

Project Documentation

Overview

Project Name

Assessment Task 2: Advanced Data Visualisation

Company

Australian Open

Project Dates

Start Date: 1905

End Date: 2025

Background

Australian Open is a tennis competition held in Melbourne, VIC, Australia

Objectives

- Exploratory data analysis
- Data visualization

Student Details

ID : 14447837

Name: Kaung Khant Kyaw

Email: KaungKhantKyaw@student.uts.edu.au

Contents

Introduction	3
Data Exploration	4
Data Cleaning and Transformation	5
Outliers	6
Gender.....	7
Top Champions.....	8
Champions.....	9
Champion vs Runner-up	12
Runner-ups	13
Conclusion.....	17

Introduction

The Australian Open, one of the most watched and popular tennis tournaments, has thrilling finals and legends like Novak Djokovic, and Serena Williams. This report details the historical data of the Australian Open, showcasing the patterns, and insights from both the men's and women's singles competition finals.

To bring these insights to life, the report leverages various powerful visual techniques, including Tree Maps, Global Maps, Scatter Plots, and Parallel Coordinates. These graphics not only enhance the clarity of complex data but also uncover hidden patterns, such as recurring rivalries, standout performances, and geographical dominance. By combining statistical analysis with interactive visual storytelling, this report aims to provide a deeper understanding of the Australian Open, celebrating both legendary champions and the many players whose consistent excellence brought them to the final stage.

Data Exploration

Name	Type	Description
Year	Integer	The year of the Australian Open held
Gender	String	The gender of the competition
Champion	String	The name of the champion who won the competition
Champion Nationality	String	A three-letter code to describe the nationality of the champion
Champion Country	String	The country of the champion
Score	String	The total score of each round of the final game
Champion Seed	Integer	The champion's ranking or placement in the tournament draw
Mins	Integer	The minutes of the total game
1st-won	Integer	The number of games won by the champion in the first round
1st-lost	Integer	The number of games lost by the champion in the first round
2nd-won	Integer	The number of games won by the champion in the second round
2nd-lost	Integer	The number of games lost by the champion in the second round
3rd-won	Integer	The number of games won by the champion in the third round
3rd-lost	Integer	The number of games lost by the champion in the third round
4th-won	Integer	The number of games won by the champion in the fourth round
4th-lost	Integer	The number of games lost by the champion in the fourth round
5th-won	Integer	The number of games won by the champion in the fifth round
5th-lost	Integer	The number of games lost by the champion in the fifth round
Runner-up	String	The name of the runner-up who became second in the competition
Runner-up Nationality	String	A three-letter code to describe the nationality of the runner-up
Runner-up Country	String	The country of the runner-up
Runner-up Seed	Integer	The runner-up's ranking or placement in the tournament draw

Data Cleaning and Transformation

During the data cleaning process, I removed columns that were either redundant or not meaningful for visualization purposes. Specifically, I deleted the "Champion Nationality" and "Runner-up Nationality" fields because they only contained three-letter country codes (e.g., "AUS", and "USA"), which Tableau was unable to recognize correctly as geographic data. Instead, I relied on the full country names provided in the "Champion Country" and "Runner-up Country" columns, which allowed for proper mapping in geographic visualizations. Additionally, I removed the "Minutes" column because it only contained data for two years, making it inconsistent and unreliable for meaningful analysis across the entire dataset. These steps ensure that the remaining data was clean, complete, and optimized for accurate and insightful visualizations in Tableau.

In addition to cleaning the dataset, I also created three columns to enhance the depth of my analysis. First, I introduced a "Wins" column, which represents the total number of games won by the champion during the final match. This was calculated by summing up the values in the "1st-won", "2nd-won", "3rd-won", "4th-won", and "5th-won" columns. I also added a "Loss" column by summing the corresponding "loss" fields for each set in the final. Finally, I derived a "Win-Rate" metric by dividing Wins by the total number of games played in the final (Wins + Losses). This percentage-based metric provides a clearer view of a champion's dominance in the final and became a key measure in several visualizations. These columns offered more precise insights and enabled more meaningful comparisons across players and countries.

