Marley Akonnor
IST 718
LAB 1

1) What is the recommended salary for the Syracuse football coach?

When Syracuse University was isolated and a regression was run to predict the salary for the head coach of Syracuse, the dollar amount was estimated to be **$2,512,146**.

```
1  #Syracuse Coach Salary
2  syr_sal = coaches_comp[coaches_comp['School'] == 'Syracuse']
3  regr.predict(syr_sal[['BonusPaid', 'Capacity', 'Revenue', 'Expenses', 'gsr']])
4
5  pred_syr_sal = regr.predict(syr_sal[['BonusPaid', 'Capacity', 'Revenue', 'Expenses', 'gs
6
7
8  # print amount
9  print("Syracuse Coach Pay should be $",(pred_syr_sal))
```

Syracuse Coach Pay should be $ [2512146.28444074]

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

I admittedly had difficulty conceptualizing how to go about this pair of questions. However, I will give my best guesses. I believe if he were to be in the Big East his salary would be **$2,806,487** and if he were to migrate to the Big 10 his salary would be **$3,345,206**. In any scenario, this gentleman is going to have a good day.

```
1  big10_sal = coaches_comp[coaches_comp['Conference'] == 'Big Ten']
2  regr.predict(big10_sal[['BonusPaid', 'Capacity', 'Revenue', 'Expenses', 'gsr']])
3
4  pred_big10_sal = regr.predict(big12_sal[['BonusPaid', 'Capacity', 'Revenue', 'Expenses',
5
6
7  # print amount
8  print("Big 10 coach Pay should be $",(pred_big10_sal))
```

Big 10 coach Pay should be $ [3345206.14525936 3280972.52211641 4928729.47937244 2904503.13
512768
 6402087.36025005 4242064.77713212 3329395.5712759 4389872.90866691
 6984837.20114298 5757930.42580036 3315684.1929491 2806487.65777642
 4966363.53920055]

3) What schools did we drop from our data and why?

This decision was made by first assessing which columns has NAN values to be removed. The following columns were returned: Capacity (15 entries), Rank (20 entries), Revenue (20), Expenses (20), state (5), gsr (5) and, fsr (7).

Marley Akonnor
IST 718
LAB 1

```
1  #Check for null values in the dataframe
2  null_check = coaches_comp.isnull().values.any()
3  tot_null = coaches_comp.isnull().sum()
4  print(null_check)
5  print(tot_null)
```

```
True
School         0
Conference     0
Coach          0
SchoolPay      0
TotalPay       0
BonusPaid      0
Capacity      15
Rank          20
Revenue       20
Expenses      20
state          5
gsr            5
fgr            7
dtype: int64
```

The image below shows universities where values for total pay are missing and were later removed.

```
1  #Colleges missing TotalPay
2  coaches_comp.loc[coaches_comp['TotalPay'] == 0]
```

	School	Conference	Coach	SchoolPay	TotalPay	BonusPaid	Capacity	Rank	Revenue	Expenses	state	gsr	
13	Baylor	Big 12	Matt Rhule	0.0	0.0	0.0	45140.0	NaN	NaN	NaN	TX	90.0	6
17	Brigham Young	Ind.	Kalani Sitake	0.0	0.0	0.0	NaN	NaN	NaN	NaN	UT	64.0	4:
92	Rice	C-USA	Mike Bloomgren	0.0	0.0	0.0	47000.0	NaN	NaN	NaN	TX	87.0	8
100	Southern Methodist	AAC	Sonny Dykes	0.0	0.0	0.0	NaN	NaN	NaN	NaN	TX	72.0	6

Marley Akonnor
IST 718
LAB 1

The image below shows universities where values for stadium capacity are missing and were later removed.

```
1  #Colleges missing Stadium Capacity
2  coaches_comp[coaches_comp['Capacity'].isna()]
```

