sql-query-deforest

September 10, 2020

SETUP

[2]:
```
import pandas as pd
```

[3]:
```
from sqlalchemy import create_engine
engine = create_engine('postgresql://murongcui:woaini520@localhost:5432/
↪deforest')
```

[41]:
```
forest_area = pd.read_csv('data/forest_area.csv')
forest_area.columns = [c.lower() for c in forest_area.columns]

land_area = pd.read_csv('data/land_area.csv')
land_area.columns = [c.lower() for c in land_area.columns]

region = pd.read_csv('data/region.csv')
region.columns = [c.lower() for c in region.columns]
```

[42]:
```
forest_area.head()
```

[42]:

	country_code	country_name	year	forest_area_sqkm
0	ABW	Aruba	2016	4.20000
1	AFG	Afghanistan	2016	13500.00000
2	AGO	Angola	2016	577311.99220
3	ALB	Albania	2016	7705.39978
4	AND	Andorra	2016	160.00000

[43]:
```
land_area.head()
```

[43]:

	country_code	country_name	year	total_area_sq_mi
0	ABW	Aruba	2016	69.50
1	AFG	Afghanistan	2016	252069.50
2	AGO	Angola	2016	481351.35
3	ALB	Albania	2016	10579.15
4	AND	Andorra	2016	181.47

[44]:
```
region.head()
```

[44]:
```
      country_name country_code                       region  \
0      Afghanistan          AFG                   South Asia
1          Albania          ALB       Europe & Central Asia
2          Algeria          DZA  Middle East & North Africa
3   American Samoa          ASM         East Asia & Pacific
4          Andorra          AND       Europe & Central Asia

           income_group
0            Low income
1   Upper middle income
2   Upper middle income
3   Upper middle income
4           High income
```

[52]:
```
forest_area.to_sql("forest_area", engine, index=None)
```

[53]:
```
land_area.to_sql('land_area', engine, index = None)
```

[54]:
```
region.to_sql('region', engine, index=None)
```

[4]:
```
%load_ext sql
```

[5]:
```
%sql postgresql://postgres:woaini520@localhost/deforest
```

Steps to Complete

1. Create a View called "forestation" by joining all three tables - forest_area, land_area and regions in the workspace.
2. The forest_area and land_area tables join on both country_code AND year.
3. The regions table joins these based on only country_code.
4. In the 'forestation' View, include the following:

- All of the columns of the origin tables
- A new column that provides the percent of the land area that is designated as forest.

5. Keep in mind that the column forest_area_sqkm in the forest_area table and the land_area_sqmi in the land_area table are in different units (square kilometers and square miles, respectively), so an adjustment will need to be made in the calculation you write (1 sq mi = 2.59 sq km).

[14]:
```
%%sql
CREATE TABLE forestation AS (
SELECT f.country_code, f.country_name, f.year, f.forest_area_sqkm, l.
 ↪total_area_sq_mi*2.59 total_area_sqkm, r.region, r.income_group,␣
 ↪forest_area_sqkm /(l.total_area_sq_mi*2.59) forest_perc
FROM forest_area f
JOIN land_area l
ON f.country_code = l.country_code
AND f.year = l.year
```

```
JOIN region r
ON l.country_code = r.country_code);
```

```
 * postgresql://postgres:***@localhost/deforest
5886 rows affected.
```

[14]:
```
[]
```

[16]:
```
%%sql
SELECT * FROM forestation
WHERE country_name = 'World'
LIMIT 5;
```

```
 * postgresql://postgres:***@localhost/deforest
5 rows affected.
```

[16]:
```
[('WLD', 'World', 2016, 39958245.9, 127354641.43569998, 'World', None,
0.3137557096430953),
 ('WLD', 'World', 2015, 39991324.6, 127344421.29569998, 'World', None,
0.31404064813438654),
 ('WLD', 'World', 2014, 40024403.3, 127344955.5868, 'World', None,
0.3142990871964209),
 ('WLD', 'World', 2013, 40057482.0, 127346324.19459999, 'World', None,
0.31455546324829536),
 ('WLD', 'World', 2012, 40090560.5, 127347417.2523, 'World', None,
0.3148125133984681)]
```

- What was the total forest area (in sq km) of the world in 1990? Please keep in mind that you can use the country record denoted as "World" in the region table.
- What was the total forest area (in sq km) of the world in 2016? Please keep in mind that you can use the country record in the table is denoted as "World."
- What was the change (in sq km) in the forest area of the world from 1990 to 2016?
- What was the percent change in forest area of the world between 1990 and 2016?
- If you compare the amount of forest area lost between 1990 and 2016, to which country's total area in 2016 is it closest to?

