

AI & CHATBOT

Para o entregável da nossa disciplina, vocês realizarão todas as etapas que fazem parte do dia a dia de um cientista de dados. São fornecidos 4 arquivos `csv` (Teams e Portal) com diferentes informações sobre poluição envolvendo plástico, e um arquivo com informações de qualidade de água para diferentes regiões. Uma coluna em comum entre todas as bases (chave primária) de dados é a coluna Entidade, que contém o nome do país associado.

Requisitos:

- **[30 pts]** Fazer uma análise exploratória de dados com os arquivos 1 a 4 (ao menos um gráfico para cada arquivo), gerando visualizações que possam trazer insights sobre a relação do plástico com o oceano e a poluição.
- **[30 pts]** Queremos treinar um modelo de ML para prever a qualidade da água do oceano com base em dados de plástico. Combine o arquivo 5 com os arquivos de 1 a 4 (se guiando pela chave primária, e lembrando que dados do mundo real possuem problemas que precisam de atenção especial). Após juntar os dados, treine um modelo para prever a qualidade da água com base nas outras colunas.
- **[40 pts]** Vídeo de no máximo 5 minutos, apresentando o projeto: mostre as visualizações geradas, faça um *storytelling* com dados, mostre a engenharia de dados para construção de uma única base à partir de 5 arquivos, e explique como o aprendizado de máquina pode ajudar a prever o impacto da poluição no oceano.

Entrega:

- Arquivo **ipynb** notebook com códigos desenvolvidos;
- Arquivo **pdf** da apresentação (se for usada no vídeo);
- Arquivo **txt** com nomes dos integrantes e link do vídeo.