

INTRO

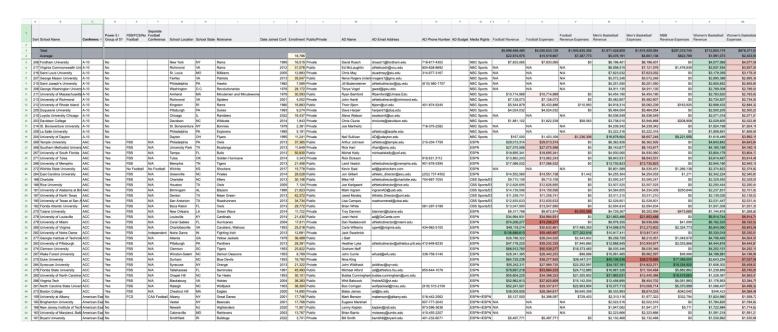
- Do off-the-court factors impact a school's men's basketball results for an individual season? If so, which factors are most important in predicting the results of a season?
- We took data from the top 8 conferences in men's college basketball and looked at a bunch of variables including revenue and enrollment





DATA SOURCE

- Compiled Data from the U.S. Department of Education Equity in Athletics Data Analysis
- Original Data was compiled for Sports Analytics Company for Research into all Division 1 College Athletics Programs
 - 50 variables, including profits for all sports
- Cut Down to 8 variables for this project



