Allegheny College
Department of Computer Science

Senior Thesis

PercenTennis

by

Wonjoon Cho

ALLEGHENY COLLEGE
COMPUTER SCIENCE

Project Supervisor: **Douglas Luman**
Co-Supervisor: **Gregory Kapfhammer**

4/5/2022

Abstract

PercenTennis is a tool that tracks and displays data in a way to assist coaching for a match in a live or post match analysis. The introduction provides background of existing tennis analytics, which primarily utilizes video analysis. The related work portion expands on identifying which data is collected and most significant. Serves will be examined further in this section. The method of approach goes into detail about the data that is collected and calculated to produce metrics. This section will continue with an overview of the UI design. Lastly, the evaluation strategy explains the process of how accurate the data is compared to a higher level analytics system and how usable the tool is.

Acknowledgment

First and foremost, many thanks to my advisor and second reader Professor Kapfhammer for helping me throughout my years at Allegheny.

For software development, Nick Sarno or "Swiftful Thinking" on YouTube, taught me the basics of SwiftUI so that I could build my application. For me, he was the best guide to learning SwiftUI.

Finally, this comp, app, and paper would not have been possible without the support of Professor Luman. He had understood the vision of my project and pushed it to existence. He saw my passion and told me to run with it. So I did. Truly, thank you.

Abbreviations

ATP	Association of Tennis Professionals
ITA	Intercollegiate Tennis Association
WTA	Women's Tennis Association

Glossary

Serve	The first shot of a tennis point. A player has two chances to serve.
Holding Serve	A service game is held when the server wins the game.
Breaking Serve	A service game is broken when the returner wins the game.
Ace	A serve that was not touched by the opponent.
Double Fault	A fault occurs when a player misses their serve. A double fault is two missed serves resulting in losing the point.
Return	The second shot of a tennis point. A player can return the ball with a forehand or backhand.
Forehand	The side which the dominant hand strikes the ball.
Backhand	The side which the non-dominant hand strikes the ball.
Approach Shot	A transitioning shot to go to the net.
Volley	A shot where the ball has never bounced. This typically occurs when a player is at the net.
Unforced Errors	An error that occurs when there is no pressure from the opponent's shot. A returnable shot.
Forced Errors	An error that occurs when there is pressure from the opponent's shot. A difficult shot.
Winner	Similar to ace, is when the opponent does not make contact with the ball.
Deuce	Deuce is when the score is 40-40, and one player must win two points in a row. The deuce side is the right side of the center hash mark.
Advantage (Ad)	Ad-in is when the server is up and requires one more point to win. Ad-out is when the returner is up. The ad side is the left side of the center has mark.

For Mom and Esther

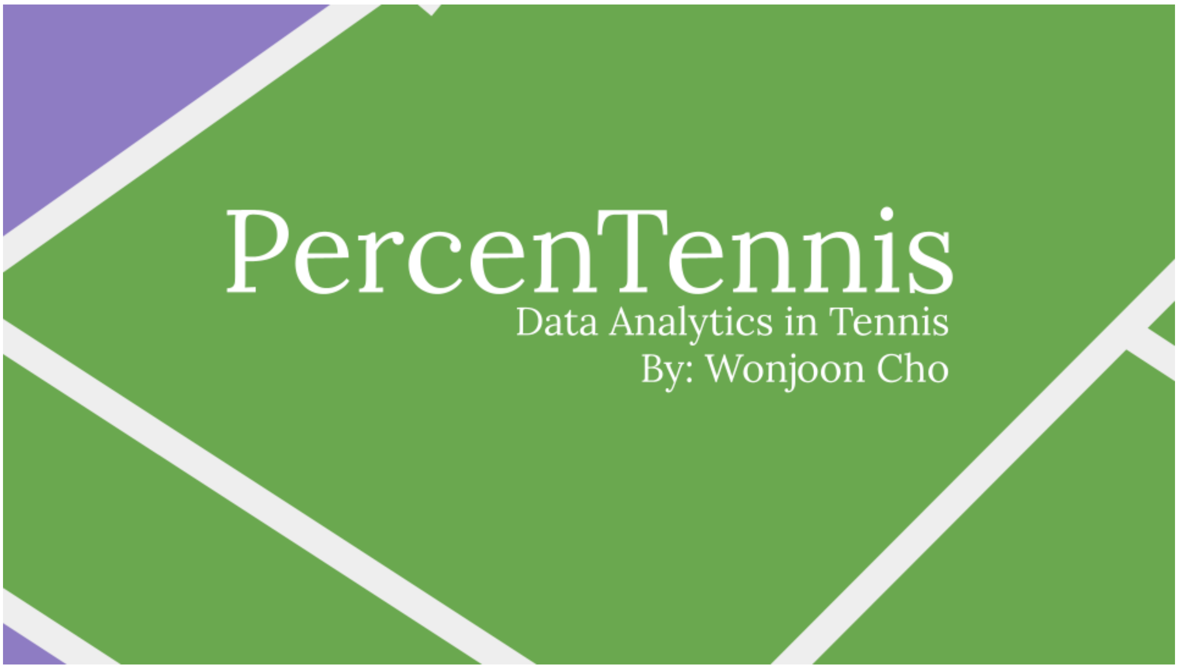
Contents

List of Figures

List of Tables

Chapter 1

Introduction



1.1 Motivation

When a player starts their career in competitive tennis, they develop their shots and gravitate towards a certain a play style. There are four major categories of play styles in tennis: Aggressive base-liner, all courter, serve-vollyer, and counter-puncher. Any particular player would fall in the range of these categories. The key difference in each style is the percentage of certain shots. An aggressive base-liner may have a low percentage in their shots, but have a high win rate. Alternatively, a counter-puncher may have a high percentage shot, and still have a high win rate. Dependent on these percentages, a player can maximize their play style to win a match. However in junior and college level tennis, many players struggle to min-max their tennis game. In other words, most players can identify their shot weapons and weaknesses, however, many players don't know how to utilize them optimally. This tendency causes a player to become predictable. Becoming more knowledgeable about the data in tennis matches can give a player just enough confidence and a solid game plan to win their match.

In the world of Big Data, a surge of data analytics in sports is becoming more prominent especially in the professional and collegiate scenes. Tennis follows this trend by using video feed and machine-learned A.I. This kind of analysis is helpful for those who can afford expensive cameras and software. Top professionals analyze their previous matches to give them a slight edge on their opponents. Players such as Novak Djokovic and Roger Federer partner with *Golden Set Analysis* to receive detailed reports about their game and more importantly, their opponent's game. While video review is slowly being incorporated in college level sports, a video analytics system in tennis is far too expensive especially at a Division 2 or Division 3 school.

PercenTennis offers a cheap alternative that hopes to match professional level analytics. This app is geared to a wide range of users from coaches, parents, to avid tennis enthusiasts. Since the data is generated through manual input, it is expected to see some margin of user error. This will not prove to be a problem due to the large amount of data points that are tracked. Tennis is a fast paced sport that takes about two hours to finish an average best out of three set match. During a match each player receives twenty seconds in between points. More often than not, most players don't use the full time. A manual tennis tracker like PercenTennis should have a simple and efficient UI to ensure the most amount of data can be tracked for each point.

1.2 Current State of the Art: Tennis Trackers

In the App Store, there are many tennis trackers that are similar to this project. The main problems for these apps is that they are either too limited with the data, or too specific causing it to have a bad user experience. In most cases these apps display long lists of percentages that give no significant information. Furthermore, some of these apps implement subscription pay-walls to provide more features and deeper analysis. PercenTennis tackles all these problems that are found in the current state. It offers good UX design, significant data portrayals, and most importantly the accessibility to anyone.

1.3 Ethics

In regards to the ethics of usage in collegiate level tennis, this application must follow |ITA rules and regulations|. The ITA organization is supportive to data collection and video review. Playfair, a partner with ITA, is the only acceptable form of video review in sanctioned tournaments. However, during team and individual competition, electronic devices may be used by coaches. These devices include phones, tablets, etc. for messaging and data purposes. It is impractical for the head coach to sit on one court to track one match. The role for a match trackers are set to a designated player assistant. DPA's are non-playing players for a dual match. These assistants must follow all the rules for coaching. Outside of collegiate tennis, match tracking is supported and encouraged in tennis organizations.

1.4 Goals of the Project

The goal of this project is to simply bring tennis analytics to the hands of players with ease of use and no cost. The statistics that will be portrayed will be similar to the ones that are shown in ATP matches. While the app itself doesn't give specific coaching advice, coaches, parent, or players must have some kind of prior knowledge of the game to understand what these metrics mean. A combination of serve placement, win percentage, and return percentage can tell a player which serves can be more advantageous in an important point. More will be explained in the method of approach section.

1.5 Thesis Outline

This project is composed of five chapters. The first chapter is the introduction. This chapter provides insight on the motivation of this project. Then a section covers the current state of tennis trackers and their faults and improvements that are made for this project. A discussion of the ethics of tennis tracking is explored. Finally, the last section talks about the goals of the project.

The second chapter is related work. In this chapter, research is conducted to understand the importance of tennis statistics. The research will provide knowledge about which metrics are the most important to track and all of the metrics that can not be tracked, but is still important to understand. For example, serve speed is impossible for a human to accurately gauge. However if a tracker notices that faster serves equate to lower percentage of points won, then a general conclusion of the matter can be an advice to slow down the pace of the serve. Based on the research conducted, the metrics are developed in a way that is efficient to track.

The third chapter is the method of approach. This chapter answers how this project will be completed. The first section explains the scoring system of a tennis match. Using SwiftUI, the app must understand how the scoring system works so that the points and server can be tracked. Data sets are provided to explain which data points are tracked and how they are stored. Once the data is collected, each data point is used to calculate metrics through algorithms. The last section of this chapter explores the UI that is developed for the app.

The fourth chapter is about the experimental results. This chapter establishes the validity of the project. In ATP matches, statistics are often displayed during the matches to provide insight about the match. Infosys ATP Stats is generally a good indicator of how a match is going. Professionals that use deeper statistics involve a lot more variables when analyzing their matches. These statistics can help the players gain clarity on how they win or lose their matches. The experiment is the comparison between ATP stats to the stats produced by PercenTennis.

The fifth chapter is discussion and future work. There will be a discussion about the results from the experiment. An example set analysis will describe a set and report findings and advice. Afterwards, plans for future development will detail possible features that will be implemented to enhance the application.

Chapter 2

Related Work

2.1 Data Types in Tennis

Data in sports take form of box-score data, tracking data, and meta-data. Box-score data is observable data that is recorded by a spectator. In the case of tennis, box-score data could be data such as score, serve percentage, and shot selection. Tracking data is data that is created by using some kind of technology to video, map, or calculate trajectories to visualize and track a ball or person depending on the sport. Tennis primarily uses a form of video analysis. By using Hawk-Eye technology, ball tracing is a popular method of tennis analytics. Another example of video analysis is tracking the physiology of a player's shot in slow motion to analyze their technique and movement. Video analysis is a powerful method of accumulating tracking data which can be used to dissect whole matches and record an immense amount of usable data. Meta-data consists of contextual factors in a sport outside the sport itself. For example, factors such as player style, court type, sponsorship, and equipment. There is no limitations as to what meta-data could be, and meta-data can be self generated from meta analysis of box-score data and tracking data. [10]

The data type that will be used in the tennis tracking app is box-score data. Key factors that are taken into consideration in data collection is serve, return, and rally. One possible use of tracking box-score data is the combination of serve percentages and score. Looking at the percentage of winning a service point, a model could predict the percentage of winning the game. While the concept of developing a predicting feature can be challenging with the limited knowledge of statistical analysis, articles written about modelling a match have been written and could be used to future development. Developing a model using Markov chain can show a predictability table of the percentage chance of winning in a game or the match. At the beginning of each game, the server starts at an increased chance of winning a game at sixty percent using the said model. The next portion will go further into detailing as to why the serve gives a player a statistical advantage. [2]

Acknowledging the probability of each point can be helpful for a player to play smarter tennis. A player that can recognize important points creates an advantage opposed to those who have a first to five points mentality. For example, a crucial point in a game is 30-30. Statistically speaking, if the server wins that point, they have a ninety-two percent chance of winning the game, as opposed to a seventy-one percent chance for the returner if they win the point. The server only needs to win one point with the advantage of their serve, while the returner must win three additional points to win the game. Understanding the percentages of these points and having insight on the highest win percentage serves allows players to play smarter in more high stakes points. [12]

2.2 What is the Most Important Factor in Tennis?

When asking the question of which shot is the most important shot in tennis, the most obvious answer is the serve. The serve is the only shot in tennis that is completely controllable by a single player. There are outside factors that can influence a player's serve, but in it's essence, a serve should provide an advantage. The advantage is more evident in professional level than the junior level counter part. A trend shows that there is a direct correlation when higher the level of tennis the more significant the serve will be. [13] Furthermore, professional players typically serve more aces regardless of gender and serve fewer double faults. This is natural as higher level tennis involves a smaller margin of errors. Another trend that was found was that professional players win more second serve returns and male professionals lose more first serve return points. It is important to understand the differences between gender and age. A coach should strategize different advice for each player, based on play style, age, height, gender, and optimizing their strengths. [5]

