COST TREND ANALYSIS OF HOUSING PRICES IN MASSACHUSETTS (2008-16) USING MINITAB

[Jan 2017-Dec 2019]

Submitted by

Nishit shah

Rohit Appandaraju

Ajit Manuel

Under the guidance of

Prof.Rajesh Juglum

Table of Contents

I. ABSTRACT

Purchasing a house is one of the major investments an individual makes in his/her life. Understanding the criticality of the decision to purchase a house, it is very important to understand the factors affecting price variation so that we can get an idea of the optimum time to make a purchase. Considering this, our study is focused on the analyzing the root cause of housing price variation over some selected years. To analyze the cause, we selected a pool of housing data for some selected cities from a website [1]. The article focuses on studying the cost trend housing prices in Boston, Lowell and Waltham using various statistical tools like time – series plot, box-plots, and One-way ANOVA. Using the tools, we can get an overview of housing prices for all the three cities and predict the changing pattern for future considerations. We use the P-diagram to depict the control factors and the noise factors that affect the housing prices. In this analysis, we compare the means of housing prices in Boston, Lowell, and Waltham. The alternate hypothesis was accepted after doing testing using one-way ANOVA.

Keywords: Time-Series Analysis, One-Way ANOVA, Box-Plots, Housing Prices.

II. METHODOLOGY

The Analysis of data started with collecting data on Housing prices. It was possible to obtain large sets of data from the website www.zillow.com. The study was carried out by taking the data like monthly prices from 2008 to 2016 for 3 major cities Boston Waltham and Lowell.

We created a P-diagram prior to starting the analysis. It is used to analyze Inputs and Outputs and Noise Factors and Control Factors. Here the P- Diagram is used to identify the various sources of variation like recession globalization crime rate and natural disasters.

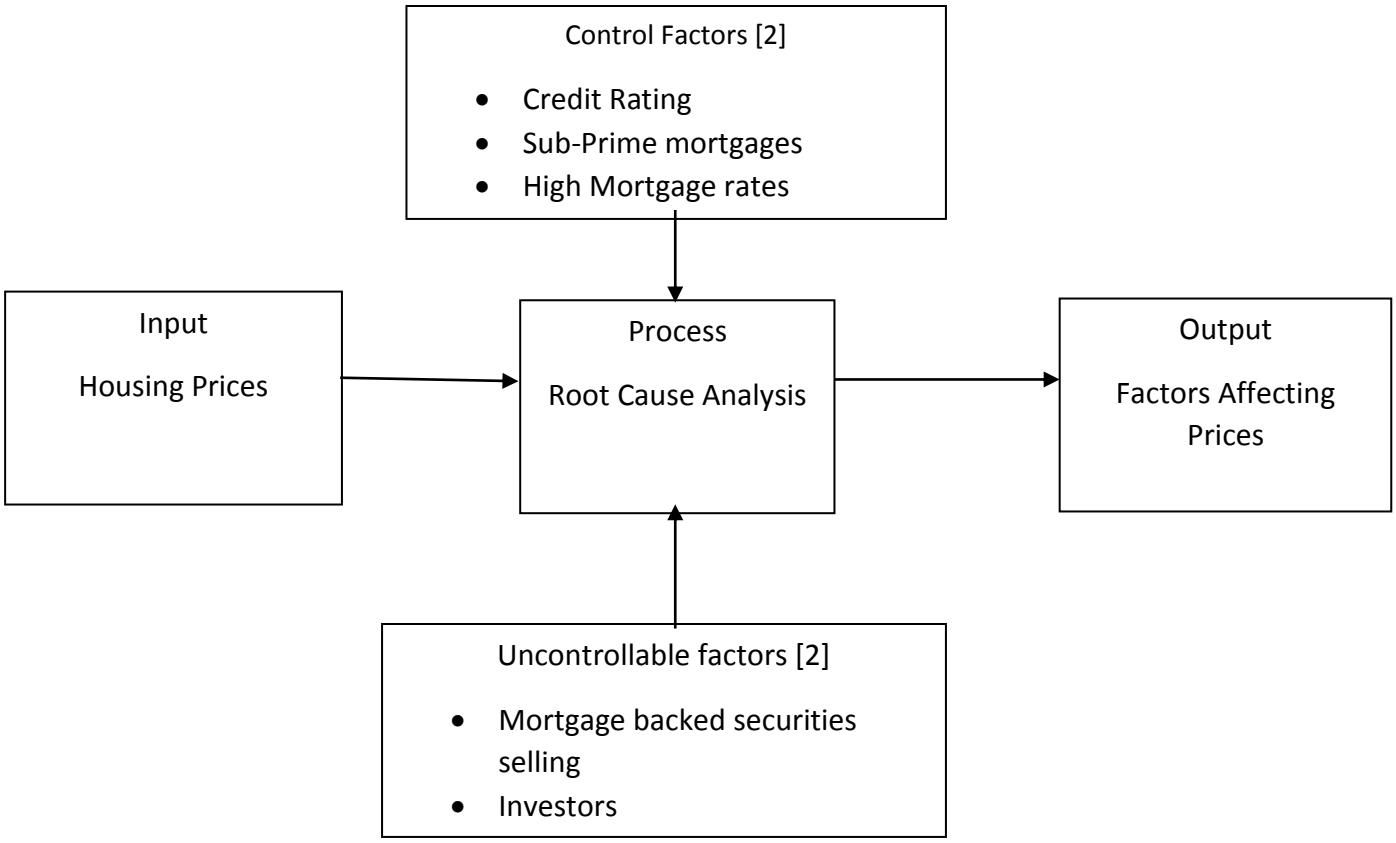
By using a time-series plot we can view the fluctuation in rental prices over the period of eight years. A time series is a set of data on a variable measured at successive time periods or over successive periods of time. The data collected in this study fits the definition of time series. Time series analysis comprises methods to discover a pattern in the historical data and then extrapolate the pattern into the future. Time series includes components like trend cyclical seasonal and irregular. A histogram is a graphical representation of the distribution of numerical data. Here the raw data is projected using a histogram and which is ultimately used to analyze and interpret the consistency of the data. The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means for eight years of 3 cities. In our analysis

