```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
68
69          **************************************************************
70          *1.Import the Bay Area House Price data set from the file Bay Area House Price.csv'.
71           Name the data set as house_price.
72          **************************************************************;
73
74         proc import datafile='/home/u63743369/week8/Bay Area House Price.csv'
75             out=work.house_price
76             dbms=csv
77             replace;
78             getnames=yes;
79         run;
```

NOTE: Unable to open parameter catalog: SASUSER.PARMS.PARMS.SLIST in update mode. Temporary parameter values will be saved to
WORK.PARMS.PARMS.SLIST.

```
80          /**********************************************************************
81          *    PRODUCT:   SAS
82          *    VERSION:   9.4
83          *    CREATOR:   External File Interface
84          *    DATE:      08MAR24
85          *    DESC:      Generated SAS Datastep Code
86          *    TEMPLATE SOURCE:  (None Specified.)
87          **********************************************************************/
88             data WORK.HOUSE_PRICE    ;
89             %let _EFIERR_ = 0; /* set the ERROR detection macro variable */
90             infile '/home/u63743369/week8/Bay Area House Price.csv' delimiter = ',' MISSOVER DSD lrecl=32767 firstobs=2 ;
91                informat address $34. ;
92                informat info $98. ;
93                informat z_address $25. ;
94                informat bathrooms best32. ;
95                informat bedrooms best32. ;
96                informat finishedsqft best32. ;
97                informat lastsolddate mmddyy10. ;
98                informat lastsoldprice best32. ;
99                informat latitude best32. ;
100               informat longitude best32. ;
101               informat neighborhood $40. ;
102               informat totalrooms best32. ;
103               informat usecode $12. ;
104               informat yearbuilt best32. ;
105               informat zestimate best32. ;
106               informat zipcode best32. ;
107               informat zpid best32. ;
108               format address $34. ;
109               format info $98. ;
110               format z_address $25. ;
111               format bathrooms best12. ;
112               format bedrooms best12. ;
113               format finishedsqft best12. ;
114               format lastsolddate mmddyy10. ;
115               format lastsoldprice best12. ;
116               format latitude best12. ;
117               format longitude best12. ;
118               format neighborhood $40. ;
119               format totalrooms best12. ;
120               format usecode $12. ;
121               format yearbuilt best12. ;
122               format zestimate best12. ;
123               format zipcode best12. ;
124               format zpid best12. ;
125           input
126                       address  $
127                       info  $
128                       z_address  $
129                       bathrooms
130                       bedrooms
131                       finishedsqft
132                       lastsolddate
133                       lastsoldprice
134                       latitude
135                       longitude
136                       neighborhood  $
137                       totalrooms
138                       usecode  $
139                       yearbuilt
140                       zestimate
141                       zipcode
142                       zpid
```

```
143              ;
144              if _ERROR_ then call symputx('_EFIERR_',1);  /* set ERROR detection macro variable */
145              run;
```

NOTE: The infile '/home/u63743369/week8/Bay Area House Price.csv' is:
      Filename=/home/u63743369/week8/Bay Area House Price.csv,
      Owner Name=u63743369,Group Name=oda,
      Access Permission=-rw-r--r--,
      Last Modified=04Mar2024:23:11:08,
      File Size (bytes)=2842323

NOTE: 11330 records were read from the infile '/home/u63743369/week8/Bay Area House Price.csv'.
      The minimum record length was 220.
      The maximum record length was 291.
NOTE: The data set WORK.HOUSE_PRICE has 11330 observations and 17 variables.
NOTE: DATA statement used (Total process time):
      real time   0.02 seconds
      user cpu time   0.02 seconds
      system cpu time   0.00 seconds
      memory   12100.31k
      OS Memory   32544.00k
      Timestamp   03/08/2024 04:36:30 AM
      Step Count   24 Switch Count 2
      Page Faults   0
      Page Reclaims   540
      Page Swaps   0
      Voluntary Context Switches   11
      Involuntary Context Switches 0
      Block Input Operations   3600
      Block Output Operations   7184

11330 rows created in WORK.HOUSE_PRICE from /home/u63743369/week8/Bay Area House Price.csv.

