ETL Project

Project Description
In this project, you will embark on a journey to extract, transform, and load (ETL) crime data to uncover insights and trends. Crime data is a valuable resource for understanding patterns, behaviors, and factors that contribute to criminal incidents. Through this project, you will gain hands-on experience in data preprocessing, manipulation, and analysis, while also honing your skills in documentation and presentation.

Dataset Description: The dataset you have contains information about incidents, offenses, and related details. The columns in your dataset provide various pieces of information about each incident:

- **INCIDENT_NUMBER:** A unique identifier for each incident.
- **OFFENSE_CODE:** Code representing the type of offense.
- **OFFENSE_CODE_GROUP:** Group/category of offenses.
- **OFFENSE_DESCRIPTION:** Description of the offense.
- **DISTRICT:** District where the incident occurred.
- **REPORTING_AREA:** Area within the district where the incident was reported.
- **SHOOTING:** Indicates whether a shooting was involved in the incident.
- **OCCURRED_ON_DATE:** Date and time when the incident occurred.
- **YEAR, MONTH, DAY_OF_WEEK, HOUR:** Temporal components of the incident date.
- **UCR_PART:** UCR (Uniform Crime Reporting) category of the offense.
- **STREET:** Street where the incident occurred.
- **Lat, Long:** Latitude and longitude coordinates of the incident location.
- **Location:** Geographical location in a format that combines latitude and longitude.

Project Setup:
1. Environment Setup: Set up the development environment. Install the necessary libraries, such as Pandas for data manipulation and transformation.
2. Data Source: Make sure you have access to the dataset. It could be stored as a CSV file, a database table, or another structured format.
3. Data Storage: Decide where you'll store the cleaned and transformed data. This could be a database, data warehouse, or even a simple flat file.
5. ETL Process: Develop separate scripts (Python files) for each phase of ETL: extraction, transformation, and loading.
6. Automated ETL Workflow (Optional): Create a load strategy to AWS
7. Documentation: Create documentation to explain the purpose of the project, the dataset's meaning, your ETL process, and the steps taken in each phase. This documentation should be placed in the **documentation/** directory.