



IACEE 2024

The 19th World Conference on Continuing Engineering Education

Continuing Engineering Education for a Sustainable Future

Comillas, Spain

May 21-24, 2024

Empowering Data Analytics Learning Leveraging Advanced Large Language Models and Visualization Tools

Jorge Valverde-Rebaza, Aram González, Octavio Navarro & Julieta Noguez



Problem

*Developing data analytics,
data science, and related
projects is challenging for
students and professionals
with a background in
computational thinking*





Problem

*Imagine the challenge for
those without this skill*

Objective

Overcome programming obstacles using

GENERATIVE AI

for agile, efficient, and robust

DATA ANALYTICS SOLUTIONS



Proposal

APPROACHES TO TACKLING A DATA ANALYTICS PROJECT

1

STANDARD PACKAGES FOR PROGRAMMING

Traditional method for individuals with computational thinking and related skills at any level.

2

CHATGPT AS A PROGRAMMING ASSISTANT

Individuals using ChatGPT (or other LLM applications) to generate data analytics solutions and explanation.

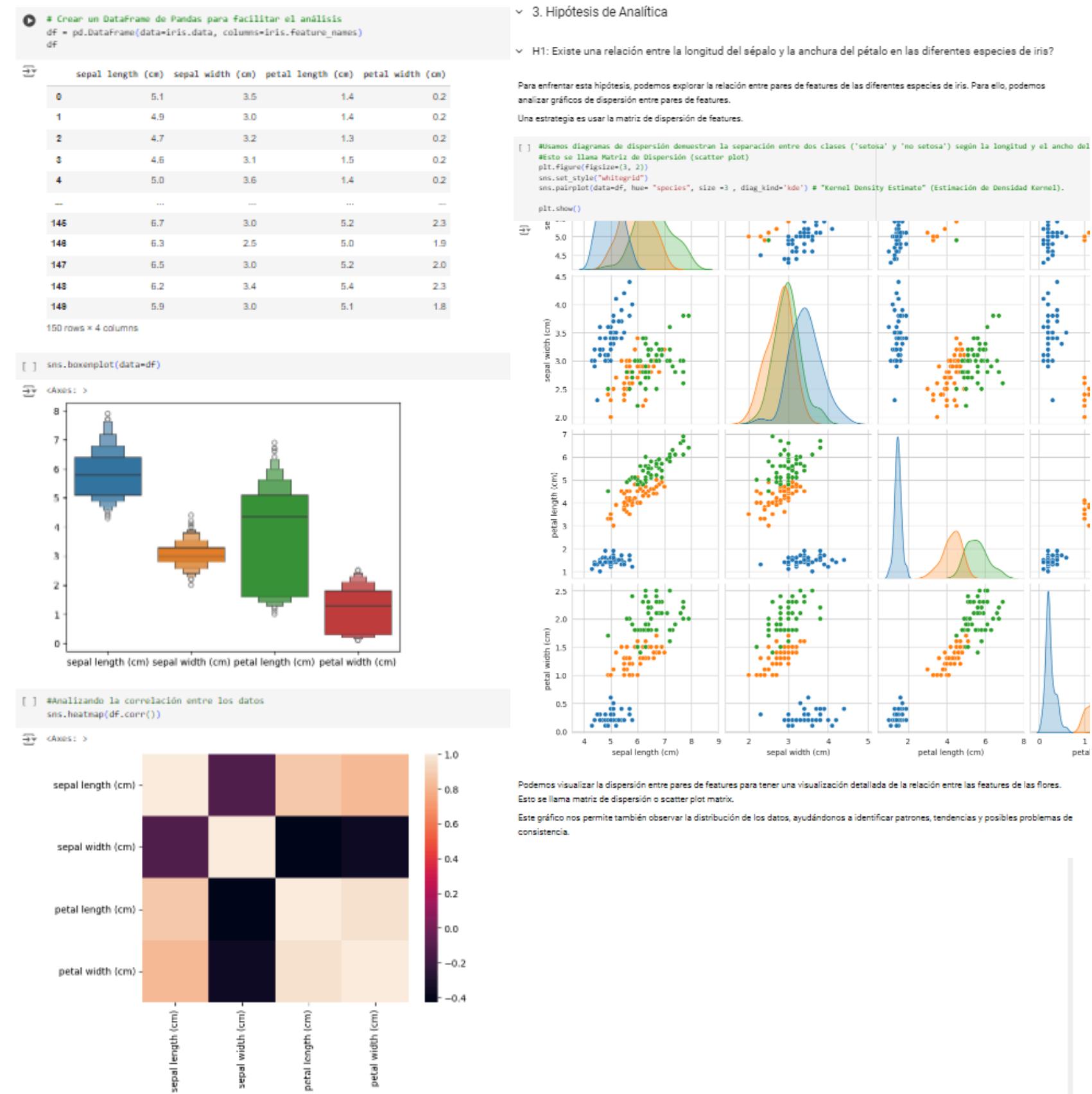
3

LIDA + GPT AS DATA ANALYTICS CONSULTANT

Use LLM applications via API to their fullest potential and obtain impactful and explainable visualizations.

1. STANDARD PACKAGES FOR PROGRAMMING

- Traditional way
- Requires that the individual has knowledge of programming, statistics, specialized packages, etc, as well as understand the data analytics pipeline
- Steep learning curve
- Difficulty in interpreting the results



2. CHATGPT AS A PROGRAMMING ASSISTANT

- New way (since November 2022)
- The individual needs to have a minimum understanding of the project requirements to build an adequate prompt
- ChatGPT can generate the solution and explain it
- Shallow learning curve
- The challenge is to integrate ChatGPT responses into a single project and check whether the "solution" is suitable



You

Estoy trabajando en Google Colab usando Python. Necesito que sobre el iris dataset puedas realizar el código para responder a la siguiente hipótesis:

Existe una relación entre la longitud del sépalo y la anchura del pétalo en las diferentes especies de iris?



