Employee Performance Analysis

DAT7 Capstone Project

Sneha Bhaskar

In [6]:

```
df = pd.read_csv ('Data/Employee_Performance_Data.csv')
df
```

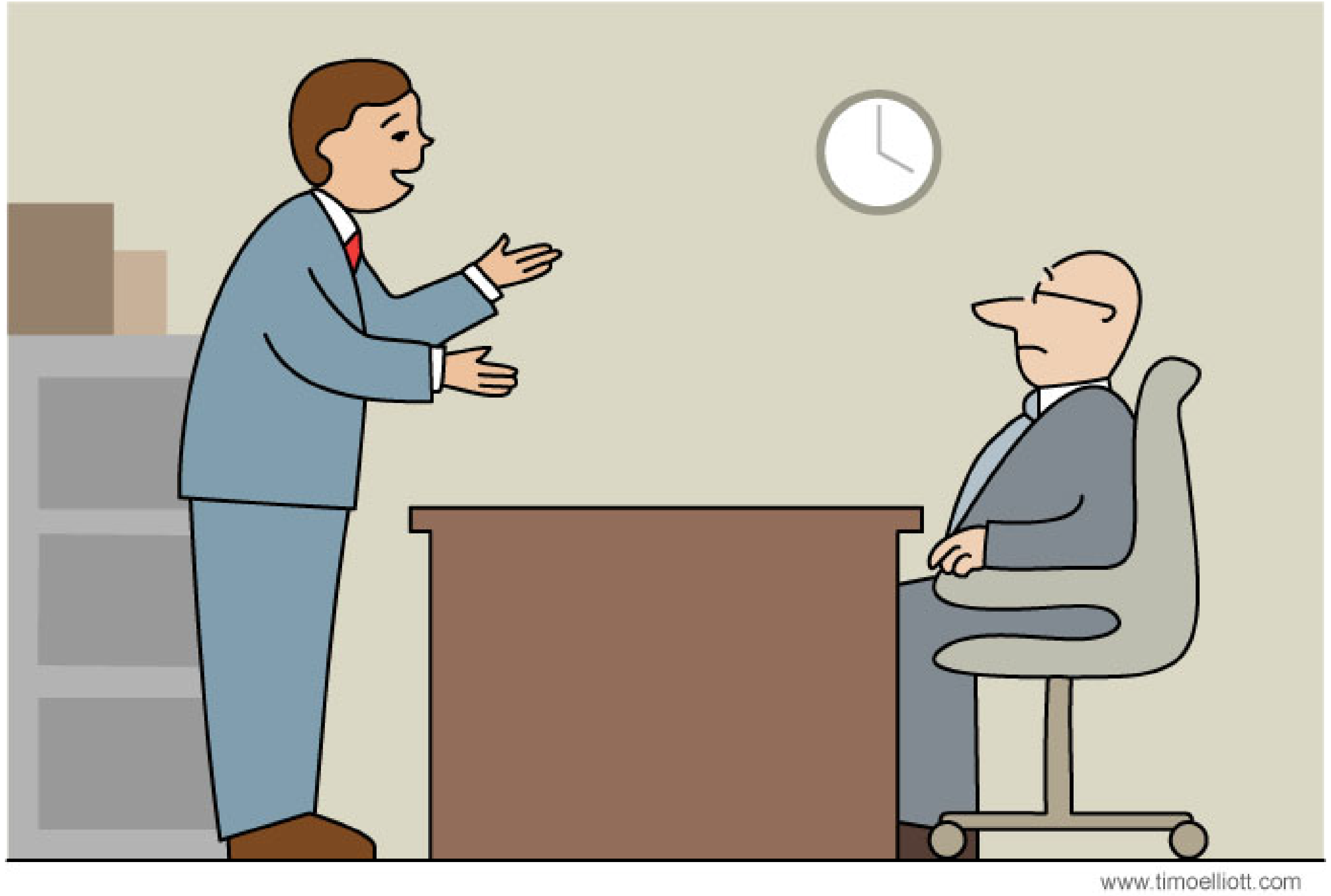
Out[6]:

