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

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

OBJECTIVES

- *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.*

Dataset

Age	Attrition	BusinessTravel	DailyRate	Department	Distance	Education	EducationField	EmployeeNo	EnvironmentSatisfaction	Gender	JobInvolv	JobLevel	JobRole	JobSatisfaction	MaritalSt	MonthlyI	NumCom	Over18	OverTime	PercentS	Performa	Relation	Standard	StockOpt	TotalWo	Training	WorkLife
41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	2	Female	3	2	Sales Executive	4	Single	5993	8	Y	Yes	11	3	1	80	0	8	0	1
49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	2	3	Male	2	2	Research Scientist	2	Married	5130	1	Y	No	23	4	4	80	1	10	3	3
37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	4	4	Male	2	1	Laboratory Technician	3	Single	2090	6	Y	Yes	15	3	2	80	0	7	3	3
33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	5	4	Female	3	1	Research Scientist	3	Married	2909	1	Y	Yes	11	3	3	80	0	8	3	3
27	No	Travel_Rarely	591	Research & Development	2	1	Medical	7	1	Male	3	1	Laboratory Technician	2	Married	3468	9	Y	No	12	3	4	80	1	6	3	3
32	No	Travel_Frequently	1005	Research & Development	2	2	Life Sciences	8	4	Male	3	1	Laboratory Technician	4	Single	3068	0	Y	No	13	3	3	80	0	8	2	2
59	No	Travel_Rarely	1324	Research & Development	3	3	Medical	10	3	Female	4	1	Laboratory Technician	1	Married	2670	4	Y	Yes	20	4	1	80	3	12	3	2
30	No	Travel_Rarely	1358	Research & Development	24	1	Life Sciences	11	4	Male	3	1	Laboratory Technician	3	Divorced	2693	1	Y	No	22	4	2	80	1	1	2	3
38	No	Travel_Frequently	216	Research & Development	23	3	Life Sciences	12	4	Male	2	3	Manufacturing Director	3	Single	9526	0	Y	No	21	4	2	80	0	10	2	3
36	No	Travel_Rarely	1299	Research & Development	27	3	Medical	13	3	Male	3	2	Healthcare Representative	3	Married	5237	6	Y	No	13	3	2	80	2	17	3	2
35	No	Travel_Rarely	809	Research & Development	16	3	Medical	14	1	Male	4	1	Laboratory Technician	2	Married	2426	0	Y	No	13	3	3	80	1	6	5	3
29	No	Travel_Rarely	153	Research & Development	15	2	Life Sciences	15	4	Female	2	2	Laboratory Technician	3	Single	4193	0	Y	Yes	12	3	4	80	0	10	3	3
31	No	Travel_Rarely	670	Research & Development	26	1	Life Sciences	16	1	Male	3	1	Research Scientist	3	Divorced	2911	1	Y	No	17	3	4	80	1	5	1	2
34	No	Travel_Rarely	1346	Research & Development	19	2	Medical	18	2	Male	3	1	Laboratory Technician	4	Divorced	2661	0	Y	No	11	3	3	80	1	3	2	3
28	Yes	Travel_Rarely	103	Research & Development	24	3	Life Sciences	19	3	Male	2	1	Laboratory Technician	3	Single	2028	5	Y	Yes	14	3	2	80	0	6	4	3
29	No	Travel_Rarely	1389	Research & Development	21	4	Life Sciences	20	2	Female	4	3	Manufacturing Director	1	Divorced	9980	1	Y	No	11	3	3	80	1	10	1	3
32	No	Travel_Rarely	334	Research & Development	5	2	Life Sciences	21	1	Male	4	1	Research Scientist	2	Divorced	3298	0	Y	Yes	12	3	4	80	2	7	5	2
22	No	Non-Travel	1123	Research & Development	16	2	Medical	22	4	Male	4	1	Laboratory Technician	4	Divorced	2935	1	Y	Yes	13	3	2	80	2	1	2	2
53	No	Travel_Rarely	1219	Sales	2	4	Life Sciences	23	1	Female	2	4	Manager	4	Married	15427	2	Y	No	16	3	3	80	0	31	3	3
