HUMAN RESOURCE DATA CLEANSING & EXPLORATORY

HR project analysis of employee data to understand workforce demographics and turnover patterns.

DATA CLEANSING

1. Data Quality Check

The original dataset contains 22,214 employee records that require cleaning and transformation before analysis.

2. Duplicate Check

No duplicate records were found in the dataset, confirming data integrity for the analysis.

3. Data Type Conversion

All date formats (birthdate, hire\_date, and termdate) were successfully converted to standard DATE format, with empty termdate values set to NULL to represent active employees.

4. Data Filtering

After applying age and employment status filters, 17,482 valid records remain representing employees over 18 years old who are currently active in the organization.

Exploratory Data Analysis

Demographic Analysis:

1. What is the gender breakdown of the employees?
2. What is the race/ethnicity breakdown of the employees?
3. What is the age distribution of the employees?

3.1. What is the gender distribution of employees by age group?

Employment Details:  
4. How many employees work at headquarters versus remote locations?  
5. What is the average length of employment for employees who have been terminated?  
6. How does the gender distribution vary across departments and job titles?  
7. What is the distribution of job titles across the company?

Turnover & Retention:  
8. Which department has the highest turnover rate?  
9. What is the distribution of employees across cities and states?  
10. How has the employee count changed over time based on hire dates?  
11. What is the tenure distribution for each department of terminated employees?

**Technical Stack:** MySQL, Data Validation, Statistical Analysis, Trend Analysis