Project Title:- IBM HR Analytics Employee Attrition & Performance
Tools:- Python,DAX(PowerQuery),Jupyternotebook,SQL
Domain:- Data Analyst & Data scientist
Data Set:- Data Folder -- Contains Raw Data Files (Data can be downloaded from kaggle)

Objective: This real-world HR analytics project focused on understanding employee attrition trends and identifying key factors influencing turnover using data-driven insights. With 35 HR variables across 1,470 employees, the analysis helped HR professionals pinpoint critical retention levers and proactively improve employee experience.

INTRODUCTION

Understanding the data:

- **The data is taken from the source** kaggle
- The data contains 1 csv file HR-Employee-Attrition.CSV

Tasks:

- **Data Cleaning:** Clean the Attrition Dataset by handling missing values,duplicates and inconssteNcies.
- **Data Aggregaton:** Aggregate the data to compute Attrition count,Attrition Rate,Active Employees and Measure like performance rating label,job satisfaction level,work life balance label,Job Involvement level,Environment Satisfaction level.

- **Visualization of Dashboard:** Create various Visualization such as:

1. EducationField & Jobrole by Department Count(clustered bar chart)
2. Gender(slicer)
3. MaritalStatus by Sum of Attrition & Relationship Education (clustered bar chart)
4. Attrition by JobRole(clustered bar chart)
5. Distance from home & Education field by Jobrole & Active employee(Ribbon Chart)
6. Avg of Job Involvement by Education Field (Funnel)

- **Visualization For Uncover The Factor Leading to Attrition:**

1. Monthly Income & Attrition By Education(line & clustered coloumn chart)
2. Performance rating label,Job satisfaction,Work life balance of employee,relationship satisfaction,environment satisfaction(matrix)
3. Avg salary hike & hourly rate by performance rating

- Build an interactive dashboard that allows users to filter by time Attrition of employee,worklife balance, performance,satisfaction etc.
- **Insights:**

1. Accurately identify 16.12% customers (237) likely to churn

2. Highest attrition from Jobrole are of Laboratory Technician(62),Sales Executive(57),Research Scientist(47)
3. Thier Average Life balance is from 2.72 to 2.89 & Age 30 - 36
4. Stock Options of Department Research & Development(133),Sales(92) and Human Resource(12)

- **Tools:**
  1. **Power BI** for dashboarding
  2. **Python (Pandas, Seaborn, scikit-learn)** for ML and EDA
  3. **SQL Server / CSV** as data source
  4. **Jupyter Notebook** for workflow and documentation

- **ASK:** By this project, I am trying to analyse Attrition performance for **Facotors leading to Employee Attrtion** company and present my insights in a comprehensive dashboard

- **PREPARE :**

1. Load the datasets to Google Sheets and PowerQuery (SQL)
2. Prepared a copy of the dataset for further reference
3. Saved the datasets by using naming conventions in a separate file folder for easy access

- **Process :**

Importing libraries

```
import numpy as np   # Its used primarily for numerical computing.It support for multidimensional arrays and matrices
import matplotlib.pyplot as plt # creating visualization
import seaborn as sns #Its used for statistical graphics
import pandas as pd #used for data manipulation,cleaning and Analysis
```

