Paypal Project

A few important things to note before you get started:

- All work should be done in SQL. Any variant is fine (e.g. MS SQL, Postgres, MySQL, etc.). If you normally use R, SAS, or other similar tools with SQL it's important that you show that you can work in SQL by itself to produce the correct answers to this challenge.
- If you are confused by a specific question, you can request clarification by replying to the message that this challenge was attached to. This is NOT intended for you to obtain technical help with solving the problems on this challenge or asking for hints; it should only be used for question clarification.
- This challenge is due back within 1 week (7 calendar days) of being sent to you.
- If you cannot answer a question, please do your best, show your work, leave comments, and let us know your general thoughts.
- We are interested in BOTH your answers and the work/code that you write to get them. Please leave plenty of comments for us to read when we review your work.
- There are some blank/null values in this set. That's how we found it and it reflects the nature of imperfect data. Please work with this and provide explanations of any issues or workarounds required to arrive at your answers.
- There are no intentional gotchas, trick questions, or traps; the challenge is intended to demonstrate some of the typical day-to-day SQL skills that the job requires.
- Some of these questions may seem nonsensical and you may find yourself asking, "why would anyone want to know that?!" They are intended purely as a measure of your SQL skills and _not_ as actual questions that we would expect to ask of this type of data set. Please take them with a grain of salt.

Instructions

- Create a SQL database using the attached CSV files.
- Use the database to answer the following questions.
- All answers that return money values should be rounded to 2 decimal points and preceded by the "$" symbol (e.g. "$1432.10").
- All answers that return percent values should be between -100.00 to 100.00, rounded to 2 decimal points and followed by the "%" symbol (e.g. "58.30%").

Code Challenge v2.11

1. Data Integrity Checking & Cleanup

- Alphabetically list all of the country codes in the continent_map table that appear more than once. Display any values where country_code is null as country_code = "FOO" and make this row appear first in the list, even though it should alphabetically sort to the middle. Provide the results of this query as your answer.

- For all countries that have multiple rows in the continent_map table, delete all multiple records leaving only the 1 record per country. The record that you keep should be the first one when sorted by the continent_code alphabetically ascending. Provide the query/ies and explanation of step(s) that you follow to delete these records.

2. List the countries ranked 10-12 in each continent by the percent of year-over-year growth descending from 2011 to 2012.

The percent of growth should be calculated as: ((2012 gdp - 2011 gdp) / 2011 gdp)

The list should include the columns:

- rank
- continent_name
- country_code
- country_name
- growth_percent

3. For the year 2012, create a 3 column, 1 row report showing the percent share of gdp_per_capita for the following regions:

(i) Asia, (ii) Europe, (iii) the Rest of the World. Your result should look something like

Asia	Europe	Rest of World
25.0%	47.0%	28.0%

4a. What is the count of countries and sum of their related gdp_per_capita values for the year 2007 where the string 'an' (case insensitive) appears anywhere in the country name?

4b. Repeat question 4a, but this time make the query case sensitive.

5. Find the sum of gpd_per_capita by year and the count of countries for each year that have non-null gdp_per_capita where (i) the year is before 2012 and (ii) the country has a null gdp_per_capita in 2012. Your result should have the columns:

- year
- country_count
- total

6. All in a single query, execute all of the steps below and provide the results as your final answer:

a. create a single list of all per_capita records for year 2009 that includes columns:

- continent_name
- country_code
- country_name
- gdp_per_capita

b. order this list by:

- continent_name ascending
- characters 2 through 4 (inclusive) of the country_name descending

c. create a running total of gdp_per_capita by continent_name

d. return only the first record from the ordered list for which each continent's running total of gdp_per_capita meets or exceeds $70,000.00 with the following columns:

- continent_name
- country_code
- country_name
- gdp_per_capita
- running_total

7. Find the country with the highest average gdp_per_capita for each continent for all years. Now compare your list to the following data set. Please describe any and all mistakes that you can find with the data set below. Include any code that you use to help detect these mistakes.

rank	continent_name	country_code	country_name	avg_gdp_per_capita
1	Africa	SYC	Seychelles	$11,348.66
1	Asia	KWT	Kuwait	$43,192.49
1	Europe	MCO	Monaco	$152,936.10
1	North America	BMU	Bermuda	$83,788.48
1	Oceania	AUS	Australia	$47,070.39
1	South America	CHL	Chile	$10,781.71