Syracuse University

IST-718 Lab 2

ThulasiRam RuppaKrishnan

IST 718

Professor Jillian Lando

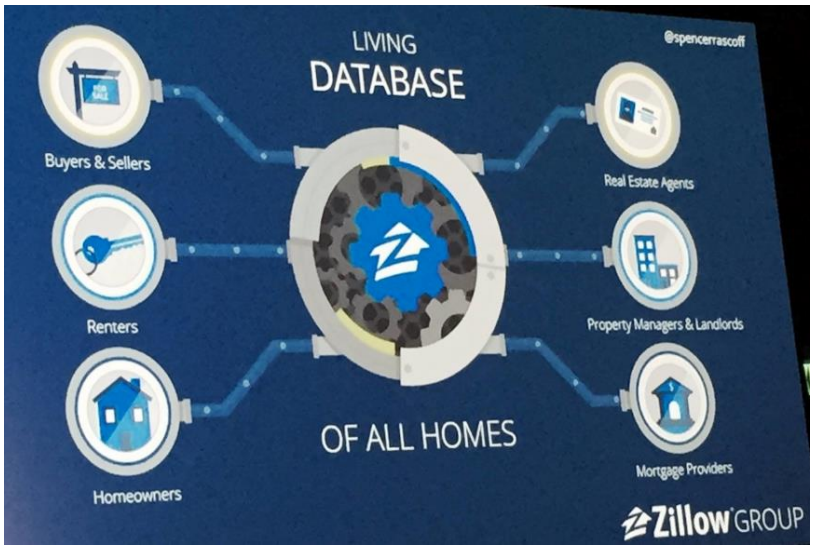
Contents

Introduction

Real estate appraisal, property valuation or land valuation is the process of developing an opinion of value, for real property (usually market value). Real estate transactions often require appraisals because they occur infrequently and every property is unique (especially their condition, a key factor in valuation), unlike corporate stocks, which are traded daily and are identical (thus a centralized Walrasian auction like a stock exchange is unrealistic). The location also plays a key role in valuation. However, since property cannot change location, it is often the upgrades or improvements to the home that can change its value. Appraisal reports form the basis for mortgage loans, settling estates and divorces, taxation, and so on. Sometimes an appraisal report is used to establish a sale price for a property.

Besides the mandatory educational grade, which can vary from Finance to Construction Technology, most, but not all, countries require appraisers to have the license for the practice. Usually, the real estate appraiser has the opportunity to reach 3 levels of certification: Appraisal Trainee, Licensed Appraiser and Certified Appraiser. The second and third levels of license require no less than 2000 experience hours in 12 months and 2500 experience hours in no less than 24 months respectively. Appraisers are often known as "property valuers" or "land valuers"; in British English they are "valuation surveyors". If the appraiser's opinion is based on market value, then it must also be based on the highest and best use of the real property. In the United States, mortgage valuations of improved residential properties are generally reported on a standardized form like the Uniform Residential Appraisal Report. Appraisals of more commercial properties (e.g., income-producing, raw land) are often reported in narrative format and completed by a Certified General Appraiser.

Zillow Group, Inc., or simply Zillow, is an American online real estate database company that was founded in 2006



Zillow has data on 110 million homes across the United States, not just those homes currently for sale. In addition to giving value estimates of homes, it offers several features including value changes of each home in a given time frame (such as one, five, or 10 years), aerial views of homes, and prices of comparable homes in the area. It can access appropriate public data, it also provides basic information on a given home, such as square footage and the number of bedrooms and bathrooms. Users can also get current estimates of homes if there was a significant change made, such as a recently remodeled kitchen. Zillow provides an application programming interface (API) and developer support network.

Analysis and Models

About the data

(files.zillowstatic.com/research/public/Zip/Zip_Zhvi_SingleFamilyResidence.csv)

The dataset (Zip_Zhvi_SingleFamilyResidence.csv) is a comma-separated files with the following variables

RegionID	Unique Identifier for a Region
RegionName	Region identified by zip code
City	City identified by zip code
State	State identified by zip code
Metro	Metro identified by zip code
CountyName	County identified by zip code
SizeRank	Size identified based on population
1996-04	Zhvi by zip code and year month
1996-05	Zhvi by zip code and year month
1996-06	Zhvi by zip code and year month
. . .	Zhvi by zip code and year month
2019-08	Zhvi by zip code and year month
2019-09	Zhvi by zip code and year month

Home Values (ZHVI)

Zillow Home Value Index (ZHVI): A smoothed, seasonally adjusted measure of the median estimated home value across a given region and housing type. It is a dollar-denominated alternative to repeat-sales indices. Zillow also publishes home value and other housing data for local markets, as well as a more detailed methodology and a comparison of ZHVI to the S&P CoreLogic Case-Shiller Home Price Indices.

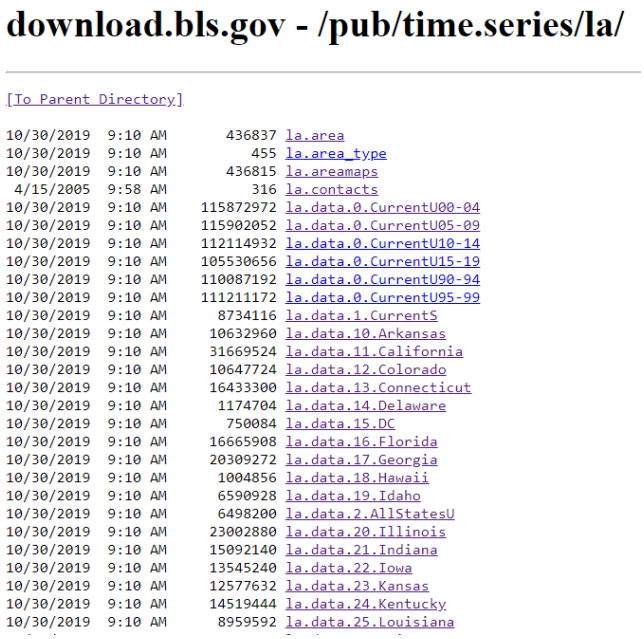
Table 1.1 given below shows the sample dataset from Zip_Zhvi_SingleFamilyResidence.csv.

RegionID	RegionNar	City	State	Metro	CountyNar	SizeRank	1996-04	1996-05	1996-06	1996-07	1996-08	1996-09	1996-10	1996-11	1996-12	1997-01	1997-02	1997-03	1997-04	1997-05	1997-06
84654	60657	Chicago	IL	Chicago-N	Cook Cour	1	337200	338200	339000	339700	340400	341000	341600	342300	343400	344900	346200	347000	347900	349100	350400
91982	77494	Katy	TX	Houston-T	Harris Cou	2	210400	212200	212200	210700	208300	205500	202500	199800	198300	197300	195400	193000	191800	191800	193000
84616	60614	Chicago	IL	Chicago-N	Cook Cour	3	502900	504900	506300	507200	507400	507000	506100	504700	503700	503200	501900	499500	497500	495900	494800
91940	77449	Katy	TX	Houston-T	Harris Cou	4	95400	95600	95800	96100	96400	96700	96800	96800	96700	96600	96400	96200	96100	96200	96300
91733	77084	Houston	TX	Houston-T	Harris Cou	5	95000	95200	95400	95700	95900	96100	96200	96100	96000	95800	95500	95300	95100	95100	95200
93144	79936	El Paso	TX	El Paso	El Paso Co	6	77300	77300	77300	77300	77400	77500	77600	77700	77700	77800	77900	77900	77800	77800	77800
84640	60640	Chicago	IL	Chicago-N	Cook Cour	7	218500	218500	218500	218400	218300	218200	218300	218500	219100	220100	221100	222000	223100	224300	225600
62037	11226	New York	NY	New York-	Kings Cour	8	161800	162200	162500	162900	163300	163900	164700	165700	166800	168000	168900	169700	170400	171100	171900
61807	10467	New York	NY	New York-	Bronx Cou	9	151900	151800	151800	151800	151700	151500	151300	151200	151300	151500	151600	151700	151800	152000	152200
92593	78660	Pflugervill	TX	Austin-Ro	Travis Cou	10	138900	138600	138400	138500	138700	139000	139300	139600	139900	140200	140600	141300	141800	142200	142400

Table 1.1 Sample Zip_Zhvi_SingleFamilyResidence dataset

Additional dataset

Unemployment rate from Bureau of Labor Statistics and Census data



download.bls.gov - /pub/time.series/la/

[To Parent Directory]

