Tourism Case Study – Assignment Guide

Below is a suggested guide to help you solve the business problem. Be aware that there are multiple solutions to this problem and that you do not need to follow the steps below. Note: we suggest you save your code as a R script (.R file) as you work through this case study.

Create the Cleaned_Tourism Table

Part 1
1. If necessary, set your working directory. Read the **tourism** table and create the **cleaned_tourism** table.
2. Remove the columns **_1995** through **_2013**.
3. Create the **Country_Name** and **Tourism_Type** columns from values in the **Country** column. Valid values for **Tourism_Type** are *Inbound tourism* and *Outbound tourism*. Remove rows that contain this labeling information and no other data.

Part 2
4. In the **Series** column, convert values to uppercase and convert ".." or " " to missing value.
5. Determine the conversion type (Mn or Thousands) that will be used to calculate values for the new **Y2014** column. **Hint**: You might want to create a new column with this information.
6. In the **_2014** column, change the data not available (values of ".." or " ") to a missing value.

Part 3
7. Create the **Y2014** column by explicitly converting character values in **_2014** to numeric and multiplying by the conversion type (millions or thousands) that is found in the **Country** column or new column, if you created one.
8. Create the new **Category** column from values in the **Country** column and change the original values to the following valid values:
   - Arrivals
   - Departures
   - Passenger Transport - US$
   - Tourism expenditure in other countries - US$
   - Tourism expenditure in the country – US$
   - Travel - US$

9. Include only **Country_Name**, **Tourism_Type**, **Category**, **Series**, and **Y2014** in the output table.

Create the Final_Tourism Table
1. Create a factor variable for the **Continent** column that labels continent IDs with the corresponding continent names:
   - 1 = North America
   - 2 = South America
   - 3 = Europe
   - 4 = Africa
   - 5 = Asia
   - 6 = Oceania
   - 7 = Antarctica

2. Join the **cleaned_tourism** table with **country_info** to create the **final_tourism** table. Include only matches in the output table. Use the new factor labels for **Continent**.

Create the NoCountryFound Table

Create the **NoCountryFound** table that has a list of countries from **Cleaned_Tourism** that are not found in the **country_info** table. This table should have only one row for each country.