Introduction

The following analysis was designed to assist the Syracuse University athletic director in understanding the key metrics that can be used to evaluate the total pay for Dino Barbers as head coach of the football club. Although Mr. Barbers recently signed a new contract with Syracuse, it is still of value to develop an understanding of where this football club is headed and at what cost to the university.

While 2018, was a phenomenal year for the football club, 2019 left a lot to be desired. Syracuse finished the 2018 season ranked 15th in the nation. 2019 was the polar opposite. The school finished with a losing record and dropped down to 71st despite being in one of the weakest conferences, the ACC.

But is it possible to build a linear model that can help to predict the total pay a coach should receive? The remainder of this analysis will attempt to answer the question.

About the Data

The primary dataset included coaches' salaries from the NCAA FBS Division 1 school. In addition, the dataset was supplemented with the following additional datasets:

- FBS Coaching Record (2019): Coach's previous season's record and all-time win-loss percentage. The number of years the coach has coached the team. Source: Supplied by the Athletic Department.
- Football Power Index (2019): The college football team ranking in 5 groups one to 5. With 1 being considered one of the top 20 schools last season and group 5 being the worse. Source: http://www.espn.com/college-football/statistics/teamratings
- Graduation School Rate (2017): While NCAA football is a big-money making venture, the majority of the students participating in the games are in school for a great education. Coaches ideally should help to drive that education. Source: https://web3.ncaa.org/aprsearch/gsrsearch
- School Stadium (2019): The school's home stadium size. https://en.wikipedia.org/wiki/List_of_NCAA_Division_I_FBS_football_stadiums
- NFL Players (2019): The number of current NFL players the school has produced. Source:https://www.ncaa.com/news/football/article/2019-09-03/colleges-most-represented-2019-nfl-rosters

Of note, the dataset's key joining fields, school or conference did not always align. Therefore the data was initially matched up using Microsoft Excel. Given the small size of the data (130 records) Excel was a good choice, if the dataset were larger, other methods such as Fuzzy Matching may have been required.

Once the data was scrubbed and merged, the final features were selected with 114 teams set for evaluation.

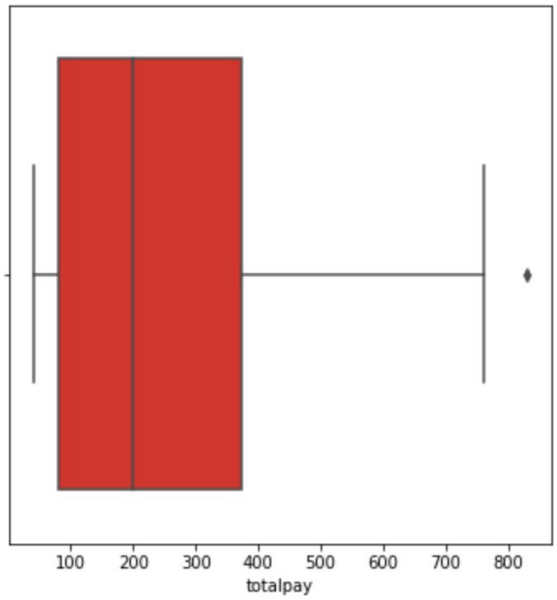
Total Pay

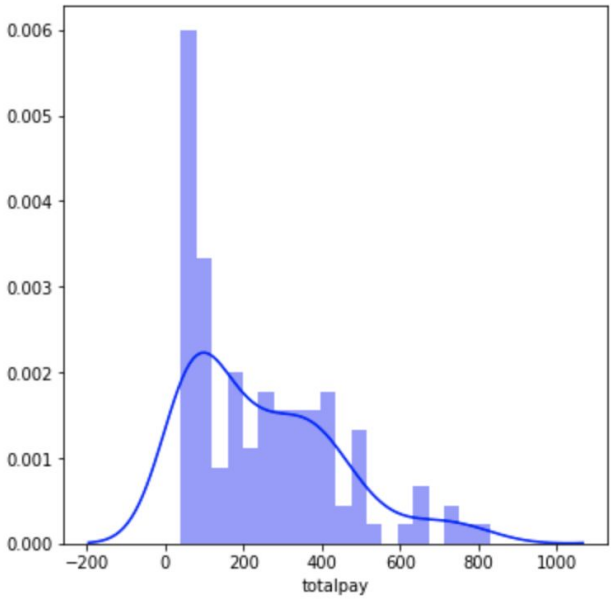
Before delving into the relationships of each feature with total pay. It would be good to understand what the average total pay per conference looks like, as illustrated below.

