

```

1      OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
NOTE: ODS statements in the SAS Studio environment may disable some output features.
69
70      *****
71      ***** PROJECT 8 *****
72      *****;
73
74      *Part 1*;
75      ***Import the dataset***;
76      PROC IMPORT DATAFILE="/home/u63559709/Week 8/Bay Area House Price.csv"
77          OUT=house_price
78          DBMS=CSV
79          REPLACE;
80      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.

```

81      /*****
82      *   PRODUCT:   SAS
83      *   VERSION:   9.4
84      *   CREATOR:   External File Interface
85      *   DATE:      20OCT23
86      *   DESC:      Generated SAS Dastep Code
87      *   TEMPLATE SOURCE: (None Specified.)
88      *****/
89      data WORK.HOUSE_PRICE ;
90      %let _EFIERR_ = 0; /* set the ERROR detection macro variable */
91      infile '/home/u63559709/Week 8/Bay Area House Price.csv' delimiter = ',' MISSOVER DSD lrecl=32767 firstobs=2 ;
92          informat address $34. ;
93          informat info $98. ;
94          informat z_address $25. ;
95          informat bathrooms best32. ;
96          informat bedrooms best32. ;
97          informat finishedsqft best32. ;
98          informat lastsolddate mmddyy10. ;
99          informat lastsoldprice best32. ;
100         informat latitude best32. ;
101         informat longitude best32. ;
102         informat neighborhood $40. ;
103         informat totalrooms best32. ;
104         informat usecode $12. ;
105         informat yearbuilt best32. ;
106         informat zestimate best32. ;
107         informat zipcode best32. ;
108         informat zpid best32. ;
109         format address $34. ;
110         format info $98. ;
111         format z_address $25. ;
112         format bathrooms best12. ;
113         format bedrooms best12. ;
114         format finishedsqft best12. ;
115         format lastsolddate mmddyy10. ;
116         format lastsoldprice best12. ;
117         format latitude best12. ;
118         format longitude best12. ;
119         format neighborhood $40. ;
120         format totalrooms best12. ;
121         format usecode $12. ;
122         format yearbuilt best12. ;
123         format zestimate best12. ;
124         format zipcode best12. ;
125         format zpid best12. ;
126     input
127         address $
128         info $
129         z_address $
130         bathrooms
131         bedrooms
132         finishedsqft
133         lastsolddate
134         lastsoldprice
135         latitude
136         longitude
137         neighborhood $
138         totalrooms
139         usecode $
140         yearbuilt
141         zestimate
142         zipcode
143         zpid

```

```

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

```

NOTE: The infile '/home/u63559709/Week 8/Bay Area House Price.csv' is:
 Filename=/home/u63559709/Week 8/Bay Area House Price.csv,
 Owner Name=u63559709,Group Name=oda,
 Access Permission=-rw-r--r--,
 Last Modified=18Oct2023:21:01:27,
 File Size (bytes)=2842323

NOTE: 11330 records were read from the infile '/home/u63559709/Week 8/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	10690.37k
OS Memory	36128.00k
Timestamp	10/20/2023 01:31:32 PM
Step Count	80 Switch Count 2
Page Faults	0
Page Reclaims	328
Page Swaps	0
Voluntary Context Switches	11
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	7176

11330 rows created in WORK.HOUSE_PRICE from /home/u63559709/Week 8/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	0.11 seconds
user cpu time	0.09 seconds
system cpu time	0.01 seconds
memory	10690.37k
OS Memory	36644.00k
Timestamp	10/20/2023 01:31:32 PM
Step Count	80 Switch Count 10
Page Faults	0
Page Reclaims	5252
Page Swaps	0
Voluntary Context Switches	80
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	7248

```

147
148      *Part 2*;
149      ***Drop selected variables from the table***;
150      PROC CONTENTS DATA=house_price;
151      RUN;

```

NOTE: PROCEDURE CONTENTS used (Total process time):

real time	0.03 seconds
user cpu time	0.03 seconds
system cpu time	0.00 seconds
memory	3099.75k
OS Memory	31152.00k
Timestamp	10/20/2023 01:31:32 PM
Step Count	81 Switch Count 0
Page Faults	0
Page Reclaims	285
Page Swaps	0
Voluntary Context Switches	0
Involuntary Context Switches	1
Block Input Operations	0
Block Output Operations	24