Name	Type	Description
Wins	Int	The total number of wins by the champion in the final game $\text{Wins} = \text{1st-won} + \text{2nd-won} + \text{3rd-won} + \text{4th-won} + \text{5th-won}$
Loss	Int	The total number of losses by the champion in the final game $\text{Loss} = \text{1st-loss} + \text{2nd-loss} + \text{3rd-loss} + \text{4th-loss} + \text{5th-loss}$
Win-Rate	String	The win percentage of the champion in the final game $\text{Win-Rate} = \text{Wins} / (\text{Wins} + \text{Loss})$

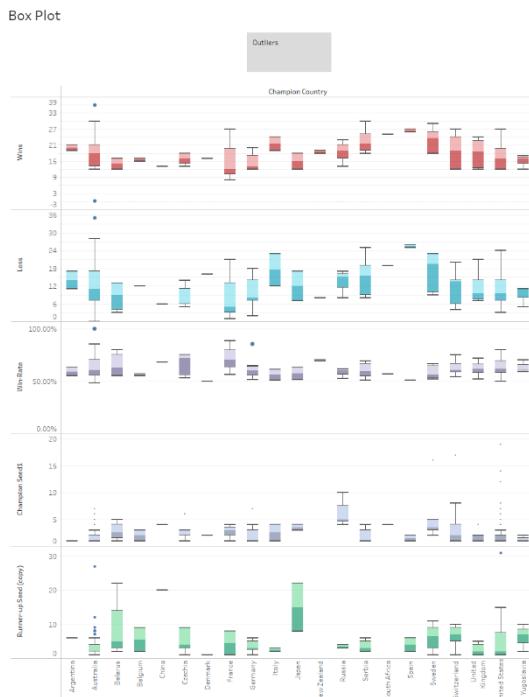
Outliers

Year	Gender	Champion	Wins	Losses	Win-rate
1925	Women's	Daphne Akhurst	15	16	48.39%
1971	Women's	Margaret Court	16	17	48.48%

- The typical win rate (median is about 0.61, but we found some players with a perfect win rate (1.0), indicating they have never lost a game in a final, which is rare and perfect performance)
- On the lower end, there are some champions who were under 49% of win rate of games in the final, indicating that they lost more games than they won which is unusual for a champion.

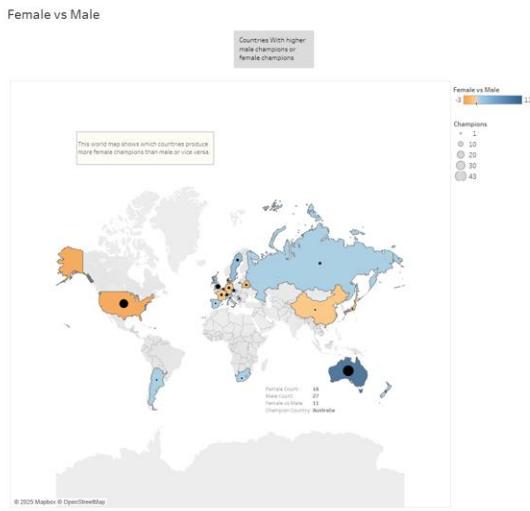
Year	Gender	Champion	Wins	Losses	Win-rate
1927	Men's	Gerald Patterson	36	35	50.70%

- Most Champions won between 13 to 21 finals. But one champion won 36 games which is higher than average.
- Likewise, that same player lost 35 games in a final which makes it a clear outlier.



This box plot provides a comprehensive comparison of champion performance by country, highlighting metrics such as wins, losses, win rate, seed ranking, and match duration. Notably, Australia and the United States stand out with outliers indicating champions with exceptionally high win counts, emphasizing their long-standing dominance in the tournament. Champion seed outliers show that a few players managed to win despite low or unexpected seedings, highlighting remarkable underdog achievements. Overall, the box plot not only visualizes performance distributions but also reveals the outliers that represent extraordinary achievements and unique stories within tennis history.

