# Uday Kumar Reddy

**LinkedIn:** https://www.linkedin.com/in/uday-kumar-426aa7178/

**GitHub:** <https://github.com/Uday9100/LinkedInn_Project>

**Gmail:** udaykumar2910reddy@gmail.com

# LinkedIn Dataset Automation and Analysis Project Summary

## Project Overview

The LinkedIn Dataset Automation and Analysis project aims to build an end-to-end automated data pipeline using modern data engineering tools. The goal is to help users find job postings based on specific skills and locations while identifying trending skills across various industries.

## Project Components

* **Data Ingestion**: Automated extraction of job postings, company data, and industry insights.
* **Data Processing**: Transformation of large datasets for analysis and recommendations.
* **Workflow Orchestration**: Scheduling and managing data workflows.
* **Data Visualization**: Creating dashboards for insights and trends.

## Steps Followed

### Step 1: Setup Environment

* Installed necessary tools and libraries, including Python, Apache Airflow, Kafka, and Databricks.
* Configured AWS S3 for data storage and management.

### Step 2: Data Ingestion

* Automated the ingestion of job postings and company data into Kafka from S3.
* Utilized Kafka producers to stream data efficiently for further processing.

### Step 3: Data Transformation

* Leveraged Databricks to transform raw datasets, perform data cleaning, and create job recommendations.
* Developed a job recommendation engine to match users with job postings based on their skills.

### Step 4: Workflow Orchestration

* Implemented Apache Airflow to orchestrate the entire data pipeline, including scheduling tasks and ensuring data flow between components.
* Created DAGs (Directed Acyclic Graphs) for managing data ingestion, transformation, and visualization tasks.

### Step 5: Data Visualization

* Used Amazon QuickSight to create interactive dashboards for visualizing trends, insights, and key metrics such as top companies hiring and trending skills.
* Developed alerts for new job postings and skill trends to keep users informed.

### Step 6: Key Outcomes

* Developed a functional job recommendation engine that assists users in finding job postings aligned with their skills.
* Implemented automated alerts for job postings and trending skills.
* Extracted and analyzed the top skills across various industries to inform users about market demand.

## Technologies Used

* **Programming Language**: Python
* **Data Streaming**: Kafka
* **Workflow Management**: Apache Airflow
* **Data Processing**: Databricks
* **Data Storage**: AWS S3
* **Data Visualization**: Amazon QuickSight

## Conclusion

This project successfully demonstrates the integration of modern data engineering tools and practices to automate the analysis of LinkedIn datasets. By leveraging automation and orchestration, the project provides valuable insights into job trends and skill demands, ultimately aiding users in their job search.