Time Series Forecasting - Retail

Tatsiana Sokalava

June 8, 2021

Table of Contents

1. Introduction

a. Data Overview

The dataset shows a public, transnational data set on transactions from 12/01/2010 and 12/09/2011 for a UK-based and registered non-store online retail company.

b. Problem Statement

The store is interested in maintaining the right inventory given its historical sales.

c. Objective

Build a model that predicts sales quantities for the 7 days from 11/27/2011 to 12/3/2011 Sun – Sat for the top three selling items. Provide an estimated number of sales of these three items, broken by country.

2. Data Exploration and Wrangling

a. Data Description

The data source is provided through a public repository: http://archive.ics.uci.edu/ml/machine-learning-databases/00352/ and consists of 541,909 observations and 8 features:

InvoiceNo: Invoice number. Nominal, a 6-digit integral number uniquely assigned to each transaction. If this code starts with the letter 'c', it indicates a cancellation.

StockCode: Product (item) code. Nominal, a 5-digit integral number uniquely assigned to each distinct product.

Description: Product (item) name. Nominal.

Quantity: The quantities of each product (item) per transaction. Numeric.

InvoiceDate: Invoice Date and time. Numeric, the day and time when each transaction was generated.

UnitPrice: Unit price. Numeric, Product price per unit in sterling.

CustomerID: Customer number. Nominal, a 5-digit integral number uniquely assigned to each customer.

Country: Country name. Nominal, the name of the country where each customer resides.

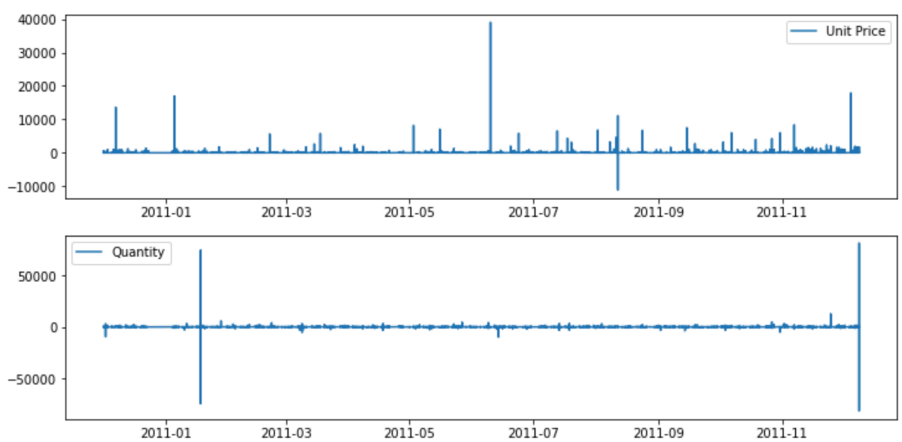
b. Data Visualization and Insights

While reviewing the dataset's summary totals, it becomes apparent there are missing values in the Description and CustomerID columns. Also, descriptions are inconsistent for the same SKUs and take different values. Knowing all CustomerID values would be useful for other modeling applications such as customer segmentation, but it is out of scope for our objective.

Based on the dataset grouping by the Country column, we see that majority of transactions are from the United Kingdom (UK).

```
United Kingdom    495478
Germany             9495
France              8557
EIRE                8196
Spain               2533
Netherlands         2371
Belgium             2069
Switzerland         2002
Portugal            1519
Australia           1259
Name: Country, dtype: int64
```

To understand how Unit Price and Quantity value change over time, we visualize it by looking at the plot. It is evident that some unit prices take on zero and negative values. Also, there are two symmetrical spikes in the quantity plot.



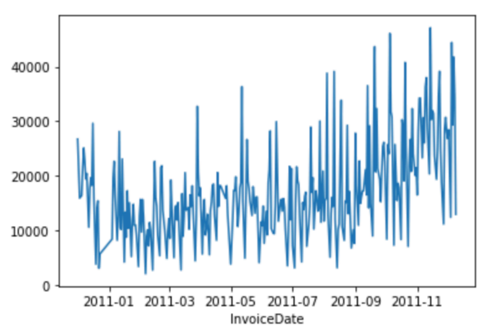
Upon further review, 2517 observations have a unit price less than zero with 1454 of these records missing a description. Here is the summary of these records:

```
check                        159
?                             47
damages                       45
damaged                       43
found                         25
                             ...
???                            1
PAPER BUNTING RETROSPOT        1
damages/display                1
amazon sales                   1
check?                         1
Name: Description, Length: 377, dtype: int64
```

While exploring the Quantity values, we see that the two spikes on the plot above are attributed to orders that were canceled on 2011/01/18 and 2011/12/09. The top five cancellations in the table below represent 61% of all canceled observations. The further review reveals that the majority is canceled the same day and therefore can remain in the dataset when aggregating by date.

	InvoiceNo	StockCode	Description	Quantity	UnitPrice	CustomerID	Country
InvoiceDate							
2011-12-09	C581484	23843	PAPER CRAFT , LITTLE BIRDIE	-80995	2.08	16446.0	United Kingdom
2011-01-18	C541433	23166	MEDIUM CERAMIC TOP STORAGE JAR	-74215	1.04	12346.0	United Kingdom
2010-12-02	C536757	84347	ROTATING SILVER ANGELS T-LIGHT HLDR	-9360	0.03	15838.0	United Kingdom
2011-04-18	C550456	21108	FAIRY CAKE FLANNEL ASSORTED COLOUR	-3114	2.10	15749.0	United Kingdom
2011-04-18	C550456	21175	GIN + TONIC DIET METAL SIGN	-2000	1.85	15749.0	United Kingdom

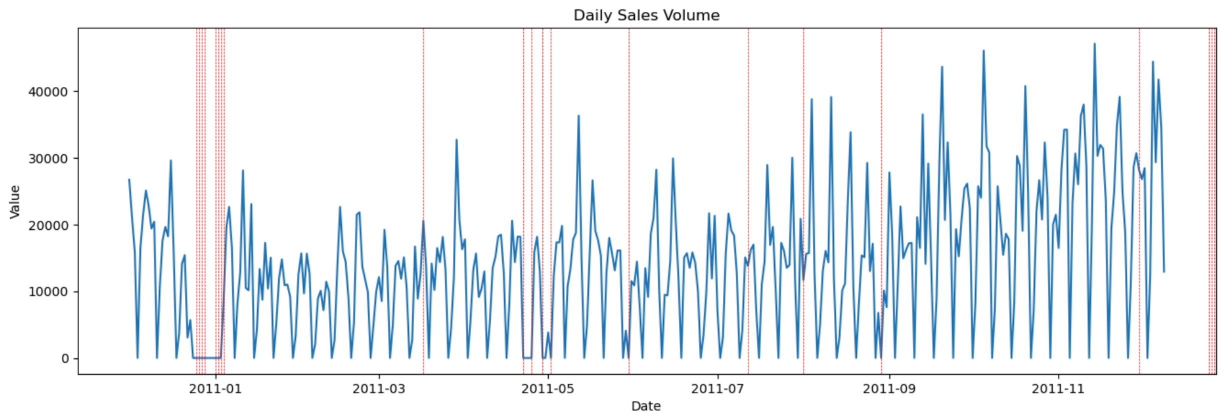
The following plot shows quantities sold by date and reveals a positive dynamic towards the end of 2021.



