1. **Sorting Data to Remove Duplicate Rows**

The pg1.np\_largeparks table contains gross acreage for large national parks. There are

duplicate rows for some locations.

a. Open and review the pg1.np\_largeparks table. Notice that there are exact duplicate rows

for some parks.

b. Create a new program. Write a PROC SORT step that creates two tables (park\_clean and

park\_dups), and removes the duplicate rows. Submit the program.

2 **Creating New Columns with Character and Date Functions**

* 1. **a.** Open a new program. Write a DATA step to create a temporary table named **eu\_occ\_total** based on the **pg1.eu\_occ** table. Create the following new columns:

The **pg1.eu\_occ** table contains individual columns for nights spent at hotels, short stay accommodations, or camps for each year and month. The **YearMon** column is character.

• **Year** – the four-digit year extracted from **YearMon**.

• **Month** – the two-digit month extracted from **YearMon**.

• **ReportDate** – the first day of the reporting month.

**Hint:** Use the MDY function and the new **Year** and **Month** columns.

• **Total** – the total nights spent at any establishment. Format the new column to display the values with commas.

**b.** Format **Hotel**, **ShortStay**, **Camp**, and **Total** with commas. Format **ReportDate** to display the values in the form JAN2018.

* 1. **c.** Keep **Country**, **Hotel**, **ShortStay**, **Camp**, **ReportDate**, and **Total** in the new table.
  2. **3. Processing Statements Conditionally with DO Groups**
  3. **a.** Create a new program. Write a DATA step to create two temporary tables named **parks** and **monuments** based on the **pg1.np\_summary** table. Read only national parks or monuments from the input table. (**Type** is either *NP* or *NM*.)
  4. **b.** Create a new column named **Campers** that is the sum of all columns containing counts of campers. Format the column to include commas.
  5. **c.** When **Type** is *NP*, create a new column named **ParkType** that is equal to *Park*, and write the row to the **parks** table. When **Type** is *NM*, assign **ParkType** as *Monument* and write the row to the **monuments** table.
  6. **d.** Keep **Reg**, **ParkName**, **DayVisits**, **OtherLodging**, **Campers**, and **ParkType** in both output tables.

Use conditional processing to split **pg1.np\_summary** into two tables: **parks** and **monuments**.

Pentru a putea realiza tema trebuie creata in prealabil libraria PG1:

libname pg1 "path\_of\_data\_folder"; unde path\_of\_data\_folder se gaseste: Files->Courses->PG1V2->data

