

**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HCM**

**KHOA CÔNG NGHỆ THÔNG TIN**

🙠🙟🕮🙝🙢



**BÁO CÁO CUỐI KỲ**

**MÔN: TƯƠNG TÁC DỮ LIỆU TRỰC QUAN**

**Đề tài:Ứng Dụng Phân Tích Dữ Liệu Trong Dự Báo Nghỉ Việc Nhân Sự**

**GVHD: TS. Lê Quang Thái**

**Mã HP : IDVI333677**

**Sinh viên thực hiện (Nhóm 9):**

1. Mai Hồng Hải 22133014

2. Nguyễn Công Đôn 22133013

3. Nguyễn Tấn Hùng 22133027

Thành phố Hồ Chí Minh, tháng 5 năm 2025

|  |  |
| --- | --- |
| BỘ GIÁO DỤC VÀ ĐÀO TẠO  **TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT THÀNH PHỐ HỒ CHÍ MINH**  **KHOA CÔNG NGHỆ THÔNG TIN**  **MÔN TƯƠNG TÁC DỮ LIỆU TRỰC QUAN** | **CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM**  **Độc lập - Tự do - Hạnh phúc**  *TP. HCM, ngày 10 tháng 5 năm 2025* |

**DANH SÁCH THÀNH VIÊN**

HỌC KỲ II NĂM HỌC 2024 - 2025

1. **Mã lớp môn học: IDVI333677**
2. **Giảng viên hướng dẫn: Lê Quang Thái**
3. **Tên đề tài : Ứng Dụng Phân Tích Dữ Liệu Trong Dự Báo Nghỉ Việc Nhân Sự**
4. **Danh sách nhóm viết báo cáo cuối kỳ:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **HỌ VÀ TÊN SINH VIÊN** | **Mã số**  **sinh viên** | **Tỉ lệ %**  **tham gia** | **Mô tả nhiệm vụ** |
| 01 | Mai Hồng Hải | 22133014 | 100% | Xác Định Các Nhóm Nhân Sự Có Nguy Cơ Nghỉ Việc Cao.  Phân tích mô hình dự đoán, huấn luyện mô hình và đánh giá kết quả |
| 02 | Nguyễn Công Đôn | 22133013 | 100% | Phụ trách Data Storytelling , trực quan hoá tình Hình Nhân Sự, báo cáo tổng hợp và trình bày cuối |
| 03 | Nguyễn Tấn Hùng | 22133027 | 100% | Phân tích EDA, xử lý dữ liệu, trực quan hóa biểu đồ phân tích Các Yếu Tố Ảnh Hưởng Đến Quyết Định Nghỉ Việc |

***Nhận xét của giáo viên***

………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………

*Ngày… tháng 5 năm 2025*

**Giáo viên chấm**

### MỤC LỤC

[**MỤC LỤC**](#_mt2d6yg5opxt) **3**

[**PHẦN MỞ ĐẦU**](#_ftrkwg25bpy0) **4**

[**PHẦN NỘI DUNG CHÍNH: XÂY DỰNG CÂU CHUYỆN DỰA TRÊN DỮ LIỆU**](#_5qtdl64b6i40) **7**

[I. Tổng Quan Tình Hình Nhân Sự](#_872t21nu4rwi) 7

[1️ Tổng số nhân viên](#_mb0zu1nwhj0) 8

[2️ Tổng số nhân viên nghỉ việc](#_61zfwowc8gr4) 9

[3️ Tỷ lệ nghỉ việc toàn công ty (%)](#_5erf7z6hhab6) 10

[4️. Tỷ lệ nghỉ việc theo Phòng ban](#_8uirpgadnm8l) 11

[5️. Tỷ lệ nghỉ việc theo Vai trò (JobRole)](#_ylokccc1dhio) 12

[6️. Tỷ lệ nghỉ việc theo Giới tính](#_dli8cxgfvkbu) 14

[II. Các Yếu Tố Ảnh Hưởng Đến Quyết Định Nghỉ Việc](#_ju0glr957swa) 15

[1.Giới Thiệu](#_e8bnh1v19xhi) 15

[2.Phân tích nhân khẩu học](#_2d0o61b8esu4) 15

[3.Yếu tố hài lòng & cảm nhận](#_xpxc6f34ecx9) 21

[III. Xác Định Các Nhóm Nhân Sự Có Nguy Cơ Nghỉ Việc Cao](#_gi3v3deq4fn8) 24

[1. Top 5 Vai trò công việc (JobRole) có tỷ lệ nghỉ việc cao nhất](#_47mq23tq02jo) 25

[2. Hành trình nghỉ việc: JobRole → OverTime→Attrition](#_4rx5mph08pf) 26

[3. Phân tích chi tiết nhóm Nhân viên Kinh doanh (Sales Representative)](#_xx372d6o1sxl) 28

[IV. Dự Đoán và Đề Xuất Hành Động](#_z494h8t64avd) 29

[1. Top 10 nhân viên có nguy cơ nghỉ việc cao nhất (dự đoán)](#_rjyu1p2xetv) 29

[2. Các yếu tố quan trọng ảnh hưởng đến dự đoán nghỉ việc (Feature Importance)](#_z03ljhhfh12r) 31

[**PHẦN KẾT LUẬN VÀ ĐỀ XUẤT**](#_9mjzug5n0sck) **33**

### PHẦN MỞ ĐẦU

**I. Giới thiệu chung:**

**Mục đích của báo cáo** Báo cáo này nhằm phân tích tình hình biến động nhân sự trong công ty thông qua bộ dữ liệu lịch sử về nhân viên. Mục tiêu chính là:

* Đánh giá tổng quan về tỉ lệ nghỉ việc trong tổ chức
* Xác định các yếu tố ảnh hưởng đến quyết định nghỉ việc của nhân viên như: độ tuổi, bộ phận công tác, thời gian làm việc, mức thu nhập,...
* Phân nhóm các đối tượng có nguy cơ nghỉ việc cao để giúp doanh nghiệp có chiến lược giữ chân nhân tài hiệu quả hơn. Từ đó đưa ra các đề xuất cải thiện về mặt chính sách nhân sự, môi trường làm việc và chế độ đãi ngộ.

**Giới thiệu công cụ sử dụng**

Trong dự án này, nhóm sử dụng **Apache Superset** – một nền tảng mã nguồn mở mạnh mẽ cho trực quan hóa dữ liệu và phân tích tương tác. Superset cho phép người dùng kết nối trực tiếp tới cơ sở dữ liệu, xây dựng dashboard, biểu đồ và khai thác dữ liệu một cách trực quan, dễ hiểu.  
 Bên cạnh đó, môi trường được triển khai bằng **Docker** – công cụ giúp đóng gói và triển khai hệ thống một cách nhất quán, thuận tiện trong việc chia sẻ, tái sử dụng và mở rộng hệ thống. Docker giúp nhóm dễ dàng cấu hình môi trường Superset, đảm bảo tính đồng nhất giữa các thành viên và tránh lỗi phụ thuộc hệ thống.

**II. Giới thiệu Apache Superset:**

**Superset là gì?**

Apache Superset là một nền tảng mã nguồn mở mạnh mẽ, chuyên dùng cho khám phá và trực quan hóa dữ liệu. Được phát triển ban đầu bởi Airbnb, Superset giúp người dùng dễ dàng xây dựng các biểu đồ và bảng điều khiển tương tác, hỗ trợ hiệu quả trong việc khai phá và phân tích thông tin từ dữ liệu thô.

**Các tính năng nổi bật của Superset:**

* **Kết nối đa dạng nguồn dữ liệu:** Superset hỗ trợ nhiều hệ thống lưu trữ dữ liệu phổ biến như SQL databases (MySQL, PostgreSQL, SQL Server), các công cụ Big Data (Hive, Presto, Spark SQL), cùng khả năng tích hợp dễ dàng các nguồn dữ liệu khác.
* **Tạo biểu đồ linh hoạt:** Superset cung cấp nhiều dạng biểu đồ trực quan như biểu đồ cột, biểu đồ tròn, heatmap, sankey, và nhiều loại biểu đồ tùy chỉnh khác.
* **Xây dựng dashboard tương tác:** Cho phép người dùng xây dựng các bảng điều khiển đa chiều, cập nhật dữ liệu theo thời gian thực, tương tác trực tiếp với dữ liệu nhằm khai phá sâu hơn các insight tiềm ẩn.

**Lợi ích khi sử dụng Superset trong phân tích dữ liệu nhân sự:**

* Giúp nhanh chóng phát hiện và trình bày trực quan các biến động về nhân sự.
* Hỗ trợ các nhà quản lý nhân sự nắm bắt thông tin nhanh chóng, từ đó ra quyết định chính xác hơn.
* Khả năng tương tác cao giúp người dùng chủ động khai phá và truy vấn thông tin, dễ dàng xác định các yếu tố tác động đến việc nghỉ việc của nhân viên, từ đó đưa ra các giải pháp giữ chân nhân tài phù hợp.

**III. Giới thiệu Docker:**

**Docker là gì?**

Docker là nền tảng mã nguồn mở hỗ trợ việc đóng gói, phân phối và triển khai ứng dụng dưới dạng **container** – một môi trường độc lập, nhẹ, và đồng nhất. Điều này giúp ứng dụng như Superset có thể chạy nhất quán trên mọi hệ thống, bất kể môi trường phần cứng hay hệ điều hành.

**Lý do sử dụng Docker trong dự án**

Trong khuôn khổ đồ án, nhóm triển khai **Apache Superset** bằng Docker để tối ưu hóa quá trình thiết lập môi trường làm việc. Cụ thể, Docker giúp nhóm:

* **Triển khai nhanh chóng:** Chỉ cần vài dòng lệnh để khởi tạo toàn bộ hệ thống Superset.
* **Tái sử dụng dễ dàng:** Các cấu hình được chuẩn hóa trong tệp docker-compose, giúp mọi thành viên có thể chạy được hệ thống mà không gặp lỗi môi trường.
* **Tách biệt dịch vụ:** Superset hoạt động cùng với PostgreSQL, Redis và các dịch vụ phụ trợ (worker, cache, scheduler), tất cả được chạy đồng bộ hóa thông qua container riêng biệt.

**Các bước triển khai Superset bằng Docker**

**1. Cấu trúc thư mục**

Dự án có thư mục superset/ chứa mã nguồn và tệp cấu hình cần thiết để chạy Docker:

A screenshot of a computer

AI-generated content may be incorrect.

**2. Khởi tạo Superset bằng Docker Compose**

Tại thư mục superset, thực hiện lần lượt 3 lệnh sau:

cd superset

docker compose -f docker-compose-non-dev.yml pull

docker compose -f docker-compose-non-dev.yml up -d

* Lệnh đầu tiên dùng để di chuyển đến thư mục cấu hình.
* Lệnh thứ hai tải về các image cần thiết cho Superset, PostgreSQL, Redis, Celery worker,...
* Lệnh thứ ba khởi chạy toàn bộ hệ thống dưới dạng container.

A screen shot of a computer

AI-generated content may be incorrect.

**3. Kiểm tra các container đang chạy**

Sau khi khởi động thành công, Docker Desktop hoặc Docker Extension trên VSCode sẽ hiển thị toàn bộ các container như:

* superset\_app: giao diện chính Superset
* superset\_db: cơ sở dữ liệu PostgreSQL
* superset\_worker, superset\_beat, superset\_cache: các dịch vụ hỗ trợ

A screenshot of a computer

AI-generated content may be incorrect.

**4. Truy cập Superset qua trình duyệt**

Mở trình duyệt và truy cập địa chỉ http://localhost:8088. Giao diện đăng nhập Superset sẽ xuất hiện, mặc định:

* **Username:** admin
* **Password:** admin

A screenshot of a computer

AI-generated content may be incorrect.

