



# EMPLOYEE ATTRITION

**Introduction**

**Data Cleaning**

**Factors**

**Factors**

**Random Forest**

**Conclusion**

# Employee Attrition

The key to success in any organization is attracting and retaining top talent.

Companies need to know what factors they can change to prevent the loss of good people.

**Aim/Study**

## **Aim :**

Determine which factors keep employees at the company and which prompt others to leave.

## Null Hypothesis

- 01** Salary = main factor
- 02** Older Employee leaves the company more
- 03** Sales department have the highest mean salary

# Data Cleaning

Check

Desc

Replace

Check null value in dataset

```
In [11]: df.isnull().sum()
```

age	0
salary	0
education	0
employeecount	0
environmentindex	0
hourlyrate	0
jobinvolvement	0
jobsecurity	0
jobtenure	0
jobvariety	0
managementpractices	0
managementpractices2	0
managementpractices3	0
managementpractices4	0
managementpractices5	0
managementpractices6	0
managementpractices7	0
managementpractices8	0
managementpractices9	0
managementpractices10	0
managementpractices11	0
managementpractices12	0
managementpractices13	0
managementpractices14	0
managementpractices15	0
managementpractices16	0
managementpractices17	0
managementpractices18	0
managementpractices19	0
managementpractices20	0
managementpractices21	0
managementpractices22	0
managementpractices23	0
managementpractices24	0
managementpractices25	0
managementpractices26	0
managementpractices27	0
managementpractices28	0
managementpractices29	0
managementpractices30	0
managementpractices31	0
managementpractices32	0
managementpractices33	0
managementpractices34	0
managementpractices35	0
managementpractices36	0
managementpractices37	0
managementpractices38	0
managementpractices39	0
managementpractices40	0
managementpractices41	0
managementpractices42	0
managementpractices43	0
managementpractices44	0
managementpractices45	0
managementpractices46	0
managementpractices47	0
managementpractices48	0
managementpractices49	0
managementpractices50	0
managementpractices51	0
managementpractices52	0
managementpractices53	0
managementpractices54	0
managementpractices55	0
managementpractices56	0
managementpractices57	0
managementpractices58	0
managementpractices59	0
managementpractices60	0
managementpractices61	0
managementpractices62	0
managementpractices63	0
managementpractices64	0
managementpractices65	0
managementpractices66	0
managementpractices67	0
managementpractices68	0
managementpractices69	0
managementpractices70	0
managementpractices71	0
managementpractices72	0
managementpractices73	0
managementpractices74	0
managementpractices75	0
managementpractices76	0
managementpractices77	0
managementpractices78	0
managementpractices79	0
managementpractices80	0
managementpractices81	0
managementpractices82	0
managementpractices83	0
managementpractices84	0
managementpractices85	0
managementpractices86	0
managementpractices87	0
managementpractices88	0
managementpractices89	0
managementpractices90	0
managementpractices91	0
managementpractices92	0
managementpractices93	0
managementpractices94	0
managementpractices95	0
managementpractices96	0
managementpractices97	0
managementpractices98	0
managementpractices99	0
managementpractices100	0

Check Description

```
In [11]: df.describe()
```

```
Out[11]:
```

	age	salary	education	employeecount	environmentindex	hourlyrate	jobinvolvement
count	1475.000000	1475.000000	1475.000000	1475.000000	1475.000000	1475.000000	1475.000000
mean	36.412914	602.667174	6.926747	2.615626	1.6	1026.855468	2.717189
std	8.170473	402.589136	0.960884	1.024185	0.8	402.861635	0.711700
min	18.000000	102.000000	1.000000	1.000000	1.0	1.000000	1.000000
25%	30.000000	440.000000	2.000000	2.000000	1.0	240.000000	2.000000
50%	36.000000	602.000000	7.000000	3.000000	1.0	1026.000000	3.000000
75%	42.000000	1107.000000	14.000000	4.000000	1.0	1955.750000	5.000000
max	60.000000	1400.000000	20.000000	5.000000	1.0	2000.000000	6.000000

