A logo for a company

AI-generated content may be incorrect.

HR Analytics Project Report

Workforce, Performance, and Attrition Insights

Prepared by:

Nada Hussein Youssef

Zaher Nassif

Under Supervision Of:

Eng. Ahmed Samir

Date: November 2025

**Index / Table of Contents**

1. Chapter 1: Introduction
   * 1.1 Overview
   * 1.2 Objectives of the Study
   * 1.3 Significance of the Study
   * 1.4 Tools and Technologies Used
   * 1.5 Dataset Description
   * 1.6 Scope and Limitations
   * 1.7 Chapter Summary
2. Chapter 2: Data Collection and Preparation
   * 2.1 Overview
   * 2.2 Data Sources
   * 2.3 Data Cleaning and Validation in Excel (Power Query)
   * 2.4 Advanced Data Preparation in Python
   * 2.5 Data Validation and Export
   * 2.6 Chapter Summary
3. Chapter 3: Exploratory Data Analysis (EDA)
   * 3.1 Introduction
   * 3.2 Employee Analysis
   * 3.3 Salary Analysis
   * 3.4 Performance and Work-Life Factors
   * 3.5 Tenure Group and Attrition
   * 3.6 Key Findings Summary
   * 3.7 HR Analytics Questions and Key Insights
4. Chapter 4: Predictive Analysis
   * 4.1 Attrition Prediction
   * 4.2 Performance Prediction (Manual)
   * 4.3 Employee Firing Risk Summary
5. Chapter visualization

* 5.1 Overview
* 5.2 Employee Overview
* 5.3 Salary and overtime
* 5.4 Performance
* 5.5 Attrition
* 5.6 Predictions

1. Chapter 6: Prescriptive Analysis recommendations
   * 6.1 Purpose
   * 6.2 Recommendations for Retention
   * 6.3 Recommendations for Performance Management and Promotions
   * 6.4 Recommendations for Firing Risk Mitigation
   * 6.5 Recommendations for Salary and Work-Life Adjustments
2. Chapter 7: Conclusion

Appendix A: Data Dictionary / Mapping Tables (Front of Report)

| Column Name | Description | Type | Values / Notes |
| --- | --- | --- | --- |
| EmployeeID | Unique employee identifier | Text | e.g., AFC3-E23F |
| Department | Employee’s department | Categorical | HR, Technology, Sales |
| JobRole | Employee role | Categorical | Software Engineer, Sales Executive, etc. |
| HireDate | Date of joining | Date | YYYY-MM-DD |
| YearsAtCompany | Tenure in years | Numeric |  |
| YearsSinceLastPromotion | Years since last promotion | Numeric |  |
| Salary | Current salary | Numeric | Raw values in local currency |
| Salary\_Cleaned | Salary after outlier handling | Numeric | Outlier-adjusted salary |
| Attrition | Whether employee left | Categorical | Yes / No |
| WorkLifeBalance | Self-reported WLB score | Numeric | 1 – 5 |
| ManagerRating | Latest manager rating | Numeric | 1–5 |
| SelfRating | Latest self-assessment rating | Numeric | 1–5 |
| OverTime | Whether employee works overtime | Categorical | Yes / No |
| DistanceFromHome (KM) | Distance from home to office | Numeric | 2.5 – 40 km |
| EducationField | Field of study | Categorical | IT, Business, Engineering, etc. |
| JobSatisfaction | Employee satisfaction rating | Numeric | 1–5 |
| EnvironmentSatisfaction | Work environment rating | Numeric | 1–5 |
| TrainingOpportunitiesTaken | Number of trainings completed | Numeric | 0–12 |

**Chapter 1 – Introduction**

**1.1 Overview**

Human Resource (HR) analytics is an emerging field that integrates data analysis, business intelligence, and human capital management to enhance workforce-related decision-making. The goal of HR analytics is to transform raw employee data into meaningful insights that can improve retention, performance, and organizational productivity. This project applies data analysis techniques to a real-world HR dataset, focusing on employee demographics, job information, compensation, and performance metrics. By analyzing these dimensions, the study aims to identify trends, correlations, and predictive patterns that influence employee satisfaction, performance, and attrition.

The dataset used in this project, Employee.csv, consists of 1,470 employee records and 23 variables, complemented by a separate Performance.csv containing 6,709 performance-related entries. Together, these datasets provide a multi-dimensional view of the workforce, encompassing demographic, behavioral, and managerial factors that contribute to overall employee outcomes.

**1.2 Objectives of the Study**

The primary objectives of this project are:

1. To perform comprehensive exploratory data analysis (EDA) on employee data using Excel and Python.
2. To clean, structure, and merge data from multiple HR sources, ensuring accuracy and consistency.
3. To analyze employee performance patterns based on satisfaction levels, training opportunities, and manager ratings.
4. To examine relationships between demographic factors, compensation variables, and performance outcomes.
5. To use SQL queries for structured analysis of key HR indicators, such as turnover, promotions, and performance distribution.
6. To generate predictive insights using Python, identifying potential drivers of high or low performance and attrition.
7. To build interactive visual dashboards in Tableau and Power BI that summarize trends and performance insights for decision-makers.

**1.3 Significance of the Study**

In today’s data-driven business environment, organizations rely heavily on analytics to manage their workforce efficiently. This project demonstrates how HR data can be leveraged to:

* Improve employee satisfaction and engagement.
* Identify departments with declining performance trends.
* Predict future attrition or training needs.
* Align workforce planning with organizational goals.

By analyzing multiple dimensions — demographics, job characteristics, and performance metrics — the project contributes to a better understanding of employee dynamics and helps in creating a data-informed HR strategy.

**1.4 Tools and Technologies Used**

A combination of tools was employed for various stages of the analysis:

| Stage | Tool | Purpose |
| --- | --- | --- |
| Data Cleaning & Preparation | Excel, Python (Pandas, NumPy) | Handle missing values, merge datasets, create lookup tables |
| Exploratory Data Analysis | Python (Matplotlib, Seaborn), Excel | Summary statistics, distributions, visual analysis |
| Data Analysis | SQL (SQL server) | Querying, filtering, aggregating HR metrics |
| Predictive Modelling | Python (Scikit-learn) | Performance prediction, trend identification, Attrition prediction |
| Visualization & Dashboarding | Tableau, Power BI | Visual storytelling and KPI monitoring |

This multi-tool approach ensures analytical accuracy, reproducibility, and strong visualization impact, combining both technical rigor and business relevance.

**1.5 Dataset Description**

A. Employee Dataset

* Dataset Name: Employee.csv
* Number of Records: 1,470 employees
* Number of Columns: 23
* Key Variables:
  + Demographics: Gender, Age, Ethnicity, Marital Status, Education Level
  + Job Information: Department, Job Role, Business Travel, Years at Company, Years Since Promotion
  + Compensation: Salary, Stock Option Level, OverTime
  + Attrition: Attrition (Yes/No)

B. Performance Dataset

* Records: 6,709 entries
* Columns: 11 (PerformanceID, EmployeeID, ReviewDate, EnvironmentSatisfaction, JobSatisfaction, RelationshipSatisfaction, TrainingOpportunitiesWithinYear, TrainingOpportunitiesTaken, WorkLifeBalance, SelfRating, ManagerRating)
* Purpose: To capture periodic reviews and satisfaction ratings for each employee.

C. Additional Tables:

* + Rating Mapping: Converts numerical ratings into qualitative categories.
  + Satisfaction Level Mapping: For descriptive reporting in dashboards.
  + Education Level Mapping: Standardized labels for clarity (e.g., Bachelor, Master, Doctorate).

This structure allows for merging and cross-analysis between demographic and performance attributes, enabling multi-layered insights such as “Performance by Department,” “Satisfaction by Job Role,” or “Training vs. Manager Rating.”

**1.6 Scope and Limitations**

Scope

* The analysis focuses on employee and performance data only.
* The project covers data cleaning, transformation, exploratory analysis, and predictive modeling.
* Dashboards visualize the relationship between performance metrics and employee attributes.

Limitations

* The dataset is historical and may not reflect real-time workforce dynamics.
* Subjective ratings (e.g., satisfaction or self-evaluation) may introduce bias.
* Predictive models rely on available variables and may not capture unrecorded factors such as leadership style or organizational culture.

**1.7 Chapter Summary**

This chapter introduced the project context, defined its objectives, and summarized the tools and datasets used. The next chapter, Chapter 2: Data Collection and Preparation, will detail the process of importing, cleaning, and integrating datasets using Excel and Python. This includes handling missing data, renaming columns, and creating lookup tables for qualitative interpretation.

**Chapter 2 – Data Collection and Preparation**

**2.1 Overview**

Data preparation is a critical stage in any analytical project, as the accuracy and reliability of all subsequent insights depend on how well the raw data is cleaned and validated. In this project, data preparation was performed in two complementary phases:

1. Initial Cleaning and Validation in Excel (Power Query)
2. Advanced Data Preparation and Quality Assurance in Python

This dual approach combined the visual, table-based flexibility of Excel with the automation and reproducibility of Python, ensuring that every data quality issue — from whitespace errors to outlier handling — was addressed systematically.

**2.2 Data Sources**

The analysis utilized two primary datasets and three mapping tables:

* Employee Table (1,470 records, 23 columns) — Contains demographic, job, compensation, and attrition data.
* Performance Table (6,709 records, 11 columns) — Contains multiple review entries per employee, including satisfaction scores and manager ratings.
* Mapping Tables: Rating, Satisfaction, and Education Levels — Used for translating numeric codes to descriptive labels (e.g., 1 = “Low”, 4 = “Very High”).