In a study performed in professional tennis matches held between 2003 and 2008 using Hawk-Eye technology to track ball placements, an analysis of the serve was extensively researched and the results show the significance of the three factors of the serve: speed, location, and spin. Firstly, it is important to note that the court surface is an key indicator of how effective a serve can be. Clay courts are slower due to the higher friction and grip when the ball would bounce on the court. As a result, this allows returners more time to react to a serve and successfully return the serve. Grass courts which are the fastest type of courts, have the highest efficiency with serves. On clay courts, a little more than half the points are finished within five shots, about fourth of the points are medium length rallies that are in the range of six to nine shots, and the remaining points are long rallies that surpass nine shots. On grass courts, nearly ninety-seven percent of points are finished within five shots due to the nature of the fast court. This proves that on faster courts, the serve and returns are a determining factor of a match, while on slower courts baseline rallies are more important. [8]

In tennis the server is allowed two serves for each point. This allows for strategies to form regarding the three factors of a serve. The first factor of the serve, speed, is what most viewers of the sport care about. Naturally, the first serve tends to have a higher velocity compared to the second serve. This can be found in figure 2.1 which shows the service speed of first and second serves for male and women professionals. The serve percentage is inversely proportional to the speed of the serve. So the faster a serve is hit, the higher percentage of faults a player would make. However, faster serves are directly linked to higher percentage chance of winning a point. Figure 2.2 shows a chart of serve speed and probability to win for male and women professionals. Players must find a balance of serve percentage and speed. The study shows that second serves

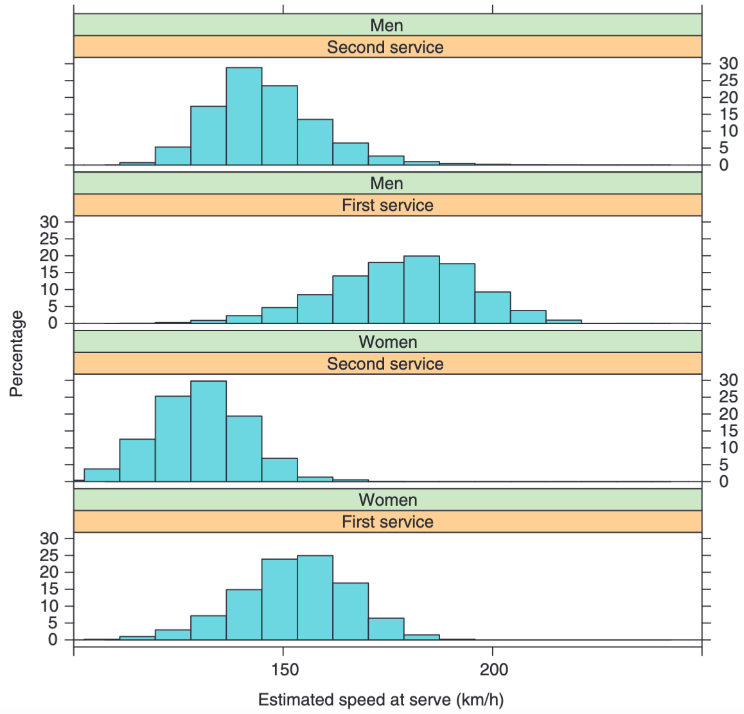


Figure 2.1: Distribution of Service Speed

even out the playing field by reducing the win percentage for the server. Therefore, when a player is competing against a better player, they could take higher risks on their first serve to hold an advantage over the opponent. In figure 2.1, while there is a correlation with the higher serve speed to the probability of winning the point, for the second serve, when serves exceed 225 km/h shows to have lower win rate than serves at 200 km/h. On the other side of the net, returns are just as important as the serve and having a high percentage of returns can result to higher percentage chance of breaking the opponents serve. [8]

The second factor is the location of the serve. A heat map of the serve position varies from first and second as shown on figures 2.3 and 2.5. The first serve, on either side of the court, are typically served away from where the opponent is standing via T or wide serves in a even distribution. A body serve may be a change-up by jamming the returner's swing. The second serve, is typically served towards the opponents backhand, which is normally considered the weaker side, via T on the deuce side and wide on the advantage side. This trend is evident in the study as most players are right-handed. The risk factor is how close to the line a player would serve towards, and similar to speed, a higher risk can be a strategy against better players. [8]

The third and final factor of the serve is spin. A serve can take in the form of slice, top-spin (kick), or flat (no spin). For a right-handed player server, slice serves would bounce and veer to the left, while a top-spin serve would bounce and "kick" to the right. First serves are typically hit flat as it generates a higher velocity. In men's tennis kick serves are the preferred serves as it would normally be aimed towards the opponents backhand. Studies show that on the first serve both high velocity and high

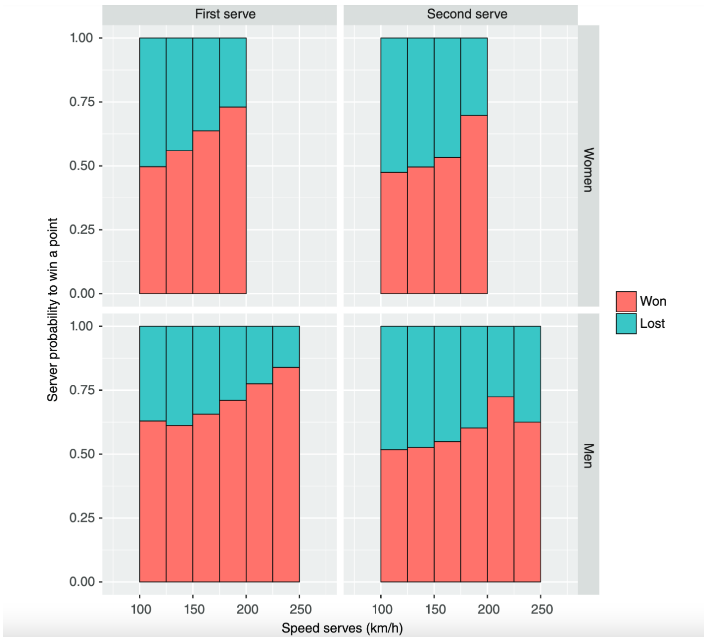


Figure 2.2: Service Speed and Probability to Win the Point

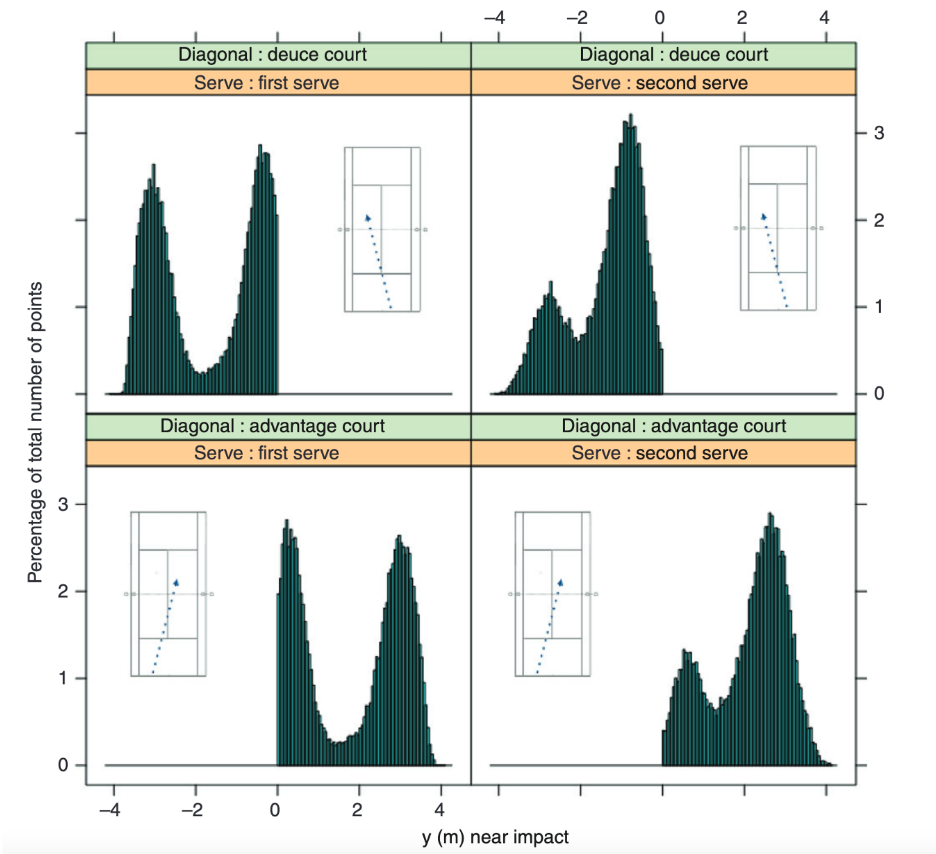


Figure 2.3: Men's Serve Placement

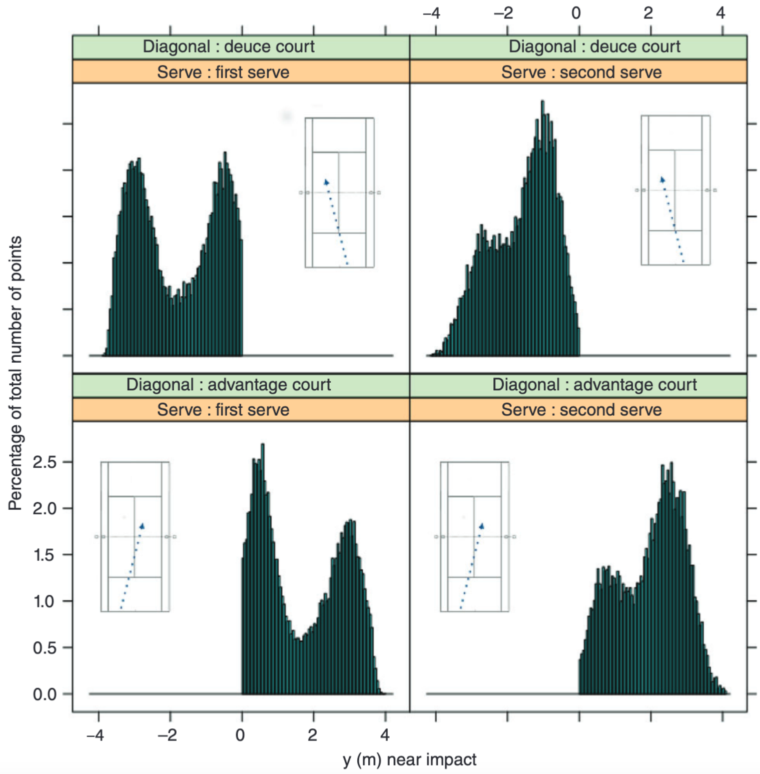


Figure 2.4: Women's Serve Placement

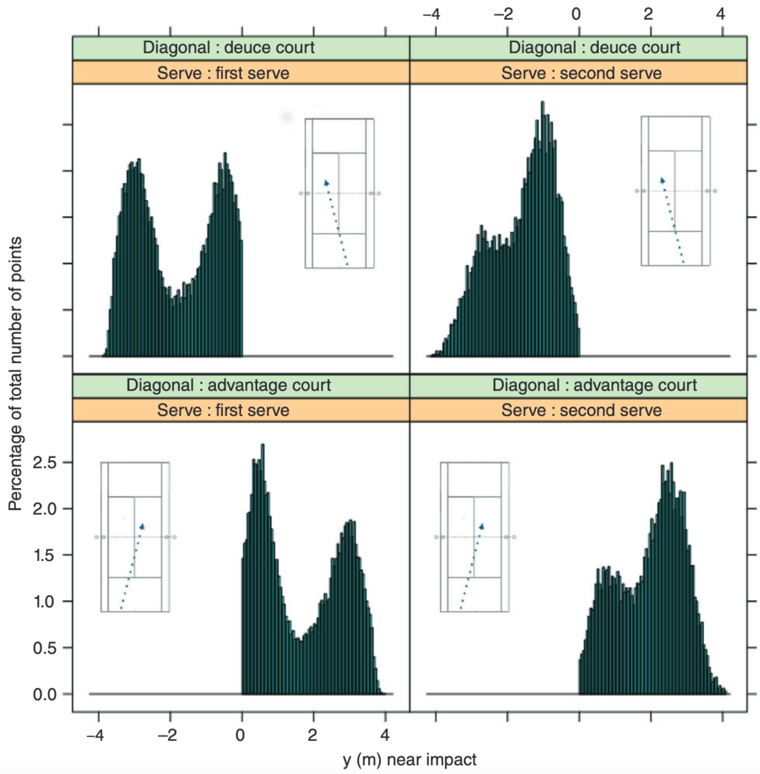


Figure 2.5: Women's Serve Placement

spin intensity have equal efficiency; however, a higher velocity is preferred. To optimize the serve, all types serves must be developed as the returner could adjust to a server's preferred serve and have a higher percentage return. [8]

2.3 Strategies for Serves

By keeping in mind the percentages and statistics of serves, a player would want to construct the most optimal strategy for winning matches. Strategies are formed during a match by exposing the opponents weaknesses and strengths. Patterns of errors may form and contribute to a player's strategy to apply pressure to those patterns. As mentioned in the previous section, the serve is a key component to a player's victory or defeat. When developing a serving strategy, a huge advantage comes to those who can play the percentages and is blessed with a little bit of luck.

