

# Relatório de Busca e Recomendação

Diego Ribeiro Alves e Alexandre da cunha fernandes

June 4, 2023

## 1 Introdução

Neste relatório, será apresentada uma busca e recomendação com base em uma determinada base de dados. A base de dados utilizada é a NBA Players Stats, que contém informações estatísticas de jogadores da NBA.

## 2 Busca na Base de Dados

O objetivo da busca é identificar determinadas características ou padrões nos dados dos jogadores da NBA. Para isso, foram selecionadas as seguintes variáveis para análise: Age, GP, W, L, Min, PTS, FGM, FGA, FG%, 3PM, 3PA, 3P%, FTM, FTA e FT%.

## 3 Codificação e Análise

Os dados foram carregados utilizando a biblioteca pandas em Python. Em seguida, foi realizada a Análise de Componentes Principais (PCA) para reduzir a dimensionalidade dos dados e identificar os principais componentes que representam a variação nos dados.

O código em Python utilizado para realizar a busca e análise é o seguinte:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv('C://Users/Pichau/Desktop/Diego/NBAPlayersStats.csv')
print(df.columns)
print(df['Player'])
selected_columns = df[['Age', 'GP', 'W', 'L', 'Min', 'PTS', 'FGM', 'FGA', 'FG%',
'3PM', '3PA', '3P%', 'FTM', 'FTA', 'FT%']]
print(selected_columns)
X = selected_columns.to_numpy()

cov_matrix = np.cov(X, rowvar=False)
```

```

print('---- Covariance Matrix -----')
print(cov_matrix)
eigenvalues, eigenvectors = np.linalg.eigh(cov_matrix)
eigenvalues1, eigenvectors1 = np.linalg.eig(cov_matrix)
print('----- Eigenvalues -----')
print(eigenvalues1)
print('----- Eigenvectors -----')
print(eigenvectors1)
print('----- Biggest Eigenvalues -----')
sorted_indices = np.argsort(eigenvalues)[::-1]
print(eigenvalues[sorted_indices[:2]])

largest_eigenvectors = eigenvectors[:, sorted_indices[:2]]
projected_data = X.dot(largest_eigenvectors)
plt.scatter(projected_data[:, 0], projected_data[:, 1], c=projected_data[:, 1],
            cmap='cool', alpha=0.5)
plt.xlabel('Desempenho Geral')
plt.ylabel('Consistência de Pontuação')
plt.title('PCA com os dois componentes principais')