**Tác động đến dự án**

Việc triển khai Superset bằng Docker mang lại nhiều lợi ích thực tiễn:

* **Tiết kiệm thời gian setup** cho toàn bộ nhóm.
* **Đảm bảo đồng bộ môi trường** giữa các thành viên.
* **Dễ dàng khởi chạy lại** hệ thống trong các lần thuyết trình, demo hay đóng gói báo cáo.

**IV. Giới thiệu Tập Dữ Liệu:**

* + Tập dữ liệu được sử dụng trong báo cáo này có nguồn gốc từ **IBM HR Analytics Employee Attrition & Performance Dataset**, một bộ dữ liệu công khai trên nền tảng Kaggle. Đây là dữ liệu mô phỏng tình huống quản lý nhân sự trong một doanh nghiệp, với mục tiêu đánh giá và phân tích các yếu tố ảnh hưởng đến tình trạng nghỉ việc của nhân viên. Dữ liệu phản ánh thông tin tổng quan về hồ sơ cá nhân, công việc, mức lương, thời gian làm việc, mức độ hài lòng và các đặc điểm hành vi khác của người lao động.
  + Một số trường dữ liệu quan trọng phục vụ trực tiếp cho phân tích gồm có: Department (phòng ban công tác), JobRole (vai trò công việc), Attrition (tình trạng nghỉ việc), Age (tuổi), YearsAtCompany (số năm làm việc), MonthlyIncome (thu nhập hàng tháng), JobSatisfaction (mức độ hài lòng với công việc), OverTime (làm thêm giờ), cùng với nhiều biến khác như DistanceFromHome, EducationField, MaritalStatus,... Những thuộc tính này đóng vai trò cốt lõi trong việc xác định mối quan hệ giữa đặc điểm nhân viên và khả năng họ nghỉ việc.
  + Về quy mô, tập dữ liệu bao gồm **1.470 bản ghi** tương ứng với từng nhân viên, với tổng cộng **35 thuộc tính**. Trước khi đưa vào phân tích và trực quan hóa bằng Apache Superset, dữ liệu đã trải qua một số bước tiền xử lý cơ bản: loại bỏ các giá trị trống (nếu có), chuẩn hóa tên cột cho dễ truy xuất (ví dụ: chuyển về chữ thường, thay khoảng trắng bằng dấu gạch dưới), và kiểm tra các ngoại lệ trong các biến định lượng như MonthlyIncome và DistanceFromHome. Ngoài ra, một số trường hợp đã được gộp nhóm như phân loại nhóm tuổi (AgeGroup) hoặc nhóm mức thu nhập (IncomeBracket) để phục vụ tốt hơn cho việc phân tích xu hướng nghỉ việc theo nhóm.

### PHẦN NỘI DUNG CHÍNH: XÂY DỰNG CÂU CHUYỆN DỰA TRÊN DỮ LIỆU

Trong một bảng tính tưởng chừng khô khan, nơi mỗi dòng là một nhân viên và mỗi cột là một thuộc tính – tuổi, giới tính, phòng ban, thu nhập, chức danh, thâm niên, mức độ hài lòng – lại ẩn chứa một câu chuyện rất con người. Dữ liệu không chỉ nói về *ai đang làm gì*, mà còn hé lộ *trạng thái của họ* – có hài lòng không, có đang làm thêm giờ không, có còn ở lại công ty không.

Tập dữ liệu nhân sự mà chúng em phân tích là một bức chân dung toàn cảnh, phản ánh đa dạng khía cạnh của người lao động: từ nền tảng học vấn **(EducationField)**, tình trạng hôn nhân **(MaritalStatus)**, đến mức độ hài lòng **(JobSatisfaction, EnvironmentSatisfaction)**, và cả dấu hiệu rời đi **(Attrition)**. Khi tất cả những yếu tố này cùng tồn tại trong một bảng dữ liệu, một câu hỏi tự nhiên xuất hiện:  
 **"Điều gì khiến một nhân viên quyết định ở lại hay rời đi?"**

Có thể đó là vì họ làm trong một bộ phận có áp lực cao, có thể vì thu nhập chưa tương xứng, hoặc vì họ làm thêm giờ quá thường xuyên. Cũng có thể là do khoảng cách xa nhà, thiếu cơ hội thăng tiến, hay đơn giản là sự thiếu gắn kết. Dữ liệu không nói rõ lý do, nhưng để lại những dấu vết.

Chính vì vậy, báo cáo này ra đời – như một hành trình lần theo các dấu hiệu trong dữ liệu, để hiểu sâu hơn về **biến động nhân sự**, xác định những **nguy cơ tiềm ẩn**, và hướng tới việc đưa ra **đề xuất cải thiện chính sách nguồn nhân lực** một cách chủ động và có cơ sở.

#### I. Tổng Quan Tình Hình Nhân Sự

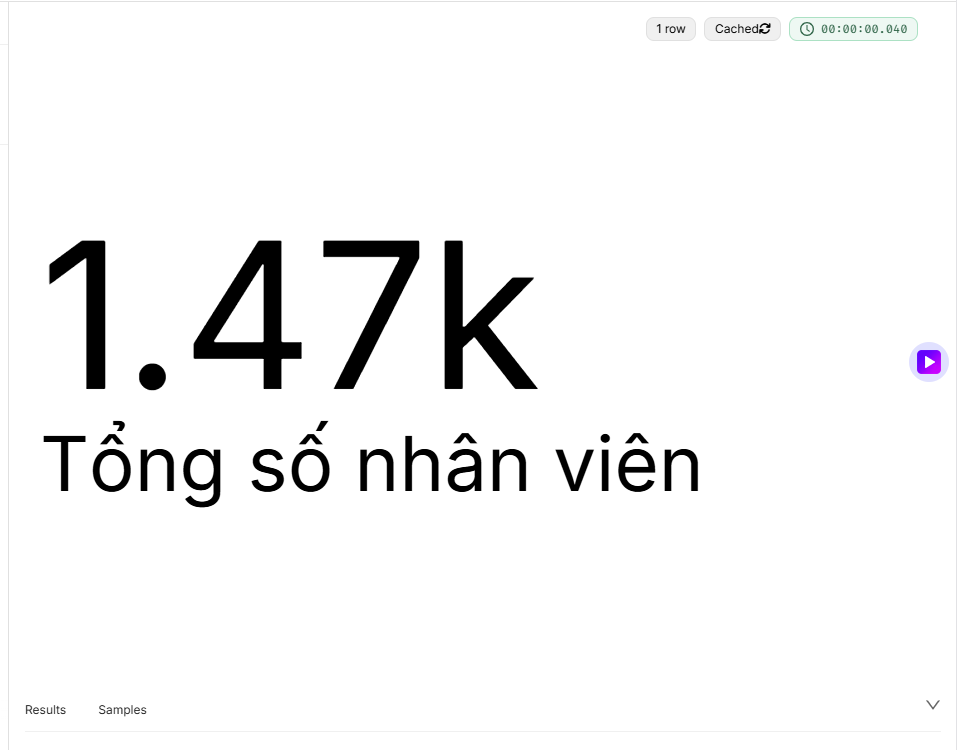
Mục tiêu: Cung cấp cái nhìn tổng thể về quy mô nhân sự và tỷ lệ nghỉ việc của công ty.

Trước khi đi sâu vào phân tích các yếu tố ảnh hưởng đến nghỉ việc, bước đầu tiên chúng tôi thực hiện là đánh giá bức tranh toàn cảnh về nhân sự của tổ chức. Câu hỏi đặt ra là: *"Công ty hiện có bao nhiêu nhân viên và mức độ biến động nhân sự hiện tại như thế nào?"*

Việc xác định quy mô nhân sự là cơ sở quan trọng để lượng hóa tỉ lệ nghỉ việc, từ đó xây dựng các chỉ số đo lường (KPI) phù hợp.

##### 1️ Tổng số nhân viên

* **Loại biểu đồ Superset:** Big Number
* **Mục đích:** Hiển thị tổng số dòng dữ liệu (tổng số nhân viên).
* **Hình ảnh biểu đồ:**

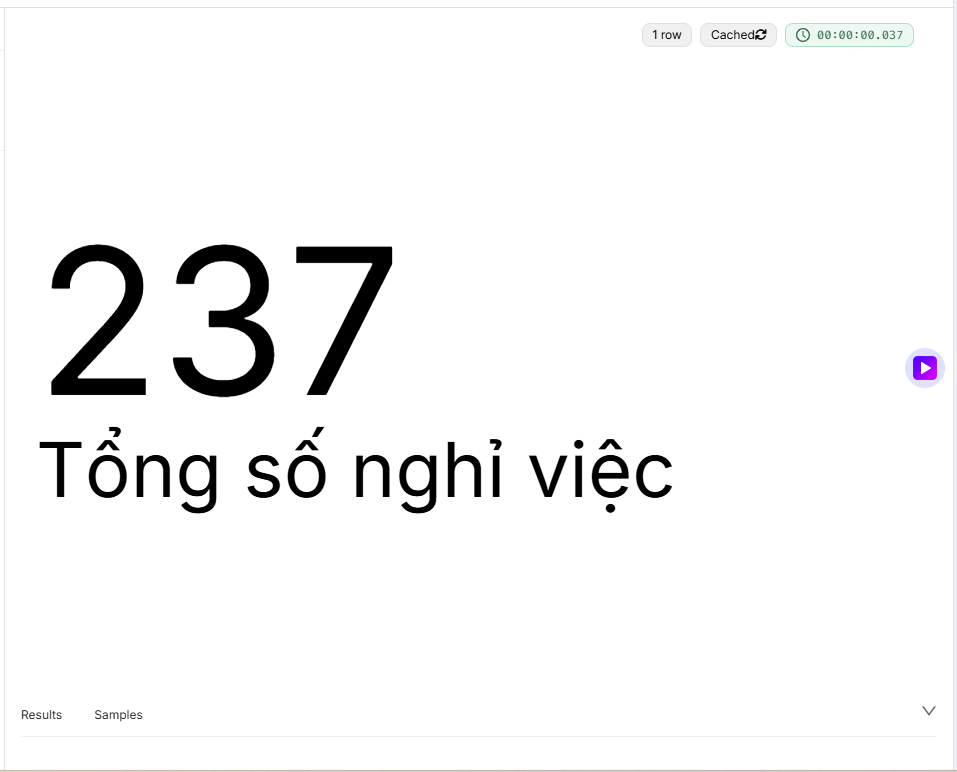
**

* **Phân tích:**

Dựa trên kết quả thống kê từ tập dữ liệu, hiện công ty đang có **1.470 nhân viên**. Tuy nhiên, con số tuyệt đối này chưa đủ để phản ánh rõ thực trạng.

