

PHÂN TÍCH CÁC YẾU TỐ ẢNH HƯỞNG ĐẾN SỰ NGHỈ VIỆC CỦA NHÂN VIÊN

NÔ LỆ TƯ BẢN

MỤC LỤC

1 Giới thiệu
về dataset

2 Các công cụ
sử dụng

3 Các bước
thực hiện

4 Các vấn đề
phân tích

5 Giải pháp
vấn đề



“

Tỷ lệ nhảy việc, nghỉ việc trên thị trường lao động hiện nay khiến doanh nghiệp đau đầu, còn nhân sự thì lại không tìm được việc phù hợp. **Vậy, yếu tố nào khiến một trong hai không thể giữ chân bên còn lại?** Có rất ít người bắt đầu và kết thúc sự nghiệp hay công việc của mình ở duy nhất 1 công ty. Phần lớn mọi người đều thay đổi nơi làm việc của mình sau một khoảng thời gian gắn bó với doanh nghiệp. **Liệu họ đang mong muốn tìm kiếm cơ hội thăng tiến mới hay do áp lực từ phía chủ doanh nghiệp?**

”

VỀ DATASET

Bộ dữ liệu được tạo ra để dự đoán tỷ lệ nghỉ việc của nhân viên tại doanh nghiệp dựa trên các thông tin liên quan, bao gồm:

1. Thông tin cá nhân (VD: tuổi, giới tính, tình trạng hôn nhân)
2. Học vị và chuyên ngành
3. Thu nhập hàng tháng
4. Mức độ hài lòng đối với công việc, môi trường làm việc
5. Tính chất công việc

Tên bộ dữ liệu: IBM HR Analytics Employee Attrition & Performance

Định dạng dữ liệu: csv

Nguồn: Kaggle



CÔNG CỤ SỬ DỤNG



python



Power BI

Xử lý và làm sạch dữ liệu trước khi
trực quan hóa dữ liệu

Xây dựng mô hình và
trực quan hóa

CÁC BƯỚC THỰC HIỆN

1. Import dữ liệu vào Python, kiểm tra và làm sạch dữ liệu Null.

```
● ments/PYTHON CLASS$ admin$ /usr/local/bin/python3 "/Users/admin/Docu
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 35 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Age              1470 non-null    int64  
 1   Attrition        1470 non-null    object  
 2   BusinessTravel   1470 non-null    object  
 3   DailyRate        1470 non-null    int64  
 4   Department       1470 non-null    object  
 5   DistanceFromHome 1470 non-null    int64  
 6   Education        1470 non-null    int64  
 7   EducationField   1470 non-null    object  
 8   EmployeeCount    1470 non-null    int64  
 9   EmployeeNumber   1470 non-null    int64  
 10  EnvironmentSatisfaction 1470 non-null    int64  
 11  Gender            1470 non-null    object  
 12  HourlyRate       1470 non-null    int64  
 13  JobInvolvement  1470 non-null    int64  
 14  JobLevel         1470 non-null    int64  
 15  JobRole          1470 non-null    object  
 16  JobSatisfaction 1470 non-null    int64  
 17  MaritalStatus    1470 non-null    object  
 18  MonthlyIncome    1470 non-null    int64  
 19  MonthlyRate      1470 non-null    int64  
 20  NumCompaniesWorked 1470 non-null    int64  
 21  Over18           1470 non-null    object  
 22  Overtime          1470 non-null    object  
 23  PercentSalaryHike 1470 non-null    int64  
 24  PerformanceRating 1470 non-null    int64  
 25  RelationshipSatisfaction 1470 non-null    int64  
 26  StandardHours    1470 non-null    int64  
 27  StockOptionLevel 1470 non-null    int64  
 28  TotalWorkingYears 1470 non-null    int64  
 29  TrainingTimesLastYear 1470 non-null    int64  
 30  WorkLifeBalance  1470 non-null    int64  
 31  YearsAtCompany   1470 non-null    int64  
 32  YearsInCurrentRole 1470 non-null    int64  
 33  YearsSinceLastPromotion 1470 non-null    int64  
 34  YearsWithCurrManager 1470 non-null    int64  
dtypes: int64(26), object(9)
memory_usage: 402.1+ KB
```

2/ Dùng lệnh Drop column để loại bỏ các cột không liên quan (Over18, Standard hour). Xuất file csv sau khi hoàn thành xóa cột.

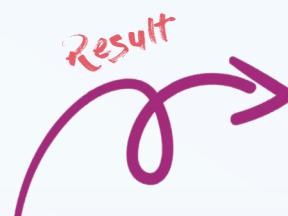
#	Column	Non-Null Count	Dtype
0	EmployeeID	1470 non-null	int64
1	Age	1470 non-null	int64
2	Attrition	1470 non-null	object
3	BusinessTravel	1470 non-null	object
4	DailyRate	1470 non-null	int64
5	Department	1470 non-null	object
6	DistanceFromHome	1470 non-null	int64
7	Education	1470 non-null	int64
8	EducationField	1470 non-null	object
9	EmployeeCount	1470 non-null	int64
10	EnvironmentSatisfaction	1470 non-null	int64
11	Gender	1470 non-null	object
12	HourlyRate	1470 non-null	int64
13	JobInvolvement	1470 non-null	int64
14	JobLevel	1470 non-null	int64
15	JobRole	1470 non-null	object
16	JobSatisfaction	1470 non-null	int64
17	MaritalStatus	1470 non-null	object
18	MonthlyIncome	1470 non-null	int64
19	MonthlyRate	1470 non-null	int64
20	NumCompaniesWorked	1470 non-null	int64
21	Overtime	1470 non-null	object
22	PercentSalaryHike	1470 non-null	int64
23	PerformanceRating	1470 non-null	int64
24	RelationshipSatisfaction	1470 non-null	int64
25	StockOptionLevel	1470 non-null	int64
26	TotalWorkingYears	1470 non-null	int64
27	TrainingTimesLastYear	1470 non-null	int64
28	WorkLifeBalance	1470 non-null	int64
29	YearsAtCompany	1470 non-null	int64
30	YearsInCurrentRole	1470 non-null	int64
31	YearsSinceLastPromotion	1470 non-null	int64
32	YearsWithCurrManager	1470 non-null	int64

