FBS FOOTBALL COACHES:
How much are they really worth?

What Should You Pay A Head Coach?

Kyle Umlang
Data Analyst

"College football head coaches themselves are likely not worth these hefty salaries, but when looked upon in the larger landscape of the success of an athletics program, they can be seen as the centerpiece."

- *Leigh Steinberg, Forbes Sports Money, 2017*

Executive Summary

Head football coaches' annual salaries for Division I FBS colleges can range anywhere from just shy of $1 million to north of $9 million. What should be determined when considering a head coaches salary? Is it just what they can deliver on the field, or are there other important deciding factors?

Background

As the number one contributor of revenue for all athletic programs, football coaches have always been the top dog when it comes to compensation. Tasked with recommending the salary of Syracuse's next head football coach myself, I had to find out why that was.

Using various data from the 130 NCAA DI FBS colleges, I will help determine and solidify what the salary should be for the next head football coach at Syracuse University.

Exploratory Data Analysis

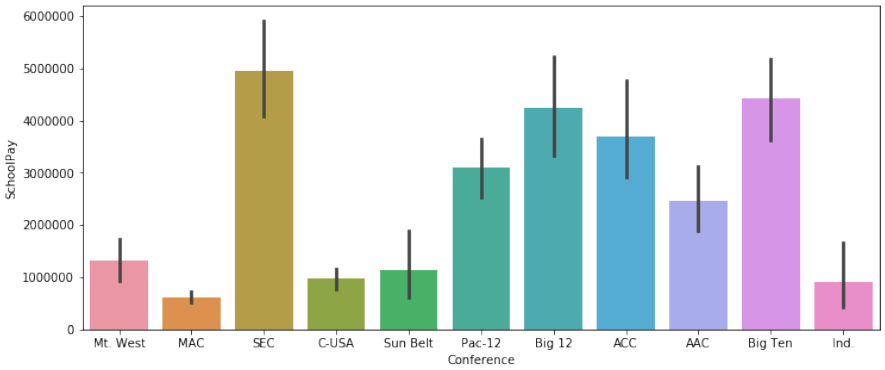
Right off the bat, I found there was missing coaching salary data for Private Schools, Independent Schools and those that do not release that information (Naval and Military Academies) so I had to remove them from the dataset.

My updated data frame now consists of 123 rows of data with 28 columns, all with valid salary data.

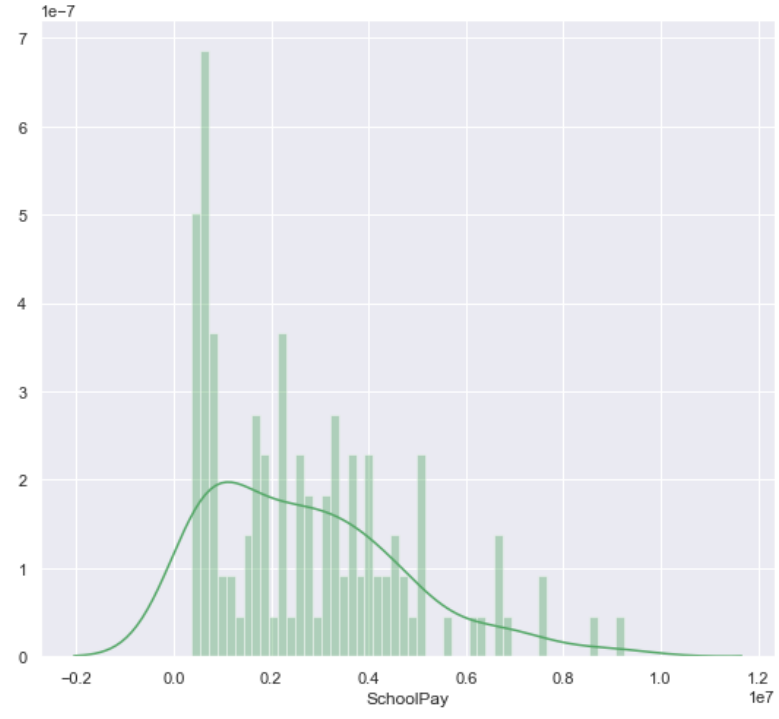
The P5 Conferences (ACC, Big Ten, Big 12, Pac 12, SEC) pay their coaches much more than the G5 conferences.

