

## Challenge Telecom X: análise de evasão de clientes

# Challengeers ONE

## Perguntas frequentes

### O que são os Challenges ONE?

É uma forma de implementar o Challenge Based Learning, ou seja, aprendizado baseado em desafios que a Apple ajudou a criar. É um mecanismo onde você pode se comprometer na resolução de um problema para depois investigar soluções com cursos, conteúdo e palestras; ou até mesmo com o conhecimento que você já tem! Finalmente, você vai agir e colocar o seu projeto no ar. Tudo isso enquanto comenta e ajuda em projetos de outros alunos e alunas.

### Se eu tiver dúvidas, com quem posso falar?

No nosso grupo oficial do Discord, temos canais específicos para os Challenges. Lá você poderá consultar suas dúvidas com os instrutores e instrutoras e com o resto dos participantes. Também teremos Lives de mentoria "Alura responde" durante os desafios para consultar as principais dúvidas.

### Obterei um certificado?

Uma vez realizada a entrega do seu projeto, você poderá baixar o seu certificado e o badge referente ao mesmo, que deverá compartilhar em suas redes profissionais como LinkedIn, etiquetando a Oracle e a Alura Latam. Lembre-se que os Challenges são parte integral do programa ONE e seu objetivo é criar projetos incríveis e incluí-los no seu portfólio para demonstrar suas habilidades.

### Quais materiais ou ferramentas preciso instalar para realizar o curso?

Todas as etapas dos Challenges terão especificações diferentes que iremos avisando conforme os desafios forem liberados. Te guiaremos durante todo o processo para que você possa finalizar o seu projeto sem preocupações.

### Como saberei sobre os novos desafios?

Enviaremos e-mails de lembrete toda vez que um novo desafio ou Live estiver próximo de começar. É importante verificar frequentemente as pastas de Spam, Promoções e Social no seu e-mail. Às vezes, os e-mails podem ir para lá.

# Importância do desafio

## Desafio Telecom X

O **desafio Telecom X** oferece uma oportunidade única para aplicar habilidades essenciais de **análise de dados** em um cenário de negócios real.

# Aplicação prática do conhecimento

A **limpeza e tratamento de dados** é uma habilidade fundamental para qualquer analista de dados. A manipulação de grandes volumes de informações exige a capacidade de identificar e corrigir inconsistências nos dados, como **valores nulos, duplicados** e dados fora do padrão. Garantir que os dados estejam prontos para análise é um passo essencial para obter resultados precisos e confiáveis.

**Análise exploratória de dados (EDA)** é uma etapa crucial no entendimento profundo dos dados. A capacidade de aplicar estatísticas descritivas e gerar **visualizações** permite que você identifique padrões, tendências e relações entre as variáveis. Isso ajuda a formar hipóteses e gerar insights que podem influenciar decisões estratégicas dentro da empresa.

Ao participar deste desafio, você aplicará conhecimentos essenciais para a análise de grandes volumes de dados em um contexto real, onde suas descobertas poderão impactar diretamente as estratégias da empresa para melhorar o principal problema que estão tendo.

Este desafio não só contribui para o seu crescimento na área de Data Science, mas também oferece a oportunidade de entender como a **ciência de dados** pode ser aplicada para resolver problemas reais enfrentados por empresas no mercado.

## Transcrição

Olá! Seja bem-vindo ou bem-vinda a mais um Desafio de Data Science. Eu sou a Ingrid Silva e faço parte do time de instrutores aqui do Programa ONE. Por questão de acessibilidade, vou me autodescrever. Sou uma mulher negra, com cabelos cacheados tingidos de vermelho. Uso óculos de grau, com armação transparente. Também estou com uma blusa preta com o logo da Alura, e atrás de mim tem uma parede branca com luz em azul e roxo.

Hoje, venho apresentar o nosso mais novo desafio, o Telecom X. Vamos nessa? Telecom X, análise de evasão de clientes. A Telecom X é uma empresa de telecomunicações, e você foi contratado como assistente de análise de dados para fazer parte da equipe de Data Science. A equipe está com um novo projeto denominado Churn de Clientes.

