

Project 6: Forecasting Sales

Project Overview

You're a supply chain analyst that creates and sells video games. our manager has tasked you to forecast monthly sales data in order to help plan out the supply with demand for the company's video games.

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.

Yes, the dataset meets criteria of a time series dataset.

The dataset is taken across a continuous time interval and in yyyy-mm format.

The dataset is sequential.

The dataset is separated by monthly intervals.

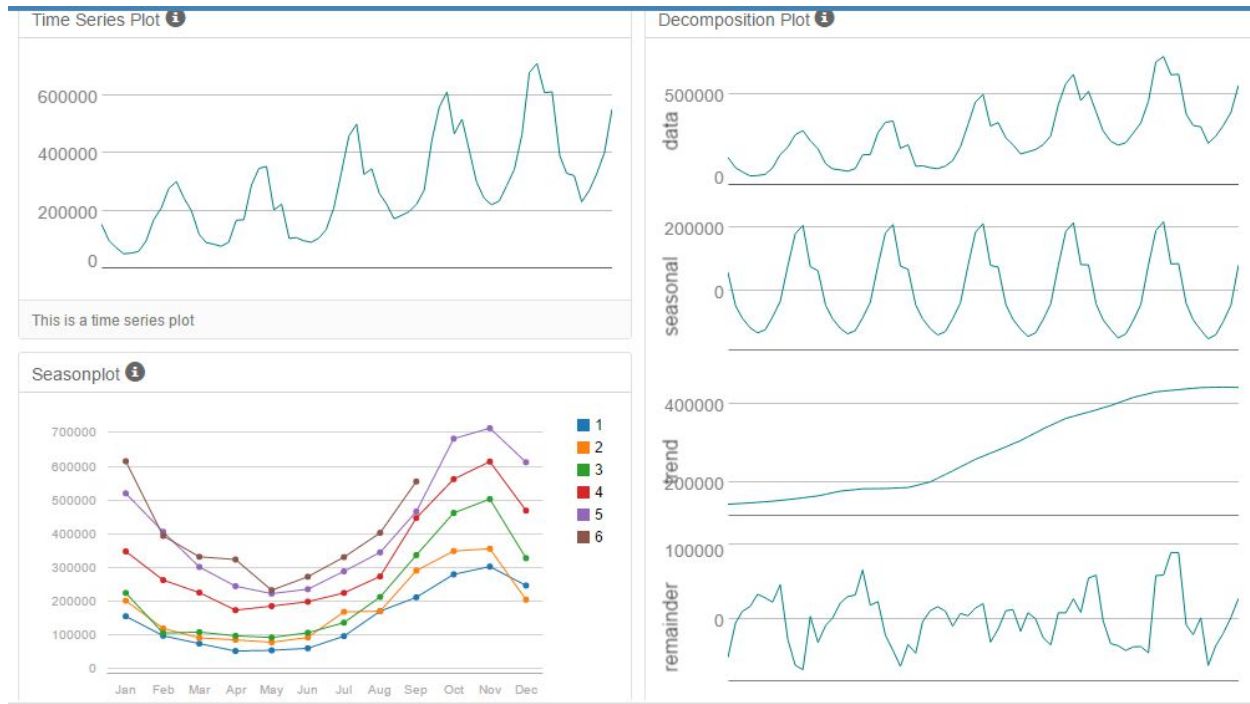
The dataset contains at most one data point for each interval.

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

The final prediction is forecast for the next 4 months of sales. The holdout sample will be the last 4 months of sales data from 2013-06 to 2013-09.

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 and decomposition plots are generated using TS plot tool.

Trend is upward and linear, Seasonality has peaks and valleys and increasing over time, Error has irregular pattern.

