SAS Base Programming 2:
Case Study Submission

Submitted by:
Mariam Ayman Moustafa Mahmoud

```
/* Data Preparation */
data cleaned_tourism;
	length Country_Name $300 Tourism_Type $20;
	retain Country_Name "" Tourism_Type "";
	set work.tourism (drop=_1995-_2013);

	/* Retain the country name if not missing */
	if A ne . then
		Country_Name = Country;

	/* Identify and set the type of tourism */
	if lowcase(Country) = 'inbound tourism' then
		Tourism_Type = "Inbound tourism";
	else if lowcase(Country) = 'outbound tourism' then
		Tourism_Type = "Outbound tourism";

	/* Exclude rows where the country name matches the tourism type */
	if Country_Name ne Country and Country ne Tourism_Type;

	/* Convert the series to uppercase */
	Series = upcase(Series);

	/* Handle missing or placeholder values in the Series variable */
	if Series = ".." then
		Series = "";

	/* Extract and process the conversion type from the country column */
	ConversionType = strip(scan(country, -1, ' '));

	/* Handle missing values in the 2014 data */
	if _2014 = '..' then
		_2014 = '.';

	/* Convert 2014 data based on the unit of measurement */
	if ConversionType = 'Mn' then
		do;
			/* Convert millions to actual values */
			if input(_2014, 16.) ne . then
				Y2014 = input(_2014, 16.) * 1000000;
			else
				Y2014 = .;
			Category = cat(scan(country, 1, '-', 'r'), " - US$");
		end;
	else if ConversionType = 'Thousands' then
		do;
			/* Convert thousands to actual values */
			if input(_2014, 16.) ne . then
				Y2014 = input(_2014, 16.) * 1000;
```

```
                else
                        Y2014 = .;
                Category = scan(country, 1, '-', 'r');
        end;

    /* Format the 2014 values with commas */
    format Y2014 comma25.;

    /* Drop unnecessary columns */
    drop A ConversionType Country _2014;
run;

/* Define a format for continents */
proc format;
    value continents 1="North America"
                     2="South America"
                     3="Europe"
                     4="Africa"
                     5="Asia"
                     6="Oceania"
                     7="Antarctica";
run;

/* Sorting the country information dataset by country name */
proc sort data=work.country_info(rename=(Country=Country_Name))
        out=Country_Sorted;
    by Country_Name;
run;

/* Merging the cleaned tourism data with the country information */
data final_tourism nocountryfound(keep=Country_Name);
    merge cleaned_tourism(in=t) country_sorted(in=c);
    by country_name;

    /* Output to final_tourism if there is a match in both datasets */
    if t = 1 and c = 1 then
        output final_tourism;

    /* Output to nocountryfound if the country is in tourism data but not in country info */
    if (t = 1 and c = 0) and first.country_name then
        output nocountryfound;

    /* Apply the continent format */
    format Continent continents.;
run;

/* Validation: List countries not found in the country_info table */
proc print data=nocountryfound;
```

```
    title "List of Countries Not Found in the Country_Info Table";
run;

/* Validation: Frequency counts for key variables */
proc freq data=final_tourism;
    tables Category Series Tourism_Type Continent / nocum nopercent;
    title "Frequency Counts of Category, Series, Tourism_Type, and Continent";
run;

/* Validation: Calculate mean, minimum, and maximum for the year 2014 */
proc means data=final_tourism mean min max maxdec=0;
    var Y2014;
    title "Mean, Minimum, and Maximum for the Year 2014";
run;
```