	School	Conference	Coach	SchoolPay	TotalPay	BonusPaid	Capacity	Rank	Revenue	Expenses	state
3	Alabama at Birmingham	C-USA	Bill Clark	900000.0	900000.0	165471.0	NaN	86.0	36494381.0	35697117.0	AL
17	Brigham Young	Ind.	Kalani Sitake	0.0	0.0	0.0	NaN	NaN	NaN	NaN	UT
22	Charlotte	C-USA	Brad Lambert	625000.0	625000.0	0.0	NaN	76.0	40866430.0	38454232.0	NC
51	Liberty	Ind.	Turner Gill	947281.0	947281.0	0.0	NaN	NaN	NaN	NaN	VA
59	Massachusetts	Ind.	Mark Whipple	500000.0	500000.0	20000.0	NaN	67.0	49461013.0	48445234.0	MA
61	Miami (Fla.)	ACC	Mark Richt	4058061.0	4058061.0	0.0	NaN	NaN	NaN	NaN	FL
62	Miami (Ohio)	MAC	Chuck Martin	524826.0	524826.0	12669.0	NaN	79.0	39644248.0	38650630.0	OH
73	Nevada-Las Vegas	Mt. West	Tony Sanchez	600000.0	600000.0	0.0	NaN	64.0	50784275.0	50445250.0	NV
77	North Carolina State	ACC	Dave Doeren	3000000.0	3000000.0	275000.0	NaN	47.0	92724548.0	90100025.0	NC
99	Southern California	Pac-12	Clay Helton	2625965.0	2625965.0	0.0	NaN	NaN	NaN	NaN	CA
100	Southern Methodist	AAC	Sonny Dykes	0.0	0.0	0.0	NaN	NaN	NaN	NaN	TX
101	Southern Mississippi	C-USA	Jay Hopson	500000.0	500000.0	15000.0	NaN	125.0	25687189.0	25687189.0	MS
107	Texas Christian	Big 12	Gary Patterson	4840717.0	4840717.0	0.0	NaN	NaN	NaN	NaN	TX
110	Texas-El Paso	C-USA	Dana Dimel	700000.0	700000.0	0.0	NaN	99.0	33481654.0	32980563.0	TX
111	Texas-San Antonio	C-USA	Frank Wilson	1100000.0	1100000.0	17500.0	NaN	97.0	33657739.0	32278508.0	TX

Marley Akonnor
IST 718
LAB 1

The image below shows universities where values for revenue are missing and were later removed.

```
1  #Colleges missing Revenue
2  coaches_comp[coaches_comp['Revenue'].isna()]
```

