

Relatório de Análise - Data Science VNL 2024

Projeto Data Science: analisando os dados da Liga das Nações de vôlei feminino 2024

Introdução

Neste projeto, analisamos os dados da Liga das Nações de Vôlei Feminino 2024 (VNL 2024) com o objetivo de identificar os principais fundamentos que contribuíram para o desempenho das seleções participantes e destacar as jogadoras mais eficientes da competição. Como atleta de vôlei, trago minha vivência esportiva para interpretar os dados e validar as conclusões.

Objetivos

- O que os dados revelam sobre o desempenho das seleções do pódio?
- Quais foram as jogadoras mais eficientes?
- Quais fundamentos (ataque, bloqueio, saque, defesa, recepção) se destacaram?
- Como cada país performou nos fundamentos?
- Os dados estão de acordo com a seleção do campeonato?
- Como estão distribuídas as posições em quadra?
- Visualizar os países participantes em um mapa interativo.

Repositório do Projeto

URL do Repositório: [<https://github.com/Jordana-Metzler/Data-Science.git>]

Dataset Utilizado

Fonte:

O conjunto de dados foi extraído do site [<https://www.kaggle.com/datasets/santibruno/volleyball-nations-league-women-2024>], contendo as estatísticas individuais das jogadoras da VNL 2024.

Variáveis Principais:

- jogadora: Nome da jogadora
- pais: País da seleção
- posicao: Posição (líbero, oposta, ponteira, etc.)

- ataque, bloqueio, saque, defesa, recepcao: Estatísticas por fundamento
- pontos_totais: Soma de pontos por jogadora

Transformações:

- Criação da coluna jogadora_completa = (nome + posição + país);
- Tradução dos valores das posições;
- Ranqueamento das top 10 jogadoras na pontuação geral;
- Ranqueamento das top 5 jogadoras por fundamento;
- Mapeamento geográfico com folium;
- Criação de um gráfico com regressão para prever desempenho por idade;

Modelo desenvolvido:

- Embora o foco principal tenha sido análise exploratória, foi aplicada uma técnica de regressão linear para estimar o **desempenho técnico das jogadoras com base na idade**.

Detalhes do Modelo:

- Tipo: Regressão Linear Simples
- Biblioteca utilizada: scikit-learn
- Objetivo: Estimar como a idade influencia o desempenho total em pontos

Etapas da modelagem:

1. Seleção das variáveis: idade (X) e pontos_totais (y)
2. Treinamento do modelo de regressão linear
3. Geração de gráfico com linha de tendência para representar a estimativa
4. Interpretação do coeficiente: impacto esperado da idade no rendimento

Análise: Relação entre Idade e Desempenho (Pontos Totais)

Com base na regressão linear aplicada aos dados da Liga das Nações Feminina de Vôlei 2024, obteve-se a seguinte equação:

$$\text{Pontos Totais} = -0.06 \times \text{Idade} + 5.88$$

O coeficiente angular negativo indica uma leve tendência de queda no desempenho à medida que a idade aumenta. No entanto, essa influência é extremamente pequena — cerca de 0,06 pontos a menos por ano — o que a torna estatisticamente irrelevante para fins práticos. Isso sugere que a idade, por si só, **não é um fator determinante no desempenho ofensivo** das atletas nesta competição. Jogadoras jovens e experientes apresentam desempenhos semelhantes em termos de pontuação, indicando que **outros fatores**, como **posição, tempo em quadra, condicionamento físico e papel tático**, exercem maior influência nos resultados individuais.

Análise: Os dados confirmam o pódio da VNL 2024?

Com base nas estatísticas da Liga das Nações de Vôlei Feminino 2024, os dados confirmam o merecido pódio composto por **Itália, Japão e Polônia**. A **Itália** apresentou o desempenho mais equilibrado em todos os fundamentos, enquanto o **Japão** liderou com folga os fundamentos defensivos, que caracterizam o estilo de jogo, e a **Polônia**, por sua vez, teve um desempenho sólido especialmente nos bloqueios.

Já a **seleção brasileira**, apesar de figurar entre as melhores em **ataques e saques**, teve um baixo aproveitamento nos **fundamentos defensivos** o que comprometeu seu rendimento, resultando na **4ª colocação**.

Nos destaques individuais, **Egonu (ITA)** lidera em pontos, **Inoue e Kojima (JPN)** se sobressaem na recepção e defesa, e **Korneluk (POL)** no bloqueio. A análise mostra que **consistência nos fundamentos** foi o diferencial decisivo para o sucesso das equipes no torneio.

Comparando a Seleção Ideal com a Seleção Oficial da VNL 2024

Além da análise por país e fundamentos, também é possível observar os destaques individuais da competição. A Federação Internacional de Voleibol (FIVB) seleciona, ao final do torneio, uma **seleção do campeonato**, composta pelas atletas que mais se destacaram tecnicamente e taticamente em suas posições.

Neste projeto, com base nos dados estatísticos da VNL 2024, construímos uma **seleção ideal baseada em desempenho quantitativo**.

A seguir, comparamos essa seleção ideal com a **seleção oficial divulgada pela FIVB**, refletindo sobre as semelhanças e diferenças entre desempenho numérico e impacto real dentro da competição.

Conclusão da Comparação dos Dream Teams

A análise mostra que, em grande parte, a **seleção oficial da VNL 2024 coincide com os dados** estatísticos. Nomes como **Paola Egonu, Alessia Orro, Sarah Fahr e Agnieszka Korneluk** aparecem em ambas as seleções, confirmando seus desempenhos de destaque.

As principais **diferenças** aparecem nas escolhas de pontreiras: enquanto os dados apontam para **Gabi Guimarães (BRA)** e **Arisa Inoue (JPN)** como mais eficientes em pontuação e regularidade, a seleção oficial optou por **Myriam Sylla (ITA)** e **Sarina Koga (JPN)**.

Essa comparação reforça que **a performance estatística é essencial**, mas **a escolha da seleção oficial também leva em conta aspectos subjetivos** e qualitativos, como equilíbrio tático, protagonismo em jogos decisivos e contribuição coletiva para o sucesso da equipe.

Limitações do Estudo e Trabalhos Futuros

Embora o projeto tenha alcançado resultados relevantes, algumas limitações devem ser consideradas:

- **Modelo simplificado:** A regressão linear foi aplicada considerando apenas a variável idade como preditora, ignorando fatores como posição, tempo de jogo, número de sets jogados e estilo tático de cada seleção.
- **Falta de normalização por tempo de quadra:** Jogadoras com maior tempo de jogo naturalmente acumulam mais pontos, o que pode enviesar os rankings.
- **Dados ausentes:** Algumas estatísticas complementares, como erros, eficiência percentual e acertos por tentativa, não estavam disponíveis no dataset original.

Trabalhos Futuros

- **Análise multivariada** incorporando variáveis como posição, altura, tempo de jogo e número de partidas.
- **Clusterização de jogadoras** com base em características de jogo para identificar estilos e perfis técnicos.
- **Modelos preditivos mais avançados**, como regressão múltipla, árvores de decisão ou redes neurais.
- **Estudos longitudinais**, comparando edições anteriores da VNL para observar evolução de desempenho.