2016-05-30

An investigation into the effectiveness of hedonic features in regression models for domestic rental prices

Peter Prunty

Technological University Dublin

Recommended Citation

Prunty, P. (2016) *An investigation into the effectiveness of hedonic features in regression models for domestic rental prices*. Technological University Dublin, March 2016.

Institiúid Teicneolaíochta Átha Cliath
Dublin Institute of Technology

An investigation into the effectiveness of hedonic features in regression models for domestic rental prices

DT228B

MSc in Computing (Data Analytics)

Peter Prunty

D13122704

Supervisor: Kristina Luus

School of Computing
Dublin Institute of Technology

06th March 2016

School of Computing
Scoil na Ríomhaireachta

Declaration

I certify that this dissertation which I now submit for examination for the award of MSc in Computing (Data Analytics), is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the test of my work.

This dissertation was prepared according to the regulations for postgraduate study of the Dublin Institute of Technology and has not been submitted in whole or part for an award in any other Institute or University.

The work reported on in this dissertation conforms to the principles and requirements of the Institute's guidelines for ethics in research.

Signed:

Peter Prunty

Date: ***06th March 2016***

Abstract

Housing is a fundamental human right. Increasing rents and rising unemployment contribute to increased rates of homelessness. Traditionally housing prices are determined by supply and demand. This project will investigate the relationship between hedonic features and domestic rental prices in California and New York, using multivariate regression models. The literature outlines a number of approaches taken to model real estate pricing using hedonic regression.

Two models were created to analyse the difference between California and New York. Features were selected using correlation analysis. Some features were derived using logarithmic and dummy feature transformations. The models themselves were evaluated by assessing the root mean square error (RMSE) and by visually inspecting the residual plots.

Despite the models not providing a high degree of accuracy in predicting rental prices, a number of valuable insights were gathered by analysing the difference between the regional models. Also, a Tableau dashboard was created to show how such models could be visualised for a data analytics novice.

Areas for future work have also been identified for those interested in expanding upon the work within this project.

Key words: *Data analytics, pricing models, hedonic regression, visualising regression models*

Acknowledgements

I would like to express my sincere thanks to all the staff of DIT who assisted with my learning during the course of the MSc in Computing. I would particularly like to thank my project supervisor, Kristina Luus, whose support and advice has been invaluable throughout this dissertation.

I would also like to thank all of my friends & family sincere thanks for all your help, support, patience and encouragement over the past few months.

This dissertation is dedicated to the memory of my much loved and greatly missed mother, Maura Prunty.

Table of Contents

Table of Figures

Table of Tables

1. Introduction

1.1 Project Background

Davenport & Harris (2007) describe data analytics as a subset of business intelligence, which is a set of technologies and processes that use data to understand and analyse business performance. Figure 1 below maps out topics of interest within the field of data analytics. The experiment in Chapters 3 and 4 will these apply these key concepts in both its design and implementation.

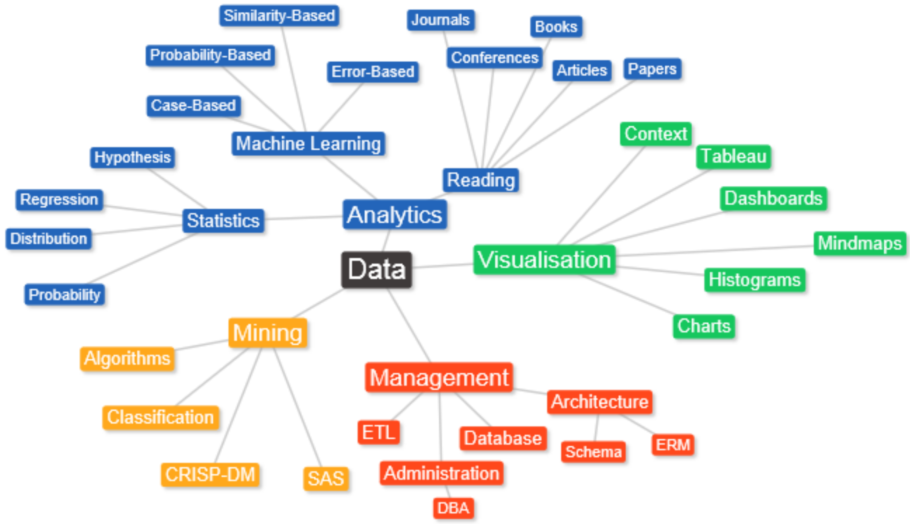


Figure 1 Data Analytics mind map

This research investigates the modelling of U.S. rental prices in the domain of data analytics. By introducing a hedonic model for rental pricing, prospective tenants will be able to determine the expected rental price they could expect to pay, considering factors other than those factored into traditional pricing models. Existing research tends to focus on economic factors impacting on supply side pricing. This research will focus on demand factors which relate to the tenant, allowing them to determine what they can expect to pay for their given circumstance. The model will utilise data gathered from the 2013 American Community Survey (ACS), conducted by the US Census Bureau.

1.2 Research Project

Building upon the research illustrated above, three predictive pricing model for rental pricing will be developed. The models will be developed using ACS data. The experiment will employ a hedonic regression model to determine the relationship between input variables associated with prospective tenants and the target domestic rental price. Multiple iterations of the models will be required to determine which combination of input variables produce the optimal solution.

The models producing the best result will then be incorporated into a Tableau workbook, which will allow users input certain parameters, to determine their expected cost of rent by geographic area. The visualisation will be published online.

The research question asks *Do hedonic features have a linear relationship with domestic rental prices?* The investigation will explore the accuracy of modelling this relationship using linear regression and also explore the regional differences between hedonic predictive features.

1.3 Research Aims and Objectives

The initial objective will be to review current research on hedonic regression models, to understand the applications and limitations with respect to domestic rental market pricing. Having completed a detailed review of the literature, the next objective will be to identify some interesting approaches to domestic rental market pricing. A further objective will be to gather a range of opinions on the suitability of hedonic regression models. Having gathered this information, a model will be designed using the American Community Survey dataset. Multiple iterations will be run to refine the model. The model will then be evaluated, and the results analysed. The final objective will be to identify limitations of the model and determine further areas of research arising from the results.

1.4 Research Methods

The research question will be addressed over the proceeding chapters.

An experiment will then be designed, based on the findings from the literature review, which can support the comparison of hedonic regression models. The experiment will use open data obtained from Kaggle.com and is provided by US Census Bureau from their 2013 American Community Survey (ACS).

An experiment will then be executed in line with the design and the principles of the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology to gain data understanding. Data cleansing and feature selection will be carried out, and multiple iterations of models will be run. Performance measures will be used to investigate the accuracy of the models and to choose the optimal models for implementation and visualisation.

Visualisations of the models will be constructed to display results of the experiment and support discussion of the models and the research question. The discussion will conclude with a summary of the findings and how this contributes to the body of knowledge. Further research topics will be summarised.

1.5 Scope and Limitations

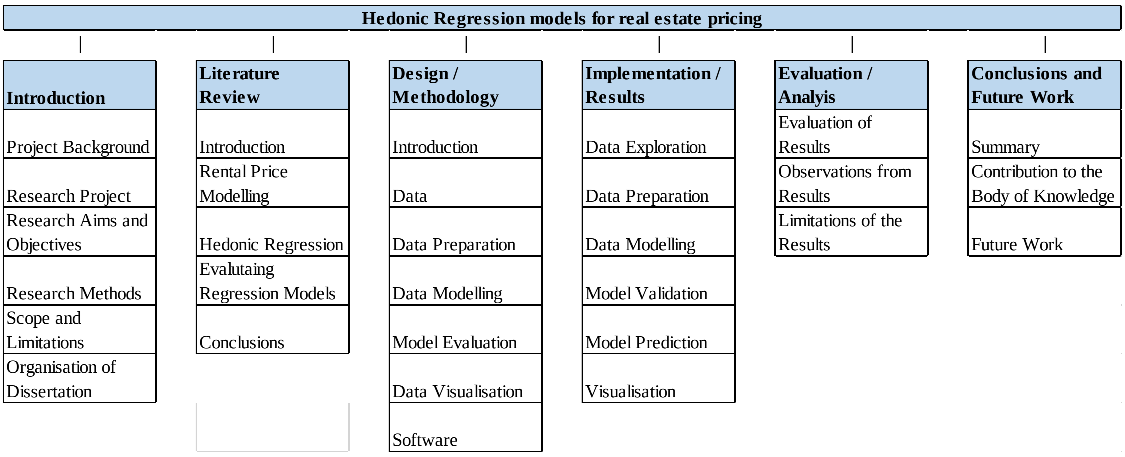
The experiment will focus on modelling pricing for New York and California domestic rental properties in 2013. Domestic properties for the purpose of the experiment are properties which contain a living space with running hot and cold water, flush toilets and no more than five bedrooms. Crawford, Bin, Kruse and Landry (2014) reported that it was not necessary to be temporally explicit when measuring the responsiveness of sales prices of a property to the view from that property. Although they do argue that the hedonic parameters used to measure view are temporally sensitive. The experiment will therefore focus on a fixed point in time - the 2013 data collection.

The dataset used for the experiment contains data on Californian and New York households in 2013, and cannot be supplemented with additional datasets at a granular level. Given that the regression model analysis requires feature selection at the granular level, the scope of the model is limited to the inherent features within the dataset, trending of changes in models over time is out of scope. While the dataset does contain data on a number of US states, two models will be created, focussing on

California and New York. The models will then be compared to discuss differences in the effects of hedonic features on rental pricing in both states. Location is to be considered when comparing the models and to gain insight into how hedonic features vary when determining rental price. A property with more than 5 bedrooms may include property which is used for commercial purposes, such as motels. Also, properties without running water and toilet facilities do not meet minimum standard requirements for living. (Economics Online, 2015) Both models will then be discussed in how they vary in both their composition and accuracy.