	EmpNumber	Age	Gender	EducationBackground	MaritalStatus	EmpDepartment	EmpJobRole	BusinessTravelFrequency	Distar
0	E1001000	32	Male	Marketing	Single	Sales	Sales Executive	Travel_Rarely	
1	E1001006	47	Male	Marketing	Single	Sales	Sales Executive	Travel_Rarely	
2	E1001007	40	Male	Life Sciences	Married	Sales	Sales Executive	Travel_Frequently	
3	E1001009	41	Male	Human Resources	Divorced	Human Resources	Manager	Travel_Rarely	
4	E1001010	60	Male	Marketing	Single	Sales	Sales Executive	Travel_Rarely	
...	...	...	...	...	...	...	...	...	
1195	E100992	27	Female	Medical	Divorced	Sales	Sales Executive	Travel_Frequently	
1196	E100993	37	Male	Life Sciences	Single	Development	Senior Developer	Travel_Rarely	
1197	E100994	50	Male	Medical	Married	Development	Senior Developer	Travel_Rarely	
1198	E100995	34	Female	Medical	Single	Data Science	Data	Travel_Rarely	

Background Knowledge About The Dataset

Business & Goal of this project

- Analyze the Employee dataset
- Identify key factors affecting performance ratings
- Accurately predict performance outcomes
- Inform HR decision-making
- Boost employee performance
- Support organizational growth.



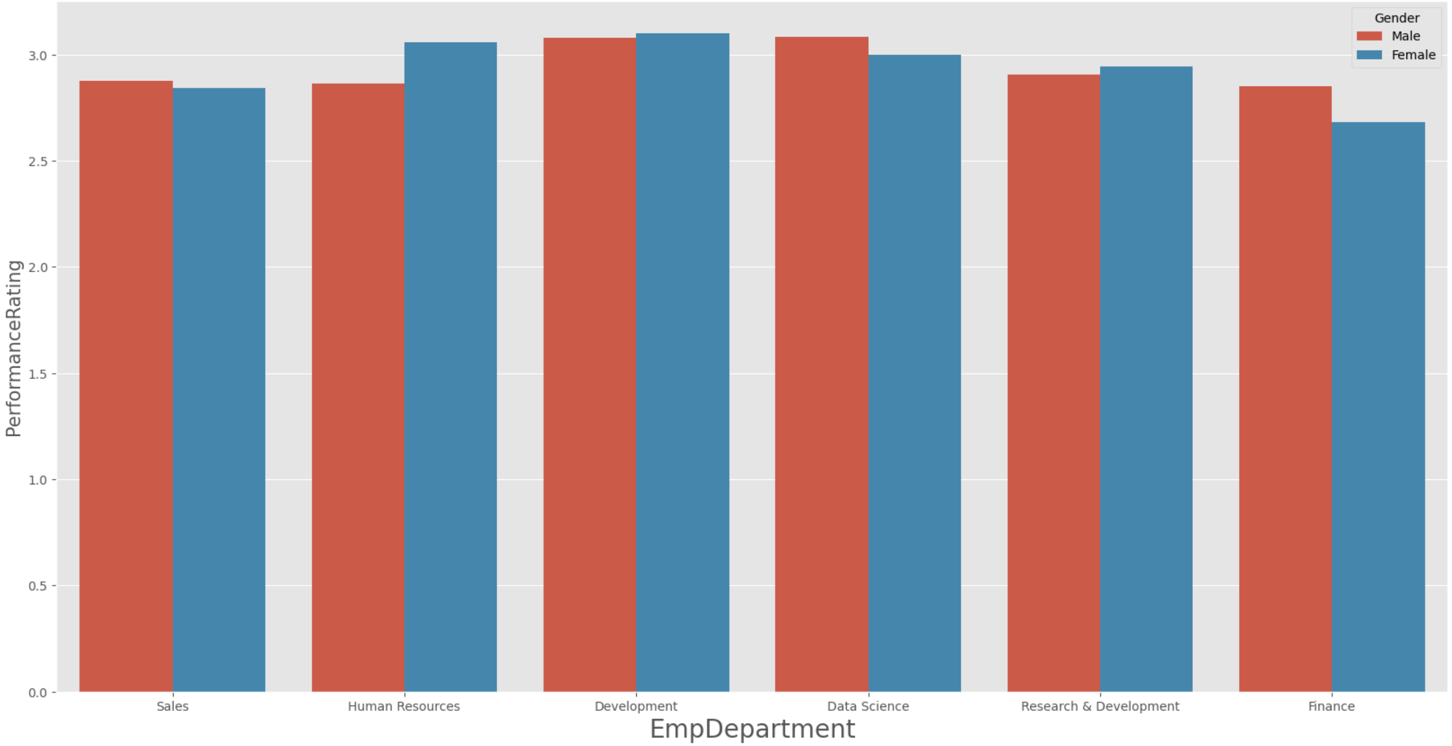
"What do you mean you don't know what I've been up to recently?! Just follow me on Twitter!"

