Project Documentation

Alex Boshnakov

Table of Contents

Introduction

This documentation goes over implementation specifics and usage instructions for the dataset command line interface. The CLI was created with the Psycopg2 (PyPi link) library for Python.

The dataset supported by the CLI is called 'Airbnb Listings & Reviews' (Kaggle page). This dataset contains Airbnb data for over 250,000 listings in 10 major US cities. Information about hosts, reviews, pricing, location, and room type is included.

The dataset is on the PostgreSQL database management system. There are a total of eight tables in the database, all of which can be managed with the CLI program. The tables' names, descriptions, and attributes are as follows:

- **Host:** This table contains information about all the hosts that have at least one listing in this dataset.
    - **Attributes:** host_id, host_location
- **Guest:** This table contains information about all the guests that have stayed at any of the listings in this dataset.
    - **Attributes:** reviewer_id, listing_id, review_id

- **Listing:** This table contains general information about each listing in this dataset.
    - **Attributes:** listing_id, host_id, property_type, price, minimum_nights, maximum_nights, instant_bookable
- **Review:** This table contains information about every review created by guests.
    - **Attributes:** review_id, date
- **Rating:** This table contains the rating information for each listing in this dataset.
    - **Attributes:** listing_id, review_scores_rating, review_scores_accuracy, review_scores_cleanliness, review_scores_checkin, review_scores_location, review_scores_communication, review_scores_value
- **Profile:** This table contains more information for each host.
    - **Attributes:** host_id, host_since, host_response_time, host_response_rate, host_acceptance_rate, host_is_superhost, host_total_listings_count, host_has_profile_pic, host_identity_verified
- **Room:** This table contains information about the room that was turned into a listing.
    - **Attributes:** listing_id, name, room_type, bedrooms, accommodates, amenities
- **Location:** This table contains location information for any listing in this dataset.
    - **Attributes:** listing_id, latitude, longitude, neighbourhood, district, city

GitHub Page

This project has a GitHub page (link: alex-bosh/CMPSC431W-Project: Name: Alex Boshnakov (github.com)). On this page, you will find a README file outlining some general information about the project, the database from PostgreSQL, and the Python file used to create the command line interface.

Project Setup

The process of setting up the project is quick and easy! The steps are listed below:

1. Download the PostgreSQL database from the GitHub link above, connect to a SQL Server (PostgreSQL recommended), and import the database to your server.
2. Download the Python file and open it in a Python interpreter (Visual Studio Code recommended). Some packages will have to be installed, which are listed below:
    - Numpy
    - Psycopg2
3. After these packages are installed, the CLI program is ready to run! Be sure that you are connected to your database server. Near the top of the main file are fields for the server credentials. These are set to default values but change them accordingly if needed.
4. Run the program and feel free to make decisions based on the prompts you are given in the terminal.

Demo Video

To facilitate the process of using the command line interface, a tutorial video is provided (YouTube link: https://www.youtube.com/watch?v=PqwRKRP3Lew).

The timestamps for each functionality are listed below:

0:45 – Insert Data

2:00 – Delete Data

3:36 – Update Data

4:49 – Search Data

5:50 – Aggregate Functions

6:39 – Sorting

7:18 – Joins

8:08 – Grouping

9:00 – Subqueries

9:57 – Transactions

10:30 – Error Handling

11:03 - Exit