

## Summary Report

**Business question:** What factors predict or contribute to the current market value of single family units?

Linear regression analysis was conducted to identify any relationship between variables within the Housing Affordability Data System dataset regarding factors that predict or contribute to the current market value of single family housing units. The conclusion of the analysis was that there are many factors that predict or contribute to the housing or current market value of single family units. Factors that have statistical significance and seem to contribute are:

- \* BEDRMS: number of bedrooms in the housing unit
- \* LMED: average median income for the area
- \* FMR: cost or fair market monthly rental rate of the housing unit
- \* BUILT: the year in which the housing unit was built
- \* ROOMS: number of rooms in the housing unit
- \* UTILITY: cost of utilities for owning the housing unit
- \* REGION: location in the United States of the housing unit; however, there is no statistical significance in the housing unit being located in the midwest region of the US as compared to the southern region and market value of the housing unit
  - \* METRO3: location to metropolitan city of the housing unit
  - \* OTHERCOST: cost or other utility costs for the occupant of the housing unit
  - \* ZINC2: annual household income of the occupant of the housing unit
  - \* AGE: Age of head of household of the housing unit
  - \* ZADEQ: Adequacy of housing unit
  - \* PER: Number of persons in household in the housing unit

Regression model has an adjusted **r-square value of 0.52**, suggesting the linear regression model explains 52% of the market value variation.

**Regression model listed below:**

$$\ln(\text{VALUE}) = \beta_0 + \beta_1 * \text{BEDRMS} + \beta_2 * \ln(\text{LMED}) + \beta_3 * \ln(\text{FMR}) + \beta_4 * \text{BUILT} + \beta_5 * \ln(\text{OTHERCOST}) + \beta_6 * \text{ROOMS} + \beta_7 * \text{UTILITY} + \beta_8 * \text{REGION\_WEST} + \beta_9 * \text{REGION\_NORTHEAST} + \beta_{10} * \text{REGION\_MIDWEST} + \beta_{11} * \text{ALL\_OTHER\_METROS} + \beta_{12} * \ln(\text{ZINC2}) + \beta_{13} * \text{AGE} + \beta_{14} * \text{PER} + \beta_{15} * \text{ZADEQ\_MODERATELY\_INADEQUATE} + \beta_{16} * \text{ZADEQ\_SEVERELY\_INADEQUATE}$$

**Interpretation of regression model predictor variables and their coefficients in relation to outcome variable (VALUE):**

- \* Intercept: has no practical managerial value
- \*  $\beta_1$ : one additional bedroom in the housing unit corresponds to a 5.72% decrease in market value of the housing unit all other variables held at their current level.
- \*  $\beta_2$ : a 1% change in area median income is associated with 0.38% increase in market value of the housing unit all other variables held at their current level.
- \*  $\beta_3$ : a 1% change in fair market monthly rent is associated with 0.65% increase in market value of the housing unit all other variables held at their current level.
- \*  $\beta_4$ : every additional year since the housing unit was built is associated with a 0.25% increase in market value of housing unit all other variables held at their current level.
- \*  $\beta_5$ : a 1% increase in the sum of other monthly costs is associated with 0.27% change in market value of the housing unit all other variables held at their current level
- \*  $\beta_6$ : one additional room in the housing unit corresponds to a 10.40% increase in market value of the housing unit all other variables held at their current level
- \*  $\beta_7$ : one dollar increase in monthly utility costs corresponds to with 0.04% increase in market value of the housing unit all other variables held at their current level
- \*  $\beta_8$ : when the housing unit is located in the western region of the US the market value of the housing unit increases by 29.12% as compared to the southern region of the US all other variables held at their current level
- \*  $\beta_9$ : when the housing unit is located in the northeastern region of the US, the market value of the housing unit increases by 14.58% as compared to the southern region of the US all other variables held at their current level
- \*  $\beta_{10}$ : there is no statistical significance for the market value of the housing unit and the housing unit being located in the midwest
- \*  $\beta_{11}$ : when the housing unit is located outside of the "Central City" metropolitan area, the market value of the housing unit increases by 8.38% as compared to the "Central City" metropolitan area all other variables held at their current level
  - \*  $\beta_{12}$ : a 1% increase in annual household income is associated with a 0.14% increase in market value of the housing unit all other variables held at their current level
  - \*  $\beta_{13}$ : a year increase in age for the head of household corresponds to 0.22% increase in market value of housing unit all other variables held at their current level
  - \*  $\beta_{14}$ : with each increase of one person in the household there corresponds to a 2.69% decrease in market value of the housing unit all other variables held at their current level
  - \*  $\beta_{15}$ : when the housing unit is in moderately inadequate condition there corresponds to a 14.97% decrease in market value of the housing unit as compared to an adequately conditioned housing unit all other variables held at their current level
  - \*  $\beta_{16}$ : when the housing unit is in severely inadequate condition there corresponds to a 10.23% decrease in market value of the housing unit as compared to an adequately conditioned housing unit

