Lab 2: Airbnb Staging Tables

Deadline: Friday, Feb. 2nd at 11:59pm.

Goals:

In this first programming assignment, we will create a database for storing and exploring Airbnb data. The database will consist of a collection of tables and their relationships. We will define the schema based on the format of the input data and visualize it through an ERD. We will create the tables in SQL and load them with the copy command. We will also create foreign key constraints for referential integrity.

Input:

-Airbnb dataset: https://storage.googleapis.com/utcs-spr2018-datasets/airbnb/austin/2017-03-07.zip

Desired Outputs:

-ER Diagram of the Airbnb tables, relationships, and attributes.

-Airbnb database created in Postgres Cloud instance.

-Airbnb tables created, 1 table per data file.

-Airbnb tables populated with input data.

-Airbnb tables contain primary key and foreign key constraints.

Tools You Need:

-GitHub

-LucidChart

-Postgres Cloud SQL instance

-Postgres psql client

Programming Style:

-Upper case the first letter of a table (e.g. Calendar, etc.).

-Lower case the attribute names (e.g. host_id, etc.).

-Place each create table statement into its own file. For example, create_calendar.sql.

-Create a main script, create_tables.sql, that calls the individual create table scripts.

-Place all copy statements into a single file, import_data.sql.

-Place all alter table statements into a single file, alter_tables.sql.

-ERD should be done in Lucidchart and downloaded locally. Name the file airbnb_ERD.png.

Implementation Hints:

-Run the command **create database airbnb;** in psql to create a new database named airbnb.

-Run the command **\c airbnb** in psql to connect to the airbnb database.

-Create each table with a primary key.

-Use **gen_ddl.py** to auto-generate the initial create table statement.

-Add the primary key constraint to the create table statement.

-Use the command **\i <filename>** in psql to run a SQL command from a script. For example, **\i create_tables.sql**

-Use the copy command to import the data.

-Before running copy, run the command **set client_encoding to 'utf8';** in your psql client to avoid character encoding issues.

-Add the foreign key constraints after the tables have been populated.

-The Listings table is has ~90 attributes, choose a small subset of attributes to include in your ERD.

Documentation:

Create table command: https://www.postgresql.org/docs/9.6/static/sql-createtable.html

Copy command: https://www.postgresql.org/docs/9.6/static/sql-copy.html

Alter table command: https://www.postgresql.org/docs/9.6/static/sql-altertable.html

Snippets:

gen_ddl.py: https://github.com/cs327e-spring2018/snippets/blob/master/gen_ddl.py

ERD example: https://github.com/cs327e-spring2018/snippets/blob/master/Best%20Buy%20Staging%20ERD.png

DDL example: https://github.com/cs327e-spring2018/snippets/blob/master/create_tables.sql

Copy example: https://github.com/cs327e-spring2018/snippets/blob/master/import_data.sql

Additional Notes:

-Create a lab2 folder in your git repo and place all your work (code and diagram) in this folder.

-Submission is done through Canvas with a submission.json file.

-The submission.json file should be in this format:

```
{
    "commit_id": "[commit id]"
}
```

-There should be one submission only per team.

-Lateness penalty is %10 reduction per late day.