



BUAN 6390.S01 S25 – ANALYTICS PRACTICUM

GROUP 6 – HEALTHCARE COSTS:
STOP GUESSING, START SAVING

PROJECT SPONSOR: VIJAY KONERU

PROFESSOR: MANDAR SAMANT

MILESTONE 2 HIGHLIGHTS: FROM COLLECTION TO CLEAN INSIGHTS

- **Data Collection:** Collected data from 8 cities using 10 CSV and 10 JSON files.
- **Architecture Implementation:** Followed the Medallion Architecture:
 - a. Bronze Layer: Stored raw ingested data
 - b. Silver Layer: Cleaned and structured the data
 - c. Gold Layer: Aggregated and enriched data for analysis
- **Data Preparation:** Performed cleaning, transformation, and aggregation.
- **Output:** Validated that our superset is ready for visualization.

MILESTONE 3: ADVANCING FROM STRUCTURE TO SOLUTION

- Orchestrated data pipeline with Apache Airflow
- Built DAGs for ingestion, cleaning, enrichment, and aggregation
- Developed dashboards to compare healthcare costs
- Finalized documentation and deliverables



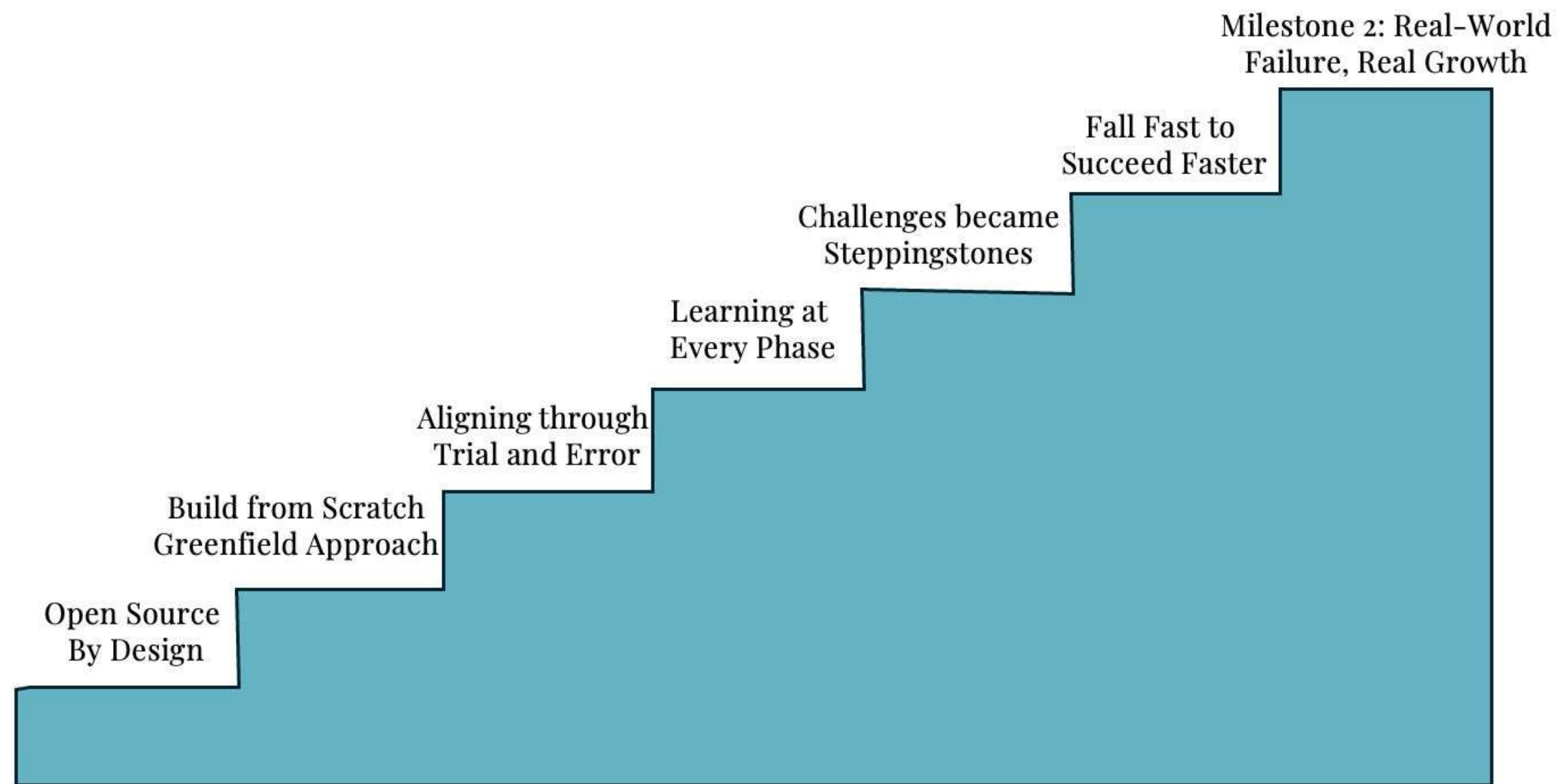
FROM DATA TO STORY: DASHBOARD DEMONSTRATION



**SEE IT IN ACTION:
OUR PROJECT IN MOTION**



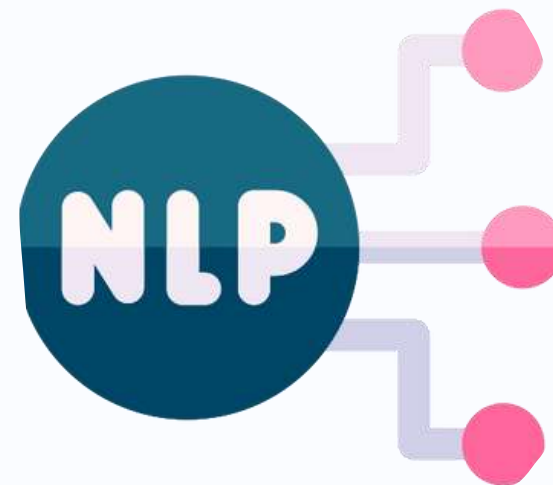
MILESTONE MOMENTS: TRACING OUR PROJECT JOURNEY



LOOKING AHEAD: FUTURE ENHANCEMENTS AND OPPORTUNITIES



Optimize and scale
Airflow pipelines,
with more refined
and professional data
modeling



Integrate NLP
models to categorize
treatments using
service descriptions



MACHINE LEARNING

Apply ML models for
cost prediction and
provider benchmarking
(e.g., Linear Regression,
Random Forest, XG
Boost)

The background is a light blue gradient. It features several medical-themed illustrations: a stethoscope on the right side, a first aid kit with a handle and a cross symbol on the bottom left, a blister pack of pills on the top left, a syringe with a cross on the top center, and two spilled pill containers with pills on the top right. There are also four plus signs scattered across the background.

THANK YOU


```
dag_id=master_healthcare_pipeline_run_id>manual__2025-05-06T16_01_33.987292+00_00_task_id=...
neels-macbook-air.local
*** Found local files:
*** * /Users/neelkalavadiya/Practicum_Project_Local/airflow/logs/dag_id=master_healthcare_pipeline/run_id>manual__2025-05-06T16:01:33.987292+00:00/task_id=enric
*** * /Users/neelkalavadiya/Practicum_Project_Local/airflow/logs/dag_id=master_healthcare_pipeline/run_id>manual__2025-05-06T16:01:33.987292+00:00/task_id=enric
[2025-05-06T11:02:13.474-0500] {taskinstance.py:1956} INFO - Dependencies all met for dep_context=non-requeueable deps ti=<TaskInstance: master_healthcare_pipelin
