Australian Open: A Century of Performance Analysis with Tableau

Ella Lee

UTS CRICOS 00099F
UTS TEQSA PRV12060

Contents

1 Executive Summary

This report presents a comprehensive visual analysis of the Australian Open tennis championships, drawing from over 120 years of historical data. The primary goal is to identify and interpret patterns in player performance, gender-based trends, and national dominance, particularly focusing on champions who have won five or more titles.

Using Tableau as the primary visualisation tool, the report leverages advanced graphic techniques including parallel coordinates, treemaps, geographic maps, scatter plots, and word clouds to reveal both macro-level tournament trends and micro-level performance distinctions. These visualisations allow readers to explore key questions: *Which countries have dominated over time? How have gender dynamics evolved? Which players have shown the greatest consistency or outlier behaviour?*

Key insights include:

Australia and the United States dominate historically in both men's and women's categories, although emerging nations like Serbia and Belarus have gained ground in recent decades.

Male champions often exhibit greater variability in win rates and losses, while female champions display more consistency.

Certain individuals, such as Novak Djokovic, Margaret Court, and Serena Williams, demonstrate sustained excellence with high win rates over time.

Geographic and performance-based visualisations also highlight performance efficiency, not just frequency of participation.

These insights are made clearer through Tableau's ability to dynamically present high-dimensional data. Its interactivity and customisation features significantly enhance the readability, accessibility, and storytelling power of the report.

In summary, this report not only showcases the rich history of the Australian Open but also demonstrates how data visualisation tools can uncover meaningful narratives hidden within complex datasets. It offers value to analysts, sports enthusiasts, and decision-makers interested in understanding performance patterns and historical trends in elite-level tennis.

2 Introduction

The Australian Open, one of the four prestigious Grand Slam tennis tournaments, has a rich history spanning over 120 years. Throughout its evolution, it has witnessed numerous legendary players, shifting global dominance, and changing gender dynamics. This report aims to analyse this comprehensive historical dataset through visual analytics to extract insights into player performance, national success patterns, and comparative dynamics by gender.

The target audience of this report includes tennis enthusiasts, sports historians, and data analysts interested in exploring trends across decades of competition. The analysis focuses particularly on identifying high-performing players—defined here as those with five or more championship wins—and understanding the distribution and evolution of success among different countries and genders.

This report uses Tableau to create interactive and multi-dimensional visualisations, offering both high-level overviews and granular performance insights. The core objective is to not only visualise the data but also to enhance interpretability through effective graphic techniques, ultimately telling a story of global sporting evolution.

3 Dataset Summary

The dataset used in this report is sourced from the file titled **“2025-Autumn-32146-Ass2-v16.xlsx”**, which includes a detailed historical archive of Australian Open results.

The key attributes in the dataset include:

- **Year** of the tournament
- **Gender** (Men’s / Women’s)
- **Champion** and **Runner-up** names
- **Nationalities** of both Champion and Runner-up
- **Champion Seed**, **Runner-up Seed**
- **Match time**, **Score**

Calculated columns for **Win Rate**, **Stage Wins (1st–5th)**, **Losses**, etc.

Data Types and Formats

- Nominal variables: Player names, countries, gender, score (text format)
- Ordinal variables: Seed rankings
- Interval/ratio variables: Match time (minutes), win/loss counts, win rate (%)

Observations

The data spans over a century, capturing more than 200 tournament years.

There is a strong concentration of wins among a few countries (e.g., AUS, USA).

A few players dominate the historical record, especially among male champions.

Some years have incomplete data (e.g., missing seeds or match durations).

4 Data Cleaning and Transformation

Prior to conducting any meaningful visual analysis, it was imperative to perform systematic data cleaning and transformation. These preprocessing tasks ensured that the dataset was structured, consistent, and analytically sound for advanced visualisation in Tableau.

4.1 Handling Missing and Inconsistent Values

- To address missing and inconsistent data entries:
- Seed values were frequently missing for players in early years of the tournament. To maintain categorical consistency, missing entries were encoded as `"Not Available"`.
- Match Time and Score variables contained null values for older records. These were retained for historical context, but numeric placeholders (such as `0`) were used where mathematical operations (e.g., averages) were necessary.
- Incomplete year entries were kept in the dataset to preserve chronological continuity but excluded from win-rate calculations to prevent distortion.

- These adjustments improved both data integrity and comparability, ensuring all dimensions could be plotted without computational errors.

4.2 Normalisation and Standardisation

To harmonise the dataset across all visualisation types:

- Player names and country codes were cleaned for inconsistent casing, spacing, and typos (e.g., "USA" vs "Usa"), thereby avoiding split categories in treemaps and scatter plots.
- A custom Win Rate metric was calculated using the formula:
- Derived fields were created for each round (e.g., 2nd-Loss, 4th-Won) to support granular performance tracking in parallel coordinates plots.
- This normalisation step enabled multidimensional analysis while reducing data fragmentation across Tableau visual layers.

4.3 Structural Adjustments for Tableau Compatibility

The dataset underwent structural realignments to support Tableau's visual logic:

- Year fields were transformed from string to numeric format, allowing accurate chronological sorting and line plotting.
- Nation codes were verified and corrected using the ISO 3166 standard to ensure reliable country mapping in geographic charts.
- Date-time alignment ensured all charts (including moving average trendlines) could be synchronised along a common x-axis.

4.4 Grouping and Analytical Categorisation

To enhance comparison and segmentation:

Players with five or more championship titles were classified as "Top Players", enabling filtered analysis in treemaps and scatter plots.

Data was categorised by Gender across all visualisations, allowing clearer comparative evaluation in line charts, parallel coordinates, and geographic distributions.

Categorical variables such as Champion Nationality and Runner-up Country were consolidated for use as colour-encoded dimensions in treemaps and maps.

