Assignment: Performing ETL with an External API

Objective:

In this assignment, you will demonstrate your ability to extract data from an external API, perform data transformations using Python's pandas library, and save the transformed data as a CSV file on your local machine.
Instructions:

Step 1: Choose an API

1. Visit RapidAPI or GitHub's Public APIs.
2. Select an API of your choice that provides interesting data for your ETL project. Ensure that the API is free to use.
3. Document the chosen API's base URL, endpoints, and any required authentication (if applicable). You will need this information to connect to the API.

Step 2: Connect to the API

1. Use Python to connect to the chosen API. You can use libraries like requests to make API requests.
2. Retrieve data from the API by sending appropriate HTTP requests to the specified endpoints.
3. Store the API response in a Python variable.

Step 3: Load Data into Pandas

1. Import the pandas library in your Python script.
2. Create a pandas DataFrame using the data obtained from the API response.
3. Explore the data by displaying the first few rows and checking the data types.

Step 4: Perform Data Transformations

1. Apply at least three data transformations to the DataFrame. Transformations can include filtering, sorting, grouping, or creating new columns.
2. Document each transformation step in your code and provide comments explaining the purpose of each transformation.

Step 5: Save as CSV

1. Save the transformed data as a CSV (Comma-Separated Values) file.
2. Choose a meaningful name for the CSV file that reflects the data and the transformations performed.
3. Specify the directory on your local machine where you want to save the CSV file.

Questions for this assignment

Were you able to connect to an API of your choice using postman?
Did you successfully check the columns and apply relevant transformations?
Were you able to save the data as a CSV?
If you run your script from end to end, will it run without any errors on the way?