Exploring Tennis Players' Earnings in SAS Visual Analytics

Industry Aligned Activity

Purpose

This activity guides students through building an interactive report aimed to better understand the current compensation structure for the sport of tennis. Students will create an informative dashboard to explore the data and create visualizations that identify potential patterns.

SAS Software

This activity is completed using SAS Visual Analytics in the 2024.03LTS release of SAS Viya 4. Other versions of our SAS Viya software can be used, but there will be minor discrepancies in the screenshots.

Industry Alignment

This activity aligns with the sports analytics industry. It explores compensation for athletes, especially tennis players, to identify aspects that lead to monetary success.

Contents

Activity Notes and Requirements

Learning Objectives

This activity provides practice with skills such as:

- Manipulating data for cleaner analysis
- Creating graphs and charts for analysis
- Adding interactivity to allow for user input
- Enhancing reports for better viewing
- Interpreting graphs to gather insights

Estimated Completion Time

Completion time will vary based upon student level. This activity should take students approximately 90-120 minutes to complete on average.

Experience Level

This problem set is targeted for beginner-intermediate users of SAS Visual Analytics.

Prerequisite Knowledge

Software

Students should be comfortable with the basics of visualizations and dashboarding in SAS Visual Analytics, utilizing SAS Viya for Learners.

Content Knowledge

Students do not need any prior knowledge about the sports industry or how athletes are paid. All sports-related references will be explained.

Data Source

Introduction

The analysis begins by working with the **Forbes_Richest_Athletes_1990-2020.csv**, which can be found on Kaggle (https://www.kaggle.com/datasets/parulpandey/forbes-highest-paid-athletes-19902019). For each year, the top 10 highest paid athletes are included. Earnings totals include salaries, bonuses, prize money, endorsements, and appearance fees. To help our exploration, the data was cleaned by adding the Team_Sport column (based on https://en.wikipedia.org/wiki/Team_sport), the column names were standardized, and observation numbers were removed.

The next dataset used is **Athletes_Earnings.xlsx**, which can also be found on Kaggle (https://www.kaggle.com/datasets/dimitrisangelide/top-10-highestpaid-athletes-tennis-nba-soccer). This data was extracted from Forbes on the top 10 highest paid athletes in tennis, NBA, and soccer from 2011-2021. To better display how athletes earn their money, their total earnings are broken down into prize money and endorsements. Data prep completed includes standardizing variable names and adding the Team_Sport column based on whether the Team variable was missing or not.

In Part 2 of the activity, we introduce the **demo_rank_matches.sas7bdat** dataset. This SAS dataset was created by combining multiple data sources from the Git repositories maintained by Jeff Sackman (https://github.com/JeffSackmann). The repositories contain master files for players in the Association of Tennis Professionals (ATP) and The Women's Tennis Association (WTA), including demographic information, historical rankings and match results for tournaments played. SAS scripts required for data manipulation, merging and links to source files can be found here: https://github.com/antonbcristina/SAS-Tennis-Asset.

All datasets for this activity can be downloaded from https://github.com/antonbcristina/SAS-Tennis-Asset.

Note: The data used in this activity was pulled from third party sources and it is beyond the scope of this activity to verify its accuracy and consistency. Datasets will be used as is, except for the changes outlined above and noted in the *Description of Variables* section below.

Description of Variables

The variables in **Forbes_Richest_Athletes_1990-2020.csv** used for this activity are:

Variable	Description
Name	Name of athlete
Nationality	Country represented when competing
Current_Rank	Earning rank for the current Year
Previous_Year_Rank	Earning rank for the previous Year
Sport	Sport athlete competes in
Year	Year data was extracted for – 1990 to 2020
Earnings	Total earnings in millions (includes salary, bonuses, prize money, endorsements and appearance fees)
Team_Sport	Indicates whether the sport competed in is compensated as a team sport (Yes vs. No)

The variables in **Athletes_Earnings.xlsx** used for this activity are:

Variable	Description
Player	Name of athlete
Nationality	Country represented when competing
Sport	Sport athlete competes in
Team	Sports team played for (missing for Tennis players)
Team_Sport	Indicates whether the sport competed in is compensated as a team sport (Yes vs. No)
Year	Year data was extracted for – 2011 to 2021
Earnings_Rank	Earning rank for the current year
Total_Earnings	Total earnings in millions (includes prize money and endorsements)
Prize_Money	Money earned from tournament wins in millions
Endorsements	Endorsement money earned in millions

The variables in **demo_rank_matches.sas7bdat** used for this activity are:

Variable	Description
Player_ID	Unique player identifier
Player	Name of athlete
DOB	Date of birth
Gender	Gender of player - M (male for ATP) or F (female for WTA)
Hand	Dominant playing hand - R (right) or L (left)
IOC	International Olympic Committee nation
Height	Height of player in cm
Year	Year data was extracted for – 2012 to 2021
Rank	Highest rank achieved in specified year
Points	Number of points achieved at highest rank in specified year
Tourneys	Number of tournaments attended
Slams	Number of Grand Slam tournaments attended
Matches	Number of matches played
W	Number of matches played and won
L	Number of matches played and lost
Win_Per	Win percentage, calculated as W/Matches

Max_Ace	Maximum number of aces in a match
Max_Df	Maximum number of double faults in a match
Total_Ace	Total number of aces in a match
Total_Df	Total number of double faults in a match

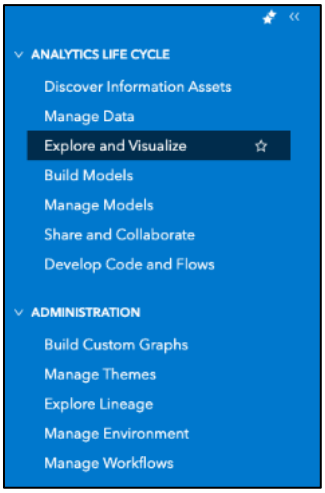
Exploring Tennis Players' Earnings

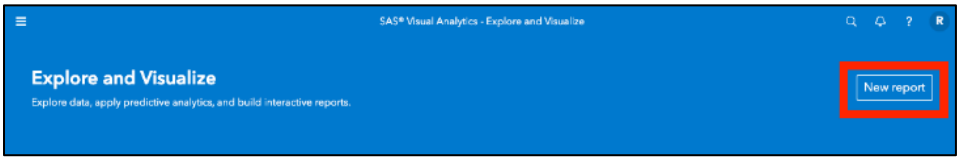
In the world of professional sports, the earnings of athletes draw significant attention and often capture headlines. Amongst individual sports, tennis stands out not only for its thrilling matches but also for the substantial earnings of its top players. As independent contractors, tennis players generate income through diverse avenues: prize money, endorsement deals, lucrative contracts, and event appearances.

We will investigate how the earnings of individual sports compare to team sports and the structure of income for tennis players. We will take a closer look at the gender pay gap for the sport of tennis and what factors are associated with monetary success. Lastly, our goal will be to shed light on the interplay between athletic achievements and financial success of top-ranked tennis players.

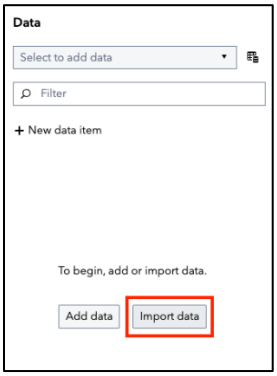
Part 1: Overall Sports Compensation Analysis

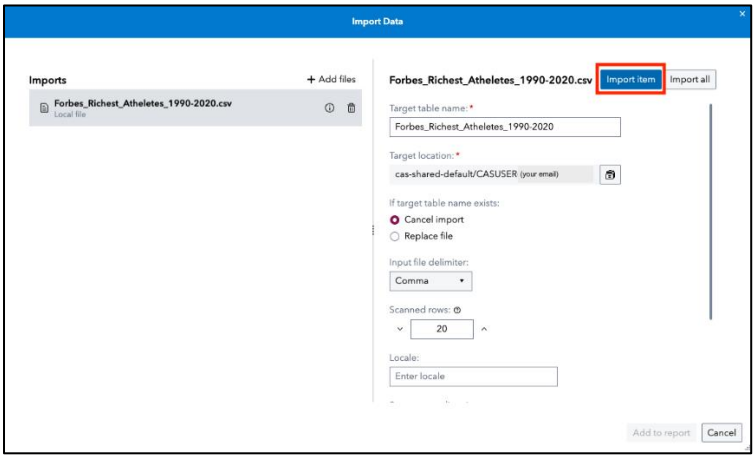
1. To begin, let's start a new report within **Explore and Visualize**! Some prompts to get you started:



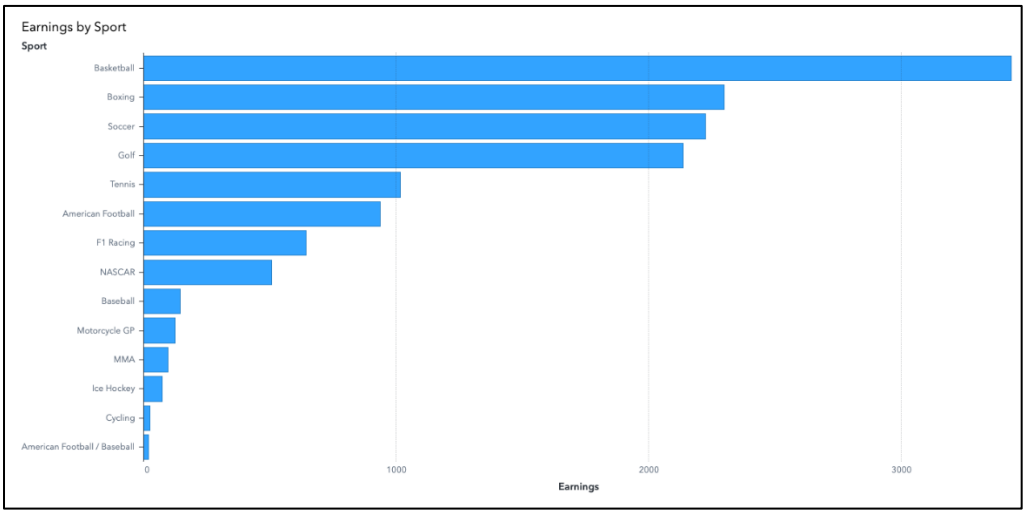


2. Our first analysis focuses on all sports, so we will import and add the **Forbes_Highest_Paid_Athletes_1990-2020.csv** dataset to the report. Place the dataset in your personal CASUSER folder. Otherwise, you can accept the defaults when importing datasets. Some helpful screenshots to guide you along:

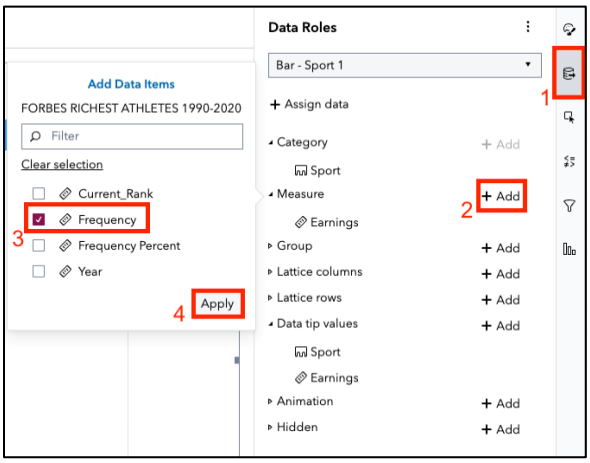




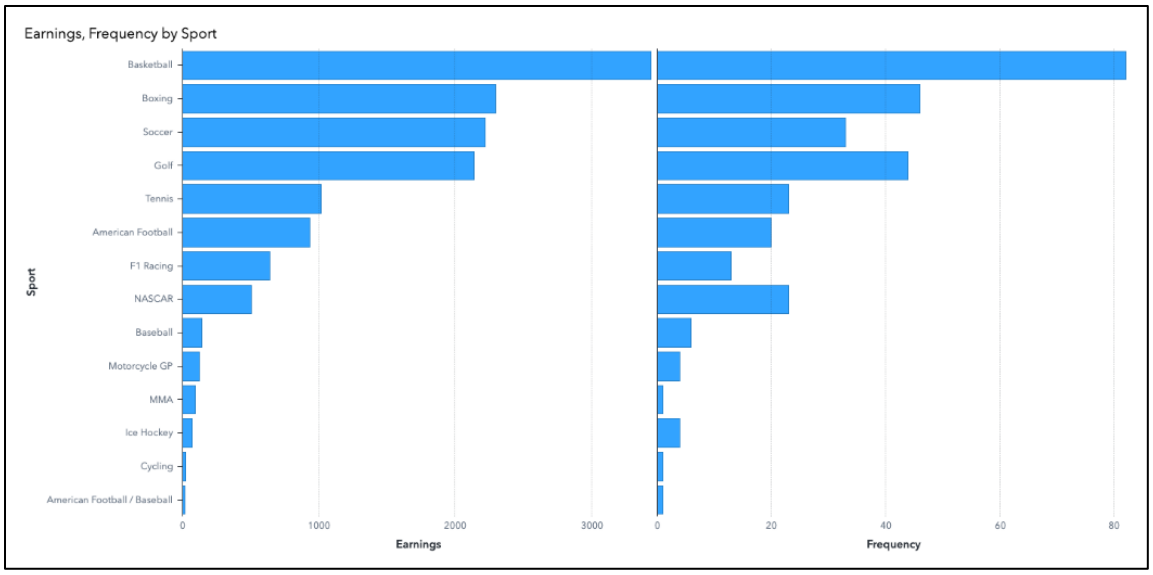
3. To find out if compensation for athletes varies depending on the sport, create a bar chart of **Earnings** by **Sport**. Our results:



4. At first glance, it appears that basketball players earn the most! To verify this, add **Frequency** to the bar chart under **Measures**, as such:

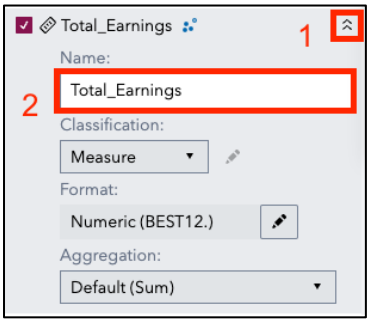


The output:

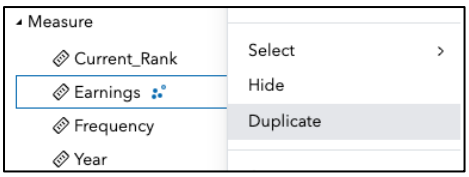


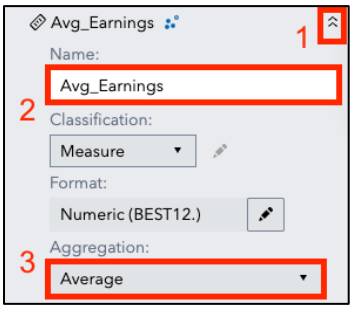
   a. Oh wait, basketball also has the highest frequency in our dataset. Turns out the default aggregation method for measure variables, like **Earnings**, is summation. This may not be good for our analysis because some sports have more players in the top 10. A better approach would be to look at average earnings by sport. We will rename the Earnings variable to **Total_Earnings** and create a new variable to display average earnings.

b. Change the name of variable **Earnings** to **Total_Earnings**.

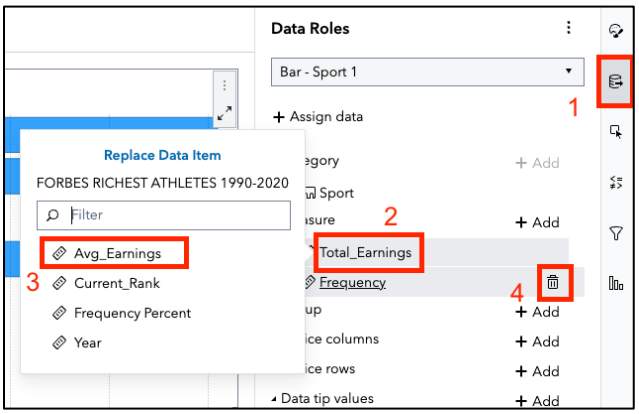


c. To create a new data item, duplicate the current variable and change the aggregation to **Average** instead of **Sum**. Name this variable **Avg_Earnings**. Some screenshots to help you along:

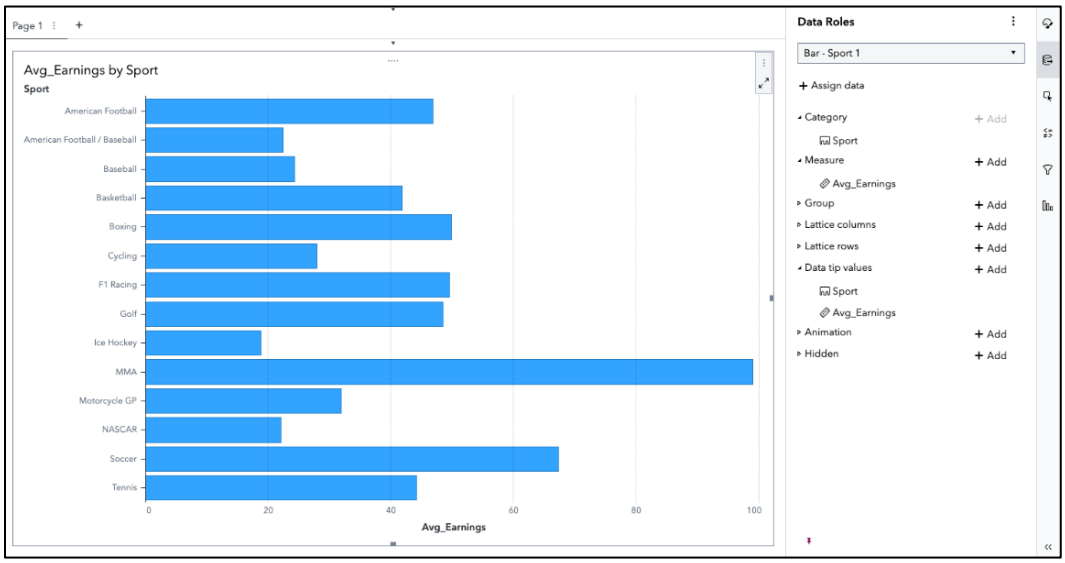




d. Using the Roles tab for the Bar Chart, replace **Total_Earnings** with **Avg_Earnings**, and remove **Frequency**. Like so:



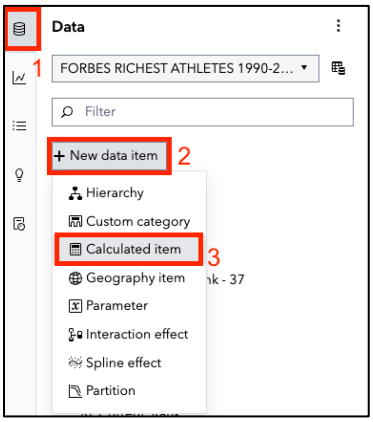
The output should look as follows:

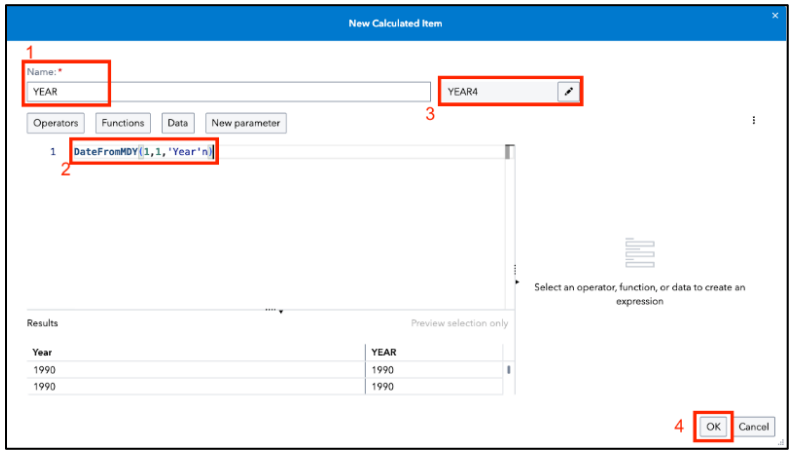


Once we account for the number of players, we see that basketball players' earnings are similar to those of tennis and football players. What stands out the most here is that MMA fighters make nearly double the average of other sports (more on this later).

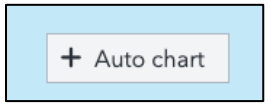
5. Another interesting analysis could be how sports compensation has changed over time, but oh no, our variable for year is a measure rather than a date value!!

    a. Let's create a new **YEAR** variable as a SAS date value using the expression **DateFromMDY(1, 1, 'Year'n)**. Even though we are storing the full date, notice that we are applying a **YEAR4** format to display the date value as a 4-digit year only. Some screenshot guidance:

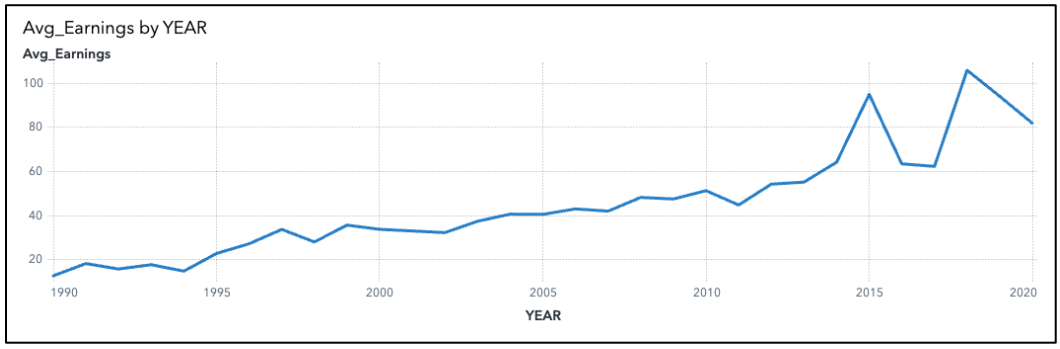




b. Now that we have our year variable as a date, we can plot **Avg_Earnings** by **Year**. An automatic chart can accomplish this perfectly – all you need to do is select both variables **Avg_Earnings** and **YEAR** and drag them onto the canvas below your bar chart! You will know you're doing it right when you see the **+ Auto chart** button below:



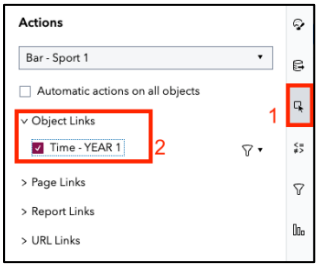
The output:



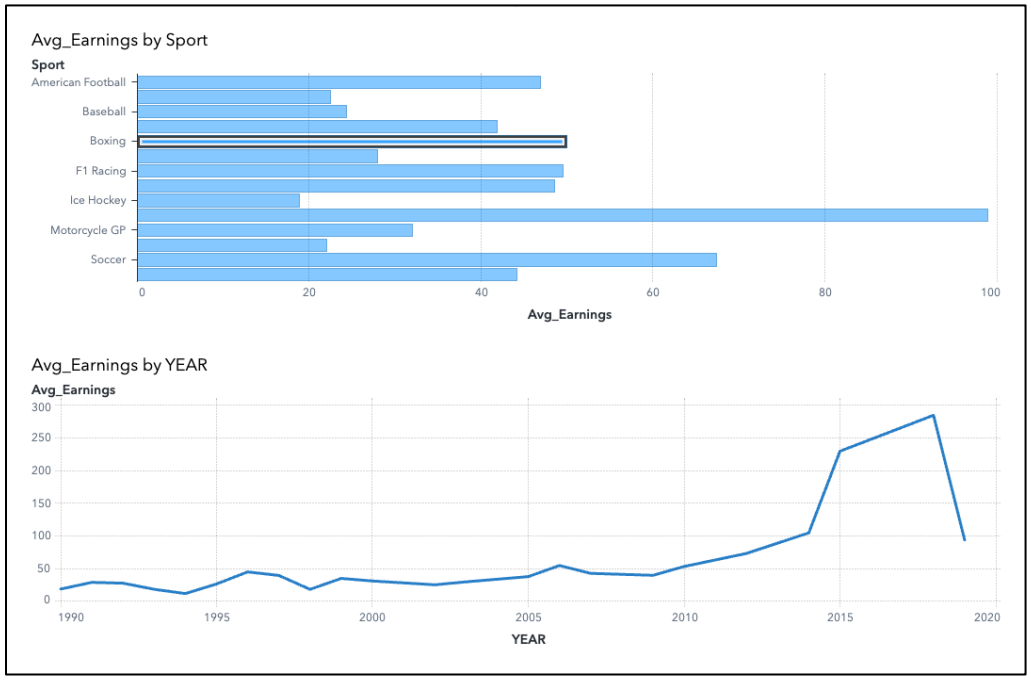
Overall, we see an upward trend in athletes' average earnings over the years. The sports industry is on the rise! We also notice some peaks and valleys starting in 2015 which might be due to some outliers in the dataset.

6. To see if this trend is true for all sports, we can add in some interactivity between the **Avg_Earnings** by **Sport** bar chart to the **Avg_Earnings** by **Year** time series plot. This will allow us to inspect changes in earnings over time for each sport.

