Forecasting Sales with Time Series Analysis

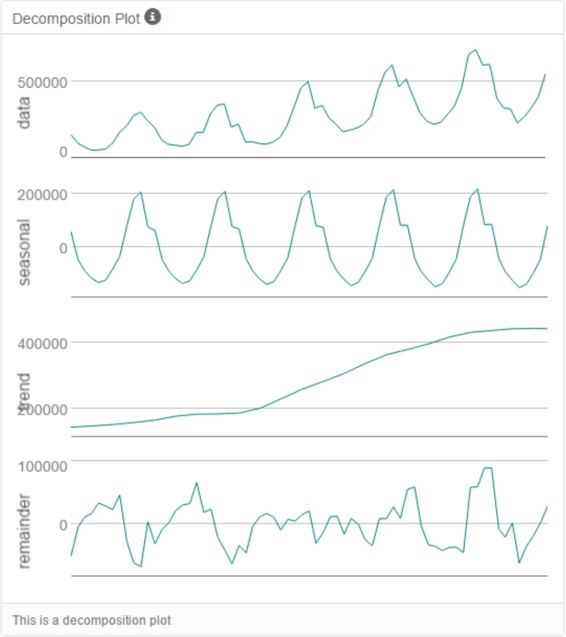
Problem Analysis

This business problem is to forecast the sales of a video game to assist the company with knowing how much to supply. The data will need to be forecasted for the next 4 months, which include October 2013, November 2013, December 2013, and January 2014.

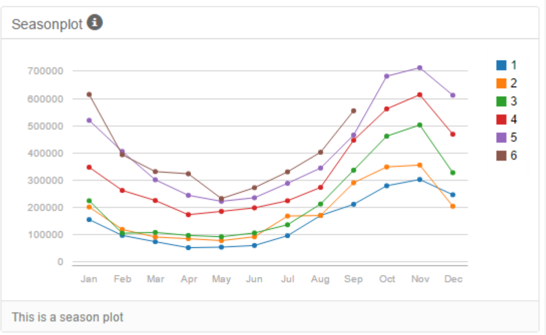
The data is continuous over a monthly interval. The data is sequential with no missing months, with no repeated months. The spacing between two consecutive data points is equally 1 month. Each month only has 1 data point, with no missing data.

The amount of holdout periods should be at least the same number of periods as we plan to forecast. In this business problem we plan on forecasting 4 periods. So therefore, there will be 4 holdout periods.

Trend, Seasonal, and Error



Trend – The trend in this plot shows an upwards trend as can be seen from the trend plot on the decomposition plot. The highest highs and the highest lows become larger over time and that indicates an uptrend time-series. This uptrend is a constant increase and therefore additive.

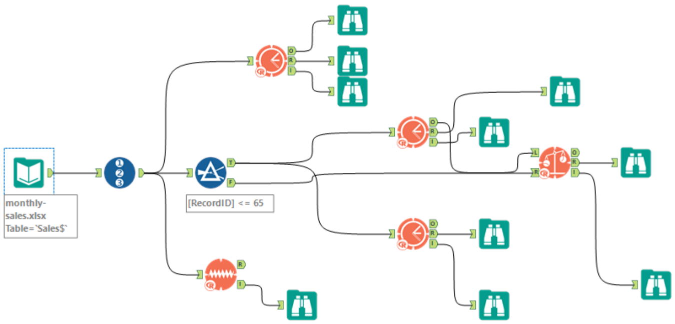


Seasonality – This time-series has a defined seasonality with the peaking occurring at its peak in November of each year and it valleys in May of each year. Over time the amount of variance for the seasonality increases, there this is multiplicative.