```

152
153      DATA house_price;
154          SET house_price;
155          DROP address info z_address neighborhood latitude longitude zpid;

```

156 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	0.01 seconds
system cpu time	0.00 seconds
memory	3277.62k
OS Memory	32432.00k
Timestamp	10/20/2023 01:31:32 PM
Step Count	82 Switch Count 2
Page Faults	0
Page Reclaims	475
Page Swaps	0
Voluntary Context Switches	13
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	2056

157

158 PROC SQL;

159 CREATE TABLE house_price AS

160 SELECT bathrooms, bedrooms, finishedsqft, lastsoldprice, lastsolddate,

161 totalrooms, usecode, yearbuilt, zestimate, zipcode

162 FROM house_price;

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.

163 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	6519.34k
OS Memory	36016.00k
Timestamp	10/20/2023 01:31:32 PM
Step Count	83 Switch Count 2
Page Faults	0
Page Reclaims	367
Page Swaps	0
Voluntary Context Switches	18
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	2056

164

165 PROC CONTENTS DATA=house_price;

166 RUN;

NOTE: PROCEDURE CONTENTS used (Total process time):

real time	0.02 seconds
user cpu time	0.03 seconds
system cpu time	0.00 seconds
memory	1898.59k
OS Memory	31152.00k
Timestamp	10/20/2023 01:31:32 PM
Step Count	84 Switch Count 0
Page Faults	0
Page Reclaims	287
Page Swaps	0
Voluntary Context Switches	0
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	16

167

168 *Part 3*;

169 ***Create new variable price_per_square_foot***;

170 DATA house_price;

171 SET house_price;

172 price_per_square_foot = lastsoldprice / finishedsqft;

173 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    0.01 seconds
memory             2981.96k
OS Memory          31920.00k
Timestamp          10/20/2023 01:31:32 PM
Step Count         85  Switch Count  2
Page Faults        0
Page Reclaims      450
Page Swaps         0
Voluntary Context Switches 13
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 2312

```

```

174
175      PROC SQL;
176          CREATE TABLE house_price AS
177          SELECT *,
178              lastsoldprice / finishedsqft AS price_per_square_foot
179          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.

```

```

180      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             6665.37k
OS Memory          36016.00k
Timestamp          10/20/2023 01:31:32 PM
Step Count         86  Switch Count  2
Page Faults        0
Page Reclaims      444
Page Swaps         0
Voluntary Context Switches 15
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 2312

```

```

181
182      *Part 4*;
183      ***Find the average of lastsoldprice by zipcode***;
184      PROC SORT DATA=house_price;
185          BY zipcode;
186      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    0.00 seconds
memory             4191.50k
OS Memory          33216.00k
Timestamp          10/20/2023 01:31:32 PM
Step Count         87  Switch Count  2
Page Faults        0
Page Reclaims      653
Page Swaps         0
Voluntary Context Switches 12
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 2312

```

```

187
188      PROC MEANS DATA=house_price NOPRINT;
189          BY zipcode;
190          VAR lastsoldprice;
191          OUTPUT OUT=avg_price_by_zipcode MEAN=AvgLastSoldPrice;
192      RUN;

```

```

NOTE: There were 11330 observations read from the data set WORK.HOUSE_PRICE.
NOTE: The data set WORK.AVG_PRICE_BY_ZIPCODE has 25 observations and 4 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            3070.62k
OS Memory         32180.00k
Timestamp         10/20/2023 01:31:32 PM
Step Count        88  Switch Count  5
Page Faults       0
Page Reclaims     442
Page Swaps        0
Voluntary Context Switches 20
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 264

```

```

193
194      PROC SQL;
195          CREATE TABLE avg_price_by_zipcode AS
196          SELECT zipcode,
197             AVG(lastsoldprice) AS AvgLastSoldPrice
198          FROM house_price
199          GROUP BY zipcode;
NOTE: Table WORK.AVG_PRICE_BY_ZIPCODE created, with 25 rows and 2 columns.

```

```

200      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           6668.93k
OS Memory         36016.00k
Timestamp         10/20/2023 01:31:32 PM
Step Count        89  Switch Count  2
Page Faults       0
Page Reclaims     318
Page Swaps        0
Voluntary Context Switches 12
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 264

```

```

201
202      *Part 5*;
203      ***Find the average of lastsoldprice by usecode, totalrooms, and bedrooms***;
204      PROC SORT DATA=house_price;
205          BY usecode totalrooms bedrooms;
206      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   0.00 seconds
memory           4198.21k
OS Memory         33216.00k
Timestamp         10/20/2023 01:31:32 PM
Step Count        90  Switch Count  2
Page Faults       0
Page Reclaims     719
Page Swaps        0
Voluntary Context Switches 10
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 2312

```

```

207
208      PROC MEANS DATA=house_price NOPRINT;
209          BY usecode totalrooms bedrooms;
210          VAR lastsoldprice;
211          OUTPUT OUT=avg_price_by_group MEAN=AvgLastSoldPrice;
212      RUN;

```

NOTE: There were 11330 observations read from the data set WORK.HOUSE_PRICE.
 NOTE: The data set WORK.AVG_PRICE_BY_GROUP has 388 observations and 6 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           3082.81k
OS Memory         32180.00k

```

