Journal of Behavioral and Applied Management
2024, Vol. 24, No. 3, 197–204
https://jbam.scholasticahq.com/

Teaching Multiple Regression Using Real Estate Data: An Experiential Exercise

Kenneth M. York
Oakland University

T.J. Wharton
Oakland University

In the era of "Big Data," data are often easy to obtain, and key management skills include analysis of the data and interpreting the statistical tests used to get an answer to research questions. In this experiential exercise, students collect and analyze Zillow.com data to determine what factors predict the list price of houses. The exercise combines data collection and data analysis and interpretation using multiple regression, to demonstrate two key skills needed by students studying marketing or management research. From a teaching standpoint the data is dynamic, allowing the exercise to be used over multiple sections and semesters with different results each time it is used.

"Let us be lovers, we'll marry our fortunes together. I've got some real estate here in my bag…." Simon and Garfunkel, America

A popular genre of television shows follows people as they look at multiple houses and then decide which house to buy. Examples of these shows are My Lottery Dream Home, Love It or List It, and House Hunters (which has had over 1900 episodes). The houses differ on a number of factors; the shows often give viewers data about the houses, reporting the number of bedrooms, the number of bathrooms, the square feet of the house, and the list price. But are the list prices of the houses actually based on these variables?

A statistical test requires data, and in the past data were often difficult to obtain or available to only a few people. However, in our modern connected world, organizations now have access to large amounts of data about their industry, their customers, and their own employees. Until recently, data about residential real estate was available only to real estate agents. Now this same data is publicly available to everyone on websites such as Realtor.com, RE/MAX.com, and Zillow.com.

In this exercise, students learn about data analysis by collecting data on houses listed for sale on a real estate website, to test whether the list price of a house can be predicted based on the data available. Students may have learned about correlation and regression from a statistics class, but dealt only with textbook homework problems to solve. The advantage of this exercise is that it is more "real"istic. They may have seen one of the many real estate shopping shows like House Hunters, they may have bought or sold a house, or they may be thinking about buying a home in the future. Instead of just being given the data and tasked with doing data analysis, students collect and analyze their own data from a real estate website where the data change on a daily basis.

Theoretical Foundation

A key management skill is analysis of data to answer a research question (Ashraf, 2020) using an existing dataset which can help a researcher obtain results quickly (Doolan, Winters, & Nouredini, 2017). A manager starts with a question, collects and analyzes data, then interprets the results to answer the question. Students need to learn how to answer a research question by using a statistical test, and interpreting what the results of the test mean. Although this experiential exercise uses real estate data, the data analysis knowledge and skills used here are the same knowledge and skills used in other industries and other functional area of business, whenever an organization needs to make a data-driven decision.

How can students best be taught these skills? Experiential exercises in which students learn-by-doing have long been used as an alternative to lectures. More than four decades ago, Certo (1976, p. 113) described the use of experiential exercises as "a growing trend in management education," where individuals actively engage in performance of a goal-oriented task and learn both from the performance itself and from getting feedback on their performance. More recently, Kolb and Kolb (2005, p. 194) suggested that to improve learning in higher education, the primary focus should be on "engaging students in a process that best enhances their learning – a process that includes feedback of the effectiveness of their learning efforts."

According to Kolb's Experiential Learning Theory (Lu, Jia, Gong, & Clark, 2007), student learning is based on two dimensions, how a person understands information (through concrete experience or abstract conceptualization), and how a person processes information (through active experimentation or reflective observation). Experiential exercises help students learn though guided experience and reflection on what they have learned, directed by the teacher. Burch, Giambatista, Batchelor, Burch, Hoo-

Article citation: York, K. M., & Wharton, T.J. (2024). Teaching multiple regression using real estate data: An experiential exercise. *Journal of Behavioral and Applied Management, 24*(3), 197-204.

ver, and Heller's (2019) meta-analysis of studies on experiential learning showed that students experienced superior learning outcomes when experiential pedagogies were employed, with learning outcomes almost a half standard deviation higher in classes employing experiential learning pedagogies versus traditional learning environments.

The classroom culture for experiential learning creates specific roles for the Teacher and the Student, which differs from a traditional classroom (Wurdinger & Carlson, 2009):

Teacher's Role: Act as a guide, allowing students to make mistakes and learn from them; provide students with the freedom to experiment to discover solutions to the problems they have been given; and provide students with resources and information when they get stuck so they can continue to move forward with their learning.

Student's Role: Have freedom in the classroom as long as they are moving forward in the learning process; may need a series of trial and errors as they attempt to complete the assignment; and understand that the problem-solving process is as important as the content they are learning.

