Brain Drain Analysis

Determining talent retainment risks for your organization through data-powered prediction

Rafael Deyta
José Sierra
Luis Vázquez
Manuel Borrego
María Inés Oñate

Index

What is Brain Drain?

Problems to Solve

Project Objective

Understanding Variables

Trend Analysis

Target Audiences

Solution

Process

Deliverable

Vision

What is Brain Drain?

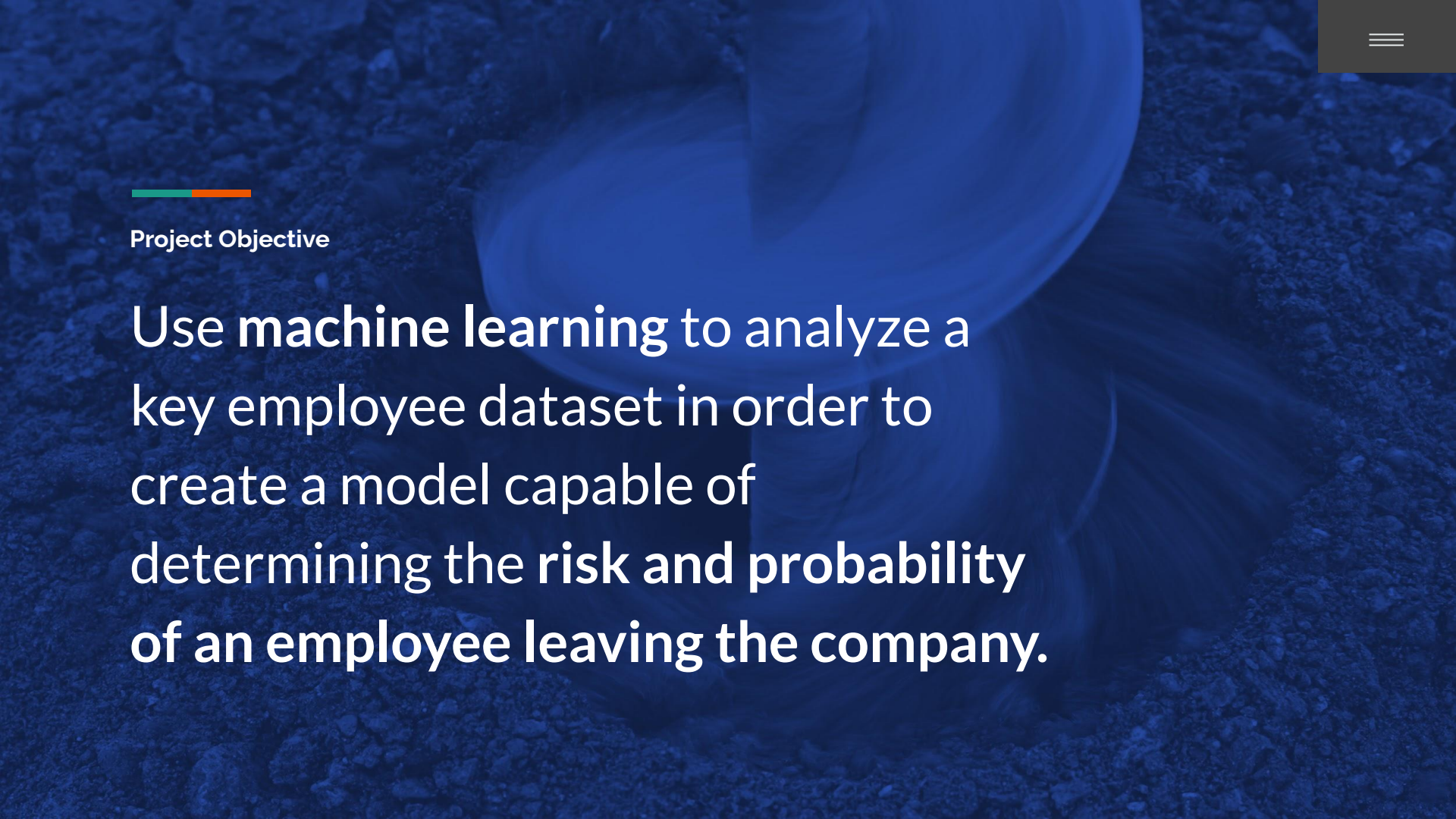
Brain drain is a slang term indicating substantial withdraw or movement of individuals. A brain drain can result from turmoil within a nation, the existence of favorable professional opportunities in other organizations or even countries, or from a desire to seek a higher standard of living. Brain drain may occur at the organizational or industrial levels when workers perceive better pay, benefits, or upward mobility within another company or industry.

