Time Series Student Project

1. Introduction

In this student project, I will employ the power of Time Series on website traffic of a popular online deal site in the Philippines. Website traffic is the amount of data sent and received by visitors to a web site. This is measured by the number of visitors and the number of pages they visit. (Source: www.en.wikipedia.org)

The subject of this project is an online deal site launched in 4th quarter of 2010. On its inaugural month, it registered a total of 84,963 page loads. Since then, it has its own share of downtime due to website traffic.

2. Data

Below are the monthly page loads from October 2010 to September 2012.

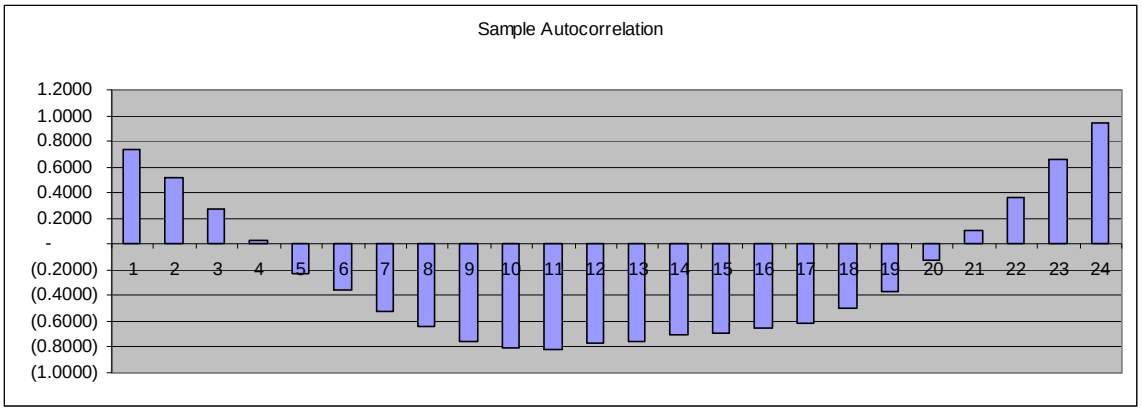
Period	Page Loads	Period	Page Loads	Period	Page Loads
		Jan-11	292,206	Jan-12	1,163,095
		Feb-11	476,601	Feb-12	1,071,425
		Mar-11	596,250	Mar-12	1,103,967
		Apr-11	602,995	Apr-12	1,040,932
		May-11	778,271	May-12	1,497,399
		Jun-11	880,066	Jun-12	1,373,883
		Jul-11	965,185	Jul-12	1,361,021
		Aug-11	1,147,406	Aug-12	1,206,463
		Sep-11	1,021,265	Sep-12	1,304,462
Oct-10	84,963	Oct-11	965,411		
Nov-10	142,679	Nov-11	1,140,811		
Dec-10	272,475	Dec-11	1,268,575		

(Source: www.statcounter.com)

