Introduction

There is a slight change of temperature. Just a bit cooler. The leaves seem to be changing. The days may still be warm, even hot, but the nights seem chilly now. You know what that means? It means its fall. And fall means College Football! Don't you just love College Football?! The excitement, the pagentry, the wondering "will we make it to a bowl game this season?" And then the big games, powerhouse football, fun to attend and watch on television with family and friends, and beer. Wow I hear those coaches are making some serious money these days. As a matter of fact, we're looking for a new coach for my team. I wonder who we'll hire? I wonder how we'll figure out how much to pay him. Wouldn't it be fun and informative to look at the process? Who would you hire?

To select that "GOAT" coach is fun and brings a dilemma. Can reviews be trusted? Are collective opinions valued? "Total Pay" and "Wins/Losses" will be evaluated in the following analysis to arrive at the coaching salary and provide confidence in your coach salary offer.

Analysis and Models

About the Data

Football is big business. The top 25 teams brought in $3B in revenues last year. Of this $812M was in donations with another $1.2B came from license fees. The average donation is $32M for the top 25 and the license fees are $49M. When compared to the spreadsheet provided, and knowing that top conferences such as the SEC have their own cable television networks it easy to see at face value how coaching salaries can grow faster than most other markets when revenue sharing is included.

When looking for a model to capture a salary for a head coach we begin with what data is available on both salaries and context. For this purpose we began with a dataset provided.

An image is shown below:

David Primrose, IST736 Section 37929

A	B	C	D	E	F	G	H	I
School	Conference	Coach	SchoolPay	TotalPay	Bonus	BonusPaid	AssistantP	Buyout
Air Force	Mt. West	Troy Calhoun	885000	885000	247000	--	$0	--
Akron	MAC	Terry Bowden	$411,000	$412,500	$225,000	$50,000	$0	$688,500
Alabama	SEC	Nick Saban	$8,307,000	$8,307,000	$1,100,000	$500,000	$0	$33,600,000
Alabama at Birr	C-USA	Bill Clark	$900,000	$900,000	$950,000	$165,471	$0	$3,847,500
Appalachian Sta	Sun Belt	Scott Satterfield	$712,500	$712,500	$295,000	$145,000	$0	$2,160,417
Arizona	Pac-12	Kevin Sumlin	$1,600,000	$2,000,000	$2,025,000	--	$0	$10,000,000
Arizona State	Pac-12	Herm Edwards	$2,000,000	$2,000,000	$3,010,000	--	$0	$8,166,667
Arkansas	SEC	Chad Morris	$3,500,000	$3,500,000	$1,000,000	--	$0	$12,500,000
Arkansas State	Sun Belt	Blake Anderson	$825,000	$825,000	$185,000	$25,000	$0	$300,000
Army	Ind.	Jeff Monken	932521	932521	--	--	$0	--
Auburn	SEC	Gus Malzahn	$6,700,000	$6,705,656	$1,400,000	$375,000	$0	$32,143,750
Ball State	MAC	Mike Neu	$435,689	$435,689	$380,000	$30,000	$0	$980,300
Baylor	Big 12	Matt Rhule	--	--	--	--	$0	--
Boise State	Mt. West	Bryan Harsin	$1,650,010	$1,650,010	$475,000	$145,000	$0	$7,784,038
Boston College	ACC	Steve Addazio	$2,514,859	$2,514,859	--	--	$0	--
Bowling Green	MAC	Mike Jinks	$437,228	$437,228	$245,000	$81,250	$0	$874,456
Brigham Young	Ind.	Kalani Sitake	--	--	--	--	$0	--
Buffalo	MAC	Lance Leipold	$455,500	$455,500	$381,000	$0	$0	$1,020,833
California	Pac-12	Justin Wilcox	$1,500,000	$1,500,000	$900,000	$75,000	$0	$7,208,500
Central Florida	AAC	Josh Heupel	$1,700,000	$1,700,000	$250,000	--	$0	$3,587,500
Central Michiga	MAC	John Bonamego	$655,000	$655,000	$415,000	$45,000	$0	$1,125,000
Charlotte	C-USA	Brad Lambert	$625,000	$625,000	$120,000	$0	$0	$556,389
Cincinnati	AAC	Luke Fickell	$2,000,000	$2,000,000	$625,000	$0	$0	$7,100,000
Clemson	ACC	Dabo Swinney	$6,205,000	$6,543,350	$1,125,000	$500,000	$0	$35,000,000

