Show Me The Money

For the college football Football Bowl Subdivision (FBS, formerly Division 1A), we wanted to discover what correlations to winning may exist by analyzing schools' financial information and team performance. We focused on three variables: coaches' salaries, school wins (including conference championships and bowl games), and total school revenue (not just athletic department). We used three data sets and focused on the years from 2010 to 2019. We included all institutions (public and private) that had available information in all of the datasets used. The results of our project can assist these athletic departments and their respective financial analysts that want to know if paying the coaches more means more wins and potentially additional revenue.

The questions we plan to answer are:

- Is it worth paying a coach a high salary (can you buy wins)?
- What is the win-to-loss ratio and expected revenue of a successful program?
- Is there any correlation between these three variables (salaries, wins, revenue)?

For FBS institutions, we expect that there will be a strong positive correlation with school revenue and winning football games. We do not expect a strong positive correlation with their coaches' salaries and wins though. For reference, FBS schools typically have the largest student enrollment; ex. University of Michigan.

Authors:

Patrick Hayes (pfhayes)

Tony Kibling (kiblian)

Dwight Ross (rossjrd)

Cleaning & Manipulation - Financial & DoE

Three data sources were used in our analysis: one with College Athletics Financial data (**Financial**), a download from the Department of Education (**DoE**), and a retrieval from CollegeFootballData.com (**CFBD**) using an API.

To bring all of the data together that we intend to use for our analyses, a common key between the Financial file and the CFBD was required. We decided to use conference-city-state as the key because there were not any duplicates. The Financial file has the institution name, subdivision, conference, and an IPEDS (Integrated Postsecondary Education Data System) ID. Unfortunately, that was not enough to link it to the CFBD. To add more identifying information to the Financial file, another source was required. The DoE file was obtained to acquire each institution's address information. Adding an address to the Financial file enables the required concatenation to create the common key. All cleaning and manipulation were done with Python.

The Financial file did not involve much upfront manipulation. One column was renamed and only columns relevant to the analysis (8 out of 31) were retained. Other changes were made once the files were merged. This financial data source does not break out total school revenue by sport. Total revenue in this data includes all sources of operating revenue.

The DoE file, on the other hand, required more cleaning and manipulation. The IPEDS ID is the common link to the Financial file and therefore the column name was changed to match the financial file (IpedsUnitIds to IPEDS ID). Similar to the Financial file, only the relevant columns were retained. The data was then filtered to include a LocationType of "Institution" only and to show six-digit IPEDS IDs. Using the power of regex on the address column, a City, State field was created. A few of the City, State values had to be replaced due to misspellings or a PO Box being shown. Also, three schools were identified as having the same City, State and were changed accordingly to differentiate them apart.

The Financial and DoE files were merged (called cfd_doe) using the IPEDS ID and relevant columns were retained. Since we are only focusing on certain years and schools that played football, the year field was filtered to include 2010-2019 and the NCAA subdivision field was filtered to exclude schools without a football program. The dollar values were converted to integers and blanks filled with zeros for future calculations. To match the CFBD file, the conferences were changed to a standard naming convention. Lastly, a common key of "**Conference - City, State**" was added for our next merge.

Cleaning & Manipulation - CFBD & showMEtheMONEY

Our final dataset also had the need for manipulation. Two dataframes were retrieved from the CFBD website: one with descriptors such as name, city, and logo; the other contained the wins per year. These two data frames were merged by team/school to give us the start to our cfbd_final data frame. A few of the cities were updated to correct their locations to match cfd_doe and two City, State combinations were changed due to duplications. A common key of "**Conference - City, State**" was then created to merge with cfd_doe.

Now that the Financial data, DoE information, and CFBD have been merged using the common key, this data frame can be manipulated to its final state for the analyses. As with the other data sets, the columns to keep were defined and unnecessary columns removed. Several calculations were added including winning percentage, average revenue per game, and what percentage of the overall school's coaches compensation was for football. The logos of each respective conference were added to show in a visualization. A column was added to identify which schools are in the "Power 5" (ACC, Big 10, Big 12, Pac 12, SEC) versus "Group of 5+" (non-Power 5) conferences. Because of their stature, Notre Dame was added to the Power 5 even though they do not have an official affiliation to one conference. Lastly, the columns were renamed to be presentable.