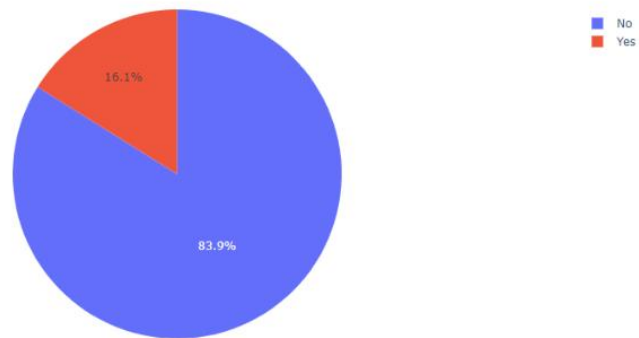
Replacing numerical values into categorical values

```
In [11]: pd.Series([1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100]).replace([1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100],['A','B','C','D','E','F','G','H','I','J','K','L','M','N','O','P','Q','R','S','T','U','V','W','X','Y','Z','AA','AB','AC','AD','AE','AF','AG','AH','AI','AJ','AK','AL','AM','AN','AO','AP','AQ','AR','AS','AT','AU','AV','AW','AX','AY','AZ','BA','BB','BC','BD','BE','BF','BG','BH','BI','BJ','BK','BL','BM','BN','BO','BP','BQ','BR','BS','BT','BU','BV','BW','BX','BY','BZ','CA','CB','CC','CD','CE','CF','CG','CH','CI','CJ','CK','CL','CM','CN','CO','CP','CQ','CR','CS','CT','CU','CV','CW','CX','CY','CZ','DA','DB','DC','DD','DE','DF','DG','DH','DI','DJ','DK','DL','DM','DN','DO','DP','DQ','DR','DS','DT','DU','DV','DW','DX','DY','DZ','EA','EB','EC','ED','EE','EF','EG','EH','EI','EJ','EK','EL','EM','EN','EO','EP','EQ','ER','ES','ET','EU','EV','EW','EX','EY','EZ','FA','FB','FC','FD','FE','FF','FG','FH','FI','FJ','FK','FL','FM','FN','FO','FP','FQ','FR','FS','FT','FU','FV','FW','FX','FY','FZ','GA','GB','GC','GD','GE','GF','GG','GH','GI','GJ','GK','GL','GM','GN','GO','GP','GQ','GR','GS','GT','GU','GV','GW','GX','GY','GZ','HA','HB','HC','HD','HE','HF','HG','HH','HI','HJ','HK','HL','HM','HN','HO','HP','HQ','HR','HS','HT','HU','HV','HW','HX','HY','HZ','IA','IB','IC','ID','IE','IF','IG','IH','II','IJ','IK','IL','IM','IN','IO','IP','IQ','IR','IS','IT','IU','IV','IW','IX','IY','IZ','JA','JB','JC','JD','JE','JF','JG','JH','JI','JJ','JK','JL','JM','JN','JO','JP','JQ','JR','JS','JT','JU','JV','JW','JX','JY','JZ','KA','KB','KC','KD','KE','KF','KG','KH','KI','KJ','KK','KL','KM','KN','KO','KP','KQ','KR','KS','KT','KU','KV','KW','KX','KY','KZ','LA','LB','LC','LD','LE','LF','LG','LH','LI','LJ','LK','LL','LM','LN','LO','LP','LQ','LR','LS','LT','LU','LV','LW','LX','LY','LZ','MA','MB','MC','MD','ME','MF','MG','MH','MI','MJ','MK','ML','MM','MN','MO','MP','MQ','MR','MS','MT','MU','MV','MW','MX','MY','MZ','NA','NB','NC','ND','NE','NF','NG','NH','NI','NJ','NK','NL','NM','NN','NO','NP','NQ','NR','NS','NT','NU','NV','NW','NX','NY','NZ','OA','OB','OC','OD','OE','OF','OG','OH','OI','OJ','OK','OL','OM','ON','OO','OP','OQ','OR','OS','OT','OU','OV','OW','OX','OY','OZ','PA','PB','PC','PD','PE','PF','PG','PH','PI','PJ','PK','PL','PM','PN','PO','PP','PQ','PR','PS','PT','PU','PV','PW','PX','PY','PZ','QA','QB','QC','QD','QE','QF','QG','QH','QI','QJ','QK','QL','QM','QN','QO','QP','QQ','QR','QS','QT','QU','QV','QW','QX','QY','QZ','RA','RB','RC','RD','RE','RF','RG','RH','RI','RJ','RK','RL','RM','RN','RO','RP','RQ','RR','RS','RT','RU','RV','RW','RX','RY','RZ','SA','SB','SC','SD','SE','SF','SG','SH','SI','SJ','SK','SL','SM','SN','SO','SP','SQ','SR','SS','ST','SU','SV','SW','SX','SY','SZ','TA','TB','TC','TD','TE','TF','TG','TH','TI','TJ','TK','TL','TM','TN','TO','TP','TQ','TR','TS','TT','TU','TV','TW','TX','TY','TZ','UA','UB','UC','UD','UE','UF','UG','UH','UI','UJ','UK','UL','UM','UN','UO','UP','UQ','UR','US','UT','UU','UV','UW','UX','UY','UZ','VA','VB','VC','VD','VE','VF','VG','VH','VI','VJ','VK','VL','VM','VN','VO','VP','VQ','VR','VS','VT','VU','VV','VW','VX','VY','VZ','WA','WB','WC','WD','WE','WF','WG','WH','WI','WJ','WK','WL','WM','WN','WO','WP','WQ','WR','WS','WT','WU','WV','WW','WX','WY','WZ','XA','XB','XC','XD','XE','XF','XG','XH','XI','XJ','XK','XL','XM','XN','XO','XP','XQ','XR','XS','XT','XU','XV','XW','XX','XY','XZ','YA','YB','YC','YD','YE','YF','YG','YH','YI','YJ','YK','YL','YM','YN','YO','YP','YQ','YR','YS','YT','YU','YV','YW','YX','YY','YZ','ZA','ZB','ZC','ZD','ZE','ZF','ZG','ZH','ZI','ZJ','ZK','ZL','ZM','ZN','ZO','ZP','ZQ','ZR','ZS','ZT','ZU','ZV','ZW','ZX','ZY','ZZ']).
```

## Factors

### Rate :

Attrition Rate



**Gender**

**Age**

**Education**

**Job Role**



## Attrition by Gender

### Question :

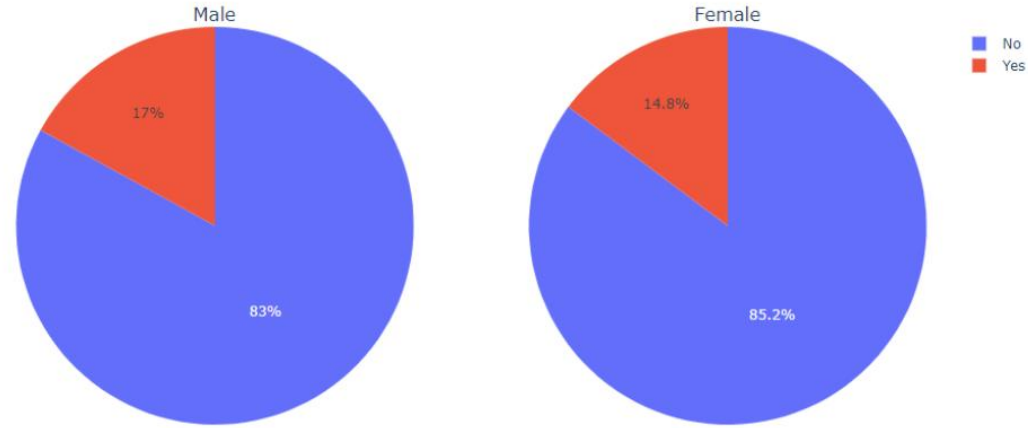
Does gender affect rate of attrition ?

**Distribution**

**Summary**

# Pie Chart

Attrition Rate by Gender



Attrition	Gender	EmployeeCount
No	Female	501
No	Male	732
Yes	Female	87
Yes	Male	150

## Summary

It seems like there's approximately 2.2% higher attrition rate for male compared to female, but it shouldn't be a big factor of employee attrition



