Data Science Challenge Question

A large segment of business data can be characterized as time series. This includes data at an individual user level or aggregate level for a business process. The attached csv file contains hypothetical evolution of a time series. For sake of example, we will refer to this as the number of support requests received at a given point in time.

1. Perform exploratory analysis of the time series using Jupyter notebook.
2. write a python function which will take as inputs:
   1. historical data in a dataframe
   2. a specified time period to forecast for and will then implement the forecasting algorithms detailed on this blog entry (<https://petolau.github.io/Forecast-double-seasonal-time-series-with-multiple-linear-regression-in-R/>) using python and scikit and modified for a 15 minute granularity
3. What are the theoretical differences between the above algorithm and convention time series forecasting algorithms such as ARIMA? What assumptions are different?
4. write pytest tests for the function and sub functions.
5. Load the historical data below and then use the method to forecast for the next six weeks following the historical data.
6. Extra Credit: How can you improve the MLR forecasting algorithm or find a better alternative algorithm?  For example, what if you need to forecast for 6 months instead of 6 weeks.  To take the solution to the next level, implement the improvement or alternative and benchmark it against the MLR algorithm.

Time series data with arrival volume for customer support requests is attached in \*.csv format.  Please note that the date times in the file are in UTC, but the support center is located in a US timezone.

Please send us the code, tests, the forecast and a visualization of the history data and the forecast together.  What do you see from looking at both the history data and the forecast data for the six weeks following?