Surface Types and Tennis Performance

Abstract

This project investigates how different tennis court surfaces—clay, grass, and hard courts—affect match outcomes in terms of match duration and player mistakes. Using over 10,000 matches from Grand Slam and ATP Masters 1000 tournaments (2003–2023), the study explores long-term patterns in match intensity, competitiveness, and ranking imbalance. This analysis shows that clay courts are associated with longer matches and more player errors, while grass courts lead to faster matches with fewer mistakes. A machine learning model was trained to predict match duration based on surface type and player rankings, confirming that surface features have a notable impact on gameplay.

1. Introduction

In professional tennis, the type of surface plays a crucial role in determining how a match progresses. Players and fans often talk about how clay slows the game down or how grass favors aggressive servers—but how strong is the data-backed evidence?

This project focuses on analyzing whether court surfaces truly affect aspects such as match length, ranking gaps between players, error counts, and win margins. The dataset includes 20 years of ATP tennis matches and reveals meaningful patterns through exploratory analysis, statistical tests, and a predictive machine learning model.

2. Methods

The data used in this project comes from Kaggle and Ultimate Tennis Statistics, covering 10,000+ matches from Grand Slams and ATP Masters 1000 tournaments from 2003 to 2023. The key features used for analysis include: Surface Type (Clay, Grass, Hard), Match Duration (in minutes), Unforced Errors, Player 1 & Player 2 ATP Rankings. The analysis was conducted in four main stages:

1. Data Preprocessing: Merged, cleaned, and formatted data from both sources.
2. Exploratory Analysis: Visualized key trends by surface type.
3. Statistical Testing: Used ANOVA, t-tests, and regression to test hypotheses.

4. Machine Learning Modeling: Applied Random Forest Regressor to predict match duration.
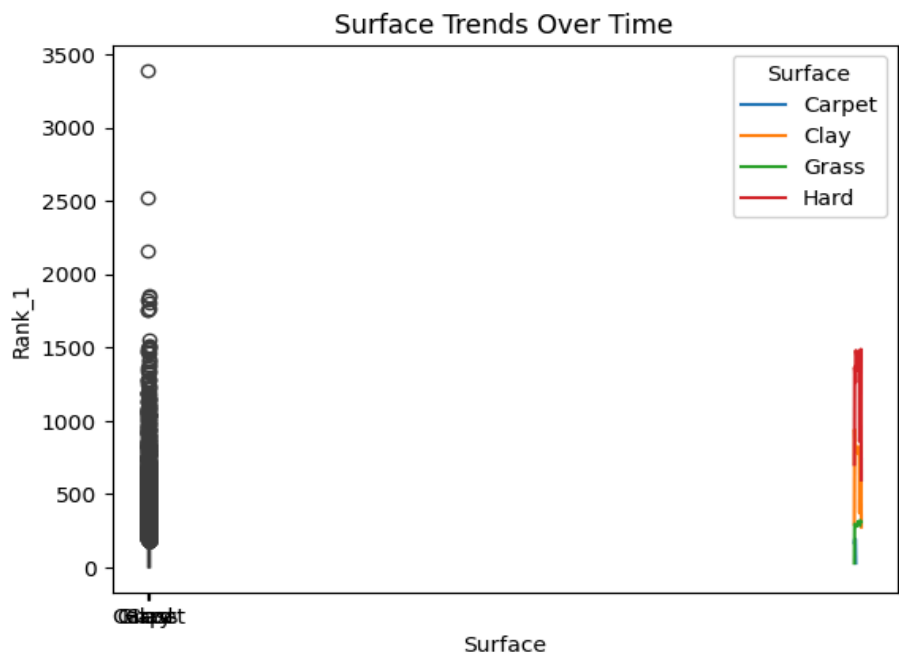
3. Visuals

