

# 2023



## HR ANALYTIC

Prepared By:

Lan Dung Do



# TABLE OF CONTENT

- I. Introduction
- II. Problem Statement and Attribute
- III. Cleaning Data
- IV. Insights
- V. Multiple Regression
- VI. Business Insights

# I. INTRODUCTION

--> Human resources (HR) are essential to the success of any organization, as they play a critical role in attracting, retaining, and developing talent. HR is responsible for managing an organization's most valuable asset - its employees - and ensuring that they are motivated, engaged, and aligned with the organization's goals and values. Effective HR practices can help organizations create a culture of innovation and growth, build high-performing teams, and foster a positive work environment that promotes productivity and collaboration. Moreover, HR is responsible for ensuring compliance with labor laws, regulations, and policies, which helps to protect both the organization and its employees. In today's highly competitive business environment, organizations that invest in HR and prioritize their employees' well-being are more likely to succeed in the long run. Therefore, it is imperative for organizations to recognize the importance of HR and to invest in building a strong HR function.

--> Turnover rate can provide several valuable business insights, including:

- Retention issues: High turnover rates can be a sign of retention issues within the organization. By analyzing the reasons why employees are leaving, businesses can identify the root cause of these retention issues and take steps to improve employee engagement, job satisfaction, and retention.
- Recruitment and training costs: High turnover rates can be costly for organizations, as they result in increased recruitment and training costs. By reducing turnover rates, businesses can save money on recruitment and training, and invest these savings into other areas of the organization.
- Employee productivity: High turnover rates can also have a negative impact on employee productivity, as it takes time for new hires to get up to speed and become productive members of the team. By reducing turnover rates, businesses can improve employee productivity and ensure that their workforce is operating at peak efficiency.
- Reputation and brand image: High turnover rates can also have a negative impact on a business's reputation and brand image. This can make it more difficult to attract and retain top talent, and can ultimately impact the bottom line.
- Skills and knowledge loss: When employees leave, they take their skills and knowledge with them. This can be particularly damaging if the employee is a key player in the organization or has specialized knowledge or skills. By reducing turnover rates, businesses can ensure that they retain their top talent and the knowledge and skills that come with them.

## II. PROBLEM STATEMENT AND ATTRIBUTES

1

### Problem Statement:

Analyzing Turnover of a Company Data

The aim of this project is to analyze the turnover data of a company and identify the key factors that contribute to employee turnover. Employee turnover is a major issue that affects many organizations, as it can lead to increased costs associated with recruitment, training, and lost productivity.

The primary objective of this project is to understand the underlying factors that contribute to employee turnover and to identify potential strategies for reducing turnover rates. To achieve this objective, we will perform exploratory data analysis to gain insights into the data and to identify patterns and trends.

The project will involve collecting and cleaning employee turnover data, as well as conducting statistical analysis and building predictive models to identify the key factors that influence turnover. We will also examine the relationships between different variables and identify any correlations that exist between them.

The project will be guided by the following research questions:

1. What are the key factors that contribute to employee turnover?
2. What is the relationship between different variables and employee turnover?
3. Can we predict employee turnover based on specific variables?
4. What are the potential strategies for reducing employee turnover?

The outcomes of this project will be useful for human resources professionals and organizational leaders who are interested in understanding the factors that contribute to employee turnover and developing strategies for reducing turnover rates.

2

### ATTRIBUTES

- 1)SATISFACTION\_LEVEL
- 2)LAST\_EVALUATION
- 3)NUMBER\_PROJECT
- 4) AVERAGE\_MONTHLY\_HOURS
- 5)TIME\_SPEND\_COMPANY
- 6)WORK\_ACCIDENT
- 7)PROMOTION\_LAST\_5YEARS
- 8)DEPARTMENT
- 9)SALARY
- 10)TARGET

03

LAN DUNG DO

# III.CLEANING DATA

- 1 Identify and remove duplicate data

```
#Data contain duplicated value that needed to remove before analyze  
df.duplicated().sum()
```

```
3008
```

```
df.drop_duplicates(inplace = True)
```

- 2 Identify and remove missing data

```
#Check if there is any NaN or Null value  
print(df.isnull().sum())
```

```
satisfaction_level      0  
last_evaluation         0  
number_project          0  
average_montly_hours   0  
time_spend_company     0  
Work_accident          0  
left                    0  
promotion_last_5years  0  
Department             0  
salary                  0  
target                 0  
dtype: int64
```

