johnniedanielsaint@gmail.com

Capstone Project: Global Data Science Salaries Analysis

Project Overview:

This capstone project aims to apply SQL to analyze a real-world dataset on global data science salaries, focusing on filtering, grouping, aggregating, and modifying data. You'll be working with the datascience_salaries_2024 dataset to derive valuable insights into the data science profession, salary distribution, and company characteristics across different countries. The project is designed to help you integrate all SQL skills you've learned.

Dataset:

The datascience_salaries_2024 dataset includes information on job titles, salaries in USD, company locations, employee residences, experience levels, and more.

Capstone Goals:

1. **Develop a deep understanding of data science salary trends.**
2. **Apply various SQL concepts such as filtering, ordering, grouping, and updating records.**
3. **Generate meaningful insights from the dataset to inform decisions.**

Tasks:

Task 1: Data Overview and Exploration

- Retrieve the first 10 rows of the dataset.
- Retrieve a list of distinct job_title values.
- Count how many unique employee locations (employee_residence) are present in the dataset.

Task 2: Salary Analysis by Job Title and Country

- Retrieve the job title and average salary for each job title, ordered from highest to lowest.
- Get the top 5 highest paying job titles.
- Calculate the average salary for employees in different countries (company_location), filtering only countries with at least 10 employees.

Task 3: Experience Level and Remote Work Impact

- Calculate the average salary based on the experience_level of employees.

- Find the average salary for employees who work fully remotely (remote_ratio = 100).
- Compare the average salary of employees who work fully remotely versus those who don’t (remote ratio less than 100).

Task 4: Salary Trends Over Time

- Find the number of employees hired per year, grouped by experience_level.
- Get the highest, lowest, and average salary for each year.

Task 5: Filtering and Logical Operators

- Retrieve the details of employees who live in the US and earn more than $150,000.
- Find all employees who work in Germany or have a salary below $80,000.
- Retrieve employees who either work remotely or have an experience level of SE (Senior).

Task 6: Updating Records

- Increase the salary by 10% for all employees who have the job title 'Data Scientist'.
- Change the job_title to 'Senior Data Analyst' for employees with more than 10 years of experience (experience_level = 'SE').

Task 7: Insights Presentation

- Prepare a report or presentation summarizing the insights derived from the SQL queries.
- Highlight key trends, such as the highest-paying countries, the impact of remote work on salaries, and the distribution of employees across different job titles and experience levels.

Final Deliverables:

1. SQL queries for each task.
2. A summary report with key findings.
3. Optional: Visualize some findings using a tool like Power BI, Excel, or Python.