**Winter 2022 Data Science Intern Challenge**

**Question 1:**

a. The unusually high AOV of $3,145.13 is the result of several outliers caused by the high amount of revenue generated by some customers.

A better way to evaluate the data is to remove all purchases made by these customers and re-estimate the AOV of the data.

Once the outliers are removed, the calculated AOV will be $302.6.

The median of this data set is $284 and is not affected by the outliers.

b. The best metric to use to protect our evaluation from the effects of outliers and errors is the **median value**. Moreover, it is close to the AOV calculated without the outliers.

c. The median is 284.

**Question 2:**

1. Answer is :54













