



UNIVERSIDADE  
DE TRÁS-OS-MONTES  
E ALTO DOURO

Escola de Ciências e Tecnologia – Departamento de Engenharias

Licenciatura em Engenharia Informática

Ano letivo 2022/2023

Introdução à Ciência dos Dados

## Protocolo do Trabalho Experimental 1 – Enunciado C

### Trabalho Experimental 1 (conclusão até 21 de abril de 2023):

Utilize o Dataset “**Life Expectancy (WHO) Fixed**” que pode ser descarregado em <https://www.kaggle.com/datasets/lashagoch/life-expectancy-who-updated>. 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 SIDE:

1. Carregue o ficheiro .csv para um DataFrame, e de seguida crie um novo DataFrame com apenas a informação da Região “Asia”. Grave este novo DataFrame num novo ficheiro .csv.
2. A partir do novo DataFrame, faça um gráfico que lhe permita visualizar convenientemente a evolução das mortes de crianças menores de cinco anos por 1000 habitantes (“Under\_five\_deaths”) nos países China, Índia, Japão e Tailândia.
3. Usando a biblioteca Matplotlib, crie um gráfico circular (‘pie chart’) que represente a média nos anos 2000 a 2015 da população total em milhões (“Population\_mln”), nos países Afeganistão, Indonésia, Filipinas e Vietnã. Coloque as legendas adequadas.
4. Crie uma função que, dado o nome do país da Região “Asia”, apresente o ano em que a esperança média de vida (“Life\_expectancy”) foi maior, bem como o respetivo valor.
5. Usando a biblioteca Seaborn, crie um gráfico de dispersão (‘scatter plot’) que permita visualizar o relacionamento entre os incidentes de HIV por 1000 habitantes dos 15 a 49 anos (“Incidents\_HIV”) e a esperança média de vida (“Life\_expectancy”). Apresente também no gráfico uma regressão linear que relacione as duas variáveis. Explique convenientemente o seu significado. Pode fazer análise para um país, uma região ou o globo.
6. Explore técnicas de Machine Learning que lhe permita fazer uma previsão da esperança média de vida (“Life\_expectancy”) no futuro. Poderá usar quaisquer colunas do Dataset que ache conveniente. Documente bem as técnicas utilizadas e decisões tomadas.