SQL Case Study: Deforestation Exploration

Introduction

You're a data analyst for ForestQuery, a non-profit organization, on a mission to reduce deforestation around the world and which raises awareness about this important environmental topic.

Your executive director and her leadership team members are looking to understand which countries and regions around the world seem to have forests that have been shrinking in size, and also which countries and regions have the most significant forest area, both in terms of amount and percent of total area. The hope is that these findings can help inform initiatives, communications, and personnel allocation to achieve the largest impact with the precious few resources that the organization has at its disposal.

You've been able to find tables of data online dealing with forestation as well as total land area and region groupings, and you've brought these tables together into a database that you'd like to query to answer some of the most important questions in preparation for a meeting with the ForestQuery executive team coming up in a few days. Ahead of the meeting, you'd like to prepare and disseminate a report for the leadership team that uses complete sentences to help them understand the global deforestation overview between 1990 and 2016.

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 **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 **1279999.9891km^2**).

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
East Asia & Pacific	25.78	26.36
Europe & Central Asia	37.28	38.04
Latin America & Caribbean	51.03	46.16
Middle East & North Africa	1.78	2.07
North America	35.65	36.04
South Asia	16.51	17.51
Sub-Saharan Africa	30.67	28.79
World	32.42	31.38

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 Federation**. This country actually increased in forest area from 1990 to 2016 by **527229.06km^2.** 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 **79200km^2**, 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
Brazil	Latin America & Caribbean	541510
Indonesia	East Asia & Pacific	282193.98
Myanmar	East Asia & Pacific	107234
Nigeria	Sub-Saharan Africa	106506
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	75.45
Nigeria	Sub-Saharan Africa	61.8
Uganda	Sub-Saharan Africa	59.13
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
1	85
2	72
3	38
4	9

The largest number of countries in 2016 were found in the **first** 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
Micronesia, Fed. Sts.	East Asia & Pacific	91.86
Gabon	Sub-Saharan Africa	90.04
Seychelles	Sub-Saharan Africa	88.41
Palau	East Asia & Pacific	87.61
American Samoa	East Asia & Pacific	87.5
Guyana	Latin America & Caribbean	83.9
Lao PDR	East Asia & Pacific	82.11
Solomon Islands	East Asia & Pacific	77.86

5. RECOMMENDATIONS

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

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

  Globaly the loss in forest area from 1990 to 2016 has been immense given that it's more than the land area of the entire country of Peru.

  Regionally the Latin America & Caribbean and Sub-Saharan Africa regions respectively had the largest loss in forest area. The Middle East & North Africa had the smallest forest area back in 1990 which consequently diminished further and still remain to have the smallest forest area in 2016.

  In terms of countries, all other countries have alot to learn from China which compared to other countries had immense increase in forest area. The United States had the second largest increase in forest area but was still far less compared to that of China.

  Iceland which is among the list of very small countries had the largest percentage increase in forest area.

  Countries that had largest decrease in forest area were, Brazil, Indonesia, Myanmar, Nigeria, and Tanzania. Countries that had largest percentage decrease were Togo, Nigeria, Uganda, Mauritania, Honduras. Nigera stands out to be in both decrease categories.

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

  Focus should be placed on the countries with largest decrease in forest area and largest percentage decrease i.e. Brazil, Indonesia, Myanmar, Nigeria, Tanzania, Togo, Uganda, Mauritania, and Honduras. Efforts should be made to derail the decrease in forest area and even convert these countries to zones with increasing forest area.

  Countries with increasing forest areas should be encouraged to keep up the increase inorder to counter the overall global decline in forest area.

