

PROJECT SPECIFICATION

Data Lake

ETL

CRITERIA	MEETS SPECIFICATIONS
The etl.py script runs without errors.	The script, etl.py, runs in the terminal without errors. The script reads song_data and load_data from S3, transforms them to create five different tables, and writes them to partitioned parquet files in table directories on S3.
Analytics tables are correctly organized on S3.	Each of the five tables are written to parquet files in a separate analytics directory on S3. Each table has its own folder within the directory. Songs table files are partitioned by year and then artist. Time table files are partitioned by year and month. Songplays table files are partitioned by year and month.
The correct data is included in all tables.	Each table includes the right columns and data types. Duplicates are addressed where appropriate.

Code Quality

CRITERIA	MEETS SPECIFICATIONS
The project shows proper use of documentation.	The README file includes a summary of the project, how to run the Python scripts, and an explanation of the files in the repository. Comments are used effectively and each function has a docstring.
The project code is clean and modular.	Scripts have an intuitive, easy-to-follow structure with code separated into logical functions. Naming for variables and functions follows the PEP8 style guidelines.