5 Data Analysis and Visualization

5.1 Parallel Coordinates

5.1.1 Multidimensional Analysis of Player Performance

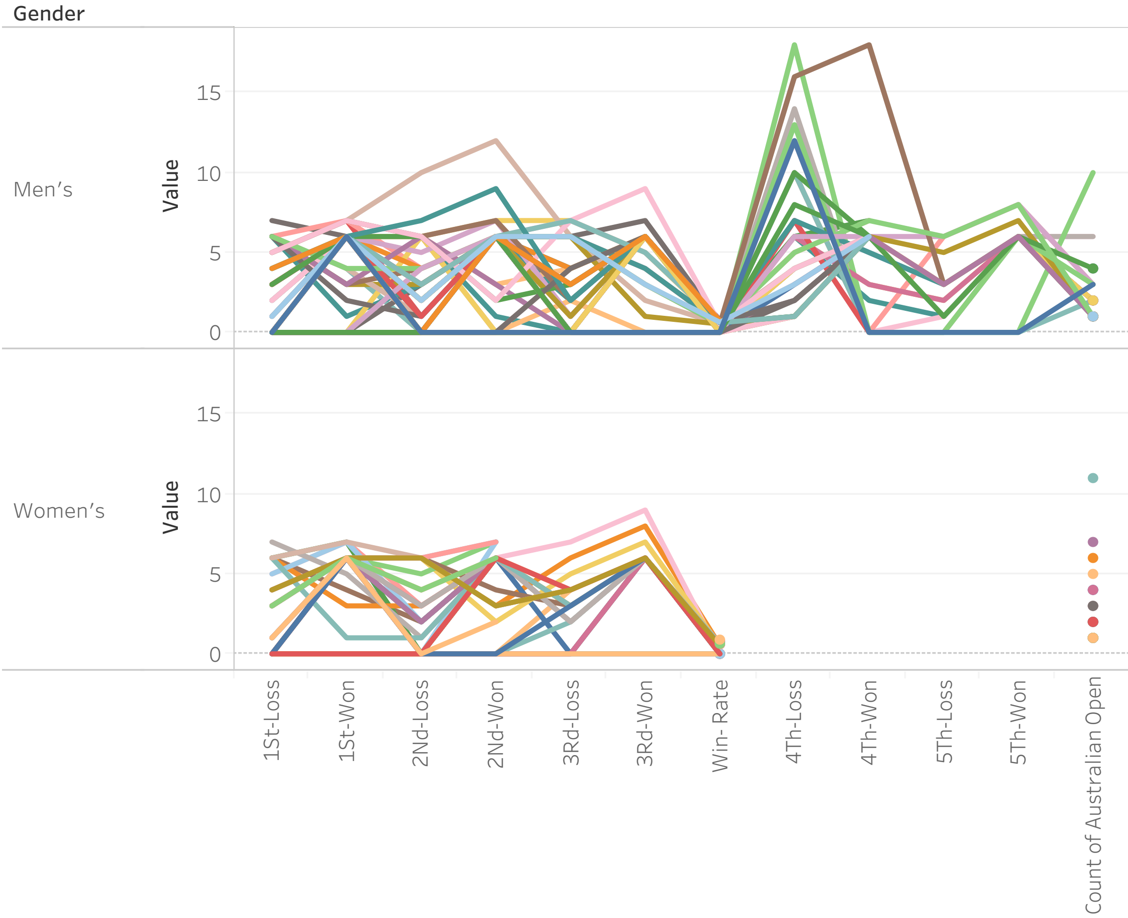

Figure 1. Match Outcomes and Win Rates by Gender (Parallel Coordinates)

This parallel coordinate chart visualises the multidimensional performance of top Australian Open champions by stage-wise match outcomes—from first to fifth round wins and losses—along with their overall win rates. The data is split by gender to enable clear visual comparison.

Each coloured line represents a unique champion, connecting their performance across each metric. The chart reveals patterns of dominance, variability, and consistency among elite players.

Key Findings:

- In the men's group, several players exhibit high consistency, showing dominance across multiple rounds with minimal losses and high win rates.
- The women's group displays tighter clustering, indicating a more uniform performance trend among top players.
- Notable outliers are visible where players have relatively fewer total wins but maintained high win rates, likely due to fewer appearances but strong outcomes.

Visualisation Value:

This chart is particularly effective in representing complex performance dimensions, offering both stage-specific and aggregated views. By using gender as a visual separator, it enhances readability and reveals subtle differences in performance consistency between male and female champions.

