

Tourism Case Study - Requirements

Your supervisor has emailed the requirements and desired outcome for this project. Be sure to read through all the requirements thoroughly.

Raw Data

The data for this case study is the **Tourism** data located in **CR** library.

A	COUNTRY	Series	1995	1996	1997	1998	1999	2000	2001	2002	2003
826	UNITED KINGDOM										
	Inbound tourism										
	Arrivals - Thousands	TF	21,719	22,936	23,215	23,710	23,341	23,212	20,982	22,307	22,787
	Tourism expenditure in the country - US\$ Mn	IMF	27,577	29,181	30,483	31,658	30,807	29,978	26,137	27,819	30,736
	Travel - US\$ Mn	IMF	20,487	21,389	22,586	23,689	22,716	21,769	18,864	20,549	22,668
	Passenger transport - US\$ Mn	IMF	7,090	7,792	7,897	7,969	8,091	8,209	7,273	7,270	8,068
	Outbound tourism										
	Departures - Thousands	TF	41,345	42,050	45,957	50,872	53,881	56,837	58,281	59,377	61,424
	Tourism expenditure in other countries - US\$ Mn	IMF	30,749	32,298	35,954	41,458	45,536	47,009	46,410	51,125	58,627
	Travel - US\$ Mn	IMF	24,926	25,962	28,529	33,452	37,034	38,262	37,931	41,744	47,853
	Passenger transport - US\$ Mn	IMF	5,823	6,336	7,425	8,006	8,502	8,747	8,479	9,381	10,774

Desired Outcome

The final table should be a combination of the **cleaned_tourism** table and the **country_info** table. See the example below:

COUNTRY_NAME	TOURISM_TYPE	CATEGORY	SERIES	Y2014	CONTINENT
UNITED KINGDOM	Inbound tourism	Arrivals	TF	32,613,000	Europe
UNITED KINGDOM	Inbound tourism	Tourism expenditure in the country - US\$	IMF	62,830,000,000	Europe
UNITED KINGDOM	Inbound tourism	Travel - US\$	IMF	46,723,000,000	Europe
UNITED KINGDOM	Inbound tourism	Passenger transport - US\$	IMF	16,107,000,000	Europe
UNITED KINGDOM	Outbound tourism	Departures	TF	60,082,000	Europe
UNITED KINGDOM	Outbound tourism	Tourism expenditure in other countries - US\$	IMF	79,935,000,000	Europe
UNITED KINGDOM	Outbound tourism	Travel - US\$	IMF	63,424,000,000	Europe
UNITED KINGDOM	Outbound tourism	Passenger transport - US\$	IMF	16,511,000,000	Europe

Data Requirements

Create the **cleaned_tourism** table with the following column requirements:

- Country_Name** contains the country name from the original **Country** column.
- Tourism_Type** contains the type of tourism from the original **Country** column. Valid values are *Inbound tourism* or *Outbound tourism*.
- Category** contains category names by extracting and modifying values from the original **Country** column. There should be six distinct values for **Category** as shown in the table below:

Values in the Country Column	New Values for the Category Column
Arrivals - Thousands	Arrivals
Departures - Thousands	Departures

Values in the Country Column	New Values for the Category Column
Passenger transport US\$ Mn	Passenger transport US\$
Tourism expenditure in other countries US\$ Mn	Tourism expenditure in other countries US\$
Tourism expenditure in the country US\$ Mn	Tourism expenditure in the country US\$
Travel US\$ Mn	Travel US\$

- **Series** All values should be in uppercase and data that is not available (coded as "..") should be changed to a missing character value.
- **Y2014** contains numeric values that are calculated from the scaled character values in the original **_2014** year column. The scaled values are multiplied by either thousands or millions (abbreviated Mn), depending on the value listed for category in the **Country** column. The new **Y2014** values should be formatted with the COMMA format.
 - Example: if the category is **Travel US\$ MN** and the value for **_2014** is 4.26, **Y2014** is equal to $4.26 * 1000000$, or 4,260,000.
 - Include only **Country_Name**, **Tourism_Type**, **Category**, **Series**, and **Y2014** in the output table.

Merge the cleaned_tourism table with the country_info table and do the following:

1. Create two new tables:
 - **final_tourism** should contain **only** merged data.
 - **nocountryfound** should contain a list of distinct countries from the **cleaned_tourism** table that do not have a match in the **country_info** table.
2. Create a format for the **Continent** column that labels continent IDs with the corresponding continent names. Permanently apply the format in the **final_tourism** table.
 - 1 = North America
 - 2 = South America
 - 3 = Europe
 - 4 = Africa
 - 5 = Asia
 - 6 = Oceania
 - 7 = Antarctica