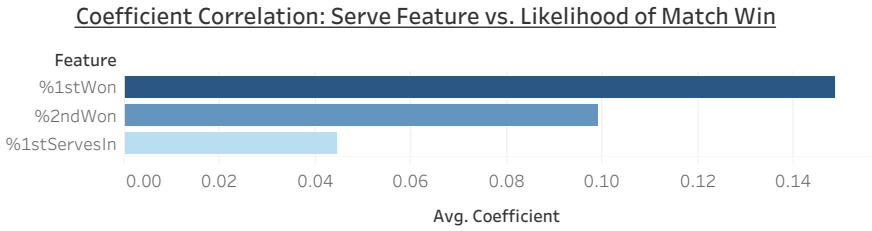
How Important is the Serve in Tennis?

This project is intended for curious tennis fans who enjoy match analyses.



In tennis, the player serving typically has the advantage in a rally. This project will explore which of the following three serve statistics has the greatest impact or correlation to the likelihood of winning a match:

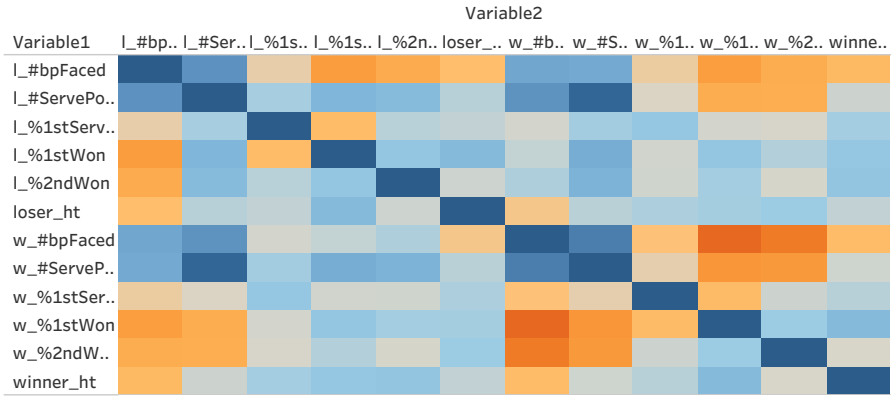
1. Percentage of First Serves in Play ('%1stServesIn)
2. Percentage of Points Won on First Serve ('%1stWon')
3. Percentage of Points Won on Second Serve ('%2ndWon')

Three Project Questions:

1. Which of the three statistics has the greatest impact on the likelihood of winnng a match on the ATP Tour?
2. Based on these three statistics, how did the impact of the serve change across the three different surfaces (clay, grass and hard courts)?
3. Roger Federer, Rafael Nadal and Novak Djokovic (known as the "Big 3") have won a vast majority of Grand Slam Tournaments between 2003-2022. Do their serve statistics at their favorite Grand Slams explain their success?

How Important is the Serve in Tennis?

Serve Stats Correlations



Initial Insights: Which of the Three Serve Statistics is Most Important?

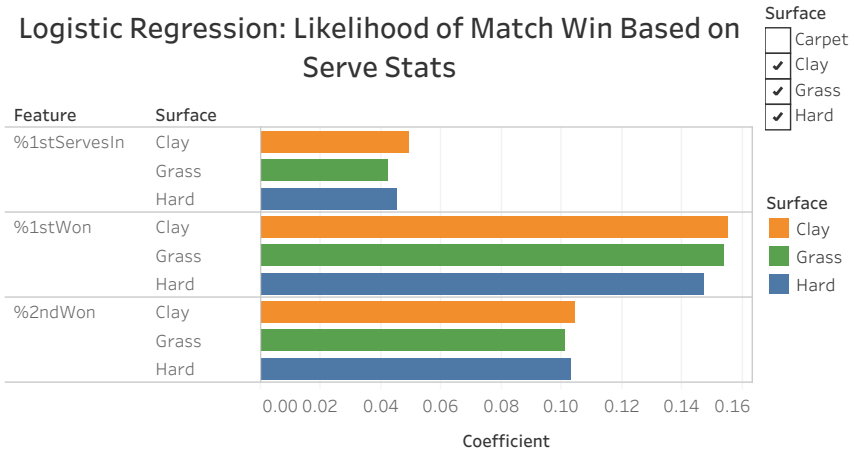
1. A player's **percentage of points won both on 1st and 2nd serve** had a **negative correlation to the number of break points** a player faced in a match, and the **negative correlation was stronger for the first serve**. This was **true for both the winner and loser** of a match, with the **correlation being stronger for the winner.**

