-- DEFORESTATION EXPLORATION PROJECT

-- 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:

-- (4a). All of the columns of the origin tables

-- (4b). 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).

CREATE VIEW forestation AS

SELECT f.country\_code AS forest\_cc,

f.country\_name AS f\_name,

f.forest\_area\_sqkm AS f\_sqkm,

f.year AS f\_year,

l.total\_area\_sq\_mi AS land\_total\_area\_sqmi,

r.region AS r\_region,

r.income\_group AS r\_income\_group,

(f.forest\_area\_sqkm/(l.total\_area\_sq\_mi\*2.59))\*100 AS percent\_forest\_area

FROM forest\_area f

JOIN land\_area l

ON f.country\_code = l.country\_code AND f.year = l.year

JOIN regions r

ON f.country\_code = r.country\_code

-- forestation view output

SELECT \*

FROM forestation

-- PART 1: GLOBAL SITUATION

-- 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 forest\_area\_sqkm

FROM forest\_area

WHERE country\_name = 'World'

AND 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 denoted as "World" in the region table.

SELECT forest\_area\_sqkm

FROM forest\_area

WHERE country\_name= 'World'

AND year=2016;

-- c. What was the change (in sq km) in the forest area of the world from 1990 to 2016?

-- this question can be solved in two different ways;

-- first solution using CASE

-- The CASE expression goes through conditions and returns a value when the first condition is met

--(like an if-then-else statement). So, once a condition is true, it will stop reading and return

--the result. If no conditions are true, it returns the value in the ELSE clause.

--If there is no ELSE part and no conditions are true, it returns NULL.

--CASE Syntax

--CASE

--WHEN condition1 THEN result1

--WHEN condition2 THEN result2

--WHEN conditionN THEN resultN

--ELSE result

--END;

SELECT Max(CASE

WHEN year = 1990 THEN forest\_area\_sqkm

END) - Max(CASE

WHEN year = 2016 THEN forest\_area\_sqkm

END) AS diff\_forest\_area

FROM forest\_area

WHERE country\_name = 'World';

-- second solution

SELECT ( sub1.forest\_area\_sqkm - sub2.forest\_area\_sqkm ) AS

diff\_forest\_area\_sqkm

FROM (SELECT f.country\_code AS cc,

f.forest\_area\_sqkm

FROM forest\_area f

WHERE f.country\_name = 'World'

AND f.year = 1990) AS sub1

JOIN (SELECT f.country\_code AS cc,

f.forest\_area\_sqkm

FROM forest\_area f

WHERE f.country\_name = 'World'

AND f.year = 2016) AS sub2

ON sub1.cc = sub2.cc;

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

SELECT Abs(( Max(CASE

WHEN year = 2016 THEN forest\_area\_sqkm

END) - Max(CASE

WHEN year = 1990 THEN forest\_area\_sqkm

END) ) / Max(CASE

WHEN year = 1990 THEN forest\_area\_sqkm

END)) \* 100 AS percent\_diff\_forest\_area

FROM forest\_area

WHERE country\_name = 'World'

--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?

--Common Table Expressions (CTE)

--We can define CTEs by adding a WITH clause directly before SELECT, INSERT, UPDATE, DELETE,

--or MERGE statement.

--After you define your WITH clause with the CTEs, you can then reference the CTEs as you would

--refer any other table. However, you can refer a CTE only within the execution scope of the

--statement that immediately follows the WITH clause.

WITH cte AS

(

SELECT l.country\_name,

l.total\_area\_sq\_mi\*2.59 AS total\_area\_sqkm,

Abs((l.total\_area\_sq\_mi\*2.59)-

(

SELECT Max(

CASE

WHEN year=1990 THEN forest\_area\_sqkm

END)- Max(

CASE

WHEN year=2016 THEN forest\_area\_sqkm

END) AS diff\_forest\_area

FROM forest\_area

WHERE country\_name= 'World')) AS diff\_total\_fa\_land

FROM land\_area l

WHERE l.year=2016

ORDER BY 3

LIMIT 1)

SELECT \*

FROM cte

WHERE diff\_total\_fa\_land IS NOT NULL;

-- PART 2: REGIONAL OUTLOOK

-- (2). 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).