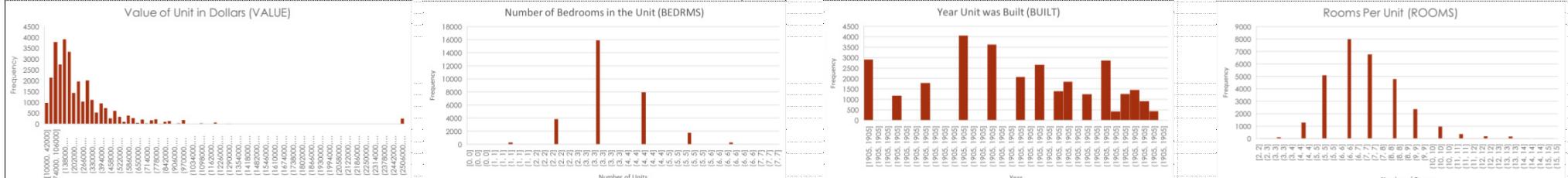
**Descriptive Statistics**  
**Single Family Houses, Flats, Apartments with Fair Market Value of at Least \$1000.00 owned in 2013**

\*LN\_VALUE, LN\_ZINC2, LN\_OTHERCOST, LN\_LMED, LN\_FMR are all natural log transformed variables

VALUE	LN_VALUE	BEDRMS	BUILT	ROOMS	UTILITY
Mean	262684.397	Mean	12.1545943	Mean	252.5376163
Standard Error	1621.170124	Standard Error	0.004558055	Standard Error	0.693492087
Median	190000	Median	12.15477935	Median	227.6666667
Mode	150000	Mode	11.91839057	Mode	192
Standard Deviation	281696.6779	Standard Deviation	0.792013648	Standard Deviation	26.57269206
Sample Variance	79353018337	Sample Variance	0.627285619	Sample Variance	1.644796837
Kurtosis	33.27347023	Kurtosis	0.895527692	Kurtosis	2.705356635
Skewness	4.838486576	Skewness	-0.040880423	Skewness	1.057836198
Range	2510000	Range	5.529429088	Range	1.057836198
Minimum	10000	Minimum	9.210340372	Minimum	2
Maximum	2520000	Maximum	14.73976946	Maximum	15
Sum	7931230000	Sum	366983.6658	Sum	204880
Count	30193	Count	30193	Count	30193
ZINC2	LN_ZINC2	LMED	LN_LMED	FMR	LN_FMR
Mean	90690.63611	Mean	11.0154779	Mean	7.11
Standard Error	495.7410009	Standard Error	0.005911485	Standard Error	0.001684137
Median	69787	Median	11.15320302	Median	7.09
Mode	99974	Mode	11.51266543	Mode	7.239932591
Standard Deviation	86140.61593	Standard Deviation	1.027187425	Standard Deviation	0.292637958
Sample Variance	7420205713	Sample Variance	1.055114006	Sample Variance	0.085636975
Kurtosis	11.48356862	Kurtosis	8.249314522	Kurtosis	-0.009038969
Skewness	2.744123496	Skewness	-1.621347339	Skewness	0.256836169
Range	1061920	Range	13.87559009	Range	1.99
Minimum	1	Minimum	0	Minimum	6.18
Maximum	1061921	Maximum	13.87559009	Maximum	8.16
Sum	2738222376	Sum	332590.3243	Sum	38790965.00
Count	30193	Count	30193	Count	30193
OTHERCOST	LN_OTHERCOST	AGE	PER		
Mean	99.54439495	Mean	4.268660126	Mean	2.650912463
Standard Error	0.602896456	Standard Error	0.004935456	Standard Error	0.008213601
Median	75	Median	4.317488114	Median	2
Mode	100	Mode	4.605170186	Mode	2
Standard Deviation	104.7600904	Standard Deviation	0.857591356	Standard Deviation	1.427206236
Sample Variance	10974.67655	Sample Variance	0.735462934	Sample Variance	2.036917639
Kurtosis	63.95248196	Kurtosis	8.569733188	Kurtosis	2.529232982
Skewness	5.751357835	Skewness	-1.190660256	Skewness	1.159564403
Range	2020.833333	Range	10.09621313	Range	19
Minimum	0.083333333	Minimum	-2.48490665	Minimum	1
Maximum	2020.916667	Maximum	7.611306483	Maximum	20
Sum	3005543.917	Sum	128883.6552	Sum	80039
Count	30193	Count	30193	Count	30193