When aggregating, we notice that some dates are missing from the dataset as there are no sales. It appears that the retailer is closed on Saturdays and some other weekdays, likely due to the holidays.

```
Saturday     53
Monday        6
Friday        4
Sunday        3
Wednesday     1
Thursday      1
Tuesday       1
Name: 0, dtype: int64
```

To visualize, we add holidays and zero sales dates to the data and plot it:



During the times corresponding to Christmas and Easter holidays (pictured in red), there are no sales.

c. Data Preparation and Wrangling

The following data preparation is performed based on the exploratory analysis described in the previous section:

- Set InvoiceDate column as an index
- Convert InvoiceNo to a string
- Remove observations that have zero or less unit price
- Remove negative quantity observations that are not part of canceled transactions

The following data wrangling is performed to prepare the data:

- Creating a data frame with grouped values by InvoiceDate, summarized by quantities sold
- Applying a defined function to ensure a consistent dates range and adding zero quantities
- Identifying the top three items sold by filtering the above data frame using dates between 2011/11/27 and 12/03, and grouping it by StockCode.

Here are the top three items based on the sales quantity:

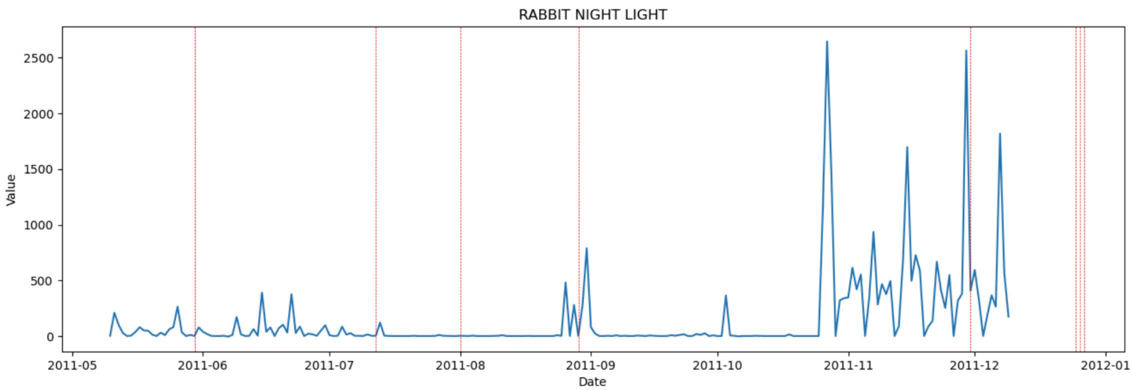
- 4588: ‘Rabbit Night Light’
- 3195: ‘Popcorn Holder’
- 1851: ‘Vintage Doily Jumbo Bag Red.’

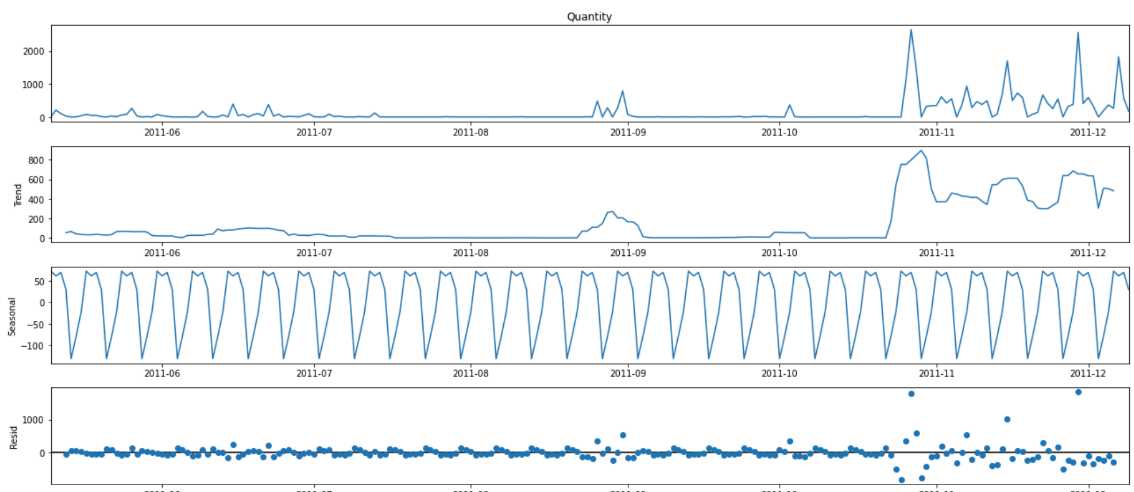
StockCode	Quantity
23084	4588
22197	3195
23582	1851

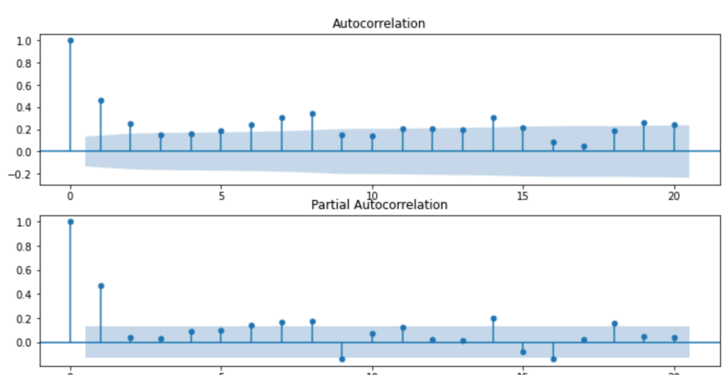
d. Reviewing the Top Three SKUs

Next, we plot the top three items' time series, decompositions, and autocorrelation functions to get familiar with their distributions:

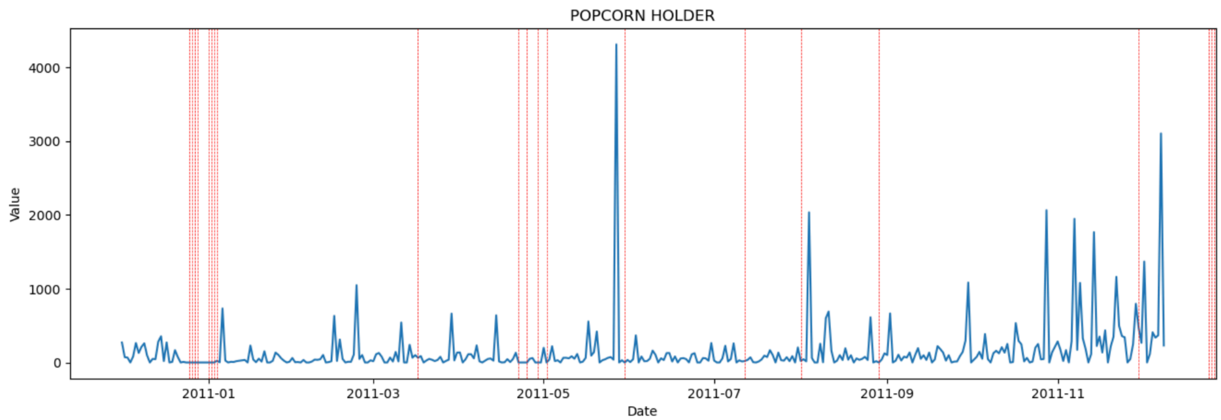
- 4588: 'Rabbit Night Light' data span is 214 days 2011/05/10 - 12/09. It has a somewhat increasing trend and shows seasonality. The residuals plot has significant variance at later observations signaling that the model cannot decompose well. The data exhibits some correlation as evident by spikes at several lags. The data may not be stationary and differencing is required to fit some models.

