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 41,282,694.9 sqkm in 1990. As of 2016, the most recent year for which data was available, that number had fallen to 39,958,245.9 sqkm, a loss of 1,324,449 sqkm, or 3.21%.

The forest area lost over this time period is slightly more than the entire land area of Peru listed for the year 2016 (which is 1,279,999.9891 sqkm).

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:

0	Region	1990 Forest Percentage	2016 Forest Percentage
1	Middle East & North Africa	1.78	2.07
2	South Asia	16.51	17.51

3	East Asia & Pacific	25.78	26.36
4	Sub-Saharan Africa	30.67	28.79
5	World	32.42	31.38
6	North America	35.65	36.04
7	Europe & Central Asia	37.28	38.04
8	Latin America & Caribbean	51.03	46.16

The only regions of the world that decreased in percent forest area from 1990 to 2016 were Latin America & Caribbean (dropped from 51.03 % to 46.16%) and Sub-Saharan Africa (30.67 % to 28.79 %). 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 527,229.062 sqkm. 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 79,200 sqkm, 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 213.66% 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 sqkm

Brazil	Latin America & Caribbean	541,510
Indonesia	East Asia & Pacific	282,193.9844
Myanmar	East Asia & Pacific	107,234.0039
Nigeria	Sub-Saharan Africa	106,506.00098
Tanzania	Sub-Saharan Africa	102,320

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	75.45
Nigeria	Sub-Saharan Africa	61.80
Uganda	Sub-Saharan Africa	59.27
Mauritania	Sub-Saharan Africa	46.75
Honduras	Latin America & Caribbean	45.03

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
< 25	87
25 - 50	73
50 - 75	38
>75	9

The largest number of countries in 2016 were found in the less than 25 or 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	98.26
Palau	East Asia & Pacific	87.61
American Samoa	East Asia & Pacific	87.50
Lao PDR	East Asia & Pacific	82.11
Solomon Islands	East Asia & Pacific	77.86
Micronesia, Fed. Sts.	East Asia & Pacific	91.86
Seychelles	Sub-Saharan Africa	88.41
Guyana	Latin America & Caribbean	83.90
Gabon	Sub-Saharan Africa	90.04

5. RECOMMENDATIONS

Write out a set of recommendations as an analyst on the ForestQuery team.

- *What have you learned from the World Bank data?*

From the above data, we realize that Sub-Saharan Africa and Latin America & Caribbean were huge outliers with regards to how their decreasing values in forestation from 1990 to 2016 affected

tremendously the world's value of forestation from 1990 to 2016. Some of the challenges faced in the calculation process stemmed from missing data and huge outliers as mentioned above. However, this should not undermine the conclusions made below.

Comparisons were made between countries using the percent forestation value and amount of forestation from 1990 to 2016. We notice that, China and the United States had huge forestation values but this was not the same in terms of percent forestation as the countries' total land area being amongst the largest in the world reduced their forestation amount in terms of percent.
94 countries in 2016, had their percent forestation higher than that of the United States.

We also note that top quartile countries or top performing countries in terms of percentage for the year 2016 are mostly from the East Asia & Pacific region. It is important to research into this region's practices in favor of forestation.

- *Which countries should we focus on over others?*

The ForestQuery team should focus on Sub-Saharan Africa, especially Nigeria. As the region and Nigeria, occurred both in our data for decreasing forestation value and percent forestation.
It is important to note that, factors like total land area for various countries can overestimate or underestimate our results.

In sum, the world is faced with a decreasing forestation value hence forestation action needs to be taken now and fast.

