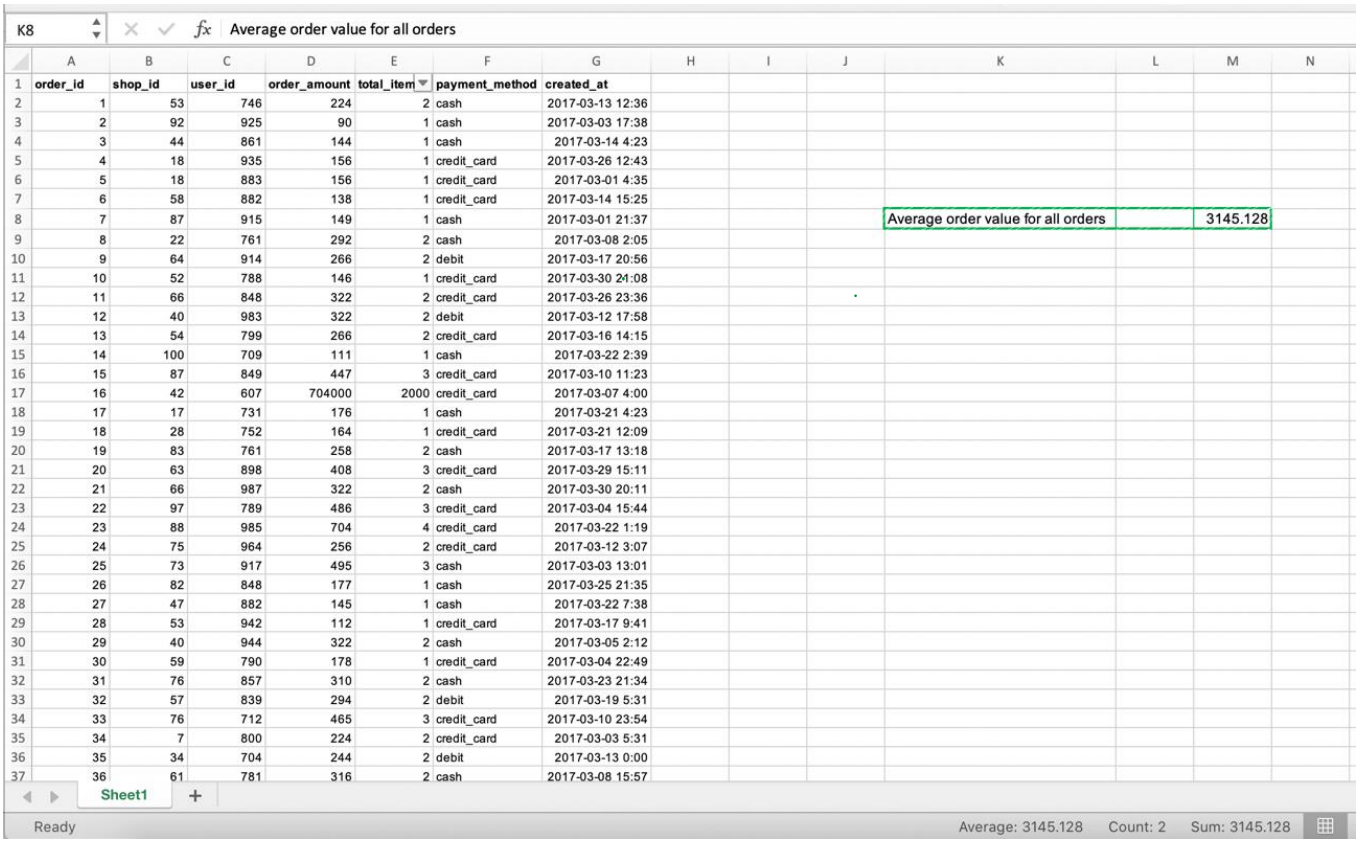
Mahtab Khan mahtabukhan16@gmail.com

Fall 2021 Data Science Intern Challenge

Question 1

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

The Average Order Value is calculated by dividing total revenue by number of orders. At first glance, a value of 3145.13 for AOV seems unreasonable for sneakers.



