Unbetting Football, Chapter I  
 The False Irregularity of Mexican Football (v2.0)

In 2013 I published an article that endeavoured to show that the perception of La Liga MX as a league in which its teams behaved in “irregular” ways was erroneous. Indeed what was really true was that the Mexican league is perhaps the most even and equitable in the world. Even though it was an acceptable job using the inputs available at the time, I felt I could have achieved more. After 8 years of collating materials, this is my second attempt.

When I was a child I was lucky enough to receive a PlayStation for Christmas and one of the first games I had was the famous *Winning Eleven*, better known as “El Winin”. (wey, muy de mexico, para UK vale?) Although I had already had a few soccer video games before Winin, I had already been trying to develop something that aroused a lot of interest in me: making my own tournaments. However, it was not until "Winin" was it finally clear that what happened in the games was realistic (or what I understood by it at least). I started first with the World Cups, including even group draws with top seeds and pools. Then I configured each game in "CPU" mode for both teams (ie they were controlled by "the machine"), and thus I watched each of the 64 games while writing down the results, statistics and some other things that caught my attention. I remember being very surprised to see how teams would change players or even lineups, depending on what was happening in the match, without any intervention on my part.

Then I thought that I could do the same with what was then called the “First Division of Mexican Soccer” and I got down to work. I took the national teams and adapted them, changing everything from the uniform to each of the player's attributes, including their appearance, to resemble what I understood Mexican soccer to be. The idea came about when I saw that the then Tecos team had practically the same uniform as Suker, Boban and Jarni's Croatia. But, there was a problem.

One of the reasons I decided to simulate my own league was because in my World Cups, almost without exception, there was almost always something that I thought spoiled my beloved tournaments. That thing might have been that a favourite team lost in the group stage, that a bad team qualified for the second round, that the two favourites faced each other in the round of 16 or that, possibly, one or even two teams appeared in the final for the first time ever. I remember that this bothered me a lot, and when the same type of strange events began to happen in my simulations of Mexican soccer, I was at a loss to understand why was that happening? Was I doing something wrong? It left me with a peculiar feeling of not understanding what was happening and I confess that I abandoned it. That was until a few years later when an unexpected occurrence helped me begin the path to understand what was happening. In fact, it wasn't one thing, but two: I read Nassim Nicholas Taleb (explicar un poco), and Leicester City, my lifelong team (oh yeah!!), won the Premier League. Twenty years later, my intention in these articles is to answer my past self why neither its World Cups, nor its leagues, nor anything that it simulated, had anything wrong with it. (wey, suena como leicester gano hace 20 años) (a que sus refeiere a "sus"?) To do this, first, we need to understand in a very simple way how betting works and what it implies. (betting viene del cielo aca. Quiza hay de vincular en la introducción con lo que quieres desarrollar)

# Unbetting Football (describe que es el concepto de unbetting, no se le quedará claro al lector y el porqu'e quieres desarrollarlo))

To produce the following study and several more to be subsequently published, I asked Joseph Buchdahl of Football-data.co.uk for permission to use his data on sports results and bets linked to them. It is a database of around 150,000 matches from 28 European leagues, to which I added results and data from the Mexican league (Liga MX). All data presented in this chapter involves only league or regular season match results.

A classic example of a bet on soccer (in Continental Europe format) for a match is shown in the table below. The final result of the match is expressed in the first row of the table, while the numbers in the second row are the Odds or *multipliers* that establish the payment of a successful bet.

|  |  |  |
| --- | --- | --- |
| Home win | Draw | Visitor win |
| 2 | 3.3 | 5 |

If a person bets $100.00 that, at the end of the game, the home team will win, if it happens, their payout will be:

In the event that the result of the match was different from the local victory, the bettor would lose their money. However, an odd is not just a multiplier, it also represents an implied probability. To express an odd in terms of probability, the following operation is required:

Which, applied to the previous example, would look like this:

|  |  |  |
| --- | --- | --- |
| Home win | Draw | Visitor win |
| 2 | 3.3 | 5 |

|  |  |  |
| --- | --- | --- |
| Home win | Draw | Visitor win |
| 0.5 (50%) | 0.3 (30%) | 0.20 (20%) |

Note: the previous example is simplified, as it does not include the margin of the bookmakers, which basically works as a kind of commission. This will be explained in the following article: "Pandora's Box" .

Now let's bring to the table three games that happened recently: the first, Liverpool vs Chelsea, was played on March 4, 2021 in England (EPL); the second, Cruz Azul vs. Monterrey, was on August 19, 2021 in Mexico (Liga MX) and the third, Boavista vs. Moreirense, on February 19, 2021 in Portugal (Primeira Liga). The three matches, expressed in probabilities are:

|  |  |  |
| --- | --- | --- |
| Liverpool | Draw | Chelsea, |
| 42% | 26% | 32% |

|  |  |  |
| --- | --- | --- |
| BLUE CROSS/BLUE SHIELD | Draw | Monterrey |
| 42% | 26% | 32% |

|  |  |  |
| --- | --- | --- |
| Boavista | Draw | Moreirense |
| 42% | 26% | 32% |

One of the previous matches ended with a 1-0 victory for the local team, another was won by the visitor 0-1 and another ended in a 1-1 draw. Of course, this is not a memory challenge or a riddle, the point here is that, even if we do not know any of the teams that played in these matches, any of the three results is perfectly applicable to any of the matches. So what is the difference between those games? One is one of the best possible matches in the world, while the other two are clearly of a more modest profile, however, this difference is only a difference in *scale* , which is only possible to appreciate when the teams compete in international tournaments. However, in terms of what happens within each game, that is, events such as fouls, shots, cards, goals or even the results, they are extremely stable. *in the long run* .

Let's imagine that the 3 matches listed above happened at the same time, would that change anything? Assuming that the teams were not affected by time zone issues and that the results of one match did not affect the other two (which is true, since the result of one tournament is independent of the others), it turns out that the turnout in the weather does not affect matches; Or if the matches had occurred 5 years apart from each other? As long as the rules of the game remained more or less intact, the matches would continue to occur in the same way. Considering the above, it turns out that we have a powerful card in our favor: how many games were there, between 2012 and 2021, in which the local teams had a 42% chance of winning? To be exact, 2484.

