**MIS-300 Enterprise Information Systems/Database**

**Richard Huang**

**Professor: Dr. Nichelle**

**December 15, 2018**

**Final Project Milestone Three**

**Enterprise-level Report**

**I. Identify the Business Question**

As a junior data specialist, I would need to provide the Access database and Enterprise-level report including executive summary to my manager. I have chosen

“What were the total 2015 sales by month?” as the business question of my report. To solve this problem, I would need to collect, compile, clean and analyze the liquor sales data from cedar county Iowa. The data should have attributes including date, transaction number, store number, store name and sales amount. I would also need to group and summarize the sales amounts by month.

**II. Obtaining the Right Data**

**A. Review the provided spreadsheets and explain why they are all required to respond to the business question and how they are all related.**

All six spreadsheets are about the liquor sale records of cedar county Iowa, and each spreadsheet is separated by the zip code. To answer the business question “What were the total 2015 sales by month?” I would need to obtain all the 2015 liquor sales records from the six spreadsheets, and need to group them with the total sales amounts by month.

**B. Explain which areas of data from these spreadsheets you will use and why. For example, which columns would you use?**

The date, transaction number, store number, store name, total sales amount of each transaction from the spreadsheets are the right data we need to obtain from the spreadsheets. For example, the columns that we need are date column, invoice number column, store number column, store name column, and total sale (Dollar) column.

**III. Analysis**

**A. Explain what the data actually is and how the data correlates back to the question you intend to answer.**

These databases are about the liquors store’s sales records in cedar county Iowa, and each file represents the sales history separate by zip code. These databases include the following information: invoice, date, store number, store name, address, category number, category name, vendor number, vendor name, item number, item name, item description, quantity amount of each pack, cost, retail price, amount sold in bottle, and total sales price. I will be able to retrieve the information from these databases above to find out the total of 2015 sales by month.

**B. Analyze the data and identify any areas that are not aligned properly. Be sure that you adjust these later to ensure that they are matched up properly.**

After reviewing these databases, “Iowa\_liquor\_sales\_52358” has no sales amount. “Iowa\_liquor\_sales\_52216” has been calculated incorrectly by the calculation of state bottle retail multiplying the bottles sold. “Iowa\_liquor\_sales\_52747” has incorrect information, whose category code does not match category name. “Iowa\_liquor\_sales\_52772” date format is different from others spreadsheets.

**C. Describe how you will pull all of the data together in order to make sure it works collectively and will report correctly for your needs.**

Based on my question: “What were the total 2015 sales by month?” I would need to collect the date column, invoice column, store number column, store name column, and total sale (dollar) column. I would need to make sure all date formats are consistent and I would need to calculate the sales amount if there was no sales amount by the calculation of state bottle retail multiplying bottle sold.

**D. Based on your analysis, provide a list of potential questions you could answer for management.**

Who were the best area liquor sales of the 2015?

Who was our biggest vendor/supplier?

What was the top sale brand in our store?

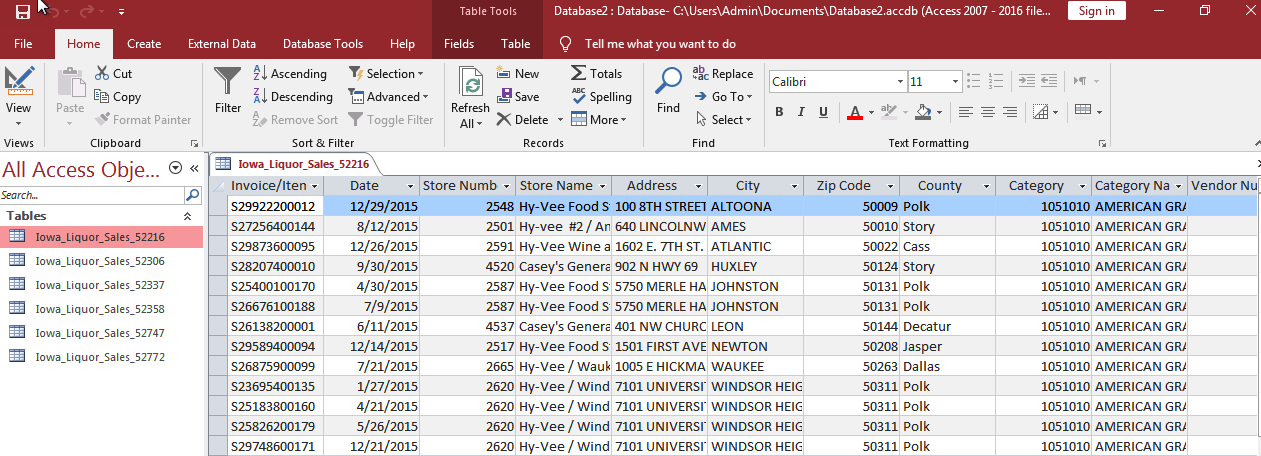
What was the top sale product in December 2015?

