***BUS 347.01 Introduction to Business Analytics***

***Homework 2 (40 Points)***

**Homework Description:**

The attached file “HW2 Data.csv” contains 500 observations from a simulated dataset. Some variables in this dataset requires data cleaning.

You need to perform a data cleaning on this dataset by following the procedure specified in each question. Your submission should be an analytics report in the word format knitted from RMarkdown code.

In your analytics report, you need clearly label the following items:

* Question Number
* R Code
* R Output
* Conclusion, if applicable

You need to submit the knitted analytics report on Blackboard. Please carefully check your work before the submission, as you can only submit your work once. Late submissions will not be accepted.

**Homework Questions (Each Question has 10 point)**

**Q1.** Use the “Winsor” method to treat the outliers of variable X. In this Winsor method, the normal range of the variable is defined by [Q1-1.5\*IQR, Q3+1.5\*IQR]. Save the winsored X as X1. Use the summary() function to provide a statistical summary of X1. (Partial points can be given if students generate the correct normal range, you need to show those normal ranges to get the partial points)

**Q2.** Use the “Winsor” method to treat the outliers of variable X. In this Winsor method, the normal range of the variable is defined by [1% percentile, 99% percentile]. Save the winsored X as X2. Use the summary() function to provide a statistical summary of X2. (Partial points can be given if students generate the correct normal range, you need to show those normal ranges to get the partial points)

**Q3.** Based on X1, perform a missing imputation through replacing the missing value by an unconditional mean of X1. Save this imputed X1 as X3. Use the summary() function to provide a statistical summary of X3. (Partial points can be given if students generate the correct unconditional mean, you need to show those normal ranges to get the partial points)

**Q4.** Based on X1, perform a missing imputation through replacing the missing value by an condition mean of X1 on variable Z. Save this imputed X1 as X4. Use the summary() function to provide a statistical summary of X4. (Partial points can be given if students generate the correct conditional means, you need to show those normal ranges to get the partial points)