1. What is the Titanic dataset?

Answer: The Titanic dataset contains information about passengers on the Titanic ship, including their age, sex, class, fare, survival status, and other characteristics. It is often used for data analysis and machine learning tasks.

---

2. What is the purpose of using the Seaborn library in this task?

Answer: Seaborn is used for data visualization. It provides a high-level interface for drawing attractive and informative statistical graphics, making it easier to explore and understand the dataset's patterns.

---

3. What is the role of `sns.load_dataset('titanic')`?

Answer: This command loads the Titanic dataset from Seaborn's built-in datasets into a Pandas DataFrame, which can then be used for analysis and visualization.

---

4. What does `df.info()` do?

Answer: The `df.info()` method provides a concise summary of the DataFrame, including the number of non-null entries, data types of each column, and memory usage.

---

5. What is the significance of `df.describe()`?

Answer: The `df.describe()` method generates summary statistics for numerical columns, including measures like mean, median, standard deviation, and percentiles.

---

6. What does `df.shape` return?

Answer: The `df.shape` method returns the dimensions of the DataFrame as a tuple (rows, columns), which helps to understand the dataset's size.

---

7. Why do we use `sns.histplot()`?

Answer: `sns.histplot()` is used to create a histogram of continuous data. In this case, it visualizes the distribution of ticket fares for Titanic passengers.

---

8. How do you visualize the distribution of fare?

Answer: By using `sns.histplot(x='fare', data=df)`, a histogram is generated that shows the distribution of ticket fares across the Titanic passengers.

---

9. How is the distribution of age visualized in the Titanic dataset?

Answer: The distribution of age is visualized using `sns.displot(x='age', data=df, bins=70, height=5, aspect=1.2)` to show the frequency of different age ranges in the dataset.

---

10. What does `sns.catplot()` do?

Answer: `sns.catplot()` is used to create categorical plots. It can visualize the distribution of categories based on other categorical variables, such as showing survival count by passenger class or sex.

---

11. How does `sns.catplot(x='survived', data=df, kind='count', hue='pclass')` work?

Answer: This command creates a count plot showing the number of survivors and non-survivors for each passenger class, with different classes distinguished by color (`hue='pclass'`).

---

12. What is the purpose of setting `sns.set(rc={'figure.figsize': (6, 4)})`?

Answer: This command sets the figure size for the plot, making the plot's visual display clearer and more readable.

---

13. What is the difference between `sns.histplot()` and `sns.displot()`?

Answer: `sns.histplot()` creates a histogram for a single variable, while `sns.displot()` can generate a variety of distribution plots, including histograms and kernel density estimates.

---

14. What does `plt.title('Distribution of Fare')` do?

Answer: This command adds a title to the plot, making it clear that the histogram represents the distribution of fares for Titanic passengers.

---

15. Why do we ignore warnings using `warnings.filterwarnings('ignore')`?

Answer: This command is used to suppress warnings in the output, helping to keep the console output clean and focused on the important information.

---

16. What does `df['fare']` represent in the Titanic dataset?

Answer: `df['fare']` refers to the ticket fare each passenger paid to board the Titanic. It is a numerical column used for analyzing passengers' financial contribution.

---

17. How do you interpret a histogram?

Answer: A histogram provides insights into the frequency of data points across different ranges. For example, a histogram of fares may show if most passengers paid low or high fares.

---

18. What does the `hue` parameter in `sns.catplot()` do?

Answer: The `hue` parameter adds color coding to the plot based on a categorical variable, allowing you to differentiate data points within each category.

---

19. How do you find the survival rate in the Titanic dataset?

Answer: You can use a count plot with `sns.catplot(x='survived', data=df, kind='count')` to visualize the number of survivors and non-survivors, and then calculate the survival rate from the proportions.

---

20. What are the possible values for the `survived` column?

Answer: The `survived` column contains binary values: 0 for non-survivors and 1 for survivors.

---

21. What does `sns.set(rc={'figure.figsize': (5, 5)})` do?

Answer: This command sets the figure size for the plots. Here, it specifies a square-shaped figure (5x5 inches) for better visual balance in the categorical plots.

---

22. What insights can you draw from the distribution of fare in the Titanic dataset?

Answer: By analyzing the fare distribution, we can determine if there are significant outliers or patterns in the data, such as most passengers paying low fares and a few paying much higher fares.

---

23. What does the `kind='count'` argument do in `sns.catplot()`?

Answer: The `kind='count'` argument in `sns.catplot()` specifies that we want to count the occurrences of each category, which is useful for visualizing categorical data like survival or class distribution.

---

24. What is the role of `hue='sex'` in `sns.catplot()`?

Answer: The `hue='sex'` argument differentiates passengers by gender, showing how survival rates differ between males and females.

---

25. Why is exploratory data analysis (EDA) important for machine learning?

Answer: EDA helps understand the underlying structure of the data, detect patterns, identify missing or outlier values, and prepare the data for further analysis or model training, improving the quality of machine learning models.