

Exploring the CIA World Factbook Database

Author : Kenneth Lucas Kusima

Contributor: DataQuest

Project Origin:

This project was made per instructions from:

The DataQuest Course: SQL Fundamentals

CIA World Factbook provides information on the history, people and society, government, economy, energy, geography, communications, transportation, military, and transnational issues for 267 world entities.

The Factbook contains demographic information like:

- population - The population as of 2015.
- population_growth - The annual population growth rate, as a percentage.
- area - The total land and water area.

Link to the database : <https://dsserver-prod-resources-1.s3.amazonaws.com/257/factbook.db>
(<https://dsserver-prod-resources-1.s3.amazonaws.com/257/factbook.db>)

Analysis will be done using the SQLite Management System

Connecting to the database:

The code below connects the jupyter file to the database

In [79]:

```
%%capture
%load_ext sql
%sql sqlite:///factbook.db
```

Out[79]:

```
'Connected: None@factbook.db'
```

Showing the tables in the database

In [80]:

```
%%sql
SELECT *
  FROM sqlite_master
 WHERE type='table';
```

Done.

Out[80]:

type	name	tbl_name	rootpage	sql
table	sqlite_sequence	sqlite_sequence	3	CREATE TABLE sqlite_sequence(name,seq)
table	facts	facts	47	CREATE TABLE "facts" ("id" INTEGER PRIMARY KEY AUTOINCREMENT NOT NULL, "code" varchar(255) NOT NULL, "name" varchar(255) NOT NULL, "area" integer, "area_land" integer, "area_water" integer, "population" integer, "population_growth" float, "birth_rate" float, "death_rate" float, "migration_rate" float)

Displaying the first 5 rows of the facts table

In [81]:

```
%%sql
SELECT *
  FROM facts
 LIMIT 5;
```

Done.

Out[81]:

id	code	name	area	area_land	area_water	population	population_growth	birth_rate
1	af	Afghanistan	652230	652230	0	32564342	2.32	38.57
2	al	Albania	28748	27398	1350	3029278	0.3	12.92
3	ag	Algeria	2381741	2381741	0	39542166	1.84	23.67
4	an	Andorra	468	468	0	85580	0.12	8.13
5	ao	Angola	1246700	1246700	0	19625353	2.78	38.78

Descriptions of some of the columns from the facts table above are as follows:

- name - The name of the country.
- area- The country's total area (both land and water).
- area_land - The country's land area in square kilometers.
- area_water - The country's waterarea in square kilometers.
- population - The country's population.
- population_growth- The country's population growth as a percentage.
- birth_rate - The country's birth rate, or the number of births a year per 1,000 people.
- death_rate - The country's death rate, or the number of death a year per 1,000 people.

Summary Statistics for the Facts table:

In [82]:

```
%%sql
SELECT MIN(population) AS Minimum_Population,
       MAX(population) AS Maximum_Population,
       MIN(population_growth) AS Minimum_Population_Growth,
       MAX(population_growth) AS Maximum_Population_Growth
FROM facts;
```

Done.

Out[82]:

Minimum_Population	Maximum_Population	Minimum_Population_Growth	Maximum_Population_Gr
0	7256490011	0.0	

The above data implies that there is a country that exists with a population of 0 and another with a population of 7,256,490,011 (i.e. 7.2 billion people)

Exploring Outliers

In [83]:

```
%%sql
SELECT *
FROM facts
WHERE population == (SELECT MIN(population) FROM facts);
```

Done.

Out[83]:

id	code	name	area	area_land	area_water	population	population_growth	birth_rate	d
250	ay	Antarctica	None	280000	None	0	None	None	

AS seen also in the CIA Factbook page, Antarctica is rightfully shown to have a population of 0 :

<https://www.cia.gov/library/publications/the-world-factbook/geos/ay.html>

(<https://www.cia.gov/library/publications/the-world-factbook/geos/ay.html>)

In [84]:

```
%%sql
SELECT *
FROM facts
WHERE population == (SELECT MAX(population) FROM facts);
```

Done.

Out[84]:

id	code	name	area	area_land	area_water	population	population_growth	birth_rate	deat
261	xx	World	None	None	None	7256490011	1.08	18.6	

Furthermore, the dataset appears to include the world as a country and therefore providing misleading statistics. This can be fixed by excluding the world from the result as follows:

Summary Statistics for the Facts table (without outlier):

In [85]:

```
%%sql
SELECT MIN(population) AS Minimum_Population,
       MAX(population) AS Maximum_Population,
       MIN(population_growth) AS Minimum_Population_Growth,
       MAX(population_growth) AS Maximum_Population_Growth
FROM facts
WHERE name IS NOT 'World'
```

Done.

Out[85]:

Minimum_Population	Maximum_Population	Minimum_Population_Growth	Maximum_Population_Gr
0	1367485388	0.0	

The above summary therefore shows that the maximum population a country has is 1,367,485,388 (i.e. 1.4 billion)

Calculating average properties

In [86]:

```
%%sql
SELECT AVG(population) AS Average_Population,
       AVG(area) AS Average_area
FROM facts
WHERE name IS NOT 'World'
```

Done.

Out[86]:

Average_Population	Average_area
32242666.56846473	555093.546184739

From the above result, the average population is 32,242,666.56846473 million people (i.e 32.2 million people) and the average country's total area is 555,093.546184739 (i.e 555,000)

Densely Populated Countries

These countries have:

- Above average values for population.
- Below average values for area.

In [87]:

```
%%sql
SELECT *
FROM facts
WHERE area < (SELECT AVG(area) FROM facts WHERE name IS NOT 'World')
      AND population > (SELECT AVG(population) FROM facts WHERE name IS NOT 'World') ;
```

Done.

Out[87]:

	id	code	name	area	area_land	area_water	population	population_growth	birth_rate
14	bg	Bangladesh	148460	130170	18290	168957745	1.6	21.14	
65	gm	Germany	357022	348672	8350	80854408	0.17	8.47	
80	iz	Iraq	438317	437367	950	37056169	2.93	31.45	
83	it	Italy	301340	294140	7200	61855120	0.27	8.74	
85	ja	Japan	377915	364485	13430	126919659	0.16	7.93	
91	ks	Korea, South	99720	96920	2800	49115196	0.14	8.19	
120	mo	Morocco	446550	446300	250	33322699	1.0	18.2	
138	rp	Philippines	300000	298170	1830	100998376	1.61	24.27	
139	pl	Poland	312685	304255	8430	38562189	0.09	9.74	
163	sp	Spain	505370	498980	6390	48146134	0.89	9.64	
173	th	Thailand	513120	510890	2230	67976405	0.34	11.19	
182	ug	Uganda	241038	197100	43938	37101745	3.24	43.79	
185	uk	United Kingdom	243610	241930	1680	64088222	0.54	12.17	
192	vm	Vietnam	331210	310070	21140	94348835	0.97	15.96	