2. A player's **percentage of first serves in play** had **no correlation to the number of break points** they faced.

Next Step: Observe the impact of each statistic on the likelihood of winning a match

* Players are potentially at a disadvantage when losing games on their own serve and should ideally limit the number of break points they face.

How Important is the Serve in Tennis?



Insights: Correlation Coefficients

Using logistic regression, the three serve statistics were tested to see which had the greatest impact on winning a tennis match.

1. On all surfaces, a player's **percentage of points won both on 1st and 2nd serve** have a **higher correlation coefficient to the likelihood of winning a tennis match**, confirming the initial insights.

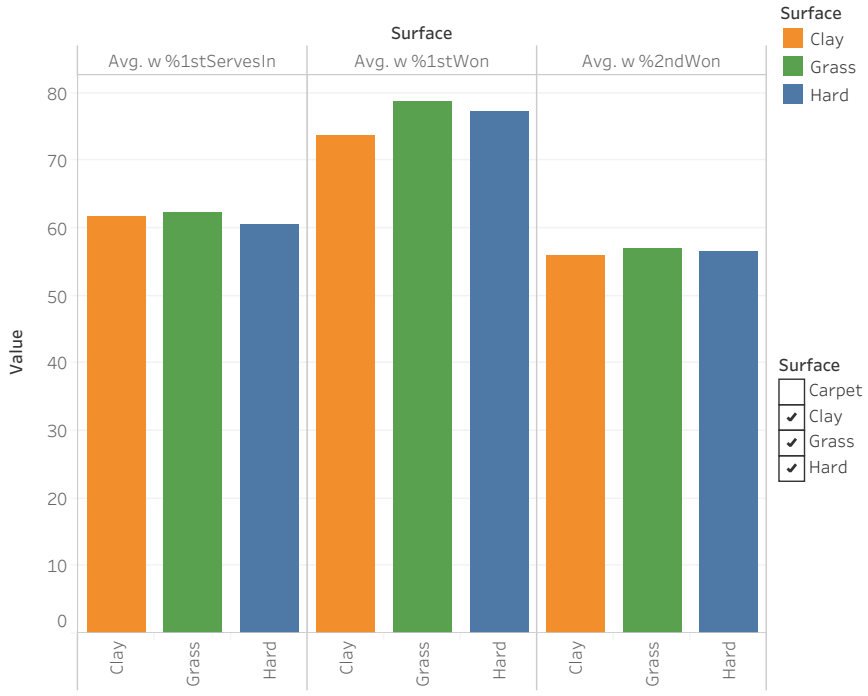
2. The **correlation coefficient was greater for 1st serve points won** than second serve points won.

3. A player's **percentage of first serves in play** had **no relationship** to their likelihood of winning a match.

Next Step: Observe differences in serve statistics across the three main surfaces

How Important is the Serve in Tennis?

