

Data Analysis

Julius

**Exploratory Data Analysis
(EDA)
Julius**



Edgar Rios Linares

Problem

▼ Problem description

Context

The original dataset contains 1000 records with 20 categorical attributes prepared by Prof. Hofmann.

In this dataset, each record represents a person taking out a loan from a bank.

Each person is classified as **good** or **bad** credit risk based on the set of attributes.

Link to the original dataset at

[UCI Machine Learning](#)

Content

The selected attributes:

- **Age** (numeric)
- **Sex** (text: male, female)
- **Job** (numeric: 0 - unskilled and non-resident, 1 - unskilled and resident, 2 - skilled, 3 - highly skilled)
- **Housing** (text: own, rent, or free)
- **Saving accounts** (text - little, moderate, quite rich, rich)
- **Checking account** (numeric, in DM - Deutsch Mark)
- **Credit amount** (numeric, in DM)
- **Duration** (numeric, in month)
- **Purpose**(text: car, furniture/equipment, radio/TV, domestic appliances, repairs, education, business, vacation/others)
- **Risk** (Value target - Good or Bad Risk)

Objective

Train a model to predict from new data whether a person applying for a loan represents a good or bad risk

Julius.Ai

Go website Julius, create account
<https://julius.ai>

ases ▾


Julius


DocsCommunitySign UpSign In


Analyze your data with computational AI.

Chat with your files and get expert-level insights in seconds.

Try Julius AI >

 Loved by over 2,000,000 users worldwide






Charts & Graphs


Create sleek looking data visualizations.

[Try now →](#)




Insights

Ask anything to your data, and get answers.




Advanced Analysis

Perform modeling and predictive forecasting.




Problem Solving

Scan-and-solve math, physics, and chemistry.



Reports

Generate polished analyses and summaries.

