ASSIGNMENT #6: WIN PROBABILITY MODEL

Your task is to estimate in-game {win probability OR scoring expectation} by fitting a Markov chain model to a dataset from the sport of your choice.

WHY ARE YOU BEING ASKED TO DO THIS?

The best way to check how well you've learned something is to try doing it yourself. We have already covered the code to implement this method in class. Successfully estimating a win probability (or scoring expectation) model on a new dataset will require you to wrangle the data and make necessary adaptations based on the unique challenges of your data. You will practice two competencies: (a) applying a model to a new dataset and (b) interpreting the results to inform decision-making in the real world.

WHAT (EXACTLY) ARE YOU BEING ASKED TO DO?

Find a dataset of sequential events from the sport of your choice, and decide whether you will model win probability or a component of it, such as scoring expectation. Create a Markov chain model for the progression of the game or competition by defining the game state based on the information available at each event. Estimate the transition probabilities by counting the number of transitions between each pair of states. Calculate the win probability or scoring expectation from each state. Using the changes in win probability or scoring expectation on event transitions, create a player evaluation metric based on the transitions for which each player is responsible. Your metric does not need to be all-encompassing—you may focus on one component of the game. Report a leaderboard of the top players according to this metric. Does the magnitude of the estimated contribution from top players make sense? Create a data visualization to tell a story about your results.

Finally, **escape from model land** by interpreting your results in the context of the real world. You need NOT address all of these, but here are examples of questions you may consider addressing: *Why might this model be an inadequate representation of reality? What decision in sport management might be affected by this analysis? How might this analysis change the way fans think about the sport?*

SUBMISSION REQUIREMENTS

- A PDF report (max 4 pages) summarizing your findings, including at minimum the following:

    - a description of the game state you defined
    - a leaderboard of the top players according to your evaluation metric
    - a data visualization that tells a story about some aspect of your results
    - an interpretation of your results in the context of the real world

- An R script that contains all of the code you used to perform the analysis

REMINDERS

- Prepare your report as if your audience is a front office executive who has not seen the assignment prompt. Write clearly and concisely, and format your report in a way that makes it easy to read.

- In this class we value exercising **creativity** on homework assignments! Look for opportunities to put your own personal touch on your work—try to do more than parrot what you've been taught.

- Please **anonymize** your submission by removing any personally identifiable information (including file paths in your R script that contain things like a username!).

How will your grade be determined?

You will get feedback on your work product based on several criteria. Within each of those criteria, the feedback will be: Missing (0%), Needs Improvement (70%), Good (85%) or Exceeds Expectations (100%). Your grade on the assignment will be the average of the grades across criteria. The criteria are:

1. **Markov chain.** Did you define a sensible game state for your Markov chain model? Did you correctly estimate transition probabilities and calculate win probability or scoring expectation from each state?

2. **Player evaluation.** Did you correctly aggregate changes in win probability or scoring expectation according to the player(s) responsible for each event? Did you report a leaderboard of the top players?

3. **Data visualization.** Did you include a plot that tells a worthwhile story about your results? Is that story easy to understand from a quick glance at your plot?

4. **Creative thinking.** Did you bring your own ideas from outside of this class to bear on the assignment?

5. **Critical thinking.** Did you escape from model land? Did you weigh evidence from multiple perspectives in forming your conclusion? Did you provide a thoughtful interpretation of your results?

6. **Written communication.** Did you write clearly and concisely? Did you organize your key ideas with the evidence supporting them? Did you format your report in a way that makes it easy to read?