The CONTENTS Procedure

Data Set Name	WORK.HOUSE_PRICE	Observations	11330
Member Type	DATA	Variables	17
Engine	V9	Indexes	0
Created	10/20/2023 09:31:32	Observation Length	312
Last Modified	10/20/2023 09:31:32	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	28
First Data Page	1
Max Obs per Page	419
Obs in First Data Page	406
Number of Data Set Repairs	0
Filename	/saswork/SAS_work974100006B40_odaws01-usw2-2.oda.sas.com/SAS_workDA9900006B40_odaws01-usw2-2.oda.sas.com/house_price.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	2147495410
Access Permission	rw-r--r--
Owner Name	u63559709
File Size	4MB
File Size (bytes)	3801088

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Informat
1	address	Char	34	$34.	$34.
4	bathrooms	Num	8	BEST12.	BEST32.
5	bedrooms	Num	8	BEST12.	BEST32.
6	finishedsqft	Num	8	BEST12.	BEST32.
2	info	Char	98	$98.	$98.
7	lastsolddate	Num	8	MMDDYY10.	MMDDYY10.
8	lastsoldprice	Num	8	BEST12.	BEST32.
9	latitude	Num	8	BEST12.	BEST32.
10	longitude	Num	8	BEST12.	BEST32.
11	neighborhood	Char	40	$40.	$40.
12	totalrooms	Num	8	BEST12.	BEST32.
13	usecode	Char	12	$12.	$12.
14	yearbuilt	Num	8	BEST12.	BEST32.
3	z_address	Char	25	$25.	$25.
15	zestimate	Num	8	BEST12.	BEST32.
16	zipcode	Num	8	BEST12.	BEST32.

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat
17	zpid	Num	8	BEST12.	BEST32.

The CONTENTS Procedure

Data Set Name	WORK.HOUSE_PRICE	**Observations**	11330
Member Type	DATA	**Variables**	10
Engine	V9	**Indexes**	0
Created	10/20/2023 09:31:32	**Observation Length**	88
Last Modified	10/20/2023 09:31:32	**Deleted Observations**	0
Protection		**Compressed**	NO
Data Set Type		**Sorted**	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information

Data Set Page Size	131072
Number of Data Set Pages	8
First Data Page	1
Max Obs per Page	1486
Obs in First Data Page	1451
Number of Data Set Repairs	0
Filename	/saswork/SAS_work974100006B40_odaws01-usw2-2.oda.sas.com/SAS_workDA9900006B40_odaws01-usw2-2.oda.sas.com/house_price.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	2147495410
Access Permission	rw-r--r--
Owner Name	u63559709
File Size	1MB
File Size (bytes)	1179648

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat
1	bathrooms	Num	8	BEST12.	BEST32.
2	bedrooms	Num	8	BEST12.	BEST32.
3	finishedsqft	Num	8	BEST12.	BEST32.
5	lastsolddate	Num	8	MMDDYY10.	MMDDYY10.
4	lastsoldprice	Num	8	BEST12.	BEST32.
6	totalrooms	Num	8	BEST12.	BEST32.
7	usecode	Char	12	$12.	$12.
8	yearbuilt	Num	8	BEST12.	BEST32.
9	zestimate	Num	8	BEST12.	BEST32.
10	zipcode	Num	8	BEST12.	BEST32.

