



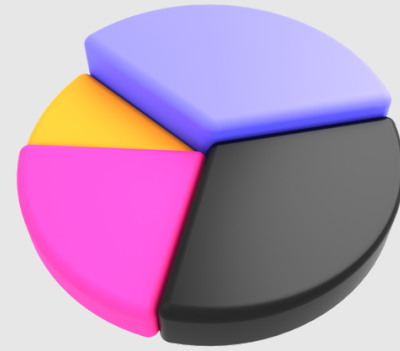
# DATA VISUALIZATION

REZA SHOKRZAD

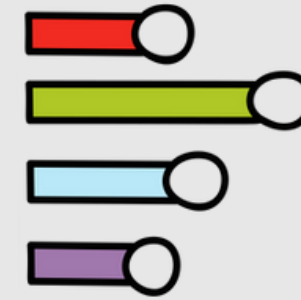


# MAIN DIAGRAMS

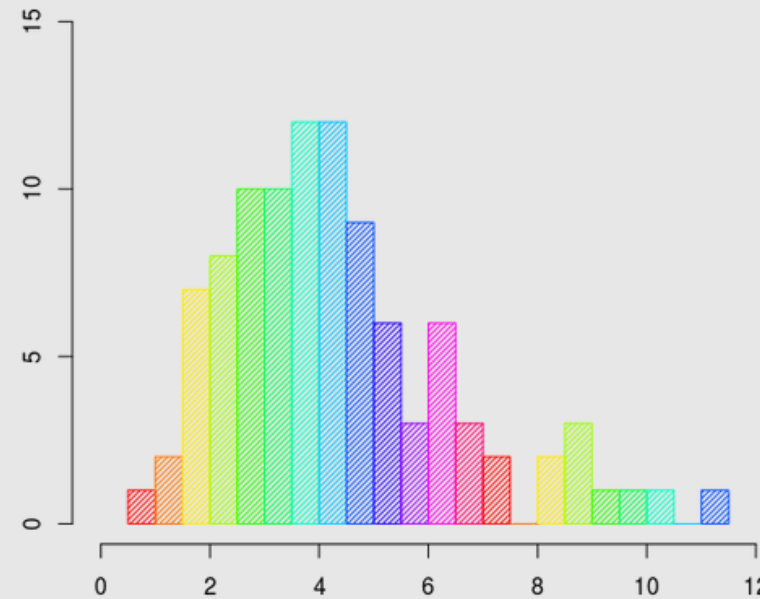
- Pie Chart



- Bar Chart



- Histogram



- Line Chart



- Scatter Plot



- Box Plot

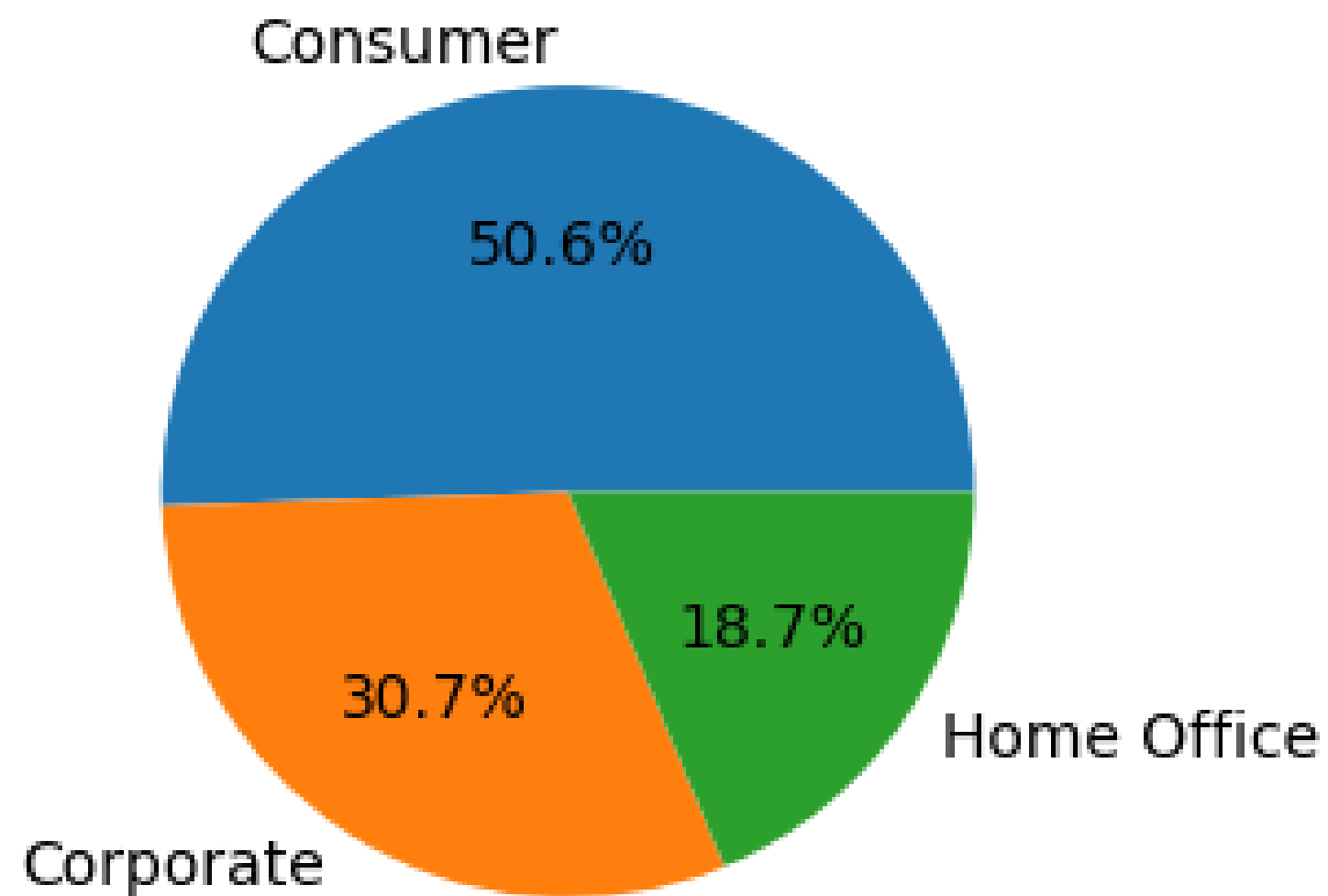


# PIE CHART

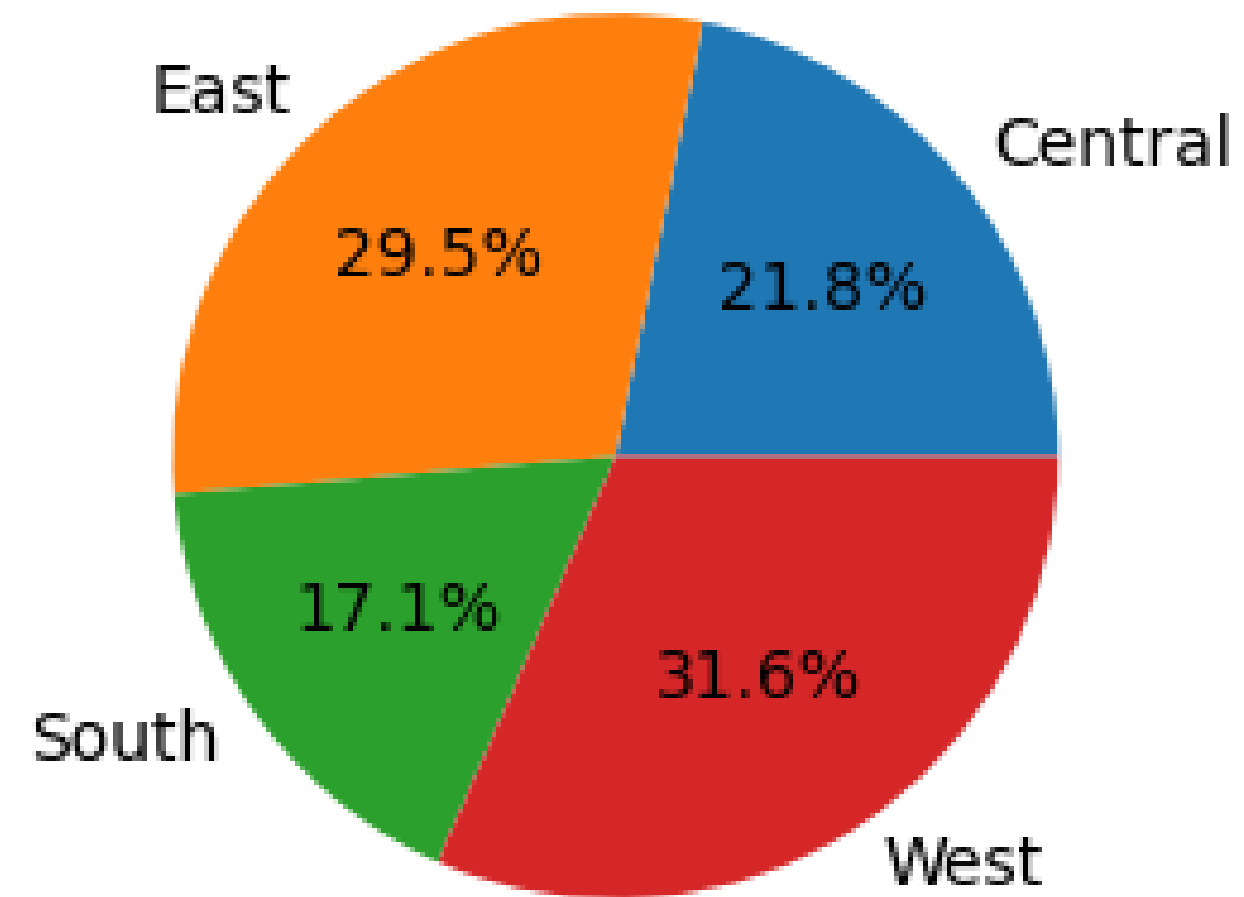
- **Function:** Ideal for showing proportions of the whole.
- **Variables:** Uses one categorical variable and one numerical variable.
- **Structure:** Each slice corresponds to a category; size of the slice represents its proportion relative to total.
- **Comparison:** Best for comparing parts of the whole (e.g., proportion of total sales per product).
- **Category Limit:** Most effective when the number of categories is small.

