Project 6: Forecast Video Game Demand

The Business Problem

You recently started working for a company as a supply chain analyst that creates and sells video games. Many businesses have to be on point when it comes to ordering supplies to meet the demand of its customers. An overestimation of demand leads to bloated inventory and high costs. Underestimating demand means many valued customers won't get the products they want. Your manager has tasked you to forecast monthly sales data in order to synchronize supply with demand, aid in decision making that will help build a competitive infrastructure and measure company performance. You, the supply chain analyst, are assigned to help your manager run the numbers through a time series forecasting model.

You've been asked to provide a forecast for the next 4 months of sales and report your findings.

Step 1: Plan Your Analysis

1. Does the dataset meet the criteria of a time series dataset?

The dataset meets the criteria of a time series dataset. In fact,

- ✓ It's over a continuous time interval from January 2008 to September 2013
- ✓ The monthly sales are presented sequentially in order
- ✓ There is equal spacing between every two consecutive measurements on monthly basis
- ✓ Each month within the time interval has one data point

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

The holdout period should be at least the forecast period, therefore we use the last 4 observations of the time series as holdout sample.

Step 2: Determine Trend, Seasonal, and Error components

What are the trend, seasonality, and error of the time series?

We use the TS plot tool in Alteryx to analyse the trend, seasonality and error of the time series.

- ✓ Trend – The general direction of the chart is upward, therefore we have an uptrend.
- ✓ Seasonality - We see that the peaks occurred in Nov while troughs in May with an amplitude that grows over time, therefore there is a seasonality in video game demand.
- ✓ Error - Finally, the error plot is variable as it grows and shrinks over time.

Step 3: Build your Models

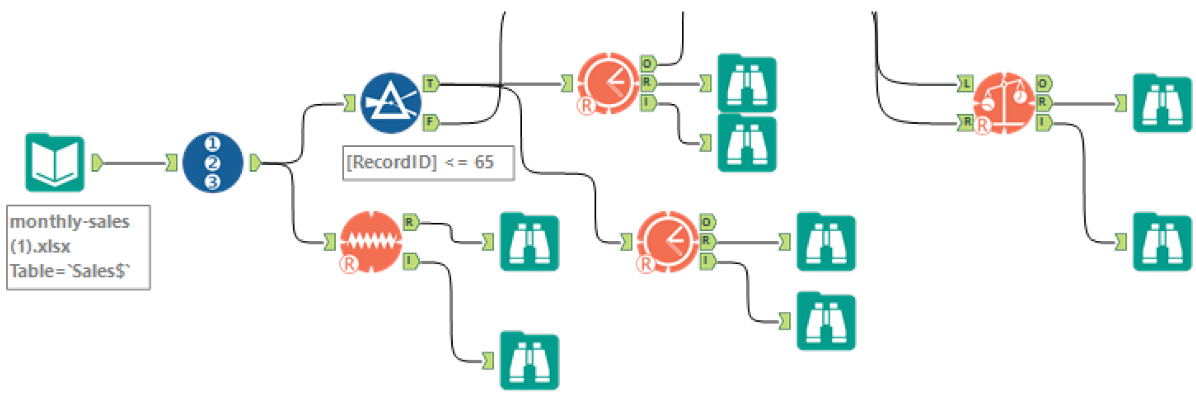
1. ETS Model

a. What are the model terms for ETS?

We use the following terms for our model ETS (M,A,M)

- ✓ The error is multiplicative as the errors are growing and shrinking over time .
- ✓ The trend is additive as it increases linearly.
- ✓ The seasonality is multiplicative as the peaks grow over time.

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



ETS (M,A,M) result without trend dampening

4 **Summary of Time Series Exponential Smoothing Model MAM**

5 Method:
ETS(M,A,M)

6 In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
3729.2947922	32883.8331471	24917.2814212	-0.9481496	10.2264109	0.3635056	0.1436491

7 Information criteria:

AIC	AICc	BIC
1634.6435	1645.9768	1669.4337

8 Smoothing parameters:

Parameter	Value
alpha	0.765251
beta	0.000103
gamma	0.001046

ETS (M,A,M) result with trend dampening

4 **Summary of Time Series Exponential Smoothing Model MAM_D**

5 Method:
ETS(A,Ad,A)

6 In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-874.8043454	36276.3502284	25774.3678735	0.4438749	12.2589823	0.3760092	0.0016254

