

SANTHAKUMAR RAMESH

Buffalo, NY | santhar1500@gmail.com | +1 716-328-4155

LinkedIn: linkedin.com/in/santhakumar-ramesh | GitHub: github.com/Santhakumarramesh

Data Analytics Profile

Results-driven Data Analyst with hands-on experience in **Python, SQL, Power BI, and machine learning**. Currently conducting research on **NHANES datasets** to study the impact of occupational heavy metal exposure on physical function using **survey-weighted regression and predictive modeling**. Proven ability to translate complex data into clear insights through **visual storytelling and dashboard development** across healthcare, aviation, and business domains. Adept at building **end-to-end data pipelines**, performing feature engineering, and delivering actionable insights that drive decision-making. Passionate about using analytics to uncover trends and improve real-world outcomes.

EDUCATION

University at Buffalo, New York — MPS in Data Science & Applications (Jan 2025 – Present) | GPA: 3.222

Coursework: Statistical Data Mining I & II, Deep Learning, Data Models & Query Language, Data Intensive Computing

Rajalakshmi Engineering College, Chennai — B.E., Electronics & Communication Engineering (Oct 2020 – May 2024) | CGPA: 7.83

TRAINING/ COURSES

- Google Data Analytics Professional Certificate — Coursera (2024)
- IBM Data Science Specialization — Coursera (2024)
- SQL for Data Analysis and Business Intelligence — Udemy (2024)
- Tableau Desktop for Data Visualization — LinkedIn Learning (2025)
- Python for Data Science & Machine Learning Bootcamp — Udemy (2023)

EXPERIENCE

Research Assistant | University at Buffalo, School of Public Health and Health Professions (SPHHP)

Buffalo, NY | Aug 2025 – Present

- Led data-driven research on the *impact of occupational heavy metal exposure on physical function and fall risk* using **NHANES 2011–2016 datasets**, supporting public health insights into workplace exposure risks.
- Processed and integrated 10K+ survey records across multiple NHANES domains (OCQ, PBCD, PFQ, DEMO) using **Python (pandas, statsmodels, seaborn)**, improving dataset consistency and analytical accuracy.
- Built and evaluated survey-weighted regression, Random Forest, and XGBoost models that identified key predictors of fall risk and physical impairment with **>85% model accuracy**.
- Designed visual analytics and calibration plots to enhance interpretability of statistical findings, directly contributing to **data-driven public health recommendations**.
- authored a forthcoming manuscript and conference presentation to disseminate findings on occupational metal exposure effects.
- Occupational Exposure of Heavy Metals- NHANES — <https://github.com/Santhakumarramesh/occupational-exposure-NHANES>

Academic Project — Flight Delay Prediction (Machine Learning Pipeline) — University at Buffalo 2025

- Developed Logistic Regression, Neural Network, and XGBoost models improving prediction accuracy by 15%.
- Performed feature engineering, class rebalancing, and model evaluation using precision-recall metrics.
- Designed Power BI dashboards and Plot charts to visualize trends and model performance.
- Flight Delay Prediction — github.com/Santhakumarramesh/flight-delay-prediction

Academic Project — Cardiovascular Risk Analysis (Predictive Modeling) — University at Buffalo 2025

- Conducted exploratory data analysis (EDA) on 2,000+ health records to identify key predictors of cardiovascular disease.
- Trained K-Nearest Neighbors and Logistic Regression models achieving AUC of 0.82.
- Generated interpretable outputs and visual summaries supporting healthcare decision insights.
- Cardiovascular Risk Prediction — github.com/Santhakumarramesh/Cardiovascular-Risk-Prediction

Academic Project — Hospital Visit Analysis Dashboard (Business Intelligence Project) — University at Buffalo 2025

- Created SQL queries and Power BI dashboards to analyze hospital visits, doctor activity, and revenue metrics.
- Automated ETL and reporting processes, reducing manual reporting time by 30%.
- Delivered interactive insights improving understanding of operational efficiency and patient inflow.
- Hospital Visit Analysis Dashboard — github.com/Santhakumarramesh/Hospital-Visit-Analysis

Personal Project — F1 Insights & Predictive Analysis (GitHub) — Independent 2025

- Built end-to-end predictive pipeline using Python (XGBoost, scikit-learn) to forecast Formula 1 race outcomes.
- Visualized driver telemetry and pit-stop efficiency using matplotlib and seaborn for performance analysis.
- F1 Insights & Predictive Analysis — github.com/Santhakumarramesh/f1-insights-predictive-analysis

SKILLS

Programming: Python, SQL, R, C; basic Java

BI: Tableau, Power BI, dashboards, reporting

ML: Logistic Regression, RandomForest, KNN, XGBoost, Neural Networks

Data Engineering: ETL pipelines, data warehousing, profiling, cleaning, feature engineering

Tools: PostgreSQL, Excel (Advanced), Git, Jupyter, WebRTC

Soft Skills: Analytical thinking, problem-solving, collaboration, communication