CSC 400, Exam #2, At Home, Spring 2024 Name _____

Instructions: This exam is in two parts: Part I is to be completed partly at home using the materials posted in the course for the at-home portion and you will answer questions about that work during the in-class portion of the exam; Part II is to be completed entirely in class. You may not use cell phones, and you may only access internet resources you are specifically directed to use.

At home, prepare for questions in Part I using R. Complete the calculations noted below. You will be asked for additional analysis and interpretation of this data in the in-class portion of the test. Print out the results of your analysis and code, and bring the pages with you to the exam. You will submit all this work along with the in-class exam.

Use the data on wine to complete the following tasks after importing the data **400exam1data.xlsx** in the file into R (this is the same dataset we used on the first exam.

1. Set aside the Type variable for now (save it before removing it from the dataset). Create a clustering model for the data set (you do not need to separate the data into test and training sets for this) using the following algorithms: K-means, Spectral, DBSCAN and mean shift. Compare models with k=2, k=3, k=4 clusters, and any other optimal model identified.

2. For each model identified above, create a confusion matrix (for the k=3 model using the Type data from the original dataset), and create appropriate model or diagnostic graphs.

3. Identify any outliers.

Import the dataset **beersales** from the TSA package.

4. Perform appropriate base time series analysis of the data such as differencing, acf and pacf graphs, decomposition, etc. Create appropriate graphs.

5. Create an ARIMA model of the time series. Create an exponential smoothing model of the time series. Use both to forecast 10 months into the future. Identify any margins of error on your predictions.

6. Redo the above analysis but remove the final 10 observations from the original data set as a test set, and using the remaining observations as a training set. Use your forecast model to predict those last 10 observations and compare the real data to your predictions. Perform appropriate diagnostics.