38	No	Travel_Rarely	371	Research & Development	2	3	Life Sciences	24	4	Male	3	1	Research Scientist	4	Single	3944	5	Y	Yes	11	3	3	80	0	6	3	3
24	No	Non-Travel	673	Research & Development	11	2	Other	26	1	Female	4	2	Manufacturing Director	3	Divorced	4011	0	Y	No	18	3	4	80	1	5	5	2
36	Yes	Travel_Rarely	1218	Sales	9	4	Life Sciences	27	3	Male	2	1	Sales Representative	1	Single	3407	7	Y	No	23	4	2	80	0	10	4	3
34	No	Travel_Rarely	419	Research & Development	7	4	Life Sciences	28	1	Female	3	3	Research Director	2	Single	11994	0	Y	No	11	3	3	80	0	13	4	3
21	No	Travel_Rarely	391	Research & Development	15	2	Life Sciences	30	3	Male	3	1	Research Scientist	4	Single	1232	1	Y	No	14	3	4	80	0	0	6	3
34	Yes	Travel_Rarely	699	Research & Development	6	1	Medical	31	2	Male	3	1	Research Scientist	1	Single	2960	2	Y	No	11	3	3	80	0	8	2	3
53	No	Travel_Rarely	1282	Research & Development	5	3	Other	32	3	Female	3	5	Manager	3	Divorced	19094	4	Y	No	11	3	4	80	1	26	3	2
32	Yes	Travel_Frequently	1125	Research & Development	16	1	Life Sciences	33	2	Female	1	1	Research Scientist	1	Single	3919	1	Y	Yes	22	4	2	80	0	10	5	3
42	No	Travel_Rarely	691	Sales	8	4	Marketing	35	3	Male	3	2	Sales Executive	2	Married	6825	0	Y	No	11	3	4	80	1	10	2	3
44	No	Travel_Rarely	477	Research & Development	7	4	Medical	36	1	Female	2	3	Healthcare Representative	4	Married	10248	3	Y	No	14	3	4	80	1	24	4	3
46	No	Travel_Rarely	705	Sales	2	4	Marketing	38	2	Female	3	5	Manager	1	Single	18947	3	Y	No	12	3	4	80	0	22	2	2
33	No	Travel_Rarely	924	Research & Development	2	3	Medical	39	3	Male	3	1	Laboratory Technician	4	Single	2496	4	Y	No	11	3	4	80	0	7	3	3
44	No	Travel_Rarely	1459	Research & Development	10	4	Other	40	4	Male	3	2	Healthcare Representative	4	Married	6465	2	Y	Yes	13	3	4	80	0	9	5	4
30	No	Travel_Rarely	125	Research & Development	9	2	Medical	41	4	Male	2	1	Laboratory Technician	3	Single	2206	1	Y	No	13	3	1	80	0	10	5	3
39	Yes	Travel_Rarely	895	Sales	5	3	Technical Degree	42	4	Male	3	2	Sales Representative	4	Married	2086	3	Y	No	14	3	3	80	1	19	6	4
24	Yes	Travel_Rarely	813	Research & Development	1	3	Medical	45	2	Male	3	1	Research Scientist	4	Married	2293	2	Y	Yes	16	3	1	80	1	6	2	2
43	No	Travel_Rarely	1273	Research & Development	2	2	Medical	46	4	Female	4	1	Research Scientist	3	Divorced	2645	1	Y	No	12	3	4	80	2	6	3	2
50	Yes	Travel_Rarely	869	Sales	3	2	Marketing	47	1	Male	2	1	Sales Representative	3	Married	2683	1	Y	Yes	14	3	3	80	0	3	2	3
35	No	Travel_Rarely	890	Sales	2	3	Marketing	49	4	Female	3	1	Sales Representative	4	Married	2014	1	Y	No	13	3	1	80	0	2	3	3
36	No	Travel_Rarely	852	Research & Development	5	4	Life Sciences	51	2	Female	2	1	Research Scientist	1	Married	3419	9	Y	Yes	14	3	4	80	1	6	3	4
33	No	Travel_Frequently	1141	Sales	1	3	Life Sciences	52	3	Female	4	2	Sales Executive	1	Married	5376	2	Y	No	19	3	1	80	2	10	3	3
35	No	Travel_Rarely	464	Research & Development	4	2	Other	53	3	Male	3	1	Laboratory Technician	4	Divorced	1951	1	Y	No	12	3	3	80	1	1	3	3
27	No	Travel_Rarely	1240	Research & Development	2	4	Life Sciences	54	4	Female	3	1	Laboratory Technician	1	Divorced	2341	1	Y	No	13	3	4	80	1	1	6	3