04

LAN DUNG DO



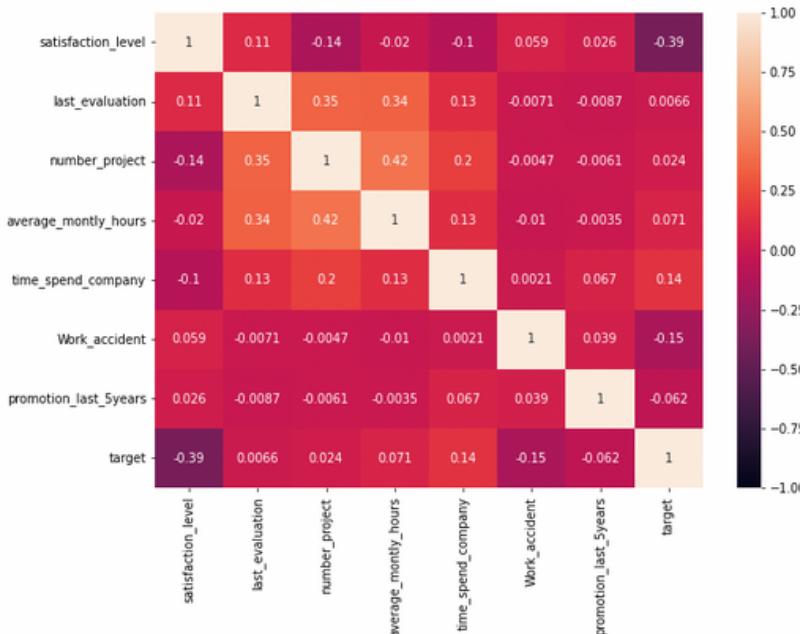
# IV. INSIGHTS

1

## Correlation heatmap

```
In [227]: plt.figure(figsize = (10,7), dpi = 70)
sns.heatmap(dfl.corr(), annot=True, vmin = -1, vmax = 1)

Out[227]: <AxesSubplot:>
```



-->. A possible conclusion that could be drawn from a heatmap correlation is that 3 strongest relationship or association with 'target' is satisfaction\_level, time\_spend\_company, work\_accident.

05

LAN DUNG DO

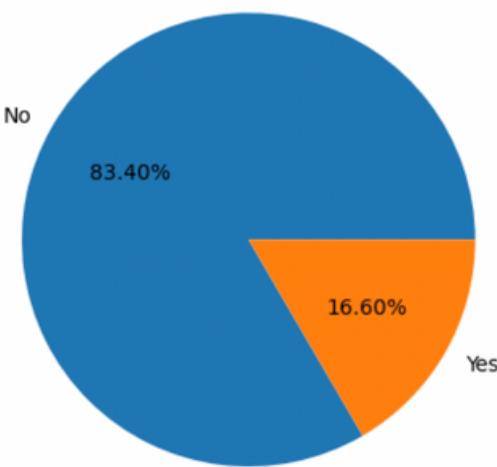
## 2

## Pie Chart

```
#Pie chart - checking whether the outcome data is balanced
print(df.left.value_counts())
df.left.value_counts().plot.pie(labels=("No", "Yes"), autopct=".2f%%", label="")
```

0 10000  
1 1991  
Name: left, dtype: int64

: <AxesSubplot:>

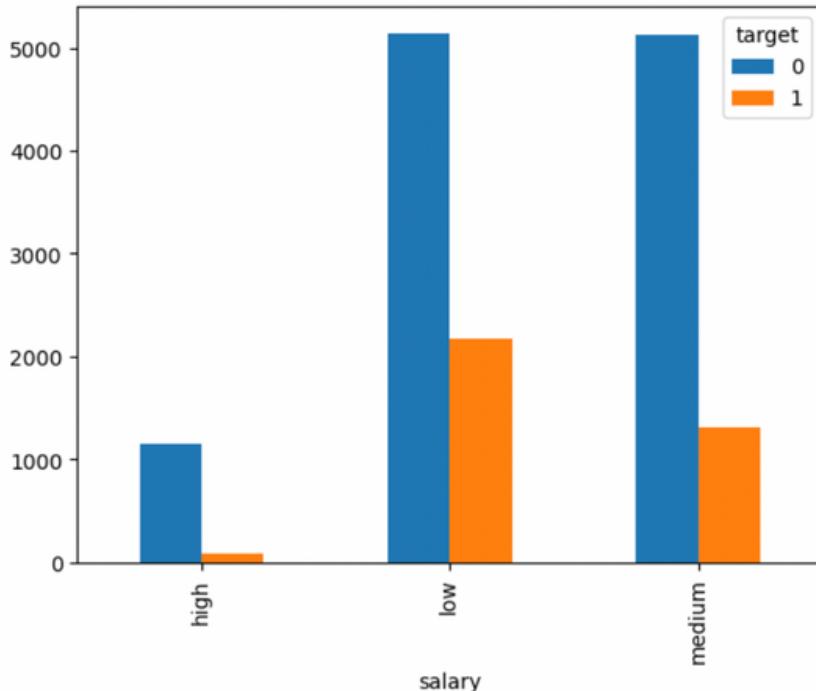


-->.The pie chart shows that there is unbalanced classes in which out of 11991 employees, 83.40% employees leave the company and 16.60% employees stay at the company.

