Alpha AS

**Overview**

This project provides a Python script for extracting, transforming, and cleaning data from an Excel file, then loading it into a SQLite database for analysis. It's designed to support data analysis and visualization tasks, specifically with Power BI. The pipeline includes a date dimension table to enhance time-based analysis and generates a data quality report in Excel format.

**Setup and Installation**

**Requirements**

* Python 3.8+
* Libraries: pandas, sqlalchemy, sklearn, openpyxl
* pip install pandas
* pip install sqlalchemy
* pip install scikit-learn
* SQLite
* DB browser for SQLite
* ODBC Data Source Administrator
* ODBC Drivers
* PowerBI

**Database Setup**

* The script automatically creates a SQLite database named optimus\_data.db in the project directory. No additional database setup is required.

**Running the Script**

* Open the script using visual studio
* Before running the code make sure the excel file has sheet named DATA and the last empty column has been removed from the excel.
* Run the script and it will create optimus\_data.db file.
* Once the file is created check the created tables using DB browser for SQLite
* Open the DB browser for SQLite and go to open database and browse the optimus\_data.db file and browse the tables.
* Then open the ODBC Data Source Administrator go to user DSN and go to add and select the SQLite3 ODBC driver and add the source name and browse the optimus\_data.db file and OK
* Open PowerBI and navigate to get data and search for ODBC and select the datasource name created in the previous step
* Then done, now you can build the dashboard