<u>Project: Forecast Video Game Demand</u>

Step 1: Plan Your Analysis

Look at your data set and determine whether the data is appropriate to use time series models. Determine which records should be held for validation later on (250 word limit).

Answer the following questions to help you plan out 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 all the criteria of a time series dataset,
   The following are the characteristics of the time series
   - *The series is over a continuous time interval: - The current time series is continuous with the data from 2008-01 to 2013-09.*
   - *Sequential measurements across that interval: - The monthly sales are the sequential measurement across the entire interval.*
   - *There is equal spacing between every two consecutive measurements: - Every measurement has one-month time interval.*
   - *Each time unit within the time interval has at most one data point: - The monthly sales data is present for every month.*

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

   A minimum of 4 months should be held as holdout sample since we need to forecast for the 4-time periods
   The following 4 months' data needs to be used as a holdout sample
   - *2013-06*
   - *2013-07*
   - *2013-08*
   - *2013-09*

Step 2: Determine Trend, Seasonal, and Error components

Graph the data set and decompose the time series into its three main components: trend, seasonality, and error. *(250 word limit)*

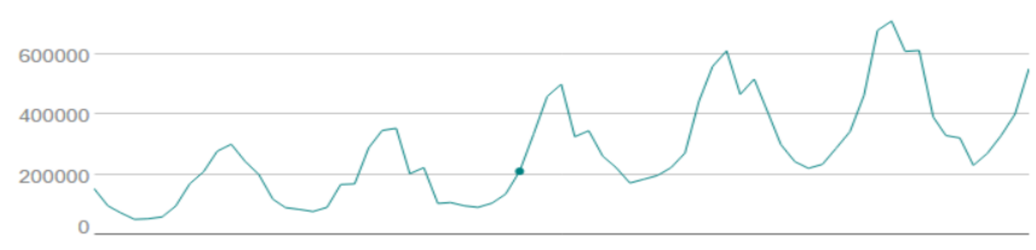
Answer this question:

a) **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 peaks and valleys for seasonality has spikes and gradually increases over the time series .*
- *The error pattern for the time series has more variance over time*
- *The trend pattern of the current time series is linear*