NOTE: WORK.HOUSE_PRICE data set was successfully created.
NOTE: The data set WORK.HOUSE_PRICE has 11330 observations and 17 variables.
NOTE: PROCEDURE IMPORT used (Total process time):
      real time   0.13 seconds
      user cpu time   0.11 seconds
      system cpu time   0.02 seconds
      memory   12100.31k
      OS Memory   33060.00k
      Timestamp   03/08/2024 04:36:30 AM
      Step Count   24 Switch Count 8
      Page Faults   0
      Page Reclaims   8085
      Page Swaps   0
      Voluntary Context Switches   98
      Involuntary Context Switches 0
      Block Input Operations   5552
      Block Output Operations   7280

146
147   **
148   *2.Drop the variables: address, info, z_address, neighborhood,
149   latitude, longitude, and zpid both using Data Statement and PROC SQL.
150   Name the new data set as house_price.
151   **;
152
153   data house_price(drop=address info z_address neighborhood latitude longitude zpid);
154   set house_price;
155   run;

NOTE: There were 11330 observations read from the data set WORK.HOUSE_PRICE.
NOTE: The data set WORK.HOUSE_PRICE has 11330 observations and 10 variables.
NOTE: DATA statement used (Total process time):
      real time   0.00 seconds
      user cpu time   0.00 seconds
      system cpu time   0.00 seconds
      memory   3279.25k
      OS Memory   28080.00k
      Timestamp   03/08/2024 04:36:30 AM
      Step Count   25 Switch Count 2
      Page Faults   0
      Page Reclaims   466
      Page Swaps   0
      Voluntary Context Switches   11
      Involuntary Context Switches 0
```

```
        Block Input Operations            0
        Block Output Operations        2064


156
157
158        proc import datafile='/home/u63743369/week8/Bay Area House Price.csv'
159            out=work.house_price
160            dbms=csv
161            replace;
162            getnames=yes;
163        run;

NOTE: Unable to open parameter catalog: SASUSER.PARMS.PARMS.SLIST in update mode. Temporary parameter values will be saved to
WORK.PARMS.PARMS.SLIST.
164        /********************************************************************
165        *    PRODUCT:   SAS
166        *    VERSION:   9.4
167        *    CREATOR:   External File Interface
168        *    DATE:      08MAR24
169        *    DESC:      Generated SAS Datastep Code
170        *    TEMPLATE SOURCE:  (None Specified.)
171        *********************************************************************/
172           data WORK.HOUSE_PRICE    ;
173           %let _EFIERR_ = 0; /* set the ERROR detection macro variable */
174           infile '/home/u63743369/week8/Bay Area House Price.csv' delimiter = ',' MISSOVER DSD lrecl=32767 firstobs=2 ;
175              informat address $34. ;
176              informat info $98. ;
177              informat z_address $25. ;
178              informat bathrooms best32. ;
179              informat bedrooms best32. ;
180              informat finishedsqft best32. ;
181              informat lastsolddate mmddyy10. ;
182              informat lastsoldprice best32. ;
183              informat latitude best32. ;
184              informat longitude best32. ;
185              informat neighborhood $40. ;
186              informat totalrooms best32. ;
187              informat usecode $12. ;
188              informat yearbuilt best32. ;
189              informat zestimate best32. ;
190              informat zipcode best32. ;
191              informat zpid best32. ;
192              format address $34. ;
193              format info $98. ;
194              format z_address $25. ;
195              format bathrooms best12. ;
196              format bedrooms best12. ;
197              format finishedsqft best12. ;
198              format lastsolddate mmddyy10. ;
199              format lastsoldprice best12. ;
200              format latitude best12. ;
201              format longitude best12. ;
202              format neighborhood $40. ;
203              format totalrooms best12. ;
204              format usecode $12. ;
205              format yearbuilt best12. ;
206              format zestimate best12. ;
207              format zipcode best12. ;
208              format zpid best12. ;
209           input
210                       address   $
211                       info   $
212                       z_address   $
213                       bathrooms
214                       bedrooms
215                       finishedsqft
216                       lastsolddate
217                       lastsoldprice
218                       latitude
219                       longitude
220                       neighborhood   $
221                       totalrooms
222                       usecode   $
223                       yearbuilt
224                       zestimate
225                       zipcode
226                       zpid
227           ;
228           if _ERROR_ then call symputx('_EFIERR_',1);  /* set ERROR detection macro variable */
229           run;
```

