

Kaustubh Sunil Khedekar

+18573507158 | kaustubh202500@gmail.com | Boston, MA | LinkedIn | GitHub | Portfolio | Medium | Tableau

Summary

Data Analyst and Engineer with 3+ years of experience in optimizing ETL pipelines, cloud infrastructure, and data visualization. Proficient in Python, SQL, Snowflake, Tableau, and AWS, with expertise in data modeling, statistical analysis, and workflow automation. Skilled in transforming datasets into actionable insights and driving cost-effective, data-driven decisions

Skills

Languages & Databases: Python (pandas, NumPy, Matplotlib, seaborn), R, SQL (MSSQL, T-SQL, PL SQL), NoSQL (MongoDB)
Cloud: AWS (EC2, Lambda, S3, RDS, Dynamo DB, Step Functions, Redshift, EventBridge, CloudWatch, SNS, SES), Snowflake, GCP, Azure
Data Engineering: ETL Pipelines, Data Cleaning, Schema Migration, Incremental Loads, Dimensional Modeling, Data Warehousing
Development Tools: Alteryx, SSIS, Docker, Git, GitHub, LangChain, OpenAI, Apache Airflow, Streamlit, Jupyter Notebooks, Visual Studio Code, Apache Kafka, ER Studio, NLP, SAS, MATLAB, MS Suite, Excel (Macros, VLOOKUP, Pivot tables)
BI Tools: Tableau (LOD Expressions, Filters, Parameters, Heatmaps), PowerBI (DAX, Slicers, Conditional Formatting), QuickSight
Data Analysis: Time-Series Analysis, Sentiment Analysis, Trend Analysis, Predictive Modeling

Experience

KeelWorks Foundation | Boston, MA | Data Analyst

Feb 2024 – Present

- Developed a chatbot application with Streamlit, Python, and BERT model to answer user queries and FAQs for website visitors
- Evaluated 500+ queries to identify behavior trends using sentiment analysis, topic modeling, and text classification techniques
- Facilitated secure data handling by automating PII extraction workflows through encrypted APIs and data masking methods, improving communication efficiency by 40%
- Streamlined employee activity reviews by analyzing activity logs, archiving 30 inactive members, and optimizing team structures
- Built dashboards in Tableau and Power BI with DAX calculations, slicers, and KPIs to track project performance and team productivity
- Migrated cloud infrastructure by dockerizing components, employing container orchestration with Docker and ECR, optimizing CI/CD pipelines, shifting infrastructure from AWS to Linode, reducing cloud costs by 67%

Bridge to Calculus | Boston, MA | Data Analyst Intern

Jun 2023 – Dec 2023

- Organized and analyzed 30-year data of students and alumni using SQL joins, data normalization, and data profiling techniques
- Leveraged pandas profiling to generate a report on 28K+ records, identifying missing values, duplicates, and schema mismatches
- Enhanced quality by 40% through cleaning, dimensional modeling, and integrating data warehouse solutions in Snowflake
- Assessed donation data in R using methods like chi-square, t-test, and ANOVA to test alumni engagement and donation patterns
- Implemented Tableau dashboards using LOD expressions, calculated fields, and filters for analyzing student performance, donations and resource allocation, improving decision making by 30%
- Proposed and improved forecasting accuracy with time-series analysis in Tableau, predicting \$37K in donations vs. \$28K manually

Hexaware | Mumbai, IN | Data Engineer

Sep 2020 – Nov 2021

- Designed and optimized ETL pipelines by using SSIS packages to migrate 15M records from a legacy system to MSSQL database
- Migrated SSIS workflows from EC2 to RDS clusters, improving efficiency and achieving 60%-time reduction and 40% cost savings
- Automated data retrieval with AWS EventBridge rules, Lambda and Step functions reducing delays with manual triggers
- Utilized SQL Server Profiler to optimize inefficient queries using execution plan analysis, indexing, and partitioning, reducing query execution time by 70%
- Managed 3 databases (Sales, Operation, Finance), configured role-based access control (RBAC), and ensured data security through audit logs using SQL server audit and data encryption
- Launched point-in-time backups on AWS RDS with multi-AZ replication ensuring data integrity and reliable recovery strategies

CitiusTech | Mumbai, IN | Data Engineer

Dec 2019 – Aug 2020

- Engineered ETL pipelines in Airflow, utilizing DAGs for task scheduling, monitoring, and integrating data from healthcare sources
- Automated healthcare inquiry workflows with batch processing, data enrichment, and Snowflake, reducing time by 50%
- Applied statistical analysis on health care training session data using SciPy, statsmodels, and feature engineering to improve training effectiveness by 10%
- Generated Tableau dashboards with filters, heatmaps, and KPI trends to assess workshop performance and resource allocation
- Conducted Python training sessions for 30+ healthcare professionals, introducing control structures, data visualization tools, and object-oriented programming concepts

Education

Northeastern University | Boston, MA | GPA: 3.7/4.0

Jan 2022 – Dec 2023

Master of Science in Industrial Engineering

Certifications: AWS Solutions Architect Associate, AWS Cloud Practitioner

Projects

Stock Market Analysis | Lambda, Docker, ECR, EventBridge, SES, Dynamo DB, Python, HTML, CSS

IPL Chatbot | OpenAI, Snowflake, Streamlit, Python, HTML