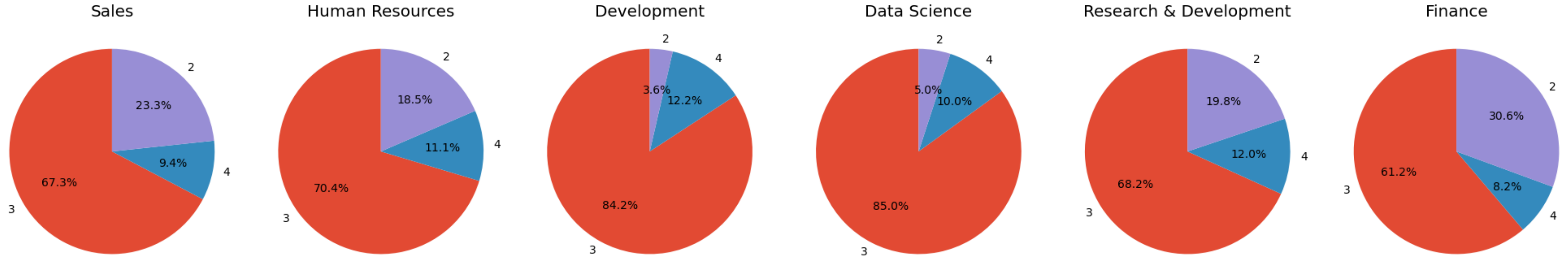
Key Findings

Department-wise Performance Rating

Top 3 Determinants Impacting Employee Performance

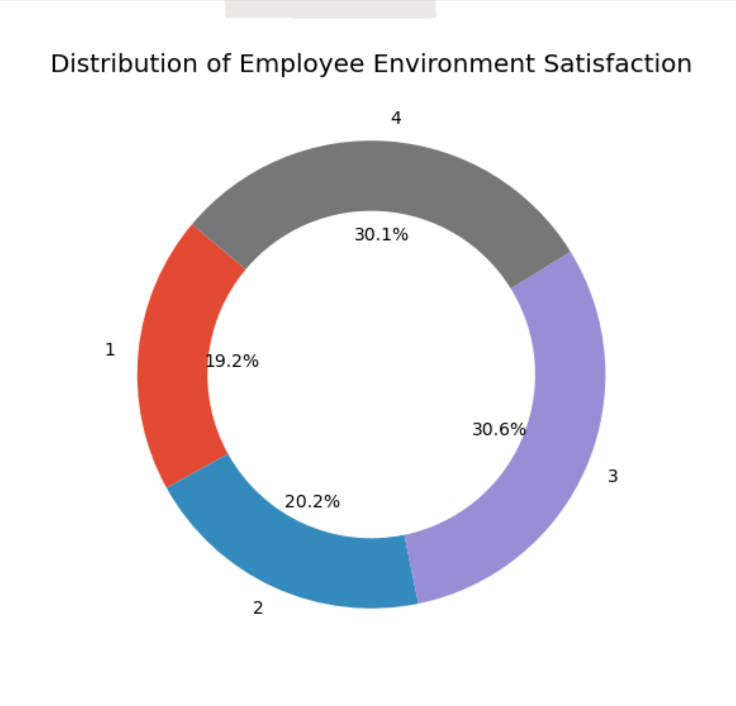
Key Takeaways for HR and Organizations





Department-wise Performance Rating

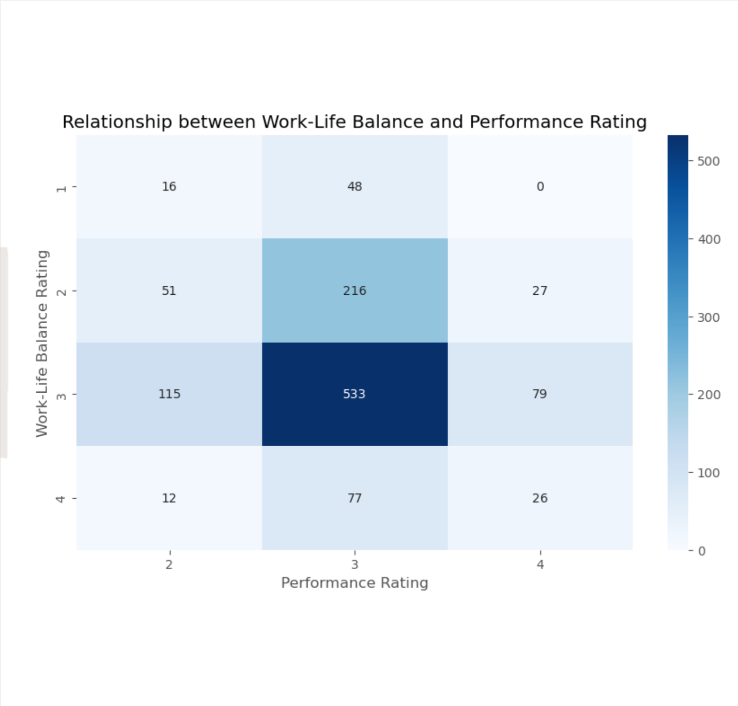
1. Low
