

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

RAFAEL SOARES IANNINI

**ANÁLISE DE DADOS SOBRE DESEMPENHO E
APOSTAS ESPORTIVAS NO FUTEBOL EUROPEU**

Belo Horizonte

2024

Identificação da base de dados

A base de dados escolhida para o desenvolvimento do trabalho foi a de *Odds* (probabilidade atribuída a diferentes resultados de um evento, como vitória de um time, empate ou derrota) de apostas esportivas em partidas do futebol europeu. A base de dados possui mais de 25 mil partidas de 2003-2024. Ela possui os atributos:

- Saison(String) Temporada correspondente a partida(e.g., 2003-2024),
- Championat(String) Nome do campeonato(e.g., “france/ligue01”),
- Date(String) Data da partida no formato “dd MMM YYYY”,
- Heure(String) Horário da partida,
- Équipe Domicile(String) Nome do time mandante,
- Score Domicile(int) Gols marcados pelo time mandante,
- Score Extérieur(int) Gols marcados pelo time visitante,
- Équipe Extérieur(String) Nome do time visitante,
- Cote 1 (int) *Odds* para vitória do time mandante,
- Cote X (int) *Odds* para o empate,
- Cote 2 (int) *Odds* para a vitória do time visitante,
- Gangnant(int) Indicação do resultado: 1 (mandante), 2 (visitante), ou 0 (empate).

Link para a base de dados original:

<https://www.kaggle.com/datasets/rayenjllassi/more-than-20k-footballsoccer-match?resource=download>

Os dados foram tratados com a retirada de linhas com valores nulos, e das colunas de nomes e quantidade de gols das equipes, temporada, data e hora que não serão utilizados durante o trabalho. Apresentação de amostra dos dados tratados:

	Championat	Cote 1	Cote X	Cote 2	Gagnant
0	france/ligue-1	1.45	3.80	5.60	1.0
1	france/ligue-1	2.94	2.61	2.45	2.0
2	france/ligue-1	1.95	3.50	2.95	1.0
3	france/ligue-1	1.47	3.43	6.20	1.0
4	france/ligue-1	2.24	3.45	2.49	1.0

As colunas deixadas de campeonato, *Odds*, e resultados serão utilizadas para a definição de problemas e análise exploratória de dados.

Definição dos problemas

Qual a relação entre as *Odds*(Cote 1, Cote X, Cote 2) e os resultados reais das partidas?

Quais campeonatos possuem maior e menor assertividade entre a *Odd* vencedora (menor multiplicador) e o resultado real da partida?

Existe um padrão entre o desempenho (quantidade de vitórias) das equipes mandantes e visitantes?

Qual campeonato possui a maior e menor porcentagem de empates?