The aforementioned factors of speed, location, and spin can allow for numerous variations of how a serve can be hit for first and second serves. However, in an unspoken rule, it is almost every case where the first serve is typically a strong serve and the second serve is a weak serve. A strong serve refers to a higher percentage of winning the point off of the serve, and a weak serve refers to the mitigation of losing a point at the cost of lowering the percentage of winning the extended point. [3] Strategies regarding the different combinations of weak and strong serves can be performed by players based on the circumstances of a match.

Like mentioned before a typical player would follow a strong first serve and weak second serve. However, Barnett suggest that a player's service strategy could change throughout a match. Looking at figure 2.6, player should play two high risk "strong" serves when the percentage of points won with a weaker serve is less than the percentage of a player's strong serve multiplied by the percentage of points won with a strong serve. This makes sense as long as a player can make their strong serve with a decent percentage. A study compared Andy Roddick and Rafael Nadal's serves. The report states that Roddick has a better chance of winning the match by serving two strong serves on grass courts which makes the serves more effective, while on clay and some hard courts he should serve a strong serve and a weak serve. Nadal's best strategy aligns with a strong serve and a weak serve on all court types. An interesting note is that though Roddick finds success with his strong serves, statistically speaking, there is no significance in playing two strong serves. A potential reasoning behind it's success is that a strong second serve could come as a surprise for Nadal, and thus give Roddick some advantage. [11]

Serve strategies can vary based on court type, the player's serve percentage, and the opponent's return percentage. After each set a good amount of data is collected to assess the serving and returning parameters to configure which serve strategy to partake. This is especially simple when a tennis tracker is involved to display the percentage of each serve and the win percentage of each serve and simple arithmetic is used to calculate the inequality to determine the serve strategy. The list and table below shows the inequalities to determine the serve strategy in a match in progress. [11]

Let:
a_{his} = percentage of high risk serves in play for player i on surface s
a_{lis} = percentage of low risk serves in play for player i on surface s (note that this percentage is not needed)
b_{his} = percentage of points won on high risk serves (conditional on them being 'in') for player i on surface s
b_{lis} = percentage of points won on low risk serves (unconditional) for player i on surface s
c_{his} = percentage of points won on return of high risk serves (conditional on them being 'in') for player i on surface s
c_{lis} = percentage of points won on return of low risk serves (unconditional) for player i on surface s
d_{hijs} = percentage of points won on high risk serves (conditional on them being 'in') for player i, for when player i meets player j on surface s
d_{lijs} = percentage of points won on low risk serves (unconditional) for player i, for when player i meets player j on surface s
c_{has} = average percentage (for all players) of points won on return of high risk serves (conditional on them being 'in') for surface s
c_{las} = average percentage (for all players) of points won on return of low risk serves (unconditional) for surface s

Serving Statistic	Inequality
First Serve Percentage	$\hat{a}_{his} > d_{lijs} \ / \ d_{hijs}$
Percentage of Points won on First Serve	$\hat{b}_{his} > d_{lijs} \ / \ a_{his}$
Percentage of Points won on Second Serve	$\hat{b}_{lis} < a_{his} * d_{hijs}$

Table 2: Inequalities for determining serving strategies during a match in progress

Figure 2.6: Serving Strategy

2.4 What Does the Data Show?

Chris Gray used a model cited from Klassen and Magnus [6] which they've built upon Newton and Keller [9]. This model is used to predict match outcomes by taking data from the top 50 players. Particularly this model looks at the probability of each player holding serve, in other words the probability of winning their own service game as shown in figure 2.7. This means that a player with a strong service game will more likely win the game. By combining two players' point probability on serve, tables in figure 2.8 are formed to show the probability of a player winning a set or a match. These tables show that typically a player with a better serve would win a set upwards to roughly 80 percent of the time and a match upwards to 90 percent of the time. This shows how tournament favorites are able to make incredible comebacks from a two set deficit at about a 40 percent chance. It is important to note that this model is based of off the top 50 male ATP players, where the serve is much more of an impact. While the model for collegiate level players is unknown without player data, an educated estimate can be taken from this model to be applied to looking at data in tennis. It is safe to assume that most players at this level fall under the 60 to 65 percentage range in winning a point on serve. It is possible using this model the app could take in serve win percentage and generated a predicting algorithm for player set win percentage on serve. [4]

A research was conducted using data collected in 2018 WTA Grand Slam tennis matches to determine which factors were most indicative for a victory when comparing two players. This study focused on the important nature of serves and returns and compared many different variables in regards to player characteristics and serve performance. While height provides a physiological advantage for player's serve speed

Table 1. Probability of holding serve

Probability of winning a point on serve	*Probability of holding serve*
0.60	0.7357
0.61	0.7562
0.62	0.7759
0.63	0.7947
0.64	0.8126
0.65	0.8296
0.66	0.8457
0.67	0.8609
0.68	0.8751
0.69	0.8884
0.70	0.9008

Note: Probabilities shown to four decimal places due to the small increments

Figure 2.7: Probability Table for Holding Serve

and power, studies find that the positive advantage doesn't continue after 6 feet for male professionals. [7] For female professionals, most players fit between 5 feet 5 inches to 6 feet tall which is above average height for women. Additionally majority of the top female players are above 6 feet tall. Height plays a bigger role in women's tennis than men's. The table in figure 2.9 shows the player characteristics and their win and loss percentages. [1] Hand orientation can also be an advantage to those who are left-handed. Loffing et al. noticed a difference in distribution of first and second serves placements. This meant left-handed players were applying different spin which right-handed players had to make an immediate adjustment. However Ma et al. found in his studies that top level left-handed players proved to have no significant advantage to their right-handed counterparts. He stated that the reasoning could be that top level players train for both right and left handed players. [7] Left-handed players at a junior or collegiate level may have an initial advantage, however as a match progresses a skilled opponent should be able to make adequate adjustments for the advantage to be non-existent.

The factors that both studies that have found is to be the most influential for winning a match is unsurprisingly serve and returns. For female professionals, the study concluded that the top four comparisons to determine favorable odds for a match are in order: percentage of first serve return won, percentage of first serve won, percentage of second serve return won. They noted that whoever has a higher percentage of first serve return won is 1.263 times more likely to win the match. Whoever has the higher percentage of first serve won is 1.244 times more likely. For second serve returns won is 1.078 and second serve won is 1.124. This shows the importance of first serves and more importantly the significance of the return of first serves. These ratios can be

Table 2. Probability of winning a set for given probabilities of winning a point

Probability of player A winning a point on serve \ Probability of player B winning a point on serve	0.60	0.61	0.62	0.63	0.64	0.65	0.66	0.67	0.68	0.69	0.70
0.60	0.50	0.53	0.57	0.60	0.63	0.66	0.69	0.72	0.75	0.77	0.79
0.61		0.50	0.53	0.57	0.60	0.63	0.66	0.69	0.72	0.74	0.77
0.62			0.50	0.53	0.57	0.60	0.63	0.66	0.69	0.72	0.74
0.63				0.50	0.53	0.57	0.60	0.63	0.66	0.69	0.71
0.64					0.50	0.53	0.57	0.60	0.63	0.66	0.69
0.65						0.50	0.53	0.56	0.60	0.63	0.66
0.66							0.50	0.53	0.56	0.59	0.62
0.67								0.50	0.53	0.56	0.59
0.68									0.50	0.53	0.56
0.69										0.50	0.53
0.70											0.50

Table 3. Probability of winning a 'best of three sets' match for given probabilities of winning a point

Probability of player A winning a point on serve \ Probability of player B winning a point on serve	0.60	0.61	0.62	0.63	0.64	0.65	0.66	0.67	0.68	0.69	0.70
0.60	0.50	0.55	0.60	0.65	0.69	0.74	0.77	0.81	0.84	0.87	0.89
0.61		0.50	0.55	0.60	0.65	0.69	0.73	0.77	0.81	0.84	0.87
0.62			0.50	0.55	0.60	0.65	0.69	0.73	0.77	0.80	0.84
0.63				0.50	0.55	0.60	0.65	0.69	0.73	0.77	0.80
0.64					0.50	0.55	0.60	0.64	0.69	0.73	0.77
0.65						0.50	0.55	0.60	0.64	0.69	0.73
0.66							0.50	0.55	0.60	0.64	0.68
0.67								0.50	0.55	0.59	0.64
0.68									0.50	0.55	0.59
0.69										0.50	0.55
0.70											0.50

Figure 2.8: Probability Table for Winning a Set and Match

Table 3. Summary for Player Characteristics

Variable	Group	Description	Won	Lose
Age	1	<= 20 years old	6 (1.21%)	10 (2.04%)
	2	21 – 30 years old	359 (72.53%)	381 (77.60%)
	3	>= 31 years old	130 (26.26%)	100 (20.37%)
Height	1	<= 160 cm	7 (1.44%)	9 (1.99%)
	2	161 – 170 cm	145 (29.87%)	155 (34.29%)
	3	171 – 180 cm	270 (55.67%)	225 (49.78%)
	4	>= 181 cm	63 (12.99%)	63 (13.94%)
Weight	1	<= 60 kg	110 (24.02%)	99 (25.38%)
	2	61 – 70 kg	289 (63.10%)	251 (64.36%)
	3	>= 71 kg	59 (12.88%)	40 (10.26%)

Figure 2.9: Table of Player Characteristics and Win/Loss Percentage

Table 8. Odds Ratio of Significant Variables

Variable	Odds ratio	Interpretation
PFSI	1.047	Players with higher percentage of first serve in are 1.047 times more likely to win the tennis match.
PFSRW	1.263	Player with higher percentage of first serve return won are 1.263 times more likely to win the tennis match.
PFSW	1.244	Player with higher percentage of first serve won are 1.244 times more likely to win the tennis match.
PSSRW	1.078	Player with higher percentage of second serve return won are 1.078 times more likely to win the tennis match.
PSSW	1.124	Player with higher percentage of second serve won are 1.124 times more likely to win the tennis match.

Figure 2.10: Table of Odds Ratio of First and Second Serve and Return Points Won

found in the figure 2.10. [1]

Some other factors that may influence a match is number of double fault, number of aces, and service speed. For ATP professionals, the losing players' average amount of double faults per match is 5.3 while the average amount of aces is 6.6. For the winning players the double fault goes up to 4.2, however the aces goes up to 8.9. [7] Following a similar trend for WTA players, winning players average 3.35 double faults and 3.44 aces per match. Losing players average 4.2 double faults and 2.32 aces. Players who mitigate the amount of double faults and be aggressive with at least their first serve consistently compared to their opponent will be more likely to win. What's interesting is that Abidin et al. found the average first serve speed is higher for winners, but their average second serve speed is lower than the losers. Although it is not by a significant amount, it makes sense for players to reduce the speed for more reliability and less double faults. [1] The tables in figures 2.11 and 2.12 show serve and other miscellany descriptions based on match outcome for female professionals [1] and male professionals [7]

Table 7. Descriptive Statistics of Serve Performance based on Match Status

Variable	Match Status	Mean	Std. Devation
PFSI	Won	62.50 (n=507)	7.424
	Lose	61.23 (n=507)	7.957
PFSW	Won	70.10 (n=507)	8.555
	Lose	57.74 (n=507)	9.497
PFSRW	Won	42.14 (n=507)	9.429
	Lose	29.70 (n=507)	8.410
PSSW	Won	51.28 (n=507)	11.314
	Lose	39.85 (n=507)	10.056
PSSRW	Won	44.93 (n=507)	10.950
	Lose	36.21 (n=507)	10.346
DF	Won	3.35 (n=507)	2.621
	Lose	4.10 (n=507)	2.652
Aces	Won	3.44 (n=507)	3.193
	Lose	2.32 (n=507)	2.366
FSS	Won	157.85 (n=347)	0.459
	Lose	156.06 (n=347)	0.503
SSS	Won	131.60(n=346)	0.463
	Lose	132.45 (n=346)	0.469

Figure 2.11: Table of Won and Loss Percentage of Serves and Returns

Table I. Descriptive statistics by match outcomes.

