NarrativeWave Data Challenge

Inside the datasets.zip file you will find files for this data challenge. Each file contains a timestamp column, a tag column and value column. The tag column value represents a hierarchy and can be interpreted as follows.:

Tag: /narrativewave/WTG13/Amb_WindSpeed_Avg
Site: narrativewave
Asset: WTG13
Column: Amb_WindSpeed_Avg

The data comes in “long” format but needs to be changed to “wide” format for analysis. See link for reference.

Objective: Create a Django App that allows you to visualize the data contained in the csv files.

1. Create Two Django models to store each asset and column contained in the files. A relationship should be created between the columns and the asset. So that if I want to know how many columns an asset contains I could easily see those columns.
    a. Asset Model
    b. Column Model

2. Create a function that reads multiple csv files and stores the data in parquet partition format. The partition should be by Asset, Year and Month.

3. Create a REST API Service, to collect data from the parquet files. Use GET Parameters to query the data that will be used for the plot and return the data in json format in record format. Query Parameters can be: start date, end date, assets, column.

4. Consume REST API Service in a View to plot a column of any given asset.

5. (Extra Points) If PySpark is used in this implementation.

NOTES:

The deliverable format may include the following: Jupyter Notebooks, Django Project, Py Files, Screenshots, Video Screen Recording, Hosted App, etc.

There are no limitations on the libraries to be used. No time limit for the delivery although the sooner the better. No need for an auth layer. Include running instructions.

Please send out the result to diego.cerda@narrativewave.com and good luck!