```
 1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
 NOTE: ODS statements in the SAS Studio environment may disable some
output features.
 62
 63
/*************************************************************************
*******************/
 64         /* Program Name:     pulkit.jain_HW12.sas    */
 65         /* Program Location: C:\Users\Pulkit
Jain\Documents\sasuniversityedition\myfolders\assign12 */
 66         /* Date Created:     10/29/2017          */
 67         /* Author:  Pulkit Jain                      */
 68         /* Purpose:   Assignment 12, converting data types /
structure    */
 69
/*************************************************************************
*******************/
 70
 71         /* Create two libname statements;   */
 72         /* Assign library to locaion of hw data with access only; */
 73         /* Assign another library with read and write access;     */
 74
 75         libname hw_data '/folders/myfolders/hw_data' access=readonly;
 NOTE: Libref HW_DATA was successfully assigned as follows:
       Engine:        V9
       Physical Name: /folders/myfolders/hw_data
 76         libname pulkit12 '/folders/myfolders/assign12';
 NOTE: Libref PULKIT12 was successfully assigned as follows:
       Engine:        V9
       Physical Name: /folders/myfolders/assign12
 77
 78         /* Specify a fileref to designate output of pdf */
 79
 80         filename HW12
'/folders/myfolders/assign12/pulkit.jain_HW12_output.pdf';
 81
 82         /* 2 Use zip_codes data as input      */
 83         /* Create temporary dataset "cleaned_up_zips" */
 84
 85         /*  retain specific variables only in resulting data*/
 86         data work.cleaned_up_zips(KEEP= zip timezone primary_city
state county estimated_population);
 87            set hw_data.zip_codes;
 NOTE: Data file HW_DATA.ZIP_CODES.DATA is in a format that is native to
another host, or the file encoding does not match the
       session encoding. Cross Environment Data Access will be used,
which might require additional CPU resources and might reduce
       performance.
 88         /*  convert type of county and estimated population    */
 89             county2 = Input(county, $31.);
 90             estimated_population2 = INPUT(estimated_population, 8.);
 91             drop county estimated_population;
 92             rename county2 = county;
```

```
 93             rename estimated_population2 = estimated_population;
 94         /*  remove observations which are decpmmissioned, and
specific states */
 95             if decommissioned = 1 then delete;
 96             if state in ('AA', 'AE','AP') then delete;
 97         /*  remove the word county, parish and Borough    */
 98             county2 = TRANWRD(county2,'County','');
 99             county2 = TRANWRD(county2,'Parish','');
 100            if FIND(county2, ' Borough ') = 0 then
 101            county2 = TRANWRD(county2,' Borough','');
 102        /*  remove underscore in time zones    */
 103            if timezone = 'America/Los_Angeles' then
 104            substr(timezone, 12, 1) = ' ';
 105            else if timezone = 'America/New_York' then
 106            substr(timezone, 12, 1) = ' ';
 107            else if timezone = 'America/Puerto_Rico' then
 108            substr(timezone, 15, 1) = ' ';
 109        /*  change labels    */
 110            label zip = 'Zip Code';
 111            label timezone = 'Time Zone';
 112            label primary_city = 'City';
 113            label state = 'State';
 114            label county = 'County';
 115            label estimated_population = 'Est. Population';
 116            label county2 = 'County';
 117            label estimated_population2 = 'Est. Population';
 118        run;

 NOTE: Character values have been converted to numeric values at the
places given by: (Line):(Column).
       95:8
 NOTE: There were 42522 observations read from the data set
HW_DATA.ZIP_CODES.
 NOTE: The data set WORK.CLEANED_UP_ZIPS has 41317 observations and 6
variables.
 NOTE: DATA statement used (Total process time):
       real time           0.28 seconds
       cpu time            0.23 seconds


 119
 120        /* 3 */
 121
 122        proc sort data = work.cleaned_up_zips;
 123        /*  sort data so that it can be processed in grouping*/
 124        /*  sort first by state and second by primary city*/
 125         by state primary_city;
 126        run;

 NOTE: There were 41317 observations read from the data set
WORK.CLEANED_UP_ZIPS.
 NOTE: The data set WORK.CLEANED_UP_ZIPS has 41317 observations and 6
variables.
