Game On: Unraveling the Rivalry Between Men's and Women's Tennis

Ava Scharfstein, Bachelor of Arts Candidate for Mathematical Data Science

Mentors: Dr. Peter Mucha and Molly LaBelle (Sportradar Data Scientist)

Abstract:

This project is motivated by gender disparities, namely the treatment and discourse of male versus female athletes in tennis' most prestigious international tournaments. After mining match-level data from the Sportradar Tennis API and historical rankings, I investigated the entertainment value of tennis matches by gender through an exploratory data analysis and hypothesis testing. Findings suggested that men's tennis showcases greater competitiveness with players having similar match-level statistics, particularly excelling in service games, while women's tennis leans towards risk-taking, producing more winners and longer games with more unforced errors. Thus, the data shows that neither genders' matches are statistically more entertaining, as they offer different dynamics and value to viewers.

Introduction:

Despite the tremendous strides made by equal pay activist Billie Jean King, there remains a striking disparity in the treatment of male and female players in professional tennis. In many highly prestigious international tournaments, women earn a smaller share of winnings compared to their male counterparts. Women also receive second billing in mixed tournaments – less desirable schedules on smaller courts. The justification for these inequities hints at the systemic gender discrimination within professional sports: the perception that men's matches hold greater attraction and appeal than women's.

Project Focus: Investigates, quantifies, and interprets the conception of *appeal* in men's versus women's tennis, using a data-driven approach.

Appeal Through Competitiveness: A competitive match is closely contested and evenly balanced, with an uncertain outcome and momentum shifts keeping spectators engaged and excited.

Big-Picture Objective: Analyzing the appeal of tennis by sex to counter gender-based biases in media rhetoric and player treatment.
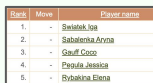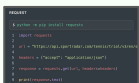
Data:

Data Extraction:

As a former data science intern at Sportradar, I was able to obtain free access to the Sportradar[1] Tennis application programming interface (API).

Match-level Data: Built a streamlined Python pipeline to navigate the complex, nested API and extract data for nearly 1000 matches.

Rankings: Scraped historical ATP (Association of Tennis Professionals, for Men) and WTA (Women's Tennis Association) rankings online to fill gaps, as tournaments only seed top players.

Data Processing:

Data Merging and Cleaning: Merged match-level data with historical ATP and WTA rankings.

Cleaning Process: Addressed missing data, normalized continuous variables, and aggregated, removed, or renamed variables to reduce noise and enhance clarity.

Iterative Refinement: The process was iterative, with adjustments made as new issues emerged during initial exploratory analyses.

Snapshot of cleaned data set

Methods:

Quantifying Competitiveness: The difference between two players individual match statistics (eg. aces, unforced errors, service points won) was called the ***stat-difference***. For example, when a certain player's stat is equivalent to their opponent – yielding a net zero difference – the players may be considered well-matched and competitive. Conversely, when one player's stat is much larger than their opponent – yielding a net non-zero difference – the match may be considered to be dominated by one player and less competitive – in regards to that specific statistic.

Additional Metrics: The total of two players individual match statistics, called ***stat-sum***, represents the degree to which the match demonstrates high skill levels. It does not necessarily reflect the competitiveness of two players, but rather is a supplementary measure to stat-difference, as outlined in the results section.

Hypothesis Testing: I compared the distributions of stat-differences for two dozen different tennis statistics by sex. I used the hypothesis tests enumerated below as well as empirical cumulative distribution functions, qq-plots, and histograms to aid in my analyses.