10/30/2019	9:10 AM	436837	la.area
10/30/2019	9:10 AM	455	la.area_type
10/30/2019	9:10 AM	436815	la.areamaps
4/15/2005	9:58 AM	316	la.contacts
10/30/2019	9:10 AM	115872972	la.data.0.CurrentU00-04
10/30/2019	9:10 AM	115902052	la.data.0.CurrentU05-09
10/30/2019	9:10 AM	112114932	la.data.0.CurrentU10-14
10/30/2019	9:10 AM	105530656	la.data.0.CurrentU15-19
10/30/2019	9:10 AM	110087192	la.data.0.CurrentU90-94
10/30/2019	9:10 AM	111211172	la.data.0.CurrentU95-99
10/30/2019	9:10 AM	8734116	la.data.1.CurrentS
10/30/2019	9:10 AM	10632960	la.data.10.Arkansas
10/30/2019	9:10 AM	31669524	la.data.11.California
10/30/2019	9:10 AM	10647724	la.data.12.Colorado
10/30/2019	9:10 AM	16433300	la.data.13.Connecticut
10/30/2019	9:10 AM	1174704	la.data.14.Delaware
10/30/2019	9:10 AM	750084	la.data.15.DC
10/30/2019	9:10 AM	16665908	la.data.16.Florida
10/30/2019	9:10 AM	20309272	la.data.17.Georgia
10/30/2019	9:10 AM	1004856	la.data.18.Hawaii
10/30/2019	9:10 AM	6590928	la.data.19.Idaho
10/30/2019	9:10 AM	6498200	la.data.2.AllStatesU
10/30/2019	9:10 AM	23002880	la.data.20.Illinois
10/30/2019	9:10 AM	15092140	la.data.21.Indiana
10/30/2019	9:10 AM	13545240	la.data.22.Iowa
10/30/2019	9:10 AM	12577632	la.data.23.Kansas
10/30/2019	9:10 AM	14519444	la.data.24.Kentucky
10/30/2019	9:10 AM	8959592	la.data.25.Louisiana

This dataset contains the following information

series_id	Unique identifier for a metric published by bureau LASST100000000000003 represents Unemployment Rate by state where the two digits (10) followed by "LASST" represents the identifier for state (10 → Delaware)
year	year when the metrics value is recorded
period	identifier for month when the metric value is recorded
value	metric value where the metric in this case is unemployment rate
footnote_codes	notes/comments if applicable

Series data file description

series_id	area_type_code	area_code	measure_code	seasonal	srd_code	series_title	footnote_codes	begin_year	begin_period	end_year	end_period
LASST100000000000003	A	ST1000000000000	03	S	10	Unemployment Rate: Delaware (S)		1976	M01	2019	M09

State file layout

srd_code	srd_text
01	Alabama
02	Alaska
04	Arizona
05	Arkansas
06	California
08	Colorado
09	Connecticut
10	Delaware
11	District of Columbia
12	Florida
13	Georgia
15	Hawaii
16	Idaho
17	Illinois
18	Indiana
19	Iowa

Table 1.2 given below shows the sample dataset from `la.data.10.Arkansas`.

series_id	year	period	value	footnote_codes
LASST050000000000003	1976	M01	7.4	
LASST050000000000003	1976	M02	7.4	
LASST050000000000003	1976	M03	7.4	
LASST050000000000003	1976	M04	7.3	
LASST050000000000003	1976	M05	7.2	
LASST050000000000003	1976	M06	7.0	
LASST050000000000003	1976	M07	6.9	
LASST050000000000003	1976	M08	6.7	
LASST050000000000003	1976	M09	6.6	
LASST050000000000003	1976	M10	6.6	
LASST050000000000003	1976	M11	6.6	
LASST050000000000003	1976	M12	6.6	
LASST050000000000003	1977	M01	6.6	
LASST050000000000003	1977	M02	6.6	
LASST050000000000003	1977	M03	6.6	
LASST050000000000003	1977	M04	6.5	
LASST050000000000003	1977	M05	6.5	
LASST050000000000003	1977	M06	6.4	
LASST050000000000003	1977	M07	6.4	
LASST050000000000003	1977	M08	6.4	

Table 1.2 Sample la.data.10.Arkansas. dataset

Answer the following business question using data from Zillow:

The research question is can we predict which three zip codes provide the best investment opportunity for the Syracuse Real Estate Investment Trust (SREIT)?

• Using the base data available from Zillow

(files.zillowstatic.com/research/public/Zip/Zip_Zhvi_SingleFamilyResidence.csv)

o Review the data – clean as appropriate

o Provide an initial data analysis to include (but not limited to):

▪ Develop time series plots for the following Arkansas metro areas:

• Hot Springs, Little Rock, Fayetteville, Searcy

• Present all values from 1997 to present

• **Average at the metro area level**

o Develop model(s) for forecasting average median housing value by zip code for 2018

o Use the historical data from 1997 through 2017 as your training data

o Integrate data from other sources (Bureau of Labor Statistics and Census data) to improve upon your base model(s)

• Answer the following questions:

o What technique/algorithm/decision process did you use to down sample?

o What three zip codes provide the best investment opportunity for the SREIT?

o Why?

After cleaning the dataset, exploratory data analysis is performed on top of this dataset to study each variable and its interaction with one another

Figure 1.1: This gives the distribution of median housing value for Hot Springs, Little Rock, Fayetteville, Searcy and its trend

Figure 1.2: Shows the boxplot and annual trend compared to each other to see the variations and any seasonal input for Fayetteville

Figure 1.3: Shows the boxplot and annual trend compared to each other to see the variations and any seasonal input for Hot Springs

Figure 1.4: Shows the boxplot and annual trend compared to each other to see the variations and any seasonal input for Little Rock

Figure 1.5: Shows the boxplot and annual trend compared to each other to see the variations and any seasonal input for Searcy

Figure 1.6: Shows the stationary observation of **Hot** Springs, Little Rock, Fayetteville, Searcy

Figure 1.7: Shows the autocorrelation and partial auto correlation details on **Hot** Springs, Little Rock, Fayetteville, Searcy