7 Information criteria:

AIC	AICc	BIC
1670.1949	1683.2162	1707.1595

8 Smoothing parameters:

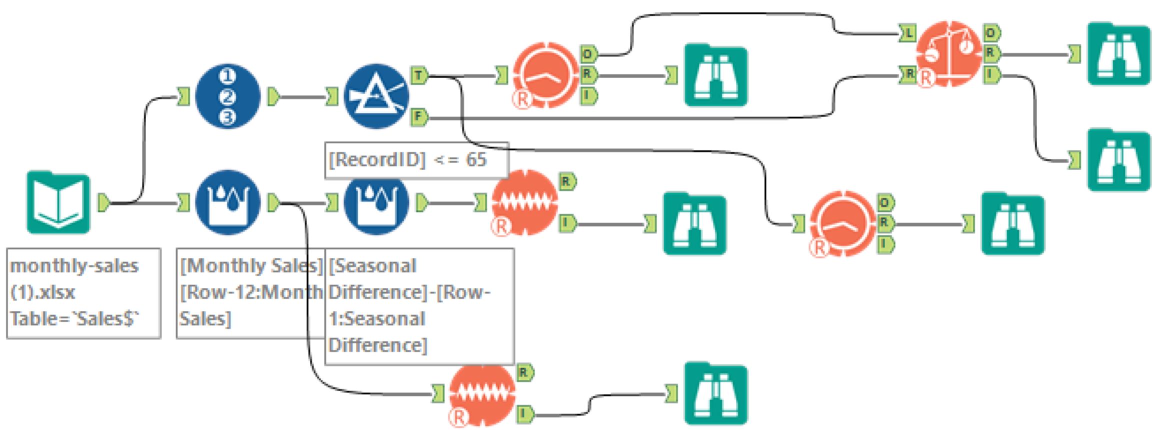
Parameter	Value
alpha	0.952012
beta	1e-04
gamma	0.000101
phi	0.979945

We run the ETS (M,A,M) without and with trend dampening.

- First, we noted that the AIC of the model without dampening is lower than the one with dampening, therefore the model without dampening provides a better fit and is less complex than the model with dampening.
- Looking at the in-sample errors, we noted that the ME of the dampened model is lower than the ME of the not-dampened model. But the dampened model has a higher RSME and MAE value than the not-dampened model. Since RSME and MAE takes respectively the squared and absolute value of the errors, they are a better measure to compare how close the forecasted values are to the actual value. Again, the ETS model without dampening seems to provide a better fit.
- Finally, both dampened model and not-dampened model have a MASE lower than 1. Since the MASE of the ETS model without dampening is lower than the one with dampening, we conclude that ETS model without dampening is the better fit.

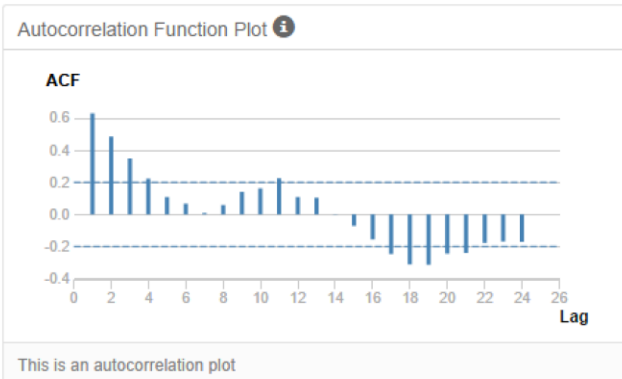
2. ARIMA Model

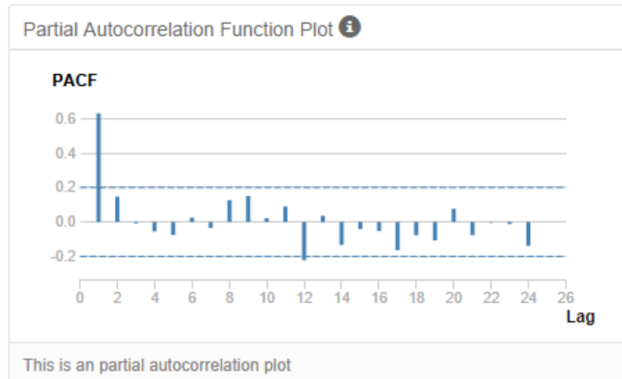
a. What are the model terms for ARIMA?



