Visual Analytics of Product Sales Across Retail Outlets – A Case Study on BigMart

Background:

BigMart operates multiple retail outlets across several cities and wants to optimize its product mix and pricing strategy. The dataset includes item-level sales, product types, MRP (Maximum Retail Price), and outlet characteristics like type, size, and establishment year. Your objective is to apply **visual analytics techniques in R** to uncover hidden trends, pricing patterns, and outlet performance insights.

Tasks to Be Completed

1. Data Preparation

- Load the dataset using read.csv().
- Clean missing values (e.g., Item_Weight, Outlet_Size).
- Convert categorical columns to factors if needed.
- Create new columns: Years_Operational from Outlet_Establishment_Year.

2. Exploratory Visual Analytics

A. Univariate Analysis

- **Bar Plot**: Count of items by Item_Type.
- **Box Plot**: Distribution of Item_MRP.
- **Pie/Donut Chart**: Share of Outlet_Type.

B. Bivariate & Multivariate Analysis

- **Box Plot**: Item_Outlet_Sales by Outlet_Size.
- **Scatter Plot**: Item_MRP vs Item_Outlet_Sales, color by Item_Type.
- **Facet Plot**: Sales distribution by Item_Type across Outlet_Location_Type.

3. Interactive Visual Analytics

- Use **plotly** to allow interactive exploration of:
    - Sales vs MRP by product type.
    - Outlet performance comparison (filter by location).
- Use **DT** for sortable/filterable tables showing top 10 items per outlet.

VIVA QUESTIONS

Section 1: Conceptual

1. What is visual analytics and how does it add value beyond static charts?
2. How did you ensure your visuals followed the principles of clarity and minimal clutter?

Section 2: Dataset-Specific

3. Which outlet type recorded the highest sales and how did you identify it visually?
4. What does your plot reveal about the relationship between MRP and item sales?
5. Did you observe any pattern in sales across different outlet sizes or locations?

Section 3: R and Package Usage

6. How did you use ggplot2 to handle grouped comparisons?
7. What advantages did plotly bring to your visual analytics?
8. How did you handle missing values in Item_Weight and Outlet_Size?

Section 4: Interpretation

9. Based on your plots, what kind of items should BigMart promote more?
10. Were there any outlets that underperformed consistently? How did you visualize this?
11. What filters did you implement in your dashboard and why?

Section 5: Advanced Insight

12. How could clustering or segmentation be integrated into your visual analytics?
13. If item visibility is too high or too low, how might that affect your interpretation of sales?