**Documentação do Projeto 1 : Dat amining**

Base de Dados: European Soccer

**Alunos**:

Adrianne Alves da Silva - 16/0047595

João Pedro Sconetto -

**Descrição do Projeto**

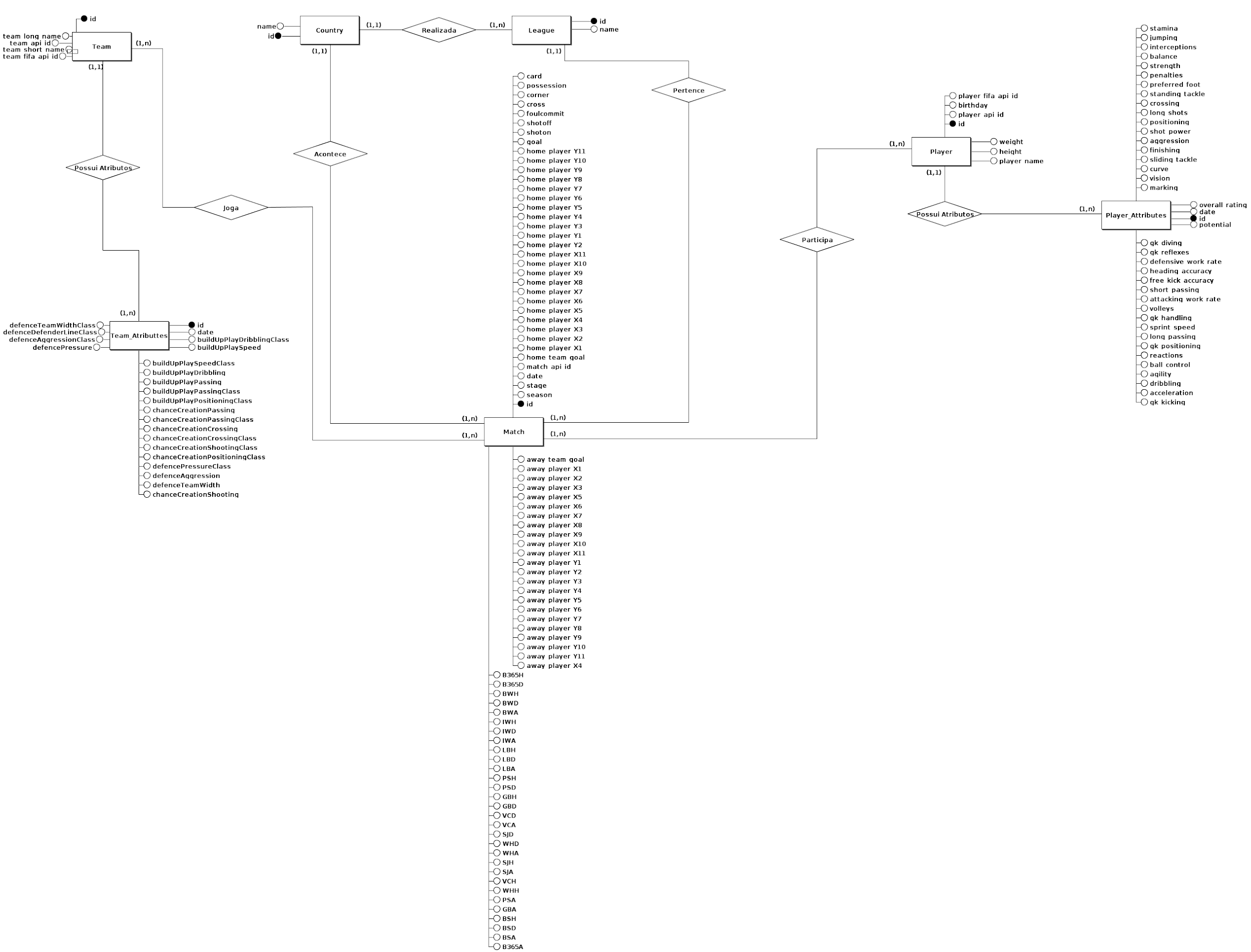
Data mining é um processo nada trivial de identificar, nos dados, padrões novos e úteis.Geralmente, esses dados estão relacionados a negócios, mercados ou pesquisas científicas. Essa atividade envolve a preparação, limpeza e seleção dos dados, além da interpretação e pode ser usada com o objetivo explicativo, confirmativo e exploratório.