Bar Chart for Bathrooms

Bar Chart for Bedrooms

Bar Chart for Usecode

Bar Chart for Total Rooms

Histogram for Last Sold Price

Boxplot for Last Sold Price

Histogram for Zestimate

Boxplot for Zestimate
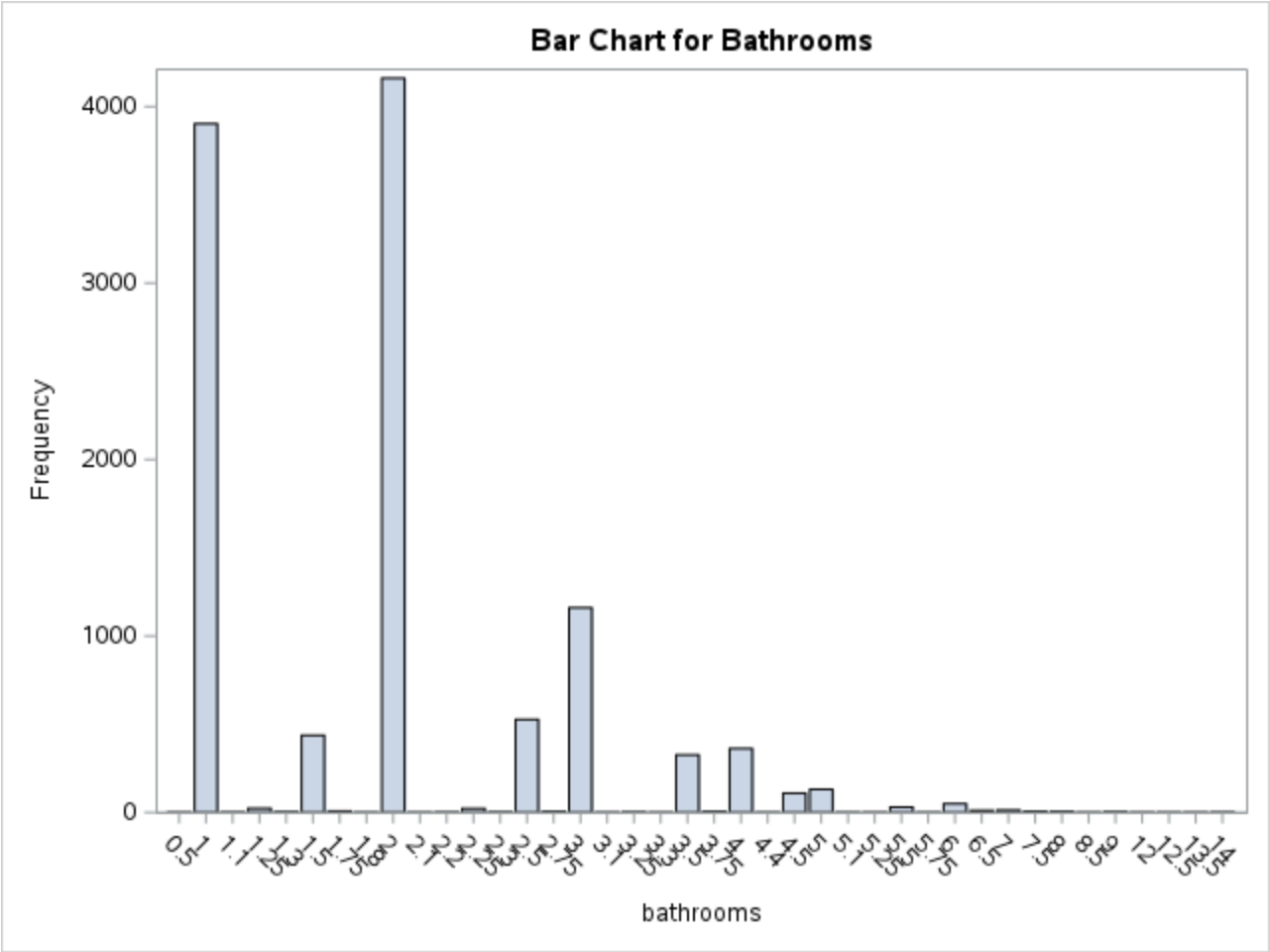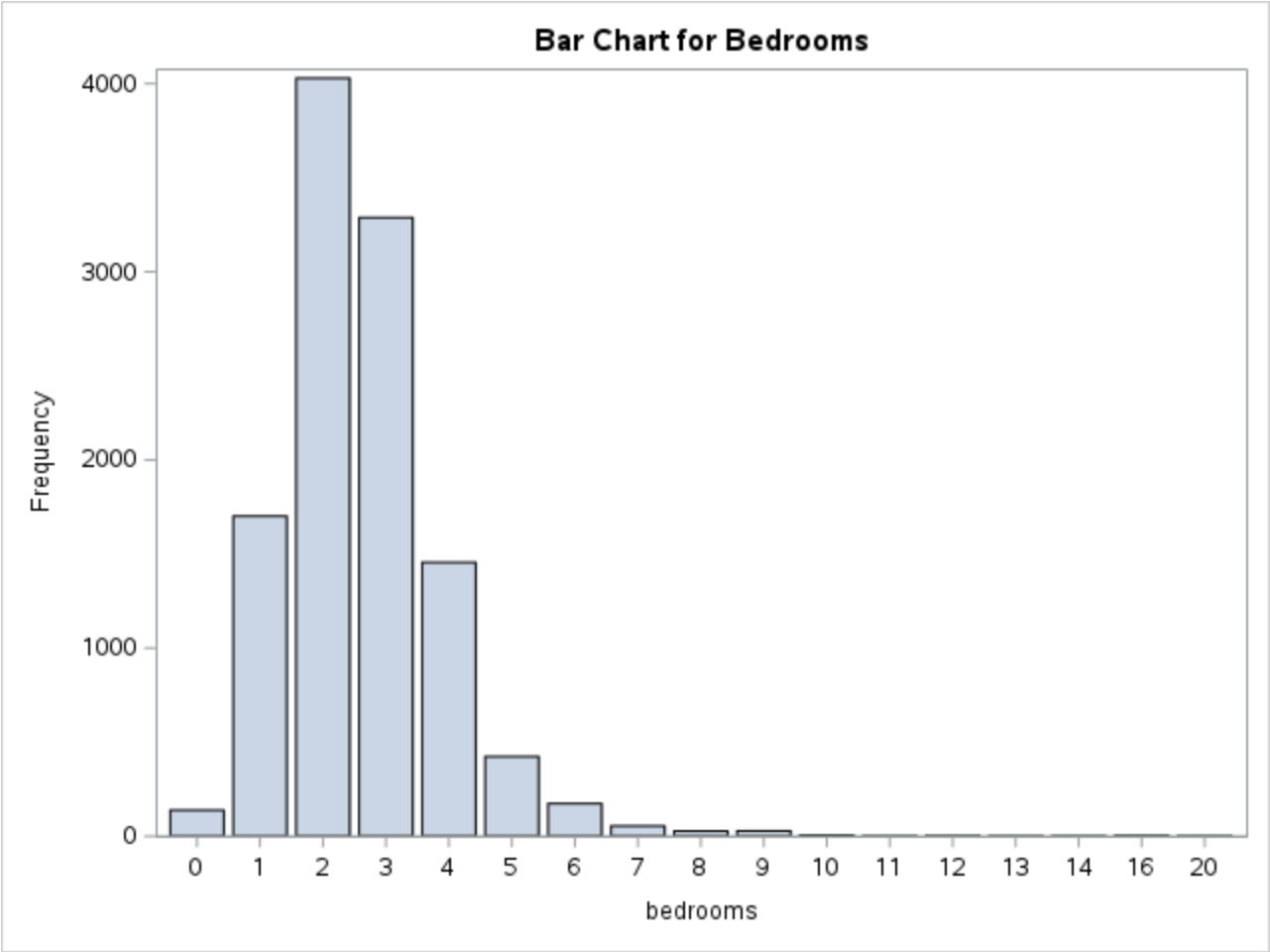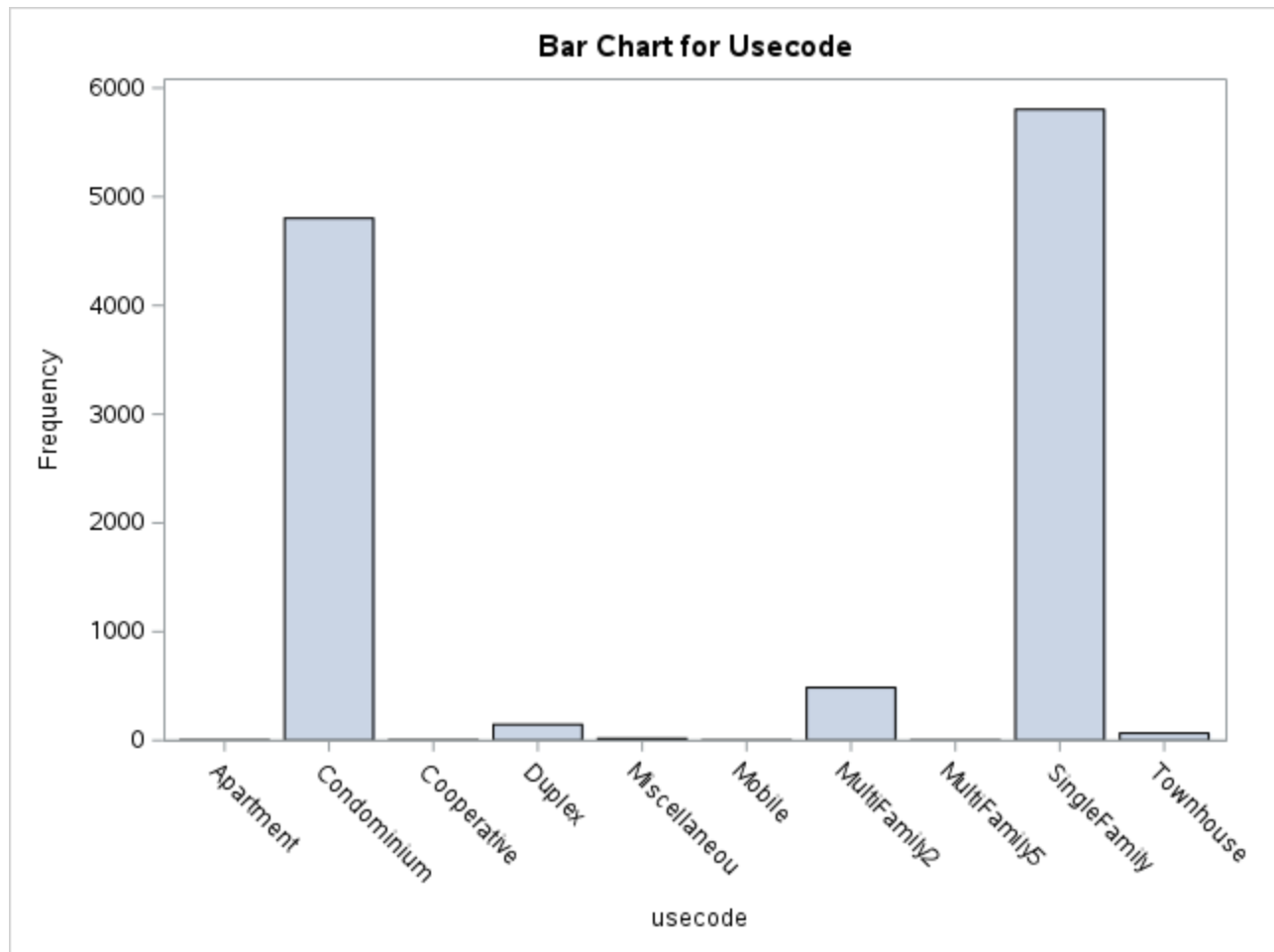