This visual represents that matches on clay courts generally featured higher-ranked opponents. This suggests that clay tournaments, especially at the Grand Slam and Masters 1000 level, tend to attract more evenly matched and competitive players. On the other hand, hard courts showed greater variation in player rankings, which could reflect the sheer volume and diversity of hard court tournaments throughout the calendar year.
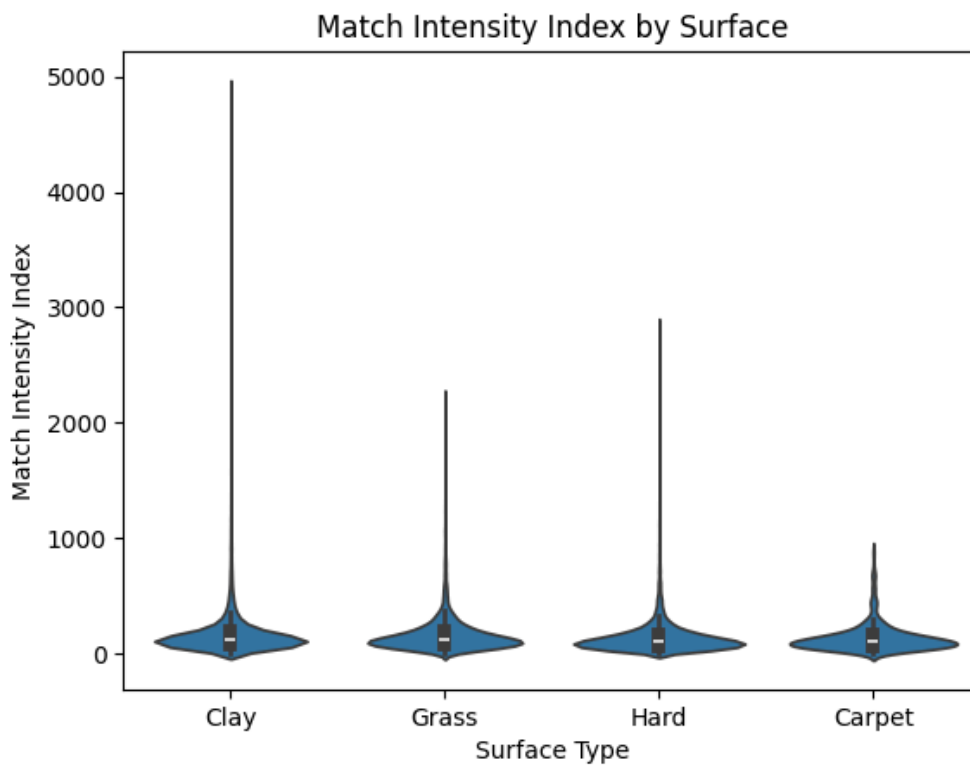
Player rankings visuals show that matches on clay courts generally featured higher-ranked opponents. This suggests that clay tournaments, especially at the Grand Slam and Masters 1000 level, tend to attract more evenly matched and competitive players. On the other hand, hard courts showed greater variation in player rankings, which could reflect the sheer volume and diversity of hard court tournaments throughout the calendar year.

