



# Como funciona...



## Dados

- Públicos ou Privados.



## Escopo e Objetivos

- Definição de desafios e problemas a serem resolvidos.



## Comunidade

- Organização em squads.
- Formações e suporte.



## Mentorias e Gestão

- Mentoria e acompanhamento de profissionais experientes.
- Gestão de projetos.



## Entrega de Valor

- Resultados, produtos, soluções.
- Transferência de conhecimento.



# Desafios

## Credit Card Fraud Detection

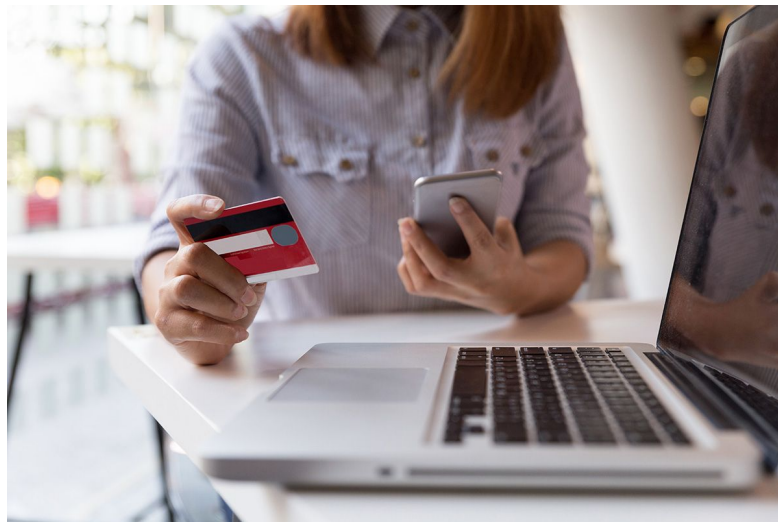
<https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>

### Objetivo

Produzir um modelo de detecção de fraude em transações de cartão de crédito.

### Entregáveis

1. Data Pipeline para automatizar as etapas de ETL.
2. Modelo de classificação, detecção de anomalias.
3. Relatório com insights gerados a partir dos dados.



# Desafios

## Cryptocurrency Historical Prices

<https://www.kaggle.com/datasets/sudalairajkumar/cryptocurrencypricehistory>

## Objetivo

Produzir uma solução para análise e predição de dados de criptomoedas.

## Entregáveis

1. Data Pipeline para automatizar as etapas de ETL.
2. Modelo de estimativa de preços, variações, tendências, sentimentos para ajudar o investidor a tomar a melhor decisão.
3. Relatório ou Dashboard com os insights gerados a partir dos dados.



# Desafios

## Daily News for Stock Market Prediction

<https://www.kaggle.com/datasets/aaron7sun/stocknews>

## Objetivo

Produzir um uma solução que use dados textuais de notícias para estimar um comportamento futuro do mercado de ações.

## Entregáveis

1. Data Pipeline para automatizar as etapas de ETL.
2. Modelo de previsão de comportamento e tendência, análise de sentimentos baseado nas notícias.
3. Relatório ou Dashboard com os insights gerados a partir dos dados.

## \* APIS

StockTwits Crypto -> <https://stocktwits.com/symbol/BTC.X>

Coin Market Api -> <https://coinmarketcap.com/api/documentation/v1/>



# Desafios

## Santander Product Recommendation

<https://www.kaggle.com/competitions/santander-product-recommendation/data>

### Objetivo

Produzir um modelo de Machine Learning para aprender a recomendar produtos aos clientes baseado em suas preferências e histórico de consumo .

### Entregáveis

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





# Desafios

## Predicting Churn for Bank Customers

<https://www.kaggle.com/datasets/adammaus/predicting-churn-for-bank-customers>

### Objetivo

Produzir um modelo de Machine Learning para aprender a identificar clientes que são propensos a deixar a empresa.

### Entregáveis

1. Data Pipeline para automatizar as etapas de ETL.
2. Modelo de detecção de churn.
3. Relatório ou Dashboard 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 executá-las.
- Comunicação e colaboração.
- Ferramentas propostas
  - Discord, Notion, Asana, Trello.



DISCORD



Trello



asana



## 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.



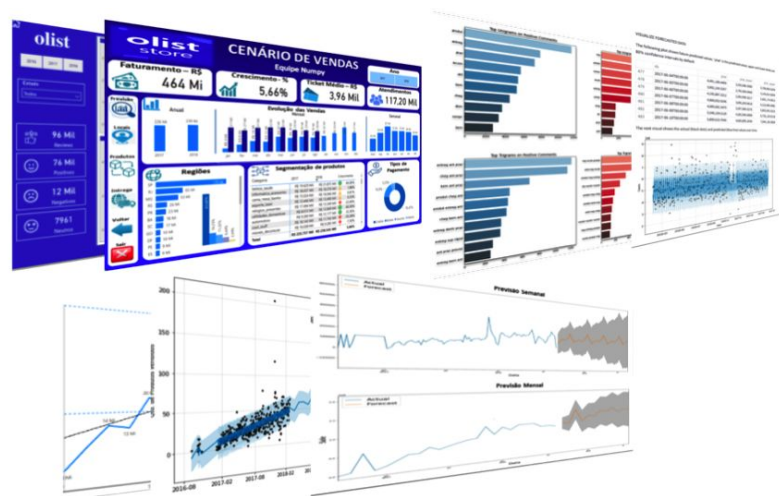
## Etapas do projeto

- **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.
- Modelo trello: <https://bit.ly/stacklabs-tarefas>



# Entregáveis

1. Data Pipeline para automatizar todas as etapas de um projeto de dados.
2. Modelo preditivo que dado um determinado conjunto de características o modelo consiga detectar se a pessoa está com alguma das doenças ou não e a sua probabilidade.
3. Código fonte e artefatos do projeto.
4. Listagem de fatores que influenciam a decisão do modelo.
5. Relatório com insights gerados a partir dos dados.
6. Dashboard para tomada de decisão.
7. Data App ou API para consumo do modelo.
8. Apresentação em formato slide ou pdf.



# Definição de Squads e Cronograma

- Limitação de membros.
  - A idéia é todos contribuírem para criar um projeto completo.
- Organização por função (Analista, Cientista, Engenheiro)
- Inscrição via formulário.
  - <https://bit.ly/inscricoes-labs-financas>
  - Lista divulgada na **segunda-feira**.
- Colaboração, networking, aprendizado contínuo.
- Canais das equipes no Discord.
- Cronograma.
  - Entregáveis, data das mentorias e apresentação.
    - 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.

- Apoio e Colaboração
  - **Giuliano Ferreira.**
- 
- 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.



## Banco de Talentos

- Parceria com empresas.
- Contato para entrevistas e processos seletivos.
- Inscrição via formulário.
  - <https://bit.ly/stack-banco-talentos>
- Dados protegidos.
- Para quem já está trabalhando ou não.

