## 

## **Phase-2**

**Student Name:poovendh G**  
**Register Number:** 611823121009  
**Institution:** PSV.college of engineering and technology  
**Department:** BIO MEDICAL ENGINEERING  
**Date of Submission:**09.05.2025  
**GitHub Repository Link:** https://github.com/Dada192270/Evaluating-employee.git

### **Problem Statement**

How can organizations effectively utilize workforce analytics to evaluate employee performance and productivity trends in order to make informed, data-driven decisions that enhance overall organizational efficiency and employee satisfaction?

### **2. Project Objectives**

The objective of this project is to design and implement a **workforce analytics system** that enables the **evaluation of employee performance and productivity trends**, with the goal of using data-driven insights to optimize organizational efficiency, enhance employee engagement, and improve overall business outcomes. Specifically, the project aims to:

1. **Develop a Data Collection Framework**:
   * Establish a comprehensive data collection process, integrating data from multiple sources such as performance reviews, time tracking systems, project management tools, employee surveys, and HR records.
   * Ensure data is captured in real-time and is accurate, relevant, and secure.
2. **Define Key Performance Indicators (KPIs)**:
   * Collaborate with key stakeholders to define clear, measurable KPIs that align with business goals, focusing on areas such as productivity, sales performance, employee engagement, task completion, and quality of work.
3. **Analyze Performance and Productivity Trends**:
   * Utilize advanced analytics techniques to analyze the collected data and identify trends, patterns, and correlations between various factors affecting performance (e.g., time spent on tasks, employee experience, and team collaboration).
   * Create predictive models to forecast future trends, such as identifying high performers or predicting potential disengagement or turnover risks.
4. **Develop Data Visualization Dashboards**:
   * Create user-friendly dashboards to visualize key performance metrics and productivity trends for managers and HR teams. These dashboards will allow for quick identification of underperforming employees, areas of improvement, and overall team productivity.
   * Ensure that data visualizations are clear, actionable, and tailored to different levels of management.
5. **Provide Personalized Feedback and Development Plans**:
   * Design a framework for delivering personalized feedback to employees based on performance insights, helping managers to provide constructive feedback.
   * Use analytics to create targeted training and development programs for employees who may need support in specific areas, enhancing their skills and improving performance.
6. **Implement Continuous Monitoring and Improvement**:
   * Set up a system for continuously monitoring employee performance and productivity, allowing for regular adjustments to performance evaluation models and data collection processes.
   * Implement an ongoing feedback loop where employees and managers can provide input on the effectiveness of the analytics system, ensuring it evolves to meet business needs.
7. **Enhance Employee Engagement and Satisfaction**:
   * Use insights from workforce analytics to identify factors that contribute to high engagement and satisfaction. Implement strategies to improve the work environment and address areas where employees may be feeling disengaged or undervalued.
   * Promote transparency and trust by communicating how performance data is being used to support employee growth and organizational success.
8. **Ensure Compliance and Ethical Data Usage**:
   * Ensure that the collection and analysis of employee data comply with all relevant privacy laws and regulations (e.g., GDPR, CCPA).
   * Establish clear policies and guidelines to ensure data is used ethically and that employee privacy is protected.

