

# 32146 Data Visualisation and Visual Analytics – Spring 2023

## Assessment Task 2: Advanced Data Visualisation

Name – Vishesh Kalsi

SID – 13731652

Submitted on – 13<sup>th</sup> October 2023

## Table of Contents

Summary .....	3
Dataset's Formats, Values and Characteristics.....	3
Data Transformations Used or Calculations Performed .....	3
Findings of the Analysis .....	5
Trends / Outliers.....	10
Advantages and Disadvantages of Tableau .....	12
Conclusion.....	13

## Summary

The following report has been published on data from the Australian Open tennis championships. It along with the Tableau report is aimed at people interested in tennis and will identify key trends across the entirety of the 118 years the Australian Open has run. I'll further be identifying key changes in the dataset, and focus on the top winners in the competition, defined as those with 5 or more wins across both genders. Lastly, I've produced visualisations that allow readers to both see high-level trends in the competition, as well as dive into specific match-related statistics for individual players and their games.

## Dataset's Formats, Values and Characteristics

The Australia Open championship match record, our dataset, contains 118 years of championship matches for both men and women between 1905 and 2023. The dataset contains several columns including year, gender, champion's name, champion's nationality, champion seed, match time, runner-up's name, runner-up's nationality, runner-up seed, and score.

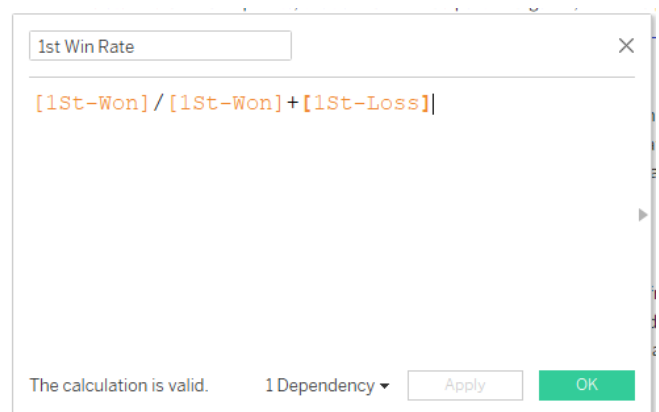
Two columns have been created and filled for champion and runner-up countries because Tableau doesn't recognise the nationality code. In the scoreboard field, data is the text format that couldn't compare. We have converted it into 10x new columns representing each set of wins and loss.

The dataset contains a set of new columns as well which have been introduced using the calculations performed that is explained below for further explanation.

## Data Transformations Used or Calculations Performed

There are mainly 4 types of calculations introduced in the dataset to make its functioning easier and more efficient across Tableau.

The first one which was done revolved around finding the "Win Rate" for a particular match played by all players. We can calculate the overall match win rate by dividing the number of games won by the total number of games played. This was done for all 5 games played during the Australia Open championship by players of both genders. The calculation involved can be seen ahead:



2nd Win Rate

$$[2\text{Nd-Won}] / ([2\text{Nd-Won}] + [2\text{Nd-Loss}])$$

The calculation is valid. 1 Dependency ▾ Apply OK

3rd Win Rate

$$[3\text{Rd-Won}] / ([3\text{Rd-Won}] + [3\text{Rd-Loss}])$$

The calculation is valid. 1 Dependency ▾ Apply OK

4th Win Rate

$$[4\text{Th-Won}] / ([4\text{Th-Won}] + [4\text{Th-Loss}])$$

The calculation is valid. 1 Dependency ▾ Apply OK

5th Win Rate

$$[5\text{Th-Won}] / ([5\text{Th-Won}] + [5\text{Th-Loss}])$$

1 Dependency ▾ Apply OK

In addition to these 2 more calculations were done: “Win Rate%” and “Set Win Rate%”. The former is self-explanatory in terms of what it means whereas the latter was calculated for each set by dividing the number of games in that set (e.g., the 1st set) by the total number of games played in that set. Mathematical explanation on this can be found ahead:

Win Rate %

$$\frac{([1\text{St-Won}] + [2\text{Nd-Won}] + \text{IFNULL}([3\text{Rd-Won}], 0) + \text{IFNULL}([4\text{Th-Won}], 0) + \text{IFNULL}([5\text{Th-Won}], 0))}{([1\text{St-Won}] + [2\text{Nd-Won}] + \text{IFNULL}([3\text{Rd-Won}], 0) + \text{IFNULL}([4\text{Th-Won}], 0) + \text{IFNULL}([5\text{Th-Won}], 0) + [1\text{St-Loss}] + [2\text{Nd-Loss}] + \text{IFNULL}([3\text{Rd-Loss}], 0) + \text{IFNULL}([4\text{Th-Loss}], 0) + \text{IFNULL}([5\text{Th-Loss}], 0))}$$

Set Win Rate%

$$\frac{(\text{IF } [1\text{St-Won}] > [1\text{St-Loss}] \text{ THEN } 1 \text{ ELSE } 0 \text{ END} + \text{IF } [2\text{Nd-Won}] > [2\text{Nd-Loss}] \text{ THEN } 1 \text{ ELSE } 0 \text{ END} + \text{IF } [3\text{Rd-Won}] > [3\text{Rd-Loss}] \text{ THEN } 1 \text{ ELSE } 0 \text{ END} + \text{IF } [4\text{Th-Won}] > [4\text{Th-Loss}] \text{ THEN } 1 \text{ ELSE } 0 \text{ END} + \text{IF } [5\text{Th-Won}] > [5\text{Th-Loss}] \text{ THEN } 1 \text{ ELSE } 0 \text{ END})}{(\text{IF } \text{ISNULL}([1\text{St-Won}]) \text{ THEN } 0 \text{ ELSE } 1 \text{ END} + \text{IF } \text{ISNULL}([2\text{Nd-Won}]) \text{ THEN } 0 \text{ ELSE } 1 \text{ END} + \text{IF } \text{ISNULL}([3\text{Rd-Won}]) \text{ THEN } 0 \text{ ELSE } 1 \text{ END} + \text{IF } \text{ISNULL}([4\text{Th-Won}]) \text{ THEN } 0 \text{ ELSE } 1 \text{ END} + \text{IF } \text{ISNULL}([5\text{Th-Won}]) \text{ THEN } 0 \text{ ELSE } 1 \text{ END})}$$

