



HR Employee Attrition Analysis

Required Library

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
```

Load Data Set

```
In [2]: Hr = pd.read_csv(r"C:\Users\dell\Downloads\HR-Employee-Attrition.csv")
```

```
In [3]: # Read top 5 rows
Hr.head()
```

```
Out[3]:
```

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome
0	41	Yes	Travel_Rarely	1102	Sales	1
1	49	No	Travel_Frequently	279	Research & Development	8
2	37	Yes	Travel_Rarely	1373	Research & Development	2
3	33	No	Travel_Frequently	1392	Research & Development	3
4	27	No	Travel_Rarely	591	Research & Development	2

5 rows × 35 columns

Cleaning & Preparation

```
In [4]: Hr.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 35 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Age                                   1470 non-null   int64
1   Attrition                           1470 non-null   object
2   BusinessTravel                       1470 non-null   object
3   DailyRate                           1470 non-null   int64
4   Department                           1470 non-null   object
5   DistanceFromHome                    1470 non-null   int64
6   Education                           1470 non-null   int64
7   EducationField                       1470 non-null   object
8   EmployeeCount                       1470 non-null   int64
9   EmployeeNumber                      1470 non-null   int64
10  EnvironmentSatisfaction              1470 non-null   int64
11  Gender                               1470 non-null   object
12  HourlyRate                           1470 non-null   int64
13  JobInvolvement                      1470 non-null   int64
14  JobLevel                            1470 non-null   int64
15  JobRole                              1470 non-null   object
16  JobSatisfaction                     1470 non-null   int64
17  MaritalStatus                       1470 non-null   object
18  MonthlyIncome                       1470 non-null   int64
19  MonthlyRate                         1470 non-null   int64
20  NumCompaniesWorked                  1470 non-null   int64
21  Over18                              1470 non-null   object
22  OverTime                            1470 non-null   object
23  PercentSalaryHike                   1470 non-null   int64
24  PerformanceRating                   1470 non-null   int64
25  RelationshipSatisfaction             1470 non-null   int64
26  StandardHours                       1470 non-null   int64
27  StockOptionLevel                    1470 non-null   int64
28  TotalWorkingYears                   1470 non-null   int64
29  TrainingTimesLastYear               1470 non-null   int64
30  WorkLifeBalance                     1470 non-null   int64
31  YearsAtCompany                      1470 non-null   int64
32  YearsInCurrentRole                  1470 non-null   int64
33  YearsSinceLastPromotion              1470 non-null   int64
34  YearsWithCurrManager                 1470 non-null   int64
dtypes: int64(26), object(9)
memory usage: 402.1+ KB

```

```

In [5]: # check duplicates
        Hr.duplicated().sum()

```

```

Out[5]: np.int64(0)

```

```

In [6]: # Check missing value
        Hr.isnull().sum().sum()

```

```

Out[6]: np.int64(0)

```

```
In [7]: for col in Hr.columns:  
        print(Hr[col].unique())  
        print("--"*50)
```

[41 49 37 33 27 32 59 30 38 36 35 29 31 34 28 22 53 24 21 42 44 46 39 43
50 26 48 55 45 56 23 51 40 54 58 20 25 19 57 52 47 18 60]

['Yes' 'No']

['Travel_Rarely' 'Travel_Frequently' 'Non-Travel']

1102	279	1373	1392	591	1005	1324	1358	216	1299	809	153	670	1346
103	1389	334	1123	1219	371	673	1218	419	391	699	1282	1125	691
477	705	924	1459	125	895	813	1273	869	890	852	1141	464	1240
1357	994	721	1360	1065	408	1211	1229	626	1434	1488	1097	1443	515
853	1142	655	1115	427	653	989	1435	1223	836	1195	1339	664	318
1225	1328	1082	548	132	746	776	193	397	945	1214	111	573	1153
1400	541	432	288	669	530	632	1334	638	1093	1217	1353	120	682
489	807	827	871	665	1040	1420	240	1280	534	1456	658	142	1127
1031	1189	1354	1467	922	394	1312	750	441	684	249	841	147	528
594	470	957	542	802	1355	1150	1329	959	1033	1316	364	438	689
201	1427	857	933	1181	1395	662	1436	194	967	1496	1169	1145	630
303	1256	440	1450	1452	465	702	1157	602	1480	1268	713	134	526
1380	140	629	1356	328	1084	931	692	1069	313	894	556	1344	290
138	926	1261	472	1002	878	905	1180	121	1136	635	1151	644	1045
829	1242	1469	896	992	1052	1147	1396	663	119	979	319	1413	944
1323	532	818	854	1034	771	1401	1431	976	1411	1300	252	1327	832
1017	1199	504	505	916	1247	685	269	1416	833	307	1311	128	488
529	1210	1463	675	1385	1403	452	666	1158	228	996	728	1315	322
1479	797	1070	442	496	1372	920	688	1449	1117	636	506	444	950
889	555	230	1232	566	1302	812	1476	218	1132	1105	906	849	390
106	1249	192	553	117	185	1091	723	1220	588	1377	1018	1275	798
672	1162	508	1482	559	210	928	1001	549	1124	738	570	1130	1192
343	144	1296	1309	483	810	544	1062	1319	641	1332	756	845	593
1171	350	921	1144	143	1046	575	156	1283	755	304	1178	329	1362
1371	202	253	164	1107	759	1305	982	821	1381	480	1473	891	1063
645	1490	317	422	1485	1368	1448	296	1398	1349	986	1099	1116	1499
983	1009	1303	1274	1277	587	413	1276	988	1474	163	267	619	302
443	828	561	426	232	1306	1094	509	775	195	258	471	799	956
535	1495	446	1245	703	823	1246	622	1287	448	254	1365	538	525
558	782	362	1236	1112	204	1343	604	1216	646	160	238	1397	306
991	482	1176	913	1076	727	885	243	806	817	1410	1207	1442	693
929	562	608	580	970	1179	294	314	316	654	168	381	217	501
650	141	804	975	1090	346	430	268	167	621	527	883	954	310
719	725	715	657	1146	182	376	571	384	791	1111	1243	1092	1325
805	213	118	676	1252	286	1258	932	1041	859	720	946	1184	436
589	760	887	1318	625	180	586	1012	661	930	342	1230	1271	1278
607	130	300	583	1418	1269	379	395	1265	1222	341	868	1231	102
881	1383	1075	374	1086	781	177	500	1425	1454	617	1085	995	1122
618	546	462	1198	1272	154	1137	1188	188	1333	867	263	938	129
616	498	1404	1053	289	1376	231	152	882	903	1379	335	722	461
974	1126	840	1134	248	955	939	1391	1206	287	1441	109	1066	277
466	1055	265	135	247	1035	266	145	1038	1234	1109	1089	788	124
660	1186	1464	796	415	769	1003	1366	330	1492	1204	309	1330	469
697	1262	1050	770	406	203	1308	984	439	793	1451	1182	174	490

