

Manual de Instruções

Luis Miguel Frazão de Sousa

31 de Julho de 2025

1 Instruções de uso

Para executar o projeto, é necessário usar o Google Collab no ambiente de execução usando o **Python**

3. Além disso é necessário instalar as seguintes bibliotecas:

Listing 1: Bibliotecas

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import chi2_contingency, pearsonr
from math import sqrt
```

E usar algumas de suas funções:

```
pd.read_csv() – Leitura do arquivo CSV como DataFrame
df.head() – Visualizacao inicial das primeiras linhas
df[['col1', 'col2']].corr() – Correlacao de Pearson entre variaveis numericas
pd.crosstab(df['col1'], df['col2']) – Tabela de contingencia entre variaveis categoricas
df['col'].value_counts(normalize=True) – Proporcoes relativas de categorias
df[df['col'] == valor] – Filtro de linhas por condicao logica
chi2_contingency(tabela) – Teste Qui-Quadrado
plt.figure(figsize=(largura, altura)) – Define o tamanho do grafico
plt.title('Titulo do grafico') – Adiciona titulo
plt.xticks(rotation=45) – Rotaciona os rotulos do eixo X
plt.show() – Exibe o grafico
sns.boxplot(x='grupo', y='nota', data=df) – Boxplot de comparacao entre grupos
```

2 Funções usadas e suas ações

Análise	Objetivo	Funções/Métodos Principais
1. Medidas Resumo	Estatísticas descritivas da nota de matemática.	mean(), median(), mode(), max(), min(), var(), std()
2. Boxplots Comparativos	Comparar notas de matemática, leitura e escrita com boxplots.	seaborn.boxplot()
