Hospital Emergency Department Data Processing

User Story

Title: Streamlining Data Processing for Hospital Emergency Department Encounters

As a Data Analyst

I want an efficient ETL (Extract, Transform, Load) process

So that I can quickly and accurately clean and prepare data for analysis and reporting

Acceptance Criteria:

1. Data Extraction

- The system should be able to read raw data from a CSV file containing hospital emergency department encounters.

2. Data Transformation

- The system should handle missing values by dropping rows with any missing data.
- The system should remove any duplicate records.
- The system should log the shape of the data before and after the transformation steps.

3. Data Loading

- The system should save the transformed data into a new CSV file in a specified directory.

4. Execution

- The ETL process should be executable via a Python script (etl_process.py).

Hospital Emergency Department Data Processing

- The ETL process should be demonstrated in a Jupyter notebook (3_etl_process.ipynb).

Scenario:

Given the data analyst has a raw CSV file named hospital-emergency-department-encounters-by-facility.csv

When the data analyst runs the etl_process.py script

Then the script should read the raw data, drop missing values, remove duplicates, and save the cleaned data to hospital_emergency_transformed.csv in the processed directory

And the Jupyter notebook should provide a step-by-step demonstration of the ETL process, showing the data before and after transformation.