IBM EMPLOYEE SURVEY

An analysis of the factors responsible for Employee Attrition at IBM

GROUP 3

COVENLABS X GIZ DATA SCIENCE TRAINING
COHORT II

GROUP MEMBERS

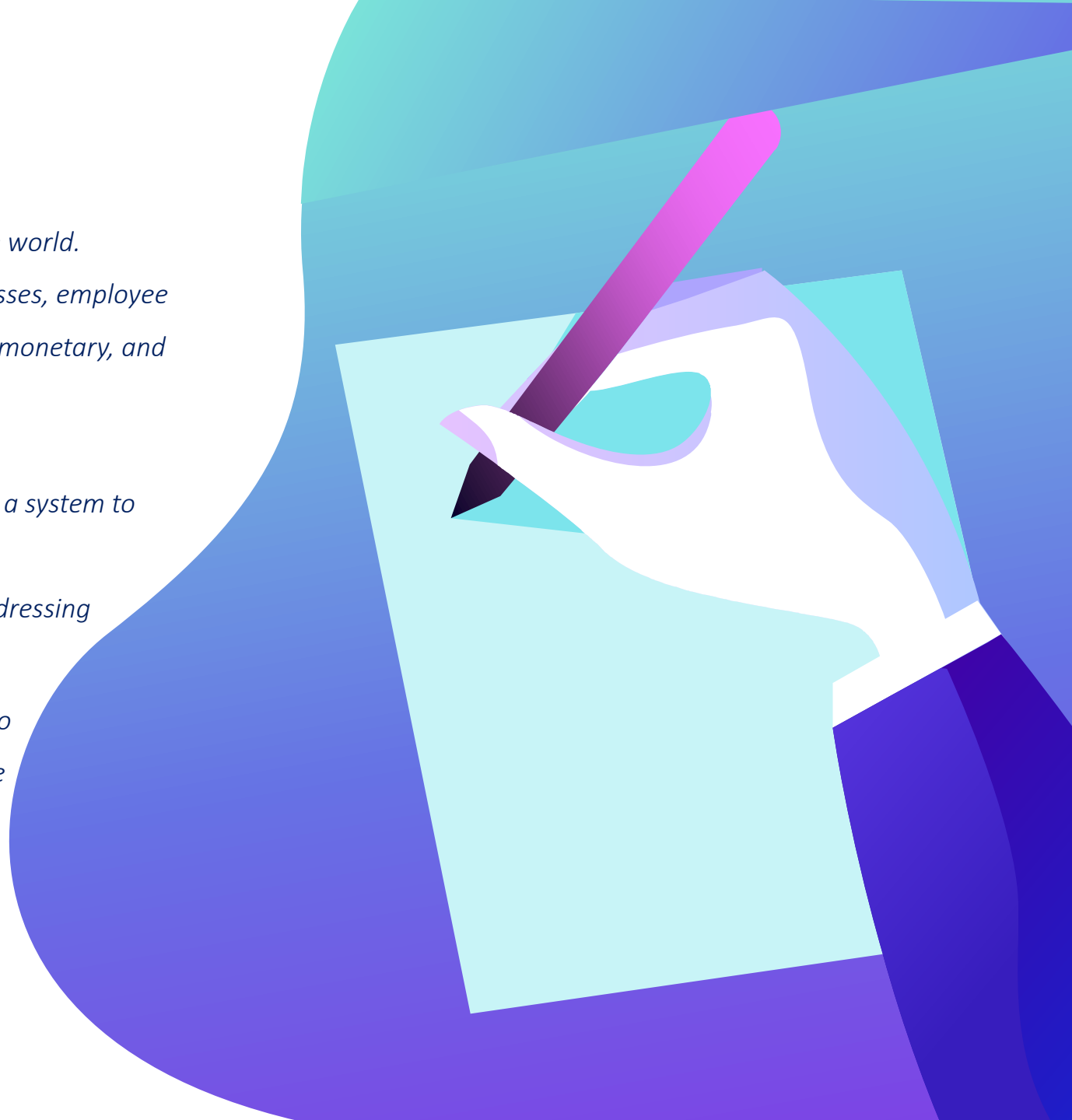
- *Ajayi Oluwatomilayo Abiodun*
- *Okondu Joseph Ifeanyi*
- *Victoria Abia Ediete*
- *Adolor Lewis*
- *Timeyin Blankson*
- *Muhammad Okunade*
- *Dele Oyapitan*
- *Ukara Evelyn Onyekachi*
- *Nsunhusi Edwin*
- *Bassey emmanuel*
- *Anthony Ewoma*

PROBLEM STATEMENT

Employee attrition is a great challenge to organizations around the world. Businesses lose talents regularly through attrition. For most businesses, employee attrition causes momentary slow-paced growth as it takes human, monetary, and time resources to fill a role.

Businesses could avoid the negative impact of attrition if they have a system to anticipate employees actions as regards resignation. Beyond that, Organizations can reduce employee turnover by identifying and addressing Factors that cause employee resignation.

This project seeks to identify and analyze the factors contributing to employee attrition and also develop model to predict the employee attrition among the employees at IBM.



BACKGROUND TO STUDY

To study the attrition rate among her employees, IBM has gathered information on employee attributes, job satisfaction, income, seniority, etc.it include data of 1470 employees.