Figure 1.8: This gives the correlation matrix between all the dependent variable

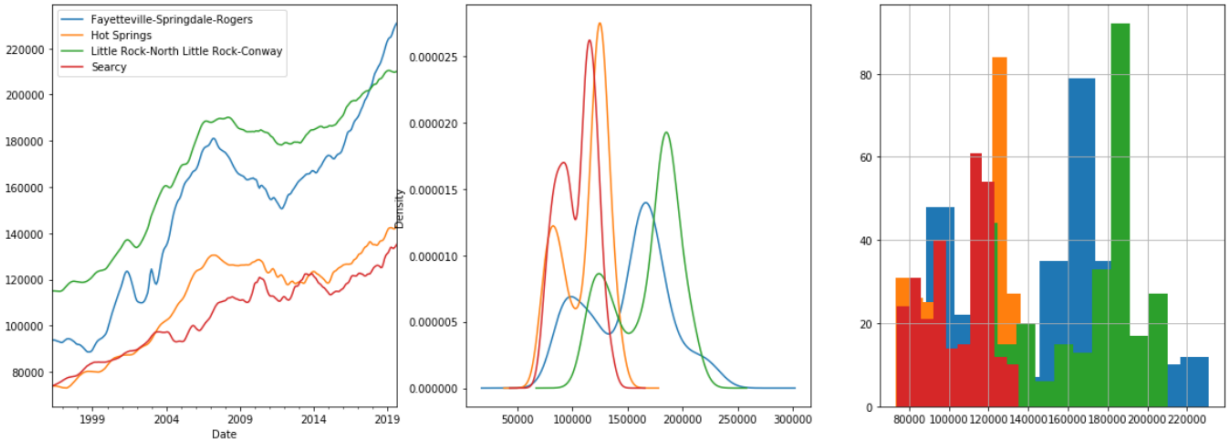


Figure 1.1

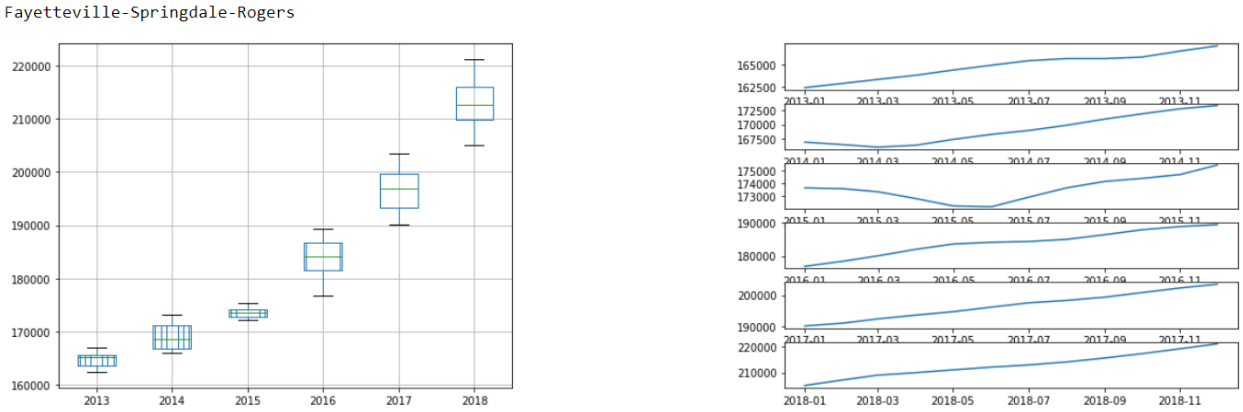


Figure 1.2

Hot Springs

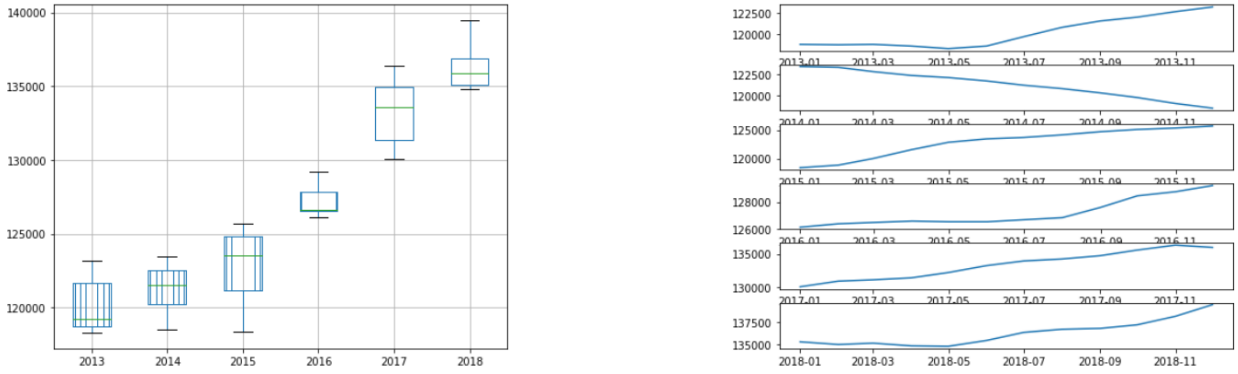


Figure 1.3

Little Rock-North Little Rock-Conway

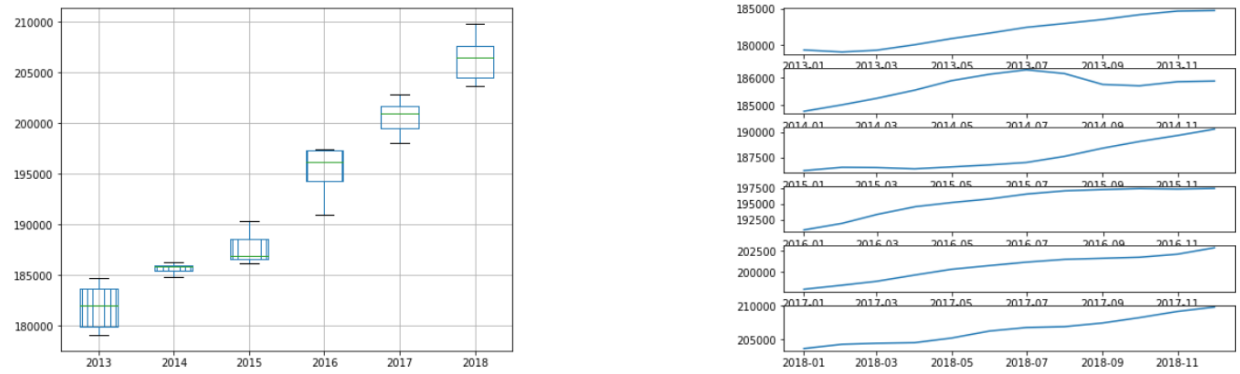


Figure 1.4

Searcy

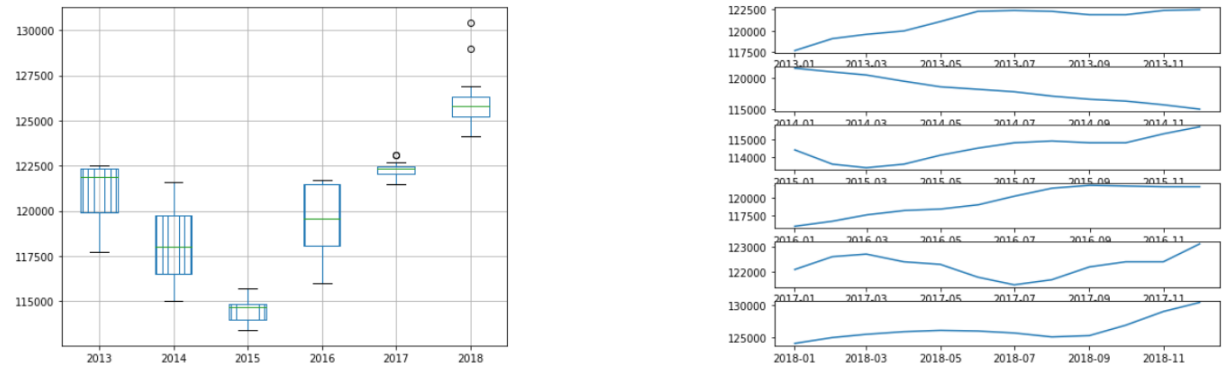


Figure 1.5

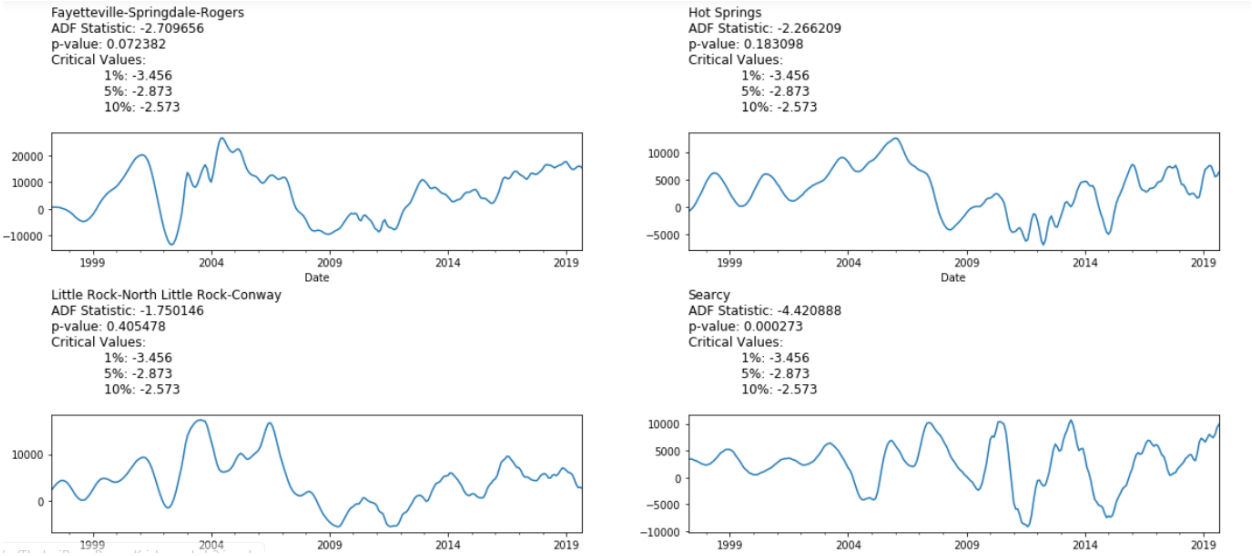


Figure 1.6

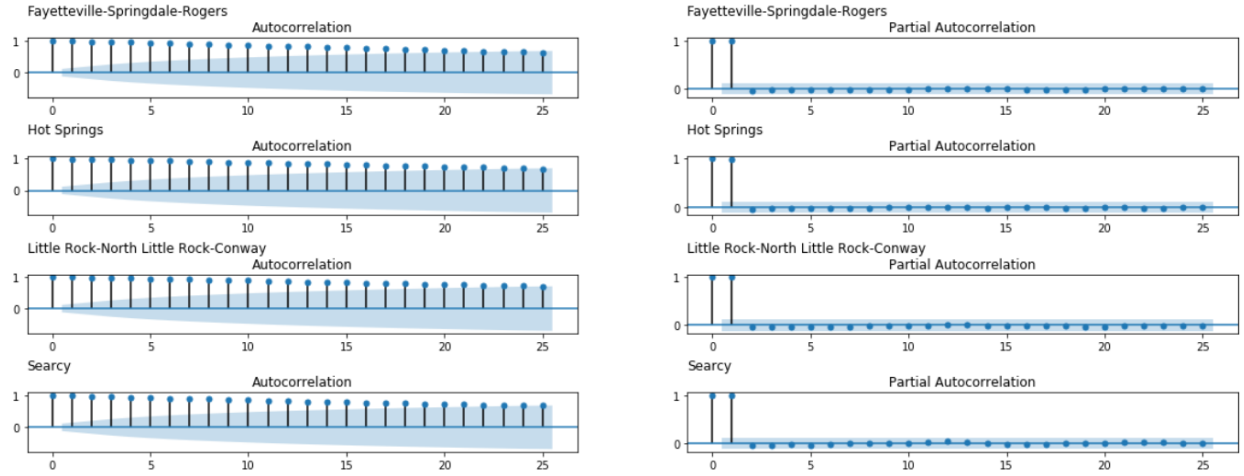


Figure 1.7

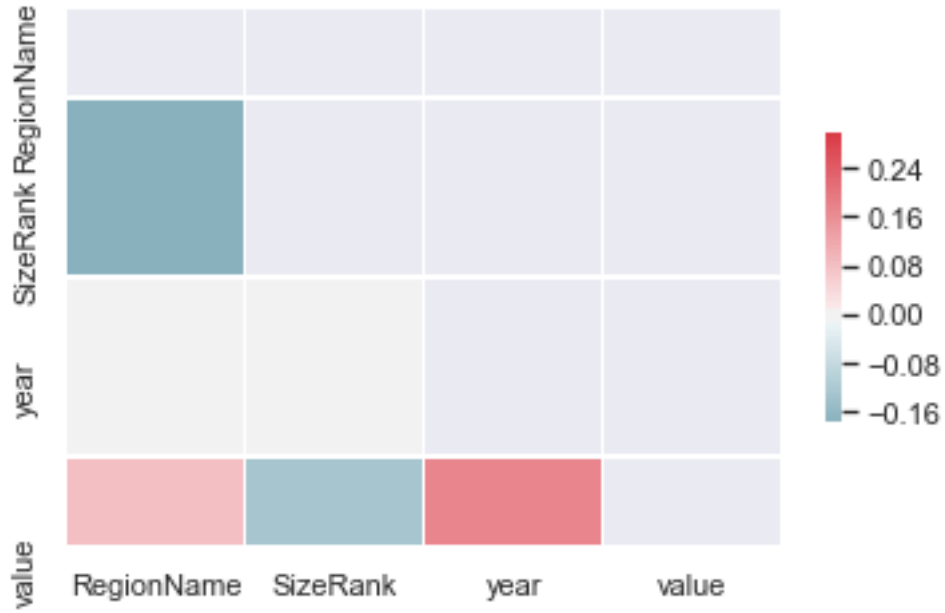


Figure 1.8

Models

In this exercise, models are developed using ARIMA and ordinary least squares.

ARIMA is an acronym that stands for AutoRegressive Integrated Moving Average. It is a class of model that captures a suite of different standard temporal structures in time series data.

OLS Ordinary least squares is a type of linear least squares method for estimating the unknown parameters in a linear **regression** model

ARIMA

An ARIMA model is a class of statistical models for analyzing and forecasting time series data.

It explicitly caters to a suite of standard structures in time series data, and as such provides a simple yet powerful method for making skillful time series forecasts.

ARIMA is an acronym that stands for AutoRegressive Integrated Moving Average. It is a generalization of the simpler AutoRegressive Moving Average and adds the notion of integration.

This acronym is descriptive, capturing the key aspects of the model itself. Briefly, they are:

AR: Autoregression. A model that uses the dependent relationship between an observation and some number of lagged observations.

I: Integrated. The use of differencing of raw observations (e.g. subtracting an observation from an observation at the previous time step) in order to make the time series stationary.

MA: Moving Average. A model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations.

Each of these components are explicitly specified in the model as a parameter. A standard notation is used of ARIMA(p,d,q) where the parameters are substituted with integer values to quickly indicate the specific ARIMA model being used.

The parameters of the ARIMA model are defined as follows:

p: The number of lag observations included in the model, also called the lag order.

d: The number of times that the raw observations are differenced, also called the degree of differencing.