A empresa vem sofrendo alto índice de evasão de clientes, porém, eles não conseguem descobrir o porquê disso estar acontecendo. E é aí que você entra como assistente de análise de dados. Você ficará responsável por uma das partes mais importantes da análise de dados. Você ficará responsável por buscar, tratar e realizar uma análise exploratória dos dados da empresa. Ou seja, você colocará em prática todos os seus conhecimentos sobre o processo de ETL.

Com os dados tratados e limpos, seus colegas mais experientes serão capazes de realizar uma análise mais profunda, e assim, criar modelos preditivos para que essa taxa de evasão diminua. Assim como nos desafios anteriores,

neste também você terá um *trello* (quadro de tarefas) que vai te ajudar no passo a passo para estar realizando esse desafio.

A parte mais importante desse desafio é a extração dos dados. Os dados da empresa estão em uma API que está armazenada dentro de um repositório no GitHub. O primeiro passo que você tem que fazer nesse desafio é descobrir a melhor maneira de fazer essa extração. Com a extração feita, você colocará em prática a parte T, ou seja, a parte de transformação. Você tem que transformar esses dados para em seguida poder analisar e gerar alguns insights dele.

Como parte final, você deverá gerar um relatório explicando detalhadamente todo o processo de ETL feito. Ou seja, você tem que começar com uma introdução, falando o objetivo da sua análise, explicando o problema da evasão de clientes, descrever o processo de importar, limpar e tratar os dados, explicar a análise exploratória, incluindo gráficos e visualização para identificar padrões e, através disso tudo, criar uma conclusão. Ou seja, resuma os principais achados e como esses dados podem ajudar a reduzir a evasão.

Também recomendamos que você ofereça sugestões baseadas nas análises para, assim, ajudar ainda mais a equipe de *data science*. Com esse desafio, você colocará em prática todo o seu conhecimento em ETL, análise exploratória, também conhecido na área de *data science* como EDA. Tudo isso utilizando Python e as suas bibliotecas nativas como Pandas, Seaborn, Matplotlib, entre outras.

Com os dados devidamente tratados e analisados, é possível identificar padrões dos clientes e, assim, evitar o *churn* (evasão) de clientes. Bom, desafio dado, mãos à obra!

## Transcrição

Olá! Seja bem-vindo ou bem-vinda a mais um Desafio de Data Science. Eu sou a Ingrid Silva e faço parte do time de instrutores aqui do Programa ONE. Por questão de acessibilidade, vou me autodescrever. Sou uma mulher negra, com cabelos cacheados tingidos de vermelho. Uso óculos de grau, com armação transparente. Também estou com uma blusa preta com o logo da Alura, e atrás de mim tem uma parede branca com luz em azul e roxo.

Hoje, venho apresentar o nosso mais novo desafio, o Telecom X. Vamos nessa? Telecom X, análise de evasão de clientes. A Telecom X é uma empresa de telecomunicações, e você foi contratado como assistente de análise de dados para fazer parte da equipe de Data Science. A equipe está com um novo projeto denominado Churn de Clientes.

A empresa vem sofrendo alto índice de evasão de clientes, porém, eles não conseguem descobrir o porquê disso estar acontecendo. E é aí que você entra como assistente de análise de dados. Você ficará responsável por uma das partes mais importantes da análise de dados. Você ficará responsável por buscar, tratar e realizar uma análise exploratória dos dados da empresa. Ou seja,

you will put into practice all your knowledge about the ETL process.

With the data cleaned and ready, your more experienced colleagues will be able to perform a deeper analysis, and thus, create predictive models so that this tax evasion rate decreases. Just like in previous challenges, in this one you will also have a *trello* (task board) that will help you in the next step to be doing this challenge.

The most important part of this challenge is the extraction of data. The data from the company is in an API that is stored within a repository on GitHub. The first step you have to do in this challenge is to discover the best way to do this extraction. Once the extraction is done, you will put into practice the T, or in other words, the transformation part. You have to transform these data so that you can then analyze and generate some insights from it.

As the final part, you will have to generate a report explaining in detail the entire ETL process. Or in other words, you have to start with an introduction, talking about the objective of your analysis, explaining the problem of tax evasion of clients, describing the process of importing, cleaning and treating the data, explaining the exploratory analysis, including graphs and visualization to identify patterns and, through all this, create a conclusion. Or in other words, summarize the main findings and how these data can help reduce tax evasion.