		Match Outcomes		
No.	Predictors	Losing Matches	Winning Matches	Independent samples t-test
1	Age (years)	25.4 (3.4)	25.2 (3.4)	−3.42***
2	Stature (cm)	184.2 (6.4)	184.7 (6.2)	6.17***
3	Mass (kg)	78.5 (6.7)	79.3 (6.6)	7.72***
4	Ranking	93.2 (101)	59.6 (85.9)	−24.17***
5	Years as a professional	6.6 (3.3)	6.8 (3.2)	3.94***
6	Aces	6.6 (5.7)	8.9 (6.4)	26.01***
7	Double faults	5.3 (3.4)	4.2 (3.1)	−21.32***
8	% 1^{st} serve	58.6 (8.4)	60.8 (8.3)	17.62***
9	% 1^{st} serve points won	66.0 (9.1)	76.3 (7.9)	82.17***
10	% 2^{nd} serve points won	43.0 (9.2)	53.4 (9.8)	74.08***
11	% 1^{st} serve return points won	23.4 (9.8)	33.5 (9.6)	69.78***
12	% 2^{nd} serve return points won	45.8 (9.9)	55.9 (9.4)	70.76***
13	% break points converted	36.0 (22.8)	47.0 (16)	36.98***
14	% break points saved	52.2 (15.9)	63.4 (22.8)	37.62***
15	% serve points won	56.4 (6.8)	67.4 (6.3)	112.38***
16	% return points won	32.1 (7.4)	42.8 (6.9)	100.16***
17	% total points won	44.5 (4.4)	54.7 (4.2)	160.65***
	Dominant hand			Chi-square
18	Right hand	84.5%	85.9%	7.09**
	Left hand	15.5%	14.1%	

Note: *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

Figure 2.12: Table of Won and Lost Matches by Miscellany Variables

Chapter 3

Method of Approach

Before the tool is created, the metrics and statistics that are used must be clearly defined. This section will detail the structure of the data sets and the data points that will be collected for calculations. Then, from the research conducted and the key metrics that clearly describes the important aspects in tennis, algorithms will be described. Most of the statistical calculations will be have similar processes, so two or three algorithms will be explained in detail. Finally, a detailed documentation of the software development process of creating the app itself will conclude this chapter.

3.1 Data Sets

As mentioned before, the three key components that my project is focusing is the serve, return, and the last ball of a rally. Unsurprisingly, there are three data sets that contain data points of certain aspects of a tennis match. All three data sets have two attributes in common: score and player. The score is kept track to identify which at which point the metrics are gathered. The player attribute determines which player the particular shot is hit. The first data set is serve. Serve is comprised of five attributes: side, serveNumber, ace, position, and result. Side determines which side the serve is coming from. The deuce side $\rightarrow$ 1 and the advantage side $\rightarrow$ 0. The serveNumber dictates which serve was hit. The first serve $\rightarrow$ 1, the second serve $\rightarrow$ 2, and a double fault (missing both serves) $\rightarrow$ 3. Ace (a nonreturnable serve that passes the opponent without contact) is shown either with a 0 or a 1 based whether or not an ace was hit. Position refers to where the serve lands on the opponent's side. T $\rightarrow$ 1, body $\rightarrow$ 2, wide $\rightarrow$ 3. Lastly, the result is shown either with a 0 or a 1 based on whether the player wins the point after they serve.

The second data set is return. The return is also comprised of five attributes: serveNumber, side, shot, mistake, result. The first attribute uses the serveNumber from the previous data set so that it could be used to show percentages between first and second serve returns. The side is also used from the previous data set. The shot attribute refers to which kind of shot was hit for the return. Forehand $\rightarrow$ 1 and backhand $\rightarrow$ 2. Mistake shows either a 1 or 0 based on whether the player misses the return. Lastly result, similar to the serve data set is shown either with a 0 or a 1 based on whether the player wins the point after they receive or if they are aced.

Table 3.1: Serve Data Set

Serve		
Attribute Name	Sample	Range
Player	1	1,2
Score	15-30	0,15,30,40
Side	1	1,2
serveNumber	2	1,2,3
Ace	0	0,1
Position	3	1,2,3
Result	0	0,1

Table 3.2: Return Data Set

Return		
Attribute Name	Sample	Range
Player	2	1,2
Score	40-0	0,15,30,40
Side	2	1,2
serveNumber	1	1,2,3
Shot	2	1,2
Mistake	1	0,1
Result	0	0,1

The third and final data set is rally. Rally has three attributes that are different from the previous two. The first attribute is the shotResult. The shotResult is determined by unforced errors (mistakes made without pressure from the opponent) $\rightarrow$ 1, forced errors (mistakes made from pressure from the opponent) $\rightarrow$ 2, and winners (a nonreturnable shot that passes the opponent without contact) $\rightarrow$ 3. Unforced errors and forced errors are subjective calls, however in most cases it is clear to see how the mistake was hit. When tracking the points, it is important to understand that when unforced or forced errors are the input the player chosen has lost the point, and when a winner is the input the player chosen has won the point. Shot type dictates which particular shot was last hit. Forehand $\rightarrow$ 1, backhand $\rightarrow$ 2, approach shot (an attacking transitioning shot to the net) 3, and net (either with a volley or overhead) $\rightarrow$ 4. The rally length is divided into three categories, short rally (a rally lasting 1-4 shots) $\rightarrow$ 1, a medium rally (5-9 shots), and a long rally (10+ shots).

Table 3.3: Rally Data Set

Rally		
Attribute Name	Sample	Range
Player	2	1,2
Score	40-0	0,15,30,40
shotResult	1	1,2,3
Shot Type	4	1,2,3,4
Rally Length	2	1,2,3

3.2 Algorithms

Just by looking at a spreadsheet of all the data sets is unhelpful when it comes to analyzing a match. To provide a clear insight of the match, algorithms take in the data sets as input and output the occurrences of certain shots and provide some percentages using the different data points. The first algorithm looks at the serve data set and calculates the first and second serve percentage. The algorithm takes in the input of serveNumber and result to calculate the first serve and second serve percentage, and also the win percentage of each serve.

Algorithm 1 Serve and Win Percentage

```
Initialize three variables to zero to indicate a count of each serve and calculate total
serve.
s1,s2,s3 = 0
totalServe = s1 + s2 + s3
for serveNumber in Dataset Serve do
    if serveNumber == 1 then
        ++s1
    end if
    if serveNumber == 2 then
        ++s2
    end if
    if serveNumber == 3 then
        ++s3
    end if
end for
for result in Dataset Serve do
    if result == 1 && serveNumber == 1 then
        ++FSW
    end if
    if result == 1 && serveNumber == 2 then
        ++SSW
    end if
end for
firstServePer = s1 ÷ totalServe
secondServePer = s2 ÷ (totalServe - s1)
firstServeWin = FSW ÷ s1
secondServeWin = SSW ÷ s2
```

The second algorithm focuses on the placement of the serve. From the research I stated that the location of the serve is one of the three factors that is important to a the serve. This algorithm again looks inside the serve data set and counts the instances of each placement of the serve, then calculates the percentages of where the serve lands.

Algorithm 2 Serve Placement

```
Initialize three variables to zero to indicate a count of each serve placement and calculate total serve.
t,b,w = 0
totalServe = t + b + w
for placement in Dataset Serve do
    if placement == 1 then
        ++t
    end if
    if placement == 2 then
        ++b
    end if
    if placement == 3 then
        ++w
    end if
end for
tServePer = t ÷ totalServe
bServePer = b ÷ totalServe
wServePer = w ÷ totalServe
```

The third algorithm moves on from the serve and calculates the percentages of returns made and which shot was hit. This algorithm looks in the return data set and takes the input from serveNumber, shot, and mistake. To calculate the first serve return percentage, the algorithm takes in the instances of forehand and backhand first serve returns and adds them. Then takes the forehand and backhand first serve return mistakes and adds them. Then divide both sums to show the percentage of first serve return. Calculating the second serve return percentage is a similar concept but using the second serve. Calculating the forehand and backhand return percentage could be a good indicator showing which return shot is weaker. To calculate, the algorithm counts the instances when a forehand and backhand was hit, then counts the instances of forehand and backhand mistakes. Afterwards dividing the two will produce the forehand and backhand return percentage.

Algorithm 3 Return Percentage

```
Initialize all forehand and backhand return variables and forehand and backhand that were made in variables.
fReturn, bReturn, fReturnMade, bReturnMade = 0
for shot in Dataset Return do
    if shot == 1 then
        ++fReturn
    end if
    if shot == 2 then
        ++bReturn
    end if
end for
for mistake in Dataset Return do
    if shot == 1 && mistake == 0 then
        ++fReturnMade
    end if
    if shot == 2 && mistake == 0 then
        ++bReturnMade
    end if
end for
fReturnPer = fReturnMade ÷ fReturn
bReturnPer = bReturnMade ÷ bReturn
```

Finally the last algorithm that will be used in my project is the rally algorithm. The special case for the rally metric is that there is no percentages or calculations that are necessary for this metric. The rally metric simply outputs the count of each attribute for a comparison between the two players. For this algorithm, it looks into the rally data set and assigns values from instances in the data set.

Algorithm 4 Rally

```
Initialize all metrics.
UFE, FE, Winner = 0
for shotResult in Dataset Rally do
    if shotResult == 1 then
        ++UFE
    end if
    if shotResult == 2 then
        ++FE
    end if
    if shotResult == 3 then
        ++Winner
    end if
end for
```

3.3 Scoring System

The most difficult section about this project will be developing a scoring system that will keep track of tennis points which is notoriously known to be complicated and confusing. Here is a link that explains the official tennis scoring system: |USTA Official Scoring|.

1. Game Scoring: On the tennis court, the right side of the center hash is called the *deuce side*, the left side is called the *ad side*. Every game starts on the deuce side for both players, with every point proceeding alternating sides. In the simplest form, a player must win four points to win a game. The scoring for a game progresses in this order 0, 15, 30, 40, and game. The only stipulation during a game is when both players are tied 40-40 which is considered *deuce*. In this case, a player must win two times in a row to win the game. When the server wins the deuce point, the score is ad-in; alternatively, if the server loses, the score is ad-out. After each game, the server changes. If the match format is no-ad, that means that the receiver will choose the side to return from, then they play one point to decide the game. For the interest of time, the initial app will always play out deuce points.

2. Set and Match Scoring: A tennis match consist of either a best out of five or best out of three sets. Typically best out of five matches are only played in men's professional matches during major tournament (i.e. Wimbledon or US Open). Professional women, college players, and competitive juniors play best out of three. A set is composed of winning 6 games. On every odd sum of games that are played, both players switch ends of the court. If the set score becomes 5-5 a player must win by two games. If the set count becomes 6-6 the competitors will play a 7-point tiebreaker. A match can be played to the full three or five sets, otherwise players would play a 10-point tiebreaker for the last set.

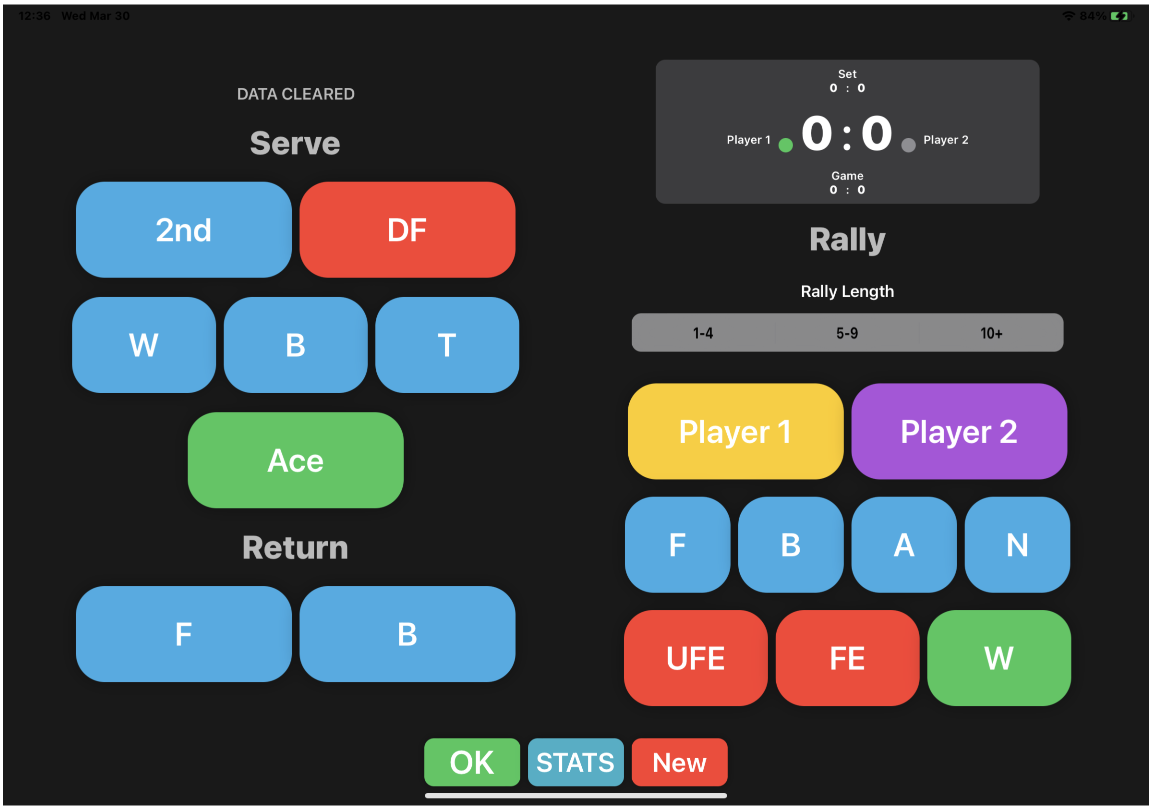
3. Tiebreakers Tiebreakers are probably the most confusing scoring system out of the three formats. Tiebreakers start off with the server serving one point on the deuce side, then alternating serves every two points. Every point the players alternate between deuce side and ad side. After a multiple of six total points (i.e. 6,12,18), the players switch sides of the court. If a 7-point tiebreaker is played during a set, the person that returned first will serve the next game in the subsequent set.

