**Texas A&M Commerce University**

**College of Business - Department of Marketing & Business Analytics**

Assignment for Course:      BUSA 511 Analytics for Managers

Submitted to:                     Dr. William J.Harris

Submitted by:                     Ruthvik Rajan Banda

                                            50326130

                                            rbanda2@leomail.tamuc.edu

Date of Submission:    10-18-2022

Title of Assignment: **Individual Project on Forecasting**

CERTIFICATION OF AUTHORSHIP: I certify that I am the author of this paper and that any assistance I received in its preparation is fully acknowledge and disclosed in the paper. I have also cited any sources from which I used data, ideas of words, whether quoted directly or paraphrased.  I also certify that this paper was prepared by me specifically for this course.

Student Signature:   \_\_\_\_\_\_\_\_\_Ruthvik Rajan Banda.\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Part II – Cutting Edge Company Case Study**

**Individual Forecasting Project Instructions & Report**

***Instructions***

This is an individual assignment and therefore it must be completed by the individual student without outside assistance of any type. Enter your answers to all questions into this document. Use only the models provided to you in this assignment – no other statistical programs or applications are permitted. Follow the instructions below in order to complete the assignment.

* **Part 1.** First, download and read the Microsoft Word case write-up for the “**Part II Cutting Edge Case Study for Ind Forecasting Project**”. Enter your answer to Part 1 below and refer to the post consolidation period only.
* **Part 2**. Second, download the Excel file “**Part II** **Cutting Edge Data\_ES\_LR Models Student V1.xlsx**” and enter your answers for Part 2 below and refer to the 18 months of data covering the consolidated center operations. Note: this is an assessment and interpretation of using time series data, casual analysis and linear regression – the use of the models with resulting calculations is to be accomplished.
* **Part 3.** Finally, provide a summary of your findings by answering the questions.

***Submittals***

1. Submit your answers to this assignment (parts 1, 2, 3) using this Microsoft Word document to the assignment drop box before the posted deadline. Be sure to complete the above first page cover sheet. Enter your answers in the pages below to include all of your answers and results of your interpretations of model calculations for the assignment questions. Your answers must be entered directly into this Word document below each question. Keep your answers concise and use the space guidelines provided in each part.
2. Submit your Excel spreadsheet(s) with calculations/ visualizations to the assignment drop box before the posted deadline. Your Excel model calculations will be used to substantiate your answers to the assignment questions herein.

***Grading***

A total of 100 percentage points is possible for this assignment. This includes the point values which are assigned to each question (point values are noted next to each question below). Ten (10) points may be deducted based on not following the prescribed assignment format and/or distracting punctuation and grammar errors. The percentage points earned on this assignment will be multiplied by 20 to obtain the final assignment grade. Use APA 7 format for any answers that the student determines to require outside references. In your written explanations, do not use underlines, highlights or italics for any part of your descriptive answers.

**Part 1 (10 points)**

In answering questions for part 1, reference the Cutting Edge case study file: “**Part II Cutting Edge Case Study for Ind Forecasting Project**”. Your answer should not be longer than 150 words.

***Question 1*** (5 points): Define a problem statement which reflects the challenges facing Mark as he addresses the new issues associated with monthly forecasting.

**Answer**:

As Mark was successful in setting up centralized data for the company and forecasting the daily call volume to overcome the complaints and issues with employee staffing. Now, Mark had noticed there is a downward trend on the monthly KPI’s. Going into details Mark had to go into detail for finding the root cause of the downward trend, and finds there are several factors affecting this based on the company’s business activity. As Cutting Edge is active in acquiring companies and selling off portions of an existing business, the employees have to face severe distress as they had to lose their job. Even, with the tax reforms and several other factors based on the company’s decision, employees need to suffer which subsequently results in failure to retain their employees. Further, handling these resignations and hiring fresh-man and training them was a tedious job that has more uncertain effects on the call center operations. Therefore, Mark analyzed there is an immediate cause and effect for every decision the company makes, Mark has gathered all the qualitative facts and sequential history of the company’s actions and is interested in studying monthly forecasts to overcome this scenario.

**Part 2 (50 points)**

In answering Part 2 questions, download and reference “**Part II** **Cutting Edge Data\_ES\_LR Models Student V1.xlsx**” which contains the historical data that was used in preparing the forecast results for your assessment. Note, in answering these questions you **need to produce scatter plots, calculate the correlation coefficients and use the two Excel forecasting model results**. Your answers are to be based on the assessment and interpretations of 18 months of data found in the Data & Correlation tab.

***Question 2a*** (10 points): This question has 5 sub-parts shown below in bold. Access the data in the Data and Correlation tab and produce 3 scatter plots using the Actual Call Volume as the dependent variable and Engineering Head Count, Administration (G&A) Head Count and Manufacturing Head Count. Your scatter plots should be visible in the provided Excel spread sheet. Do the scatter plots show linear relationships?

