

# **DATA 6560 – Sports Analytics**

Checkpoint 1: Proposal Memo

Nicholas Condos

October 31, 2025

Tennis Analytics

## **Project Title & Topic Summary**

Title: Predicting ATP Match Outcomes through Serve and Other Player Metrics

This project will examine how player metrics and statistics influence match outcomes across different playing surfaces on the ATP Tour. Using data from 2008 and on, the goal is to identify which performance factors — such as serve accuracy, aces, and break points saved — most strongly correlate with match success (among other factors).

## **Problem Statement & Relevance**

In professional tennis, data analytics increasingly inform coaching, preparation, and player development strategies. Measurable insights are in high demand to enhance player performances.

### **Central Question/Proposal:**

Which serve-related and/or performance factor(s) most significantly predict ATP match success across different surfaces (hard, clay, and grass)?

This analysis will be relevant to several key stakeholders in the tennis and sports world:

# DATA 6560 – Sports Analytics

- **Coaches and high-performance staff**, who can use statistical insights to tailor match preparation and refine serve and other strategies.
- **Players and development teams**, who can identify specific areas of improvement and optimize their training focus.
- **Sports data analysts and performance consultants**, who can integrate predictive models into scouting reports or player dashboards.
- **Media and fan engagement professionals**, who can translate analytical findings into storytelling and performance context during live coverage.

By translating raw performance data into actionable insights, this project contributes to a more evidence-based understanding of success factors in professional tennis competition (ATP).

These insights will help refine strategies, enhance player preparation, and support evidence-based performance evaluation across multiple professional settings.

The findings will offer value to coaches, analysts, and players by revealing how measurable serve metrics and other variables contribute to match outcomes. These insights can help refine strategies, enhance preparation, and support evidence-based performance evaluation.

## Intended Dataset (Source & Description)

The project will use the “tennis\_atp” dataset by Jeff Sackmann, a publicly available, well-established repository hosted on GitHub. This dataset compiles detailed ATP match results from 1991 to the present, including:

- Player biographical data (age, height, handedness, nationality)
- ATP ranking information (rank and ranking points at match time)

# DATA 6560 – Sports Analytics

- Match-level performance statistics (aces, double faults, first-serve percentage, break points saved, etc.)

The dataset I selected is very comprehensive, well documented (matches\_data\_dictionary.txt), free and available for research. It seems ideal for statistical modeling and analysis of performance metrics. It offers a variety of metrics (tournament, surface, rounds, serve statistics, etc) to analyze.

## Initial Plan or Method

I will begin by performing data cleaning and exploratory data analysis (EDA) to identify trends and relationships between serve performance and match outcomes.

Next, I will conduct:

### *Descriptive Analytics*

- Summarize and visualize serve performance across surfaces.

### *Predictive Modeling*

- Build a logistic regression model to predict match outcomes (win/loss) using variables such as first-serve percentage, aces, double faults, and surface type.

### *Comparative Analysis*

- Examine whether key performance metrics differ between top-ranked and lower-ranked players.

The goal is to generate insights that highlight the most influential factors driving match success.

## Next Steps

# **DATA 6560 – Sports Analytics**

- 1) Download and inspect the tennis\_atp dataset from GitHub.
- 2) Clean and filter data to focus on ATP main-draw matches from 2008–2024.
- 3) Conduct preliminary summary statistics and visualizations by surface type.
- 4) Draft the Topic Justification Memo for Checkpoint 2, refining the research question and analytical focus.

## **Verification**

The tennis\_atp dataset by Jeff Sackmann is confirmed as a trusted, open-source dataset widely used in professional and academic sports analytics. It is hosted publicly:

[https://github.com/JeffSackmann/tennis\\_atp](https://github.com/JeffSackmann/tennis_atp)