



주차 수요 예측

김경덕, 안지현,
이현주, 박영원

데이터 전처리

1. 인근 지하철역 / 버스 정류장 결측치 처리

2. 임대료 / 임대보증금 결측치 처리 및 공급유형 그룹화

3. 등록 차량 수 순위에 따른 지역 그룹화

4. 기타 전처리

인근 지하철역 / 버스 정류장 결측치 처리



인근 지하철역 및 버스 정류장이 결측값인 경우
→ 0으로 간주

두 피처는 단지의 교통 접근성을 나타내는 같은 성질
→ '대중교통' 이라는 하나의 피처로 통일

임대료 / 임대보증금 결측치 처리

임대료 및 임대보증금이 결측

임대상가 (전용면적 별 세대수가 1)

공공 분양

국민 임대

행복 주택



대체로 임대료가 저렴하거나,
정부 보조금을 통해 일부 지원됨

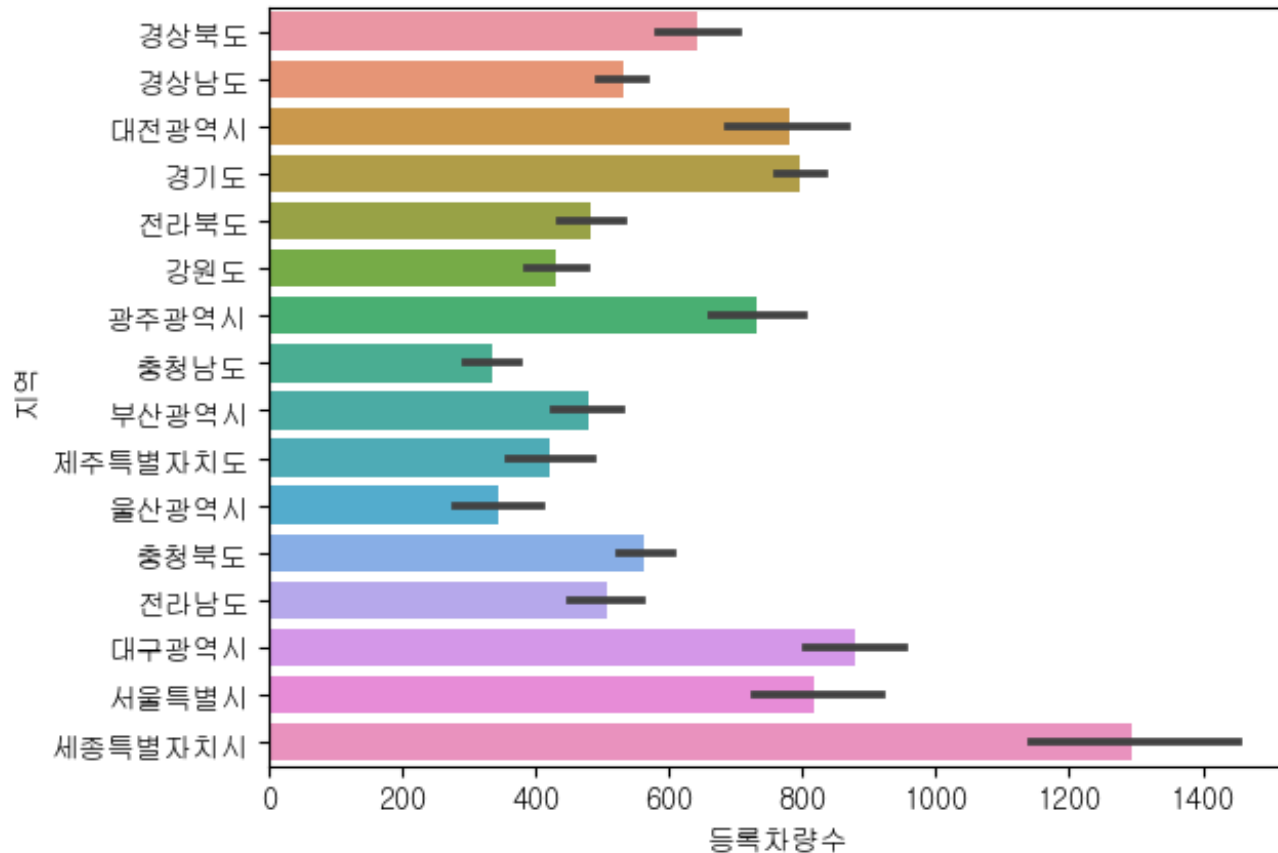


따라서 무료는 아니지만,
전처리 과정에서 0으로 대체

공급유형 그룹화

Type A	장기공공임대	50년			
Type B	단기공공임대	5년	10년	분납	분양
Type C	저소득층	행복주택	영구임대	장기전세	
Type D	국민임대	국민임대			
Type E	임대상가	임대상가			

