Advanced Data Visualization Lab
Sem 7 B.Tech. Program Elective
Advait Ravi Sapkal 2021700055

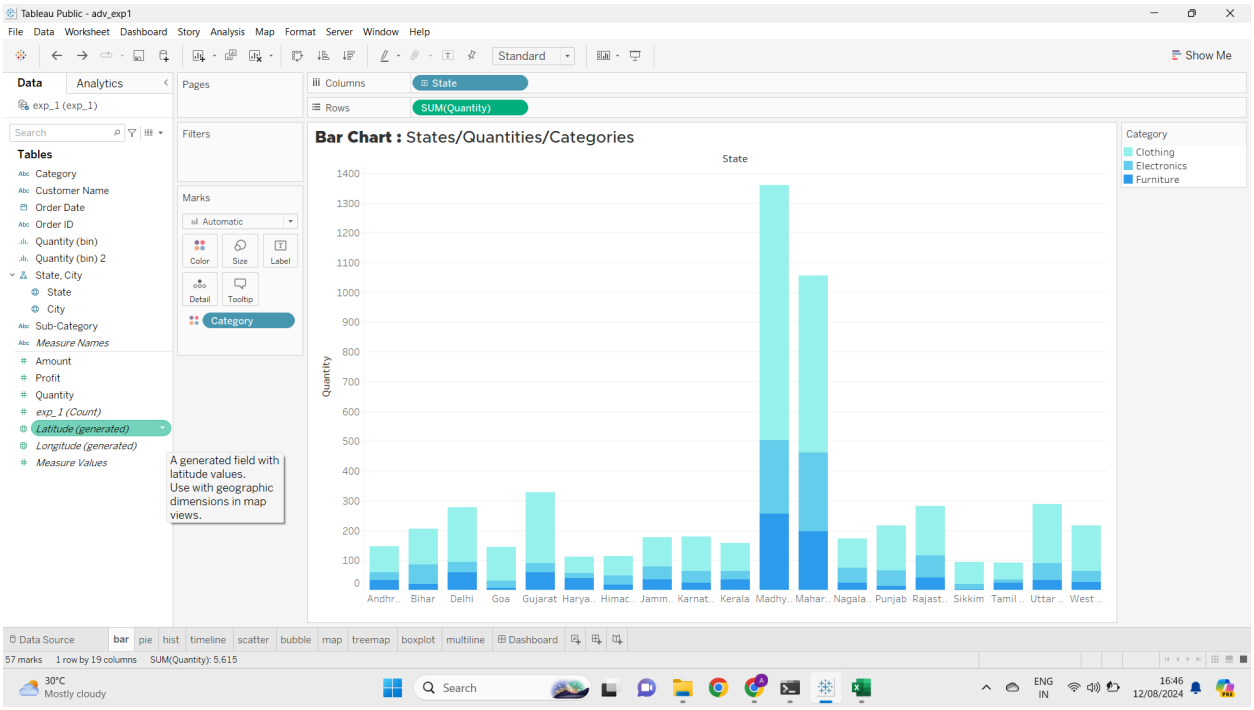
Expt 1: E-Commerce

Problem Statement: Create basic charts using Tableau / Power BI / R / Python / D3.js to be performed on the dataset of Ecommerce field

- Complete all plots on practice dataset and reproduce on e-commerce dataset.
- Basic - Bar chart, Pie chart, Histogram, Timeline chart, Scatter plot, Bubble plot
- Calculate Product wise sales, region wise sales
- Write observations from each chart

