

Analysis of Property Values

Given the dataset for the indicators of property values of the given region, we look to gather useful insights from the same.

Codebook:

Variable	Definition	Type
propid	Property unique ID	Quantitative
township	Township region	Categorical (1 = Eastern, 2 = Central, 3 = Southern, 4 = Northern, 5 = Western)
assessor	ID of the assessor	Categorical
saleval	Sale value of the house	Quantitative
lastval	Value at last appraisal	Quantitative
time	Time since last appraisal	Quantitative

Selecting a random sample:

In order to make our analysis more efficient we use pick a random sample from the dataset before proceeding.

Calculation of the sample size with the objective to estimate the population-mean:

Our acceptable margin of error is $\epsilon = 5\%$ (our estimate should be within 5% of the actual mean) with 5% level of significance ($\alpha = 5\%$)

We have the equations:

$$P[|\bar{y}_n - \bar{Y}_N| \leq \epsilon \bar{Y}_N] = 1 - \alpha$$

Finally n is given by:

$$n = \frac{n_0}{1 + n_0/N}$$

Where:

$$d = \epsilon \bar{Y}_N$$
$$n_0 = \frac{Z_{\alpha/2}^2 \times S^2}{d^2}$$

As the population parameters are unknown we use their unbiased estimators to calculate the value. To proceed further we randomly pick a sample of size 50 to estimate \mathbf{s} and \bar{y}_n .

For the initial 50 samples we have:

$$\bar{y}_n = 173.17$$

$$s = 60.90232$$

$$d = 8.6585$$

$$Z_{\alpha/2} = 1.96$$

$$N = 1000$$

We have:

$$n = 160$$

We now draw $160 \cdot 50 = 110$ more samples from the data to obtain our final sample.

Analyzing our Sample:

We start by drawing Descriptive Statistics for the quantitative variables namely Saleval and Lastval.

Steps:

