

Football Training

1 Introduction

Welcome to the Football Training! This exercise will serve as a self-assessment of your data science skills. The questions presented here represent typical types of problems that you might encounter when cleaning data, and you should feel prepared to take these on after completing an introductory course using any data science tool.

To simulate a real research or professional data project, these questions are framed as open-ended tasks to answer questions about the data. There may be places where there appear to be multiple ways to answer a question, or the question may not warn you of potential issues in the data. In these cases, make a judgment call to clean the data, and explain in your answer how that choice affected your results. You may work with a partner and you may use any resources available to you, printed or online.

2 Football Training

1. This project will investigate how college conference realignment will impact university finances. To start, import the dataset entitled “Revenues All Sports and Men’s Women’s and Coed Teams.csv”. This contains revenues for all university sports by university between 2003-2021. For this project, you will need total revenues from football, basketball, baseball, and hockey, as well as the grand total revenues for each university, so identify these so that you can use them later.

Optional: rename the variables for ease to more straightforward names, and drop any extraneous or unnecessary variables to streamline your work.

2. For this project, we will only look at data from 2013-present, so drop any observations not in this time frame. In addition, drop any colleges that don’t have an NCAA Division I, Division II, or Division III classification.
3. Next, answer the following exploratory questions about the football revenues:
 - What is the average total annual revenue of college football NCAA Division I, II, and III programs? What is the average annual revenue for NCAA Division I FBS or FCS teams?
 - What share of total revenue did football comprise in 2019?
 - For comparison, what shares of total revenue did basketball, baseball, and hockey comprise in 2019 (bonus points for using a loop!)?
4. Save this data so that you can come back to this later.¹
5. Next we will use a different dataset to get more information about college sports conferences. Import the college conference data entitled “NCAA Conferences”. This dataset contains the college conferences of each NCAA Division I, Division II, and Division III college between 2013-2023. In the next step, you will merge this dataset to the revenues dataset.

After importing, convert the data so that it is unique at the college-year level. Then, save it so that it is in clean format for merging.

6. Merge the conference data with the revenue data. Drop any observations that don’t merge, and ensure that the data are still at the college-year level. Then, save the merged data and answer the following questions.
 - How many unique college-year pairs are in the merged data?
 - How many unique colleges are in the merged data?
 - How many college conferences in the merged data?
7. Aggregate the data to the year-conference level, taking the sum of football revenue and total revenue.
 - What were the top 5 earning college conferences in 2018?
 - How much did the SEC make in 2021 as a share of total revenue in 2021? How much did the top 5 earning college conferences make in 2021 as a share of total revenue in 2021?
8. Shifting to analysis, the University of Texas at Austin (Unit ID 228778) and the University of Oklahoma (Unit ID 207500) are moving to the SEC college conference in 2024. To understand how this will affect university finances, use the merged data and change the conference designations of the University of Texas and the University of Oklahoma to “SEC” in 2021. Aggregate to the conference level again, and report the new SEC revenues as a share of total college sports revenues. Is this very different?
9. Create a simple line graph of revenues of the top 5 earning conferences over time.
10. Suppose that a journalist reaches out to you, an education policy expert, looking for information on conference realignment. They ask how important football is to university sports revenues and how conference realignment will affect college finances in the future. In a short paragraph, summarize your findings from analyzing these

¹You may save the data as frame or in working memory if your data science tool allows it, but this is optional.

datasets, and comment on how you think conference realignment will affect the Big 12 conference and the SEC in the future. Use the percentages that you just calculated to quantify how SEC and Big 12 finances will likely change.