K8 | Average order value for all orders

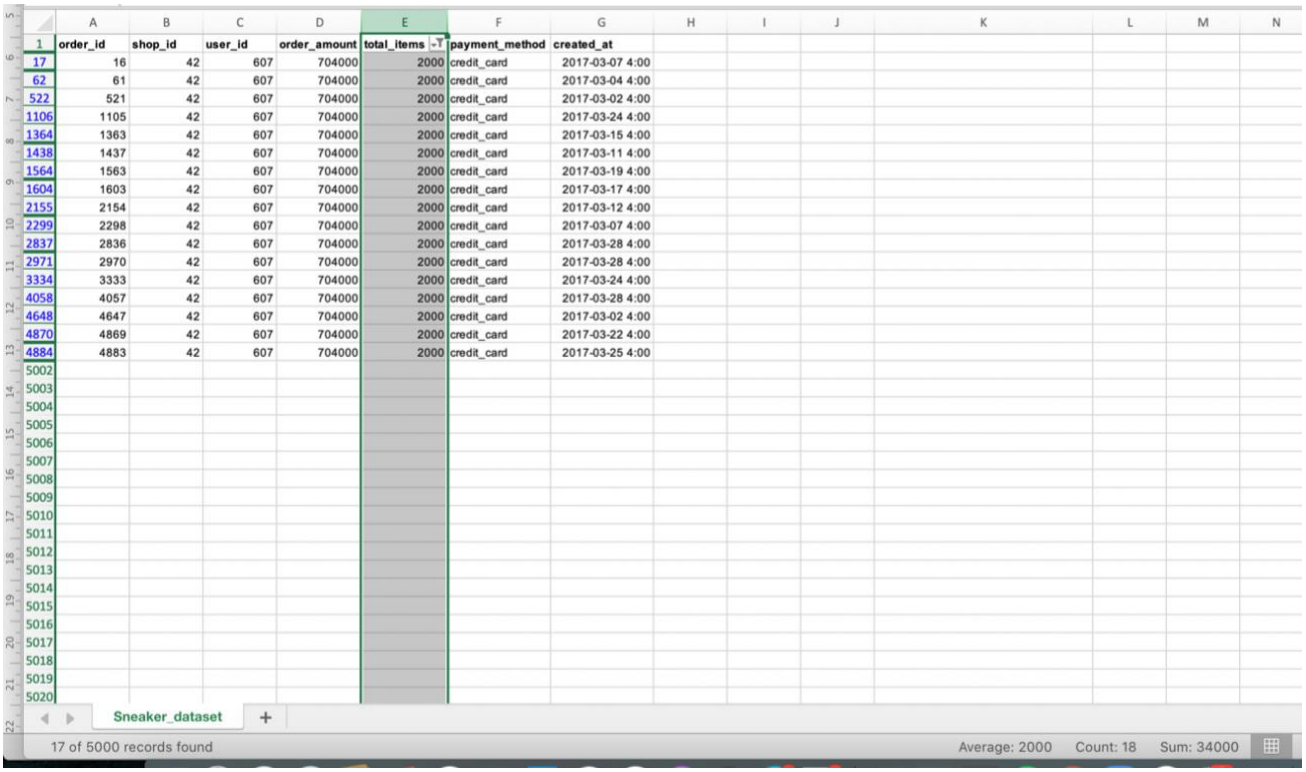
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	order_id	shop_id	user_id	order_amount	total_item	payment_method	created_at							
2	1	53	746	224	2	cash	2017-03-13 12:36							
3	2	92	925	90	1	cash	2017-03-03 17:38							
4	3	44	861	144	1	cash	2017-03-14 4:23							
5	4	18	935	156	1	credit_card	2017-03-26 12:43							
6	5	18	883	156	1	credit_card	2017-03-01 4:35							
7	6	58	882	138	1	credit_card	2017-03-14 15:25							
8	7	87	915	149	1	cash	2017-03-01 21:37				Average order value for all orders		3145.128	
9	8	22	761	292	2	cash	2017-03-08 2:05							
10	9	64	914	266	2	debit	2017-03-17 20:56							
11	10	52	788	146	1	credit_card	2017-03-30 21:08							
12	11	66	848	322	2	credit_card	2017-03-26 23:36							
13	12	40	983	322	2	debit	2017-03-12 17:58							
14	13	54	799	266	2	credit_card	2017-03-16 14:15							
15	14	100	709	111	1	cash	2017-03-22 2:39							
16	15	87	849	447	3	credit_card	2017-03-10 11:23							
17	16	42	607	704000	2000	credit_card	2017-03-07 4:00							
18	17	17	731	176	1	cash	2017-03-21 4:23							
19	18	28	752	164	1	credit_card	2017-03-21 12:09							
20	19	83	761	258	2	cash	2017-03-17 13:18							
21	20	63	898	408	3	credit_card	2017-03-29 15:11							
22	21	66	987	322	2	cash	2017-03-30 20:11							
23	22	97	789	486	3	credit_card	2017-03-04 15:44							
24	23	88	985	704	4	credit_card	2017-03-22 1:19							
25	24	75	964	256	2	credit_card	2017-03-12 3:07							
26	25	73	917	495	3	cash	2017-03-03 13:01							
27	26	82	848	177	1	cash	2017-03-25 21:35							
28	27	47	882	145	1	cash	2017-03-22 7:38							
29	28	53	942	112	1	credit_card	2017-03-17 9:41							
30	29	40	944	322	2	cash	2017-03-05 2:12							
31	30	59	790	178	1	credit_card	2017-03-04 22:49							
32	31	76	857	310	2	cash	2017-03-23 21:34							
33	32	57	839	294	2	debit	2017-03-19 5:31							
34	33	76	712	465	3	credit_card	2017-03-10 23:54							
35	34	7	800	224	2	credit_card	2017-03-03 5:31							
36	35	34	704	244	2	debit	2017-03-13 0:00							
37	36	61	781	316	2	cash	2017-03-08 15:57							