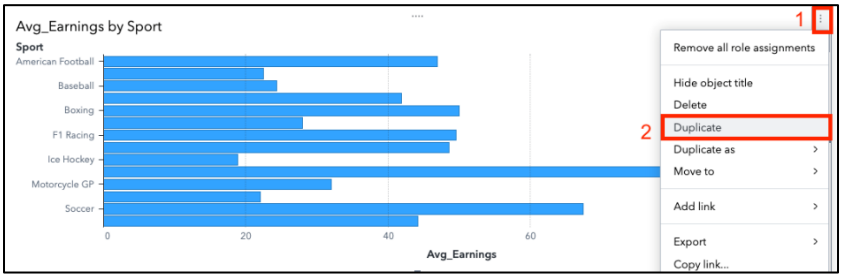
    a. With the bar chart selected, open the **Actions** pane from the right. Under the **Object Links** section, click the check box next to **Time – YEAR 1**. Some help:

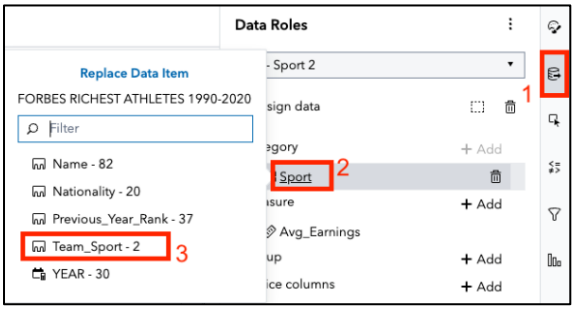


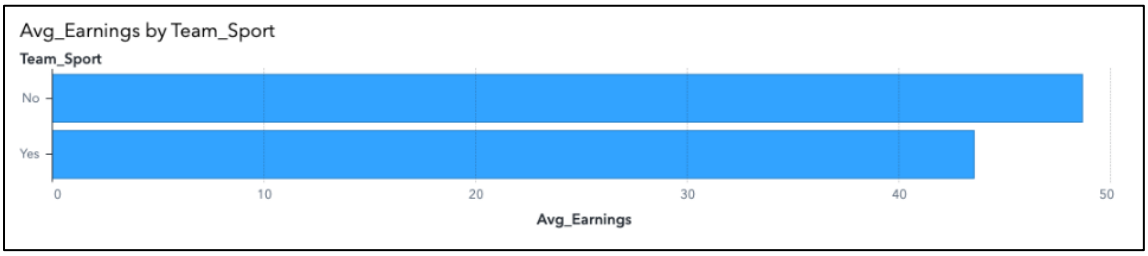
    b. Now, clicking on individual bars will show how that sport's players have been compensated over time. For example, professional boxers have certainly had increased earnings – and it looks like these individuals may be the reason for the overall spikes we see in 2015 and 2018! Our work:



7. Going back to our bar chart, seeing all the sports is helpful, but you know what would be even better for our analysis – seeing the sports grouped into whether they are a team sport or not! **Duplicate** the bar chart displaying **Avg_Earnings** by **Sport** and in your new bar chart, swap **Sport** with **Team Sport**. Some guardrails:

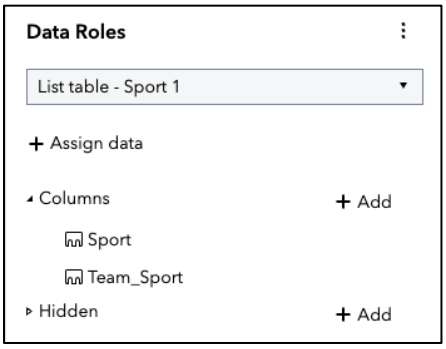






The bar chart shows that team sports have lower average earnings than individual sports. We might initially think that team sports tend to pay their players a consistent salary and therefore are more likely to have higher average earnings. However, keep in mind that we are working with datasets that capture the highest earners in each sport and there are a few outliers in the individual sport category driving the average earnings up.

Note: If you want to see which sports are counted as **Team Sports**, create a **List table** with **Sport** and **Team Sport**. For example:

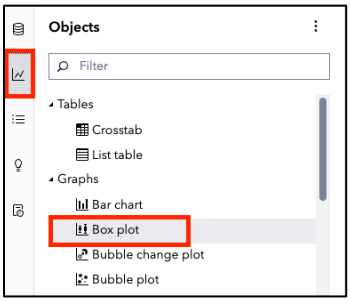


Sport	Team Sport
American Football	Yes
American Football / Baseball	Yes
Baseball	Yes
Basketball	Yes
Boxing	No
Cycling	Yes
F1 Racing	Yes
Golf	No
Ice Hockey	Yes
MMA	No
Motorcycle GP	Yes
NASCAR	Yes
Soccer	Yes
Tennis	No

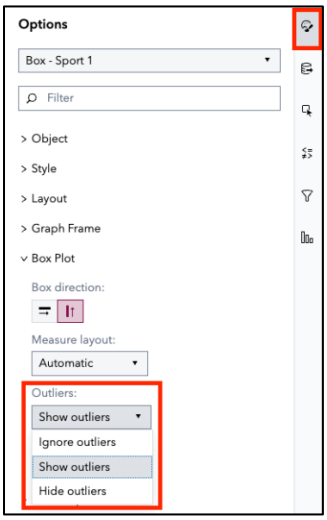
8. To better understand the distribution of earnings for each sport and if there are any outliers in our data, we can create box plots.

    a. This page is getting a little crowded, so we'll add a new page to the report.

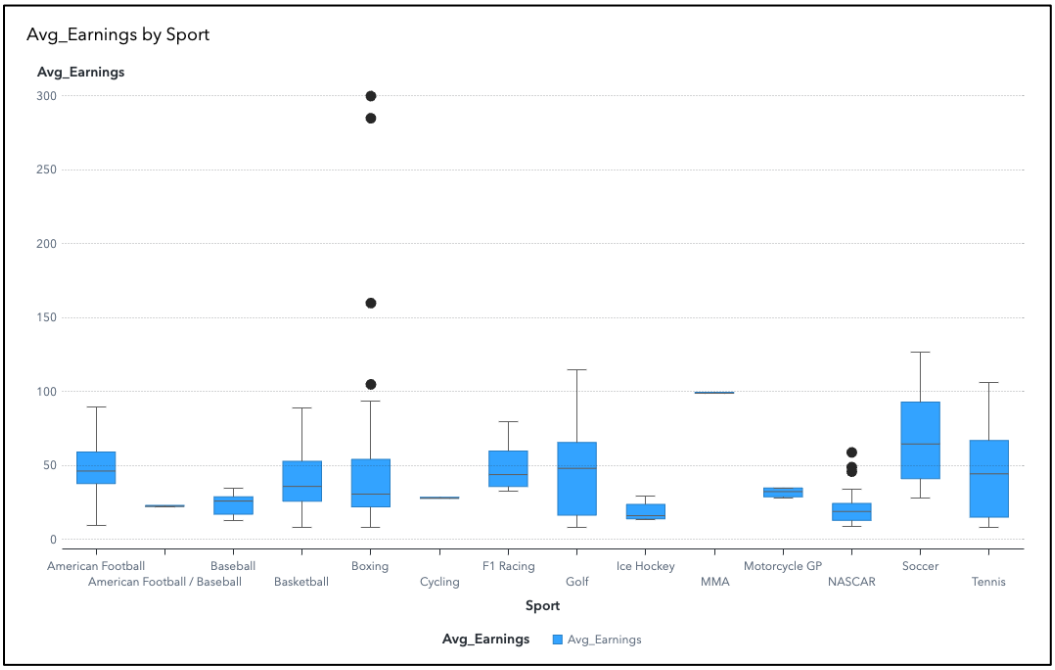
    b. Drag the **Box plot** object onto the new page and assign the following roles:
        - Category: **Sport**
        - Measure: **Avg_Earnings**



    c. To see if there are outliers, we will need to explicitly choose to show them. From the **Options** pane, find the **Box Plot** section. Change the **Outliers** option to **Show outliers**. Our work:

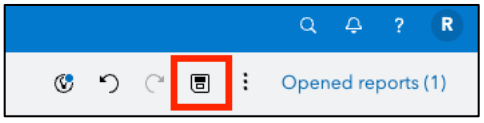


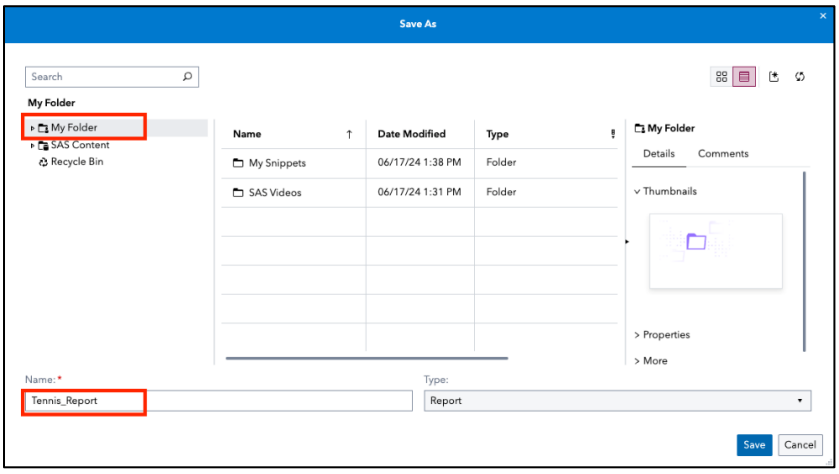
Your resulting box plot should look like the following:



Most sports seem to have a compact distribution for average earnings; however, boxing and NASCAR have outliers present in the data, with points far above the rest. These outliers might also help explain the peaks and valleys we observed earlier in our Avg_Earning by Year line chart.

9. Now is a great time to save your report. Give it a memorable name, like *Tennis_Report*, and save it in **My Folder**.



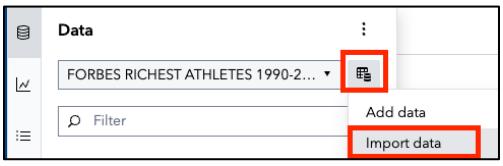


Don't forget to save along the way!

Part 2: The Structure of a Player's Compensation

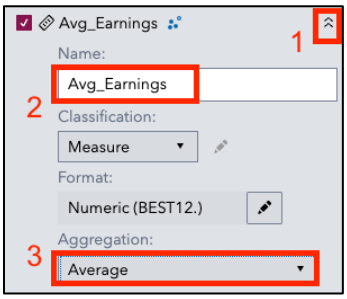
Since we saw a difference in the earnings of individual and team sports athletes, it's worthwhile digging deeper to understand the compensation structure. For this, we'll introduce a new dataset!

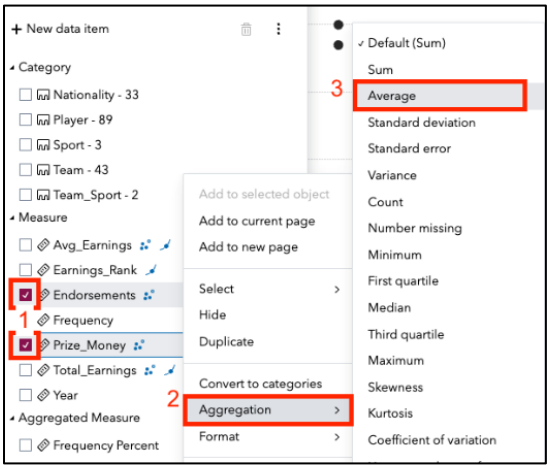
1. Import **Athletes_Earnings.xlsx** dataset and add it to your report. Check out the screenshots below for adding another dataset to your report. For help with the import itself, refer to Part 1.



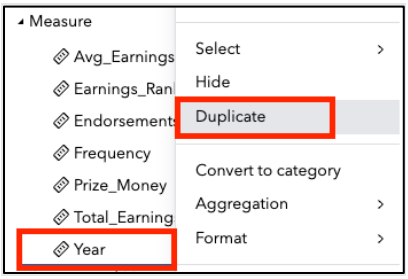
   Notice that along with **Total_Earnings**, we also have **Endorsements** and **Prize_Money** in our data; however, we don't have average earnings.