Problems to Solve

1. Avoid key employees leaving the organization after a successful career and significant company investment.

2. Determine key variables inside the organization and whether there is relationship with an employee's decision of leaving a company.

3. Establish a path to success for key employees and track the evolution regularly in order to prevent premature resignation.



Project Objective

Use **machine learning** to analyze a key employee dataset in order to create a model capable of determining the **risk and probability of an employee leaving the company.**

OUR VARIABLES

1 Dataset Variables
Employee Type

There are different employee types in the organization:

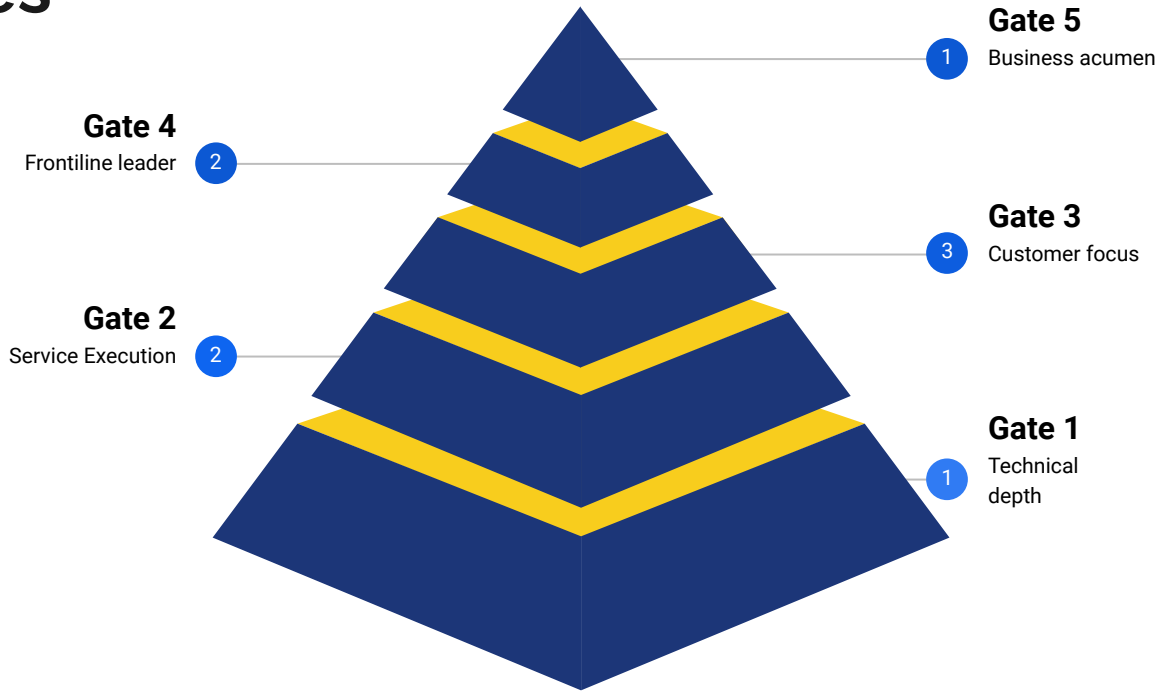
1. Regular.
2. Commuter.
3. Expat.
4. Expat-ShortTerm/Dev.

Implications for client
The employee type has different salary scales and different priorities for the organization.

2 Dataset Variables
Gate

Gate is a variable to determine the path to become a manager. Each employee has a gate value in relationship with degree and technical knowledge.

You enter the process based on your current job title and previous experience. You will only need to have assessments completed from your current Phase forward.



3 Dataset Variables
Gender

There are several employee types in the organization:

1. Male.
2. Female.

Implications for client
Oil industry has a historically higher male occupation.

4 Dataset Variables
Country

Oil company operates in a large number of countries:

- India.
- Malaysia.
- Brunei.
- Singapore.
- Japan.
- Myanmar.
- Thailand.
- Vietnam.
- South Korea.
- Bangladesh.

5 Dataset Variables
Scale

Hierarchy and importance inside the organization

1. Value from 4 to 31.

Observations:
31 is the highest value.

6 Dataset Variables
Comparatio

1. Value > 1 salary is higher than industry.
2. Value < 1 salary is lower than industry

Observations:

Compensation metric that compares the salary an employee is paid to the midpoint of the salary range for their position or similar positions at other companies.

