Project: Forecasting Video Game Demand

By: Anthony Nguyen

Project Overview

The client has requested for a monthly sales data forecast in order to help plan out the supply with demand for the company's video games.

Business Problem

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.

The client has requested a monthly sales data forecast in order to synchronize supply with demand, aid in decision making that will help build a competitive infrastructure and measure company performance. Assistance is needed to run the numbers through a time series forecasting model.

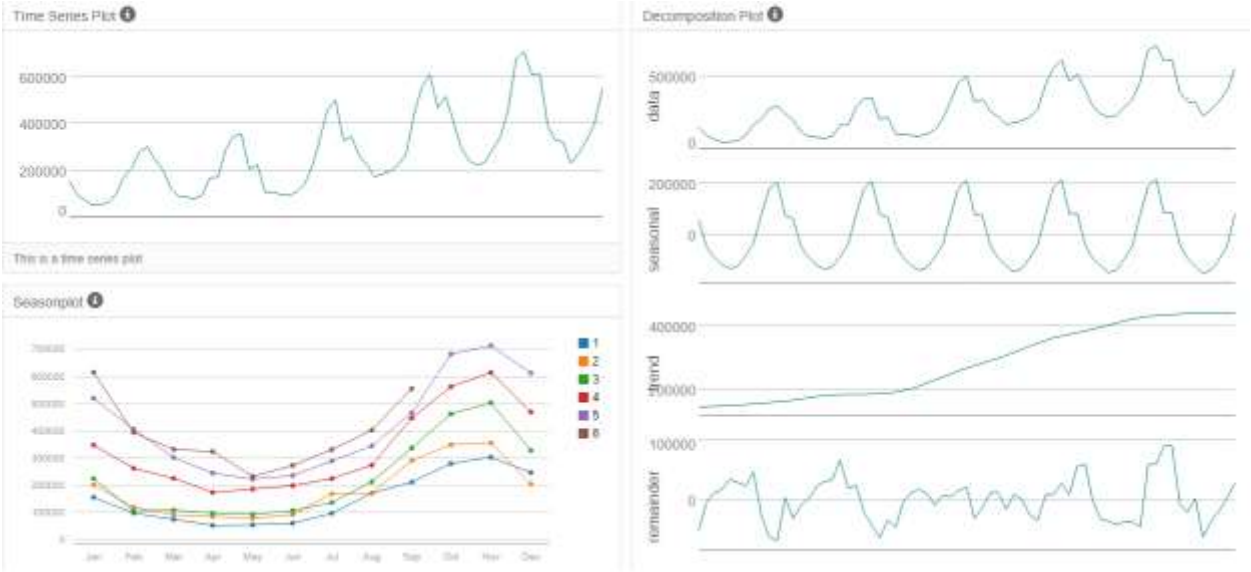
The forecasting model should predict the next four months of sales.

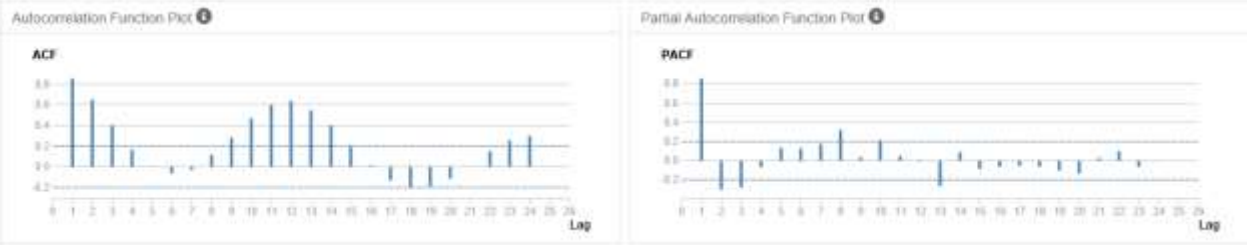
Step 1: Plan Your Analysis

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.
   - Yes the dataset meets the criteria of a time series data set by meeting all 4 key characteristics of a time series data by:
     - Cover a continuous time interval
     - Has sequential measurements across the time interval
     - Has equal spacing between every 2 consecutive measurements
     - Each time unit within the time interval has at most 1 data point
2. Which records should be used as the holdout sample?
   - The records to be used as the holdout sample are 4 of the most recent months:
     - 2013-06
     - 2013-07
     - 2013-08
     - 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 trend plot has a linear progression suggesting additive terms. The seasonal plot has peaks and valleys that increase with the level of time series suggesting multiplicative terms. The error plot fluctuates between large and small errors over time suggesting multiplicative terms.



