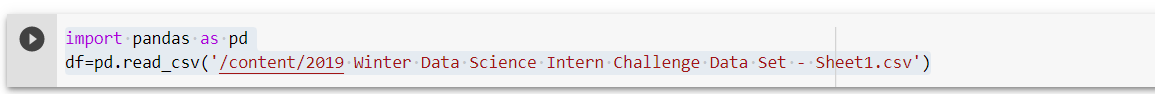
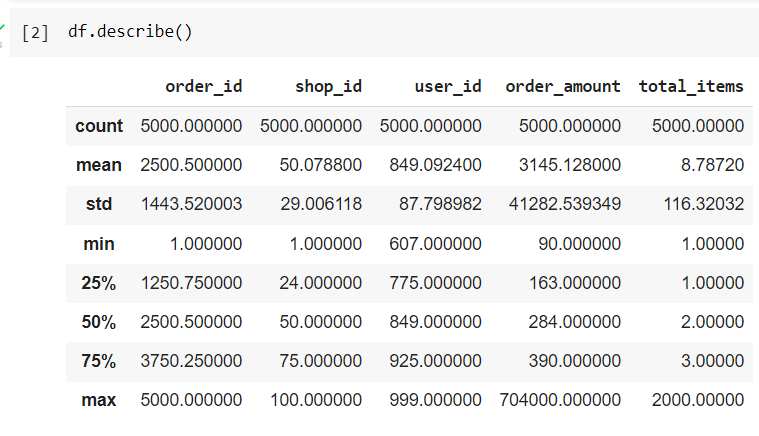
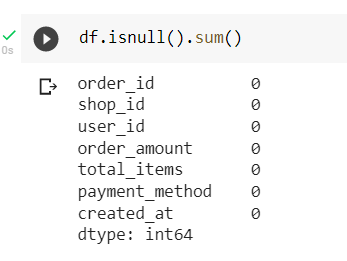
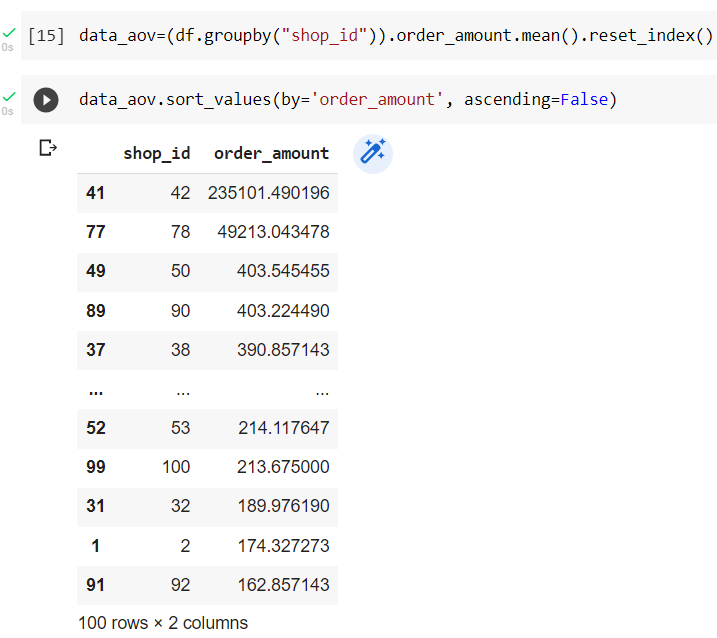
# Fall 2022 Data Science Intern Challenge

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](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.**   
   Reading in the data  
     
     
   As shown above, the mean of order\_amount is **3145.128.  
   **No null values in the data.  
   The mean that we calculated above shows the mean value of the total order amount for all the shops, but we are trying to find the average order value (AOV), which should be calculated as the average amount spent in each sneaker shop i.e, AOV by shop\_id.  
     
   From the above info, Shops with shop\_id 42 and 78 are the outliers. This could be due to that they sell expensive shoes only since each shop sells only one type of sneakers
2. **What metric would you report for this dataset?**  
   Average Order Value per shop.
3. **What is its value?**  
   Attached csv in the github [repo](https://github.com/27vikram/Shopify---Fall-2022-Data-Science-Intern-Challenge-Solution/blob/main/AOV_Per_Shop.csv).

**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?  
   SELECT Shippers.ShipperName, COUNT(Orders.OrderID) FROM Orders LEFT JOIN Shippers ON Orders.ShipperID = Shippers.ShipperID GROUP BY ShipperName HAVING ShipperName='Speedy Express';  
   **54**
2. What is the last name of the employee with the most orders?  
   SELECT Employees.LastName, COUNT(Orders.OrderID) FROM Orders LEFT JOIN Employees ON Orders.EmployeeID = Employees.EmployeeID GROUP BY LastName ORDER BY COUNT(Orders.OrderID) DESC LIMIT 1;  
   **Peacock 40**
3. What product was ordered the most by customers in Germany?  
   SELECT Products.ProductName, Customers.Country, COUNT(Orders.OrderID) FROM Customers JOIN Orders ON Customers.CustomerID = Orders.CustomerID JOIN OrderDetails ON Orders.OrderID = OrderDetails.OrderID JOIN Products ON OrderDetails.ProductID = Products.ProductID WHERE Country = 'Germany' GROUP BY ProductName ORDER BY COUNT(Orders.OrderID) Desc LIMIT 1;  
   **Gorgonzola Telino 5**