



**INSTITUTO
FEDERAL**
Norte de Minas Gerais

Introdução a Sistemas Inteligentes

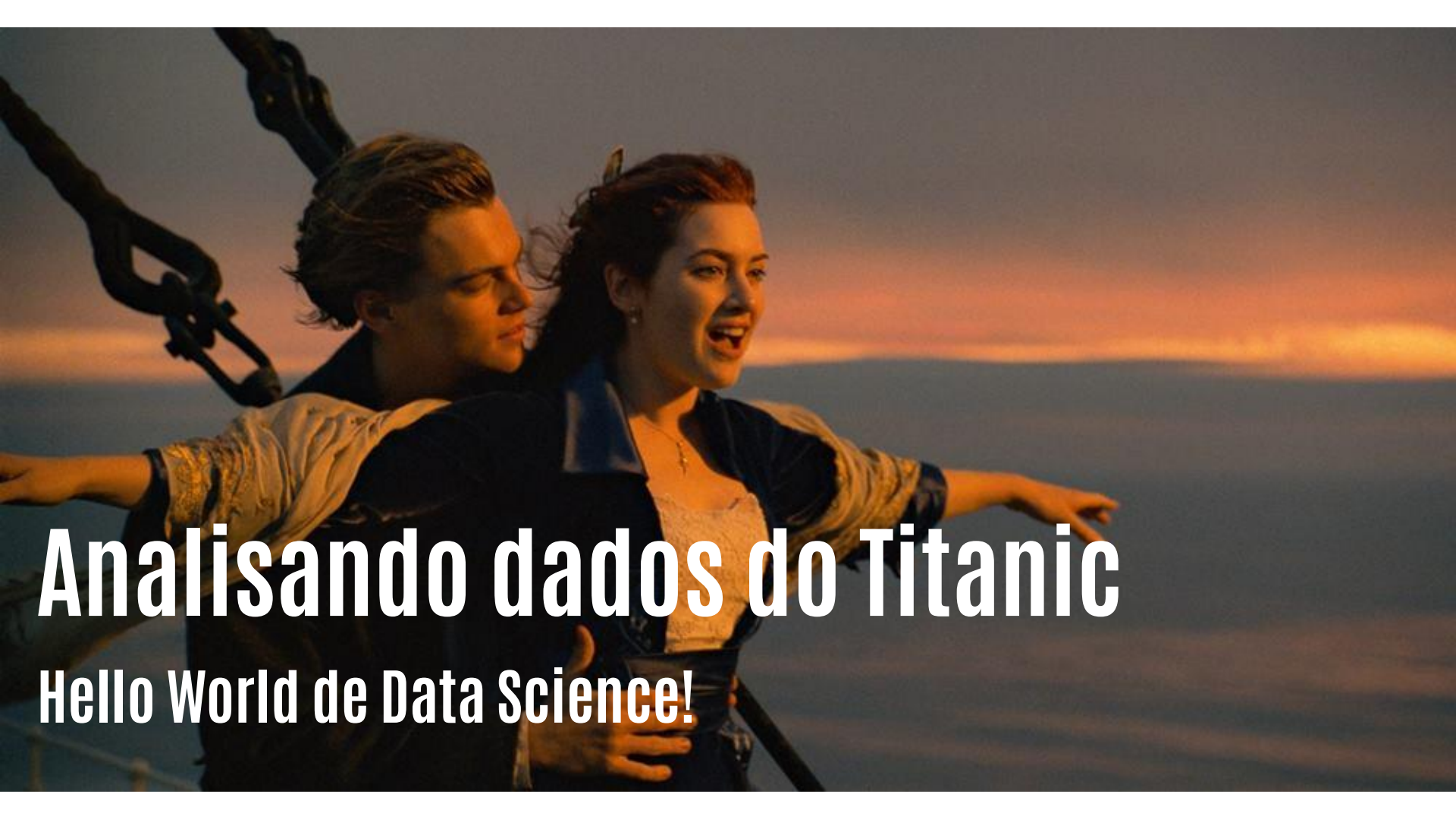
Análise Exploratória

Prof^a. Suzana Mota



Análise Exploratória de Dados





Analizando dados do Titanic

Hello World de Data Science!

Dataset

Conjunto de Dados

É uma coleção organizada de dados.

Geralmente possui **features** e **labels**.

Label: É o rótulo!