# PIE CHART

Sales Distribution by Segment



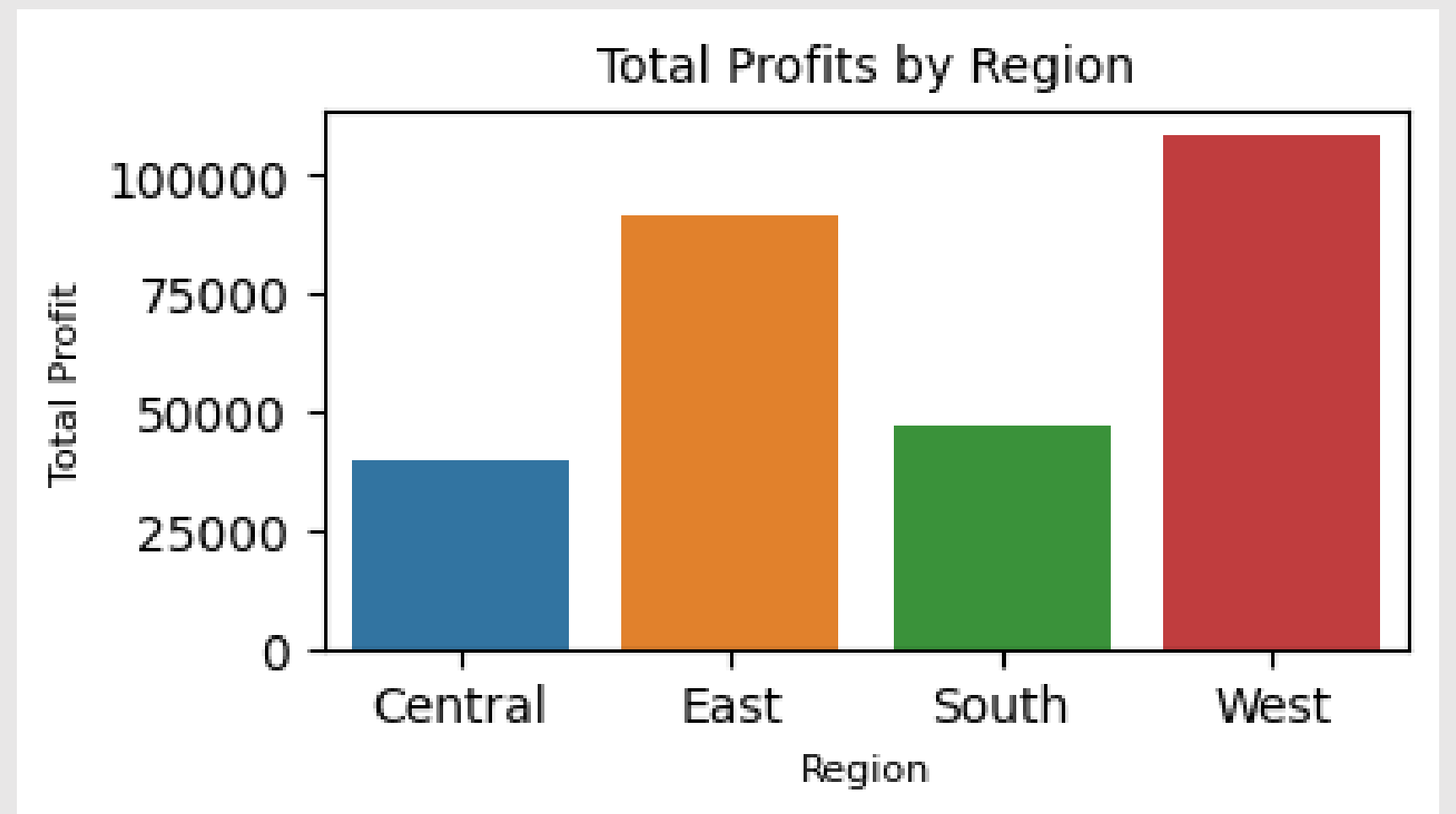
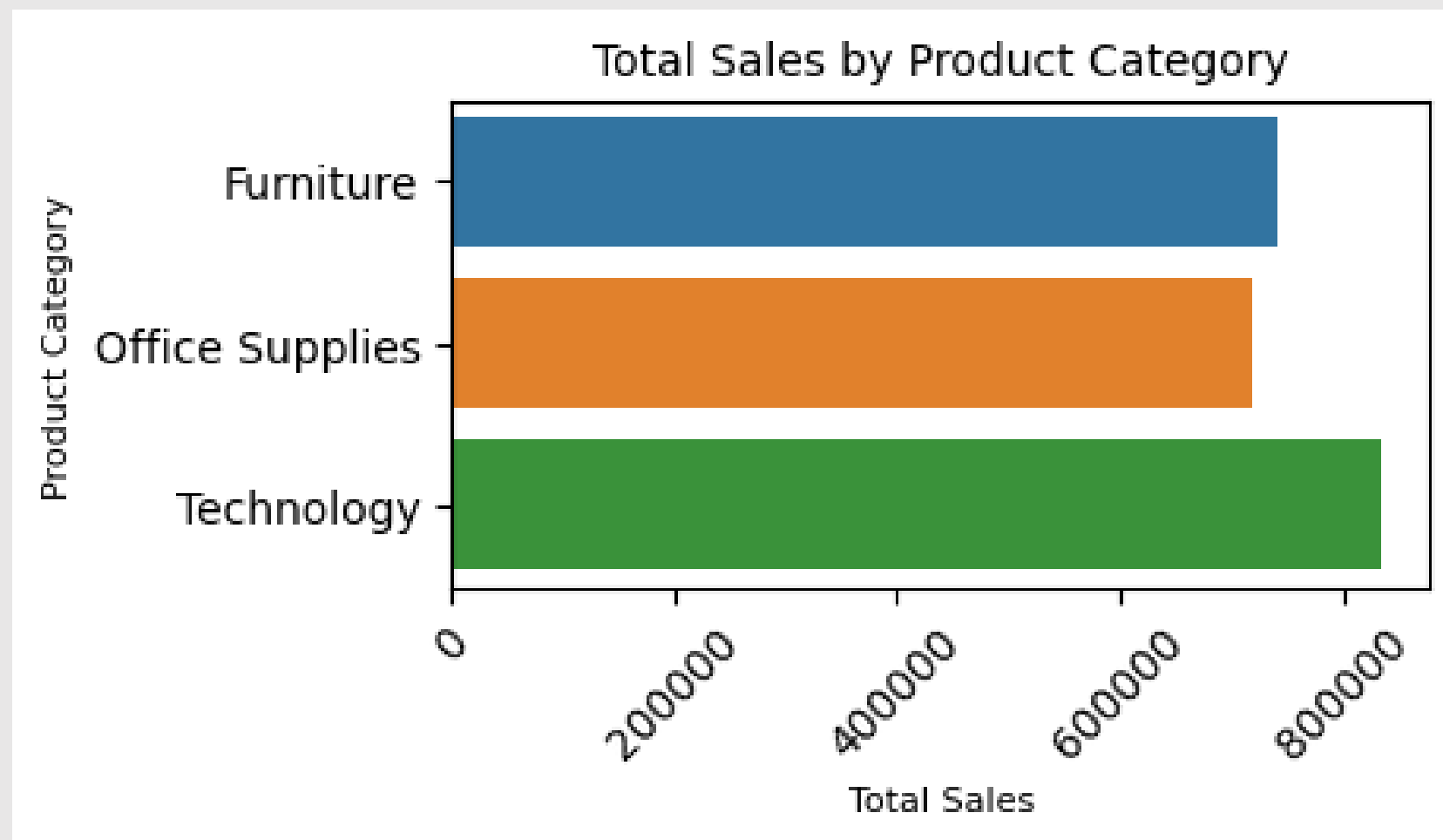
Sales Distribution by Region