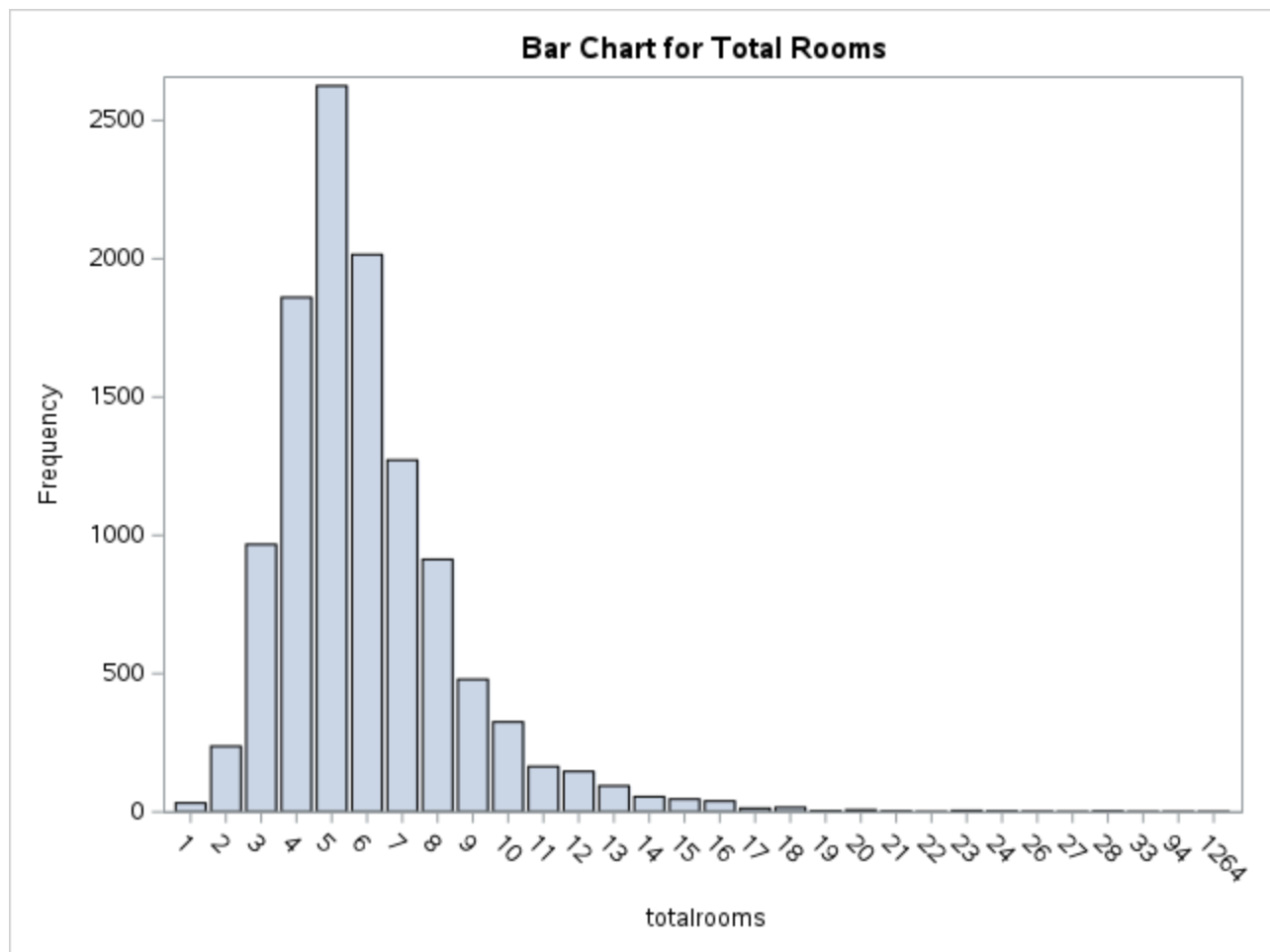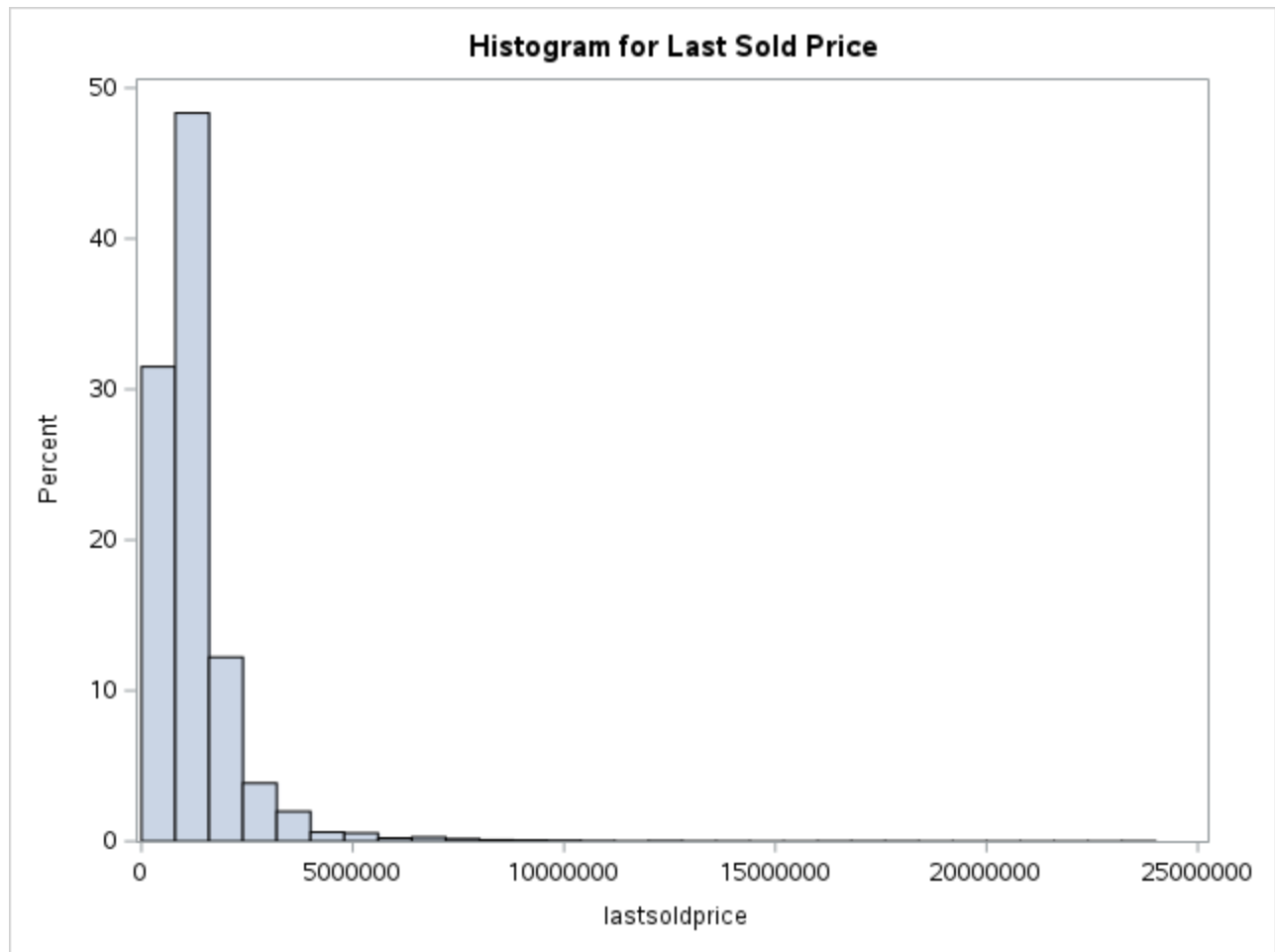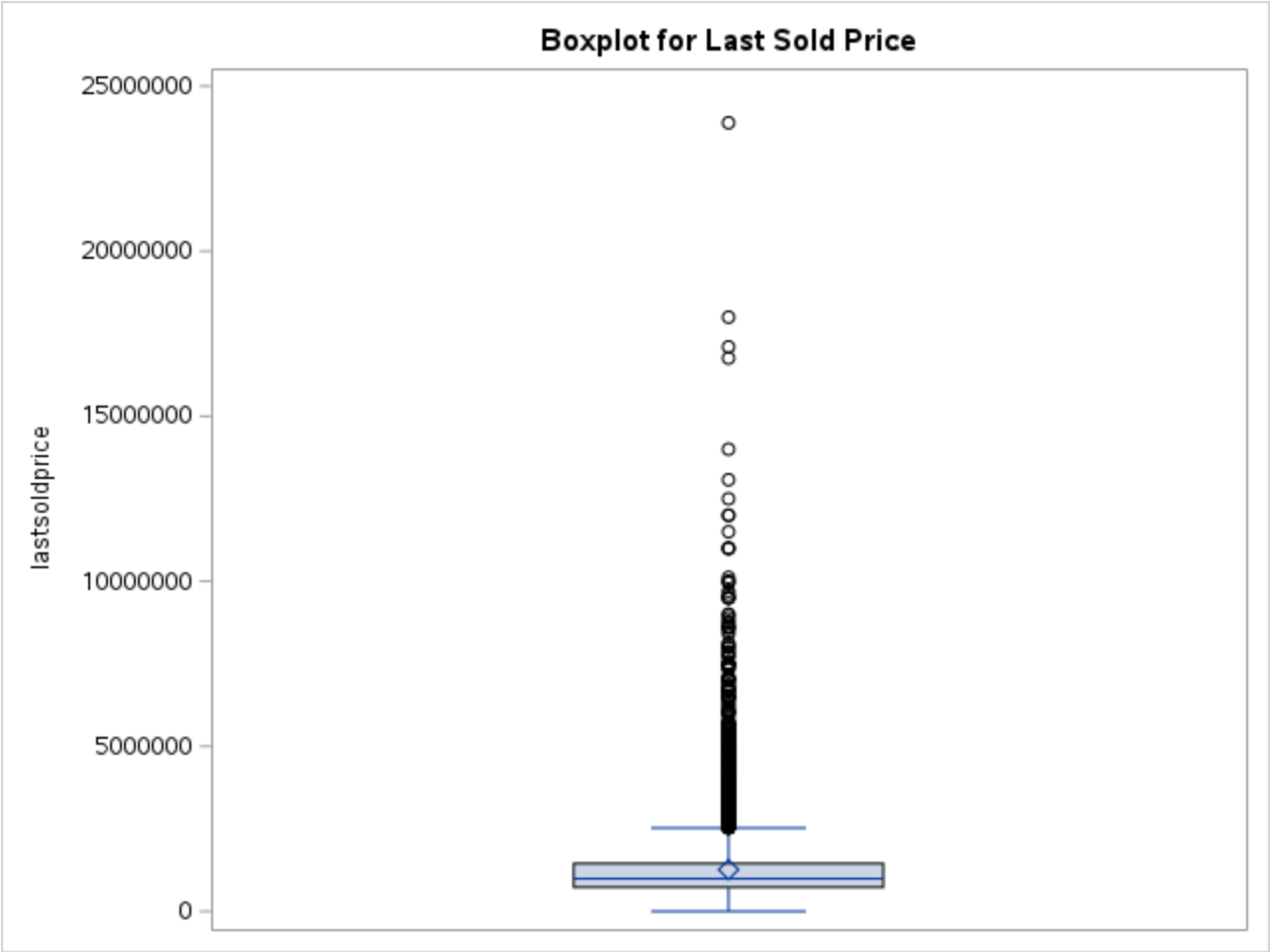