CREATE VIEW regional\_table

AS

(SELECT r.region,

l.year,

SUM(f.forest\_area\_sqkm) AS

total\_fa\_sqkm,

SUM(l.total\_area\_sq\_mi \* 2.59) AS

total\_la\_sqkm,

( SUM(f.forest\_area\_sqkm) / SUM(l.total\_area\_sq\_mi \* 2.59) ) AS

percent\_fa\_region

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

GROUP BY r.region,

l.year

--ORDER BY r.region, l.year

)

-- regional\_table view output

SELECT \*

FROM regional\_table

-- (2a.1). What was the percent forest of the entire world in 2016?

-- to 2 decimal places?

-- first solution

SELECT ROUND(CAST(SUM(total\_fa\_sqkm)/SUM(total\_la\_sqkm) AS numeric) \*100, 2)

AS perc\_fa\_region

FROM regional\_table

WHERE region = 'World'

AND year = 2016;

-- second solution

-- The CAST() function converts a value (of any type) into a specified datatype.

SELECT ROUND(CAST(percent\_fa\_region AS numeric) \* 100, 2) AS perc\_fa\_region

FROM regional\_table

WHERE region = 'World'

AND year = 2016;

-- (2a.2) Which region had the HIGHEST percent

--forest in 2016, and which had the LOWEST,

--to 2 decimal places?

-- regional\_table view output

SELECT \*

FROM regional\_table

SELECT region, ROUND(CAST(SUM(total\_fa\_sqkm)/SUM(total\_la\_sqkm) AS numeric)

\*100, 2) AS highest\_per\_fa

FROM regional\_table

WHERE year = 2016

GROUP BY region

ORDER BY highest\_per\_fa DESC;

-- (2b.1) What was the percent forest of the entire world in 1990?

--to 2 decimal places?

SELECT ROUND(CAST(SUM(total\_fa\_sqkm)/SUM(total\_la\_sqkm) AS numeric) \*100, 2)

AS perc\_fa\_region

FROM regional\_table

WHERE region = 'World'

AND year = 1990;

-- (2b.2) Which region had the HIGHEST percent forest in 1990, and which had the LOWEST,

--to 2 decimal places?

-- regional\_table view output

SELECT \*

FROM regional\_table

-- year 1990 % forest area

SELECT region, ROUND(CAST(SUM(total\_fa\_sqkm)/SUM(total\_la\_sqkm) AS numeric)

\*100, 2) AS highest\_per\_fa

FROM regional\_table

WHERE year = 1990

GROUP BY region

ORDER BY highest\_per\_fa DESC;

-- year 2016 % forest area

SELECT region, ROUND(CAST(SUM(total\_fa\_sqkm)/SUM(total\_la\_sqkm) AS numeric)

\*100, 2) AS highest\_per\_fa

FROM regional\_table

WHERE year = 2016

GROUP BY region

ORDER BY highest\_per\_fa DESC;

-- (2c) Based on the table you created, which regions of the world DECREASED

--in forest area from 1990 to 2016?

--solution using a Subquery WITH placement

WITH table\_1990 AS (

SELECT region, ROUND(CAST(SUM(total\_fa\_sqkm)/SUM(total\_la\_sqkm) AS numeric)

\*100, 2) AS highest\_per\_fa\_1990

FROM regional\_table

WHERE year = 1990

GROUP BY region),

table\_2016 AS (

SELECT region, ROUND(CAST(SUM(total\_fa\_sqkm)/SUM(total\_la\_sqkm) AS numeric)

\*100, 2) AS highest\_per\_fa\_2016

FROM regional\_table

WHERE year = 2016

GROUP BY region)

--ORDER BY highest\_per\_fa DESC

SELECT table\_1990.region, highest\_per\_fa\_1990, highest\_per\_fa\_2016

FROM table\_1990

JOIN table\_2016

ON table\_1990.region = table\_2016.region

WHERE highest\_per\_fa\_1990 > highest\_per\_fa\_2016;

-- PART 3. COUNTRY-LEVEL DETAIL

-- (3a). Which 5 countries saw the largest amount decrease in forest area from 1990 to 2016? What was

--the difference in forest area for each?