 NOTE: PROCEDURE SORT used (Total process time):
```

```
      real time           0.03 seconds
      cpu time            0.03 seconds


 127
 128
 129
 130        data work.summary_zips(DROP = estimated_population zip
timezone);
 131        set work.cleaned_up_zips;
 132        /* set labels and maximum length */
 133        length zip_codes $1700;
 134        label zip_codes = 'Zip Codes';
 135        label est_city_population = 'Est. City Population';
 136        /* group and create summary statistics*/
 137        by state primary_city;
 138        if First.primary_city = 1 then do;
 139        est_city_population = 0;
 140        zip_codes = '';
 141        end;
 142        retain est_city_population 0;
 143        retain zip_codes '0';
 144        est_city_population = sum(est_city_population,
estimated_population);
 145        zip_codes = CATX(',', zip_codes, zip);
 146        if Last.primary_city = 1;
 147        /* remove observations where population is zero and change
its format */
 148        if est_city_population = 0 then delete;
 149        format est_city_population comma10. ;
 150        run;

 NOTE: There were 41317 observations read from the data set
WORK.CLEANED_UP_ZIPS.
 NOTE: The data set WORK.SUMMARY_ZIPS has 21404 observations and 5
variables.
 NOTE: DATA statement used (Total process time):
      real time           0.06 seconds
      cpu time            0.06 seconds

 150      !

 151
 152
 153        /* 4 PDF output file so that bookmarks are not created*/
 154
 155        ods pdf file = HW12 bookmarkgen= no;
 NOTE: Writing ODS PDF output to DISK destination "HW12", printer "PDF".
 156
 157        /*  5 Print the two data steps contents and output for
limited cities*/
 158
 159        title '4.1 Descriptor Portion of Cleaned Zip Code Data Set';
 160        Proc contents data = work.cleaned_up_zips;
```

```
 161        run;

 NOTE: PROCEDURE CONTENTS used (Total process time):
       real time           0.11 seconds
       cpu time            0.11 seconds


 162
 163        title '4.2 Cleaned Zip Codes from Selected Cities';
 164        Proc print data = work.cleaned_up_zips label;
 165        var zip primary_city state timezone county
estimated_population;
 166        where propcase(primary_city) IN ('Buffalo', 'Center', 'Las
Vegas', 'Bristow',
 167         'Athens', 'Carolina', 'Auke Bay', 'Muleshoe',
 168         'Washington');
 169        run;

 NOTE: There were 489 observations read from the data set
WORK.CLEANED_UP_ZIPS.
       WHERE PROPCASE(primary_city) in ('Athens', 'Auke Bay', 'Bristow',
'Buffalo', 'Carolina', 'Center', 'Las Vegas', 'Muleshoe',
       'Washington');
 NOTE: PROCEDURE PRINT used (Total process time):
       real time           1.12 seconds
       cpu time            1.12 seconds


 170
 171        title '4.3 Descriptor Portion of Summarized Zip Codes Data
Set';
 172        Proc contents data = work.summary_zips;
 173        run;

 NOTE: PROCEDURE CONTENTS used (Total process time):
       real time           0.07 seconds
       cpu time            0.08 seconds


 174
 175
 176
 177        title '4.4 Summarized Zip Codes from Selected Cities';
 178        Proc print data = work.summary_zips label;
 179        var primary_city state county zip_codes est_city_population;
 180        where propcase(primary_city) IN ('Buffalo', 'Center', 'Las
Vegas', 'Bristow',
 181         'Athens', 'Carolina', 'Auke Bay', 'Muleshoe',
 182         'Washington');
 183        run;

 NOTE: There were 61 observations read from the data set
WORK.SUMMARY_ZIPS.
```

```
      WHERE PROPCASE(primary_city) in ('Athens', 'Auke Bay', 'Bristow',
'Buffalo', 'Carolina', 'Center', 'Las Vegas', 'Muleshoe',
      'Washington');
 NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.19 seconds
      cpu time            0.20 seconds


 184
 185        ods pdf close;
 NOTE: ODS PDF printed 20 pages to
/folders/myfolders/assign12/pulkit.jain_HW12_output.pdf.
 186        ods listing;
 187
 188        OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
 201
```