q: The size of the moving average window, also called the order of moving average.

A linear regression model is constructed including the specified number and type of terms, and the data is prepared by a degree of differencing in order to make it stationary, i.e. to remove trend and seasonal structures that negatively affect the regression model.

A value of 0 can be used for a parameter, which indicates to not use that element of the model. This way, the ARIMA model can be configured to perform the function of an ARMA model, and even a simple AR, I, or MA model.

Adopting an ARIMA model for a time series assumes that the underlying process that generated the observations is an ARIMA process. This may seem obvious, but helps to motivate the need to confirm the assumptions of the model in the raw observations and in the residual errors of forecasts from the model.

OLS Regression

OLS chooses the parameters of a linear function of a set of explanatory variables by the principle of least squares: minimizing the sum of the squares of the differences between the observed dependent variable (values of the variable being predicted) in the given dataset and those predicted by the linear function.

Geometrically, this is seen as the sum of the squared distances, parallel to the axis of the dependent variable, between each data point in the set and the corresponding point on the regression surface – the smaller the differences, the better the model fits the data. The resulting

estimator can be expressed by a simple formula, especially in the case of a simple linear regression, in which there is a single regressor on the right side of the regression equation.

The OLS estimator is consistent when the regressors are exogenous, and optimal in the class of linear unbiased estimators when the errors are homoscedastic and serially uncorrelated. Under these conditions, the method of OLS provides minimum-variance mean-unbiased estimation when the errors have finite variances. Under the additional assumption that the errors are normally distributed, OLS is the maximum likelihood estimator.

OLS is used in fields as diverse as economics (econometrics), data science, political science, psychology and engineering (control theory and signal processing).

Model 1: # ARIMA timeseries model for forecasting ZHVI

ARIMA model is constructed for the few metro areas using p, d and q ranging from various input values as given below

p_values = range(0, 7)

d_values = range(0, 3)

q_values = range(0, 7)

