

ASSIGNMENT #1: SPORTS DATA

Your task is to go out and survey the landscape of data in the sport of your choice.

WHY ARE YOU BEING ASKED TO DO THIS?

We will be needing publicly available sports data throughout this semester for assignments and the project. We will share our findings with each other so that we are all aware of data available across many sports.

WHAT (EXACTLY) ARE YOU BEING ASKED TO DO?

Choose a sport. If you are undecided between two sports, choose the one that you think is less popular. Scour the internet to find publicly available datasets for your chosen sport, and summarize the state of data as high, medium, low or none, in each of three categories: box score data, event data and tracking data.

None: Not even professional teams have access to these data.

Low: Teams have access, but these data aren't available publicly for any meaningful sample of games.

Medium: There is at least one meaningful (e.g. a historical season of data) dataset publicly available.

High: Present-day data are publicly available as they are generated.

For the medium and high categories (if you have any), provide a short description of the publicly available data (year(s), league(s), sample size) and a URL link to the data.

SUBMISSION REQUIREMENTS

- A one-page PDF summarizing your findings (see sample below)

REMINDER

- Please anonymize your submission by removing any personally identifiable information.

HOW WILL YOUR GRADE BE DETERMINED?

This assignment will be graded for completion, meaning that as long as you complete the assignment, your grade will be 100%.