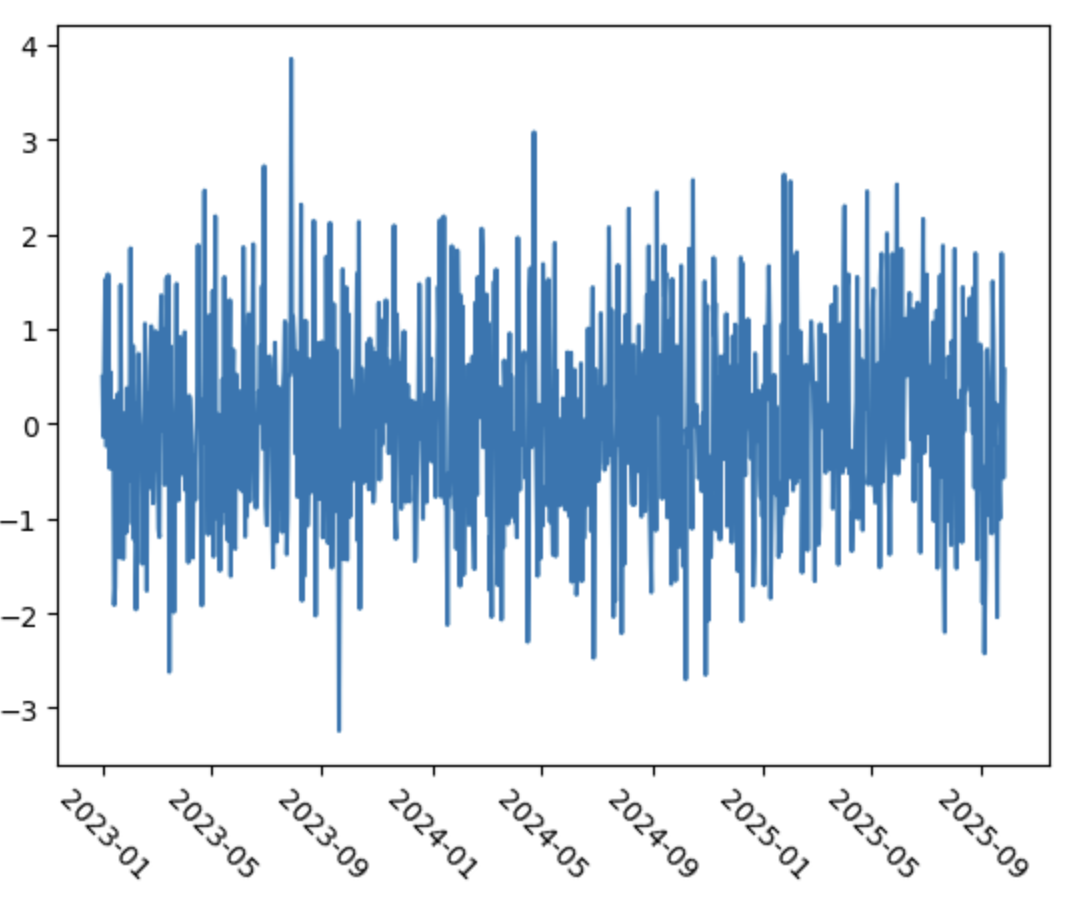
MS\_HW3\_1

이상윤

1. 시계열 데이터 분석(추세, 계절성, 잔차)
2. 추세

시계열 데이터를 생성하기 위해 기간에 따라 무작위로 값을 부여하였다. 추세에서는 증감을 통해 장기적인 변동패턴을 파악하고, 계절성 그래프에서는 주기적인 패턴을 분석하여 반복되는 형태를 분석, 마지막 잔차는 앞서 이야기한 요소를 제외한 변동요소를 분석한 것이다.

텍스트, 스크린샷, 폰트, 그래프이(가) 표시된 사진

자동 생성된 설명

1. 계절성

계절성 성분의 그래프를 통해 데이터의 주기적인 패턴과 변동을 시각적으로 파악할 수 있습니다. 특정 기간에 데이 터 값이 어떻게 변화하는지 확인할 수 있으며, 계절성 패턴을 이해하여 데이터의 특성을 파악하는 데 도움이 됩니다.

1. 자기상관성

이 그래프는 시계열 데이터의 자기 상관성을 보여줍니다. X축은 Lag(시차)를 나타내며, Y축은 자기 상관성의 값입니다. 시차(Lag)는 현재 데이터와 이전 데이터 사이의 간격을 의미합니다.

텍스트, 스크린샷, 그래프, 도표이(가) 표시된 사진

자동 생성된 설명텍스트, 스크린샷, 라인, 그래프이(가) 표시된 사진

자동 생성된 설명

1. 시계열 데이터 구성 요소 3가지(상향, 하향, 평평한 추세)

텍스트, 라인, 스크린샷, 도표이(가) 표시된 사진

자동 생성된 설명텍스트, 스크린샷, 라인, 그래프이(가) 표시된 사진

자동 생성된 설명텍스트, 스크린샷, 라인, 폰트이(가) 표시된 사진

자동 생성된 설명

1. 시계열 데이터 탐색
2. (Scatter, Bar plot, Box plot, Histogram, Heatmap)

텍스트, 스크린샷이(가) 표시된 사진

자동 생성된 설명텍스트, 스크린샷, 라인, 평행이(가) 표시된 사진

자동 생성된 설명텍스트, 스크린샷, 도표, 직사각형이(가) 표시된 사진

자동 생성된 설명도표, 텍스트, 그래프, 스크린샷이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 직사각형, 패턴이(가) 표시된 사진

자동 생성된 설명

1. 시계열 분해(추세, 계절성, 불규칙성)

텍스트, 폰트, 친필, 라인이(가) 표시된 사진

자동 생성된 설명

MS\_HW3\_2

이상윤

1. 시계열 데이터 결측치 위치 확인
2. 위치 확인

Print(data.isnull())을 이용하여 True, False형태로 결측치의 위치를 확인할 수 있다.