Ho: $\mu1=\mu2=\mu3=\mu4=\mu5=\mu6=\mu7=\mu8$

Ha: at least one mean is different. (Where 1 23…8 are years respectively). Results were summarized in the form of tables. Graphs and suitable charts were used to present the data graphically so that interpretation could be better.

s

III. P-DIAGRAM



IV. DATA

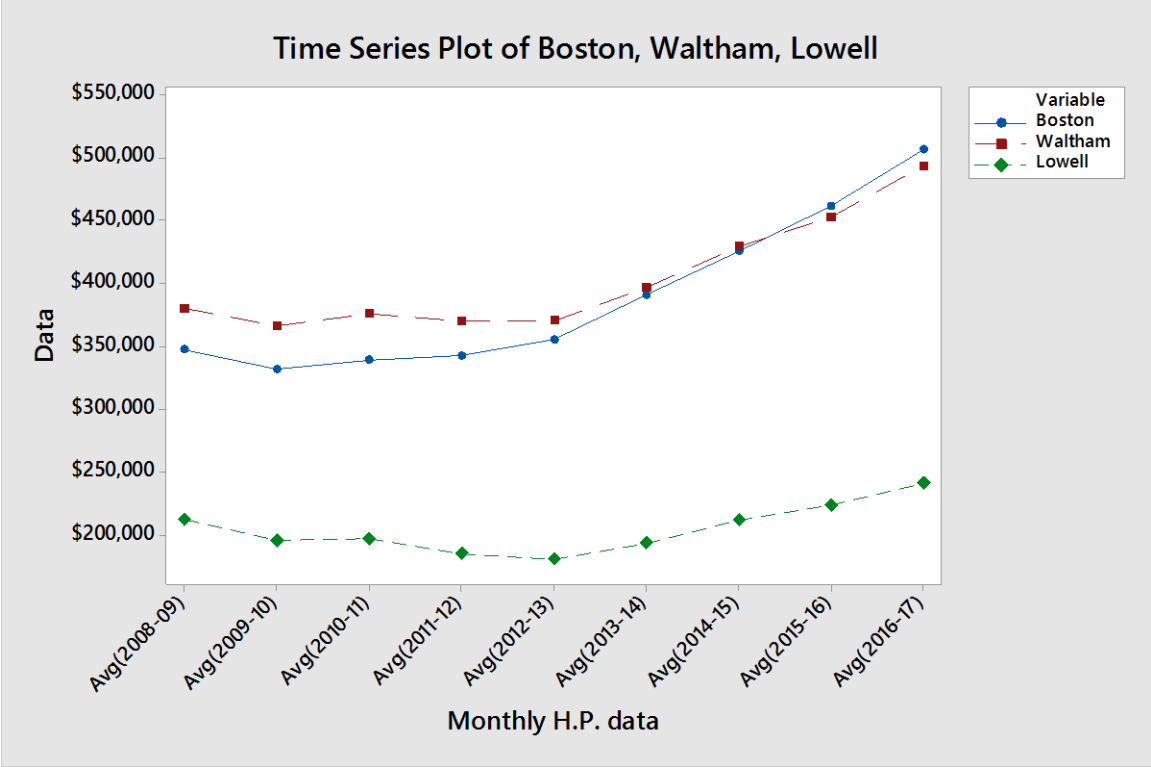
Monthly H.P. data	Boston	Waltham	Lowell
Jan 2008	$354,200	$394,800	$226,800
Feb 2008	$352,900	$392,500	$223,600
Mar 2008	$351,200	$389,600	$220,000
Apr 2008	$349,200	$385,000	$216,400
May 2008	$347,800	$379,800	$213,400
Jun 2008	$347,300	$375,500	$210,400
Jul 2008	$347,100	$373,900	$208,000
Aug 2008	$347,500	$374,500	$207,000
Sep 2008	$347,000	$375,100	$206,600
Oct 2008	$344,500	$374,500	$205,500
Nov 2008	$340,800	$372,600	$204,000
Dec 2008	$337,100	$369,800	$202,600
Avg(2008-09)	$347,217	$379,800	$212,025
Jan 2009	$333,000	$367,400	$201,700
Feb 2009	$328,700	$365,500	$201,100
Mar 2009	$325,300	$362,900	$199,800
Apr 2009	$323,400	$360,900	$197,700
May 2009	$324,400	$361,000	$195,500
Jun 2009	$328,100	$363,000	$193,800
Jul 2009	$332,100	$365,500	$192,600
Aug 2009	$335,200	$367,400	$191,900
Sep 2009	$337,200	$368,200	$191,700
Oct 2009	$337,900	$368,800	$191,800
Nov 2009	$337,500	$369,700	$191,900
Dec 2009	$336,900	$372,000	$193,900
Avg(2009-10)	$331,642	$366,025	$195,283
Jan 2010	$335,300	$373,900	$196,700
Feb 2010	$332,100	$373,700	$197,200
Mar 2010	$331,800	$373,300	$196,900
Apr 2010	$335,000	$374,700	$197,800
May 2010	$337,700	$376,700	$198,000
Jun 2010	$339,500	$378,400	$198,100

