

## Assignment: Forecasting Analytics

This deliverable has 50% weightage in the Consolidated Total Score.

### Deliverables:

A report (**A pdf file**) & 1 R code file.

### General Instructions:

1. This is an individual assignment.
2. Do NOT submit **.zip** files otherwise the submission will not be considered.
3. Any late submission will attract a penalty as mentioned in the course outline.
4. The honour code for this submission is **2N-a**.

### Questions:

1. **Consider the data set SouvenirSales.xls (1995 Jan -2001 Dec) that gives the monthly sales of souvenir at a shop in New York. Back in 2001, an analyst was appointed to forecast sales for the next 12 months (Year 2002). The analyst portioned the data by keeping the last 12 months of data (year 2001) as validation set, and the remaining data as training set. Answer the following questions. Use R to solve the questions.**
  - a. Plot the time series of the original data. Which time series components appear from the plot.
  - b. Fit a linear trend model with additive seasonality (Model A) and exponential trend model with multiplicative seasonality (Model B). Consider January as the reference group for each model. Produce the regression coefficients and the validation set errors. Remember to fit only the training period.
  - c. Which model is the best model considering RMSE as the metric? Could you have understood this from the line chart? Explain. Produce the plot showing the forecasts from both models along with actual data. In a separate plot, present the residuals from both models (consider only the validation set residuals).

- d. Examine the additive model. Which month has the highest average sales during the year. What does the estimated trend coefficient in the model A mean?
- e. Examine the multiplicative model. What does the coefficient of October mean? What does the estimated trend coefficient in the model B mean?
- f. Use the best model type from part (c) to forecast the sales in January 2002. Think carefully which data to use for model fitting in this case.
- g. Plot the ACF and PACF plot until lag 20 of the residuals obtained from training set of the best model chosen. Comment on these plots and think what AR(p) model could be a good choice?
- h. Fit an AR(p) model as you think appropriate from part (h) to the training set residuals and produce the regression coefficients. Was your intuition at part (h) correct?
- i. Now, using the best regression model and AR(p) model, forecast the sales in January 2002. Think carefully which data to use for model fitting in this case.

## **2. Short answer type questions:**

- a. Explain the key difference between cross sectional and time series data.
- b. Explain the difference between seasonality and cyclicity.
- c. Explain why centered moving average is not-considered suitable for forecasting.
- d. Explain stationarity and why is it important for some time series forecasting methods?
- e. How does an ACF plot help to identify whether a time series is stationary or not?
- f. Why partitioning time series data into training, validation, and test set is not recommended? Briefly describe two considerations for choosing the width of validation period.
- g. Both smoothing and ARIMA method of forecasting can handle time series data with missing value. True/False. Explain

- h. Additive and multiplicative decomposition differ in the way the trend is computed. True /False. Explain.
- i. After accounting for trend and seasonality in a time series data, the analyst observes that there is still correlation left amongst the residuals of the time series. Is that a *good* or a *bad* news for the analyst? Explain.

**Deadline: 9<sup>th</sup> June 2021, 11:55 pm**