It's also worth noting, the number of active NFL players per conference is a strong indicator of how lucrative the coaching job is. More on that later.

Conference	Sum of TotalPay	Count of Coach	Min of TotalPay	Max of TotalPay	nfl_roster	teams	avg
SEC	$65,008,379.00	14	$2,350,000.00	$8,307,000.00	339	14	$4,643,455.64
Big Ten	$60,256,192.00	14	$1,830,000.00	$7,600,000.00	253	14	$4,304,013.71
ACC	$48,073,154.00	14	$1,831,580.00	$6,543,350.00	215	11	$4,370,286.73
Big 12	$36,163,301.00	10	$1,701,109.00	$5,500,000.00	131	10	$3,616,330.10
Pac-12	$34,681,433.00	12	$1,500,000.00	$4,377,500.00	189	12	$2,890,119.42
AAC	$16,562,677.00	11	$1,000,000.00	$2,600,000.00	100	14	$1,183,048.36
Mt. West	$12,071,254.00	12	$486,504.00	$1,800,000.00	65	12	$1,005,937.83
C-USA	$10,921,091.00	14	$500,000.00	$1,425,000.00	59	14	$780,077.93
MAC	$6,959,534.00	12	$412,500.00	$1,125,000.00	46	12	$579,961.17
Sun Belt	6506500	10	$390,000.00	$850,000.00	21	10	$650,650.00
Ind.	$4,929,080.00	6	$419,640.00	$2,129,638.00	35	6	$821,513.33

With the exception of a few notable programs, the majority of coaches make close to 2 million per year. It should be expected that going into negotiations, the baseline for an experienced coach within the Power 5 (first 5 schools listed above) conferences will start there.

