Throughout my work experience, I supported various business teams by using different software packages. But the one software package I used the most is Excel. I used Excel for data cleanup, data validation, and data classification. So, I thought it would be fun to use my professional experience of Excel and apply it to my personal interest.

Tennis has always been my favorite sport to watch and follow. There is lots of data involved in the sport with the number of tournaments played, number of titles won, wins, losses, and winning percentile range. So, I knew that the best way to organize this data is in Excel. I came up with the idea to collect data, organize it in an Excel spreadsheet, and analyze that data in several different ways. I will break down this data analysis project into smaller pieces to show you guys how to put this together step-by-step.

**PART 1**

First, I collected the data for all the currently active women’s singles Grand Slam champions from the Women’s Tennis Association (WTA) website (<https://www.wtatennis.com/>). I created Excel workbooks for each woman. When I organized each woman’s data in an Excel workbook, I split each workbook into worksheets. I named the worksheet with the year that each woman started making her debut on the WTA tour. In each worksheet, I had the following headings listed:

* TOURNAMENT
* SURFACE
* ROUND
* OPPONENT
* RESULT
* SCORE

After I organized my data, I formatted the “RESULT” column using cell styles. Cell styles gave my spreadsheet a polished look and made my data easy to interpret. In my work experience, I always use the “Color Categories” in Outlook to prioritize my tasks because it helps me keep track of what I get done. Red means a task is high priority and needs to be done, yellow means that a task is pending on hold, and green means that a task has been 100% completed. So, I applied this same tactic to my Excel spreadsheet. In my “RESULT” column, the “Win” result is highlighted in green, while any “Loss” result is highlighted in red. Green is positive and red is negative.

Wins and losses are straight-forward results. But a walkover is different. A walkover occurs when a player withdraws before the start of a match, or a player wins a match without facing their opponent on court. So, the result doesn’t count at all towards a player’s record of wins and losses. When my “SCORE” column had the result of “Walkover”, my “RESULT” column had one of two results – “Won” or “Withdrew”. In my “RESULT” column for the cells that had “Won” or “Withdrew” with the “SCORE” of “Walkover”, I customized those cells with a style of “Neutral” in yellow.

**Part 2**

Based on the worksheets that I created for each year, I created a new worksheet called “YTD Stats”. I organized the “YTD Stats” worksheet into a table with the following headings:

* YEAR
* NUMBER OF TOURNAMENTS PLAYED
* NUMBER OF TITLES WON
* WINS
* LOSSES
* WINNING PERCENTILE RANGE

After I filled in my table with the data, I made calculations for the TOURNAMENTS PLAYED, TITLES WON, WINS, LOSSES, and WINNING PERCENTILE RANGE headings. This is where I used formulas. Formulas are expressions that calculate the values of cells. When I created my formulas, I always started them with an equal sign (=). The first formula I created was the sum, which adds up all the values in all the cells. My sum for TOURNAMENTS PLAYED looked like this:

=SUM(B2:B25)

I can manually enter in this formula for the TITLES WON, WINS, LOSSES, and WINNING PERCENTILE RANGE headings. But it would take a long time. So, the easiest, quickest way to copy this formula would be to use the fill down tool. The fill down tool lets me quickly copy formulas into adjacent cells by dragging the fill handle. The fill down tool is a black cross located in the lower right-hand corner of a cell. It can be used vertically down or horizontally to the right when a formula is copied over.

I followed the same step to calculate the average for each heading. Average calculates the average of numbers provided. But the difference is which function comes after the equal sign. The average formula looks like this:

=AVERAGE(B2:B25)

After I created my average formula, I created a new formula for the winning percentile range. To calculate the winning percentile range, I took the difference between the wins and losses and divided that amount by the total number of wins. So, my winning percentile range formula looks like this:

=(D2-E2)/D2

Just like when I created my sum and average, formulas, I followed the same step in using the fill down tool to drag my winning percentile range formula down vertically. That way, the formula would copy down from the first year up until the current year.

After I created my formulas, I formatted my values in the WINNING PERCENTILE RANGE column as percentages rounded off to the nearest whole number. To do this, I highlighted my values in the WINNING PERCENTILE RANGE column, navigated to the “Home” tab, and went to the “Number” group. In the “Number” drop-down menu, I selected “Percentage”. When I clicked on “Percentage”, my values in the WINNING PERCENTILE RANGE column had a percent sign at the end. To round off the values to the nearest whole number, I clicked on “Decrease Decimal” twice to get rid of the two places after the decimal point. I repeated this process to calculate the winning percentile ranges for the SUM and AVERAGE rows. Those formulas looked like this:

* SUM: =(D26-E26)/D26
* AVERAGE: =(D27-E27)/D27

**Part 3**

Now that I completed my formulas and made my necessary calculations, I am ready to add more customizing to my YTD Stats worksheet. This is where I used conditional formatting. Conditional formatting lets me analyze data and identify patterns and trends in that specific set of data. I knew that the winning percentile range would be the best column, where I can apply conditional formatting. To do this, I highlighted my percentages in the WINNING PERCENTILE RANGE column, navigated to the “Home” tab, and went to the “Styles” group. In the “Styles” group, I clicked on “Conditional Formatting”. In the “Conditional Formatting” drop-down menu, I have multiple options to choose from. I chose the “Icon Sets” option. In the “Icon Sets” option, I have four different categories - “Directional”, Shapes”, “Indicators”, and “Ratings”. I felt that it would be the easiest and best choice to see three different icons for my winning percentile ranges. So, I chose the “3 Symbols - Uncircled” option. When I clicked on the “3 Symbols - Uncircled” option, I can see that the WINNING PERCENTILE RANGE column got formatted with green checkmarks, yellow exclamation points, and red “X” marks.