Import Data

```
data = pd.read_csv(r"C:\Users\kanak\Desktop\UNIFIED MENTOR PROJ\DAproj_1IN\WA_Fn-UseC_-HR-Employee-Attrition.csv")
data.head()
data.shape
(1470, 35)

import pandas as pd

print(f'Number of duplicated data: {data.duplicated().sum()}')
print(data.isnull().sum() / len(data)*100)
Number of duplicated data: 0
Age                 0.0
Attrition            0.0
```

```
BusinessTravel              0.0
DailyRate                   0.0
Department                  0.0
DistanceFromHome            0.0
Education                   0.0
EducationField              0.0
EmployeeCount               0.0
EmployeeNumber              0.0
EnvironmentSatisfaction     0.0
Gender                      0.0
HourlyRate                  0.0
JobInvolvement              0.0
JobLevel                    0.0
JobRole                     0.0
JobSatisfaction             0.0
MaritalStatus               0.0
MonthlyIncome               0.0
MonthlyRate                 0.0
NumCompaniesWorked          0.0
Over18                      0.0
OverTime                    0.0
PercentSalaryHike           0.0
PerformanceRating           0.0
RelationshipSatisfaction    0.0
StandardHours               0.0
StockOptionLevel            0.0
TotalWorkingYears           0.0
TrainingTimesLastYear       0.0
WorkLifeBalance             0.0
YearsAtCompany              0.0
YearsInCurrentRole          0.0
YearsSinceLastPromotion     0.0
YearsWithCurrManager        0.0
dtype: float64
```

```
data.dtypes
data.describe
```

```
<bound method NDFrame.describe of       Age Attrition     BusinessTravel  DailyRate
Department  \
0      41       Yes      Travel_Rarely       1102                 Sales
```

```
1      49       No  Travel_Frequently       279  Research & Development
2      37      Yes      Travel_Rarely      1373  Research & Development
3      33       No  Travel_Frequently      1392  Research & Development
4      27       No      Travel_Rarely       591  Research & Development
...   ...      ...                ...       ...                     ...
1465   36       No  Travel_Frequently       884  Research & Development
1466   39       No      Travel_Rarely       613  Research & Development
1467   27       No      Travel_Rarely       155  Research & Development
1468   49       No  Travel_Frequently      1023                   Sales
1469   34       No      Travel_Rarely       628  Research & Development

      DistanceFromHome  Education EducationField  EmployeeCount  \
0                    1          2  Life Sciences              1
1                    8          1  Life Sciences              1
2                    2          2          Other              1
3                    3          4  Life Sciences              1
4                    2          1        Medical              1
...                ...        ...            ...            ...
1465                23          2        Medical              1
1466                 6          1        Medical              1
1467                 4          3  Life Sciences              1
1468                 2          3        Medical              1
1469                 8          3        Medical              1

      EmployeeNumber  ...  RelationshipSatisfaction StandardHours  \
0                  1  ...                         1            80
1                  2  ...                         4            80
2                  4  ...                         2            80
3                  5  ...                         3            80
4                  7  ...                         4            80
...              ...  ...                       ...           ...
1465            2061  ...                         3            80
1466            2062  ...                         1            80
1467            2064  ...                         2            80
1468            2065  ...                         4            80
1469            2068  ...                         1            80

      StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  \
0                    0                  8                      0
```