From here we added other characteristics which may have impact on coaching salaries. In this analysis we added the following to create the dataframe shown below:

```
In [ ]:

In [521]: susitemdf.head()
```

Out[521]:

	School	Conference	Coach	SchoolPay	TotalPay	Bonus	AssistantPay	Buyout	ninethyr	fivyrtotal	...	win	loss	perc	Yr2019
0	Missouri	SEC	Barry Odom	2350000.0	2350000.0	2350000.0	4729500.0	2350000.0	54160.0	274472.0	...	6	6	0.500000	38471523.0
1	Louisville	ACC	Bobby Petrino	3980434.0	3980434.0	3980434.0	3820000.0	3980434.0	49913.0	249458.0	...	8	5	0.615385	47327709.0
2	Old Dominion	C-USA	Bobby Wilder	654667.0	654667.0	654667.0	1138911.0	654667.0	18234.0	98221.0	...	1	11	0.083333	11415018.0
3	Notre Dame	Ind.	Brian Kelly	2129638.0	2129638.0	2129638.0	0.0	2129638.0	76288.0	393122.0	...	11	2	0.846154	115510518.0
4	Virginia	ACC	Bronco Mendenhall	3550000.0	3550000.0	3550000.0	3400000.0	3550000.0	47863.0	210180.0	...	9	5	0.642857	33726985.0

5 rows × 25 columns

```
In [ ]:
```

Assistant Pay – this was found on a current USAToday coaching dataset and used to approximate the pay scale for coaches. In reviewing the data one can see that most coaches obviously earn more than their assistant coaches except where the coach is young, new, or both.

Win-Loss-Win Percentage record

Game attendance by school for 2019

Average game attendance for the past five years

Sports revenue for the college

Total enrollment (Dept of Ed)

Pre-season ranking for 2019 (from year ending 2018)

Post-season ranking for 2019 (from year ending 2019)

Associated Press rankings for the top teams for 2018 and 2918 (zero's were entered for other teams)

When assembled we reviewed statistical data and also to clean and adjust:

```
print(susitemdf.describe())
```

```
          SchoolPay      TotalPay         Bonus   AssistantPay        Buyout  \
count  5.000000e+01  5.000000e+01  5.000000e+01  5.000000e+01  5.000000e+01
mean   2.870548e+06  2.887210e+06  2.887210e+06  2.662350e+06  2.887210e+06
std    2.143305e+06  2.149499e+06  2.149499e+06  2.294234e+06  2.149499e+06
min    4.110000e+05  4.125000e+05  4.125000e+05  0.000000e+00  4.125000e+05
25%    1.106250e+06  1.106250e+06  1.106250e+06  1.001687e+06  1.106250e+06
50%    2.275000e+06  2.275000e+06  2.275000e+06  1.795438e+06  2.275000e+06
75%    3.995108e+06  4.005308e+06  4.005308e+06  4.212450e+06  4.005308e+06
max    8.307000e+06  8.307000e+06  8.307000e+06  7.541277e+06  8.307000e+06

            ninethyr      fivyrtotal      FiveYrAvg         PerCap       GSR  \
count      50.000000       50.000000      50.000000      50.000000  50.00000
mean    44963.720000   225088.520000   45017.640000   47069.980000  69.10000
std     28238.261971   141702.220581   28340.411627    6409.431607  11.53035
min      9924.000000    45874.000000    9175.000000   38426.000000  41.00000
25%     20638.750000   102361.750000   20472.000000   43073.000000  62.50000
50%     40388.000000   205168.500000   41033.500000   45180.000000  67.00000
75%     54125.250000   279732.500000   55946.250000   50416.750000  75.00000
max    111459.000000   554421.000000  110884.000000   67322.000000  97.00000

       ...        win       loss       perc        Yr2019        Enroll  \
count  ...  50.000000  50.000000  50.000000  5.000000e+01     50.000000
mean   ...   7.460000   5.440000   0.566304  4.459869e+07  30559.880000
std    ...   3.387914   2.696634   0.233631  3.529649e+07  14513.719917
min    ...   0.000000   1.000000   0.000000  6.901879e+06   4682.000000
25%    ...   5.000000   3.000000   0.416667  1.287620e+07  21219.500000
50%    ...   8.000000   5.000000   0.615385  4.084752e+07  28204.000000
75%    ...  11.000000   7.000000   0.785714  6.314056e+07  38642.250000
max    ...  14.000000  12.000000   0.933333  1.230737e+08  75756.000000

          Rank18   APRank18       Rank19   APRank19   runiform
count  50.000000  50.000000    50.000000  50.000000  50.000000
```

