Summer 2022 Data Science Intern Challenge - Solutions

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: **Please Check the .ipynb notebook titled summerchallenge.ipynb to find thorough analysis on the data which is also more interactive. Found on the GitHub Challenge Link that is attached. You need the .ipynb to understand entire thought process.**

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

   - **The main reason why the calculation was going wrong is because of the outliers. Through analysis of the data we find out that roughly 2% of the data are outliers and those 2% contribute to 90% of the revenue. So we we cannot omit them or impute them with a some value because it would highly bias the data. The better thing to do instead of Average Order Value over a Rolling 30 day period is to use the Median as a Central Tendency Measure rather than the mean.**

b. What metric would you report for this dataset?

   Even though Conversion Rate and Revenue per User are metrics that are used in evaluating eCommerce data. I would choose to use a measure called the median over value over rolling 30 day periods.

c. What is its value?

   The value is found to be approximately 319.5

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?

SQL Query:

```
SELECT COUNT(*) as COUNT_OF_ORDERS from Orders where Orders.ShipperID = (SELECT ShipperID from Shippers where ShipperName = 'Speedy Express');
```

Total Number of Orders Shipped = 54

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

SQL Query:

```
SELECT TOP 1 Employees.LastName from Employees, Orders where Employees.EmployeeID = (SELECT TOP 1 EmployeeID from Orders GROUP BY EmployeeID ORDER BY count(EmployeeID) DESC);
```

LastName: Peacock

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

SQL Query:

```
Select TOP 1 Customers.CustomerName, sum(OrderDetails.Quantity) as SUM_OF_ORDERS from Customers, Orders, OrderDetails where Customers.CustomerID = Orders.CustomerID and Orders.OrderID = OrderDetails.OrderID and Customers.Country = 'Germany' GROUP BY Customers.CustomerName ORDER BY sum(OrderDetails.Quantity) DESC;
```

Customer Name: QUICK-Stop