1 Summary

Among the fans of any sport, the most heated debates occur while discussing who the Greatest Of All Time (G.O.A.T.) in the sport is. We define the G.O.A.T. of a sport as the most skilled, capable, consistent, and popular sportsperson that the sport has ever seen. Each of these can be determined using multiple different metrics.

By creating models, we aimed to objectively determine who the G.O.A.T. of any sport is. In order to do so, we categorised sports into individual and team sports. To simplify the modelling process, we first found the greatest player for an individual sport within a specific time period. For this, we used Women's Tennis in 2018, considering competition success, popularity, consistency as well as the players' official ranking.

Once we found the greatest Women's Tennis player for 2018, we expanded the model to find the G.O.A.T. for an individual sport. For this, we used Golf and considered the number of Majors won, the total number of wins as well as the career span. We disregarded popularity as it is difficult to measure for players across eras without using social media following since some players may have played in times before the advent of social media platforms.

Following this, we looked into team sports such as Basketball, where we considered how a player contributes to team success, as well as the factors mentioned above.

Our results are as follows:

- Simona Halep was the greatest female Tennis player in 2018
- Jack Nicklaus is the G.O.A.T. of Golf
- Michael Jordan is the G.O.A.T. of Basketball

We then generalised our models to be used for any individual or team sport, and mentioned how to incorporate factors into the models as needed.

Towards the end of the paper, schematic diagrams, code, and data sources are provided to show how we arrived at our models and to allow for easy understanding.

Contents

2 Introduction

Among the fans of any sport, the most heated debates occur while discussing who the Greatest Of All Time (G.O.A.T.) in the sport is. We define the G.O.A.T. of a sport as the most skilled, capable, consistent, and popular sportsperson that the sport has ever seen. Each of these can be determined using multiple different metrics. Before going into the modelling, we narrowed down the types of sports to Individual Sports and Team Sports.

Definition of Individual Sports An individual sport is a sport where an individual competes against one or more other players either "one-on-one" or against an inanimate standard set by competitors, timings or rankings. Competitors do not necessarily physically interact with others. Winners in competitions for these sports are the individuals themselves.

Definition of Team Sports A group of individuals compete against another group of individuals and the competition includes multiple interactions between players physically, strategically, or through equipment. Winners in such competitions are the teams as a collective unit and not the individuals themselves.

Based on these definitions, we started identifying general assumptions for all sports.

3 General Assumptions

In this paper, the general assumptions that are made are as follows:

1. Everyone is healthy throughout the duration of their careers
2. Rules of the game are kept constant throughout time
3. Data cited is accurate
4. Advancements in training methods and Sports Science are based on the training habits and styles of the great players of the past
5. Modern game analysis techniques are based on the best games played by past greats.

Assumption 1: Everyone is healthy throughout their careers.

The first assumption is important as it allows us to compare different players against each other by assuming that their careers are uninterrupted throughout. This is essential when making effective comparisons as we can now focus on the achievements of the players during the time they were active without worrying about differences in circumstances they faced.

Assumption 2: Rules of the game are unchanged.

Changes in the rules would result in differences in the gameplay, hence affecting the performances of the players. This leads to an unfair comparison. Thus making this assumption key as it allows us to compare players over different eras.

Assumption 3: Data cited is accurate.

The third assumption is warranted as it allows us to use and manipulate the data cited to properly and reliably determine the G.O.A.T. of a sport. Thus, we made sure to take our data from the official websites of the Womens' Tennis Association (WTA), National Basketball Association (NBA) and the Professional Golfers Association (PGA), which are the organisations in charge of the sports that we used as examples. In addition to this, we used the data provided in the attachment to the problem statement, which we assume is reliable.

Assumption 4: Advancements in Sports Science and Technology are based on past greats.

This is a justifiable assumption as the difference in technology available to players often affects their performance. In the cases of injuries, access to advanced technology can reduce the impact of previously career-ending injuries, allowing players to reach greater heights in the time bought. Almost all of the advancements in Sports Science and training methods are dependent on the methods employed by the great players in the past, making this a valid assumption.

Assumption 5: Modern game analysis techniques are based on games played by past greats.

This is another necessary assumption as the analysis techniques have changed over time. We assume that analysis techniques are based on the best games of the greatest players of the past. As a result, making this assumption allows us to compare players across eras.

4 Key Factors and Variables

We decided that the key factors that determine the greatness of a player include their success in competitions, official ranking, popularity and their win rate.

4.1 Success in Competitions

The greatness of a player can be determined by their skill. As such, we need to find a way to objectively quantify a player's skill. To do this, we can look at how far they have progressed in competitions. Since a player with superior skill defeats more opponents, they will progress further in the competition and

possibly win the title. With this, we notice that the greatness of a player is proportional to their success in competitions.