1. **Kolmogorov–Smirnov test:** a nonparametric test which assesses the whether two distributions are derived from the same underlying distribution
2. **T-test:** assess the equality of the means between two groups, assuming normativity
3. **Mann Whitney U test:** a non-parametric test used to compare two independent samples, generally as assessing equality of the medians
4. **Levene test:** assess the equality of the variances between two or more groups

Results:

Levene Tests: For stat-differences, there were statistically significant differences in variance between male and female distributions in **21 out of 24 features**, with men showing smaller variance. Since variance quantifies the dispersion of player similarity, smaller variance means that stats are clustered closer around the mean. This indicates a more consistent performance or style of play among the players and **suggests that men's matches are generally more competitive**. Notably, men exhibited significantly smaller variance in unforced errors (p = 6.74449e-16) and winners (p = 2.54959e-10), while only aces had greater variability for men (p = 1.82336e-10).
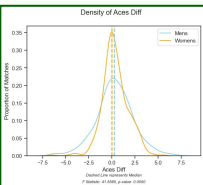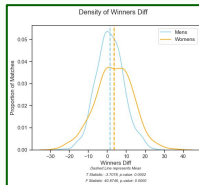
T-Tests: For stat-differences distributions, the mean represents the average similarity between opponents; a positive mean indicates higher ranked players dominate on average. T-tests for unforced errors (p = 0.00281319) and winners (p = 0.000221226) revealed statistically significant differences across gender. Unforced errors, which are errors that occur without pressure from the opponent, can be due to player inconsistency *or* risk-taking. A positive mean for women and negative mean for men indicates that **higher-ranked female players might be more willing to take risks compared to their male counterparts**, while lower-ranked male players take more risks than their female counterparts. Higher and positive mean for winners for women than men suggests that **higher ranked female players are more likely to hit winners than their opponents compared to men**.

Mann Whitney U: Mann-Whitney U tests showed statistically **significant differences in all serving-related features for stat-sums** (e.g., first serve points won, second serve points won, aces, service games) between the genders, with men having higher medians. This aligns with the well-known fact that men are stronger servers and more likely to hold their service games, which is crucial for maintaining a lead or staying even in a match. On the other hand, women have a greater median for total break points, indicating more challenges in holding serve and suggesting **more unpredictable service games**. This might be due to aggressive returns or the ability to exploit opponents' weaknesses. Women also had more unforced errors and winner totals, reinforcing **higher risk-taking**. Additionally, women's matches had higher median points per game and average game length, highlighting **greater intensity and competitiveness at the game level**.

Results (continued):
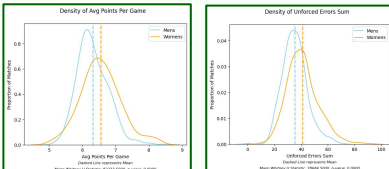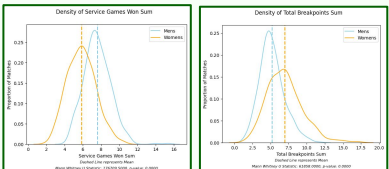
Stat-sum/Avg Feature	P-value	Men > Women
1st serve points won	9.6482e-23	True
2nd serve points won	1.1289e-12	True
Aces	8.3515e-65	True
Service games won	1.6928e-54	True
Total break points	1.9145e-32	False
Unforced errors	4.4393e-15	False
Winners	.01198	False
Avg game length	.00238	False
Avg points per game	2.648e-12	False

Mann Whitney U P-value Table

Conclusion:

While the results show that men's tennis tends to exhibit greater competitiveness in terms of players having similar statistics, it's worth noting that dominance can manifest in various ways, and one player may excel in certain aspects while another excels in others. Men typically outperform in service games and exhibit greater consistency, but are more prone to forced errors. Conversely, women tend to outperform in risk-taking, producing more winners and longer, more unpredictable and closer games, albeit with more unforced errors. Ultimately, I claim that **neither gender's matches are inherently more entertaining; they simply offer different dynamics and value to viewers**.

Future Work:

With more time and resources, I would have liked to explore additional statistics such as viewership, match attendance, prize money, and various match dynamics, including deuces, match points, and the percentage of points won on the opponent's serve. Analyzing the frequency of upsets and variability across Grand Slam tournaments, as well as differences in early versus later rounds, could reveal patterns in competition and unpredictability. Moreover, focusing on the top 20 players from each gender could uncover insights into their playing styles, strengths, weaknesses, and overall consistency, enhancing my understanding of factors contributing to success and entertainment value in tennis by sex.

Acknowledgements: