# Shopify - Winter 2022 Data Science Intern Challenge

Name: Juan Garcia

Date: Sep 19, 2021

I, Juan Garcia, certify that this material is my original work. No other person's work has been used without due acknowledgement.

**Question 1:**

**a.** An AOV of $3145.13 is not necessarily wrong, but needs additional details to get the full picture. Given that the shops sell sneakers, which are relatively low in price, the first thing that comes to my mind is to check on quantities, as the value of an order is often calculated as ***item price * quantity.*** If we got such a high AOV, then definitely someone bought a huge number of sneakers in a single order. A better way to evaluate this data is to add a calculation for items per order.

**b.** I would definitely report Items per Order, and would add the data range to help detect possible outliers.

**c.** From the dataset and using the field total_items:

Items per Order = 8.8
min value = 1
max value = 2000

From these results, we see at least one order has 2000 items. We would need a more detailed analysis (scatter plot, etc.) and historical data from past months to check if this is an expected value for certain shop/customer.

To answer this question, I used Qlikview® Personal Edition, as it is fast in loading data and producing KPIs and charts. See below screenshots for details:

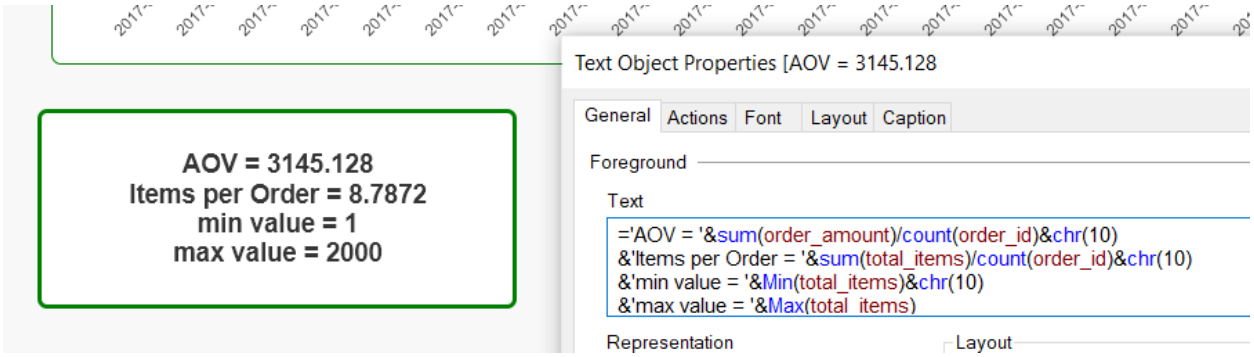
Edit Script [C:\Users\juanf\Desktop\Shopify.qvw*]

