

Bài tập 1: Cho một bảng Employee (file employee_01.csv đính kèm) chứa thông tin tất cả employees và manager của các employees này. Mỗi employee có một Id và thông tin manager Id của người quản lý

Dùng pandas liệt kê ra các employees có salary cao hơn manager của mình.

Ví dụ với table employee như sau

| Id | Name | Salary | ManagerId |
|----|-------|--------|-----------|
| 1 | Joe | 70000 | 3 |
| 2 | Henry | 80000 | 4 |
| 3 | Sam | 60000 | |
| 4 | Max | 90000 | |

Lưu ý có một vài nhân viên không có manager, lúc này ManagerId sẽ có giá trị rỗng.

Với table trên kết quả mong đợi là một table output có format và nội dung như sau

| Employee |
|----------|
| Joe |

Giải thích: theo bảng employee nhân viên Joe có managerId = 3 tương ứng là Sam. Joe có salary là 70000 và Sam có salary là 60000 do đó Joe có lương cao hơn manager của mình. Henry có manager là Max và lương Henry thấp hơn lương của Max. Sam và Max không có manager nên không so sánh được lương với manager của mình nên không xuất ra output.

Bài tập 2: Cho bảng Employee (file employee_02.csv đính kèm file này) chứa thông tin tất cả employees. Format bảng như sau

| Id | Name | Salary | DepartmentId |
|----|-------|--------|--------------|
| 1 | Joe | 70000 | 1 |
| 2 | Jim | 90000 | 1 |
| 3 | Henry | 80000 | 2 |
| 4 | Sam | 60000 | 2 |
| 5 | Max | 90000 | 1 |

Một bảng Department (file deparment.csv đính kèm file này) chứa thông tin tất cả departments của công ty

| Id | Name |
|----|-------|
| 1 | IT |
| 2 | Sales |

Dùng pandas viết truy vấn tìm ra các employees có mức lương cao nhất của mỗi phòng ban. Với ví dụ trên, output mong muốn như sau

| Department | Employee | Salary |
|------------|----------|--------|
| IT | Max | 90000 |
| IT | Jim | 90000 |
| Sales | Henry | 80000 |

Giải thích: Cả Max và Jim đều có mức lương cao nhất IT department nên cần liệt kê cả hai người này, trong khi đó phòng Sales chỉ có Henry có mức lương cao nhất nên chỉ liệt kê Henry