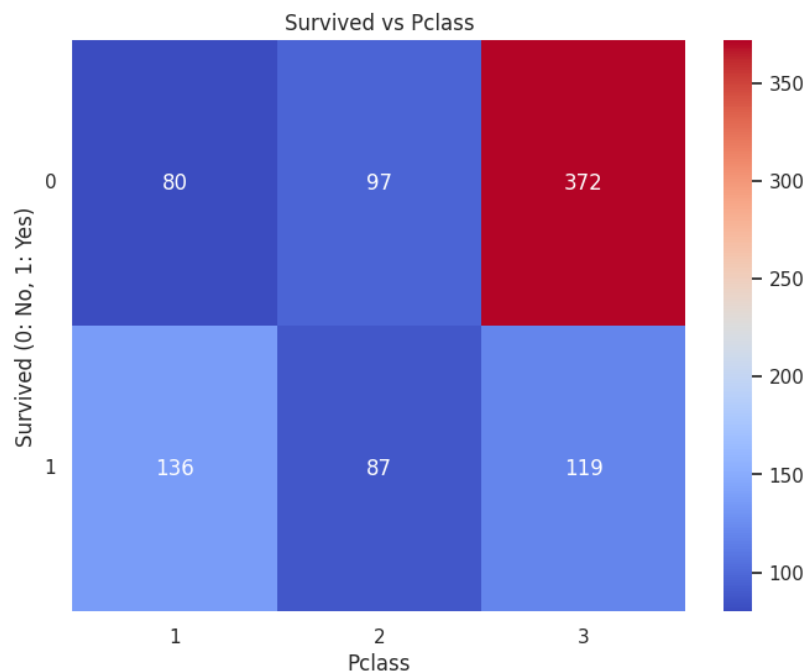


How to interpret contingency table using heatmap using bivariate analysis



A. Understanding the Components of a Contingency Table Heatmap:

- **Axes (X and Y):** The horizontal axis represents one categorical variable ("Pclass" with categories 1, 2, and 3), and the vertical axis represents the other categorical variable ("Survived" with categories 0 for No and 1 for Yes).
- **Cells:** Each cell in the heatmap represents the **count** of observations that fall into the intersection of the categories for the two variables.¹ For example, the top-left cell shows the number of passengers who did not survive (Survived=0) and were in Pclass 1.
- **Color Intensity:** The color intensity of each cell is mapped to the count within that cell. A color bar on the right provides the scale. In this image, a blue-to-red color scheme is used, where:
 - **Blue hues** represent lower counts.
 - **Red hues** represent higher counts.
 - Intermediate colors represent counts in between.

- **Numerical Values:** The numerical value within each cell explicitly shows the count for that combination of categories.

B. Interpreting the Relationship Between Survived and Pclass:

By examining the heatmap, we can understand how survival rates varied across different passenger classes:

- **Pclass 1 (First Class):**
 - The top cell (Survived=0, Pclass=1) has a count of 80 (lighter blue), indicating a relatively lower number of first-class passengers did not survive.
 - The bottom cell (Survived=1, Pclass=1) has a count of 136 (lighter blue), indicating a higher number of first-class passengers survived.
 - This suggests a higher survival rate for first-class passengers.
- **Pclass 2 (Second Class):**
 - The top cell (Survived=0, Pclass=2) has a count of 97 (light blue), indicating a moderate number of second-class passengers did not survive.
 - The bottom cell (Survived=1, Pclass=2) has a count of 87 (light blue), indicating a slightly lower number of second-class passengers survived compared to those who did not.
 - The survival rate for second-class passengers appears lower than for first class.
- **Pclass 3 (Third Class):**
 - The top cell (Survived=0, Pclass=3) has a count of 372 (deep red), indicating a very high number of third-class passengers did not survive.
 - The bottom cell (Survived=1, Pclass=3) has a count of 119 (light blue), indicating a lower number of third-class passengers survived compared to those who did not.

- This strongly suggests a much lower survival rate for third-class passengers.

C. Overall Interpretation:

The heatmap clearly shows a strong association between passenger class and survival outcome. First-class passengers had the highest number of survivors and a relatively lower number of non-survivors. In contrast, third-class passengers had the highest number of non-survivors and a relatively lower number of survivors. Second-class passengers fell somewhere in between. The color intensity reinforces this observation, with the deep red cell highlighting the high number of non-survivors in third class.

Contingency table heatmaps are the best choice for visualizing the relationship between two categorical variables when you want to:

- Display the frequency distribution of the categories of two variables simultaneously. Each cell clearly shows the number of occurrences for a specific combination of categories.
- Quickly identify patterns and associations between the categories. Differences in color intensity make it easy to spot combinations with high or low counts.
- Compare the distribution across different categories. You can visually assess if certain categories of one variable are more likely to occur with certain categories of the other variable.
- Visualize the results of a cross-tabulation or frequency table in an intuitive way. The heatmap format makes it easier to grasp the relationships compared to just looking at the raw numbers.
- Highlight cells with the most or least frequent combinations. The color scale effectively draws attention to these areas.
- Present the relationship between two categorical variables in a visually appealing and easily understandable format.

In summary, a contingency table heatmap is a powerful tool for exploring and presenting the association between two categorical variables by visualizing the counts of their co-occurrence. It allows for quick identification of patterns and differences in the distribution of the categories.