Milestone 1

Racket Sciences

Réza Machraoui, Luna Ralet, Leila Sidjanski

March 21, 2025

1 Dataset

We found multiple data sources containing various files over the years for both men and women, covering multiple tournaments.

The datasets :

1. Jeff Sackmann's ATP Matches Dataset : This dataset provides ATP match outcomes, player statistics (titles, aces, first serves, win percentage), and details for major tournaments (Grand Slams, Masters 1000, Finals, Olympics).

2. Jeff Sackmann's ATP Doubles Matches Dataset : Similar to the ATP singles dataset, this focuses on doubles match outcomes and team performance, including player statistics like titles, aces, and first serve percentage.

3. Jeff Sackmann's WTA Matches Dataset : This dataset provides WTA match outcomes and player statistics (titles, aces, first serves, win percentage), with details for major tournaments (Grand Slams, Masters 1000, Finals, Olympics).

4. Jeff Sackmann's Github Match Charting Project : This Github compiles detailed shot-by-shot data from professional tennis matches (ATP and WTA). The datasets include shot types, directions, errors, statistics, and match metadata. We utilized the dataset on Net Points to identify the best players at the net.

5. Ultimate Tennis Matches Dataset : This dataset from Kaggle includes data from men's and women's tennis tournaments, providing details about matches and players. We used it to extract information on the playing surfaces of these matches.

6. Huge Tennis Database : This dataset from Kaggle offers comprehensive men player rankings, match results, and statistics, providing an extensive resource for tennis analytics.

7. Association of Tennis Professionals Matches : comprises detailed records of ATP tournament matches from 2000 to 2017. Each entry includes comprehensive match statistics, such as player information, match outcomes, and various performance metrics, providing a rich resource for analyzing professional men's tennis over this period.

2 Problematic

Tennis has long been the stage for debates about greatness, with fans and analysts constantly comparing players across different eras. With the emergence of statistical analysis and data-driven insights, we can move beyond subjective opinions to a more objective approach in determining the Greatest of All Time (GOAT) in various aspects of the sport.

This project aims to provide a visual exploration of the GOAT debate in tennis, using data to assess player performances across multiple dimensions. Our goal is to make complex statistics accessible and engaging for both casual tennis enthusiasts and dedicated followers of the sport.

We will develop visualizations that allow users to explore:

1. **The Overall GOAT Debate** – Comparing Djokovic, Federer, and Nadal using key performance indicators such as Grand Slam titles, total wins, head-to-head records, and longevity.
2. **The Best Server in History** - Analyzing ace counts, first serve percentage, service games won, and break points saved.
3. **GOAT by Surface** - Identifying the most dominant players on clay, grass, and hard courts based on win percentages and titles.
4. **GOAT in Doubles** - Highlighting legendary doubles players through Grand Slam titles and team chemistry.
5. **GOAT of Specific Tournaments** - Investigating who has dominated Grand Slam events and Masters 1000 tournaments over the years.
6. **The Best at Net Play** - Evaluating the greatest volleyers based on net points won and serve-and-volley efficiency.

Our objective is to provide a data-driven perspective on tennis greatness, using interactive visualizations to allow users to compare, filter, and interpret key insights. By leveraging historical data, we aim to offer a compelling tool that sparks discussion and deepens the understanding of what truly defines a tennis legend.

3 Exploratory Data Analysis

See notebooks. We used different notebooks to show the processing performed on the different datasets.

4 Related work

We are utilizing multiple datasets from various sources, with a significant portion coming from Jeff Sackmann's GitHub repository. His dataset has also been used for Ultimate Tennis Statistics, a comprehensive platform that provides in-depth analysis of professional tennis. This platform explores and analyzes a vast number of matches from both the ATP and WTA tours, offering insights into player performance, rankings, and match statistics.

We believe our approach is original because it goes beyond traditional rankings (such as Grand Slam titles or ATP rankings) by incorporating multiple characteristics to evaluate the GOAT in a more comprehensive way. While numerous analyses have been conducted on

tennis players, we wanted to create something more interactive and engaging, where users can directly compare the greatest players and see what makes each of them unique. Our visualization will allow users to explore strengths and weaknesses across different aspects, (in addition to an overall overview of who is the true GOAT)

To our knowledge, while overall GOAT rankings exist, but few (if any) offer a detailed breakdown by specific categories, enabling a nuanced and customizable comparison. Our approach makes the GOAT debate more interactive, insightful, and fun for fans, analysts, and tennis enthusiasts alike.