CÁC BƯỚC THỰC HIỆN

3/ Import file csv từ Python vào DBeaver. Sau đó, đối chiếu với bộ dữ liệu, dùng CASE WHEN để chuyển đổi các mức độ đánh giá từ số sang chuỗi..

```

129 •   select
130   case
131     when education = 1 then "Below College"
132     when education = 2 then "College"
133     when education = 3 then "Bachelor"
134     when education = 4 then "Master"
135     when education = 5 then "Doctor"
136   end "Education_Rating",
137   case
138     when EnvironmentSatisfaction = 1 then "Low"
139     when EnvironmentSatisfaction = 2 then "Medium"
140     when EnvironmentSatisfaction = 3 then "High"
141     when EnvironmentSatisfaction = 4 then "Very High"
142   end "Environment_Satisfaction",
143   case
144     when JobInvolvement =1 then 'Low'
145     when JobInvolvement =2 then 'Medium'
146     when JobInvolvement =3 then 'High'
147     when JobInvolvement =4 then 'Very High'
148   end "Job_Involvement",
149   case
150     when JobLevel =1 then "Entry-level"
151     when JobLevel =2 then "Intermediate or experienced (senior staff)"
152     when JobLevel =3 then "First-level management"
153     when JobLevel =4 then "Middle management"
154     when JobLevel =5 then "Executive or senior management"
155   end as "Job_Level",
156   case
157     when JobSatisfaction = 1 then 'Low'
158     when JobSatisfaction = 2 then 'Medium'
159     when JobSatisfaction = 3 then 'High'
160     when JobSatisfaction = 4 then 'Very High'
161   end "Job_Satisfaction",
162   case
163     when PerformanceRating = 1 then 'Low'
164     when PerformanceRating = 2 then 'Good'
165     when PerformanceRating = 3 then 'Excellent'
166     else 'Outstanding'
167   end as "Performance Rating"
168   ,case
169     when RelationshipSatisfaction = 1 then 'Low'
170     when RelationshipSatisfaction = 2 then 'Medium'
171     when RelationshipSatisfaction = 3 then 'High'
172     else 'Very High'
173   end as RelationshipSatisfaction,
174   case
175     when WorkLifeBalance=1 then "Bad"
176     when WorkLifeBalance=2 then "Good"
177     when WorkLifeBalance=3 then "Better"
178     when WorkLifeBalance=4 then "Best"
179   end as WorkLifeBalance
180   from hr_employee_attrition
  
```

GIỚI THIỆU VỀ
DATASETCÁC CÔNG CỤ
SỬ DỤNGCÁC BƯỚC
THỰC HIỆNCÁC VẤN ĐỀ
PHÂN TÍCHGIẢI PHÁP
VẤN ĐỀ

Education	Environment	JobInvolvement	JobLevel	JobSatisfia	Performance	Re
2	2	3	2	4	3	
1	3	2	2	2	4	
2	4	2	1	3	3	
4	4	3	1	3	3	
1	1	3	1	2	3	
2	4	3	1	4	3	
3	3	4	1	1	4	
1	4	3	1	3	4	
3	4	2	3	3	4	
3	3	3	2	3	3	
3	1	4	1	2	3	
2	4	2	2	3	3	
1	1	3	1	3	3	
2	2	3	1	4	3	
3	3	2	1	3	3	
4	2	4	3	1	3	
2	1	4	1	2	3	

Education_Rati	Environment	Job_Involv	JobLevel	Job_Satisfi	Performance	RelationshipSa	WorkLifeB'
College	Medium	High	Intermediate or experien	Very High	Excellent	Low	Bad
Below College	High	Medium	Intermediate or experien	Medium	Outstanding	Very High	Better
College	Very High	Medium	Entry-level	High	Excellent	Medium	Better
Master	Very High	High	Entry-level	High	Excellent	High	Better
Below College	Low	High	Entry-level	Medium	Excellent	Very High	Better
College	Very High	High	Entry-level	Very High	Excellent	High	Good
Bachelor	High	Very High	Entry-level	Low	Outstanding	Low	Good
Below College	Very High	High	Entry-level	High	Outstanding	Medium	Better
Bachelor	Very High	Medium	First-level management	High	Outstanding	Medium	Better
Bachelor	High	High	Intermediate or experien	High	Excellent	Medium	Good
Bachelor	Low	Very High	Entry-level	Medium	Excellent	High	Better
College	Very High	Medium	Intermediate or experien	High	Excellent	Very High	Better
Below College	Low	High	Entry-level	High	Excellent	Very High	Good
College	Medium	High	Entry-level	Very High	Excellent	High	Better
Bachelor	High	Medium	Entry-level	High	Excellent	Medium	Better
Master	Medium	Very High	First-level management	Low	Excellent	High	Better
College	Low	Very High	Entry-level	Medium	Excellent	Very High	Good
College	Very High	Very High	Entry-level	Very High	Excellent	Medium	Good

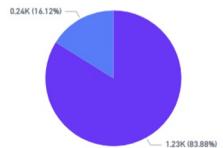
our dashboard

Total Number of Employees

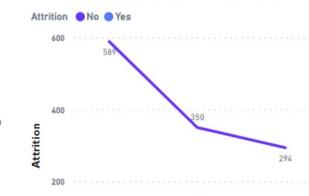
1470



Percent of Attrition



Attrition by Marital Status



Attrition by Education



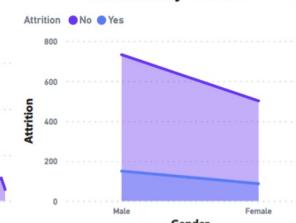
Attrition by Job Role



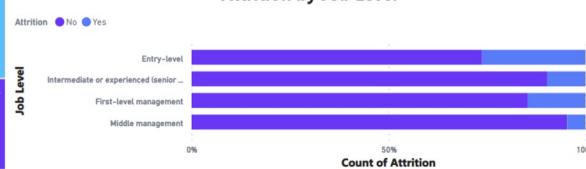
Attrition by Age



Attrition by Gender



Attrition by Job Level



WOE AND IV

Step 1

```
import numpy as np
import pandas as pd
```

Step 2

```
def calculate_woe_iv(data, target, var):
    """
    Calculate the WOE and IV for a categorical variable.
    """
    grouped = data.groupby(var, as_index=False)
    agg = grouped.agg({target: ['count', 'sum']})
    agg.columns = [var, 'total', 'bad']
    agg['good'] = agg['total'] - agg['bad']
    agg['good_pct'] = agg['good'] / agg['good'].sum()
    agg['bad_pct'] = agg['bad'] / agg['bad'].sum()
    agg['woe'] = np.log(agg['good_pct'] / agg['bad_pct'])
    agg['iv'] = (agg['good_pct'] - agg['bad_pct']) * agg['woe']
    return agg[[var, 'woe', 'iv']]
```

Step 3

```
# create a dictionary to map genders and attrition to numeric values
gender_to_num = {"Male": 0, "Female": 1, "Non-binary": 2}
attrition_to_num = {"Yes": 1, "No": 0}
# convert the gender and Attrition column to numeric values
data["Gender"] = data["Gender"].map(gender_to_num)
data["Attrition"] = data["Attrition"].map(attrition_to_num)
```

Step 4

```
data = pd.read_csv("hr_employee_attrition.csv")
```

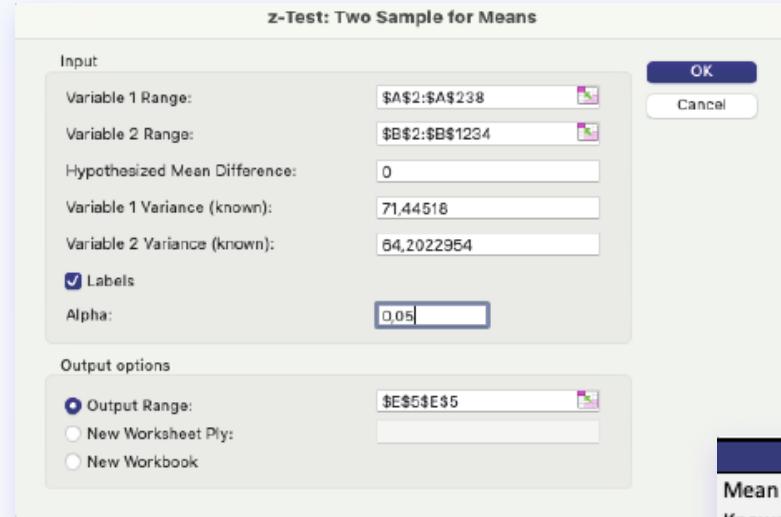
Step 5

```
calculate_woe_iv(data,'Attrition', 'Gender')
```

