

HR Analysis Project Report

Data Structure

Overview of the Dataset: The data provided consists of multiple related tables

I. Table: Employee

- **Columns and Data Types:**

- EmployeeID (object): **Unique identifier for employees.**
- FirstName (object): **Employee's first name.**
- LastName (object): **Employee's last name.**
- Gender (object): **Gender of the employee.**
- Age (int64): **Age of the employee.**
- BusinessTravel (object): **Frequency of business travel.**
- Department (object): **Department in which the employee works.**
- DistanceFromHome (int64): **Distance from home in kilometers.**
- State (object): **State of residence.**
- Ethnicity (object): **Employee's ethnicity.**
- Education (int64): **Educational level identifier ([links to the Education Level table](#)).**
- EducationField (object): **Employee's field of education.**
- JobRole (object): **Employee's job role.**
- MaritalStatus (object): **Marital status of the employee.**
- Salary (int64): **Annual salary of the employee.**
- StockOptionLevel (int64): **Stock option level of the employee.**
- OverTime (object): **Whether the employee works overtime.**
- HireDate (object): **Date of hire.**
- Attrition (object): **Indicates if the employee has left the company.**
- YearsAtCompany (int64): **Number of years the employee has been at the company.**
- YearsInMostRecentRole (int64): **Number of years in their current role.**
- YearsSinceLastPromotion (int64): **Number of years since their last promotion.**
- YearsWithCurrManager (int64): **Number of years working with their current manager.**

2. Table: Performance Rating

- **Columns and Data Types:**

- PerformanceID (object): **Unique identifier for performance records.**
- EmployeeID (object): **Links to the Employee table.**
- ReviewDate (object): **Date of performance review.**
- EnvironmentSatisfaction (int64): **Satisfaction level with the environment (1-5).**
- JobSatisfaction (int64): **Job satisfaction level (1-5).**
- RelationshipSatisfaction (int64): **Satisfaction with workplace relationships (1-5).**
- TrainingOpportunitiesWithinYear (int64): **Training sessions offered within a year.**
- TrainingOpportunitiesTaken (int64): **Training sessions attended.**
- WorkLifeBalance (int64): **Work-life balance level (1-5).**
- SelfRating (int64): **Employee's self-assessment rating (1-5).**
- ManagerRating (int64): **Manager's rating of the employee (1-5).**

3. Table: Education Level

- **Columns and Data:**

- EducationLevelID (int): **Identifier for the education level.**
 - 1: No Formal Qualifications
 - 2: High School
 - 3: Bachelors
 - 4: Masters
 - 5: Doctorate
- EducationLevel (object): **Description of the education level.**

4. Table: Rating Level

- **Columns and Data:**

- RatingLevelID (int): **Identifier for the performance rating.**
 - 1: Unacceptable
 - 2: Needs Improvement
 - 3: Meets Expectation
 - 4: Exceeds Expectation
 - 5: Above and Beyond
- RatingLevel (object): **Description of the rating.**

5. Table: Satisfaction Level

- **Columns and Data:**

- SatisfiedLevelID (int): **Identifier for satisfaction level.**
 - 1: Very Dissatisfied

- 2: Dissatisfied
- 3: Neutral
- 4: Satisfied
- 5: Very Satisfied
- SatisfiedLevel (object): **Description of the satisfaction level.**

Relationships and Keys

1. EmployeeID is the primary key in the Employee table and is referenced in the Performance Rating table.
2. Education in the Employee table links to EducationLevelID in the Education Level table.
3. Satisfaction and rating levels are qualitative indicators used in performance review and satisfaction evaluation.

Data Cleaning Steps for the Dataset

1. Handle Missing Values:

- Identify and handle any missing or null values in all tables.
- Fill missing numeric values with appropriate imputed values (mean, median, or mode) or use domain knowledge.
- For categorical columns, fill missing values with the most frequent category or a placeholder (e.g., "Unknown").

2. Validate Data Types:

- Ensure all columns have the correct data types:
 - Convert dates (e.g., HireDate, ReviewDate) to datetime format.
 - Ensure numeric columns (e.g., Age, Salary, YearsAtCompany) are integers.
 - Ensure categorical columns (e.g., Gender, Department, EducationLevel) are stored as string.