Now the most important moment has come: let's take that set of matches and ask them questions. For example, what was the goal average in that *stage* ? Having scored a total of 6,185 goals in those 2,484 games, the average number of goals per game was:

But is that too little or too much? Or let's put the question from another angle: how do we know if someone is taller than another person? What we do is that we take both individuals and compare them with what we have defined as a measure (previously agreed upon and known by both parties), in this case the *meter* . If one of our hypothetical friends were 2 times as tall as a meter and the other 1.5 times, we will have solved the dilemma and we will know who is the tallest.

Returning to the average of 2.49 goals in the scenario of 42% probability of victory of the home team, how can we know if this is too much or too little? Comparing the average of that scenario with the other 99, 😎 .Let's observe, please, very calmly the following graph:



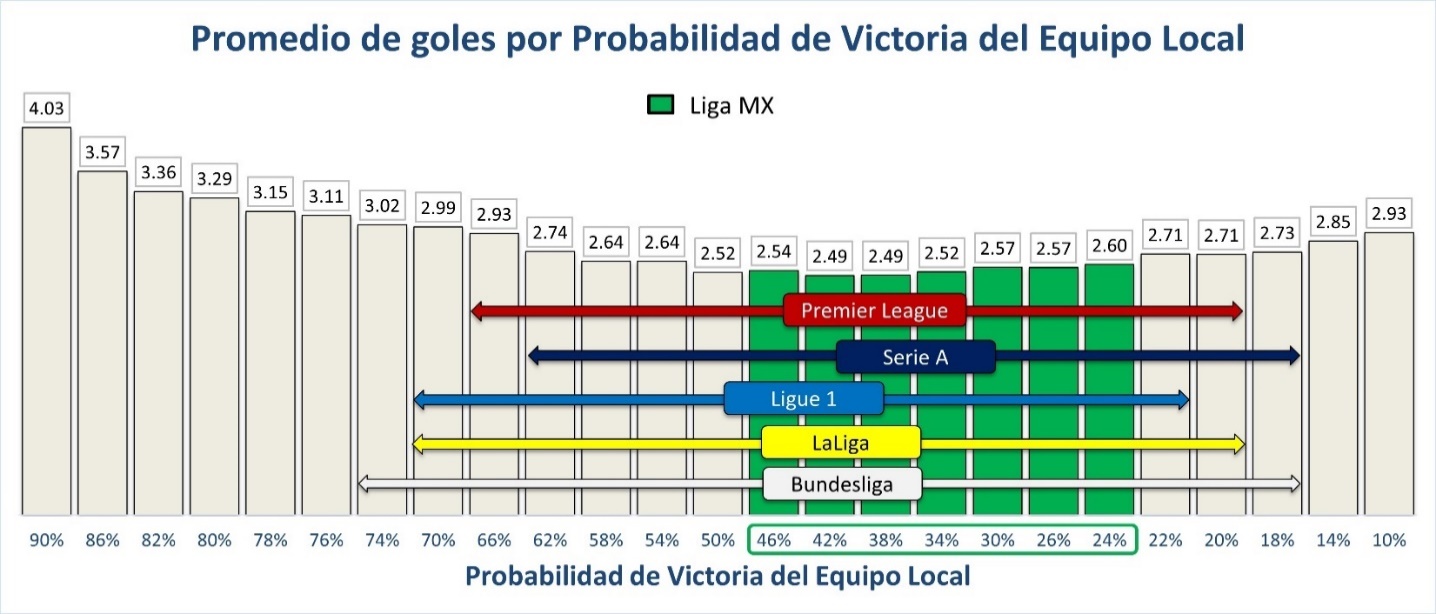
The graph above has some very interesting features. In the first place, we see that the graph describes a U-shaped curvature, that is to say that the average number of goals per scenario (the height of the bar) is compressed towards the interval 42% to 38%, reaching what seems to be a minimum value. . To find out why this happens (our hypothesis) you will have to wait a bit, to a later chapter of *Unbetting Football* , and which will be called "The times of all plays, in The Play of All Times". But, for now, let's bring our beloved Liga MX into this context. If we remember one of the 3 matches that we mentioned before:

|  |  |  |
| --- | --- | --- |
| BLUE CROSS/BLUE SHIELD | Draw | Monterrey |
| 42% | 26% | 32% |

The match above is a perfect example of what it's like to compete in Liga MX. For the following measurement, all Liga MX matches between August 2012 and December 2021 were taken and the odds of victory for each team were averaged. Where are the Liga MX teams located in the range of scenarios presented in the previous graph? Here:



Is this little or much? Right or wrong? Let's see the following graph to compare Liga MX with the top 5 leagues in the world:



At first glance it seems as if Liga MX's detractors have found their Promised Land, but will they? Let's ask more questions, but let's ask all the parties. Let us consider matches that fall into the following scenario:

|  |
| --- |
| Home win |
| 42% |

We know that the matches in which the local teams have a 42% probability of victory, the average number of goals per game is 2.49, but this average must necessarily be made up of the sum of the averages of the local teams plus the average of the visitors:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Average goals on stage 42% |  | Average goals of the  Local |  | Visitor Average Goals |  |  |
|  |  |

Let's talk a little about what the numbers tell us, what they mean. Of course, according to the rules of the game, it is impossible to score 1.4 or 1.09 goals in a game, but what if it was 100 games? How many goals *we would expect* What would both teams score if they played a total of 100 games under the same circumstances?

