ETL Telecom Project - Documentation

This document summarizes how to deploy and run the full ETL Telecom project.

1. Open SQL Server Management Studio (SSMS).
2. Execute the SQL scripts in the following order:
3. 1) 01. Create database.sql
4. 2) 02. Create fact transaction table.sql
5. 3) 03. Create error destination output.sql
6. 4) 04. Create dim imsi.sql
7. 5) 05. Create error source output.sql
8. 6) 06_create_staging_and_sp.sql
9. Edit process_etl.py: update INCOMING_DIR, PROCESSED_DIR, FAILED_DIR, and DATABASE_CONNECTION with your environment settings.
10. Install dependencies: pip install pyodbc
11. Place input CSV files into the INCOMING_DIR folder.
12. Run the ETL script: python process_etl.py
13. The script loads data into staging (stg_telecom_raw) and calls usp_ProcessStagingTelecom.
14. Valid rows go into fact_transaction; invalid rows go into error tables.
15. Processed files are moved to PROCESSED_DIR; failed files are moved to FAILED_DIR.
16. You can schedule process_etl.py with Task Scheduler (Windows) or cron (Linux) for automation.

Business Rules implemented in usp_ProcessStagingTelecom:

- IMSI must not be NULL.
- CELL and LAC must not be NULL.
- EVENT_TS must be a valid date (dd/MM/yyyy or yyyy-MM-dd HH:mm:ss).
- Subscriber ID is looked up from dim_imsi_reference; if not found, set to -99999.
- TAC = first 8 chars of IMEI if IMEI length >= 14, else -99999.
- IMEI stored as left 14 chars if valid, else -99999.
- SNR = right 6 chars of IMEI if valid, else -99999.

This completes the ETL Telecom project setup. You can extend it by adding scheduling, monitoring, and reporting.