6. Appendix: SQL Queries Used

```
-- create a View called “forestation” by joining all three tables - forest_area, land_area and regions
CREATE OR REPLACE VIEW forestation
(country_code, country_name, year, forest_area_sqkm, land_area_country_name,
total_area_sqkm, regions_country_name, region, income_group, Percent_forest_area)
AS
SELECT forest_area.*, land_area.country_name, land_area.total_area_sq_mi*2.59,
regions.country_name, regions.region, regions.income_group,
(forest_area.forest_area_sqkm/(land_area.total_area_sq_mi*2.59))*100
FROM forest_area, land_area, regions
WHERE forest_area.country_code = land_area.country_code AND forest_area.year =
land_area.year AND forest_area.country_code = regions.country_code

-- total forest area (in sq km) of the world in 1990
SELECT forest_area_sqkm
FROM forestation
WHERE country_name = 'World' AND year = 1990;


-- total forest area (in sq km) of the world in 2016
SELECT forest_area_sqkm
FROM forestation
WHERE country_name = 'World' AND year = 2016


-- change (in sq km) in the forest area of the world from 1990 to 2016
SELECT year, forest_area_sqkm,
LAG(forest_area_sqkm) OVER (ORDER BY year)-forest_area_sqkm  AS change_in_sqkm
FROM forestation
WHERE country_name = 'World' AND year IN (1990, 2016)
ORDER BY year


-- percent change in forest area of the world between 1990 and 2016
SELECT year, forest_area_sqkm,
((LAG(forest_area_sqkm) OVER (ORDER BY year)-forest_area_sqkm)/LAG(forest_area_sqkm)
OVER (ORDER BY year))*100  AS percent_change
FROM forestation
WHERE country_name = 'World' AND year IN (1990, 2016)
```

```
ORDER BY year


-- 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?
SELECT country_name, total_area_sqkm
FROM forestation
WHERE total_area_sqkm < (SELECT change_in_sqkm
From (SELECT year, forest_area_sqkm,
LAG(forest_area_sqkm) OVER (ORDER BY year)-forest_area_sqkm  AS change_in_sqkm
FROM forestation
WHERE country_name = 'World' AND year IN (1990, 2016)
ORDER BY year)sub1
WHERE  year IN (2016))
ORDER BY total_area_sqkm DESC
LIMIT 1


-- table that shows the Regions and their percent forest area (sum of forest area divided by sum of land area) in 1990 and 2016.
SELECT region, year,
ROUND(((SUM(forest_area_sqkm)/SUM(total_area_sqkm))*100)::numeric, 2) as region_percent
INTO   TEMP TABLE temp_tbl
FROM  forestation
WHERE year IN (1990, 2016)
GROUP BY region, year;

-- percent forest of the entire world in 2016  -- 31.38
SELECT region, year, region_percent
FROM temp_tbl
WHERE  year = 2016 AND region = 'World'

-- percent forest of the entire world in 1990 -- 32.42
SELECT region, year, region_percent
FROM temp_tbl
WHERE  year = 1990 AND region = 'World'

-- region that had the HIGHEST percent forest in 2016  -- Latin America & Caribbean  46.16%
SELECT region, year, ROUND(region_percent::numeric, 2) AS region_percent2
FROM temp_tbl
WHERE  year = 2016
ORDER BY region_percent DESC
LIMIT 1
```

```
-- region that had the LOWEST percent forest in 2016  -- Middle East & North Africa          2016
2.07%
SELECT region, year, ROUND(region_percent::numeric, 2) AS region_percent2
FROM temp_tbl
WHERE  year = 2016
ORDER BY region_percent
LIMIT 1



-- region that had the LOWEST percent forest in 1990 Middle East & North Africa  1990    1.78%
SELECT region, year, ROUND(region_percent::numeric, 2) AS region_percent2
FROM temp_tbl
WHERE  year = 1990
ORDER BY region_percent
LIMIT I

-- region that had the HIGHEST percent forest in 1990 Latin America & Caribbean           1990
51.03
SELECT region, year, ROUND(region_percent::numeric, 2) AS region_percent2
FROM temp_tbl
WHERE  year = 1990
ORDER BY region_percent DESC
LIMIT 1



-- regions of the world that DECREASED in forest area from 1990 to 2016
SELECT T1.*
FROM temp_tbl AS T1
JOIN temp_tbl AS T2
   ON T1.region = T2.region
   AND T1.year != T2.year
   AND T1.region_percent < T2.region_percent
WHERE T1.year = 2016



--  5 countries saw the largest amount decrease in forest area from 1990 to 2016
WITH temp as (SELECT country_name, region, year, forest_area_sqkm
FROM forestation