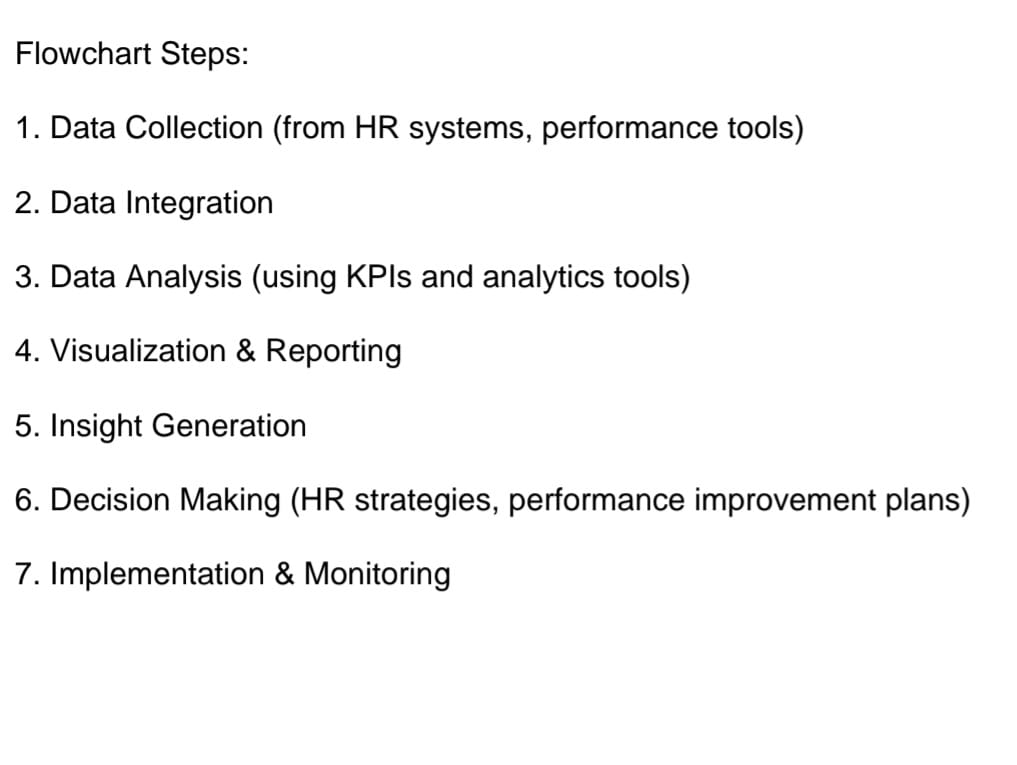
### Expected Outcomes:

* A comprehensive, data-driven approach to employee performance management.
* Increased accuracy and fairness in performance evaluations.
* Clear identification of high-performing and underperforming employees, allowing for targeted interventions.
* Enhanced employee development programs and a more effective training strategy.
* Improved employee satisfaction and engagement, leading to better retention rates.
* Optimized organizational productivity, with better resource allocation and goal alignment.

### Timeline:

1. **Phase 1**: Data Collection Framework (1-2 months)
2. **Phase 2**: KPI Definition and Analytics Model Development (2-3 months)
3. **Phase 3**: Dashboard Creation and Data Visualization (1-2 months)
4. **Phase 4**: Feedback and Development Plan Design (1 month)
5. **Phase 5**: Continuous Monitoring and Improvement Setup (Ongoing)

### **Flowchart of the Project Workflow**



### 

### 

### 

### **4. data description**

#### 1. **Employee Demographics**

* **Employee ID** (unique identifier)
* **Name** (anonymized or coded)
* **Department**
* **Role/Job Title**
* **Location/Region**
* **Date of Joining**
* **Employment Type** (Full-time, Part-time, Contract)

#### 2. **Performance Data**

* **Performance Rating** (e.g., 1–5 scale or descriptive grades)
* **Review Period** (Quarterly, Annual)
* **Manager Feedback**
* **Goal Completion Rate (%)**
* **Peer Review Scores** (if applicable)

#### 3. **Productivity Metrics**

* **Hours Worked** (from time-tracking systems)
* **Task Completion Rate**
* **Project Involvement** (number and type of projects)
* **Time to Task Completion**
* **Utilization Rate** (billable vs. non-billable hours)

#### 4. **Attendance & Behavior**

* **Absenteeism Rate**
* **Tardiness/Delay Records**
* **Overtime Hours**
* **Leave Type and Frequency** (Sick, PTO, Unpaid)

#### 5. **Learning & Development**

* **Training Hours Completed**
* **Courses Attended**
* **Skill Improvement Metrics**
* **Certifications Acquired**

#### 6. **Engagement & Satisfaction**

* **Employee Survey Scores** (e.g., satisfaction, engagement)
* **Participation in Company Events**
* **Retention Likelihood Score** (if modeled)

#### 7. **Attrition/Turnover Data**

* **Resignation Date**
* **Reason for Leaving** (exit interview data)
* **Tenure at Exit**

### **Data Preprocessing**

#### 1. **Data Integration**

* **Combine data** from multiple sources:
  + HRIS (Human Resource Information Systems)
  + Time-tracking tools
  + Performance review systems
  + Survey platforms (e.g., employee engagement surveys)
* Use **Employee ID** as the key to link datasets.

