INST327: Database Design & Modeling Department of Information Studies, University of Maryland, College Park Team Project Proposal - Friday October 6, 2023 Section 0102 Group members: Kamal Patel, Beza Ermias, John Seman, Bemenet Berhanu, Fariha Prapti

Team Project Proposal

Introduction:

The IT job posting is a relational database from Poland that has been storing information related to job posts. There are many types of jobs roles within the IT sector that have been posted about. It includes different componental attribute aspects that are common on most of the job posts. Just to summarize a few, it includes about job titles, country of job, city, workplace type, skills required, experience levels, salary, and other informational components one can find on a job posting. The dataset given has stored data about different job postings from February to November 2022. We only plan to include like a month or two of data to limit the scope and so for example in practice wherever our database is used, there is only the latest and up-to-date information to give to users to avoid confusion and for better job search results.

The goal of this semester-wide project is to get more familiar with relational databases and how you can add or extract data or create new databases from an existing database. The task is to create a new and improved database for Polish IT job postings, including only the most important and useful of the attributes and data components. Only including information from a certain time period scope, like the most recents one to two months only. This will make it more efficient to read and understand. It would also allow the programmer the flexibility to add newer data easily and quickly. Combining some of the columns into one column is something that we can do to reduce the size of the database frame. In our new/refined database: these are some of the potential tables (entities) to include: Job Type, Company, Location Information, Salary, Skills, Experience. This covers the most important information for any job posts. In our new/refined database: these are some of the potential columns (attributes) we might choose to include into those tables: Job title, Experience level, Skills required, Job publish date, Job type: remote/onsite, City, country, Salary under different tables where they will be most represented, be well summarized of their purpose.

Sometimes out of data information and or extra information is less efficient and useful. There is a need for up-to-date information databases, especially for fields like Information technology because the world is getting more digital and technology based each passing day, so more IT jobs will be available everywhere. This database is based on Europe's job posting, done in poland. It could be used generally everywhere if some changes are implemented. A central place to store the latest IT job posts. Information Science is related to this topic, so we choose as it's more relevant to us. Sometimes it's difficult for one to find a right job or the company hiring to find right candidates because of outdated listings that people don't see the news ones available right away.

Target Audience:

Job seekers: Can make use of a normalized database, so that whenever they filter/ put a query on a job platform, they can find more relevant job feeds matching their skills, experiences and other specified filters. The query would match the database entities, attributes, or data elements. **Employers, Recruiters:** Would help to find better qualified candidates matching their criteria for their job posts.

Researchers, Government, Educational Institutions, and Media: Could make use of such a database to study most updated trends in the rising IT sector and better understand the job prospects and behaviors of the field and professionals in it.

Potential entities/tables with potential attributes/columns:

Table name: Job_Title

Columns/attributes names: Job Title, Department, Qualifications, Job Type, Education Level

Job Title Department Qualifications	Job Type	Education Level
-------------------------------------	----------	-----------------

Software	Software	SQL,Python,CS	Full- Time	Bachelor
Eligineer	Development	S,III ML,JAVA		

Table name: Location

Columns/attributes names: Address, City, State, Zip_Code, Country

Address	City	State	Zip_Code	Country
751 Mowatt Ln	College Park	MD	20704	US

Table name: Skills

Columns/attributes names: Skill_ID, Description, Relevant_Tech, Skill_Level, Certifications

Skill_ID	Description	Relevant_Tech	Skill_Level	Certifications
JP564	is in charge of designing and enhancing a product or website's overall usability.	Figma	Intermediate	AWS

Table name: Interview

Columns/attributes names: Interview_ID, Interview_Date, Interview_Time, Interview_Type,

Techincal_Assement

Interview_ID	Interview_Date	Interview_Time	Interview_Type	Techincal_Asse ment
SB435	11-04-2024	1:55 PM EST	Remote	Coding

Table name: Company_Size

Columns/attributes names: Company_ID, Company_Size, Number_Empolyee

Company_ID	Company_Size	Number_Empolyee
A234	Large	15,000

Table name: Published_Date

Columns/attributes names: Publish_Date, Due_Date,

Publish_Date	Due_Date	Days_to_apply
10-31-2023	11-31-2023	31 Days

Table name: Contact_Info

Columns/attributes names: Contact_Name, Email_Adress, Phone_Number

Contact_Name	Email_Adress	Phone_Number
John Doe	Jdoe@amazon.com	240-654-3564

Entities/tables you will not include in the database:

Table name: Salary

Columns/attributes names: Employee_position, Employee_salary, Net_Salary

Employee_position	Employee_salary	Net_Salary
UX Designer	0	0

Table name: Currency_exchange

Columns/attributes names:

Exchange_Rate	Targret_currency	Source_currency
0.85	EURO	USD

Sample data:

In order to keep the project proper and achievable, it is essential to define the scope of the data included in it. Starting with a clear definition of the project's objectives and goals can help you to reduce the scope. What particular issue are we attempting to solve with the data? Who will make use of or profit from the gathered information? Understanding the purpose or the audience's needs will help guide the data selection. Moreover, to reduce the amount of data in a large dataset, the following can be considered. Rather than analyzing an entire dataset, selecting a random or systematic sample that represents the larger population might reduce data volume while maintaining statistical validity. Or aggregate data, such as summarizing daily recordings into monthly averages; can simplify the dataset without compromising important insights. The new database will only keep records from the latest time and date range, so whoever might use this type of database to extract latest job post information can get quick access and up-to-date information.

Questions that the database will be able to answer:

- 1 Name the top 5 companies with the most job listings.
- 2 What is the total number of job postings in the database ?
- 3- How many jobs in the database require employees to physically come to work?
- 4 What abilities are usually demanded by senior software engineers?
- 5 What kind of work environment are you looking for in terms of company size?
- 6 How many companies require their employees to interview remotely?
- 7 Which country is having the most job openings?
- 8 Can you provide a breakdown of job postings by job titles?
- 9 Which job title doesn't require any programming skills?
- 10 Can you filter the database to the jobs that are closest to my location?

Diversity, Equity, and Inclusion Considerations:

We must take into account capturing and retaining a variety of job postings in the technology industry when evaluating the database design for diversity, equity, and inclusion. When it comes to problems that could be fixed, bias and a lack of inclusivity may be of concern. For instance, diversity makes sense because a corporation needs to make some kind of diversity in order to be more inclusive. Even in the database, gender diversity is not specified, which is crucial for promoting diversity. Another would be salary equity, which would take into account justice and inclusivity regardless of race, gender, and many other factors. There are a variety of general solutions to this issue, but for everyone to feel included and advanced in pursuing their goals, society as a whole needs to reflect the diversity and equity aspects of the tech industry.

Data Privacy, Fair Use, Other Ethical Considerations"

One thing that we are concerned about when it comes to fair use is demographics. If users have to complete a questionnaire when signing up for their account, AI could potentially discriminate against different groups of people. People of different races, ethnicities, genders, and sexualities can be given different job opportunities, and that is not fair. In order to best tailor results to the individual, questions like disability status may come up. In terms of data privacy, these companies should not have access to this information whatsoever. One thing that all companies can aim to do is to avoid using AI that reinforces stereotypes, and biases. Additionally, they can overall make their platforms more secure in order to keep their users' data private.