```
NOTE: The infile '/home/u63743369/week8/Bay Area House Price.csv' is:
      Filename=/home/u63743369/week8/Bay Area House Price.csv,
      Owner Name=u63743369,Group Name=oda,
      Access Permission=-rw-r--r--,
      Last Modified=04Mar2024:23:11:08,
      File Size (bytes)=2842323

NOTE: 11330 records were read from the infile '/home/u63743369/week8/Bay Area House Price.csv'.
      The minimum record length was 220.
      The maximum record length was 291.
NOTE: The data set WORK.HOUSE_PRICE has 11330 observations and 17 variables.
NOTE: DATA statement used (Total process time):
      real time             0.02 seconds
      user cpu time         0.02 seconds
      system cpu time       0.01 seconds
      memory                10699.75k
      OS Memory             32800.00k
      Timestamp             03/08/2024 04:36:30 AM
      Step Count                        26  Switch Count  2
      Page Faults                       0
      Page Reclaims                     308
      Page Swaps                        0
      Voluntary Context Switches        11
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           7184


11330 rows created in WORK.HOUSE_PRICE from /home/u63743369/week8/Bay Area House Price.csv.



NOTE: WORK.HOUSE_PRICE data set was successfully created.
NOTE: The data set WORK.HOUSE_PRICE has 11330 observations and 17 variables.
NOTE: PROCEDURE IMPORT used (Total process time):
      real time             0.10 seconds
      user cpu time         0.10 seconds
      system cpu time       0.02 seconds
      memory                10699.75k
      OS Memory             33316.00k
      Timestamp             03/08/2024 04:36:30 AM
      Step Count                        26  Switch Count  10
      Page Faults                       0
      Page Reclaims                     5501
      Page Swaps                        0
      Voluntary Context Switches        77
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           7200


230
231        proc sql;
232            create table house_price as
233            select *
234            from house_price
235            (drop=address info z_address neighborhood latitude longitude zpid);
WARNING: This CREATE TABLE statement recursively references the target table. A consequence of this is a possible data integrity
         problem.
NOTE: Table WORK.HOUSE_PRICE created, with 11330 rows and 10 columns.

236        quit;
NOTE: PROCEDURE SQL used (Total process time):
      real time             0.00 seconds
      user cpu time         0.00 seconds
      system cpu time       0.00 seconds
      memory                6958.87k
      OS Memory             32688.00k
      Timestamp             03/08/2024 04:36:30 AM
      Step Count                        27  Switch Count  2
      Page Faults                       0
      Page Reclaims                     476
      Page Swaps                        0
      Voluntary Context Switches        12
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           2056


237
```

```
238          **************************************************************
239          *3.Add a new variable price_per_square_foot defined by
240          lastsoldprice/finishedsqft both using Data Statement and PROC SQL.
241          **************************************************************;
242
243          data house_price;
244              set house_price;
245              price_per_square_foot = lastsoldprice / finishedsqft;
246          run;
```

NOTE: There were 11330 observations read from the data set WORK.HOUSE_PRICE.
NOTE: The data set WORK.HOUSE_PRICE has 11330 observations and 11 variables.
NOTE: DATA statement used (Total process time):
      real time   0.00 seconds
      user cpu time   0.00 seconds
      system cpu time   0.00 seconds
      memory   2985.50k
      OS Memory   28336.00k
      Timestamp   03/08/2024 04:36:30 AM
      Step Count   28 Switch Count 2
      Page Faults   0
      Page Reclaims   382
      Page Swaps   0
      Voluntary Context Switches   11
      Involuntary Context Switches   0
      Block Input Operations   0
      Block Output Operations   2320