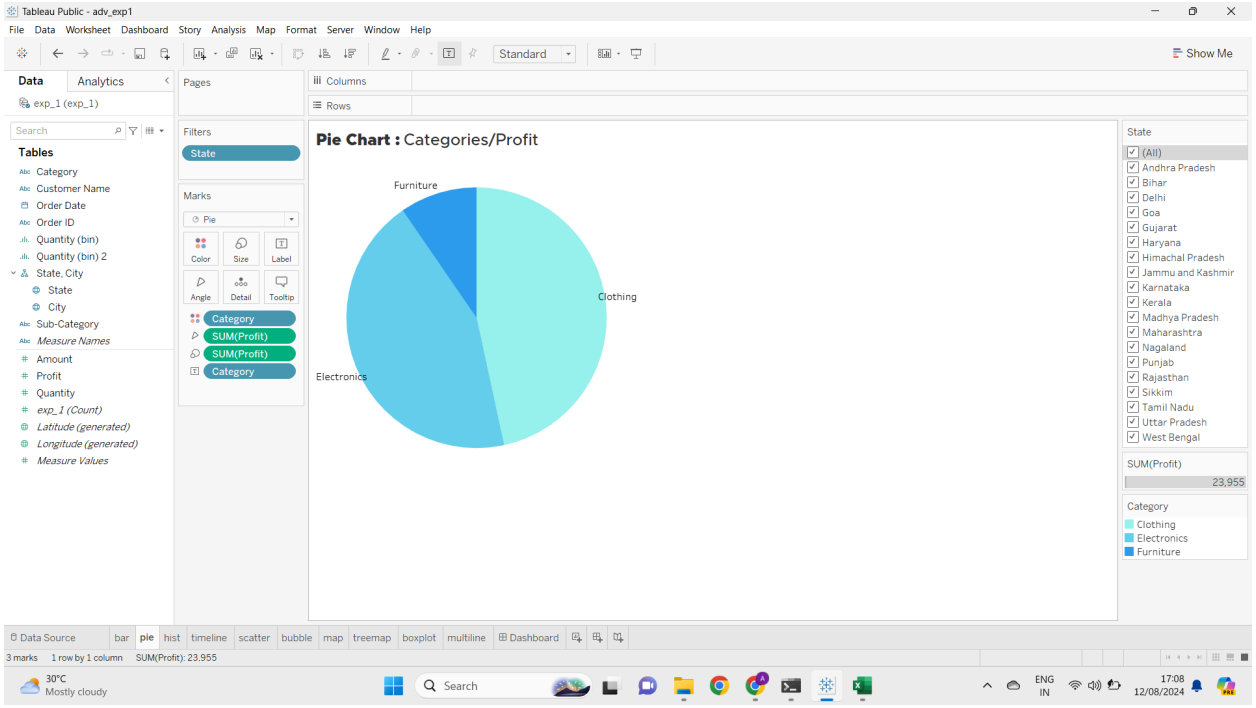
Software Used: Tableau Public, Pandas
Dataset Used: [E-Commerce Data]

1. Bar Chart



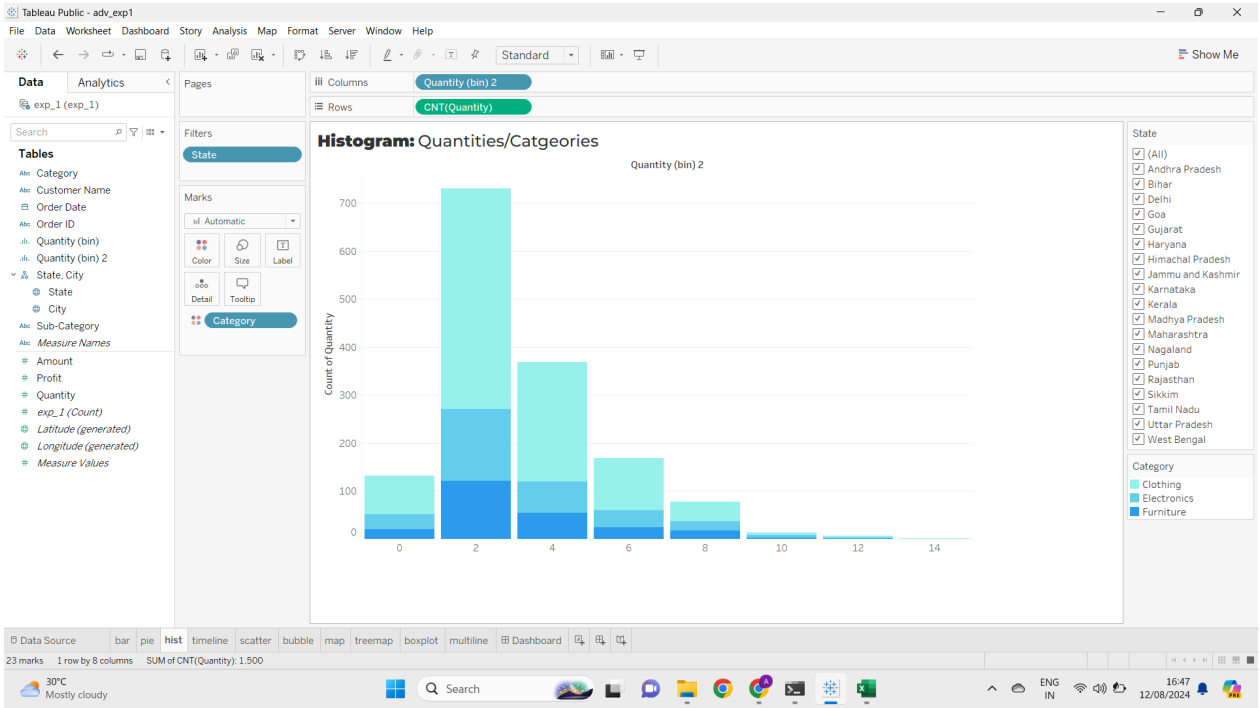
Here it is clearly seen that Madhya Pradesh and Maharashtra have significantly higher quantities ordered compared to other states. The general trend showing clothing is the most seeked category in all states. The second most seeked is mainly electronics with some exceptions like Delhi, Haryana, Tamil Nadu where it is Furniture.