As with any rule, there were exceptions. In the case of the financial information, not all institutions reported their data (mainly private schools). Years with missing data are filtered out of our analyses.
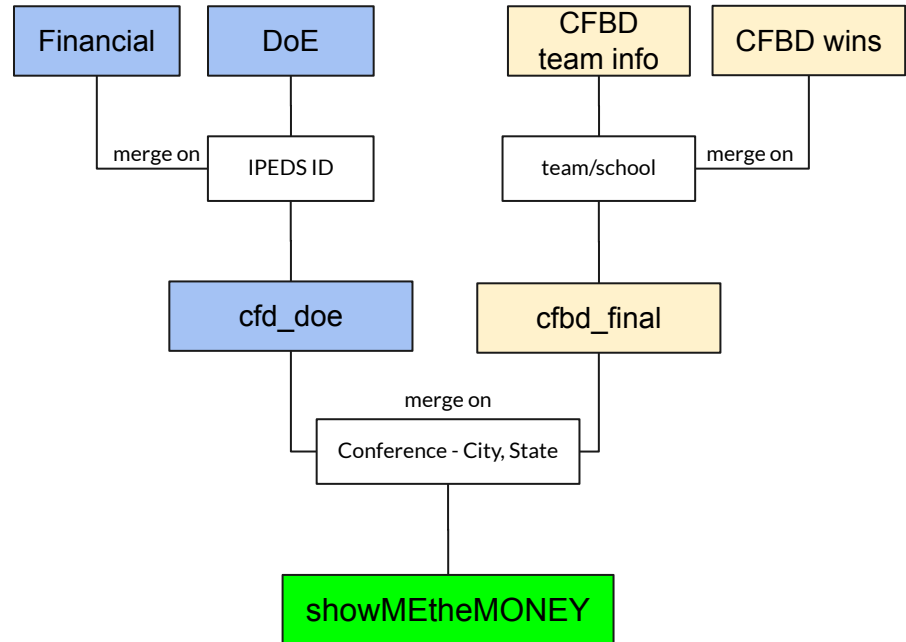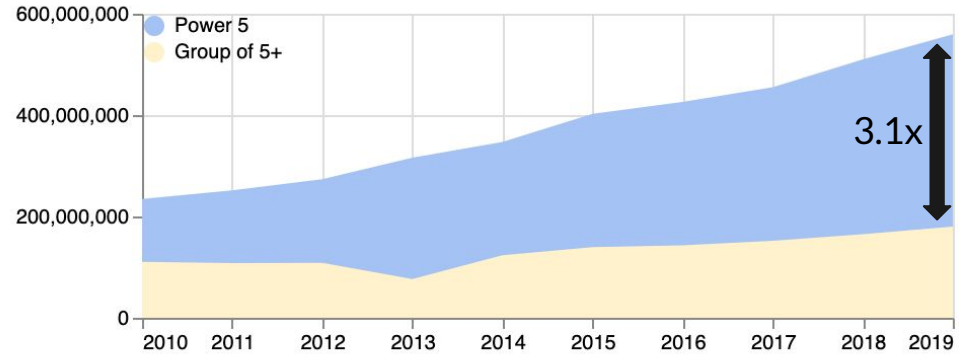
Analysis & Viz - Groupings

To begin our analysis, we find it essential to understand and illustrate how total revenue and coaching salary spending has grown over the decade of 2010-2019 for the FBS and the Power 5 conferences and the Group of 5+ (all other FBS conferences).

There are **significant differences** between the two groups of conferences, and this is a critical theme to keep in mind for the rest of the project.

As the chart shows, coaching salaries more than doubled over the ten years. Also, it is important to note that the Power 5 pays **3.1 times more** ($560M to $180M for the Group of 5+). Total FBS revenue almost **doubled** ($4.7B to $9B) with the Power 5 generating $4.5B more than the Group of 5+ which is **2.9 times** as much.

Part of the growth is due to conference realignments, the creation of additional conference championship games, TV media rights being re-signed, and the introduction of conference television networks.


