CS677 HW4

Create a Jupyter notebook **CS677_HW4_lastName** and add the code and output cells for the following:

Part1

Use the Dow Jones Industrials dataset shown in the following link:
http://people.bu.edu/kalathur/datasets/DJI.csv

1) Read the data frame parsing the dates and the Date as the index in a single statement. Show the tail of the resulting dataframe.

2) Create a dataframe, weekly_mean_volume, with the weekly averages of the Volume attribute only. Show the time series plot for this data. What do you infer from the plot.

3) Create a dataframe, monthly_mean_volume, with the montly averages of the Volume attribute only. Show the time series plot for this data. What do you infer from the plot.

4) Using statsmodels seasonal decompose, find the trend, seasonal and residual components for the weekly_mean_volume dataframe. Show the plots for the observed and the trend values, the seasonality, and the residual components. Show also the seasonal plot for the year 2018 only. Summarize the plots with brief descriptions.

Part2

Use the bank marketing dataset shown in the following link (note that separator is ;)
http://people.bu.edu/kalathur/datasets/bank-full.csv

The description of the dataset can be found here:
https://archive.ics.uci.edu/ml/datasets/bank+marketing

5) Build a decision tree for conversion (last column) using Gini and Entropy separately using the entire dataset. Show how good your model predicts the entire dataset using the appropriate measures.

6). Using the last four digits of your BU id as the random state, use 70-30 split for training and test data. Build the previous models using the training data and predict the outcomes on the test data. Show how good your model predicts the test data using the appropriate measures.

7). Use Logistic regression model to fit the conversion (last column) using the entire dataset. Show how good your model predicts the entire dataset using the appropriate measures.

8) Use Gaussian Naïve Bayes classifier to build the model for the conversion (last column) using the entire dataset. Show how good your model predicts the entire dataset using the appropriate measures.