A “resposta” que vamos tentar prever

Features: São as características que temos disponíveis no dataset. É dentro delas, que vamos procurar padrões!

Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

Tipos de Dados

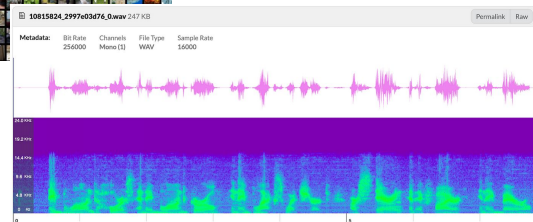
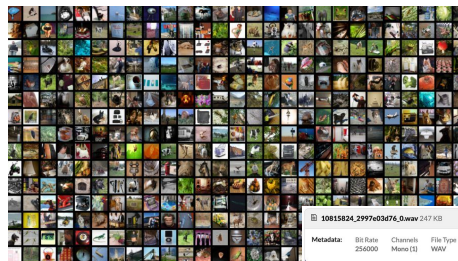
Dados Estruturados: Dados estruturados são organizados em um formato definido e previsível, geralmente em tabelas, com linhas e colunas. Cada coluna contém um tipo específico de dado, como números, datas, ou strings, e cada linha representa um registro individual.

Ex: Dados em CSV, em tabelas, em Banco de dados.

Nome	CPF	Endereço	Telefone
Marcela Freitas	11111	Rua A, nº 1	101010
João Augusto	22222	Rua B, nº 2	202020
Pablo Silva	33333	Rua C, nº 3	303030
André Mendes	44444	Rua D, nº 4	404040
Juliana Freitas	55555	Rua E, nº 5	505050

Dados não Estruturados: Dados não estruturados não têm um formato específico ou organização fixa. Eles podem ser de natureza variada e geralmente não seguem um esquema predefinido.

Ex: Imagens, Vídeos, Áudio, Mensagens de redes sociais



Google Colab

É um ambiente online de notebook interativo.
Permite que você misture código e texto.

<https://colab.research.google.com/>

E assim é possível ter um “*paper executável*”, um artigo que indica o passo a passo de uma pesquisa e também traz seu código fonte, seus dados e tudo que for necessário, para que a pesquisa seja Reprodutível.

Exemplo:

https://github.com/suzanasvm/SmileFace/blob/master/deliver/JoyFace_ExecutablePaper.ipynb

JoyFace

The JoyFace HMI considers the displacement of the user's face about a reference region. The face is identified by a regular webcam and verify the face positions. Each position is associated with a movement control of the wheelchair.

JoyFace was implemented in Python language and uses face detection based on the Viola-Jones classifiers incorporated into the OpenCV library (Viola and Jones, 2001). These classifiers use Haar Cascade features that are applied to images in real time (Papageorgiou et al., 1998). After detection of the user's face, we observed the last 40 frames. From there the average face position is calculated, and a reference region demarcated. This reference region will remain static when using JoyFace and can be viewed as a white rectangle.

The centroid of the face detection square is calculated in real time and receives a green circle to highlight the displayed image. This way the user can send commands through the displacement of his nose that has the same position of the calculated centroid. Figure 2 shows how JoyFace HMI it works. If the user puts the nose above the reference region, the wheelchair begins to move front, if the nose is positioned to the right or left the wheelchair moves to the corresponding side and the smile interrupts the movement of the wheelchair.

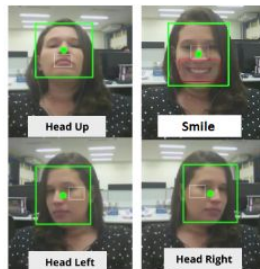


Figure 2: JoyFace facial expressions commands used to control the robotized wheelchair.

Results

This work proposed a solution that will offer to the person with the motor disability, independence, and autonomy of locomotion, in addition to the control of the environment where it is inserted, thus, improving its quality of life and social inclusion. The results obtained can be replicated through the simulation code available in the annexes of this paper.

Although the results achieved are promising, these upsides rely on the feedback screen with the user's face image captured by the notebook webcam. During the navigation, the operator tends to look at the screen to monitor if he is positioning the face landmark (centroid) out of the reference square and in-deed actuating on the wheelchair. The user feels forced to look at the screen during the most of the navigation which is a limitation, as he can not move his head freely without sending undesirable commands to the wheelchair.

