Forecasting Video Game Sales

Step 1: Plan Your Analysis

1. Does the dataset meet the criteria of a time series dataset? Make sure to explore all four key characteristics of a time series data.

Sequential and equal intervals with one data point only.
Continuous time interval.
Monthly data from January 2008 to September 2009

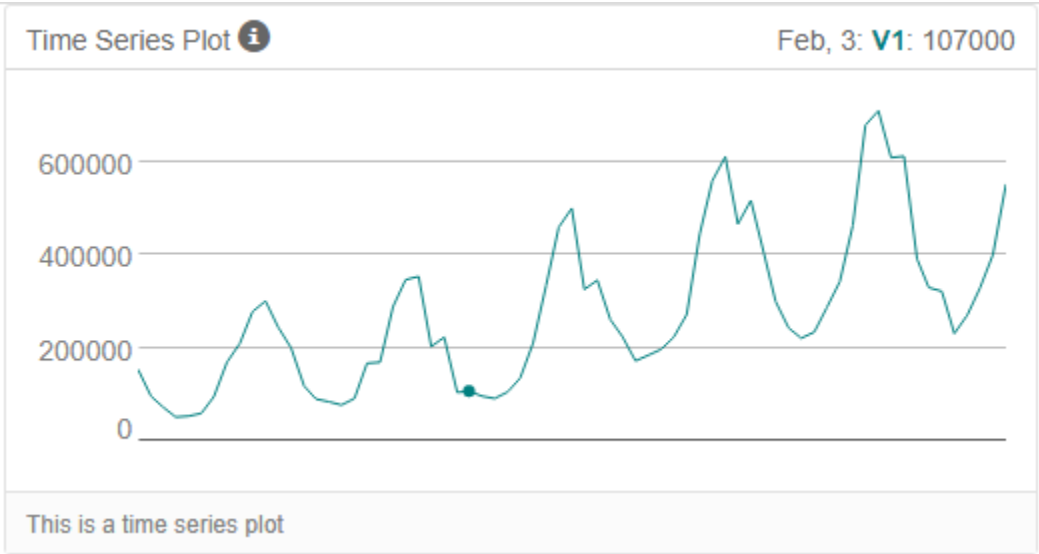
2. Which records should be used as the holdout sample?

A forecast for the next 4 months will be provided, therefore a holdout sample of 4 months will be used (June 2013 – September 2013).