--SUCCESS STORIES

WITH top\_5\_countries

AS (SELECT fa\_1990.country\_name,

Sum(fa\_2016.forest\_area\_sqkm) - Sum(fa\_1990.forest\_area\_sqkm) AS

diff\_fa

FROM forest\_area fa\_1990

JOIN forest\_area fa\_2016

ON fa\_1990.country\_code = fa\_2016.country\_code

WHERE fa\_1990.year = 1990

AND fa\_2016.year = 2016

GROUP BY fa\_1990.country\_name

--ORDER BY 2

)

SELECT t.country\_name, r.region,

t. diff\_fa

FROM top\_5\_countries t

JOIN regions r

ON r.country\_name = t.country\_name

WHERE t. diff\_fa IS NOT NULL

AND t.country\_name != 'World'

ORDER BY 3 DESC

--LARGEST CONCERNS

WITH top\_5\_countries

AS (SELECT fa\_1990.country\_name,

Sum(fa\_2016.forest\_area\_sqkm) - Sum(fa\_1990.forest\_area\_sqkm) AS

diff\_fa

FROM forest\_area fa\_1990

JOIN forest\_area fa\_2016

ON fa\_1990.country\_code = fa\_2016.country\_code

WHERE fa\_1990.year = 1990

AND fa\_2016.year = 2016

GROUP BY fa\_1990.country\_name

--ORDER BY 2

)

SELECT t.country\_name, r.region,

t. diff\_fa

FROM top\_5\_countries t

JOIN regions r

ON r.country\_name = t.country\_name

WHERE t. diff\_fa IS NOT NULL

AND t.country\_name != 'World'

ORDER BY 3

LIMIT 5;

--LARGEST CONCERNS IN %

WITH top\_5\_countries

AS (SELECT fa\_1990.country\_name,

Sum(fa\_2016.forest\_area\_sqkm) - Sum(fa\_1990.forest\_area\_sqkm) AS

diff\_fa,

Sum((fa\_2016.forest\_area\_sqkm) - (fa\_1990.forest\_area\_sqkm))/ Sum(fa\_1990.forest\_area\_sqkm) \* 100 AS per\_diff\_fa

FROM forest\_area fa\_1990

JOIN forest\_area fa\_2016

ON fa\_1990.country\_code = fa\_2016.country\_code

WHERE fa\_1990.year = 1990

AND fa\_2016.year = 2016

GROUP BY fa\_1990.country\_name

--ORDER BY 2

)

SELECT t.country\_name, r.region,

t. diff\_fa, t.per\_diff\_fa

FROM top\_5\_countries t

JOIN regions r

ON r.country\_name = t.country\_name

WHERE t. diff\_fa IS NOT NULL

AND t.country\_name != 'World'

ORDER BY 3

LIMIT 5;

SELECT \*

FROM forest\_area

-- (3b). 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 top\_5\_countries AS

(

SELECT fa\_1990.country\_code,

fa\_1990.country\_name,

fa\_1990.forest\_area\_sqkm - fa\_2016.forest\_area\_sqkm AS fa\_diff\_askdlnl,

(fa\_2016.forest\_area\_sqkm - fa\_1990.forest\_area\_sqkm)/fa\_1990.forest\_area\_sqkm\*100 AS percent\_diff

FROM forest\_area fa\_1990

JOIN forest\_area fa\_2016

ON fa\_1990.country\_code=fa\_2016.country\_code

WHERE fa\_1990.year =1990

AND fa\_2016.year= 2016

--ORDER BY 2

)

SELECT t.country\_name,

r.region,

fa\_diff\_askdlnl,

percent\_diff

FROM top\_5\_countries t

JOIN regions r

ON r.country\_code=t.country\_code

WHERE fa\_diff\_askdlnl IS NOT NULL

AND t.country\_name != 'World'

ORDER BY percent\_diff

LIMIT 5;

-- (3c). If countries were grouped by percent forestation in quartiles, which group had the most countries

--in it in 2016?