Jul 2010	$341,500	$378,300	$198,200
Aug 2010	$341,600	$377,000	$197,700
Sep 2010	$340,900	$376,500	$197,200
Oct 2010	$342,500	$376,500	$196,300
Nov 2010	$343,900	$375,700	$194,100
Dec 2010	$344,000	$375,000	$192,000
Avg(2010-11)	$338,817	$375,808	$196,683
Jan 2011	$344,900	$374,500	$190,100
Feb 2011	$346,800	$374,000	$188,600
Mar 2011	$346,700	$374,100	$188,000
Apr 2011	$345,800	$375,700	$188,300
May 2011	$344,900	$376,300	$188,800
Jun 2011	$343,600	$374,300	$188,200
Jul 2011	$341,600	$370,200	$186,300
Aug 2011	$339,300	$366,400	$183,600
Sep 2011	$337,300	$363,900	$180,600
Oct 2011	$336,500	$362,600	$178,900
Nov 2011	$338,600	$361,900	$178,400
Dec 2011	$342,700	$362,600	$177,900
Avg(2011-12)	$342,392	$369,708	$184,808
Jan 2012	$346,600	$364,200	$177,400
Feb 2012	$349,600	$364,900	$177,200
Mar 2012	$352,600	$365,500	$176,800
Apr 2012	$355,000	$366,400	$176,900
May 2012	$355,700	$367,600	$177,800
Jun 2012	$354,900	$368,700	$179,200
Jul 2012	$353,600	$370,300	$180,600
Aug 2012	$354,400	$371,600	$181,400
Sep 2012	$356,700	$372,700	$182,100
Oct 2012	$358,400	$373,900	$183,500
Nov 2012	$359,900	$376,600	$185,900
Dec 2012	$362,500	$378,900	$186,900
Avg(2012-13)	$354,992	$370,108	$180,475
Jan 2013	$366,800	$381,100	$187,100
Feb 2013	$373,100	$384,700	$187,700
Mar 2013	$378,800	$388,400	$188,100
Apr 2013	$382,700	$389,800	$188,000
May 2013	$386,100	$391,000	$188,800
Jun 2013	$389,600	$392,800	$190,800

Jul 2013	$393,200	$395,000	$192,900
Aug 2013	$397,500	$398,400	$195,100
Sep 2013	$402,000	$403,400	$197,500
Oct 2013	$405,100	$408,100	$199,300
Nov 2013	$406,100	$411,600	$200,400
Dec 2013	$406,200	$414,000	$201,200
Avg(2013-14)	$390,600	$396,525	$193,075
Jan 2014	$407,800	$417,000	$202,500
Feb 2014	$410,500	$421,400	$205,200
Mar 2014	$413,500	$425,000	$208,500
Apr 2014	$418,100	$427,600	$210,900
May 2014	$424,000	$430,400	$212,100
Jun 2014	$428,500	$432,200	$212,700
Jul 2014	$429,300	$432,800	$213,000
Aug 2014	$429,000	$433,400	$213,900
Sep 2014	$430,400	$432,300	$214,800
Oct 2014	$433,200	$431,100	$215,000
Nov 2014	$437,000	$432,000	$214,800
Dec 2014	$442,200	$433,400	$215,300
Avg(2014-15)	$425,292	$429,050	$211,558
Jan 2015	$445,700	$434,200	$215,600
Feb 2015	$447,900	$436,600	$216,000
Mar 2015	$450,700	$440,700	$217,500
Apr 2015	$452,900	$444,500	$219,300
May 2015	$454,700	$447,800	$220,900
Jun 2015	$458,100	$451,200	$221,900
Jul 2015	$461,300	$454,900	$223,300
Aug 2015	$463,900	$458,700	$225,600
Sep 2015	$467,400	$461,600	$227,600
Oct 2015	$472,600	$463,900	$229,000
Nov 2015	$478,400	$466,400	$231,100
Dec 2015	$483,400	$469,000	$232,900
Avg(2015-16)	$461,417	$452,458	$223,392
Jan 2016	$487,200	$471,400	$234,100
Feb 2016	$491,500	$473,000	$235,100
Mar 2016	$496,100	$475,500	$236,400
Apr 2016	$499,100	$479,600	$237,600
May 2016	$500,500	$484,000	$239,000
Jun 2016	$502,100	$491,700	$240,400

Jul 2016	$505,600	$500,200	$241,300
Aug 2016	$510,500	$504,400	$242,400
Sep 2016	$514,800	$506,000	$244,100
Oct 2016	$518,200	$508,800	$245,800
Nov 2016	$521,900	$510,500	$247,300
Dec 2016	$526,200	$513,400	$248,300
Avg(2016-17)	$506,142	$493,208	$240,983
Jan 2017	$531,200	$517,900	$249,400