& Dataset Variables
Other Variables

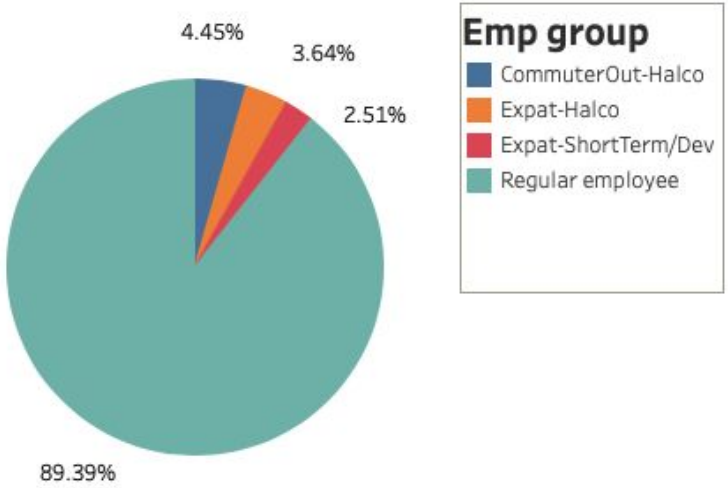
Hierarchy and importance inside the organization

1. Promotions.
2. International Assignment.
3. Age.
4. Seniority.
5. Dependents.
6. Job Code.

Employee Group

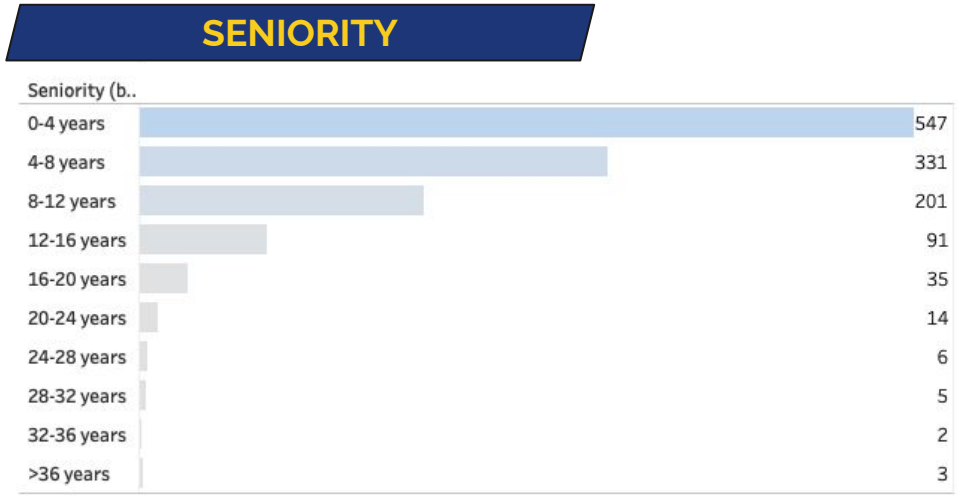
1.Commuter Out
2.Expat
3.Expat Short term
4.Regular Employee

