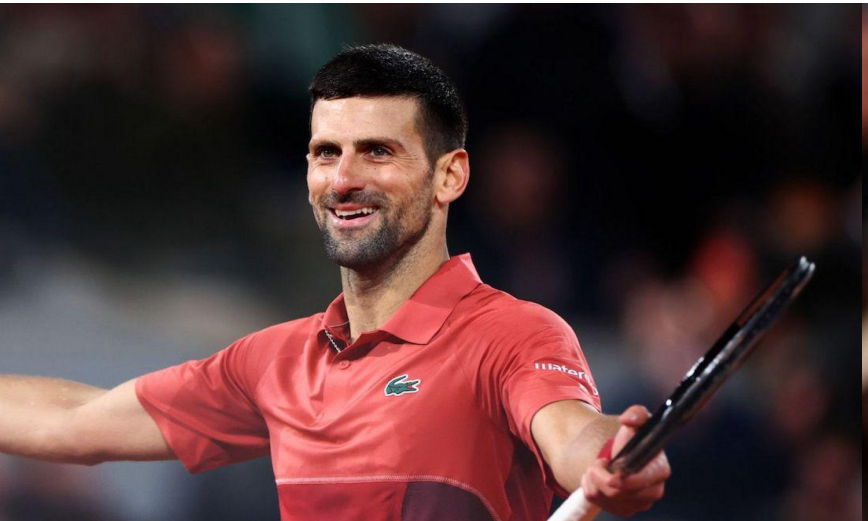
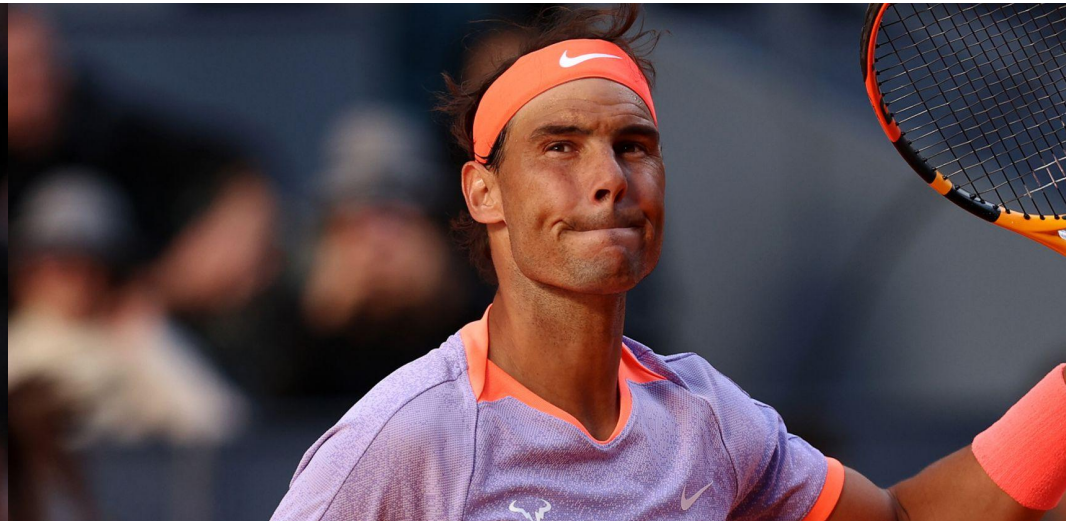
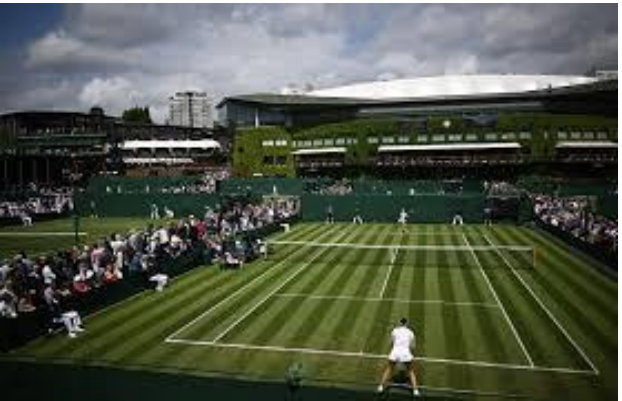
Using data science and ML to predict tennis outcomes