Total Football Coaching Salary Growth ($)

Power 5
Group of 5+

3.1x

Total Revenue Growth ($)

2.9x
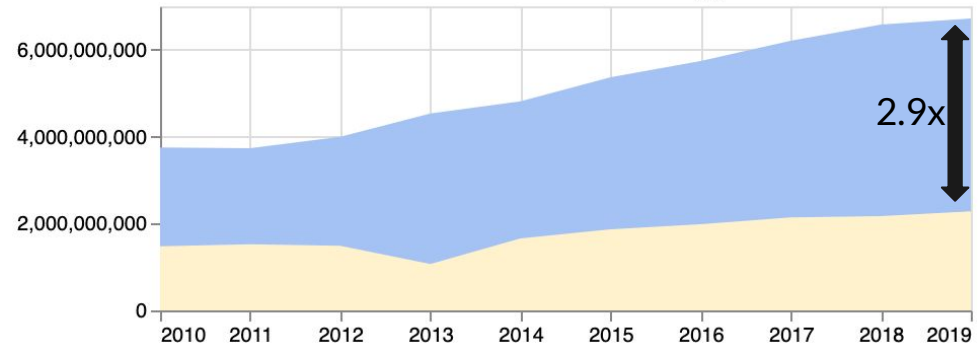
Analysis & Viz - Groupings Cont'd

Continuing our review of growth over the ten years from 2010-2019, the data shows an **increase of $350,000 for Power 5 football coach's per game salaries** compared to a **$111,000 increase for all other conferences**. A trend line is shown in the visualization below to represent the mean salary.
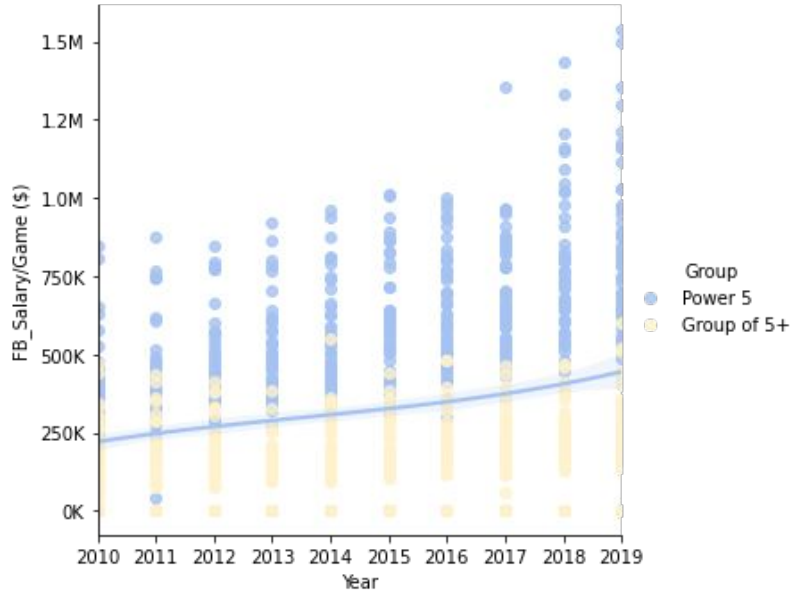
The following statistics were calculated to better understand the differences in the conference groups and will be used later to determine if there is any correlation between universities and their football coaches' salary, winning football games, and overall revenue.

- Coaches Salary / Wins Per Year
    - **Power 5:** **$827K**
    - **Group of 5+:** **$360K**
- Coaches Salary / School Revenue
    - Power 5: 7.4%
    - Group of 5+: 7.3%
- School Revenue / Wins Per year
    - **Power 5:** **$11.3M**
    - **Group of 5+:** **$4.9M**