From this dataset we can uncover the factors that lead to employee attrition and explore important questions such as how does the distance from home to work affect employee attrition or how does the monthly income relate to employee attrition

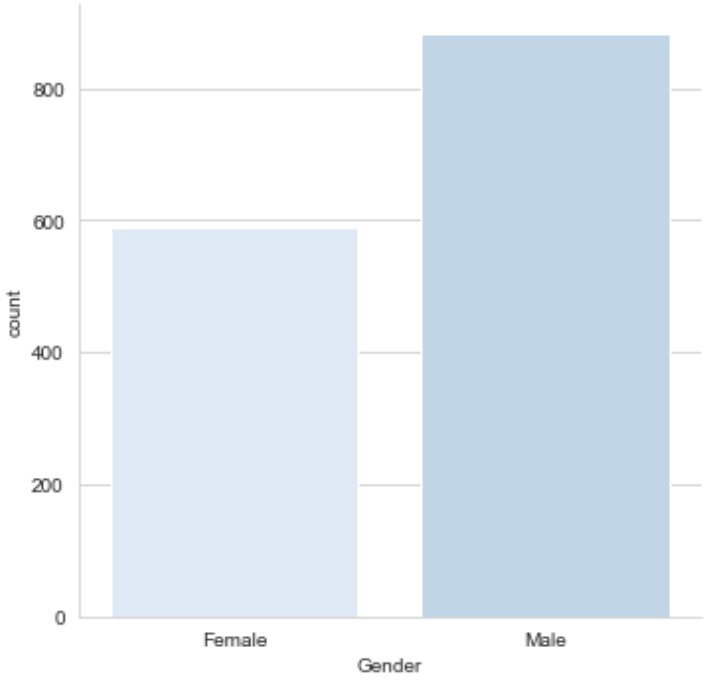
DEFINITION OF TERMS

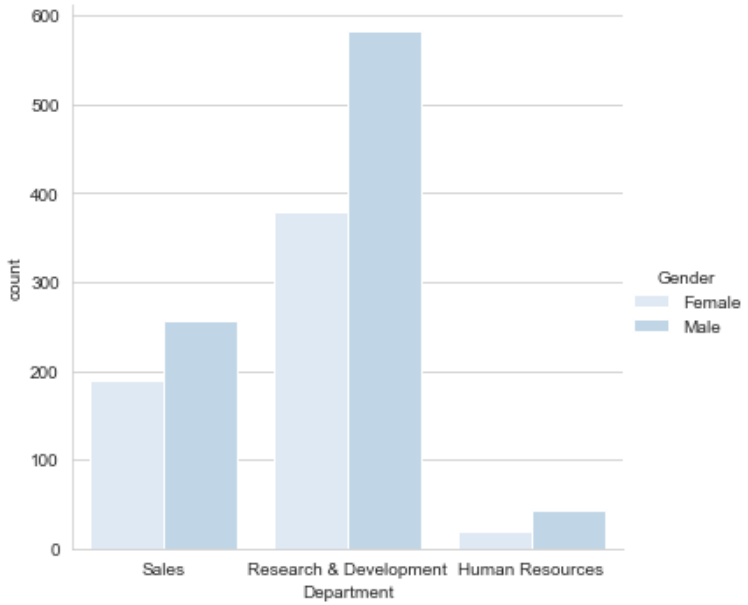
Attrition: *Attrition refers to the loss of employees through a natural process such as retirement, resignation, elimination of a position, personal health, or other similar reasons. Companies sometimes lose some of their best staff due to this or sometimes wrongly invest in training of some employees and some staff who have been groomed for leadership positions over the years could abruptly leave their jobs. Employees who are likely to stay long at the company could be laid off while those who might leave eventually could be retained instead.*

AGENDA

- *To conduct a demographic study of the employees at IBM and visualize the results*
- *To analyze and visualize the possible factors leading to employee attrition such as career stagnation, job satisfaction, work-life balance, etc.*
- *To analyze other phenomenon such as promotional pattern across departments, Gender equality, etc.*
- *Train a Machine Learning model that will be able to help HR predict if an employee is likely to leave the company so appropriate measures can be taken.*

Demographics

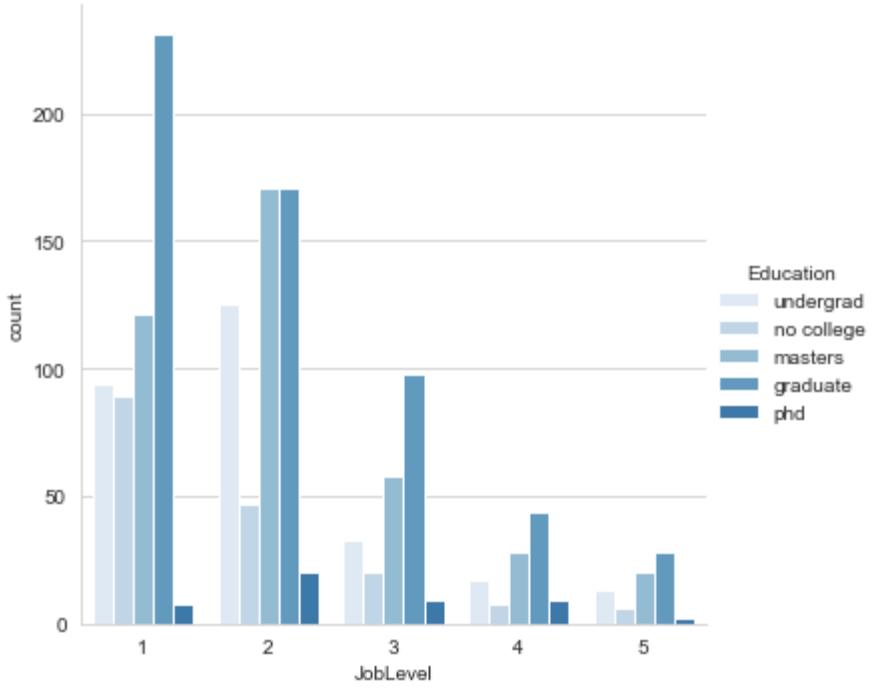


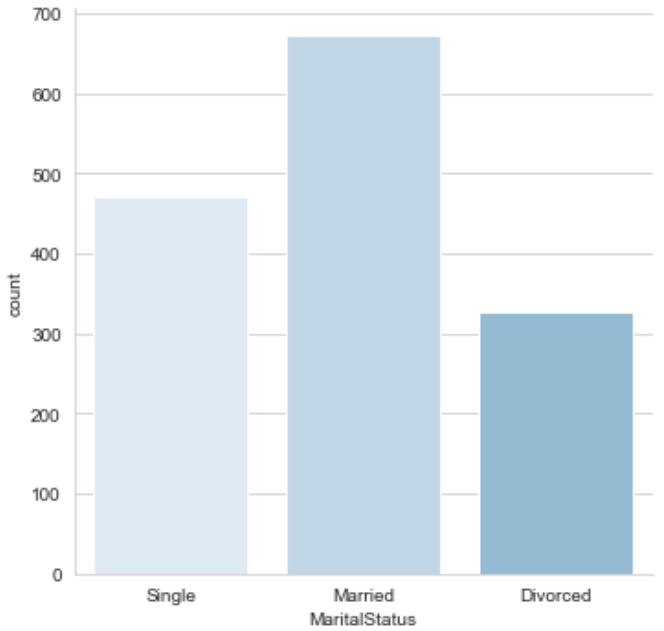


The dataset consists of 1470 employees. There are 588 female staff and about 882 male staff