# Findings of the Analysis

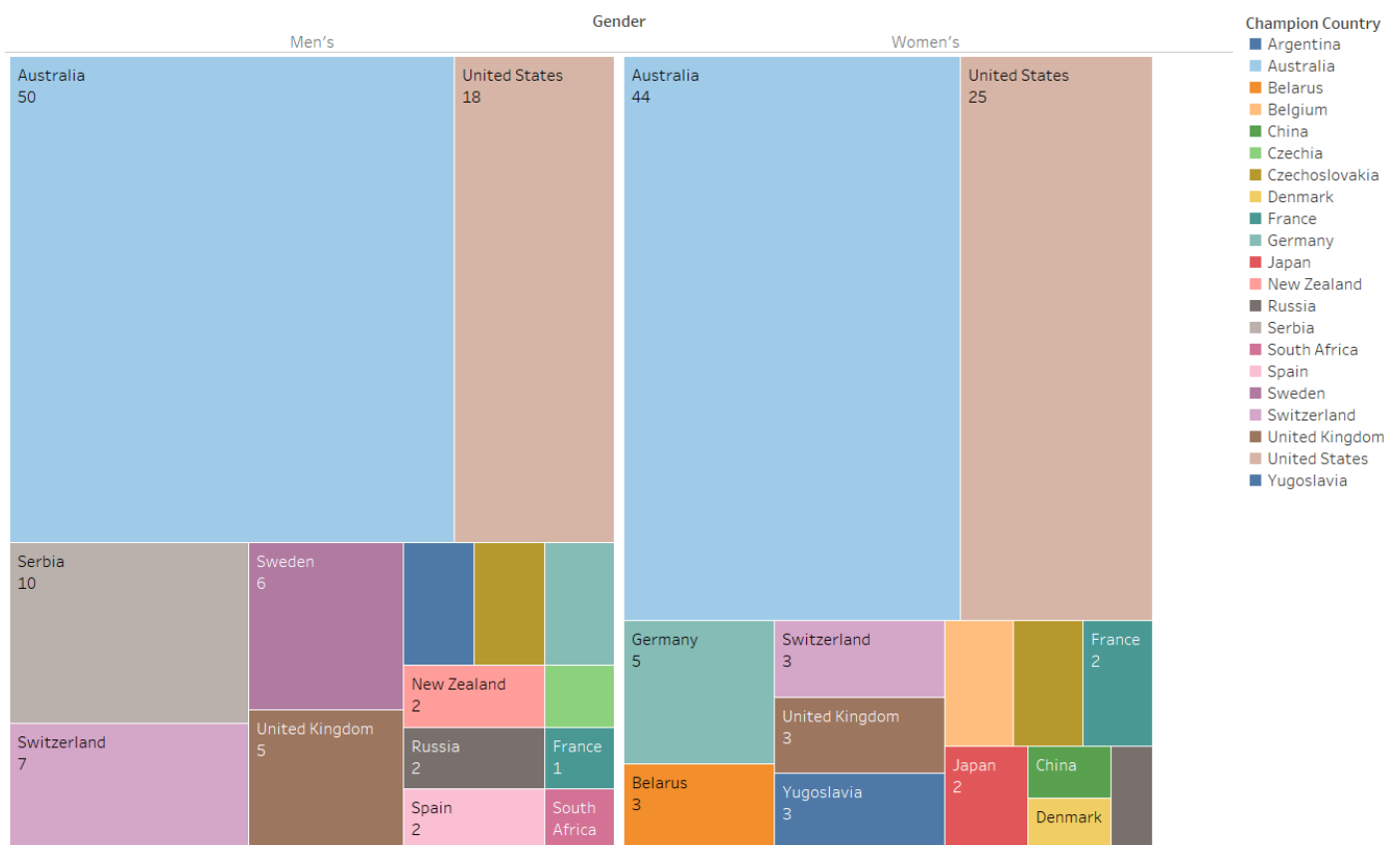
I've encapsulated a total of 2 dashboards and 10 sheets for analyzing the dataset provided. It was tested against various parameters on lots of filters. All the visualizations created in the attached Tableau file have been elucidated upon below and more information can be obtained using the file itself.

The first visualisation created was a Treemap. A treemap displays structured hierarchical data as a set of nested rectangles (or leaves) by using different sizes and colours. The leaf size illustrates the data value and the colours show separate data categories.

In the treemap on the below, you can see the results of the Countries during the entire period for both genders. Left side represents men's and right women's. The size of the rectangle represents the count, i.e. the number of Australian Open championships won.

The readability of this report was enhanced using the different colours for both genders which made it easier for the user to read. Along with legend was introduced for the sake of reference to categories.

