

# Introduction to Data Analysis (DATA 1200)

## Assignment #1 – Preliminary Analysis (10% of Total Grade)

### Professor: Sam Plati

Mr. John Hughes has been reviewing the Hitters-Cleaned.xlsx dataset. The variables are as follows:

#### Independent Variables

AtBat - Number of times at bat in 1986  
 Hits - Number of hits in 1986  
 HmRun - Number of home runs in 1986  
 Runs - Number of runs in 1986  
 RBI - Number of runs batted in in 1986  
 Walks - Number of walks in 1986  
 Years - Number of years in the major leagues  
 CAtBat - Number of times at bat during his career  
 CHits - Number of hits during his career  
 CHmRun - Number of home runs during his career  
 CRuns - Number of runs during his career  
 CRBI - Number of runs batted in during his career  
 CWalks - Number of walks during his career  
 League - A and N indicating player's league at the end of 1986  
 Division - E and W indicating player's division at the end of 1986  
 PutOuts - Number of put outs in 1986  
 Assists - Number of assists in 1986  
 Errors - Number of errors in 1986

#### Dependent Variable

Salary - 1987 annual salary on opening day in thousands of dollars

#### The Ask:

#### Part A: Python Code in HTML Document (.html) – 2%

Using Python create the **key statistics** (i.e. mean, median, etc.) for the Hitters-Cleaned dataset (i.e. `.describe()` function)

#### Part B: PowerPoint (.ppt or pptx) - 8%

1. Create a PowerPoint (PPT) presentation that includes the following:
  - a. Cover Page (Title, Name (1<sup>st</sup> and last) and Student Number)
  - b. Select any **four (4) independent variables** and explain/justify key insights that were generated in Part A. For example: Mean, Standard Deviation, etc.

**Hint: Leverage the Wk2b-BlueJaysExercise**

<b>Rubric</b>				
<b>Question</b>	<b>Exemplary (9-10)</b>	<b>Proficient (7-8)</b>	<b>Incomplete (5-6)</b>	<b>Needs Improvement (0-4)</b>
Part A (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.
Part B (8%)	Four (4) independent variables with insights presented with detailed explanation/justification	Four (4) independent variables with insights presented with high-level explanation/justification.	Less than four (4) independent variables with selective insights and/or Missing explanation/justification.	Answer is missing or incorrect.

**Please post your Word Document (.doc or .docx) and HTML Python Code via assignments under Assignment #1 by Friday, January 28<sup>th</sup>, 2022 @ 11:59 p.m.**