**Question 1:**

Data visualization brings to light insights that might otherwise be concealed in raw data. By visually showing data, analysts can discover correlations, outliers, and critical linkages, allowing firms to learn important lessons that drive growth and expansion initiatives. In this question, I have used Sample Sales Data to compare and contrast the effectiveness of Line Charts, Bar Charts, and Scatter Plots as visualization techniques. The Sample Sales Data contains Order Information, Sales, Customer, Shipping, and Other Information, Used for Segmentation, Customer Analytics, Clustering, and More. Motivated for retail analytics. It was originally written by Mara Carina Roldán, Pentaho Community Member and Business Intelligence Consultant from Argentina.

**Line Chart**

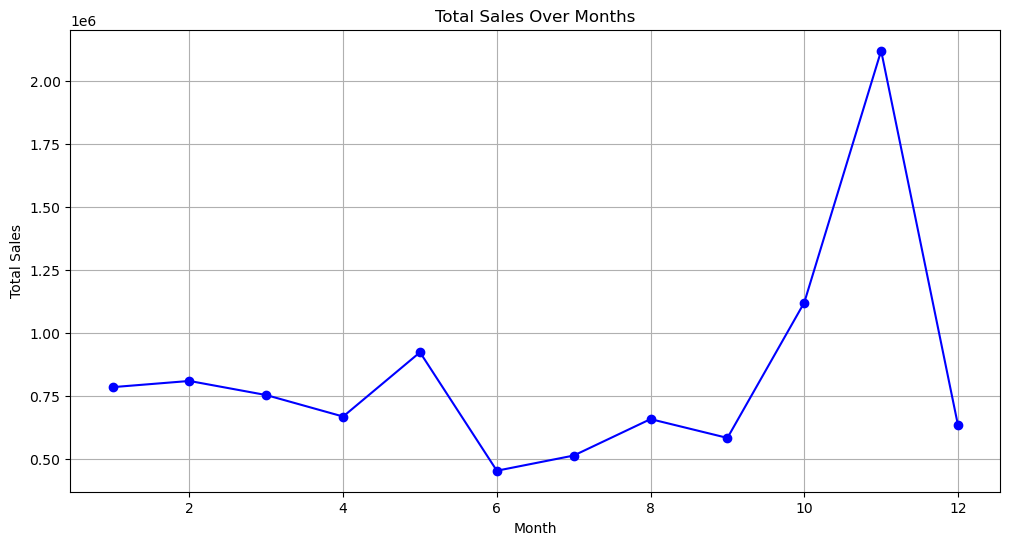
A line chart is a type of chart that displays data as a succession of data points known as 'markers' connected by straight line segments. It's similar to a scatter plot, however the measurement points are arranged (usually by x-axis value) and connected with straight line segments. A line chart is frequently used to depict a pattern in data over time intervals - a time series - hence the line is frequently drawn chronologically.

**Effectiveness**

Line charts are useful for tracking changes over time. A line graph is often constructed with time periods on the x-axis and the number of occurrences on the y-axis.

They are also useful for tracking minor changes. A graph's range can be modified to better zoom in on data that does not vary too much. In comparison to other types of charts, a line graph can be made with extremely small increments on the y-axis, making it easier to see how minor changes have occurred over time.

*Example:* ***Line Chart (Total Sales Over Months)***

****

**Advantages:**

* Suitable for visualizing trends and changes over time.
* Useful for examining how streaming metrics (streams) change over different months or years.
* It is useful in data sets that are continuously updated. Because a line graph is based on a single strain of unbroken data, at least one variable must be continuous.

**Limitations:**

* May not effectively show relationships between multiple variables.
* Less suitable for comparing categorical data.

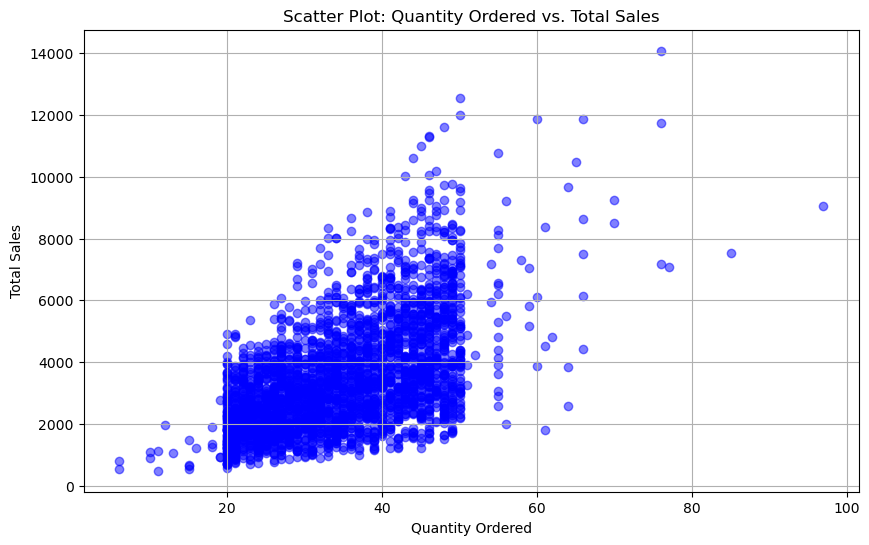
Scatter Plot

A scatter plot is a sort of graphic that displays values for two variables in Cartesian coordinates for a set of data. One more variable can be displayed if the points are coded (color/shape/size). The data is represented as a collection of points, with the value of one variable determining the horizontal axis and the value of the other variable determining the vertical axis.

