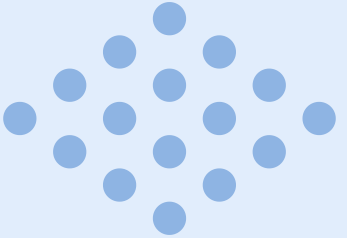


BÁO CÁO ĐỒ ÁN

**Phân tích Nghỉ việc và Hiệu suất
của nhân viên**



Nội dung



**Thu thập
dữ liệu**

**Đặt những câu hỏi
có ý nghĩa**

Phản hồi

**Khám phá
dữ liệu**

**Mô hình hóa
dữ liệu**

Nguồn tham khảo



Thu Thập Dữ Liệu

Giới thiệu

Bộ dữ liệu IBM HR Analytics Employee Attrition & Performance là kết quả của một cuộc khảo sát được thực hiện bởi các nhà khoa học dữ liệu của IBM trên 1470 tình nguyện viên. Bộ dữ liệu này chứa thông tin chi tiết về các tình nguyện viên tham gia, bao gồm các yếu tố nhân khẩu học, thông tin liên quan đến công việc, kết quả khảo sát về mức độ hài lòng, và tình trạng nghỉ việc (attrition).

Mục tiêu

Phục vụ cho việc phân tích nhân sự, giúp hiểu rõ hơn về các yếu tố ảnh hưởng đến quyết định nghỉ việc của nhân viên và đánh giá hiệu suất làm việc.

Nguồn thu thập dữ liệu & Chứng chỉ

kaggle

Chứng chỉ: Database: Open Database,
Contents: Database Contents



Khám Phá Dữ Liệu

Bao nhiêu dòng, cột

Bộ dữ liệu này có 1470 dòng và 35 cột

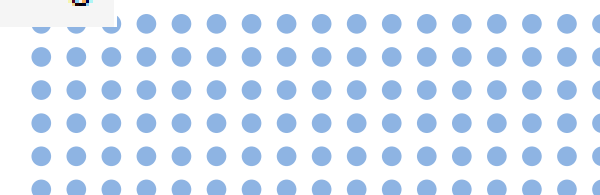
Ý nghĩa của mỗi dòng

Mỗi dòng trong dữ liệu đại diện cho thông tin chi tiết của một nhân viên trong công ty, bao gồm các yếu tố như mức độ hài lòng, thu nhập, tình trạng công tác, và nhiều thông tin liên quan đến công việc của họ.

Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber
41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1

EmployeeNumber	...	RelationshipSatisfaction	StandardHours	StockOptionLevel	TotalWorkingYears	TrainingTimesLastYear
1	...	1	80	0	8	0

WorkLifeBalance	YearsAtCompany	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
1	6	4	0	5

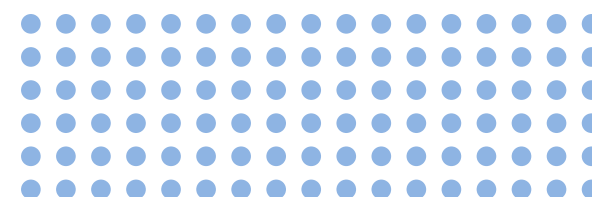


Khám Phá Dữ Liệu

Sơ lược về các cột trong dữ liệu

Bộ dữ liệu có tổng cộng 35 cột đặc trưng, gồm các nhóm cột như sau:

1. Thông tin chi tiết nhân viên: Age, Gender, Over18, MaritalStatus, Education, EducationField, NumCompaniesWorked, TotalWorkingYears
2. Thông tin về công việc: Attrition, Department, JobRole, JobLevel, BusinessTravel, OverTime, DistanceFromHome, YearsAtCompany, YearsInCurrentRole, YearsSinceLastPromotion, YearsWithCurrManager, TrainingTimesLastYear, EmployeeCount, EmployeeNumber, StandardHours
3. Lương thưởng và đãi ngộ: MonthlyIncome, DailyRate, HourlyRate, MonthlyRate, PercentSalaryHike, StockOptionLevel
4. Đánh giá và sự hài lòng: PerformanceRating, JobInvolvement, JobSatisfaction, EnvironmentSatisfaction, RelationshipSatisfaction, WorkLifeBalance



Khám Phá Dữ Liệu

Kiểm tra trùng lặp dòng

```
num_duplicates = ibm_df.duplicated().sum()  
print(f"Số dòng trùng lặp: {num_duplicates}")
```

Số dòng trùng lặp: 0

Không tồn tại trùng lặp dòng

Loại bỏ các cột không cần thiết

Loại bỏ 3 cột đặc trưng trong bộ dữ liệu chỉ gồm 1 giá trị duy nhất: "Over18", "EmployeeNumber", "EmployeeCount"

Missing value

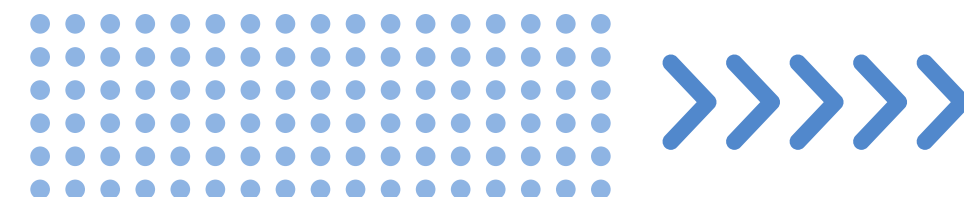
```
missing_values = ibm_df.isnull().sum().sum()  
missing_values
```

0

Không tồn tại missing value trong bộ dữ liệu

Kiểu dữ liệu

Ép kiểu các cột object và các cột định tính thành category



Đặt Những Câu Hỏi Có Ý Nghĩa

- > Mục tiêu của câu hỏi
- > Các bước phân tích
- > Kết luận





Câu hỏi 1: Mức thu nhập có thực sự là yếu tố quan trọng dẫn đến quyết định nghỉ việc không?

Mục tiêu:

- Giúp tổ chức hiểu rõ hơn về vai trò của thu nhập trong việc giữ chân nhân viên, từ đó cải thiện chính sách lương thưởng và điều kiện làm việc.
- Giúp tổ chức hiểu sâu hơn về các yếu tố ảnh hưởng đến sự gắn bó của nhân viên, từ đó xây dựng chiến lược nhân sự hiệu quả hơn.

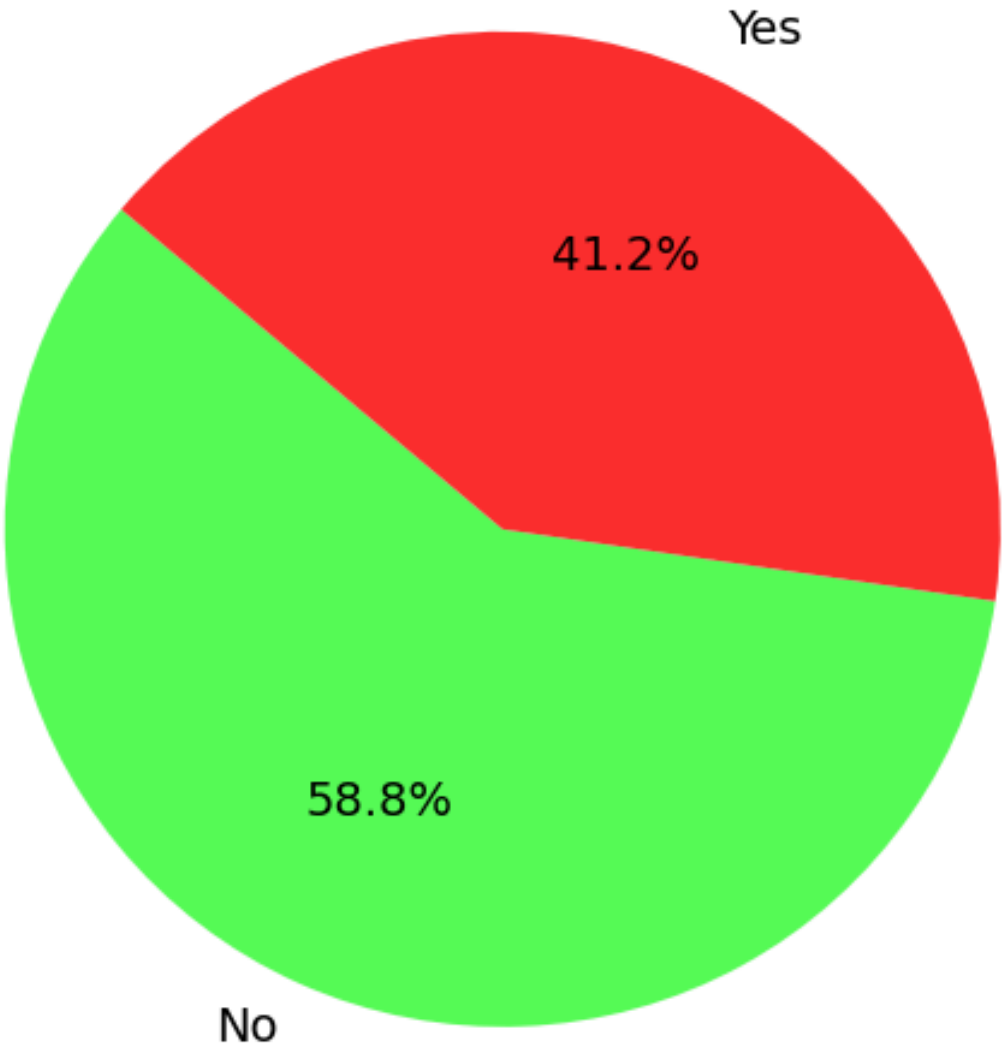
Các bước phân tích:

- Trung bình thu nhập theo trạng thái nghỉ việc.
- Trung bình thu nhập hàng tháng giữa nhóm nhân viên đã nghỉ việc và nhóm nhân viên không nghỉ việc theo từng mức độ hài lòng trong công việc.
- Chênh lệch về thu nhập trung bình hàng tháng giữa nhóm nhân viên đã nghỉ việc và nhóm nhân viên không nghỉ việc theo từng vai trò công việc.



Câu hỏi 1: Mức thu nhập có thực sự là yếu tố quan trọng dẫn đến quyết định nghỉ việc không?

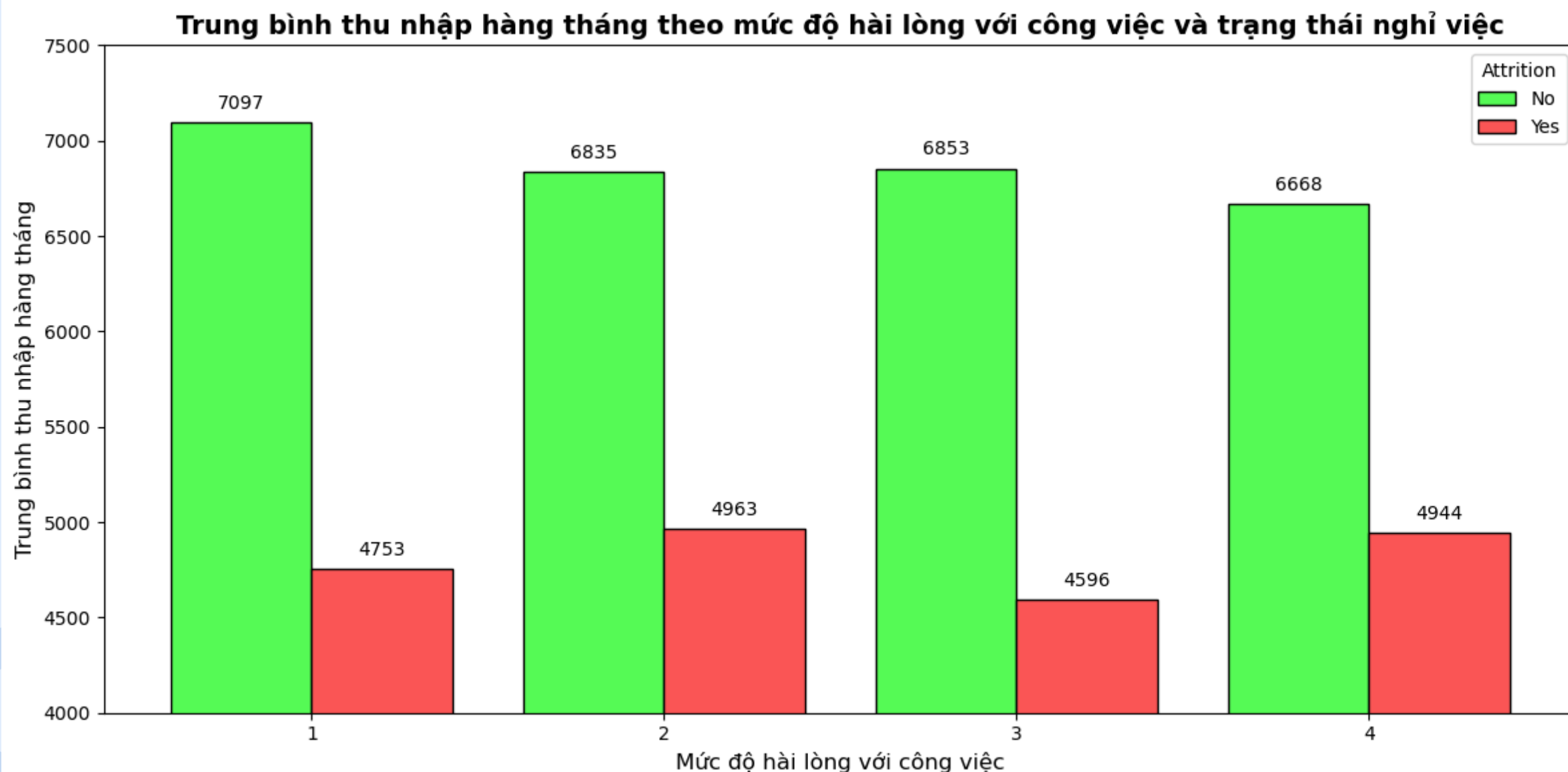
Thu nhập trung bình theo trạng thái nghỉ việc



Thu nhập trung bình của nhân viên đã nghỉ việc thấp hơn đáng kể so với nhân viên không nghỉ việc.