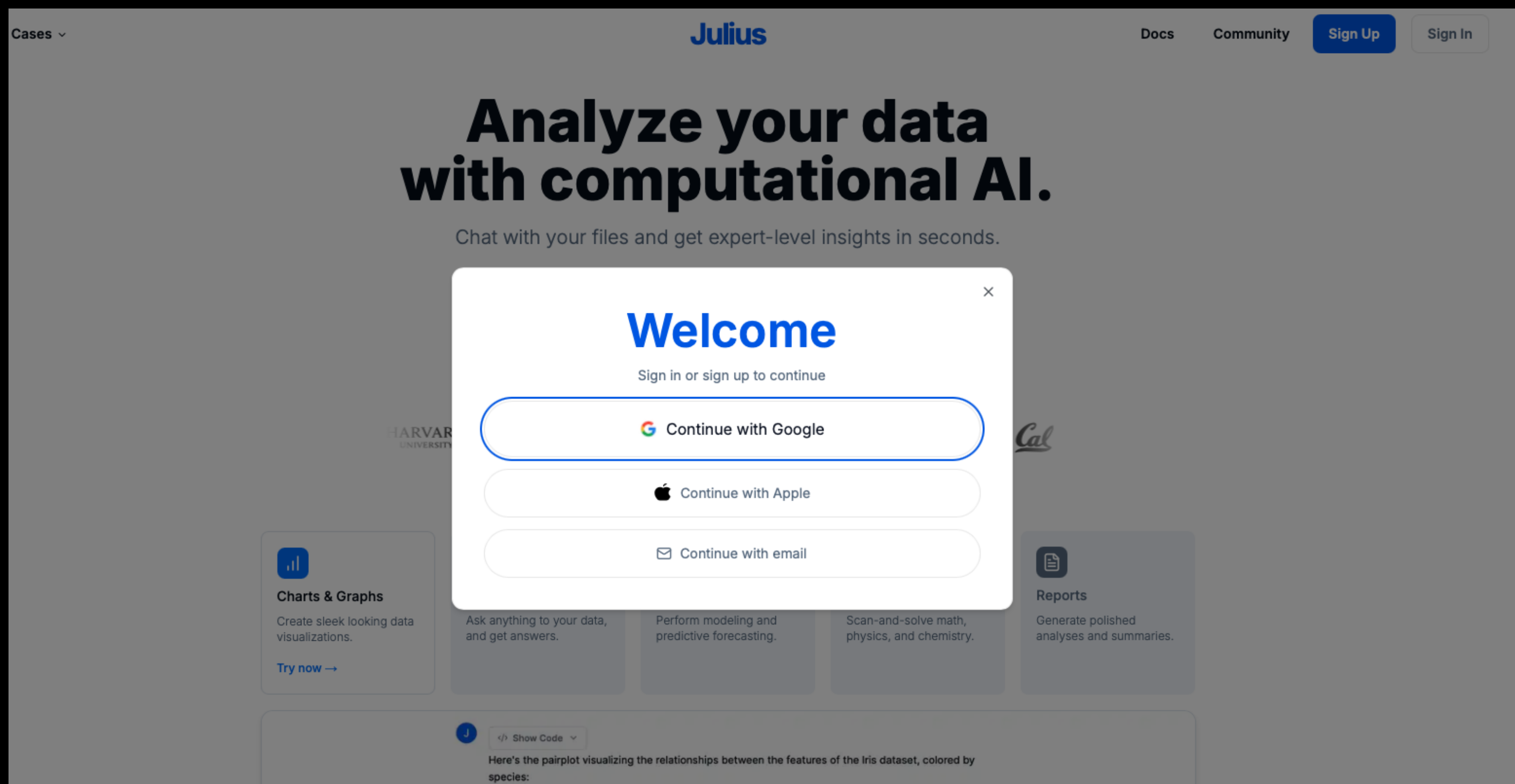


Show Code ▾

Here's the pairplot visualizing the relationships between the features of the Iris dataset, colored by species:

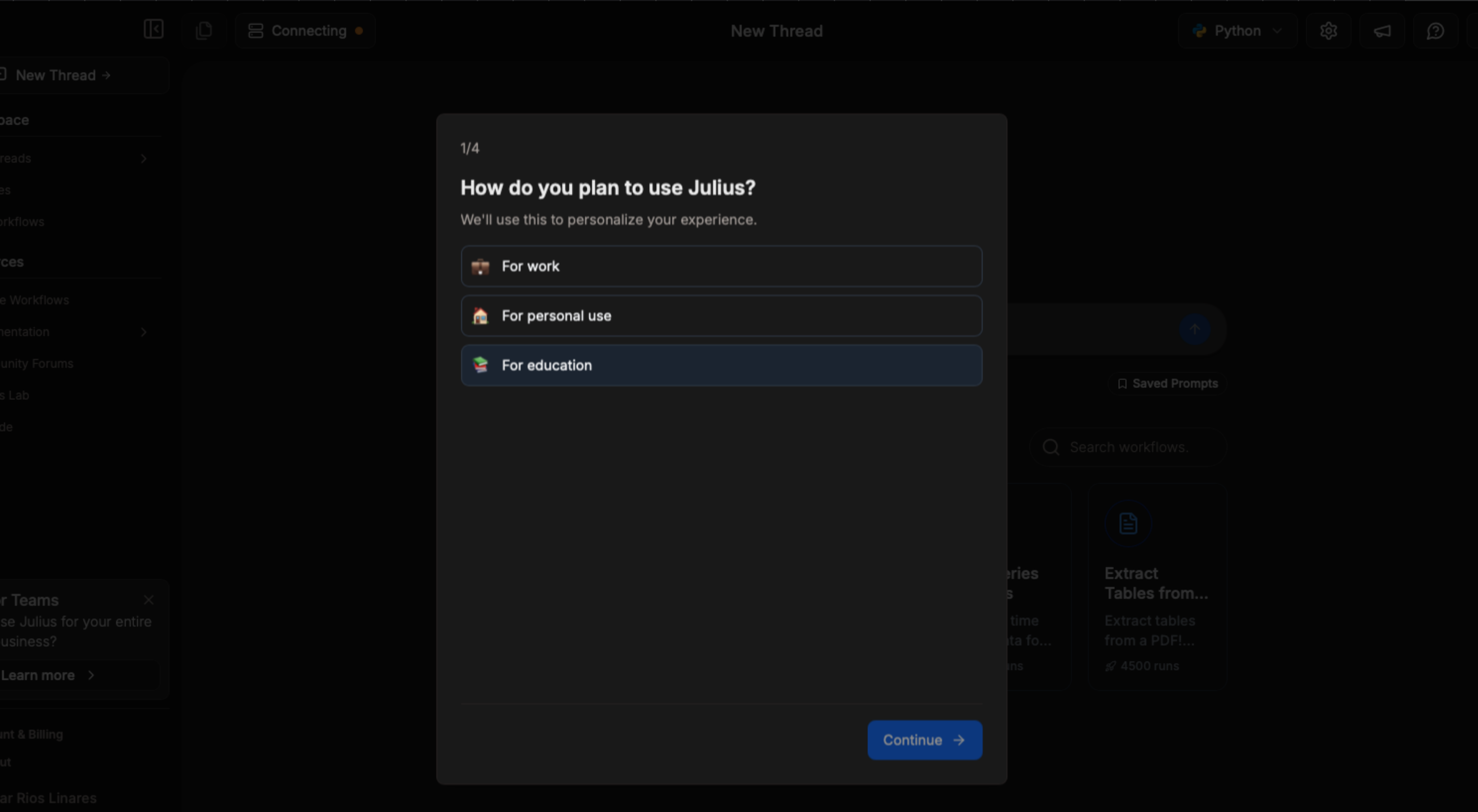
Create Account

Go website Julius, create account
<https://julius.ai>



Create Account

Go website Julius, create account
<https://julius.ai>



New Project

Julius

Connected

New Thread →

Workspace

My Threads

My Files

My Workflows

Resources

Explore Workflows

Documentation

Community Forums

Models Lab

Upgrade

Julius for Teams

Want to use Julius for your entire team or business?

Learn more

Account & Billing

Log Out

Edgar Rios Linares

Google • erlinares@gmail.com

New Thread

Python

What do you want to analyze today?

Add a file or start a conversation now and add files later...

Default

Tools

Advanced Reasoning

Saved Prompts

Or start from ready workflows

Search workflows.

Quick Visualization

Visually explore a spreadshee...

43081 runs

Data Cleaner

Methodically clean your da...

11299 runs

Significance Testing

Run significance...

5668 runs

Time Series Analysis

Forecast time series data fo...

8820 runs

Extract Tables from...

Extract tables from a PDF!...

4500 runs

Load File

Julius

New Thread →

Workspace

My Threads (1)

Search content...

Today

My Files

Inbox

My Workflows

Resources

Explore Workflows

Documentation

Community Forums

Models Lab

Julius for Teams

Want to use Julius for your entire team or business?

Learn more >

Account & Billing

Log Out

My Files

All files that have been uploaded to Julius, and created by Julius.

Drag files to upload

Upload file

Link google sheet

All files are deleted after 1 hour of inactivity.
[Upgrade to increase this →](#)

Search files...

Download (0)

Delete (0)

Chat with files (0)

	Name	Size	Upload Date	
<div>Hey, it looks like you haven't uploaded any files yet. Upload some files to start chatting with them!</div>				

Chat with files

Julius

New Thread →

Workspace

My Threads (1)

Search content...

Today

New Thread

Older

Wine Quality Dataset Overvi...

My Files

Inbox

My Workflows

Resources

Explore Workflows

Documentation

Community Forums

Models Lab

Julius for Teams

Want to use Julius for your entire team or business?

