



Data Science

With a side of Chaturanga

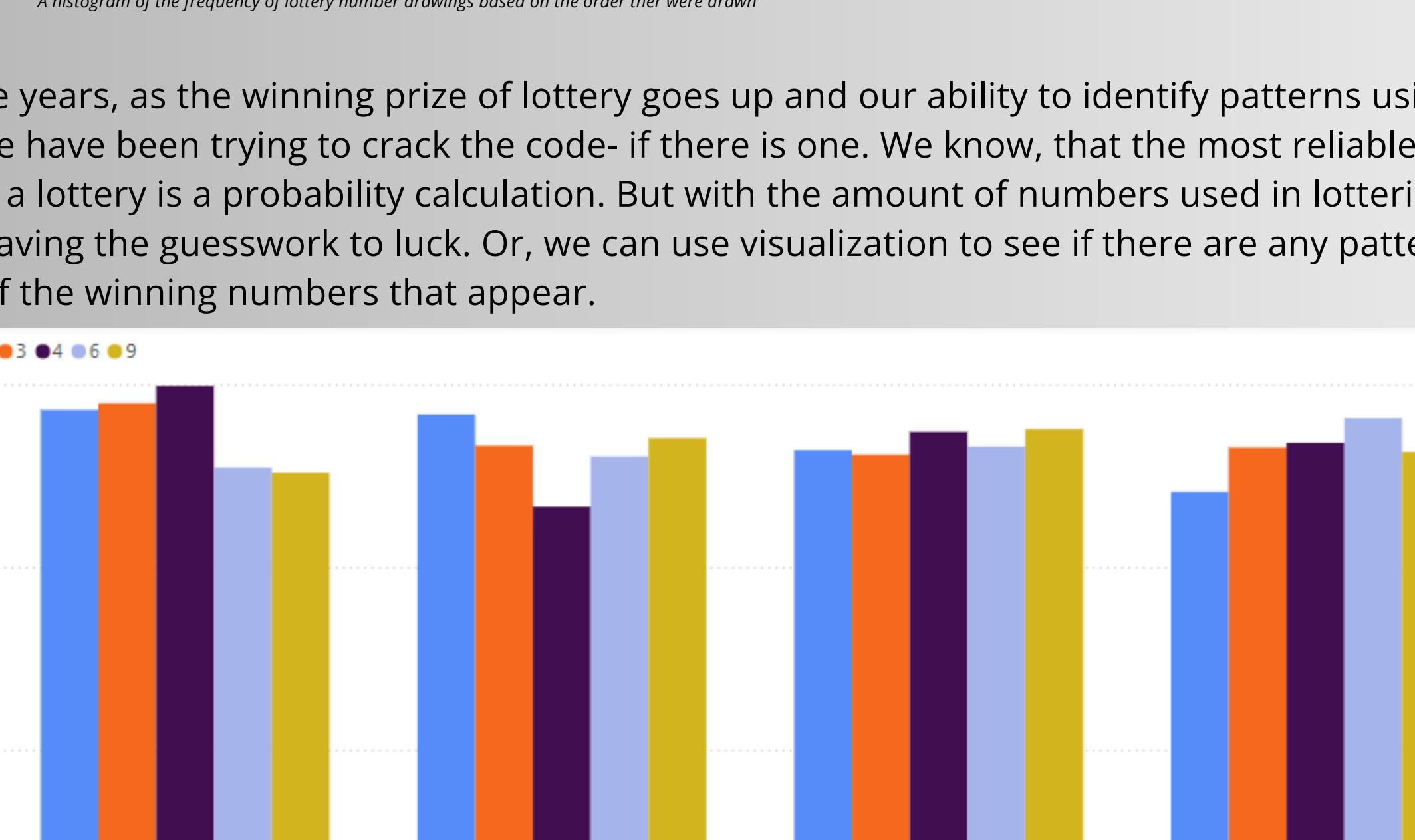
The Art of Analysis - Turning Randomosity to Insights



Ai Paula Couto

January 25, 2025

It's often been said that Data Science is both an art and a science. In this blog, I will be using information from Powerball, Megamillions, and Pick 10 lotteries in New York to apply scientific and artistic principles to understand patterns about the bi-weekly numbers drawn that seems to have the power to change people's lives, for better or worse.

