

Summer 2022 Data Science Intern Challenge by Valeria Bricio.

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

After analyzing the data and running tests on the information, I was able to find 2 shops that were not adding positive value to the rest of the data pool. By data that didn't match the trends, we were getting AOV and medians of value that were over the average normal purchase.

A better way is to include a separate independent analysis for shops and businesses that are disrupting the trends we are familiar with. This foresees how these specific shops behave and how we can serve them better on our platform.

- b. What metric would you report for this dataset?

Order_amount would be the metric I would report. In this scenario, both order_amount and total_items were showing the high values that clouded out data; however, if order_amount was broken down also into the value of the unit we could have gotten a good start on cleaning the information and getting the one we needed.

By reporting, I would not get rid of this valuable insight into trends, other than add more detail that could be beneficial to understanding the shops and the kind of sneakers they sell.

- c. What is its value?

The value that is looking to be added is to be able to look further and more detailed into some of the information that is not following the standards we are familiar with. All deviations have the potential of unlocking something that has not been explored yet.

By segmenting the shops and continuing to know how to process that specific set of information, we are looking for our next great idea.

Question 2: For this question you'll need to use SQL. [Follow this link](#) to access the data set required for the challenge. Please use queries to answer the following questions. Paste your queries along with your final numerical answers below.

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

SELECT * FROM [Orders] WHERE ShipperID = 1	Number of Records 54
--	----------------------

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

SELECT EmployeeID, COUNT(*) AS Sales FROM ORDERS GROUP BY EmployeeID	Peacock
---	---------

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

SELECT c.country, p.productid, p.productname, o.quantity FROM OrderDetails AS o Join products AS p JOIN Customers AS c ORDER BY c.country, p.productid	
---	--