2. Pie Chart



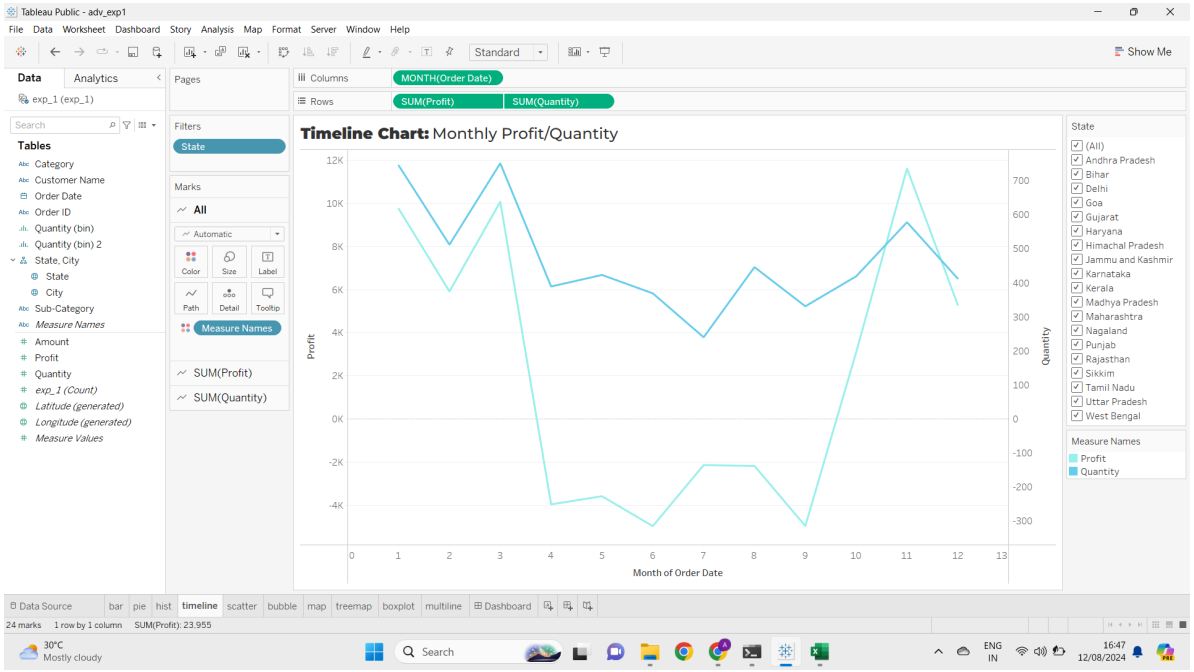
Clothing and Electronics are found to be equally contributing to the profit, whereas the profit margin in furniture is significantly lesser. These trends vary statewise.

