Data Narrative-3

Pranav Patil, 22110199
Computer Science department,
IIT Gandhinagar
Gandhinagar, India
pranav.patil@iitgn.ac.in

Prof. Shanmuganathan Raman
Electrical Engineering and *Computer Science department,*
IIT Gandhinagar
Gandhinagar, India
shanmuga@iitgn.ac.in

I. Overview of the data set

The dataset is related to the tennis matches played during the 2013 tennis season. The dataset consists of 8 files and each file has 42 columns and a minimum of 76 rows.

The tennis match dataset contains information on various aspects of professional tennis matches, such as player names, match results, number of sets won by each player, serving statistics, and more. The dataset can be used to explore different factors that may contribute to a player's success in a tennis match, such as their serving accuracy, the number of unforced errors they make, and the number of breakpoints they win. It can also be used to compare the performance of different players, as well as to analyze trends over time or across different tournaments. Overall, this dataset provides a rich source of information for anyone interested in the sport of tennis and the factors that determine success in professional matches.

II. Scientific Questions/Hypotheses

1. What is the relationship between a player's first serve percentage and the number of aces they hit in the Australian Open Men's 2013 tournament?
2. What is the impact of the first serve percentage (FSP) on the result of a tennis match?
3. How does the ratio of winners to unforced errors of the Wimbledon 2013 finalists (men and women) compare, and how does it impact the result?
4. On an average men serve more aces than women in a match (hypothesis).
5. Which players defeated the Big 3 i.e. Roger Federer, Novak Djokovic and Rafael Nadal at Grand Slams in 2013? You among the Big 3 held the edge over the other two that year?
6. How does the number of aces, winners, break points, net points depend on the nature of the court on which the match is played on?

7. Which male player(s) was the most consistent in the Grand Slams in 2013?
8. How do the different parameters compare for the French Open 2013 semi-finalists?

III. Details of libraries and functions

1. Pandas: python library for data manipulation and analysis
2. Matplotlib.pyplot: python library for data visualization
3. Seaborn: a popular Python library used for creating informative and visually appealing statistical graphics, including heatmaps, scatterplots, and bar charts, making it a valuable tool for data visualization and exploratory data analysis.

IV. Answers to the questions

1. What is the relationship between a player's first serve percentage and the number of aces they hit in the Australian Open Men's 2013 tournament?

From the output, we can see a scatter plot showing the relationship between the first serve percentage and the number of aces in the Australian Open men's tennis tournament in 2013. We can see that as the first serve percentage increases, the number of aces also tends to decrease. The red line represents the linear regression model that has been fitted to the data, which shows the best linear approximation of the relationship between the two variables. Overall, we can conclude that there is a negative relationship between first serve percentage and the number of aces in the tournament. One possible explanation to the result obtained could be that as a player focuses on increasing their first serve percentage, they may be sacrificing some power or speed on their serve. This could result in fewer

aces, as the serve is not as strong or difficult for the opponent to return. Additionally, a player with a higher first serve percentage may be more likely to aim for the service box rather than the corners or lines, which could also lead to a decrease in aces.
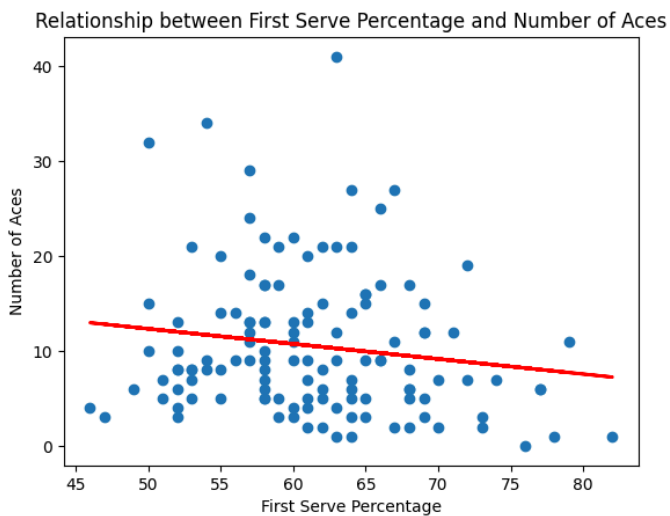