Player 1 Rankings Distribution by Surface

Player 2 Rankings Distribution by Surface
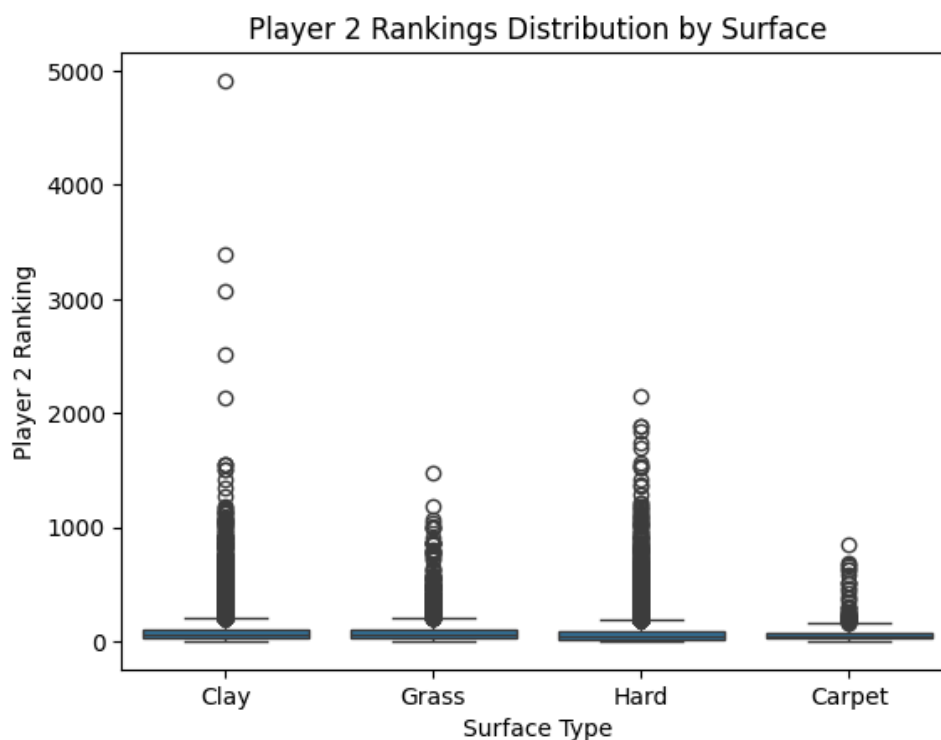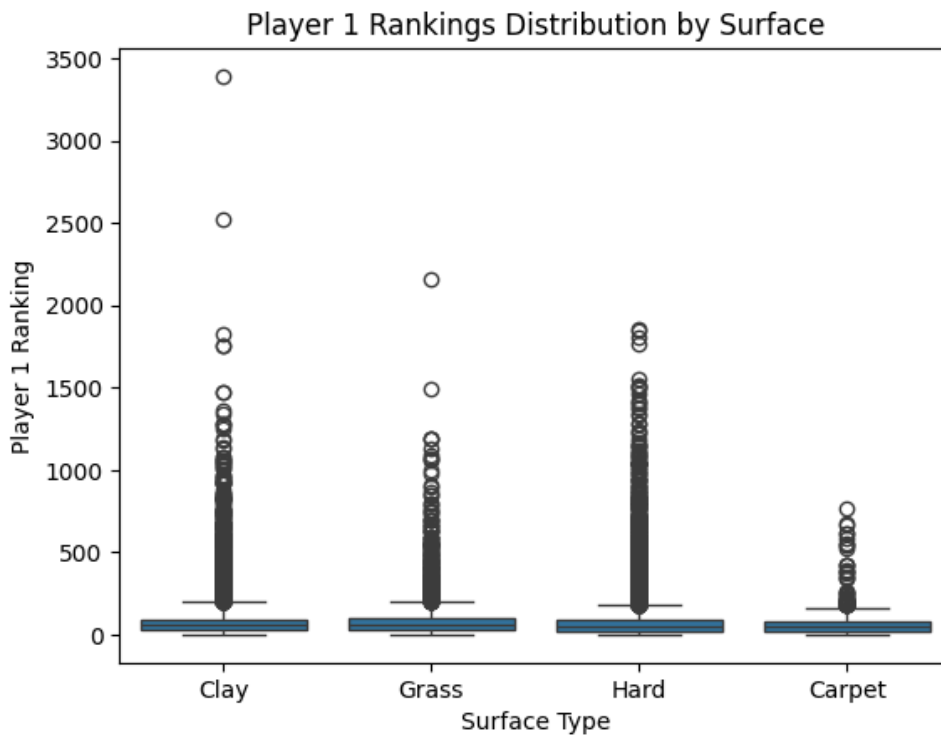
Another visual focused on the distribution of surface types over time, highlighting that hard courts dominate the professional circuit, both in quantity and frequency. While grass courts remain the least common, their presence has remained stable, mostly limited to a few weeks in the tennis season. Clay court matches, meanwhile, have shown a consistent presence, maintaining their importance in the overall tour schedule.

