

Sathwik Bollepalli

860-781-1558 | sathwik238@gmail.com | [LinkedIn](#) | [Portfolio](#) | [GitHub](#)

SUMMARY

Data Analyst with a strong foundation in data visualization, statistical analysis, and business intelligence. Proven track record of improving model accuracy by 15% and reducing data processing time by 70%. Skilled in Python, SQL, and Tableau, with a focus on uncovering insights that support data-driven decisions and improve business performance.

PROFESSIONAL EXPERIENCE

Plymouth Rock Assurance - Boston, USA

Data Scientist

May 2024 – December 2024

- Analyzed neighborhood-level property loss data and developed over **800 derived features** to support risk segmentation and pricing analysis for home insurance. Built an XGBoost model to enhance risk segmentation and improve pricing accuracy.
- Conducted detailed exploratory data analysis (**EDA**) to uncover **trends in claim severity** across geographic regions, contributing to improved targeting of high-risk areas like flood zones and high-crime neighborhoods.
- Applied resampling and adjustment techniques to **analyze large-loss fire events** in imbalanced datasets, improving the precision of high-risk segment identification and supporting underwriter recommendations.
- Performed a **coverage analysis** of a new dataset within our six-state footprint to evaluate data completeness, identify regional gaps, and assess alignment with existing underwriting criteria.
- Influenced key strategic decisions by delivering insights to senior management, resulting in the integration of two high-value third-party data sources, improving model performance and decision-making processes.

MAQ Software Pvt Ltd - Hyderabad, India

Reporting Analyst/ Data Engineer

May 2022 – August 2023

- Built **Power BI dashboards** from scratch, bringing together data from **multiple sources** to provide a clear, interactive view of key metrics across the organization which led to **35%** increase in customer engagement.
- Managed the entire **data pipeline**—from extracting raw data using Azure Data Factory to storing it in Azure Data Lake and preparing it for reporting.
- Created meaningful **metrics using SQL**, helping teams better understand trends, track performance, and make faster decisions.
- Cleaned and merged data from different systems to ensure the dashboards told a consistent, accurate story across departments.
- Improved the speed and performance of dashboards by **optimizing SQL queries** with PySpark and Python, cutting load times and making reports more responsive.
- Worked closely with stakeholders to understand reporting needs and turn them into practical solutions, leading to better adoption and more informed **business conversations**.

EDUCATION

University of Connecticut – Storrs, CT

August 2023 - December 2024

Master of Science (M.S) in Data Science, GPA : 3.7

Relevant Courses – Predictive Modeling, Machine Learning, Statistical Analysis, Causal Inference, Data Visualization, Deep Learning.

ACADEMIC PROJECTS

- Bank Customer Churn Prediction | R | ML** – Utilized logistic regression, decision trees, and random forests for predictive analysis of customer churn in the banking sector, incorporating various performance metrics to identify high-risk customers, contributing to the development of effective retention strategies, showcasing skills in data analysis and problem-solving.
- Email Spam Detection | Python | ML** – Developed an advanced email spam detection system using Python and ML techniques, achieving 95% accuracy in classifying emails into 'spam' or 'ham'. Boosted data management and classification efficiency, enhancing workflow processes and minimizing false positives.
- Energy Demand Forecasting | Python | Deep Learning** – Developed an RNN-based energy demand forecasting model using weather data, leveraging Conv1D, LSTM layers, dropout, and activation functions in Keras. Achieved high accuracy with feature engineering, walk-forward validation, early stopping, and validated using mean absolute error.

TECHNICAL SKILLS

Python (Pandas, Keras, NumPy, PySpark, Scikit-learn, Matplotlib, Seaborn), SQL (MySQL, PostgreSQL, Snowflake), R, Data Analysis, Data Wrangling, ETL Pipelines, Data Visualization (Tableau, Power BI, Matplotlib, ggplot2), Statistical Modeling, A/B Testing, Hypothesis Testing, Bayesian Analysis, Transformers (BERT, GPT), Tableau, Microsoft Office Suite (Excel, PowerPoint, Word), Machine Learning (XGBoost, Random Forest, SVM, K-Means, KNN), Predictive Modeling, Deep Learning (CNN, RNN, LSTM), Negative Binomial Regression, Linear & Logistic Regression, AWS (S3, EC2, SageMaker), Azure (Data Factory, Databricks).

ACHIEVEMENTS AND CERTIFICATIONS

- Modern Big Data Analysis with SQL Specialization
- Google Data Analytics
- Hacker Rank certified Python Programmer and Advanced SQL Programmer
- JP Morgan & Chase CADP Certification