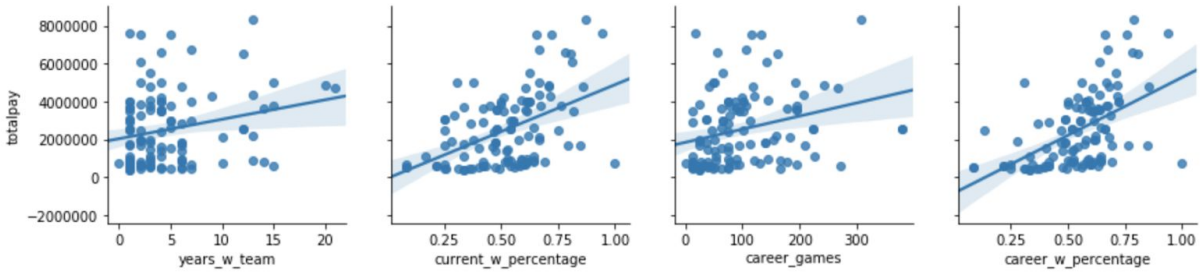




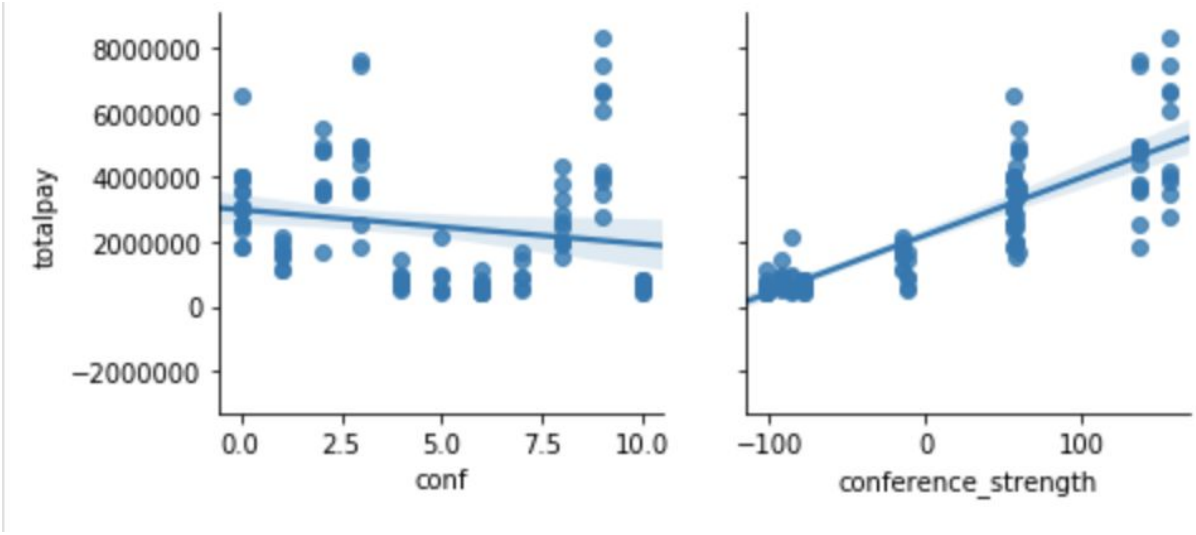
Relationships

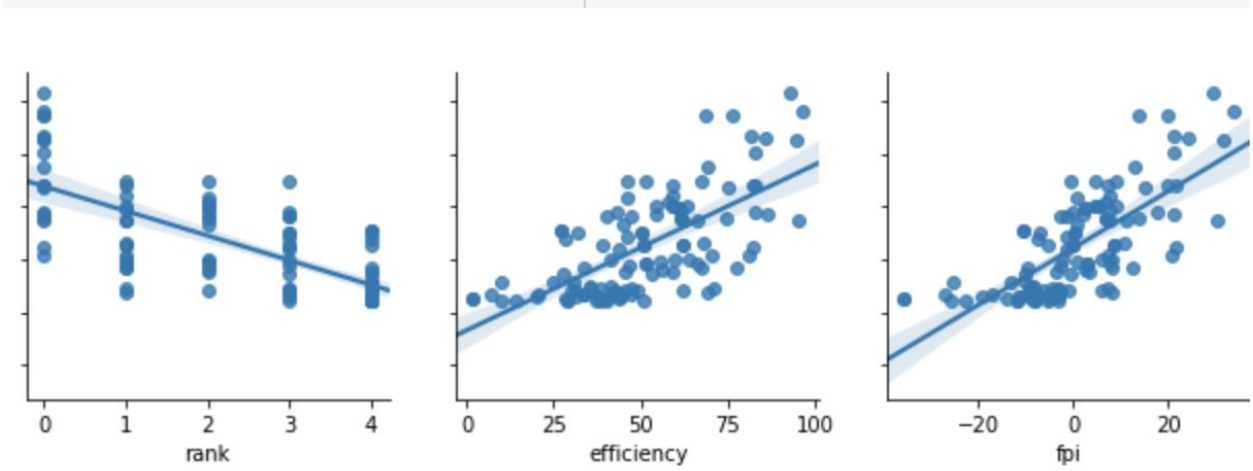
In order to identify the final features to be used for linear modeling. The dataset was analyzed to determine which features had a relationship with the dependent variable total pay.

Coaches Record:
It appears there is a relationship between two of the features within a coach's record and total pay. The two that jump out are current_w_percentage and career_w_percentage. Both account for the win percentages both from last season and over the course of a coach's career.