```
718 433 773 603 874 367 199 481 647 1384 902 819 862 1457
977 942 1402 1421 1361 917 200 150 179 696 116 363 107 1465
458 1212 1103 966 1010 326 1098 969 1167 694 1320 536 373 599
251 131 237 1429 648 735 531 429 968 879 640 412 848 360
1138 325 1322 299 1030 634 524 256 1060 935 495 282 206 943
523 507 601 855 1291 1405 1369 999 1202 285 404 736 1498 1200
1439 499 205 683 1462 949 652 332 1475 337 971 1174 667 560
172 383 1255 359 401 377 592 1445 1221 866 981 447 1326 748
990 405 115 790 830 1193 1423 467 271 410 1083 516 224 136
1029 333 1440 674 1342 898 824 492 598 740 888 1288 104 1108
479 1351 474 437 884 1370 264 1059 563 457 1313 241 1015 336
1387 170 208 671 711 737 1470 365 763 567 486 772 301 311
584 880 392 148 708 1259 786 370 678 146 581 918 1238 585
741 552 369 717 543 964 792 611 176 897 600 1054 428 181
211 1079 590 305 953 478 1375 244 511 1294 196 734 1239 1253
1128 1336 234 766 261 1194 431 572 1422 1297 574 355 207 706
280 726 414 352 1224 459 1254 1131 835 1172 1266 783 219 1213
1096 1251 1394 605 1064 1337 937 157 754 1168 155 1444 189 911
1321 1154 557 642 801 161 1382 1037 105 582 704 345 1120 1378
468 613 1023 628]
```

```
-----
-----
['Sales' 'Research & Development' 'Human Resources']
-----
-----
```

```
[ 1  8  2  3 24 23 27 16 15 26 19 21  5 11  9  7  6 10  4 25 12 18 29 22
14 20 28 17 13]
```

```
-----
-----
[2 1 4 3 5]
-----
-----
```

```
['Life Sciences' 'Other' 'Medical' 'Marketing' 'Technical Degree'
'Human Resources']
-----
-----
```

```
[1]
-----
-----
```

```
[  1    2    4 ... 2064 2065 2068]
```

```
-----
-----
[2 3 4 1]
-----
-----
```

```
['Female' 'Male']
-----
-----
```

```
[ 94 61 92 56 40 79 81 67 44 84 49 31 93 50 51 80 96 78
 45 82 53 83 58 72 48 42 41 86 97 75 33 37 73 98 36 47
 71 30 43 99 59 95 57 76 87 66 55 32 52 70 62 64 63 60
100 46 39 77 35 91 54 34 90 65 88 85 89 68 69 74 38]
```

[3 2 4 1]

[2 1 3 4 5]

['Sales Executive' 'Research Scientist' 'Laboratory Technician'
'Manufacturing Director' 'Healthcare Representative' 'Manager'
'Sales Representative' 'Research Director' 'Human Resources']

[4 2 3 1]

['Single' 'Married' 'Divorced']

[5993 5130 2090 ... 9991 5390 4404]

[19479 24907 2396 ... 5174 13243 10228]

[8 1 6 9 0 4 5 2 7 3]

['Y']

['Yes' 'No']

[11 23 15 12 13 20 22 21 17 14 16 18 19 24 25]

[3 4]

[1 4 2 3]

[80]

[0 1 3 2]

[8 10 7 6 12 1 17 5 3 31 13 0 26 24 22 9 19 2 23 14 15 4 29 28
21 25 20 11 16 37 38 30 40 18 36 34 32 33 35 27]

[0 3 2 5 1 4 6]

```
[1 3 2 4]
```

```
[ 6 10  0  8  2  7  1  9  5  4 25  3 12 14 22 15 27 21 17 11 13 37 16 20  
40 24 33 19 36 18 29 31 32 34 26 30 23]
```

```
[ 4  7  0  2  5  9  8  3  6 13  1 15 14 16 11 10 12 18 17]
```

```
[ 0  1  3  2  7  4  8  6  5 15  9 13 12 10 11 14]
```

```
[ 5  7  0  2  6  8  3 11 17  1  4 12  9 10 15 13 16 14]
```

Exploratory Data Analysis And Visualization