Gender



This map provides a geographical breakdown of gender dominance among Australian Open singles champions by country. Each country is color-coded to represent whether it has produced more male or female champions: shades of blue indicate a higher number of female champions, while orange tones represent male dominance. The intensity of the color corresponds to the degree of gender imbalance. Additionally, the size of the black circles in each country marks the total number of champions produced by that nation, offering a quick visual cue for champion volume. This visualization effectively reveals not only which countries contribute most to the tournament's history but also highlights fascinating trends in gender representation across tennis nations.

The analysis of gender representation among Australian Open champions reveals distinct geographical trends. Countries like Australia and the United States lead in overall champion production. Australia shows a stronger presence of female champions, whereas the United States slightly favors male champions. A global map visualization emphasizes these patterns using color-coded regions and marker sizes to reflect both the total number of champions and the degree of gender imbalance. In contrast, countries such as Germany and Russia demonstrate a more balanced contribution of male and female champions, offering a nuanced perspective on gender dynamics within international tennis success.

Top Champions

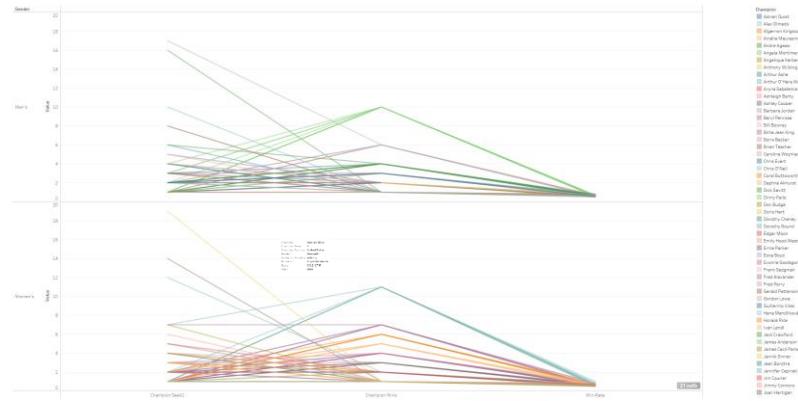


This map highlights the countries that have produced the most dominant champions in Australian Open finals, based on the highest win rate. Each country is color-coded and labeled with the name of the top-performing champion from that nation. By focusing on final win rates, this map shows those who were the most efficient and successful when it mattered most in the championship match. For example, Margaret Court of Australia emerges as a standout figure, representing a nation with many champions and an exceptional record in finals. Other notable

players, such as Maria Sharapova (Russia), Li Na (China), and Steffi Graf (Germany), demonstrate global excellence in tennis beyond the traditional powerhouses. This visualization offers an engaging perspective on how tennis greatness is distributed around the world, not just by the number of champions, but by their ability to perform in the final.

Focusing on final match win rates rather than sheer title counts, the top champions map highlights the most efficient and dominant players in Australian Open history. Margaret Court of Australia stands out with both high win rates and numerous titles, symbolizing an era of dominance. Champions like Steffi Graf (Germany), Maria Sharapova (Russia), and Li Na (China) further emphasize global excellence, proving that top-tier performance is not confined to traditional tennis powerhouses. This visualization shifts the narrative toward players who performed exceptionally well under the highest pressure—on championship day.

