

A STUDY ON DATA SCIENCE APPLIED TO SPORTING EVENTS: AN EXPLORATORY AND DESCRIPTIVE ANALYSIS

Gregori Alisson de Sena Ramos¹, Paulo Henrique Ferreira da Silva¹

¹Federal University of Bahia

The present work aims to study, analyze and generate insights from the odds of sporting events (more specifically, football matches) in online bookmakers. We will also see how statistical fallacies can arise in the middle of our analyzes and how we are exposed to random and high-impact events.



gregori_sr@hotmail.com



4th Conference on
**Statistics and
Data Science**
Salvador, Brazil (online)
December 1-3, 2022

Data extraction and methods used

The work makes use of a database of more than 1.1 million records of football matches from around the world, dated from 2003 to 2022. The data were collected through a web crawler.

And the methods used were:

- C#, SQL Server, Python(Pandas, Matplotlib e Numpy)
- Return on Investment (ROI)
- House Edge
- Brier Score



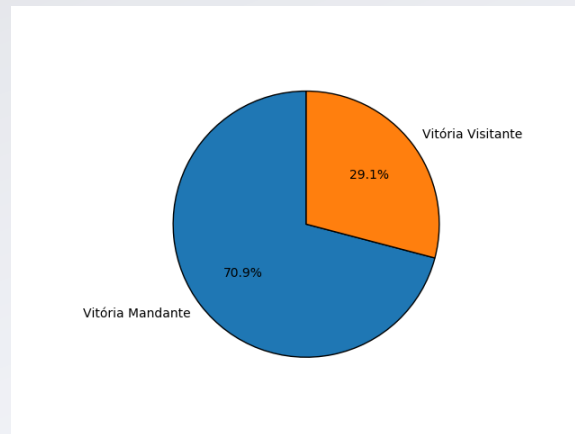
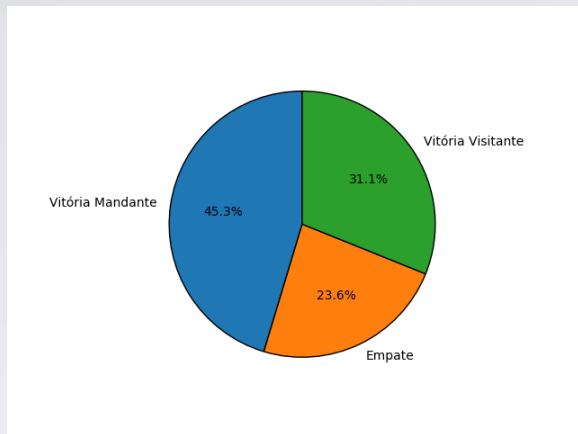
gregori_sr@hotmail.com



4th Conference on
**Statistics and
Data Science**
Salvador, Brazil (online)
December 1-3, 2022

Results

- The home win bias



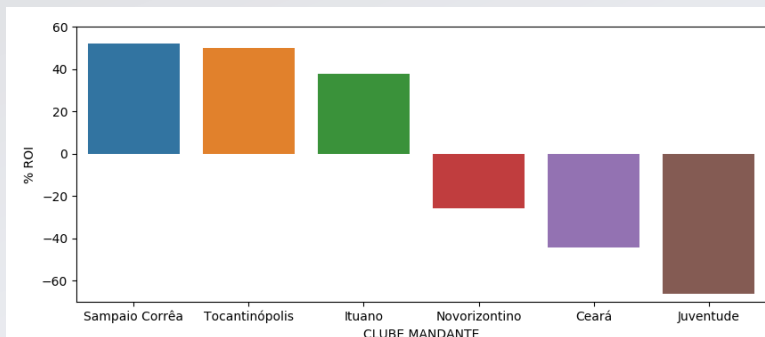
gregori_sr@hotmail.com



4th Conference on
**Statistics and
Data Science**
Salvador, Brazil (online)
December 1-3, 2022

Results

- Sports betting and traditional investing



CDB BANCO MÉDIO	8,26%
TESOURO SELIC	7,16%
CDB BRANCO GRANDE	5,50%
POUPANÇA	4,95%
DIVI	3,97%
TESOURO PREFIXADO	3,79%
IBOVESPA	3,65%
IFIX	0,75%
TESOURO IPCA	0,48%
SMLL	-8,81%
DÓLAR	-10,32%
OURO	-12,45%
IVVB11	-20,01%
BDRX	-21,68%
BITCOIN	-54,32%
ETHEREUM	-55,48%