What was the best selling product by month?

**IV. Compiling and Cleaning the Data**

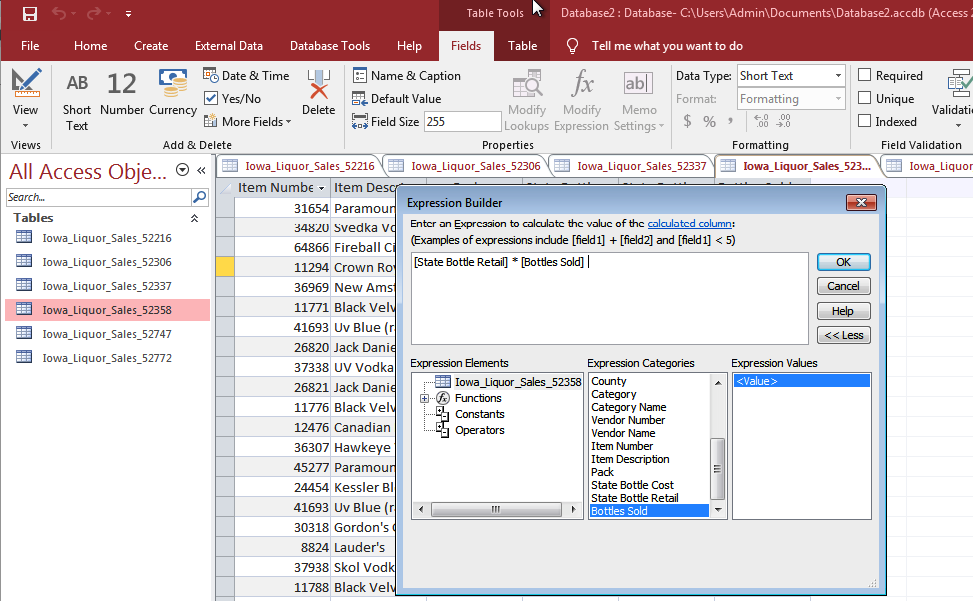
**A. Compile the spreadsheets you have analyzed and load them into Access in order to create a table.**