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

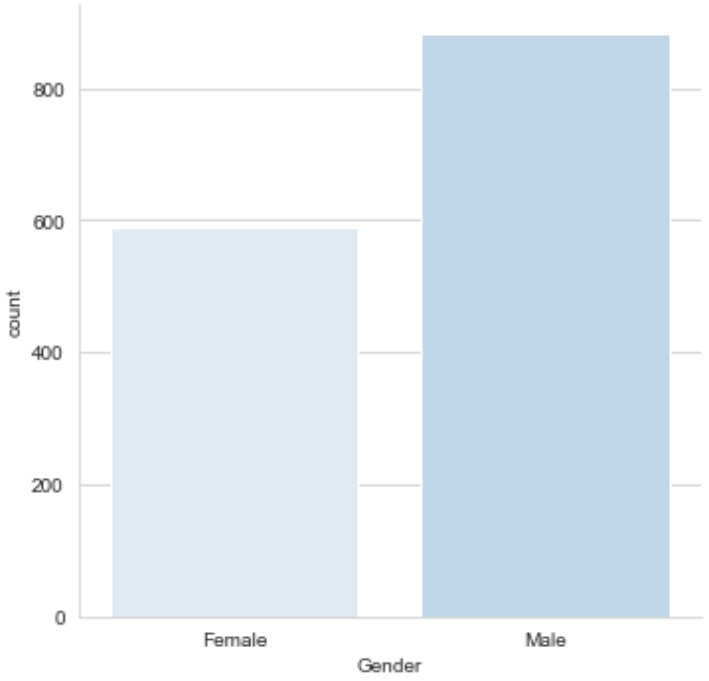
Distance Travelled

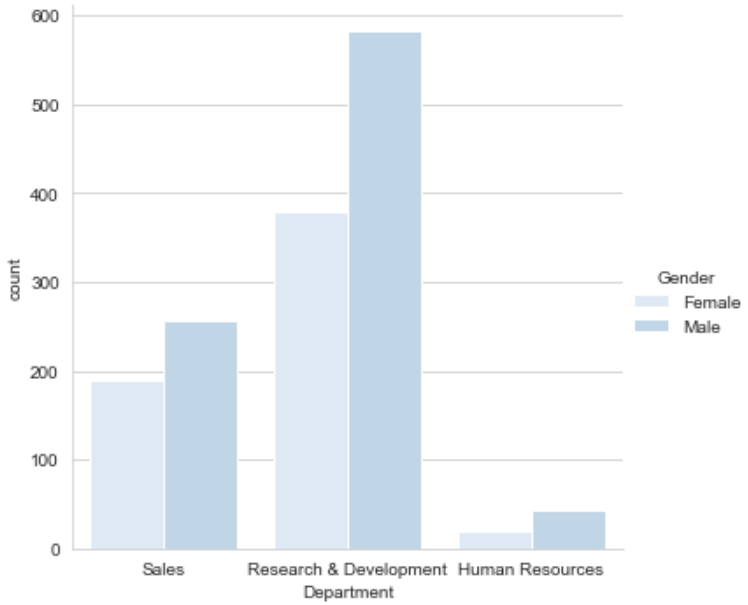
Less than 10 km is Short Distance, 11-19 is moderately far while 20 and above is far