	School	Conference	Coach	SchoolPay	TotalPay	BonusPaid	Capacity	Rank	Revenue	Expenses	state	
10	Army	Ind.	Jeff Monken	932521.0	932521.0	0.0	38000.0	NaN	NaN	NaN	NaN	
13	Baylor	Big 12	Matt Rhule	0.0	0.0	0.0	45140.0	NaN	NaN	NaN	TX	
15	Boston College	ACC	Steve Addazio	2514859.0	2514859.0	0.0	44500.0	NaN	NaN	NaN	MA	
17	Brigham Young	Ind.	Kalani Sitake	0.0	0.0	0.0	NaN	NaN	NaN	NaN	UT	
29	Duke	ACC	David Cutcliffe	2540928.0	2540928.0	0.0	40004.0	NaN	NaN	NaN	NC	
51	Liberty	Ind.	Turner Gill	947281.0	947281.0	0.0	NaN	NaN	NaN	NaN	VA	
61	Miami (Fla.)	ACC	Mark Richt	4058061.0	4058061.0	0.0	NaN	NaN	NaN	NaN	FL	
70	Navy	AAC	Ken Niumatalolo	2163000.0	2163000.0	0.0	34000.0	NaN	NaN	NaN	NaN	
80	Northwestern	Big Ten	Pat Fitzgerald	3619775.0	3619775.0	0.0	47130.0	NaN	NaN	NaN	IL	1
81	Notre Dame	Ind.	Brian Kelly	2129638.0	2129638.0	0.0	77622.0	NaN	NaN	NaN	IN	
90	Pittsburgh	ACC	Pat Narduzzi	3017718.0	3017718.0	0.0	65500.0	NaN	NaN	NaN	PA	
92	Rice	C-USA	Mike Bloomgren	0.0	0.0	0.0	47000.0	NaN	NaN	NaN	TX	
99	Southern California	Pac-12	Clay Helton	2625965.0	2625965.0	0.0	NaN	NaN	NaN	NaN	CA	
100	Southern Methodist	AAC	Sonny Dykes	0.0	0.0	0.0	NaN	NaN	NaN	NaN	TX	
102	Stanford	Pac-12	David Shaw	4311543.0	4311543.0	0.0	50000.0	NaN	NaN	NaN	CA	
107	Texas Christian	Big 12	Gary Patterson	4840717.0	4840717.0	0.0	NaN	NaN	NaN	NaN	TX	
114	Tulane	AAC	Willie Fritz	1629000.0	1629000.0	0.0	30000.0	NaN	NaN	NaN	LA	
115	Tulsa	AAC	Philip Montgomery	1518177.0	1518177.0	0.0	35542.0	NaN	NaN	NaN	OK	
119	Vanderbilt	SEC	Derek Mason	2812523.0	2812523.0	0.0	40550.0	NaN	NaN	NaN	TN	
122	Wake Forest	ACC	Dave Clawson	1831580.0	1831580.0	0.0	31500.0	NaN	NaN	NaN	NC	

Marley Akonnor
IST 718
LAB 1

The image below shows universities where values for expenses are missing and were later removed.

```
1  #Colleges missing Expenses
2  coaches_comp[coaches_comp['Expenses'].isna()]
```

	School	Conference	Coach	SchoolPay	TotalPay	BonusPaid	Capacity	Rank	Revenue	Expenses	state
10	Army	Ind.	Jeff Monken	932521.0	932521.0	0.0	38000.0	NaN	NaN	NaN	NaN
13	Baylor	Big 12	Matt Rhule	0.0	0.0	0.0	45140.0	NaN	NaN	NaN	TX
15	Boston College	ACC	Steve Addazio	2514859.0	2514859.0	0.0	44500.0	NaN	NaN	NaN	MA
17	Brigham Young	Ind.	Kalani Sitake	0.0	0.0	0.0	NaN	NaN	NaN	NaN	UT
29	Duke	ACC	David Cutcliffe	2540928.0	2540928.0	0.0	40004.0	NaN	NaN	NaN	NC
51	Liberty	Ind.	Turner Gill	947281.0	947281.0	0.0	NaN	NaN	NaN	NaN	VA
61	Miami (Fla.)	ACC	Mark Richt	4058061.0	4058061.0	0.0	NaN	NaN	NaN	NaN	FL
70	Navy	AAC	Ken Niumatalolo	2163000.0	2163000.0	0.0	34000.0	NaN	NaN	NaN	NaN
80	Northwestern	Big Ten	Pat Fitzgerald	3619775.0	3619775.0	0.0	47130.0	NaN	NaN	NaN	IL 1
81	Notre Dame	Ind.	Brian Kelly	2129638.0	2129638.0	0.0	77622.0	NaN	NaN	NaN	IN
90	Pittsburgh	ACC	Pat Narduzzi	3017718.0	3017718.0	0.0	65500.0	NaN	NaN	NaN	PA
92	Rice	C-USA	Mike Bloomgren	0.0	0.0	0.0	47000.0	NaN	NaN	NaN	TX
99	Southern California	Pac-12	Clay Helton	2625965.0	2625965.0	0.0	NaN	NaN	NaN	NaN	CA
100	Southern Methodist	AAC	Sonny Dykes	0.0	0.0	0.0	NaN	NaN	NaN	NaN	TX
102	Stanford	Pac-12	David Shaw	4311543.0	4311543.0	0.0	50000.0	NaN	NaN	NaN	CA
107	Texas Christian	Big 12	Gary Patterson	4840717.0	4840717.0	0.0	NaN	NaN	NaN	NaN	TX
114	Tulane	AAC	Willie Fritz	1629000.0	1629000.0	0.0	30000.0	NaN	NaN	NaN	LA
115	Tulsa	AAC	Philip Montgomery	1518177.0	1518177.0	0.0	35542.0	NaN	NaN	NaN	OK
119	Vanderbilt	SEC	Derek Mason	2812523.0	2812523.0	0.0	40550.0	NaN	NaN	NaN	TN
122	Wake Forest	ACC	Dave Clawson	1831580.0	1831580.0	0.0	31500.0	NaN	NaN	NaN	NC

