

## **Lottery Numbers - A Data-Driven Narrative on Lottery Randomness**

### **Audience**

The intended audience for this data story is the **New York State Gaming Commission Board of Directors**. They are the primary stakeholders, and although they have a high-level understanding regarding data in a business context (less so with statistics), we will explain all the complex terminologies clearly. They are very familiar with the significance of the lottery data, and their primary concern is maintaining "the public's trust," conforming to regulatory scrutiny, and protecting sustainable revenue. Making a presentation to this audience, as opposed to the public, will allow for richer storytelling related to operational integrity issues and strategic risks, instead of simply repeating myths based on lucky numbers. The Board of Directors is also the only audience with the authority to implement the final call to action, based on whoever manages to be the targeted audience for our data-based proposal.

### **Purpose**

This presentation seeks board support and funding for a proactive public education campaign. This campaign would send our players a clear message: while players may see patterns in winning numbers, the lottery is a fair random game of chance by all statistics. The board needs to leave knowing that the best way to mitigate the danger of public distrust is to recognize those perceptions held by players and then use clear, data-driven rationales to address those perceptions. Moreover, this is not just an informational or educational presentation but a recommendation about a new strategic direction—moving from a reactive to a proactive position on player education and game integrity.

### **Call to Action**

With the information above supporting the objective randomness of the lottery, while acknowledging the patterns people believed to see, we recommend: **Approve a budget for creating and launching a multi-platform public education and transparency campaign.** It's not a marketing initiative

but an investment in the integrity and revenue sources of the New York Lottery, both short and long term. The campaign will focus on creating straightforward, engaging content (like infographics) explaining randomness concepts and advertising the security protocols that ensure you have a fair draw every time. By addressing player superstitions atomically and using the facts as persuasive evidence, we can build trust between players and providers, encourage responsible gaming, and have a stronger defense against the narratives the games may perceive as "rigged."

### **Medium**

Recognizing the importance of a board, the format of choice is a series of eye-catching infographics and data visualizations that we can easily distribute through social media, public service announcements, and community presentations. A clear written report (this paper) supports the infographics and data visualizations to provide context, nuances, and the call to action. Multi-platforms create a digestible and shareable format for further public engagement and impact opportunities.

### **Design**

The design approach we took centers on clarity, impact, and emotion. While we utilized several elements to clarify understanding, we used some Gestalt principles to guide the viewer's eye intentionally.

- **Color:** The use of a professional palette comprised of blues and greys promotes a tone of authority and trustworthiness, and we use a splash of a bright orange contrasting color for a Focal Point to indicate key findings, such as highlighting the highest and lowest frequency numbers in the histogram, which immediately draws the board's attention to the section of the story we are talking about.
- **Text:** Text is kept to a minimum, and a clear font is chosen for screen viewability when it is used. Title fonts are treated as action-oriented headlines (i.e., "Long-Term Averages Confirm Randomness). This was done to support the audience's understanding of the key thing we want

them to take away from the description of each slide and, subsequently, use short bullet points to support the audience's cognitive load.

- **Alignment & Sizing:** Throughout the presentation, I maintained strict grid alignment, consistent use of white space and lots of it, and an organized structure to provide clean and uncluttered aesthetics to put the data at ease. In the bubble chart, the size of the bubble offers a direct proportion to the frequency of the draw, making the correlation an immediate visual cue that is again more easily comprehended than reading data in a table.
- **Spacing & Proximity:** The principle of proximity is integral to the histogram, as the overall structure places the bars close together and invites the reader to see the overall distribution, not just individual number frequencies. Further, small multiples utilize spacing and alignment similarly; consistent usage creates one singular grid. After the board compared the months, they quickly saw no recurring patterns.

### **Ethical Considerations**

- **Data Changes:** The data were only changed structurally as a prerequisite for data analysis. The "Winning Numbers" string was parsed to integers and unpivoted from a wide to long format; no raw numbers were changed, added, or selectively removed - all source data retained its original structural integrity.
- **Legal or Regulatory Guidelines:** The data source came from New York's public data portal and is public record, so legal guidelines are absent. However, all visualizations were developed to align with responsible gaming guidelines, emphasizing the value of randomness laid under threads of pattern spotting versus the endorsement of false patterns that deem gambling strategies.
- **Risks & Mitigation:** The main risk comes from the visuals that someone may perceive as "hot numbers," which could be misrepresented when referring to gambling using poor strategy. This risk is mitigated by how the presentation is framed, immediately after the visuals is a conclusive

representation of randomness with the density plot and long-term average, with the depiction of "hot number" remaining only as statistical noise. The final call to action - an educational campaign- is the ultimate mitigation strategy to curb these risks that were even made known to the public.

- **Assumptions and Filtering:** The only filtering done was the explicit and labelled separation of "Regular" numbers from special "Mega Ball" or "Powerball" numbers. This was an essential analytical decision to guarantee we were comparing like with like, and it was noted in the visuals.
- **Data Sourcing and Verification:** The credibility of the Official New York Lottery data is very high, as it was taken from the official government number data.ny.gov site. This verified, canonical source of New York lottery data was obtained ethically, as the data is public for public usage.