Metro areas for which the model is used are as follows

- Fayetteville
- HotSprings
- LittleRock
- Searcy

Model Outcome

Fayetteville **HotSprings** **LittleRock** **Searcy**

```
ARIMA(0, 1, 1) RMSE=941.535
ARIMA(0, 2, 1) RMSE=611.970
ARIMA(1, 1, 0) RMSE=649.378
ARIMA(1, 1, 2) RMSE=617.768
ARIMA(1, 1, 3) RMSE=589.769
ARIMA(1, 2, 0) RMSE=645.333
ARIMA(1, 2, 1) RMSE=630.915
ARIMA(2, 1, 0) RMSE=641.561
ARIMA(2, 1, 1) RMSE=615.212
ARIMA(2, 1, 2) RMSE=588.622
ARIMA(2, 2, 0) RMSE=625.696
ARIMA(2, 2, 1) RMSE=631.356
ARIMA(2, 2, 2) RMSE=613.672
ARIMA(3, 1, 0) RMSE=613.143
ARIMA(3, 1, 1) RMSE=614.405
ARIMA(3, 1, 2) RMSE=610.783
ARIMA(3, 2, 0) RMSE=633.472
ARIMA(3, 2, 1) RMSE=634.762
ARIMA(4, 1, 1) RMSE=621.536
ARIMA(4, 1, 2) RMSE=618.485
ARIMA(4, 2, 0) RMSE=611.747
ARIMA(4, 2, 1) RMSE=598.107
ARIMA(4, 2, 2) RMSE=598.953
ARIMA(4, 2, 3) RMSE=595.875
ARIMA(5, 1, 1) RMSE=608.230
ARIMA(5, 2, 0) RMSE=597.608
ARIMA(5, 2, 1) RMSE=587.425
ARIMA(6, 2, 0) RMSE=600.213
ARIMA(6, 2, 1) RMSE=590.840
Best ARIMA(5, 2, 1) RMSE=587.425
```

```
ARIMA(0, 2, 5) RMSE=362.614
ARIMA(0, 2, 6) RMSE=365.010
ARIMA(1, 1, 0) RMSE=447.838
ARIMA(1, 1, 1) RMSE=366.768
ARIMA(1, 1, 2) RMSE=351.799
ARIMA(1, 1, 3) RMSE=357.517
ARIMA(1, 1, 4) RMSE=363.251
ARIMA(1, 2, 0) RMSE=415.050
ARIMA(1, 2, 1) RMSE=386.957
ARIMA(2, 0, 2) RMSE=351.109
ARIMA(2, 0, 3) RMSE=356.745
ARIMA(2, 0, 4) RMSE=361.522
ARIMA(2, 1, 0) RMSE=381.234
ARIMA(2, 1, 1) RMSE=362.301
ARIMA(2, 1, 2) RMSE=360.613
ARIMA(2, 2, 0) RMSE=364.570
ARIMA(2, 2, 1) RMSE=364.006
ARIMA(3, 1, 0) RMSE=354.073
ARIMA(3, 2, 0) RMSE=365.083
ARIMA(3, 2, 1) RMSE=363.698
ARIMA(4, 1, 1) RMSE=355.451
ARIMA(4, 2, 0) RMSE=367.935
ARIMA(4, 2, 1) RMSE=365.629
ARIMA(5, 1, 1) RMSE=360.663
ARIMA(5, 2, 0) RMSE=360.361
ARIMA(5, 2, 1) RMSE=363.959
ARIMA(5, 2, 2) RMSE=369.395
ARIMA(6, 1, 1) RMSE=357.156
ARIMA(6, 2, 0) RMSE=364.503
ARIMA(6, 2, 1) RMSE=365.904
ARIMA(6, 2, 2) RMSE=371.107
Best ARIMA(2, 0, 2) RMSE=351.109
```

```
ARIMA(0, 2, 1) RMSE=266.625
ARIMA(1, 1, 0) RMSE=308.246
ARIMA(1, 1, 1) RMSE=259.175
ARIMA(1, 1, 2) RMSE=269.896
ARIMA(1, 1, 4) RMSE=263.871
ARIMA(1, 1, 5) RMSE=271.334
ARIMA(1, 1, 6) RMSE=271.988
ARIMA(1, 2, 0) RMSE=308.860
ARIMA(1, 2, 1) RMSE=277.708
ARIMA(1, 2, 2) RMSE=277.254
ARIMA(2, 0, 1) RMSE=258.327
ARIMA(2, 0, 2) RMSE=268.940
ARIMA(2, 1, 1) RMSE=272.906
ARIMA(2, 1, 2) RMSE=273.860
ARIMA(2, 1, 3) RMSE=278.493
ARIMA(2, 2, 0) RMSE=295.356
ARIMA(2, 2, 1) RMSE=277.853
ARIMA(3, 2, 0) RMSE=285.897
ARIMA(3, 2, 1) RMSE=276.563
ARIMA(4, 1, 1) RMSE=271.060
ARIMA(4, 2, 0) RMSE=272.121
ARIMA(4, 2, 1) RMSE=276.697
ARIMA(5, 1, 1) RMSE=271.330
ARIMA(5, 1, 2) RMSE=272.894
ARIMA(5, 2, 0) RMSE=275.004
ARIMA(5, 2, 1) RMSE=277.926
ARIMA(5, 2, 2) RMSE=278.248
ARIMA(6, 1, 1) RMSE=271.049
ARIMA(6, 2, 0) RMSE=276.744
ARIMA(6, 2, 1) RMSE=276.581
ARIMA(6, 2, 2) RMSE=276.385
Best ARIMA(2, 0, 1) RMSE=258.327
```

```
ARIMA(0, 1, 1) RMSE=664.614
ARIMA(0, 2, 1) RMSE=590.695
ARIMA(0, 2, 2) RMSE=593.016
ARIMA(0, 2, 5) RMSE=576.952
ARIMA(0, 2, 6) RMSE=534.385
ARIMA(1, 0, 0) RMSE=1016.690
ARIMA(1, 1, 0) RMSE=610.288
ARIMA(1, 1, 1) RMSE=547.180
ARIMA(1, 1, 2) RMSE=562.855
ARIMA(1, 1, 3) RMSE=541.655
ARIMA(1, 2, 0) RMSE=621.070
ARIMA(1, 2, 1) RMSE=591.554
ARIMA(2, 1, 0) RMSE=575.773
ARIMA(2, 1, 1) RMSE=557.451
ARIMA(2, 2, 0) RMSE=574.833
ARIMA(2, 2, 1) RMSE=574.132
ARIMA(3, 1, 0) RMSE=561.857
ARIMA(3, 1, 1) RMSE=556.506
ARIMA(3, 2, 0) RMSE=578.562
ARIMA(3, 2, 1) RMSE=575.224
ARIMA(4, 1, 0) RMSE=552.577
ARIMA(4, 1, 1) RMSE=556.196
ARIMA(4, 2, 0) RMSE=589.199
ARIMA(4, 2, 1) RMSE=584.312
ARIMA(5, 1, 0) RMSE=560.168
ARIMA(5, 1, 1) RMSE=557.069
ARIMA(5, 2, 0) RMSE=559.461
ARIMA(5, 2, 1) RMSE=563.322
ARIMA(6, 1, 0) RMSE=547.238
ARIMA(6, 1, 1) RMSE=548.679
ARIMA(6, 2, 0) RMSE=565.678
ARIMA(6, 2, 1) RMSE=568.164
Best ARIMA(0, 2, 6) RMSE=534.385
```

Model 2: # OLS Regression for predicting ZHVI using Zillow data

```
                            OLS Regression Results
==============================================================================
=
Dep. Variable:                   value   R-squared:                       0.98
6
Model:                             OLS   Adj. R-squared:                  0.98
6
Method:                  Least Squares   F-statistic:                 2.449e+0
4
Date:                 Sun, 10 Nov 2019   Prob (F-statistic):               0.0
0
Time:                         10:00:53   Log-Likelihood:            -3.3450e+0
6
No. Observations:              286604   AIC:                             6.692e+0
6
Df Residuals:                  285766   BIC:                             6.700e+0
6
Df Model:                         837
Covariance Type:            nonrobust
==============================================================================
==========================================
                                                coef    std err
t      P>|t|      [0.025      0.975]
```

```
------------------------------------------------------------------------------
----------------------------------------
Intercept                                        9.335e+05   1.91e+04
48.917      0.000    8.96e+05    9.71e+05
State[T.AL]                                     -1697.1159   3241.207
-0.524      0.601   -8049.791    4655.560
State[T.AR]                                      4.027e+04   3812.401
10.562      0.000    3.28e+04    4.77e+04
.
.
.
State[T.WA]                                      7.183e+04   6954.182
10.329      0.000    5.82e+04    8.55e+04
State[T.WI]                                      2.012e+04   1613.876
12.468      0.000     1.7e+04    2.33e+04
State[T.WV]                                     -1.427e+04   4044.592
-3.527      0.000   -2.22e+04   -6338.180
State[T.WY]                                       3.48e+04   9614.084
3.620      0.000     1.6e+04    5.36e+04
Metro[T.Abilene]                                 5073.9595   3748.282
1.354      0.176   -2272.569    1.24e+04
Metro[T.Ada]                                     2671.0174   6141.204
0.435      0.664   -9365.572    1.47e+04
.
.
.
Metro[T.Yuma]                                    4806.1553   2699.941
1.780      0.075    -485.653    1.01e+04
Metro[T.Zanesville]                               853.9452   3610.300
0.237      0.813   -6222.144    7930.034
RegionName                                         -1.1717      0.143
-8.211      0.000      -1.451      -0.892
SizeRank                                           -0.0563      0.014
-4.026      0.000      -0.084      -0.029
year                                             -444.0871      8.957
-49.580      0.000    -461.643    -426.532
previous_value                                      1.0400      0.000   3
609.201      0.000       1.039       1.041
==============================================================================
=
Omnibus:                   254070.345   Durbin-Watson:                   0.87
9
Prob(Omnibus):                  0.000   Jarque-Bera (JB):       3799396378.16
5
```

```
Skew:                          2.565   Prob(JB):                         0.0
0
Kurtosis:                    567.032   Cond. No.                      2.95e+2
0
==============================================================================
=

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correc
tly specified.
[2] The smallest eigenvalue is 3.16e-25. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

Model 3: OLS Regression for predicting ZHVI using dataset from Zillow and Centre for Bureau and Labor Statistics