1.6 Organisation of Dissertation

The layout of the dissertation is shown in Figure 2 below. There are six chapters which cover Introduction, Literature Review, Design, Experiment, Evaluation and Conclusions & Future Work. Each chapter has a number of sections, which are also summarised in the diagram.



Hedonic Regression models for real estate pricing

Introduction	Literature Review	Design / Methodology	Implementation / Results	Evaluation / Analyis	Conclusions and Future Work
Project Background	Introduction	Introduction	Data Exploration	Evaluation of Results	Summary
Research Project	Rental Price Modelling	Data	Data Preparation	Observations from Results	Contribution to the Body of Knowledge
Research Aims and Objectives	Hedonic Regression	Data Preparation	Data Modelling	Limitations of the Results	Future Work
Research Methods	Evalutaing Regression Models	Data Modelling	Model Validation		
Scope and Limitations	Conclusions	Model Evaluation	Model Prediction		
Organisation of Dissertation		Data Visualisation	Visualisation		
		Software			

Figure 2 Structure of dissertation

Chapter two will outline the literature review which will cover traditional approaches to modelling rental pricing. It will closely examine the role of hedonic regression in this area, with a review of some such models. The chapter will also examine the role of data visualisation, with particular focus on the visualisation of regression models and spatial data. The literature review will conclude by identifying gaps in the current approaches to modelling pricing data and inform an approach to designing the experiment and visualisation.

Chapter three will detail the underlying data set. Approaches to cleansing and modelling of the data in preparation for the experiment will be outlined. The chosen type of visualisation will be explained. The software used for all aspects of the experiment will be discussed. The chapter will end with a discussion on how the results of the experiment are to be measured and evaluated.

Chapter four will include details of how the data was explored in preparation for modelling the data. The details of how the data was modelled and adjusted will too be discussed. The chapter will conclude with a discussion on the executed hedonic regression models were run, adjusted and finalised.

Chapter five will discuss the results from the experiment. The hedonic regression models will be evaluated for accuracy using the coefficient of determination, the standard error of the regression, and visually by plotting the residuals. The regional models compared to each other and the national model. The findings will be compared to the knowledge found in the literature in chapter two.

Chapter six will summarise the findings of the research undertaken during the dissertation. Conclusions will be drawn on the findings and how this contributes to the body of knowledge. The chapter will conclude with areas for further research being identified.

2. Literature Review

2.1 Introduction

This chapter will review the literature relating to the domestic rental market in the US, how hedonic models work, how hedonic models may be applied to the prediction of domestic rental pricing, how models are evaluated and visualised. This literary review will outline past research in these areas and shape the experiment design and implementation.

2.2 Rental Price modelling

Economists regularly measure the elasticity, or responsiveness, of the supply and demand of a good with respect to changes in price. The availability of substitute goods, how necessary a good is, and the percentage of a consumer's income spent on a good are all key demand driven determinants in setting price. Similarly, availability of raw materials, the length of time taken to produce a good, a producer's spare production capacity are all key supply driven determinants in setting price. Traditionally, both elasticity curves are drawn and where supply and demand intersect is price equilibrium. Excess supply or demand occur at all other points, as illustrated by Economics Online (2015).

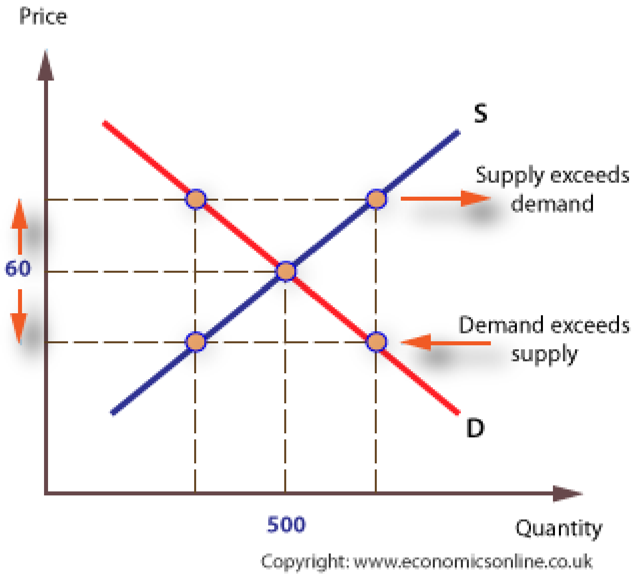


Figure 3 How is equilibrium established?

Source: (Economics Online, 2015)

The principle of market equilibrium underpins pricing models for normal goods. Demand is lower at higher prices, as consumers spend a higher portion of their wages on the good and the good is relatively more expensive to substitute goods. Supply increases with price, as producers expect to make more profit as prices increase. When supply exceeds demand, producers must lower price to attract buyers. Inversely, when demand outstrips supply, producers can increase price until market equilibrium is achieved. Whelan (2015) outlines the continued excess demand for rental property in the USA. Rents for apartments rose nationally for 23 straight quarters, since the first quarter of 2009, and were 15.2% higher than they were at the end of the recession in 2009. This is attributed to the classic economic argument of poor supply to meet increasing demand.

While the rental crisis in the US may well be a simple matter of under supply, it does not help a prospect tenant determine what they can expect to pay for a particular standard of housing. Nor does it assist a landlord when trying to decide how best to maximise the profits from an existing portfolio. This paper will use data analytics to predict rental pricing based on factors relating to the rental property. As such, a tenant can understand what they can expect to pay for a property with such features. Also, a landlord can estimate what price they can expect to receive for their property and the effect of adding additional features will have on that price.

2.3 Hedonic Regression

2.3.1 Hedonic Regression Pricing

Rosen (1974) devised a model to differentiate products based upon the hedonic hypothesis that goods are valued for their utility-bearing characteristics. As these characteristics can be quantified and measured, a regression analysis can be completed to estimate the associated price of the good. The hedonic price model does not typically identify supply nor demand, but rather the price based on the input characteristics.

Hedonic regression has drawn criticisms however. Reis *et al.* (2006) argue that modelling price based on hedonic features alone can lead to inaccuracies as the quality of the product is not taken into account. They found that pricing models which failed to take quality of the product into account resulted in an overestimation of the number of unit sold.

According to Kuminoff *et al.* (2010), the hedonic property value model is among our foremost tools for evaluating the economic consequences of policies that target the supply of local public goods, environmental services and urban amenities. They found that accuracy of the model could be improved by moving from the standard linear pricing model to one that a more flexible framework

The Hedonic Price Method (HPM) is a revealed preference method of valuation. The hedonic price method of environmental valuation uses surrogate markets for placing a value on environmental quality. The real estate market is the most commonly used surrogate in hedonic pricing of environmental values because the word "hedonic" comes from a Greek origin, which means, "pleasure". Hence, the hedonic pricing method relies on information provided by households when they make their location decisions. People derive pleasure by living in nice places. (Gundimeda, 2006)

Bao & Wan (2007) assert that the experience of real estate professionals often provides them with insight into the likely values of true parameters in the hedonic pricing model. They consider improved estimators to allow real estate practitioners to introduce potentially useful information about the parameter values into the estimation of the hedonic pricing model.

Brunauer *et al.* (2010) seek to address two common challenges in hedonic price modelling: nonlinear price functions and the inherent spatial heterogeneity in real estate markets. Accounting for spatial heterogeneity in a very general way, their approach permits higher accuracy in prediction and allows for location-specific nonlinear rent index construction.

Brunauer *et al.* (2013) analyse house price data belonging to three hierarchical levels of special units. They found that hedonic models allow for more precise prediction intervals and with it more reliable risk management.

Liang *et al.* (2011) estimate the determinants of the retail space rent in Shanghai using both hedonic and spatial regression models. They found that the significant explaining variables were the age, the area of retail space, the distance to the Jing An CBD centre, the type of the retail and the district of the property.

Through the results generated by a hedonic pricing model for apartments in Tirana, it was found that, besides residential area location, there are a number of structural features of the apartments, which affect their value, as the flat surface related to the number of rooms, view, the opportunity for parking and furniture. Other features such as residential floor, partial furniture, the age of the apartment, the presence of more than one toilet, number of balconies, the presence of central heating and orientation are estimated to have insignificant impact jointly on the price of apartment. (Boçe, 2015)

Schlapfer *et al.* (2015) suggest that systematic hypothesis testing and reporting of correlations may contribute to consistent explanatory patterns in hedonic pricing estimates for landscape amenities.

The findings outlined above illustrate that the area of hedonic modelling has been a widely studied and researched domain. To this end, this research will set out to define how hedonic modelling may be applied to domestic rental pricing. The key outcome from this will be to define a hedonic regression model which will predict a rental price, for a prospective tenant.

2.3.2 Variables

Chen, Clapp and Tirtiroglu (2011) considered the responsiveness of house sale prices in two districts of China to hedonic variables. They reported that there was a price elasticity of demand for house prices relative to the size and type of housing unit. They infer that developers allocate floor area per housing unit based on the expected return from buyers. Floor area and unit type will therefore be considered as part of the modelling process in later chapters.

Krupka and Donaldson (2013) consider the effects of Quality of Life (QoL) factors such moving costs and wages on regional rents. “Housing supply becomes the main other determinant of regional rents” (Krupka and Donaldson, 2013, p. 844). The experiment in later chapters will give consideration to some QoL factors such as utility bills and access to internet.

Kelleher *et al.* (2015) state that a benefit of using linear regression models is that the weights of the descriptive features in the model describe the effect each feature has on the target feature. However, they also note that it is a mistake to infer the importance of a descriptive feature simply by taking the weight in isolation. Instead they advocate for analysing the t-statistics and p-values of each feature to determine if they are statistically significant (pp. 347-349).

2.3.2 Model Comparison

"Housing studies regarding Chinese cities are limited because of the short history of China's free housing market" (Liao and Wang, 2012;2011;, p. 16).