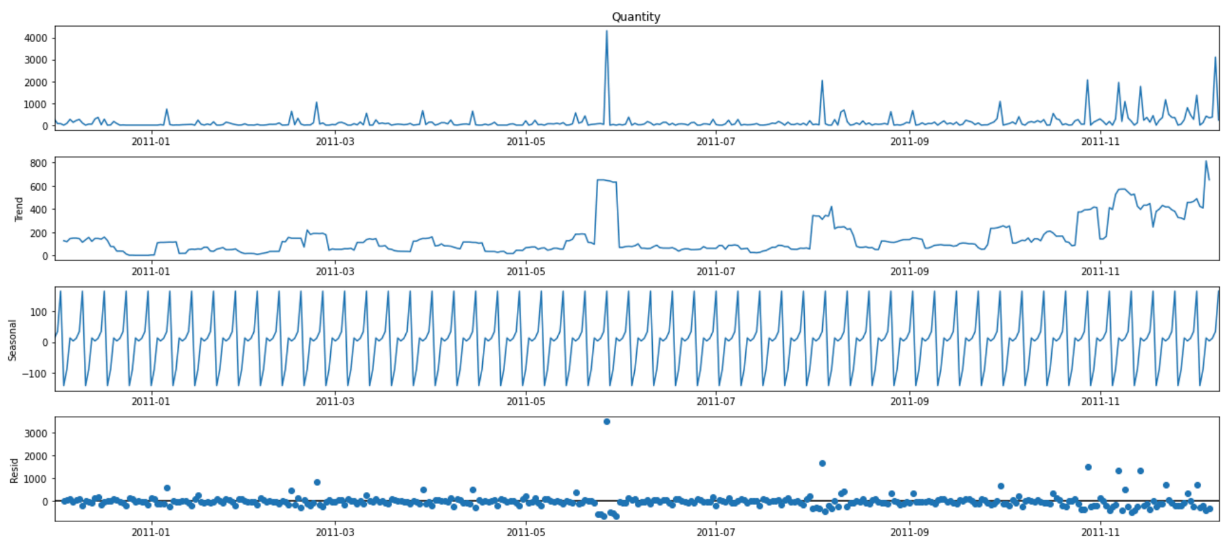


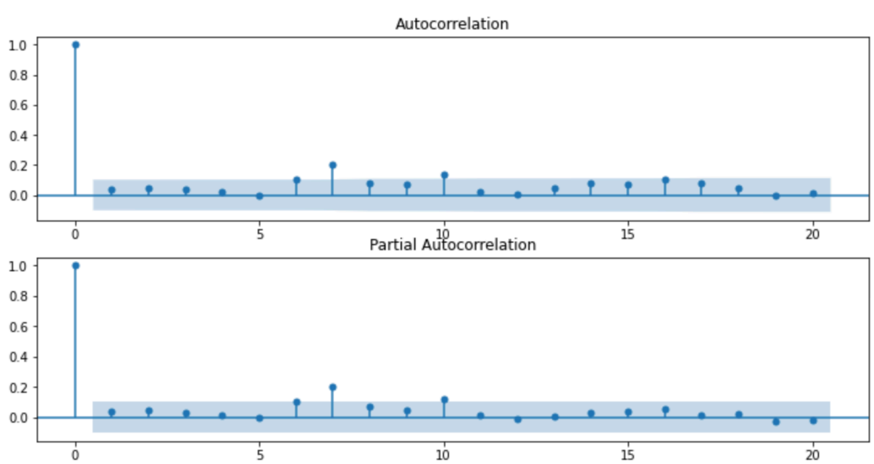




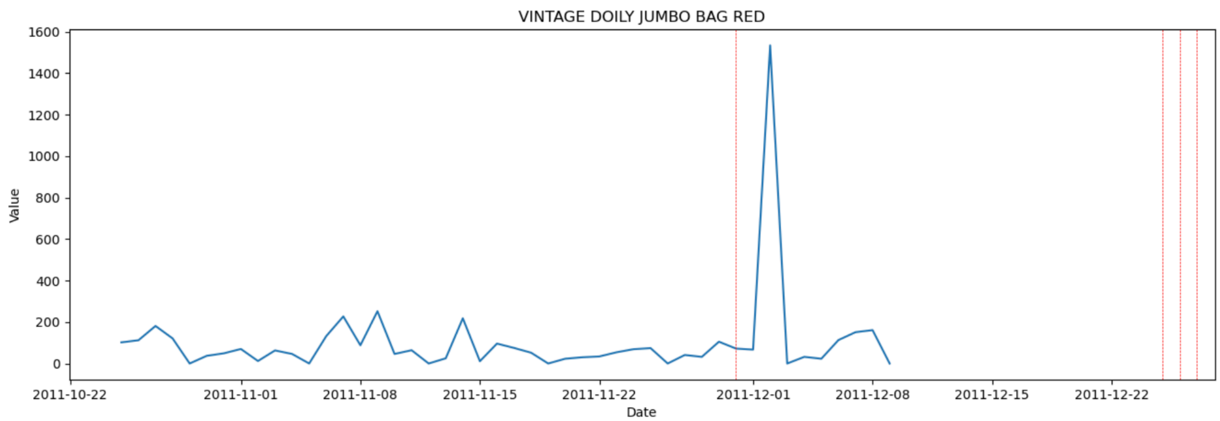
- 3195: ‘Popcorn Holder’ data span is 374 days 2010/12/01 - 12/09. It also shows a somewhat increasing trend and seasonality. The data exhibits some correlation as evident by the spikes at seventh and tenth lags.