```
247
248
249          proc sql;
250              create table house_price as
251              select *,
252                  lastsoldprice / finishedsqft as price_per_square_foot
253              from house_price;
```
WARNING: This CREATE TABLE statement recursively references the target table. A consequence of this is a possible data integrity
         problem.
WARNING: Variable price_per_square_foot already exists on file WORK.HOUSE_PRICE.
NOTE: Table WORK.HOUSE_PRICE created, with 11330 rows and 11 columns.

```
254          quit;
```
NOTE: PROCEDURE SQL used (Total process time):
      real time   0.00 seconds
      user cpu time   0.00 seconds
      system cpu time   0.00 seconds
      memory   6669.40k
      OS Memory   32688.00k
      Timestamp   03/08/2024 04:36:30 AM
      Step Count   29 Switch Count 2
      Page Faults   0
      Page Reclaims   417
      Page Swaps   0
      Voluntary Context Switches   11
      Involuntary Context Switches   0
      Block Input Operations   0
      Block Output Operations   2312


```
255
256
257          **************************************************************
258          *4.Find the average of lastsoldprice by zipcode both using Data
259           Statement and PROC SQL.
260          **************************************************************;
261
262          proc sort data=house_price;
263              by zipcode;
264          run;
```

NOTE: There were 11330 observations read from the data set WORK.HOUSE_PRICE.
NOTE: The data set WORK.HOUSE_PRICE has 11330 observations and 11 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time   0.00 seconds
      user cpu time   0.01 seconds
      system cpu time   0.01 seconds
      memory   4177.06k
      OS Memory   29888.00k
      Timestamp   03/08/2024 04:36:30 AM
      Step Count   30 Switch Count 2
      Page Faults   0

```
                Page Reclaims                    673
                Page Swaps                       0
                Voluntary Context Switches       13
                Involuntary Context Switches     0
                Block Input Operations           0
                Block Output Operations          2312


265
266            proc means data=house_price noprint;
267                by zipcode;
268                output out=avg_lastsoldprice(drop=_type_ _freq_) mean=average_lastsoldprice;
269            run;

NOTE: There were 11330 observations read from the data set WORK.HOUSE_PRICE.
NOTE: The data set WORK.AVG_LASTSOLDPRICE has 25 observations and 2 variables.
NOTE: PROCEDURE MEANS used (Total process time):
                real time               0.00 seconds
                user cpu time           0.00 seconds
                system cpu time         0.00 seconds
                memory                  3060.62k
                OS Memory               28852.00k
                Timestamp               03/08/2024 04:36:30 AM
                Step Count                       31  Switch Count  5
                Page Faults                      0
                Page Reclaims                    598
                Page Swaps                       0
                Voluntary Context Switches       20
                Involuntary Context Switches     0
                Block Input Operations           0
                Block Output Operations          264


270
271
272            proc sql;
273                create table avg_lastsoldprice_zipcode as
274                select zipcode,
275                       mean(lastsoldprice) as average_lastsoldprice
276                from house_price
277                group by zipcode;
NOTE: Table WORK.AVG_LASTSOLDPRICE_ZIPCODE created, with 25 rows and 2 columns.

