

ATHARVA YEOLA

+1 8582203508 | [✉ atharvayeola12@gmail.com](mailto:atharvayeola12@gmail.com) | [in/atharvayeola](https://www.linkedin.com/in/atharvayeola) | [/atharvayeola](https://github.com/atharvayeola) [/atharvayeola](https://www.youtube.com/channel/UCv33333333333333333333)

SUMMARY

Data Engineer with experience designing scalable data pipelines, building analytics infrastructure, and transforming raw behavioral and operational data into high-impact reporting systems. Skilled in cross-functional collaboration, real-time data processing, and delivering reliable data assets for decision-making and product optimization.

EDUCATION

University of California San Diego (UCSD)

Master of Science (M.S) – Data Science & Machine Learning

Sep 2023 – Jun 2025

San Diego, USA

Pune Institute of Computer Technology (PICT)

Bachelor of Engineering (B.E) – Electronics & Telecommunication

Aug 2019 – Jun 2023

Pune, India

WORK EXPERIENCE

Data Engineer

Jul 2025 – Present

San Diego Natural History Museum

San Diego, USA

- Designed a modular star schema in dbt, transforming over 2 million specimen records from disparate MS Access databases into MySQL, enabling unified biodiversity analytics across 6+ scientific departments.
- Optimized ELT workflows by implementing field-level dbt tests and schema validations, reducing pipeline errors by at least 40% and improving downstream query performance for curators and researchers by 1.5x.

Data Scientist, Analytics

Jul 2024 – Jun 2025

Qualcomm Institute - UC San Diego Division of Calit2

San Diego, USA

- Headed the causal inference analysis of impact of the U.S. sports betting on gambling addiction via synthetic control and time series forecasting; quantifying a 23% nationwide increase in addiction.
- Created HIVTrends.org, a live dashboard that tracks weekly uptake of at-home HIV test kits; adopted by internal product and public health partners, and steering campaign-budget decisions.

Data Engineer, Product Analytics

Aug 2022 – Apr 2023

RhythmFlows Solutions Pvt. Ltd

Pune, India

- Engineered batch ingestion workflows to consolidate user event logs and in-app feedback; built training-ready datasets for recommender system optimization, improving engagement metrics (NDCG@10 boosted by 18%).
- Automated A/B test tracking for 20K+ user sessions, integrating result outputs into reporting pipelines; enabled rapid evaluation of ranking algorithms with CTR boosted by 33% and 20% increase in session time.

Data Scientist

Jan 2022 – Jun 2022

Atomic Loops

Pune, India

- Developed real-time data ingestion pipeline for conveyor object detection using Raspberry Pi and edge inference logs; tracked operational KPIs and supported data streams that led to a 30% increase in customer footfall.
- Built dashboards on top of structured output tables and integrated conveyor telemetry; improved latency-resilient synchronization across teams and drove a 12% MoM revenue boost with 5x food waste reduction.

PROJECTS

Superstore Analytics Pipeline | [Python](#), [Pandas](#), [PySpark](#), [SQLite](#), [Dash](#)

- Developed a medallion-architected ETL pipeline (Bronze→Silver→Gold) on 10k+ Superstore rows, automating daily parquet and SQLite table refreshes.
- Modeled a star schema in SQLite (dim_date, dim_category, dim_region, fact_sales), powering an interactive Plotly Dash dashboard of sales trends, margins, and top customers.

E-commerce Funnel Analysis | [PySpark](#), [SQL](#), [Plotly Dash](#), [Cohort Analysis](#), [RFM Segmentation](#)

- Processed 5 nested ZIPs of click-stream data with PySpark (2M+ events), computed session-level funnel metrics (view→cart→purchase) and identified a 96% cart-to-purchase drop-off point.
- Built advanced analytics—RFM segmentation, cohort retention tables, and 6-month sales forecast with Holt–Winters—exposed via a self-serve Dash app replacing manual Looker reports.

TECHNICAL SKILLS

Programming – SQL, Python, Bash, PySpark, Pandas, NumPy

Pipeline Design – Airflow, dbt, Snowflake, BigQuery, Redshift, MySQL, PostgreSQL, ELT/ETL

Visualization Tools – Power BI, Tableau, Plotly, Metabase, Looker

Cloud & Deployment – AWS, GCP, Docker, Git, GitHub Actions, CI/CD, Data Studio

Other – Cosmos, Apache Spark, Kafka, Flink