

# Bank Marketing ETL Pipeline using Apache Airflow & PostgreSQL

## Project Overview

This project demonstrates an end-to-end ETL pipeline using Apache Airflow, Python, and PostgreSQL. The pipeline extracts raw bank marketing data, transforms and validates it, and loads analytics-ready tables.

## Technologies Used

- Apache Airflow
- PostgreSQL
- Python (Pandas, NumPy, SQLAlchemy)
- Docker & Docker Compose
- pgAdmin4

## ETL Process

**Extract:** Data is extracted from raw.bank\_marketing using SQLAlchemy.

**Transform:** Data cleaning, feature engineering, and schema normalization are applied.

**Validate:** Assertions ensure data quality and integrity.

**Load:** Cleaned datasets are loaded into the cleaned schema in PostgreSQL.

## Airflow Orchestration

A PythonOperator DAG orchestrates the full ETL flow: Extract → Transform → Validate → Load.

## Outputs

- cleaned.bank\_client
- cleaned.bank\_campaign
- cleaned.bank\_economics

## Skills Demonstrated

- Data Engineering & ETL Design
- Workflow Orchestration with Airflow
- Data Validation & Feature Engineering
- PostgreSQL & Dockerized Pipelines

## Author

Raffy Alcazar