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

Demographics

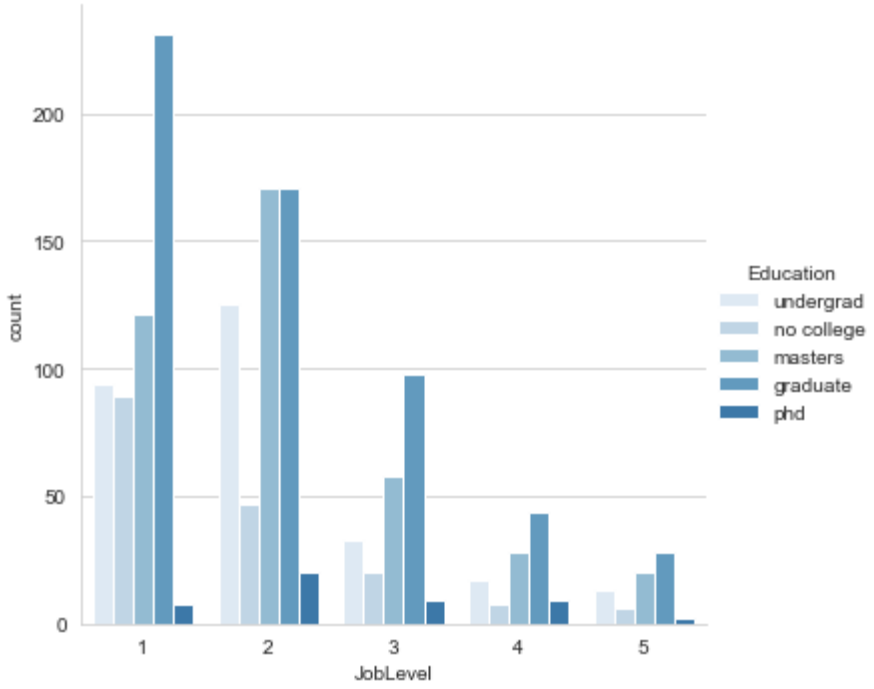


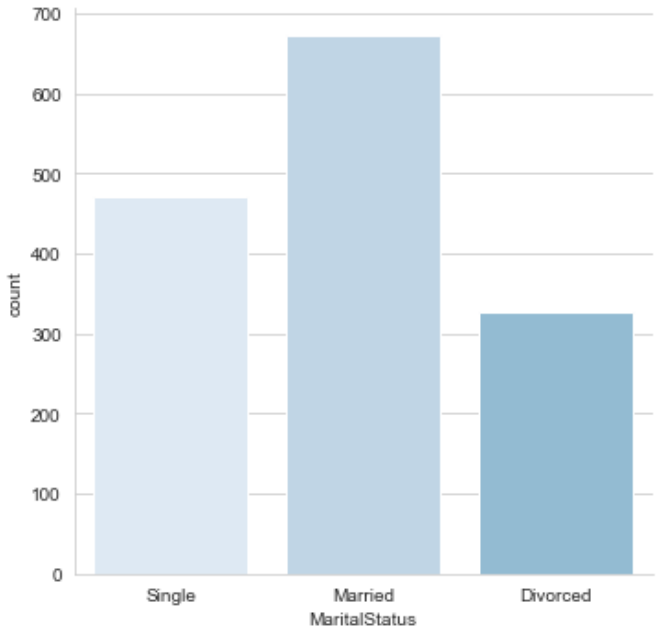


The dataset consists of 1470 employees. There are 588 female staff and about 882 male staff