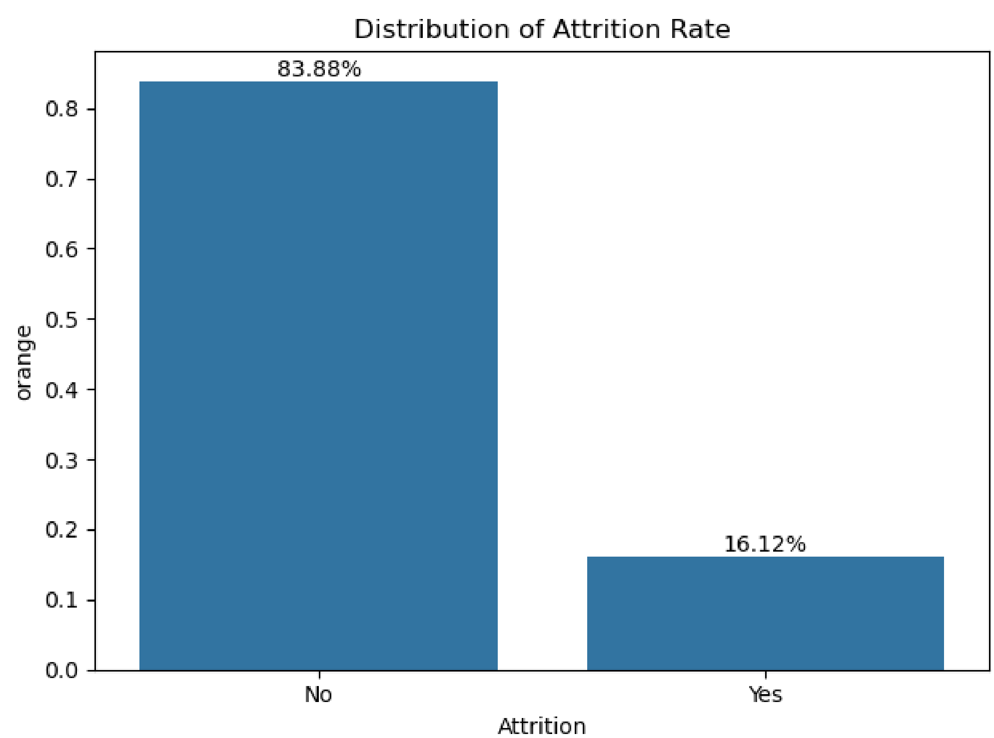
```
1       1      10        3
2       0       7        3
3       0       8        3
4       1       6        3
...    ...     ...      ...
1465    1      17        3
1466    1       9        5
1467    1       6        0
1468    0      17        3
1469    0       6        3

     WorkLifeBalance  YearsAtCompany YearsInCurrentRole  \
0          1            6            4
1          3           10            7
2          3            0            0
3          3            8            7
4          3            2            2
...       ...          ...          ...
1465       3            5            2
1466       3            7            7
1467       3            6            2
1468       2            9            6
1469       4            4            3

     YearsSinceLastPromotion  YearsWithCurrManager
0                 0              5
1                 1              7
2                 0              0
3                 3              0
4                 2              2
...              ...            ...
1465              0              3
1466              1              7
1467              0              3
1468              0              8
1469              1              2

[1470 rows x 35 columns]>
```

To find Attrition Rate

```
data['Attrition'].value_counts(normalize=True)
```

```
Attrition
No      0.838776
Yes     0.161224
Name: proportion, dtype: float64
```

```
import matplotlib.pyplot as plt
import seaborn as sns
Attrition = data['Attrition'].value_counts(normalize=True)
ax = sns.barplot(x=Attrition.index, y=Attrition)
for p in ax.patches:
    ax.annotate(f'{p.get_height() * 100:.2f}%',
                (p.get_x() + p.get_width() / 2., p.get_height()),
                ha='center', va='center', fontsize=10, color='black', xytext=(0, 5),
                textcoords='offset points')
plt.title('Distribution of Attrition Rate')
plt.xlabel('Attrition')
plt.ylabel('Percentage')
plt.tight_layout()
plt.show()
```



To Find Average tenure

```
avg_tenure = data['YearsAtCompany'].mean()
print(f'Average years of employee to leave the company years is')
print(avg_tenure)
```