A proposta do projeto consiste em selecionar uma base de dados e aplicar a tecnologia descrita. Assim, a base de dados escolhida foi a **europeanSoccer**, encontrada na plataforma **kaggle** (https://www.kaggle.com/hugomathien/soccer). Essa base possui informações sobre futebol europeu, com dados sobre partidas, jogadores, ligas, times e locais em que ocorreram.

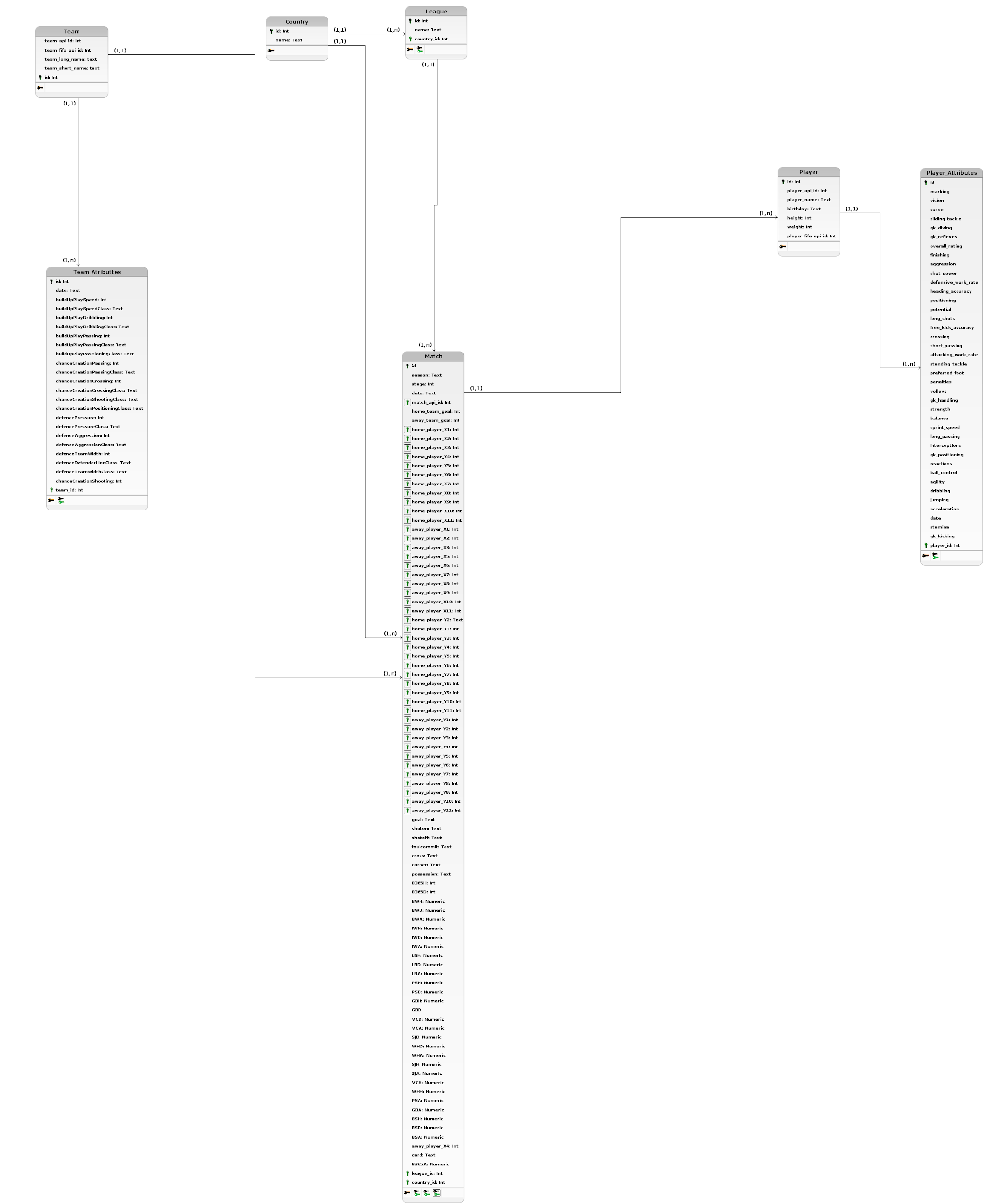
Pretende-se realizar a limpeza dos dados, selecionando por meio do software weka, aqueles que serão utilizados na mineração. Tendo um objetivo definido será feita a escolha do algoritmo que melhor se adequa e posteriormente as informações obtidas no processo serão analisadas.(978 caracteres)