Champions

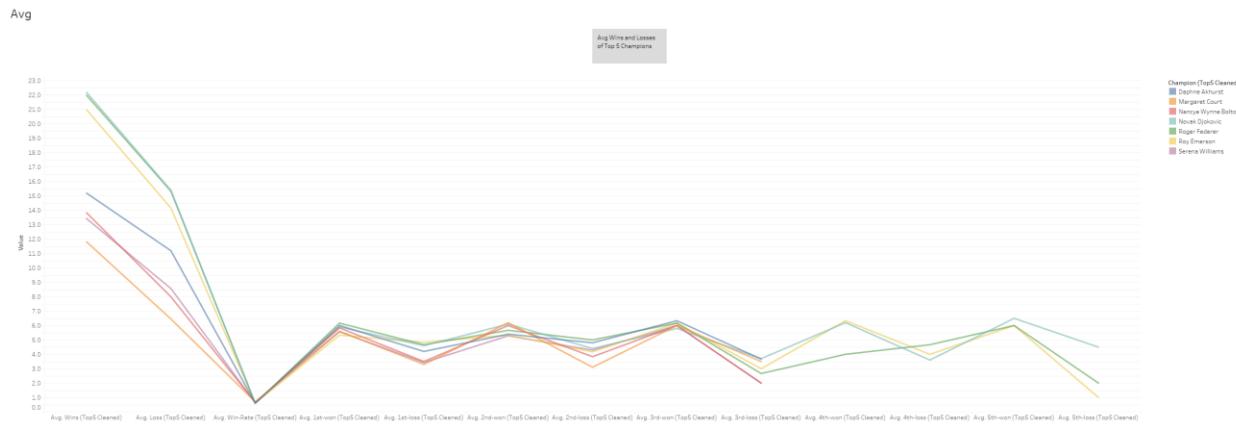


This parallel coordinate plot offers a dynamic visualization of Australian Open champions by gender, and plotted across three key performance dimensions: Champion Seed, Champion Wins, and Final Win Rate. Each line represents a unique player, colored individually, allowing us to trace their performance journey across these axes.

In the men's category, several champions have achieved high win counts and win rates, even though they were not always top seeds. This indicates a strong tournament performance that exceeds expectations. In contrast, the women's category shows a denser cluster near the lower seed numbers, suggesting that top-seeded players more frequently convert their advantage into victories. However, both groups feature striking outliers: certain players with low seeds have managed to win consistently, as reflected by their higher positions in the "Champion Wins" and "Win Rate" axes.

Notably, the steep slope seen in many lines between the "Champion Wins" and "Win Rate" axes indicates players who have consistently performed well in finals, with some achieving near-perfect records. This visualization not only reflects performance but also invites deeper analysis into the success of underdogs, the reliability of seeds, and gender-

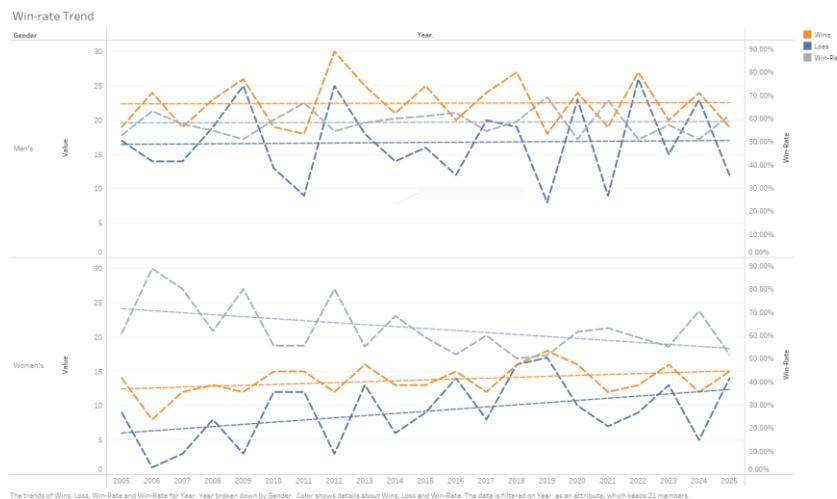
based patterns in tennis dominance at the Australian Open.