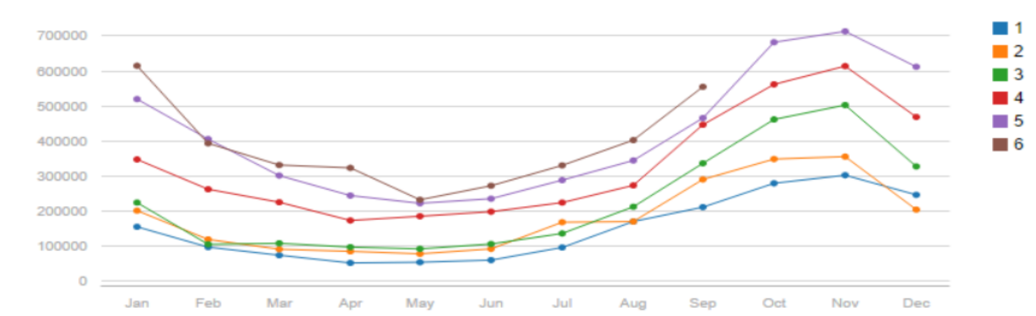
Time Series Plot

Aug, 3: **V1**: 211000



This is a time series plot

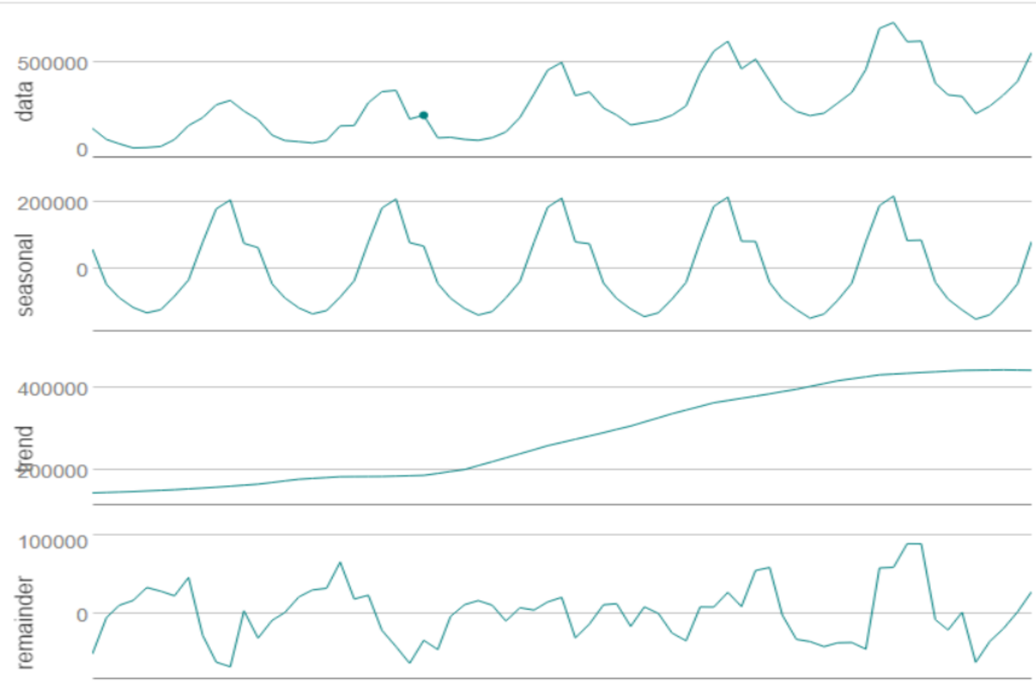
Seasonplot



This is a season plot

Decomposition Plot

Jan, 3: **data**: 223000



This is a decomposition plot

Step 3: Build your Models

Analyze your graphs and determine the appropriate measurements to apply to your ARIMA and ETS models and describe the errors for both models. (500 word limit)

Answer these questions:

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**

- *Even though the seasonal pattern looks constant over time, when we zoom over the peaks and valleys they are not same, hence we use it multiplicatively.*
- *The error pattern is fluctuating over time, so we use it multiplicatively*
- *The trend pattern is linear hence we use it additively.*

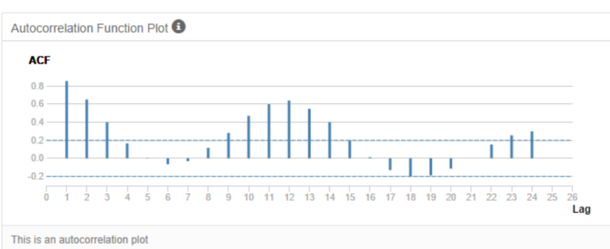
In-sample error measures:

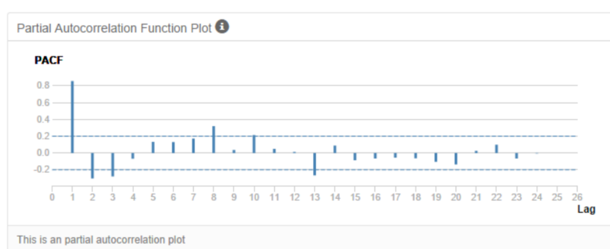
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
3729.2947922	32883.8331471	24917.2814212	-0.9481496	10.2264109	0.3635056	0.1436491

The MPE of the ETS model is -0.941 and it means that the forecasts are negatively biased
The RMSE which represents the standard deviation of the differences of predicted values and observed values is 32883.833
The MASE value is 0.3635 which is less than 1 and is acceptable

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.**

The ACF and PACF plots when we do not consider any lag is given below, we can observe that the data are highly corelated and cannot be used as such,





We know that the time series given to us is seasonal and, we have to remove the seasonality to have a better understanding in the ARIMA model,

We will first do a seasonal differencing to remove the seasonality and then check the ACF and PACF plots, after the seasonal differencing we can still see that the we still have correlation in the lags and all the lags are significant,

Since we have a significant correlation, we have to make the series more stationary

When we do the seasonal first differencing, all the significant lags are smoothened and this can be seen in the ACF and PACF plots, so there is no need for further differencing

We have D (1) and d (1), since we were able to remove the seasonality at the first differencing

The ACF and PACF plots show that there is a strong negative correlation at lag 1, this suggests that the MA (1) for the non-seasonal component

When we look at the difference ACF-PACF plots the lags at 12 and 24 are not very significant, so there is no need for the seasonal AR or MA terms

ARIMA (p,d,q) (P,D,Q) m
Non-seasonal (p,d,q) is (0,1,1) and
Seasonal (P, D, Q) is (0,1,0)
m is 12

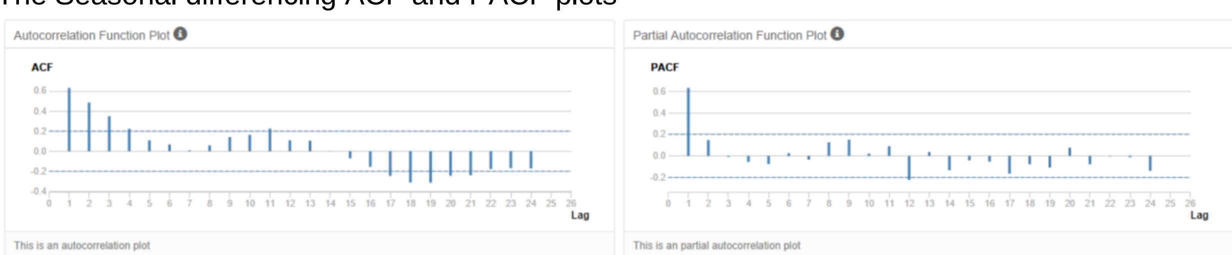
a) Describe the in-sample errors. Use at least RMSE and MASE when examining results