ChatGPT

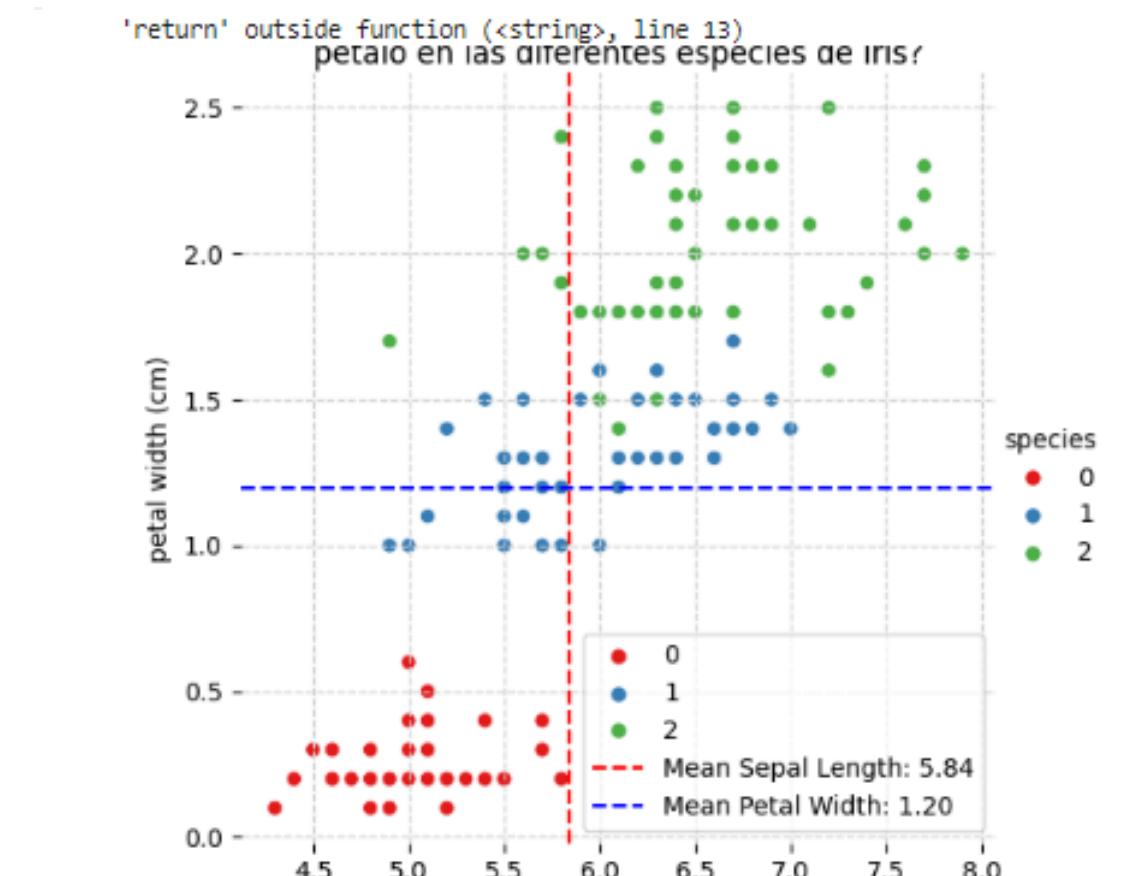
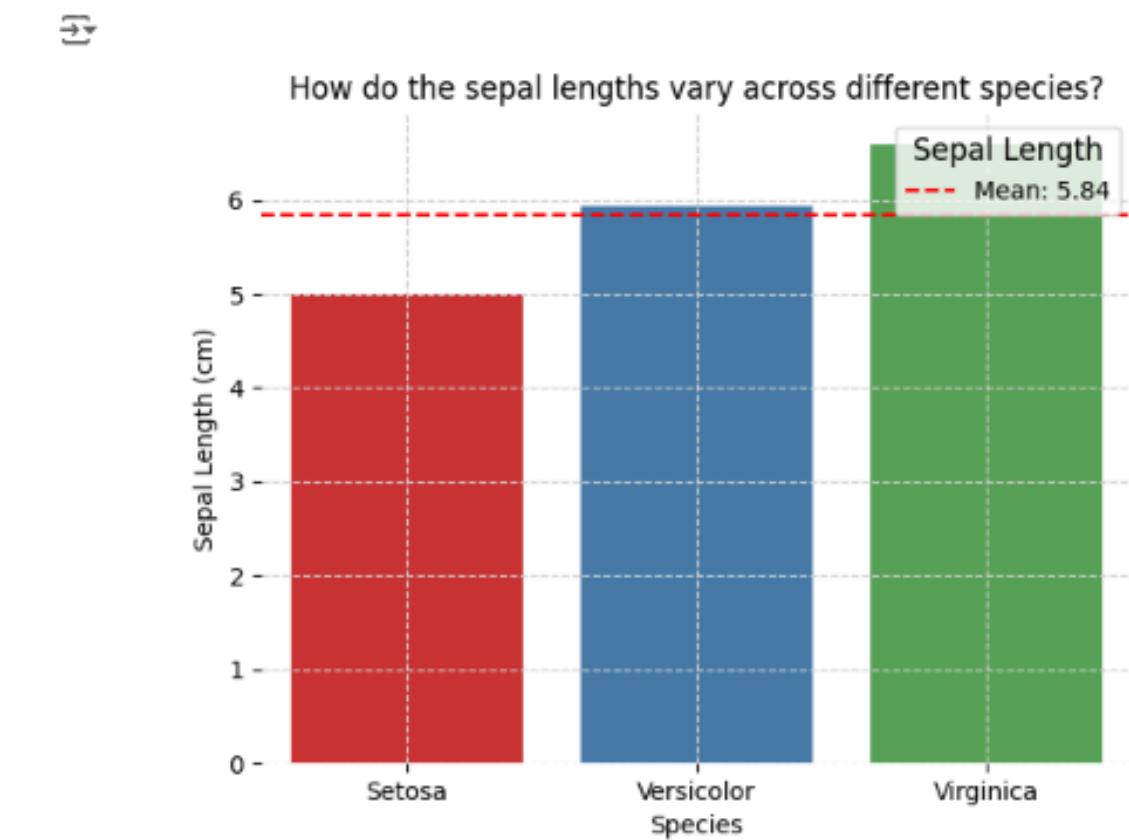
¡Claro! Para abordar esa hipótesis, podemos utilizar el dataset de Iris y realizar un análisis de correlación entre la longitud del sépalo y la anchura del pétalo para cada especie de iris. Aquí tienes un código básico en Python usando Google Colab:

```
python Copy code  
  
import seaborn as sns  
import pandas as pd  
  
# Cargar el dataset de Iris  
iris = sns.load_dataset('iris')  
  
# Calcular la correlación entre la longitud del sépalo y la anchura del pétalo para cada especie  
correlation = iris.groupby('species').apply(lambda x: x['sepal_length'].corr(x['petal_width']))  
  
# Mostrar los resultados  
print("Correlación entre longitud del sépalo y anchura del pétalo para cada especie:  
print(correlation)
```

