Keerthi Samhitha Kadaveru

(774) 346-4640 | k.samhitha23@gmail.com | LinkedIn | Buffalo, NY

Education

University at Buffalo - SUNY, Buffalo, NY — *M.S. in Data Science* | August 2024 - December 2025
Gurunank Institute of Technology, India — *B.Tech, Electronics & Communication* | June 2017 - May 2021

Work Experience

Associate | Ingram Micro, Buffalo, NY | *June 2025 - December 2025*
- Maintain weekly **Power BI dashboards** tracking order KPIs for 15+ stakeholders, consolidating data from multiple sources (**CRM**, **Excel**) that reduced **ad-hoc reporting** requests by **20%**
- Resolved 40+ high-priority customer-facing issues through **root cause analysis** using **Excel** pivots and CRM data, documented findings that decreased similar incidents by 25%
- Extracted order and revenue data from CRM using **SQL**, cleaned and standardized 15+ weekly datasets in Excel (**handled missing values**, **duplicates**, format inconsistencies using **Pivot Tables** and **VLOOKUP**) to ensure accurate month-end financial reconciliation
- Analyzed **forecast variance** by querying CRM order data and financial actuals using **SQL**, built **Power BI reports** segmenting gaps by SKU, pricing errors, and rejection codes to identify top contributors—**improved** forecast accuracy by **11%**

Data Analyst | Capgemini, Bengaluru, India | *October 2021 - August 2024*
- Built automated **ETL pipelines** using **Apache Airflow** processing data from 10+ sources, cut manual data processing time by 25% and delivered real-time business reporting for 50+ stakeholders
- Engineered data cleansing pipelines using **Python** and **SQL** to handle missing values, **outliers** and **inconsistencies**, implementing **standardized validation rules** that improved data completeness to **45%**
- Optimized **PostgreSQL databases** (50M+ records) through **SQL query tuning** and **index restructuring**, boosting query execution time by 30% for **ad-hoc analysis** and dashboards
- Analyzed **compliance** data across 20+ client IT infrastructures using statistical sampling and automated validation scripts, identified **300+ security vulnerabilities** improving client compliance scores to 47%
- Developed **Power BI dashboards** with **DAX measures** visualizing system performance and compliance KPIs, reducing decision cycles from 15 days to 3 days across 15+ clients

Projects

RetinaScan-AI (Diabetic Retinopathy Detection System) | Python, TensorFlow, OpenCV, Flask | **GitHub**
- Built **CNN model** trained on 35K+ retinal images to classify diabetic retinopathy across 5 severity levels, achieving **0.70 kappa score** using dual-eye fusion architecture
- Implemented **preprocessing** pipeline using **OpenCV** (CLAHE, denoising) with **strategic augmentation** for class imbalance; deployed via **Flask REST API** for real-time predictions
- Addressed class imbalance through **selective augmentation** (rotations, mirroring) on minority classes, applied two-phase sampling strategy to **prevent overfitting** on rare disease categories

Crime Analysis Dashboard | Power BI, DAX, Power Query, SQL | **GitHub**
- Analyzed **2,500+ crime** incidents across 13 countries using **Power Query** ETL and DAX; built **relational data model** connecting 7,461 records tracking 70% resolution rate
- Created **DAX measures** tracking **year-over-year** trends (405 to 1,150 cases), highlighted 30% unresolved backlog across 12 crime categories for investigation prioritization
- Designed **interactive dashboards** with drill-through and **heatmap visualizations** showing crime patterns by weekday, month, and location for resource allocation

Healthcare Performance Analytics | Power BI, DAX, Statistical Modeling | **Link**
- Analyzed **HCAHPS survey data** (43,200+ responses over 9 years) using **DAX** and **statistical trend analysis**, identifying baseline satisfaction metrics (62.9%) and post-2020 decline patterns
- Built **BI framework** with regional benchmarking across 15+ states, automated **anomaly detection** for satisfaction scores and predictive indicators for patient experience risk factors
- Implemented **data quality** assurance protocols and **validation rules** ensuring 99.8% **data accuracy**, supporting regulatory compliance and evidence-based improvement strategies

Technical Skills

Data Analysis: Excel (LOOKUP functions, Pivot Tables, VBA, Macros), Data Wrangling, Exploratory Data Analysis (EDA)
Data Science & Modeling: Statistical Analysis, Predictive Modeling, Machine Learning, Data Mining, Trend Analysis, Hypothesis Testing, A/B Testing, Regression Analysis
Programming & Database: Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn), SQL (PostgreSQL, MySQL, Advanced Queries), R Programming, ETL Development, Data Warehousing, Agile/Scrum, KPI Development
Visualization & BI: Power BI (DAX, Power Query), Tableau, Interactive Dashboard Visualizations

Certifications

Google Data Analytics Professional Certificate | Power BI Data Analyst Associate | SQL Server Certification