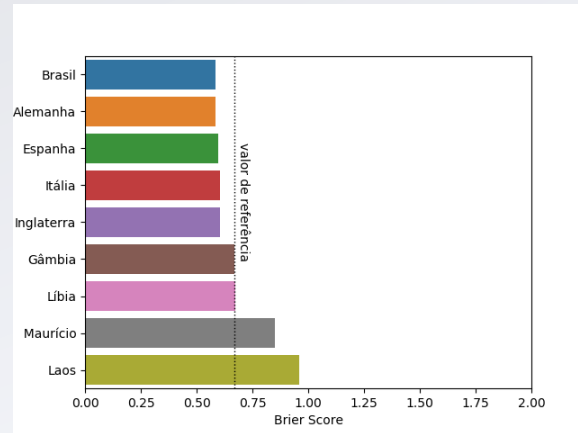
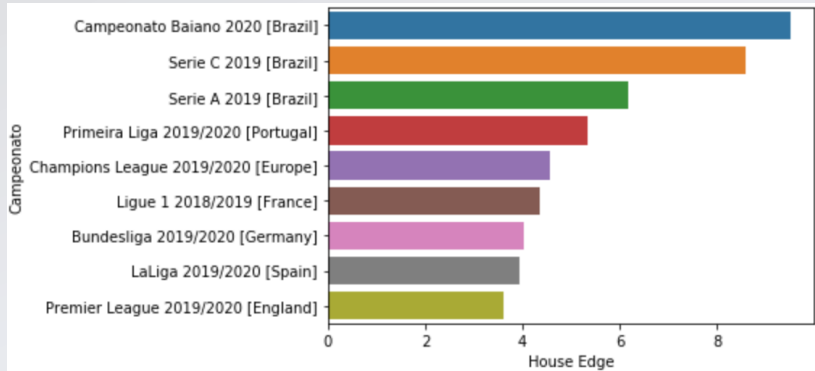
gregori_sr@hotmail.com



4th Conference on
**Statistics and
Data Science**
Salvador, Brazil (online)
December 1-3, 2022

Results

- How bookmakers make money and Brier Score and the Wisdom of Crowds

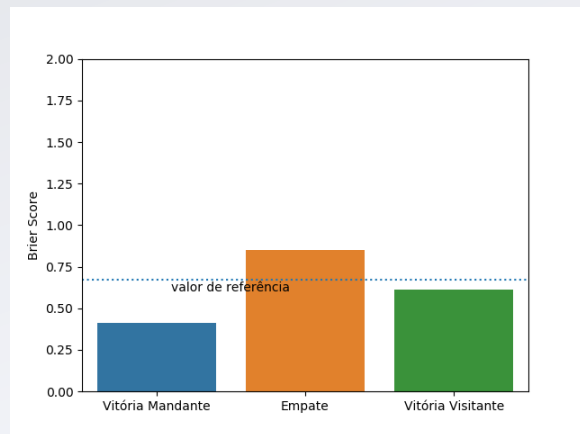
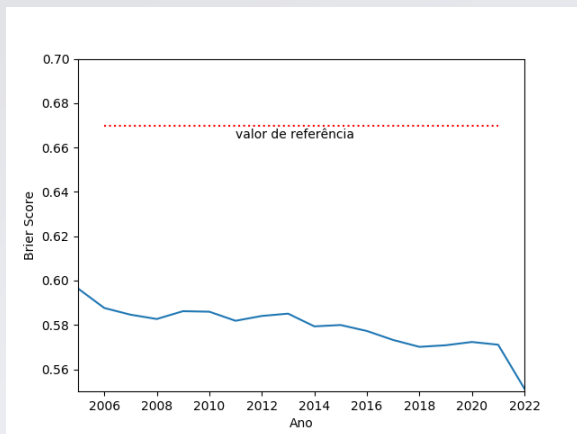