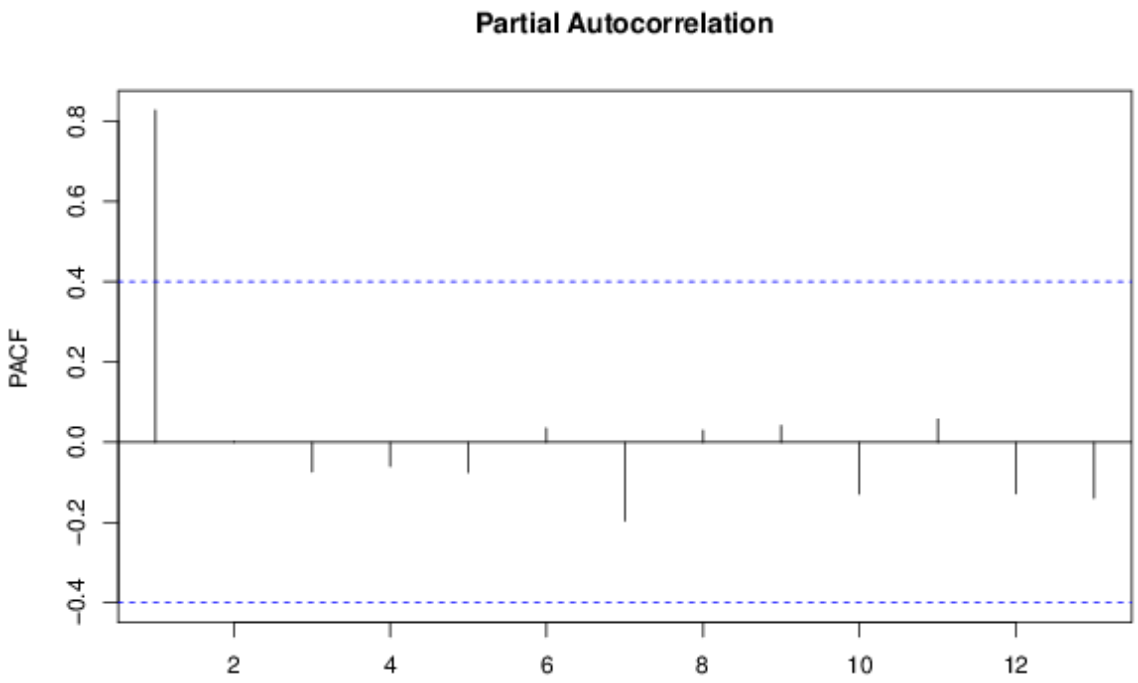
3. Analysis

The graph of the monthly page loads from October 2010 to September 2012, above is shown below.

Based on this graph it appears that seasonality of data is absent for this web site. This observation can be confirmed upon further examination of its sample autocorrelation against its lag as shown below



The above correlogram suggests that the process is not an MA process. The steady decline and the latter rise of the values implicate that the model is best characterize by AR(p). A look on it partial autocorrelation is given below:



(Source: http://www.wessa.net/rwasp_autocorrelation.wasp/)

The sudden drop to zero suggests that AR(1) is would best fit the time series data.

AR(1) Model Analysis

Excel Regression Analysis tool provides the following data:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.93967654
R Square	0.88299199
Adjusted R Square	0.87742018
Standard Error	132.69937
Observations	23

ANOVA

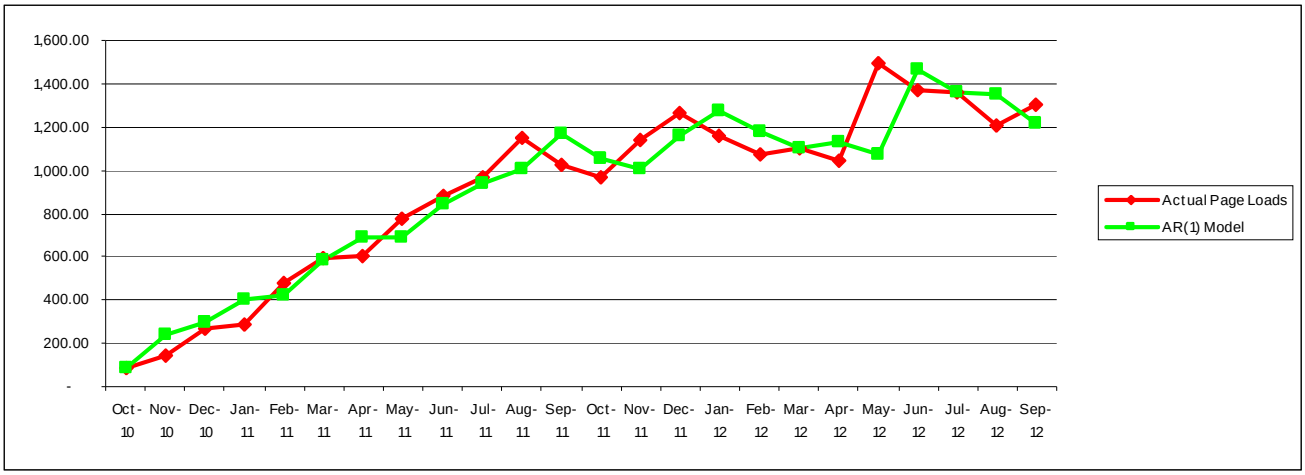
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	2790603.963	3E+06	158.475	2.996E-11
Residual	21	369791.5788	17609		
Total	22	3160395.542			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	169.957576	67.30285135	2.525	0.01967	29.99361	309.921541	29.99360988	309.9215413
X Variable 1	0.86850438	0.068990893	12.59	3E-11	0.7250299	1.01197882	0.725029934	1.011978818

Thus, the model is Y(t) = 169.957576 + 0.86850438 Y(t-1)

Note that the coefficient is less than 1, hence the model is stationary.

A plot of the actual data against the model (AR(1)) is shown below:



4. Conclusion

This project demonstrates time series modeling using website traffic data. Based on the collected data from October 2010 to September 2012, it is inferred that the best model to depict the data is an AR(p) with p = 1, with the following parameters:

$$Y(t) = 169.957576 + 0.86850438\ Y(t-1)$$