## Attrition by Age

### Question :

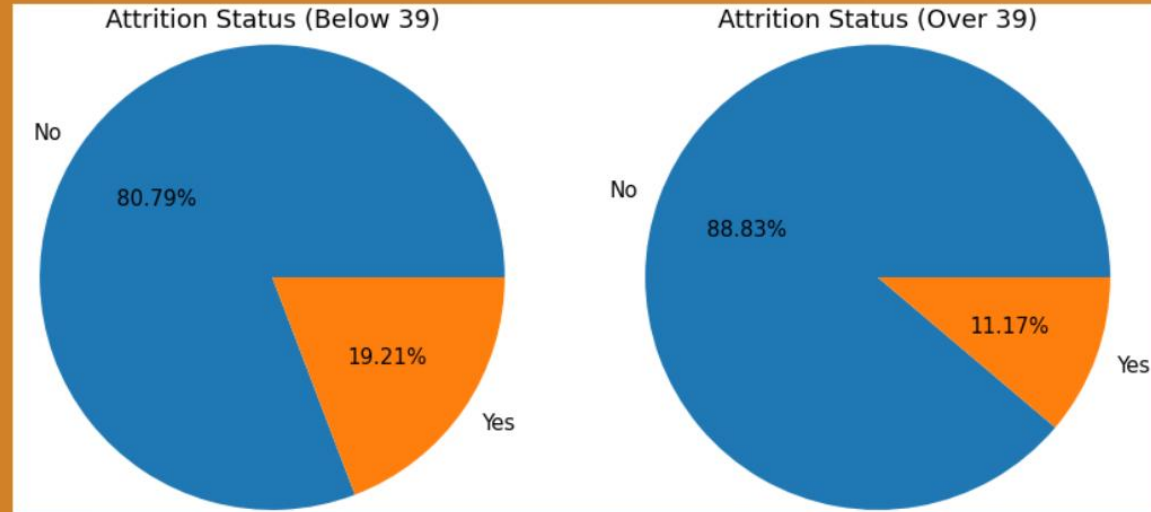
What is the range of age which leaves the organization more frequently than the others?

**Distribution**

**Summary**

# Age Distribution with Attrition

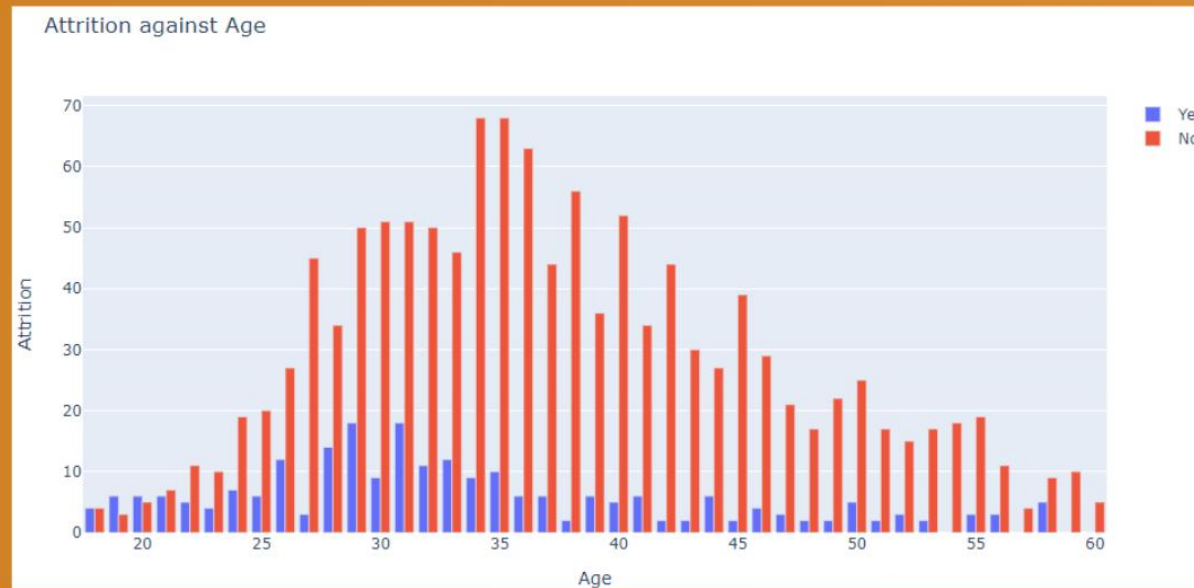
**Total  
Working  
Years**



People who are young (Millennial) seems to have a higher attrition rate than the older ones (Boomers).

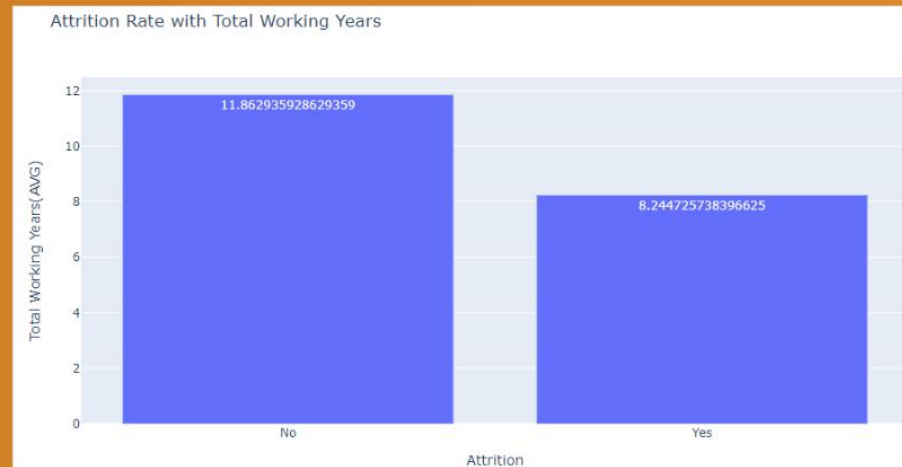
**Detailed  
Age**

# Detailed



Range of most people who leave  
the organization : 25-35

# Total Working Years



It seems like those who have less total working years leaves the organization which is true as Millennial has a higher attrition rate.

# Summary

It seems that millennial are the ones with the highest turnover rate, followed by the boomers. What does this tell us? The newer generation which are the millennial opt to look more easy for other jobs that satisfy the needs on the other side we have the boomers which are approximating retirement and could be one of the reasons why the turnover rate of boomers is the second highest.

## Education Level

### Question :

Is Education Level a factor of  
employee attrition?