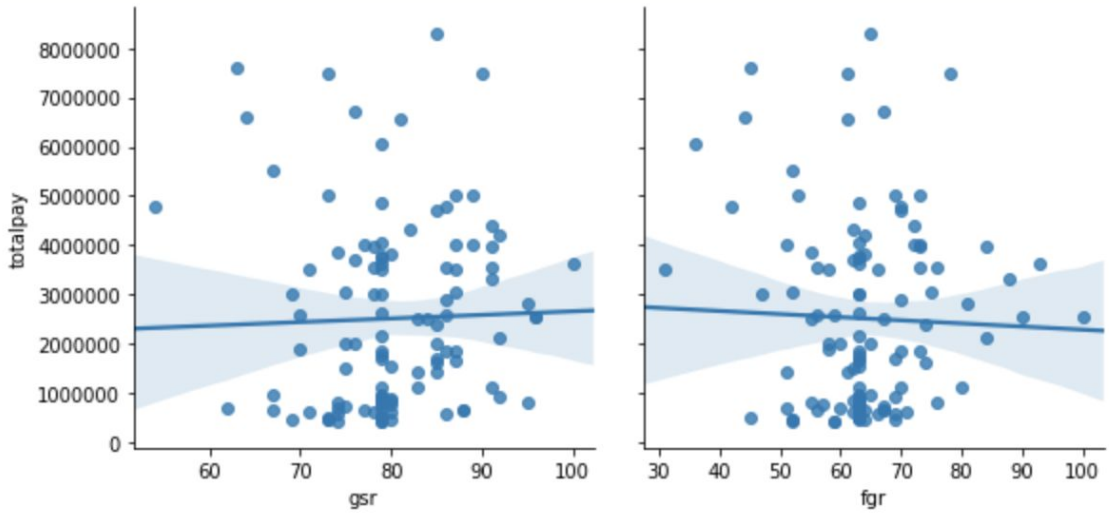


Teams Performance:
How the team performs is a strong indicator of total pay for the coach. All 5 values here should be looked at in the modeling phase.

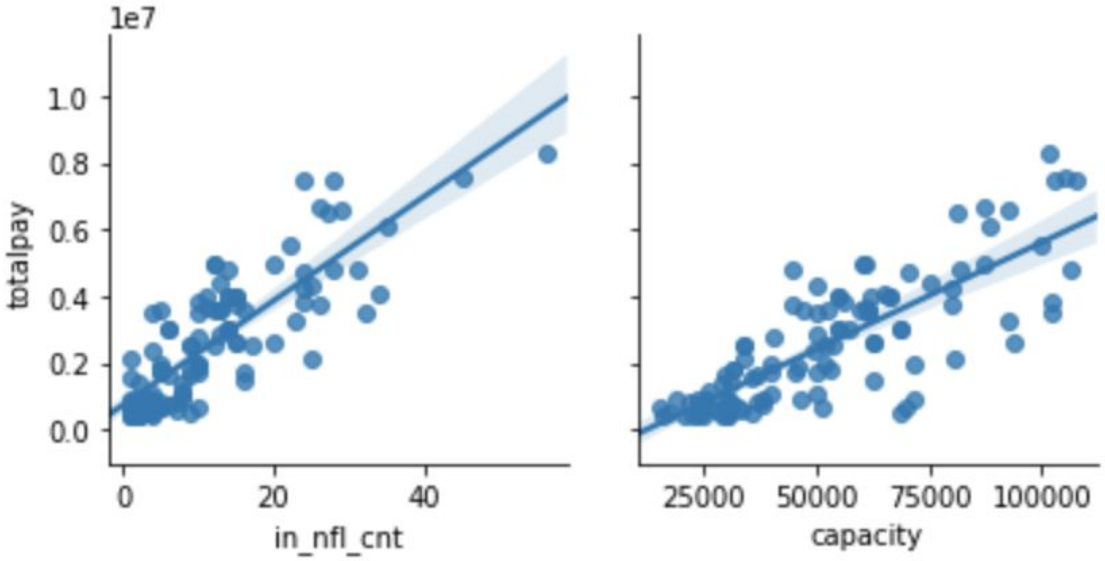




Graduation Rate:
Based upon the plot there does not appear to be a strong relationship between graduation rate and a head coach's pay. Not exactly surprising given the head coach's primary job is to win football games.



