Kyle Bailey

Marvel Mart Python Project

Part 1:

After my import statements, I read in the MM\_Sales.csv and saved a copy to begin cleaning. First, I tested for columns where there were no values so I could fill those later. I then began testing Countries for any numbers that were not characters but floats and indexed them to replace with a NULL value. I did a similar process with Order ID using indexing and then filled the empty rows in Item Type and Order Priority with NULL values as well. I then retested missing values with the .isna().sum() code and checked for unique values that looked out of place.

Part 2:

First, I appended the columns into new lists to form a dictionary for easier organization and created a new data frame. I chose to do this as it made it easier for me later on to save results to the .txt files and was more intuitive for me with sorting the values and getting counts. I then got the count of Sales Channels and Order Priorities, made a pie chart using Seaborn, and wrote them into MM\_Rankings.txt using the dictionary indices in a for loop.

I again used Seaborn to create a boxplot of the distributions and created a data frame of Item Type and Total Profit using a .groupby line with the sum of Profit. I created another Seaborn chart of a bar plot showing profit sums by item type and again used a similar .groupby code to rank the top 3 selling Item Types. To print these, I created two lists to combine into a dictionary that will be iterated through and appended to MM\_Rankings.txt.

For the last part of Part 2 I wanted to use a function so I defined a function called calc that will write a specified heading and display the later calculated stats. I created a data frame with the required values for calculations and then set three variables to calculate the sum, mean, and max from each of the 5 columns in the data frame. I plotted these values using log scale line plots for both to show a cleaner graph. Finally, I printed with a function call at the end.

Part 3:

For the cross reference statistics, I started by reading in the cleaned data and creating a data frame with Region and Country as the values. I then filtered the data frame for unique regions. Next, I created an empty list to iterate through with a for loop to locate the unique regions and countries that corresponded. Finally, I created a data frame that used region as the key and then concatenated the corresponding countries by column. Last, I saved the data frame to a .csv.