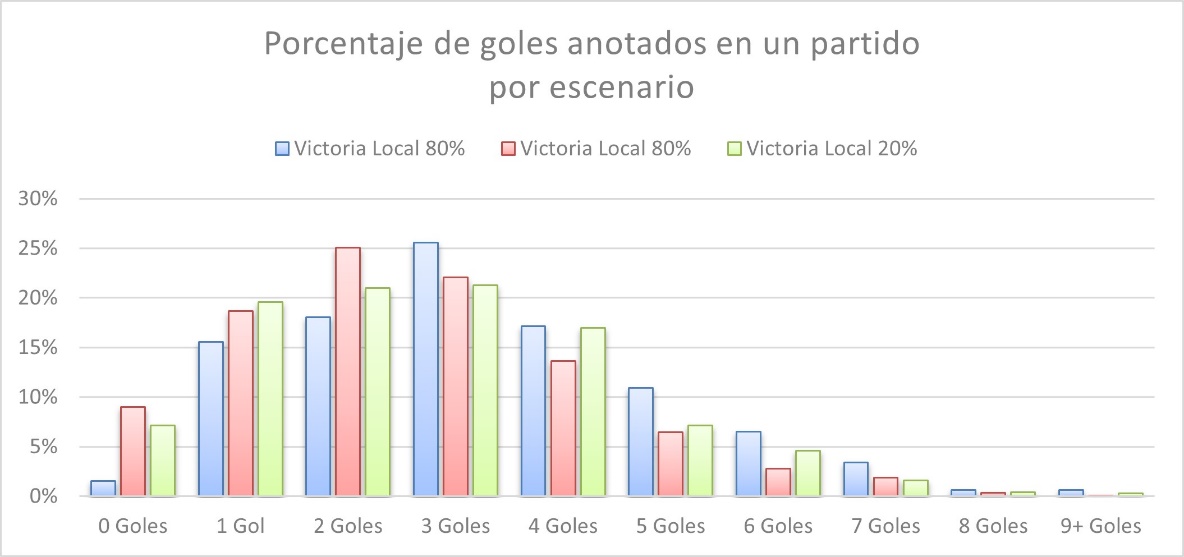
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Average goals on stage |  | Home Goal Average |  | No. of parties |  | Visitor Average Goals |  | No. of parties |  | goals |
|  |  |  |  |

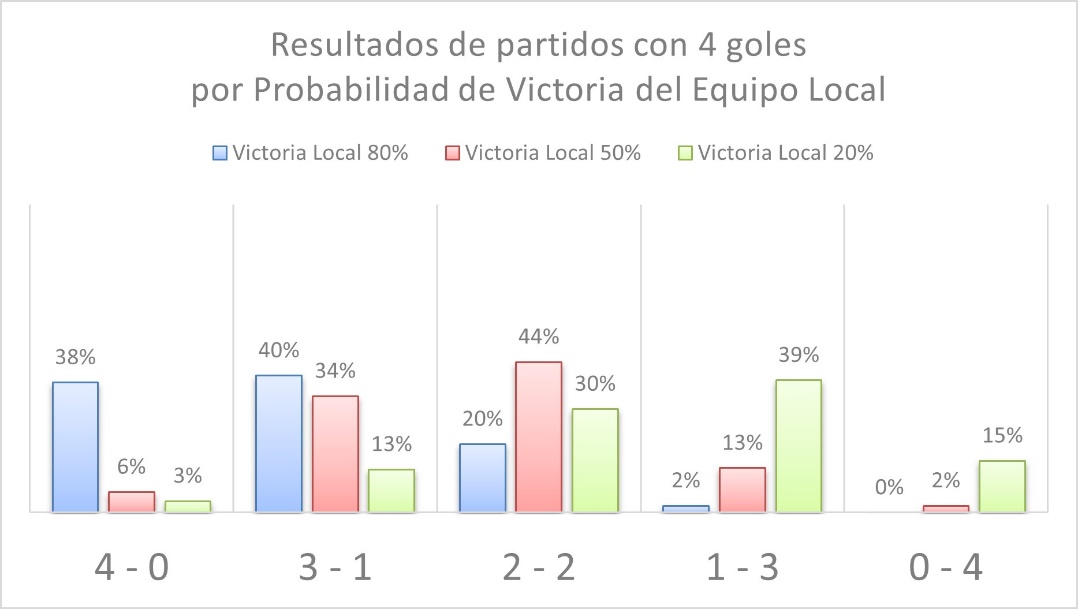
What we are saying in this case is that, for every 100 games played in this circumstance, we would expect a total of 249 goals to be scored. But let's look at another curious detail in the operation we just did, what percentage of the total goals were scored by the home teams and the away teams? Let's observe:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Average goals on stage |  | **Home Goal Average** |  | No. of parties |  | Visitor Average Goals |  | No. of parties |  | goals |
|  |  |  |  |
|  |  | **140  local goals**  56% | | |  | **109  away goals**  44% | | |  |  |
|  |  |  |  |

Is this difference a little or a lot? How can we know if it is too little or too much? That's right, we have to compare. Let's put together 3 types of matches from their odds:

|  |  |  |
| --- | --- | --- |
| Home Win Probability | Home Win Probability | Home Win Probability |
| 80% | 50% | 20% |

Now let's look at the following graph that shows how the number of goals is distributed according to each scenario:

Of course, there are several interesting observations that we can make about the graph above. But, to understand the weight of the scenarios in the matches we must look more precisely. Let's get closer to the parties and ask another question. If we consider only the matches in which 4 goals were scored, we would have 5 possible outcomes. Let's observe how they are distributed according to the scenario:

From the graph above we can make some observations. First, it is clear that the blue bars are loaded towards the results that favor the local (to the left). For its part, the green bars are loaded towards the results that favor visitors (to the right). The red bars, on the other hand, seem to be more balanced towards the tie (towards the center). Which of these three types of scenario will most closely resemble what happens most often in Liga MX? Without a doubt, the one with the tendency towards the center. Now let us recall the following graph:



The further away a scenario is from the interval between 46% and 38%, the more the goals per game will tend to increase, however (and this is my opinion), personally I prefer to see games that face more even rivals where they tend to result more balanced or, at least, that they were not humiliating. Again, in my opinion, scoring to the point of humiliating an opponent is not a sign of competitiveness: it is proof of its absence.

If we compare it with the Box, watching a match like Bayern Munich against practically any other team in the Bundesliga is something like the sporting equivalent of seeing Tyson Fury (2.06 meters, 126 kilos) face El Canelo Álvarez (1.75 meters). , 76 kilos) or Gervonta Davies (1.66 meters, 61 kilos).