The employees are spread across the 3 departments in the company which are: Sales, Research & Development and Human Resources. The Research & Development department recruits the most employees and this highlights the company's preference for in-house Research and Development

Demographics



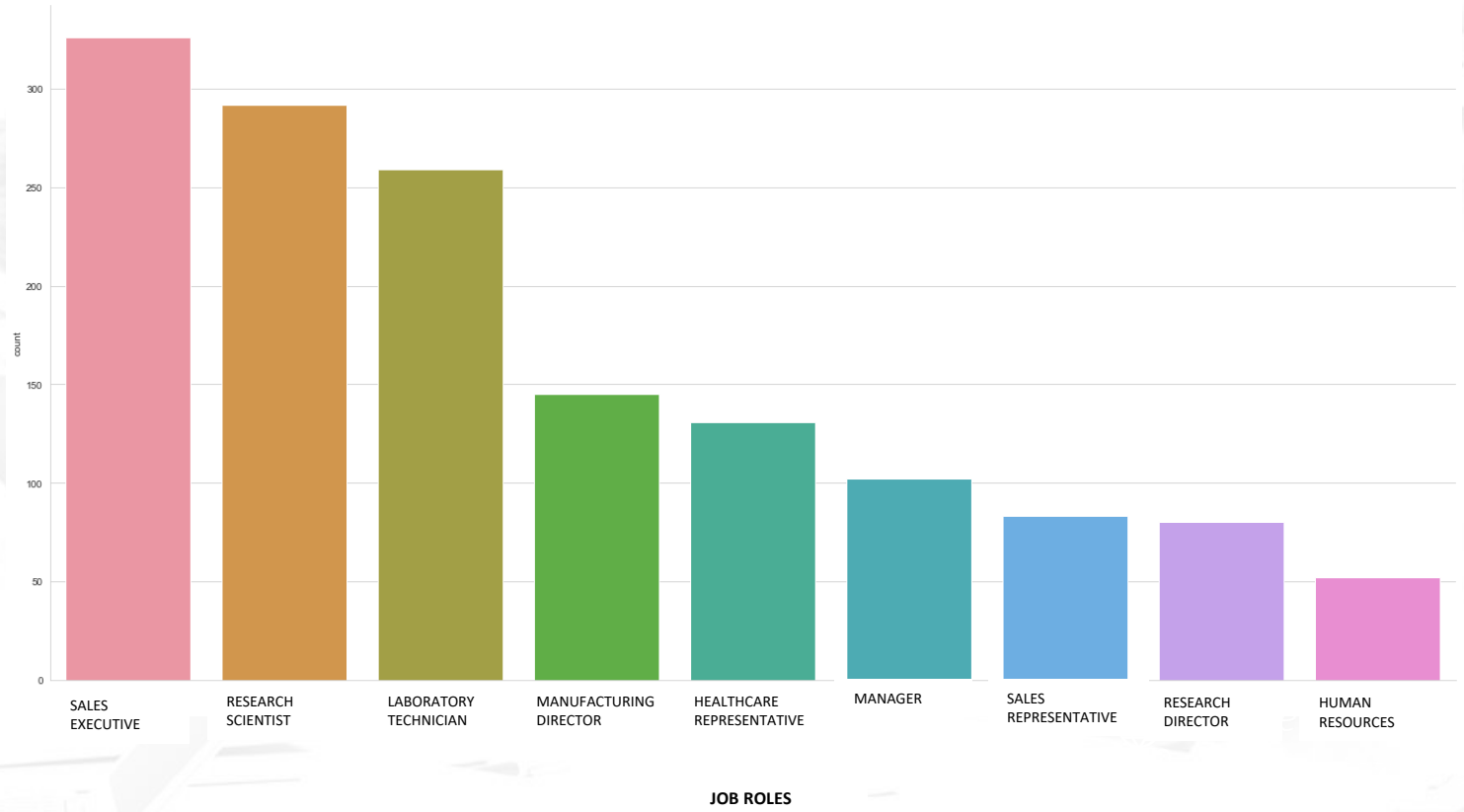


The qualifications of the employees ranges from Bachelors, Masters and Ph.D degree. People with no college degrees and college students are also employed

The job level ranges from 1 to 5 with 1 being the most junior and 5 being the most senior.

Most employees are married but there are also Single and Divorced people

Demographics



The job roles include: sales executive, research scientist, laboratory technician, Manufacturing Director, Healthcare Representatives, Managers, Sales representatives and Human resources. Most Employees are hired for Sales related roles and research roles

Feature Description

Education

1 'Below College

' 2 'College

' 3 'Bachelor

' 4 'Master

' 5 'Doctor

Environment Satisfaction

1 'Low'

2 'Medium'

3 'High'

4 'Very High'

Job Involvement

1 'Low'

2 'Medium'

3 'High'

4 'Very High'

Job Satisfaction

1 'Low'

2 'Medium'

3 'High'

4 'Very High'

Performance Rating

1 'Low'

2 'Good'

3 'Excellent'

4 'Outstanding'

Relationship Satisfaction

1 'Low'

2 'Medium'

3 'High'

4 'Very High'

WorkLife Balance

1 'Bad'

2 'Good'

3 'Better'

4 'Best'

Years at the company have been classified under:

'Less than 10 years'

'11-20 years'

'21-30 years'

'Above 30 years'

Salary Level:

Employees who earn less than 5000 are categorized as Junior Salary,5000-10000 as Middle Level,10000-15000 as Senior level and above 15000 as Manager Level