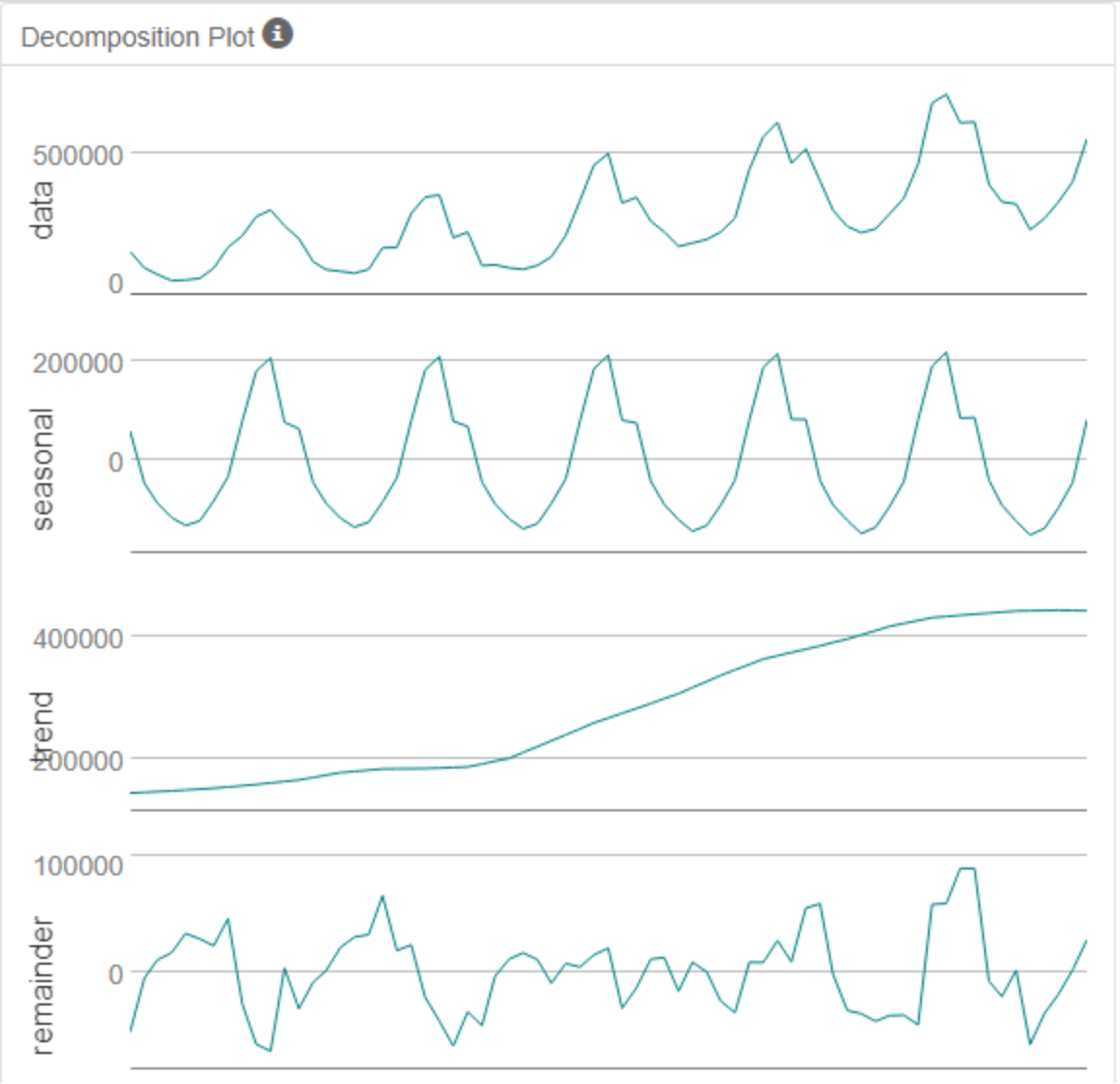
Step 2: Determine Trend, Seasonal, and Error components

1. What are the trend, seasonality, and error of the time series? Show how you were able to determine the components using time series plots. Include the graphs.

The Time Series Plot shows a rising upward trend and a spike in the later part of each year which shows seasonality.



The Error plot shows inconsistent changes over time (Multiplicative).
The Decomposition Plot below shows a rising Trend (Additive).
Seasonality is visible over time with a slight increase (Multiplicative).

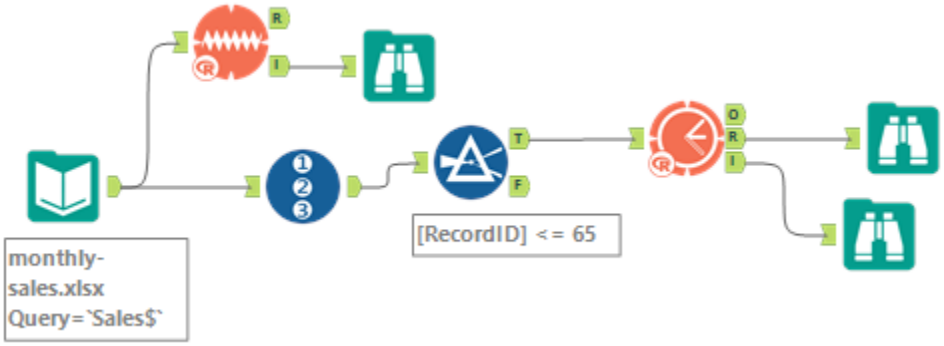


Step 3: Build your Models

1. What are the model terms for ETS? Explain why you chose those terms.
    a. Describe the in-sample errors. Use at least RMSE and MASE when examining results

The Decomposition Plot below shows a rising Trend therefore an additive method will be used. Seasonality shows a slight increase over time therefore a multiplicative method will be used. The Error plot shows inconsistent changes over time therefore a multiplicative method will be used.