3. Assimetria e Curtose	Medir assimetria e curtose das notas.	scipy.stats.skew(), scipy.stats.kurtosis()
4. Percentual por Etnia	Distribuição percentual por grupo étnico.	value_counts(normalize=True)
5. Discretização de Notas	Transformar nota de matemática em faixas qualitativas.	.apply(), value_counts(), cumsum()
6. Associação Gênero/Curso	Associação entre gênero e curso preparatório.	pd.crosstab(), chi2_contingency()
7. Força da Associação	Medir força da associação com Phi e V de Cramer.	math.sqrt()
8. Correlação entre Notas	Correlação linear entre notas de matemática e leitura.	.corr(method='pearson')
9. Nota vs. Escolaridade dos Pais	Comparar notas de matemática por escolaridade dos pais.	seaborn.boxplot(), plt.xticks()
10. Nota vs. Curso Preparatório	Comparar notas de escrita com/sem curso preparatório.	seaborn.boxplot()

Figura 1: Funções usadas

3 Passo a passo de instalação e importação

3.1 Primeiro acesse o link a seguir:

<https://www.kaggle.com/datasets/spscientist/students-performance-in-exams>

The screenshot shows the Kaggle website interface. On the left is a sidebar with navigation links: Home, Competitions, Datasets (highlighted), Models, Benchmarks, Code, Discussions, Learn, and More. The main content area displays the dataset page for 'Students Performance in Exams' by JAKKI SESHAPANPU, updated 7 years ago. The page has 4724 votes and a 'Code' button. Below the title, it says 'Marks