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

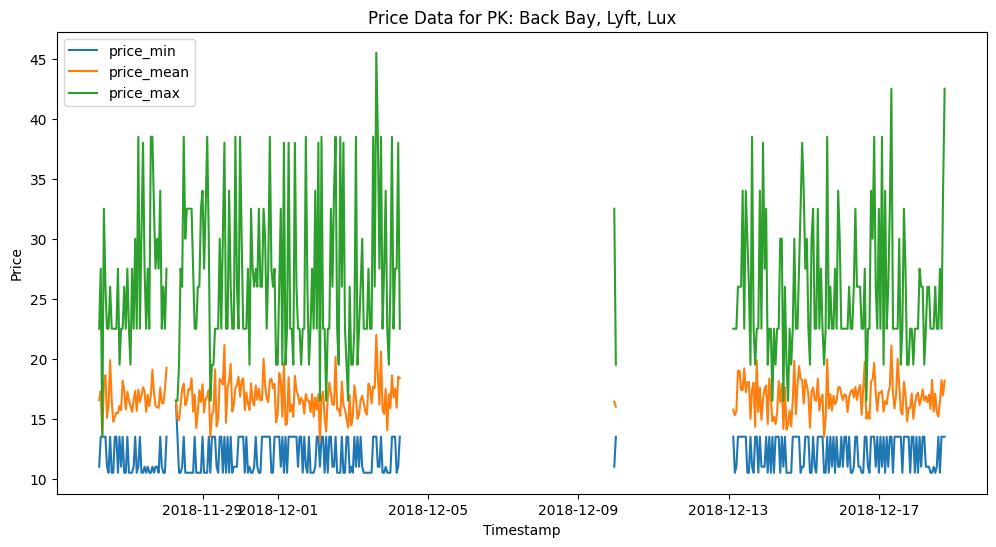
Rideshare : 2018년 11월 26일~2018년 12월 18일 뉴욕의 다양한 지역에서 Uber 및 Lyft 승차 공유 서비스와 관련된 속성의 다양한 시간별 시계열