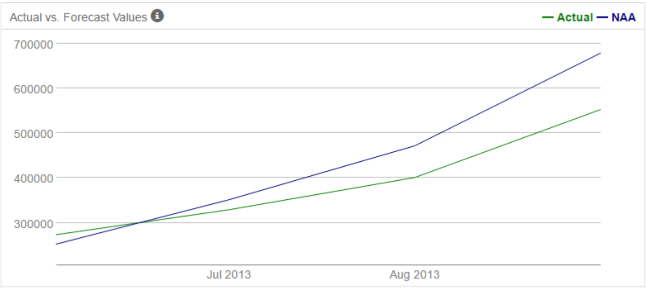
Error – The error or “remainder chart shows the variance over time shrinks then grows again. Based on this, the error is multiplicative.

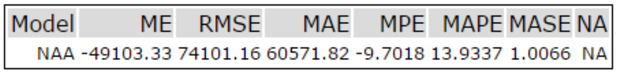
Models

ETS



Within the model a holdout group of 4 is taken. We set the target field to Monthly sales and ensure the frequency is set to 'Monthly.' Based on the analysis from above we set the error to be multiplicative, the trend to be additive, and the seasonality to multiplicative. After comparing the model to the actual data the following was discovered:



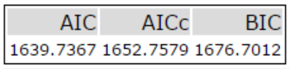


Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
NAA	-49103.33	74101.16	60571.82	-9.7018	13.9337	1.0066	NA

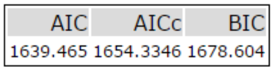
The RMSE (Root Mean Squared Error) of this model is 74101.16, which is the standard deviation from the mean in the model. The model with a lower number here indicates a more consistent and more accurate model for prediction.

The MASE (Mean Absolute Scaled Error) here is 1.0066. The MASE ideally should be less than 1 to be considered effective. In the case of this model although it is close to 1, we should see if the other model out-performs this one.

The Akaike Information Criterion is used to compare models, the AIC of this model is 1639.74. The lower this number the better the model is at predicting the time-series. When the model was run with dampening the AIC dropped to 1639.47 which is the number we will use to compare.