Câu hỏi 1: Mức thu nhập có thực sự là yếu tố quan trọng dẫn đến quyết định nghỉ việc không?

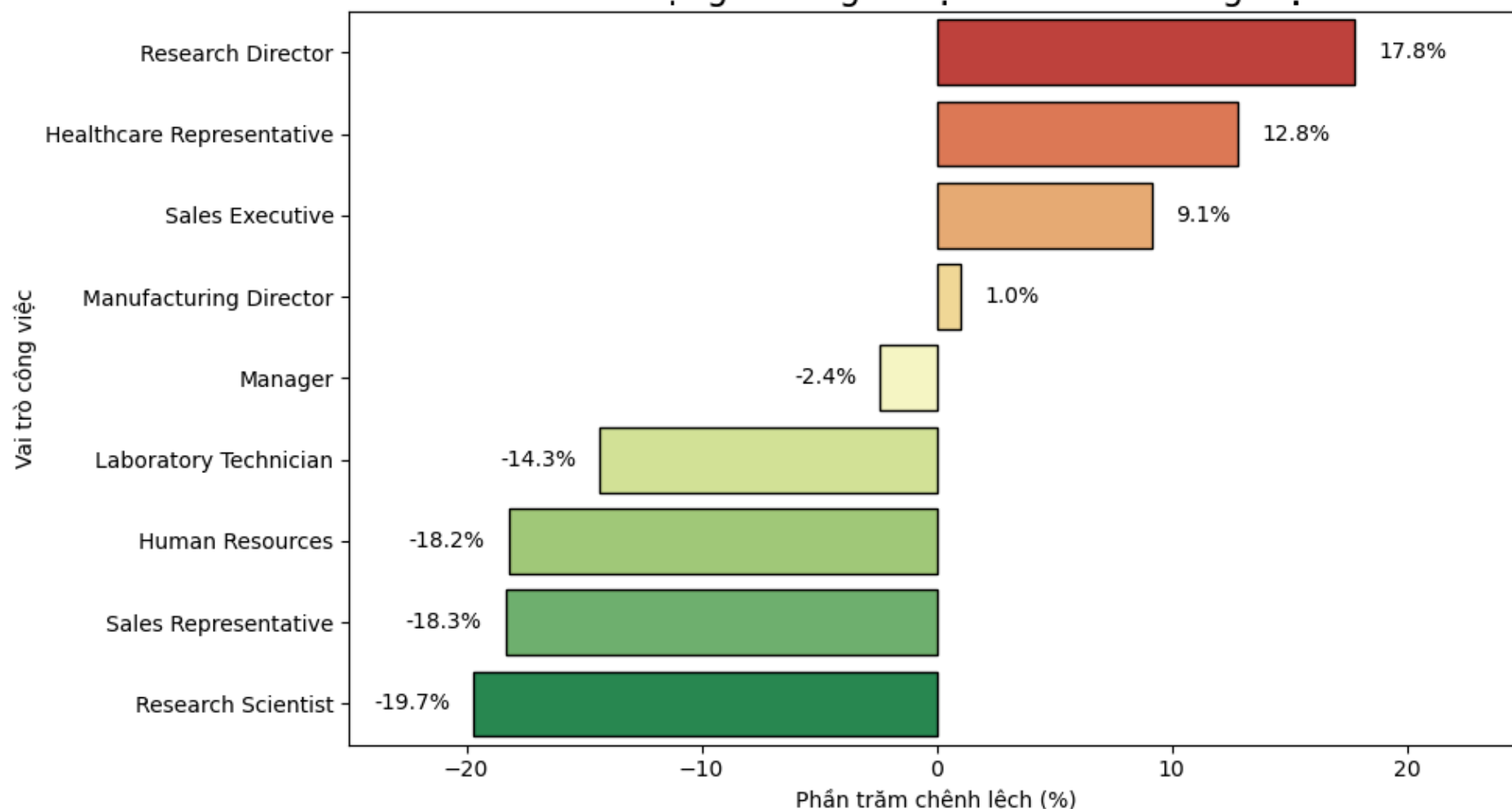


- Thu nhập trung bình của nhóm chưa nghỉ việc luôn cao hơn nhóm đã nghỉ việc ở tất cả các mức độ hài lòng công việc.
- Khi mức độ hài lòng công việc tăng lên, mức chênh lệch thu nhập giữa hai nhóm có xu hướng giảm xuống:
 - * Mức độ hài lòng 1: Chênh lệch thu nhập lớn nhất. Nhóm chưa nghỉ việc có thu nhập trung bình cao hơn đáng kể so với nhóm đã nghỉ việc.
 - * Mức độ hài lòng 2 và 3: Chênh lệch thu nhập vẫn tồn tại nhưng đã thu hẹp hơn so với mức 1.
 - * Mức độ hài lòng 4: Chênh lệch thu nhập là nhỏ nhất.



Câu hỏi 1: Mức thu nhập có thực sự là yếu tố quan trọng dẫn đến quyết định nghỉ việc không?

Phần trăm chênh lệch về thu nhập trung bình hàng tháng theo trạng thái nghỉ việc và vai trò công việc



- Nhóm có thu nhập cao hơn nghỉ việc nhiều hơn: Ở các vị trí Research Director (17.8%), Healthcare Representative (12.8%), và Sales Executive (9.1%), những người nghỉ việc có thu nhập trung bình cao hơn đáng kể so với những người ở lại.
- Nhóm có thu nhập thấp hơn nghỉ việc nhiều hơn: Ở các vị trí Research Scientist (-19.7%), Sales Representative (-18.3%), Human Resources (-18.2%), Laboratory Technician (-14.3%), và Manager (-2.4%), những người nghỉ việc có thu nhập trung bình thấp hơn so với những người ở lại.



Câu hỏi 2: Môi trường làm việc ảnh hưởng như thế nào đến tỷ lệ nghỉ việc của nhân viên trong tổ chức?

Mục tiêu:

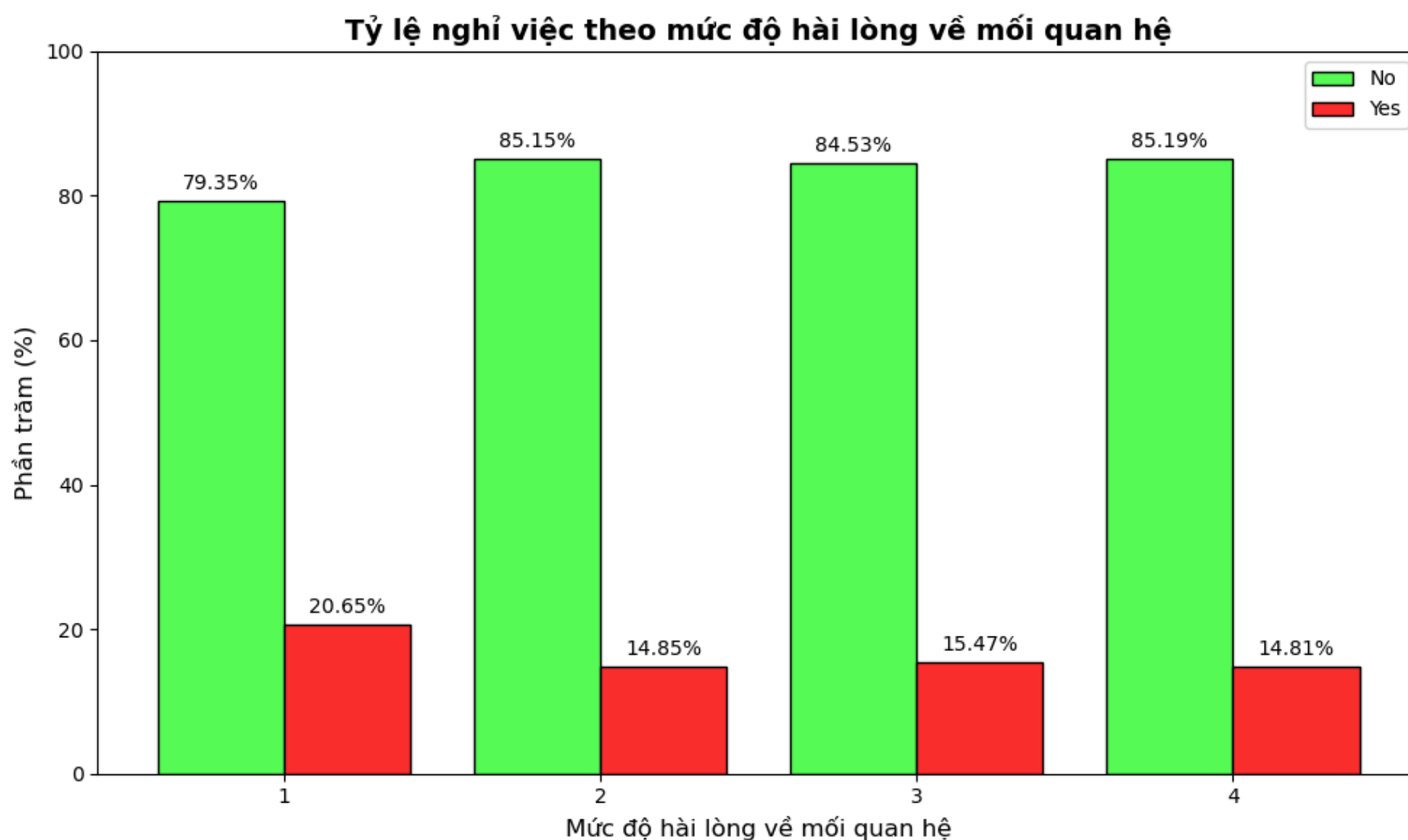
- Đánh giá các tác động của môi trường làm việc đến tỷ lệ nghỉ việc của nhân viên.
- Phân tích các yếu tố trong môi trường làm việc cần cải thiện để nâng cao hiệu quả công việc.

Các bước phân tích:

- Tỷ lệ nghỉ việc theo mức độ hài lòng về các mối quan hệ tại nơi làm việc.
- Tỷ lệ nghỉ việc theo từng vai trò công việc.
- Mức độ hài lòng với môi trường làm việc theo từng vai trò công việc.



Câu hỏi 2: Môi trường làm việc ảnh hưởng như thế nào đến tỷ lệ nghỉ việc của nhân viên trong tổ chức?

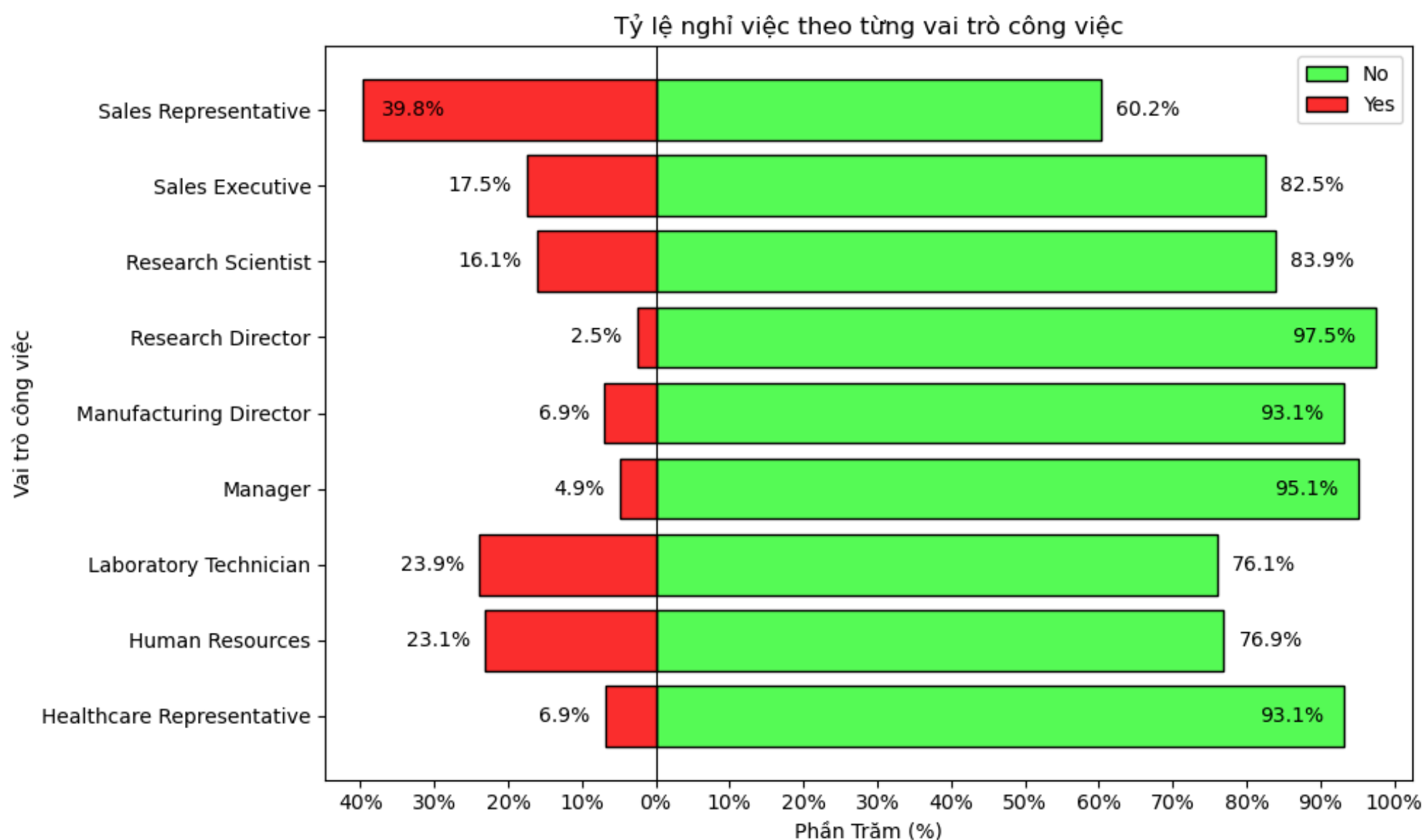


Tỷ lệ nghỉ việc (Yes) thấp hơn ở những nhóm có mức độ hài lòng cao hơn:

- Nhóm có mức độ hài lòng về mối quan hệ là 1 (thấp nhất) có tỷ lệ nghỉ việc cao nhất (20.65%).
- Nhóm có mức độ hài lòng về mối quan hệ là 2, 3 và 4 (cao hơn) có tỷ lệ nghỉ việc thấp hơn và tương đương nhau (lần lượt là 14.85%, 15.47%, và 14.81%).



