1. 데이터셋 개요

2. 전처리 및 특징 분석

3. 모델 서칭 및 실험 방법

4. 성능 및 실행시간

5. 결과 분석

6. 결론

1. 데이터셋 개요

Tourism : 월별, 분기별, 연간 관광객

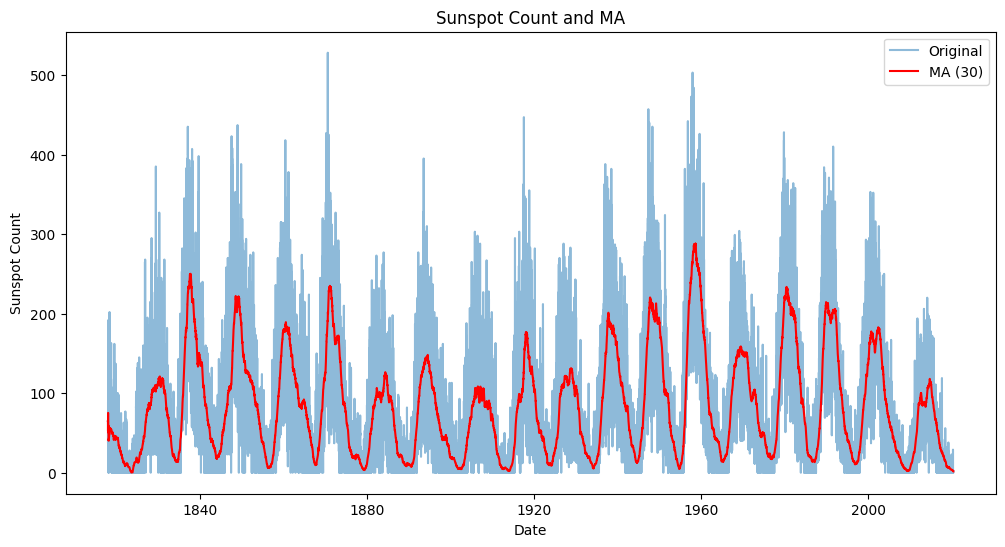
Rideshare : 2018년 11월 26일~2018년 12월 18일 뉴욕의 다양한 지역에서 Uber 및 Lyft 승차 공유 서비스와 관련된 속성의 다양한 시간별 시계열