등록 차량 수 순위에 따른 지역 그룹화



지역에 따른 등록 차량 수 중앙값

세종특별자치시

경기도

서울특별시

⋮

제주특별자치도

충청남도

울산광역시

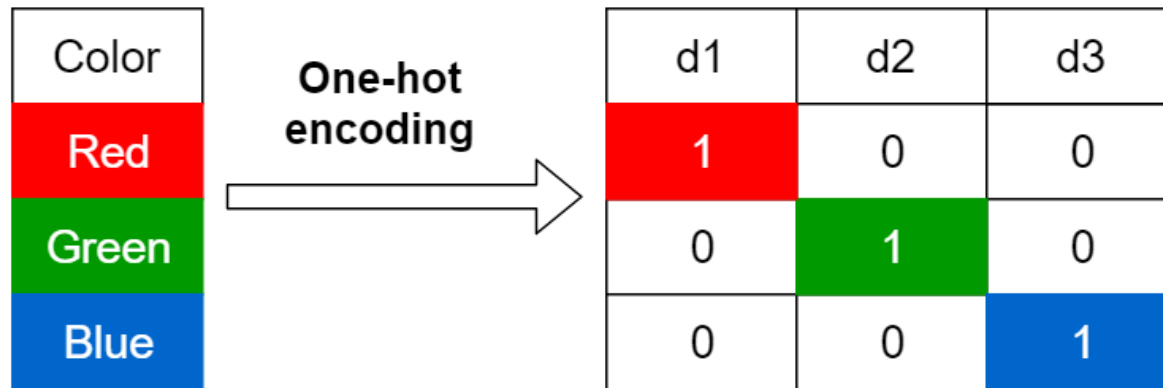
등록 차량 수 순위에 따른 지역 그룹화

1	서울특별시	경기도	세종특별자치시	
2		경상북도	광주광역시	대구광역시
3	전라남도	전라북도	충청북도	
4		강원도	울산광역시	경상남도
5	대전광역시	제주특별자치도	충청남도	부산광역시

원-핫 인코딩

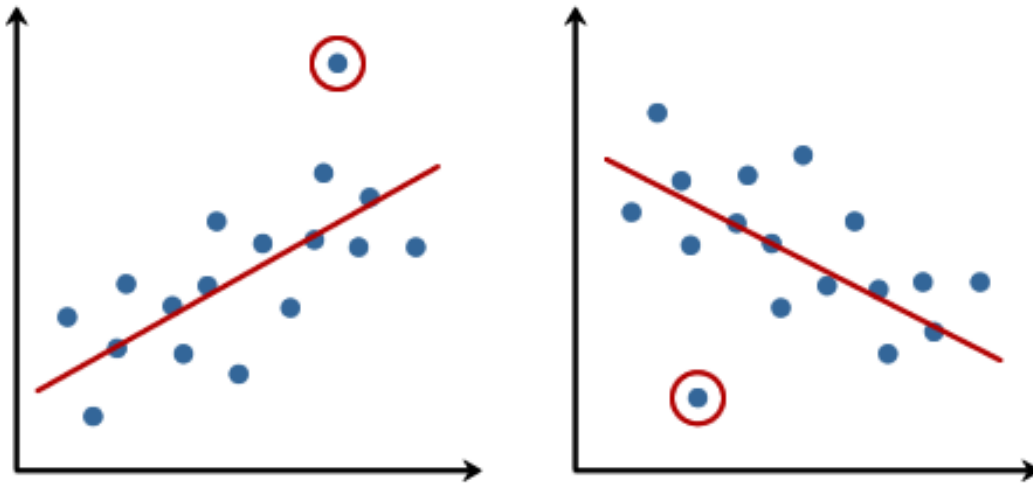
등록 차량 수 순위에 따라 그룹화한 지역 피처와 달리,
공급유형 피처는 유사 범주에 따라 그룹화됨

