

# **Serving Analysis in ATP dataset(2021-2023)**

## **1.Introduction**

### **1.1 Background**

Tennis stands as one of the most widely adored sports globally, characterized as a solo player game that rivals team sports in popularity. Tennis is a complex sport. Each score requires multiple strokes to achieve, and among them, there is one shot that is unaffected by the opponent: the serve. So the serve allows the player to begin a point with a proactive advantage. Furthermore, the tennis match is divided into service games and receiving games. So that an athlete's serving statistics during a match directly impact the outcome of the game.

In this project, I try to utilize the ATP dataset and R programming language to find out the relationship between serving statistics and the result of game.

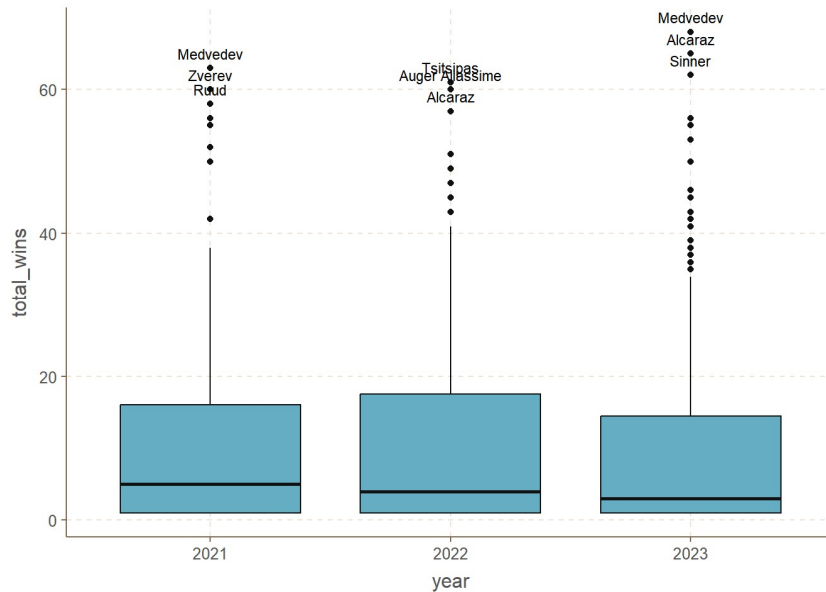
### **1.2 Dataset**

The project's ATP dataset is available on GitHub, posted by JeffSackmann. <sup>[1]</sup> The dataset consists of master ATP player file, and match stats file. In the player file, it focuses on the players' information about id, name, the racket-holding hand, birth date, country, and height (cm). In match stats file, it includes fundamental match details of ATP tour events, such as match duration, court surface, participating players, match outcomes, etc. And it also contains critical technical statistics, such as first and second serve data, and holding and breaking serves data, etc.

## **2.Methods**

### **2.1 Player performance**

Tennis is a form of competitive sports game played in a single-elimination format. Consequently, I've gathered the number of wins in the last three years to showcase the performance of ATP tennis players. I've specified the years by extracting data from tournament games' dates, then calculated the total wins per year for each player. Subsequently, I grouped the players by the sum of their wins per year and obtained their last names by left joining the player tables. From the output, it's evident that the performance of top players significantly exceeds the median. This indicates that in ATP tournaments, their performance surpasses that of other players. Therefore, just a few players control the overall superiority in the circuit.



## 2.2 Serve performance

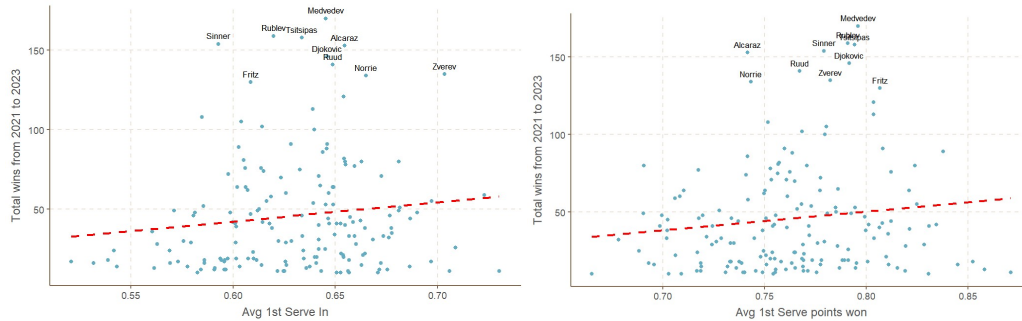
In the ATP official statistical analysis, there exists a serve score function that displays players' serve performance.