```

Timestamp          10/20/2023 01:31:32 PM
Step Count          91  Switch Count  5
Page Faults         0
Page Reclaims       438
Page Swaps          0
Voluntary Context Switches 17
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 264

```

```

213
214      PROC SQL;
215          CREATE TABLE avg_price_by_group AS
216          SELECT usecode, totalrooms, bedrooms,
217                 AVG(lastsoldprice) AS AvgLastSoldPrice
218          FROM house_price
219          GROUP BY usecode, totalrooms, bedrooms;
NOTE: Table WORK.AVG_PRICE_BY_GROUP created, with 388 rows and 4 columns.

```

```

220      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             6683.06k
OS Memory          36016.00k
Timestamp          10/20/2023 01:31:32 PM
Step Count          92  Switch Count  2
Page Faults         0
Page Reclaims       314
Page Swaps          0
Voluntary Context Switches 11
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 264

```

```

221
222      *Part 6*;
223      ***Plot the bar charts for bathrooms, bedrooms, usecode, totalrooms***;
224      ods listing gpath='/home/u63559709/Week 8';
225      ods graphics / outputfmt=png imagename='bedrooms';
226
227      PROC SGPLOT DATA=house_price;
228          VBAR bathrooms;
229          TITLE "Bar Chart for Bathrooms";
230      RUN;

```

```

NOTE: PROCEDURE SGPLOT used (Total process time):
real time          0.15 seconds
user cpu time       0.06 seconds
system cpu time     0.01 seconds
memory             15337.62k
OS Memory          42292.00k
Timestamp          10/20/2023 01:31:32 PM
Step Count          93  Switch Count  3
Page Faults         0
Page Reclaims       3814
Page Swaps          0
Voluntary Context Switches 216
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 712

```

NOTE: Listing image output written to /home/u63559709/Week 8/bedrooms2.png.
NOTE: There were 11330 observations read from the data set WORK.HOUSE_PRICE.

```

231
232      PROC SGPLOT DATA=house_price;
233          VBAR bedrooms;
234          TITLE "Bar Chart for Bedrooms";
235      RUN;

```

```

NOTE: PROCEDURE SGPLOT used (Total process time):
real time          0.10 seconds
user cpu time       0.04 seconds
system cpu time     0.01 seconds
memory             3524.00k
OS Memory          42420.00k
Timestamp          10/20/2023 01:31:32 PM
Step Count          94  Switch Count  3

```

```

Page Faults          0
Page Reclaims        1042
Page Swaps            0
Voluntary Context Switches 203
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 432

```

NOTE: Listing image output written to /home/u63559709/Week 8/bedrooms3.png.
 NOTE: There were 11330 observations read from the data set WORK.HOUSE_PRICE.

```

236
237     ods graphics off;
238
239     PROC SGPLOT DATA=house_price;
240         VBAR usecode;
241         TITLE "Bar Chart for Usecode";
242     RUN;

```

NOTE: PROCEDURE SGPLOT used (Total process time):

```

real time          0.10 seconds
user cpu time      0.05 seconds
system cpu time    0.01 seconds
memory             3609.50k
OS Memory          42416.00k
Timestamp          10/20/2023 01:31:32 PM
Step Count                     95  Switch Count  3
Page Faults                    0
Page Reclaims                  873
Page Swaps                      0
Voluntary Context Switches     234
Involuntary Context Switches   0
Block Input Operations         0
Block Output Operations        456

```

NOTE: Listing image output written to /home/u63559709/Week 8/SGPlot6.png.
 NOTE: There were 11330 observations read from the data set WORK.HOUSE_PRICE.

```

243
244     PROC SGPLOT DATA=house_price;
245         VBAR totalrooms;
246         TITLE "Bar Chart for Total Rooms";
247     RUN;

```

NOTE: PROCEDURE SGPLOT used (Total process time):

```

real time          0.10 seconds
user cpu time      0.04 seconds
system cpu time    0.00 seconds
memory             3728.37k
OS Memory          42932.00k
Timestamp          10/20/2023 01:31:32 PM
Step Count                     96  Switch Count  3
Page Faults                    0
Page Reclaims                  927
Page Swaps                      0
Voluntary Context Switches     206
Involuntary Context Switches   0
Block Input Operations         0
Block Output Operations        448

```

NOTE: Listing image output written to /home/u63559709/Week 8/SGPlot7.png.
 NOTE: There were 11330 observations read from the data set WORK.HOUSE_PRICE.

