**Briefing Fictício de Ciência de Dados: Otimização do Custo de Produção na Indústria Química**

**Cliente:** QuimTech S.A. (Empresa fictícia do setor químico)

**Contexto:** A QuimTech S.A. é uma empresa de médio porte que produz diversos produtos químicos para diferentes setores da indústria. Nos últimos anos, a empresa tem observado uma volatilidade nos seus custos de produção, dificultando o planejamento financeiro e a precificação dos produtos. A empresa busca uma solução para prever e otimizar seus custos de produção, tornando-se mais competitiva no mercado.

**Problema:** A QuimTech S.A. enfrenta dificuldades em prever com precisão os custos de produção de seus produtos químicos. Fatores como variações na temperatura e pressão durante o processo produtivo, tempo de inatividade das máquinas, quantidade produzida e até mesmo o número de funcionários envolvidos parecem influenciar o custo final. A falta de um modelo preditivo robusto impede a empresa de:

* **Realizar orçamentos precisos:** Dificultando a definição de preços competitivos e a gestão financeira.
* **Identificar oportunidades de redução de custos:** Impossibilitando a otimização dos processos produtivos.
* **Antecipar potenciais problemas:** Impedindo a tomada de ações preventivas para evitar aumentos inesperados nos custos.

**Objetivo:** Desenvolver um modelo de *machine learning* que preveja o custo de produção por lote, com base em variáveis do processo produtivo, com alta precisão (idealmente com um erro médio absoluto percentual - MAPE - inferior a 5%). O modelo deverá permitir:

* **Prever o custo de produção futuro:** Com base em parâmetros conhecidos do processo.
* **Identificar os fatores que mais influenciam o custo:** Auxiliando na tomada de decisões para otimizar o processo produtivo e reduzir custos.
* **Simular cenários:** Avaliando o impacto de mudanças nos parâmetros de produção no custo final.

**Dados Disponíveis:**

* Dados históricos de produção da QuimTech S.A., contidos no arquivo CSV "dados\_industria\_quimica\_800.csv" (gerado anteriormente), contendo as seguintes variáveis:
  + **Data:** Data da produção.
  + **Produto:** Nome do produto químico produzido.
  + **Lote:** Número de identificação do lote.
  + **Quantidade Produzida (kg):** Quantidade do produto fabricada.
  + **Temperatura Média (°C):** Temperatura média durante o processo.
  + **Pressão Média (atm):** Pressão média durante o processo.
  + **Tempo de Reação (horas):** Tempo da reação química.
  + **Pureza (%):** Percentual de pureza do produto final.
  + **Número de Funcionários:** Quantidade de funcionários envolvidos.
  + **Tempo de Inatividade da Máquina (horas):** Tempo de inatividade das máquinas.
  + **Custo de Produção (R$):** Custo total da produção do lote (variável alvo).

**Entregas:**

* **Modelo de *machine learning* treinado e avaliado:** Com documentação detalhada sobre o processo de desenvolvimento, incluindo a escolha do algoritmo, as métricas de avaliação e a performance do modelo.
* **Relatório de análise:** Contendo a análise exploratória dos dados, a identificação dos fatores que mais influenciam o custo de produção e as recomendações para a QuimTech S.A.
* **Interface de usuário (opcional):** Uma interface simples (ex: uma planilha, um script Python ou um aplicativo web básico) para permitir que a equipe da QuimTech S.A. utilize o modelo para realizar previsões e simulações.

**Métricas de Avaliação:**

* **Erro Médio Absoluto Percentual (MAPE):** Principal métrica. Busca-se um MAPE inferior a 5%.
* **Raiz do Erro Quadrático Médio (RMSE):** Métrica secundária para avaliar a magnitude dos erros.
* **R-quadrado (R²):** Métrica para avaliar o ajuste do modelo aos dados.

**Restrições:**

* O modelo deve ser interpretável, ou seja, deve ser possível entender a influência das variáveis no custo de produção. Modelos de caixa preta com baixa interpretabilidade devem ser evitados, a menos que ofereçam ganhos significativos em precisão.
* O desenvolvimento deve ser realizado utilizando Python com bibliotecas como pandas, scikit-learn e matplotlib (ou outras bibliotecas relevantes).

**Cronograma:**

* A ser definido em conjunto com a QuimTech S.A.

**Observações:**

Este briefing é fictício e tem como objetivo fornecer um contexto para a aplicação de técnicas de ciência de dados e *machine learning*. Os dados fornecidos (arquivo "dados\_industria\_quimica\_800.csv") são simulados e podem não representar a realidade de uma indústria química.

Este briefing serve como ponto de partida para um projeto de ciência de dados. Detalhes adicionais podem ser definidos em reuniões com o cliente (QuimTech S.A.).

Este briefing, juntamente com os dados gerados anteriormente, oferece um excelente cenário para praticar suas habilidades em ciência de dados, desde a análise exploratória até a construção e avaliação de modelos de *machine learning*. Você pode escolher diferentes algoritmos de regressão (regressão linear, árvores de decisão, *random forest*, etc.) e comparar seus resultados. Lembre-se de realizar uma boa análise exploratória dos dados para entender as relações entre as variáveis e preparar os dados adequadamente para o treinamento do modelo.