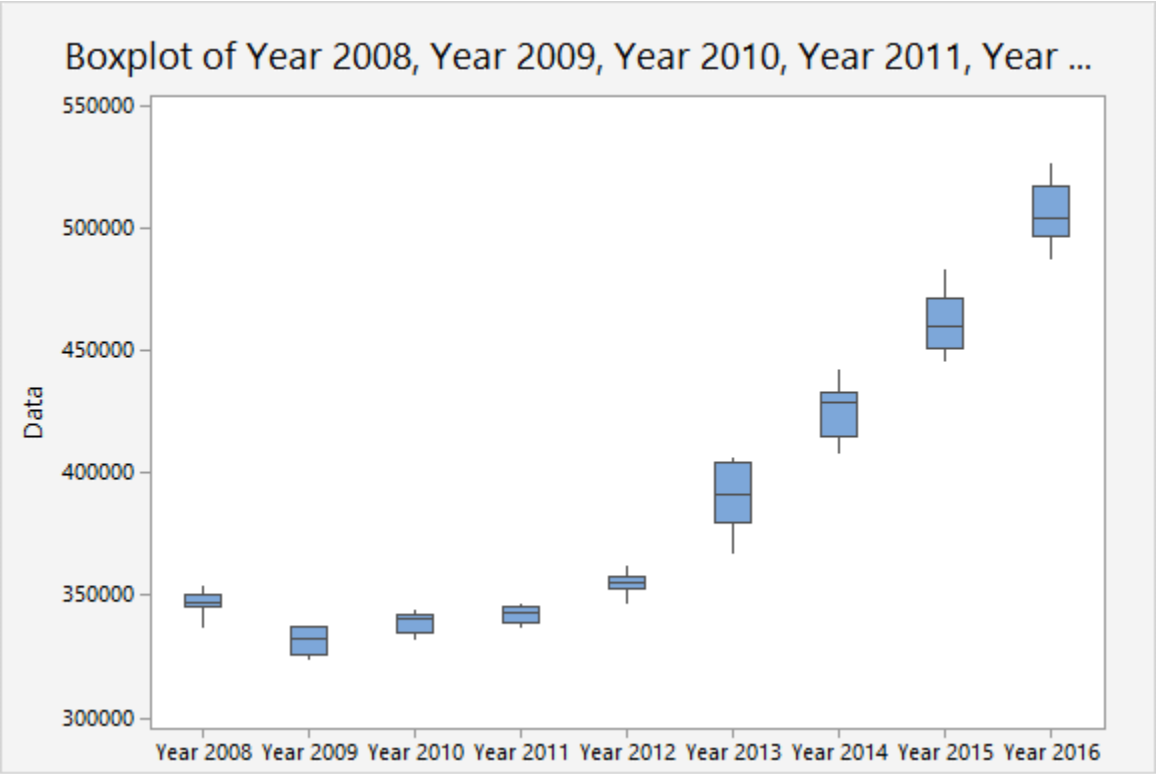
V. TIME SERIES



The time series plot depicted above shows average yearly housing price variation for Boston, Waltham and Lowell spread across 2008-09 to 2016-17.The y-axis represents the average of housing prices of cities and the x-axis represents time in years.