Fig. 1. Relationship between first serve percentage and number of aces

2. The impact of the first serve percentage (FSP) on the result of a tennis match.

We computed the number of matches won by the player who had a higher first serve percentage (FSP) and the number of matches won by the player who had a lesser FSP. Then a pie chart is plotted showing the proportion of matches won by the player with higher FSP and the player with lesser FSP. From the result, we can conclude that having a higher first serve percentage (FSP) increases the likelihood of winning a match. The pie chart shows that 58% of matches were won by the player with a higher FSP, while 42% were won by the player with a lesser FSP.

The following result could be because a high first serve percentage is indicative of a player's ability to consistently get their first serve in, which can give them an advantage in the match. When a player gets their first serve in, they have the opportunity to control the point and potentially win it outright with an ace or a winner. Additionally, a high first serve percentage can put pressure on their opponent to play more conservatively on their own serve, which can give the player with the higher first serve percentage more opportunities to break their opponent's serve and ultimately win the match.
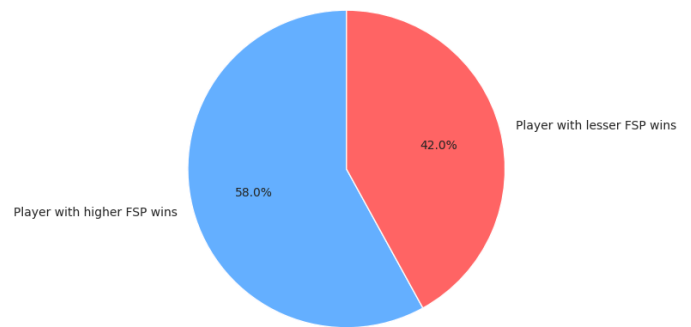

Fig. 2. Impact of FSP on the result of a match

3. How does the ratio of winners to unforced errors of the Wimbledon 2013 finalists compare, and how does it impact the result?

In general, the ratio of winners to unforced errors should be greater than 1 in order to win a match. This means that a player is hitting more winners (shots that are not returned by the opponent) than unforced errors (shots that result in a point for the opponent due to the player's mistake). A ratio less than 1 would mean that the player is making more unforced errors than winners, which is not a winning strategy in most cases.

The output also confirms the above theory. The output consists of two sets of plots. The first set is for the men's Wimbledon 2013 final and the second set is for the women's Wimbledon 2013 final. In both sets of plots, the ratio of winners to unforced errors for Player 1 and Player 2 is calculated and plotted against the player names. It can be observed that for both the finals, the player having a higher winners to unforced errors ratio than the other player, wins the match.
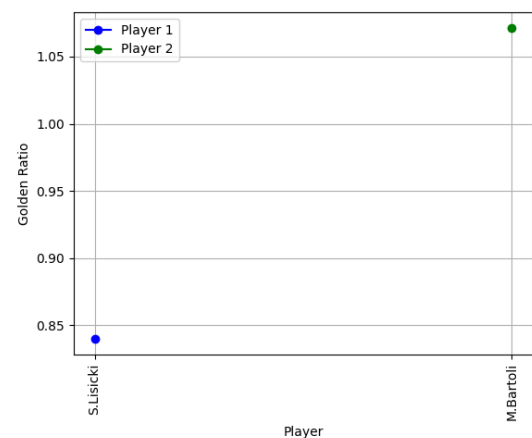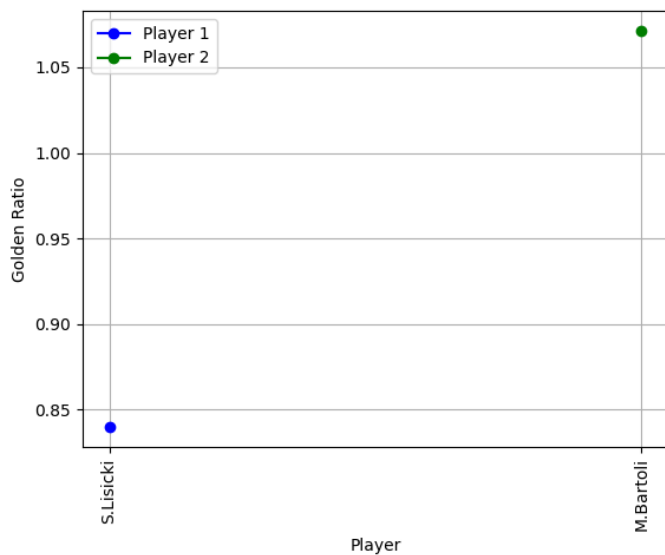
Fig. 3. ratio of winners to unforced errors of the Wimbledon 2013 finalists

