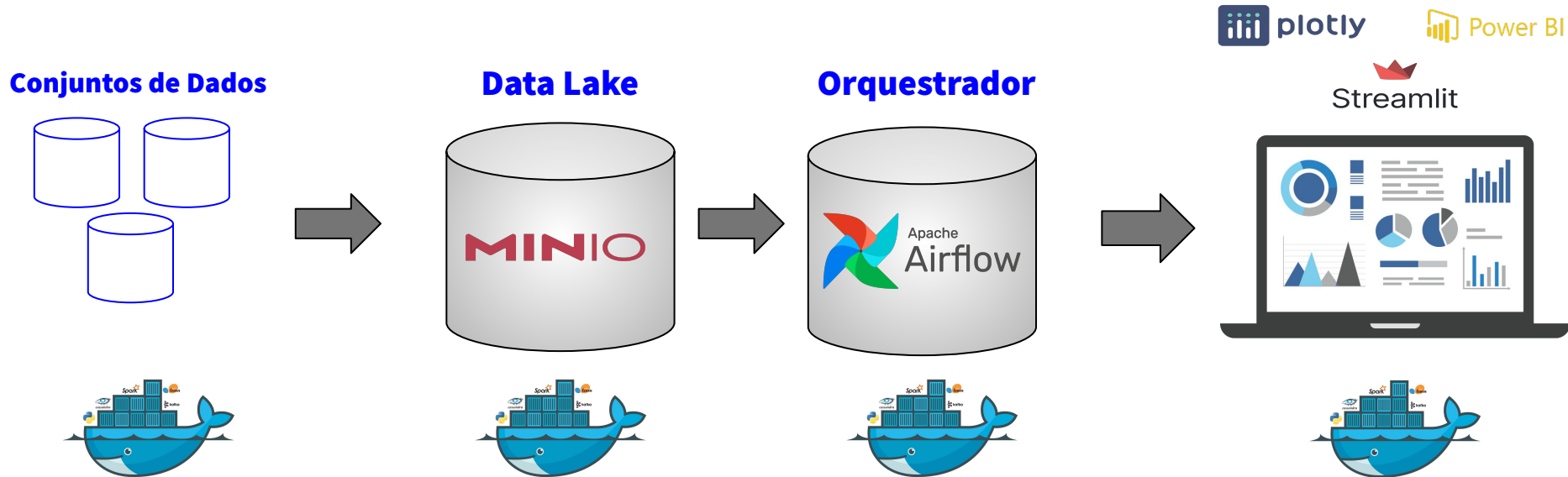


# Stack Labs



# Desafios

#olistpaganos

InfoMoney <a href="#">Como vai funcionar o Open Finance?</a> • <a href="#">Saia da Poupança</a> • <a href="#">CDB 1% ao mês</a> • <a href="#">Altas e Baixas</a> • <a href="#">IM+</a>									
+3,91%	ITUB4 R\$ 21,93 +1,01%	ABEV3 R\$ 16,10 +0,12%	GGBR4 R\$ 29,63 +2,30%	IBOVESPA 108.326 pts +0,83%	DÓLAR R\$ 5,67 -0,72%	BITCOIN R\$ 275.577 -1,63%	IFIX 2.681 pts +0,19%	MGLU3 R\$ 6,40 +3,73%	

Comércio eletrônico

## Olist se torna mais novo unicórnio brasileiro, incluindo empreendedores no e-commerce

Startups cresceu três vezes em 2021, conectando vendedores a marketplaces como Amazon e Mercado Livre. Futuro é de expansão de soluções e de geografias



Por Mariana Fonseca

15 dez 2021 15h05



→ Mais sobre

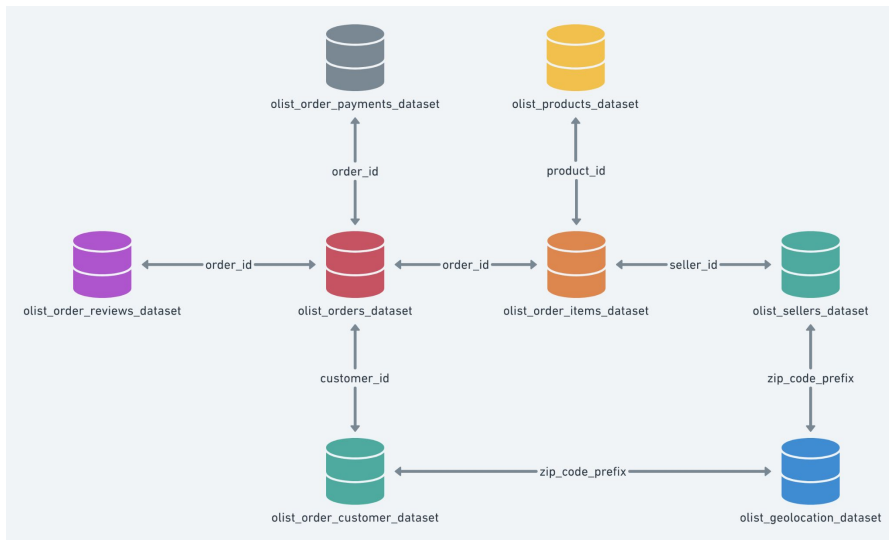
[Do Zero ao Topo](#)  
[Comércio Eletrônico](#)  
[E-commerce](#)  
[Startup](#)  
[Unicórnios](#)  
[Venture Capital](#)