## 3

## Bar Chart

```
: # Bar chart  
pd.crosstab(df1.salary, df1.target).plot(kind = 'bar')  
: <AxesSubplot:xlabel='salary'>
```



--> Based on the bar chart, it appears that employees who earn higher salaries are more likely to stay with the company compared to those with lower salaries. Additionally, the data suggests that the likelihood of staying with the company is similar for the medium and low salary groups. However, within these two groups, the data indicates that employees with lower salaries are more likely to leave compared to those with medium salaries.

## 4

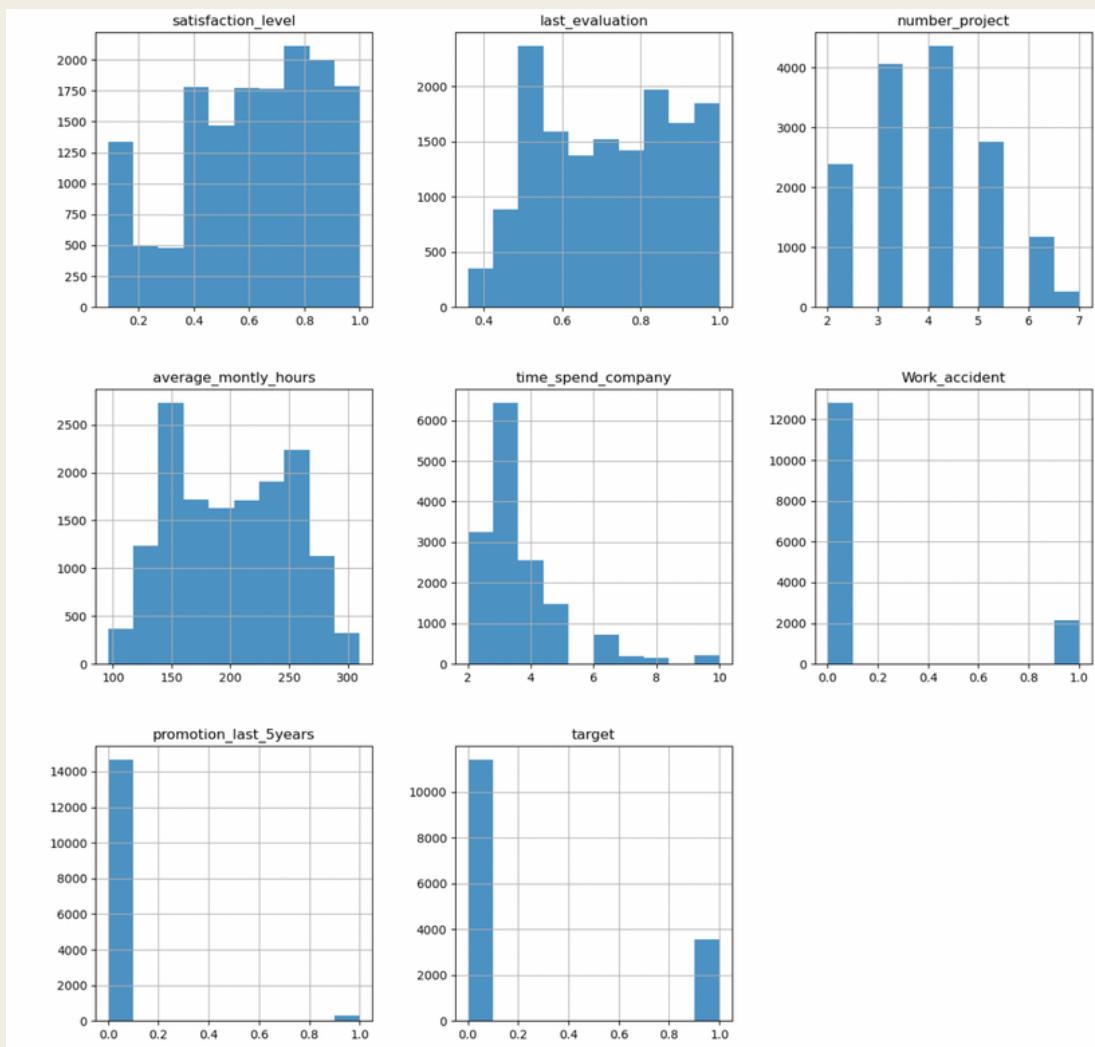
## Caculate the mean based on target

df1.groupby('target').mean()							
	satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	promotion_last_5years
target							
0	0.666810	0.715473	3.786664	199.060203	3.380032	0.175009	0.026251
1	0.440098	0.718113	3.855503	207.419210	3.876505	0.047326	0.005321