Sunspot : 1818년~2020년 일별 흑점데이터 세트

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

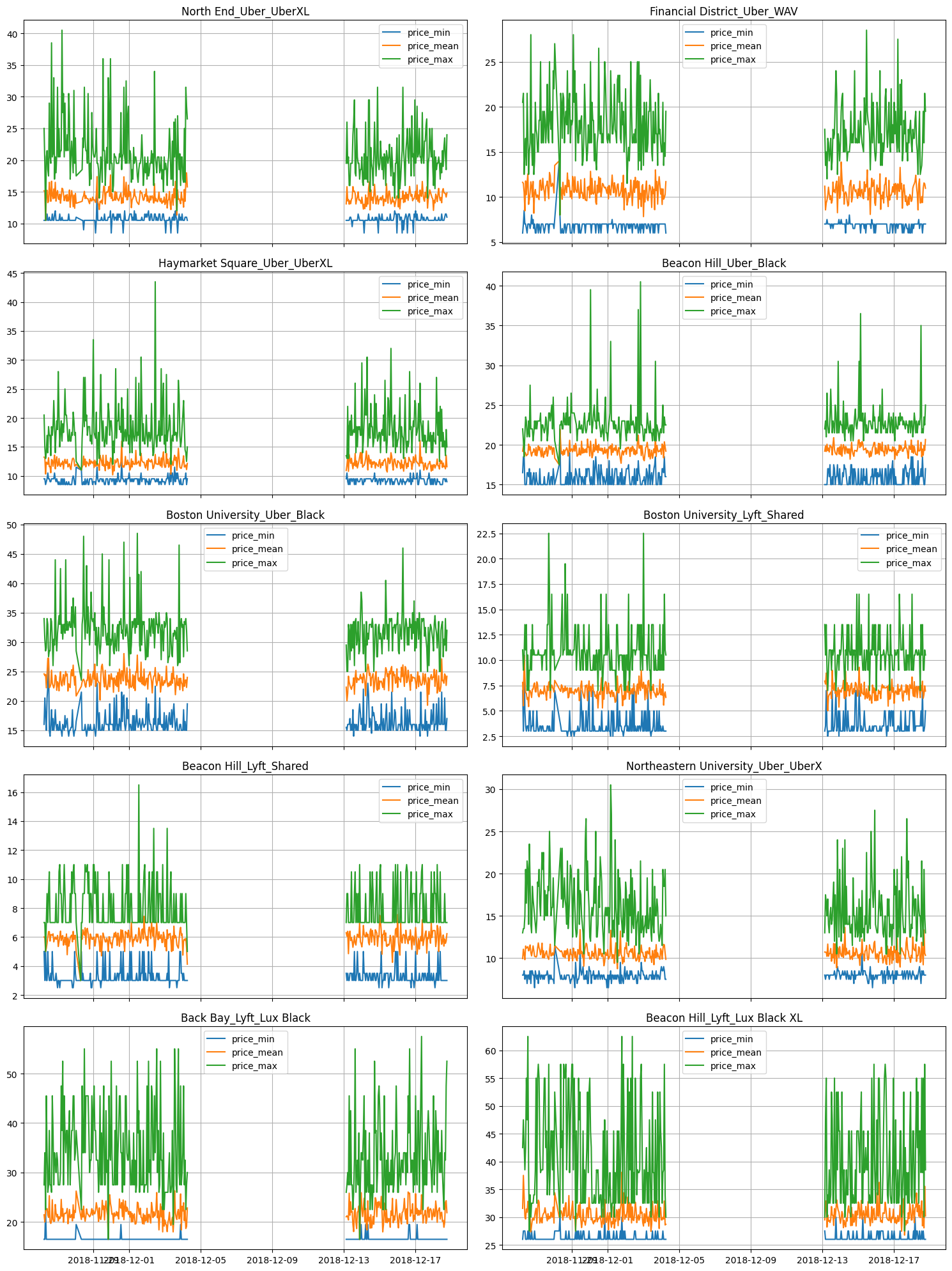
1. 전처리 및 특징 분석

Rideshare : 11월 26일~12월 18일 1시간단위 다변량 데이터



다변량 정보에는 시계열을 구분할 수 있는 Location, Provider, Service type 칼럼이 있음 이 세 칼럼을 1개의 시계열로 구분

