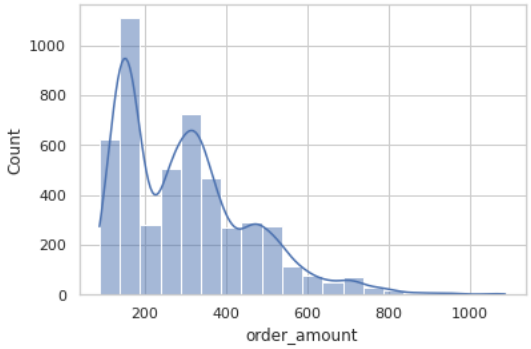
Shopify Technical Data Science Challenge

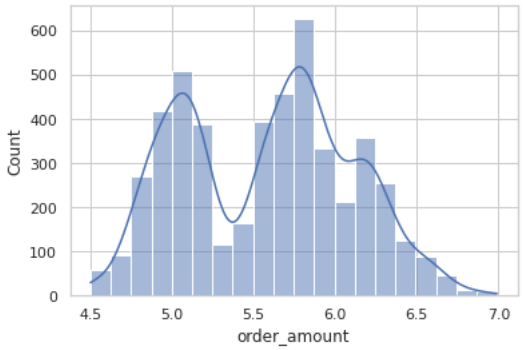
Apratim Mishra

Ans1) All the Answers are presented in the Colab attached as Well

Initial Right Skewed Histogram with Order_Amount



Histogram with Log of Order Amounts



1. Things wrong with initial Average Order Value (AOV) were outliers and the different kind of selling Activity for some shops

Shop id ‘42’ and ‘78’ had some outliers, for which details are loaded in the Colab file attached.

2. New Metric: Average of Log of Order Amount of shops. The median is a good statistic, but for any machine learning task -> log of order amount is a better stat. This is also because it transforms the target into a normal distribution; additionally, for each shop, we do not have large number of variables, so median is not a better stat than the average of log amounts. Below is the new dataset with metric provided. The average of log of order amount is 5.578

3. Finally, Shop '42' -> sells numerous items at once and Shop '78' has luxury goods sold; so, they will report their activity separately

Ans 2)

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

```
SELECT count(distinct o.OrderID)
FROM Orders o
Join Shippers s
on o.ShipperID = s.ShipperID
where s.ShipperID = 1
```

➔ total orders shipped is 54

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

```
select cte.LastName
from (
select e.EmployeeID, e.LastName, Count(o.OrderID)
from Employees e
join Orders o
where e.EmployeeID = o.EmployeeID
group by 1
order by 3 desc
limit 1 ) cte
```

➔ lastname is Peacock

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

```
select cte.ProductName from (
select p.ProductName, c.Country, sum(od.Quantity)
from Products p
join OrderDetails od on p.ProductID = od.ProductID
join Orders o on od.OrderID = o.OrderID'
join Customers c on o.CustomerID = c.CustomerID
where c.Country = 'Germany'
group by 1
order by 3 desc
limit 1 ) cte
```

'Boston Crab Meat' was ordered the most in Germany -> 160 times