```
In [66]: sns.set_theme(style='dark')
```

```
In [9]: Hr.describe().T
```

Out[9]:

	count	mean	std	min	25%	
Age	1470.0	36.923810	9.135373	18.0	30.00	
DailyRate	1470.0	802.485714	403.509100	102.0	465.00	
DistanceFromHome	1470.0	9.192517	8.106864	1.0	2.00	
Education	1470.0	2.912925	1.024165	1.0	2.00	
EmployeeCount	1470.0	1.000000	0.000000	1.0	1.00	
EmployeeNumber	1470.0	1024.865306	602.024335	1.0	491.25	1
EnvironmentSatisfaction	1470.0	2.721769	1.093082	1.0	2.00	
HourlyRate	1470.0	65.891156	20.329428	30.0	48.00	
JobInvolvement	1470.0	2.729932	0.711561	1.0	2.00	
JobLevel	1470.0	2.063946	1.106940	1.0	1.00	
JobSatisfaction	1470.0	2.728571	1.102846	1.0	2.00	
MonthlyIncome	1470.0	6502.931293	4707.956783	1009.0	2911.00	4
MonthlyRate	1470.0	14313.103401	7117.786044	2094.0	8047.00	14
NumCompaniesWorked	1470.0	2.693197	2.498009	0.0	1.00	
PercentSalaryHike	1470.0	15.209524	3.659938	11.0	12.00	
PerformanceRating	1470.0	3.153741	0.360824	3.0	3.00	
RelationshipSatisfaction	1470.0	2.712245	1.081209	1.0	2.00	
StandardHours	1470.0	80.000000	0.000000	80.0	80.00	
StockOptionLevel	1470.0	0.793878	0.852077	0.0	0.00	
TotalWorkingYears	1470.0	11.279592	7.780782	0.0	6.00	
TrainingTimesLastYear	1470.0	2.799320	1.289271	0.0	2.00	
WorkLifeBalance	1470.0	2.761224	0.706476	1.0	2.00	
YearsAtCompany	1470.0	7.008163	6.126525	0.0	3.00	
YearsInCurrentRole	1470.0	4.229252	3.623137	0.0	2.00	
YearsSinceLastPromotion	1470.0	2.187755	3.222430	0.0	0.00	
YearsWithCurrManager	1470.0	4.123129	3.568136	0.0	2.00	

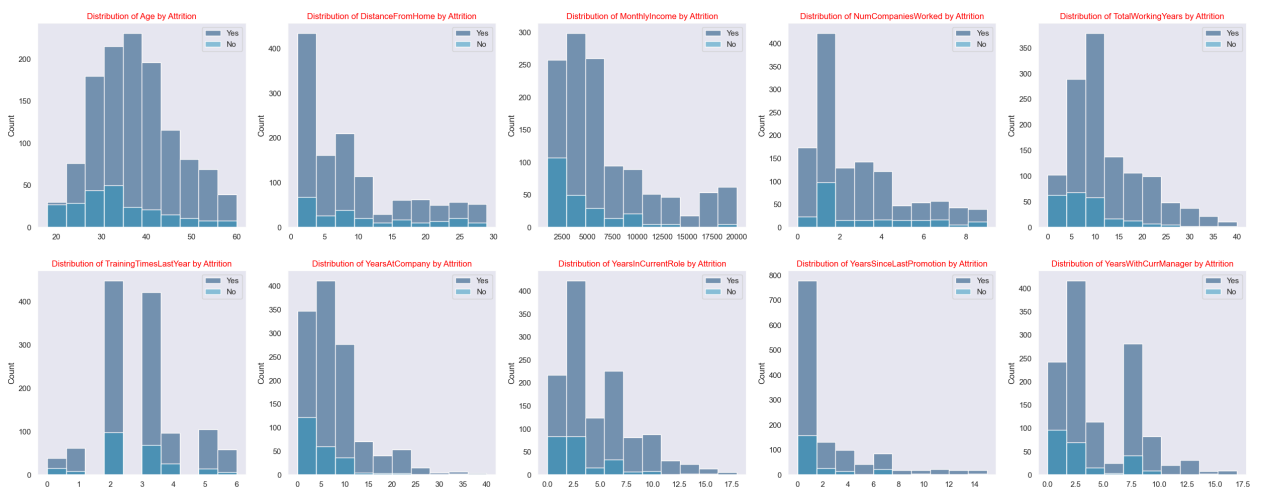
In [64]: Hr.describe(include="object").T

Out[64]:

	count	unique	top	freq
Attrition	1470	2	No	1233
BusinessTravel	1470	3	Travel_Rarely	1043
Department	1470	3	Research & Development	961
EducationField	1470	6	Life Sciences	606
Gender	1470	2	Male	882
JobRole	1470	9	Sales Executive	326
MaritalStatus	1470	3	Married	673
Over18	1470	1	Y	1470
OverTime	1470	2	No	1054

Age Distribution