278            quit;
NOTE: PROCEDURE SQL used (Total process time):
                real time               0.00 seconds
                user cpu time           0.01 seconds
                system cpu time         0.00 seconds
                memory                  6665.81k
                OS Memory               32688.00k
                Timestamp               03/08/2024 04:36:30 AM
                Step Count                       32  Switch Count  2
                Page Faults                      0
                Page Reclaims                    334
                Page Swaps                       0
                Voluntary Context Switches       11
                Involuntary Context Switches     0
                Block Input Operations           0
                Block Output Operations          264


279
280
281            ************************************************************
282            *5.Find the average of lastsoldprice by usecode, totalrooms,
283             and bedrooms both using Data Statement and PROC SQL.
284            ************************************************************;
285
286            proc sort data=house_price;
287                by usecode totalrooms bedrooms;
288            run;

NOTE: There were 11330 observations read from the data set WORK.HOUSE_PRICE.
NOTE: The data set WORK.HOUSE_PRICE has 11330 observations and 11 variables.
NOTE: PROCEDURE SORT used (Total process time):
                real time               0.00 seconds
                user cpu time           0.00 seconds
                system cpu time         0.00 seconds
                memory                  4198.03k
                OS Memory               29888.00k
                Timestamp               03/08/2024 04:36:30 AM
```

```
             Step Count                    33  Switch Count  2
             Page Faults                   0
             Page Reclaims                 683
             Page Swaps                    0
             Voluntary Context Switches    10
             Involuntary Context Switches  0
             Block Input Operations        0
             Block Output Operations       2312


289
290        proc means data=house_price noprint;
291            by usecode totalrooms bedrooms;
292            output out=avg_lastsoldprice(drop=_type_ _freq_) mean=average_lastsoldprice;
293        run;

NOTE: There were 11330 observations read from the data set WORK.HOUSE_PRICE.
NOTE: The data set WORK.AVG_LASTSOLDPRICE has 388 observations and 4 variables.
NOTE: PROCEDURE MEANS used (Total process time):
             real time            0.00 seconds
             user cpu time        0.01 seconds
             system cpu time      0.00 seconds
             memory               3077.62k
             OS Memory            28852.00k
             Timestamp            03/08/2024 04:36:30 AM
             Step Count                    34  Switch Count  5
             Page Faults                   0
             Page Reclaims                 438
             Page Swaps                    0
             Voluntary Context Switches    20
             Involuntary Context Switches  0
             Block Input Operations        0
             Block Output Operations       264


294
295
296
297        proc sql;
298            create table avg_lastsoldprice as
299            select usecode, totalrooms, bedrooms,
300                  mean(lastsoldprice) as average_lastsoldprice
301            from house_price
302            group by usecode, totalrooms, bedrooms;
NOTE: Table WORK.AVG_LASTSOLDPRICE created, with 388 rows and 4 columns.

303        quit;
NOTE: PROCEDURE SQL used (Total process time):
             real time            0.00 seconds
             user cpu time        0.00 seconds
             system cpu time      0.00 seconds
             memory               6676.81k
             OS Memory            32688.00k
             Timestamp            03/08/2024 04:36:30 AM
             Step Count                    35  Switch Count  2
             Page Faults                   0
             Page Reclaims                 313
             Page Swaps                    0
             Voluntary Context Switches    10
             Involuntary Context Switches  0
             Block Input Operations        0
             Block Output Operations       272


304
305
306        **************************************************************
307        *6.Plot the bar charts for bathrooms, bedrooms, usecode, totalrooms
308        respectively, and save the bar chart of bedrooms as bedrooms.png.
309        **************************************************************;
310
311        proc sgplot data=house_price;
312            title 'Bar Chart for Bathrooms';
313            vbar bathrooms;
314        run;

NOTE: PROCEDURE SGPLOT used (Total process time):
             real time            1.96 seconds
             user cpu time        0.05 seconds
             system cpu time      0.01 seconds
             memory               10034.50k
```

```
        OS Memory              34100.00k
        Timestamp              03/08/2024 04:36:32 AM
        Step Count                       36  Switch Count  3
        Page Faults                      0
        Page Reclaims                    3077
        Page Swaps                       0
        Voluntary Context Switches       301
        Involuntary Context Switches     0
        Block Input Operations           0
        Block Output Operations          536

 NOTE: There were 11330 observations read from the data set WORK.HOUSE_PRICE.