Stadium Size and Number of Current NFL Players:
Both Stadium and the number of current NFL players point towards both the popularity of the school and the number of players that will go on to become even more popular within the NFL. It would stand to reason, if you are training and producing NFL caliber players, you will win games and attract large-number of paying fans to watch live and on a device.



Analysis

3 different models were used for the regression analysis, Ordinary Least Squares Regression, Mixed Linear Regression, and Quantile Regression.

Model 1 - Ordinary Least Squares Regression (all features): To determine the statistically significant features all features were included within the model. Of the features selected, 5 turned out to be statistically significant. Where the significance is defined as having a p-score of five-percent or below. Many of the features associated with the coach's record, career wins and a career winning percentage did not turn out to be as significant as the features associated with the team's performance such as FPI, efficiency, and rank. All three team performance features penalized coaches with low to negative scores. Another key driver was the number of players currently in the NFL. Lastly stadium capacity proved to be significant.

OLS Regression Results

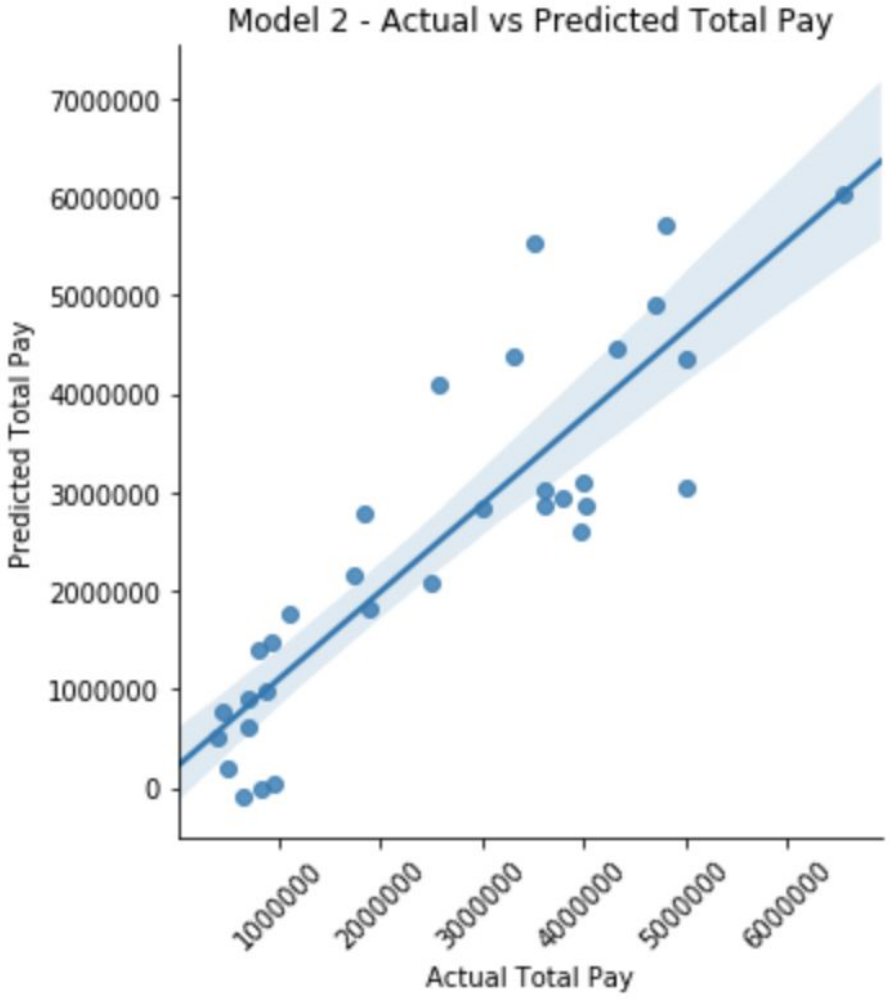
Dep. Variable:	totalpay	R-squared:	0.837
Model:	OLS	Adj. R-squared:	0.809
Method:	Least Squares	F-statistic:	29.60
Date:	Sun, 26 Jan 2020	Prob (F-statistic):	1.60e-22
Time:	08:24:32	Log-Likelihood:	-1230.7
No. Observations:	82	AIC:	2487.
Df Residuals:	69	BIC:	2519.
Df Model:	12		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	6.472e+06	2.43e+06	2.668	0.009	1.63e+06	1.13e+07
gsr	1945.3467	2.13e+04	0.091	0.928	-4.06e+04	4.45e+04
fgr	1641.3927	1.59e+04	0.104	0.918	-3e+04	3.33e+04
years_w_team	6.416e+04	3.46e+04	1.855	0.068	-4842.149	1.33e+05
current_w_percentage	-3.557e+06	1.58e+06	-2.254	0.027	-6.71e+06	-4.08e+05
career_games	-3058.4440	1943.754	-1.573	0.120	-6936.127	819.239
in_nfl_cnt	8.874e+04	1.93e+04	4.608	0.000	5.03e+04	1.27e+05
career_w_percentage	2.278e+06	1.5e+06	1.516	0.134	-7.19e+05	5.28e+06
conf	-5.715e+04	3.33e+04	-1.717	0.090	-1.24e+05	9236.953
rank	-4.822e+05	1.77e+05	-2.723	0.008	-8.36e+05	-1.29e+05
efficiency	-8.039e+04	3.04e+04	-2.640	0.010	-1.41e+05	-1.96e+04
fpi	1.31e+05	4.81e+04	2.725	0.008	3.51e+04	2.27e+05
capacity	15.6125	8.324	1.876	0.065	-0.994	32.219

The model returned an R-Squared Adjusted value of 80% in training and 80% with the test dataset. This translates to 80% of the model's variance can be attributed to the features included in the model. The next model will carry over Rank, Efficiency, FPI, Capacity, and in_nfl_cnt. We will also include conference encoded feature.

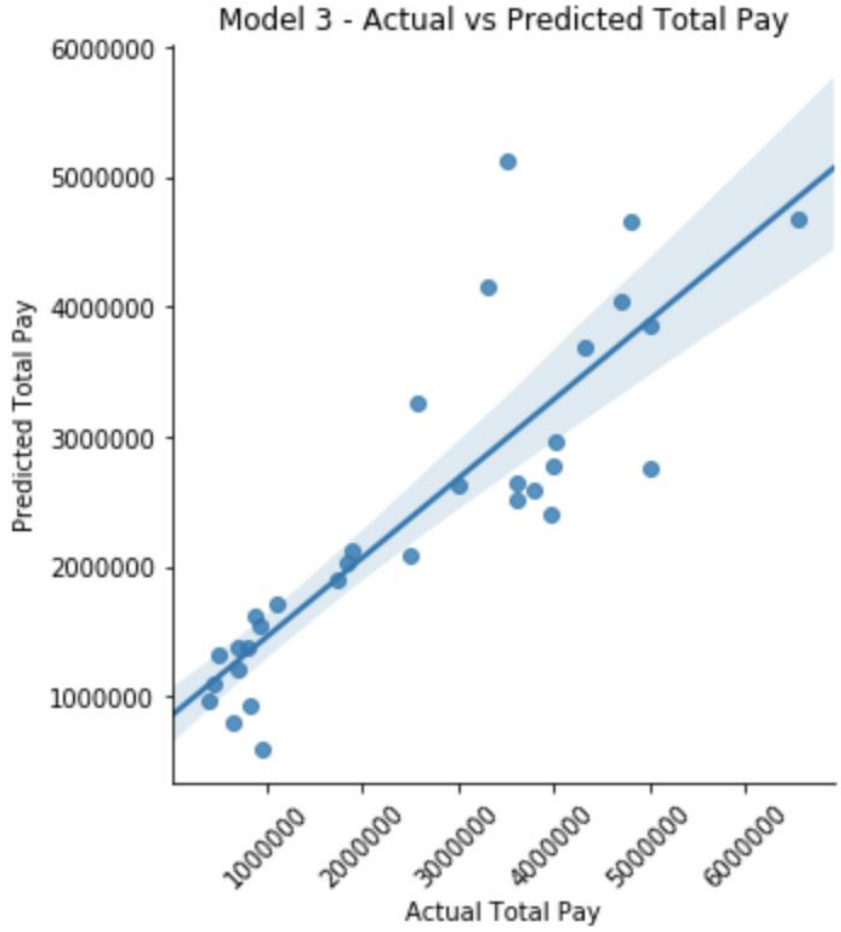
Model 2 - Ordinary Least Squares Regression(all statistically significant features): While only using statistically significant features, the model's R-squared value dropped. This is to be expected, as having more features is a known way to trick the R-squared value in improving. The underlying RMSE slightly dropped for model 2, with the MAE remaining the same. In terms of prediction, model 1 was closer to the actual total pay. This can be explained but the level of "noise" included in model by the less significant features.

```
Model 2 R-squared:  0.7512790359847417
Model 2 RMSE:  854346.291109091
Model 2 MAE:  690290.407867036
```



