Report for ForestQuery into Global Deforestation, 1990 to 2016

ForestQuery is on a mission to combat deforestation around the world and to raise awareness about this topic and its impact on the environment. The data analysis team at ForestQuery has obtained data from the World Bank that includes forest area and total land area by country and year from 1990 to 2016, as well as a table of countries and the regions to which they belong.

The data analysis team has used SQL to bring these tables together and to query them in an effort to find areas of concern as well as areas that present an opportunity to learn from successes.

1. **GLOBAL SITUATION**

According to the World Bank, the total forest area of the world was **41282694.9km^2** in 1990. As of 2016, the most recent year for which data was available, that number had fallen to **39958245.9km^2**, a loss of 1324449km^2, or **1.63%**

The forest area lost over this time period is slightly more than the entire land area of **Peru (PER)** listed for the year 2016 (which is **1279999.99**).

2. **REGIONAL OUTLOOK**

In 2016, the percent of the total land area of the world designated as forest was **31.38%**. The region with the highest relative forestation was **Latin America & Caribbean** with **46.16%,** and the region with the lowest relative forestation was **Middle East & North Africa,** with **2.07%** forestation.

In 1990, the percent of the total land area of the world designated as forest was **32.42%**,. The region with the highest relative forestation was **Latin America & Caribbean**, with **51.03%**, and the region with the lowest relative forestation was **Middle East & North Africa**, with **1.78%** forestation.

Table 2.1: Percent Forest Area by Region, 1990 & 2016:

Region	1990 Forest Percentage	2016 Forest Percentage
Middle East & North Africa	2.69	3.19
South Asia	20.68	21.60
Europe & Central Asia	26.33	28.31
North America	29.95	30.20
World	32.42	31.38
Sub-Saharan Africa	35.26	31.28
Latin America & Caribbean	43.35	41.64
East Asia & Pacific	47.38	50.09

The only regions of the world that decreased in percent forest area from 1990 to 2016 were **Sub-Saharan Africa** (dropped from **35.26%** to **31.28%**) and **Latin America & Caribbean** (**43.35% to 41.64%).** All other regions actually increased in forest area over this time period. However, the drop in forest area in the two aforementioned regions was so large, the percent forest area of the world decreased over this time period from **32.42%** to **31.38%.**

3. **COUNTRY-LEVEL DETAIL**

A. SUCCESS STORIES

There is one particularly bright spot in the data at the country level, **China**. This country actually increased in forest area from 1990 to 2016 by **527229.06**. It would be interesting to study what has changed in this country over this time to drive this figure in the data higher. The country with the next largest increase in forest area from 1990 to 2016 was the **United States**, but it only saw an increase of **79200.00**, much lower than the figure for **China**.

China and **United States** are of course very large countries in total land area, so when we look at the largest *percent* change in forest area from 1990 to 2016, we aren't surprised to find a much smaller country listed at the top. **Iceland** increased in forest area by **51.65%** from 1990 to 2016.

B. LARGEST CONCERNS

Which countries are seeing deforestation to the largest degree? We can answer this question in two ways. First, we can look at the absolute square kilometer decrease in forest area from 1990 to 2016. The following 3 countries had the largest decrease in forest area over the time period under consideration:

Table 3.1: Top 5 Amount Decrease in Forest Area by Country, 1990 & 2016:

Country	Region	Absolute Forest Area Change
Brazil	Latin America & Caribbean	541510
Indonesia	East Asia & Pacific	282193.9844
Myanmar	East Asia & Pacific	107234.0039
Nigeria	Sub-Saharan Africa	106506.00098
Tanzania	Sub-Saharan Africa	102320

The second way to consider which countries are of concern is to analyze the data by percent decrease.

Table 3.2: Top 5 Percent Decrease in Forest Area by Country, 1990 & 2016:

Country	Region	Pct Forest Area Change
Togo	Sub-Saharan Africa	-60.57
Nigeria	Sub-Saharan Africa	-44.72
Uganda	Sub-Saharan Africa	-41.97
Mauritania	Sub-Saharan Africa	-30.50
Honduras	Latin America & Caribbean	-29.06

When we consider countries that decreased in forest area the most between 1990 and 2016, we find that four of the top 5 countries on the list are in the region of **Sub-Saharan Africa** The countries are **Togo, Nigeria, Uganda, and Mauritania.** The 5th country on the list is **Honduras**, which is in the **Latin America & Caribbean** region.