**DE-R**

Segue abaixo o DE-R desenvolvido a partir da base dados original baixada no kaggle em formato **sqlite**. Procurou-se representar as relações como são apresentadas na base original, diferente das modificações realizadas posteriormente, já registrando o andamento do processo de data mining.



**Diagrama Lógico**

****

**DICIONÁRIO DE DADOS**

Os tipos dos dados desse dicionário estão adaptados para o mysql.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Entidade: Team** | | | | |
| **Descrição: Define os dados de um time de futebol europeu.** | | | | |
| **Atributo** | **Propriedades do atributo** | **Tipo de dado** | **Tamanho** | **Descrição** |
| **id** | **Chave primária**  **obrigatório** | **Int** |  | **Identificador do time** |
| **team\_long\_name** | **obrigatório** | **varchar** | **255** | **Nome do time** |
| **team\_api\_id** | **opcional** | **Int** |  | **Identificador do time na api** |
| **team\_short\_name** | **obrigatório** | **varchar** | **3** | **Nome resumido do time** |
| **team\_fifa\_api\_id** | **obrigatório** | **Int** |  | **Identificador do time de acordo a api da fifa** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Entidade: Team\_Attributes** | | | | |
| **Descrição: Tabela que contém os atributos de um time em uma determinada data, acompanhando o seu desempenho.** | | | | |
| **Atributo** | **Propriedades do atributo** | **Tipo de dado** | **Tamanho** | **Descrição** |
| **id** | **Obrigatório**  **Chave Primária** | **Int** |  | **Identificador do grupo de atributos do time** |
| **date** | **Obrigatório** | **datetime** |  | **Data em que foram registrados esses atributos** |
| **buildUpPlaySpeed** | **Obrigatório** | **Int** |  | **Velocidade de jogo** |
| **buildUpPlaySpeedClass** | **Obrigatório** | **varchar** | **8** | **Classe da velocidade do jogo** |
| **buildUpPlayDribbling** | **Obrigatório** | **Int** |  | **Jogar drible** |
| **buildUpPlayDribblingClass** | **Obrigatório** | **varchar** | **6** | **Classe de drible da** |
| **buildUpPlayPassing** | **Obrigatório** | **Int** |  |  |
| **buildUpPlayPassingClass** | **Obrigatório** | **varchar** | **5** |  |
| **buildUpPlayPositioningClass** | **Obrigatório** | **varchar** | **9** |  |
| **chanceCriationPassing** | **Obrigatório** | **Int** |  |  |
| **chanceCriationPassingClass** | **Obrigatório** | **varchar** | **6** |  |
| **chanceCriationCrossing** | **Obrigatório** | **Int** |  |  |
| **chanceCriationCrossingClass** | **Obrigatório** | **varchar** | **6** |  |
| **chanceCriationShootingClass** | **Obrigatório** | **varchar** | **6** |  |
| **chanceCriationPositioningClass** | **Obrigatório** | **varchar** | **9** |  |
| **defencePressure** | **Obrigatório** | **Int** |  |  |
| **defensePressureClass** | **Obrigatório** | **varchar** | **6** |  |
| **defenceAgression** | **Obrigatório** | **Int** |  |  |
| **defenceAggressionClass** | **Obrigatório** | **varchar** | **7** |  |
| **defenceTeamWidth** | **Obrigatório** | **Int** |  |  |
| **defenceDefenderLineClass** | **Obrigatório** | **varchar** | **12** |  |
| **defenceTeamWidthClass** | **Obrigatório** | **Int** |  |  |
| **chanceCriationShooting** | **Obrigatório** | **Int** |  |  |