4.2 Official Ranking

Another way to measure a player's skill would be to look at their official ranking from their respective official sports associations, which is an indicator of how great the player is recognised as by the respective associations. However, since this is based on the association, it may not agree with the views of the general public. This factor is not the only factor that quantifies a player's skill. Nonetheless, the lower a player's ranking is, the more skilled they are perceived to be, with rank 1 being the top ranked player. Therefore, the greatness of a player can be said to be inversely proportional to their official ranking.

4.3 Popularity

The impact of a player on the public can be quantified by their popularity. In the modern-day and age, we can measure this by looking at their social media following. As their popularity amongst the general public increases, their social media following increases. However, it is important to note that when considering the greatest player of all time, we must take into account people who played before the advent of modern mass communication technologies. As such, the size of a player's following can only be used for determining the popularity of a player in a period that is after the invention of social media.

4.4 Win Rate

Another factor that makes a player great would be their consistency. This can be seen in their win rate. Win rate can either be defined as the proportion of wins out of the total number of games played or the total number of wins per unit time in their overall career length. It shows us how consistently player defeats his opponents in the sport. Hence, it can be said that the higher the win rate, the greater the player.

4.5 Definition of Variables

The variables used in this report, as well as their representation are as follows:

Variables	Meaning
P	Prize Money (million £)
F	Following on Instagram (millions)
W_p	Win Percentage
W	Number of Wins
U	Average Number of Games lost per Set
R	Ranking
M	Number of Golf Majors
T	Career Span (Years)
S	Number of Points scored
A	Number of Assists
E	Number of Steals
B	Number of Blocks
GP	Number of Games Played
C	Number of Championships Won
G_x	Greatness Constant of player x
GI_x	Greatness Index of player x

Note: G_x and GI_x are the only dependent variables, all other variables are independent.

5 Individual Sports

5.1 Women's Tennis 2018

In Tennis, the greatest player should be the individual who is most skilled, consistent, and popular among the public. This can be measured using various metrics such as the win rate of the player, WTA world ranking, popularity, and how far a player progresses in the Grand Slams, which are the main Tennis tournaments.

5.1.1 Specific Assumptions

For Tennis, the specific assumptions made are as follows:

1. Value of Currency remains constant
2. Supporters follow player on Social Media (Instagram)
3. Best players make it to the Round of 16

Specific Assumption 1: Value of Currency was constant

The first specific assumption is critical so that we can compare the different prize money values over the whole year without accounting for changes in the currency's value due to inflation or otherwise. Doing so allows us to make better comparisons as we can now directly compare the prize money won in different competitions.

Specific Assumption 2: Supporters follow player on Social Media

The second specific assumption is imperative because it allows us to quantify the following and popularity of a player. Our choice for social media platform is Instagram. Comparing between the top 3 Social Media websites, which are Facebook, Twitter and Instagram, we ruled out Twitter as it has a low user base of about 191 million. While Facebook has a larger user base of about 2.5 billion, many Tennis players do not have public Facebook accounts. Hence, we decided to use Instagram as it has a large user base of just over 1 billion people and most Tennis players have public Instagram accounts, allowing for a fair comparison.

Specific Assumption 3: Best player make it to the Round of 16

The third specific assumption was made as we only used the data provided which was about the matches from the Round of 16 onwards in all 4 grand slams.

5.1.2 Win Rate

Consistency is an important factor that affects the greatness of a player. When a player is deemed to be consistent, they are perceived to perform at their best regularly. If a player with talent and skill that is greater than the opponent consistently performs to the best of her abilities throughout the game, she, in most cases, will be the winner. This implies that when a player shows greater consistency overall, she wins more games, resulting in a greater overall win rate (W_p). As aforementioned, consistency is key to greatness. Hence, there is a positive relationship between win rate and greatness, leading us to the relationship $G_x \propto W_p$.

5.1.3 Average Number of Games Lost per Set

In Tennis, looking at the average number of games lost per set also gives us an indication of how consistent and skilled a player is in defeating their opponents in every single game. This allows us to determine how dominant a player is on the court over the period of the match. Hence, we see that the relationship between the average number of games lost per set (U)and the greatness of the player is negative. Therefore, we arrive at the conclusion that $G_x \propto \frac{1}{U}$

5.1.4 WTA World Ranking

The WTA World Rankings are based on a 52-week cumulative system, determined by players' results in tournaments. As such, the lower the ranking of a player (R), the more skilled they can be said to be as they have been able to win more matches against their competitors. Hence, they can be said to be greater than the other competitors. Thus, the lower the ranking, the greater the player, giving us the relationship $G_x \propto \frac{1}{R}$.