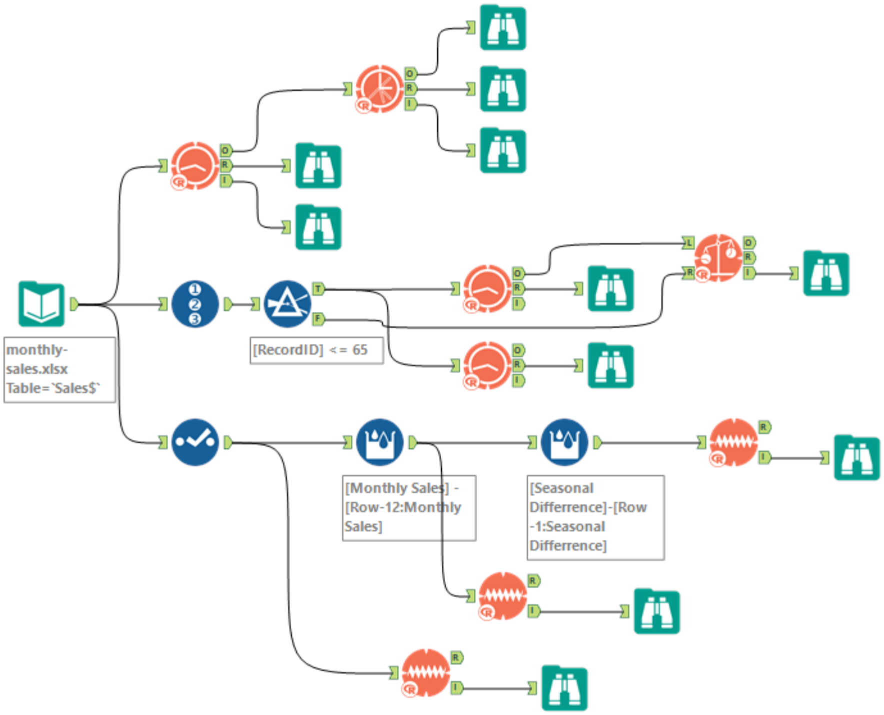


AIC	AICc	BIC
1639.7367	1652.7579	1676.7012

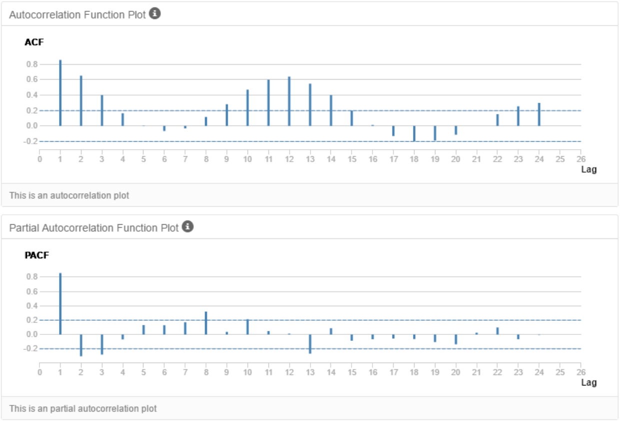


AIC	AICc	BIC
1639.465	1654.3346	1678.604

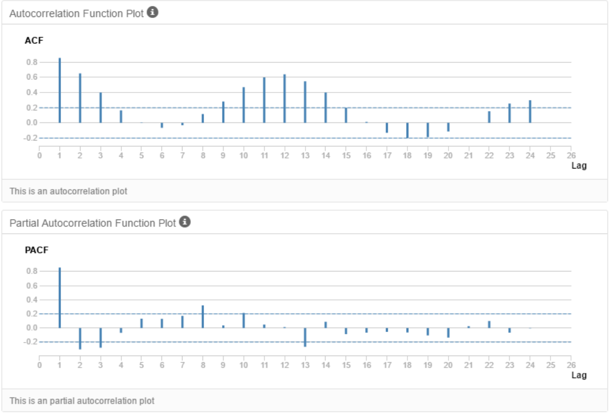
ARIMA


monthly-sales.xlsx
Table=`Sales$`
[RecordID] <= 65
T
F
O
R
I
L
[Monthly Sales] - [Row-12:Monthly Sales]
[Seasonal Differrence]-[Row -1:Seasonal Differrence]

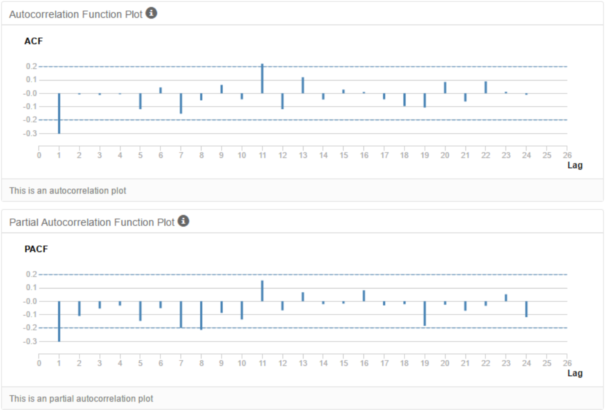
In the ARIMA model, it needs to be decided what values to use for the AM, I, and MA values. In order to do this, first we need to differentiate the values to check for correlation and partial correlation among the values in the monthly sales variable. The first image below shows the Auto-Correlation Function (ACF) and Partial Autocorrelation Function Plots (PACF) charts, these were used to decide on how to implement the ARIMA model.



Before differentiation



After differentiation



After seasonal differentiation

Integrated Component Term

Because the data was non-stationary, we need to use differing to make the time-series stationary. This was done once to make the trend stationary and once to make the seasonality stationary. Because of the need to do this, we set the integrated component terms to 1, and since there was differentiating done on the seasonality we will set the seasonal integrated component to 1 as well.

Moving Average Component

The reason to have MA components include negative correlation in the first lag in the ACF, where the PACF gradually decreases to 0. This model shows the negative correlation in the ACF and gradual decrease in the PACF. We will be setting the MA component to 1. As there is only auto correlation in lag one, no seasonality is seen and the MA seasonality will be set to 0.

