# OSN analytics: Inguiring into Gambling Related Channels On Telegram

**CHEN XIAOWEI, 3200105621**

*Compiled May 19, 2023*

**This report analyzes gambling-related channels on a social media platform. The report describes the data collection methodology and dataset used in the analysis, and presents the findings of the data analysis, including the prevalence of different types of gambling activities, the behavior of users, and the potential risks associated with online gambling. The report concludes with recommendations for increased regulation of online gambling.**

## 1. SUMMARY

For centuries, gambling has persisted as a popular pastime, enjoyed by people across the globe in diverse forms. In the modern era, advancements in technology have enabled the proliferation of gambling activities on various social platforms, where thousands of users participate in various forms of online betting and gaming.

This laboratory study undertakes a comprehensive inquiry into Telegram gambling channels, involving the accumulation of 20,506 messages pertaining to gambling activities, sourced from 1,470 channel participants across 155 distinctive channels on the Telegram platform. The collected messages are methodically archived and sorted for analytical examination. Through conducting frequency and behavior analyses on the data set, this study endeavors to reveal the complex structure of the online Chinese gambling industry and identify the key stakeholders who shape it.

## 2. BACKGROUND

In this section, we delve into the topic of Telegram, a prominent instant messaging platform that has recently gained notoriety for its lax approach towards content moderation. Unfortunately, this has led to a proliferation of illegal activities and criminal behavior on the platform. Through the aid of the Super Index Bot, we were able to uncover over 1,000 Chinese online gambling channels.

The majority of these channels are dedicated to promoting online gambling and seek to lure new gamblers to their platforms. Typically, these channels feature a barrage of messages that highlight the platform's unique features, tutorials, benefits, and contact information. The organizers of these channels



**Fig. 1.** One of the largest online gambling platforms in China.

often make claims of trustworthiness and emphasize the scale of their platform, as well as the privacy and security of bank transactions. They also employ enticing benefits, such as free cash, to attract users.

Another significant category of these channels is focused on providing gambling information in specific areas, such as sports. These channels furnish users with the latest news, analysis, and advice on sports events such as football, basketball, and e-sports, in addition to offering betting guidance.

## 3. DATA COLLECTION METHODOLOGY

The present laboratory study employed a meticulous data collection methodology that utilized the Super Index Bot and custom-built Python scripts. Leveraging the Super Index Bot, we were able to extract a substantial corpus of channel URLs. Subsequently, we employed the Telethon library and an automated Python script to seamlessly join these channels. In order to accommodate the request limit of the Telegram

API, the URLs were stored across five distinct plain text files, facilitating streamlined processing. Once these channels were joined, another custom Python script was utilized to extract the most recent 2000 messages from each channel and subsequently stored in a database, paving the way for further analysis. After eliminating redundant and irrelevant data yielded by the API, 20,506 valid messages contributed by 1470 distinct users are acquired.

The data collection process was carried out over a specific period of time to ensure that the data collected was representative of the channels during that period. To maintain anonymity and confidentiality, no personal information of the users or channels was collected.

This method of data collection allowed for a large amount of data to be collected efficiently and accurately, providing a robust dataset for the subsequent analysis of gambling-related channels on the platform.

## 4. DATASET DESCRIPTION

In total, we managed to collect a large dataset of 20,506 gambling-related messages, from 155 distinct publicly accessible channels. Despite collecting the most recent 2,000 messages from each channel, the total number of valid messages obtained amounted to 20,506. This is primarily due to the prevalence of duplicate and invalid messages within the channels.

**Table 1.** Breakdown of gambling activities on category

| Category | Mentions |
| --- | --- |
| gambling games | 1933(37.2%) |
| betting | 2956(60.0%) |
| lottery | 299 (2.8%) |

**Table 2.** Breakdown of message senders

| Messages Sent | Accounts | Sum of Messages |
| --- | --- | --- |
| >1000 | 4 | 6235 |
| 500–1000 | 0 | 0 |
| 100–500 | 32 | 1329 |
| <100 | ... | 12942 |

Table 1 presents a comprehensive breakdown of messages, categorizing them based on their respective gambling activities. The analysis reveals three primary categories: gambling games, betting, and lottery, accounting for approximately 37%, 60%, and 3% of the total messages, respectively. Within the category of betting, further subcategories will be explored and discussed in the data analysis section of this paper.

