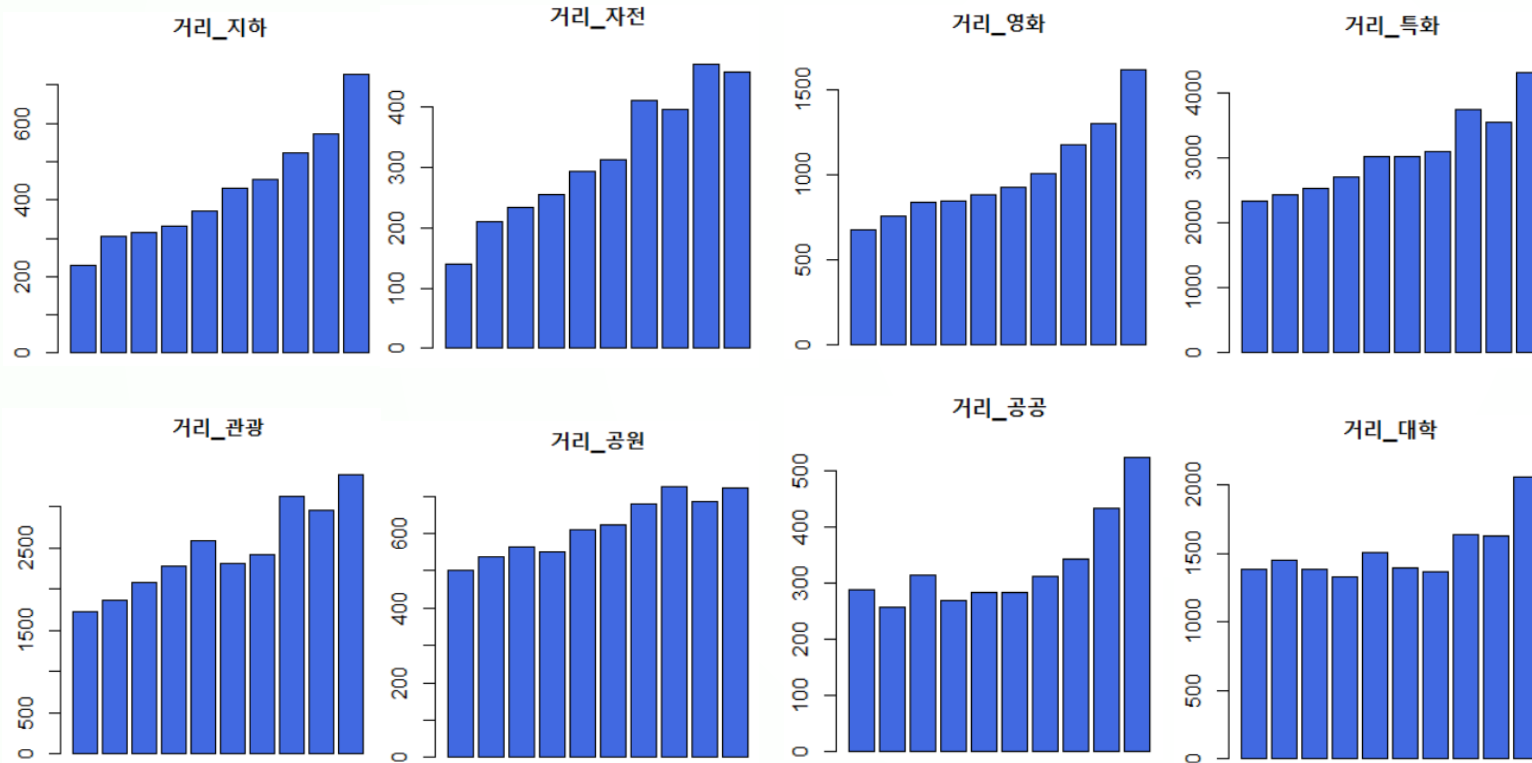
-> 즉, **20/30/40대**는 더 많은 **이동**을 하므로
이들과 관련된 변수가 이용량 예측에 중요할 것

[유동인구 & 주거인구]

20/30/40대 유동인구는 주거인구와 상관성이 적음
0/10/50/60/70대 유동인구는 주거인구와 상관성이 큼

변수들의 분포(거리+)

EDA

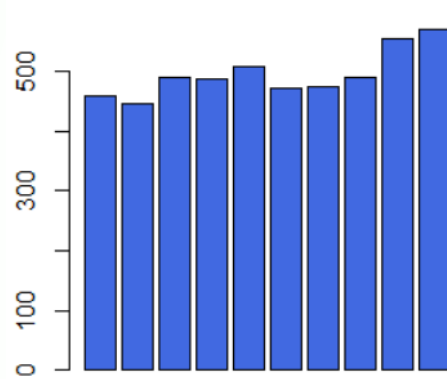


거리_대학							0.38
거리_관광						0.22	0.38
거리_영화					0.26	0.19	0.41
거리_공공				0.17	0.17	0.22	0.22
거리_지하			0.32	0.36	0.19	0.15	0.25
거리_공원		0.02	0.14	0.03	0.29	0.05	0.14
거리_자전	0.16	0.09	0.01	-0.01	-0.04	-0.13	0.03
	거리_공원	거리_지하	거리_공공	거리_영화	거리_관광	거리_대학	거리_특화

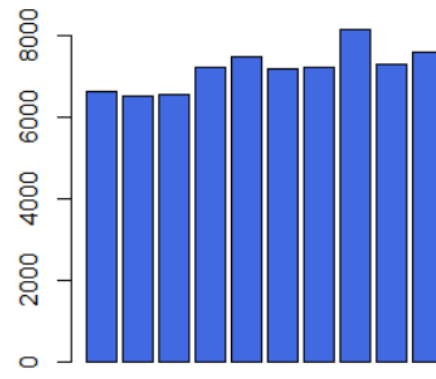
변수들의 분포(거리)