5.1.2 Loss Patterns and Win Rates by Gender

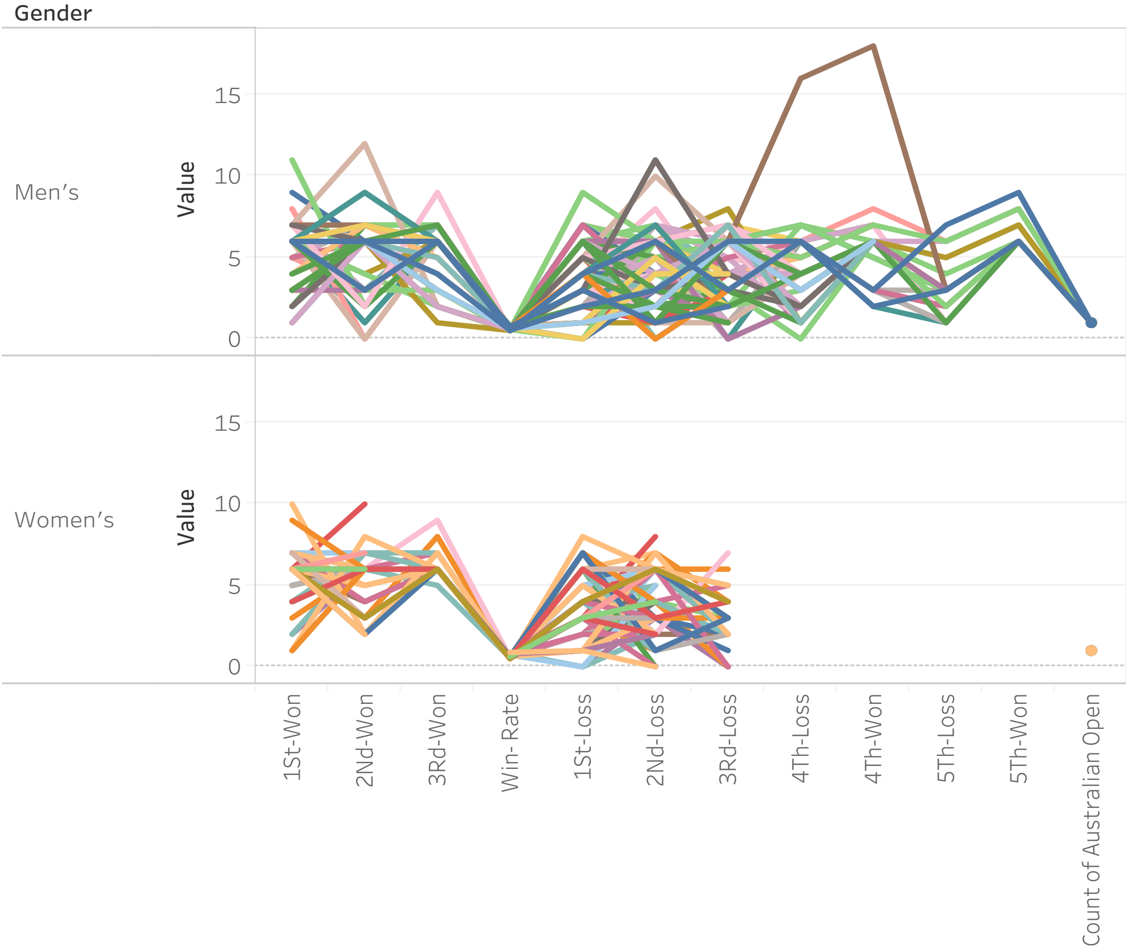

Figure 2. Loss Patterns and Win Rates by Gender (Parallel Coordinates)

This second parallel coordinates visualisation focuses on the **loss trajectory** of champions across tournament stages and relates this to their win rate. It offers a complementary view to the win-focused perspective in Figure 1, emphasising how champions handle setbacks.

Key Findings:

- Male players show **greater variation** in later-round losses, with some maintaining high win rates despite multiple losses, suggesting **resilience** or **longer careers**.
- Female players exhibit **tighter trends**, suggesting **greater consistency** and lower variability in performance across tournaments.
- Some champions are clear **outliers**, combining high losses with above-average win rates or vice versa.

Visualisation Value:

This chart deepens the performance analysis by shifting the lens from "how often players win" to "how they manage losses." It reveals that success is not always linear and that understanding how losses are distributed offers a richer perspective on what makes a champion.

5.2 Geographic Maps

5.2.1 Championship Counts by Country

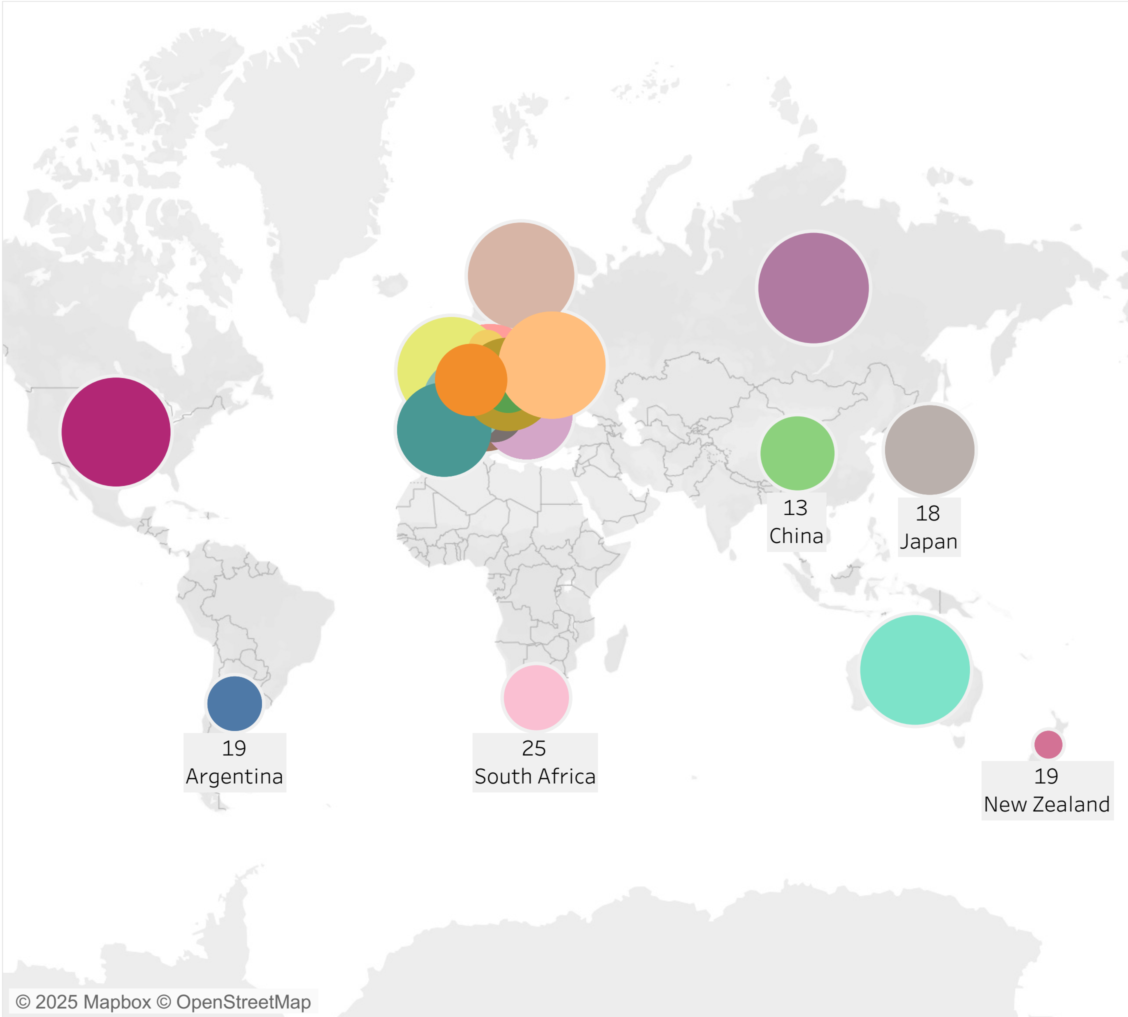

Figure 3. Championship Counts by Country (Geographic Map)

This geographic map visualisation illustrates the total number of Australian Open champions from each country. The size of the circle represents the number of championships, allowing a quick visual comparison of nations' overall success in the tournament.

