1. To find the number of unique values corresponding to each column in the Data Frame
2. find the **percentage of missing values** in the dataset
3. plot missing values
4. Find and Impute the missing values
5. Remove unwanted Rows and Columns, Find the Duplicate?
6. Plot and Find Outliers?

**What is the distribution of positive axillary nodes in the patients?**

1. **How is the dataset distributed across different marital statuses, and what percentage of patients belong to each category?**
2. **What is the cumulative distribution of age groups in the Cancer Survival Prediction dataset, and how does it illustrate the overall distribution of patients across different age groups?**
3. **How many patients were operated in each year for each Age group?**
4. Does the distribution of ages vary between patients who survived and those who did not? Are there any noticeable differences in the median age or the spread of ages for different survival outcomes?
5. Are there any trends or patterns in the distribution of the year of operation for patients based on their survival status? Does the boxplot reveal any significant differences in the years of operation for patients who survived compared to those who did not?
6. How does the distribution of positive axillary nodes differ between patients with different survival outcomes? Are there any clear differences in the spread of positive axillary nodes for survivors and non-survivors?
7. How does the age distribution vary between patients who survived and those who did not?
8. What insights can be gained regarding the year of operation for patients with different survival outcomes?
9. How is the distribution of positive axillary nodes different for patients with varying survival outcomes?
10. **Is there a correlation between Tumor\_Size and the number of Positive\_Axillary\_Nodes?**
11. **How is the distribution of age\_groups different for patients who survived and those who did not?**