



Lisbon School
of Economics
& Management
Universidade de Lisboa



Projeto de Grupo: De Dados Abertos à Business Intelligence

1. Enquadramento do Projeto

Neste projeto de grupo, os estudantes deverão desenvolver uma solução completa de dados, partindo de dados públicos reais e transformando-os, progressivamente, numa base de dados relacional e numa solução de Business Intelligence. O projeto deverá demonstrar a capacidade de recolher, compreender, modelar, armazenar, consultar, transformar, analisar e comunicar informação baseada em dados reais.

Os dados utilizados devem provir de fontes públicas, oficiais ou reputadas, tais como Eurostat, Instituto Nacional de Estatística, Portal de Dados Abertos da União Europeia, Portal de Dados Abertos da Administração Pública Portuguesa ou outras fontes equivalentes, desde que previamente aprovadas pelo docente.

2. Objetivo do Projeto

Cada grupo deverá selecionar um domínio ou problema do mundo real, identificar conjuntos de dados públicos adequados e construir uma solução que permita responder a questões analíticas relevantes. O objetivo não é apenas analisar um ficheiro de dados, mas conceber uma arquitetura completa que mostre como dados estatísticos em bruto podem ser transformados numa base de dados estruturada e, posteriormente, num modelo dimensional adequado à análise em Business Intelligence.

Exemplos de domínios possíveis incluem turismo, transportes, emprego, população, educação, saúde, energia, agricultura, atividade empresarial, desenvolvimento regional, comércio internacional, digitalização, tecnologia, ambiente e sustentabilidade.

3. Etapas do Projeto

3.1 Identificação das Fontes de Dados

O grupo deverá identificar e justificar as fontes de dados utilizadas. Para cada conjunto de dados, devem ser indicados a fonte, o URL, a instituição responsável, a descrição, o período abrangido, a cobertura geográfica, as principais variáveis, a unidade de medida, a frequência das observações e as condições de licença ou utilização, quando disponíveis. Os dados artificiais não podem ser utilizados como fonte principal do projeto.

3.1.1 Exemplos de Fontes de Dados

As fontes de dados devem ser escolhidas em função do problema de investigação ou de análise definido pelo grupo. Não basta selecionar uma fonte conhecida; é necessário garantir que os dados têm granularidade suficiente, cobertura temporal adequada, documentação clara e possibilidade de integração com outros conjuntos de dados.

Por exemplo, projetos sobre turismo podem recorrer ao INE para obter dados sobre dormidas, hóspedes, estabelecimentos hoteleiros e capacidade de alojamento, podendo complementar esta informação com indicadores regionais disponíveis na PORDATA ou no Eurostat. Projetos sobre emprego e mercado de trabalho podem combinar estatísticas do INE com dados comparáveis do Eurostat, permitindo analisar diferenças entre Portugal, regiões portuguesas e outros países europeus. Trabalhos sobre energia e sustentabilidade podem articular dados da Direção-Geral de Energia e Geologia com indicadores ambientais disponibilizados pela Agência Portuguesa do Ambiente ou por portais europeus de dados abertos.

Outras possibilidades incluem a utilização de dados do Portal Nacional de Dados Abertos, especialmente quando o projeto envolve administração pública, mobilidade, educação, saúde ou território; dados do Banco Mundial ou da OCDE, quando o objetivo é comparar países ou estudar indicadores internacionais; e dados geográficos da Direção-Geral do Território, quando a análise exige uma componente espacial. A escolha de cada fonte deve ser justificada no relatório, explicando a sua relevância para as perguntas analíticas, a qualidade dos dados, as limitações existentes e a forma como os dados serão integrados no modelo de base de dados.

3.2 Compreensão e Preparação dos Dados

Antes da modelação da base de dados, os estudantes deverão analisar os dados originais, identificando variáveis relevantes, identificadores primários, possíveis relações entre conjuntos de dados, variáveis categóricas, medidas, datas, dimensões geográficas, valores em falta, duplicados, inconsistências, diferenças de unidades ou formatos e problemas de integração. As transformações realizadas devem ser documentadas, preservando-se separadamente os dados originais e os dados tratados.

4. Modelação Conceptual, Relacional e Implementação

Com base no domínio escolhido, cada grupo deverá criar um modelo conceptual sob a forma de diagrama de classes, incluindo classes, atributos, relações, multiplicidades e eventuais generalizações ou especializações. Este modelo deverá ser depois transformado num modelo relacional, identificando tabelas, atributos, chaves primárias, chaves estrangeiras, chaves candidatas, relações e restrições de integridade referencial.

