Bharatiya Vidya Bhavan's
Sardar Patel Institute of Technology
(Autonomous Institute Affiliated to University of Mumbai)
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058-India

Name:	**Vaishnavi Borkar**
UID:	**2021300016**
Batch:	**COMPS A (Batch G)**
Exp No. :	**1**

Aim:
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 or any other reports
- Write observations from each chart

Dataset Description:

This is a fictional dataset created for helping the data analysts to practice exploratory data analysis and data visualization. The dataset has data on orders placed by customers on a grocery delivery application.

The dataset is designed with an assumption that the orders are placed by customers living in the state of Tamil Nadu, India.

This dataset consists of various details related to customer orders, including sales transactions and product information. The **Order ID** uniquely identifies each transaction. **Customer_Name** provides the name of the customer who placed the order, while **Category** and **Sub_Category** classify the products, indicating broader product groups and more specific sub-groups respectively. The **City** and **State** columns specify the customer's location, while **Region** refers to the geographical area of the market. **Order_Date** records the date the order was placed, and **Sales** represents the monetary value of each transaction. The **Discount** column shows any price reduction applied, and **Profit** reflects the revenue generated from each order after deducting costs.

Dataset Link:

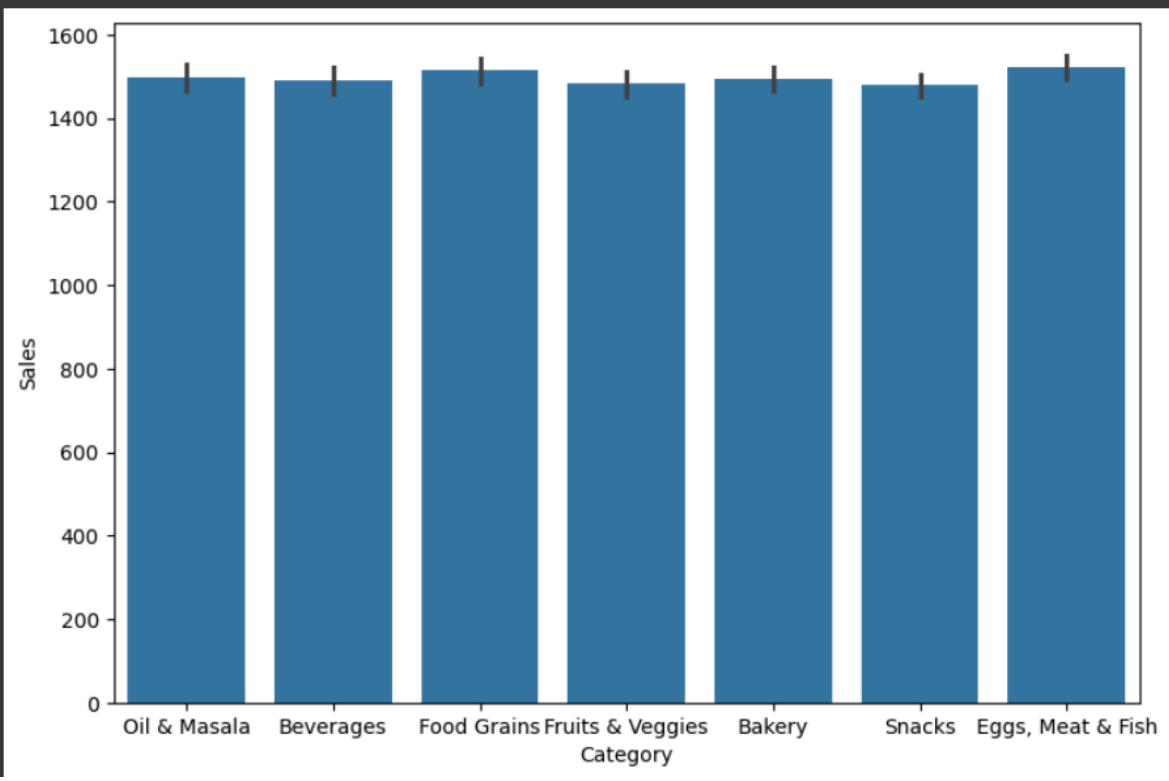
https://www.kaggle.com/datasets/mohamedharris/supermart-grocery-sales-retail-analytics-dataset