3.4 Using SwiftUI

Once the data sets and algorithms are fully planned out, the next step is to create and design the tool itself. The software of choice for this project is SwiftUI. SwiftUI is an Apple IOS software language which allowed for creating applications on the iPhone or iPad. An advantage to SwiftUI is its dynamic nature and easy to learn syntax. There are many resources online available to learn Swift coding.

This section will lay out the development process of this project. The architectural style of this project follows the model-view-viewmodel or MVVM. Loosely speaking, this pattern is the developmental process of facilitating the GUI (view) and the back-end logic (model) into separate files and connect them to one another through the the view-model. The whole process can be divided into five parts: UI, data management, scoring, data calculations, and data visualization.

3.4.1 Designing the UI



During the initial stages of the application, there were different layouts when designing the placement of the buttons. Eventually through a process of testing and tweaking, a natural and intuitive flow formed the placements of the buttons. The flow of tracking the point goes from top to bottom, left to right. Each view consists of vertical and horizontal stacks of various buttons. The button's locations are managed on the screen in relation to other buttons with axial stacks which can be nested. For example, the main view is divided into three parts: serve and return, rally, and the settings bar. The serve and return v-stack (vertical) consists of the input text, an h-stack (horizontal) of serve number buttons, an h-stack of serve location buttons, the ace button, and finally an h-stack of the buttons for the return shot. The stacks are dynamic which means it would calculate the spacing of the buttons in relation to one another, not in relation to the screen.

The first section, as described before is the serve and return buttons. At the top, the user can select second serve or double fault. If the first serve is made in, neither buttons are pressed. The first serve is initially input by default. If a double fault occurs, then no other input is necessary, and the confirmation button should pressed. The serve placement is tracked with either wide, body, or T. These buttons change position based on the side that the serve is played. Next the return tracks the forehand or backhand. If a return mistake is hit, a down slide gesture onto the forehand or backhand return button is used. For a return winner an up slide gesture is used. When a point is played out into the rally, the rally length will default to the 5-9 rally length. The segmented slider for the rally length will determine how long the point is played. Lastly, whoever hit the last shot is tracked with which shot and what kind of mistake or winner is played.

The second section, is the v-stack that contains the score box indicating the set, game, and point scoring and which player is serving. Underneath the score, the rally length is determined using a segmented picker style. The two blank buttons indicate which player has hit the last ball of the rally, then the four buttons underneath indicate which type of shot was hit. Lastly, the three buttons in the bottom describes whether and error or winner was played.

Finally at the bottom of the main view there are three buttons. The first button on the left is the confirmation button which will take all the inputs of the point, and save them in the data-base. When a person tracks the point, the buttons that are pressed change opacity to show which buttons have been selected. The middle button titled 'STATS' will open a pop-up screen of the statistics that have been tracked so far in the match. Next, the 'New' button will clear all existing data and open a pop-up screen to enter two names that will populate the player buttons and the scoreboard. In a more finished product, the new button would probably take in another confirmation notification and open a view with more setting options to customize the scoring system for the match.

3.4.2 Establishing the Data

The data is stored in the app using core data. Core data saves data locally within the device in a persistent container. First the data base is established by creating entities with the attributes of the data points that are tracked. The attributes ask for a data type which in most cases are integer values following the data sets established in the previous section. Once the persistence is created, a model file which contains the variables of serve and return data initializes the values. While three models are used: serve and return, rally, and score, a more efficient data base could combine all the tracking data into one data base. However, with the current state of the application the data is saved using these models, and it would be more work and risk to refactor the code. One special data type is used for the return shot and shot type in the serve and return and rally entities. Enumerations are used for these data variables so that when a certain shot type or return is called, it would ask for the case type instead of an integer. To save the data in the application, a fetch call receives the data from the input, then saves it once the confirm button is selected. The raw data collection can be accessed in the match timeline tab in the 'STAT' view. Once a match is completed, the 'New' button clears all data stored in the application.

3.4.3 Coding the Scoring

This section was by far the most challenging in terms of programming. As mentioned before, the scoring in tennis is complicated and a headache to code with all it's stipulations and changes in server. The scoring logic takes in the pointResult value from the data base to calculate the score. This means that the code will reside in the view-model file. The point logic goes as follows. If the point result is true and the server is 1 then player 1 receives a point (p1 += 1). Alternatively if the point result is false and the server is 1 then player 2 receives a point (p2 += 2). If the server is 2 then the opposite point result will receive the point. Next the side is determined by the parity of the point total. If the point total is even, the side is deuce indicated by 1; and if the point total is odd, the side is advantage indicated by 2.

```
// Decide Point Result
    if serveReturn.pointResult {
        if serveReturn.server == 1 {
            p1 += 1
        } else {
            p2 += 1
        }
    } else {
        if serveReturn.server == 1 {
            p2 += 1
        } else {
            p1 += 1
        }
    }

    // Determine Side
    if (p1+p2) % 2 == 0 {
        side = 1
    } else {
        side = 2
    }
```

Before displaying the game score, the code checks to see if the two player's games are at 6-6 which would prompt a tiebreaker. The challenge with coding the tiebreaker is the change of server once after the first point and then after every two points. At the start of a first set tiebreaker, whoever started serving the set will start serving in the tiebreaker. The total games carries values between sets so that if a player started serving in the first set and finishes the set with an odd number, then the second player would start the next game. The code segment below shows the conditional statement that after one instance changes the server after every two points. If the tiebreaker proceeds past 6 points for both players, one of the players must win two points in a row.

```
    // Tiebreaker Conditional
    if p1g == 6 && p2g == 6 {
        p1Score = "\(p1)"
        p2Score = "\(p2)"
        // Assigning Server
        if (totalGame) % 2 != 0 {
```

```
            if (p1+p2) == 1 || ts == 0 || ts == 1 {
                server = 2
                ts += 1
            } else if ts == 2 {
                server = 1
                ts += 1
            } else if ts == 3 {
                server = 1
                ts = 0
            }
        } else if (totalGame) % 2 == 0 {
            if (p1+p2) == 1 || ts == 0 || ts == 1 {
                server = 1
                ts += 1
            } else if ts == 2 {
                server = 2
                ts += 1
            } else if ts == 3 {
                server = 2
                ts = 0
            }
        }
        if p1 > 6 && p1 >= p2+2 {
            p1s += 1
            ts = 0
            resetGame()
        }
        if p2 > 6 && p2 >= p1+2 {
            p2s += 1
            ts = 0
            resetGame()
        }
```

The game scoring is an else statement if the tiebreaker condition does not meet. First the score that is shown in the app must follow the scoring format that is universally accepted. Simply put, if p1 $=1$ then the score is 15, p1 $=2$ the score is 30, etc. To win a game, a player must win four points while the opponent has less than three points. Otherwise a Deuce Ad conditional statement is met. A player must then win two points in a row.

```
    else {
        // Player 1 Point Conditionals
        if p1 == 1 {
            p1Score = "15"
        } else if p1 == 2 {
            p1Score = "30"
```

```
        } else if p1 == 3 {
            p1Score = "40"
        }
        // Player 2 Point Conditionals
        if p2 == 1 {
            p2Score = "15"
        } else if p2 == 2 {
            p2Score = "30"
        } else if p2 == 3 {
            p2Score = "40"
        }
        // Game Conditionals
        if p1 == 4 && p2 < 3 {
            p1g += 1
            resetPoints()
        }
        if p2 == 4 && p1 < 3 {
            p2g += 1
            resetPoints()
        }
        // Deuce Ad Conditional
        if p1 >= 3 && p2 >= 3 {
            // if Advantage
            if p1 > p2 {
                p1Score = "Ad"
                p2Score = "P1"
            } else {
                p1Score = "Ad"
                p2Score = "P2"
            }
            // if Deuce
            if p1 == p2 {
                p1Score = "40"
                p2Score = "40"
            }
            // if player wins game
            if p1 == p2 + 2 {
                p1g += 1
                resetPoints()
            }
            if p2 == p1 + 2 {
                p2g += 1
                resetPoints()
            }
        }
    }
```

Lastly, the reset functions are provided below that occur when a game or set is won. This will reset the points and add game and set values to the score box in the main view.

```
func resetPoints() {
    // Determine Server
    print("\(totalGame)")
    if totalGame == 0 {
        server = 1
    } else if totalGame % 2 == 0 {
        server = 1
    } else {
        server = 2
    }
    totalGame += 1
    p1 = 0
    p2 = 0
    ts = 0
    p1Score = "0"
    p2Score = "0"
    side = 1
    // set over
    if p1g == 6 && p2g < 5 {
        p1s += 1
        resetGame()
        return
    }
    if p2g == 6 && p1g < 5 {
        p2s += 1
        resetGame()
        return
    }
    // Win by Two Conditional
    if p1g > 4 && p2g > 4 {
        if p1g == p2g + 2 {
            p1s += 1
            resetGame()
            return
        }
        if p2g == p1g + 2 {
            p2s += 1
            resetGame()
            return
        }
    }
}
```

```
func resetGame() {
    p1g = 0
    p2g = 0
    resetPoints()
    // calculate values for stat entity
    // save set statistics
}
func resetSet() {
    p1s = 0
    p2s = 0
    resetGame()
    // calculate values for match stat
    // save match stat
}
```

3.4.4 Data Calculations

Since the calculations require data information similarly to the scoring system, the algorithms also take place in the view-model file. There are several calculations that are performed at the same time which can be accessed in the 'STAT' view. Each calculations returns an array of values which will be specifically called in the respective analysis views. Nearly all the functions follow a similar pattern of specific conditionals and counting each instance after every point. As more complex conditionals were being developed, the lines of code grew exponentially and the process became tedious.

The code below is a good example of how most of the functions process calculations. The parameters that are set on the functions ask for a player integer and a return index. The return index may cause confusion in terminology. The return index is the parameter for which integer array is desired. Within the views where these functions are called also specifies the index of the return array. More detail about data visualization will be in the next section. Initially the functions would simply ask for a player number, however as the number of return values increased, the array became organized with the addition of the return index.

Like the algorithms described in the early stages of this project, the variables that are being tracked are initialized. Then the function looks in the data base for the rally entity to query certain conditions. In this case the shotResult, shotType, and player attributes are being examined. Since the player attribute is the input of the function, their statistics for each shot type's result is the output.

```
func calculateRally(player: Int, index: Int) -> [Int] {
    var unforcedErrorF = 0
    var unforcedErrorB = 0
    var unforcedErrorA = 0
    var unforcedErrorN = 0
    var forcedErrorF = 0
    var forcedErrorB = 0
    var forcedErrorA = 0
    var forcedErrorN = 0
```

```
    var winnerF = 0
    var winnerB = 0
    var winnerA = 0
    var winnerN = 0
    var inRally:[Int] = []

    for entity in rEntity {
        if (entity.shotResult == 1) {
            if (entity.shotType == 1 && entity.player == player) {
                unforcedErrorF += 1
            } else if (entity.shotType == 2 && entity.player == player) {
                unforcedErrorB += 1
            } else if (entity.shotType == 3 && entity.player == player) {
                unforcedErrorA += 1
            } else if (entity.shotType == 4 && entity.player == player) {
                unforcedErrorN += 1
            }
        } else if (entity.shotResult == 2) {
            if (entity.shotType == 1 && entity.player == player) {
                forcedErrorF += 1
            } else if (entity.shotType == 2 && entity.player == player) {
                forcedErrorB += 1
            } else if (entity.shotType == 3 && entity.player == player) {
                forcedErrorA += 1
            } else if (entity.shotType == 4 && entity.player == player) {
                forcedErrorN += 1
            }
        } else if (entity.shotResult == 3) {
            if (entity.shotType == 1 && entity.player == player) {
                winnerF += 1
            } else if (entity.shotType == 2 && entity.player == player) {
                winnerB += 1
            } else if (entity.shotType == 3 && entity.player == player) {
                winnerA += 1
            } else if (entity.shotType == 4 && entity.player == player) {
                winnerN += 1
            }
        }
    }
    if index == 1 {
        inRally = [unforcedErrorF, unforcedErrorB,
                   unforcedErrorA, unforcedErrorN]
    } else if index == 2 {
        inRally = [forcedErrorF, forcedErrorB, forcedErrorA, forcedErrorN]
    } else if index == 3 {
        inRally = [winnerF, winnerB, winnerA, winnerN]
    }
    return inRally
}
```

3.4.5 Data Visualization

When each point is confirmed, the input data is stored and the metrics are calculated and portrayed in the data summary view which can be accessed from the 'STATS' button in the main view. The data summary view is composed of five tabs: general summary, serve analysis, return analysis, rally analysis, and match timeline. The general summary contains the important metrics that can be indicative to how the match is going for each player. The design of these analytics are meant to be a comparative analysis of the two players. The serve and return analysis view provides a deeper analysis for the deuce and ad side. The rally analysis provides some insight on which shots create the most error or winners. The figures below shows all the tabs in the statistic view.

