Tristan Savella
CareerFoundry: Achievement 6
TableauStoryboard Overview

How Important is the Serve on the ATP Tour?

Table of Content

Project Overview:

The serve is one of the most important strokes in tennis. The player who is serving usually has the advantage in a rally. Using data from the ATP Tour (professional men's tennis), this project will explore three of the most common statistics, listed below, regarding the serve and their impact on the likelihood of winning a match:

1) Percentage of Points Won on First Serve
2) Percentage of Points Won on Second Serve
3) Percentage of First Serves in Play

Second, this project will also explore how or if each of the three statistics varies between the three most commonly used court surfaces: grass, hard and clay.

Last, this project will explore the serve statistics of Roger Federer, Rafael Nadal and Novak Djokovic at three of the four Grand Slam Tournaments; Wimbledon, Roland Garros and the Australian Open. These three players are arguably the three greatest players of all time and collectively known as the "Big 3". Each of them has the record number of titles at one of the aforementioned Grand Slams (Federer at Wimbledon, Nadal at Roland Garros and Djokovic at the Australian Open). Each of these slams is also played on one of the three aforementioned surfaces. Their serve statistics will then be compared to the average match winner, taking court surface into consideration.

Project Questions:

1) Of the three listed serve statistics, which has the greatest impact on the likelihood of winning a tennis match on the ATP Tour?
2) Did the impact of the most impactful serve stat (determined in question 1) change depending on the court surface?
3) Novak Djokovic, Roger Federer and Rafael Nadal are collectively known as the BIg 3 and as of May 2024, they hold the record number of most Grand Slam Tournaments Won (Djokovic at 24, Nadal at 22, and Federer at 20). How did the Big 3's serve statistics compare to average winners at their most successful Grand Slam Tournaments (which are the Australian Open, Wimbledon and Roland Garros, respectively)?

Introduction: Why This Project

Who is this project for:
This project is intended for tennis enthusiasts who follow the ATP Tour and enjoy match analyses.

Why is this being built:
The purpose of this project is to provide tennis enthusiasts with a better understanding of how the quality of a player's serve throughout a match impacts their likelihood of winning, and which elements of the serve in particular to take note of.

What information will this storyboard provide:

This project will explore three main questions:

1) Which of the three serve statistics (listed in the "Project Overview" section) has the greatest impact on the likelihood of winning a match on the ATP Tour?
2) Did any of the serve statistics (or the correlation between one of these stats and likelihood of winning a match) change depending on the court's surface?
3) Do the Big 3's serve statistics at the Grand Slam Tournaments explain their success?

Where will this dashboard be hosted
This dashboard will be hosted on Tableau Public

Data Limitations:
1) This dataset only focus on the serve; serve alone cannot summarize the outcome of matches
2) There are no match statistics from matches prior to 1991
3) Match statistics are also missing from certain tournaments/matches

Slide 2: Correlation Heatmap & Initial Insights

Dataframe/Subset Used: “df_matchstats2’ (removing all irrelevant and non-numerical values from the main dataframe.

Goal: An initial exploration of the correlation between different serve statistics (by both the winner and loser of a match).

Insights: Players are at a potentially big disadvantage when losing games on their own serve and should try to reduce the number of breakpoints that they face.

1) A player’s percentage of points won on both 1st and 2nd serve had negative correlation with the number of breakpoints they faced. The negative correlation was stronger for the winner than the loser of a match
2) There was no correlation between a player’s percentage of 1st serves in play and the number of breakpoints that they faced

Next Steps: Wrangle the data in order to observe the correlation between each of the three serve statistics and the likelihood of winning a match

Slide 3: Logistic Regression - Impact of Serve Stats on Match Outcome

Dataframe/Subset Used: 'coefficients_df' (wrangled from Notebook 3b)

Goal: To check the correlation between each of the serve statistics and the likelihood of winning a match on the ATP tour

Insights:

1) The percentage of points won on 1st serve had the strongest correlation to the likelihood of match win (the coefficient is approximately 0.15)
2) The percentage of points won on 2nd serve also had a weak correlation to the likelihood of match win (approximately 0.10)
3) There was no correlation between match win and the percentage of 1st serves in play (approximately 0.05)

* The same graph can be found on the introduction slide without taking court surface into consideration.

Next Steps: Observe differences in serve statistics across the three surfaces.

Slide 4: Serve Statistics by Surface

Dataframe/Subset Used: Main dataframe, ‘df_matchstats’

Goal: To explore if serve statistics vary by surface and gain insight into the differences of the impact of the serve on each surface.

Insights:

1) Players won the highest percentage of 1st and 2nd serve points on grass courts, followed by hard courts. The lowest percentage was on clay courts
2) The percentage points won on 1st serve varied greater by surface than the percentage of 2nd serve points won.

.

Next Steps: Observe the statistics of the Big 3 at their favorite grand slams.

Slide 5: The Big 3's Serve Statistics at Their Favorite Grand Slams

Dataframe/Subset Used: 'result_df'

Goal: Federer has the record number of titles at Wimbledon (grass courts); Nadal at Roland Garros (clay courts); and Djokovic at the Australian Open (hard courts). The dashboard on this slide allows the user to observe each player's average percentage of points won on 1st serve at each of the Grand Slams. The filters provided allow users to click on certain years, in order to observe each player's average statistic during a specific edition (year) of each tournament.

Additionally, I have listed all of the titles each player won at one of these three tournaments (the last recorded tournament is Wimbledon 2022), which can help users compare serve statistics during the years when a player won the tournament versus when they did not.

Slide 6: Conclusion

Goal: To provide final insights and suggestions for further analysis