TYPE OF EMPLOYMENT



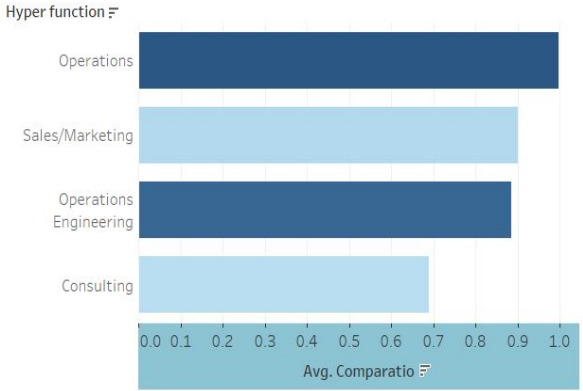
Seniority

0-4 years: 547
4-8 years: 331
8-12 years: 201
12-16 years: 91

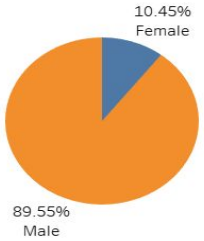


Other Variables

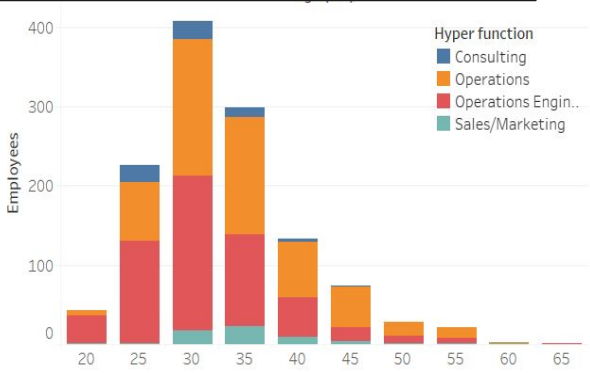
COMPARATIO



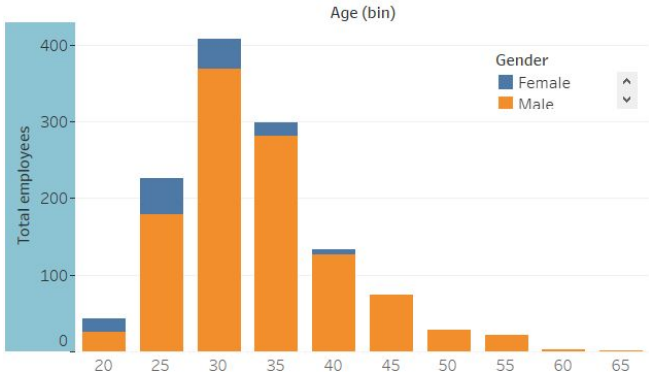
GENDER



COMPARATIO PER AGE

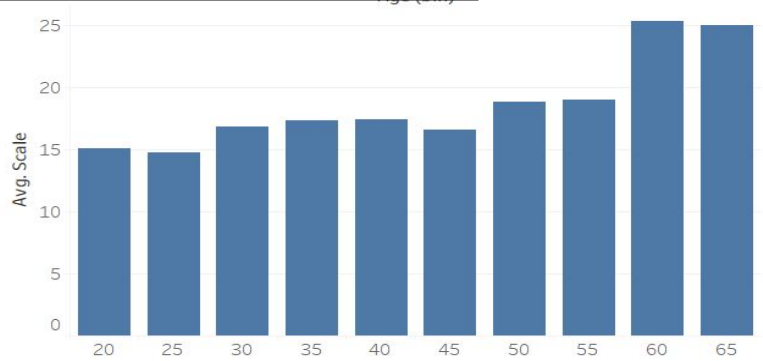


GENDER, AGE, CLASS

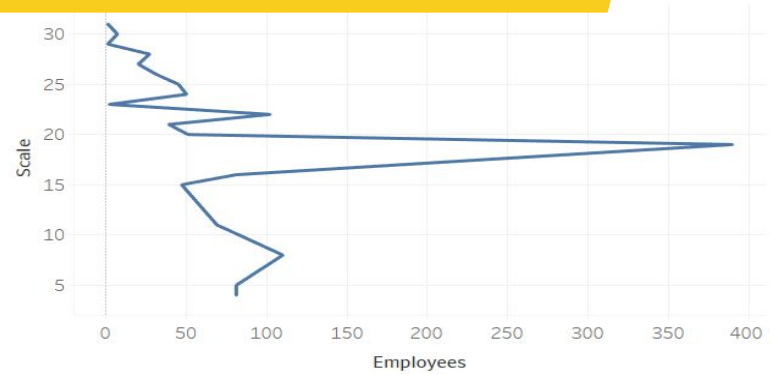


Other Variables

RATIO PER AGE



RATIO PER EMPLOYEE



Machine Learning

MODEL SELECTION

Random Forest

After running a Logistic Regression in first place we find accurate results using random forest model.

Numerical Data

Transform non numerical in numerical columns

Visualize correlation

Create a heatmap to identify the correlations between variables

Training and Testing the Model

X_train, X_test, y_train, y_test = train_test_split

EVALUATE

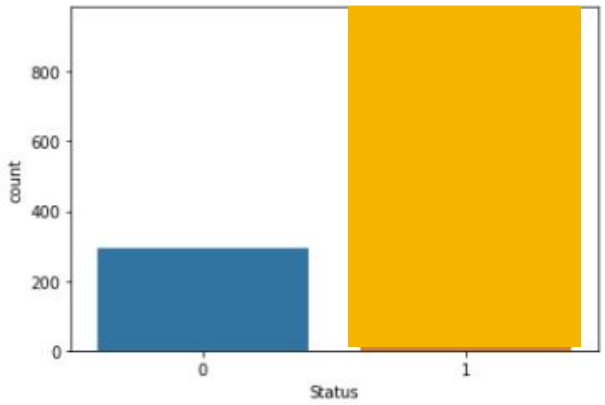
Evaluate Accuracy, Recall and confusion Matrix.