2. Good
3. **Excellent** - All the departments have almost 60-80% excellent employee performance rate
4. Outstanding



In [22]:
```
# 2.Employee last salary hike percent
pd.crosstab(df['EmpLastSalaryHikePercent'],df['PerformanceRating'],margins=True)
```

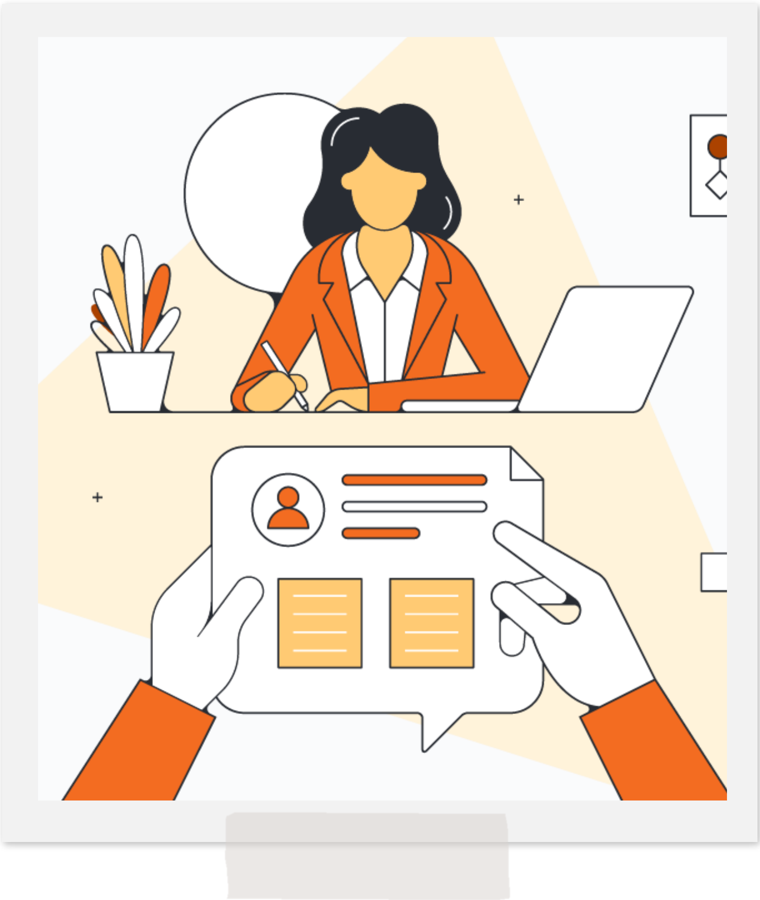
Out[22]:

PerformanceRating	2	3	4	All
EmpLastSalaryHikePercent				
11	28	140	1	169
12	30	123	2	155
13	27	138	3	168
14	28	140	4	172
15	11	67	4	82
16	12	54	2	68
17	9	55	3	67
18	10	63	0	73
19	10	51	2	63
20	9	14	27	50
21	4	6	24	34
22	7	13	27	47
23	4	2	15	21
24	2	5	11	18
25	3	3	7	13
All	194	874	132	1200



Top 3 Determinants Impacting Employee Performance

Key Takeaways for HR and Organizations



- **Focus on Employee Environment Satisfaction**
- **Review Compensation Regularly**
- **Promote Work-Life Balance**
- **Continuous Learning & Development**
- **Feedback Mechanism**
- **Data-Driven Decision Making**
- **Promote Collaboration**
- **Recognition and Reward System**
- **Career Path Clarity**

Regression Analysis Insights

```
Test RMSE:  0.3836557985067259
```

Out[104…] OLS Regression Results

Dep. Variable:	PerformanceRating	**R-squared:**	0.532
Model:	OLS	**Adj. R-squared:**	0.493
Method:	Least Squares	**F-statistic:**	13.68
Date:	Fri, 06 Oct 2023	**Prob (F-statistic):**	3.40e-96
Time:	21:48:07	**Log-Likelihood:**	-351.30
No. Observations:	900	**AIC:**	842.6
Df Residuals:	830	**BIC:**	1179.
Df Model:	69		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	2.0227	0.160	12.629	0.000	1.708	2.337
Age	-0.0014	0.002	-0.793	0.428	-0.005	0.002
DistanceFromHome	-0.0031	0.002	-1.940	0.053	-0.006	3.59e-05
EmpEducationLevel	0.0191	0.013	1.522	0.129	-0.006	0.044
EmpHourlyRate	-0.0015	0.001	-2.472	0.014	-0.003	-0.000
EmpJobInvolvement	0.0059	0.019	0.316	0.752	-0.031	0.042
EmpJobLevel	-0.0002	0.019	-0.008	0.993	-0.037	0.037
EmpJobSatisfaction	-0.0029	0.012	-0.247	0.805	-0.026	0.020
NumCompaniesWorked	-0.0003	0.006	-0.056	0.956	-0.012	0.011
TotalWorkExperienceInYears	0.0010	0.003	0.308	0.758	-0.005	0.007
TrainingTimesLastYear	-0.0089	0.018	-0.482	0.630	-0.045	0.027
ExperienceYearsAtThisCompany	0.0026	0.005	0.546	0.585	-0.007	0.012

In [125]:
```
LR2 = LogisticRegression(C=0.1,penalty='l2',solver='newton-cg')
LR2.fit(x_sm,y_sm)
y_predict_lr2 = LR.predict(x_test)
accuracy_score(y_test,y_predict_lr)
```

Out[125]: 0.8627700127064803

In [126]:
```
#When we use logistic regression, we're trying to put things into categories. A confusion matrix and classification report h
from sklearn.metrics import classification_report,confusion_matrix
print("Confusion Matrix----",confusion_matrix(y_test,y_predict_lr2))
print("Classification Report----",classification_report(y_test,y_predict_lr2))
```

```
Confusion Matrix---- [[247  16  15]
 [ 31 190  29]
 [  4  13 242]]
Classification Report----               precision    recall  f1-score   support

           2       0.88      0.89      0.88       278
           3       0.87      0.76      0.81       250
           4       0.85      0.93      0.89       259

    accuracy                           0.86       787
   macro avg       0.86      0.86      0.86       787
weighted avg       0.86      0.86      0.86       787
```

Thank You

- Email: Sneha.bhaskar1995@gmail.com
- GitHub: @Sdec30
- LinkedIn: https://www.linkedin.com/in/sneha-bhaskar-8aa158186/