- > On average, employees who stayed had higher satisfaction\_level than that of employees who left the company.
- > Average working hours and number of projects of employees leaving company are higher than those who decided to stay.
- > Employees who were promoted in the last 5 years were more likely to stay

## 5

## Histogram



- According to the graph, the satisfaction level between 0.7 and 0.8 is the most frequent, with around 2100 occurrences. The range of 0.45 to 0.55 is the most common for the last evaluation score, with approximately 2500 occurrences. The majority of employees work around 150 hours per month, as shown by the peak in the graph at that point.

# V. MULTIPLE REGRESSION

1

```
# Create a linear regression object and fit the model to the data
check = [0,1,2,3,5,6]
X = df1.iloc[1:1000, check].values
y = df1.iloc[1:1000, 4].values
model = LinearRegression().fit(X, y)

# Print the coefficients and intercept
print("Coefficients:", model.coef_)
print("Intercept:", model.intercept_)

Index(['satisfaction_level', 'last_evaluation', 'number_project',
       'average_montly_hours', 'time_spend_company', 'Work_accident',
       'promotion_last_5years', 'Department', 'salary', 'target'],
      dtype='object')
Coefficients: [ 1.74753851  1.56095943  0.14398624  0.00322129  0.07916078 -0.2456209 ]
Intercept: 0.7543232993437279
```

Equation of multiple regression on time spend with company (with all numerical variables)

$$Y = 0.754 + 1.747x_1 + 1.56x_2 + 0.14x_3 + 0.0032x_4 + 0.079x_5 - 0.245x_6$$

Intercept of equation:  $B_0 = 0.754$

--> Most 3 important features that affect time spend with company in order is satisfaction level , last evaluation, work accident

10

LAN DUNG DO

2

```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score, mean_squared_error

# assume X_train, y_train, X_test, and y_test are already defined
# 80% of instances are used to train model and the remaining 20% is hold-out set.
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42, stratify = y)
# create a Linear Regression model
model = LinearRegression()

# train the model using the training data
model.fit(X_train, y_train)

# make predictions using the test data
y_pred = model.predict(X_test)

# calculate R-squared value
r2 = r2_score(y_test, y_pred)

# calculate mean squared error
mse = mean_squared_error(y_test, y_pred)

print("R-squared value:", r2)
print("Mean squared error:", mse)
```

R-squared value: 0.7296332161310328  
Mean squared error: 0.25898434226808364

-->R-square value of model: 0.7296

-->Mean square error of model: 0.258

11

LAN DUNG DO

3

```
#Classification report
Treereport = classification_report(y_test, Treepredict)
print(Treereport)
```

	precision	recall	f1-score	support
2	0.00	0.00	0.00	7
3	0.95	0.95	0.95	176
4	0.82	0.86	0.84	99
5	0.64	0.67	0.65	96
6	0.13	0.09	0.11	22
accuracy			0.80	400
macro avg	0.51	0.51	0.51	400
weighted avg	0.78	0.80	0.79	400

-->The f1 accuracy score of TreeDecision Regression is 0.80, indicating that the model is performing well in terms of both precision and recall.

12

LAN DUNG DO

# BUSINESS INSIGHTS

1

dummy_variables.sum()	
Department_IT	976
Department_RandD	694
Department_accounting	621
Department_hr	601
Department_management	436
Department_marketing	673
Department_product_mng	686
Department_sales	3239
Department_support	1821
Department_technical	2244

