Question 1: Given some sample data, write a program to answer the following: click here to access the required data set

> On Shopify, we have exactly 100 sneaker shops, and each of these shops sells only one model of shoe. We want to do some analysis of the average order value (AOV). When we look at orders data over a 30 day window, we naively calculate an AOV of $3145.13. Given that we know these shops are selling sneakers, a relatively affordable item, something seems wrong with our analysis.

a. Think about what could be going wrong with our calculation. Think about a better way to evaluate this data.

After doing analysis, we found that the $3145.13 is the mean of order_amount. And there are several outliers that are highly affect the calculation. We found the abnormal order comes from the same shop (Shop_id 42). It seems like fraudulent as it all happens at 4:00:00 in all different day during March. It doesn't appear to be human behavior. So, we filter the Shop_42 out. Also, the Shop_78 seems like a store that sells luxury brand shoes with a price at $25725.0 per shoes, which is far above all other stores. As it can not represent the overall performance of the sneaker shops on Shopify, we also filter it out.
After Exclude these two stores, the **Corrected Average Order Amount is roughly $300.157 with an average 1.996 items ordered per time.** This is reasonable as the price for each item is (300.157/1.996) = $150.379, Which matches our assumption that "these shops are selling sneakers, a relatively affordable item".

b. What metric would you report for this dataset?

I will consider the median of the order_amount and the trim_mean of the order_amount. This two metrics can more intuitively reflect the dataset.

c. What is its value?

The 2.5% Trimmed Mean of order_amount is 300.23, which is relative same with our Corrected Average Order Amount

The Median is 284, which is also another good metrics to evaluate the performance of sneak shop on Shopify.