

Introduction:

Our team created a database that contains information about a specific dataset we selected labeled “Polish IT Jobs Listings”. With the use of this database, the Department of Defense group decided to base the database off the likes of Ukrainian job opportunities and narrow down the data to see if the jobs listed fit their style and if eligible to work. The ultimate objective in the development of this database relative to the sample data we utilized is to view all the listed jobs throughout Poland and even some nations in Europe like the United Kingdom, France, Ukraine, Slovakia, and many other nations based in Europe and a select few in Asia like the United Arab Emirates.

Our database contained 36 total columns worth of sample data and was reduced to about 21 columns as there were a lot of repeating values and irrelevant columns of data needed to help the user understand the specific job listings. The main aspects of our dataset we ranked as the most informative is the job, location, experience level, preferred skill, open to hire the likes of Ukrainian immigrants due to the invasion, if you can get a remote interview, and salary details. We improved the dataset by narrowing it down from 37,787 rows of data to about 1,000 rows due to the fact that importing large amounts of data can take a long time when creating the database and importing specific datasets into the tables. The major problem from the original dataset is the columns provided were just confusing and not helpful especially if you put yourself in a position as a user searching for job opportunities. Our database will be able to help users navigate through job listings without coming across vague information.

Database Description:

Our database is designed based on the location and the type of job the user may be searching for. The database’s purpose is to be a viable method to display jobs through the users filtering. Our database contains 8 tables along with 1,000 job results based in the likes of Poland and other European nations.

All the data restructured from the original dataset will be able to guide the users to answer specific questions such as the details specific to the company like location, company size,

salary details, probability of job landing, company size and salary comparisons, and the main aspect of being able to hire Ukrainians. During the invasion that began in March of 2022 in Ukraine, millions of Ukrainians have fled their native land to the likes of many European countries and this dataset will help provide all the necessary information about the job listings. Utilizing the likes of the dataset in SQL workbench will be able to provide the information of the company they are looking into and whether or not it fits their style and supports the needs.

Logical Design:

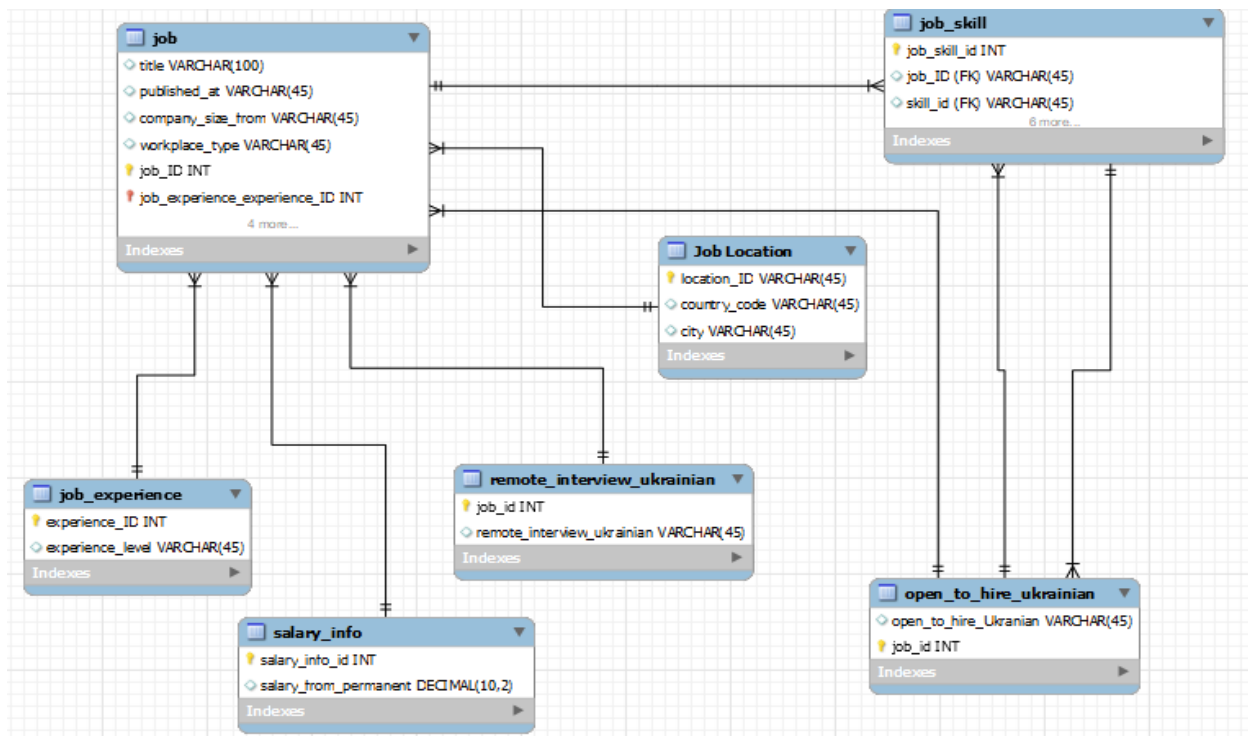


Image 1. Entity Relationship Diagram for Polish IT Job Listings database.

The main objective of the logical design is to make sure there were balanced tables and no irrelevant columns in each specific table. There are a total of seven tables created which allows users to be able to discover all the specific types of data very easily and at a quick rate as there aren't many columns of data.

Our group decided that the job_id foreign key is very crucial to helping users know what job will be displayed as we took the approach to make the likes of an integer value equal the

same value as the other primary keys from the database to easily display specific data for the user in their queries. The tables “job_location” and “job_experience” received their own primary keys in the form of “location_id” and “experience_id” as the data relationships between the two tables along with the “job” table. We made sure that all tables don’t display any potential null values which is why we changed the value of the likes of the Ukrainian aspect of the data to display varchar values rather than a boolean statement.

Physical Design:

Our database functionality is proposed to meet the requirements and specific needs for users that are interested in IT work or currently looking for available positions that suit their necessary skills. The dataset utilized in this database was narrowed down to the most relevant information to guide users and if there were any repeating values in the set. The idea of the dataset understood the needs for the likes of Ukrainian refugees seeking work throughout Europe due to unfortunate circumstances which is why we implemented the likes of a ‘job_location’ table which will display where the job is located by city and country code. This would be the most important aspect of the data analysis along with the ‘job’ table that will be able to lay out if it is a remote job or not, the title of the position, company size, and when the job was listed. The users may have specific issues where they would not be able to travel to the specific nation the job was listed for which is why we included two tables being remote_interview_Ukrainian and open_to_hire_Ukrainian which will be able to automatically showcase if a Ukrainian would have a smooth hiring process as there are listings that may only hire citizens of the likes of Poland. Salary is a table we decided not to make a huge factor as most of the job listings in the 1,000 rows of jobs only displayed zeros, but luckily there are a good amount of listings that have actually listed permanent salaries. One of the queries will be able to represent the likes of the salaries above the value zero to make it much easier for users who are curious about the likes of salary.

Sample Data:

The data imported into the created database was retrieved from Kaggle, a platform for data science competitions, machine learning projects, and datasets, where users can collaborate, learn, and compete to solve real-world problems using data analysis and predictive modeling. A csv file was provided that included an estimated 38,000 job listings which was narrowed down to about 1,000 listings to make data importing much more efficient. The sample data included information such as job title, job type, location, skills required, salary information, a few boolean columns, and the estimated size of the company the position was listed as. For the boolean columns like `remote_interview` and `open_to_hire_Ukrainians`, we adjusted it from the boolean value to a varchar due to there being issues with data importing and making it look more appealing by displaying the likes of “Open to hire Ukrainians” and “Open to Remote Interview” rather than a true or false value. Below are two images representing samples of data from the two tables labeled “job” and “job_location”.

	job_id	title	published_at	company_size_from	workplace_type
▶	1	Senior Frontend Software Engineer	2022-07-31	25	remote
	2	Sr. Backend Engineer	2022-07-07	20	remote
	3	Sr. Backend Engineer	2022-05-31	20	remote
	4	Sr. Backend Engineer	2022-05-25	20	remote
	5	Python Developer (4 days/week)	2022-06-30	100	remote
	6	Python Developer (4 days/week)	2022-06-15	100	remote

Image 2. A snippet of the ‘job’ table displaying the job post and minor details.

	location_id	country_code	city
▶	1	GB	London
	2	DE	Berlin
	3	PL	Katowice
	4	GB	London
	5	PL	ystok
	6	PL	Kraka

Image 3. A snippet of the ‘job_location’ table displaying the location of the job, location_id is equal to value of job_id.

Views and Queries:

Query Name	JOIN	FILTER	AGGREGATE	LINKING	SUB-QUERY
view_engineer_salaries	X	X	X	X	
view_lowest_salaries	X	X			X
view_job_counts	X		X		
view_job_count_by_countries			X		
view_senior_jobs	X	X		X	
view_open_remote_ukrainian_jobs	X	X		X	
Total	5	4	3	3	1

The following list describes each query developed for the database and the function:

Query 1: Creates a view displaying average salaries for senior engineers, obtained through joins, filtered by experience levels, and grouped by job title.

Query 2: Creates a view showing job titles with lowest salaries using joins and subqueries, filtering for positive salaries.

Query 3: Creates a view presenting job counts per experience level, utilizing joins and aggregation on job and experience tables.

Query 4: Creates a view presenting job counts per country, with translated country codes and ordered by job count.

Query 5: Creates a view displaying senior-level job titles and experience levels, filtered by publication date.

Query 6: Creates a view filtering jobs open to Ukrainians with remote interview availability, joining multiple tables for relevant information.

Changes from Original Design:

In relation to the original design of the Polish IT Job Listings database, we had to modify a lot of portions in terms of the normalization process and the ERD model showcased in the project progress report. From the feedback received from the course instructors, included the likes of there being one huge middle table that connects all the other tables. One issue with this kind of design is that, for example, partial dependencies are not resolved. This design required further normalization. Some issues with primary keys. Some tables have multiple connections with the main table, which should be avoided. Some fields repeat more than once. We decided in our normalization process along with the new model rebuild to remove a lot of columns and primary keys in the job_id table. After readjusting the tables and its keys, we made sure there was only one primary key per table and still have the display of the original 8 tables in the database.

Lessons Learned:

One of the key lessons taken from this group project was how effective group communication can bring to a group's potential. Our group has come to an understanding that communication comes first compared to the likes of skills and even work ethic in a group setting. This became an issue as one of the group's biggest obstacles was relative to the likes of communication. The obstacle formed when most group members were not able to attend breakout rooms after class period to discuss anything relative to the project which can cause the likes of not receiving any member's intuition nor ideas while following through the stages of the project. How the group overcame the obstacle of the lack of communication after class period with the group was the practice of effective communication and understanding each member's schedule. Ultimately our group got into the habit of communicating effectively on the application 'GroupMe' and would utilize the modern group work application 'Zoom'. Time is a very important factor in the form of group work as there are deadlines and it is preferable to understand specific material before beginning the next task. One example during the project that

caused problems was the normalization process which is arguably the most important task of the database project. The task was done two days prior to the due date which caused a lot of stress between the group and led to intensive learning of the normalization process. Understanding the material before getting started weeks prior would have been the better route in this situation. In terms of the creation of the database, fortunately the material covered in the likes of the lectures and labs helped us significantly understand the concepts of importing and exporting data along with hard coding specific portions of a dataset to make it more easier to understand and visually appealing.

Potential Future Work:

A potential route we will take from this database is to include more than just 1,000 job listings from the random listings chosen out of the 37,787 listings. One downside in implementing large chunks of data is that it can be very time consuming to import the likes of even 5,000 listings as we want to create more opportunity for Ukrainians and other ethnic groups to search for the jobs that suit their wants and needs. Another potential step we can take to maximize the potential of the database is working with the likes of salaries and the type of currency based on each job as there are different work locations. An example would be to include Euros, British Pounds, Slovak Koruna, Polish Zloty, and many more currencies throughout the European union. To make it more effective we would need to implement a query concept that can convert a specific currency to another currency from the user's choice as it can benefit the job search more efficiently. Our group in particular thinks that the concept of job listings from a csv file isn't the best way to present and visualize the likes of job searching as there are more efficient ways to present a job on the internet. Of course there are many valuable sources in the realm of jobs and in particular LinkedIn is the perfect modern tool to utilize. The likes of images and well-organized layouts make this application so valuable in today's modern technology world. It is even more effective as it can be utilized as an app from the use of your smartphone which can make job searching very easy and with instant, fast results.

INST 327, Section: WB-21

Project Report

08/07/2023

Department of Defense: Mert Inak, Miguel Rodriguez, Jacob Mattheis, Aneesa Munir, Xuhui Zhou

Work Cited:

Just join it. Just Join IT: #1 Job Board for tech industry in Europe. (n.d.).

<https://justjoin.it/>

RSKriegs. (2022, November 20). Polish it job board data from 2022. Kaggle.

<https://www.kaggle.com/datasets/kriegsmaschine/polish-it-job-board-data-from-2022>