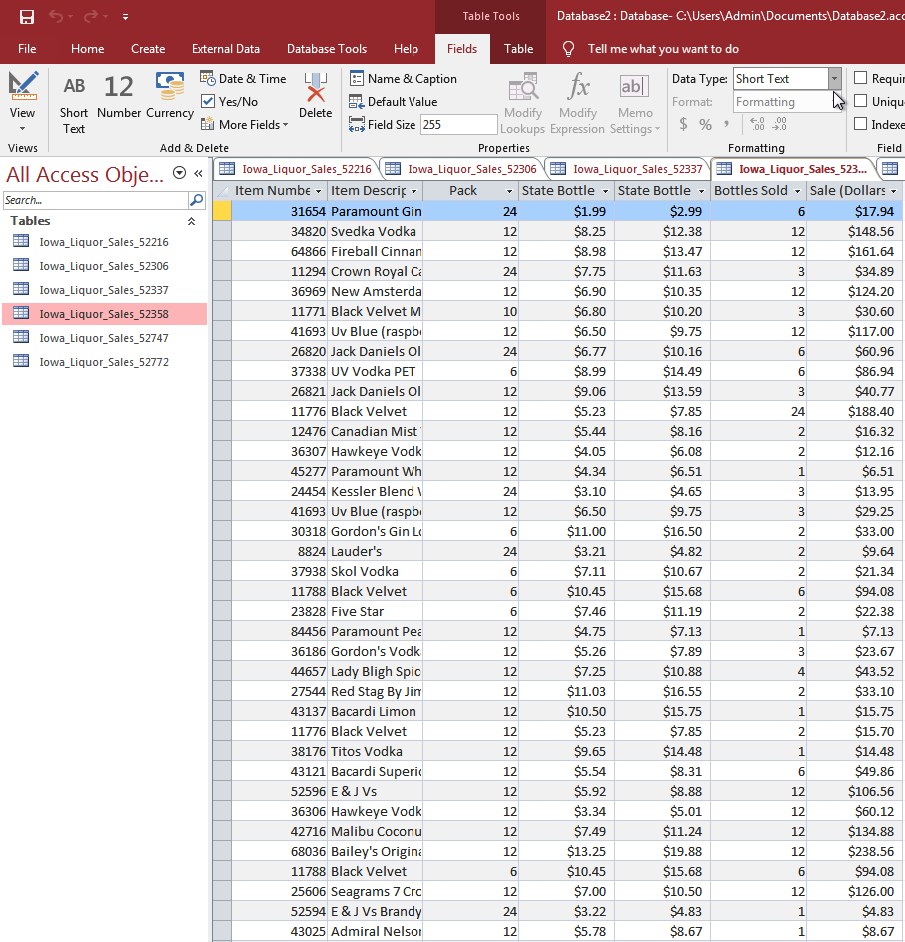
I have loaded all the excel spreadsheets into the access.

****

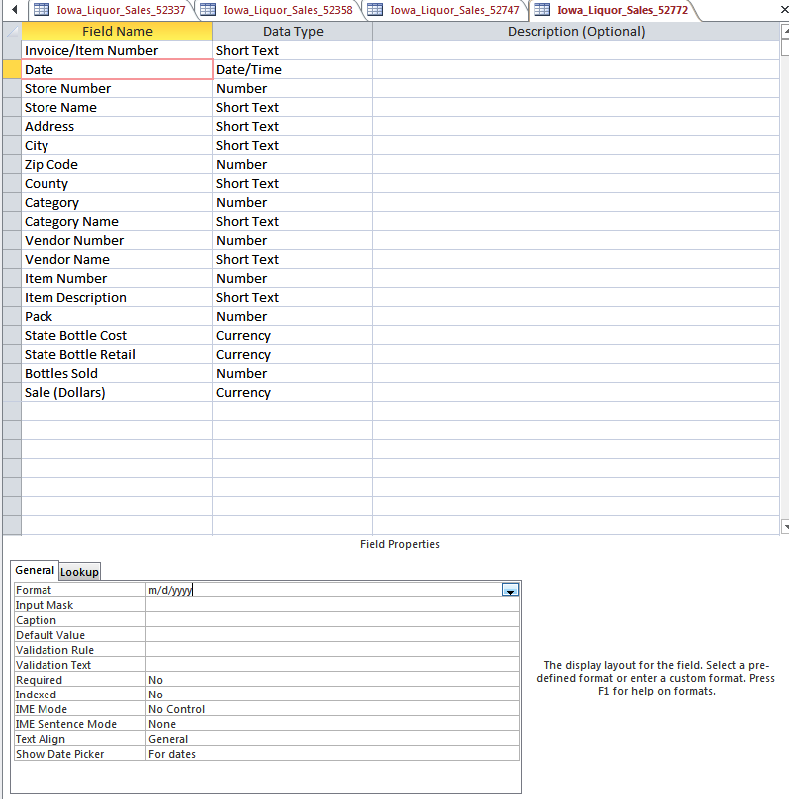
**B. Create a query to maintain the integrity of the original data and clean the data. For example, if you have different names within your table that actually mean the same thing, you will use the query to make the necessary changes.**