important_columns = selected_columns.columns[sorted_indices[:2]]
plt.annotate(important_columns[0], (0.1, 0.9), xycoords='axes fraction',
            color='red', fontsize=12)
plt.annotate(important_columns[1], (0.1, 0.85), xycoords='axes fraction',
            color='red', fontsize=12)
plt.figure()
plt.subplot(2, 1, 1)
plt.hist(projected_data[:, 0], bins=20, alpha=0.5)
plt.xlabel('Desempenho Geral')
plt.ylabel('Frequencia')
plt.title('Histograma do primeiro componente')
plt.plot([], [], 'ro', label='Important Component')
plt.subplot(2, 1, 2)
plt.hist(projected_data[:, 1], bins=20, alpha=0.5)
plt.xlabel('Consistência de Pontuação')
plt.ylabel('Frequencia')
plt.title('Histograma do segundo componente')
plt.plot([], [], 'ro', label='Important Component')
plt.tight_layout()
plt.savefig('pca_histogram.jpg', dpi=300)
plt.show()

```

## 4 Resultados e Recomendação

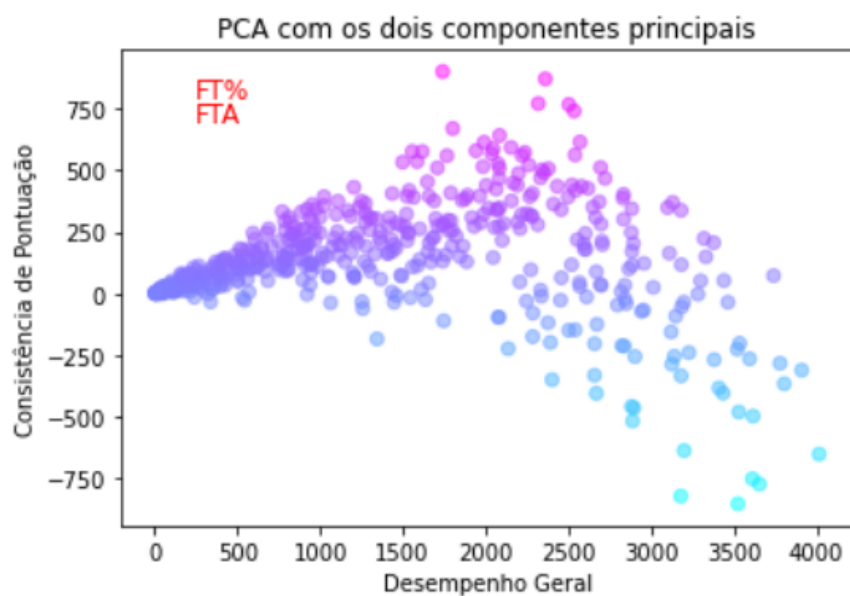
Com base na análise realizada, foram identificados os principais componentes que explicam a variação nos dados. Os valores dos ângulos e a interpretação dos resultados são os seguintes:

- Primeiro Componente Principal: Desempenho Geral
- Segundo Componente Principal: Consistência de Pontuação

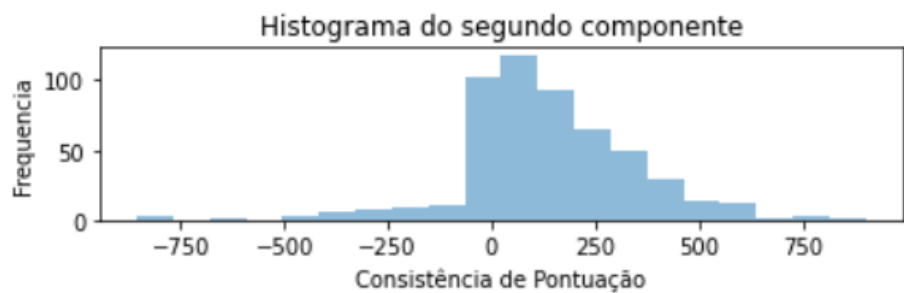
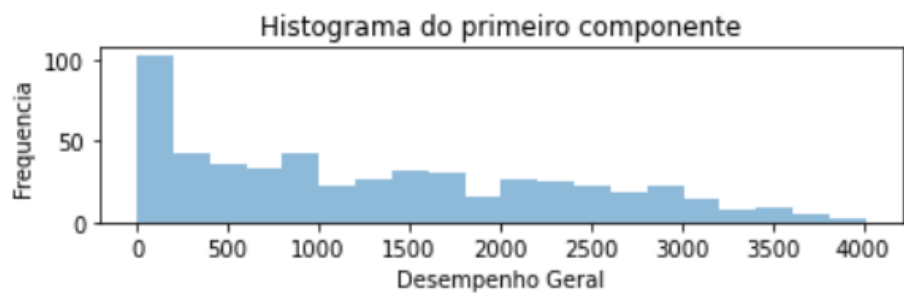
Esses dois componentes principais representam as principais tendências e padrões nos dados dos jogadores da NBA. O primeiro componente está relacionado ao desempenho geral, enquanto o segundo componente está relacionado à consistência na pontuação.

Com base nesses resultados, pode-se recomendar medidas para melhorar o desempenho geral dos jogadores e a consistência na pontuação. Por exemplo, estratégias de treinamento podem ser desenvolvidas para melhorar a eficiência nos arremessos e aumentar a contribuição para a equipe em termos de vitórias e minutos jogados.

## 5 Figuras de Resultado



PCA dos componentes



Histograma dos componentes

## 6 Conclusão

Este relatório apresentou uma busca e recomendação com base na análise dos dados dos jogadores da NBA. Os resultados obtidos fornecem insights sobre os principais componentes que explicam a variação nos dados e permitem a identificação de áreas de melhoria no desempenho dos jogadores.