Câu hỏi 2: Môi trường làm việc ảnh hưởng như thế nào đến tỷ lệ nghỉ việc của nhân viên trong tổ chức?



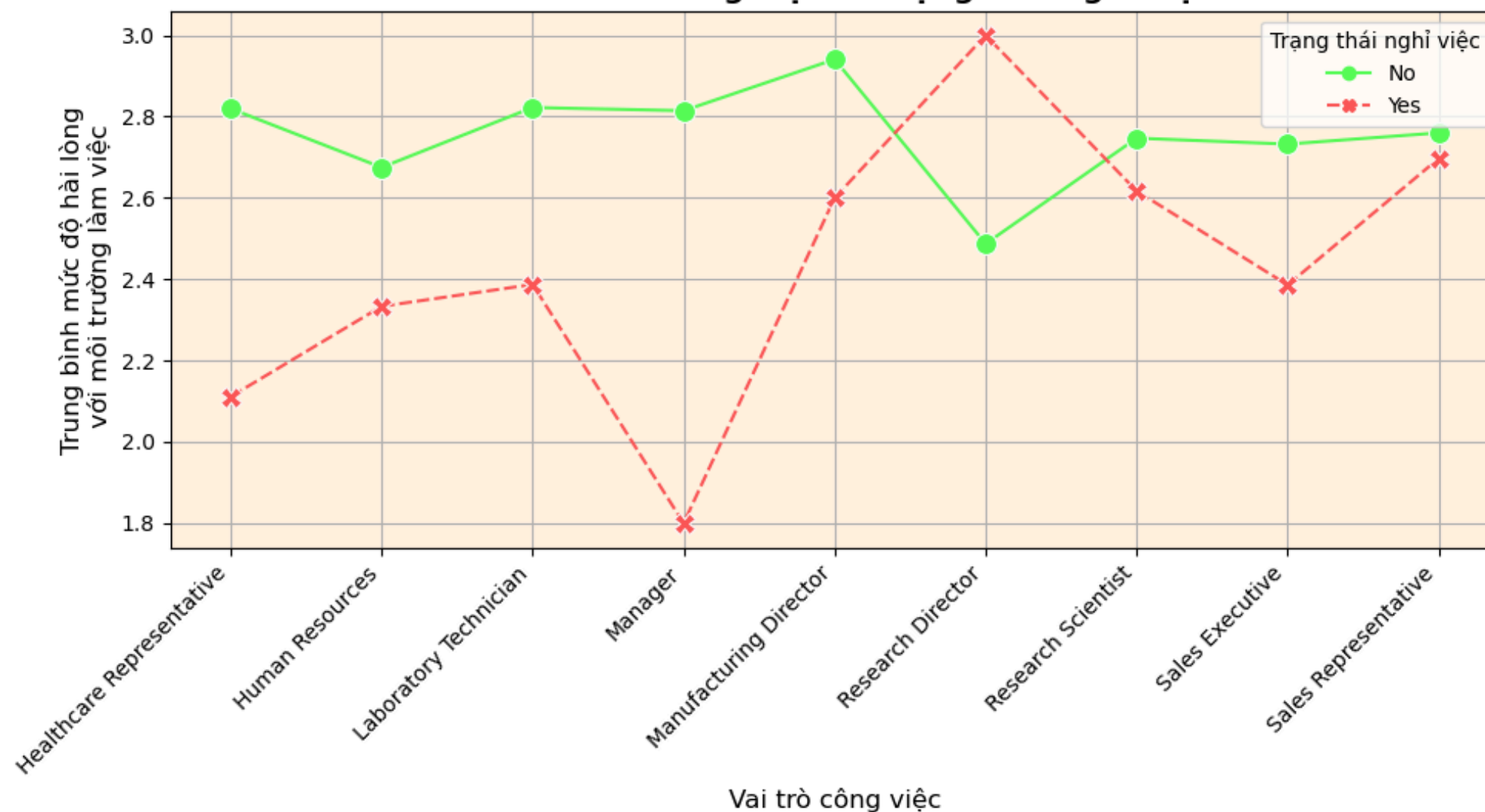
Biểu đồ tỷ lệ nghỉ việc theo từng vai trò công việc cho thấy sự khác biệt lớn về tỷ lệ nghỉ việc giữa các vị trí. Các vị trí chịu áp lực cao như Sales Representative và Sales Executive có tỷ lệ nghỉ việc cao, lần lượt là 39.8% và 17.5%, có thể do áp lực công việc lớn, mức độ hài lòng thấp với công việc.

Ngược lại, các vị trí quản lý như Manager, Healthcare Representative và Manufacturing Director có tỷ lệ nghỉ việc thấp hơn, đều dưới 7%, cho thấy tổ chức thực hiện quản lý nhân sự, giữ chân nhân tài tốt với các vai trò này.



Câu hỏi 2: Môi trường làm việc ảnh hưởng như thế nào đến tỷ lệ nghỉ việc của nhân viên trong tổ chức?

Mức độ hài lòng với môi trường làm việc theo vai trò công việc và trạng thái nghỉ việc



Biểu đồ mức độ hài lòng với môi trường làm việc cho thấy sự tương quan đáng kể giữa mức độ hài lòng và trạng thái nghỉ việc. Nhìn chung, nhân viên không nghỉ việc có mức độ hài lòng cao hơn đáng kể so với nhóm nghỉ việc ở hầu hết các vai trò, ngoại trừ Research Director.

Tuy nhiên, mức độ tác động của sự hài lòng với môi trường làm việc lên quyết định nghỉ việc có sự khác biệt giữa các vai trò. Ở vị trí Manager, mức độ hài lòng thấp (1.8 trên thang 4 ở nhóm nghỉ việc) có thể xuất phát từ áp lực công việc cao.

Đối với Sales Representative và Sales Executive, dù mức độ hài lòng của nhóm nghỉ việc không thấp hơn nhiều so với các vai trò khác, nhưng tỷ lệ nghỉ việc vẫn cao.



Câu hỏi 3: Trong các nhóm nhân viên được phân loại theo trình độ học vấn (Education Level), mức độ đào tạo (TrainingTimesLastYear) và các cơ hội phát triển (StockOptionLevel) được phân bổ như thế nào, và những yếu tố này có ảnh hưởng ra sao đến tỷ lệ nghỉ việc của công ty?

Mục tiêu:

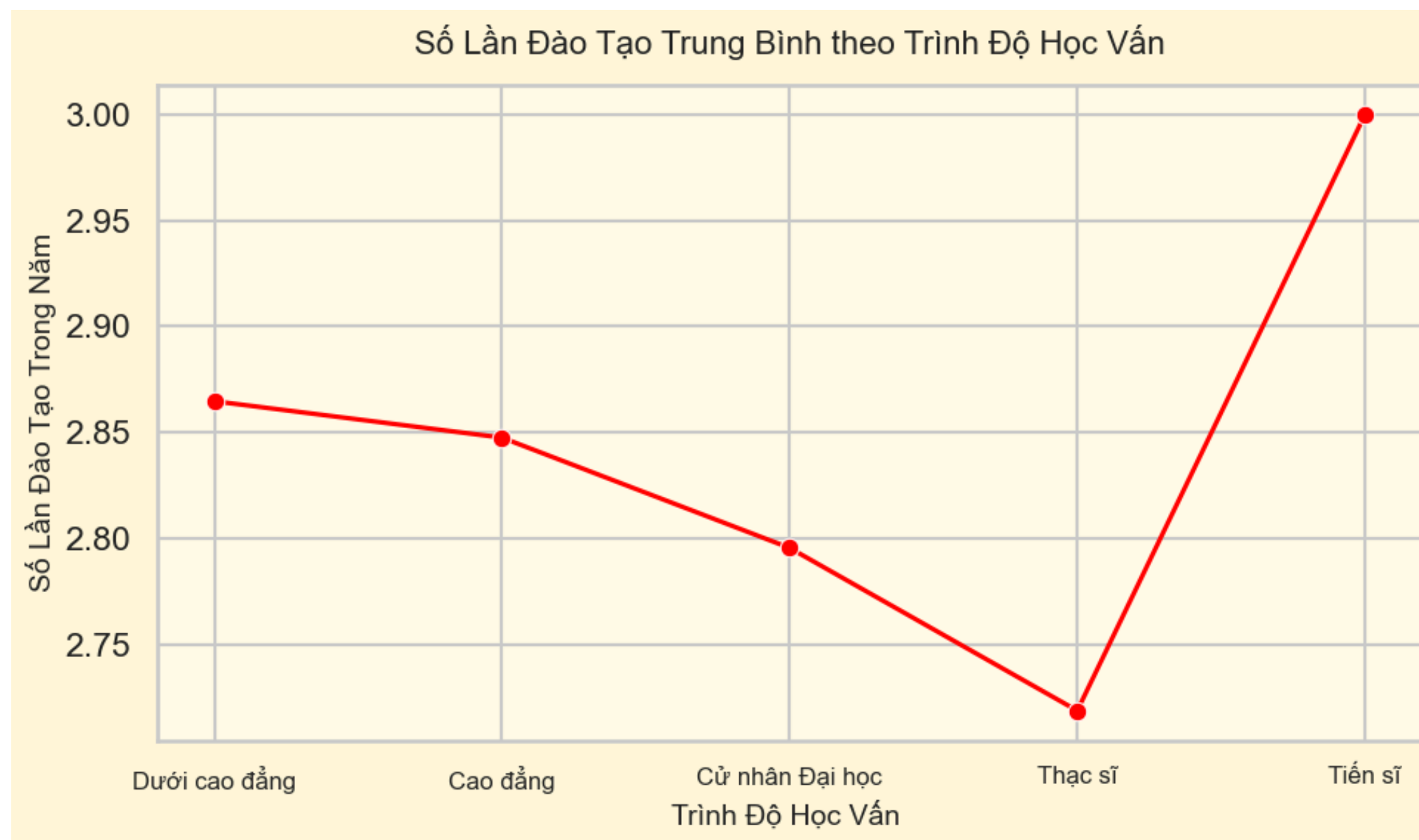
- Xác định mối quan hệ giữa trình độ học vấn và số lần đào tạo, quyền chọn cổ phiếu.
- Đánh giá tác động của việc đào tạo và cơ hội phát triển đến số năm gắn bó với công ty.
- Tìm hiểu liệu công ty có đang tập trung đào tạo và phát triển cho một số nhóm trình độ học vấn nhất định, từ đó đề xuất chiến lược phát triển nhân sự.

Các bước phân tích:

- Phân tích số lần tham gia đào tạo trong năm theo trình độ học vấn
- Phân tích mức quyền chọn cổ phiếu theo trình độ học vấn
- Phân tích tỷ lệ nghỉ việc dựa theo trình độ



Câu hỏi 3: Trong các nhóm nhân viên được phân loại theo trình độ học vấn (Education Level), mức độ đào tạo (TrainingTimesLastYear) và các cơ hội phát triển (StockOptionLevel) được phân bổ như thế nào, và những yếu tố này có ảnh hưởng ra sao đến tỷ lệ nghỉ việc của công ty?



1. Xu hướng giảm dần:

- Số lần đào tạo trung bình trong năm giảm dần từ nhóm Dưới Cao đẳng (2.86) đến nhóm Thạc sĩ (2.73).

