

# Data Science aplicada a otimização do planejamento da manutenção

## Sumário

[Sumário](#)

[Situação](#)

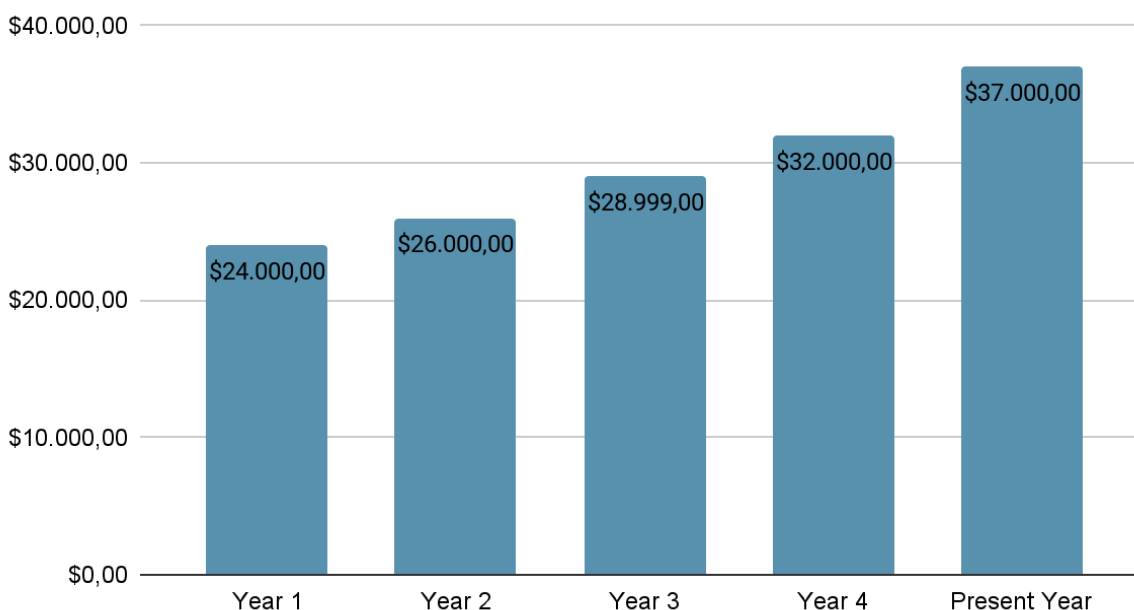
[Sobre a base de dados](#)

[Atividades do desafio](#)

## Situação

Uma nova consultoria de *data science* foi contratada para resolver e melhorar o planejamento da manutenção de uma empresa terceirizada de transporte de caminhões. A empresa conta uma frota média de caminhões para realizar entregas por todo país, porém nos últimos 3 anos vem percebendo um grande aumento nos gastos referentes à manutenção do sistema de ar de seus veículos, mesmo mantendo o tamanho de sua frota constante. O custo com manutenção deste sistema específico é mostrado abaixo em dólar:

Custo com manutenção do sistema de ar



Seu objetivo como consultor é diminuir os custos de manutenção deste sistema nos caminhões. Os custos de manutenção para o sistema de ar podem variar de acordo com o real estado do caminhão.

- Caso um caminhão seja mandado para manutenção, porém o mesmo não apresente nenhum defeito neste sistema, será cobrado em torno de \$10 pelo tempo gasto durante a vistoria pela equipe especializada.
- Caso um caminhão seja mandado para manutenção e o mesmo tenha defeito neste sistema será cobrado \$25 pela realização do serviço de reparo preventivo.
- Caso um caminhão com defeitos no sistema de ar não seja enviado diretamente para manutenção, a empresa desembolsa em \$500 para realizar a manutenção corretiva do mesmo, considerando a mão de obra, troca de peças e outros possíveis transtornos (caminhão quebrou no meio da pista por exemplo).

Durante a reunião de alinhamento com os responsáveis do projeto e a equipe de TI da empresa, algumas informações lhe foram passadas:

- A equipe técnica lhe informou que todas as informações referentes ao sistema de ar dos caminhões serão disponibilizadas para você, porém por motivos burocráticos referente aos contratos da companhia, todas as colunas precisavam ser codificadas.
- A equipe técnica também lhe informou que dado a recente digitalização da empresa, algumas informações podem estar ausentes da base de dados que lhe foi enviada.

Por último, a equipe técnica te informou que a fonte de informações advém do setor de manutenção da empresa, onde eles criaram uma coluna na base de dados chamada **class**: “pos” seriam aqueles caminhões que apresentavam defeitos no sistema de ar e “neg” seriam aqueles caminhões que apresentavam defeito em qualquer outro sistema que não fosse o de ar.

Os responsáveis pelo projeto estão bem animados com a iniciativa e ao pedir uma prova técnica colocou como requisitos principais:

- Conseguimos diminuir nossos gastos com este tipo de manutenção utilizando técnicas de AI?
- Você consegue me apresentar os principais fatores que alertam uma possível falha neste sistema?

Estes pontos, segundo eles, são importantes para convencer a diretoria executiva a abraçar a causa e aplicá-la para outros sistemas de manutenção no próximo ano.

## Sobre a base de dados

Dois arquivos serão enviados para você:

- *air\_system\_previous\_years.csv*: Arquivo contendo todas as informações do setor de manutenção dos anos anteriores com 178 colunas.
- *air\_system\_present\_year.csv*: Arquivo contendo todas as informações do setor de manutenção no ano vigente.
- Qualquer valor faltante na base de dados é denotado por *na*.

Os resultados finais que serão apresentados para a diretoria executiva precisam ser avaliados com *air\_system\_present\_year.csv*.

## Atividades do desafio

Para resolver este problema desejamos que você responda às seguintes perguntas:

1. Quais passos você realizaria para solucionar este problema? Por favor descreva da forma mais completa e clara possível todos os passos que você enxerga como essenciais para a resolução do problema.
2. Qual métrica **técnica** de *data science* você utilizaria para solucionar este desafio? Ex: erro absoluto, rmse, etc.
3. Qual métrica de **negócio** você utilizaria para solucionar o desafio?
4. Como as métricas técnicas se relacionam com a de negócio?
5. Quais tipos de análises você gostaria de realizar na base de dados do cliente?
6. Quais técnicas você utilizaria para reduzir a dimensionalidade do problema?
7. Quais técnicas você utilizaria para selecionar variáveis para seu modelo preditivo?
8. Quais modelos preditivos você utilizaria ou testaria para este problema? Por favor indique pelo menos 3.
9. Como você avaliaria qual dos modelos treinados é o melhor?

10. Como você explicaria o resultado de seu modelo? É possível saber quais variáveis são mais importantes?
11. Como você avaliaria o impacto financeiro do modelo proposto?
12. Quais técnicas você utilizaria para realizar a otimização de hiperparâmetros do modelo escolhido?
13. Quais riscos ou cuidados que você levaria para o cliente antes de colocar este modelo em produção?
14. Caso o seu modelo preditivo seja aprovado, como você colocaria ele em produção?
15. Caso o modelo esteja em produção, como você faria seu acompanhamento?
16. Caso o modelo esteja em produção, como você saberia o momento de retreinar o modelo?