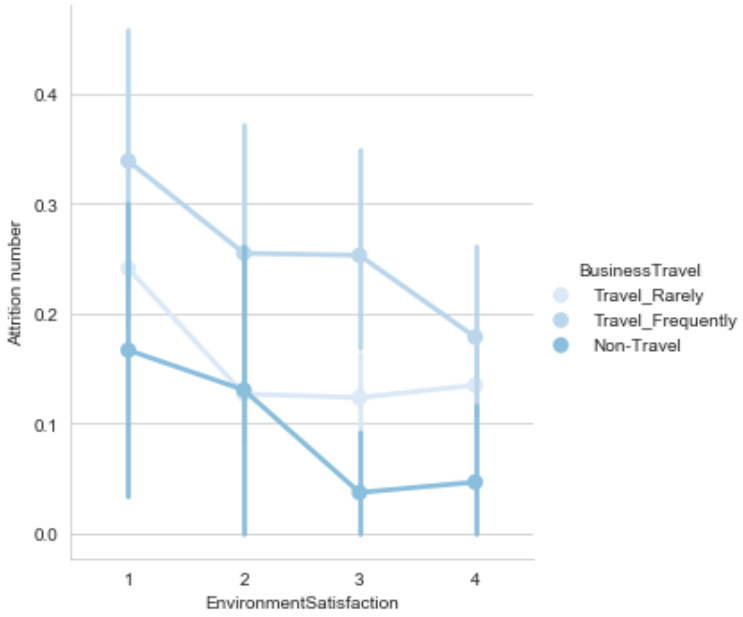
Distance Travelled

Less than 10 km is Short Distance, 11-19 is moderately far while 20 and above is far

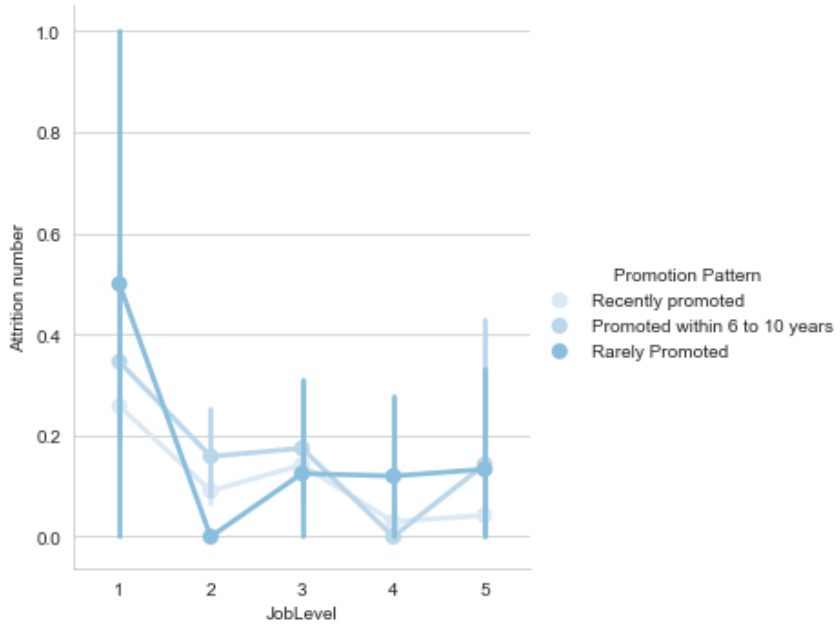
Years spent at the company

Little experience is less than 3 years middle experience is 3-6 years while highly experienced id above 6 years

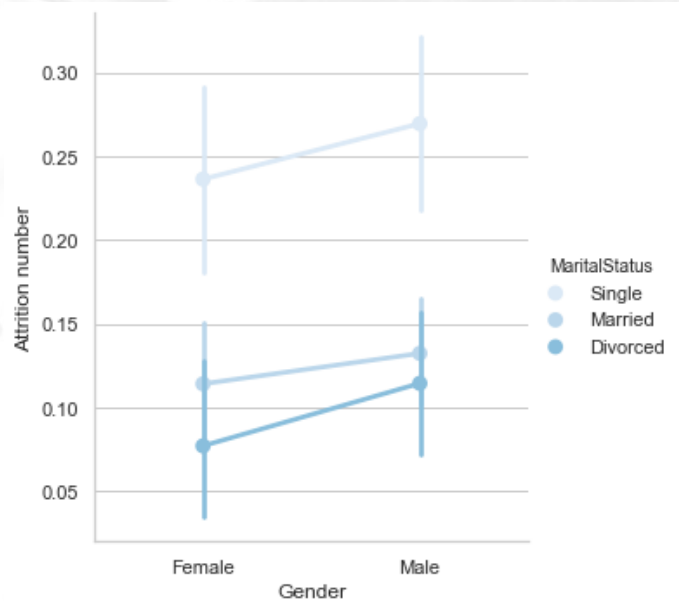
Factors responsible for Employee Attrition



From the chart, employees who travel frequently and are least satisfied with the environment at work are more likely to leave their jobs. Those who do not travel at all and are satisfied with the work environment are less likely to leave their jobs

