IE 360 Course Project
Due Date: May 22, 2023

**For the project you are required to form groups of four students. Please form
your groups as soon as possible, and send your group information to Emre
Kara by April 3, 2023.**

Please download the data given in the file "IE360-ProjectData.xls". This file contains
quarterly gasoline and diesel sales (in 1000 m^3) of a major distributor between 2000-2006,
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 task will be to forecast the sales (for two products) for every quarter of 2007
by using two methods: Method A: Forecasting with time series analysis, Method B:
Forecasting with regression. 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. Specifi-
cally, do you think the time series are stationary?

2. Plot the autocorrelation functions of the time series UGS and DGS (get autocorrela-
tions for at least 8 lags). What do you think the autocorrelation values at different
lags indicate?

METHOD A: FORECASTING WITH TIME SERIES ANALYSIS

1. You should first check if you need a preliminary transformation to induce stationarity (such as logarithmic).

2. You should utilize time series plot, ACF and PACF plots in order to determine regular and seasonal differencing to be applied.

3. You should come up with an initial ARIMA model based on inspection of ACF and PACF plots.

4. You should performs a neighborhood search of the initial model. That is, a search around initial AR, MA, and seasonal AR, seasonal MA terms by increasing or decreasing them. Limit yourself to six such candidate models.

5. These candidate models should always be accompanied with ACF and PACF plots.

6. You should decide on the "best" model to use in forecasting.

7. You should present all the statistical evidence supporting the validity of your analysis.

8. Forecast gasoline and diesel sales in 2007 (all quarters) with the selected method.

METHOD B: FORECASTING WITH REGRESSION

1. Decide if you need a preliminary transformation to induce stationarity (such as logarithmic).

2. 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.

3. You should try various regression models and come up with a model that explains sales in terms of independent and trend/seasonality variables.

4. You should present all the statistical evidence supporting the validity of your analysis (residual analysis, Durbin-Watson statistics, colinearity checks, significance of coefficients, etc.).

5. Using the fitted model, forecast unleaded and diesel gasoline sales for 2007 (all quarters).

WHAT YOU SHOULD SUBMIT?

1. An introductory section describing the observations from the time series plot and autocorrelation plots.

2. A report containing all your analysis, tables, figure and forecasts (separate sections for Method A and Method B).

3. A section that compares results obtained by two methods.

4. A final set of forecasts for 2007, based on your conclusions on both methods and your assessment.