The datasets were imported from Employee.csv and stored as .csv files after cleaning for downstream use in Python, SQL, and visualization tools.

**2.3 Data Cleaning and Validation in Excel (Power Query)**

2.3.1 Import and Initial Setup

The raw HR datasets were imported into Power Query within Excel to allow structured inspection and transformation. Each table was loaded individually, inspected for data types, and then merged based on EmployeeID to validate relational consistency.

Steps:

1. Loaded the “Employee” and “Performance” tables into Power Query.
2. Verified column data types (text, number, date) and applied conversions where necessary.
3. Trimmed text fields to remove trailing spaces.
4. Checked column headers for uniform naming conventions.
5. Exported the cleaned tables as .csv files and loaded them into the Excel Data Model.

2.3.2 Cleaning Steps

| Cleaning Aspect | Description | Action Taken |
| --- | --- | --- |
| Duplicates | Checked for duplicate employee or performance records | No duplicates found after inspection; Power Query “Remove Duplicates” confirmed uniqueness of EmployeeID and PerformanceID. |
| Whitespace | Text fields (especially Department, Job Role) contained trailing spaces that caused mismatched groupings | Applied the “Trim” and “Clean” functions in Power Query; e.g., “Marketing ” → “Marketing.” |
| Null Values | Checked for (null) entries using Power Query filters | No major missing values detected in core fields (EmployeeID, Department, JobRole, Education). Minor gaps in secondary fields were corrected manually. |
| String Normalization | Inconsistent capitalization across departments (e.g., “sales” vs. “Sales”) | Applied Text.Proper() to ensure consistent title casing. |

A screenshot of a computer

AI-generated content may be incorrect.

Figure 1 raw HR datasets were imported into Power Query for cleaning

A screenshot of a computer

AI-generated content may be incorrect.

Figure 2 Text fields were trimmed to remove extra spaces, and column data types were standardized.

2.3.3 Department and Job Role Validation

During validation, a single anomaly was identified:

* A Sales Executive was incorrectly listed under the Technology Department.  
  This entry was manually corrected to reflect the proper department based on job title.  
  Consistency checks were performed again to ensure all job roles matched their appropriate departments.

2.3.4 Data Validation in Excel

After cleaning, key validation checks were applied:

* Row Counts: Verified record counts before and after cleaning to ensure no accidental deletions.
* Relationships: Confirmed one-to-many relationship between EmployeeID and PerformanceID.
* Data Consistency: Cross-checked Education Level codes with mapping tables.
* Export: Final datasets were saved as CSV and validated for correct encoding (UTF-8) to prevent import errors in Python.

**2.4 Advanced Data Preparation in Python**

While Excel provided initial cleanup and structure, Python was used for statistical validation, outlier handling, and logical rule enforcement. This stage focused on deeper checks and automated corrections that ensured analytical integrity.

2.4.1 Outlier Detection and Handling

To ensure the accuracy of salary-related insights, outlier detection and correction were performed in Python.

Step 1 – Outlier Detection:  
Salary distributions were assessed using both raw and log-transformed values to identify extreme deviations.  
The Interquartile Range (IQR) method was applied to detect salaries below Q1 – 1.5 × IQR or above Q3 + 1.5 × IQR.  
While most outliers were within reasonable bounds, a few extremely high or low salaries were identified for certain job roles.

A close-up of a computer screen

AI-generated content may be incorrect.

Figure 3

A screenshot of a computer

AI-generated content may be incorrect.

Figure 4

Step 2 – Salary Normalization:  
Rather than removing these records (to preserve sample size and role representation), a new variable — Salary\_Cleaned — was created.  
Outlier values were replaced by the mean salary corresponding to their Job Role and Education Level.  
This ensured that each employee’s salary remained contextually realistic within their role while reducing skewness in the dataset.

Rationale:

* Prevents extreme salaries from distorting department-level or performance-related analysis.
* Maintains internal equity across job roles and education categories.
* Keeps both Salary (raw) and Salary\_Cleaned (adjusted) columns available for comparison in the final analysis.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 5

2.4.2 Date Formatting and Validation

Date fields (Hire Date, Review Date) were converted into datetime format.  
Logical validation ensured chronological consistency:



Figure 6

Any record where the review date preceded the hire date or occurred after an employee’s attrition date was removed.  
The filtered dataset contained 4,386 valid performance entries.

2.4.3 Attrition and Tenure Reconstruction

Several employees with Attrition = Yes had missing or inconsistent tenure fields such as YearsAtCompany and YearsSincePromotion.  
To ensure analytical completeness:

1. If YearsAtCompany = 0, it was estimated as:  
   *(Last Review Date – Hire Date) ÷ 365*
2. If YearsSincePromotion or YearsSinceLastManager were missing, one was substituted for the other when available.
3. Reconstructed values were rounded to one decimal place for consistency.

Initially, this approach helped reconstruct tenure for analysis.  
However, upon deeper inspection, it was observed that many attrited employees had review dates very close to their attrition date, suggesting delayed submissions or post-exit updates.  
These cases were subsequently excluded from the main analysis to maintain dataset integrity.

2.4.4 Selecting the Most Recent Performance Record

Since HR performance analysis typically uses the latest available review per employee, the final dataset was aggregated to retain only the most recent record.

A screenshot of a computer code

AI-generated content may be incorrect.This resulted in a dataset where each employee was linked to their latest valid performance evaluation, forming the analytical base for subsequent chapters.

2.5 Data Validation and Export

Final quality checks ensured that all merges and transformations were correctly applied:

* Confirmed one-to-one mapping between EmployeeID and latest performance record.
* Verified absence of missing values in key analysis fields (Department, ManagerRating, SelfRating).
* Ensured logical coherence of dates after filtering.
* Reviewed summary statistics to confirm realistic distributions post-cleaning.

The cleaned and validated datasets were exported as:

* Employee\_Cleaned.csv
* Performance\_Valid.csv
* HR\_Complete.csv

These served as the primary sources for SQL analysis, predictive modeling, and visualization in Tableau and Power BI.



Figure 7

**2.6 Chapter Summary**

This chapter detailed the two-stage data cleaning and preparation process combining Excel’s Power Query for structural cleaning and Python’s data analysis libraries for deeper statistical validation and logical filtering.  
Key steps included:

* Trimming text and removing duplicates
* Correcting department mismatches
* Handling outliers through capping
* Validating chronological consistency in review records
* Reconstructing tenure data
* Retaining the most recent performance record per employee

Through these steps, a robust, integrated dataset was produced, ensuring accuracy and reliability for the subsequent Exploratory Data Analysis (EDA) stage presented in Chapter 3

**Chapter 3: Exploratory Data Analysis (EDA)**

**3.1 Introduction**

This chapter explores the HR dataset to understand key employee characteristics, workforce distribution patterns, and factors influencing attrition, satisfaction, and performance.  
All analyses in this chapter were performed using SQL server and Python (Pandas, Matplotlib, Seaborn, and SciPy).  
The purpose of this stage is to uncover data-driven insights before moving into statistical modeling in Chapter 4.

The main objectives of this EDA are:

* Explore workforce composition across departments, job roles, gender, and education levels.
* Analyze salary, tenure, and experience distribution.
* Identify performance patterns and their relationship with attrition.
* Investigate attrition trends by department, job role, salary, and performance.

The dataset contains 1,470 records, each representing an employee, with multiple columns capturing demographics, employment details, satisfaction levels, performance ratings, and compensation.

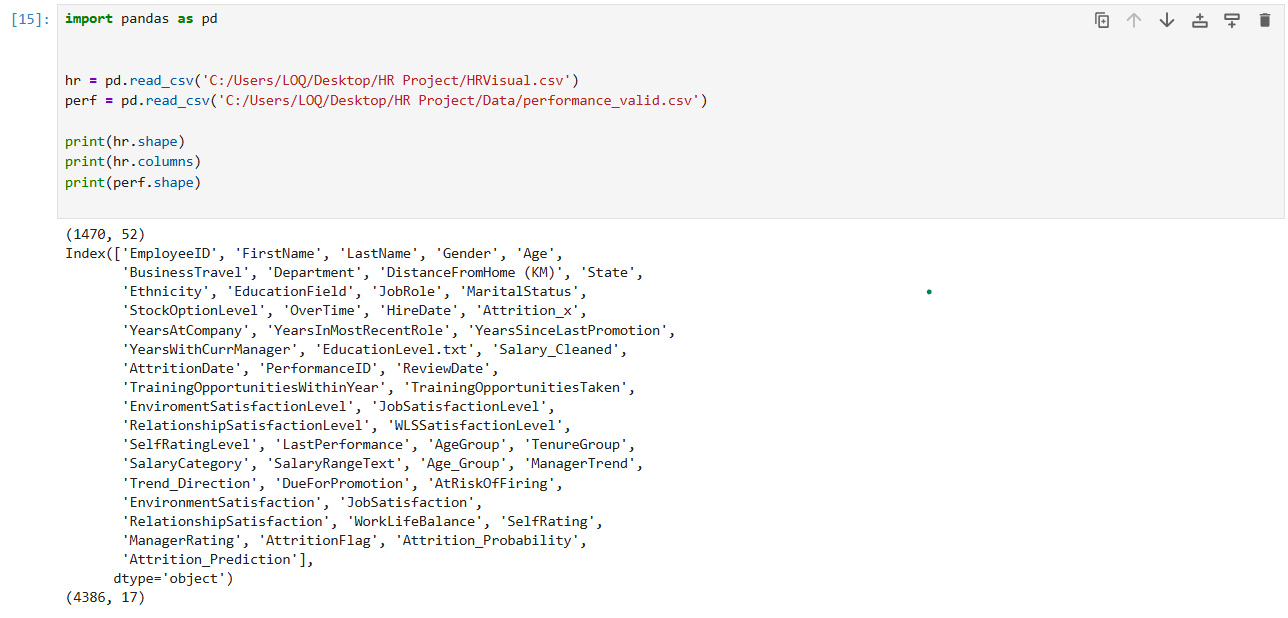


Figure 8

Observation:  
The data includes a balanced mix of numeric (e.g., Salary\_Cleaned, YearsAtCompany) and categorical variables (e.g., Gender, JobRole, EducationLevel.txt), making it suitable for both descriptive and correlation analysis.