```
In [67]: Num_Column = ['Age', 'DistanceFromHome', 'MonthlyIncome', 'NumCompaniesWorked',  
                        'TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany',  
                        'YearsInCurrentRole', 'YearsSinceLastPromotion', 'YearsWithCurrManager']  
plt.figure(figsize=(25,10))  
for x, col in enumerate(Num_Column):  
    plt.subplot(2,5,x+1)  
    sns.histplot(x = col, data= Hr,  
                 hue= "Attrition",palette=["#2596be","#063970"],bins=10)  
    plt.title(f"Distribution of {col} by Attrition",color = "Red")  
    plt.xlabel(" ")  
    plt.legend(Hr["Attrition"])  
plt.tight_layout()  
plt.show()
```



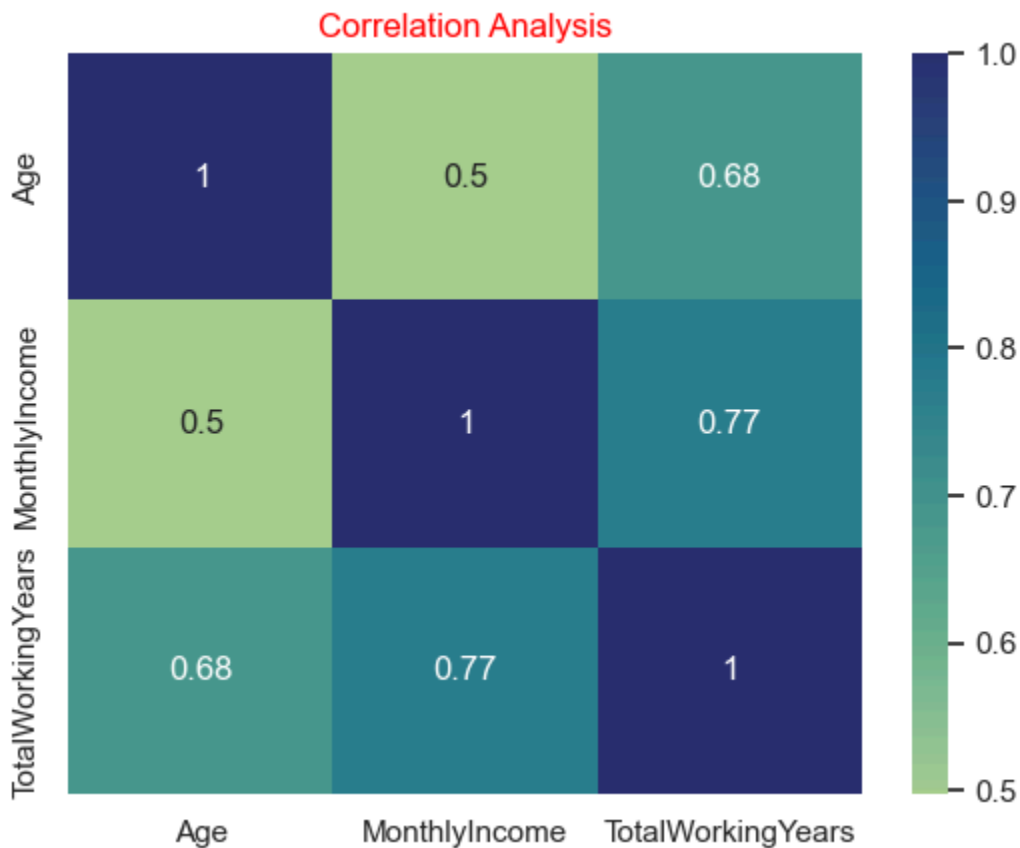
Insights

- **1 Age** Employee attrition is higher among younger employees (approximately 25–35 years). As age increases, attrition steadily declines, indicating that mid-career and senior employees are more stable and less likely to leave.
- **2 Distance From Home** Employees living farther from the workplace show higher attrition compared to those living closer. This suggests that long commute distance contributes to employee dissatisfaction and turnover.
- **3 Monthly Income** Attrition is concentrated in the lower income ranges, while employees earning higher salaries demonstrate stronger retention. This indicates that compensation plays a significant role in employee attrition.
- **4 Number of Companies Worked** Employees who have worked at multiple companies in the past are more likely to leave again. This pattern reflects repeated job-hopping behavior.
- **5 Total Working Years** Attrition is higher among employees with lower overall work experience, whereas highly experienced employees tend to remain with the organization longer, showing greater career stability.
- **6 Training Times Last Year** No clear linear relationship is observed between training frequency and attrition. This suggests that training alone does not guarantee retention, and factors such as role growth and recognition are also important.
- **7 Years at Company** Attrition peaks during the early tenure period (0–5 years). Employees with longer tenure show significantly lower attrition, indicating stronger organizational attachment over time.
- **8 Years in Current Role** Employees with shorter tenure in their current role are more likely to leave. This highlights the importance of role clarity, job satisfaction, and internal mobility.
- **9 Years Since Last Promotion** Employees who have not received promotions for a long time exhibit higher attrition. Delayed career progression appears to be a major driver of employee exits.

- **10 Years With Current Manager** Higher attrition is observed among employees with shorter tenure under their current manager. Stable manager-employee relationships contribute positively to retention.

Correlation Insights

```
In [71]: Corr_Column = ['Age', 'MonthlyIncome', 'TotalWorkingYears']
sns.heatmap(Hr[Corr_Column].corr(), annot=True, cmap="crest")
plt.title("Correlation Analysis", color="Red")
plt.show()
```



Insights

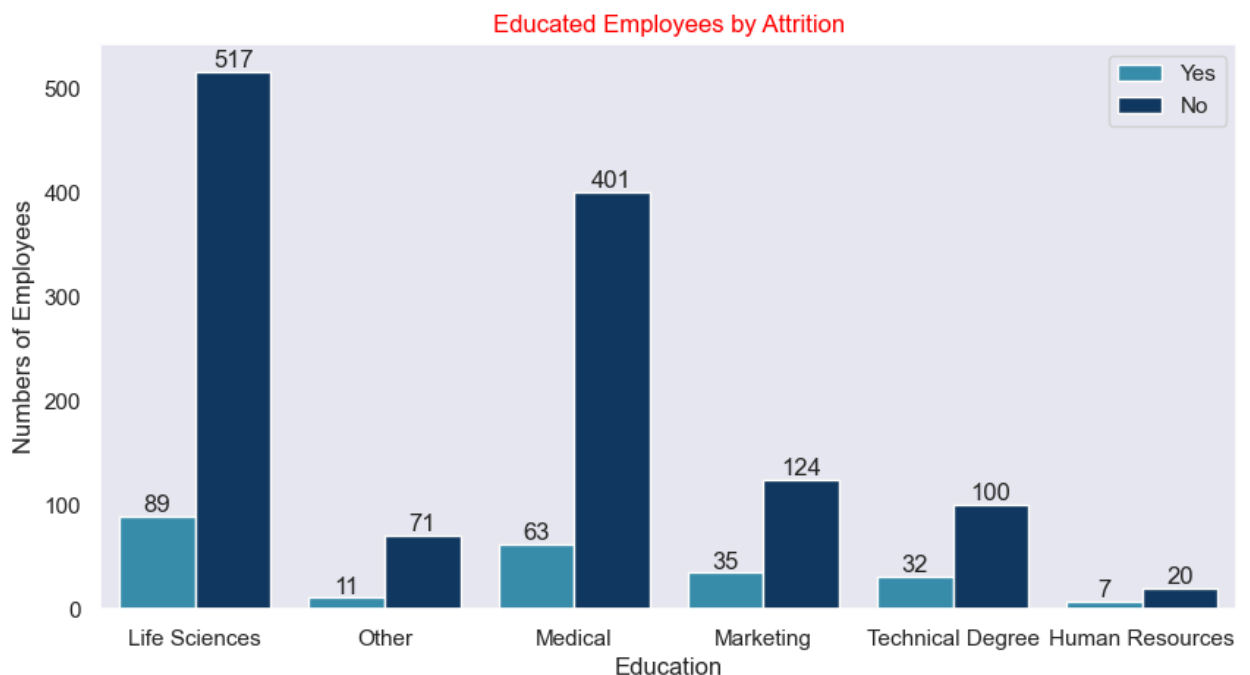
The correlation analysis reveals a moderate to strong positive relationship among age, monthly income, and total working years.

- Age and Monthly Income show a moderate positive correlation (~0.50), indicating that employee income generally increases as age increases.
- Age and Total Working Years have a strong positive correlation (~0.68), reflecting natural career progression as employees gain experience over time.

- Monthly Income and Total Working Years exhibit the strongest correlation (~ 0.77), suggesting that work experience is a key driver of salary growth.

Education Analysis