```
                            OLS Regression Results
==============================================================================
====
Dep. Variable:                  value   R-squared:                       0
.987
Model:                            OLS   Adj. R-squared:                  0
.987
Method:                 Least Squares   F-statistic:                 2.682
e+04
Date:                Sun, 10 Nov 2019   Prob (F-statistic):
0.00
Time:                        22:27:06   Log-Likelihood:            -3.3319
e+06
No. Observations:              286604   AIC:                         6.666
e+06
Df Residuals:                  285765   BIC:                         6.674
e+06
Df Model:                         838
Covariance Type:            nonrobust
==============================================================================
==============================================
                                                    coef    std err
t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
------------------------------------------
Intercept                                       -6.814e+04   1.92e+04
-3.545      0.000   -1.06e+05   -3.05e+04
State[T.AL]                                     -1.925e+04   3098.951
-6.212      0.000   -2.53e+04   -1.32e+04
State[T.AR]                                      2.308e+04   3644.415
6.334      0.000    1.59e+04    3.02e+04
.
```

```
.
.
Metro[T.Yuba City]                                    463.9552   2089.655
0.222      0.824   -3631.710    4559.621
.
.
.
Metro[T.Yuma]                                        3173.4506   2579.937
1.230      0.219   -1883.155    8230.056
Metro[T.Zanesville]                                  1208.2214   3449.810
0.350      0.726   -5553.311    7969.753
RegionName                                             -1.2081      0.136
-8.860      0.000      -1.475      -0.941
SizeRank                                               -0.0588      0.013
-4.399      0.000      -0.085      -0.033
year                                                   78.5126      9.126
8.603      0.000      60.625      96.400
u_rate                                               -5174.1171     31.368
-164.949      0.000   -5235.598   -5112.637
previous_value                                          1.0389      0.000
3771.851      0.000       1.038       1.039
==============================================================================
====
Omnibus:                   286911.663   Durbin-Watson:                   0
.758
Prob(Omnibus):                  0.000   Jarque-Bera (JB):       5432354376
.739
Skew:                           3.291   Prob(JB):
0.00
Kurtosis:                     677.432   Cond. No.                     5.95
e+20
==============================================================================
====