Gender	woe	iv
0	-0.064000	0.002511
1	0.101553	0.003985

Z-TEST

Step 1



Step 2

	Variable 1	Variable 2
Mean	10,63291139	8,915652879
Known Variance	71,44518	64,2022954
Observations	237	1233
Hypothesized Mean	0	
z	2,888183122	
P(Z<=z) one-tail	0,001937371	
z Critical one-tail	1,644853627	
P(Z<=z) two-tail	0,003874742	
z Critical two-tail	1,959963985	

RESULT

WOE & IV

Z-TEST

CÁC YẾU TỐ CẦN PHÂN TÍCH	LOẠI DỮ LIỆU	INFORMATION VALUE (IV)
RelationshipSatisfaction	var	0,024
Department	var	0,05
WorkLifeBalance	var	0,07
Environment_Satisfaction	var	0,1
Job_Satisfaction	var	0,1
BusinessTravel	var	0,12
Job_Involvement	var	0,13
MaritalStatus	var	0,22
JobLevel	var	0,4
OverTime	var	0,4

CÁC YẾU TỐ CẦN PHÂN TÍCH	LOẠI DỮ LIỆU	Z-VALUES
Age	numberic	(5,82)
DistanceFromHome	numberic	2,89
MonthlyIncome	numberic	7,48
TotalWorkingYears	numberic	7,01
YearsWithCurrManager	numberic	6,63
YearsAtCompany	numberic	(5,28)
YearsInCurrentRole	numberic	6,84

Các nhà quản lý

ảnh hưởng như thế nào đến **sự nghỉ việc** của nhân viên?

%GT Attrition(Yes) by YearsWithCurrManager



%GT Attrition(No) by YearsWithCurrManager

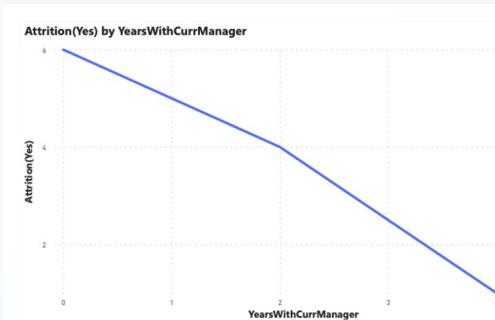


Dựa vào biểu đồ dễ dàng nhận ra có 3 đợt tăng đáng kể về tỷ lệ nghỉ việc. Ở năm đầu tiên, tỷ lệ % nghỉ việc **biến động mạnh** lên tới **35,85%**. Sau đó tỷ lệ nghỉ việc bắt đầu giảm nhẹ theo thời gian ở hai đợt tiếp theo Đợt 1 (Từ năm thứ nhất tới năm thứ 3) và Đợt 2 (Từ năm thứ 6 tới năm thứ 7)

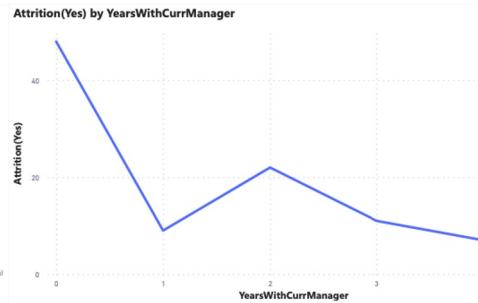
Các nhà quản lý ở từng phòng ban

ảnh hưởng như thế nào đến sự nghỉ việc của nhân viên?

Điều gì đã tác động đến số lượng nhân viên nghỉ việc nhiều trong khoảng thời gian dưới 5 năm của nhân viên dưới sự tác động của nhà quản lý ?



HR Department



R&D Department

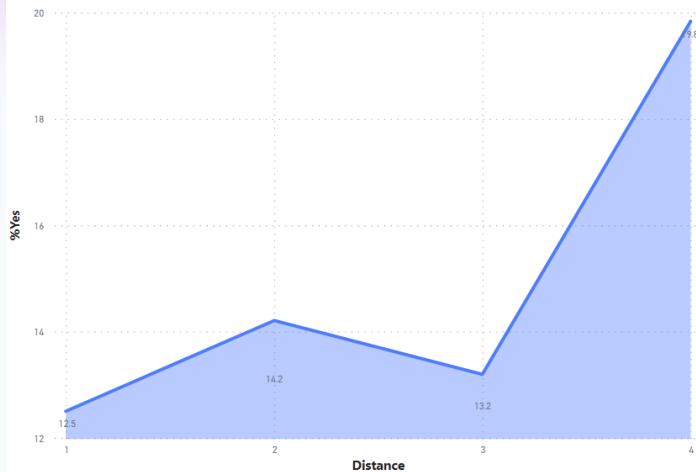


Sales Department

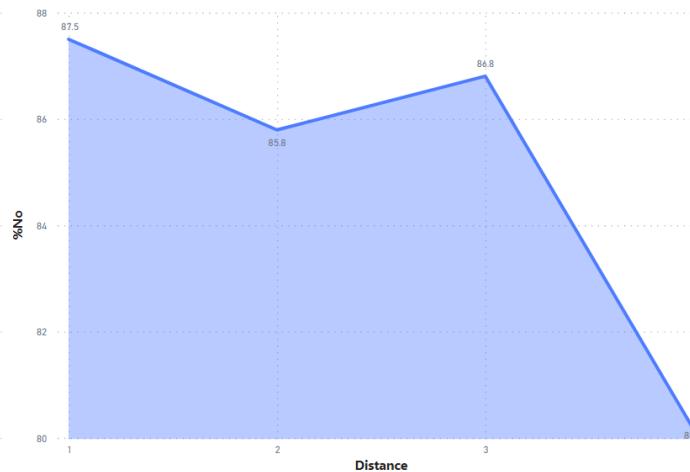
Dường như do tính chất quản lý, sự áp lực của chuyên môn của từng phòng ban đã tác động đến số lượng nhân viên nghỉ việc.

Khoảng cách từ nhà đến công ty ảnh hưởng như thế nào đến sự nghỉ việc của nhân viên?

%Yes by Distance



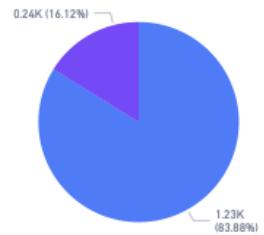
%No by Distance



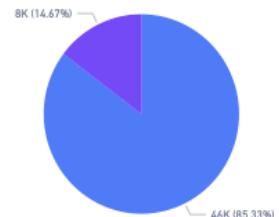
- Những nhân viên sống xa công ty của họ nhất có **xu hướng nghỉ việc nhiều nhất** lên tới **19,8%**
- Khoảng cách là một yếu tố quan trọng ảnh hưởng đến sự hao hụt nhân viên

Độ tuổi ảnh hưởng như thế nào đến sự nghỉ việc của nhân viên?