# Desafios

Conjuntos de Dados: <https://www.kaggle.com/olistbr/brazilian-ecommerce>

- Agrupamento de clientes.
- Análise de sentimentos.
- Previsão de vendas.
- Classificação automática de produtos.



# Desafios

## Agrupamento e Segmentação de Clientes

### Objetivo

Produzir um modelo de agrupamento para segmentar clientes em categorias como cliente ouro, prata ou bronze baseado em atributos como **receita média**, **quantidade de pedidos**, **retenção**, **recência**.

### Entregáveis

1. Data Pipeline para automatizar as etapas de ETL.
2. Modelo de agrupamento que dado um determinado conjunto de teste o modelo consiga agrupar o cliente em um determinado grupo.
3. Relatório com insights gerados a partir dos dados.



# Desafios

## Análise de Sentimentos/NLP

### Objetivo

Produzir um modelo de Machine Learning para classificar o sentimento dos usuários em relação às compras de acordo com os reviews dos produtos.

### Entregáveis

1. Data Pipeline para automatizar as etapas de ETL.
2. Modelo de classificação que dado um determinado conjunto de teste o modelo consiga retornar a classificação em positivo, negativo ou neutro.
3. Relatório com os insights gerados a partir dos dados.



# Desafios

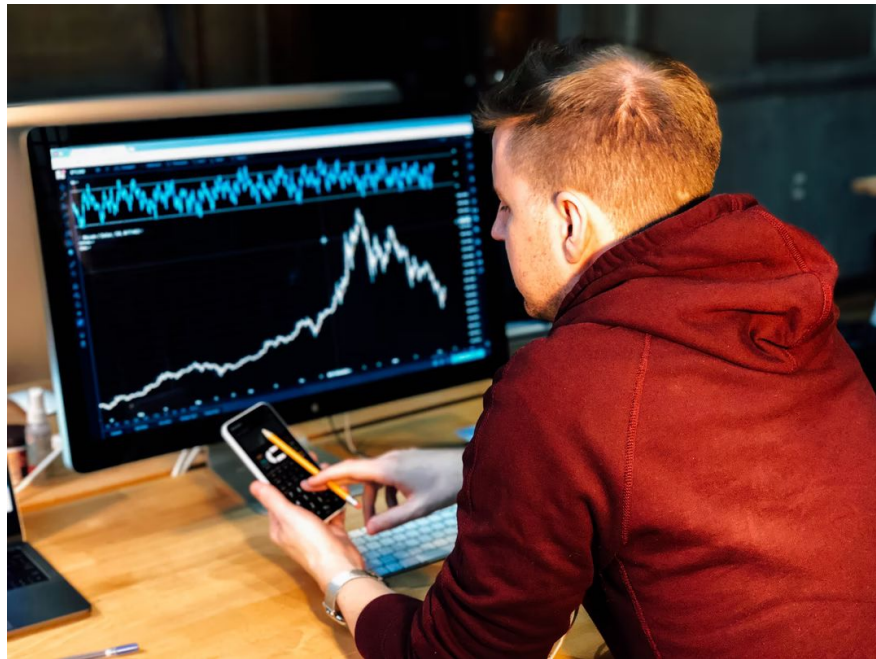
## Previsão de Vendas

### Objetivo

Produzir um **modelo de previsão baseado nas vendas dos produtos** da base de dados com o objetivo de prever o comportamento ao longo do tempo.

### Entregáveis

1. Data Pipeline para automatizar as etapas de ETL.
2. Modelo de previsão de vendas que dado um determinado período entregue o quantitativo de vendas e volume.
3. Relatório com os insights gerados a partir dos dados.