4. On an average men serve more aces than women in a match (hypothesis).

There can be several reasons why men serve more aces than women in tennis. One of the main reasons is the difference in physical attributes between men and women. Men, on average, have greater upper body strength, which allows them to serve the ball faster and with more spin than women. Additionally, men tend to be taller than women, giving them an advantage in terms of serving angle and power. Furthermore, men tend to play best-of-five sets matches at major tournaments, while women play best-of-three sets. This means that men have more opportunities to serve and hit aces over the course of a match.

The output also confirms that there is sufficient conclusive data to accept the hypothesis that men serve more aces in a match than women.

	Average number of Aces served in a match
Men	17.76
Women	5.15

Table. 1. Average number of aces served in a match by men and women

5. Which players defeated the 'Big 3' i.e. Roger Federer, Novak Djokovic and Rafael Nadal at Grand Slams in 2013? You among the Big 3 held the edge over the other two that year?

The "Big Three" of tennis refer to Roger Federer, Rafael Nadal, and Novak Djokovic. Defeating the Big 3 in Grand Slams is considered a huge achievement because they are three of the greatest tennis players of all time. They have dominated men's tennis for more than a decade and have won a staggering number of Grand Slam titles, with Djokovic, Nadal, and Federer occupying the top three spots for most Grand Slam singles titles won by a male player. Not only do they have exceptional skills, but they are also known for their mental toughness, which allows them to perform at their best even under pressure. Their consistency and longevity at the top of the game are remarkable, and they have set the bar very high for the rest of the players on the tour. Therefore, defeating one of the Big 3 in a Grand Slam requires not only exceptional skills and physical fitness but also a strong mental game, determination, and the ability to perform at one's best under intense pressure.