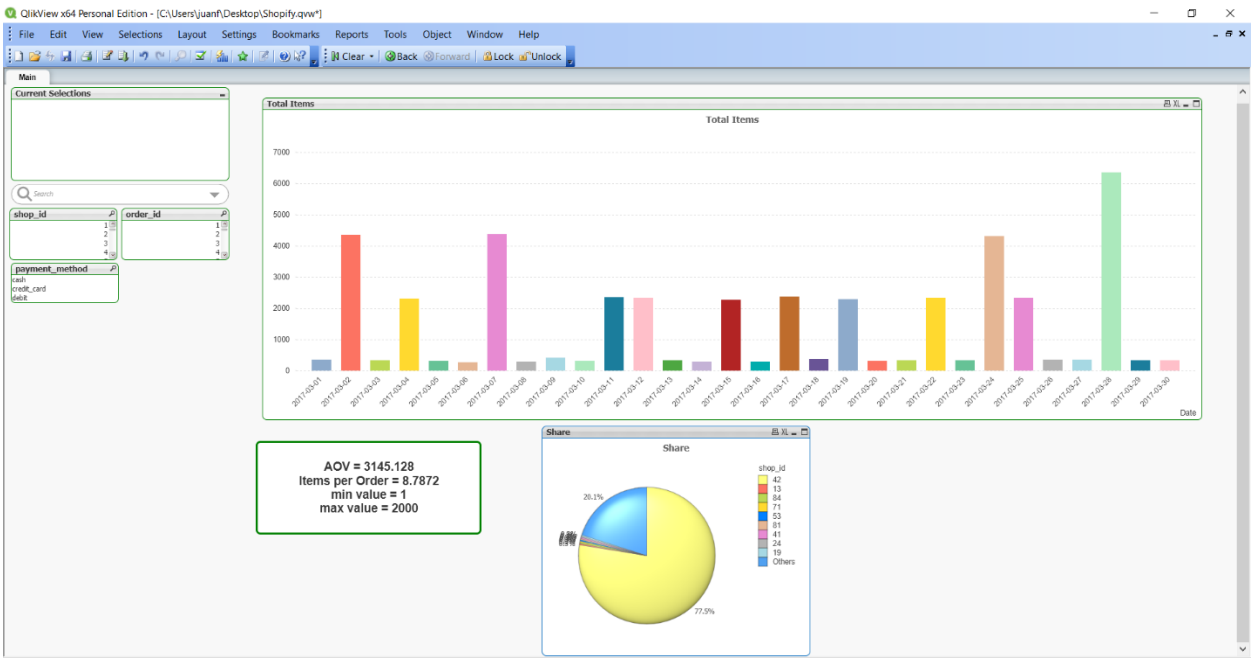
File Edit Insert Tab Tools Help

Reload Debug Tabs Main

Main

```
LOAD order_id,
     shop_id,
     user_id,
     order_amount,
     total_items,
     payment_method,
     created_at
FROM
[C:\Users\juanf\Desktop\Shopify.xlsx]
(ooxml, embedded labels);
```






## Question 2:

**a.** 54 orders in total were shipped by Speedy Express. See below image for query and result. I could have filtered the ShipperName to ‘Speedy Express’ but I think it limits the analysis, as some records could have a wrong or null ShipperID, so I prefer looking at the full picture first, to make sure all data is complete before giving an answer:

```
SELECT Orders.ShipperID, ShipperName, COUNT(OrderID) as
TotalOrders FROM [Orders] JOIN Shippers ON Shippers.ShipperID =
Orders.ShipperID GROUP BY Orders.ShipperID;
```

SQL Statement:

```
SELECT Orders.ShipperID, ShipperName, COUNT(OrderID) as TotalOrders FROM [Orders] JOIN Shippers ON Shippers.ShipperID = Orders.ShipperID GROUP BY Orders.ShipperID;
```

Edit the SQL Statement, and click "Run SQL" to see the result.

Run SQL »

Result:

Number of Records: 3

| ShipperID | ShipperName | TotalOrders |
|---|---|---|
| 1 | Speedy Express | 54 |
| 2 | United Package | 74 |
| 3 | Federal Shipping | 68 |

**b.** Last name of the employee with most orders is Peacock with 40 orders. See below image for query and result (included first name just because I like to know and call people by their first name):

```
SELECT LastName, FirstName, COUNT(OrderID) as TotalOrders FROM [Orders] JOIN Employees on Employees.EmployeeID = Orders.EmployeeID GROUP BY LastName ORDER BY TotalOrders DESC;
```

SQL Statement:

```
SELECT LastName, FirstName, COUNT(OrderID) as TotalOrders FROM [Orders] JOIN Employees on Employees.EmployeeID = Orders.EmployeeID GROUP BY LastName ORDER BY TotalOrders DESC;
```

Edit the SQL Statement, and click "Run SQL" to see the result.

Run SQL »

