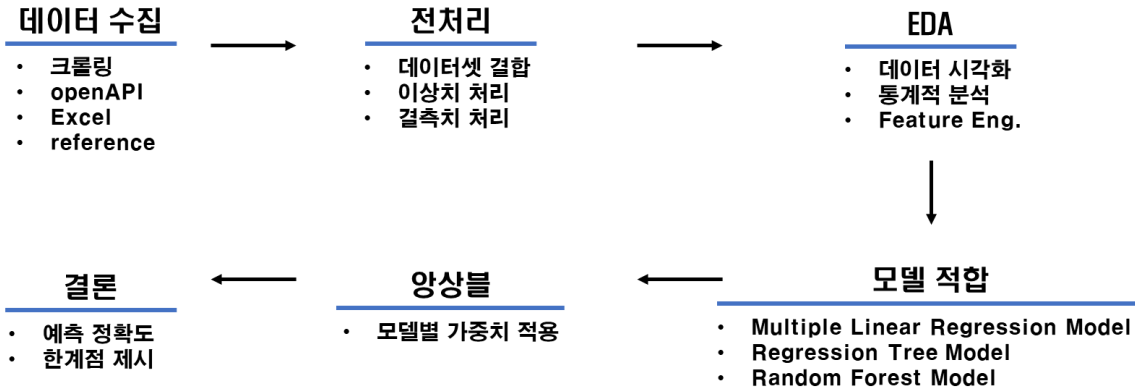
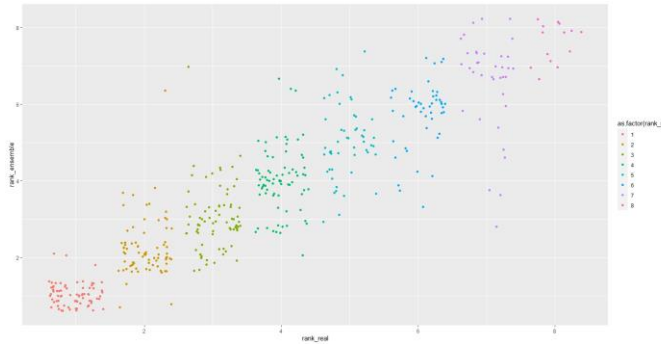


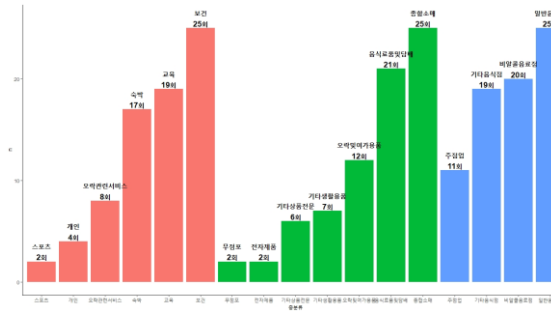
서울시 행정구역 및 업종별 매출 상위 Top3 업체 제시



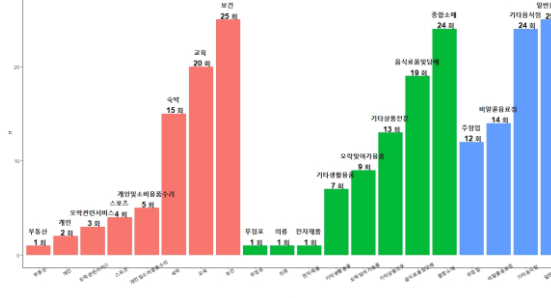
20.3분기 실제 업체 Rank vs 예측 업체 Rank



20.3분기 예측 Rank count bar plot



20.3분기 실제 Rank count bar plot



프로젝트 설명 :
코로나와 같은 예측불가능한 팬데믹에서 신규 창업자가 안정적인 수익 창출이 가능한 업체 제시를 위하여, 15.1~20.2분기 데이터 분석 후 업체 예측 모델을 제작하여 행정구역 및 업종별 Top3 업체를 제시해주는 프로젝트

언어 : R

머신기술 스택:
Multi Linear Regression
Regression Tree
Random Forest

통계기술 스택:
T검정
아노바 검정
상관분석

- bias 최소화를 위해 점포수가 5개 이하인 데이터 삭제
- 15.1~20.2에서 한 분기라도 매출액이 임대료보다 적은 경우 삭제
- 모델 적합 전 기본 변수들에 대한 EDA를 통해 목표 변수와 연관성이 있는 Feature Engineering
➔ 토요일 10,40,50대 매출비율
- Model별 Coef(가중치)를 통한 앙상블 결과 행정구역 및 업종별 Top3 Rank 정확도 2.3% 향상 (89.3% ➔ 91.6%)