**Answer (yes/no):** \_\_YES\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Now, using the Excel function CORREL in the provided spreadsheet Data & Correlation tab, calculate both the Correlation Coefficient and the Coefficient of Determination for the remaining three (3) independent variables to the dependent variable. Provide the resulting values and describe the strength, quality and relationship of each to the dependent variable.

**Engineering Head Count** v**alues and descriptive relationship answer**:

The correlation between call volume and engineering headcount is a very strong relationship with a 0.9161 Pearson correlation coefficient and 84% coefficient of determination. As the number of employees on this trait is constantly rising irrespective of the business decisions, there seems to be a very correlation and strong positive relationship between the variables.

**Administrative (G&A) Head Count** v**alues and descriptive relationship answer**:

The correlation between Administrative(G&A) and call volume is 0.730 correlation coefficient and 54% coefficient of determination. As the headcount in this trait seems to be varying with the company’s decisions. But, still, there is a high correlation and substantial positive relationship between the variables.

**Manufacturing Head Count** v**alues and descriptive relationship answer**:

The correlation between Manufacturing and call volume is 0.790 correlation coefficient, and 63% of the coefficient of determination. The headcount was initially high for a few months and later the count started to fall due to uncertainties or external factors on the employees. Again, after acquiring other companies the situation got better, and the correlation between the variables was with high correlation and substantial positive relationship.

Next, take note of the first Correlation Coefficient (Actual Call Volume and Total Co. Employee Head Count) that is provided for you to use in comparison to the other calculated coefficients.

**Which Independent variable impacts the dependent variable the most?**

**Answer:**

Engineering Headcount with an 84% coefficient of determination having a very strong relationship between the variables impacts the most.

**Question *2b*** (10 points): Access the data in the Data & Correlation tab and copy the Actual Call Volume data (column D) into the **Exponential Smoothing tab** to calculate a forecast for time period (month) 19. Note and **use the value of 24,279 for the initial value in cell J10**. Use trial and error to determine the best alpha value for this Exponential Smoothing Model. Provide the new forecast from the Exponential Smoothing model (at the end of column D) along with the model accuracy values and enter them below.

**Month 19 Forecast**: \_\_31,851\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Alpha Value Selected**: \_\_0.9\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**MAD Accuracy value**: \_\_\_\_1,231\_\_\_\_\_\_\_\_\_\_\_\_\_

**MAPE average Percent Error value**: \_\_\_\_\_4.60%\_\_\_\_\_\_\_\_\_\_\_\_

**MSE value**: \_\_\_\_\_3,167,407\_\_\_\_\_\_\_\_\_\_\_\_

**ME Bias value**: \_\_\_\_467.40\_\_\_\_\_\_\_\_\_\_\_\_\_

For the following, keep your answer to less than 150 words. Note the resulting line graph in the Exponential Smoothing tab and along with the qualitative facts from the Data & Correlation tab, answer the following: What are at least two (2) of the major causes (events) of the trend from period 9 through period 18 which had the greatest impact on headcount and call volume?

**Causal descriptions with supporting facts/data Answer**:

After selling the printer division, the headcount on three of the traits seems to fall from the 9th period resulting in a surge in the forecast for the call volume (ActualValue:23456, ForecastValue:27010). Between the 10th to 12th periods, the forecast seems to have an uncertain trend and has been affected by a lower forecast than the actual call volume. From the 11th period, we see the company headcounts are improving but there is a lower forecast for those periods. For the 15th period again after acquiring Paxton Enterprises the head count was strong on all of the traits but the call volume is a bit higher than the forecasted value, there is the possibility of other factors affecting the forecast. From the 16th period onwards there seems a stable trend. But still, this forecast model has MAPE of 4.60% Error in the trend and constitutes biases Mean Error of 467.40 looks makes the forecast inaccurate.

***Question 2c*** (10 points): Access the data in the Data and Correlation tab and **copy the dependent variable and the Total Co. Employee Head Count independent variable** into Linear Regression tab data columns and assess the results. **Use the value from the Data & Correlation cell (E22) for the Linear Regression estimator cell (x value, cell L8)**. Provides a forecast of the Call Volume based on the Total Co. Employee Head Count. Also, provide the model quality values below.

**Month 19 Predicted value (y) Forecast based on Total Co. Employee Head Count:** \_\_\_\_\_\_32,017\_\_\_\_\_\_\_\_\_\_

**MAD Accuracy value**: \_\_\_\_990\_\_\_\_\_\_\_\_\_\_\_\_\_

**MAPE average Percent Error value**: \_\_\_3.82%\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**MSE value**: \_\_\_\_\_\_1,557,888\_\_\_\_\_\_\_\_\_\_\_

**ME Bias value**: \_\_\_\_\_0.00\_\_\_\_\_\_\_\_\_\_\_\_

***Question 2d*** (10 points): Again, access the data in the Data & Correlation tab and, **using the same dependent variable, copy the independent variable that has the strongest impact and highest quality** to the dependent variable into the data column of the Linear Regression tab. **Use the Correlation Coefficient and Coefficient of Determination values from the Data & Correlation tab** as the basis of your selection. Also, **use the supporting head count value from row 22** in the Linear Regression estimator cell (x value) L8. Provides a forecast of the Call Volume based on the Head Count that has the greatest impact. Also, provide the model quality values below.