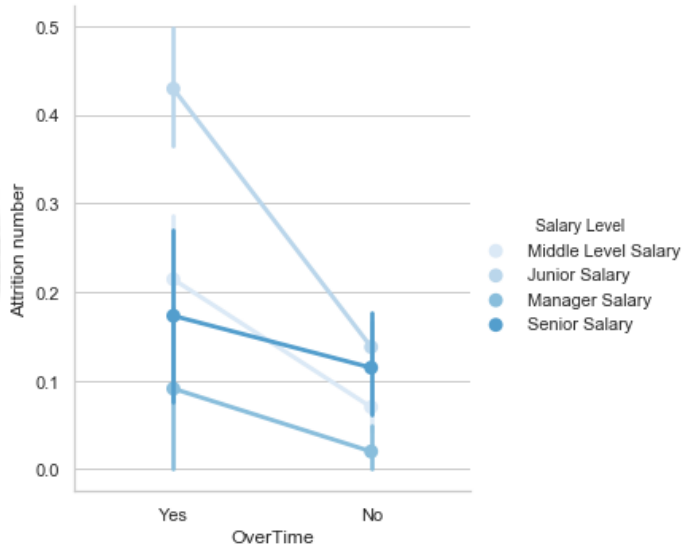


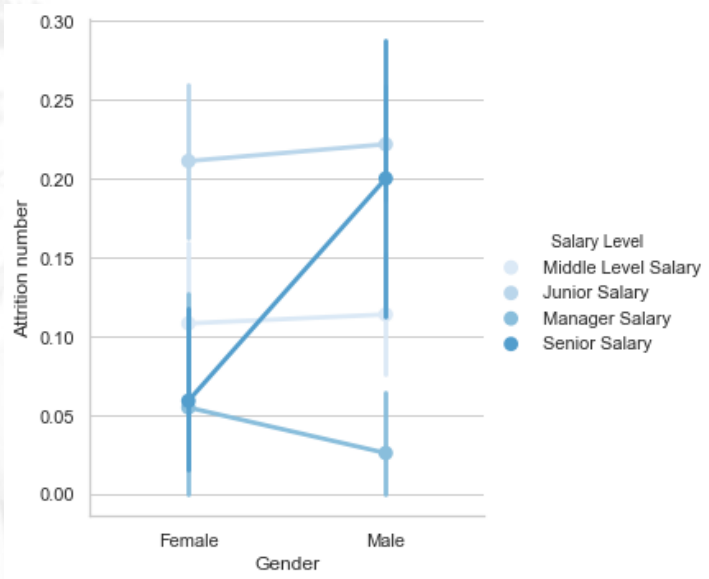
Employees who rarely get promoted and are at the bottom of the job ladder will most likely eventually leave their jobs compared to manager level staff who just got promoted recently



Single men have the least loyalty to the company. Divorced women are less likely to leave their jobs maybe because they are single parents and need stable income. Married people are also less likely to leave their jobs compared to single people who have some level of financial freedom

Factors responsible for Employee Attrition

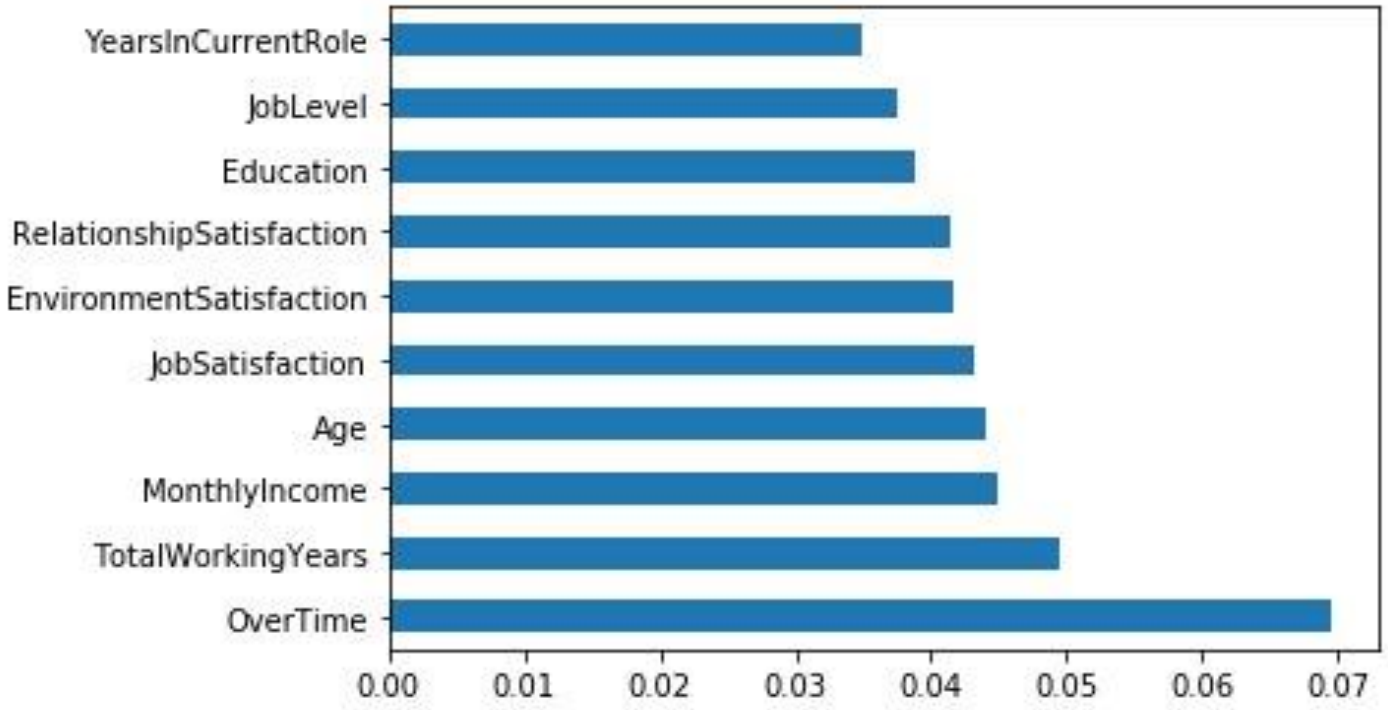




Employees who are paid the lowest and still have to work overtime will be more likely to leave their jobs. The people least likely to quit are those who are paid the highest and do not have to work overtime

