George Smith

IST 718

Week 3 Lab

**Objective**

• This exercise provides an opportunity to demonstrate our ability to combine data sets and

produce meaningful analysis. Specifically, we would like to provide a decision maker with more

than just data – we want to provide insights, understanding, and wisdom. This exercise allows

the student an opportunity to demonstrate progress (or mastery) of learning objectives 1, 2, 3,

4, and 5:

1) Obtain data and understand data structures and data elements.

2) Scrub data using scripting methods, to include debugging, for data manipulation in R and other

tools.

3) Explore data using essential qualitative analysis techniques including descriptive statistics.

4) Model relationships between data using the appropriate analytical methodologies matched to

the information and the needs of clients and users.

5) Interpret the data, model, analysis, and findings. Communicate the results in a meaningful way.

**Instructions**

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 –https://github.com/2SUBDA/IST\_718

• 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?

**Data**

The below datasets were used to construct my data frame. Data was merged using the SCL\_UNITID. NA values were removed from the data set and rows that contained the value “– “were changed to 0.

<https://github.com/2SUBDA/IST_718>

<https://www.icpsr.umich.edu/web/ICPSR/studies/26801/summary>

Initial data analysis was conducted to confirm the authenticity of the data. The below visualizations were created to look for trends in the data:

Chart, box and whisker chart

Description automatically generated

Chart, scatter chart

Description automatically generated

**Regression Model**

A regression model was developed to predict potential coaching salaries based on a number of predictive variables, as a result TotalPay was used as the response variable. After multiple iterations of models the following predicting variables were considered significant :

* SCL\_HBCU
* SCL\_PRIVATE
* FED\_RATE\_SA
* GSR\_N\_SA
* GSR\_SA
* FED\_N\_SB
* FED\_RATE\_SB

**Questions**

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

A: The recommend salary for the Syracuse football coach is $2,794,031.

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

A: If were still in the Big Ten the salary would be 3,840,393.

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

A: Air Force, Army, and Navy were dropped from out data frame as the SCL\_UNITID were inconsistent across data frames.

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

A: Based on our model student athlete graudation rate has a negative effect on projected salary with a value of -1.087 e+05. I believe this makes sense as exceptional players would likely go pro before completing University, however student body graduation rates have the opposite effect on projected salary.

**How good is our model?**

A: Our model is not great as the predicting variables can predict only 51 % of the response variable, this is based on the r squared value.

**What is the single biggest impact on salary size?**

A: The single biggest impact on salary size was number of students (Student Body (SB)) in federal cohort with a coefficient of 52.85.