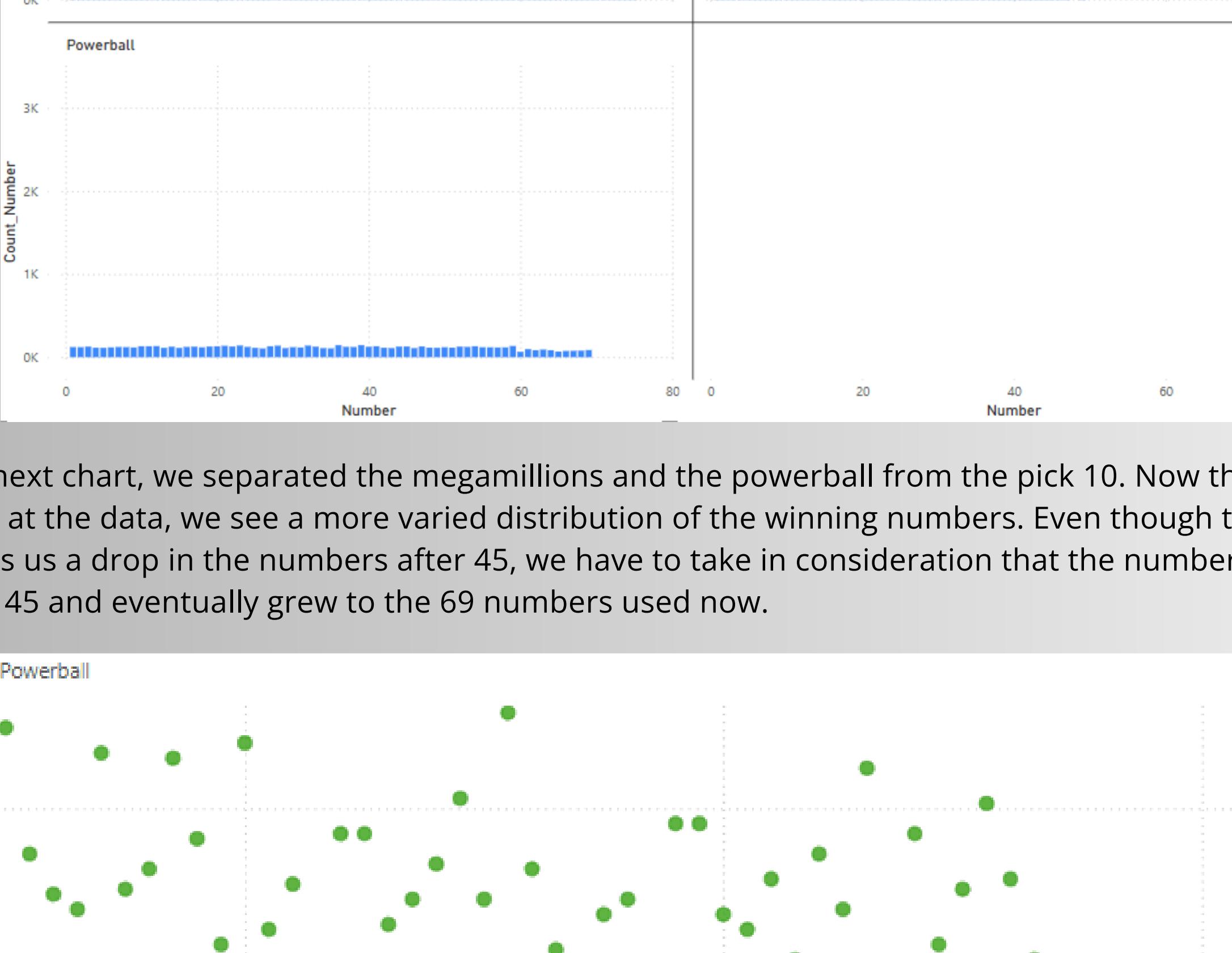


A histogram of the frequency of lottery number drawings based on the order they were drawn