3. Histogram



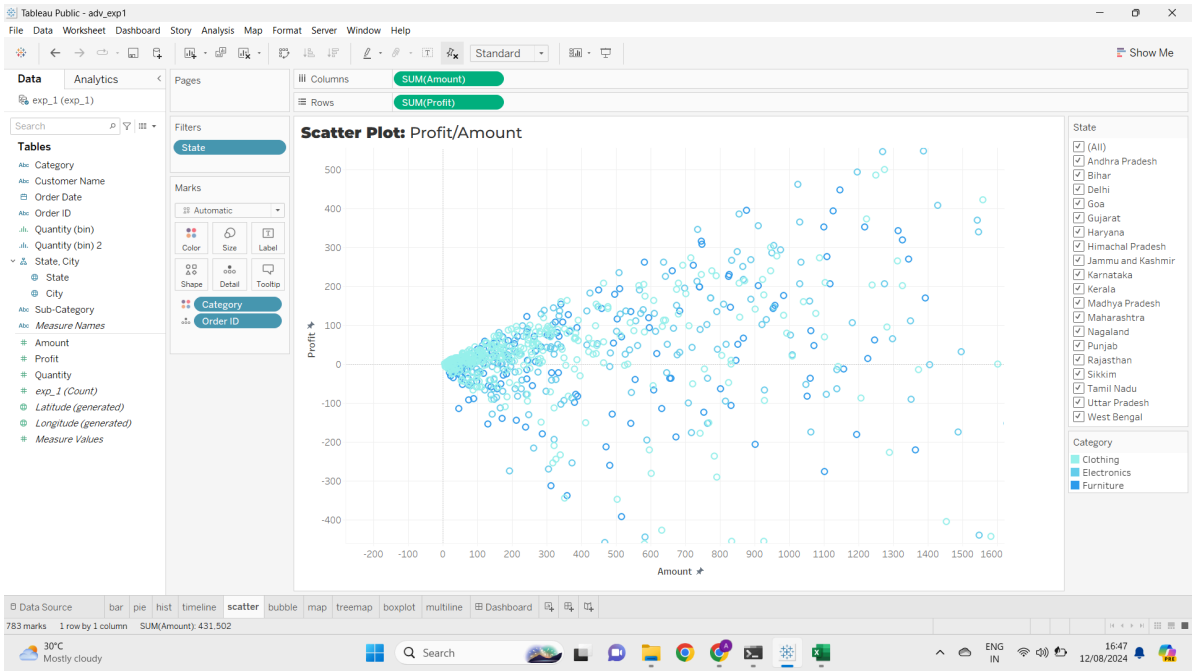
People tend to buy products in quantities 2-4, followed by 4-6. The distribution is approximately normal, which is to be expected of a natural process.

4. Timeline based graph



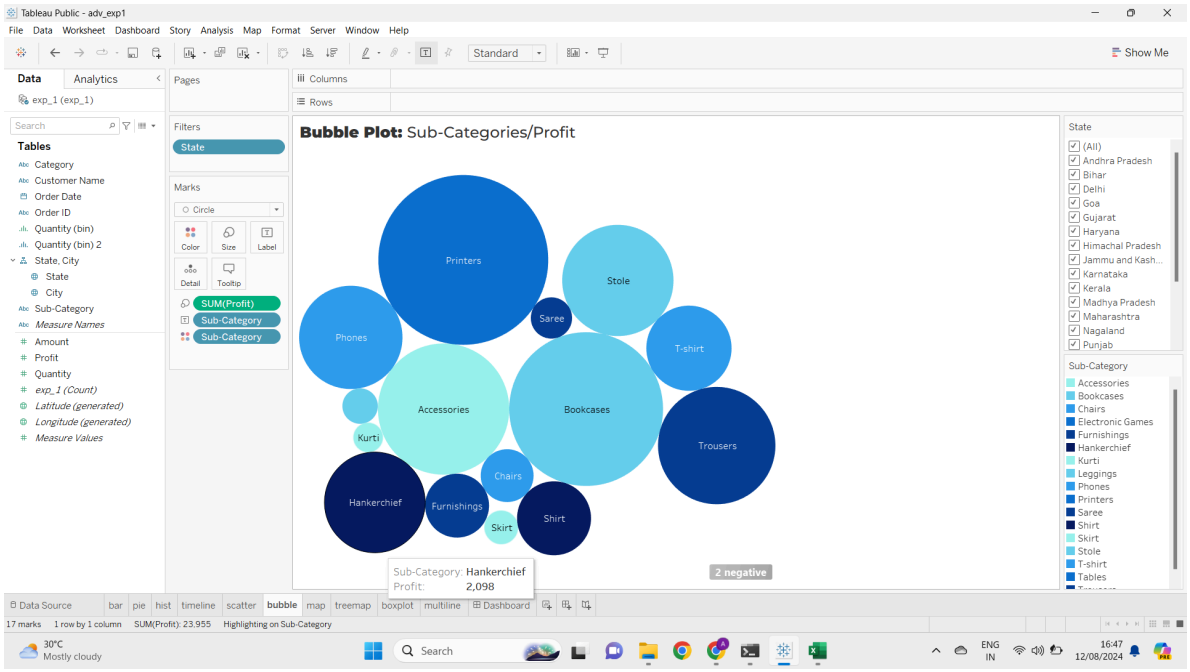
There is a sharp dip in order quantities in the months April-September. The business must revise the strategies they use in these months. The Profit closely follows the trend in the quantities.There are sharp peaks during early March and late October.