The plot shows the number of times each player defeated the Big 3 in Grand Slam matches during the year 2013. From the plot, we can see that only seven players were able to defeat the Big 3 during the year 2013. Rafael Nadal, the member of the Big 3, defeated the other two members three times (most by any player), while Stanislas Wawrinka was second on the list with two victories over Big 3 in that year. This plot shows the rarity of defeating the Big 3 in Grand Slam matches, which is considered a significant achievement in tennis.
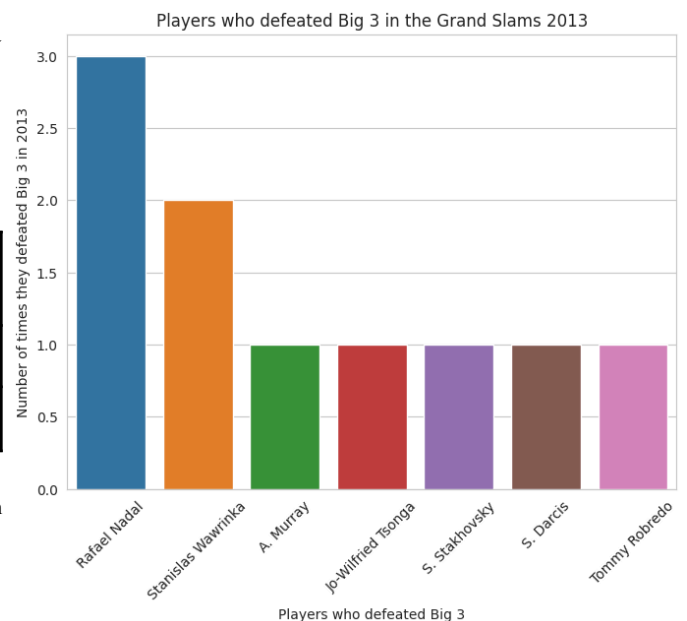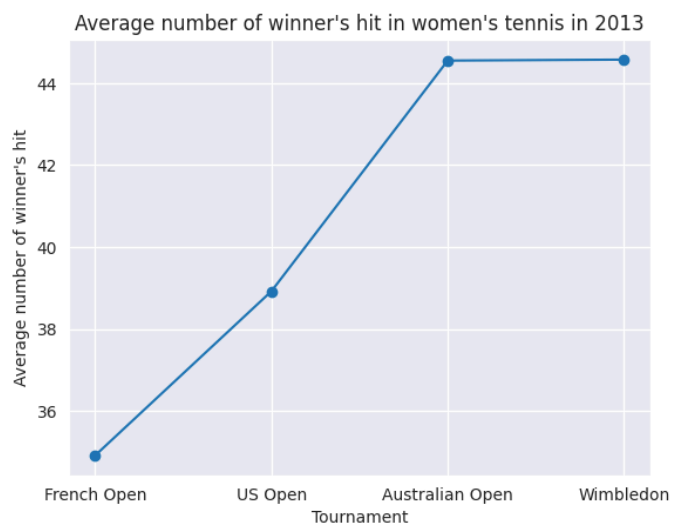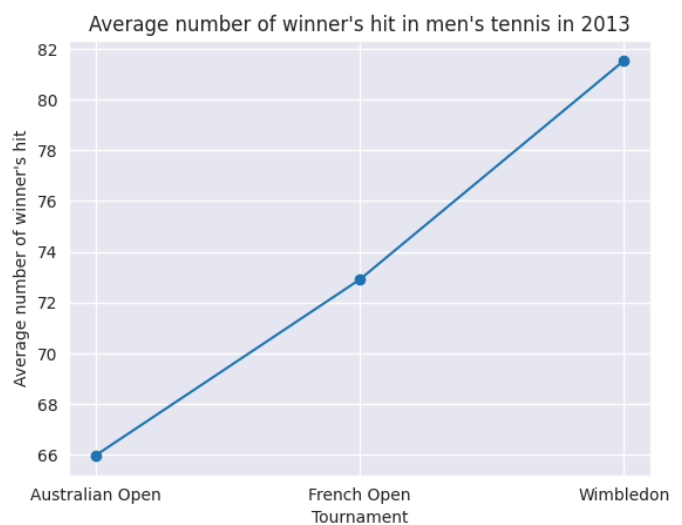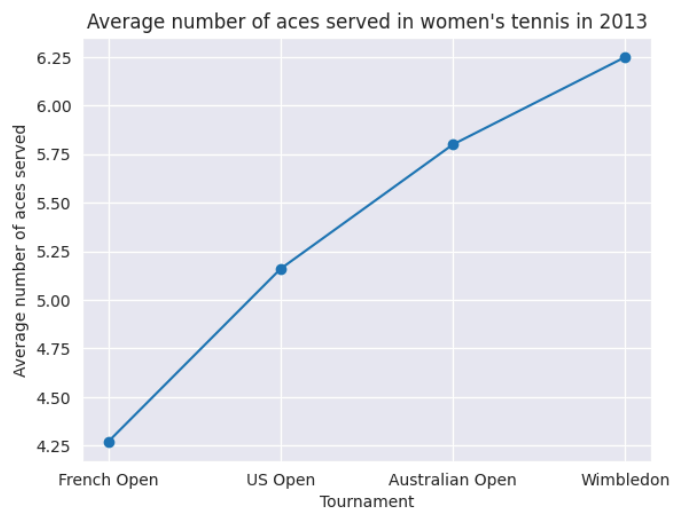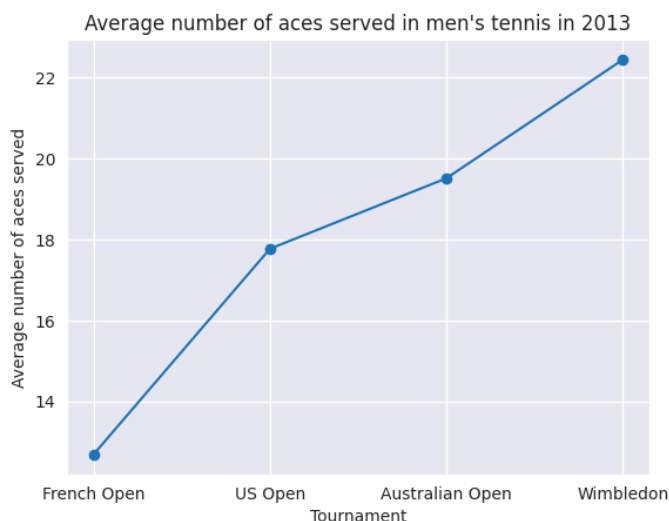

