

# HR DATA

```
In [9]: import numpy as np
import pandas as pd

import seaborn as sns
import matplotlib.pyplot as plt

import warnings
warnings.filterwarnings('ignore')

In [2]: pd.options.display.max_rows = 4000
pd.options.display.max_columns = 1000
```

## Read file

```
In [4]: data = pd.read_csv('HR_data.csv')
data.head()
```

Out[4]:

	Attrition	Business Travel	CF_age band	CF_attrition label	Department	Education Field	emp no	Employee Number	Gender	Job Role	Marital Status	Over Time	Over18	Training Times Last Year	Age	CF_cur Empl
0	Yes	Travel_Rarely	35 - 44	Ex-Employees	Sales	Life Sciences	STAFF-1	1	Female	Sales Executive	Single	Yes	Y	0	41	
1	No	Travel_Frequently	45 - 54	Current Employees	R&D	Life Sciences	STAFF-2	2	Male	Research Scientist	Married	No	Y	3	49	
2	Yes	Travel_Rarely	35 - 44	Ex-Employees	R&D	Other	STAFF-4	4	Male	Laboratory Technician	Single	Yes	Y	3	37	
3	No	Travel_Frequently	25 - 34	Current Employees	R&D	Life Sciences	STAFF-5	5	Female	Research Scientist	Married	Yes	Y	3	33	
4	No	Travel_Rarely	25 - 34	Current Employees	R&D	Medical	STAFF-7	7	Male	Laboratory Technician	Married	No	Y	3	27	

## Inspecting Data

```
In [26]: print(f'Number of Rows - {data.shape[0]}')
print(f'Number of Columns - {data.shape[1]}')

Number of Rows - 1470
Number of Columns - 39

In [5]: data.describe()
```

Out[5]:

	Hourly Rate	Job Involvement	Job Level	Job Satisfaction	Monthly Income	Monthly Rate	Num Companies Worked	Percent Salary Hike	Performance Rating	Relationship Satisfaction	Standard Hours	Stock Op L
1470.000000	1470.000000	1470.000000	1470.000000	1470.000000	1470.000000	1470.000000	1470.000000	1470.000000	1470.000000	1470.000000	1470.0	1470.000
65.891156	2.729932	2.063946	2.728571	6502.931293	14313.103401	2.693197	15.209524	3.153741	2.712245	80.0	0.793	
20.329428	0.711561	1.106940	1.102846	4707.956783	7117.786044	2.498009	3.659938	0.360824	1.081209	0.0	0.852	
30.000000	1.000000	1.000000	1.000000	1009.000000	2094.000000	0.000000	11.000000	3.000000	1.000000	80.0	0.000	
48.000000	2.000000	1.000000	2.000000	2911.000000	8047.000000	1.000000	12.000000	3.000000	2.000000	80.0	0.000	
66.000000	3.000000	2.000000	3.000000	4919.000000	14235.500000	2.000000	14.000000	3.000000	3.000000	80.0	1.000	
83.750000	3.000000	3.000000	4.000000	8379.000000	20461.500000	4.000000	18.000000	3.000000	4.000000	80.0	1.000	
100.000000	4.000000	5.000000	4.000000	19999.000000	26999.000000	9.000000	25.000000	4.000000	4.000000	80.0	3.000	

Min age of employee is 18years || Max age of employee is 60years

Max travelling distance of an employee from home to office is 29KM

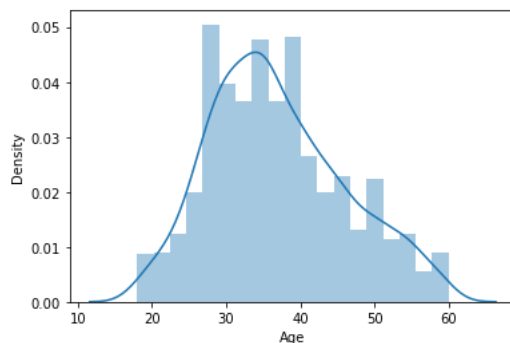
Min salary hike is 11% || Max salary hike is 25%

```
In [6]: data.isnull().sum().sort_values(ascending=False)
```

```
Out[6]: Attrition                0
Performance Rating              0
Job Involvement                0
Job Level                     0
Job Satisfaction               0
Monthly Income                 0
Monthly Rate                   0
Num Companies Worked            0
Percent Salary Hike            0
Relationship Satisfaction       0
Environment Satisfaction       0
Standard Hours                 0
Stock Option Level             0
Total Working Years            0
Work Life Balance              0
Years At Company               0
Years In Current Role          0
Years Since Last Promotion     0
Hourly Rate                    0
~ .. ~ .. ~ .. ~ .. ~ .. ~ ..
```

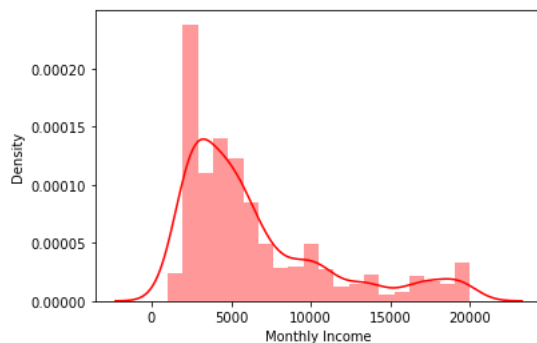
no null value is present

```
In [10]: sns.distplot(data['Age'])
plt.show()
```



```
In [14]: sns.distplot(data['Monthly Income'], color='red')
plt.show()

print(f"Monthly Income is Left skewed")
```



Monthly Income is Left skewed

```
In [20]: data.columns
```

```
Out[20]: Index(['Attrition', 'Business Travel', 'CF_age band', 'CF_attrition label',
'Department', 'Education Field', 'emp no', 'Employee Number', 'Gender',
'Job Role', 'Marital Status', 'Over Time', 'Over18',
'Training Times Last Year', 'Age', 'CF_current Employee', 'Daily Rate',
'Distance From Home', 'Education', 'Employee Count',
'Environment Satisfaction', 'Hourly Rate', 'Job Involvement',
'Job Level', 'Job Satisfaction', 'Monthly Income', 'Monthly Rate',
'Num Companies Worked', 'Percent Salary Hike', 'Performance Rating',
'Relationship Satisfaction', 'Standard Hours', 'Stock Option Level',
'Total Working Years', 'Work Life Balance', 'Years At Company',
'Years In Current Role', 'Years Since Last Promotion',
'Years With Curr Manager'],
dtype='object')
```

```
In [22]: sns.distplot(data['Percent Salary Hike'], color='red')
plt.show()

print(f"Percent Salary Hike is Left skewed")
```



Percent Salary Hike is Left skewed

```
In [ ]:
```