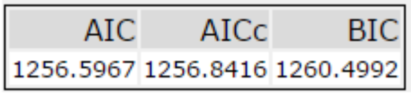
Autoregressive Component

The reason to have AR components include positive correlation in the first lag in the PACF, where the ACF gradually decreases to 0. Although the first lag in the PACF is negative, the ACF gradually decreases to 0. We will test if the model is better with or without AR components. As there is only auto correlation in lag one, no seasonality is seen and the AR seasonality will be set to 0.

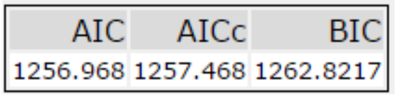
Based on the above analysis we will be testing the following models.

ARIMA(1,1,1)(0,1,0) and ARIMA(0,1,1)(0,1,0)

After doing a holdout on 4 values for testing the ARIMA(0,1,1)(0,1,0) was superior with an AIC of 1256.60, whereas ARIMA(1,1,1)(0,1,0) had an AIC of 1256.97. We will be using ARIMA(0,1,1)(0,1,0) for comparison due to this increased accuracy.

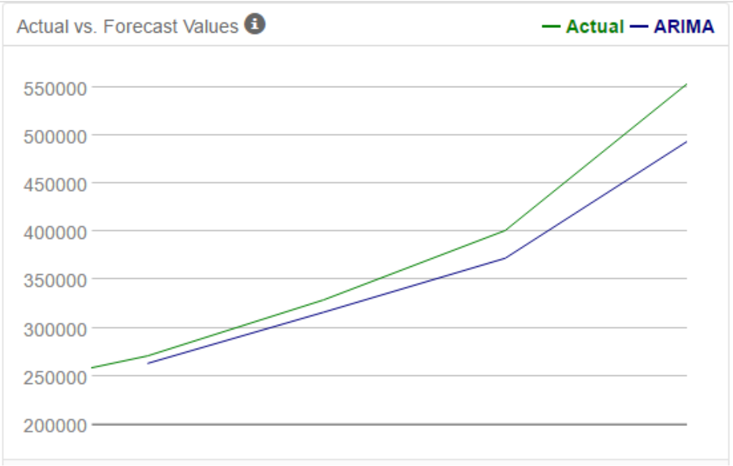


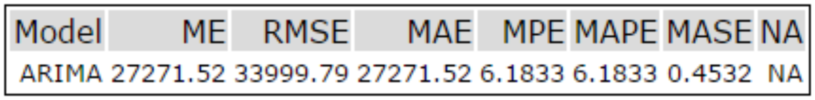
AIC	AICc	BIC
1256.5967	1256.8416	1260.4992



AIC	AICc	BIC
1256.968	1257.468	1262.8217

After checking for accuracy by comparing to the holdouts, these were the results:





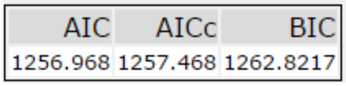
Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
ARIMA	27271.52	33999.79	27271.52	6.1833	6.1833	0.4532	NA

The RMSE (Root Mean Squared Error) of this model is 33999.79, which is the standard deviation from the mean in the model. The model with a lower number here indicates a more consistent and more accurate model for prediction.

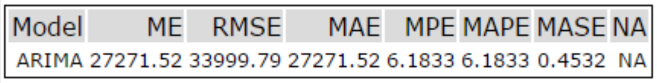
The MASE (Mean Absolute Scaled Error) here is 0.4532. The MASE ideally should be less than 1 to be considered effective. In the case of this model it is in a range to be considered effective.

The Akaike Information Criterion is used to compare models, the AIC of this model is 1256.60. The lower this number the better the model is at predicting the time-series.

