SIDDHARTH BHATTACHARYA (61510787) **Assignment 1**

Solution to Chapter 2

1. The time plot of the data is given below.
2. Now if we take the log of sales and again plot log(sales) against time we get a graph which looks like the following.

The logarithmic time plot looks most linear.

1. Comparing the two time plots it can be said that taking log of sales and fitting it against time makes the fluctuations in the data series much smaller and it is approaching a near linear trend. Although, fluctuations remain the fluctuations are much less compared to the original time plot and are also showing a distinct increasing trend upwards. Due to this the trend line fits much better than the previous case.

Solution Chapter 3

1. The data was partitioned so that after creating the model it can be tested on a fresh test set provided by the test/validation set. We build the model using the traing set but keep the last 12 months data intact so that a new data set can be used to test the model. Further, the problem of overfitting (where the same data used to create the model is used to test it resulting in very high but inaccurate performance) can be avoided. The results are much more accurate since it is tested with fresh set of data.
2. The analyst should choose 12 months for the validation period since that is the smallest unit of the forecasting horizon which simulates the actual predictive performance well. Choosing a longer forecast or a shorter validation set will not be a good strategy here. If we have 1 year’s fresh data, we can get a good picture of what will happen on testing in real world.
3. The naïve forecast for the validation period will have the following objective function.

Ft+1 = Yt-M+1 (where M here will correspond to 1 year).

1. The RMSE for the forecast is given by 9542.34, while the MAPE is given by 27.279.(Calculations shown in Excel).



1. The plot of the Errors is given below.

The time plots for respectively the Actual Sales and the Naïve Forecasts are given below.

**Actual**:

**Naïve Forecasts**

The naïve forecast seems to give a better value compared to the actual values .It is steeper than the actual hinting more sensitivity to change in time.

f). From the above calculations we can say that the analyst found the 1 year prior model to be quite accurate and hence should use this model to predict results for 2002 as well. The 2002 values are based on the 2001 results, each month’s prediction based on the previous month’s prediction.

Solution to Questions Chapter 9

T*ips & Suggested Steps*

1. First we plot the demand (aggregated) versus days, the following graph appears. From the graph it is clear that there is a clear seasonality in the data occurring after every 7 day periods. There are a total of 3 such seasons occurring over the period. It is also a global pattern since it is occurring uniformly over the period with stipulated periods of time.



Next we do a plot of Demand versus Time(minutes in a day),we find that (given in the trend below) that the demand is the highest around 7PM and falls drastically again by 9PM.It is also true that since there is no service after 10PM there is no demand between the period 10PM to 6:30 AM next day.



If we look at the data closely we find that since the huge downward decline happens every seven days the demand is falling precipitously at repeated periodic intervals. Now to investigate this further it should be observed that over a three week period there will be 3 weekends and hence maybe the demand is falling during the weekends but this needs to be analyzed from the data given.

1. When we plot Demand versus Weekday plot we see that this is exactly what is happening(the figure below substantiates this further).On Weekdays(as shown in the graph) the demand is 7x times the demand on weekends(Saturdays and Sundays).Thus the cyclical drop and rise in rise in demand can be explained by almost no services being availed on the weekends and very high services on weekdays. Now a practical implication of this could be that may be passengers don’t avail the transport option much. Thus may be more promotion and marketing needs to be done so that customers are attracted to this option. Providing attractive fairs over the weekend may be an option to increase demand(have separate fairs for weekdays and weekends).
2. The comparison for weekdays and weekends is given in the table below. It is clear that the weekends have almost very low to no demand while the weekdays have almost 7xtimes that demand. To understand the situation better we should create separate demand functions for week days and week ends and then forecast the demand for future period using regression technique. Thus running separate regression techniques for both week days and weekends should give a deeper idea about the underlying pattern of demand. Also a correlation analysis can be carried out to check the correlation between a weekday’s demand and a correspondingly same day next week and a weekend’s demand and correspondingly next weekend’s demand.

. 

1. The missing values in this data set could correspond to lean periods(for which sufficient data is not available ) or even more granular breakdowns of data. There is also lack of data available to substantiate the drastic fall in demand during weekends and rise on weekdays. The factors that affect demand are not clearly visible from the data available. Additionally the unusual rise in demand at around 7PM (as shown in 1 graph earlier) needs further investigation as this is not a very convenient time for such sharp increase in demand. Models such as regression and other models can be useful. Also a sensitivity and correlation analysis can give us more insights.