



UNIVERSIDADE
DE TRÁS-OS-MONTES
E ALTO DOURO

Escola de Ciências e Tecnologia – Departamento de Engenharias

Licenciatura em Engenharia Informática

Ano letivo 2024/2025

Introdução à Ciência dos Dados

Protocolo do Trabalho Experimental 1 – Enunciado D2

Trabalho Experimental 1 (conclusão até 30 de março de 2025):

Utilize o seguinte Dataset “**Global Water Consumption Dataset (2000-2024)**” que pode ser descarregado em <https://www.kaggle.com/datasets/atharvasoundankar/global-water-consumption-dataset-2000-2024>. Neste link pode também encontrar informação detalhada sobre os diversos campos do Dataset.

Em função deste Dataset realize as seguintes tarefas em linguagem Python, que devem ficar devidamente documentadas no relatório deste Trabalho Experimental 1 a ser submetido no NONIO (ou seja, devem submeter um ficheiro ZIP contendo dois ficheiros, um com o relatório em formato PDF e outro com o código da linguagem Python):

1. Carregue o ficheiro .csv para um DataFrame, e de seguida crie um novo DataFrame com apenas a informação dos países Italy, Japan, Spain, UK e USA. Grave este novo DataFrame num novo ficheiro .csv.
2. A partir do novo DataFrame, faça um gráfico que lhe permita visualizar convenientemente a evolução do consumo total de água (“Total Water Consumption”) por ano nos referidos países.
3. Usando a biblioteca Matplotlib, crie um gráfico circular (‘pie chart’) que represente em a relação entre as percentagens de utilização de água na agricultura (“Agricultural Water Use”), utilização de água industrial (“Industrial Water Use”) e utilização doméstica da água (“Household Water Use”) no ano 2020 para o país Spain. Coloque as legendas adequadas.
4. Crie uma função que, dado o nome do país (“Country”), apresente o ano (“Year”) em a percentagem de utilização de água na agricultura (“Agricultural Water Use”) foi menor, bem como este valor.
5. Usando o DataFrame completo, crie um gráfico de dispersão (“scatter plot”) que permita visualizar o relacionamento entre a utilização de água industrial (“Industrial Water Use”) e a taxa de esgotamento das águas subterrâneas (“Groundwater Depletion Rate”). Apresente também no gráfico uma regressão linear que relacione as duas variáveis. Explique convenientemente o seu significado.
6. Usando o DataFrame completo, explore técnicas de Machine Learning que lhe permita fazer uma previsão da utilização de água per capita (“Per Capita Water Use”) em função do país (“Country”), da utilização de água na agricultura (“Agricultural Water Use”) e do impacto da precipitação (“Rainfall Impact”). Documente bem as técnicas utilizadas, as decisões tomadas e os resultados obtidos.