Experiential learning has been used to teach a wide variety of skills and concepts, such as group problem-solving (Hedges & Pedigo, 2002), creativity (Armstrong, 1999), diversity (Merta, Stringham, & Ponterotto, 1988), attribution theory (Paglis, 2008), leadership (Reilly, & Ehlinger, 2007), managing intergroup behavior (Hunsaker, 2004), team building (Yanson & Mann, 2013), change management (McDonald & Mansour-Cole, 2000), and conflict management (Anakwe & Purohit, 2006). The active-learning technique encourages students to become more involved with the course content through application of theory to real-life situations. Freeman (2020) described a course at Carnegie Mellon University that provides a circumscribed opportunity for statistical practice, and which showed clear gains in the students' knowledge of, and attitudes toward, doing statistics. Shatzkin, Chen, Greisler, & Kratz (2023) described the development of an introductory statistics course, where students received instruction on the Define, Collect, Organize, Visualize, and Analyze (DCOVA) framework to integrate data with business-related decision-making. They found that this approach positively impacted the students' ability to apply statistics, as well as their attitudes regarding their statistics abilities. Anderson and Williams (2019), created a semester-long project in which students asked and attempted to answer questions about themselves by collecting and analyzing data, using applied descriptive analytics, regression, and data visualization techniques to answer their posed questions. Hakeem (2010), examined the effect of a semester-long, active-learning project on student learning in a business statistics course and found that compared with the traditional lecture format, the active-learning technique encouraged students to become more involved with the course content through application of theory to real-life situations. Thus, the active-learning students gained a better understanding of statistics than did the students in the traditional class. In brief, the purpose of experiential learning is to go beyond just giving students information, students learn through experience: The "Learn-Do" Model.

The Exercise

Learning Objectives

Learning Objective 1: Learn how to collect secondary data that can be used to find answers to research questions.

Learning Objective 2: Learn how to use regression analysis to answer research questions.

Learning Objective 3: Describe the limitations of the data analysis, and suggest further analysis that might be done to get a better answer to the research questions.

Learning Objective 4: Build critical thinking skills by interpreting the meaning of the results of regression analysis.

Learning Objective 5: Describe the ethical issues of a realtor (but not a customer) knowing that the regression model indicates a house is "overpriced," or using data on crime rates or school district quality to predict list price.

Instructions for Running the Exercise

In this experiential exercise, student teams collect real estate data from Zillow.com to answer the research questions, "Can you predict the list price of houses from the number of bedrooms, the number of bathrooms, and the square feet?" and "What factors drive the list price of a home for sale?" The Step by Step Instructions for the students are given in Appendix A.

The instructor can give the teams zip codes to search for listings from the area around the college/university so that the students are somewhat familiar with the neighborhoods, or the instructor can give the teams zip codes from different cities or regions of the country, where the various zip codes will have very different housing markets, (e.g., Beverly Hills, CA, 90210; Detroit, MI, 48210; Anchorage, AK, 99505; Boston, MA, 02110; Grand Island, NB, 68801).

For example, in comparing and contrasting hursing markets in Beverly Hills, to Detroit, and Grand Island, Beverly Hills has median age of 47.5, 10.2% poverty rate, median household income of $116,771, homeownership rate 42.9%, a median property value of $2.0 million, and the five largest ethnic groups are White (Non-Hispanic) (73.2%), Asian (Non-Hispanic) (12.1%), Two+ (Non-Hispanic) (4.66%), White (Hispanic) (2.81%), and Two+ (Hispanic) (2.32%), whereas Detroit has a median age of 35, 31.5% poverty rate, median household income of $37,761, homeownership rate 48.8%, a median property value of $66,700, and the five largest ethnic groups are Black or African American (Non-Hispanic) (77.4%), White (Non-Hispanic) (10.1%), Other (Hispanic) (3.7%), Two+ (Non-Hispanic) (2.58%), and White (Hispanic) (2.14%). Grand Island has a median age of 35, 13.7% poverty rate, median household income of $59,061, home ownership rate 58.6%, a median property value of $183,600, and the five largest ethnic groups are White (Non-Hispanic) (59.2%), Other (Hispanic) (16.1%),

White (Hispanic) (9.69%), Two+ (Hispanic) (6.68%), and Black or African American (Non-Hispanic) (3.51%). (DataUSA, 2004). Changing the locale of the assignment can clearly lead to an enrichment of class discussion.

To be prepared for this exercise, students should have taken an introductory course in statistics which included how to do regression analysis using Microsoft Excel or a statistics package such as IBM's SPSS or Grafiti's SYSTAT. Based on this prerequisite, the exercise is well suited for an upper-level undergraduate course focusing on research and data analysis (e.g., marketing and/or management research).

