PATHNOSTICS
MOVING CARE FORWARD

Take-Home Test for Senior Data Scientist

Given the attached CSV file "urine_test_data.csv", which contains 1000 samples of urine test data, your task is to perform some exploratory data analysis. Each row represents a single sample and contains:

- Cell counts for 10 different organisms (labeled as Organism_1 to Organism_10)
- Resistance (R/S) for 17 different antibiotics (labeled as Antibiotic_1 to Antibiotic_17)
- Presence (1) or absence (0) of 5 different resistance genes (labeled as Gene_1 to Gene_5)

Your tasks are:

1. **Data Visualization**: Generate matplotlib/seaborn plots that show:
    - The percentage of samples resistant (R) and sensitive (S) to each antibiotic.
    - The presence of each resistance gene across the samples.
    - The distribution of cell counts for each organism.

The plots should have appropriate legends and labeling.

2. **Unit Testing**: Write a function that adds new data to the existing dataframe and a corresponding unit test that verifies whether the data has been added correctly.
3. **Documentation**: Provide clear and concise documentation for all your code. This includes, but is not limited to, comments in your code, usage examples for your functions, and explanations of your chosen plotting method.
4. **Version Control with Git and GitHub**:
    - Initialize a new repository on GitHub and clone it to your local machine.
    - Add your Jupyter notebook and any other relevant files to the repository.
    - Make at least one pull request during your process.
    - Document your process, code, and results in a README file in Markdown format.

Please submit your GitHub repository link as your submission. The quality of your code, the clarity of your explanations, the correctness of your outputs, and your use of Git and GitHub will be evaluated.