Average years of employee to leave the company years is
7.0081632653061225

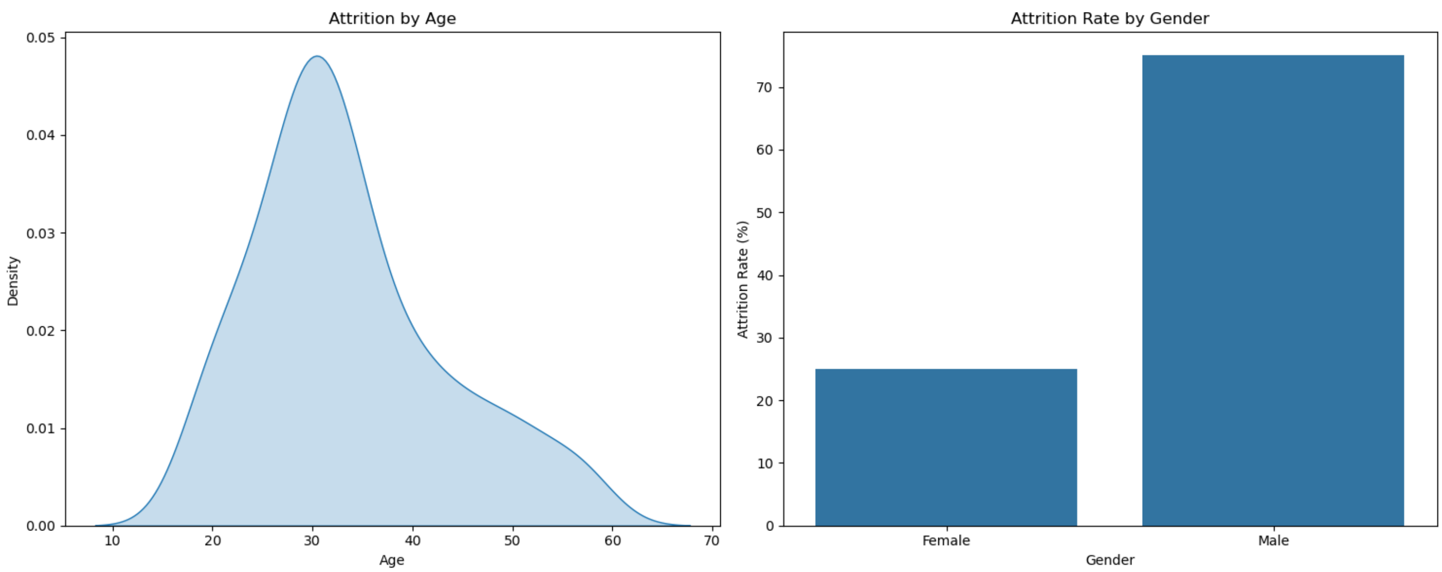
Plotting Distribution of Employee by Age,Employee by Gender and Employee by Department

```
fig, axes = plt.subplots(nrows=1, ncols=3, figsize=(15,5))
sns.histplot(data=data, x='Age', kde=True, ax=axes[0])
axes[0].set_title('Distribution Employee by Age')
axes[0].set_xlabel('Age')
axes[0].set_ylabel('Count')
sns.countplot(data=data, x='Department', ax=axes[2])
axes[2].set_title('Distribution Employee by Department')
axes[2].set_xlabel('Department')
axes[2].set_ylabel('Count')
sns.countplot(data=data, x='Gender', ax=axes[1])
axes[1].set_title('Distribution Employee by Gender')
axes[1].set_xlabel('Gender')
axes[1].set_ylabel('Count')
plt.tight_layout()
```