2. Sự gia tăng đột ngột:

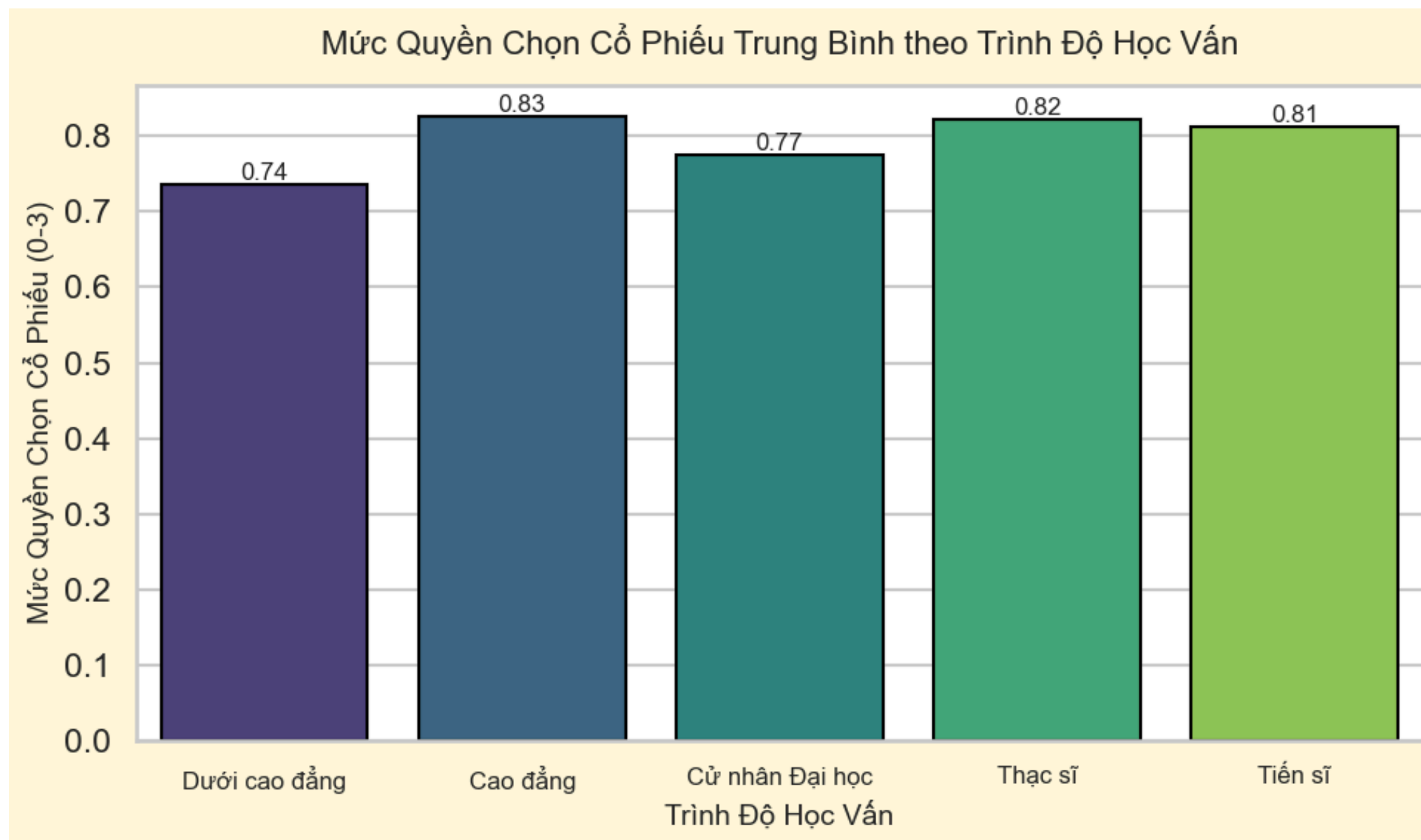
- Ở nhóm Tiến sĩ, số lần đào tạo trung bình tăng mạnh lên mức cao nhất, đạt 3.00.

3. Kết luận:

- Nhân viên có trình độ Tiến sĩ được đào tạo nhiều nhất.
- Các nhóm từ Cao đẳng đến Thạc sĩ có ít cơ hội đào tạo hơn, cho thấy sự giảm đầu tư vào đào tạo khi trình độ học vấn tăng, ngoại trừ nhóm Tiến sĩ.



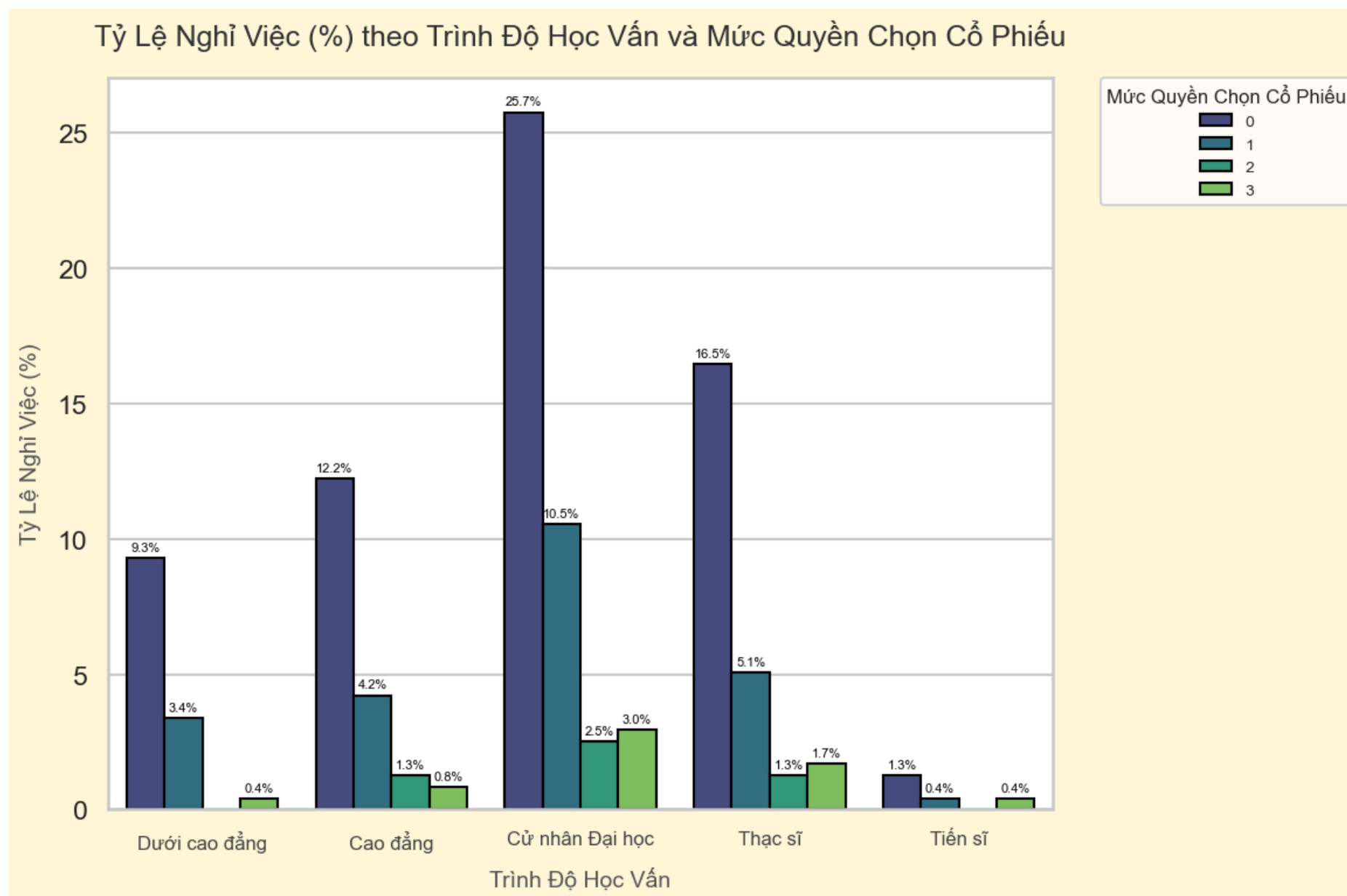
Câu hỏi 3: Trong các nhóm nhân viên được phân loại theo trình độ học vấn (Education Level), mức độ đào tạo (TrainingTimesLastYear) và các cơ hội phát triển (StockOptionLevel) được phân bổ như thế nào, và những yếu tố này có ảnh hưởng ra sao đến tỷ lệ nghỉ việc của công ty?



1. Nhóm Cao đẳng có mức quyền chọn cổ phiếu trung bình cao nhất (0.83).
2. Mức quyền chọn cổ phiếu trung bình của nhóm Thạc sĩ (0.82) và nhóm Tiến sĩ (0.81) gần như tương đương nhau.
3. Nhóm Dưới cao đẳng có mức quyền chọn cổ phiếu thấp nhất (0.74), cho thấy ít cơ hội hơn so với các nhóm còn lại.
4. Kết luận:
 - Mức quyền chọn cổ phiếu trung bình khá đồng đều giữa các nhóm trình độ học vấn, cho thấy chính sách phân bổ quyền chọn cổ phiếu không có sự khác biệt lớn dựa trên trình độ học vấn.



Câu hỏi 3: Trong các nhóm nhân viên được phân loại theo trình độ học vấn (Education Level), mức độ đào tạo (TrainingTimesLastYear) và các cơ hội phát triển (StockOptionLevel) được phân bổ như thế nào, và những yếu tố này có ảnh hưởng ra sao đến tỷ lệ nghỉ việc của công ty?



1. Tỷ lệ nghỉ việc cao nhất:

- Nhóm Cử nhân Đại học và mức quyền chọn cổ phiếu 0 có tỷ lệ nghỉ việc cao nhất (25.7%), cho thấy đây là nhóm dễ nghỉ việc nhất.

2. Tỷ lệ nghỉ việc giảm dần khi mức quyền chọn cổ phiếu tăng:

- Ở tất cả các trình độ học vấn, tỷ lệ nghỉ việc giảm rõ rệt với mức quyền chọn cổ phiếu 2 và 3, chứng tỏ quyền chọn cổ phiếu cao có tác động tích cực giữ chân nhân viên.

3. Nhóm có trình độ cao ít nghỉ việc:

- Nhóm Thạc sĩ và Tiến sĩ có tỷ lệ nghỉ việc thấp hơn đáng kể (dưới 5%) ở mọi mức quyền chọn cổ phiếu.

4. Mức quyền chọn cổ phiếu thấp ảnh hưởng lớn:

- Ở mọi trình độ, mức quyền chọn cổ phiếu 0 dẫn đến tỷ lệ nghỉ việc cao hơn đáng kể.



Câu hỏi 4: Liệu tình trạng làm thêm giờ và sự cân bằng công việc - cuộc sống trong các phòng ban có ảnh hưởng đến tỷ lệ nghỉ việc của nhân viên không?

Mục tiêu:

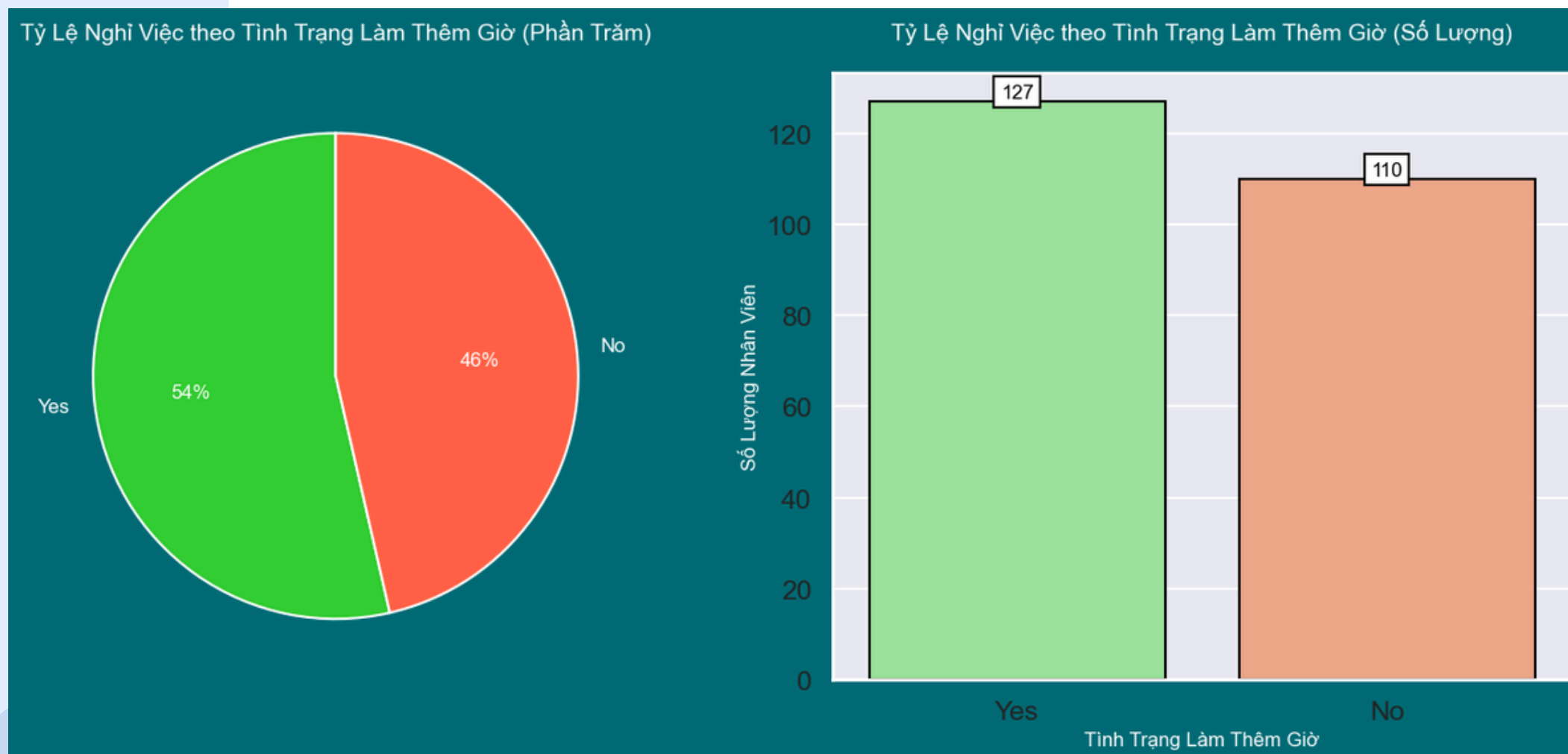
Đánh giá tác động của tình trạng làm thêm giờ (Overtime) và mức độ cân bằng công việc - cuộc sống (Work-Life Balance) trong các phòng ban đến tỷ lệ nghỉ việc của nhân viên. Từ đó, đề xuất giải pháp cải thiện môi trường làm việc, giảm áp lực và tăng sự gắn bó.

Các bước phân tích:

- **Phân tích tỷ lệ nghỉ việc theo tình trạng làm thêm giờ**
- **Phân tích số lượng nhân viên nghỉ việc dựa vào mức độ cân bằng công việc - cuộc sống của các phòng ban**



Câu hỏi 4: Liệu tình trạng làm thêm giờ và sự cân bằng công việc - cuộc sống trong các phòng ban có ảnh hưởng đến tỷ lệ nghỉ việc của nhân viên không?



- **Biểu đồ tròn (Pie Chart):**

- Tỷ lệ nghỉ việc của nhân viên làm thêm giờ (54%) cao hơn nhân viên không làm thêm giờ (46%), cho thấy làm thêm giờ có thể là một yếu tố ảnh hưởng đến việc nghỉ việc.
- Mặc dù chênh lệch không quá lớn, nhưng làm thêm giờ có thể gây áp lực và làm giảm sự hài lòng của nhân viên.