# Desafios

## Análise de Dados e Entendimento de Negócio

### Objetivo

Produzir um relatório visual e dinâmico detalhado com uma análise de negócio feita a partir dos dados com o foco em explorar a base de dados e **identificar padrões, oportunidades, possibilidade de criação de produtos e soluções.**

### Entregáveis

1. Data Pipeline para automatizar as etapas de ETL.
2. Relatório com os insights gerados a partir dos dados.





## Etapas e gestão do projeto

- Definição das equipes e funções de cada membro.
  - Quem fará a função de Gestor de Projeto?
  - Quem será o Analista, Cientista e Engenheiro de Dados?
- Definição de tarefas e prazos.
  - Definir as tarefas a serem realizadas e qual o prazo para executar.
- Comunicação e colaboração.
- Ferramentas propostas
  - Discord, Notion, Asana, Trello.



DISCORD



Trello



asana



Stack



## Etapas do projeto

- **Documentação**
  - Desenhar a arquitetura da solução e seus componentes.
- **Criação de containers para infraestrutura.**
  - Banco de Dados, Minio, Airflow.
- **Download dos conjuntos de dados e persistência no Data Lake.**
  - Subir os arquivos para um bucket no Minio.
- **Processamento e limpeza de dados.**
  - Data Cleaning.
  - Selecionar colunas.
    - Relevância, Junções, Remoção.
  - Transformar dados
    - Formatos (data, hora, data types)
    - Missing values.
    - Distribuições.
  - Padronizar formato para parquet.
  - Armazenar os dados processados em outro bucket.



# Etapas do projeto

- **Análise Exploratória de Dados.**
  - Destacar insights e padrões descobertos.
  - Documentar toda a análise de dados.
- **Modelagem e pré-processamento de dados.**
  - Separar as features para criação de modelos.
  - Aplicar técnicas e métodos de pré-processamento de dados.
  - Testar e avaliar a performance de algoritmos.
  - Persistir melhor modelo no Data Lake.
- **Desenvolver pipeline para automação de todas as etapas do projeto.**
  - Desenvolvimento de Dags
- **Criar Dashboard ou App de Visualização de dados:**
  - Plotly
  - PowerBI, Tableau, Streamlit.



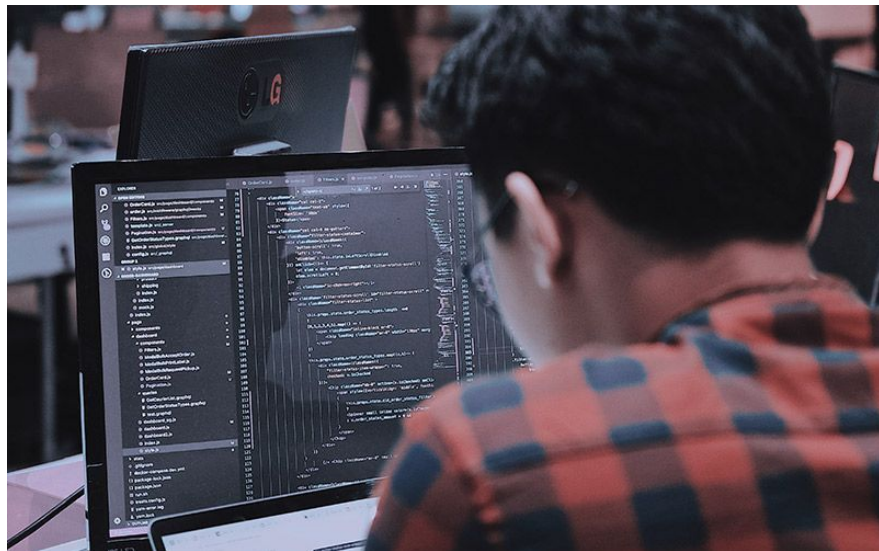
# Definição de Squads e Cronograma

- Limitação de membros.
  - A ideia é todos contribuírem para criar um projeto completo.
- Organização por função (Analista, Cientista, Engenheiro)
- Inscrição via formulário.
  - <https://forms.gle/d5sERwcPmHinpCMg6>
  - Lista divulgada na **segunda-feira 20/12**.
- Colaboração, networking, aprendizado contínuo.
- Canais das equipes no Discord.
- Data para mentorias: **06/01 e 20/01**.
- Data de apresentação: **27/01**.
  - Apresentar a solução desenvolvida com sua arquitetura, componentes, tecnologias utilizadas e o resultado.
  - Tempo de 15 à 20 minutos.



## Premissas e pontos para levar em consideração.

- Organização
  - Código.
  - Artefatos.
  - Documentações.
- Tecnologias.
  - Explore as diferentes tecnologias e saiba o porquê de cada uma na solução desenvolvida.
- Foco na resolução do problema.
  - Melhor que aprofundar em apenas um item é desenvolver todo o projeto.
  - O seu projeto SEMPRE terá melhorias.
- Se possível trazer algo funcionando.
  - Uma API, Dashboard, App.
- Vamos divulgar o seu trabalho.



# Dúvidas?