So, this is how I analyzed Serena’s data with conditional formatting. Hopefully, this video was helpful, and I will see you in the next one!

**Part 4**

Welcome back! We are in the final stages of completing a polished Excel spreadsheet of the data. The last piece we have left is creating charts. Charts are powerful tools that let me visually display data in a variety of different charts.

The first chart I created was to show each woman’s YTD wins and losses. For that, I created a clustered column chart. A clustered column chart compares values across categories. I highlighted the YEAR, WINS, and LOSSES columns, navigated to the “Insert” tab, went to the “Charts” group, and clicked on “Recommended Charts”. When I clicked on “Recommended Charts”, I got a “Insert Chart” pop-up window. I clicked on the “Clustered Column Chart” option. When I selected the “Clustered Column Chart” option, I got a clustered column chart on my “YTD Stats” worksheet. I also got two new tabs at the top of my Excel window Ribbon - “Chart Design” and “Format”. In my clustered column chart, my legend is formatted with default colors to represent wins and losses. But I wanted to customize the colors to show wins in green and losses in red. To do this, I selected the chart, right clicked on the “WINS”, and set the “Fill” and “Outline” colors to “Green”. I repeated the same step for “LOSSES”. But for the losses, I set the Fill” and “Outline” colors to “Red” instead. At the top of my chart, I have “Chart Title” in the center. I changed the title so that it reflects Serena’s YTD wins and losses. I selected the chart, clicked on the “Chart Title” text box, and changed the title to each woman’s name. Next, I added titles for my Axes. The X axis for the year goes horizontally from left to right, while the Y axis for the number of matches goes vertically from bottom to top. I selected the clustered column chart, navigated to the “Chart Design” tab, and went to the “Chart Layouts” group. In the “Add Chart Element” drop-down menu, I clicked on “Axis Titles”. For the X axis, I clicked on “Primarily Horizontal”, and I named it as “YEAR”. I repeated this step for the Y axis too. But I clicked on “Primarily Vertical” instead and I named it as “NUMBER OF MATCHES”. I followed this process to create a second chart to analyze Serena’s winning percentile range. But instead of a clustered column chart, I created a line chart. A line chart is used to show trends over a certain period. Just like when I created the clustered column chart, I still highlighted the “YEAR” column. But this time, I selected the “WINNING PERCENTILE RANGE” column instead. I followed the same steps when I navigated to the “Insert” tab, went to the “Charts” group, and clicked on “Recommended Charts”. But unlike the first chart when I chose “Clustered Column”, I selected the “Line Chart” option instead. I set the “Fill” and “Outline” colors to “Green” for the line that displays for all the years. Then, I named my line chart as “23-time Grand Slam Champion Serena Williams (USA): Winning Percentile Range”. Just like my first clustered column chart, I followed the same steps to name my X and Y axes. I still named my X axis as “YEAR”. But my Y axis was “PERCENTAGE”. The last piece I added to my line chart was a trendline. A trendline can tell me an upward slope or trend in my data and it can help me predict the future. I wanted to see Serena’s winning percentile range for the next years in 2022. For me to add a trendline, I selected the line chart, navigated to the “Chart Design” tab, and went to the “Chart Layouts” group. In the “Add Chart Element” drop-down menu, I clicked on “Trendline”. When I clicked on “Trendline”, I had multiple options to choose from. I wanted to customize my trendline. So, I clicked on “More Trendline Options”. When I clicked on “More Trendline Options”, I got a “Format Trendline” sidebar on the right side of my spreadsheet. I can see that my trendline option is set to “Linear” as default. I will leave the option as it is. Next, I will go to the “Forecast” section. This is where I want to see Serena’s winning percentile range for 2021 and 2022. In the “Forecast section”, I typed in “1” for “Forward” so that my line chart will display the winning percentile range for the next year. When I typed in “1” for “Forward”, a new line was added on my line chart. That new line shows Serena’s winning percentile range for 2022. I customized the color for the new trendline to show in “Green”. This analysis is telling me that Serena’s winning percentile range for 2022 is between 80% and 90%.

So, this is how I analyzed Serena’s data. To my family and friends out there who are tennis fans, if you want to try this with either 17-time Grand Slam Champion Novak Djokovic or 19-time Grand Slam Champion Rafael Nadal, I say go for it! You never know how amazing data analysis in tennis can be! I hope all of you enjoyed my videos and I really hope to see you soon!