**IST 718: Big Data Analytics**

**Description**

The research question is how can we recommend the best salary (totalpay or schoolpay – your

choice) for our next head football coach?

• Start with the data Coaches

• Review the data – clean as appropriate

• Consider the base worksheet and additional data such as:

o Stadium size – available via internet search

o Graduation rate :

Available from https://www.icpsr.umich.edu/web/ICPSR/studies/26801/summary

or

From here https://web3.ncaa.org/aprsearch/gsrsearch

Use the most recent cohort available and include both GSR and FGR

o Annual donations to program – if available via internet search

o School’s win/loss record from last available year

o Other data as you determine might be applicable

• Build a data frame for your analysis

• Conduct an initial data analysis – exploratory data analysis – develop appropriate visualizations

• Fit a regression model with the salary as the response and the relevant predictors (i.e., you will

need more than one predictor)

• Answer the following questions in your report:

o What is the recommended salary for the Syracuse football coach?

o What would his salary be if we were still in the Big East? What if we went to the Big

Ten?

o What schools did we drop from our data and why?

o What effect does graduation rate have on the projected salary?

o How good is our model?

o What is the single biggest impact on salary size?

**Required Tools**

* Python
* MS Word

**Table of Contents**

1. README.md
   * Overview Document detailing repository contents
2. IST\_718\_Williams\_Jennifer\_Lab1
   * Lab Report
3. IST\_718\_Williams\_Jennifer\_PythonCode
   * Script used to conduct analysis and visualizations
4. IST\_718\_Williams\_Jennifer\_Data\_Base
   * Base Data for the Lab
5. IST\_718\_Williams\_Jennifer\_Data\_Supplemental\_Conferences
   * Supplemental Data for the Lab
6. IST\_718\_Williams\_Jennifer\_Supplemental\_Conferences
   * Supplemental Data for the Lab