Statistical and Predictive Modeling II (DATA 2204)
Assignment #1 – *k*-NN Regression (10% of Final Grade)
Professor: Sam Plati

Mr. John Hughes has been reviewing the abalone dataset (**abalone.csv)** and he would like you to create a ***standard and optimized k-NN regression model***.

The **abalone.csv** dataset contains the following variables:

Independent Variables:

Length - mm / Longest shell measurement
Diameter - mm / perpendicular to length
Height - mm / with meat in shell
Whole weight - grams / whole abalone
Shucked weight – grams / weight of meat
Viscera weight - grams / gut weight (after bleeding)
Shell weight - grams / after being dried

Dependent Variable:

Rings - gives the age in years

The Ask:

1. Create a PowerPoint (PPT) presentation that includes the following:
   a. Cover Page (Title, Name (1st and last) and Student Number)
   b. Analysis Statement (i.e. Create *k*-NN models to….) – ***2%***
   c. Present the Correlation Heatmap and explain **two (2)** insights – ***2%***
   d. Present the Learning Curve for the *k*-NN standard model and explain **two (2)** insights – ***2%***
   e. Present and explain **all evaluation results (i.e. Adj. R^2, MAE, RMSE)** for both the Standard and Optimized *k* -NN Regression model – ***2%***
   f. State and explain **two (2) recommendations** for Mr. John Hughes for next steps. ***2%***

Note: Please ensure that all key facts are in your slides and not in the notes section

Hint: Leverage the code from Wk2-kNNReg
Random State = 100 for all section

2. Provide an HTML copy of your python code

Please post your PowerPoint Document(.ppt) and HTML Python Code via assignments under Assignment #1 by 11:59 p.m. on Friday, May 21st, 2021

Grading Rubric

	Above Average	Average	Needs Improvement	Comments
1. Create a PowerPoint (PPT) presentation that includes the following: a. Cover Page (Title, Name (1st and last) and Student Number) b. Analysis Statement (i.e. Create *k*-NN models…) ***- 2%*** *c.* Present the Correlation Heatmap and explain **two (2)** insights ***– 2%*** *d.* Present the Learning Curve for the *k*-NN standard model and explain **two (2)** insights ***– 2%*** e. Present and explain **all evaluation results (i.e. Adj. R^2, MSE, RMSE)** for both the Standard and Optimized *k* -NN Regression model ***– 2%*** *f.* State and explain **two (2)** recommendations do you have for Mr. John Hughes for next steps. ***– 2%*** 2. HTML copy of your Python Code				

Needs Improvement –Missing the minimum requirements stated in the assignment requirements.
Average –Meets the minimum requirements stated in the assignment requirements.
Above Average –Exceeds the requirements that are stated in the assignment requirements.