Key Insights:

- The United States and Australia dominate the visual with the largest circles, reflecting their historical dominance and deep tennis tradition.
- European countries such as Germany, France, and Switzerland also appear prominently, showing a sustained contribution to the tournament's history.
- Interestingly, smaller nations like New Zealand (19 titles), South Africa (25 titles), and Argentina (19 titles) are also strongly represented, suggesting standout individual performances rather than depth.

Visualisation Value:

This map effectively provides a **global overview** of the distribution of champions across the world.

It helps identify both dominant and emerging tennis nations, supporting geographic analysis of performance concentration.

5.2.2 Average Win Rates by Country

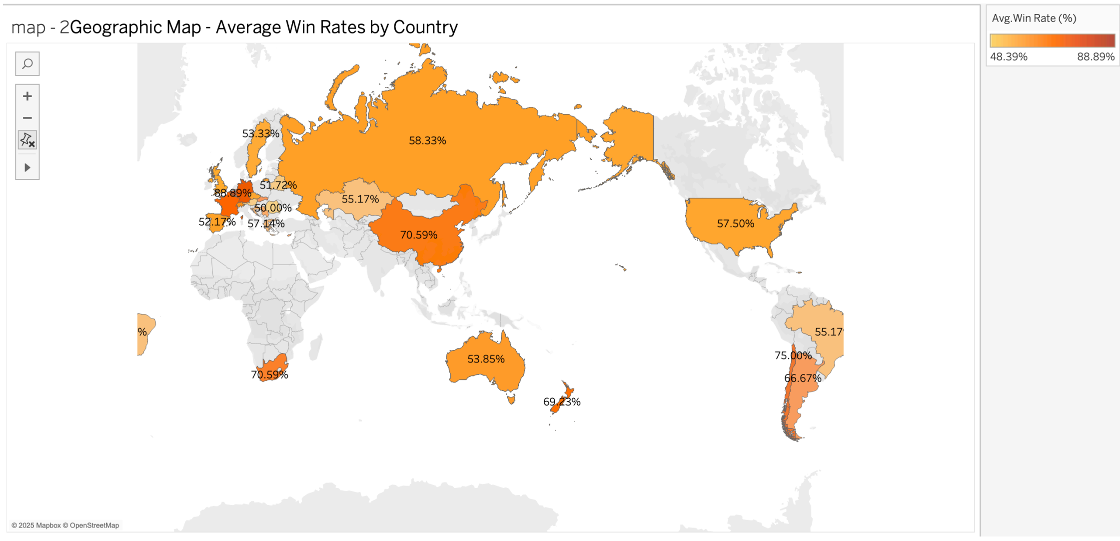

Figure 4. Average Win Rate of Champions by Country

This choropleth map visualises the average win rates of players from each country who became Australian Open champions. Darker shades indicate higher win rates, and lighter shades show relatively lower success ratios.

Key Insights:

- Switzerland (88.89%) and Argentina (75%) stand out with the highest win rates, suggesting their champions were highly efficient when participating.
- While Australia and the U.S. had high total wins, their win rates were more moderate (AUS: ~53.85%, USA: ~57.5%), possibly due to larger player pools and higher match volumes.
- European nations such as France, Germany, and Russia show moderate win rates, reflecting consistent but not overwhelmingly dominant performances.

Visualisation Value:

This map provides a qualitative layer to the previous figure by shifting focus from quantity to efficiency.

It supports performance quality analysis by showing which countries produced champions who were not only frequent participants but also highly successful competitors.