:

With this information a total correlation was created:

```python
# check correlations
correlation = susitemdf.loc[:, [ 'School', 'Conference','Coach', 'SchoolPay', 'TotalPay', 'Bonus','AssistantPay', 'Buyout',\
        'ninethyr', 'fivyrtotal', 'FiveYrAvg', 'PerCap', 'GSR', 'FSR', 'Capa', 'win', 'loss', 'perc', 'Yr2019',\
        'Enroll', 'Rank18', 'APRank18', 'Rank19',  'APRank19']]
print(correlation.corr())
```

```
              SchoolPay  TotalPay     Bonus  AssistantPay    Buyout  ninethyr  \
SchoolPay      1.000000  0.999418  0.999418      0.765315  0.999418  0.921124
TotalPay       0.999418  1.000000  1.000000      0.769794  1.000000  0.921503
Bonus          0.999418  1.000000  1.000000      0.769794  1.000000  0.921503
AssistantPay   0.765315  0.769794  0.769794      1.000000  0.769794  0.729549
Buyout         0.999418  1.000000  1.000000      0.769794  1.000000  0.921503
ninethyr       0.921124  0.921503  0.921503      0.729549  0.921503  1.000000
fivyrtotal     0.914006  0.915258  0.915258      0.740240  0.915258  0.994271
FiveYrAvg      0.914006  0.915258  0.915258      0.740241  0.915258  0.994271
PerCap        -0.220647 -0.226146 -0.226146     -0.243233 -0.226146 -0.208870
GSR            0.315314  0.313331  0.313331     -0.026357  0.313331  0.300558
FSR            0.197600  0.197761  0.197761     -0.103634  0.197761  0.191974
Capa           0.903040  0.904761  0.904761      0.722792  0.904761  0.979863
win            0.580457  0.581662  0.581662      0.483294  0.581662  0.622958
loss          -0.607344 -0.607478 -0.607478     -0.490052 -0.607478 -0.659386
perc           0.585603  0.585542  0.585542      0.474774  0.585542  0.635869
Yr2019         0.861943  0.860396  0.860396      0.634472  0.860396  0.943190
Enroll         0.310585  0.310812  0.310812      0.422427  0.310812  0.336475
Rank18        -0.741768 -0.741587 -0.741587     -0.583363 -0.741587 -0.793178
APRank18       0.226107  0.222209  0.222209      0.061893  0.222209  0.290813
Rank19        -0.751767 -0.750343 -0.750343     -0.590705 -0.750343 -0.783814
APRank19       0.490363  0.485931  0.485931      0.269689  0.485931  0.510280

              fivyrtotal  FiveYrAvg    PerCap       GSR  ...      Capa  \
SchoolPay       0.914006   0.914006 -0.220647  0.315314  ...  0.903040
```

