Module Code

MANG1041

Assignment Title

Individual Coursework

Student ID

33245215

/*Task 1 {Import the data from names.csv to SAS}*/

```
PROC IMPORT DATAFILE='/home/u63290591/sasuser.v94/names.csv' DBMS=csv
   /*PROC IMPORT tell SAS to import "names.csv" files*/
   /*DBMS specifies the type of data to import is CSV*/
OUT=work.names REPLACE;
   /*OUT option tells SAS to put the new the new SAS dataset "names" in the Work library*/
   /*REPLACE statement tells SAS  if there is an existing dataset in SAS's memory with this
     name, it should be overwritten*/
RUN;
```

/*Task 2 {Create a new data set(female_name)that only contains those names which were given to female babies. Include only the variables name and count}*/

```
DATA work.female_names;
   /*DATA tells SAS to create a "female_names" dataset in the Work library*/
SET work.names;
   /*The function of the SET statement is to process existing "names" SAS datasets in Work
     library as input for DATA step above*/
WHERE gender='F';
   /*WHERE statement tell SAS only selects the observation that "gender" is "female"  in
     "names" SAS datasets only*/
DROP gender;
   /*DROP option tells SAS to drop variable "gender" from a "names" dataset*/
RUN;
```

/*Task 3 {In the data female_names:knowing that the total number of female babies is 180623104,create a new variable(fraction)that is equal to count divided by the total number of female babies}*/

```
DATA work.female_names;
   /*DATA tells SAS to create a "female_names" dataset in the Work library*/
SET work.names;
   /*The function of the SET statement is to process existing "names" SAS datasets in Work
     library as input for DATA step above*/
WHERE gender='F';
   /*WHERE statement tell SAS only selects the observation that "gender" is "female"  in
     "names" SAS datasets only*/
DROP gender;
   /*DROP option tells SAS to drop variable "gender" from a "names" dataset*/
Fraction=count/180623104;
   /*Create a new variable "Fraction" that  is equal to "count" divided by the total number
     of female babies "180623104" */
RUN;
```

/*Task 4 {Sort the data set female_name by count(order:largest to smallest)}*/

```
DATA work.female_names;
   /*DATA tells SAS to create a "female_names" dataset in the Work library*/
SET work.names;
   /*The function of the SET statement is to process existing "names" SAS datasets in Work
     library as input for DATA step above*/
WHERE gender='F';
   /*WHERE statement tell SAS only selects the observation that "gender" is "female" in
     "names" SAS datasets only*/
DROP gender;
   /*DROP option tells SAS to drop variable "gender" from a "names" dataset*/
Fraction=count/180623104;
   /*Create a new variable "Fraction" that is equal to "count" divided by the total number
     of female babies "180623104" */
RUN;

PROC SORT DATA=work.female_names;
   /*PROC SORT DATA tell SAS sort data from "female_names" dataset in the Work library
     */
BY DESCENDING Count;
   /*The data is sorted first BY "count", in DESCENDING order.*/
RUN;
```

/*Task 5 {Once the data set female_names is sorted, print the first 35 observations from the data set to find the most popular female names}*/

```
DATA work.female_names;
   /*DATA tells SAS to create a "female_names" dataset in the Work library*/
SET work.names;
   /*The function of the SET statement is to process existing "names" SAS datasets in Work
     library as input for DATA step above*/
WHERE gender='F';
   /*WHERE statement tell SAS only selects the observation that "gender" is "female" in
     "names" SAS datasets only*/
DROP gender;
   /*DROP option tells SAS to drop variable "gender" from a "names" dataset*/
Fraction=count/180623104;
   /*Create a new variable "Fraction" that is equal to "count" divided by the total number
     of female babies "180623104" */
RUN;

PROC SORT DATA=work.female_names;
   /*PROC SORT DATA tell SAS sort data from "female_names" dataset in the Work library
     */
BY DESCENDING Count;
   /*The data is sorted first BY "count", in DESCENDING order.*/
RUN;

PROC PRINT DATA=work.female_names (OBS=35);
   /*OBS options tell SAS which range of first 35 observation numbers to PRINT*/
RUN;
```