3. LIDA + GPT AS DATA ANALYTICS CONSULTANT

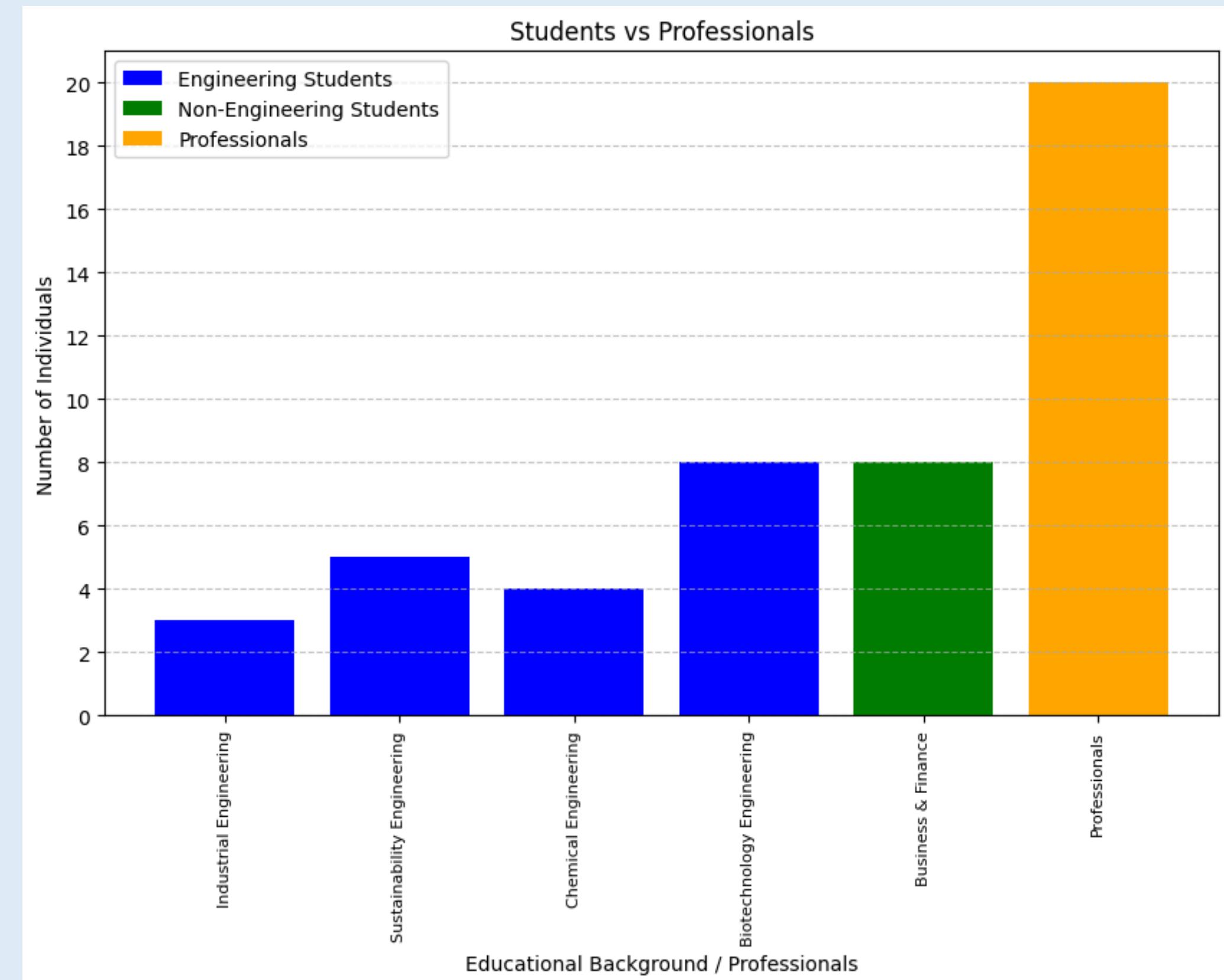
- New way (since November 2023)
- LIDA provide connection between data analytics projects and GPT via API
- LIDA understand the project objective and provides data protection, information security, as well as integrate GPT responses directly into the code
- Visualizations and interpretations are of high quality.
- J-curve of learning

```
[ ] i = 0 #orden/identificador del GOAL
library = "seaborn"
charts = lida.visualize(summary=summary, goal=goals[i], textgen_config=textgen_config, library=library)
display(charts[0])
```