Liao and Wang (2012;2011;) also argue that the reason there are large variances in results of hedonic pricing models for housing is that each study is specific to a target market. Thus it is difficult to generalise to a universal housing price model. The experiment outlined in the following chapters will compare two region specific models – California and New York, to investigate if the Chinese findings hold true in the USA.

Redfearn (2009) assessed hedonic models focussing on the effects of proximity to light rail on property prices in Los Angeles. It was discovered that such models were highly unstable. In order to account for such variance a number of variables for other local amenities need to be included in the model. Hence, locally-weighted regression models were found to be more robust than standard hedonic models. The experiment in proceeding chapters will compare two localised hedonic regression models with a national hedonic regression model.

2.4 Evaluating Regression Models

In order to assess the accuracy of the regression models, a metric is required to calculate the difference between the predicted value the model produces and the actual target value. The root mean squared error (RMSE) is one such metric and has the added advantage of producing values in the same unit as the target feature. (Kelleher *et al.*, 2015, pp. 443-444).

Frost (2012) states that it is essential to assess residual plots as randomness and unpredictability are crucial components of any regression model. The basic structure of a regression model is:

Response = (Constant + Predictors) + Error

Regression models explain the deterministic portion, (Constant + Predictors) of the response. However it is also imperative to analyse the stochastic error, or randomness of the response. By plotting residuals against the fitted values, a visual representation of the stochastic error is achieved.

The models in the later chapters will be evaluated using a visual inspection of the residuals and also accessing the Root Mean Squared Error of each model.

2.5 Conclusions

This chapter outlined how hedonic regression has been utilised and how it can relate to rental price modelling. The literature also identified some key features which should be considered when modelling for rental price. Limitations to hedonic regression modelling were also identified, but will be factored into the analysis of the experiment results.

The literature also provided insight into how features may be selected for building the hedonic regression model. It showed how to measure the significance of features to produce better models.

An approach for evaluating the models was also provided by the literature. Details of metrics which can be used to measure the hedonic regression models was outlined, and will be implemented in the proceeding chapters.

The research question asks *Do hedonic features have a linear relationship with domestic rental prices?* The literature review helped refine the research question, by highlighting how pricing has been modelled using hedonic variables. Much of the findings of the literature review were based on datasets outside of the US. This inspired the use of similar hedonic features on the dataset comprised solely of US domestic rental properties. The literature review also provided guidance on how best to set up the experiment and measure the results, to accurately determine if hedonic features have a linear relationship on price.

3. Design / Methodology

3.1 Introduction

This chapter will outline the design of the experiment in support of the research question. It will also outline the methodologies employed in evaluating and presenting the findings of the experiment. The composition of the underlying dataset will be explained, as will the process by which the data was explored, prepared, analysed and visualised. An overview of the tools used to complete the experiment will also be provided, addressing the strengths and limitations of same.

The research question asks how hedonic features affect rental prices in California and New York. The purpose of the experiment was to investigate how well hedonic features within the data set were in predicting rental prices. The experiment investigated the relationship between hedonic features and rental price.

3.2 Data

The housing dataset (Kaggle, 2015) was downloaded from Kaggle.com and was provided by US Census from their 2013 American Community Survey (ACS). The data was split over two CSV files, containing 1,476,313 unique records in total. There were a total of 231 features within the dataset. All features were integers, except for RT (Record Type), which was a redundant feature as it was a uniform feature with value ‘H’ (House).

The dataset contained SERIALNO, a unique identifier for each record. It also contained geographical information such as region and state of each respondent. Financial metrics relating to each respondent was included, from income to average monthly costs. Details of the rental property were included, from type of property to the facilities provided, such as internet access and number of bedrooms. 56 of the 231 features were flags, which mostly related to data completeness.

The features outlined in Table 1 represent those utilised in the experiment. The steps involved in the refinement of the dataset are outlined in the next section.

Field Name	Feature Description	Data Type	Unit
ACCESS	Access to the internet	Categorical	Integer
ACR	Size of lot	Categorical	Integer
BDSP	Number of bedrooms in the household	Numerical	Integer
ELEP	Monthly electricity cost	Numerical	US Dollar
HFL	House heating fuel	Categorical	Integer
NP	Number of people in the household	Numerical	Integer
RMSP	Number of rooms in the household	Numerical	Integer
RNTP	Monthly rent for the household	Numerical	US Dollar
ST	State code	Categorical	Integer
YBL	Year household was built	Categorical	Integer

Table 1 Features used in experiment

3.3 Data Preparation

A number of steps were required to prepare the data for investigation and modelling. The two source .csv files needed to be merged to create a master list of records. The experiment assessed the effectiveness of hedonic features in predicting rental prices in both California and New York. This identified differentials in models for the Californian and New York housing markets. The source files stored all features as varchar(254) strings. Once redundant records had been removed, each feature was then converted to the appropriate data type for optimal performance of the data exploration and modelling.

Each feature within the dataset was then analysed. Uniform features were removed, as they are of no value in the regression model. Similarly all null features were removed. Where a feature only had null values and one other value, these features were also removed

Once the redundant records and features had been removed, a full assessment of the remaining features was carried out to determine which features would be selected for the final dataset. Features were selected having analysed the minimum, mean, maximum and standard deviation values. The relationships between the target feature and each independent feature were analysed, as well as the relationships between the independent features themselves. Scatterplots were run to visually inspect the residuals. The p-value and correlation coefficient, r, were also used to determine strength of the relationships.

Similarly, scatterplots were generated between independent features to test for multicollinearity. For redundant multicollinear features, only one was used in the model. Histograms for each model were inspected to get an understanding of variance and outliers. Where a feature appeared skewed, the feature was normalised using a log transformation.

Categorical features existed within the dataset and were stored as numeric values. Dummy features were created by transforming categorical variables.

Having analysed the features within the dataset, a number of predictive features were selected to be included with the target feature, to produce the final dataset.

3.4 Data Modelling

The data was modelled using Alteryx to produce an Analytics Base Table (ABT). Once the ABT had been constructed, the hedonic regression models were designed in Alteryx. Two models were designed and compared: a California model and a New York model. Both models were run within the same workflow within Alteryx, by splitting the dataset on the state feature. The linear regression model algorithm in Alteryx utilises the open source R algorithm, lm, for linear regression.

The dataset for both models was split to train, test and evaluate the models. The split was consistent for both models, with 40% used for training, 20% for validation and fine tuning, and the remaining 40% for testing the models.

A Stepwise Regression was also run on both models to determine if each variable was statistically significant.

3.5 Model Evaluation

The models were evaluated using a number of measures. The coefficient values, t-statistics and p-values for each model were analysed. A large t-statistic implied that the coefficient is likely different than zero. A low p-value implied that the coefficient was statistically significant. Also the Standard Error and R-Squared were analysed. Standard Error of the regression was the average distance of the data points from the regression line in dependent feature units. Standard Error was evaluated with respect to the weights. (Kelleher *et al*., 2015, pp. 347-444) In the case of the models being run in this experiment, the units were US Dollars. The R-Squared value determined how much of the variation in rental price was accounted for by variation in the independent features. Each of these measures were analysed to determine which model is better.

3.6 Data Visualisation

Once models had been generated for both California and New York, Tableau was utilised to create an interactive visualisation. The visualisation consisted of a horizontal bar chart, which updated in real time as input parameters were adjusted. The simple design allowed the viewer to quickly and easily determine difference in rental prices between both states.

3.7 Software

Two software packages were needed to complete the experiment – Alteryx and Tableau.

Firstly, Alteryx was used to explore the data, model the data, run and edit the regression models, and ultimately export the data in .tde format, which is the proprietary Tableau file format. Alteryx was chosen as it has a large number of features which lend itself to the experiment. While other software packages may address elements of the experiment, it was possible to complete all of the data exploration and modelling within a single workflow in Alteryx. Alteryx provided a graphical interface to run regression models, which were based on the R programming language. There were a number of model evaluation nodes in Alteryx which could be used to assess the effectiveness of each model. Alteryx also integrated seamlessly with Tableau. Another advantage of using Alteryx was that there was a large online support community, with active forums. There were a few disadvantages in using Alteryx compared to running the models directly in R. While the predictive model nodes were based on R they did not have the same complete flexibility as writing code directly. Similarly, the options for model evaluation were limited to those approved by Alteryx.

Tableau was the software chosen to visualise the results of the hedonic regression models. Tableau could easily consume the .tde files produced by Alteryx. Its use of

VisQL to explore data was a core feature, promoting the concept of visual analytics. Parameters could be easily created within Tableau to pass values to the underlying models. Tableau also allowed for outputs to be published to its online server, Tableau Online, for free. Like Alteryx, Tableau too had a large online community of users, who actively participate in the community forums, providing support on all queries global developers may have. There were some limitations to Tableau. The type of visualisations was limited when compared to the flexibility offered by some packages available in R. The data once loaded to Tableau Online, was readily accessible by other users, who could download the content for free.

4. Implementation / Results

4.1 Data Exploration

The dataset was filtered to only include records from California and New York, using the state code found in the ST feature. As the experiment focuses on private rental accommodation, records which did not have running and hot and cold water (RWAT), a flush toilet (TOIL) or kitchen facilities (KIT) were removed. Given that the target feature was monthly rental price (RNTP), all records where RNTP was null were also removed, as they were not useful for the purposes of this experiment. This reduced the dataset to 77,445 records for CA and NY, 26,920 records for NY and 50,525 records for CA.