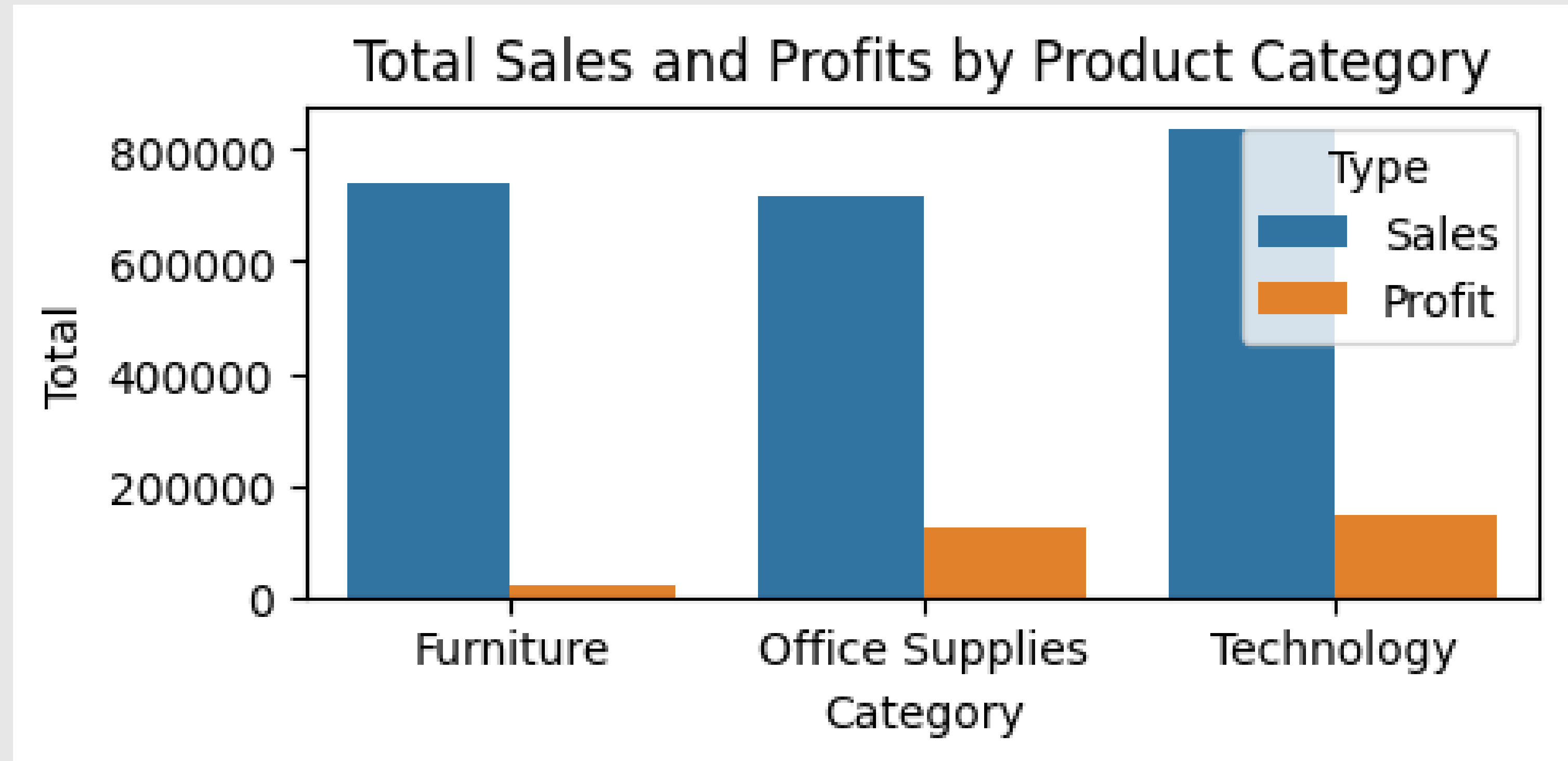
# BAR CHART

- **Function:** Ideal for comparing categories of data.
- **Axes:** X-axis represents categories; Y-axis represents numerical values.
- **Variables:** Can use one or more categorical variables.
- **Comparison:** Good for comparing numerical variables across categories (e.g., sales per product).
- **Time-Series:** Useful for showing changes over time for different categories (if X-axis is time).