5.3 Treemap Visualisation

5.3.1 Individual Champions’ Win Rates by Gender

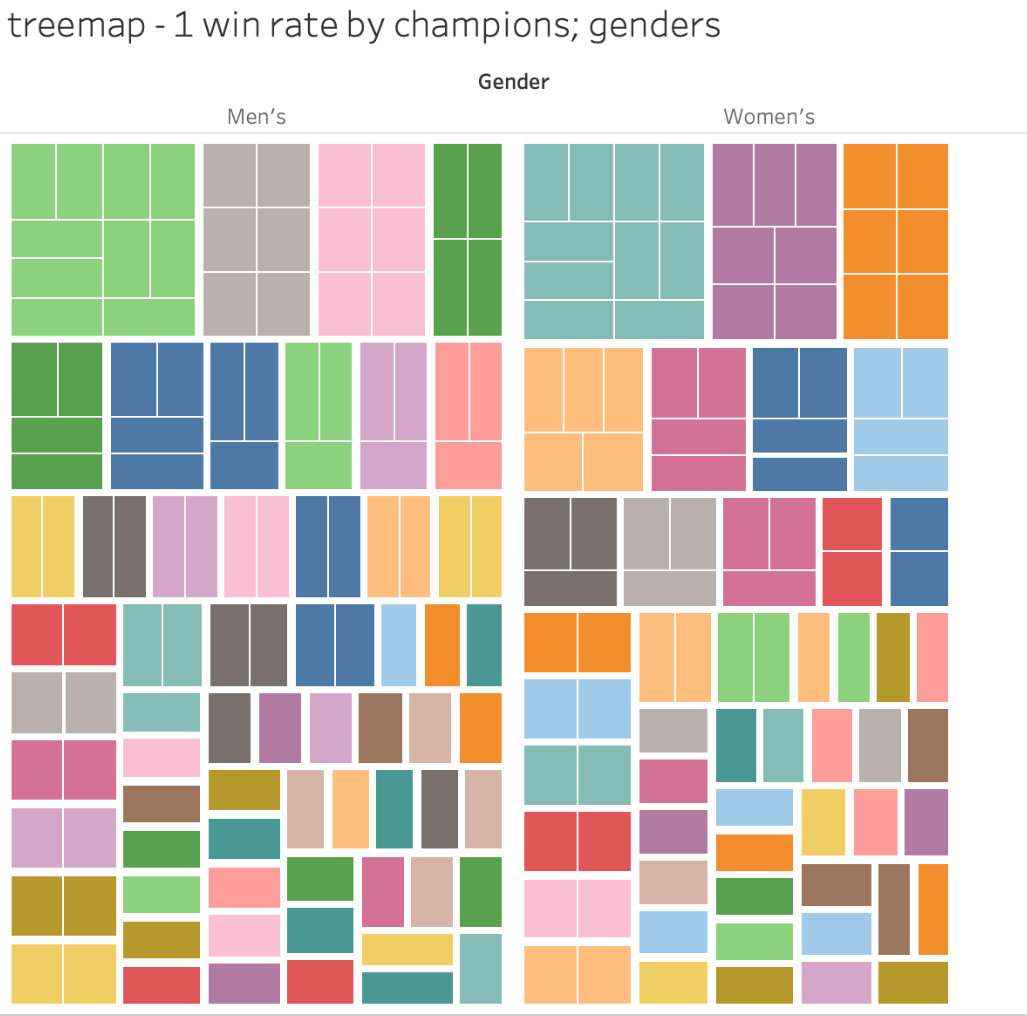

Figure 5. Individual Champions’ Win Rates by Gender (Treemap)

This treemap visualises each champion's individual win rate, grouped by gender. Each rectangle represents a player, with its size corresponding to their win rate. Colours vary by player identity, offering a visual separation for comparative purposes.

Key Insights:

- Male champions display a more dispersed win rate distribution, with a wider range of rectangle sizes, suggesting variability in performance.
- Female champions, while still diverse, tend to cluster more closely in size, indicating relatively more consistent win rates among top-performing women.
- Several large blocks stand out in both groups, representing dominant players with exceptional consistency over their appearances at the Australian Open.

Visualisation Value:

This view effectively allows rapid identification of high-performing individuals, as well as general performance stability across gender lines.

It supports comparative analysis of outstanding consistency vs. short bursts of success.

5.3.2 National Win Rate Share by Gender

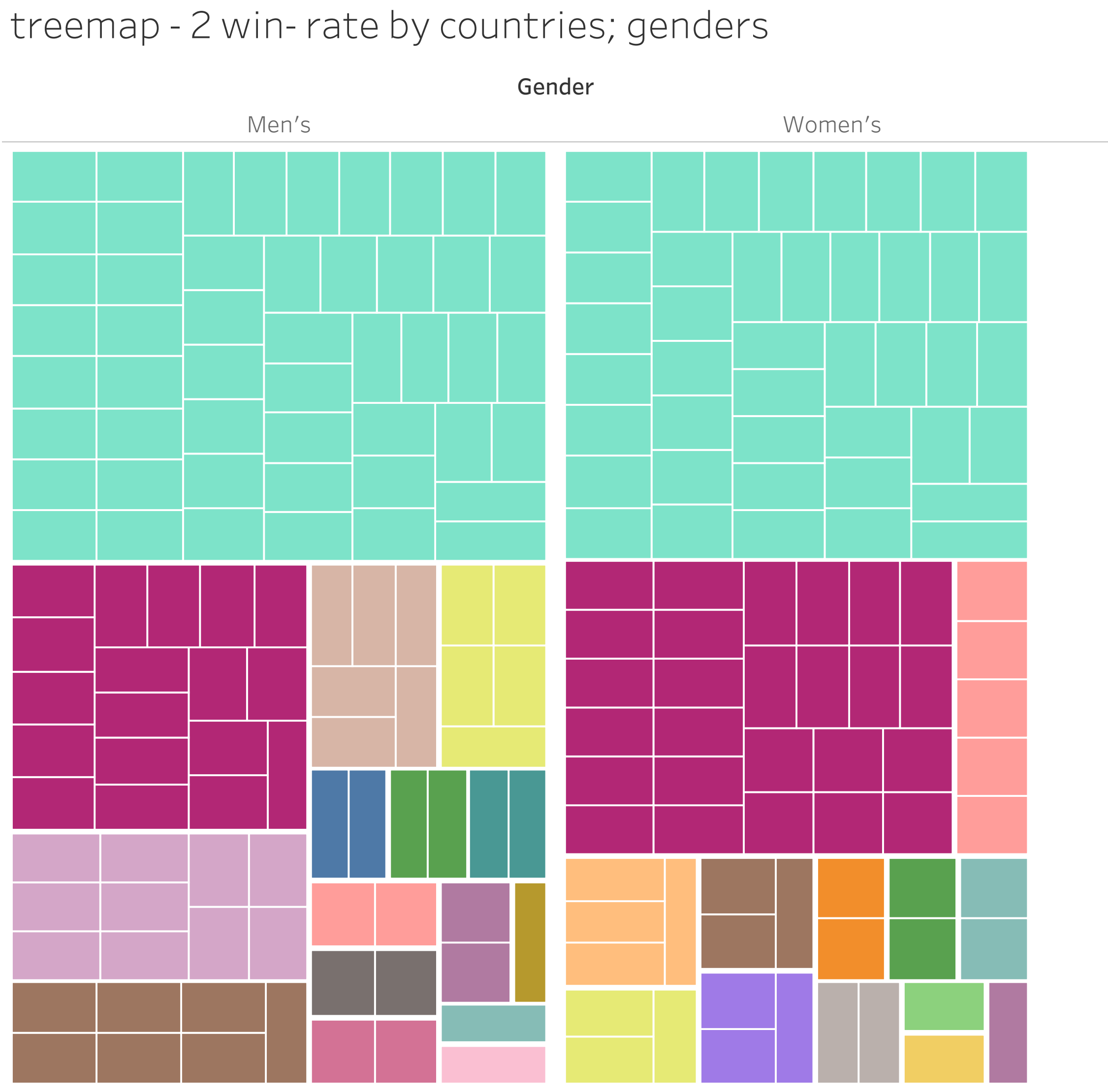

Figure 6. National Win Rate Share by Gender (Treemap)

This second treemap aggregates win rates at the **country level**, split again by gender. The size of each rectangle reflects the total number of wins by players from that country, and the colours represent different nationalities.

