

SARADHI SOMAROUTHU

☎ 812-671-2352 ✉ ssomaro@iu.edu [in linkedin.com](https://www.linkedin.com/in/ssomaro)

EDUCATION

Indiana University Bloomington

August 2022 – May 2024

Master of Science in Data Science

GPA: 3.93

- **Coursework:** Data structures and Algorithms, Machine Learning, Artificial Intelligence, Statistics, Database management, Software Engineering, Data Visualization, Knowledge Graphs for LLMs, Exploratory Data Analysis, Graph analytics

EXPERIENCE

Market Research Data Analyst | eBay

January 2024 – Present

Contract, Remote

- **Data Warehousing:** Designed a Data Mart to supporting causal inference and data visualization tasks. Built ETL pipelines to automate data refresh, Leveraged SQL and PySpark to process over 2 billion rows, ensuring data accuracy and faster updates through incremental data load.
- **Advanced SQL:** Expertly utilized advanced SQL techniques, including CTEs, window functions, views, and stored procedures, to optimize query performance, enhancing the efficiency of ETL pipelines and supporting complex data operations at scale.
- **Causal Inference:** Conducted in-depth causal inference analysis using R to assess the impact of Promoted Listings on seller performance metrics, employing methodologies such as Average Treatment Effect, Causal Trees, Causal Random Forests, Difference-in-Difference estimates, and Propensity Score Matching.

Data Science Intern | Novelis

June 2023 – December 2023

Atlanta, Georgia

- Developed **retrieval augmented generation** features for EagleGPT leveraging Azure OpenAI API's for chat and embeddings.
- **Business Intelligence tool:** Developed and deployed a containerized BI tool using Plotly Dash, Databricks notebooks and workflows, transforming data from Azure data Lake. Hosted front-end as an dockerized image on Azure Container Registry and integrated Single Sign-on for added security and optimized load time by 25%
- Architected full-cycle forecasting system for LME contract prices, spanning algorithmic formulation, configuration of automated pipelines on Databricks to refresh data, and deployment of the application on Azure App Service.

Data Science Senior Analyst | Accenture

July 2020 – Aug 2022

Bengaluru, India

- Developed automated framework to forecast demand at SKU-Store-Week granularity with an accuracy of 85% for the supply chain team at Apple Inc.
- Implemented pipelines using Azure Data Factory, Databricks and Analysis Services to detect and flag anomalies in the global supply chain for the Real Time Visibility and Analytics team at Unilever.
- Collaborated with business stakeholders in understanding technical requirements, and performed complex transformations and analytic functions on data to make it available in a reporting dashboard for downstream business consumers.
- Designed and developed multiple complex dashboards using Power BI that queries 10 million records of data from Azure Synapse to calculate 160+ KPIs related to warehousing, logistics and customer service.
- Worked on a vendor management tool to optimize the expenditure on procurement by selecting the best benchmark vendors, Showed opportunity cost of 1.5 million USD over 2 quarters.
- Developed 15+ KPIs and visualization for a Demand planning and inventory watch tower using React, Python(flask).

TECHNICAL SKILLS

Programming Languages: Python, R, SQL, PySpark

Cloud Technologies: Azure data factory, Databricks, Analysis services, Synapse, ML Studio, Airflow

Databases: MySQL, SQL Server, PostgreSQL, MongoDB, Neo4j

Data Visualization: Power BI, Tableau, Matplotlib, Seaborn, Dash, Streamlit

PROJECTS

BeatMe - AI Game play bot | Search algorithms, Reinforcement Learning, Game Theory

- Implemented an AI Bot that plays customized version of chess, Shortest path maze, 5 x 5 board game of Go against humans using reinforcement learning, mini-max, alpha-beta pruning, A*, heuristics and other search algorithms that makes the optimal move under 10 seconds.

Dreamship | Full stack MERN application

- Developed a Delivery Management system to choose optimal service, place, track and manage Orders integrating MFA, Google maps API and OpenAI chat completion API for enhanced chat-bot experience with Retrieval augmented generation.

Hoosier Gas Project | PostgreSQL, Python, PowerBI, Airflow

- Designed and orchestrated robust data pipelines, extracting hourly GasBuddy metrics using GraphQL, leveraging Apache Airflow and Spark on Databricks for transformation, storing in a PostgreSQL, and visualizing real-time trends via PowerBI dashboards and reports.