[17]:
```
%%sql
SELECT forest_area_sqkm FROM forestation
WHERE country_name = 'World' AND year = 1990;
```

```
 * postgresql://postgres:***@localhost/deforest
1 rows affected.
```

[17]:
```
[(41282694.9,)]
```

[26]:
```
%%sql
SELECT year, forest_area_sqkm FROM forestation
WHERE country_name = 'World' AND year = 1990;
```

```
 * postgresql://postgres:***@localhost/deforest
1 rows affected.
```

[26]:
```
[(1990, 41282694.9)]
```

[25]:
```
%%sql
SELECT year, forest_area_sqkm FROM forestation
WHERE country_name = 'World' AND year = 2016;
```

```
 * postgresql://postgres:***@localhost/deforest
1 rows affected.
```

[25]:
```
[(2016, 39958245.9)]
```

[47]:
```
%%sql
WITH t1 AS
(SELECT a.forest_area_sqkm forest_1990, b.forest_area_sqkm forest_2016
FROM forestation a,
 forestation b
WHERE a.year = 1990
 AND b.year = 2016
 AND a.country_name = 'World'
 AND b.country_name = 'World')

SELECT forest_2016 - forest_1990 forest_loss
FROM t1;
```

```
 * postgresql://postgres:***@localhost/deforest
1 rows affected.
```

[47]:
```
[(-1324449.0,)]
```

[48]:
```
%%sql
WITH t1 AS
(SELECT a.forest_area_sqkm forest_1990, b.forest_area_sqkm forest_2016
FROM forestation a,
 forestation b
WHERE a.year = 1990
 AND b.year = 2016
 AND a.country_name = 'World'
 AND b.country_name = 'World')

SELECT (forest_2016 - forest_1990)/forest_1990 forest_loss_perc
FROM t1;
```

```
 * postgresql://postgres:***@localhost/deforest
1 rows affected.
```

[48]:
```
[(-0.0320824258980244,)]
```

[82]:
```
%%sql

SELECT country_name, total_area_sqkm FROM forestation
WHERE year = 2016 AND total_area_sqkm <= (

WITH t1 AS
(SELECT a.forest_area_sqkm forest_1990, b.forest_area_sqkm forest_2016
FROM forestation a,
 forestation b
WHERE a.year = 1990
 AND b.year = 2016
 AND a.country_name = 'World'
 AND b.country_name = 'World')
SELECT ABS(forest_2016 - forest_1990) forest_loss_perc
FROM t1)
ORDER BY 2 DESC
LIMIT 1 ;
```

```
 * postgresql://postgres:***@localhost/deforest
1 rows affected.
```

[82]:
```
[('Peru', 1279999.9891)]
```

- What was the percent forest of the entire world in 2016? Which region had the HIGHEST percent forest in 2016, and which had the LOWEST, to 2 decimal places?

[84]:
```
%%sql
SELECT * FROM forestation
WHERE country_name = 'World' AND year = 2016;
```

```
 * postgresql://postgres:***@localhost/deforest
1 rows affected.
```

[84]:
```
[('WLD', 'World', 2016, 39958245.9, 127354641.43569998, 'World', None,
0.3137557096430953)]
```

[100]:
```
%%sql
SELECT region, SUM(forest_area_sqkm) forest_sum, SUM(total_area_sqkm)␣
 ↪total_sum, ROUND(CAST(SUM(forest_area_sqkm)/SUM(total_area_sqkm)*100 AS␣
 ↪NUMERIC), 2) forest_perc
FROM forestation
WHERE year = 2016
GROUP BY 1
```

```
ORDER BY 4 DESC
;
```

```
 * postgresql://postgres:***@localhost/deforest
8 rows affected.
```

[100]:
```
[('Latin America & Caribbean', 9250585.884135248, 20039364.446500003,
Decimal('46.16')),
 ('Europe & Central Asia', 10438609.30732392, 27440113.6114, Decimal('38.04')),
 ('North America', 6573934.063, 18240983.9864, Decimal('36.04')),
 ('World', 39958245.9, 127354641.43569998, Decimal('31.38')),
 ('Sub-Saharan Africa', 6115290.9152861, 21242361.0679, Decimal('28.79')),
 ('East Asia & Pacific', 6421326.3921158, 24361338.4462, Decimal('26.36')),
 ('South Asia', 835310.4846399999, 4771604.0344, Decimal('17.51')),
 ('Middle East & North Africa', 232131.004009593, 11223465.984499997,
Decimal('2.07'))]
```

- What was the percent forest of the entire world in 1990? Which region had the HIGHEST percent forest in 1990, and which had the LOWEST, to 2 decimal places?