Place students into teams of 4 to 6 students and assign each team a different zip code from which to collect data on houses for sale on the Zillow website. The instructor may also give students a link to an episode of House Hunters, to make the problem more realistic. Ask the students to write a report on their results, including:

1) A description of the type of homes their zip code contains (i.e., the mean number of bathrooms, bedrooms, and square feet, and mean list price);

2) The statistical tests they did and the results they obtained; and

3) Their conclusions about the research questions.

When students do their data analysis, they should do a multiple regression analysis using number of bedrooms, number of bathrooms, and square feet as independent variables and list price as the dependent variable. To test whether the model is a good predictor of list price, students should report the Multiple R, and the probability value for the model (what Excel calls "Significance F"). If the p-value is .05 or less, list price is predictable from this set of variables. The Multiple R is the multiple correlation of the set of independent variables with the dependent variable; it is an index number, not a percent. The Multiple Correlation Squared is the variance accounted for in list price by the set of independent variables, this is a percent. The Adjusted Multiple Correlation Squared is the variance accounted for in list price adjusted for the number of independent variables in the model. Each of the independent variables can be tested for statistical significance. If the independent variables are highly correlated with each other there is multicollinearity. Typically, square feet is correlated with the number of bedrooms and the number of bathrooms (and will therefore have a low Tolerance). A sample regression analysis done in Excel is shown in Appendix B.

Predicted list prices of houses can be generated using the regression equation, with the general form $Y = mX + b$, where the predicted list price equals the values of each independent variable(s) times the regression coefficients for the variable(s) plus the intercept. Using the regression output in Appendix B, a house with 3 bedrooms, 2 bathrooms, and 1800 square feet has a predicted list price of $389,667. If the predicted list price of the house is more than the actual list price, the house is listed "under market," according to the regression model.

The key advantage of the exercise is that the data are constantly changing, so students cannot reuse data from a previous semester, the data and the results will be different. Also, student teams may get very different results because each zip code may be a different housing market, i.e., one zip code may be an established neighborhood with moderately priced houses, another zip code may be an upscale neighborhood with larger homes. The main disadvantage of the exercise is that the data can change even daily, so there can be some small differences between what the students turn in and the instructor's results—in the time between when the students and the instructor collected the data some listings may be added, while others are removed. List price will usually be predictable from these three independent variables, but the best predictor of list price can vary by zip code. Typically, the best predictor of list price is square feet, although sometimes the number of bathrooms is the best predictor.

Debriefing and Discussion Questions

1. Houses vary in their number of bedrooms, number of bathrooms, and square feet. Is this what realtors use to determine what the list price should be? Although the model with these independent variables fits these data well, the model does not perfectly predict the list price, so the model is "underspecified." Collecting more data on additional variables may improve the model. The researcher must then make a practical decision as to when the cost of collecting more data is not worth the increase in the fit of the model to the data.
2. When realtors say that a house is "overpriced" or "underpriced," what is this based on? The regression equation could be used to generate a predicted list price for a house based on number of bedrooms, number of bathrooms, and square feet, to see whether a given house is over- or under-predicted (i.e., a good value or pricey). Flipping is a real estate strategy that involves buying homes, renovating them, and quickly selling them for a profit. Flippers might use the regression model to improve their profits by finding a house that is underpriced, and then renovating the house and setting the list price higher than other houses in the neighborhood.
3. Do these predictors of list price differ by zip code? If you had data from different zip codes, what statistical test could you do to determine whether neighborhood is also a factor in the list price of a home? Zip code be used as an independent variable by using dummy variables for zip code.
4. What should be done about "odd" data points? For example, one house listed for sale in a zip code has much less square footage (really a cottage) than the average house but has a lakefront and acreage. Should that data be kept in the dataset or deleted ("cleaned") because it is an outlier? This often happens in real world datasets. Although there is ongoing debate (the Mendel-Fisher controversy), Gregor

Mendel may or may not have cleaned his data by omitting results that did not fit his theory well (Kalina, 2014).