```

248
249
250     *Part 7*;
251     ***Plot the Histogram, boxplot for lastsoldprice, zestimate respectively***;
252     PROC MEANS DATA=house_price NOPRINT;
253         VAR lastsoldprice zestimate;
254         OUTPUT OUT=stats
255             MEDIAN=Median_LastsoldPrice Median_Zestimate
256             SKEWNESS=Skewness_LastsoldPrice Skewness_Zestimate;
257     RUN;

```

NOTE: There were 11330 observations read from the data set WORK.HOUSE_PRICE.
 NOTE: The data set WORK.STATS has 1 observations and 6 variables.
 NOTE: PROCEDURE MEANS used (Total process time):

```

real time          0.01 seconds
user cpu time      0.01 seconds
system cpu time    0.01 seconds
memory             8987.53k
OS Memory          48596.00k

```

```

Timestamp          10/20/2023 01:31:32 PM
Step Count          97  Switch Count  3
Page Faults         0
Page Reclaims       1948
Page Swaps          0
Voluntary Context Switches  34
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations 264

```

```

258
259      PROC SGPLOT DATA=house_price;
260          HISTOGRAM lastsoldprice;
261          TITLE "Histogram for Last Sold Price";
262      RUN;

```

```

NOTE: PROCEDURE SGPLOT used (Total process time):
real time          0.11 seconds
user cpu time       0.05 seconds
system cpu time     0.01 seconds
memory              3651.56k
OS Memory           43060.00k
Timestamp           10/20/2023 01:31:32 PM
Step Count          98  Switch Count  6
Page Faults         0
Page Reclaims       852
Page Swaps          0
Voluntary Context Switches  228
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations 1392

```

NOTE: Listing image output written to /home/u63559709/Week 8/SGPlot8.png.
 NOTE: There were 11330 observations read from the data set WORK.HOUSE_PRICE.

```

263
264      PROC SGPLOT DATA=house_price;
265          VBOX lastsoldprice;
266          TITLE "Boxplot for Last Sold Price";
267      RUN;

```

```

NOTE: PROCEDURE SGPLOT used (Total process time):
real time          0.19 seconds
user cpu time       0.08 seconds
system cpu time     0.02 seconds
memory              3720.50k
OS Memory           43060.00k
Timestamp           10/20/2023 01:31:32 PM
Step Count          99  Switch Count  6
Page Faults         0
Page Reclaims       840
Page Swaps          0
Voluntary Context Switches  2553
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations 1408

```

NOTE: Listing image output written to /home/u63559709/Week 8/SGPlot9.png.
 NOTE: There were 11330 observations read from the data set WORK.HOUSE_PRICE.

```

268
269      PROC SGPLOT DATA=house_price;
270          HISTOGRAM zestimate;
271          TITLE "Histogram for Zestimate";
272      RUN;

```

```

NOTE: PROCEDURE SGPLOT used (Total process time):
real time          0.11 seconds
user cpu time       0.06 seconds
system cpu time     0.00 seconds
memory              3690.81k
OS Memory           43060.00k
Timestamp           10/20/2023 01:31:33 PM
Step Count          100 Switch Count  6
Page Faults         0
Page Reclaims       831
Page Swaps          0
Voluntary Context Switches  222
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations 1448

```


NOTE: Listing image output written to /home/u63559709/Week 8/SGPlot10.png.
 NOTE: There were 11330 observations read from the data set WORK.HOUSE_PRICE.

```

273
274      PROC SGPLOT DATA=house_price;
275          VBOX zestimate;
276          TITLE "Boxplot for Zestimate";
277      RUN;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.17 seconds
      user cpu time       0.07 seconds
      system cpu time     0.02 seconds
      memory              3764.37k
      OS Memory           43060.00k
      Timestamp           10/20/2023 01:31:33 PM
      Step Count          101  Switch Count   6
      Page Faults         0
      Page Reclaims       822
      Page Swaps          0
      Voluntary Context Switches 2990
      Involuntary Context Switches 0
      Block Input Operations 0
      Block Output Operations 1472

```

NOTE: Listing image output written to /home/u63559709/Week 8/SGPlot11.png.
 NOTE: There were 11330 observations read from the data set WORK.HOUSE_PRICE.

```

278
279      *The median for lastsoldprice is 990000. The median for zestimate is 1230758.
280      They are skewed.*;
281
282      *Part 8*;
283      ***Compare the average of zestimate for any two different zipcodes***;
284      PROC SQL;
285          CREATE TABLE unique_zip AS
286          SELECT DISTINCT zipcode
287          FROM house_price
288          ORDER BY zipcode;
NOTE: Table WORK.UNIQUE_ZIP created, with 25 rows and 1 columns.

```

```

289      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              6658.84k
      OS Memory           47280.00k
      Timestamp           10/20/2023 01:31:33 PM
      Step Count          102  Switch Count   2
      Page Faults         0
      Page Reclaims       369
      Page Swaps          0
      Voluntary Context Switches 11
      Involuntary Context Switches 0
      Block Input Operations 0
      Block Output Operations 272

```

```

290
291      PROC PRINT DATA=unique_zip;
292      RUN;

```

NOTE: There were 25 observations read from the data set WORK.UNIQUE_ZIP.
 NOTE: The PROCEDURE PRINT printed page 17.
 NOTE: PROCEDURE PRINT used (Total process time):