Step 3: Build your Models

1. What are the model terms for ETS? Explain why you chose those terms.

ETS Model Terms

Error - Multiplicative - Irregular pattern

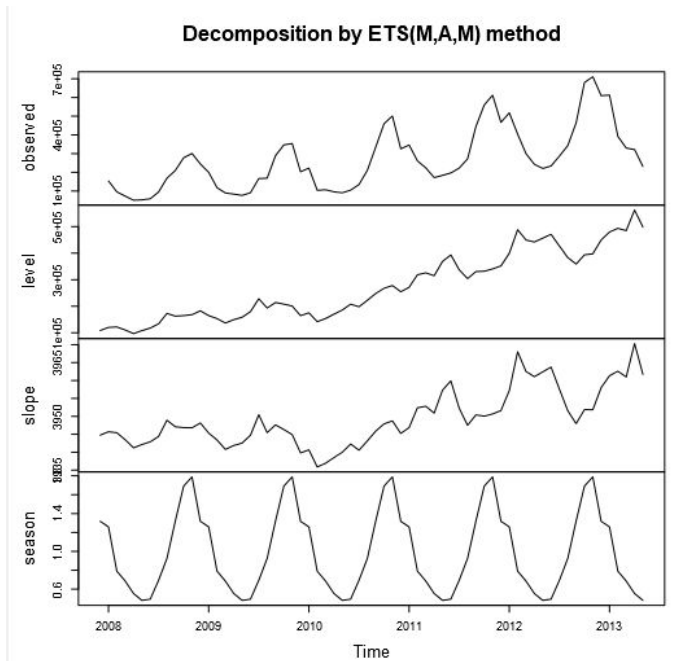
Trend - Additive - Upward and Linear

Seasonality - Multiplicative - Increasing peaks and valleys

ETS Model (M,A,M) is chosen.

ETS model is run with a holdout sample of 4 months.

ETS model



Summary of Time Series Exponential Smoothing Model MAM

Method:
ETS(M,A,M)

In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
2818.2731122	32992.7261011	25546.503798	-0.3778444	10.9094683	0.372685	0.0661496

Information criteria:

AIC	AICc	BIC
1639.7367	1652.7579	1676.7012

Two key components to look at are the RMSE, which shows the in-sample standard deviation, and the MASE which can be used to compare forecasts of different models.

ETS model has RMSE 32992.7 and MASE 0.37 and AIC 1639.7

MASE value is under the 1.0 threshold (commonly accepted threshold for model accuracy), which is good.

ETS model

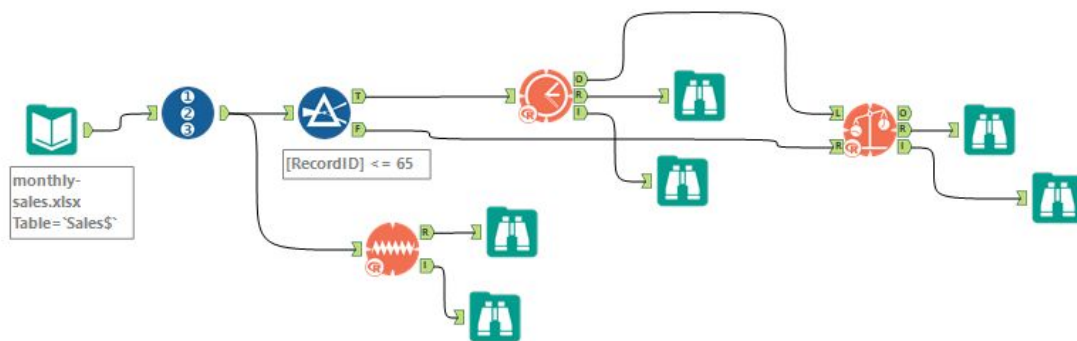
Actual and Forecast Values:

Actual	MAM
271000	248063.01908
329000	351306.93837
401000	471888.58168
553000	679154.7895

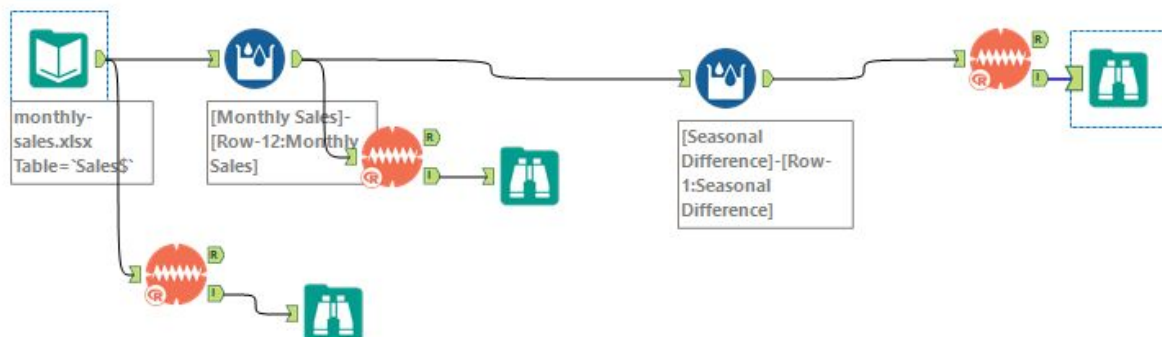
Accuracy Measures:

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

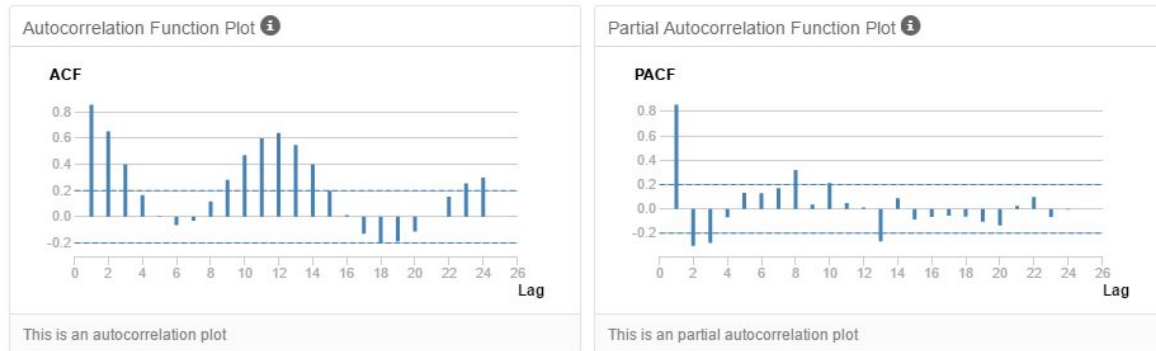
Alteryx Workflow ETS model



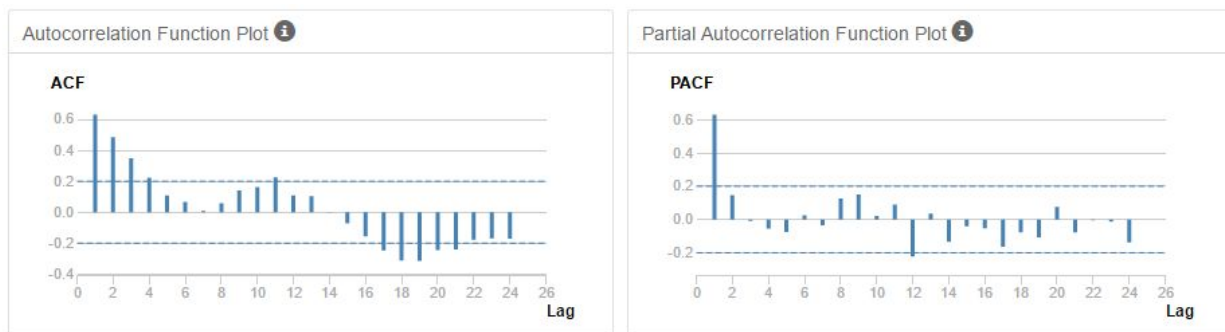
2. What are the model terms for ARIMA? Explain why you chose those terms. Graph the AutoCorrelation Function (ACF) and Partial Autocorrelation Function Plots (PACF) for the time series and seasonal component and use these graphs to justify choosing your model terms.



