Univariate Analysis for Categorical Variables

1. What is Univariate Analysis for Categorical Variables?

Univariate analysis for categorical variables involves analyzing a single categorical variable. It helps in understanding the distribution of categories, identifying dominant categories, and detecting imbalances in the dataset.

Common categorical plots:
- Count Plot (Bar Chart)
- Pie Chart

2. Plots for Categorical Data

2.1 Count Plot (Bar Chart)

A count plot (bar chart) is used to visualize the frequency of different categories. It helps in identifying the dominant categories and class imbalances.

```
import seaborn as sns
import matplotlib.pyplot as plt

sns.countplot(x=df["sex"])
plt.title("Gender Count")
plt.show()
```

2.2 Pie Chart (Category Proportion)

A pie chart helps in understanding the proportion of each category relative to the whole dataset.

```
df["sex"].value_counts().plot.pie(autopct="%1.1f%%", colors=["lightblue", "pink"])
plt.title("Gender Distribution")
plt.ylabel("")
plt.show()
```

3. Conclusion

Count plots and pie charts are essential for analyzing categorical variables. Count plots help visualize category frequencies, while pie charts show relative proportions. These visualizations are crucial for detecting class imbalances and understanding dataset composition.