IBM HR Analytics Employee Attrition & Performance

Understanding and Reducing Employee Attrition Using Data Analytics

Presented By:

Subrata Dey
Data Analyst Intern
Tools Used: Python, SQL, Tableau
Unified Mentor Private Limited
Date:

Executive Summary

- Overall attrition rate is **16.12%**, impacting workforce stability.
- Research & Development department (**56.12%**) contributes the highest attrition volume.
- Single employees show the highest attrition rate (Male = **30.8%** & Female = **19.83%**).
- Employees living closer to the office **(0–5 km)** still show notable attrition, indicating non-commute factors.
- Job satisfaction and engagement are stronger drivers than age alone.

Problem Statement and Objectives

Problem Statement

Employee attrition leads to loss of talent, increased recruitment costs, and reduced productivity. Understanding who is leaving and why is essential for retention planning.

Objectives of the Analysis

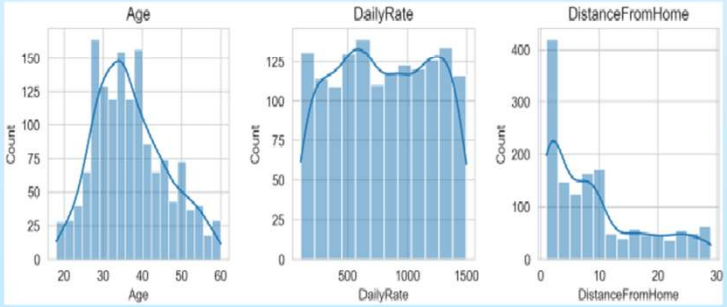
- Measure employee attrition and turnover rate.
- Identify demographic and job-related attrition patterns.
- Analyze engagement, satisfaction, and commute impact.
- Enable data-driven HR decisions using dashboards.

Data Preparation and Cleaning

Data Loading & Cleaning

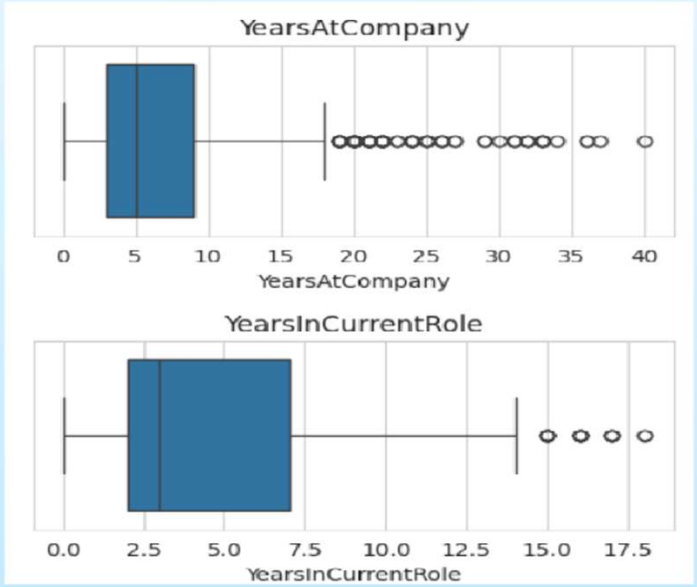
```
# Load the dataset:
df = pd.read_csv('/content/HR-Employee-Attrition-dataset.csv')

# Looking for any Missing Values:
df.isnull().sum()

# Check number of duplicate data:
print('Number of Duplicate Data:', df.duplicated().sum())

# Remove Constant columns & Irrelevant Columns
df2 = df.drop(columns=['EmployeeCount','Over18','StandardHours'],
errors='ignore')
df2

# Generate descriptive statistics for numerical columns
df2.describe().T      # T = Transpose
```

Understanding Some Data Distributions



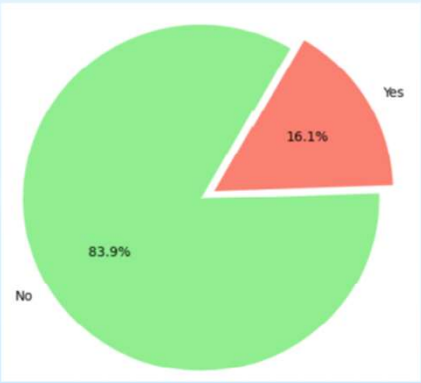
Identifying Outliers

```
# Identifying Outliers

for i in df2.select_dtypes(include='number').columns:
  plt.figure(figsize=(5, 2))
  sns.boxplot(data=df2, x=i)
  plt.title(i)
  plt.show()
```