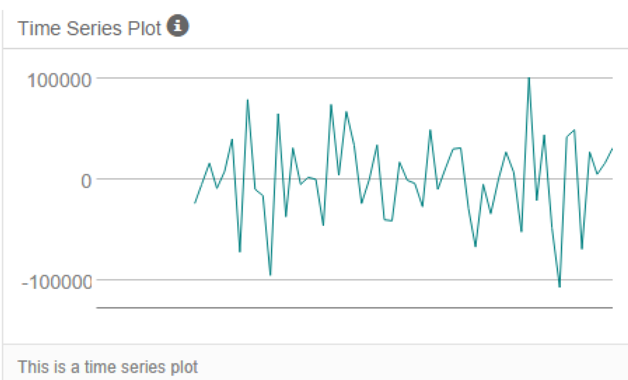
As we discussed in step 2, there is a seasonality in the dataset. But, after we calculated the seasonal differencing, we noted that the time series plot and ACF plot shows that the dataset is not stationary. Therefore, we will need to use the first seasonal difference to make the dataset stationary.

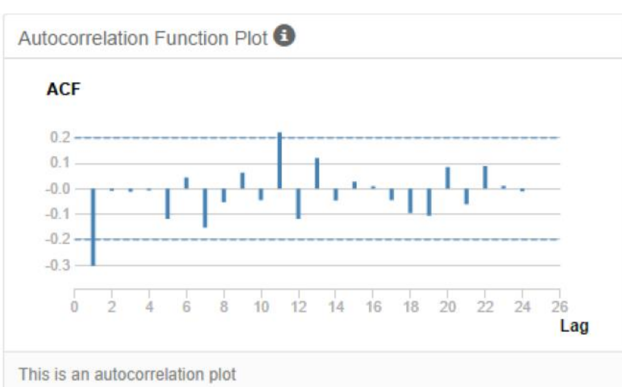
Time series plot, ACF and PACF of the seasonal difference dataset.

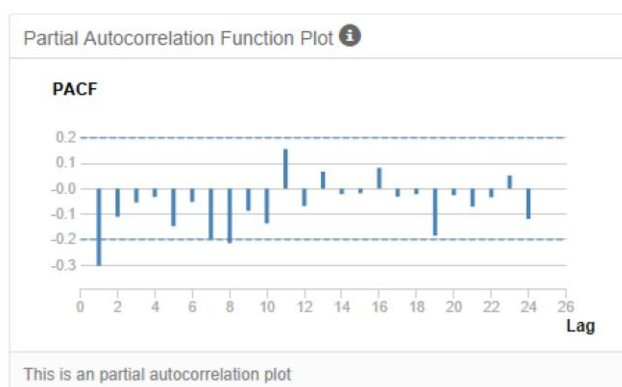




Time series plot, ACF and PACF of the first seasonal difference dataset







- We note a negative auto-correlation at lag 1 in the ACF and PACF plot, and the partial autocorrelation drops after lag 1 and gradually with no other significant autocorrelation, which suggest a MA model therefore we choose “p=0”, “q=1”, “P=0”, “Q=0”.
- We use the seasonal difference and first seasonal difference to make our dataset stationary, therefore we choose “d=1” and “D=1”. We could have continue the differencing, but we noticed that it did not make much difference to the ACF and PACF graphs.
- We choose “m=12” as it is the number of period between each period.

Finally, the ARIMA model is ARIMA(0,1,1)(0,1,0)12.

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

In addition to the chosen model above, we also run a model adding a seasonal moving average component - ARIMA(0,1,1)(0,1,1)12. Below are the in-sample errors results.