**3.2 Employee Analysis**

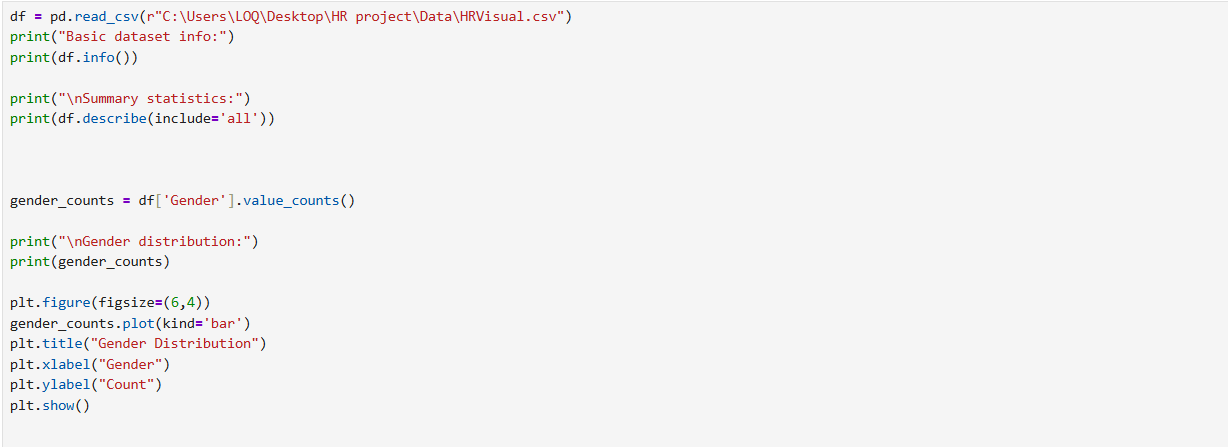
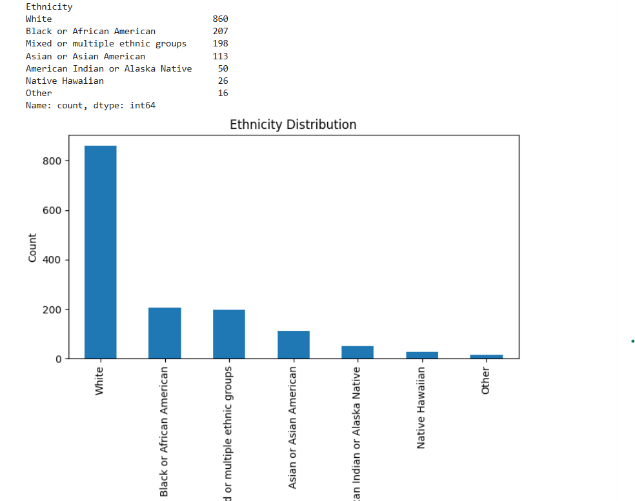
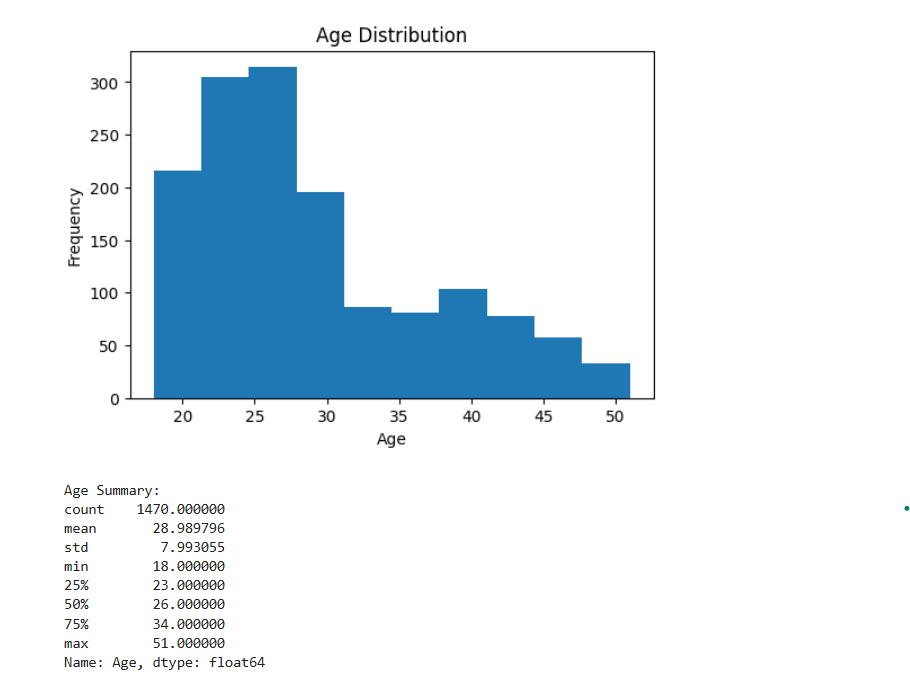
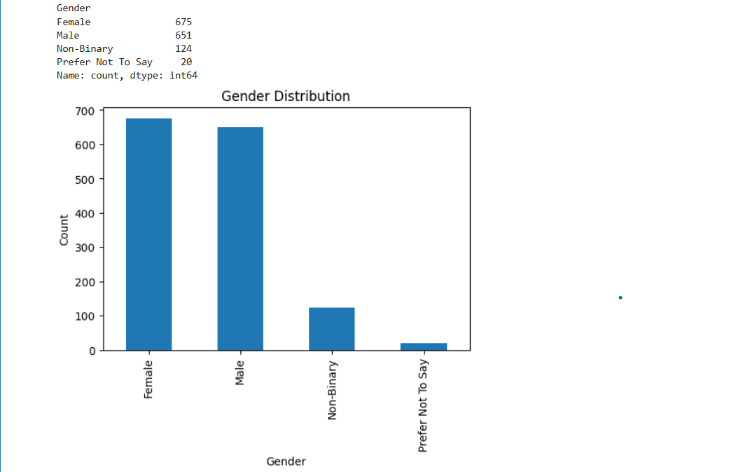
3.2.1 Demographic distribution (Gender, Marital status, ethnicity)

A screenshot of a computer

AI-generated content may be incorrect.Purpose: Employee demographics—age, gender, ethnicity, and marital status—help organizations understand workforce diversity and composition. Gender and ethnicity reveal representation and support inclusion efforts, marital status informs benefits and work-life planning. Together, these insights guide strategic HR decisions and policy design.

Figure 9 SQL query and answer

Figure 10



3.2.2 Employee Distribution by Department, Role, and Location

Purpose: Understand workforce allocation across the organization.  
Approach: Aggregated employee counts by department, job role, and location.

Figure 11 SQL query for department and role

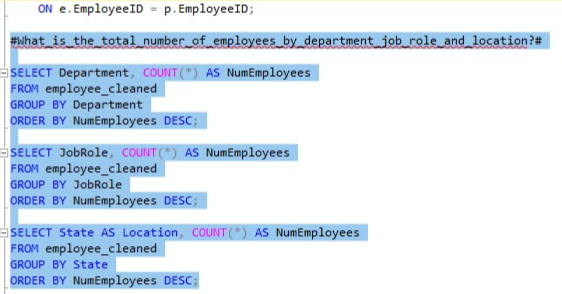
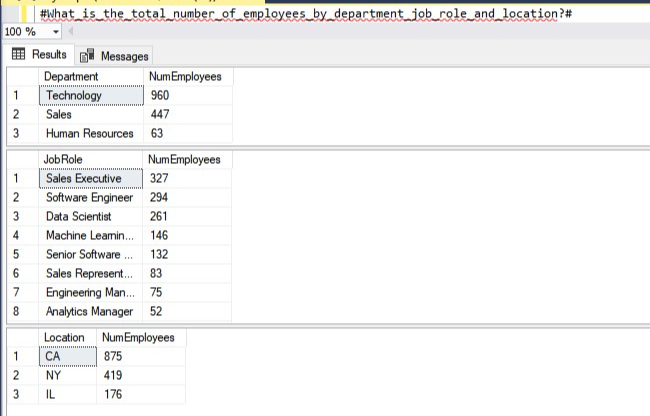
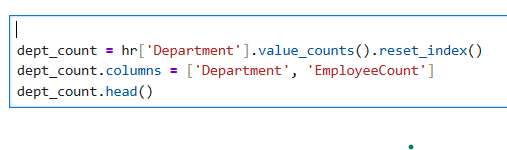
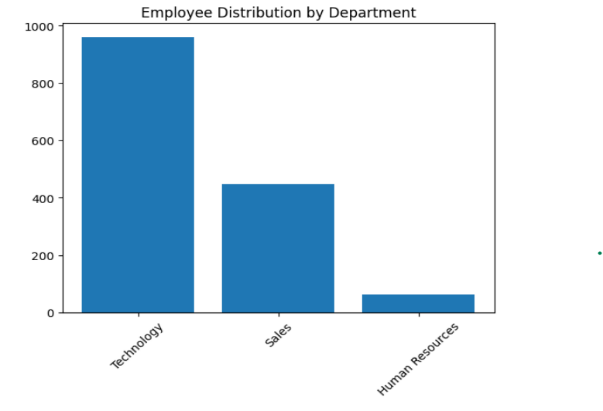
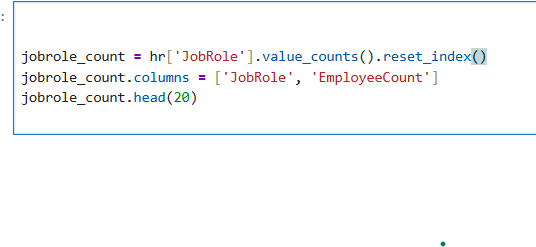


