Manoj Kumar Ashok

(312)-284-9898 | mashok@depaul.edu | linkedin.com/in/MKASHOK/ | GitHub | Portfolio | New York NY / Chicago IL

Professional Summary

Data Scientist with strong SQL and Python expertise, experienced in building predictive models, analytics pipelines, and automated reporting workflows. Hands-on experience with regression-based forecasting, statistical testing, and data quality validation across large, messy datasets. Skilled at translating model outputs into clear, actionable insights for technical and non-technical stakeholders.

Technical Skills

Concepts: Predictive Modeling, Regression, Forecasting, Statistical Testing, Data Cleaning, ETL, Reporting Automation
Programming: Python, SQL, R, C++
Data Tools: Pandas, NumPy, Scikit-learn, PySpark, Athena, MySQL, PostgreSQL
Visualization: Tableau, Power BI, Matplotlib, Seaborn

Professional Experience

Research Assistant – Data Science Jan 2025 – Present
DePaul University *Chicago, IL*

- Performed data preprocessing, cleaning, and feature engineering on large-scale datasets to support predictive modeling and statistical analysis.
- Developed Python-based workflows to automate data validation, transformation, and reproducible analysis pipelines.
- Analyzed model outputs, evaluated performance metrics, and documented results to support measurement accuracy and reporting.

Technical Support Engineer – Data Engineering Jan 2023 – Jul 2023
Zoho Corporation *Chennai, India*

- Wrote and optimized SQL queries to clean, transform, and validate large CRM and financial datasets across staging and production systems.
- Improved SQL query performance by 35% through indexing, schema tuning, and query refactoring, enabling faster analytics and reporting.
- Built Python-based ETL workflows to automate ingestion, transformation, and KPI reporting pipelines.
- Partnered with analysts and stakeholders to ensure data accuracy and consistency for dashboards and operational metrics.

Projects

NYC Taxi Tip Percentage Prediction 2024
Python, SQL, PySpark, AWS Athena

- Cleaned and transformed 3M+ NYC taxi trip records using Python and PySpark for feature engineering and exploratory analysis.
- Built regression-based forecasting models to predict tip percentage, improving baseline RMSE by 18%.
- Validated datasets and model inputs using SQL queries to support reliable reporting and downstream analysis.

HAM10000 Skin Lesion Classification 2025
Python, Scikit-learn, Pandas

- Preprocessed 10k+ dermoscopic images with augmentation, normalization, and label encoding.
- Evaluated multiple ML classifiers and improved accuracy using feature engineering and tuning.
- Automated training and evaluation using modular Python scripts.

A/B Testing Simulation Engine 2025
Python, SQL, NumPy

- Generated synthetic experiment datasets simulating user behavior across variants.
- Built SQL-based evaluation pipelines to compute CTR, retention, lift, and experiment metrics.
- Evaluated experiment stability and statistical significance to support data-driven decision-making.

Education

DePaul University Chicago, IL
M.S. in Data Science – Computational Methods *Jan 2024 – Dec 2025*

- **Coursework:** Mining Big Data, Database Processing, Advanced ML, Regression, NLP, Neural Networks.

Bharathiar University Coimbatore, India
Bachelor of Computer Applications – ML & AI *Jul 2020 – Nov 2023*

- **Coursework:** Data Structures, Python Programming, AI & ML, Statistics, NLP, Computer Vision.