5. Scatter Plot



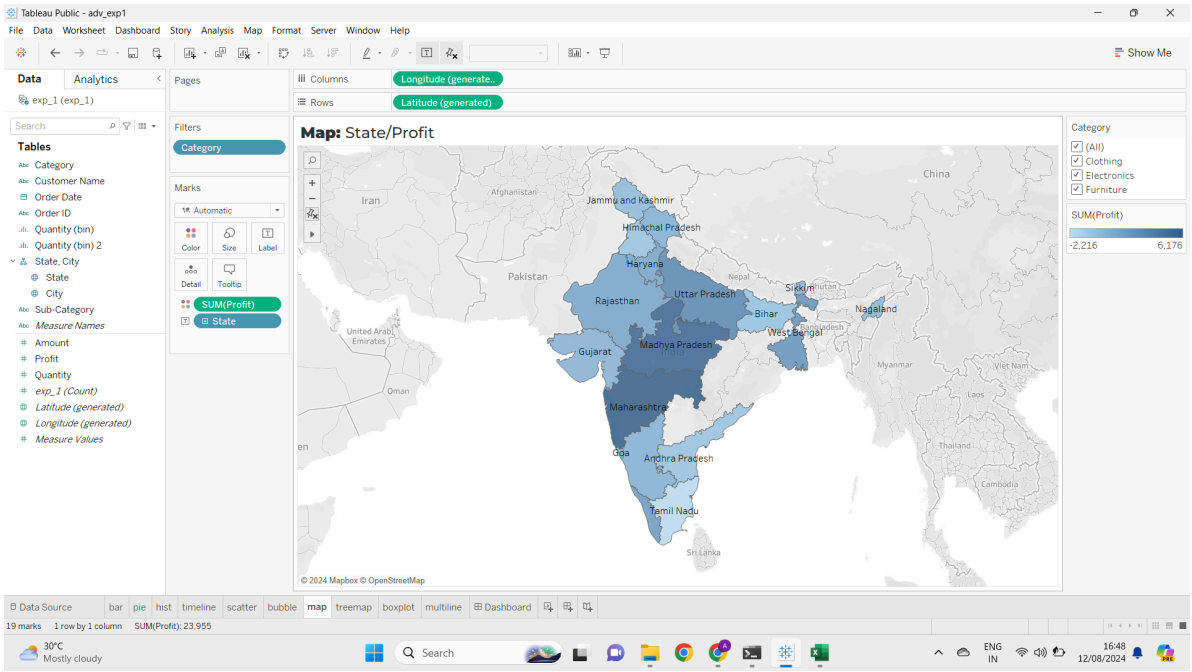
The amount has approximately a linear relationship with either positive or negative profit (loss). The clothing category however seems less affected by this trend.

