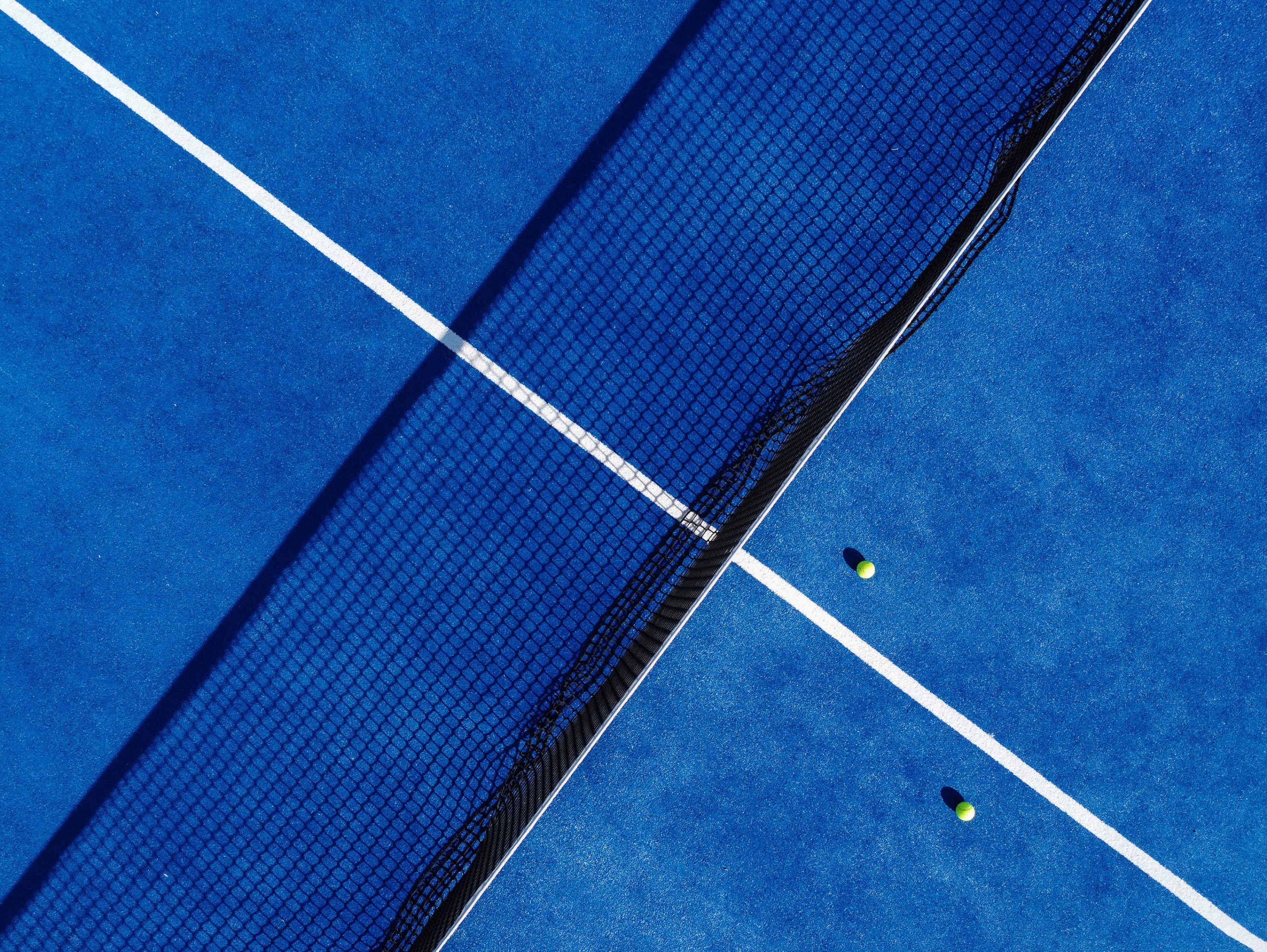
**A picture containing font, text, graphics, logo