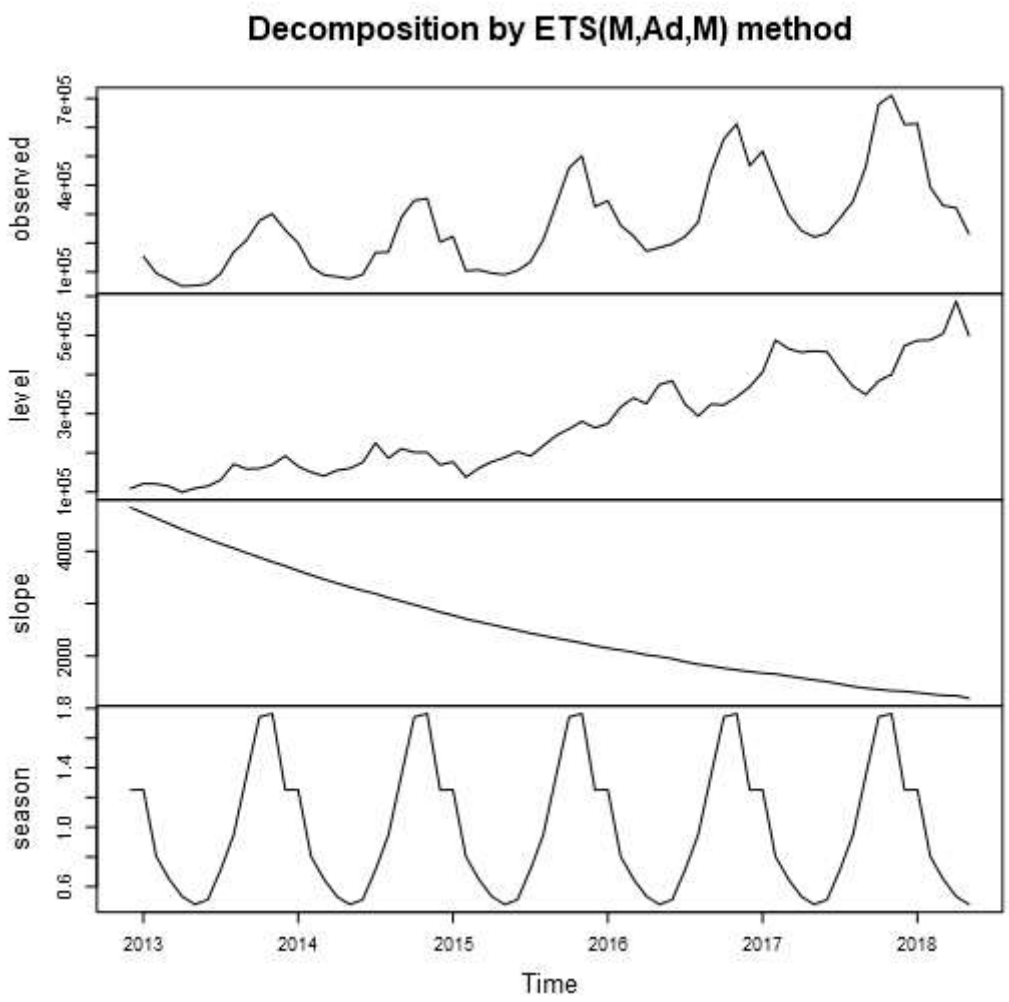


Step 3: Build your Models

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 TS Plot helped determine that the model term for the ETS Model will be multiplicative, additive, and multiplicative (MAM).
- ETS vs. ETS Dampen In-Sample Errors Results:
    - ETS:
        - AIC = 1634.64
        - RMSE = 32883.8331
        - MASE = 0.363
    - ETS Dampen:
        - AIC = 1636.53
        - RMSE = 33302.042
        - MASE = 0.375
- The RMSE determines the width of the confidence intervals for predictions and the MASE is a measure of the relative reduction in error compared to the naïve model.
- While the ETS Dampen model does have the highest in-sample value for both RMSE + MASE (33302.042/0.375) compared to the ETS model (32883.83/0.363), both models are relatively close in scale and have a difference of 2% between each other. The difference of 2% is not significant while comparing the two models against each other.
- ETS Dampen model was selected over the ETS model due to having a forecast value for each period which is closer to the holdout/validation sample.