Plots and Inference:

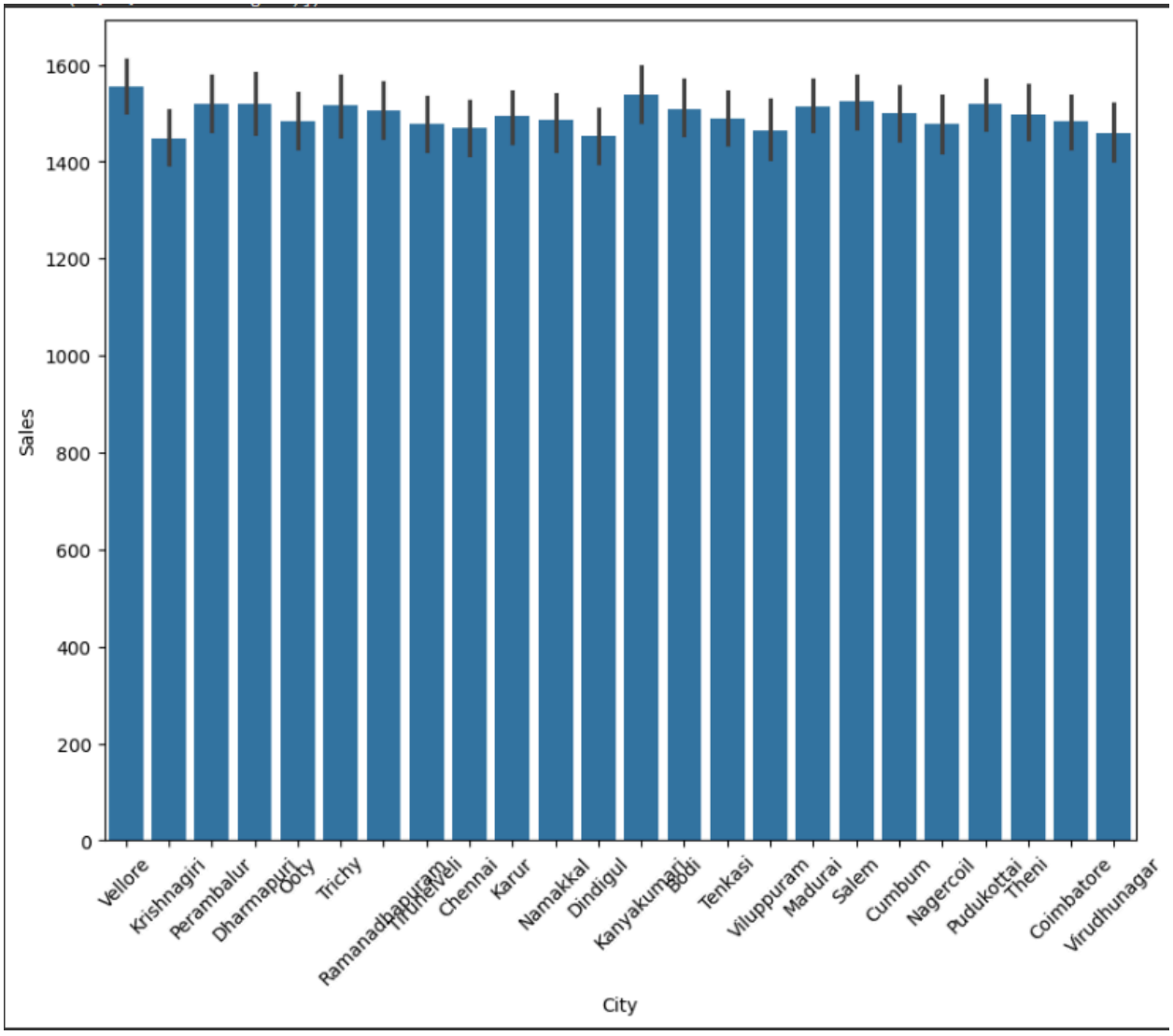
1. **Bar chart:**

This shows product/category wise sales and Grain and Meat-Fish category items were bought frequently in TamilNadu.

```
plt.figure(figsize=(9, 6))  # Adjust the width and height as needed
sns.barplot(data, x='Category', y='Sales')
plt.show()
#inference = Grain and Meat-Fish category items were bought frequently in TamilNadu
```

