**Name:** Gowtham Loganathan

**Milestone 2: Data, Methods and Analysis Reporting**

**Project Title:** Exploring Data Science Job Trends through Data Visualization

**Project Type:** Individual Project

**Data Identification:** I am using the publicly available "Data Science Job" dataset from Kaggle. The dataset contains job postings for data science roles, including job titles, locations, salaries, required skills and company names. Since the dataset is freely accessible, there are no privacy concerns or IRB requirements. I successfully downloaded the dataset as a CSV file and loaded it for analysis.

**Methods and Initial Processing:** I have started the exploratory data analysis (EDA) and data pre-processing steps. The dataset was examined for missing values, duplicates, and inconsistencies. Below are the key steps taken so far:

1. **Data Cleaning:**
   * Checked for missing values in key columns (job title, salary, skills, etc.).
   * Identified some missing salary data, which might require imputation or exclusion from analysis.
   * Removed duplicate job postings to avoid skewed analysis.
2. **Data Transformation:**
   * Extracted relevant skills from job descriptions to create a structured format.
   * Converted salary ranges into a consistent numerical format for better analysis.
   * Standardized job location information to facilitate regional comparisons.
3. **Exploratory Data Analysis (EDA):**
   * Generated summary statistics for salary distributions.
   * Created initial visualizations to observe trends in job demand, required skills, and regional distribution.
   * Identified outliers in salary data that may need further investigation.

**Issues Encountered and Handling Strategies:**

* **Missing Salary Data:** Since some job postings do not include salary information, I am considering filling in missing values using median salary estimates based on job title and location.
* **Inconsistent Skill Listings:** The skills were listed as unstructured text, requiring careful extraction and cleaning to categorize them properly.
* **Different Salary Units:** Some salaries were listed annually, while others were monthly or hourly. I standardized these to a common annual format.