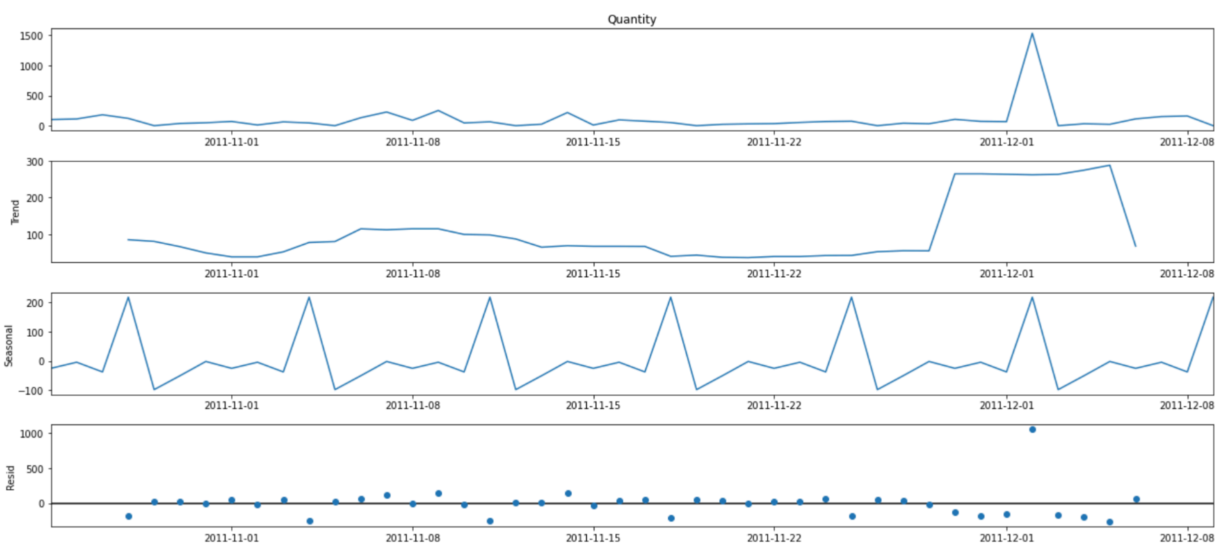


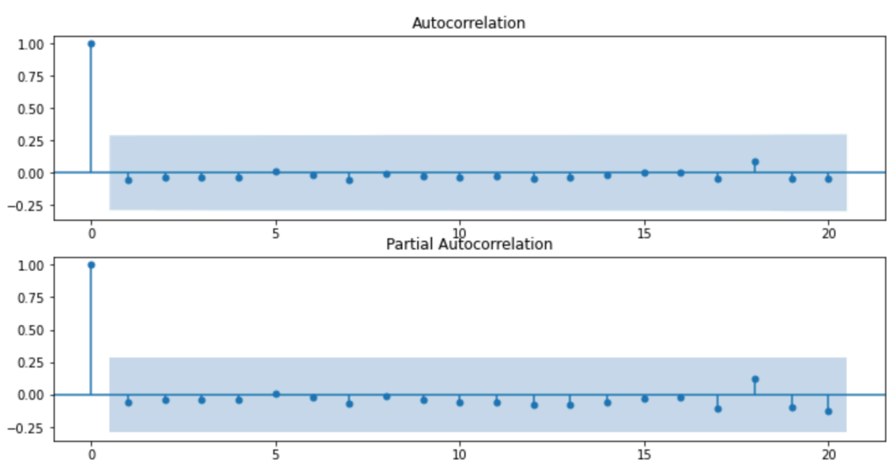




- 1851: ‘Vintage Doily Jumbo Bag Red’ data span is 46 days 2011/10/25 - 12/09. It also shows a somewhat increasing trend and seasonality. The data doesn’t exhibit correlation with lags of itself. Yet, a large spike of sales observation at the beginning of December shows that there will be challenges fitting a model using our historical data.





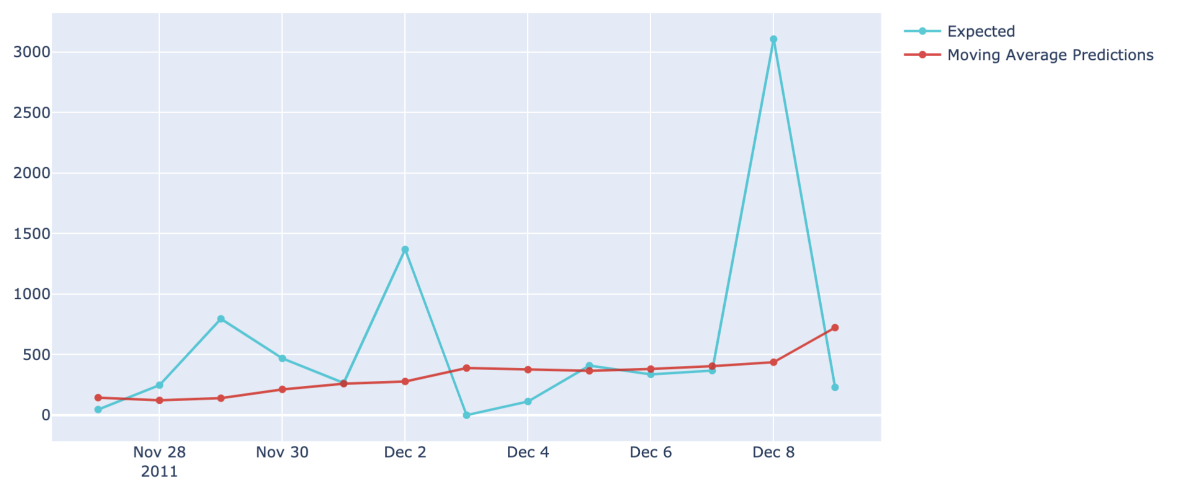


3. Fitting Models

Since 3195: 'Popcorn Holder' has the largest span of data, and, based on the plots, shows properties somewhere in between the other two SKUs, we will use it for model evaluation and fitting. We select seven different models that are popular for time series forecasting. The underlying goal for selecting seven models is to test these and pick the ones that perform best. Please note that some models are fit in their simplistic form without parameter optimization, cross-validation, or additional feature engineering. For this regression problem, root mean squared error (RMSE) is used to measure model performance. Predicted values are compared to the test values defined as observations from 2011/11/27 and forward.

1. Moving Average

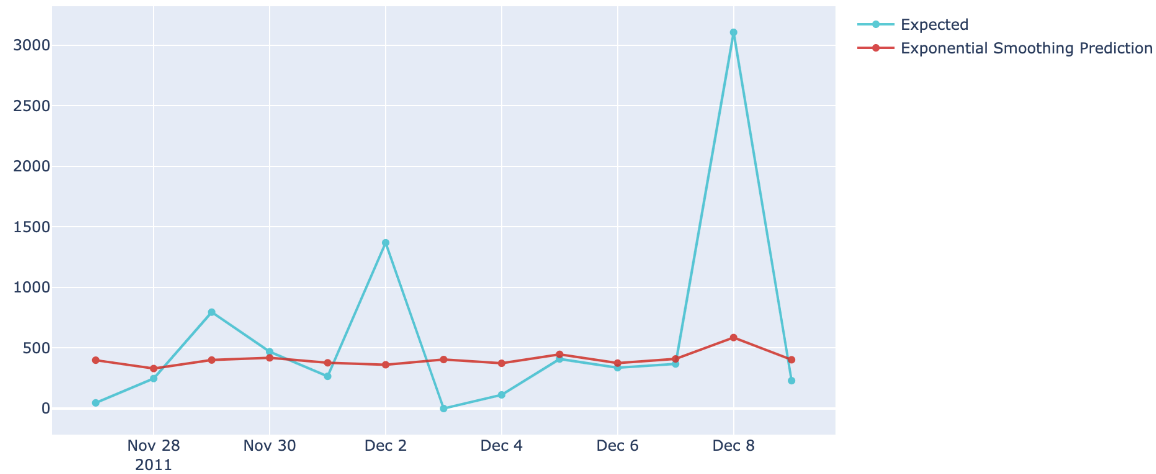
It is a naïve model where we use moving averages to form a prediction. We iterate the model across 12 possible windows to select a window with the lowest RMSE on the train data. RMSE on the test set is 746, which is pretty high. The existing trend and seasonality have a negative effect on the model's performance.



Models Popcorn Holder	RMSE	Parameters
Moving Average	746	Window size: 10

2. Exponential Smoothing

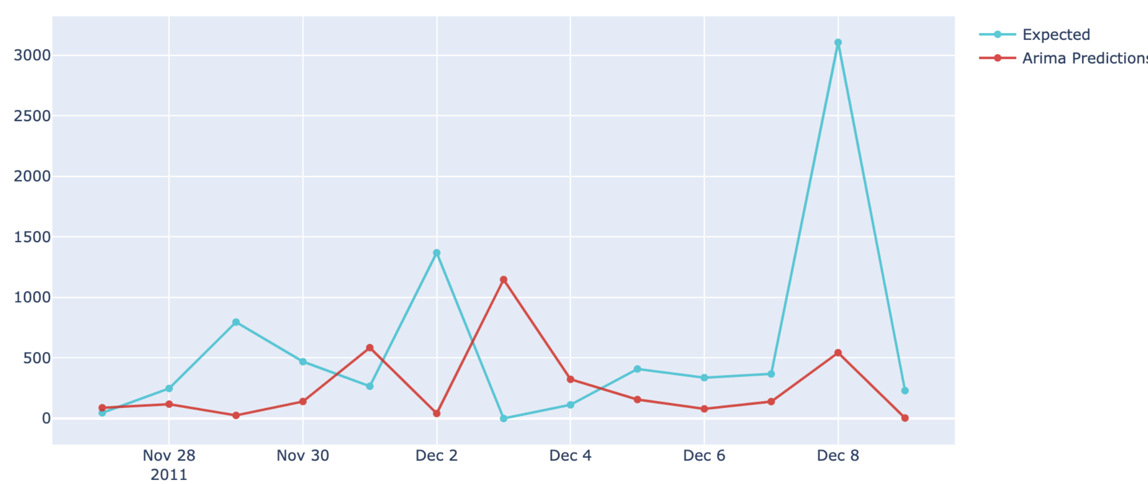
We apply Holt's Winters seasonal exponential smoothing model with 52 seasonal periods. This model allows for the level, trend, and seasonality patterns to change over time. The test RMSE for this model increased to 781.

