ETL Pipeline Project Report

1. Introduction

This project implements an ETL pipeline that extracts employee data from a CSV file, transforms it using Python and Pandas, and loads it into a SQLite database.

2. Tools & Technologies

Python, Pandas, SQLite, Jupyter Notebook, Matplotlib

3. ETL Architecture

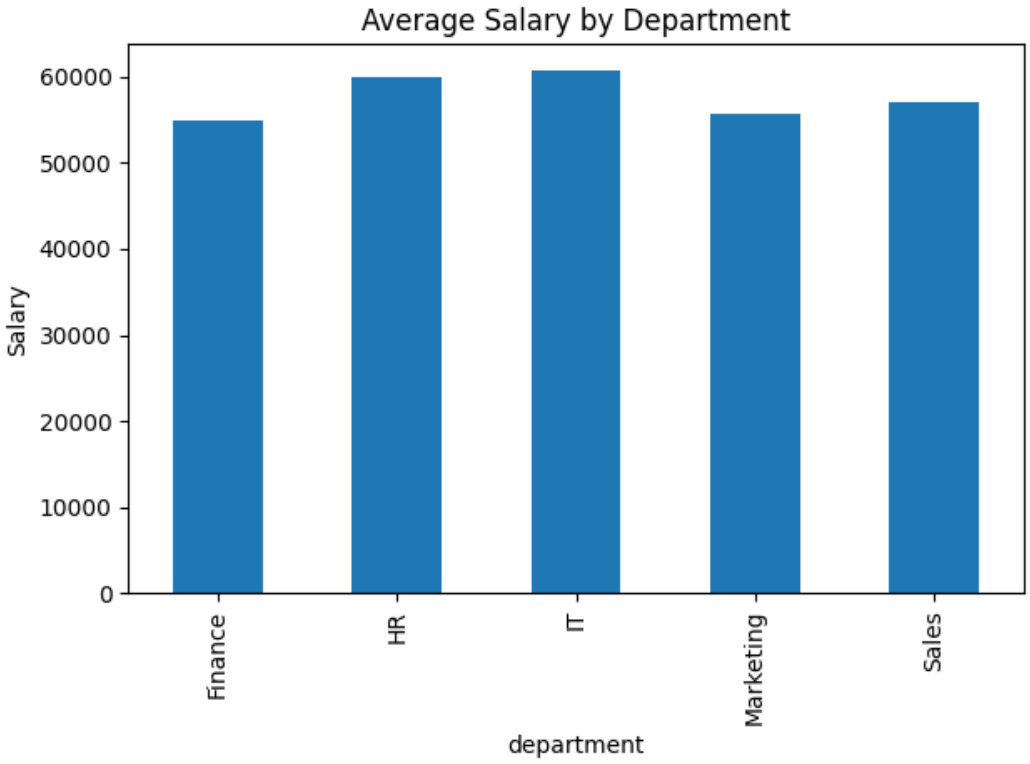
CSV → Extract → Transform → Load → Database

4. Data Overview

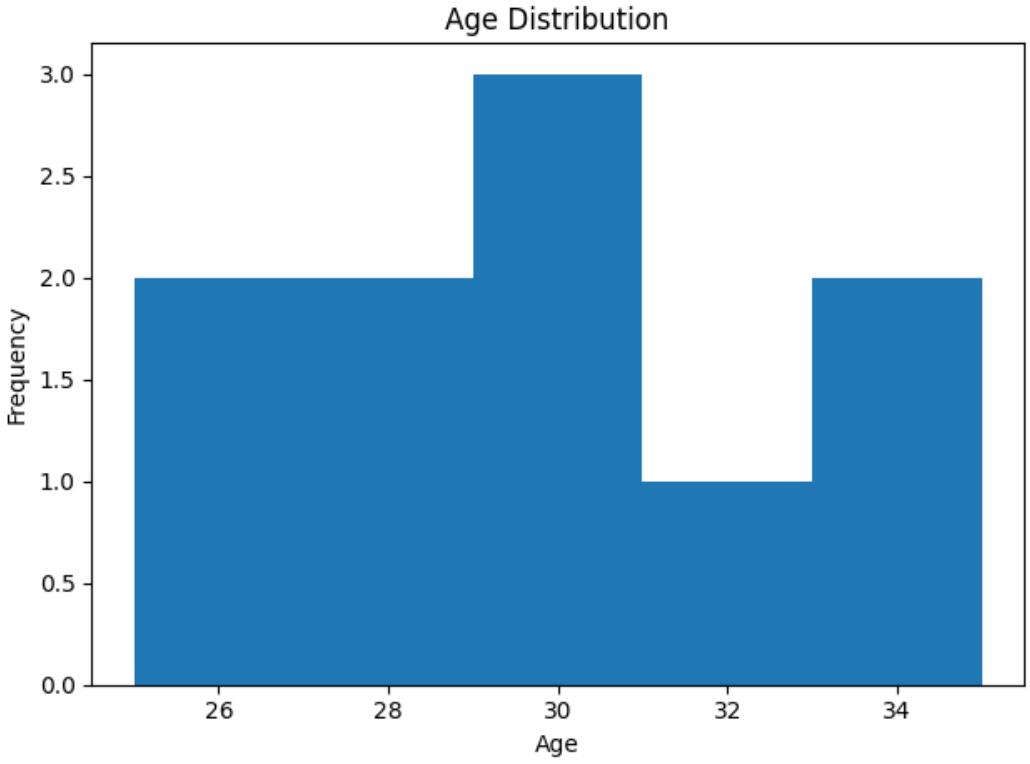
id	name	age	department	salary	joining_date
1	Alice Johnson	25	Sales	50000	2023-01-01
2	Bob Smith	30	Marketing	55000	2023-02-01
3	Charlie Davis	29	HR	60000	2023-03-01
4	David Lee	28	Finance	52000	2023-04-01
5	Eva Brown	27	IT	56500	2023-05-01
6	Frank Miller	35	Sales	72000	2023-06-01
7	Grace Kim	31	IT	65000	2023-07-01
8	Henry Wilson	29	Finance	58000	2023-08-01
9	Ivy Adams	33	Marketing	56500	2023-09-01
10	Jack Turner	26	Sales	49000	2023-10-01

5. Visualizations

Average Salary by Department



Age Distribution



6. Conclusion

The ETL pipeline successfully processes and stores clean data, demonstrating core data engineering concepts suitable for academic evaluation.