**Technologies Used**

**Data Cleaning and Analysis**

[An Attempt](https://github.com/ksdisch/nfl-ml/blob/main/Scraping.ipynb) was made to scrape data from Football Outsiders automatically using a jupyter notebook and python. For now, the individual yearly tables were manually downloaded as csv files. The dataset from Kaggle containing the game information and totals did not require any scraping. The input data from Kaggle and Football Outsiders were joined and cleaned into cumulative csv's using an [r script](https://github.com/ksdisch/nfl-ml/blob/main/clean_transform_inputdata.R).

**Database Storage**

The cumulative dvoa, cumulative pace, and cleaned game data were imported into pgAdmin and joined into one dataset using [postgreSQL](https://github.com/ksdisch/nfl-ml/blob/main/database_schema.sql). This dataset could be imported directly into the Machine Learning Model, but another [r script](https://github.com/ksdisch/nfl-ml/blob/main/feature_engineering.R) was used to engineer features. This dataset was exported to the Machine Learning Model.

**Machine Learning**

[SciKitLearn](https://scikit-learn.org/stable/) is the ML library we'll be using to create our classification model that will attempt to correctly classify NFL games into hitting the over or under on their game total (as determined by betting markets). We've been testing several algorithms and comparing their results as we continue to improve and refine the dataset and model.