| **team\_id** | **Obrigatório**  **Chave Estrangeira** | **Int** |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Entidade: Country** | | | | |
| **Descrição: Contém as informações dos países em que acontecem os jogos.** | | | | |
| **Atributo** | **Propriedades do atributo** | **Tipo de dado** | **Tamanho** | **Descrição** |
| **id** | **Obrigatório**  **Chave Primária** | **Int** |  | **Identificador do País** |
| **name** | **Obrigatório**  **Único** | **varchar** | **255** | **Nome do país** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Entidade: League** | | | | |
| **Descrição: Guarda as informações das ligas.** | | | | |
| **Atributo** | **Propriedades do atributo** | **Tipo de dado** | **Tamanho** | **Descrição** |
| **id** | **Obrigatório**  **Chave Primária** | **Int** |  | **Identificador da liga** |
| **name** | **Obrigatório** | **varchar** | **255** | **Nome da liga** |
| **country\_id** | **Obrigatório**  **Chave Estrangeira** | **Int** |  | **Atributo que identifica o país ao qual a liga pertence** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Entidade: Player** | | | | |
| **Descrição: Guarda as informações dos jogadores no banco de dados.** | | | | |
| **Atributo** | **Propriedades do atributo** | **Tipo de dado** | **Tamanho** | **Descrição** |
| **id** | **Obrigatório**  **Chave Primária** | **Int** |  | **Identificador do jogador na base de dados** |
| **player\_api\_id** | **Obrigatório**  **Único** | **Int** |  | **Identificador do jogador na api** |
| **player\_name** | **Obrigatório** | **varchar** | **255** | **Nome do jogador** |
| **birthday** | **Obrigatório** | **datetime** |  | **Data de nascimento do jogador** |
| **height** | **Obrigatório** | **float** |  | **Altura do jogador** |
| **weight** | **Obrigatório** | **float** |  | **Peso do jogador** |
| **player\_fifa\_api** | **Obrigatório**  **Único** | **Int** |  | **Identificador do jogador na api da fifa** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Entidade: Player\_Attributes** | | | | |
| **Descrição: Guarda as informações sobre o desempenho dos jogadores com o tempo.** | | | | |
| **Atributo** | **Propriedades do atributo** | **Tipo de dado** | **Tamanho** | **Descrição** |
| **id** | **Obrigatório**  **Chave Primária** | **Int** |  | **Identificador do conjunto dos atributos na base de dados** |
| **marking** | **Obrigatório** | **Int** |  |  |
| **vision** | **Obrigatório** | **Int** |  |  |
| **curve** | **Obrigatório** | **Int** |  |  |
| **sliding\_tackle** | **Obrigatório** | **Int** |  |  |
| **gk\_diving** | **Obrigatório** | **Int** |  |  |
| **gk\_reflexes** | **Obrigatório** | **Int** |  |  |
| **overall\_rating** | **Obrigatório** | **Int** |  |  |
| **finishing** | **Obrigatório** | **Int** |  |  |
| **aggression** | **Obrigatório** | **Int** |  |  |
| **shot\_power** | **Obrigatório** | **Int** |  |  |
| **defensive\_work\_rate** | **Obrigatório** | **varchar** | **6** |  |
| **heading\_accuracy** | **Obrigatório** | **Int** |  |  |
| **positioning** | **Obrigatório** | **Int** |  |  |
| **potential** | **Obrigatório** | **Int** |  |  |
| **long\_shots** | **Obrigatório** | **Int** |  |  |
| **free\_kick\_accuracy** | **Obrigatório** | **Int** |  |  |
