Scenario

You are a business analyst for an organization that distributes pizza ingredients to various restaurants. These ingredients include dough, sauces, and toppings. Your organization wants to expand its customer base but does not know what kind of restaurant or U.S. region to target. You have been given a customer sales data set that is a representative sample of the sales at your customers’ restaurants for 2015 and 2016. This also contains household income data for the restaurant zip codes.

Often, data required to answer stakeholders’ questions are found scattered across multiple data sets or even software systems. In some cases, adding that data to your analysis is costly in terms of resources, but including it can make or break your analysis. In the earlier activities in this course, you analyzed the data provided to solve business problems. In this assignment, you will use the data sets to analyze more variables that can enhance the quality of recommendations and projections. You will also explain to the IT department, in a memo, why capturing these variables is important and how the additional cost incurred is justified.

You will be given the following data sets:

* [Clean Customer Sales Data Set](https://learn.snhu.edu/content/enforced/1510737-QSO-560-X3347-OL-TRAD-GR.24TW3/Course%20Documents/QSO%20560%20Clean%20Customer%20Sales%20Data%20Set.xlsx?ou=1510737)
* [Customer Sales Data Set With Region](https://learn.snhu.edu/content/enforced/1510737-QSO-560-X3347-OL-TRAD-GR.24TW3/Course%20Documents/QSO%20560%20Customer%20Sales%20Data%20Set%20With%20Region.xlsx?ou=1510737)

Prompt

Analyze the given data sets using MS Excel and create a memo to the IT department advocating for the inclusion of data that is not currently part of the data set but can be procured to ensure a better analysis.

First, create a pivot table using existing variables in the customer sales data set and analyze the data. Then, use the customer sales data set with region to create a pivot table and analyze the data. Finally, compare the results of both these operations and determine the resource spending need so that these variables are captured by the restaurants to ensure better quality projections. Include relevant screenshots in your memo.

Specifically, you must address the following rubric criteria:

* **Impact of Existing Variables:**Analyze the different types of orders in the customer sales data set.
  + Create a pivot table of the customer sales data set.
    - Determine the following for each restaurant category:
      * The average overall order total.
      * The average dollar amount of each order for each of the three order types:
        + Delivery
        + Dine-in
        + Carryout
  + Explain whether disaggregating by restaurant category impacts the analysis based on average order amount.
* **Impact of a New Variable:** Analyze the different types of orders in the customer sales data set with region.
  + Create a pivot table of the customer sales data set with region.
  + Using the region variable, create a pivot table and perform the following analyses:
    - Determine the average dollar amount of an order:
      * in each of the four U.S. regions
      * within each region, for each order type
      * within each region, for each restaurant category
  + Explain the differences you see between regions in the data.
    - How will these affect your recommendations for sales teams going forward?
* **Justify Your Rationale for Adding New Data:** Justify the inclusion of new data for a better analysis.
  + What can you say about the benefit of analyzing your data by four regions instead of by 50 states? Is this helpful?
    - Does it outweigh any cost in resources involved in creating the region variable, either within the IT system or within your analytics group?