SQL Questions for Data Analytics Project

Data Preprocessing & Cleaning

1. Identify and count missing values in each column.
2. Remove duplicate job postings based on job_title, company_name, and job_location.
3. Standardize job location names (e.g., ensure country names are consistent).
4. Fill missing salary values using the median salary for the same job title.
5. Convert job_posted_date to a proper datetime format.
6. Extract the year and month from job_posted_date.

Exploratory Data Analysis (EDA)

1. What are the top 10 most common job titles?
2. How many job postings are remote (job_work_from_home = True)?
3. What percentage of jobs require no degree (job_no_degree_mention = True)?
4. What are the top 5 most in-demand skills based on job_skills?
5. Count jobs by country and visualize the distribution.
6. Find the average salary per country for jobs with valid salary_year_avg.
7. Show the number of job postings by company, listing the top 10 employers.
8. Determine the most common job_schedule_type.

Salary Analysis

1. Find the highest-paying job title based on salary_year_avg.
2. Compare the average salaries between full-time and part-time jobs.
3. Identify the salary trends over time (e.g., monthly average salaries).
4. Find the top 5 countries with the highest average salaries.
5. Compare average salaries for remote vs. non-remote jobs.
6. Identify the correlation between salary and required skills (e.g., do Python jobs pay more?).

Industry Trends & Insights

1. Identify the top industries hiring for data-related jobs.
2. Find the most common job locations for Data Engineers.
3. Analyze skill trends by country (e.g., most popular skills in the USA vs. Europe).
4. Identify the companies that offer health insurance the most (job_health_insurance = True).

5. Find the percentage of jobs offering health insurance per country.
6. How many jobs require cloud skills (job_type_skills contains 'cloud')?
7. Determine the most in-demand database skills (job_type_skills contains 'sql', 'nosql', etc.).
8. Compare the demand for Python vs. R vs. SQL in job postings.

Time-Based Analysis

1. Show the trend of job postings per month for the past year.
2. Determine the busiest hiring months across different years.
3. Identify which companies are consistently hiring over time.
4. Analyze how job postings changed before and after COVID-19 (if applicable).