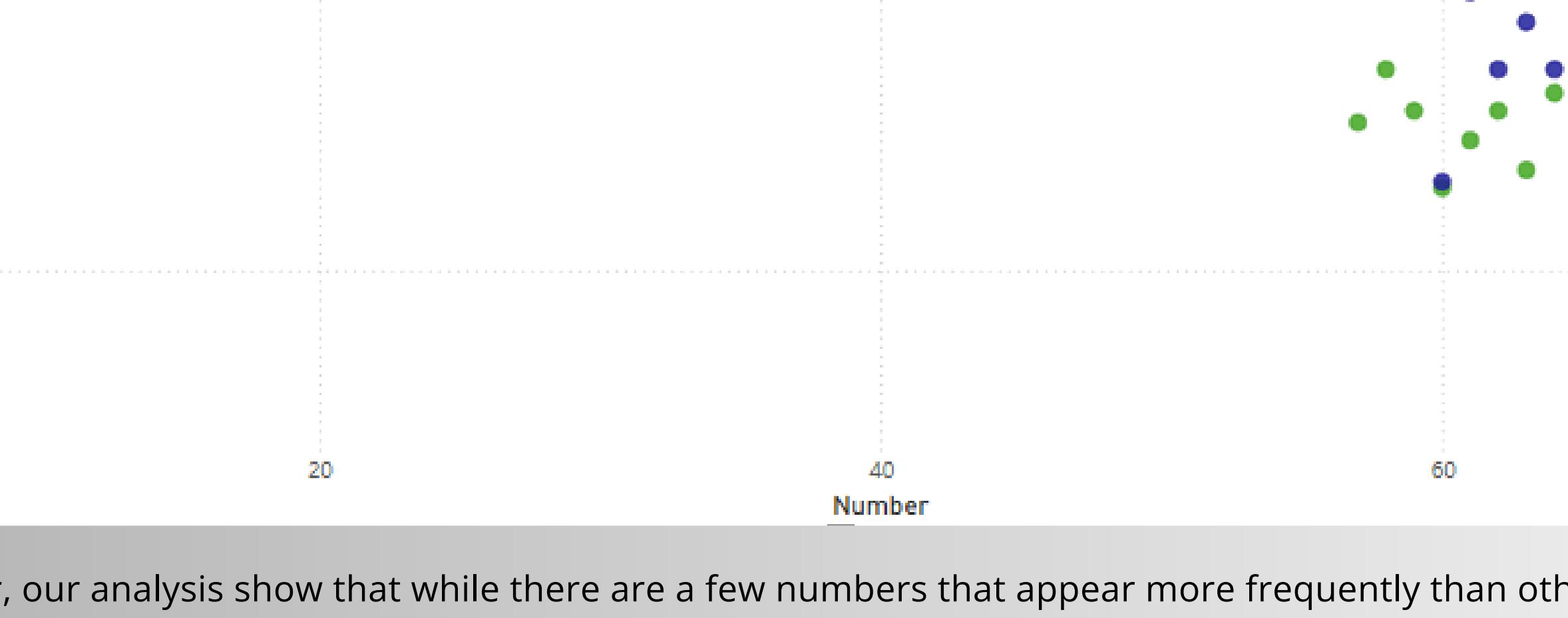
Over the years, as the winning prize of lottery goes up and our ability to identify patterns using data, many people have been trying to crack the code- if there is one. We know, that the most reliable source of outcome in a lottery is a probability calculation. But with the amount of numbers used in lotteries, we are better off leaving the guesswork to luck. Or, we can use visualization to see if there are any patterns to the frequency of the winning numbers that appear.