3. Standardize Categorical Values:

- Ensure consistent formatting for categorical columns (e.g., Gender, BusinessTravel, MaritalStatus) by standardizing case (e.g., title case or uppercase).
- Replace inconsistent or misspelled values with the correct terms.

4. Remove Duplicate Records:

- Check for and remove duplicate rows in all tables, especially in Employee and Performance Rating tables.

5. Validate Key Relationships:

- Ensure foreign keys (e.g., EmployeeID, Education, SatisfiedLevelID) in the related tables match valid entries in their respective primary tables.

6. Address Outliers:

- Identify and address outliers in numeric columns like Salary, DistanceFromHome, and Age.
- Use domain knowledge or statistical methods (e.g., interquartile range) to handle outliers.

7. Normalize Column Names:

- Standardize column names across tables for readability and consistency (e.g., replace spaces with underscores, use consistent casing).
- Rename columns with unclear names to more descriptive ones.

8. Validate Unique Identifiers:

- Ensure EmployeeID in the Employee table is unique.
- Validate that PerformanceID and other IDs (e.g., EducationLevelID, RatingLevelID) are unique within their respective tables.

9. Convert Numeric Levels to Descriptive Labels:

- Replace numeric codes in columns (e.g., Education, EnvironmentSatisfaction, JobSatisfaction, RatingLevelID) with their descriptive labels using the associated mapping tables.

10. Align Data Ranges:

- Verify the ranges of numeric columns (e.g., Age, Salary, YearsAtCompany) to ensure they fall within realistic and expected values.
- Ensure rating-related columns (e.g., SelfRating, ManagerRating, WorkLifeBalance) are within their expected ranges (1-5).

11. Check for Consistency Across Tables:

- Cross-verify that each EmployeeID in the Performance Rating table exists in the Employee table.
- Ensure all values in foreign key columns (e.g., Education, SatisfiedLevelID) have valid matches in the corresponding lookup tables.

12. Drop Irrelevant Columns:

- Remove columns that do not add value to the analysis or are redundant (e.g., FirstName, LastName).

13. Format Dates:

- Standardize all date columns (e.g., HireDate, ReviewDate) to a consistent format (e.g., MM-DD-YYYY).

14. Create Derived Fields:

- Add calculated columns such as Tenure (difference between ReviewDate and HireDate).
- Categorize continuous variables (e.g., Age into age groups, Salary into salary bands).

15. Document Assumptions and Transformations:

- Maintain a log of all cleaning steps and assumptions made during the data cleaning process for transparency.

Main KPIs from this data:

1. Demographics and Workforce Distribution

- **Age Group Distribution:** Percentage of employees in each AgeGroup.
- **Distance Status Breakdown:** Percentage of employees in each DistanceStatus category.
- **State Representation:** Number or percentage of employees by State.
- **Education Level Distribution:** Percentage of employees by EducationLevel.
- **Salary Band Distribution:** Percentage of employees in each SalaryBand.

2. Financial Metrics

- **Average Monthly Salary:** Average of MonthlySalary across employees.
- **Salary Band Analysis:** Average Salary for each SalaryBand.

3. Tenure and Retention

- **Hire Year Trend:** Number of employees hired by HireYear.
- **Tenure Category Breakdown:** Percentage of employees in each TenureCategory.
- **Average Years Since Last Activity:** Average value of YearsSinceLastActivity.

4. Satisfaction and Engagement

- **Overall Employee Satisfaction:** Average EmployeeSatisfactionIndex.
- **Satisfaction Breakdown:** Distribution across EnvironmentSatisfactionLevel, JobSatisfactionLevel, and RelationshipSatisfactionLevel.
- **Work-Life Balance Levels:** Distribution across WorkLifeBalanceLevel.

5. Performance

- **Average Self-Rating:** Average value of SelfRatingLevel.
- **Average Manager Rating:** Average value of ManagerRatingLevel.
- **Performance Rating Analysis:** Average PerformanceRating across the workforce.
- **Rating Gap Analysis:** Distribution of positive, negative, and neutral RatingGap values.

6. Training and Development

- **Training Utilization Rate:** Average value of TrainingUtilization.
- **Training Effectiveness:** Correlation between TrainingUtilization and PerformanceRating.