 315
 316        proc sgplot data=house_price;
 317            title 'Bar Chart for Bedrooms';
 318            vbar bedrooms;
 319        run;

 NOTE: PROCEDURE SGPLOT used (Total process time):
        real time              0.10 seconds
        user cpu time          0.03 seconds
        system cpu time        0.00 seconds
        memory                 3404.17k
        OS Memory              34484.00k
        Timestamp              03/08/2024 04:36:32 AM
        Step Count                       37  Switch Count  3
        Page Faults                      0
        Page Reclaims                    745
        Page Swaps                       0
        Voluntary Context Switches       143
        Involuntary Context Switches     0
        Block Input Operations           0
        Block Output Operations          360

 NOTE: There were 11330 observations read from the data set WORK.HOUSE_PRICE.

 320
 321        proc sgplot data=house_price;
 322            title 'Bar Chart for Use Code';
 323            vbar usecode;
 324        run;

 NOTE: PROCEDURE SGPLOT used (Total process time):
        real time              0.11 seconds
        user cpu time          0.03 seconds
        system cpu time        0.01 seconds
        memory                 3511.64k
        OS Memory              34224.00k
        Timestamp              03/08/2024 04:36:32 AM
        Step Count                       38  Switch Count  3
        Page Faults                      0
        Page Reclaims                    630
        Page Swaps                       0
        Voluntary Context Switches       175
        Involuntary Context Switches     0
        Block Input Operations           0
        Block Output Operations          384

 NOTE: There were 11330 observations read from the data set WORK.HOUSE_PRICE.

 325
 326        proc sgplot data=house_price;
 327            title 'Bar Chart for Total Rooms';
 328            vbar totalrooms;
 329        run;

 NOTE: PROCEDURE SGPLOT used (Total process time):
        real time              0.11 seconds
        user cpu time          0.04 seconds
        system cpu time        0.00 seconds
        memory                 3597.39k
        OS Memory              34484.00k
        Timestamp              03/08/2024 04:36:32 AM
        Step Count                       39  Switch Count  3
        Page Faults                      0
        Page Reclaims                    681
        Page Swaps                       0
        Voluntary Context Switches       143
        Involuntary Context Switches     0
        Block Input Operations           0
```

```
        Block Output Operations            384

NOTE: There were 11330 observations read from the data set WORK.HOUSE_PRICE.

330
331
332
333        ************************************************************
334        *7.Plot the Histogram, boxplot for lastsoldprice, zestimate respectively.
335         Are they normal or skewed? What's the median of the lastsoldprice?
336         What's the median of the zestimate?
337        ************************************************************;
338
339        proc univariate data=house_price noprint;
340            var lastsoldprice;
341            histogram lastsoldprice / normal;
342            inset median / position=ne;
343        run;

NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time             0.32 seconds
      user cpu time         0.07 seconds
      system cpu time       0.01 seconds
      memory                8551.62k
      OS Memory             39432.00k
      Timestamp             03/08/2024 04:36:32 AM
      Step Count                      40  Switch Count  0
      Page Faults                     0
      Page Reclaims                   2255
      Page Swaps                      0
      Voluntary Context Switches      794
      Involuntary Context Switches    1
      Block Input Operations          0
      Block Output Operations         392


344
345
346
347        proc sgplot data=house_price;
348            vbox lastsoldprice;
349            title 'Box Plot of Last Sold Price';
350        run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time             0.21 seconds
      user cpu time         0.05 seconds
      system cpu time       0.02 seconds
      memory                3459.81k
      OS Memory             35124.00k
      Timestamp             03/08/2024 04:36:33 AM
      Step Count                      41  Switch Count  6
      Page Faults                     0
      Page Reclaims                   667
      Page Swaps                      0
      Voluntary Context Switches      2490
      Involuntary Context Switches    1
      Block Input Operations          0
      Block Output Operations         1352

NOTE: There were 11330 observations read from the data set WORK.HOUSE_PRICE.

351
352
353
354        OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
364
```