- The highest salary is $9.26 million
- The lowest salary is $360 thousand
- The average salary is $2.7 million

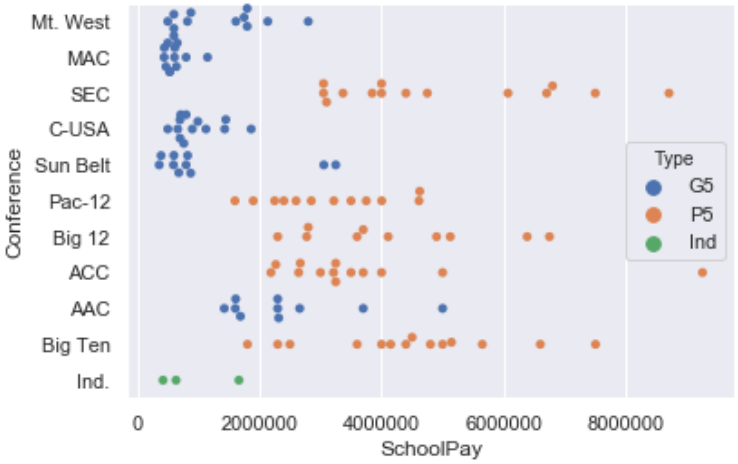
Total Salary per Conference



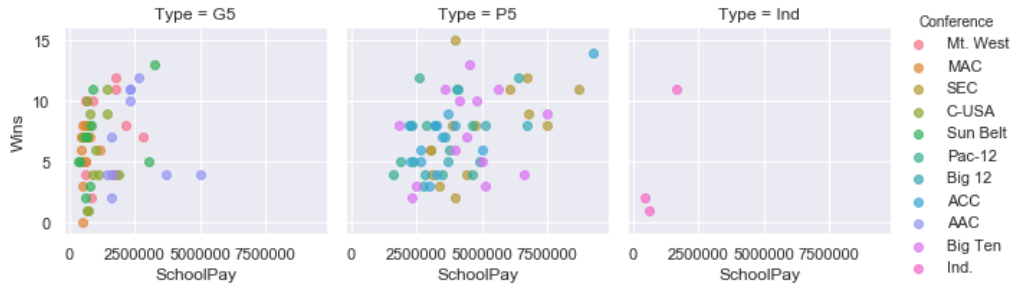
Distribution of Coaches Salary



Distribution of Coaches Salary



School Pay vs. Wins Per Conference



Correlation

There are 8 strongly correlated values:

- Capacity 0.775571
- Buyout 0.767884
- Ticket Sales 0.738704
- Assistant Pay 0.724128
- Rights/Licensing 0.707415
- Total Revenue 0.706850
- Contributions 0.681954
- Other Income 0.639930

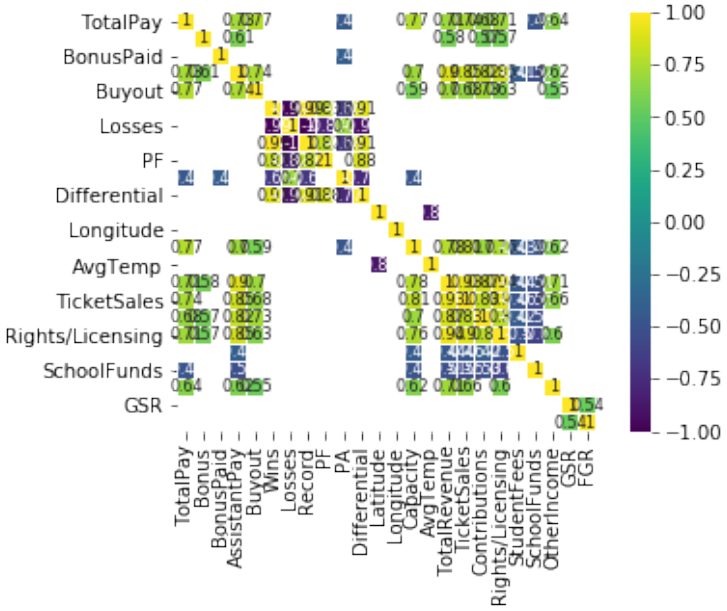
It was too difficult to interpret feature-to-feature relationships of the other attributes using distributions, so I used a heatmap to showcase any other relationships, which helped me reconfigure the dataset by removing columns that mean the same thing, which allows me to work with a smaller set of variations, leading to a theoretically higher accuracy.