| **crossing** | **Obrigatório** | **Int** |  |  |
| **short\_passing** | **Obrigatório** | **Int** |  |  |
| **attacking\_work\_rate** | **Obrigatório** | **varchar** | **6** |  |
| **standing\_tackle** | **Obrigatório** | **Int** |  |  |
| **preferred\_foot** | **Obrigatório** | **varchar** | **5** |  |
| **penalties** | **Obrigatório** | **Int** |  |  |
| **volleys** | **Obrigatório** | **Int** |  |  |
| **gk\_handling** | **Obrigatório** | **Int** |  |  |
| **strength** | **Obrigatório** | **Int** |  |  |
| **balance** | **Obrigatório** | **Int** |  |  |
| **sprint\_speed** | **Obrigatório** | **Int** |  |  |
| **long\_passing** | **Obrigatório** | **Int** |  |  |
| **interceptions** | **Obrigatório** | **Int** |  |  |
| **ball\_control** | **Obrigatório** | **Int** |  |  |
| **agility** | **Obrigatório** | **Int** |  |  |
| **dribbling** | **Obrigatório** | **Int** |  |  |
| **jumping** | **Obrigatório** | **Int** |  |  |
| **acceleration** | **Obrigatório** | **Int** |  |  |
| **date** | **Obrigatório** | **datetime** |  |  |
| **stamina** | **Obrigatório** | **Int** |  |  |
| **gk\_kicking** | **Obrigatório** | **Int** |  |  |
| **player\_id** | **Obrigatório**  **Chave estrangeira** | **Int** |  | **Identificador do jogador que possui esse conjunto de atributos** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Entidade: Match** | | | | |
| **Descrição: Guarda as informações sobre as partidas de futebol europeu.** | | | | |
| **Atributo** | **Propriedades do atributo** | **Tipo de dado** | **Tamanho** | **Descrição** |
| **id** | **Obrigatório**  **Chave Primária** | **Int** |  | **Identificador da partida na base de dados** |
| **season** | **Obrigatório** | **varchar** | **9** |  |
| **stage** | **Obrigatório** | **Int** |  |  |
| **date** | **Obrigatório** | **datetime** |  |  |
| **match\_api\_id** | **Obrigatório**  **Único** | **Int** |  | **Identificador da partida na api** |
| **home\_team\_goal** | **Obrigatório** | **Int** |  |  |
| **away\_team\_goal** | **Obrigatório** | **Int** |  |  |
| **home\_player\_x1** | **opcional** | **Int** |  |  |
| **home\_player\_x2** | **opcional** | **Int** |  |  |
| **home\_player\_x3** | **opcional** | **Int** |  |  |
| **home\_player\_x4** | **opcional** | **Int** |  |  |
| **home\_player\_x5** | **opcional** | **Int** |  |  |
| **home\_player\_x6** | **opcional** | **Int** |  |  |
| **home\_player\_x7** | **opcional** | **Int** |  |  |
| **home\_player\_x8** | **opcional** | **Int** |  |  |
| **home\_player\_x9** | **opcional** | **Int** |  |  |
| **home\_player\_x10** | **opcional** | **Int** |  |  |
| **home\_player\_x11** | **opcional** | **Int** |  |  |
| **away\_player\_x1** | **opcional** | **Int** |  |  |
| **away\_player\_x2** | **opcional** | **Int** |  |  |
| **away\_player\_x3** | **opcional** | **Int** |  |  |
| **away\_player\_x4** | **opcional** | **Int** |  |  |
| **away\_player\_x5** | **opcional** | **Int** |  |  |
| **away\_player\_x6** | **opcional** | **Int** |  |  |
| **away\_player\_x7** | **opcional** | **Int** |  |  |
| **away\_player\_x8** | **opcional** | **Int** |  |  |
| **away\_player\_x9** | **opcional** | **Int** |  |  |
| **away\_player\_x10** | **opcional** | **Int** |  |  |