5. What other variables might predict the list price of a home? On real estate shows, potential buyers often talk about noise level (near a busy street or in a cul-de-sac) or how big the lot is (or acreage), recent renovations, whether there is a pool, how long their commute to work would be, the quality of school district, or how much the property taxes or homeowners association (HOA) fees are. These variables could be added to the model, where the data are available.
6. In any statistical analysis there is the issue of statistical significance versus practical significance. Even a very small effect can be statistically significant, if there are enough observations. In the example in Appendix B, there is statistical significance, $F(2, 27) = 23.88$, $p < .001$, and there is practical significance because the variance accounted for in list price by the set of three independent variables is R Square = .75, 75% of the variance is accounted for.
7. Real estate websites show list price, how does that differ from sale price? Sometimes "motivated" sellers accept an offer that is less than the list price, and sometimes there is a "bidding war" and the sale price is higher than the list price. Because of this, the regression model for the dependent variable sale price may not fit the data as well for the dependent variable list price. If houses that have been on the market longer have more motivated sellers, then adding the variable "days on Zillow" may improve the model.
8. Some of the variables may be misleading. For example, an owner might increase the number of bedrooms by splitting one bedroom into two, with the same square footage. Two houses may have the same number of square feet, but one house has a better layout or flow and therefore has more "usable space." This is a good example of where it is difficult to measure what you really want to measure, but easier to measure something else that is correlated with it.
9. What are the ethical implications of using crime rates or school district quality as predictors of list price? What are the ethical implications of a realtor using a regression model to predict list price, when potential buyers or sellers do not have the data that the regression analysis is based on?

Variations on the Exercise

Student Presentations of Results

Instead of having students write a report of their results, the instructor can have the students make presentations, and show the results they obtained and conclusions they made about the research questions. The advantage of having students present their results is that they can see the results of other teams (and different neighborhoods can give very different results) and respond to questions about the results they found.

Student Reflection or Journaling

After the students or teams have completed their analysis and made their conclusions, students can be asked to reflect on their process, including the variables they used in their model, and what the results mean. This can be especially useful when students have worked in teams, they may reflect on the soft skills they developed while they collaborated, and what analytics skills they developed when they analyzed the data and drew their conclusions.

Adjusted R Square

Zillow has more data about houses for sale in addition to number of bedrooms, number of bathrooms, and square feet. It also has data on the year the home was built, the HOA fee, property taxes, the lot size, etc. Students could choose which of these variables to include in the regression model, so that each student or team creates a different regression model with a different R Square. The instructor could hold a contest among students or teams to find the highest R Square and the highest Adjusted R Square. This could be the basis of a discussion of the diminishing returns from adding additional variables to the model, as shown by the difference between R Square and Adjusted-R Square.

Instructor Data Entry

The exercise requires that students enter the data themselves, because the first step in data analysis is collecting, entering, and cleaning the data. Alternatively, the instructor can collect the data from Zillow and post it for students so that students do not spend time entering data. The advantages of this variation are that the instructor and the students have the same data and should get the same statistical output, and the students are focused only on data analysis and interpretation.

Incorporating Other Issues in Regression Modeling

If the students have a good background in statistics and data analysis, the instructor may also want to discuss other issues related to the regression modeling process such as testing the assumptions of the model, prediction versus estimation, and dealing with multicollinearity. Students could analyze the data using other forms of regression, such as Logistic Regression to predict whether the houses are above or below the median price for houses in a zip code, or Cox Proportional Hazards Regression to predict how long a home has been on the market ("days on Zillow"). Ridge Regression may be used when there is much multicollinearity, and typically the number of bedrooms, number of bathrooms, and square feet are highly correlated. Stepwise Regression is tempting to do because the "best" variable is entered in the model first and additional variables are added until some criterion is reached where adding an additional variable no longer improves the model. But the results can be unstable because the method finds the optimal variables with this sample and with these data, but the optimal variables could change with the next sample, sometimes called "overfitting" the model. A detailed understanding of mul-

tivariable regression is essential for correct interpretation of these regression models, and a statistics package that can do these other types of regression.

Regression Modeling and Management Education

It is important to teach multiple regression as part of a modern business education because data literacy has an ever-increasing importance, not just in management, but also in accounting, finance, operations management, supply chain management, human resources, and marketing. Students need to build skills in using big data to answer questions, using such tools as Data Mining, Classification Modeling, Time Series, Predictive Analytics, and Regression Analysis. Based on their survey of Canadian and U.S. marketing respondents with experience in big data in marketing, Haverila, Haverila, Mohiuddin, and Su (2022), found that technology and information quality are related to the market and financial performance, and the level of deployment had a significant impact on both the technology and information quality in big data marketing analytics. According to Goldstein, Spatt, and Ye (2021), big data is revolutionizing the finance industry and has the potential to significantly shape future research in finance, and pushing the frontier on fundamental questions across areas in finance—including corporate finance, market microstructure, and asset pricing. In their literature review of big data and real estate, Oluwunmi, Role, Akinwale, Oladayo, and Afolabi (2019), observed that the impact of big data includes digitization of records, information on user preferences, sensor information on the urban environment and sensor information on movement. The authors conclude that the relevance of big data to the real estate profession cannot be over-emphasized.

