

```

1      OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
68
69      /*Exploring data files*/
70
71      PROC CONTENTS DATA=MYDATA.Listings; RUN;

```

NOTE: PROCEDURE CONTENTS used (Total process time):

real time	0.07 seconds
user cpu time	0.08 seconds
system cpu time	0.00 seconds
memory	3091.90k
OS Memory	33456.00k
Timestamp	04/20/2022 03:58:54 AM
Step Count	64 Switch Count 0
Page Faults	0
Page Reclaims	295
Page Swaps	0
Voluntary Context Switches	6
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	32

```

72
73      PROC FREQ DATA=MYDATA.Listings;
74      TABLES neighbourhood_group_cleansed;
75      RUN;

```

NOTE: There were 49056 observations read from the data set MYDATA.LISTINGS.

NOTE: PROCEDURE FREQ used (Total process time):

real time	0.02 seconds
user cpu time	0.02 seconds
system cpu time	0.00 seconds
memory	2188.34k
OS Memory	33456.00k
Timestamp	04/20/2022 03:58:54 AM
Step Count	65 Switch Count 5
Page Faults	0
Page Reclaims	338
Page Swaps	0
Voluntary Context Switches	23
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	280

```

76
77      PROC FREQ DATA=MYDATA.LISTINGS ORDER=FREQ;
78      TABLES neighbourhood_group_cleansed;
79      RUN;

```

NOTE: There were 49056 observations read from the data set MYDATA.LISTINGS.

NOTE: PROCEDURE FREQ used (Total process time):

real time	0.02 seconds
user cpu time	0.02 seconds
system cpu time	0.01 seconds
memory	2178.65k
OS Memory	33456.00k
Timestamp	04/20/2022 03:58:54 AM
Step Count	66 Switch Count 5
Page Faults	0
Page Reclaims	328
Page Swaps	0
Voluntary Context Switches	24
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	264

```

80
81      PROC FREQ DATA=MYDATA.LISTINGS ORDER=FREQ;
82      TABLES neighbourhood_group_cleansed * room_type /
83      NOCOL NOROW NOPERCENT OUT=room_freq;
84      RUN;

```

NOTE: There were 49056 observations read from the data set MYDATA.LISTINGS.

NOTE: The data set WORK.ROOM_FREQ has 16 observations and 4 variables.

NOTE: PROCEDURE FREQ used (Total process time):

```

real time          0.04 seconds
user cpu time      0.04 seconds
system cpu time    0.01 seconds
memory            2961.03k
OS Memory          33976.00k
Timestamp          04/20/2022 03:58:54 AM
Step Count                67  Switch Count  9
Page Faults              0
Page Reclaims           442
Page Swaps              0
Voluntary Context Switches  50
Involuntary Context Switches 44
Block Input Operations    0
Block Output Operations   792

```

```

85
86      PROC UNIVARIATE DATA=MYDATA.LISTINGS;
87      VAR Price;
88      HISTOGRAM;
89
90      RUN;

```

NOTE: PROCEDURE UNIVARIATE used (Total process time):

```

real time          0.27 seconds
user cpu time      0.13 seconds
system cpu time    0.01 seconds
memory            14157.40k
OS Memory          44040.00k
Timestamp          04/20/2022 03:58:54 AM
Step Count                68  Switch Count  3
Page Faults              0
Page Reclaims          3203
Page Swaps              0
Voluntary Context Switches 147
Involuntary Context Switches 0
Block Input Operations    0
Block Output Operations   672

```

```

91
92      PROC MEANS DATA=MYDATA.LISTINGS;
93      VAR price accommodates bathrooms;
94      RUN;

```

NOTE: There were 49056 observations read from the data set MYDATA.LISTINGS.

NOTE: PROCEDURE MEANS used (Total process time):

```

real time          0.04 seconds
user cpu time      0.04 seconds
system cpu time    0.01 seconds
memory            7791.50k
OS Memory          47552.00k
Timestamp          04/20/2022 03:58:54 AM
Step Count                69  Switch Count  4
Page Faults              0
Page Reclaims           1557
Page Swaps              0
Voluntary Context Switches 39

```

```

Involuntary Context Switches      1
Block Input Operations             0
Block Output Operations            0

```

```

95
96      PROC MEANS DATA=MYDATA.LISTINGS N NMISS MIN MAX MEAN MEDIAN STDDEV;
97      CLASS room_type;
98      VAR price accommodates bathrooms;
99      OUTPUT OUT=list_stats;
100
101      RUN;

```

NOTE: There were 49056 observations read from the data set MYDATA.LISTINGS.

NOTE: The data set WORK.LIST_STATS has 20 observations and 7 variables.

NOTE: PROCEDURE MEANS used (Total process time):

```

real time          0.09 seconds
user cpu time      0.11 seconds
system cpu time    0.01 seconds
memory            9395.51k
OS Memory          48068.00k
Timestamp          04/20/2022 03:58:54 AM
Step Count                    70  Switch Count   6
Page Faults                  0
Page Reclaims               1709
Page Swaps                  0
Voluntary Context Switches   69
Involuntary Context Switches 29
Block Input Operations       0
Block Output Operations     264

```

```

102
103      PROC SUMMARY DATA=MYDATA.LISTINGS;
104      VAR price accommodates bathrooms;
105      OUTPUT OUT=list_sum;
106

```

NOTE: There were 49056 observations read from the data set MYDATA.LISTINGS.