A regression analysis was run for all variables

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               TotalPay   R-squared:                       0.999
Model:                            OLS   Adj. R-squared:                  0.998
Method:                 Least Squares   F-statistic:                     1207.
Date:                Sat, 17 Oct 2020   Prob (F-statistic):           6.90e-23
Time:                        17:24:43   Log-Likelihood:                -447.91
No. Observations:                  36   AIC:                             933.8
Df Residuals:                      17   BIC:                             963.9
Df Model:                          18
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
Intercept     1.659e+06   1.75e+06      0.951      0.355   -2.02e+06    5.34e+06
SchoolPay        0.9831      0.027     36.498      0.000       0.926       1.040
AssistantPay     0.0089      0.022      0.404      0.691      -0.037       0.055
ninethyr        -5.7808      8.175     -0.707      0.489     -23.029      11.468
fivyrtotal    8796.5777   1.87e+04      0.471      0.643   -3.06e+04    4.82e+04
FiveYrAvg    -4.398e+04   9.33e+04     -0.471      0.643   -2.41e+05    1.53e+05
PerCap          -1.3916      3.750     -0.371      0.715      -9.303       6.520
GSR            841.0070   4088.254      0.206      0.839   -7784.456    9466.470
FSR            827.9468   4364.384      0.190      0.852   -8380.097       1e+04
Capa             5.1051      6.419      0.795      0.437      -8.438      18.648
win            1.04e+05   5.28e+04      1.969      0.065   -7425.145    2.15e+05
loss         -1.506e+05   1.34e+05     -1.125      0.276   -4.33e+05    1.32e+05
perc         -3.124e+06   1.93e+06     -1.618      0.124    -7.2e+06    9.49e+05
Yr2019          -0.0015      0.002     -0.927      0.367      -0.005       0.002
Enroll           0.4039      1.462      0.276      0.786      -2.681       3.489
Rank18         -27.9783   1053.915     -0.027      0.979   -2251.545    2195.588
APRank18     -1648.5816   3516.878     -0.469      0.645   -9068.545    5771.382
Rank19         754.1283   1746.332      0.432      0.671   -2930.311    4438.568
APRank19      -567.9285   4390.562     -0.129      0.899   -9831.205    8695.348
```

Looking at the results and running multiple times we narrowed the predictor variables to the following:

```
# refining the model and trying a few combinations. This one came out better even though I was adding in something I had
# previously taken out
np.random.seed(1234)
susitemdf['runiform'] = uniform.rvs(loc = 0, scale = 1, size = len(susitemdf))
susitemdf_train = susitemdf[susitemdf['runiform'] >= 0.33]
susitemdf_test = susitemdf[susitemdf['runiform'] < 0.33]
# check training data frame
#print('\SusItems_train data frame (rows, columns): ',susitemdf_train.shape)
#print(susitemdf_train.head())
# check test data frame
#print('\SusItems_test data frame (rows, columns): ',susitemdf_test.shape)
#print(susitemdf_test.head())

# specify a simple model with bobblehead entered last
my_model = str('TotalPay ~  Bonus + Capa + win')

# fit the model to the training set
train_model_fit = smf.ols(my_model, data = susitemdf_train).fit()
# summary of model fit to the training set
print(train_model_fit.summary())
```

With the following results:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               TotalPay   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 7.998e+30
Date:                Sat, 17 Oct 2020   Prob (F-statistic):               0.00
Time:                        16:36:00   Log-Likelihood:                 661.54
No. Observations:                  36   AIC:                            -1315.
Df Residuals:                      32   BIC:                            -1309.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept   1.164e-10   1.35e-09      0.086      0.932   -2.64e-09    2.87e-09
Bonus          1.0000   4.95e-16   2.02e+15      0.000       1.000       1.000
Capa        4.086e-14    4.2e-14      0.973      0.338   -4.47e-14    1.26e-13
win        -1.019e-10   1.72e-10     -0.593      0.557   -4.52e-10    2.48e-10
==============================================================================
Omnibus:                        4.833   Durbin-Watson:                   0.468
Prob(Omnibus):                  0.089   Jarque-Bera (JB):                4.329
Skew:                          -0.846   Prob(JB):                        0.115
Kurtosis:                       2.854   Cond. No.                     1.14e+07
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.14e+07. This might indicate that there are
strong multicollinearity or other numerical problems.
```

While the R-squared appears to over influenced by the bonus the results for a coaching salary are as follows:

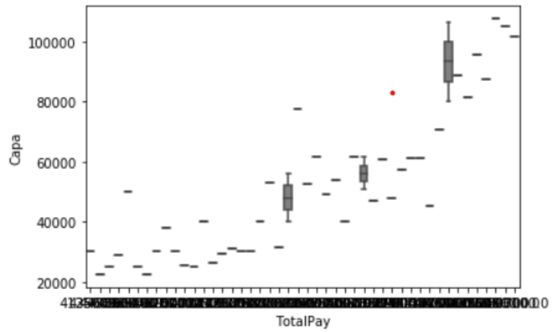
Total Pay = 1.164e-10 + 1.0 * bonus + 4.086e -14 * Capacity -1,019e-10 * win

David Primrose, IST736 Section 37929

Visual approach:

Looking at Total Pay vs Capacity

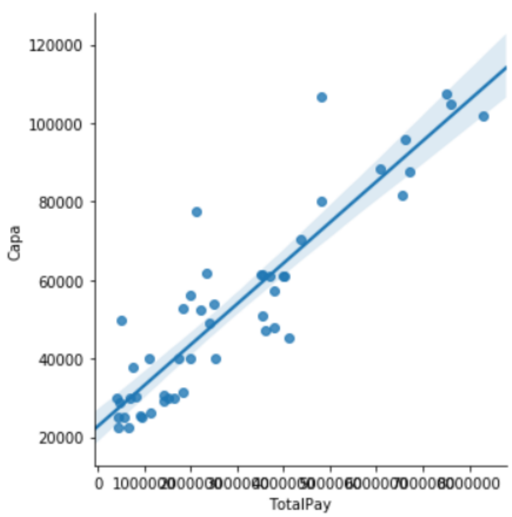
```
# Going back to correlations to capture a boxplot.  Obviously bonus is completely correlated with TotalPay so trying the othe
# Trying with Capa (capacity)
sns.boxplot(x="TotalPay", y="Capa", data=susitemdf, color = "gray");
plt.show()
```



Another view via seaborn

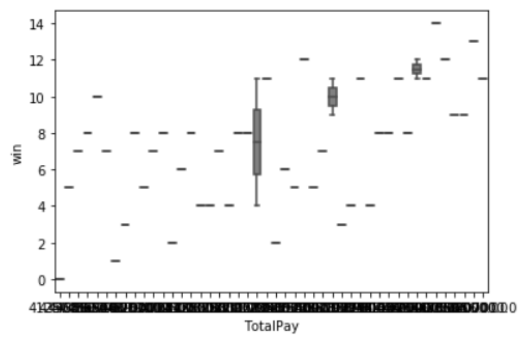
```
#Total Pay vs Capacity
sns.lmplot(x='TotalPay', y='Capa', data=susitemdf)
```

```
]: <seaborn.axisgrid.FacetGrid at 0x1c8d1d98630>
```



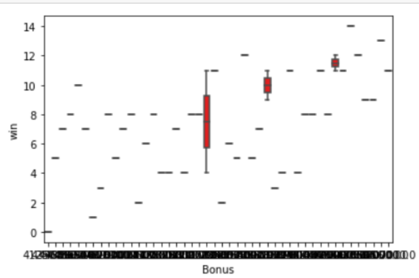
Also looking at Total Pay vs. wins:

```
# Going back to correlations to capture a boxplot.  Obviously bonus is completely correlated with TotalPay so trying the othe
# Trying with win (# of wins)
sns.boxplot(x="TotalPay", y="win", data=susitemdf, color = "gray");
plt.show()
```