Much of the statistics of the general summary tab is reflected in Infosys ATP stats shown on their app or website. The general summary can be partitioned into three sections serve, return, and rally in that order of significance. The first serve percentage is shown as an indicator of an advantage of higher win percentage on serve. The second serve percentage is not as important to include in the general summary typically because the context of the second serve percentage is indicative enough of performance. The win percentage of the serves provide important context of serve performance. Ace and double faults are also an indicator of serve performance. The placement of serves provide more useful information to the perspective of the returner. Knowing the percentage of the placement of serves allows the returner to anticipate and recognize a certain tendency towards comfort serves which they can attack. Serve return percentages indicate return performance which directly affects the return win percentages. The return win percentages can be simply thought as the inverse of the opponent's serve win percentage. The rally section, though brief, shows the unforced errors and winners of the match. The amount of unforced errors shows which player is dictating the rally points.

The serve analysis view has a unique way of portraying the deuce and ad side. At the top of the view a button toggles each side statistics to show. When the toggle occurs, a function takes the side and returns an index for a returned array. This view's main purpose is to obtain a better understanding of the patterns and percentages of the serve on both sides with a focus on placement. In a similar regard, the return analysis view also utilizes the deuce and ad button. The return analysis view's main purpose is to provide insight on the return percentages with a focus on forehand and backhand percentages. The rally analysis provides a table of the different shot results specific to each shot. The player mistake value is a measurement of various conclusions dependent on context. The important context to take into consideration is the number of mistakes and the ratio of unforced errors and forced errors. The formula for this metric is the amount of unforced errors divided by the total amount of errors. Many percentages in every analysis view could benefit with some context regarding the amount of points and how the actual match is played to confidently provide some conclusion. The final important stat in the rally analysis is the rally length win percentage. A longer length of a rally can be advantage to the players who identify as a counter-puncher.

Chapter 4

Experimental Results

4.1 Experimental Design

To test validity of the project, the app's statistics are compared to the statistics compiled by Infosys ATP Stats. It is important to note that there are certain metrics in tennis that are subjective calls. Unforced and forced errors are subjective in nature since different perspectives could conclude different results. In most cases these errors are clear and easy to identify, however scenarios where a player's lack of footwork or shot selection can be difficult to confidently assume the error. Another statistics that is not compared is the serve placement. The ATP statistics accessed on a desktop show an in-depth analysis of the serve placement and other statistics that would otherwise be impossible to track by a human. ATP uses high-end camera and AI system to determine the serve position verses that to human estimations is not an acceptable comparison. The statistics that are acceptable and concrete to compare are the following: serve percentage, serve win percentage, ace, double faults, return percentage, and return win percentage. Instead of directly comparing the percentages, the fraction is compared. This compares which points may have been incorrectly input. For example, take the first serve percentage. Using the algorithms described earlier, the first serve percentage is a fraction of first serve made in to the total amount of serves. When a discrepancy is found in the experiment results, the specific deviant values are easily identifiable.

The formula for calculating the the error percentage goes as follows. Looking at the total points win percentage, add the $|NumeratorPT - NumeratorATP|$ to the $|DenominatorPT - DenominatorATP|$ of both players. This would produce the number of points deviant from the desired statistic. Then take that number and divide it to the total points. This percentage should not be read as the absolute accuracy of the experiment, however it provides some kind of relational information about the percentage of points that can produce an error in all the statistics.

The experiment process takes two forms of tracking matches: pre-recorded matches and live matches. Pre-recorded matches are tracked to ensure validity to the calculations in the code. Since the matches are available as a whole the use of pauses, rewinds, and time skips can produce fast and accurate results. When the results are satisfied, live tests are conducted to ensure validity to the usage of the app. Without the ability to rewind or pause, the live tests test the efficiency of the UI to ensure sufficient time when tracking each point.

8 sets will be tracked. The first five sets are pre-recorded matches which will be tracked carefully with pauses and rewinds available. The next three sets are to be tracked as if they were played live. While live matches are the preferred method of evaluating UI validity, they are dependent on which tournament is active. Time zones and scheduling makes it difficult for allocating time to track these matches. Fortunately during the time of this project, the 2022 ATP 1000s Miami Open has time slots that are available to track. The 2022 ATP 1000s Rolex Monte-Carlos Masters are played throughout April 10th-17th which could potentially have matches available to track.

4.2 Evaluation

In the early stages of the application, it was quickly found out that the return percentages were being incorrectly tracked. The percentages were vastly different in the ATP Stats and it was clear that there was some discrepancy. It was also at this point where the evaluation process of comparing the match's statistics had to be reconsidered by looking at the amount of points rather than comparing percentages. For example, if a player had a 25 percent second return percentage and the actual return percentage is 50 percent. It could be the case that the player only had four returns and one was mistakenly input as first serve. There is a 25 percent difference, which is undeniably a significant difference due to the small amount of second serve points.

To calculate the error percentage, the amount of point discrepancy in the total points is compared to the total amount of points played. While this will produce a percentage error, the severity of the error can have vastly different effects on the percentages for each statistic. For instance, by simply not tracking a second serve on a single point not only affects the second serve percentage, but it also affects the first serve percentage and both first and second serve win percentages. This is shown on figure 4.3 The accuracy of the tracking completely depends on the amount and type of errors that the user makes when tracking the match. On average a typical tennis match takes 40 minutes to complete the first set. It is common for a user during this time to make at least one error, and it is unlikely but possible to make more than one error dependent on the user. That being said, a "good" match tracked is one that has an error percentage of roughly less than 10 percent. 10 percent may be a significant difference, however by spreading through multiple different metrics, the percentage is typically a small difference.

Table 4.1 provides an example of a pre-recorded match and a live match. These sets were tracked during the first and third sets of the quarter finals of the Miami Open played by Casper Ruud and Alexander Zverev. The pre-recorded match shows promising results with an error percentage of 9 percent. All the serve statistics were correctly identified and correctly computed by the app. The discrepancy can be found in the serve return win percentage. Looking at the total points In the PT stats, Ruud won 28 out of 43 points, and Zverev won 18 out of 45 points. In the ATP stats, Ruud won 28 out of 46 points, and Zverev won 18 out of 46 points. $|28\text{-}28| + |43\text{-}46| + |18\text{-}18| + |45\text{-}46| = 4$ point differential. Take the 4 points and divide it by the total amount of points and the error percent is 9 percent.

Table 4.1: Table of Stats comparing PT & ATP: Ruud Vs. Zverev

PT/ATP	C. Ruud	A. Zverev	C. Ruud	A. Zverev	Pre- Recorded
Ace	1	3	1	3	
Double Fault	0	0	0	0	
1st Serve Per	17/22	14/24	17/22	14/24	
1st Serve W	16/17	11/14	16/17	11/14	
2nd Serve W	4/5	5/10	4/5	5/10	
1st S R W	3/12	1/16	3/14	1/17	
2nd S R W	5/9	1/5	5/10	1/5	
Total S W	20/22	16/24	20/22	16/24	
Total R W	8/21	2/21	8/24	2/22	Error Percent
Total Pts	28/43	18/45	28/46	18/46	4/46 | 9.00%

For the live set in table 4.2, even better results were calculated. A user mistake can be found in the ace row. PercenTennis tracked no aces for Zverev, while the ATP stats showed that Zverev had made one ace. This user mistake doesn't affect the win percentages in the serve and returns. The only discrepancy that made an impact in the error percentage is the first serve return win percentage. Again looking at the total points row, PT tracked Ruud winning 25 out of 43 points and Zverev winning 18 out of 40 points. ATP stats tracked Ruud winning 25 out of 43 points and Zverev winning 18 out of 43 points. $|25\text{-}25| + |43\text{-}43| + |18\text{-}18| + |40\text{-}43| = 3$ point differential. 3 divided by 43 equals to a 7 error percent.

Table 4.2: Table of Stats comparing PT & ATP: Ruud Vs. Zverev

PT/ATP	C. Ruud	A. Zverev	C. Ruud	A. Zverev	Live
Ace	3	0	3	1	
Double Fault	0	0	0	0	
1st Serve Per	18/24	13/19	18/24	13/19	
1st Serve W	15/18	11/13	15/18	11/13	
2nd Serve W	5/6	3/6	5/6	3/6	
1st S R W	2/13	3/15	2/13	3/18	
2nd S R W	3/6	1/6	3/6	1/6	
Total S W	20/24	14/19	20/24	14/19	
Total R W	5/19	4/21	5/19	4/24	Error Percent
Total Pts	25/43	18/40	25/43	18/43	3/43 | 7%

In reference to 4.3, like mentioned before when certain points are tracked incorrectly, a more significant difference can occur. In this case, a mistake was made tracking the wrong serve number for Tiafoe. In the PT stats, Tiafoe wins 13 first serve points and 11 second serve points. In ATP stats, Tiafoe wins 14 first serve points and 10 second serve points. This simple mistake causes complications in the calculation of the first serve percentage and the first and second serve returns. PT stats tracked Tiafoe winning a second serve point, when in reality Tiafoe had won that point on his first serve. The error percentage for this set is 8.97 percent.

Table 4.3: Table of Stats comparing PT & ATP: Tiafoe Vs. F. Cerundolo

PT/ATP	F. Tiafoe	F. Cerundolo	F. Tiafoe	F. Cerundolo	Pre-Recorded
Ace	1	2	1	2	
Double Fault	0	2	0	2	
1st Serve Per	17/40	29/38	18/40	29/38	
1st Serve W	13/18	16/29	14/18	16/29	
2nd Serve W	11/22	4/9	10/22	4/9	
1st S R W	13/28	5/17	13/29	4/18	
2nd S R W	3/6	11/22	5/9	12/22	
Total S W	24/40	20/38	24/40	20/38	
Total R W	16/39	16/39	18/38	16/40	Error Percent
Total Pts	40/79	36/77	42/78	36/78	7/78 | 8.97%

The next live sets are shown in tables 4.4 and 4.5 for the semifinal round competed by Casper Ruud and Francisco Cerundolo. Similar to the mistake made on Tiafoe's serve, a mistake was made tracking the serve number for Cerundolo. The same statistics were affected for Cerundolo. The difference is that a user mistake had accidentally counted Cerundolo winning the first serve point. When in actuality Cerundolo won a second serve point. The expected discrepancy in the allocation of points can be found by comparing the two data sets. The error percentage for this set is 12.5 percent. Since the total amount of points were lower, the error percentage is expected to be higher.

Table 4.4: Table of Stats comparing PT & ATP: Ruud Vs. F. Cerundolo Set 1

PT/ATP	C. Ruud	F. Cerundolo	C. Ruud	F. Cerundolo	Live Set 1
Ace	2	0	2	0	
Double Fault	1	2	1	2	
1st Serve Per	18/29	22/35	18/29	21/35	
1st Serve W	14/18	11/22	14/18	11/21	
2nd Serve W	4/11	8/13	4/11	8/14	
1st S R W	11/22	4/16	10/21	4/18	
2nd S R W	3/11	6/10	6/14	7/11	
Total S W	18/29	19/35	18/29	19/35	
Total R W	14/33	10/26	16/35	11/29	Error Percent
Total Pts	32/62	29/61	34/64	30/64	8/64 | 12.50%

In the second set of the semi final match, seemingly no user mistakes could reasonably explain the point discrepancies. Upon further examination of this set and previous sets, the mistake was not produced by user error. Instead the error is detected from the difference in which the return win percentage is tracked in ATP stats. PT was not accounting for double faults and aces for this percentage. ATP stats counted aces into the total return points, and double faults as winning the second serve point. This was confirmed by mistake in reference to table 4.2 during the live set. While Ruud's aces are not counted in PT stats, Zverev's missing ace resulted in "accurate" results.

