

Analise da performance de jogos em Casa vs Fora

Felipe Schafhauser Lubke
GRR 20233918
03-Dez-2023 · Curitiba/PR/Brasil

- Analise da performance de jogos em Casa vs Fora
 - * Introdução
 - * Os dados
 - * Metodologia
 - * Análise Exploratória
 - * Teste A/B
 - * Conclusões
 - * Recomendações e próximos passos
 - * Tabelas
 - * Glossário

Introdução

O futebol é um esporte dinâmico, com vários fatores afetando o jogo de diversas formas e uma das que mais se fala é em relação ao fator "casa", mas será mesmo que temos uma grande predominância no aspecto casa, ou isso se aplica apenas para alguns times ou momentos?

Os dados

Foi selecionado o site <https://fbref.com/>, para ser feito o web scraping, o motivo de sua seleção foi a variedade de dados contidos e não ser um site fechado para a captura dos dados, inicialmente foi selecionado a tabela de dados gerais, como vitórias, empates, gols a favor e contra, além de algumas outras métricas já fornecidas pelo site, outra tabela selecionado foi a diferença entre os jogos em casa e fora, que também incluem os mesmos dados citados anteriormente. Os dados foram coletados no início da 37 rodada de 2023, faltando 2 rodadas para finalizar o campeonato e pode ser refeito assim que o campeonato acabar.

Metodologia

O ambiente utilizado para trabalhar com o Python foi o PyCharm C.E., com os pacotes, "requests", "BeautifulSoup", "statistics", "numpy", "scipy", "pandas" e "seaborn". Com os dados já coletados, foi iniciado um processo de pré-visualização e higienização desses dados, esse processo se deu para poder eliminar colunas que não seriam utilizadas e corrigir possíveis problemas identificados. Esse processo foi realizado para nas duas extrações realizadas, a primeira extração com os dados de classificação dos times sem levar em conta nenhuma segmentação, a segunda extração contendo os mesmos dados sendo segmentados por jogos em casa e fora. Os dados contidos nas tabelas podem ser conferidos no

Glossário. Todas as colunas retiradas e ajustes realizados podem ser acompanhados no decorrer do código.

Análise Exploratória

Após o processo de scraping e higienização do data-frame, pode-se verificar alguns insights para progredir com a análise. Inicialmente o principal foco do estudo é a diferença entre a quantidade vitórias comparando com a posição do time, foi tomado esse ponto de partida para entender se temos nos times mais bem posicionados um valor maior nesse indicador. Com uma comparação descritiva entre vitórias, temos que a variação de vitórias em casa é a menor, outra questão é a média de vitórias, temos 9 times acima da média para vitórias em casa e fora de casa. Enquanto fora de casa se tem 5 vitórias em média no campeonato, em casa temos 8 vitórias, 3 vitórias a menos, e ao relacionar com a colocação do time percebemos que, os 11 primeiros times mais bem posicionados estão acima da média de vitórias em casa, mas em relação a vitórias fora temos um comportamento diferente, times em diferentes posições estão acima da média.