WHERE year IN (1990, 2016) AND forest_area_sqkm IS NOT NULL AND total_area_sqkm IS
NOT NULL)

SELECT T1.*, ROUND((T2.forest_area_sqkm - T1.forest_area_sqkm)::numeric, 2) AS decrease
```

```
FROM temp AS T1
JOIN temp AS T2
   ON T1.country_name = T2.country_name
   AND T1.year != T2.year
   AND T1.forest_area_sqkm < T2.forest_area_sqkm
WHERE T1.year = 2016 AND  T1.country_name != 'World'
ORDER BY 5 DESC
LIMIT 5


--  5 countries saw the largest amount increase in forest area from 1990 to 2016
WITH temp as (SELECT country_name, region, year, forest_area_sqkm
FROM forestation
WHERE year IN (1990, 2016) AND forest_area_sqkm IS NOT NULL AND total_area_sqkm IS
NOT NULL)

SELECT T1.*, ROUND((T1.forest_area_sqkm - T2.forest_area_sqkm)::numeric, 2) AS increase
FROM temp AS T1
JOIN temp AS T2
   ON T1.country_name = T2.country_name
   AND T1.year != T2.year
   AND T1.forest_area_sqkm > T2.forest_area_sqkm
WHERE T1.year = 2016 AND  T1.country_name != 'World'
ORDER BY 5 DESC
LIMIT 5

--5 countries saw the largest percent increase in forest area from 1990 to 2016
WITH temp as (SELECT country_name, region, year, forest_area_sqkm
FROM forestation
WHERE year IN (1990, 2016) AND forest_area_sqkm IS NOT NULL AND total_area_sqkm IS
NOT NULL)

SELECT T1.*, ROUND((((T1.forest_area_sqkm -
T2.forest_area_sqkm)/T2.forest_area_sqkm)*100)::numeric, 2) AS percent_increase
FROM temp AS T1
JOIN temp AS T2
   ON T1.country_name = T2.country_name
   AND T1.year != T2.year
   AND T1.forest_area_sqkm > T2.forest_area_sqkm
WHERE T1.year = 2016 AND T1.country_name != 'World'
ORDER BY 5 DESC
LIMIT 5
```

```
--5 countries saw the largest decrease in forest area from 1990 to 2016
WITH temp as (SELECT country_name, region, year, forest_area_sqkm
FROM forestation
WHERE year IN (1990, 2016) AND forest_area_sqkm IS NOT NULL AND total_area_sqkm IS
NOT NULL)

SELECT T1.*, ABS((T2.forest_area_sqkm - T1.forest_area_sqkm)) AS decrease
FROM temp AS T1
JOIN temp AS T2
   ON T1.country_name = T2.country_name
   AND T1.year != T2.year
   AND T1.forest_area_sqkm < T2.forest_area_sqkm
WHERE T1.year = 2016 AND T1.country_name != 'World'
ORDER BY 5 DESC
LIMIT 5


--5 countries saw the largest percent decrease in forest area from 1990 to 2016
WITH temp as (SELECT country_name, region, year, forest_area_sqkm
FROM forestation
WHERE year IN (1990, 2016) AND forest_area_sqkm IS NOT NULL AND total_area_sqkm IS
NOT NULL)

SELECT T1.*, ROUND((((T2.forest_area_sqkm -
T1.forest_area_sqkm)/T2.forest_area_sqkm)*100)::numeric, 2) AS decrease
FROM temp AS T1
JOIN temp AS T2
   ON T1.country_name = T2.country_name
   AND T1.year != T2.year
   AND T1.forest_area_sqkm < T2.forest_area_sqkm
WHERE T1.year = 2016 AND T1.country_name != 'World'
ORDER BY 5 DESC
LIMIT 5


-- countries grouped by percent forestation in quartiles
SELECT COUNT(country_name), quartile
FROM
(SELECT country_name, year,
CASE WHEN Percent_forest_area >= 75 THEN 4
    WHEN Percent_forest_area < 75 AND Percent_forest_area >= 50 THEN 3
    WHEN Percent_forest_area < 50 AND Percent_forest_area >=25 THEN 2
    ELSE 1
    END AS quartile
```

```
FROM forestation
WHERE year = 2016 AND forest_area_sqkm IS NOT NULL AND total_area_sqkm IS NOT
NULL AND country_name != 'World' ) sub
GROUP BY quartile
ORDER BY 1
```

```
-- countries in top quartile of percent forestation
SELECT country_name, region, ROUND(Percent_forest_area::numeric, 2) AS
percent_forest_area
FROM
(SELECT country_name, year, region, Percent_forest_area,
CASE WHEN Percent_forest_area >= 75 THEN 4
    WHEN Percent_forest_area < 75 AND Percent_forest_area >= 50 THEN 3
    WHEN Percent_forest_area < 50 AND Percent_forest_area >=25 THEN 2
    ELSE 1
    END AS quartile
FROM forestation
WHERE year = 2016 AND forest_area_sqkm IS NOT NULL AND total_area_sqkm IS NOT
NULL) sub
WHERE quartile = 4
ORDER BY 3 DESC
```

```
-- countries that had a percent forestation higher than the United States in 2016
SELECT COUNT(country_name)
FROM forestation
WHERE year = 2016  AND Percent_forest_area > (SELECT  Percent_forest_area
FROM forestation
WHERE year = 2016  AND country_name ='United States')
```