Play Your Mega Millions

Sta 523 - Final Project

AUTHORS
Yanjiao Yang
Huiying Lin
Hanyue Liang
Ruixin Pang

Introduction

Mega Millions is an American multi-jurisdictional lottery game. Players pick six numbers from two separate pools of numbers - five different numbers from 1 to 70, and one number from 1 to 25 - or select Easy Pick/Quick Pick. You win the jackpot by matching all six winning numbers in a drawing. The drawing is at 11 p.m. Eastern Time every Tuesday and Friday evening, beginning from 2002.

This project uses the lottery data from Texas Mega Millions and designs a shiny app for users to retrieve past drawings and explores statistics via tables and plots. Based on the statistics, users are also allowed to generate new potential winning numbers. The remaining of the write up is structured as follows: the data section will introduce how the data is obtained, followed by the coding logic and design of the app in the method section. The result section will provide an overview of each dashboard in the shiny app and the discussion section will conclude the project as well as propose future improvements.

Lottery Data

This section will elaborate on how the lottery data is obtained. We scrape past drawings of Mega Millions from https://www.texaslottery.com, as in get_lottery.R. Specifically, we first get all the past drawings from 2003 to 2022 and store it in a data frame. To clean the data and facilitate the analysis in the shiny app, we create columns of year, month and day and all the five winning numbers for each drawing. The unit of the Estimated.jackpot column is unified to be million dollars. Since each drawing is linked to a webpage that records the detailed information, we choose to scrape Number.Correct, Prize.Amount, Total.Texas.Winners from each page, save them as a list, and combine with the original data frame. The first 5 rows of the lottery data and the first record of detail information are presented below. The tidy data is saved as lottery.rds for future usage.

Table 1: 1

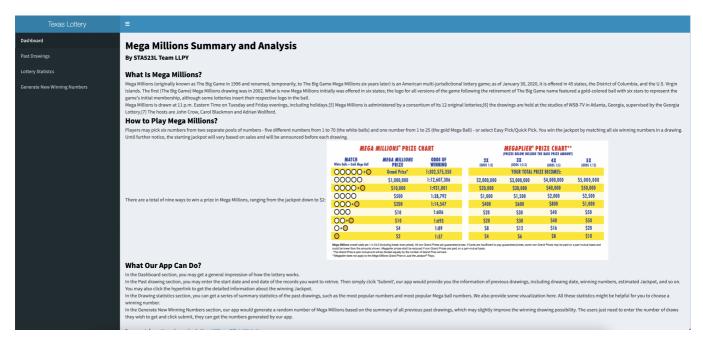
month	day	year	first	second	third	fourth	fifth	Mega.Ball	Megaplier	Jackpot.Winners	Estimated.Jac
12	16	2022	8	35	40	53	56	11	3	Roll	429
12	13	2022	14	22	48	58	68	6	2	Roll	400
12	09	2022	8	19	53	61	69	19	4	Roll	379
12	06	2022	15	16	19	28	47	13	3	Roll	354
12	02	2022	1	21	36	46	52	16	3	Roll	333

Table 2: The first detailed information of the lottery

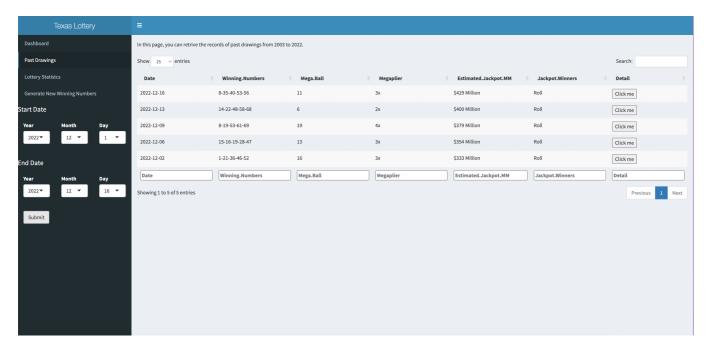
Number.Correct	Prize.Amount	Total.Texas.Winners
5 of 5 w/Mega Ball	\$429 Million	Roll
5 of 5	\$1 Million	0
4 of 5 w/Mega Ball	\$10,000	1
4 of 5	\$500	32
3 of 5 w/Mega Ball	\$200	118
3 of 5	\$10	2,761
2 of 5 w/Mega Ball	\$10	2,586
1 of 5 w/Mega Ball	\$4	21,277
0 of 5 w/Mega Ball	\$2	53,373

Shiny App

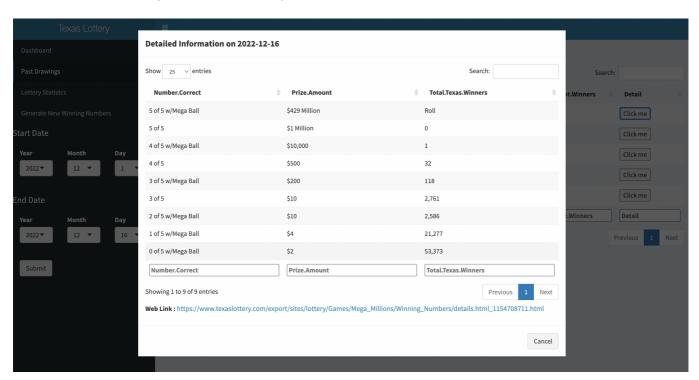
The Shiny App is built using the collected lottery data, as in <code>shinyapp.R</code>. The shiny app includes four sections. In the <code>Dashboard</code> section, a brief introduction of Mega Millions and the app is presented. This page facilitates users to understand the Mega Millions and the function of the shiny app.