**Count**

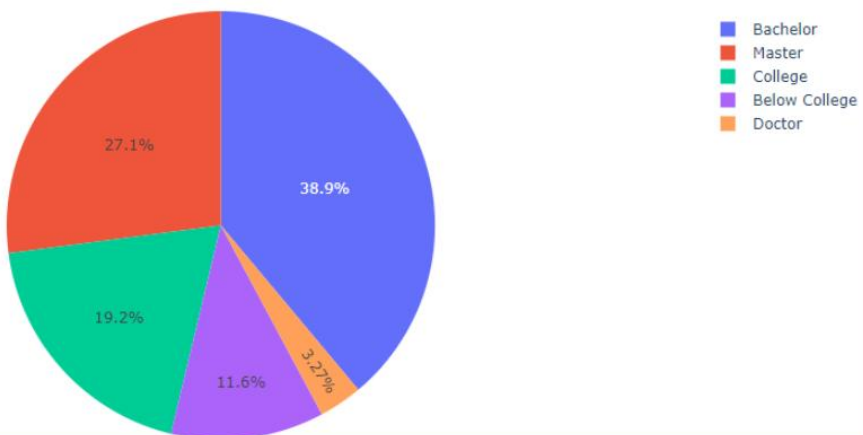
**Attrition**

**Summary**



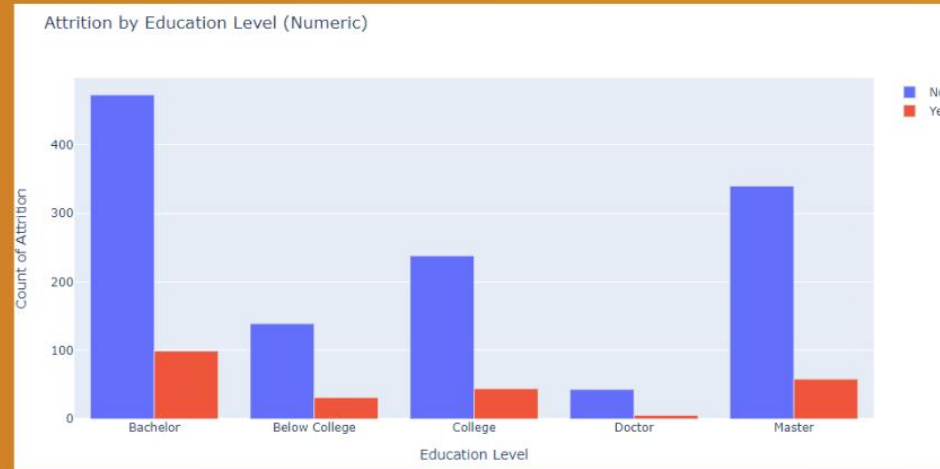
# Count

Percentage of each Education Level



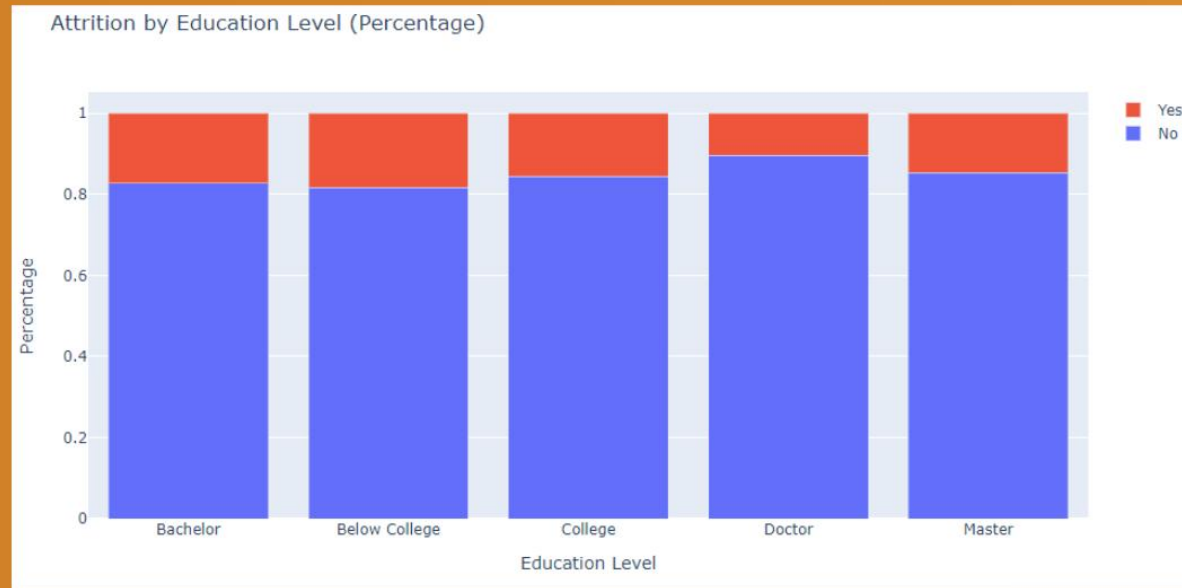
**Percentage**

**Attrition**



Looks like Bachelor holder have a higher Attrition Rate

# Percentage



It seems that the rate of attrition are quite the same



## Summary

It seems like education is a main factor of Attrition at first. However, if we compare it into percentage, they are quite the same. Therefore, Education Level is not a main factor of Employee Attrition

## Job Role

### Question :

How many employees are there in each Job Role? Does it affect the attrition rate?