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

Method: ARIMA(0,1,1)(0,1,0)[12]

Call:
Arima(Monthly.Sales, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 0), period = 12))

Coefficients:

	ma1
Value	-0.378032
Std Err	0.146228

sigma^2 estimated as 1689257799.31927: log likelihood = -626.29834

Information Criteria:

AIC	AICc	BIC
1256.5967	1256.8416	1260.4992

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

Model 2 - ARIMA(0,1,1)(0,1,1)12 result

Method: ARIMA(0,1,1)(0,1,1)[12]

Call:
Arima(Monthly.Sales, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1), period = 12))

Coefficients:

	ma1	sma1
Value	-0.382258	0.011666
Std Err	0.162557	0.198094

sigma^2 estimated as 1688970569.06489: log likelihood = -626.29661

Information Criteria:

AIC	AICc	BIC
1258.5932	1259.0932	1264.447

In-sample error measures:

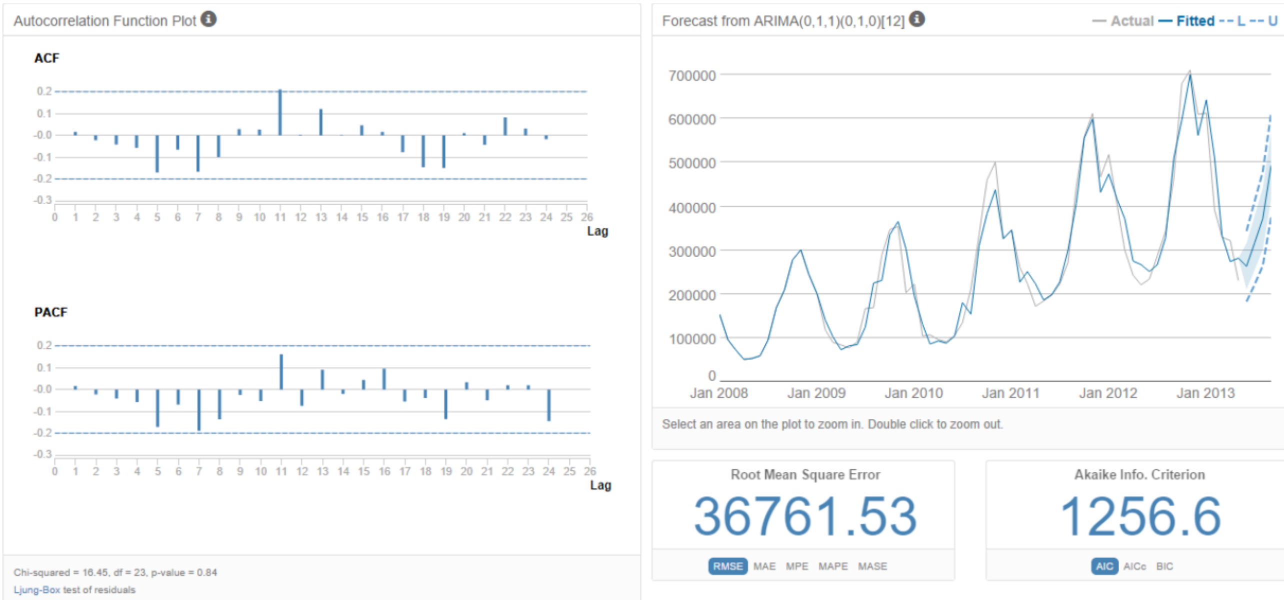
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-358.1274828	36758.4027043	24996.5435416	-1.800917	9.8272386	0.3646619	0.0166958

- First, we noted that the AIC of the ARIMA model 1 is lower than the model 2, which means that the model 1 provides a better fit and is less complex than the model 2.
- Looking at the in-sample errors, the model 1 has a lower RSME and MAE values than model 2.
- Finally, since the MASE of the model 1 is lower than the model 2, we conclude that ARIMA model 1 is the better fit.