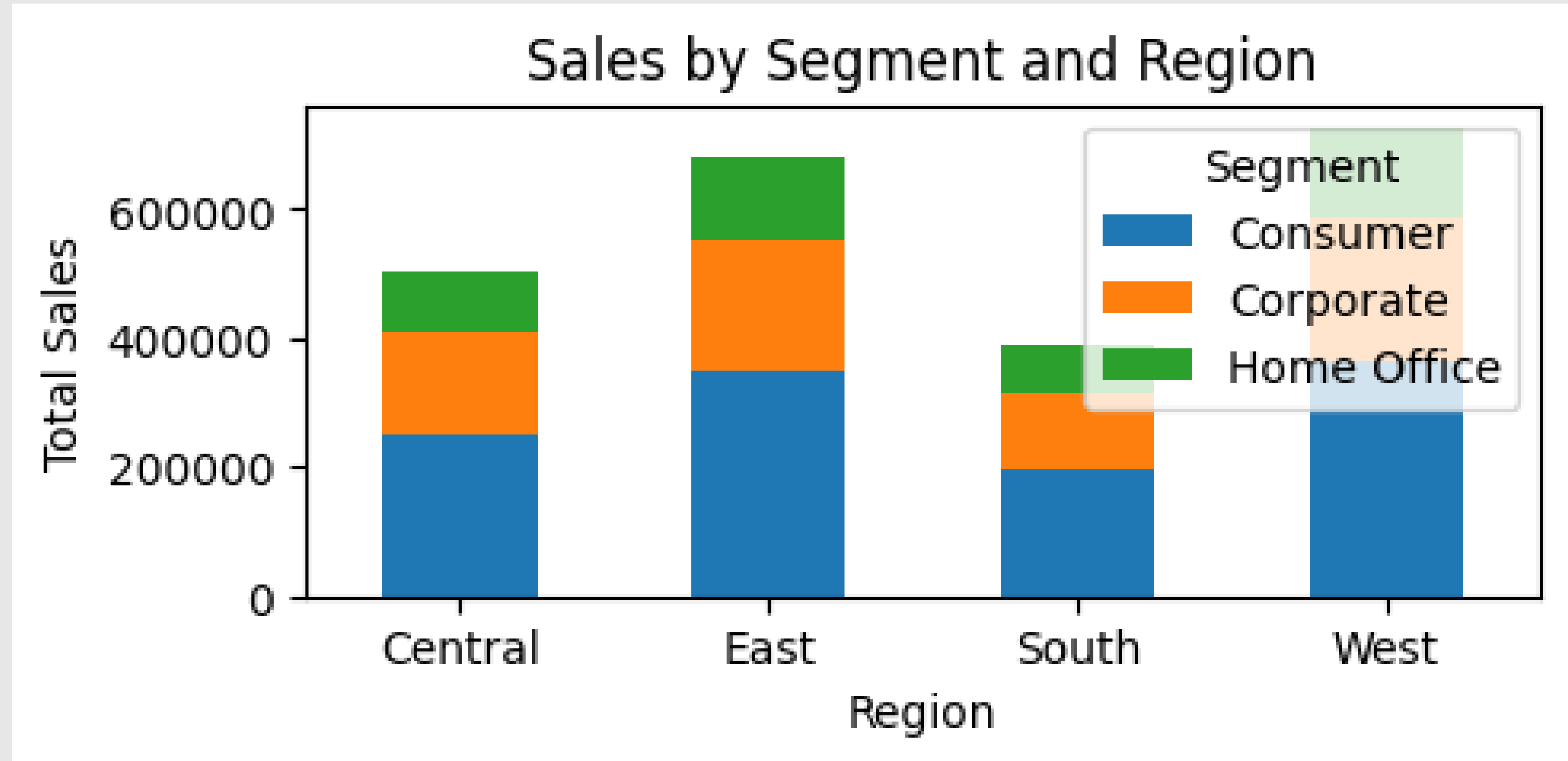
# BAR CHART



# CLUSTERED BAR



# STACKED BAR



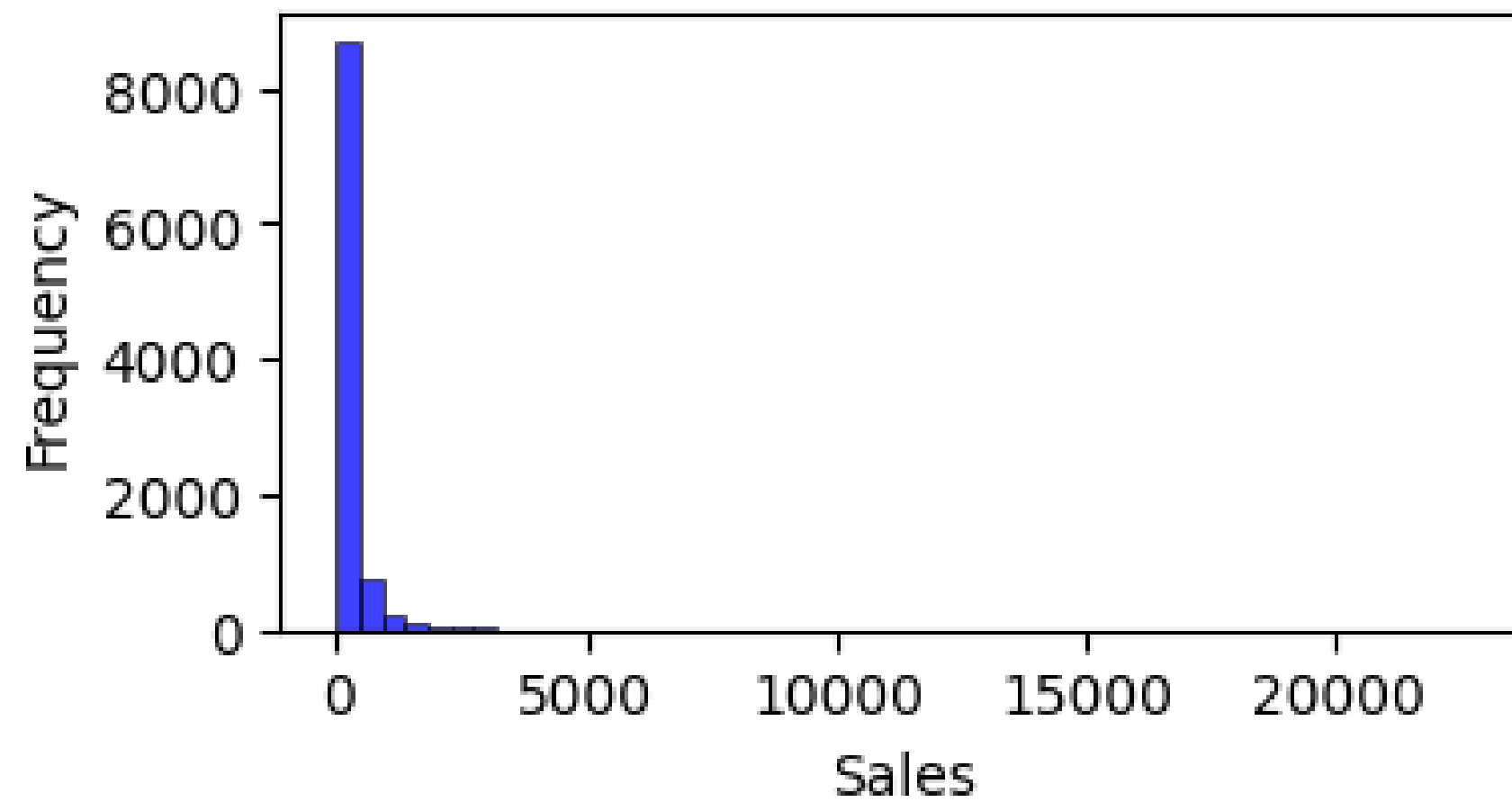


# HISTOGRAM

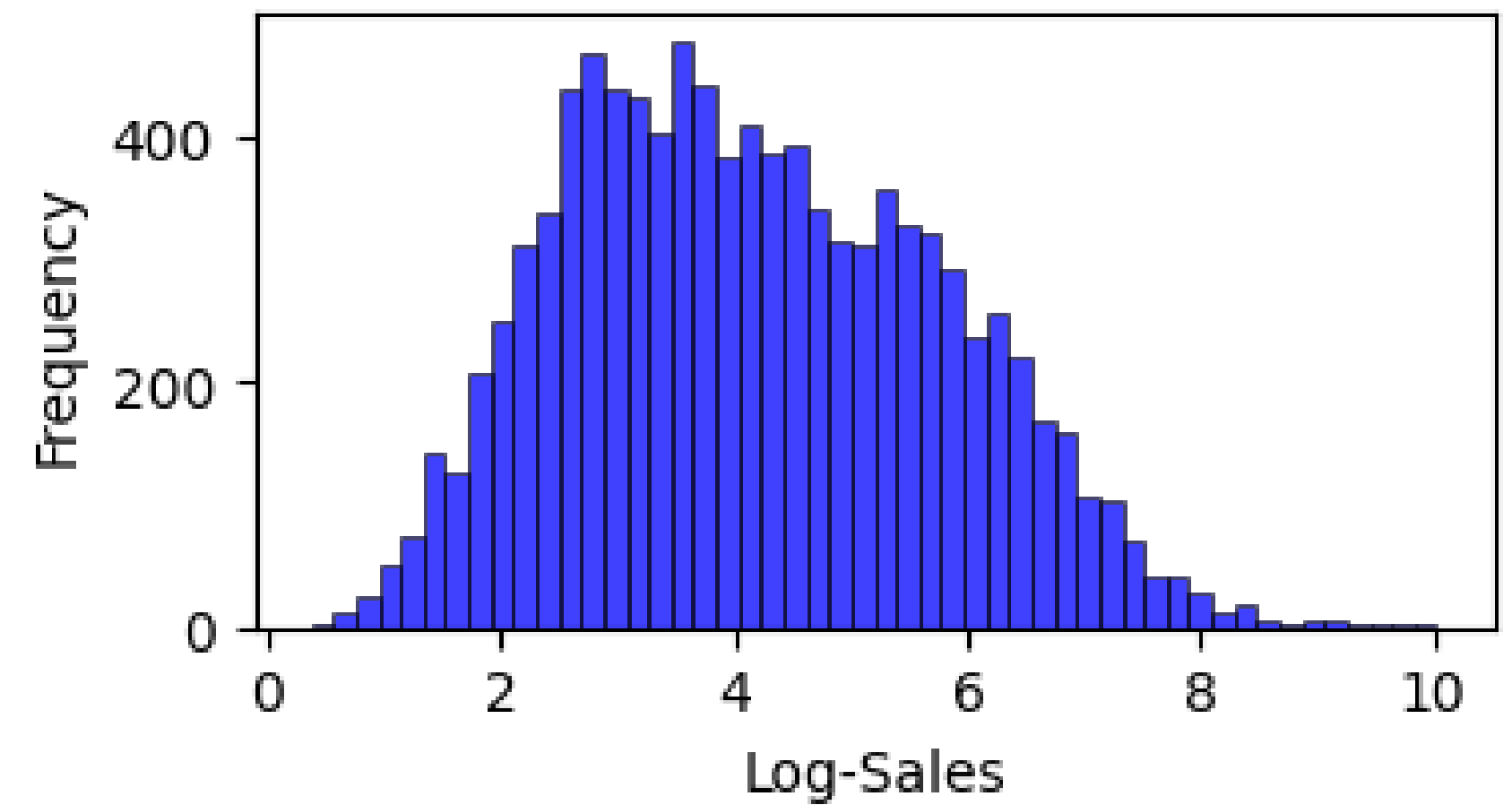
- **Use Case:** Ideal for displaying the distribution of a single numerical variable.
- **Number of Variables:** Histograms typically represent one numerical variable.
- **Type of Data:** Histograms are used with numerical data.
- **Comparison:** Histograms make it easy to understand the distribution of data, identify central tendencies, and spot any outliers or skewness.
- **Considerations:** The choice of the number of bins in a histogram can significantly affect the resulting visualization. Too few bins may oversimplify the data, while too many bins may overcomplicate the data.