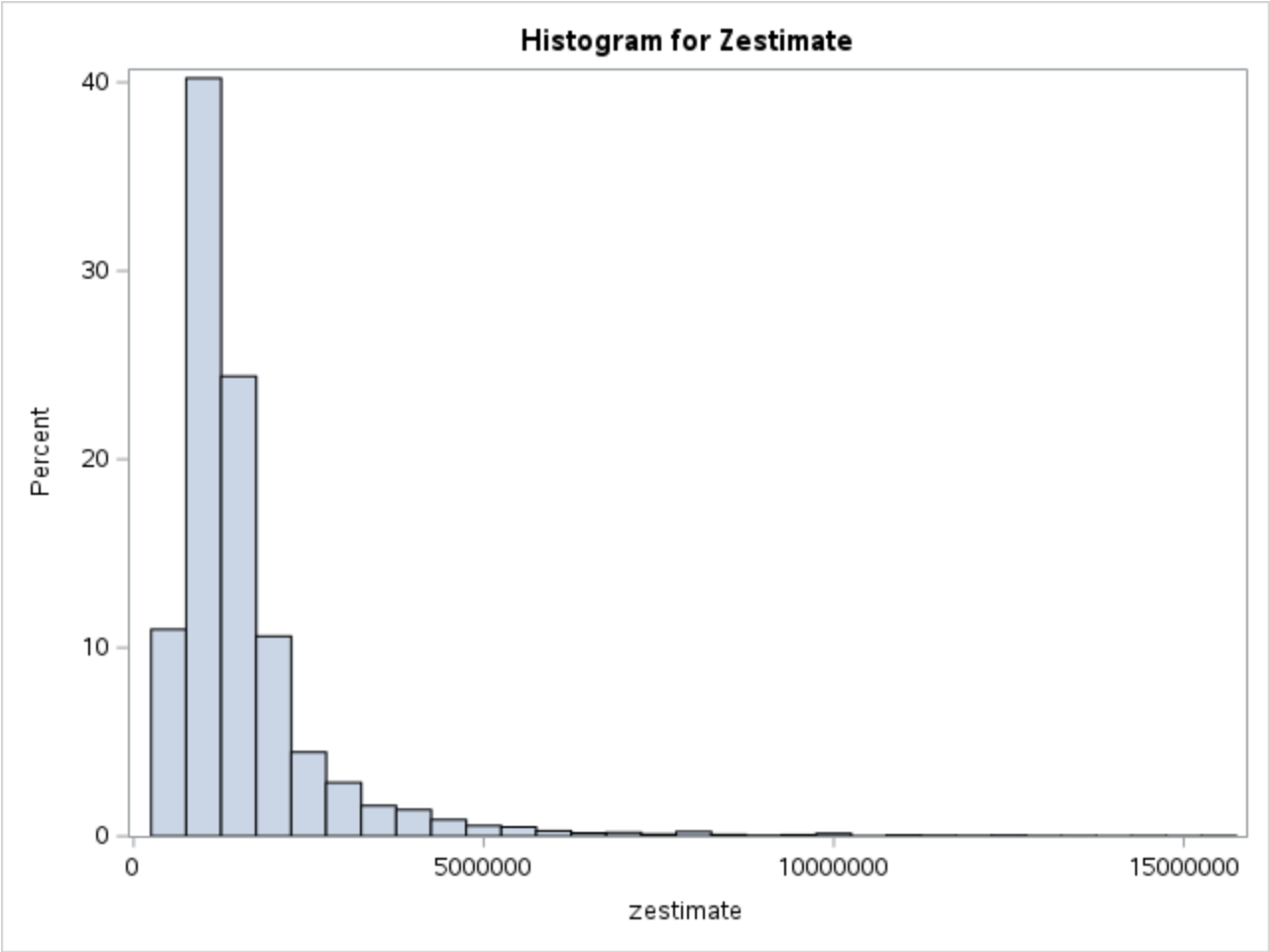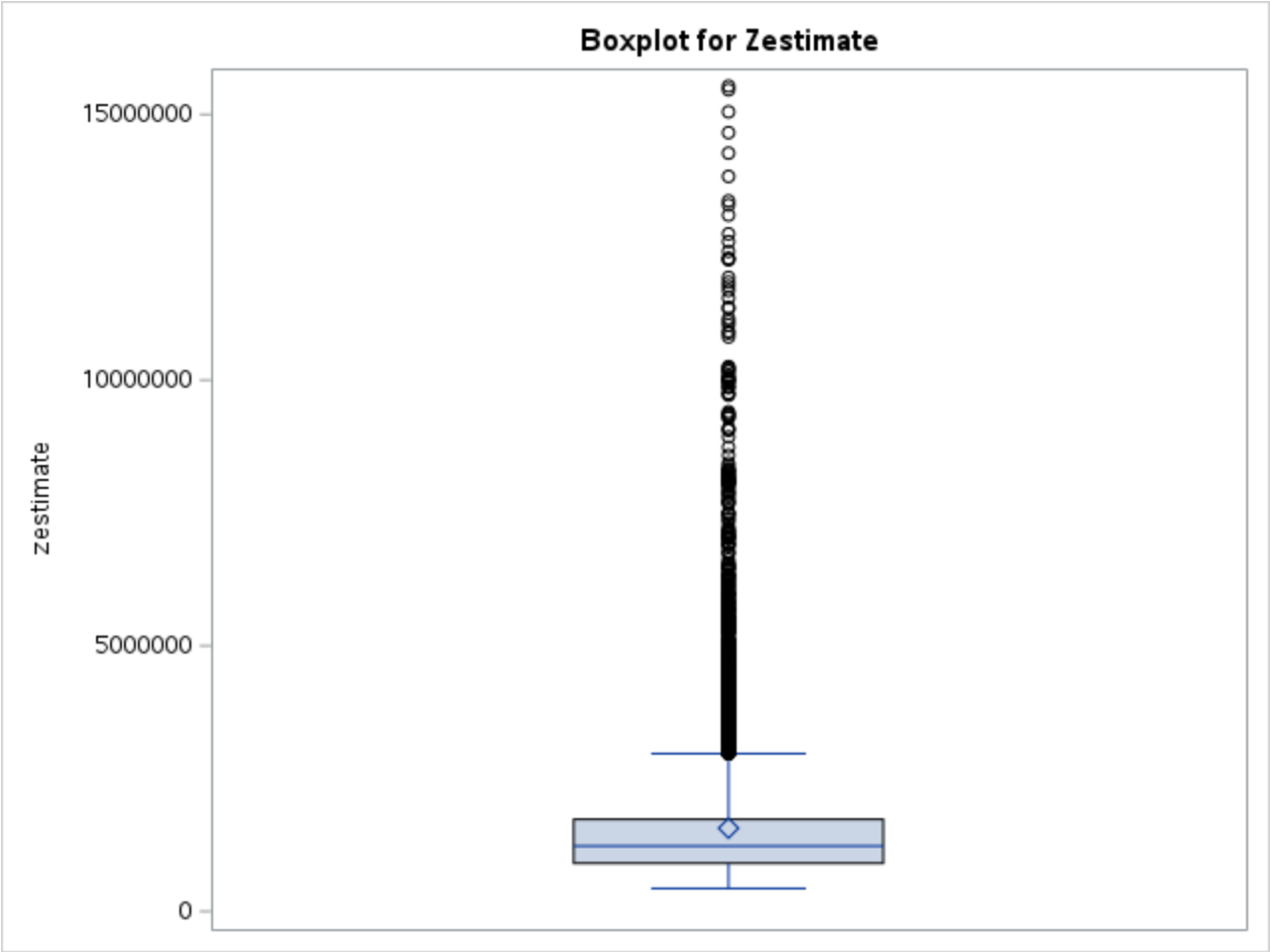
Boxplot for Zestimate