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is cor
rectly specified.
[2] The smallest eigenvalue is 7.76e-26. This might indicate that there ar
e
strong multicollinearity problems or that the design matrix is singular.
```

Results

Test data set results are captured for each model specification and are given below

Model 1: # ARIMA timeseries model for forecasting ZHVI

Best ARIMA model for Fayetteville, HotSprings, LittleRock and Searcy are as follows

Fayetteville: Best ARIMA(5, 2, 1) RMSE=587.425

HotSprings: Best ARIMA(2, 0, 2) RMSE=351.109

LittleRock: Best ARIMA(2, 0, 1) RMSE=258.327

Searcy: Best ARIMA(0, 2, 6) RMSE=534.385

Figure 2.1: Shows the scatter plot on root mean squared value for Fayetteville for different p, d and q values

Figure 2.2: Shows the scatter plot on root mean squared value for HotSprings for different p, d and q values

Figure 2.3: Shows the scatter plot on root mean squared value for LittleRock for different p, d and q values

Figure 2.4: Shows the scatter plot on root mean squared value for Searcy for different p, d and q values

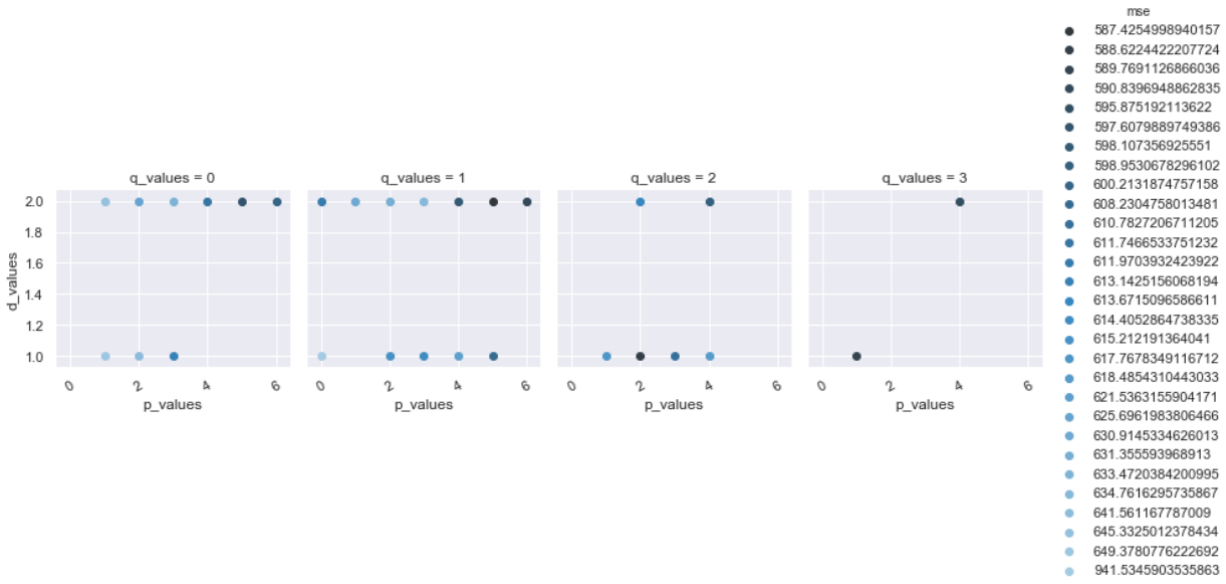


Figure 2.1

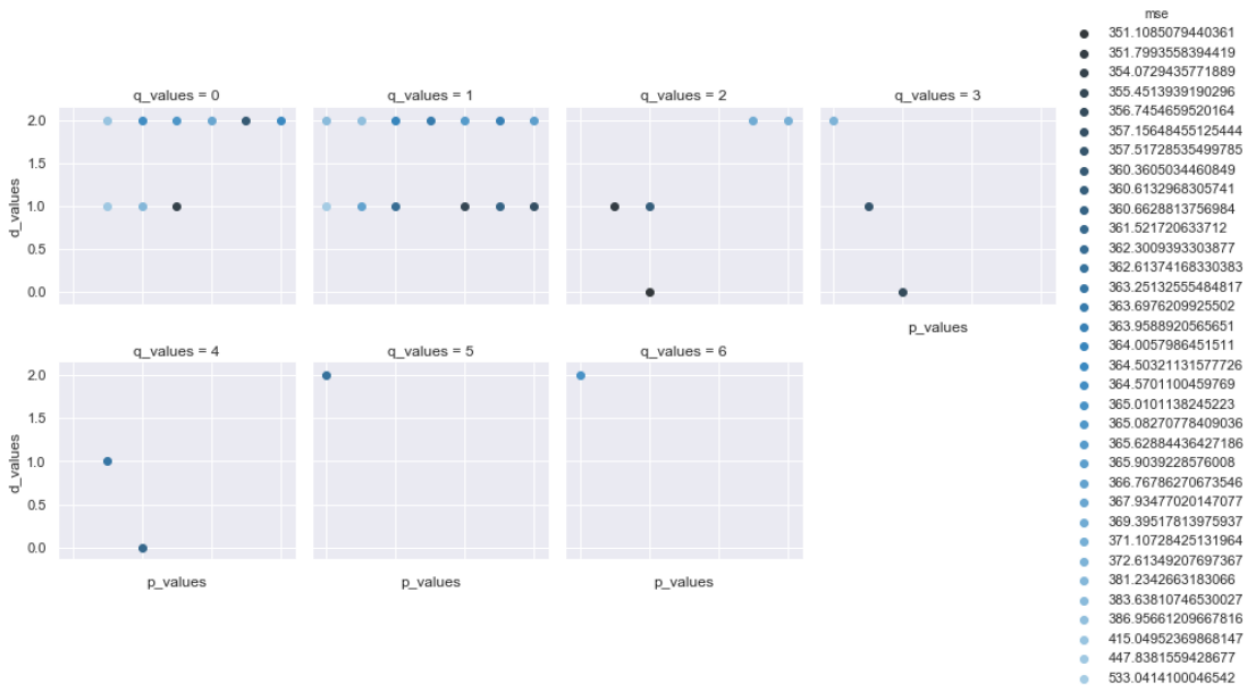


Figure 2.2

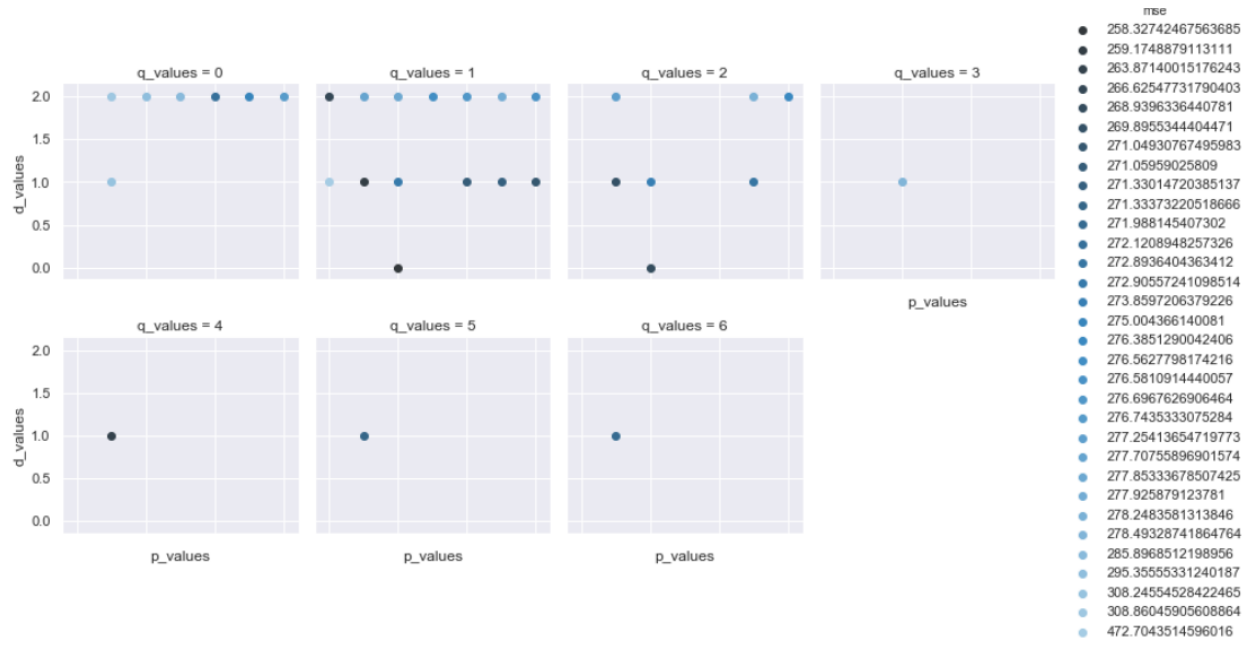


Figure 2.3

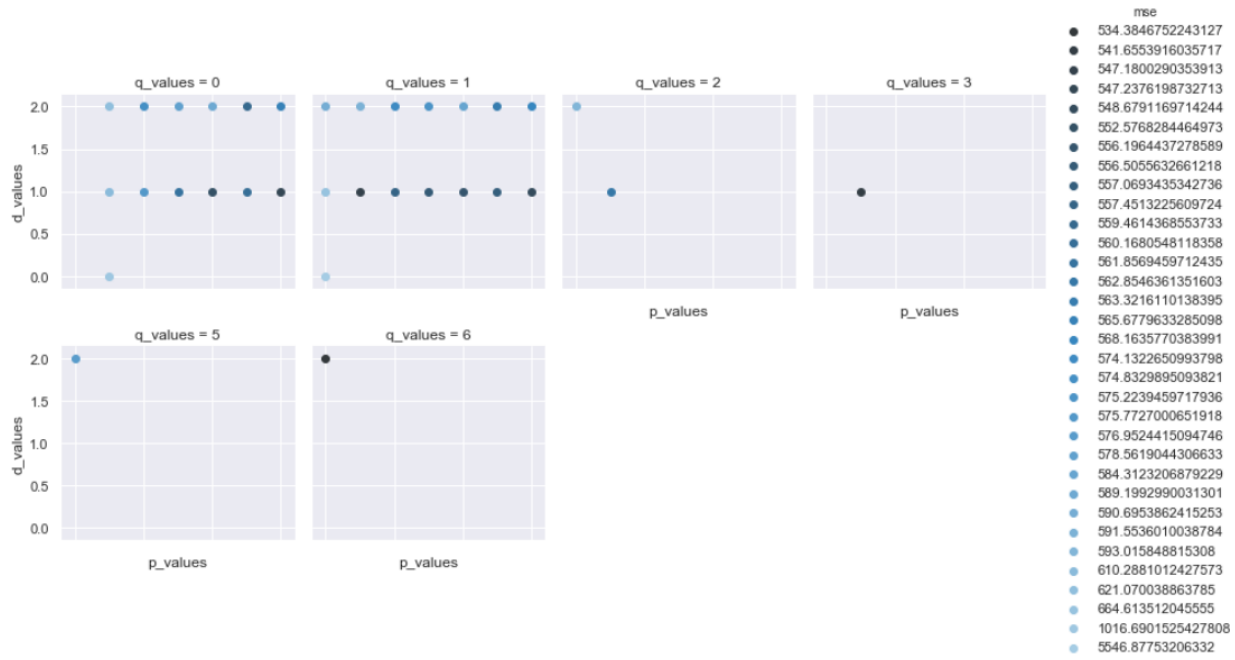


Figure 2.4

Model 2: # OLS Regression for predicting ZHVI using Zillow data

Proportion of Test Set Variance Accounted for: 0.995

Figure 2.5: Shows the top 10 zip codes which yielded consistently higher returns in the past

Figure 2.6: Shows the top 10 zip codes with highest return in 2018

Table 2.1: It gives the top 3 zipcodes which are best for investment for the year 2018 based on the maximum log return

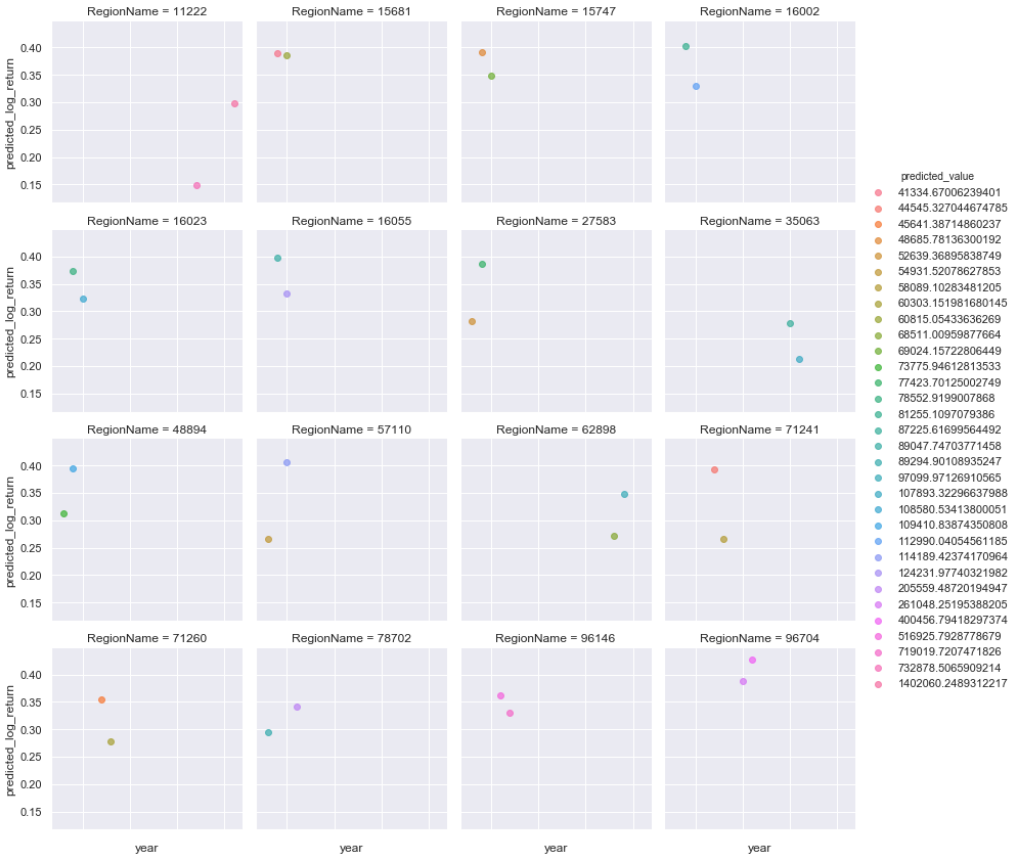


Figure 2.5

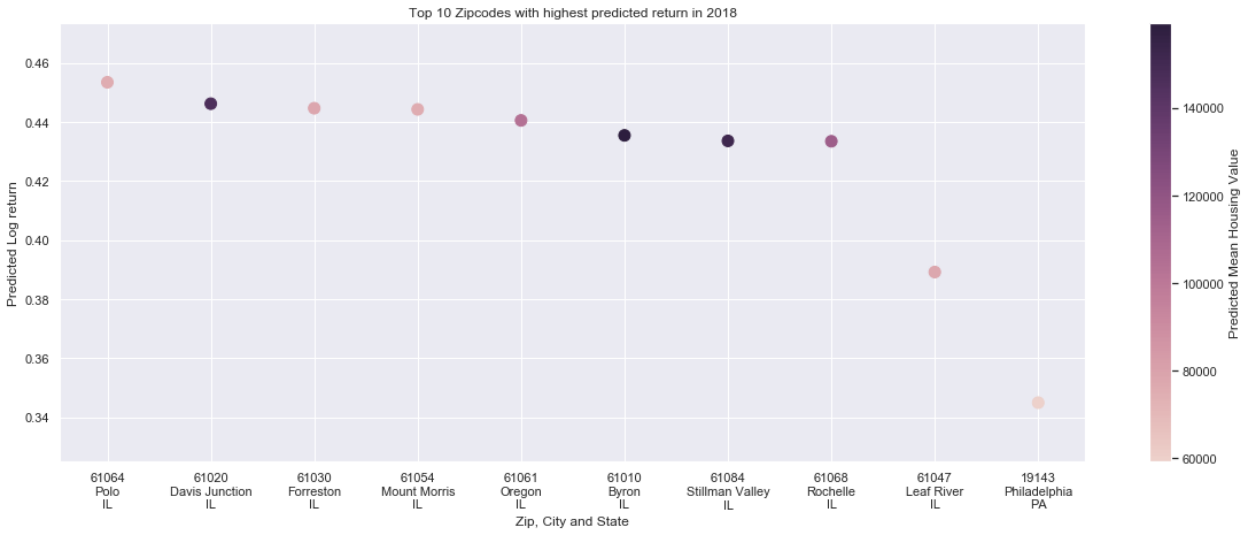


Figure 2.6

RegionName	City	State	Metro	CountyName	SizeRank	year	value	log_return	previous_value	predicted_value	predicted_log_return
61064\nPolo\nIL	Polo	IL	Rochelle	Ogle County	12303	2018	81083.333333	-0.013069	82150.0	75972.815547	0.453430
61020\nDavis Junction\nIL	Davis Junction	IL	Rochelle	Ogle County	13597	2018	147516.666667	-0.014358	149650.0	146150.829194	0.446178
61030\nForreston\nIL	Forreston	IL	Rochelle	Ogle County	14252	2018	83808.333333	-0.011468	84775.0	78632.896214	0.444642

