```
/*************************************************************************
*******************/
/* Program Name:     pulkit.jain_HW12.sas
                                          */
/* Program Location: C:\Users\Pulkit
Jain\Documents\sasuniversityedition\myfolders\assign12 */
/* Date Created:     10/29/2017
                                          */
/* Author:              Pulkit Jain
                                       */
/* Purpose:             Assignment 12, converting data types / structure
                                    */
/*************************************************************************
*******************/

/* Create two libname statements;                              */
/* Assign library to locaion of hw data with access only; */
/* Assign another library with read and write access;     */

libname hw_data '/folders/myfolders/hw_data' access=readonly;
libname pulkit12 '/folders/myfolders/assign12';

/* Specify a fileref to designate output of pdf */

filename HW12 '/folders/myfolders/assign12/pulkit.jain_HW12_output.pdf';

/* 2 Use zip_codes data as input      */
/* Create temporary dataset "cleaned_up_zips" */

/*  retain specific variables only in resulting data*/
data work.cleaned_up_zips(KEEP= zip timezone primary_city state county
estimated_population);
   set hw_data.zip_codes;
/*  convert type of county and estimated population    */
    county2 = Input(county, $31.);
    estimated_population2 = INPUT(estimated_population, 8.);
    drop county estimated_population;
    rename county2 = county;
    rename estimated_population2 = estimated_population;
/*  remove observations which are decpmmissioned, and specific states */
    if decommissioned = 1 then delete;
    if state in ('AA', 'AE','AP') then delete;
/*  remove the word county, parish and Borough    */
    county2 = TRANWRD(county2,'County','');
    county2 = TRANWRD(county2,'Parish','');
    if FIND(county2, ' Borough ') = 0 then
      county2 = TRANWRD(county2,' Borough','');
/*  remove underscore in time zones      */
    if timezone = 'America/Los_Angeles' then
      substr(timezone, 12, 1) = ' ';
    else if timezone = 'America/New_York' then
      substr(timezone, 12, 1) = ' ';
    else if timezone = 'America/Puerto_Rico' then
      substr(timezone, 15, 1) = ' ';
```

```
/*  change labels       */
    label zip = 'Zip Code';
    label timezone = 'Time Zone';
    label primary_city = 'City';
    label state = 'State';
    label county = 'County';
    label estimated_population = 'Est. Population';
    label county2 = 'County';
    label estimated_population2 = 'Est. Population';
run;

/* 3 */

proc sort data = work.cleaned_up_zips;
/*  sort data so that it can be processed in grouping*/
/*  sort first by state and second by primary city*/
      by state primary_city;
run;



data work.summary_zips(DROP = estimated_population zip timezone);
      set work.cleaned_up_zips;
/*    set labels and maximum length */
      length zip_codes $1700;
      label zip_codes = 'Zip Codes';
      label est_city_population = 'Est. City Population';
/*    group and create summary statistics*/
      by state primary_city;
      if First.primary_city = 1 then do;
            est_city_population = 0;
            zip_codes = '';
      end;
      retain est_city_population 0;
      retain zip_codes '0';
      est_city_population = sum(est_city_population,
estimated_population);
      zip_codes = CATX(',', zip_codes, zip);
      if Last.primary_city = 1;
/*    remove observations where population is zero and change its format
*/
      if est_city_population = 0 then delete;
      format est_city_population comma10. ;
run;


/* 4 PDF output file so that bookmarks are not created*/

ods pdf file = HW12 bookmarkgen= no;

/*  5 Print the two data steps contents and output for limited cities*/

title '4.1 Descriptor Portion of Cleaned Zip Code Data Set';
Proc contents data = work.cleaned_up_zips;
```

```
run;

title '4.2 Cleaned Zip Codes from Selected Cities';
Proc print data = work.cleaned_up_zips label;
      var zip primary_city state timezone county estimated_population;
      where propcase(primary_city) IN ('Buffalo', 'Center', 'Las Vegas',
'Bristow',
                                                      'Athens',
'Carolina', 'Auke Bay', 'Muleshoe',
                                                      'Washington');
run;

title '4.3 Descriptor Portion of Summarized Zip Codes Data Set';
Proc contents data = work.summary_zips;
run;



title '4.4 Summarized Zip Codes from Selected Cities';
Proc print data = work.summary_zips label;
      var primary_city state county zip_codes est_city_population;
      where propcase(primary_city) IN ('Buffalo', 'Center', 'Las Vegas',
'Bristow',
                                                      'Athens',
'Carolina', 'Auke Bay', 'Muleshoe',
                                                      'Washington');
run;

ods pdf close;
ods listing;
```