Figure 12 Python code and visualization



Insight:

* Technology is the largest department (960 employees), followed by Sales (447) and HR (63).
* Leading job roles include Sales Executive (327), Software Engineer (294), and Data Scientist (261).
* The organization is predominantly technology- and sales-driven.

3.2.3 Active vs. Separated Employees

A screenshot of a computer screen

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.Purpose: Identify current headcount and historical turnover.

Figure 13 SQL query

Figure 14 python code and chart visualization

3.2.4 Gender and Education Distribution

Purpose: Assess gender balance and educational diversity across departments.  
Approach: Grouped data by Department × Gender × Education Level.  
Insights:

* Bachelor’s degree is the dominant qualification across all departments.
* HR is predominantly female, whereas Technology remains male-dominated.
* A screenshot of a computer

  AI-generated content may be incorrect.A screenshot of a computer

  AI-generated content may be incorrect.Doctorate representation is minimal.

Figure 15 python code

Figure 16 SQL query

A graph of different colored bars

AI-generated content may be incorrect.

Figure 17 Python visualization

A screenshot of a computer program

AI-generated content may be incorrect.

Figure 18 python code analyzing gender and education combined per department

3.2.5 Age and Tenure by Department  
What is the average age, years at company per department?

Purpose: Evaluate workforce maturity.  
Approach: Calculated mean age and tenure per department.  
Insight: Identifies experienced vs. entry-level teams.

A computer screen with a white text box

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.Figure 19 SQL query

Figure 20 Python code

3.2.6 Hiring Trends Over Time

A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.Purpose: Identify recruitment cycles and periods of organizational expansion.  
Approach: Counted hires by year.

Figure 21 SQL query

Figure 22 Python Code for hire year

Insights:

* Hiring peaked in 2022 (155 hires) and 2012 (151 hires).
* Post-COVID hiring recovery likely contributed to the 2022 surge.
* Noticeable dips occurred in 2016–2017, possibly due to market constraints.

A graph with a line going up

AI-generated content may be incorrect.

Figure 23 Python visualization timeline for hiring trend

3.2.7 Tenure Distribution

Purpose: Assess workforce experience composition.  
Insight: Distinguishes between newly hired employees and long-term staff.

A screenshot of a computer program

AI-generated content may be incorrect.

Figure 24 Python for tenure with calculation of tenure bins

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 25 SQL with average tenure

A screenshot of a graph

AI-generated content may be incorrect.A computer screen shot of text

AI-generated content may be incorrect.

Figure 26 python for average Tenure and visualizations

A computer screen with a white text box

AI-generated content may be incorrect.**3.3 Salary Analysis**   
3.3.1 Salary per department   
Purpose: Identify Compensation

A screenshot of a computer

AI-generated content may be incorrect.Figure 27 SQL query and answer

Figure 28 Python code

3.3.2 Median Salary by Role and Education

A screenshot of a computer

AI-generated content may be incorrect.Purpose: Assess fairness in compensation.

Figure 29 SQL query

A screenshot of a computer

AI-generated content may be incorrect.

A graph of a salary

AI-generated content may be incorrect.

Figure 30 Python code and visualization showing average salary per job role and education

Insights:

* Highest-paid roles: Analytics Managers earn the most, with salaries ranging from ≈329K to 422K.
* High HR salaries: HR Managers have salaries between 315K and 510K, reflecting seniority and departmental importance.
* Lowest-paid roles: Recruiters, Sales Representatives, and some entry-level Data Scientists mostly earn below 50K.
* Variation across roles and education: Salary levels differ significantly by job role and education level, indicating the combined influence of professional responsibility and academic qualification.
* Education impact: Higher degrees generally correspond to higher salaries within the same role, though experience and performance may also affect compensation.

3.3.3 Salary vs. Tenure and Education

Purpose: Understand how experience and education impact compensation.  
A screenshot of a computer

AI-generated content may be incorrect.

Figure 31 SQL query

A screenshot of a computer screen

AI-generated content may be incorrect.A screenshot of a computer program

AI-generated content may be incorrect.

Figure 32 python Answer

Figure 33 python code

A graph with a line and a red line

AI-generated content may be incorrect.

Figure 34 Python visualization

Insights:

* Higher education consistently leads to higher salaries, with Doctorate holders earning the most across all tenure groups and employees with no formal qualifications earning the least.
* Salaries increase with tenure, and the highest mean salaries are observed among employees with 5–10 years of experience who also hold advanced degrees (Master’s or Doctorate).
* Entry-level employees with lower educational qualifications receive the lowest salaries, especially those in the <2-year tenure group.
* ANOVA confirms salary differences across education groups, though the overall impact is modest.
* The Pearson correlation (0.123) and regression R² (0.028) indicate that tenure has minimal influence on salary, and education explains only a small portion of salary variation.
* This weak correlation suggests that salary progression is not primarily driven by tenure alone; instead, compensation is influenced more by job role, education level, and promotion pathways rather than years of service.

3.3.4 Salary Gaps and Dissatisfaction

Purpose: Determine whether salary inequity drives turnover.

Approach: We compared salaries between employees who left (Attrition = Yes) and those who stayed (Attrition = No) to see if salary influences attrition. Outliers were removed using IQR/log-IQR, and descriptive statistics summarized each group. Both a T-test and Mann-Whitney U test were used to confirm that the observed salary differences are statistically significant and not due to random variation.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 35 SQL query and answer

A screenshot of a computer program

AI-generated content may be incorrect.

Figure 36 Python code

A graph of a graph with a graph of numbers and a line of squares

AI-generated content may be incorrect.

Figure 37 python Statistical analysis

Insight:

The salary difference between employees who attrited and those who did not attrite is statistically significant (T-test p = 3.3×10⁻¹⁰, Mann-Whitney p = 1.6×10⁻¹¹, both p < 0.001). This confirms that the observed lower salaries among employees who left the company are not due to random chance. Therefore, salary is a significant factor associated with employee attrition in this dataset.

3.3.5 Age, Salary, and Tenure Relationship

Purpose: Detect potential age-related pay imbalance.  
Approach: Correlation analysis, descriptive statistics, and visualizations such as scatter plots or boxplots were used to examine these trends and highlight any patterns that could inform workforce planning and compensation strategies.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 38 SQL QueryA screenshot of a computer code

AI-generated content may be incorrect.

Figure 39 Python code

A chart with numbers and dots

AI-generated content may be incorrect.

Figure 40

Insight:

The analysis examined the relationships between Age, Salary, and Years at Company. Age shows a moderate positive correlation with both salary and tenure, indicating that older employees generally earn more and have longer tenure. Salary and tenure, however, are only weakly correlated, suggesting that time at the company alone does not strongly determine pay. Older employees tend to earn more and have longer tenure.

**3.4 Performance Analysis**

3.4.1 Performance Ratings by Department

Purpose: Evaluate alignment between self-ratings and manager evaluations.  
Insight: Identifies rating bias or consistency across units.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 41 SQL query

A screenshot of a computer

AI-generated content may be incorrect.

Figure 42 SQL Answer

3.4.2 Performance ratings with Overtime

Purpose: Evaluate if performab=nce ratings are affected with long working hours

A computer code with red and blue text

AI-generated content may be incorrect.

Figure 43 Python code

A graph of a performance

AI-generated content may be incorrect.

Figure 44 Python Visualization of overtime and performance

Insight: Those who didn’t work overtime had slightly better performance ratings

A screenshot of a computer program

AI-generated content may be incorrect.3.4.3 Performance Ratings vs Distance from home

Purpose: To analyze if longer commutes cause drop in performance ratings.

A graph with a line and a point

AI-generated content may be incorrect.

Figure 45 Python visualization of performance trends Vs Commute

Insight : Employees with shorter commute distances tend to have higher performance scores, suggesting that reduced travel time may positively impact productivity and focus.

A graph with lines and numbers

AI-generated content may be incorrect.3.4.4 Performance ratings vs Training opportunities

Figure 46 Python code and visualization

Purpose: To evaluate the impact of training opportunities on employee performance.