Treemap

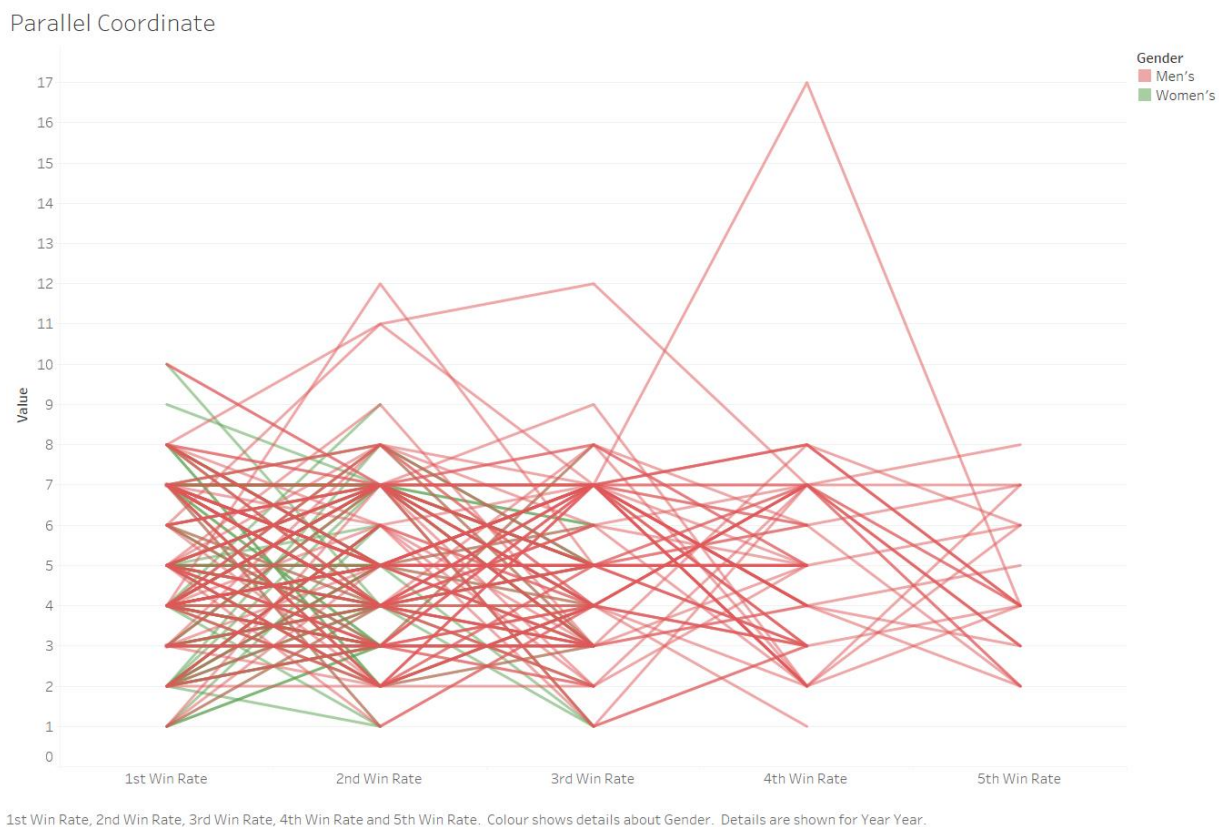


Champion Country and count of Australian Open broken down by Gender. Colour shows details about Champion Country. Size shows details about count of Australian Open. The marks are labelled by Champion Country and count of Australian Open.

The next visualization created was a Parallel coordinate. These are a popular tool for high dimensional data visualisation. The chart represents each data dimension as a parallel axis and draws polylines between independent axes at appropriate values.

The value measured against the axis was the 5-win rates calculated earlier against each other. This was further divided by Gender with the legends being shown below:

The readability of this report was enhanced using the different colours for both genders which made it easier for the user to read. Along with legend was introduced for the sake of reference to categories.