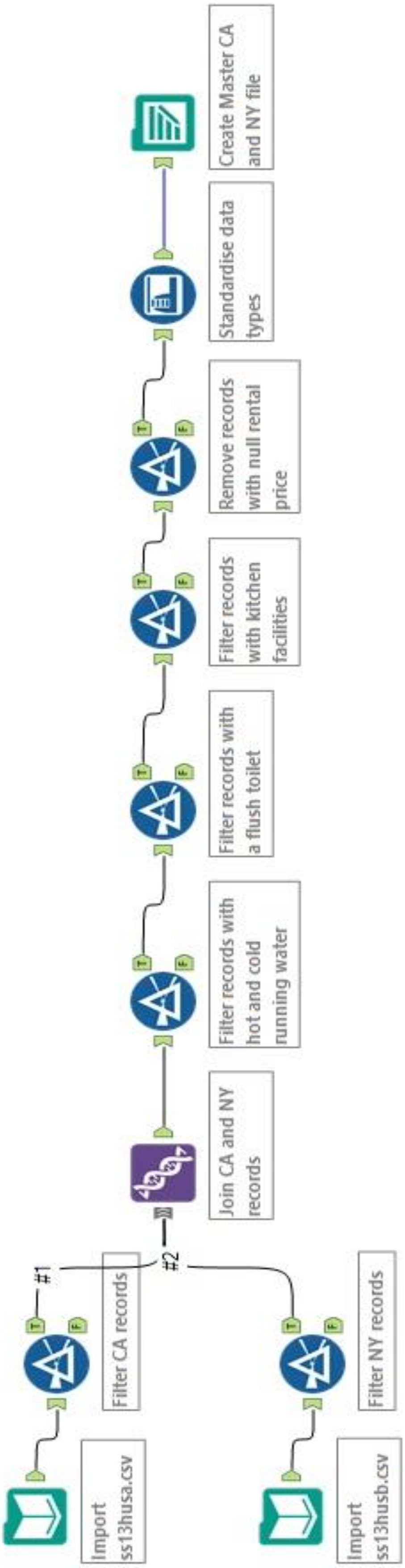

Import ss13husa.csv
Filter CA records
Import ss13husb.csv
Filter NY records
#1
#2
Join CA and NY records
Filter records with hot and cold running water
Filter records with a flush toilet
Filter records with kitchen facilities
Remove records with null rental price
Standardise data types
Create Master CA and NY file

Figure 4 Alteryx workflow for Data Preparation

Correlation analysis was run on each independent feature within the dataset against the target feature to determine relationships. Table 2 shows that Internet access (ACCESS) had the strongest negative correlation with rental price (RNTP). ACCESS contained values {1, 2, 3}, where 1 and 2 indicated access to internet existed, with and without an internet subscription respectively. 3 indicated the household did not have access to internet. The strongest positive correlations were number of bedrooms (BDSP) and number of rooms (RMSP). Overall correlations were not very strong, with largest correlation in absolute terms of 0.3. Some transformations outlined in the next section were carried out on dependent features to address this issue.

Feature Name	**Correlation with RNTP**
ACCESS	-0.30
ACR	-0.13
BDSP	0.31
ELEP	0.16
HFL	-0.12
NP	0.13
RMSP	0.28
YBL	0.11

Table 2 Correlation of dependent features versus target feature

4.2 Data Preparation

The flag ACCESS_IND, as shown in Figure 3, was created grouping values where ACCESS = {1, 2}, as 1, indicating Internet access was available in the property. ACCESS=3 indicated no internet access was available. For these records ACCESS_IND=0.

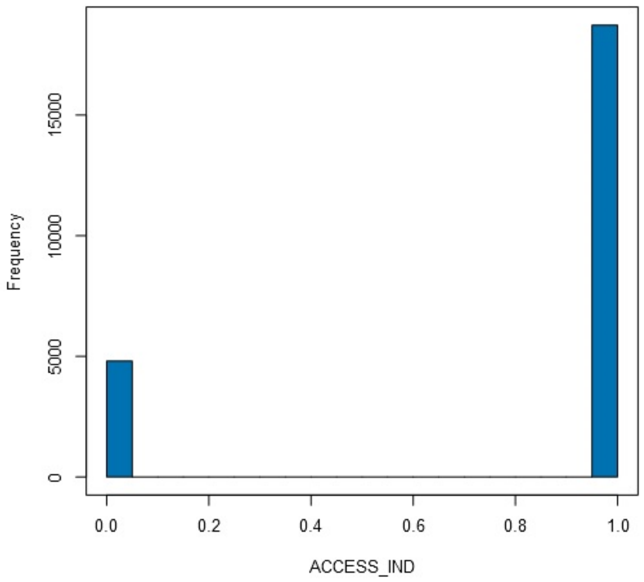


Figure 3 Histogram of ACCESS_IND

Similarly, HPL was used to create a FOSSIL_FUEL_IND. Where HPL indicated property was fuelled by gas, electricity, oil, coal or wood, FOSSIL_FUEL_IND=1. Where HPL indicated property was fuelled by electricity, solar or other methods FOSSIL_FUEL_IND=0.

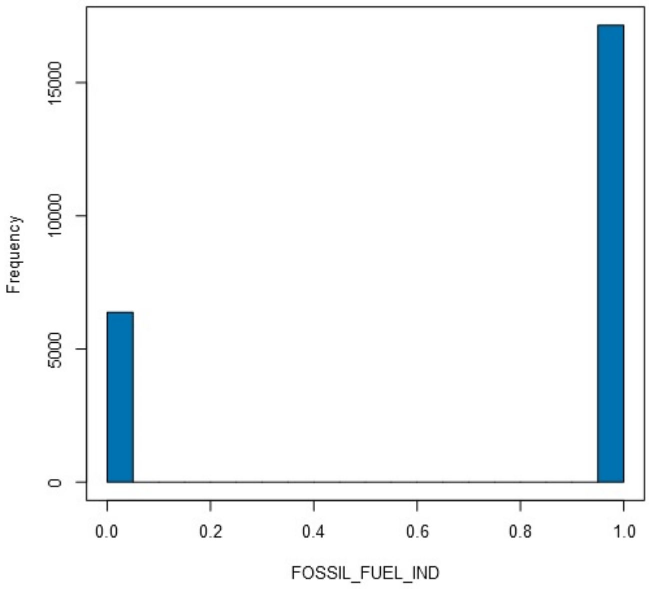


Figure 4 Histogram of FOSSIL_FUEL_IND

As indicated in Figure 5, ELEP proved to have a high positive skew, which necessitated the feature being normalised to be used in the models.

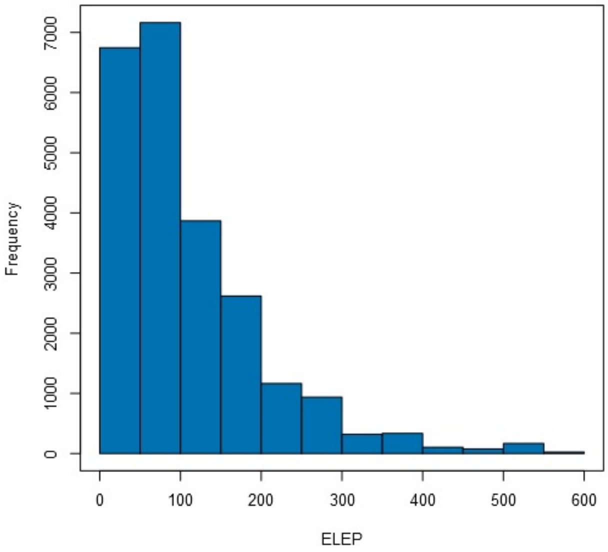


Figure 5 Histogram of ELEP

Normalised_ELEP, in Figure 6, was created by running a log transformation on ELEP

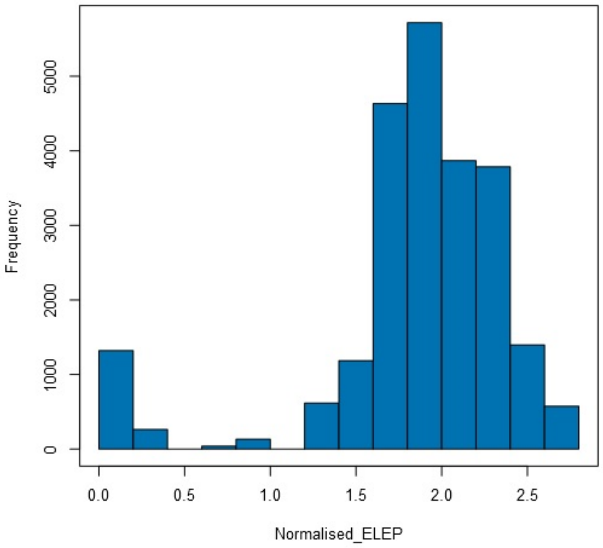


Figure 6 Histogram of Normalised_ELEP

BDSP had a relative normal distribution with a small number of outliers. Records with more than 5 bedrooms were removed from the dataset.

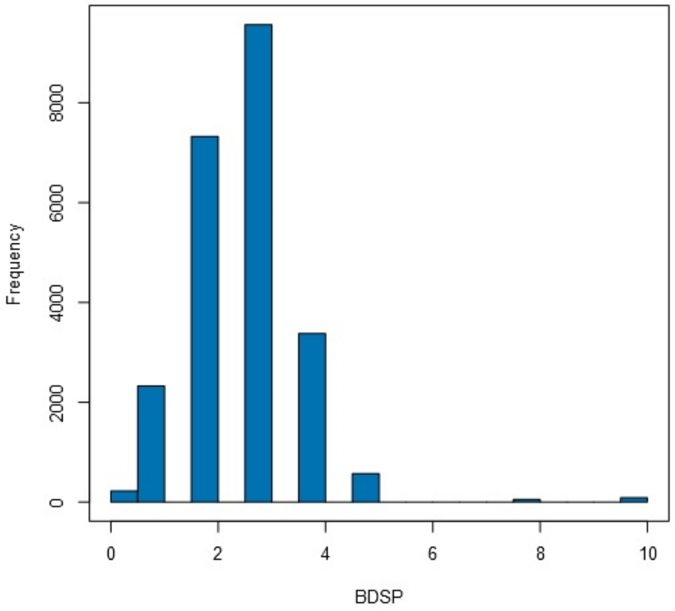


Figure 7 Histogram of BDSP

Figure 8 shows the workflow used to generate the finalised Analytics Base Table (ABT). Records with null values for ACR and ACCESS were removed. Outliers for BDSP were removed in line with scope of the experiment. ACR was transformed into the dummy feature LARGE_PLOT_IND. Where ACR indicated a lot size of less than 1 acre, LARGE_PLOT_IND was set to 0. For all other values where ACR indicated a lot size of greater than 1 acre, LARGE_PLOT_IND was set to 1. Similarly YBL was used to create the OLD_PROPERTY_IND dummy feature. Where YBL indicated the property was built prior to 1950, the value of OLD_PROPERTY_IND was set to 1, and where the property was built in or after 1950, the value was set to 0.