Between the Apertura 2012 and Apertura 2021 tournaments, 25 teams participated in Liga MX, scoring 7,425 of the 7,395 expected goals in the regular phase, that is, 100.4% of goals, with León and América being the teams that scored the most goals, 502, which it represents 6.7% of the total goals; On the other hand, in the same period and with the same number of games in the Bundesliga, 8,028 goals were scored out of 7,265 expected goals, that is, 110.5%, with Bayern Munich being the team that scored the most goals, 698, representing 8.7%. The difference between the percentage of León and América compared to Bayern Munich seems very small, but what implications will it have? To appreciate it, it's time to go on vacation.

# We take it to Monte Carlo

When I was at the University, at the Monterrey Campus, I took a subject called “Signal Analysis and Communication Systems” with the great Doc Olvera. I remember feeling a mixture of admiration and deep respect for My Teacher back then. The Doc, who was not short on charisma, was brilliant in many ways, but, in particular, when it came time to solve something in class (real math puzzles) he had a generous deck of "old tricks" that nowadays we don't have. I know if they were science or magic, but at that moment I remember that it seemed surprising to me how telecommunications could be described using Transforms and Theorems that until that moment seemed to me to be the most unconnected with reality. One of the old tricks that El Doc used the most (or, surely, the one I remember the most) was what is known as "L'Hôpital's Theorem" and, every time he applied it to solve something, he told us "and now how do we do it? Well, we took him to the hospital!”

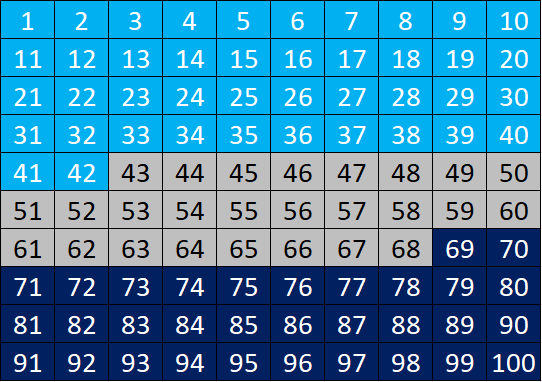
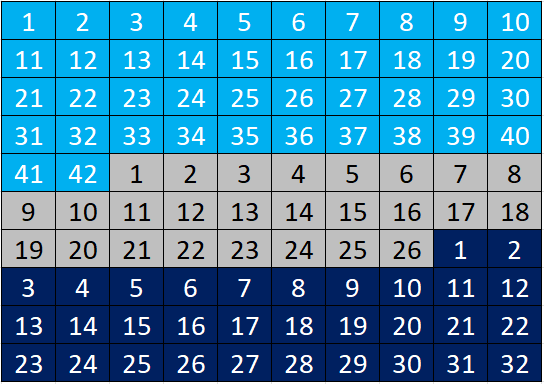
So, for the end of this chapter, and with The Doc's permission, let's go to a most enchanting destination. To get there we need a very simple machine, but at the same time very special, it is a box that gives us *random* a ball that has a printed number between 1 and 100. We do not need more.

Now let's remember the game we mentioned a few paragraphs ago:

|  |  |  |
| --- | --- | --- |
| Victory  Blue Cross | Draw | Victory  Monterey |
| 42% | 26% | 32% |

What we will do is distribute those percentages between the numbers from 1 to 100, assigning 42 numbers to Cruz Azul, 26 to Empate and 32 to Monterrey:

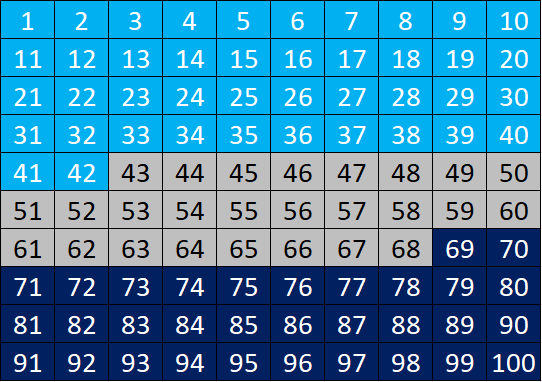
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Victory  Blue Cross | Draw | Victory  Monterey |  | Victory  Blue Cross | Draw | Victory  Monterey |
| 42% | 26% | 32% |  | 1 to 42 | 43 to 68 | 69 to 100 |

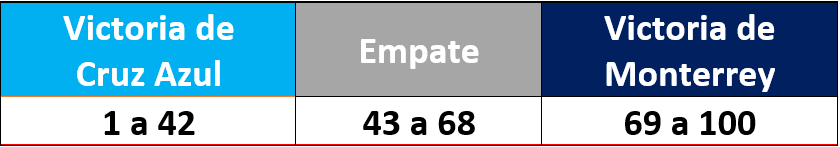


In such a way that the percentages would now be translated into three ranges of values. With this ready, now we ask our box to give us a ball, which turns out to be the following:

As the number of the ball was within the range of numbers that we assigned to Cruz Azul's victory, then, we will say that the Celeste team won the match. Thus:







But that was only one game and in Liga MX there are, every season, 17 days of 9 games that gives a total of 153 games. The results determine the positions in the table at the end of the regular season awarding 3 points per game won and, in case of a tie, 1 point is distributed to each team. To simulate a possible tournament, thanks to the Matches and Bets Database of https://www.football-data.co.uk/ that they kindly allowed me to use, we would only have to repeat the operation of the ball to each of the 153 matches. Finally, with the results obtained, we accumulate the points and, in the end, we would have our general table:







And if we're here, how about we simulate another tournament? Or better once they are 4:









The five tournaments simulated so far are possible tournaments that end up forming a set that we can ask questions of, but why fall short? What if we follow him? A Liga MX tournament is made up of 153 matches and each of them has 3 possible results (victory of team A, draw and victory of team B). If we compare it, for example, with the group stage of the FIFA World Cup, in a group of 4 teams there are a total of 6 games (3 days with 2 games each) and each one has 3 possible results. Hence, the total of possible outcomes is calculated as which is equal to 729 possible combinations of results in 6 games, then How many possible tournaments are there in our beloved Liga MX if its season has 153 games? To be exact: which is… a figure of more than 70 numbers and for which there is not even a word to name it in any language of the world.

That's how big football is.