**Distribution**

**Attrition**

# Distribution

Major Job Roles in the Organization

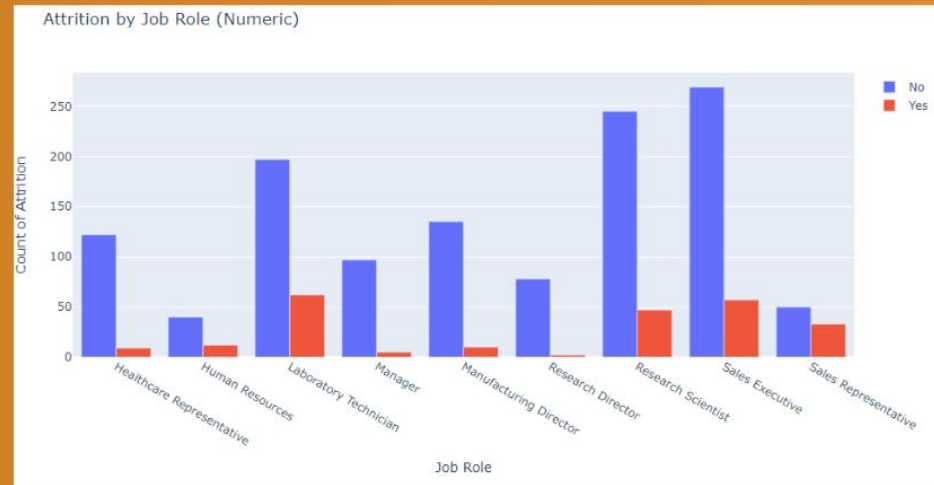


Most Employee work as Sales Executive in the organization



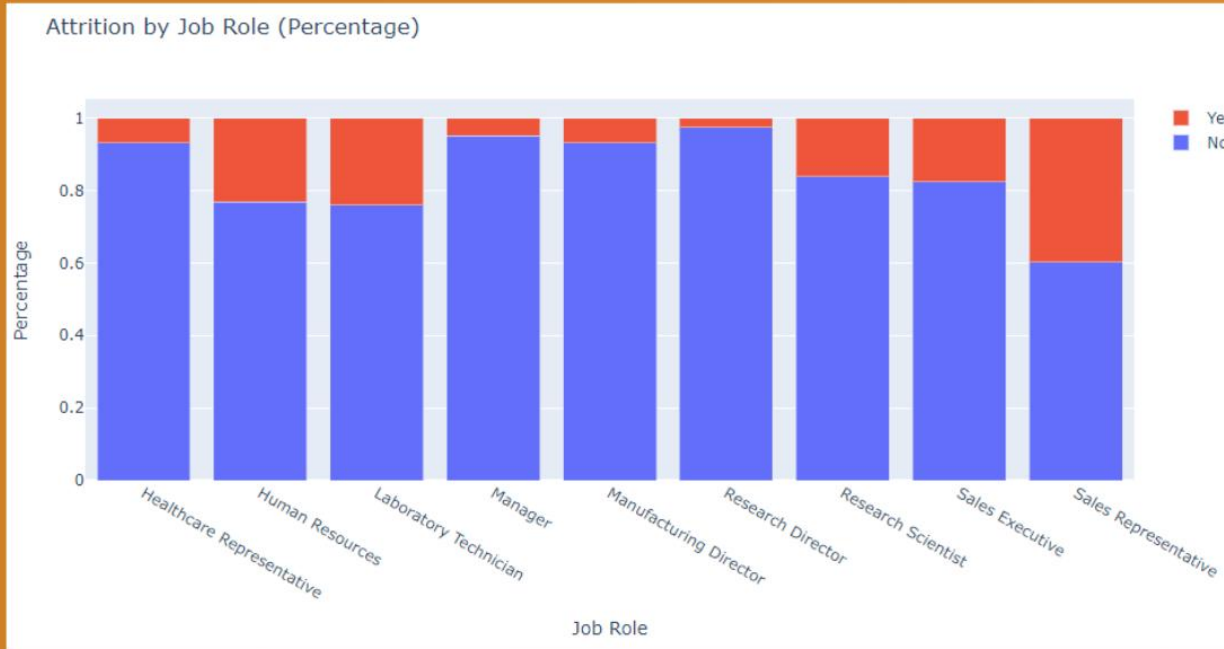
## Percentage

## Attrition



Laboratory Technician , Sales Executive  
and Research Scientist seems to have  
a higher rate of attrition

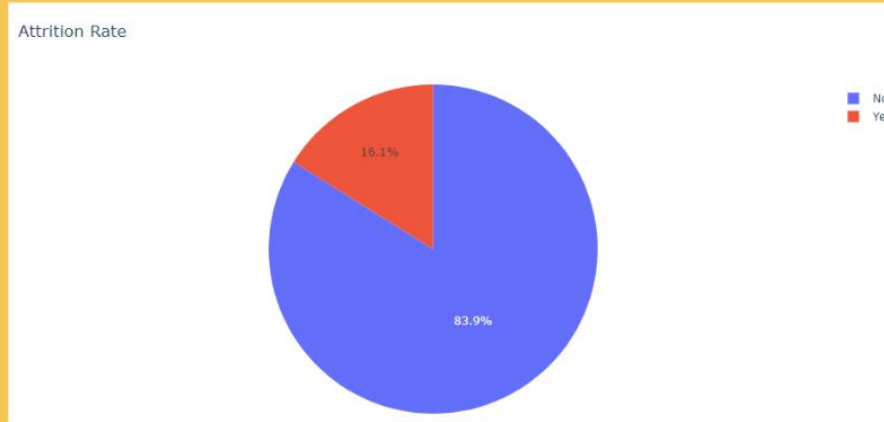
# Percentage



**Summary :** Sales Representatives, Human Resources and Laboratory Technician have the highest attrition rates. This could give us a hint that in these departments we are experiencing certain issues with employees.

## Factors

### Rate :



**Satisfaction**

**Work Life Balance**

**Salary**

**Overtime**



## **Attrition by Satisfaction**

### **Question :**

How does satisfaction affect  
attrition?

**Environment**

**Job**

# Attrition by Environment Satisfaction

## Question :

What's the working environment by job  
role?  
Is this a major factor of employee  
attrition?

**Attrition**



# Attrition



**Summary :** As expected, managers and healthcare representatives are dealing with a lower working environment however, we don't see the same with sales representatives that could be because most sales representatives work outside the organization. Those who have lower environment satisfaction tend to leave the organization too compared to those who are satisfied.