manual__2025-05-06T16:01:33.987292+00:00 [queued]>
[2025-05-06T11:02:13.477-0500] {taskinstance.py:1956} INFO - Dependencies all met for dep_context=requeueable deps ti=<TaskInstance: master_healthcare_pipeline.en
manual__2025-05-06T16:01:33.987292+00:00 [queued]>
[2025-05-06T11:02:13.477-0500] {taskinstance.py:2170} INFO - Starting attempt 1 of 1
[2025-05-06T11:02:13.482-0500] {taskinstance.py:2191} INFO - Executing <Task(PythonOperator): enrich_bronze_table> on 2025-05-06 16:01:33.987292+00:00
[2025-05-06T11:02:13.490-0500] {standard_task_runner.py:60} INFO - Started process 9990 to run task
[2025-05-06T11:02:13.494-0500] {standard_task_runner.py:87} INFO - Running: ['airflow', 'tasks', 'run', 'master_healthcare_pipeline', 'enrich_bronze_table', 'manu
id', '535', '--raw', '--subdir', 'DAGS_FOLDER/Master_DAG.py', '--cfg-path', '/var/folders/zh/dswmqvz11lg5pwpvlk3b9zz80000gn/T/tmpqggu1e37']
[2025-05-06T11:02:13.496-0500] {standard_task_runner.py:88} INFO - Job 535: Subtask enrich_bronze_table
[2025-05-06T11:02:13.500-0500] {local_task_job_runner.py:234} INFO - Task exited with return code Negsignal.SIGKILL
[2025-05-06T11:02:13.515-0500] {taskinstance.py:3280} INFO - 0 downstream tasks scheduled from follow-on schedule check
[2025-05-06T11:02:13.752-0500] {scheduler_job_runner.py:1766} ERROR - Detected zombie job: {'full_filepath': '/Users/neelkalavadiya/Practicum_Project_Local/airflo
Users/neelkalavadiya/Practicum_Project_Local/airflow/dags', 'msg': '{"DAG Id": "master_healthcare_pipeline", "Task Id": "enrich_bronze_table", "Run Id": "manual__
'neels-macbook-air.local'}', 'simple_task_instance': <airflow.models.taskinstance.SimpleTaskInstance object at 0x11932c4c0>, 'is_failure_callback': True} (See htt
stable/core-concepts/tasks.html#zombie-undead-tasks)
```

ChatGPT 4o

Share

🧟 Why It Became a "Zombie Task"

Airflow calls a task a "zombie" when:

- The task was running, but the process disappeared without finishing cleanly, and
- The scheduler doesn't get a completion status (success/failure).

This usually happens because of:

1. **Out-of-memory kill (OOMKill):** macOS or Python forcibly killed your task because it consumed too much RAM (common with Spark tasks).
2. **Manual kill:** You accidentally killed the process (`kill -9`) or force-quit Python.
3. **System interruption:** MacBook went to sleep or ran out of system resources.