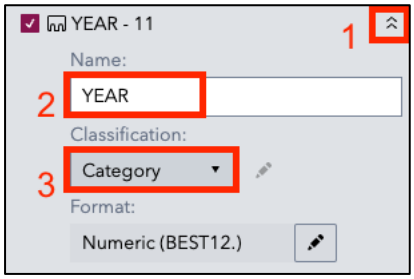
2. Recreate the **Avg_Earnings** variable in this dataset by duplicating **Total_Earnings** and changing the name and aggregation. While we are at it, change the aggregation of **Endorsements** and **Prize_Money** to Average as well! Like so:



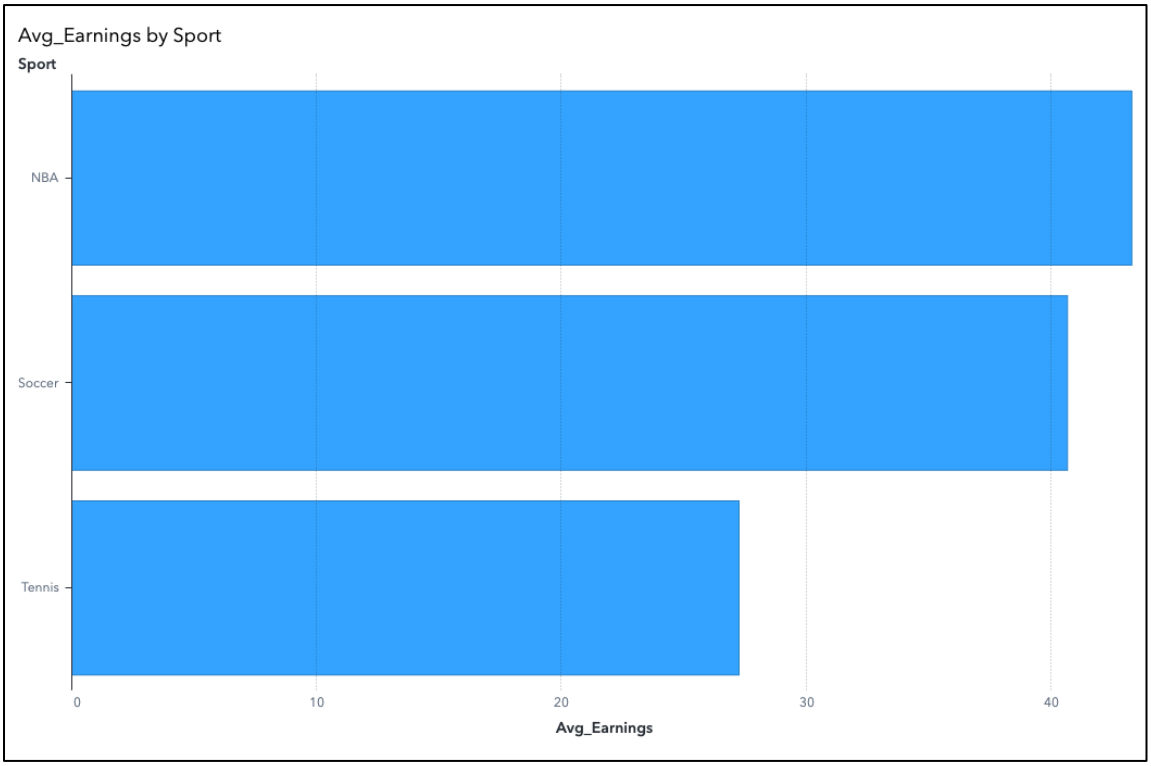


3. As a bit of data prep, we should duplicate the **Year** variable and change it to a **Category**. Some guidance:





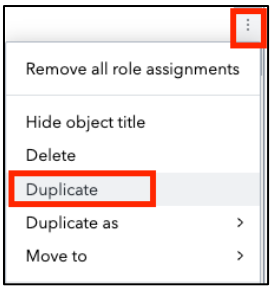
4. Now onto the analysis! On a **New page**, create a **Bar Chart** to compare **Avg_Earnings** by **Sport** using the **Athlete Earnings** dataset. You should see the following:

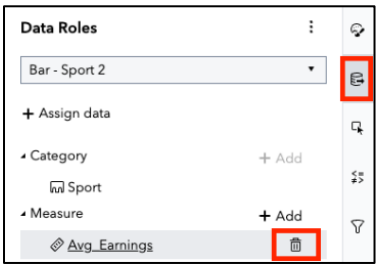


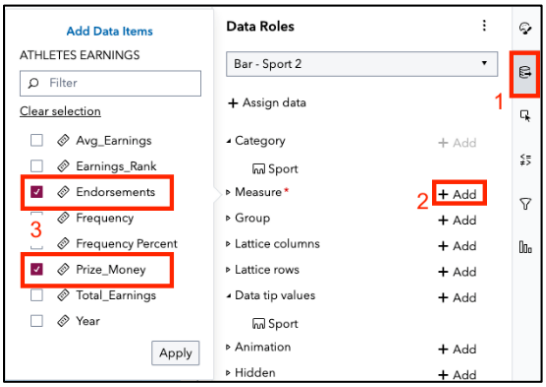
From this dataset, we see that tennis players make less on average than basketball or soccer players. This contradicts our findings from Part 1, where we observed that tennis and basketball players have similar average earnings. However, keep in mind that the two datasets cover different time periods and that the sports industry landscape has changed considerably over time.

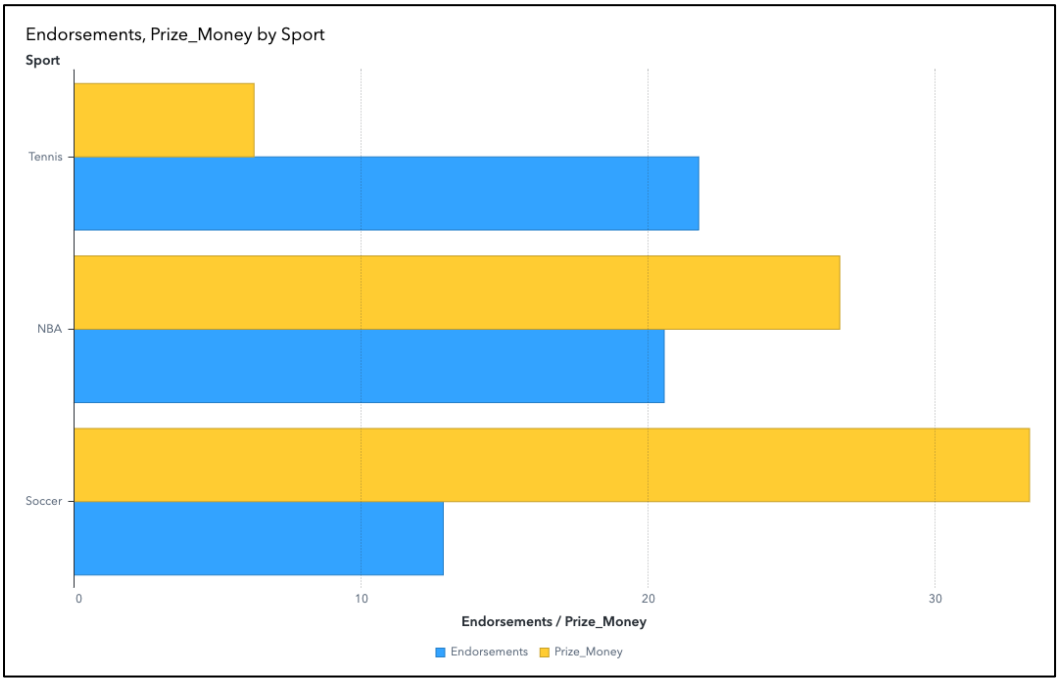
5. To get a better sense for the components of a player's compensation, we'll break down earnings into prize money and endorsements. Prize money is going to capture the players' salary or winnings from sporting events (i.e., on-the-field earnings). Endorsements on the other hand, will give us a sense of the monetary compensation athletes receive for brand deals, campaigns, licenses, and social events (i.e., off-the-field earnings).

   Duplicate the **Bar chart** and swap Avg_Earnings for **Endorsements** and **Prize_Money**. Some guardrails:









Notice that both basketball and soccer players earn more in prize money (i.e. yearly contract pay) than in endorsements. However, for tennis players this is reversed. One reason for this might be that tennis players are not paid a consistent salary and even top-ranked players only bring home large winnings at high-profile tournaments, which are rare. To be able to sustain their expenses (travel, coaches, trainers, etc.), they must supplement their income with brand deals, sponsorships, and event appearances.

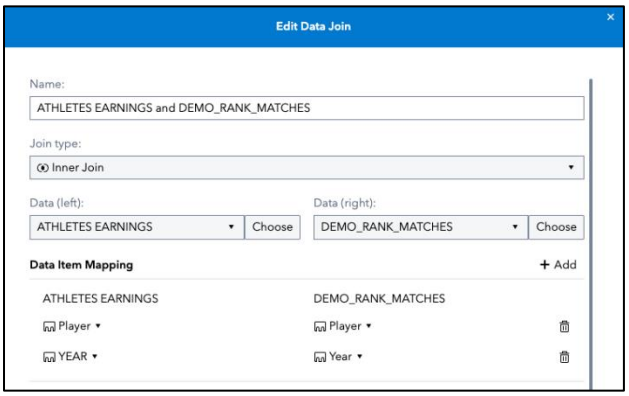
6. Although this immediate realization is helpful, it'd be nice to have a few more statistics to better understand nuances of sports compensation. Good news – we have another dataset that can provide these further details!

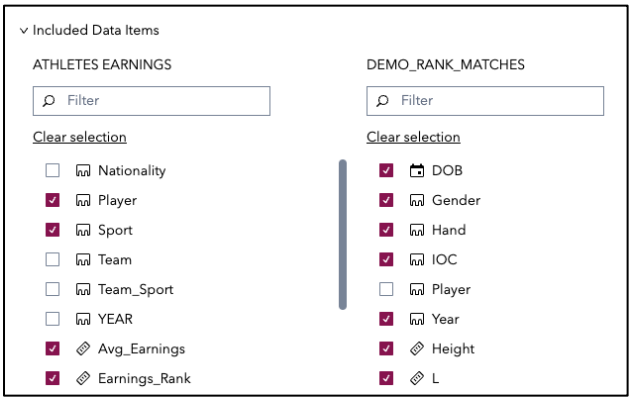
   Create a new dataset from an **inner join** between **Athlete Earnings** and **Demo_Rank_Matches**.

Some specifics for making this join successful:

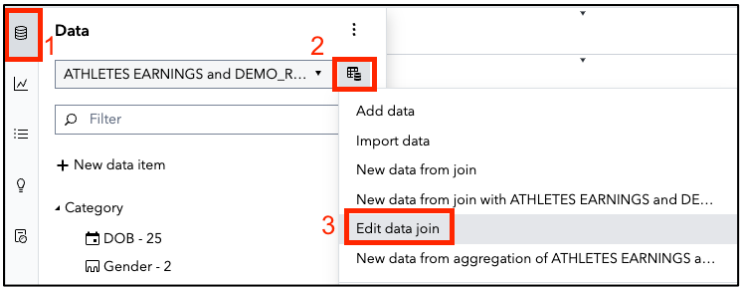
a. Import **Demo_Rank_Matches** into your VA report in the usual way.
b. Change the following settings:
    - Dataset mapping: **Player** and **Year** (category)
    - Variables to exclude from Athlete Earnings: **Team, Team_Sport**, **YEAR** and **Nationality**.

Some screenshots to help:





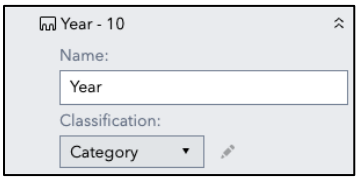
Note: You can make edits to your data join from the **Data** pane:



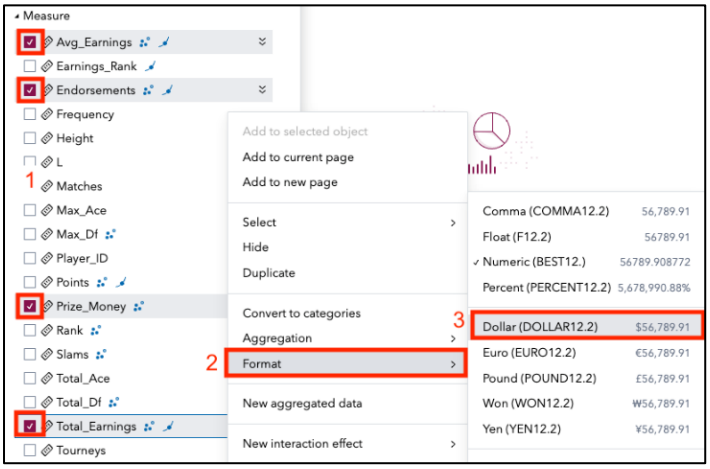
7. Let's check in on our variables:

    a. Are the aggregations correct? You need to change the **Aggregation** for **Avg_Earnings** and **Win_Per**, plus any others you think necessary!

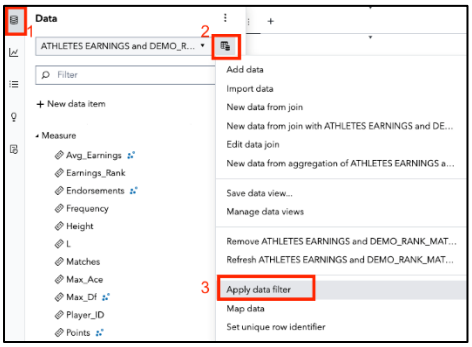
    b. Change the name of the **Year (DEMO_RANK_MATCHES)** variable to remove the dataset reference.

