Module 9 Challenge

**Background**

It’s been two weeks since you were hired as a new data engineer at Pewlett Hackard (a fictional company). Your first major task is to do a research project about people whom the company employed during the 1980s and 1990s. All that remains of the employee database from that period are six CSV files.

For this project, you’ll design the tables to hold the data from the CSV files, import the CSV files into a SQL database, and then answer questions about the data. That is, you’ll perform data modeling, data engineering, and data analysis, respectively.

**Before You Begin**

1. Create a new repository for this project called sql-challenge. **Do not add this assignment to an existing repository**.
2. Clone the new repository to your computer.
3. Inside your local Git repository, create a directory for this Challenge. Use a folder name that corresponds to the Challenge, such as EmployeeSQL.
4. Note that you’ll add your files to this folder and push the changes to GitHub.

**Files**

Download the following files to help you get started:

[Module 9 Challenge filesLinks to an external site.](https://static.bc-edx.com/data/dl-1-2/m9/lms/starter/Starter_Code.zip)

**Instructions**

This Challenge is divided into three parts: data modeling, data engineering, and data analysis.

**Data Modeling**

Inspect the CSV files, and then sketch an Entity Relationship Diagram of the tables. To create the sketch, feel free to use a tool like [QuickDBDLinks to an external site.](http://www.quickdatabasediagrams.com/" \t "_blank).

**Data Engineering**

1. Use the provided information to create a table schema for each of the six CSV files. Be sure to do the following:
   * Remember to specify the data types, primary keys, foreign keys, and other constraints.
   * For the primary keys, verify that the column is unique. Otherwise, create a [composite keyLinks to an external site.](https://en.wikipedia.org/wiki/Compound_key), which takes two primary keys to uniquely identify a row.
   * Be sure to create the tables in the correct order to handle the foreign keys.
2. Import each CSV file into its corresponding SQL table.

**HINT**

To avoid errors, import the data in the same order as the corresponding tables got created. And, remember to account for the headers when doing the imports.

**Data Analysis**

1. List the employee number, last name, first name, sex, and salary of each employee.
2. List the first name, last name, and hire date for the employees who were hired in 1986.
3. List the manager of each department along with their department number, department name, employee number, last name, and first name.
4. List the department number for each employee along with that employee’s employee number, last name, first name, and department name.
5. List first name, last name, and sex of each employee whose first name is Hercules and whose last name begins with the letter B.
6. List each employee in the Sales department, including their employee number, last name, and first name.
7. List each employee in the Sales and Development departments, including their employee number, last name, first name, and department name.
8. List the frequency counts, in descending order, of all the employee last names (that is, how many employees share each last name).

**Requirements**

**Data Modeling (10 points)**

* Entity Relationship Diagram is included or table schemas provided for all tables (10 points)

**Data Engineering (70 points)**

* All required columns are defined for each table (10 points)
* Columns are set to the correct data type (10 points)
* Primary Keys set for each table (10 points)
* Correctly references related tables (10 points)
* Tables are correctly related using Foreign Keys (10 points)
* Correctly uses NOT NULL condition on necessary columns (10 points)
* Accurately defines value length for columns (10 points)

**Data Analysis (20 points)**

* List the employee number, last name, first name, sex, and salary of each employee (2 points)
* List the first name, last name, and hire date for the employees who were hired in 1986 (2 points)
* List the manager of each department along with their department number, department name, employee number, last name, and first name (2 points)
* List the department number for each employee along with that employee’s employee number, last name, first name, and department name (2 points)
* List first name, last name, and sex of each employee whose first name is Hercules and whose last name begins with the letter B (2 points)
* List each employee in the Sales department, including their employee number, last name, and first name (2 points)
* List each employee in the Sales and Development departments, including their employee number, last name, first name, and department name (4 points)
* List the frequency counts, in descending order, of all the employee last names (that is, how many employees share each last name) (4 points)

**Submission**

Before submitting your Challenge assignment, make sure that you’ve done the following:

* Create an image file of your ERD.
* Create a .sql file of your table schemata.
* Create a .sql file of your queries.
* (Optional) Create a Jupyter notebook of the bonus analysis. (Note: there will be no extra points for completing this.)
* Ensure that your repository has regular commits and a thorough README.md file

To submit your Challenge assignment, click Submit, and then provide the URL of your GitHub repository for grading.

**References**

Data generated by [Mockaroo, LLCLinks to an external site.](https://mockaroo.com/" \t "_blank), (2022). Realistic Data Generator.