THE LOTTERY LOT

Natsuura Fukuda

DO YOU HAVE WHAT IT TAKES TO WIN A MILLION DOLLARS? (OR TEN)

1

2

3

4

5

6

You can’t win the lottery without buying a ticket

Winning the lottery is commonly understood as a test of luck, one big chance to turn it around. But what if it wasn’t?

What if certain people have statistically higher chances of winning the lottery? This study attempts to find variables that can lead to increase in chance of winning.

Failing to plan is planning to fail

1
2
3
4
5
6

By analyzing and aggregating information about lottery winners and the retailers that sell the tickets, we can find trends and explain aspects that can change odds of winning.

Using graphs and charts to visualize statistics and inspecting each variable to find what works and what doesn’t.

One small purchase for man, one giant loss for the lottery

By identifying what makes up a lottery winner, we uncover what people can do to be more likely to win the lottery.

The lottery will be more evenly distributed between players if everyone has this information.

If this information is kept to few people, certain people could benefit from the extra chance to win the lottery.

Demographics that are more likely to win the lottery may be because of other reasons like vulnerability to gambling.

Data is the new oil

1
2
3
4
5
6

The dataset we will be using is the “Winners List of Texas Lottery Prizes” which lists information about prize winners (not necessarily jackpots, includes small prizes like $1).

There’s 2,856,283 data entries which makes the data very accurate because of its miniscule margin of error.

However, there were many missing values in entries so we had to artificially fill them in which may produce different results.

Due to the massive size of the data, not much can be said with a preliminary analysis but It appears that US Citizens are more likely to win the lottery, and higher scratch ticket prices and MegaMillion seem to yield the highest prize pools

One goal is the starting point of another

1
2
3
4
5
6

The next steps would be to compare the aggregated statistics to real life numbers.

To take into account the proportion of for example, US citizens and non-US citizens to the amount of US citizen lottery winner and non-US citizen lottery winners.

Frequency models will be best for the dataset since most variables are categorical, not numerical.

Thank you