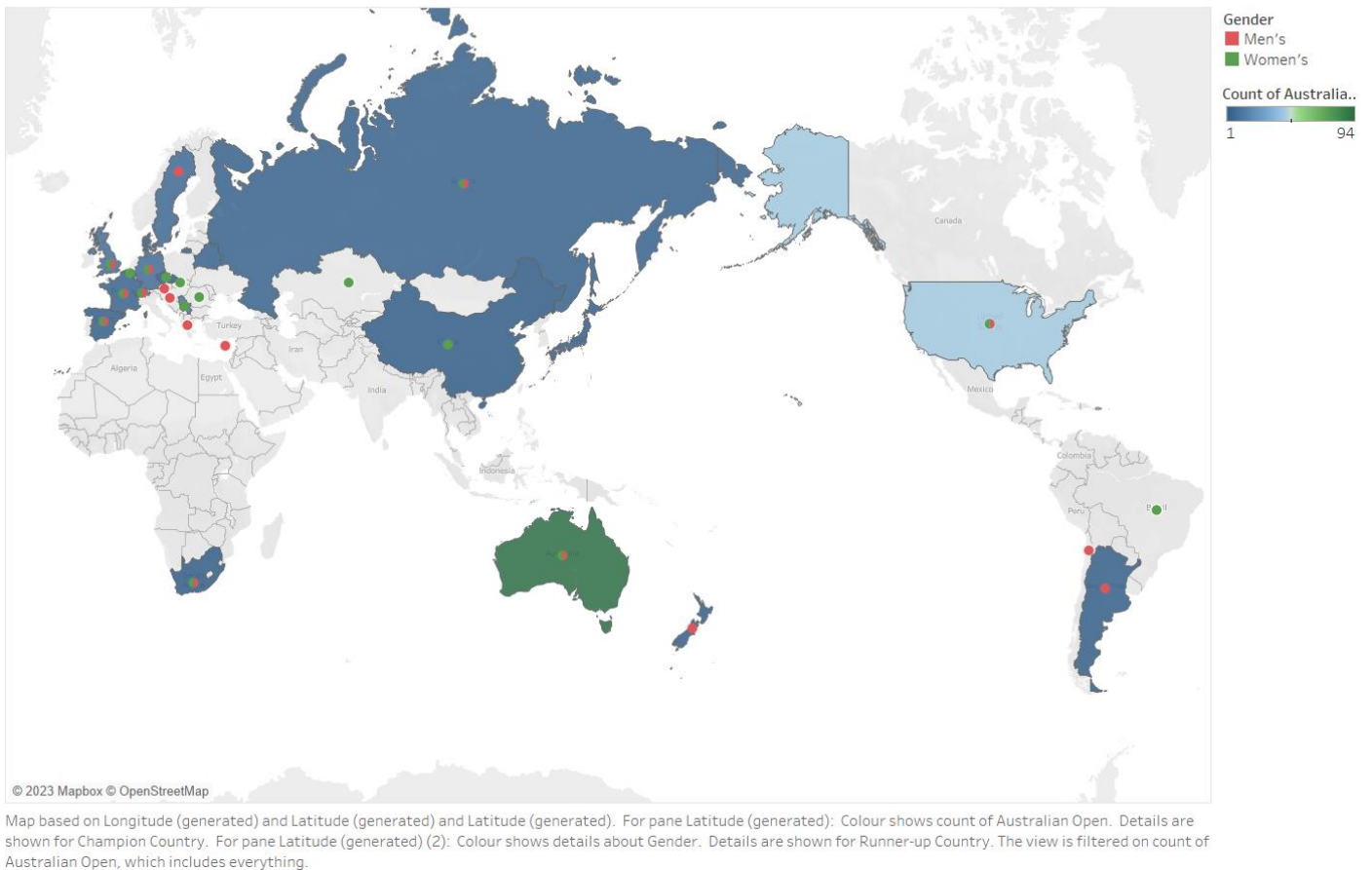


The next visualization was a Geographic Map. Geographic Information System (GIS) is a conceptualised framework that provides the ability to capture and analyse spatial and geographic data. GIS applications (or GIS apps) are computer-based tools that allow the user to create interactive queries (user-created searches), store and edit spatial and non-spatial data, analyse spatial information output, and visually share the results of these operations by presenting them as maps.

In the Geographic Map below, one can find data about the country's championships depicted on a scale of colours from blue to green, as on the right hand side in legend against each other. Also, you can find out the split of runner-up trophies won by the country using the pie chart.

The readability of this report was enhanced using the different colours for all the countries, which made it easier for the user to read. Along with legend was introduced for the sake of reference to categories for both visualisations.