Count of EmployeeID by Attrition

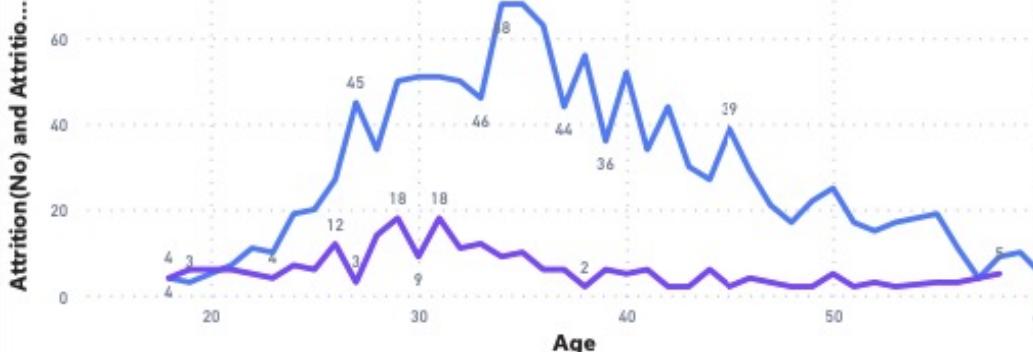


Sum of Age by Attrition



Attrition(No) and Attrition(Yes) by Age

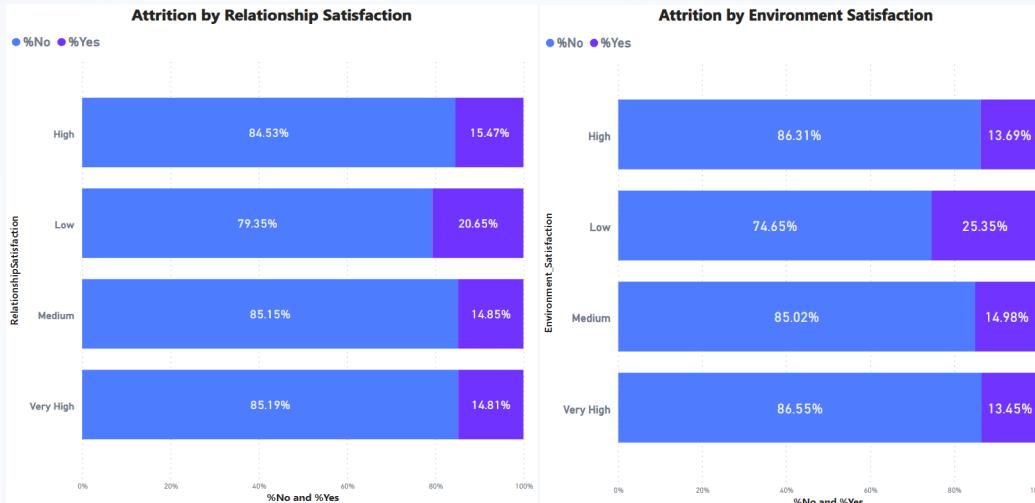
Attrition(No) Attrition(Yes)



- Tỷ lệ nghỉ việc **cao nhất** trong khoảng từ 28 đến 36.
- Khi mọi người già đi, tỷ lệ nghỉ **giảm** vì họ tìm kiếm sự ổn định trong công việc.

Môi trường làm việc

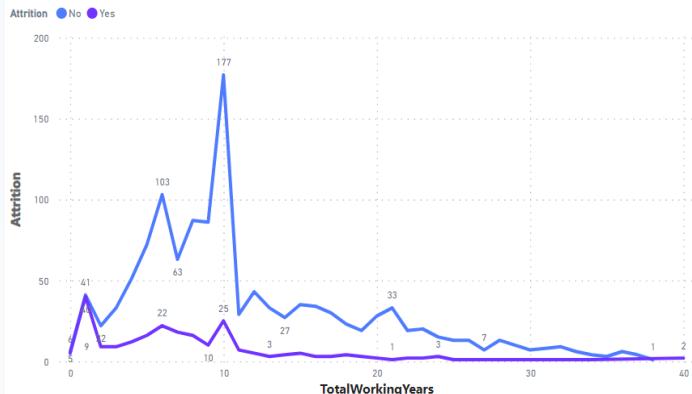
ảnh hưởng như thế nào đến sự nghỉ việc của nhân viên?



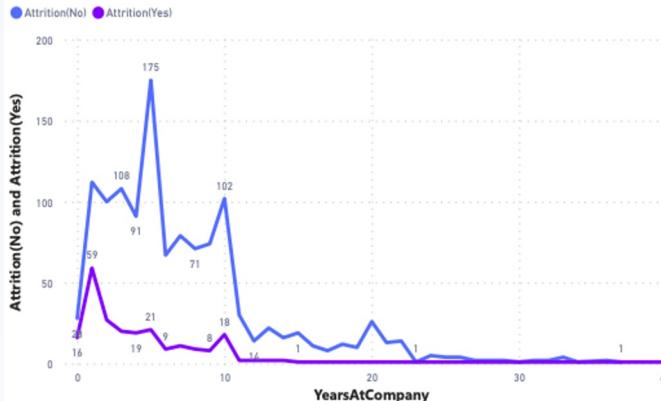
- Ở các mức độ hài lòng thấp và trung (Low và Medium) khả năng mọi người **rời bỏ tổ chức cao** (20,65% và 15,47%).
- Ngược lại, ở các mức độ hài lòng cao đến rất cao (High sang Very High), phần lớn nhân viên có xu hướng **gắn bó với công ty lâu hơn**.

Yếu tố thể hiện sự cam kết với công ty ảnh hưởng như thế nào đến **sự nghỉ việc** của nhân viên?

Attrition (No) and Attrition (Yes) by Total Working Years



Attrition(No) and Attrition(Yes) by YearsAtCompany

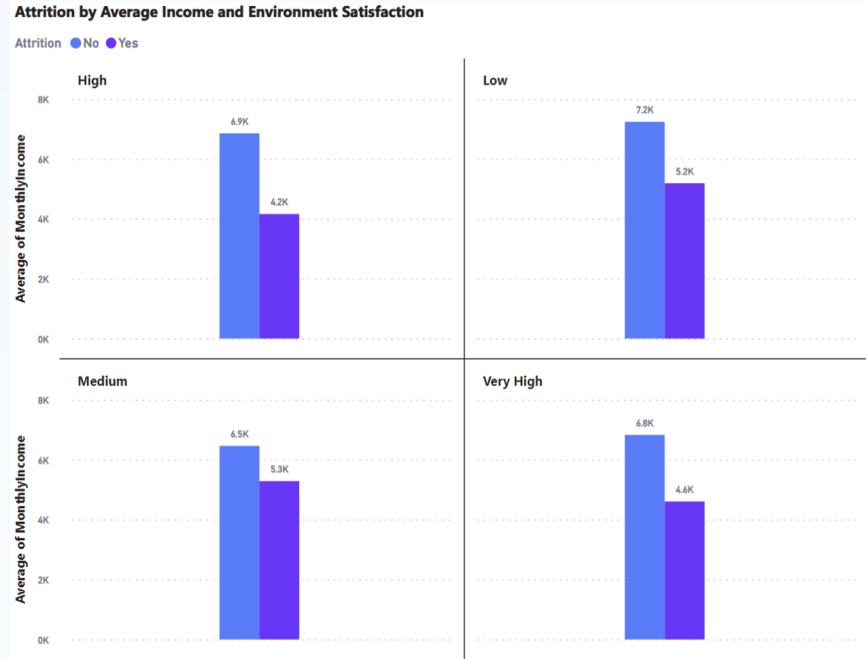


- Nhân viên bắt đầu sự nghiệp của họ với công ty và những người mới bắt đầu làm việc trong công ty, có xu hướng rời bỏ tổ chức cao hơn sang một công ty khác.
- Những người đã dành nhiều năm trong công ty và những người đã có nhiều kinh nghiệm khi làm việc ở nhiều công ty có xu hướng ở lại.