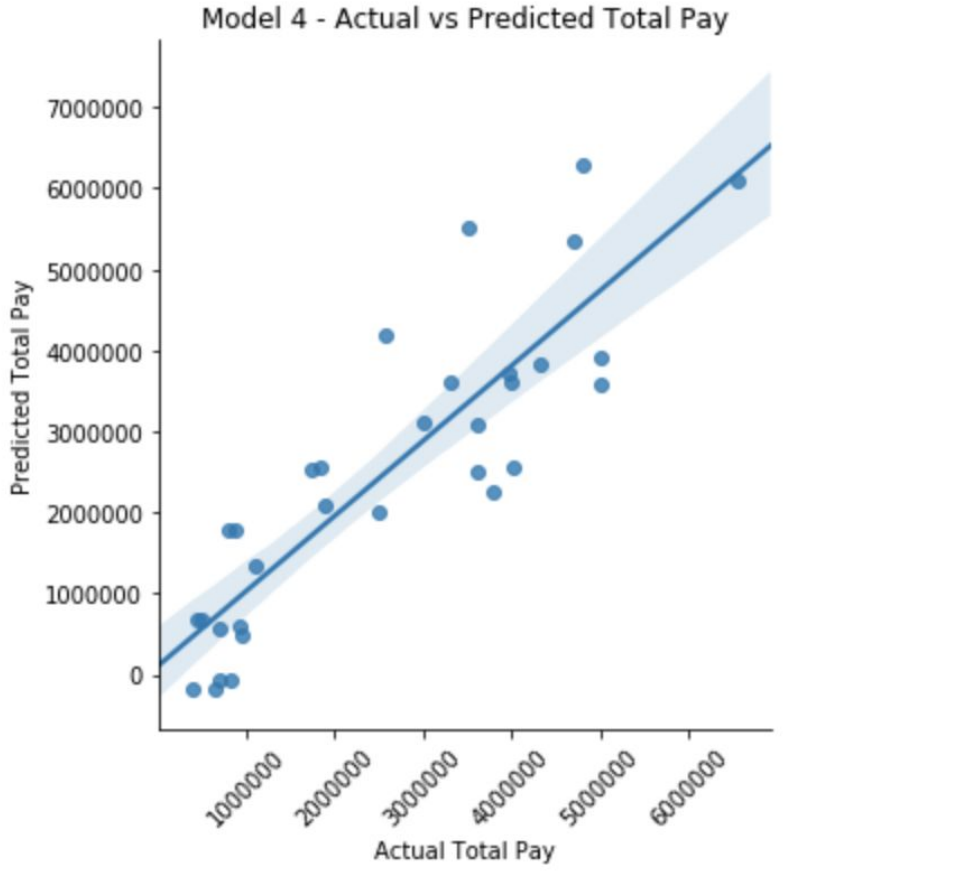
Model 3 - Mixed Model Linear Model: (all statistically significant features grouped with the conference): Of the first 3 models, this model performed the worse. Again the drop in R-squared can be explained by removing a feature. However, the MAE has gone up 774k. This means that our model is off by 774k when making predictions.

```
Model 3 R-squared:  0.7052233570397024
Model 3 RMSE:  930088.5041794796
Model 3 MAE:  774661.5131787339
```



