HR DATA SET ANALYSIS – SQL & PANDAS

Project Portfolio Documentation

Prepared By

Abdul Rahman

Aspiring Data Analyst

Technologies Used

- **Python (Pandas, Jupyter Notebook)**
- **MySQL**
- **SQLAlchemy**
- **Excel (Raw Data Sources)**

Project Category: Data Cleaning & SQL Analytics

Date: August 2025

PROJECT OVERVIEW

This project focuses on analyzing HR data to derive insights about workforce distribution, employee attendance, performance factors, and salary trends. The goal is to support data-driven HR decision-making through SQL queries, data cleaning using Pandas, and visualization support.

OBJECTIVES

- Clean and transform raw HR datasets
- Load structured data into MySQL database
- Perform SQL-based analytical queries
- Identify workforce patterns such as
    - Attendance status
    - Team and manager-wise performance
    - Job level distribution
    - Gender diversity
    - Salary averages

DATASETS USED

Dataset Name		Format
Employees	-	Excel
Departments	-	Excel
Attendance	-	Excel

TECHNOLOGIES & SKILLS APPLIED

Skill	Usage
Python – Pandas	Data cleaning & transformation
SQLAlchemy	DB connection & table creation
MySQL (SQL Queries)	Aggregation & HR analytics
Jupyter Notebook	Code execution & reporting

KEY SQL QUERIES & ANALYTICS PERFORMED

The following SQL queries were executed on HR datasets stored in MySQL to extract meaningful business insights:

Workforce Distribution

- Total number of employees in each department
- Number of employees reporting to each manager

Attendance Insights

- Present vs Absent count per employee
- Average attendance days by department
- Filter attendance by specific month & year

Organizational Role Insights

- Employees belonging to HR department
- Job-level distribution across teams

Salary-Based Analysis

- Salary comparison across different job levels
- Identify employees earning above department average

These queries helped in supporting HR decision-making

- Performance evaluation
- Workforce planning
- Team-level productivity

Benefits to Business

- Improved decision-making for manager evaluation
- Identification of high-performing teams
- Better planning of workload & hiring
- Ensuring fair salary allocation

ATTACHMENTS & REFERENCES

Includes all project artifacts supporting this analysis:

- Jupyter Notebook (Python Execution)
- MySQL SQL Queries
- Query Result CSV Files
- Raw HR Datasets – Excel Files
- Documentation – PDF & DOCX formats

These files validate the transformation steps and analytical results.

CONCLUSION

This project successfully cleaned and structured HR data from raw Excel files into a MySQL relational database using Python and Pandas.
SQL-based analytical queries helped uncover important workforce insights such as attendance trends, manager performance, job distribution and salary variations.

This work demonstrates strong capabilities in:

- ETL (Extract → Transform → Load)
- SQL Analytics & Business Insight Generation
- Practical Data Handling & Reporting

FUTURE SCOPE

To improve the project further:

- Add Power BI dashboard for visual insights
- Automate query execution and data refresh
- Integrate real-time HR data APIs
- Extend analysis with predictive modeling for attrition analysis