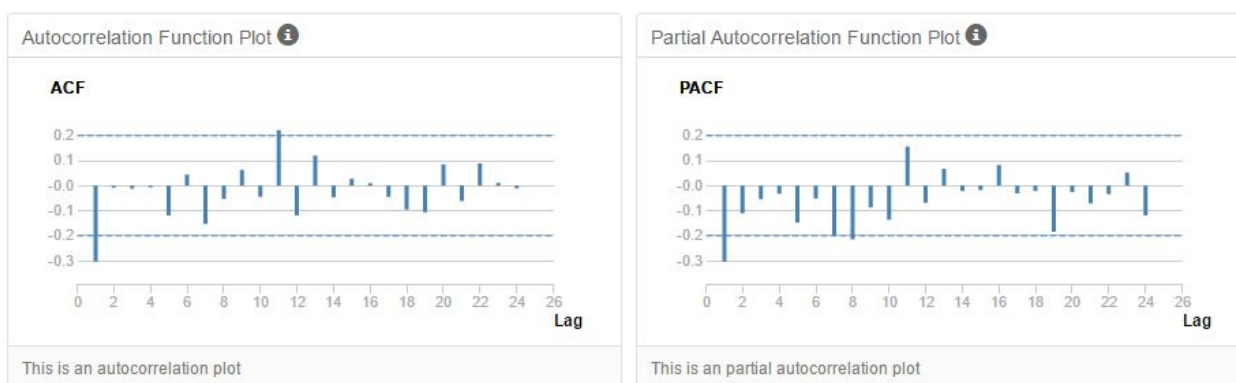
The original time series plot shows seasonality.



So differencing for the 12 months is done. This will remove seasonality.
ACF and PACF plot after seasonal difference is applied.



A seasonal first difference is applied and ACF plot does not show strong correlation.



Information Criteria:

AIC	AICc	BIC
1256.5967	1256.8416	1260.4992

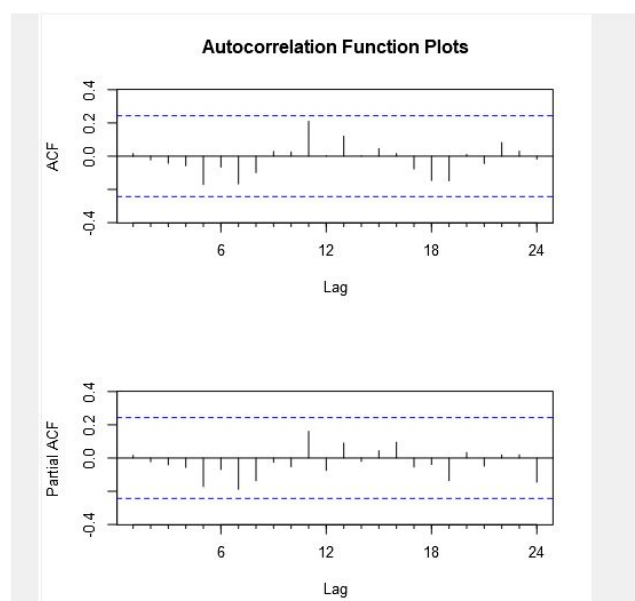
In-sample error measures:

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

Two key components to look at are the RMSE, which shows the in-sample standard deviation, and the MASE which can be used to compare forecasts of different models.

ARIMA model has RMSE 36761.5 and MASE 0.36 and AIC 1256.59

MASE is at 0.36 which means it under the commonly accepted threshold of 1.0



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

Step 4: Forecast

Answer these questions.

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

ETS model

Actual and Forecast Values:	
Actual	MAM
271000	248063.01908
329000	351306.93837
401000	471888.58168
553000	679154.7895

Accuracy Measures:							
Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
MAM	-49103.33	74101.16	60571.82	-9.7018	13.9337	1.0066	NA

ARIMA model

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

ARIMA model is better at forecasting sales using holdout sample.

The RMSE for ARIMA is **33999.79** compared to ETS' RMSE at **74101.16**. ARIMA's MASE value of **0.4532** is also lower than ETS' MASE value of **1.0066**. ARIMA model has lower AIC of **1256.59** and ETS has higher AIC of **1639.73**

ARIMA

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	

ETS

Summary of Time Series Exponential Smoothing Model MAM

Method:

ETS(M,A,M)

In-sample error measures:

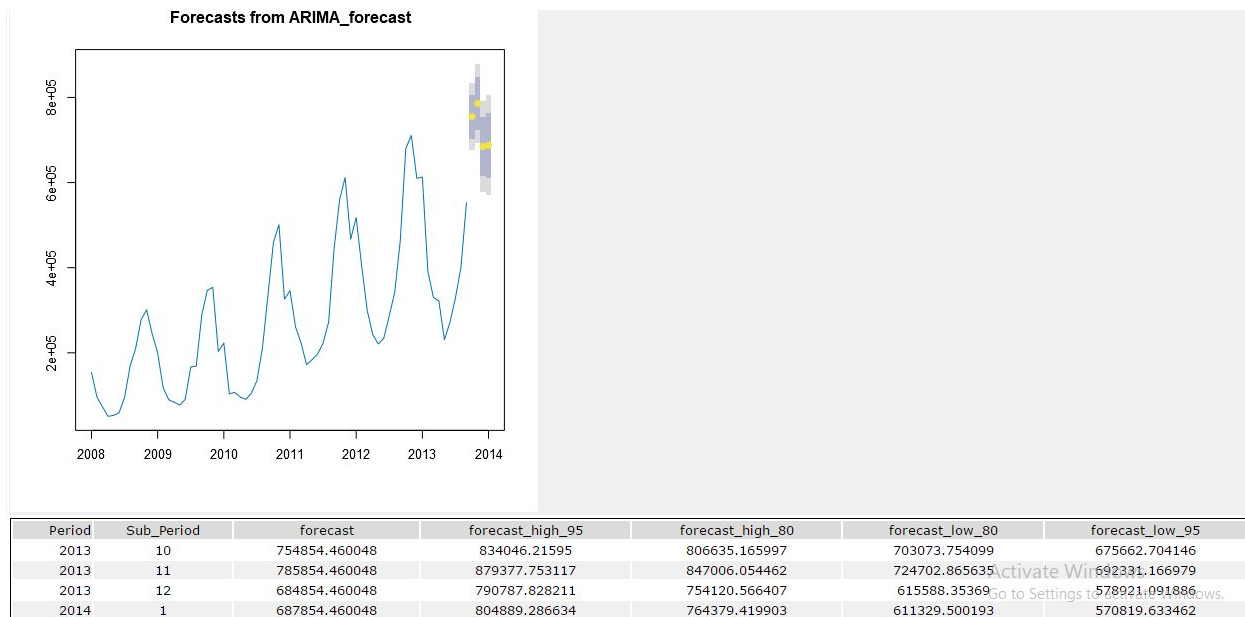
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
2818.2731122	32992.7261011	25546.503798	-0.3778444	10.9094683	0.372685	0.0661496

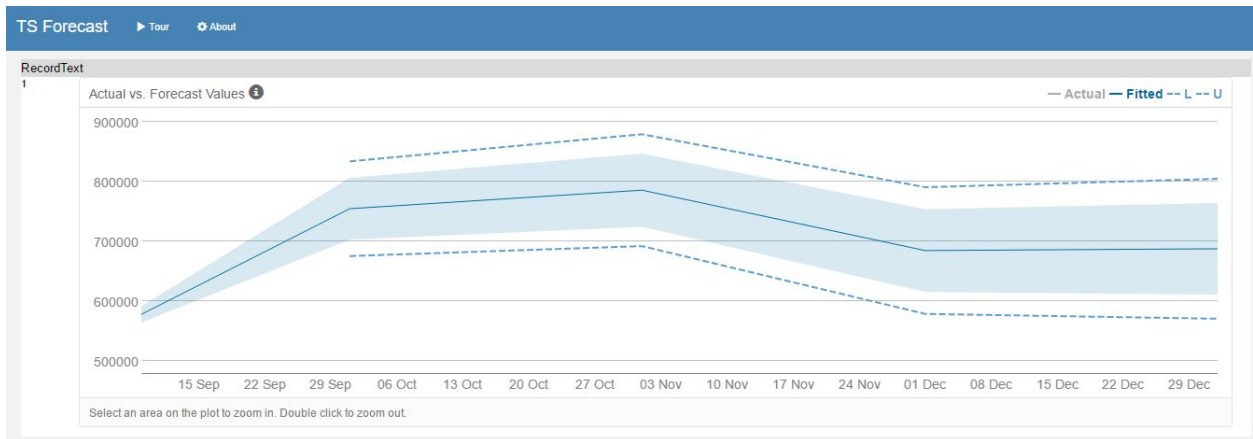
Information criteria:

AIC	AICc	BIC
1639.7367	1652.7579	1676.7012

Further investigation shows that the MAPE and ME of the ARIMA model are lower than the ETS. This suggests that, on average, the ARIMA model misses its forecast by a lesser amount. ARIMA model AIC is lower than that of ETS. Lower AIC reflects better model accuracy. Hence, ARIMA model is chosen.

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





Alteryx Workflow ARIMA model