A base de dados relacional deverá ser implementada em Microsoft Access ou noutro sistema de gestão de bases de dados aprovado, como MySQL, PostgreSQL, SQLite ou SQL Server. A implementação deve conter tabelas, chaves primárias, chaves estrangeiras, tipos de dados adequados, relações, restrições de integridade e índices quando relevantes. A base de dados deverá ser povoada com os dados reais recolhidos.

5. SQL e Modelo Dimensional

O projeto deverá incluir um conjunto de consultas SQL que respondam a perguntas relevantes sobre o domínio escolhido. As consultas devem demonstrar seleção simples, filtragem, ordenação, agregação, agrupamento, junções entre tabelas, análise temporal, combinação de filtros e agregações, e pelo menos uma consulta mais complexa envolvendo várias tabelas ou subconsultas. As consultas não devem existir apenas para ilustrar sintaxe; cada uma deverá responder a uma questão analítica significativa.

A partir da base de dados relacional, os estudantes deverão construir um modelo dimensional adequado à Business Intelligence, podendo optar por um esquema em estrela ou em floco de neve. O modelo deverá identificar tabelas de factos, tabelas de dimensões, medidas, chaves, relações e a granularidade da tabela de factos. A granularidade deve ser explicitamente definida, por exemplo: cada registo representa o valor de um indicador para uma determinada região, ano e setor económico.

6. Power BI e Questões Analíticas

O modelo dimensional deverá ser ligado ao Microsoft Power BI para criação de uma solução interativa de Business Intelligence. O dashboard deverá incluir um modelo de dados adequado, relações, medidas, indicadores calculados, visualizações interativas, filtros, segmentações, análise temporal e visualização geográfica sempre que os dados o permitam. As visualizações devem contar uma história analítica coerente e responder às questões definidas no início do projeto.

Antes da construção do dashboard, cada grupo deverá definir pelo menos cinco questões analíticas relevantes. Estas questões deverão orientar as consultas SQL e as visualizações em Power BI, garantindo coerência entre o problema escolhido, os dados utilizados, os modelos desenvolvidos e os resultados apresentados.

7. Entregáveis

1. **Relatório do projeto**, incluindo introdução, definição do problema, fontes de dados, preparação dos dados, modelo conceptual, modelo relacional, implementação da base de dados, consultas SQL, modelo dimensional, implementação dimensional, solução Power BI, principais resultados, limitações, conclusões e referências. (maximo 15 paginas).
Versão em DOC e em PDF
2. **Base de dados relacional** completa, com tabelas povoadas e relações definidas.
3. **Ficheiro SQL** com as consultas desenvolvidas.
4. **Modelo dimensional** implementado, em estrela ou floco de neve.
5. **Ficheiro Power BI** com modelo de dados, medidas e dashboards.
6. **Dados originais** ou instruções precisas para acesso às fontes, quando a redistribuição não for permitida.

8. Apresentação do Projeto

Cada grupo apresentará o projeto à turma, demonstrando a trajetória completa dos dados: dados públicos, dados em bruto, modelo conceptual, modelo relacional, base de dados, SQL, modelo dimensional, Power BI e principais conhecimentos obtidos. A apresentação deve privilegiar as decisões de modelação e os resultados analíticos, em vez de se limitar à exibição de capturas de ecrã. É obrigatória a presença de todos os estudantes.

9. Requisitos Importantes

- Utilização obrigatória de dados reais e publicamente disponíveis.
- Identificação clara das fontes de dados e dos indicadores utilizados.
- Integração de múltiplos conjuntos de dados sempre que possível.
- Coerência entre modelo conceptual, modelo relacional, base de dados e modelo dimensional.
- Documentação das transformações realizadas e preservação dos dados originais.
- Reprodutibilidade do processo, permitindo que outra pessoa compreenda a origem dos dados e os principais passos realizados.

10. Resultado Esperado

No final do projeto, cada grupo deverá ter desenvolvido uma solução completa de dados, demonstrando a capacidade de transformar dados públicos reais em informação estruturada, consultável e visualmente comunicável através de tecnologias de bases de dados e Business Intelligence.

Data de submissão de discussão de acordo com planejamento da UC