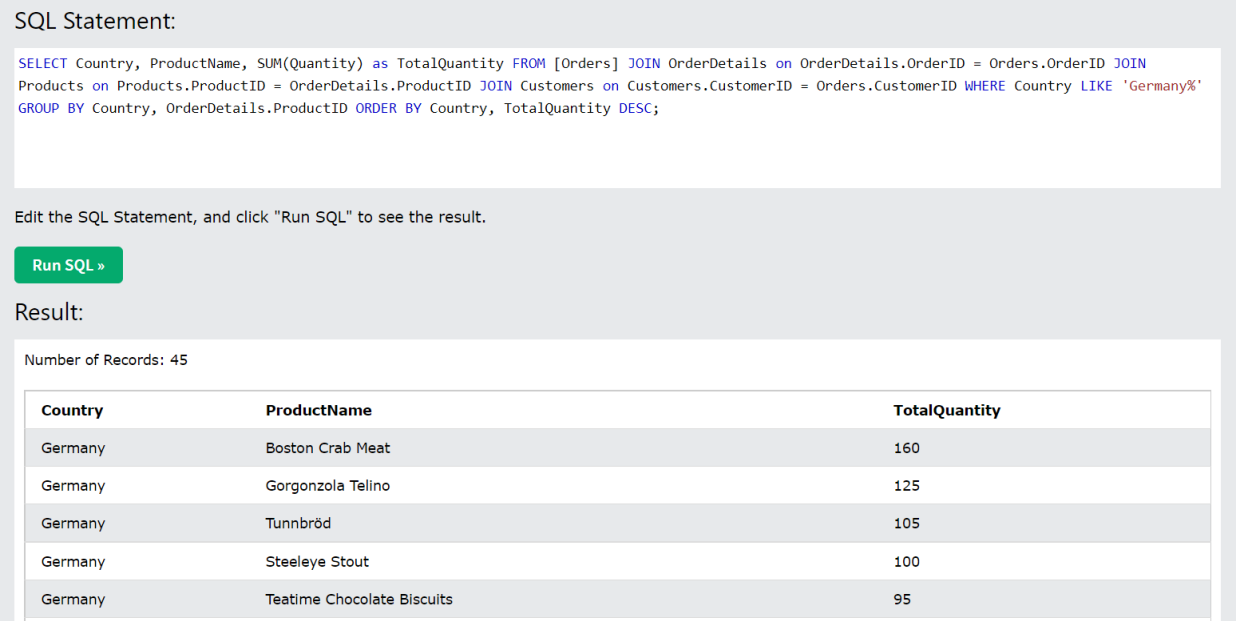
Result:

Number of Records: 9

| LastName | FirstName | TotalOrders |
|---|---|---|
| Peacock | Margaret | 40 |
| Leverling | Janet | 31 |
| Davolio | Nancy | 29 |
| Callahan | Laura | 27 |
| Fuller | Andrew | 20 |
| Suyama | Michael | 18 |
| King | Robert | 14 |
| Buchanan | Steven | 11 |

**c.** The product ordered the most by customers in Germany was Boston Crab Meat with a total 160 ordered. See below image with query and result (filtered the Country with a % at the end due to past experience with records that had spaces at the end due to some system glitches):

```
SELECT Country, ProductName, SUM(Quantity) as TotalQuantity FROM
[Orders] JOIN OrderDetails on OrderDetails.OrderID =
Orders.OrderID JOIN Products on Products.ProductID =
OrderDetails.ProductID JOIN Customers on Customers.CustomerID =
Orders.CustomerID WHERE Country LIKE 'Germany%' GROUP BY
Country, OrderDetails.ProductID ORDER BY Country, TotalQuantity
DESC;
```



SQL Statement:

```
SELECT Country, ProductName, SUM(Quantity) as TotalQuantity FROM [Orders] JOIN OrderDetails on OrderDetails.OrderID = Orders.OrderID JOIN
Products on Products.ProductID = OrderDetails.ProductID JOIN Customers on Customers.CustomerID = Orders.CustomerID WHERE Country LIKE 'Germany%'
GROUP BY Country, OrderDetails.ProductID ORDER BY Country, TotalQuantity DESC;
```

Edit the SQL Statement, and click "Run SQL" to see the result.

Run SQL »

Result:

Number of Records: 45

| Country | ProductName | TotalQuantity |
|---|---|---|
| Germany | Boston Crab Meat | 160 |
| Germany | Gorgonzola Telino | 125 |
| Germany | Tunnbröd | 105 |
| Germany | Steeleye Stout | 100 |
| Germany | Teatime Chocolate Biscuits | 95 |

Thanks for the challenge, it was fun to do. Looking forward to hearing from you soon.

Juan Garcia