E-Commerce
Data Set (Tata)

Presented by:

Avinash Kumar Kurrey

1. Introduction:

E-commerce has revolutionized the way businesses operate, offering unprecedented opportunities for growth and global reach. In this digital era, vast amounts of data are generated through online transactions, customer interactions, and website usage, presenting a treasure trove of insights for businesses to leverage. The analysis of e-commerce datasets plays a pivotal role in understanding customer behaviour, optimizing marketing strategies, and enhancing operational efficiency.

In this chapter performed descriptive analysis, we analyze some important variables applied histograms, bar plots and boxplots. Also, we retrieve the descriptive statistics of the variables draws good insights.

In chapter four we written managerial implications and insights by getting all the insights from descriptive analysis. finally, we conclude the report with what we have done and what we can do in future studies.

2. Descriptive Analysis:

Descriptive statistics describe, show, and summarize the basic features of a dataset found in each study, presented in a summary that describes the data sample and its measurements. It helps analysts to understand the data better. Descriptive statistics represent the available data sample and does not include theories, inferences, probabilities, or conclusions. That's a job for inferential statistics.

3. Data understanding:

We have a data contains total **223** records having **25** columns like Product Name, Customer Id, Channel, Logistic Name, Shipping state, shipping city, Billing state, Billing pin code, and Order status. below figure shows the sample data.

	SlrName	OrderId	TransactionId	Ussid	SKU	ProductName	CustomerId	Channel	OrderDate	P1SlaveId	...	P1LogisticsID	F
0	Aureole Atelier Private Limited	'230527-026-537061'	'125260035876948'	'125260MAYS1020204'	MAYS-1020-204	Mayhem MAYS-1020-204 Grey Aviatorlenscoloreyew...	1030631091	Web	00:20:14	125260-AANOIDA	...	DL	
1	Aureole Atelier Private Limited	'230107-024-259584'	'125260032491445'	'125260MAYS1020204'	MAYS-1020-204	Mayhem MAYS-1020-204 Grey Aviatorlenscoloreyew...	1029804416	Web	00:12:44	125260-AANOIDA	...	DL	
2	Aureole Atelier Private Limited	'230404-025-694594'	'125260034725233'	'125260MAYS1020204'	MAYS-1020-204	Mayhem MAYS-1020-204 Grey Aviatorlenscoloreyew...	1046637229	Web	00:29:47	125260-AANOIDA	...	BD	
3	Aureole Atelier Private Limited	'230509-026-203436'	'125260035391265'	'125260MAYS1020204'	MAYS-1020-204	Mayhem MAYS-1020-204 Grey Aviatorlenscoloreyew...	1049786162	Mobile	00:07:14	125260-AANOIDA	...	DL	
4	Aureole Atelier Private Limited	'231215-029-986594'	'125260040565269'	'125260MAYS1020204'	MAYS-1020-204	Mayhem MAYS-1020-204 Grey Aviatorlenscoloreyew...	1003451931	Mobile	00:17:56	125260-AANOIDA	...	XC	
...	...	...	...	...	...	...	...	...	...	...	...	...	
218	Aureole Atelier Private Limited	'231204-029-818936'	'125260040357351'	'125260MAYS9008060'	MAYS-9008-060	Mayhem MAYS-9008-060 Purple Square Sunglassesl...	1025601133	Mobile	00:04:51	125260-AANOIDA	...	XC	
219	Aureole Atelier Private Limited	'230225-025-176659'	'125260034076107'	'125260MAYS9008060'	MAYS-9008-060	Mayhem MAYS-9008-060 Purple Square Sunglassesl...	1036710437	Mobile	00:50:12	125260-AANOIDA	...	BD	
220	Aureole Atelier Private Limited	'230319-025-468989'	'125260034437071'	'125260MAYS9008060'	MAYS-9008-060	Mayhem MAYS-9008-060 Purple Square Sunglassesl...	1000861250	Mobile	00:21:49	125260-AANOIDA	...	DL	
221	Aureole Atelier Private Limited	'230609-026-813870'	'125260036252350'	'125260MAYS9008060'	MAYS-9008-060	Mayhem MAYS-9008-060 Purple Square Sunglassesl...	1034148602	Mobile	00:30:41	125260-AANOIDA	...	DL	
222	Aureole Atelier Private Limited	'230609-026-809759'	'125260036247162'	'125260MAYS9008060'	MAYS-9008-060	Mayhem MAYS-9008-060 Purple Square Sunglassesl...	1046807158	Mobile	00:33:57	125260-AANOIDA	...	DL	

