**데이터 준비**

1. **날짜 데이터 변환**

날짜 데이터를 datetime 형식으로 변환 (연/월/일 분리 가능하도록 변환)

1. **컬럼명 정리**
2. **데이터 전처리**

결측치 처리 (제거 또는 대체 방법 적용)

이상치 분석 및 제거 (z-score 등 활용)

데이터 정규화 (스케일 조정)

1. **연속형 데이터 상관관계 분석**

heatmap을 사용하여 상관관계를 시각화

높은 상관계수를 가진 독립변수 제거 (다중공선성 문제 고려)

1. **선형 관계 분석**

연속형 변수들과 중요 변수 간의 관계를 regplot(산점도) 또는 barplot(막대 차트)로 시각화

1. **명목형 변수 분석**

범주형 변수와 중요 변수 간의 관계를 boxplot 등을 이용하여 분석

**피처 엔지니어링**

1. **파생 변수 생성**

변수 간의 관계를 고려하여 새로운 파생변수 생성

생성된 파생변수 간의 상관계수 분석 (heatmap 활용)

1. **범주형 변수 인코딩**

get\_dummies를 활용하여 범주형 변수를 수치형 데이터로 변환

**모델링**

1. **데이터 분할**

훈련 데이터, 검증 데이터, 테스트 데이터로 나누기 (train\_test\_split 활용)

1. **기본 모델 학습**

회귀 모델을 사용하여 학습 진행

훈련 데이터와 검증 데이터의 점수(score) 계산

1. **앙상블 학습 적용**

랜덤 포레스트, LGBM 회귀, XGBoost 회귀 모델 학습

각 모델의 훈련 데이터 및 검증 데이터 점수 계산

RMSE 값 분석하여 최적 모델 선정

회귀 평가 지표(r2 score, MSE, MAE 등) 고려

**하이퍼파라미터 탐색**

1. **최적 모델 선택 후 하이퍼파라미터 튜닝**

GridSearchCV 또는 RandomizedSearchCV 활용

1. **최적 하이퍼파라미터로 모델 재학습**

**최종 모델 평가**

테스트 데이터를 이용하여 최종 성능 평가 (r2 score 등 활용)

**계획에 추가해야할 점**

뉴스 감성 분석 또는 정책 발표를 NLP 기법으로 분석하고 모델에 통합하는 방법. 대화형 대시보드 방법

하이퍼 파라미터 사용 방법.