Marley Akonnor
IST 718
LAB 1

The image below shows universities where values for graduation success rate are missing and were later removed.

```
1  #Colleges missing Graduate Success Rate
2  coaches_comp[coaches_comp['gsr'].isna()]
```

	School	Conference	Coach	SchoolPay	TotalPay	BonusPaid	Capacity	Rank	Revenue	Expenses	state	g
9	Arkansas State	Sun Belt	Blake Anderson	825000.0	825000.0	25000.0	30964.0	92.0	35550592.0	35550592.0	NaN	Na
10	Army	Ind.	Jeff Monken	932521.0	932521.0	0.0	38000.0	NaN	NaN	NaN	NaN	Na
26	Colorado	Pac-12	Mike MacIntyre	2878500.0	2878500.0	297000.0	53750.0	46.0	94935198.0	98413284.0	NaN	Na
54	Louisiana-Monroe	Sun Belt	Matt Viator	390000.0	390000.0	0.0	30427.0	178.0	15568952.0	16927856.0	NaN	Na
70	Navy	AAC	Ken Niumatalolo	2163000.0	2163000.0	0.0	34000.0	NaN	NaN	NaN	NaN	Na

4) What effect does graduation rate have on the projected salary?

From an initial scatterplot look at the directional relationship between the graduation success rate and school pay, it appears there is none. The dots are all very spread apart.

```
1  #Plot "gsr" against "SchoolPay" to see relationship
2  plt.scatter(coaches_comp.gsr, coaches_comp.SchoolPay,  color='orange')
3  plt.xlabel("Graduation Success Rate")
4  plt.ylabel("SchoolPay")
5  plt.show()
```

Marley Akonnor
IST 718
LAB 1
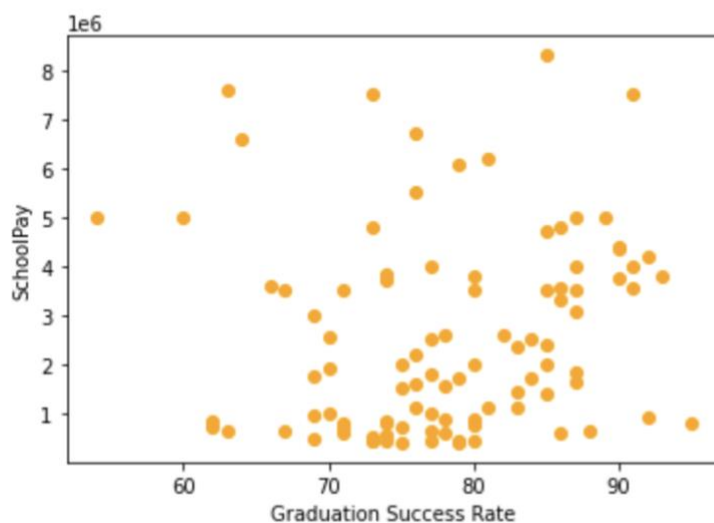
In the image below. The regression summary indicated that the graduation success rate has a positive directional influence on the school pay of a coach.

```
                        OLS Regression Results
==============================================================================
Dep. Variable:             SchoolPay   R-squared:                       0.818
Model:                           OLS   Adj. R-squared:                  0.810
Method:                Least Squares   F-statistic:                     104.6
Date:               Sun, 24 Jul 2022   Prob (F-statistic):           1.49e-33
Time:                       19:28:00   Log-Likelihood:                -1476.6
No. Observations:                 98   AIC:                             2963.
Df Residuals:                     93   BIC:                             2976.
Df Model:                          4
Covariance Type:           nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept    -1.158e+06   8.53e+05     -1.357      0.178   -2.85e+06    5.36e+05
Capacity       23.2531      8.543      2.722      0.008       6.287      40.219
Revenue         0.0183      0.011      1.630      0.107      -0.004       0.041
Expenses        0.0068      0.012      0.574      0.567      -0.017       0.030
gsr          3122.0315    1.07e+04      0.291      0.772   -1.82e+04    2.44e+04
==============================================================================
Omnibus:                       2.937   Durbin-Watson:                   1.867
Prob(Omnibus):                 0.230   Jarque-Bera (JB):                2.281
Skew:                          0.301   Prob(JB):                        0.320
Kurtosis:                      3.442   Cond. No.                     1.39e+09
==============================================================================
```