# HISTOGRAM

Sales Distribution



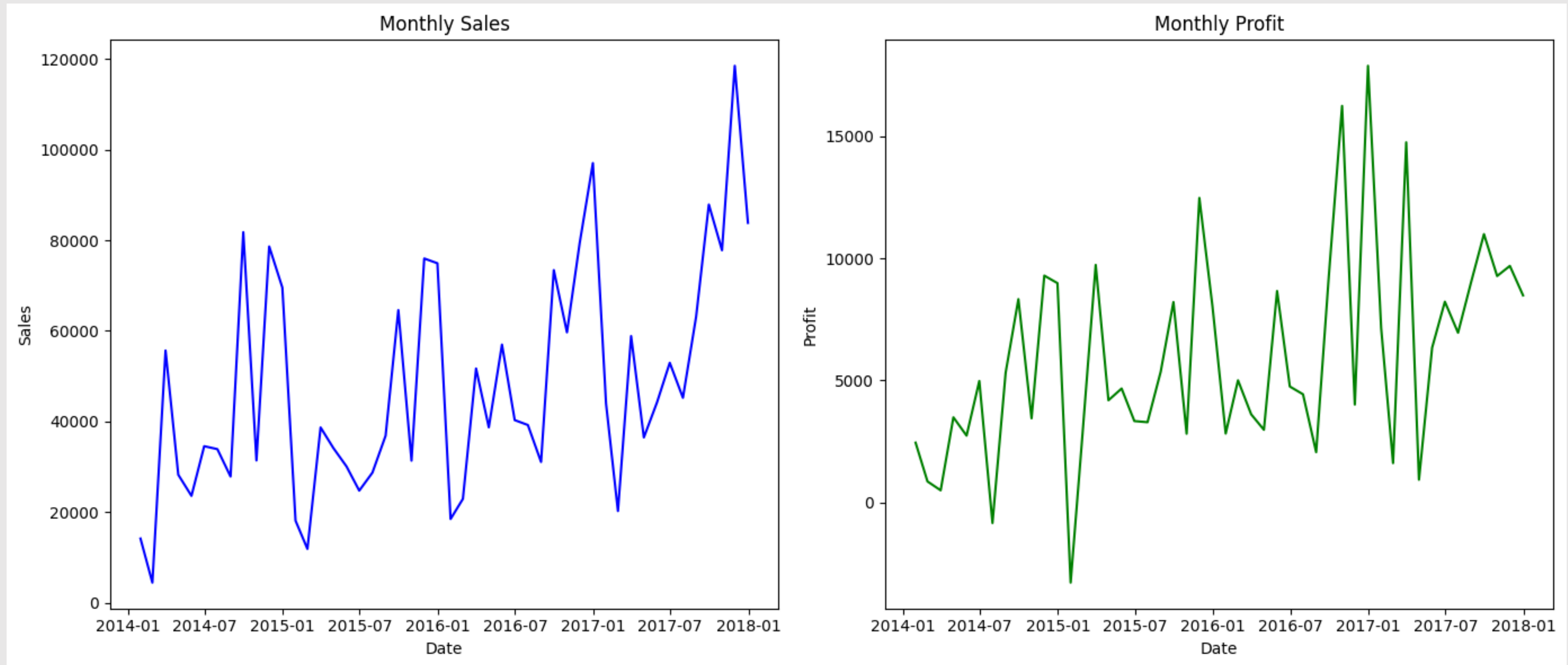
Log-Sales Distribution



# LINE CHART

- **Use Case:** Line charts are best for showing trends over time or any other continuous ordered variable.
- **Number of Variables:** Line charts can represent multiple numerical variables.
- **Type of Data:** Line charts are often used with time series data but can be used with any continuously ordered data.
- **Comparison:** Line charts make it easy to compare changes over time. They are good for showing trends, patterns, and fluctuations.
- **Considerations:** Line charts can become cluttered if there are too many lines (i.e., variables).