[101]:
```
%%sql
SELECT * FROM forestation
WHERE country_name = 'World' AND year = 1990;
```

```
 * postgresql://postgres:***@localhost/deforest
1 rows affected.
```

[101]:
```
[('WLD', 'World', 1990, 41282694.9, 127328467.43959999, 'World', None,
0.3242220355756894)]
```

[102]:
```
%%sql
SELECT region, SUM(forest_area_sqkm) forest_sum, SUM(total_area_sqkm)␣
↪total_sum, ROUND(CAST(SUM(forest_area_sqkm)/SUM(total_area_sqkm)*100 AS␣
↪NUMERIC), 2) forest_perc
FROM forestation
WHERE year = 1990
GROUP BY 1
ORDER BY 4 DESC
```

```
 * postgresql://postgres:***@localhost/deforest
8 rows affected.
```

[102]:
```
[('Latin America & Caribbean', 10242341.796304759, 20071224.450900003,
Decimal('51.03')),
 ('Europe & Central Asia', 10199847.602310268, 27357215.0411, Decimal('37.28')),
 ('North America', 6507240.0, 18252523.9904, Decimal('35.65')),
 ('World', 41282694.9, 127328467.43959999, Decimal('32.42')),
 ('Sub-Saharan Africa', 6515615.1999664, 21241391.086999997, Decimal('30.67')),
```

```
 ('East Asia & Pacific', 6280252.8421379, 24364639.97100001, Decimal('25.78')),
 ('South Asia', 789187.09961, 4779833.0601, Decimal('16.51')),
 ('Middle East & North Africa', 199292.595698698, 11226230.006599998,
Decimal('1.78'))]
```

- Based on the table you created, which regions of the world DECREASED in forest area from 1990 to 2016?

[108]:
```
%%sql
SELECT a.region, a.forest_perc perc_90, b.forest_perc perc_16, b.forest_perc-a.
 ↪forest_perc perc_change

FROM (SELECT region, SUM(forest_area_sqkm) forest_sum, SUM(total_area_sqkm)␣
 ↪total_sum, ROUND(CAST(SUM(forest_area_sqkm)/SUM(total_area_sqkm)*100 AS␣
 ↪NUMERIC), 2) forest_perc
FROM forestation
WHERE year = 1990
GROUP BY 1
ORDER BY 4 DESC) a

JOIN (SELECT region, SUM(forest_area_sqkm) forest_sum, SUM(total_area_sqkm)␣
 ↪total_sum, ROUND(CAST(SUM(forest_area_sqkm)/SUM(total_area_sqkm)*100 AS␣
 ↪NUMERIC), 2) forest_perc
FROM forestation
WHERE year = 2016
GROUP BY 1
ORDER BY 4 DESC) b

ON a.region = b.region
ORDER BY 4
;
```

```
 * postgresql://postgres:***@localhost/deforest
8 rows affected.
```

[108]:
```
[('Latin America & Caribbean', Decimal('51.03'), Decimal('46.16'),
Decimal('-4.87')),
 ('Sub-Saharan Africa', Decimal('30.67'), Decimal('28.79'), Decimal('-1.88')),
 ('World', Decimal('32.42'), Decimal('31.38'), Decimal('-1.04')),
 ('Middle East & North Africa', Decimal('1.78'), Decimal('2.07'),
Decimal('0.29')),
 ('North America', Decimal('35.65'), Decimal('36.04'), Decimal('0.39')),
 ('East Asia & Pacific', Decimal('25.78'), Decimal('26.36'), Decimal('0.58')),
 ('Europe & Central Asia', Decimal('37.28'), Decimal('38.04'), Decimal('0.76')),
 ('South Asia', Decimal('16.51'), Decimal('17.51'), Decimal('1.00'))]
```

- Which 5 countries saw the largest amount decrease in forest area from 1990 to 2016? What was the difference in forest area for each?

