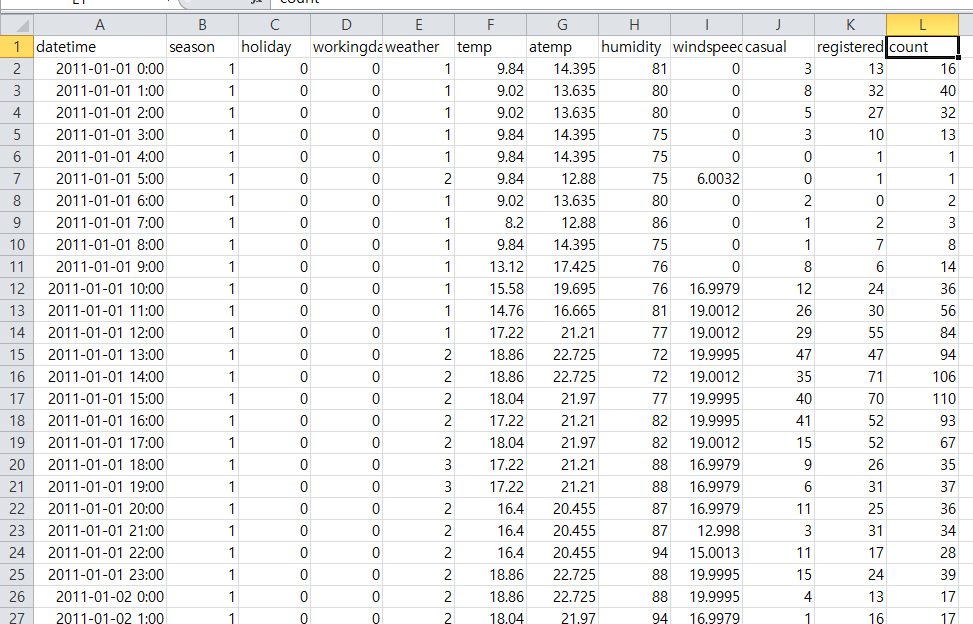
<Quiz 13>

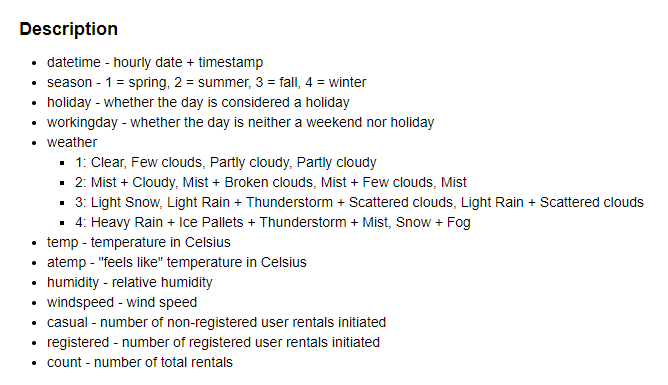
이번 시간에는 Bike sharing demand 데이터(이하 bike 데이터)를 분석해보고, 자전거 공유 시스템의 수요를 예측할 수 있는 회귀모델을 구축하고자 한다.

Bike Sharing Demand 데이터셋은 워싱턴 DC의 Capital Bikeshare 프로그램에서 자전거 대여 수요를 예측하는 Kaggle Competition에서 이용되었던 데이터셋이다. (https://www.kaggle.com/c/bike-sharing-demand)

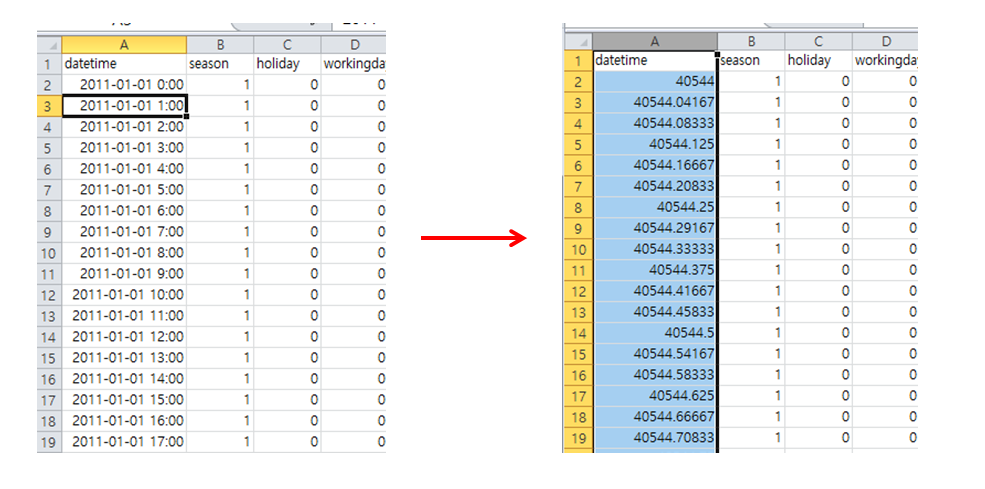
우선 train.csv 파일을 열어보면 다음과 같은 형태로 데이터셋이 주어진 것을 알 수 있다.