How good is our model?

The two models performed relatively well. The first model had an R-squared of ~82%. Which means the model explains about 82% of the variance in school pay when the factors of stadium capacity, revenue, expenses, and graduation success rate were included in the model.

```
                        OLS Regression Results
==============================================================================
Dep. Variable:             SchoolPay   R-squared:                       0.818
Model:                           OLS   Adj. R-squared:                  0.810
Method:                Least Squares   F-statistic:                     104.6
Date:               Sun, 24 Jul 2022   Prob (F-statistic):           1.49e-33
Time:                       19:28:00   Log-Likelihood:                -1476.6
No. Observations:                 98   AIC:                             2963.
Df Residuals:                     93   BIC:                             2976.
Df Model:                          4
Covariance Type:           nonrobust
```

The second model performed slightly better with an R-Squared of ~85% implying the combination of bonus paid, stadium capacity, revenue, expenses, and graduation success rate were able to explain ~85% of the variance in school pay for coaches.

Marley Akonnor
IST 718
LAB 1

```
1  y_pred=regr.predict(X_test)
2  mse = mean_squared_error(y_test, y_pred)
3  mae = mean_absolute_error(y_test, y_pred)
4  score=r2_score(y_test, y_pred)
5  #Display summary statistics for regression
6  print("Mean absolute error : " + str(mae))
7  print("Mean squared error : " + str(mse))
8  print("r2_score : " + str(score))
```

```
Mean absolute error : 565661.8126893027
Mean squared error : 589858918954.0946
r2_score : 0.8542541751117876
```

5) What is the single biggest impact on salary size?

Upon initial examination it appears Graduation Success Rate (gsr) has the largest impact on
salary size with a coefficient of 3122. However, when we move to the right, we see the p-value
for gsr is 0.772 which is greater than the normal alpha of 0.5. Ascending to Capacity, the p-value
is 0.008 which is smaller than 0.05 so we accept and decide Capacity has the biggest influence
on salary size with a coefficient of 23.25.

```
----------------------------------------------------------------------------
               coef      std err        t       P>|t|      [0.025     0.975]
----------------------------------------------------------------------------
Intercept   -1.158e+06   8.53e+05    -1.357     0.178     -2.85e+06   5.36e+05
Capacity      23.2531       8.543     2.722     0.008         6.287     40.219
Revenue        0.0183       0.011     1.630     0.107        -0.004      0.041
Expenses       0.0068       0.012     0.574     0.567        -0.017      0.030
gsr         3122.0315    1.07e+04     0.291     0.772     -1.82e+04   2.44e+04
----------------------------------------------------------------------------
```