EDA

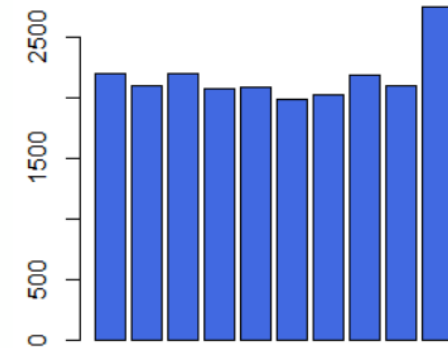
거리_주차



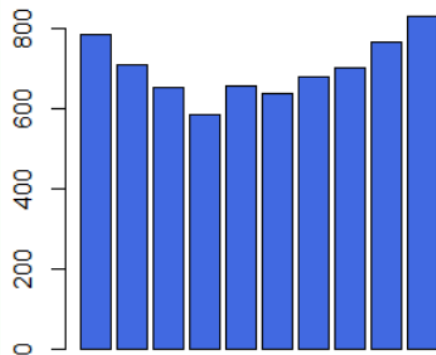
거리_교통



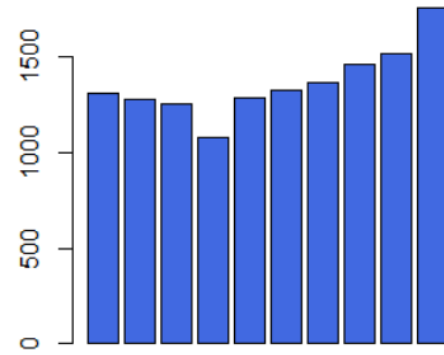
거리_체육



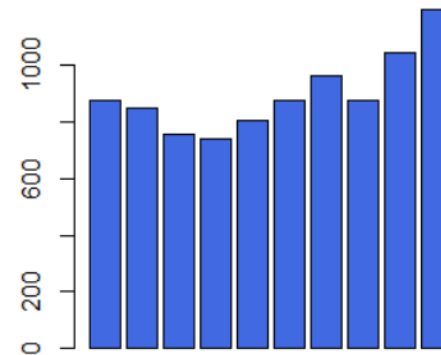
거리_문화



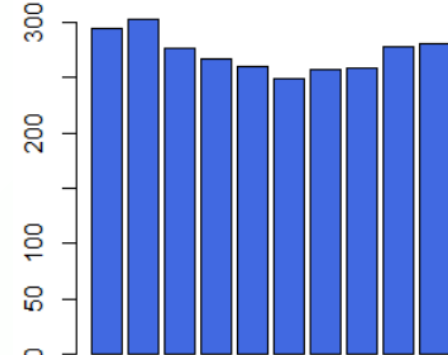
거리_의료



거리_상업

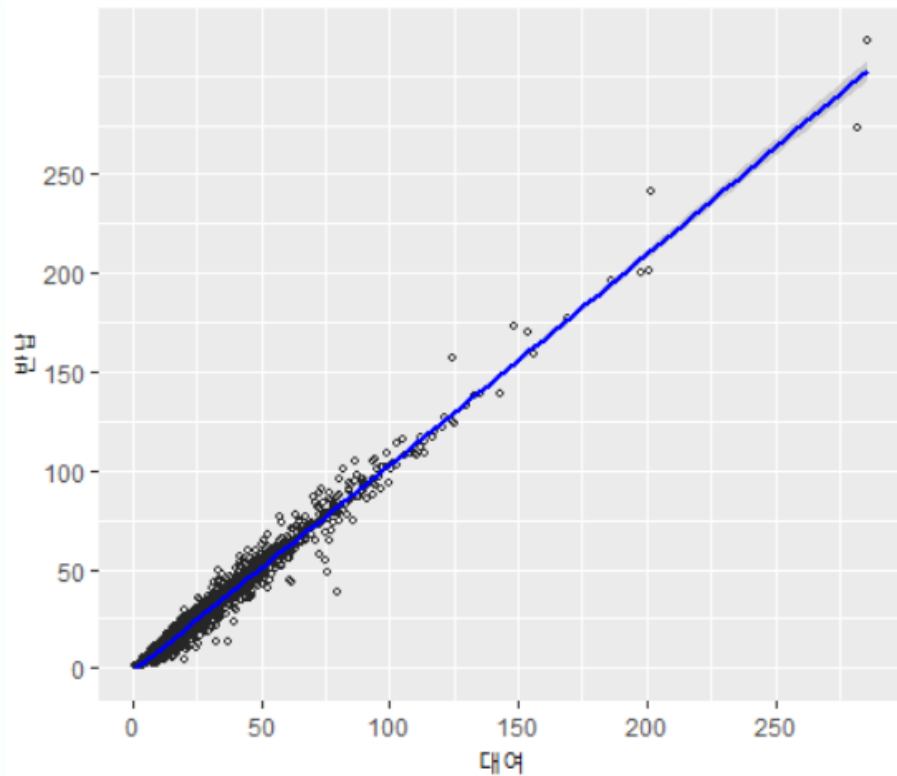


