# Documentation for Marvel Mart Project:

Part 1: Cleaning the data  
I started this section by using the lectures from the class as the basis. The data cleaning methods I learned from the lectures can be easily applied to this assignment, requiring only a few changes, so it was quite straightforward. I bumped into some trouble when removing the NULL or 0 values, because there is a space in some of the names. I fixed that by changing it into an underscore. At first, I tried to delete the empty space, but it caused some trouble in Part 2 so I used underscore instead.

## Part 2: Exploratory Data Analysis with Reports & Visualizations

I also used the lectures from the class as a reference for this section. When the code doesn’t work, I usually go back the previous lectures to see what was missing/wrong. If it’s not on the previous lectures, I make a couple educated or random guesses on what the code should be and if it still doesn’t work, I just look it up on the internet. Things were going smoothly until I made my first graph. The numbers on the graph had scientific notations (1e6, etc.), but I wanted ordinary numbers. So, I looked online and picked the first solution, which was using ticklabel\_format. The code didn’t immediately work when I pasted it over, so I had to make some adjustments. Another problem I encountered was that the pie charts only showed percentages, but I wanted both the percentage and the actual value. I searched online and the solution was to create my own formatting function, so I used it and it worked wonderfully. While creating the lineplot for Question 4, I changed the underscores from Part 1 back into empty space, so that the graph labels and the text file have the correct name.

## Part 3: Cross-Reference Statistics

I solved this problem by grouping the clean dataframe by the unique Regions and Countries, applying the list function, and then turning it into a dictionary. I followed the tips from the hint page, turning the dictionary into a dataframe. I first tried turning it into a csv to see what would happen, and the csv file didn’t look good. There were two problems: the results are aligned horizontally, unlike the picture from the project page. I solved that by googling and finding out the transpose function which reflect the dataframe by writing rows as columns and vice-versa. The second problem was the index numbers (0,1,2,…), which I solved by removing them with index=false.

## Cleaning up

After I wrote all the code, I cleaned them up. There were some redundant sections with repetitive code that can be trimmed by creating functions, so that’s what I did. I also tried to remove some part of the code and re-run it (after restarting the kernel) to see if I could get a similar result with less. I added comments where necessary and went over everything again once I was done.