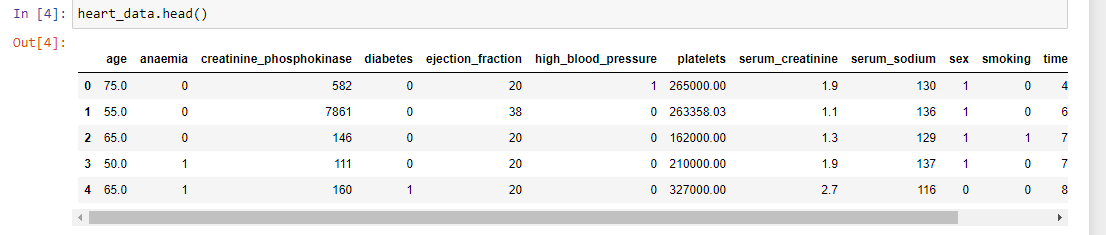
idea # 1:

<http://archive.ics.uci.edu/ml/datasets/Heart+failure+clinical+records#>

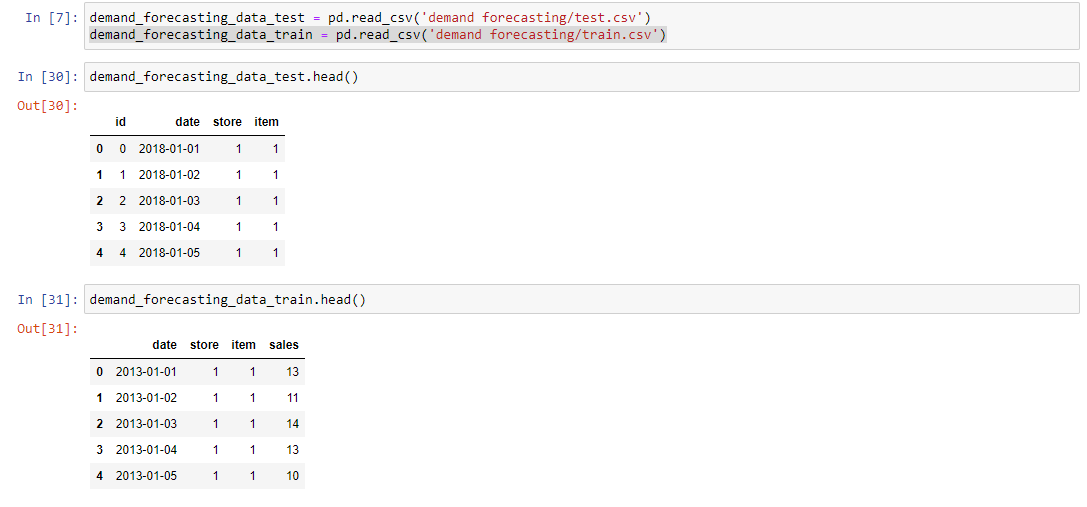
Using the Heart failure clinical records data set to predict which patients died upon follow up. This data set contains 13 attributes about for each of the 299 patients. The first 12 attributes contain the patient’s age and various clinical features, while the 13th attribute is the follow up on the study containing a Boolean about whether the patient survived or not. The goal here would be to find which attributes are best at predicting which subjects died. The data looks clean and straight forward.



idea # 2:

<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 # 3:

<http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Prognostic%29>

Using the Breast Cancer Wisconsin (Prognostic) Data Set to predict recurrence of breast cancer in patients. The first column contains the patient ID. The second column contains the binary presence or absence of recurrence. The rest of the columns contain various clinical features. However, none of the columns are labeled and there are multiple datasets that are different in ways I cannot easily determine.