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



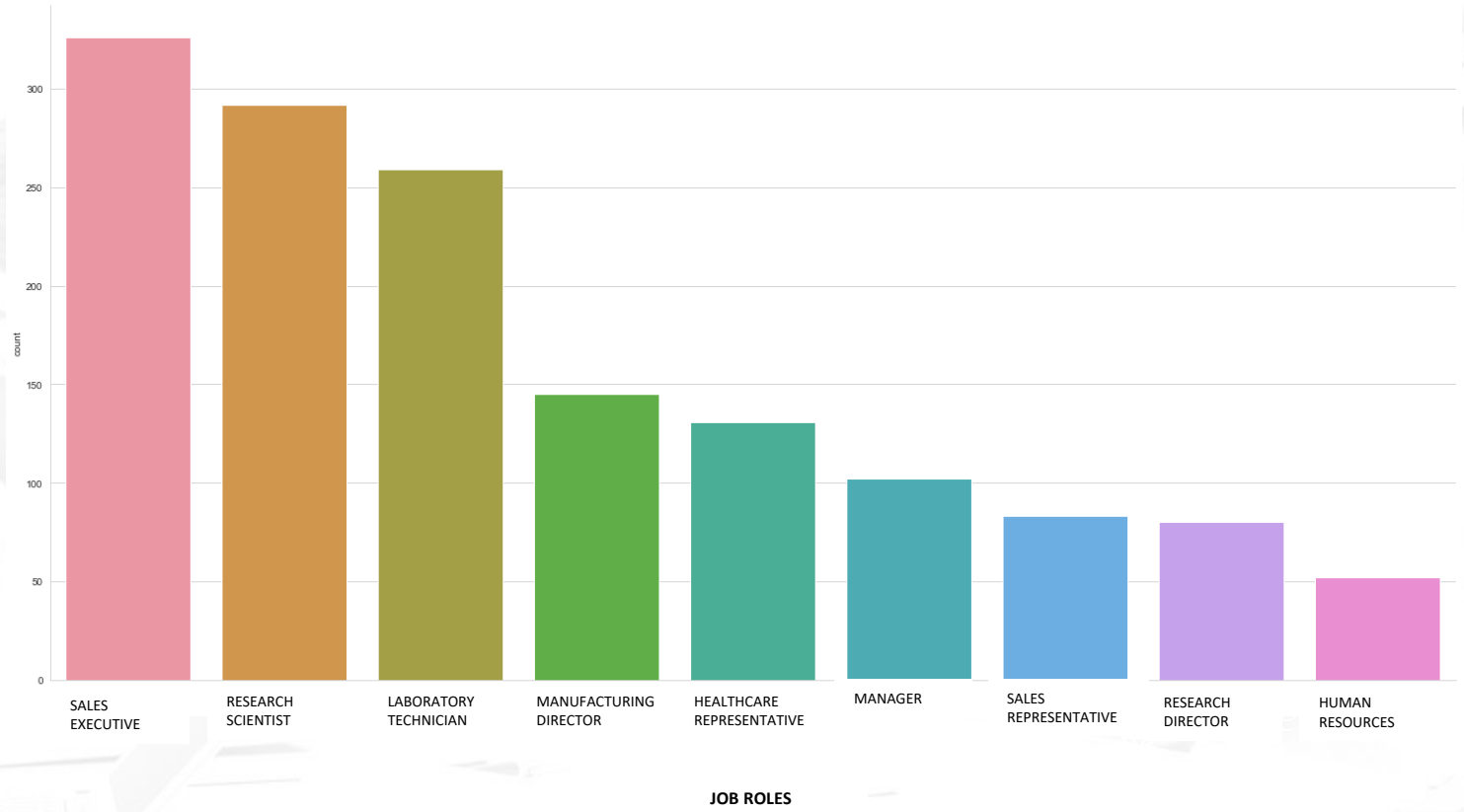


The qualifications of the employees ranges from Bachelors, Masters and Ph.D degree. People with no college degrees and college students are also employed

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

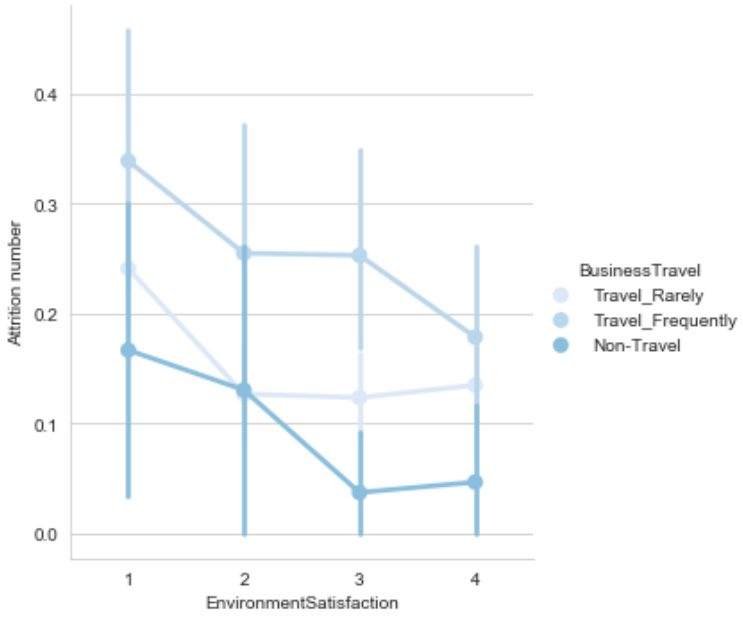
Most employees are married but there are also Single and Divorced people

Demographics

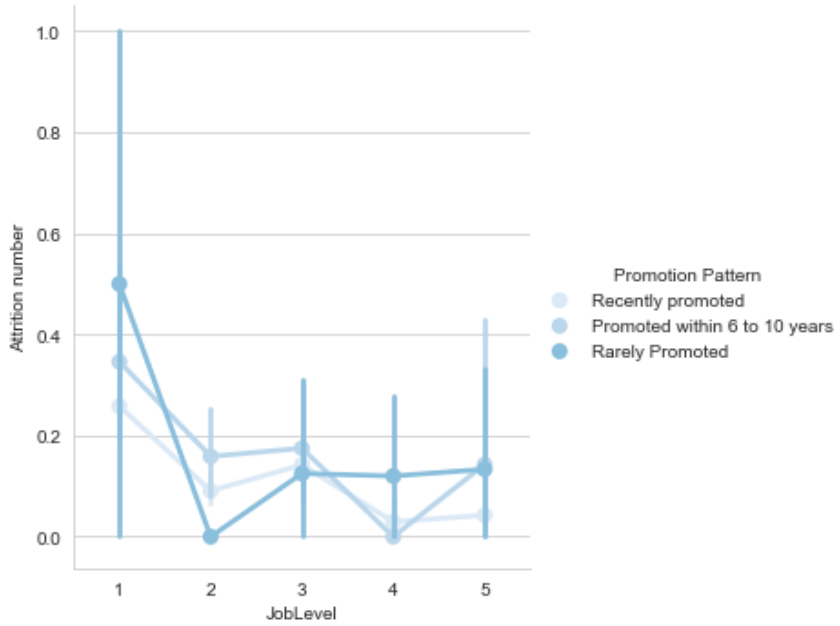


The job roles include: sales executive, research scientist, laboratory technician, Manufacturing Director, Healthcare Representatives, Managers, Sales representatives and Human resources. Most Employees are hired for Sales related roles and research roles

