Tennis Trends Over Time

Christopher Lee and Ryan Oscherwitz

Overview: This project explores various tennis trends and metrics over time, as well as forecasts them to better understand how the future of tennis might look. We analyze three metrics – US Open duration, serve rating, and break points won. The dataset was gathered from Kaggle and contains over 350,000 individual matches in the ATP from 1991 to 2018. Some of the questions we hope to answer with our data are "*Are matches getting shorter compared to before?*", "*How has the importance of the serve changed over time?*", "*Are today's players better servers?*", and "*Are servers more dominant than returners?*".

Methods: To historically analyze the data, we group observations by year and record the average metric for that year, resulting in an average metric over the years. With this summarized data, we use the locally estimated scatterplot smoothing (LOESS) technique to tune out noise in the data and visualize a general trend. We also employ the autoregressive integrated moving average technique to generate a forecast for each metric.

Results: Using LOESS, it is clear that the average duration of matches at the US Open has been decreasing. While there may be erroneous data, the average duration of a match at the US open has been around 2.25 to 2.50 hours, while more recent years (post 2010) have been around two hours. The LOESS curve also shows that the rate at which the duration has been decreasing is slowing, indicating that matches may normalize around two hours.

As for serve rating, court surface greatly impacts a players' serve rating. There appears to be a decline in serve rating overall after 2005. This could be due to the implementation of hawk eye – an automated line judge. In more recent years, serve ratings appear to be trending upwards.

Break points won, similar to duration, also appears to be declining as time progresses, stabilizing around 2010. This indicates that the server likely dictates the point and game more, leading to fewer break points.

Using the ARIMA model, we find visuals of how the average match duration at the U.S. Open has changed over the past 27 years and is forecasting the next five years. We found three outliers, but other than those there is not a noticeable trend. For this reason, the forecasted values are not very insightful as it is essentially a horizontal line.

From there, we move to examining the average serve rating over time. There is a slight bell curve shape in our plot. Again, the forecasted estimates are just a horizontal line. However, the prediction intervals are more interesting as they start small and become wider over time.

We finally look at average break points won over time. This plot shows a clear downward trend. As such, the forecasted values are also trending downwards.

Conclusion: We analyzed match duration, serve rating, and break points won from 1991 to 2018. We found that match duration has recently plateaued around two hours in the U.S. Open; however, has significantly decreased compared to the 1990s. Serve rating was increasing during the 1990s, but declined around 2010. Recently, however, serve rating has been in an uptrend, indicating that the metric could return to its 2005 highs. Finally, break points have been a clear downtrend, falling from around 3 per match to 2.7. While these differences are statistically significant when grouping every 10 years, there is not a huge meaningful impact on the game. In the future, we would like to see if there are any interesting differences in the trends between men's and women's tennis. Additionally, it would be insightful to look at how prize money has changed over time.