## Graphs and Charts

### Histograms





## Statistical Tests

Linear Regression						
SUMMARY OUTPUT						
<b>Regression Statistics</b>						
Multiple R	0.719911309					
R Square	0.518272292					
Adjusted R Square	0.51808075					
Standard Error	0.549819075					
Observations	30193					
<b>ANOVA</b>						
	df	SS	MS	F	Significance F	
Regression	12	9815.562783	817.9635652	2705.791662	0	
Residual	30180	9123.444625	0.302301015			
Total	30192	18939.00741				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-5.212718285	0.378945442	-13.75585431	6.33067E-43	-5.955467491	-4.46996908
BEDRMS	-0.07372714	0.00695517	-10.60033689	3.3017E-26	-0.087359568	-0.060094712
LN_LMED	0.393561215	0.032702615	12.03454873	2.78808E-33	0.329462697	0.457659734
LN_FMR	0.637435388	0.02369989	26.89613299	1.7584E-157	0.590982595	0.683888182
BUILT	0.002644208	0.00013016	20.31501275	3.86028E-91	0.002389088	0.002899327
LN_OTHERCOST	0.279841082	0.003982998	70.25890793	0	0.272034237	0.287647928
ROOMS	0.108573891	0.00302337	35.9154162	1.3554E-276	0.102647957	0.114499826
UTILITY	0.000344244	2.96175E-05	11.62297768	3.67018E-31	0.000286192	0.000402295
REGION_WEST	0.293517654	0.010897311	26.93486957	6.3435E-158	0.27215846	0.314876848
REGION_NORTHEAST	0.14548898	0.010950342	13.28622577	3.609F-40	0.124025804	0.166952155
REGION_MIDWEST	-0.021175247	0.009325269	-2.270738449	0.02316983	-0.002897322	-0.039453172
ALL_OTHER_METROS	0.089691474	0.007944088	11.29034287	1.67856E-29	0.074120724	0.105262224
LN_ZINC2	0.120575898	0.003345376	36.0425602	1.4711E-278	0.11401882	0.127132977

conclusion: p-values for all variables are less than alpha = 0.05; all variables are statistically significant