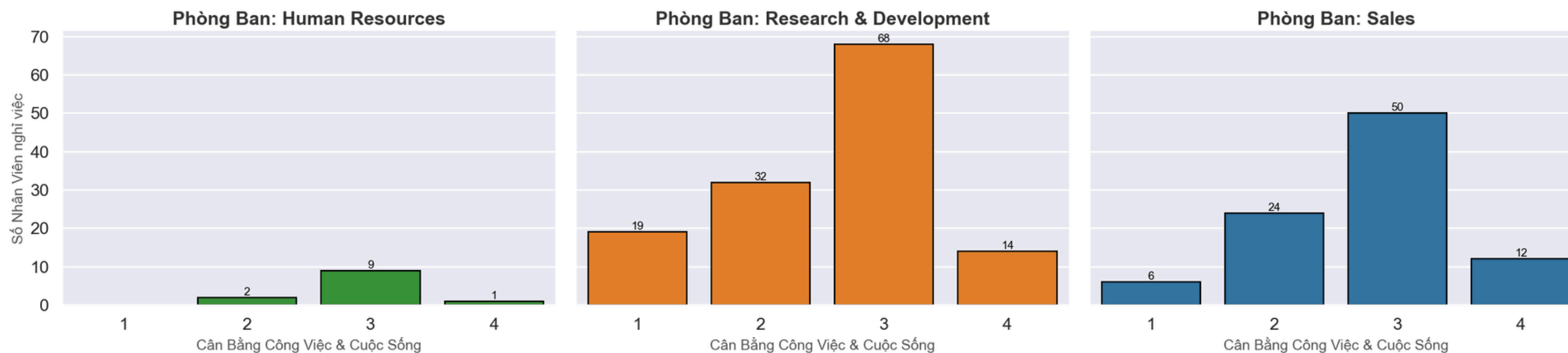
- **Biểu đồ cột (Bar Chart):**

- Nhân viên làm thêm giờ chiếm số lượng nghỉ việc lớn hơn (127 nhân viên) so với nhân viên không làm thêm giờ (110 nhân viên).
- Điều này cho thấy, việc làm thêm giờ có thể làm gia tăng nguy cơ nghỉ việc, đặc biệt đối với những nhân viên phải làm thêm giờ liên tục trong thời gian dài.



Câu hỏi 4: Liệu tình trạng làm thêm giờ và sự cân bằng công việc - cuộc sống trong các phòng ban có ảnh hưởng đến tỷ lệ nghỉ việc của nhân viên không?

Có Môi Trường Cân Bằng Công Việc và Cuộc Sống Không?





Câu hỏi 4: Liệu tình trạng làm thêm giờ và sự cân bằng công việc - cuộc sống trong các phòng ban có ảnh hưởng đến tỷ lệ nghỉ việc của nhân viên không?

- Phòng ban Human Resources:
 - Chủ yếu nghỉ việc ở mức 3 (9 nhân viên).
 - Một số ít ở mức 2 (2 nhân viên) và mức 4 (1 nhân viên).
 - → Kết luận: Số lượng nghỉ việc thấp nhưng vẫn tồn tại tình trạng mất cân bằng.
- Phòng ban Research & Development:
 - Nghỉ việc tập trung ở mức 3 (68 nhân viên) và mức 2 (32 nhân viên).
 - Mức 4 có 14 nhân viên.
 - → Kết luận: Phòng ban này có số lượng nghỉ việc cao nhất, áp lực công việc còn lớn.
- Phòng ban Sales:
 - Mức 3 chiếm đa số (50 nhân viên).
 - Mức 2 (24 nhân viên) và mức 4 (12 nhân viên).
 - → Kết luận: Có sự cân bằng tương đối nhưng tình trạng mất cân bằng vẫn đáng kể.
- Tổng Kết:
 - Mức độ cân bằng công việc trung bình (mức 3) chiếm ưu thế ở cả 3 phòng ban.
 - Research & Development có số lượng nghỉ việc cao nhất, đặc biệt ở mức mất cân bằng (mức 2).
 - Sales cần cải thiện để giảm tỷ lệ nghỉ việc do mất cân bằng công việc.
 - Human Resources có số nghỉ việc ít nhưng thiếu sự hài lòng cao về Work-Life Balance.
- Đề xuất: Tăng cường chính sách hỗ trợ Work-Life Balance để giảm áp lực và tỷ lệ nghỉ việc, đặc biệt tại các phòng Research & Development và Sales.



Câu hỏi 5: Khoảng cách từ nơi làm việc và tần suất đi công tác có phải là yếu tố chính dẫn đến tình trạng nghỉ việc hay không

Mục tiêu:

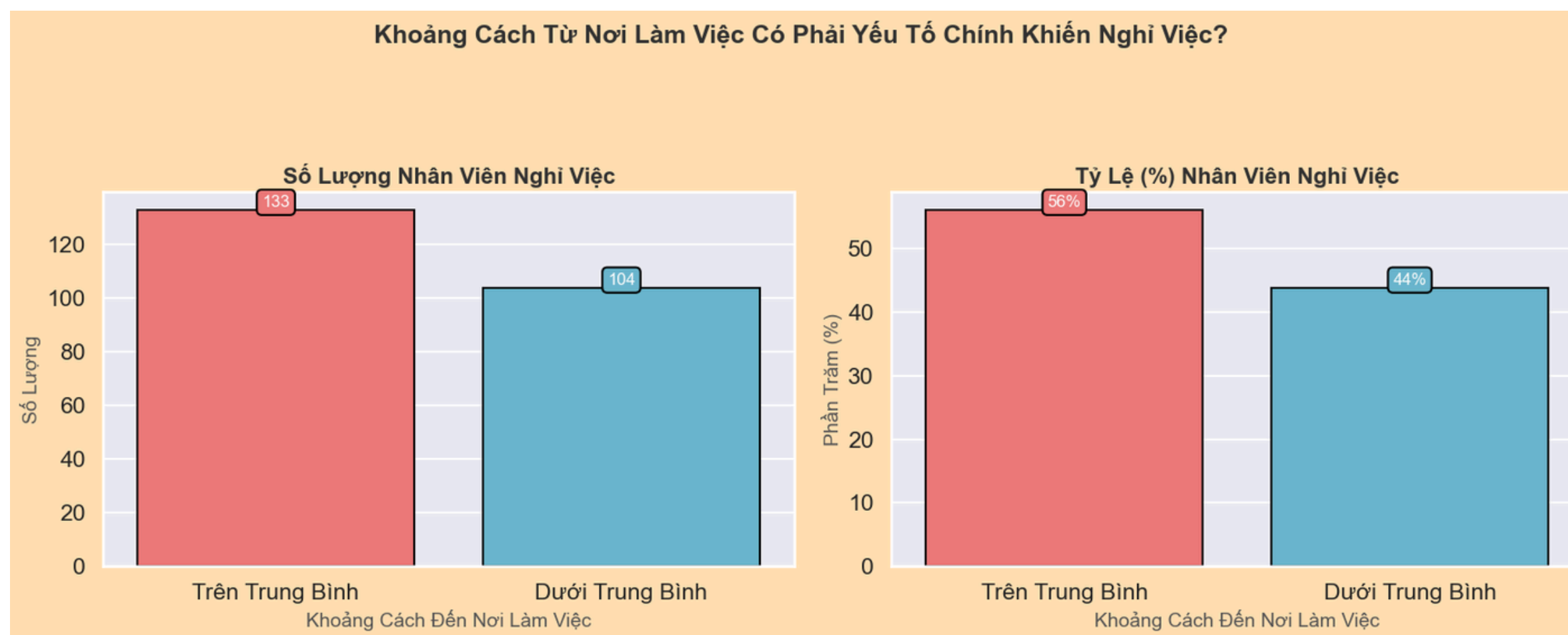
- Phân tích mối liên hệ giữa khoảng cách từ nơi làm việc và tần suất đi công tác với tình trạng nghỉ việc.
- Đánh giá xem các yếu tố trên có ảnh hưởng đáng kể đến quyết định nghỉ việc của nhân viên hay không để giảm tỷ lệ nghỉ việc.

Các bước phân tích:

- Phân tích tỷ lệ nghỉ việc theo khoảng cách
- Phân tích tỷ lệ nghỉ việc theo tần suất đi công tác



Câu hỏi 5: Khoảng cách từ nơi làm việc và tần suất đi công tác có phải là yếu tố chính dẫn đến tình trạng nghỉ việc hay không



- **Số lượng nhân viên nghỉ việc:**

- Những người sống trên trung bình về khoảng cách từ nơi làm việc có 133 người nghỉ việc, cao hơn so với 104 người ở dưới trung bình.

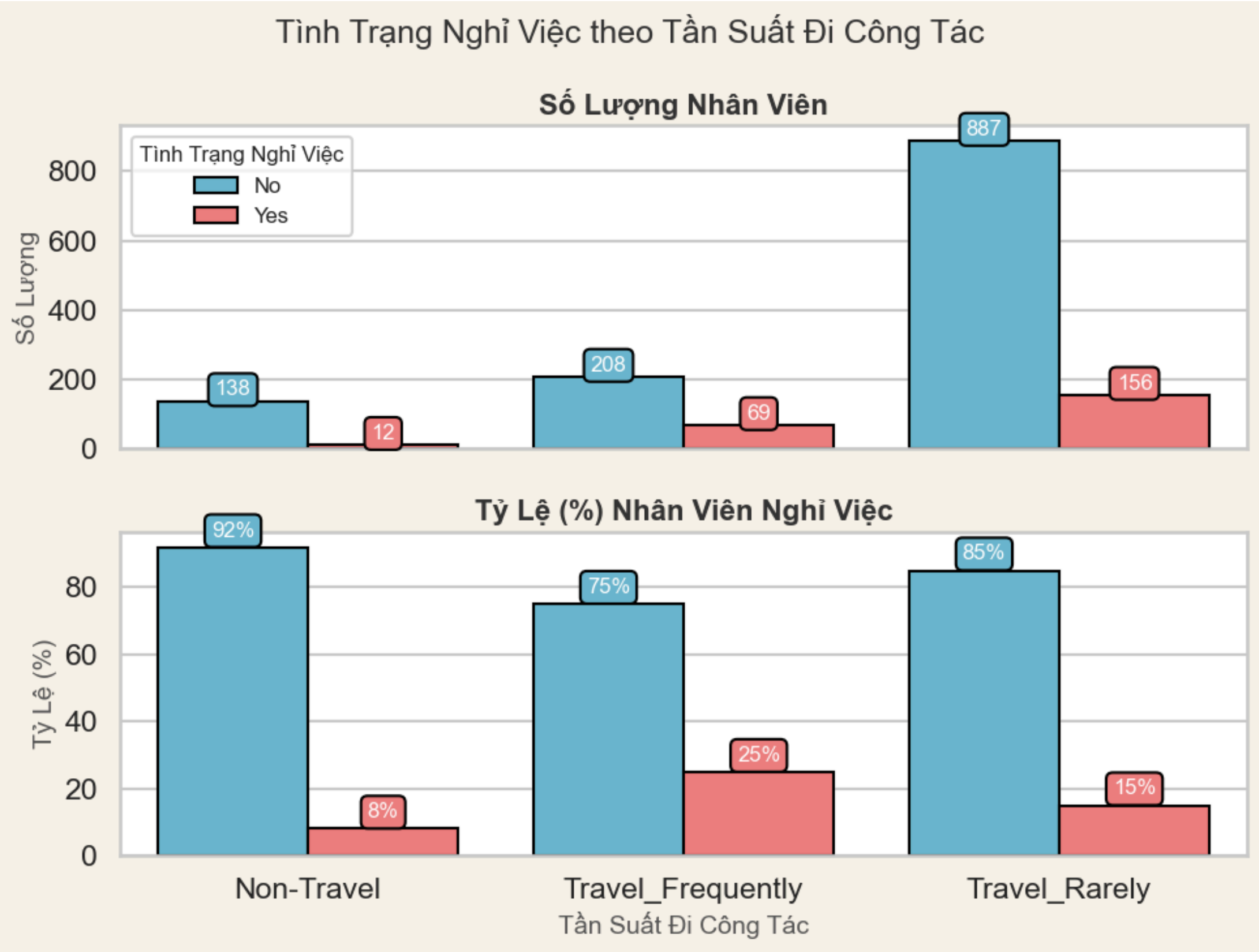
- **Tỷ lệ nhân viên nghỉ việc:**

- 56% nhân viên sống xa nơi làm việc trên trung bình nghỉ việc, cao hơn so với 44% nhân viên ở gần nơi làm việc dưới trung bình.

Kết luận: Khoảng cách đến nơi làm việc có vẻ là một yếu tố quan trọng, vì tỷ lệ nghỉ việc cao hơn ở nhóm có khoảng cách xa. Điều này có thể liên quan đến chi phí di chuyển, thời gian đi lại, và sự cân bằng giữa công việc và cuộc sống.



Câu hỏi 5: Khoảng cách từ nơi làm việc và tần suất đi công tác có phải là yếu tố chính dẫn đến tình trạng nghỉ việc hay không





Câu hỏi 5: Khoảng cách từ nơi làm việc và tần suất đi công tác có phải là yếu tố chính dẫn đến tình trạng nghỉ việc hay không

- Số lượng nhân viên nghỉ việc:
 - Nhóm không đi công tác (Non-Travel): Chỉ có 12 nhân viên nghỉ việc trong tổng số 150 người (tỷ lệ thấp).
 - Nhóm đi công tác thường xuyên (Travel Frequently): 69 nhân viên nghỉ việc (tỷ lệ nghỉ việc cao nhất).
 - Nhóm thỉnh thoảng đi công tác (Travel Rarely): 156 nhân viên nghỉ việc, chiếm số lượng cao nhất về số lượng tuyệt đối.
- Tỷ lệ nhân viên nghỉ việc:
 - Nhóm không đi công tác: Tỷ lệ nghỉ việc chỉ 8%, thấp hơn nhiều so với nhóm đi công tác thường xuyên (25%) và nhóm đi công tác ít (15%).