Key Insights:

- Australia and the United States occupy the largest space in both gender categories, reflecting their historical dominance in the tournament.
- Eastern European countries such as Russia and Serbia are highly visible, particularly among women, showing the emergence of strong female players in more recent years.
- Some smaller nations (e.g., New Zealand, Switzerland) have relatively fewer wins but still appear due to high efficiency or standout individuals.

Visualisation Value:

Compared to individual analysis, this treemap summarises national-level success, allowing quick geopolitical comparisons.

The gender separation further supports understanding of whether success is skewed by gender within a country.

5.4 Scatter Plot Analysis: Performance of Top Champions

5.4.1 Win Rates of Top Players

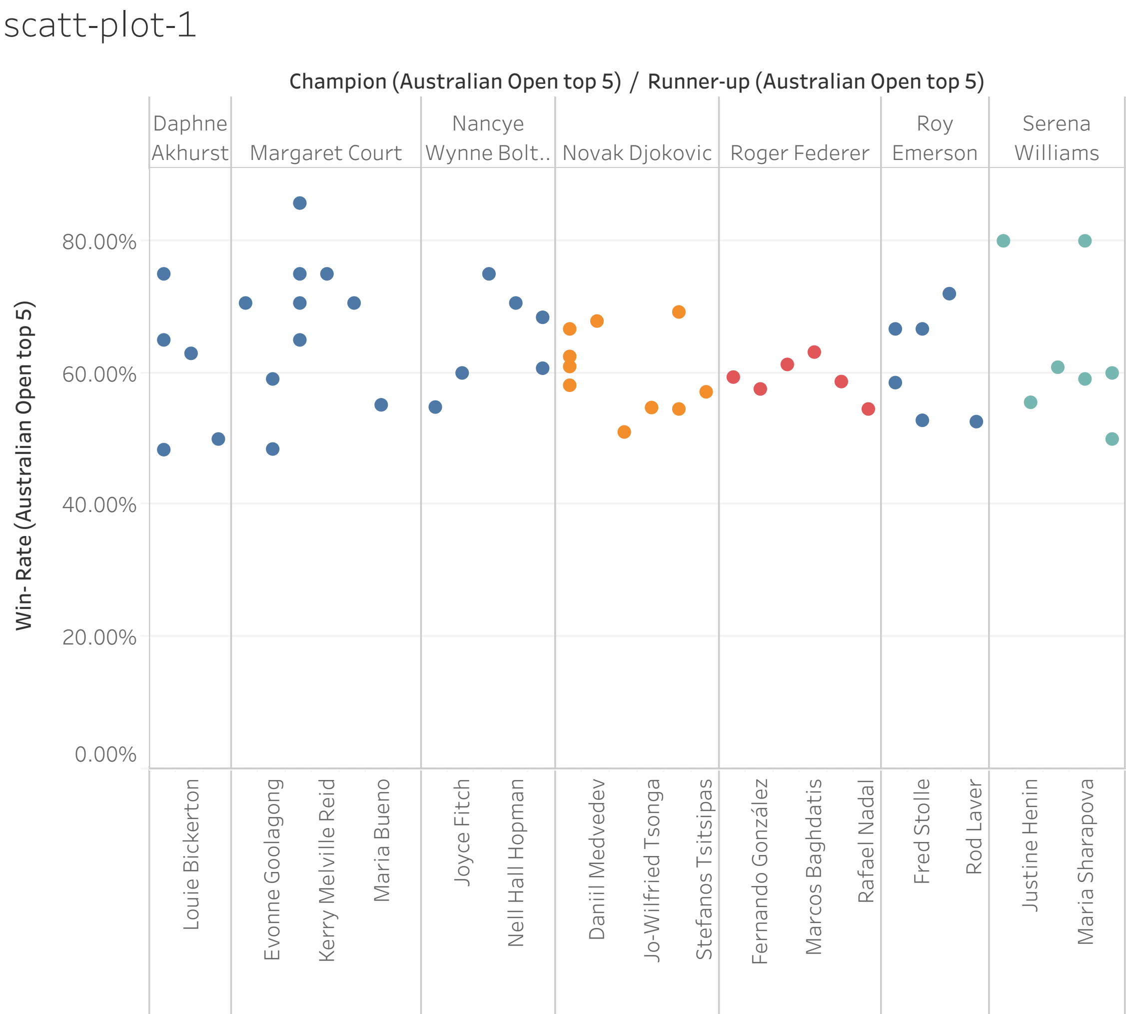

Figure 7. Win Rates of Top Players

This scatter plot visualises the win rates of selected top champions and runners-up who have either won or reached the finals five or more times in the Australian Open. Each dot represents an individual tournament participation, with the y-axis reflecting win rate percentage.

Key Insights:

- Margaret Court and Serena Williams dominate the women's group with consistently high win rates, often exceeding 70–80%.
- Novak Djokovic shows tight clustering of points above 60%, highlighting exceptional consistency in match wins over the years.
- In contrast, players like Maria Sharapova and Roger Federer exhibit more variance, indicating fluctuations in performance throughout their careers.
- A few male players show multiple appearances with win rates below 50%, suggesting periods of decline or challenging match-ups in later years.

Visualisation Contribution:

The plot supports longitudinal comparison by placing individual performances side by side.

It highlights both sustained excellence and performance variability, essential in evaluating career consistency.