223 rows × 25 columns

Figure 1. Sample data

There are **5** numerical variables, **20** categorical variables shown in below figure 2.

```
SlrName                object
OrderId                object
TransactionId          object
Ussid                  object
SKU                    object
ProductName            object
CustomerId              int64
Channel                object
OrderDate              object
P1SlaveId              object
parentproductName      object
Price                   int64
ShippingCharge          int64
IsCOD                  object
FulfillmentType        object
P1LogisticsID          object
P1LOGISTICNAME         object
ReturnLogisticsID      object
ShippingCity           object
ShippingState          object
ShippingPincode         int64
BillingCity            object
BillingState           object
BillingPincode          int64
OrderStatus            object
```

Figure 2. Data types of Tata

4. Missing Values Analysis:

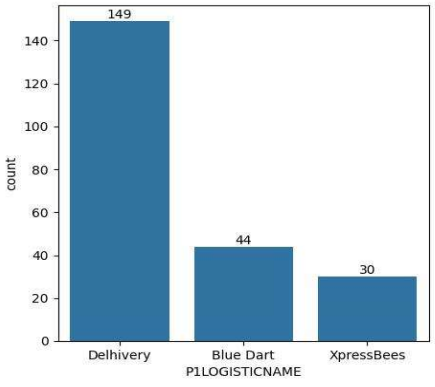
```
SlrName                0
OrderId                0
TransactionId          0
Ussid                  0
SKU                    0
ProductName            0
CustomerId             0
Channel                0
OrderDate              0
P1SlaveId              0
parentproductName      0
Price                  0
ShippingCharge         0
IsCOD                  0
FulfillmentType        0
P1LogisticsID          0
P1LOGISTICNAME         0
ReturnLogisticsID    180
ShippingCity           0
ShippingState          0
ShippingPincode        0
BillingCity            0
BillingState           0
BillingPincode         0
OrderStatus            0
dtype: int64
```

Figure 3. Missing values

- When we analyze the data we found that there is only one column ,which have some missing value.
- But we see that the count of missing values is **180** out of **223** in ReturnLogistics ID column.
- We see that is 80% of the value that's why we decide that to drop this column because it is no possible to fill that too much missing value.
- Except this column no one column have missing values.

5. Channel, Logistic Name and Cash on Delivery (COD) Analysis:

Channel, Logistic name and COD are three variables in this data set, these all are categorical variable.

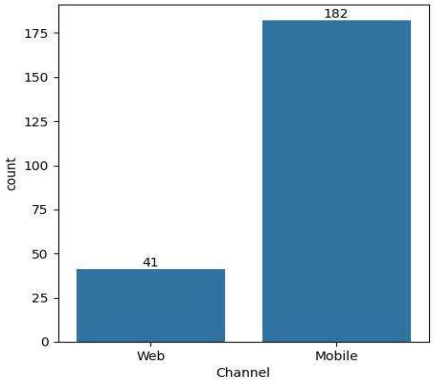


Sr. No.	Logistic Name	Percentage (%)
1.	Delhivery	66.82
2.	Blue Dart	19.73
3.	XpressBees	13.45

Figure 4. Bar Graph of Logistic Name

From the above bar graph, we conclude that there is total three logistic way to deliver the product. These are the following

1. Delhivery: From Delhivery we deliver **149** products from the total.
2. Blue Dart: From Blue Dart we deliver **44** products from the total.
3. XpressBess: From XpressBess we deliver **30** products from the total.

