

Solution of Fall 2022 Data Science Intern Challenge

Please complete the following questions, and provide your thought process/work. You can attach your work in a text file, link, etc. on the application page. Please ensure answers are easily visible for reviewers!

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.

1. Think about what could be going wrong with our calculation. Think about a better way to evaluate this data.
2. What metric would you report for this dataset?
3. What is its value?

Solution 1:

1. The problem is, shop 42 sold bulk amount of items 17 times while the products price was very high in shop 78 compared to others which makes a bias in the dataset and hide the actual scenario. Therefore, if we separate the AOV of the bulk products of shop 42 (\$704000) and AOV of shop 78 (\$49213.05), then we get the actual AOV (\$1760).
2. Number of customers in a month, Average amount spent per customer are the metrics for this dataset.
3. Number of customers in a month is 301. Average amount spent by user 607 is \$704000 for bulk purchase while average amount spent in shop 78 is \$50306.67 per customer. Hence, average amount spent per customer is \$4979.47 for others.

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.

1. How many orders were shipped by Speedy Express in total?
2. What is the last name of the employee with the most orders?
3. What product was ordered the most by customers in Germany?

Solution 2:

1. 54 orders were shipped by Speedy Express in total.
2. "Peacock" is the last name of the employee with the most orders.
3. Product "Gorgonzola Telino" (Product ID 31) was ordered the most by customers in Germany.