Learn more >

My Files

All files that have been uploaded to Julius, and created by Julius.

Drag files to upload

Upload file

Link google sheet


All files are deleted after 1 hour of inactivity.
[Upgrade to increase this →](#)

Search files...

Download (1)

Delete (1)

Chat with files (1)

	Name	Size	Upload Date	
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>  german_credit_data.csv	52.14 KB	a few seconds ago	<div><div></div><div></div></div>

Prompt

Create an exploratory data analysis (EDA)

Dataset overview: Provides an informative summary of the dataset, including the number of rows and columns, the data types in each column, and the number of non-null values.

Statistical description: Generates a statistical summary of the dataset, including metrics such as mean, median, standard deviation, minimum, maximum, and quartiles for numeric columns.

Correlation matrix: Calculates the correlation matrix for numeric variables in the dataset.

Heatmap: Visualizes the correlation matrix using a heatmap, making sure that the correlation values are readable and the graph is well labeled.

Summary Dataset

Provides an informative summary of the dataset, including the number of rows and columns, the data types in each column, and the number of non-null values

J Julius

Let's start by loading and taking a look at the head of the `german_credit_data.csv` file. One moment, please.



You

Provides an informative summary of the dataset, including the number of rows and columns, the data types in each column, and the number of non-null values



J Julius

</> Loaded and Analyzed German Credit Data

Python

Rerun code

Edit code



```
1 import pandas as pd
2
3 df = pd.read_csv('german_credit_data.csv', encoding='ascii')
4
5 # Print basic info: shape, dtypes and non null counts
6 print('Shape:', df.shape)
7 print('Data types:')
8 print(df.dtypes)
9 print('\n
10 Non-null counts:')
11 print(df.count())
12
13 # Confirm first 5 rows
14 print('\n
15 Head:')
16 print(df.head())
17
18 print(df.tail())
```

Summary Dataset



	0
Unnamed: 0	1000
age	1000
sex	1000
job	1000
housing	1000
saving_accounts	817
checking account	606
credit amount	1000
duration	1000
purpose	1000
risk	1000

and details from Data types:

).

- **Head:** The first few rows of the dataset provide examples of the provided data (



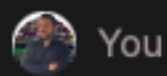
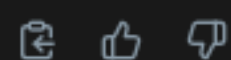
Optional: Click on columns or rows to focus on specific data

⬇️ Export ⓘ

	Unnamed: 0	age	sex	job	housing	saving_accounts	checking account	credit amount	duration	
0	0	67	male	2	own	NULL	little	1169	6	radi
1	1	22	female	2	own	little	moderate	5951	48	radi
2	2	49	male	1	own	little	NULL	2096	12	edu
3	3	45	male	2	free	little	little	7882	42	furn
4	4	53	male	2	free	little	little	4870	24	car

Statistical Summary

Generates a statistical summary of the dataset, including metrics such as mean, median, standard deviation, minimum, maximum, and quartiles for numeric columns.



Statistical description: Generates a statistical summary of the dataset, including metrics such as mean, median, standard deviation, minimum, maximum, and quartiles for numeric columns.



Let me generate a statistical summary of the numeric columns in the dataset:

</> Loaded and analyzed German credit data

Python

Rerun code


Edit code





```
1 import pandas as pd
2 import numpy as np
3
4 # Load the dataset
5 df = pd.read_csv('german_credit_data.csv')
6
7 # Generate statistical summary for numeric columns
8 numeric_summary = df.describe(include=[np.number])
9 print(numeric_summary)
10
11 # Count unique values for categorical columns
12 categorical_columns = df.select_dtypes(include=['object']).columns
13 print("\nUnique values in categorical columns:")
14 for col in categorical_columns:
15     print(f"{col}: {df[col].nunique()} unique values")
16
17 # Check for missing values
18 missing_values = df.isnull().sum()
19 print("\nMissing values per column:")
20
```

Statistical Summary

values in categorical columns, and checks for missing values.