- Point Differential has a positive correlation with Points For (88%)
- Point Differential has a negative correlation with Points Against (-70%)

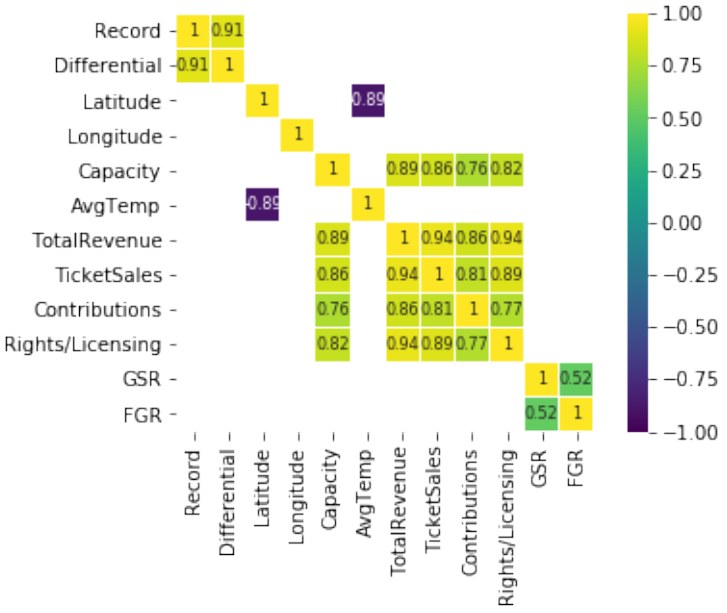
After removing similar columns, as well as schools who didn't report revenue, I'm left with 109 rows of data with 16 columns and a cleaner correlation heatmap. After tidying up some incorrect data types as well, I'm left with my remaining variables that I want to build a regression model from

- Record (W/(W+L))
- Differential (PF-PA)
- Capacity
- AvgTemp
- TotalRevenue
- TicketSales
- Contributions
- Rights/Licensing
- GSR
- FGR

Feature to Feature Relationships



Updated Feature to Feature Relationships

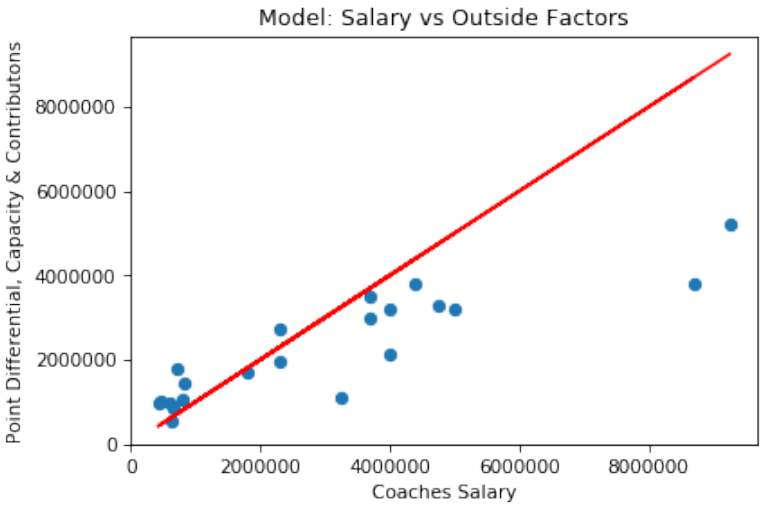


Regression Model

After splitting my data into training and testing data, I built my model and got the following results

- p-value
  - Capacity and Contributions appear to be the most statistically significant
- R squared 0.776
- Adjusted R squared 0.746

I dropped the insignificant variables and built a new model. My R squared is still great and I don't have a single variable with a p-value higher than 0.05 so I will plot my findings below:



Looking at the plot, it appears that these 3 variables can really predict a coach's salary in schools in DI FBS.