```

      real time           0.01 seconds
      user cpu time       0.01 seconds
      system cpu time     0.00 seconds
      memory              963.34k
      OS Memory           41388.00k
      Timestamp           10/20/2023 01:31:33 PM
      Step Count          103  Switch Count   0
      Page Faults         0
      Page Reclaims       64
      Page Swaps          0
      Voluntary Context Switches 0
      Involuntary Context Switches 0
      Block Input Operations 0
      Block Output Operations 0

```

```

293
294     DATA filtered_data;
295         SET house_price;
296         IF zipcode IN (94110, 94131);
297     RUN;

```

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

NOTE: The data set WORK.FILTERED_DATA has 1622 observations and 11 variables.

NOTE: DATA statement used (Total process time):

```

real time          0.00 seconds
user cpu time      0.01 seconds
system cpu time    0.00 seconds
memory            2118.84k
OS Memory          42416.00k
Timestamp          10/20/2023 01:31:33 PM
Step Count         104  Switch Count  2
Page Faults        0
Page Reclaims      290
Page Swaps         0
Voluntary Context Switches  16
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  520

```

```

298
299     PROC TTEST DATA=filtered_data;
300         CLASS zipcode;
301         VAR zestimate;
302         TITLE "Average Zestimate between Zip Codes 94110 and 94131";
303     RUN;

```

NOTE: The PROCEDURE TTEST printed page 18.

NOTE: PROCEDURE TTEST used (Total process time):

```

real time          0.03 seconds
user cpu time      0.03 seconds
system cpu time    0.00 seconds
memory            1748.92k
OS Memory          41388.00k
Timestamp          10/20/2023 01:31:33 PM
Step Count         105  Switch Count  1
Page Faults        0
Page Reclaims      52
Page Swaps         0
Voluntary Context Switches  7
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  16

```

```

304
305     *We reject the null hypothesis that there is no significant difference between the two
306     zipcodes in regard to zestimate. The p value is <.0001 for two methods.
307
308     *Part 9*;
309     ***Do you agree that there is no difference between the average
310     zestimate and the average of lastsoldprice statistically?***;
311     PROC TTEST DATA=house_price;
312         PAIRED zestimate*lastsoldprice;
313         TITLE "Average Zestimate vs. Last Sold Price";
314     RUN;

```

NOTE: The PROCEDURE TTEST printed page 19.

NOTE: PROCEDURE TTEST used (Total process time):

```

real time          0.01 seconds
user cpu time      0.02 seconds
system cpu time    0.01 seconds
memory            2181.60k
OS Memory          42156.00k
Timestamp          10/20/2023 01:31:33 PM
Step Count         106  Switch Count  1
Page Faults        0
Page Reclaims      244
Page Swaps         0
Voluntary Context Switches  9
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  0

```

315

```

316      *We do not agree the average zestimate and the average lastsoldprice
317      have no difference statistically. The p value is <.0001 with a large F-value.
318
319      *Part 10*;
320      ***Do you agree that the number of bedrooms is associated with the usecode?***;
321      PROC GLM DATA=house_price;
322          CLASS usecode;
323          MODEL bedrooms = usecode;
324          TITLE 'Simple Linear Regression of Bedrooms and Usecode';
325      RUN;

326
327      *We disagree the number of bedrooms is associated with usecode. The p-value is
328      under .0001 with a large F-value.*;
329
330      *Part 11*;
331      ***Do you agree that the number of bedrooms is associated with the number of bathrooms?***;
332      ods graphics off;

```

NOTE: The PROCEDURE GLM printed pages 20-21.

NOTE: PROCEDURE GLM used (Total process time):

```

real time      0.03 seconds
user cpu time   0.03 seconds
system cpu time 0.00 seconds
memory         3264.37k
OS Memory      43196.00k
Timestamp      10/20/2023 01:31:33 PM
Step Count     107  Switch Count  4
Page Faults    0
Page Reclaims  413
Page Swaps     0
Voluntary Context Switches  34
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  304

```

```

333      PROC REG DATA=house_price;
334          MODEL bedrooms=bathrooms;
335          title 'Simple Linear Regression';
336      RUN;

```

```

337
338      *We agree, the parameter estimate for "bathrooms" is 0.88620, and the associated p-value
339      is <.0001. This suggests that for each additional bathroom, the number of bedrooms
340      increases by approximately 0.88620 on average.*;
341
342      *Part 12*;
343      ***Calculate the correlation coefficients of all numerical variables with
344      the variable zesitmate, and plot the scatter plot and matrix.***;

```

NOTE: The PROCEDURE REG printed page 22.

NOTE: PROCEDURE REG used (Total process time):

