IE 360 Statistical Forecasting and Time Series
Homework 2, due May 11th, 2022

Instructions: Please solve the following exercises using R (http://www.r-project.org/) or Python (https://www.python.org/). You are expected to use GitHub Classroom and present your work as an html file (i.e. web page) on your progress journals. There are alternative ways to generate an html page for you work:

- A Jupyter Notebook including your codes and comments. This works for R and Python, to enable using R scripts in notebooks, please check:
  - https://docs.anaconda.com/anaconda/navigator/tutorials/r-lang/
  - https://medium.com/@kyleake/how-to-install-r-in-jupyter-with-irkernel-in-3-steps-917519326e41

  Things are little easier if you install Anaconda (https://www.anaconda.com/). Please export your work to an html file. Please provide your *. ipynb file in your repository and a link to this file in your html report will help us a lot.

- A Markdown html document. This can be created using RMarkdown for R and Python-Markdown for Python

Note that html pages are just to describe how you approach to the exercises in the homework. They should include your codes. You are also required to provide your R/Python codes separately in the repository so that anybody can run it with minimal change in the code. This can be presented as the script file itself or your notebook file (the one with *.ipynb file extension).

The last and the most important thing to mention is that academic integrity is expected! Do not share your code (except the one in your progress journals). You are always free to discuss about tasks but your work must be implemented by yourself. As a fundamental principle for any educational institution, academic integrity is highly valued and seriously regarded at Boğaziçi University.

Task:
Please download the data given in the file "IE360_Spring22_HW2_data.csv". This file contains quarterly gasoline and diesel sales (in 1000 m3) of a major distributor between 2000 and 2006, and a number of potential independent variables. These variables are summarized as follows:

UGS: Unleaded gasoline sale in a given quarter,
DGS: Diesel gasoline sale in a given quarter,
RNUV: An index indicating the rate of new unleaded gasoline using vehicles being added to the traffic in a quarter,
PU: Average price (adjusted with an index) of a liter of unleaded gasoline in a quarter,
PG: Average price (adjusted with an index) of a liter of diesel gasoline in a quarter,
NUGV: Number of unleaded gasoline using vehicles in the traffic,
NDGV: Number of diesel gasoline using vehicles in the traffic (per 1000 people),
GNPA: Agriculture component of Gross National Product (adjusted with an index),
GNPC: Commerce component of Gross National Product (adjusted with an index),
GNP: Grand total for GNP (agriculture, commerce and other components total).

Your ultimate task will be to forecast the sales (for UGS and DGS) for every quarter of 2007 by using time series regression methods:

Before you start with forecasting, you should perform and show the following steps:
1. Plot the time series of UGS and DGS. Comment on the shapes of the plots. Specifically, do you think the time series is stationary with respect to its mean and variance?
2. Plot the autocorrelation functions of the time series UGS and DGS (get autocorrelations for at least 8 lags). What do you think the autocorrelation values at different lags indicate?

FORECASTING WITH REGRESSION
3. In addition to existing variables, you may need to define seasonality and trend related variables. You may also want to include "lagged" variables. That is, you may want to use Y_{t-1} (or any X_{t-1}) to explain Y_t.
4. You should try various regression models and come up with a model that explains sales in terms of independent and trend/seasonality variables.
5. You should present all the statistical evidence supporting the validity of your analysis (residual analysis, significance of coefficients, etc.). This should be done only for your selected model.
6. Using the fitted model, forecast unleaded and diesel gasoline sales for 2007 (all quarters).