The MPE of the ARIMA model is -1.802 and it means that the forecasts are negatively biased
The RMSE which represents the standard deviation of the differences of predicted values and observed values is 6761.5281
The MASE value is 0.3646109 which is less than 1 and is acceptable

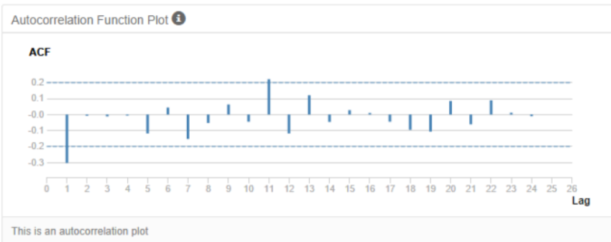
In-sample error measures:

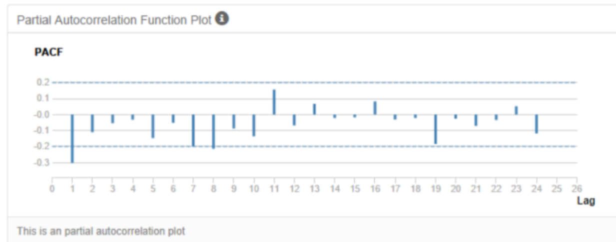
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-356.2665104	36761.5281724	24993.041976	-1.8021372	9.824411	0.3646109	0.0164145

The Seasonal differencing ACF and PACF plots



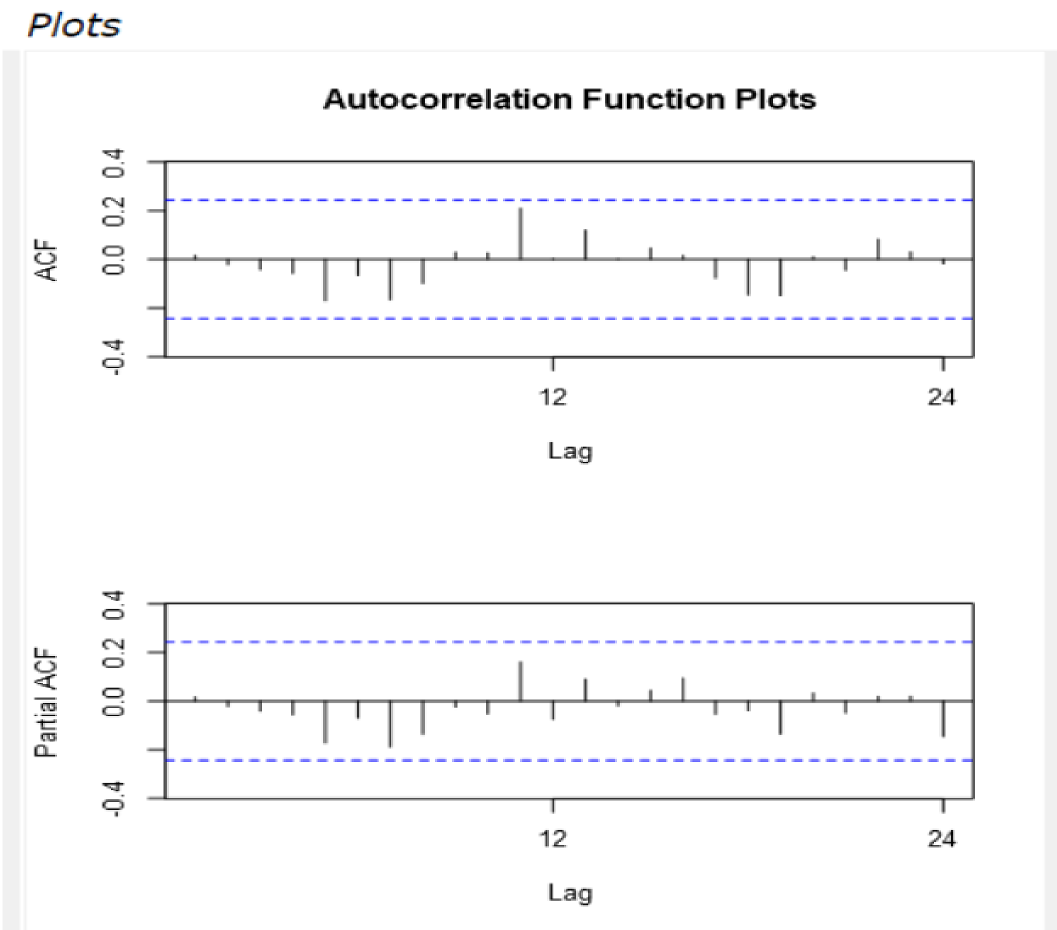
The Seasonal first differencing ACF and PACF plots