Coaches Salary = -0.00002184 + 35.3756 x (Capacity) + 0.0575 x (Contributions)

First regression model

Dep. Variable:	SchoolPay	**R-squared:**	0.776
Model:	OLS	**Adj. R-squared:**	0.746
Method:	Least Squares	**F-statistic:**	25.93
Date:	Wed, 22 Jul 2020	**Prob (F-statistic):**	1.81e-20
Time:	20:44:57	**Log-Likelihood:**	-1302.7
No. Observations:	86	**AIC:**	2627.
Df Residuals:	75	**BIC:**	2654.
Df Model:	10		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	6.14e+05	2.07e+06	0.296	0.768	-3.51e+06	4.74e+06
Record	-1.807e+06	1.25e+06	-1.448	0.152	-4.29e+06	6.79e+05
Differential	2660.9611	1902.287	1.399	0.166	-1128.588	6450.510
Capacity	23.0855	10.179	2.268	0.026	2.809	43.362
AvgTemp	-4707.0391	1.68e+04	-0.280	0.780	-3.82e+04	2.88e+04
TotalRevenue	0.0117	0.010	1.201	0.234	-0.008	0.031
TicketSales	-0.0087	0.021	-0.423	0.673	-0.050	0.032
Contributions	0.0351	0.016	2.146	0.035	0.003	0.068
Rights/Licensing	0.0046	0.011	0.418	0.677	-0.017	0.027
GSR	9679.4799	1.75e+04	0.552	0.582	-2.52e+04	4.46e+04
FGR	-5505.2977	1.14e+04	-0.483	0.630	-2.82e+04	1.72e+04

Omnibus:	21.272	**Durbin-Watson:**	2.280
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	29.782
Skew:	1.103	**Prob(JB):**	3.41e-07
Kurtosis:	4.855	**Cond. No.**	2.22e+09

Second regression model

Dep. Variable:	SchoolPay	**R-squared:**	0.748
Model:	OLS	**Adj. R-squared:**	0.742
Method:	Least Squares	**F-statistic:**	123.0
Date:	Wed, 22 Jul 2020	**Prob (F-statistic):**	1.52e-25
Time:	20:49:27	**Log-Likelihood:**	-1307.7
No. Observations:	86	**AIC:**	2621.
Df Residuals:	83	**BIC:**	2629.
Df Model:	2		
Covariance Type:	nonrobust		

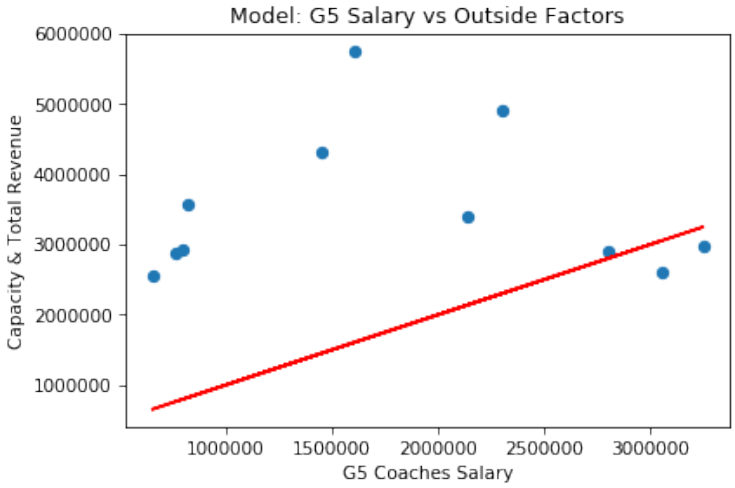
	coef	std err	t	P>|t|	[0.025	0.975]
const	-2.184e+05	2.76e+05	-0.790	0.432	-7.68e+05	3.31e+05
Capacity	35.3756	7.138	4.956	0.000	21.178	49.573
Contributions	0.0575	0.011	5.186	0.000	0.035	0.080

Omnibus:	16.865	**Durbin-Watson:**	2.423
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	20.573
Skew:	0.971	**Prob(JB):**	3.41e-05
Kurtosis:	4.402	**Cond. No.**	5.98e+07

I decided to redo the model for G5 schools only, After splitting my data into training and testing data, I built my model and got the following results

- p-value
  - Capacity and TotalRevenue appear to be the most statistically significant
- R squared 0.474
- Adjusted R squared 0.309

I dropped the insignificant variables and built a new model. My R squared is still a little low but I don't have a single variable with a p-value higher than 0.05 so I will plot my findings below:



Looking at the plot, it appears that these 2 variables can somewhat predict a coach's salary in schools in DI FBS.

