INTRODUCTION

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, (d) irregular nature of the data, (e) presence of invoice cancellations (9,288 cancellations).

Key to my analysis is decomposing time series data to identify trends at different frequencies: year, month, day, day of week, day of year, week of year, quarter.

2. METHOD

In this section, I describe the set of transformations used to isolate the top three selling items and fit a predictive model.

Sales recorded after 11/26/11 were not used for training and were held out for validation/testing.

2.1 Identifying top three selling items

I found the best selling items over this period by revenue, which was calculated as total units multiplied by unit price. The best selling times are 23084, 22197, and 23582.

Note: DOT refers to postage and was therefore excluded from this analysis.

Best Sellers by Stockcode between 11/27/11 through 12/3/11
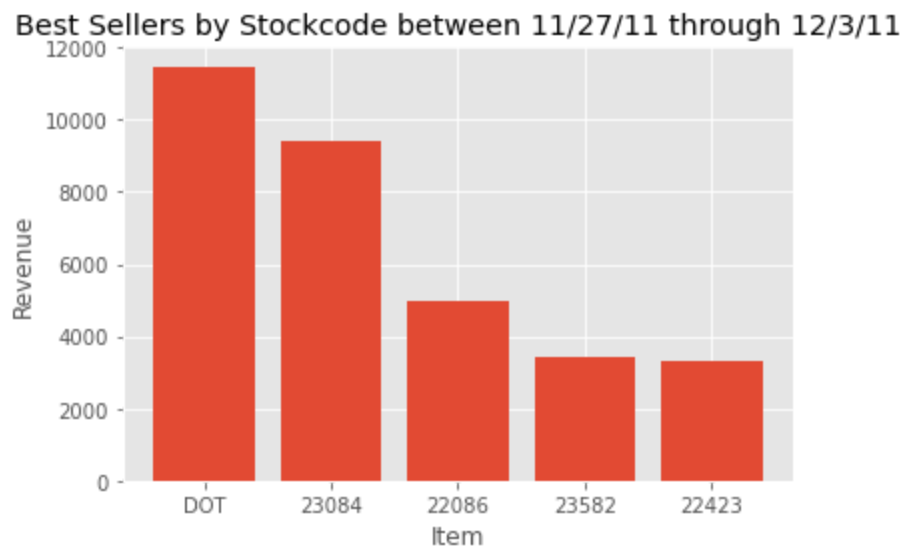
2.2 Fitting a Predictive Model

The absence of seasonal trends, coupled with sparse and irregular data, makes it difficult to fit a generalizable time series model across the three products. As a result, I fit three distinct models, one for each product.

Importantly, it appears that each of the three products had a unique combination of yearly, weekly, daily trends. Therefore, key to my analysis was viewing sales cycles from different frequencies or cycles: year, month, day, day of week, day of year, week of year, quarter.

2.2.1 Item 23582

At first glance there did not seem to be any discernible trend for item 23582.

Historical Monthly Sales for Item 23582

Decomposing into weekly shows some evidence of weekly trends:


Historical Weekly Sales for Item 23582

However, there seems to be a major trend when we look at the day of the year. The majority of sales happened on a single day of the year (32% of historical sales happened on the 336th day of the year).

Historical Daily Sales for Item 23582
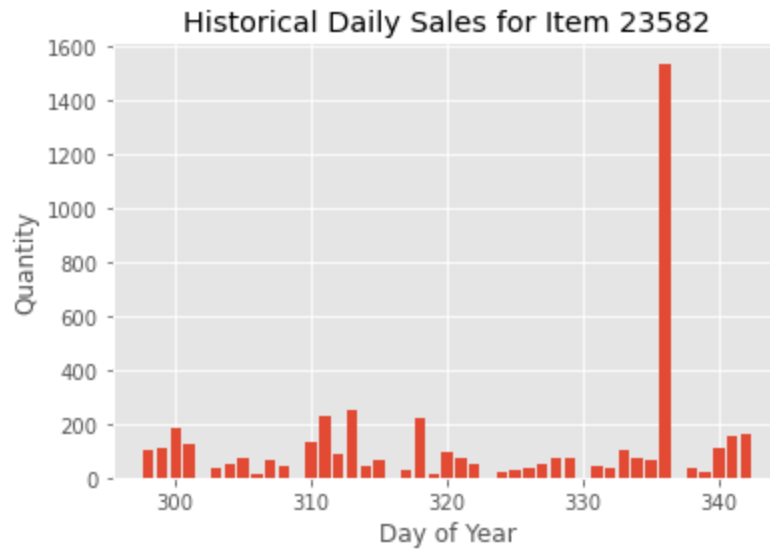
2.2.2 Item 22197

There was evidence of monthly and seasonal trends for item 22197, noticeably every three months or so. From this cycle, we can see it is 4th quarter heavy.

