



**BHARATIYA VIDYA BHAVAN'S**  
**SARDAR PATEL INSTITUTE OF TECHNOLOGY**  
(Empowered Autonomous Institute Affiliated to University of Mumbai)  
[Knowledge is Nectar]  
**Department of Computer Engineering**

Name: Shivam Patil  
UID: 2021300097  
Batch: G  
Exp: 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
- Write observations from each chart

**Theory:**

**Dataset:** <https://www.kaggle.com/datasets/rohitsahoo/sales-forecasting>

**Dataset Description:**

The dataset represents transactional sales data for a retail Superstore. It includes detailed information on sales transactions, customer demographics, and logistical details like shipping dates and methods. The data covers various products, categorized by both broad categories and more specific sub-categories, providing insights into sales performance, customer segmentation, and distribution channels. The dataset can be used for a variety of analyses, including sales trends, customer behavior, and shipping logistics.

This dataset includes the following variables:

1. **Category** - The broad category under which each product falls (e.g., Furniture, Office Supplies, Technology).
2. **City** - The city where the customer is located.
3. **Country** - The country where the customer is located.
4. **Customer ID** - A unique identifier for each customer.
5. **Customer Name** - The name of the customer.



**BHARATIYA VIDYA BHAVAN'S**  
**SARDAR PATEL INSTITUTE OF TECHNOLOGY**  
(Empowered Autonomous Institute Affiliated to University of Mumbai)  
[Knowledge is Nectar]

**Department of Computer Engineering**

6. **Order Date** - The date when the order was placed.
7. **Order ID** - A unique identifier for each order.
8. **Postal Code** - The postal code of the customer's address.
9. **Product ID** - A unique identifier for each product.
10. **Product Name** - The name of the product sold.
11. **Region** - The region where the customer is located (e.g., East, West, Central, South).
12. **Row ID** - A unique identifier for each row in the dataset.
13. **Sales** - The revenue generated from each sale.
14. **Segment** - The customer segment to which the customer belongs (e.g., Consumer, Corporate, Home Office).
15. **Ship Date** - The date when the order was shipped.
16. **Ship Mode** - The method used for shipping the order (e.g., Standard Class, Second Class, First Class, Same Day).
17. **State** - The state where the customer is located.
18. **Sub-Category** - A more specific classification of products within the broader Category (e.g., Chairs, Phones, Binders).

## 1. Bar Charts

**Theory:** Bar charts are designed to compare quantities across different categories. They use rectangular bars where the length of each bar represents the value of the category it corresponds to.

### Strengths:

- **Clarity:** Easily compare different categories.
- **Versatility:** Can be used for both categorical and numerical data.
- **Comparison:** Effective for comparing individual categories or showing changes over time with grouped bars.

### Limitations:

- **Overcrowding:** Can become cluttered if too many categories are displayed.
- **Precision:** Less effective for showing precise relationships or trends compared to other charts.

### Usage Examples:



**BHARATIYA VIDYA BHAVAN'S**  
**SARDAR PATEL INSTITUTE OF TECHNOLOGY**  
(Empowered Autonomous Institute Affiliated to University of Mumbai)  
[Knowledge is Nectar]