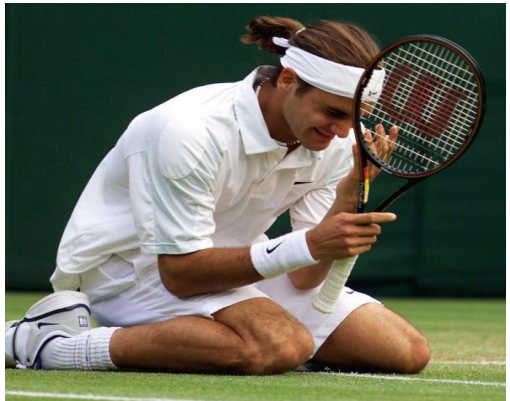


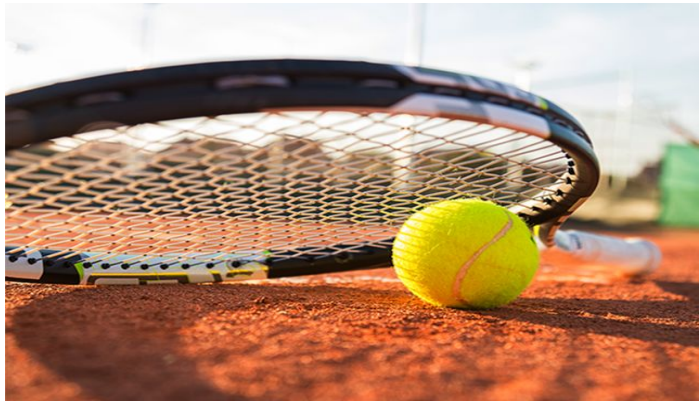


The Problem Area

- One of the most popular sports in the world
- It has millions of followers and many people offering their thoughts on who will win a match
- Predicting the winner of a tennis match is hard
- Many variables that go into predicting a winner

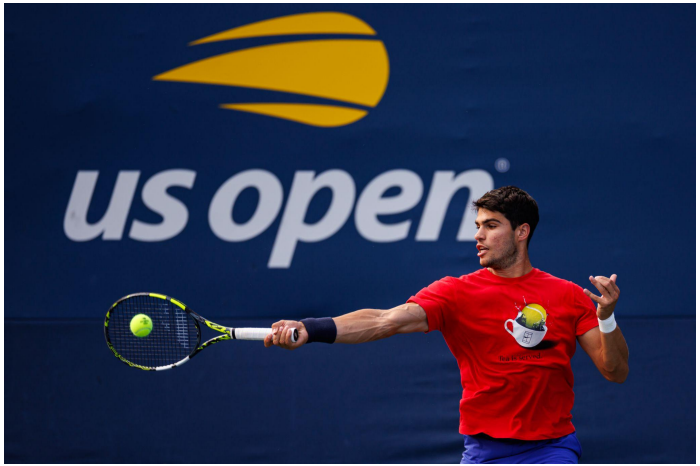






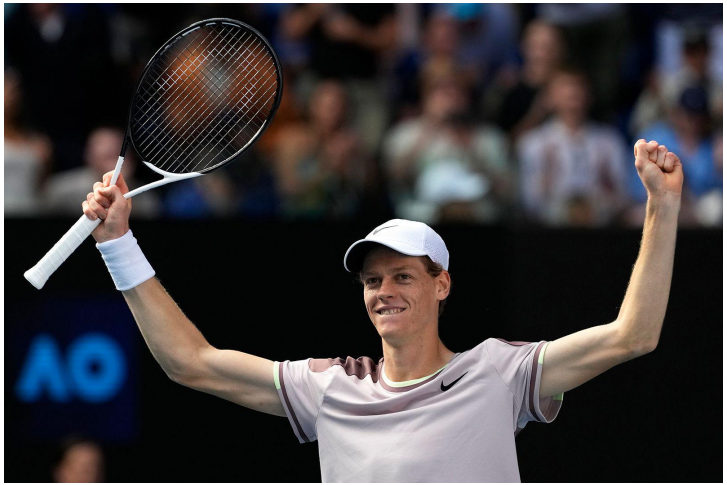
Data Science Solution

- Build a model that can predict who will win the match
  - As a baseline compare against simply picking the higher ranked player
- Explore the variables that do and don’t function as good predictors of match outcomes



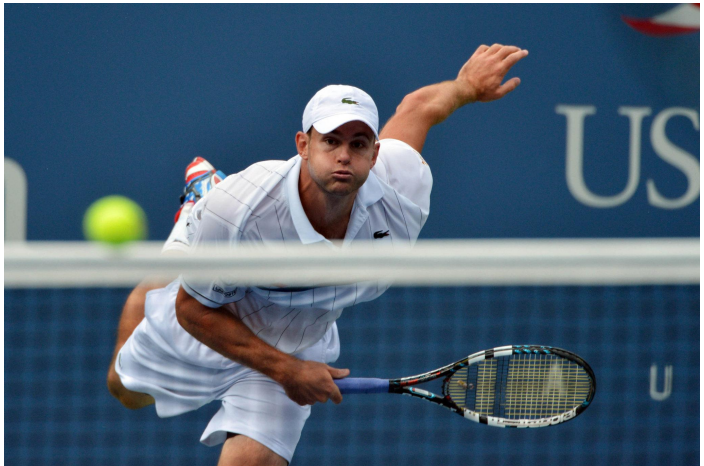
Impact

- Tennis Analysis
  - Shed light on the good and bad variables that predict match outcomes
  - Help to drive better analysis
- Sports betting
  - Using ML to “beat the house”



Dataset & Quality concerns

- JeffSackmann Github
  - Data is sourced from Github user
  - Data contains all tennis matches and outcomes from 1970 to present
  - No major quality concerns that I found after my initial EDA
  - Data is mostly clean with some null values



Next Steps

- Simplest model: Picking the higher ranked player
  - Higher ranked player wins 68% of the time
  - This is the number that I am trying to beat
- Perform further analysis on data
  - Looking for columns that are correlated with winning in the dataset
  - Start with linear modeling and then see if other ML techniques can help to predict outcomes