Improving Linear Regression Model via Stepwise Regression						
SUMMARY OUTPUT						
<b>Regression Statistics</b>						
Multiple R	0.72211662					
R Square	0.521452412					
Adjusted R Square	0.521246272					
Standard Error	0.548010336					
Observations	30193					
<b>ANOVA</b>						
	df	SS	MS	F	Significance F	
Regression	13	9875.791102	759.6762386	2529.595282	0	
Residual	30179	9063.216307	0.300315329			
Total	30192	18939.00741				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-5.986125292	0.381626767	-15.68581089	3.12711E-55	-6.73413001	-5.238120575
BEDRMS	-0.0648688	0.006960454	-9.319623218	1.24446E-20	-0.078511584	-0.051226015
LN_LMED	0.40125763	0.032599564	12.30868093	9.85599E-35	0.337361096	0.465154164
LN_FMR	0.631802786	0.023625273	26.74266648	9.8514E-156	0.585496245	0.678109326
BUILT	0.002858197	0.000130609	21.88359415	2.4591E-105	0.002602198	0.003114197
LN_OTHERCOST	0.275567289	0.003981349	69.21454656	0	0.267763675	0.283370903
ROOMS	0.10407726	0.003030107	34.34771913	1.1795E-253	0.098138122	0.110016399
UTILITY	0.00034758	2.9521E-05	11.77399887	6.24274E-32	0.000289718	0.000405443
REGION_WEST	0.291512367	0.010862385	26.83686505	8.3464E-157	0.27022163	0.312803105
REGION_NORTHEAST	0.144539378	0.010914544	13.2482284	6.41755E-40	0.123146406	0.16593235
REGION_MIDWEST	-0.019925856	0.00929501	-2.143715237	0.032063478	-0.038144472	-0.001707239
ALL_OTHER_METROS	0.08402712	0.00792805	10.59871188	3.35937E-26	0.064887804	0.095566436
LN_ZINC2	0.135044672	0.003487389	38.72372236	0	0.128209242	0.141880102
AGE	0.003074568	0.000217106	14.16157707	2.21436E-45	0.002649031	0.003500106

conclusion: BEDRMS is first variable to enter regression model and its p-value is less than alpha-to-exit value, AGE variable added to regression model

SUMMARY OUTPUT						
<b>Regression Statistics</b>						
Multiple R	0.723289709					
R Square	0.523148003					
Adjusted R Square	0.522926784					
Standard Error	0.547047683					
Observations	30193					
<b>ANOVA</b>						
	df	SS	MS	F	Significance F	
Regression	14	9907.9039	707.7074214	2364.84883	0	
Residual	30178	9031.103508	0.299261167			
Total	30192	18939.00741				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-5.90889649	0.381029331	-15.50772135	4.97381E-54	-6.655730409	-5.162062971
BEDRMS	-0.05724288	0.006987116	-8.192633194	2.65569E-16	-0.070937925	-0.043547835
LN_LMED	0.387248634	0.032570386	11.8895929	1.58481E-32	0.451087979	0.32340929
LN_FMR	0.645255399	0.0236195	27.31875771	2.4073E-162	0.598960173	0.691550625
BUILT	0.002866	0.000130382	21.9815851	2.9523E-106	0.002610446	0.003121554

LN_OTHERCOST	0.273765234	0.003978161	68.81703224	0	0.265967869	0.281562599	0.265967869	0.281562599
ROOMS	0.104222012	0.003024816	34.4556492	3.3028E-255	0.098293243	0.110150781	0.098293243	0.110150781
UTILITY	0.00040646	3.00123E-05	13.54311929	1.15199E-41	0.000347635	0.000465286	0.000347635	0.000465286
REGION_WEST	0.292643337	0.010843854	26.98702386	1.6044E-158	0.271388922	0.313897732	0.271388922	0.313897732
REGION_NORTHEAST	0.145761421	0.01089601	13.37750404	1.06922E-40	0.124404777	0.167118065	0.124404777	0.167118065
REGION_MIDWEST	-0.015595451	0.009288095	-1.679709683	0.093146868	-0.033800513	0.00260961	-0.033800513	0.00260961
ALL_OTHER_METROS	0.084472676	0.00791424	10.67330383	1.51036E-26	0.0689760427	0.099984924	0.0689760427	0.099984924
LN_ZINC2	0.139663237	0.003509697	39.79352663	0	0.132784008	0.146542393	0.132784008	0.146542393
AGE	0.002240804	0.000231188	9.69255173	3.49411E-22	0.001787665	0.002693942	0.001787665	0.002693942
PER	-0.027262542	0.002631797	-10.35890606	4.20108E-25	-0.032420977	-0.022104107	-0.032420977	-0.022104107

conclusion: BEDRMS is first variable to enter regression model and its p-value is less than alpha-to-exit value, PER variable added to regression model

#### SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.723712232
R Square	0.523759394
Adjusted R Square	0.523506881
Standard Error	0.546714991
Observations	30193

#### ANOVA

	df	SS	MS	F	Significance F
Regression	16	9919.483046	619.9676904	2074.183103	0
Residual	30176	9019.524362	0.298897281		
Total	30192	18939.00741			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-5.730871444	0.361905809	-15.00598135	1.02286E-50	-6.479423099	-4.982319788	-6.479423099	-4.982319788
BEDRMS	-0.057191648	0.006983289	-8.189786711	<b>2.71912E-16</b>	-0.070879193	-0.043504104	-0.070879193	-0.043504104
LN_LMED	0.38405834	0.032556409	11.79363205	4.79568E-32	0.320233885	0.447857782	0.320233885	0.447857782
LN_FMR	0.645287739	0.023406576	27.33508447	1.5571E-162	0.59017845	0.691557633	0.59017845	0.691557633
BUILT	0.0020803033	0.000130695	21.44717048	2.7727E-101	0.0025464865	0.0030592	0.0025464865	0.0030592
LN_OTHERCOST	0.273639233	0.003975814	68.82596966	0	0.2658446469	0.281431998	0.2658446469	0.281431998
ROOMS	0.104007825	0.00302321	34.40310544	1.8871E-254	0.098082204	0.109933447	0.098082204	0.109933447
UTILITY	0.000407035	2.99942E-05	13.57043677	7.95671E-42	0.000348245	0.000465825	0.000348245	0.000465825
REGION_WEST	0.291236609	0.010839666	26.86767355	3.7179E-157	0.269990401	0.312482816	0.269990401	0.312482816
REGION_NORTHEAST	0.145782817	0.010890463	13.38628244	9.50777E-41	0.124437045	0.167128589	0.124437045	0.167128589
<b>REGION_MIDWEST</b>	<b>-0.016945619</b>	<b>0.009286885</b>	<b>-1.824682698</b>	<b>0.686058808</b>	<b>-0.035148309</b>	<b>0.001257072</b>	<b>-0.035148309</b>	<b>0.001257072</b>
ALL_OTHER_METROS	0.083817082	0.007910224	10.597604372	3.45624E-26	0.068312705	0.099321458	0.068312705	0.099321458
LN_ZINC2	0.138542588	0.003512144	39.445602	0	0.131658439	0.145426736	0.131658439	0.145426736
AGE	0.002212675	0.000231092	9.574852946	1.09508E-21	0.001759725	0.002645626	0.001759725	0.002645626
PER	-0.026888764	0.002630883	-10.22043471	1.76081E-24	-0.032045406	-0.021732122	-0.032045406	-0.021732122
ZADEQ_MODERATELY_INA	DEQUATE	-0.149719223	0.027844908	-5.37689767	7.63425E-08	-0.20429643	-0.095142016	-0.20429643
ZADEQ_SEVERELY_INADEQ	UATE	-0.102331419	0.031717495	-3.226339843	0.00125518	-0.164499061	-0.040163777	-0.164499061
ZADEQ_MODERATELY_INA	DEQUATE	-0.06700063	-0.017767679	-0.032657905	-0.02461822	-0.068753469	-0.032350318	-0.028769792
ZADEQ_SEVERELY_INADEQ	UATE	-0.033707449	-0.0270283	0.025296016	0.00221287	-0.056613618	-0.015835515	-0.032432992

conclusion: BEDRMS p-value is less than alpha-to-exit value, ZADEQ dummy variables added to regression model, Excel does not accommodate additional variables