The ETS (M,A,M) model will be used.



The Error Measures show

In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
3243.4703524	31474.3668886	24188.2167878	-0.572395	10.3052041	0.3528697	0.0087402

RMSE – Shows the standard deviation of the sample, 31474 in this case.

MASE – One of the best metrics for error measurement. Ideally this value should be significantly lower than 1, the table above shows a MASE value of 0.35.

2. What are the model terms for ARIMA? Explain why you chose those terms. Graph the Auto-Correlation Function (ACF) and Partial Autocorrelation Function Plots (PACF) for the time series and seasonal component and use these graphs to justify choosing your model terms.

    a. Describe the in-sample errors. Use at least RMSE and MASE when examining results

The ARIMA error measures show the model accuracy. The RMSE shows the in-sample standard deviation as 36761.
The MASE value is 0.36 which is well below the threshold for model accuracy.

In-sample error measures:

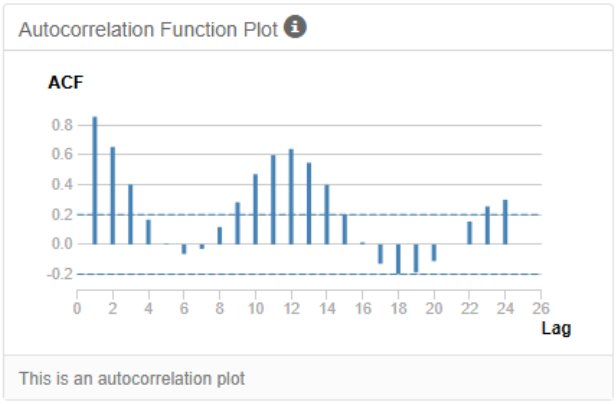
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-356.2665104	36761.5281724	24993.041976	-1.8021372	9.824411	0.3646109	0.0164145

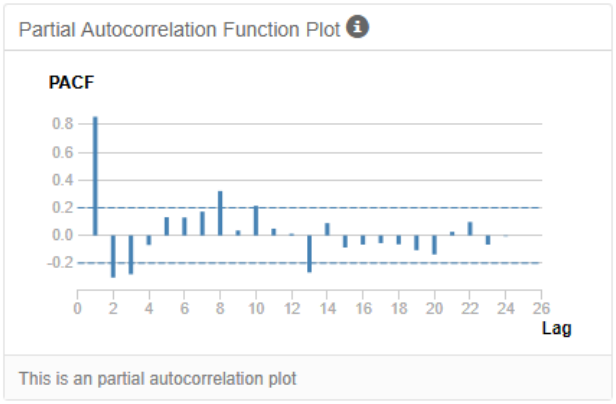
b. Regraph ACF and PACF for both the Time Series and Seasonal Difference and include these graphs in your answer.

The time series exhibits seasonality therefore the ARIMA (p,d,q)(P,D,Q)12 will be used for modelling.

Time Series ACF and PACF

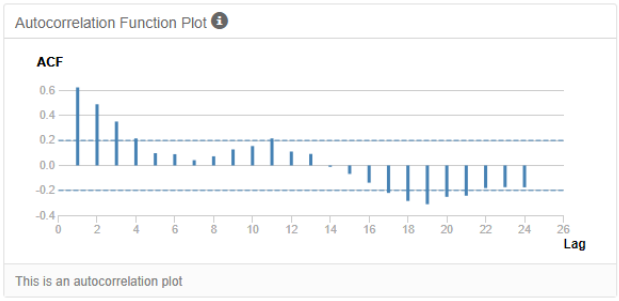
The ACF has a slow decay towards zero correlation with seasonal lags, values are more correlated to recent values than values in the past. This suggests the series is not stationary and will need to be differenced to reach a stationary series.

