Name:	Chaitya Arun Dobariya
UID:	2021600017
Course:	Advanced Data Visualization

Experiment 1

Aim:	Create basic charts using Tableau / Power BI / R / Python / D3.js to be performed on the dataset of the E-commerce field Complete all plots on the practice dataset and reproduce them on the e-commerce dataset. Basic - Bar chart, Pie chart, Histogram, Timeline chart, Scatter plot, Bubble plot Calculate Product sales, region sales Write observations from each chart

1. **Dataset**

You can find the dataset here.

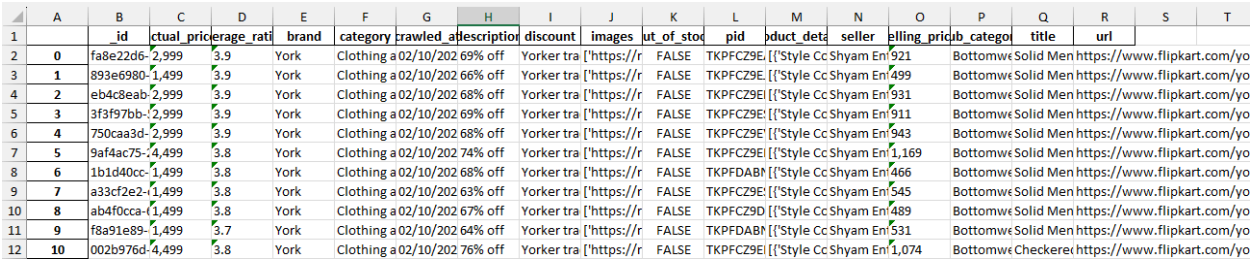
Description

E-commerce Product dataset - 5.7 million + records Flipkart dataset
Flipkart is the largest e-commerce website in India. The pre-crawled dataset has more than 5.7 million records.

Fields

Url: URL from Flipkart (String)
Name: Name of the product (String)
_id: Unique identifier (String)
Crawled_at: Date and Time when the product data was crawled (DateTime)
Selling_price: Current selling price (Integer)
Brand: Brand or Manufacturer of the product (String)
Seller_name: Name of the seller offering the product (String)
Seller_rating: Rating of the seller (Float)
Images: Links to the product images (String)
Product_details: Additional details about the product (JSON)

Pid: product ID (String)
Description: Description of the product (String)
Out_of_stock: flag indicating whether the product is currently out of stock (Boolean)



