**BACKGROUND**

This Guided Project was initiated as part of Dataquest’s course entitled “SQL Fundamentals”. The data used was compiled from the [CIA World Factbook](https://www.cia.gov/library/publications/the-world-factbook/) (with figures updated only up to 2015). The data set is a collection of statistics describing the countries of the Earth and can be accessed from Gerald Bauer’s github page [factbook.db](https://github.com/factbook/factbook.sql/releases). While the project accessed data using SQL, Python’s sqlite3 module was used for querying the database.

**DESCRIPTION**

Initiating the project involves importing the necessary modules (sqlite3 and pandas) for querying and analysis. A connection object was then created to represent the database. The database contains two tables for querying:

1. sqlite\_sequence; and
2. facts

The country statistics is found in the table *facts*. pandas.read\_sql\_query() was used extensively to query the *facts* table. Data analysis were focused on the following:

1. Calculating some summary statistics such as the minimum and maximum population and minimum and maximum population growth;
2. Creating histograms for the columns population, population\_growth, birth\_rate, and death\_rate;
3. Calculating the population densities per country and creating a histogram for the data;
4. Finding the country with the highest population density;
5. Finding the countries with the highest ratio of water to land; and
6. Finding the countries that have more water than land.

**FILES**

The following files are included in this project:

|  |  |
| --- | --- |
| FILE NAME | DESCRIPTION |
| GP\_Factbook.ipynb | The Jupyter Notebook version of the project. |
| factbook | The database file. |
| GP\_Factbook.html | The project in html format for easy viewing |
| GP\_Factbook\_ReamMe.docx | Short documentation for the project. |