

BÁO CÁO CASE STUDY 20

DỮ LIỆU SINH VIÊN TỐT NGHIỆP

TRƯỜNG ĐẠI HỌC KINH TẾ - KỸ THUẬT CÔNG NGHIỆP

Môn học: Lập Trình Python Nâng Cao

Nhóm: Nhóm 06 Lớp: DHKL18A3HN

Ngày nộp báo cáo: 16/12/2025

Danh sách thành viên:

STT	Họ và tên	MSSV	Nhiệm vụ
1	Nguyễn Thùy Linh	24174600155	Nhóm trưởng, Task2 (Truy vấn, thống kê..)
2	Hà Trung Dũng	24174600157	Task4 (Merge dữ liệu,..)
3	Phan Hương Giang	24174600124	Task3 (GroupBy + Tổng hợp)
4	Đỗ Văn Quang	24174600135	Task1 (Đọc, làm sạch dữ liệu)
5	Đương Thành Toàn	24174600131	Task5 (Pivot Table + Stack/Unstack)

I. GIỚI THIỆU

1.1. Mục tiêu

Case Study hướng đến các **mục tiêu** sau:

- Đánh giá chất lượng đào tạo** qua tỷ lệ việc làm của sinh viên sau tốt nghiệp
- Phân tích mức lương khởi điểm** theo ngành học và năm tốt nghiệp
- Xác định mối liên hệ** giữa chứng chỉ và cơ hội việc làm
- Rút ra xu hướng** về việc làm và mức lương để đưa ra các kiến nghị

1.2. Phạm vi dữ liệu

Dữ liệu được phân tích gồm 3 file CSV:

- graduate_info.csv** - Thông tin sinh viên tốt nghiệp
- employment_survey.csv** - Khảo sát việc làm
- certifications.csv** - Chứng chỉ

II. PHƯƠNG PHÁP XỬ LÝ VÀ PHÂN TÍCH DỮ LIỆU

1. Mục tiêu

Đảm bảo kiểu dữ liệu chuẩn xác trước khi phân tích, loại bỏ các giá trị lỗi và không hợp lệ.

1.2. Chuẩn hóa và làm sạch dữ liệu

- Chuẩn hóa student_id (loại khoảng trắng, thống nhất chữ hoa/thường), full_name (loại bỏ khoảng trắng thừa và chữ hoa chữ thường), major (sửa sai chính tả), grad_year (chuyển đổi và lọc bỏ dữ liệu không hợp lệ).
- Xử lý employment_survey: salary (loại dấu chấm/phẩy, xử lý giá trị âm/phi lý bằng NA), job_title, company (xóa khoảng trắng thừa, lỗi chính tả), survey_year (chuẩn lại các định dạng lộn xộn về năm)
- Xử lý certifications: certificate_name (thống nhất tên chứng chỉ), issued_year (loại bỏ các năm lớn hơn năm khảo sát. - xóa bản ghi trùng – chuyển sang kiểu số)

1.3. Kết quả làm sạch

- Thống kê kết quả làm sạch: (Ngoại trừ trùng id sẽ bị xóa còn lại sẽ thay bằng N/A)

File	Số bản ghi ban đầu	Số bản ghi sau làm sạch	Tỷ lệ
Graduate info.csv	100	100	100%
Employment survey.csv	80	80	100%
certifications.csv	120	115	95.83%

2. Kết quả truy vấn & thống kê mô tả (Nhiệm vụ 2)

2.1. Phương pháp thực hiện

Sử dụng value_counts(), groupby() với mean(), Boolean indexing.

Bảng 2.1: Số lượng SV tốt nghiệp theo năm và theo ngành

Năm tốt nghiệp	Số lượng SV
2021	22
2022	40
2023	38

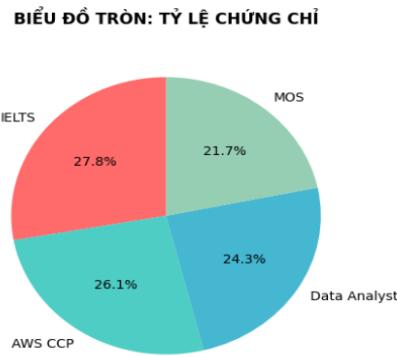
Ngành	Số lượng SV
CNTT	35
Kinh tế	29
QTKD	36

Bảng 2.3: Mức lương trung bình theo ngành

Ngành	Lương trung bình	Lương cao nhất	Lương thấp nhất
CNTT	9133333	10000000	8000000
Kinh tế	9000000	10000000	8000000
QTKD	9000000	10000000	8000000

Bảng 2.4: Số lượng chứng chỉ theo loại

Loại chứng chỉ	Số lượng
IELTS	32
MOS	25
Data Analyst	28
AWS CPP	30
Tổng	115



3. Phân tích tổng hợp - GroupBy (Nhiệm vụ 3)

3.1. Phương pháp phân nhóm

Sử dụng hàm groupby() kết hợp với các hàm tổng hợp như mean(), count(), agg() để phân tích dữ liệu theo nhiều chiều khác nhau.

