

Visualization



- Visualization is crucial to summarize, see distribution & relation in data
- Multiple type of graph can be generated
- Each visualization has its own purpose

Visualization

- Bar Chart → Compare Values Across Categories
- Line Chart → Trends over the time
- Pie Chart → Proportion / Percentage distribution of whole
- Scatter Plot → Show relation between numeric variables
- Box Plot → Summarize distribution and detect outlier
- Heatmap → Visualize data density

Question

What is proper visualization for:

1. Total Sales for each product category
2. Trend of Product A total sales over the time
3. Contribution of each Product Category on total Revenue
4. Showing relation between Sales and Revenue
5. Sales intensity for each product
6. Distribution of customer satisfaction

Bar Chart

```
grouped_data = df.groupby("Product")["Sales"].sum()
grouped_data.plot(kind="bar", title="Total Sales by Product")
plt.ylabel("Total Sales")
plt.xticks(rotation=0)
plt.show()
```

Line Chart

```
df_PA = df[df.Product == "A"]  
df_PA_trend = df_PA.groupby("Month")["Sales"].sum()  
df_PA_trend.plot(kind= "line", x="Month", y="Sales")  
plt.show()
```

or without groupby set_index as Month and plot of sales

or df.plot(x="Month", y="Sales", kind="line")

Q. Is it the proper trend as month isn't ordered. (convert to category)

Q. What if we want to show trend of multiple product? (label='')

Pie Chart

```
df_grouped = df.groupby("Product")["Revenue"].sum()  
df_grouped.plot(kind= "pie", x="Month", y="Sales")  
plt.show()
```

Scatter Plot

```
df.plot(kind="scatter", x="Revenue", y="Sales")  
plt.show()
```

Box Plot

```
df.plot(kind='box', column='CustomerSatisfaction', by='Product')  
plt.show()
```

Heat Map

- No direct support
- We will cover in Seaborn

Histogram

```
df.plot(kind="hist", y="CustomerSatisfaction", bins=6)  
plt.show()
```



Exercise