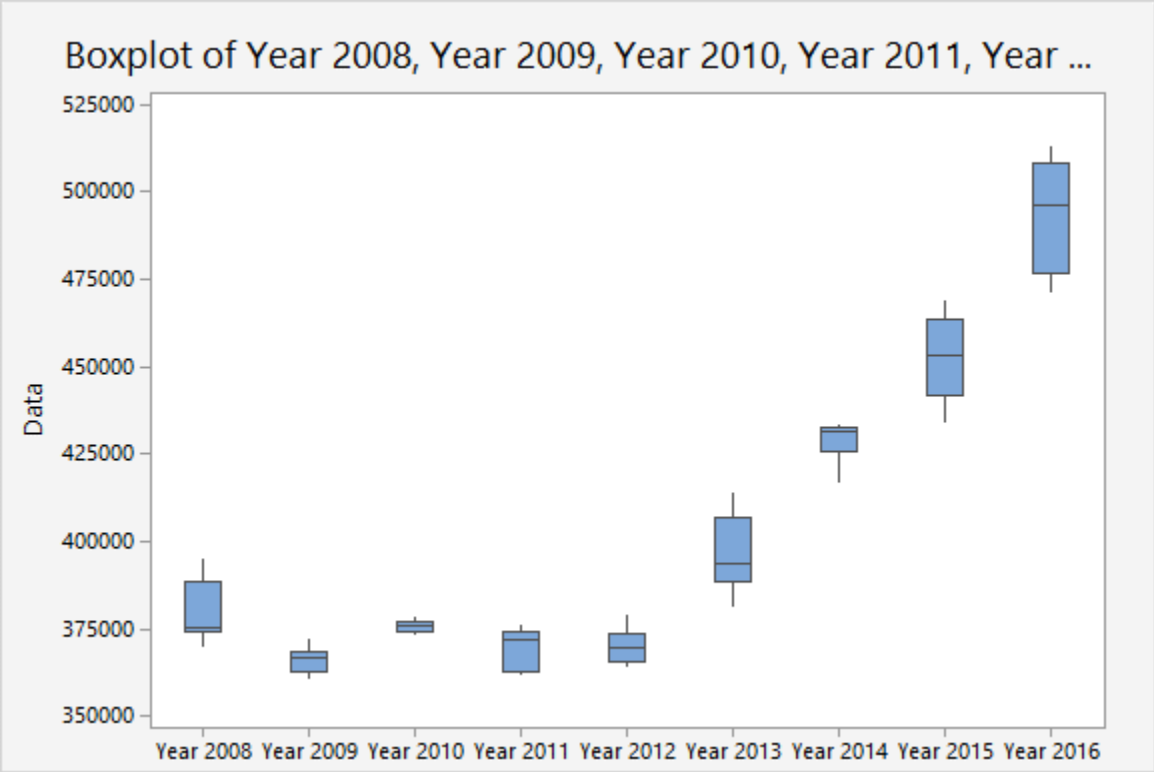
VI. BOX-PLOTS

BOSTON



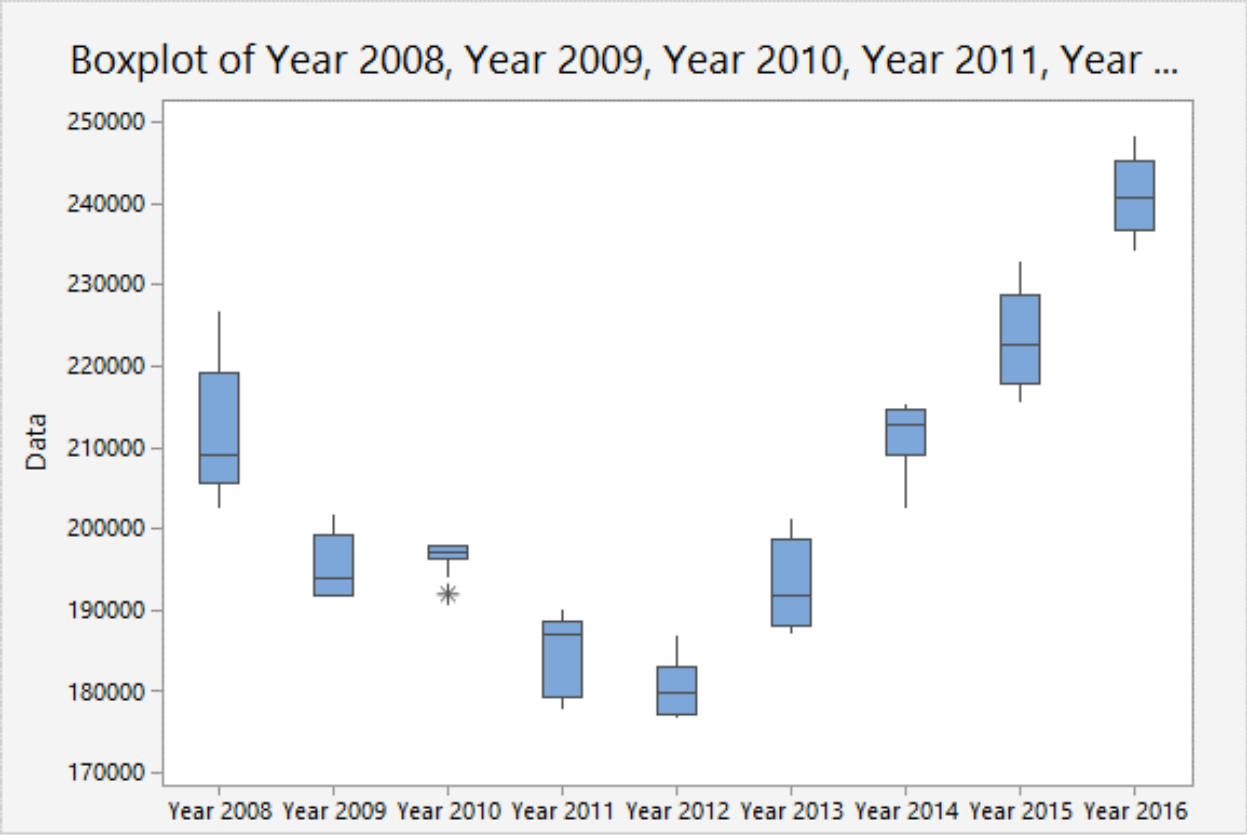
The box plot shown above helps to understand the upper limit and lower limit of the housing prices for Boston in years from 2008-2016. With the help of box plots, we can find the inter-quartile ranges for all the cities.

BOX PLOT: WALTHAM



The box plot shown above helps to understand the upper limit and lower limit of the housing prices for Waltham in years from 2008-2016. With the help of box plots, we can find the inter-quartile ranges for all the cities.

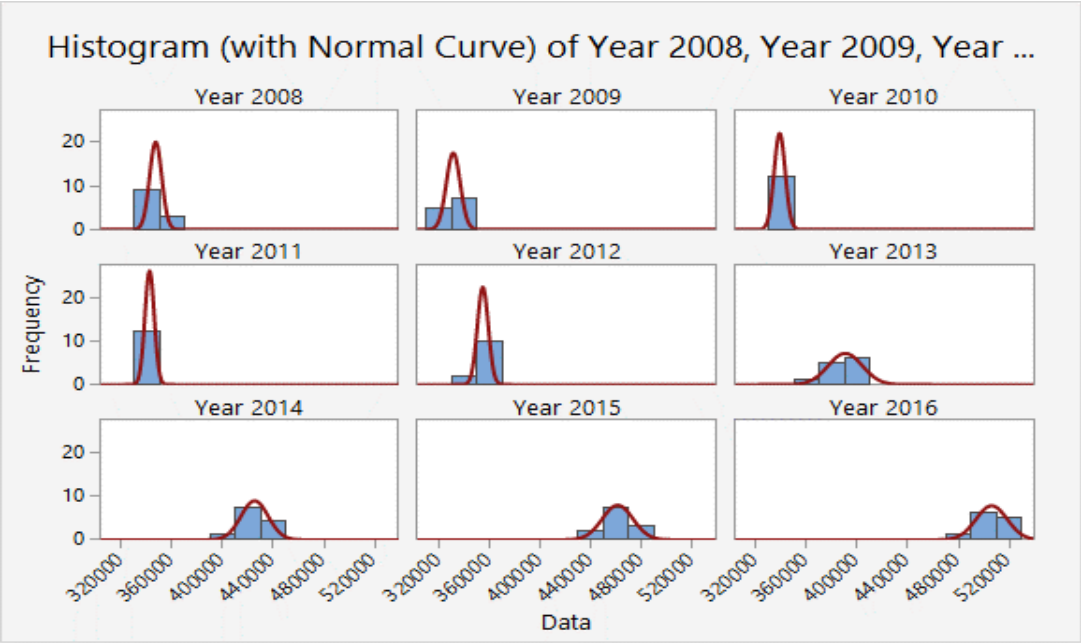
BOX PLOT: LOWELL



The box plot shown above helps to understand the upper limit and lower limit of the housing prices for Lowell in years from 2008-2016. With the help of box plots, we can find the inter-quartile ranges for all the cities. However, in this case we note that there is a presence of one outlier point in the year 2010.These points represent the prices that fall far outside the typically expected variation.