This report provides an overview of the performance of the top 5 Australian Open champions, using a parallel coordinates plot to compare their match statistics over the years. The focus is on average wins and losses for both men's and women's champions. The men's champions, including Novak Djokovic and Roger Federer, have consistently shown high performance. Their lines on the chart are higher, indicating more wins compared to losses. This demonstrates their dominance in the tournament over the years. The women's champions, such as Serena Williams and Margaret Court, are represented in a different color group on the chart. Although their lines are lower than the men's, they still show strong performances with more wins than losses. This highlights their significant achievements in the tournament. This visualization helps us compare how each top player performs over time and how their wins and losses relate to one another. It clearly shows that these top champions have maintained high winning records with fewer losses in their championship won throughout their careers.

The Champions section features two key visualizations that provide an overview of player performance at the Australian Open. The first chart, a parallel coordinates plot, illustrates each champion's tournament journey by comparing their seed ranking, total wins, and final match win rate. This visualization reveals some intriguing trends: men's champions often succeed despite not being top seeds, showcasing their ability to surpass expectations, while women's champions generally align more closely with their seed rankings, indicating a stronger correlation between high seeding and victory. The chart also highlights outliers, such as Madison Keys in 2025, whose exceptional performance challenges typical seed-based patterns.

Complementing this chart, the second visualization focuses on the average performance of the top five champions across their Australian Open appearances. It compares key match statistics, such as average wins and losses, and visually contrasts the dominance of legendary male players like Novak Djokovic and Roger Federer, whose lines rise sharply due to consistently high win counts. Female champions like Serena Williams and Margaret Court also demonstrate strong performance, although their averages are slightly lower. Together, these visualizations provide a detailed picture of what makes a great champion, whether it's consistent dominance, outperforming expectations, or maintaining a high level of play over multiple years.



This trend line plot shows the yearly performance of top players at the Australian Open, broken down by gender. It tracks three main things from 2005 to 2025: the number of wins (in orange), losses (in grey), and the win-rate percentage (in blue dashed lines). For the men's section at the top, we can see that wins tend to stay higher than losses almost every year, and the win rate mostly stays above 60%, even reaching close to 80% in some years. This shows that the top male players have been consistently successful throughout the years. There's a slight dip in some years, but the trend remains strong overall. In the women's section at the bottom, the winning figures are generally lower than in the men's, and the losses are a bit higher. The win-rate here is more varied, with some noticeable drops below 40%, especially in the early 2010s and mid-2020s. However, there have also been improvements in more recent years, showing that performance can bounce back.

Overall, this plot gives a clear picture of how the top male and female players have performed over time. While men show a steadier and higher win-rate trend, the women's trend shows more ups and downs but also signs of recovery and improvement in later years.