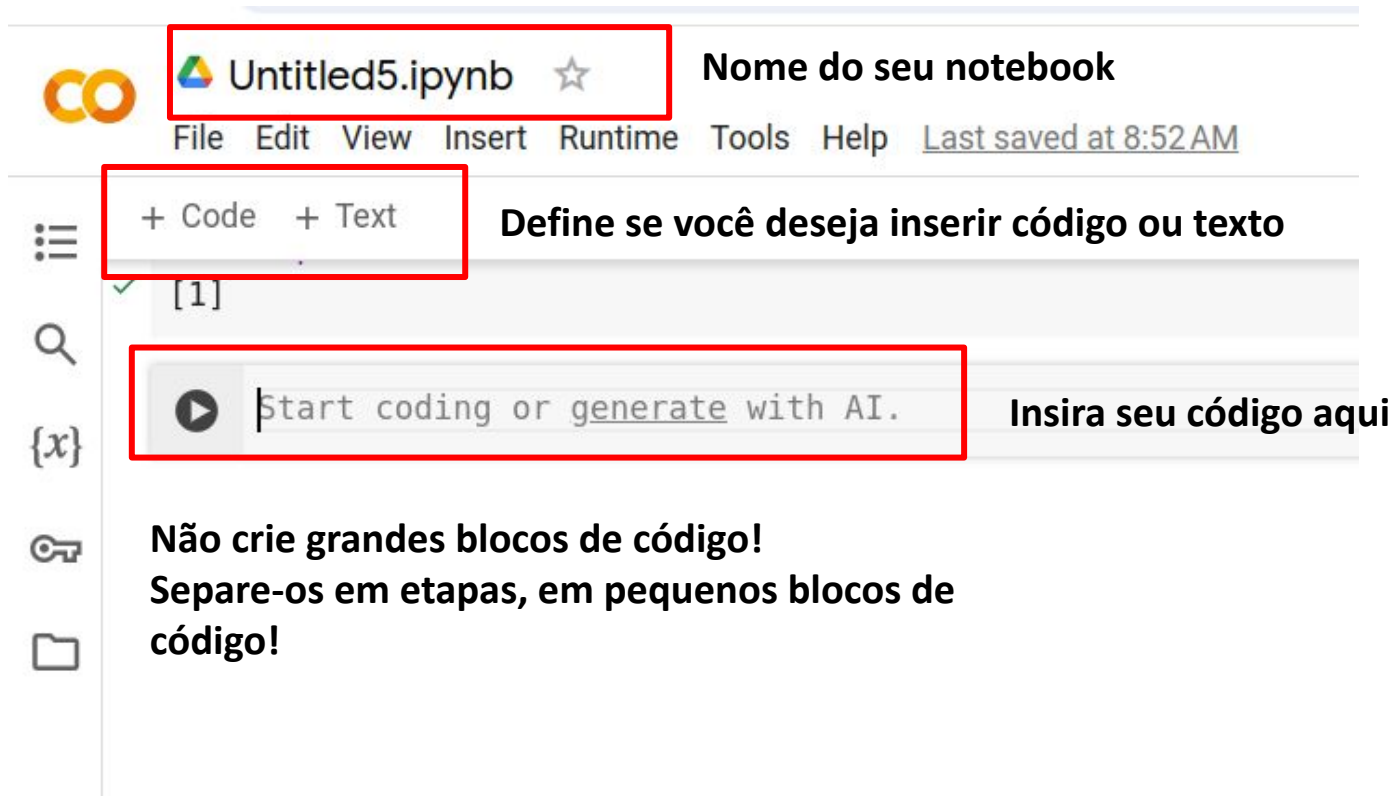
As future work, we hope to associate a new facial expression to turn on and off the wheelchair movement system. In this way, the user will have more freedom to do their activities and choose the moment when they want to move the wheelchair.

