Victoria Okoro
1/18/2022

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

> 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.
    a. Something that could be wrong with your calculation is that you are reporting the mean of the order amount. The mean of the order amount is 3145, which is due to a huge outlier value in the total items data set. The maximum number of items a customer has ordered was 2000, significantly larger than the other values in the total items data column. For example, a user with the id of 969 had an order amount of 432 and 2000 total items, which brings the cost to $864,000, another reason why the data is skewed.
b. What metric would you report for this dataset?
    a. The metric that I would report instead of in this data set is the median because it would give us the middle-value amount of what people would pay for their order.
c. What is its value?
    a. The value of the median is 284.

Victoria Okoro
1/18/2022

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?
    a. select count(*) from orders join shippers on orders.shipperid = shippers.shipperid where shippers.shipperid = 1;
    b. 54

b. What is the last name of the employee with the most orders?
    a. select lastname from employees e join orders o on e.employeeid = o.employeeid group by lastname order by count(*) desc limit 1;
    b. Peacock

c. What product was ordered the most by customers in Germany?
    c. Select productname from orders join customers on orders.customerid = customers.customerid join orderdetails on orderdetails.orderid = orders.orderid join products on orderdetails.productid = products.productid where country = 'Germany' group by productname order by quantity desc limit 1;
    d. **Steeleye Stout**