



Data Science Challenge

Data & Analytics Team

O desafio consiste em 2 exercícios independentes que vão desde análise exploratória, machine learning ou o design de uma solução de data science.

O que avaliamos?

O desafio busca avaliar diferentes aspectos como:

- Capacidade analítica e exploração de dados
- Visualização de resultados
- Conhecimentos de técnicas de geração de features e modelagem
- Boas práticas de desenvolvimento
- Design e implementação de Machine Learning

Algumas regras e recomendações:

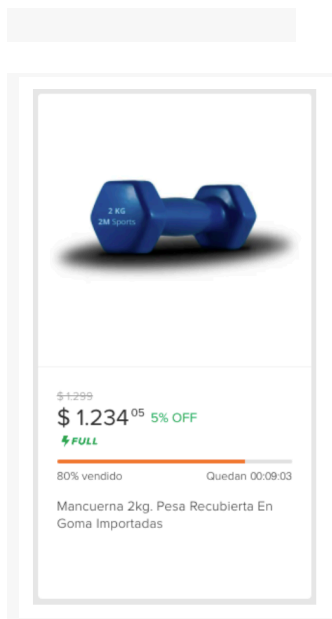
A maioria dos exercícios deve ser resolvida em Jupyter notebooks e recomendamos que você os suba para um repositório público no GitHub para compartilhar os resultados. Não deixe de nos fazer perguntas sobre qualquer dúvida com os enunciados. O desafio é analisado de acordo com a senioridade do candidato e levando em conta também as necessidades particulares da posição.

1. Explore ofertas relâmpago. Que insights você pode gerar?

Descrição

Em conjunto com o desafio compartilhamos um arquivo chamado "relampago_relampago.csv" que contém informações sobre os resultados de ofertas do tipo relâmpago para um período de tempo e um país específico.

Essas ofertas no mercadolibre podem ser vistas da seguinte forma :



Ou seja, são ofertas que têm duração definida de algumas horas e um percentual de unidades (estoque) comprometidas .

O objetivo deste desafio é fazer uma EDA sobre estes dados procurando insights sobre este tipo de ofertas.

As colunas do conjunto de dados são autoexplicativas, mas você pode nos fazer qualquer pergunta.

Entregável :

O resultado deste desafio é um notebook Jupyter com EDA.

2. Previsão de Falha

Descrição

Os armazéns no mercado livre possuem uma frota de máquinas que transmitem diariamente medidas sobre seu status e funcionamento .

As técnicas de manutenção preditiva são projetadas para ajudar a determinar a condição do equipamento de manutenção em serviço para prever quando a manutenção deve ser realizada. Esta abordagem promete economia de custos em relação à manutenção preventiva de rotina ou baseada no tempo, porque as tarefas são executadas apenas quando justificada E m média uma interrupção por falha custa 4x a mais do que uma interrupção preventiva.

O arquivo "full_devices.csv" possui os valores diários de 9 atributos dos dispositivos ao longo do tempo e a coluna que você está tentando prever é chamada de 'failure' com o valor binário 0 para nenhuma falha e 1 para falha, a coluna do dispositivo possui o id do dispositivo

Entregável

O objetivo é gerar um notebook Jupyter com um modelo para prever a falha do dispositivo **antes** de uma possível falha. Tente calcular e maximizar a possível economia gerada pelo modelo.