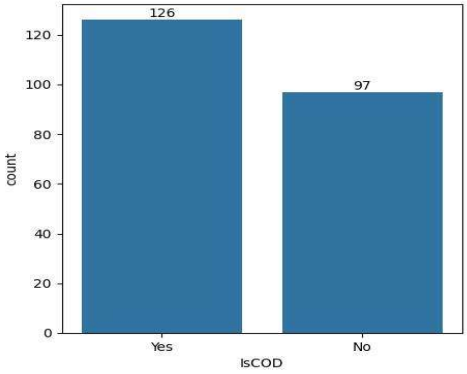


Sr. No.	Channel	Percentage (%)
1.	Web	18.39
2.	Mobile	81.61

Figure 5. Bar Graph of Channel

From the above bar graph, we conclude that there is total two way by which the customer placed their orders. These are the following

1. Web: By Web there are total **41** people placed their orders.
2. Mobile: By Mobile there are total **182** people placed their orders.



Sr. No.	COD	Percentage (%)
1.	Yes	56.50
2.	No	43.50

Figure 6. Bar Graph of COD

From the above bar graph, we conclude that there is total **126** customer to placed their order in COD and **97** customers are not select COD option i.e. they select other option to pay the bill.

6. Price and Sipping Charge Analysis:

Price and Shipping charge are two variables in this data set, these both columns are numerical variable. We have different price range and according to their price different shipping charge applicable on that product.

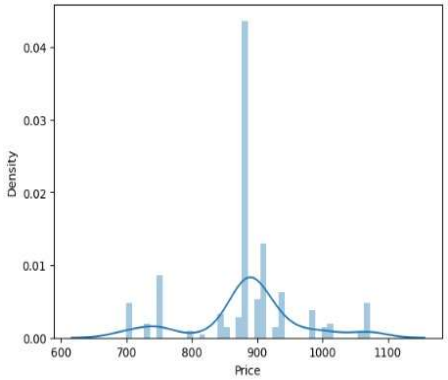


Figure 7. Histogram of Price

- From this histogram we conclude that the price range of our product is in between **600 to 1200** from that we know most of the product bought by the costumer in between **800 to 900** price range.

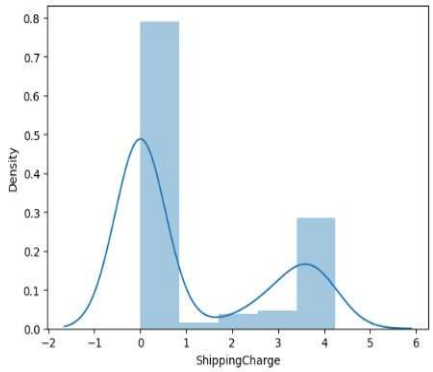


Figure 8. Histogram of Shipping Charge

- From this histogram we apply shipping charge on the product in between **0** to **6** rupees, from this we conclude that we didn't apply any charges on most of the product that is on **144 (80%)** product we not apply any shipping charges on that product.

7. Box Plot:

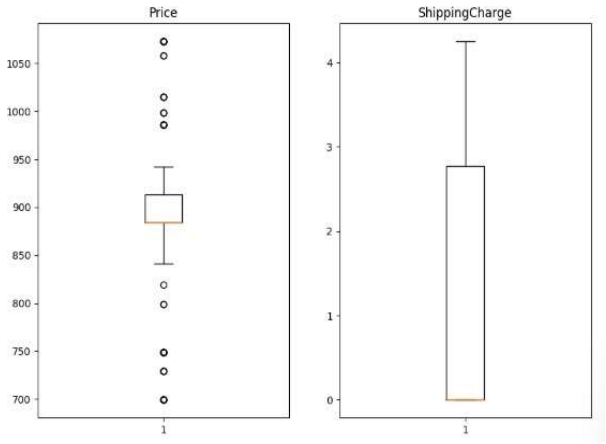


Figure 9. Boxplot of Shipping Charge and price

- From these two boxplots we conclude that:
- From the price boxplot we clearly see that there are some outliers but these outliers are discrete value so no need to remove the outliers.
- From the shipping charge boxplot, we see that there are no outliers that means our data is normal.