#### 2. **Data Cleaning**

* **Remove duplicates** (e.g., repeated entries due to sync errors).
* **Handle missing values**:
  + Drop columns/rows if data is minimal and missing at random.
  + Impute values using:
    - Mean/median (for numerical data)
    - Mode or constant (for categorical data)
* **Correct inconsistent formatting**:
  + Standardize date formats (YYYY-MM-DD)
  + Normalize department/job titles (e.g., “HR” vs. “Human Resources”)

#### 3. **Data Transformation**

* **Encode categorical variables**:
  + One-hot encoding (for nominal data: department, location)
  + Label encoding (for ordinal data: performance ratings)
* **Create new features** (feature engineering):
  + Tenure = Today’s date - Date of Joining
  + Productivity Ratio = Tasks Completed / Hours Worked
  + Engagement Score = Weighted average of survey results

#### 4. **Outlier Detection and Treatment**

* Use statistical techniques:
  + Z-score or IQR for numerical features (e.g., hours worked)
* Cap or remove outliers depending on business rules.

#### 5. **Normalization/Scaling**

* Apply scaling to numeric fields if using algorithms sensitive to scale (e.g., KNN, SVM):
  + **Min-Max Scaling** or **StandardScaler**

#### 6. **Data Splitting (if modeling is involved)**

* Split into:
  + **Training Set** (70-80%)
  + **Test Set** (20-30%)
* Optionally, create a **validation set** for tuning.

#### 7. **Save Preprocessed Data**

* Export to a clean CSV or database table
* Document preprocessing steps for reproducibility

### **6. Exploratory Data Analysis (EDA)**

#### **1. Understand the Data Structure**

* **Shape of the dataset** (df.shape)
* **Data types** of each column (df.dtypes)
* **Missing values** (df.isnull().sum())
* **Basic statistics** (df.describe())

#### **2. Univariate Analysis**

* **Categorical variables**:
  + Count plots (e.g., Department, Job Title, Location)
  + Pie charts for proportion (e.g., Employment Type)
* **Numerical variables**:
  + Histograms (e.g., Hours Worked, Task Completion Rate)
  + Boxplots (to spot outliers, e.g., Overtime Hours)

#### **3. Bivariate Analysis**

* **Categorical vs Numerical**:
  + Boxplots/Violin plots of Performance Rating across Departments
  + Average Productivity by Job Role
* **Numerical vs Numerical**:
  + Scatter plots (e.g., Hours Worked vs Task Completion Rate)
  + Correlation heatmap (df.corr())
* **Categorical vs Categorical**:
  + Cross-tabulations (e.g., Department vs Performance Tier)

#### **4. Time-Based Trends**

* **Line plots over time**:
  + Performance Rating or Task Completion over months/quarters
  + Overtime/Absenteeism trends by date
* Use Date of Joining or Review Period to group by time

#### **5. Outlier Detection**

* Use boxplots or Z-score methods to detect:
  + Unusually high or low hours worked
  + Abnormal performance ratings
  + Long tenure with low performance

#### **6. Insights and Questions Raised**

* Are certain departments underperforming?
* Is high overtime linked to low satisfaction or attrition?
* Do employees with more training show higher performance?
* Any gender/location bias in performance ratings?

#### **7. Tools for EDA**

* **Python (Pandas, Seaborn, Matplotlib, Plotly)**
* **Excel/Google Sheets**
* **Power BI/Tableau** for interactive dashboards

#### **7** **Data Collection & Integration**

* **HRIS Systems** (e.g., SAP SuccessFactors, Workday, BambooHR)
* **Time Tracking Tools** (e.g., Toggl, Harvest, Clockify)
* **Survey Platforms** (e.g., Google Forms, SurveyMonkey, Qualtrics)
* **APIs & Connectors** (e.g., Zapier, Power Automate, Python scripts with requests or pandas)

### **8. Team Members and Contributions**

|  |  |  |  |
| --- | --- | --- | --- |
| S.no | Team members |  | contribution |
| 1. | SANTHOSH.M |  | Project Manager |
| 2. | YUVARAJ.S |  | DATA ANALYST |
| 3. | SANTHOSH.M |  | MACHINE LEARNING ENGINEERING |
| 4. | **POOVENDHAN.G** |  | DATA ENGINEER |