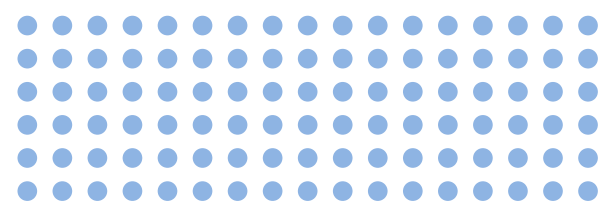
Kết luận: Tần suất đi công tác có tác động đáng kể đến tình trạng nghỉ việc. Nhân viên phải đi công tác thường xuyên hoặc thỉnh thoảng có xu hướng nghỉ việc cao hơn, có thể do áp lực công việc, mất cân bằng cuộc sống, hoặc ảnh hưởng từ việc di chuyển nhiều.

Tổng kết:

- Khoảng cách từ nơi làm việc và tần suất đi công tác đều có tác động đáng kể đến tình trạng nghỉ việc.
- Các công ty nên:
 - a. Xem xét hỗ trợ chi phí đi lại hoặc cung cấp chỗ ở gần cho nhân viên ở xa để giảm tỷ lệ nghỉ việc.
 - b. Đánh giá lại tần suất đi công tác, giảm bớt công tác không cần thiết hoặc hỗ trợ nhân viên trong việc sắp xếp công việc khi công tác.

Cả hai yếu tố này đều nhấn mạnh tầm quan trọng của việc duy trì môi trường làm việc thuận lợi và cân bằng để giảm tỷ lệ nghỉ việc.

**Phân nhóm nhân viên
dựa trên các đặc trưng
cá nhân và công việc để
tối ưu hóa quản lý nhân
sự**



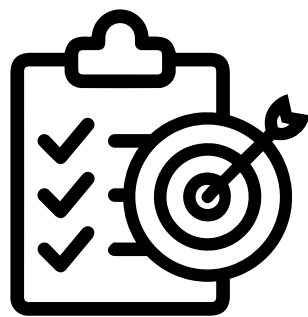
Các bước thực hiện >>>>>

- > Phát biểu vấn đề
- > Đề xuất phương pháp giải quyết
- > Chuẩn bị dữ liệu
- > Xác định số lượng cụm tối ưu
- > Phân tích kết quả các cụm
- > Kết luận



Phát biểu vấn đề

Mục tiêu



Phân nhóm nhân viên dựa trên các đặc điểm cá nhân, hiệu suất làm việc và sự hài lòng để tối ưu hóa chiến lược nhân sự

Lợi ích

- Tối ưu hóa đào tạo, thăng tiến và giữ chân nhân viên
- Giảm tỷ lệ nghỉ việc và cải thiện hiệu suất làm việc

Đề xuất phương pháp giải quyết

Để giải quyết bài toán phân nhóm nhân viên với các đặc điểm tương tự nhau và các nhóm được chia phải đồng nhất

Phương pháp

- Chọn ra những đặc trưng phù hợp, liên quan đến bài toán (tuổi, giới tính, hiệu suất, ...)
- Áp dụng mô hình K-means Clustering để tìm các nhóm nhân viên tương đồng

Tại sao lại là K-means

- Đơn giản, hiệu quả với dữ liệu định lượng
- Hỗ trợ xây dựng các chiến lược nhân sự dựa trên đặc điểm từng nhóm



Các bước chuẩn bị dữ liệu

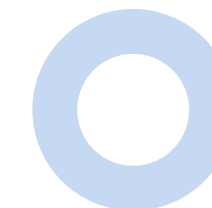
Dựa trên tập dữ liệu đã được xử lý trước đó để lựa chọn ra các đặc trưng thích hợp

Đặc trưng đầu vào

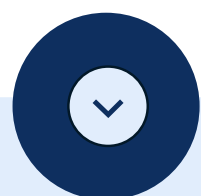
- Mang tính cá nhân
- Hiệu suất và sự hài lòng
- Trách nhiệm công việc
- Thâm niên và kinh nghiệm

Xử lý dữ liệu

- Mã hóa các đặc trưng phân loại
- Chuẩn hóa dữ liệu bằng MinMaxScaler



Xác định số lượng cụm tối ưu



Elbow method

Dựa trên biểu đồ quan hệ giữa số cụm k và tổng bình phương sai số (Inertia). Số cụm tối ưu là tại điểm "khuỷu", nơi mà việc tăng thêm số cụm không giảm đáng kể sai số.



Silhouette Score

Đánh giá chất lượng cụm dựa trên mức độ tách biệt giữa các cụm và sự gắn kết bên trong mỗi cụm. Giá trị k tối ưu là khi Silhouette Coefficient đạt giá trị cao nhất, cho thấy các cụm rõ ràng nhất.

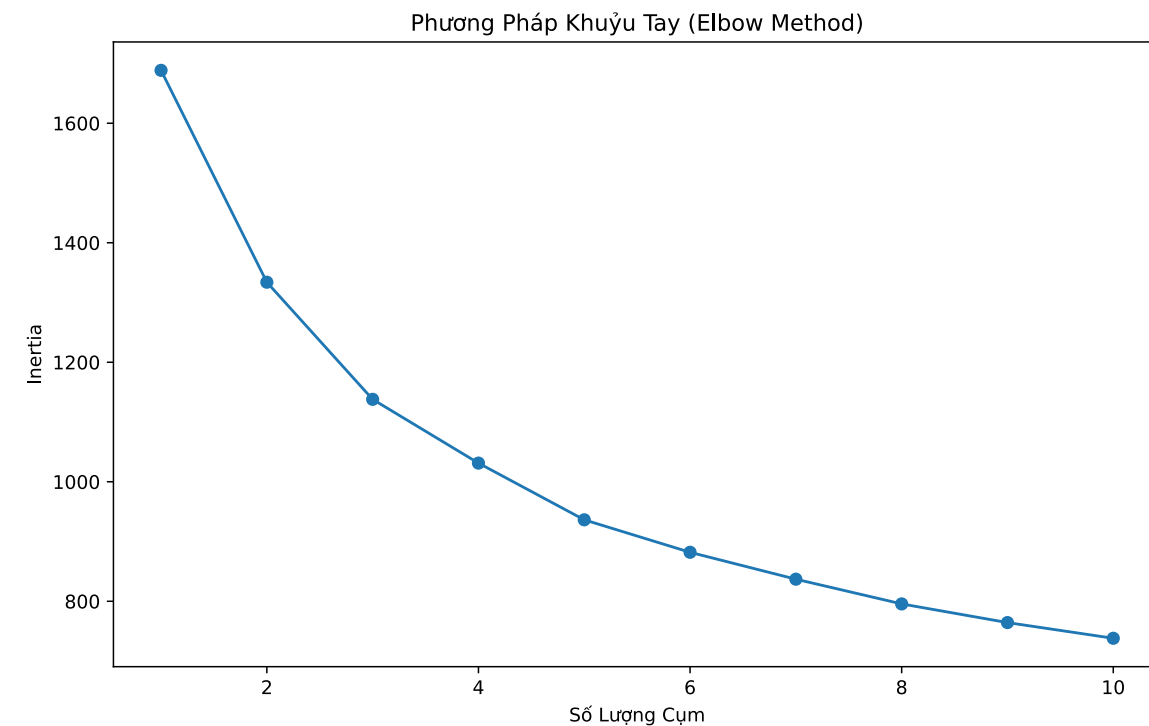


Gap statistic

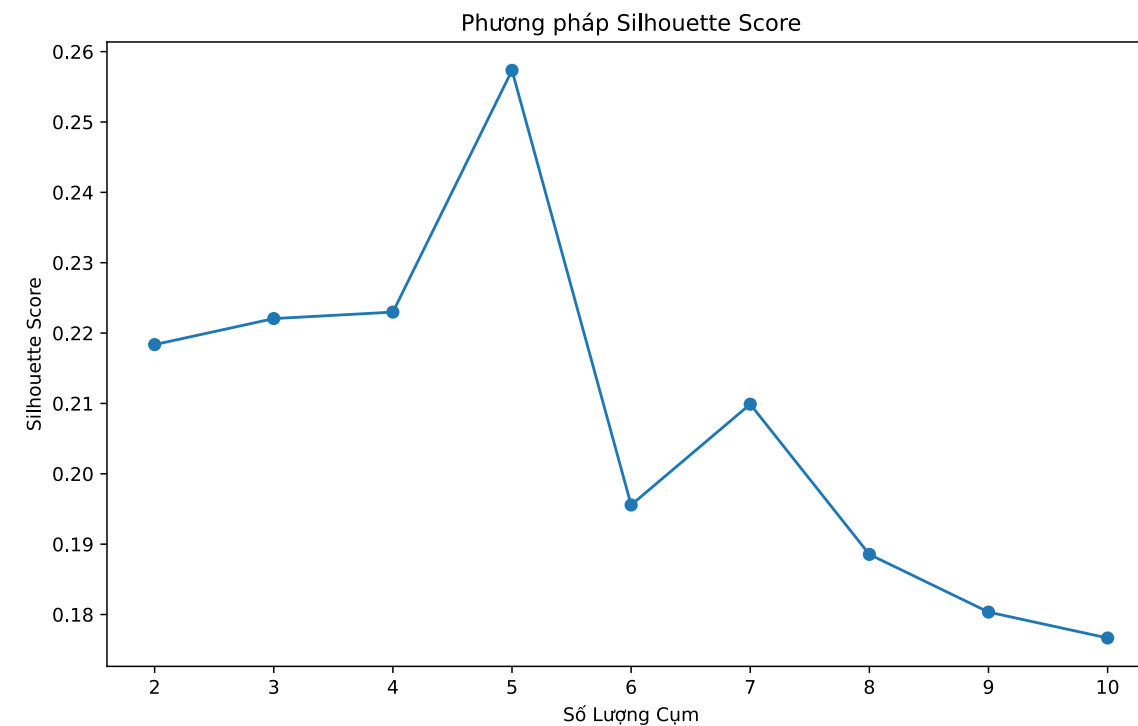
So sánh độ chặt chẽ của cụm trong dữ liệu thật với dữ liệu ngẫu nhiên trong cùng kích thước. Số cụm tối ưu là tại giá trị k có khoảng cách (gap) lớn nhất giữa dữ liệu thật và dữ liệu ngẫu nhiên.



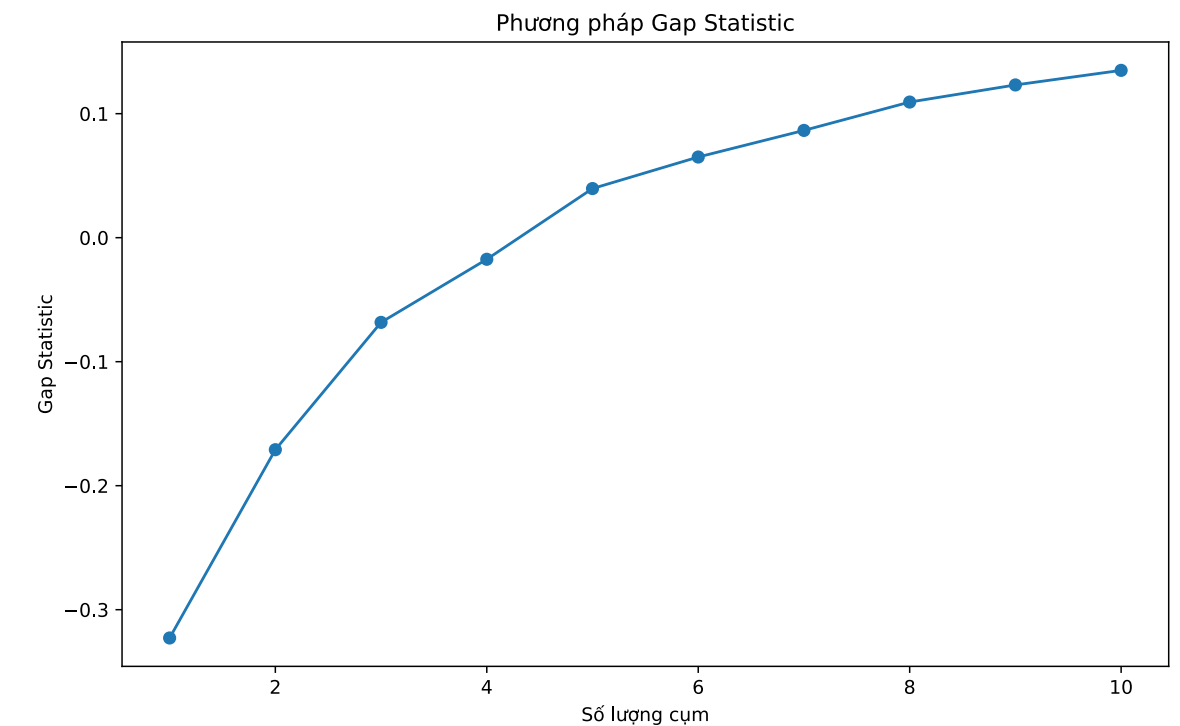
Xác định số lượng cụm tối ưu



Biểu đồ Elbow Method cho thấy có một "khuỷu tay" xảy ra ở khoảng 5-6 cụm, gợi ý rằng số lượng cụm tối ưu nằm trong khoảng này



