Description of SLURM Script, Python Code, and Toy Dataset

SLURM Script (abhishek_slurmscript.sh):

The SLURM script is used to submit a job for running the Python code on a SLURM cluster. It automates the process of setting up the required environment, installing dependencies, and executing the Python code. The script includes the following steps:

1. Defines the necessary SLURM directives such as the partition, nodes, CPU/memory requirements, job name, and output file.

2. Creates a temporary virtual environment to isolate the Python environment for this job.

3. Installs the required packages (pandas) within the virtual environment using pip.

4. Activates the virtual environment to ensure the script uses the correct Python interpreter and dependencies.

5. Changes to the directory where the Python code is located.

6. Executes the Python code using the `python` command.

7. Deactivates the virtual environment.

8. Removes the temporary virtual environment and its associated files.

Python Code (abhishek_python_script.py):

The Python code processes a provided dataset using the pandas library. It performs data analysis based on the input data. The code follows these main steps:

The code follows these main steps:

1. Dataset Loading: The code reads the provided CSV dataset using the **pandas** library's **read_csv** function.
2. Data Processing: The code performs various data processing tasks, including converting date columns to the appropriate data types using **pd.to_datetime** function, grouping the data based on patient IDs and drug start dates, and generating a new column for lot advancement.
3. Treatment Analysis: The code analyzes the treatments for each patient, identifies unique treatments, and determines the lot advancement based on treatment history.
4. Result Output: The code generates a final dataframe with patient ID, treatment details, drug start and end dates, and the corresponding lot numbers.

Toy Dataset (lot_example.csv):

The toy dataset (CSV format) serves as an example input for the Python code. It contains sample data related to patient treatments, including the treatment start and end dates. The dataset includes the following columns:

- "patid": Patient ID.

- "treatment": Treatment information.

- "drug_start": Start date of the treatment.

- "drug_end": End date of the treatment.