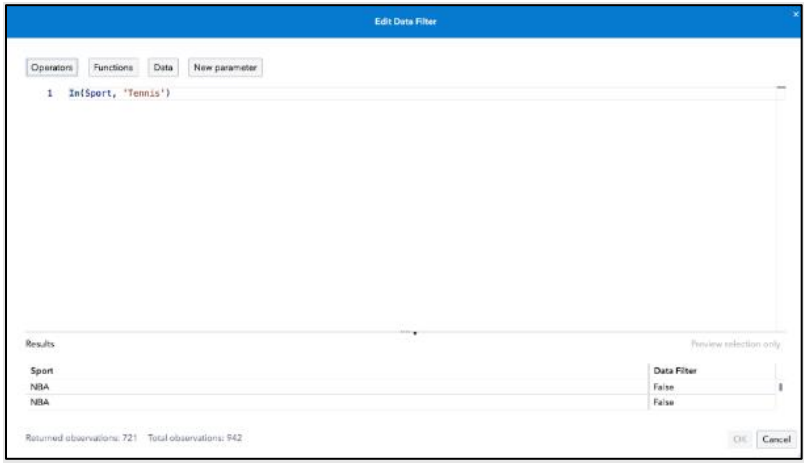


    c. To help with interpreting monetary values, change the format for all the earnings variables to **DOLLAR12.2** using the steps below:



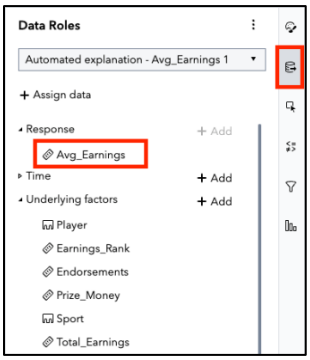
8. With our data prepped and ready, it's time to dive into our analysis of tennis earnings. **Apply a data filter** to only look at observations where **Tennis** is the **Sport**. Note: there are multiple ways of doing this – and here is one of them:





9. At this point, we don't know what we don't know – so let's give Visual Analytics some freedom to analyze the **Avg_Earnings** variable in an **Automated Explanation**. This object reveals the most important underlying features for a target variable and provides additional details, such as minimum/maximum, outliers, and specifics about the relationship between each of the other variables in a dataset.

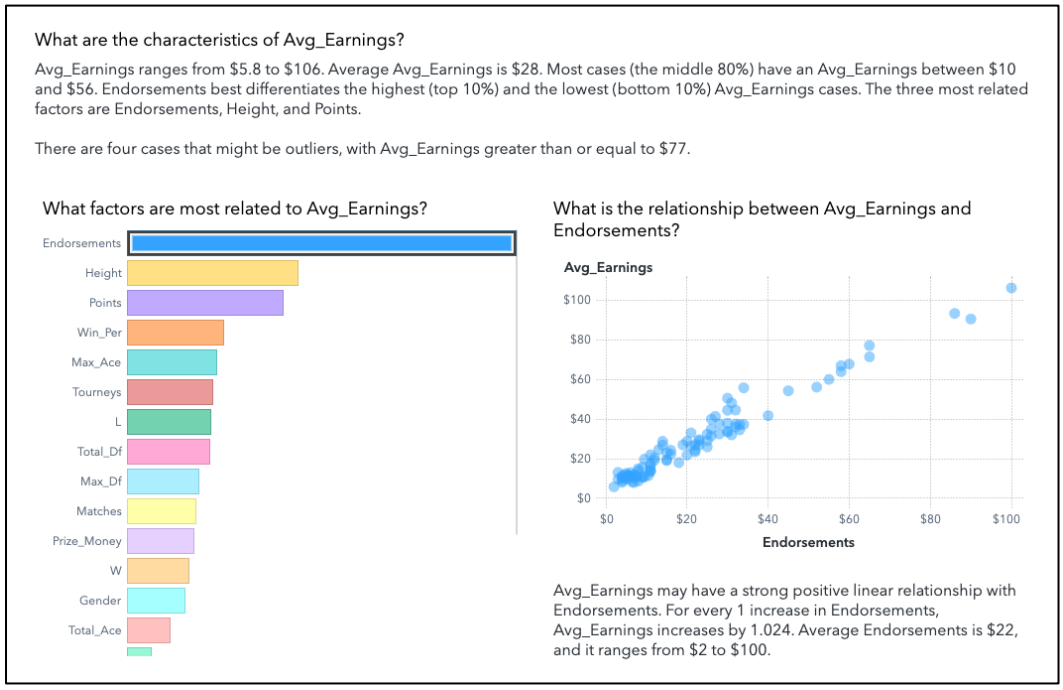
    a. From the **Objects** pane, add **Automated Explanation** to a **New page**. Add **Avg_Earnings** as the **Response** and the rest of the variables will be added as underlying factors by default.



    b. Some variables added by default aren't helpful though. For instance, the **Earnings_Rank** variable was created based on how much a player earns; thus, it is redundant and does not give us any additional information about **Avg_Earnings**. **Player_ID** also does not provide any helpful information since this is used as an identifier.

    Remove the following variables from **Underlying factors**:
    - **Earnings_Rank**
    - **Player_ID**
    - **Player**
    - **IOC**

c. An Automated Explanation can be a great starting point for exploration. Click around and see how the explanations and graphics change.



There are some obvious findings here:

- Endorsements is the factor most related to earnings.
- Prize money, not so much! Remember that most of a tennis player's earnings are obtained through endorsements.

But there are also some surprising results:

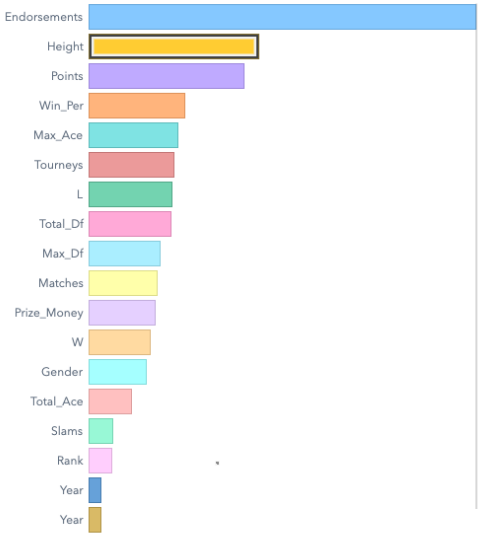
- Height is the second most important factor related to earnings; however, it is a weak relationship.
- Tennis rank is one of the least important factors related to earnings! In this sport, we often see players rise up the ranks and garner attention as they win big tournaments as underdogs and score lucrative sponsorships in the process. Do you see this too?

What are the characteristics of Avg_Earnings?

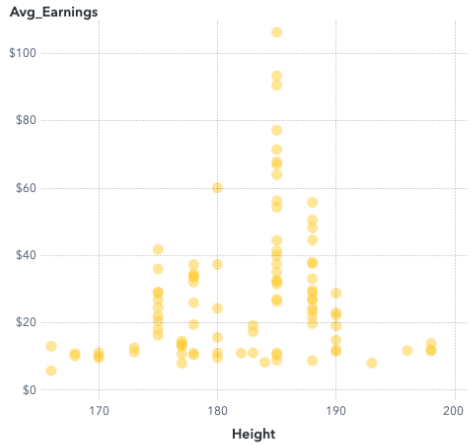
Avg_Earnings ranges from $5.8 to $106. Average Avg_Earnings is $28. Most cases (the middle 80%) have an Avg_Earnings between $10 and $56. Endorsements best differentiates the highest (top 10%) and the lowest (bottom 10%) Avg_Earnings cases. The three most related factors are Endorsements, Height, and Points.

There are four cases that might be outliers, with Avg_Earnings greater than or equal to $77.

What factors are most related to Avg_Earnings?



What is the relationship between Avg_Earnings and Height?



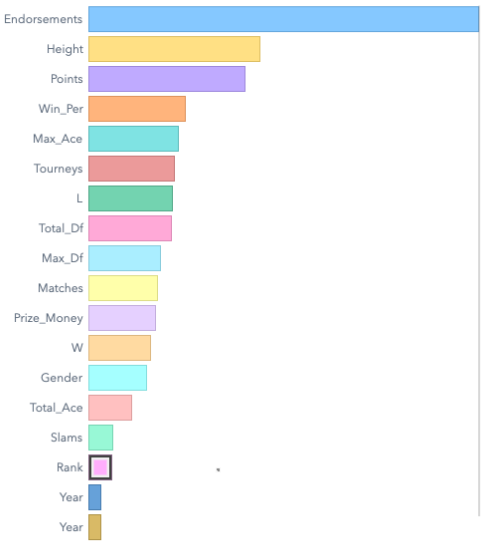
Avg_Earnings may have a weak relationship with Height. Average Height is 182, and it ranges from 166 to 198.

What are the characteristics of Avg_Earnings?

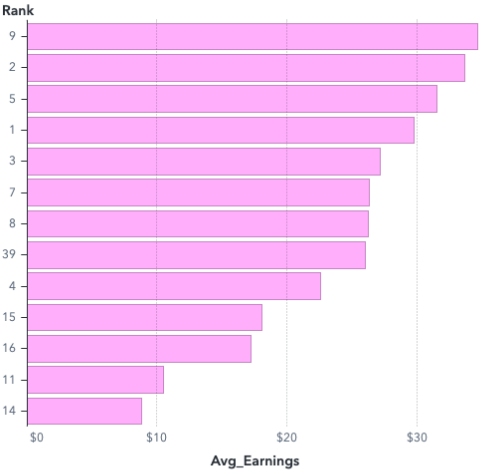
Avg_Earnings ranges from $5.8 to $106. Average Avg_Earnings is $28. Most cases (the middle 80%) have an Avg_Earnings between $10 and $56. Endorsements best differentiates the highest (top 10%) and the lowest (bottom 10%) Avg_Earnings cases. The three most related factors are Endorsements, Height, and Points.

There are four cases that might be outliers, with Avg_Earnings greater than or equal to $77.

What factors are most related to Avg_Earnings?



What is the relationship between Avg_Earnings and Rank?



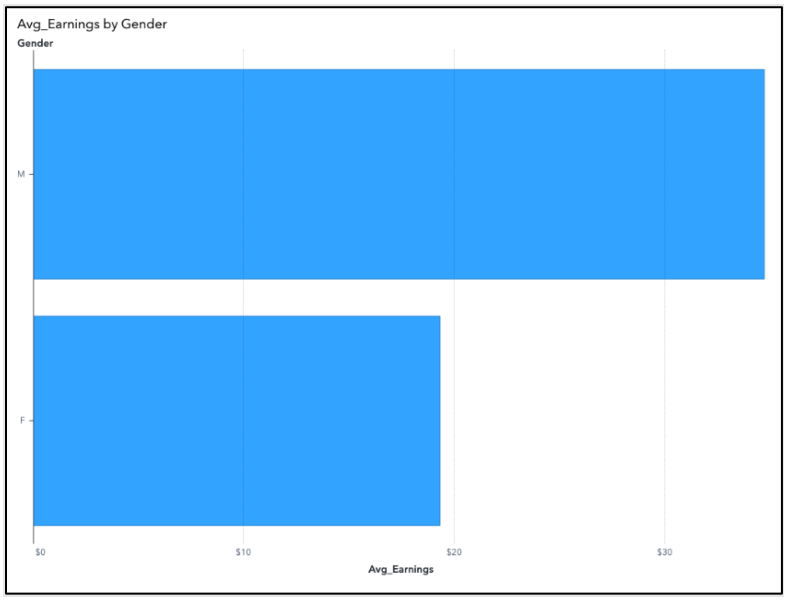
When Rank is 9, 2, 5, 1 or 3, the average of Avg_Earnings is a high value. When Rank is 11 or 14, the average of Avg_Earnings is a low value. The most common Rank value is 1.

Part 3: Gender Differences

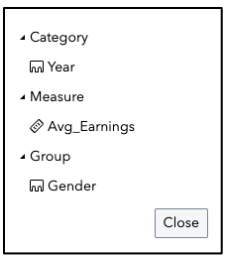
One of the biggest topics in sports related to earnings is the gender pay gap (source: https://en.wikipedia.org/wiki/Gender_pay_gap_in_sports). While tennis has made notable progress towards pay equity by awarding equal prize money at its highest-profile events (Grand Slam tournaments), the disparity persists for smaller tournaments. Let's investigate this further!