To Calculate Attrition by Age and Attrition Rate by Gender

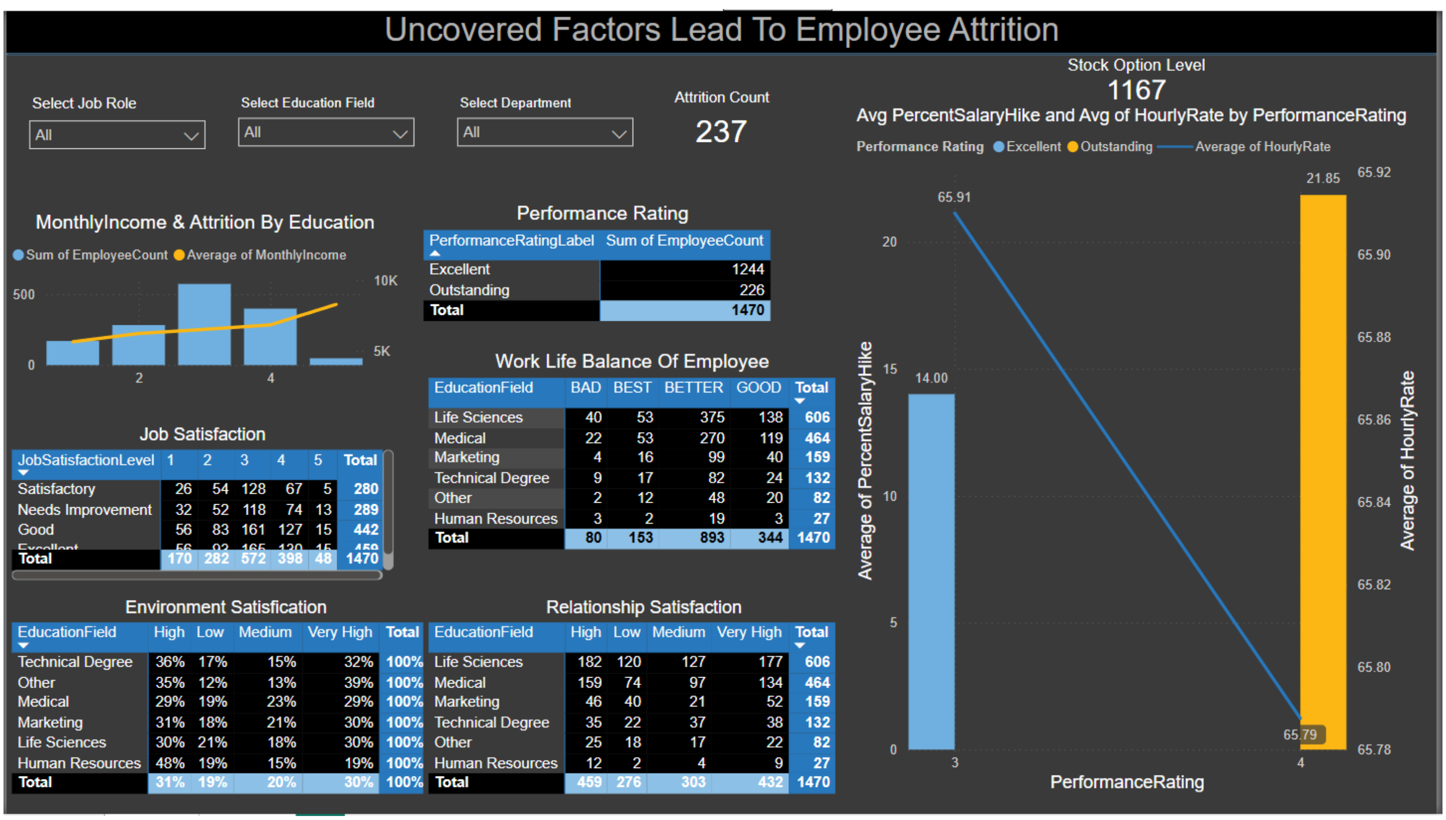
```
def calculate_Attrition_rate(data, column):
    attrition_counts = data.groupby([column, 'Attrition']).size().unstack(fill_value=0)
    attrition_rate = attrition_counts['Yes'] / attrition_counts.sum(axis=1) * 100
    attrition_rate_data = attrition_rate.reset_index()
    attrition_rate_data.columns = [column, 'Attrition Rate']
```

```
    return attrition_rate_data
fig, axes = plt.subplots(nrows=1, ncols=2, figsize=(15,6))
# Plot 1: KDE plot of Age with Attrition hue
sns.kdeplot(data=data_attrition, x='Age', fill=True, ax=axes[0])
axes[0].set_title('Attrition by Age')
axes[0].set_xlabel('Age')
axes[0].set_ylabel('Density')
# Plot 2: Bar plot of Gender count with Attrition hue
attrition_rate_data = calculate_attrition_rate(data, 'Gender')
sns.barplot(data=attrition_rate_data, x='Gender',y='AttritionRate', ax=axes[1])
axes[1].set_title(f'Attrition Rate by Gender')
axes[1].set_xlabel('Gender')
axes[1].set_ylabel('Attrition Rate (%)')
plt.tight_layout()
plt.show()
```



Share :


IBM
DASHBOARD
OVERVIEW
Average of MonthlyIncome
6.50K
Years At Company
10K
Avg of work Life Balance
2.76
Employee
1K
Attrition
237
Avg Age
36.92
Avg Tenure
7.01
Attrition Rate
16.12%
Avg Income
6.50K
Active Employee
1K
Job Satisfaction
Business Travel
Education Field and JobRole
EducationField
Count of Department
Female
Male
Relations of Active Employee
MaritalStatus
Married
Single
Divorced
Attrition by Job Role
JobRole
Distance From Home & Attrition of employee
Count of JobRole
Active employees
DistanceFromHome
Avg of Job Involvement by EducationField
Life Sciences 606
Medical 464
Marketing 159
Technical Degree 132
Other 82
Human Resources 27


