

## CHAPTER 1

### INTRODUCTION

Human Resources (HR) Analytics, also known as People Analytics, refers to the application of data-driven methodologies and statistical analysis to gain insights into various aspects of human resource management. It is an emerging field that helps organizations optimize their workforce and improve overall employee performance. This project, “HR Analytics to Track Employee Performance,” aims to leverage data analytics techniques to assess and track employee performance using relevant data from a dataset containing employee records.

The primary goal of this project is to analyze key performance indicators (KPIs) such as attendance, training, job satisfaction, and compensation, and to correlate these factors with employee performance outcomes. By examining such data, organizations can gain valuable insights into patterns and trends that contribute to high or low performance levels. HR Analytics allows organizations to identify employees who need additional support, recognize high performers, and predict potential issues such as employee turnover or disengagement.

For this project, a dataset containing 5,000 employee records will be analyzed. The dataset includes various attributes such as employee demographics, performance ratings, department information, training details, and compensation data. Using Python and libraries like Pandas, Matplotlib, and Seaborn, the data will be cleaned, visualized, and analyzed through a series of graphs to uncover insights about the organization’s workforce.

By utilizing advanced data visualization techniques, this project aims to provide a comprehensive understanding of how different factors impact employee performance. The findings can assist HR professionals in making informed decisions related to performance management, talent development, and employee retention. Ultimately, this project highlights the potential of HR Analytics in transforming human resource functions and driving data-driven decision-making processes in organizations.

## CHAPTER 2

### LITERATURE REVIEW

Human Resource (HR) Analytics, also known as people analytics, is an emerging field that leverages data analysis techniques to enhance HR decision-making. With the increasing availability of employee data, organizations are now able to use analytical tools to gain actionable insights into workforce performance, productivity, and engagement.

A number of studies emphasize the significance of data-driven decision-making in HR. Bassi and McMurrer (2016) highlighted that firms using advanced analytics in HR practices outperform their competitors in talent acquisition, retention, and development. Similarly, Rasmussen and Ulrich (2015) pointed out that the integration of HR data with business outcomes can significantly improve strategic alignment and performance tracking.

In the context of employee performance, analytics can help identify high performers, predict attrition risks, and optimize training and development plans. Techniques such as regression analysis, clustering, and machine learning are frequently used to analyze variables like job satisfaction, appraisal scores, absenteeism, and employee engagement.

Seong and Hong (2018) further explored the impact of HR analytics in performance management, suggesting that predictive modeling enhances the ability to make objective and proactive HR decisions. Additionally, real-time dashboards and visualizations allow managers to monitor performance trends and intervene early when necessary.

Overall, the literature strongly supports the adoption of HR analytics to transform traditional personnel management into a strategic, evidence-based function. This project builds upon these findings to explore key HR metrics and visualize employee performance using real-world data and Python-based analytical tools.

## CHAPTER 3

### AIMS AND OBJECTIVES

#### 3.1 Aim of the Project

1. **Enhance Decision-Making:** By utilizing HR data analytics, organizations can make more informed decisions regarding employee performance, promotions, retention strategies, and workforce planning, ensuring that HR actions are data-driven rather than based on assumptions.
2. **Optimize Development:** HR analytics helps identify skill gaps and areas where employees need further development, enabling the creation of targeted training and development programs that align with individual and organizational goals.
3. **Boost Efficiency:** Performance data analysis enables HR departments to streamline their processes, enhance recruitment, improve staff allocation, and make more efficient staffing decisions to increase productivity and minimize wasted resources.
4. **Predict Trends:** By leveraging historical data on employee performance and behavior, HR analytics can forecast future trends such as employee turnover, career growth patterns, and workforce needs, allowing the organization to take proactive measures.

#### 3.2 Specific Objectives

1. **Analyze Performance Data:** HR analytics tracks a variety of performance indicators such as job satisfaction, performance reviews, productivity metrics, and employee engagement scores to assess overall employee performance and identify areas for improvement.
2. **Identify Performers:** Using data, HR can pinpoint high performers who may be considered for promotions or leadership roles, while also identifying low performers who may require additional support, training, or role reassignment to enhance their contribution.
3. **Evaluate Training Impact:** HR analytics helps assess the effectiveness of training programs by measuring changes in employee performance, engagement, and productivity before and after the training, ensuring that investments in development are yielding the desired outcomes.
4. **Monitor Attrition:** Analyzing patterns and reasons behind employee turnover allows HR to uncover root causes of attrition and identify trends, enabling the organization to implement retention strategies that reduce turnover rates.

5. Predict Performance: By analyzing historical data, predictive analytics can be used to anticipate potential performance challenges or success factors for employees, allowing HR to take preemptive actions to address or capitalize on these trends.