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

<u>The complete ACF and PACF plot of the timeseries when developed using the above ARIMA configuration</u>



Step 4: Forecast

Compare the in-sample error measurements to both models and compare error measurements for the holdout sample in your forecast. Choose the best fitting model and forecast the next four periods. (250 words limit)

ETS

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

ARIMA

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

Information Criteria:

AIC	AICc	BIC
1256.5967	1256.8416	1260.4992

The mean error of the ARIMA model is less than that of the ETS model

The MPE of the ARIMA model is less than that of the ETS model

The AIC of the ARIMA is 1256.5967 which is less than that of the ETS (1634.6435)

The ARIMA model is used for forecasting since the AIC is very less when compared to the ETS 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.**

Forecast error measurements against the holdout sample for the ETS and ARIMA Models are given below

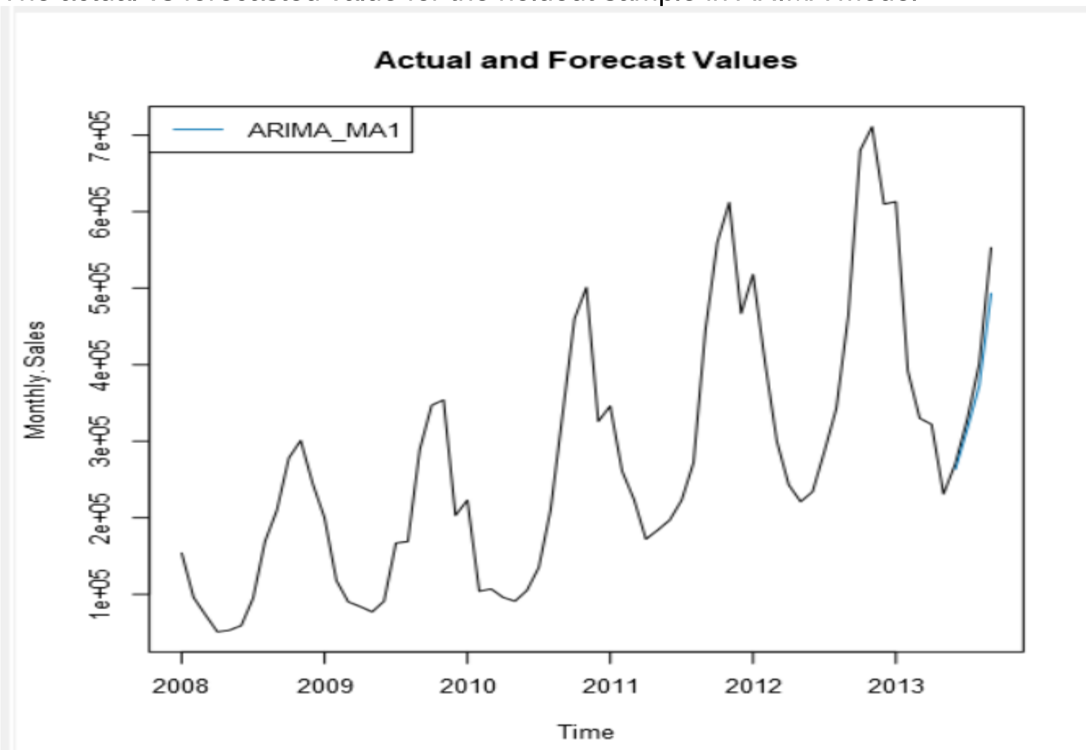
Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
ETS_MAM	-68257.47	85623.18	69392.72	-15.2446	15.6635	1.1532	NA

Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
ARIMA_MA1	27271.52	33999.79	27271.52	6.1833	6.1833	0.4532	NA

The model chosen for the forecast is ARIMA (0,1,1) (0,1,0) 12, because of the following factors, because the RMSE and the MASE are much lower for the ARIMA model

Also, the AIC is low for the ARIMA model

The actual vs forecasted value for the holdout sample in ARIMA model



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 given below

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

The forecast graph of the entire time series, zoomed in to show the year 2013-2014 forecase