From the above analysis, we see that Nigeria is the only country that ranks in the top 5 both in terms of absolute square kilometer decrease in forest as well as percent decrease in forest area from 1990 to 2016. Therefore, this country has a significant opportunity ahead to stop the decline and hopefully spearhead remedial efforts.

C. QUARTILES

Table 3.3: Count of Countries Grouped by Forestation Percent Quartiles, 2016:

Quartile	Number of Countries
0%-25%	85
25%-50%	72
50%-75%	38
75%-100%	9

The largest number of countries in 2016 were found in the **0%-25%** quartile.

There were **9** countries in the top quartile in 2016. These are countries with a very high percentage of their land area designated as forest. The following is a list of countries and their respective forest land, denoted as a percentage.

Table 3.4: Top Quartile Countries, 2016:

Country	Region	Pct Designated as Forest
Suriname	Latin America & Caribbean	98.58
Micronesia, Fed. Sts.	East Asia & Pacific	91.36
American Samoa	East Asia & Pacific	89.74
Seychelles	Sub-Saharan Africa	88.41
Palau	East Asia & Pacific	86.35
Gabon	Sub-Saharan Africa	85.98
Guyana	Latin America & Caribbean	84.34
Solomon Islands	East Asia & Pacific	80.45
Lao PDR	East Asia & Pacific	75.37

5. RECOMMENDATIONS

Write out a set of recommendations as an analyst on the ForestQuery team.
- *What have you learned from the World Bank data?*

I have learned that, datasets will often come with blank rows and how you deal with them impacts your result.
- *Which countries should we focus on over others?*

Countries with high decrease in forested land of the time period 1990 – 2016 should be consulted but the focus should be on countries with a decrease over this time period. There is a need to utilize the approaches used by countries with an increase in their forested land over this same time period to implement these practices to the decreased countries. This will be a good way to fight global warming and consequently climate change

Appendix
SQL Queries used to Answer questions
Part 1 1. GLOBAL SITUATION

CREATE VIEW forestation AS
(SELECT f.country_code f_country_code, f.country_name f_country_name, f.year f_year,
f.forest_area_sqkm forest_area_sqkm, l.country_code   l_country_code, l.country_name
l_country_name, l.year l_year,   l.total_area_sq_mi total_area_sqmi, r.country_name
r_country_name, r.country_code r_country_code, r.region region, r.income_group income_grp,
ROUND((f.forest_area_sqkm / (l.total_area_sq_mi*2.59) * 100)::NUMERIC , 2) AS
percentage_forest_land
 FROM forest_area f
 JOIN land_area l
 ON f.country_code = l.country_code AND f.year = l.year
 JOIN regions r
 ON l.country_code = r.country_code)

a. 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.
SELECT *
FROM forestation
WHERE region = 'World' AND l_year = 1990

b. 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."
SELECT *
FROM forestation
WHERE region = 'World' AND l_year = 2016

c**. What was the change (in sq km) in the forest area of the world from 1990 to 2016?**
SELECT DISTINCT(
    (SELECT forest_area_sqkm FROM forestation
     WHERE region = 'World' AND l_year = 1990) -
    (SELECT forest_area_sqkm FROM forestation
     WHERE region = 'World' AND l_year = 2016)
) AS Difference
FROM
  Forestation

