DSCI 510 – Final Project Proposal

Student Name: Bryan Ayala-Rivera, bayalari@usc.edu

- Project Title: THAT'S NOT A STRIKE UMP: Measuring MLB Umpire Accuracy
- Project Summary
    - Major League Baseball generates a lot of data. I would like to analyze the accuracy of MLB Umpires when it comes to calling balls and strikes. When umpires miss a call, players sometimes argue and get ejected by the umpire. I want to examine if low accuracy umpires are associated with player arguments and ejections. I would also like to possibly explore if some umpires are more likely to eject players.
- Data Sources table

Data Source #	Description	Source URL	Type	List of Fields	Format	Have you tried to access/collect data with python	Estimated data size/number of data points
1	Retro sheet has detailed information on more than 19,000 ejections in the MLB. A complete listing of known ejections is available for download	https://www.retrosheet.org/eject.htm	-file	GAMEID, Date, Umpire, Reason for Ejection, Inning Ejected	csv	Not yet, will use read and write functions to process data	CSV has 19,000 instances but I plan on using a subset, years 2015-2025

2	Baseball Savant	https://baseballsavant.mlb.com/statcast_search	API	Pitchlocation, called ball/strike, actually ball/strike	csv	No, I am not confident with scrapping data with API yet	>19,000, not sure, don't know how to call API
3	@UmpScorecards is an online platform dedicated to measuring the accuracy, consistency, and favor of MLB umpires. It is not affiliated with Major League Baseball.	https://umpscorecards.com/	API	Umpire information, accuracy of calls	csv	No, I am not confident with scrapping data with API yet	>19,000, not sure, don't know how to call API

- Description of analysis
    - I will find 3 umpires that have at least 10 ejections caused by arguing and strikes. I will get the GAMEID from those games and I will use Baseball Savant to get the pitch data location and calculate the accuracy of the calls. I will also compare the accuracy of games for the umpire where there was no ejection and see if the accuracy was better. I will then use UmpScorecard to collect data on the umpires and use it to compare data

DSCI 510 – Final Project Proposal

Comment: I am not sure about the scope of how much I should take, baseball generates a lot of data, I would like some feedback as to what a good goal for a beginner would be learning how to scrape data from an API. I am currently most confident manipulating the data in the csv with baseball ejections.