```

real time      0.02 seconds
user cpu time   0.02 seconds
system cpu time 0.00 seconds
memory         3626.87k
OS Memory      43716.00k
Timestamp      10/20/2023 01:31:33 PM
Step Count     108  Switch Count  2
Page Faults    0
Page Reclaims  454
Page Swaps     0
Voluntary Context Switches  37
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  48

```

```

345      PROC CORR DATA=house_price PLOTS(MAXPOINTS=none)=scatter;
346          VAR zestimate bedrooms bathrooms finishedsqft lastsoldprice totalrooms yearbuilt zipcode;
347          WITH zestimate;
348          TITLE 'Correlation Coefficients and Scatter Plot Matrix';
349      RUN;

```

WARNING: You must enable ODS graphics before requesting plots.

NOTE: The PROCEDURE CORR printed page 23.

NOTE: PROCEDURE CORR used (Total process time):

```

real time      0.03 seconds
user cpu time   0.03 seconds
system cpu time 0.00 seconds
memory         2241.31k

```

```

OS Memory          42156.00k
Timestamp          10/20/2023 01:31:33 PM
Step Count         109  Switch Count  1
Page Faults        0
Page Reclaims      247
Page Swaps         0
Voluntary Context Switches  8
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  16

```

```

350
351      *Part 13*;
352      ***Find a regression model for zestimate with the first three most correlated variables***;
353      PROC CORR DATA=house_price NOPRINT OUTP=CorrOut;
354          VAR zestimate bedrooms bathrooms finishedsqft lastsoldprice totalrooms yearbuilt zipcode;
355      WITH zestimate;
356      RUN;

```

NOTE: The data set WORK.CORROUT has 4 observations and 10 variables.

NOTE: PROCEDURE CORR used (Total process time):

```

real time          0.00 seconds
user cpu time      0.01 seconds
system cpu time    0.00 seconds
memory            2070.78k
OS Memory          42416.00k
Timestamp          10/20/2023 01:31:33 PM
Step Count         110  Switch Count  2
Page Faults        0
Page Reclaims      313
Page Swaps         0
Voluntary Context Switches  11
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  264

```

```

357
358      PROC PRINT DATA=CorrOut;
359      RUN;

```

NOTE: There were 4 observations read from the data set WORK.CORROUT.

NOTE: The PROCEDURE PRINT printed page 24.

NOTE: PROCEDURE PRINT used (Total process time):

```

real time          0.01 seconds
user cpu time      0.01 seconds
system cpu time    0.00 seconds
memory            868.53k
OS Memory          41388.00k
Timestamp          10/20/2023 01:31:33 PM
Step Count         111  Switch Count  0
Page Faults        0
Page Reclaims      64
Page Swaps         0
Voluntary Context Switches  1
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  0

```

```

360
361      PROC REG DATA=house_price;
362          MODEL zestimate = bathrooms finishedsqft lastsoldprice;
363      RUN;

```

```

364
365      *Part 14*;
366      *Find a regression model for zestimate with the first five most correlated variables*;

```

NOTE: The PROCEDURE REG printed page 25.

NOTE: PROCEDURE REG used (Total process time):

```

real time          0.02 seconds
user cpu time      0.03 seconds
system cpu time    0.00 seconds
memory            3603.50k
OS Memory          43716.00k
Timestamp          10/20/2023 01:31:33 PM
Step Count         112  Switch Count  1
Page Faults        0
Page Reclaims      454
Page Swaps         0

```

```

Voluntary Context Switches      39
Involuntary Context Switches    0
Block Input Operations           0
Block Output Operations          64

```

```

367      PROC REG DATA=house_price;
368          MODEL zestimate = bathrooms finishedsqft lastsoldprice bedrooms yearbuilt;
369      RUN;

370
371      *Part 15*;
372      ***Compare the adjusted R^2 in the two models from question 13 and 14***;
373      *The first model's r-squared is .8320 and the second's is .8329. The second is better.*;
374
375      *Part 16*;
376      ***Use the better model from question 15 to predict the house prices given the values
377      of independent variables.*;

```

NOTE: The PROCEDURE REG printed page 26.

NOTE: PROCEDURE REG used (Total process time):

```

real time      0.02 seconds
user cpu time   0.03 seconds
system cpu time 0.00 seconds
memory         3648.68k
OS Memory      43716.00k
Timestamp      10/20/2023 01:31:33 PM
Step Count     113  Switch Count  1
Page Faults    0
Page Reclaims  453
Page Swaps     0
Voluntary Context Switches  65
Involuntary Context Switches 0
Block Input Operations      0
Block Output Operations     64

```

```

378      DATA new_houses;
379          INPUT bathrooms finishedsqft lastsoldprice bedrooms yearbuilt;
380          DATALINES;

```

NOTE: The data set WORK.NEW_HOUSES has 4 observations and 5 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         669.90k
OS Memory      41388.00k
Timestamp      10/20/2023 01:31:33 PM
Step Count     114  Switch Count  2
Page Faults    0
Page Reclaims  92
Page Swaps     0
Voluntary Context Switches  13
Involuntary Context Switches 0
Block Input Operations      0
Block Output Operations    264

```

```

385      ;
386      RUN;
387
388      PROC REG DATA=house_price OUTEST=estimates;
389          MODEL zestimate = bathrooms finishedsqft lastsoldprice bedrooms yearbuilt;
390      RUN;
391

```

NOTE: The data set WORK.ESTIMATES has 1 observations and 11 variables.