Model 4 -Quantile Regression: (all statistically significant): Similar to mixed-model regression, quantile regression did not perform as well as oridinary least squares.

```
Model 4 R-squared:  0.730257950952954
Model 4 RMSE:  889717.3678647808
Model 4 MAE:  737894.6294061621
```



Of the 4 models, model 2 was chosen to complete the remaining predictions.

Conclusion

2019 was not a particularly strong year for Syracuse's football club. As such, key coefficients such as rank, efficiency, and FPI penalized Dino Barbers when predicting total pay. Put another way, for the 2020 season, if Barbers is able to improve the team's efficiency, FPI score and therefore move up in rank, his predicted total pay will improve. Based upon the recent staff changes, most notably firing his defensive coach, clearly, Barber understands this.

What's really nice about using the team predictors to evaluate the coach's total pay, is these metrics can be tracked and predicted weekly on a game by game basis. Armed with this data, Syracuse's athletic director can set-up KPI's that can be tied into Coach Barber's next contract.

Still, the linear model can be improved by including both historical data and looking at additional metrics such as offensive and defensive efficiency. In addition to tracking the number of current NFL players, another metric may be to track which key team members are injured or unable to play (academic, graduation or NFL draft) weekly.

Lastly, It is important to note this analysis represents a snap-shot and it does not provide a complete picture of what happened in 2019. For example, perhaps key team members from the 2018 run were injured, graduated or left the school.

Questions:
What is the recommended salary for the Syracuse football coach?

Suggested Salary for Syracuse Head Coach is $2.22 million with the actual salary $2.4 million. 200k is a significant drop in value. However, this can be explained by each of the team performance coefficients. For example as the team's FPI improves, it will have a positive impact on Coach Barber's Salary.

```
Estimated Effect of all features on Total Pay:  Intercept     7810833.0
in_nfl_cnt     101436.0
conf           -86750.0
rank          -521724.0
efficiency     -96009.0
fpi            142183.0
dtype: float64
```

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

Suggested Salary for Syracuse Head Coach is $2.22 million with the actual salary $2.4 million. This can be due to overfitting or many of the same teams in the Big East are not in the same conference as Syrcuse.

What if we went to the Big Ten?

Using the quantile regression model the results were the same for both ACC and Big Ten: 2,396,002.37. This suggest the model may be overfitting.

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

These schools were dropped either because they did not have total pay, their career or current win percentage was 0.

Akron
Coastal Carolina
Louisiana,ÄìMonroe
SMU
Baylor
Rice
BYU

How good is our model?

Based on the Adjusted R-square and MAE values, the model is stastically significant. But it be improved using historical data. For example, testing Buffalo, which is in the MAC conference. The model predicated 536k and the coach's pay is 455k. This is a difference of 80k, however 536k would place the coach's pay closer to the median within the conference.

What is the single biggest impact on salary size?

The team's rank represented the biggest impact on the salary independent of a conference. As mentioned previously, rank impacted salary negatively. That said, this year, Syracuse's FPI was extremely low. Had this been the 2018 season, ironically the season Dino Barber received his contract, Syracuse's FPI was much higher