VII. GRAPHICAL SUMMARY

BOSTON



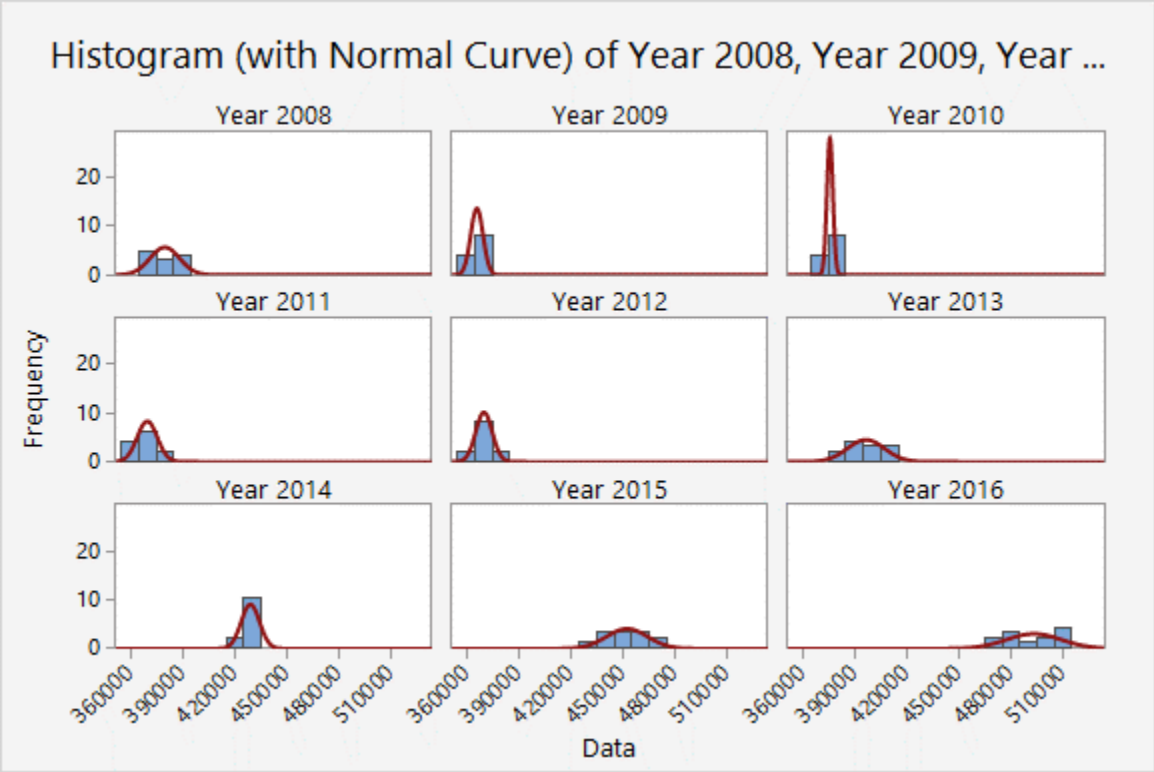
Descriptive Statistics: Year 2008, Year 2009, Year 2010, Year 2011, Year 2012, Year 2013, Year 2014, Year 2015, Year 2016

Statistics

Variable	N	N*	Mean	SE Mean	StDev	Variance	CoefVar	Minimum	Q1	Median	Q3	Maximum
Year 2008	12	0	347217	1382	4789	22932424	1.38	337100	345125	347400	350700	354200
Year 2009	12	0	331642	1578	5465	29866288	1.65	323400	326000	332550	337125	337900
Year 2010	12	0	338817	1257	4354	18959697	1.29	331800	335075	340200	342275	344000
Year 2011	12	0	342392	1062	3680	13540833	1.07	336500	338775	343150	345575	346800
Year 2012	12	0	354992	1240	4296	18455379	1.21	346600	352850	354950	357975	362500
Year 2013	12	0	390600	3867	13395	179434545	3.43	366800	379775	391400	404325	406200
Year 2014	12	0	425292	3100	10740	115355379	2.53	407800	414650	428750	432500	442200
Year 2015	12	0	461417	3495	12105	146541515	2.62	445700	451250	459700	471300	483400
Year 2016	12	0	506142	3548	12292	151097197	2.43	487200	496850	503850	517350	526200

The Graphical summary represents the histograms of housing prices over the span of 8 years for Boston. The Histograms for each of the eight years show the basic information about the data set related to central location, width of spread and shape. We cannot see considerable skewness in any of the histograms for the Housing prices. The plot for the years 2008-2016 shows the price statistics to be approximately normally distributed.

