Home Assignment

We are an innovative startup that develops a platform for planning and operation, supporting the migration into autonomous manufacturing and retail.
Our core product relies on strong statistical and ML-based engines, to analyze our customers' data and bring them action-driven insights in real time.

As part of our product, we're developing a few algorithmic pipelines that "feed" the product, to tackle problems such as customer segmentation, asset utilization, supplier risk analysis and more.
In this assignment, we'll be focusing on a different task - demand forecasting. The ability to properly predict the demand into the future is essential when coming to answer the question of "how much to produce and when?", and in this assignment, you'll be given the chance to try and do just that.

The following link will redirect you to a google drive containing data and description files. The data itself represents the sales of different SKUs of a big retailer. You are required to build a model (or several - your choice) that can best predict the sales of the different SKUs sold at different stores.

https://drive.google.com/drive/folders/1QJH5meAbT101xuY-p9K1iEzuHX-QCecz

Note:
The dataset contains a lot of different SKUs. For practical reasons, feel free to sample just a subset of them, if you feel it represents the general solution. If you find value in using more SKUs - go for it :)

Required outputs:
1. The code implementing the solution you've developed. Please make sure to:
   a. Add a *requirements.txt* file in case certain dependencies were used.
   b. List any open-source code you use. Must be free and licensed for commercial use.
   c. We should be able to run your code by calling your main function. The code needs to build / train the model, and then use it to make predictions on the SKUs picked.
   d. Zip the source code with all relevant files you find to be complementary (graphs, tables, etc.)
2. Preferably, a document describing your assumptions and general solution approach. This should not be more than one page.

Stuff to consider during the work:
1. What kind of relevant EDA can you use?
2. What types of data pre-processing to use? If at all?
3. How to evaluate my models? Relevant losses, metrics, etc.
4. What time horizons to use? How far and how well into the future can I forecast?
5. Are there any interesting events that can be observed in the data?

You are welcome to contact us for any questions.

Good luck!