



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.

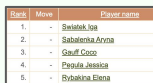
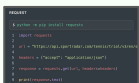
Data:

Data Extraction:

As a former data science intern at Sportradar, I was able to obtain free access to the Sportradar 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.

```
name_home  aces_home  breakpoints_won_home  ...  name_away  aces_away  breakpoints_won_away
0  Kudermertova, Veronika  7.500000  ...  Kangni, Kaia  1.500000  0.000000
1  Samsonova, Liudmila  2.500000  ...  Bogdan, Ana  3.000000  1.500000
2  Paric, Aljicia  2.000000  ...  Friedsam, Anna-Lena  1.000000  0.000000
3  Rakhimova, Kanilla  0.666667  ...  Bucsa, Cristina  2.000000  2.000000
4  Pegula, Jessica  1.000000  ...  Davis, Lauren  0.666667  0.666667
```

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.

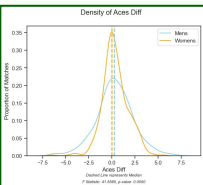
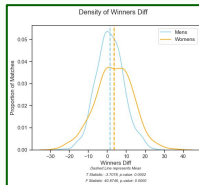
Additional Metrics: The total of two players individual match statistics, called **stat-sum**, represents the degree to which the match demonstrates high skill levels. It does not necessarily reflect the competitiveness of two players, but rather is a supplementary measure to stat-difference, as outlined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

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

Results:

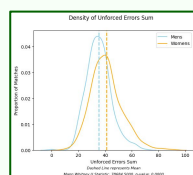
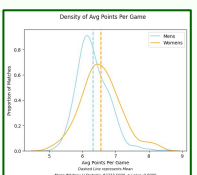
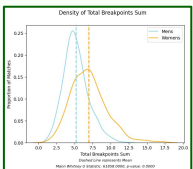
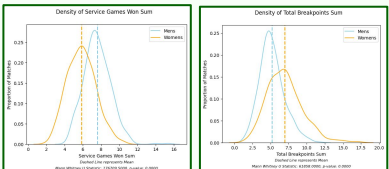
Levene Tests: For stat-differences, there were statistically significant differences in variance between male and female distributions in **21 out of 24 features**, with men showing smaller variance. Since variance quantifies the dispersion of player similarity, smaller variance means that stats are 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 ($p = 6.74449\text{e-}16$) and winners ($p = 2.54959\text{e-}10$), while only aces had greater variability for men ($p = 1.82336\text{e-}10$).



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 (eg., 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-name/Avg Feature	P-value	Men > Women
1st serve points won	9.6462e-23	True
2nd serve points won	1.1289e-12	True
Aces	8.3515e-47	True
Service games won	1.6928e-34	True
Total break points	1.9145e-12	False
Unforced errors	4.4395e-11	False
Winners	.01198	False
Avg game length	.00238	False
Avg points per game	2.646e-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:

Sportradar is a leading sports technology company that I interned for from January-March (23W) and June to August 2023 (23X). They kindly granted me access to their Tennis API for free. Thank you to Molly LaBelle and the NY Office at Sportradar for their help in facilitating this project and helping me resolve API bugs throughout the process. Thank you also to Peter Mucha for his support despite taking leave during 24W and Eugene Demidenko for his feedback on my initial draft.

Data Sources: [ATP](#), [WTA Tennis Explorer](#), and [Sportradar](#).



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

Ava Scharfstein, BA in Mathematical Data Science

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

Abstract:

This project is motivated by gender disparities in tennis, namely the treatment and discourse surrounding male versus female athletes in tennis' most prestigious international tournaments. After mining match-level data from a Sportradar API and historical rankings from ATP and WTA, I investigate the entertainment value of tennis matches by gender through an exploratory data analysis and hypothesis testing. Findings suggested [Include results]. [Say something about the impact] (...I don't have a concise story quite yet...)

Motivation:

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. This project takes a data-driven approach to investigate, quantify, and interpret the conception of 'appeal' through a variety of factors. I considered **competitiveness** (player similarity) and **upsets** (when a lower ranked player triumphs) as two potential predictors of entertainment.

Data:

Data Extraction: I obtained access to Sportradar's Tennis API, which contains detailed match information and an array of supplementary data. With help from the API documentation, I built a streamlined pipeline in python that navigated Sportradar's complex and nested API to extract match-level data for close to 1000 Grand Slam matches. I also scraped historical ATP (Association of Tennis Professionals, for Men) and WTA (Women's Tennis Association) data from their websites to build a time-stamped dataset of men's and women's rankings.