1. On a new page, create a **Bar Chart** and assign the following roles:
   - Measure: **Avg_Earnings**
   - Category: **Gender**

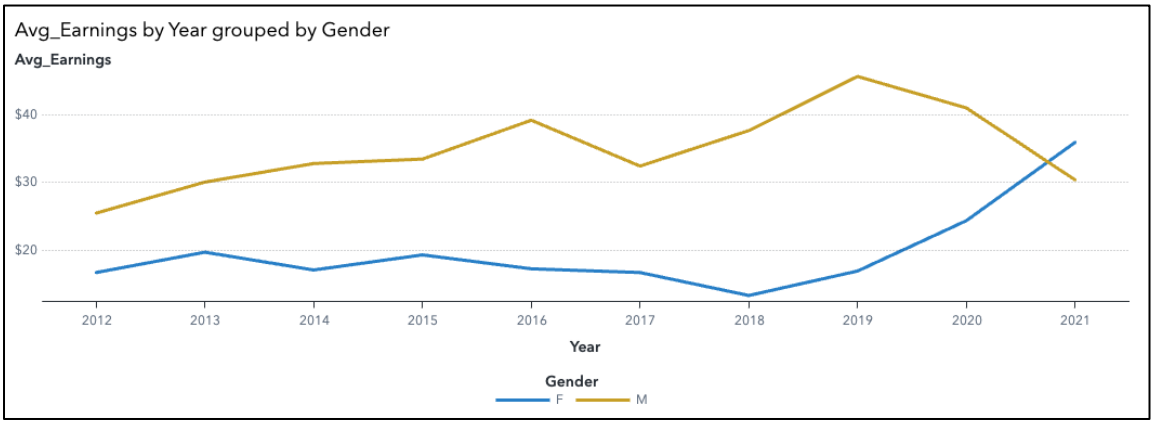
   It should appear as:



2. Hey, what gives? I thought we were closing the gap! Let's build a **Line Chart** to see what patterns we notice over time. Assign the following roles for the line chart:
   - Category: **Year**
   - Measure: **Avg _Earnings**
   - Group: **Gender**



The output:

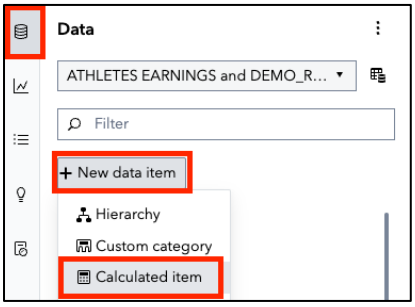


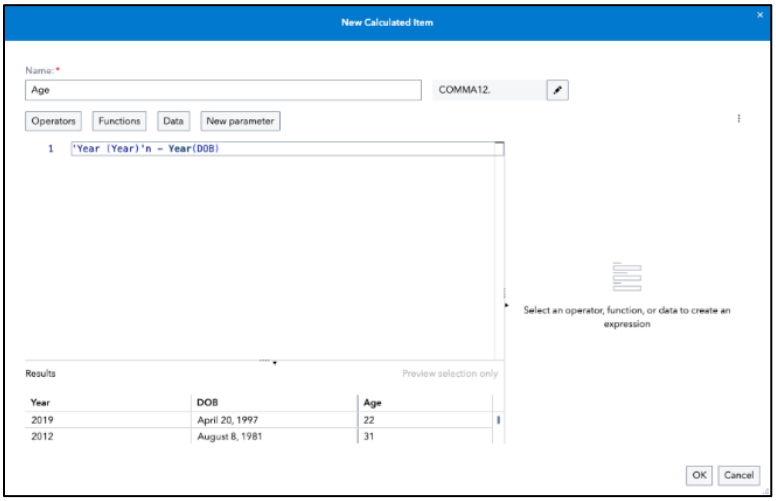
The proof's in the pudding! When we look at average earnings over the entire decade, it seems that male tennis players make on average almost twice as much as female players. However, the progression throughout the years shows that the gender pay gap is indeed closing.

Since 2007, the four Grand Slam tournaments (Australian Open, French Open, Wimbledon, US Open) have awarded equal prize money to both men and women, with many 1000-level and 500-level tournaments following suit. This has in turn given more recognition and increased publicity for female tennis players, leading to more lucrative endorsements which drive overall earnings.

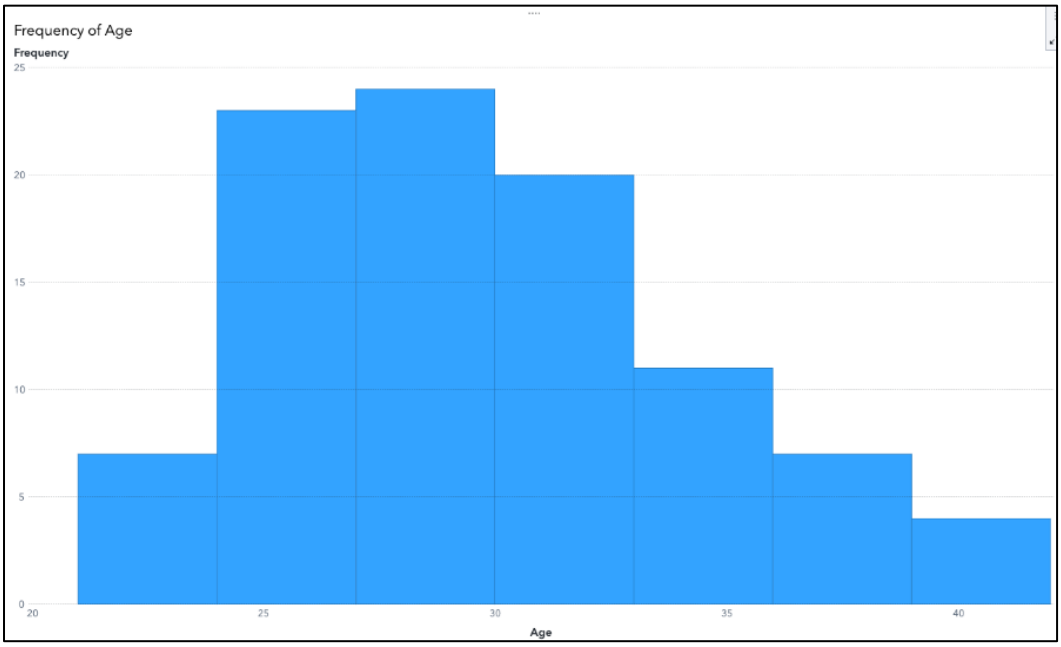
3. Since this data does show statistics over a decade, there may be some importance behind when players are entering/exiting the game based on their age.

    a. Since age is not already in our dataset, let's create it using the expression **'Year (Year)'n – Year(DOB)**. We want age to be displayed in years, so we'll assign a format that shows no decimals. It'll resemble the following:





b. Now let's see what information we can glean from Age! Drag the **Age** variable (without any others selected) onto your canvas to see a histogram of the values.

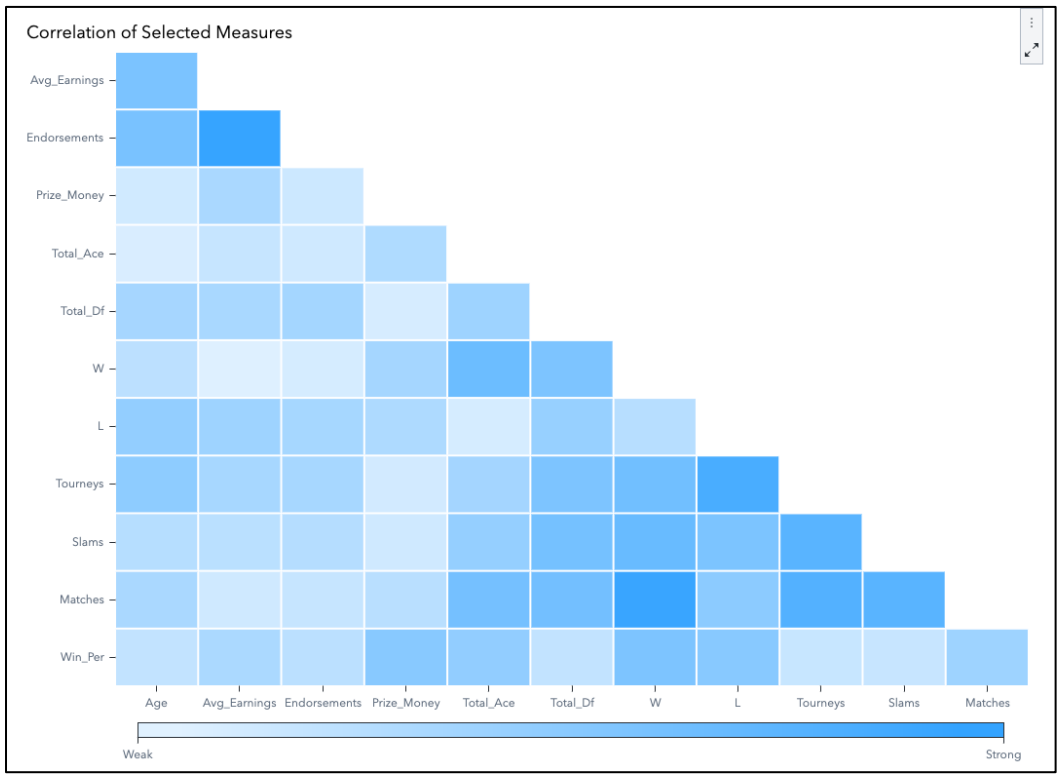


Looking at the distribution of **Age,** we see the most players are in their mid to late 20's, but some players are as young as 21 and as old as 42! Keep in mind that this dataset includes the top 10 highest paid players of 2012-2021.

10. Finally, to identify which tennis metrics lead to monetary success we will introduce a **Correlation analysis**.

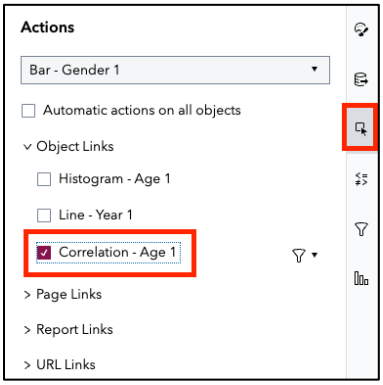
    a. Select the following measures, and drag them onto your canvas:
    - **Age**
    - **Avg_Earning**

- **Endorsements**
- **Prize_Money**
- **Total_Ace**
- **Total_Df**
- **W**
- **L**
- **Tourneys**
- **Slams**
- **Matches**
- **Win_Per**

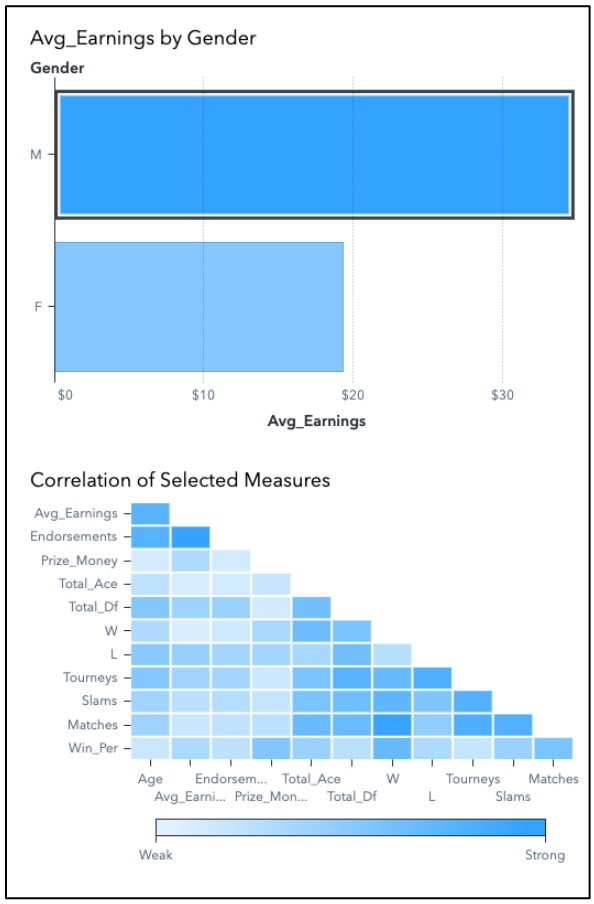


Age seems to have the highest correlation with **Avg_Earnings** and **Endorsements**, but **Prize_Money** is most correlated with win percentage (**Win_Per**). Makes sense! It's also no surprise that we see a high correlation between wins (**W**) and **Matches**, …and **Tourneys**! The more tournaments you attend as a player, the more matches you play and the greater your chances of winning. You can't win if you don't play 😊. However, winning isn't everything in the sport of tennis. Wins is not highly correlated with **Avg_Earnings**, because it isn't highly correlated with **Endorsements**.

