Introduction to Data Analysis (DATA 1200)
Assignment #2 – Predictive Modeling (15% of Final Grade)
Professor: Ritwick Dutta

John Hughes is reviewing the **swiss.csv** dataset and is looking to create a regression model.

The dataset has 47 observations and 6 variables:

Independent Variables:

Fertility – Ig, common standard fertility measure
Agriculture - % of males involved in agriculture as occupation
Examination - % of draftees receiving highest mark on army examination
Education - % education beyond primary school for draftees
Catholic - % "catholic" (as opposed to "protestant")

Dependent Variable:

InfantMortality – live births who live less than 1 year

The Ask:

1. **Create a Python Script using Jupyter Notebook (then convert to .html)** – *2%*
   a) Using Python develop a **Multivariate/Multiple Regression Algorithm** script to predict Infant Mortality. Attach the HTML copy of your Python Code with your submission

   Note: All steps need to be annotated (i.e. As per the Wk4b-MLRExample)

2. **Create a PowerPoint (PPT or PPTX) presentation that includes the following:**
   a) Cover Page (Title, Name (1st and last) and Student Number)
   b) Present the QQ Plots and Explain **three (3) key insights** – *3%*
   c) Present the written form of the Regression Model (i.e. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$) and Explain all the coefficients (i.e. weights and effect) – *4%*
   d) Present the Regression Model Output and Explain **three (3) key insights** from the Model metrics (i.e., Adj. R^2, MAE and RMSE) – *3%*
   e) Present and Explain **three (3) ways** to help improve the performance of the Regression model. Please justify each of your answers. – *3%*

Hint: Leverage Wk4d-Tutorial-MLR

Please post your PowerPoint Document (.ppt or .pptx) and Jupyter Notebook in HTML (.html) format via assignments under Assignment #2 by

Tuesday, October 17th, 2023 @ 11:59 p.m.

Grading Rubric				
	Exemplary (14-15)	**Proficient (10-13)**	**Incomplete (7-9)**	**Needs Improvement (0-6)**
Python Code *(2%)*	Python HTML file is complete	Python HTML file is mostly complete. Missing headings or structure.	Python HTML file is incomplete. Incorrect use of heading or code.	Python HTML file is missing or incorrect.
PPT *(13%)*	Cover Page Complete QQ Plots presented including three (3) key insights with detailed explanations/ justifications Regression Model presented and explained in detail (i.e. coefficient meanings) Regression Model output presented with Three (3) key insights from Model metrics (i.e. Adj. R^2, MAE and RMSE) that are fully evaluated/justified Three (3) ways to improve the model have been identified with detailed explanations.	Cover Page Complete QQ Plots presented including three (3) key insights with high-level explanation/ justification Regression Model presented and with high-level explanations (i.e. coefficient meanings) Regression Model output presented with Three (3) key insights from Model metrics (i.e. Adj. R^2, MAE and RMSE) with high-level evaluations Three (3) ways to improve the model have been identified with only high-level explanations.	Cover Page missing a least one element QQ Plots presented including less three (3) key insights and/or Missing explanation/ justification Regression Model not presented and/or missing detailed explanations (i.e. coefficient meanings) Regression Model output presented with less than Three (3) key insights from Model metrics (i.e. Adj. R^2, MAE and RMSE) and/or incomplete evaluations Less than Three (3) ways to improve the model have been identified and/or incomplete explanations.	Cover Page Missing QQ Plots and/or Insights are missing or incorrect. Regression Model and/or Insights are missing or incorrect Regression Model output and/or explanation missing or incorrect Missing ways to improve the model and/or incorrect.