| **away\_player\_x11** | **opcional** | **Int** |  |  |
| **home\_player\_y1** | **opcional** | **Int** |  |  |
| **home\_player\_y2** | **opcional** | **Int** |  |  |
| **home\_player\_y3** | **opcional** | **Int** |  |  |
| **home\_player\_y4** | **opcional** | **Int** |  |  |
| **home\_player\_y5** | **opcional** | **Int** |  |  |
| **home\_player\_y6** | **opcional** | **Int** |  |  |
| **home\_player\_y7** | **opcional** | **Int** |  |  |
| **home\_player\_y8** | **opcional** | **Int** |  |  |
| **home\_player\_y9** | **opcional** | **Int** |  |  |
| **home\_player\_y10** | **opcional** | **Int** |  |  |
| **home\_player\_y11** | **opcional** | **Int** |  |  |
| **away\_player\_y1** | **opcional** | **Int** |  |  |
| **away\_player\_y2** | **opcional** | **Int** |  |  |
| **away\_player\_y3** | **opcional** | **Int** |  |  |
| **away\_player\_y4** | **opcional** | **Int** |  |  |
| **away\_player\_y5** | **opcional** | **Int** |  |  |
| **away\_player\_y6** | **opcional** | **Int** |  |  |
| **away\_player\_y7** | **opcional** | **Int** |  |  |
| **away\_player\_y8** | **opcional** | **Int** |  |  |
| **away\_player\_y9** | **opcional** | **Int** |  |  |
| **away\_player\_y10** | **opcional** | **Int** |  |  |
| **away\_player\_y11** | **opcional** | **Int** |  |  |
| **goal** | **opcional** | **LONGTEXT** |  |  |
| **shoton** | **opcional** | **LONGTEXT** |  |  |
| **shotoff** | **opcional** | **LONGTEXT** |  |  |
| **foulcommit** | **opcional** | **LONGTEXT** |  |  |
| **cross** | **opcional** | **LONGTEXT** |  |  |
| **corner** | **opcional** | **LONGTEXT** |  |  |
| **possession** | **opcional** | **LONGTEXT** |  |  |
| **B365H** | **opcional** | **float** |  |  |
| **B365D** | **opcional** | **float** |  |  |
| **BWH** | **opcional** | **float** |  |  |
| **BWD** | **opcional** | **float** |  |  |
| **BWA** | **opcional** | **float** |  |  |
| **IWH** | **opcional** | **float** |  |  |
| **IWD** | **opcional** | **float** |  |  |
| **IWA** | **opcional** | **float** |  |  |
| **LBH** | **opcional** | **float** |  |  |
| **LBD** | **opcional** | **float** |  |  |
| **LBA** | **opcional** | **float** |  |  |
| **PSH** | **opcional** | **float** |  |  |
| **PSD** | **opcional** | **float** |  |  |
| **GBH** | **opcional** | **float** |  |  |
| **GBD** | **opcional** | **float** |  |  |
| **VCD** | **opcional** | **float** |  |  |
| **VCA** | **opcional** | **float** |  |  |
| **SJD** | **opcional** | **float** |  |  |
| **WHD** | **opcional** | **float** |  |  |
| **WHA** | **opcional** | **float** |  |  |
| **SJH** | **opcional** | **float** |  |  |
| **SJA** | **opcional** | **float** |  |  |
| **VCH** | **opcional** | **float** |  |  |
| **WHH** | **opcional** | **float** |  |  |
| **PSA** | **opcional** | **float** |  |  |
| **GBA** | **opcional** | **float** |  |  |
| **BSH** | **opcional** | **float** |  |  |
| **BSD** | **opcional** | **float** |  |  |
| **BSA** | **opcional** | **float** |  |  |
| **card** | **opcional** | **float** |  |  |
| **B365A** | **opcional** | **float** |  |  |
| **league\_id** | **Obrigatório**  **Chave Estrangeira** | **Int** |  |  |
| **country\_id** | **Obrigatório**  **Chave Estrangeira** | **Int** |  |  |