## **Overview**

This repository contains two data engineering projects that focus on extracting, transforming, and loading (ETL) data from various sources into PostgreSQL and Google Cloud Platform (GCP). The projects utilize Python, Pandas, PostgreSQL, Google Cloud Storage (GCS), and BigQuery for data processing and storage.

## **Project 1: Data to PostgreSQL**

Project 1 is designed to extract data from CSV and JSON sources, clean it, and insert it into a PostgreSQL database.

### *Workflow*

1. **Extract Data:**
   * Reads a CSV file (ev\_population\_data.csv).
   * Fetches JSON data from an API (https://data.cityofchicago.org/resource/ydr8-5enu.json).
2. **Transform Data:**
   * Basic cleaning and standardizing the data using Pandas.
3. **Load Data into PostgreSQL:**
   * Inserts cleaned data into specific PostgreSQL tables:
     + ev\_population\_data.ev\_data
     + permits\_data.permits

## **Project 2: PostgreSQL to Google Cloud (GCS & BigQuery)**

Project 2 extends the data pipeline by extracting data from CSV, JSON, and PostgreSQL, uploading it to Google Cloud Storage (GCS), and loading it into BigQuery.

### *Workflow*

1. **Extract Data:**
   * Reads CSV and JSON files similar to Project 1.
   * Fetches data from PostgreSQL
2. **Transform Data:**
   * Cleans and standardizes data.
3. **Upload to GCS:**
   * Converts data to Parquet format and uploads it to a GCS bucket.
4. **Load into BigQuery:**
   * Loads the data from GCS into BigQuery tables:
     + ev\_population\_data.ev\_data
     + permits\_data.permits
     + postgres\_data.permits\_postgres

## **Conclusion**

These projects demonstrate a basic ETL pipeline for moving data from local and remote sources into cloud storage and databases, enabling further analytics and processing in Google Cloud.