**ANÁLISE EXPLORATÓRIA FUTEBOL**

Elisa de Oliveira Soares, Gabrielly Esteves Pinheiro Chácara,   
João Vitor Tomaz Alves Ferreira, Luiz Eduardo Bravin, Nícolas Mateus Spaniol

**RESUMO**

*Neste artigo analisamos dados do futebol europeu, incluindo informações detalhadas sobre partidas, clubes, jogadores e eventos ocorridos em campo extraídos do site Transfermarkt e disponibilizados publicamente no Kaggle. Procuramos responder, usando Python e bibliotecas de visualização e manipulação de dados, à hipóteses desenvolvidas pelo grupo acerca do dataset. Exploramos correlações entre o desempenho de jogadores e seu custo, entre a performance dos clubes em partidas e o fato destes estarem jogando dentro/fora do país e as distribuições dos jogadores por mês de nascimento e de cartões recebidos por posição em campo. Com relação ao mês de nascimento, vimos que a quantidade de jogadores no dataset cai linearmente ao longo do ano, com dezembro tendo cerca de metade dos jogadores de janeiro. A distribuição dos cartões, similarmente, se concentrou nas posições defensivas. Descobrimos que, em média, clubes que vendem e depois compram um mesmo jogador ficam com saldo positivo, e tem o intervalo entre as transações de três anos. Quanto ao desempenho dos jogadores e clubes, ….*

**Palavras-chave:** Futebol, Python, visualização de dados, gráfico, estatística, análise de dados

**1 INTRODUÇÃO**

Por meio do uso da linguagem Python, foram analisados diferentes aspectos dos times e jogadores, com base em alguns fatores que podem ou não influenciar o desempenho dos mesmos.

**2 DESENVOLVIMENTO**

**2.1** **Há uma diferença na performance dos times quando estes jogam em casa, como visitantes e fora do país?**

Ddjfhte

**2.2** **Quais foram as compras de jogadores com melhores e piores custo-benefício registradas?**

Ggreryd

**2.3 A posição dos jogadores em campo influencia na quantidade de cartões que estes recebem?**

Para analisar essa hipótese, utilizamos as tabelas *“game\_events.csv”* e *“game\_lineups.csv”*. Foi feita uma tabela com os cartões por posição em que cada cartão, amarelo ou vermelho, foi colocado na posição de acordo com a posição em que o jogador estava atuando no jogo em quelevou o cartão. Posições que tinham uma quantidade insuficiente de dados (menos de 150 cartões em todos os jogos) foram removidas da análise.

Primeiramente, foram filtrados apenas os eventos de recebimento de cartões, e a posição em que o jogador estava atuando no jogo

Foram feitas dois gráficos para uma visualização mais precisa dos dados: um gráfico que mostra a quantidade total de cartões recebidos por cada posição em campo (Figura 1), e um que mostra a proporção de cartões vermelhos e amarelos por posição (Figura 2).

**Figura 1 –** Gráfico de cartões por posição

Uma imagem com texto, captura de ecrã, diagrama, Tipo de letra

Descrição gerada automaticamente

Fonte: elaborado pelo autor

**Figura 2 -** Proporção de cartões por posição

Uma imagem com captura de ecrã, texto, amarelo, Retângulo

Descrição gerada automaticamente

Fonte: elaborado pelo autor

Enfim, pudemos observar que as posições defensivas, como Centre-Back, Defensive Midfield, Left-Back e Right-Back são as que mais tomam cartão. Enquanto isso, as que menos recebem cartões são mais bem espalhadas, contendo pessoas de todas as áreas do jogo, como o Second Striker (ataque), Goalkeeper (defesa), Left Midfield e Right Midfield (ambos do meio de campo). Já sobre a proporção de cartões, é possível ver que o goleiro é o que recebe uma maior porcentagem de cartões vermelhos.

**2.4** **Vendas e compras posteriores de um jogador por um mesmo time costumam gerar lucro para o time?**

Para a análise dessa hipótese, usamos as tabelas *“players.csv”* e *“transfers.csv”*. Primeiro, conferimos se um jogador foi vendido por um time e posteriormente comprado pelo mesmo time. Em seguida, analisamos se, após a sequência de transações, o time teve lucro ou prejuízo com esse jogador.

