**Actions:**

* Using the ToyotaCorolla.xlsx dataset, the prepared and evaluated data set (Clean Data) your team created for assignment one, and XLMiner, apply the linear regression algorithm.  
  IMPORTANT: Be certain that your variables are correct before you start a regression model. Read the comments made on your team’s Assignment 1.
* Partitioning at the time of the model run is required. Not before.
* Use appropriate partitioning 60/40 and **all variables** for the first run.
* The subsequent runs will exclude variables of your choice based on a valid/justified reason.
* Using metrics (see below) assess the performance of the model in predicting the variable price.
* When stating number results do not use more than 4 decimal places and remember to check the ANOVA option in the XLMiner regression model.

**Submission:**

* The Excel file name ASG2ToyotaRegressionLastname.xlsx with the XLMiner produced worksheets.
  + The first worksheet in the workbook must always be the updated Data Description sheet.
  + The second worksheet should be the Original Data.
  + The third worksheet, Clean Data, should be the data ready to be mined; this means all data is numeric and missing data is imputed or deleted.
  + Subsequent labeled worksheets will be the output to the **first regression run with all variables** and then the **best run with variables excluded to improve performance**.  
    If you run more than two models remove those worksheets to runs that you do not wish to submit. Do not remove any worksheets from the runs you are keeping.
* The Regression model interpretation at a minimum requires the following items for each run:
* **Enter the assignment questions and the answers on the Data Description worksheet.**
* **First run** interpretation
  1. Is the Model significant? Why?
  2. What is the R2 and what does it mean?
  3. Which variables are significant at the .05 Alpha level?
  4. State the Regression Equation in terms of the variables of the problem (not x and y).
  5. Using the Regression Equation, enter the data values from any one record, predict the Price for that record and the residual.
  6. State the following two Metrics
     + RMSE
     + MAPE (use the information on the MLR\_ValidationScore worksheet to calculate the Mean Absolute Percent Error). See Module 4.
* **Best run** interpretation

1. State which variables were removed for the best run and why.
2. Is the Model significant? Why?
3. What is the R2 and what does it mean?
4. Which variables are significant at the .05 Alpha level?
5. State the Regression Equation in terms of the variables of the problem (not x and y).
6. Using the Regression Equation, enter the data values from the same record used in the first run prediction, predict the Price for that record and the residual.
7. State the following two Metrics
   * + RMSE
     + MAPE (using the information on the MLR\_ValidationScore worksheet calculate the Mean Absolute Percent Error).

* Provide a brief contrast and comparison of the two runs.
* This is an opportunity for you to explore and understand a prediction model, reducing a dataset using regression and comparing models with metrics.