The question from the provided list I aim to answer is the following: what were the total sales by vendor? In order to solve this problem, I will first need to identify all relevant information regarding the vendors and their total sales numbers. By obtaining this data, I can compile it into one software, such as Microsoft Access, and discover patterns and trends. As there are six spreadsheets in total, it can be difficult to analyze information fragmented in various locations. This is why assembling the data and normalizing its column headers are vital to ensuring a superior response to the question and promoting the reduction of errors when investigating it. The areas of data required to answer this would be numerical and transactions by vendor. From this, the collective accumulation of all sales by their respective vendors can be organized and analyzed in one location to determine their performance.

Each spreadsheet is required to respond to the business question, as they all exist as fragments of the overall ledger for transactions in 2015. Only utilizing portions of the overall data can skew one's analysis and ultimately provide an incorrect report. For example, if only one spreadsheet were analyzed and one particular vendor outperformed the other, the total sales would appear to be in their favor. However, by gathering each spreadsheet as a whole, It can provide a richer understanding of each vendor's performance to illustrate their overall performance when presented with all possible data regarding them.

The areas of data from these spreadsheets I will use are vendor name, bottles sold, and state bottle retail. These specific areas directly adhere and respond to the question of total sales by vendor. While additional information such as zip code or vendor number can provide enhanced business intelligence, they create bloat in management's evaluation of the data, taking their attention away from the most important points within the spreadsheets. Vendor name will directly answer by whom the total sales were attributed to and the bottles sold and state bottle

retail work cooperatively to calculate the total transaction number per purchase. While the column for sales exists within numerous spreadsheets, these numbers may be erroneous. This is why utilizing these two columns for calculation can minimize improperly computed numbers where they exist.

Within the accumulated 2015 sales data within Cedar County, Iowa, the information provided exists as entries for alcohol sales throughout the year. From the store's location of the sale to the quantity purchased, these spreadsheets aim to act as a ledger for all transactions of these beverages within 2015. All transactions and their individual identifiers are required to respond to the business question, as they provide additional information on the data regarding sales and the vendor. For example, store name and address can provide further insight into the vendor's transaction by understanding the location. After analyzing each data set, each data category can be compiled into three sections: product information, geographic location and date, and vendor identification. These three sections demonstrate how all the data within each spreadsheet are related, as each spreadsheet is a portion of the overall ledger of cedar county's alcohol sales in 2015. This advances the categorization of data within the spreadsheets by allowing for an improved comprehension of how they interrelate, and how patterns can ensue from this knowledge.

Regarding the question of what the total sales by vendor were, most of the information present does not exist to establish this answer. As stated previously, this information may provide further insight into the transaction record that occurred per vendor, but does not provide specific answers for the business question, and merely exists to produce bloat for management. There exists data within areas of each spreadsheet that are improperly aligned, such as spreadsheet 52358 not including a column for sales, or 52337 including an extra column for county number.

These spreadsheets will be matched up properly when they are all compiled into Access for analysis. Such areas not appropriately aligned with the question at hand are item information, store information, and location of the sale, date, and category number.

To ensure all of the data works collectively and reports for my needs, establishing each spreadsheet's layout is identical is paramount. As errors existed within two spreadsheets already mentioned, their columns will either be added or removed in a newly created spreadsheet, in order not to lose the original documents. By guaranteeing that each spreadsheet matches, transferring their data into Access will ensure minimal to no errors. This process is known as normalization, as everything from identification number to bottles sold must be aligned and in identical format to decrease the risk of error when compiling the data. This way, the produced report will correctly output the proper data to answer the business question.

Potential questions for management my evaluation of vendor sales could assist in answering can be statistical differences between vendor sales, such as the difference between the highest and lowest performing vendor, geographic information which can assist in understanding where the highest volume of sales occurred, and the average quantity of alcohol purchased at a time assisting in inventory allocation. This information can answer other questions for the final report, such as sales by category, brand, or month. However, the main question to which the report will be directed will be a comprehension of sales by vendor.

After compiling and cleaning the data through the queries and other functions used to manipulate the records, this information is prepared to answer the business question effectively. In order to effectively answer the question of what the total sales by vendor were, the table compiled of all Iowa liquor sales spreadsheets will be queried by vendor name and sales to produce an output of the sum by individual record. Each vendor will be grouped by name within

the query to eliminate redundancy and normalize the data. Once this is accomplished, a view of each vendor in correspondence to their overall sales across Excel sheets will be present for analysis and determining what patterns among the information exist, which vendor sold the most or the least, and the differences between each vendor's sales. With this, an explanation of the business question can be established and presented to management for further proceedings.

Within the report for management, the main focus circulated the following question: What were the total sales by vendor? Recognizing that this would represent each vendor within the six spreadsheets, I compiled them into one table in Access. This allowed me to condense the information presented among the data in one object instead of numerous ones. Existing in numerous data sheets increased the chance of error through misreading it, having to navigate multiple times to identify patterns and ensure each record and column was standardized. After compiling the data sets into Access, I then went through the columns and multiple records to ensure integrity and that nothing was lost when transferring information between software. This was done to ensure that the information within my final report was not erroneous and did not leave records such as sales or vendor data out for final evaluation.

Once the integrity of the information was verified, I added an additional column labeled "sales," which would be computed by multiplying the state bottle retail by the bottles sold to determine the total amount of the final sale per record. This was done to identify the final value of the sale per vendor in an independent column to not lose track or incorrectly calculate the sum of sales within the report. This would also ensure that when the query was produced, there was a set column to transfer and not the two columns of bottles sold and state bottle retail, which would have to be multiplied anyway to find the sale amount. I then produced a query that retrieved the vendor name, sales, and bottles sold. The sale and bottle amounts were summed to ensure that

only one instance of vendor name was present within the query. If this was not executed, then the query would have produced a result of each instance of vendor name, and the total amount of sales per vendor would be very challenging to identify, along with how many bottles were sold.

After the query was executed, I began to produce the final report. This report was produced from the information within the query since it held the sum of sales along with each vendor name individually. I did not include the number of bottles sold as this was not crucial for management to understand. To adhere strictly to data that would answer the total sales by vendor, I included vendor name, sum of sales, and the percentage of the grand total of all sales among all vendors. This third column was included in the report because it could provide further insight into total sales and the distribution of which vendor accumulated what percentage of funds compared to each vendor as a whole. This was done in anticipation of questions that management may ask, such as by what percentage of sales each vendor made

Within the report exists the four following categories of data: vendor name, sum of sales, percentage of grand total, and grand total at the bottom of the report. The vendor name identifies each specific vendor included within the six spreadsheets to categorize their respective sales numbers in descending order, illustrating the highest and lowest-performing ones. The sum of sales was included to visualize all instances of each vendor's sales. This allows management to see a grand total per vendor in one area instead of calculating each record of vendor transactions individually. The grand total percentage includes what portion of sales each vendor attributed and how far they were apart. For example, Diageo Americas held 20% of all sales totaled, with Mango Bottling Inc at only 0.02%. Lastly, at the bottom of the report lies the grand total of every sale made among the six spreadsheets. This identifies every sale per vendor in one number for management to witness, compare, and identify altogether accumulated revenue.