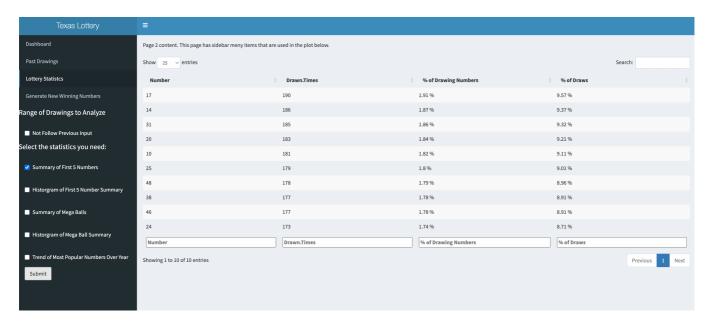
In the *Past Drawings* section, users can specify the start date and end date at the left panel and corresponding past drawings will be shown as a data table after the *Submit* button is clicked. The following figure shows drawing between 2022-12-01 and 2022-12-16.



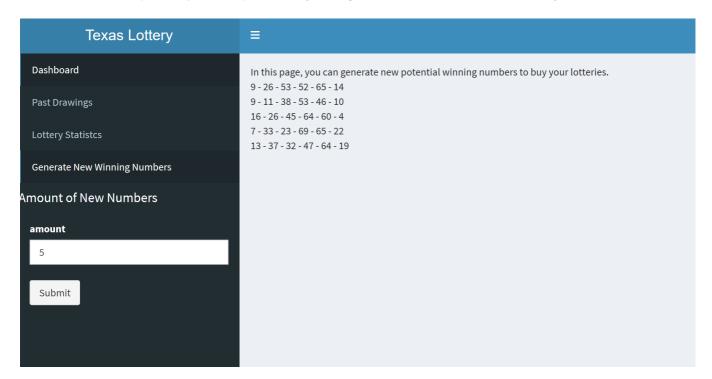
Moreover, a modal dialogue will pop up if users click the button inside the Detail column. The modal dialogue presents users with detailed information on corresponding date and provides a web link to direct them to the original website if they would like to see more information.



As for the *Lottery statistics* section, users are allowed to select interested statistics and results will then be shown in the main panel. The statistics involve the most popular numbers, most popular Mega ball numbers and corresponding visualizations. As shown in the figure below, if the *Not Follow previous input* is not selected, the default statistics would correspond to the time range in the *Past Drawings* section. Otherwise, users can re-specify the start date and end date by clicking *Not Follow previous input* and the statistics will be re-calculated.



In the *Generate New Winning Numbers* section, we generate new random numbers based on the amount of user input, and the probability of each number is slightly modified according to the previous results. We use drawings after 2018 since previous rules regarding range of numbers were different. If a number's frequency is below the first quantile (Q1), we decrease its probability by 20% while if a number's frequency is above the third quantile (Q3), we increase its probability by 20%. Based on this rule, we adjust the probability and generate the numbers randomly.



Discussion & Conclusions

In Nov 7th, 2022, the \$2.04 billion Powerball Jackpot found a winner at California. This massive prize is the world biggest lottery prize right now. If we pay more attention to the lottery field, we may find that in the past five years, there are 5 times that a lottery jackpot crossing the billion-dollar threshold. Motivated by these big amount of lottery jackpot, we wish to design an app to provide some basic information of one of the most famous lottery Mega Millions, retrieve previous drawing records, find related summary statistics, and generate random new numbers.

In our project, we find there is not obvious pattern of the most popular numbers or Megaball numbers, while sometimes certain number does have more appearance than others. Based on this, we could believe the winning number are randomly selected. In our generation of new number section, we based on the appearance percentage of numbers to adjust the probability that one number might be drawn. In our future job, we might add more kinds of lotteries, such as Powerball and so on into our app. In addition, we may continue to digging digger in the possibility that using machine learning to predict the winning numbers.

Reproducibility

We provide the following files:

get_lottery.R: the code for scraping the lottery data

lottery.rds: store the lottery data

shinyapp.R: the code for designing and running the shiny app

project. Rmd: all the write-up for the project

dashboard.png, past.png, dialogue.png, statistics.png, new numbers.png:images for illustrating the shiny app

Run the get_lottery.R to obtain the latest lottery data and the shiny app will be presented by running the shinyapp.R.

References

https://www.megamillions.com/ https://en.wikipedia.org/wiki/Mega_Millions https://www.texaslottery.com/export/sites/lottery/Games/Mega_Millions/index.html https://stackoverflow.com/questions/58821886/adding-buttons-to-shiny-dt-to-pull-up-modal https://stackoverflow.com/questions/42047422/create-url-hyperlink-in-r-shiny