→ 레이블 인코딩 대신 원-핫 인코딩



타깃 변수 이상치 대체

등록 차량 수가 1500 이상인 경우, 이상치로 간주
→ 공급유형 / 지역 그룹 별 중앙값으로 대체



최종 선택된 피처

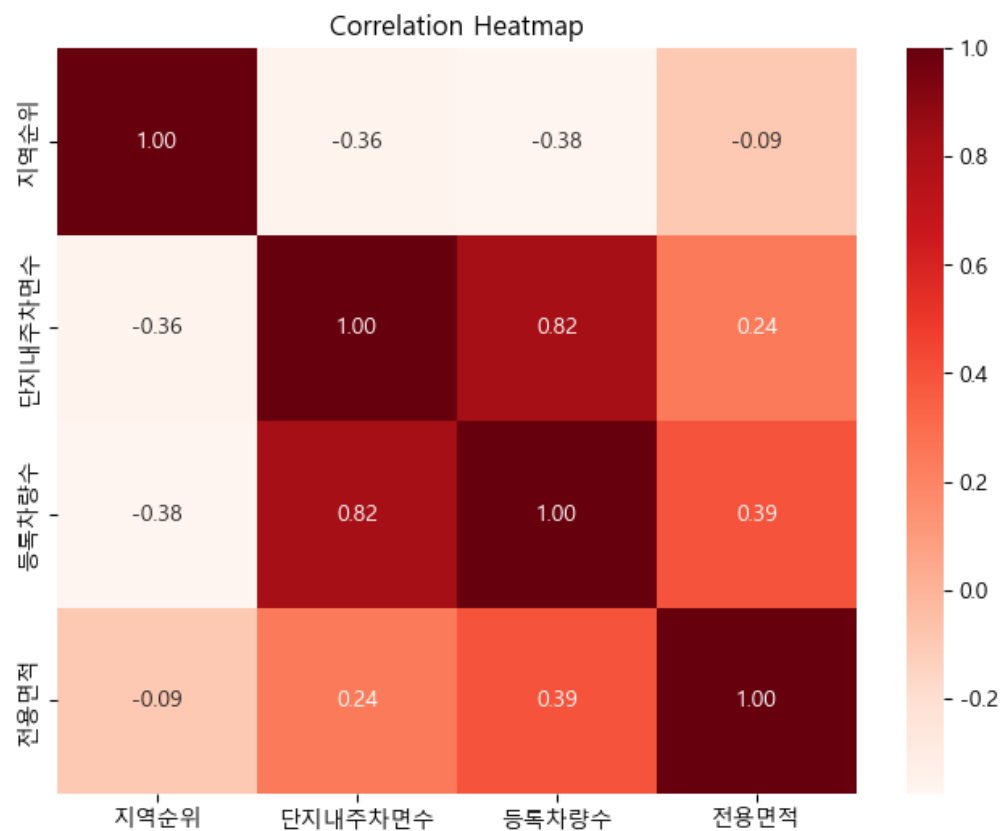
자체적으로 회귀 진단을 시행해 본 결과,

최종 선택

공급유형	지역순위
단지내주차면수	
전용면적	

총세대수	대중교통수	
공가수	임대료	삭제
	자격유형	