Champion vs Runner-up

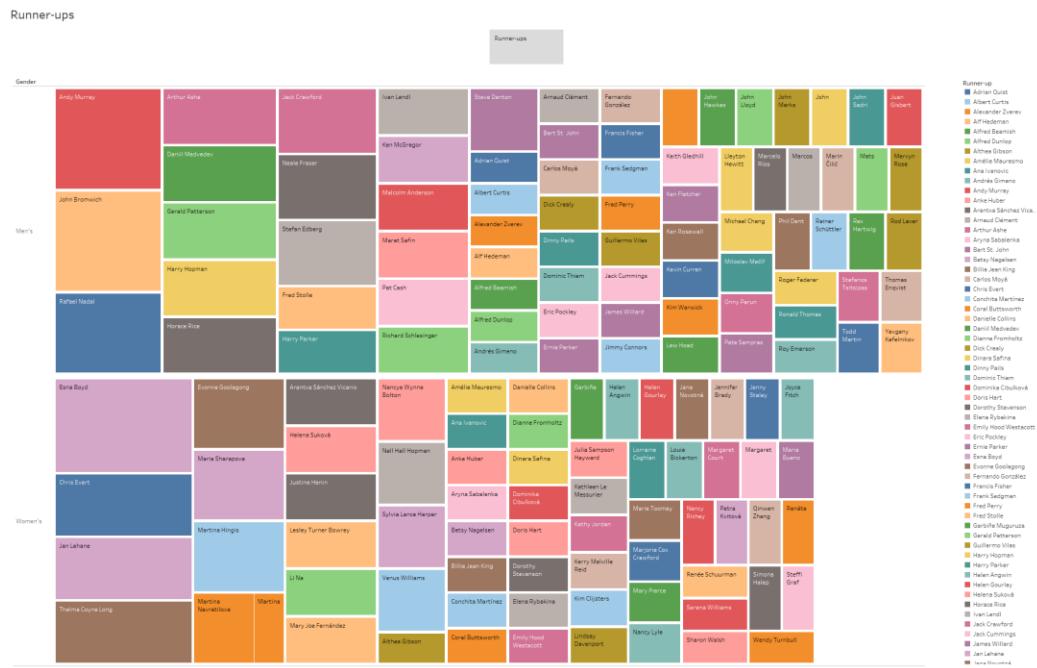
Champion vs Runner-up



This scatter plot titled “Champion vs Runner-up” highlights the dominant rivalries in Australia Open finals by showcasing champions who have defeated the same runner-up more than once. Each color dot represents a champion’s win rate against a specific runner-up, with champions distinguished by color and runners-up listed along the x-axis. Notably, Margaret Court and Serena Williams stand out with high win rates against multiple opponents, reflecting their dominance across eras. Serena’s repeated victories over Venus Williams show one of the most iconic rivalries between sisters in the tennis world. Also, players like Novak Djokovic and Roy Emerson also demonstrate sustained superiority over certain runners-up. The chart effectively captures patterns of repeated success, revealing how champions maintained a competitive edge in recurring final matchups.

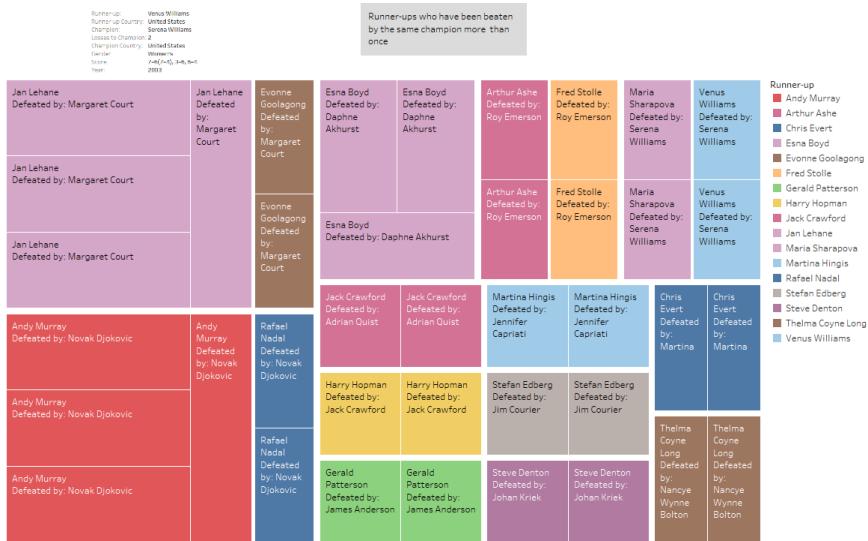
This chart is created by using calculated fields to identify and highlight champions who have repeatedly defeated the same runner-up in Australian Open finals. By creating a calculated field that counts how many times each champion beats a specific runner-up, this visualization filters for cases where this count exceeds one. This makes the visualization more about champions and their rivalries in the final of Australian Open.