Obs	zipcode
1	94102
2	94103
3	94104
4	94105
5	94107
6	94108
7	94109
8	94110
9	94111
10	94112
11	94114
12	94115
13	94116
14	94117
15	94118
16	94121
17	94122
18	94123
19	94124
20	94127
21	94131

Obs	zipcode
22	94132
23	94133
24	94134
25	94158

Average Zestimate between Zip Codes 94110 and 94131

The TTEST Procedure

Variable: zestimate

zipcode	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
94110		935	1453542	607715	19874.4	534440	7406262
94131		687	1697687	840116	32052.4	477816	6164949
Diff (1-2)	Pooled		-244144	715403	35949.4		
Diff (1-2)	Satterthwaite		-244144		37714.0		

zipcode	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
94110		1453542	1414539	1492546	607715	581364	636587
94131		1697687	1634754	1760619	840116	797920	887059
Diff (1-2)	Pooled	-244144	-314657	-173632	715403	691597	740918
Diff (1-2)	Satterthwaite	-244144	-318138	-170151			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	1620	-6.79	<.0001
Satterthwaite	Unequal	1186.1	-6.47	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	686	934	1.91	<.0001

Average Zestimate vs. Last Sold Price

The TTEST Procedure

Difference: zestimate - lastsoldprice

N	Mean	Std Dev	Std Err	Minimum	Maximum
11330	301767	591958	5561.3	-9616379	11142925

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
301767	290866	312668	591958	584350	599768

DF	t Value	Pr > |t|
11329	54.26	<.0001

Simple Linear Regression of Bedrooms and Usecode

The GLM Procedure

Class Level Information		
Class	**Levels**	**Values**
usecode	10	Apartment Condominium Cooperative Duplex Miscellaneou Mobile MultiFamily2 MultiFamily5 SingleFamily Townhouse

Number of Observations Read	11330
Number of Observations Used	11330

Simple Linear Regression of Bedrooms and Usecode

The GLM Procedure

Dependent Variable: bedrooms

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	4790.87867	532.31985	420.24	<.0001
Error	11320	14339.14746	1.26671		
Corrected Total	11329	19130.02613			

R-Square	Coeff Var	Root MSE	bedrooms Mean
0.250438	43.04810	1.125482	2.614475

Source	DF	Type I SS	Mean Square	F Value	Pr > F
usecode	9	4790.878666	532.319852	420.24	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
usecode	9	4790.878666	532.319852	420.24	<.0001

Simple Linear Regression

The REG Procedure
Model: MODEL1
Dependent Variable: bedrooms

Number of Observations Read	11330
Number of Observations Used	11330

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	1	9759.79303	9759.79303	11799.0	<.0001
Error	11328	9370.23310	0.82717		
Corrected Total	11329	19130			

Root MSE	0.90949	**R-Square**	0.5102
Dependent Mean	2.61447	**Adj R-Sq**	0.5101
Coeff Var	34.78677		

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	0.85961	0.01828	47.03	<.0001

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
bathrooms	1	0.88620	0.00816	108.62	<.0001

Correlation Coefficients and Scatter Plot Matrix

The CORR Procedure

1 With Variables:	zestimate
8 Variables:	zestimate bedrooms bathrooms finishedsqft lastsoldprice totalrooms yearbuilt zipcode

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
zestimate	11330	1565695	1229417	1.77393E10	432385	15533253
bedrooms	11330	2.61447	1.29946	29622	0	20.00000
bathrooms	11330	1.98023	1.04736	22436	0.50000	14.00000
finishedsqft	11330	1585	921.97825	17962819	1.00000	27275
lastsoldprice	11330	1263928	1042079	1.43203E10	535.00000	23889000
totalrooms	11330	6.11156	12.12582	69244	1.00000	1264
yearbuilt	11330	1948	37.91120	22076484	1860	2016
zipcode	11330	94117	9.40088	1066344613	94102	94158

Pearson Correlation Coefficients, N = 11330
Prob > |r| under H0: Rho=0

	zestimate	bedrooms	bathrooms	finishedsqft	lastsoldprice	totalrooms	yearbuilt	zipcode
zestimate	1.00000	0.47768 <.0001	0.60316 <.0001	0.75849 <.0001	0.87694 <.0001	0.11199 <.0001	-0.21256 <.0001	0.03453 0.0002

Correlation Coefficients and Scatter Plot Matrix

Obs	_TYPE_	_NAME_	zestimate	bedrooms	bathrooms	finishedsqft	lastsoldprice	totalrooms	yearbuilt	zipcode
1	MEAN		1565695.26	2.61	1.98	1585.42	1263928.19	6.11	1948.50	94116.91
2	STD		1229417.07	1.30	1.05	921.98	1042079.05	12.13	37.91	9.40
3	N		11330.00	11330.00	11330.00	11330.00	11330.00	11330.00	11330.00	11330.00
4	CORR	zestimate	1.00	0.48	0.60	0.76	0.88	0.11	-0.21	0.03

Correlation Coefficients and Scatter Plot Matrix

The REG Procedure
Model: MODEL1
Dependent Variable: zestimate

Number of Observations Read	11330
Number of Observations Used	11330

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1.424625E16	4.748749E15	18693.6	<.0001

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Error	11326	2.877154E15	2.540309E11		
Corrected Total	11329	1.71234E16			

Root MSE	504015	R-Square	0.8320
Dependent Mean	1565695	Adj R-Sq	0.8319
Coeff Var	32.19112		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-136521	10421	-13.10	<.0001
bathrooms	1	25373	6566.18158	3.86	0.0001
finishedsqft	1	419.68249	8.25507	50.84	<.0001
lastsoldprice	1	0.78058	0.00602	129.75	<.0001

Correlation Coefficients and Scatter Plot Matrix

The REG Procedure
Model: MODEL1
Dependent Variable: zestimate

Number of Observations Read	11330
Number of Observations Used	11330

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1.426167E16	2.852334E15	11286.8	<.0001
Error	11324	2.86173E15	2.527137E11		
Corrected Total	11329	1.71234E16			

Root MSE	502706	R-Square	0.8329
Dependent Mean	1565695	Adj R-Sq	0.8328
Coeff Var	32.10755		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	193236	270189	0.72	0.4745
bathrooms	1	53029	7615.19000	6.96	<.0001
finishedsqft	1	447.00624	9.13460	48.94	<.0001
lastsoldprice	1	0.77160	0.00612	126.10	<.0001
bedrooms	1	-45902	5930.75133	-7.74	<.0001
yearbuilt	1	-152.15986	137.06924	-1.11	0.2670