Description automatically generated**

**PROJETO APLICADO EM CIÊNCIA DE DADOS I -**

**RELATÓRIO**

Licenciatura em Ciência de dados

**Base de Dados ATP Players -**

**Brasil**

Grupo 5:

Eliane Susso Efraim Gabriel, Nº 103303

Marco Delgado Esperança, Nº 110451

Maria João Ferreira Lourenço, Nº 104716

Umeima Adam Mahomed, Nº 992397

Docentes: Diana Aldea Mendes, Sérgio Moro

19 de abril de 2023

**Índice:**

[1. Introdução 3](#_Toc132377810)

[2. Business Understanding 3](#_Toc132377811)

[ATP Tour 3](#_Toc132377812)

[Terminologia 3](#_Toc132377813)

[3. Data Understanding 4](#_Toc132377814)

[Jogadores 4](#_Toc132377815)

[Variável Altura 4](#_Toc132377816)

[Variável País de naturalidade 4](#_Toc132377817)

[Variável Destro ou Esquerdino 4](#_Toc132377818)

[Variável Data de nascimento 4](#_Toc132377819)

[Jogos 4](#_Toc132377820)

[Variável Nome do jogo? 4](#_Toc132377821)

[Variável Data 4](#_Toc132377822)

[Variável Tipo de piso 4](#_Toc132377823)

[Variável Prémio monetário 4](#_Toc132377824)

[Variável Round 4](#_Toc132377825)

[Variável Número de sets 4](#_Toc132377826)

[4. Data Preparation 4](#_Toc132377827)

[5. Modelling 4](#_Toc132377828)

[6. Evaluation/ Deployment 4](#_Toc132377829)

[7. Conclusões 4](#_Toc132377830)

[8. Bibliografia 4](#_Toc132377831)

# Introdução

Objetivo do trabalho

Como vamos fazer isso: Metodologia CRISP-DM

# Business Understanding

## ATP Tour

Com o objetivo de prever o número de sets para a conclusão de um jogo de ténis profissional no Brasil, começamos por perceber o contexto do problema.

A Associação de Tenistas Profissionais, *Association of Tennis Professionals* (*ATP*), é a entidade responsável pelos circuitos de ténis profissionais masculinos, o ATP Tour, ATP Challenger Tour e ATP Champions Tour.

Esta foi constituída em setembro de 1972 por Donald Dell, Jack Kramer e Cliff Drysdale com o objetivo de proteger os interesses dos tenistas profissionais. A denominação desta organização foi mudada ao longo dos anos, passando por ATP Tour, ATP e por fim retomou para o nome original que prevalece atualmente.

A organização está sediada em Londres, tendo as suas filiais noutros continentes: na América em Ponte Vedra Beach, no estado norte americano na Flórida, na Europa tem a sua filial no Mónaco e a internacional que cobre Ásia, África e Australásia em Sidney. Esta é gerida pelo atual diretor executivo Massimo Calvelli e uma equipa de seis membros.

Os torneiros mais prestigiosos na modalidade desportiva ténis são os de Grand Slam, ou *majors*- Australian Open, Roland Garros (ou French Open) e Wimbledon e US Open, que não são sob a alçada de ATP, no entanto, são atribuídos pontos do *ranking* ATP.

Relativamente aos torneios que ocorrem no Brasil, o mais famoso é o Rio Open, que teve início em 2014 e ocorre anualmente, apesar de não ter ocorrido em 2021, devido à pandemia. Este é considerado o maior torneio de tênis da América do Sul e pertence ao ATP 500, logo é um torneio bastante prestigiado a nível internacional, onde participam os grandes nomes do ténis a nível mundial.

Também podemos mencionar o Brasil Open que se realizou de 2001 a 2019, apenas para atletas masculinos e ocorreu na Costa do Sauipe, São Paulo e Rio de Janeiro e pertencia ao ATP Tour 250; o ATP São Paulo que ocorreu de 2001 a 2012, também só para atletas masculinos, sendo substituído pelo Rio Open.

As competições do tipo BRA vs. X são jogos que ocorrem na Davis Cup, nomeadamente na World Group Play-Offs, onde jogadores de países diferentes se defrontam.