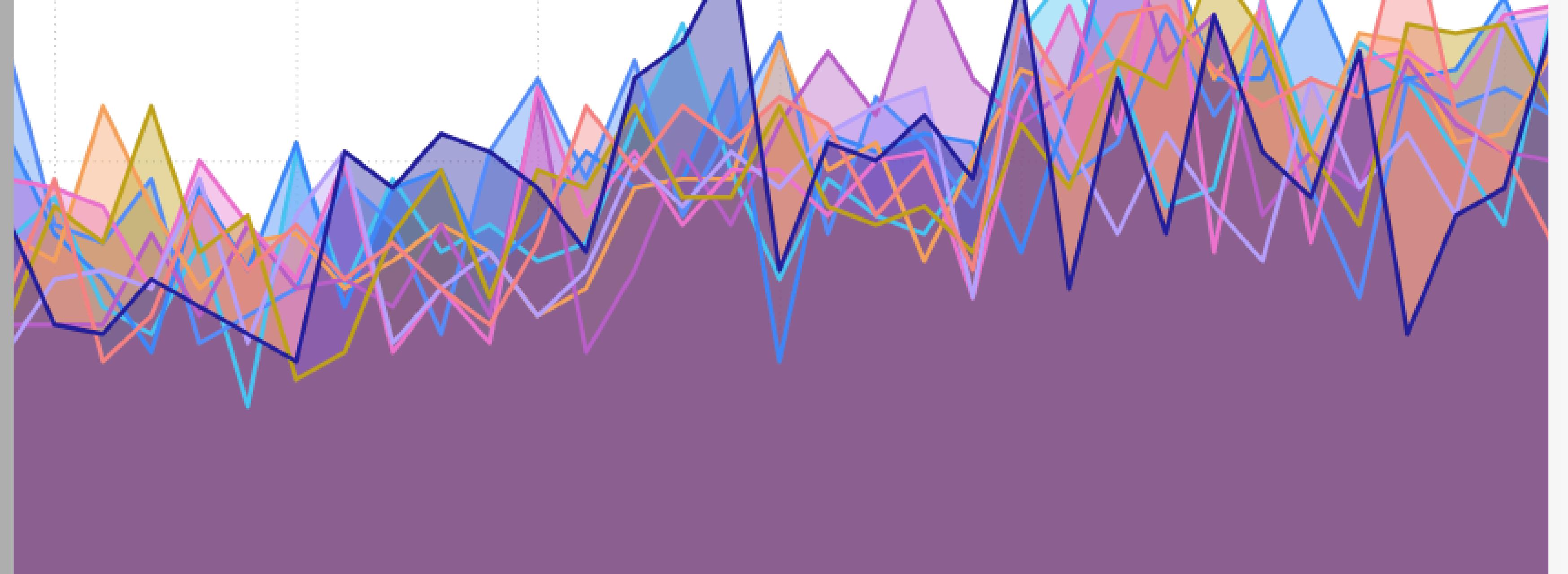
In this chart, I filtered the top 5 numbers that appear in every lottery winning since 1984. These numbers are 2, 3, 4, 6, and 9. Then, using a histogram, I plotted the frequency of these appearances by Quarter. This information tells us that if we are to buy a lottery ticket in the first quarter of the year, the number 4 would be the most likely to appear. Unfortunately, both the megaball and the powerball requires 6 numbers and the Pick 10 requires 10 to win. However, for the powerball and the mega million, if you get one number right - you at least get the money you betted on. So, using statistics, what is the most likely number to appear as the mega ball? The Data tells us that for all the numbers available, the number 7 is the lucky number.



For a bigger chance of winning, we would need to predict more numbers. I tried to see the frequency of the winning numbers separated by the type of lottery. However, The distribution of the Pick 10 winnings is higher on the smaller numbers than larger. To have a better analysis, we would need to isolate the powerball and the megamillions from the pick10.



In the next chart, we separated the megamillions and the powerball from the pick 10. Now that we have a better look at the data, we see a more varied distribution of the winning numbers. Even though the scatter chart shows us a drop in the numbers after 45, we have to take in consideration that the numbers from 1992, stopped at 45 and eventually grew to the 69 numbers used now.



So far, our analysis show that while there are a few numbers that appear more frequently than others, the chances of winning the lottery is still left to chance or probability as we call it in statistics. With all these data, what can we do to it? I think, with the right color and shape... Maybe something beautiful.

