Thank you for your interest CleanChoice Energy's Data Science internship program.The following test is a required part of the application process. The test is designed to measure your ability to evaluate and manipulate data, think creatively and analytically, and communicate your research effectively.

Section 1
To complete this exam, you'll need to provide the SQL code necessary to create a couple of tables and to retrieve particular information from these tables.

The first table you need to create holds Mail Leads and should track the following information:

- First Name
- Last Name
- Mailing Address
- The customer's electricity utility
- The electricity utility id number
- State id
- County id
- Census Block id

Because leads can be mailed multiple times, we use a separate table to track each mailing. Mailings will be inserted into the table when created, so new mailings will continually be added to the table. Create a table to track the following information:

- A persistent table id
- The name of the mailing
- The date the mailing was sent
- The utility id that the mail was sent to
- The number of pieces of mail sent
- Census tracking code

Deliverable 1: SQL Code
Write SQL code that creates the two tables described above, and any additional tables that could increase the data integrity of these two tables.

Section 2
Based on the mailing leads tables created above, write queries that return the following information from your created tables.

1. Number of customer leads by state
2. Number of distinct addresses in the file
3. Number of customer leads in each utility, grouped by state
4. A concatenated census tracking code from the state, county, and census block ids.
5. The total number of mail pieces sent in each mailing

Deliverable 2: SQL Code

Section 3
The attached file contains some energy usage data for a set of customers. The data are a panel of invoices over a period of 2 years for a set of customers. Here is a description of the fields included:

Column name	**Description**
'Customer_id'	Unique customer identifier
'Invoicefromdt'	Start date of bill period
'Invoicetodt'	End date of bill period
'Invoicedate'	Date of invoice
'Kwh'	Energy usage in kWh
'Geoid1'	Geographical identifier
'Geoid2'	Geographical identifier
'Score'	Customer propensity score

To complete this portion of the exam, please use python to:

1. Summarize the data using descriptive statistics and/or plots.
    a. Do you observe any noteworthy patterns or insights?
2. Build a model to predict a customer's next invoice amount (kWh). We are interested in seeing the steps you would take to build a good predictive model. We are not necessarily concerned that you have actually built the best possible model. If you run out of time, please describe what steps you would take to improve on your model if you had more time.

Deliverable 3: Python Code
An annotated jupyter notebook that answers questions 1 and 2 above.