Data Engineering

Provide the following:

1. Schema and PNG for table relationships
2. Specify data types, primary keys, foreign keys, etc.
    a. Verify columns are unique, otherwise create a composite key
    b. Create tables in correct order to handle foreign keys

Table/Relationship Review

1. For the schema and picture of the tables and their relationships refer to
    a. Here for the schema
    b. Here for the png

The starter code included 6 excel files with the following tables:

1. Departments
2. Employees
3. Dept_emp
4. Dept_manager
5. Salaries
6. Titles

However when examining the columns in each table, it's important to note that there are two primary tables being used: Departments and Employees. Departments has a primary key of ***dept_no*** and Employees has ***emp_no***. The other remaining tables use these keys as foreign keys to relate back to these primary tables.

When looking at relationships in these databases (refer to the png), this is what follows:

- Departments has a one to many relationship with the Dept_emp table.
    - This means that an employee (as referenced by the emp_no) can be a part of different departments over the lifetime of their employment at the company. But it also means that many employees fall under different departments

- Departments has a one to many relationship with the Dept_managers table.
    - This means that each department can have multiple managers but an employee can only be a manager to a single department.
- Employees has a one to many relationship with both titles and salaries.
    - This means that each employee can hold a number of different titles and different salaries over the course of their time at the company. However this also means that an employee may only hold a single title or salary level at one time.
- Employees has a many to one relationship with both Dept_emp and Dept_managers
    - This refers to the idea that each department can have many employees but employees can only belong to a single department at any given point in time. Similarly, each department manager has many employees under their leadership but each employee only answers to their respective departments' manager.

All 6 tables use similar/shared types of data in their columns:

1. **VARCHAR** is used for columns that include employee names or department names
2. **DATE** is used to handle all columns relating to a hire/start date or birth date.
3. **INT** is used for emp_no since these are all numeric identifications
    a. In contrast dept_no is VARCHAR because it has a set structure of d###