BÀI TẬP THỰC HÀNH PANDAS VÀ MATPLOTLIB

Sử dụng lại file salaries.csv với những dữ liệu đã được tính toán từ bài tập Pandas.

Câu 1. Vẽ biểu đồ histogram biểu diễn sự phân bố mức lương theo USD (salary_in_usd).

Câu 2. Vẽ biểu đồ histogram biểu diễn sự phân bố mức lương theo VND (salary_in_vnd).

Câu 3. Vẽ biểu đồ histogram biểu diễn sự phân bố mức độ hoàn thành công việc từ xa (remote_ratio) (tính theo %).

Câu 4. Thêm cột gender (giới tính) và sau đó random cho giá trị 0 và 1, trong đó 0 đại diện cho nam giới và 1 đại diện cho nữ giới. Vẽ biểu đồ tròn biểu diễn tỉ lệ phân bố giới tính và cho biết giới tính nào chiếm tỉ lệ cao hơn.

Câu 5. Vẽ biểu đồ miền biểu diễn tỉ lệ phân bố giới tính (gender) qua các năm.

Câu 6. Vẽ biểu đồ phân tán biểu diễn mối quan hệ giữa giới tính (gender) và mức lương theo VND (salary_in_vnd). Quan sát sự phân bố và xem xét liệu rằng giữa hai đại lượng này có mối quan hệ tương quan với nhau hay không. Nếu có, hãy cho biết mối quan hệ giữa hai đại lượng này có phải là mối quan hệ tuyến tính hay không?

Câu 7. Vẽ biểu đồ phân tán biểu diễn mối quan hệ giữa ngành nghề (job_title) và mức lương theo VND (salary_in_vnd). Quan sát sự phân bố và xem xét liệu rằng giữa hai đại lượng này có mối quan hệ tương quan với nhau hay không. Nếu có, hãy cho biết mối quan hệ giữa hai đại lượng này có phải là mối quan hệ tuyến tính hay không?

Câu 8. Vẽ biểu đồ tròn biểu diễn sự phân bố quy mô công ty (company_size).

Câu 9. Vẽ biểu đồ cột biểu diễn sự phân bố hình thức làm việc (employment_type).

Câu 10. Vẽ biểu đồ cột biểu diễn số lượng nhân sự của các ngành nghề (job_title) có liên quan đến lĩnh vực Khoa học và Kỹ thuật dữ liệu (trong tên ngành nghề có từ "Data").

Câu 11. Tính trung bình mức lương theo USD (salary_in_usd) của tất cả những người được khảo sát theo từng năm. Vẽ biểu đồ đường biểu diễn sự thay đổi trung bình mức lương theo USD (salary_in_usd), từ đó đưa ra dự đoán về xu hướng thay đổi mức lương theo USD (salary_in_usd) trong tương lai.

Câu 12. Tính trung bình tỉ lệ hoàn thành công việc từ xa (remote_ratio) (tính theo %) của tất cả những người được khảo sát theo từng năm. Vẽ biểu đồ đường biểu diễn sự thay đổi tỉ lệ

hoàn thành công việc từ xa (remote_ratio), từ đó đưa ra dự đoán về xu hướng thay đổi tỉ lệ hoàn thành công việc từ xa (remote_ratio) trong tương lai.

Câu 13. Tính trung bình mức lương theo USD (salary_in_usd) của 500 người đầu tiên trong bộ dữ liệu, sau đó thống kê số lượng nhân sự trong số 500 người này theo hai nhóm sau:

- "Muc luong thap": Những người có mức lương thấp hơn mức trung bình.
- "Muc luong cao": Những người có mức lương từ mức trung bình trở lên.

Vẽ biểu đồ tròn biểu diễn sự phân bố mức lương và rút ra nhận xét.