Coaches Salary = -0.000001089 + 23.6407 x (Capacity) + 0.0351 x (TotalRevenue)

First G5 only regression model

Dep. Variable:	SchoolPay	**R-squared:**	0.474
Model:	OLS	**Adj. R-squared:**	0.309
Method:	Least Squares	**F-statistic:**	2.878
Date:	Sat, 25 Jul 2020	**Prob (F-statistic):**	0.0112
Time:	11:45:12	**Log-Likelihood:**	-637.19
No. Observations:	43	**AIC:**	1296.
Df Residuals:	32	**BIC:**	1316.
Df Model:	10		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-1.441e+06	2.28e+06	-0.633	0.532	-6.08e+06	3.2e+06
Record	-7.415e+05	1.3e+06	-0.568	0.574	-3.4e+06	1.92e+06
Differential	911.5291	2284.812	0.399	0.693	-3742.480	5565.539
Capacity	21.4749	11.761	1.826	0.077	-2.481	45.431
AvgTemp	9526.1740	1.68e+04	0.568	0.574	-2.46e+04	4.37e+04
TotalRevenue	0.0248	0.016	1.562	0.128	-0.008	0.057
TicketSales	-0.0378	0.099	-0.380	0.706	-0.241	0.165
Contributions	0.0575	0.066	0.876	0.388	-0.076	0.191
Rights/Licensing	0.0247	0.048	0.515	0.610	-0.073	0.122
GSR	5311.4272	1.96e+04	0.271	0.788	-3.46e+04	4.52e+04
FGR	-3882.2602	8716.036	-0.445	0.659	-2.16e+04	1.39e+04

Omnibus:	34.982	**Durbin-Watson:**	1.827
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	122.025
Skew:	1.915	**Prob(JB):**	3.18e-27
Kurtosis:	10.310	**Cond. No.**	8.22e+08

Second G5 only regression model

Dep. Variable:	SchoolPay	**R-squared:**	0.413
Model:	OLS	**Adj. R-squared:**	0.383
Method:	Least Squares	**F-statistic:**	14.05
Date:	Sat, 25 Jul 2020	**Prob (F-statistic):**	2.39e-05
Time:	11:51:54	**Log-Likelihood:**	-639.54
No. Observations:	43	**AIC:**	1285.
Df Residuals:	40	**BIC:**	1290.
Df Model:	2		
Covariance Type:	nonrobust		

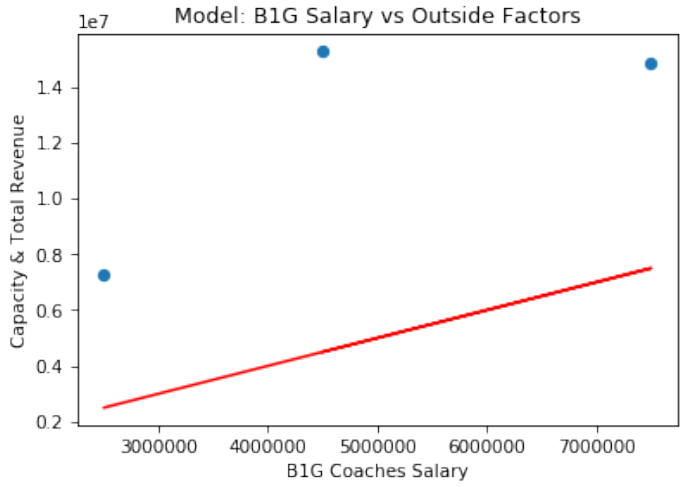
	coef	std err	t	P>|t|	[0.025	0.975]
const	-1.089e+06	4.33e+05	-2.516	0.016	-1.96e+06	-2.14e+05
Capacity	23.6407	9.606	2.461	0.018	4.225	43.056
TotalRevenue	0.0351	0.012	2.951	0.005	0.011	0.059

