

Egypt University of Informatics

Computer and Information Systems

Data Analysis Course

The Analysis of Sports Betting Outcomes

Submitted by: Marwan Khaled: 23-101296, Hany Yasser:23-101107, Mohamed Emad: 23-101018

26/5/2024

**Data Analysis Report: Sports Betting Outcomes**

# Introduction

# This project studies data from sports bets to find out what factors affect whether a bet wins or loses. By using graphs and basic statistical tests, we examine which sports are more profitable, if certain odds make winning more or less likely, and whether betting larger amounts leads to higher returns. The main goal is to give useful insights about what influences success in sports betting and figure out if it is worth to invest your money in or not.

# Research Question:

# How do the type of sport, betting odds, and the amount staked affect the likelihood of winning and the profit or loss from sports bets?

# Hypothesis:

1. Sport Profitability Hypothesis: Is there a difference in profitability (Gross Gaming Revenue) between different sports?  
2. Odds and Winning Probability Hypothesis: Are higher odds associated with a lower probability of winning?

3. Proportion of Winning vs. Losing Bets: Is the proportion of winning bets significantly different from losing bets?

4. Stake Size and Gain Hypothesis: Does staking more money on a bet increase the net gain?

# **Null and Alternative Hypotheses:**

1. Sport Profitability  
 - Null Hypothesis (H₀): There is no significant difference in mean Gross Gaming Revenue (GGR) between different sports.  
 - Alternative Hypothesis (H₁): There is a significant difference in mean GGR between different sports.  
  
2. Odds and Win Probability  
 - Null Hypothesis (H₀): Betting odds are not associated with the probability of winning.  
 - Alternative Hypothesis (H₁): Betting odds are associated with the probability of winning.

3.Proportion of Winning vs. Losing Bets

- Null Hypothesis (H₀): The proportion of winning bets is equal to the proportion of losing bets, so the probability of winning is 0.5.

- Alternative Hypothesis (H₁): The proportion of winning bets is different from the proportion of losing bets, so the probability of winning is not 0.5

4. Stake Size and Gain  
 - Null Hypothesis (H₀): The stake size is not correlated with net gain.  
 - Alternative Hypothesis (H₁): The stake size is correlated with net gain.

# Population of Interest

The population of interest consists of 5000 individuals placing sports bets, as represented in the provided dataset. Each record corresponds to a unique bet made by a user on a particular sport.

# Sampling Method

The dataset contains 100000 bets sampled from a sports betting platform. It is assumed that these bets represent real user activity, without choosing based on user details or types of bets.

# Bias Identification

- Self-Selection Bias: The dataset includes only users who chose to place bets. People who do not bet at all are not part of the dataset; therefore, it doesn’t represent everyone.  
- Reporting Bias: Because the dataset was obtained from Kaggle and the selection process is not clearly described, certain types of bets, such as small bets or specific results, may have been excluded.

- Cofounding Variables: Unmeasured factors such as gambler experience, for example, a bettor expert who has a higher chance of winning compared to a beginner.

# Survey Questions / Collected Data / Dataset

The dataset contains the following features:  
- bet\_id: Unique identifier for each bet  
- user\_id: User identifier  
- bet\_type: Type of bet (single or multiple)  
- sport: Type of sport (e.g., Football, Tennis, Ice Hockey)  
- Odds: Betting odds for the bet  
- is\_win: Whether the bet was won (True/False)  
- stake: The amount of money staked  
- Gain: Amount gained from the bet  
- GGR: Gross Gaming Revenue (stake minus gain)

# Analysis:

## Sport Profitability

A graph with blue bars

AI-generated content may be incorrect.

This bar chart illustrates the total Gross Gaming Revenue (GGR) generated by each sport in the dataset. GGR represents the net amount retained by the betting platform after paying out winnings, effectively showing the overall revenue from bets placed on each sport.

Hypothesis Test:

* - F-statistics: 1.02
* - P-value: 0.427

Decision:

* Fail to reject the null hypothesis.
* There is no significant difference in mean Gross Gaming Revenue (GGR) between sports.

## Odds and Winning Probability

A graph showing the number of the number of individuals

AI-generated content may be incorrect.

This histogram displays the distribution of betting odds for all bets in the dataset, separated by win or loss. Most bets were placed at lower odds, where winning bets are also most frequent. As the odds increase, both the total number of bets and the number of winning bets drop sharply, indicating that gamblers tend to the safe bets with low odds.

Hypothesis Test:

* Chi-squared: 22904.02
* P-value: 0.000

Decision:

* Reject the null hypothesis.
* There is a significant association between betting odds category and the probability of winning.

## Proportion of Winning vs. Losing Bets:

A pie chart with numbers and a percentage

AI-generated content may be incorrect.

This bar chart compares the proportion of winning bets to losing bets in the dataset. If betting outcomes were purely random, we would expect roughly equal proportions of wins and losses. However, the results show a strong imbalance between the two categories.

Hypothesis Test:

* Test: One-sample Proportion Z-test
* Z-statistic: -89.05
* P-value: 0.000

Decision:

* Reject the null hypothesis.
* The proportion of winning bets is significantly different from the proportion of losing bets.

## Stake Size and Gain

A graph showing a graph of a winning bet

AI-generated content may be incorrect.

This scatter plot shows the relationship between stake amount and gains for winning bets. As the stake increases, the gain generally rises, as shown by the upward red regression line. This suggests that higher stakes typically lead to higher winnings, although there is a wide variation in the actual gains for similar stake amounts.

Hypothesis Test:

* Slope= 2.47
* Correlation (r)= 0.591
* P-value=0.000

Decision:

* Reject the null hypothesis
* There is a significant linear relationship between stake and gain among winning bets.

## Conclusion:

This analysis found that while the type of sport did not affect profitability, betting odds strongly influenced win probability, and higher stakes led to higher gains. The win/loss ratio was also unbalanced; winning probability is much bigger. These findings offer useful insights but should be viewed with caution due to potential bias. The goal of our project is to figure out if betting is worth it or not, it is pretty obvious from the data visualization that it is not, people can invest their money in safer ways such as real estate and gold and keep themselves out of the risk.