Imbalance data > SMOTE

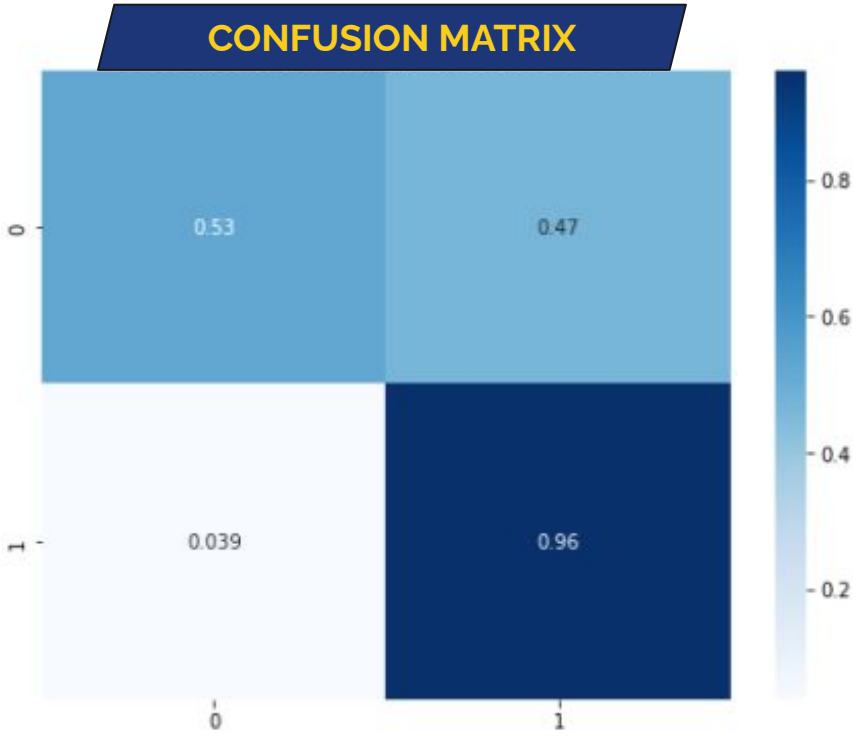
After analyze the output it was necessary to balance the data since the Active samples where considerable higher than the withdraw.

CLASSIFY, FIT, PREDICT

First Scenario (No Smote)




```
1    939
0    295
Name: Status, dtype: int64
```

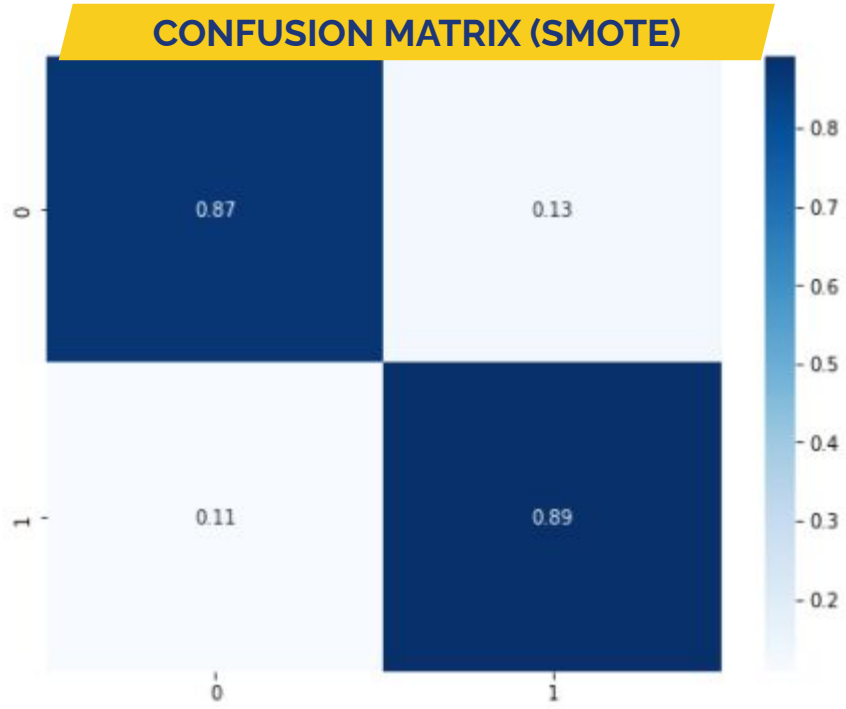


Second Scenario

```
In [25]: # summarize class distribution
         counter = Counter(y)
         print(counter)

         Counter({1: 939, 0: 939})
```

```
Accuracy = 0.88
Recall = 0.89
```



Conclusions

- Regular employees are prone to leave the company between 29 -35 years old
- Between 7-10 years of seniority the employees look other options
- Women are leaving the company after 35 year old
- Employees with more than 2 dependents prefer to stand at the company
- International assignations help to retain talent
- The company needs to focus on Gate 1 and 2, 85% of the separations are in these gates.

Thanks.