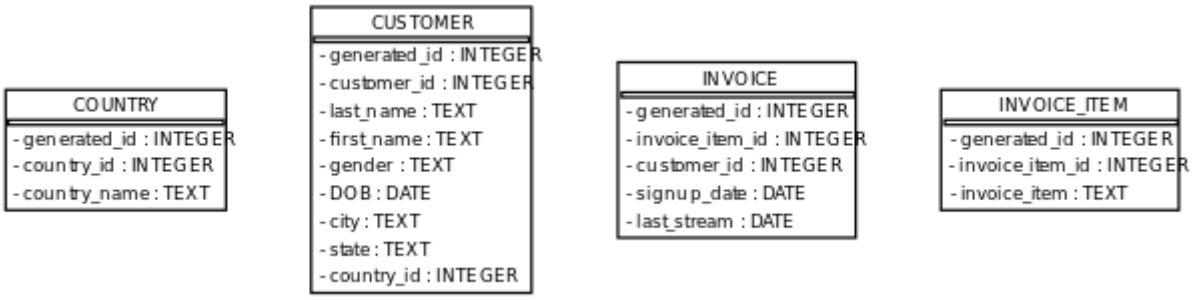
PART 1: Gathering the data

The following is an Entity Relationship Diagram (ERD) that details the tables and contents of the database.




```
## all the imports you will need for this case study
import os
import pandas as pd
import numpy as np
import sqlite3

```

Much of the data exist in a database. You can connect to is using the *sqlite3* Python package with the function shown below. Note that is is good practice to wrap your connect functions in a try-except statement to cleanly handle exceptions.

```
def connect_db(file_path):
    try:
        conn = sqlite3.connect(file_path)
        print("...successfully connected to db\n")
    except Error as e:
        print("...unsuccessful connection\n",e)

    return(conn)

```

```
## make the connection to the database
conn = connect_db('./data/aavail-customers.db')

## print the table names
tables = [t[0] for t in conn.execute("SELECT name FROM sqlite_master WHERE type
  ='table';")]
print(tables)

```

...successfully connected to db

['CUSTOMERS', 'INVOICES', 'INVOICE_ITEMS']

QUESTION 1:

Extract the relevant data from the DB

Query the database and extract the following data into a Pandas DataFrame.

- Customer ID (integer)
- Last name
- First name
- DOB
- City
- State
- Country (the name NOT the country_id)
- Gender

Remember that that SQL is case-insensitive, but it is traditional to use ALL CAPS for SQL keywords. It is also a convention to end SQL statements with a semi-colon.

Resources

- W3 schools SQL tutorial
- W3 schools SQL joins

An answer key has been provided in the form of an online Jupyter Notebook for your to review upon completion of this exercise.

QUESTION 2:

Extract the relevant data from the CSV file

For each *customer_id* determine if a customer has stopped their subscription or not and save it in a dictionary or another data container.

```
df_streams = pd.read_csv(r"./data/aavail-streams.csv")
df_streams.head()
```

	customer_id	stream_id	date	subscription_stopped
0	1	1356	2018-12-01	0
1	1	1540	2018-12-04	0
2	1	1395	2018-12-11	0
3	1	1255	2018-12-22	0
4	1	1697	2018-12-23	0

An answer key has been provided in the form of an online Jupyter Notebook for your to review upon completion of this exercise.