5.1.5 Prize Money

When a player progresses further in a tournament, they will receive higher prize money (P) as opposed to competitors who did not progress as far. Thus, the more prize money a player receives in a tournament, the more successful in the competition they can be said to be, which in turn makes them greater than their opponents. From here, we see that the more prize money one wins in a tournament, the greater the player. Thus, we can say that $G_x \propto P$.
Note: We used prize money data for games from only the Round of 16 onwards for each Grand Slam.

5.1.6 Popularity

The greatness of a player can be said to be directly proportional to their popularity, as the more popular a player is, the more people consider them as 'great'. Hence, we need a way to quantify the following of a player (F), which can be done by taking their social media following based on the assumption that the supporters of an athlete will follow said player on social media, which will provide a good representation of how popular a player is, and thus how 'great' they are. We can therefore say that $G_x \propto F$

5.1.7 The Model

Combining the proportionalities from above, we get the equation:

$$G_x = \frac{PFW_p}{UR}$$

Since we want to compare this greatness to that of other players to find the greatest player, we can simply take the ratio:

$$GI_x = \frac{G_x}{\sum G_x},$$

which gives us the probability that a random fan will say that player x is the greatest female Tennis player for the 2018 season. As such, the player with the highest GI value can be said to be the greatest player.

5.1.8 Results

We used the code found in the Appendix to determine the greatest player for the 2018 season. From the code, we generated the following table which shows the variables used in the model, the greatness constants, as well as the greatness index for each player.

Name	Prize Money	Followers	Win Percentage	Avg. Games Lost Per Match	Ranking	Greatness Constant (Using Prize money in £100000)	Greatness Index
SIMONA HALEP	3.090	1.600000	0.875	3.650	1	1185.205479	0.502846
SERENA WILLIAMS	2.664	12.900000	0.666	3.333	16	429.183348	0.182089
NAOMI OSAKA	3.039	2.100000	0.800	3.545	5	288.040621	0.122207
CAROLINE WOZNIACKI	2.515	1.400000	0.800	4.167	3	225.325974	0.095599
ANGELIQUE KERBER	3.093	0.651000	0.777	4.167	2	187.727731	0.079647
SLOANE STEPHENS	1.349	0.500000	0.667	3.615	6	20.741886	0.008800
ELINA SVITOLINA	0.459	0.716000	0.333	4.000	4	6.839903	0.002902
MADISON KEYS	1.457	0.199000	0.625	3.938	17	2.706874	0.001148
MARIA SHARAPOVA	0.538	4.100000	0.333	6.000	29	4.221445	0.001791
GARBINE MUGURUZA	0.492	0.822000	0.667	3.750	18	3.996308	0.001696
KAROLINA PLISKOVA	0.782	0.398000	0.400	5.182	8	3.003049	0.001274

Figure 1: GI for Women's Tennis 2018

After these first 11 players, none of them had a greatness index of more than 0.001. From Figure 1, it is clear that Simona Halep is the greatest player of Women's Tennis in 2018 with the highest GI value of 50.3% to 3 significant figures (s.f.).

5.1.9 Justification

This result is justified through this model because Simona Halep has the highest win percentage, prize money and best ranking among the competitors. This shows that she was the most consistent player throughout the year with her outstanding performance, and is thus deserving of the title of Greatest Women's Tennis player in 2018.

Moreover, our results are consistent with Sportskeeda's List of Top 5 Women's Tennis Players of 2018. They take into account their consistency in playing well throughout the year in different Grand Slam tournaments and the challenges they faced in doing so. The Tennis players they ranked as the top 5 in the year 2018 fall within the top 8 players in the ranking from our model. Hence, it is shown our model is accurate.

5.2 Golf

Similar to Tennis, the G.O.A.T. of Golf should be the most skilled and consistent. Since we are dealing with the greatest player of all time, it is not possible to simply measure Instagram followings to understand the effect of every professional player's popularity as we need to consider the players who were active before the invention of social media. As such, this factor has been excluded.

In addition, another important factor that determines the greatness of a player would be the number of majors they have won in the span of their career. This is because if a player has more Golf majors, they are seen as the best out of all their competitors, which makes them greater. Since we are dealing with the G.O.A.T., we need to consider their career as a whole, and compare that against those of other players. As such, it is not as effective to compare the overall ranking of players since many of them have already retired. In addition, win rate cannot be effectively considered as the player could have played professionally for a long time and as a result played a large number of games.

Hence, we are tweaking our model used for Women's Tennis in 2018 to account for the points mentioned above.

5.2.1 Specific Assumptions

For Golf, the specific assumption made is as follows:

1. The greatest players have made it to the top 10 highest number of PGA tour wins

5.2.2 First Iteration of the Model

Based on the factors that make a Golf player great, we see that the more Majors (M) a player wins, the greater they are. Hence, $G_x \propto M$.

If a player wins more games (W), they are seen as more skilled and thus greater than their opponents. Thus $G_x \propto W$.