NOTE: The data set WORK.LIST_SUM has 5 observations and 6 variables.

NOTE: PROCEDURE SUMMARY used (Total process time):

```

real time          0.03 seconds
user cpu time      0.01 seconds
system cpu time    0.01 seconds
memory            8000.84k
OS Memory          48068.00k
Timestamp          04/20/2022 03:58:54 AM
Step Count                    71  Switch Count   6
Page Faults                  0
Page Reclaims               1601
Page Swaps                  0
Voluntary Context Switches   43
Involuntary Context Switches 0
Block Input Operations       0
Block Output Operations     264

```

```

107      PROC SORT DATA=MYDATA.LISTINGS OUT=list_sort;
108      BY host_id;
109      RUN;

```

NOTE: There were 49056 observations read from the data set MYDATA.LISTINGS.

NOTE: The data set WORK.LIST_SORT has 49056 observations and 54 variables.

NOTE: PROCEDURE SORT used (Total process time):

```

real time          0.06 seconds
user cpu time      0.04 seconds
system cpu time    0.03 seconds
memory            25803.34k
OS Memory          66184.00k

```

```

Timestamp          04/20/2022 03:58:54 AM
Step Count          72  Switch Count  5
Page Faults         0
Page Reclaims       5862
Page Swaps          0
Voluntary Context Switches 24
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 42512

```

```

110
111 PROC SUMMARY DATA=list_sort;
112 BY host_id;
113 VAR price accommodates bathrooms;
114 OUTPUT OUT=list_sum SUM=;
115 RUN;

```

NOTE: There were 49056 observations read from the data set WORK.LIST_SORT.

NOTE: The data set WORK.LIST_SUM has 37511 observations and 6 variables.

NOTE: PROCEDURE SUMMARY used (Total process time):

```

real time          0.08 seconds
user cpu time      0.06 seconds
system cpu time    0.03 seconds
memory             4464.71k
OS Memory          45492.00k
Timestamp          04/20/2022 03:58:54 AM
Step Count          73  Switch Count  42
Page Faults         0
Page Reclaims       618
Page Swaps          0
Voluntary Context Switches 115
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 3592

```

```

116
117 DATA WORK.TEST;
118 SET MYDATA.Listings;
119 WHERE accommodates le 4;
120 RUN;

```

NOTE: There were 42565 observations read from the data set MYDATA.LISTINGS.

WHERE accommodates<=4;

NOTE: The data set WORK.TEST has 42565 observations and 54 variables.

NOTE: DATA statement used (Total process time):

```

real time          0.03 seconds
user cpu time      0.01 seconds
system cpu time    0.02 seconds
memory             3557.31k
OS Memory          44464.00k
Timestamp          04/20/2022 03:58:54 AM
Step Count          74  Switch Count  5
Page Faults         0
Page Reclaims       502
Page Swaps          0
Voluntary Context Switches 18
Involuntary Context Switches 1
Block Input Operations 0
Block Output Operations 36872

```

```

121
122 DATA WORK.TEST;
123 SET MYDATA.LISTINGS;
124 IF accommodates le 4;
125 RUN;

```

NOTE: There were 49056 observations read from the data set MYDATA.LISTINGS.

NOTE: The data set WORK.TEST has 42565 observations and 54 variables.

NOTE: DATA statement used (Total process time):

real time	0.03 seconds
user cpu time	0.01 seconds
system cpu time	0.03 seconds
memory	3699.68k
OS Memory	44464.00k
Timestamp	04/20/2022 03:58:54 AM
Step Count	75
Switch Count	5
Page Faults	0
Page Reclaims	494
Page Swaps	0
Voluntary Context Switches	21
Involuntary Context Switches	1
Block Input Operations	0
Block Output Operations	36872

```
126
127      /* Split data into TRAIN and TEST datasets at an 80/20 split */
128      PROC SURVEYSELECT DATA=Mydata.listings SAMPRATE=0.20 SEED=42
129      OUT=Full OUTALL METHOD=SRS;
130      RUN;
```

NOTE: The data set WORK.FULL has 49056 observations and 55 variables.

NOTE: PROCEDURE SURVEYSELECT used (Total process time):

real time	0.04 seconds
user cpu time	0.02 seconds
system cpu time	0.03 seconds
memory	3617.62k
OS Memory	44464.00k
Timestamp	04/20/2022 03:58:54 AM
Step Count	76
Switch Count	5
Page Faults	0
Page Reclaims	492
Page Swaps	0
Voluntary Context Switches	22
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	43288

```
131
132      DATA TRAIN TEST;
133      SET FULL;
134      IF SELECTED=0 THEN OUTPUT TRAIN; ELSE OUTPUT TEST;
135      DROP Selected;
136
137      OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
147
```