However, using "Big Data" can be intimidating for students, so it is helpful if they can start with a simple problem with a limited dataset, then move on to solving more complex problems using more advanced methods. In this experiential exercise, students are given a practical problem that they are likely to be familiar with: What is the list price of houses based on?

When using regression to find whether number of bathrooms, number of bedrooms, and square feet of a house predicts list price, students will generally find that the best predictor is square feet. But different zip codes can have very different housing markets, and the best predictor in one housing market may not be the best predictor in another. For example, in a zip code listing for sale modest homes (Means: List Price = $213,757, Bedrooms = 3.3, Bathrooms = 2.2, Square feet = 1,741), list price is predictable from the set of three independent variables (Multiple R = .82, $F(3, 20) = 12.03$, $p < .001$), and there is one significant predictor of list price (Square feet t Stat = 2.68, $p = .016$). But in a more upscale zip code (Means: List Price = $1,613,161, Bedrooms = 3.9, Bathroom = 4.7, Square feet = 5,568), list price is equally predictable from the set of three independent variables (Multiple R = .82, $F(3, 49) = 31.49$, $p < .001$), but Bathrooms is the one significant predictor of list price (t Stat = 2.72, $p = .009$). Modeling list price across different housing markets can be difficult, the data sampled affect the best fitting model. Inescapably, realtors need to know the different housing markets and what drives the List Price in each market.

The best regression models will have independent variables that have high correlations with the dependent variable, and low inter-correlations among the set of independent variables, as shown in Figure 1. If the independent variables are correlated with each other, then they may have shared variance with the dependent variable, so that each independent variable by itself predicts the dependent variable, but when taken as a set in the regression analysis, one independent variable "gets" most of the variance in the dependent variable, and there is little variance left to predict for the other independent variables, as shown in Figure 2. In the example in Appendix B, the univariate correlations of the independent variables with list price are Bedrooms = .71, Bathrooms = .79, and Square feet = .86. When three separate regression analyses are done, all three independent variables are significant predictors of list price: Bedrooms $F(1, 26) = 27.14$, $p < .05$; Bathrooms $F(1, 26) = 44.45$, $p < .05$; Square feet $F(1, 26) = 75.16$, $p < .05$. But the inter-correlations among the three independent variables range from .79 to .88, and when all independent variables are used, Square feet is the only significant predictor (t stat 3.05, $p < .05$). The shared variance effect can be seen in the variance accounted for in list price. In the regression model using just Square feet (the best predictor of list price), the R Square is .74, and in the three-variable model, the R Square is .75, a modest increase despite the addition to the model of two independent variables that are also correlated with list price, because the independent variables are highly correlated with each other.

Figure 1

Visual Representation of Multiple Regression with Three Uncorrelated Independent Variables

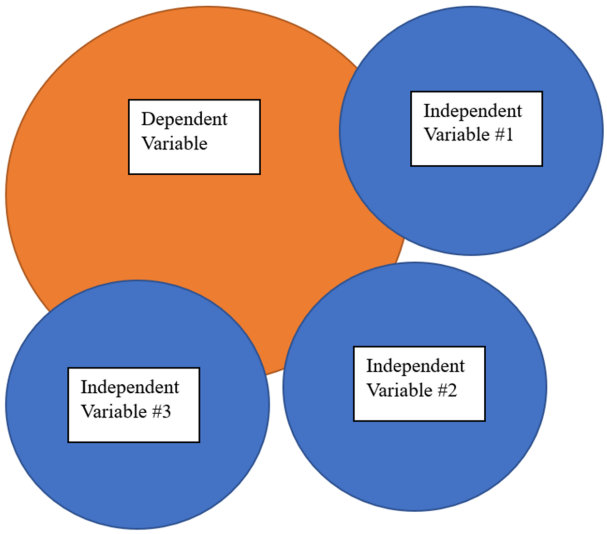
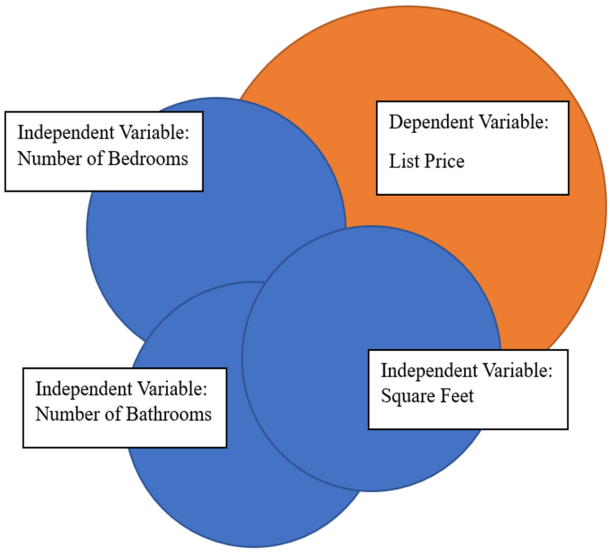