Historical Quartery Sales for Item 22197

Historical Monthly Sales for Item 22197

Historical Weekly Sales for Item 22197
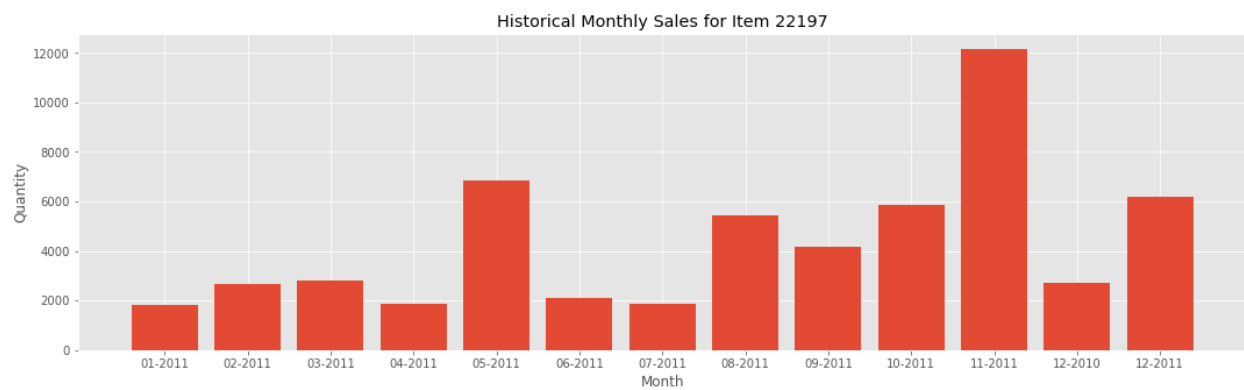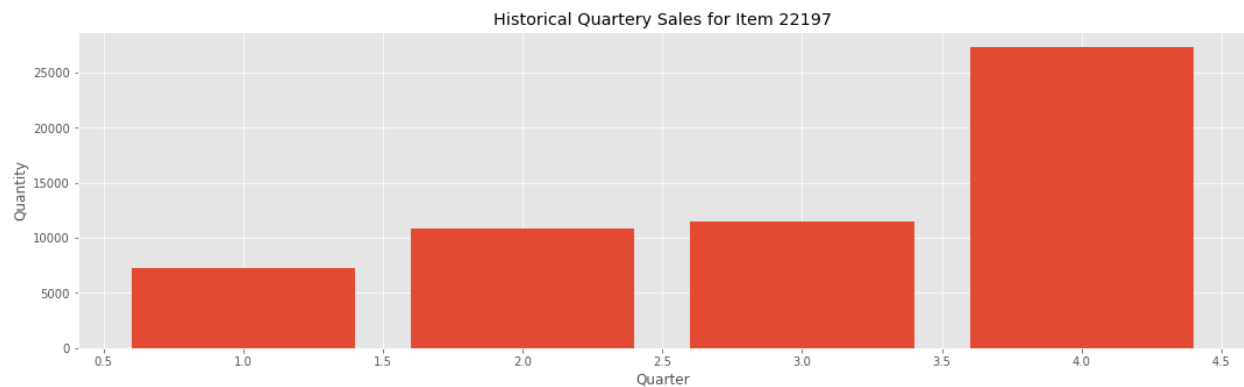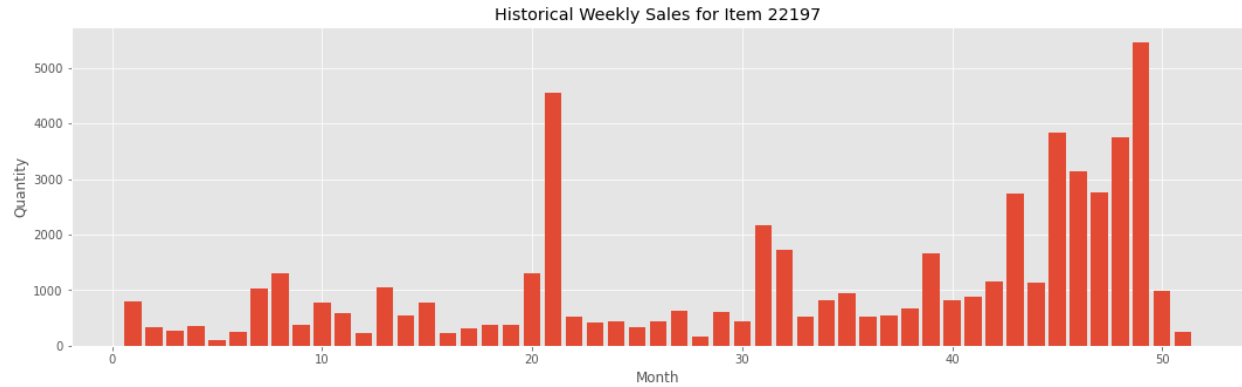
2.2.2 Item 23084