Employee Attrition Percentage

```
# Count attrition values
counts = df3['Attrition'].value_counts()

# Create pie chart
plt.figure(figsize=(6,6))
plt.pie(counts, labels=counts.index,
        autopct='%1.1f%%',  # show percentages
        startangle=60,
        colors=['lightgreen','salmon'],
        explode=(0.09, 0))
# Add title
plt.title("Employee Attrition Percentage")
plt.show()
```

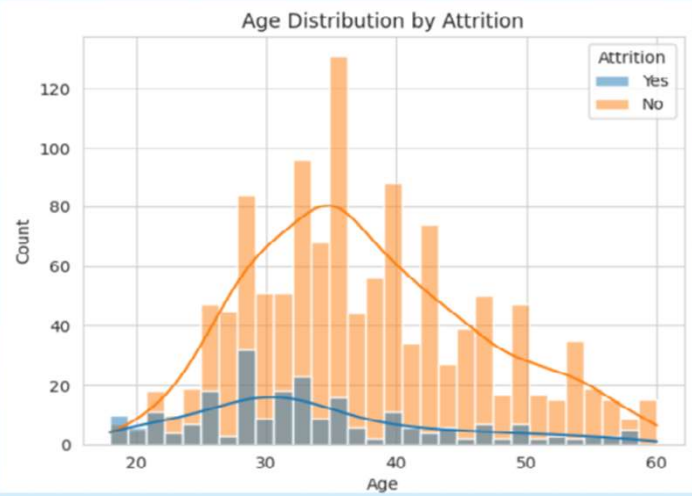


- From the pie plot we can represent in the percentage format that employees who have not attrited ('No') is 83.9% and employees who have attrited ('Yes') is 16.1%.

Exploratory Data Analysis (EDA) with Python

Age Distribution with Attrition

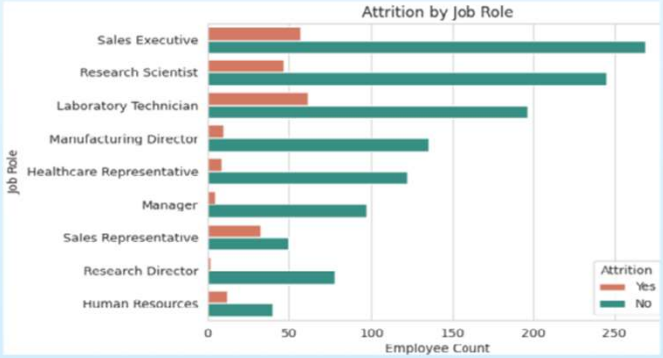
```
# Plot age distribution by attrition
sns.histplot(data=df3, x='Age', hue='Attrition',
             bins=30, kde=True)
plt.title('Age Distribution by Attrition')
plt.show()
```



- From the above charts we can observe that, employees in younger age groups, appear to have a higher propensity to attrite ('Yes') compared to older employees.
- Older employees tend to be more stable and less likely to leave the company.

Attrition by Job Role

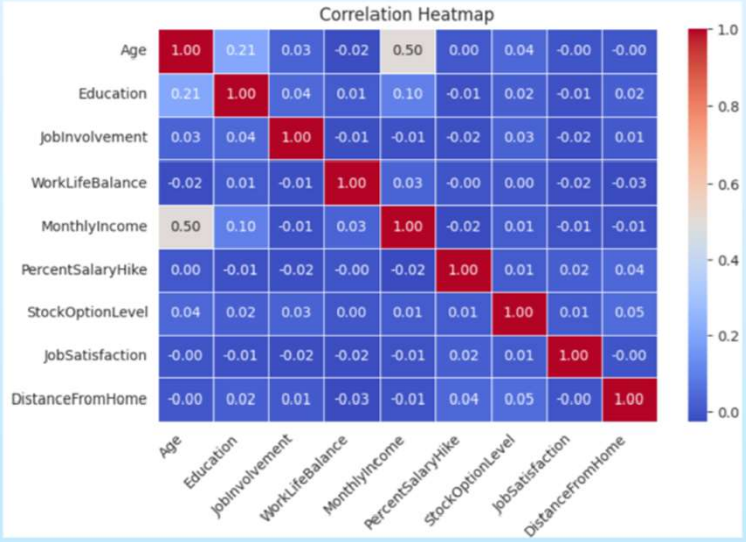
```
sns.set_style("whitegrid")
# Create horizontal grouped bar chart for Job
Role vs Attrition
sns.countplot(
    y='JobRole',
    hue='Attrition',
    data=df3,
    palette={'No': '#2A9D8F', 'Yes': '#E76F51'}
)
plt.title("Attrition by Job Role")
plt.xlabel("Employee Count")
plt.ylabel("Job Role")
plt.legend(title="Attrition")
plt.show()
```



