US24STT180 Group 5 Project

August 2nd, 2024

Contents

Setting up:

```
library(tidyverse)
library(ggplot2)
library(knitr)
library(infer)
```

Introduction

In college football discourse, many people view recruiting as one of the most important jobs a coaching staff has. During the off season, fans obsess over whether their school is attracting top prospects. Head coaches are often hired or fired based on their recruiting record. And schools invest millions in state of the art facilities and Name, Image, and Likeness (NIL) deals to attract top recruits. So clearly, both schools and fans must hold recruiting in high regard.

But how important is recruiting really? Is the relationship between a team's offseason recruiting performance and in season record really that strong? Is it possible to build a winning football program without cream of the crop recruits? In this project, we set out to investigate these questions that lie at the heart of college football.

We anticipate that recruiting will play an important role when it comes to a team's success. After all, having more talented players than your opponent is a big advantage. However, because a team's offensive and defensive strategies as well as player development abilities will vary between teams, we expect there to be a decent number of teams with weaker recruiting classes that still go on to perform decently well. For example, in 2021 the University of Cincinnati reached the four-team national playoffs with the 50th overall rated recruiting class. As such, we expect recruiting ranking to be an important factor in team success but certainly not definitive.

Dataset

To conduct our investigation, we have merged two different datasets. The first is the college football dataset provided in the project directions titled "College Football Team Stats Seasons 2013 to 2023" curated by Jeff Gallini Link. This dataset provides a comprehensive collection of college football stats per team per team from 2013 to 2023. For this study we have only chosen to include 5 years from 2019-2023 since college football is a dynamic sport that changes quite frequently (i.e stats from 5+ years ago may not be relevant to understanding the modern game.) We have then merged this dataset with recruiting data sourced from collegefootball.com. To source this data, we wrote a script in python that makes API calls for recruiting data for each year, translates the JSON output into a pandas dataframe, and then exports the data to CSV format.

We further filtered the data to include only Power 5 teams, which are the teams from the top conferences in division 1 college football and therefore represent the highest level of competition in the game. We feel that filtering the dataset to include only these teams is justified for two main reasons. First, top recruits are usually only interested in the most prestigious programs and so recruiting is probably more important as a factor among these schools. Second, schools at this level of competition usually have similar strength of schedule, which makes comparing the number of wins between schools more meaningful (since we dont have to worry as much about some teams having substantially easier opponents than other schools, influencing their win count regardless of recruiting performance.)

Methods

Results

Conclusion

References