ETL Mini Project: A Collaborative Endeavor

Throughout our ETL mini project, our team of three collaborated to gain hands-on experience in constructing an ETL (Extract, Transform, Load) pipeline using Python, Pandas, and various techniques like Python dictionaries and regular expressions for data extraction and transformation.

Extract and Transform

We began by extracting raw data and converting it into a structured format. We then utilized Python and Pandas for data manipulation, using Python dictionaries and regular expressions to handle intricate data extraction and transformation tasks. We were able to efficiently clean and normalize the data, getting it ready for the next steps.

Create CSV Files

After properly transforming the data, we generated four CSV files, each representing a different aspect of our dataset. The CSV files provided a solid basis for conducting in-depth analysis and seamlessly integrating them into the database.

Create an Entity-Relationship Diagram (ERD) and Table Schema

After obtaining the transformed data, we created an Entity-Relationship Diagram (ERD) to visually represent the connections between various data entities. The ERD provided valuable guidance in creating a strong table schema, guaranteeing the integrity and relational structure of the data during its transfer to the database.

Importing Data into Postgres Database

To complete our project, we loaded the data from the CSV files we had generated into a Postgres database. We needed to use SQL commands and PostgreSQL features to ensure the data insertion process was smooth and accurate. After this phase, our data was seamlessly incorporated into the database, poised for thorough examination and analysis.

In conclusion,

Overall, this mini project gave us practical experience building an ETL pipeline. The project's collaborative approach enabled us to harness the unique strengths of each team member, resulting in a thorough grasp of ETL processes and database management.