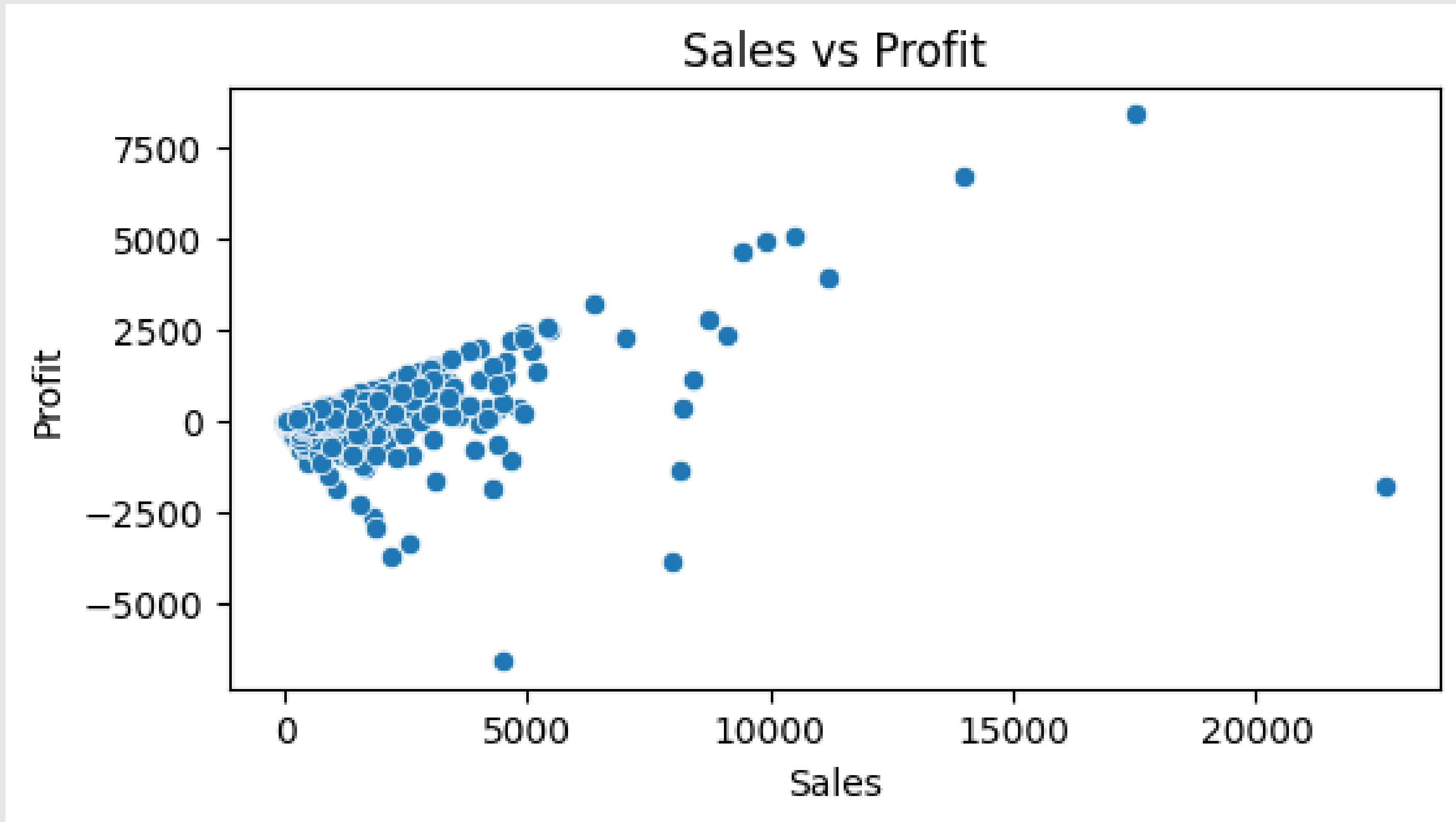
# LINE CHART



# SCATTER PLOT

- **Use Case:** Scatter plots are used to find relationships or correlations between two numerical variables.
- **Number of Variables:** Scatter plots represent two (or three, if color, size, or shape is used to represent another variable) numerical variables.
- **Type of Data:** Scatter plots are used with numerical data.
- **Comparison:** Scatter plots make it easy to compare two numerical variables and identify any trends, clusters, or outliers.
- **Considerations:** Scatter plots require sufficiently large and diverse datasets to be meaningful. They might not be effective if the dataset is too small or if the variables are not numerically continuous.