If a player reaches a certain number of majors and wins in a shorter time span (T), it is seen as more impressive and thus greater. Therefore $G_x \propto \frac{1}{T}$.

Combining the above proportionalities, it would seem that the model to use should be $G_x = \frac{MW}{T}$.

5.2.3 Sensitivity Analysis and Further Refinement

However, upon sensitivity analysis, we realised that changing M while keeping W and T constant did not produce a large enough change in G_x as opposed to changing W and keeping M and T constant. This is less ideal because majors

are more difficult to achieve than a win, as such there needs to be more emphasis on the majors. We then realized that simply squaring the number of majors produced a large enough emphasis on them. Thus, the suitable model to be used in finding out the G.O.A.T. of Golf is given by:

$$G_x = \frac{M^2 W}{T}$$

$$GI_x = \frac{G_x}{\sum G_x}$$

Using the equations stated above, we can find the greatness index for every Golf player, and find the G.O.A.T. of Golf.

5.2.4 Results

Taking the top 11 players of all time based on PGA tour wins, we ran the code found in the Appendix to determine the G.O.A.T. of Golf. From the code, we generated the following table which shows the variables used in the model, as well as the respective greatness constants and greatness indices for the various players.

Name	Wins (num)	Majors (num)	Career Span (yrs)	Greatness Constant	Greatness Index
Jack Nicklaus	73	18	25	946.080000	0.344778
Tiger Woods	82	15	24	768.750000	0.280154
Walter Hagen	45	11	23	236.739130	0.086274
Ben Hogan	64	9	22	235.636364	0.085872
Arnold Palmer	64	7	19	165.052632	0.060150
Sam Snead	82	7	30	133.933333	0.048809
Tom Watson	39	8	25	99.840000	0.036384
Byron Nelson	52	5	17	76.470588	0.027868
Phil Mickleson	44	5	29	37.931034	0.013823
Billy Casper	51	3	20	22.950000	0.008364
Cary Middlecoff	39	3	17	20.647059	0.007524

Figure 2: GI for Golf

From the table, it can be seen that Jack Nicklaus is the G.O.A.T. of Golf with a GI value of 34.5% to 3s.f..

5.2.5 Justification

This result is justified as Jack Nicklaus has the most majors, as well as the second highest number of wins, while doing so in an average time period,

which shows his consistent outstanding performance in his career, making the G.O.A.T. of Golf.

In addition, our results are consistent with a ranking by Golfnews.co.uk, which ranked their top 4 players as Jack Nicklaus, Tiger Woods, Ben Hogan and Walter Hagen. Comparing these to our results, we realise that Ben Hogan and Walter Hagen have switched places, which is understandable because their GI_x values are 8.59% and 8.63% respectively, to 3s.f.. These are within 0.04% of each other, which is a negligible difference. As such our model is accurate and reliable in finding the G.O.A.T. of Golf.

6 Team Sports

The G.O.A.T. of a team sport has all the characteristics mentioned previously and a few more. We consider Skill, Consistency, Success, Team Success and ability to elevate the performance of his teammates as the important characteristics. We decided to use Basketball as the sport to model due to the availability of complete statistics and the fact that it allows players to incorporate elements of all the characteristics mentioned above in their game.

6.1 Basketball

In Basketball, the main facets of the game are Offense and Defense.

Offense Offense is the scoring of points by the team as a whole and relies on the offensive skill of the players. We can tell the offensive impact of an individual player on the games they have played by looking at the points and assists that they have racked up over the course of their careers. Points (S) show us the overall scoring of a player. Assists (A) show us how many times the player passed to his teammates who then scored because of the pass within the same possession. These two cover both the possible aspects of Offensive contributions by a single player.

Defense Defense refers to the team doing their best to prevent their opponents from scoring. The defensive impact of the players can be measured using blocks and steals. Blocks (B) are when the player directly stops the ball from entering the basket by deflecting it. Steals (E) refer to when the player retrieves the ball from the opponents during play, resulting in the opposing team being unable to score in that possession.

It must be noted that the career totals for each of the categories is dependent on the number of games played by the player (GP). The players who are consistently great should require lesser games to put up the same statistics than if they were less consistent. Thus, it makes sense to take the number of games played into consideration as well.

In addition to individual achievements of the player, it is necessary to look at the overall success of the teams that the player has played on and the impact that his team has made in general. The best way to quantify team success is to look at the number of NBA championships (C) the team has won as championships are what differentiates the best teams from the good teams.

