```
/*************************************************************************/
/*Program Name- HW12.sas*/
/* Date Created: November 1 2017 */
/* Author: Dilip Lalwani */
/* Purpose: SAS and working with datasets */

/*1 Output filename and libname statements*/
libname data "/folders/myfolders/data" access=readonly ;
libname hw12 "/folders/myfolders/HW12" ;
filename output "/folders/myfolders/HW12/dilip.k.lalwani_HW12_output.pdf";
data hw12.zip_codes(drop=estpopulation lastword_county decommissioned);

/*2a Reduce the length of the county variable to 31 */
length county $ 31;
set data.zip_codes(keep=county decommissioned estimated_population primary_city state timezone zip
rename=(estimated_population=estpopulation));

/*2 Remove decommissioned zip codes*/
if decommissioned eq 1 then delete;

/*2 Remove observations with states equal to AE, AA or AP*/
if state in ('AE', 'AA', 'AP') then delete;


/*2a Use manipulation functions to modify county values*/
lastword_county=scan(county,-1);
if UPCASE(lastword_county) in ('COUNTY','PARISH','BOROUGH') then do;
county=substr(county,1,length(county)-length(scan(county,-1)));
end;
```

```
/*2b Convert estimated_population variable from character to numeric*/
estimated_population=input(estpopulation, 8.);

/*2c Replace underscores with blank space in timezone variable*/
if timezone eq "America/New_York" then
do;
        substr(timezone,12,1)=' ';
end;
else if timezone eq "America/Los_Angeles" then
do;
        substr(timezone,12,1)=' ';
end;
else if timezone eq "America/Puerto_Rico" then
do;
        substr(timezone,15,1)=' ';
end;
else if timezone eq "America/Indiana/Tell_City" then
do;
        substr(timezone,21,1)=' ';
end;
else if timezone eq "America/North_Dakota/Center" then
do;
        substr(timezone,14,1)=' ';
end;

label zip="Zip Code"
primary_city="City"
state="State"
```

```
timezone="Time Zone"
county="County"
estimated_population="Est Population";
run;

/*3a Use a sort procedure to sort the clean data set - sorting by state and primary_city*/
proc sort data=hw12.zip_codes;
by state primary_city;
run;

/*3b Remove zip and estimated_population from the dataset*/
data work.zipstats(drop=zip estimated_population timezone);
set hw12.zip_codes;
retain zip_codes;
length zip_codes $1700;
by state primary_city;

/*3c Store total of estimated population values for each city in est_city_population*/
if First.primary_city then do;
est_city_population = 0;
zip_codes='';
end;
est_city_population+estimated_population;

/*3d List all zips into Zip Codes variable and create summary*/
zip_codes=catx(',',zip_codes, zip);
if Last.primary_city;
label est_city_population="Est. City Population"
zip_codes="Zip Codes"
```

```
primary_city="City"
state="State"
county="County";
format est_city_population COMMA10.;

/*3e Output cities with estimated city population greater than 0*/
if est_city_population<=0 then delete;
run;

/*4 Open PDF Destination and output observations only for selected cities*/
ods pdf file=output bookmarkgen=no;
proc contents data=hw12.zip_codes;
title "4.1 Descriptor Portion of Cleaned Zip Code Data Set";
run;
proc print data =hw12.zip_codes label;
title "4.2 Cleaned Zip Codes from Selected Cities";
where (primary_city ='Center'
or primary_city='Buffalo'
or primary_city='Las Vegas'
or primary_city='Bristow'
or primary_city='Muleshoe'
or primary_city='Athens'
or primary_city='Carolina'
or primary_city='Auke Bay'
or primary_city='Washington');
var zip primary_city state timezone county estimated_population;
run;
proc contents data=work.zipstats;
title "4.3 Descriptor Portion of Summarized Zip Codes Data Set";
```

```
run;
proc print data =work.zipstats label;
title "4.4 Summarized Zip Codes from Selected Cities";
where (primary_city ='Center'
or primary_city='Buffalo'
or primary_city='Las Vegas'
or primary_city='Bristow'
or primary_city='Muleshoe'
or primary_city='Athens'
or primary_city='Carolina'
or primary_city='Auke Bay'
or primary_city='Washington');
var primary_city state county zip_codes est_city_population;
run;
ods pdf close;
```