

SANTOSH GAIRE

gaire.2002@gmail.com | GitHub

SUMMARY

Aspiring Analytics Engineer with hands-on experience in SQL-based data transformations, ETL/ELT pipeline development, and data visualization. Skilled in Python, SQL, and data modeling with a growing proficiency in modern analytics engineering tools including dbt, Snowflake, and Git workflows. Passionate about building governed, high-quality datasets that drive business decisions, and committed to continuous learning in data engineering and financial technology.

EXPERIENCE

Data Engineering Training and Personal Projects

- Designed and implemented a data warehouse using Star Schema and Medallion Architecture (Bronze → Silver → Gold), while evaluating its scalability compared to cloud-native platforms like Snowflake, BigQuery, and GCP for enterprise analytics readiness
- Built ETL pipelines using Python and SQL to extract, transform, and load data from CRM and ERP systems, ensuring data models align with stakeholder requirements through an agile-inspired development process to facilitate cross-functional collaboration
- Applied data modeling concepts including dimensional modeling, star schemas, and surrogate keys to support reporting, analytics, and business decision-making
- Managed version control and CI/CD workflows using Git and GitHub, incorporating branching strategies and pull request reviews to maintain high-quality, deployable code
- Developed proficiency in Docker, PostgreSQL, and T-SQL through hands-on data engineering projects and self-directed learning
- Gained foundational exposure to dbt and modern analytics engineering practices including data transformations, testing, and documentation
- Troubleshoot and resolved inconsistent data quality issues in CRM and ERP CSV files by analyzing source patterns, validating assumptions, and implementing custom Python cleaning routines to ensure accurate downstream transformations

EDUCATION

Birendra Multiple Campus, Nepal – Nepal	2026
Bachelor of Science in Computer Science and Information Technology (Pursuing)	

LEARNING & CERTIFICATIONS

Data Engineering Program – Dlytica	completed
------------------------------------	-----------

SKILLS

Python, SQL, Data Pipelines, Matplotlib, Seaborn, Pandas, NumPy, MySQL, PostgreSQL, Cleaning, Transformation, Statistical Analysis, Problem-Solving, Critical Thinking, Communication, Attention to Detail, Data Modeling, Medallion Architecture, Data Warehousing, Version Control (Git/GitHub)

PROJECTS

Data Warehouse on SQL

- Designed a data warehouse using Star Schema with Medallion Architecture (Bronze → Silver → Gold) layers for scalable data transformation and analytics-readiness

- Built fact tables, dimension tables, and surrogate keys to support structured data modeling and business reporting
- Optimized SQL queries for efficient data retrieval, transformation, and warehouse performance

Diwali Market Analysis (Data Visualization)

- Cleaned and analyzed seasonal sales data using Pandas to identify key product demand trends and sales patterns
- Built data visualizations using Matplotlib and Seaborn to present business insights and seasonal market trends

Student Marks Analysis (Data Visualization)

- Processed and visualized student performance data using pandas and numpy
- Generated insights to identify top-performing and at-risk students

ETL Pipeline (Learning Project)

- Developed ETL pipelines using Python, SQL, and Apache Spark to extract, transform, and load data from CRM and ERP sources, utilizing AWS S3-compatible storage (MinIO) for scalable data staging and orchestrating workflows with Apache Airflow
- Performed data cleaning, transformation, and loading operations to prepare analytics-ready datasets for downstream reporting
- Researching MongoDB architectural patterns to understand its role in handling semi-structured data within modern data engineering workflows