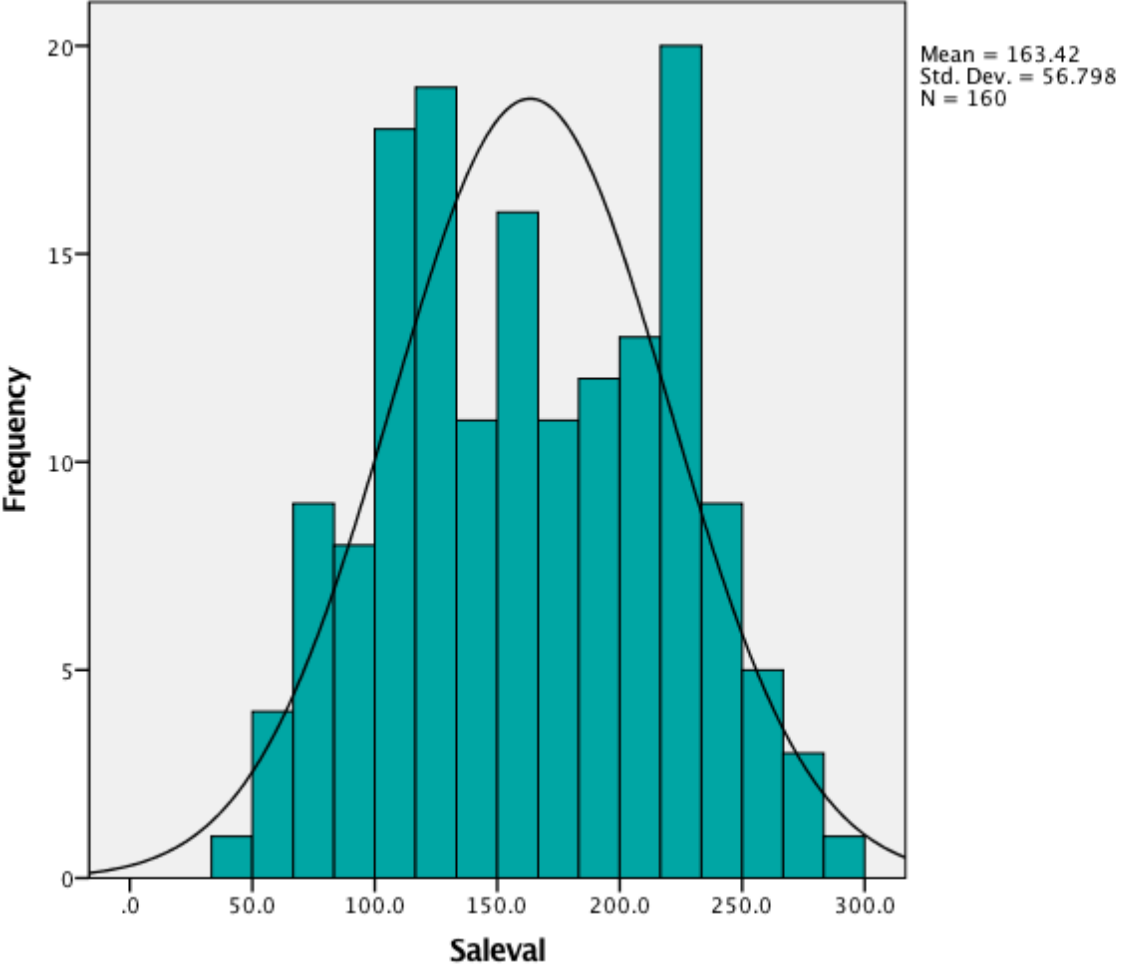
Descriptive Statistics

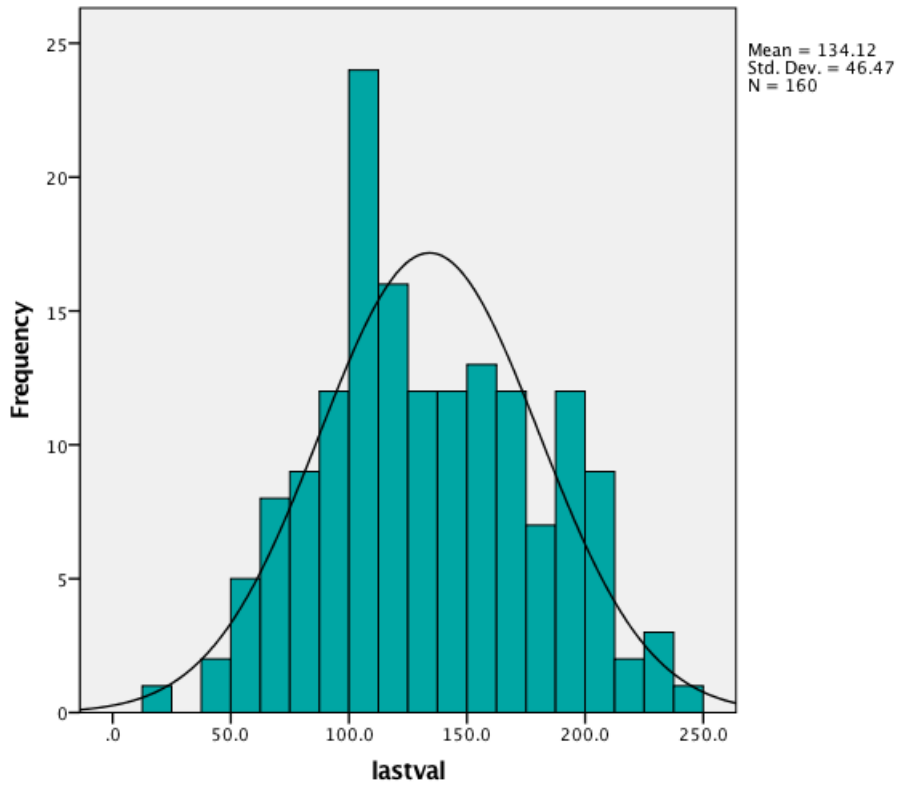
	N	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Saleval	160	163.415	56.7985	.066	.192	-.970	.381
lastval	160	134.121	46.4698	.167	.192	-.663	.381
Valid N (listwise)	160						

Township

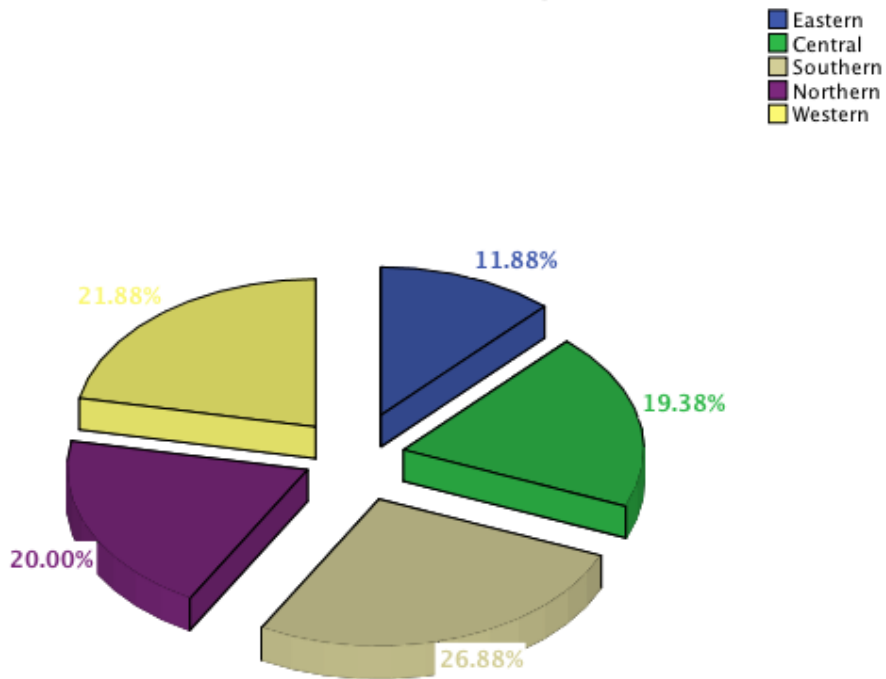
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Eastern	19	11.9	11.9	11.9
	Central	31	19.4	19.4	31.2
	Southern	43	26.9	26.9	58.1
	Northern	32	20.0	20.0	78.1
	Western	35	21.9	21.9	100.0
	Total	160	100.0	100.0	

Both the variables have a positive, albeit very low, coefficient of skewness, thus we should expect to see a right-tailed distribution for the same. Looking at the figures for kurtosis we see that both the variables have a high negative value for the same hence we should expect to see a Platykurtic distribution for the same. We now draw histograms for the same.