Posição X Vitórias

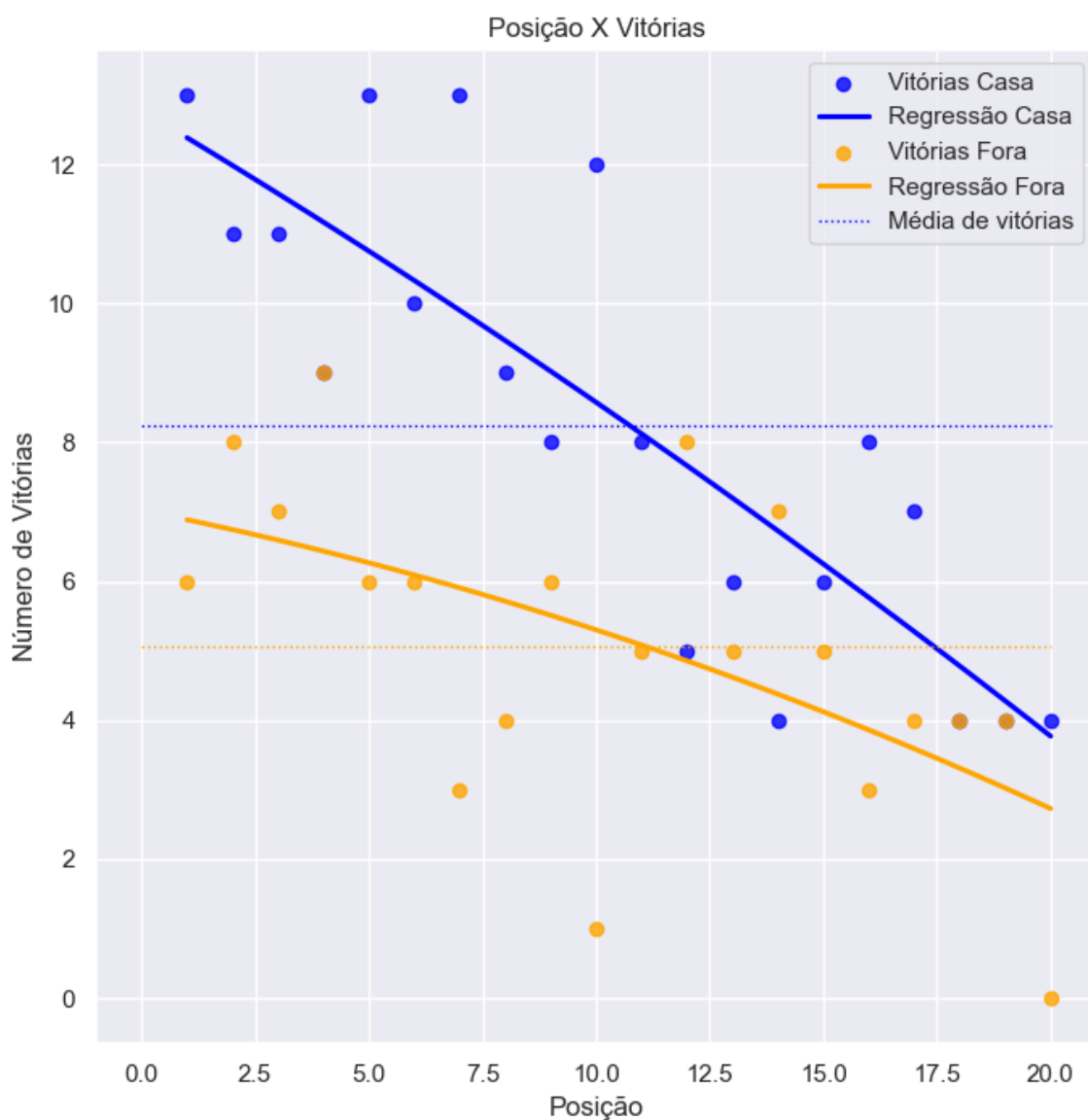


As vitórias em casa caem após a 11ª posição, enquanto as fora de casa variam mais.

Aplicamos uma regressão de ordem 2 para entender melhor o comportamento dos dados, e percebemos de forma visual que as vitórias em casa realmente têm uma declividade bem mais acentuada que em relação a vitórias fora. Com as vitórias de casa tendo um desvio de 4, temos uma variabilidade bem maior já que as vitórias fora de casa têm desvio de 2, assim fica mais claro que vencer em casa acaba sendo

um diferencial. Outro ponto analisado para colaborar com a importância de vencer em casa é a correlação de -0,84 com a posição, a correlação das vitórias fora com a posição é de -0,57.