secured by the students in various subjects'. There are tabs for 'Data Card' (selected), 'Code (1450)', 'Discussion (33)', and 'Suggestions (1)'. The 'About Dataset' section includes: Context (Marks secured by the students), Content (This data set consists of the marks secured by the students in various subjects.), Acknowledgements (http://roycekimmons.com/tools/generated_data/exams), and Inspiration (To understand the influence of the parents background, test preparation etc on students performance). On the right, there are metadata fields: Usability (7.06), License (Unknown), Expected update frequency (Not specified), and Tags (Education, Standardized Testing, Data Visualization).

Figura 2: Página inicial

The screenshot shows the Kaggle dataset page for "Students Performance in Exams" by JAKKI SESHAPANPU. The left sidebar contains navigation links: Home, Competitions, Datasets (selected), Models, Benchmarks, Code, Discussions, Learn, and More. The main content area includes a search bar, a "Sign In" button, and a "Register" button. The dataset title "Students Performance in Exams" is prominently displayed, followed by the subtitle "Marks secured by the students in various subjects". Below this, there are tabs for "Data Card" (selected), "Code (1450)", "Discussion (33)", and "Suggestions (1)". The "About Dataset" section is visible, containing "Context", "Content", "Acknowledgements", and "Inspiration" subsections. On the right