임대보증금		

피처 간 상관관계수



단지내주차면수 0.82

전용면적 0.39

지역순위 -0.38

회귀진단

다중공선성 진단

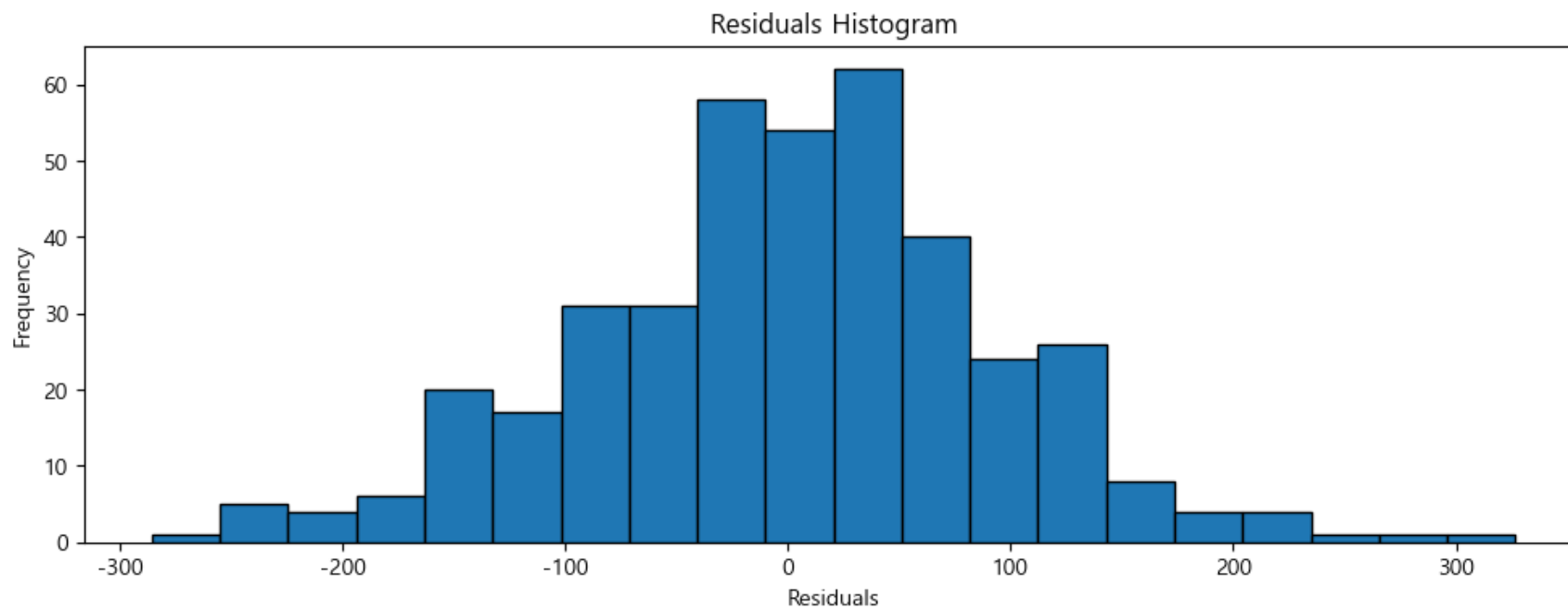
모든 피처의 VIF가 5 미만
→ 다중공선성 문제 없음

오차항의 자기 상관

더빈 왓슨 통계량 = 1.952
→ 2와 가까우므로 자기 상관 문제 없음

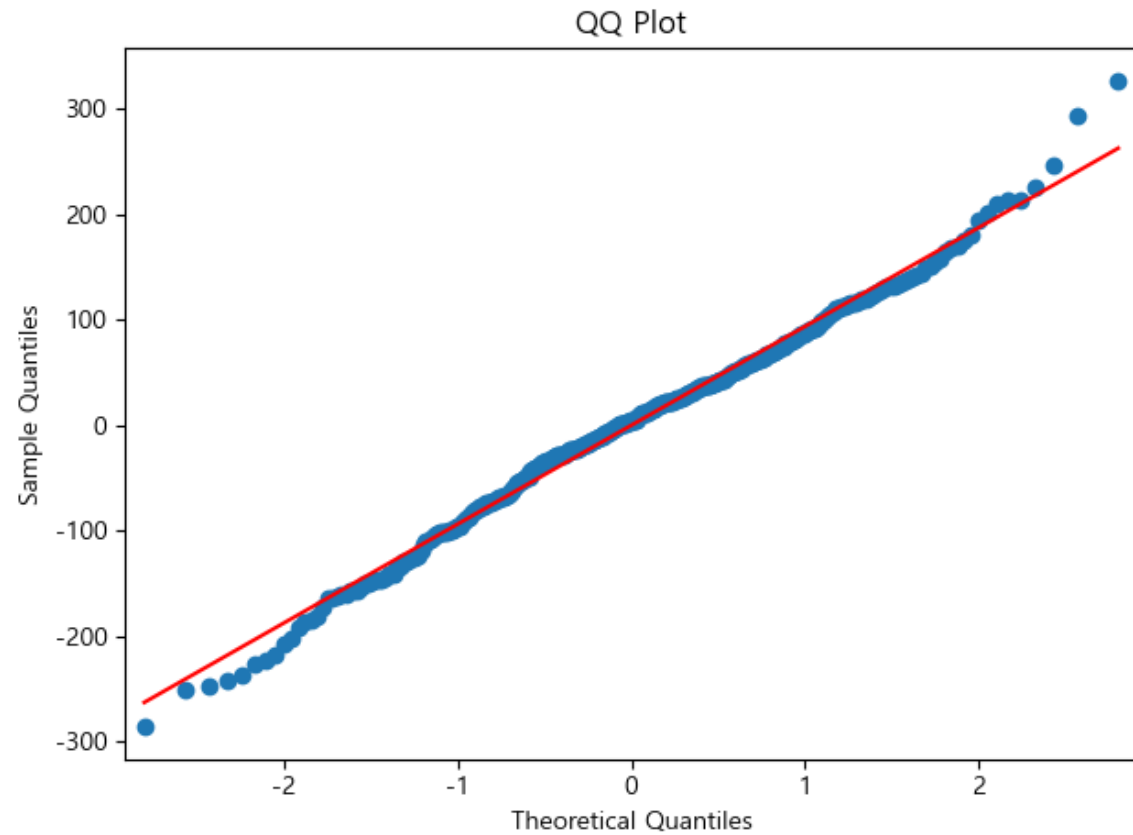
