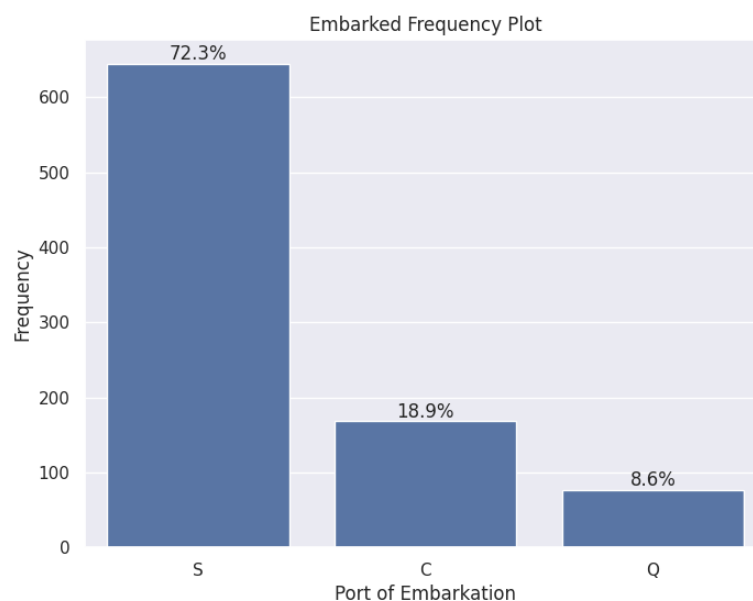


How to interpret frequency plots (%) ?



A. Interpretation of the Frequency Plot Components:

- **Horizontal Axis (X-axis):** Represents the different categories of the categorical variable "Embarked". Here, the categories are 'S', 'C', and 'Q', likely representing different ports of embarkation.
- **Vertical Axis (Y-axis):** Represents the **frequency** or **count** of observations within each category. The scale ranges from 0 to 600+.
- **Bars:** Each vertical bar corresponds to one of the categories of the "Embarked" variable.
 - The **height** of each bar indicates the number of observations that fall into that specific category.
 - The labels above each bar explicitly show the **percentage** of the total observations that fall into that category:
 - 'S' has 72.3%.
 - 'C' has 18.9%.
 - 'Q' has 8.6%.

B. Interpreting the "Embarked" Distribution (Percentages):

This frequency plot, displaying percentages, shows the relative proportion of passengers based on their port of embarkation:

- **'S' (72.3%):** A very large majority (72.3%) of the passengers embarked from the port represented by 'S'.
- **'C' (18.9%):** A smaller but still notable proportion (18.9%) of the passengers embarked from the port represented by 'C'.
- **'Q' (8.6%):** A relatively small proportion (8.6%) of the passengers embarked from the port represented by 'Q'.

This plot emphasizes the proportion of each category relative to the total number of observations, making it easy to see the dominant category and the relative sizes of the others.

Frequency plots showing percentages are the best choice in the following scenarios:

- **Emphasizing Relative Proportions:** When the primary goal is to understand the share or proportion of each category within the whole dataset, displaying percentages directly above the bars makes this information immediately accessible.
- **Comparing Categories to the Whole:** It clearly shows how each category contributes to the total distribution.
- **When the Exact Counts Are Less Important:** If the focus is on the relative size of groups rather than the absolute number of observations, percentage plots are more effective.
- **Facilitating Comparisons Across Different Datasets (Normalized View):** When comparing the distribution of the same categorical variable across datasets of different sizes, using percentages normalizes the data, allowing for a more meaningful comparison of the relative proportions of each category.
- **Communicating Insights to a General Audience:** Percentages are often more intuitive for a broader audience to grasp relative sizes compared to raw counts.

In contrast to frequency plots showing raw counts:

- **Percentage plots highlight the relative contribution of each category**, which can be more insightful when the total number of observations isn't the primary focus.
- **Raw count plots emphasize the absolute number of observations** in each category.

Similarities to Pie and Donut Charts:

- Like pie and donut charts, percentage frequency plots show the parts of a whole. However, bar plots generally offer better readability and allow for easier comparison of category sizes, especially when there are many categories or when proportions are similar.

In summary, frequency plots displaying percentages are the best choice when the focus is on the relative distribution of categories within a categorical variable. They provide a clear and easily understandable way to see the proportion each category represents of the total dataset, making them useful for emphasizing relative sizes and facilitating comparisons.