**Serve score= percentage of first serves in\*100+percentage of first serves win\*100+ percentage of second serves win\*100+ percentage of games win\*100+number of ace-number of double faults.**

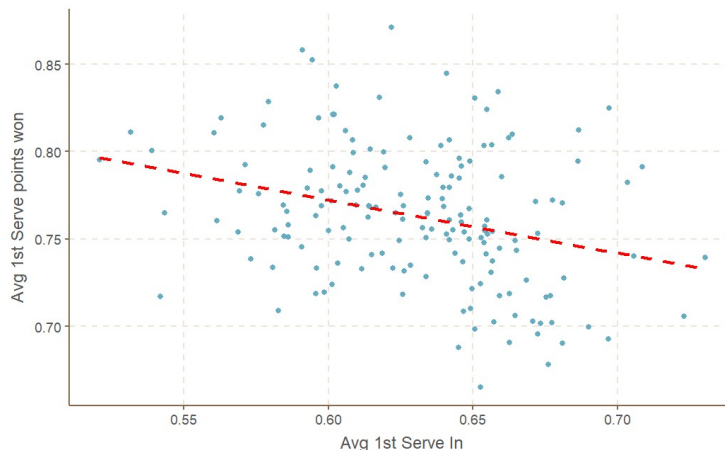
Therefore, based on the serve score function, I calculated each serve indicator within the function by excluding null values and subsequently added the related columns with the calculation results. Following the aggregation of serve indicators over the last 3 years, grouped by winners, I filtered the win count of games to be greater than 10 to avoid potential outlier situations due to a small number of wins. Finally, I generated a plot between serve indicators and the number of wins for players, and highlighted the data of players whose number of wins ranks within the top ten, in order to try to find out the implied relationship between them.

### 2.2.1 First serves

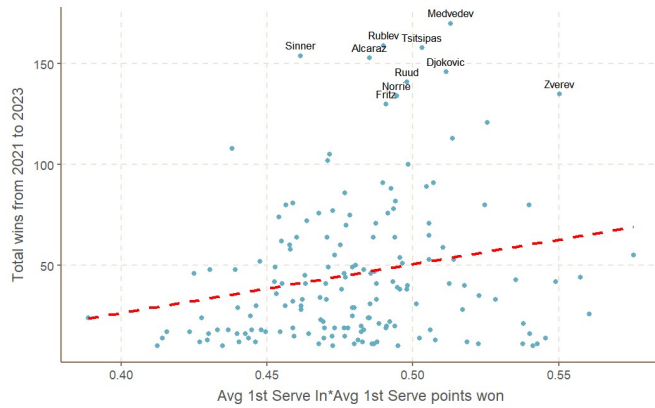
After examining both the scatter plot depicting the relationship between the percentage of first serves and total wins, I didn't observe a notably strong correlation. While there seemed to be a general positive trend. Interestingly, the top players didn't exhibit an absolute dominance as their performance in terms of total wins.



But looking at it from a different angle, there's an intriguing discovery regarding the relationship between the percentage of first serves in and winning, which shows a negative correlation. From a broader perspective, solely pursuing a high percentage of successful first serves might decrease the threat posed by serves, resulting in a failure to establish an advantage in the serving aspect. Consequently, it may become more challenging to win points, leading to a decline in first serve points win. This could explain why the positive correlation between the percentage of first serves in and the number of wins isn't as prominent, as players might sometimes avoid second serves, consequently reducing the effectiveness of their first serves.



Based on the analytical results, I attempted to use the first serve percentage multiplied by the first serve points win percentage as an indicator in comparison to the number of wins. It appears that most of the top players have a value exceeding 50% in this indicator. From the graph, it appears that there are considerably fewer players who can reach this high level compared to the previous two individual serve statistics. This might be a crucial factor contributing to their victories—maintaining a high first serve rate while maximizing the threat posed by their serves, thus winning crucial points.



