Fall 2021 Data Science Intern Challenge

Question 1:

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

Some of the users bought large quantities of sneakers, which skewed up the AOV. Those users or customers can be considered as business owners not regular customers. I am thinking about two approaches. First approach is calculating the mode (the most frequently occurring order value), which will help us to find the most frequent amount of money spent by customers. Second approach is filtering the customers or users that buy large number of sneakers such as user_id 607, acted as outlier and might be considered as business owner not regular customer. Then calculate the median, which is less affected by outliers.

b. What metric would you report for this dataset?

I will use two approaches. First approach, Calculate the mode (the most frequently occurring order value) of order_amount. Second approach is filtering the customers who buy large quantities of sneakers. Those customers could be considered as business owner (I will assume the business owner is a customer buying more than 50 amounts in one order) then calculate the median of order_amount.

c. What is its value? (The code is saved in different file)

First approach Mode = 153

Second approach median = 284

Question 2:

a. How many orders were shipped by Speedy Express in total?

Count = 54

b. What is the last name of the employee with the most orders?

Last name = Peacock

c. What product was ordered the most by customers in Germany?

Product name = Gorgonzola Telino