Data Cleaning: After merging the two data sets, I undertook a lengthy data-cleaning process, which included handling missing data, normalizing continuous variables, and aggregating, removing, or renaming variables to minimize noise and improve clarity. The process became iterative, especially as new issues arose during the initial exploratory analyses.

Methods:

When a player's stat is equivalent to its opponent [yielding a net zero difference], the players may be considered well-matched and the match **competitive**. When a player's stat is much larger than its opponent [yielding a net non-zero difference], the match may be considered to be **dominated** by one player and less 'competitive'. For simplicity, I denote this difference between a player's stats as **stat-difference** where the stat might be the number of aces, breakpoints, or first serve points won (stats are normalized by the number of sets played). In my exploratory data analysis, I look at the distribution of these stat-differences between home and away players for a set of matches. First, I considered the absolute value of the stat-differences, thereby ignoring which player was dominating (indicated by the sign). However, this yielded a set of folded distributions (all data points to the left of $x=0$), which have a plethora of limitations. As a result, I relabeled the cleaned dataset to ensure that the home-player is always the higher ranked player and the away-player is lower ranked. For twenty four different features (match stats), I plotted the empirical cumulative distribution functions, qq-plots, and histograms for both male and female matches. I also conducted four hypothesis tests to compare the male and female distributions:

1. **Kolmogorov-Smirnov test:** a nonparametric test which assesses the whether distributions are derived from the same underlying distribution
 2. **T-test:** assesses equality of means between two groups, assumes normality
 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:** assesses equality of variances between two or more groups
- I conducted a similar analysis for the distribution of **stat-sums** [stat totals for both players] to aid in the interpretation of the stat-differences.

An **upset** or unexpected win occurs when a lower ranked player triumphs over a higher ranked player, particularly if the difference in their ranking is significant...

Results: Disclaimer - major work in progress

Across almost all the features for stat-differences, the Levene tests revealed statistically significant differences between the variance for male and female distributions, and variance was smaller for men's stat-differences. The variance is the spread of players' similarity, where the larger the spread, the less similar the opponents tend to be, for a given stat. This suggests that men's matches may be more competitive than women's. For example, unforced errors-difference and winners-difference had p-values of $6.7449e-16$ and $2.54959e-10$ respectively for the Levene test. Aces-difference is the only feature for which the opposite was the case. That is, the female distribution has a smaller variance, with p-value of $1.82336e-10$. For a selected number of features, there is a statistically significant difference between the mean for male and female distributions. The mean represents the average similarity of opponents; a positive mean being more home-dominant and a negative mean being more away-dominant. For example, the unforced errors-difference is statistically significantly distinct for men and women, where men have a negative mean and women a positive with a p-value of $.0028132$ for the t-test. Since unforced errors are commonly associated with risk-taking, this result suggests that higher ranked female players on average are more likely to take risks compared to men. Across all of the serving-related features for stat-sums, the Mann Whitney U tests revealed statistically significant differences between the medians of the male and female distributions, with the median higher for men's stat-sum. This is an expected result, since men are known to be stronger servers and more likely to hold. On the other hand, the Mann Whitney U tests revealed significant differences between the medians of the male and female distributions, with the median higher for women for several key features: total break points, unforced errors, winners, game length, and points per game. These results suggest that there is more unpredictability in service games, more risk taking, more great shots, and longer games for women compared to men.

Conclusion:

Tbd on results

References:

Sportradar is a sports data analytics company that I interned for 23W and 23X. They kindly gave me access to their Tennis API for free.

Evaluating Tennis Entertainment By Gender

Ava Scharfstein, BA in Mathematical Data Science



Mentors: Peter Mucha (primary), Soroush Vosoughi (secondary),

Molly LaBelle (Sportradar data scientist)

Abstract:

This project addresses gender disparities in tennis, focusing on the unequal treatment of male and female athletes in prestigious international tournaments. Using data from Sportradar's tennis API and historical rankings, I assess the entertainment value of matches by gender, considering various competitiveness metrics. Through unsupervised clustering and dimensionality reduction, I aim to objectively compare the entertainment levels. The results may inform decisions regarding scheduling and court assignments to promote gender equity in tennis and enhance the fan experience.

Motivation:

In tennis, there remains a striking disparity in the treatment of male and female athletes. In highly prestigious international tournaments, women continue to receive a smaller share of the winnings compared to their male counterparts. Women also receive second billing in mixed tournaments – less desirable schedules on smaller courts. A glaring example from the 2022 French Open underscores this inequity. Amélie Mauresmo, the tournament director, notably allocated just one women's match to the prestigious nighttime slot, while nine men's matches took precedence. Her justification, that the men's game

Motivation (continued):

held greater attraction and appeal, only served to highlight the prevailing bias. Indeed, the chairman and CEO of the Women's Tennis Administration claims that the market values men's tennis more than women's, particularly in the context of sponsorships and media rights.