Análise exploratória de dados

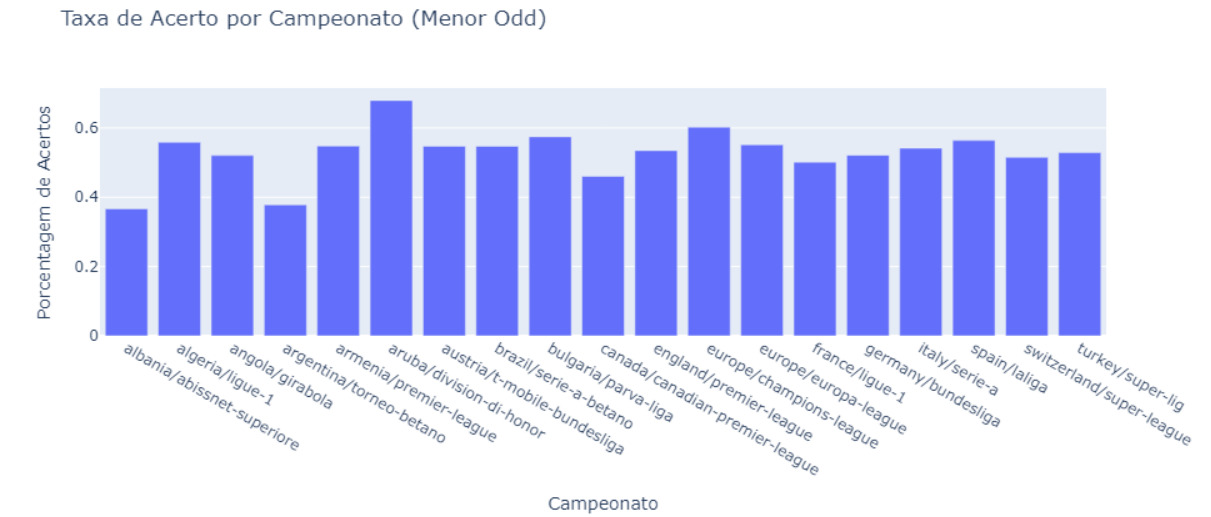
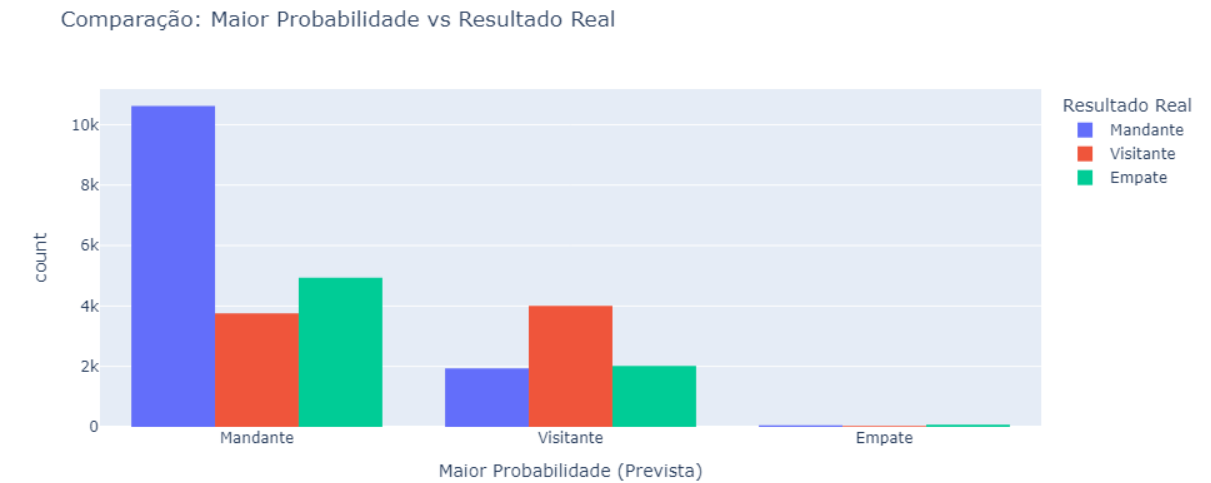
Em média 53% das vezes a *Odd* cotada como mais provável foi o resultado correto da partida. E a maior parte dos erros acontecem quando a *Odd* mais provável era de mandante e o resultado ser empate (38.79%) e a *Odd* mais provável ser mandante e o resultado ser visitante (29.53%).

O campeonato que possui maior assertividade entre a *Odd* vencedora e o resultado real é o Aruba / *Division Di Honoraruba* acertando 67.94% das vezes o resultado correto. Já o campeonato que possui menor assertividade é o Albania / *Abissnet SuperioreAlbania* com apenas 36.67% de assertividade entre a *Odd* vencedora e o resultado real.

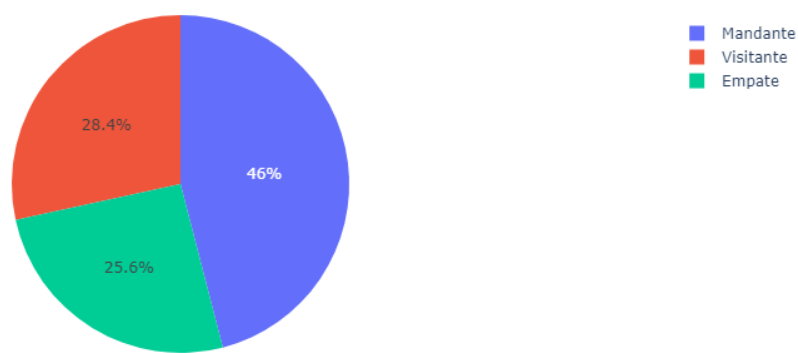
Em média levando em conta todos os campeonatos, 46% das partidas terminam com o time mandante vencedor, 28,4% com o time visitante vencedor e apenas 25,6% em empate. O único campeonato onde a porcentagem de vitória do visitante é maior que a do mandante é o Aruba / *Division Di Honoraruba* com 45,9% de vitória pro visitante e 43,1% pro mandante.