ETS DAMPEN MODEL



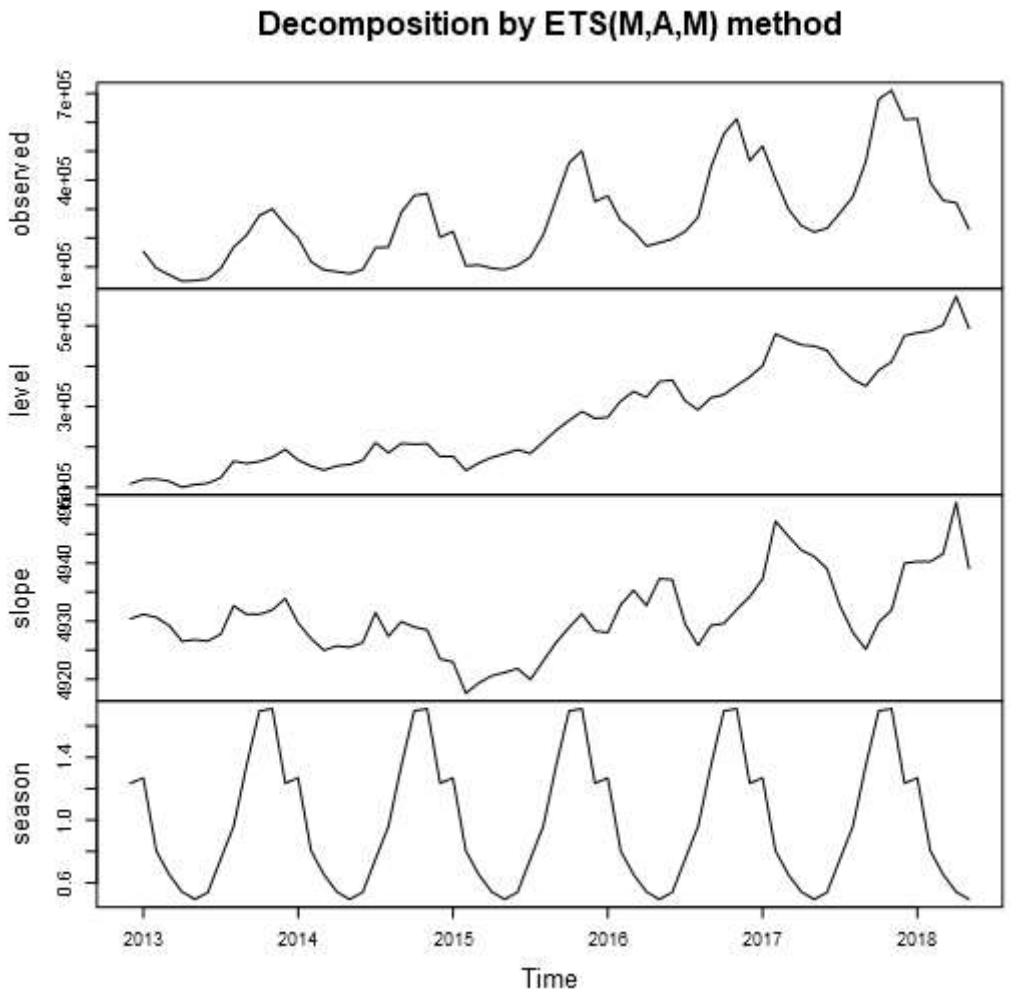
In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
5572.6821018	33302.042717	25725.4553044	0.1900065	10.54361	0.3752957	0.100576

Information criteria:

AIC	AICc	BIC
1636.5328	1649.554	1673.4973

ETS MODEL (without Trend Dampening)



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

Information criteria:

AIC	AICc	BIC
1634.6435	1645.9768	1669.4337

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 RMSE of 36073.53 and MASE value of 0.3632 are a result of the ARIMA (0,1,1)(0,1,0)12 model terms. The RMSE and MASE values were obtained from the in-sample error measures.

Information Criteria:

AIC	AICc	BIC
1350.0112	1350.2376	1354.0619

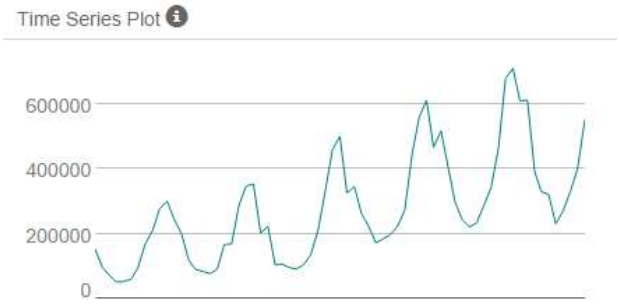
In-sample error measures:

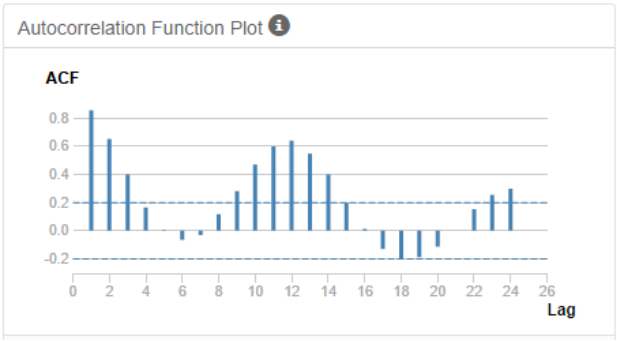
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
718.4459867	36073.5331869	24594.4458891	-1.4412026	9.5014904	0.3632763	0.016125

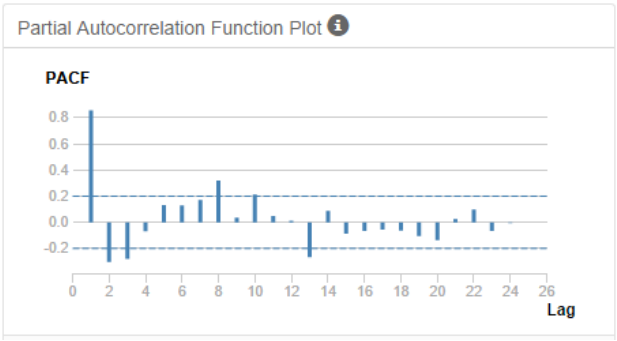
b. Regraph ACF and PACF for both the Time Series and Seasonal Difference and include these graphs in your answer.

TIME SERIES:

Without differencing, the initial time series plot plots confirmed that the series is not stationary.

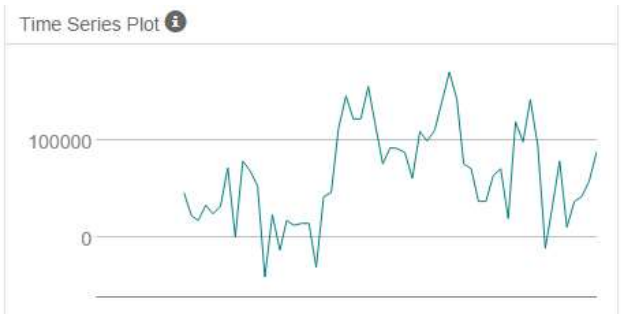


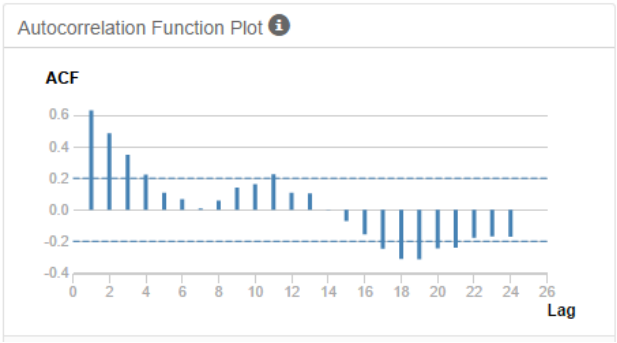


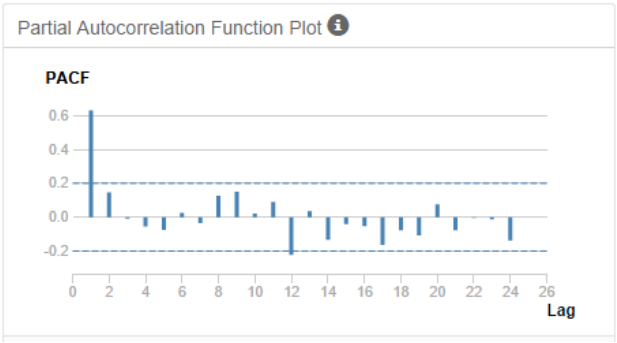


SEASONAL DIFFERENCE:

After running a seasonal difference, the time series plot still shows that the series is not stationary. The ACF plot shows that there are many strong correlations reflected. The PACF plot does not show a strong correlation after a seasonal difference is applied as it drops off quickly after lag 1.