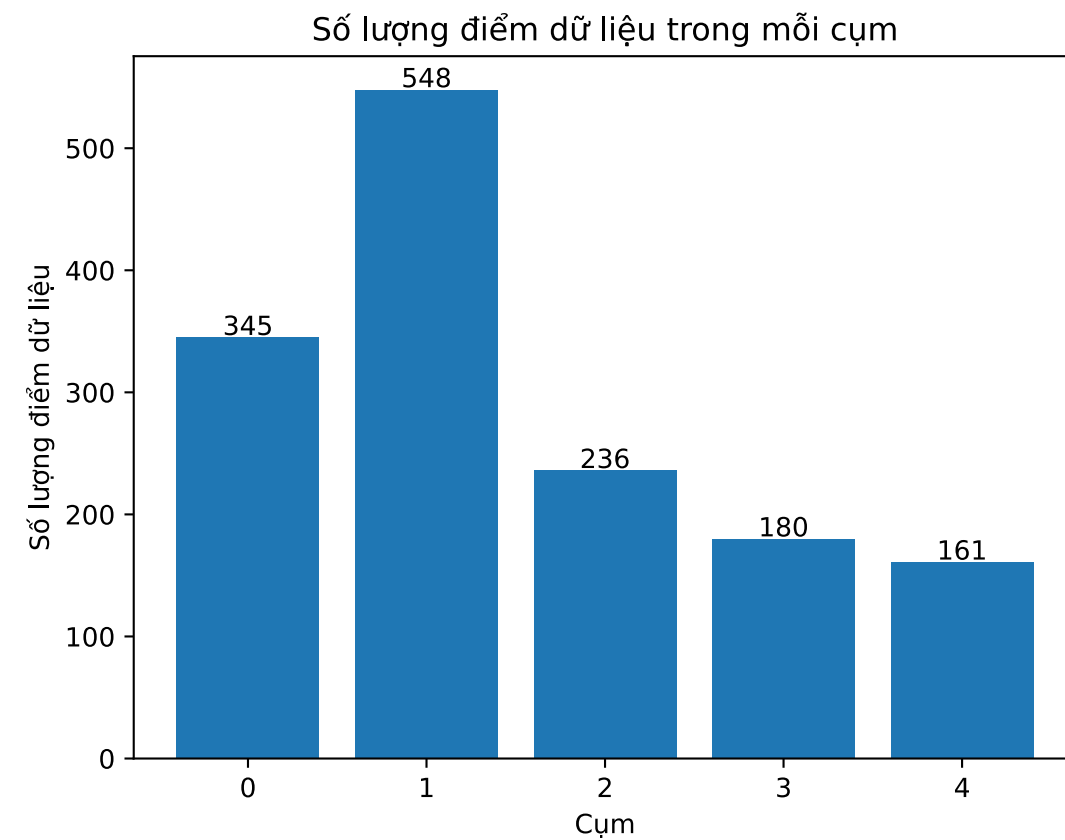
Biểu đồ Silhouette đạt giá trị cao nhất ở 5 cụm, cho thấy cách phân nhóm này là tối ưu nhất



Biểu đồ Gap Statistic cho thấy giá trị Gap tăng dần khi số lượng cụm tăng lên, có sự gia tăng rõ rệt ở 5 cụm. Điều này cũng gợi ý rằng 5 cụm là số lượng tối ưu

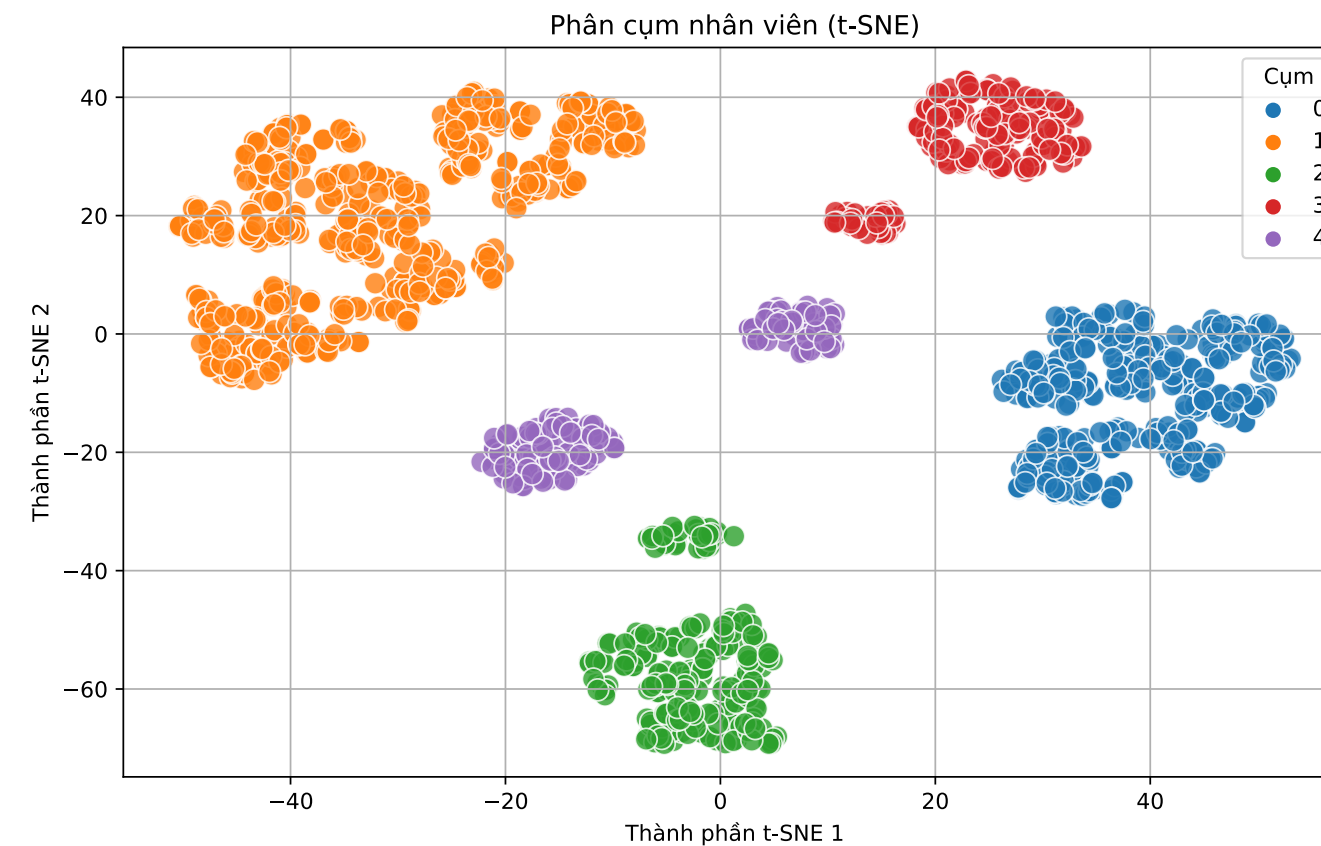


Phân tích kết quả phân cụm



Dữ liệu nhân viên đã được chia thành 5 cụm rõ ràng.

Cho thấy sự không đồng đều về kích thước giữa các cụm.

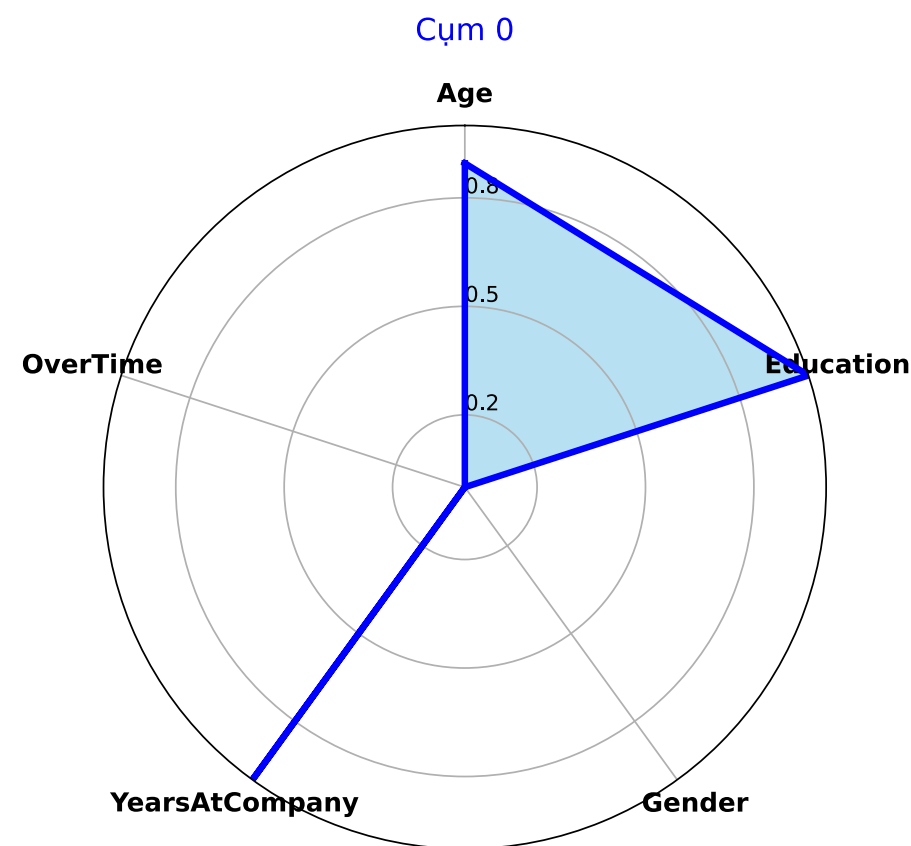


Các cụm được phân tách rõ ràng trong không gian 2 chiều khi giảm chiều dữ liệu bằng t-SNE.

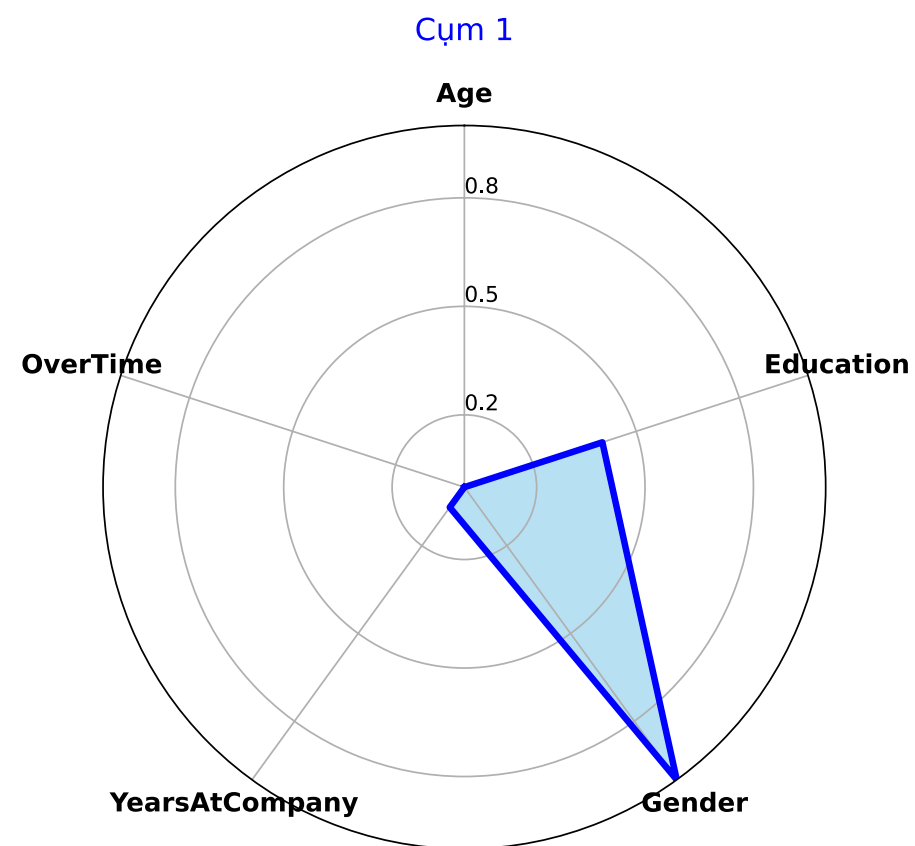
Cụm 1 và cụm 0 có mật độ dày và chiếm diện tích lớn. Các cụm nhỏ hơn nằm tách biệt và không có sự chồng lấn đáng kể, chứng tỏ việc phân cụm khá hiệu quả.

Quá trình phân cụm đã xác định được 5 nhóm nhân viên khác nhau với kích thước và đặc điểm riêng biệt. Sự khác biệt rõ ràng giữa các cụm cho phép dễ dàng phân tích sâu sắc hơn về đặc điểm và nhu cầu của từng nhóm nhân viên để đưa ra chiến lược quản lý phù hợp.