Correlation Coefficients and Scatter Plot Matrix

The REG Procedure
Model: MODEL1
Dependent Variable: zestimate

Number of Observations Read	11330
Number of Observations Used	11330

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1.426167E16	2.852334E15	11286.8	<.0001
Error	11324	2.86173E15	2.527137E11		
Corrected Total	11329	1.71234E16			

Root MSE	502706	R-Square	0.8329
Dependent Mean	1565695	Adj R-Sq	0.8328
Coeff Var	32.10755		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	193236	270189	0.72	0.4745
bathrooms	1	53029	7615.19000	6.96	<.0001
finishedsqft	1	447.00624	9.13460	48.94	<.0001
lastsoldprice	1	0.77160	0.00612	126.10	<.0001
bedrooms	1	-45902	5930.75133	-7.74	<.0001
yearbuilt	1	-152.15986	137.06924	-1.11	0.2670

Correlation Coefficients and Scatter Plot Matrix

Obs	bathrooms	finishedsqft	lastsoldprice	bedrooms	yearbuilt	MODEL1
1	2	1230	900000	3	1997	1101984.02
2	3	2200	1300000	4	2002	1850587.07
3	1	845	700000	2	1989	769656.57
4	4	2500	2200000	5	2011	2684888.45

Average of zestimate by bathrooms

Obs	bathrooms	_TYPE_	_FREQ_	zestimate
1	0.5	0	1	1261265
2	1	0	3903	995595.31745
3	1.1	0	2	1115667
4	1.25	0	23	1391978.6957
5	1.3	0	4	1268753
6	1.5	0	437	1320067.6934
7	1.75	0	8	1392044.125
8	1.8	0	1	1064590
9	2	0	4161	1438119.1995

Obs	bathrooms	_TYPE_	_FREQ_	zestimate
10	2.1	0	1	1341075
11	2.2	0	2	1061038.5
12	2.25	0	22	1939354.8182
13	2.3	0	3	2200753.3333
14	2.5	0	527	2018993.6186
15	2.75	0	6	2063902.1667
16	3	0	1160	1972618.2828
17	3.1	0	1	4016477
18	3.25	0	3	2023643.3333
19	3.3	0	1	1577169
20	3.5	0	326	3093116.5951
21	3.75	0	6	2867713.6667
22	4	0	362	2801971.4061
23	4.4	0	1	3110155
24	4.5	0	109	4806925.9358
25	5	0	131	3538158.1145
26	5.1	0	1	11035724
27	5.25	0	1	6767479
28	5.5	0	30	5612292.5667
29	5.75	0	1	6924608
30	6	0	49	4810541.8367
31	6.5	0	11	9576799.2727
32	7	0	13	7226854.1538
33	7.5	0	7	9010615.2857
34	8	0	6	7898563.3333
35	8.5	0	1	10232127
36	9	0	4	4997167.5
37	12	0	1	2144324
38	12.5	0	1	5965458
39	13.5	0	1	1864778
40	14	0	2	808641

Average of zestimate by bedrooms

Obs	bedrooms	_TYPE_	_FREQ_	zestimate
1	0	0	138	785880.25362
2	1	0	1701	895983.4274
3	2	0	4029	1262074.8253
4	3	0	3288	1633544.1709
5	4	0	1455	2238060.0454
6	5	0	423	3006164.4208
7	6	0	173	3747993.7861
8	7	0	54	4026877.963
9	8	0	27	3265500.7778
10	9	0	28	4199407.1071
11	10	0	5	3204089.8
12	11	0	1	1903680
13	12	0	2	3742127.5

Obs	bedrooms	_TYPE_	_FREQ_	zestimate
14	13	0	1	9304469
15	14	0	1	1322365
16	16	0	3	4718716
17	20	0	1	1032780

Average of price_per_square_foot by zipcode

Obs	zipcode	_TYPE_	_FREQ_	price_per_square_foot
1	94102	0	318	928.85
2	94103	0	360	3018.24
3	94104	0	1	866.67
4	94105	0	340	2856.13
5	94107	0	857	1659.79
6	94108	0	114	1001.95
7	94109	0	653	919.28
8	94110	0	935	3474.67
9	94111	0	38	823.54
10	94112	0	877	573.12
11	94114	0	651	1012.22
12	94115	0	534	940.64
13	94116	0	655	650.60
14	94117	0	473	2691.56
15	94118	0	466	834.23
16	94121	0	527	695.39
17	94122	0	601	642.40
18	94123	0	430	1188.12
19	94124	0	393	431.21
20	94127	0	435	710.00
21	94131	0	687	3703.84
22	94132	0	292	3190.92
23	94133	0	183	7832.94
24	94134	0	439	499.77
25	94158	0	71	21914.18

Average of zestimate by totalrooms

Obs	totalrooms	_TYPE_	_FREQ_	zestimate
1	1	0	32	864987.34375
2	2	0	237	790535.90717
3	3	0	967	872228.56774
4	4	0	1860	1096498.8914
5	5	0	2625	1269967.613
6	6	0	2015	1530213.5012
7	7	0	1272	1738700.6682
8	8	0	913	1927580.8258
9	9	0	479	2472959.428
10	10	0	325	2840516.6615

Obs	totalrooms	_TYPE_	_FREQ_	zestimate
11	11	0	164	3267640.75
12	12	0	146	3047366.2466
13	13	0	94	3692985.0638
14	14	0	55	3530035.1091
15	15	0	46	4273266.8913
16	16	0	39	4120780.2821
17	17	0	13	3178413.9231
18	18	0	16	4178196.1875
19	19	0	4	3739224.25
20	20	0	7	6404786.8571
21	21	0	3	2012261
22	22	0	1	5684941
23	23	0	5	3106397.4
24	24	0	3	3662498.3333
25	26	0	2	8765272
26	27	0	1	5430463
27	28	0	3	8381745
28	33	0	1	9304469
29	94	0	1	502285
30	1264	0	1	1591608