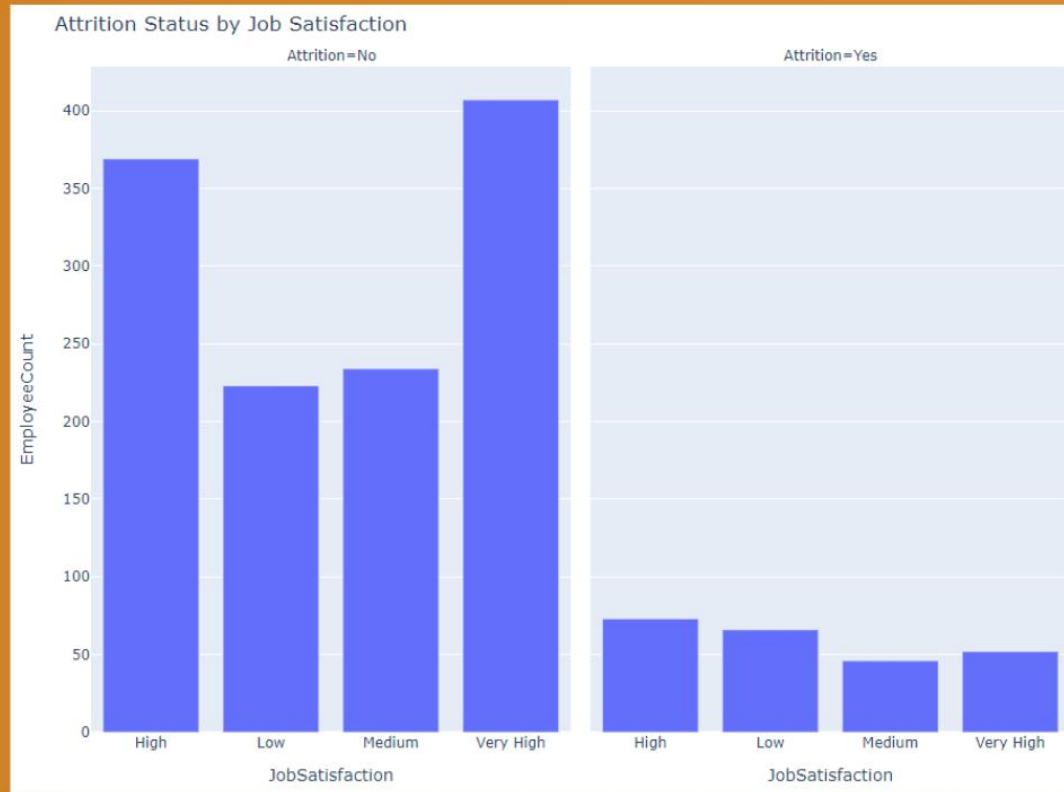
# Attrition by Job Satisfaction

## Question :

Does Job Satisfaction affect the attrition rate?  
Are individuals with a lower satisfaction quit the organization than the ones who are more satisfied?

## Attrition

# Attrition



**Summary :** Job Satisfaction seems to be quite balance for those who leaves the company and who doesn't.



## Attrition by WorkLifeBalance

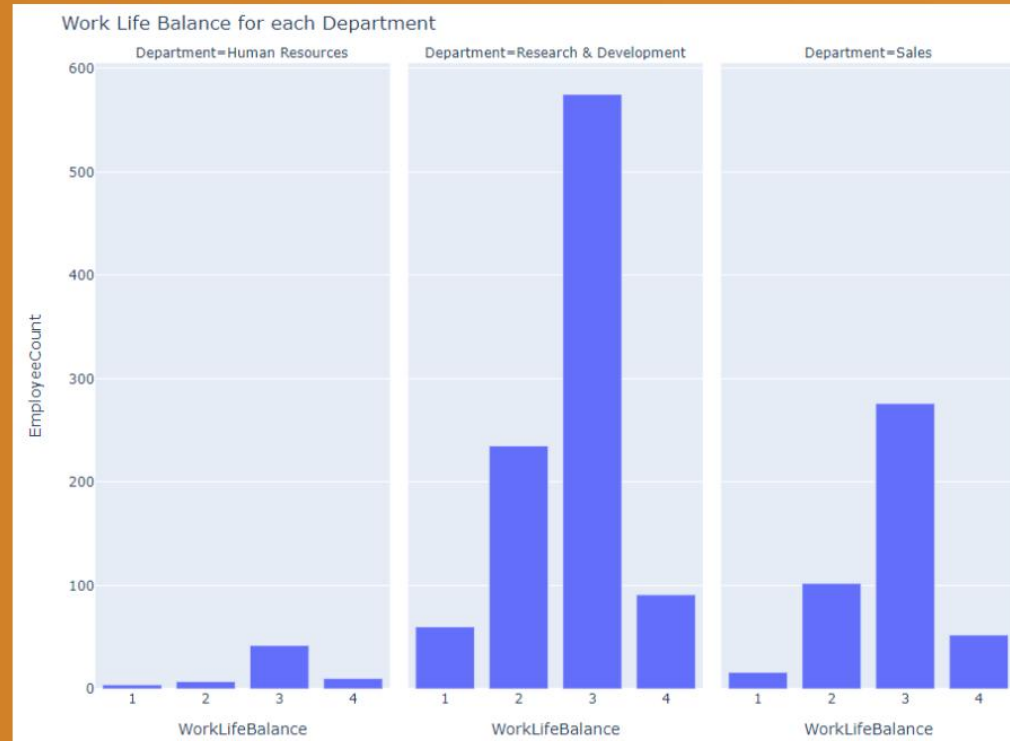
### Question :

Is there Work Life Balance in these  
department?  
Is Work Life Balance a factor of Employee  
Attrition?

**Distribution**

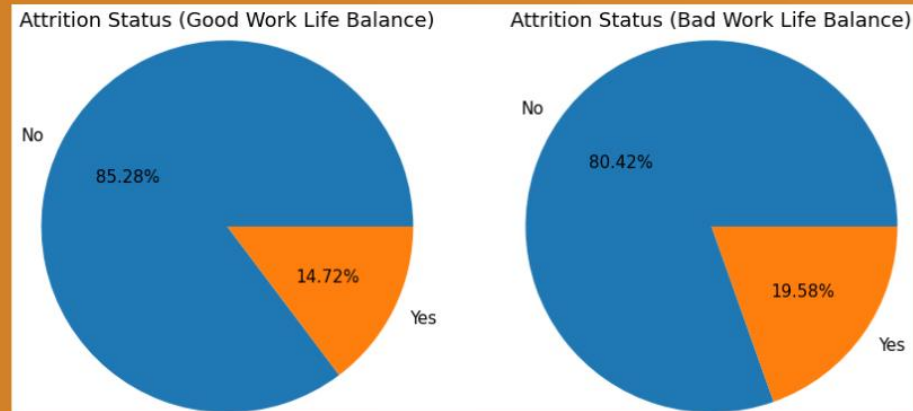
**Attrition**

# Distribution of WorkLifeBalance by Department



Work Life Balance does exist in these Department. However, there are still many of them who stated that they didn't have Work Life Balance, it might be a factor of attrition, lets dive deeper into it.

# Attrition



**Summary :** 19.58% of those who have bad work life balance and 14.72% of those who have good work life balance leaves the company. Therefore, this shouldn't be a major factor of employee attrition as both of the rate of attrition are quite similar.