Figure 2

Visual Representation of Multiple Regression with Three Intercorrelated Independent Variables



Many variables could be used to test a regression model to predict list price, but which variables should be used? The best answer is that the variables to include in the model are driven by theory or seem to make logical sense. In the case of list price, the show House Hunters always gives number of bedrooms, number of bathrooms, and square feet, and sometimes gives others (e.g., commuter time, street noise, new build/fixer-upper). Absent any theoretical or logical reason to include some variables in the model, a stepwise approach could be taken (a kind of dust bowl empiricism). There are statistics packages which allow you to do stepwise regression quickly and easily, and the program will add independent variables based on criteria that the user selects, such as variance accounted for, or number of independent variables. One method starts with the independent variable that has the highest correlation with the dependent variable, and adds independent variables to the model one at a time that add the most to the multiple correlation or other criteria for model fit (forward selection), another method starts with all the variables in the model and removes the independent variable that reduces the multiple correlation or other criteria the least (backward elimination), and yet another method iteratively removes and adds independent variables to get the best model (bidirectional elimination). Stepwise methods reduce the role of theory in model building; the program just builds a model out of the data it is given. The next step is to test the model on a new set of data, not the data that was used to develop the model, i.e., build the list price model on a set of zip codes in one metropolitan area, then test how well the model works on the data from another metropolitan area.

In our era of big data where data is cheap and easy to get, researchers may be tempted to put into the model all the data that they have, regardless of whether it makes sense to include it. Stepwise methods are used as a kind of data reduction, to find which variables are important. They might also find spurious correlations, meaningless in theory and in the real world. The data analyst must balance accuracy and simplicity: more variables may make the model more accurate, but too many variables in the model make it difficult to use. On the other hand, a dust bowl empiricist would say that theory may make researchers' focus too narrow, so that they do not include a variable which could turn out to be a good predictor. For example, in the mid 1800's, John Snow collected data on the cause of cholera which included which water company supplied each home with water (driven by his theory that cholera was caused by something in the water, although it was too small for him to see), which turned out to be a good predictor. A better predictor of cholera deaths would have been the number of Vibrio cholerae bacteria in the water, but he did not have the technology to collect that data (Goldstein, & Goldstein, 1978).

Conclusion

Data do not speak for themselves. Analyzing data to answer questions helps students understand what the data mean, and allows them to explore ways that the data analysis could be improved to get better answers to the research questions under study. This experiential exercise ties together the process of data collection with the analysis and interpretation of the data collected. The students should come away from this exercise with an appreciation of issues related to data collection as well as an understanding of the issues in the use of multiple regression as an analytical tool.

Appendix A

Handout for Students – Step by Step Instructions

When you want to find out whether a set of variables predicts a variable of interest, you use Multiple Regression: multiple independent variables and one dependent variable. Can the list prices of houses be predicted from data about the houses, such as the number of bedrooms, the number of bathrooms, and the square feet?

Zillow (https://www.zillow.com) claims to be the most-visited real estate website in the United States, and offers data for selling, buying, renting, and financing houses. The data from the website can be used to test whether the list prices of houses can be predicted from the data Zillow provides.

Research Questions

Can you predict the list price of houses from the number of bedrooms, the number of bathrooms, and the square feet?

What is the best predictor of list price?

What other factors might you consider in predicting the list price of a house?

To Do

1. Collect data on houses for sale on Zillow.com in the zip code assigned to you. On the Zillow web page, enter the zip code assigned to your team. Select the option "For Sale," and no other filters. Ignore listings for lots only. This will give you data on all the houses for sale in that zip code.
2. For each house listed, use an Excel spreadsheet to enter the data for the number of bedrooms, the number of bathrooms, the square feet, and list price. There may be more than one page of listings. As you start to enter data, it will look something like that:

	A	B	C	D	E	F	G	H	I	J
1	**Zillow**	**ZipCode:**		**48309**	**Rochester Hills**		**Zillow**			
2	*Address*	*bds*	*ba*	*sqft*	*ListPrice*		Address	House address		
3	3226 Fantail Dr	4	4	2,190	$429,900		bds	Number of bedrooms		
4	136 Waltonshire Ct	4	5	2,931	$515,000		ba	Number of bathrooms		
5	More....						sqft	Square feet		
6							ListPrice	Offer price		
7										
8							Zip Codes			
9							48309	Rochester Hills		
10							48043	Mount Clemens		
11							48362	Lake Orion		
12							48304	Bloomfield Hills		
13							48317	Shelby/Utica		
14										
15							https://www.zillow.com/			
16										

3. To get descriptive statistics on the houses, use the =AVERAGE function to find the means of the variables, and the =STDEV function to get the standard deviation.
4. Analyze the data to answer the research questions. Use the Data Analysis Tool Regression. The dependent variable (X Range) is List Price, the independent variables (Y Range) are Bedrooms, Bathrooms, and Square Feet.
5. Write a short memo reporting what you did: 1) Describe the type of homes the zip code contains, i.e., the mean and standard deviation for number of Bedrooms, number of Bathrooms, Square Feet, and List Price, 2) The Regression output you obtained, the Multiple R, *p*-value (Significance F) and R Square for the regression model, and the *p*-values of the independent variables, 3) Your answers to the research questions, and 4) Any other variables you might search for on Zillow, such as the school district, year built, fireplace, finished basement, days listed, lot size, etc., that might improve your prediction of List Price.
6. Submit a Word file (memo) and the Excel file (data).

Appendix B

Sample Excel Data Analysis and Interpretation of the Results

Research Questions

1. Can you predict the list price of houses from the number of bedrooms, the number of bathrooms, and the square feet? Yes, because the multiple correlation is statistically significant, R = .87, p < .05, and the variance accounted for was 75%.
2. What is the best predictor of list price? Square feet, of the three independent variables, only Square Feet is statistically significant (p < .05). Refer to Picture 2 at the end of the document.

	A	B	C	D	E	F	G	H	I	J
1	**Zillow**	**ZipCode:**		**48309**	**Rochester Hills**		**Zillow**			
2	*Address*	*bds*	*ba*	*sqft*	*ListPrice*		Address	House address		
3	3226 Fantail Dr	4	4	2,190	$429,900		bds	Number of bedrooms		
4	136 Waltonshire Ct	4	5	2,931	$515,000		ba	Number of bathrooms		
5	More....						sqft	Square feet		
6							ListPrice	Offer price		
7										
8							Zip Codes			
9							48309	Rochester Hills		
10							48043	Mount Clemens		
11							48362	Lake Orion		
12							48304	Bloomfield Hills		
13							48317	Shelby/Utica		
14										
15							https://www.zillow.com/			
16										

References

Anakwe, U. P., & Purohit, Y. S. (2006). Using conflict-management surveys to extricate research out of the "Ivory Tower": An experiential learning exercise. *Journal of Management Education, 30*(3), 501-524. https://doi.org/10.1177/1052562905280089

Anderson, J. S., & Williams, S. K. (2019). Turning data into better decision making: Asking questions, collecting and analyzing data in a Personal Analytics Project. *Decision Sciences Journal of Innovative Education, 7*(2), 126-145. http://dx.doi.org/10.1111/dsji.12173

Armstrong, A. (1999). Teaching creativity: an "EEE" experiential exercise. *Journal of Management Education, 23*(2), 174-179. https://doi.org/10.1177/105256299902300206

Ashraf, F. (2020). Rise of big data analytics and the role of human resource management. *American Journal of Management, 20*(2), 90-95. https://www.proquest.com/scholarly-journals/rise-big-data-analytics-role-human-resource/docview/2437939511/se-2

Burch, G. F., Giambatista, R., Batchelor, J. H., Burch, J. J., Hoover, J. D., & Heller, N. A. (2019). A meta-analysis of the relationship between experiential learning and learning outcomes. *Decision Sciences Journal of Innovative Education, 17*(3), 239-273. https://doi.org/10.1111/dsji.12188

Certo, S. C. (1976). The experiential exercise situation: A comment on instructional role and pedagogy evaluation. *Academy of Management Review, 1*(3), 113-116. https://doi.org/10.5465/amr.1976.4400877

DataUSA (2024). Beverly Hills, California https://datausa.io/profile/geo/beverly-hills-ca?redirect=true; Detroit, Michigan https://datausa.io/profile/geo/detroit-mi/; Grand Island, Nebraska https://datausa.io/profile/geo/grand-island-ne/

Doolan, D. M., Winters, J., & Nouredini, S. (2017). Answering research questions using an existing data set. *Medical Research Archives, 5*(9), 1-14. https://esmed.org/MRA/mra/article/view/1543

Excel (n.d.). Microsoft Excel, https://microsoft.com.