Once all features had been processed, the finalised ABT was exported as an Alteryx database file, to be run in a separate workflow for running the regression models.

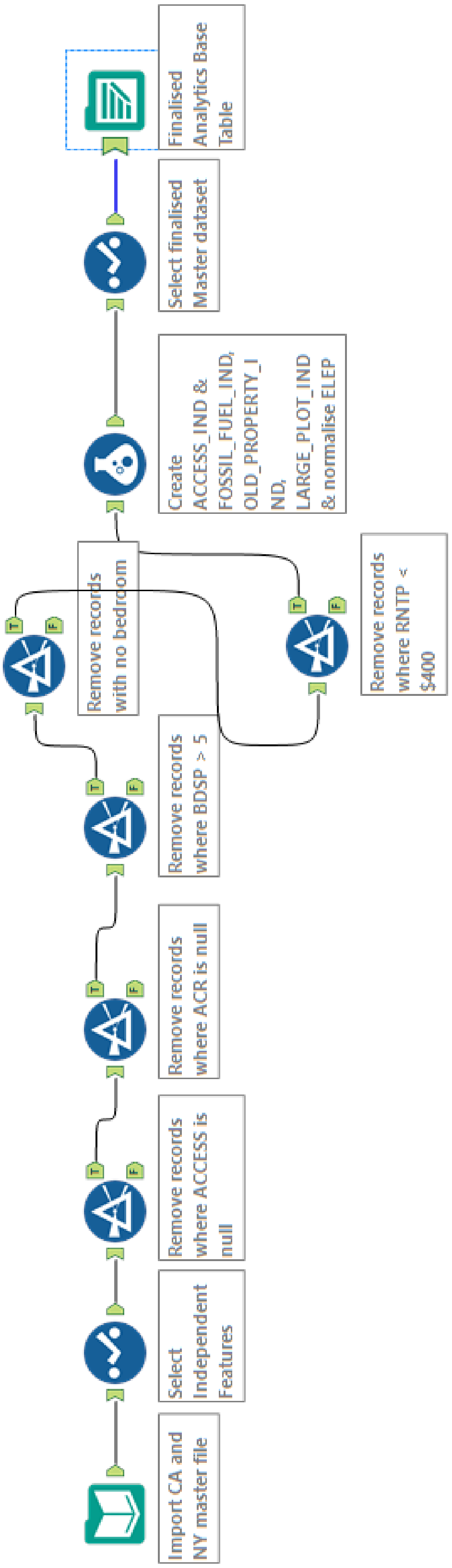

Import CA and NY master file
Select Independent Features
Remove records where ACCESS is null
Remove records where ACR is null
Remove records where BDSP > 5
Remove records with no bedroom
Remove records where RNTP < $400
Create ACCESS_IND & FOSSIL_FUEL_IND, OLD_PROPERTY_IND, LARGE_PLOT_IND & normalise ELEP
Select finalised Master dataset
Finalised Analytics Base Table

Figure 5 Workflow for creation of Analytics Base Table

4.3 Data Modelling

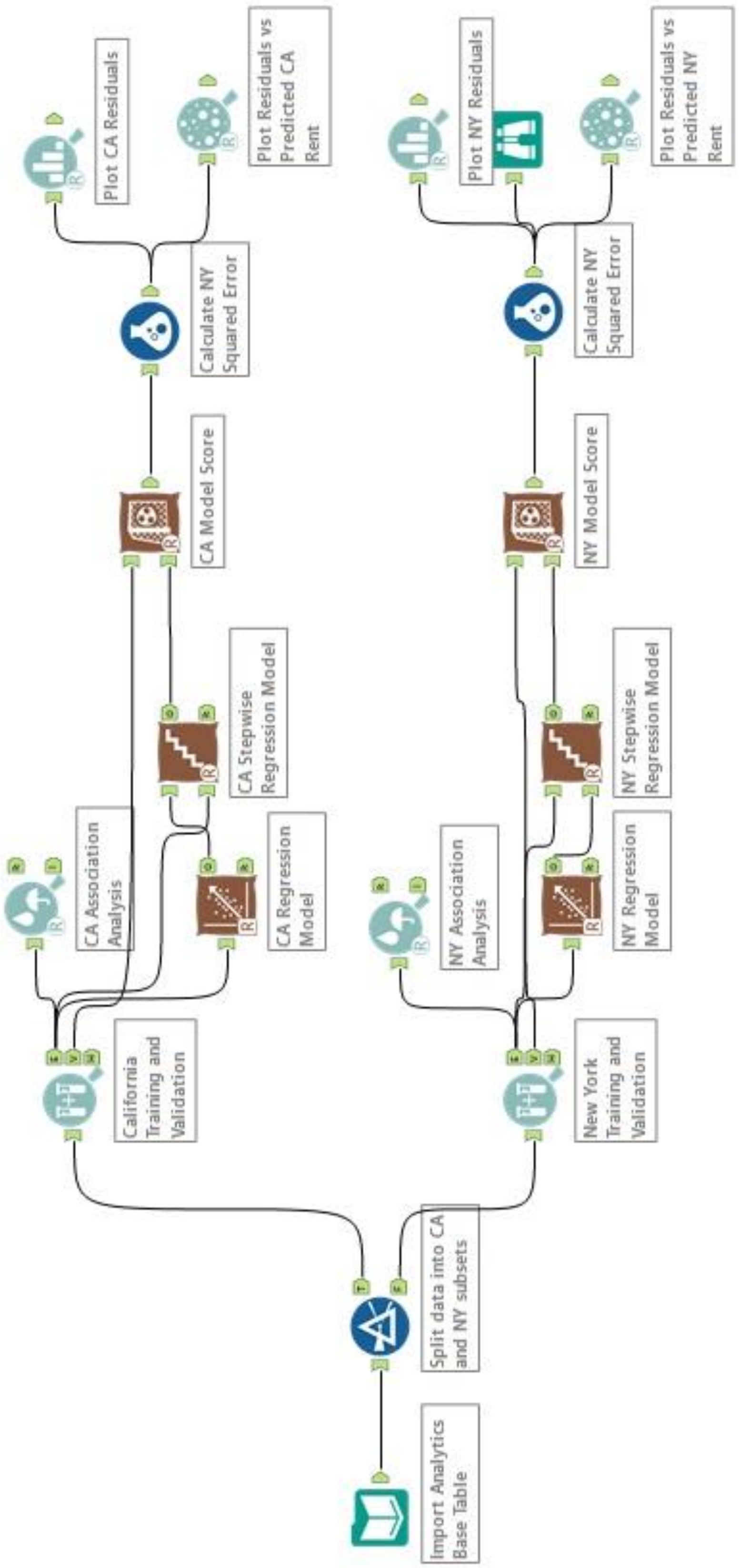

Import Analytics Base Table
Split data into CA and NY subsets
T
F
California Training and Validation
New York Training and Validation
E
V
H
CA Association Analysis
CA Regression Model
CA Stepwise Regression Model
CA Model Score
Calculate NY Squared Error
Plot CA Residuals
Plot Residuals vs Predicted CA Rent
NY Association Analysis
NY Regression Model
NY Stepwise Regression Model
NY Model Score
Calculate NY Squared Error
Plot NY Residuals
Plot Residuals vs Predicted NY Rent

Figure 6 Creation and evaluation of the CA and NY models

Once the Analytics Base Table had been created, a new workflow in Alteryx, shown in Figure 4, was created for creation and evaluation of the CA and NY models. The data was split on ST to create a perfect subset for both California and New York. The relevant subsets were then further divided into training, validation and test sets.
An association analysis was run on both training subsets to investigate the relationship between the target RNTP and the independent features.

	Association Measure	p-value	
BDSP	0.36	0.00	***
RMSP	0.31	0.00	***
ACCESS_IND	0.28	0.00	***
Normalised_ELEP	0.13	0.00	***
LARGE_PLOT_IND	-0.10	0.00	***
OLD_PROPERTY_IND	-0.08	0.00	***
FOSSIL_FUEL_IND	0.08	0.00	***
NP	0.07	0.00	***
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			

Table 3 Association analysis of CA training set

Feature Name	NP	BDSP	RMSP	RNTP	ACCESS_IND	Normalised_ELEP	FOSSIL_FUEL_IND	OLD_PROPERTY_IND	LARGE_PLOT_IND
NP	1.00	0.36	0.19	0.07	0.04	0.16	-0.06	-0.06	-0.01
BDSP	0.36	1.00	0.66	0.36	0.16	0.28	0.05	-0.23	-0.01
RMSP	0.19	0.66	1.00	0.31	0.15	0.22	0.04	-0.15	0.01
RNTP	0.07	0.36	0.31	1.00	0.28	0.13	0.08	-0.08	-0.10
ACCESS_IND	0.04	0.16	0.15	0.28	1.00	0.10	0.07	-0.07	-0.03
Normalised_ELEP	0.16	0.28	0.22	0.13	0.10	1.00	0.01	-0.06	0.02
FOSSIL_FUEL_IND	-0.06	0.05	0.04	0.08	0.07	0.01	1.00	0.01	-0.02
OLD_PROPERTY_IND	-0.06	-0.23	-0.15	-0.08	-0.07	-0.06	0.01	1.00	-0.02
LARGE_PLOT_IND	-0.01	-0.01	0.01	-0.10	-0.03	0.02	-0.02	-0.02	1.00

Table 4 Correlation analysis of CA training set

Correlation Analysis for California showed that number of bedrooms (BDSP), number of rooms (RMSP) and internet access (ACCESS_IND) were had the strongest relationship with rental price. A further seven features had a significant p-value (< 0.001), if somewhat weaker relationship with rental price. However, upon inspection for collinearity, it was discovered the BDSP and RMSP had a score of 0.66. As BDSP had a slightly higher correlation with RNTP, RMSP was discounted from the model. All feature with p-value greater than 0.001 were also discounted.