##### 2️ Tổng số nhân viên nghỉ việc

* **Loại biểu đồ Superset:** Big Number (filter Attrition = Yes)
* **Mục đích:** Nhấn mạnh quy mô nhân viên đã nghỉ việc.
* **Hình ảnh biểu đồ:**

**

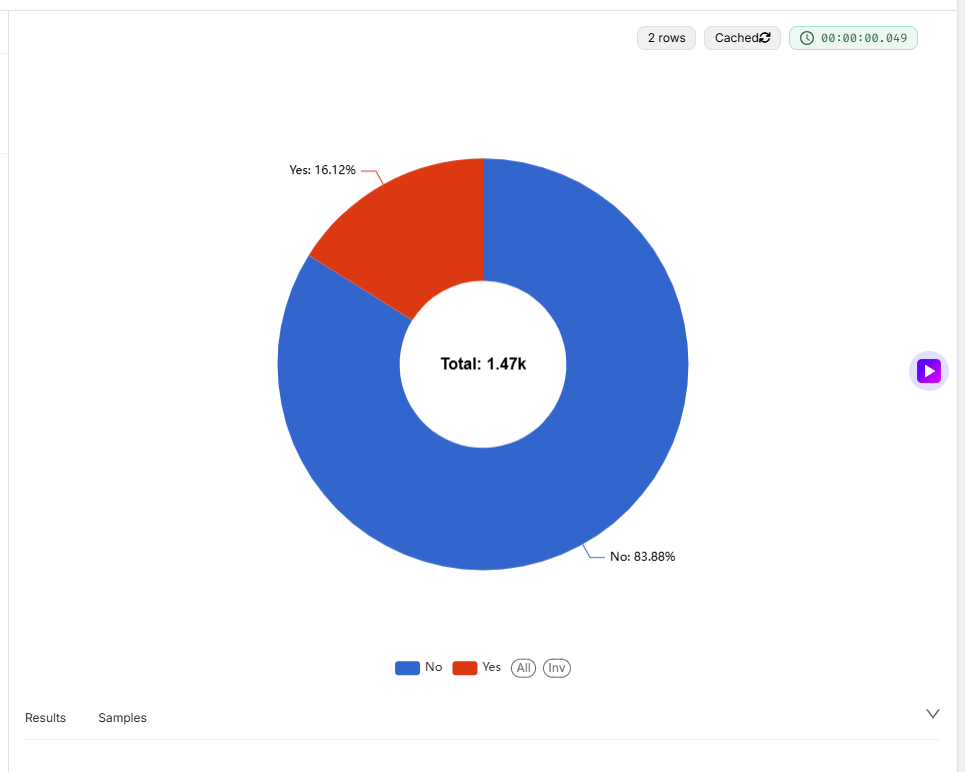
* **Phân tích:**

*Để hiểu rõ hơn mức độ biến động nhân sự, chúng tôi tiếp tục bóc tách số liệu tuyệt đối: trong tổng số 1.470 nhân viên, có* ***237 trường hợp đã nghỉ việc****. Đây là con số đáng chú ý vì nó phản ánh quy mô thực tế của tình trạng “chảy máu chất xám” trong tổ chức – không còn là một tỷ lệ phần trăm khô khan, mà là* ***237 con người cụ thể đã rời khỏi công ty****.*

*Việc biết chính xác số người nghỉ việc giúp doanh nghiệp hình dung được tác động của hiện tượng này đến hoạt động vận hành: từ khối lượng công việc bị dồn lại, chi phí tuyển dụng – đào tạo lại, đến ảnh hưởng tới tinh thần những người ở lại. Nó cũng là nền tảng để đánh giá hiệu quả của chính sách giữ chân nhân viên và đo lường rủi ro mất người trong tương lai.*

##### 3️ Tỷ lệ nghỉ việc toàn công ty (%)

* **Loại biểu đồ Superset:** Pie chart
* **Mục đích:** Đo lường chỉ số nghỉ việc chính của toàn công ty.
* **Hình ảnh biểu đồ:**

**

* **Phân tích:**

*Để bắt đầu bức tranh toàn cảnh về nguồn lực lao động của công ty, chúng tôi tiến hành thống kê tổng số nhân viên hiện tại trong tổ chức. Theo dữ liệu ghi nhận, tổng số nhân viên là* ***1.470 người****. Tuy nhiên, điều đáng chú ý là trong số đó đã có* ***237 người nghỉ việc****, tương đương với* ***16.12% tổng số nhân sự****.*

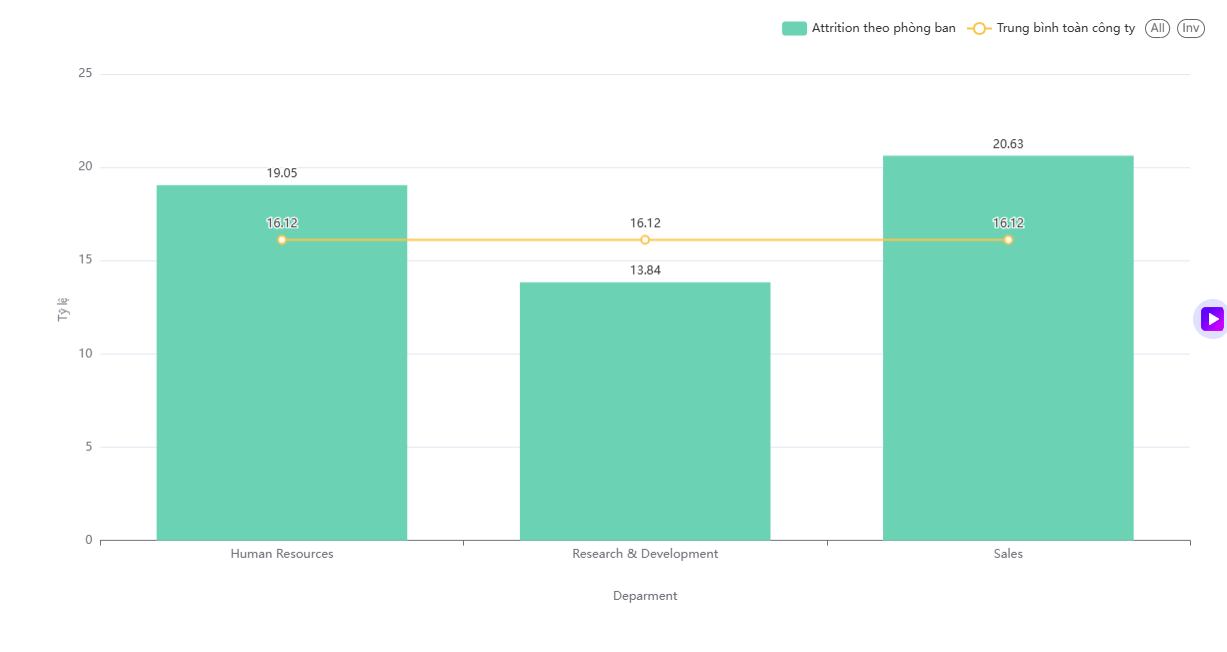
*Tỷ lệ nghỉ việc 16.12% không phải là con số nhỏ và có thể phản ánh nhiều vấn đề tiềm ẩn trong tổ chức như: môi trường làm việc thiếu ổn định, chính sách chưa phù hợp, khối lượng công việc cao hoặc thiếu sự gắn bó giữa nhân viên và doanh nghiệp.*

*Biểu đồ sau đây thể hiện rõ tỷ lệ giữa nhân viên vẫn đang làm việc (83.88%) và nhân viên đã nghỉ việc (16.12%), qua đó giúp ban lãnh đạo có cái nhìn trực quan về mức độ biến động nhân sự hiện tại.*

*Tuy nhiên, đây mới chỉ là cái nhìn cấp cao. Câu hỏi tiếp theo được đặt ra là:* ***"Những nhân viên nào đang nghỉ việc nhiều nhất? Họ thuộc bộ phận nào, làm công việc gì, và có đặc điểm gì nổi bật?".*** *Để giải đáp, chúng tôi sẽ đi sâu vào phân tích tỷ lệ nghỉ việc (Attrition Rate) theo từng* ***phòng ban (Department)*** *và* ***chức danh công việc (Job Role)*** *ở phần tiếp theo.*

##### 4️. Tỷ lệ nghỉ việc theo Phòng ban

* **Loại biểu đồ Superset:** Stacked Bar (Department + Attrition)
* **Mục đích:** So sánh tỷ lệ nghỉ việc giữa các phòng ban, xác định phòng ban có nhiều người nghỉ.
* **Hình ảnh biểu đồ:**

**

* **Phân tích:**

*Sau khi xác định tỷ lệ nghỉ việc toàn công ty là* ***16.12%****, bước tiếp theo trong quá trình phân tích là đặt câu hỏi: “Phòng ban nào đang có tỷ lệ nghỉ việc cao hơn mức trung bình này? Và liệu có những khu vực nào cần được chú ý đặc biệt?”*

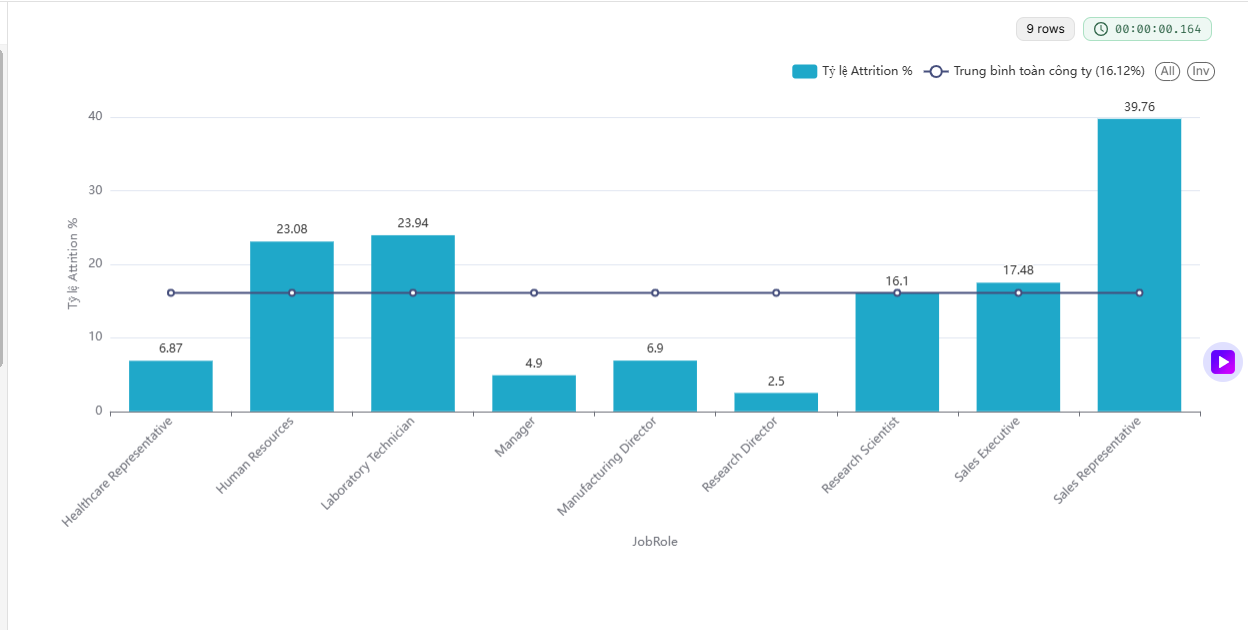
*Biểu đồ trên minh họa rõ ràng sự khác biệt về tỷ lệ nghỉ việc giữa các phòng ban:*

* ***Phòng Sales*** *có tỷ lệ nghỉ việc cao nhất, lên tới* ***20.63%****, vượt xa mức trung bình toàn công ty. Đây có thể là một điểm nóng cần được ưu tiên phân tích sâu thêm để tìm hiểu nguyên nhân – có thể do áp lực doanh số, khối lượng công việc lớn, hay chính sách đãi ngộ chưa phù hợp.*
* ***Phòng Nhân sự (Human Resources)*** *cũng ghi nhận tỷ lệ nghỉ việc cao hơn trung bình, ở mức* ***19.05%****. Điều này gây chú ý vì đây chính là bộ phận đóng vai trò giữ chân và phát triển con người – việc người làm nhân sự cũng rời bỏ công ty là một tín hiệu không mấy tích cực.*
* *Ngược lại,* ***phòng Nghiên cứu và Phát triển (Research & Development)*** *lại có tỷ lệ nghỉ việc thấp nhất, chỉ* ***13.84%****, thấp hơn đáng kể so với mặt bằng chung. Điều này cho thấy môi trường làm việc ở đây có thể ổn định hơn, hoặc nhân viên có xu hướng gắn bó lâu dài hơn với công việc chuyên môn.*

*Từ phân tích này, có thể thấy* ***tỷ lệ nghỉ việc không phân bố đồng đều giữa các bộ phận****. Việc nhận diện chính xác các khu vực có “nguy cơ cao” như Sales và HR giúp doanh nghiệp khoanh vùng vấn đề, phân bổ nguồn lực phù hợp và triển khai các giải pháp cải thiện chính sách giữ chân nhân sự một cách có trọng tâm*