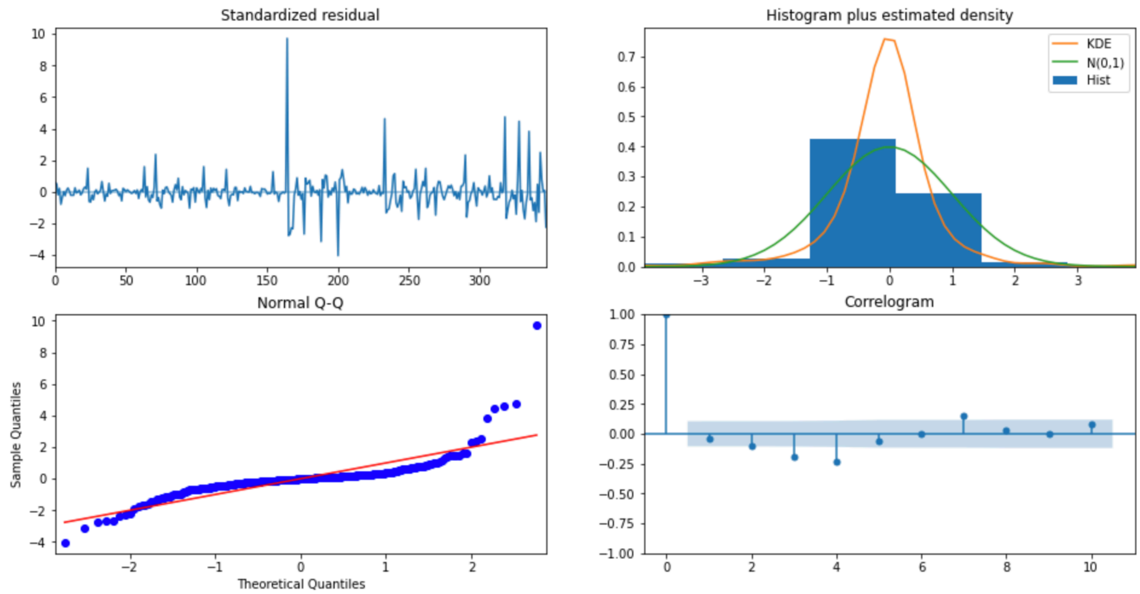


Models Popcorn Holder	RMSE	Parameters
Moving Average	746	Window size: 10
Exponential Smoothing	781	Seasonal_periods =52

3. ARIMA

ARIMA, which stands for Autoregressive (AR) Integrated (I) Moving Average (MA), is frequently used to model time series data. The downside of this model is that it has poor performance on data that is not stationary, has outliers, and shows weekly or daily seasonality. We apply the seasonal ARIMA (or SARIMA) model to our training data using auto_arima function that helps to determine the best model parameters such as p, d, q and P, D, Q. ARIMA (3, 1, 0) (2, 1, 0) [12] is fitted and used for prediction on the test data. The model yields one of the highest RMSE of 909. By looking at the residuals and autocorrelation function plots, we see that it does not fit well.

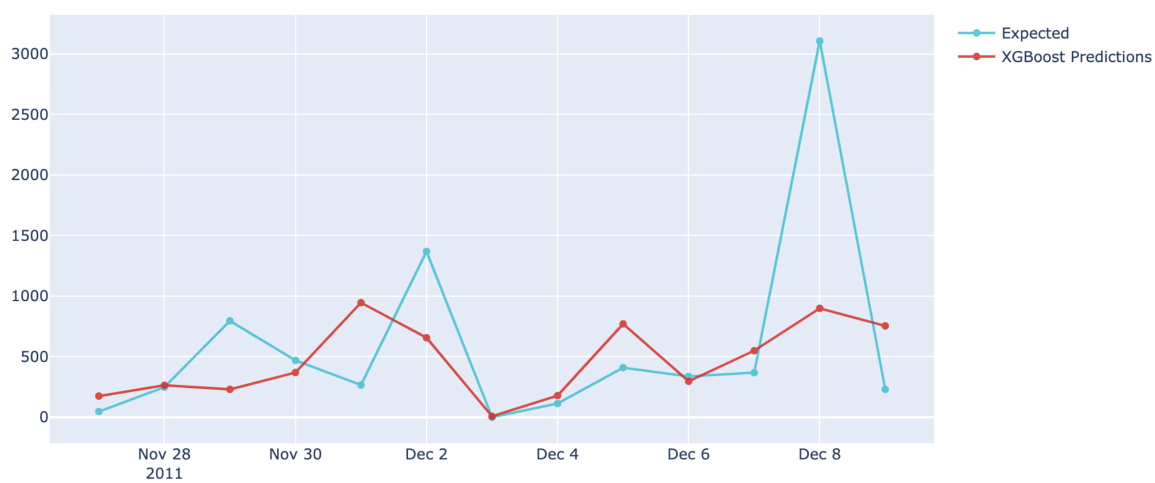




Models Popcorn Holder	RMSE	Parameters
Moving Average	746	Window size: 10
Exponential Smoothing	781	Seasonal_periods =52
Arima	909	ARIMA(3,1,0)(2,1,0)[12]

4. XGBoost

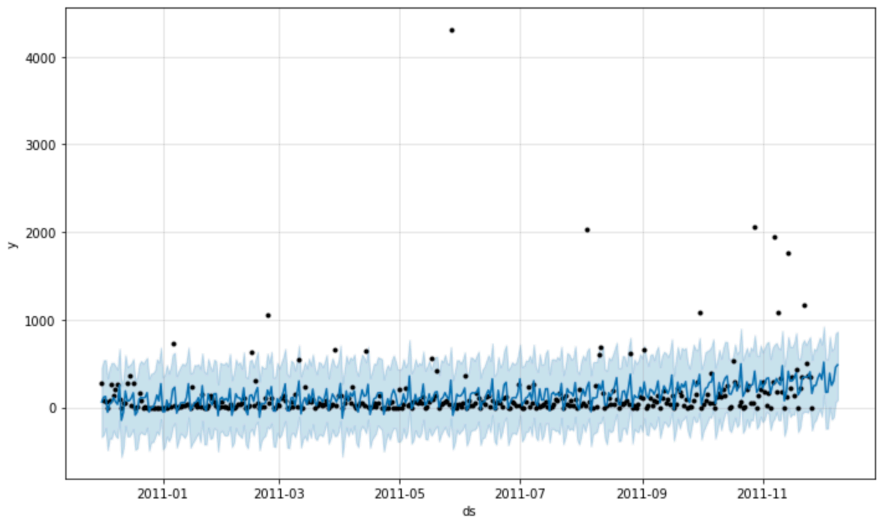
XGBoost is one of the well-known machine learning models that is also applied to time series data. It uses the gradient boosting framework and has a variety of tuning parameters for cross-validation, regularization, trees, etc. We define functions to transform, split, forecast, and test data to receive an RMSE of 714 without much tuning. So far, that is the best RMSE.

