Data

Nature:
The data is focus on NFL (National Football League) games, providing detailed information about plays, players, game outcomes, and in-game events, specifically surrounding tackles. The data consists of structured information related to games, plays, players, tackles, and player tracking details. It includes many data types such as numeric, text, and categorial variables.

Size:	Attributes
games.csv has 136 rows	games.csv has 9 columns
players.csv has 1683 rows	players.csv has 7 columns
plays.csv has 12.5k rows	plays.csv has 35 columns
tackles.csv has 17.4k rows	tackles.csv has 7 columns
tracking_week_1.csv has 1.14m rows	tracking_week_1.csv has 17 columns
tracking_week_2.csv has 1.37m rows	tracking_week_2.csv has 17 columns
tracking_week_3.csv has 1.42m rows	tracking_week_3.csv has 17 columns
tracking_week_4.csv has 1.42m rows	tracking_week_4.csv has 17 columns
tracking_week_5.csv has 1.48m rows	tracking_week_5.csv has 17 columns
tracking_week_6.csv has 1.25m rows	tracking_week_6.csv has 17 columns
tracking_week_7.csv has 1.29m rows	tracking_week_7.csv has 17 columns
tracking_week_8.csv has 1.41m rows	tracking_week_8.csv has 17 columns
tracking_week_9.csv has 1.15m rows	tracking_week_9.csv has 17 columns

The data may be too large, it consists of 4 training csv files and 9 possible testing or training csv files.

Overview

Goals:
The primary goal is to develop a machine learning model which accurately predicts/classifies the outcome of a pass play in NFL games. By accurately classifying these outcomes, teams can gain insights into game strategies and improve gameday decisions. This is interesting to us as athletes here at URI and football fans. Showing that this insight is not just interesting to analysts, but fans as well. The challenge will be the size of the data and the amount of attributes to work with.

Strategy

- What are the key problems you can immediately identify?
There is an excess amount of data we do not need all of the data so we will need to clean it.

- Is something wrong with the data?

Nothing is wrong with the data, it just needs to be cleaned.

- Does the nature of the data make it difficult to discern between instances?
- What is your strategy for addressing these problems?

We are going to clean the data by deleting extra columns.

- Are any of the methods we've covered so far conducive to solving the problem?

Yes we are going to use classification. To help us identify the outcome of the play pass.

- If not, how will you gain the knowledge to solve this problem? - Which ML framework do you plan to use. Why?

Notes:

- Pick a publicly available dataset
- View the following for inspiration:
    - [Papers With Code](#)
    - [Stanford C299 Projects](#)