Table 4.5: Table of Stats comparing PT & ATP: Ruud Vs. F. Cerundolo Set 2

PT/ATP	C. Ruud	F. Cerundolo	C. Ruud	F. Cerundolo	Live Set 2
Ace	4	1	4	1	PT Calc Error
Double Fault	2	3	2	3	PT & User Error
1st Serve Per	19/32	12/22	19/32	12/22	User Error
1st Serve W	17/19	6/12	17/19	6/12	
2nd Serve W	4/13	3/10	4/13	3/10	
1st S R W	6/11	2/15	6/12	2/19	
2nd S R W	4/7	7/11	7/10	9/13	
Total S W	21/32	9/22	21/32	9/22	
Total R W	10/18	9/26	13/22	11/32	Error Percent
Total Pts	31/50	18/48	34/54	20/54	15/54 | 27.78%

Refer back to the previous tables and observe the highlighted regions where discrepancies are found. The red highlights represents user made error, the yellow highlights represent PercenTennis return win percentage calculation errors, and the orange highlight represent a mix of the two. Table 4.5 shows an error percentage of 27.78 percent. While this raises high concern to the validity to the project, no user made errors were made in this match.

With a couple lines of code to include ace and double faults, a quick pre-recorded match was tracked. In this quarter final round, Sinner had retired the match after five games. This was the perfect match to track to limit user error. The error percentage for this match is zero. Meaning the calculations in PercenTennis is validated.

Table 4.6: Table of Stats comparing PT & ATP: F. Cerundolo Vs. J. Sinner

PT/ATP	F. Cerundolo	J. Sinner	F. Cerundolo	J. Sinner	Pre-Recorded
Ace	3	1	3	1	
Double Fault	2	0	2	0	
1st Serve Per	15/20	6/13	15/20	6/13	
1st Serve W	11/15	4/6	11/15	4/6	
2nd Serve W	3/5	3/7	3/5	3/7	
1st S R W	2/6	4/15	2/6	4/15	
2nd S R W	4/7	2/5	4/7	2/5	
Total S W	14/20	7/13	14/20	7/13	
Total R W	6/13	6/20	6/13	6/20	Error Percent
Total Pts	20/33	13/33	20/33	13/33	0/33 | 0%

4.3 Threats to Validity

One of the risks with the current version of the app is the absence of an undo button. When tracking live matches, if a user mistake occurs and the incorrect point result is allocated, the whole set is invalid and the match must be reset. The biggest threat to the validity of the application is primarily sourced by user made errors. This can occur through various reasons: a user may have incorrectly identified a point, a user may have incorrectly input the wrong button, or a user may have completely missed a point via a lapse of focus. Whether or not the average user of this app would have a difficult time tracking is hard to gauge. It is believed that through practice on one or two matches, the use of the app is intuitively easy to learn.

Chapter 5

Discussion and Future Work

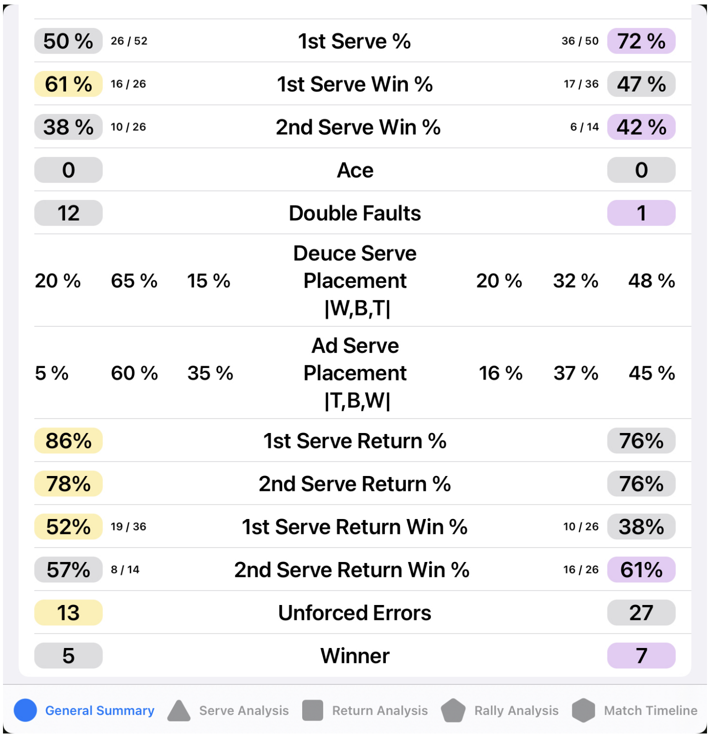
5.1 Summary of Results

It is important to reiterate the usage of this tool. This app is primarily developed for the use for college and junior level players. The research that was accomplished for this app was conducted using professional players as there is a lack data for college and junior level players. This being case, it is safe to assume that the small margin of percentage, especially in serve and return, become increasingly more impactful as the level of the players increase. In ATP matches, every point is analyzed by advanced camera and AI systems, so the accuracy of the statistics is infallible. With generally lower stakes when users of this app track matches, the accuracy matters a little less. Knowing user errors will undoubtedly occur, the small percentage difference is acceptable.

The results provided by the experiment prove to be satisfactory. During the experimental phase the pre-recorded sets was tracked under the utmost highest level of scrutiny to ensure low user mistakes. To find error percentages upwards to nearly 30 percent error caused for alarm to the integrity of the algorithms in PercenTennis. Upon learning the the way ATP stats tracked their points, the user error percentage yielded significantly lower percentages. 5.1 table displays which tracking method was conducted, amount of user errors, the user error percentage, the error percentage, and the total points.

Table 5.1: Table of User Error Percentage

Set	Tracking	User Error	User Error Percent	Error Percent	Total points
1	PR	0	0%	9.00%	46
2	L	1	2.33%	7%	43
3	PR	1	1.28%	8.97%	78
4	PR	2	2.63%	3.95%	76
5	L	1	1.56%	11.54%	64
6	L	0	0%	27.78%	54
7	PR	0	0%	0%	33
8	L	0	0%	0%	64



50 %	26 / 52		1st Serve %		36 / 50	72 %
61 %	16 / 26		1st Serve Win %		17 / 36	47 %
38 %	10 / 26		2nd Serve Win %		6 / 14	42 %
0			Ace			0
12			Double Faults			1
20 %	65 %	15 %	Deuce Serve Placement |W,B,T|	20 %	32 %	48 %
5 %	60 %	35 %	Ad Serve Placement |T,B,W|	16 %	37 %	45 %
86%			1st Serve Return %			76%
78%			2nd Serve Return %			76%
52%	19 / 36		1st Serve Return Win %		10 / 26	38%
57%	8 / 14		2nd Serve Return Win %		16 / 26	61%
13			Unforced Errors			27
5			Winner			7

Figure 5.1: General Analysis View

5.2 Example Analysis

This last section will analyze a set that was tracked during a practice match. Player 1 defeated Player 2 in a set tiebreaker. The analysis process goes as follows: look at the general stats and seek percentages that seem significant, then go to the respective analysis view and examine the metrics to come to a conclusion. Figure 5.1 shows the general analysis of the set. Quickly a few metrics stand out. Concerning serves, player 1 had much lower first serve percentage but also had higher first serve win percentage. Conversely, player 2 has a better second serve win percentage. A huge difference can be found in the amount of double faults. For returns, player 1 performed better in first and second serve percentage as well as the first serve return win percentage. Player 2 had a higher second serve return win percentage. This can be contributed to the amount of double faults player 1 had hit. Another big difference can be found in unforced errors where player 2 has more than double the amount of player 1.

The serve analysis views are shown in figures 5.2 and 5.3. On the deuce side, player 2 had better first and second serve percentages, but player 1 performed better first and second serve win percentages. The same trend is found on the ad side. However, when compared to the stats generated by the general analysis, player 2 had overall better second serve win percentage. A reason may be due to the fact that player 2 hit much less second serves. Overall player 1 had a higher total serve points won than player 2.

The return analysis views are shown in figures 5.4 and 5.5. The return analysis offer an interesting analysis. First on the deuce side, player 2 has a very high first serve return percentage, but low second serve return percentage. This could mean that player 2 is taking higher risk in the second serve return. Player 1 also has a high first

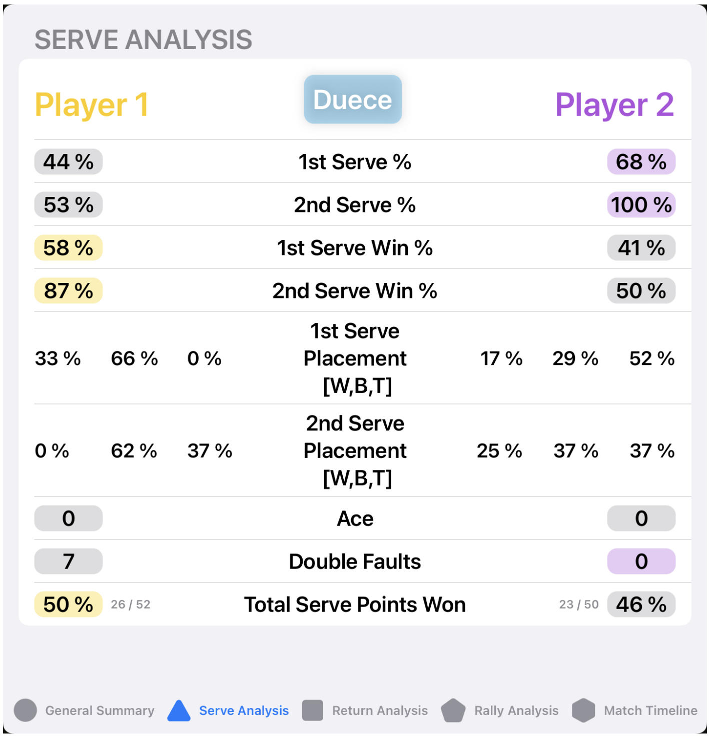


Figure 5.2: Deuce Serve Analysis View

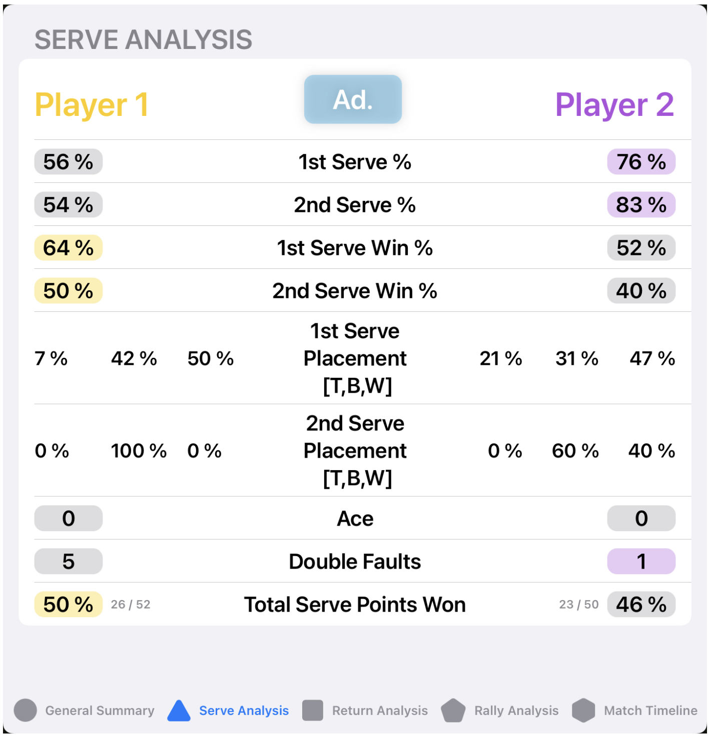


Figure 5.3: Ad Serve Analysis View

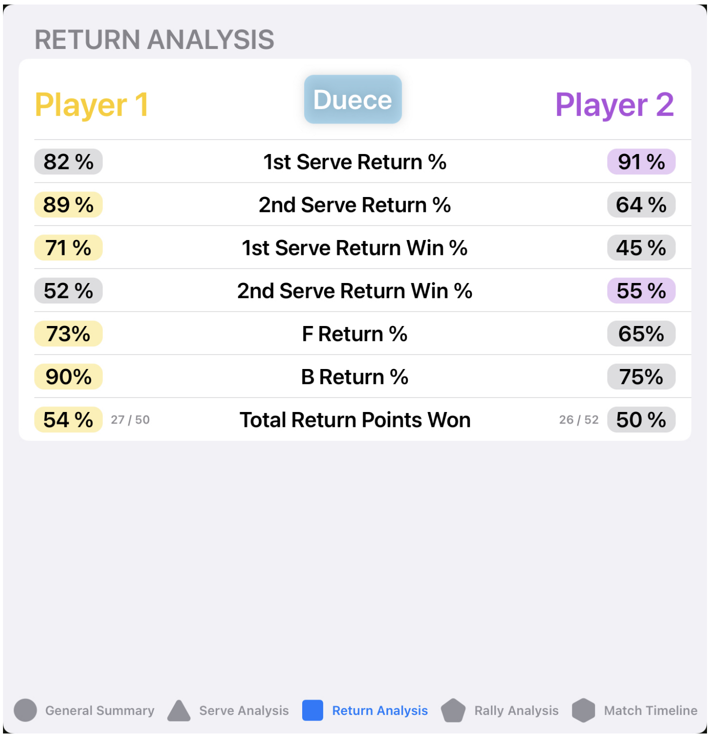


Figure 5.4: Deuce Return Analysis View

serve return percentage, but additionally has a high second serve return percentage. Player 1 performed much better in first serve return win points than player 2. Player 2 performed slightly better in second serve return win points. In this case, player 2's higher risk second serve returns are not worth it in regards to win percentages. On the ad-side, a massive difference in first serve return percentage can be observed. Second serve returns on this side remains to be fairly equal. Player 1 overall performed better on the ad-side compared to player 2. Player 1 had a higher total return points won than player 2.