Yếu tố nào

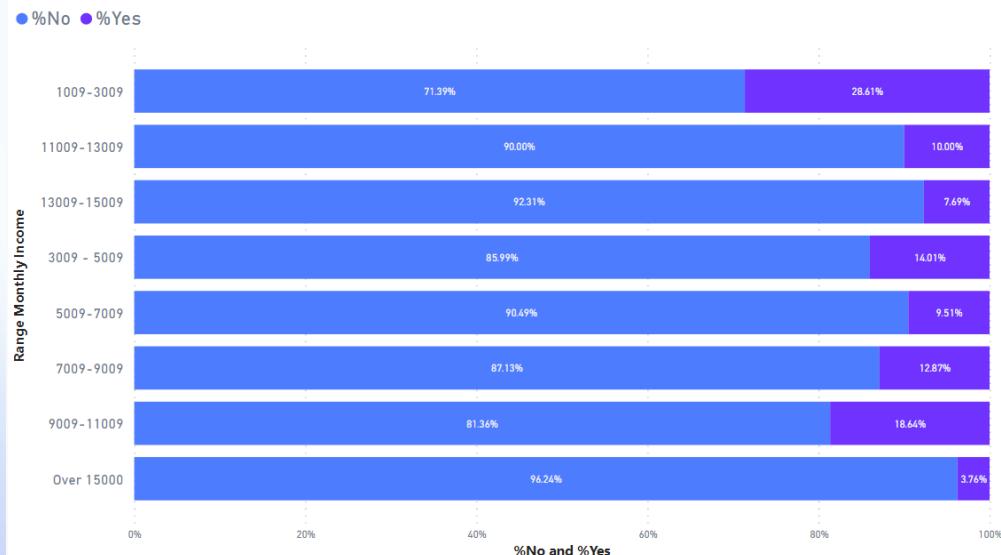
khiến nhân viên có độ hài lòng cao lại quyết định nghỉ việc?

- Thu nhập hàng tháng của nhóm nhân viên có độ hài lòng cao lại thấp hơn (khoảng 23%) so với 2 nhóm còn lại
- Mức lương thấp là **yếu tố quan trọng** khiến họ quyết định rời khỏi công ty



Liệu thu nhập thấp có phải là nguyên nhân dẫn đến tỷ lệ nghỉ việc tăng?

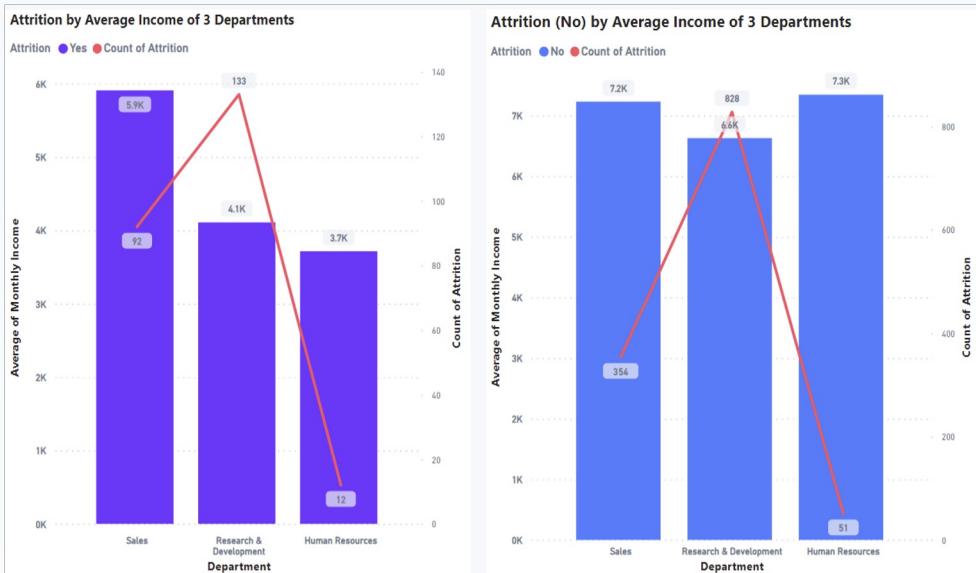
Attrition by Range Monthly Income



Thu nhập hàng tháng được chia thành 8 nhóm, từ \$1009 đến hơn \$15000.

- Nhìn chung, thu nhập càng cao, tỷ lệ nghỉ việc càng thấp.
- Tuy nhiên, nhóm \$9009 - \$11009 thu nhập cao nhưng tỷ lệ nghỉ việc cao thứ 2 so với các nhóm còn lại.

Thu nhập trung bình có ảnh hưởng đến tỷ lệ nghỉ việc theo từng phòng ban?

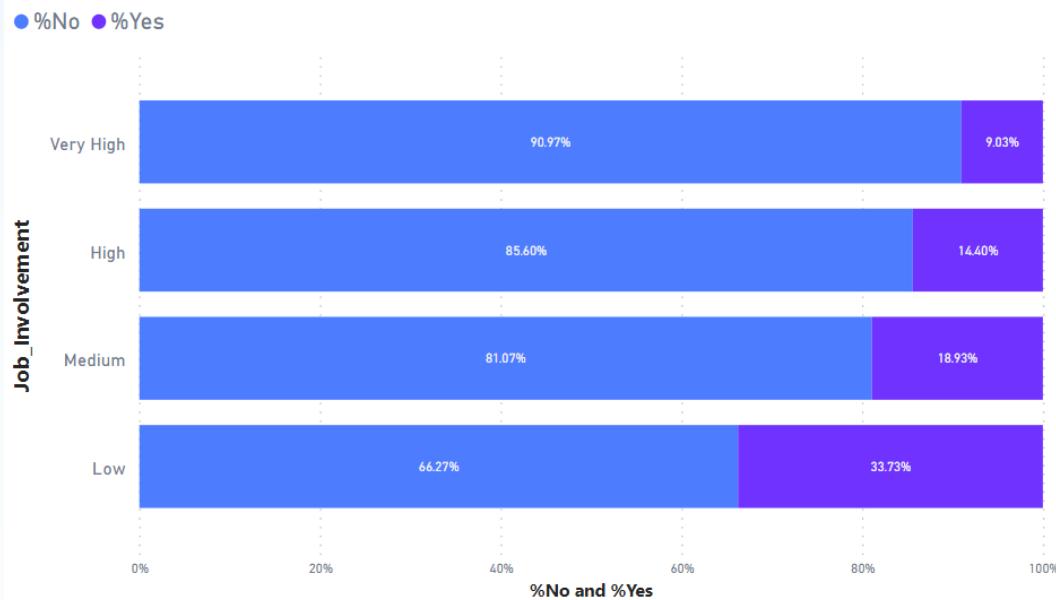


- HR có **mức lương khá ổn định**, tỷ lệ nghỉ việc cũng ở **mức thấp nhất**.
- Sales có **mức thu nhập cao nhất**, số **nhân viên nghỉ việc nhiều hơn**.