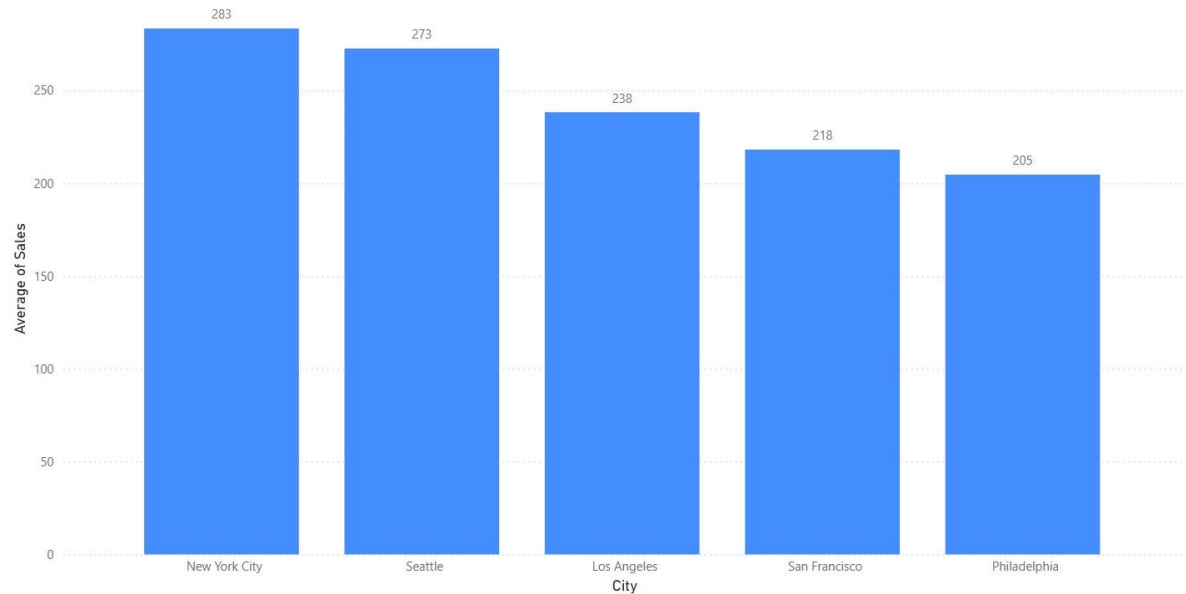
**Department of Computer Engineering**

- Comparing sales across different regions.
- Showing the number of students in various classes.

**Region-wise Sales:**

○ **Chart:**

Average of Sales by City



- **Observation:** The bar chart shows the average sales by city, with New York City leading at 283, followed closely by Seattle with an average of 273. Los Angeles is next with an average of 238, followed by San Francisco at 218. Philadelphia has the lowest average sales among the cities at 205. This indicates that while New York City and Seattle have the highest average sales, the difference between the cities is less pronounced compared to the total sales figures, suggesting relatively consistent sales performance across these cities.