Table 2.1

Model 3: OLS Regression for predicting ZHVI using dataset from Zillow and Centre for Bureau and Labor Statistics

Proportion of Test Set Variance Accounted for: 0.995

Figure 2.7: Shows the top 10 zip codes with highest return in 2018

Table 2.2: It gives the top 3 zipcodes which are best for investment for the year 2018 based on the maximum log return

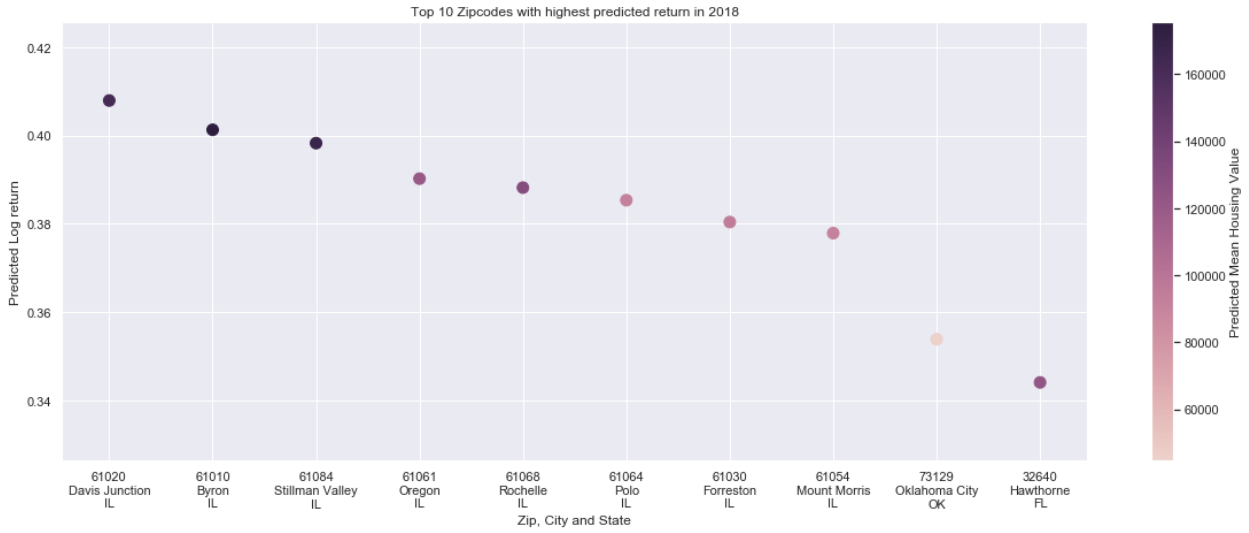


Figure 2.7

RegionName	City	State	Metro	SizeRank	year	previous_value	log_return	u_rate	value	predicted_value	predicted_log_return
61020\nDavis Junction\nIL	Davis Junction	IL	Rochelle	13597	2018	149650.000000	-0.014358	4.002833	147516.666667	162354.916283	0.407905
61010\nByron\nIL	Byron	IL	Rochelle	9726	2018	161950.000000	0.003903	4.002833	162583.333333	175372.585255	0.401271
61084\nStillman Valley\nIL	Stillman Valley	IL	Rochelle	13230	2018	155358.333333	0.010405	4.002833	156983.333333	168229.345987	0.398257

Table 2.2

Conclusion

ARIMA time series plots are developed for the Arkansas metro areas Hot Springs, Little Rock, Fayetteville, Searcy using all the data from 1997 till present and the results are given as follows

Best ARIMA model

Fayetteville: Best ARIMA(5, 2, 1) RMSE=587.425

HotSprings: Best ARIMA(2, 0, 2) RMSE=351.109

LittleRock: Best ARIMA(2, 0, 1) RMSE=258.327

Searcy: Best ARIMA(0, 2, 6) RMSE=534.385

Model(s) for forecasting average median housing value by zip code for 2018 is constructed using OLS regression without any down sampling

What three zip codes provide the best investment opportunity for the SREIT?

The threes zip codes which are best for the investment opportunity is as follows and they are chosen because of the high predicted log return value.

RegionName	City	State	Metro	SizeRank	year	previous_value	log_return	u_rate	value	predicted_value	predicted_log_return
61020\nDavis Junction\nIL	Davis Junction	IL	Rochelle	13597	2018	149650.000000	-0.014358	4.002833	147516.666667	162354.916283	0.407905
61010\nByron\nIL	Byron	IL	Rochelle	9726	2018	161950.000000	0.003903	4.002833	162583.333333	175372.585255	0.401271
61084\nStillman Valley\nIL	Stillman Valley	IL	Rochelle	13230	2018	155358.333333	0.010405	4.002833	156983.333333	168229.345987	0.398257

BONUS

State Average Housing Value is plotted in the US MAP and the figure is shown below