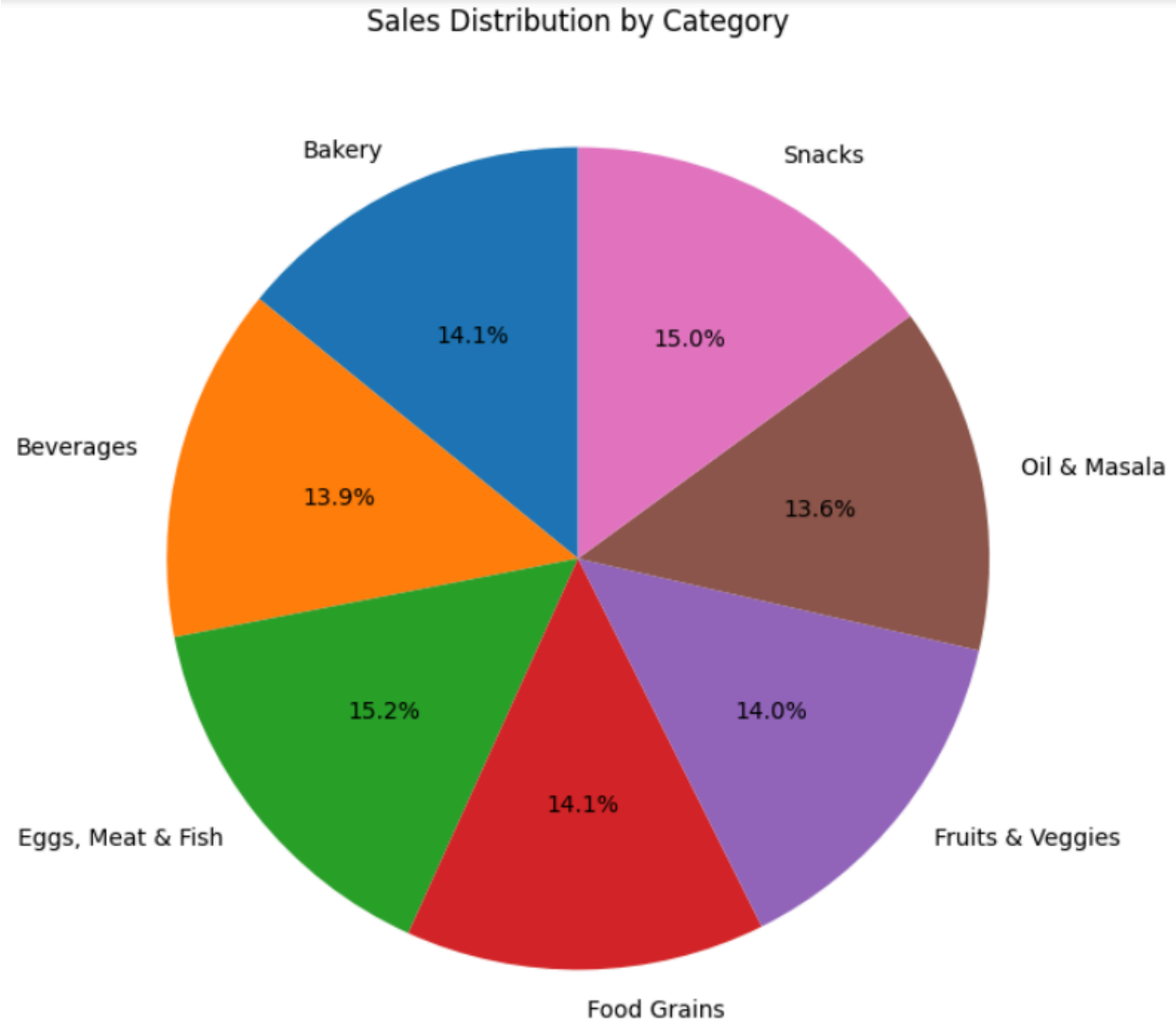


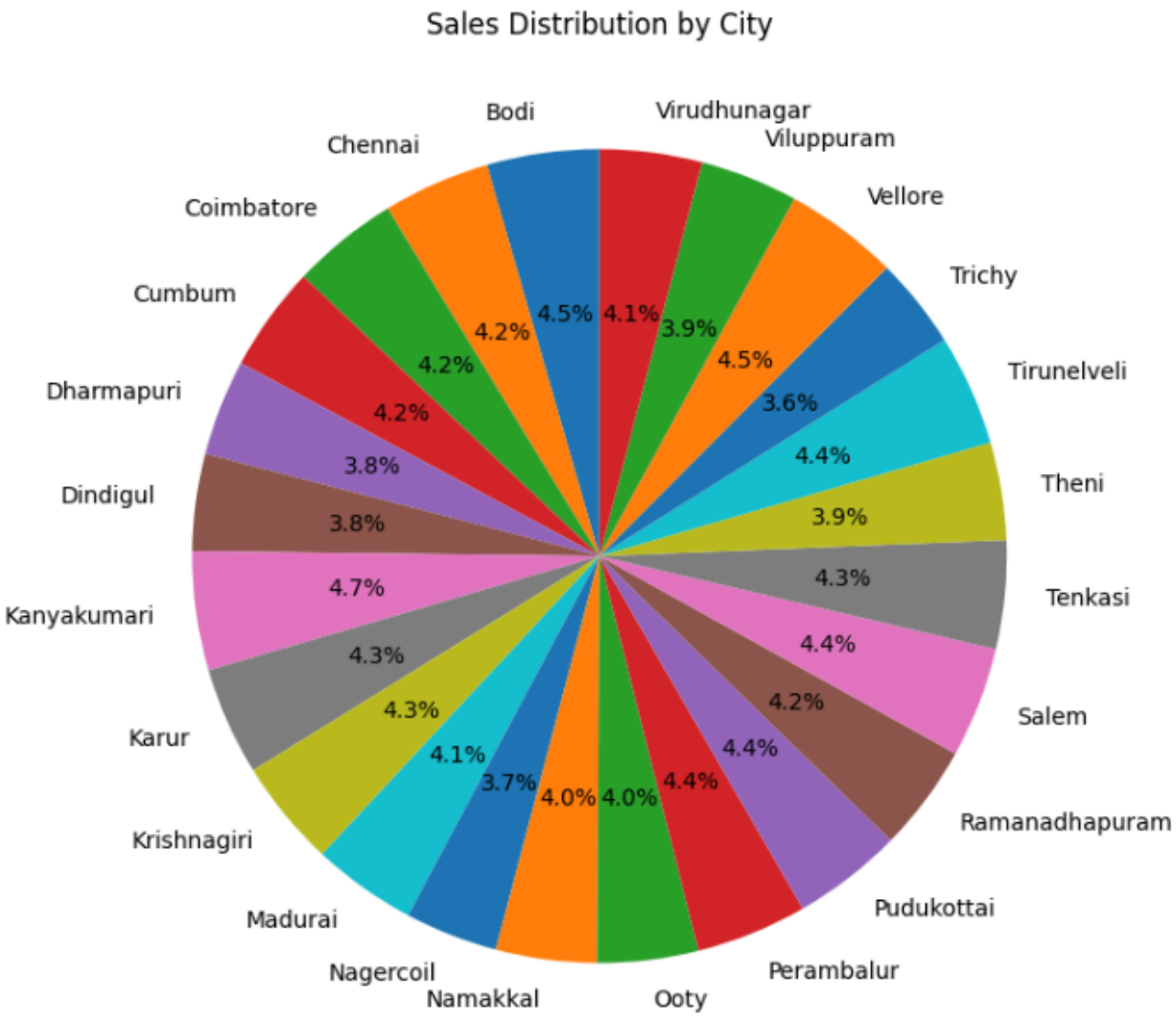
This gives me city wise sales and from the bar graph we can see that Vellore has the highest sales.



2. **Pie chart:**

Snacks and Eggs, Meat and Fish Category generated more revenue.