회귀진단

잔차 히스토그램



회귀진단

QQ Plot

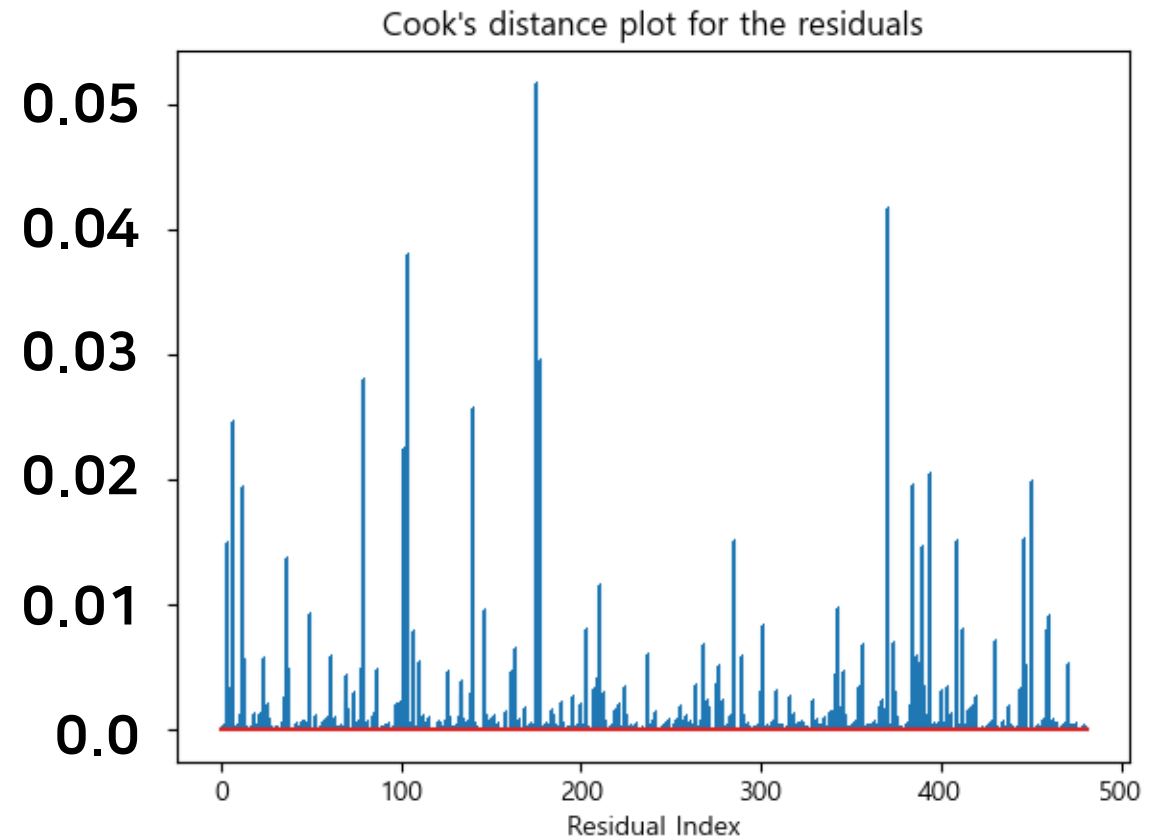


회귀진단

영향력이 큰 점 탐지)

Cook's Distance 그래프

→ 기준이 되는 threshold 값을
넘긴 샘플은 제외



최종 MAE

훈련 데이터 예측 MAE

117.66

(Cook's Distance) threshold 값을 넘긴 데이터 제외

72.38

감사합니다.