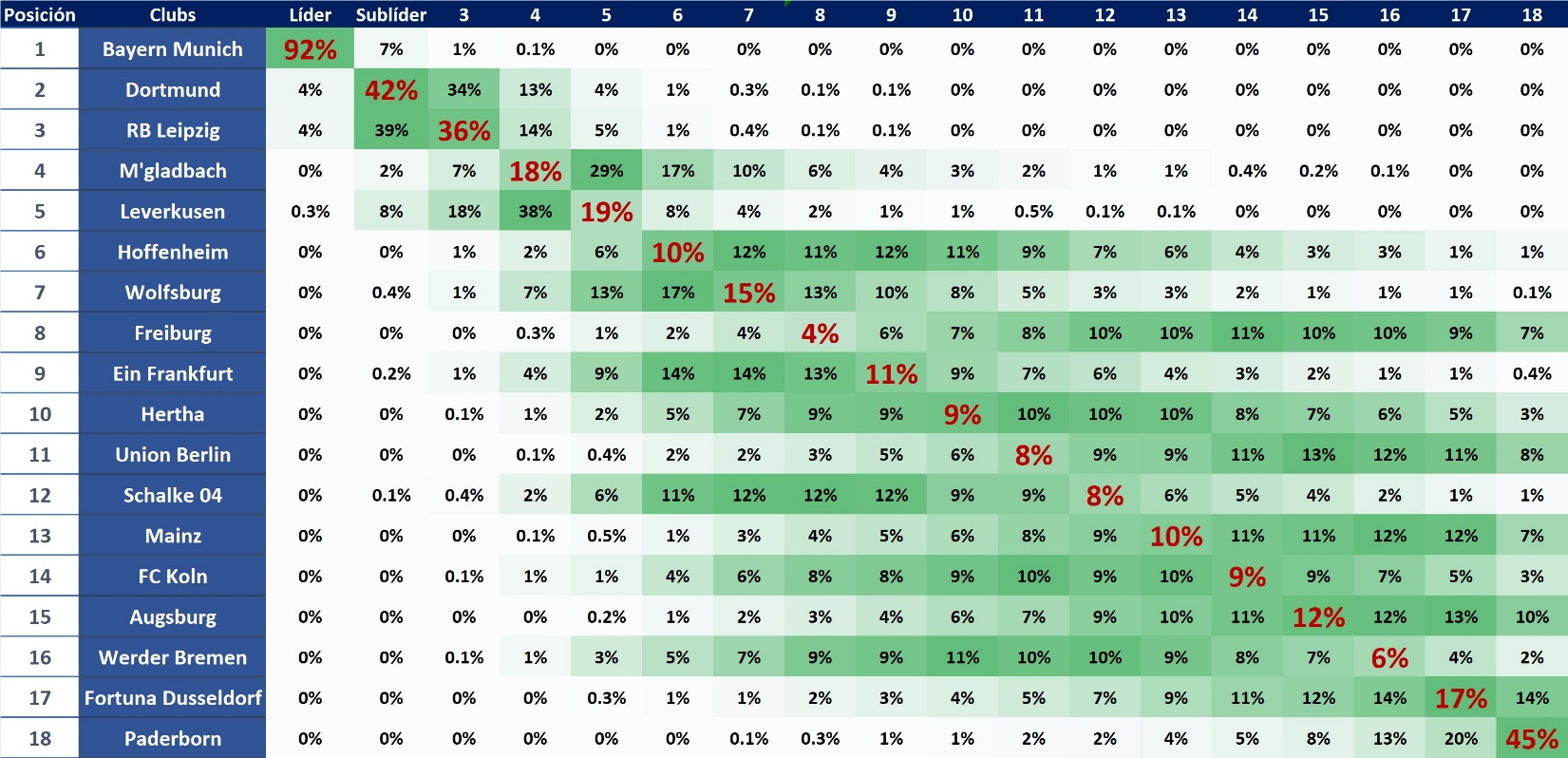
However, since listing all those possible tournaments and calculating their odds would take a long time, and since it would be preferable to get results in less than a human's lifetime, and since no one here wants to put Dr. Strange out of a job , let's lock the number of simulations to 5000.

The following table shows the result of what started with a PlayStation and some world cups at El Winin. To understand the table, consider the following points. I recommend that, after reading each point, try to understand it in the table:

* In the table, the 18 teams appear in vertical order, according to the position they occupied in the Apertura 2021 tournament.
* The Leader, Sub-Leader, 3, 4, 5, 6, etc. columns represent the 18 possible positions in Liga MX at the end of the season. Each team, in principle, could be in any of the 18 positions.
* The percentages in each of the boxes represent the percentage of times that, out of 5,000 simulated tournaments, they were in that position.
  + For example, of the 5,000 simulations, the Atlas was the Sub-Leader on 639 occasions, that is, 13%.
* The numbers in red that are diagonal represent the percentage that the 18 teams had of being in the position they occupied in the Apertura 2021.
  + For example, of the 5,000 simulations, in 10% of the cases, Tigres was in position 4.
* The color in the cells is a scale of tones that, the more green, the greater the percentage that the team occupied that position in the table. On the other hand, the whiter the box, the lower the percentage of the team occupying that position.
  + For example, Mazatlán had a 22% chance of finishing last overall and, being its highest percentage, appears in the darkest shade of green. At the same time, Mazatlan had a 0.1% chance of being the Leader of the tournament and, being its lowest percentage, it appears in white.
* Note that the percentages are rounded.
  + Rounding to 1 decimal, 0.05% rounds to 0.1%, while 0.04% rounds to 0%.



For fear of sounding even more repetitive, remember that it's not enough to just get results. To put the previous table in perspective, take a look at the following table that corresponds to the same type of simulation, but applied to the Bundesliga in its 2019-2020 season.



Thousands of tournaments later, what we have built is a *Probability Approximation* using what is known as *Monte Carlo simulation* and, with these tables in hand, it is necessary to ask them questions:

In which of the two tables appear more boxes with 0% probability? What interpretation would you give to that detail? Or another way of appreciating the above: in which of the two tables appear more boxes colored with some shade of green? What could this tell us? Another way to appreciate the idea of this chapter: in which of the two tables are there more teams that have a 0% chance of coming in last? Or, the other way around, in which of the two tables are there more teams that have a 0% chance of reaching first place?