Fig. 4. Players who defeated 'Big 3' in the Grand Slams 2013

6. How does the number of aces, winners, break points, net points depend on the nature of the court on which the match is played on?

The number of aces, winners, break points, and net points can depend on the nature of the court on which the match is played. For example, on a fast surface like grass (Wimbledon is played on grass courts), the ball tends to bounce lower and skid through the court, which favors big servers and aggressive players who can hit flat shots and approach the net to finish points quickly. As a result, we might expect to see more aces, winners, and net points on grass courts, and perhaps fewer break points as the server has an advantage. On a slow surface like clay (French Open is played on red clay courts), the ball tends to bounce higher and slower, giving players more time to set up their shots and defend against aggressive play. This can lead to longer rallies and more opportunities to break serve. We might expect to see fewer aces and winners on clay courts, but more breakpoints and a higher percentage of points won by players at the net. On hard courts (Australian Open and US Open are played on hard courts), which are medium-fast surfaces, the ball tends to bounce true and allow players to hit a variety of shots.It's worth noting that other factors can also influence the number of aces, winners, break points, and net points, including the skill and style of the players themselves, the weather conditions, and the balls and equipment being used. Nevertheless, the nature of the court surface is one important factor to consider when analyzing these statistics in tennis. The plots support the comments made above.
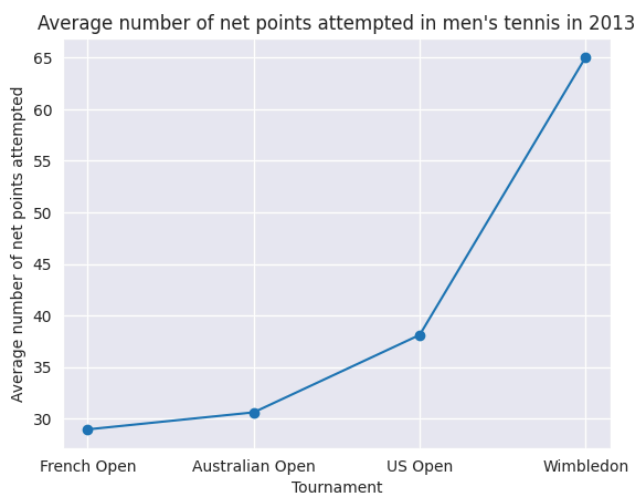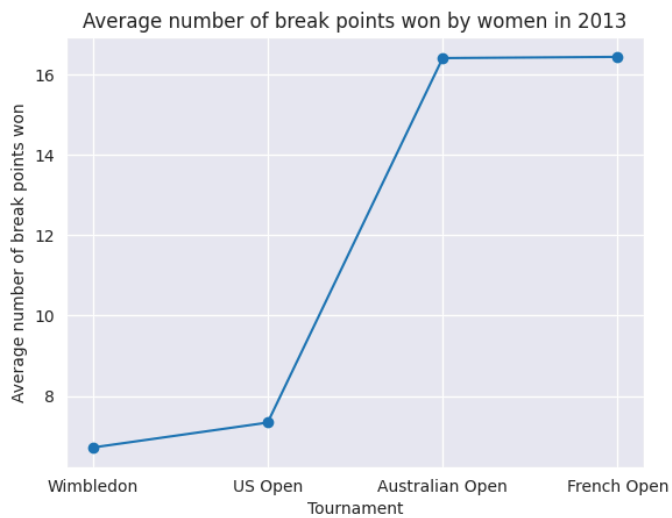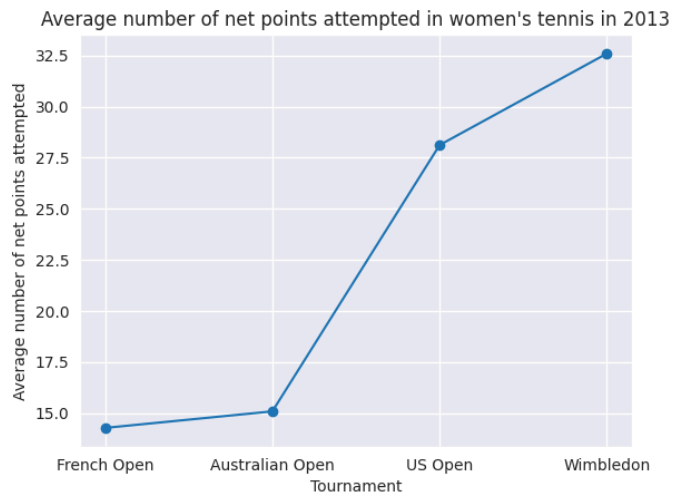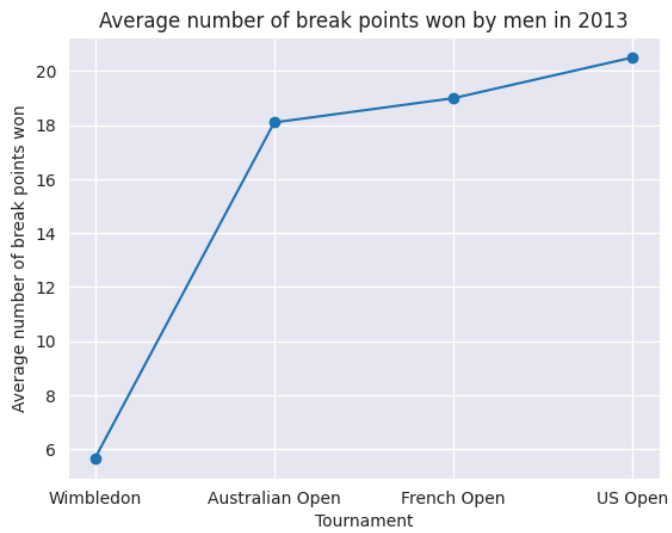
Average number of aces served in women's tennis in 2013