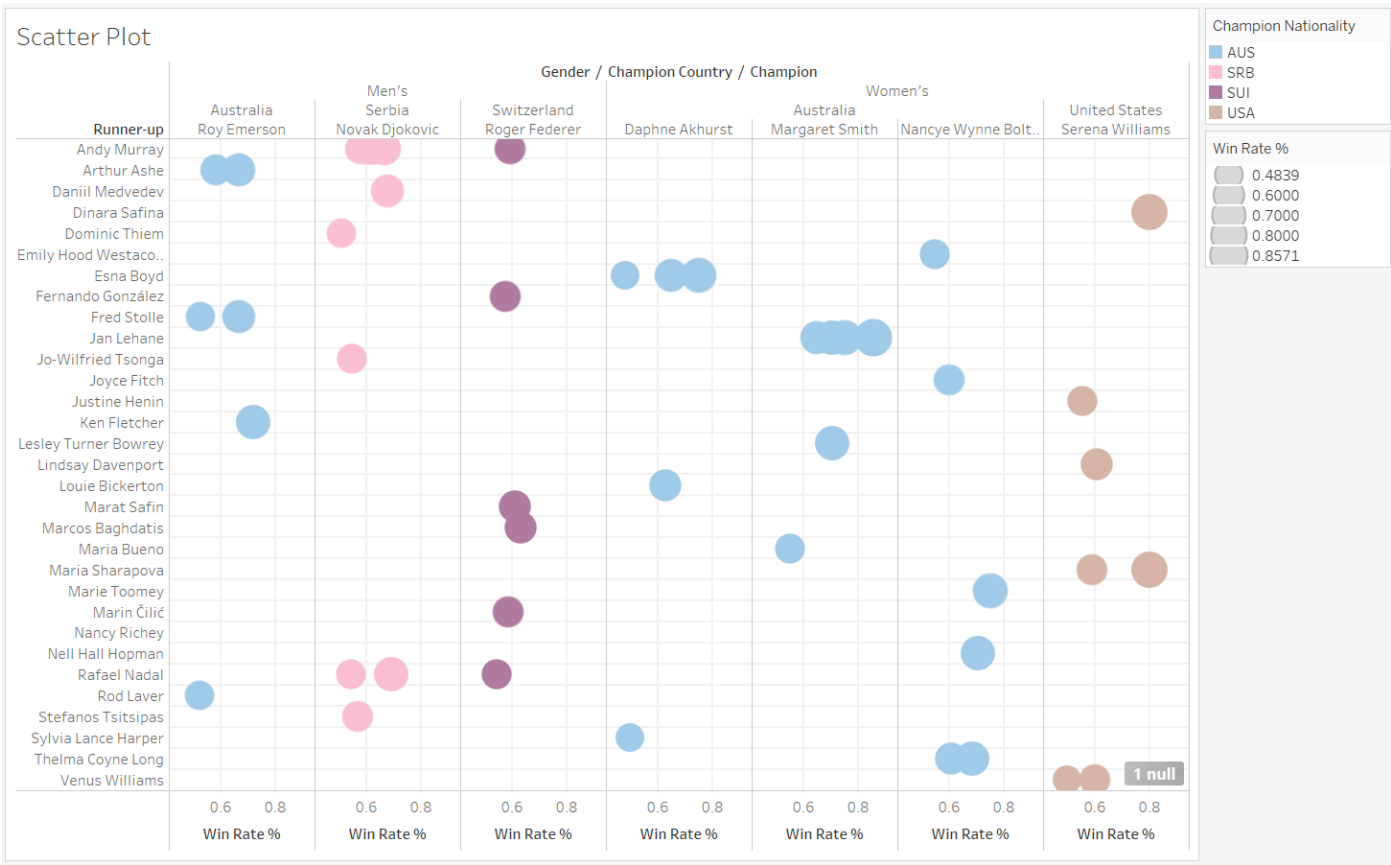
Geographic Map



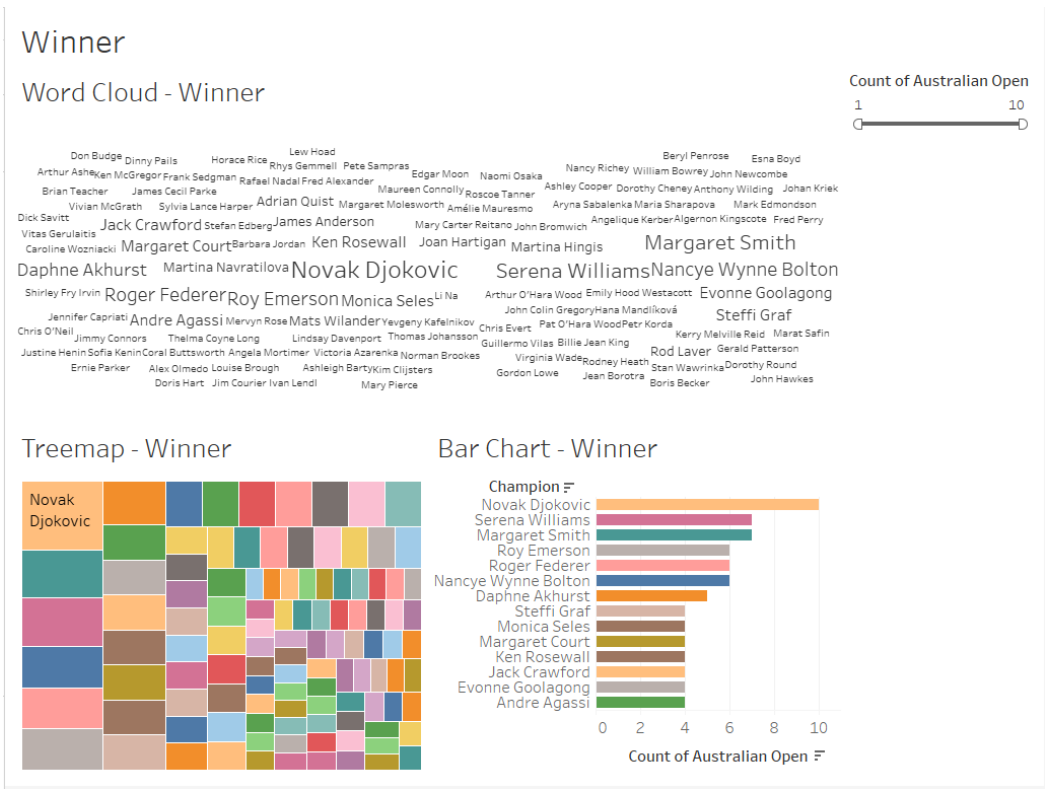
The next visualization created was a Scatter plot. Scatter plots are a handy way to visualise numerical data. They typically plot the relationship between two numerical data points, plotting one along the X axis (horizontal, referred to as "Columns" in Tableau) and one along the Y axis (vertical, referred to as "Rows" in Tableau).

In the Scatter plot below, you can find out data about the player's win rate against each competitor. This tells us how each player on x-axis performed against each competitor(runner-up) on y-axis. Also, we can see that this visualisation provides us with an insight into the top 5 players for the period across both genders.

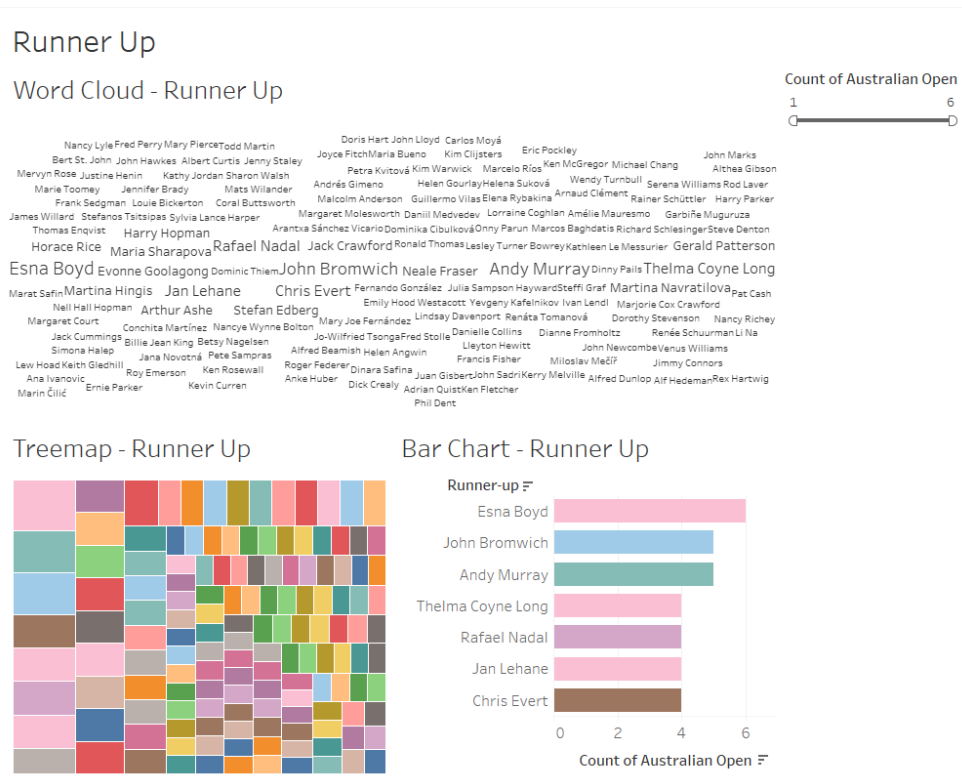
The readability of this report was enhanced using the different colours for all the players, which made it easier for the user to read. Along with legend was introduced for the sake of reference to categories.



