

# Prathyusha Kodumagulla

prathyuk24@gmail.com | (813) 682-7119

## SUMMARY

---

Senior data and software professional with over 8 years of experience in architecting and optimizing data platforms, enhancing application performance, and leading cross-functional teams. Proven expertise in developing scalable data pipelines and implementing data governance practices to ensure data quality and security.

## SKILLS

---

Python (Pandas, Keras, Tensorflow, PyTorch, Scikit-learn, Jupyter), SQL, AWS (Glue, S3, EC2, EMR, Lambda), GCP, Terraform, Data Governance, Data Visualization (Tableau, Power BI), ETL, Machine Learning, Data Pipelines, NLP, BigQuery, Apache Airflow, Cross-Functional Collaboration, Predictive Analytics

## WORK EXPERIENCE

---

### UnitedHealth Group

Aug 2022 - Present

Senior Data Engineer

- Architected and automated scalable data pipelines using AWS Step Functions and Terraform, improving build latency by 70% and enhancing security and traceability.
- Led a team to implement data governance practices, ensuring compliance and improving data quality across the organization.
- Optimized cloud infrastructure on AWS and GCP, reducing costs by 30% and improving application performance by 25%.
- Deployed over 20 ML models, managing data pipelines using Glue and Lambda, which improved data processing efficiency by 30%.
- Facilitated collaboration among cross-functional teams, developing a cloud Data Warehouse that migrated petabyte-scale data, leading to 30% annual savings.
- Spearheaded multiple projects to optimize data pipelines using GCP BigQuery, significantly reducing query execution times and storage costs by implementing partitioning and clustering strategies.

### Bristol Myers Squibb

Jan 2022 - May 2022

Data Scientist Intern

- Engineered and deployed a real-time NLP model to analyze clinical trial reports, achieving 85% accuracy in identifying patient responses.
- Implemented a preprocessing pipeline that improved model performance by 30%, utilizing AWS Lambda and API Gateway to deploy the NLP model as a serverless REST API.

### University of South Florida

Aug 2021 - Dec 2021

Research Data Analyst

- Automated data collection and analysis using Reddit API and Pandas, reducing team effort by 25%.
- Increased efficiency and accuracy of ETL operations by 50% using Pandas and PySpark to fetch COVID data.

### L'Oreal

Jan 2014 - Jan 2020

Senior Software Engineer

- Developed and maintained scalable software solutions using Python and SQL, reducing system downtime by 25%.
- Implemented ML algorithms for product recommendations, increasing repeat purchaser rates by 18% and reducing reporting time by 40%.
- Led a cross-functional team to implement data governance practices, ensuring compliance with industry standards and improving data quality across the organization.
- Collaborated with stakeholders to establish data access controls and metadata management processes, enhancing security and enabling better data discoverability.

## PROJECTS

---

### Humana-Mays Analytics Case Competition

- Led a team in analyzing factors influencing Covid vaccine hesitancy, building a Random Forest Classifier model with a predictive accuracy of 82%.

## EDUCATION

---

University of South Florida

Jan 2021 - Aug 2022

GPA: 4.0/4.0

**CERTIFICATIONS**

---

**SAS Certified Specialist**  
Base Programming Using SAS 9.4