8. Top Selling Products:

ProductName	count
Mayhem MAYS-8019-003 Brown Rectangular Sunglasseslenscoloreyewearbrownsizeeyewearlarge	67
Mayhem MAYS-8019-003 Brown Rectangular Sunglasseslenscoloreyewearbrownsizeeyewearfreesize	23
Mayhem MAYS-1042-104 Purple Butterfly Sunglasseslenscoloreyewearpurplesizeeyewearlarge	14
Mayhem MAYS-1022-201 Brown Rectangular Sunglasseslenscoloreyewearbrownsizeeyewearfreesize	13
Mayhem MAYS-8017-109 Grey Rectangular Sunglasseslenscoloreyeweargreysizeeyewearfreesize	6

Figure 10. Data Frame of Top selling product

- From the above data we clearly see that first product from the data is highly demandable items. We sale total **67** units of this product out of our total selling unit **223**.
- From the above data we see that second product from the data is also demandable items. We sale total **23** units of this product out of our total selling unit **223**.
- Except these two items rest items are also in demand but above two items are more demandable.

9. State with their Delivered counts:

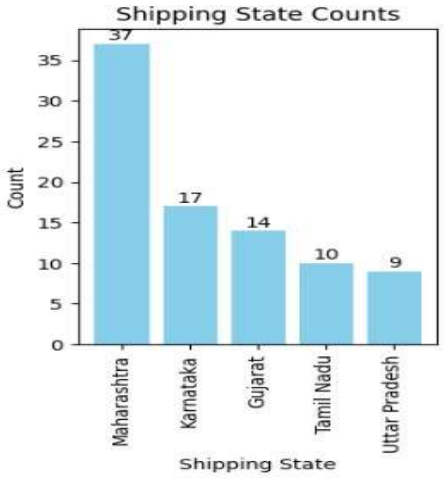


Figure 11. bar graph of delivered counts

- From this bar graph we clearly see that Maharashtra is the top most state in which our product is most delivered and the delivered count is **37**.
- After that Karnataka, Gujrat, Tamil Nadu and utter Pradesh are the top state in which our product is most delivered and there are **17**, **14** , **10**, **9.**

10. Most Ordered States:

ShippingState	All
All	223
Maharashtra	61
Karnataka	29
Gujarat	19
Uttar Pradesh	15
Telangana	11
Tamil Nadu	11
Delhi	10
Odisha	9
Andhra Pradesh	9

- From the above data we see that total **223** product are ordered.
- In the **223** product **61** product are just ordered by the Maharashtra customer i.e. In Maharashtra our product is highly demandable.
- After that Maharashtra the Karnataka are the second state in which our product it also in demand, and total **29**-unit items are bought by the Karnataka Customers.
- In Gujarat, Uttar Pradesh, Telangana and some other states like Tamil Nadu, Delhi ,Odisha, and Andhra Pradesh our product are in demand and ordered by the customers of these states.

Figure 12. Most Ordered Product

11. Product Count with Price and Order Status with COD Analysis:

Price	Count
884	91
913	27
749	18
942	13
899	11

- From these data we clearly see that the **91**-products cost price is **884** out of **223** products, that means **884** is very common price in all the products.
- After that we see that **27**-product cost is **913**. This is the second most common price in all the products.
- After these two the common price is **749**, **942** and **899**. These are also common price in this dataset.

Figure 13. Product Count and Price

OrderStatus	IsCOD
Delivered	138
QC PASS	25
Closed on cancellation	19
RTO_DELIVERED	19
Return Delivered to Seller	9

- This is the data of order status and that order is COD or not.
- So, we clearly see that most of the delivered items are ordered on COD, that is **138** items are ordered on COD out of **223** items.
- Most of the QC PASS the item are also ordered on COD.
- QC PASS, closed on cancellation, RTO_DELIVERED this ordered status are also ordered on the COD and their count are **19**, **19**, **9**.

Figure 14. Order Status with COD Analysis

12. Conclusion:

Understanding the e-commerce dataset is very useful in any online business. There might be influence this many factors, understanding the good factor is very important in this report we understood that which state costumers are most like our product, which product are most demandable. from this report we also conclude that which price range is more liked by the costumer. After this we also easily understand which color product, size and shape are mostly demand by the costumer. After that we also easily know that which state people most buy product on COD. From all above the data we finally understand that how the easily increase our sale and how we gain mor number of costumer and that is very useful for our business growth.