Along with the fore mentioned visualltions a dashboard was created each for the winners and losers. The dashboard compared the winners and losers, respectively against each other by use of various techniques. This was also supported by use of filters in the dashboard so that it can be made interactive for the user to play with. Both dashboards can be found ahead:







We can see that this is made up of 3 sheets each. The first one being a word cloud that depicts name of the winner based on the number of titles won. More number of titles depict the size being bigger. Also, a comparison was established using the bar charts that can be found next to the treemap. These 2 compared the number of titles won/runnered-up for each player from a more analytical perspective.

Word Cloud - Winner



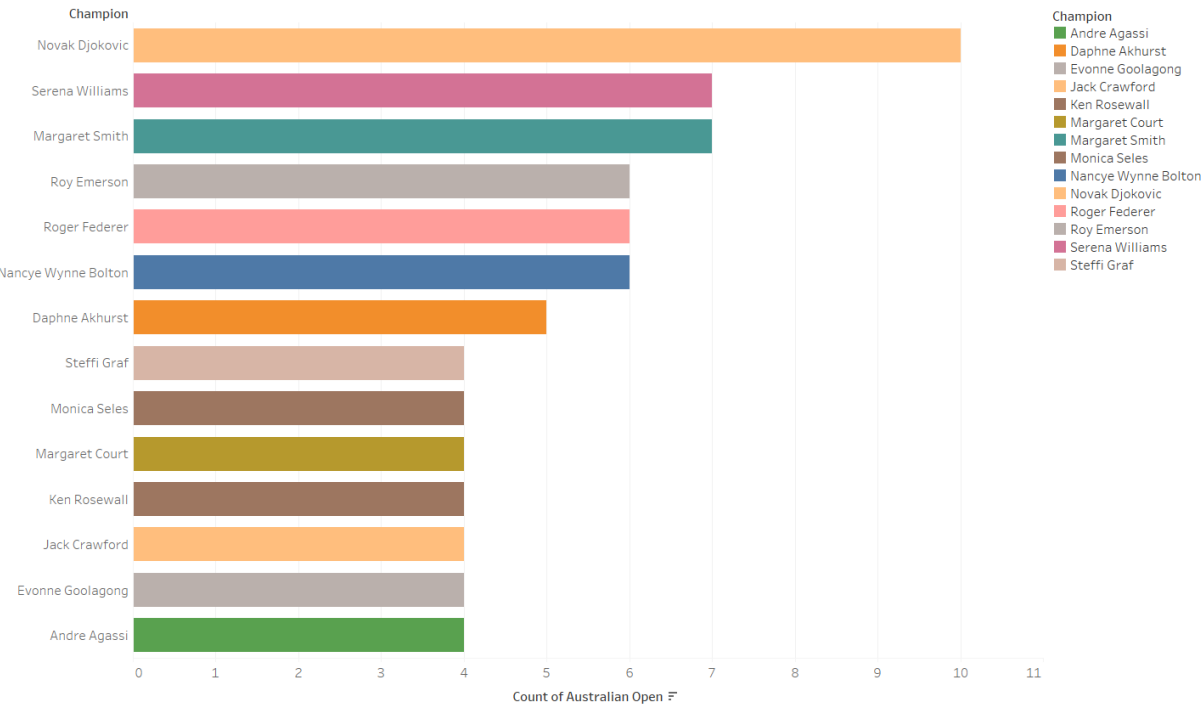
Champion - Size shows count of Australian Open