[136]:
```
%%sql
SELECT a.country_name, a.region, a.forest_area_sqkm-b.forest_area_sqkm␣
 ↪forest_loss

FROM (SELECT country_name, region, forest_area_sqkm
FROM forestation
WHERE year = 1990) a
JOIN (SELECT country_name, region, forest_area_sqkm
FROM forestation
WHERE year = 2016) b
ON a.country_name = b.country_name
ORDER BY 3 DESC
LIMIT 20;
```

```
 * postgresql://postgres:***@localhost/deforest
20 rows affected.
```

[136]:
```
[('Hong Kong SAR, China', 'East Asia & Pacific', None),
 ('Ethiopia', 'Sub-Saharan Africa', None),
 ('Gibraltar', 'Europe & Central Asia', None),
 ('Kosovo', 'Europe & Central Asia', None),
 ('Sint Maarten (Dutch part)', 'Latin America & Caribbean', None),
 ('Macao SAR, China', 'East Asia & Pacific', None),
 ('Qatar', 'Middle East & North Africa', None),
 ('South Sudan', 'Sub-Saharan Africa', None),
 ('Sudan', 'Sub-Saharan Africa', None),
 ('St. Martin (French part)', 'Latin America & Caribbean', None),
 ('Monaco', 'Europe & Central Asia', None),
 ('Nauru', 'East Asia & Pacific', None),
 ('San Marino', 'Europe & Central Asia', None),
 ('Curacao', 'Latin America & Caribbean', None),
 ('World', 'World', 1324449.0),
 ('Brazil', 'Latin America & Caribbean', 541510.0),
 ('Indonesia', 'East Asia & Pacific', 282193.98439999996),
 ('Myanmar', 'East Asia & Pacific', 107234.00390000001),
 ('Nigeria', 'Sub-Saharan Africa', 106506.00098),
 ('Tanzania', 'Sub-Saharan Africa', 102320.0)]
```

- Which 5 countries saw the largest percent decrease in forest area from 1990 to 2016? What was the percent change to 2 decimal places for each?

[138]:
```
%%sql
SELECT a.country_name, a.region, (a.forest_area_sqkm-b.forest_area_sqkm)/a.
 ↪forest_area_sqkm forest_perc_loss

FROM (SELECT country_name, region, forest_area_sqkm
FROM forestation
```

```
WHERE year = 1990) a
JOIN (SELECT country_name, region, forest_area_sqkm
FROM forestation
WHERE year = 2016) b
ON a.country_name = b.country_name
ORDER BY 3 DESC
LIMIT 20;
```

```
 * postgresql://postgres:***@localhost/deforest
20 rows affected.
```

[138]:
```
[('Monaco', 'Europe & Central Asia', None),
 ('Qatar', 'Middle East & North Africa', None),
 ('Sudan', 'Sub-Saharan Africa', None),
 ('Sint Maarten (Dutch part)', 'Latin America & Caribbean', None),
 ('Gibraltar', 'Europe & Central Asia', None),
 ('Kosovo', 'Europe & Central Asia', None),
 ('Curacao', 'Latin America & Caribbean', None),
 ('South Sudan', 'Sub-Saharan Africa', None),
 ('Macao SAR, China', 'East Asia & Pacific', None),
 ('Ethiopia', 'Sub-Saharan Africa', None),
 ('St. Martin (French part)', 'Latin America & Caribbean', None),
 ('Hong Kong SAR, China', 'East Asia & Pacific', None),
 ('San Marino', 'Europe & Central Asia', None),
 ('Nauru', 'East Asia & Pacific', None),
 ('Togo', 'Sub-Saharan Africa', 0.7544525592700729),
 ('Nigeria', 'Sub-Saharan Africa', 0.6179993093884182),
 ('Uganda', 'Sub-Saharan Africa', 0.5912860347295306),
 ('Mauritania', 'Sub-Saharan Africa', 0.4674698795180723),
 ('Honduras', 'Latin America & Caribbean', 0.4503441494591937),
 ('Pakistan', 'South Asia', 0.43450732093391375)]
```

- If countries were grouped by percent forestation in quartiles, which group had the most countries in it in 2016?

[144]:
```
%%sql
SELECT ntile, COUNT(*)
FROM
(SELECT country_name, forest_perc, NTILE(4) OVER (PARTITION BY forest_perc)
FROM forestation
WHERE year = 2016 AND forest_perc IS NOT NULL) sub
GROUP BY 1;
```

```
 * postgresql://postgres:***@localhost/deforest
4 rows affected.
```

[144]:
```
[(3, 3), (4, 3), (2, 3), (1, 209)]
```

- List all of the countries that were in the 4th quartile (percent forest > 75%) in 2016.