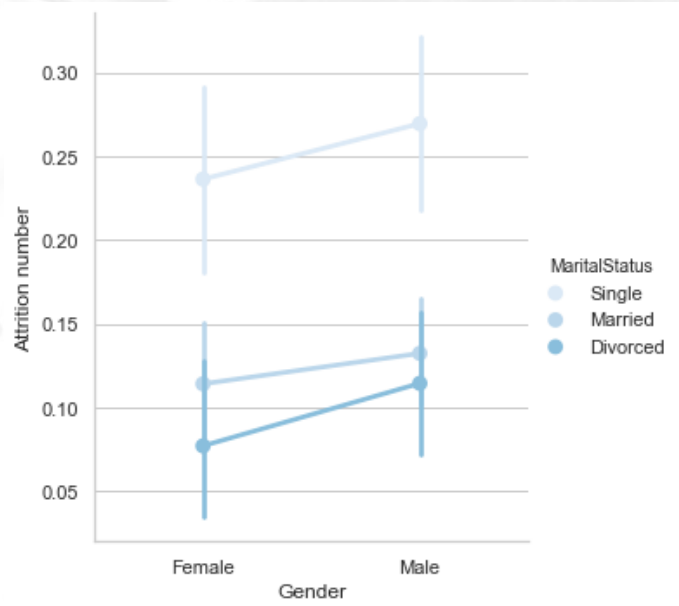
Factors responsible for Employee Attrition



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

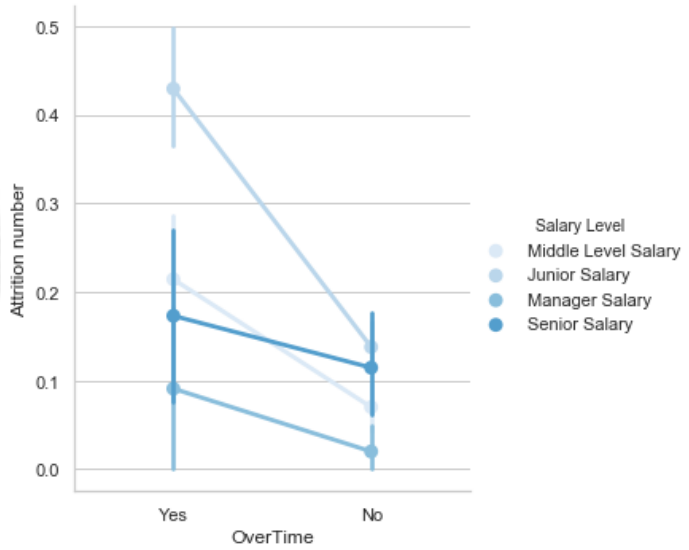


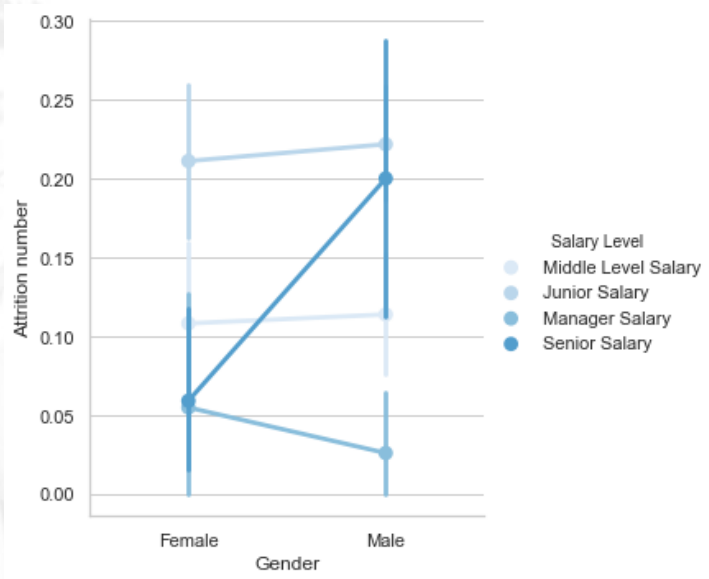
Employees who rarely get promoted and are at the bottom of the job ladder will most likely eventually leave their jobs compared to manager level staff who just got promoted recently