Foram feitos dois gráficos para a análise mais precisa desses dados: Um boxplot com o saldo final do time, para visualizar qual a tendência em operações de compra pós venda (Figura 3), e um boxplot do intervado entre a venda e a compra (Figura 4). Foram calculadas também medidas de tendência, como a mediana do saldo (R$ 225.000,00), o desvio padrão do saldo (R$ 13.198.979,93), e as médias da idade de venda (23,1), da idade de compra (26,3) e do intervalo ente a venda e a compra (3,2) .

**Figura 4 –** Saldo final do time

Uma imagem com file, diagrama, captura de ecrã, Retângulo

Descrição gerada automaticamente

Fonte: elaborado pelo autor

**Figura 4 –** Intervalo entre venda e compra

Uma imagem com Retângulo, file, diagrama, captura de ecrã

Descrição gerada automaticamente

Fonte: elaborado pelo autor

Assim, podemos observar que os times costumam ter um pequeno lucro após a venda e posterior compra de jogadores, mas isso não é algo certo, visto que muitos times podem ter pequenos prejuízos, grandes prejuízos ou até mesmo grandes lucros, visto que o desvio padrão é alto.

**2.5** **Jogadores com preço fora do comum tem o desempenho proporcional?**

Para realizar a análise proposta, primeiro definimos os conceitos que utilizamos, que são: preços fora do comum e desempenho do jogador.

Com as colunas “*date*” e “*market\_value\_in\_eur*” da tabela “*player\_valuations.csv*”*,* criamos o conceito de preços fora do comum. Primeiramente, corrigimos os valores de “*market\_value\_in\_eur*” de acordo com a inflação decorrida de *“date”* até os dias atuais. Em seguida, calculamos a média ponderada dos preços dos jogadores ao longo de suas carreiras, considerando os pesos iguais ao tempo, em dias, que o jogador permaneceu com aquela cotação. Após isto, calculamos o limite superior – segundo a fórmula LS = Q3 + 1,5DI, onde DI = Q3 – Q1 e Q1 e Q3 são, respectivamente, os primeiro e terceiro quartis – da coluna *“mean\_price”* da tabela *“performance.csv”* onde os dados foram inseridos. Finalmente, os jogadores considerados com preço fora do comum são aqueles cujo preço é maior ou igual a LS.

Para definir o desempenho do jogador, usamos as colunas “*yellow\_cards*”, “*red\_cards*”, “*goals*”e “*assists*” da tabela “*appearances.csv*”

**2.6** **Pessoas que nascem na primeira metade do ano tem mais chance de se tornarem jogadores profissionais?**

O jornalista britânico Malcolm Gladwell, em seu livro Outliers: The Story of Success (2008), escreveu sobre a relação entre o mês de nascimento dos jogadores de Hoquei no Canadá e sua ascenção no esporte, e percebeu que uma tendência dos jogadores profissionais nascerem majoritariamente na primeira metade do ano.

Com base nisso, nos questionamos se esse padrão é válido também para outros tipos de esportes. Para realizar essa análise, utilizamos a coluna *“date\_of\_birth”* da tabela *“players”* para criar uma nova tabela com a quantidade e a frequência de jogadores que fazem aniversário por mês. Com esses dados, foi feito um gráfico das frequências por mês (Figura 5) e também o cálculo do desvio relativo dos aniversários, que é de 653,32.

**Figura 5 –** Frequência de jogadores que fazem aniversário por mês

Uma imagem com texto, captura de ecrã, Saturação de cores, diagrama

Descrição gerada automaticamente

Fonte: elaborado pelo autor

A partir da análise do gráfico e do desvio relativo, foi possível perceber que, assim como no Hóquei canadense, há uma relação entre os meses de aniversário mais frequentes no futebol europeu. De acorco com Gladwell, essa relação tem causa na forma que os jogadores são selecionados para os treinos e cluber infantis, desde a infância. Essas seleções tem, em sua maioria, limite de idade para se inscrever, e crianças que nasceram e janeiro e dezembro de um mesmo ano apresentam grandes diferenças de desenvolvimento.

**3 CONSIDERAÇÕES FINAIS**

REGRrygxcbgszdegsr

**REFERÊNCIAS**

**GLADWELL, Malcolm.** Outliers: The Story of Success. 1. ed. Nova York: Little, Brown and Company, 2008.

**CARIBOO, David.** Player Scores. Kaggle, 2023. Disponível em: <https://www.kaggle.com/datasets/davidcariboo/player-scores>. Acesso em: 02 out. 2024.