Word Cloud - Runner Up



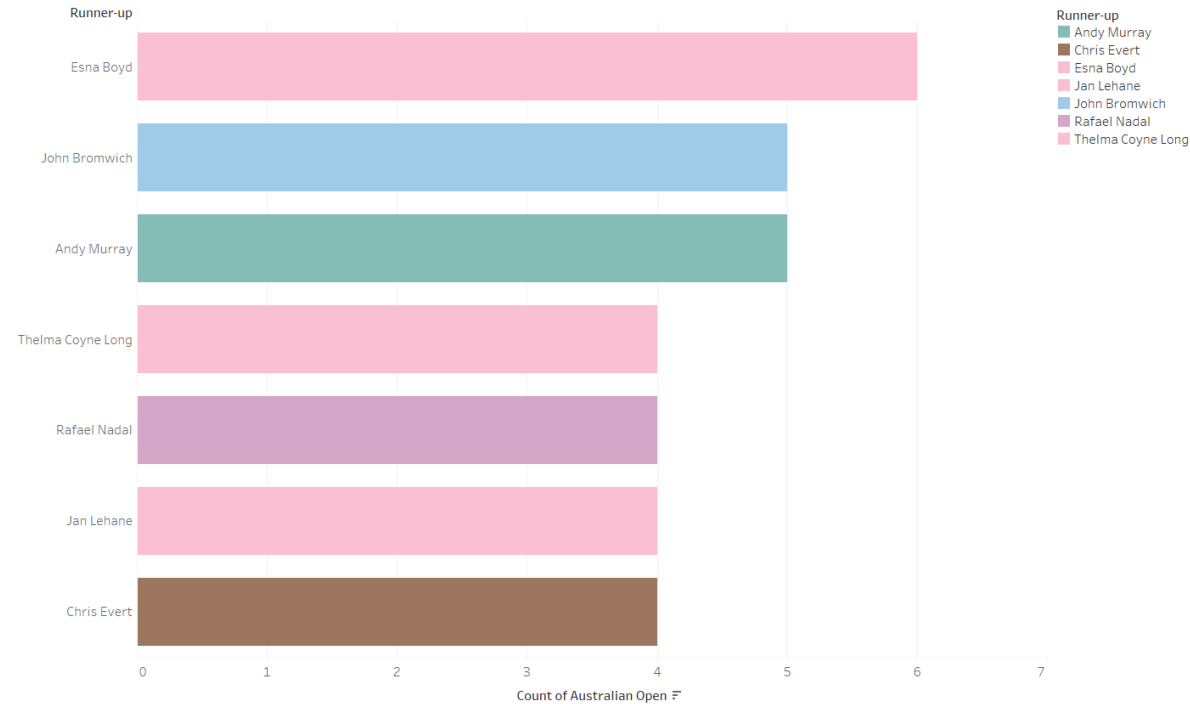
Runner-up - Size shows count of Australian Open

Bar Chart - Winner

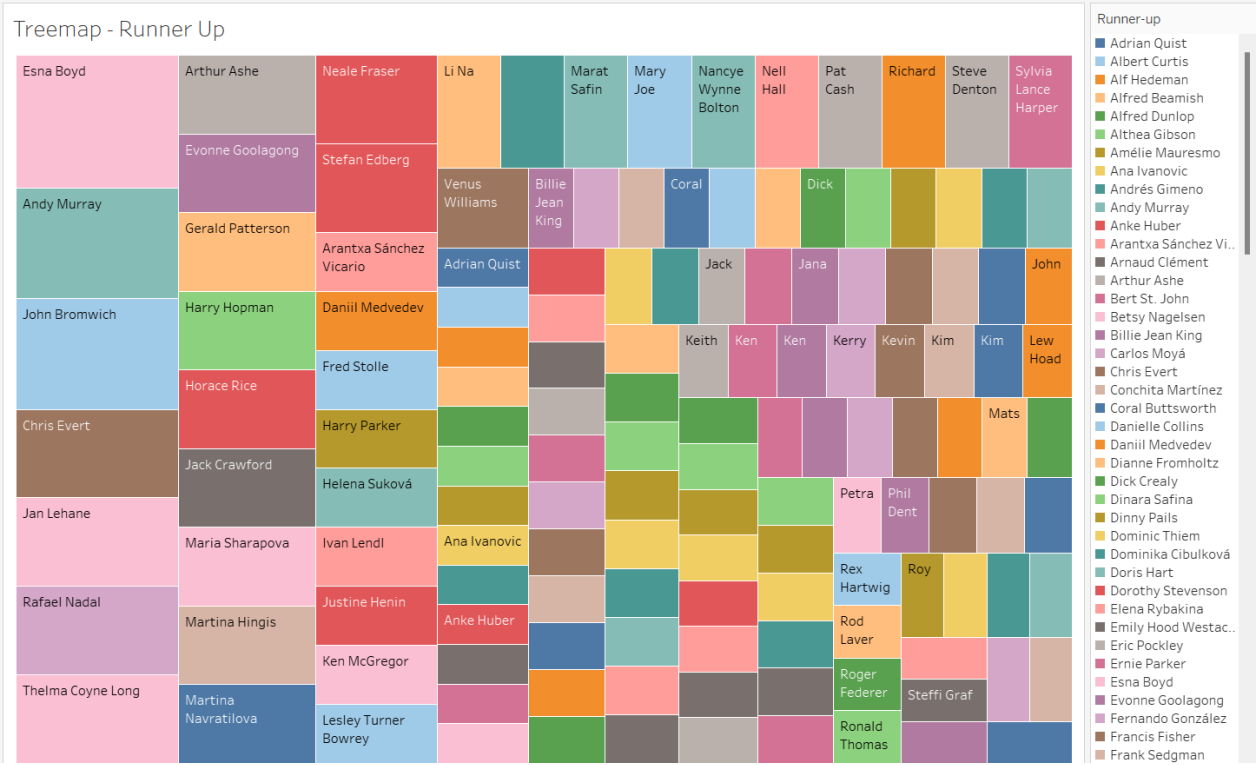
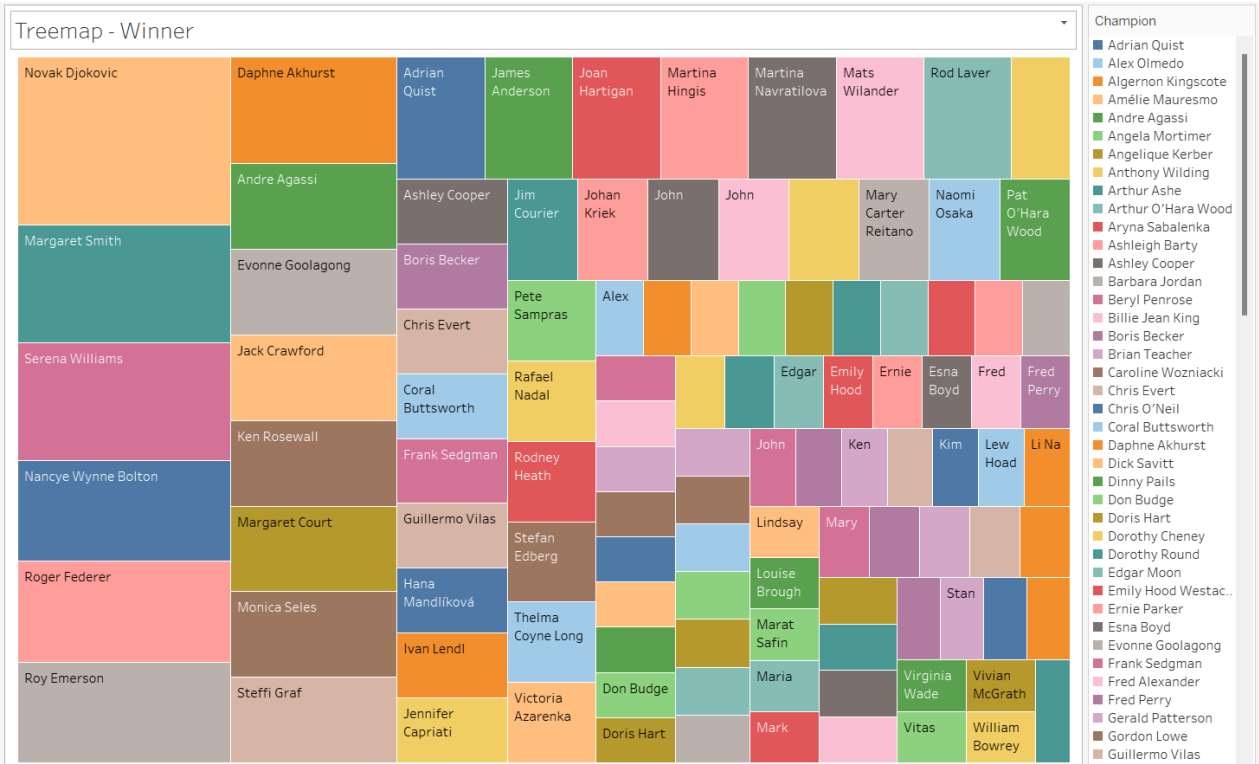


