|  |  |
| --- | --- |
| **| 문제의도** | * 데이터 필터링을 통해 필요한 데이터를 선택하고, 이를 바탕으로 선형 회귀 모델을 학습하여 예측 결과를 평가하는 과정을 학습하는 것이 목표입니다. * train\_test\_split을 이용해 학습 데이터와 테스트 데이터를 분리하고, R² 점수를 통해 모델 성능을 확인하는 과정을 학습합니다. |
| **| 사전지식** | * pandas를 사용한 데이터 로딩 및 필터링 * scikit-learn의 선형 회귀 모델 사용법 * 학습 데이터와 테스트 데이터의 개념 및 분리 방법 |
| **| 문제핵심** | * Boolean indexing을 사용하여 데이터를 필터링하는 과정 * 학습 데이터와 테스트 데이터를 나누는 방법 * R² 점수를 이용한 모델 성능 평가 |

|  |  |
| --- | --- |
| **| 문제해설** | |
| **| 해설** | 1. 데이터 불러오기: pandas의 read\_excel 함수를 사용하여 데이터를 불러옵니다. 이 함수를 사용하면 엑셀 파일을 데이터프레임으로 읽어올 수 있습니다. 파일 경로는 주어진 경로에 맞게 수정하여 사용하세요.  2. 데이터 필터링: Boolean indexing을 사용하여 원하는 조건에 맞는 데이터만 선택합니다. 여기서는 '성산로(금화터널)' 지점의 '유입' 방향 데이터를 필터링합니다. Boolean indexing은 조건이 True인 데이터만 선택하는 방법입니다.  3. 피처와 타깃 값 설정: '요일'을 독립 변수로, '0시' 시간대의 교통량을 종속 변수로 설정합니다. 피처는 모델이 학습할 때 사용하는 입력값이고, 타깃 값은 모델이 예측하는 대상입니다.  4. 학습 데이터와 테스트 데이터 분할: train\_test\_split 함수를 사용하여 데이터를 학습용과 테스트용으로 나눕니다. 이때 test\_size=0.2로 지정하여 전체 데이터의 20%를 테스트용 데이터로 사용합니다.  5. 선형 회귀 모델 정의 및 학습: LinearRegression 클래스를 사용하여 선형 회귀 모델을 정의하고, 학습 데이터를 사용해 모델을 학습시킵니다. fit 메서드는 모델을 학습시키는 함수입니다.  6. 모델 평가: 모델의 성능을 R² 점수로 평가합니다. R² 점수는 모델이 데이터를 얼마나 잘 설명하는지 나타내며, 1에 가까울수록 성능이 좋음을 의미합니다.  실행 결과로, R² 값은 0.2835616050573254로 출력되었습니다. 이 값은 모델이 0시 교통량을 설명하는 데 있어 요일만을 사용한 예측의 성능이 낮음을 의미합니다. 이는 요일만으로는 0시 교통량을 충분히 예측하기 어렵다는 것을 나타냅니다. 요일 외에 다른 변수들(예: 날씨, 시간대, 공휴일 등)을 추가하여 예측 모델의 성능을 개선할 수 있을 것입니다. |