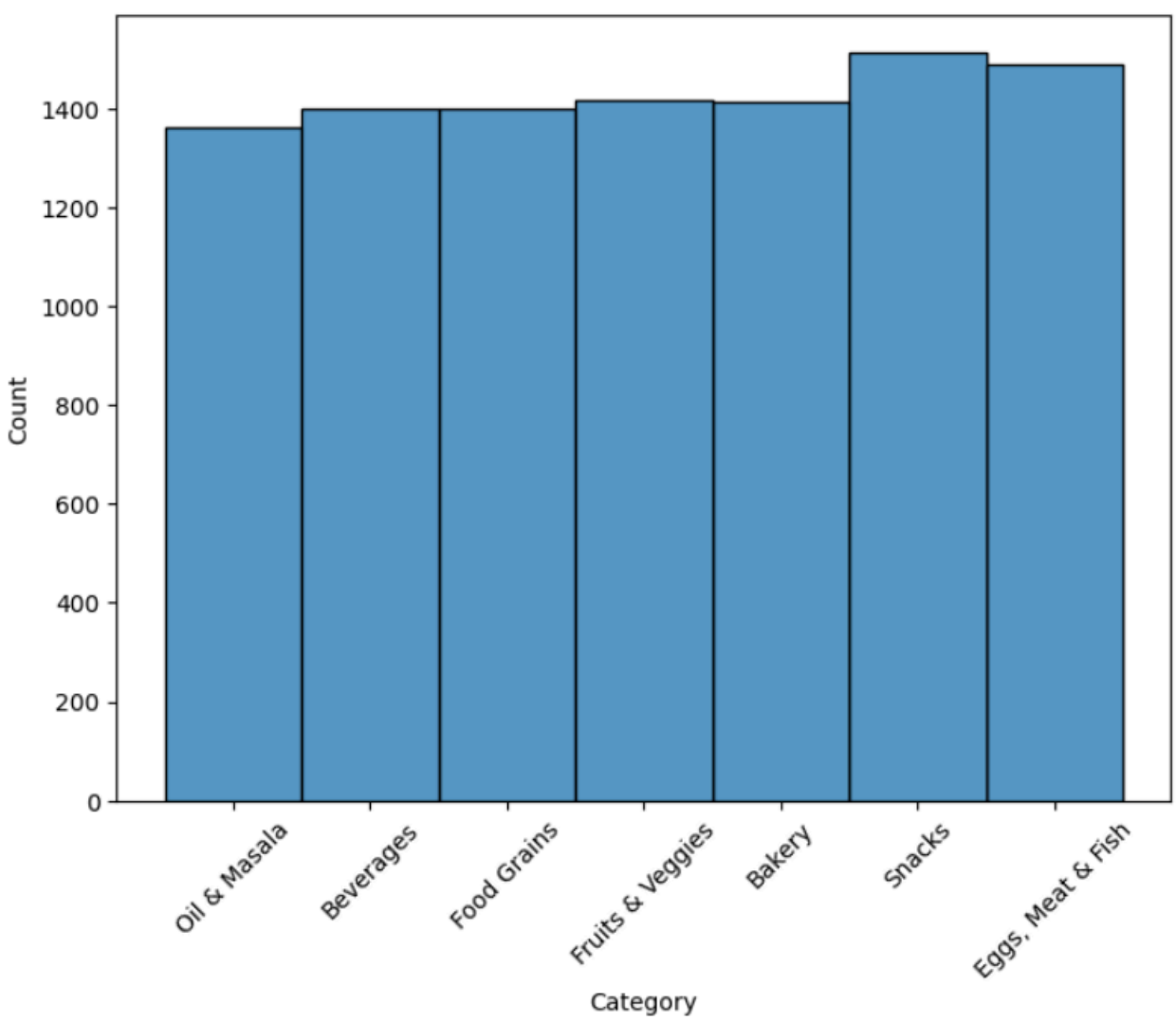




3. Histogram:

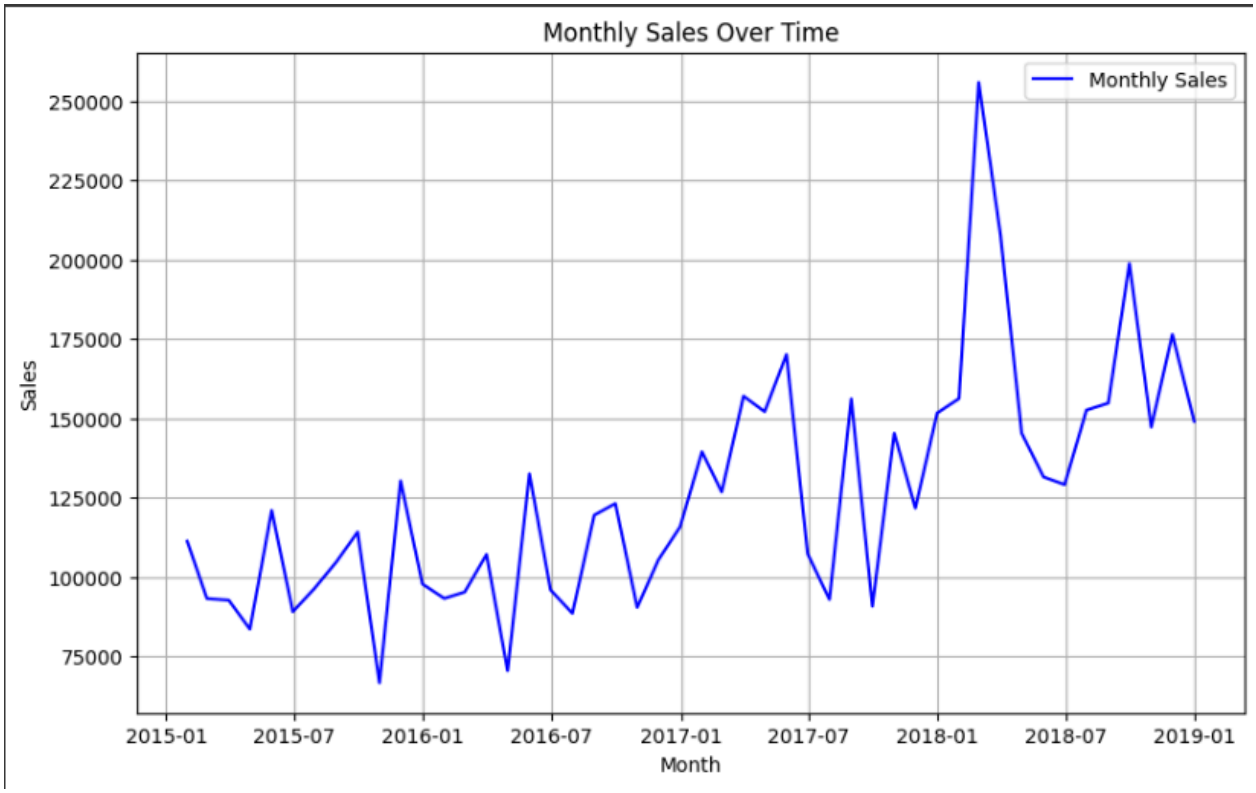
```
[7]  plt.figure(figsize=(8, 6))
     sns.histplot(data ,x = "Category",bins=5)
     plt.xticks(rotation=45)
```

```
([0, 1, 2, 3, 4, 5, 6],
 [Text(0, 0, 'Oil & Masala'),
  Text(1, 0, 'Beverages'),
  Text(2, 0, 'Food Grains'),
  Text(3, 0, 'Fruits & Veggies'),
  Text(4, 0, 'Bakery'),
  Text(5, 0, 'Snacks'),
  Text(6, 0, 'Eggs, Meat & Fish')])
```



There is not a drastic difference in the popularity of different categories, but Snacks and Eggs, Meat & Fish are slightly more favored by customers.