	Association Measure	p-value	
ACCESS_IND	0.24	0.00	***
NP	0.20	0.00	***
BDSP	0.16	0.00	***
RMSP	0.14	0.00	***
LARGE_PLOT_IND	-0.14	0.00	***
OLD_PROPERTY_IND	-0.11	0.00	***
Normalised_ELEP	0.06	0.07	.
FOSSIL_FUEL_IND	-0.01	0.66	
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			

Table 5 Association analysis of NY training set

Feature	NP	BDSP	RMSP	RNTP	ACCESS_IND	Normalised_ELEP	FOSSIL_FUEL_IND	OLD_PROPERTY_IND	LARGE_PLOT_IND
NP	1.00	0.44	0.32	0.20	0.09	0.19	0.00	0.05	-0.02
BDSP	0.44	1.00	0.74	0.16	0.10	0.18	0.03	0.05	0.05
RMSP	0.32	0.74	1.00	0.14	0.07	0.17	0.06	0.09	0.08
RNTP	0.20	0.16	0.14	1.00	0.24	0.06	-0.01	-0.11	-0.14
ACCESS_IND	0.09	0.10	0.07	0.24	1.00	0.07	-0.02	-0.03	-0.01
Normalised_ELEP	0.19	0.18	0.17	0.06	0.07	1.00	-0.08	0.02	0.01
FOSSIL_FUEL_IND	0.00	0.03	0.06	-0.01	-0.02	-0.08	1.00	0.06	-0.01
OLD_PROPERTY_IND	0.05	0.05	0.09	-0.11	-0.03	0.02	0.06	1.00	-0.13
LARGE_PLOT_IND	-0.02	0.05	0.08	-0.14	-0.01	0.01	-0.01	-0.13	1.00

Table 6 Correlation analysis of NY training set

Correlation Analysis for New York showed that internet access (ACCESS_IND) and number of housemates had the strongest relationship with rental price. Four other features had a significant p-value (< 0.001). Normalised_ELEP and FOSSIL_FUEL_IND were not statistically significant.

4.4 Model Validation

California Model

RNTP ~ 403.91 - 22.85*NP + 239.75*BDSP + 367.24*ACCESS_IND + 37.53*Normalised_ELEP + 62.11*FOSSIL_FUEL_IND - 229.82*LARGE_PLOT_IND

Results from the initial regression model are summarised in the table below:

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	398.54	39.12	10.19	<0.00	***
NP	-22.91	4.69	-4.88	<0.00	***
BDSP	240.90	9.90	24.34	<0.00	***
ACCESS_IND	367.64	20.78	17.70	<0.00	***
Normalised_ELEP	37.50	16.08	2.33	0.02	*
FOSSIL_FUEL_IND	61.89	17.53	3.53	0.00	***
OLD_PROPERTY_IND	11.17	20.19	0.55	0.58	
LARGE_PLOT_IND	-229.30	31.99	-7.17	<0.00	***
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 596.71 on 5587 degrees of freedom Multiple R-squared: 0.19, Adjusted R-Squared: 0.19 F-statistic: 187.5 on 7 and 5586 DF, p-value: < 0.00					

Table 7 Initial CA model

The Adjusted R-Squared value of 0.19 indicated quite poor predictive rate of the CA Model. A stepwise model was run on this model, which removed OLD_PROPERTY_INDICATOR, and returned the same Adjusted R-Squared value. Adjusted weightings for the Stepwise model are summarised below:

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	403.91	37.89	10.66	0.00	***
NP	-22.85	4.69	-4.87	0.00	***
BDSP	239.75	9.68	24.78	0.00	***
ACCESS_IND	367.24	20.76	17.69	0.00	***
Normalised_ELEP	37.53	16.08	2.34	0.02	*
FOSSIL_FUEL_IND	62.11	17.52	3.55	0.00	***
LARGE_PLOT_IND	-229.82	31.97	-7.19	0.00	***
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard Error: 596.71 on 5587 degrees of freedom Multiple R-squared: 0.19, Adjusted R-Squared: 0.19 F-statistic: 218.7 on 6 and 5587 DF, p-value: < 0.00					

Table 8 Stepwise CA Model

New York Model

RNTP ~ 619.98 + 52.94*NP + 63.01*BDSP + 303.63*ACCESS_IND - 169.76*OLD_PROPERTY_IND - 244.61*LARGE_PLOT_IND

Initial results from these models are summarised in the table below:

Coefficients:

	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**	
(Intercept)	628.83	91.74	6.85	0.00	***
NP	52.78	12.47	4.23	0.00	***
BDSP	62.98	21.59	2.92	0.00	**
ACCESS_IND	303.27	42.81	7.08	0.00	***
Normalised_ELEP	2.34	27.87	0.08	0.93	
FOSSIL_FUEL_IND	-14.15	57.19	-0.25	0.80	
OLD_PROPERTY_IND	-169.32	36.08	-4.69	0.00	***
LARGE_PLOT_IND	-244.72	47.17	-5.19	0.00	***
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 563.36 on 1021 degrees of freedom Multiple R-squared: 0.13, Adjusted R-Squared: 0.12 F-statistic: 21.82 on 7 and 1021 DF, p-value: < 0.00					

Table 9 Initial NY model

The Adjusted R-Squared value of 0.12 indicated quite poor predictive rate of the NY Model. A stepwise model was run on this model, which removed Normalised_ELEP and FOSSIL_FUEL_IND. The stepwise returned a marginally better Adjusted R-Squared value of 0.13. Results for the Stepwise model are summarised below:

Coefficients:

	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**	
(Intercept)	619.98	62.85	9.87	<0.00	***
NP	52.94	12.35	4.29	0.00	***
BDSP	63.01	21.44	2.94	0.00	**
ACCESS_IND	303.63	42.71	7.11	0.00	***
OLD_PROPERTY_IND	-169.76	35.99	-4.72	0.00	***
LARGE_PLOT_IND	-244.61	47.12	-5.19	0.00	***
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 562.83 on 1023 degrees of freedom Multiple R-squared: 0.13, Adjusted R-Squared: 0.13 F-statistic: 30.59 on 5 and 1023 DF, p-value: < 0.00					

Table 10 Stepwise NY model

4.5 Model Prediction

Models were evaluated by examining the plots of residuals and determining the root mean squared errors.

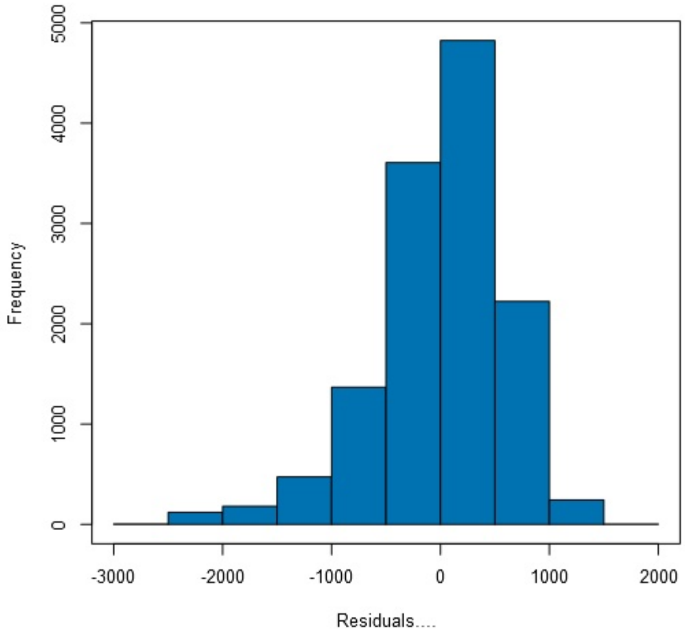


Figure 7 Plot of CA Residuals

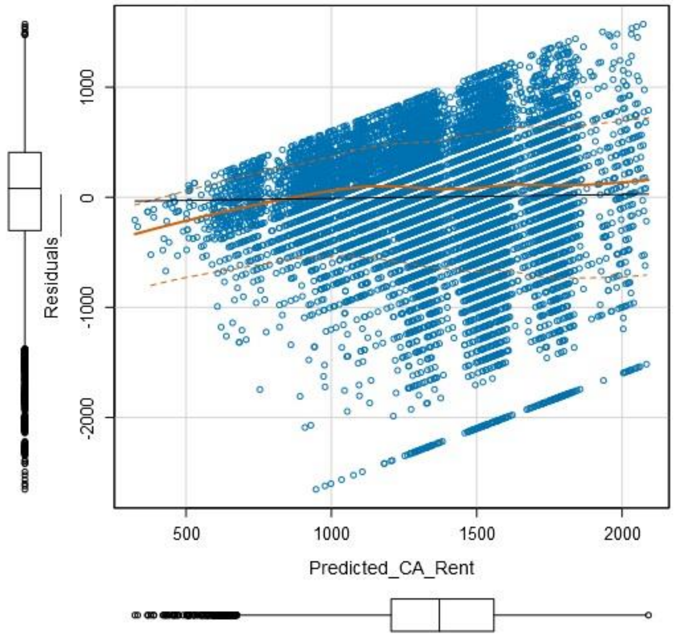


Figure 8 CA Residuals vs Fit Plot

The RMSE for the CA model was $593.

