Sports Odds

Santonastaso

2022-08-01

# Sports Odds

This is inspired by an analysis of whether the Liverpool upset over Barcelona in 2016 was the greatest upset of all time. Here, we evaluate the sports books to see who is the least likely winners in 2016.

The data and story for them comes from : Winston, Wayne. *Danalytics Stories: Using Data to make Good Things Happen.*  Indianapolis, John Wiley & Sons, Inc., 2021.

#Set up our libraries  
library(readxl)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(tidyverse)

## ── Attaching packages  
## ───────────────────────────────────────  
## tidyverse 1.3.2 ──

## ✔ ggplot2 3.3.6 ✔ purrr 0.3.4  
## ✔ tibble 3.1.8 ✔ stringr 1.4.0  
## ✔ tidyr 1.2.0 ✔ forcats 0.5.1  
## ✔ readr 2.1.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

#Establish path and import the spreadsheet  
setwd('C:/Users/tsant/Documents/Data Science/Data\_Stories/Chapter 04')  
#I wrap it in 'as.data.frame' otherwise it comes out as a tibble which isn't very friendly to manipulation  
df <- as.data.frame(read\_excel("Leicester.xlsx", sheet = "RawData"))

The formula for figuring out the odds comes from published moneylines. As with all probabilities, the probability of winning plus the probability of losing must equal 0.

((x - p) \* (-x)) + (p \* y) = 0

where x = the amount to lose y = the amount to win

For the money line, this often is published in $100 USD increments. For example, the moneyline for Arsenal in 2016 was 240. Where a wager of $100 would net $240. The above equation simplifies to:

p = 100/(100+240)

The Moneyline will require a correction of 1.07 as this is the percentage a booking agent adds to ensure a profit.

The underlying equation I’ll be using then is:

### P = (100/(100 + x))/1.07

Where x is the user input which is the moneyline and can be taken from any sporting publication.

If, on the otherhand, the entity is favored to win, the equation will be:

### (x/(100 + x))/1.07

probs <- c()  
  
for(elem in df$Odds){ #iterate through the Odds  
 if(elem > 0) {  
 var <- 100/(100 + elem)/1.07\*100 #this formula if odds are positive  
 probs <- append(probs, var) #build probability vector  
} else if(elem <= 0) {  
 var <- (elem\*-1/(100 + elem\*-1))/1.07\*100 #this formula if odds are negative  
 probs <- append(probs, var)  
}  
}  
  
df1 <- df   
df1['Probabilities'] <- probs #slap probabilities to dataframe

dfsort <- df1[order(df1$Probabilities, decreasing = FALSE), ]  
  
print("The 10 least likely to win teams are:")

## [1] "The 10 least likely to win teams are:"

head(dfsort,10)

## Team Odds Money Line Probability True Prob Probabilities  
## 13 Newcastle 750000 0.0001333156 0.0001245940 0.01245940  
## 15 Aston Villa 750000 0.0001333156 0.0001245940 0.01245940  
## 17 Sunderland 750000 0.0001333156 0.0001245940 0.01245940  
## 18 Bournemouth 750000 0.0001333156 0.0001245940 0.01245940  
## 20 Watford 600000 0.0001666389 0.0001557373 0.01557373  
## 14 West Brom Albion 500000 0.0001999600 0.0001868785 0.01868785  
## 19 Norwich 500000 0.0001999600 0.0001868785 0.01868785  
## 10 Stoke City 250000 0.0003998401 0.0003736823 0.03736823  
## 9 Swansea City 150000 0.0006662225 0.0006226379 0.06226379  
## 16 Leicester City 150000 0.0006662225 0.0006226379 0.06226379

We should notice that Liverpool is not in the top 10. So it isn’t the biggest upset ever…by a long shot.