How To Run, assumptions and outputs

How to run

1. Please install the following python dependencies via pip:
   a. pandas==1.1.4
   b. tabulate==0.8.9
   c. dnspython==2.1.0
   d. pymongo==3.12.0
2. Please place the 'raw_data.csv' and 'run.py' in the same folder.
3. Run the python code using the following command: 'python run.py'
4. You can chk the databse directly by logging into mongodb cloud database.
   link - https://www.mongodb.com/cloud/atlas
   email – soumyadipauddy.incubyte@protonmail.com
   Password - Photo@2021

Assumptions

1. Since mongo is used here, we need to declare a primary key for all the tables. Here 'customer_id' is used as a primary key. The primary key in mongo is named as '_id'. So '_id' in the tables means 'customer_id'
2. Table schemas:

```
                              customer_id
{
    "_id": "C006",
    "country": "India",
    "customer_name": "Jennifer",
    "dob": 19980429,
    "dr_name": "Sandra",
    "is_active": "Y",
    "last_consulted_date": 20209202,
    "open_date": 20199202,
    "state": "Maharashtra",
    "vaccination_id": "V006"
}
```

Please login to Mongodb cloud databse to see the outputs.

link - https://www.mongodb.com/cloud/atlas

email – soumyadipauddy.incubyte@protonmail.com

Password - Photo@2021

In my raw data there are 5 countries: 'India', 'Canada', 'Uganda', 'Spain', 'Poland'. So 5 tables have been created.

1. India

2. Canada

incubyteDb.table_Canada

Documents   **Aggregations**   **Schema**

FILTER   { field: 'value' }

ADD DATA ▾   ⬆   VIEW   ☰   {}   ⊞

```
_id: "C002"
country: "Canada"
customer_name: "Christopher"
dob: 19990121
dr_name: "Jasmine"
is_active: "Y"
last_consulted_date: 20219520
open_date: 20209520
state: "Alberta"
vaccination_id: "V002"
```

```
_id: "C007"
country: "Canada"
customer_name: "Joshua"
dob: 19981221
dr_name: "Tracy"
is_active: "Y"
last_consulted_date: 20209424
open_date: 20199424
state: "Manitoba"
vaccination_id: "V007"
```

Similarly, rest of the countries have their separate table.