Anexes


The code utilized in this works it's available here:

```
81: import time
import cv2
import matplotlib.pyplot as plt
from IPython import display
```

Conhecendo o Colab



The screenshot shows the Google Colab interface. At the top, the Colab logo is on the left, followed by a red box containing the notebook name 'Untitled5.ipynb' and a star icon. To the right of this box is the text 'Nome do seu notebook'. Below the name is a menu bar with 'File', 'Edit', 'View', 'Insert', 'Runtime', 'Tools', and 'Help', followed by the text 'Last saved at 8:52 AM'. On the left side, there is a sidebar with icons for a menu, search, variables, and a folder. The main area has a red box around the '+ Code' and '+ Text' buttons, with the text 'Define se você deseja inserir código ou texto' to their right. Below this, another red box highlights a play button icon and the text 'Start coding or generate with AI.', with the text 'Insira seu código aqui' to its right. At the bottom, there is a key icon and a folder icon, followed by the text: 'Não crie grandes blocos de código! Separe-os em etapas, em pequenos blocos de código!'.

 **Untitled5.ipynb** ☆ Nome do seu notebook

File Edit View Insert Runtime Tools Help [Last saved at 8:52 AM](#)

⋮ + Code + Text Define se você deseja inserir código ou texto

✓ [1]

🔍 {x}

▶ Start coding or generate with AI. Insira seu código aqui

🔑 Não crie grandes blocos de código!
Separe-os em etapas, em pequenos blocos de código!

📁

Conhecendo o Colab

Open notebook

Examples >

Recent >

Google Drive >

GitHub >

Upload >

Search notebooks



Title

Last opened ↓

First opened ↑↓



[Untitled5.ipynb](#)

8:53 AM

8:52 AM



[Algoritmos de ordenação.ipynb](#)

8:52 AM

March 12



[SI1_AnaliseExploratoria](#)

September 1

September 1



[Welcome To Colab](#)