Por fim, podemos referir as competições que fazem parte da *ITF Men’s Circuit*: Brazil Fx - apenas para atletas masculinos e estão divididas em diferentes níveis consoante o nível dos atletas e prémio, sendo principalmente para os jogadores ganharem experiência e melhorarem os seus rankings -, M15 São Paulo - só para atletas masculinos, cujo prémio é de 15000$ e ocorre em São Paulo, permitindo o melhoramento dos rankings e a qualificação para o ATP Challenger Tour e ATP World Tour - e M25 Rio de Janeiro - com as mesmas características que a última, só que o prémio é de 25000$ e ocorre no Rio de Janeiro.

Relativamente a alguns nomes do tênis brasileiro podemos apontar Gustavo Kuerten, Thiago Monteiro, Thomaz Belluci, Thiago Alves e Ricardo Mello.

## Terminologia

Esta modalidade apresenta termos técnicos que têm que ser compreendidos para garantir a criação de um bom modelo que responda ao objetivo proposto, para isso estes vão ser explorados de forma breve.

Relativamente ao sistema de pontuação, temos os termos “*point*”, “*game*”, “*set*” e “*match*”. De forma simples, o primeiro jogador a ganhar quatro pontos (“points”) ganha um “*game*”, e o primeiro jogador a ganhar seis “*games*” ganha um “*set*”, dependendo do tipo de torneiro, o primeiro que ganhar dois ou três “*sets*” ganha o “*match*”.

O primeiro ponto é chamado de “15”, o seguinte de “30”, “40” e por fim “*game*”.

Referente ao tipo de piso, este será explorado futuramente, mas adiantamos que existem quatro tipos principais: relva artificial, barro/argila artificial, cimento e borracha, sendo o último considerado o melhor.

# Data Understanding

usar só para ver os que podemos dizer

O nosso dataset tem 1308835 linhas e 16 colunas

## Jogadores

### Variável Altura

### Variável País de naturalidade

### Variável Destro ou Esquerdino

### Variável Data de nascimento

## Jogos

### Variável Nome do jogo?

### Variável Data

### Variável Tipo de piso

### Variável Prémio monetário

### Variável Round

### Variável Número de sets

MOSTRAR EXPLORAÇÃO DOS DADOS

# Data Preparation

Os passos que seguimos para a limpeza dos dados seguem a seguinte ordem:

### Uniformização da variável Prize

### Criação das variáveis DateStart e DateEnd

### Valores omissos

### Imputação dos valores omissos na variável DateEnd

### Substituição de valores omissos na variável Height

### Substituição de valores omissos na variável GameRank

### Renomear a variável GameRank

### Criação da função que associa o jogador ao seu rank por ano

### Criação da função que busca o valor do ranking e da coluna RankPlayer

### Uniformização do Score

### Criação da coluna NumberSets

### Campo \_id

### Criação da nova variável City a apartir da variável Location

### Importação do dataset Worldcities

### Duplicados de cidades em países diferentes

### Valores omissos em Country

### Países com escritas diferentes

### Formas diferentes de escrita de Brasil

### Uniformização da escrita para “Brasil”

A partir daqui, podemos trabalhar apenas com os dados para a o Brasil, uma vez que já obtivemos uma base de dados mais coerente a respeito do país em que nos iremos focar.