Sheet1 +

Ready | Average: 3145.128 | Count: 2 | Sum: 3145.128

The total items sold for 17 orders exceed 2000 orders worth 704000, this could be an anomaly in the dataset. The orders are from the same store and by the same customer. This could be due to a business to business transaction, which massively affects the AOV.

Mahtab Khan mahtabukhan16@gmail.com

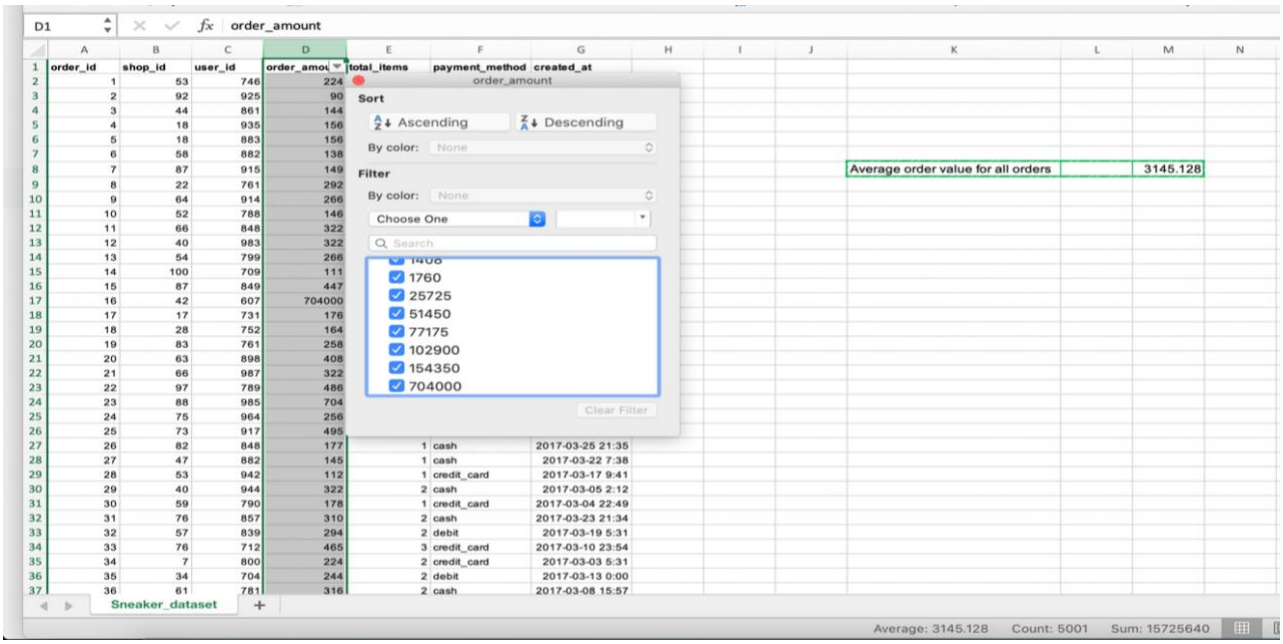


	order_id	shop_id	user_id	order_amount	total_items	payment_method	created_at
17	16	42	607	704000	2000	credit_card	2017-03-07 4:00
62	61	42	607	704000	2000	credit_card	2017-03-04 4:00
522	521	42	607	704000	2000	credit_card	2017-03-02 4:00
1106	1105	42	607	704000	2000	credit_card	2017-03-24 4:00
1364	1363	42	607	704000	2000	credit_card	2017-03-15 4:00
1438	1437	42	607	704000	2000	credit_card	2017-03-11 4:00
1564	1563	42	607	704000	2000	credit_card	2017-03-19 4:00
1604	1603	42	607	704000	2000	credit_card	2017-03-17 4:00
2155	2154	42	607	704000	2000	credit_card	2017-03-12 4:00
2299	2298	42	607	704000	2000	credit_card	2017-03-07 4:00
2837	2836	42	607	704000	2000	credit_card	2017-03-28 4:00
2971	2970	42	607	704000	2000	credit_card	2017-03-28 4:00
3334	3333	42	607	704000	2000	credit_card	2017-03-24 4:00
4058	4057	42	607	704000	2000	credit_card	2017-03-28 4:00
4648	4647	42	607	704000	2000	credit_card	2017-03-02 4:00
4870	4869	42	607	704000	2000	credit_card	2017-03-22 4:00
4884	4883	42	607	704000	2000	credit_card	2017-03-25 4:00

Sneaker_dataset

17 of 5000 records found Average: 2000 Count: 18 Sum: 34000

