Artificial Intelligence – Week 2 (Homework Assignment)

Instructor: Teerapong Panboonyuen

Course Repository:
https://github.com/kaopanboonyuen/SC310005_ArtificialIntelligence_2025s1

Dataset: Fictional Characters

🎯 Objective

This week, you will practice fundamental **data engineering and exploratory data analysis** (EDA) techniques using pandas on the Fictional Characters dataset. Your goal is to create and solve 5 meaningful problems that extract insights or new features from the data using Python.

Additionally, you have an **optional bonus challenge** to leverage modern AI tools (e.g., Hugging Face transformers) to create an innovative feature or analysis for extra credit!

📚 Dataset Description

The dataset contains fictional characters with attributes such as:

- Character Name

- Media Type, Source, Genre

- Role (e.g., Hero, Antagonist)

- Personality Traits, Skills/Abilities

- Alignment (Hero, Villain, Neutral)

- Description and Dialogue Examples

📝 Assignment Instructions

You must:

1. Load the dataset into a pandas DataFrame.

2. Define **5 problems/questions** that you find interesting or relevant for data engineering/EDA. Examples include counting roles, analyzing text length, extracting new numerical features, or summarizing categorical distributions.

3. For each problem, write clear Python code that **solves** it and outputs meaningful results.

4. Use pandas and built-in Python libraries only (no external AI libraries required).

5. Include comments and explanations for your code and results.

6. (Bonus) Use Hugging Face or other modern AI tools to engineer **at least one new feature or insight** (e.g., sentiment score, named entity extraction, text embeddings).

🔥 Examples of Possible Problems

(You don't have to do these exact ones.)

- Count the number of characters by Role (Hero, Antagonist, etc.).

- Calculate the average Description length per Alignment.

- Create a new feature: number of skills listed per character.

- Find the top 3 most common Personality Traits across all characters.

- Analyze how many characters belong to each Media Type and visualize the result.

- (Bonus) Use Hugging Face transformers to analyze the sentiment of the Description text.

💡 Deliverables

Submit a Jupyter notebook or Colab, or Python script containing:

- The code solves your 5 problems.

- Comments explaining your logic.

- Clear outputs showing results (tables, charts if you want).

- If you do the bonus, add an extra section clearly labeled "Bonus: AI Feature Engineering".

🎯 Evaluation Criteria

- Correctness and completeness of solutions.

- Creativity and insightfulness of the problems chosen.

- Use of pandas efficiently and effectively.

- Bonus AI feature adds substantial value and is implemented correctly.

🚀 Getting Started: Dataset Code Snippet

https://github.com/kaopanboonyuen/SC310005_ArtificialIntelligence_2025
s1/raw/main/dataset/fictional_characters.xlsx

📅 Submission Deadline

The deadline for this assignment will be announced and agreed upon during class. Please make sure to take note of the finalized date as discussed in our session.