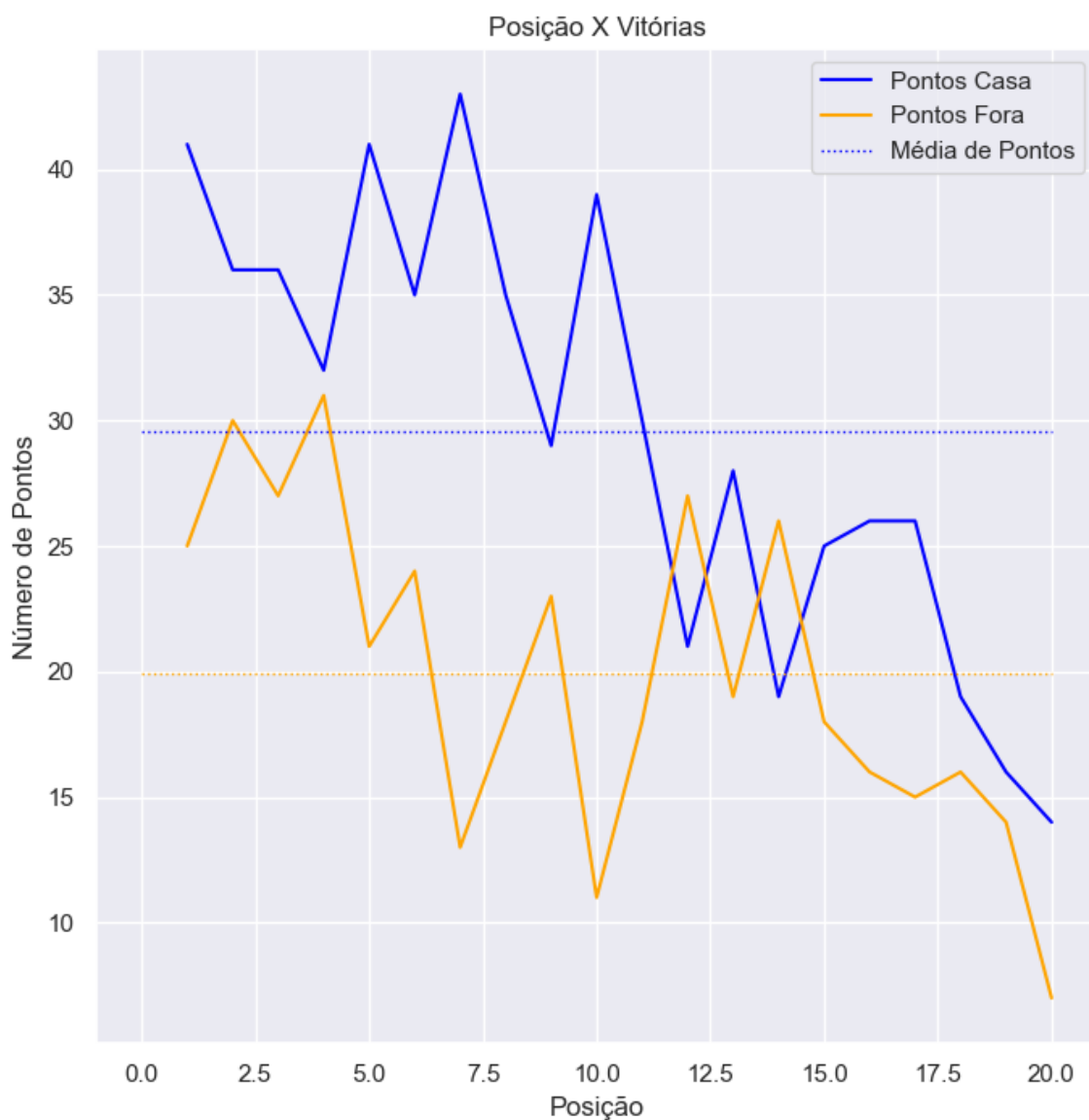
Posição X Vitórias com Regressão



Com maior variabilidade a regressão de vitórias em casa tem uma declividade bem mais acentuada.

Fizemos as mesmas análises considerando a pontuação, afim de verificar se temos o empate com um fator decisivo ou que possa alterar a dinâmica já observada. De início o que já chama atenção é que a variação, quando olhamos apenas em casa, a pontuação tem 8,7 enquanto vitórias tem 8,3. Ao olhar fora de casa, a variação muda mais, pontuação tem 6,5 e vitórias tem 5. Outro padrão também já identificado é que os 11 primeiros times mais bem posicionados estão acima da média de pontos em casa, e em relação a pontos fora temos o mesmo comportamento de diferentes posições estarem acima da média.