side, a "Download" button is highlighted, and a dropdown menu is open, showing options: "Download dataset as zip (9 kB)" and "Export metadata as Croissant". The "Tags" section on the right lists "Education", "Standardized Testing", and "Data Visualization".

Figura 3: Aba de download aberta

This screenshot is identical to the one above, showing the Kaggle dataset page for "Students Performance in Exams". The "Download" button is highlighted, and the dropdown menu is open, showing the option "Download dataset as zip (9 kB)" which is now highlighted with a blue border. The "Export metadata as Croissant" option is also visible. The "Tags" section on the right lists "Education", "Standardized Testing", and "Data Visualization".

Figura 4: Clique na opção de download marcada

The screenshot shows a file explorer window with a dark background. The top bar displays the "archive" logo, the date and time "28/07/2025 15:41", the name of the folder "Pasta compactada", and the size "9 KB". The main area shows a single folder icon representing the compacted folder.

Figura 5: Acesse a pasta compactada

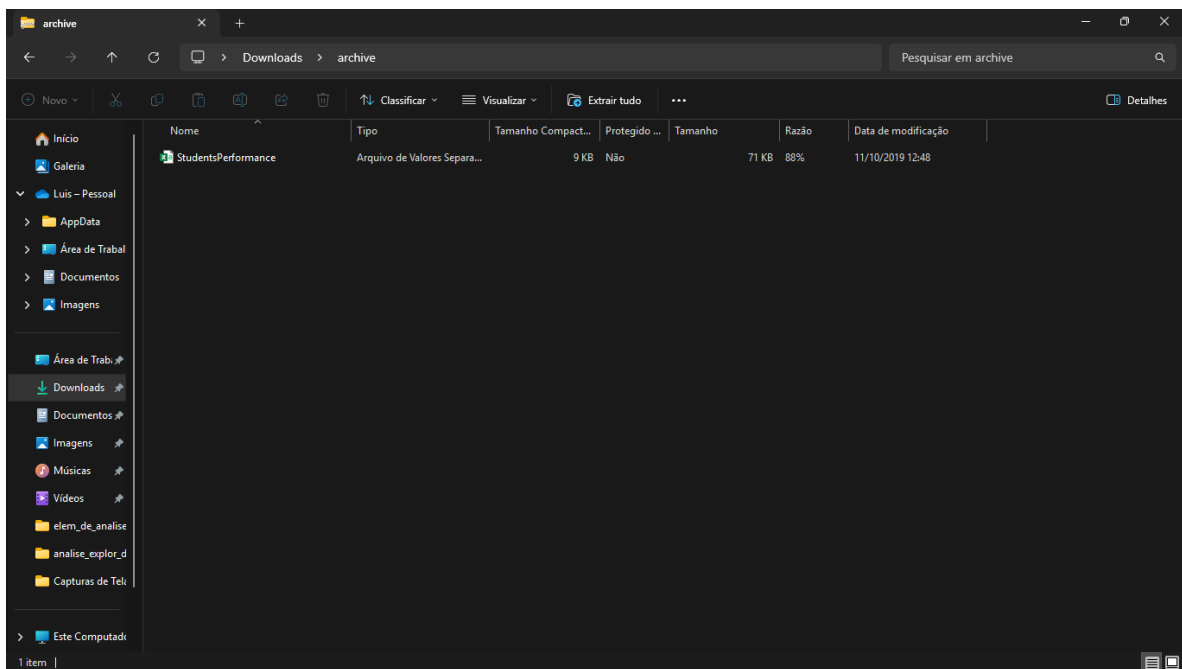


Figura 6: Selecione o arquivo do conjunto de dados

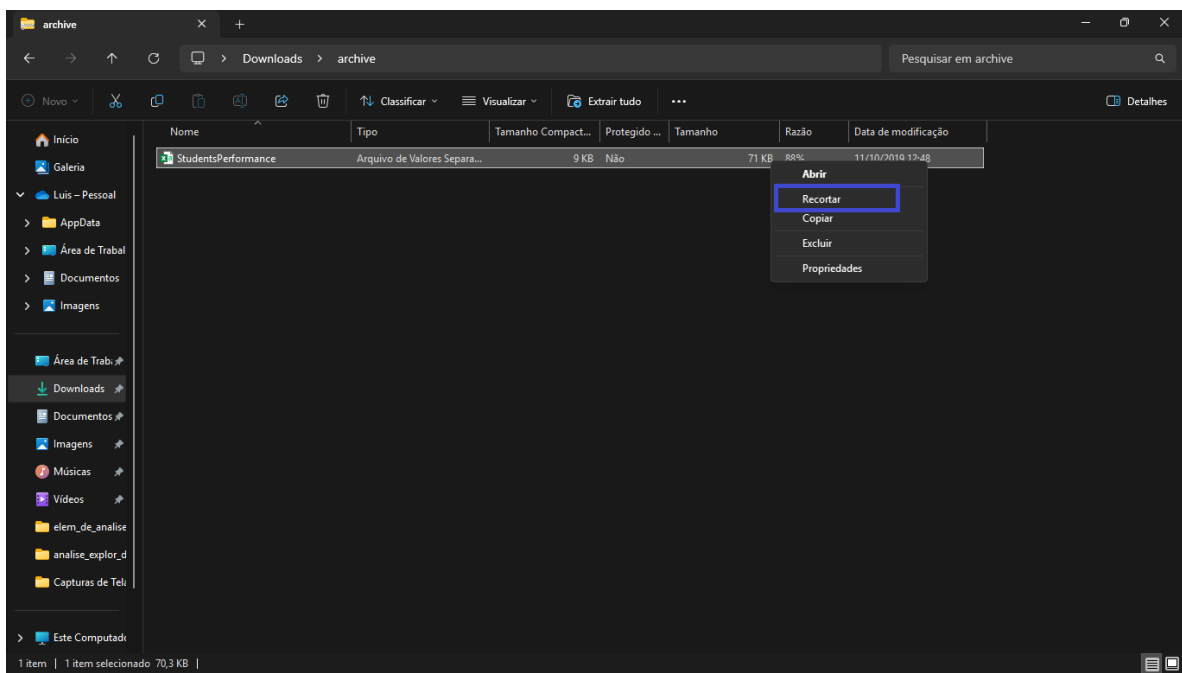


Figura 7: Descompacte e salve o arquivo .csv

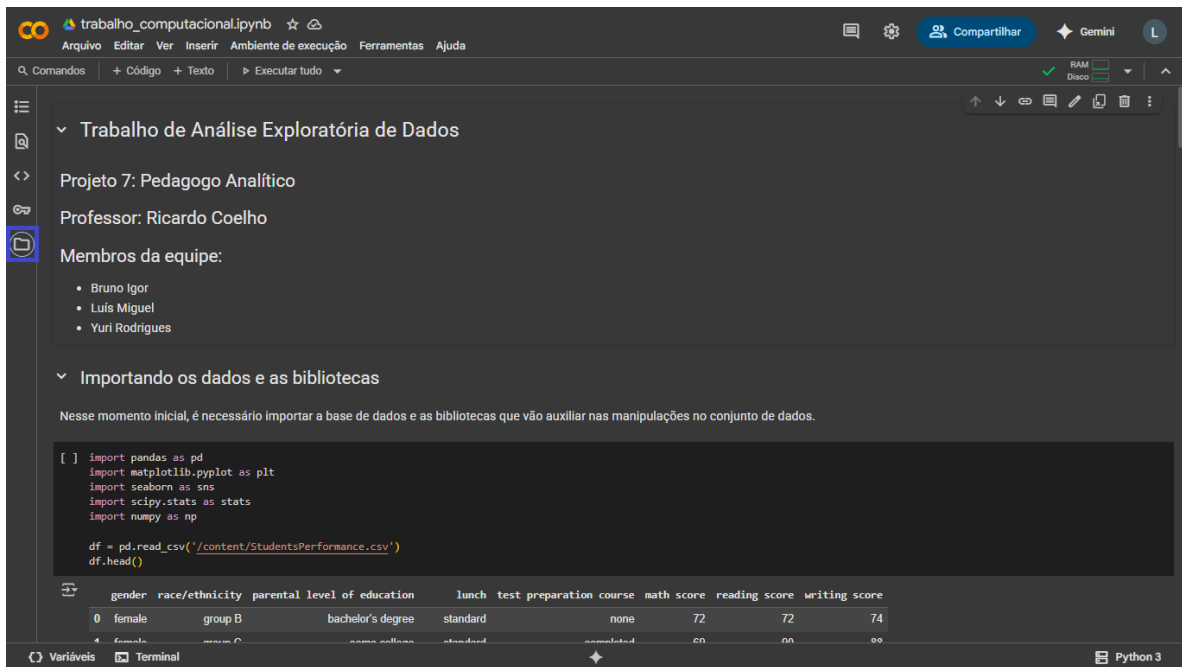


Figura 8: Selecione o ícone para abrir a aba

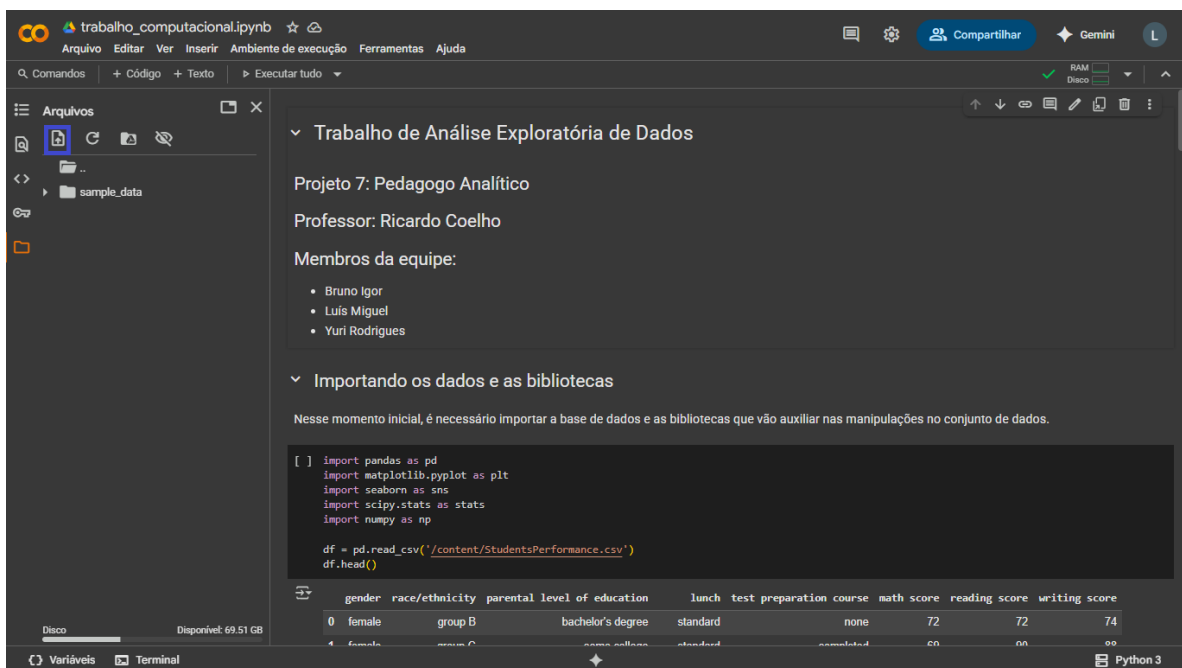


Figura 9: Importe o arquivo