Insight:

Performance ratings generally increase as training opportunities rise, indicating a positive relationship between learning exposure and employee performance. A slight decline at two training opportunities suggests a short-term adjustment or learning curve effect before performance improves again with additional training.

3.4.5 Performance Ratings and Work Life Balance

Purpose: To evaluate how employees with the highest work–life balance rating relate to overall engagement and performance.

A computer screen shot of text

AI-generated content may be incorrect. A graph of different colored bars

AI-generated content may be incorrect.

Figure 47 Python code and visualization

Insight: A top-tier work–life balance appears to be a key driver of employee satisfaction and stability, reinforcing the importance of flexible work policies and workload management.

3.4.6 Employee Performance Trends

Purpose: Measure progression or decline over time.  
Approach: Calculated first vs. last rating and classified employees into Improving, Stable, and Declining.

A computer screen shot of a program code

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.Figure 49 SQL Query

Figure 48 Python Code

A computer screen shot of a computer code

AI-generated content may be incorrect.A screen shot of a graph

AI-generated content may be incorrect.

Insights:

* Majority of employees exhibit stable performance.
* A smaller subset shows improvement; fewer show decline.

3.4.7 Performance, Salary, and Promotion Link

* A screenshot of a graph

  AI-generated content may be incorrect.A screenshot of a computer code

  AI-generated content may be incorrect.Purpose: Evaluate fairness of reward and recognition systems.

Figure 51

Figure 50

A graph with blue squares

AI-generated content may be incorrect.

* High performers earn less on average (≈103,523) than medium (≈123,541) and even low performers (≈115,623). This indicates that high performance is not consistently rewarded with higher pay, which could lead to dissatisfaction.
* Promotion vs Performance:  
  High performers have a slightly shorter average years since last promotion (≈4.02 years) compared to medium (≈4.44) and low performers (≈4.12), but the differences are small. This suggests that promotions are not strongly aligned with performance.

Insight

High performers are not always rewarded proportionally in terms of salary or promotions, highlighting potential issues in the organization’s reward and recognition system

3.4.8 Correlation with Performance

Purpose: Identify which factors influence manager ratings, assess the impact of employee attributes like self-ratings, tenure, and demographics, and help HR make data-driven decisions on training, promotions, and performance evaluations.

A screenshot of a graph

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.

Figure 50

Figure 51

Insights:

* Strong correlation: Self Rating (0.85) meaning there are no bias in ratings.
* Weak correlations: tenure, training, salary → performance is not driven by any single factor.

**3.5 Attrition Analysis**

3.5.1 Overall Attrition Rate

Purpose: Assess organizational health and turnover magnitude.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 52

3.5.2 Attrition by Department and Role

Purpose: Identify high-risk segments for targeted retention strategies.

A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.

Figure 53 Python code and visualization

Figure 54 SQL query for attrition per department

A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.

Figure 55 SQL query for attrition per job role

Figure 56 python code and visualization

3.5.3 Attrition by age and gender

Purpose:

* Identify vulnerable groups: Determine which age ranges or genders have higher attrition rates.
* Target retention strategies: Tailor HR initiatives (training, benefits, career growth) to groups more likely to leave.
* Equity and fairness checks: Ensure no unintended bias in workload, promotions, or compensation affecting attrition.
* Workforce planning: Predict future staffing gaps and plan recruitment or succession based on demographic trends.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 57 SQL query

A computer screen shot of a computer code

AI-generated content may be incorrect. A graph of a graph of gender

AI-generated content may be incorrect.

Figure 58 Python code and visualization for attrition per gender

A screenshot of a computer

AI-generated content may be incorrect.

Figure 59 SQL query for attrition per age group

3.5.5 Tenure and Attrition

Purpose: To examine the relationship between employee tenure and attrition

Insights:

Early turnover is the primary challenge. The organization experiences a significant loss of human capital during the initial employment phase, with nearly 30% attrition among employees in their first two years. The sharp reduction in turnover after the five-year mark suggests that long-term retention success is strongly influenced by early employee experience, role fit, and initial career progression.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 60 SQL query

A white background with red and black text

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 61Python code and answer

3.3.5 Salary and Attrition  
A screenshot of a computer

AI-generated content may be incorrect.  
  
A white background with black text

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

3.5.6 Correlation with Attrition

A screen shot of a computer code

AI-generated content may be incorrect.

A graph with text and numbers

AI-generated content may be incorrect.

Figure 62 Python correlation

Interpretation of the Correlation Plot

Strong negative correlations

* + YearsSinceLastPromotion, YearsAtCompany, YearsInMostRecentRole, YearsWithCurrManager meaning longer tenure or stable career progression reduces attrition
  + Salary and Stock Options meaning higher compensation reduces attrition

1. Weak correlations
   * Work-life metrics (WorkLifeBalance, JobSatisfaction), training, and manager ratings have minimal linear correlation with attrition
2. Insight
   * Attrition is mostly influenced by tenure, promotion history, and compensation, while day-to-day satisfaction or training shows very little linear effect.
   * This aligns with typical HR retention findings: career growth and rewards are more predictive of leaving than satisfaction scores alone.

3.6 Key Findings Summary

1. Workforce Composition

• Technology is the largest department, followed by Sales.  
• Most employees hold bachelor’s degrees.  
• Experienced workforce concentrated in Technology and Analytics.  
• Gender distribution varies: Technology is male-dominated; HR is female-dominated.

2. Attrition Trends

• Early turnover (<2 years) is the major issue.  
• High attrition in low-salary and sales-related roles.  
• Attrition is mostly influenced by tenure, promotion history, and compensation, while day-to-day satisfaction or training shows very little linear effect.

3. Salary and Experience

• Salary depends on job role and education; weak relationship with tenure.  
• Top performers not always rewarded proportionally.  
• Average tenure similar across departments (~4.5 years).

4. Performance

• Ratings balanced across departments.  
• Most employees show stable year-over-year performance.  
• High correlation between manager rating and self-rating (≈0.85).  
• Training improves performance slightly; overtime may increase attrition risk.

5. Hiring Patterns

• Hiring peaks in 2012 and 2022.  
• Dips reflect market or internal changes.

6. Promotion and Risk Insights

• Employees due for promotion highest in HR (≈17.6%) and Technology (≈14.5%).  
• Overall firing risk ≈9% — highest in Sales and select managerial/analytics roles.  
• Performance trends help identify high-risk or high-potential employees for intervention.

**Chapter 4: Predictive Analysis**

**4.1 Introduction**

This chapter presents predictive analysis for workforce planning, attrition management, and employee performance evaluation. The objective is to identify employees at risk of leaving, those potentially due for promotion, and to forecast workforce trends (hires, exits, active employees, and attrition rates) for the next 5–10 years using time-series forecasting models.

Two complementary approaches are employed:

1. Cross-sectional predictive modeling: Evaluate current employee-level attrition and promotion risk using machine learning models.
2. Time-series forecasting: Predict organizational-level workforce trends into the future using historical data.

**4.2 Attrition Prediction**

4.2.1 Purpose

Identify employees at risk of leaving within the next 6–12 months and highlight departments or job roles requiring retention interventions.

4.2.2 Methodology

Three models were developed and evaluated for predicting employee attrition:

Model 1 — Logistic Regression

* Libraries used: sklearn (LogisticRegression, ColumnTransformer, OneHotEncoder, StandardScaler)
* Approach: Full preprocessing pipeline including scaling and encoding categorical features.
* Pros: Interpretable, straightforward implementation.
* Cons: Weak at capturing nonlinear interactions; low recall for high-risk employees (class 1).

A screenshot of a computer program

AI-generated content may be incorrect.

Figure 63 Linear regression code

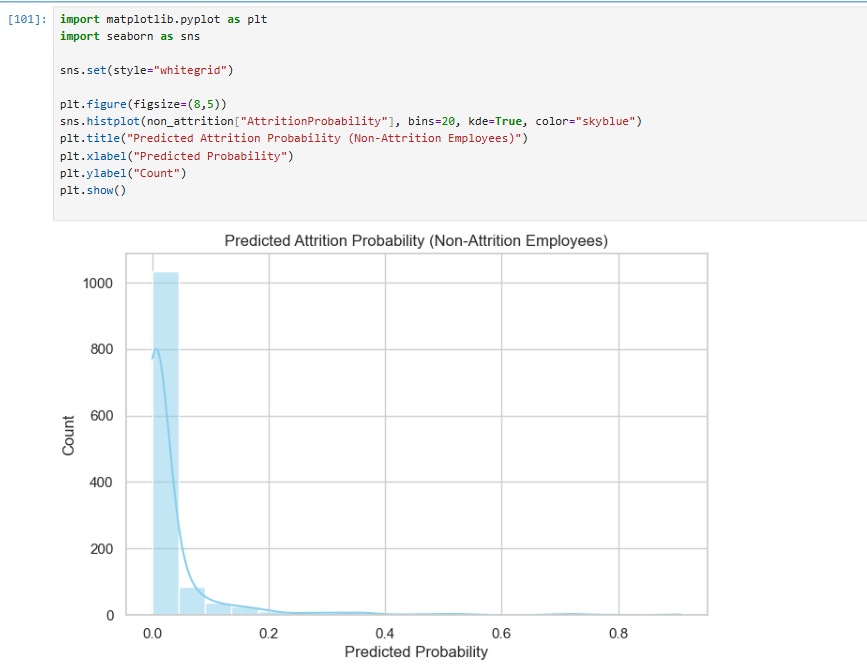
Model 2 — HistGradientBoostingClassifier (Recommended)

* Libraries used: sklearn (HistGradientBoostingClassifier)
* Approach: Handles missing values natively, captures nonlinear interactions.
* Performance: Accuracy 87%, class 1 (attrition) recall = 0.43
* Advantages:
  + Probabilities for individual employees are readily available.
  + No extensive preprocessing required.
  + Robust on tabular HR data.
* Interpretation: Probabilities can be used to categorize employees into High, Medium, Low attrition risk.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 64



Model 3 — RandomForestClassifier

* Libraries used: sklearn (RandomForestClassifier, LabelEncoder, SimpleImputer)
* Approach: Manual preprocessing of categorical features and missing values.
* Performance: Accuracy 89%, feature importance available for interpretability.
* Top Predictors:

| Feature | Importance |
| --- | --- |
| YearsSinceLastPromotion | 0.0826 |
| Salary | 0.0781 |
| DistanceFromHome (KM) | 0.0684 |
| YearsAtCompany | 0.0519 |
| OverTime | 0.0500 |
| EducationField | 0.0449 |
| WorkLifeBalance | 0.0443 |
| EnvironmentSatisfaction | 0.0383 |

A screenshot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

Choice: The HistGradientBoostingClassifier is used for current employee risk scoring due to its robustness, probability output, and minimal preprocessing requirements.