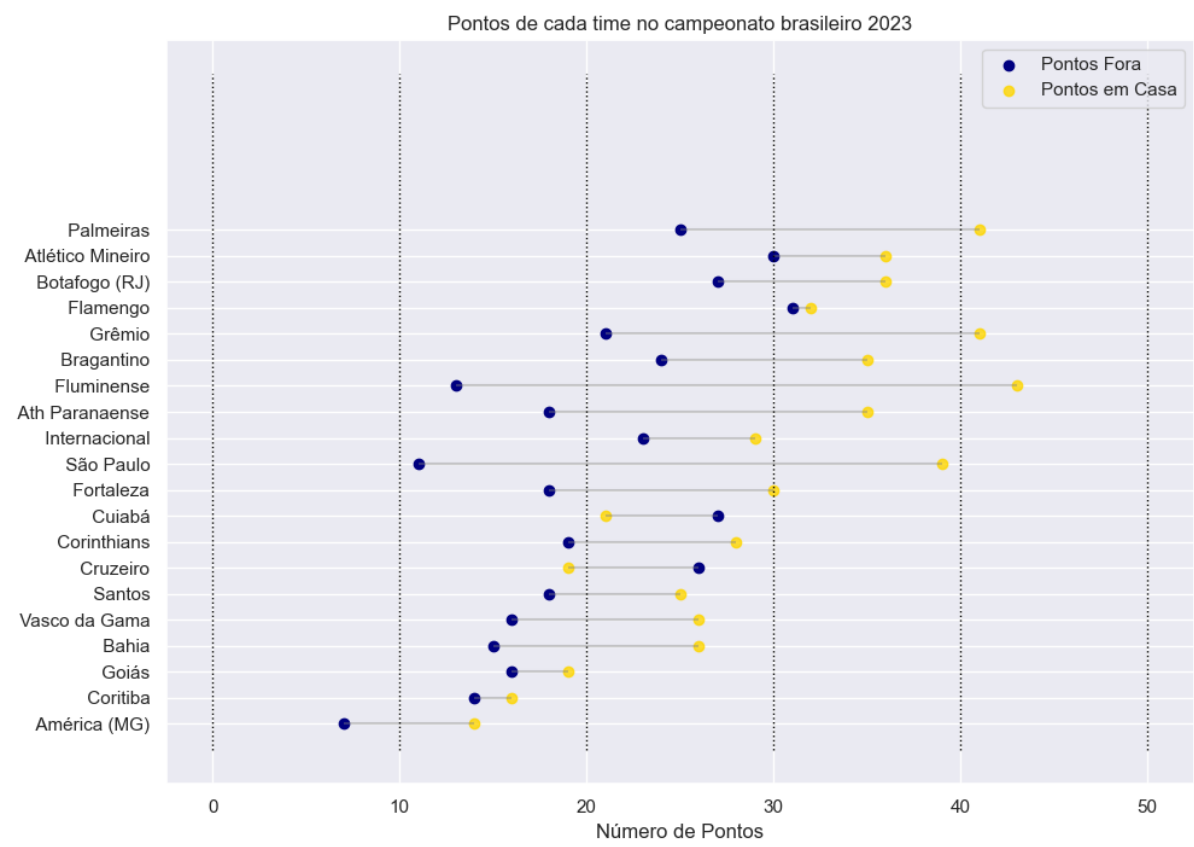
Posição X Pontos



Da mesma forma que o gráfico anterior, temos o mesmo comportamento que as vitórias, demonstrando que empates não são um bom resultado.

Trazendo o time de forma individual, mas ainda categorizado por sua classificação, vemos o quão predominante é a pontuação em casa para os primeiros colocados, de forma que alguns times como o Fluminense e o São Paulo são 2 dos 3 piores times fora de casa e são o melhor e 4º melhor, respectivamente, pontuando em casa, por conta disso conseguiram ainda ficar na metade de cima da tabela.