 **Table**
Click to view

 **Table**
Click to view

 **Text**
Click to view

Here is the statistical summary and additional details obtained:

- Numerical Summary:**
The descriptive statistics for the numeric columns are displayed below (





Optional: Click on columns or rows to focus on specific data

 Export 

	Unnamed: 0	age	job	credit amount	duration
count	1000	1000	1000	1000	1000
mean	499.5	35.546	1.904	3271.258	20.903
std	288.8194360957	11.3754685743	0.6536139619	2822.7368759604	12.0588144528
min	0	19	0	250	4
25%	249.75	27	2	1365.5	12
50%	499.5	33	2	2319.5	18
75%	749.25	42	2	3972.25	24
max	999	75	3	18424	72

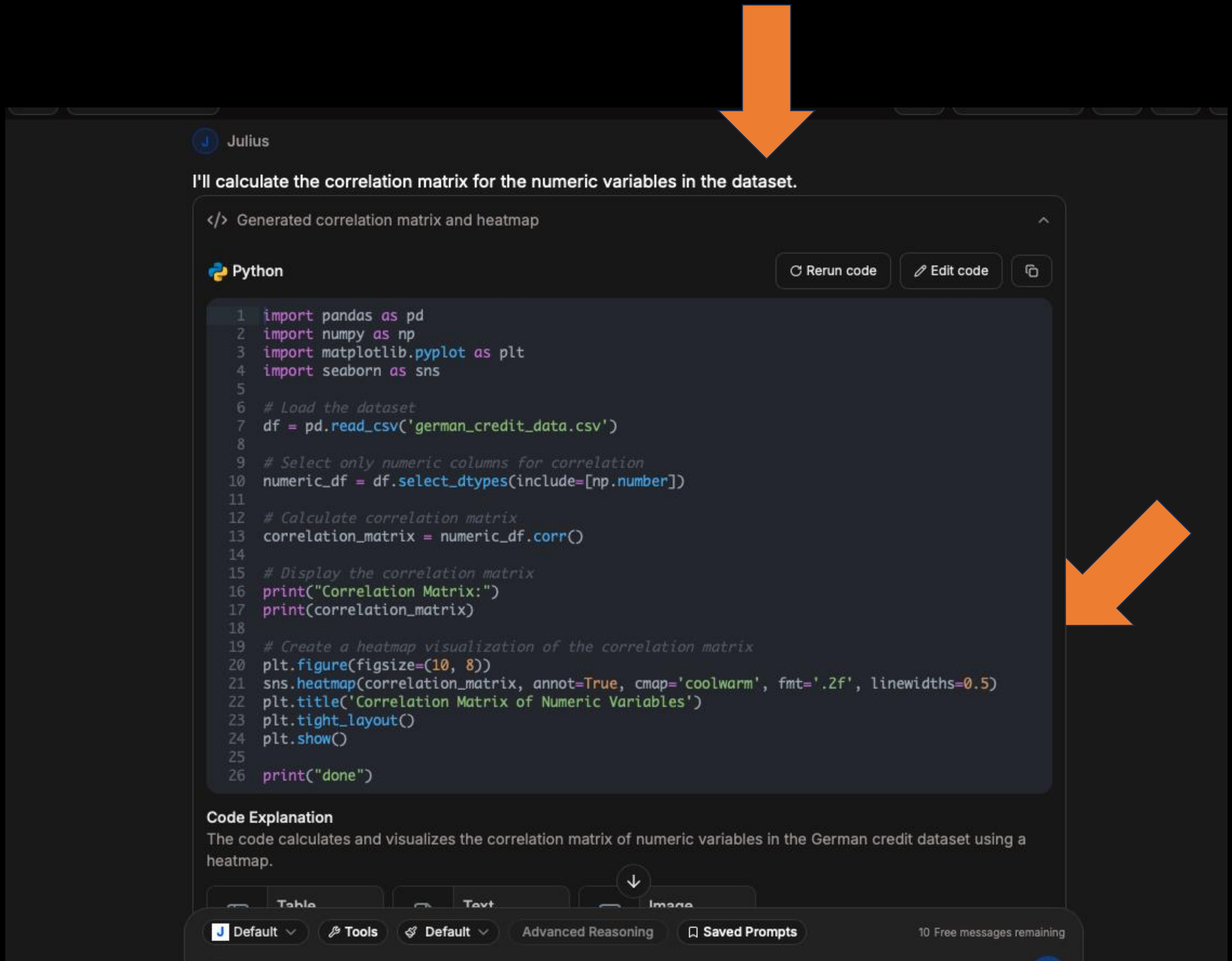
). This summary includes metrics such as mean, standard deviation, minimum, maximum, and quartiles.