거리_초중



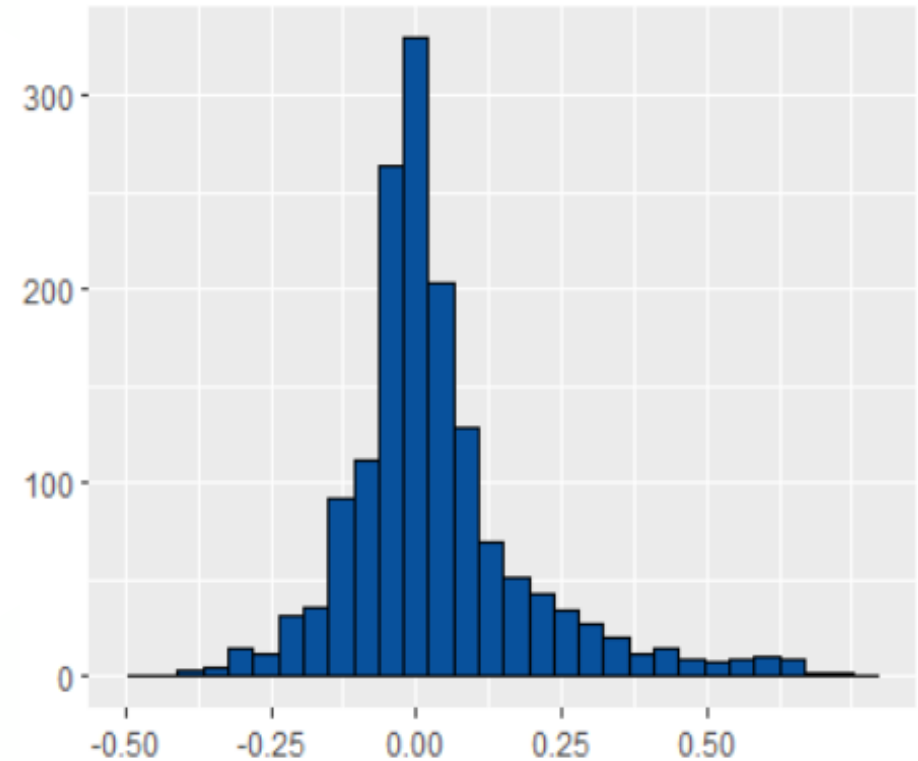
대여~반납

EDA



뚜렷한 양의 상관관계
 $\text{Corr}(\text{대여}, \text{반납}) = 0.9878$

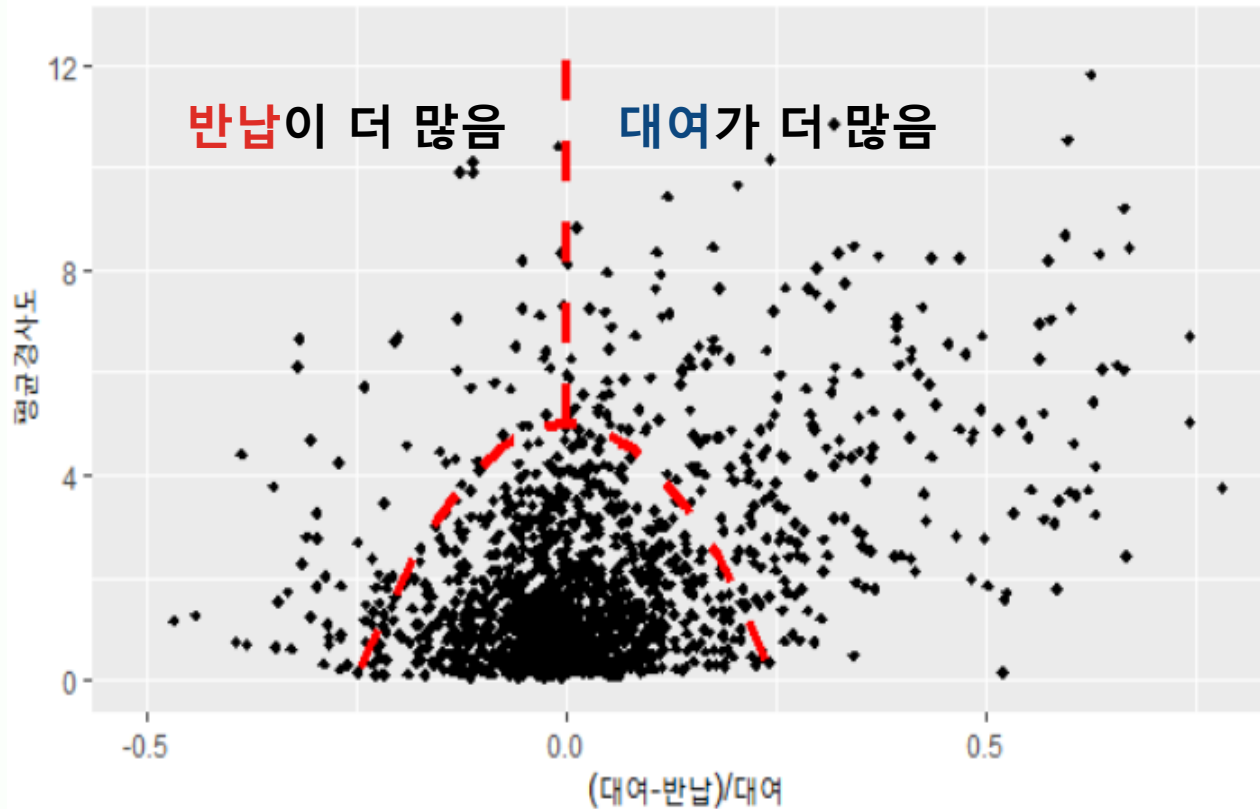
대여와 반납의
상대적인 비교



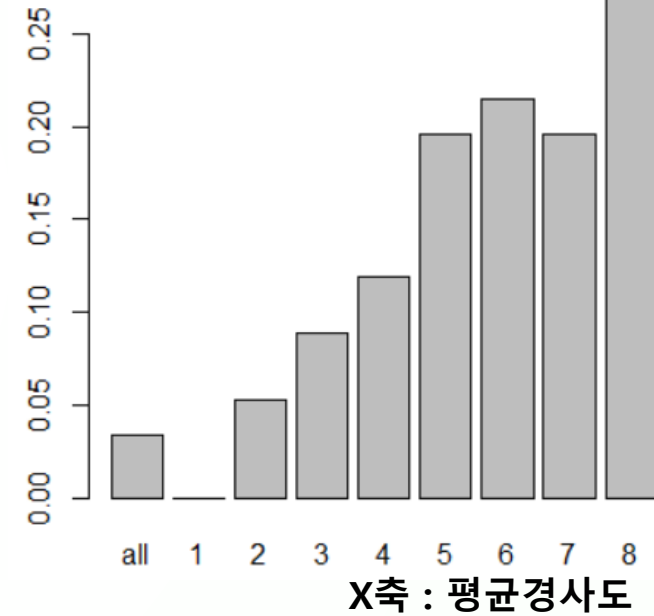
$$X = \frac{\text{대여} - \text{반납}}{\text{대여}} : \text{대여-반납 비율}$$

