**PROCESSO DE SELEÇÃO MORNING STAR**

**ANÁLISE DE DADOS**

***CASE DE TREINAMENTO***

A MStar Oil, empresa especializada em perfuração e extração de petróleo, com décadas de experiência no setor, em vista da enorme quantidade de dados que gerou ao longo dos anos, decide iniciar um projeto de análise de dados. O intuito é identificar informações que auxiliem a companhia na tomada de decisões estratégicas em relação aos seus projetos de perfuração.

Foi atribuído a você o desenvolvimento de um modelo de correlações de poços. O trabalhoconsiste em encontrar poços correlatos, ou seja, poços que tenham características muito próximas umas das outras.

Agora, vamos às definições:

Quando falamos de poços, estamos nos referindo a projetos de perfuração de poços. Esses, por sua vez, serão definidos por um conjunto de características. Logo, **entenda poço como o conjunto de dados do projeto de criação do mesmo.**

Características de um poço:

Nome: Nome que identifica o projeto do poço.

Fase: É o número da fase. Um projeto é dividido em fases.

Tipos de Poço: Existem aqui dois tipos de poços - **VERTICAL** e **HORIZONTAL**.

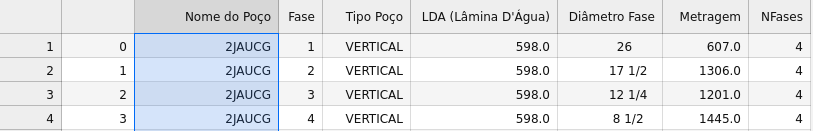
LDA (Lâmina D'Água): LDA é o valor em metros que corresponde a distância entre a superfície do mar até o ponto de perfuração em poços subaquáticos.

Diâmetro Fase: Para cada fase da perfuração, pode-se usar uma broca com diâmetro diferente. Aqui, o diâmetro é descrito em polegadas.

Metragem: Valor em metros que corresponde ao total perfurado naquela fase.

NFases: Corresponde ao número total de fases que o projeto exigiu.

A imagem a seguir ilustra o que seria o conjunto de dados que caracteriza um projeto de poço. O exemplo em questão possui quatro fases, e cada linha contém os dados de cada fase. **Lembre-se de que você encontrará projetos com números de fases distintos. Portanto, fique atento!**

  
Você receberá um arquivo com uma grande amostra de dados chamado **dados\_pocos.csv.**

Agora que você já tem uma ideia do que é um poço, considere o exemplo da imagem anterior como sendo a entrada de teste. Quais seriam os poços com as características mais próximas do mesmo? A resposta para tal pergunta será uma lista de poços correlatos.

**Bem, esta é a sua missão neste *case*:** utilizando a linguagem *Python* e sua vasta caixa de ferramentas de análise de dados (Pandas, NumPy, Sklearn etc.), bem como os diversos algoritmos já existentes para resolução de problemas, desenvolva um modelo treinado com os dados contidos no arquivo **dados\_pocos.csv**. Você deverá definir um poço como entrada, fazendo com que seja retornada uma lista com os quatro poços mais correlatos ao primeiro.

**Observações:**

* Você está livre para usar qualquer ferramenta de *Machine Learning* e análise de dados, desde que utilize a linguagem *Python*;
* Você pode enviar o seu projeto na forma que preferir: *Script Python*, *Jupyter Notebook* ou *Google Colab Notebook*;
* Toda boa prática terá peso positivo na avaliação. Lembre-se de organizar e comentar bem o seu trabalho;
* Faça as transformações que julgar necessárias na base de dados;
* Caso você tenha dificuldades em implementar códigos *Python*, toda e qualquer explicação teórica/matemática que leve à resolução do problema será levada em consideração.

**BOA SORTE!**