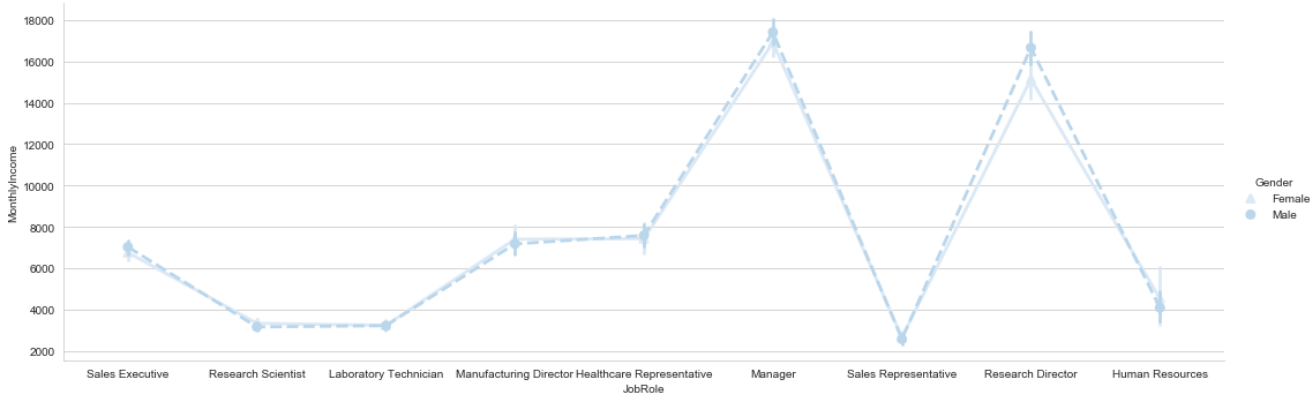
Females are less likely to quit their jobs than their male counterparts on the same salary level. Although men who have achieved the highest salary level are the least likely to quit

Feature Importance

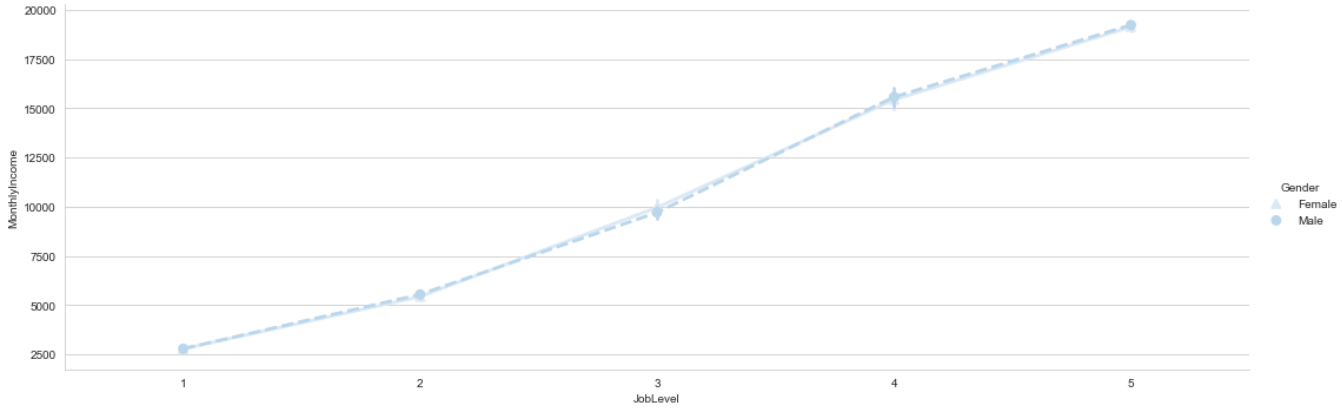


The major reason for employee attrition was determined to be the constant amount of overtime by some employees. After spending a large amount of time at the company, employees also tend to leave either to start their own companies or due to the retirement age.
The monthly income is also a major factor in employee attrition. Some employees feel they are not properly compensated for the work they do at the company compared to some o their fellow employees

Salary Distribution



Managers and research directors earn the most while the sales representatives who are the most in the company earn the least. Research scientists and laboratory technicians are also not paid much



The employees on level 5 earn close to 10 times as much as their counterparts on level 1

Predictions

Using Logistic Regression

```
import warnings
warnings.filterwarnings('ignore')
from sklearn.linear_model import LogisticRegression as logreg
logreg=LogisticRegression()
logreg.fit(X_train,y_train)
```

```
LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
                   intercept_scaling=1, l1_ratio=None, max_iter=100,
                   multi_class='warn', n_jobs=None, penalty='l2',
                   random_state=None, solver='warn', tol=0.0001, verbose=0,
                   warm_start=False)
```

```
confusionmatrix=confusion_matrix(y_test, ypred)
confusionmatrix
```

```
array([[406,   5],
       [ 51,  28]], dtype=int64)
```

The model has been able to predict if an employee is going to leave his or her job. '1' means the employee will leave and '0' means the employee will stay.
The accuracy of the model was tested and returned a score of 89%

```
array([0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
       1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0])
```

Predictions

Using Decision Tree Classifier

```
]: from sklearn.tree import DecisionTreeClassifier
   dtcf=DecisionTreeClassifier(criterion='entropy',random_state=0)
   dtcf.fit(X_train,y_train)

]: DecisionTreeClassifier(class_weight=None, criterion='entropy', max_depth=None,
                          max_features=None, max_leaf_nodes=None,
                          min_impurity_decrease=0.0, min_impurity_split=None,
                          min_samples_leaf=1, min_samples_split=2,
                          min_weight_fraction_leaf=0.0, presort=False,
                          random_state=0, splitter='best')
```

The model has been able to predict if an employee is going to leave his or her job. '1' means the employee will leave and '0' means the employee will stay.
The accuracy of the model was tested and returned a score of 80%

```
In [41]: ypred=dtcf.predict(X_test)
         ypred

Out[41]: array([1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0,
                0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0,
                0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0,
                0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
                0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 1, 1, 0,
                1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1,
                1, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1,
                0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
                0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1,
                1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1,
                0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
                0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0,
                0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
                0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0,
                0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
                0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
                0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
                0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0,
                0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
                0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0,
                0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
                0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0,
                0, 0, 0, 0, 0, 0], dtype=int64)

In [43]: from sklearn import metrics

In [44]: print("Accuracy:",metrics.accuracy_score(y_test, ypred))

         Accuracy: 0.7959183673469388
```

Conclusion and Recommendations

- *The factor most likely to lead to employee attrition is overtime. HR could control this by reducing the amount of extra hours employees have to work outside working hours and adequately compensating the workers who will still have to work overtime*
- *Older employees can be given incentives to make them retire quicker so younger employees can be promoted to senior positions*
- *The disparity in the income of the lowest employees is too much from that of the most senior employee*