대여~반납 & 경사도

EDA



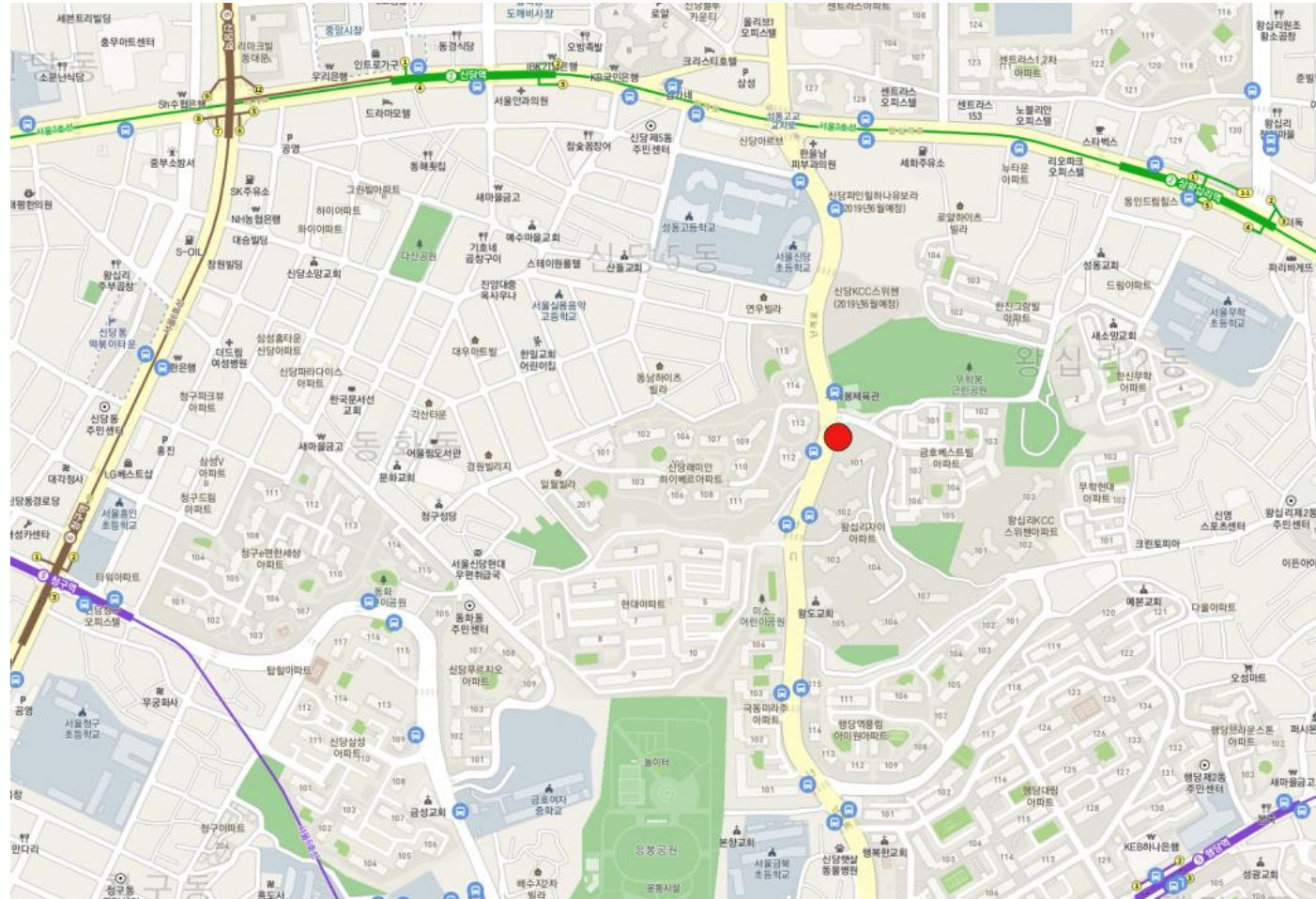
Y축 : 대여반납비율



-> 경사도가 높을수록 반납에 비해서 대여가 많음

대여~반납 & 경사도

EDA



왕십리 자이아파트

[반납 하위 TOP5]

[대여 평균 수준]

[상위 10%의 평균경사도]