The rally analysis view is shown in figure 5.6. Here is where the most significant difference can be found. Player 2 had made a lot of forehand unforced errors. Most of the points lost by player 2 can be found in the medium length rally. 79 percent of 34 errors were unforced errors made by player 2, while 59 percent of 22 errors were unforced errors made by player 1. This is where majority of the points were lost by player 2.

The questions to be asked are why and how did player 1 win? What can each player do to perform better in future matches? Player 1 won because of their the high return percentage which helped them get into the rally where player 2 performed poorly. The obvious advice for player 1 would be to reduce the amount of double faults and raise the first serve percentage. This would allow player 1 to have more opportunities to play out the point to the rally where they were performing better. The advice for player 2 is a little more nuanced. First, the amount of unforced errors in forehand must be reduced. The forehand was clearly a strong weapon which can be found in the winners section in figure 5.6, but losing 15 points for 5 winners is not worth the risk. Second, the serve could have had a bigger impact. High serve percentages is usually a

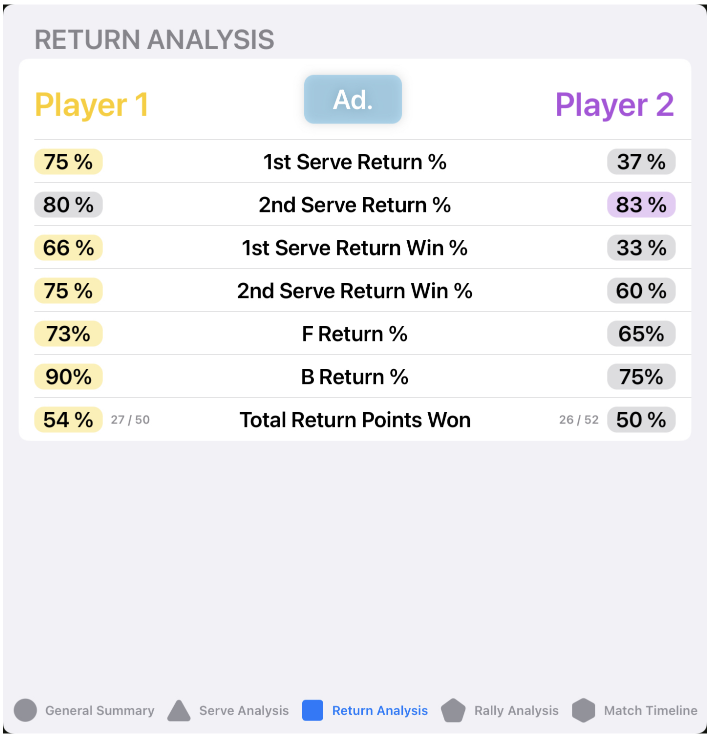


Figure 5.5: Ad Return Analysis View

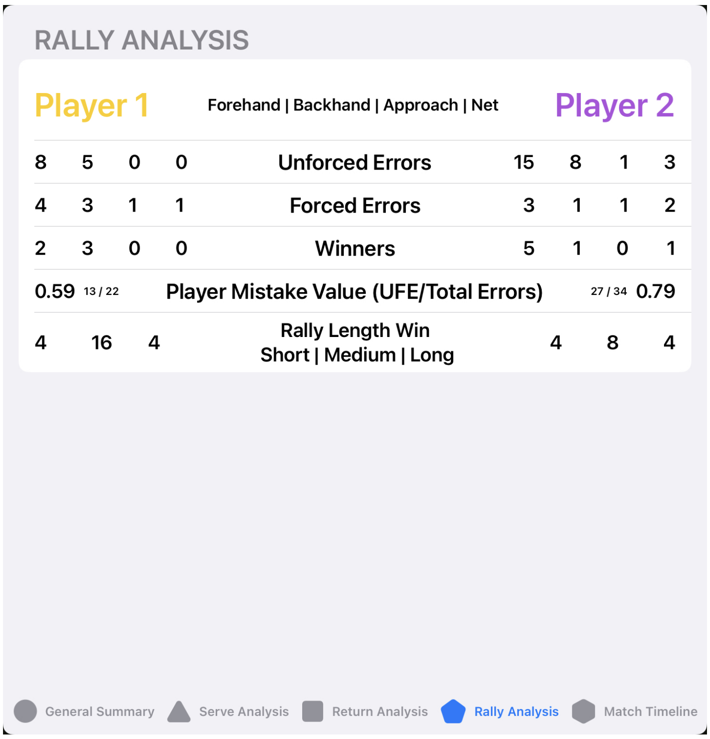


Figure 5.6: Rally Analysis View

good thing, but if no pressure is applied from the first serves, then taking bigger risks could contribute to higher first serve win percentages. Third, similar to the previous advice, the serve could have had "riskier" placement. The first serve could have been played closer to the lines which would be a more difficult serve to return. In general analysis in figure 5.1, most serves were hit towards the backhand where Player 1 had a higher percentage return than their forehand. It is important to note that in figures 5.4 and 5.5, the forehand and backhand percentage is the total of both sides and does not portray different metrics for each side. This is one of the tweaks that would be changed so that the percentages will account for each side. It could be argued that because most of the serves were hit to the backhand Player 1 had more of an opportunity to return off the backhand, but when a player is returning 90 percent of serves from one side, a change in serve strategy should be formed. Whether that is to increase speed or play closer to the lines. This is one of many ways that PercenTennis aims to clarify strengths and expose weaknesses for players.

5.3 Future Work

The vision for this app had ambitious goals for a quality tennis tracking app. While goals which were established for this project have been met, there is room for improvement in user experience, data visualizations, and clearer data portrayal. However, the immediate actions for the future development of this project is to tackle the issues and bugs that directly affect the usage of the app. There remains a bug where the server does not correctly switch when a set is won with the loser having less than 4 games. While this is a major problem as a product itself, the experiment results will not be compromised as they are tracked by sets.

In terms of user experience the much needed feature to undo the last point will be a great implementation. This will follow a similar approach to the 'New' button. Instead of deleting all the data, this button will simply clear the last entry. The 'New' button will also have added features. This button will display a notification to confirm either to save or clear the existing data and start a new match in the 'New Match' view. This view will still allow users to add names to the players. They can also customize the scoring format for the rules they wish to follow. The remaining scoring rules are 8-game pro sets, 10-point tie-breakers, and no ad scoring. The scoring code would require refactoring to accommodate the other scoring rules and allow for flexibility. Another notable feature is the ability to save completed matches and look at the statistics on hand. Along with saving entire matches, saving each set adds another level of match analysis. This would require a deeper knowledge in the use of core data.

Considering data visualization, the lack of graphics in the app was disappointing to what the proposal had originally envisioned. However, bar graphs may not be the most optimal method of tennis analysis. Presenting the percentages and fractions to compare one another had more-or-less held the same information. ATP stats use small arrows to show which player had better statistics in comparison. PercenTennis will instead highlight the better stat. This approach provides a subtle emphasis on the percentages. The use of pie charts for serve placements would be another great use of data visuals. Figure 5.7 shows a pie chart generated by Google Sheets during the preliminary testings

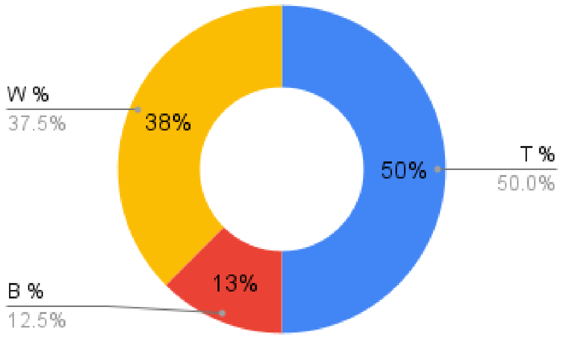


Figure 5.7: Pie Chart of Serve Placement

of the algorithms. When the percentages are linearly placed for serve placement, it takes more effort to comprehend the each percentage and placement. A serve placement pie chart would be a better representation of how often a player serves to certain locations. The most ambitious addition would be an update to the match timeline view. Currently this view displays a hideous and impossible to comprehend table of each point tracked. Instead, a line graph of the points progressing on the x axis with a drop window with general summary stats for the y axis can provide an insightful comparison between two players and a trend throughout the set.

In regards to clearer data portrayal, an initial clean up with how the calculations functions are written is necessary. All functions will use the return index that are used in bigger functions. This way when adding new data to the analysis views, figuring out the index becomes much easier to manage. It was stated earlier that much of the analysis of the match requires context, and that can be accomplished by adding to all the percentage metrics their representative fraction. Another statistic that will benefit with added context is the rally length win percentage. There will be a line with three sections to indicate short, medium, and long rallies. On top of the segmented line is the higher percentage in the color of the player.

There are numerous amounts of minor tweaks in regards to code formatting, visuals, and data management. As software development skills are enhanced through practice, research, and time, the end goal for this project is to scrap major portions of the code and start once more from scratch with the new gained knowledge. If an opportune circumstance appears, there will be further development with the proposed implementations.

Bibliography

[1] Abidin, A. W. Z., and Ruslan, N. A. S. Exploring the importance of players' characteristics and performance on serve and return of serve in winning the women's singles grand slam tennis tournaments. *Journal of Physics: Conference Series 1496*, 1 (mar 2020), 012008.

[2] Barnett Tristan, S. C. Using microsoft excel to model a tennis match. *Swinburne Research Bank No. 1: 63-68* (2002).

[3] George, S. L. Optimal strategy in tennis: A simple probabilistic model. *Journal of the Royal Statistical Society. Series C (Applied Statistics) 22*, 1 (1973), 97–104.

[4] Gray, C. Game, set and stats. *Significance 12*, 1 (2015), 28–31.

[5] Hizan, H., Whipp, P., and Reid, M. Comparison of serve and serve return statistics of high performance male and female tennis players from different age-groups. *International Journal of Performance Analysis in Sport 11*, 2 (2011), 365–375.

[6] Klaassen, F. J. G. M., and Magnus, J. R. Are points in tennis independent and identically distributed? evidence from a dynamic binary panel data model. *Journal of the American Statistical Association 96*, 454 (2001), 500–509.

[7] Ma, S.-M., Liu, C.-C., Tan, Y., and Ma, S. Winning matches in grand slam men's singles: An analysis of player performance-related variables from 1991 to 2008. *Journal of sports sciences 31* (03 2013).

[8] Mecheri, S., Rioult, F., Mantel, B., Kauffmann, F., and Benguigui, N. The serve impact in tennis: First large-scale study of big hawk-eye data: The serve impact in tennis. *Statistical Analysis and Data Mining: The ASA Data Science Journal 9* (06 2016).

[9] Newton, P. K., and Keller, J. B. Probability of winning at tennis i. theory and data. *Studies in Applied Mathematics 114*, 3 (2005), 241–269.

[10] Perin, C., Vuillemot, R., Stolper, C. D., Stasko, J. T., Wood, J., and Carpendale, S. State of the art of sports data visualization. *Computer Graphics Forum 37(3): 663-487* (2018).

[11] Pollard, G., Pollard, G., Barnett, T., and Zeleznikow, J. Applying tennis match statistics to increase serving performance during a match in progress. *Medicine and Science in Tennis 14*, 3 (2009), 16–19.

[12] Roure, C. What are the key points to win in tennis? *The Official Tennis Coaching and Sports Science Issue 64: 14-15* (2014).

[13] Stephanie A Kovalchik, M. R. Comparing matchplay characteristics and physical demands of junior and professional tennis athletes in the era of big data. *J Sports Sci Med 16(4): 489–497* (2017).