Which of the two tables shows a more competitive tournament?

The False Irregularity of Mexican Soccer

All soccer leagues in the world are regular in the long run. The perception of irregularity in Liga MX is an illusion resulting from the combination of a high degree of parity among Mexican clubs and an Anchorage Bias ( *Anchoring Bias* , the tendency to overvalue the latest information found or available).

- Jaime G. Melendez

The word " [word](http://etimologias.dechile.net/?palabra) *”* comes from the Greek *parabolê* which means “to draw a parallel between *”* , that is to say: *compare* .

Twitter: [@JamesGMelendez](https://twitter.com/JamesGMelendez)  
 Web:<https://academiathi.com/>  
 E-mail: [james@academiathi.com](mailto:james@academiathi.com)  
 LinkedIn: [jaimemelendezthi](https://www.linkedin.com/in/jaimemelendezthi/)  
 GitHub: @JamesGMelendez

Follow us on Enrachados Sports  
 podcast: [Enrachados Thursday 9:00 p.m.](https://www.facebook.com/EnrachadosDeportes)  
 Twitter and Instagram: [@Enrachados\_deportes](https://www.instagram.com/enrachados_deportes/?hl=en)

Previous article: [The False Irregularity of Mexican Soccer](https://www.futbolsapiens.com/?p=156562?utm_source=&utm_medium=&utm_campaign=) .

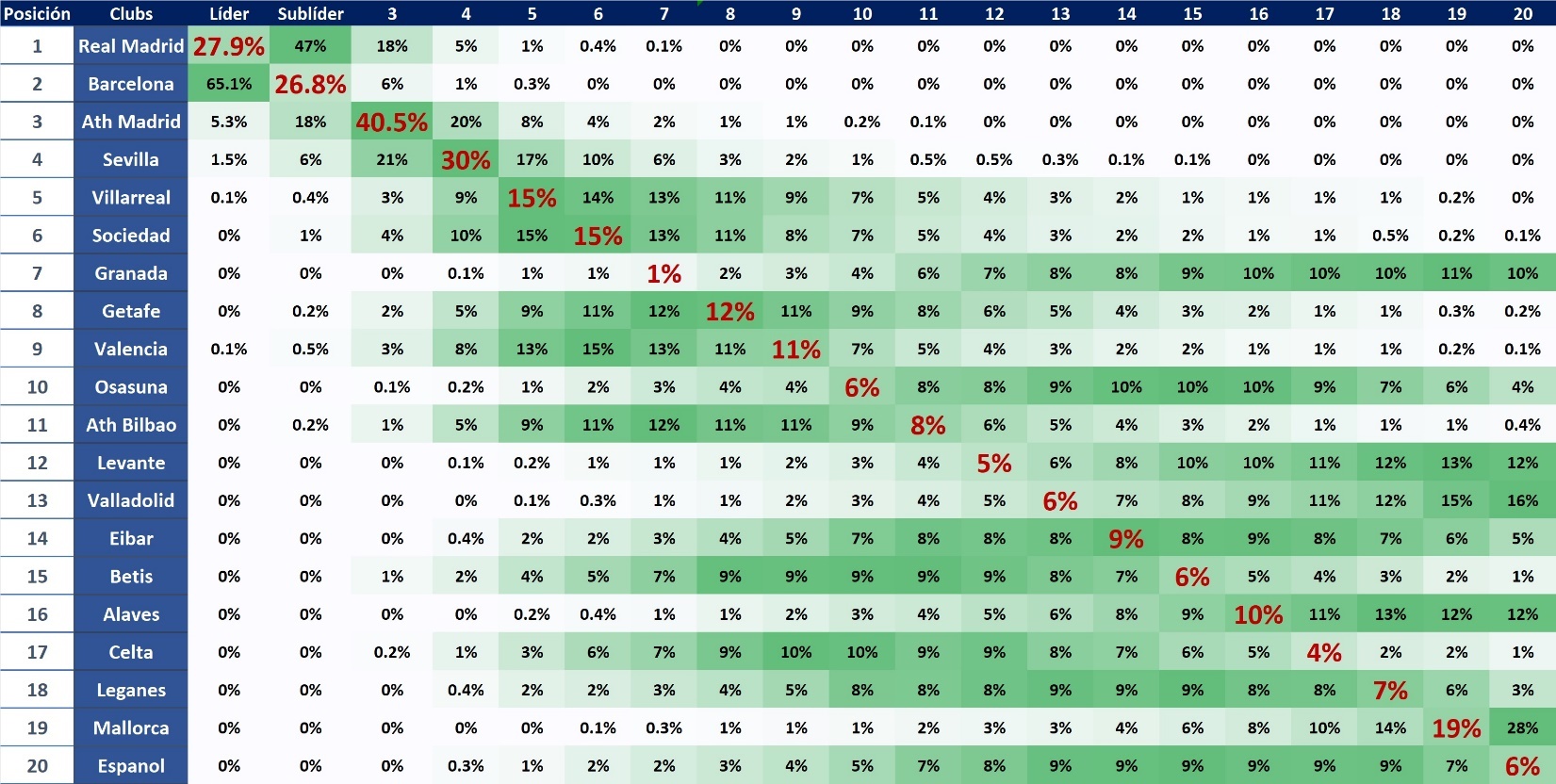
# Appendix

Note 1: Readers should excuse a bit of drama, of course it is possible to work with circumstances involving amounts as large as without resorting to computational brute force. It is enough to invert the exponential function to transform it into a logarithmic function and do the necessary calculations without falling into *overflows* (Basically, when we crash a processor for handling numbers that are too large, something like the computational equivalent of Roberto Durán's famous “No más”). But, I guess you'll agree that saying "That's how great football is" sounds a lot cooler.

Note 2: find a probability table by performing *n* number of simulations, of course that is not the only way to discover the probability of a set of events. Of course it is possible to do it analytically (i.e. figure out the formula that calculates the percentage that should be inside each cell), but to be honest, although I consider myself to be moderately proficient in algebra, I am not an algebraist nor do I claim to be. An Approximation of Probabilities by Monte Carlo Simulation is not only good enough to communicate the idea of this chapter, it is simply revealing.