```
In [12]: plt.figure(figsize=(10,5))
ax= sns.countplot(x = "EducationField", data= Hr
, hue="Attrition", palette=["#2596be", "#063970"])
for bar in ax.containers:
    ax.bar_label(bar)
plt.title("Educated Employees by Attrition", color = "Red")
plt.xlabel("Education")
plt.ylabel("Numbers of Employees")
plt.legend()
plt.show()
```



Insight

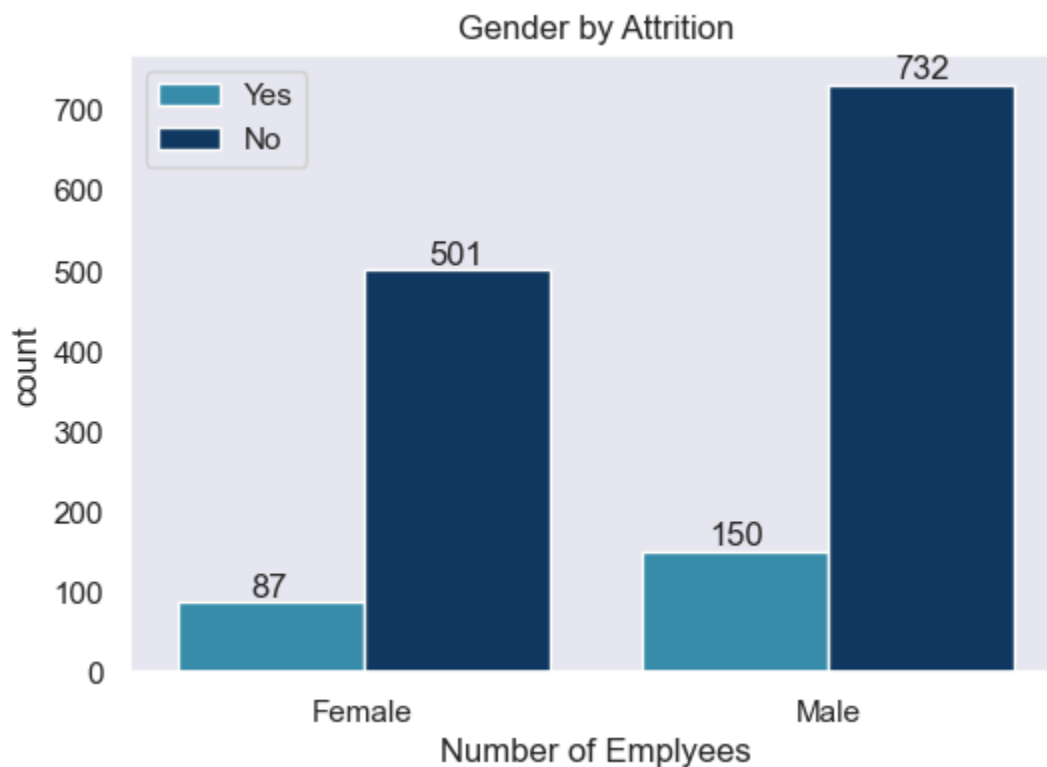
- The analysis shows that employees from Life Sciences and Medical backgrounds constitute the largest portion of the workforce and therefore record the highest number of attrition cases in absolute terms. However, their attrition remains proportionate to their overall employee count, indicating no unusually high attrition risk in these fields.
- Employees with Marketing and Technical Degree backgrounds display a

relatively higher attrition compared to their group size, suggesting greater job mobility or stronger external market opportunities.

- In contrast, employees from Human Resources and 'Other' education fields exhibit lower attrition levels, reflecting stronger retention or limited role-switching opportunities.

Gender Analysis

```
In [13]: plt.figure(figsize=(6,4))
ax = sns.countplot(x = "Gender", data=Hr,
hue="Attrition",palette=["#2596be", "#063970"])
for bar in ax.containers:
    ax.bar_label(bar)
plt.title("Gender by Attrition")
plt.xlabel("Number of Employees")
plt.legend()
plt.show()
```



Insights

- The analysis shows that male employees have a higher number of attrition cases compared to female employees. However, this difference is largely influenced by the fact that the overall male workforce is larger than the female workforce.

- When comparing attrition within each gender group, both male and female employees show similar retention patterns, indicating that gender alone is not a strong driver of attrition in this dataset.

Job Role Analysis

