

Venu Madhav S



+91 8309458792



with.u.venu@gmail.com



Hyderabad, IN



[venumadhav.s](#)



[Portfolio](#)

Senior Data Engineer

SUMMARY

Senior Data Engineer with 4+ years of experience architecting scalable data products and AI-driven solutions using LangChain, RAG, and FastAPI. Proven track record in automating end-to-end ETL pipelines, optimizing team workflows with LLM-based agents (Claude), and engineering robust data validation tools. Expert in leveraging cloud infrastructure (AWS/S3) and modern orchestration (Argo Workflows) to drive operational efficiency and data quality. Passionate about building innovative, high-impact data ecosystems.

KEY SKILLS

- Data Engineering • Machine Learning Methodologies • Optimization Techniques
- Predictive & Statistical Modelling • Sentiment Analysis • Account Mining • Data Analysis • Langchain
- Data Visualization • Project Delivery • Predictive Modelling & Analytics • Programming • Process Improvement • RAG
- Team Coordination & Leadership • Debugging • Adaptability • Transfer of Learning • Active Learning/Application • Resourcefulness

TECHNICAL SKILLS

- Languages: Python, HTML, LangChain framework
- Cloud Computing: AWS, S3
- Machine/Deep Learning: ANN, CNN, LSTM
- Database: MySQL, MongoDB
- Statistics/ML: Linear/Logistic Regression, Ensemble Trees, Gradient Boosted trees

PROFESSIONAL EXPERIENCE

Senior Data Engineer

Dec '21 - Present

Factly Media and Research

Hyderabad, IN

Factly is a prominent Indian data journalism and fact-checking initiative focused on making public data and government information accessible, meaningful, and transparent to citizens.

Roles and responsibilities:

My position at Factly entails collaborating with diverse stakeholders to develop datasets. This process includes extracting data from various sources through scraping, cleaning the raw information, and applying standardization in accordance with company protocols. Once processed, these refined datasets are uploaded to a Wasabi S3 bucket to ensure they are accessible via the Factly product, [Dataful](https://dataful.in/)

Key Professional Accomplishments

- Delivered a sophisticated internal **research assistant persona** in under 3 months using **Claude** and integrating **LangChain, & Retrieval-Augmented Generation systems with React, postgresql, vector embeddings suitable for LLM** used resulting in a 25% increase in productivity
- Architected an automated task scheduling system leveraging **Claude and Google Chat integration** to streamline task execution, automate pull request generation, and optimize team-based peer review workflows.
- Spearheaded the integration of custom **Claude-reusable domain expertise (Skills)** for a multi-source data integration project, optimizing Extract, Transform, Load pipeline efficiency & streamlining data processing operations
- Optimized internal product development using the **FAST-API framework & Python** to enhance data accuracy & streamline the validation process for uploaded datasets
- Streamlined processes & enhanced productivity by developing internal validation & standardization libraries using **Python programming**, resulting in improved data accuracy
- Modified a workflow generation tool utilizing Argo Workflows & Python scripting to automate the **Extract, Transform, Load pipeline**, reducing processing time by **10%** & significantly enhancing workflow efficiency
- Enhanced data cleaning scripts by incorporating **application programming interface functionality** using Python, resulting in a 15% increase in data accuracy
- Executed the cleaning of 25% of datasets featured in the Datafull product using Python, enhancing **product performance, customer insights, & boosting user engagement**

NAAC accreditation is a quality assurance process by the National Assessment and Accreditation Council for Indian higher education institutions.

- Facilitated continuous improvement initiatives by analyzing three years of student retention data & tracking a **10% increase in student satisfaction scores** using Excel, which contributed to a 15% rise in overall institutional efficiency
- Utilized Excel to identify learning gaps & improve student performance through comprehensive data analysis, contributing to **enhanced educational outcomes**
- Conducted engaging workshops to highlight feature opportunities in data science, emphasizing skills in **Python, data visualization, & machine learning**, resulting in increased student interest & enhanced career awareness

EDUCATION

- **Course Modules:**
 - Data Analysis in Excel | Analytics Problem Solving | Data Analysis using SQL | Introduction to Python
 - Programming in Python | Python for Data Science | Inferential Statistics | Hypothesis Testing
 - Introduction to Machine Learning and Linear Regression | Logistic Regression | Tree Models
- **Secured 90%**

KEY DATA SCIENCE PROJECTS

- Domain: Telecom sector | Tech Stack: Python | May 2021
 - Conducted extensive research on predicting churn and churn analytics to retaining high profitable customers in the telecom sector.
 - Applied various imputation techniques like SMOTE to handle data imbalance, as typical churn rate is about 5 to 10%
 - Recommended strategies to manage customer churn.
- Domain: Real-estate | Tech Stack: Python | March 2021
 - Developed an advanced regression model using Lasso and Ridge regression technique to find significant variables in predicting price of a house.
 - Applied data transformation techniques to remove skewness of data.
 - Collected analytical evidence how well given variables describe the price of a house.
- Domain: Bike renting company | Tech Stack: Python | Feb 2021
 - Designed a multiple linear regression model to predict the demand of bikes on a particular day
 - Exercised RFE and VIF to find a strong predictor which can explain the demand of bikes on typical day
 - Analysed the error in predicted demand with residual analysis to finalize a model.
- Domain: NGO | Tech Stack: Python | Jan 2021
 - Divided given list of countries into two clusters based on GDP, Income and Child mortality rate using K-Means & hierarchical clustering