

# Fall 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](#)

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.

Answer: There are few transactions which have caused the data to be skewed such as shop\_id =42 and user\_id=607 are having multiple big transactions so average is around 3145.13, but as they are multiple transaction anomaly is ruled out. When can either remove such transaction or change the measure. Median would be a better measure as compared to mean as we would get a better picture of how the shops are performing.

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

Answer: Median order value

- c. What is its value?

Answer: 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(*) as Total_order_from_Speedy_Express
FROM orders,shippers
where orders.shipperid=shippers.shipperid
and shippers.ShipperName='Speedy Express'
```

Answer: 54

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

```
SELECT lastname
from orders as o,employees as e
where o.employeeid=e.employeeid
group by o.employeeid
order by count(*) DESC
limit 1
```

Answer: Peacock

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

```
SELECT Productname
from orders as o, customers as c, products as p,orderdetails as od
where o.customerid=c.customerid
and o.orderid=od.orderid and od.productid=p.productid
and country='Germany'
order by quantity DESC
limit 1
```

Answer: Steeleye Stout