THE LOTTERY LOT

Natsuura Fukuda

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

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

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 the frequencies of each variable, we can tinspect trends and unexpected results to find what could potentially boost chances of winning the lottery.

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.

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

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

Variable	Description	Type
Amount Won	Amount won from the lottery (any prize doesn't have to be jackpot)	Float
Price of Ticket	Price of the purchased ticket	Float
US Citizen Flag	Marks if winner is a US citizen	Categorical
City	City where claimer resides	Categorical
State	State where claimer resides	Categorical
County	County where claimer resides	Categorical
Retail Name	Name of retailer that sold ticket	Categorical
Retail City	City where retailer sold ticket	Categorical
Scratch or Draw	Identifies lottery as a draw game (like PowerBall or MegaMillion) or a scratch-off	Categorical
Date Ticket Sold	The date the ticket was sold	Date

Data is the new oil

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.

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

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.

I could potentially employ a neural network to determine what kind of person has the best chance of winning, identifying variables that aid in lottery winning.

Thank you