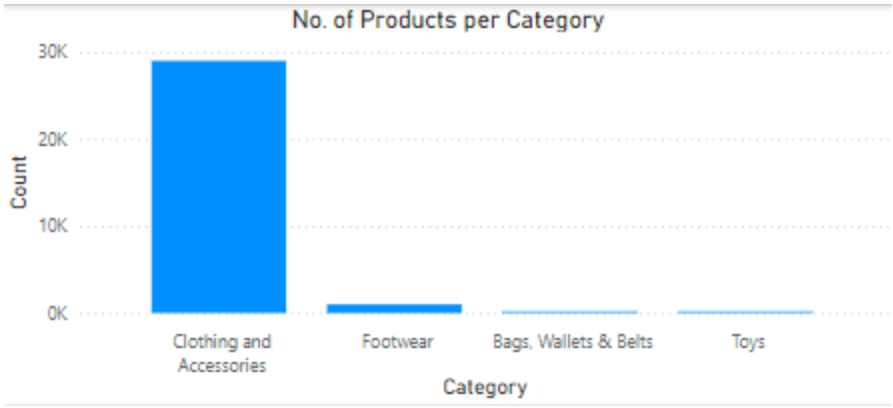
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1		_id	ctual_pric	erage_rati	brand	category	crawled_at	description	discount	images	ut_of_stoc	pid	oduct_deta	seller	elling_pric	ub_categor	title	url		
2	0	fa8e22d6-	2,999	3.9	York	Clothing a	02/10/202	69% off	Yorker tra	['https://r	FALSE	TKPFCZ9E	[{'Style Co	Shyam Ent	921	Bottomwe	Solid Men	https://www.flipkart.com/yo		
3	1	893e6980-	1,499	3.9	York	Clothing a	02/10/202	66% off	Yorker tra	['https://r	FALSE	TKPFCZ9E	[{'Style Co	Shyam Ent	499	Bottomwe	Solid Men	https://www.flipkart.com/yo		
4	2	eb4c8eab-	2,999	3.9	York	Clothing a	02/10/202	68% off	Yorker tra	['https://r	FALSE	TKPFCZ9E	[{'Style Co	Shyam Ent	931	Bottomwe	Solid Men	https://www.flipkart.com/yo		
5	3	3f3f97bb-	2,999	3.9	York	Clothing a	02/10/202	69% off	Yorker tra	['https://r	FALSE	TKPFCZ9E	[{'Style Co	Shyam Ent	911	Bottomwe	Solid Men	https://www.flipkart.com/yo		
6	4	750caa3d-	2,999	3.9	York	Clothing a	02/10/202	68% off	Yorker tra	['https://r	FALSE	TKPFCZ9E	[{'Style Co	Shyam Ent	943	Bottomwe	Solid Men	https://www.flipkart.com/yo		
7	5	9af4ac75-	4,499	3.8	York	Clothing a	02/10/202	74% off	Yorker tra	['https://r	FALSE	TKPFCZ9E	[{'Style Co	Shyam Ent	1,169	Bottomwe	Solid Men	https://www.flipkart.com/yo		
8	6	1b1d40cc-	1,499	3.8	York	Clothing a	02/10/202	68% off	Yorker tra	['https://r	FALSE	TKPFDABI	[{'Style Co	Shyam Ent	466	Bottomwe	Solid Men	https://www.flipkart.com/yo		
9	7	a33cf2e2-	1,499	3.8	York	Clothing a	02/10/202	63% off	Yorker tra	['https://r	FALSE	TKPFCZ9E	[{'Style Co	Shyam Ent	545	Bottomwe	Solid Men	https://www.flipkart.com/yo		
10	8	ab4f0cca-	1,499	3.8	York	Clothing a	02/10/202	67% off	Yorker tra	['https://r	FALSE	TKPFCZ9D	[{'Style Co	Shyam Ent	489	Bottomwe	Solid Men	https://www.flipkart.com/yo		
11	9	f8a91e89-	1,499	3.7	York	Clothing a	02/10/202	64% off	Yorker tra	['https://r	FALSE	TKPFDABI	[{'Style Co	Shyam Ent	531	Bottomwe	Solid Men	https://www.flipkart.com/yo		
12	10	002b976d-	4,499	3.8	York	Clothing a	02/10/202	76% off	Yorker tra	['https://r	FALSE	TKPFCZ9E	[{'Style Co	Shyam Ent	1,074	Bottomwe	Checkere	https://www.flipkart.com/yo		

2. Data Preprocessing

- Removed the columns not required for visualization.

3. Charts & Plots

3.1 Bar Chart:

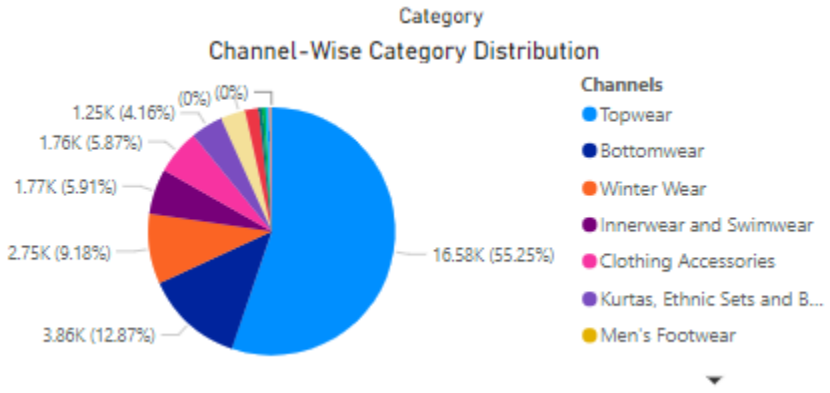


Observation:

- The bar chart shows the total sales of each product category
- The category ‘Clothing and Accessories’ has the highest sales
- The category ‘Toys’ has the lowest sales

This suggests that the ‘Clothing and Accessories’ category has the most sales and should be focused more.