Vậy ngoài ảnh hưởng từ mức lương hàng tháng, yếu tố nào khiến nhân viên phòng Sales nghỉ nhiều hơn R&D và HR?

Nhân viên nghỉ việc có phải do họ không cống hiến hết mình?

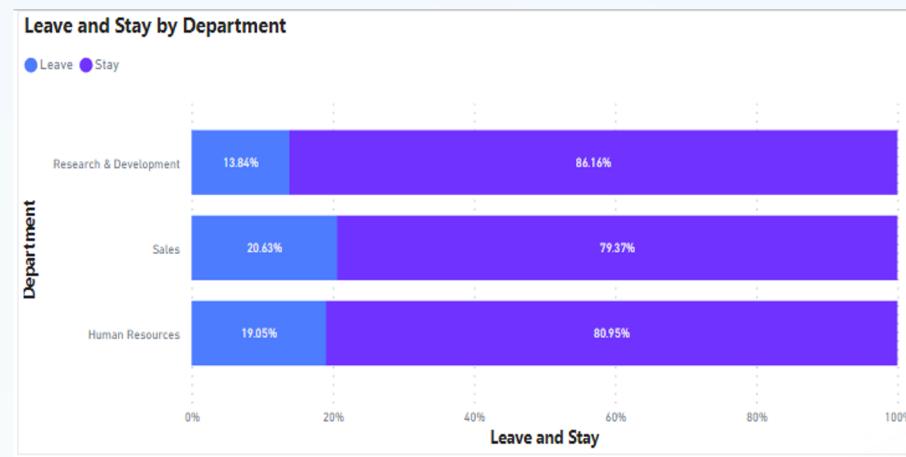
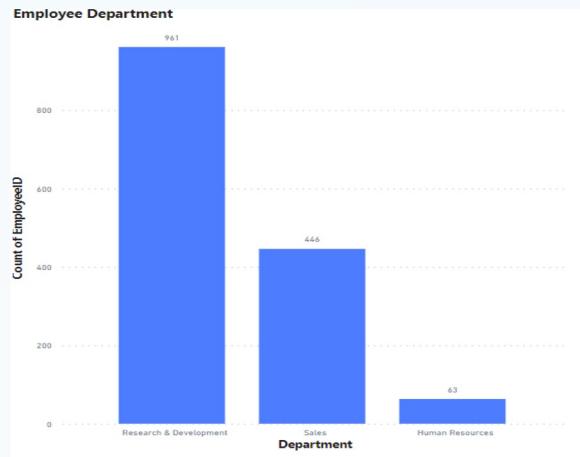
Attrition by Job Involvement



- Mức độ cống hiến của nhân viên được chia thành 4 nhóm, từ thấp đến rất cao.
- Biểu đồ cho thấy nhân viên thể hiện sự đóng góp càng thấp, **xu hướng nghỉ việc của họ càng cao** và ngược lại.

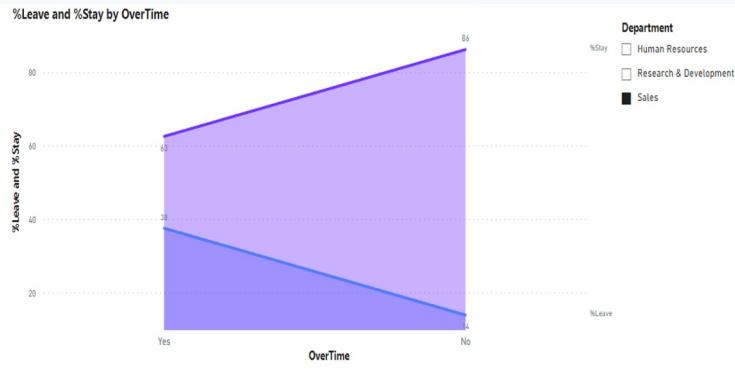
Tính chất công việc

ảnh hưởng như thế nào đến sự nghỉ việc của nhân viên?



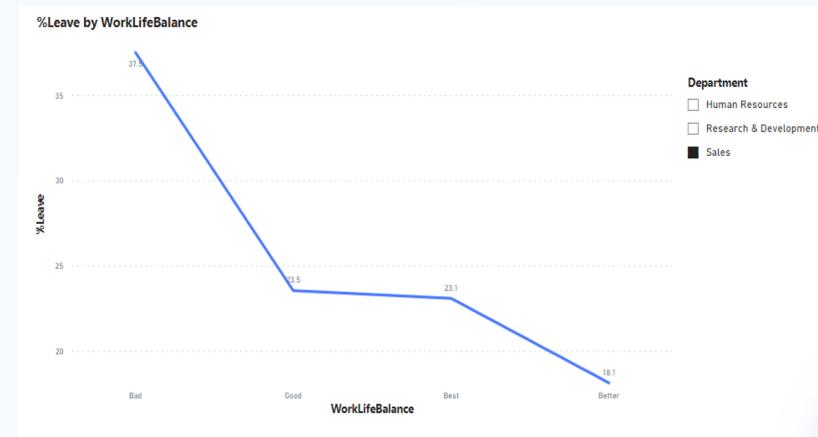
Sales là phòng ban chịu nhiều áp lực từ các mốc KPI, doanh số nên không khó hiểu khi có tỷ lệ nghỉ việc cao nhất (**20,63%**), tiếp theo là Phòng Nhân sự (**19,05%**). Mặt khác, Phòng R&D, tuy có số lượng nhân sự cao nhất nhưng lại có tỷ lệ nghỉ việc ít nhất, điều này cho thấy sự ổn định nhân sự của bộ phận (**13,83%**).

Các yếu tố trong công việc tác động đến sự nghỉ việc của nhân viên?



