Exercise 1

Historical Data Transformation

Objective: Transform current employee data from a columnar format into a historical, row-based versioning format suitable for database storage.

Task Overview: Your task is to convert an input CSV file containing employee data into a structured format representing historical records of employee compensation, engagement, and performance reviews. The new format requires transforming columnar data into a row-based historical versioning system for insertion into our data warehouse.

Key Instructions:

1. **Effective and End Dates**:
    - Derive 'Effective Date' and 'End Date' for each historical record.
    - Ensure the 'End Date' is one day before the next 'Effective Date' to avoid overlap.
    - For the latest record of an employee, assign a far-future date (e.g., 2100-01-01) as the 'End Date'.
2. **Data Transformation**:
    - Transform columnar data related to compensation, engagement, and review into a row-based format.
    - Each row should represent a specific period with consistent data.
    - If data for a range is missing, inherit values from the most recent past record for the same employee.
3. **Data Copying**:
    - Maintain unchanged values for fields without associated dates across different records.
    - Ensure all relevant data from the input file is accurately reflected in the output format.
4. **Output Format**:
    - The output should be a CSV file formatted for historical data analysis, including fields for employee identifiers, compensation, dates, performance ratings, and engagement scores.
5. **Documentation**:
    - Briefly document your approach and any assumptions made during the transformation process.

Deliverables:

1. A transformed CSV file containing the historical data.
2. A short documentation of your methodology and assumptions.

Evaluation Criteria:

- Accuracy of the transformation based on the provided rules.
- Clarity and efficiency of the documentation.
- Ability to handle missing data and date ranges appropriately.

Files:

- Input File
- Output File

Exercise 2

Context

You have a data set which is a response to a recently run survey in a company on different themes or "Drivers".

As a data analyst, you have to analyse the data set to give useful slice & dice to the HR leader so that they can take actions

Data set

There are two main tables:

1. Employees - The master database with a list of employees
    1. `id` - Primary key
    2. `name` - Employee Name
    3. `department` - Department
    4. `location` - Location
    5. `gender` - Male/Female/Others
    6. `age`
    7. `manager_id` - Self-referenced with employee table for Managers
2. Responses - These are the responses by employees for the survey
    1. `id` - Primary Key

2. `driver_name` - Drivers or themes (Eg: Role Clarity, Career Growth, Policies, etc)
3. `score` - The numerical score they have given from 1-5
4. `employee_id` - The employee who gave this score, foreign key from employee table primary key

- Use the db-fiddle link to see the schema and data- https://www.db-fiddle.com/f/jQv1JemWTruj8iWHNhGgZe/25
- Fork it and write SQL for the problems added to the above fiddle.
- Submit the db-fiddle link of the final solution