Frequency Distribution
Township



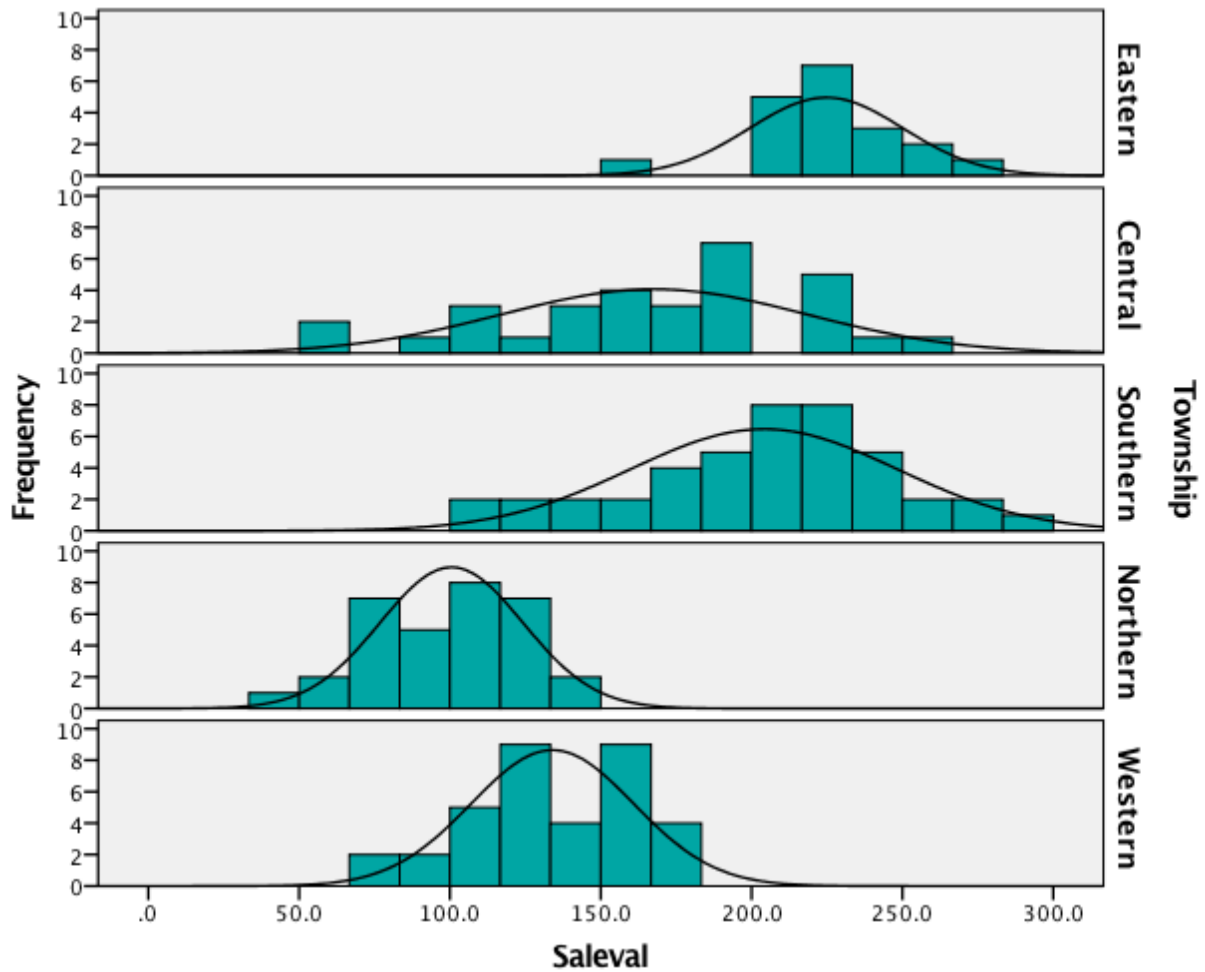
Let's now take a look at the Sale Values and the Last assessed values according to the various regions in our population.

To do so we use the following plots:

- Histogram (Area and Bar shaped)
- Boxplot

Sale Value

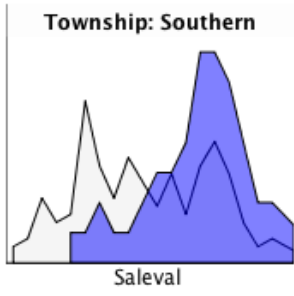
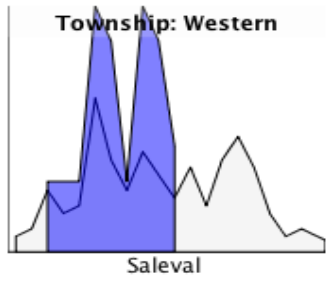
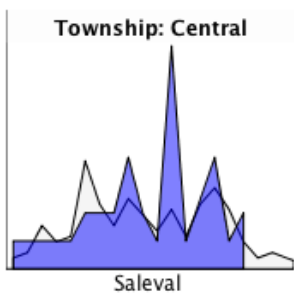
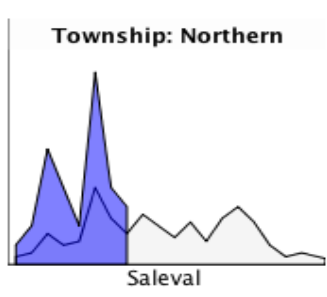
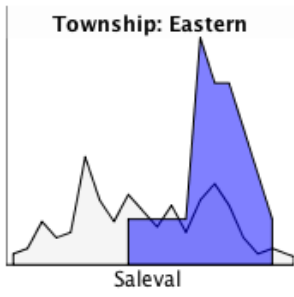
Bar Shaped Histograms:



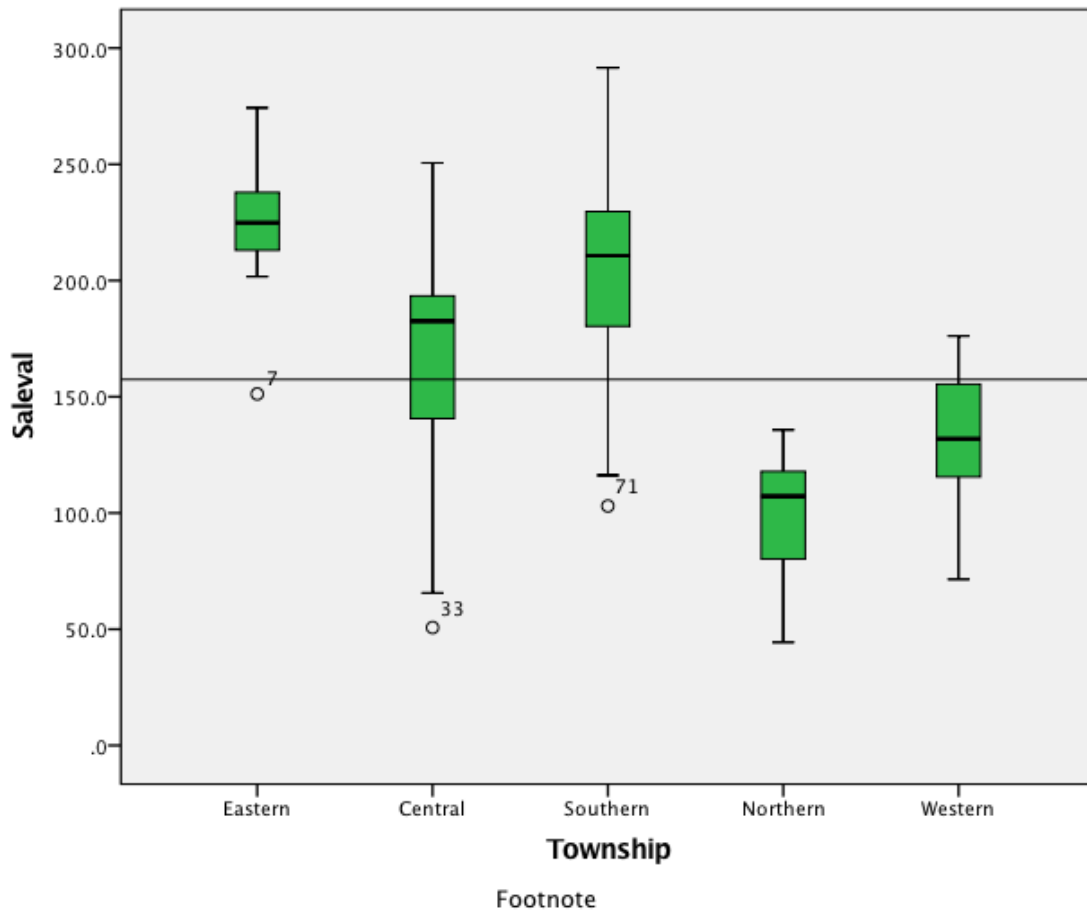
Area Shaped Histograms:

Chart Information

Settings	Value
Subgroups Defined by	Township
Missing Value Treatment	Variable By Variable
Color for Entire Sample	Whitesmoke
Color for Subgroups	Blue
Pattern for Entire Sample	Solid
Pattern for Subgroups	Solid



