Name : Vikram Reddy Sankepally

Question 1

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.

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

AOV tells the average of all orders. “Average” order value will only give you a partial picture of your customer's purchase behavior. We need to look at other statistics before making any decision.

```
data['order_amount'].describe()
✓ 0.1s
count       5000.000000
mean        3145.128000
std        41282.539349
min           90.000000
25%          163.000000
50%          284.000000
75%          390.000000
max       704000.000000
Name: order_amount, dtype: float64
```

1. Mean/Average will be biased towards the outliers.As you can see the data is not evenly distributed.
2. Mean =3145.12 ,Whereas 50% of orders are less than $184 and 75% of orders are less than $390. These values differ by a significant amount with AOV is $3145.12.
3. This tells us the AOV is highly influenced by the (large orders) outlier values and does not reflect the behavior of the entire data.

b. What metric would you report for this dataset?

As all the shops sell the same item and it's affordable,I would concentrate more on the **mode and median** of the purchases to better assess the orders.
Mode: Frequent orders in all the shops over the period of 30 days is $153.
Median: Where does the 50% of our orders stand?
50 % of the orders are less than $284
75% of our orders are less than $390
These values are far away from mean and these values represent the sales performance of the item.

c. What is its value?

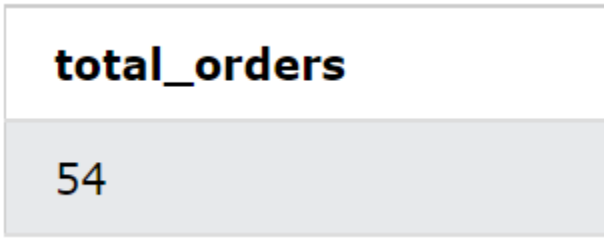
Median : $284
Mode : $153

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?

```
select count(*) as total_orders
From Orders
where ShipperID=(SELECT ShipperID FROM [Shippers] where ShipperName="Speedy Express")
```

-- a) ans: 54



total_orders
54

--

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

```
with max_order_emp as
(
SELECT EmployeeID,count(*) as cnt
FROM [Orders]
      group by EmployeeID
      order by cnt DESC
)
SELECT emp.LastName,max_ord.cnt as OrderCount
FROM [Employees] emp
      join max_order_emp max_ord
      on emp.EmployeeID=max_ord.EmployeeID
      order by max_ord.cnt DESC
```

-- b) ans: LastName:Peacock

LastName	OrderCount
Peacock	40
Leverling	31
Davolio	29
Callahan	27
Fuller	20
Suyama	18
King	14
Buchanan	11
Dodsworth	6

--

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

```
with germany_orders as
(
SELECT  ord.OrderID as OrderID
FROM [Orders]  ord
      join Customers cus
      on  ord.CustomerID=cus.CustomerID and cus.Country="Germany"
),

germany_productids as (
            SELECT OD.ProductID,count(*) cnt
            FROM [OrderDetails] od join germany_orders go
                  on  go.OrderID=od.OrderID
                  group by od.ProductID
                  order by cnt DESC
)
select p.ProductName ,germ_prod.cnt
From Products p join germany_productids germ_prod
      on p.ProductID =germ_prod.ProductID  order by germ_prod.cnt DESC
```

-- c) ans: ProductName:Gorgonzola Telino

ProductName	cnt
Gorgonzola Telino	5