House Price Analysis
Spring 2022

Background

Using the House Prices data, determine the factors that influence the price of a home.

Resources

Use the dataset SCM 651 Homework 1 House Prices spreadsheet.

House Price Data

ID	unique identifier
Price	price of home in dollars
SqFt	square feet of area of home
Bedrooms	number of bedrooms
Bathrooms	number of bathrooms
Offers	number of offers received on home before sale
Brick	Yes/No on brick construction
Neighborhood	location of home in east, north or west quadrants of city

Assignment

What's due:

For this assignment, you will enter your answers on BlackBoard. The assignment is due no later than **8:00 AM, Friday, February 25, 2022**.

The following instructions will step you through problems and list the questions that you will be asked on BlackBoard. Each problem will include several questions for that problem. Each question is worth 2 points. There are 50 questions for a total of 100 points.

This is a group assignment but will be scored individually. You may work with your group on the assignment, but each student must enter your answers into BlackBoard. Scores will be automatically posted after the deadline. Assignments that are submitted late will receive a score of zero. Note: you must enter answers exactly as requested. Misspelled answers or errors in numbers will be scored as wrong.

Questions:

Problem #1: Pivot Table

Create a pivot table with Neighborhood for row labels, Brick for column labels, Sum of Price for values.

Problem #1 questions:

1. For Brick houses (Brick-Yes), which Neighborhood (East, North, or West) had the highest sales (sum of price)?
2. For non-Brick houses (Brick-No), which Neighborhood (East, North, or West) had the highest sales (sum of price)?
3. For Brick houses, which Neighborhood (East, North, or West) had the lowest sales (sum of price)?
4. For non-Brick houses, which Neighborhood (East, North, or West) had the lowest sales (sum of price)?
5. What are the sales (sum of price) for Brick houses in the East Neighborhood (enter the exact number, no commas)?

Problem #2: Pivot Table

Create a pivot table with Neighborhood for row labels, Brick for column labels, Average of SqFt for values.

Problem #2 questions:

6. For Brick houses, which Neighborhood (East, North, or West) had the highest average square feet?
7. For non-Brick houses, which Neighborhood (East, North, or West) had the highest average square feet?
8. For Brick houses, which Neighborhood (East, North, or West) had the lowest average square feet?
9. For non-Brick houses, which Neighborhood (East, North, or West) had the lowest average square feet?
10. What is the average square feet for non-Brick houses in the East Neighborhood (enter the whole number, no decimal places)?

Problem #3: Descriptive Statistics

Use descriptive statistics for the following questions.

Problem #3 questions:

11. What is the mean price for a house (enter number with no decimals)?
12. What is the median price for a house (enter number with no decimals)?
13. What is the minimum sq ft for a house (enter number with no decimals)?
14. What is the maximum sq ft for a house (enter number with no decimals)?
15. What is the minimum number of bedrooms (enter number with no decimals)?
16. What is the maximum number of bedrooms (enter number with no decimals)?
17. What is the minimum number of bathrooms (enter number with no decimals)?
18. What is the maximum number of bathrooms (enter number with no decimals)?
19. What is the minimum number of offers (enter number with no decimals)?
20. What is the maximum number of offers (enter number with no decimals)?

Problem #4: Correlation

Use correlation for the following questions. Create the correlation table only for Price, SqFt, Bedrooms, Bathrooms, and Offers.

Problem #4 questions:

21. When looking at correlations with Offers, which factor (Price, SqFt, Bedrooms, Bathrooms) has the strongest correlation?
22. When looking at correlations with Offers, which factor (Price, SqFt, Bedrooms, Bathrooms) has the weakest correlation?
23. When looking at all variables (Price, SqFt, Bedrooms, Bathrooms, Offers), what is the value of the strongest correlation (enter number to two decimal places x.xx)?
24. When looking at all variables (Price, SqFt, Bedrooms, Bathrooms, Offers), what is the value of the weakest correlation (enter number to two decimal places x.xx)?
25. Which variable (SqFt, Bedrooms, Bathrooms, Offers) has a negative correlation with Price?

Problem #5: Linear Regression

Use regression for the following questions. Create the regression where Price is the dependent variable (Y) and SqFt, Bedrooms, Bathrooms, and Offers are the independent variables (X's). Use multiple regression to solve this problem.

Problem #5 questions:

26. Is the equation statistically significant (Yes or No)?
27. What is the R^2 (enter number to four decimal places 0.xxxx)?
28. What is the intercept (enter number with no decimals)?
29. What is the coefficient of SqFt (enter number with no decimals)?
30. What is the coefficient of Bedrooms (enter number with no decimals)?
31. What is the coefficient of Bathrooms (enter number with no decimals)?
32. What is the coefficient of Offers (enter number with no decimals)?
33. Is the intercept statistically significant (Yes or No)?
34. Is the coefficient of SqFt statistically significant (Yes or No)?
35. Is the coefficient of Bedrooms statistically significant (Yes or No)?
36. Is the coefficient of Bathrooms statistically significant (Yes or No)?
37. Is the coefficient of Offers statistically significant (Yes or No)?
38. If one bedroom is added to a house, how much value would you expect it to add to the price of the house (enter number with no decimals)?
39. If one bathroom is added to a house, how much value would you expect it to add to the price of the house (enter number with no decimals)?
40. How much of the change of the price of a house, <u>in percent</u>, is explained by the change in the variables SqFt, Bedrooms, Bathrooms, and Offers (enter as percent with two digits, no decimal places, no percent sign)?

Problem #6: Linear Regression with dummy variables

Use regression for the following questions.

First, create the following dummy variables:

- In column I, create a variable called Brick.
  - 1 if BrickConstruction = Yes
  - 0 if BrickConstruction = No
- In column J, create a variable called East
  - 1 if Neighborhood = East
  - 0 if Neighborhood is not East
- In column K, create a variable called North
  - 1 if Neighborhood = North
  - 0 if Neighborhood is not North

Create the regression where Price is the dependent variable (Y) and SqFt, Bedrooms, Bathrooms, Offers, Brick, East, and North are the independent variables (X's). Use multiple regression to solve this problem.

Problem #6 questions:

41. Is the equation statistically significant (Yes or No)?
42. How much of the change of the price of a house, in percent, is explained by the change in the variables SqFt, Bedrooms, Bathrooms, Offers, and the dummy variables for BrickConstruction and Neighborhood (enter as percent with two digits, no decimal places, no percent sign)?
43. What is the intercept (enter number with no decimals)?
44. Is the intercept statistically significant (Yes or No)?
45. Which region (East, North, or West) has the highest price?
46. Which region (East, North, or West) has the lowest price?
47. Which type of construction (Brick or Not-Brick) has the highest price?
48. What is the price difference between Brick and Not-Brick (enter a positive number with no decimals)?
49. What is the intercept for a Brick house in the West (enter number with no decimals)?
50. What is the intercept for a Non-Brick house in the East (enter number with no decimals)?