September 1

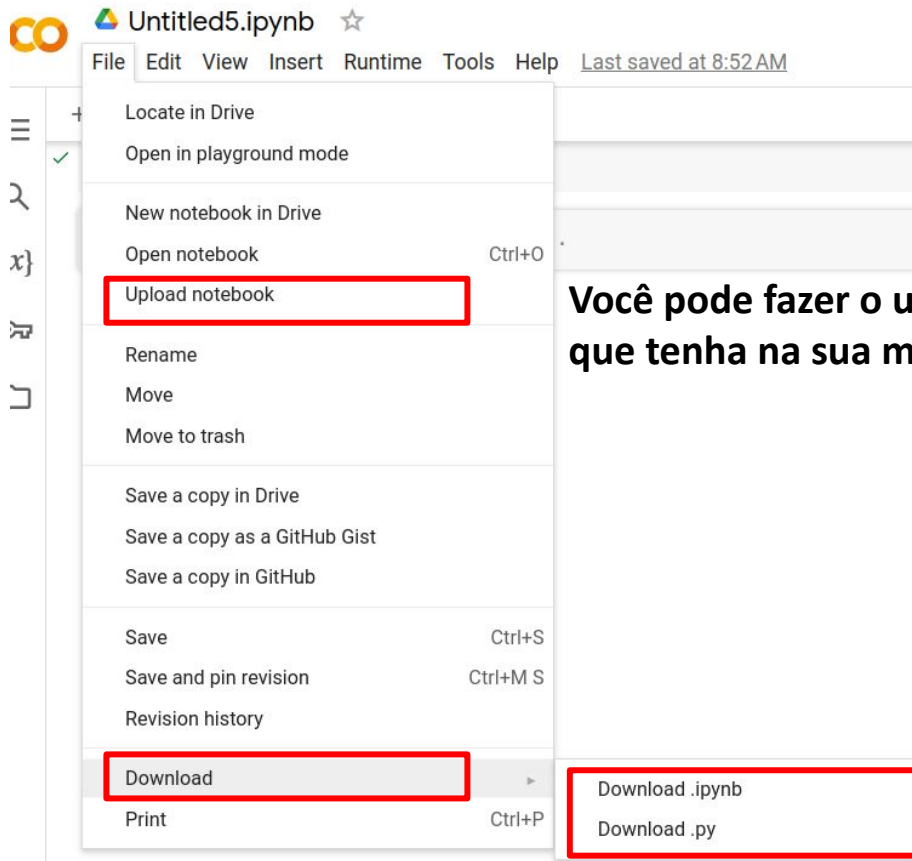
Jan 24, 2020



+ New notebook

Cancel

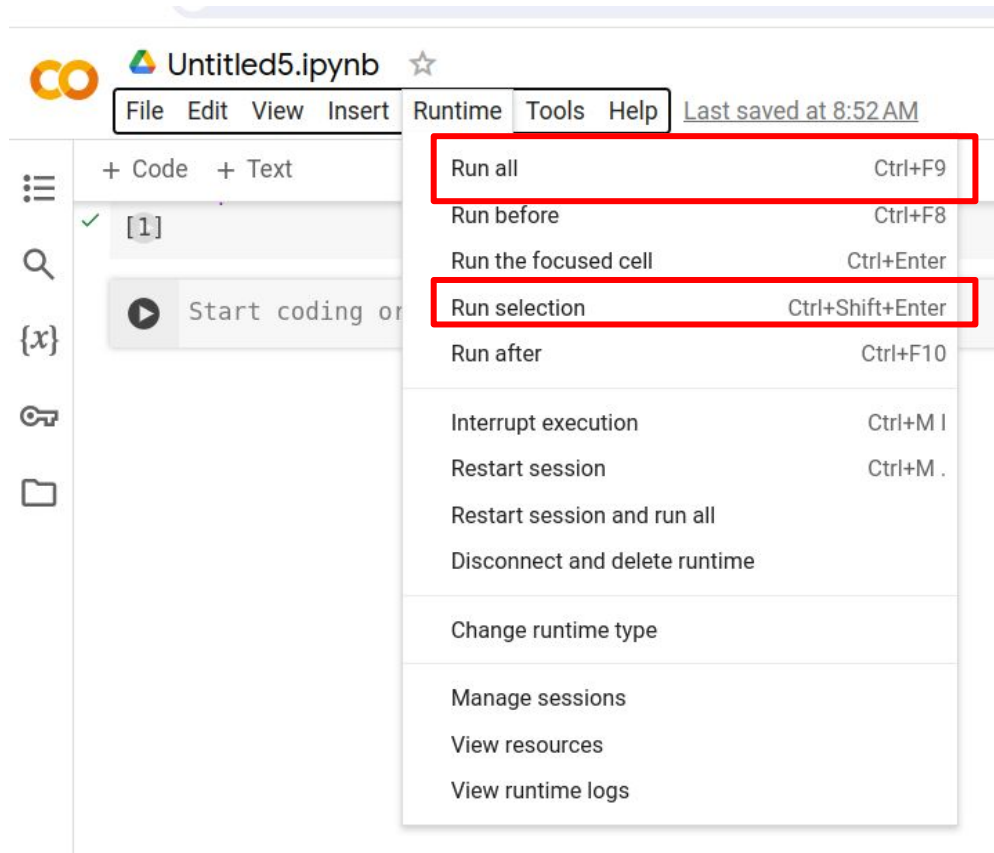
Conhecendo o Colab



Você pode fazer o upload de algum notebook que tenha na sua máquina localmente

Você pode fazer o download do seu código, para rodá-lo na sua máquina localmente

Conhecendo o Colab



Roda TODAS as células do notebook de uma vez! Atalho: Ctrl + F9

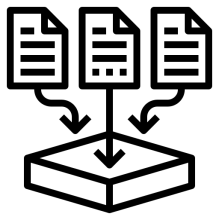
**Roda a célula selecionada
Atalho: Shift + Enter**

**SEMPRE QUE POSSÍVEL, USE ATALHOS DO
TECLADO!
POIS TE TORNAM MUITO MAIS PRODUTIVO**

Passos para análise



**Configuração de
Ambiente**



**Exploração e
Entendimento
dos Dados**



**Limpeza e
Preparação dos
Dados**



**Análise
Exploratória
dos Dados**



**Conclusões &
Insights**

Configuração de Ambiente

1. Crie um novo notebook no Google Colab
2. Faça os imports necessários:
 - Pandas: Biblioteca capaz de ler os dados e manipulá-los para obtermos insights
 - Matplotlib: Biblioteca capaz de criar visualizações gráfica
 - Seaborn: Biblioteca que cria visualizações mais modernas

import nome_biblioteca as xy:

A palavra import, importa a lib

e o **as**, cria um “apelido” para a biblioteca

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Instalando uma biblioteca

!pip install pandas

Documentações Bibliotecas Úteis

pandas: <https://pandas.pydata.org/>

polars: <https://pola.rs/>

matplotlib: <https://matplotlib.org/>

seaborn: <https://seaborn.pydata.org/>

plotly: <https://plotly.com/python/>

Exploração e Entendimento dos Dados

Dados geralmente estão disponíveis em CSV:

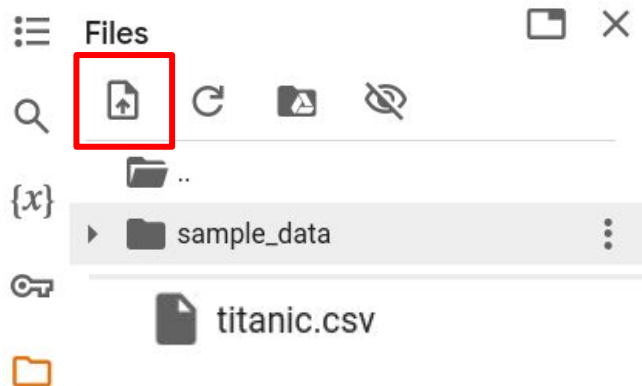
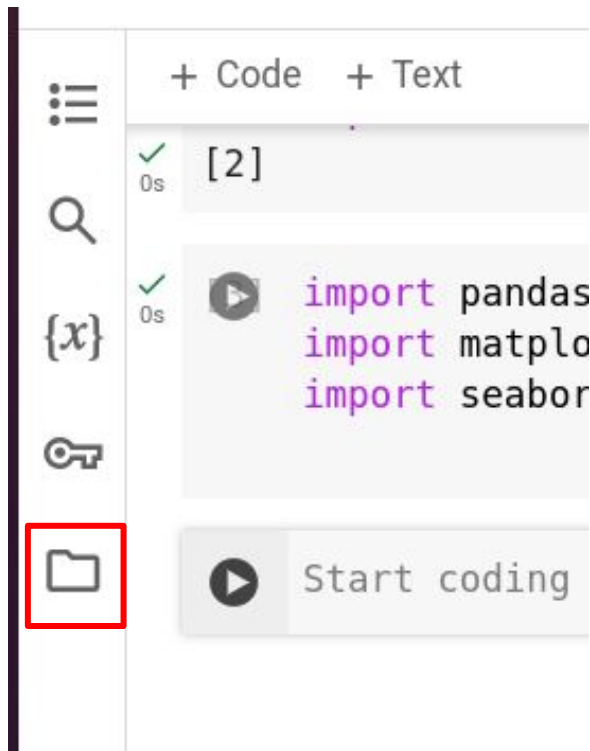
- Comma-separated values: Valores separados por vírgula
- Um arquivo que se parece com uma planilha do excel

Obtenha os dados do titanic.csv, aqui:

https://drive.google.com/drive/folders/12i9yBm0EOz6br_M-BYCx8gxUgwgxc4EJ?usp=sharing

Inserindo dados - Direto no notebook

Você pode fazer o upload dos dados no notebook.
Mas se a sessão for encerrada, você perde os dados!



Implemente uma linha como esta, no seu notebook

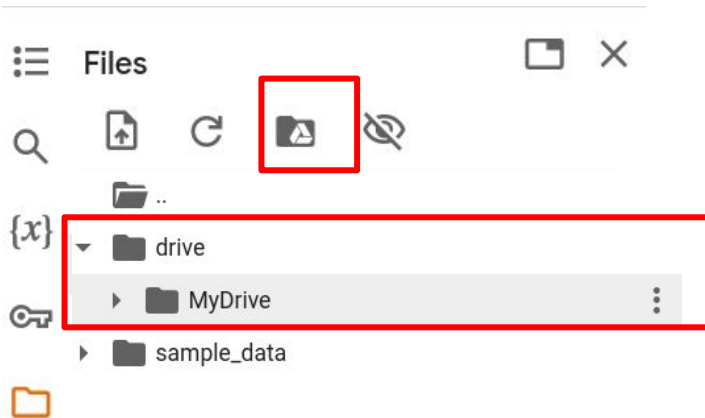
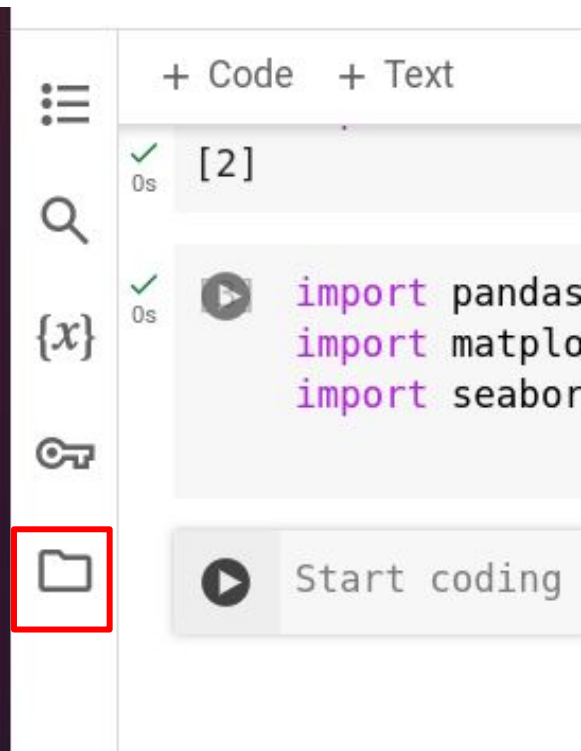
```
[7] file = 'titanic.csv'
     df = pd.read_csv(file)
```