6. Bubble Plot



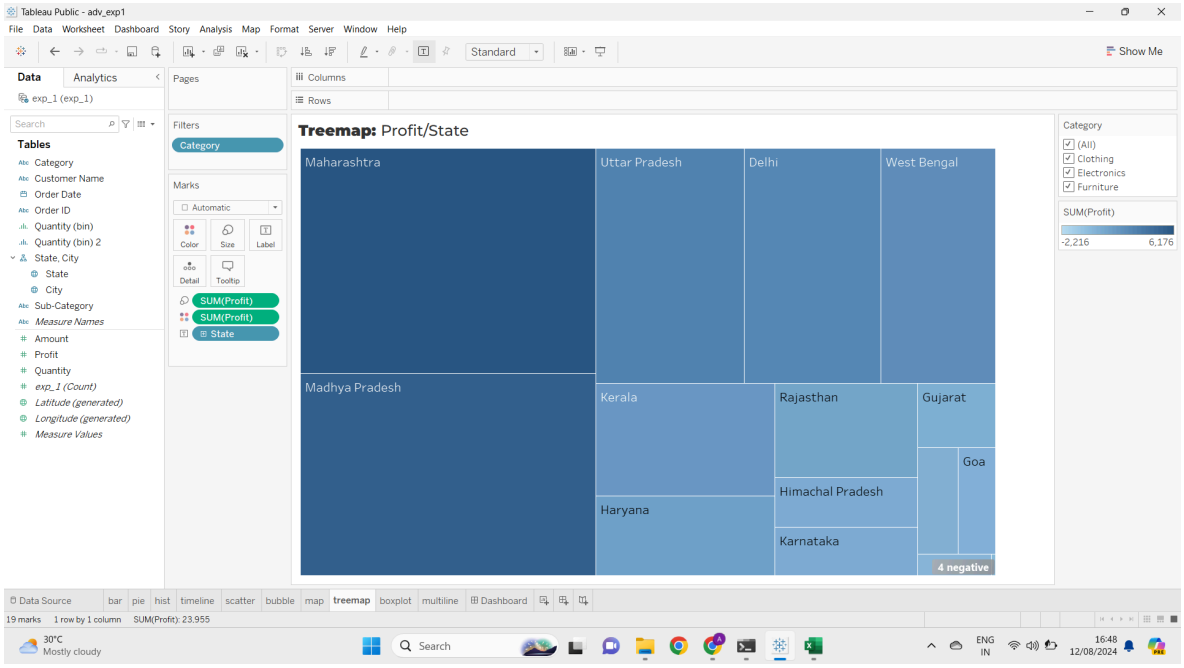
In the sub categories we see that the Printers, Bookcases and Accessories yield the most profit margin. Their sales and marketing must be encouraged.