6.1.1 Specific Assumptions

For Basketball, the specific assumption made is as follows:

1. The players most likely to be the G.O.A.T. have made it to our list of the top 10 players.

6.1.2 First Iteration of the Model

Based on what has been said earlier, we arrive at a series of proportionalities that allow us to develop our model. These are:

$$G_x \propto C$$

$$G_x \propto (S + A + E + B)$$

$$G_x \propto \frac{1}{GP}$$

Combining these proportionalities, we get the equation:

$$G_x = \frac{C(S + A + E + B)}{GP}$$

However, using this equation would be erroneous when no championships have been won by the player as their G_x, and as a result, their GI_x would be 0, which may not be an accurate comparison as a lot of players have not won a championship. Hence, the model should be adjusted to include $C + 1$ instead of C so that there is a positive value of $G_x \ \forall \ C \in \mathbb{N}_0$. Leading us to the equation:

$$G_x = \frac{(C + 1)(S + A + E + B)}{GP}$$

After this, the GI_x is calculated using the formula:

$$GI_x = \frac{G_x}{\sum G_x},$$

which shows the probability that a random Basketball fan would say that player x is the G.O.A.T.

6.1.3 Sensitivity Analysis and Further Refinement

Upon conducting a sensitivity analysis on the first iteration of the model, it was found that the results of the model are extremely sensitive to changes in the total number of championships won by a player. Any increase to the number of championships won will result in a large increase in the calculated G_x. In order to reduce the effect of small changes of C on the G_x calculated, we decided to take the square root of $C+1$ as that maintains the importance of championships won in the model while reducing the excessive sensitivity of the model to it.

Hence, the model is now refined to be:

$$G_x = \frac{\sqrt{C+1}(S+A+E+B)}{GP}$$

6.1.4 Results

Using the NBA website, we compiled a list of 10 players most likely to be the greatest of all time. The 10 players we considered were Michael Jordan, Kareem Abdul-Jabbar, LeBron James, Magic Johnson, Kobe Bryant, Shaquille O'Neal, Larry Bird, Stephen Curry, Hakeem Olajuwon, and Karl Malone. We then used the code found in the Appendix to determine the greatness indices for these 10 players. We got the following result for each player in the list.

Name	Games Played	Points	Assists	Blocks	Steals	Championships Won + 1	Greatness Constant	Greatness Index
Michael Jordan	1072	32292	5633	893	2514	7	95.804827	0.143329
Kareem Abdul-Jabbar	1560	38387	5660	3189	1160	7	80.111993	0.119852
LeBron James	1302	35189	9636	979	2057	5	78.664253	0.117686
Magic Johnson	906	17707	10141	374	1724	6	76.301876	0.114152
Kobe Bryant	1346	33643	6036	640	1944	6	73.373683	0.109771
Shaquille O'Neal	1207	28596	3026	2732	739	5	63.643645	0.095214
Larry Bird	897	21791	5695	755	1556	4	62.967670	0.094203
Stephen Curry	736	17387	4854	159	1254	4	60.869565	0.091064
Hakeem Olajuwon	1238	26946	3058	3830	2162	3	47.336193	0.070818
Karl Malone	1476	36928	5248	1145	2085	1	29.350271	0.043910

Figure 3: GI for Basketball

It is observed that Michael Jordan is the G.O.A.T. of Basketball as he has the highest GI value of 14.3% to 3s.f..

6.1.5 Justification

Our results are consistent with Bleacher Report's List of Top 50 NBA players of all time. Bleacher report took into account more advanced statistics such as box plus/minus (available from the 1973-74 season on) and win shares per

48 minutes. Pace and playing-time-adjusted numbers were also used to determine the G.O.A.T. of NBA Basketball. The top 4 players on our greatness index: Michael Jordan (14.3%), Kareem Abdul-Jabbar(12.0%), LeBron James (11.8%), and Magic Johnson (11.4%) are the same with the Bleacher report list albeit with LeBron James and Kareem Abdul-Jabbar switching positions. This is understandable as the difference in greatness index between Kareem Abdul Jabbar and LeBron James is negligible at about 0.2%. Hence, it shows that our model is accurate in determining that Michael Jordan is the G.O.A.T. followed by Kareem Abdul-Jabbar and LeBron James.

7 Generalisation of Models

7.1 Individual Sports

For individual sports as we saw with Golf earlier, we can find the greatness constant G_x of player x by using the number of championships or equivalents won, the length of their careers and the win percentage or win totals. The championship totals in the model can be any high level competition wins that are sought after in the sport and considered the highest achievement available as in the case of Majors in Golf. The length of the career can be measured in games played or years active if necessary. More factors can be included as per the convenience of the users by considering the relationship between greatness and the factor that is to be included. A basic model for this is given below.

$$G_x = \frac{C^2 W}{T}$$

Where,

$$G_x \propto \frac{1}{T},$$

$$G_x \propto C^2,$$

$$G_x \propto W$$

From here, we can calculate the Greatness Index:

$$GI_x = \frac{G_x}{\sum G_x},$$

which gives the probability that a random person would say that player x is the G.O.A.T. of the sport.