WALTHAM



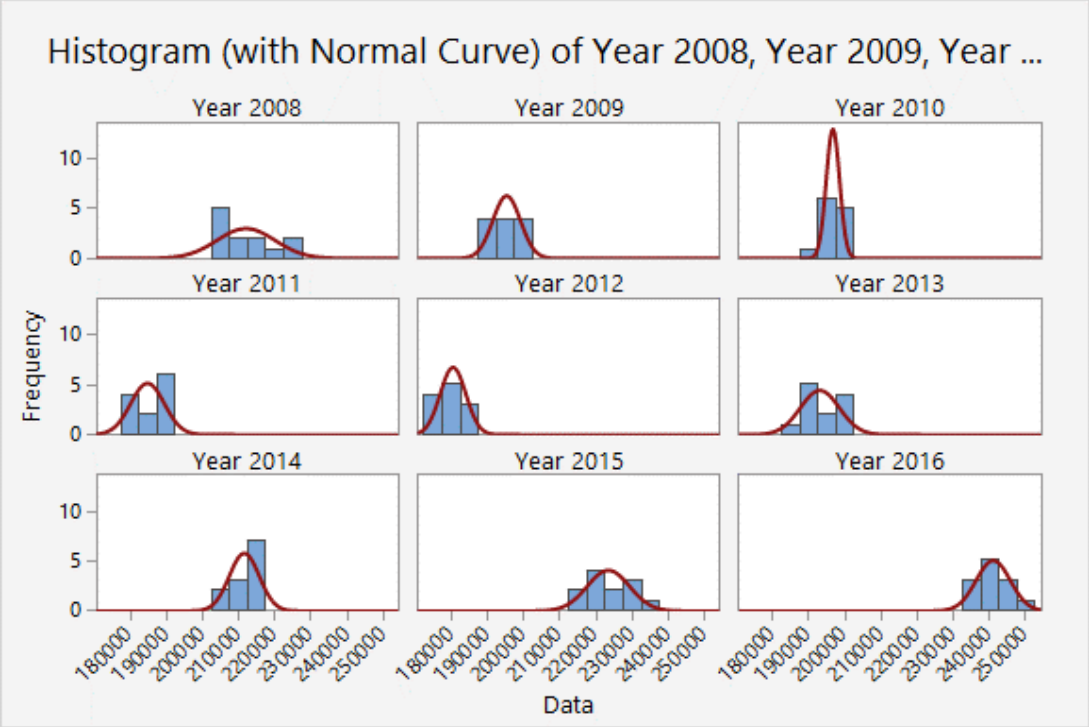
Descriptive Statistics: Year 2008, Year 2009, Year 2010, Year 2011, Year 2012, Year 2013, Year 2014, Year 2015, Year 2016

Statistics

Variable	N	N*	Mean	SE Mean	StDev	Variance	CoefVar	Minimum	Q1	Median	Q3	Maximum
Year 2008	12	0	379800	2451	8489	72070909	2.24	369800	374050	375300	388450	394800
Year 2009	12	0	366025	1017	3521	12400227	0.96	360900	362925	366450	368650	372000
Year 2010	12	0	375808	495.2	1715.4	2942652	0.46	373300	374100	376100	376925	378400
Year 2011	12	0	369708	1670	5784	33459015	1.56	361900	362925	372100	374450	376300
Year 2012	12	0	370108	1371	4749	22553561	1.28	364200	365725	369500	373600	378900
Year 2013	12	0	396525	3081	10673	113920227	2.69	381100	388750	393900	406925	414000
Year 2014	12	0	429050	1531	5303	28122727	1.24	417000	425650	431550	432675	433400
Year 2015	12	0	452458	3412	11821	139739015	2.61	434200	441650	453050	463325	469000
Year 2016	12	0	493208	4570	15833	250671742	3.21	471400	476525	495950	508100	513400

The Graphical summary represents the histograms of housing prices over the span of 8 years for Waltham city. The Histograms for each of the eight years show the basic information about the data set related to central location, width of spread and shape. We cannot see considerable skewness in any of the histograms for the Housing prices. The plot for the years 2008-2016 shows the price statistics to be approximately normally distributed.

LOWELL



Descriptive Statistics: Year 2008, Year 2009, Year 2010, Year 2011, Year 2012, Year 2013, Year 2014, Year 2015, Year 2016

Statistics

Variable	N	N*	Mean	SE Mean	StDev	Variance	CoefVar	Minimum	Q1	Median	Q3	Maximum
Year 2008	12	0	212025	2313	8013	64203864	3.78	202600	205775	209200	219100	226800
Year 2009	12	0	195283	1103	3820	14588788	1.96	191700	191900	193850	199275	201700
Year 2010	12	0	196683	534.1	1850.2	3423333	0.94	192000	196400	197200	197950	198200
Year 2011	12	0	184808	1341	4645	21571742	2.51	177900	179325	187150	188525	190100
Year 2012	12	0	180475	1025	3550	12602045	1.97	176800	177250	179900	183150	186900
Year 2013	12	0	193075	1559	5399	29152955	2.80	187100	188025	191850	198850	201200
Year 2014	12	0	211558	1194	4137	17117197	1.96	202500	209100	212850	214800	215300
Year 2015	12	0	223392	1695	5871	34464470	2.63	215600	217950	222600	228650	232900
Year 2016	12	0	240983	1368	4740	22470606	1.97	234100	236700	240850	245375	248300

The Graphical summary represents the histograms of housing prices over the span of 8 years for Lowell city. The Histograms for each of the eight years show the basic information about the data set related to central location, width of spread and shape. We cannot see considerable skewness in any of the histograms for the Housing prices. The plot for the years 2008-2016 shows the price statistics to be approximately normally distributed.

VIII. ANALYSIS OF VARIANCE (ANOVA)

BOSTON

One-Way ANOVA: Year 2008, Year 2009, Year 2010, Year 2011, Year 2012, Year 2013, Year 2014, Year 2015, Year 2016

Method

Null hypothesis H_0: All means are equal
Alternative hypothesis H_1: At least one mean is different

Equal variances were assumed for the analysis.

