**Name:** Gowtham Loganathan

**Milestone 3:** Final Draft of Data Science Job Trends Project

**Project Title:** Exploring Data Science Job Trends through Data Pre-processing and Visualization

**Project Type:** Individual Project

**Research Questions:**

1. How does city development index influence job-seeking behaviour in the data science field?
2. What are the key trends in experience levels, education, and major disciplines among data science job seekers?
3. How does training impact job-seeking trends in data science?

**Data and Methods:** development

* **Dataset:** The dataset used for this project is the publicly available "Data Science Job" dataset from Kaggle. It contains job seekers' details, including education, experience, training hours and city index. Since the dataset is openly accessible and anonymized, there are no privacy concerns.
* **Data Pre-processing and Transformation:**
  + Handled missing values by filling numerical columns (experience, training hours, city development index) with median values.
  + Categorical columns (gender, enrolled university, education level, major discipline, company size, company type) were filled with mode values.
  + Removed duplicate records to avoid bias in the analysis.
* **Exploratory Data Analysis (EDA):**
  + Analyzed experience distributions using box plots and histograms.
  + Examined demand for data science roles across education levels using bar charts.
  + Identified+ the most common major disciplines among job seekers.
  + Investigated training hours distribution and its potential impact on job readiness.
  + Explored the correlation between city development index and job seekers using heatmaps

**Results and Key Findings:**

* Job seekers from higher city development index areas have a higher tendency to seek jobs.
* Most job seekers have a graduate degree, with STEM fields being the most common major discipline.
* Candidates with more training hours tend to be more proactive in seeking job opportunities.
* Experience levels vary widely, but fresh graduates and mid-level professionals form the majority of job seekers.

**Visualizations Used:**

* Box plots and histograms for experience distribution analysis.
* Bar charts to visualize demand across education levels and major disciplines.
* Heatmaps to analyse correlations between city development index and job-seeking intent.
* Histograms to study the impact of training hours on job readiness.

**Next Steps and Recommendations:**

* Further investigate how company size and type influence job-seeking behaviour.
* Explore additional external datasets to analyse salary trends and skill requirements in the data science field.
* Study the impact of remote job availability on candidate preferences**.**

**Conclusion:** This project provides a detailed exploration of data science job seekers' trends, offering valuable insights into how education, experience, and city development affect job-seeking behaviour. These findings can help recruiters and policymakers tailor their strategies to better match industry demands.