NOTE: The PROCEDURE REG printed page 27.

NOTE: PROCEDURE REG used (Total process time):

```

real time      0.02 seconds
user cpu time   0.03 seconds
system cpu time 0.00 seconds
memory         3917.46k
OS Memory      43976.00k
Timestamp      10/20/2023 01:31:33 PM
Step Count     115  Switch Count  3
Page Faults    0
Page Reclaims  521
Page Swaps     0
Voluntary Context Switches  65

```

```

Involuntary Context Switches      0
Block Input Operations             0
Block Output Operations            312

```

```

392      PROC SCORE DATA=new_houses SCORE=estimates OUT=PredictedValues TYPE=parms;
393          VAR bathrooms finishedsqft lastsoldprice bedrooms yearbuilt;
394      RUN;

```

NOTE: There were 4 observations read from the data set WORK.NEW_HOUSES.
 NOTE: There were 1 observations read from the data set WORK.ESTIMATES.
 NOTE: The data set WORK.PREDICTEDVALUES has 4 observations and 6 variables.
 NOTE: PROCEDURE SCORE used (Total process time):

```

real time      0.00 seconds
user cpu time   0.00 seconds
system cpu time 0.01 seconds
memory         1104.56k
OS Memory      41908.00k
Timestamp      10/20/2023 01:31:33 PM
Step Count     116  Switch Count  2
Page Faults    0
Page Reclaims  151
Page Swaps     0
Voluntary Context Switches  14
Involuntary Context Switches 0
Block Input Operations      0
Block Output Operations     264

```

```

395
396      PROC PRINT DATA=PredictedValues; RUN;

```

NOTE: There were 4 observations read from the data set WORK.PREDICTEDVALUES.
 NOTE: The PROCEDURE PRINT printed page 28.
 NOTE: PROCEDURE PRINT used (Total process time):

```

real time      0.00 seconds
user cpu time   0.01 seconds
system cpu time 0.00 seconds
memory         855.96k
OS Memory      41388.00k
Timestamp      10/20/2023 01:31:33 PM
Step Count     117  Switch Count  0
Page Faults    0
Page Reclaims  66
Page Swaps     0
Voluntary Context Switches  0
Involuntary Context Switches 0
Block Input Operations      0
Block Output Operations     16

```

```

397
398      *Part 17*;
399      ***Export the predictive values from question 16) as an excel file***;
400      PROC EXPORT DATA=PredictedValues
401          OUTFILE="/home/u63559709/Week 8/prediction.xlsx"
402          DBMS=xlsx
403          REPLACE;
404      RUN;

```

NOTE: The export data set has 4 observations and 6 variables.
 NOTE: "/home/u63559709/Week 8/prediction.xlsx" file was successfully created.
 NOTE: PROCEDURE EXPORT used (Total process time):

```

real time      0.01 seconds
user cpu time   0.00 seconds
system cpu time 0.00 seconds
memory         3413.96k
OS Memory      44084.00k
Timestamp      10/20/2023 01:31:33 PM
Step Count     118  Switch Count  0
Page Faults    0
Page Reclaims  649
Page Swaps     0
Voluntary Context Switches  23
Involuntary Context Switches 0
Block Input Operations      16
Block Output Operations     16

```

```

405
406      *Part 18*;
407      ***Create a macro named average with two parameters category and price.***;

```

```

408      %macro average(category, price);
409
410          PROC SORT DATA=house_price;
411              BY &category;
412          RUN;
413
414          PROC MEANS DATA=house_price NOPRINT;
415              BY &category;
416              VAR &price;
417              OUTPUT OUT=averageprice MEAN=;
418          RUN;
419
420          PROC PRINT DATA=averageprice;
421              TITLE "Average of &price by &category";
422          RUN;
423      %mend average;
424
425      %average(bathrooms, zestimate);

```

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	0.00 seconds
memory	4179.09k
OS Memory	44480.00k
Timestamp	10/20/2023 01:31:33 PM
Step Count	119 Switch Count 2
Page Faults	0
Page Reclaims	652
Page Swaps	0
Voluntary Context Switches	12
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	2312

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

NOTE: The data set WORK.AVERAGEPRICE has 40 observations and 4 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	3062.87k
OS Memory	43444.00k
Timestamp	10/20/2023 01:31:33 PM
Step Count	120 Switch Count 17
Page Faults	0
Page Reclaims	447
Page Swaps	0
Voluntary Context Switches	54
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	272

NOTE: There were 40 observations read from the data set WORK.AVERAGEPRICE.