Tăng ca

Tăng ca cũng chính là 1 trong những nguyên nhân chính dẫn đến quyết định xin nghỉ việc của nhân viên



Cân bằng cuộc sống

Những nhân viên có cân bằng giữa công việc và cuộc sống ở mức kém có xu hướng nghỉ việc nhiều hơn

Cân bằng giữa cuộc sống và công việc

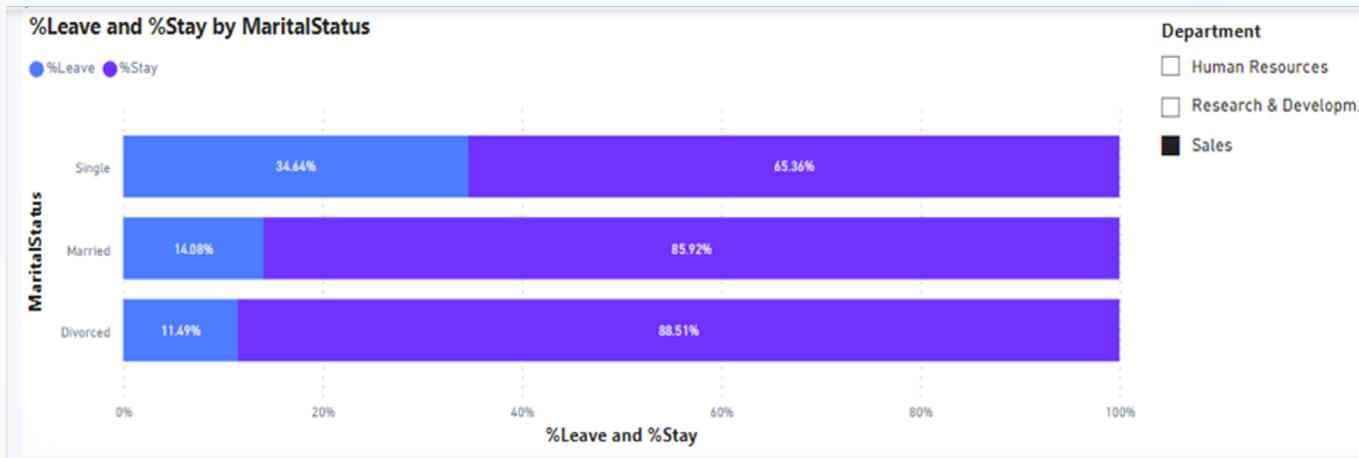
có phải yếu tố ảnh hưởng đến quyết định nghỉ việc của nhân viên?



- 40% vẫn quyết định nghỉ việc khi sự cân bằng trong cuộc sống là tốt.
- Khi việc cân bằng giữa cuộc sống và công việc không tốt thì sẽ có xu hướng nghỉ việc với tỉ lệ 42% nghỉ việc dù họ không phải tăng ca

Tình trạng hôn nhân

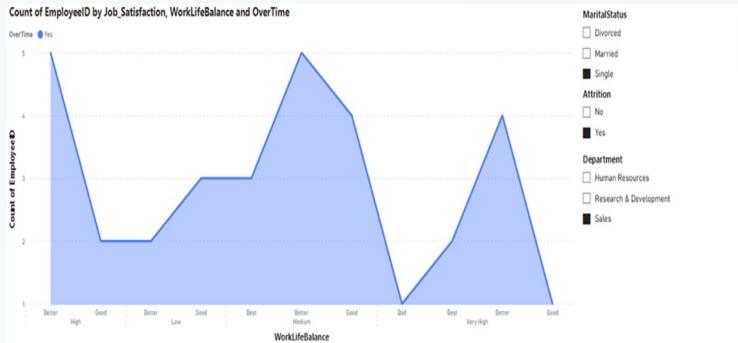
tác động như thế nào đến **sự nghỉ việc** của nhân viên?



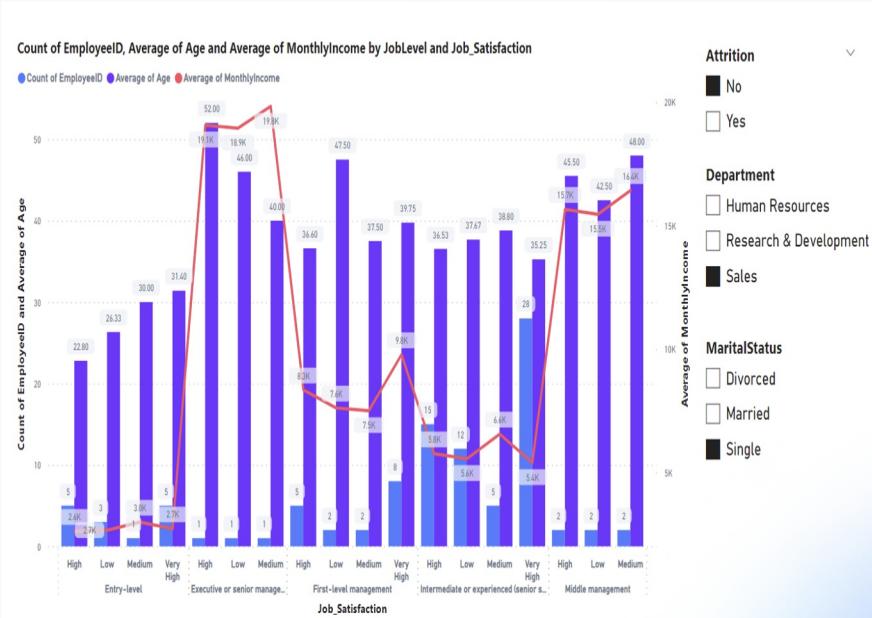
Những nhân viên đã có gia đình và ly hôn thì họ có tỷ lệ rời đi ít hơn khi lần lượt là **14,08%** và **11,49%**.

Các yếu tố trong công việc

tác động như thế nào đến **sự nghỉ việc** của nhân viên?



Đa số tập trung ở độ tuổi 24-37, cấp bậc Entry-level và Intermediate or experienced (senior staff) mức thu nhập cũng ở mức trung bình thấp => nhóm đưa ra dự đoán như đề cập ở phần mở đầu, **ra đi để tìm cơ hội tốt hơn, lương cao hơn**



Công tác

tác động như thế nào đến **sự nghỉ việc** của nhân viên?

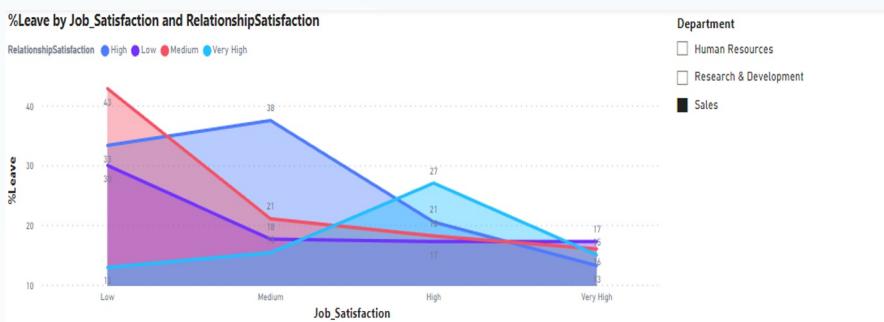
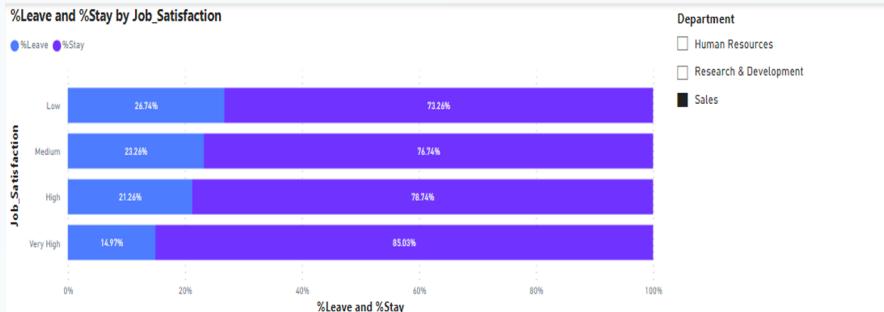


Xét giá trị tuyệt đối, phần lớn **nhân viên nghỉ việc** đến từ nhóm **ít đi công tác** (60 lao động).