- Categorical Columns Unique Values:**
The number of unique values for the categorical columns is as follows:
 - sex: 2 unique values (sex: 2 unique values)
 - housing: 3 unique values ({outputs_dict['979ce

Correlation matrix

Calculates the correlation matrix for numeric variables in the dataset



The screenshot shows a code editor interface with a dark theme. At the top, a user profile icon and the name 'Julius' are visible. Below this, a message states: 'I'll calculate the correlation matrix for the numeric variables in the dataset.' The code editor displays Python code for calculating and visualizing a correlation matrix. The code includes imports for pandas, numpy, matplotlib, and seaborn, followed by loading a dataset, selecting numeric columns, calculating the correlation matrix, and creating a heatmap. The heatmap is titled 'Correlation Matrix of Numeric Variables'. Below the code, there is a 'Code Explanation' section that describes the code's purpose. The bottom of the interface shows a toolbar with icons for 'Table', 'Text', and 'Image', along with a 'Default' dropdown menu, 'Tools', 'Advanced Reasoning', 'Saved Prompts', and a status bar indicating '10 Free messages remaining'.

```
</> Generated correlation matrix and heatmap

Python

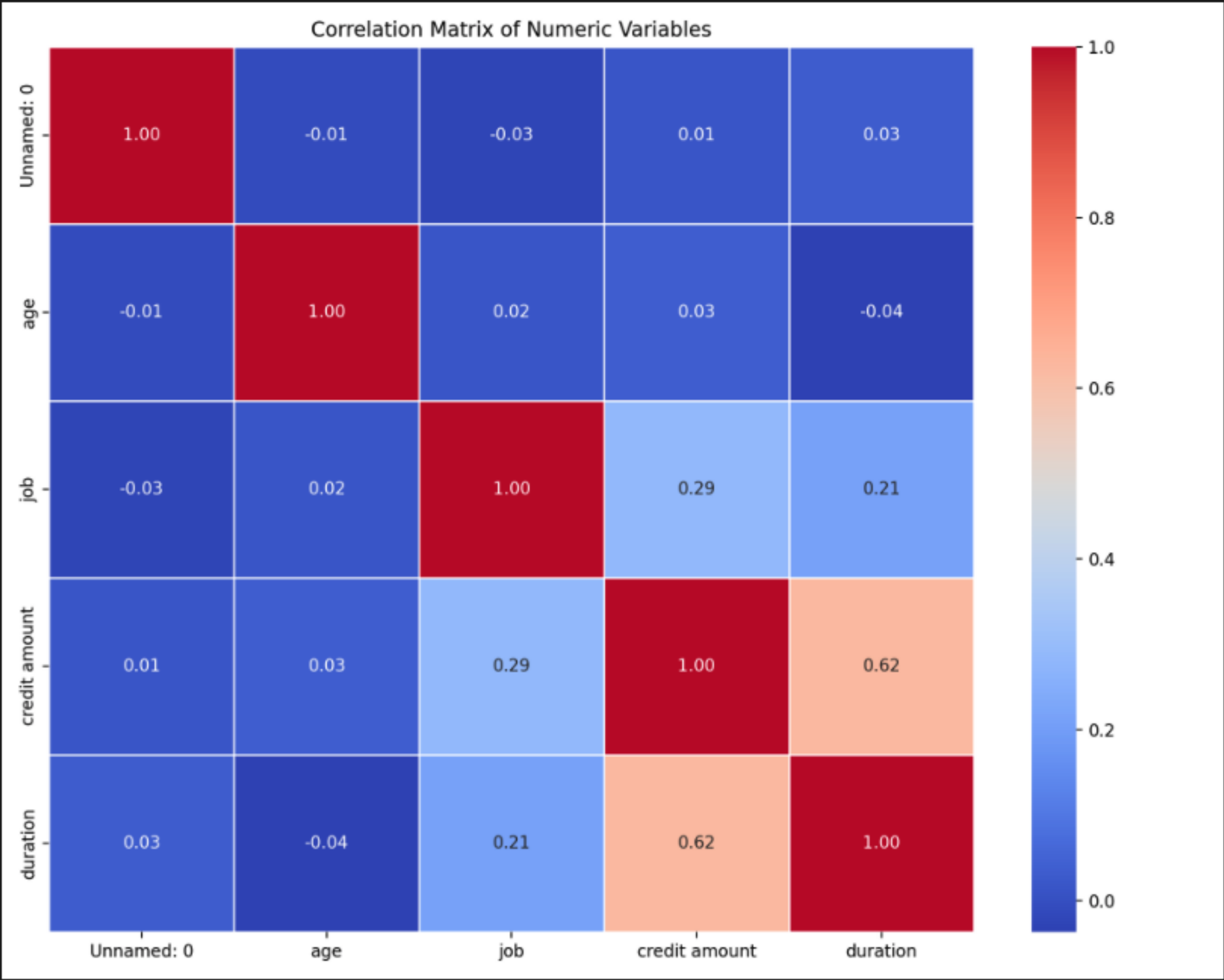
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import seaborn as sns
5
6 # Load the dataset
7 df = pd.read_csv('german_credit_data.csv')
8
9 # Select only numeric columns for correlation
10 numeric_df = df.select_dtypes(include=[np.number])
11
12 # Calculate correlation matrix
13 correlation_matrix = numeric_df.corr()
14
15 # Display the correlation matrix
16 print("Correlation Matrix:")
17 print(correlation_matrix)
18
19 # Create a heatmap visualization of the correlation matrix
20 plt.figure(figsize=(10, 8))
21 sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f', linewidths=0.5)
22 plt.title('Correlation Matrix of Numeric Variables')
23 plt.tight_layout()
24 plt.show()
25
26 print("done")
```

Code Explanation
The code calculates and visualizes the correlation matrix of numeric variables in the German credit dataset using a heatmap.

Table Text Image

Default Tools Default Advanced Reasoning Saved Prompts 10 Free messages remaining