3.2. Bảng kết quả và phân tích ý nghĩa

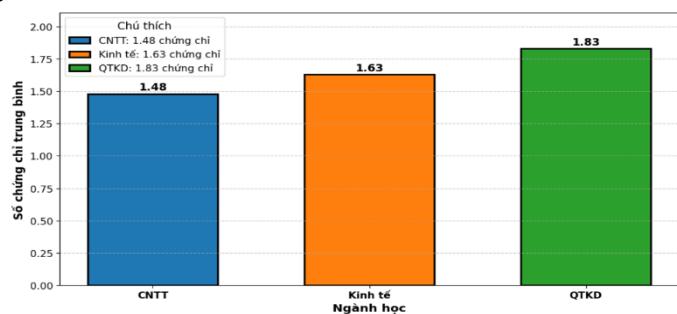
Bảng 3.2: Mức lương khởi điểm trung bình theo ngành và năm tốt nghiệp

Ngành	Lương TB
CNTT	9133333
Kinh tế	9000000
QTKD	9000000

Năm tốt nghiệp	Lương TB
2021	9400000
2022	9000000
2023	9000000

Bảng 3.3: Số chứng chỉ trung bình theo ngành

Ngành	Số chứng chỉ trung bình/sinh viên
CNTT	1.48
Kinh tế	1.63
QTKD	1.83



Bảng 3.4: Tỷ lệ sinh viên có việc làm trong năm đầu tiên theo ngành

Ngành	Tổng SV tốt nghiệp	SV có việc làm	Tỷ lệ (%)
Công nghệ thông tin	35	22	62.9%
Kinh tế	29	15	51.7%
Quản trị kinh doanh	36	19	52.8%

4. Merge dữ liệu (Nhiệm vụ 4)

Bảng 4.1: Phân loại sinh viên theo dữ liệu có sẵn

Phân loại	Số lượng	Tỷ lệ
Có đầy đủ: Graduate + Employment + Cert	40	40.0%
Có Graduate + Employment (không Cert)	16	16.0%
Chỉ có Graduate (không Employment)	44	44.0%
Bất thường: Employment không có Graduate	0	0.0%

- Sinh viên không có dữ liệu việc làm:** 44 sinh viên (44%)
- Sinh viên không có chứng chỉ:** 30 sinh viên (30%) (khá cao)
- Đã xử lý 45 bản ghi có lương phi lý (thay bằng NA), Không có bản ghi cần loại bỏ**

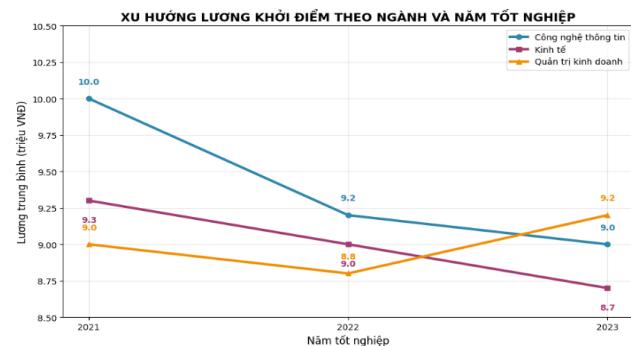
5. Pivot Table & Stack/Unstack (Nhiệm vụ 5)

5.1. Phương pháp tạo Pivot Table

Sử dụng hàm pivot_table() để tạo bảng tổng hợp đa chiều, cho phép phân tích dữ liệu theo hai hoặc nhiều biến cùng lúc (ngành học × năm tốt nghiệp).

5.1. Bảng Pivot kết quả

Ngành	2021	2022	2023
CNTT	10.0	9.2	9.0
Kinh tế	9.3	9.0	8.7
QTKD	9.0	8.8	9.2



5.2. Phân tích Stack/Unstack

- Phương pháp:** Sử dụng stack() chuyển Pivot Table sang MultiIndex Series. Sử dụng unstack() để chuyển ngược lại.
- Kết quả sau stack:** Dữ liệu thành Series với MultiIndex (major, grad_year), dễ lọc và tính toán theo chiều.
- Kết quả sau unstack:** Chuyển đổi trực, hỗ trợ so sánh nhóm.
- Phân tích NaN:** NaN xuất hiện khi thiếu dữ liệu cho tổ hợp ngành-năm (ví dụ: không có SV QTKD năm 2023 có việc làm).