Forecast

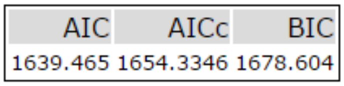


AIC	AICc	BIC
1256.968	1257.468	1262.8217

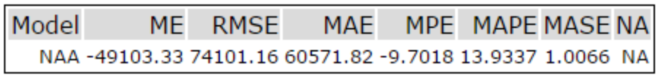


Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
ARIMA	27271.52	33999.79	27271.52	6.1833	6.1833	0.4532	NA

ARIMA



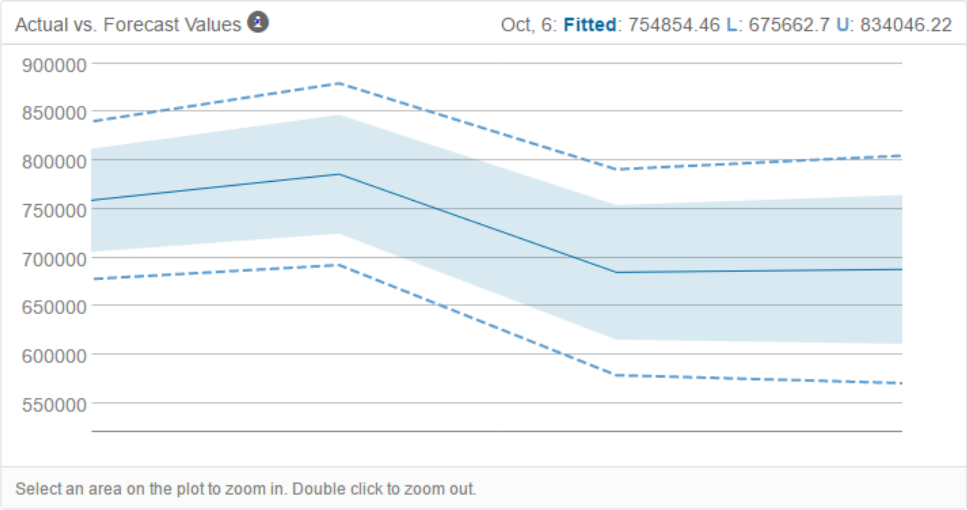
AIC	AICc	BIC
1639.465	1654.3346	1678.604



Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
NAA	-49103.33	74101.16	60571.82	-9.7018	13.9337	1.0066	NA

ETS

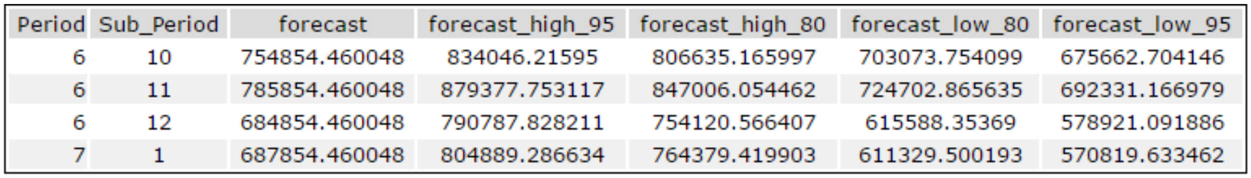
Based off the lower (indicating higher accuracy) AIC, RMSE and the MASE the ARIMA(0,1,1)(0,1,0) model was chosen to forecast the values for October 2013, November 2013, December 2013, and January 2014. The model was then run on the whole set of 69 cases (the 4 holdouts were returned) with the same parameters used for accuracy testing.



Actual vs. Forecast Values

Oct, 6: Fitted: 754854.46 L: 675662.7 U: 834046.22

Select an area on the plot to zoom in. Double click to zoom out.



Period	Sub_Period	forecast	forecast_high_95	forecast_high_80	forecast_low_80	forecast_low_95
6	10	754854.460048	834046.21595	806635.165997	703073.754099	675662.704146
6	11	785854.460048	879377.753117	847006.054462	724702.865635	692331.166979
6	12	684854.460048	790787.828211	754120.566407	615588.35369	578921.091886
7	1	687854.460048	804889.286634	764379.419903	611329.500193	570819.633462

Forecasted Values:
October 2013 – 754,854
November 2013 – 785,854
December 2013 – 684,854
January 2014 – 687,854