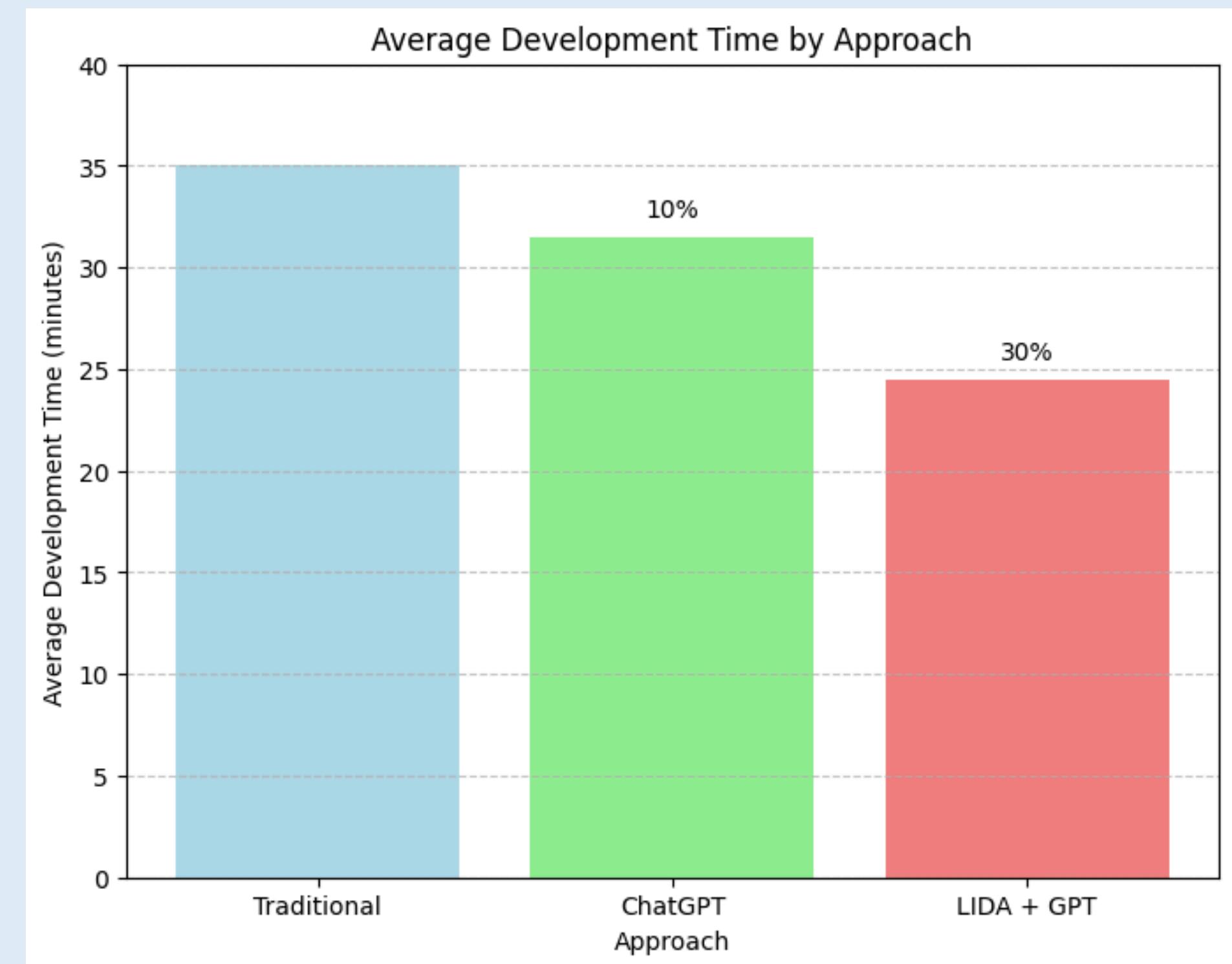
Case Study

- 30 students + 20 professionals
- Diverse non-computational backgrounds
- The case: develop the same data analytics project by using the 3 approaches previously showed
- 90% of the population had no previous experience in data analytics or even programming.



