# DAT 325 Project Two Template

# Executive Summary Report

## Data Set Anomalies

| **Key Value** | **Description of Anomaly** | **Plan for Resolution** |
| --- | --- | --- |
| 704-48-3927 | Missing city | Delete row |
| 778-71-5554 | Date is missing | Delete row |
| 252-56-2699 | Unit price is negative (-43.17) | Remove “-“ symbol. Cost of goods sold is $431.90. Unit price is $43.19. (43.19 \* 10 quantity = 431.90) |
| 347-34-2234 | Cost of goods sold is - $9999 when it should be $495.63 | Change the number to the right data Unit price \* Quantity (F64 \* G64 = $495.63) |

## Data Types

| **Header Name From File** | **Data Types Note** |
| --- | --- |
| Customer type | Convert into INT |
| Gender | Convert into INT |
| Time | Convert into INT |
| City | Change to only 1 character (A, B, or C based on the location) |
| Rating | Won’t use |
| gross margin percentage | Place 2 decimals to the left |
| All (except Gender and Quantity) | Manually change header to match Bruce’s data set |
| Payment | Replace “Credit card” with “CC” |
| TaxApplied | Create column  Using data dictionary, calculate the tax |
| TotalOrder | Create column  Using data dictionary, calculate the total amount |
| GrossIncome | Create column  Using data dictionary, calculate the gross income |

## Specific Transformations Needed to Join the Data

| **Header Name From File** | **Excel Function One** | **Excel Function Two** | **Excel Function Three** |
| --- | --- | --- | --- |
| City | IF(B2=”Yangon, Myanmar”, “A”, IF(B2=”Mandalay, Myanmar”, “B”, IF(B2=”Naypyitaw, Myanmar”, “C”))) then drag bottom right corner through all the rows in the column | Select column  Find city, state  Replace to just city (Example, “Find Yangon, Myanmar”, Replace it with “Yangon”)  IF(B2=”Yangon”, “A”, IF(B2=”Mandalay”, “B”, IF(B2=”Naypyitaw”, “C”))) then drag bottom right corner through all the rows in the column | Select column  Find city, state  Replace to just city (Example, “Find Yangon, Myanmar”, Replace it with “Yangon”)  =VLOOKUP(B2,'SM Data Dictionary from Bruce'!$G$3:$H$5, 2, FALSE) |
| gross margin percentage | =L2/100 then drag bottom right corner through all the rows in the column |  |  |
| Customer type | =IF(C2="Member", 1, IF(C2="Normal", 0)) then drag bottom right corner through all the rows in the column | Create a table with member/normal and the INT with it in a different sheet 'SM Data Dictionary from Bruce'  =VLOOKUP(I2, \*table from 'SM Data Dictionary from Bruce', 2, FALSE) |  |
| Gender | =IF(D2="Male", 0, IF(D2="Female", 1, IF(D2="N/A", 2))) then drag bottom right corner through all the rows in the column | Create a table with male, female, and N/A and the INT with it in a different sheet 'SM Data Dictionary from Bruce'  =VLOOKUP(I2, \*table from 'SM Data Dictionary from Bruce', 2, FALSE) |  |
| Time | Create a table with the times: 07:00, 11:01, and 15:01. Table2 is a table in a different sheet 'SM Data Dictionary from Bruce'  =VLOOKUP(I2, Table2, 2, TRUE) |  |  |
| TaxApplied | =G2\*(.05\*F2) then drag bottom right corner through all the rows in the column |  |  |
| TotalOrder | =(G2\*F2)+(TaxApplied in that row) then drag bottom right corner through all the rows in the column |  |  |
| GrossIncome | =L2\*(TotalOrder in that row) then drag bottom right corner through all the rows in the column |  |  |
| Payment | Home tab, “Editing” drop down menu, select “Find and Select” and then select “Replace”. Find “Credit card” and replace it with “CC” |  |  |
| cogs | =(Unit price) \* (Quantity) |  |  |

## Executive Summary

As we can see here, there are some anomalies in Wayne Enterprises’ Supermarket data set. I used Excel “find” function to find the blanks in the data set and came across one row that is missing a date. We will have to remove that row since we will not be able to find out exactly when that purchase had occurred. I looked through the data set and came across 3 more anomalies.

* One row was missing a city and only showed "Myanmar”. To solve this problem, we will need to remove this row.
* One row had a negative unit price. While the calculation was right, the negative symbol is not needed so that will need to be removed.
* One row had a negative cost of goods sold amount. To solve this problem, we will need to calculate the quantity \* unit price to get the right calculation. To ensure cogs column contains the right calculation for each row, I would use a formula and apply it to every cell in that column. If you click on a cell, it appeared that the number was put in manually.

Comparing Wayne Enterprises’ data set with Bruce, INC data set and data dictionary, Wayne Enterprises’ data set is missing 3 columns: TaxApplied, TotalOrder, and GrossIncome. Thankfully, the data dictionary had the calculation I needed to be able to create a column for each header in Wayne Enterprises’ data set so it can properly merge with Bruce, INC’s data set. We will need to fix the calculation for gross margin percentage by moving the decimal point 2 places to the left.

Majority of the header names do not match with Bruce, INC’s data set so each one would have to be manually change to match. Some columns used a different data type so that would need to be changed as well (Customer type, Gender, and Time needs INT variable). In Wayne Enterprises’ data set, City column contains city, state when Bruce, INC’s Location column contains A, B, or C. So, we will need to match the city with A, B, or C based on the location table in Bruce, INC’s data dictionary. Rating is the only column in Wayne Enterprises data set that will need to be removed since we will not be using it.

References:

Online Training for Everyone. (2020, Jan 2). EXCEL VLOOKUP: How to Lookup Data From Another Sheet [Video]. Youtube. <https://www.youtube.com/watch?v=h5abzTVqMiE>

Tugwell, Chester. (2021, July 27). If Time is Within Range | IF Time is Greater Than or Less Than, Return Value | Lookup Time in Range [Video]. Youtube. <https://www.youtube.com/watch?v=kyLLgxCCWWo&t=480s>