Zhongheng Sun

I selected the dataset from Metropolitan Museum of Art because, to put it simply, I enjoy appreciating Art. I wanted to pay a sort of homage to the museum even though I was swiftly rejected when I applied for its competitive summer internships. Working with the dataset is challenging because of its sheer size. There are over 500,000 items on display at the Met, and parsing through them is taxing on both my eyes and my machine. This is a credible data set because it came from the organization’s github account. This dataset can be used to filter for the type of artifact that the user is looking for. If the user types in “Painting” at the prompt, the program will search the entire database for paintings and return a list of Paintings, sorted by their date of origin. If the user enters “all” at the prompt, the program will return the entire Met collection, albeit with some data scrubbing. Some headings were taken out based on my discretion. For instance, I took out object IDs from the dataset because I judged that those are not necessary in my resulting dataset. Some important headers that I kept were the object types, location in the Met, and names of the artists. Also, I added a count for the number of objects returned at the end of the program, as a data point. The data is also sorted according to the date the artwork were created on.

Admittedly, there are flaws in the dataset that I produced. There are gaps in the data that can be better processed, such as the method shown during class. Personally, I was not particularly concerned with having white space in the data set, though I can see why one would take out the white space and fill it with a holder value such as “NIL”. Also, the sorting could be improved with a custom sorting algorithm and more precise data scrubbing. For now, the in-built sorting will place unknown values at the top of the list, followed by the later pieces, and then the earlier pieces. It might be more presentable to either remove pieces without a known date, or list them at the end of the list. One of tests I ran was the median value on a range of the object date of paintings which you can see in the Histogram document. The return value is 1960 to 1990, which suggest that a statistically significant amount of paintings done were in the period of 1960s to 1990s. This result is surprising to me as many would associate Eurocentric paintings that are displayed at the Met to the period of Romanticism. Many people would also agree that Romanticism ended in the 18th century. Something can be said about how impressions are not a reliable indication of the truth. Therefore, it would seem that either there are more art pieces produced after this particular age than I thought, or there might be an error in my data scrubbing and calculation. I conclude that more work needs to be done to process these datasets.

Source: <https://media.githubusercontent.com/media/metmuseum/openaccess/master/MetObjects.csv>