Thank You

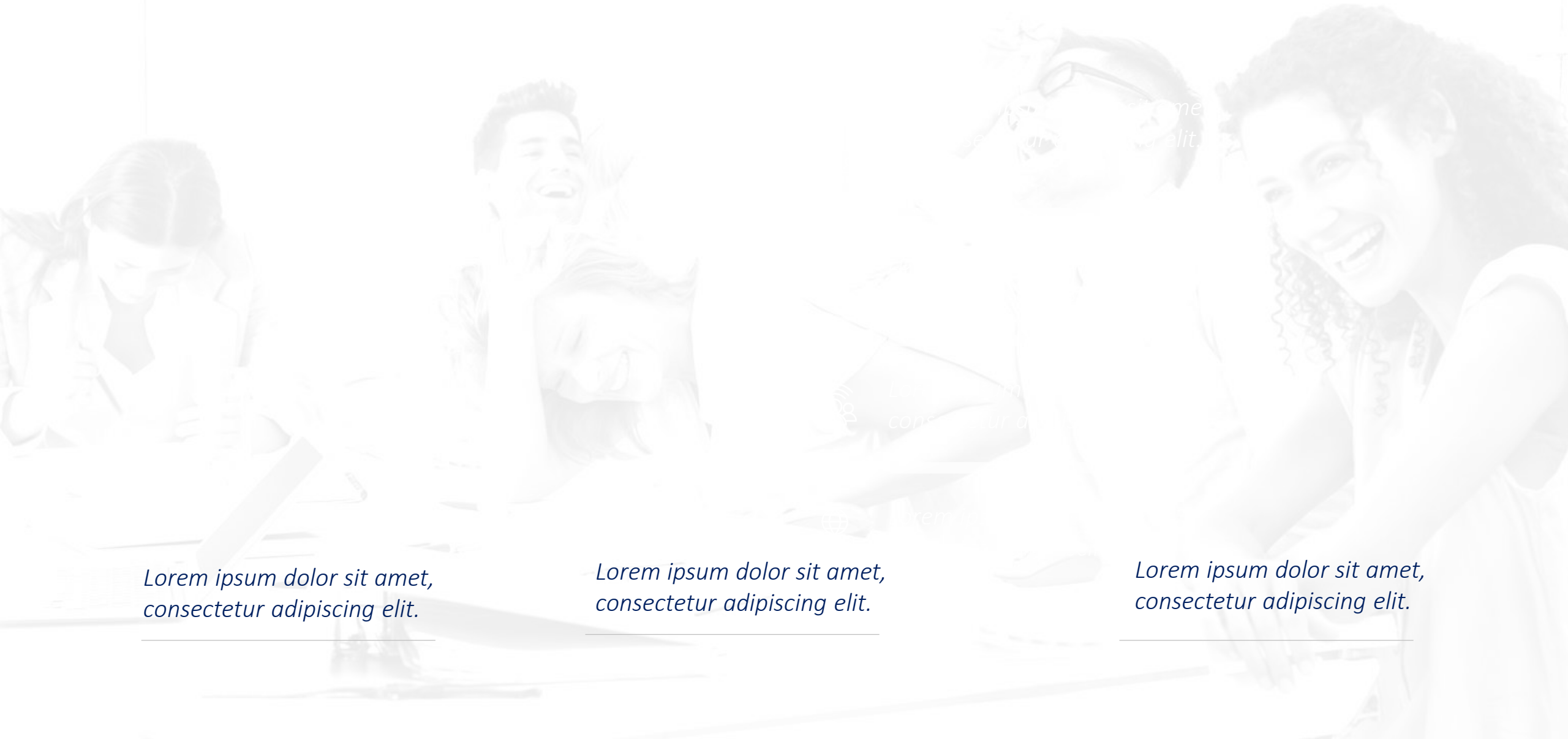
PROBLEM STATEMENT

Attrition refers to the loss of employees through a natural process such as resignation, elimination of a position, personal health, or other similar reasons. Companies sometimes lose some of their best staff due to this or sometimes wrongly invest in training of some employees and some staff who have been groomed for leadership positions over the years could abruptly leave their jobs. Employees who are likely to stay long at the company could be laid off while those who might leave eventually could be retained instead.

To study the attrition rate among her employees, IBM has gathered information on employee attributes, job satisfaction, income, seniority, etc.it include data of 1470 employees.

From this dataset we can uncover the factors that lead to employee attrition and explore important questions such as how does the distance from home to work affect employee attrition or how does the monthly income relate to employee attrition