Runner-ups



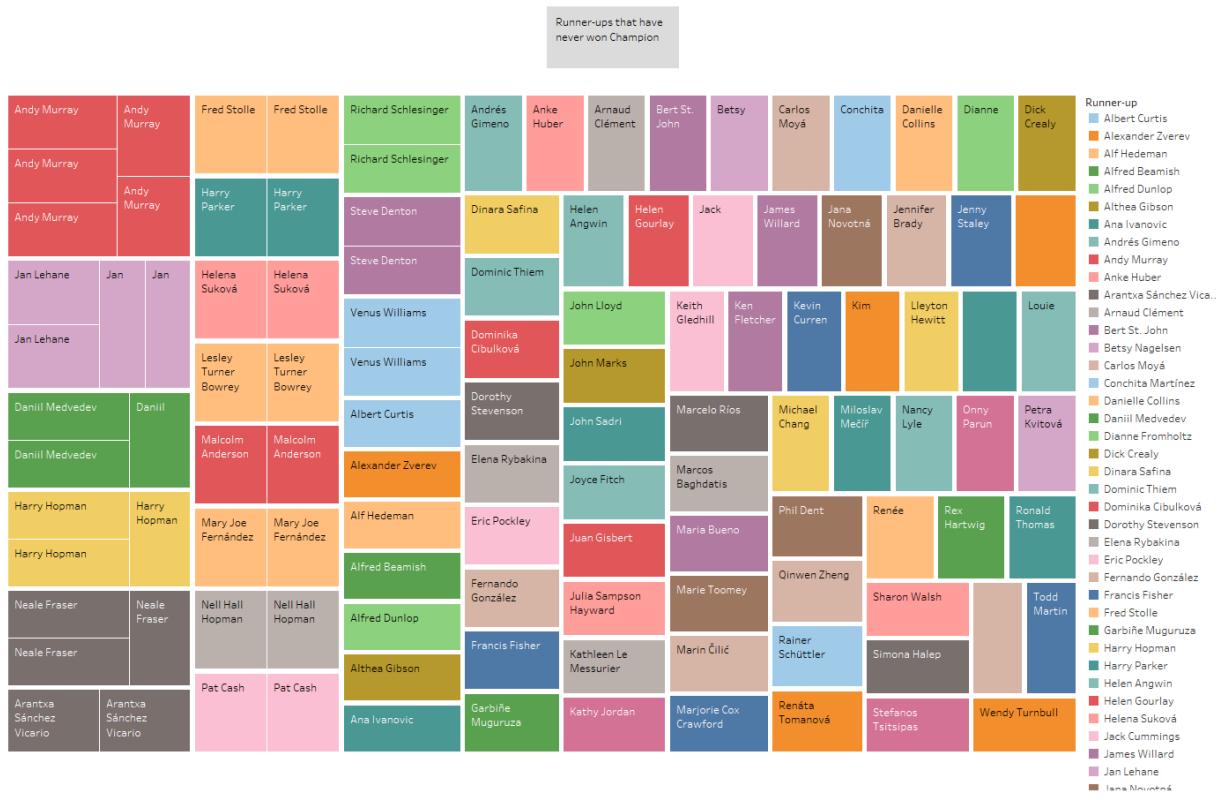
This tree map visualizes the distribution of runners-up in the Australian Open, separated by gender and colored by individual players. Each block represents a runner-up, with the size corresponding to the number of times they have finished in second place. This design allows for a quick comparison of how frequently different players have reached the finals without winning. Notable figures such as Andy Murray, John Bromwich, and Rafael Nadal stand out in the men's category, while Esna Boyd and Chris Evert are prominent among the women. By organizing the data this way, the visualization highlights the persistence and consistency of players who frequently made it to the final round, even if they didn't always emerge victorious. It also offers a unique perspective on the history of near victories in the tournament.

Runner-ups detailed



This tree map, titled “Runner-ups Detailed,” provides a deeper narrative by not only highlighting runner-ups who have lost multiple finals to the same champion, but also specifying the opponents who defeated them. At the forefront is Andy Murray, who was defeated three times by Novak Djokovic, showing one of the most notable rivalries in tennis. Jan Lehane’s repeated losses to Margaret Court and Esna Boyd’s multiple defeats by Daphne Akhurst reflect similar historical patterns of dominance. The inclusion of legends like Serena Williams, who bested both Maria Sharapova and Venus Williams multiple times, and Roy Emerson, who defeated both Fred Stolle and Arthur Ashe more than once, reveals the extent to which a few champions consistently outperformed their rivals in finals. This visualization effectively shows patterns of dominance, illustrating how certain players repeatedly proved to be undefeatable for others at the final games.