NOTE: The PROCEDURE PRINT printed page 29.

NOTE: PROCEDURE PRINT used (Total process time):

real time	0.02 seconds
user cpu time	0.03 seconds
system cpu time	0.00 seconds
memory	742.65k
OS Memory	41388.00k
Timestamp	10/20/2023 01:31:33 PM
Step Count	121 Switch Count 1
Page Faults	0
Page Reclaims	65
Page Swaps	0
Voluntary Context Switches	6
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	16

```

426      %average.bedrooms, zestimate);

```

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    0.00 seconds
memory            4179.09k
OS Memory          44480.00k
Timestamp          10/20/2023 01:31:33 PM
Step Count         122  Switch Count  2
Page Faults        0
Page Reclaims      652
Page Swaps         0
Voluntary Context Switches 12
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 2312

```

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

NOTE: The data set WORK.AVERAGEPRICE has 17 observations and 4 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            3058.87k
OS Memory          43444.00k
Timestamp          10/20/2023 01:31:33 PM
Step Count         123  Switch Count  9
Page Faults        0
Page Reclaims      448
Page Swaps         0
Voluntary Context Switches 35
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 264

```

NOTE: There were 17 observations read from the data set WORK.AVERAGEPRICE.

NOTE: The PROCEDURE PRINT printed page 30.

NOTE: PROCEDURE PRINT used (Total process time):

```

real time          0.01 seconds
user cpu time      0.02 seconds
system cpu time    0.00 seconds
memory            706.96k
OS Memory          41388.00k
Timestamp          10/20/2023 01:31:33 PM
Step Count         124  Switch Count  1
Page Faults        0
Page Reclaims      64
Page Swaps         0
Voluntary Context Switches 8
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 0

```

```

427
428      *Part 19*;
429      ***Call the macro %average(category=zipcode, price=price_per_square_foot)***;
430      %average(category=zipcode, price=price_per_square_foot);

```

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            4179.25k
OS Memory          44480.00k
Timestamp          10/20/2023 01:31:33 PM
Step Count         125  Switch Count  2
Page Faults        0
Page Reclaims      653
Page Swaps         0
Voluntary Context Switches 13
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 2312

```

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

NOTE: The data set WORK.AVERAGEPRICE has 25 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	0.01 seconds
memory	3062.71k
OS Memory	43444.00k
Timestamp	10/20/2023 01:31:33 PM
Step Count	126 Switch Count 23
Page Faults	0
Page Reclaims	445
Page Swaps	0
Voluntary Context Switches	72
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	264

NOTE: There were 25 observations read from the data set WORK.AVERAGEPRICE.

NOTE: The PROCEDURE PRINT printed page 31.

NOTE: PROCEDURE PRINT used (Total process time):

real time	0.01 seconds
user cpu time	0.01 seconds
system cpu time	0.00 seconds
memory	742.21k
OS Memory	41388.00k
Timestamp	10/20/2023 01:31:33 PM
Step Count	127 Switch Count 1
Page Faults	0
Page Reclaims	65
Page Swaps	0
Voluntary Context Switches	7
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	16

431

432 *Part 20*;

433 ***Call the macro %average(category=totalrooms, price=zestimate)***;

434 %average(category=totalrooms, price=zestimate);

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	4179.09k
OS Memory	44480.00k
Timestamp	10/20/2023 01:31:33 PM
Step Count	128 Switch Count 2
Page Faults	0
Page Reclaims	652
Page Swaps	0
Voluntary Context Switches	12
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	2312

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

NOTE: The data set WORK.AVERAGEPRICE has 30 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	3058.87k
OS Memory	43444.00k
Timestamp	10/20/2023 01:31:33 PM
Step Count	129 Switch Count 14
Page Faults	0
Page Reclaims	443
Page Swaps	0
Voluntary Context Switches	53
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	264

NOTE: There were 30 observations read from the data set WORK.AVERAGEPRICE.

NOTE: The PROCEDURE PRINT printed page 32.

```
NOTE: PROCEDURE PRINT used (Total process time):  
  real time          0.02 seconds  
  user cpu time      0.02 seconds  
  system cpu time    0.00 seconds  
  memory             706.96k  
  OS Memory          41388.00k  
  Timestamp           10/20/2023 01:31:33 PM  
  Step Count          130  Switch Count  1  
  Page Faults         0  
  Page Reclaims       64  
  Page Swaps          0  
  Voluntary Context Switches  8  
  Involuntary Context Switches 0  
  Block Input Operations 0  
  Block Output Operations 0
```

```
435  
436      OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;  
446
```