Steps to Complete Your Project

1. Understanding the Data

- Load the datasets (d1.xlsx, d2.xlsx, d3.json) and inspect them to understand their structure.
- Identify key attributes related to job satisfaction, mental health, and salary differences.
- Check for missing values, duplicate records, and inconsistencies.

2. Storing Data in a Database

- Choose a **database system**:
    - Use **MongoDB** to store d3.json (semi-structured JSON).
    - Use **PostgreSQL/MySQL** to store structured Excel data (d1.xlsx and d2.xlsx).
- Write scripts in **Python** to load data:
    - Use pymongo to insert JSON data into MongoDB.
    - Use pandas and SQLAlchemy to insert Excel data into PostgreSQL.

3. Preprocessing and Transformation

- Convert salary to a common currency (e.g., USD).
- Standardize job roles and company sizes.
- Convert categorical values (e.g., experience levels) into numerical formats if needed.
- Handle missing values using appropriate imputation techniques.

4. Data Analysis

- Compare **salary differences** between public and private sector jobs.
- Analyze **mental health trends** (if relevant data exists).
- Evaluate **job satisfaction** by sector, experience level, or remote work ratio.
- Identify **correlations** between salary, job satisfaction, and mental health.

5. Data Visualization

- Use **Matplotlib, Seaborn, and Plotly** to create:
    - Salary distribution across sectors.
    - Job satisfaction levels comparison.
    - Mental health trends (if applicable).
    - Correlation heatmaps.
- Develop an **interactive dashboard** using **Dash or Streamlit**.

6. Report and Documentation

- Write a structured **IEEE-format report** including:
    - Introduction, methodology, findings, and conclusions.
- Include **separate visualizations** in the report.

Tools Required

- **Python** (for data processing, analysis, and visualization)
- **MongoDB** (for semi-structured data storage)
- **PostgreSQL/MySQL** (for structured data storage)
- **Pandas, NumPy** (for data manipulation)
- **Matplotlib, Seaborn, Plotly** (for visualization)
- **Dash/Streamlit** (for interactive dashboards)
- **Jupyter Notebook/PyCharm** (for coding environment)