In this project, I aim to explore if men's tennis is truly more entertaining than women's. That is, evaluate the entertainment value of tennis matches by gender.

Data and Methods:

I just got access to the Sportradar tennis API (Sat) and have started extracting relevant data (where relevance is still TBD). I have also begun the process of mining historical tennis ranking data from WTA/ATP sites.

I may define competitiveness on the basis of number of games/set (or simply total number of games), number of breakpoints (refers to an opportunity for a player serving to lose their service game), difference in seed/rank (similar rank would probably be a more interesting match), average set length, average point length, whether the players are famous/top 20/won a grand slam before, match point in the middle of the match. Once acquiring, merging, and cleaning the data, I plan to take an unsupervised clustering approach to group matches by competitiveness and visualize the

Data and Methods (continued):

differences between male and female matches. In addition to various clustering techniques, I anticipate utilizing principal component analysis to perform dimensionality reduction for visualization purposes.

It may also be a point of interest to compare competitiveness between Grand Slam (where men play to five sets and women play to three) and lower level tournaments (where both genders play best of three). I anticipate that Grand Slam matches will be evaluated with a higher level of entertainment, which will then beg the question whether women should be playing five-set matches at Grand Slam tournaments too.

Results:

This will all depend on the data. Read [here](#) for more specific ideas

Conclusion:

The results of my analysis may be used to inform business decisions, such as rearranging court schedules to play more women's matches on center courts.

References:

Sportradar is a sports data analytics company that I interned for 23W and 23X. They are giving me access to their Tennis API for free.

Evaluating Tennis Entertainment By Gender

Ava Scharfstein, BA in Mathematical Data Science



Mentors: Dr. Peter Mucha (Major Advisor)

Connor Lewis & Molly LaBelle (Sportradar data scientists)

Abstract:

This project addresses gender disparities in tennis, focusing on the unequal treatment of male and female athletes in prestigious international tournaments. Using data from Sportradar's tennis API and historical rankings, I assess the entertainment value of matches by gender, considering various competitiveness metrics. Through unsupervised clustering and dimensionality reduction, I aim to objectively compare the entertainment levels. The results may inform decisions regarding scheduling and court assignments to promote gender equity in tennis and enhance the fan experience.

Motivation:

In tennis, there remains a striking disparity in the treatment of male and female athletes. In highly prestigious international tournaments, women continue to receive a smaller share of the winnings compared to their male counterparts. Women also receive second billing in mixed tournaments – less desirable schedules on smaller courts. A glaring example from the 2022 French Open underscores this inequity. Amélie Mauresmo, the tournament director, notably allocated just one women's match to the prestigious nighttime slot, while nine men's matches took precedence. Her justification, that the men's game

Motivation (continued):

held greater attraction and appeal, only served to highlight the prevailing bias. Indeed, the chairman and CEO of the Women's Tennis Administration claims that the market values men's tennis more than women's, particularly in the context of sponsorships and media rights.

In this project, I aim to explore if men's tennis is truly more entertaining than women's. That is, evaluate the entertainment value of tennis matches by gender.

Data and Methods:

I plan to acquire the data using Sportradar's tennis API as well as through mining historical tennis ranking data from WTA/ATP, and perhaps even tv ratings/spectator data.

I may define competitiveness on the basis of number of games/set (or simply total number of games), number of breakpoints (refers to an opportunity for a player serving to lose their service game), difference in seed/rank (similar rank would probably be a more interesting match), average set length, average point length, whether the players are famous/top 20/won a grand slam before, match point in the middle of the match. Once acquiring, merging, and cleaning the data, I plan to take an unsupervised clustering approach to group matches by competitiveness and visualize the

Data and Methods (continued):

differences between male and female matches. In addition to various clustering techniques, I anticipate utilizing principal component analysis to perform dimensionality reduction for visualization purposes.

It may also be a point of interest to compare competitiveness between Grand Slam (where men play to five sets and women play to three) and lower level tournaments (where both genders play best of three). I anticipate that Grand Slam matches will be evaluated with a higher level of entertainment, which will then beg the question whether women should be playing five-set matches at Grand Slam tournaments too.

Results:

This will all depend on the data. Read [here](#) for more specific ideas

Conclusion:

The results of my analysis may be used to inform business decisions, such as rearranging court schedules to play more women's matches on center courts.

References:

Sportradar is a sports data analytics company that I interned for 23W and 23X. They are giving me access to their Tennis API for free.