

MEGHNA REDDI

(973) 906-9462 | mreddi2001@gmail.com | [/in/meghnareddi](https://in/meghnareddi) | github.com/meghnareddi | Jersey City, NJ

SUMMARY

Data Scientist & Data Engineer with 2+ years of experience building high-impact predictive models and data solutions across government and equity-focused organizations. Developed housing demand models informing statewide infrastructure decisions and spearheaded data standardization and visualization efforts that enabled a 40% increase in financial support for women inventors. Skilled in ETL pipelines, machine learning, Power BI dashboards, and cloud platforms, driving actionable insights to support strategic decision-making.

PROFESSIONAL EXPERIENCE

Maryland Department of Planning | Data Scientist & Engineer – *Baltimore, MD*

03/2025 – Present

- Developed predictive models (Linear Regression, XGBoost) on ACS data using mortgage rates, family size, and income to forecast regional housing demand ($R^2 = 0.81$). Informed statewide funding and land use decisions for infrastructure planning teams and legislators
- Engineered an ETL pipeline in Python (requests, JSON) to extract and transform 3.8M ACS records via Census API into SQL Server using GitLab CI/CD. Supports annual data ingestion cycles powering predictive models and Power BI dashboards for planning teams
- Designed an interactive Power BI dashboard visualizing KPIs population growth %, dependency ratio, diversity index to guide housing and infrastructure planning. Accessed year-round by Maryland policymakers to justify funding assumptions in state and federal meetings

New Jersey Equity in Commercialization Collective | Data Scientist & Engineer – *Newark, NJ*

01/2024 – 12/2024

- Deployed clustering pipeline (DBSCAN, TF-IDF, cosine similarity) to standardize 5,000 inventor names across 9 universities, reducing false positives by 60% in USPTO data. Enabled timely demographic analysis, securing funding for women inventors at NJ equity conferences
- Validated 1,000 inventor profiles by retrieving gender data via LinkedIn and university forums using REST APIs, correcting inaccuracies from the `gender_it` library. Improved report accuracy and female inventor representation in NJECC's equity-focused patent analysis
- Created Power BI dashboard tracking KPIs women inventor ratio YoY, Patents granted, citations enabling NJECC board to review trends before the 2024 NSF ADVANCE-supported "Innovation for All" conference. Informed a 40% increase in financial aid for women inventors

New Jersey Institute of Technology | Teaching Assistant – *Newark, NJ*

01/2024 – 12/2024

- Progressed assignment scores by 15% through 6 lab exercises in RStudio on CNNs, Data Analysis and Large Language Models (LLMs)
- Conducted workshop on pandas, seaborn and statistical models for 30+ students, improving project quality and reproducibility

Zenoti India Pvt Ltd | *Data Analyst – Hyderabad, IND*

07/2022 – 07/2023

- Developed 10+ DAX churn KPIs in Power Query, including MRR, average invoice value, and system adoption metrics, to analyze high-risk customer accounts across North America. Enabled PMs to conduct root cause churn analysis and drive targeted retention strategies
- Reduced churn by 12% in SMB and strategic accounts within one month by developing a BI dashboard tracking MRR, average invoice value, and system adoption metrics. Enabled targeted interventions that boosted customer retention
- Built 4+ automated Excel reports with Pivot Tables for weekly customer training course updates, integrating with Power Automate to deliver insights to product specialists and PMs. Saved 3 hours per week by eliminating manual reporting

ACADEMIC PROJECTS

Minimum Norm Solutions & Neural Collapse in Deep Network Training (Current): Designed a layer-wise and end-to-end training framework for neural networks, implementing Coordinate Descent (CD) and Stochastic Gradient Descent (SGD). Achieved 97.8% accuracy with CD and 97.06% with SGD on high-dimensional datasets, and analyzed minimum norm solutions to assess model generalization

Bias Detection and Mitigation in LLM Generated Text: Developed a bias detection pipeline using AIF360 to analyze LLM (GPT-2, BERT) outputs, assessing fairness metrics across demographic factors. Reduced bias by 10% using Adversarial Debiasing, enhancing model trustworthiness and fairness in AI-generated content

Predictive modeling & Time Series Forecasting on Divvy Bicycle Sharing System: Implemented an ETL pipeline using AWS Glue to preprocess 25M+ records from S3, conducted EDA with Athena and QuickSight, and trained time series models (ARIMA, VAR, Prophet) on SageMaker, reducing forecasting error by 20%. Deployed models via Docker with CI/CD using GitHub Actions for reproducibility

Azure AI-Powered Document Processing and QA System: Built an Azure data processing pipeline using AI Vision to extract text from 30GB+ documents in Blob Storage and AI Language Service for sentiment analysis, named entity recognition, and review categorization. Deployed a Question Answering system using Azure AI Language, reducing response time by 35% through knowledge base integration

TECHNICAL SKILLS

Programming Languages: Python (Scikit-learn, TensorFlow, Pandas, NumPy, Seaborn) | R | SQL

Machine Learning: Linear & Logistic Regression | Decision Tree | XGBoost | Random Forest | Ensemble Methods | Feature Engineering | Hyperparameter Tuning | Model Evaluation | Time Series (ARIMA, Prophet) | A/B Testing | Hypothesis Testing

NLP & Deep Learning: BERT | Transformers (Hugging Face) | spaCy | GPT-3.5 | CNN | RNN | LSTM

Generative AI: Retrieval-Augmented Generation (RAG) | Prompt Engineering | LLM Fine-tuning | Adversarial Debiasing (AIF360)

Data Engineering & MLOps: ETL | Data Wrangling | Spark | REST APIs | Docker | CI/CD | Git | Model Deployment

Cloud Platforms: AWS (SageMaker, EC2, S3, EMR, Redshift) | Azure (AI Services, Data Factory, Databricks, Dataflow Gen2) | GCP (BigQuery)

Data Visualization & Business Intelligence: Power BI (Power Query, DAX) | Tableau | Quicksight

Databases & Automation: MS SQL Server | MySQL | Power Automate

Project Management: JIRA | Confluence | Agile | Scrum

Certifications: [AWS Certified Cloud Practitioner](#) | [Generative AI with Large Language Models](#) | [Google Data Analytics Professional](#)

EDUCATION

New Jersey Institute of Technology, Newark, NJ – Master of Science (M.S.) – Data Science, Concentration in Statistics – GPA: 3.8

Mahatma Gandhi Institute of Technology, Hyderabad, IND – Bachelor of Technology (B.Tech.) – Computer Science – GPA: 3.5