O campeonato com maior porcentagem de empates é o Argentina / *Torneo Betano* com 41,18% dos jogos terminando em empate, e o campeonato com menor é o Aruba / *Division Di Honoraruba* com apenas 11% dos jogos terminando em empate.

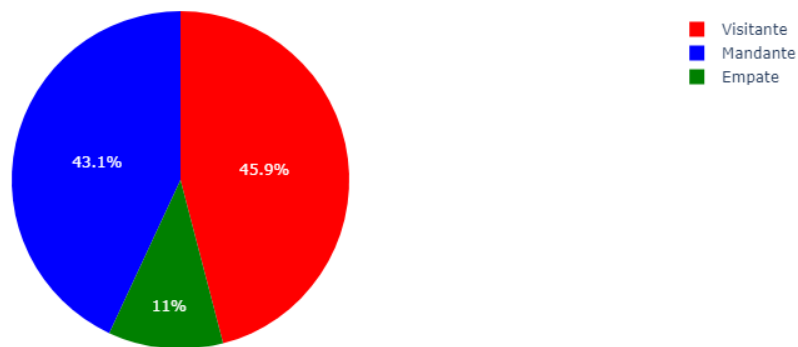
Visualização dos dados



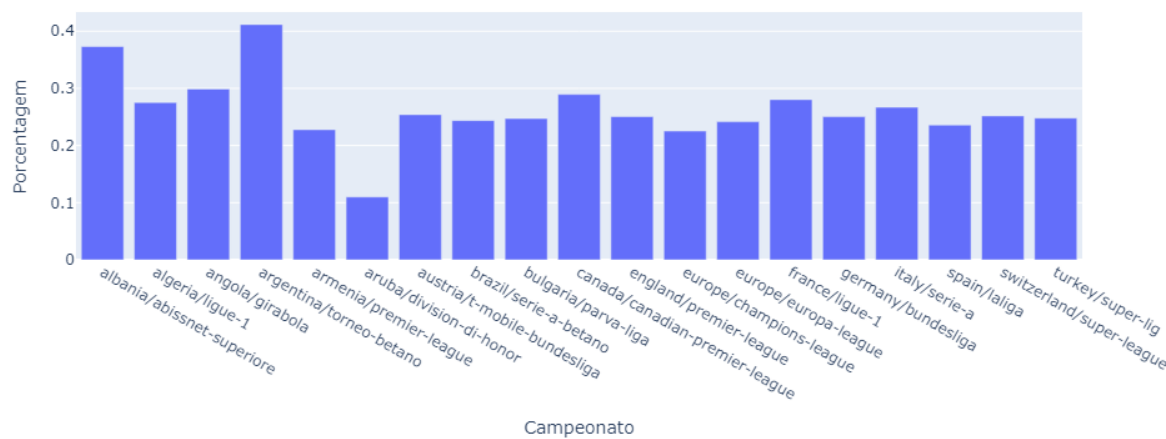
Distribuição de Resultados: Mandante, Visitante e Empate



Distribuição de Resultados: aruba/division-di-honor



Porcentagem de Empates por Campeonato



Análise crítica e conclusão

A partir da análise dos dados é possível perceber em quais campeonatos faz mais sentido apostar a favor da *Odd* (apostar em quem a *Odd* acha que irá vencer), já que ela normalmente vai estar correta. E em quais campeonatos é mais vantajoso apostar contra a *Odd* (apostar em quem a *Odd* acha que irá perder), já que muitas das vezes ela estará errada.

Além disso é possível perceber em quais campeonatos vale a pena apostar no empate que normalmente tem uma *Odd* mais consistente, já que em campeonatos como o Argentina / *Torneo Betano* mais de 40% dos jogos terminam em empate.

Assim como que na maioria dos campeonatos apostar a favor do time mandante faz mais sentido já que em média mais de 40% dos jogos eles vencem contra 28,6% dos visitantes. Apesar de alguns campeonatos específicos essa diferença ser bem menor ou até os visitantes vencerem mais, como é o caso do Aruba / *Division Di Honoraruba*.

Todas as imagens e os códigos que foram feitos em Python e Jupyter Notebook com pandas para tratar e visualizar os dados e plotly para criar os gráficos estão no github: <https://github.com/rafaeliannini/Trabalho-Ciencia-de-Dados>