##### 5️. Tỷ lệ nghỉ việc theo Vai trò (JobRole)

* **Loại biểu đồ Superset:** Stacked Bar và Line Chart
* **Mục đích:** Làm nổi bật các vai trò công việc (JobRole) có nguy cơ nghỉ việc cao.
* **Hình ảnh biểu đồ:**

**

* **Phân tích:**

*Sau khi đánh giá tỷ lệ nghỉ việc giữa các phòng ban, chúng tôi tiếp tục đào sâu thêm một lớp để trả lời câu hỏi: "Những chức danh công việc cụ thể nào đang có tỷ lệ nghỉ việc cao vượt trội? Và ai đang rời bỏ công ty nhiều nhất?"*

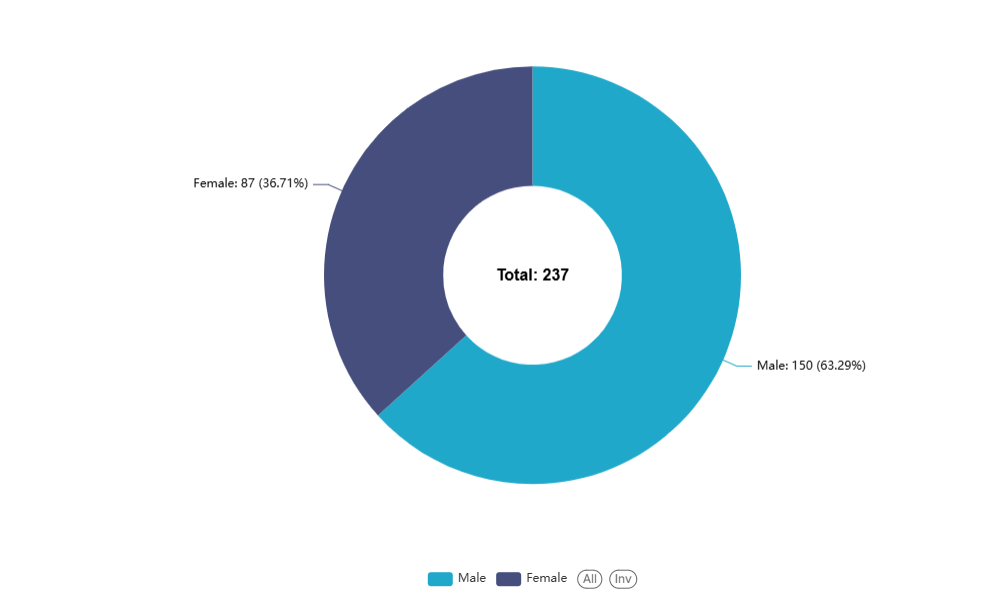
*Biểu đồ trên thể hiện tỷ lệ nghỉ việc tương ứng với từng chức danh công việc, cùng với đường trung bình toàn công ty là* ***16.12%*** *để làm chuẩn so sánh. Kết quả cho thấy có sự chênh lệch rõ rệt giữa các vai trò:*

* ***Sales Representative (Nhân viên kinh doanh)*** *là nhóm có tỷ lệ nghỉ việc cao nhất với con số đáng báo động:* ***39.76%*** *– gần gấp 2.5 lần mức trung bình. Điều này phản ánh mức độ biến động rất lớn trong lực lượng bán hàng trực tiếp, nơi nhân viên có thể chịu nhiều áp lực từ chỉ tiêu, khách hàng và môi trường cạnh tranh.*
* *Một số chức danh khác cũng có tỷ lệ nghỉ việc cao hơn trung bình, bao gồm* ***Laboratory Technician (23.94%)*** *và* ***Human Resources (23.08%)****. Những vai trò này tuy không mang tính chất kinh doanh như Sales, nhưng lại cho thấy dấu hiệu về sự mệt mỏi hoặc thiếu hài lòng trong môi trường làm việc hiện tại.*
* *Ngược lại, các vị trí như* ***Research Director (2.5%)****,* ***Manager (4.9%)****, và* ***Manufacturing Director (6.9%)*** *có tỷ lệ nghỉ việc rất thấp. Điều này có thể đến từ sự ổn định của vai trò lãnh đạo hoặc chuyên môn cao, cùng với mức độ cam kết lâu dài với tổ chức.*

*Qua biểu đồ, có thể thấy rõ* ***sự mất cân bằng trong tỷ lệ nghỉ việc giữa các chức danh****, từ đó giúp doanh nghiệp xác định các vị trí cần được hỗ trợ, chăm sóc hoặc cải tổ. Nhóm “Sales Representative” là đối tượng nổi bật cần được quan tâm đặc biệt trong các chiến lược giữ chân nhân viên sắp tới.*

##### 6️. Tỷ lệ nghỉ việc theo Giới tính

* **Loại biểu đồ Superset:** Donut Chart
* **Mục đích:** So sánh tỷ lệ nghỉ việc giữa nam và nữ.
* **Hình ảnh biểu đồ:**

**

* **Phân tích:**

*Một trong những khía cạnh thường bị bỏ qua nhưng lại rất quan trọng trong phân tích nhân sự là yếu tố* ***giới tính****. Câu hỏi đặt ra ở đây là: Liệu có sự khác biệt đáng kể nào về hành vi nghỉ việc giữa nam và nữ?*

*Biểu đồ tròn trên thể hiện cơ cấu giới tính trong nhóm* ***237 nhân viên đã nghỉ việc****. Kết quả cho thấy:*

* ***Nam giới chiếm đa số*** *với* ***150 người (63.29%)***
* ***Nữ giới chiếm 87 người (36.71%)***

*Sự chênh lệch này cho thấy nam giới có xu hướng nghỉ việc nhiều hơn so với nữ trong tổ chức.*

#### II. Các Yếu Tố Ảnh Hưởng Đến Quyết Định Nghỉ Việc

##### 1.Giới Thiệu

*Sau khi có cái nhìn tổng thể về tình hình nghỉ việc trong công ty thông qua các biểu đồ tổng quan, phần này sẽ đi sâu vào phân tích các yếu tố cụ thể ảnh hưởng đến quyết định nghỉ việc của nhân viên. Mỗi nhân viên rời đi không chỉ là một dòng dữ liệu biến mất, mà là một câu chuyện, một nguyên nhân và một dấu hiệu cảnh báo cho tổ chức. Vì vậy, việc truy vết các yếu tố liên quan đến nghỉ việc là rất quan trọng để doanh nghiệp không chỉ phản ứng, mà còn chủ động trong chiến lược giữ chân nhân tài.*

*Phần này được chia thành 3 nhóm chính:*

* *Phân tích theo các yếu tố nhân khẩu học: độ tuổi, thâm niên, mức thu nhập*
* *Phân tích theo hành vi làm việc: làm thêm giờ, thời gian gắn bó*
* *Phân tích theo yếu tố cảm nhận: mức độ hài lòng, sự cân bằng cuộc sống – công việc*

*Mục tiêu của phần phân tích này là:*

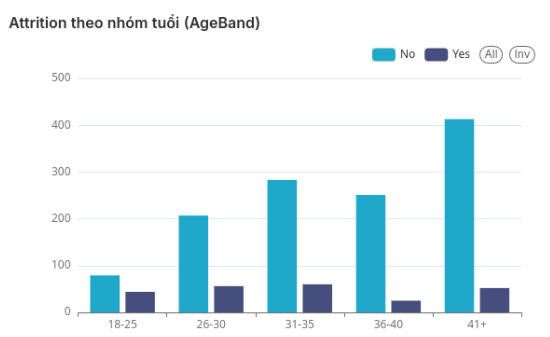
* *Xác định rõ nhóm nhân sự rủi ro cao*
* *Tìm hiểu mối liên hệ giữa đặc điểm nhân viên và khả năng nghỉ việc*
* *Gợi ý các hành động cải thiện cụ thể dựa trên dữ liệu thực tế*

##### 2.Phân tích nhân khẩu học

Mục tiêu:

Xác định các đặc điểm nền tảng như độ tuổi, thâm niên và thu nhập có ảnh hưởng như thế nào đến tỷ lệ nghỉ việc, từ đó phát hiện nhóm nhân sự có nguy cơ cao.

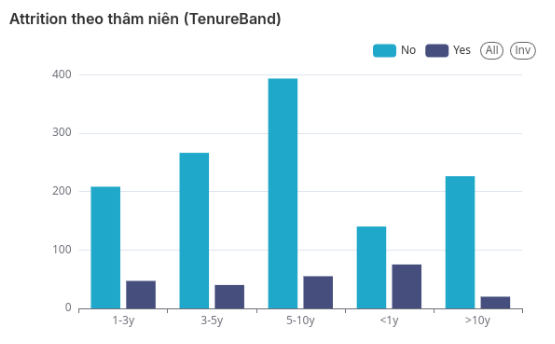
Biểu đồ 1: Tỷ lệ nghỉ việc theo Nhóm tuổi (AgeBand)

****

**Nhận định:**

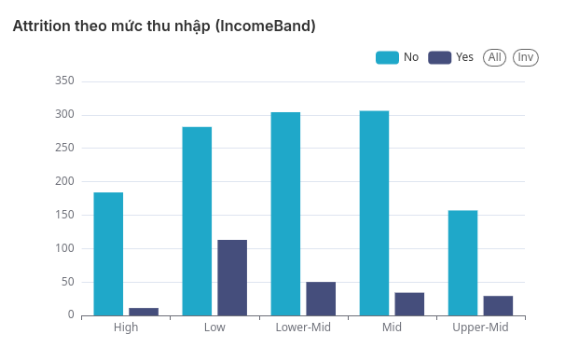
* Nhóm tuổi **18–25** có tỷ lệ nghỉ việc cao nhất so với các nhóm khác.
* Điều này phản ánh nhóm nhân viên trẻ thường có xu hướng thay đổi công việc cao hơn, có thể do còn tìm kiếm định hướng nghề nghiệp phù hợp.
* Nhóm tuổi từ **31–40** có xu hướng ổn định hơn, tỷ lệ nghỉ việc giảm nhẹ.
* Tuy nhiên, tỷ lệ nghỉ việc lại tăng nhẹ trở lại ở nhóm **41+**, có thể do các yếu tố về cân bằng cuộc sống, sức khỏe hoặc hướng đến nghỉ hưu sớm.

Biểu đồ 2: Tỷ lệ nghỉ việc theo Thâm niên làm việc (TenureBand)

****  
 **Nhận định:**

* Nhân viên có thâm niên **< 1 năm** có tỷ lệ nghỉ việc cao nhất trong tất cả các nhóm.
* Điều này cho thấy giai đoạn đầu thử việc/nhập môn là rủi ro lớn về giữ chân nhân sự.
* Các nhóm có **3–10 năm** thể hiện sự ổn định cao hơn, với tỷ lệ nghỉ việc thấp đáng kể.
* Bất ngờ là nhóm **> 10 năm** lại có một số trường hợp nghỉ việc, có thể do họ đã đạt đến ngưỡng phát triển hoặc sẵn sàng chuyển hướng sự nghiệp.

Biểu đồ 3: Tỷ lệ nghỉ việc theo Mức thu nhập (IncomeBand)



**Nhận định:**

* Tỷ lệ nghỉ việc **cao đột biến ở nhóm thu nhập Low**, cho thấy chính sách đãi ngộ ảnh hưởng trực tiếp đến khả năng giữ chân nhân sự.
* Ngược lại, nhóm **High và Upper-Mid** gần như không ghi nhận nhiều ca nghỉ việc, chứng minh rằng thu nhập là yếu tố then chốt trong quyết định ở lại công ty.
* Các nhóm **Lower-Mid** và **Mid** vẫn có hiện tượng nghỉ việc, nhưng thấp hơn nhóm Low.