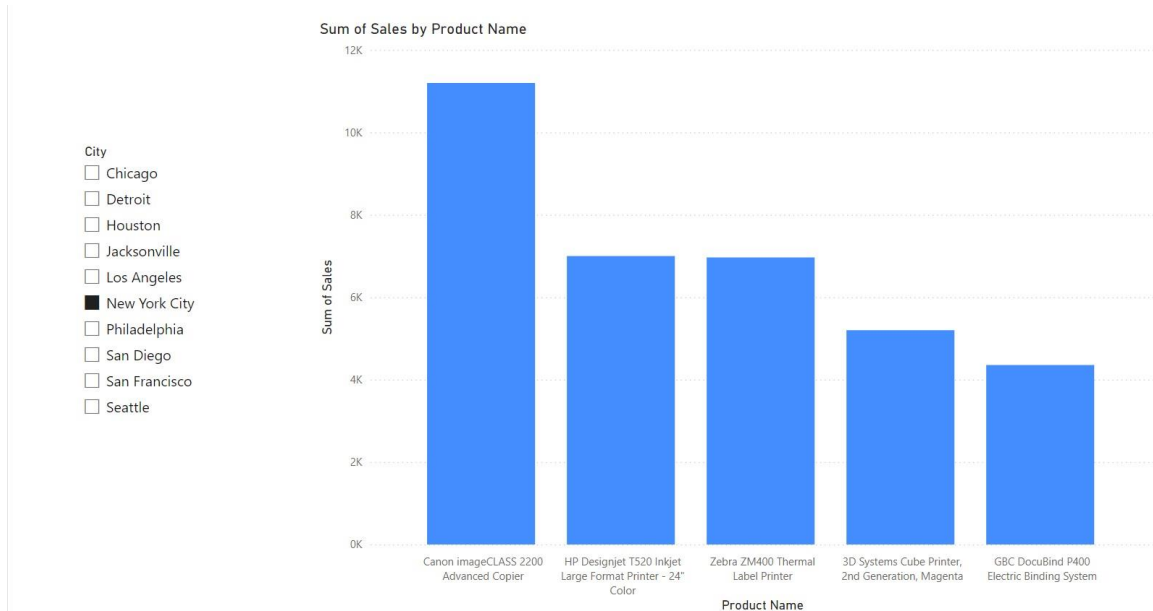
**Product-wise Sales:**

○ **Chart:**



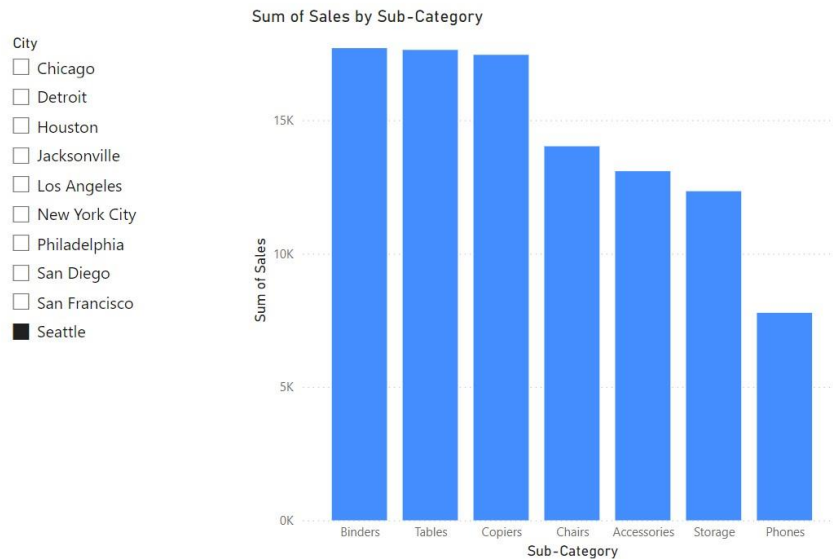
**BHARATIYA VIDYA BHAVAN'S**  
**SARDAR PATEL INSTITUTE OF TECHNOLOGY**  
(Empowered Autonomous Institute Affiliated to University of Mumbai)  
[Knowledge is Nectar]

**Department of Computer Engineering**



- **Observation:** As we saw earlier that New York has the highest average sales, we will now analyze the products which were sold the most. From this chart, we observe that Canon Copier was sold the most in New York, followed by HP Printers. Overall, we can see that in the top 3, there are printers, which shows that it may have occurred due to New York being a corporate hub, with some of the biggest companies in the world having their headquarters there.

**Chart:**



- **Observation:** We see, for Seattle which was ranked 2nd in average sales, the bar chart depicts the sum of sales by sub-category. Binders, Tables, and Copiers are the top-performing sub-categories, each with sales just above 15K. Chairs follow closely, while Accessories and Storage show moderately lower sales around the 10K mark. Phones have the lowest sales, significantly trailing the other sub-categories. This suggests that in Seattle, office supplies like Binders, Tables, and Copiers are the most in-demand, while Phones see comparatively less sales activity.

## 2. Pie Charts

**Theory:** Pie charts represent proportions of a whole. Each slice of the pie corresponds to a category's contribution to the total, providing a visual representation of percentage-based data.

### Strengths:

- **Simplicity:** Easy to understand at a glance.
- **Proportion:** Good for showing relative sizes of parts to a whole.

### Limitations:

- **Precision:** Difficult to compare sizes of slices accurately.



**BHARATIYA VIDYA BHAVAN'S**  
**SARDAR PATEL INSTITUTE OF TECHNOLOGY**  
(Empowered Autonomous Institute Affiliated to University of Mumbai)  
[Knowledge is Nectar]  
**Department of Computer Engineering**

