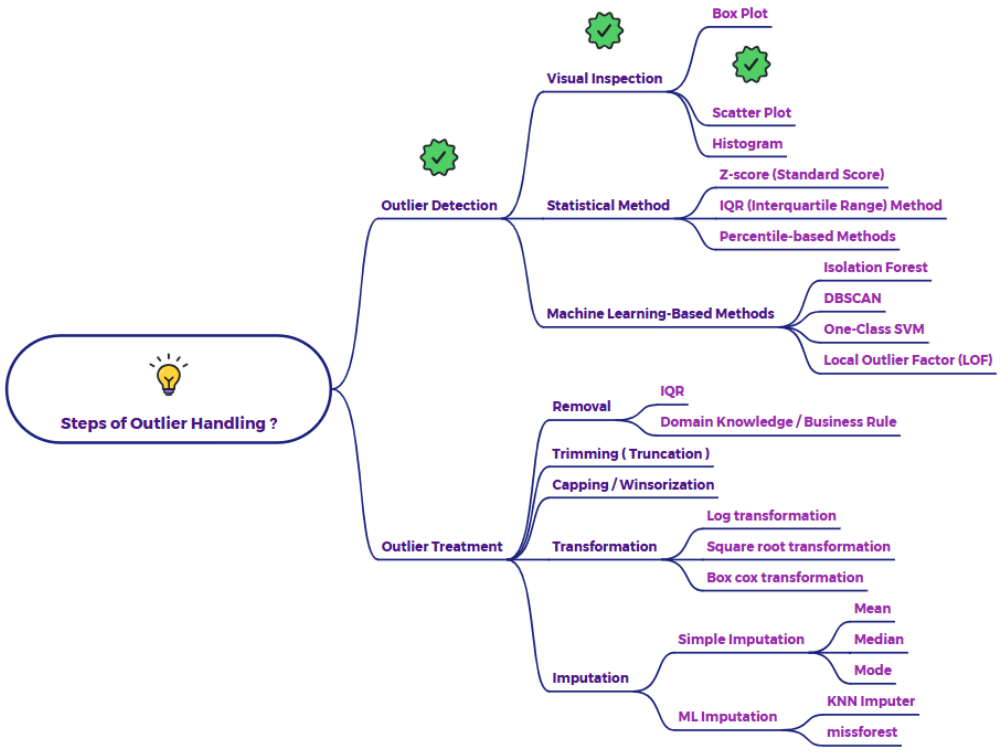
Explain Outlier detection through visual inspection (Scatter plot)



Outlier Detection: Visual Inspection - Scatter Plot

A scatter plot is a graphical tool used to visualize the relationship between two numerical variables. It displays data points as individual dots on a two-dimensional plane, where the position of each dot corresponds to its values for the two variables. Scatter plots are excellent for identifying potential outliers in bivariate data.

How to Interpret a Scatter Plot for Outliers

In a scatter plot, most data points tend to form a general pattern or cluster. Outliers, in this context, are data points that deviate significantly from this pattern. They appear as points that are far away from the main cluster of data points.

Example:

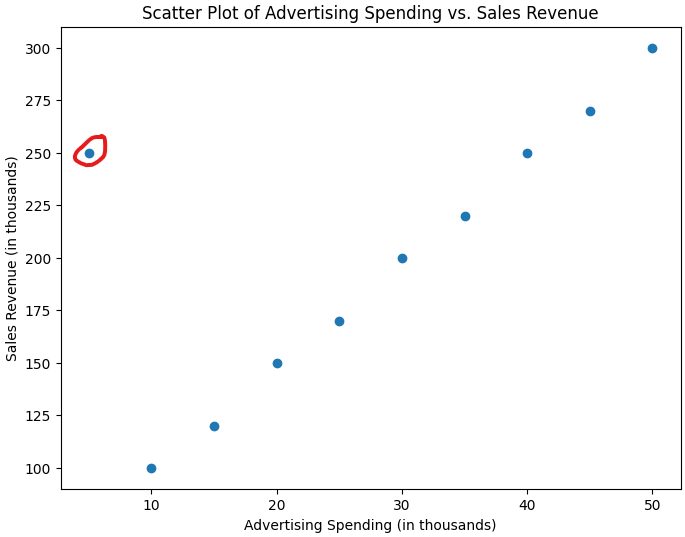
Let's consider a dataset that shows the relationship between "Advertising Spending" and "Sales Revenue" for a company:

Advertising Spending (in thousands)	Sales Revenue (in thousands)
10	100
15	120
20	150
25	170
30	200
35	220
40	250
45	270
50	300
5	250

In this example, we expect that as advertising spending increases, sales revenue also increases. The data points from (10, 100) to (50, 300) show this positive relationship. However, the last data point (5, 250) seems unusual. It has very low advertising spending but relatively high sales revenue, which deviates from the general trend.

In a scatter plot of this data:

- The data points (10, 100) through (50, 300) would form a roughly linear, upward-sloping pattern.
- The point (5, 250) would be located far away from this pattern, likely appearing in the upper-left corner of the plot. This isolated position would visually indicate that (5, 250) is a potential outlier.



Benefits of Using Scatter Plots for Outlier Detection:

- **Visual Identification:** Scatter plots make it easy to visually identify data points that deviate from the general relationship between two variables.
- **Bivariate Analysis:** They are specifically designed for exploring relationships and detecting outliers in two-dimensional data.
- **Pattern Recognition:** They help in understanding the overall pattern of the data and how outliers deviate from that pattern.

Limitations:

- **Two Variables Only:** Scatter plots are limited to visualizing the relationship between only two variables at a time.
- **Subjectivity:** Identifying outliers in a scatter plot can be somewhat subjective, especially when the deviation is not very extreme.