Winter 2022 Data Science Intern Challenge

Name: Samuel Nguyen  
Email: [sam.vinhnguyenn@gmail.com](mailto:sam.vinhnguyenn@gmail.com)

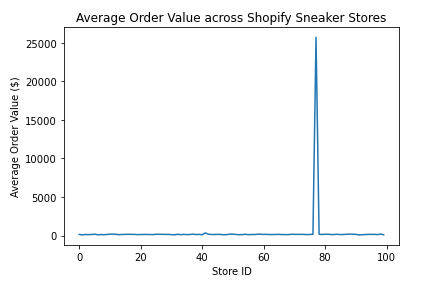
**Question 1:**

**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.**

**Question 1a:**

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

Assuming that the calculation of AOV = $3145.13 is numerically correct, there are a few things that could be going wrong with our calculations. While sneakers are generally affordable items, there are shoes such as designer shoes that could cost significantly more than your average running shoes from Nike or Adidas. This is observed in the figure below when looking at store #78, as the cost of ONE pair of sneakers is $25,725. Compared to the rest of the data, this is a huge outlier, which is skewing the average order value.



A better way to evaluate to this would be to categorize the stores based on price categories, or if possible, style of shoe (ie. Runners, baskeball shoes, lifestyle, athletic). However, categorizing by style of shoe would require additional data. These additional categories would help provide a more accurate analysis of average order value, since you are now comparing Nikes to Nikes, and not Nikes to Balenciaga’s. Alternatively, you could also simply remove the outliers from the dataset, and then calculate the average order value.

**Question 1b:**

**What metrics would you report for this dataset?**

Due to the data quality presented, it seems that the most insightful metric would still be average order value, as we are unable to calculate perhaps the cost of acquisition (COA) or return on investment (ROI) as an example. In order to calculate the average order value, you could simply group the data by Shop ID, and then divide the total order amount by the total items for each store.

**Question 1c:**

As mentioned above, the simplest method to calculate the true average order value would be to remove the outliers from the dataset. After removing store #78, I found that the AOV was $152.26, which is definitely more reasonable for a pair of sneakers.

**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.**

**Question 2a:**

**How many orders were shipped by Speedy Express in total?**

***Query:***

SELECT COUNT(\*)

FROM Orders

JOIN Shippers ON Shippers.ShipperID = Orders.ShipperID

WHERE ShipperName = "Speedy Express";

***Answer:***

A total of **54** orders were shipped by Speedy Express.

**Question 2b:**

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

***Query:***

SELECT OrderID,

Orders.EmployeeID,

LastName,

COUNT(\*) as total\_orders

FROM Orders

JOIN Employees ON Employees.EmployeeID = Orders.EmployeeID

GROUP BY Orders.EmployeeID

ORDER BY total\_orders DESC;

***Answer:***

The employee with the last name “Peacock” had the most orders (40).