- Sales Executives and Lab Technicians face the highest attrition, while Research Scientists shows comparatively stronger retention.
- Monthly income rises moderately with age, while other variables show little to no correlation.

Correlation Heatmap

```
# Keep only important numeric columns for attrition analysis
col = [ 'Age', 'Education', 'JobInvolvement', 'WorkLifeBalance',
       'MonthlyIncome','PercentSalaryHike','StockOptionLevel',
        'JobSatisfaction', 'DistanceFromHome']
corr = df4[col].corr()
plt.figure(figsize=(8,5))
sns.heatmap(
    corr,
    cmap="coolwarm",
    annot=True,      # show correlation values
    fmt='.2f',       # format to 2 decimal places
    linewidth=0.5,   # add grid lines between cells
    cbar=True        # show color bar
)
plt.title("Correlation Heatmap")
# Tilt x-axis labels for readability
plt.xticks(rotation=45, ha='right')
plt.show()
```



SQL Analysis for Insights

1) Top 5 Job Roles with the Lowest Average Job Satisfaction

```
Q. Top 5 Job Roles with the Lowest Average Job Satisfaction
 select
     JobRole,
     round(avg(JobSatisfaction), 2) as avg_job_satis,
     rank() over(
         order by avg(JobSatisfaction) asc
     ) as satis_rank
 from hr_employee_attrition
 group by JobRole
 order by satis_rank
 limit 5;
```

Result Grid | Filter Rows:

JobRole	avg_job_satis	satis_rank
Human Resources	2.56	1
Manufacturing Director	2.68	2
Laboratory Technician	2.69	3
Research Director	2.70	4
Manager	2.71	5

Result 2

2) Job roles with highest attrition rate (%)

```
select
    JobRole,
    Round(
        sum(case when Attrition = 'Yes' then 1 else 0 end) * 100.0 / count(*), 2
    ) as att_rate_perc
from hr_employee_attrition
group by JobRole
order by att_rate_perc desc;
```

Result Grid | Filter Rows:

JobRole	att_rate_perc
Sales Representative	39.76
Laboratory Technician	23.94
Human Resources	23.08
Sales Executive	17.48
Research Scientist	16.10
Manufacturing Director	6.90
Healthcare Representative	6.87
Manager	4.90
Research Director	2.50

3) How Work-life balance differ for employees who left vs stayed?

```
select
    WorkLifeBalance,
    Attrition,
    count(*) as employee_count
from hr_employee_attrition
group by WorkLifeBalance, Attrition
order by WorkLifeBalance;
```

Result Grid | Filter Rows:

WorkLifeBalance	Attrition	employee_count
1	No	55
1	Yes	25
2	No	286
2	Yes	58
3	No	766
3	Yes	127
4	No	126
4	Yes	27

4) Compare average monthly income for employees who left vs stayed

```
select
    Attrition,
    round(avg(MonthlyIncome),2)  -- upto 2 decimal places
    as avg_monthly_income
from hr_employee_attrition
group by Attrition;
```

Result Grid | Filter Rows:

Attrition	avg_monthly_income
Yes	4787.09
No	6832.74

5) Department having the highest attrition count?

```
select
    Department,
    count(*) as attrition_count
from hr_employee_attrition
where Attrition = 'Yes'
group by Department
order by attrition_count desc;
```

Result Grid | Filter Rows:

Department	attrition_count
Research & Development	133
Sales	92
Human Resources	12