## Attrition by Salary

### Question :

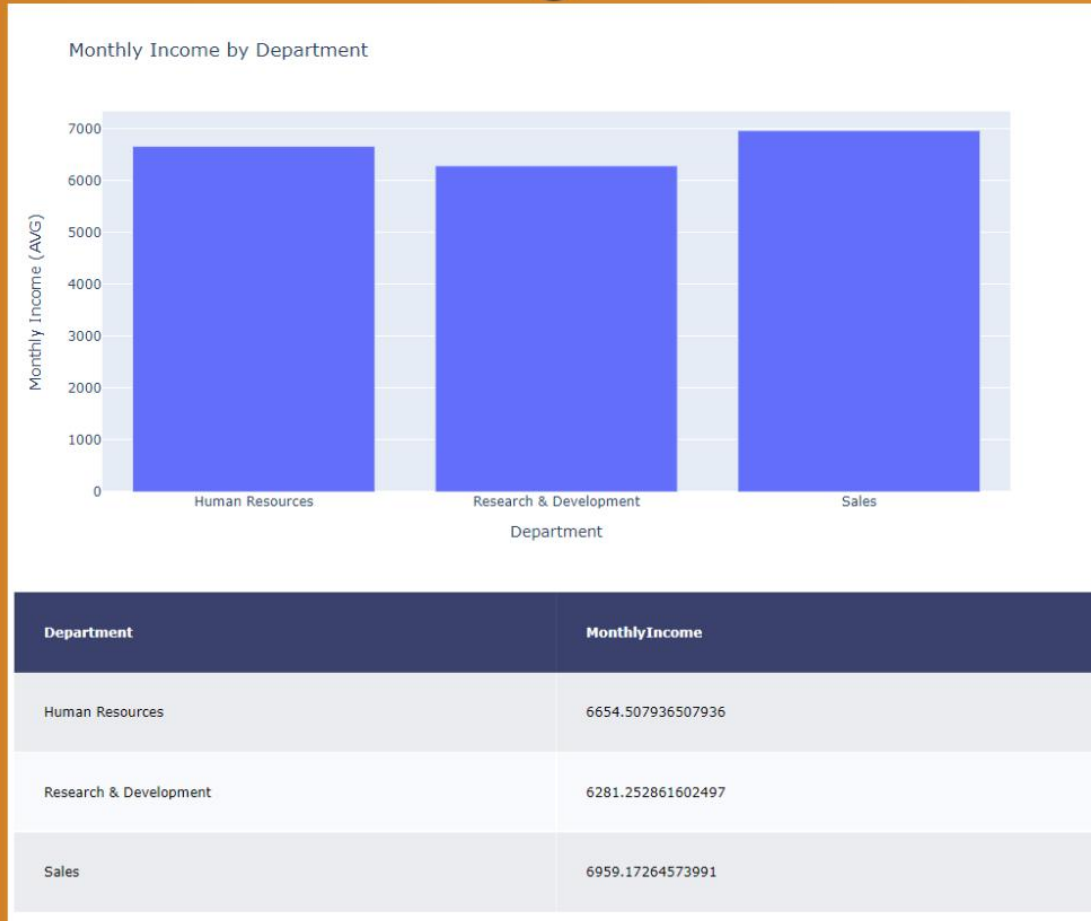
How does the salary varies from different department? Is salary a major factor of employee attrition?

**Distribution**

**Attrition**

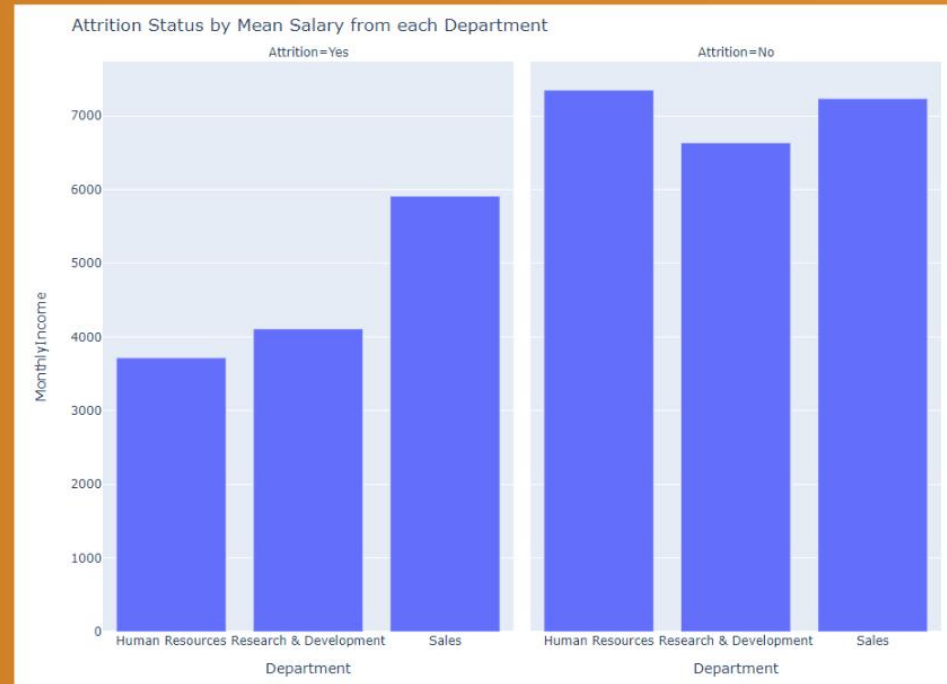


# Salary Distribution by Department



Sales Department have the highest mean monthly salary followed by Human Resource, lastly Research and Development

# Attrition



**Summary :** As we can see, for those who leave the organization, they have a significant lower mean salary compared to those who stay. Therefore, we could clearly see that this is a major factor of Employee Attrition.