As the numbers show, revenue and salaries for the two conference groups are vastly different. But there is a similarity between how much of the total school revenue is being used for the football staff.
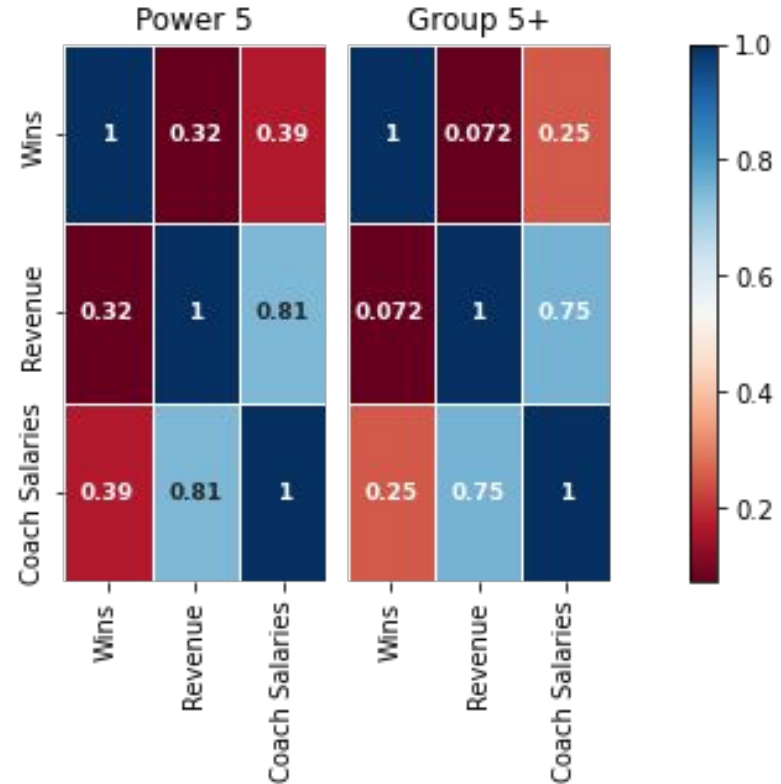
Analysis & Viz - Correlation Analysis

Using the three variables we isolated (coaching salaries, wins, and school revenue) a regression analysis was performed across each of them to determine if any correlation exists using the Pearson correlation coefficient. The conference group results are plotted in the heatmap to the right, and FBS level outputs are below.

- Coaches Salary & Wins Per Year (all FBS)
    - 0.372 correlation, 0.138 r^2 value

- **Coaches Salary / School Revenue (all FBS)**
    - **0.916 correlation, 0.839 r^2 value**

- School Revenue / Wins Per Year (all FBS)
    - 0.309 correlation, 0.104 r^2 value

We find that winning **does not** have a positive relationship with school revenue. It is known that there are teams that will always be top 5 in revenue but have not been successful with winning.

However, school revenue and coaches' salaries do **show a strong correlation** indicating that if a coach is being paid more, the school produces enough income to do so.

Analysis & Viz - Correlation Results

We are saving the best analysis for last. It allows us to answer the question we asked at the beginning of "Is it worth paying a coach a high salary (can you buy wins)?

**We discovered that paying for coaches does not directly lead to winning.**

The data set we analyzed suggests that a school cannot buy wins, as better visualized by the scatterplot on the right.

Across all schools, the r^2 value for the relationship of these two variables is 0.138, which is not indicative of decisive influence.