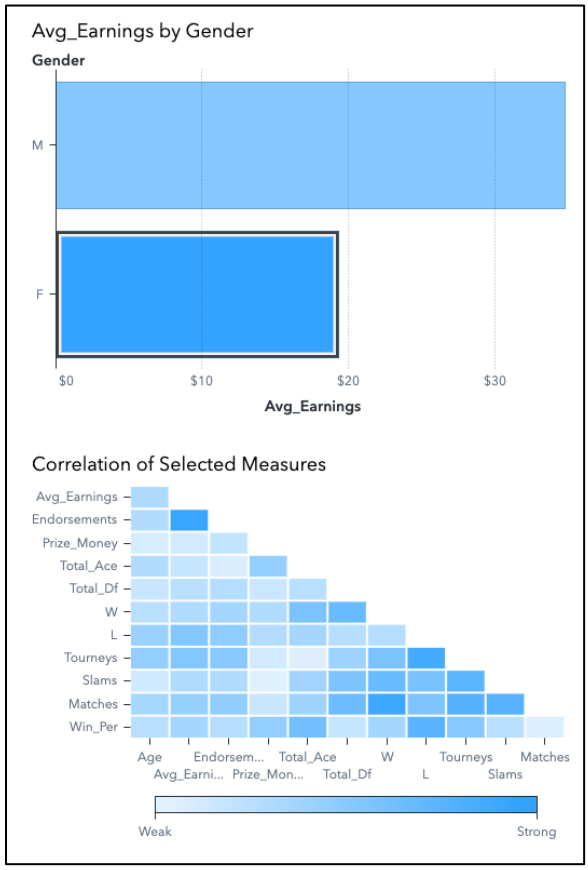
11. You may be interested to know if any of these correlations are gender-specific. We can easily add an **Action** within Visual Analytics so that the correlation matrix can be filtered based on gender. To do this:
    a. Select the **Gender** bar chart to make it active.
    b. Select the **Actions** pane on the right sidebar.
    c. Under **Object Links**, click the check box for the **Correlation matrix**, as such:



    d. Click the M bar in the bar chart to see the correlation matrix update for males:



e. Click the F bar in the bar chart to see the correlation matrix update for females:

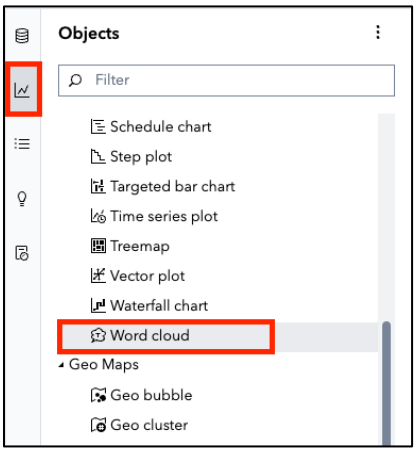


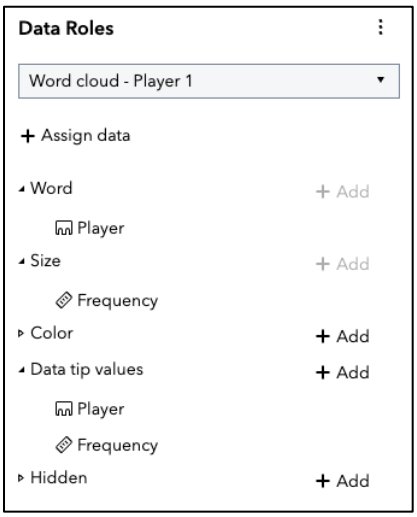
It seems as though the trends we observed above hold true for each individual gender as well.

Part 4: Individual Player Analysis

We've done some great work looking at an overall analysis for the sport of tennis, but there are some standout players that have changed the tennis game. Who are these players? Let's find out!

1. On a **New page**, let's add a **Word Cloud** with **Player** as the Word variable to see which names come up the most as top earners.



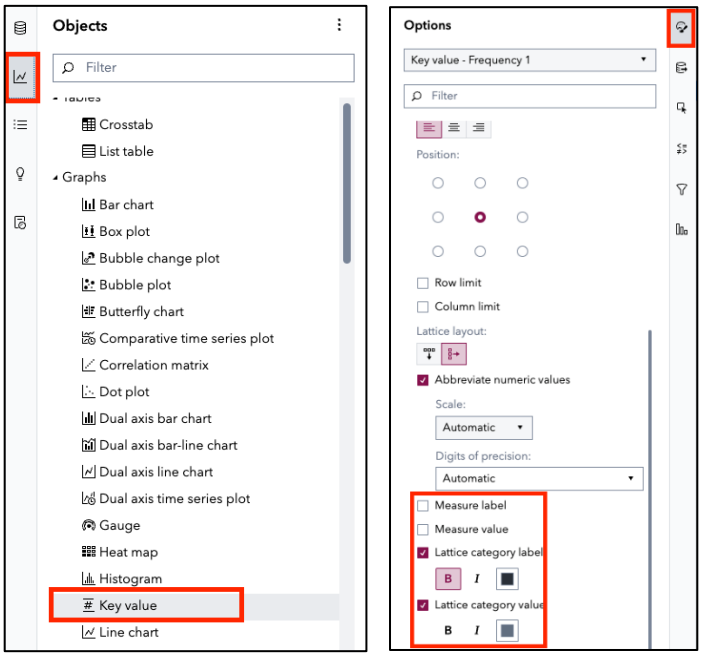


Your output should appear as:

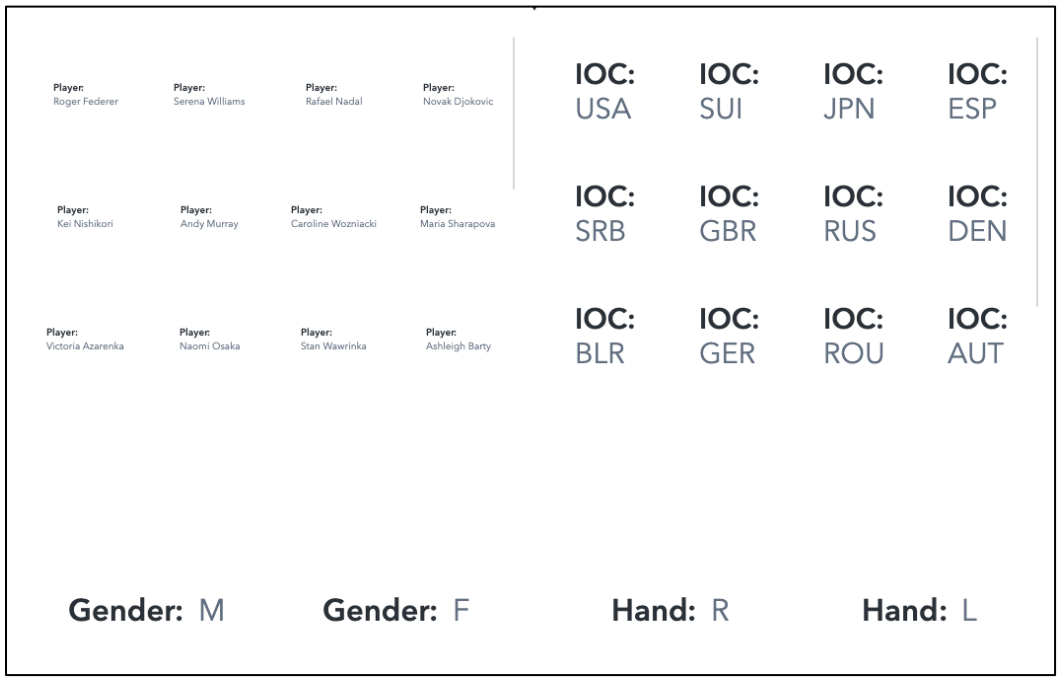
2. We can use a **Hidden page** to capture individual player statistics. On a **New page**, we'll create **Key Value** objects.

    a. Use the following components as the Lattice category in each of the Key Value objects:
    - **Player**
    - **IOC**
    - **Gender**
    - **Hand**

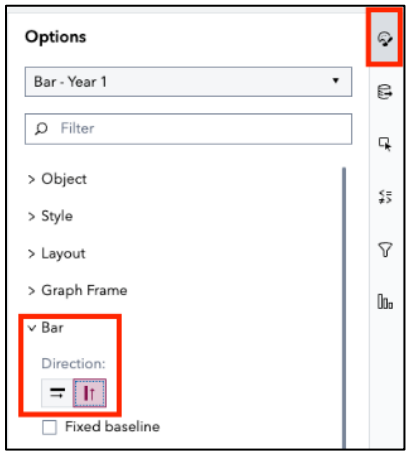
    Note that we care about the category rather than the measure, so use **Frequency** for simplicity. Each key value should only show the Lattice category label and Lattice category value. Some guidance:

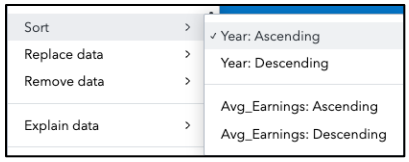


The top of your page should look similar to this:

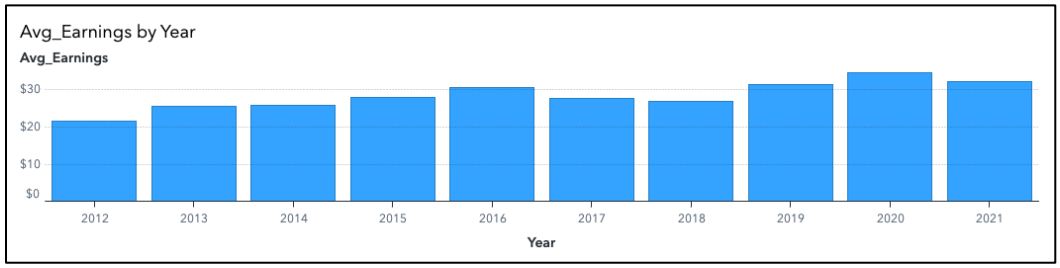


b. Below your key values, create a **Bar Chart** showing **Avg_Earnings** by **Year**. The default for bar charts is to display horizontal bars and to sort in descending order by Avg_Earnings. But how helpful is that (hint: not very!!). Change the **Direction** to **Vertical bars** and right-click the bar chart to sort by **Year: Ascending**.





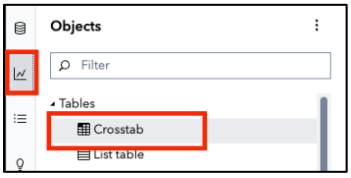
Your updated bar chart should look like this:

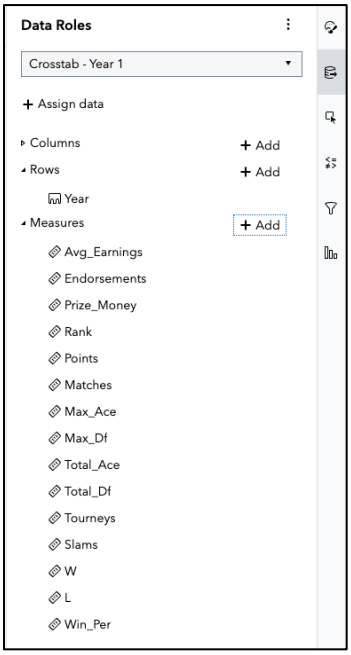


c. All these quick views and graphics are great – but sometimes we want the nitty gritty details! To satisfy the in-depth data-lovers, create a crosstab next to your bar chart showing yearly numbers for:

- **Avg_Earnings**
- **Endorsements**
- **Prize Money**
- **Rank**
- **Points**
- **Matches**
- **Max_Ace**
- **Max_Df**

- **Total_Ace**
- **Total_Df**
- **Tourneys**
- **Slams**
- **W**
- **L**
- **Win_Per**

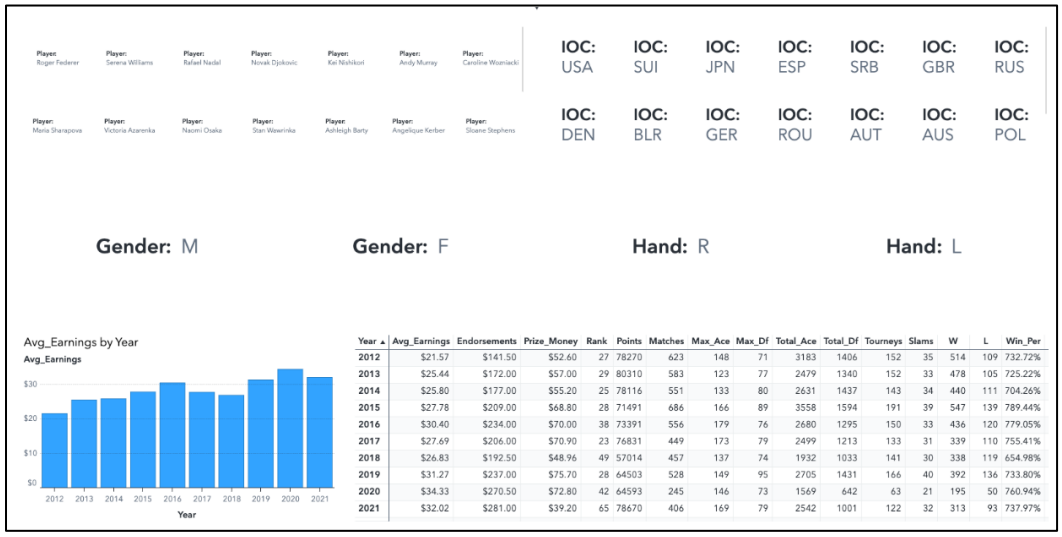




The output:

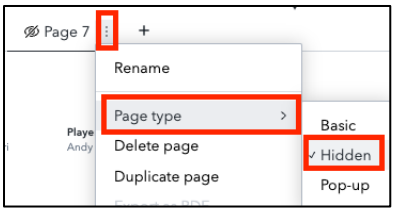
Year ▲	Avg_Earnings	Endorsements	Prize_Money	Rank	Points	Matches	Max_Ace	Max_Df	Total_Ace	Total_Df	Tourneys	Slams	W	L	Win_Per
2012	$21.57	$141.50	$52.60	27	78270	623	148	71	3183	1406	152	35	514	109	732.72%
2013	$25.44	$172.00	$57.00	29	80310	583	123	77	2479	1340	152	33	478	105	725.22%
2014	$25.80	$177.00	$55.20	25	78116	551	133	80	2631	1437	143	34	440	111	704.26%
2015	$27.78	$209.00	$68.80	28	71491	686	166	89	3558	1594	191	39	547	139	789.44%
2016	$30.40	$234.00	$70.00	38	73391	556	179	76	2680	1295	150	33	436	120	779.05%
2017	$27.69	$206.00	$70.90	23	76831	449	173	79	2499	1213	133	31	339	110	755.41%
2018	$26.83	$192.50	$48.96	49	57014	457	137	74	1932	1033	141	30	338	119	654.98%
2019	$31.27	$237.00	$75.70	28	64503	528	149	95	2705	1431	166	40	392	136	733.80%
2020	$34.33	$270.50	$72.80	42	64593	245	146	73	1569	642	63	21	195	50	760.94%
2021	$32.02	$281.00	$39.20	65	78670	406	169	79	2542	1001	122	32	313	93	737.97%

And you full page should look something like this:

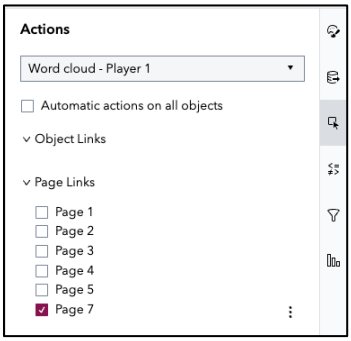


Year	Avg_Earnings	Endorsements	Prize_Money	Rank	Points	Matches	Max_Ace	Max_Df	Total_Ace	Total_Df	Tourneys	Slams	W	L	Win_Per
2012	$21.57	$141.50	$52.60	27	78270	623	148	71	3183	1406	152	35	514	109	732.72%
2013	$25.44	$172.00	$57.00	29	80310	583	123	77	2479	1340	152	33	478	105	725.22%
2014	$25.80	$177.00	$55.20	25	78116	551	133	80	2631	1437	143	34	440	111	704.26%
2015	$27.78	$209.00	$68.80	28	71491	686	166	89	3558	1594	191	39	547	139	789.44%
2016	$30.40	$234.00	$70.00	38	73391	556	179	76	2680	1295	150	33	436	120	779.05%
2017	$27.69	$206.00	$70.90	23	76831	449	173	79	2499	1213	133	31	339	110	755.41%
2018	$26.83	$192.50	$48.96	49	57014	457	137	74	1932	1033	141	30	338	119	654.98%
2019	$31.27	$237.00	$75.70	28	64503	528	149	95	2705	1431	166	40	392	136	733.80%
2020	$34.33	$270.50	$72.80	42	64593	245	146	73	1569	642	63	21	195	50	760.94%
2021	$32.02	$281.00	$39.20	65	78670	406	169	79	2542	1001	122	32	313	93	737.97%

d. Yes, it's messy right now, which is why we'll make this a **Hidden** page so that it only appears when we have a specific player selected. As such:



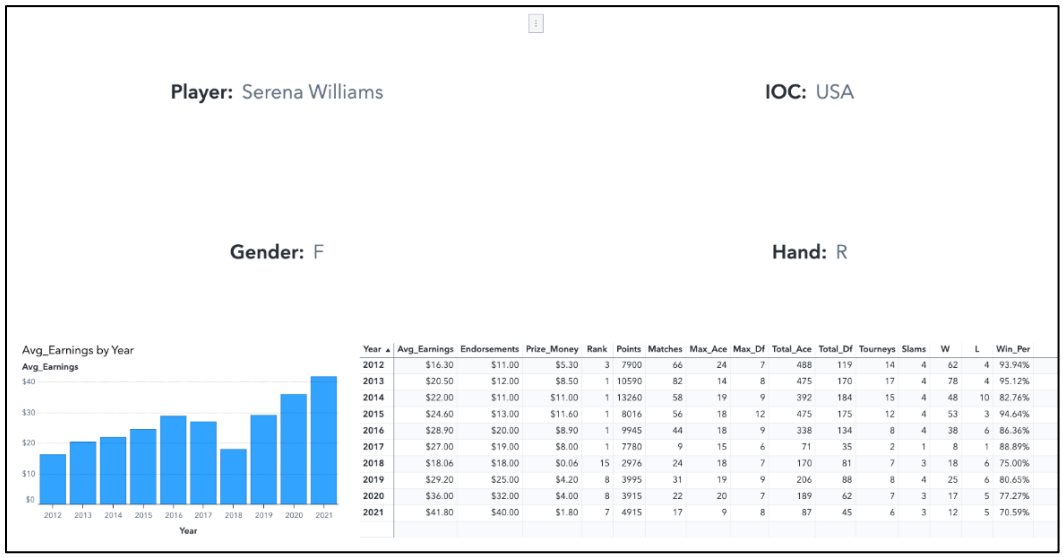
3. To make the hidden page appear, we will add an **Action** on the **Word Cloud** to reveal this page with specific details when a player is selected.



4. Time to test it out! Go to **View mode** (that little pencil in the upper corner) and see your hidden page disappear. In your word cloud, double-click the following players to make your hidden page appear. We'll walk through a few examples:

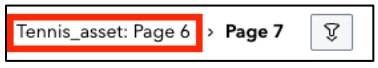
   a. Serena Williams

   Serena Williams, one of the greatest female tennis players of all time, has for many years maintained the highest total career prize money earnings in the WTA. Throughout her time as a dominating player, she gained a lot of publicity and endorsements to boot. Her retirement in 2022 had tennis fans in tears but her legacy remains.



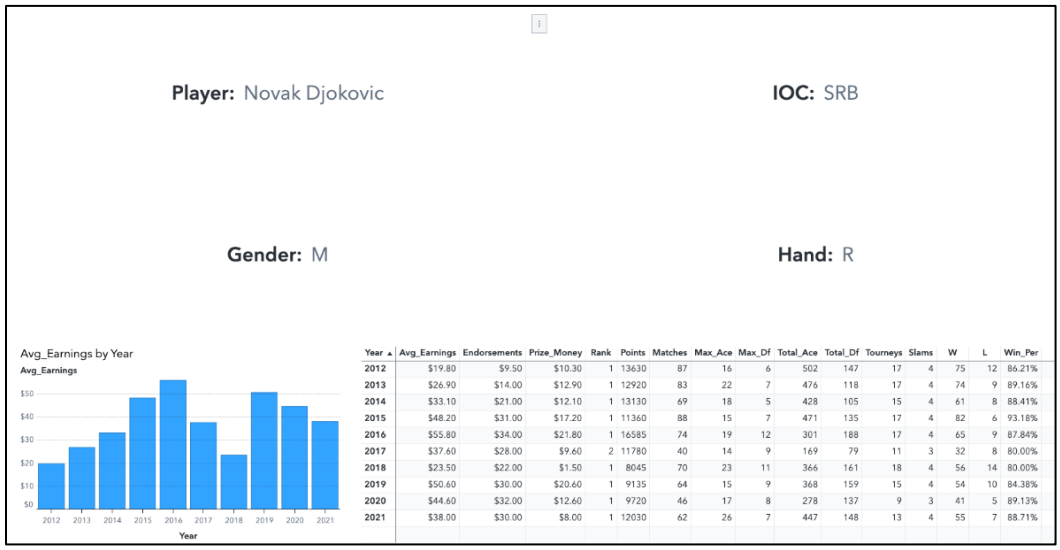
Year	Avg_Earnings	Endorsements	Prize_Money	Rank	Points	Matches	Max_Ace	Max_Df	Total_Ace	Total_Df	Tourneys	Slams	W	L	Win_Per
2012	$16.30	$11.00	$5.30	3	7900	66	24	7	488	119	14	4	62	4	93.94%
2013	$20.50	$12.00	$8.50	1	10590	82	14	8	475	170	17	4	78	4	95.12%
2014	$22.00	$11.00	$11.00	1	13260	58	19	9	392	184	15	4	48	10	82.76%
2015	$24.60	$13.00	$11.60	1	8016	56	18	12	475	175	12	4	53	3	94.64%
2016	$28.90	$20.00	$8.90	1	9945	44	18	9	338	134	8	4	38	6	86.36%
2017	$27.00	$19.00	$8.00	1	7780	9	15	6	71	35	2	1	8	1	88.89%
2018	$18.06	$18.00	$0.06	15	2976	24	18	7	170	81	7	3	18	6	75.00%
2019	$29.20	$25.00	$4.20	8	3995	31	19	9	206	88	8	4	25	6	80.65%
2020	$36.00	$32.00	$4.00	8	3915	22	20	7	189	62	7	3	17	5	77.27%
2021	$41.80	$40.00	$1.80	7	4915	17	9	8	87	45	6	3	12	5	70.59%

Note: To exit out of the hidden page, click your report name in the top left.



b. Novak Djokovic

Novak Djokovic, one of the "Big Three", has topped the ATP ranks on 13 different years and till this day is ranked as the No. 1 tennis player in the world! He also holds the record for highest total career prize money earnings and most majors won. His most successful year was 2015, where he reached 15 finals and won 10 titles in a single season.



Year	Avg_Earnings	Endorsements	Prize_Money	Rank	Points	Matches	Max_Ace	Max_Df	Total_Ace	Total_Df	Tourneys	Slams	W	L	Win_Per
2012	$19.80	$9.50	$10.30	1	13630	87	16	6	502	147	17	4	75	12	86.21%
2013	$26.90	$14.00	$12.90	1	12920	83	22	7	476	118	17	4	74	9	89.16%
2014	$33.10	$21.00	$12.10	1	13130	69	18	5	428	105	15	4	61	8	88.41%
2015	$48.20	$31.00	$17.20	1	11360	88	15	7	471	135	17	4	82	6	93.18%
2016	$55.80	$34.00	$21.80	1	16585	74	19	12	301	188	17	4	65	9	87.84%
2017	$37.60	$28.00	$9.60	2	11780	40	14	9	169	79	11	3	32	8	80.00%
2018	$23.50	$22.00	$1.50	1	8045	70	23	11	366	161	18	4	56	14	80.00%
2019	$50.60	$30.00	$20.60	1	9135	64	15	9	368	159	15	4	54	10	84.38%
2020	$44.60	$32.00	$12.60	1	9720	46	17	8	278	137	9	3	41	5	89.13%
2021	$38.00	$30.00	$8.00	1	12030	62	26	7	447	148	13	4	55	7	88.71%

That concludes our guided analysis – but look, you have a great dashboard now and your curiosity doesn't have to stop here.

Appendix

Appendix A: Access Software

Steps to accessing SAS Viya for Learners for the first time:

1. Navigate to the SAS Viya for Learners webpage, https://www.sas.com/viyaforlearners.
2. Click the "Access for Educators" or "Access for Students" button based on your role.
3. Log in with your SAS Profile that is linked to an email address affiliated with an academic institution. If you don't have a SAS Profile, click here to set one up. Accept the terms and conditions to access the software.
4. The Viya for Learners launch page opens: https://learn.sas.com/vfl.
5. Launch the software by clicking the blue "Launch Software" button.

For future sessions with VFL, you can access VFL by visiting https://learn.sas.com/vfl then logging in with your SAS profile.

Contact us at academic@sas.com if you have any questions about access.

Appendix B: Helpful Documentation

There are several helpful online resources that can help you better understand predictive modeling in SAS Visual Analytics. Below are some helpful links to get you started:

- Need reasons to get excited about SAS Viya for Learners 4 (VFL4)? Start Here!
- Explore and Visualize Data with SAS Visual Analytics - SAS Viya Quick Start - SAS Video Portal (14 min)
- Unveiling the Latest Upgrades: SAS Visual Analytics Empowers the Academic Community Part 1 of 3
- Unveiling the Latest Upgrades: SAS Visual Analytics Empowers the Academic Community Part 2 of 3
- Unveiling the Latest Upgrades: SAS Visual Analytics Empowers the Academic Community Part 3 of 3

Appendix C: Recommended Learning

The SAS Global Academic Program offers free e-learning courses for students to learn SAS through the SAS Skill Builder for Students portal. The following e-learning courses and paths available are recommended to help with this activity:

- SAS Visual Analytics 1 for SAS Viya: Basics

- SAS Visual Analytics 2 for SAS Viya: Advanced

Faculty and Students can also gain access to these courses via the SAS Learning Subscription. With a free trial, learners are granted access to an extensive library of SAS eLearning courses.