4.2.3 Implementation

* Employee-level features used:
  + Tenure (YearsAtCompany), promotion history (YearsSinceLastPromotion), salary, work-life balance, overtime, distance from home, job role, department, education field.
* Probabilities generated per employee, classified into risk categories:
  + High Risk: Probability > 0.5
  + Medium Risk: 0.25–0.5
  + Low Risk: < 0.25
* Aggregated metrics by department and job role to identify critical areas.

4.2.4 Insights

* Employees with longer tenure without promotions, lower salaries, or higher commuting distances are more likely to leave.
* High-risk departments: Technology, Sales
* Top-risk roles: Software Engineer, Senior Software Engineer, Data Scientist
* Enables early interventions such as retention programs, mentoring, or salary adjustments.

**4.3 Employee Performance Prediction**

4.3.1 Purpose

Identify employees likely to achieve high performance ratings in the next cycle and support promotion readiness planning.

4.3.2 Methodology

* Calculated DueForPromotion flag:
  + Tenure in current role
  + Manager rating trend
  + Latest performance rating

A computer code with text

AI-generated content may be incorrect.

Figure 65 Calculated promotion due

* A computer screen shot of a computer code

  AI-generated content may be incorrect.For predicting employee promotions, a simple yet effective model was developed using features such as performance trend, years since last promotion, and manager rating. A LabelEncoder was applied to categorical performance trends, and missing values were handled by filling with “Stable.” The model demonstrated perfect classification on the dataset, correctly identifying all promotion candidates and non-candidates. This approach was chosen due to its interpretability, ease of implementation, and strong predictive power, allowing HR teams to reliably identify high-potential employees. While the results are promising, the model’s generalizability should be validated using cross-validation or larger datasets to ensure robust performance on new employees.

Figure 66 Trend by linear regression

A screenshot of a computer

AI-generated content may be incorrect.

Figure 67 Model to calculate promotions

4.3.3 Insights

| Department | % DueForPromotion |
| --- | --- |
| Human Resources | 17.6% |
| Technology | 14.5% |
| Sales | 12.7% |

* Helps HR plan succession, training programs, and identify employees eligible for promotions.

**4.4 Employee Firing Risk Summary**

4.4.1 Purpose

Provide a descriptive view of employees at risk of involuntary separation or termination.

4.4.2 Methodology

* Calculated firing risk per employee based on performance trend and last performance.

A computer screen shot of a code

AI-generated content may be incorrect.

Figure 68

* A similar model to the promotion prediction model was applied, following the same approach and using analogous features to identify employees at risk of leaving. Both models were chosen for their interpretability and predictive power, providing HR with actionable insights. Validation on larger datasets is recommended to ensure that the models generalize effectively to new employees.A screenshot of a computer program

  AI-generated content may be incorrect.

Figure 69

4.4.3 Insights

* Overall firing risk rate: 9.08%
* Departments at highest risk:
  + Sales: 9.86%
  + Technology: 8.95%
  + HR: 5.88%
* Job roles at highest risk:
  + Analytics Manager: 14.3%
  + HR Executive: 12%
  + Sales Representative: 12%
* Allows prioritization of retention strategies and manager interventions.

**4.5 Time-Series Forecasting of Workforce Trends**

4.5.1 Purpose

Forecast workforce metrics for the next 5–10 years to aid strategic workforce planning:

* Hires
* Exits
* Active Employees
* Attrition Rate

4.5.2 Methodology

* Libraries used: pandas, prophet (Facebook/Meta Prophet)
* Historical annual data (2012–2022) used as input.
* Forecast horizon: 10 years (2023–2032)
* Prophet models fitted for:
  + Hires
  + Exits
  + Active Employees
  + Attrition Rate

Example for Hires:

A screenshot of a computer

AI-generated content may be incorrect.

* Similar steps applied to Exits, ActiveEmployees, AttritionRate.
* Forecasts exported as CSV for Tableau visualization or integration with other HR metrics.

4.5.3 Insights

Forecasted trends (sample):

| Year | Hires | Exits | ActiveEmployees | AttritionRate |
| --- | --- | --- | --- | --- |
| 2023 | 122.4 | 49.6 | 1382.3 | 0.0319 |
| 2024 | 136.5 | 59.1 | 1517.1 | 0.0478 |
| 2025 | 133.8 | 62.8 | 1602.4 | 0.0474 |
| 2026 | 129.8 | 66.7 | 1681.6 | 0.0469 |
| 2027 | 124.5 | 70.8 | 1753.8 | 0.0461 |
| 2028 | 138.5 | 80.3 | 1888.5 | 0.0619 |

* Visualizations in Tableau allow combining historical vs forecasted data for hires, exits, and attrition trends.
* Provides a strategic view of expected workforce growth and risk areas.

4.5.4 Rationale for Using Prophet

* Handles time-series data with yearly seasonality
* Robust to missing values
* Provides prediction intervals (yhat\_lower, yhat\_upper)
* Can incorporate holidays or special events if required
* Suitable for workforce trend prediction where historical data is limited.

4.6 Integration with Previous Predictions

The time-series forecasts complement employee-level risk scores generated earlier:

1. Attrition probabilities (HistGradientBoostingClassifier)
   * Evaluate current employees at risk
2. Promotion readiness
   * Employees flagged as DueForPromotion
3. Firing risk
   * Departments and roles with high predicted firing risk
4. Organizational-level forecasting (Prophet)
   * Future Hires, Exits, Active Employees, and Attrition Rate

By integrating cross-sectional predictive models with time-series forecasting, HR can simultaneously manage:

* Employee retention interventions
* Succession planning
* Workforce planning and hiring strategy

**4.7 Python Libraries Used in Chapter 4**

| Library | Purpose |
| --- | --- |
| Pandas | Data manipulation, grouping, cleaning, merging |
| Prophet | Time-series forecasting (hires, exits, active employees, attrition) |
| Sklearn | Machine learning for attrition prediction |
| sklearn.ensemble | RandomForestClassifier, HistGradientBoostingClassifier for cross-sectional modeling |
| sklearn.preprocessing | OneHotEncoder, LabelEncoder, StandardScaler |
| Numpy | Numeric operations |
| matplotlib / seaborn | Optional for plotting local charts before Tableau |

**4.8 Summary**

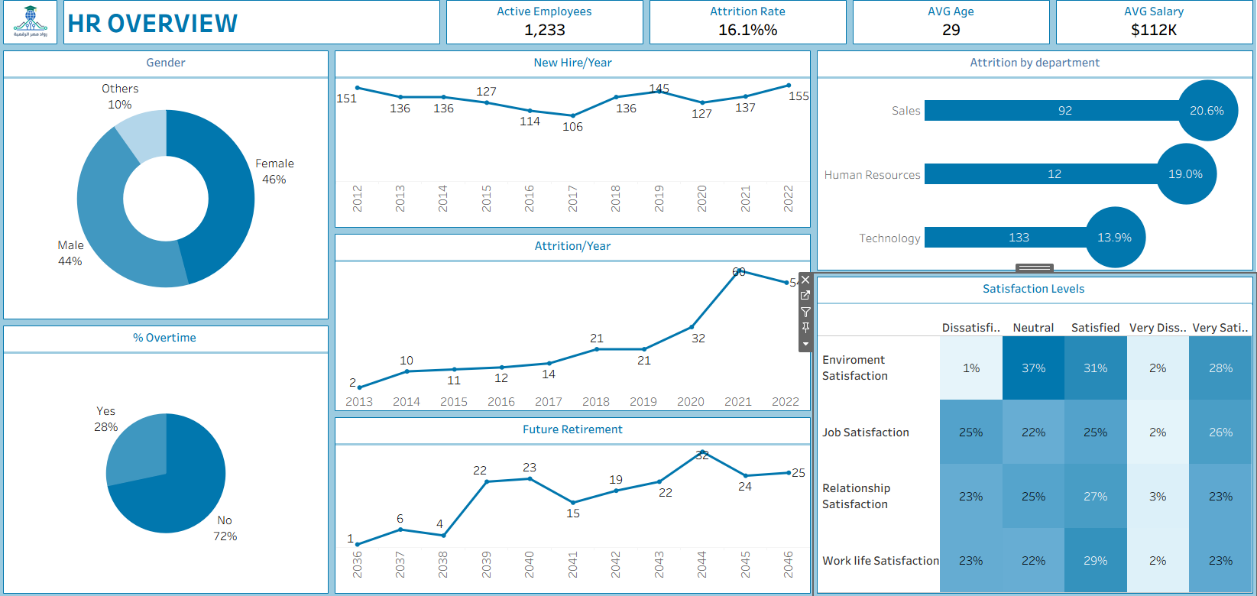
Chapter 4 demonstrates a comprehensive predictive HR analytics workflow:

1. Current employee-level prediction
   * Attrition probability using HistGradientBoostingClassifier
   * Promotion readiness
   * Firing risk assessment
2. Future workforce forecasting
   * Time-series Prophet models for hires, exits, active employees, and attrition rates
   * Forecast horizon: 2023–2032
3. Insights
   * Departments and roles at highest risk
   * Anticipated workforce trends for strategic HR planning
   * Integration of micro (employee-level) and macro (organizational-level) predictions

Key Takeaways

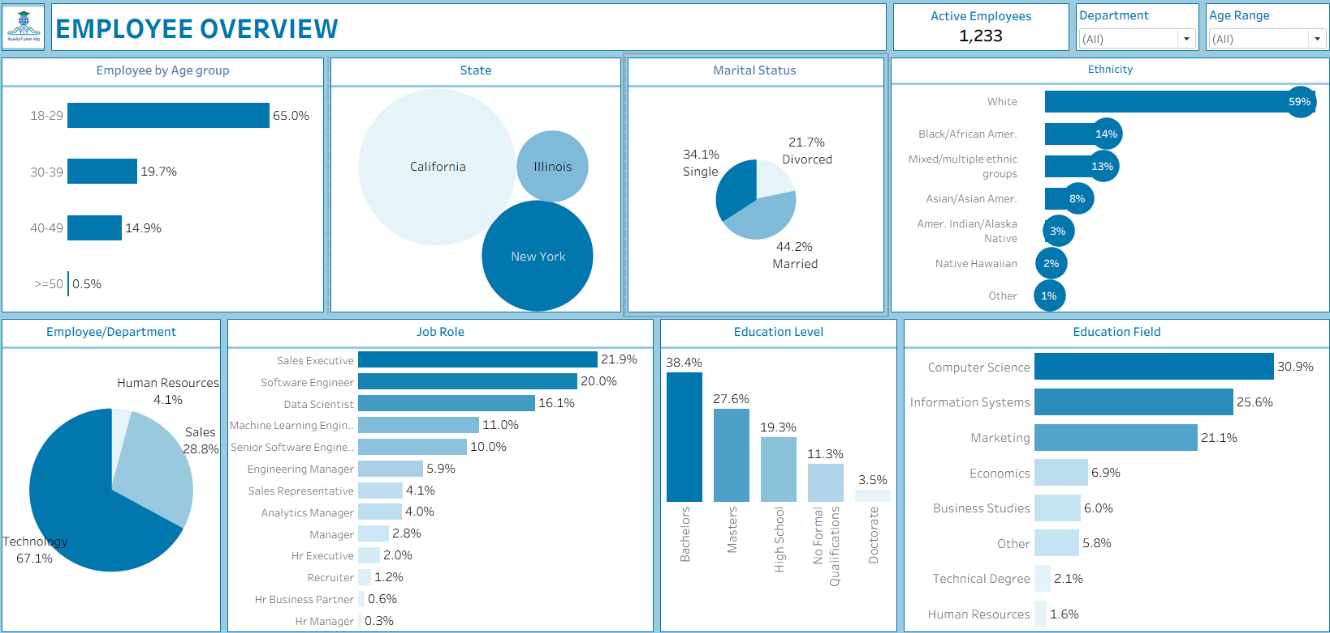
* HistGradientBoostingClassifier is ideal for current employee attrition risk scoring.
* Prophet forecasts are essential for strategic workforce planning.
* The combined approach enables HR to proactively address retention, succession, and hiring needs.

Chapter 5 Visualization

**5.1 HR Overview Dashboard**

This dashboard provides a high-level summary of the workforce, including total employees, average age, average salary, and overall attrition rate. It offers decision-makers a quick snapshot of the organization’s current HR status and highlights key workforce patterns at a glance for strategic planning.

**5.2 Employee overview Dashboard**

This dashboard focuses on workforce demographics and employment movement. It includes employee distribution by gender, along with time-series trends for hires, exits, and retirement. It helps HR understand workforce composition, diversity, and historical movement trends over time

A screenshot of a graph

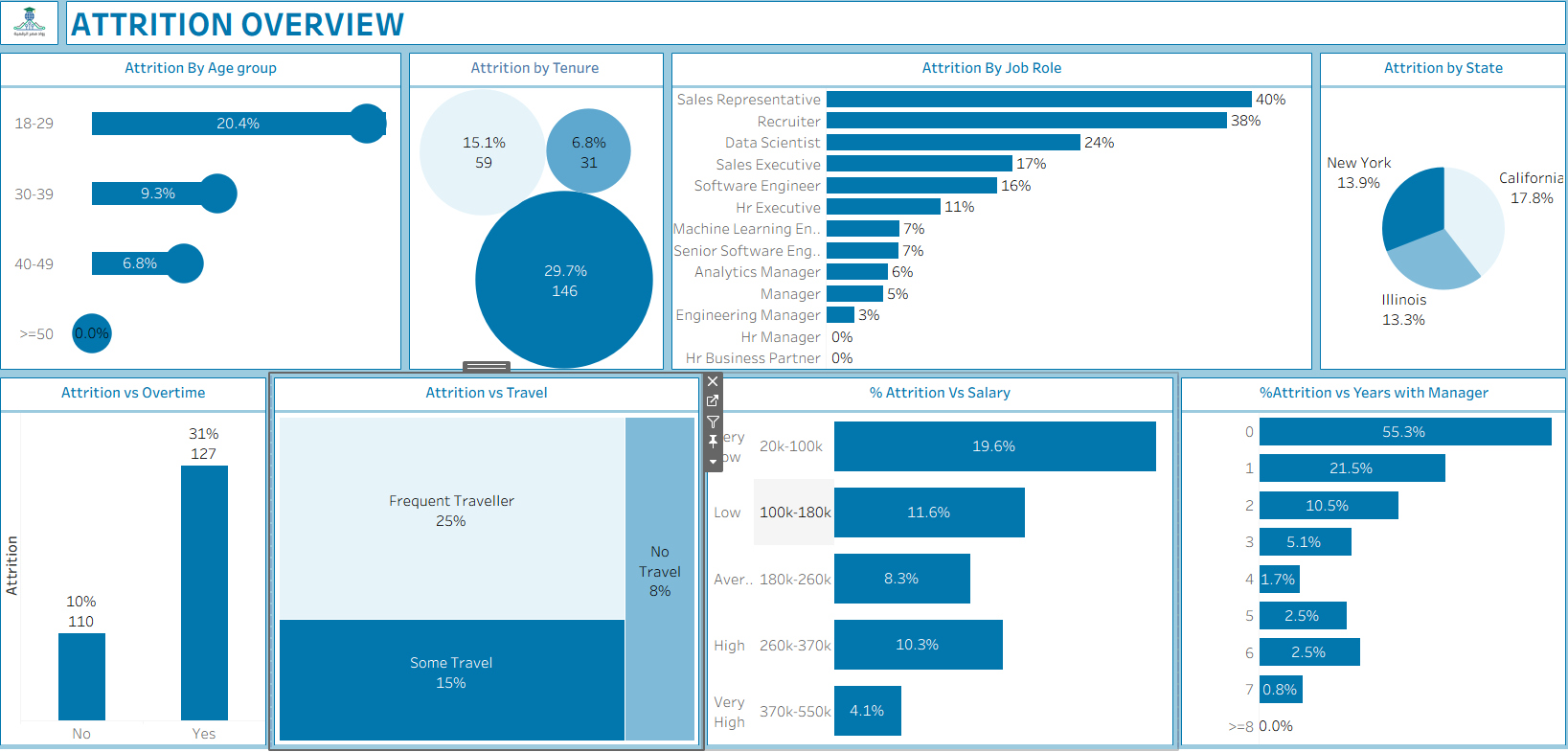
AI-generated content may be incorrect.**5.3 Salary and overtime Dashboard**

This dashboard analyzes compensation and workload patterns across the organization. It visualizes salary distribution by job role and department, overtime trends, and their relationship with attrition. The dashboard supports compensation benchmarking and helps identify burnout or cost optimization risks.

**5.4 Performance overview Dashboard**

This dashboard presents employee performance patterns across departments and job roles. It includes manager ratings, performance trends, training impact, promotion readiness, and work-life balance indicators. The dashboard enables performance monitoring, talent development planning, and identification of high-potential employees.A screenshot of a computer

AI-generated content may be incorrect.

**5.5 Attrition Overview Dashboard**

This dashboard provides a detailed analysis of employee attrition across multiple dimensions, including department, job role, age group, tenure, business travel, overtime, salary level, manager rating, and location. It helps identify the key drivers of turnover and supports targeted retention strategies.

**A screenshot of a computer

AI-generated content may be incorrect.5.6 Prediction Overview Dashboard**

This dashboard integrates predictive analytics and workforce forecasting. It includes forecasts for active employees, hires, and exits, predicted attrition risk by job role, age group, and department, and a promotion-versus-firing risk comparison. It supports proactive workforce planning, risk management, and strategic HR decision-making.