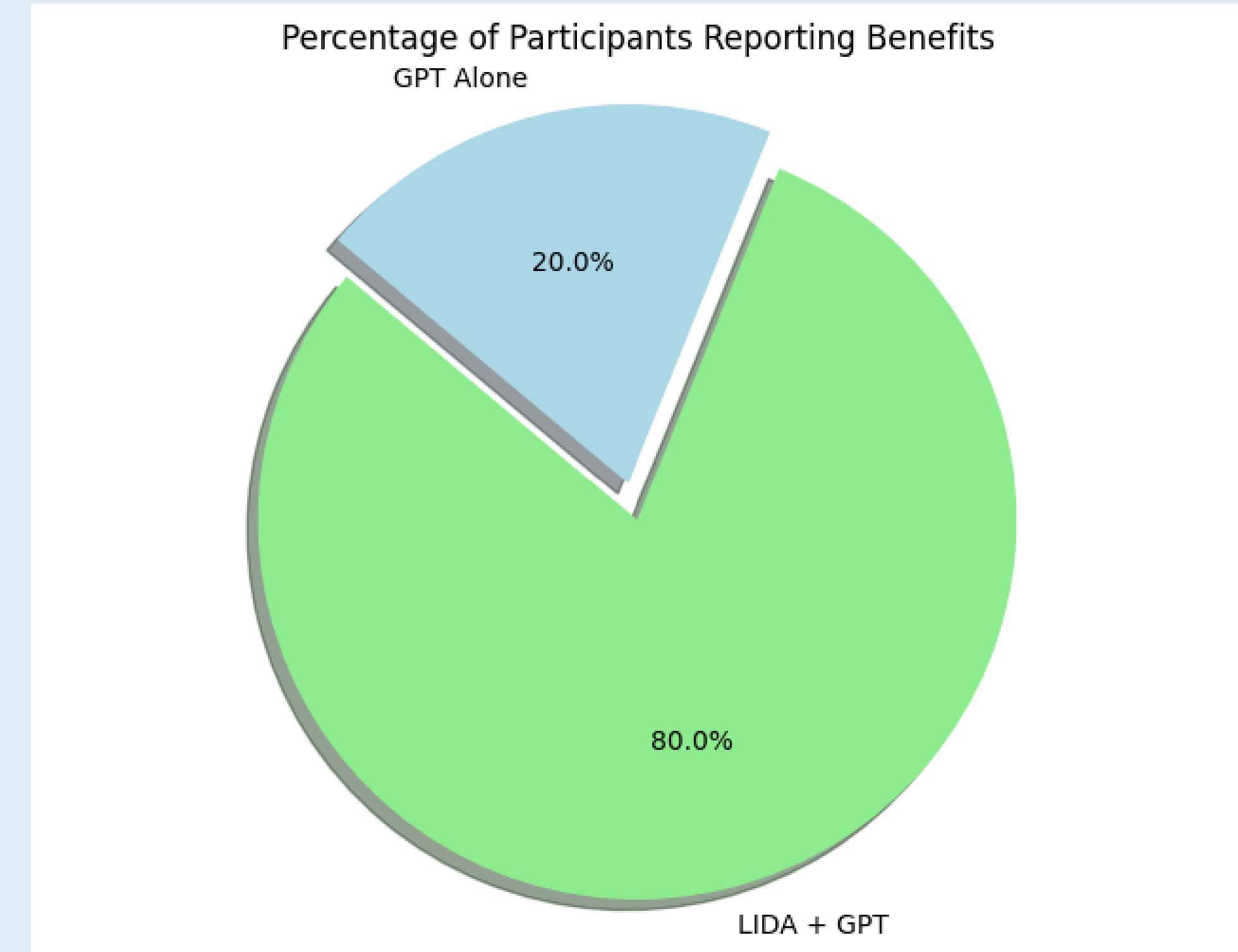
Case Study

- Take 30-45 minutes to build the solution using the traditional approach
- When using ChatGPT speed was 10% faster
- When using LIDA+GPT the developing time was accelerated by up to 30%