Can You Buy Wins? No.
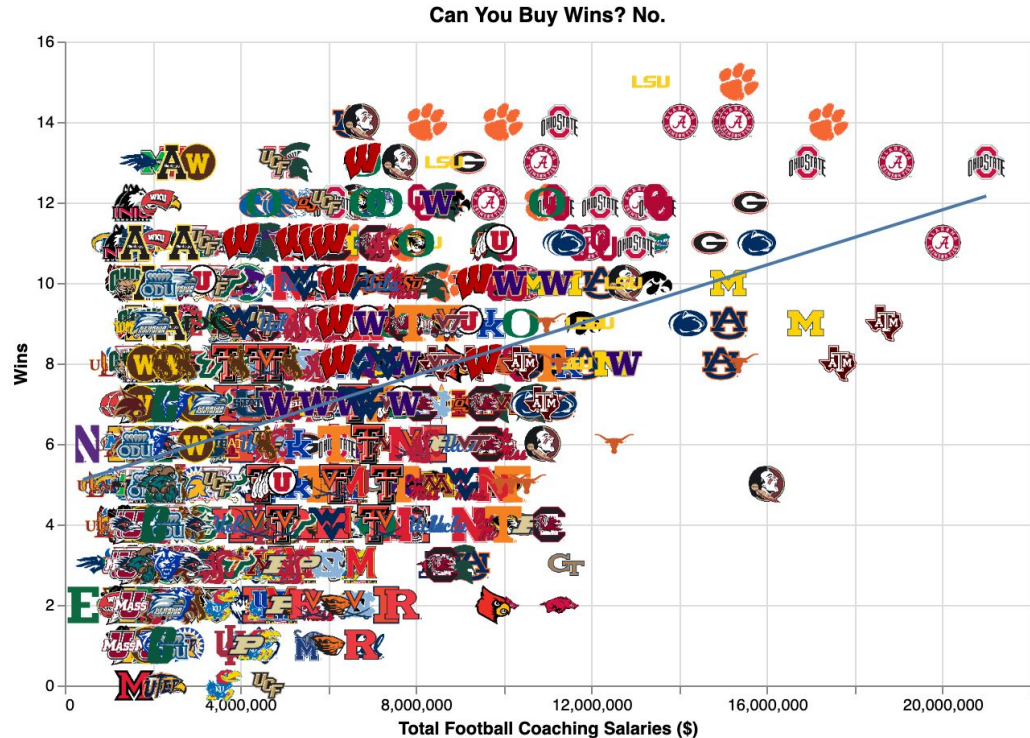
Analysis & Viz - Determining Success

Given how each institution defines success provided their unique constraints, finding an objective way to calculate the win-to-loss ratio and expected revenue of a successful program is not probable. Therefore, we defined **success as total wins minus total expected wins** using data from the CFBD.
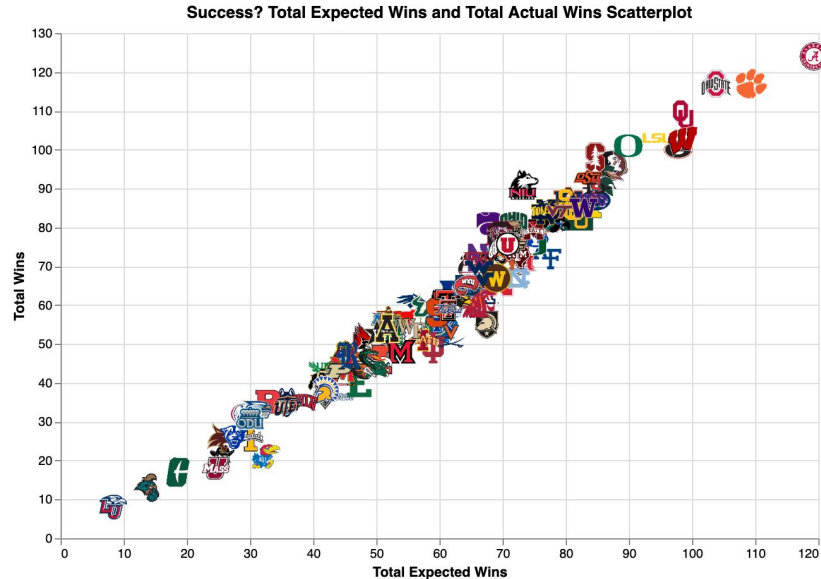
As a related approach, we reviewed school- specific coefficients for wins and salaries. Four schools had a strong positive correlation (>.66), which would indicate success in paying for wins. On the other hand, nine schools had a strong negative correlation (<-.66) and seemed to be wasting resources. These nine could be labeled as paying to lose.

