Olá, tudo bem?! Esperamos que sim. Ficamos felizes em saber que você chegou a esta fase e pelo interesse na Oncase. Neste passo, gostaríamos de conhecer um pouco mais sobre você através das suas habilidades em Ciência de Dados. Para tal, avaliamos os pontos a seguir, mas salientamos que os quesitos não estão, necessariamente, em ordem de prioridade:

* Sua capacidade na compreensão e modelagem dos dados e problemas;
* Qualidade e proficiência codificado na linguagem de sua escolha (R, Python);
* Qualidade das soluções encontradas para responder às perguntas;
* Comunicação na hora de questionar itens que não estejam claros;
* Comunicação visual eficaz (escrita e gráficos) no embasamento as respostas;
* Utilização de métodos adequados para cada problema;
* Análise de estudo e performance quando a resposta envolver a criação de modelos;

A avaliação será composta por duas macro tarefas, por favor não se limite a estes pontos, mas sinta-se livre para explorar e pôr em prática toda sua criatividade para solucionar os problemas.

**Obs.:** Crie um repositório no github e deixe-o público para podermos avaliar o desafio.

1. Análise descritiva dos dados (EDA)
2. Construa uma análise descritiva extraindo conhecimento das variáveis e apresentando quais *insights* podem ser obtidos a partir delas;
3. Mostre-nos um caminho para selecionar **graficamente** as variáveis mais ou menos importantes para cada problema, como elas se relacionam e porquê.
4. Em cada problema descreva quais outras técnicas poderiam ser aplicadas e porquê você não as escolheu.
5. Utilize os dados: **eda\_receitas\_data.zip**
6. Teste técnico de modelagem

Nessa parte, será necessário implementar um algoritmo de acordo com o paradigma do problema. Cada problema tem um conjunto de métricas que são requeridas. A variável alvo sempre será a coluna de nome *“target”*, exceto para os problemas não-supervisionados.

1. Construa um classificador e identifique quais variáveis exercem maior impacto sobre o “*target”* e **informe o porquê interpretando os resultados** obtidos.
   1. **métricas:** *precision, recall* e *F1-score;*
   2. **dados:** classification\_data.zip;
2. Crie um sistema de recomendação de receitas considerando que um dado cliente irá informar 5 tags. O sistema deve conter um score das melhores receitas para cada cliente. E, caso as tags mudem, as receitas recomendadas e scores também devem mudar. Além disso, na análise apresente resultados para as top 5 receitas para 3 clientes diferentes. **Interprete os resultados**.
   1. Utilize os dados: **eda\_receitas\_data.zip;**
3. Construa um regressor e identifique quais variáveis exercem maior impacto sobre o “*target”* e **informe o porquê interpretando os resultados** obtidos.
   1. **métricas:** RMSE, R2 e cor(target\_observado, target\_predito);
   2. **dados**: regression\_data.zip.