DATA PREVIEW

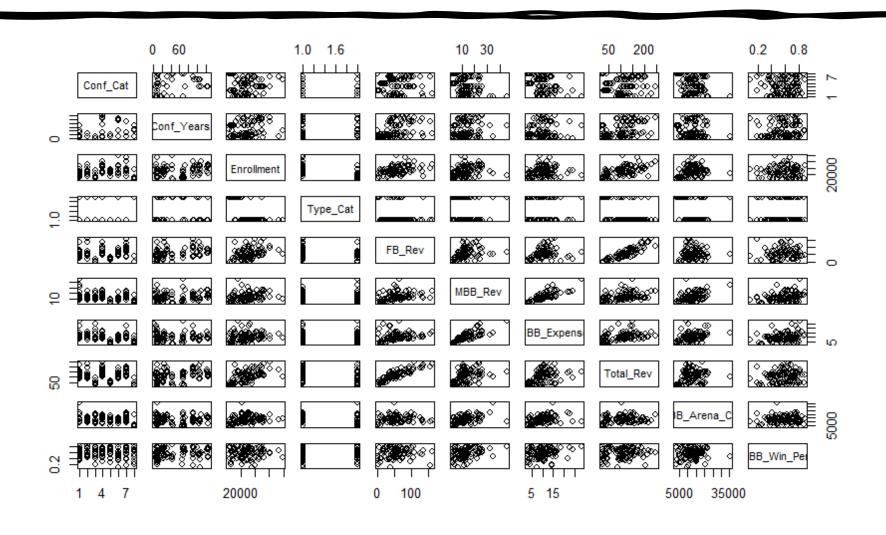
	A	В	С	D	E	F	G ∢	▶ I	J	К	L
1	School Name	Conf_Cat	Years in Confere nce	Enrollment	Type_Cat	Football Revenue	Men's Basketball Revenue	MBB Profit	Line 14: Grand Total Revenues	Men's Basketball Arena Capacity	MBB Winning %
2	Clemson University	1	69	25,822	1	\$68,912,760	\$8,035,346	\$0	\$140,436,882	9000	51.52%
3	Boston College	1	17	14,890	2	\$38,009,926	\$8,333,993	-\$340,040	\$95,703,917	8606	39.39%
4	Georgia Institute of Technology	1	43	36,489	1	\$28,788,322	\$9,258,729	\$1,048,678	\$93,696,369	8600	37.50%
5	University of Miami	1	18	17,811	2	\$74,206,277	\$9,978,355	\$41,699	\$148,497,805	7972	70.27%
6	Wake Forest University	1	69	8,789	2	\$28,341,365	\$10,061,485	\$98,888	\$84,889,487	14665	71.43%
7	University of Notre Dame	1	9	12,681	2	\$136,688,613	\$10,817,411	\$0	\$215,302,668	9149	68.57%
8	Virginia Tech	1	18	36,383	1	\$52,962,813	\$12,456,695	\$4,051,993	\$109,216,783	10052	63.89%
9	University of Pittsburgh	1	9	28,391	1	\$47,178,220	\$12,968,545	\$2,033,868	\$122,722,495	12508	34.38%
10	University of Virginia	1	69	25,018	1	\$48,118,214	\$14,598,575	\$2,324,773	\$128,298,742	14593	60.00%
11	North Carolina State University	1	69	36,304	1	\$52,241,521	\$15,377,713	\$5,370,999	\$102,387,569	19557	74.36%
12	Florida State University	1	31	45,493	1	\$78,697,218	\$16,987,326	\$5,882,862	\$162,146,012	12100	54.84%
13	University of Louisville	1	8	21,430	1	\$34,994,931	\$21,853,488	\$0	\$139,978,924	22090	40.63%
14	University of North Carolina at Chapel Hill	1	69	30,101	1	\$55,604,225	\$31,965,031	\$19,519,663	\$119,569,409	21750	74.36%
15	Syracuse University	1	9	21,322	2	\$50,242,311	\$34,157,191	\$19,334,920	\$105,631,408	35446	48.48%
16	Duke University	1	69	16,780	2	\$64,725,236	\$45,108,538	\$17,088,650	\$150,517,681	9314	82.05%
17	West Virginia University	2	10	25,474	1	\$19,787,210	\$8,371,424	-\$1,356,711	\$97,067,706	14000	48.48%
18	Kansas State University	2	26	20,229	1	\$50,528,678	\$10,163,782	\$2,510,671	\$100,822,204	12528	45.16%
19	Oklahoma State University	2	26	24,660	1	\$49,465,735	\$11,360,616	\$3,862,472	\$101,526,357	13611	50.00%
20	Texas Christian University (TCU)	2	10	11,938	2	\$71,767,270	\$12,670,472	\$2,292	\$138,998,636	8500	61.76%
21	Baylor University	2	26	20,626	2	\$48,492,390	\$13,308,914	\$0	\$111,131,098	10347	79.41%
22	University of Oklahoma	2	26	28,052	1	\$132,659,502	\$14,034,622	\$1,628,567	\$186,948,657	11528	54.29%
23	Texas Tech University	2	26	40,666	1	\$60,218,700	\$15,290,989	\$3,432,306	\$103,708,791	15300	72.97%
24	Iowa State University	2	26	30,708	1	\$55,861,324	\$15,676,249	\$8,233,209	\$86,775,037	14376	62.86%
25	University of Kansas	2	26	27,685	1	\$39,077,010	\$17,126,971	\$1,618,077	\$124,842,549	16300	85.00%
26	University of Texas at Austin	2	26	51,892	1	\$161,532,860	\$23,637,212	\$7,782,313	\$230,503,008	10763	64.71%
27	DePaul University	3	9	21,922	2	N/A	\$7,614,927	\$0	\$33,627,815	10387	48.39%

DATA & VARIABLES

- 9 independent variables: conference (Conf_Cat), years in conference (Conf_Years), school enrollment (Enrollment), private/public university (Type_Cat), football revenue (FB_Rev), men's basketball revenue (MBB_Rev), total revenue (Total_Rev), men's basketball expenses (MBB_Expense), men's basketball arena capacity (MBB_Arena_Cap)
 - oquantified conferences 1–8 and public schools (1) vs private schools (2)
- dependent variable: men's basketball winning percentage

Conferences	& Categories
ACC	1
Big 12	2
Big East	3
BIG-10	4
Mountain West	5
PAC-12	6
SEC	7
WCC	8