We also recommend that you offer suggestions based on the analyses, so, to help even more the *data science* team. With this challenge, you will put into practice all your knowledge in ETL, exploratory analysis, also known in the *data science* area as EDA. All this using Python and its native libraries like Pandas, Seaborn, Matplotlib, among others.

With the data properly treated and analyzed, it is possible to identify patterns of clients and, thus, avoid the *churn* (tax evasion) of clients. Well, challenge accepted, let's get to work!

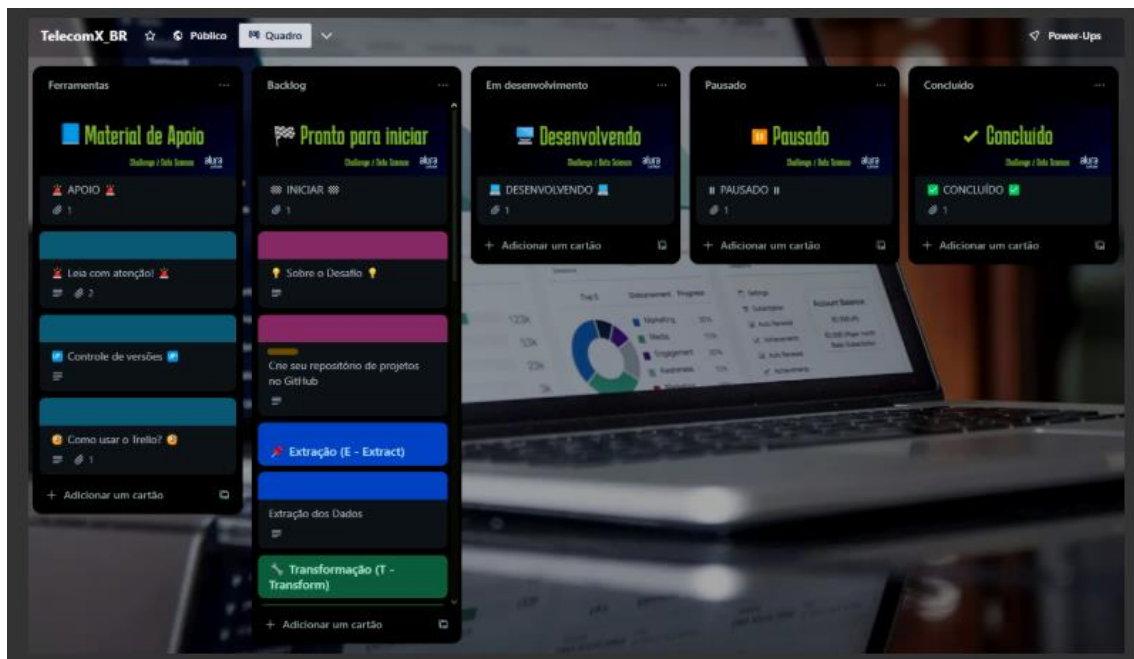
## Trello do desafio

The **Trello** is a collaboration and project management tool that allows you to organize tasks in cards. You will use this tool to manage the development of your project.

## Acesse o quadro do desafio

Log in to your Trello account and access the challenge board through the link below:

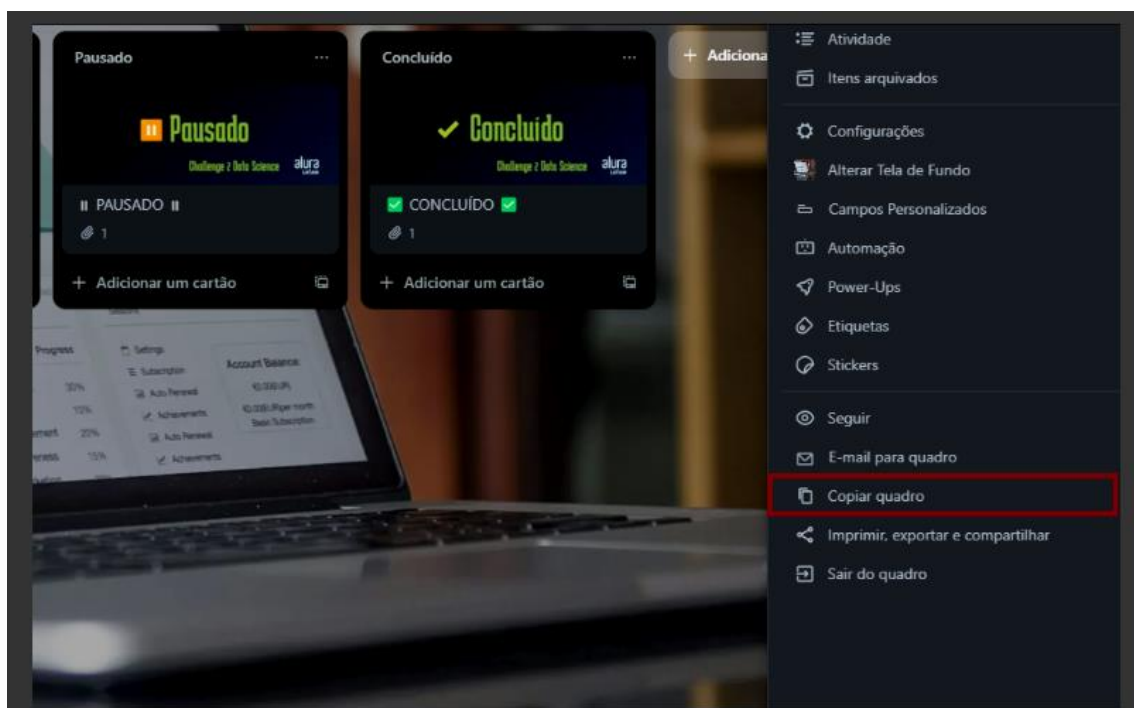
[Trello do desafio](#)



## Criando sua cópia do Trello

Para começar a gerenciar seu progresso, siga os passos abaixo:

1. Acesse o quadro do desafio;
2. **Crie um quadro a partir deste modelo:**
  - Na parte superior do quadro, clique em '**Criar quadro com base em template**' ou, se preferir, nos três pontinhos no topo e selecione '**Copiar quadro**'..



Agora você tem uma cópia do quadro, incluindo **listas, cartões e etiquetas**. Utilize essa estrutura para aplicar a metodologia ágil e acompanhar seu progresso movendo os cartões conforme avança no desafio.

## Sobre o Desafio

### Telecom X - Análise de Evasão de Clientes

Você foi contratado como assistente de análise de dados na **Telecom X** e fará parte do projeto "**Churn de Clientes**". A empresa enfrenta um alto índice de cancelamentos e precisa entender os fatores que levam à perda de clientes.

Seu desafio será **coletar, tratar e analisar os dados**, utilizando **Python** e suas principais bibliotecas para extrair insights valiosos. A partir da sua análise, os demais colegas da equipe de Data Science poderá avançar para modelos preditivos e desenvolver estratégias para reduzir a evasão.

### O que você vai praticar:

- ✓ **Importar e manipular dados** de uma API de forma eficiente.
- ✓ **Aplicar os conceitos de ETL** (Extração, Transformação e Carga) na preparação dos dados.
- ✓ **Criar visualizações de dados** estratégicas para identificar padrões e tendências.
- ✓ **Realizar uma Análise Exploratória de Dados (EDA)** e gerar um relatório com insights relevantes.

Use seus conhecimentos para transformar dados brutos em informações estratégicas e ajudar a **Telecom X** a reter mais clientes.

### Crie seu repositório de projetos no GitHub

No desenvolvimento de projetos, sabemos como é essencial organizar seu trabalho desde o início. Por isso, neste desafio, você deverá criar um repositório no GitHub para armazenar e versionar seu projeto.

Mesmo que você ainda não tenha desenvolvido nenhum código, o objetivo é que você crie uma estrutura inicial para o projeto. À medida que avançar, você poderá atualizar e adicionar arquivos ao repositório.

#### Materiais de Apoio

Se precisar de ajuda, confira os seguintes recursos:

- [Git - Configuração Inicial do Git](#)
- [Iniciando um repositório com Git | Alura](#)

Organizar seu projeto desde o início facilita o desenvolvimento e garante boas práticas no controle de versões!