```
In [14]: Job_Role = Hr.groupby(["JobRole", "Attrition"])["MonthlyIncome"].mean().reset_index()
plt.figure(figsize=(19,8))
sns.lineplot(x = "JobRole", y = "MonthlyIncome",
hue="Attrition",data=Job_Role,palette=["#2596be","#063970"])
plt.title("Average Monthly Income by Job Role and Attrition Status", color = "red")
plt.legend()
plt.ylabel("Income")
plt.xlabel("Job")
plt.show()
```

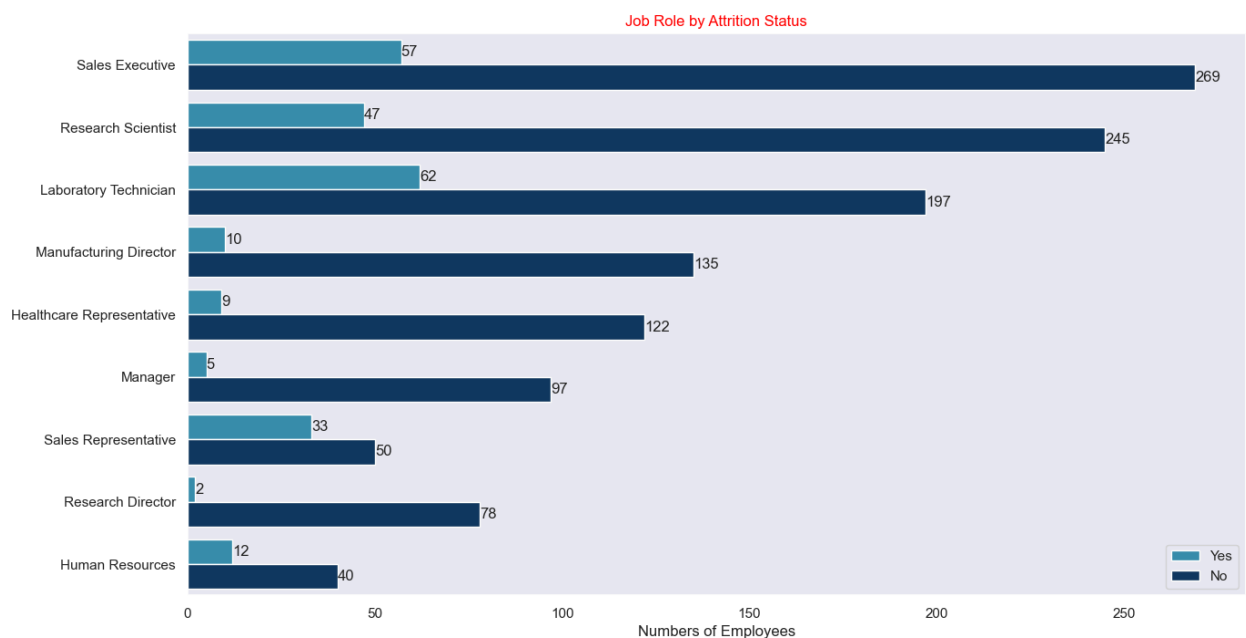


Insights

- The analysis reveals clear differences in average monthly income between employees who left and those who stayed across job roles. In several roles—such as Laboratory Technician, Human Resources, Research Scientist, and Sales Representative—employees who left the organization tend to have lower average income compared to retained employees, suggesting that compensation may be a contributing factor to attrition in these roles.
- For senior and high-paying roles like Manager and Research Director, the income gap between attrition and retention is minimal, and overall attrition is lower, indicating greater job stability and satisfaction at higher compensation levels.

- In roles such as Sales Executive, the income difference between leavers and stayers is relatively small, implying that factors beyond salary (e.g., performance pressure or incentives) may influence attrition.