### Valores omissos

# Modelling

# Evaluation/ Deployment

# Conclusões

# Bibliografia

[ATP. (s.d.).](https://d.docs.live.net/932760ed81e76c92/Ambiente de Trabalho/PACDI - PROJETO/PACDI - Projeto_semanal/ATP. (s.d.). History. Obtido de atptour: https:/www.atptour.com/en/corporate/historyATP. (s.d.). Senior Leadership Team. Obtido de ATP: https:/www.atptour.com/en/corporate/managementHadlich, G. (s.d.). How Does Tennis Scoring Work?%20(By%20Former%20PRO).%20Obtido%20em%2013%20de%2004%20de%202023,%20de%20mytennishq:%20https://mytennishq.com/tennis-scoring-rules-the-ultimate-guide-explained/Kothari,%20R.%20(24%20de%2002%20de%202022).%204%20Types%20of%20Surfaces%20Used%20for%20Tennis%20Court.%20Obtido%20em%2014%20de%2004%20de%202023,%20de%20Rubcorp:%20https://www.rubcorp.com/types-of-tennis-court-surfaces/https://rioopen.com/o-torneio/) *[History](https://d.docs.live.net/932760ed81e76c92/Ambiente de Trabalho/PACDI - PROJETO/PACDI - Projeto_semanal/ATP. (s.d.). History. Obtido de atptour: https:/www.atptour.com/en/corporate/historyATP. (s.d.). Senior Leadership Team. Obtido de ATP: https:/www.atptour.com/en/corporate/managementHadlich, G. (s.d.). How Does Tennis Scoring Work?%20(By%20Former%20PRO).%20Obtido%20em%2013%20de%2004%20de%202023,%20de%20mytennishq:%20https://mytennishq.com/tennis-scoring-rules-the-ultimate-guide-explained/Kothari,%20R.%20(24%20de%2002%20de%202022).%204%20Types%20of%20Surfaces%20Used%20for%20Tennis%20Court.%20Obtido%20em%2014%20de%2004%20de%202023,%20de%20Rubcorp:%20https://www.rubcorp.com/types-of-tennis-court-surfaces/https://rioopen.com/o-torneio/)*[. Obtido de atptour: https://www.atptour.com/en/corporate/history](https://d.docs.live.net/932760ed81e76c92/Ambiente de Trabalho/PACDI - PROJETO/PACDI - Projeto_semanal/ATP. (s.d.). History. Obtido de atptour: https:/www.atptour.com/en/corporate/historyATP. (s.d.). Senior Leadership Team. Obtido de ATP: https:/www.atptour.com/en/corporate/managementHadlich, G. (s.d.). How Does Tennis Scoring Work?%20(By%20Former%20PRO).%20Obtido%20em%2013%20de%2004%20de%202023,%20de%20mytennishq:%20https://mytennishq.com/tennis-scoring-rules-the-ultimate-guide-explained/Kothari,%20R.%20(24%20de%2002%20de%202022).%204%20Types%20of%20Surfaces%20Used%20for%20Tennis%20Court.%20Obtido%20em%2014%20de%2004%20de%202023,%20de%20Rubcorp:%20https://www.rubcorp.com/types-of-tennis-court-surfaces/https://rioopen.com/o-torneio/)

[ATP. (s.d.).](https://d.docs.live.net/932760ed81e76c92/Ambiente de Trabalho/PACDI - PROJETO/PACDI - Projeto_semanal/ATP. (s.d.). History. Obtido de atptour: https:/www.atptour.com/en/corporate/historyATP. (s.d.). Senior Leadership Team. Obtido de ATP: https:/www.atptour.com/en/corporate/managementHadlich, G. (s.d.). How Does Tennis Scoring Work?%20(By%20Former%20PRO).%20Obtido%20em%2013%20de%2004%20de%202023,%20de%20mytennishq:%20https://mytennishq.com/tennis-scoring-rules-the-ultimate-guide-explained/Kothari,%20R.%20(24%20de%2002%20de%202022).%204%20Types%20of%20Surfaces%20Used%20for%20Tennis%20Court.%20Obtido%20em%2014%20de%2004%20de%202023,%20de%20Rubcorp:%20https://www.rubcorp.com/types-of-tennis-court-surfaces/https://rioopen.com/o-torneio/) *[Senior Leadership Team](https://d.docs.live.net/932760ed81e76c92/Ambiente de Trabalho/PACDI - PROJETO/PACDI - Projeto_semanal/ATP. (s.d.). History. Obtido de atptour: https:/www.atptour.com/en/corporate/historyATP. (s.d.). Senior Leadership Team. Obtido de ATP: https:/www.atptour.com/en/corporate/managementHadlich, G. (s.d.). How Does Tennis Scoring Work?%20(By%20Former%20PRO).%20Obtido%20em%2013%20de%2004%20de%202023,%20de%20mytennishq:%20https://mytennishq.com/tennis-scoring-rules-the-ultimate-guide-explained/Kothari,%20R.%20(24%20de%2002%20de%202022).%204%20Types%20of%20Surfaces%20Used%20for%20Tennis%20Court.%20Obtido%20em%2014%20de%2004%20de%202023,%20de%20Rubcorp:%20https://www.rubcorp.com/types-of-tennis-court-surfaces/https://rioopen.com/o-torneio/)*[. Obtido de ATP: https://www.atptour.com/en/corporate/management](https://d.docs.live.net/932760ed81e76c92/Ambiente de Trabalho/PACDI - PROJETO/PACDI - Projeto_semanal/ATP. (s.d.). History. Obtido de atptour: https:/www.atptour.com/en/corporate/historyATP. (s.d.). Senior Leadership Team. Obtido de ATP: https:/www.atptour.com/en/corporate/managementHadlich, G. (s.d.). How Does Tennis Scoring Work?%20(By%20Former%20PRO).%20Obtido%20em%2013%20de%2004%20de%202023,%20de%20mytennishq:%20https://mytennishq.com/tennis-scoring-rules-the-ultimate-guide-explained/Kothari,%20R.%20(24%20de%2002%20de%202022).%204%20Types%20of%20Surfaces%20Used%20for%20Tennis%20Court.%20Obtido%20em%2014%20de%2004%20de%202023,%20de%20Rubcorp:%20https://www.rubcorp.com/types-of-tennis-court-surfaces/https://rioopen.com/o-torneio/)