Uncovered Factors Lead To Employee Attrition
Select Job Role
Select Education Field
Select Department
Attrition Count
237
Stock Option Level
1167
Avg PercentSalaryHike and Avg of HourlyRate by PerformanceRating
Performance Rating
Excellent
Outstanding
Average of HourlyRate
MonthlyIncome & Attrition By Education
Sum of EmployeeCount
Average of MonthlyIncome
Performance Rating
PerformanceRatingLabel Sum of EmployeeCount
Excellent 1244
Outstanding 226
Total 1470
Work Life Balance Of Employee
EducationField BAD BEST BETTER GOOD Total
Life Sciences 40 53 375 138 606
Medical 22 53 270 119 464
Marketing 4 16 99 40 159
Technical Degree 9 17 82 24 132
Other 2 12 48 20 82
Human Resources 3 2 19 3 27
Total 80 153 893 344 1470
Job Satisfaction
JobSatisfactionLevel 1 2 3 4 5 Total
Satisfactory 26 54 128 67 5 280
Needs Improvement 32 52 118 74 13 289
Good 56 83 161 127 15 442
Total 170 282 572 398 48 1470
Environment Satisfaction
EducationField High Low Medium Very High Total
Technical Degree 36% 17% 15% 32% 100%
Other 35% 12% 13% 39% 100%
Medical 29% 19% 23% 29% 100%
Marketing 31% 18% 21% 30% 100%
Life Sciences 30% 21% 18% 30% 100%
Human Resources 48% 19% 15% 19% 100%
Total 31% 19% 20% 30% 100%
Relationship Satisfaction
EducationField High Low Medium Very High Total
Life Sciences 182 120 127 177 606
Medical 159 74 97 134 464
Marketing 46 40 21 52 159
Technical Degree 35 22 37 38 132
Other 25 18 17 22 82
Human Resources 12 2 4 9 27
Total 459 276 303 432 1470
Average of PercentSalaryHike
Average of HourlyRate
PerformanceRating

- **ACT (Actions to Take)**

Based on the analysis of the HR attrition data, the following actions are recommended:

1. Employees in Sales and R&D, which show the highest attrition rates, should be offered flexible hours and remote work opportunities to reduce stress and improve work-life balance.
2. Since over 53% of employees who left were working overtime, it's crucial to track workloads and redistribute tasks to prevent employee burnout.
3. Job role satisfaction is a strong attrition driver. Run frequent anonymous employee engagement surveys to monitor satisfaction and act early on dissatisfaction.
4. Introduce bonuses and stock options for high performers, especially in roles like Sales Executive and Laboratory Technician which have higher attrition rates.
5. Establish clear and personalized career paths with defined promotion criteria. This will retain mid-career professionals who form a major chunk of attrition.

- **Suggestions**

1. Address Department-Specific Needs:
   Tailor retention strategies for departments like Sales and R&D, where turnover is highest.
2. Target High-Risk Age Group:
   Focus on employees aged 28–39 years with personalized development plans, as they form 50% of total attrition cases.
3. Support Low-Income Employees:
   Many employees earning below $3,200 per month are leaving. Consider salary adjustments or non-monetary benefits like learning stipends or flexible hours.
4. Improve Manager-Employee Relations:
   Offer leadership training to managers to improve team engagement and communication.
5. Use Predictive Analytics:
   Build and deploy a model using this dataset to predict attrition risk and act proactively with one-on-one interventions.

- **Conclusion :**

1. Attrition is strongly tied to employee workload, satisfaction, and compensation.
2. Employees in Sales, younger age groups, or those doing overtime are most at risk.
3. Implementing better work-life policies, performance-linked rewards, and Managerial engagement can improve retention.
4. ML models provide reliable prediction to support HR decision-making.