Winter 2021 Data Science Intern Challenge

*Alexander Polus’s Submission*

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?

***This calculation was done incorrectly. As I viewed a Pandas description of the data, I saw the $1345.13 figure that was naively calculated. What’s actually important here is the average purchase price for the shoe. Therefore, for each order amount, I divided it by the total items for that order. This gave a price per shoe for that transaction. Then, I used this in combination with the number of shoes purchased in each transaction in order to come up with a weighted average shoe price: $357.92***

**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?

Query:

SELECT COUNT(OrderID) FROM Orders  
WHERE ShipperID=1;

Result: 54

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

Query:

SELECT LastName FROM Employees  
WHERE EmployeeID=4;

Result: Peacock

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

ProductID: 40 Total Quantity: 160

Query:

SELECT ProductName FROM Products  
WHERE ProductID=40

Result: Boston Crab Meat

\*\* Additional work / thought process can be found in “ShopifyChallengeWork.txt”