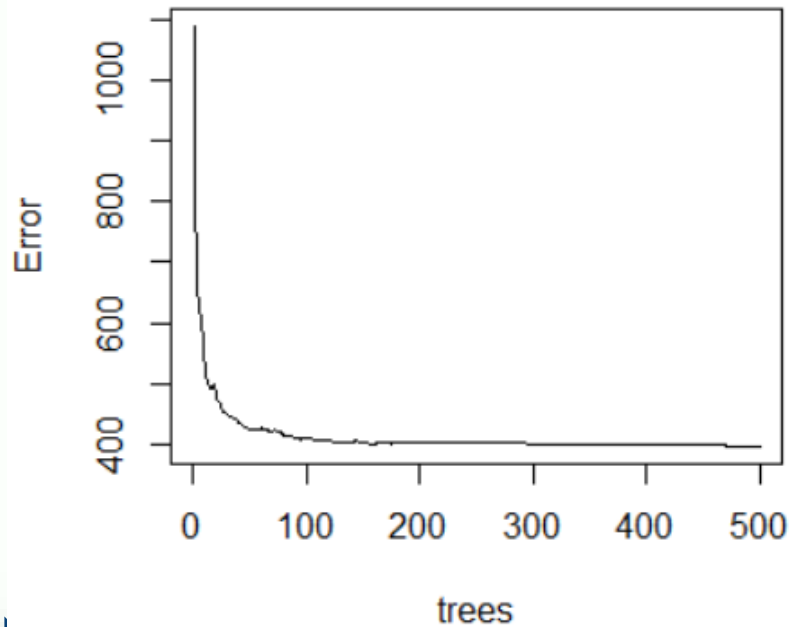
분석

하이퍼 파라미터

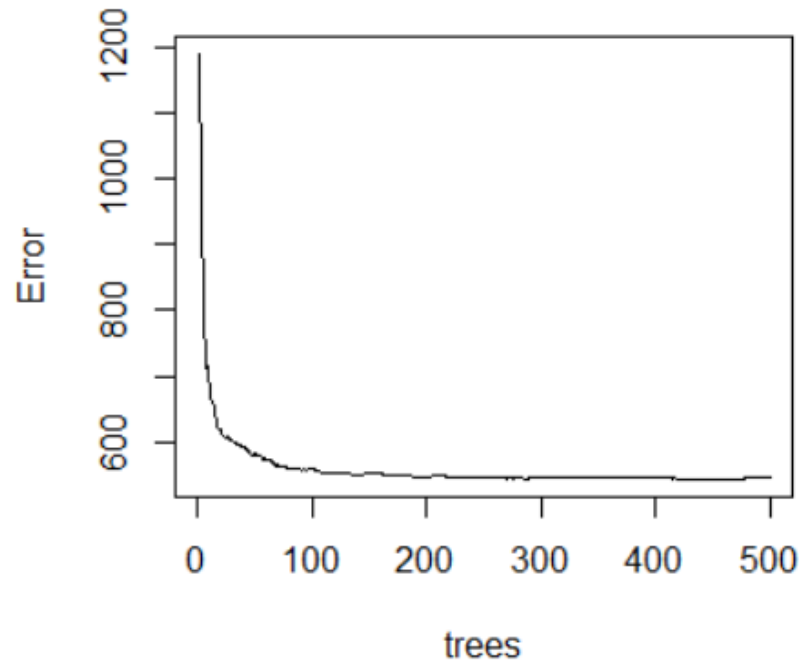
분석

ntree = 500 : Number of trees to grow.

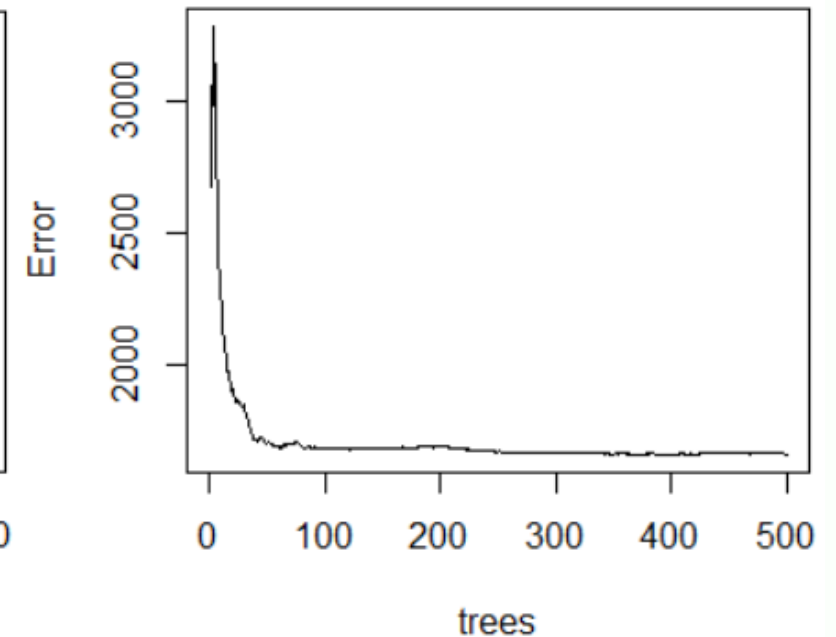
대여



반납



이용량

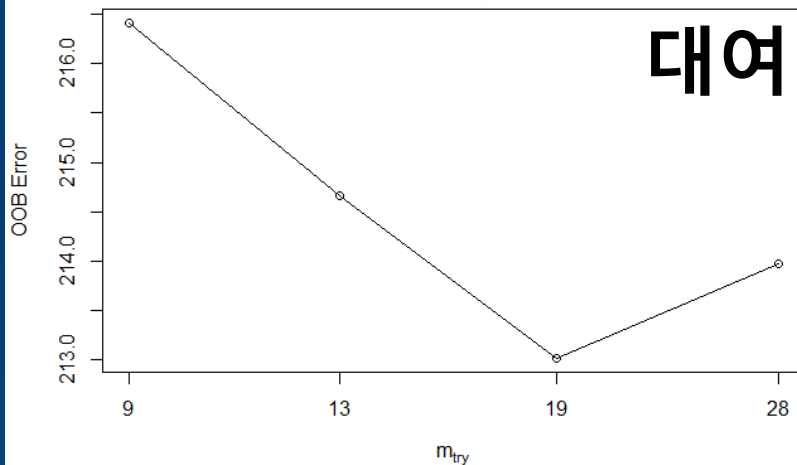


하이퍼 파라미터

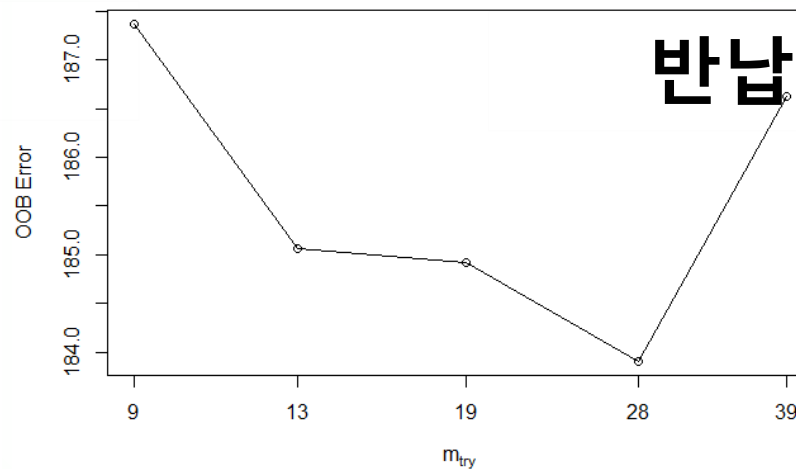
분석

Mtry : Number of variables randomly sampled at each split.

