Fall 2021 Data Science Intern Challenge

By:

Arshdip Hans, B.Eng.Sci.
MBA candidate class of 2022

Question 1

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

The data has a few notable issues that allow it to be skewed upwards from a value that would be beneficial when looking at the general representation of sneaker shop owners. First, the value is skewed upwards as shop id number 42 sold 2000 shoes per transaction at the price per show at $352.00 17 times in the last 30 days to user id 607 at the order amount of $704,000 each time. This extremely large value in comparison to other sales resulted in a skew upwards of the AOV. Also, it should be taken into consideration that store id 78 is selling each sneaker at a notable higher price then its competitors at $25,725 (refer to figure 1 for scatter plot) again skewing the data upwards.

If the metric of AOV in total were to be taken and the outliers stated above should be taken out of the data set the new AOV would be $302.58. This value seems a lot more reasonable than the original value of $3,145.13

When looking at better ways to calculate the data, we should take into consideration that the calculations do not address the needs for each store but lumps them all into the same broad category making generic assumptions. These assumptions are false as the sneakers have different competitive scopes and competitive advantages (could be thought as different sections on Porter's Generic Strategies). An example of this would be that store 78 as it is selling shoes at a high cost, would have an extremely high AOV but have lower sales. This is shown as it sold 88 sneakers compared to the average sneakers sold per store which is 439. As we do not know the competitive scope and advantages of the stores, we cannot subgroup them to allow a better representation of their competitive categories, but a more detailed metric should be introduced rather than a general total.

b) *What metric would you report for this dataset?*

When thinking of metrics, a few could be taken into consideration such as the previously mentioned average order value or revenue per visitor (RPV). Another powerful metric that we do not have the data to calculate is the conversion rate as it represents the percentage of visitors that make purchases an important metric in the ecommerce sector. When looking at both AOV and RPV, they should be taken as per store values rather than generic totals, which would allow a more tailored metric that is much more beneficial to the stores as it is a representation on how they are doing. While both average order value and revenue per visitor have their benefits and drawbacks, I would pick revenue per visitor as it is able to take into consideration repeat customer. The per shop value is necessary for shop as Shopify businesses in the sneaker industry may have different scopes and competitive advantages to each other. I believe this metric would be ideal to help shop owners successfully grow businesses as even with RPV data taken with the entire set that has the outliers and represent the same customer once even if they are shopping at different stores which would not benefit the shop owners as much as specific data about their store. In an ideal situation, we would know subcategories of the sneakers in which its RPV would also be calculated and store owners can compare their values to that giving them even more information. Ultimately, the metric I would report with this dataset is RPV per store.

c) *What is its value?*

Please refer to figure 2 in appendix (note: I also added AOV per store for fun)

Question 2

a) How many orders were shipped by Speedy Express in total?

Speedy Express shipped a total of **54** orders.

```
SELECT COUNT( * ) as "Number of Rows"
FROM [Orders]
Join Shippers
On Orders.ShipperID = Shippers.ShipperID
Where ShipperName= 'Speedy Express';
```

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

The last name of the employee with the most orders is **Peacock**.

```
Select LastName
FROM [Orders]
Join Employees
On Orders.EmployeeID = Employees.EmployeeID
GROUP BY LastName
Order BY count(Orders.OrderID) DESC
LIMIT 1;
```

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

The most popular product ordered by customers in Germany is **Gorgonzola Telino**.

```
SELECT ProductName
From Products
INNER JOIN OrderDetails
ON Products.ProductID= OrderDetails.ProductID
INNER JOIN Orders
ON OrderDetails.OrderID=Orders.OrderID
INNER Join Customers
On Orders.CustomerID=Customers.CustomerID
WHERE Country = 'Germany'
Group BY ProductName
Order BY count(ProductName) DESC
LIMIT 1;
```

A time I messed up and what I learned about it.

Recently, at my tenure as a product manager intern, I was constantly given information with regards to multiple products I was working on and had to analyze it to help with the decision making process. While usually I was independently able to make correct assumption to reach my objective, there was a time where wrong assumptions were made in how the data should be displayed for a stakeholder presentation making it difficult to reach my objective in time. To remedy the situation, I had a discussion with a senior product manager who helped clarify some confusion I had and I worked overtime to get the objective out in time. This not only improved my ability in the decision making progress but also cemented the team effort regarding the positive impact reaching out to senior members can be in help growth and success in reaching not only the business goals but make myself better at the job.