[Hadlich, G. (s.d.).](https://d.docs.live.net/932760ed81e76c92/Ambiente de Trabalho/PACDI - PROJETO/PACDI - Projeto_semanal/ATP. (s.d.). History. Obtido de atptour: https:/www.atptour.com/en/corporate/historyATP. (s.d.). Senior Leadership Team. Obtido de ATP: https:/www.atptour.com/en/corporate/managementHadlich, G. (s.d.). How Does Tennis Scoring Work?%20(By%20Former%20PRO).%20Obtido%20em%2013%20de%2004%20de%202023,%20de%20mytennishq:%20https://mytennishq.com/tennis-scoring-rules-the-ultimate-guide-explained/Kothari,%20R.%20(24%20de%2002%20de%202022).%204%20Types%20of%20Surfaces%20Used%20for%20Tennis%20Court.%20Obtido%20em%2014%20de%2004%20de%202023,%20de%20Rubcorp:%20https://www.rubcorp.com/types-of-tennis-court-surfaces/https://rioopen.com/o-torneio/) *[How Does Tennis Scoring Work? (By Former PRO)](https://d.docs.live.net/932760ed81e76c92/Ambiente de Trabalho/PACDI - PROJETO/PACDI - Projeto_semanal/ATP. (s.d.). History. Obtido de atptour: https:/www.atptour.com/en/corporate/historyATP. (s.d.). Senior Leadership Team. Obtido de ATP: https:/www.atptour.com/en/corporate/managementHadlich, G. (s.d.). How Does Tennis Scoring Work?%20(By%20Former%20PRO).%20Obtido%20em%2013%20de%2004%20de%202023,%20de%20mytennishq:%20https://mytennishq.com/tennis-scoring-rules-the-ultimate-guide-explained/Kothari,%20R.%20(24%20de%2002%20de%202022).%204%20Types%20of%20Surfaces%20Used%20for%20Tennis%20Court.%20Obtido%20em%2014%20de%2004%20de%202023,%20de%20Rubcorp:%20https://www.rubcorp.com/types-of-tennis-court-surfaces/https://rioopen.com/o-torneio/)*[. Obtido em 13 de 04 de 2023, de mytennishq: https://mytennishq.com/tennis-scoring-rules-the-ultimate-guide-explained/](https://d.docs.live.net/932760ed81e76c92/Ambiente de Trabalho/PACDI - PROJETO/PACDI - Projeto_semanal/ATP. (s.d.). History. Obtido de atptour: https:/www.atptour.com/en/corporate/historyATP. (s.d.). Senior Leadership Team. Obtido de ATP: https:/www.atptour.com/en/corporate/managementHadlich, G. (s.d.). How Does Tennis Scoring Work?%20(By%20Former%20PRO).%20Obtido%20em%2013%20de%2004%20de%202023,%20de%20mytennishq:%20https://mytennishq.com/tennis-scoring-rules-the-ultimate-guide-explained/Kothari,%20R.%20(24%20de%2002%20de%202022).%204%20Types%20of%20Surfaces%20Used%20for%20Tennis%20Court.%20Obtido%20em%2014%20de%2004%20de%202023,%20de%20Rubcorp:%20https://www.rubcorp.com/types-of-tennis-court-surfaces/https://rioopen.com/o-torneio/)

