Project Name: Understanding Correlation Between College Football Statistics and Performance

Team Members: Tara Moore, Matt Powers, Clare McAuliffe

Data Set:

We will be using the College Football Team Stats Seasons 2013 - 2020: https://www.kaggle.com/jeffgallini/college-football-team-stats-2019?select=cfb13.csv

9 CSV Files: Each CSV contains data of over 100 football teams for each season from 2013 - 2020.

Project Overview and Technique:

We will use this dataset to understand which teams and conferences have been the most dominant in college football over the last decade. As we have been excited to see Michigan Football do well this season, we have an interest in learning if there are specific statistics within college football that correlate with the success or failure of other teams. In addition we want to explore if the relationships among variables are stronger based on the conference a team plays in. We also hope to answer questions such as is there a correlation between offensive/defensive performance and conference? Which statistics can predict that a team will be successful the most?

Within each csv there is a data set including over 120 columns giving statistics for each team like games played, games won, offensive rank, defensive rank, yards allowed, points allowed, points scored, etc. There will be some cleaning before we begin to run our analyses. For example, we will have to create a conference column and split it up from the team's school since the two are combined right now. Once we have cleaned the data, our next step will be to run analyses on the different variables to understand which relationships are the strongest. While the obvious correlation will be offensive/defensive rank and games won, we are interested in looking at variables such as pass attempts per game and seeing if there is any correlation there.

Milestones:

Milestone 1: Data Collection and Cleaning

Due Date: 11/14

Milestone 2: Data Analysis

Due Date: 11/21

Milestone 3: Data Visualizations and Correlations

Due Date: 11/30

Sketching Out Presentation:

1) **Introducing the Project & Topic:**
    a) Explaining our interest and motivation to the topic
    b) Relate to how the season is going currently
2) **Explaining Dataset in Real World Context:**
    a) Talking through what the columns mean and what the data points represent at a high-level
3) **Explaining Manipulation of Data:**
    a) Walk through how we manipulated code to get it to the point where we were able to create visualizations
    b) Talk about hard aspects of the working with code, outside resources we used, and how we were able to overcome the challenges
4) **Showing Visualizations & Findings:**
    a) This is where we show our insights and findings and talking about what they mean and how they could potentially be applied to next steps
5) **Q&A:**
    a) Answering questions from audience and facilitating further discussion