Homework Assignment

In this assignment, you will analyze the dataset on data science salaries from Kaggle. You will complete this assignment using only (generic) Python. No additional Python modulaes like Pandas, Numpy, SciPy are allowed. You will get 0 for this assignment if you use any of these modules.

For the dataset, we use data salaries dataset from Kaggle:
`https://www.kaggle.com/datasets/ruchi798/data-science-job-salaries`

This data set with 608 rows and 11 columns:

1. work_year:

2. experience_level: EN Entry-level / Junior MI Mid-level / Intermediate SE Senior-level / Expert EX Executive-level / Director

3. employment_type: PT Part-time FT Full-time CT Contract FL Freelance

4. job_title:

5. salary:

6. salary_currency:

7. salaryinusd: The salary in USD

8. employee_residence: Employee's primary country of residence

9. remote_ratio:

10. company_location:

11. company_size: The average number of people that worked for the company during the year: S less than 50 employees (small) M 50 to 250 employees (medium) L more than 250 employees (large)

Questions:

1. load the dataset csv file as a list of lines using Python

2. compute average salary by year?

3. which year was the lowest and which year was the highest?

4. for each experience level, compute the average salary (over 3 years) for each job title?

5. For which job title was it the minimum and maximum?

6. for each year and job title compute the annual salary?

7. which job title experienced the maximum change in salary? Which job title had the minimum change.

8. compute the average salaries for each remote ratio? how many entries are there?

9. which company locations pay the highest salary and lowest salaries.

10. Would you change your resume after analyzing this dataset?