1. I would need to calculate the sales amount since there was no sales amount on “Iowa\_Liquor\_Sales\_52358” by inserting calculated field with [State Bottle Retail] \* [Bottle Sold]

****

****

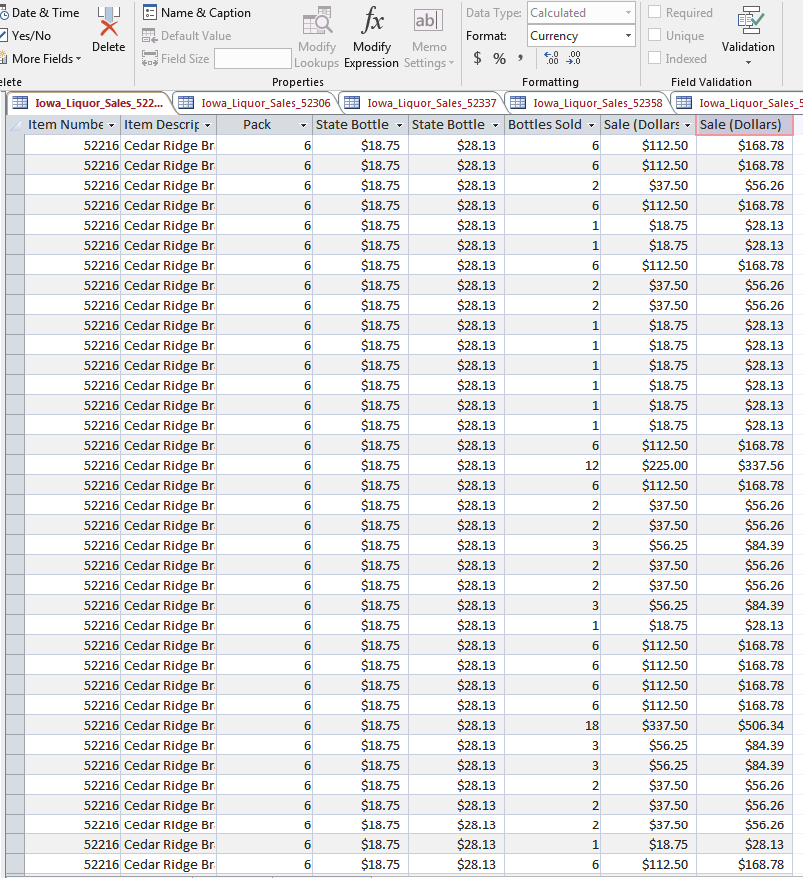
2. I would need to correct date format on “Iowa\_Liquor\_Sales\_52772” from “d-mmm-yyyy@” to “ m/d/yyyy” in design view section.

****

3. I would need to clean all other unnecessary fields, for example: “########” field and other extra fields that are not in the spreadsheets.

**C. While cleaning the data, be sure to look for and fix any incorrect, incomplete, and erroneous data.**

I have used the same method that I used on question B to insert the correct calculation of sales amount into “Iowa\_Liquor\_Sales\_52216” table. I would also need to delete the incorrect “Sale (Dollars)” field.

****

“Iowa\_liquor\_sales\_52747” has incorrect information, and the Blended Whiskies’ category number should be 1011100 not 1011200. We need to update the category number to 1011100 on all Blended Whiskies’ records in “Iowa\_liquor\_sales\_52747” spreadsheet.

**D. Based on your initial analysis and your compiled and cleaned data, explain how it will effectively answer the business question.**

After initial analysis and compiling and cleaning data, the data has been correct and consistent. It would greatly increase the compatibility of the data, and it would calculate the accurate data to answer the business questions.

**V. Report for Management**

Please refer to “Final Project Enterprise-Level Report” file.