

## PDI

O objetivo era identificar possíveis relações entre os conjuntos de dados, áreas de melhoria e, de forma geral, determinar a melhor abordagem para integrá-los numa base de dados e administrar os dados por meio de um sistema de base de dados. Isso envolveria garantir a integridade dos dados e realizar normalização para evitar redundâncias.

O conteúdo dos arquivos na pasta de conjuntos de dados, observamos que todos os arquivos são em formato Excel, a maioria deles relacionados ao "Inquérito ao Emprego no Ensino Superior Público" (IEESP) e ao inquérito estatístico "Registo Biográfico de Docentes do Ensino Superior" (Rebides). Todos esses arquivos têm um nome geral, que é "Lista Nominativa de Docentes do Ensino Superior".

Ao explorar os arquivos da "Lista Nominativa de Docentes do Ensino Superior", notamos que, nos anos letivos de 2003/04 a 2015/16, os documentos referem-se exclusivamente à lista Rebides, ou seja, uma lista de docentes do ensino superior em instituições privadas (inquérito Rebides). No entanto, a partir do ano letivo 2016/17 até 2018/19, todos os dados foram incluídos em uma única lista, que manteve os dados do inquérito Rebides referentes a esse ano, mas também incluiu o inquérito IEESP. Isso significa que tanto os docentes do ensino superior em instituições públicas quanto privadas estão presentes, mas organizados em folhas separadas.

De 2003/04 a 2015/16:

Arquivo Excel com 3 folhas: capa, carreira Rebides e habilitações Rebides.

De 2016/17 a 2018/19:

Arquivo Excel com 8 folhas: capa, categoria (ou carreira) IEESP, habilitações IEESP, habilitação mais elevada, nota metodológica IEESP, categoria (ou carreira) Rebides, habilitações Rebides e nota metodológica Rebides.

É importante que certos campos nas folhas mudaram de nome ao longo dos anos, indicando uma atualização na forma como esses dados foram interpretados para se adaptar a um sistema mais recente. Também notamos que alguns campos podem conter valores nulos em registros específicos.

É crucial identificar esses aspetos para encontrar uma solução viável para a normalização da base de dados, garantindo que não haja redundâncias, campos desnecessários ou tabelas em excesso. Devemos determinar o que pode ser consolidado entre as diferentes páginas e arquivos de todos os anos letivos fornecidos.

Existem 4 arquivos Excel relacionados ao "Observatório do Emprego Científico e Docente" para os anos de 2019, 2020, 2021 e 2022. Esses arquivos consistem em 5 páginas, que incluem a capa, contrato, habilitações mais elevadas, atividades e nota metodológica. Embora haja semelhanças com os arquivos mencionados anteriormente, eles parecem conter dados de tipos diferentes em alguns campos.

Em princípio a nossa base de dados deve incluir as seguintes tabelas:

#### **Docente**

Id\_docente

Nome\_docente

Data\_nascimento

Faculdade\_trabalho

Faculdade\_estudou

Curso

Pais\_curso

grau

#### **Faculdade**

Id\_faculdade

Nome\_faculdade

regime

Data\_criação

#### **Pais**

Id\_pais

Pais\_curso

#### **Curso**

Id\_curso

Nome\_curso

nome\_faculdade

grau