/*Task 6 {Create a new data set (male_names) that only contains those names which were given to male babies. Include only the variables name and count}*/

```
DATA work.male_names;
   /*DATA tells SAS to create a "male_names" dataset in the Work library*/
SET work.names;
   /*The function of the SET statement is to process existing "names" SAS datasets in Work
     library as input for DATA step above*/
WHERE gender='M';
   /*WHERE statement tell SAS only selects the observation that "gender" is "male"  in
     "names" SAS datasets only*/
DROP gender;
   /*DROP option tells SAS to drop variable "gender" from a "names" dataset*/
RUN;
```

/*Task 7 {In the data set male_names: knowing that the total number of male babies is 184775046, create a new variable (fraction) that is equal to count divided by the total number of male babies}*/

```
DATA work.male_names;
   /*DATA tells SAS to create a "male_names" dataset in the Work library*/
SET work.names;
   /*The function of the SET statement is to process existing "names" SAS datasets in Work
     library as input for DATA step above*/
WHERE gender='M';
   /*WHERE statement tell SAS only selects the observation that "gender" is "male"  in
     "names" SAS datasets only*/
DROP gender;
   /*DROP option tells SAS to drop variable "gender" from a "names" dataset*/
        Fraction=count/184775046;
   /*Create a new variable "Fraction" that  is equal to "count" divided by the total number
     of male babies "184775046" */
RUN;
```

/*Task 8 {Sort the data set male_names by fraction (order: largest to smallest)}*/

```
DATA work.male_names;
   /*DATA tells SAS to create a "male_names" dataset in the Work library*/
SET work.names;
   /*The function of the SET statement is to process existing "names" SAS datasets in Work
     library as input for DATA step above*/
WHERE gender='M';
   /*WHERE statement tell SAS only selects the observation that "gender" is "male"  in
     "names" SAS datasets only*/
DROP gender;
   /*DROP option tells SAS to drop variable "gender" from a "names" dataset*/
        Fraction=count/184775046;
   /*Create a new variable "Fraction" that  is equal to "count" divided by the total number
     of male babies "184775046" */
RUN;

PROC SORT DATA=work.male_names;
   /*PROC SORT DATA tell SAS sort data from "male_names" dataset in the Work library */
BY DESCENDING fraction;
   /*The data is sorted first BY "fraction", in DESCENDING order.*/
RUN;
```

/*Task 9 {Once the data set male_ names is sorted, print the data where fraction is greater than 0.005 to find
the most popular male names}*/

```
DATA work.male_names;
   /*DATA tells SAS to create a "male_names" dataset in the Work library*/
SET work.names;
   /*The function of the SET statement is to process existing "names" SAS datasets in Work
     library as input for DATA step above*/
WHERE gender='M';
   /*WHERE statement tell SAS only selects the observation that "gender" is "male"  in
     "names" SAS datasets only*/
DROP gender;
   /*DROP option tells SAS to drop variable "gender" from a "names" dataset*/
        Fraction=count/184775046;
   /*Create a new variable "Fraction" that  is equal to "count" divided by the total number
     of male babies "184775046" */
RUN;

PROC SORT DATA=work.male_names;
   /*PROC SORT DATA tell SAS sort data from "male_names" dataset in the Work library */
BY DESCENDING fraction;
   /*The data is sorted first BY "fraction", in DESCENDING order.*/
RUN;

PROC PRINT DATA=work.male_names;
   /*PROC PRINT DATA tell SAS to print data from "male_names" dataset in the Work
   library*/
WHERE fraction>0.005;
   /*WHERE statement tell SAS only selects the observation "fraction" is "greater than 0.005"
     from "male"  in "names" SAS datasets */
RUN;
```