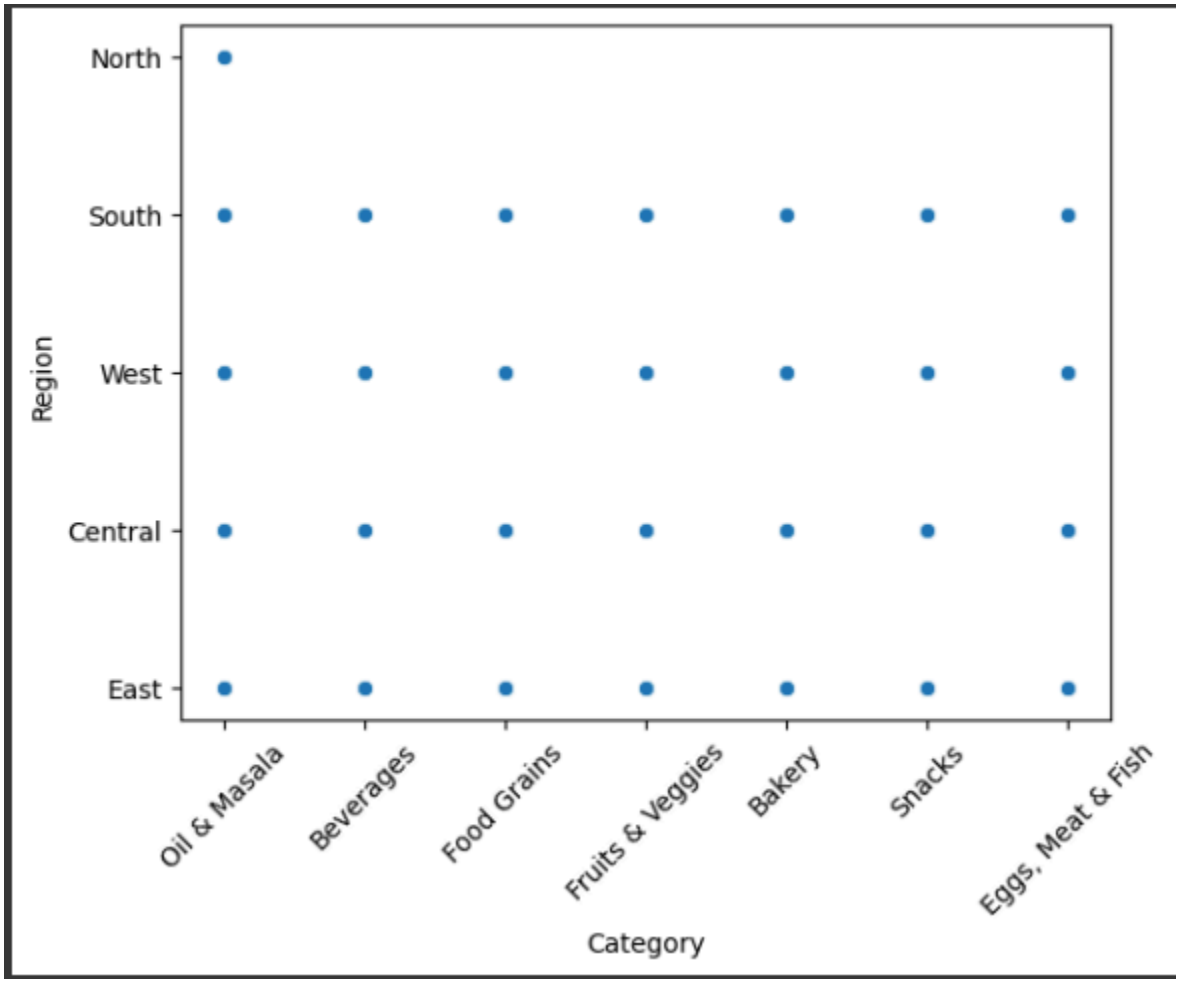
4. **Timeline chart:**



1. General Trend: There is an upward trend in sales over time, with noticeable fluctuations.
2. Sharp Increase in 2018: Around early 2018, there is a sharp spike in sales, peaking at over 250,000 units. This could be due to a seasonal or promotional event that significantly boosted sales.
3. Seasonal Dips: The chart shows several dips, particularly noticeable at various points in 2016 and 2018, indicating periods of reduced sales, possibly due to off-season trends or fewer promotions during those times.
4. Post-2018 Fluctuations: After the peak in 2018, sales show sharp rises and falls, indicating high variability and potentially unstable market conditions in this period.

Overall, despite fluctuations, the sales are on an upward trajectory with major spikes around 2018.