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

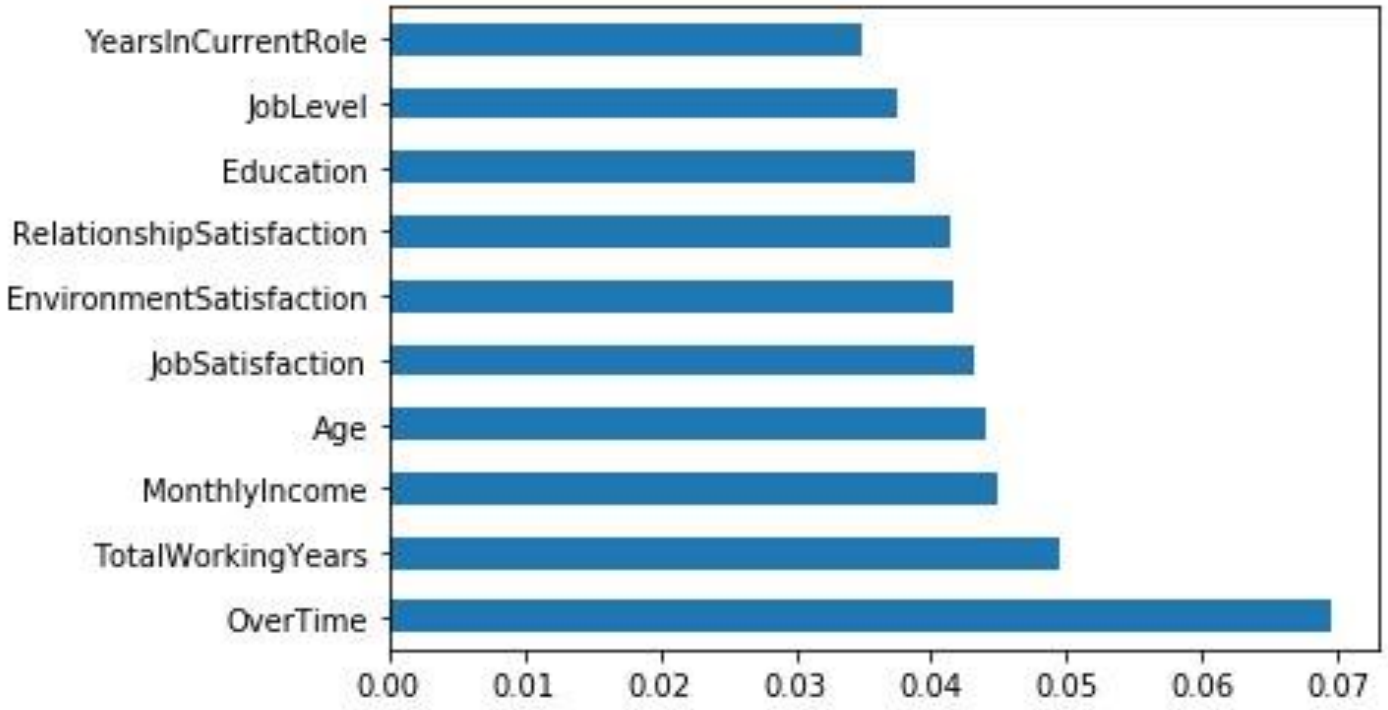




Employees who are paid the lowest and still have to work overtime will be more likely to leave their jobs. The people least likely to quit are those who are paid the highest and do not have to work overtime

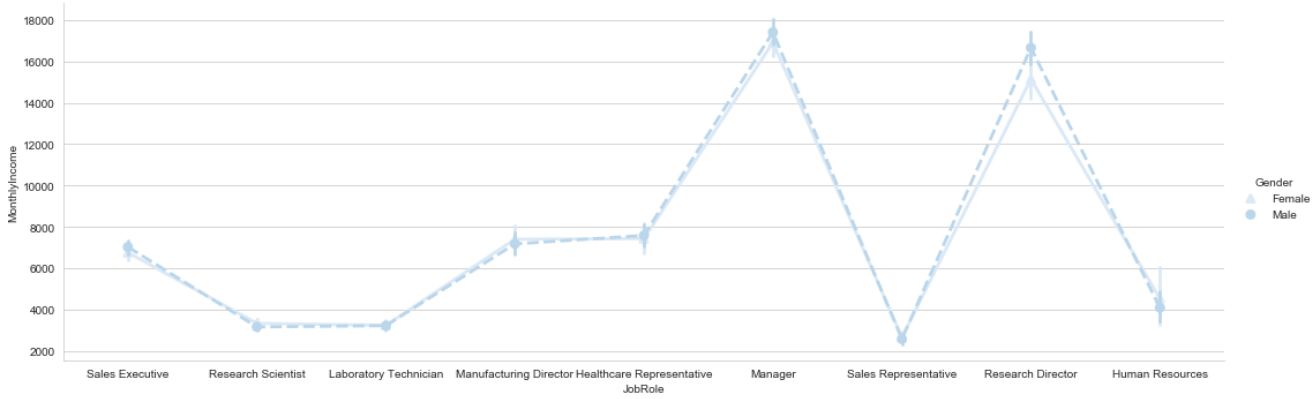
Females are less likely to quit their jobs than their male counterparts on the same salary level. Although men who have achieved the highest salary level are the least likely to quit