Xét giá trị phần trăm, nhóm **đi công tác thường xuyên** **nghỉ nhiều hơn** so với 2 nhóm còn lại (khoảng 33%).

Các yếu tố trong công việc

tác động như thế nào đến **sự nghỉ việc** của nhân viên?



- Sự hài lòng trong công việc ngày càng tăng, tỷ lệ nghỉ việc càng giảm
- Tỷ lệ nghỉ việc cao nhất ở mức **Low** với **26,7 %**.
- Ngoài các yếu tố như phải thường xuyên đi công tác cũng như làm việc tăng ca thì việc nhân viên cảm thấy không thoải mái với công việc hiện tại cũng có thể xuất phát từ việc không có mối quan hệ tốt với đồng nghiệp hoặc quản lý.

LOGISTIC REGRESSION

Model

```
Age           int64
Attrition    object
BusinessTravel object
Department   object
DistanceFromHome int64
Environment_Satisfaction object
Job_Involvement object
JobLevel     object
Job_Satisfaction object
MaritalStatus object
MonthlyIncome int64
OverTime      object
RelationshipSatisfaction object
TotalWorkingYears int64
WorkLifeBalance object
YearsAtCompany int64
YearsInCurrentRole int64
YearsWithCurrManager int64
dtype: object
```

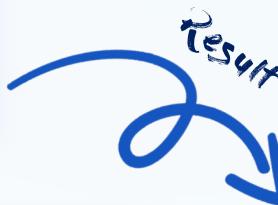
Step 1

```
[5] yes_no_columns=['OverTime','Attrition']
for col in yes_no_columns:
    df1[col].replace({'Yes':1, 'No':0},inplace=True)
```

```
df2=pd.get_dummies(data=df1,columns=['Maritalstatus','JobLevel',
                                      'Department','BusinessTravel',
                                      'WorkLifeBalance','Environment_Satisfaction',
                                      'Job_Involvement','Job_Satisfaction',
                                      'RelationshipSatisfaction'])
```

df2.columns

Step 2



Step 3

```
[36] cols_to_scale=['Age','DistanceFromHome','MonthlyIncome',
                    'TotalWorkingYears','YearsAtCompany',
                    'YearsInCurrentRole','YearsWithCurrManager']
from sklearn.preprocessing import MinMaxScaler
scaler=MinMaxScaler()
df2[cols_to_scale]=scaler.fit_transform(df2[cols_to_scale])
```

```
[16] X=df2.drop('Attrition',axis='columns')
ytestLabels=df2.Attrition.astype(np.float32)
from imblearn.over_sampling import SMOTE, SVMSMOTE
sm = SMOTE()
X_train_resample, y_train_resample = sm.fit_resample(X,y)
X_train,X_test,y_train,y_test=train_test_split(X_train_resample, y_train_resample,test_size=0.2,random_state=70)

model=LogisticRegression()
model.fit(X_train,y_train)
acc=model.score(X_test,y_test)
print("Accuracy",acc,"\\n")
y_pred=model.predict(X_test)
print("Preds",y_pred,"\\n")
cl_rep=classification_report(y_test,y_pred)
print(cl_rep)
```

Step 4

classification report

	precision	recall	f1-score	support
0.0	0.88	0.97	0.92	243
1.0	0.96	0.87	0.92	251
accuracy			0.92	494
macro avg	0.92	0.92	0.92	494
weighted avg	0.92	0.92	0.92	494

LOGISTIC REGRESSION

$$Y = W_0 + W_1 * x_1 + \dots + W_n * x_n$$

$$W_0 = 16.23933784$$

Xn: các giá trị đầu vào

Wx

[-0.7482178 , 0.66018137, -1.10940806, 1.33762384, -0.17559184, 1.22144384, -
 1.12331773, -1.01746384, -2.67427847, -2.24195229, -1.32842794, -1.05084404, -
 1.0135989 , -1.32982526, -2.38360717, -2.23772682, -1.46023095, -2.14368201, -
 1.25132983, -2.3760345 , -1.03003554, -1.4743937 , -1.26277859, -2.2489297 , -
 2.40607259, -2.19889462, -2.33530066, -1.54366346, -2.34324903, -2.39291195, -
 2.03115126, -0.68808292, -1.7659241 , -2.48954835, -2.20229476, -1.64445071, -
 2.22255368, -2.84353742, -2.16574177, -1.63935684, -2.23026843, -2.12448453]



$$P_{\text{leave}} = \frac{1}{1+e^{-y}}$$

P_{leave}
 < 0.5

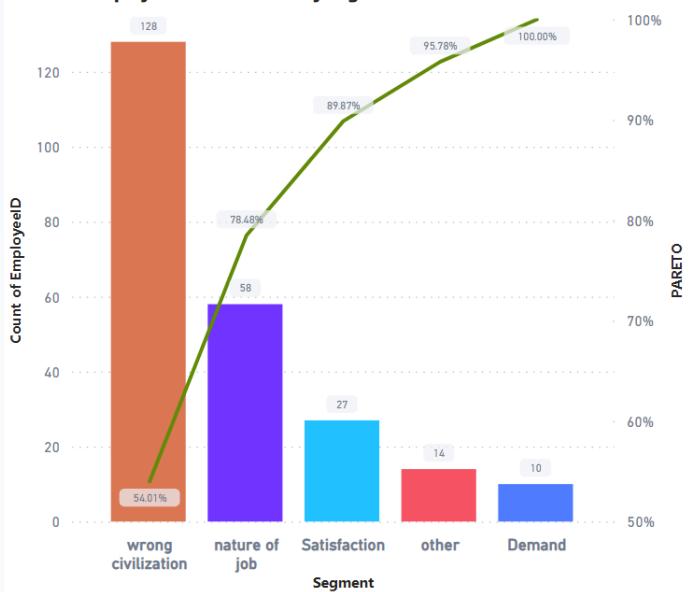
- Nhân viên có nguy cơ nghỉ việc

P_{leave}
 >= 0.5

- Nhân viên chưa có nguy cơ nghỉ việc

Các yếu tố cần tập trung xử lý để giảm thiểu sự nghỉ việc của nhân viên?

Count of EmployeeID and PARETO by Segment



PARETO by Segment

Segment ● Demand ● nature of job ● other ● Satisfaction ● wrong civilization



5 phân khúc đại diện cho 5 yếu tố cần
tập trung xử lý.

Trong đó, **Wrong Civilization và Nature
of Job** là 2 yếu tố cần được cải thiện
nhất.

ĐỀ XUẤT GIẢI PHÁP

“

Doanh nghiệp nên hỗ trợ về mặt vật chất lẫn tinh thần để giữ chân nhân viên.

Thực hiện khảo sát ý kiến nhân viên để hiểu rõ và đáp ứng nhu cầu của họ.

Xây dựng môi trường làm việc tích cực, linh hoạt.

Điều chỉnh mức lương và chính sách đãi ngộ phù hợp với từng vị trí, năng lực.

Bên cạnh đó, quy trình tuyển dụng cần được cải thiện để sàng lọc ứng viên tiềm năng hơn.

”

THANK YOU