3.2 Pie Chart

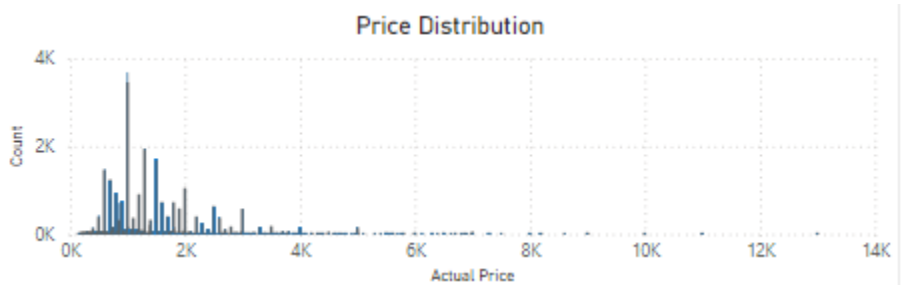


Observation:

- The pie chart shows how the subcategories are distributed among the sales
- Most sales are from the ‘Topwear’ subcategory followed by ‘Bottomwear’ then ‘Winter Wear’ and so on.

This suggests that the ‘Topwear’ subcategory has the most potential for sales and should be advertised more and manage its inventory accordingly.

3.3 Histogram

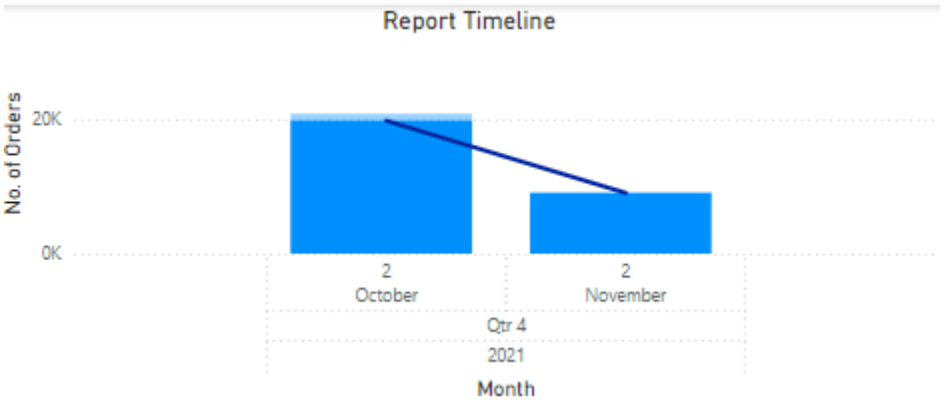


Observation:

- The histogram shows the distribution of the actual price per subcategory
- Most products lie between 1K to 2K

This suggests that most products are priced reasonably and are affordable to customers.

3.4 Timeline Chart



Observation:

- The time-based chart shows the monthly orders of the company
- The sales have dropped from month of October to November for the 4th Quarter of 2021

This suggests that the month of October was the most profitable for Flipkart due to the Diwali season sales.

3.5 Scatter Plot

Observation:

- The scatter plot shows the relationship between the quantity of categories sold and the amount of sales
- The ‘Clothing and Accessories’ have contributed the most

3.6 Bubble Plot

Observation:

- The bubble plot shows the relationship between the number of categories sold and the number of sales, with the size of the bubble representing the brand of the product

Conclusion

In this experiment, I learned how to work with PowerBI and also create basic charts on the e-commerce dataset. I created the following plots on the practice dataset and reproduced them on the e-commerce dataset: bar chart, pie chart, histogram, timeline

chart, scatter plot and bubble plot. I have also written observations for each chart to gain insights into the data.