```
In [50]: plt.figure(figsize=(15,8))
ax = sns.countplot(y = "JobRole",hue="Attrition",
data=Hr,palette=["#2596be","#063970"])
for bar in ax.containers:
    ax.bar_label(bar)
plt.title("Job Role by Attrition Status", color = "Red")
plt.legend()
plt.xlabel("Numbers of Employees")
plt.ylabel("")
plt.show()
```



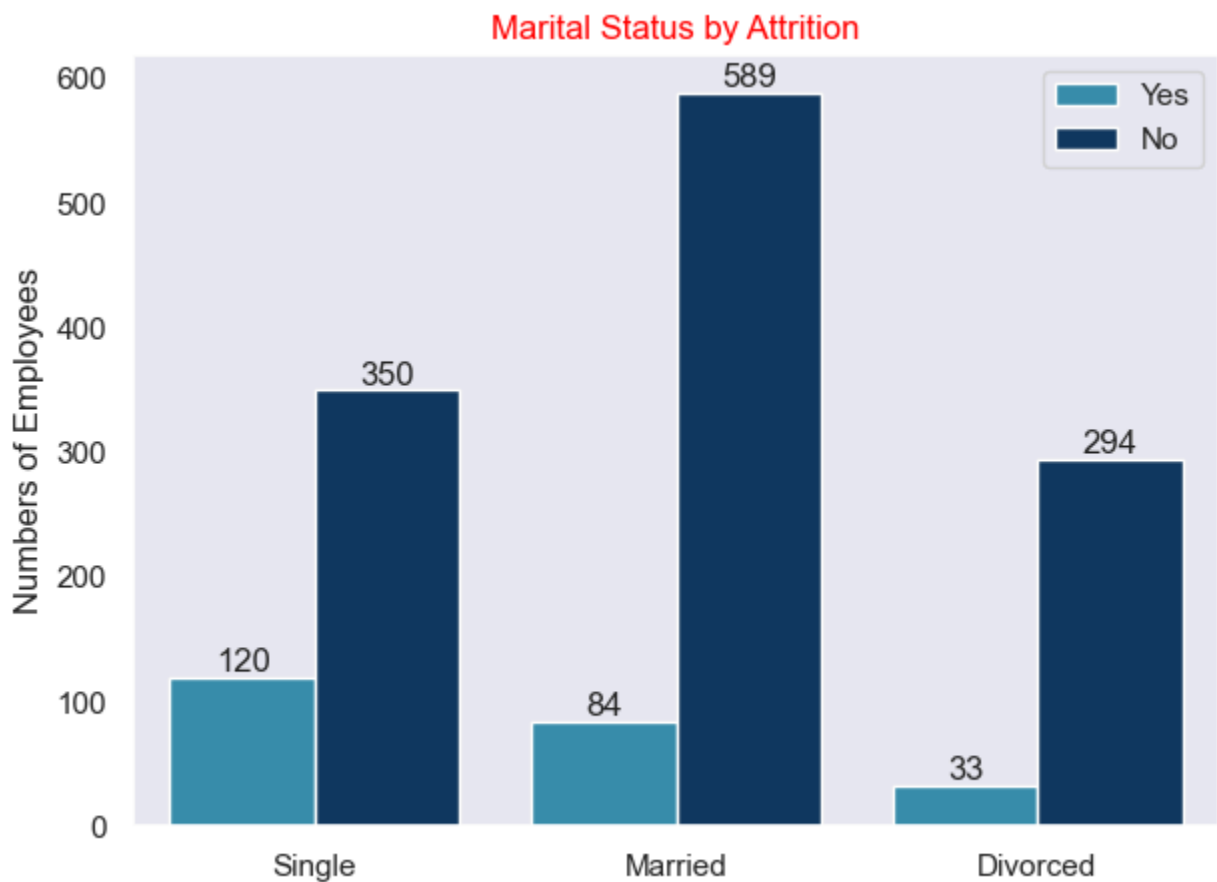
Insights

- The analysis shows that attrition is not evenly distributed across job roles. Roles such as Laboratory Technician, Sales Executive, and Research Scientist record the highest number of employees leaving, primarily because these roles also have a larger workforce size.
- In contrast, senior and specialized roles like Manager, Manufacturing Director, and Research Director exhibit significantly lower attrition counts, indicating stronger job stability, higher satisfaction, or better compensation and growth opportunities.
- Roles such as Human Resources and Sales Representative show moderate attrition, suggesting that while employees do leave, attrition

is relatively controlled compared to operational and sales-heavy roles.

MaritalStatus Analysis

```
In [47]: plt.figure(figsize=(7,5))
ax = sns.countplot(x = 'MaritalStatus',data= Hr,
hue= "Attrition",palette=["#2596be","#063970"])
for bar in ax.containers:
    ax.bar_label(bar)
plt.title("Marital Status by Attrition",color = "Red")
plt.xlabel(" ")
plt.ylabel("Numbers of Employees")
plt.legend()
plt.show()
```



Insights

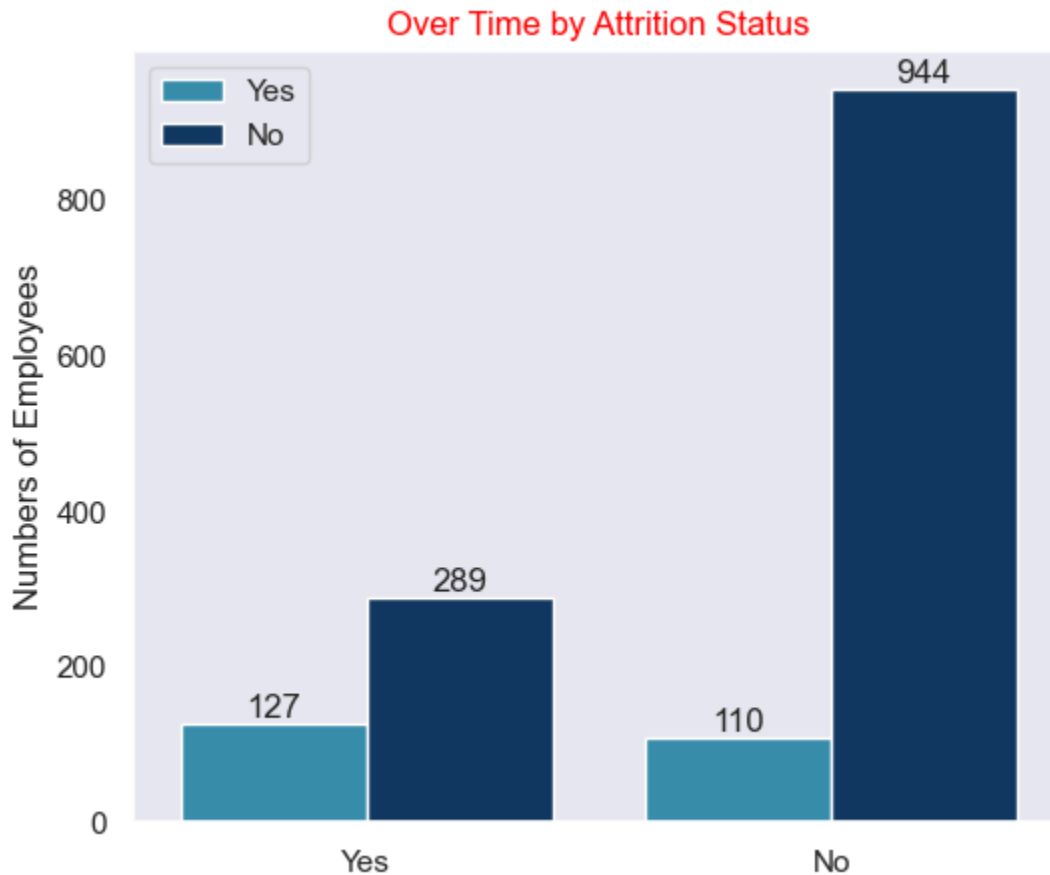
- The analysis indicates that single employees experience the highest attrition compared to married and divorced employees. Despite married employees forming the largest portion of the workforce, their attrition count is relatively lower, suggesting stronger retention and stability

among married employees.

- Divorced employees show the lowest attrition, indicating comparatively higher job continuity within this group.

Over Time Analysis