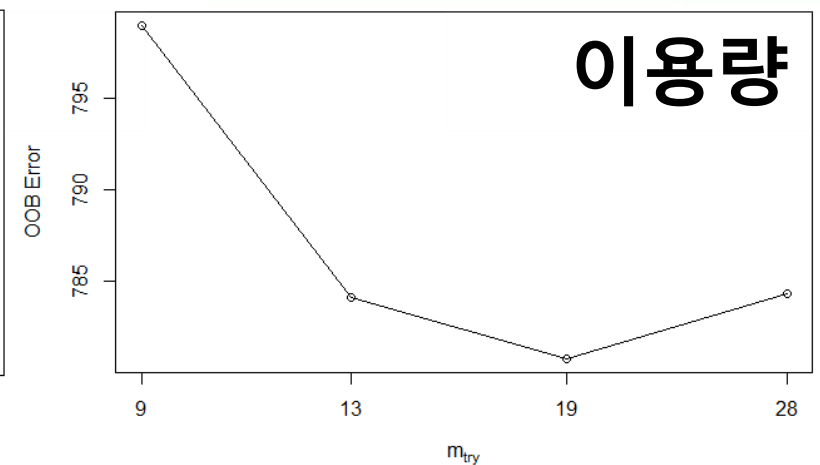
OOB ERROR : BootStrap에 포함되지 않은 데이터를 Decision Tree를 통해 나온 예측값과 실제 값의 차이