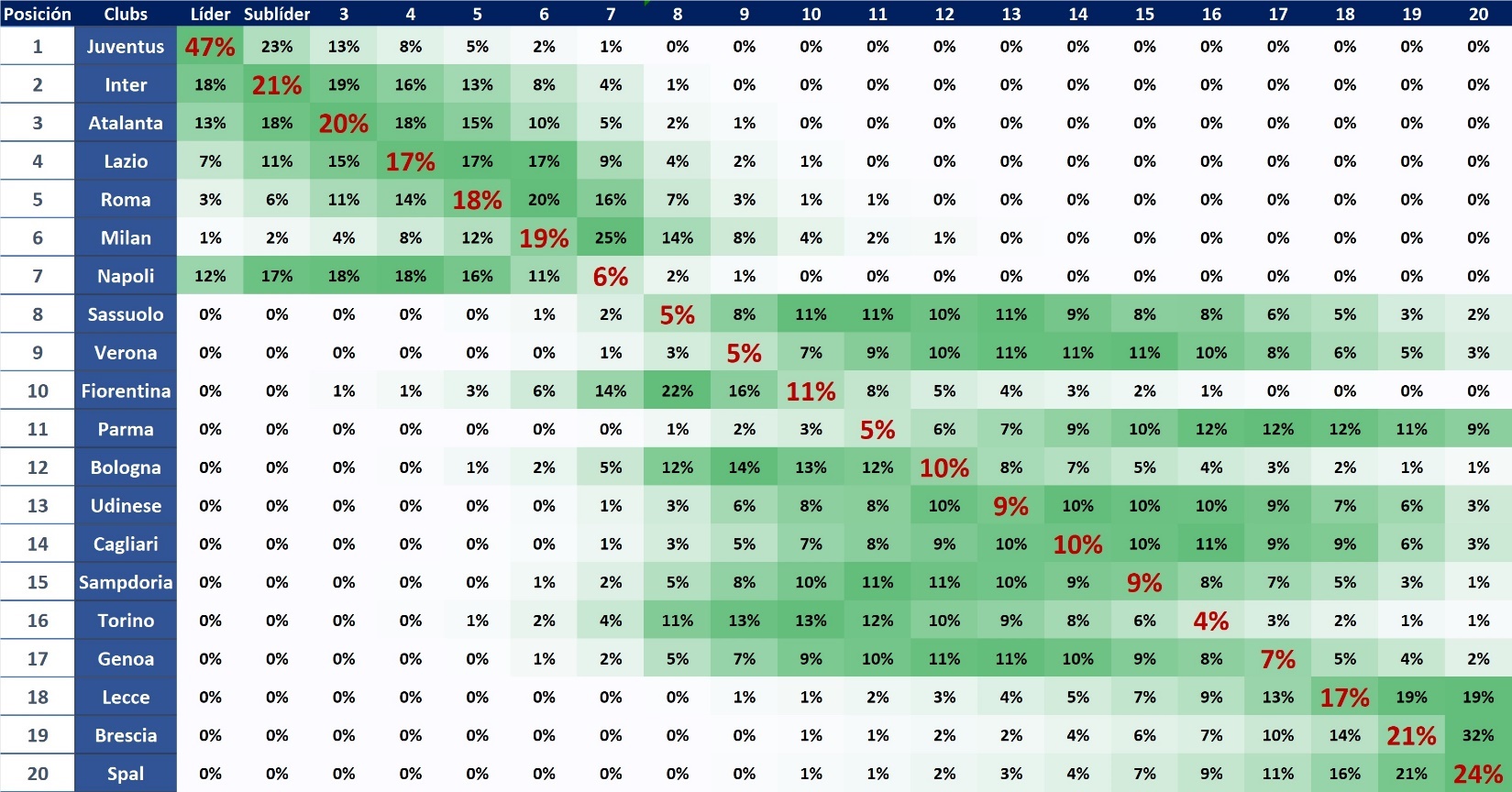
Note 3: below, I add the same Probability Approximations expressed in tables for LaLiga, Serie A, English Premier League and Ligue 1 in the 2019-2020 season. I leave you to take a tour.

Title: LaLiga, 2019-2020 season.



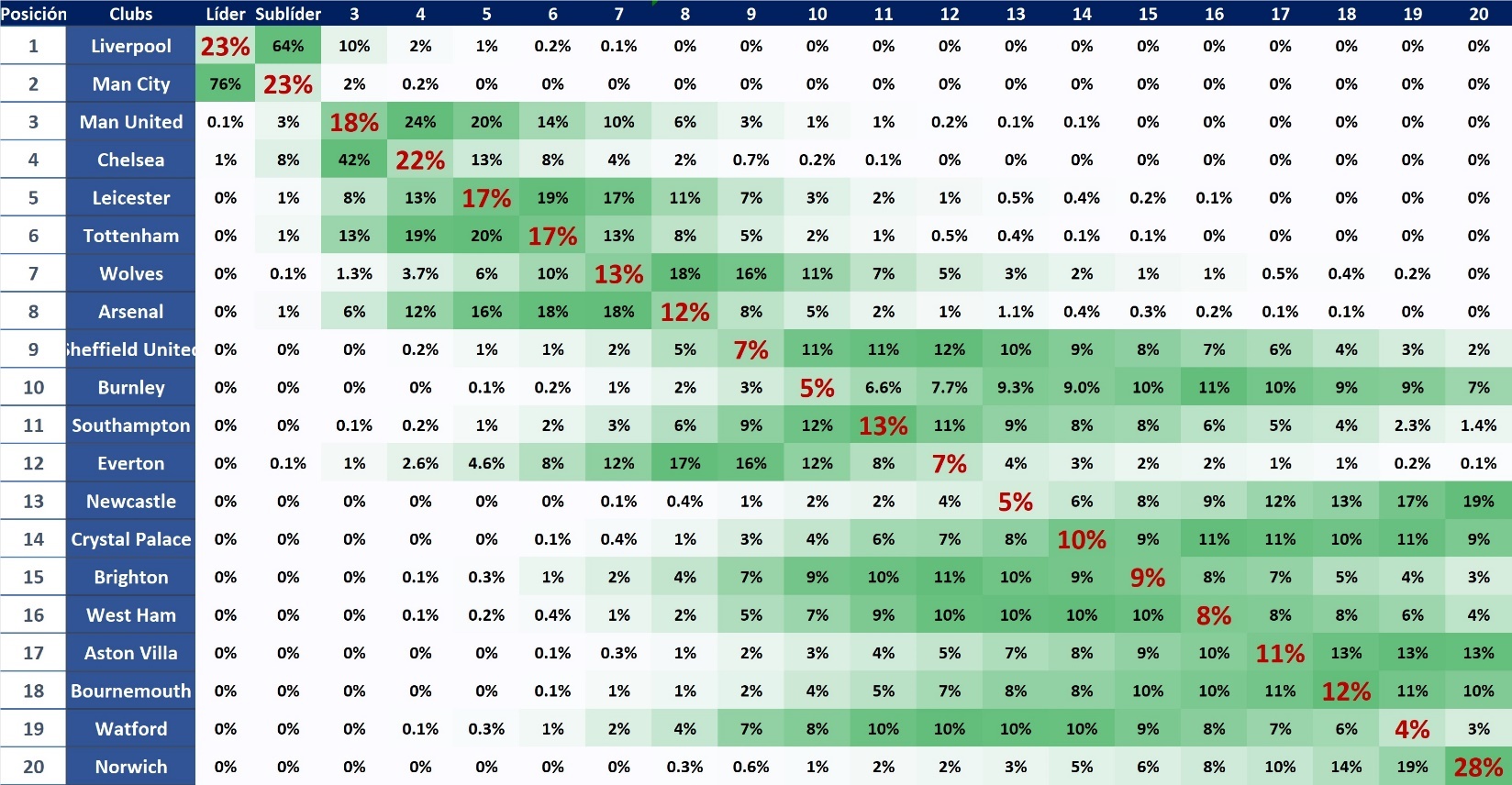
Description of work: Duel to the death with knives.