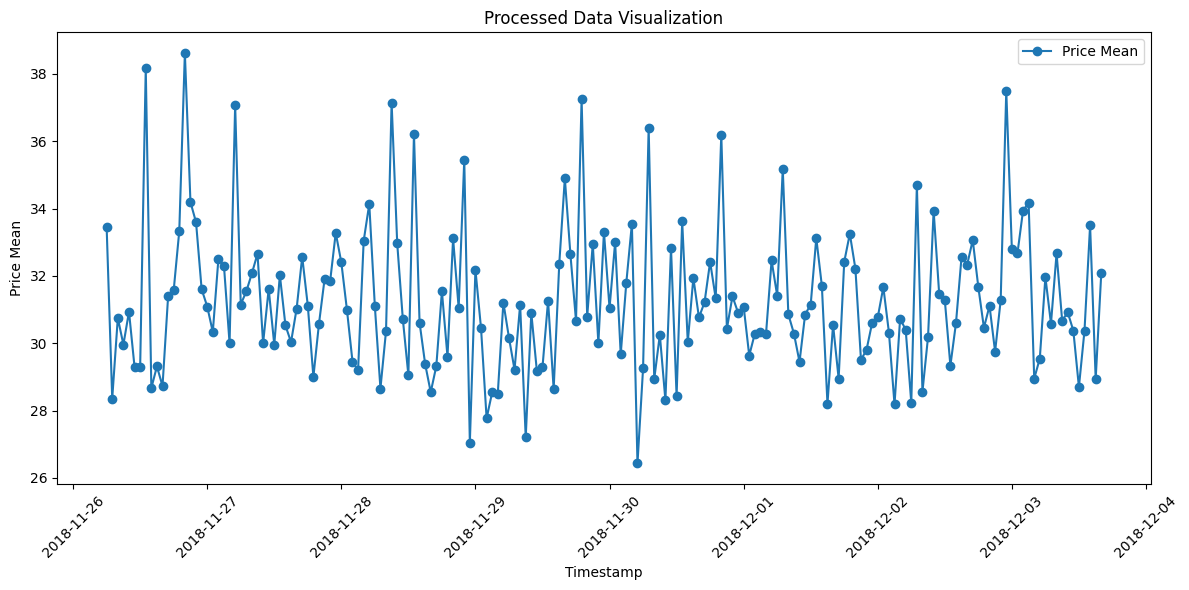
한 시계열 당 541개의 데이터

총 144개의 시계열

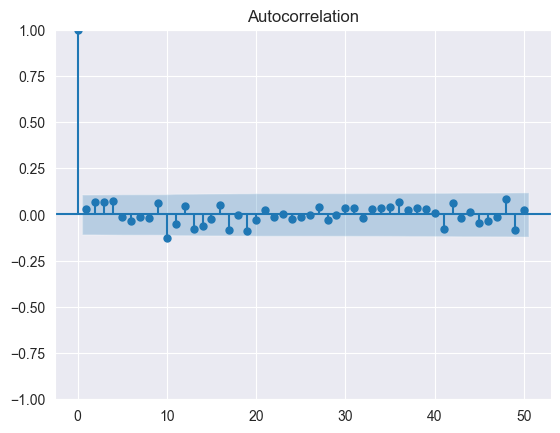
대부분의 시계열이 비슷한 결측 구간을 보임

결측 구간이 상당해서 보간이나 중앙값 대체가 애매함, 해결책으로 그냥 데이터 있는 큰 두 구간 붙여버림

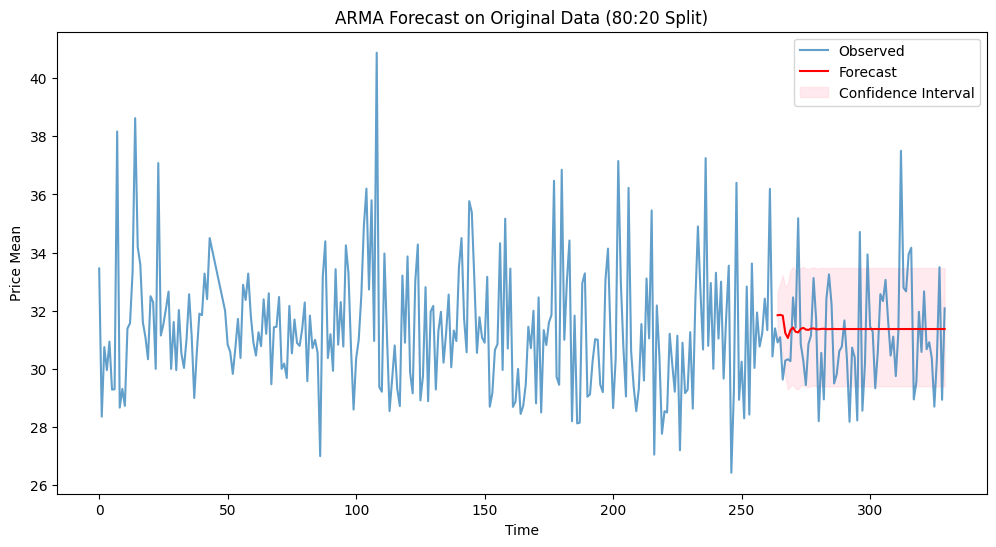
왜? -> 어차피 1달간의 짧은 시계열이고, 시계열 분포 자체가 일정한 Band가 있는 White Noise임



시계열은 Stationary. Auto CORR보면 전 구간에서 거의 상관관계가 0에 가까움 White noise



ADF, KPSS검정 결과 정상성 시계열



ARMA로 임의의 시계열을 하나 골라 예측한 결과. 거의 예측을 못하고 있음.

모델 SES, Theta, TBATS, ETS, ARIMA family, Prophet 이용

성능 및 실행시간

결과분석 벤치마크(MASE)

텍스트, 스크린샷, 폰트, 라인이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

벤치마크 대비 더 낮은 MASE를 보이나 애초에 학습이 제대로 안됐기 때문에 의미 없을 듯