

Summer 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](#)

Program can be found in file titled “data science intern challenge”.

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.

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

The AOV of \$3145.13 is the mean order amount. Since the data has many massive outliers, the AOV is skewed. To better evaluate the data, a measure of central tendency that does not weigh the outliers as much should be used.

- b. What metric would you report for this dataset?

I would report the median order amount. Since the median is simply the middle value in a set of data, the magnitudes of any outliers are not weighed as heavily.

- c. What is its value?

The median order amount is \$284.

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.

- a. How many orders were shipped by Speedy Express in total?

```
SELECT COUNT(*)
FROM shippers INNER JOIN orders
      ON shippers.shipperid = orders.shipperid
WHERE shippers.shippername = 'Speedy Express';
```

Final Answer: 54

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

```
SELECT TOP 1 lastname
FROM employees INNER JOIN orders
      ON employees.employeeid = orders.employeeid
GROUP BY lastname
ORDER BY COUNT(lastname) DESC;
```

Final Answer: Peacock

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

```
SELECT TOP 1 p.productname
FROM (((Customers as c
      INNER JOIN orders as o
            ON c.customerid = o.customerid)
      INNER JOIN orderdetails as od
            ON o.orderid = od.orderid)
      INNER JOIN products as p
            ON od.productid = p.productid)
WHERE Country = "Germany"
GROUP BY p.productname
ORDER BY SUM(od.quantity) DESC;
```

Final Answer: Boston Crab Meat