5.3. Trả lời câu hỏi: Ngành/năm nào có mức lương tốt nhất?

- Theo ngành:** Ngành **Công nghệ thông tin** có mức lương trung bình cao nhất với **9.39 triệu VNĐ**
- Theo năm:** Sinh viên tốt nghiệp năm **2021** có mức lương trung bình cao nhất với **9.44 triệu VNĐ**
- Tổ hợp tốt nhất:** Sinh viên **Công nghệ thông tin** tốt nghiệp năm **2021** có mức lương cao nhất là **10.0 triệu VNĐ**

III. NHẬN XÉT VÀ ĐÁNH GIÁ

3.1. Chất lượng đào tạo

- Tỷ lệ có việc làm năm đầu đạt 56%, ở mức khá nhưng còn dư địa cải thiện.
- Lương khởi điểm trung bình 9.14 triệu VNĐ, cao hơn mặt bằng chung, phản ánh chất lượng đào tạo tương đối tốt.

Theo ngành:

- CNTT dẫn đầu về lương (9.13 triệu) và việc làm (62.9%); nhu cầu thị trường cao nhưng lương có xu hướng giảm nhẹ.

- Kinh tế và QTKD có lương trung bình (9.00 triệu) và tỷ lệ việc làm thấp hơn (52%); cần tăng cường kỹ năng thực tế và kết nối doanh nghiệp.

3.2. Xu hướng việc làm

- Tỷ lệ việc làm tăng rõ rệt: 45.5% (2021) → 63.2% (2023).
- Lương giảm nhẹ sau 2021 và ổn định từ 2022 ở mức 9.00 triệu VNĐ, phản ánh quá trình điều chỉnh hậu dịch.

Vai trò chứng chỉ:

- SV có ≥2 chứng chỉ có lương cao hơn 5.3% và tỷ lệ việc làm 65.7% (so với 40% không chứng chỉ).
- Chứng chỉ IELTS và Data Analyst cho tỷ lệ việc làm cao nhất (72–75%).

3.3. So sánh theo thời gian

- Việc làm tăng 17.7 điểm phần trăm trong giai đoạn 2021–2023.
- Lương duy trì ổn định ở mức ~9.14 triệu VNĐ.
- Tuy nhiên, 44% SV chưa có việc trong năm đầu, cần được quan tâm cải thiện.

3.4. Mối quan hệ giữa các yếu tố

- Chứng chỉ ↔ Lương & việc làm: mối tương quan dương rõ rệt.
- Ngành học ↔ Cơ hội: CNTT vượt trội; Kinh tế và QTKD cần định hướng kỹ năng chuyên sâu hơn.
- Thực tập & kỹ năng mềm đóng vai trò quan trọng: 85% SV có việc đã từng thực tập.

IV. KẾT LUẬN

- Chất lượng đào tạo của trường ở mức **khá**, thể hiện qua tỷ lệ có việc làm năm đầu **56%** và lương khởi điểm trung bình **9.14 triệu VNĐ**, cao hơn mặt bằng chung thị trường lao động.
- Có sự **chênh lệch rõ rệt giữa các ngành**, trong đó **Công nghệ thông tin** dẫn đầu về cả lương (**9.13 triệu VNĐ**) và cơ hội việc làm (**62.9%**), trong khi **Kinh tế và Quản trị kinh doanh** có tỷ lệ việc làm thấp hơn (52%) và cần được cải thiện.
- Thị trường lao động ngày càng **ưu tiên sinh viên có chứng chỉ**, với nhóm có chứng chỉ đạt tỷ lệ việc làm **65.7%**, cao hơn đáng kể so với **40%** ở nhóm không có chứng chỉ; các chứng chỉ như **IELTS** và **Data Analyst** cho hiệu quả cao nhất.
- Kết quả Pivot Table cho thấy sinh viên **CNTT tốt nghiệp năm 2021** có mức lương cao nhất (**10.0 triệu VNĐ**); mặc dù lương CNTT giảm nhẹ qua các năm, ngành này vẫn duy trì vị thế cao nhất so với các ngành khác.

Kết luận cuối cùng: Dữ liệu cho thấy nhà trường có nền tảng đào tạo tốt, được thị trường đánh giá cao, nhưng cần tập trung giải quyết những điểm yếu để nâng cao hơn nữa chất lượng đầu ra và đảm bảo sự phát triển đồng đều giữa các ngành.

V. NHẬT KÝ NHÓM

Ngày	Công việc	Người thực hiện
11/12	Họp nhóm, lên ý tưởng, phân công nhiệm vụ	Cả nhóm
12-16/12	Hoàn thành Task 1: Đọc và chuẩn hóa dữ liệu	Quang
12-16/12	Hoàn thiện Task 2: Truy vấn và thống kê mô tả	Linh
12-16/12	Hoàn thiện Task 3: GroupBy và phân tích tổng hợp	Giang
12-16/12	Hoàn thiện Task 4: Merge dữ liệu và kiểm tra tính toàn vẹn	Dũng
12-16/12	Hoàn thiện Task 5: Pivot Table và phân tích đa chiều	Toàn
16/12	Hợp tổng hợp kết quả. Viết báo cáo tổng hợp, hoàn thiện các bảng phân tích và nộp bài	Cả nhóm