APPENDIX: SQL queries used

```
--DEFORESTATION DATA EXPLORATION USING SQL
--Start by creating my virtual table 'forestation'. this table will contain all columns I need for further use.

CREATE VIEW forestation
AS
SELECT --selected only unique columns, what I need, to avoid duplicated columns as in the case of country_name which cuts across all tables
    r.country_name,
    r.country_code,
    r.region,
    r.income_group,
    f.year,
    f.forest_area_sqkm,
    l.total_area_sq_mi,
```

```
   CASE --creating the column with forest area as a percentage of land area. Use of case statement clears off division by zero error
        WHEN l.total_area_sq_mi = 0 OR l.total_area_sq_mi IS NULL THEN 0
        ELSE ROUND((SUM(f.forest_area_sqkm) * 100/(SUM(l.total_area_sq_mi) * 2.59))::numeric, 2)
        END AS  forest_as_percentage_of_land
FROM regions r
FULL JOIN forest_area f
ON r.country_code = f.country_code
FULL JOIN land_area l
ON f.country_code = l.country_code
AND f.year = l.year
GROUP BY 1,2,3,4,5,6,7
ORDER BY 5 DESC;

--PART 1  GLOBAL SITUATION
--Using my virtual table 'forestation' to answer the following questions:
--a) What was the total forest area (in sq km) of the world in 1990?
SELECT year, region, SUM(forest_area_sqkm) AS total_forest_area_sqkm
FROM forestation
WHERE region = 'World' AND year = '1990'
GROUP BY 1,2;

--b) What was the total forest area (in sq km) of the world in 1990?
SELECT year, region, SUM(forest_area_sqkm) AS total_forest_area_sqkm
FROM forestation
WHERE region = 'World' AND year = '2016'
GROUP BY 1,2;

--c) What was the change (in sq km) in the forest area of the world from 1990 to 2016?
--d) What was the percent change in forest area of the world between 1990 and 2016?
--Getting the difference and the percent difference from the two queries above using the WITH statement
WITH t1 AS
    (SELECT f.year, f.region, SUM(f.forest_area_sqkm) AS total_forest_area_1990
     FROM forestation f
     WHERE region = 'World' AND year = '1990'
     GROUP BY 1,2),
    t2 AS
    (SELECT f.year, f.region, SUM(f.forest_area_sqkm) AS total_forest_area_2016
     FROM forestation f
     WHERE region = 'World' AND year = '2016'
     GROUP BY 1,2)
```

```
--this query returns negative values depicting a negative change in the world's forest area from 1990 to 2016
SELECT t1.total_forest_area_1990, t2.total_forest_area_2016, (t2.total_forest_area_2016 - t1.total_forest_area_1990) AS change_in_world_forest_area,
      ROUND(((t2.total_forest_area_2016 - t1.total_forest_area_1990)*100/t1.total_forest_area_1990)::numeric, 2) AS percent_change
FROM t1
JOIN t2
ON t1.region = t2.region
GROUP BY 1,2;

--e) 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 f.year, f.region, SUM(f.forest_area_sqkm) AS total_forest_area_1
     FROM forestation f
     WHERE region = 'World' AND year = '1990'
     GROUP BY 1,2),
    t2 AS
    (SELECT f.year, f.region, SUM(f.forest_area_sqkm) AS total_forest_area_2
     FROM forestation f
     WHERE region = 'World' AND year = '2016'
     GROUP BY 1,2)

SELECT f.year, f.country_name, SUM(f.total_area_sq_mi), SUM(f.total_area_sq_mi) * 2.59 AS total_area_sqkm
FROM forestation f
WHERE year = 2016
GROUP BY 1,2
HAVING SUM(f.total_area_sq_mi) * 2.59 <=
   (SELECT (t1.total_forest_area_1 - t2.total_forest_area_2) AS change_in_world_forest_area
   FROM t1 JOIN t2 ON t1.region = t2.region)
ORDER BY 4 DESC
LIMIT 1;

--PART 2  REGIONAL OUTLOOK
--Create a table that shows the Regions and their percent forest area (sum of forest area divided by sum of land area) in 1990 and 2016. (Note that 1 sq mi = 2.59 sq km)
--This code displays data for the coulumns; region, percent forest area for 1990, percent forest area for 2016, and percent change in forest area
--We can then answer specific questions or pull specific data by using the WHERE filter
-- the codes in the latter part of the regional outlook answer specific questions. comment or uncomment them to see results
WITH t1 AS
```

```
    (SELECT f.year, f.region, SUM(f.forest_area_sqkm) * 100/(SUM(f.total_area_sq_mi) * 2.59)
AS percent_forest_area_1990
     FROM forestation f