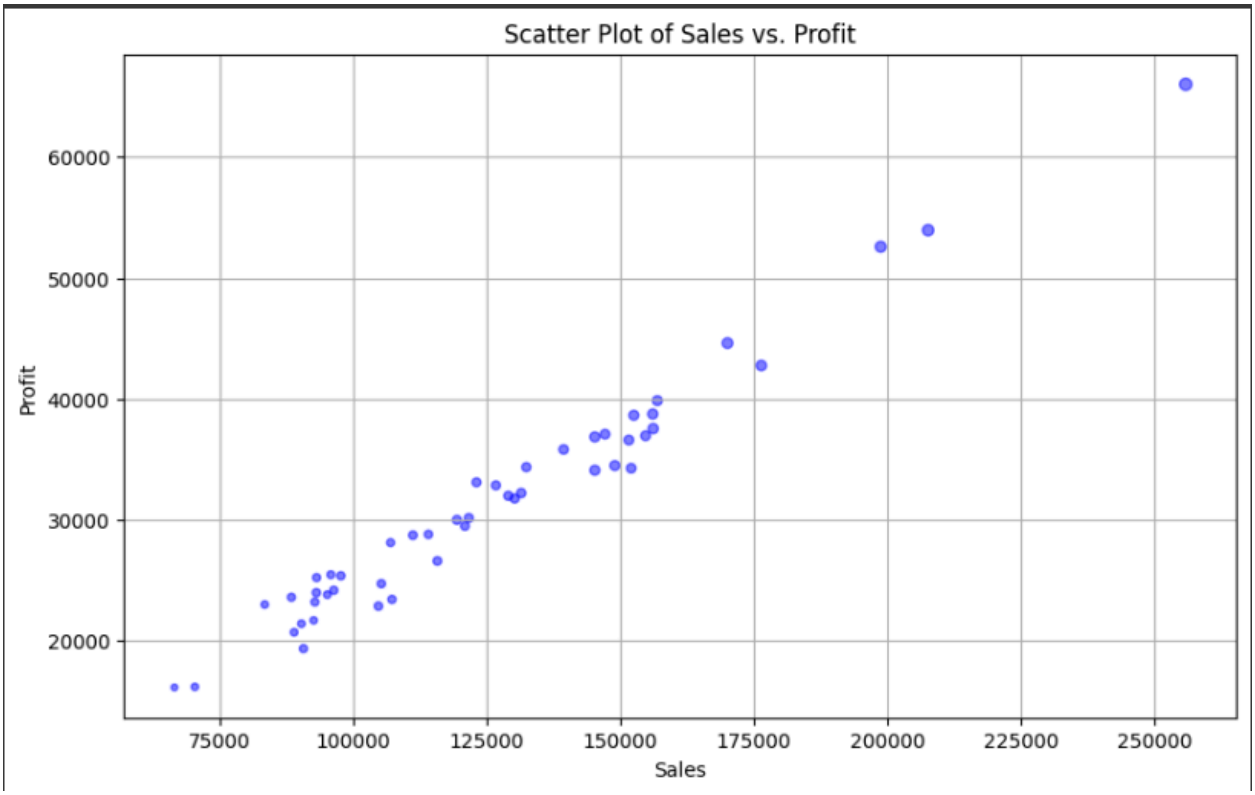
5. **Scatter plot:**



This plot shows that the Oil and Masala Category is only famous or has the highest sales only in the North India region of Tamil Nadu.

All other category products are available in all the regions of Tamil Nadu.

6. **Bubble plot:**



1. Positive Correlation: There is a clear positive linear relationship between Sales and Profit. As sales increase, profits also increase. This suggests that higher sales tend to generate higher profits.
2. Clustered Data Points: Most of the data points are clustered in the 75,000 to 150,000 range for Sales and 20,000 to 40,000 range for Profit. This indicates that the majority of sales and profit data falls within these ranges.
3. Outlier: There seems to be an outlier where Sales are around 250,000, resulting in a Profit above 60,000. This might indicate a particularly successful instance where a high level of sales led to disproportionately high profits.
4. Gradual Increase: The points gradually spread out as sales increase, indicating that profit grows at a relatively steady pace in relation to sales.
5. No Saturation Point Observed: There is no indication of diminishing returns or a saturation point, meaning that, at least in this dataset, profit keeps increasing with sales without flattening out.