텍스트, 폰트, 스크린샷, 화이트이(가) 표시된 사진

자동 생성된 설명

1. 히트맵 시각화

사각형, 직사각형, 스크린샷, 다채로움이(가) 표시된 사진

자동 생성된 설명

1. 결측치 개수 확인

텍스트, 폰트, 라인, 스크린샷이(가) 표시된 사진

자동 생성된 설명

1. 시계열 데이터 전처리
2. 결측치가 있는 행 삭제

결측치는 데이터를 분석하는데 무의미한 상관관계를 만들어낸다. 따라서 이런 결측치를 어떻게 다루느냐가 시계열 데이터를 분석하는데 중요한 요소 중 하나이다. 선형, 스플라인, 최근접 이웃 보간법을 사용하기 위해 결측치가 있는 행을 먼저 삭제해 준다.

1. 선형 보간법

텍스트, 스크린샷, 라인, 그래프이(가) 표시된 사진

자동 생성된 설명

'Original'은 원본 데이터를 나타내며, 'Linear Interpolation'은 선형 보간된 데이터를 나타냅니다.

x축은 날짜(Date)를 나타내고, y축은 값(Value)을 나타냅니다. 원본 데이터의 경우, 결측치가 있어서 일부 값이 비어있는 것 을 볼 수 있습니다. 선형 보간된 데이터는 결측치를 선형적으로 보간하여 채운 결과입니다.

선형 보간된 데이터는 원본 데이터의 경향을 따라가면서 결측치를 채웠기 때문에 부드러운 곡선으로 이어져 있습니다.

1. 스플라인 보간법

텍스트, 스크린샷, 도표, 폰트이(가) 표시된 사진

자동 생성된 설명

스플라인 보간법은 유효한 데이터에 대해서만 적용이 가능합니다. 그래서 결측치를 제거하거나 대처하는 작업이 필요합니다.

'Original': 원본 데이터를 나타내는 선 그래프입니다. 원본 데이터는 100번째부터 200번째까지의 구간에 결측치를 가지고 있습니다. 따라서 해당 구간은 비어 있는 상태로 표시되며, 다른 데이터들은 원본 값을 따라 나타납니다.

Spline Interpolation': 스플라인 보간법을 적용하여 결측치를 보간한 결과를 나타내는 X 표시입니다. 스플라인 보간법은 주어진 데이터를 이용하여 부드러운 곡선으로 보간하는 방법입니다. 따라서 원본 데이터의 추세를 잘 따라가며 결측치를 채워 넣은 것을 확인할 수 있습니다.

1. 최근접 이웃 보간

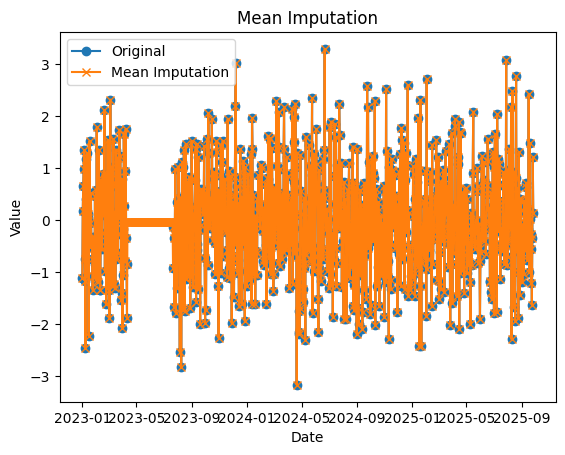
텍스트, 스크린샷, 도표, 폰트이(가) 표시된 사진

자동 생성된 설명'

Original : 이 선은 원본 데이터를 나타냅니다. 원본 데이터에 는 결측치가 포함되어 있어서 몇몇 지점에서 비어있는 값이 있습니다. 이 선은 결측치를 가지고 있으며, 데이터의 원래 패턴과 변동성을 보여줍니다.

Nearest Neighbor Interpolation': 이 선은 최근접 이웃 보간을 적용한 결과를 나타냅니다. 최근접 이웃 보간은 결측치를 주 변의 가장 가까운 이웃 데이터로 대체하는 방법입니다. 따라 서 이 선은 결측치가 채워진 데이터를 보여줍니다. 최근접 이웃 데이터와 동일한 값을 가지며, 데이터의 패턴을 최대한 유지하려고 합니다.

1. 평균값 대처



평균값 대체는 간단하고 효과적인 방법이지만, 데이터의 특성과 분석 목적에 맞게 다른 결측치 처리 방법을 고려해 야 합니다. 상황에 따라서는 다른 통계적 기법이나 머신러닝 모델을 활용하여 결측치를 예측하거나, 도메인 지식을 활용하여 대체하는 것이 더 적절할 수 있습니다.

1. ARIMA 모델 활용하여 결측치 예측과 대처 실습

텍스트, 스크린샷, 도표, 폰트이(가) 표시된 사진

자동 생성된 설명

ARIMA(자동회귀누적이동평균)는 시계열 데이터의 패턴을 파악하여 미래 값을 예측하는 모델입니다.