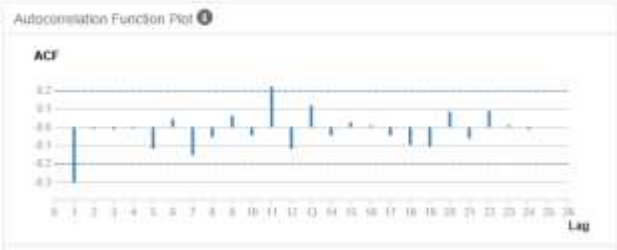


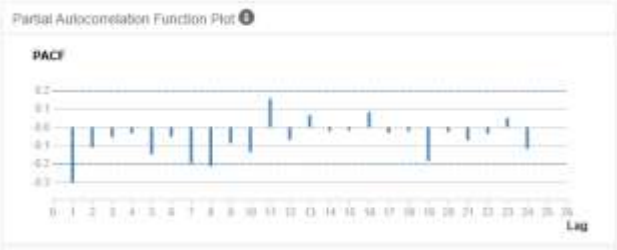




FIRST DIFFERENCE:

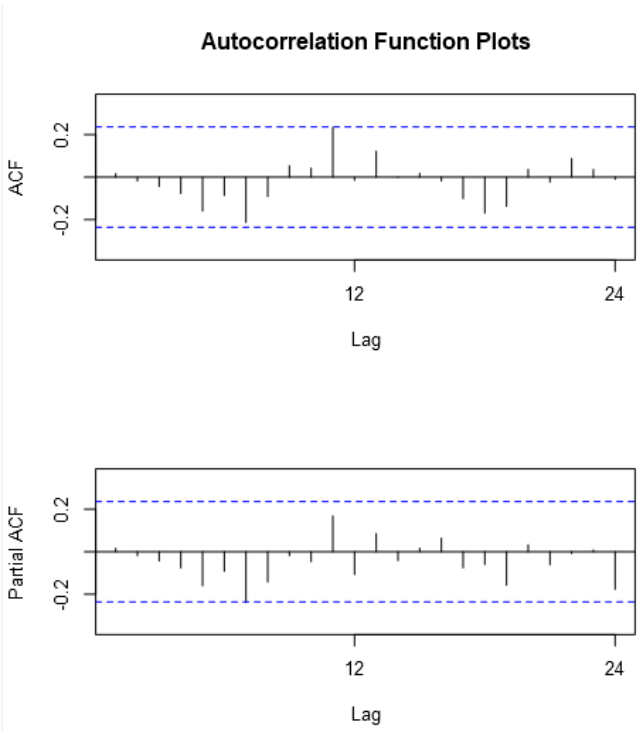
A seasonal first difference is applied and the time series plot reflects that the series is now stationary. The ACF and PACF show that there is no high amount of correlation as a result of the seasonal first difference.





ARIMA MODEL RESULT:

The ARIMA model will follow (0,1,1) for the non-seasonal component and (0,1,0) for the seasonal component with a period of 12. Both ACF & PACF plots for ARIMA model (0,1,1)(0,1,0)12 below. This analysis mirrors the automated model term function of the ARIMA tool. The ACF & PACF plots do not show significant correlation and no additional AR or MA terms are required.



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.
   - The ARIMA model was chosen over the ETS Dampen model because the forecasted values are closer to the holdout/validation sample as compared to the ETS Dampen model.
   - The both the RMSE and MASE value of the ETS Dampen model (RMSE 74839.3/MASE 1.0147) are much higher than the ARIMA model (RMSE 33999.79/MASE 0.4532).
   - The MASE value of the ARIMA model (0.4532) compared to the ETS Dampen model (1.0147) suggest that the ARIMA model is more fitted to the time series data. Ideally, the MASE value is significantly less than 1. The ETS Dampen model's MASE value is greater than 1 while the ARIMA model's MASE value is significantly less than 1.

ETS DAMPENING IN-SAMPLE ERROR MEASURES

In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
5572.6821018	33302.042717	25725.4553044	0.1900065	10.54361	0.3752957	0.100576

ARIMA IN-SAMPLE ERROR MEASURES

In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
718.4459867	36073.5331869	24594.4458891	-1.4412026	9.5014904	0.3632763	0.016125

Actual and Forecast Values:

Actual	ETS_MAM_Dampen
271000	255827.81186
329000	357675.13305
401000	475924.99669
553000	678449.06298

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
ETS_MAM_Dampen		-53469.25	74839.3	61055.35	-11.1217	13.921	1.0147	NA

Accuracy Measures:

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

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 periods are as follows:
     - 2013 – 10 = 754854.46
     - 2013 – 11 = 785854.46
     - 2013 – 12 = 684854.46
     - 2014 – 01 = 687854.46

Forecast results using 95% and 80% confidence intervals: