

# Project 3 Proposal

## Project Goal:

**Which are the most valued data science skills?**

### 1. Group Members

- Haoming Chen
- James Chun
- Pricilla Nakyzze

### 2. Collaboration Tools

- Communication: **Email**
- Code sharing: **Github**
- Project Documentation: **Google Docs**

### 3. Data Source

- <https://www.kaggle.com/datasets/arshkon/linkedin-job-postings/data>
- <https://www.kaggle.com/datasets/asaniczka/data-science-job-postings-and-skills/data>
- Github

### 4. Relational Database and ER diagram

- Use **MySQL** to create normalized tables. (tables, columns, primary keys, foreign keys)

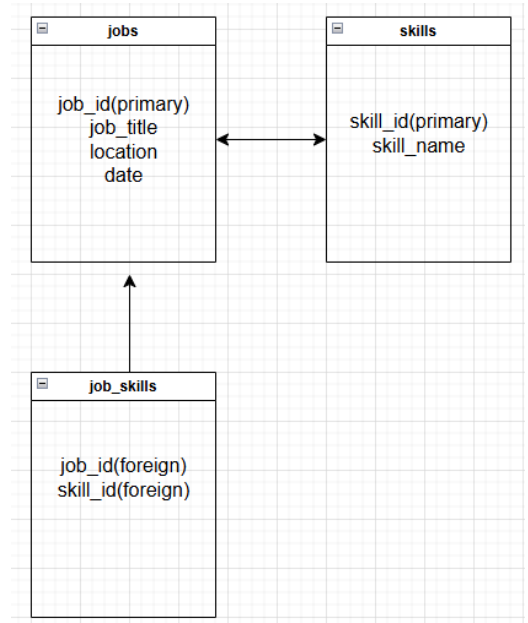
Table Name	Purpose	Fields
jobs	Contain jobs	job_id, job_title, location, date
skills	Skill names	skill_id, skill_name
job_skills	a table shows jobs and skills	job_id, skill_id

- Use draw.io or [dbdiagram.io](https://dbdiagram.io) to draw ER diagram

## 5. Load data into R and Data Handling

- Use packages like tidyverse, DBI, Rmysql to input data to R
- Perform data cleaning, tidying, transformation
- Visualize and analyze
- Summarize findings

## 6. Presentation



Project Timeline:

- Project proposal: 10/12
- Code submission: 10/19
- Presentation: 10/22