Success? Total Expected Wins and Total Actual Wins Scatterplot
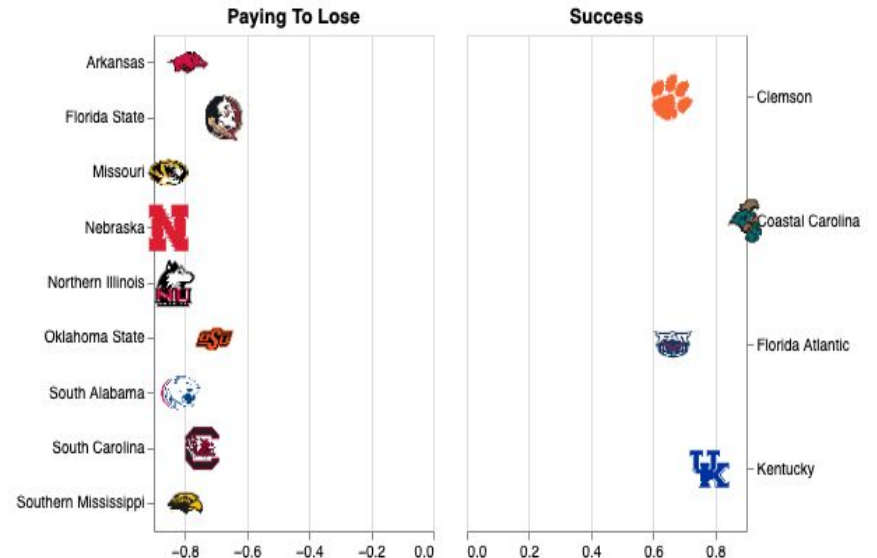
Summary and Future Considerations

Our journey began with the intention to answer three questions about paying coaches and whether that resulted in more wins and increased revenue from 2010-2019 in the FBS. Along the way, we discovered many other factors to consider about these items. We also found significant differences between the Power 5 conferences and all other FBS conferences.

Paying coaches to develop a system and recruit the best student-athletes to execute that system are factors that can increase overall football salaries. The data shows that coaches in the Power 5 conferences make five times more than their counterparts in the other FBS conferences. However, our regression analysis indicates that paying coaches a higher salary, regardless of conference, does not directly correlate to winning more games. Strength of schedule, injuries, and transfers do impact getting wins, though.

The Financial file reveals that revenue is four and a half times as much in the Power 5 than the other FBS conferences. TV contracts, conference affiliation, and making bowl games are some sources of revenue. There is a strong positive correlation between overall revenue and coaches salaries, thus the results provide proof that schools making higher revenues can pay their coaching staffs more salary. We also expected a strong positive correlation between school revenue and winning football games. The data proved otherwise.

Future Considerations:

As mentioned, there are other factors that can be used to advance this project. For instance, even though recruitment rankings are subjective and come from various sources, there could be an increase in wins and revenue due to a player being added to a roster. The caliber of student(s) via the transfer portal may also have an impact. State laws surrounding Name, Image, and Likeness (NIL) deals can influence where a student enrolls too.

Do other metrics better define success? Metrics that could be analyzed further include SP+ ratings, Elo ratings, SRS ratings, talent compositing ratings, win expectancies, etc.

Expanding the study into whether the conference correlates to coaches salaries and adding FCS (Football Championship Series, formerly Division 1AA) schools can also be accomplished with additional resources.

Statement of Work and Data Sources

Statement of Work:

-All: project proposal, code debugging and reviews, data cleaning and manipulation, analyses
-Patrick Hayes: initiated notebook with deepnote, obtained CFBD data, visualizations
-Tony Kibling: obtained Financial and DoE data sets, final project template and layout
-Dwight Ross: project summary, ethical issues, visualizations

Data Sources:

-College Financial data file, 5296 x 31, 1.5MB, CSV, http://cafidatabase.knightcommission.org/reports
Self-reported data by institutions on NCAA financial reports and on reports required by the federal government (Equity in Athletics Disclosure Act and the Integrated Postsecondary Education Data System)
-Methodology for College Financial data file,
https://www.usatoday.com/story/sports/college/2021/08/04/2019-ncaa-athletic-department-revenue-database-methodology/5482815001/
The definitions for all fields and any changes made throughout the years.
-College Football Database, 1263 x 13, API, https://pypi.org/project/cfbd/ https://collegefootballdata.com
Sports statistics and analytics website with no direct affiliation to the NCAA, its member conferences, or its member teams
-Department of Education file, 33603 x 14, 6.3MB, CSV, https://ope.ed.gov/dapip/#/download-data-files
Database of Accredited Postsecondary Institutions and Programs (DAPIP) contains the full institution name, IPEDS ID, and address information