

Desafio Técnico – Hackathon Forecast Big Data 2025



Objetivo

Você deverá desenvolver um **modelo de previsão de vendas (forecast)** para apoiar o varejo na reposição de produtos. A tarefa é prever a quantidade semanal de vendas por **PDV (Ponto de Venda) /SKU Stock Keeping Unit**

(ou **Unidade de Manutenção de Estoque**) para as **cinco semanas de janeiro/2023**, utilizando como base o histórico de vendas de **2022**.

Esse é um problema real, baseado no nosso produto **One-Click Order**.



Dados Disponíveis

Você terá acesso a um conjunto de dados:

- **Dados de treino (2022)**
 - **Este é o conjunto de dados que você e sua equipe irão usar para trabalhar, criar o modelo, fazer testes e desenvolver a solução final.**
 - Transações: Data, PDV, Produto, Quantidade, Faturamento.
 - Cadastro de produtos: Produto, Categoria, Descrição, + até 4 atributos.
 - Cadastro de PDVs: PDV, On/Off Prem, Categoria (c-store, g-store, liquor etc.), Zipcode.
- **Dados de teste (Jan/2023) - Não será compartilhado com os participantes.**
 - Esse é o conjunto de dados em que sua solução será avaliada. Vamos comparar a sua previsão com o dado real.
 - Mesma estrutura dos dados de treino.
 - **Não será compartilhado com os participantes.**
 - Usado apenas pela Big Data para avaliar as previsões enviadas.



Entregáveis

Sua equipe deve submeter:

Arquivo de previsão no formato **CSV ou Parquet**, com as seguintes colunas:

semana	pdv	produto	quantidade
1	1023	123	120
2	1045	234	85
3	1023	456	110

No caso do csv, utilize “;” como caractere separador (exemplo: 1;1023;123;120) e encoding UTF-8.

1. **semana (número inteiro)**: número da semana (1 a 4 de janeiro/2023)
2. **pdv (número inteiro)**: código do ponto de venda
3. **produto (número inteiro)**: código do SKU
4. **quantidade (número inteiro)**: previsão de vendas

2. Repositório público no GitHub com:

- Código completo e documentação da solução.
- Instruções claras de execução (README).



Submissões

- Cada participante/equipe poderá realizar **até 5 submissões** durante o período do desafio.
- Apenas o **melhor resultado** será considerado para efeito de ranking.
- A submissão é feita pelo site oficial do Hackathon [link aqui](#).
- Ao submeter, sua solução será analisada e comparada com os números reais dos dados de teste.
- O leaderboard será atualizado em até 20 minutos após cada submissão, exibindo nome do participante e/ou codinome da equipe, WMAPE (%) e posição no ranking — ordenado em crescente (quanto menor, melhor).



Avaliação

As soluções serão avaliadas com base em:

1. **Performance do modelo** – a métrica oficial de avaliação será divulgada no site do hackathon.
2. **Qualidade técnica da entrega** – clareza, organização e documentação do código.
3. **Criatividade na abordagem** – estratégias de modelagem e tratamento de dados.
4. **Comparação com baseline da Big Data** – o modelo precisa superar a solução interna da empresa para ser considerado válido.

⚠ Importante: mesmo que sua solução esteja bem posicionada no leaderboard, ela poderá ser invalidada se não atender aos critérios de execução (código não executável, resultado inconsistente ou incompleto).



Cronograma

- **Divulgação e inscrições:** até 08/09
- **Lançamento do desafio:** 09/09
- **Submissões:** 09 a 21/09
- **Validação técnica:** 22 a 26/09
- **Anúncio dos vencedores:** 29/09



Premiação

- **1º lugar:** R\$ 30.000
- **2º lugar:** R\$ 10.000



Comunicação

- Nosso **canal oficial de comunicação** será o e-mail **hackathon@bdtech.ai**: fique atento à sua caixa de entrada.
- Também vamos compartilhar conteúdos e bastidores nas redes sociais da Big Data.

 **Boa sorte! Esse é o momento de mostrar todo o seu talento em Ciência de Dados em um problema real de mercado.**