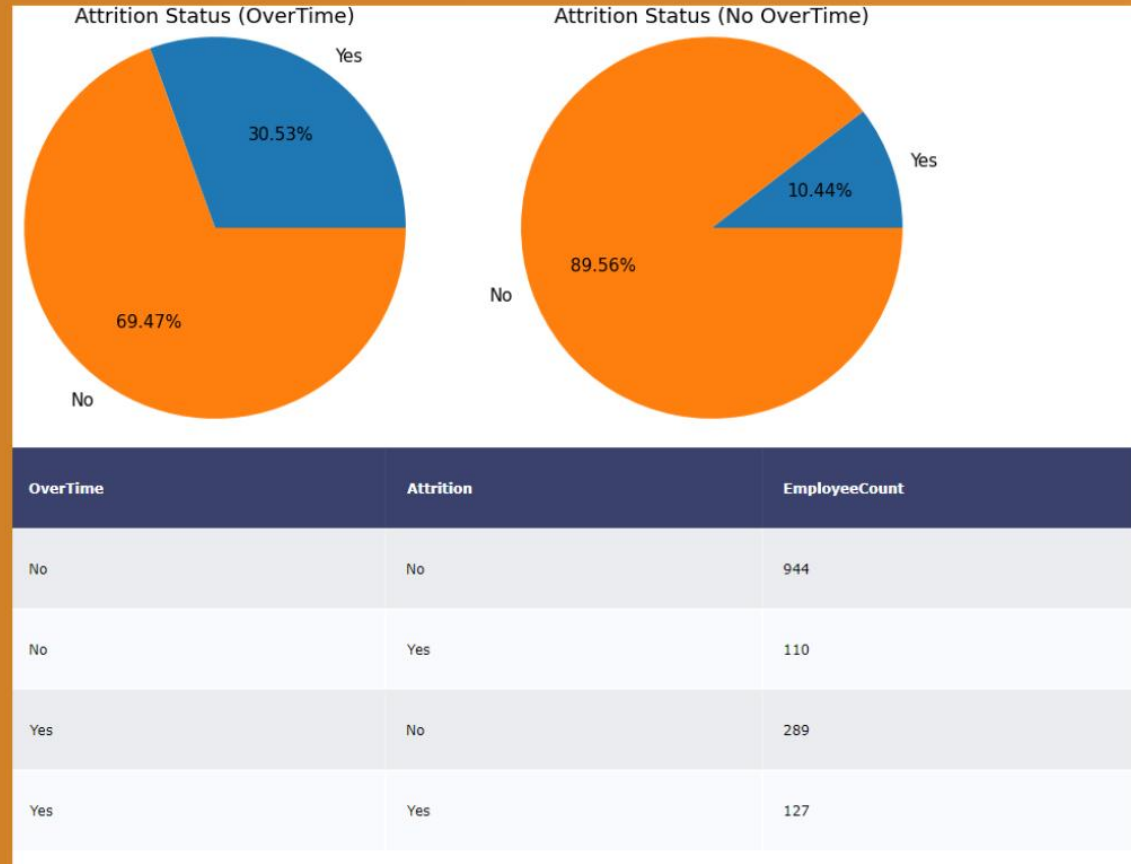
## Overtime

### Question :

Will overtime status affect the rate of attrition?

## Attrition

# Attrition Rate



**Summary :** Over 30% of workers who left the organization worked overtime compared to those who didn't worked overtime! Will this be a reason why employees are leaving?



## Random Forest Classifier

### Aim :

- Get importance of each features
- Build a model to predict employee Attrition

**Features & labels**

**Importance**

**Model**

## Features and label

# Data Cleaning

## Drop

**Drop unwanted features**

```
df = df.drop(['unlabeled_data', 'unlabeled_data', 'unlabeled_data', 'unlabeled_data', 'unlabeled_data'])
```

df.info()

	age	sex	education	experience	department	department_recode	education_recode	education_recode2	gender	hourcode
0	35	M	Post-Graduate	270	Team 1	...	...	...	2	Person
1	30	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
2	30	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
3	35	M	Post-Graduate	270	Research & Development	...	...	...	4	Person
4	35	M	Post-Graduate	270	Research & Development	...	...	...	1	Person
5	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
6	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
7	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
8	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
9	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
10	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
11	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
12	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
13	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
14	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
15	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
16	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
17	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
18	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
19	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
20	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
21	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
22	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
23	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
24	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
25	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
26	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
27	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
28	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
29	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
30	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
31	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
32	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
33	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
34	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
35	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
36	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
37	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
38	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
39	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
40	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
41	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
42	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
43	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
44	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
45	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
46	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
47	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
48	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person
49	35	M	Post-Graduate	270	Research & Development	...	...	...	3	Person

5000 rows x 11 columns

## Get

```

# Get categorical features
cat_features = [col for col in df.columns if df[col].dtype == 'object']

# Get numerical features
num_features = [col for col in df.columns if df[col].dtype == 'float64' or df[col].dtype == 'int64']

# Print the number of categorical and numerical features
print('Number of categorical features: %d' % len(cat_features))
print('Number of numerical features: %d' % len(num_features))

# Print the number of rows and columns
print('Number of rows: %d' % df.shape[0])
print('Number of columns: %d' % df.shape[1])

# Print the first 5 rows of the dataset
df.head()

```

## Convert

```

# Convert Categorical Features into numerical
In [26]: df['smoker'] = df['smoker'].replace({'no': 'non-sm',
df['charter'] = df['charter'].replace({'yes': 'charter',
df['class'] = df['class'].replace({'business', 'premium', 'first', 'economy', 'business_extra',
df['cabin'] = df['cabin'].replace({'nan': 'no_cabin', 'other_cabin': 'other_cabin'})
df['island'] = df['island'].replace({'nan': 'no_island'})

In [27]: df.groupby('PassengerId').head(5)
Out[27]:
   SibSp   Survived   PassengerId  CabinNumber   Gender   Age     Name      Sex
0      1         0           1         54.00000    female   26.0  Mrs. O. S. W.
1      0         1           2         59.69633    female   27.0  Mrs. O. S. W.
2      1         1           3         58.59226    female   22.0  Mrs. O. S. W.
3      0         0           4         26.33597    female   14.0  Mrs. O. S. W.
4      0         0           5         19.00000    female   17.0  Mrs. O. S. W.

Name: 0, dtype: object

In [28]: df.groupby('PassengerId').head(5)
Out[28]:
   SibSp   Survived   PassengerId  CabinNumber   Gender   Age     Name      Sex
0      1         0           1         54.00000    female   26.0  Mrs. O. S. W.
1      0         1           2         59.69633    female   27.0  Mrs. O. S. W.
2      1         1           3         58.59226    female   22.0  Mrs. O. S. W.
3      0         0           4         26.33597    female   14.0  Mrs. O. S. W.
4      0         0           5         19.00000    female   17.0  Mrs. O. S. W.

Name: 0, dtype: object

```

# Final

```
In [67]: # Selecting features and label
x = df.drop(['Attrition'], axis = 1)
y = df['Attrition']
```



# Importance



# Model

## Split

```
In [71]: # Splitting data
x_train,x_test,y_train,y_test = train_test_split(x_scaled,y,test_size=0.1,random_state=42)
print("training shape: ",x_train.shape)
print("testing shape: ",x_test.shape)

training shape: (1829, 39)
testing shape: (201, 39)
```

## Fit

```
In [75]: # Fit model and accuracy score

clf = RandomForestClassifier()

clf.fit(x_train, y_train)

Out[75]: RandomForestClassifier()
```

## Predict

```
In [76]: pred = clf.predict(x_test)

print(clf.__class__.__name__, " : ", accuracy_score(y_test,pred))

RandomForestClassifier : 0.850498038109023
```

## Conclusion

### Top Reasons why Employees leave the Organization:

**Overtime :** Those who worked overtime tend to leave the organization as they are exhausted in their job.

**Monthly Income :** As expected, Income is a huge factor as why employees leave the organization in search for a better salary.

**Age :** This could also be expected, since people who are aiming to retire will leave the organization.

Knowing the most likely reasons why employees leave the organization, can help the organization take action and reduce the level of Attrition inside the organization.