Mtry = 19
OOB Error = 213.0



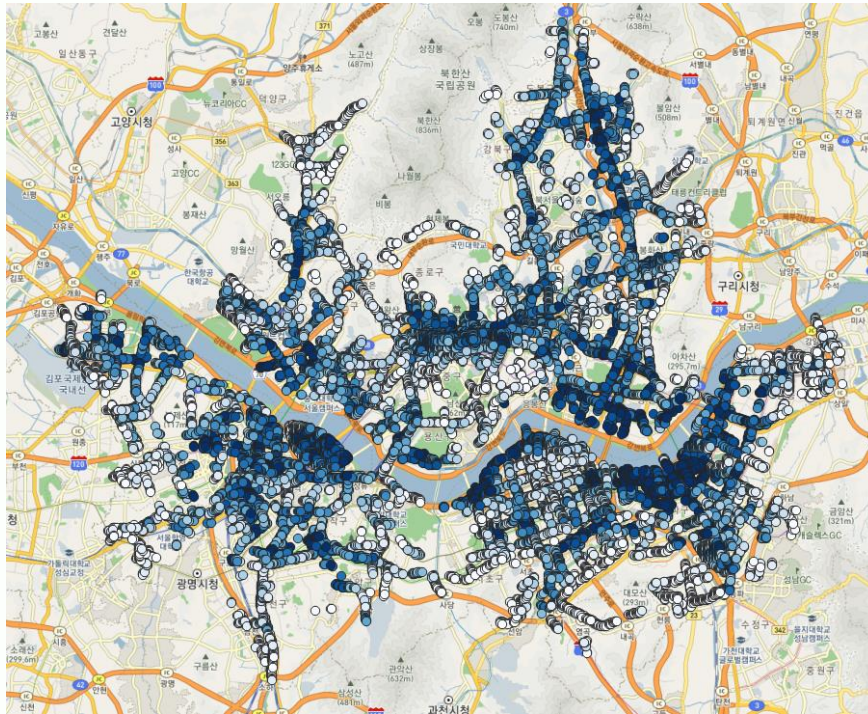
Mtry = 28
OOB Error = 184.0



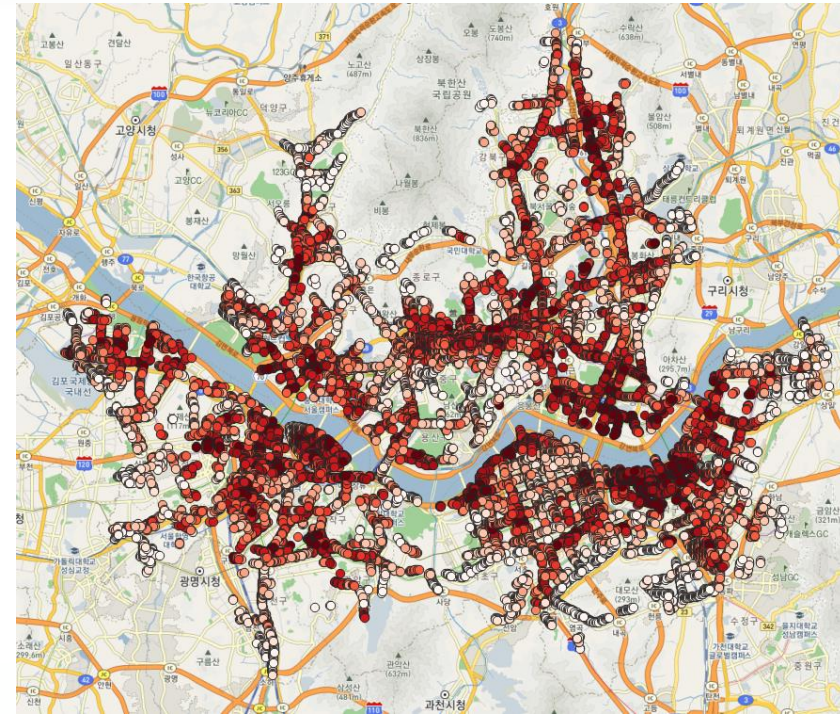
Mtry = 19
OOB Error = 780.8

이용량 vs 대여+반납

분석



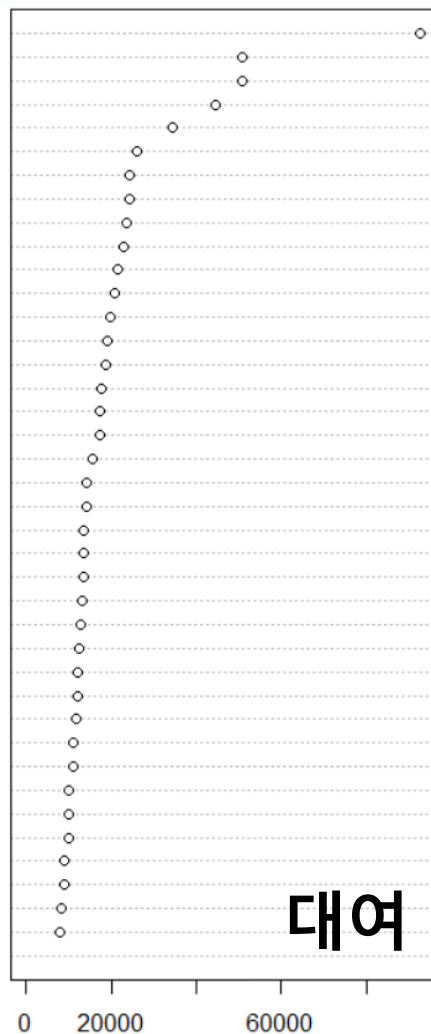
이용량
MSE = 1860.589



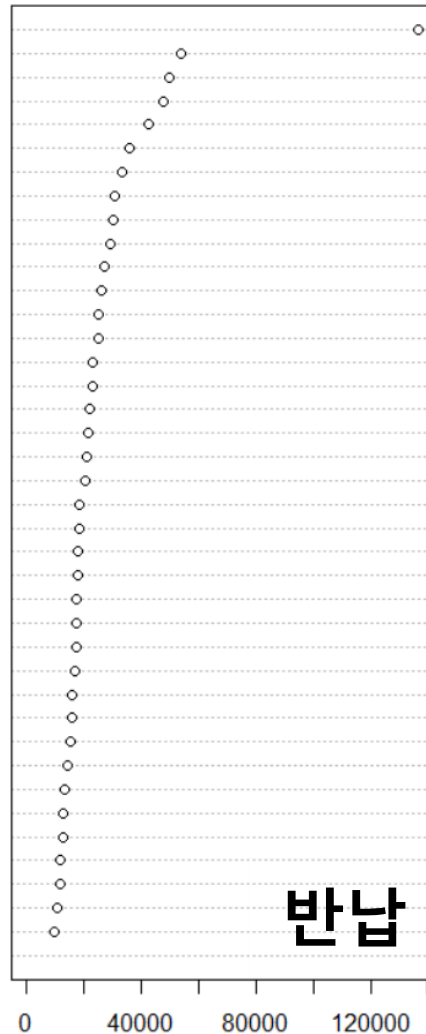
대여+반납
MSE = 947.052

변수 중요도

평균 경사
거리_지하
유동20대
지하철_승
버스_승객
거리_관광
거리_하천
거리_자전
유동30대
유동10대
주거_40대
거리_대학
유동10대미
거리_영화
거리_상업
거리_특화
거리_문화
유동40대
거리_주차
주거_20대
거리_공원
거리_의료
거리_교통
유동50대
유동70대미
거리_체육
near
면적_아파
유동60대
버스_경유
거리_공공
면적_주거
주거_10대
주거_30대
주거_60대
far
거리_초중
medium



평균 경사
유동20대
거리_지하
지하철_승
거리_관광
버스_승객
거리_하천
유동30대
주거_40대
거리_자전
거리_대학
유동10대
거리_문화
거리_상업
거리_영화
거리_특화
거리_공원
유동40대
주거_50대
주거_20대
주거_30대
거리_의료
거리_교통
유동50대
거리_체육
주거_10대
거리_공공
유동70대미
거리_주차
유동10대미
유동60대
near
거리_초중
면적_아파
면적_주거
주거_10대
주거_30대
주거_60대
far
면적_주거
medium
주거_60대