Inserindo dados - Lendo do Google Drive

Você pode logar ao seu Google Drive!

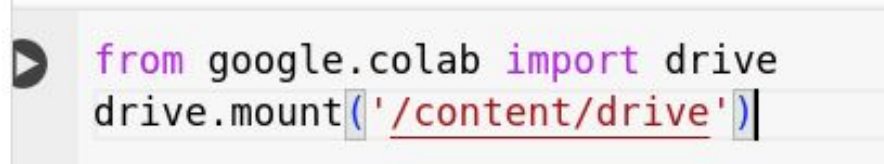
E ele sempre conseguirá visualizar seus dados

É possível apenas, que ele pensa a confirmação da autenticação

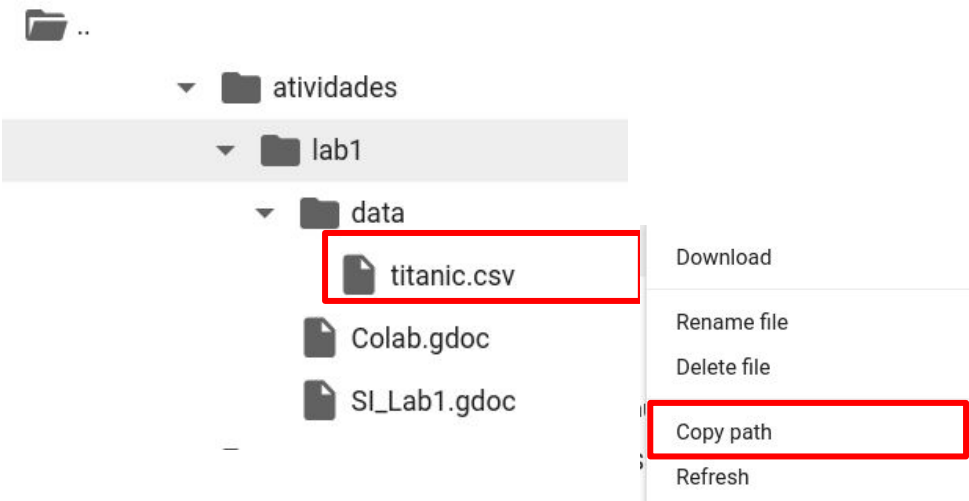


O Google drive precisa ser montado!

Adicione essas linhas ao notebook



Inserindo dados - Lendo do Google Drive



Copie o caminho do seu arquivo, no seu Google Drive

Crie uma linha como esta no seu notebook



```
file = '/content/drive/MyDrive/IFNMG/02 2024/Sistemas_Inteligentes/atividades/lab1/data/titanic.csv'
df = pd.read_csv(file)
```

Limpeza e Preparação dos Dados

Conheça os seus dados!

- Quantos dados você tem?
- Quais são as colunas?
- Todas estão preenchidas?
- Os dados estão padronizados?
- Os dados são balanceados?
- **O que fazer com os valores vazios?** Existem várias técnicas e a escolhida, vai depender do tipo de dado, do tipo de modelo, da situação, do contexto

Limpeza e Preparação dos Dados

Conheça os seus dados!