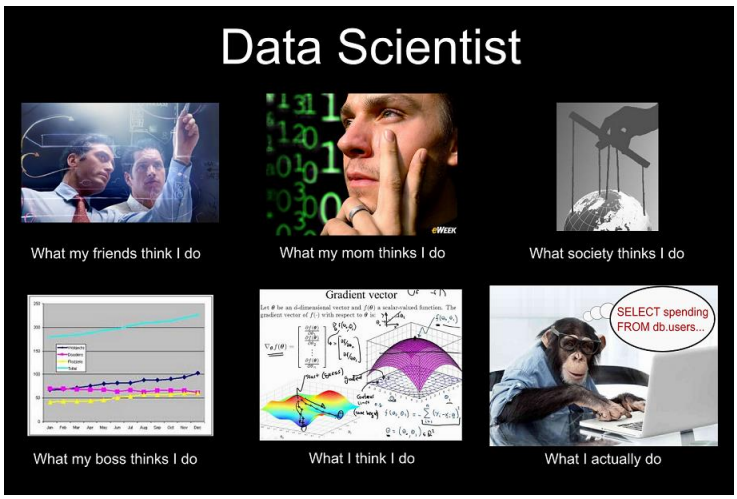
If you were to open a Shopify store, what would you sell and why

If I were to open a Shopify store, it would be a store focusing on selling martial arts equipment. The type of equipment I would sell would be typical self-defense training equipment such as pads, sparring equipment and training dummies. I would open this store because I trained in martial arts since I was 9 years old, in varying styles such as Shotokan, Jit-Jitsu, Judo and Wing Chun, later even becoming an instructor in Shotokan after I received my black belt at 14. Over a decade learning martial arts resulted in myself having a passion for not only the technique but teaching such as perseverance in actions I am doing as effort in the actions I am doing ultimately represents if I am going to be successful or not. This point has transcended martial arts for me and opening a Shopify store would allow me to promote other generations of people that are not only interested in martial arts the way I was but also people that want to learn it to stay healthy.

Any last thoughts? (i.e. proud accomplishments, cool ideas, a joke, etc.)

A proud accomplishment that I achieved recently was in the past year I was elected by my peer graduate students as a board member for the Graduate Student Association, a not for profit to help improve student life for us graduate students. I believe this not only represents the trust I am able to gather from my peers but also that wherever I am I try to promote a positive environment, something I am hoping to do at Shopify.

