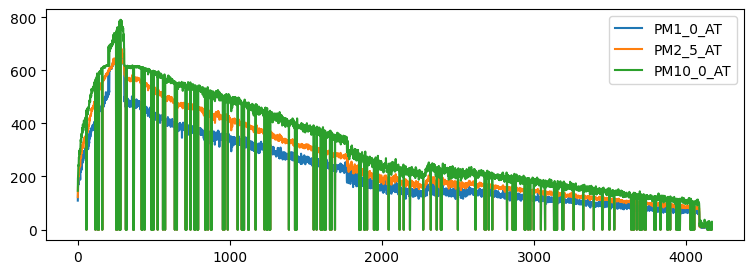
미세먼지 데이터 정제 내역

미세먼지 센서를 통해 측정한 데이터는 센서의 결함에 따라 결측치 등 오류가 있을 수 있으며, 측정 시간이 일정하지 않아 머신 러닝 처리에 오류가 발생할 수 있음. 그렇기 때문에 측정한 데이터에서 결측치를 제거하는 과정과 데이터 보정, 시간 간격을 일관되게 하는 전처리 과정을 필요로 함.

  
센서에서 측정된 원시 데이터를 그래프로 확인했을 때 다음과 같이 결측치가 발생하는 것을 확인할 수 있으며, 데이터를 직접 표로 확인해보면 일관된 시간에 측정되지 않는 것을 확인할 수 있음.  
따라서 결측치를 제거하고, 선형 회귀 등의 머신러닝을 통한 균일화 과정을 거쳐 데이터가 일정한 시간마다 배치되도록 하였으며 더 정확한 데이터를 확보하기 위해 노이즈를 제거함, 다음과 같은 전처리를 거쳐 사용할 수 있는 데이터를 다음과 같이 확인할 수 있음.  