Pontos por Time

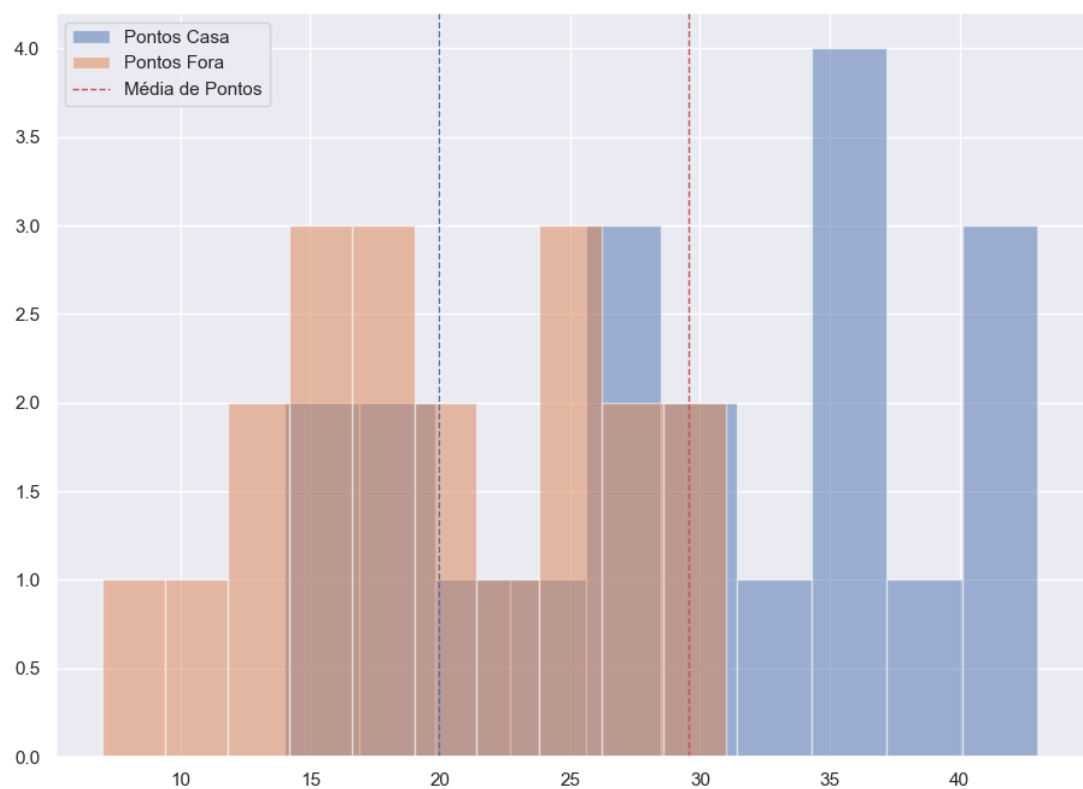


Da metade para baixo do gráfico não temos times com mais de 30 pontos conquistados em casa, mas da metade para cima temos alguns times com menos de 20 pontos conquistados fora de casa.

Teste A/B

Vamos utilizar o teste A/B para confirmar a hipótese de que os jogos em casa geram mais pontos, para isso vamos ter a versão A sendo pontos em casa e a versão B sendo pontos fora de casa. Com isso temos o valor de p em 0, indicando que foi observado diferenças estatísticas entre as duas versões testadas, pontos em casa e pontos fora. O valor positivo de t, indica que a versão A (pontos em casa) é estatisticamente superior a versão B (pontos fora de casa), como o valor de t ficou em 3,9, temos que é provável que seja estatisticamente significativo.

Teste A/B



Conclusões

O time precisa corresponder em casa para evitar posições muito inferiores na tabela, ao menos ficar próximo da média da pontuação e para isso ocorrer é necessário que os jogos sejam ganhos, o empate em si não traz tantos benefícios para compor a pontuação, mas é claro que acaba colaborando como um todo no resultado final mas nunca deve ser o objetivo. Fora de casa é um desafio para todos os times, mas fica claro que para disputar as primeiras colocações é necessário ter uma excelente performance.

Recomendações e próximos passos

- Analisar os times de forma individual;
- Analisar a capacidade do estádio;
- Analisar média de público;
- Correlacionar a média de público com a performance do adversário;
- Analisar distância viajada para jogar;
- Analisar distância a data da viagem para jogar;

- Tabelas
- Glossário

Referência bibliográfica

FBREF. **Série A estatísticas**. Disponível em: <https://fbref.com/pt/comps/24/Serie-A-Estatisticas>. Acesso em: 03 de Dez. de 2023 as 22:30.