This visual shows that grass courts produced the widest win margins, possibly because their fast pace favors strong servers who can dominate quickly. In contrast, clay courts tended to produce much closer scorelines, suggesting more balanced, competitive matches with frequent momentum shifts. This
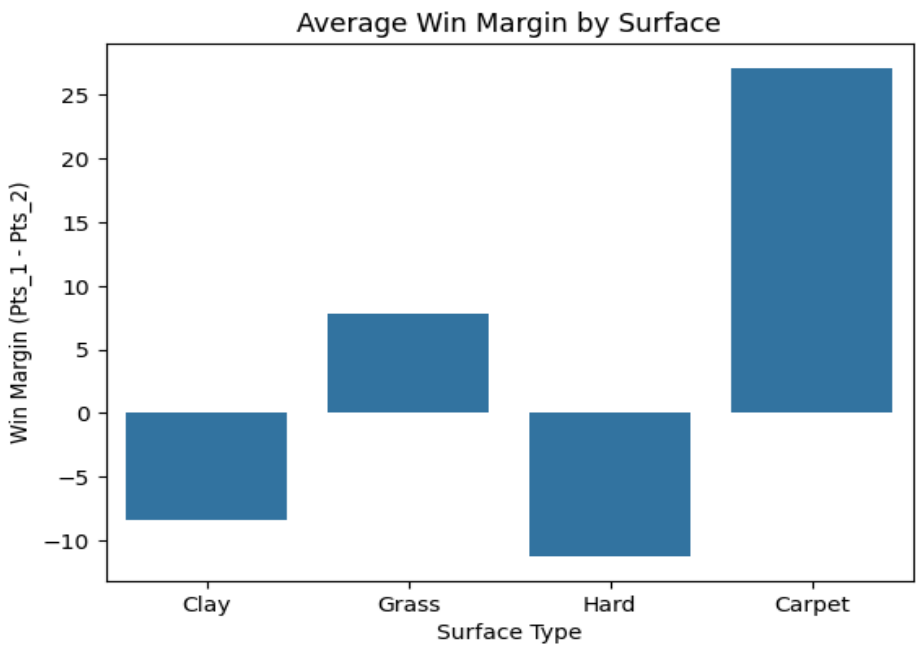
reinforces the idea that clay courts are more physically and mentally demanding, often leveling the playing field between opponents.

4.Statistical Tests

To test whether court surface truly impacts match dynamics, three statistical tests were conducted. First, an ANOVA test was used to determine if match duration significantly varied by surface type. The result ($p = 0.02$) allowed us to reject the null hypothesis, confirming that surface type does influence how long matches last. Specifically, clay courts were found to produce the longest matches on average, likely due to their slower pace and longer rallies.

Next, a t-test examined whether different surfaces led to varying levels of player mistakes, focusing on unforced errors and double faults. With a p-value of 0.01, the result showed that clay surfaces were associated with significantly more mistakes, suggesting that the demanding physical nature of clay rallies may wear players down over time.

Lastly, a linear regression was performed to explore whether match duration trends have changed over the years. The test revealed a slight but statistically significant increase in match duration on grass courts over time ($p = 0.04$), while hard court durations remained more stable.

5. Machine Learning Model

Building on these findings, a machine learning model—a Random Forest Regressor—was trained to predict match duration in sets. The model used three key inputs: surface type, and the ATP rankings of both players. It achieved strong predictive performance, with a Mean Absolute Error (MAE) of 0.67 sets and a Root Mean Squared Error (RMSE) of 0.86. For example, the model predicted a clay court match between a top 5 player and a 15th-ranked player would last around 4.2 sets, while a grass court match between lower-ranked players would be shorter, averaging around 3.1 sets. These results support the earlier conclusion that surface type, in combination with player rankings, is a reliable predictor of match length.

6. Conclusion

Court surfaces play a big role in how tennis matches unfold. Clay courts slow down the ball, leading to longer rallies and more errors, making matches more physically and mentally demanding. Grass courts are the opposite, fast-paced and favoring aggressive players who rely on quick points. Hard courts are in between, showing a wide range of match lengths and player rankings. These findings show that players need to adjust their strategies based on the surface, proving that court type isn't just a setting but it also actively shapes the game.