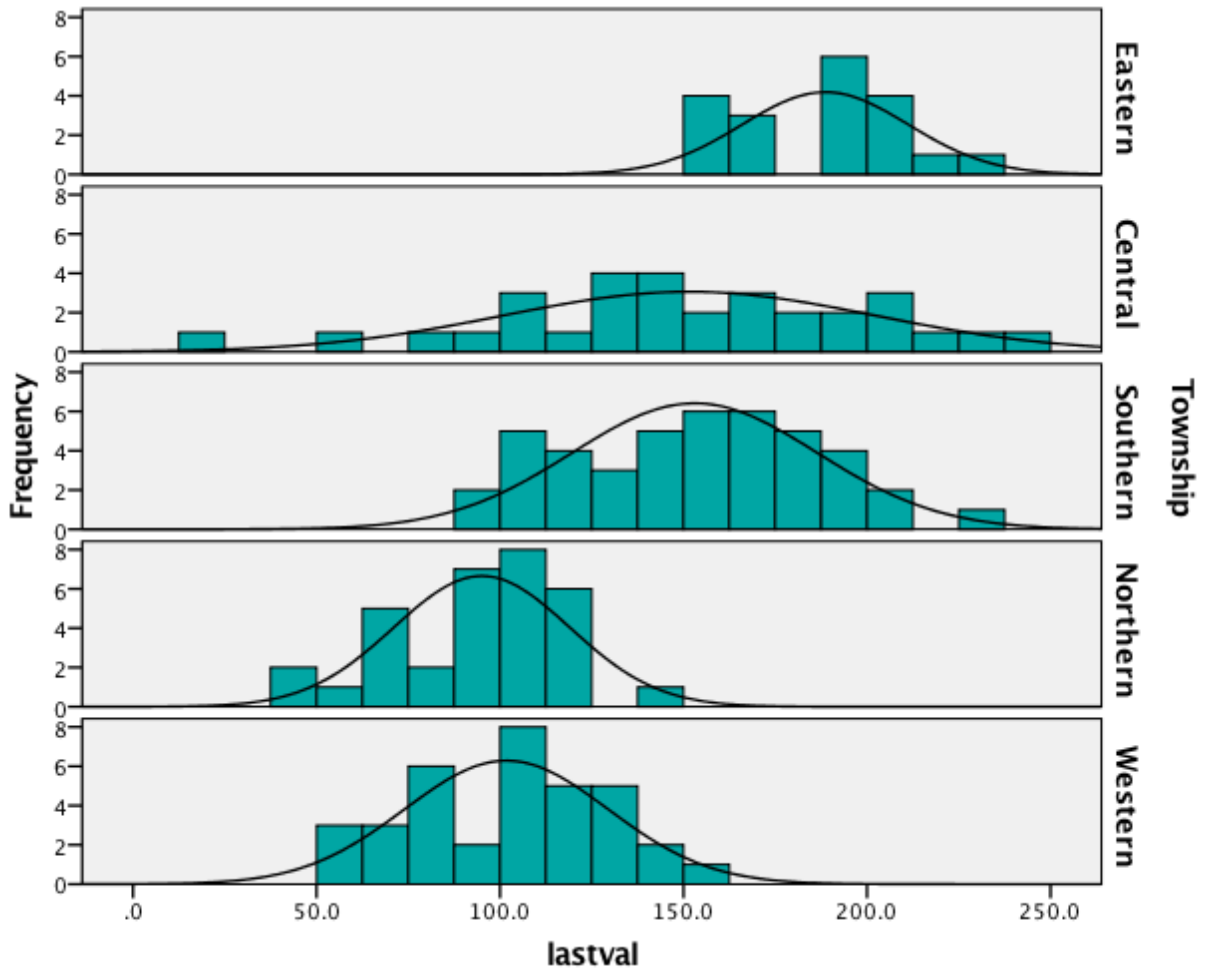
Boxplot:



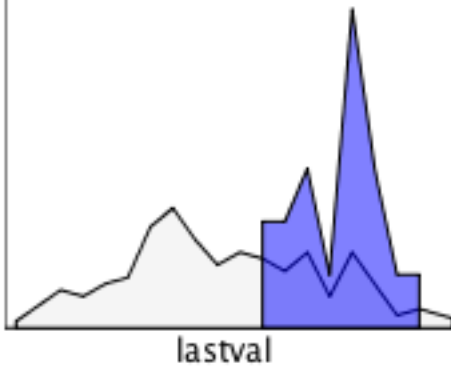
Albeit different in nature, the three plots tell us a very similar story about our sample, following are our inferences from the same:

- The sale values of the houses are the highest and least skewed in the Eastern region.
- The prices in the Northern region are concentrated towards the lower interval of prices.
- The median prices in the Northern and Western region are below the median price of the total sample (represented by the horizontal line in the boxplot).
- A peculiar highlight of the depictions is the distribution of the sale prices over the different regions; we examine the significance of these differences using a one-sample t-test further in our study.

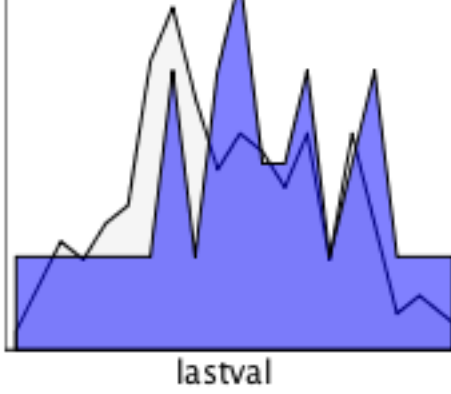
Last Evaluation:



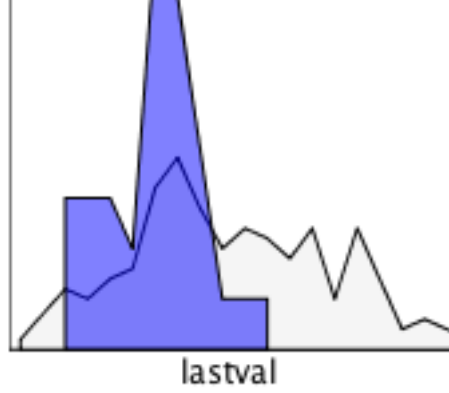
Township: Eastern



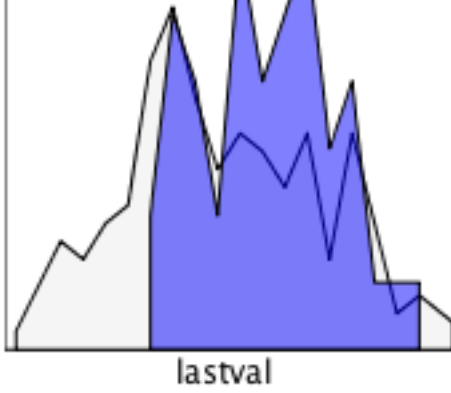
Township: Central



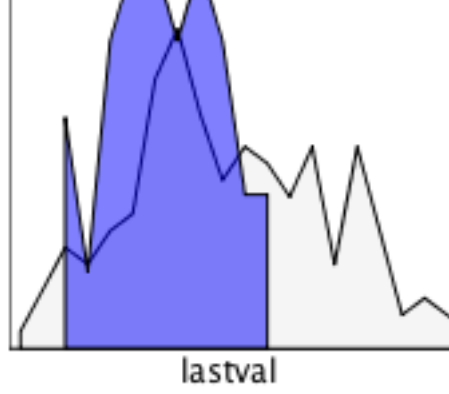
Township: Northern

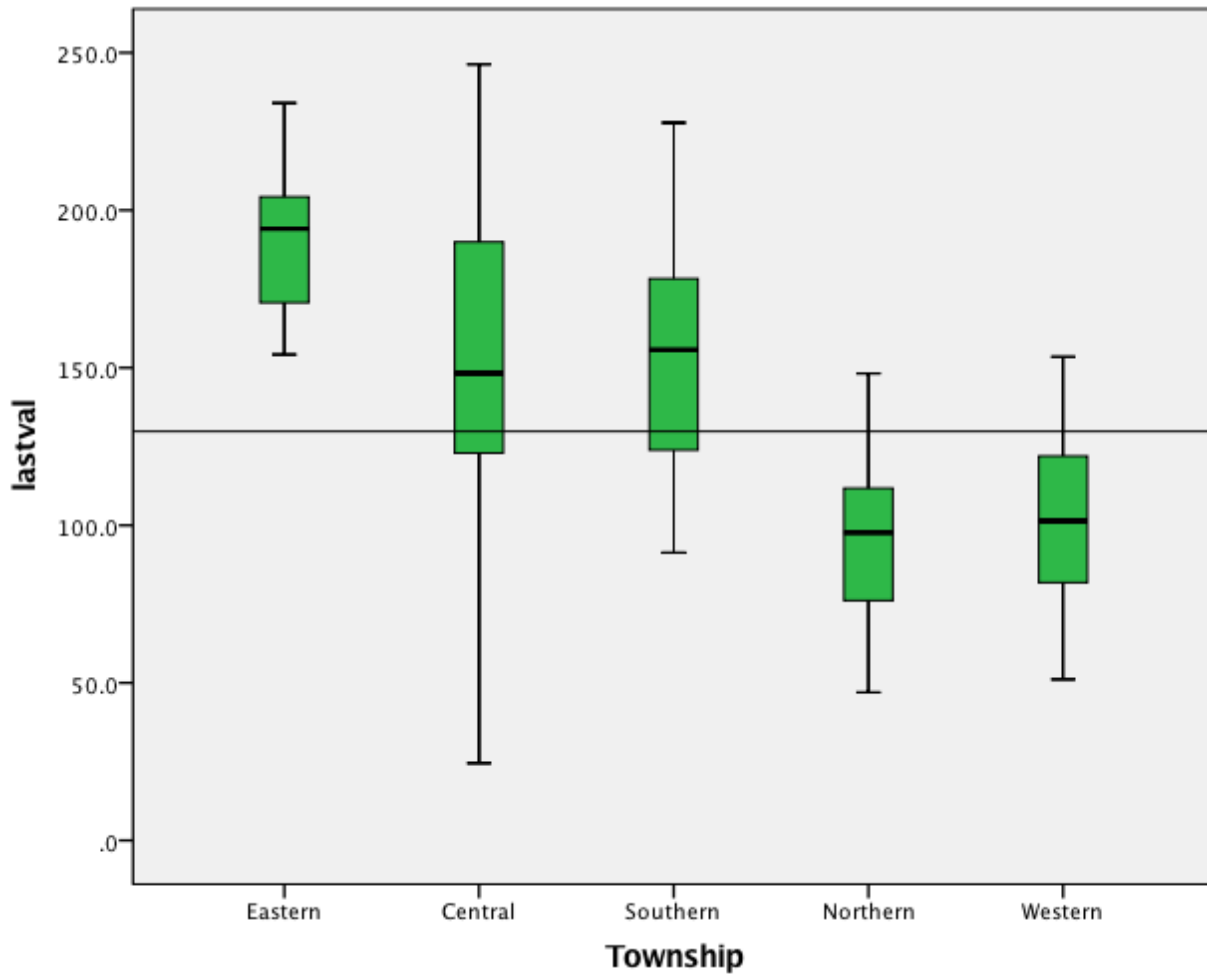


Township: Southern



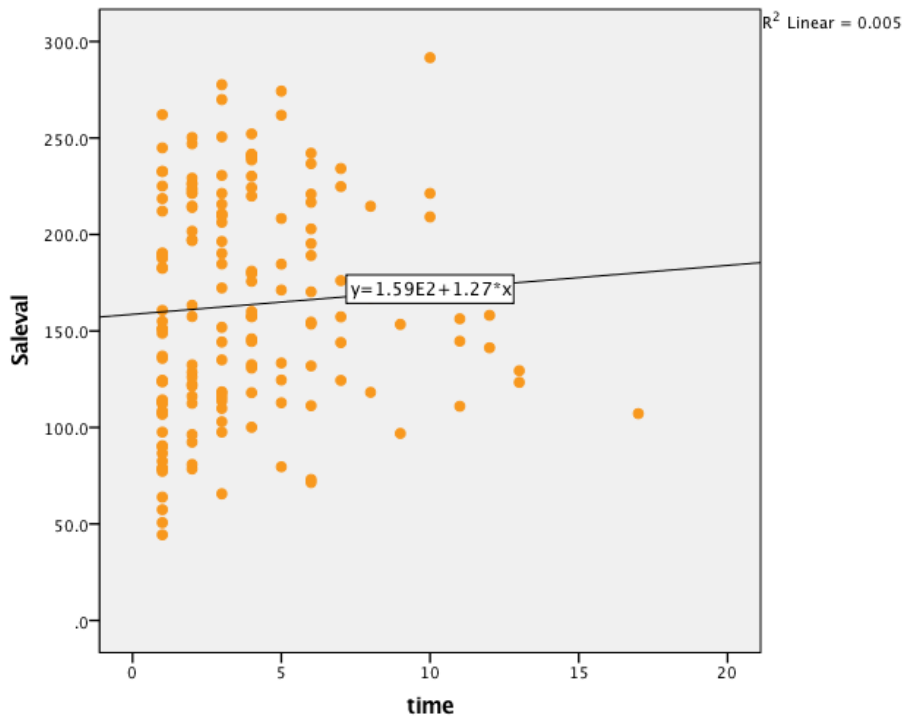
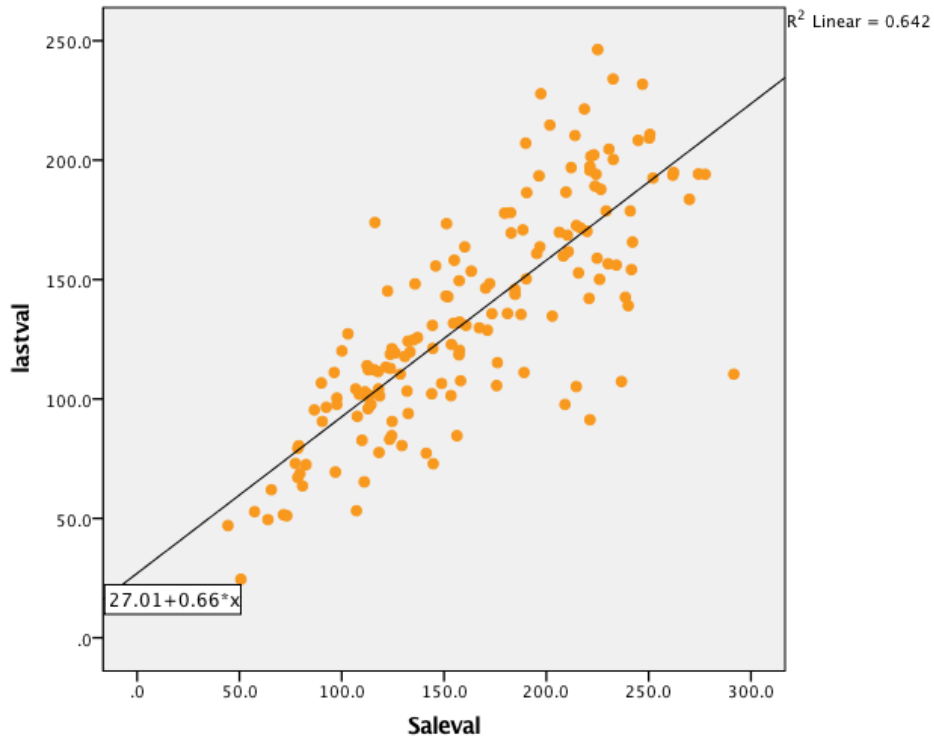
Township: Western





The following conclusions can be drawn on the sample with the help of the above three graphs. The last appraisal value has been plotted against each township.

1. The last appraisal placed the median eastern township at the highest valuation with a negatively skewed distribution whereas the median of the northern township was the least. The medians of the eastern, central and southern townships are above the average medians.
2. The curve of the central township is platykurtic, thus the central township has lesser fluctuations in the last appraisal value and the data is, more or less, structured. All the other townships have a mesokurtic curve.



Null Hypothesis, H₀:

The two variables are not significantly correlated.

Alternate Hypothesis, H₁:

The two variables are significantly correlated.

Level of significance:

$$\alpha = 0.05$$

Result:

		Saleval	lastval	time
Saleval	Pearson Correlation	1	.801**	.067
	Sig. (2-tailed)		.000	.397
	N	160	160	160
lastval	Pearson Correlation	.801**	1	-.315**
	Sig. (2-tailed)	.000		.000
	N	160	160	160
time	Pearson Correlation	.067	-.315**	1
	Sig. (2-tailed)	.397	.000	
	N	160	160	160

** . Correlation is significant at the 0.01 level (2-tailed).

Interpretation:

1. Sale value and last value are significantly correlated as H₀ gets rejected at 1% level of significance.
2. Sale value and time are not significantly correlated as H₀ gets accepted at 1% level of significance.
3. Last value and time are significantly correlated as H₀ gets rejected at 1% level of significance.

Hypothesis testing:

Test 1 (One-Way ANOVA):

Null Hypothesis, H₀:

There is no significant relationship between the effect due to township and Sale value.

Alternate Hypothesis, H₁:

There is a significant relationship between effect due to township and Sale value.

