SHAIK KAREEM BASHA

Data Analyst | AI Automation Engineer | Python Developer
Phone: 9573434890 || Email: kareembashashaik2401@gmail.com
LinkedIn || Portfolio

PROFESSIONAL SUMMARY

Data Analyst with hands-on experience in analytics, automation, and machine learning. Skilled in building real-time dashboards, data pipelines, and predictive models. strong in Python, SQL, EDA, visualization, and cloud workflows.

CORE SKILLS

- **Programming & Data Analysis:** Python, Pandas, NumPy, SQL, MySQL
- **Visualization & BI Tools:** Power BI, Tableau, Plotly, Streamlit
- **Machine Learning & Automation:** Scikit-learn, GitHub Actions, Supabase
- **Generative AI & Copilot Tools:** Microsoft Copilot, ChatGPT (Prompt Engineering, Workflow Automation)
- **Data Handling & Optimization:** Advanced Excel, ETL Pipelines, Parquet Conversion
- **Version Control & Collaboration:** Git, GitHub

PROFESSIONAL EXPERIENCE

Carelon Global Solutions — Claims Analyst

- Processed **600 CMS-1500 claims per week** with a **98% accuracy rate**, ensuring compliance with payer policies and reducing resubmission delays.
- Identified top recurring error patterns and enhanced validation checks, contributing to a **10–12% reduction in rework** within the team.
- Collaborated with QA and operations to establish data-quality checkpoints, reducing average claim resolution time from **7 to 4 days**.
- Used advanced Excel functions (VLOOKUP, conditional rules, pivot reports) to streamline claim verification and improve weekly reporting efficiency.

PROJECTS

AI-Driven Farm Intelligence Dashboard

• Built a **real-time farm analytics system** using Python, Streamlit, and Supabase to assist farmers with daily decision-making.
• Created an **automated ETL pipeline** using GitHub Actions to fetch weather and market data every morning at 6 AM IST.
• Integrated **Open-Meteo API** and commodity market feeds to generate actionable insights (irrigation, spraying, sell/hold).
• Designed an **interactive dashboard** with price trends, weather conditions, and AI-generated recommendations.
• Improved data freshness and accessibility through **cloud storage**, automated updates, and API-driven workflows.

Data Cleaning Assistant (Web App)

• Developed a **Stream lit web application** to automate data cleaning, profiling, and metadata generation.
• Implemented **missing value handling, outlier detection (IQR), text standardization, filtering, sorting**, and grouping tools.
• Added **metadata editor**, enabling analysts to modify column types and export JSON/Excel metadata.
• Included **quick EDA features** like summary statistics, missing value charts, and correlation heatmaps.
• Improved data preparation speed with **undo/redo**, action logs, dataset merging, and export (CSV/Excel/JSON).

Billing Loss & AI Insights Dashboard – Power BI Project

- Developed a complete Power BI solution to analyze **1.2M+ rows** of employee timesheet data, identifying **1,000+ hours** of unbilled work and major revenue leakage.
- Designed a star schema data model that improved query performance by **40%** and reduced report load time.
- Built 12+ DAX measures to calculate **billing loss**, **utilization %**, **revenue loss**, and **project-level impact**, increasing reporting accuracy by **98%**.
- Integrated AI features (Forecasting, Key Influencers, What-If Analysis, Smart Narrative, Q&A) that reduced manual analysis time from **3–4 hours to under 10 minutes**.
- Helped simulate revenue improvements with what-if models, enabling managers to project up to **₹14M** in recoverable revenue after optimization.
- Automated refresh and insights delivery, reducing MIS team workload by **60%** and improving decision-making speed across departments.

Healthcare Fraud Detection (Machine Learning)

• Built **Logistic Regression** and **Random Forest** models to classify fraudulent healthcare claims.
• Performed **EDA and data cleaning**, identifying class imbalance issues common in fraud detection.
• Evaluated models using **Accuracy, Precision, Recall, F1-Score, Confusion Matrix, and ROC-AUC**.
• Identified low recall for minority fraud class and proposed **SMOTE, class weighting**, and boosting models to improve detection rates.
• Demonstrated practical challenges in **fraud analytics**, emphasizing the importance of recall in high-risk domains.

Customer Churn Prediction

• Built churn prediction models using **Logistic Regression** and **Random Forest** to classify high-risk customers.
• Cleaned and transformed data with **encoding, scaling, and exploratory analysis** to improve model robustness.
• Evaluated performance using **Accuracy, Confusion Matrix, and ROC-AUC** (≈0.84).
• Identified customer behavior patterns impacting churn, enabling **data-driven retention strategies**.
• Compared model performance to understand linear vs non-linear relationships influencing customer churn.

Crop Yield Prediction (1 million Rows)

• Processed a **1,000,000-row agricultural dataset** to predict crop yield using Linear Regression.
• Converted heavy CSV files to **compressed Parquet format**, reducing file size and improving load times.
• Performed **feature encoding, train-test split, and correlation analysis** to identify key predictors.
• Achieved strong baseline accuracy with **$R^2 = 0.913$** and low MSE, demonstrating effective regression modeling.
• Showcased capability to handle **large datasets, preprocessing pipelines, and agricultural analytics workflows**.

INTERNSHIPS

Accenture Virtual Internship – Data Analytics

Tata Virtual Internship – Business & Data Analytics

CERTIFICATIONS

Python for Data Analysis (Free Code Camp)

Data Analyst Essentials (Cisco)

Python Programming (GUVI) Prompt Engineering (IBM)

EDUCATION

B.Sc. Computer Applications – Rayalaseema University (2018–2022)