7. Map



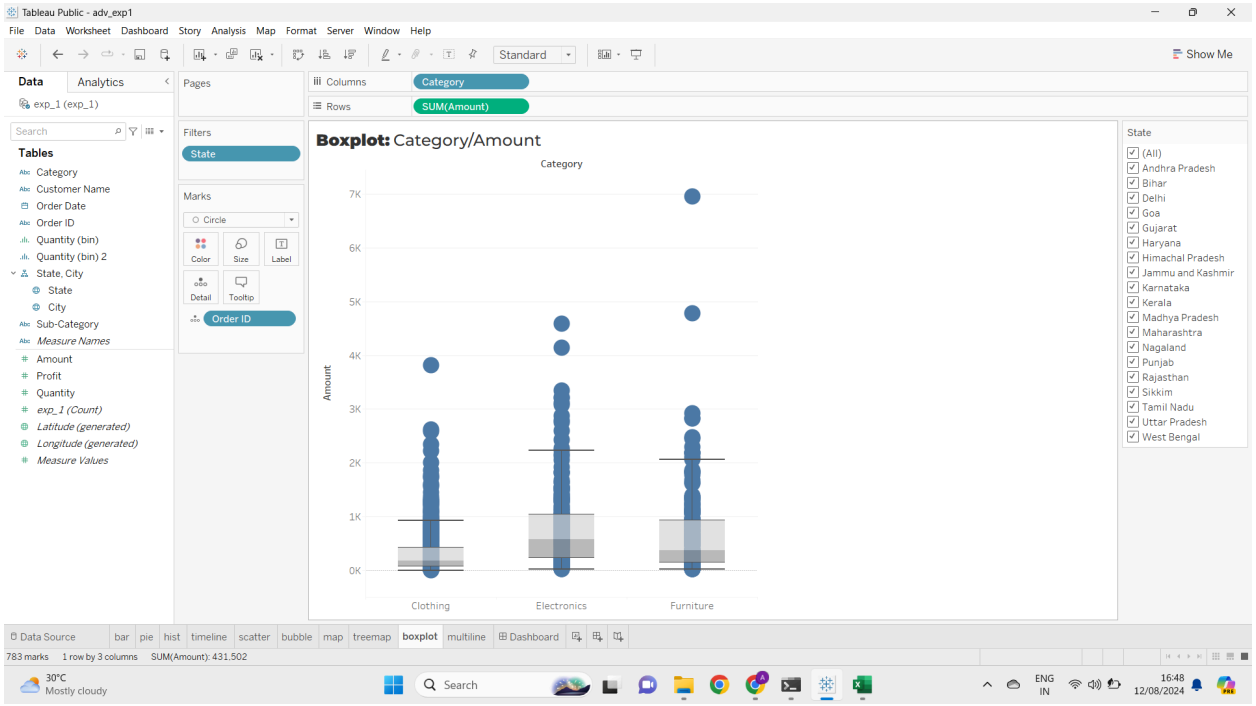
Maharashtra, Madhya Pradesh, Uttar Pradesh, West Bengal are the highest contributors to the profit. The sales must be expanded more in East India like Jharkhand, Odisha, Telangana.

8. Treemap



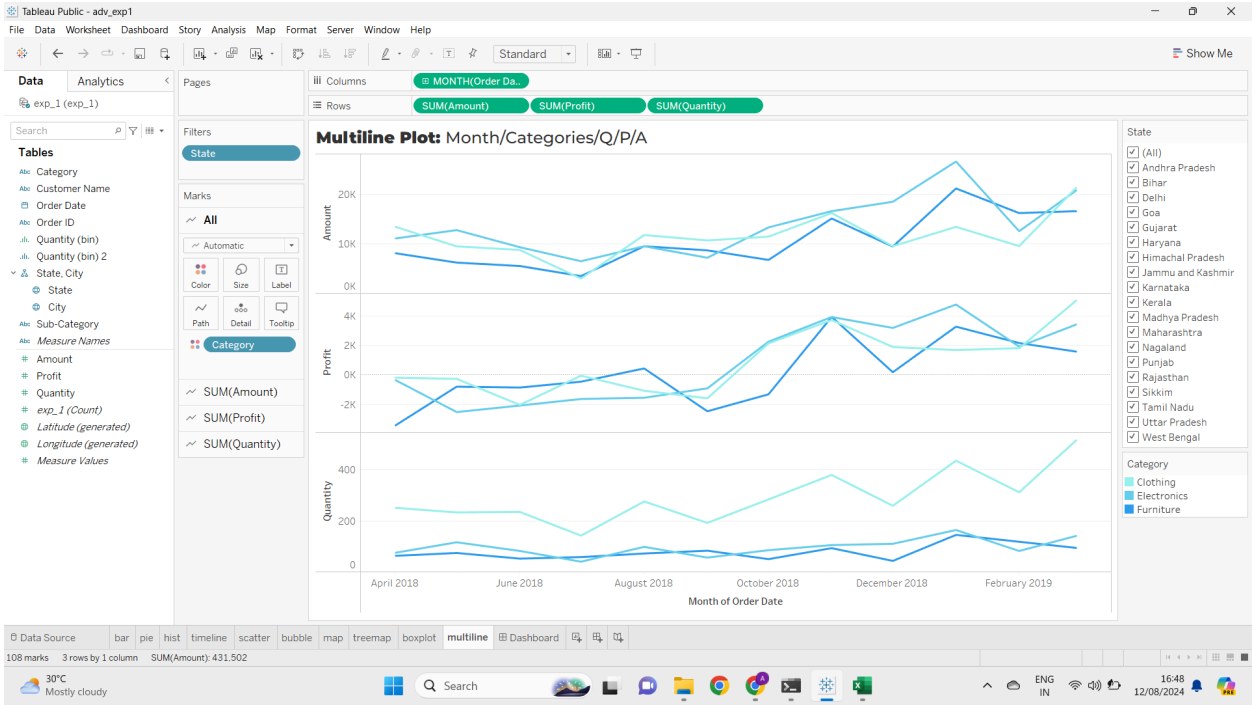
This is a different way to visualize the above data/observation.

