Analyzing CIA Factbook Data using SQL ¶

In this project, I will be working with data from the CIA world Factbook, a compendium of statistics about all the countries of the world.

Connecting to Jupyter Notebook

Below codes will be use to connect the jupyter Notebook to the database file:

In [1]:
```
%%capture
%load_ext sql
%sql sqlite:///factbook.db
```

Out[1]:
```
'Connected: None@factbook.db'
```

Overview of the Table information

I will query the database to return information on the tables to give an idea of the data I am workig with.

In [2]:
```
%%sql
SELECT *
  FROM facts
 LIMIT 5;
```

```
Done.
```

Out[2]:

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

Checking for outliers

I will proceed to calculate some summary statistics and look for any outlier countries.

In [4]:
```
%%sql
SELECT MIN(population) 'Minimum Population',
       MAX(population) 'Maximum Population',
       MIN(population_growth) 'Minimum Population Growth',
       MAX(population_growth) 'Maximum Population Growth'
  FROM facts;
```

```
Done.
```

Out[4]:

Minimum Population	Maximum Population	Minimum Population Growth	Maximum Population Growth
0	7256490011	0.0	4.02

Above shows that there is a country with 0 population and a country with 7.2bn persons, These are clearly not true.

I will use subqueries to zoom in on these countries using the specific values:

In [6]:
```
%%sql
SELECT *
  FROM facts
 WHERE population = (SELECT MIN(population)
                       FROM facts
                    );
```

```
Done.
```

Out[6]:

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

In [7]:
```
%%sql
SELECT *
  FROM facts
 WHERE population = (SELECT MAX(population)
                       FROM facts
                    );
```

```
Done.
```

Out[7]:

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

The table contains Antarctica with zero population and world with 7.2bn population. I will exclude the 'world' rows from our data which is an outlier and recalculate the summary statistics as below:

In [12]:
```
%%sql
SELECT MIN(population) 'Minimum Population',
       MAX(population) 'Maximum Population',
       MIN(population_growth) 'Minimum Population Growth',
       MAX(population_growth) 'Maximum Population Growth'
  FROM facts
WHERE name <> 'World';
```

Done.

Out[12]:

Minimum Population	Maximum Population	Minimum Population Growth	Maximum Population Growth
0	1367485388	0.0	4.02

Calculating Average Population and Area

In [14]:
```
%%sql
SELECT AVG(population) 'Average Population',
        AVG(area) 'Average Area'
  FROM facts
 WHERE name <> 'World';
```

Done.

Out[14]:

Average Population	Average Area
32242666.56846473	555093.546184739

Average population is 32 million and average area stands at 555 thousands sq km.

Exploring Densely Populated Countries

These will be countries with above average population and below average values for Area:

In [19]:
```
%%sql
SELECT *
  FROM facts
 WHERE population > (SELECT AVG(population)
                       FROM facts
                      WHERE name <> 'World'
                    )
   AND area < (SELECT AVG(area)
                 FROM facts
                WHERE name <> 'World'
)
    ORDER BY "population" DESC
```

Done.

Out[19]:

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

Country with the highest population

In [20]:
```
%%sql
SELECT *
  FROM facts
 WHERE population = (SELECT MAX(population)
                       FROM facts
                      WHERE name <> 'World'
                    )
```

Done.

Out[20]:

id	code	name	area	area_land	area_water	population	population_growth	birth_rate	dea
37	ch	China	9596960	9326410	270550	1367485388	0.45	12.49	

China has the highest population with 1.36bn persons.

Country with Highest Growth Rate

In [24]:
```
%%sql
SELECT *
  FROM facts
 WHERE population_growth = (SELECT MAX(population_growth)
                       FROM facts
                      WHERE name <> 'World'
                           )
```

Done.

Out[24]:

id	code	name	area	area_land	area_water	population	population_growth	birth_rate	dea
162	od	South Sudan	644329	None	None	12042910	4.02	36.91	

South Sudan is country with the highest population growth rate of 4.02.

Countries with more Death Rate than Birth Rate

In [44]:
```
%%sql
SELECT name, population,
    CAST(death_rate AS FLOAT) / CAST(birth_rate AS FLOAT) Death_Birth_Rate
  FROM facts
    WHERE name <> 'World'
    AND CAST(death_rate AS FLOAT) / CAST(birth_rate AS FLOAT) > 1
    ORDER BY Death_Birth_Rate DESC
```

Done.

Out[44]:

name	population	Death_Birth_Rate
Bulgaria	7186893	1.6188340807174888
Serbia	7176794	1.5044052863436124
Latvia	1986705	1.431
Lithuania	2884433	1.4128712871287128
Hungary	9897541	1.3897379912663756
Monaco	30535	1.3894736842105262
Slovenia	1983412	1.350356294536817
Ukraine	44429471	1.3488805970149254
Germany	80854408	1.3482880755608027
Saint Pierre and Miquelon	5657	1.3099730458221026
Romania	21666350	1.3019693654266957
Croatia	4464844	1.288888888888889
Greece	10775643	1.2806004618937643
Belarus	9589689	1.2485981308411216
Japan	126919659	1.1992433795712485
Portugal	10825309	1.1887810140237325
Russia	142423773	1.1801724137931033
Estonia	1265420	1.1798287345385348
Italy	61855120	1.165903890160183
Bosnia and Herzegovina	3867055	1.0992108229988726
Czech Republic	10644842	1.0737279335410175
Moldova	3546847	1.0491666666666666
Poland	38562189	1.0462012320328542
Austria	8665550	1.0010626992561105

Bulgaria, Serbia and Latvia are the top three countries with highest death rate to birth rate.

Countries with more Water Area than Land Area

In [43]:
```
%%sql
SELECT name, population, area, area_land, area_water
  FROM facts
    WHERE name <> 'World'
    AND area_water > area_land
```

Done.

Out[43]:

name	population	area	area_land	area_water
British Indian Ocean Territory	None	54400	60	54340
Virgin Islands	103574	1910	346	1564

British Indian Ocean Territory and Virgin Islands are the only two countries with more water area than land area.

In [48]:
```
%%sql
SELECT *
  FROM facts
 WHERE population_growth > (SELECT AVG(population_growth)
                       FROM facts
                      WHERE name <> 'World'
                    )
ORDER BY population_growth DESC
LIMIT 10;
```

Done.

Out[48]:

id	code	name	area	area_land	area_water	population	population_growth	birth_rate	c
162	od	South Sudan	644329	None	None	12042910	4.02	36.91	
106	mi	Malawi	118484	94080	24404	17964697	3.32	41.56	
29	by	Burundi	27830	25680	2150	10742276	3.28	42.01	
128	ng	Niger	None	1266700	300	18045729	3.25	45.45	
182	ug	Uganda	241038	197100	43938	37101745	3.24	43.79	
141	qa	Qatar	11586	11586	0	2194817	3.07	9.84	
27	uv	Burkina Faso	274200	273800	400	18931686	3.03	42.03	
109	ml	Mali	1240192	1220190	20002	16955536	2.98	44.99	
219	cw	Cook Islands	236	236	0	9838	2.95	14.33	
80	iz	Iraq	438317	437367	950	37056169	2.93	31.45	

Above are the top 10 countries that will add more people to their population next year.

Conclusion

In conclusion, I was able to analyze the CIA Factbook data using SQL to explore countries with most densed population, more water area than land, higher death rate than birth rate, most populated and the countries that will add more people to its population in the next year.