Case Study

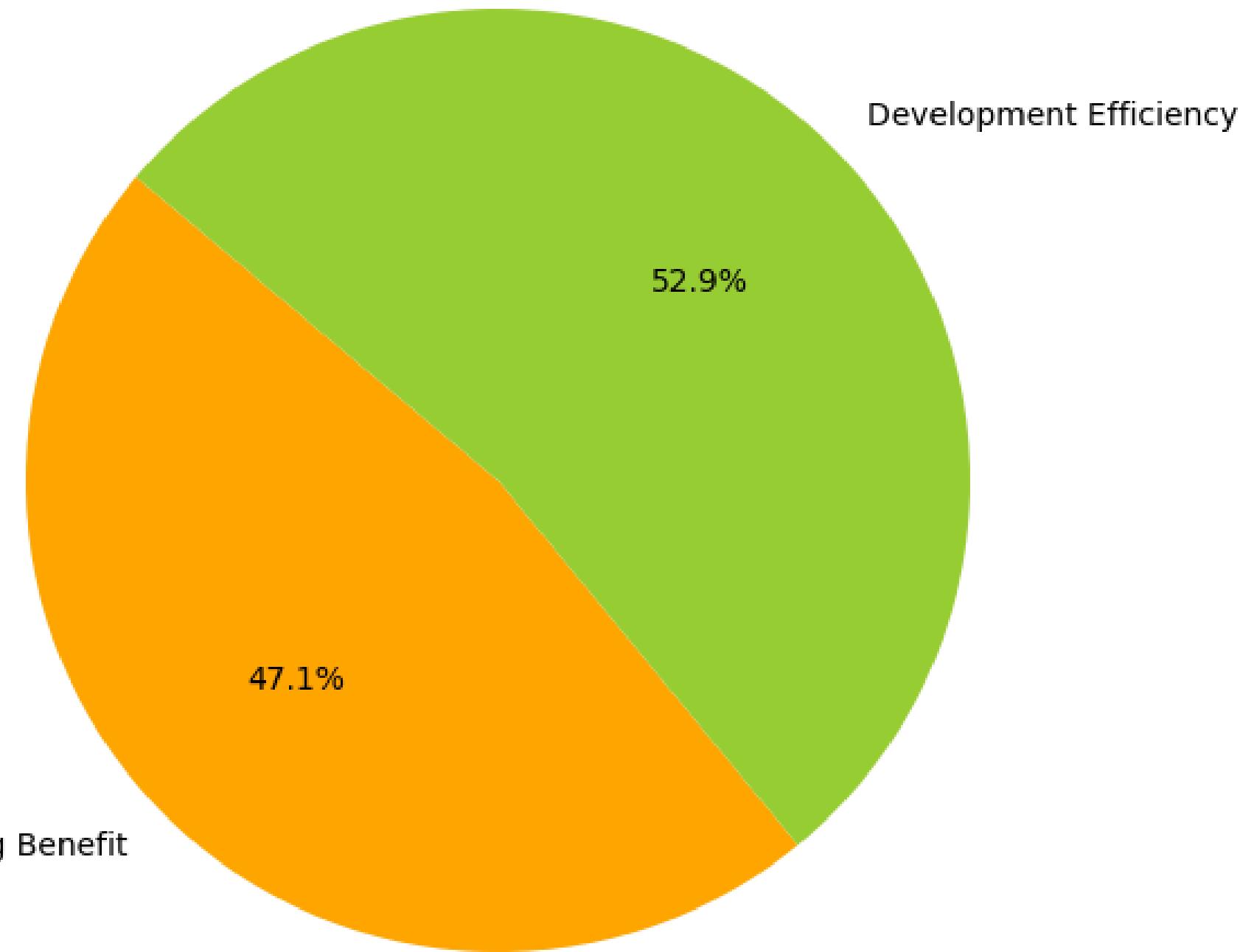
- 80% of participants reported greater learning benefits from using LIDA+GPT
- The remaining 20% stated that ChatGPT yielded better results for them



Case Study

- Participants highlighted the notable capabilities of the LIDA+GPT approach as effective tool to support in balancing learning benefit and development efficiency.

Percentage of Participants Reporting Benefits



Conclusion

- It is possible to extend standard functions of LLMs by using specialized tools as LIDA
- It is possible to enhance skills related to accessibility to computational resources, data literacy, programming education, debugging, and interpreting visualizations by using LLMs in a more effective way



LIDA

Generate visualizations from data





IACEE 2024

The 19th World Conference on Continuing Engineering Education

Continuing Engineering Education for a Sustainable Future

Comillas, Spain

May 21-24, 2024

THANK YOU

jvalverr@tec.mx



/jvalverr



Innovate or Die:
IA, ML e LLMs