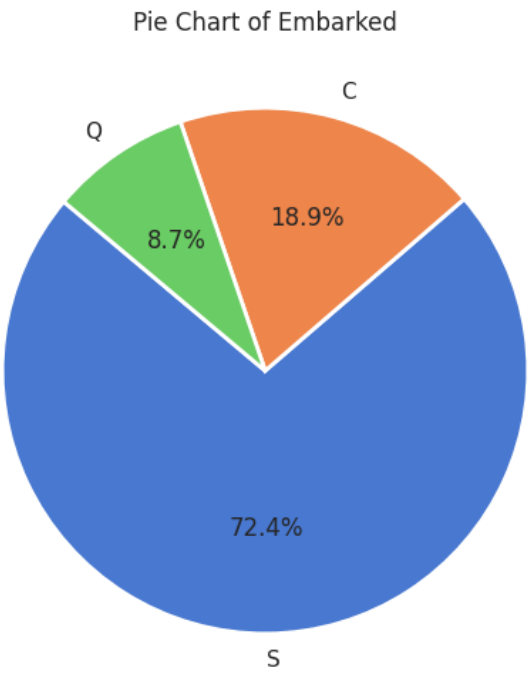
How to interpret Pie Chart ?



A. Interpretation of the Pie Chart Components:

- **The Circle:** The entire circle represents the whole dataset or 100% of the observations for the "Embarked" variable.
- **Slices:** The circle is divided into different slices, each representing a category of the "Embarked" variable ('S', 'C', and 'Q').
- **Size of the Slices:** The size (area and angle) of each slice is proportional to the proportion or percentage of observations belonging to that category relative to the whole dataset.
- **Labels:** Each slice is labeled with the category it represents ('S', 'C', 'Q') and the corresponding percentage of the total.
    - 'S' (blue) occupies the largest slice and represents 72.4% of the embarked passengers.
    - 'C' (orange) occupies the second largest slice and represents 18.9% of the embarked passengers.
    - 'Q' (green) occupies the smallest slice and represents 8.7% of the embarked passengers.

B. Interpreting the "Embarked" Distribution:

The pie chart visually shows the relative proportions of passengers embarking from each port:

- **'S' (72.4%):** The vast majority of passengers (almost three-quarters) embarked from the port represented by 'S'.
- **'C' (18.9%):** A smaller but still significant proportion of passengers embarked from the port represented by 'C'.
- **'Q' (8.7%):** The smallest proportion of passengers embarked from the port represented by 'Q'.

The pie chart provides a quick visual understanding of the relative contribution of each port of embarkation to the total number of passengers.

Pie charts are best used in a limited set of scenarios:

- **Showing Parts of a Whole:** Their primary strength is to illustrate how a whole is divided into its constituent parts, emphasizing the proportion of each category relative to the total.
- **Simple Datasets with Few Categories:** Pie charts work best when there are only a few (typically 2-5) distinct categories. With too many slices, the chart becomes cluttered and difficult to read.
- **Highlighting Large Proportions:** They can be effective at quickly conveying which category represents a large majority or a small minority of the data. In the example, it's immediately clear that 'S' represents the dominant embarkation port.

However, it's important to note that bar plots are generally preferred over pie charts in most situations for the following reasons:

- **Difficulty in Comparing Sizes:** It can be harder for the human eye to accurately compare the sizes of slices (especially by area or angle) than to compare the lengths of bars in a bar plot. This makes it difficult to precisely determine the relative proportions, especially for slices that are close in size.
- **Handling Many Categories:** Pie charts become very cluttered and unreadable with more than a few categories, whereas bar plots can

accommodate more categories (though readability might still be a concern with a very large number).

- **Difficulty in Judging Small Differences:** Small differences in proportions can be very hard to discern in a pie chart.
- **Lack of Precise Values:** While percentages are often included, the underlying counts are not directly represented in the same way as in a bar plot.

In summary, while pie charts can be visually appealing for simple datasets with a few distinct and significantly different proportions, bar plots are generally a more effective and versatile choice for displaying the distribution of categorical variables due to their ease of comparison and ability to handle more complex data. Avoid using pie charts when precise comparisons are needed, when there are many categories, or when the proportions are similar. In many cases, a bar plot would convey the information more clearly and accurately.