PAIRED DATA



ORIGINAL MODEL

```
Call:
lm(formula = MBB\_Win\_Perc \sim ., data = data)
Residuals:
    Min
              10
                  Median
                               3Q
                                       Max
-0.44357 -0.09320 0.00254 0.13578 0.29816
Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)
              4.407e-01 1.287e-01
                                    3.424 0.00103 **
Conf_Cat
              4.518e-03 1.005e-02
                                    0.450 0.65442
           2.257e-05 5.537e-04
Conf Years
                                    0.041
                                          0.96759
Enrollment
           2.555e-09 1.738e-06
                                    0.001 0.99883
            1.596e-02 5.645e-02
                                    0.283 0.77817
Type_Cat
FB Rev
             -1.022e-03 1.226e-03
                                   -0.834
                                           0.40727
                                           0.08488 .
MBB_Rev
        8.970e-03 5.133e-03
                                    1.747
MBB_Expense -3.520e-03 7.637e-03
                                   -0.461
                                           0.64628
Total_Rev
            7.857e-04 1.007e-03
                                    0.780
                                           0.43795
MBB_Arena_Cap -1.725e-06 4.732e-06
                                   -0.365 0.71650
Signif. codes: 0 '*** 0.001 '** 0.01 '* 0.05 '. '0.1 ' '1
Residual standard error: 0.1674 on 71 degrees of freedom
  (15 observations deleted due to missingness)
Multiple R-squared: 0.1151, Adjusted R-squared: 0.002878
F-statistic: 1.026 on 9 and 71 DF, p-value: 0.4284
```

• only 2 significant predictors: intercept & MBB revenue

• Interpretation of Slope: As total revenue increases by \$1 million, the winning percentage is expected to increase by 0.07857%, holding all other variables in the model constant.

PROCESS TO REDUCED MODEL: BEST SUBSET REGRESSION

X <chr></chr>	r2 <dbl></dbl>	adjr2 «dbl»	AIC <dbl></dbl>	BIC <dbl></dbl>	r2press <dbl></dbl>	Cp <dbl></dbl>
MBB_Rev	0.1029232	0.093379792	-73.39222	-65.69918	0.06739601	-4.2247161
Type_Cat MBB_Rev	0.1091770	0.090019531	-72.06382	-61.80643	0.05015541	-2.8366298
Conf_Cat Type_Cat MBB_Rev	0.1122669	0.083319121	-70.39739	-57.57564	0.03465980	-1.1389661
Conf_Cat Conf_Years Type_Cat MBB_Rev	0.1130322	0.074044600	-68.48018	-53.09409	0.01385079	0.7861566
Conf_Cat Conf_Years Type_Cat MBB_Rev MBB_Expense	0.1133328	0.064073514	-66.51272	-48.56228	-0.01340984	2.7567434
Conf_Cat FB_Rev MBB_Rev MBB_Expense Total_Rev MBB_Arena_Cap	0.1137947	0.041940162	-54.29030	-35.13471	-0.07828786	4.1010231
Conf_Cat Type_Cat FB_Rev MBB_Rev MBB_Expense Total_Rev MBB_Arena_Cap	0.1150331	0.030173247	-52.40358	-30.85354	-0.12876150	6.0016622
Conf_Cat Conf_Years Type_Cat FB_Rev MBB_Rev MBB_Expense Total_Rev MBB_Arena_Cap	0.1150538	0.016726421	-50.40547	-26.46098	-0.16017053	8.0000022
Conf_Cat Conf_Years Enrollment Type_Cat FB_Rev MBB_Rev MBB_Expense Total_Rev MBB_Arena_Ca	0.1150538	0.002877527	-48.40547	-22.06653	-0.18417435	10.0000000

9 rows | 1-1 of 7 columns

- Best R²: model 8 & 9 (highest) Best Radj²: model 1 (highest) Best Rpress²: model 1 (highest) Best Cp: model 8 (closest to k+1) Best AIC: model 1 (lowest) Best BIC: model 1 (lowest)
- We choose model 1 as our best model because it fits 4/6 optimality criterion

PROCESS TO REDUCED MODEL: STEPWISE SELECTION AND BACKWARD ELIMINATION

- Stepwise Selection
 - Only added MBB Revenue
- Backward Elimination
 - Removed all variables except for MBB Revenue

REDUCED MODEL: SIMPLE LINEAR REGRESSION

```
Call:
lm(formula = MBB_Win_Perc ~ MBB_Rev, data = data)
Residuals:
    Min
              10 Median
-0.44098 -0.10004 0.00514 0.13058 0.28528
Coefficients:
           Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.482490
                    0.034033 14.177 < 2e-16 ***
                    0.002247
           0.007378
                               3.284 0.00144 **
MBB_Rev
Signif. codes: 0 '*** 0.001 '** 0.01 '* 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.1617 on 94 degrees of freedom
Multiple R-squared: 0.1029, Adjusted R-squared: 0.09338
F-statistic: 10.78 on 1 and 94 DF, p-value: 0.001438
Analysis of Variance Table
Response: MBB_Win_Perc
         Df Sum Sq Mean Sq F value Pr(>F)
         1 0.28204 0.282045 10.785 0.001438 **
Residuals 94 2.45830 0.026152
Signif. codes: 0 '*** 0.001 '** 0.01 '* 0.05 '.' 0.1 ' ' 1
```