Match Winner's Serve Statistics on Different Surfaces



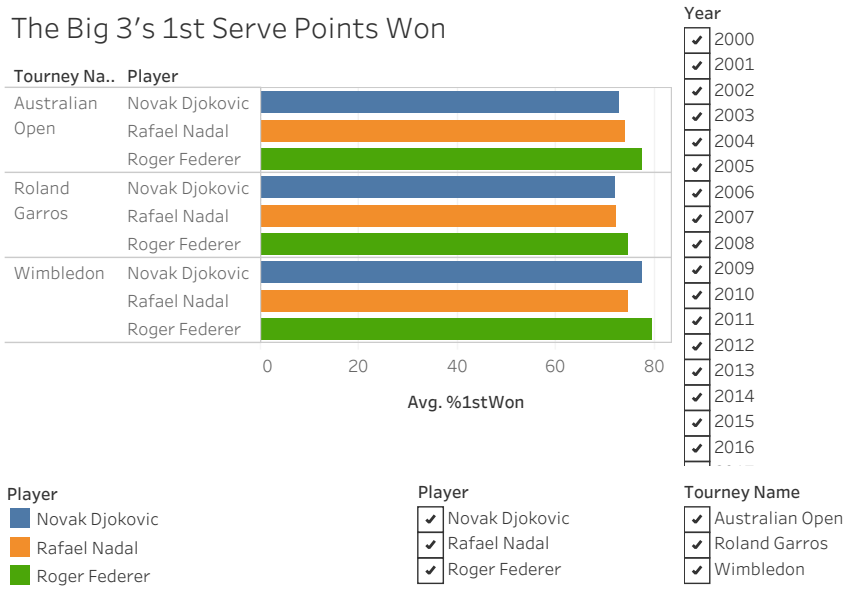
Insights

1. On grass courts, match winners won the highest percentage of points on both 1st and 2nd serve.

2. On clay courts, match winners won the lowest percentage of points on both 1st and 2nd serve.

3. The average points won by the match winner on 1st serve varied greater by surface than points won on 2nd serve

How Important is the Serve in Tennis?



Grand Slam Titles by the Big 3 (Excluding the US Open)

Australian Open (hard court)
Novak Djokovic: 9 Titles ('08, '11, '12, '13, '15, '16, '19, '20 and '21)
Roger Federer: 6 Titles ('04, '06, '07, '10, '17 and '18)
Rafael Nadal: 2 Titles ('09 and '21)

Roland Garros (clay court)
Rafael Nadal: 14 Titles ('05, '06, '07, '08, '10, '11, '12, '13, '14, '17, '18, '19, '20 and '22)
Novak Djokovic: 2 Titles ('16 and '21)
Roger Federer: 1 Title ('09)

Wimbledon (grass court)
Roger Federer: 8 Titles ('03, 04, '05, '06, 07, '09, '12 and '17)
Novak Djokovic: 7 Titles ('11, '14, '15, '18, '19, '21, and '22)
Rafael Nadal: 2 Titles ('08 and '10)

* The data does not include tournaments played after July 2022

How Important is the Serve in Tennis?

Final Insights

1. Of the three serve statistics, a player's percentage of points won on first serve had the greatest correlation to both the number of break points a player faced in a match, as well as the match outcome.

2. The player's percentage of first serves in play did not have an observable correlation to the match outcome.

3. Match winner's won the highest percentage of points on their first serves on grass courts and the lowest on clay courts.

4. The correlation between percentage of points won and likelihood of a match win was similar accross the 3 surfaces

5. Of the Big 3 Federer had the highest percentage of first serve points won. Nadal and Djokovic averaged lower at the Grand Slams than the average ATP tour match winner (on the corresponding surfaces).

6. The Big 3 all won the highest percentage of 1st serve points at Wimbledon (grass courts) and the lowest percentage at Roland Garros (clay courts). This is consistent with the average match winner's statistics on those surfaces.

Suggestions for Further Analysis

As this project is strictly for the curiosity of tennis fans, there are infinite ways to conduct further analyses depending on one's specific interests. I am providing the following suggestions:

1. Focusing solely on the serve cannot explain match outcomes. A more comprehensive study and dataset involving other match statistics (such as groundstroke winners and unforced errors)

2. One can more comprehensively analyze the Big 3's serve statistics at their favorite grand slams by observing if their serve statistics were better during the years they won a specific grand slam (or if they made it far into the tournament versus when they lost early on).

3. Matches played in Grand Slam Tournaments are played best out of 5 sets while in a majority of other tournaments, they are best out of 3 sets. It would be interesting to see if serve statistics differ depending on the length of the match.

4. Various elements can be tested to see if they are affected by court surface. Such elements include: use of spin, rally length, ratio of winners to errors, etc.

5. This project only used one of the original datasets CSV files. One of the other files included information on player ranking. One could ob explore how the importance of the serve varies depending on the level of a match (based on ranking).