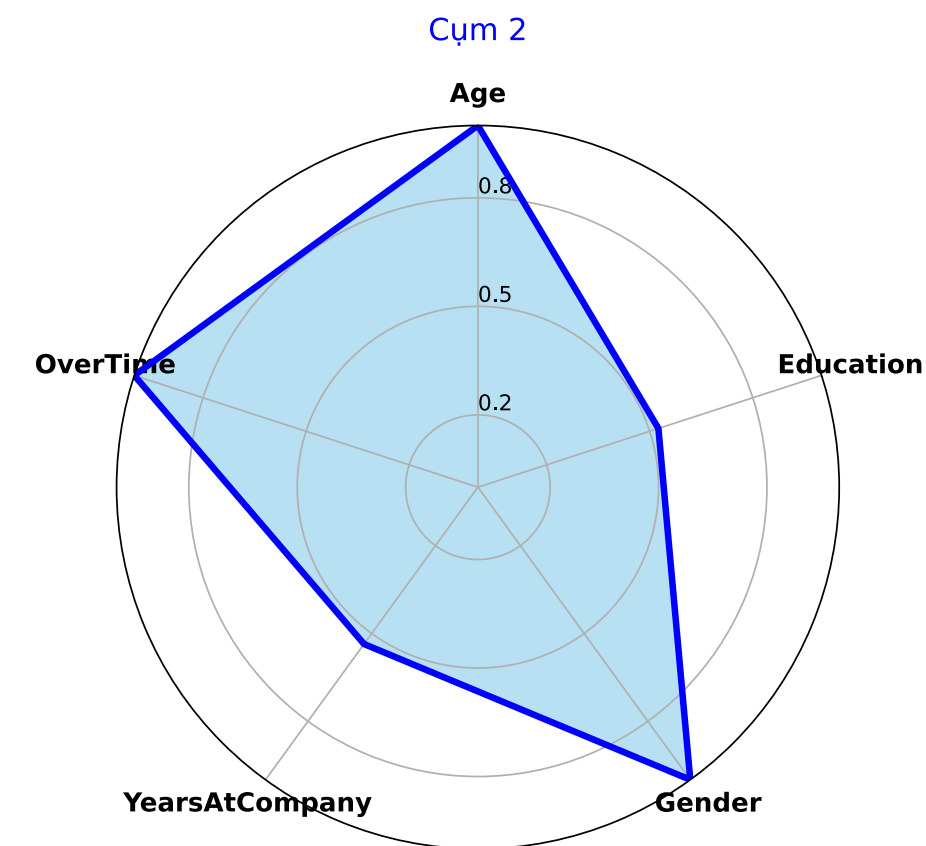
Phân tích kết quả phân cụm



Nhóm 0: Đa số là những nhân viên nữ lâu năm tại công ty, có mức độ học vấn cao



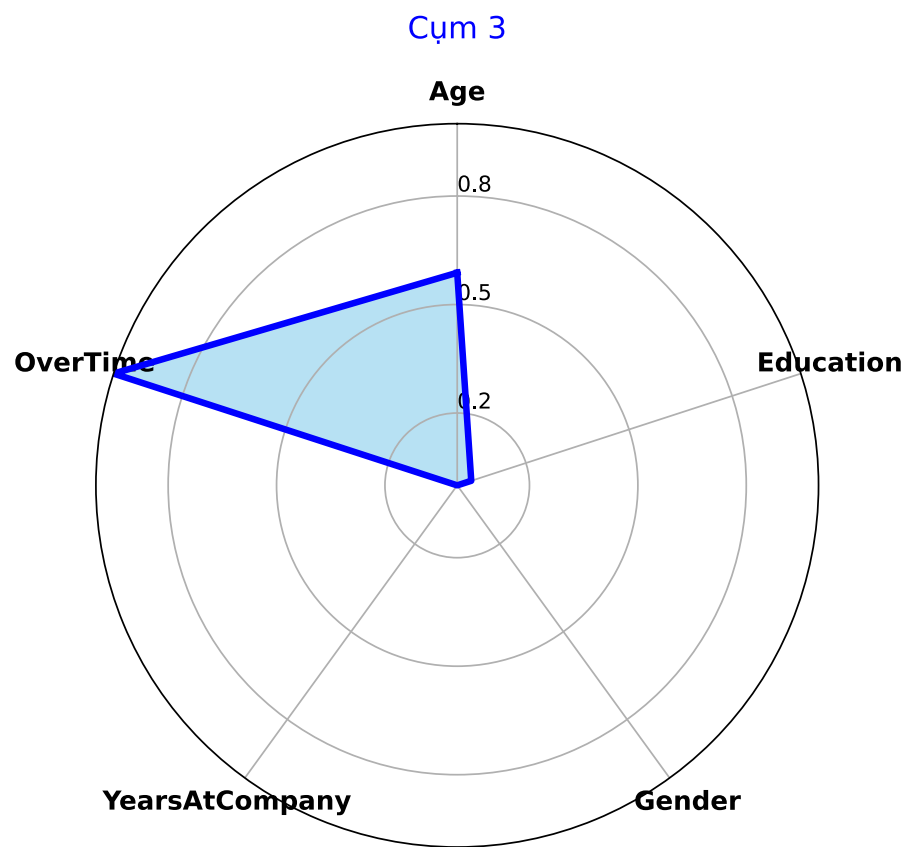
Nhóm 1: Gồm những nhân viên nam có mức độ học vấn trung bình và số năm làm việc tại công ty ít



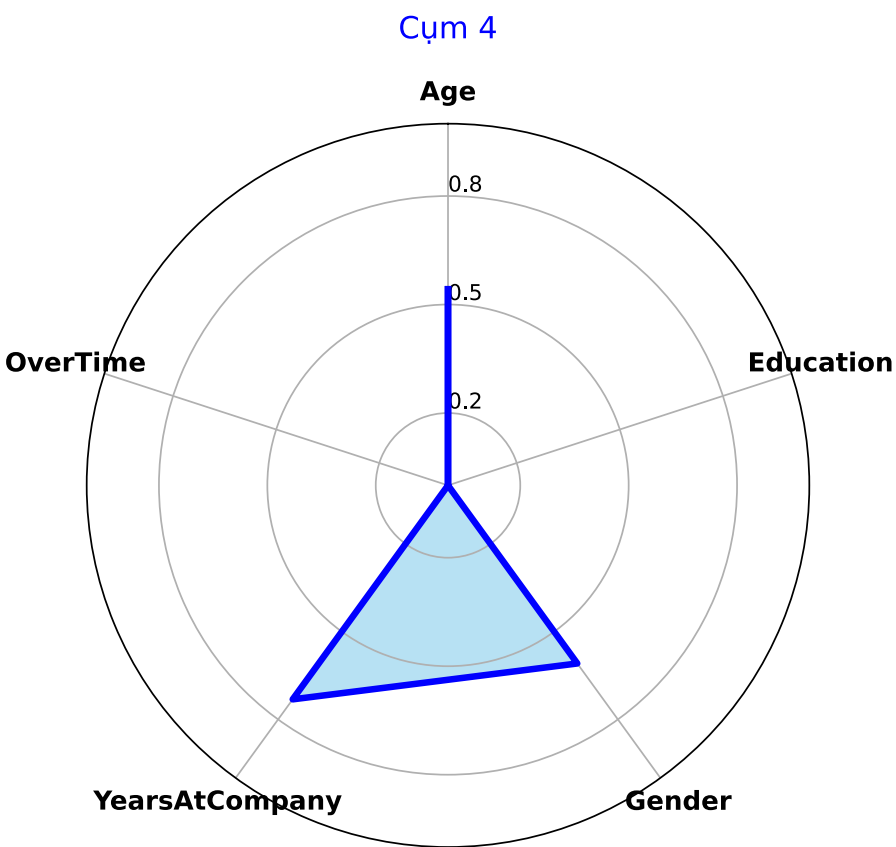
Nhóm 2: Nhân viên nam, lớn tuổi, thường xuyên tăng ca



Phân tích kết quả phân cụm



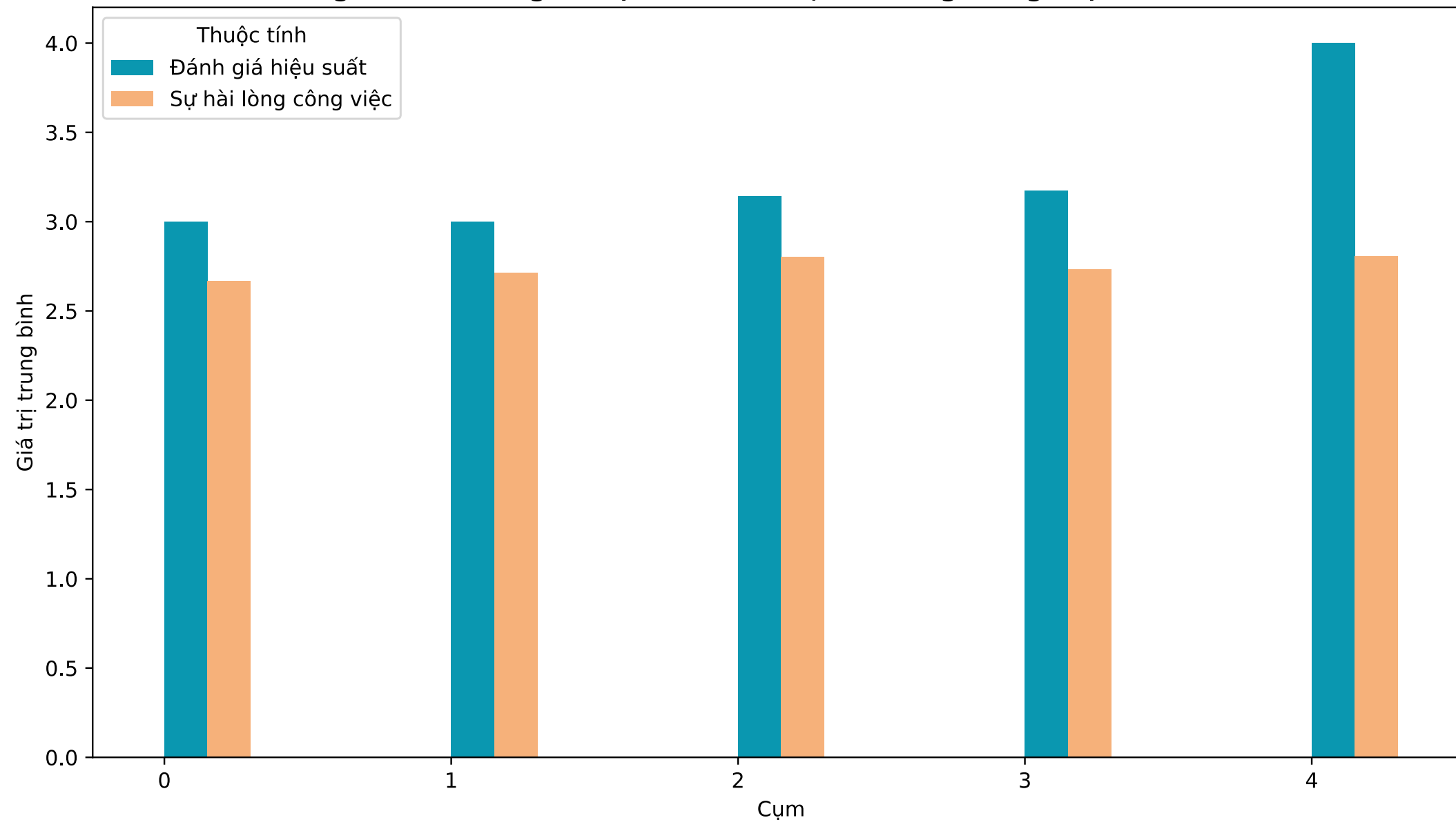
Nhóm 3: Nhân viên nữ mới, hay tăng ca, có tiềm năng phát triển



Nhóm 4: Nhóm này là tổ hợp giữa nam và nữ, những người có số năm gắn bó và độ tuổi trung bình

Phân tích kết quả phân cụm

Trung bình đánh giá hiệu suất và sự hài lòng công việc theo nhóm



Từ biểu đồ, ta thấy được các nhóm nhân viên có mức độ hài lòng với công việc là như nhau, không phân biệt vai trò, giới tính, hôn nhân hay vị trí công việc.

Nhưng về hiệu suất của từng nhóm thì lại có sự khác biệt khá khác:

- Nhóm 4 cho thấy được hiệu suất cực kì tốt của nhóm nhân viên này
- Trái ngược với nhóm 4 là nhóm 0, với những nhân viên kỳ cựu lâu năm và trình độ học vấn cao lại có hiệu suất làm việc là thấp nhất trong công ty
- Nhóm 3 có thể được gọi là nhóm có triển vọng nhất vì nhóm này chỉ gồm các nhân viên mới vào nhưng lại đạt hiệu suất cao thứ hai
- Còn lại nhóm 1 và 2 chỉ bao gồm những nhân viên nam, không phân biệt độ tuổi hay số lượt tăng ca mà lại có mức hiệu suất gần bằng nhau





Kết luận

Kết luận rút ra

- Hiệu suất làm việc không tỷ lệ thuận với mức độ hài lòng của nhân viên.
- Thâm niên và trình độ không phải là yếu tố quyết định hoàn toàn hiệu suất.
- Cần phân nhóm nhân viên dựa trên sự kết hợp nhiều yếu tố, bao gồm cá nhân, công việc và môi trường.

Đề xuất chiến lược

- Cá nhân hóa các chương trình đào tạo và phúc lợi phù hợp với từng nhóm.
- Tăng cường cải thiện môi trường làm việc cho các nhóm có mức độ hài lòng thấp
- Tập trung khai thác tiềm năng từ nhóm nhân viên mới và nhóm có hiệu suất cao để phát triển lâu dài

PHẢN HỒI



1. Khó khăn của mỗi thành viên:

- Khó khăn về thời gian: Không đủ thời gian để hoàn thành công việc một cách toàn diện.
- Khó khăn kỹ thuật: Một số thành viên chưa quen với các công cụ vẽ biểu đồ, git/github. Phải tìm hiểu về những phương pháp mới như Silhouette Score và Gap Statistic để áp dụng cho mô hình K-means, chưa có nhiều kinh nghiệm trong việc huấn luyện mô hình.

2. Mỗi thành viên học được:

- Kỹ năng làm việc nhóm: Cải thiện khả năng giao tiếp và hợp tác với các thành viên khác.
- Quản lý thời gian: Biết cách phân bổ thời gian hợp lý hơn cho các nhiệm vụ quan trọng.
- Kỹ năng mới: Học thêm các công cụ hoặc phương pháp làm việc mới trong quá trình thực hiện nhiệm vụ.

3. Nếu có thêm thời gian, nhóm em sẽ làm gì?

- Dành thêm thời gian để sử dụng tối ưu các cột đặc trưng có liên quan để trả lời các câu hỏi.
- Phân tích sâu hơn: Thực hiện nghiên cứu bổ sung hoặc cải thiện kết quả cuối cùng.
- Rèn luyện kỹ năng: Thử nghiệm và áp dụng thêm các công cụ hoặc phương pháp mới để nâng cao hiệu suất làm việc.

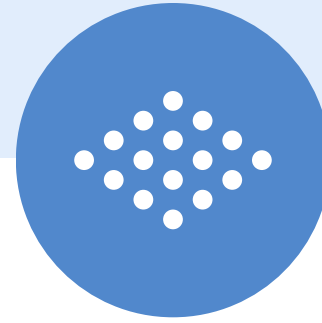


NGUỒN THAM KHẢO

Outlier!!! The Silent Killer

IBM Attrition Analysis and Prediction

HR Analytics and Prediction of Employee Attrition



**THANK
YOU!**