- Best Subset Regression, Stepwise Selection, and Backward
 Elimination all chose model 1 (with only 1 predictor: MBB revenue)
 as the best model
- Interpretation of Estimates:
 - o If a school had \$0 revenue from men's basketball, they would be expected to win 48.25% of their games.
 - As men's basketball revenue increases by \$1 million, the winning percentage is expected to increase by as much as 0.7%.

TESTING THE MODEL: FULL VS REDUCED

- H_0 : Reduced model is adequate.
 - H_a : Full model is adequate.
- F test statistic: $\frac{(2.2458-1.988)/(9-1)}{1.988/(96-(9+1))} = 1.394$
- F critical value: qf(0.05, 8, 86, lower.tail=F) = 2.048
- ullet Decision: F test < F critical so we fail to reject $H_{f 0}$
- Conclusion: We do not have enough evidence to show the full model is adequate, so we can use the reduced model.

TESTING THE MODEL: SIGNIFICANCE OF ESTIMATES

- H_0 : $\beta_j = 0$; the j^{th} predictor is not significant H_a : $\beta_j \neq 0$; the j^{th} predictor is significant
- MBB_Rev p-value: 0.00144
- Decision: p-value much less than $\alpha = 0.05$, so we reject H_0
- Conclusion: We have evidence that MBB revenue is a significant predictor.

PREDICTIONS

- using Villanova as a test predictor: MBB revenue was \$20.513366 million
- our model predicted Nova to have a win percentage of 63.38%
 - actual was 78.95%, so 19.7% error

```
#Predict using NOVA!
```{r}
test = data.frame(MBB_Rev=20.513366)
predict(new_fit, test)

