MLDS 430 | Fall 2025 | Data Warehousing Final Project

Project: End-to-End Data Pipeline

Due Date: Tuesday, 12/9/25 at 11:59pm

Objective

Build a complete, end-to-end data pipeline and analytics project. Choose a topic you are genuinely interested in and that you can include in your portfolio. Your project should demonstrate your ability to extract, transform, load, and visualize data using modern tools and best practices.

Requirements

Tools

- **Required**: (Fivetran **OR** API), Snowflake, dbt, (Tableau **OR** Streamlit)
- **Optional**: Airflow (extra credit)

Pipeline Expectations

- **Extraction**: Use Fivetran to pull data from at least 1 dataset (API, S3 bucket, database, etc.) or connect directly to an API via Python
- **Storage**: Load the data into Snowflake for staging
- **Transformation**: Transform the data using at least 2 dbt models (each dbt model = one table in Snowflake)
- **Analytics / Visualization**: Build a dashboard in Tableau or an app in Steamlit that contains at least 2 meaningful components (visualization, interactivity, etc.)
- **Optional Automation**: Use Airflow to orchestrate your pipeline for extra credit

Deliverables

- **Github Repository**: Include all Python scripts, SQL scripts, dbt models and documentation explaining your project and workflow
- **Recording**: 2-3 minute video of you giving an overview of your pipeline and demoing your final dashboard or app, as you would explain to an interviewer

FAQs

- **Scope Minimums**
  - 1 dataset (not one used in class, such as Open Meteo or Wiki)
  - 2 dbt models
  - 1 dashboard / app
  - 2 meaningful visuals / components
- **Data**
  - It's fine to use a snapshot of a dataset
  - Using continuous or live-updating data with Airflow is extra credit
- **Best Practices**
  - Aim to use best practices throughout, including tests, documentation, well-organized folders, etc.
- **Video Recording**
  - Focus on explaining your project – storytelling is ***not required***
  - Make sure your video is ***no longer than 3 minutes***

Submission Details

Due: Tuesday, 12/9/25 at 11:59pm

Submit via Canvas:

- Github repo link
- Video recording

IMPORTANT: Grant adashofdata (instructor) & rsuhendra (TA) access to your Github repo

Grading Rubric – 20 points

Category	Points	Notes
Fivetran / API	3	Extract 1+ dataset correctly
Snowflake	3	Load & stage data properly
dbt	5	2+ models, correct transformations
Tableau / Streamlit	5	2+ visuals / components, interactive
Documentation	2	Comments, README, well-organized GitHub repo
Video Recording	2	2-3 min walkthrough of pipeline & dashboard / app
Extra Credit: Airflow	2	Optional orchestration
TOTAL	**20 + 2 EC**	20 core points, up to 2 extra credit points