Moving on to Table 2, it provides insights into the distribution of message senders based on the number of gambling-related messages they have sent during the data collection period. A more comprehensive discussion will be presented later in this article.

## 5. DATA ANALYSIS AND FINDINGS

In this section, we undertake a meticulous analysis of the content details extracted from the collected messages. The primary objective is to uncover underlying issues and potential scams, while also investigating the terminologies associated with Chinese offshore online gambling. The aim is to shed light on various misconducts prevalent within gambling-related Telegram channels.

A comprehensive set of word clouds was generated, highlighting various aspects related to gambling-related messages. Furthermore, a comprehensive behavioral analysis was conducted on the entirety of the collected messages, aiming to unveil potential patterns and insights into the general discourse surrounding gambling-related messages. The objective was to shed light on the common dialogues and themes prevalent within the realm of gambling-related discussions.

### A. CATEGORIES



**Fig. 2.** Top sports-related categories.

Having previously discussed the three primary categories provided by Chinese offshore gambling platforms, this section will now primarily concentrate on the realm of sports-related gambling activities. In the messages we have gathered, approximately 60% of the content pertains to betting. Furthermore, within this specific category, a significant majority of the betting activities revolve around sports. Notably, basketball and football games occupy a substantial portion of the sports-related bets. This observation suggests that individuals participating in Chinese offshore online betting are primarily sports enthusiasts, with a higher likelihood of being senior sports enthusiasts.

Specifically, we have observed that within the realm of sports, although football betting represents a smaller proportion compared to basketball betting, the betting platform demonstrates a significantly higher level of attentiveness towards football bettors. The platform provides extensive and meticulously categorized options for football wagers, along with concise yet comprehensive betting guidance, including professional sporting analysis. This observation potentially suggests that while the number of individuals placing football bets may be fewer, the potential benefits they can offer to the betting platform are considerably higher. Moreover, it implies that these bettors exhibit a noteworthy level of professionalism.

**Fig. 3.** Top betted-on football teams.

### B. NAMES

Upon examining the senders of the collected messages, the data strongly indicates that the Chinese off-shore online gambling industry exhibits a polarized structure. A significant proportion of messages are attributed to prominent players, indicating their substantial influence within the market. Additionally, another notable portion is contributed by small-scale players who have sent less than 100 messages, suggesting their active presence in the market and limited space for middle players to operate effectively.



**Fig. 4.** Top mentioned contacts.



**Fig. 5.** Top mentioned domains.

### C. DIALOGUES ANALYSIS

Throughout our investigation process, we have identified certain dialogue characteristics that are commonly employed by gambling platforms to entice users. One prevailing characteristic is the consistent use of emojis and visually striking images, often in vibrant colors such as red and orange. Some platforms even resort to explicit depictions of women to captivate users' attention. Additionally, these platforms employ various benefits to allure users, including but not limited to free cash, complimentary points, and other tangible advantages.

### 6. MISCELLANEOUS

In this laboratory study, we conducted a comprehensive investigation on gambling-related Telegram channels. A total of 20,506 messages pertaining to gambling activities were meticulously collected from a diverse pool of 1,470 channel participants spanning across 155 distinct channels within the Telegram platform. Leveraging this extensive dataset, we meticulously carried out a rigorous data analysis procedure to meticulously dissect the information and unveil key terminologies, as well as prominent entities associated with Chinese offshore online gambling.

This laboratory study delves into the intricate process of data crawling and fundamental data analysis of Telegram messages. It accomplishes this by synergistically integrating the telethon library with Python's comprehensive data analysis framework. Furthermore, the study successfully employs advanced visualization techniques, such as generating word clouds, to unveil pivotal aspects present within the messages. These word clouds effectively uncover crucial attributes of Internet Services complaints and the associated merchants, shedding light on significant findings within the research domain.