Linear Regression Project Proposal

- Phillip Richardson

Question/need:

- What is the framing question of your analysis, or the purpose of the model/system you plan to build?
- The MLB has come to me to create a model to predict stadium attendance by game (outside of the first 5 games of the season due to non-representation of future attendance)
- Who benefits from exploring this question or building this model/system?
- The MLB can launch social media campaigns and marketing campaigns for teams with lower predicted attendance (potentially below certain thresholds), as well as allowing the MLB to potentially estimate stadium revenue by using the predicted attendance with concession consumption and avg ticket prices.

Data Description:

- What dataset(s) do you plan to use, and how will you obtain the data?
- I will web-scrape 2 websites to get 2 different datasets
  - MLB schedule/results data for 2019 (here)
  - Weather data for home city by day (here)
- What is an individual sample/unit of analysis in this project?

- An individual sample here is 1 home game in the 2019 season (~2350 games)
- What characteristics/features do you expect to work with?
  - W/L %
  - Avg Runs Scored (per game)
  - Avg Runs Allowed (per game)
  - Temperature (daily high)
  - Precipitation
  - Opponent W/L %
  - Prior game (if home) had extra innings
  - Weekend/weekday indicator
  - Day or night game
    - interaction between D/N and Weekend (?)
  - Current W/L Streak (5 wins in a row = 5, 5 losses in a row = -5)
  - Games behind leader
- If modeling, what will you predict as your target?
- Total Stadium attendance by game

Tools:

- How do you intend to meet the tools requirement of the project?
  - Will use pandas/numpy for exploratory analysis
  - BeautifulSoup for web scraping
  - Scikit learn for actual regression
  - Will likely use plotly for visualizations, seaborn if things go south

- Are you planning in advance to need or use additional tools beyond those required?
  - I would like to explore plotly for interactive plots in this project (looking towards portfolio implementation in the future)

MVP:

- A basic linear model with a few impactful features