# Winter 2021 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](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.

Answer:

The data set has 7 features and 5,000 instances. An AOV of $3,145.13 seems unreasonable given that we are dealing with sneakers. To be able to suggest a better way to evaluate this data, I first had to perform data exploration. Please refer to the R file named Shopify Challenge Q1 to see how this was conducted. In summary, I found out that outliers were skewing the results. These outliers had a significant effect on the AOV. These outliers should be further investigated to verify their veracity. In this case, I decided to get rid of these outliers and then calculated a new AOV.

1. What metric would you report for this dataset?

Answer:

Once I eliminated the outliers that were skewing the results, the resulting dataset is more realistic of a 30 day window and, therefore, it is okay with utilizing AOV as our metric. This metric helps Shopify keep track of its revenue.

However, a few of other metrics that would be of interest would be:

-Conversion rate: This would be calculated by dividing the total number of conversions by the total number of visitors to the Shopify>Sneakers category. The larger this conversion rate is the better it is for Shopify.

-Customer Lifetime Value: This metric would tell us how much Shopify can expect a customer to spend on its products during their relationship with the brand. However, to calculate this, I would need to know what the average customer lifespan is.

Some other interesting metrics would be average profit margin and returning customer orders vs new customer orders.

1. What is its value?

Answer:

An AOV of $302.58 is calculated for this dataset. Please refer to the R file named Shopify Challenge Q1 to see how this value was calculated. This value makes sense if we take into consideration that the average number of items/ order is 2 sneakers. This AOV is telling us, on average, how much a customer spends on a single order. The goal will be that next month’s AOV will be larger than $302.58.

AOV can help Shopify identify trends and buying patterns as well insights on how to come up with more appropriate pricing strategies.

**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?

Answer:54 orders

SELECT o.ShipperID, s.ShipperName, COUNT( DISTINCT o.OrderID) AS ['Orders Shipped']

FROM Orders AS o, Shippers AS s

WHERE s.ShipperName='Speedy Express' AND o.ShipperID=s.ShipperID



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

Answer: Peacock

SELECT o.EmployeeID, e.LastName, COUNT(o.OrderID) as ['Number of Orders']

FROM Orders AS o, Employees AS e

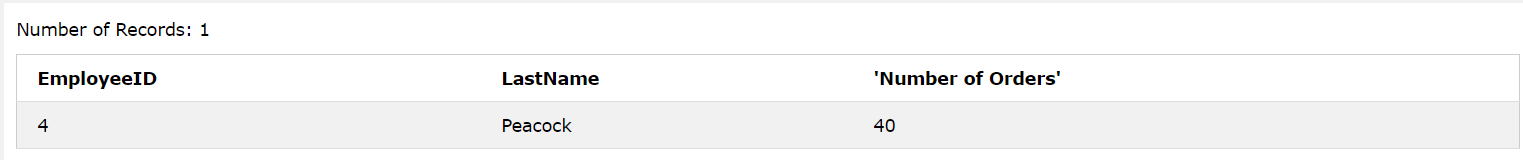
WHERE o.EmployeeID=e.EmployeeID

GROUP BY e.EmployeeID

ORDER BY ['Number of Orders']

DESC

LIMIT 1;



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

Answer: Boston Crab Meat

select c.ProductID, sum(c.Quantity) as [Quantity] ,d.ProductName

from Orders b join Customers a on b.CustomerID=a.CustomerID

join OrderDetails c on b.OrderID=c.OrderID

join Products d on c.ProductID=d.ProductID

where a.Country='Germany' group by ProductName order by Quantity

desc

limit 1