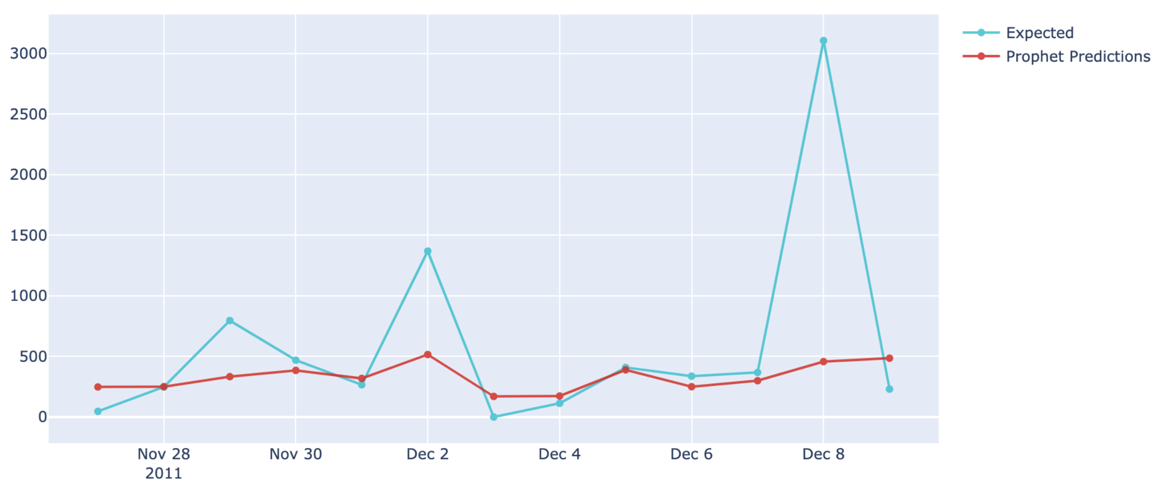


Models Popcorn Holder	RMSE	Parameters
Moving Average	746	Window size: 10
Exponential Smoothing	781	Seasonal_periods =52
Arima	909	ARIMA(3,1,0)(2,1,0)[12]
XGBoost	714	n_estimators=1000, max_depth=5,min_child_weight=1

5. FB Prophet

FB Prophet, released by Facebook, has gained popularity for its ability to forecast time series data modeling non-linear trends with yearly, weekly, and daily seasonality (using Fourier order), plus holiday effects. We add the UK holiday calendar along with seasonality parameters to the model to get an RMSE of 790.



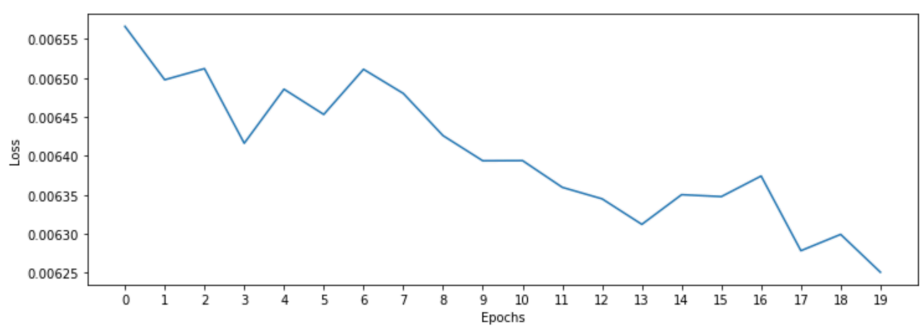


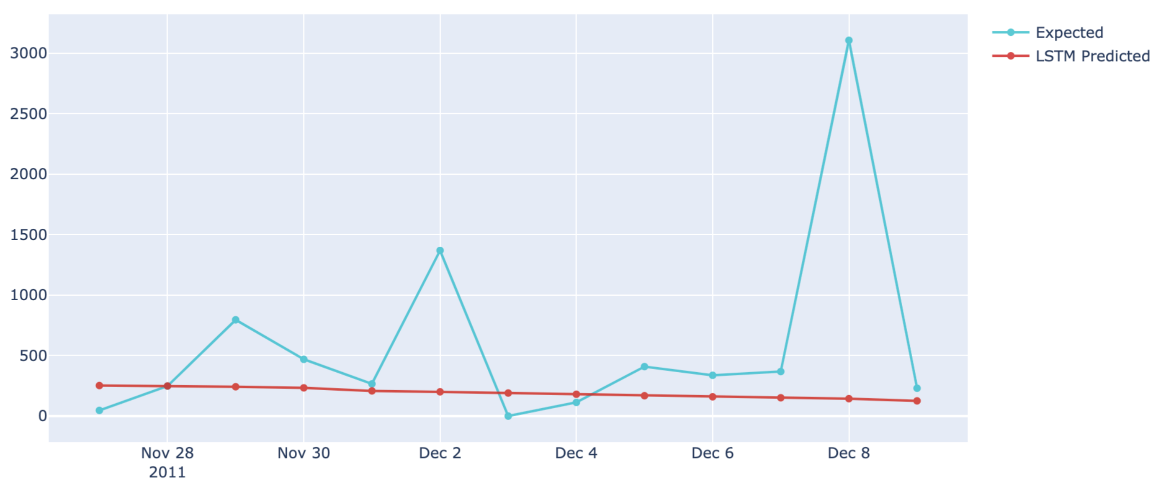
Models Popcorn Holder	RMSE	Parameters
Moving Average	746	Window size: 10
Exponential Smoothing	781	Seasonal_periods =52
Arima	909	ARIMA(3,1,0)(2,1,0)[12]
XGBoost	714	n_estimators=1000, max_depth=5,min_child_weight=1
Prophet	790	Holidays, Seasonality: W, D, Y

6. LSTM

LSTM stands for long short-term memory and it belongs to a family of recurrent neural networks (RNNs) capable of learning order dependence in sequence prediction problems such as time series. LSTM is a complex area of deep learning that requires time investment and expertise when fitting the model to leverage great results. A simple application of LSTM is applied to the scaled data resulting in a test RMSE of 919.

