Yinhui Yang （Kami）

IBUAN 4210-01

March 14th, 2021

Python Project - Marvel Mart Project

Preliminary work

First of all I would like to input all of the csv, numpy, pandas, seaborn, matplotlib.pyplot into the beginning of the project start for individual coding. And input the pd.set\_option script to help me avoid the Python automatically doing the large number sums with any of floats in produces in the scientific notation while I am working on the project.

Part 1: Cleaning the Data

Input initial ‘df’ data frame and read MM\_Sales.csv file then follow the project requirements to find out any incorrect or missing data which is in the text type for that column in MM\_Sales.csv file and then change it to the string "NULL" or number 0 (or 0.0 if it's a float). Writing code to clean the Order Priority, Item Type, individually by using the isnull() check out NULL and and filling any missing values with "NULL". Accordingly, for the required Country cleaning data which will need to apply the (lambda x: x if isinstance(x, str) else "NULL") into the code. Lastly, since the Order ID carries the numeric in the data columns then I am writing the errors='coerce' to avoid the invalid parsing errors which will be found in the data and set as NaN at the same time. Next, I am going to copy the data and double check on the NULL and 0 which I have completely cleaned out then output a new file to the MM\_Sales\_Clean.csv file for the rest of the project.

Part 2: Exploratory Data Analysis with Reports & Visualizations

Input the new MM\_Sales\_Clean.csv file for coding work, counting the sales transitions and identify the top 10 countries, next step create the new countries list carry the 'Trinidad and Tobago', 'Guinea', 'Maldives' to found out the shipping center build list for country from the country in top 10 countries but not in the countries list and then the solution written into the Rankings file. Next I am going to use the new MM\_Sales\_Clean.csv file for the online and offline sales channel count then write the solution written into the Rankings file. Input chart value group by and count from new MM\_Sales\_Clean.csv file, in order to get the pic chart. The particular reason for the circumstances is that I am doing similar logical coding work for the next two questions Order Priority and other Sales Channel Pie charts. Import warnings and write the action='ignore' for the Part 2 Boxplot of Total Profits DISTRIBUTION avoid the error warnings. Then comes to the part 2 Q3 I am using the value group by and cum from a clean new MM\_Sales\_Clean.csv file, in order to get a bar chart displaying the result. Using the previous data result, write into a dictionary and get the top 3 Item Types from Total profit then write the solution into the Randing file. Finally, markdown the Boxplot of Total Profits DISTRIBUTION Discussion. Part 2 Q4 follows the project requirements determining the sum, average, maximum of Units Sold, Unit Cost, Total Revenue, Total Cost, Total Profit then write each solution into MM\_Calc.txt file individually. Lastly, using the calculation data display the Line Plot.

Part 3: Cross-Reference Statistics

Import warnings and write the action='ignore' for the Part 3 dictionary data to avoid the error warnings. After inputting all of the dictionary data with a header for the data of the region and then convert the dictionary data to the CSV file, create a CSV file called Countries\_By\_Region.csv .