**Dataset Description**

**Summary of data**

Here, you'll find a summary of each data set in the 2026 NFL Big Data Bowl, a list of key variables to join on, and a description of each variable. The tracking data is provided by the NFL Next Gen Stats team.

**Files**

* **train/**
  + **input\_2023\_w[01-18].csv**
    - The input data contains tracking data before the pass is thrown
    - **game\_id**: Game identifier, unique (numeric)
    - **play\_id**: Play identifier, not unique across games (numeric)
    - player\_to\_predict: whether or not the x/y prediction for this player will be scored (bool)
    - **nfl\_id**: Player identification number, unique across players (numeric)
    - frame\_id: Frame identifier for each play/type, starting at 1 for each game\_id/play\_id/file type (input or output) (numeric)
    - play\_direction: Direction that the offense is moving (left or right)
    - absolute\_yardline\_number: Distance from end zone for possession team (numeric)
    - player\_height: player height (ft-in)
    - player\_weight: player weight (lbs)
    - player\_birth\_date: birth date (yyyy-mm-dd)
    - player\_position: the player's position (the specific role on the field that they typically play)
    - player\_side: team player is on (Offense or Defense)
    - player\_role: role player has on play (Defensive Coverage, Targeted Receiver, Passer or Other Route Runner)
    - x: Player position along the long axis of the field, generally within 0 - 120 yards. (numeric)
    - y: Player position along the short axis of the field, generally within 0 - 53.3 yards. (numeric)
    - s: Speed in yards/second (numeric)
    - a: Acceleration in yards/second^2 (numeric)
    - o: orientation of player (deg)
    - dir: angle of player motion (deg)
    - num\_frames\_output: Number of frames to predict in output data for the given game\_id/play\_id/nfl\_id. (numeric)
    - ball\_land\_x: Ball landing position position along the long axis of the field, generally within 0 - 120 yards. (numeric)
    - ball\_land\_y: Ball landing position along the short axis of the field, generally within 0 - 53.3 yards. (numeric)
  + **output\_2023\_w[01-18].csv**
    - The output data contains tracking data after the pass is thrown.
    - **game\_id**: Game identifier, unique (numeric)
    - **play\_id**: Play identifier, not unique across games (numeric)
    - **nfl\_id**: Player identification number, unique across players. (numeric)
    - frame\_id: Frame identifier for each play/type, starting at 1 for each game\_id/play\_id/ file type (input or output). The maximum value for a given game\_id, play\_id and nfl\_id will be the same as the num\_frames\_output value from the corresponding input file. (numeric)
    - **x: Player position along the long axis of the field, generally within 0-120 yards. (TARGET TO PREDICT)**
    - **y: Player position along the short axis of the field, generally within 0 - 53.3 yards. (TARGET TO PREDICT)**
* **test\_input.csv**
  + Player tracking data before the pass is thrown for the held-out test plays. Same structure as the input files. Provided without ground truth outputs.
* **test.csv**
  + A mock test set representing the structure of the unseen test set.  
    Contains the prediction targets as rows with columns (game\_id, play\_id, nfl\_id, frame\_id) representing each position that needs to be predicted.
* **sample\_submission.csv**
  + The required submission format showing the expected structure for predictions. Contains columns (id, x, y) where: id is a concatenated identifier in format {game\_id}\_{play\_id}\_{nfl\_id}\_{frame\_id}