Detecting Multicollinearity via Correlation Matrix										
	LN_VALUE	BEDRMS	LMED	FMR	BUILT	LN_OTHERCOST	ROOMS	UTILITY	REGION_WEST	REGION_NORTHEAST
	1									
LN_VALUE	1									
BEDRMS	0.349356187	1								
LMED	0.379978108	0.100796211	1							
FMR	0.540971847	0.478373433	0.683909079	1						
BUILT	0.17975282	0.141359142	-0.157950553	-0.009533726	1					
LN_OTHERCOST	0.484012152	0.169594825	0.14451998	0.245000668	0.178934023	1				
ROOMS	0.425075464	0.744466882	0.123035791	0.349642535	0.141897797	0.234027022	1			
UTILITY	0.310049876	0.320632022	0.218660209	0.320721059	-0.020008622	0.189162798	0.3548668529	1		
REGION_WEST	0.215607225	0.036113558	0.017154818	0.331268873	0.090392725	0.034742166	0.003311656	-0.017137005	1	
REGION_NORTHEAST	0.204057969	0.013066553	0.541856153	0.330526633	-0.20683294	0.045296847	0.037211134	0.228194654	-0.249802734	1
REGION_MIDWEST	-0.239443141	-0.034247467	-0.152123242	-0.38028809	-0.090052872	-0.11604382	-0.021119765	-0.144323799	-0.285996469	-0.372037884
ALL_OTHER_METROS	0.056750391	0.032054368	0.007692986	-0.052916661	0.151153934	0.019210434	0.048873071	-0.024215311	-0.088456584	0.059083872
LN_ZINC2	0.379124888	0.254196112	0.149561678	0.225325028	0.147312933	0.213016462	0.308939265	0.226916307	0.050399576	0.060399251
AGE	-0.021840518	-0.108483595	-0.010359123	-0.040291848	-0.138524948	0.005428025	-0.04620475	-0.04956857	-0.006164336	0.019912854
PER	0.141882168	0.33377307	0.073111081	0.218773205	0.088180339	0.061161075	0.26843246	0.28928055	0.032760463	0.034952261
ZADEQ_MODERATELY_INA	DEQUATE	-0.06700063	-0.017767679	-0.032657905	-0.02461822	-0.068753469	-0.032350318	-0.028769792	-0.005714947	-0.017945522
ZADEQ_SEVERELY_INADEQ	UATE	-0.033707449	-0.0270283	0.025296016	0.00221287	-0.056613618	-0.015835515	-0.032432992	-0.00235269	-0.021413189
ZADEQ_MODERATELY_INA	DEQUATE	-0.017767679	-0.032657905	-0.02461822	-0.068753469	-0.032350318	-0.028769792	-0.005714947	-0.017945522	0.033212875
ZADEQ_SEVERELY_INADEQ	UATE	-0.033707449	-0.0270283	0.025296016	0.00221287	-0.056613618	-0.015835515	-0.032432992	-0.00235269	0.002755414
ZADEQ_MODERATELY_INA	DEQUATE	-0.017767679	-0.032657905	-0.02461822	-0.068753469	-0.032350318	-0.028769792	-0.005714947	-0.017945522	-0.019246132
ZADEQ_SEVERELY_INADEQ	UATE	-0.033707449	-0.0270283	0.025296016	0.00221287	-0.056613618	-0.015835515	-0.032432992	-0.00235269	-0.002755414
ZADEQ_MODERATELY_INA	DEQUATE	-0.017767679	-0.032657905	-0.02461822	-0.068753469	-0.032350318	-0.028769792	-0.005714947	-0.017945522	-0.019246132
ZADEQ_SEVERELY_INADEQ	UATE	-0.033707449	-0.0270283	0.025296016	0.00221287	-0.056613618	-0.015835515	-0.032432992	-0.00235269	-0.019246132

conclusion: no variables are correlated by 80% or more; which means no variables are highly correlated to each other, with high correlation being a sign of possible multicollinearity