Fall 2022 Data Science Intern Challenge(Shopify)

**Question 1:** Given some sample data, write a program to answer the following: [click here to access the required data set](https://docs.google.com/spreadsheets/d/16i38oonuX1y1g7C_UAmiK9GkY7cS-64DfiDMNiR41LM/edit#gid=0)

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?
4. The AOV which is calculated is the total order amount for all shops due to which the price for shoe seems higher than the normal price. However, average order value should be calculated per shop. After calculating it like this the amount was $407 which seems still higher than the normal price. I checked for duplicate values but there were none. Further I analyses that there are outliers in the dataset as shop\_id 42 is selling order\_amount of 70400 which further turned out to be reasonable as the items were 2000.

Shop\_id 78 had the order amount of $15430 which seemed suspicious as the total items purchased were 6.

1. Rather than mean value, I would report median or mode as they do not make that much difference by extremely large values in dataset
2. The value is $153

**Question 2:** For this question you’ll need to use SQL. [Follow this link](https://www.w3schools.com/SQL/TRYSQL.ASP?FILENAME=TRYSQL_SELECT_ALL) 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?**

**Graphical user interface, text, application, email

Description automatically generated**

#### What is the last name of the employee with the most orders?

#### Ans : Peacock

#### Graphical user interface, application Description automatically generated

#### What product was ordered the most by customers in Germany?

#### Graphical user interface, text, application, email Description automatically generated