- Quantos dados você tem?

```
df.shape
```

- Quais são as colunas?

```
df.columns
```

- Todas estão preenchidas?

```
df.isnull().sum()
```

- Os dados estão padronizados?

```
df['Cabin'].unique()
```

- Os dados são balanceados?

```
df['Survived'].value_counts()
```

```
df['Survived'].value_counts(normalize=True)
```

Limpeza e Preparação dos Dados

- O que fazer com os valores vazios?

```
df['Age'].fillna(df['Age'].median(), inplace=True)
```

Preenchendo os dados ausentes com a mediana

A mediana é o valor que separa a metade superior e a metade inferior dos dados

`inplace=True`: Significa que vamos implementar a mudança em todo o dataset, sem ter que reatribuir os dados atualizados, ou seja, já vai rodando e atualizando

Limpeza e Preparação dos Dados

Para ver todos os dados

```
| df
```

Para ver os 5 primeiros dados

```
df.head()
```

Para ver os 100 primeiros dados

```
) df.head(100)
```

Análise Exploratória dos Dados



Seja curioso e faça perguntas:

- Qual foi a porcentagem dos passageiros sobreviventes?
- Qual era a faixa etária dos passageiros que estavam no Titanic?
- Houveram mais crianças ou mais adultos que sobreviveram?
- O gênero influenciou na sobrevivência?

Análise Exploratória dos Dados

Seja curioso e faça perguntas:

- Qual foi a porcentagem dos passageiros sobreviventes?

```
df['Survived'].value_counts(normalize=True) * 100
```

proportion

Survived

0	61.616162
1	38.383838

Análise Exploratória dos Dados

Seja curioso e faça perguntas:

- Qual foi a porcentagem dos passageiros sobreviventes?

```
contagem_sobreviventes = df['Survived'].value_counts()  
categorias = ['Não Sobreviventes', 'Sobreviventes']  
percentuais = contagem_sobreviventes.sort_index().values
```

Survived	
0	61.616162
1	38.383838
..	..

```
contagem_sobreviventes.sort_index().values  
  
array([61.61616162, 38.38383838])
```

Análise Exploratória dos Dados

Seja curioso e faça perguntas:

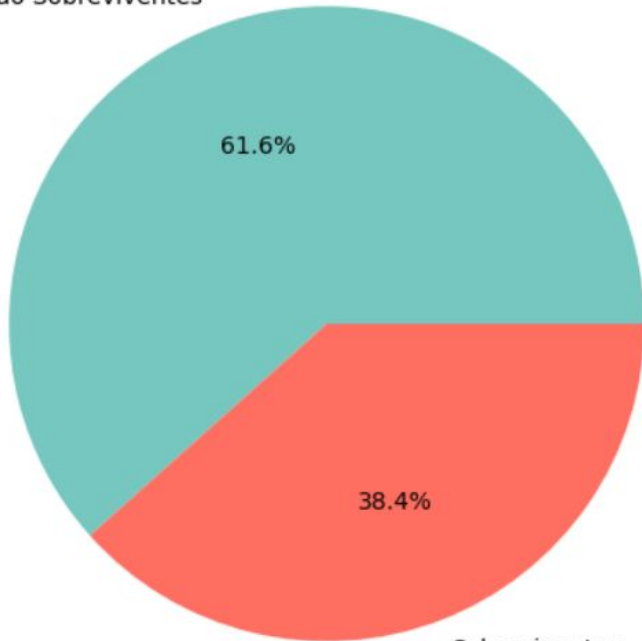
- Qual foi a porcentagem dos passageiros sobreviventes?

```
# Cria o gráfico de pizza
plt.figure(figsize=(8, 6))
plt.pie(percentuais, labels=categorias, autopct='%1.1f%%', colors=['#76c7c0', '#ff6f61'])
plt.title('Porcentagem de Passageiros Sobreviventes no Titanic')
plt.show()
```

Análise Exploratória dos Dados

Porcentagem de Passageiros Sobreviventes no Titanic

Não Sobreviventes



Sobreviventes

Análise Exploratória dos Dados



Seja curioso e faça perguntas:

- ~~Qual foi a porcentagem dos passageiros sobreviventes?~~
- Qual era a faixa etária dos passageiros que estavam no Titanic?
- Houveram mais crianças ou mais adultos que sobreviveram?
- O gênero influenciou na sobrevivência?
- E deixe a curiosidade fluir, para fazer outras perguntas!

Conclusões e Insights

- Quais conclusões você obteve após a análise?
- O que mais influenciou na taxa de sobrevivência?
- O que não teve nenhuma influência na taxa de sobrevivência?

