**Project Name:** Career Pathway Recommender System

**Group Name: Team 2**

**Group Members: Anusha Boppa**

**Sunil Kumar Bandili**

**TharunTeja Bavandlapelli**

**Venkateswarlu Jampani**

**Sairamlaxman Reddy Burugapalli**

**2. Problem Statement**

**Optimizing Career Recommendations for Job Seekers**

The goal of this system is to optimize career recommendations for job seekers by leveraging structured job market data, skill assessments, and salary insights. The system will match job seekers' current skills, education, and experience with potential career opportunities while also identifying skill gaps for improvement.

The objective is to:  
Maximize career alignment between job seekers and available job opportunities.  
Optimize salary predictions to help users make informed career choices.  
Minimize skill gaps by recommending relevant learning paths.

**3. System Components and Relevant Data**

|  |  |  |
| --- | --- | --- |
| **Component of Problem Statement** | **Module to Include** | **Relevant Data (Variables/Fields) Produced by the Module** |
| Career Alignment | Job Matching Engine | Job Title, Industry, Required Skills, Experience Level |
| Salary Optimization | Salary Estimation Model | Job Title, Location, Experience Level, Expected Salary |
| Skill Gap Minimization | Skill Assessment & Recommendation | Current Skills, Missing Skills, Suggested Courses |
| Industry Trends | Market Analysis | Job Growth Rate, Demand for Skills, Top Hiring Companies |

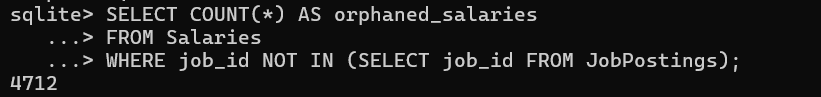
**4. Data Tables and Fields**

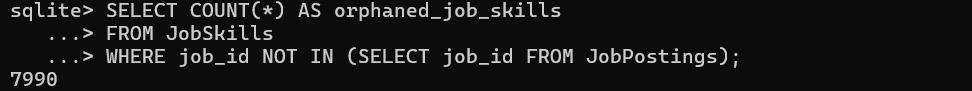
Below are the specific datasets used, their fields, and data types:

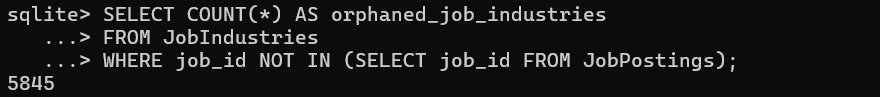
|  |  |  |  |
| --- | --- | --- | --- |
| **Table Name** | **Field Name** | **Data Type** | **Description** |
| **Job Postings** | job\_id | Integer | Unique Job ID |
|  | company\_title | String | Name of the Hiring Company |
|  | title | String | Job Title |
|  | description | Text | Job Description |
|  | max\_salary | Float | Maximum Salary for the Job Role |
|  | pay\_period | String | Payment Type (Yearly/Hourly) |
| **Salaries** | job\_title | String | Job Title |
|  | location | String | Job Location |
|  | salary | Float | Expected Salary for the Role |
| **Industries** | industry\_id | Integer | Unique Industry ID |
|  | industry\_name | String | Industry Name |
|  | top\_companies | String | Top Companies Hiring in this Industry |
| **Job Industries** | job\_id | Integer | Unique Job ID |
|  | industry\_id | Integer | Industry Category for the Job |
| **Job Skills** | job\_id | Integer | Unique Job ID |
|  | skill\_id | Integer | Skill Required for the Job |
| **Benefits** | job\_id | Integer | Unique Job ID |
|  | benefit\_name | String | Type of Benefit Offered |

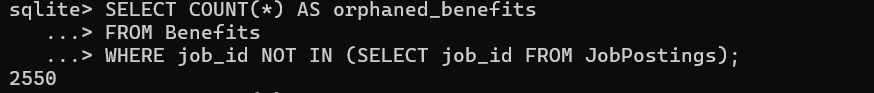
**Referential integrity**

Referential integrity ensures that relationships between tables are maintained properly. In the Career Pathway Recommender System, job\_id is a key identifier used across multiple tables. If a job\_id exists in dependent tables (Salaries, JobSkills, JobIndustries, and Benefits) but not in the primary table (JobPostings), it is considered an orphaned record.







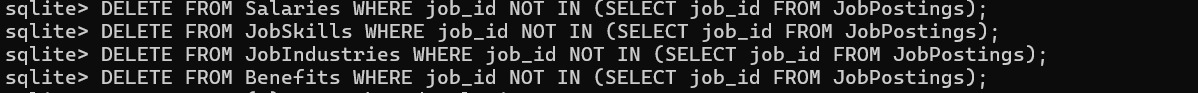


The output from these queries showed:

* **4,712 orphaned records** in Salaries
* **7,990 orphaned records** in JobSkills
* **5,845 orphaned records** in JobIndustries
* **2,550 orphaned records** in Benefits

**Delete the orphaned records**

To maintain database integrity, orphaned records should be removed. The following queries delete records where job\_id is not present in JobPostings:



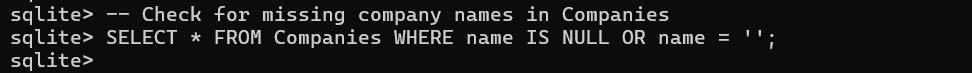
This ensures that all records in the dependent tables (Salaries, JobSkills, JobIndustries, and Benefits) have a valid reference in JobPostings.

**Field-Level Integrity Checks**

Field-level integrity ensures that data in each column follows expected rules, such as no missing values, valid formats, and appropriate constraints.

**Check for Missing Company Names**

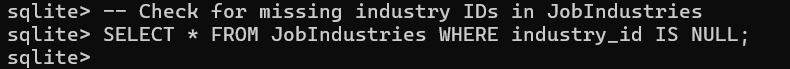
The following SQL query detects company records with missing names in the Companies table:



* This ensures that every company in the database has a **valid name**.
* Missing company names can cause issues in job postings, salary analysis, and company-based insights.

**Check for Missing Industry IDs**

This query identifies missing industry IDs in the JobIndustries table:



Every job should be associated with an industry category.

Missing industry IDs might result from incorrect data entry or broken foreign key relationships.

**Check for duplicate records**

Duplicate records can lead to incorrect data aggregation, misleading insights, and performance issues.

A black screen with white text

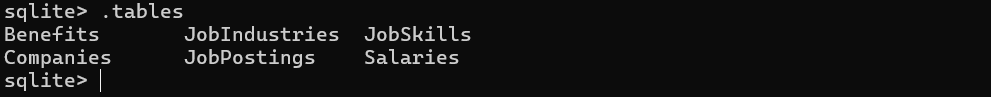
AI-generated content may be incorrect.

Manual changes:  
  
deleted some columns completely which contained mostly null values, and are not useful for my project

**Process Used**

Created tables for all my csv files using in this project

List the tables in the database



First 5 rows from job postings

A screen shot of a computer screen

AI-generated content may be incorrect.

List the first 5 rows from salaries

A screen shot of a computer

AI-generated content may be incorrect.

View the first 5 rows from job skills

A black screen with white text

AI-generated content may be incorrect.

First 5 rows from Jobb industries

A black screen with white text

AI-generated content may be incorrect.

First 5 rows from benefits

A black screen with white text

AI-generated content may be incorrect.

Count the number of rows from all the tables

A computer screen shot of white text

AI-generated content may be incorrect.

**Generative AI Used: chatgpt**

Asked for the query to check the number of orphaned records