6. Benchmark: HR analytics compares internal performance metrics with industry standards or competitor data to ensure that the organization remains competitive in talent management, helping HR align performance expectations and strategies to industry best practices.

## CHAPTER 4

### METHODOLOGY

#### 4.1 DATA COLLECTION & DATASET DESCRIPTION

The dataset used in this project contains 5000 employee records, capturing various HR-related attributes such as Age, Department, Salary, Education, Job Role, Years at Company, Performance Ratings, etc. These attributes are crucial in analyzing and predicting employee performance and retention.

Key features include:

- Demographics: Age, Gender, Education
- Professional Details: Department, Job Role, Years at Company, Salary
- Performance Metrics: Performance Rating, Promotion History

Objective: To gain insights into employee performance patterns, trends, and factors influencing productivity and retention.

Sample Code:

```
Df = pd.read_csv("5000 HRA Records.csv")  
Print(df.head())
```

#### 4.2 DATA ANALYSIS – PREPARATION & PREPROCESSING

Initial exploration helps understand data distribution, detect missing values, and recognize patterns. Preprocessing includes handling categorical variables and preparing data types for analysis. The raw data is first inspected for structure and completeness. This helps to shape further cleaning and transformation efforts.

Tasks:

- Check for null values
- Understand column data types
- Review value distributions

Sample Code:

```
Print(df.info())
```

```
Print(df.isnull().sum())
```

```
Print(df['Department'].value_counts())
```

### 4.3 DATA CLEANING & MANIPULATION

This step involves removing duplicates, handling missing values, correcting inconsistencies, and filtering irrelevant data to ensure quality and accuracy in the analysis.

Methods include:

- Duplicate removal
- Missing value imputation
- Type conversion (e.g., dates)
- Renaming for clarity

Sample Code:

```
Df = df.drop_duplicates()
```

```
Df['Salary'].fillna(df['Salary'].mean(), inplace=True)
```

```
Df['Hire Date'] = pd.to_datetime(df['Hire Date'])
```

```
Df.rename(columns={'YearsAtCompany': 'Years_At_Company'}, inplace=True)
```

### 4.4 DATA TRANSFORMATION

Transforming the data to make it suitable for visualization and modeling, e.g., encoding categorical variables, scaling, creating derived metrics.

To support visualizations and modeling:

- Categorical encoding is applied
- Features like normalized performance scores and age groups are created
- Derived insights like tenure bands or department-wise summaries are calculated

Code:

Encode categorical variables

```
df['Gender'] = df['Gender'].map({'Male': 1, 'Female': 0})
```

Create new column: Performance Score normalized (if needed)

```
from sklearn.preprocessing import MinMaxScaler
```

```
scaler = MinMaxScaler()
```

```
df['Normalized_Performance'] = scaler.fit_transform(df[['PerformanceRating']])
```

Create bins for Age group

```
df['Age_Group'] = pd.cut(df['Age'], bins=[18, 30, 40, 50, 60],  
                        labels=["18-30", "31-40", "41-50", "51-60"])
```

## **CHAPTER 5**

### **EXPLORATORY DATA ANALYSIS**



## CHAPTER 6

### RESULTS AND DISCUSSIONS

This HR Analytics project aimed to explore and visualize employee performance patterns using a dataset of 5000 HR records. Based on the visualizations created (10 graphs), several key insights were observed:

#### 6.1 Employee Count by Department

**Insight:**

The largest number of employees belong to departments like Sales and Research & Development. Smaller departments such as Human Resources and Legal had comparatively fewer employees.

#### 6.2 Age Distribution

**Insight:**

Most employees are in the age group of 30 to 45, indicating a predominantly mid-career workforce.

#### 6.3 Gender Ratio

**Insight:**

A slight gender imbalance was noted, with a higher number of male employees. Gender diversity could be a focus area for HR.

#### 6.4 Salary vs. Performance

**Insight:**

No direct correlation was observed between salary and performance rating, suggesting that performance evaluation may not always align with compensation.

#### 6.5 Heatmap of Correlation Matrix

**Insight:**

Features such as "Years at Company", "Years in Current Role", and "Total Working Years" showed positive correlations with performance, suggesting experience plays a key role in performance.

#### 6.6 Attrition vs. Tenure

**Insight:**

Attrition rates were higher among employees with shorter tenures, pointing toward onboarding and early engagement challenges.

### 6.7 Promotion History by Department

**Insight:**

Departments like R&D showed higher promotion rates. Some areas had stagnant promotion trends, possibly indicating limited growth opportunities.