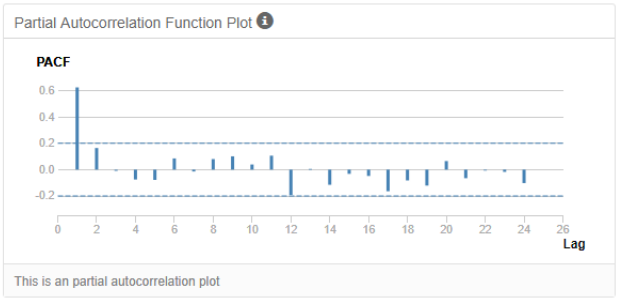




Seasonal Difference ACF and PACF

The ACF has a slow decay towards zero correlation with seasonal lags. The series is not stationary and will need to be differenced further.



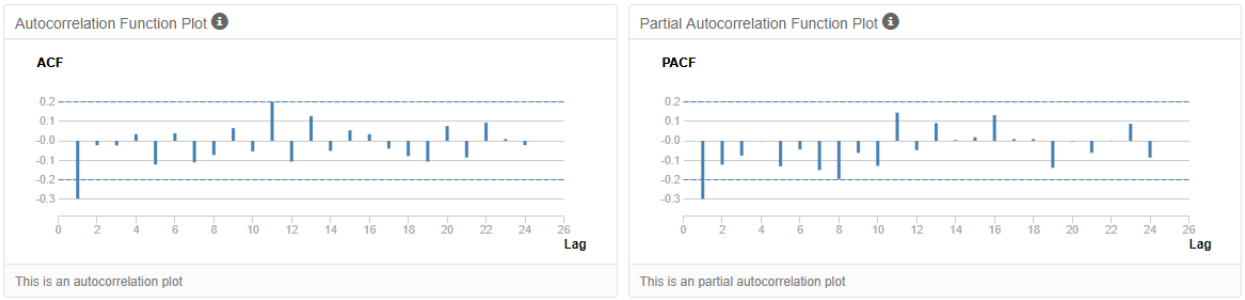


Seasonal First Difference ACF and PACF

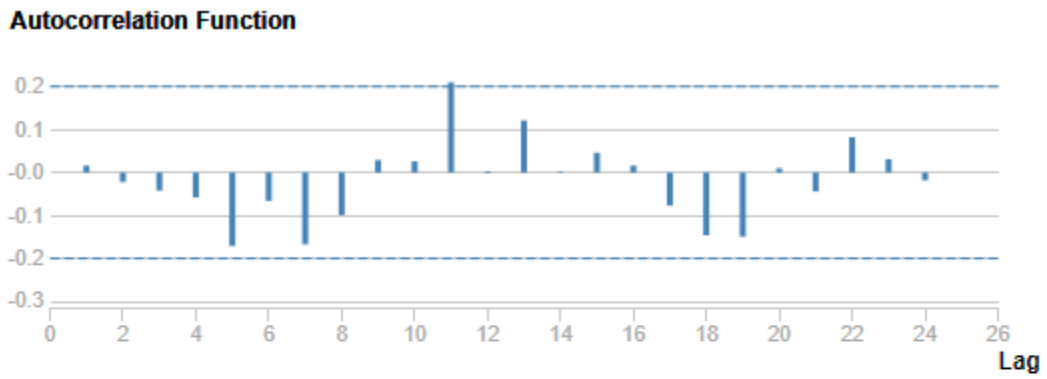
ACF drop after Lag 1 is significant and negatively correlated suggesting MA1 terms, the significance after lag one is much less suggesting a stationary series and no need for further differencing. Terms will be d=1 and D=1.