Biểu đồ 4: Phân phối mức lương theo Vai trò công việc (Chỉ tính nhân viên đã nghỉ việc)



**Nhận định:**

* Trong số các nhân viên đã nghỉ việc, **Sales Representative và Laboratory Technician** là hai nhóm chiếm tỷ trọng lớn và có xu hướng tập trung ở nhóm thu nhập **Low và Lower-Mid**.
* Ngược lại, các nhóm như **Manager** hay **Manufacturing Director** có thu nhập cao và số lượng nghỉ việc rất thấp.
* Biểu đồ phản ánh rõ ràng mối liên hệ giữa vai trò công việc, mức thu nhập và tỷ lệ nghỉ việc – những vị trí tuyến đầu thường chịu áp lực cao nhưng lại có thu nhập chưa tương xứng.

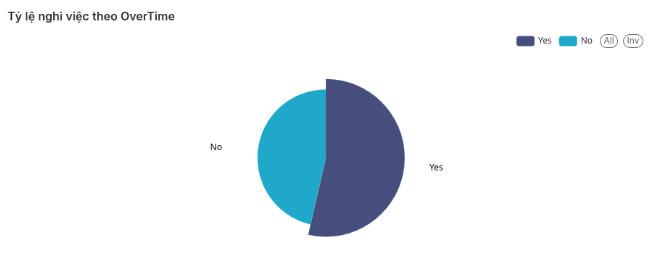
Tổng kết insight nhân khẩu học:

Những nhân viên có độ tuổi trẻ, thâm niên thấp, và thu nhập thấp là nhóm có nguy cơ nghỉ việc cao nhất. Việc xác định đúng các đặc điểm này sẽ giúp doanh nghiệp sớm nhận diện được nhóm rủi ro, từ đó triển khai các chính sách thu hút, đào tạo và giữ chân phù hợp hơn.

**3. Phân tích Hành vi & Thời gian làm việc**

Mục tiêu: Làm rõ mối quan hệ giữa các hành vi trong công việc và yếu tố thời gian gắn bó với tỷ lệ nghỉ việc, từ đó xác định các điểm "nóng" cần can thiệp.

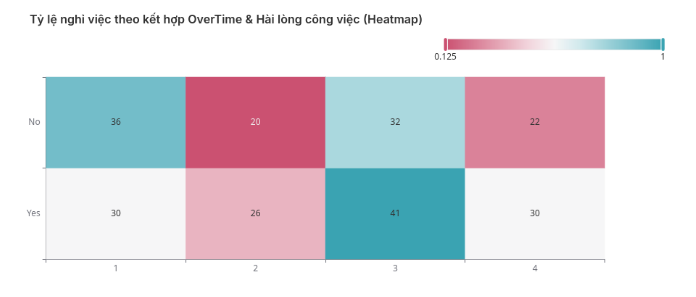
Biểu đồ 1: Tỷ lệ nghỉ việc theo làm thêm giờ (OverTime)



Nhận định:

* Nhóm nhân viên làm thêm giờ (OverTime = Yes) chiếm tỷ lệ nghỉ việc cao rõ rệt so với nhóm không làm thêm.
* Biểu đồ tròn cho thấy OverTime là một yếu tố rủi ro cao và có thể đo lường dễ dàng.
* Các doanh nghiệp nên xét lại chính sách làm thêm đối với nhóm nhân viên có nguy cơ nghỉ cao.

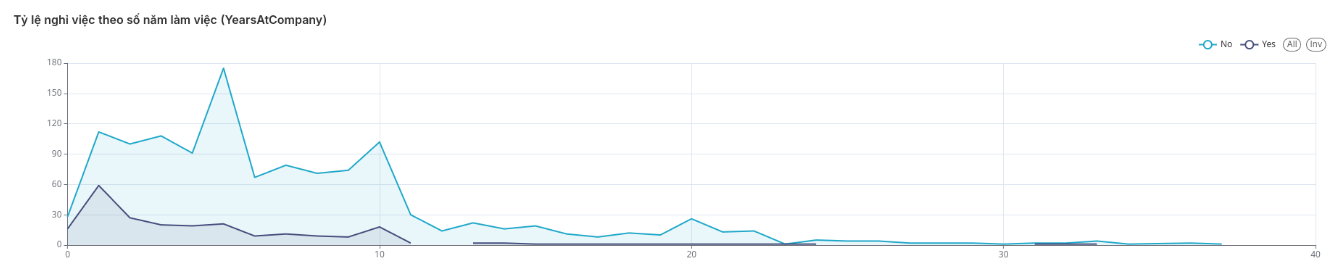
Biểu đồ 2: Tỷ lệ nghỉ việc theo kết hợp OverTime & Mức hài lòng công việc (Heatmap)



Nhận định:

* Nhóm làm thêm giờ và đánh giá JobSatisfaction dưới 3 có số ca nghỉ việc cao nhất (41 trường hợp).
* Trong khi nhóm không làm thêm nhưng hài lòng cao (3 hoặc 4) thường có tỷ lệ nghỉ thấp.
* Sự kết hợp giữa OverTime và đánh giá JobSatisfaction giúc xác định nhóm nguy cơ nghỉ việc một cách rõ ràng hơn so với khi nhìn rời rạc.

Biểu đồ 3: Tỷ lệ nghỉ việc theo số năm làm việc (YearsAtCompany)



Nhận định:

* Giai đoạn <1 năm và từ 1–3 năm là điểm "nóng" với số ca nghỉ việc nhiều nhất.
* Sau mốc 3 năm, tỷ lệ nghỉ việc bắt đầu giảm mạnh, cho thấy sự gắn bó cao hơn.
* Tuy nhiên, vẫn ghi nhận một số trường hợp nghỉ việc ở các nhóm >10 năm, có thể do bị bão hoà, thiếu cơ hội phát triển hoặc định hướng về hưu sớm.

Tổng kết insight hành vi:

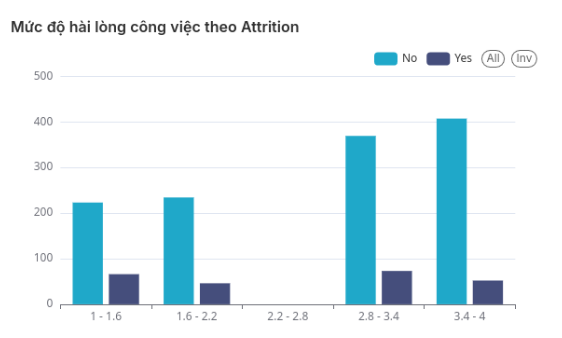
* Việc làm thêm giờ và đều hài lòng công việc là hai yếu tố tương tác mạnh lằng dẫn đến nghỉ việc.
* Thời gian gắn bó dưới 3 năm, đặc biệt <1 năm, là nhóm rủi ro cao cần quan tâm đặc biệt.
* Doanh nghiệp có thể cân nhắc các chính sách onboarding, mentoring đối với nhóm nhân viên mới và cân đối lại khối lượng làm thêm giờ, đặc biệt với nhóm có JobSatisfaction thấp.

##### 3.Yếu tố hài lòng & cảm nhận

**Mục tiêu:** Phân tích ảnh hưởng của các yếu tố cảm xúc và đánh giá chủ quan đến quyết định nghỉ việc, gồm:

* Mức độ hài lòng công việc (JobSatisfaction)
* Mức độ hài lòng môi trường làm việc (EnvironmentSatisfaction)
* Cân bằng công việc – cuộc sống (WorkLifeBalance)

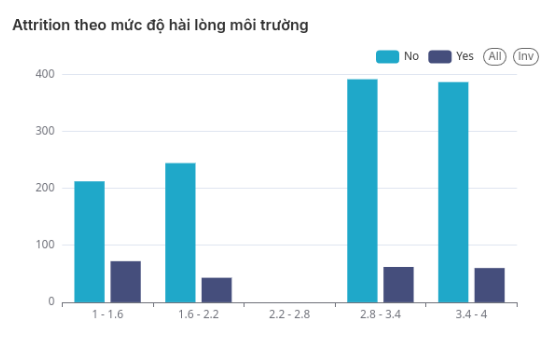
**Biểu đồ 1: Mức độ hài lòng công việc theo Attrition**

****

**Nhận định:**

* Nhân viên có mức hài lòng công việc từ 1 đến dưới 2.5 chiếm tỷ lệ nghỉ việc cao hơn rõ rệt so với các mức còn lại.
* Mức độ hài lòng cao (từ 2.8 trở lên) đi kèm với số lượng nghỉ việc giảm đáng kể.
* Điều này cho thấy cảm nhận chủ quan về công việc có ảnh hưởng trực tiếp đến quyết định rời bỏ công ty.

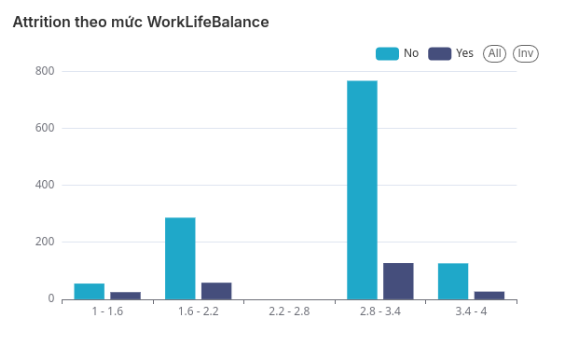
**Biểu đồ 2: Attrition theo mức độ hài lòng môi trường**

****

**Nhận định:**

* Những nhân viên không hài lòng với môi trường làm việc (mức 1.0 – 2.2) có xu hướng nghỉ việc nhiều hơn.
* Ngược lại, môi trường làm việc được đánh giá tốt hơn (mức 2.8 – 4.0) có tỷ lệ nghỉ việc thấp hơn, cho thấy tầm quan trọng của môi trường làm việc đối với sự gắn bó.
* Có thể bao gồm các yếu tố như đồng nghiệp, cấp trên, không gian, văn hóa tổ chức...

**Biểu đồ 3: Attrition theo mức WorkLifeBalance**

****

**Nhận định:**

* Nhân viên có cảm nhận cân bằng công việc – cuộc sống thấp (1.0 – 2.2) ghi nhận tỷ lệ nghỉ việc cao hơn hẳn.
* Mức độ WorkLifeBalance từ 2.8 trở lên giúp giữ chân nhân viên tốt hơn, phản ánh rằng sự hài hòa giữa công việc và cuộc sống cá nhân đóng vai trò quan trọng trong sự gắn bó lâu dài.
* Đây là yếu tố đặc biệt quan trọng trong bối cảnh nhân viên ngày càng ưu tiên sức khỏe tinh thần và thời gian cá nhân.

**Tổng kết insight yếu tố hài lòng & cảm nhận:**

Nhân viên có mức độ hài lòng thấp trong công việc, môi trường và cân bằng cuộc sống – công việc là nhóm rủi ro cao về nghỉ việc. Việc nâng cao trải nghiệm nhân viên thông qua cải thiện môi trường làm việc, xây dựng văn hóa tích cực và hỗ trợ cân bằng công việc – cuộc sống là hướng đi then chốt giúp doanh nghiệp giữ chân nhân tài bền vững.