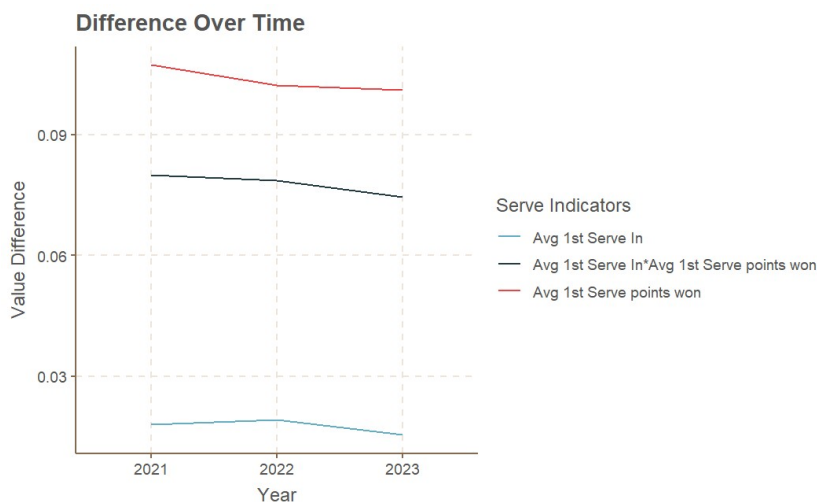
### 2.2.2 Second serves

Compared to the first serve, it's evident that the second serve poses much less threat in terms of the direct risk of losing points due to double faults. Therefore, the second serve percentage is no longer a primary factor in assessing serve performance, while the second serve points win rate often reflects a player's advantage in other competitive aspects.

### 2.2.3 The difference between winner and loser

Tennis is a competitive sport where game outcomes are influenced not only by the winner's performance but also by the performance of their opponent. Thus, by categorizing matches annually, I calculate the differences in key serving metrics between winners and losers. The aim is to observe trends in the discrepancies between winners and losers in their serving aspects.

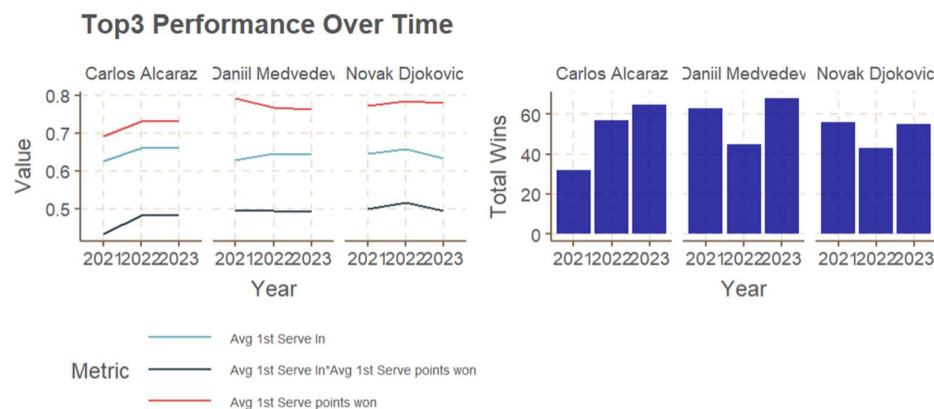
The following output clearly indicates a decline in difference of serving performance between the winner and loser, which means most of ATP players pay more attention on serving. The gap in serving performance between the winner and loser is narrowing.



### 2.3 Top3 players' performance

From the ATP official rankings as of the end of 2023, "Novak Djokovic," "Carlos Alcaraz," and "Daniil Medvedev" stand as the top 3 players. These players have also maintained high rankings in victories over the past three years. From the statistical results of serving performances, they also have demonstrated exceptional skills in serving. Especially, they can manage well in balance the percentage of first serve in and the percentage of first serve win. In order to validate the positive correlation between serving performance and the number of match victories among the top 3 players, I have separately collected data on match victories and serving performances for the top 3 players over the past three years.

From the graphs, it appears that the performance fluctuations of these three players over the past three years align closely with the fluctuations in their serving data. An interesting observation is that Carlos Alcaraz, being a player who rapidly ascended in the ATP rankings over the past three years, shows the most noticeable improvement in his serving performance data. On the other hand, for Novak Djokovic and Daniil Medvedev, whose serving levels have been at a relatively high standard, achieving substantial improvement within a short period might be more challenging compared to their match victories.



### 3. Conclusion & Further work

There is indeed a certain positive correlation between serving performance and match results. Simply hitting a high percentage of serves within the service box isn't sufficient for ATP players. They must optimize their chances of winning a point while serving by finding the right balance between successfully landing a serve and converting that into a winning point. Additionally, players must appropriately adjust the balance between these two factors based on their opponent's situation.

The serving aspect of the game is receiving increasing attention from ATP players. Within the participants of ATP tours, the disparity in this technical aspect is

narrowing. For athletes experiencing an upward trend in their performance, enhancing their serving capabilities proves to be an effective and swift means to elevate their overall achievements. However, as players reach higher levels, the room for improvement in serving becomes increasingly limited.

In future research, I aim to gather data on the serving speeds of winning and losing players, using serve speed as an indicator of a player's serving proficiency. I also plan to delve deeper into how athletes with varying serving speeds balance the threat and risk associated with their serves.

#### **Appendix:**

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

[2]Code: <https://github.com/Caroljiang114/DSCI401/blob/main/Project401.Rmd>