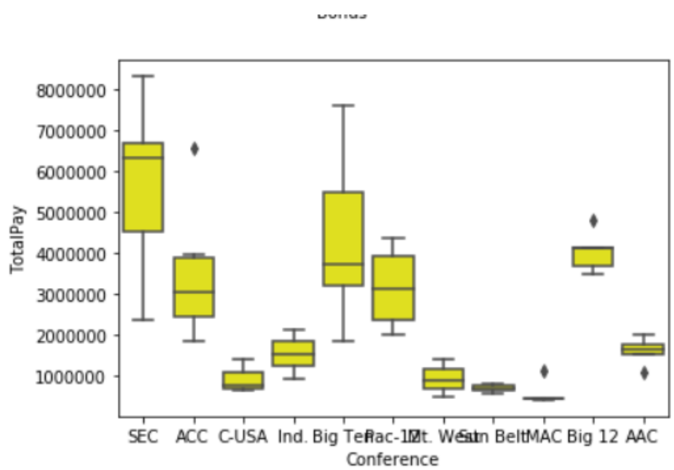
Bonus vs Win:

```
np.mean(susitemdf['SchoolPay'])
np.min(susitemdf['TotalPay'])
np.max(susitemdf['AssistantPay'])
np.mean(susitemdf['ninethyr'])
np.min(susitemdf['Capa'])
np.max(susitemdf['Enroll'])

sns.boxplot(x='Bonus', y="win", data=susitemdf, color = "red")
plt.show()

# Total Pay by Conference
sns.boxplot(x="Conference", y="TotalPay", data=susitemdf, color = "yellow");
plt.show()

# School vs TotalPay shows nothing
sns.boxplot(x="School", y="TotalPay", data=susitemdf, color = "gray")
plt.show()
```



David Primrose, IST736 Section 37929

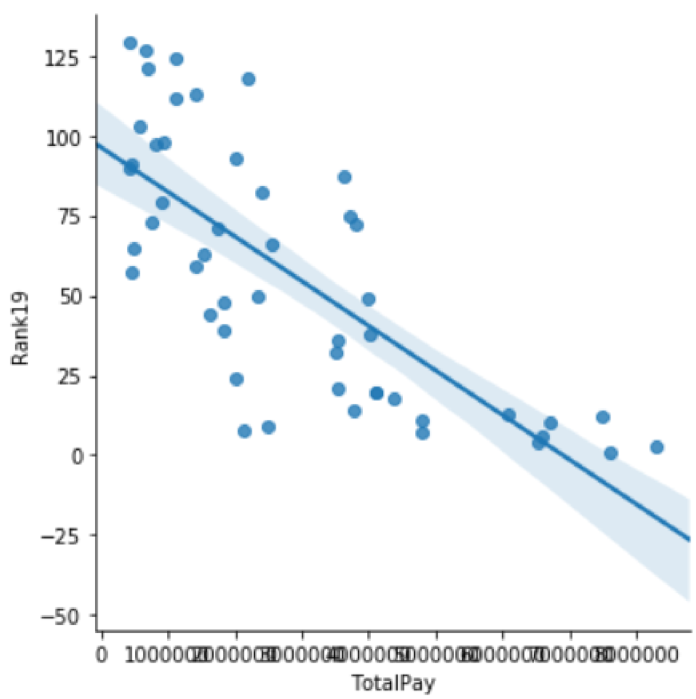
Conference vs Total Pay:



Total Pay vs Final Team Rankings after the 2019 season

```
sns.lmplot(x='TotalPay', y='Rank19', data=susitemdf)
```

<seaborn.axisgrid.FacetGrid at 0x1c8cc33e748>



Results

Results show the effects of bonus, capacity, and wins on coaching salaries. These appear to be the most reliable indicators of coaching salaries. Interestingly rankings while appearing initially to be a contributing factor actually showed negative correlation. The p-values were not significant for this exercise. However using the data the salary for a new coach would be

Total Pay = 1.164e-10 + 1.0 * bonus + 4.086e -14 * Capacity -1,019e-10 * win

Further analysis is necessary. If this formula were true we would need to offer a prospective candidate a very large bonus for them to achieve the appropriate level of compensation.
I would suggest that an total pay of an amount in excess of $4M to account for the bonus requirement as well as the limited impact of capacity of stadium and coefficient of wins.

Conclusion.

These are unprecedented times. A grand reopening of American society is occurring in real time. This includes some of our favorite football teams playing in a limited season. While this will have anticipated impact downward on coaching salaries due to lack of upside potential the future still looks bright. Post our current distress there will be future student populations who enjoy football and future alums who are willing to donate money to football and college programs. But the capacity, wins and bonus structures are still the future of coaching.