### Extração dos Dados

Para iniciar sua análise, você precisará importar os dados da **API da Telecom X**. Esses dados estão disponíveis no formato **JSON** e contêm informações essenciais sobre os clientes, incluindo dados demográficos, tipo de serviço contratado e status de evasão.

### 📌 Link da API:

🔗 [challenge2-data-science/TelecomX\\_Data.json at main · ingridcrith/challenge2-data-science](https://challenge2-data-science/TelecomX_Data.json)

🔗 [GitHub - ingridcrith/challenge2-data-science](https://github.com/ingridcrith/challenge2-data-science)

## O que você deve fazer:

- ✅ Carregar os dados diretamente da API utilizando **Python**.
- ✅ Converter os dados para um **DataFrame do Pandas** para facilitar a manipulação.

Este é o primeiro passo para transformar os dados brutos em insights valiosos.

## Conhecendo o Dataset

Agora que você extraiu os dados, é fundamental entender a estrutura do dataset e o significado de suas colunas. Essa etapa ajudará a identificar quais variáveis são mais relevantes para a análise de evasão de clientes.

Para te auxiliar nesse processo, criamos um **dicionário de dados** com a descrição de cada coluna. Embora não seja obrigatório utilizá-lo, ele pode facilitar a compreensão das informações disponíveis.

## Dicionário de dados

- **customerID**: número de identificação único de cada cliente
- **Churn**: se o cliente deixou ou não a empresa
- **gender**: gênero (masculino e feminino)
- **SeniorCitizen**: informação sobre um cliente ter ou não idade igual ou maior que 65 anos
- **Partner**: se o cliente possui ou não um parceiro ou parceira
- **Dependents**: se o cliente possui ou não dependentes
- **tenure**: meses de contrato do cliente
- **PhoneService**: assinatura de serviço telefônico
- **MultipleLines**: assinatura de mais de uma linha de telefone
- **InternetService**: assinatura de um provedor internet
- **OnlineSecurity**: assinatura adicional de segurança online
- **OnlineBackup**: assinatura adicional de backup online
- **DeviceProtection**: assinatura adicional de proteção no dispositivo
- **TechSupport**: assinatura adicional de suporte técnico, menos tempo de espera
- **StreamingTV**: assinatura de TV a cabo
- **StreamingMovies**: assinatura de streaming de filmes
- **Contract**: tipo de contrato
- **PaperlessBilling**: se o cliente prefere receber online a fatura
- **PaymentMethod**: forma de pagamento
- **Charges.Monthly**: total de todos os serviços do cliente por mês
- **Charges.Total**: total gasto pelo cliente

## O que você deve fazer:

- ✓ Explorar as colunas do dataset e verificar seus tipos de dados.
- ✓ Consultar o dicionário para entender melhor os significados das variáveis.
- ✓ Identificar as colunas mais relevantes para a análise de evasão.

📌 Dicas:

- [pandas.DataFrame.info — pandas 2.3.3 documentation](#)
- [pandas.DataFrame.dtypes — pandas 2.3.3 documentation](#)

## Verificando Inconsistências nos Dados

Neste passo, **verifique se há problemas nos dados** que possam afetar a análise. Fique atento a valores ausentes, duplicados, erros de formatação e inconsistências nas categorias. Esse processo é essencial para garantir que os dados estejam prontos para as próximas etapas.

📌 Dicas:

- [pandas.unique — pandas 2.3.3 documentation](#)
- [pandas.Series.dt.normalize — pandas 2.3.3 documentation](#)

# Foco no Processo de ETL

O principal objetivo deste desafio é desenvolver suas habilidades em **ETL (Extract, Transform, Load)** com Python. Os dados para este desafio estão disponíveis em uma **API**.

### Acesse os dados da API

- [TelecomX Dados](#)

Neste desafio, o foco está no processo de **extração** de dados da API, **limpeza** e **transformação**. Após essa etapa de tratamento, você deverá organizar os dados de maneira a possibilitar análises mais profundas e visualizações.