중요도 ↑

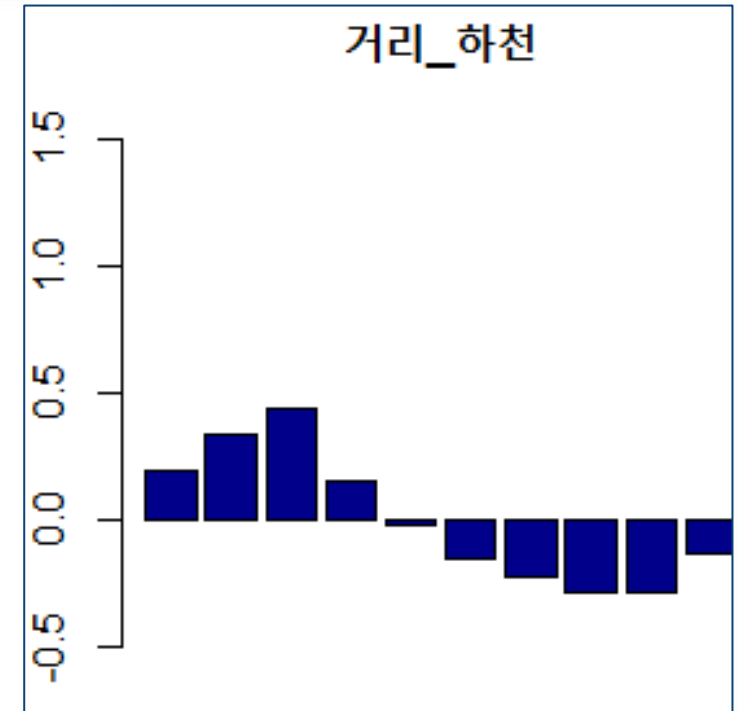
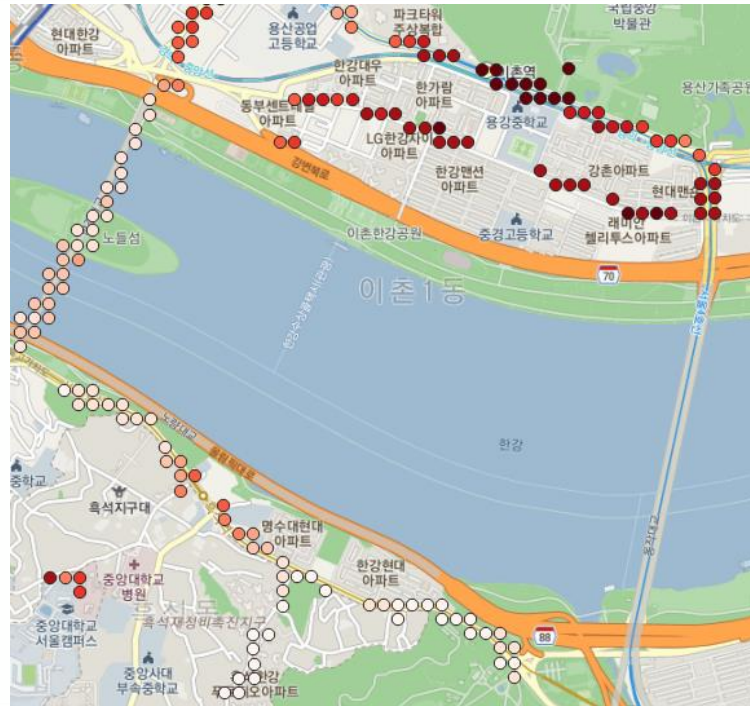
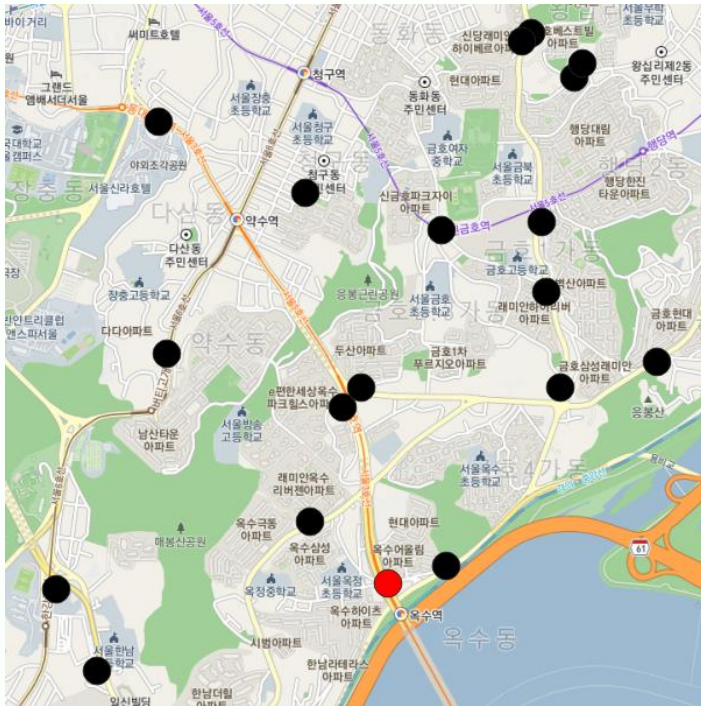
중요도 ↓

평균경사도
10/20/30대 유동인구
지하철 거리, 대중교통 승객
하천과의 거리
자전거도로와의 거리
대학과의 거리

Near/Far/Medium
초/중/고등학교와의 거리
주거지역 면적
아파트 면적
주거 60대

하천

결론

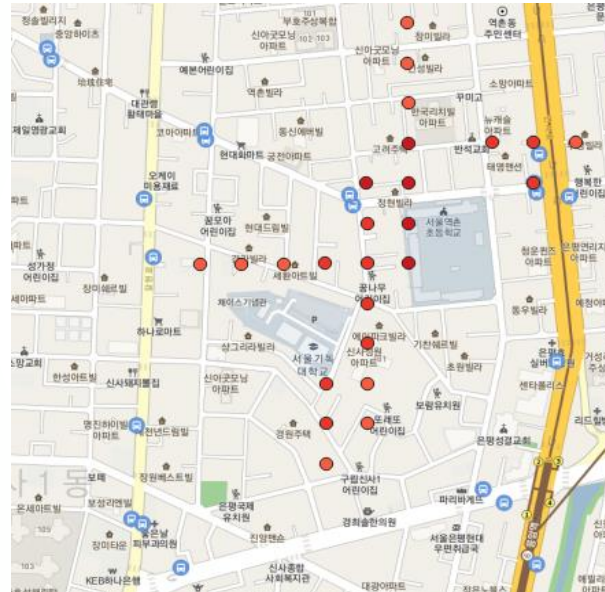


[하천과의 거리]

이용량에 영향을 많이 주는 변수이지만
하천 내 공원의 존재여부, 접근성이 중요해 보임

입지 선정

결론



비슷한 수준의 예측 이용량을 가진 장소들이 모여 있다
모두 최적의 위치로 보아서 대여소를 전부 설치할 수 없으므로 **대표장소**를 지정해 주어야한다

—

감사합니다

서울시립대 통계학과 김태현