Effectiveness

A scatter plot is useful when one continuous variable is under the control of the researcher and the other is reliant on it, or when both continuous variables are independent. If one parameter is systematically increased and/or decremented by another, it is typically plotted along the horizontal axis. Traditionally, the measured or dependent variable is plotted along the vertical axis. In the absence of a dependent variable, each type of variable can be displayed on either axis, and a scatter plot will merely show the degree of correlation between two variables.

*Example: Scatter Plot (Quantity Ordered vs. Total Sales)*



**Advantages:**

* Effective for visualizing relationships between two continuous variables.
* Useful for identifying correlations between variables, such as Quantity Ordered vs. Total Sales

**Limitations:**

* Scatter plots are limited to visualizing two variables at a time.
* Scatter plots are not suitable for categorical or discrete variables without additional encoding.

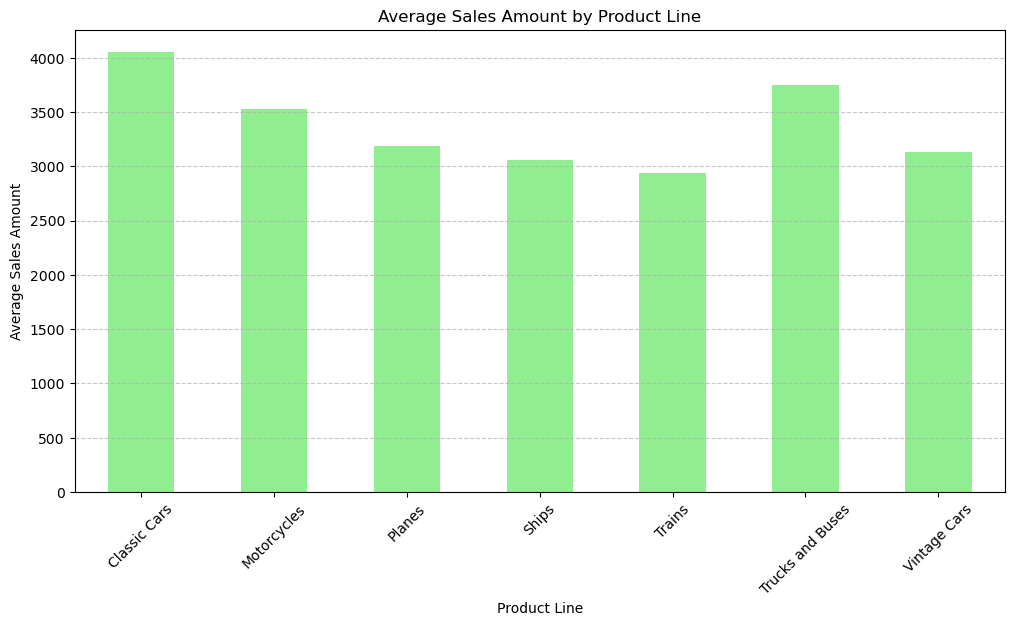
Bar Charts

A bar chart is a chart that displays categorical data using rectangular bars with heights or lengths proportional to the values they represent. The bars can be plotted vertically or horizontally. One axis of the chart depicts the specific categories being compared, while the other axis indicates a measured value.

Effectiveness

Bar charts are useful for displaying categorical data such as months of the year and age groups. Because these categories are typically qualitative, categories are displayed along the horizontal axis in a bar chart, while the vertical axis corresponds to the value of each category. With grouped bar charts and stacked bar charts, bar charts may also be used to illustrate more sophisticated data comparisons.

*Example:Bar Charts(Average Sales Amount by Product Line)*

**Advantages:**

* Bar charts are a user-friendly tool for comparing values and identifying trends, making them suitable for a wide range of audiences.
* Bar charts can be customized to suit user needs, such as color, width, and height, and can be enhanced with labels and annotations for additional information.
* They can handle large amounts of data, providing a clear representation through their ability to be narrow or wide, and the use of color or patterns to distinguish between data points.
* Bar charts are particularly useful for comparing values between categories or data points, enabling quick identification of differences and similarities, facilitating decision-making and comparison of data points.

**Limitations:**

* Bar charts are not suitable for displaying continuous data like temperature or time
* Bar charts can be misleading if the scale is not appropriate or if the data is presented in an intentionally misleading manner. If the y-axis is truncated, for example, the discrepancies between the bars may appear larger than they are.
* Multivariate data has a limited scope: Because bar charts can only show one or two variables at a time, they are ineffective for displaying multivariate data. A scatter plot or heat map may be more suited in such circumstances.

**References**

Bar chart. (2023, August 6). In *Wikipedia*. https://en.wikipedia.org/wiki/Bar\_chart