9. Boxplot



Box plots can be used to detect outliers and understand the general distribution of the numerical variables, such as median, 1st quartile, 3rd quartile, minimum and maximum based on IQR, etc.

10. Multiline Plot



There seems to be an overall positive trend for the e-commerce platform in terms of amount, profit and quantity engaged.

11. Product wise sales

```
In [9]: lst.groupby("Category").sum()
```

Out[9]:

Category	Amount	Profit	Quantity
Clothing	139054.0	11163.0	3516
Electronics	165267.0	10494.0	1154
Furniture	127181.0	2298.0	945

```
In [12]: lst.groupby("Sub-Category").sum()
```

Out[12]:

Sub-Category	Amount	Profit	Quantity
Accessories	21728.0	3559.0	262
Bookcases	56861.0	4888.0	297
Chairs	34222.0	577.0	277
Electronic Games	39168.0	-1236.0	297
Furnishings	13484.0	844.0	310
Hankerchief	14608.0	2098.0	754
Kurti	3361.0	181.0	164
Leggings	2106.0	260.0	186
Phones	46119.0	2207.0	304
Printers	58252.0	5964.0	291
Saree	53511.0	352.0	782
Shirt	7555.0	1131.0	271
Skirt	1946.0	235.0	248
Stole	18546.0	2559.0	671
T-shirt	7382.0	1500.0	305
Tables	22614.0	-4011.0	61
Trousers	30039.0	2847.0	135

12. Region wise sales:

```
In [10]: lst.groupby("State").sum()
```

Out[10]:

State	Amount	Profit	Quantity
Andhra Pradesh	13256.0	-496.0	146
Bihar	12943.0	-321.0	206
Delhi	22531.0	2987.0	277
Goa	6705.0	370.0	145
Gujarat	21058.0	465.0	328
Haryana	8863.0	1325.0	111
Himachal Pradesh	8666.0	656.0	113
Jammu and Kashmir	10829.0	8.0	177
Karnataka	15058.0	645.0	180
Kerala	13459.0	1871.0	157
Madhya Pradesh	105140.0	5551.0	1360
Maharashtra	95348.0	6176.0	1056
Nagaland	11903.0	148.0	173
Punjab	16786.0	-609.0	216
Rajasthan	21149.0	1257.0	282
Sikkim	5276.0	401.0	93
Tamil Nadu	6087.0	-2216.0	91
Uttar Pradesh	22359.0	3237.0	288
West Bengal	14086.0	2500.0	216

```
In [11]: lst.groupby("City").sum()
```

Out[11]:

City	Amount	Profit	Quantity
Ahmedabad	14230.0	-880.0	235
Allahabad	16857.0	3081.0	138
Amritsar	4507.0	544.0	52
Bangalore	15058.0	645.0	180
Bhopal	23583.0	871.0	247
Chandigarh	21142.0	172.0	275
Chennai	6087.0	-2216.0	91
Delhi	25019.0	3508.0	306
Gangtok	5276.0	401.0	93
Goa	6705.0	370.0	145
Hyderabad	13256.0	-496.0	146
Indore	79069.0	4159.0	1084
Jaipur	10076.0	-753.0	167
Kashmir	10829.0	8.0	177
Kohima	11903.0	148.0	173
Kolkata	14086.0	2500.0	216
Lucknow	5502.0	156.0	150
Mumbai	61867.0	1637.0	727
Patna	12943.0	-321.0	206
Pune	33481.0	4539.0	329
Simla	8666.0	656.0	113
Surat	6828.0	1345.0	93
Thiruvananthapuram	13459.0	1871.0	157
Udaipur	11073.0	2010.0	115

These are the results for the required aggregate queries, using Pandas dataframe and function.