Câu 14. Thống kê tỉ lệ hoàn thành công việc từ xa (remote_ratio) (tính theo %) của 500 người đầu tiên trong bộ dữ liệu theo hai nhóm sau:

- "Lam viec tu xa khong hieu qua": Những người có tỉ lệ hoàn thành công việc từ xa nhỏ hơn 50%.
- "Lam viec tu xa hieu qua": Những người có tỉ lệ hoàn thành công việc từ xa từ 50% trở lên.

Vẽ biểu đồ tròn biểu diễn sự phân bố tỉ lệ hoàn thành công việc từ xa và rút ra nhận xét.

Câu 15. Vẽ biểu đồ cột biểu diễn số lượng công ty đặt tại mỗi quốc gia (company_location).

Câu 16. Vẽ biểu đồ hộp biểu diễn sự phân bố số lượng nhân sự của các ngành nghề (job_title) trong năm 2021 và rút ra nhận xét.

Câu 17. Vẽ biểu đồ hộp biểu diễn sự phân bố mức lương theo USD (salary_in_usd) trung bình theo từng cấp độ kinh nghiệm làm việc (experience_level) trong năm 2020 và rút ra nhận xét.

Câu 18. Vẽ biểu đồ cột nhóm biểu diễn mức lương theo USD (salary_in_usd) trung bình theo từng quy mô nhân sự của công ty (company_size) qua các năm và rút ra nhận xét.

Câu 19. Vẽ biểu đồ cột nhóm biểu diễn số lượng nhân sự theo từng ngành nghề (job_title) qua các năm và rút ra nhận xét.

Câu 20. Vẽ biểu đồ hộp biểu diễn sự phân bố mức lương theo USD (salary_in_usd) của các ngành nghề (job_title) trong năm 2022 và rút ra nhận xét.

Câu 21. Tính trung bình tỉ lệ hoàn thành công việc từ xa (remote_ratio) (tính theo %) và trung bình mức lương theo USD (salary_in_usd) của nghề Kỹ sư dữ liệu (job_title = 'Data Engineer') qua các năm, sau đó vẽ biểu đồ kết hợp cột và đường biểu diễn hai đại lượng trên

theo mô tả sau:

- Mỗi cột tương ứng với một mốc thời gian (work_year) với chiều cao mỗi cột bằng trung bình mức lương theo USD (salary_in_usd).
- Trung bình tỉ lệ hoàn thành công việc từ xa (remote_ratio) được biểu diễn bằng đường.

Câu 22. Tính trung bình tỉ lệ hoàn thành công việc từ xa (remote_ratio) (tính theo %) và trung bình mức lương theo USD (salary_in_usd) của các công ty đặt tại châu Mĩ (company_location = 'US') qua các năm, sau đó vẽ biểu đồ kết hợp cột và đường biểu diễn hai đại lượng trên theo mô tả sau:

- Mỗi cột tương ứng với một mốc thời gian (work_year) với chiều cao mỗi cột bằng trung bình mức lương theo USD (salary_in_usd).
- Trung bình tỉ lệ hoàn thành công việc từ xa (remote_ratio) được biểu diễn bằng đường.

Câu 23. Vẽ biểu đồ miền biểu diễn tỉ lệ phân bố quy mô nhân sự (company_size) qua các năm.

Câu 24. Vẽ biểu đồ phân tán biểu diễn mối quan hệ giữa địa điểm đặt công ty (company_location) và quy mô nhân sự (company_size). Quan sát sự phân bố và xem xét liệu rằng giữa hai đại lượng này có mối quan hệ tương quan với nhau hay không, từ đó rút ra nhận xét về địa điểm đặt các công ty có quy mô nhân sự lớn?

Câu 25. Vẽ Word Cloud (đám mây từ vựng) biểu diễn sự phân bố số lượng nhân sự của các ngành nghề (job_title) trong năm 2022 (size chữ của từ miêu tả ngành nghề sẽ biểu diễn số lượng nhân sự của ngành nghề) và rút ra nhận xét.