WITH country\_fa

AS (SELECT f.country\_code,

f.country\_name,

(SUM(f.forest\_area\_sqkm) / SUM(l.total\_area\_sq\_mi \* 2.59 )) \* 100 AS country\_percent

FROM forest\_area f

JOIN land\_area l

ON f.country\_code = l.country\_code

AND f.year = l.year

WHERE f.year = 2016 AND l.year = 2016

AND f.country\_name != 'World'

AND (f.forest\_area\_sqkm IS NOT NULL AND l.total\_area\_sq\_mi IS NOT NULL)

GROUP BY f.country\_code, f.country\_name),

quartiles

AS (SELECT country\_code,

country\_name,

--NTILE(4) OVER(ORDER BY country\_percent desc) AS quartile

CASE

WHEN country\_percent >= 75 THEN 4

WHEN country\_percent < 75

AND country\_percent >= 50 THEN 3

WHEN country\_percent < 50

AND country\_percent >= 25 THEN 2

ELSE 1

END AS quartile

FROM country\_fa)

SELECT quartile,

Count(quartile) AS number\_of\_countries

FROM quartiles

GROUP BY quartile

ORDER BY quartile;

-- (3d). List all of the countries that were in the 4th quartile (percent forest > 75%) in 2016.

WITH country\_fa AS (

SELECT f.country\_code,f.country\_name,f.year,f.forest\_area\_sqkm,l.total\_area\_sq\_mi\*2.59 AS land\_area

FROM forest\_area f

JOIN land\_area l

ON f.country\_code = l.country\_code AND f.year=l.year

WHERE f.year=2016 and f.country\_name != 'World'

),

country\_percent\_fa AS(

SELECT country\_code,country\_name,year,(forest\_area\_sqkm/land\_area) \*100 AS country\_percent

FROM country\_fa

),

quartiles AS (

SELECT country\_code,country\_percent,

country\_name,year,

CASE WHEN country\_percent >= 75 THEN 4

WHEN country\_percent < 75 AND country\_percent >= 50 THEN 3

WHEN country\_percent < 50 AND country\_percent >=26 THEN 2

ELSE 1

END AS quartile

FROM country\_percent\_fa

)

SELECT quartile,q.country\_name,r.region,country\_percent

FROM quartiles q

JOIN regions r

ON r.country\_code=q.country\_code

WHERE quartile = 4

ORDER BY 3 DESC;

-- (3e). How many countries had a percent forestation higher than the United States in 2016?

WITH country\_fa

AS (SELECT f.country\_code,

f.country\_name,

f.year,

f.forest\_area\_sqkm,

l.total\_area\_sq\_mi \* 2.59 AS land\_area

FROM forest\_area f

JOIN land\_area l

ON f.country\_code = l.country\_code

AND f.year = l.year

WHERE f.year = 2016

AND f.country\_name != 'World'),

country\_percent\_fa

AS (SELECT country\_code,

country\_name,

year,

( forest\_area\_sqkm / land\_area ) \* 100 AS country\_percent

FROM country\_fa),

quartiles

AS (SELECT country\_code,

country\_name,

year,

--NTILE(4) OVER(ORDER BY country\_percent desc) AS quartile

CASE

WHEN country\_percent >= 75 THEN 4

WHEN country\_percent < 75

AND country\_percent >= 50 THEN 3

WHEN country\_percent < 50

AND country\_percent >= 26 THEN 2

ELSE 1

END AS quartile

FROM country\_percent\_fa)

SELECT COUNT(\*) AS number\_of\_countries

FROM country\_percent\_fa

WHERE year = 2016

AND country\_percent > (SELECT country\_percent

FROM country\_percent\_fa

WHERE country\_name = 'United States'

AND year = 2016);