[Kothari, R. (24 de 02 de 2022).](https://d.docs.live.net/932760ed81e76c92/Ambiente de Trabalho/PACDI - PROJETO/PACDI - Projeto_semanal/ATP. (s.d.). History. Obtido de atptour: https:/www.atptour.com/en/corporate/historyATP. (s.d.). Senior Leadership Team. Obtido de ATP: https:/www.atptour.com/en/corporate/managementHadlich, G. (s.d.). How Does Tennis Scoring Work?%20(By%20Former%20PRO).%20Obtido%20em%2013%20de%2004%20de%202023,%20de%20mytennishq:%20https://mytennishq.com/tennis-scoring-rules-the-ultimate-guide-explained/Kothari,%20R.%20(24%20de%2002%20de%202022).%204%20Types%20of%20Surfaces%20Used%20for%20Tennis%20Court.%20Obtido%20em%2014%20de%2004%20de%202023,%20de%20Rubcorp:%20https://www.rubcorp.com/types-of-tennis-court-surfaces/https://rioopen.com/o-torneio/) *[4 Types of Surfaces Used for Tennis Court](https://d.docs.live.net/932760ed81e76c92/Ambiente de Trabalho/PACDI - PROJETO/PACDI - Projeto_semanal/ATP. (s.d.). History. Obtido de atptour: https:/www.atptour.com/en/corporate/historyATP. (s.d.). Senior Leadership Team. Obtido de ATP: https:/www.atptour.com/en/corporate/managementHadlich, G. (s.d.). How Does Tennis Scoring Work?%20(By%20Former%20PRO).%20Obtido%20em%2013%20de%2004%20de%202023,%20de%20mytennishq:%20https://mytennishq.com/tennis-scoring-rules-the-ultimate-guide-explained/Kothari,%20R.%20(24%20de%2002%20de%202022).%204%20Types%20of%20Surfaces%20Used%20for%20Tennis%20Court.%20Obtido%20em%2014%20de%2004%20de%202023,%20de%20Rubcorp:%20https://www.rubcorp.com/types-of-tennis-court-surfaces/https://rioopen.com/o-torneio/)*[. Obtido em 14 de 04 de 2023, de Rubcorp: https://www.rubcorp.com/types-of-tennis-court-surfaces/](https://d.docs.live.net/932760ed81e76c92/Ambiente de Trabalho/PACDI - PROJETO/PACDI - Projeto_semanal/ATP. (s.d.). History. Obtido de atptour: https:/www.atptour.com/en/corporate/historyATP. (s.d.). Senior Leadership Team. Obtido de ATP: https:/www.atptour.com/en/corporate/managementHadlich, G. (s.d.). How Does Tennis Scoring Work?%20(By%20Former%20PRO).%20Obtido%20em%2013%20de%2004%20de%202023,%20de%20mytennishq:%20https://mytennishq.com/tennis-scoring-rules-the-ultimate-guide-explained/Kothari,%20R.%20(24%20de%2002%20de%202022).%204%20Types%20of%20Surfaces%20Used%20for%20Tennis%20Court.%20Obtido%20em%2014%20de%2004%20de%202023,%20de%20Rubcorp:%20https://www.rubcorp.com/types-of-tennis-court-surfaces/https://rioopen.com/o-torneio/)

O Torneio – Rio Open. (n.d.). Retrieved April 16, 2023, from https://rioopen.com/o-torneio/