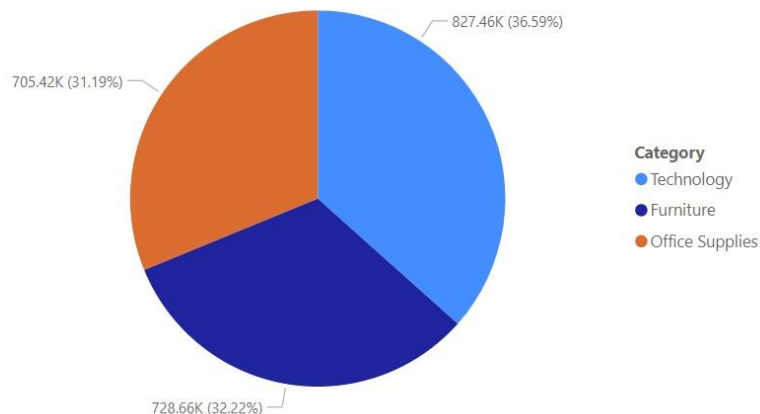
- **Limited Categories:** Best used with a small number of categories; too many slices can make the chart cluttered and hard to read.

#### Usage Examples:

- Showing market share percentages.
- Illustrating budget allocation.

#### Chart:

Sum of Sales by Category



- **Observation:** The pie chart shows the sum of sales by category, divided into three segments: Technology, Furniture, and Office Supplies. Technology leads with 36.59% of total sales, amounting to 827.46K. Furniture follows closely, contributing 32.22% with sales totaling 728.66K. Office Supplies make up the smallest portion, with 31.19% of sales, totaling 705.42K. The chart indicates a fairly balanced distribution of sales across the three categories, with Technology having a slight edge over the others.

### 3. Scatter Plots

**Theory:** Scatter plots display values for two variables, using dots to represent data points. They help visualize the relationship between these variables and identify patterns, correlations, or outliers.



**BHARATIYA VIDYA BHAVAN'S**  
**SARDAR PATEL INSTITUTE OF TECHNOLOGY**  
(Empowered Autonomous Institute Affiliated to University of Mumbai)  
[Knowledge is Nectar]  
**Department of Computer Engineering**

### Strengths:

- **Correlation:** Effective for identifying relationships and trends between variables.
- **Distribution:** Useful for observing the distribution of data points.

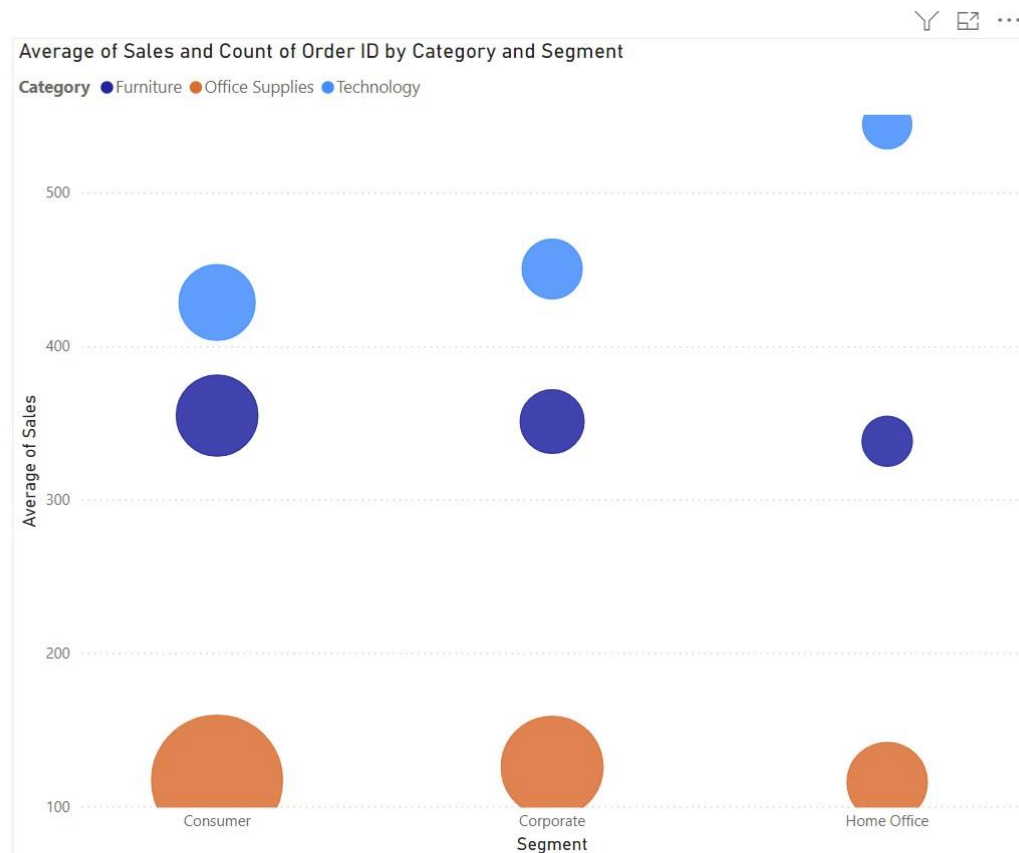
### Limitations:

- **Complexity:** Can become cluttered with large datasets.
- **Trends:** Does not explicitly show trends without additional analysis or fitting lines.

### Usage Examples:

- Analyzing the correlation between hours studied and exam scores.
- Plotting geographical locations and their attributes.

### Chart:





**BHARATIYA VIDYA BHAVAN'S**  
**SARDAR PATEL INSTITUTE OF TECHNOLOGY**  
(Empowered Autonomous Institute Affiliated to University of Mumbai)  
[Knowledge is Nectar]

**Department of Computer Engineering**

- **Observation:** The scatter plot illustrates the relationship between the average sales and the count of orders across different categories and segments. The categories include Furniture, Office Supplies, and Technology, represented by different shades of blue and orange. The size of each bubble indicates the count of orders, while the position on the vertical axis reflects the average sales.
  - Technology has the highest average sales, particularly in the Corporate and Consumer segments, with larger bubbles indicating a higher count of orders.
  - Furniture shows moderate average sales across all segments, with the Corporate segment having the largest bubble, signifying a higher order count.
  - Office Supplies have the lowest average sales, especially in the Home Office segment, which also has a smaller bubble, indicating fewer orders.Overall, Technology dominates in both average sales and order count, while Office Supplies lag behind, particularly in the Home Office segment.

## 4. Histograms

**Theory:** Histograms display the frequency distribution of numerical data by grouping values into bins or intervals. Each bar represents the count of data points within a particular range.

### Strengths:

- **Distribution:** Ideal for showing the distribution of a single variable.
- **Frequency:** Highlights patterns, such as skewness or modality.

### Limitations:

- **Range Selection:** The choice of bin width can significantly affect the appearance and interpretation of the data.
- **Detail Loss:** Can obscure details if bins are too broad.

### Usage Examples:

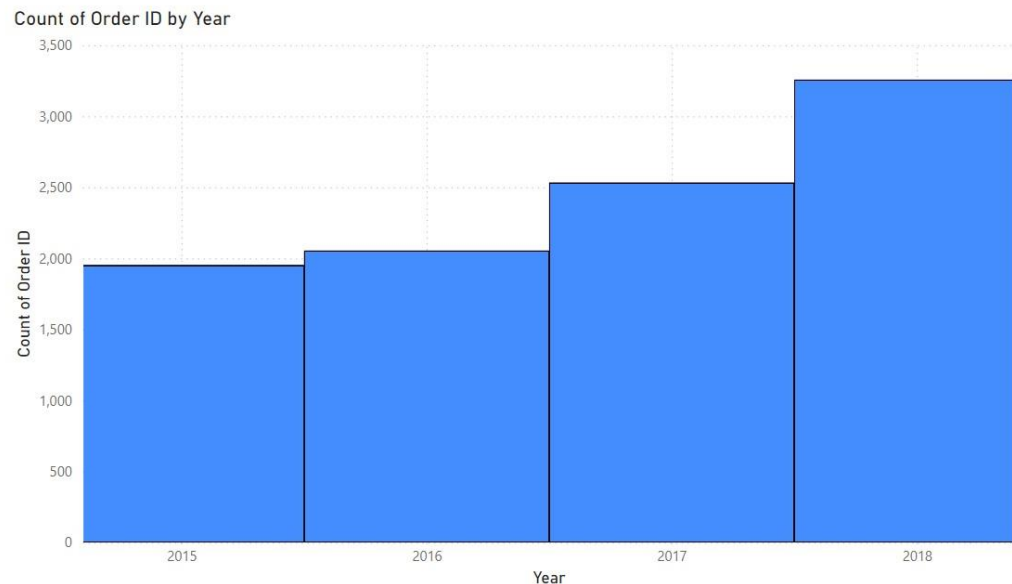
- Visualizing the distribution of ages in a population.
- Showing the frequency of test scores within certain ranges.