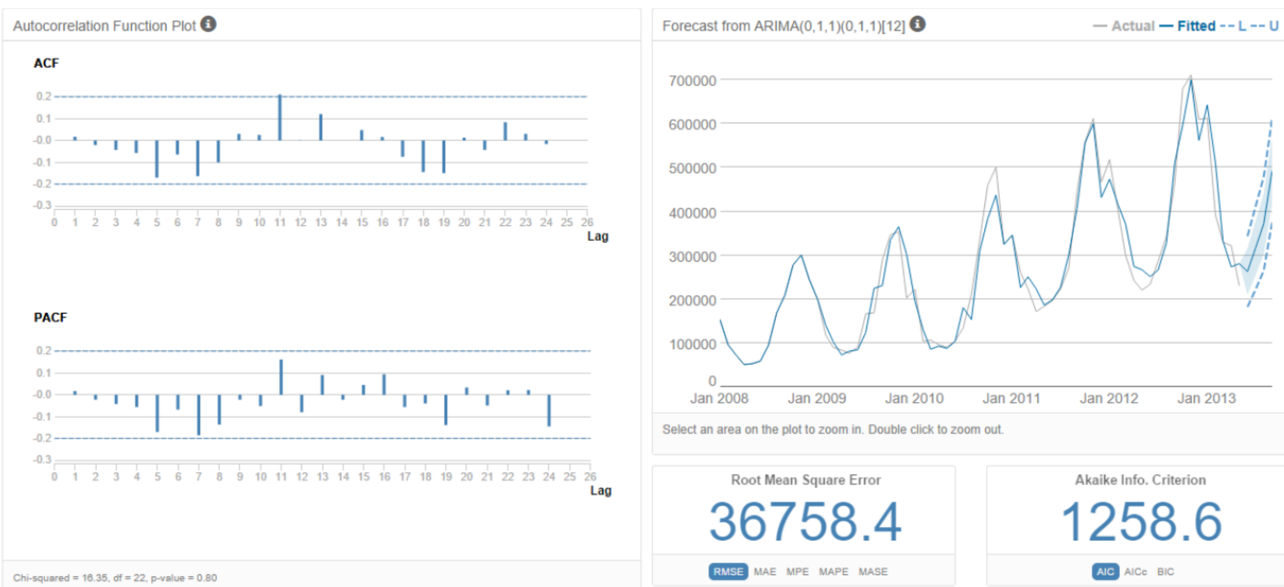
c. Regraph ACF and PACF for both the Time Series and Seasonal Difference and include these graphs in your answer.

Below are the ACF and PACF plots for both ARIMA models. We noted that for both there are no significant correlation for all the lags (the one in lag 11 is really minor), which means that we have use all the information in the time series for our models and no further adjustment is needed.

Model 1 - ARIMA(0,1,1)(0,1,0)12 – ACF and PACF plots



Model 2 - ARIMA(0,1,1)(0,1,1)12 – ACF and PACF plots



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.**

Below are the forecast error measurement of both chosen ETS and ARIMA models against the holdout sample.

Comparison of Time Series Models

Actual and Forecast Values:

Actual	ETS_MAM
271000	268729.50166
329000	378187.04023
401000	488199.64792
553000	691913.69155

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
ETS_MAM	-68257.47	85623.18	69392.72	-15.2446	15.6635	1.1532	NA

Comparison of Time Series Models

Actual and Forecast Values:

Actual	ARIMA
271000	263228.48013
329000	316228.48013
401000	372228.48013
553000	493228.48013

Accuracy Measures:

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

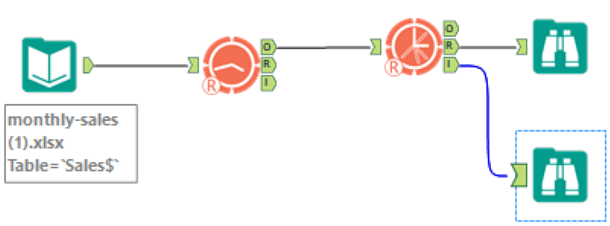
Based on the in-sample error measurement and the forecast error measurement results, we choose the ARIMA model because

- The AIC of ARIMA model is lower than the AIC of ETS model
- Though in the in-sample errors, ETS model's RMSE is lower than ARIMA model's RMSE and both MAE and MASE are around the same
- However, it appears clearly that on the forecast error measurement, ARIMA model fit better with a MASE lower than 1 while ETS model's MASE is higher than 1.

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

Finally, we make the forecast for the next four period from October 2013 to January 2014.

We use the whole dataset including previous holdout sample for our forecast with the model ARIMA(0,1,1)(0,1,0)12.



Below are the forecast results using 95% and 80% confidence intervals.

Record #	Period	Sub_Period	forecast	forecast_high_95	forecast_high_80	forecast_low_80	forecast_low_95
1	2013	10	754854.460048	833335.856133	806170.686679	703538.233418	676373.063963
2	2013	11	785854.460048	878538.837645	846457.517118	725251.402978	693170.082452
3	2013	12	684854.460048	789837.592834	753499.24089	616209.679206	579871.327263
4	2014	1	687854.460048	803839.469806	763692.981576	612015.938521	571869.450291