Average number of winner's hit in men's tennis in 2013

Average number of winner's hit in women's tennis in 2013

Average number of aces served in men's tennis in 2013

Fig. 7. Impact of the court surface on the different parameters

7. Which male player(s) was the most consistent in the Grand Slams in 2013?

Staying consistent in all the Grand Slams is very difficult, as it requires a high level of physical fitness, mental toughness, and skill. The Grand Slams are the most prestigious and physically demanding tournaments in tennis, and players must compete at a high level for two weeks in each tournament. In addition, the surface of each tournament differs, with the Australian Open and the US Open being played on hard courts, the French Open on clay, and Wimbledon on grass. This requires players to adjust their game to the different surfaces, which can be challenging. Furthermore, players must be able to maintain their form over the course of the entire year, as the Grand Slam events are spread out over several months. This means that players must be able to manage their schedules and avoid injuries and fatigue. Only a few players have been able to achieve consistency in all the Grand Slam events, which is a testament to how difficult it is to accomplish.

In the following analysis, the player who managed to reach the quarter-finals (round 5) of all the four Grand Slams is termed as the 'consistent player'.
From the analysis it was concluded that only two players satisfied this criteria. The players who were consistent in the Grand Slams in 2013 were Novak Djokovic and David Ferrer.

Breakdown of Novak Djokovic's 2013 Grand Slam Journey:

Grand Slam	Result
Australian Open	Quarter-Finalist (round 5)
French Open	Semi-Finalist (round 6)
Wimbledon	Finalist (round 7)
US Open	Finalist (round 7)

Table. 2. Novak Djokovic's 2013 Grand Slam Journey

Breakdown of David Ferrer's 2013 Grand Slam Journey:

Grand Slam	Result
Australian Open	Quarter-Finalist (round 5)
French Open	Finalist (round 7)
Wimbledon	Quarter-Finalist (round 5)
US Open	Quarter-Finalist (round 5)