5.4.2 Win Rates by Gender

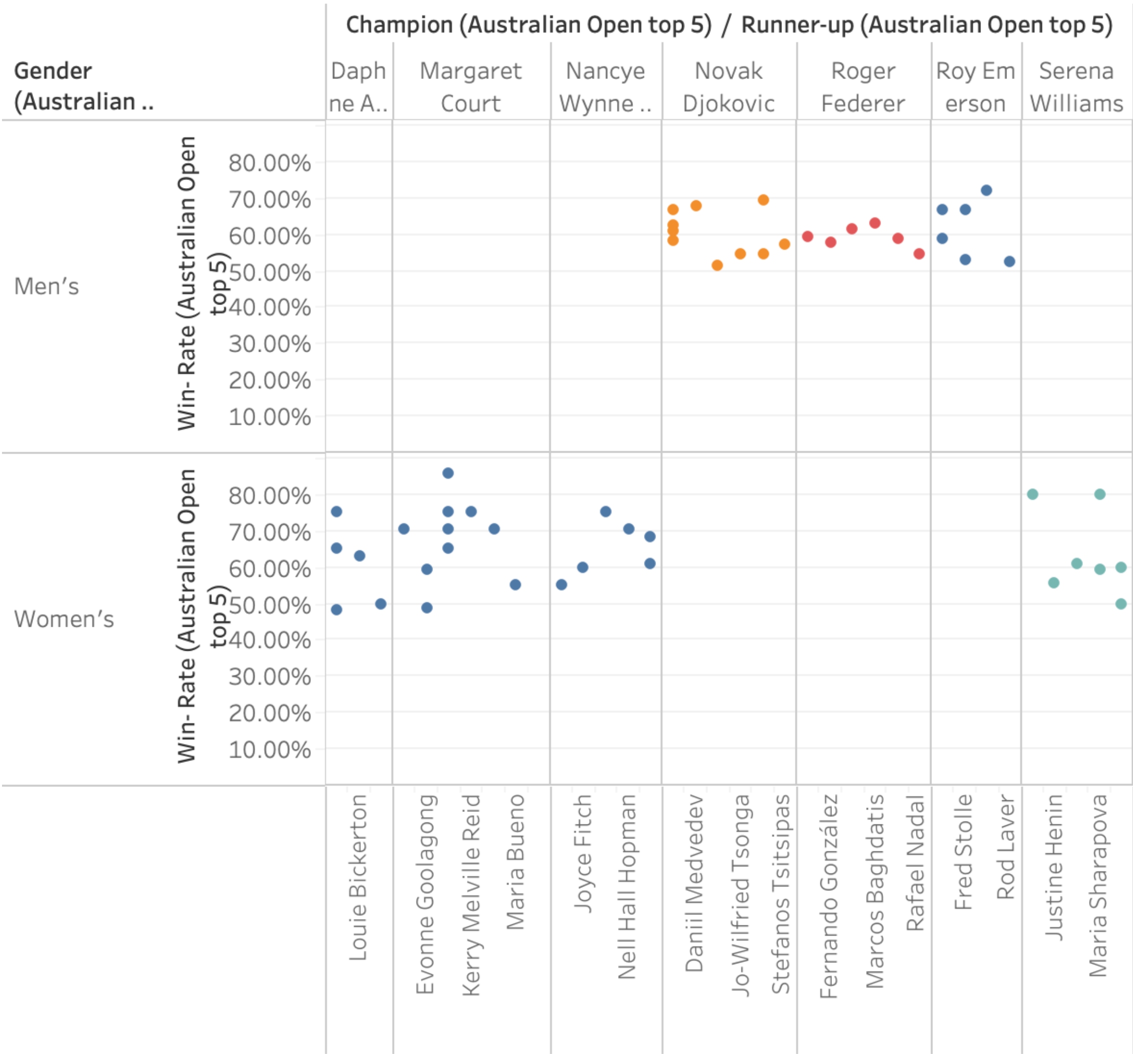

Figure 8. Win Rates by Gender (Segmented Scatter Plot)

This scatter plot builds upon Figure 7 by adding a gender segmentation, displaying win rate distributions separately for men and women. This enhances the interpretability of gender-based differences and comparative performance profiles.

Key Observations:

- The women's group shows greater consistency with a tight clustering of points between 60% and 80%, particularly among players such as Court, Williams, and Evonne Goolagong.
- The men's group displays more spread, with certain champions maintaining high win rates (e.g., Djokovic, Roy Emerson), while others show more variability.
- Outliers such as Rod Laver and Justine Henin reflect strong performances in fewer appearances, pushing their win rates into the higher band despite limited data points.

Visualisation Value:

This gender split provides clearer visibility of performance patterns, enabling a better understanding of whether dominance is player-specific or gender-systemic.

It also assists in evaluating historical versus modern-era trends within each gender group.

5.5 5-Year Moving Average of Win Rates by Gender (1907–2023)

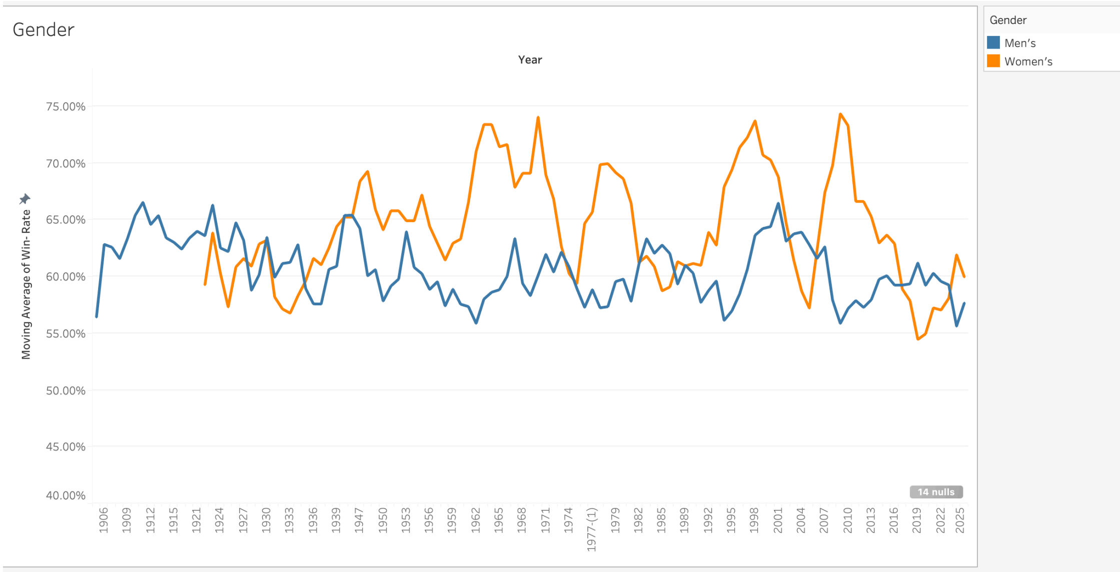

Figure 9. 5-Year Moving Average of Win Rates by Gender (1907–2023)

To explore long-term performance dynamics at the Australian Open, a five-year moving average of win rates was calculated and visualised by gender. This smoothing technique effectively mitigates short-term fluctuations and enhances visibility of broader competitive trends across the tournament's 120-year history.

