Data Scientist Salaries

Introduction

Dataset

The dataset (DataSalaries.csv) includes remuneration information for the merging world of data science. Variables included are listed below:

- **work_year:** The year in which the employee started working for the company.
- **experience_level:** The level of experience of the employee, such as entry-level (EN), mid-level (MI), senior (SE), etc.
- **employment_type:** The type of employment, such as full-time (FT), part-time (PT), contract (CT), etc.
- **job_title:** The job title of the employee.
- **salary:** The salary of the employee in their local currency.
- **salary_currency:** The currency in which the salary is given.
- **salary_in_usd:** The salary of the employee converted to US dollars.
- **employee_residence:** The country in which the employee is currently residing.
- **remote_ratio:** The percentage of time the employee works remotely.
- **company_location:** The country in which the company is located.
- **company_size:** The size of the company, such as small (S), medium (M), or large (L).

The dataset is provided in a CSV file format and can be loaded into statistical software for analysis.

Task

Using the provided dataset, you are required to:

1. Visualize and describe all (appropriate) numeric variables in the dataset.
2. Visualize and describe all (appropriate) categorical variables in the dataset.
3. Use the appropriate visualizations/summaries to find all variables that are associated with salaries.
4. Are the variables "work_year" and "experience_level" Dependent? Explain what this means in plain language.
5. Are the variables "work_year" and "job_title" Dependent? Explain what this means in plain language.
6. Create a time series plot showing how the median income for data scientists has changed over the years included in the dataset. Is there a trend? How would you personally forecast the salaries over the next 20 years?
7. Which locations have the highest median salaries?
8. Order the variance's for salaries across various experience levels.
9. Create one additional question to be answered in full detail with appropriate visualizations/summaries.

Your report should be written in LaTeX and include all code and output used for analysis. Submit your report as a PDF file.