Level of significance,

$$\alpha = 0.05$$

Results

ANOVA

Saleval

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	300030.690	4	75007.672	54.605	.000
Within Groups	212913.554	155	1373.636		
Total	512944.244	159			

Interpretation:

Null Hypothesis is rejected since $p < \alpha$

This suggests a relationship indicating a strong significant association between Township and Sale value

Test 2 (Paired Sample t-test) :

We now wish to examine if increase in prices are significant or not, in order to do so we use a paired sample t-test.

Null Hypothesis, H₀:

$$\mu_i = \mu_j$$

Alternate Hypothesis, H₁:

$$\mu_i > \mu_j$$

Test Criteria:

Level of significance, $\alpha = 0.05$

Reject H₀ at α level of significance if $p < \alpha$

Result

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Eastern_saleval	224.737	19	25.4601	5.8410
	Eastern_lastval	188.605	19	22.6121	5.1876
Pair 2	Central_saleval	167.668	31	50.6868	9.1036
	Central_lastval	151.468	31	50.4797	9.0664
Pair 3	Southern_saleval	195.509	35	47.4120	8.0141
	Southern_lastval	146.800	35	32.4559	5.4860
Pair 4	Northern_saleval	100.641	32	23.7500	4.1984
	Nothern_lastval	95.016	32	23.9722	4.2377
Pair 5	Western_saleval	134.174	31	26.6539	4.7872
	Western_lastval	101.274	31	25.9543	4.6615

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	eastern_saleval & eastern_lastval	19	.190	.436
Pair 2	central_saleval & central_lastval	31	.927	.000
Pair 3	southern_saleval & southern_lastval	35	.389	.021
Pair 4	northern_saleval & nothern_lastval	32	.935	.000
Pair 5	western_saleval & western_lastval	31	.704	.000

A crucial aspect involved in a paired-sample t-test is a high level of co-relation between the two variables, all of the above the pairs show high significant co-relation except for the first one, hence we exclude the same from our further calculations.

Paired Samples Test

	Paired Differences					t	df	Sig. (1-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair 2 central_saleval - central_lastval	16.2000	19.3254	3.4709	9.1114	23.2886	4.667	30	.000
Pair 3 southern_saleval - southern_lastval	48.7086	45.8563	7.7511	32.9564	64.4608	6.284	34	.000
Pair 4 northern_saleval - northern_lastval	5.6250	8.5931	1.5191	2.5269	8.7231	3.703	31	.000
Pair 5 western_saleval - western_lastval	32.9000	20.2644	3.6396	25.4670	40.3330	9.039	30	.000

The output for the same is returned as a two-tailed test in SPSS, we convert the same into a one tailed test in the table above.

Interpretation:

Our results are significant for all the pairs, indicating that the housing prices have indeed gone up for the whole region.

Saleval	Equal variances assumed	9.127	.004	4.555	48	.000	57.069	12.5277	31.8804	82.257
	Equal variances not assumed			5.276	46.617	.000	57.069	10.8163	35.3048	78.833

The difference between the mean is significant as $p < \alpha$. We reject H_0 at α level of significance.

2. Results for Township Central and Southern:

Group Statistics

Township	N	Mean	Std. Deviation	Std. Error Mean
Central	31	167.668	50.6868	9.1036
Southern	43	204.051	44.2182	6.7432

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Saleval	Equal variances assumed	.825	.367	-3.284	72	.002	-36.3834	11.0790	-58.4689	-14.2979
	Equal variances not assumed			-3.212	59.21	.002	-36.3834	11.3290	-59.0510	-13.7158

The difference between the mean is significant as $p < \alpha$. We reject H_0 at α level of significance.

3. Results for Township Eastern and Southern:

Group Statistics

Township	N	Mean	Std. Deviation	Std. Error Mean
Eastern	19	224.737	25.4601	5.8410
Southern	43	204.051	44.2182	6.7432

Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Saleval Equal variances assumed	5.960	.018	1.899	60	.062	20.6857	10.8914	-1.1003	42.4717
Equal variances not assumed			2.319	55.615	.024	20.6857	8.9212	2.8117	38.5597

The difference between the mean is not significant as $p > \alpha$ when equal variance is assumed and is significant otherwise. We reject H_0 at α level of significance.

4. Results for Township Northern and Western:

Group Statistics

Township	N	Mean	Std. Deviation	Std. Error Mean
Northern Western	32	100.509	23.7025	4.1900
	35	133.949	26.9174	4.5499
Northern Western	32	95.016	23.9722	4.2377
	35	101.711	27.7574	4.6919

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Saleval	Equal variances assumed	.465	.498	-5.375	65	.000	-33.4392	6.2210	-45.8633	-21.0151

Equal variances not assumed			-5.406	64.915	.000	-33.4392	6.1853	-45.7924	-21.0860
--------------------------------	--	--	--------	--------	------	----------	--------	----------	----------

The difference between the mean is significant as $p < \alpha$. We reject H_0 at α level of significance.

Summary

Our analysis provides a deep insight into the nature and trend of the underlying prices of the property and the real estate market in general. Our conclusions provide actionable intelligence for concerned agents in the industry. Using the tools of hypothesis testing we see strong evidence for the following observations:

- The difference between the average prices over different regions as shown in our sample is significant hence housing prices in certain regions are higher than the other.
- The difference between the last evaluated prices and the current sale prices has shown to be significant throughout various regions indicating an upward trend in the prices of the housing market.