Key Observations:

- Men's win rates have remained relatively stable, hovering around the 60% mark with moderate variation across decades.
- Women's win rates exhibited more pronounced fluctuations, with clear performance peaks observed during the 1960s and again in the early 2000s—likely aligned with dominant eras led by iconic female players.
- In more recent years, the win rates for both genders appear to converge slightly below 60%, suggesting a growing parity in competitive balance.

Visualisation Value:
This moving average plot provides a high-level view of evolving performance trends across genders. By removing short-term noise, the visual enables clearer identification of periods of dominance and competitive shifts. It also supports historical comparison and contextualisation of performance consistency over time.

5.6 WordCloud

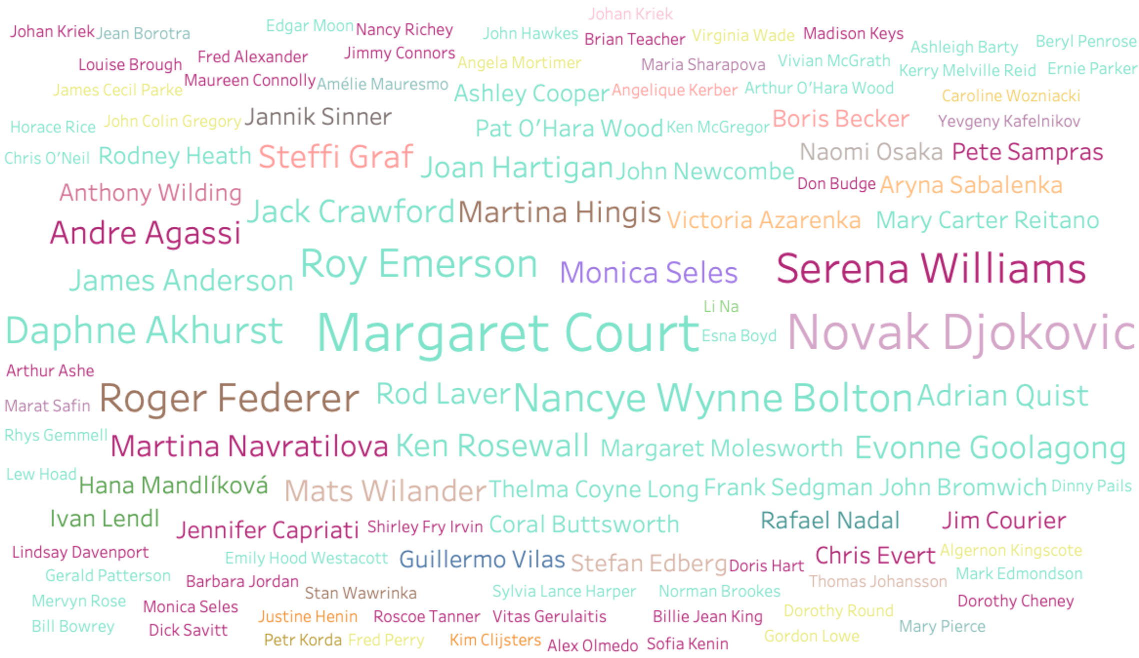

Figure 10. Word Cloud of Champion Names

To conclude the visual analysis, a word cloud was generated to represent the most prominent Australian Open champions. Player names were scaled based on the frequency of championship wins or runner-up finishes.

Key Observations:

- Margaret Court, Novak Djokovic, and Serena Williams are prominently featured, reflecting their exceptional historical impact and sustained success.
- Australian legends such as Roy Emerson, Rod Laver, and Evonne Goolagong are also notable, underscoring Australia's foundational contribution to the tournament.
- The word cloud illustrates gender diversity, with both male and female players represented across various levels of prominence.

Visualisation Value:
While not numerically precise, the word cloud offers a qualitative yet impactful overview of champion legacy. It effectively summarises career influence and public recognition, serving as a complementary visual that reinforces the statistical findings presented throughout this report.

6 Conclusion

This report presented a comprehensive visual analysis of the Australian Open's historical dataset using Tableau, uncovering meaningful patterns across player performance, national trends, and gender-based dynamics over the past 120 years.

The analysis demonstrated that while countries such as Australia and the United States have historically dominated the tournament in volume of wins, their efficiency—as measured by average win rates—varied significantly when compared to smaller nations like Switzerland and Argentina. At the individual level, athletes such as Novak Djokovic, Margaret Court, and Serena Williams emerged as consistent and dominant performers, while others displayed variable or underdog trajectories.

A wide range of visualisation techniques—including treemaps, geographic maps, scatter plots, parallel coordinates, and moving averages—enabled both macro-level trend identification and micro-level player comparisons. Gender-segmented visuals added further depth by highlighting temporal shifts in competitive balance and performance consistency between male and female athletes.

Tableau's strengths in interactive, multi-dimensional visualisation significantly enhanced the accessibility and narrative power of the dataset. While minor limitations were encountered—such as managing label precision and complex table calculations—these were outweighed by the platform's visual clarity and ease of exploration.

Overall, this report successfully transformed a complex dataset into an engaging analytical story. It not only delivered insight into historical performance trends but also demonstrated how thoughtful data visualisation can elevate strategic understanding and decision-making in sport analytics.