### Chart:





**BHARATIYA VIDYA BHAVAN'S**  
**SARDAR PATEL INSTITUTE OF TECHNOLOGY**  
(Empowered Autonomous Institute Affiliated to University of Mumbai)  
[Knowledge is Nectar]  
**Department of Computer Engineering**



- **Observation:** The histogram shows the count of Order IDs by year from 2015 to 2018. There is a clear upward trend in the number of orders each year. The count starts at around 2,000 in 2015 and shows steady growth, increasing slightly in 2016 and more noticeably in 2017. The most significant increase occurs in 2018, where the count approaches 3,500 orders. This indicates a consistent and accelerating growth in order volume over the four years.

## 5. Timeline Charts

**Theory:** Timeline charts visualize events or data points over time, arranging them chronologically. They help in understanding the sequence and duration of events or trends over a period.

### Strengths:

- **Temporal Analysis:** Excellent for showing changes over time and historical events.
- **Trend Identification:** Useful for tracking progress or historical trends.

### Limitations:



**BHARATIYA VIDYA BHAVAN'S**  
**SARDAR PATEL INSTITUTE OF TECHNOLOGY**  
(Empowered Autonomous Institute Affiliated to University of Mumbai)  
[Knowledge is Nectar]

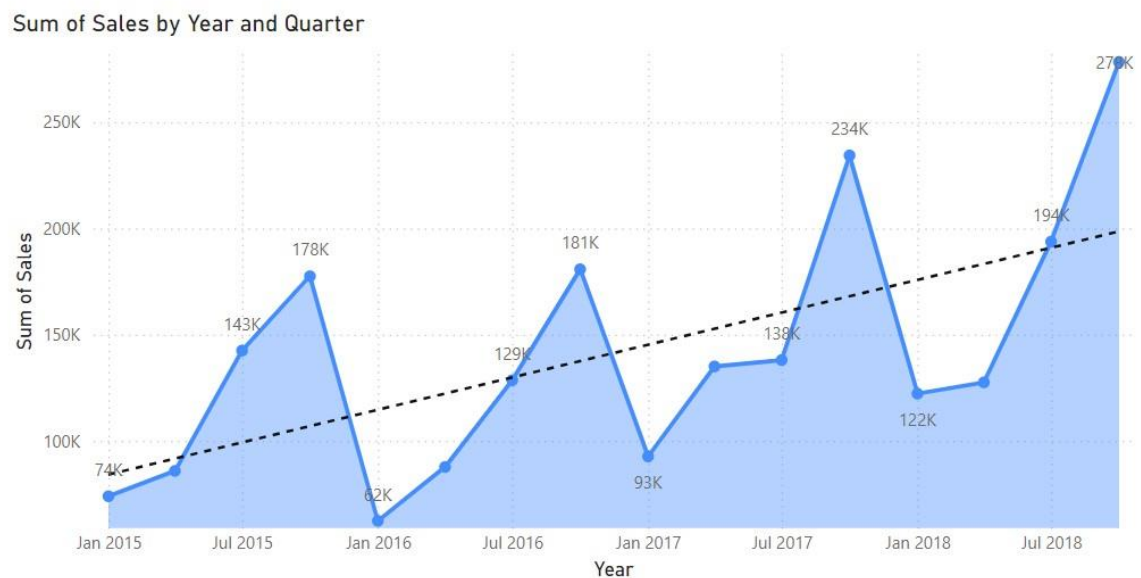
**Department of Computer Engineering**

- **Complexity:** Can become cluttered with too many events or a long time span.
- **Detail Limitation:** May not show detailed information on each event unless designed carefully.

**Usage Examples:**

- Project management timelines.
- Historical event sequencing.

