Practical Exam: Career Atlas Inc.

Career Atlas Inc. provides salary insights to help companies benchmark compensation and guide job seekers in making informed career decisions. Recently, there have been inconsistencies in the reported salary data. Your team is tasked with cleaning and analyzing a dataset containing job salary details to support a reliable salary trend report.

Instructions

- You should start with the dataset: **ds_salaries_dirty.csv**.

- Submit your work as a **notebook (.ipynb)** with **code and markdown cells explaining your process, decisions, and findings**. This is just as important as your code.

- When you're done, upload your notebook using this submission link
- Deadline: **18/06/2025**

Task 1: Clean and Validate the Salary Data

Start by examining and cleaning the dataset. Some records contain missing or inconsistent values.

- You should start with the data in the file "ds_salaries_dirty.csv".

- Your output should be a DataFrame named clean_salaries.

Apply the following rules:

Column Name	Criteria
work_year	Integer. Year the salary was reported. Must not be missing.
experience_level	Categorical. One of 'EN', 'MI', 'SE', 'EX'. Replace missing/invalid with 'EN'.

employment_type	Categorical. One of 'FT', 'PT', 'CT', 'FL'. Replace missing values with 'FT'.
job_title	String. Remove entries with less than 3 characters.
salary_in_usd	Numeric. Replace any values below 10,000 or above 1,000,000 with median.
company_location	Should be a valid 2-letter country code. Replace missing with 'US'.
company_size	One of 'S', 'M', 'L'. Replace missing or invalid with 'M'.

Task 2: Salary Trends by Experience

Investigate how experience level affects average salary across employment types.

- Group the data by experience_level and employment_type, and note observations in markdown.

- Compute average salary_in_usd.

- Your result should be a DataFrame named salary_trends and Notes stating your findings in markdown

Task 3: Remote Work Analysis

Analyze and discuss the effect of remote work on salaries.

- Compare average salaries between remote (remote_ratio = 100) and non-remote (remote_ratio = 0) roles.

- Output should be a DataFrame named remote_salary_comparison with two rows:

  - Remote

  - On-site