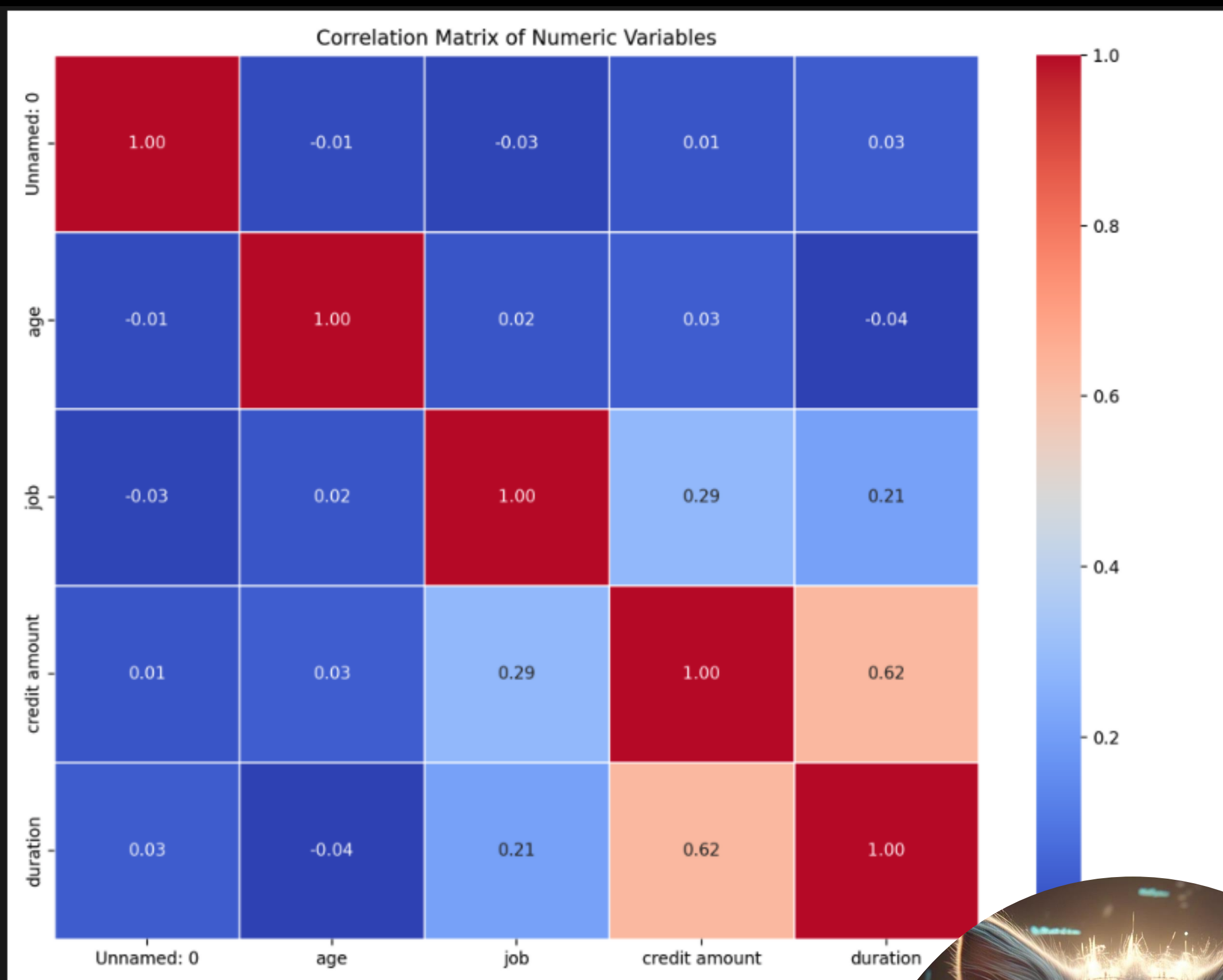
Correlation matrix



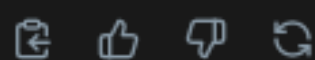
Note that the heatmap uses a "coolwarm" color palette and annotations for ease of interpretation.

The work is done, great job!"

You have a basic EDA quickly



Note that the heatmap uses a "coolwarm" color palette and annotations for ea



Educator in AI

**Artificial
Intelligence**

Data Engineering



Machine Learning

Data Science

📌 **Linkedin** —> <https://www.linkedin.com/in/erlinares/>

👋 **Follow us on X**: <https://x.com/erlinares>^[SEP]

💻 **GitHub**: https://github.com/erlinares/365_AI_Journey/

💬 **Discord**: <https://discord.gg/5fFM2zh8>



Edgar Rios Linares