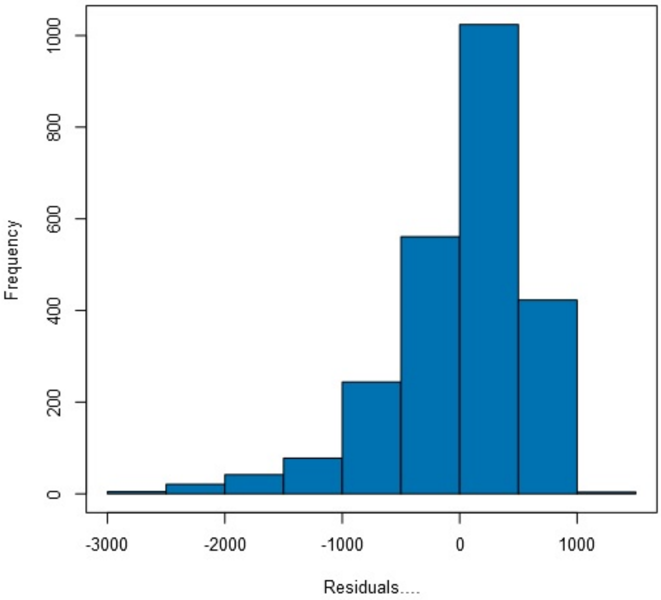


Figure 9 Plot of NY Residuals

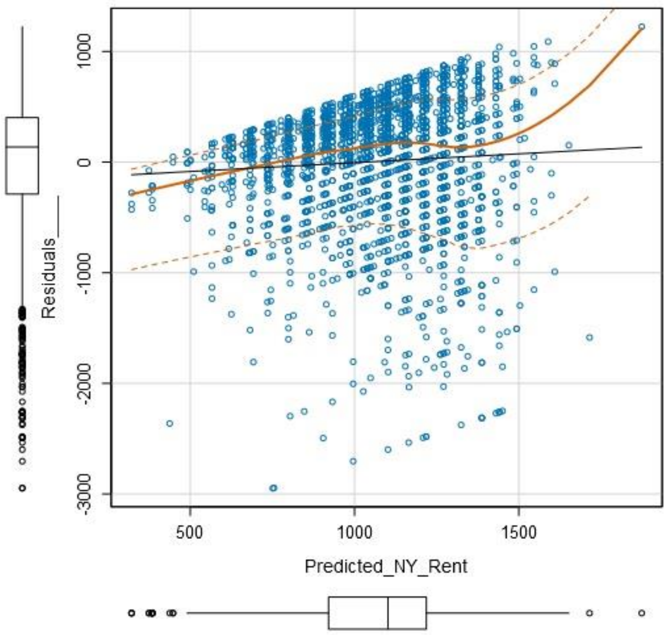
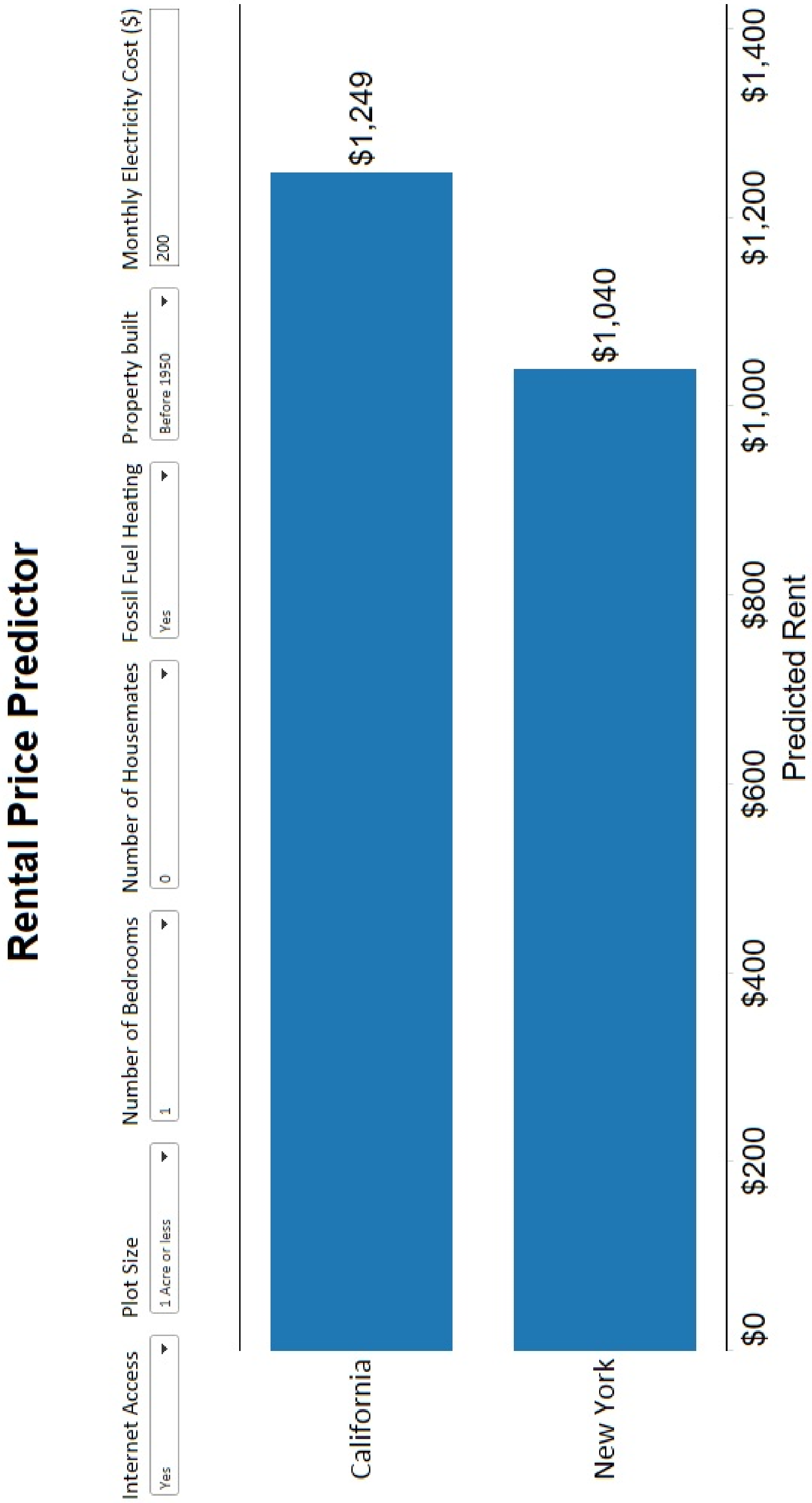


Figure 10 NY Residuals vs Fit Plot

The RMSE for the New York model was $586.

4.6 Visualisation

The selected models for New York and California were coded in Tableau. Parameters were created to adjust for selected values for each of the hedonic variables, with the visualisation producing the predicted value for both New York and California.



Rental Price Predictor

Internet Access: Yes

Plot Size: 1 Acre or less

Number of Bedrooms: 1

Number of Housemates: 0

Fossil Fuel Heating: Yes

Property built: Before 1950

Monthly Electricity Cost ($): 200

Figure 11 Tableau Visualisation of Models

5. Evaluation / Analysis

This chapter will evaluate the experiment results. Recommendations will be made of the best model based on experiment results in line with the learnings gathered from the literature review. Analysis of the models will highlight differences between the regional models. Limitations of the models will also be discussed.

5.1 Evaluation of Results

Table 11 outlines the summary of results from the Californian and New York models. Both models have poor prediction rates. The Californian model's adjusted R squared is 0.19, which means only 19% of variance in rental price can be explained by the model. The New York model's adjusted R squared is even weaker at 0.13. The poor performance can be attributed to the poor correlations of the dependent features used in the model. The removal of missing values may have contributed to the poor performance of the models. This may be addressed if there was a higher volume of complete data. The literature also indicated that hedonic regression pricing model's performance can be impeded by not taking account of the quality of the item being modelled.

A visual analysis of the residuals shows that they are slightly negatively skewed for both models. The scatterplots of both models residuals versus predicted values also highlight a poor fit, as there appears to be a linear pattern in both cases.

Model	Training Set Percentage	Validation Set Percentage	Adjusted R Squared	Standard Error	RMSE
Stepwise CA Model	30%	70%	0.19	$596	$593
Stepwise NY Model	30%	70%	0.13	$563	$586

Table 11 Summary of model results

5.2 Observations from the Results

The research question sought to determine the relationship of hedonic features with rental pricing across different locations. While the results of the models did not indicate a strong relationship, it is interesting to note that the poor performance of the hedonic features is consistent in both models. This supports the literature learning that hedonic regression pricing models are vulnerable when not taking account of quality of the underlying asset.

Another interesting observation is the differences in the dependent variables in both models. Both models contain common dependent features. These are number of people in the household (NP), number of bedrooms (BDSP), internet access (ACCESS_IND) and size of lot of household (LARGE_PLOT_IND). However, these common dependent features have different effects on rental price in both California and New York. NP has a negative impact on price in California, yet in New York it has a positive impact on rental price. In California each additional housemate reduces rental price by $22.85, while in New York an additional housemate increases rental price by $52.94. In California an additional bedroom increases rental price by $239.75, almost four times the amount an additional bedroom in New York increases rental price, $63.01. Internet access also has a greater impact on Californian rental prices by a factor of almost 21%

Both models also have dependent features which were statistically significant in one region. Properties built prior to 1950 (OLD_PROPERTY_IND) can expect to attract €169.76 less in rent in New York. OLD_PROPERTY_IND is not statistically significant in the Californian model. In California properties which are heated by fossil fuels (FOSSIL_FUEL_IND) attract an additional $62.11. Monthly electricity costs (Normalised_ELEP) also increase the rental asking price in California. Neither FOSSIL_FUEL_IND nor Normalised_ELEP are statistically significant in the New York model.