Freeman, P. E. (2020). Facilitating authentic practice for early undergraduate statistics students. *The American Statistician, 75*(4), 433-444. https://doi.org/10.1080/00031305.2020.1844293

Goldstein, M., & Goldstein, I. F. (1978). *How we know: An exploration of the scientific process*. New York.

Goldstein, I., Spatt, C. S., & Ye, M. (2021). Big data in finance. *The Review of Financial Studies, 34*(7), 3213-3225. https://doi.org/10.1093/rfs/hhab038

Hakeem, S. A. (2001). Effect of experiential learning in business statistics. *Journal of Education for Business, 77*(2), 95-98. https://doi.org/10.1080/08832320109599056

Haverila, M., Haverila, K. C., Mohiuddin, M., & Su, Z. (2022). The impact of quality of big data marketing analytics (BDMA) on the market and financial performance. Journal of *Global Information Management, 30*(1), 1-21. https://doi.org/10.4018/JGIM.315646

Hedges, P., & Pedigo, K. (2002). What's happening in your neighborhood? An experiential exercise for teams. *Journal of Management Education, 26*(3), 322-333. https://doi.org/10.1177/10562902026003007

Hunsaker, P. L. (2004). Learning to manage intergroup dynamics in changing task environments: An experiential exercise. *Journal of Management Education, 28*(6), 790-805. https://doi.org/10.1177/1052562904264230

Kalina, J. (2014). Gregor Mendel, his experiments and their statistical evaluation. *Acta Musei Moraviae, Scientiae Biologicae, 99*, 87-99. http://www.ammbiol.com/fileadmin/user_upload/08kalina.pdf

Kolb, A. Y., & Kolb, D. A. (2005). Learning styles and learning spaces: Enhancing experiential learning in higher education. *Academy of Management Learning & Education, 4*(2), 193-212. https://doi.org/10.5465/amle.2005.17268566

Lu, H., Jia, L., Gong, S. H., & Clark, B. (2007). The relationship of Kolb learning styles, online learning behaviors and learning outcomes. *Journal of Educational Technology & Society, 10*(4), 187-196. https://www.jstor.org/stable/jeductechsoci.10.4.187

McDonald, K. S., & Mansour-Cole, D. (2000). Change requires intensive care: An experiential exercise for learners in university and corporate settings. *Journal of Management Education, 24*(1), 127-148. https://doi.org/10.1177/105256290002400108

Merta, R. J., Stringham, E. M., & Ponterotto, J. G. (1988). Simulating culture shock in counselor trainees: An experiential exercise for cross-cultural training. *Journal of Counseling & Development, 66*(5), 242-245. https://doi.org/10.1002/j.1556-6676.1988.tb00858.x

Oluwunmi, A. O., Role, B. A., Akinwale, O. M., Oladayo, O. P., & Afolabi, T. O. (2019). Big data and real estate: A review of literature. *Journal of Physics: Conference Series, 1378*(3), 32015. https://doi.org/10.1088/1742-6596/1378/3/032015

Paglis, L. L. (2008). The "It's Not My Fault!" exercise: Exploring the causes and consequences of managers' explanations for poor performance. *Journal of Management Education, 32*(5), 613-628. https://doi.org/10.1177/1052562908314528

Reilly, A. H., & Ehlinger, S. (2007). Choosing a values-based leader: An experiential exercise. *Journal of Management Education, 31*(2), 245-262. https://doi.org/10.1177/1052562906292565

Shatzkin, M. P., Chen, W., Greisler, D. S., & Kratz, C. (2023). The Data to Decision Project: An experiential approach to teaching undergraduate business statistics. *Journal of Education for Business, 99*(3), 187–194. https://doi.org/10.1080/08832323.2023.2286444

SPSS (IBM, n.d.). IBM SPSS Statistics, https://www.ibm.com/products/spss-statistics.

SYSTAT (n.d.). Grafiti SYSTAT, https://grafiti.com/systat/.

Wurdinger, S. D., & Carlson, J. A. (2009). *Teaching for experiential learning: Five approaches that work*. R&L Education.

Yanson, R. C., & Mann, M. J. (2013). Building towers to build better teams: An innovative approach to teaching team building. *Journal of Organizational Behavior Education, 6*, 41-58. Retrieved from: https://www.thecasecentre.org/products/view?id=120972

Zillow (n.d.). https://www.zillow.com.

Kenneth M. York (york@oakland.edu)

T.J. Wharton (wharton@oakland.edu)
