

Kaustubh Sunil Khedekar

+18573507158 / khedekarskaustubh@gmail.com / LinkedIn / GitHub / Tableau

Education

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

Jan 2022 – Dec 2023

Master of Science in Industrial Engineering

Coursework: Database Management & Database Design, Intro to Programming for Data Science, Data Mining, Computation & Visualization, Economic Decision Making, Probability & Statistics

Certification: AWS Solutions Architect Associate, AWS Cloud Practitioner

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
- Analyzed patterns across 500+ queries to identify behavior trends and popular topics, increasing content relevance by 35%
- Facilitated PII extraction via chatbot, allowing coordinators to reach out directly, improving communication efficiency
- Streamlined employee activity reviews by analyzing activity logs, archiving 30 inactive members, and optimizing team structures
- Analyzed project KPIs using PowerBI to evaluate results, assess effectiveness, identify improvements, and inform decision-making
- Optimized cloud infrastructure by dockerizing components, 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 to check program effectiveness and enhance fundraising efforts
- Leveraged Python for pandas profiling to generate a report on 28K+ records, identifying missing values, duplicates, and outliers
- Boosted quality by 40% through cleaning, standardization, normalization, and dimensional modeling in MSSQL and ER Studio
- Assessed donation data using statistical methods like chi-square, t-test to test alumni engagement and donation patterns using R
- Created Tableau dashboards for student performance, donations and resource allocation, improving decision-making by 30%
- Proposed and improved forecasting accuracy by using Tableau's trend analysis, predicting \$37K in donations vs. \$28K manually

Milestone Solutions PLM | Mumbai, IN | Data Engineer

Sep 2020 – Nov 2021

- Designed ETL pipelines by using SSIS packages to migrate 15M records from a legacy system to MSSQL database
- Optimized migration by moving SSIS package from EC2 to RDS clusters achieving 60% -time reduction and 40% cost savings
- Automated data retrieval by creating EventBridge rules and Lambda for CRON schedules reducing delays with manual triggers
- Utilized SQL Server Profiler to optimize inefficient queries, reducing query execution time by 70% through indexes and CTEs
- Managed 3 databases (Sales, Production, Finance), configured access levels, and ensured data security through SQL Server Audits
- Launched point-in-time backups on AWS RDS, ensuring data integrity and protection against unauthorized access

Robokart | Mumbai, IN | Data Engineer

Dec 2019 – Aug 2020

- Engineered ETL pipelines in Airflow to integrate data from 3 sources: student performance, workshop, and customer feedback data
- Automated student inquiry data workflows with Airflow and enhanced data quality using Snowflake, cutting manual time by 50%
- Applied statistical analysis on 1-on-1 session data improving teaching methods and increasing student participation by 10%
- Generated Tableau dashboards on student and teacher performance metrics, workshop effectiveness for robotics skills analysis
- Boosted sales team productivity by analyzing KPIs like vendor suitability, best-selling equipment and cost analysis in PowerBI
- Led 1-on-1 sessions in Python, JavaScript game design, and application building sessions, impacting 25 students' learning

Skills

Languages & Databases: Python, R, SQL (MSSQL, T-SQL, PL SQL), NoSQL (MongoDB)

Cloud Platforms: AWS (EC2, Lambda, S3, RDS, Dynamo DB, ECR, Redshift, EventBridge, CloudWatch, SNS, SES), GCP, Azure

Development Tools: Alteryx, SSIS, Docker, Snowflake, Git, GitHub, Apache Airflow, Streamlit, Jupyter Notebooks, Visual Studio Code, Apache Kafka, ER Studio, NLP, SAS, MATLAB, MS Suite, Excel (Macros, VLOOKUP, Pivot tables)

Visualization/BI Tools: Tableau, PowerBI, QuickSight

Projects

Stock Market Analysis | Lambda, Docker, ECR, EventBridge, SES, Dynamo DB, Python, HTML Feb 2024 – Mar 2024

- Developed a Dockerized Lambda function to analyze stock market data from API for various timeframes (7, 30, 180, 365 days, and YTD)
- Integrated Amazon SES with Lambda to email HTML-formatted stock market insights to recipients, with addresses stored in DynamoDB
- Configured Amazon EventBridge to trigger Lambda function every morning at 8 AM EST, ensuring timely and consistent data retrieval
- Incorporated data validation and error handling in Lambda function to handle API failures and ensure reliable data processing

Accidents in NYC | Python, MSSQL, Alteryx, Tableau

Jun 2023 – Sep 2023

- Performed comprehensive data profiling using pandas profiling to check data quality, followed by staging 10M records in MSSQL
- Executed ETL pipelines in Alteryx with effective parallel processing solution, achieving a 60% reduction in execution time
- Produced a snowflake dimensional model with 2 facts and 12 dimensions for data organization and analysis
- Designed 3 Tableau dashboards to analyze causes and patterns of accidents, for developing effective preventive measures