Gender policy

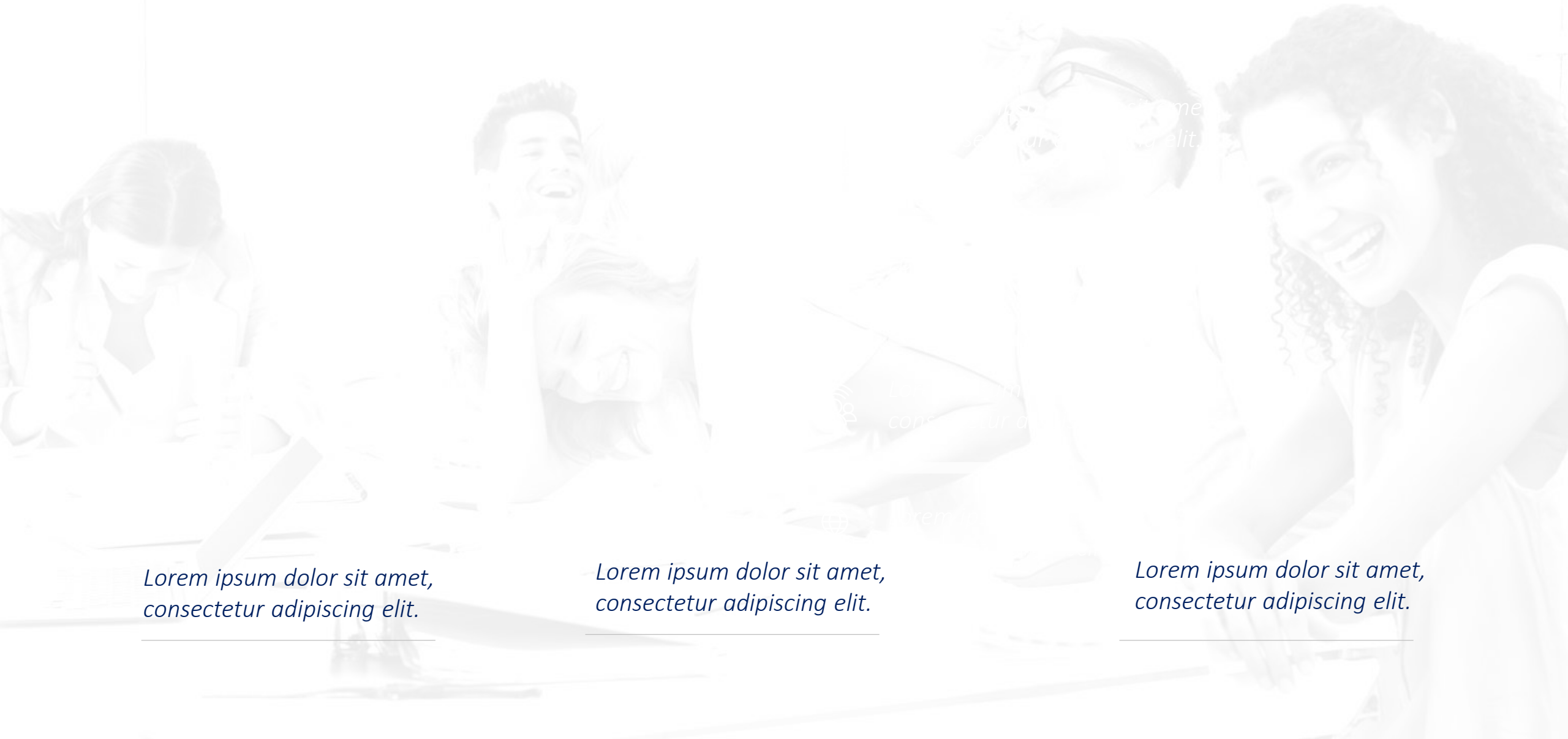


Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Promotion Pattern

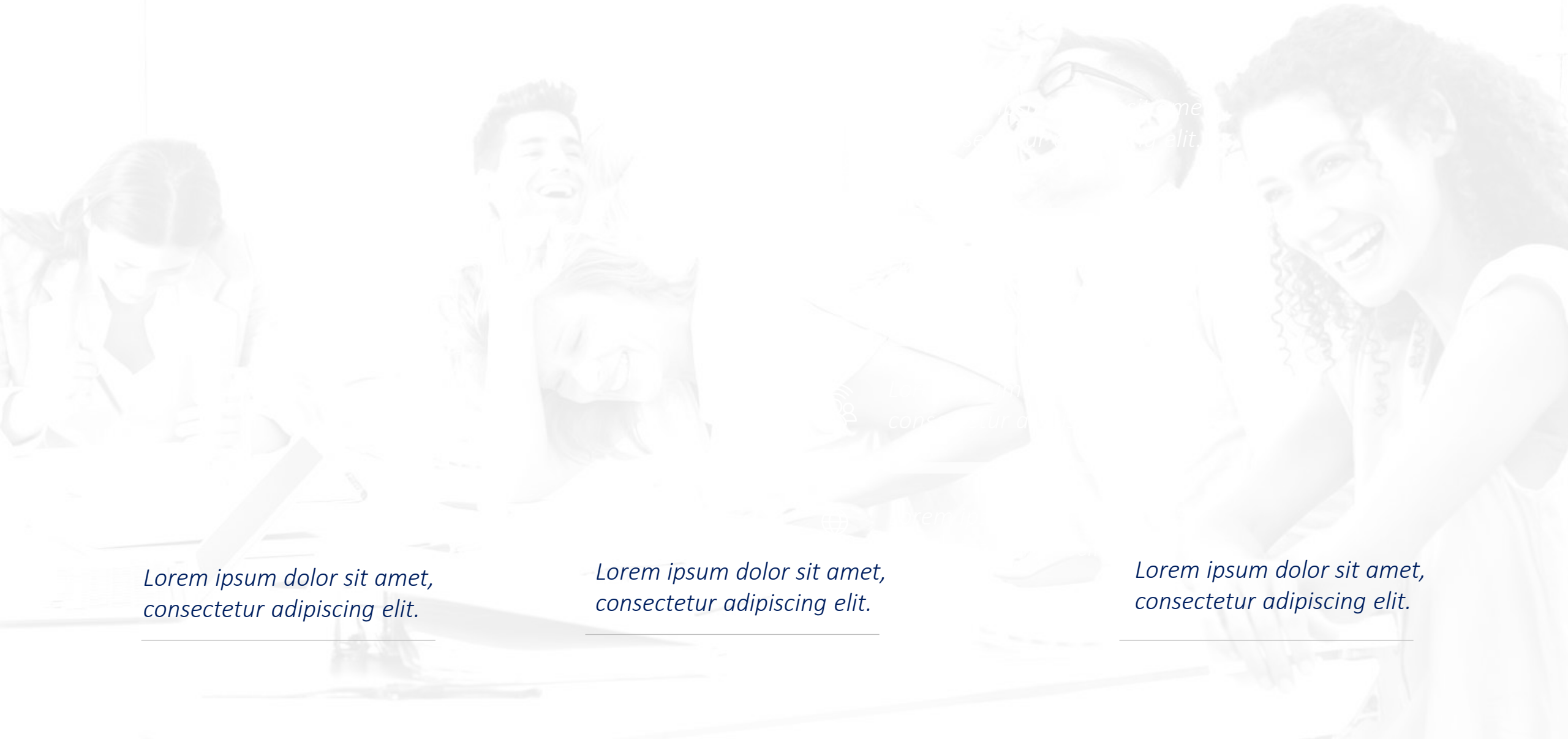


Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Benefits and compensation



Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Lorem ipsum dolor sit amet, consectetur adipiscing elit.