**Shopify : Summer 2022 Data Science Intern Challenge**

**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.

*Answer: Before calculating and analyzing Average Order value, it is essential to check for outliers, duplicates and errors. When the data is filtered by order amounts greater than 24999, orders of very high values have been generated from shop IDs 78 and 42.*

*Shop\_ID 78 seems to be selling products worth 25725 a piece. This could mean that either they are not a sneaker store and there is an error in the data, or they could be selling special edition sneakers which are a lot more expensive than the regular ones. Therefore, they may be considered as an outlier or maybe grouped with other stores that sell special edition sneakers or other items of that value.*

*Shop\_Id 42 has several orders of the same quantity (2000 units) from the same user\_ID across several days. The orders are placed at the same time of the day too. There are multiple orders on the same day as well. These orders may need to be investigated to see if they are correct orders, or if these are duplicates or errors. We may also need to check if this customer regularly orders this quantity through the year or if it is a one-time thing which causes the AOV of March to be abnormally skewed as compared to the rest of the months.*

*Eliminating these outliers and duplicates may give a more accurate insight into the AOV.*

*The program for the same can be seen* [*here*](https://github.com/Njraman/InternshipChallenge)

1. What metric would you report for this dataset?

*Answer: There are many possible metrics that can be reported with this dataset that address different Key performance Indicators. Some examples are:*

*Average order value (transaction size) by shop\_ID : Comparison of order values across stores*

*Average order value by payment\_method: Comparison of order values across payment methods*

*Average order value by customer ID: Comparison of order values across customers.*

*Average units sold by shop\_ID : Identifying which shop has the most sales volume.*

*Number of unique user\_Ids by Shop\_ID- Determining where most users buy sneakers- the most popular shop.*

*Sales Volume by day of the week*

*Sales Volume by time of the day*

*These metrics could be calculated across several months to study how buyer behavior and store popularity change over a period.*

*However, the key metric that I would report is Sales Value by shop and the most popular shop ie. Average order value by shop ID and Number of unique user IDs by Shop ID. This would help determine how well a shop is performing.*

1. What is its value?

*Answer:*

*Average Order value by shop\_ID (Top 5 values):*

*Text

Description automatically generated*

*Number of unique user IDs by Shop ID:*

*A picture containing table

Description automatically generated*

*The program for the same can be seen* [*here*](https://github.com/Njraman/InternshipChallenge)

**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?

*Answer:*

*Query:*

*SELECT ShipperID from Shippers where ShipperName like"%Speedy Express%";*

*SELECT count(OrderID) FROM Orders where ShipperID = 1;*

*Result: 54*

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

*Answer:*

*Query:*

*Select e.LastName*

*from Employees as e*

*right JOIN Orders as o*

*on e.EmployeeID = o.EmployeeID*

*GROUP BY o.EmployeeID, e.LastName*

*ORDER BY count(o.OrderID) desc ;*

*Result: Last name of employee with most orders: Peacock*

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

*Answer:*

*Query:*

*SELECT d.OrderID, d.ProductID, d.Quantity, c.CustomerID, p.ProductName, c.Country*

*from (((OrderDetails as d*

*LEFT JOIN Products as p ON d.ProductID = p.ProductID)*

*LEFT JOIN Orders as o ON d.OrderID = o.OrderID)*

*LEFT JOIN Customers as c ON o.CustomerID = c.CustomerID )*

*where c.Country = 'Germany'*

*order by d.Quantity desc;*

*Result: Product ordered the most by customers in Germany is Steeleye Stout*