7. Trends and Patterns

- **Monthly Hiring Trends:** Employee counts grouped by HireMonth.
- **Promotion and Role Activity:** Correlation between YearsSinceLastActivity and PerformanceRating.

Calculated Columns Added:

From Employee Table:

- **AgeGroup**

=IF([@Age]<25,"Under25",IF([@Age]<35,"25-34",IF([@Age]<45,"35-44",IF([@Age]<55,"45-54","55+"))))

- **DistanceStatus**

=IF([@[DistanceFromHome (KM)]]<=10,"Near-by",IF([@[DistanceFromHome (KM)]]<=30,"Far","Very Far"))

- **State**

=IF([@StateAbbreviation]="IL","Illinois",IF([@StateAbbreviation]="CA","California","New York"))

- **EducationLevel**

=IF([@Education]=1,"NoFormalQualifications",IF([@Education]=2,"HighSchool",IF([@Education]=3,"Bachelors",IF([@Education]=4,"Masters","Doctorate"))))

- **SalaryBand**

=IF([@Salary]<50000,"Entry",IF([@Salary]<100000,"MidLevel",IF([@Salary]<150000,"Senior","Executive")))

- **MonthlySalary**

=[@Salary]/12

- **HireYear**

=YEAR([@HireDate])

- **HireMonth**

=TEXT([@HireDate],"mmm")

- **TenureCategory**

=IF([@YearsAtCompany]<2,"New",IF([@YearsAtCompany]<5,"Mid Tenure","Veteran"))

- **YearsSinceLastActivity**

=MAX([@YearsSinceLastPromotion],[@YearsInMostRecentRole])

From Performance Rating Table:

- **EnvironmentSatisfactionLevel**

=IF([@EnvironmentSatisfaction]=1,"VeryDissatisfied",IF([@EnvironmentSatisfaction]=2,"Dissatisfied",IF([@EnvironmentSatisfaction]=3,"Neutral",IF([@EnvironmentSatisfaction]=4,"Satisfied","Very Satisfied"))))

- **JobSatisfactionLevel**

=IF([@JobSatisfaction]=1,"VeryDissatisfied",IF([@JobSatisfaction]=2,"Dissatisfied",IF([@JobSatisfaction]=3,"Neutral",IF([@JobSatisfaction]=4,"Satisfied","Very Satisfied"))))

- **RelationshipSatisfactionLevel**

=IF([@RelationshipSatisfaction]=1,"VeryDissatisfied",IF([@RelationshipSatisfaction]=2,"Dissatisfied",IF([@RelationshipSatisfaction]=3,"Neutral",IF([@RelationshipSatisfaction]=4,"Satisfied","Very Satisfied"))))

- **EmployeeSatisfactionIndex**

=([@EnvironmentSatisfaction] + [@JobSatisfaction] + [@RelationshipSatisfaction])/3

- **TrainingUtilization**

=[@TrainingOpportunitiesTaken]/[@TrainingOpportunitiesWithinYear]

- **WorkLifeBalanceLevel**

=IF([@WorkLifeBalance]=1,"Poor",IF([@WorkLifeBalance]=2,"Fair",IF([@WorkLifeBalance]=3,"Good",IF([@WorkLifeBalance]=4,"Very Good","Excellent"))))

- **SelfRatingLevel**

=IF([@SelfRating]=1,"Unacceptable",IF([@SelfRating]=2,"NeedsImprovement",IF([@SelfRating]=3,"Meets Expectation",IF([@SelfRating]=4,"Exceeds Expectation","Above and Beyond"))))

- **ManagerRatingLevel**

=IF([@ManagerRating]=1,"Unacceptable",IF([@ManagerRating]=2,"NeedsImprovement",IF([@ManagerRating]=3,"Meets Expectation",IF([@ManagerRating]=4,"Exceeds Expectation","Above and Beyond"))))

- **PerformanceRating**

=([@SelfRating] + [@ManagerRating])/2

- **RatingGap**

=[@ManagerRating]-[@SelfRating]

Key Analysis Questions for Pivot Tables:

Demographics Analysis:

1. What is the age distribution across departments?
2. How does gender diversity vary by job role?
3. What is the educational background distribution?
4. How does ethnicity representation vary across departments?