5.3 Limitations of the Results

There were a number of limitations to the dissertation. The main limitation is the poor accuracy of both models. This may be due to poor selection of dependent features, poor preparation of the data or improperly handling null values. Imputing values for nulls of dependent features may have improved the accuracy of the models. Accuracy may be improved if additional hedonic features were sourced and merged. An alternative methodology to regression modelling may also have produced more accurate results.

6. Conclusions and Future Work

This chapter completes the dissertation by summarising the findings in relation to the research question. The dissertation sought to investigate the relationship between hedonic features and domestic rental prices in California and New York utilising a regression model. *Do hedonic features have a linear relationship with domestic rental prices?* was the research question being asked. The question was asked to determine if modelling rental prices purely based on hedonic features can result in an accurate model, different to the traditional models used by economists. The next section will outline the findings for each objective of the dissertation

6.1 Summary

The initial objective of this dissertation was to review current research on hedonic regression models. Chapter two detailed findings from the literature. The objective was carried out by providing a summary of some applications of hedonic regression models. Some limitations were identified, along with key features for consideration when modelling rental prices with hedonic features.

Having gathered the information in the literature review, a hedonic regression model was designed using the American Community Survey dataset. Chapter three provided a detailed description of the experiment design and the composition of the dataset. It outlined the steps undertaken to investigate the data. All data preparation steps were also outlined in order to create the Analytics Base Table. Detailed schematics of workflows from Alteryx were also included to demonstrate the work completed in preparing both the data and the models. The techniques employed in evaluating the accuracy of the hedonic regression models were also outlined. The approach to visualising results was also mentioned. A detailed analysis of the software used in the experiment completed the design chapter.

Chapter four outlined the results of carrying out the data exploration steps identified in the design. Features identified in the literature review were selected for modelling purposes. The chapter also detailed the results of data transformations and the correlation analysis of the selected dependent features. Models were then constructed and validated for accuracy. Tableau was utilised to demonstrate how the models may be visualised for an end user not familiar with the mechanics of the underlying model.

Chapter five reviewed the results from the experiment. The accuracy of the hedonic regression models were evaluated using the standard error of the regression, the root mean squared error and visually by analysing the residual plots. While the accuracy of the models was not very good, they did offer insight into how hedonic features vary differently depending on location. The regional models were compared to each other. The insights gathered from identifying key differences in both models was then discussed.

6.2 Contribution to the body of knowledge

There were a number of contributions to the body of knowledge from this dissertation. Firstly, the literature review identified approaches to hedonic regression and discussed their findings. The weak results of the models indicates that there are further features which need to be identified to accurately model rental pricing. The variance in dependent hedonic features across locations is a significant finding. Similarly the variance in magnitude of significance of hedonic dependent features is noteworthy. As too is the inverse correlations which some dependent features have on rental price depending on location. A review of this dissertation could inspire further investigation of hedonic features as predictors of rental pricing using alternative modelling methodologies.

6.3 Future Work

A number of areas of future work have been identified while completing this dissertation. Scope and limitations of the experiment could be expanded to consider a wider range of hedonic dependent features for the regression models. Likewise the same dependent features could be used with an alternative modelling technique to see if more accurate models could be constructed. Models could also be derived for other locations, such as the other remaining states within the dataset. A national model for the US could also be constructed, including location by state as a hedonic feature. Alternative datasets could be gathered to check models using the hedonic features identified. So too could the models be re-evaluated when the next data is released by the American Community Survey (ACS).

Bibliography

Bao, H, & Wan, A 2007, 'Improved Estimators of Hedonic Housing Price Models', *Journal Of Real Estate Research*, 29, 3, pp. 267-301, EconLit with Full Text, EBSCO*host*, viewed 30 August 2015.

Boçe, M.T. 2015, "Application of a Hedonic Pricing Model for Assessment of Apartments in Tirana, Albania", Journal of Economic Development, Management, I T, Finance, and Marketing, vol. 7, no. 1, pp. 75.

Brunauer, W, Lang, S, & Umlauf, N 2013, 'Modelling house prices using multilevel structured additive regression', *Statistical Modelling: An International Journal*, 13, 2, pp. 95-123, Academic Search Complete, EBSCO*host*, viewed 30 August 2015.

Brunauer, W.A., Lang, S., Wechselberger, P. & Bienert, S. 2010, "Additive Hedonic Regression Models with Spatial Scaling Factors: An Application for Rents in Vienna", The Journal of Real Estate Finance and Economics, vol. 41, no. 4, pp. 390-411.

Chen, Y., Clapp, J., & Tirtiroglu, D. (2011). Hedonic estimation of housing demand elasticity with a markup over marginal costs. *Journal of Housing Economics, 20*(4), 233-248. doi:10.1016/j.jhe.2011.07.001

Crawford, T. W., Bin, O., Kruse, J. B., & Landry, C. E. (2014). On the importance of time for GIS view measures and their use in hedonic property models: Does being temporally explicit matter? *Transactions in GIS, 18*(2), 234-252. doi:10.1111/tgis.12036

Davenport, T.H. & Harris, J.G. 2007, Competing on analytics: the new science of winning, Harvard Business School Press, Boston, Mass.

Economics Online,. (2015). *Equilibrium*. Retrieved 8 December 2015, from http://www.citizensinformation.ie/en/housing/renting_a_home/repairs_maintenance_and_minimum_physical_standards.html

Economics Online,. (2015). *Equilibrium*. Retrieved 8 December 2015, from http://www.economicsonline.co.uk/Competitive_markets/Market_equilibrium.html

Frost, J. (2012, April 5). Why You Need to Check Your Residual Plots for Regression Analysis: Or, To Err is Human, To Err Randomly is Statistically Divine [Blog post]. Retrieved from http://blog.minitab.com/blog/adventures-in-statistics/why-you-need-to-check-your-residual-plots-for-regression-analysis

Gundimeda, H. 2006. Hedonic Price Method-A Concept Note. Madras, Tamil Nadu: Centre for Excellence (COE), Madras School of Economics.

Kaggle. (2015). *2013 American Community Survey. Kaggle.* Retrieved 23 August 2015, from https://www.kaggle.com/census/2013-american-community-survey

Kelleher, John D., MacNamee, B. & D'Arcy, A. (2015). *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies.* Cambridge, MA: The MIT Press.

Krupka, D. J., & Donaldson, K. N. (2013). Wages, rents, and heterogeneous moving costs. *Economic Inquiry,51*(1), 844-864. doi:10.1111/j.1465-7295.2012.00475.x

Kuminoff, N.V., Parmeter, C.F. & Pope, J.C. 2010, "Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities?", Journal of Environmental Economics and Management, vol. 60, no. 3, pp. 145-160.

Liang, J., Wilhelmsson, M., Centra, KTH, Centrum för bank och finans,Cefin & Skolan för arkitektur och samhällsbyggnad (ABE) 2011, "The value of retail rents with regression models: a case study of Shanghai", Journal of Property Investment & Finance, vol. 29, no. 6, pp. 630-643.

Liao, W., & Wang, X. (2012;2011;). Hedonic house prices and spatial quantile regression. *Journal of Housing Economics, 21*(1), 16-27. doi:10.1016/j.jhe.2011.11.001

Osland, L. (2013). The importance of unobserved attributes in hedonic house price models. *International Journal of Housing Markets and Analysis, 6*(1), 63-78. doi:10.1108/17538271311306020

Redfearn, C. L. (2009). How informative are average effects? hedonic regression and amenity capitalization in complex urban housing markets. *Regional Science and Urban Economics, 39*(3), 297-306. doi:10.1016/j.regsciurbeco.2008.11.001

Reis, Hugo J.; Silva, J. M. C. Santos (2006). "Hedonic Price Indexes for New Passenger Cars in Portugal (1997–2003)". *Economic Modelling* 23 (6): 890–906

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal Of Political Economy, 82*(1), 34-55.

Saphores, J., & Li, W. (2012;2011;). Estimating the value of urban green areas: A hedonic pricing analysis of the single family housing market in los angeles, CA. *Landscape and Urban Planning, 104*(3-4), 373-387. doi:10.1016/j.landurbplan.2011.11.012

Schlapfer, F., Waltert, F., Segura, L. & Kienast, F. 2015, "Valuation of landscape amenities: A hedonic pricing analysis of housing rents in urban, suburban and periurban Switzerland", Landscape and Urban Planning, vol. 141, pp. 24-40.

Shen, L. (2012). Are house prices too high in china? *China Economic Review, 23*(4), 1206-1210. doi:10.1016/j.chieco.2012.03.008

Sue, E. D. W., & Wong, W. (2010). The political economy of housing prices: Hedonic pricing with regression discontinuity. *Journal of Housing Economics, 19*(2), 133-144. doi:10.1016/j.jhe.2010.04.004

Sunding, D. L., & Swoboda, A. M. (2010). Hedonic analysis with locally weighted regression: An application to the shadow cost of housing regulation in southern California. *Regional Science and Urban Economics, 40*(6), 550-573. doi:10.1016/j.regsciurbeco.2010.07.002

Whelan, R. (2015). *Apartment Rents Are Rising Steadily and Quickly. WSJ.* Retrieved 8 December 2015, from http://www.wsj.com/articles/apartment-rents-are-rising-steadily-and-quickly-1412220601

Winters, J. V. (2013). Differences in quality of life estimates using rents and home values. *The Annals of Regional Science, 51*(2), 377-409. doi:10.1007/s00168-013-0551-7

Witkowska, D. (2014). An application of hedonic regression to evaluate prices of polish paintings. *International Advances in Economic Research, 20*(3), 281-293. doi:10.1007/s11294-014-9468-x

Zhang, Y., Hua, X., & Zhao, L. (2012). Exploring determinants of housing prices: A case study of Chinese experience in 1999-2010. *Economic Modelling, 29*(6), 2349-2361. doi:10.1016/j.econmod.2012.06.025