Omnibus:	32.203	**Durbin-Watson:**	1.564
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	102.432
Skew:	1.771	**Prob(JB):**	5.72e-23
Kurtosis:	9.680	**Cond. No.**	1.57e+08

I decided to redo the model for Big Ten schools only to see if Conference specific data would affect the model. Before starting this specific one, I removed TicketSales, Contributions and Rights/Licensing from the variables as I could not find that data for Syracuse and therefore would not be able to use them when calculating head coach salary at the end. After splitting my data into training and testing data, I built my model and got the following results

- p-value
  - Capacity and Total Revenue appear to be slightly significant
- R squared 0.918
- Adjusted R squared 0.6333

I dropped the insignificant variables and built a new model. My R squared dropped but I don't have a single variable with a p-value higher than 0.05 so I will plot my findings below:



Looking at the plot, it appears that these 2 variables can somewhat predict a coach's salary in schools in DI FBS.

Coaches Salary = -0.000004248 + 93.3897 x (Capacity) - 0.0495 x (Total Revenue)

First Big Ten only regression model

Dep. Variable:	SchoolPay	**R-squared:**	0.918
Model:	OLS	**Adj. R-squared:**	0.633
Method:	Least Squares	**F-statistic:**	3.214
Date:	Sat, 25 Jul 2020	**Prob (F-statistic):**	0.258
Time:	15:00:03	**Log-Likelihood:**	-143.28
No. Observations:	10	**AIC:**	302.6
Df Residuals:	2	**BIC:**	305.0
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	5.249e+07	1.5e+07	3.507	0.073	-1.19e+07	1.17e+08
Record	6.731e+06	7.35e+06	0.915	0.457	-2.49e+07	3.84e+07
Differential	-1958.7736	8075.929	-0.243	0.831	-3.67e+04	3.28e+04
Capacity	133.8281	40.228	3.327	0.080	-39.257	306.913
AvgTemp	-1.635e+05	1.69e+05	-0.967	0.436	-8.91e+05	5.64e+05
TotalRevenue	-0.1366	0.045	-3.054	0.093	-0.329	0.056
GSR	-8.983e+05	3.93e+05	-2.287	0.149	-2.59e+06	7.92e+05
FGR	6.183e+05	3.04e+05	2.031	0.179	-6.91e+05	1.93e+06

Omnibus:	0.154	**Durbin-Watson:**	1.515
Prob(Omnibus):	0.926	**Jarque-Bera (JB):**	0.101
Skew:	0.108	**Prob(JB):**	0.951
Kurtosis:	2.558	**Cond. No.**	7.19e+09

Second Big Ten only regression model

Dep. Variable:	SchoolPay	**R-squared:**	0.403
Model:	OLS	**Adj. R-squared:**	0.232
Method:	Least Squares	**F-statistic:**	2.360
Date:	Sat, 25 Jul 2020	**Prob (F-statistic):**	0.165
Time:	16:12:06	**Log-Likelihood:**	-152.88
No. Observations:	10	**AIC:**	311.8
Df Residuals:	7	**BIC:**	312.7
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	4.248e+06	2.96e+06	1.434	0.195	-2.76e+06	1.13e+07
Capacity	93.3897	48.554	1.923	0.096	-21.422	208.201
TotalRevenue	-0.0495	0.041	-1.201	0.269	-0.147	0.048

Omnibus:	0.567	**Durbin-Watson:**	2.387
Prob(Omnibus):	0.753	**Jarque-Bera (JB):**	0.042
Skew:	0.141	**Prob(JB):**	0.979
Kurtosis:	2.856	**Cond. No.**	9.84e+08

Conclusions

Based on my analysis and regression models, the new head football coach at Syracuse should be compensated $7.48M.

After further analysis and making a new G5 only model to see if the compensation would change if Syracuse was still in the Big East Conference, I found that if that was the case, I would be recommending a compensation of $1.49M.

After further analysis and making a new Big Ten only model to see if the compensation would change if Syracuse was in the Big Ten Conference, I found that if that was the case, I would be recommending a compensation of $4.14M.

I'm most confident in the first model that involves the majority of schools. I do not think conference specific models tell us as much as those involving all the schools.

The single biggest impact on salary size appears to be stadium capacity. Total Revenue and Contributions are also important, but capacity shows to be significant in every model.