7.2 Team Sports

Compared to individual sports, in team sports, deciding who the G.O.A.T. is depends on an extra factor, which is team success. We can quantify the different components making the G.O.A.T. by looking at their total important career statistics, the number of championships or equivalents won as a measure of team

success, and the number of games played to show the pace and consistency of the player. Like the Individual Sports model, the users can add factors as they wish by considering the relationships between those factors and the Greatness of the player. Keeping the process of creating the Basketball model in mind, we arrive at the model:

$$G_x = \frac{\sqrt{C+1}\sum \text{Career Totals}}{GP}$$

Where,

$$G_x \propto \sqrt{C+1},$$

$$G_x \propto \sum \text{Career Totals},$$

$$G_x \propto \frac{1}{GP}$$

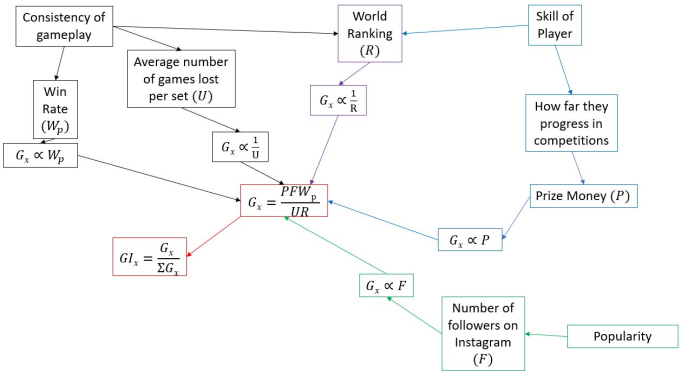
As with all the other models in this report, from the greatness constant, we use the following formula to find the greatness index of the player x:

$$GI_x = \frac{G_x}{\sum G_x}$$

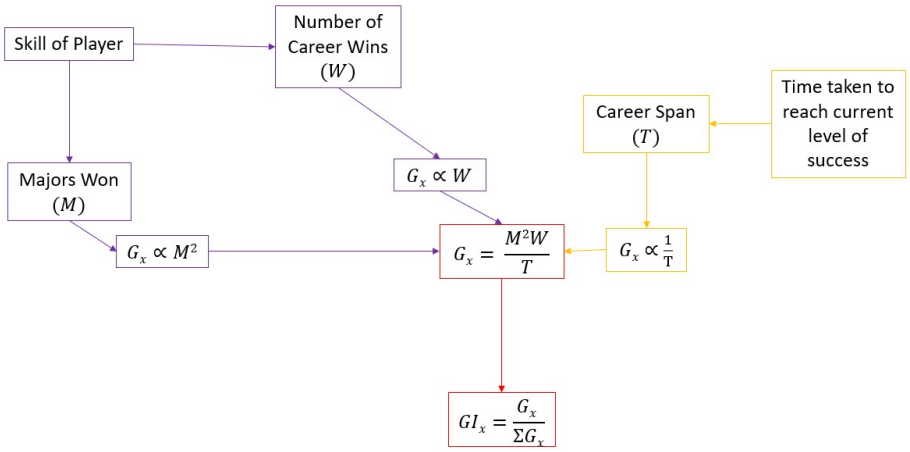
8 Schematic Diagrams for Models

The Schematic Diagrams for all of our models are as follows:

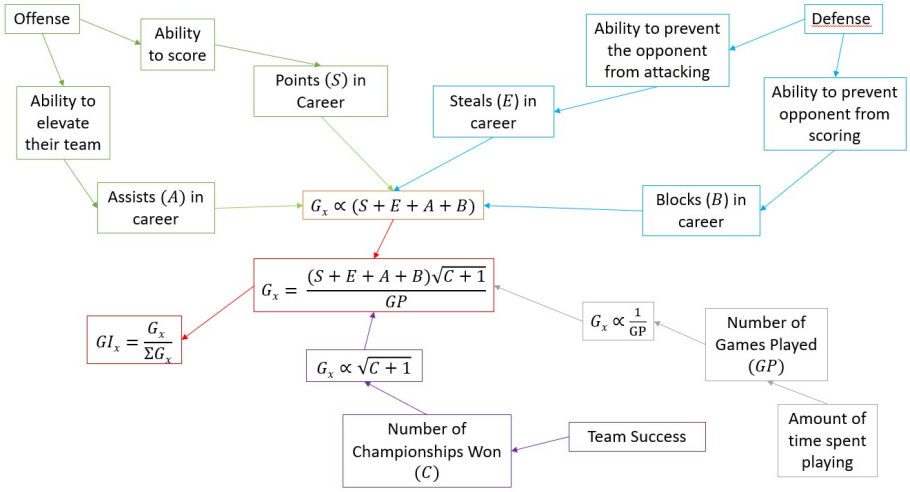
8.1 Schematic Diagram of Women's Tennis 2018 Model



