**GSBA 524 Homework 1**

(50 points total)

**LA Real Estate**. Real estate pricing is very complicated. It is misleading to compare properties only by assessing List Prices. Other factors need to be considered, as well, such as Location, Size, and Type of Home. The data set “LA Real Estate.csv” gives the List Price (in $), SQFT, and Location (Downtown LA, Santa Monica, and Beverly Hills) for 217 properties in Los Angeles that were listed for sale recently.

1. Report the following summary statistics for the variable List Price.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| MEAN | MEDIAN | STANDARD DEVIATION | RANGE | IQR |
|  |  |  |  |  |

1. Report the % of properties that were listed for more than $20,000,000. Hint: I created a new variable called “HIGH” that takes on the value of 1 only if a property was listed for more than $20,000,000. Use Radiant to obtain summary statistics about this variable to answer this.
2. What is the shape of the distribution of the variable List Price? Answer the question without creating a histogram. Select one:
   1. Right Skewed
   2. Left Skewed
   3. Symmetrical

Explain your answer in one short sentence.

1. Create a histogram to verify your answer in c). Use 15 bins and include a graph in your solution.
2. Suppose one of the properties is accidentally recorded as $7,000,000 when it should have been listed as $17,000,000. Which measure (mean vs median) is less influenced by the recording error? Explain why in one short sentence (no computation is needed).
3. Report the % of properties that are designated Single Family Residential (vs Condo/Coop).
4. Report the % of properties for each of the locations (Downtown LA, Santa Monica, and Beverly Hills).
5. To see how Location and Home Type are related to each other, you need to examine the joint distribution of Location and Home Type. Use Pivot Tables in Excel or Radiant to create a two-way table describing this joint distribution and report the % of the grand total.
6. Report the following summary statistics for the variable SQFT for each of the three locations: Mean, SD, Q1, and Q3. What are the IQR values for each location?  
     
   Graphical user interface, application, table

   Description automatically generated  
     
   BH IQR = 4350  
   DTLA IQR = 308.5  
   SM IQR = 1746.25
7. What do the values you found in the above question indicate about differences between the locations?  
     
   From the sampled data, BH has the most SQFT and most observations in that area also have that same trend of large SQFT from looking at the mean and median.
8. Compute the correlation between List Price and Sqft (for all locations and property types combined).   
     
   A picture containing table

   Description automatically generated
9. Create a Side-by-Side boxplot of List Price by Location (comparative boxplots). Include a graph in your solution and comment on any interesting patterns that you observe.  
   Chart, box and whisker chart

   Description automatically generated

Chart, box and whisker chart

Description automatically generated

1. Create a Side-by-Side boxplot of Sqft by Location (comparative boxplots). Include a graph in your solution and comment on any interesting patterns that you observe.

Chart, box and whisker chart

Description automatically generated

1. Create a scatter plot of List Price vs Sqft. Include a graph in your solution and comment on any interesting patterns that you observe.

Chart, scatter chart

Description automatically generated

1. Which do you think is more related to the List Price of a property: Sqft, Home Type, or Location? Justify your response by citing values from your analysis. Note: no definitive conclusion can be reached at this stage. We need to use multiple regression models to do a rigorous study. Just briefly comment based on the above plots and your intuitions (in 2-3 short sentences).  
     
   Based on the sampled and limited data, especially in dtla, my initial thoughts lean towards sqft being more of a driver of list price because of:

the high correlation coefficient between list price and sqft

the rejection of the null hypothesis (no correlation).

However, that is not to say that geography does not play a large factor in the real world. In fact, if there were an equivalent sqft listing in Huntington Park or Watts, it would change my initial thoughts lean towards geography. Overall, there are many other confounding factors at play impacting list price.