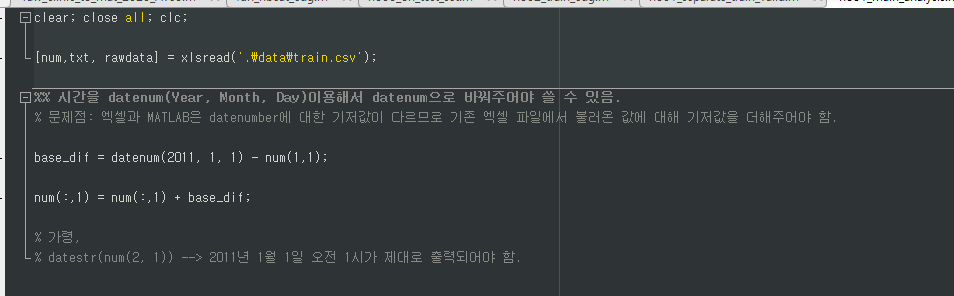
우리가 feature로 사용할 수 있는 데이터는 Datetime(날짜), season(계절. 1: 봄, 2: 여름, 3: 가을, 4: 겨울), holiday(국경일 여부), workingday(주중 여부), weather(날씨), temp(온도), atemp(체감온도), humidity(습도), windspeed(풍속)까지 이다. 뒤의 casual, registered, count는 해당 날짜에 자전거를 빌려간 횟수에 해당된다. casual은 비회원 대여 횟수, registered는 회원 대여 횟수, count는 비회원, 회원 대여 횟수를 합친 값을 의미한다.



MATLAB에서 시간 데이터를 쉽게 불러올 수 있게 하기 위해 셀 서식 조정을 통해 아래와 같이 datetime을 숫자형식으로 수정해주도록 하자.

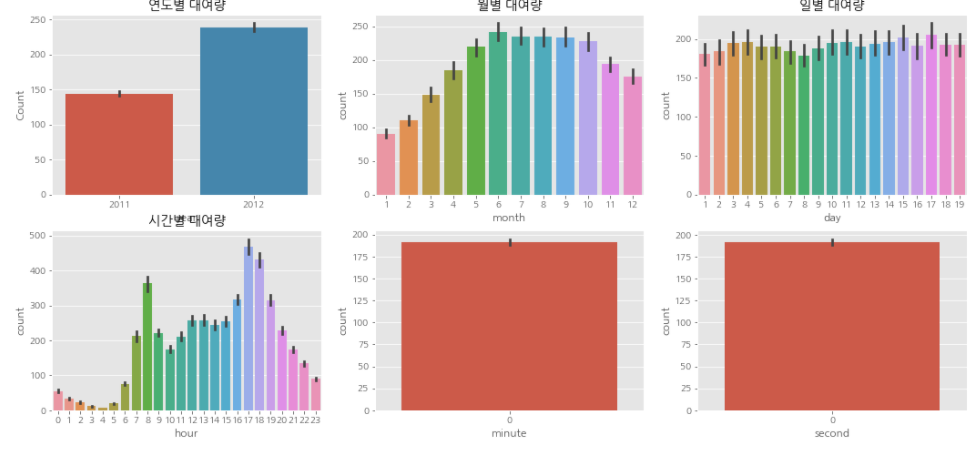


그런 다음 MATLAB에서 xlsread() 함수를 이용해 데이터셋을 불러온 뒤, 아래의 코드를 이용해 MATLAB에서 시간 데이터를 이용할 수 있도록 기저 시간을 수정해주도록 하자.



연도, 월, 일, 시간, 분, 초를 나눠서 새로운 feature로 만들어주도록 하자.

그런 다음, 여러가지 시간 단위에 따른 대여량의 차이를 확인해보도록 하자.



여러 조건에 따른 월별 count의 비교