Compensation Analysis:

5. What is the average salary by department and job role?
6. How does salary correlate with years of experience?
7. What is the pay gap analysis across different demographics?
8. How do stock option levels relate to salary bands?

9. What are the Salary trends across age groups?

Performance & Satisfaction Metrics:

10. What is the relationship between performance ratings and satisfaction levels?

11. How does work-life balance correlate with performance?

12. What is the impact of training on performance ratings?

13. How do manager ratings compare to self-ratings?

Attrition Analysis:

14. What is the attrition rate by department?

15. How does satisfaction level impact attrition?

16. What factors correlate most strongly with attrition?

17. How does distance from home impact attrition?

Career Development:

18. What is the average time between promotions?

19. How does education level impact career progression?

20. What is the relationship between training and promotion?

21. How do years with current manager impact satisfaction?

Work Patterns:

22. How does business travel impact satisfaction?

23. What is the distribution of overtime across departments?

24. How does distance from home affect work-life balance?

25. What is the relationship between overtime and performance?

26. How do different departments compare in work-life balance scores?

Proposed Enhancements for Employee Performance and Work Environment Development

First: Salary and Compensation Improvements

I. Salary Structure

- Develop a fair salary system based on:
 - Academic qualifications
 - Years of experience
 - Performance levels
 - Job responsibilities
- Conduct bi-annual salary reviews.

- Reduce gender pay gaps.

2. Rewards and Incentives System

- Create a performance-based rewards system.
- Offer additional incentives for outstanding employees.
- Introduce bonuses for successful projects.
- Establish a profit-sharing program for long-term employees.

Second: Professional Development and Training

1. Training Programs

- Design tailored training programs for each department.
- Provide courses in both technical and soft skills.
- Develop clear career development paths.
- Encourage continuous learning opportunities.

2. Career Advancement

- Establish clear promotion plans.
- Offer cross-department mobility opportunities.
- Launch mentoring programs for new employees.
- Support postgraduate studies and professional certifications.

Third: Work Environment Enhancement

1. Work-Life Balance

- Implement flexible working hours.
- Allow remote work options.
- Minimize overtime hours.
- Provide additional paid leave for exceptional employees.

2. Physical Work Environment

- Improve workspace designs.
- Create comfortable rest areas.
- Upgrade equipment and technology.
- Enhance lighting and ventilation systems.

Fourth: Enhancing Job Satisfaction

1. Communication and Transparency

- Hold regular meetings with management.
- Implement a consistent feedback system.
- Ensure transparency in decision-making.
- Maintain open communication channels with senior management.

2. Recognition and Appreciation

- Launch a monthly "Employee of the Month" program.
- Celebrate individual and team achievements.
- Recognize long service years.
- Distribute certificates of appreciation and awards.

Fifth: Health and Wellness

I. Health Programs

- Provide comprehensive health insurance.
- Offer fitness programs.
- Include free psychological consultations.
- Conduct health awareness campaigns.

2. Psychological Well-Being

- Organize stress management programs.
- Offer counseling sessions.
- Arrange social and recreational activities.
- Support mental health balance.

Sixth: Leadership Development

I. Manager Training

- Equip managers with effective leadership skills.
- Train on team management techniques.
- Focus on effective communication.
- Teach conflict resolution strategies.

2. Administrative Skills Enhancement

- Develop decision-making abilities.
- Strengthen strategic planning skills.
- Train on performance management.
- Encourage motivational and guiding techniques.

Seventh: Performance Evaluation System

I. Evaluation Process Improvement

- Ensure objective and fair evaluations.
- Define clear evaluation criteria.
- Provide continuous feedback.
- Create individual development plans.

2. Progress Monitoring

- Conduct monthly goal reviews.
- Measure key performance indicators (KPIs).

- Regularly update goals.
- Implement continuous improvement plans.

Eighth: Talent Retention

1. Loyalty Programs

- Introduce long-term rewards.
- Create employee savings programs.
- Offer additional insurance benefits.
- Provide exclusive privileges.

2. Talent Management

- Identify top-performing employees.
- Design special development plans.
- Offer leadership opportunities.
- Assign unique and challenging project.

By : Aml Wajeh Mamdouh Abuzeid.