     WHERE year = '1990'
     GROUP BY 1,2),
    t2 AS
    (SELECT f.year, f.region, SUM(f.forest_area_sqkm) * 100/(SUM(f.total_area_sq_mi) * 2.59)
AS percent_forest_area_2016
     FROM forestation f
     WHERE year = '2016'
     GROUP BY 1,2)

SELECT t1.region, t1.percent_forest_area_1990, t2.percent_forest_area_2016,
ROUND((t1.percent_forest_area_1990)::numeric, 2) AS rounded_percent_1990,
ROUND((t2.percent_forest_area_2016)::numeric, 2) AS rounded_percent_2016,
ROUND(((t2.percent_forest_area_2016 -
t1.percent_forest_area_1990)*100/t1.percent_forest_area_1990)::numeric, 2) AS
percent_change_in_forest_area
FROM t1
JOIN t2
ON t1.region = t2.region
--WHERE t1.region = 'World' --(filter for data on world)
GROUP BY 1,2,3
ORDER BY 2,3;
--ORDER BY 3 DESC LIMIT 1; --(highest percent forest area 2016)
--ORDER BY 3 ASC LIMIT 1; --(lowest percent forest area 2016)
--ORDER BY 2 DESC LIMIT 1; --(highest percent forest area 1990)
--ORDER BY 2 ASC LIMIT 1; --(lowest percent forest area 1990)
--ORDER BY 6 ASC;

--PART 3 COUNTRY-LEVEL DETAIL
--the code below privides country level data. to answer questions regarding countries, we can
navigate or order by a column name and then set a limit for cases where top five countries are of
interest to us.
WITH t1 AS
    (SELECT f.year, f.region, f.country_name, SUM(f.forest_area_sqkm) AS forest_area_1990,
SUM(f.forest_area_sqkm) * 100/(SUM(f.total_area_sq_mi) * 2.59) AS percent_forest_area_1990
     FROM forestation f
     WHERE year = '1990'
     GROUP BY 1,2,3),
    t2 AS
    (SELECT f.year, f.region, f.country_name, SUM(f.forest_area_sqkm) AS forest_area_2016,
SUM(f.forest_area_sqkm) * 100/(SUM(f.total_area_sq_mi) * 2.59) AS percent_forest_area_2016
     FROM forestation f
```

```
    WHERE year = '2016'
    GROUP BY 1,2,3)

SELECT f.region, t1.country_name, t1.forest_area_1990, t2.forest_area_2016,
t1.percent_forest_area_1990, t2.percent_forest_area_2016,
ROUND(((t2.percent_forest_area_2016 -
t1.percent_forest_area_1990)*100/t1.percent_forest_area_1990)::numeric, 2) AS
percent_change_in_forest_area, (t2.forest_area_2016 - t1.forest_area_1990) AS
difference_forest_area
FROM t1
JOIN t2
ON t1.country_name = t2.country_name
JOIN forestation f
ON t2.region = f.region
--WHERE (t2.forest_area_2016 - t1.forest_area_1990) IS NOT NULL --(ORDER BY 8 DESC
LIMIT 2; after the GROUP BY clause while removing ORDER BY 7,8 will return the top 2 countries
with forestation increase from 1990 to 2016)
--*GROUP BY 1,2,3,4,5,6
--*HAVING ROUND(((t2.percent_forest_area_2016 -
t1.percent_forest_area_1990)*100/t1.percent_forest_area_1990)::numeric, 2) IS NOT NULL
--*ORDER BY 7 DESC LIMIT 1;
--uncommenting the query with * returns the country with the largest percent change in forestation
GROUP BY 1,2,3,4,5,6
ORDER BY 7,8;
--ORDER BY 8 ASC LIMIT 6; --returns top 5 Amount Decrease in Forest Area by Country
including 'World', 1990 & 2016
--ORDER BY 7 ASC LIMIT 5; -- returns top 5 Percent Decrease in Forest Area by Country, 1990
& 2016

--finding quartiles
SELECT sub.country_name, sub.region, sub.forest_as_percentage_of_land, sub.quartiles,
      COUNT(sub.country_name) OVER (PARTITION BY sub.quartiles ORDER BY sub.quartiles)
AS country_count
FROM
   (SELECT f.country_name, f.region, f.forest_as_percentage_of_land,
    CASE WHEN f.forest_as_percentage_of_land <=25 THEN '<25'
        WHEN f.forest_as_percentage_of_land BETWEEN 25 AND 50 THEN '25 - 50'
        WHEN f.forest_as_percentage_of_land BETWEEN 50 AND 75 THEN '50 - 75'
        ELSE '> 75' END AS quartiles
   FROM forestation f
   WHERE year = 2016 AND f.forest_as_percentage_of_land IS NOT NULL
   GROUP BY 1, 2, 3) AS sub
--WHERE sub.quartiles = '> 75' --Returns the top Quartile Countries
GROUP BY 1, 2, 3, 4
```

```
ORDER BY 5 DESC;

--How many countries had a percent forestation higher than the United States in 2016?
SELECT COUNT(f.country_name) AS country_count
FROM forestation f
WHERE year = 2016 AND f.forest_as_percentage_of_land >
    (SELECT f.forest_as_percentage_of_land
     FROM forestation f
     WHERE f.year = 2016 AND f.country_name = 'United States')
;
```