Feature Importance

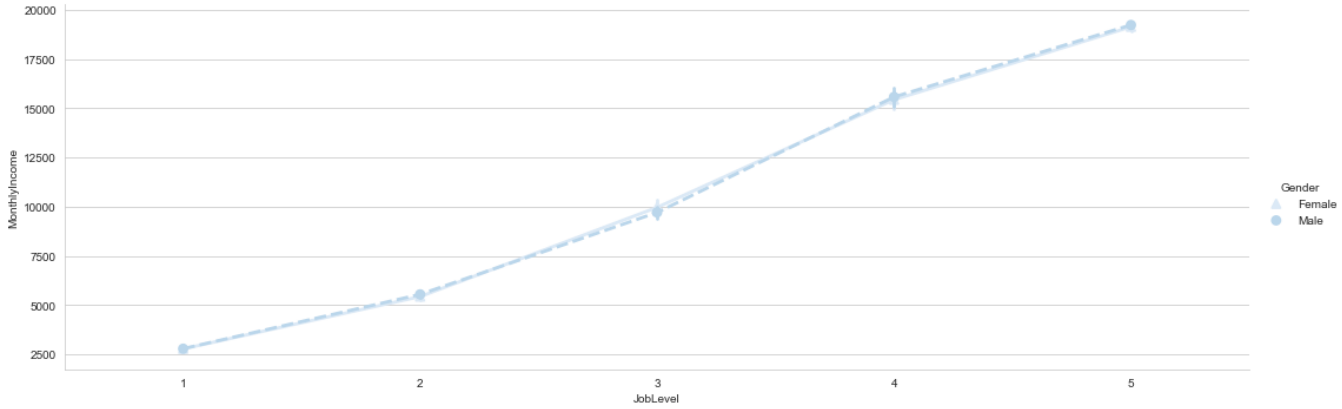


The major reason for employee attrition was determined to be the constant amount of overtime by some employees. After spending a large amount of time at the company, employees also tend to leave either to start their own companies or due to the retirement age.
The monthly income is also a major factor 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, 0, 1, 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, 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, 1, 0, 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, 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, 1, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0])
```

	precision	recall	f1-score	support
0	0.89	0.99	0.94	411
1	0.85	0.35	0.50	79
accuracy			0.89	490
macro avg	0.87	0.67	0.72	490
weighted avg	0.88	0.89	0.87	490

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')
```

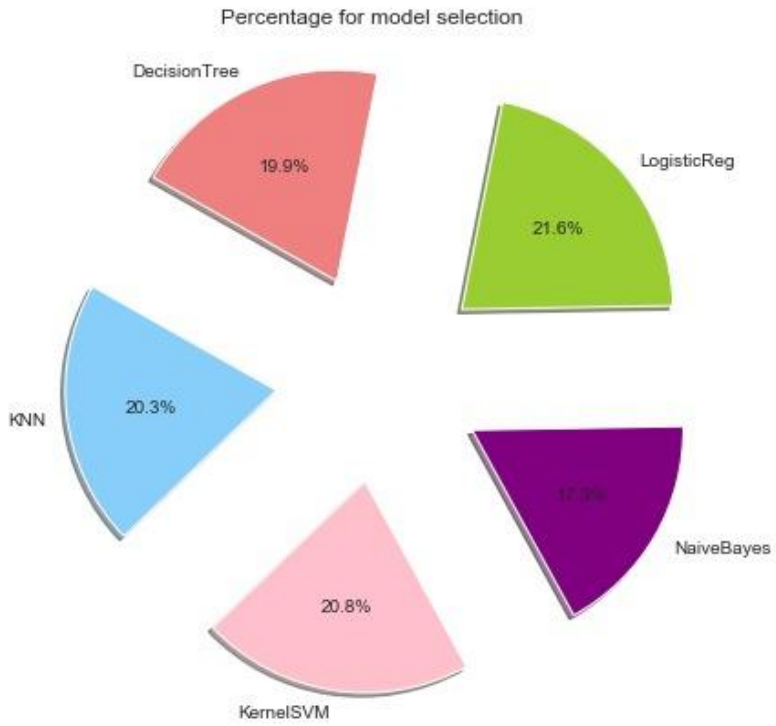
The model has been able to predict if an employee is going to leave his or her job. '1' means the employee will leave and '0' means the employee will stay.
The accuracy of the model was tested and returned a score of 80%

```
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
```

Predictions

```
LogisticReg: 0.870408 (0.028149)
DecisionTree: 0.810204 (0.027074)
KNN: 0.818367 (0.028425)
KernelSVM: 0.838776 (0.025734)
NaiveBayes: 0.695918 (0.031218)
```

The model with the highest accuracy score is th
logistic Regression followed by the Kernel SVM
then the Decision Tree Classifier



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

CREDITS

Dataset: https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset

Presentation Template: *PowerPoint online Templates*