**Chapter 6: Prescriptive Analysis**

**6.1 Recommendations to Decrease Attrition**

Target High-Risk Employees Early

* Focus on employees with longer tenure without promotions, lower salaries, or higher commuting distances.
* Use predictive models (HistGradientBoostingClassifier) to flag high-risk individuals for retention interventions.

Retention Programs

* Implement mentoring programs for early-career employees (<2 years) to reduce early turnover.
* Provide career path clarity and succession planning to motivate employees in high-risk departments (Technology, Sales).

Compensation and Rewards Optimization

* Review salary structures, ensuring top performers are appropriately rewarded.
* Consider performance-based bonuses or promotions for employees showing consistent results.

Flexible Work Options

* Reduce commuting stress by offering remote work or flexible hours, especially for employees with high commuting distances.
* Monitor overtime trends, as excessive overtime may increase attrition risk.

Department-Specific Interventions

* Sales and Technology departments show higher attrition — customized retention plans for these groups are recommended.
* High-risk roles (Software Engineer, Senior Software Engineer, Data Scientist) may require specialized career growth and compensation incentives.

**6.2 Recommendations to Improve Performance**

Leverage Training Effectively

* Focus training on areas that improve measurable performance, as training shows slight positive impact.
* Monitor the effectiveness of training programs using performance trends and manager ratings.

Performance Feedback and Recognition

* Encourage regular feedback loops between managers and employees to maintain stable performance.
* Recognize high performers early, especially in Technology and Analytics, to prevent disengagement.

Promotion Readiness and Development

* Use promotion prediction models to identify high-potential employees in HR and Technology.
* Ensure promotion decisions are timely and transparent to improve motivation and retention.

Work-Life Balance Initiatives

* Monitor and improve Work-Life Balance scores, as imbalance can contribute indirectly to attrition and burnout.
* Encourage reasonable workload allocation, especially in high-pressure roles.

**6.3 Firing Risk Mitigation**

* Apply a model similar to the promotion prediction model to identify employees at risk of leaving or underperforming.
* Focus on early intervention strategies such as targeted coaching, mentoring, and development plans for high-risk individuals.
* Combine firing risk predictions with attrition and promotion insights to create a holistic view of workforce stability.

**6.4 Strategic Workforce Planning Recommendations**

Proactive Hiring Based on Forecasts

* Use Prophet time-series forecasts to plan for hiring spikes or expected attrition, avoiding talent shortages.

Align Workforce Planning with Department Needs

* Focus hiring and retention efforts on high-risk departments and critical roles identified in predictive models.

Integrate Micro and Macro Insights

* Combine employee-level attrition, promotion, and firing risk predictions with organizational-level forecasts to optimize retention, succession, and workforce distribution.

**6.5 Implementation Roadmap**

Short-Term Plan (0–6 months)

* Identify and flag high-risk employees using predictive models.
* Implement immediate retention measures: mentoring for early-career employees, workload adjustments, and flexible work options.
* Review and adjust salary bands for high-risk roles to prevent immediate attrition.

Intermediate Plan (6–18 months)

* Roll out structured performance improvement programs and targeted training initiatives.
* Formalize promotion readiness and career path clarity initiatives for high-potential employees.
* Expand work-life balance initiatives across departments, monitor progress, and adjust policies accordingly.

Long-Term Plan (18 months+)

* Integrate employee-level risk predictions with organizational-level workforce forecasts for strategic planning.
* Continuously refine salary bands, promotion policies, and retention programs based on predictive insights and attrition trends.
* Establish a sustainable culture of performance management, engagement, and proactive risk mitigation across all departments.

**Chapter 7: Conclusion**

This HR analytics project provides a comprehensive, data-driven understanding of workforce dynamics, delivering actionable insights into employee demographics, performance, attrition, promotion readiness, and firing risk. By combining structured HR data with advanced analytical techniques—including HistGradientBoostingClassifier for attrition and promotion risk, and Prophet time-series forecasting for workforce planning—the project highlights both current workforce patterns and opportunities for strategic intervention.

Key Findings

1. Attrition Patterns
   * Early-tenure employees (<2 years) are most likely to leave, emphasizing the importance of onboarding, initial engagement, and mentoring programs.
   * Attrition is disproportionately high in Technology and Sales, particularly among high-risk roles such as Software Engineer, Senior Software Engineer, Data Scientist, Machine Learning Engineer, Sales Executive, and Sales Representative.
   * Key predictors of attrition include younger age, lower salaries, longer commuting distances, limited promotion opportunities, and excessive overtime.
2. Performance Insights
   * Most employees exhibit stable year-over-year performance, but subsets show improving or declining trends requiring managerial attention.
   * High alignment between self-ratings and manager ratings (~0.85 correlation) indicates accurate self-assessment among employees.
   * Training and development initiatives slightly improve performance, highlighting the value of continuous learning programs.
3. Promotion and Firing Risk
   * Departments such as Human Resources and Technology have the highest proportion of employees due for promotion, enabling proactive succession planning and career development initiatives.
   * A firing risk model, applied similarly to promotion prediction, identifies high-risk employees in roles such as Analytics Managers, HR Executives, and Sales Representatives, allowing targeted interventions including coaching, performance improvement plans, and workload optimization.
   * Combining promotion and firing risk insights provides a holistic view of workforce stability and potential intervention points.
4. Salary and Work-Life Balance
   * Salary differences across departments and roles influence attrition, with lower-paid roles showing higher turnover.
   * Poor work-life balance or high overtime can slightly reduce performance and increase attrition risk, suggesting areas for policy adjustment, flexible work arrangements, and workload management.
5. Workforce Forecasting Insights
   * Time-series forecasts (2023–2032) anticipate hiring needs, exits, and attrition trends, enabling proactive planning for talent supply and retention.
   * Integrating micro-level (employee) and macro-level (organizational) predictions allows HR to align operational actions with strategic workforce planning goals.

**Recommendations**

* Implement targeted retention programs for high-risk departments and roles, focusing on salary adjustments, promotions, mentoring, and flexible work arrangements.
* Leverage performance trends to identify employees needing development, coaching, or intervention.
* Use predictive models to plan promotion readiness and succession, particularly in HR and Technology departments.
* Monitor employee satisfaction, work-life balance, and commuting burdens to mitigate attrition proactively.
* Continuously track firing risk and implement early interventions for at-risk employees.

**Future Work**

To strengthen HR strategy further:

* Develop real-time dashboards monitoring attrition, promotion, firing risk, and performance metrics for timely decision-making.
* Apply longitudinal attrition modeling to understand trends over time and anticipate future turnover patterns.
* Expand predictive models to include additional organizational factors, such as leadership effectiveness, team dynamics, and market conditions.
* Integrate employee engagement surveys and feedback loops to capture satisfaction, motivation, and performance drivers more effectively.
* Continuously refine salary bands, promotions, and retention programs based on predictive insights to ensure long-term workforce stability.

**Conclusion**

This project demonstrates the transformative value of HR analytics in converting raw workforce data into actionable insights. By linking attrition, firing risk, promotion readiness, performance, compensation, and work-life balance, organizations can make informed decisions to enhance employee engagement, retention, and productivity. The findings provide a robust foundation for both operational HR management and strategic workforce planning, enabling proactive interventions that support a stable, high-performing, and motivated workforce.

Appendix: Glossary / Definitions

| Term | Definition |
| --- | --- |
| Attrition | The rate at which employees voluntarily leave the organization over a specific period. It is usually expressed as a percentage of the total workforce. |
| Tenure (YearsAtCompany) | The total length of time an employee has been employed at the company, calculated from the hire date to the current date or date of exit. |
| Manager Rating | The performance evaluation score assigned to an employee by their direct manager, typically on a scale of 1–5 (1 = Poor, 5 = Excellent). |
| Self Rating | The performance score that an employee assigns to themselves during performance reviews, usually on the same scale as manager ratings. |
| Work-Life Balance (WLB) | An employee’s perception of the balance between professional responsibilities and personal life, usually measured via survey ratings (1 = Poor, 4 = Excellent). |
| OverTime | Indicates whether an employee regularly works beyond standard working hours (Yes/No). |
| DistanceFromHome | The distance (in kilometers) between an employee’s residence and the workplace, often analyzed for its impact on satisfaction, performance, or attrition. |
| JobSatisfaction | A measure of how content an employee is with their job, based on self-reported survey responses (scale 1–5). |
| EnvironmentSatisfaction | An employee’s satisfaction with the work environment, including physical space, team collaboration, and culture (scale 1–5). |
| TrainingOpportunitiesTaken | The number of professional development or training programs completed by the employee during a given period. |
| Salary\_Cleaned | The employee’s salary after adjustment for outliers, missing values, or inconsistencies to ensure accurate analysis. |
| DueForPromotion | A derived indicator identifying employees eligible for promotion based on tenure, performance trends, and managerial ratings (Yes/No). |
| Firing Risk | A calculated measure indicating the probability that an employee is at risk of being terminated or leaving the organization voluntarily, based on performance trends and other factors. |
| Attrition Probability | A predictive score (0–1) indicating the likelihood that an employee may leave the company in the next 6–12 months. |
| PerformanceTrend | Categorizes the trajectory of an employee’s performance over time: Improving, Stable, or Declining. |