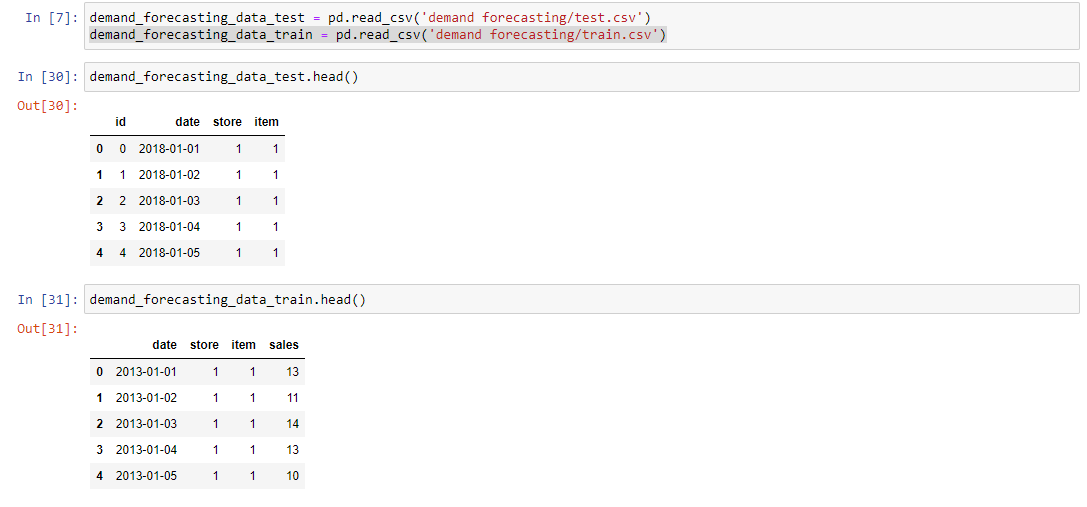
idea # 1:

<https://www.kaggle.com/c/demand-forecasting-kernels-only/overview>

Using store-item sales data to predict 3 months of sales for 50 different items at 10 different stores. This comes from a Kaggle competition. Aside from predicting 3 months of sales, other questions to explore: what is the best way to deal with seasonality?; should stores be modeled separately or together?; Does deep learning work better than ARIMA?; Can either beat xgboost? I think I could tackle the initial problem of simply predicting sales and possibly if the stores should be modeled together or separately, but I think deep learning might be above my level right now. The data looks clean and simple, but it is contained in 2 different dataframes: test and train. Also, the train file contains an additional ‘sales’ column. I am not quite sure how this all fits together, nor do I know why the dataframe is already split with train and test.

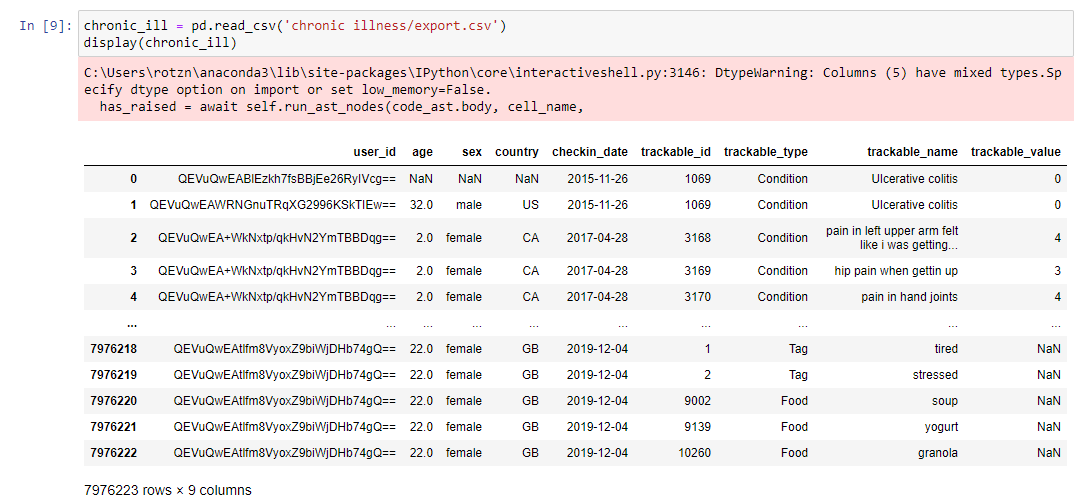


idea # 2:

<https://www.kaggle.com/flaredown/flaredown-autoimmune-symptom-tracker>

The data here is derived from the Flaredown app, which helps patients of chronic autoimmune and invisible illnesses improve their symptoms by avoiding triggers and evaluating their treatments. Each day, patients track their symptom severity, treatments and does and any potential environmental triggers (foods, stress, allergens, weather, etc.). Using this data, there is a few different questions would could address. Does X treatment affect Y symptom positively/negatively/not at all? What are the most strongly-correlated symptoms and treatments?

* Are there subsets within our current diagnoses that could more accurately represent symptoms and predict effective treatments?
* Can we reliably predict what triggers a flare for a given user or all users with a certain condition?
* Could we recommend treatments more effectively based on similarity of users, rather than specific symptoms and conditions? (Netflix recommendations for treatments)
* Can we quantify a patient’s level of disease activity based on their symptoms? How different is it from our existing measures?
* Can we predict which symptom should be treated to have the greatest effect on a given illness?
* How accurately can we guess a condition based on a user’s symptoms?
* Can we detect new interactions between treatments?



idea # 3:

<https://www.kaggle.com/fedesoriano/stroke-prediction-dataset>

From Kaggle we have a stroke prediction dataset. The main goal would be to be able predict the risk of stroke based on the given attributes. It is unclear whether the final value being predicted needs to be binary or can be a floating value. I also wonder if a simple program could be developed to assess the risk of stroke by inputting the various attributes.