There are 4984 orders below 8 total items per transaction, which could be business to costumer transactions. Some order amounts range from 90 to 1760 could be purchases by individual customers, whereas order amounts in the range 25725 to 154350 could be purchases of designer/limited edition shoes by individuals or businesses. Any orders above that are purchases solely by businesses as mentioned earlier.



Mahtab Khan mahtabukhan16@gmail.com

1b. What metric would you report for this dataset?

By judging the dataset values, the distribution is skewed towards lower values, which would make the median value for order amounts a suitable fit for the AOV.

1c. What is its value?

Using the median method as shown in my program, the median value is equal to 224.

Question 2

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

```
SELECT COUNT(*) FROM [Orders]
WHERE ShipperID=1
```

Total number of orders shipped by Speedy Express is 54.

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

```
SELECT LastName FROM [Employees]
WHERE (EmployeeID = (SELECT EmployeeID FROM [Orders]
GROUP BY EmployeeId
ORDER BY COUNT(EmployeeId) DESC LIMIT 1))
```

The last name “Peacock” had the highest number of orders at 40.

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

```
SELECT [Customers].Country,
   [OrderDetails].ProductID,
   SUM([OrderDetails].Quantity) AS "TotalOrdered"
FROM [Orders]
JOIN [Customers]
   ON [Customers].CustomerID = [Orders].CustomerID
JOIN [OrderDetails]
   ON [OrderDetails].OrderID = [Orders].OrderID
WHERE [Customers].Country = 'Germany'
GROUP BY [OrderDetails].ProductID
ORDER BY TotalOrdered DESC
```

Following the product ID, we can see that Boston Crab Meat has the highest orders at 160.