Lab 3 - Data Engineering & EDA Submission

Student Name: Ali Cihan Ozdemir

Student ID: 9091405

Date: 2026-02-06

GitHub Repository

https://github.com/alicih4n/WranglingWorkshop.git

Project Summary

1. Cloud Database & Advanced Architecture: Established a PostgreSQL connection on Neon. Designed a multi-table schema with 'employees' and 'departments' connected via Foreign Keys to simulate enterprise data complexity. Created a Python script (lab3_sdg.py) to generate 500 synthetic employee records linked to 5 departments, utilizing 20% 'dirty data' logic.

2. Data Wrangling & Cleaning: Developed a Jupyter Notebook to extract data from multiple SQL tables. Used Pandas to inspect quality, impute missing salaries, standardize job titles, and resolve logic errors. Performed a SQL-style Inner Join in Pandas to enrich employee records with departmental metadata.

3. Feature Engineering & Scaling: Created 'start_year' and 'years_of_service' features. Applied Z-Score standardization (StandardScaler) to salary data to prepare it for potential Neural Network applications.

4. Visual Intelligence: Built advanced visualizations including a Grouped Bar Chart for salary trends and a FacetGrid Heatmap for departmental salary distributions using the joined dataset.

Contribution Validation

Please note that this project was completed entirely by Ali Cihan Ozdemir.
Group partner Roshan was absent from class and did not contribute to this lab submission.