```
In [48]: plt.figure(figsize=(6,5))
ax = sns.countplot(x = 'OverTime',data= Hr,
hue= "Attrition",palette=["#2596be","#063970"])
plt.title("Over Time by Attrition Status",color = "Red")
for bar in ax.containers:
    ax.bar_label(bar)
plt.xlabel(" ")
plt.ylabel("Numbers of Employees")
plt.legend()
plt.show()
```



Insights

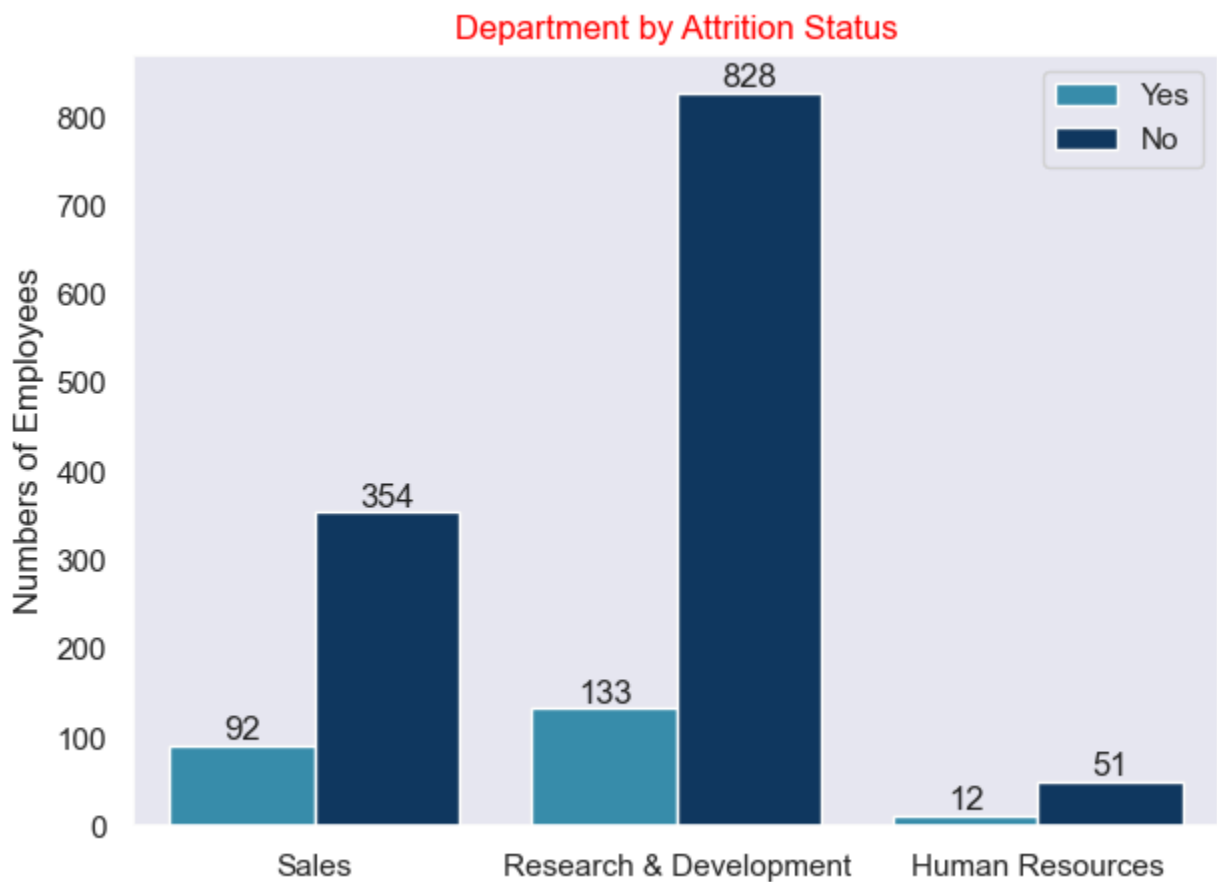
- The analysis clearly shows that employees who work overtime have significantly higher attrition compared to those who do not work

overtime. A noticeably larger proportion of employees who left the organization belong to the overtime = Yes category.

- In contrast, employees who do not work overtime demonstrate much stronger retention, with a substantially higher number remaining in the organization.

Department Analysis

```
In [49]: plt.figure(figsize=(7,5))
ax = sns.countplot(x = 'Department',data= Hr,
hue= "Attrition",palette=["#2596be","#063970"])
plt.title("Department by Attrition Status",color = "Red")
for bar in ax.containers:
    ax.bar_label(bar)
plt.xlabel(" ")
plt.ylabel("Numbers of Employees")
plt.legend()
plt.show()
```



Insights

- The analysis shows that Research & Development (R&D) experiences the highest number of attrition cases, largely due to it being the largest department in terms of workforce size. Sales also shows a notable level of attrition, indicating comparatively higher employee movement in revenue-facing roles.
- In contrast, the Human Resources department records the lowest attrition, suggesting stronger employee stability or fewer external job-switching opportunities.