Factor Information

Factor	Levels	Values
Factor	9	Year 2008, Year 2009, Year 2010, Year 2011, Year 2012, Year 2013, Year 2014, Year 2015, Year 2016

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	8	374020736296	46752592037	604.40	<0.0001
Error	99	7658015833	77353695		
Total	107	381678752130			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
8795.09	97.99%	97.83%	97.61%

Means

Factor	N	Mean	StDev	95% CI
Year 2008	12	347217	4789	(342179, 352254)
Year 2009	12	331642	5465	(326604, 336679)
Year 2010	12	338817	4354	(333779, 343854)
Year 2011	12	342392	3680	(337354, 347429)
Year 2012	12	354992	4296	(349954, 360029)
Year 2013	12	390600	13395	(385562, 395638)
Year 2014	12	425292	10740	(420254, 430329)
Year 2015	12	461417	12105	(456379, 466454)
Year 2016	12	506142	12292	(501104, 511179)

Pooled StDev = 8795.09

One-Way ANOVA: WALTHAM

One-Way ANOVA: Year 2008, Year 2009, Year 2010, Year 2011, Year 2012, Year 2013, Year 2014, Year 2015, Year 2016

Method

Null hypothesis	H_0: All means are equal
Alternative hypothesis	H_1: At least one mean is different

Equal variances were assumed for the analysis.

Factor Information

Factor	Levels	Values
Factor	9	Year 2008, Year 2009, Year 2010, Year 2011, Year 2012, Year 2013, Year 2014, Year 2015, Year 2016

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	8	193626875741	24203359468	322.29	<0.0001
Error	99	7434680833	75097786		
Total	107	201061556574			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
8665.90	96.30%	96.00%	95.60%

Means

Factor	N	Mean	StDev	95% CI
Year 2008	12	379800	8489	(374836, 384764)
Year 2009	12	366025	3521	(361061, 370989)
Year 2010	12	375808	1715.4	(370845, 380772)
Year 2011	12	369708	5784	(364745, 374672)
Year 2012	12	370108	4749	(365145, 375072)
Year 2013	12	396525	10673	(391561, 401489)
Year 2014	12	429050	5303	(424086, 434014)
Year 2015	12	452458	11821	(447495, 457422)
Year 2016	12	493208	15833	(488245, 498172)

Pooled StDev = 8665.90

ONE-WAY ANOVA: LOWELL

One-Way ANOVA: Year 2008, Year 2009, Year 2010, Year 2011, Year 2012, Year 2013, Year 2014, Year 2015, Year 2016

Method

Null hypothesis	H_0: All means are equal
Alternative hypothesis	H_1: At least one mean is different

Equal variances were assumed for the analysis.

Factor Information

Factor	Levels	Values
Factor	9	Year 2008, Year 2009, Year 2010, Year 2011, Year 2012, Year 2013, Year 2014, Year 2015, Year 2016

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	8	36424423519	4553052940	186.60	<0.0001
Error	99	2415545000	24399444		
Total	107	38839968519			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
4939.58	93.78%	93.28%	92.60%

Means

Factor	N	Mean	StDev	95% CI
Year 2008	12	212025	8013	(209196, 214854)
Year 2009	12	195283	3820	(192454, 198113)
Year 2010	12	196683	1850.2	(193854, 199513)
Year 2011	12	184808	4645	(181979, 187638)
Year 2012	12	180475	3550	(177646, 183304)
Year 2013	12	193075	5399	(190246, 195904)
Year 2014	12	211558	4137	(208729, 214388)
Year 2015	12	223392	5871	(220562, 226221)
Year 2016	12	240983	4740	(238154, 243813)

Pooled StDev = 4939.58

IX. RESULTS AND DISCUSSION

- ANOVA was conducted to evaluate the significance of factors on the variation of Housing prices. After conducting the One-Way ANOVA for the Housing prices for Boston using Minitab, we find that the p value is very small, almost approximating to 0. Since $0<0.05$ (α value), we reject the null hypothesis and accept the alternate hypothesis that states that at least one mean is different.

- Similarly, as per our findings from the results of the One-Way ANOVA test for the Housing prices for Waltham and Lowell, we find that the p value is very small, approximating to 0. Since $0<0.05$ (α value), we reject the null hypothesis and accept the alternate hypothesis that states that at least one mean is different.

- Influence of the Control factors on the Housing prices were prominently noticed during years 2008 to 2011. This was because of Financial crisis due to banks all over United States

- After year 2011, we see increase in the prices which is comparable to the yearly mean. But in Boston, there is a greater increase in prices than Waltham and Lowell. This is because of increase in GDP for Boston is high compare to that of other cities

X. SCOPE OF WORK

The study of the variation of the Housing prices over the 8 years can be of substantial use to the Buyers, Investors and Banks. The statistics and graphs show the trend of variation from the mean price over the considered time span. These variations result from both the control factors as well as uncontrollable factors, and housing price being a common problem depends on lot of uncontrollable factors which result in variation. Going by the trend, banks need to careful while lending loans to people buying houses. Investors also must be cautious and should not blindly believe in rating agencies. Buyers must also understand the value of interest rate and buy accordingly. Too much price fluctuations affects the housing prices as it results in dispute between buyers and sellers. Thus, the analysis of the variation of Housing prices provides a basis for evaluating the trend of economics of the housing in Massachusetts.

XI.LEARNING OUTCOMES

- How to retrieve and store useful data from different databases.
- How to manipulate the data according to the user's requirement.
- How to analyze the data using MINITAB and ANOVA to do root cause analysis of a problem.
- How to do hypothesis testing practically using confidence intervals on large data sets.
- How to make decisions using charts and graph
- Perform Graphical summaries of data sets to get a brief overview of the data variation.
- Use of statistical tools like MINITAB to study time-series plot, histograms, normal distribution in practice.

XII.REFERENCES

[1] www.zillow.com

[2] https://www.youtube.com/watch?v=GPOv72Awo68