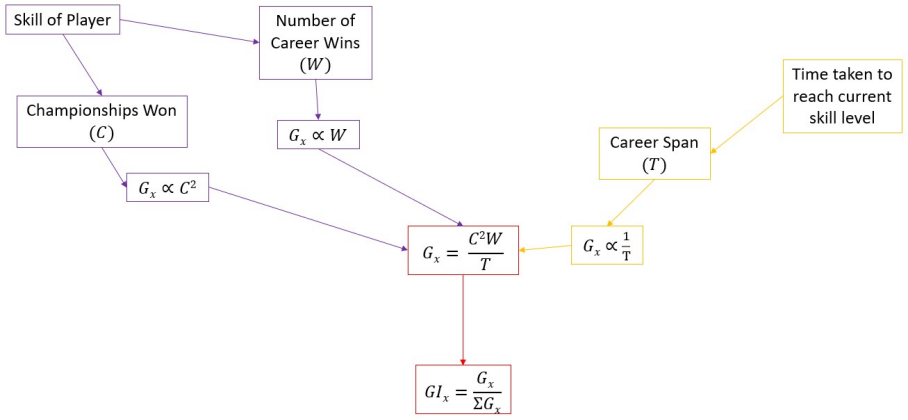
8.2 Schematic Diagram of Golf Model



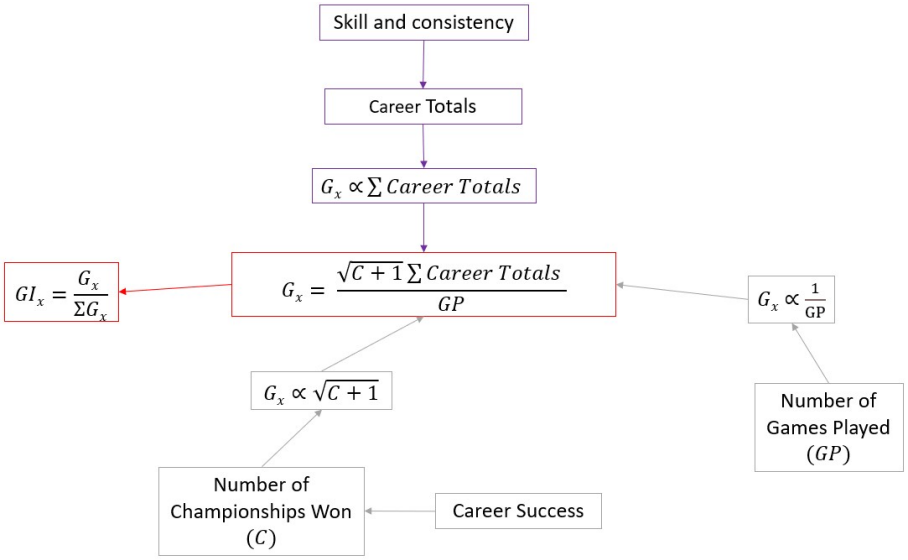
8.3 Schematic Diagram of Basketball Model



8.4 Schematic Diagram of Generalized Individual Sports Model



8.5 Schematic Diagram of Generalized Team Sport Model



9 Conclusion

Firstly, our models have a few limitations:

1. Popularity was not considered when comparing across eras
2. Intangibles such as Team Chemistry were not considered
3. Every professional player to ever play was not considered

Nonetheless, our model to determine the G.O.A.T. is reliable as we considered important factors:

1. Popularity (Within the same time-frame)
2. Pace
3. Consistency
4. Skill
5. Career Totals
6. Official Ranking
7. Team Success (Team Sports)

Using our models, our findings are as follows:

1. Simona Halep was the greatest Women's Tennis player in 2018
2. Jack Nicklaus is the G.O.A.T. of Golf
3. Michael Jordan is the G.O.A.T. of Basketball

These findings are also consistent with common consensus.

10 Letter to Director of *Top Sport*

Dear Sir/Madam,

Our team has taken a look at some of the articles and magazines from *Top Sport*, and found ourselves particularly interested in the debate of who is the greatest Golf player of all time, or the "G.O.A.T." of Golf.

In order to settle this debate once and for all, our team has done extensive research to determine how to create a mathematical model in order to objectively determine the G.O.A.T.. We found that in order to be considered the greatest in Golf one should have the highest number of major wins, total wins, and as short a career span as possible. These reiterate that the greatest player is consistent in their outstanding performances, and accumulates achievements at an astounding pace.