Runner-ups



This tree map shows the runner-up players who have never won a championship. Each colored box represents a player, and the size of the box shows how many times they have finished as runner-up in the tournament. Some players appear more than once, like Andy Murray, Jan Lehane, and Venus Williams, which means they reached the finals many times but could not win the title. This chart tells us that becoming a champion is very difficult, even for talented and hardworking players who make it to the final round more than once. It also shows how competitive and challenging the sport is. These players were good enough to reach the finals, sometimes even multiple times, but they always fell just short of victory. The tree map helps us appreciate their efforts and gives us a better understanding of how many players have come close but were not able to become champions.

In this visualization, I combined all the champion and runner-up names into a single field called "Player Name". Then, I created two sets: one for champions and one for runners-up. To highlight the players who were only runners-up and never became champions, I created a calculated dimension that filters for names that appear in the Runner-up Set but not in the Champion Set. This allows the visualization to focus specifically on players who reached the finals but never won a championship.

The Runner-ups section presents a multi-faceted view of players who consistently reached the Australian Open finals but fell short of securing the title. The first tree map visualizes all runner-ups, segmented by gender and colored by player, with block size indicating the number of times everyone finished second. Prominent figures such as Andy Murray, John Bromwich, and Rafael Nadal among the men, and Esna Boyd and Chris Evert among the women, emerge as persistent finalists. The second tree map, titled “Runner-ups Detailed,” adds narrative depth by highlighting players who lost multiple finals to the same champion, such as Murray’s four defeats to Djokovic and Jan Lehane’s repeated losses to Margaret Court, showing how specific rivalries shaped final outcomes. Finally, the third chart focuses on players who never won a title despite multiple final appearances, underscoring the difficulty of converting a finals berth into victory. Figures like Venus Williams, Jan Lehane, and Andy Murray reappear here, symbolizing resilience and competitive excellence, even in the absence of a trophy. Together, these visualizations pay tribute to the often-overlooked stories of perseverance and near-triumph in Australian Open history.

Conclusion

The comparison analysis of player performances at the Australian Open uncovered several meaningful patterns and outliers that add depth to our understanding of tennis history. Consistent trends revealed that top-seeded women more frequently converted their positions into victories, while male champions often defied seed expectations, winning from lower ranks. Players like Margaret Court, Serena Williams, and Novak Djokovic demonstrated sustained dominance, while others such as Andy Murray, Jan Lehane, and Venus Williams stood out for their repeated runner-up finishes, highlighting stories of both triumph and perseverance.

To bring clarity to these insights, a range of graphic techniques was used. Tree maps effectively showcased frequency-based comparisons, especially for runners-up. Global maps add geographical context to gender distribution and champion origins. Scatter plots and parallel coordinates allowed for multi-dimensional analysis of performance metrics like seed, win rate, and match statistics. These visualizations not only made the data more engaging but significantly enhanced readability by allowing viewers to detect patterns and anomalies briefly.

Using Tableau offered several advantages: its drag-and-drop interface, real-time filtering, and strong geographic mapping capabilities enabled intuitive and rapid visualization development. However, Tableau also presented some limitations, particularly in handling non-standardized data formats and restrictions on using three-letter country codes for maps, which required preprocessing. Additionally, while highly visual, Tableau can be less flexible for complex calculations or custom scripts, where programming-based tools like Python offer more control.

In conclusion, this report provided a data-driven exploration of player performances at the Australian Open, revealing how excellence, rivalry, and persistence have shaped the tournament's legacy. Through meaningful visual storytelling and analytical depth, the report celebrates not only the champions but also those whose journeys left a lasting impact on the sport, even without lifting the trophy.