

Final Project Proposal

Ankit Sanghi and Bichuan Ren

Link to dataset:

https://drive.google.com/file/d/1_qsFUzIUGTDopcDrBBnhxBim1A0-MJ0W/view?usp=sharing

We are using a dataset describing Major League Soccer matches. Our two primary research questions are “Which venues have the highest attendance?” and “Which teams score the most points on average?”. For our first question, we will be using the attendance (quantitative) variable as our response variable and the venue (categorical) variable as our explanatory variable. For our second question, we will be using the home and away teams (categorical) as our explanatory variable and their respective scores in matches (quantitative) as our response variable. We obtained the data from Kaggle at [this](#) link. You can view the source csv file [here](#).