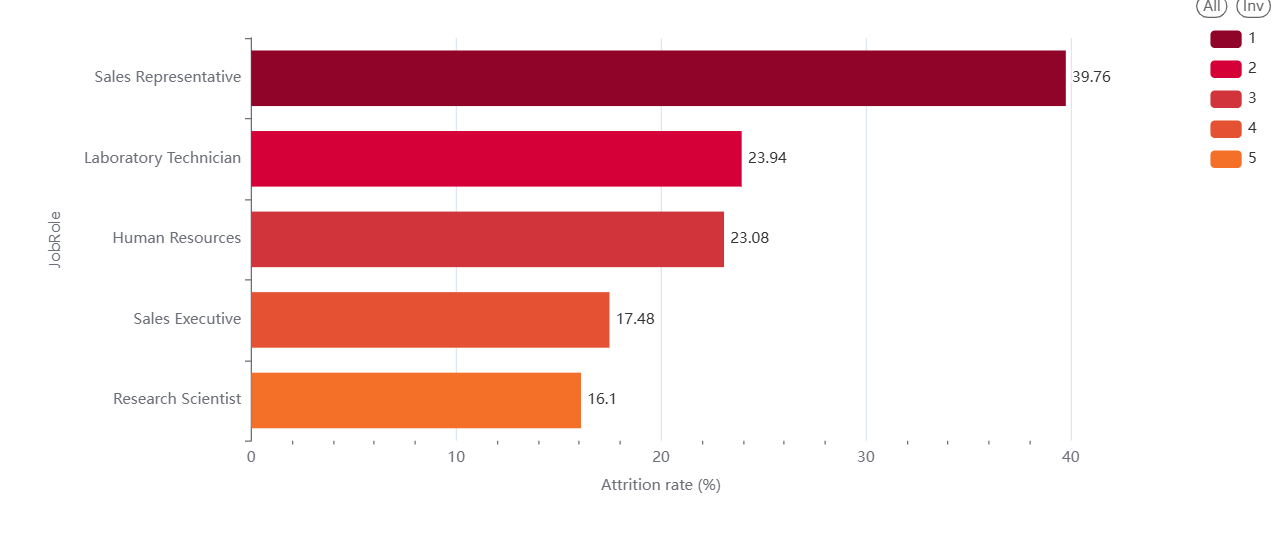
#### III. Xác Định Các Nhóm Nhân Sự Có Nguy Cơ Nghỉ Việc Cao

Sau khi hiểu được toàn cảnh về thực trạng nghỉ việc trong tổ chức cũng như những nguyên nhân , nhóm phân tích đã bắt đầu đi sâu vào những điểm nóng – nơi các chỉ số bắt đầu phát tín hiệu đỏ. Và hành trình đó bắt đầu từ một câu hỏi quen thuộc nhưng không dễ trả lời: "Ai đang có nguy cơ rời bỏ chúng ta cao nhất?"

##### 1. Top 5 Vai trò công việc (JobRole) có tỷ lệ nghỉ việc cao nhất

* **Loại biểu đồ Superset:** Word Cloud, Bar Chart
* **Mục đích:** Nhấn mạnh các nhóm vai trò công việc rủi ro nhất.
* **Hình ảnh biểu đồ:**

****

**

Trên biểu đồ Word Cloud và Bar Chart, những cái tên hiện lên đậm nét nhất không phải ai xa lạ. Sales Representative – vai trò tưởng chừng quen thuộc và năng động – lại là nhóm có tỷ lệ nghỉ việc gần 40%, cao nhất trong toàn bộ các vị trí. Đây không còn là một lời cảnh báo mà là một tiếng chuông báo động cần sự can thiệp ngay lập tức.

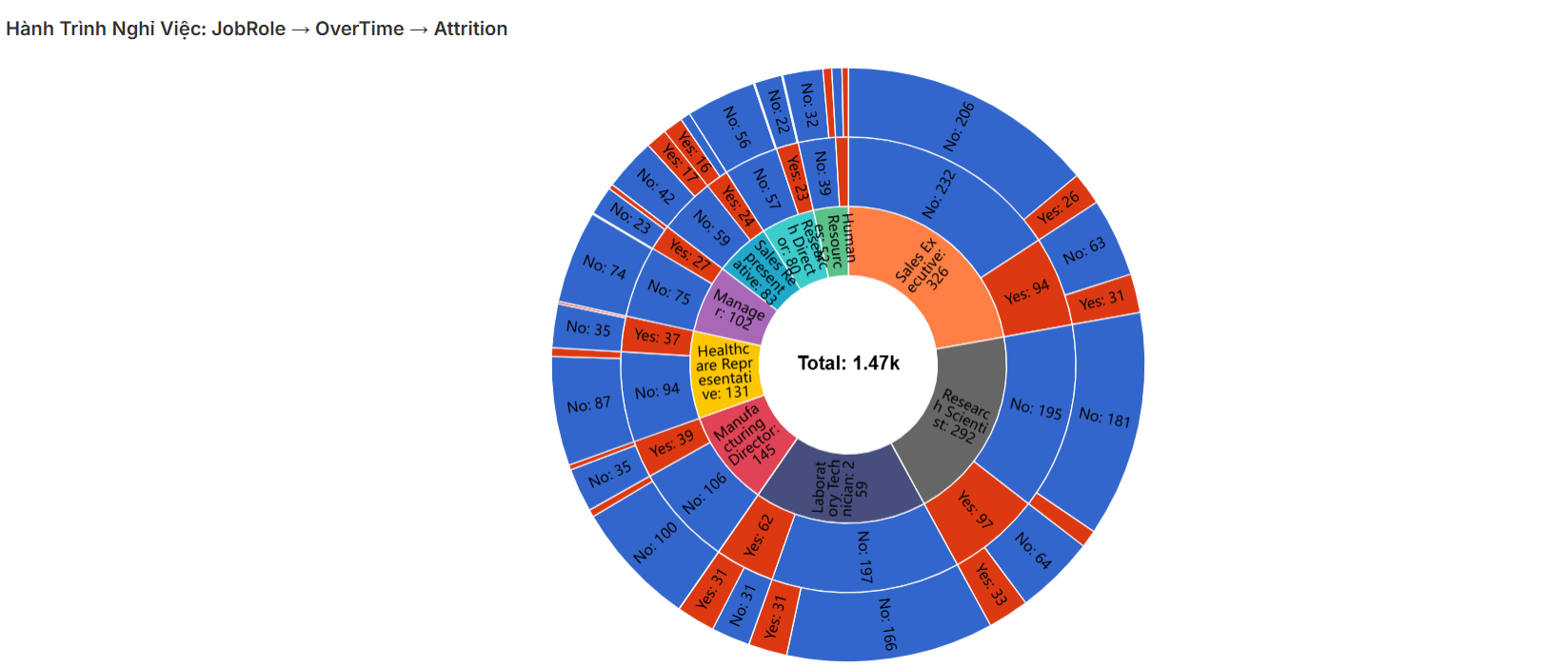
Không chỉ nhóm Sales, Laboratory Technician và Research Scientist – đại diện cho các khối kỹ thuật – cũng nằm trong nhóm rủi ro cao. Có thể là do áp lực công việc, thiếu cơ hội phát triển, hoặc đơn giản là cảm giác không còn động lực để tiếp tục. Đáng ngạc nhiên hơn, Human Resources – bộ phận được kỳ vọng là "người giữ người" – lại đứng thứ ba trong danh sách. Điều này cho thấy ngay cả những người lo cho người khác cũng cần được quan tâm đúng mực.

Hai vị trí liên quan đến bán hàng chiếm đến 2 trong 5 vị trí đầu bảng, cho thấy tác động rõ rệt của KPIs và biến động thị trường. Câu chuyện không chỉ nằm ở con số, mà là câu chuyện về sự quá tải, thiếu hỗ trợ, và có thể là thiếu định hướng phát triển.

* **Hành động đề xuất:**
* Xem xét lại chế độ lương thưởng, KPI và hỗ trợ nhân viên ở các nhóm Sales.
* Tạo lộ trình phát triển rõ ràng hơn cho nhóm kỹ thuật và nhân sự.
* Triển khai khảo sát chuyên sâu để xác định nguyên nhân rời bỏ cụ thể trong từng nhóm.

##### 2. Hành trình nghỉ việc: JobRole → OverTime→Attrition

* **Loại biểu đồ Superset:** Sunburst Chart
* **Mục đích:** Tái hiện dòng chảy dữ liệu nghỉ việc, xem xét sự kết hợp của các yếu tố.
* **Hình ảnh biểu đồ:**

**

Chúng tôi muốn nhìn thấy bức tranh toàn diện hơn – không chỉ là ai nghỉ việc, mà là lý do và con đường nào dẫn đến quyết định đó. Biểu đồ Sunburst là nơi chúng tôi tái hiện lại "hành trình nghỉ việc", bắt đầu từ vai trò công việc, đến việc có làm thêm giờ, và cuối cùng là nghỉ việc hay không.

Ngay tại tâm biểu đồ là con số 1.470 nhân viên – mỗi người là một câu chuyện riêng. Khi mở rộng ra các vòng tròn tiếp theo, ta thấy rõ ràng: những người làm Sales Executive và Research Scientist có số lượng nghỉ việc cao nhất, đặc biệt nếu họ thuộc nhóm thường xuyên làm thêm giờ.

OverTime – tưởng chừng là một yếu tố tất yếu trong môi trường cạnh tranh – hóa ra lại là "người bạn đồng hành" với quyết định nghỉ việc. Trong hầu hết các vai trò, phần lớn những người nghỉ việc đều nằm trong nhóm làm thêm nhiều.

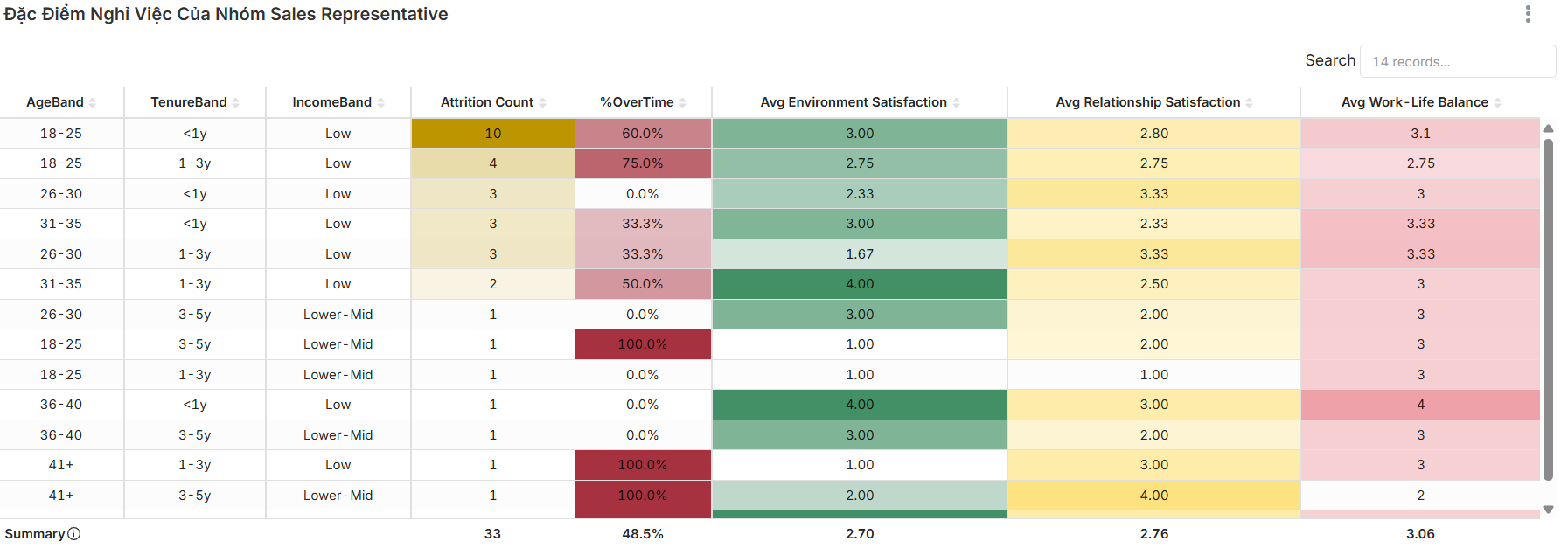
Tuy nhiên, không phải lúc nào OverTime cũng là thủ phạm duy nhất. Một ví dụ điển hình là nhóm Manager – dù có ít người làm thêm, nhưng vẫn có tỷ lệ nghỉ việc đáng kể. Đây là manh mối cho thấy còn nhiều yếu tố phi thời gian đang âm thầm ảnh hưởng: áp lực lãnh đạo, môi trường văn hóa, hoặc thiếu sự công nhận.