[163]:
```
%%sql
WITH t1 AS (SELECT sub.country_name, sub.forest_perc, NTILE(4) OVER (PARTITION␣
 ↪BY sub.forest_perc)
FROM (
SELECT country_name, forest_perc
FROM forestation
WHERE year = 2016 AND forest_perc IS NOT NULL) sub)
SELECT ntile, COUNT(*)
FROM t1
GROUP BY 1;
```

```
 * postgresql://postgres:***@localhost/deforest
1 rows affected.
```

[163]:
```
[(1, 205)]
```

- How many countries had a percent forestation higher than the United States in 2016?

[169]:
```
%%sql
SELECT ntile, COUNT(*)
FROM (SELECT country_name, forest_perc, NTILE(4) OVER (ORDER BY forest_perc)
FROM forestation
WHERE year = 2016 AND forest_perc IS NOT NULL) sub
GROUP BY 1;
```

```
 * postgresql://postgres:***@localhost/deforest
4 rows affected.
```

[169]:
```
[(3, 51), (4, 51), (2, 51), (1, 52)]
```

[131]:
```
%%sql
SELECT a.country_name, a.forest_area_sqkm forest_90, b.forest_area_sqkm␣
 ↪forest_16, a.forest_area_sqkm-b.forest_area_sqkm forest_loss

FROM (SELECT country_name, forest_area_sqkm
FROM forestation
WHERE year = 1990) a
JOIN (SELECT country_name, forest_area_sqkm
FROM forestation
WHERE year = 2016) b
ON a.country_name = b.country_name
ORDER BY 4
LIMIT 2;
```

```
 * postgresql://postgres:***@localhost/deforest
2 rows affected.
```

[131]:
```
[('China', 1571405.938, 2098635.0, -527229.0619999999),
 ('United States', 3024500.0, 3103700.0, -79200.0)]
```

[132]:
```
%%sql
SELECT a.country_name, a.forest_perc forest_90, b.forest_perc forest_16, a.
 ↪forest_perc-b.forest_perc forest_loss

FROM (SELECT country_name, forest_perc
FROM forestation
WHERE year = 1990) a
JOIN (SELECT country_name, forest_perc
FROM forestation
WHERE year = 2016) b
ON a.country_name = b.country_name
ORDER BY 4
LIMIT 1;
```

```
 * postgresql://postgres:***@localhost/deforest
1 rows affected.
```

[132]:
```
[('French Polynesia', 0.15027294895389126, 0.42349649250642085,
-0.2732235435525296)]
```

[175]:
```
%%sql
WITH t1 AS (SELECT country_name, forest_perc,
CASE WHEN forest_perc > 0.75 THEN 4
     WHEN forest_perc <= 0.75 AND forest_perc > 0.5 THEN 3
     WHEN forest_perc <= 0.5 AND forest_perc > 0.25 THEN 2
     WHEN forest_perc <= 0.25 THEN 1
END AS level

FROM forestation
WHERE year = 2016)

SELECT level, COUNT(*)
FROM t1
GROUP BY 1
```

```
 * postgresql://postgres:***@localhost/deforest
5 rows affected.
```

[175]:
```
[(None, 13), (3, 38), (4, 9), (2, 73), (1, 85)]
```

[178]:
```
%%sql
WITH t1 AS (SELECT country_name, region, forest_perc,
CASE WHEN forest_perc > 0.75 THEN 4
     WHEN forest_perc <= 0.75 AND forest_perc > 0.5 THEN 3
```

```
    WHEN forest_perc <= 0.5 AND forest_perc > 0.25 THEN 2
    WHEN forest_perc <= 0.25 THEN 1
END AS level

FROM forestation
WHERE year = 2016)

SELECT country_name, region, forest_perc
FROM t1
WHERE level = 4
```

```
 * postgresql://postgres:***@localhost/deforest
9 rows affected.
```

[178]:
```
[('American Samoa', 'East Asia & Pacific', 0.875000875000875),
 ('Micronesia, Fed. Sts.', 'East Asia & Pacific', 0.9185723907152479),
 ('Gabon', 'Sub-Saharan Africa', 0.9003764187005651),
 ('Guyana', 'Latin America & Caribbean', 0.8390144891106817),
 ('Lao PDR', 'East Asia & Pacific', 0.8210823176408608),
 ('Palau', 'East Asia & Pacific', 0.8760680854912034),
 ('Solomon Islands', 'East Asia & Pacific', 0.7786351779450664),
 ('Suriname', 'Latin America & Caribbean', 0.9825769396765779),
 ('Seychelles', 'Sub-Saharan Africa', 0.8841113673857888)]
```

[ ]:
```

```