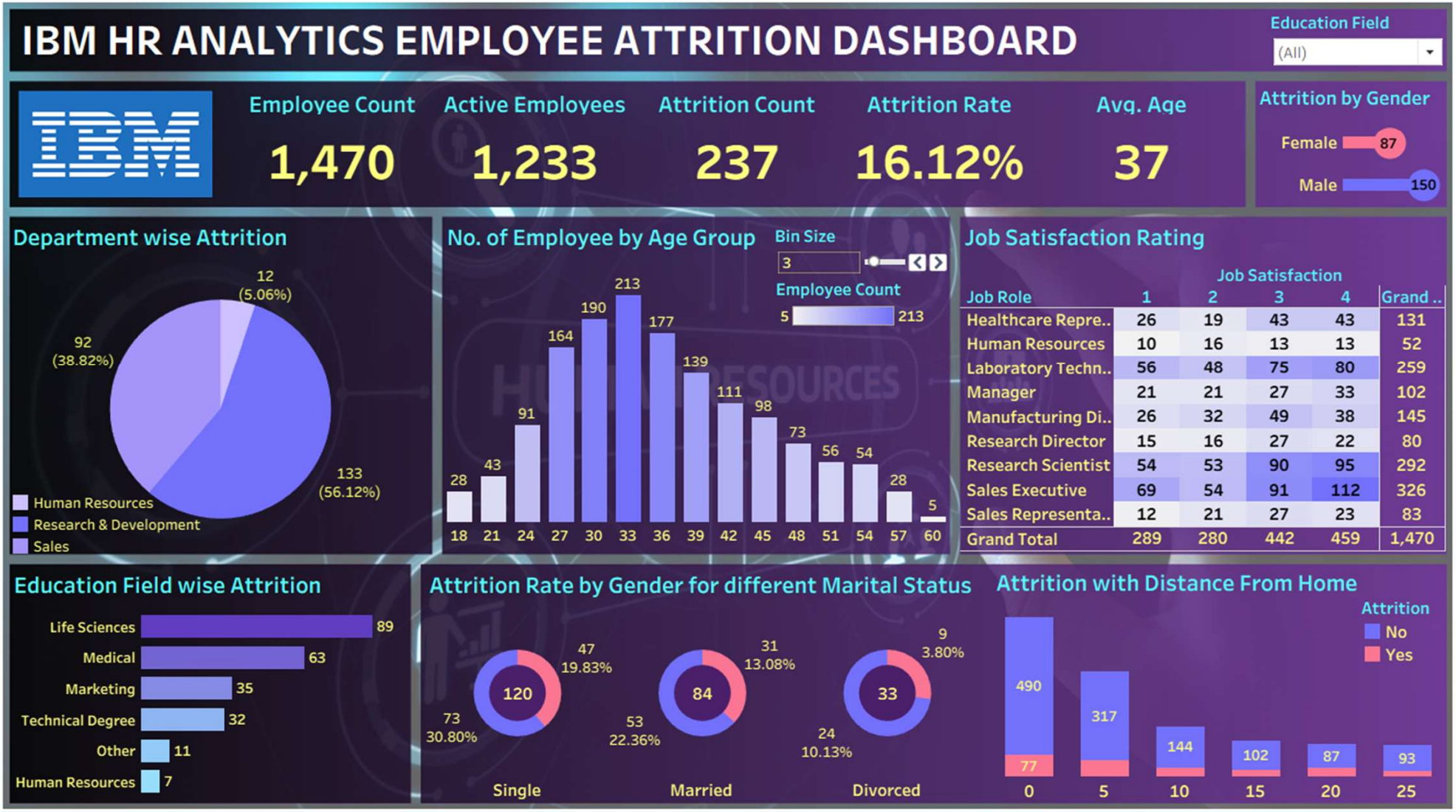
6) Average age of employees who left vs stayed

```
select Attrition, round(avg(Age), 2) as avg_age
from hr_employee_attrition
group by Attrition;
```

Result Grid | Filt

Attrition	avg_age
Yes	33.61
No	37.56

Tableau Dashboard Insights



Key Insights and Recommendations

Key Insights

- Engagement and satisfaction outweigh age as attrition drivers.
- Single employees and certain job roles need focused attention.
- Benefits like stock options and work-life balance reduce attrition risk.

Strategic Recommendations

- Improve work-life balance policies.
- Introduce targeted engagement programs.
- Review compensation and long-term benefits.
- Use dashboards for continuous HR monitoring.

Conclusion

This project demonstrates end-to-end analytics using Python, SQL, and Tableau to deliver actionable HR insights and support effective retention strategies.