VENKATA SUBRAHMANYA SRI HARSHA KALAGA

Boston, MA | +1 (857)-396-8947 | Kalaga.v@northeastern.edu | https://www.linkedin.com/in/vsshk

EDUCATION

Northeastern University, Boston, MA **September 2023 – April 2025**
Master of Professional Studies in Analytics **CGPA: 3.8/4.0**
Relevant Course Work: Python, SQL, R Programming, Data Mining, Predictive Analytics, Applications of AI, Probability and Statistics, Database Management System

TECHNICAL SKILLS

Languages	Python, R
Cloud Technologies	Azure (Databricks, Machine Learning Studio), AWS (S3, SageMaker)
Databases	SQLite, PostgreSQL, Microsoft SQL Server
Productivity and Design	Microsoft Excel, Google Sheets, Canva, Adobe Illustrator
Visualization	Tableau, Power BI, Qlik
Statistics	Descriptive Statistics, Hypothesis Testing, ANOVA,
Machine Learning	Linear Regression, Logistic Regression, Decision Trees, Random Forest
Libraries	NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn, Statsmodels, SQLite3, PySpark

PROFESSIONAL EXPERIENCE

Client: Kenvue Inc. (via Tata Consultancy Services) | Data Analyst
New Brunswick, NJ | January 2021 – June 2023

- Remodeled database architecture for Global Inventory data using **SQL**
- Configured faster Big Data ETL pipelines through **PySpark**
- Reduced execution time for transformation workflows through parallelization
- Led integration of inventory data from 15 regions across SAP, BPCS, and MOVEX from EDW to Azure, enabling faster centralized analysis using **SQL** and **Azure Databricks**

ACADEMIC PROJECTS

SQL query generator from natural language prompts **January 2025 – March 2025**

- Designed and implemented a middleware system that leverages Large Language Models (LLMs) to automatically generate SQL queries from natural language inputs, aiming to simplify access to urban mobility data and enable data-driven insights without requiring manual coding.
- Integrated LLM into a user-centric interface that translates natural language prompts into analytical SQL queries, empowering non-technical stakeholders to conduct complex urban mobility analysis by removing the barrier of programming knowledge.
- Established a continuous learning mechanism by developing a feedback loop stored in a dedicated database, which tracks user interactions and corrections to iteratively refine the LLM's outputs—enhancing the system's adaptability and supporting the evolution of scalable smart city planning tools.

Census-Income Analysis **May 2024 – June 2024**

- Predicted income categories by analyzing demographic and employment data using decision trees and logistic regression models, with the goal of uncovering patterns relevant to socioeconomic stratification.
- Utilized SQLite as the data management system to facilitate efficient handling of large datasets and reduce preprocessing complexity, supporting streamlined machine learning workflows.
- Developed and validated predictive models to support data-driven policy discussions and provide a foundation for future studies on income inequality and demographic shifts.
- Employed Python in conjunction with SQLite and Visual Studio Code to build, test, and deploy predictive models while ensuring modular and readable code practices.

Event Management Database System **November 2023 – December 2023**

- Built a relational database system using SQLite to manage operational workflows for over 50 diverse events, aiming to centralize planning, coordination, and cost tracking.
- Created structured tables and query systems to monitor key event attributes, such as budgets, timelines, and vendor interactions, thereby facilitating more organized event execution.
- Developed intuitive interfaces for event managers to enter, update, and review records, ensuring the solution met real-time usability and operational decision-making needs.
- Designed the system with SQL-based logic to ensure data integrity, minimize redundancy, and enable consistent reporting for performance and satisfaction analysis.

CERTIFICATIONS

Azure Fundamentals | Azure Data Fundamentals | Python Certified Associate Programmer