Equipped with the above criteria, our team went to the Professional Golfers Association (PGA) website to gather data on the top 11 players ranked by wins in the PGA tour. For each of these players, we compiled their career major wins, as well as the career span so that we can calculate each player's greatness. After which, we used this data in our mathematical model which emphasizes more on the number of major wins a player has, then takes into account the total wins and the career span. Following this, we went on to calculate the probability that a random Golf fan would choose the player as the G.O.A.T. when provided with the list of the 11 players with the most PGA wins, who make up our base data set.

From our model and data, we found that the G.O.A.T. of Golf is Jack Nicklaus, which is justified because he has the most majors, as well as the second highest number of wins, while doing so in an average time period. This shows his consistent outstanding performance over the course of his career, which makes him deserving of the title of the G.O.A.T.

After determining this, we realised that, for any individual sport, the greatest player is often characterized by outstanding performances throughout their career that sets them apart from the rest of the players in the sport.

We hope you find our insights useful.

Most Sincerely,
The IMMC Team

References

[1] Andy Bailey. *Bleacher Report's All-Time Player Rankings: NBA's Top 50 Revealed*. URL: https://bleacherreport.com/articles/2854727-bleacher-reports-all-time-player-rankings-nbas-top-50-revealed.

[2] GolfNews.co.uk. *The Top 10 Greatest Golfers of All Time*. URL: https://www.golfnews.co.uk/news/the-top-10-greatest-golfers-of-all-time/.

[3] NBA.com. *All Time Leaders in Basketball*. URL: https://www.nba.com/stats/alltime-leaders/.

[4] PGATour.com. *PGA Most Career Wins List*. URL: https://www.pgatour.com/stats/stat.ATR727.html.

[5] Rajdeep Singh. *Top 5 women's tennis players of 2018*. URL: https://www.sportskeeda.com/tennis/top-5-women-s-tennis-players-of-2018-ss/3.

[6] Tennis365.com. *Comparing Australian Open, French Open, Wimbledon and US Open prize money*. URL: https://www.tennis365.com/grand-slam/comparing-the-australian-open-french-open-wimbledon-and-us-open-prize-money/.

[7] WTATennis.com. *WTA World Ranking 2018*. URL: https://www.wtatennis.com/rankings/singles.

Appendix

Note: NaN values are returned if G_x and GI_x values are too close to 0. Tennis.csv, Golf.csv, and Basketball.csv are the datasets that we used.

Women's Tennis 2018

In [ ]:

```
import pandas
data = pandas.read_csv(r"tennis.csv")
```

In [ ]:

```
prize_money = list(data["Prize Money"])
followers = list(data["Followers"])
win_pct = list(data["Win Percentage"])
loss_rate = list(data["Avg. Games Lost Per Match"])
ranking = list(data["Ranking"])
greatness_sum = 0
for i in range(0, 11):
  greatness_constant = (prize_money[i]*followers[i]*win_pct[i]*1000)/((ranking[i])*loss
_rate[i])
  data.at[i, 'Greatness Constant (Using Prize money in £100000)'] = greatness_constant
  greatness_sum += greatness_constant
```

In [ ]:

```
greatness_consts = list(data['Greatness Constant (Using Prize money in £100000)'])
for i in range(0, 11):
  greatness_index = greatness_consts[i]/greatness_sum
  data.at[i, 'Greatness Index'] = greatness_index
```

Golf

In [ ]:

```
import pandas
data = pandas.read_csv(r"golf.csv")
```

In [ ]:

```
majors = list(data["Majors (num)"])
span = list(data["Career Span (yrs)"])
wins = names = list(data["Wins (num)"])
greatness_sum = 0
for i in range(0, 11):
  greatness_constant = (majors[i]**2)*wins[i]/span[i]
  data.at[i, 'Greatness Constant'] = greatness_constant
  greatness_sum += greatness_constant
```

In [ ]:

```
greatness_consts = list(data['Greatness Constant'])
for i in range(0, 11):
  greatness_index = greatness_consts[i]/greatness_sum
  data.at[i, 'Greatness Index'] = greatness_index
```

Basketball

In [ ]:

```
data = pandas.read_csv(r"basketball.csv")
```

In [ ]:

```
gp = list(data["Games Played"])
points = list(data["Points"])
assists = list(data["Assists"])
blocks = list(data["Blocks"])
steals = list(data["Steals"])
c = list(data['Championships Won + 1'])
greatness_sum = 0
for i in range(0, 10):
  greatness_constant = (c[i]**0.5)*(points[i]+assists[i]+blocks[i])/gp[i]
  data.at[i, 'Greatness Constant'] = greatness_constant
  greatness_sum += greatness_constant
```

In [ ]:

```
greatness_consts = list(data['Greatness Constant'])
for i in range(0, 10):
  greatness_index = greatness_consts[i]/greatness_sum
  data.at[i, 'Greatness Index'] = greatness_index
```