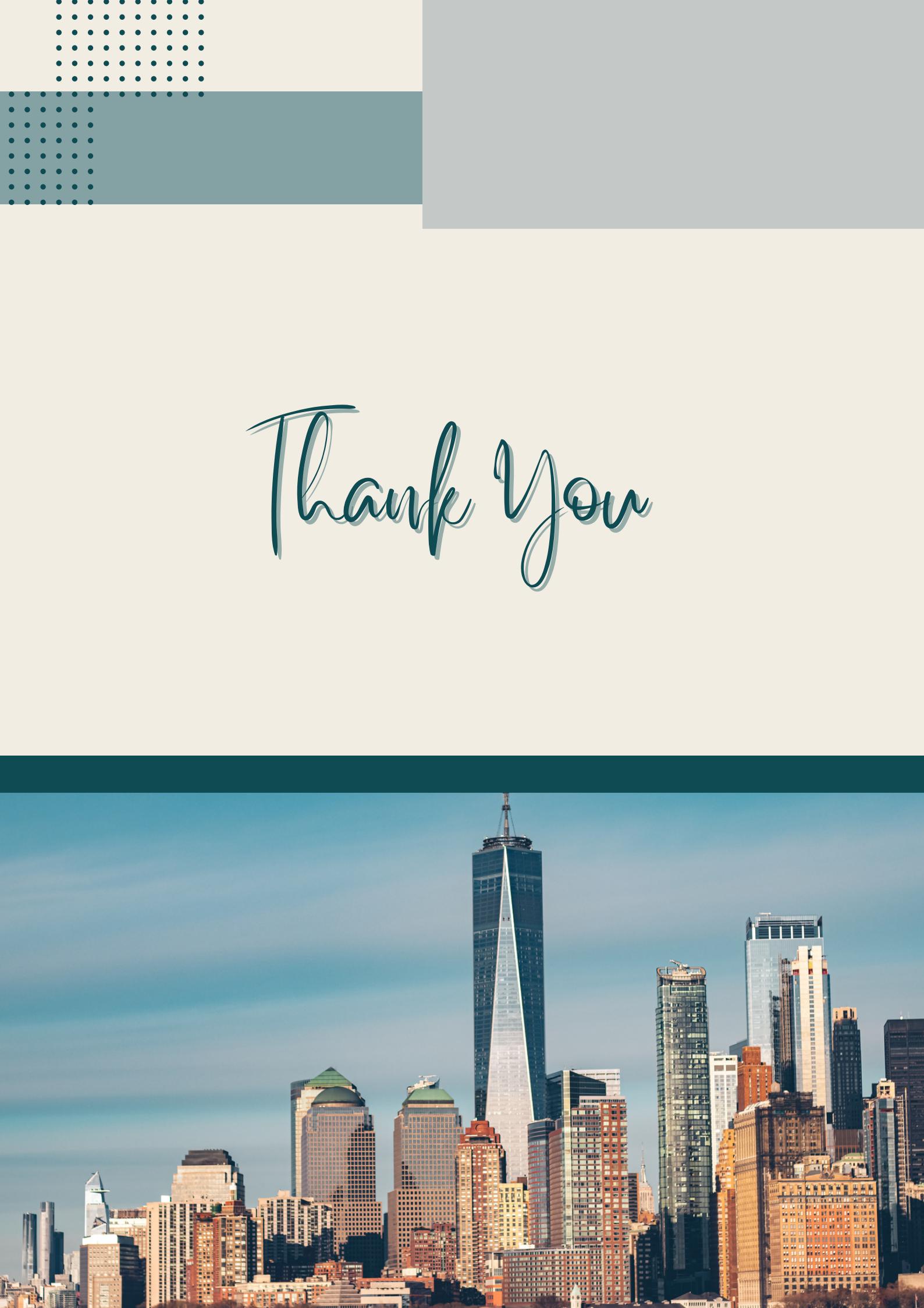
-->From the dummy variables stats, we observe 3 most department that have highest turnover rate are sale , technical and support

13

LAN DUNG DO

## What company can do to improve retention?

1. Improve employee engagement: Engaged employees are more likely to stay with a company. Companies can improve engagement by providing clear promotion opportunities, recognition and rewards programs, and opportunities for employees to share feedback and ideas.
2. Offer competitive compensation and benefits: Compensation and benefits are important drivers of employee retention. Companies should ensure that their compensation packages are competitive with industry standards and offer attractive benefits packages that meet employee needs.
3. Provide opportunities for learning and development: Employees are more likely to stay with a company that invests in their professional development. Companies can offer training programs, mentorship opportunities, and other resources to help employees grow and develop their skills.
4. Foster a positive company culture: A positive company culture can help employees feel valued and connected to the company. Companies can create a positive culture by promoting open communication, transparency, and collaboration.
5. Provide flexible work arrangements: Flexible work arrangements, such as remote work or flexible schedules, can help employees achieve work-life balance and improve job satisfaction.
6. Conduct stay interviews: Stay interviews can help companies identify the reasons why employees stay with the company and what could potentially drive them away. Companies can use this information to make targeted improvements and retain their top talent.



Thank You