**Which independent variable was selected:** \_\_\_\_Engineering Head Count\_\_\_\_\_\_\_\_\_\_\_\_\_

**Month 19 Predicted value (y) Forecast based on selected Head Count variable:** \_\_\_32,903\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**MAD Accuracy value**: \_\_\_\_\_811\_\_\_\_\_\_\_\_\_\_\_\_

**MAPE average Percent Error value**: \_\_\_\_3.18%\_\_\_\_\_\_\_\_\_\_\_\_\_

**MSE value**: \_\_\_1,123,756\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**ME Bias value**: \_\_\_0.00\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Part 3 (40 points):**

In Part 3, you will use your answers from parts 1 and two 2 to summarize your findings and provide recommendations based on analytic interpretation of data and facts.

***Question 3a*** (10 points):

For the following, keep your answer to less than 150 words. In terms of how you used each model, what was the difference in using Exponential Smoothing verses Linear Regression in assessing causation from the data and model calculations? In other words, what *assessment* *information* did each model provide you that the other did not?

**Answer**:

Exponential smoothing is a simple model which had no absolute forecast for the call made on monthly basis, it is based only on the historical call volume but has not taken any cause and effect/qualitative facts into consideration. Whereas, Linear Modeling is the best model where we could see first where we are trying to normalize by taking the slope and intercept into consideration for forecasting the call volume. Exponential smoothing uses a smoothing constant, which is not the appropriate way to keep a check on the values and select the model with the lowest MAD, It is almost the manual interpretation where we have a lot of uncertainty about the model we have chosen. Linear modeling is based on the calculative fields of having the highest correlation to forecast the call volume, which in turn uses a standard format of normalizing the curve by taking the slope and intercept for the forecast.

***Question 3b*** (10 points): For the following, keep your answer to less than 150 words. Given the accuracy, percent error, and bias, which model provides the best forecast for period 19 and why?

**Answer**:

I choose to select the Linear Regression Forecast Model for predicting the call volume for the problem faced by Mark. Where we can refer to the model calculative fields taking into consideration working with engineering headcount, with the lowest MAD value of 811, and the MAPE is at least 3.18% with ZERO biases. And, when tried with other traits such as Administrative & Manufacturing headcounts we still have Zero biases, but when comparing the MAD & MAPE seems to be on the higher side and not satisfying as what Engineering Headcount does. Whereas, The Exponential Smoothing when chosen with Alpha=0.9 smoothing constant we derived the lowest MAD value with 1231, & MAPE with 4.60%, and when coming to biases the facts are distressing with biases of 467.40 for this model. So, I choose to go with model which has the lowest MAD & with the lowest Error Percentage (MAPE) i.e, Linear Regression Model for forecasting the call volume.

***Question 3c*** (5 points): What is your forecast for period 19? Provide justification and/or rationale in terms of facts and keep your answer to less than 200 words..

**Call Volume Forecast for period 19 (My forecast**): \_\_\_32,903\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Explanation and Justification of Your Forecast Method**:

Ahead of choosing, the Linear Regression Forecasting model as the best fit for predicting the monthly call volume, then swiftly started to identify which independent variables work the best in predicting the forecast. Considering the Engineering Headcount which has the highest correlation among the other variables, With a MAD value of 811 and & MAPE of 3.18%. And again, considering the second-highest correlated variable for the prediction the outcomes seem to have a MAD value of 1,269 & MAPE of 4.91% has a forecast of 31,736 for period 19. When considered with Administrative(G&A) the MAPE is all-time highest with 5.28% which is unacceptable to consider anymore. Though interestingly, all variables seem to have no biases, when considering the error percentages and MAD values, Engineering headcount seems to be the appropriate independent variable for forecasting the monthly call volume. Therefore, the best outcome for the monthly call volume forecast for period 19 is 32,903.

***3d Summary*** (15 points): Provide process and/or method recommendations to allow Mark to provide the most accurate month-to-month forecast. Respond in terms of what he should do and what methods he should use. Keep your descriptive answer to 500 words or less below.

I recommend Mark, using Linear Regression Modelling for his monthly-monthly call volume forecast would solve the problem to have an accurate call volume prediction. Having Zero Biases is the most important advantage to consider in the linear regression model. And the next apex factor is to find which variable has the best correlation among the other variables and start working with it to have a better forecast. Irrespective of the qualitative facts or the company decisions the Linear Regression model normalizes the curve and tries to get the best forecast. With this said, Mark can staff the employees accordingly to the forecasted call volume and be extensive in hiring the employees and training them in advance to suffice the surge in call volume and maintain the KPI standards set by the company. However, Mark had forecasted the daily call volume forecast successfully and now this monthly forecast helps him to better compare them effectively and cautiously prepare for staffing the employees efficiently and cater to the needs of the business.