

# Moukthika Gunapaneedu

Boulder, CO | [LinkedIn](#) | 303-332-1899 | [moukthikagunapaneedu@gmail.com](mailto:moukthikagunapaneedu@gmail.com)

## SUMMARY

**Data Scientist** with 2+ years of experience in developing analytical systems that transform complex data into strategic insights and measurable outcomes. Skilled in Python, SQL, and cloud-based data platforms, with strong foundations in data engineering, statistical modeling, and machine learning. Experienced in building scalable pipelines, designing interpretable models, and delivering actionable insights that influence business strategy. Known for combining technical rigor with clear communication and a focus on real-world impact, driving solutions that are both data-driven and human-centered.

## WORK EXPERIENCE

### University of Colorado Boulder | CO | Data Engineer

January 2025 - Present

- Automated end-to-end ETL pipelines for 3,500+ NOAA tsunami microfilm records, improving metadata accuracy and researcher accessibility; designed a deduplication workflow that reduced redundant archival data by 30%
- Built and deployed Tableau and Power BI dashboards to visualize archival and metadata trends, enabling leadership to monitor data quality metrics, track ingestion rates, and make informed decisions about research priorities
- Applied statistical hypothesis testing (Chi-square, t-tests) to validate metadata consistency across datasets, improving the integrity of research outputs and providing confidence in archival data reliability
- Delivered SQL-based reporting solutions and reusable technical documentation, creating standardized data validation processes that supported collaboration between research, analytics, and IT teams

### Studio Zenoya | India | Database Intern

May 2024 - July 2024

- Designed and implemented scalable ETL workflows to integrate 100K+ API records into analytics pipelines, ensuring data quality, consistency, and availability for downstream BI dashboards and reporting systems
- Developed Python automation scripts to streamline recurring transformations and integrity checks, reducing manual workload for the BI team by 40% and accelerating project delivery timelines
- Optimized MongoDB queries and indexing strategies, which improved dashboard response times and reporting retrieval speed by 25%, enhancing the user experience for business stakeholders
- Partnered with the BI engineering team to design star-schema data models that improved query performance, simplified reporting logic, and ensured scalability of the analytics infrastructure
- Produced Excel and Power BI dashboards highlighting API data trends, anomalies, and performance metrics, giving business stakeholders actionable insights to guide product and customer engagement strategies

### Solar Secure Solutions | India | Data Scientist

September 2023 - December 2023

- Designed and deployed machine learning pipelines using Scikit-learn and XGBoost that reduced fraud false positives by 20%, while improving overall detection accuracy and enabling more efficient fraud investigation processes
- Built and maintained ETL workflows to process large volumes of structured and unstructured financial transaction data, increasing BI readiness by 35% and ensuring consistent availability of clean datasets for modeling and analysis
- Created interactive Tableau dashboards that visualized fraud risks, flagged suspicious anomalies, and communicated fraud detection outcomes in a clear, business-friendly format for non-technical stakeholders
- Conducted A/B testing and regression analysis to evaluate the effectiveness of fraud detection strategies, delivering actionable recommendations that enhanced model performance and informed product security measures
- Collaborated with cross-functional teams, including operations and compliance, to translate ML-driven insights into fraud prevention rules, operational alerts, and real-time monitoring systems, strengthening overall security posture

## SKILLS

**Programming & Systems:** Python (Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch), Java, C++, SQL (MySQL, PostgreSQL, SQL Server), Linux, Bash; Data Structures & Algorithms; Parallel & Distributed Systems

**Machine Learning & AI:** Supervised & Unsupervised Learning, Regression & Classification, Clustering, Deep Learning (CNNs, RNNs, Transformers), Generative AI (LLMs, Diffusion Models), Reinforcement Learning (RLHF), Recommendation Systems, NLP

**Data Engineering & Big Data:** ETL/ELT Pipelines, Apache Spark, Hadoop, Kafka, Data Lakes, Feature Stores, Data Warehousing (Snowflake, Redshift, BigQuery), Workflow Orchestration (Airflow, Prefect)

**MLOps & Deployment:** Docker, Kubernetes, AWS (S3, SageMaker, Lambda), GCP, CI/CD Pipelines, Monitoring & Logging (MLflow, Datadog, CloudWatch), Experiment Tracking & Version Control (Git/GitHub)

**Mathematics & Statistics:** Probability & Statistics, Linear Algebra, Optimization, Causal Inference, Hypothesis Testing, A/B Testing

**Visualization & Business Intelligence:** Tableau, Power BI, Matplotlib, Seaborn, Plotly; Communication with Stakeholders

**Tools & Workflow Management:** Jupyter, VS Code, GitHub Actions, DVC, Weights & Biases, Streamlit, Flask, FastAPI, Notion, Jira, Confluence

## EDUCATION

**University of Colorado Boulder**  
MS in Data Science (GPA: 3.9)

Boulder, CO