d. What was the percent change in forest area of the world between 1990 and 2016?

```
WITH t1 AS (
    SELECT DISTINCT  f_country_code, f_country_name,
            ((SELECT forest_area_sqkm
              FROM forestation
              WHERE region = 'World' AND l_year = 1990) -
              (SELECT forest_area_sqkm
               FROM forestation
               WHERE region = 'World' AND l_year = 2016))    difference
    FROM
        Forestation
    WHERE region = 'World'
),
t2 AS (
     SELECT DISTINCT f_country_code, f_country_name,
            ((SELECT forest_area_sqkm
              FROM forestation
              WHERE region = 'World' AND l_year = 1990)+
              (SELECT forest_area_sqkm
               FROM forestation
               WHERE region = 'World' AND l_year = 2016))  AS  total_forest_area_1990_2016
    FROM
        Forestation
    WHERE region = 'World'
)
SELECT t1.f_country_code, t1.f_country_name,
ROUND(((difference/total_forest_area_1990_2016)*100 )::Numeric, 2) AS percent
FROM t1
JOIN  t2
ON  t1.f_country_code = t2.f_country_code
```

e. 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?
```
WITH t1 AS (
  SELECT l_country_code, l_country_name, ROUND(((total_area_sqmi)* 2.59)::Numeric,2) AS
total_area_in_sqkm
  FROM forestation)
SELECT l_country_code,l_country_name, total_area_in_sqkm
FROM t1
WHERE total_area_in_sqkm <= 1324449
ORDER BY 3 DESC
LIMIT 1
```
PART TWO: REGIONAL OUTLOOK QUERIES
**This first query bellow, contains all inform but for exact extraction, the queries Q2a.1,
Q2a.2, Q2b.1, Q2b.1 and Q2c are applicable**

```
WITH CT AS (
SELECT f_country_code, region,
ROUND(((SUM(forest_area_sqkm)/SUM(total_area_sqmi*2.59))*100)::Numeric, 2)  AS
current_percent_forest
FROM forestation
WHERE l_year = 2016
GROUP BY 1,2
),
ET AS (
 SELECT f_country_code, region,
ROUND(((SUM(forest_area_sqkm)/SUM(total_area_sqmi*2.59))*100)::Numeric, 2)  AS
early_percent_forest
FROM forestation
WHERE l_year = 1990
GROUP BY 1,2
)
SELECT CT.region, ROUND((AVG(CT.current_percent_forest))::Numeric, 2)
current_percent_forest , ROUND((AVG(ET.early_percent_forest))::Numeric, 2)
early_percent_forest
FROM CT
JOIN ET
ON CT.f_country_code = ET.f_country_code
GROUP BY 1
```

**a. 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?**
```
Q2a.1.
SELECT region,
ROUND(((SUM(forest_area_sqkm)/SUM(total_area_sqmi*2.59))*100)::Numeric, 2)  AS
percent_forest
FROM forestation
WHERE l_year = 2016 AND region = 'World'
GROUP BY 1
ORDER BY 2  DESC
Q2a.2.
SELECT region,
ROUND(((SUM(forest_area_sqkm)/SUM(total_area_sqmi*2.59))*100)::Numeric, 2)  AS
percent_forest
FROM forestation
WHERE l_year = 2016
GROUP BY 1
ORDER BY 2  DESC
```

b. 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?

Q2b.1.

```
SELECT region,
ROUND(((SUM(forest_area_sqkm)/SUM(total_area_sqmi*2.59))*100)::Numeric, 2)  AS
percent_forest
FROM forestation
WHERE l_year = 1990 AND region = 'World'
GROUP BY 1
ORDER BY 2  DESC
```

Q2b.2

```
SELECT region,
ROUND(((SUM(forest_area_sqkm)/SUM(total_area_sqmi*2.59))*100)::Numeric, 2)  AS
percent_forest
FROM forestation
WHERE l_year = 1990
GROUP BY 1
ORDER BY 2  DESC
```

Q2c.

```
WITH CT AS (
SELECT f_country_code, region,
ROUND(((SUM(forest_area_sqkm)/SUM(total_area_sqmi*2.59))*100)::Numeric, 2)  AS
current_percent_forest
FROM forestation
WHERE l_year = 2016 AND f_country_name != 'World'
GROUP BY 1,2
),
ET AS (
 SELECT f_country_code, region,
ROUND(((SUM(forest_area_sqkm)/SUM(total_area_sqmi*2.59))*100)::Numeric, 2)  AS
early_percent_forest
FROM forestation
WHERE l_year = 1990 AND f_country_name != 'World'
GROUP BY 1,2
)
SELECT CT.region, ROUND((AVG(ET.early_percent_forest))::Numeric, 2)
early_percent_forest, ROUND((AVG(CT.current_percent_forest))::Numeric, 2)
current_percent_forest
FROM CT
JOIN ET
ON CT.f_country_code = ET.f_country_code
GROUP BY 1
```