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm_1 (LSTM)                (None, 100)               40800
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 101
=================================================================
Total params: 40,901
Trainable params: 40,901
Non-trainable params: 0
_________________________________________________________________
```

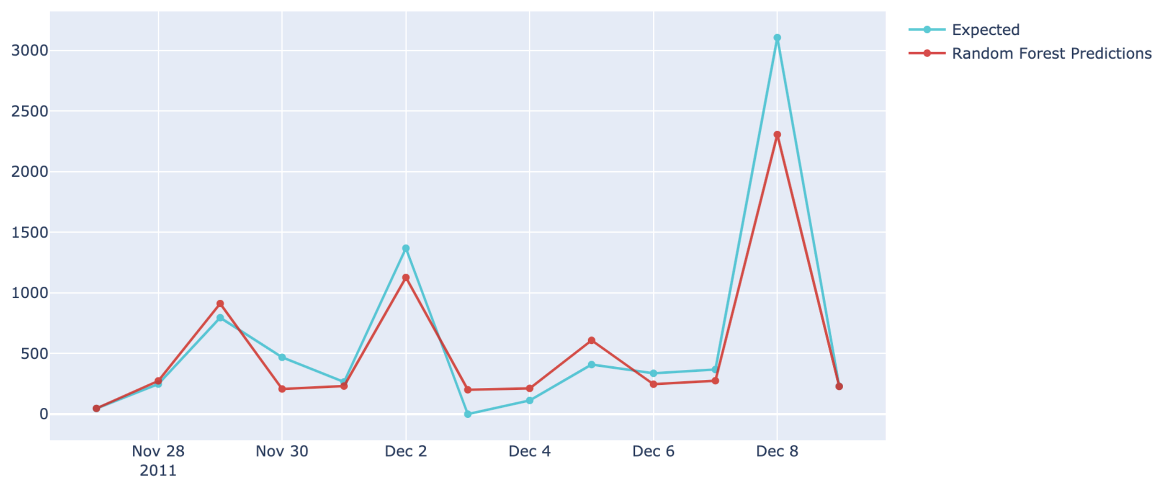




Models Popcorn Holder	RMSE	Parameters
Moving Average	746	Window size: 10
Exponential Smoothing	781	Seasonal_periods =52
Arima	909	ARIMA(3,1,0)(2,1,0)[12]
XGBoost	714	n_estimators=1000, max_depth=5,min_child_weight=1
Prophet	790	Holidays, Seasonality: W, D, Y
LSTM	838	LSTM: 100, act = 'relu', input_shape: 16x1

7. Random Forest

Random Forest is a popular machine learning algorithm based on an ensemble of decision trees at its core. A prediction on the regression problem such ours is based on the average of the prediction across the trees. We use differenced values and extend the data by adding four lags and sixteen observations in a rolling mean window. We get the lowest RMSE of 261.



Models Popcorn Holder	RMSE	Parameters
Moving Average	746	Window size: 10
Exponential Smoothing	781	Seasonal_periods =52
Arima	909	ARIMA(3,1,0)(2,1,0)[12]
XGBoost	714	n_estimators=1000, max_depth=5,min_child_weight=1
Prophet	790	Holidays, Seasonality: W, D, Y
LSTM	838	LSTM: 100, act = 'relu', input_shape: 16x1
RF	261	Differenced: 1, Lags: 4 , RM Window: 16

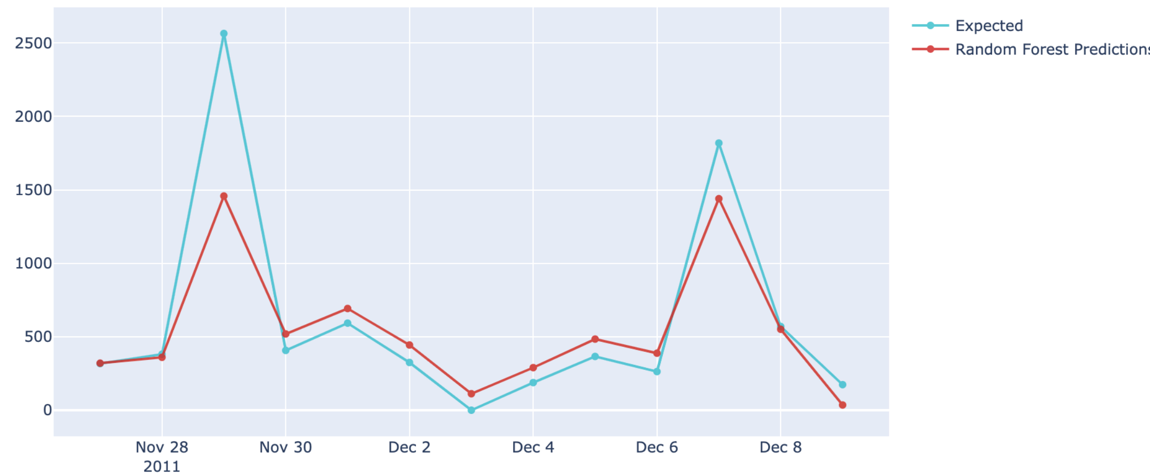
It seems that the application of the Random Forrest model yields the best result across all the models tested. Therefore, we will use it for the other two top-selling items 4588: 'Rabbit Night Light' and 1851: 'Vintage Doily Jumbo Bag Red.'

4. Model Results for the Top 3

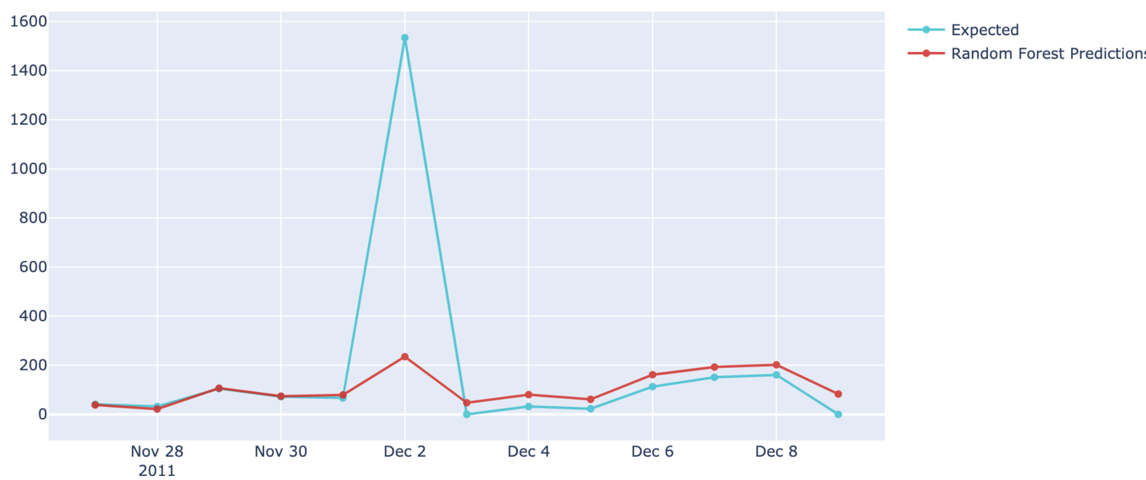
a. Fitting models for the remaining SKUs

We define a function to fit the Random Forest model adjusting parameters for lags and window size.

- For 4588: 'Rabbit Night Light,' we use 4 lags and no moving average in the model to get RMSE of 337.



- For 1851: 'Vintage Doily Jumbo Bag Red,' we use 4 the same parameters to get RMSE of 362.



As expected, the model underperforms for the December 2nd observation – the highest quantity sold in the history of the SKU.

b. Predicted Orders for the Time Period

Based on the predictions for the time period of 2011/11/27-12/03 using the Random Forest model, the following order quantity is suggested for each item:

Item	Predicted	Actual
4588 : 'Rabbit Night Light'	3906	4588
3195: 'Popcorn Holder'	3001	3195
1851: 'Vintage Doily Jumbo Bag Red'	602	1851

c. Predicting Orders by Country

To come up with predicted orders by country, one way is to regress quantities over countries and dates and use the model's coefficients as multiplying factors. Another way is to use a naïve approach by defining a function that picks the last month of sales data and determines the factors. Let's review the application of the latter for each SKU:

- 4588: 'Rabbit Night Light' Prediction by Country

Country	Historical	Actual	Ratio	Predictions
United Kingdom	8784	1985	0.51	1992.0
Netherlands	2616	0	0.15	586.0
France	2326	383	0.14	547.0
Australia	1632	0	0.10	391.0
Japan	1080	2040	0.06	234.0
Germany	192	72	0.01	39.0
Belgium	108	0	0.01	39.0
Finland	96	48	0.01	39.0
Sweden	84	0	0.00	0.0
EIRE	48	0	0.00	0.0
Iceland	48	0	0.00	0.0
Italy	48	0	0.00	0.0
Denmark	24	12	0.00	0.0
Portugal	18	48	0.00	0.0
Unspecified	12	0	0.00	0.0

While we predict well for the UK, the approach underestimates some other countries. There is a bias considering the inconsistency of ordering quantities by the country. For instance, the table below shows all orders from Japan in the time series for this item. The order of 2040 on 11/29 is the largest order this country has.

InvoiceDate	InvoiceNo	StockCode	Description	Quantity	UnitPrice	CustomerID	Country
2011-06-22	557670	23084	RABBIT NIGHT LIGHT	288	1.79	12798.0	Japan
2011-10-26	572869	23084	RABBIT NIGHT LIGHT	960	1.79	12798.0	Japan
2011-11-17	576923	23084	RABBIT NIGHT LIGHT	120	1.79	12753.0	Japan
2011-11-29	579498	23084	RABBIT NIGHT LIGHT	2040	1.79	12798.0	Japan
2011-12-06	C580832	23084	RABBIT NIGHT LIGHT	-7	1.79	12753.0	Japan

- 3195: 'Popcorn Holder' Prediction by Country

Country	Historical	Actual	Ratio	Predictions
United Kingdom	12783	3083	0.98	2941.0
Italy	100	0	0.01	30.0
EIRE	92	12	0.01	30.0
France	54	0	0.00	0.0
Belgium	36	0	0.00	0.0
Spain	36	0	0.00	0.0
Denmark	0	0	0.00	0.0
Greece	0	0	0.00	0.0
Germany	0	0	0.00	0.0
Finland	0	0	0.00	0.0
European Community	0	0	0.00	0.0
Unspecified	0	0	0.00	0.0
Hong Kong	0	0	0.00	0.0
Cyprus	0	0	0.00	0.0
Channel Islands	0	0	0.00	0.0

The prediction quantity is underestimated by 142 for the UK and overestimated for Italy and Ireland (EIRE).

