

```
In [2]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [7]: employee_df=pd.read_csv('Employee_Salary_Dataset.csv')
employee_df.head()
```

```
Out[7]:
```

	ID	Experience_Years	Age	Gender	Salary
0	1	5	28	Female	250000
1	2	1	21	Male	50000
2	3	3	23	Female	170000
3	4	2	22	Male	25000
4	5	1	17	Male	10000

```
In [9]: department_df=pd.read_csv('Department_Dataset.csv')
department_df.head()
```

```
Out[9]:
```

	ID	Dept_name	location	travel_required
0	1	HR	Pune	yes
1	2	Finance	Bangalore	no
2	3	Finance	Bangalore	no
3	4	Finance	Pune	no
4	5	Tech	Mumbai	no

```
In [12]: employee_df.shape
```

```
Out[12]: (35, 5)
```

```
In [15]: merge_df = pd.merge(employee_df,department_df,on="ID")
```

```
In [17]: merge_df.head()
```

```
Out[17]:
```

	ID	Experience_Years	Age	Gender	Salary	Dept_name	location	travel_required
0	1	5	28	Female	250000	HR	Pune	yes
1	2	1	21	Male	50000	Finance	Bangalore	no
2	3	3	23	Female	170000	Finance	Bangalore	no
3	4	2	22	Male	25000	Finance	Pune	no
4	5	1	17	Male	10000	Tech	Mumbai	no

```
In [20]: #Average Salary of all the employees
```

```
salary_avg = merge_df['Salary'].mean()
print("average salary of all employees is :",salary_avg)
```

average salary of all employees is : 2059147.142857143

```
In [23]: #Department having highest number of employees
```

```
highest_employee = merge_df['Dept_name'].value_counts().idxmax()
print(highest_employee," department has highest number of employees")
```

Tech department has highest number of employees

```
In [26]: #Gender distribution of employees
```

```
m = merge_df['Gender'].value_counts()['Male']
f = merge_df['Gender'].value_counts()['Female']
print("Number of Male Employees are ",m)
print("Number of Female Employees are ",f)
```

Number of Male Employees are 17
Number of Female Employees are 18

```
In [28]: #Average Salary by department
```

```
merge_df['Dept_name'].value_counts()
```

```
Out[28]:
```

Tech	14
HR	12
Finance	9

Name: Dept_name, dtype: int64

```
In [33]: Tech_avg = 0
```

```

HR_avg = 0
Finance_avg = 0
for i in merge_df.index:
    if merge_df['Dept_name'][i] == "Tech":
        Tech_avg += merge_df['Salary'][i]
    elif merge_df['Dept_name'][i] == "HR":
        HR_avg += merge_df['Salary'][i]
    else:
        Finance_avg += merge_df['Salary'][i]
Tech_avg = Tech_avg/merge_df['Dept_name'].value_counts()['Tech']
HR_avg = HR_avg/merge_df['Dept_name'].value_counts()['HR']
Finance_avg = Finance_avg/merge_df['Dept_name'].value_counts()['Finance']
print("Average Salary of Tech employees are ",Tech_avg)
print("Average Salary of HR employees are ",HR_avg)
print("Average Salary of Finance employees are ",Finance_avg)

```

Average Salary of Tech employees are 2077257.142857143
 Average Salary of HR employees are 1958208.3333333333
 Average Salary of Finance employees are 2165561.1111111111

In [34]: *#Highest Average Salary by department*

```

maxm = max(Tech_avg, HR_avg, Finance_avg)
if maxm == Tech_avg:
    print("\nTech department has highest average salary")
elif maxm == HR_avg:
    print("\nHR department has highest average salary")
else:
    print("\nFinance department has highest average salary")

```

Finance department has highest average salary

In [36]: *#Correlation between Years of Experience and Salary*

```
print(merge_df['Experience_Years'].corr(merge_df['Salary']))
```

0.6855999775494617

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js