1
0.6338475
```

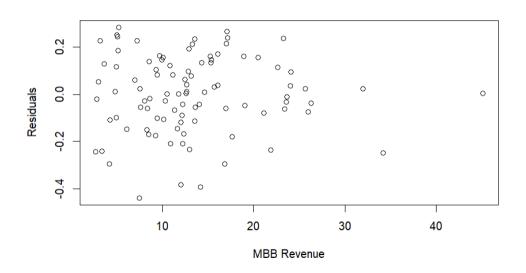
- Prediction Interval: We are 95% confident that the true winning percentage of an NCAA men's basketball team with \$21,513,366 revenue due to men's basketball is between 30.95% and 95.82%.

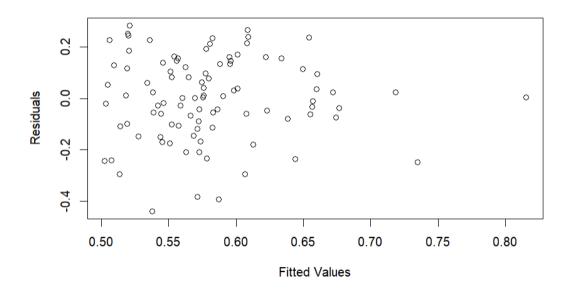
  1 0.6338475 0.3094646 0.9582305
- Confidence Interval: We are 95% confident that the true increase in winning percentage per \$1 million increase in men's basketball revenue is between 58.78% and 67.99%.

  1 0.6338475 0.5877517 0.6799433

## **ASSUMPTIONS**

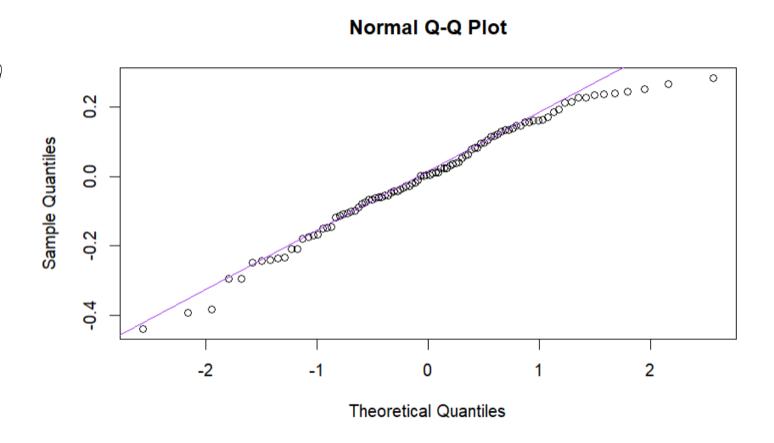
• Linearity & Constant Variance satisfied





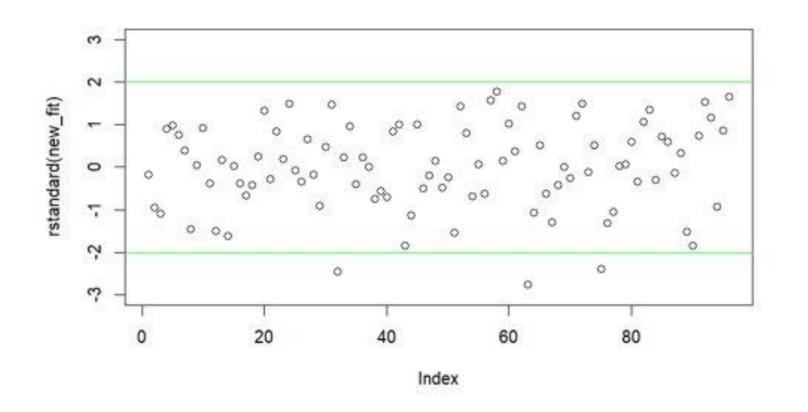
#### NORMALITY

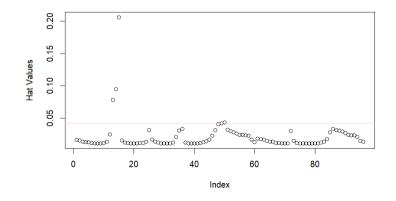
- Normality satisified
  - Wilk-Shapiro value = 0.97759
  - op-value = 0.099

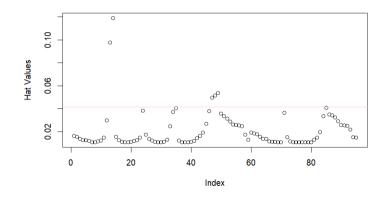


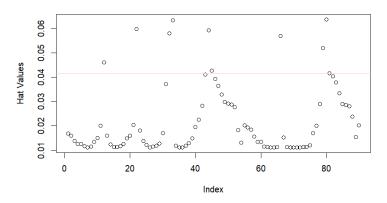
## OUTLIERS

- 3 outliers, but not super far
- None were also influential









#### INFLUENTIAL POINTS

- 4 influential points initially, then remove the most influential
- now 5 influential points, so we removed all 5
- then many more influential points
  - Back to the original data set since the overall model was generally the same

#### SUMMARY AND CONCLUSIONS

- Reconsider Initial Research Questions:
  - Do off-the-court factors impact a school's men's basketball results for an individual season?
  - If so, which factors are most important in predicting the results of a season?

#### SUMMARY AND CONCLUSIONS CONT'D

- Linear regression indicates that it is possible to predict the results of a season through off-court values
- Win Percentage = 0.48249 0.007378(MBB Revenue)
- However, we cannot predict a team's winning percentage with much precision
  - Our adjusted R<sup>2</sup> value for the linear regression model is only about 9.3%
  - We can say that a team's men's basketball revenue and their winning percentage are lightly linearly correlated
  - Our tests indicated that no other variables were able to predict men's basketball winning percentage with any significance

#### LOOKING FORWARD

- Look at time-series data to see how revenues over time are correlated to a team's success over the years
- Incorporate more variables and smaller schools (schools that are in smaller conferences with less money) in order to help those schools with tighter budgets
- We plan to investigate this relationship with additional skills and model capabilities in the future

#### THANK YOU AND WORKS CITED

#### Works Cited

"2021-2022 Division 1 Athletics Information." *Equity in Athletics*, ope.ed.gov/athletics/#/. Accessed 11 Dec. 2023.