Table. 3. David Ferrrer's 2013 Grand Slam Journey

8. How do the different parameters compare for the French Open 2013 semi-finalists?

We compare different parameters like first serve percentage, break points won, winners hit, aces hit and unforced errors made by each semi-finalist of French Open 2013.
All these parameters are analyzed as these have a direct influence on the result of the match. And in almost every plot, we can see that Rafael Nadal has outperformed his rivals and this is the reason why he managed to win the French Open that year.
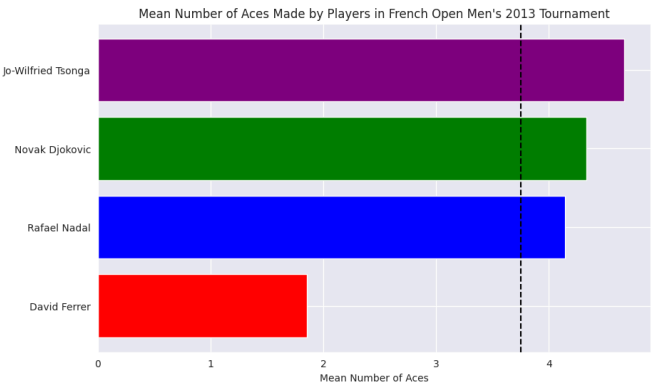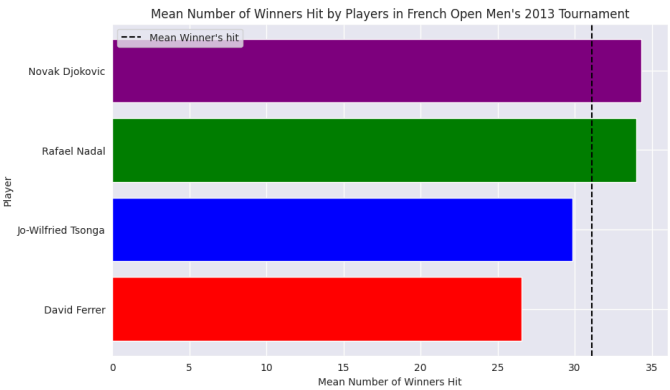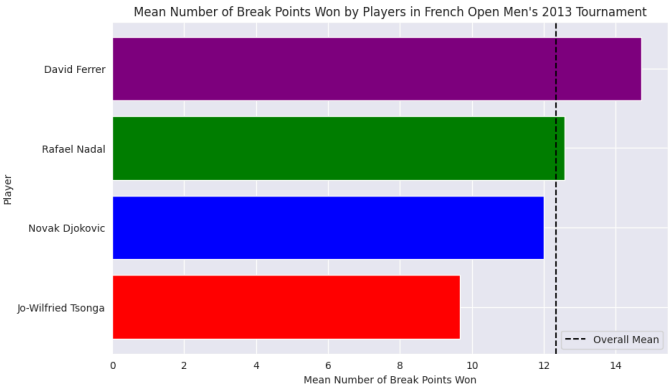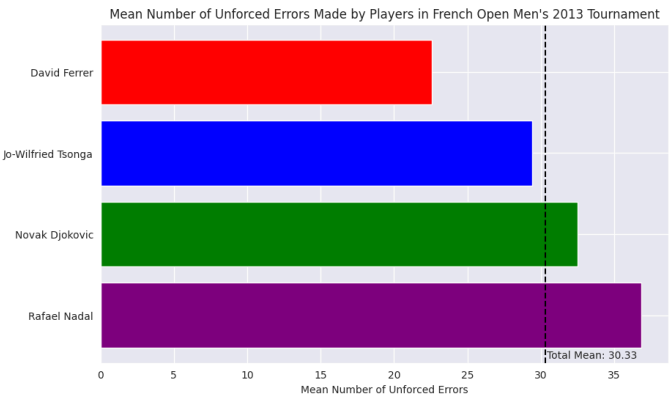
Mean Number of Unforced Errors Made by Players in French Open Men's 2013 Tournament
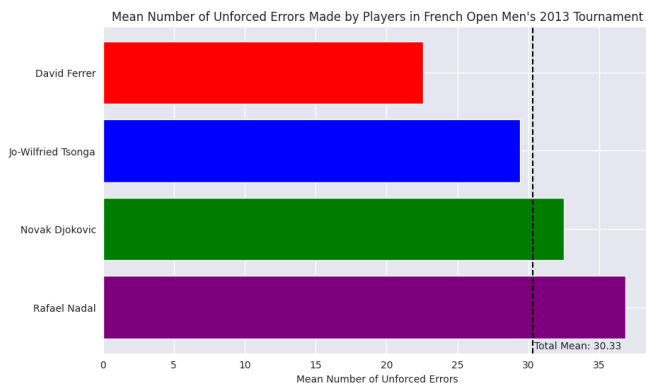
Fig. 9(1). Comparison of different parameters for the French Open 2013 semi-finalists

V. Summary of the observations

1. As the first serve percentage of a player increases, the number of aces hit tends to decrease.
2. Having a higher first serve percentage (FSP) increases the likelihood of winning a match.
3. Hitting more winners than unforced errors greatly improves one's chance to win the match.
4. On an average men serve more aces than women in a match.
5. Rafeal Nadal was the best player among the 'Big 3' in the year 2013. He won 2 out of the 4 Grand Slams whereas the other two players failed to win any. Also in the head to head record he led them by 3-0 (in 2013).
6. The nature of the court surface is one important factor to consider when analyzing the statistics in tennis.
7. Novak Djokovic and David Ferrer were the two most consistent male players in 2013.

VI. REFERENCES

1. https://www.geeksforgeeks.org/introduction-to-pandas-in-python/
2. https://matplotlib.org/
3. https://pandas.pydata.org/docs/
4. https://www.w3schools.com/python/pandas/default.asp