

# Gouri Menon

United States | +1-716-817-3964 | [gourimenon8@gmail.com](mailto:gourimenon8@gmail.com) | [LinkedIn](#) | [GitHub](#)

## Summary

MS Data Science grad with experience in building scalable data pipelines, machine learning models, and real-time dashboards. Skilled in SQL, Python, and cloud platforms (AWS, GCP, Snowflake), with hands-on expertise in analytics for healthcare, marketing, and fintech. Passionate about turning raw data into actionable insights and deploying end-to-end solutions that create measurable impact.

## Experience

### Graduate Research Assistant | University at Buffalo

Jan 2025 – July 2025

- Built modular preprocessing services in Python to handle 50GB+ of biomedical imaging data, enabling reproducible pipelines.
- Deployed PyTorch-based segmentation models behind REST APIs to automate drusen detection, improving accuracy by 14% compared to a U-Net baseline.
- Designed statistical services in Python to calculate progression metrics, integrated into downstream research workflows.
- Collaborated with cross-functional teams to align model outputs with software systems and databases.

### Associate Engineer | Muse Production House, Pune, India

May 2022 – Oct 2023

- Built serverless ETL pipelines using AWS Glue and Athena to clean, transform, and load multi-channel marketing data into Snowflake, enabling real-time dashboards and analytics for campaign performance.
- Developed RESTful APIs using FastAPI to expose campaign KPIs (engagement, conversion) to internal dashboards.
- Automated compliance classification services with Regex parsers in Python, reducing manual overhead by 35%.
- Implemented time-series forecasting modules and exposed results via APIs, improving ROI-based decision making.
- Integrated Tableau dashboards with Snowflake and APIs to provide real-time insights for 2M+ users.

### Software Engineering Intern | Skiploop Labs, Bangalore, India

Sep 2021 – April 2022

- Developed real-time anomaly detection microservices in Python & Flask, consuming device event streams from Kafka, detecting irregularities, and triggering alerts for fitness app users.
- Leveraged PySpark to design clustering services, powering segmented chatbots for personalized healthcare engagement.
- Automated ETL validation services in Python before ingesting device data into BigQuery, improving reliability across devices.
- Designed interactive web dashboards with Plotly to surface model metrics and predictive alerts for client-facing applications.

## Education

### University at Buffalo, SUNY – MS. in Engineering Science (Data Science)

Jan 2024 – May 2025

(Coursework: Statistics and Data Mining, Machine Learning, Time Series Analysis, Python & Data Structures, Probability, Data Model and Query Language, Data Intensive Computing, Machine Learning Ops, Numerical Mathematics, Hypothesis Testing)

### University of Pune – BS in Physics (minor: Mathematics)

June 2018- June 2021

## Projects

### NYC Taxi Forecasting System | *AWS, Snowflake, XGBoost, FastAPI, Tableau*

Built scalable data pipelines on AWS to process millions of transportation records. Trained forecasting models (XGBoost, LightGBM) and deployed predictions via FastAPI services and Snowflake dashboards, improving forecast accuracy by 20%.

### Salary Prediction Web App | *Python, scikit-learn, Streamlit, Flask*

Designed ML models to predict salaries by role & location. Deployed as an interactive Streamlit + Flask app, enabling real-time predictions and scenario testing for users.

### Blockchain Analytics Platform | *SQL, Flask APIs, Dashboards*

Developed SQL databases and ETL pipelines for millions of Bitcoin transactions. Exposed insights through Flask APIs and dashboards, helping uncover on-chain patterns and anomalies.

- Final Thesis for my Undergrad:** Built high-fidelity physics simulations in Python/VPython modeling orbital mechanics, projectile motion, and charged particle dynamics. Implemented numerical integration (Euler–Cromer) to solve complex multi-body systems and optimized computation for real-time 3D visualization. Validated results against theoretical models, demonstrating skills in physics based modelling, algorithm design.

## Technical Skills

- Languages:** Python, SQL, Java, R, Bash, HTML/CSS, Git
- Data Science & Machine Learning:** PySpark, Hadoop, Scikit-learn, PyTorch, TensorFlow, Hugging Face, MLflow, Feature Engineering, A/B Testing, Clustering, XGBoost, LightGBM
- Cloud & MLOps :** AWS, GCP, Docker, GitHub Actions, CI/CD, Kubernetes, Terraform
- Analytics & Visualization:** Power BI, Tableau, Streamlit, Excel (advanced), Google Sheets, Looker, KPI Dashboards
- Tools & Platforms:** Jupyter Notebook, Google Colab, Docker, Flask, REST APIs, VS Code, Jira, Confluence, Slack, GitHub
- Database & Storage:** PostgreSQL, MySQL, SQLite, Redshift