**Kết luận & đề xuất:**

* Kiểm soát thời gian làm việc, đặc biệt cho các nhóm áp lực cao như Sales, Lab, Research.
* Cải thiện điều kiện làm việc: hỗ trợ tinh thần, tạo động lực, giảm áp lực KPI.
* Với các nhóm ít làm thêm nhưng vẫn nghỉ việc, cần phân tích sâu hơn về lương, văn hóa, sự gắn kết và vai trò lãnh đạo.

##### 3. Phân tích chi tiết nhóm Nhân viên Kinh doanh (Sales Representative)

* **Loại biểu đồ Superset:** Table (Bảng) kết hợp với Filter theo JobRole = "Sales Representative"
* **Mục đích:** Khám phá sâu một nhóm cụ thể có tỷ lệ nghỉ việc cao (ví dụ: Sales Representative).
* **Hình ảnh biểu đồ/bảng:**

**

Sau khi nhận diện được những vai trò công việc có nguy cơ nghỉ việc cao và tìm ra mối liên hệ đáng kể giữa việc làm thêm giờ với tỷ lệ nghỉ việc, nhóm phân tích quyết định đi sâu vào một vai trò cụ thể đang nằm trong "vùng đỏ" của rủi ro nhân sự – đó chính là Sales Representative.

Hành trình tiếp tục như một cuộc điều tra thực địa. Khi "zoom" kỹ vào nhóm này, một bức tranh hiện ra rõ nét hơn. Trong tổng số 33 nhân viên Sales Representative đã nghỉ việc, gần một nửa thường xuyên làm thêm giờ, với tỷ lệ trung bình lên tới 48.5%. Con số này lập tức gợi lại kết luận trước đó: OverTime là một "chỉ báo nguy hiểm" về khả năng nghỉ việc.

Nhưng không dừng ở đó, khi soi chiếu các chỉ số về mức độ hài lòng, thêm một tín hiệu cảnh báo được bật lên: Trung bình, nhóm này chỉ đạt mức 2.70 cho môi trường làm việc, 2.76 cho quan hệ công sở, và 3.06 cho cân bằng công việc – cuộc sống. Đây đều là những mức thấp so với mặt bằng chung và cho thấy sự thiếu gắn bó và môi trường không tích cực đang âm thầm đẩy họ ra xa tổ chức.

Và điều đáng lo hơn nữa – chính là độ tuổi và thâm niên: Gần một phần ba số người nghỉ việc nằm ở nhóm tuổi từ 18 đến 25, với thời gian làm việc dưới 1 năm. Những người trẻ, mới vào nghề, lẽ ra là tài sản tương lai, lại đang là nhóm rời bỏ nhiều nhất. Không chỉ vậy, thu nhập cũng là một yếu tố nổi bật: Phần lớn những người rời bỏ thuộc nhóm thu nhập thấp hoặc trung bình thấp, và không có trường hợp nào ở nhóm thu nhập cao phải nói lời chia tay.

Từ tất cả những mảnh ghép dữ liệu này, nhóm phân tích đưa ra những khuyến nghị rõ ràng và gắn kết với bối cảnh:

* Hạn chế yêu cầu làm thêm giờ với nhân viên mới, nhất là trong 1–2 năm đầu làm việc, để họ có thời gian thích nghi và phát triển.
* Cải thiện môi trường và văn hóa làm việc, không chỉ bằng lời hứa mà bằng những hành động cụ thể như xây dựng chính sách feedback minh bạch, tạo không gian gắn kết nội bộ.
* Cân nhắc điều chỉnh thu nhập khởi điểm hợp lý, đặc biệt cho nhóm nhân viên tuyến đầu như Sales Representative – những người đối mặt trực tiếp với áp lực doanh số và khách hàng.
* Và cuối cùng, thiết kế các chương trình onboarding và mentoring bài bản, để nhân viên trẻ không cảm thấy đơn độc trong hành trình nghề nghiệp. Một người hướng dẫn phù hợp có thể giúp họ không chỉ hiểu công việc mà còn thấy mình thuộc về một nơi – điều kiện tiên quyết để giữ chân nhân tài.

#### IV. Dự Đoán và Đề Xuất Hành Động

Sau khi phân tích dữ liệu và hiểu rõ những yếu tố ảnh hưởng đến tình trạng nghỉ việc, nhóm chúng tôi tiếp tục triển khai một mô hình dự đoán để hỗ trợ doanh nghiệp chủ động trong việc giữ chân nhân sự. Mục tiêu của mô hình là dự báo khả năng một nhân viên có thể nghỉ việc trong tương lai, từ đó đưa ra các hành động phù hợp trước khi điều đó thực sự xảy ra.

##### 1. Top 10 nhân viên có nguy cơ nghỉ việc cao nhất (dự đoán)

* **Loại biểu đồ Superset:** Table (Bảng)
* **Mục đích:** Trích xuất danh sách nhân viên có nguy cơ nghỉ việc cao từ mô hình dự đoán.
* **Hình ảnh bảng:**

**

Từ kết quả dự đoán, chúng tôi trích xuất danh sách 10 nhân viên có xác suất nghỉ việc cao nhất để phân tích sâu. Bảng kết quả được hiển thị trên Superset dưới dạng biểu đồ Table, cung cấp thông tin chi tiết về các đặc điểm của từng nhân viên.

Một điều đáng chú ý là 8/10 nhân viên trong danh sách này đều có tỷ lệ làm việc ngoài giờ (OverTime) lên đến 100%. Điều này cho thấy họ đang phải làm thêm rất nhiều – có thể do thiếu nhân sự, khối lượng công việc cao, hoặc kỳ vọng hiệu suất lớn từ cấp quản lý. Việc làm thêm quá nhiều trong thời gian dài chắc chắn ảnh hưởng tiêu cực đến sức khỏe, tinh thần, và sự gắn bó của nhân viên.

Đáng lo ngại hơn, 6/10 người đến từ bộ phận Sales – một bộ phận vốn đã có đặc thù áp lực cao về doanh số và hiệu suất. Việc nhiều nhân viên Sales nằm trong nhóm nguy cơ nghỉ việc cao đặt ra tín hiệu cảnh báo cho ban lãnh đạo về khả năng suy giảm hiệu quả kinh doanh nếu tình trạng này kéo dài.

Không chỉ vậy, điểm số về sự hài lòng – bao gồm JobSatisfaction và EnvironmentSatisfaction – của các nhân viên này đều khá thấp, chủ yếu dao động từ 1 đến 2 (trên thang điểm 4 hoặc 5). Mức điểm trung bình cho JobSatisfaction chỉ khoảng 1.6, và với EnvironmentSatisfaction là 2.1, cho thấy một sự bất mãn khá rõ rệt trong nhóm nhân viên này.

Khi phân tích kỹ hơn về nguyên nhân rủi ro nghỉ việc (RiskReason) được mô hình ghi nhận, phần lớn lý do xoay quanh các cụm vấn đề như: "Thu nhập thấp, Làm thêm nhiều", hoặc kết hợp nhiều yếu tố như "Tuổi trẻ; Không hài lòng công việc; Làm thêm; Môi trường chưa tốt". Điều này củng cố giả thuyết rằng không chỉ một, mà nhiều yếu tố cộng hưởng đang đẩy nhân viên đến quyết định rời đi.

Mặc dù chỉ số trung bình về WorkLifeBalance là 3, một mức trung bình khá, nhưng tỷ lệ OverTime cao mâu thuẫn với chỉ số này, cho thấy có thể tồn tại sự chênh lệch đáng kể giữa các cá nhân – một vấn đề mà các chỉ số trung bình không thể hiện rõ.

