

Upgrading Skills

Primavera
ACADEMY

Projeto

Análise de RH - Rotatividade e Desempenho dos Colaboradores

Parte 1 - Modelação e Visualizações

AiDAPT01

A. APRESENTAÇÃO

Tema	Análise de RH - Rotatividade e Desempenho dos Colaboradores
Objetivos	Análise exploratória dos dados Normalização da base de dados Consultas SQL Caso de uso com NoSQL Preparação de dados Modelação de dados Identificação de perguntas de negócio, como por exemplo, 'Que fatores contribuem mais para a saída de colaboradores com alto desempenho?' Relatório dos indicadores analíticos
Dados	Dataset público 'IBM HR Analytics Employee Attrition & Performance'
Recursos	Linguagem R e VSCode SQL Server, SQL Server Management Studio e T-SQL MongoDB e MongoDB Compass Power Query e Linguagem M Power BI Desktop e Linguagem DAX

B. CONTEXTO

O departamento de Recursos Humanos procura compreender os fatores que levam à saída de colaboradores e que são responsáveis pela sua alta rotatividade, para além dos elementos que influenciam o seu desempenho. Esta análise pode ajudar a antecipar comportamentos, identificar perfis de desempenho e descobrir processos de retenção de colaboradores.

C. CONJUNTO DE DADOS

O conjunto de dados sugerido chama-se 'IBM HR Analytics Employee Attrition & Performance', que inclui informações sobre colaboradores, como satisfação no trabalho, idade, cargo, tempo de empresa, avaliação de desempenho, entre outros. O dataset está disponível no Kaggle.

D. ETAPAS

1. Analisar o dataset

- Descarregar o dataset do [Kaggle](#).
- Analisar o dataset e criar um dicionário de dados (ver exemplo dataset Iris).

2. Importação de dados e análise exploratória de dados com R

- Carregar o dataset no R.
- Realizar a respetiva análise estatística descritiva:
 - Distribuição das variáveis (médias, medianas, desvio padrão, ...).
 - Identificação de dados ausentes ou inconsistentes.
 - Gerar gráficos básicos (histogramas, boxplots, correlações, ...).

3. Normalização de dados em modelo relacional com SQL Server

- Normalizar o dataset original com uma estrutura de base de dados relacional.
- Aplicar até à 3ª Forma Normal (3FN).
- Justificar as decisões de normalização.

4. Criação de Consultas com T-SQL

- Criar scripts de T-SQL, documentados, para:
 - Criar a base de dados.
 - Criar as tabelas.
 - Inserir dados de exemplo do negócio.
 - Realizar consultas simples, agregadas e com múltiplas tabelas.

5. Definição dum caso de uso usando MongoDB

- Representar os dados dos colaboradores no formato JSON compatível com MongoDB.
- O caso de uso deve permitir guardar o histórico de desempenho dos colaboradores.
- Inserir documentos numa coleção MongoDB chamada 'colaboradores'.
- Realizar consultas simples.

6. Transformação de dados no Power Query

- Importar os dados no Power BI Desktop.
- Usar Power Query para efetuar as necessárias transformações, como:
 - Remoção de colunas desnecessárias
 - Remoção de duplicados
 - Renomeação de colunas, alteração de tipos de dados, entre outros.
- Criar colunas derivadas relevantes para análise.

7. Implementação do modelo de dados OLAP no Power BI Desktop

- Gerar um modelo OLAP usando as convenções das tabelas dimensão e facto.
- Gerar uma tabela de calendário em linguagem M.
- Criar um modelo dimensional com o esquema estrela.
- Relacionar corretamente as tabelas do modelo.
- Verificar cardinalidades e direções de filtro.

8. Definição de métricas e criação de relatório no Power BI Desktop

- Gerar o layout do relatório em PowerPoint.
- Gerar ou adaptar um tema com as cores a usar no relatório.
- Criar o relatório com foco em:
 - Distribuição de colaboradores por género, idade e departamento.
 - Percentagem de colaboradores que saíram da empresa (attrition rate).
 - Relação entre performance e tempo de empresa.
 - Outras métricas pertinentes à análise.

9. Melhoria do modelo no Power BI Desktop

- Nomear objetos de forma amigável para o utilizador final.
- Formatar corretamente as colunas numéricas.
- Categorizar dados:
 - A tabela calendário.
 - Dados geográficos.
- Criar hierarquias relevantes para a análise.
- Criar pastas para organizar as colunas.
- Ocultar colunas de chave e colunas de suporte.

E. RELATÓRIO DO PROJETO

É esperado um relatório final de projeto, elaborado em PowerPoint. Este relatório deve estar organizado por páginas temáticas e ter um índice.

Devem ser incluídos todos os elementos pertinentes para a análise do projeto, nomeadamente:

- Apresentação, conclusão e recomendações com base nos dados.
- Dicionário de dados do dataset.
- Resumo da análise exploratória dos dados.
- Processo de normalização, com a apresentação da 3NF.
- Imagem do diagrama ER do modelo relacional proposto.
- Excertos dos principais scripts T-SQL.
- Resumo explicado do caso de uso NoSQL.
- Comparação visual dos dados antes e depois das transformações.
- Imagem do modelo OLAP em estrela.
- Lista das métricas criadas, com a respetiva explicação.
- Imagens das principais visualizações.

São bem-vindos quaisquer elementos adicionais, desde que devidamente fundamentados.

F. ENTREGA DO PROJETO

O projeto deve ser entregue numa pasta zipada com todos os ficheiros de suporte. Incluam um documento com o índice de ficheiros entregues, identificando o seu nome e uma breve descrição.

Data de início	19/06/2025
Data de entrega	20/07/2025
Ficheiro .zip	Incluir todos os ficheiros do projeto e documento com índice de ficheiros
Email para envio	milena.reis.sousa@hotmail.com
Assunto do email	CA_AiDAPT01_P1_Grupoxx
Anexo do email	CA_AiDAPT01_P1_Grupoxx.zip

Sugere-se a criação e partilha do repositório do projeto em GitHub (serve de portfólio profissional).