

Dinesh Sammeta

Data Engineer | SQL, Python | Technology

London, UK | sammetadineshkumar@gmail.com | +44 (0) 7442056804 | [LinkedIn](#)

Data Engineer with over 2 years of industry experience and recent academic training in cloud-native data solutions. Delivered an end-to-end real-time sentiment analysis pipeline using Apache Spark, Airflow, and AWS tools, simulating business applications for social data insights. Proficient in Python, SQL, and cloud platforms including AWS and Azure, with hands-on experience in data modelling, pipeline orchestration, and LLM integration. Strong foundation in scalable ETL design, stakeholder-facing dashboards, and cross-functional project delivery.

AREAS OF EXPERTISE

- Python, SQL, Bash
- Java, Scala
- Apache Spark, Redshift
- EMR, Kinesis, Hadoop
- Apache Airflow (basic)
- Pandas, Excel
- Google Data Studio (basic)
- Google BigQuery
- PostgreSQL, Numpy
- MySQL, Neo4J, MongoDB
- Scikit-learn, TensorFlow
- Deep Learning (LSTM, CNN)
- Hyperparameter Tuning
- Data Validation
- Log Analysis (basic)
- Azure Data Factory
- Azure Synapse
- AWS S3, AWS EMR
- AWS Kinesis
- Databricks (familiar)
- Anomaly Detection
- Outlier Handling
- AWS CDK (basic)
- Git, REST APIs
- AWS SDK (boto3)
- Jupyter, Flask
- Data Warehousing Concepts
- React, Linux
- Statistical Analysis
- Data Storytelling

PROFESSIONAL EXPERIENCE & PROJECTS

Inventory & Data Quality Analyst (ICQA) | Amazon | Mansfield, UK

Aug 2024 – Present

- Investigated over 120 monthly stock discrepancies using Amazon's WMS, RF scanners, and Fishbone diagrams, enabling faster issue flagging and improved root cause resolution.
- Built and maintained cycle count dashboards in QuickSight and Excel, supporting KPI reviews across 3 fulfilment zones and informing Ops team decisions during weekly audit cycles.
- Applied Pareto analysis and A3 templates to streamline issue workflows, reducing average investigation time by 5 days and maintaining a 98% SLA compliance rate.
- Queried inventory movement logs with SQL and supported exception handling in robotic zones, contributing to the resolution of 75+ Pod Bin misplacements within a quarter.

Data Analyst – Business Insight Engine Project | University of Leicester | Leicester, UK

Sep 2023 – Dec 2023

- Designed and deployed a full-stack sentiment analysis pipeline combining real-time data ingestion, deep learning, and LLM-based reasoning, enabling business stakeholders to monitor customer sentiment at scale with high data reliability.
- Built and orchestrated an end-to-end pipeline using Python, Apache Spark, and Airflow, with custom DAGs, monitoring, and alert systems to ensure data validation and timely transformation of Twitter data streams.
- Simulated a cost-effective cloud architecture using AWS S3, Redshift, Lambda, Glue, and boto3 SDK, converting unstructured JSON logs into clean, structured datasets suitable for downstream analytics.
- Applied advanced models (LSTM, CNN, hybrid LSTM-CNN) and integrated GPT-4 via LangChain to produce sentiment labels and justifications, with enhanced accuracy and generalisability in low-data contexts through zero-shot, one-shot, and few-shot learning.
- Developed a responsive user interface using Flask and React, enabling real-time prediction and reasoning, supported by built-in anomaly detection and missing data handling to maintain consistent performance.
- Followed Agile development practices, tracked project milestones via Azure DevOps boards, and maintained a modular, well-documented codebase to ensure code quality, reusability, and collaboration efficiency.

Data Analyst – Housing Price Prediction Project | University of Leicester | Leicester, UK

Jan 2023 – Apr 2023

- Consolidated over 25,000 records of housing market data from Kaggle and public APIs, cleaning and standardising datasets across 15+ regions using Python (Pandas) to build a foundation for accurate trend modelling.
- Conducted in-depth EDA using Matplotlib and Seaborn, identifying high-value predictors such as proximity to city centres, school ratings, and floor area, uncovering insights that could inform real estate investment strategies.
- Developed and optimised Random Forest and XGBoost regression models to predict housing prices, achieving ~90% R^2 score, with model outputs capable of informing property valuation and pricing decisions for estate agents and developers.
- Presented data-driven recommendations and regional pricing forecasts via interactive dashboards and visual reports, highlighting areas with up to 15% projected year-on-year appreciation, supporting portfolio planning for investors and mortgage brokers.
- Delivered a proof-of-concept that could reduce market research time by 30+ hours per analyst per month, demonstrating tangible potential for automation of pricing models in proptech or residential lending environments.

Data Engineer | Wipro Technologies | Hyderabad, India

June 2014 – July 2016

- Engineered and maintained ETL pipelines to process ~5M records per day of transaction and customer behaviour data for a Tier 1 telecom client using Informatica and SQL Server, which reduced reporting lag by 72 hours.
- Collaborated with client-side business analysts to develop customer segmentation logic for digital marketing campaigns, resulting in a regional 12% uplift in upsell conversions.
- Diagnosed and resolved data quality issues affecting monthly finance reports by leading root cause investigations, building SQL validation scripts and working cross-functionally with QA and DevOps teams.
- Participated in the early-stage design of a cloud migration proof of concept supporting schema mapping and transformation testing, which helped shape the client's move to a hybrid architecture.

EDUCATION

MSc in Advanced Computer Science | University of Leicester – Leicester, UK

B.Tech in Electrical and Electronics Engineering | VIT University – Vellore, India