Table of Contents

# Step 1: Create a correlation table.

1. **Variables**: The variables listed in the table include “Indoor pool,” “Competitor rooms,” “Distance to competitor,” “Office space,” “Median income,” “Distance to downtown,” and “Operating Margin.”
2. **Correlation Coefficients**: The numbers in the table represent the correlation coefficients between pairs of variables. Each cell in the table indicates the strength and direction of the linear relationship between the corresponding variables. The correlation coefficient ranges from -1 to 1:
   * A correlation coefficient close to 1 indicates a strong positive linear relationship, meaning that as one variable increases, the other variable also tends to increase.
   * A correlation coefficient close to -1 indicates a strong negative linear relationship, meaning that as one variable increases, the other variable tends to decrease.
   * A correlation coefficient close to 0 indicates little to no linear relationship between the variables.
3. **Interpretation**:
   * For example, the correlation coefficient between “Indoor pool” and “Competitor rooms” is approximately -0.027. This suggests a very weak negative linear relationship between having an indoor pool and the number of competitor rooms nearby.
   * The correlation coefficient between “Office space” and “Operating Margin” is approximately 0.618. This indicates a strong positive linear relationship between the amount of office space available in the hotel and its operating margin.
4. **Analysis**:
   * Correlation coefficients close to 1 or -1 indicate variables that may be strongly related to each other, suggesting potential areas of influence or dependency.
   * Correlation coefficients close to 0 suggest little to no linear relationship between the variables.
   * It’s important to note that correlation does not imply causation. While two variables may be correlated, it does not necessarily mean that changes in one variable cause changes in the other. Other factors or variables may influence the relationship.

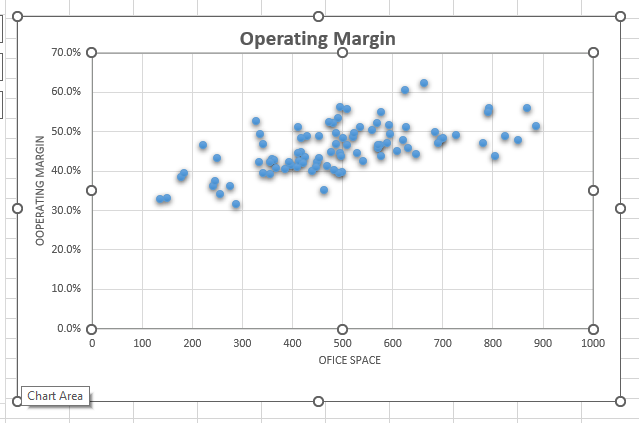
## • Highlight the correlation values for operating margin related to each independent variable.

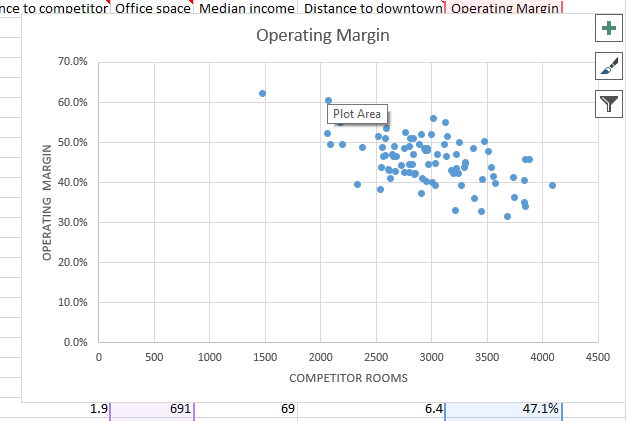
1. **Indoor Pool**: Correlation coefficient = 0.369
   * This positive correlation suggests that there is a moderate positive linear relationship between the presence of an indoor pool in the hotel and its operating margin. Hotels with indoor pools tend to have higher operating margins.
2. **Competitor Rooms**: Correlation coefficient = -0.572
   * This negative correlation indicates a strong negative linear relationship between the number of competitor rooms nearby and the operating margin. Hotels located in areas with a high number of competitor rooms may experience lower operating margins.
3. **Distance to Competitor**: Correlation coefficient = -0.275
   * This negative correlation suggests a moderate negative linear relationship between the distance to competitors and the operating margin. Hotels located closer to competitors may face more competitive pressure, resulting in lower operating margins.
4. **Office Space**: Correlation coefficient = 0.618
   * This strong positive correlation indicates that there is a significant positive linear relationship between the amount of office space available in the hotel and its operating margin. Hotels with more office space tend to have higher operating margins.
5. **Median Income**: Correlation coefficient = -0.096
   * This correlation value suggests a weak negative linear relationship between the median income of the area and the operating margin of the hotel. Higher median incomes in the surrounding area may not necessarily translate to higher operating margins for the hotel.
6. **Distance to Downtown**: Correlation coefficient = 0.149
   * This positive correlation indicates a weak positive linear relationship between the distance to downtown and the operating margin. Hotels located farther from downtown areas may have slightly higher operating margins, although the relationship is not very strong.

# • Choose the two independent variables that have the highest correlation values (negative or positive) with Operating Margin.

1. **Office Space**: Correlation coefficient = 0.618 (positive correlation)
2. **Competitor Rooms**: Correlation coefficient = -0.572 (negative correlation)

# • Create a scatterplot with operating margin on the vertical axis and one of the independent variables you picked on the horizontal axis.





# Step 3: Create a linear regression for one independent variable.

1. **R Square (R²)**: The R square value represents the proportion of the variance in the dependent variable that is explained by the independent variables in the model. In this case, an R square of approximately 0.657 means that about 65.7% of the variability in the dependent variable is accounted for by the independent variables. A higher R square indicates a better fit of the model to the data, suggesting that the independent variables collectively provide a good explanation for the variation in the dependent variable.
2. **Coefficient**: The coefficients (or slopes) in the regression model indicate the change in the dependent variable for a one-unit change in the corresponding independent variable, holding all other variables constant. For example, in the provided output, the coefficient for “Competitor rooms” is approximately -6.68345E-05. This means that for each additional unit increase in the number of competitor rooms, the dependent variable decreases by approximately 6.68345E-05 units, assuming all other variables remain constant. Similarly, the coefficient for “Office space” is approximately 0.000209167, indicating that for each additional unit increase in office space, the dependent variable increases by approximately 0.000209167 units, holding other variables constant.
3. **P-value**: The p-value associated with each coefficient indicates the statistical significance of that coefficient. In hypothesis testing, the null hypothesis typically states that the coefficient is equal to zero (i.e., the independent variable has no effect on the dependent variable). A low p-value (usually below a predetermined significance level, such as 0.05) suggests that the coefficient is statistically significant, meaning that it is unlikely to observe such an extreme value if the true coefficient were zero. In the provided output, the p-values associated with the coefficients for “Competitor rooms” and “Office space” are both very close to zero (1.00096E-12 and 2.43524E-14, respectively), indicating that these coefficients are statistically significant predictors of the dependent variable.