gregori_sr@hotmail.com



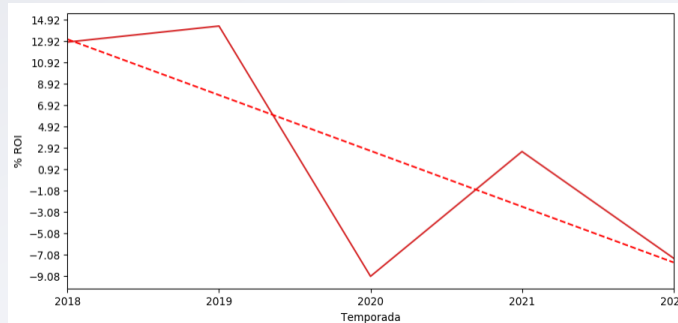
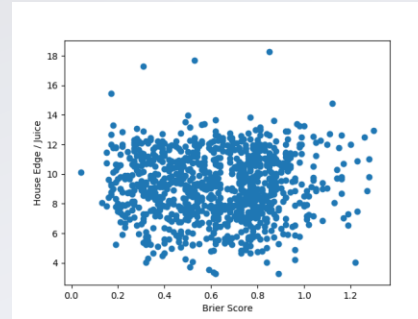
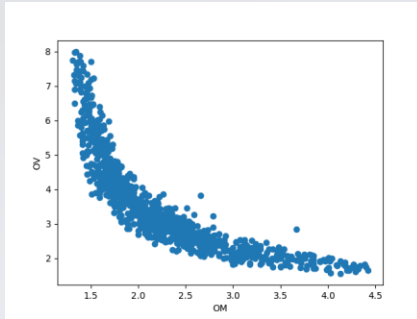
Results

- Brier Score and the quality of quotes



Results

- False Causality Fallacy and Gambler's fallacy



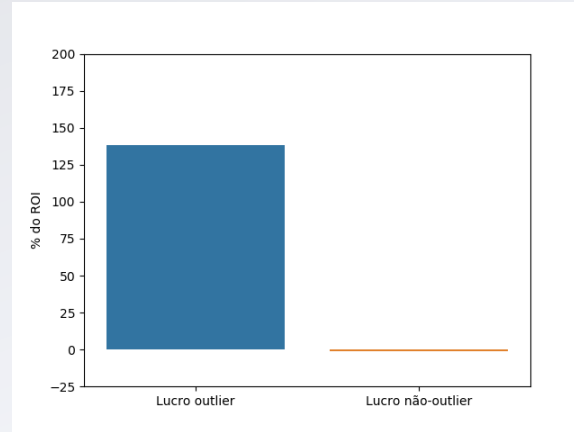
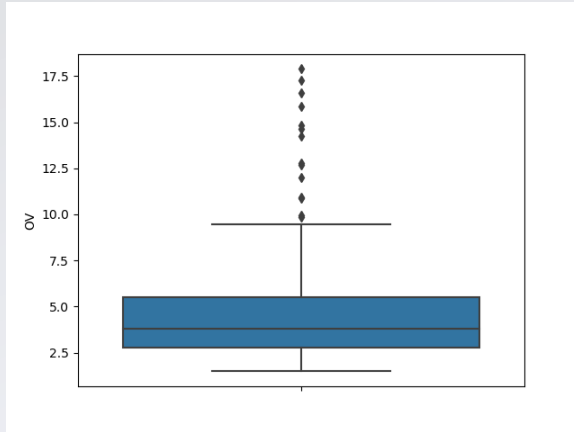
gregori_sr@hotmail.com



4th Conference on
**Statistics and
Data Science**
Salvador, Brazil (online)
December 1-3, 2022

Results

- Black Swan Logic: The Impact of the Highly Improbable



gregori_sr@hotmail.com



4th Conference on
**Statistics and
Data Science**
Salvador, Brazil (online)
December 1-3, 2022

Conclusions

Sports betting, like other statistical events, also deals with uncertainty; however much the individual tries to predict the future, there will always be a dose of uncertainty (or randomness). That way, one can be exposed to black swans. The next possible steps, using the exploratory and descriptive analyzes of this work, would be the development of predictive models considering the quotations (odds) as predictive variables.



gregori_sr@hotmail.com



4th Conference on
**Statistics and
Data Science**
Salvador, Brazil (online)
December 1-3, 2022