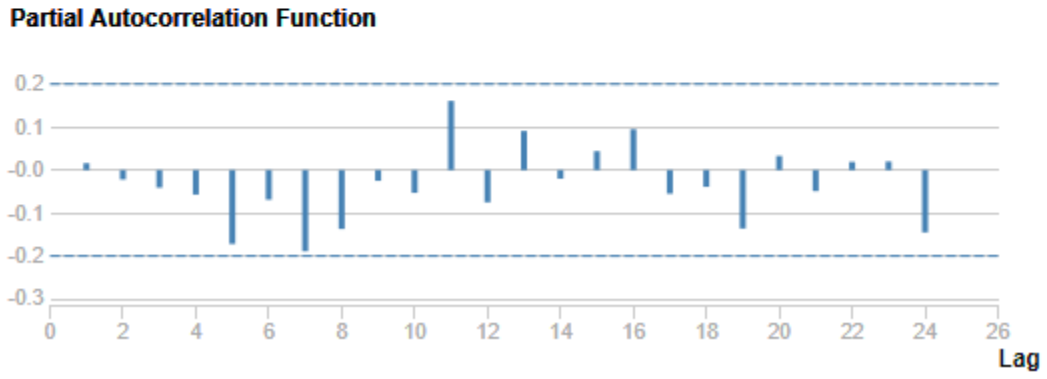
PACF decreases gradually towards zero this also suggests MA1 terms.

Seasonal autocorrelation is negative.



ARIMA Model (0,1,1)(0,1,0)12





Step 4: Forecast

1. Which model did you choose? Justify your answer by showing: in-sample error measurements and forecast error measurements against the holdout sample.

ETS(M,A,M)

In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
3243.4703524	31474.3668886	24188.2167878	-0.572395	10.3052041	0.3528697	0.0087402

ARIMA(0,1,1)(0,1,0)12

In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-356.2665104	36761.5281724	24993.041976	-1.8021372	9.824411	0.3646109	0.0164145

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE
ETS_MAM	-33469.61	53828.48	41542.75	-6.3476	9.3266	0.6904
ARIMA	27271.52	33999.79	27271.52	6.1833	6.1833	0.4532

The MAPE and MASE values are lower in both cases for the ARIMA model, therefore the ARIMA model is used for the forecast.

2. What is the forecast for the next four periods? Graph the results using 95% and 80% confidence intervals.

The forecast for the next four months sales are:

October 2013; 754854.46

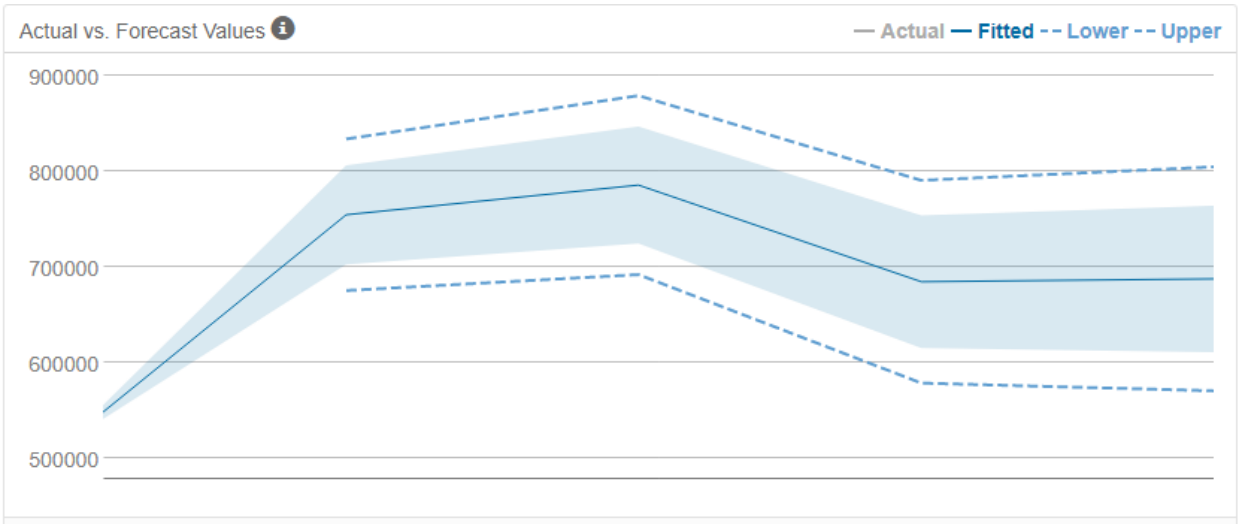
November 2013; 785854.46

December 2013; 684854.46

January 2014; 687854.46

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

Graph showing the results using 95% and 80% confidence intervals:



Forecast Video Game Sales Workflow