Como **auxílio adicional**, criamos um **caderno base opcional** para ajudá-lo a estruturar melhor o desafio. O caderno fornece uma base inicial, sugerindo um fluxo de trabalho organizado e práticas recomendadas para o processo de ETL. Você pode usá-lo para se orientar durante a execução do desafio, mas também pode optar por não utilizá-lo e estruturar a solução por conta própria.

### Acesse o caderno base (opcional)

- [Caderno Base - Telecom X](#)

Ao completar esse desafio, você terá aplicado competências essenciais em ETL, o que é fundamental para o trabalho de um analista de dados e as demais áreas em Data Science.



## Para saber mais

Este desafio foi pensado para que você possa aplicar de forma prática os conhecimentos adquiridos nos seguintes cursos:

- [NumPy: análise numérica eficiente com Python](#)
- [Pandas: conhecendo a biblioteca](#)
- [Pandas I/O: trabalhando com diferentes formatos de arquivos](#)
- [Pandas: transformação e manipulação de dados](#)
- [Data Visualization: criando gráficos com bibliotecas Python](#)

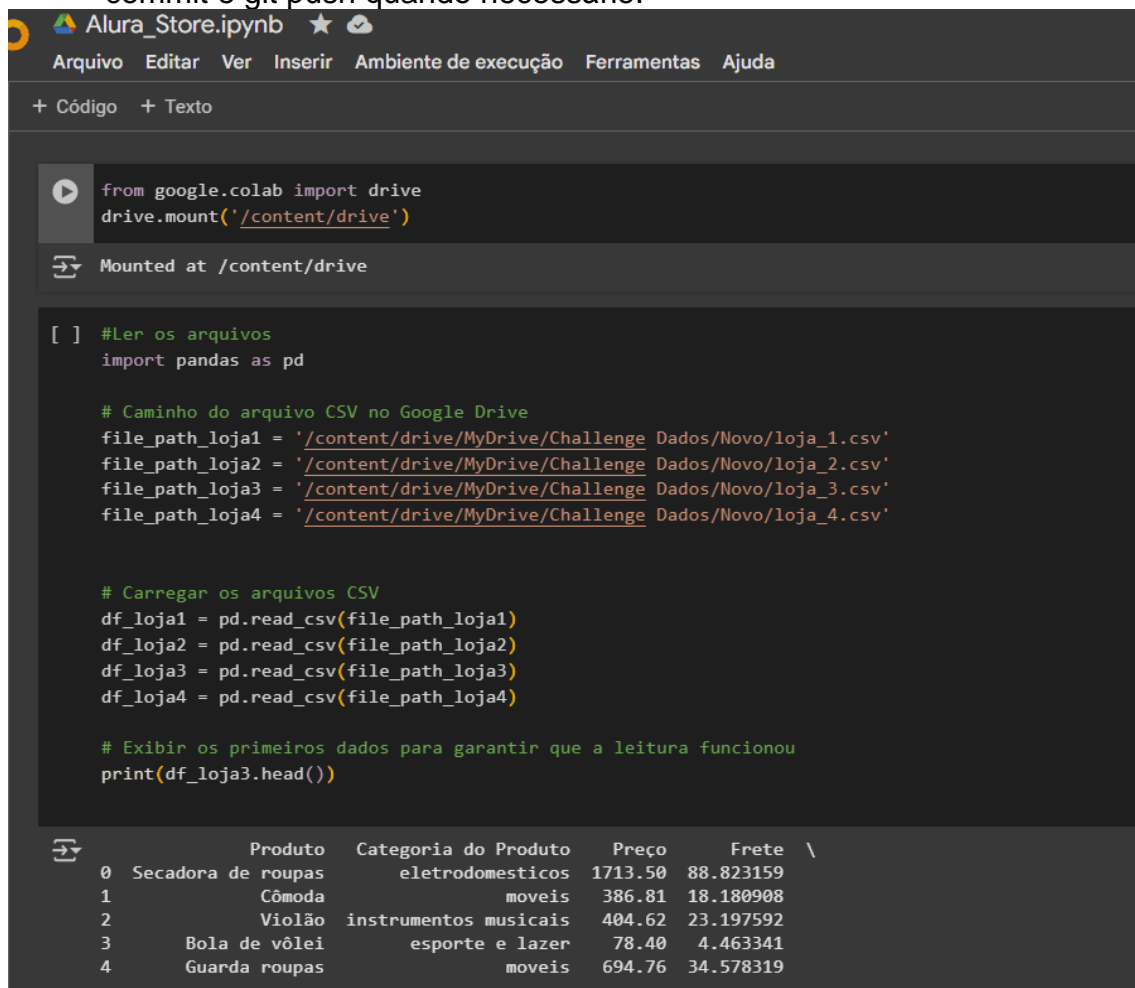
## Suba seu projeto no GitHub

Git e GitHub são ferramentas essenciais para qualquer analista de dados, permitindo versionamento e compartilhamento eficiente dos projetos.