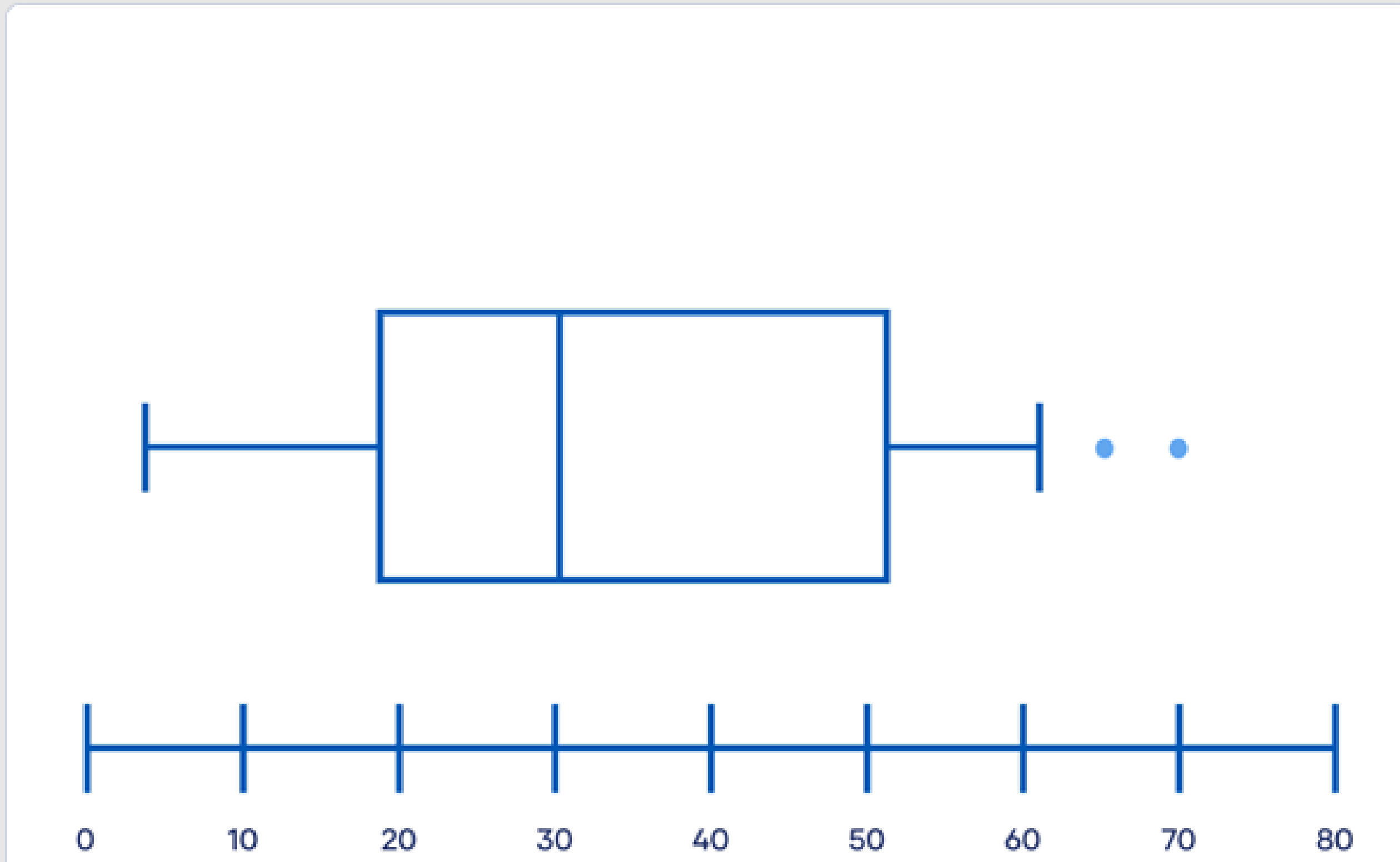
# SCATTER PLOT



# BOXPLOT

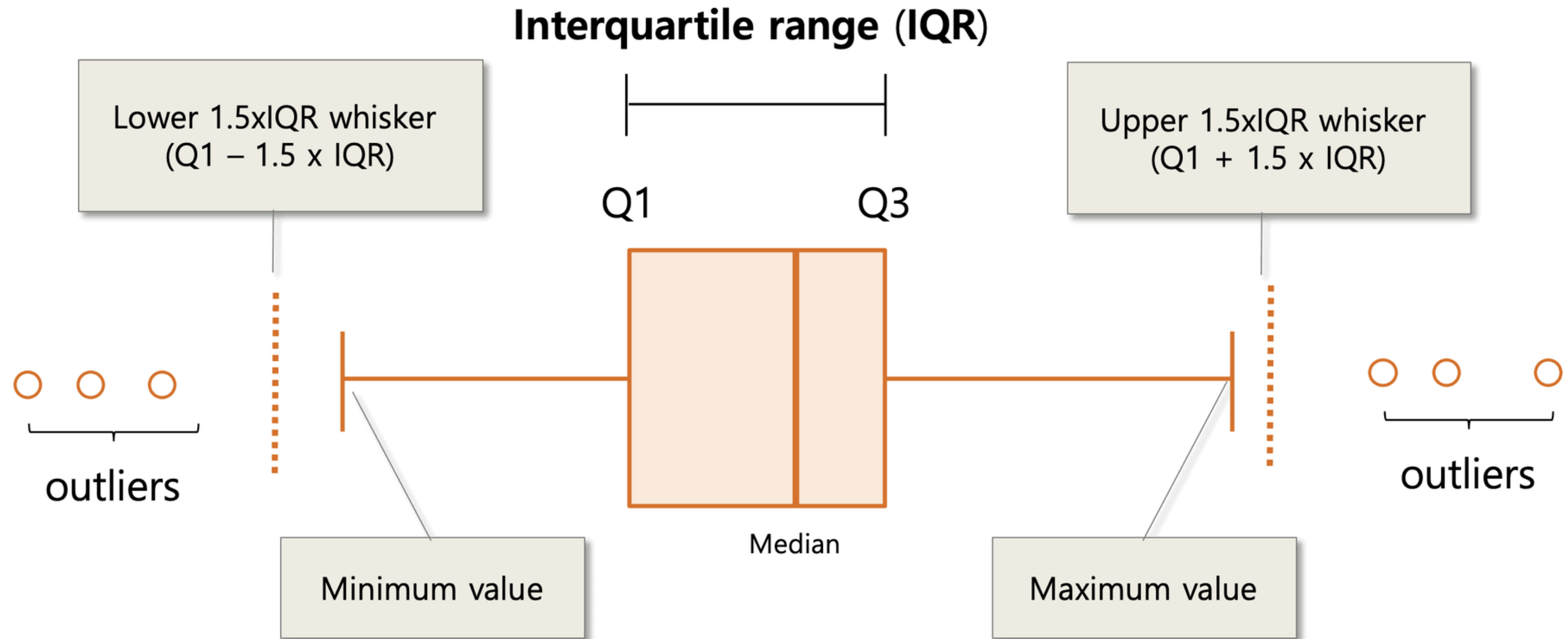
- **Use Case:** Box plots are ideal for displaying the distribution of a numerical variable across different categories.
- **Number of Variables:** Box plots can represent one or more numerical variables and one or more categorical variables.
- **Type of Data:** Box plots are used with one numerical and one categorical variable.
- **Comparison:** Box plots make it easy to compare the distribution of a numerical variable across different categories. They show the median, quartiles, and potential outliers.
- **Considerations:** Box plots provide a more detailed summary of data distribution compared to bar charts, but they can become cluttered if there are too many categories.

# BOXPLOT





# BOXPLOT



# BOXPLOT

