Executive Summary

This section provides an overview to the project.

It should briefly touch on the motivation, data question, data to be used, along with any known assumptions and challenges.

Soccer is colloquially known as "the world's game" and I was lucky to have grown up playing the sport into my early adulthood. My childhood dream was to play in the world cup for the USA but now I am more interested in how I might continue to add value to the game, but from an analytical perspective.

Motivation

Here you will go into more detail about why you have chosen this project.

Grew up playing soccer in the deep south ; curious how weather might affect outcomes of games: more scores? More cards? More injuries?

Data Question(s)

Weather/Temperature vs. Soccer Game - If the weather is hotter, are teams more apt to draw, have lower scores, etc. (Data sets you'd need--> weather and soccer data). Hypothesis - I think there will be more draws when the temperature is hotter.

Schedule (through 6/25/2020)

1. Get the Data (5/29) data.world EPL : English premier league
2. Clean & Explore the Data (6/5)
3. Create Presentation and Shiny App (6/12)
4. Internal Demos (6/19/2020)
5. Demo Day (6/25/2020)

Data Sources

Document the data you use and the source of that data (provide a link to the data when possible)

Soccer/   EPL 2015-2016

https://data.world/chas/2016-2017-premier-league-matches

Weather: https://metoffice.apiconnect.ibmcloud.com/metoffice/production/node/174

Known Issues and Challenges

Explain any anticipated challenges with your project, and your plan for managing them.

Unit of measure?
Celsius?
What is the equivalent of a zip code in UK?
City of home team not readily transparent.

Only one year dataset, too narrow?

Joining - unknown exact type of join