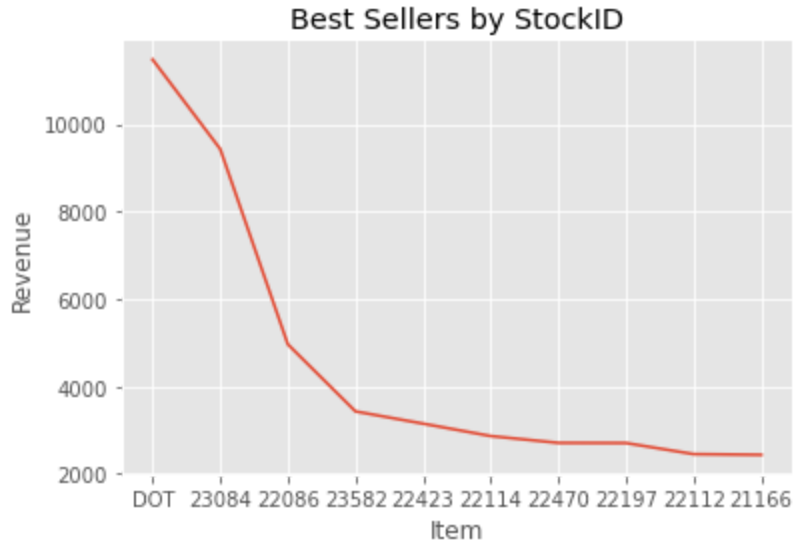
Objective

The goal is to predict sales for the period 11/27/11 through 12/3/11 for best-selling items. Best selling items are items with the greatest total revenue over this period.

This forecasting problem is not straightforward due to (a) the lack of strong yearly/seasonal trends, (b) the lack of data (dating back to 12/1/10), (c) the high degree of sparsity, and (d) irregular nature of the data.

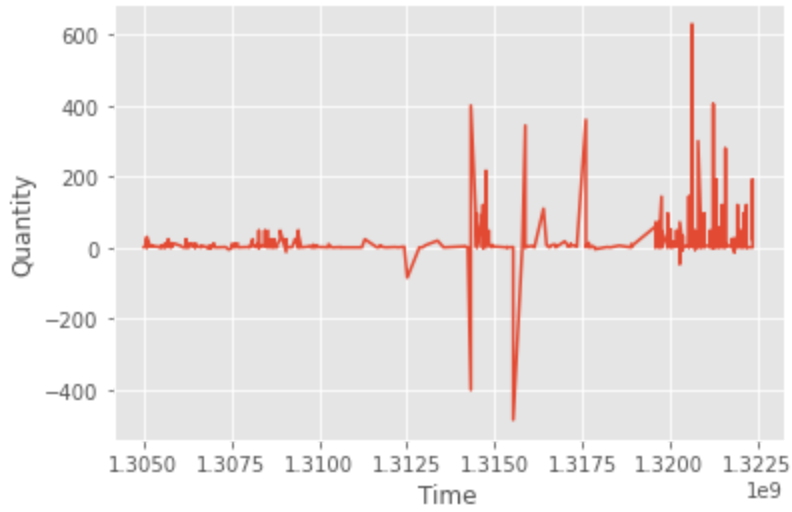
Methodology

First, I found the best selling items over this period by revenue, which was calculated as total units multiplied by unit price.



Then, I prepared the training data by limiting the dataset to include (a) stock ID corresponding to the three best selling items and (b) records through 11/26/11. Sales recorded after 11/26/11 were used as hold-out data for validation/testing.

Finally, I fit a model. Due to the lack of strong yearly and seasonal trends, it was exceedingly difficult to capture trends.



Conclusion

Predicted sales (broken out by country) for the top three selling items is provided in the attached CSV files. There is a large error when compared to actual observations, which is not unexpected given the nature of sparse and irregular time series data.

Future Steps

If I had more time, I would have liked to find reasonable ways to impute missing data.