Randy’s Angels

The model itself is a Seasonal Naïve model. A naïve model sets all forecasts to be the same value as the previous observation. Similarly, a seasonal naïve model sets the forecast as the same as the corresponding value from the previous season. For example, in a quarterly forecast the first day is equal to the first day of the previous quarter, the second is equal to the previous second, and so on. This model was chosen as a result of Microsoft Azure Auto Machine Learning results. A SARIMA model was considered, however there exists either a bug in the library or a memory limit in our machines that caused kernel failure. Thus, a seasonal naïve model was the next available model with the best results.

Our model is entirely contained in the naïve.py file included with our deliverables. The time elapsed from the beginning of the file to the end of its running is 128 seconds, a large amount of that being when the data is read in. The accuracy function calculates the total RMSE per SKU. The function pysnaive helps create a seasonal naïve model, which is a function developed by Sandeep Pawar. We edited the function to better fit our needs. The function requires training data, a “lag” on the model, which we set to be 7, or one week, and a forecast horizon which is the number of days to predict for, which we also set to 7, since we are predicting the next week. The all\_sku function splits the data and tests each SKUs individually using the pysnaive function. The return is the sum of the RMSE’s for all SKUs. To test the data on the validation set, we enact the all\_sku\_valid function. First, since not all SKUs are included in the validation set, the function creates a list of skus that are in the validation set and runs over the training data from the last 60 days to generate predictions to compare to the validation set.