**Data Model:**

To implement LinkedIn job posting analysis, here are some of the key entities mentioned which will be helpful in drawing meaning full insights. The two major tables used after de-normalization are,

• Job Postings

• Companies

**Job postings:**

It is a central repository for job listings and is commonly used in job search platforms. The attributes used in this table are job id , company id , title, description, work\_type, location, remote allowed, sponsored, salary details: currency, max salary, med salary, min salary, pay\_period and compensation type, benefits: inferred and types, formatted work type, views, original listed time, listed time, expiry, job posting url, application type, industry\_ids, skills, skills desc, formatted experience level, applies, application url ,closed time, posting domain.

PRIMARY KEY: job id FOREIGN KEY: company id

id (ObjectId):

The unique identifier for each document in the collection. It's an automatically generated ObjectId by MongoDB.

job\_id (Number):

A numerical identifier for the job posting, unique for each job.

title (String):

The job title or position name.

description (String):

A text field containing the job description, providing details about the responsibilities, requirements, and other relevant information about the job.

work\_type (String):

Describes the type of work associated with the job (e.x., full-time, part-time, contract).

location (String):

Represents the geographical location where the job is based.

remote\_allowed (Number):

Indicates whether remote work is allowed or not. It's represented as a number, boolean with 1 for true and 0 for false.

sponsored (Number):

Indicates whether the job posting is sponsored. Like remote\_allowed, it's represented as a number.

formatted\_work\_type (String):

A formatted or human-readable representation of the work type. Ex: Full-time

views (Number):

The number of views or impressions the job posting has received.

original\_listed\_time (String):

The original time when the job was listed.

listed\_time (String):

The current listed time for the job.

expiry (String):

Represents the expiration date or time for the job posting.

job\_posting\_url (String):

The URL or link to the job posting.

application\_type (String):

Describes the type of application process (e.x., OffsiteApply, ComplexOnsiteApply).

salary\_details (Object):

An object containing details about the salary, such as currency, maximum salary, minimum salary, pay period, compensation type, and median salary.

currency (String):

Represents the currency in which the salary is denominated. For example, it could be "USD" for U.S. dollars or "EUR" for Euros.

max\_salary (Number):

Represents the maximum salary offered for the job posting. It indicates the highest amount an employee can earn for this position.

min\_salary (Number):

Represents the minimum salary offered for the job posting. It indicates the lowest amount an employee can earn for this position.

pay\_period (String):

Describes the period at which the salary is paid. Common values might include per year, per month,per week or per hour. (e.x., yearly, monthly, hourly).

compensation\_type (String):

Describes the type of compensation structure associated with the job posting. ex: BASE\_SALARY.

med\_salary (Number):

Represents the median salary for the job posting. The median is the middle value in a set of numbers and is useful for understanding the central tendency of the salary distribution.

industry\_ids (Array of Numbers):

An array of numerical identifiers representing the industries associated with the job.

skills (Array of Strings):

An array of strings representing the skills required for the job.

benefits (Array of Objects):

An array of objects representing benefits associated with the job. Each object may have inferred and types fields.

Inferred (Number): Whether the benefit was explicitly tagged or inferred through text by LinkedIn

type(Array of Strings): Type of benefit provided (401K, Medical Insurance, etc)

company\_id (Number):

A numerical identifier representing the company associated with the job posting.

applies (Number):

The number of job applications received for the job posting.

application\_url (String):

The URL or link to the job application.

formatted\_experience\_level (String):

formatted representation of the experience level required for the job.

Ex: entry, associate, executive,

posting\_domain (String):

The domain where the job posting is originally posted.

closed\_time (String):

The time when the job posting was closed.

skills\_desc (String):

A description details related to the required skills.

**Companies:**

The companies table contains data about various companies and serves as a reference for job

postings. This table is typically linked to the job postings table to associate job postings with specific companies. It includes details about the company\_id, company’s name, description, company size, url, address\_details: address, city, state, country and zip\_code, employe\_counts: employee\_count, follower\_count, time\_recorded, specialities, industries.

PRIMARY KEY: company id

Along with these primary key’s while executing queries in mongodb it by default creates the primary key “\_id” for the tables.

\_id (ObjectId):

The unique identifier for each document in the collection. It's an automatically generated ObjectId by MongoDB.

company\_id (Number):

A numerical unique identifier representing the company.

name (String):

The name of the company.

description (String):

A text field containing a description of the company.

company\_size (Number):

Indicates the size of the company, represents the number of employees.

url (String):

The URL associated with the company, which could be the company's website.

address\_details (Object):

An object containing details about the company's address, including fields like address, city, state, country, and zip code.

address (String):

Represents the street address or location of the company.

city (String):

Represents the city where the company is located.

state (String):

Represents the state or province where the company is located.

country (String):

Represents the country where the company is located.

zip\_code (String):

Represents the ZIP code associated with the company's address.

specialities (Array of Mixed):

An array of mixed types representing the specialities of expertise associated with the company.

Ex: advertising, recruiting etc.

industries (Array of Strings):

An array of strings representing the industries in which the company operates.

employee\_counts (Array of Objects):

An array of objects representing employee counts and follower counts for the company. Each object may have fields such as employee\_count and follower\_count.

employee\_count (Number): Number of employees at company.

time\_recorded (String):

A string representing the time when the record was recorded. Unix time of data collection.

follower\_count (Number): Number of company followers on LinkedIn.