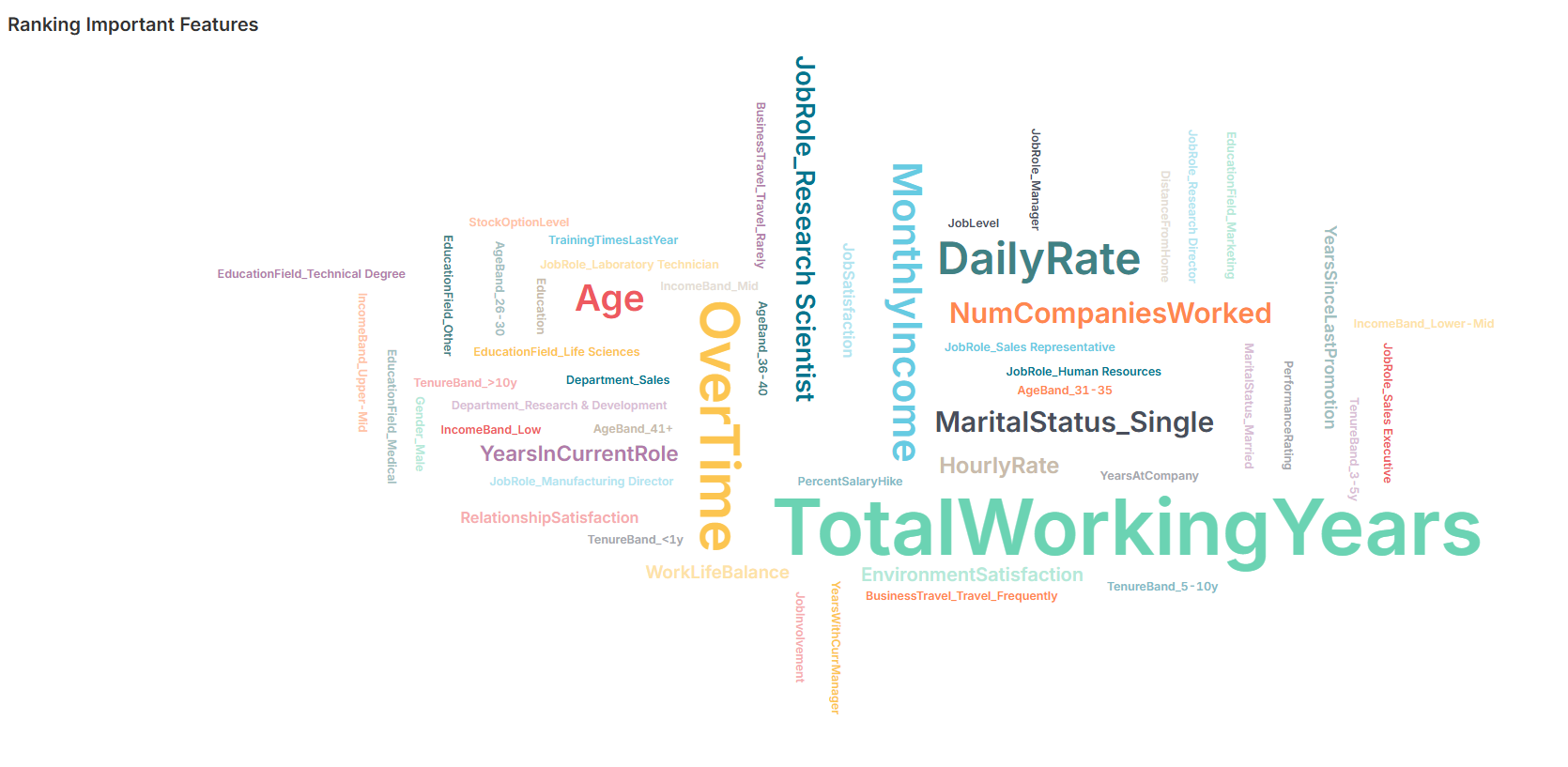
* **Đề xuất hành động :**

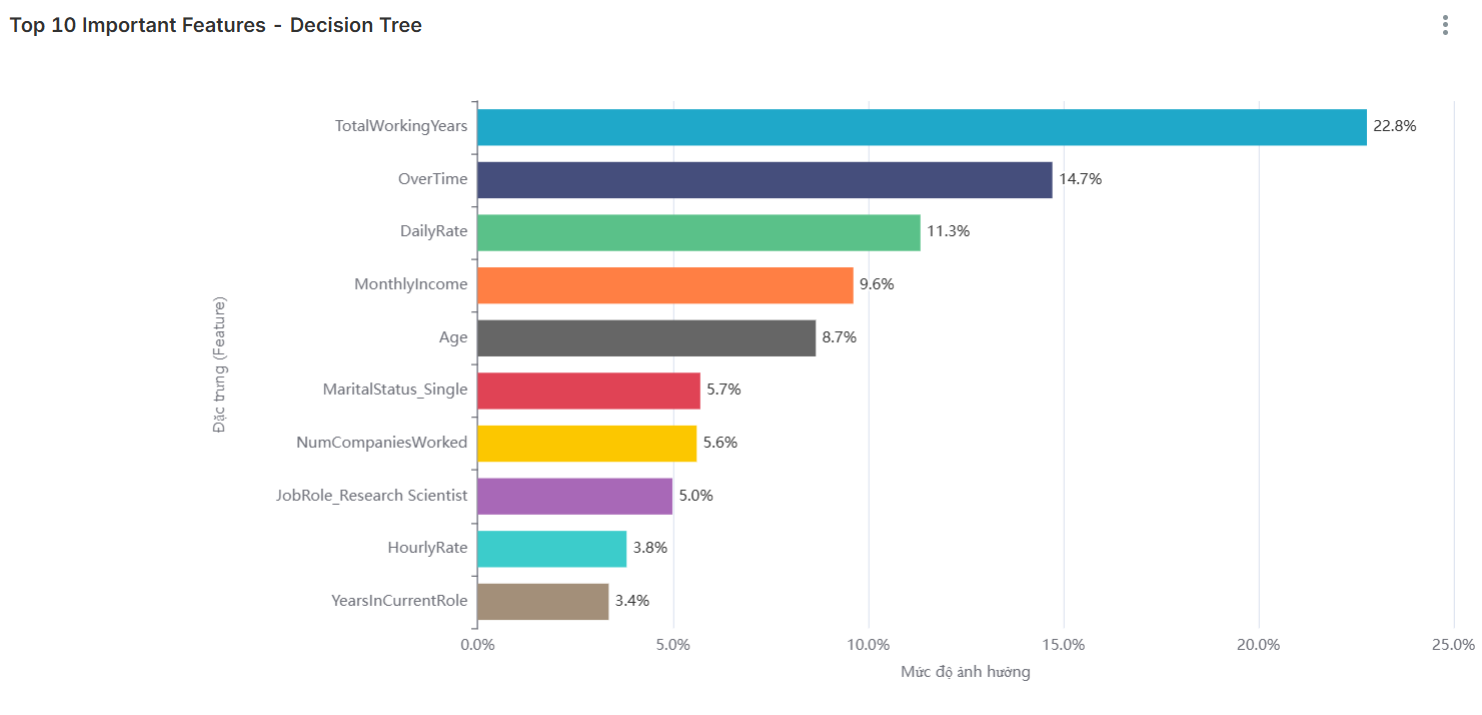
|  |  |
| --- | --- |
| **Hành động** | **Mục tiêu cụ thể** |
| Đối thoại 1:1 | Lắng nghe lý do bất mãn, xây dựng niềm tin, cam kết phát triển. |
| Điều chỉnh khối lượng làm việc | Hạn chế làm thêm, cân bằng cuộc sống – công việc. |
| Cải thiện lương thưởng | Xem xét tăng lương, phúc lợi, thưởng theo hiệu suất để giữ chân người giỏi. |
| Nâng cao môi trường làm việc | Cải thiện không gian làm việc, đào tạo quản lý, văn hóa tích cực. |
| Phát triển nhân viên trẻ | Đào tạo, xây dựng lộ trình nghề nghiệp và chương trình mentor. |

##### 2. Các yếu tố quan trọng ảnh hưởng đến dự đoán nghỉ việc (Feature Importance)

Để mô hình không chỉ là một “hộp đen”, nhóm chúng tôi đã tiến hành phân tích các yếu tố ảnh hưởng mạnh nhất đến kết quả dự đoán thông qua biểu đồ Word Cloud và Biểu đồ cột ngang (Horizontal Bar Chart) trong Superset.

* **Hình ảnh biểu đồ:**

**

**

Đứng đầu danh sách là TotalWorkingYears (tổng số năm làm việc) – với tỷ trọng lên đến 22.8%. Điều này có thể phản ánh rằng những nhân viên có nhiều kinh nghiệm thường có nhiều lựa chọn hơn trên thị trường lao động và sẵn sàng rời đi nếu cảm thấy không được ghi nhận hoặc không còn cơ hội phát triển.

Tiếp theo là OverTime (14.7%) – điều đã được thấy rõ trong nhóm Top 10 nguy cơ cao. Đây là một yếu tố mang tính cảnh báo mạnh mẽ và dễ can thiệp nhất nếu công ty muốn thay đổi.

Yếu tố về thu nhập như DailyRate (11.3%) và MonthlyIncome (9.6%) cũng có ảnh hưởng rõ rệt. Điều này càng nhấn mạnh vai trò quan trọng của chính sách đãi ngộ trong việc giữ chân nhân tài.

Ngoài ra, các yếu tố nhân khẩu học như Age (8.7%) và MaritalStatus\_Single (5.7%) cũng góp phần giải thích hành vi nghỉ việc. Những người trẻ và độc thân có thể linh hoạt và sẵn sàng tìm kiếm cơ hội mới.

Một điểm thú vị là JobRole\_Research Scientist cũng xuất hiện trong top 10 yếu tố quan trọng (5.0%). Điều này đặt ra yêu cầu cho công ty cần đánh giá kỹ hơn từng vai trò công việc cụ thể trong chiến lược giữ người.

Cuối cùng, yếu tố như YearsInCurrentRole chỉ chiếm 3.4%, cho thấy thời gian làm ở vị trí hiện tại không ảnh hưởng lớn bằng các yếu tố khác như mức lương, làm thêm, và môi trường tổng thể.

### PHẦN KẾT LUẬN VÀ ĐỀ XUẤT

1. **Tóm tắt các phát hiện chính:**

Phân tích dữ liệu nhân sự đã chỉ ra một số kết quả quan trọng, bao gồm:

* Tỷ lệ nghỉ việc chung của công ty ở mức đáng chú ý là 16.12%, với 237 trên tổng số 1,470 nhân viên đã nghỉ việc.
* **Phòng ban có tỷ lệ nghỉ việc cao:** Phòng Sales (20.63%) và phòng Nhân sự (19.05%) có tỷ lệ nghỉ việc cao hơn mức trung bình toàn công ty. Ngược lại, phòng Nghiên cứu và Phát triển có tỷ lệ thấp nhất (13.84%).
* **Vai trò có tỷ lệ nghỉ việc cao:** Vị trí Nhân viên Kinh doanh (Sales Representative) có tỷ lệ nghỉ việc cao nhất một cách đáng báo động (39.76%). Các vị trí khác có tỷ lệ cao hơn trung bình bao gồm Kỹ thuật viên Phòng thí nghiệm (Laboratory Technician) (23.94%) và Nhân sự (Human Resources) (23.08%).
* **Các yếu tố chính ảnh hưởng đến quyết định nghỉ việc:**
  + **Tuổi và Thâm niên:** Nhóm nhân viên trẻ (18-25 tuổi) và nhân viên có thâm niên dưới 1 năm có tỷ lệ nghỉ việc cao nhất.
  + **Làm thêm giờ (OverTime):** Nhân viên làm thêm giờ có tỷ lệ nghỉ việc cao hơn rõ rệt. Đặc biệt, 8 trong số 10 nhân viên được dự đoán có nguy cơ nghỉ việc cao nhất đều làm thêm giờ 100%.
  + **Mức độ hài lòng:** Mức độ hài lòng thấp về công việc, môi trường làm việc và sự cân bằng công việc-cuộc sống đều liên quan đến tỷ lệ nghỉ việc cao hơn. Nhân viên có điểm hài lòng công việc và môi trường thấp thường nằm trong nhóm nguy cơ nghỉ việc cao.
  + **Thu nhập:** Tỷ lệ nghỉ việc cao đột biến ở nhóm thu nhập thấp. Nhân viên Kinh doanh và Kỹ thuật viên Phòng thí nghiệm nghỉ việc thường tập trung ở nhóm thu nhập thấp và trung bình thấp.
  + **Giới tính:** Nam giới có xu hướng nghỉ việc nhiều hơn nữ giới (63.29% so với 36.71% trong số những người đã nghỉ).

1. **Đề xuất hành động:**

Dựa trên các phát hiện trên, các đề xuất cụ thể nhằm giảm tỷ lệ nghỉ việc và cải thiện tình hình nhân sự bao gồm:

* **Đối với các phòng ban/vai trò có tỷ lệ nghỉ việc cao (đặc biệt là Sales Representative, Sales Executive, Laboratory Technician, Research Scientist và Human Resources):**
  + Tổ chức các buổi phỏng vấn thôi việc (exit interview) sâu hơn để tìm hiểu nguyên nhân gốc rễ.
  + Xem xét lại khối lượng công việc, cơ hội phát triển nghề nghiệp, chính sách lương thưởng và phúc lợi đặc thù cho các nhóm này.
  + Cải thiện điều kiện làm việc, hỗ trợ tinh thần, tạo động lực và giảm áp lực KPI.
  + Với nhóm Sales Representative, cần hạn chế yêu cầu làm thêm giờ với nhân viên mới, cải thiện môi trường và văn hóa làm việc, cân nhắc điều chỉnh thu nhập khởi điểm và thiết kế chương trình onboarding, mentoring bài bản.
* **Đối với các yếu tố ảnh hưởng như làm thêm giờ, mức độ hài lòng:**
  + Xây dựng chính sách làm thêm giờ hợp lý, khuyến khích cân bằng công việc-cuộc sống.
  + Triển khai các chương trình khảo sát sự hài lòng của nhân viên định kỳ, lắng nghe và cải thiện dựa trên phản hồi.
  + Tăng cường các hoạt động ghi nhận, khen thưởng và xây dựng văn hóa doanh nghiệp tích cực.
  + Đặc biệt quan tâm đến nhóm nhân viên làm thêm giờ và có mức độ hài lòng công việc thấp.
* **Nếu có kết quả từ mô hình dự đoán:**
  + Xây dựng kế hoạch tiếp cận cá nhân hóa và các biện pháp giữ chân chủ động đối với các nhân viên được xác định có nguy cơ nghỉ việc cao. Điều này có thể bao gồm đối thoại 1:1, điều chỉnh khối lượng công việc, cải thiện lương thưởng, nâng cao môi trường làm việc và phát triển nhân viên trẻ.

1. **Hạn chế của phân tích:**

- **Giới hạn của dữ liệu:** Phân tích dựa trên bộ dữ liệu IBM HR Analytics Employee Attrition & Performance Dataset, là dữ liệu mô phỏng. Dữ liệu có thể thiếu một số thông tin quan trọng hoặc chỉ phản ánh tình hình trong một khoảng thời gian nhất định.

- **Phương pháp phân tích:** Mô hình dự đoán có độ chính xác nhất định và kết quả phân tích dựa trên các giả định của mô hình. Các yếu tố được xác định có ảnh hưởng quan trọng (feature importance) dựa trên kết quả của mô hình.