Title: Serie A, 2019-2020 season.



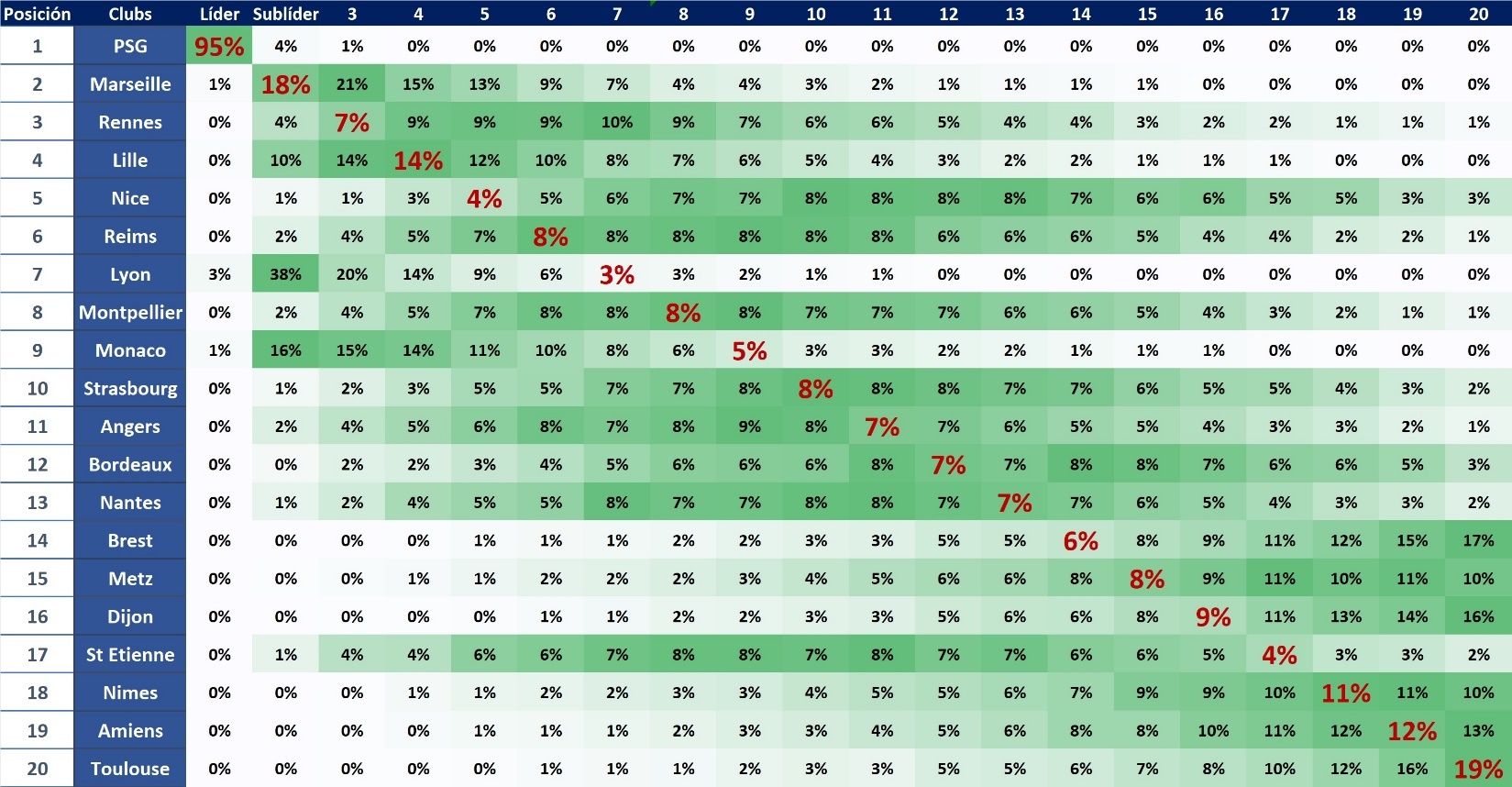
Description of work: El Senado, El Pueblo and 3 Serie B teams.

Title: English Premier League, 2019-2020 season.



Description of work: The Big ~~Six~~ Two.

Title: Ligue 1, 2019-2020 season.



Description of work: Give them cakes.