With limited data, it appears 23084 has a strong yearly trend but not a seasonal, monthly, weekly or daily one. It is likely item 23084 is an item associated with an annual holiday (do the British celebrate Thanksgiving?)

Historical Monthly Sales for Item 23084

Historical Weekly Sales for Item 23084
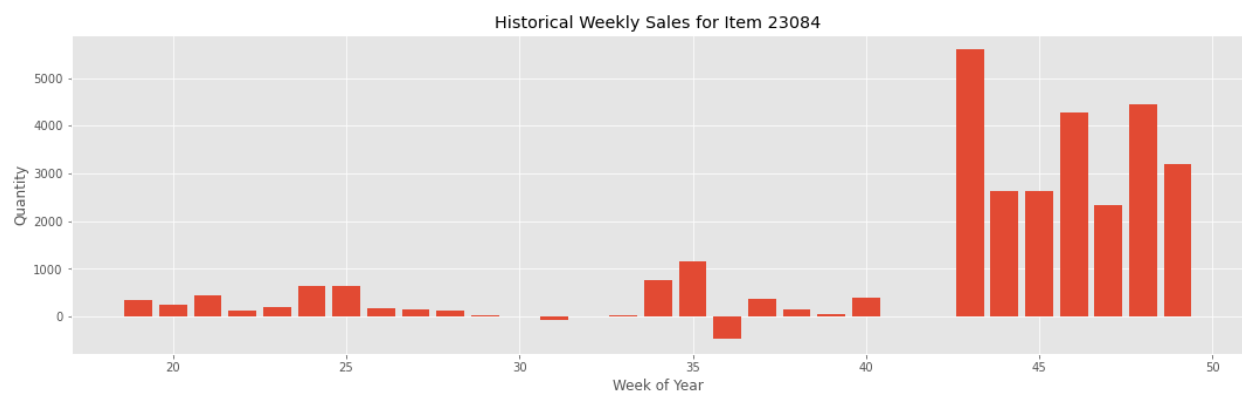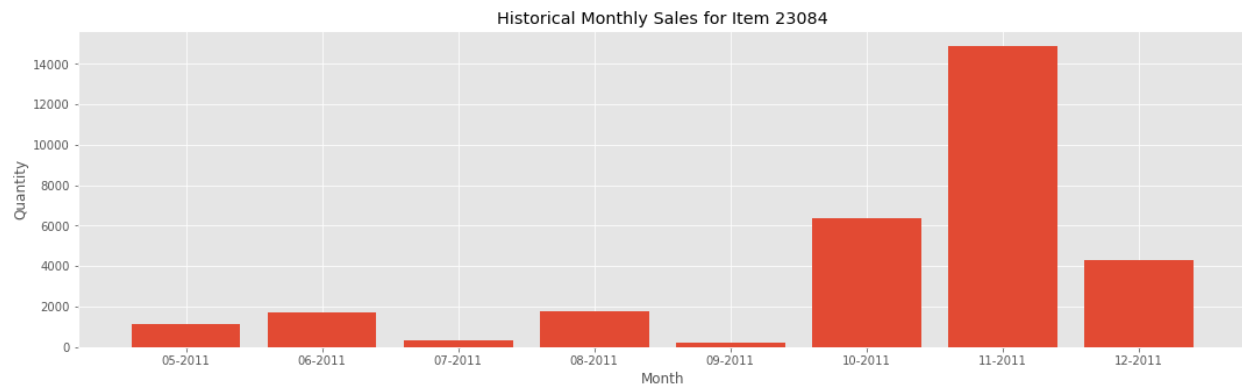
CONCLUSION

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 work should focus on imputing missing data.