Shareable Link

Submitted on November 16, 2022

PROMPT

Question 1 (5 points)

Create a pivot table to left join the social media data to the sales data by using VLOOKUP function, in a new spreadsheet named leftJoin. Before joining the social media data, the pivot table must meet the following requirements:

- There should be one row for each day.
- The date range should be between January 02, 2018 to December 31, 2018.
  For each day, calculate the sum of quantity, discounts, and netTotal and the average of price, purchase number, and days since last purchase. Feel free to include other variables if you wish.
- Subtotal and grand total should be removed.
- Once you create that pivot table, use VLOOKUP function to join the social media data.

Note: you must submit the pivot table in an .xlsx file.

DOD1.xlsx

RUBRIC

The date range is between January 02, 2018 to December 31, 2018.

- 0 points
  No, the date range is not between January 02, 2018 to December 31, 2018.
- 1 point
  Yes, the date range is between January 02, 2018 to December 31, 2018.

All menu items are aggregated by day.

- 0 points
  No
- 1 point
  Yes

Subtotal and grand total are removed.

- 0 points
  No
- 1 point
  Yes

The VLOOKUP is used to left join the social media data into the sales data.

- 0 points
  No, the data is not left joined.
- 1 point
  Yes, the data is left joined, but not through VLOOKUP function.
- 2 points
  Yes, the social media data is left joined to sales data using VLOOKUP function.

PROMPT

Question 2 Part 1 (7 points)

1. Create descriptive statistics using the left joined data from Question 1 in a new worksheet named Descriptive Statistics. The descriptive statistics must meet the following requirements:

- The data is grouped by columns.
- Include numeric values for the following descriptive statistics: mean, standard error, median, mode, standard deviation, sample variance, kurtosis, skewness, range, minimum, maximum, sum, and count.

2. Tidy up the descriptive statistics to meet the following requirements:

- The column headers appear over their corresponding numeric values.
- Delete extra row labels and only keep the first column of row labels.
- Adjust the decimal point to 1 and use 1,000 separator.

3. Create a histogram for the netTotal column using the left joined data from Question 1. Copy and paste the histogram under descriptive statistic table.

Note: you must submit the pivot table in an .xlsx file.

DOD2.xlsx

RUBRIC

The data is grouped by columns and the following descriptive statistics are included:

- Mean
- Standard error
- Median
- Mode
- Standard deviation
- Sample variance
- Kurtosis
- Skewness
- Range
- Minimum
- Maximum
- Sum
- Count

0 points
No descriptive statistics are provided.

1 point
Yes, the descriptive statistics are created using the left joined data from Question 1. However, either the data is not grouped by columns, or not all of the descriptive statistics are provided.

2 points
Yes, the descriptive statistics are created using the left joined data from Question 1, while the data is grouped by columns, and the all of the descriptive statistics are provided.

The column headers appear over their corresponding numeric values.

0 points
No

1 point
Yes

The extra row labels are deleted.

0 points
No

1 point
Yes

The decimal point is adjusted to 1.

0 points
No

1 point
Yes

1,000 separator is used.

0 points
No

1 point
Yes

A histogram is created for the netTotal column using the left joined data from Question 1.

0 points
No

1 point
Yes

PROMPT

Question 2 Part 2 (3 points)

Write a short summary to compare the histogram with its corresponding descriptive statistics. In the short summary, you should:

- Compare mean and median to discuss how the mean and median shapes the histogram you create.

- Discuss the values of skewness and standard deviation of netTotal. What do they tell us?

Mean is 3250.9 greater than median 3086.5 This tells us that this data is going to be slightly skewed to the right. we've got a much larger tail on the right, indicating that there are some larger values on right side.Low standard deviation means data are clustered around the mean, and high standard deviation indicates data are more spread out.