- 1851: 'Vintage Doily Jumbo Bag Red' Prediction by Country

Country	Historical	Actual	Ratio	Predictions
United Kingdom	2125	1830	0.94	566.0
Portugal	40	0	0.02	12.0
France	35	10	0.02	12.0
Germany	20	10	0.01	6.0
Finland	20	0	0.01	6.0
Italy	10	0	0.00	0.0
Channel Islands	10	0	0.00	0.0
Netherlands	1	0	0.00	0.0
Denmark	0	0	0.00	0.0
European Community	0	0	0.00	0.0
EIRE	0	0	0.00	0.0
Unspecified	0	0	0.00	0.0
Hong Kong	0	0	0.00	0.0
Czech Republic	0	0	0.00	0.0
Cyprus	0	0	0.00	0.0

For this item, we are underestimating by 1264 units for the UK. This SKU has a short time span of data and therefore difficult to model. Oftentimes, more information is needed from subject matter experts (SMEs) to adjust the ordering quantity.

d. Results Review and Adjustments

So far, we predict the following total quantities for each SKU for the time period defined in the objective:

StockCode	Quantity	Description	Predicted_Order_ML
23084	4588	Rabbit_Night_Light	3906
22197	3195	Popcorn_Holder	3001
23582	1851	Vintage_Doily_Jumbo_Bag	602

There are some significant differences in the actual vs. predicted, especially for 1851: 'Vintage Doily Jumbo Bag Red' that may result in loss of sales. Sometimes, adding a safety stock to the order presents a workable solution for many retailers.

*Safety Stock (SS) = STD of delivery_lead_time * importance_factor*

Since we don't have that data, we can substitute with the following:

*Safety Stock (SS) = STD of daily sales quantity * number of predicted days * factor*,

where factor ranging from 0 to 1 could be either developed by an analyst using methods such as cluster analysis of SKUs or based on the SMEs feedback.

The caveat here is that if your data is not stationary, the standard deviation may not be used as a descriptive statistic for the data.

Therefore, our adjusted prediction could be the following:

Final Predicted Order = Predicted Order ML + SS

To come up with the safety stock parameter, we arbitrarily assign the following factors for each top-selling SKU and multiply it by the historical deviation of the data:

Item	Predicted	Actual	Factor
4588 : 'Rabbit Night Light'	3906	4588	0.3
3195: 'Popcorn Holder'	3001	3195	0.1
1851: 'Vintage Doily Jumbo Bag Red'	602	1851	0.9

The orders predicted using this method is much closer to the observed values.

StockCode	Quantity	Description	Safety_Stock	Predicted_Order_ML	Final_Order
23084	4588	Rabbit_Night_Light	645	3906	4551
22197	3195	Popcorn_Holder	224	3001	3225
23582	1851	Vintage_Doily_Jumbo_Bag	1201	602	1803

The table below shows predictions for top-selling SKUs summarized by country. It is evident that there is still an opportunity to optimize the model.

Country	Hist_rabbbit_light	Act_rabbit_light	R_rabbit_light	Preds_rabbit_light	Hist_popcorn_holder	Act_popcorn_holder	R_popcorn_holder	Preds_popcorn_holder	Hist_vintage_bag	Act_vintage_bag	R_vintage_bag	Preds_vintage_bag
United Kingdom	8784	1985	0.51	2321.0	12783	3083	0.98	3160.0	2125	1830	0.94	1695.0
Netherlands	2616	0	0.15	683.0	0	0	0.0	0.0	1	0	0.0	0.0
France	2326	383	0.14	637.0	54	0	0.0	0.0	35	10	0.02	36.0
Australia	1632	0	0.1	455.0	0	0	0.0	0.0	0	0	0.0	0.0
Japan	1080	2040	0.06	273.0	0	0	0.0	0.0	0	0	0.0	0.0
Germany	192	72	0.01	46.0	0	0	0.0	0.0	20	10	0.01	18.0
Belgium	108	0	0.01	46.0	36	0	0.0	0.0	0	0	0.0	0.0
Finland	96	48	0.01	46.0	0	0	0.0	0.0	20	0	0.01	18.0
Sweden	84	0	0.0	0.0	0	0	0.0	0.0	0	0	0.0	0.0
EIRE	48	0	0.0	0.0	92	12	0.01	32.0	0	0	0.0	0.0
Iceland	48	0	0.0	0.0	0	0	0.0	0.0	0	0	0.0	0.0
Italy	48	0	0.0	0.0	100	0	0.01	32.0	10	0	0.0	0.0
Denmark	24	12	0.0	0.0	0	0	0.0	0.0	0	0	0.0	0.0
Portugal	18	48	0.0	0.0	0	0	0.0	0.0	40	0	0.02	36.0
Unspecified	12	0	0.0	0.0	0	0	0.0	0.0	0	0	0.0	0.0
Norway	12	0	0.0	0.0	0	100	0.0	0.0	0	0	0.0	0.0
Switzerland	12	0	0.0	0.0	0	0	0.0	0.0	0	0	0.0	0.0
Spain	6	0	0.0	0.0	36	0	0.0	0.0	0	1	0.0	0.0

5. Challenges and Future work

Time series forecasting is a special area of machine learning. As evident through this work, predicting retail inventory orders can be a challenging task due to various factors including small quantities, low frequency, short-term retail dynamics, business arrangements, promotions, and others. Several models are fitted to the data and many underperform as a result of these dynamics. Granted, additional feature engineering, parameter optimization, and cross-validation should be included in future work. Furthermore, to determine better orders by country, regression and/or Bayesian approaches could be applied to define coefficients for each country. To ensure there are no lost sales due to understocking, it is plausible to develop a model utilizing a safety stock mechanism based on a cluster analysis of items or allowing for input from SMEs such as the retail team.