ORDER BY 2, 3

Part 3. COUNTRY-LEVEL DETAIL
Q3a. Which 5 countries saw the largest amount decrease in forest area from 1990 to 2016? What was the difference in forest area for each?
Q3a.1

```
WITH t1 AS (
  SELECT  f_country_code, f_country_name, region,   forest_area_sqkm,
percentage_forest_land
  FROM forestation
 WHERE f_year = 1990 AND forest_area_sqkm IS NOT NULL AND total_area_sqmi IS NOT
NULL
  ),
t2 AS (
  SELECT  f_country_code, f_country_name, region, forest_area_sqkm,
percentage_forest_land
  FROM forestation
 WHERE f_year = 2016 AND forest_area_sqkm IS NOT NULL AND total_area_sqmi IS NOT
NULL
)
SELECT t1.f_country_name, t1.region,  ROUND((SUM(t2.forest_area_sqkm) -
SUM(t1.forest_area_sqkm))::Numeric, 2) Absolute_forest_area_change
FROM t1
JOIN t2
ON t1.f_country_code = t2.f_country_code
where t1.f_country_name !='World'
GROUP BY 1,2
ORDER BY 3
limit 5
```

Q3b. 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?

```
WITH t1 AS
 (SELECT f_country_code, f_country_name, region, forest_area_sqkm
 FROM forestation
 WHERE l_year = 1990 AND forest_area_sqkm IS NOT NULL),
t2 AS
```

```
(SELECT f_country_code, f_country_name, region, forest_area_sqkm
 FROM forestation
 WHERE l_year = 2016 AND forest_area_sqkm IS NOT NULL)
SELECT t1.f_country_name, t1.region, ROUND(((t2.forest_area_sqkm -
t1.forest_area_sqkm)/(t2.forest_area_sqkm + t1.forest_area_sqkm)*100)::Numeric, 2)
percentage_decrease_forest_area
FROM t1
JOIN t2
ON t1.f_country_code = t2.f_country_code
ORDER BY 3
```

Q3c. If countries were grouped by percent forestation in quartiles, which group had the most countries in it in 2016?.
```
WITH T1 AS
    (SELECT f_country_name,
     f_year,
    (SUM(forest_area_sqkm) / SUM(total_area_sqmi*2.59))*100   percent_forestation
    FROM forestation
    WHERE f_year = 2016 AND forest_area_sqkm IS NOT NULL AND  total_area_sqmi IS NOT
NULL
    GROUP BY 1, 2)
SELECT Distinct(quartiles),
 count(f_country_name)Over(PARTITION BY quartiles)
FROM
 (SELECT f_country_name,
 CASE
 WHEN percent_forestation<25 THEN '0%-25%'
 WHEN percent_forestation>=25
 AND percent_forestation<50 THEN '25%-50%'
 WHEN percent_forestation>=50
 AND percent_forestation<75 THEN '50%-75%'
 ELSE '75%-100%'
 END AS quartiles
 FROM T1
 WHERE percent_forestation IS NOT NULL
 AND f_year = 2016) sub
```

Q3d. List all of the countries that were in the 4th quartile (percent forest > 75%) in 2016.
```
SELECT f_country_name, region, ROUND(((SUM(forest_area_sqkm) /
SUM(total_area_sqmi*2.59))*100)::Numeric, 2)   percent_forestation
FROM forestation
```

where f_year = 2016 AND forest_area_sqkm IS NOT NULL AND total_area_sqmi IS NOT NULL
GROUP BY 1, 2
ORDER BY 3 DESC
LIMIT 9

Q3e. How many countries had a percent forestation higher than the United States in 2016?

SELECT COUNT(f_country_name)
FROM forestation
where f_year = 2016 and percentage_forest_land > (SELECT percentage_forest_land
                                                    FROM forestation
                                                    WHERE f_year = 2016 and
                                                    f_country_name =
                                                    'United States')