Count of Australian Open for each Champion. Colour shows details about Champion. The view is filtered on count of Australian Open, which ranges from 4 to 10.

Bar Chart - Runner Up



Count of Australian Open for each Runner-up. Colour shows details about Runner-up. The view is filtered on count of Australian Open, which ranges from 4 to 6.



## Trends / Outliers

A couple of trends / outliers were observed during analyzing the dataset both using Excel and Tableau. They are listed ahead as follows:

- Australia won most titles throughout the entire period for championships. In the men's category they won 50 whereas 44 was the total of the women's. This was for off by their closest competitor was USA which had 18 and 25, respectively for men's and women's category.
- Australia also was the country with most runner-up titles. In the men's category they had 56 whereas 42 was the total of the women's. This was for off by their closest competitor was USA which had 11 and 22, respectively for men's and women's category.
- Despite of the fore stated facts, winners from men's and women's category were not from Australia. Rather Novak Djokovic and Serena Williams hailed from Serbia and the United States, respectively.
- Men's category in 1927 had the highest win rate ever achieved across genders over the years of 17. This was observed for the 4<sup>th</sup> win rate. Rest of the categories did have majorly men dominating the win rate with this being the highest.
- European countries such as France, Austria, Spain and the ones around did have a high participation rate. Even though the titles won by these was little to the States however, it depicted a higher participation rate.
- The number of titles won by men's winner was close to 1.67 times that of achieved by women's. This showed us that women in general were less involves in the sport when compared to men over the tenure.

## Advantages and Disadvantages of Tableau

The advantages can be found ahead:

- Data visualization – Tableau is primarily a tool for data visualization. Therefore, it's technology supports intricate calculations, data blending, and dashboarding so that stunning visualizations may be produced that provide insights that are difficult to obtain by simply glancing at a spreadsheet. Due to its commitment to achieving this goal, it has risen to the top of the data visualization heap.
- Quickly Create Interactive visualizations - With Tableau's drag-and-drop features, the user may quickly construct a highly interactive display. The interface can accommodate many modifications while also preventing you from producing charts that go against recommended guidelines for data display. The Tableau Gallery has some of the most incredible images ever made on display.
- Ease of Implementation – Tableau offers a wide range of options for visualization that improve the user experience. with addition, Tableau is far simpler to learn than Python,

Business Objects, or Domo; someone without any programming experience can quickly become proficient with Tableau.

The disadvantages can be found below:

- Scheduling or notification of reports – Tableau does not offer the option of using scheduling to automatically refresh reports. Tableau does not offer a schedule option. Therefore, if users need to change data in the back end, some human labor is always needed.
- No Custom Visual Imports – Tableau isn't a fully functional open tool. Developers can produce bespoke graphics that are simple to import into Tableau, unlike competing tools like Power BI. Any new visuals must therefore be produced from scratch rather than imported.
- Custom formatting in Tableau - Users complain about Tableau's conditional formatting and its 16-column table display limitations. Additionally, there is no direct means for a user to apply the same formatting to all fields at once. Users must complete that manually for each field, which takes a lot of time.

## Conclusion

To conclude, an immersive experience has been aimed to provide the user with for the provided dataset. It was a great opportunity given to analyse the dataset and drive insights using the various fore mentioned techniques. Furthermore, it showed that there is a need for women to be more engaged in playing the game along with Australia to continue dominating its effect.