Neste desafio, você deverá **subir seu caderno do Colab para um repositório no GitHub**. Isso garantirá que seu progresso esteja salvo e acessível de qualquer lugar.

### O que você precisa fazer:

1. Crie um **novo repositório no GitHub** para armazenar seu projeto.
2. Exporte seu caderno do Colab como um arquivo .ipynb.
3. Faça o **upload** do caderno para o repositório.
4. Mantenha seu trabalho atualizado utilizando `git pull`, `git add`, `git commit` e `git push` quando necessário.



```
from google.colab import drive
drive.mount('/content/drive')

# Ler os arquivos
import pandas as pd

# Caminho do arquivo CSV no Google Drive
file_path_loja1 = '/content/drive/MyDrive/Challenge Dados/Novo/loja_1.csv'
file_path_loja2 = '/content/drive/MyDrive/Challenge Dados/Novo/loja_2.csv'
file_path_loja3 = '/content/drive/MyDrive/Challenge Dados/Novo/loja_3.csv'
file_path_loja4 = '/content/drive/MyDrive/Challenge Dados/Novo/loja_4.csv'

# Carregar os arquivos CSV
df_loja1 = pd.read_csv(file_path_loja1)
df_loja2 = pd.read_csv(file_path_loja2)
df_loja3 = pd.read_csv(file_path_loja3)
df_loja4 = pd.read_csv(file_path_loja4)

# Exibir os primeiros dados para garantir que a leitura funcionou
print(df_loja3.head())
```

	Produto	Categoria do Produto	Preço	Frete	\
0	Secadora de roupas	eletrodomesticos	1713.50	88.823159	
1	Cômoda	moveis	386.81	18.180908	
2	Violão	instrumentos musicais	404.62	23.197592	
3	Bola de vôlei	esporte e lazer	78.40	4.463341	
4	Guarda roupas	moveis	694.76	34.578319	

**Caso precise relembrar conceitos, confira os recursos abaixo:**

- [Começando com Git: Aprendendo a versionar](#)
- [GitHub: diferentes maneiras de compartilhar seu projeto](#)
- [Como escrever um README incrível no seu GitHub](#)

O **README** é essencial para documentar seu projeto, explicando sua finalidade, funcionalidades e instruções de uso.

Como **desafio adicional**, crie um README.md para seu projeto **Telecom X**, incluindo:

1. O propósito da análise realizada.
2. Estrutura do projeto e organização dos arquivos.
3. Exemplos de gráficos e insights obtidos.
4. Instruções para executar o notebook.

Isso tornará seu projeto mais profissional e acessível para qualquer pessoa que queira entender suas análises.

## Entrega do Challenge

Parabéns por completar seu challenge, temos certeza de que você desenvolveu um projeto incrível!

### IMPORTANTE!

- Verifique a URL do seu projeto antes de enviar.
- O sistema ACEITA APENAS URL do GitHub.
- **Você deve enviar apenas o link do repositório GitHub do seu projeto, de preferência o repositório sem deploy.**
- **Após adicionar a URL do seu projeto, baixe o badge que aparecerá ao clicar no botão de envio e, em seguida, envie seu projeto.**
- Você tem 5 tentativas para entregar seu projeto.

**Após a entrega do desafio, VOCÊ DEVE CONCLUIR O CURSO E GERAR SEU CERTIFICADO!**

Baixe sua badge, compartilhe no LinkedIn e em todas as suas redes sociais marcando #AluraLatam e #oraclenexteducation!

Não se esqueça de dar uma olhada neste vídeo que preparamos para você com o passo a passo descrito acima.