**Chart:**



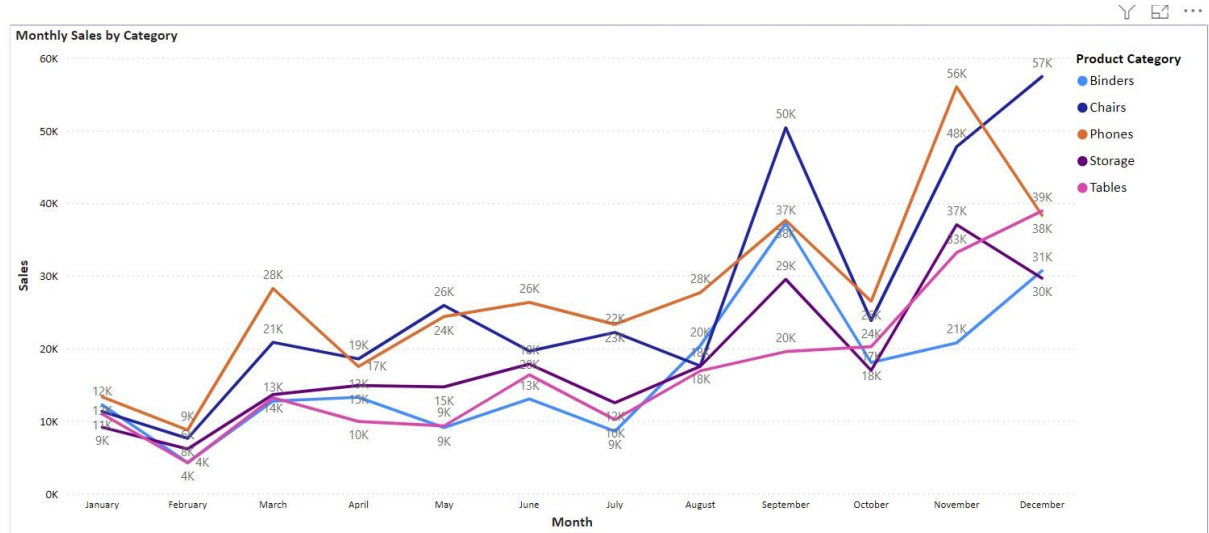
- **Observation:** The graph shows the trend in sales over a period from January 2015 to July 2018, with noticeable fluctuations across quarters. While the overall trend, as indicated by the dotted black line, shows a steady increase in sales, there are significant peaks and troughs within each year. The most notable spikes in sales occur around mid-2016, mid-2017, and mid-2018, with the highest sales recorded in July 2018 at 278K. Conversely, the lowest sales dip occurs around January 2016 at 62K. Despite these fluctuations, the general upward trend suggests growth in sales over the period.



**BHARATIYA VIDYA BHAVAN'S**  
**SARDAR PATEL INSTITUTE OF TECHNOLOGY**  
(Empowered Autonomous Institute Affiliated to University of Mumbai)  
[Knowledge is Nectar]

**Department of Computer Engineering**

**Chart:**



- **Observation:** The line chart illustrates monthly sales trends for five product categories: Binders, Chairs, Phones, Storage, and Tables throughout the year. Phones consistently show the highest sales, particularly peaking in September, November, and December at 56K, 48K, and 57K, respectively. Chairs also demonstrate strong sales, with significant spikes in September and December at 50K and 48K. Binders and Tables maintain moderate sales, with Binders peaking in December at 39K, while Storage generally lags behind, with lower sales throughout the year, peaking at 38K in November. The chart highlights significant sales growth towards the end of the year across most categories.

### Conclusion:

The experiment conducted in Power BI served as an introductory exploration of the tool's capabilities, focusing on creating basic charts and visualizations using the Superstore dataset. Through this process, I familiarized myself with essential features like drag-and-drop field placements, axis settings, and the creation of simple visual comparisons, such as sales trends over time and category-wise breakdowns. This exercise provided a foundational understanding of how to navigate Power BI, manipulate data, and present insights visually, laying the groundwork for more advanced analysis in the future.