### 6.8 Education Level vs. Performance

**Insight:**

Higher education levels (Master's or above) were not necessarily linked to higher performance scores, suggesting other factors like skills or experience might play a bigger role.

### 6.9 Department-wise Average Salary

**Insight:**

Finance and Management departments had the highest average salaries, aligning with industry trends.

### 6.10 Performance Rating Distribution

**Insight:**

Most employees were clustered around average to good performance scores. Few employees were rated poor or excellent, indicating a standard bell curve approach in evaluations.

## CHAPTER 7

### FUTURE ENHANCEMENTS

A future enhancement of this project could be adding machine learning and interactive dashboards to predict employee performance and provide real-time insights for HR decisions.

Although this project already gives useful insights into employee performance, several improvements and additions can make it even better:

#### 7.1 Predictive Analytics

- What it means: Use machine learning models to predict outcomes like employee attrition (who might leave the company) or future performance scores.
- How to do it: Train models using features like salary, experience, education, etc.
- Tools: Use `scikit-learn`, `XGBoost`, or `LightGBM`.

#### 7.2 Interactive Dashboards

- What it means: Create user-friendly dashboards where HR teams can filter, click, and view data easily.
- How to do it: Use tools like `Plotly Dash`, `Streamlit`, or `Power BI`.
- Benefit: Helps managers make decisions quickly with visual summaries.

#### 7.3 Automated Reports

- What it means: Automatically generate performance reports monthly or quarterly.
- How to do it: Schedule scripts using Python (`cron jobs` or `Task Scheduler`) that export charts and summaries to PDFs or Excel files.

#### 7.4 Sentiment Analysis from Employee Feedback

- What it means: Analyze employee comments or feedback (if available) to detect satisfaction or dissatisfaction.
- How to do it: Use Natural Language Processing (NLP) libraries like `TextBlob` or `NLTK`.

### **7.5 Time-Series Analysis**

- What it means: Analyze how performance or attrition trends change over time.
- How to do it: Use time-based plots and forecasting models like ARIMA.

### **7.6 Integration with Real HR Systems**

- What it means: Connect this analysis tool with real HR software (like SAP, Workday, or Excel sheets) for live data updates.
- How to do it: Use APIs or connectors to fetch data directly from HR systems.

### **7.7 More Advanced Visualizations**

- What it means: Add heatmaps, interactive maps, pie charts, or scatter plots for deeper insights.
- How to do it: Use `Plotly`, `Bokeh`, or `Altair`.

### **7.8 Employee Clustering**

- What it means: Group employees based on similar characteristics (like K-Means clustering).
- Use case: Identify top performers, at-risk employees, or training needs.

## CONCLUSION

This HR Analytics project aimed to explore and understand key patterns in employee performance using data-driven visualizations and analysis. By examining a dataset containing 5000 employee records, we were able to gain valuable insights into how various factors like age, department, salary, education, job role, and years at the company relate to performance ratings.

The project began with importing the required libraries, loading the dataset, and setting up the environment for data visualization. We then carried out data cleaning, manipulation, and transformation to prepare the data for accurate and meaningful analysis. This included handling missing values, correcting data types, and creating new features like age groups and normalized performance scores.

Using various graphs and charts, we visualized important trends such as the distribution of employees by department, age, and gender, as well as the relationship between salary and performance. We also identified how factors like education level and years at the company influence employee ratings and promotion history.

Through these insights, HR professionals can better understand workforce dynamics and make informed decisions in areas such as talent management, employee engagement, and retention strategies. This project demonstrates how Python and data visualization tools like Seaborn and Matplotlib can be effectively used to support human resource decision-making with clear, visual storytelling.

In conclusion, HR analytics helps organizations turn employee data into actionable insights. By applying this process, companies can improve performance evaluations, identify potential high-performers, and create a more productive and engaged workforce.

## REFERENCE

- HR Dataset Source  
Kaggle Datasets  
[<https://www.kaggle.com/datasets>](<https://www.kaggle.com/datasets>)
- HR Standards and Practices  
[<https://www.shrm.org/>](<https://www.shrm.org/>)
- Data Handling Library  
[<https://pandas.pydata.org/>](<https://pandas.pydata.org/>)
- Data Visualization Libraries  
Matplotlib: [<https://matplotlib.org/>](<https://matplotlib.org/>)  
Seaborn: [<https://seaborn.pydata.org/>](<https://seaborn.pydata.org/>)
- Machine Learning and Preprocessing  
[<https://scikit-learn.org/stable/>](<https://scikit-learn.org/stable/>)
- People Analytics Course  
Wharton School, Coursera  
[<https://www.coursera.org/learn/wharton-people-analytics>]
- Book Reference  
Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2010). Multivariate Data Analysis (7<sup>th</sup> ed.).