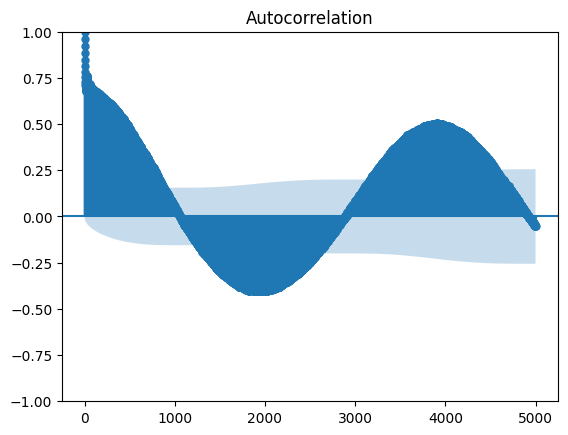
Sunspot : 1818년~2020년 일별 흑점데이터 세트

US birth : 1969년~1988년 일별 미국에서 태어난 아기 수

1. 전처리 및 특징 분석

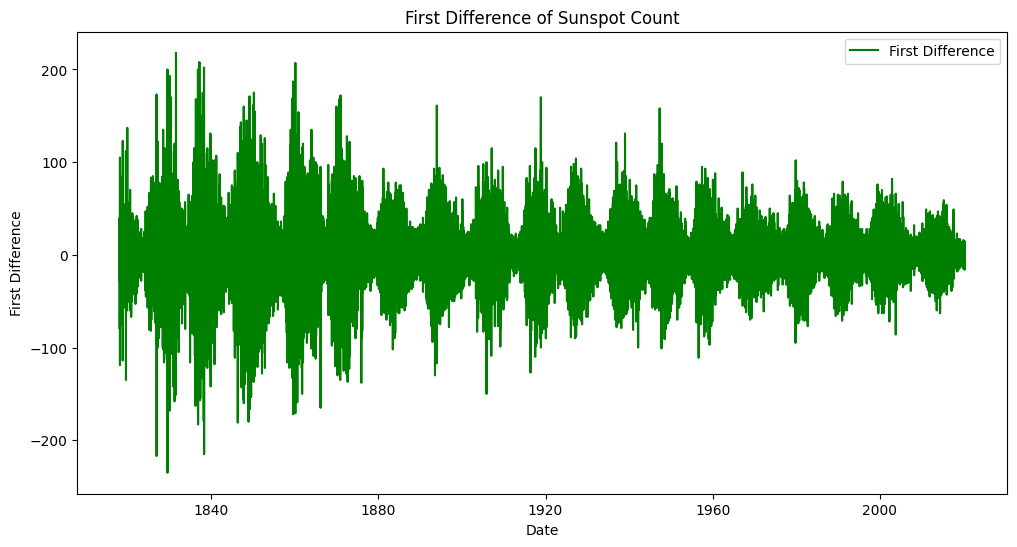
sunspot : 1818~2020 간 일별 흑점 개수 데이터 세트. 일반적으로 태양의 흑점의 개수는 10~11년 주기가 있음. 결측값은 해당 가장 마지막에 발생한 시점의 데이터를 계속 이용.(forward fill)





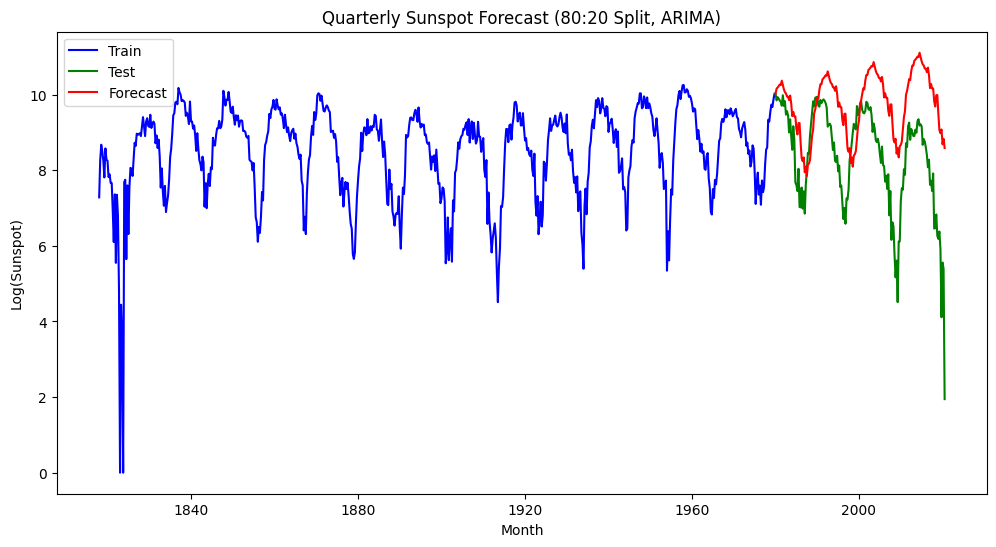
AutoCORR확인 결과 실제로 4000일(약 11년) 주기가 관측

ADF, KPSS검정 결과 ADF는 Stationary, KPSS는 Non-stationary로 판단.  
1차 차분 후는 모두 Stationary로 판단. -> 원본 데이터에 아주 약간의 추세가 관측된듯.



원본 데이터가 정상과 비정상의 경계에 있는 데이터라고 가정하고 SARIMA에 주기를 넣어 학습

원본 데이터는 학습에 시간이 너무 오래 걸려서 데이터를 3, 1개월 단위로 리샘플해 학습.



대략 예측은 잘 되는데 하필 추세가 전환되어 예측이 살짝 틀린듯. 데이터가 조금 더 있다면 추세도 잘 맞출 수 있을 듯.

처음엔계속 이래나옴

텍스트, 스크린샷, 그래프, 도표이(가) 표시된 사진

자동 생성된 설명

벤치마크(MASE)

텍스트, 폰트, 스크린샷, 라인이(가) 표시된 사진

자동 생성된 설명

New

벤치마크 대비 예측결과가 잘 안나온 이유 -> train set과 test셋을 분할하는 과정에서 장기 계절성이 관측되기 전에 데이터가 분할되어 모델이 계절성을 추세로 인식한듯. 그래서 test셋에 대한 예측을 보면 주기는 맞으나 추세 방향이 다름

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

Old(참고만) 최근 결과와 통제되지 않은 환경에서 나온 결과임. 실행 시간만 참고.

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명