Also, I hope you find these Data science memes funny 😊



Appendix

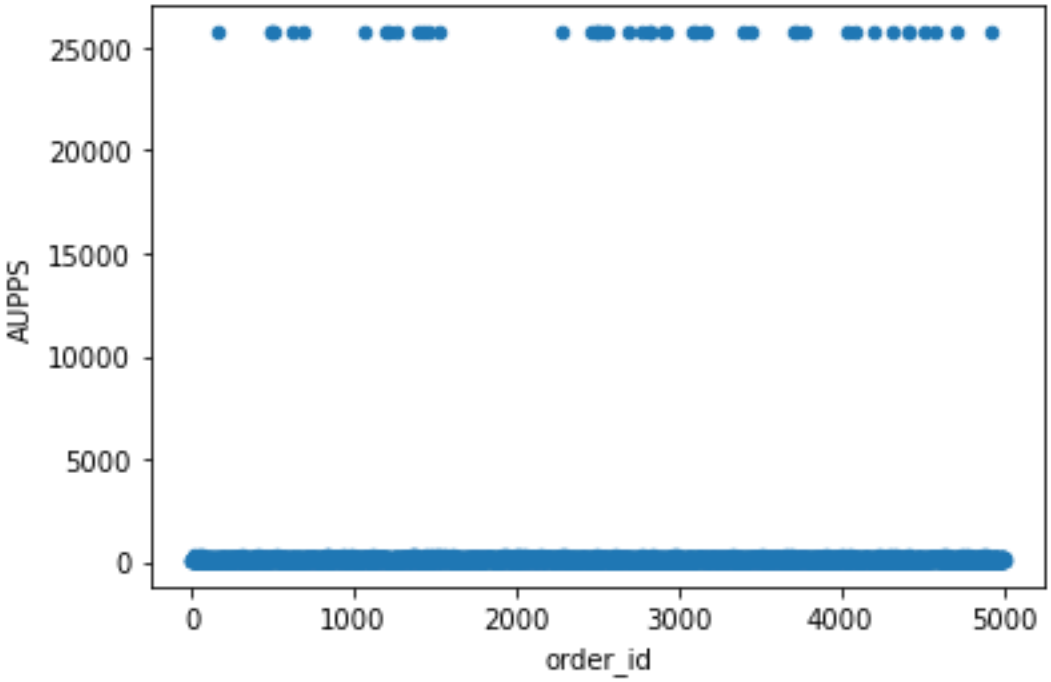


Figure 1

	Store_id	Amount_of_Transactions	Revenue_per_store $	Average_Order_Value_per_store $	Revenue_per_visitor $
0	1.0	44.0	13588.0	308.82	323.52
1	2.0	55.0	9588.0	174.33	188.0
2	3.0	48.0	14652.0	305.25	305.25
3	4.0	51.0	13184.0	258.51	292.98
4	5.0	45.0	13064.0	290.31	296.91
5	6.0	59.0	22627.0	383.51	426.92
6	7.0	56.0	12208.0	218.0	259.74
7	8.0	46.0	11088.0	241.04	264.0
8	9.0	59.0	13806.0	234.0	246.54
9	10.0	53.0	17612.0	332.3	366.92
10	11.0	49.0	17480.0	356.73	364.17
11	12.0	53.0	18693.0	352.7	397.72
12	13.0	63.0	21760.0	345.4	388.57
13	14.0	58.0	14036.0	242.0	259.93
14	15.0	52.0	16065.0	308.94	327.86
15	16.0	41.0	11076.0	270.15	270.15
16	17.0	53.0	17600.0	332.08	366.67
17	18.0	51.0	17472.0	342.59	371.74
18	19.0	64.0	20538.0	320.91	342.3
19	20.0	52.0	13081.0	251.56	278.32
20	21.0	46.0	14200.0	308.7	346.34
21	22.0	48.0	13140.0	273.75	298.64
22	23.0	55.0	17472.0	317.67	336.0
23	24.0	55.0	17640.0	320.73	339.23
24	25.0	48.0	11180.0	232.92	248.44
25	26.0	49.0	16720.0	341.22	355.74
26	27.0	54.0	18083.0	334.87	347.75
27	28.0	43.0	13776.0	320.37	320.37
28	29.0	58.0	19234.0	331.62	384.68
29	30.0	56.0	16524.0	295.07	344.25
30	31.0	47.0	12642.0	268.98	268.98
31	32.0	42.0	7979.0	189.98	204.59
32	33.0	40.0	15051.0	376.28	385.92
33	34.0	50.0	11712.0	234.24	260.27
34	35.0	52.0	17056.0	328.0	341.12
35	36.0	50.0	12740.0	254.8	265.42
36	37.0	48.0	16330.0	340.21	371.14
37	38.0	35.0	13680.0	390.86	402.35
38	39.0	41.0	10988.0	268.0	313.94
39	40.0	48.0	14168.0	295.17	308.0
40	41.0	59.0	14986.0	254.0	282.75
41	42.0	51.0	11990176.0	235101.49	386779.87
42	43.0	58.0	19367.0	333.91	372.44
43	44.0	39.0	10224.0	262.15	300.71
44	45.0	58.0	15620.0	269.31	312.4
45	46.0	43.0	14940.0	347.44	347.44
46	47.0	47.0	12180.0	259.15	276.82
47	48.0	40.0	9711.0	242.78	262.46
48	49.0	53.0	14835.0	279.91	302.76
49	50.0	44.0	17756.0	403.55	443.9
50	51.0	46.0	16643.0	361.8	387.05
51	52.0	41.0	12994.0	316.93	324.85
52	53.0	68.0	14560.0	214.12	242.67
53	54.0	50.0	13832.0	276.64	288.17
54	55.0	48.0	15732.0	327.75	349.6
55	56.0	37.0	8073.0	218.19	237.44
56	57.0	53.0	15729.0	296.77	314.58
57	58.0	59.0	15042.0	254.95	294.94
58	59.0	60.0	21538.0	358.97	398.85
59	60.0	47.0	16461.0	350.23	382.81
60	61.0	50.0	17222.0	344.44	391.41
61	62.0	43.0	13280.0	308.84	340.51
62	63.0	58.0	15368.0	264.97	307.36
63	64.0	43.0	11704.0	272.19	278.67
64	65.0	54.0	17864.0	330.81	364.57
65	66.0	53.0	16583.0	312.89	360.5
66	67.0	37.0	10087.0	272.62	288.2
67	68.0	47.0	11968.0	254.64	272.0
68	69.0	60.0	15851.0	264.18	288.2
69	70.0	59.0	20241.0	343.07	361.45
70	71.0	66.0	21320.0	323.03	343.87
71	72.0	46.0	14240.0	309.57	339.05
72	73.0	58.0	19470.0	335.69	389.4
73	74.0	38.0	11628.0	306.0	314.27
74	75.0	42.0	10112.0	240.76	252.8
75	76.0	42.0	13485.0	321.07	345.77
76	77.0	50.0	14040.0	280.8	286.53
77	78.0	46.0	2263800.0	49213.04	50306.67
78	79.0	54.0	17738.0	328.48	354.76
79	80.0	45.0	13485.0	299.67	328.9
80	81.0	59.0	22656.0	384.0	419.56
81	82.0	42.0	14691.0	349.79	367.28
82	83.0	42.0	10449.0	248.79	267.92
83	84.0	59.0	20196.0	342.31	381.06
84	85.0	35.0	11524.0	329.26	338.94
85	86.0	52.0	14430.0	277.5	313.7
86	87.0	52.0	15198.0	292.27	316.62
87	88.0	50.0	17776.0	355.52	395.02
88	89.0	61.0	23128.0	379.15	413.0
89	90.0	49.0	19758.0	403.22	439.07
90	91.0	54.0	17600.0	325.93	359.18
91	92.0	42.0	6840.0	162.86	171.0
92	93.0	59.0	12654.0	214.47	238.75
93	94.0	45.0	13400.0	297.78	311.63
94	95.0	39.0	12432.0	318.77	355.2
95	96.0	51.0	16830.0	330.0	358.09
96	97.0	48.0	15552.0	324.0	370.29
97	98.0	58.0	14231.0	245.36	268.51
98	99.0	54.0	18330.0	339.44	374.08
99	100.0	40.0	8547.0	213.68	237.42

Figure 2