# *Caprae Capital Project Report*

**Project Title:** Job Scraping and Analysis Tool

**Author:** Shravani More

Objective:

This project automates job data extraction from Naukri.com based on user-defined keywords such as job roles and locations. The extracted data is analyzed to gain insights into hiring trends, skill demands, and experience requirements. The goal is to provide structured and valuable insights for job seekers and recruiters.

Approach:

Project follows the **Quality First** approach, improving data accuracy, preprocessing, and usability:

- **Web Scraping**: Buit the web scraper tool by using tools like Selenium & Python. Extracted job details (title, company, location, experience, salary, skills, link).
- **Data Cleaning**: Removed duplicates, handled missing values, extracted structured insights.
- **Feature Engineering**: Converted skills into lists, split experience into min/max years, and analyzed demand for job roles.
- **Web App**: built an interactive web app using Streamlit, where users can search for job listings by role and location and download the structured data in excel format.

Tech Stack & Implementation:

- **Scraping**: Selenium & Python for extracting job listings.
- **Data Processing**: Pandas & NumPy for cleaning and structuring the dataset.
- **Web App**: Streamlit for an interactive UI to display scraped job data.

Model Selection & Performance Evaluation:

Since this project primarily involves data scraping and preprocessing, no predictive ML model was implemented. However, the cleaned dataset can be extended for future applications, such as:

- Predictive Job Matching (using NLP models like TF-IDF, BERT).
- Salary Prediction (Regression-based ML models).

Conclusion:

This project provides a real-world job insights tool that automates data extraction and analysis. It helps job seekers identify in-demand skills and companies hire efficiently. The next steps include integrating more job portals and applying machine learning models for predictive insights.