Here's a simplified explanation of your Titanic dataset visualization code, along with potential viva questions and easy-to-understand answers:

### \*\*Code Explanation (Simple Version)\*\*

1. \*\*Loading Data\*\*

```python

df = sns.load\_dataset('titanic')

```

- Loads built-in Titanic passenger dataset (891 rows)

- Contains info like fares, ages, survival status, passenger class

2. \*\*Basic Histogram (Matplotlib)\*\*

```python

plt.hist(df['fare'], bins=30)

```

- Shows fare distribution

- Most fares clustered at left (cheaper tickets)

- Few high fares (luxury suites) → right tail

3. \*\*Enhanced Distribution Plot (Seaborn)\*\*

```python

sns.distplot(df['fare'], bins=30)

```

- Histogram + smooth curve (KDE)

- Better shows fare concentration around £10-30

4. \*\*Modern Histplot\*\*

```python

sns.histplot(df['fare'], kde=True)

```

- Cleaner version of distplot

- Blue bars + orange density line

5. \*\*Joint Plot (Fare vs Age)\*\*

```python

sns.jointplot(x='fare', y='age', data=df)

```

- Scatter plot + histograms

- Shows:

- Most children (low age) had cheap fares

- Expensive fares → mostly adults

6. \*\*Rug Plot\*\*

```python

sns.rugplot(data=df)

```

- Tiny lines show exact fare values

- Reveals clusters at £0-10, £10-20, etc.

7. \*\*Bar Plot (Class vs Avg Fare)\*\*

```python

sns.barplot(x='class', y='fare', data=df)

```

- 1st class: £84 average

- 2nd class: £21

- 3rd class: £13

- Clear class hierarchy in pricing

---

---

### \*\*Technical Questions\*\*

\*\*Q1: Why set `bins=30` in the histogram?\*\*

A1: It divides the fare range into 30 equal intervals. More bins show finer details, but too many can make the plot messy.

\*\*Q2: What does `edgecolor='black'` do in the histogram?\*\*

A2: It adds black borders to each bar, making them visually distinct.

\*\*Q3: Why use both Matplotlib and Seaborn?\*\*

A3: Matplotlib is the base library (more control), while Seaborn builds on it with prettier defaults and advanced plots.

\*\*Q4: What does `kde=True` do in `histplot`?\*\*

A4: It overlays a Kernel Density Estimate (smooth curve) on the histogram to show the probability distribution.

\*\*Q5: How is `jointplot` different from a regular scatter plot?\*\*

A5: A jointplot adds histograms/KDEs on the margins, showing distributions of both variables simultaneously.

---

### \*\*Data Interpretation Questions\*\*

\*\*Q6: What does the right-skewed fare distribution imply?\*\*

A6: Most passengers paid low fares, with a few paying very high amounts (luxury passengers).

\*\*Q7: Why does the fare-age jointplot show vertical lines?\*\*

A7: Many passengers paid the same fare (e.g., groups/families buying tickets at fixed prices).

\*\*Q8: What does the rugplot reveal about fare distribution?\*\*

A8: Dense "ticks" at £0-30 confirm most passengers were in this range, with sparse ticks beyond £100.

\*\*Q9: How does passenger class relate to fare in the barplot?\*\*

A9: 1st class (highest fare) > 2nd class > 3rd class (lowest fare), reflecting the Titanic's social hierarchy.

\*\*Q10: What might outliers in fare represent?\*\*

A10: Luxury suites or group bookings (e.g., families purchasing multiple tickets together).

---

### \*\*Practical Application Questions\*\*

\*\*Q11: How would you visualize survival rates by passenger class?\*\*

A11:

```python

sns.barplot(x='class', y='survived', data=df, estimator=lambda x: sum(x)/len(x))

```

\*\*Q12: How to compare age distributions between survivors and non-survivors?\*\*

A12:

```python

sns.boxplot(x='survived', y='age', data=df)

```

\*\*Q13: What plot would show gender-wise survival counts?\*\*

A13:

```python

sns.countplot(x='sex', hue='survived', data=df)

```

\*\*Q14: How to check for missing values visually?\*\*

A14:

```python

sns.heatmap(df.isnull(), cbar=False)

```

\*\*Q15: How would you visualize embarkation port vs fare?\*\*

A15:

```python

sns.violinplot(x='embark\_town', y='fare', data=df)

```

---

### \*\*Conceptual Questions\*\*

\*\*Q16: When would you use a histogram vs a boxplot?\*\*

A16:

- Histogram: To see the full distribution shape.

- Boxplot: To compare groups and identify outliers.

\*\*Q17: Why is exploratory data analysis (EDA) important?\*\*

A17: It helps uncover patterns, detect anomalies, and guide feature selection before modeling.

\*\*Q18: What insights could you derive from fare vs survival analysis?\*\*

A18: Higher fare payers had better survival rates (likely due to priority access to lifeboats).

\*\*Q19: How would you improve these visualizations?\*\*

A19:

- Add titles/axis labels

- Use consistent color schemes

- Annotate key points (e.g., average fare)

\*\*Q20: What other Titanic features would be interesting to visualize?\*\*

A20:

- Survival by age group

- Family size (sibsp + parch) vs survival

- Cabin location (if data available)

---

### \*\*Bonus: Pro Tip for Interviews\*\*

When asked to interpret plots, follow this structure:

1. \*\*Describe\*\* ("The histogram shows...")

2. \*\*Patterns\*\* ("Most values cluster around...")

3. \*\*Exceptions\*\* ("Notably, there's an outlier at...")

4. \*\*Implications\*\* ("This suggests that...")

Example for the fare histogram:

\*"The right-skewed distribution indicates most passengers paid under £50, with a few paying £200+. This reflects the socioeconomic divide on the Titanic, where wealthier passengers could afford luxury cabins."\*

---

Let me know if you'd like me to generate code for any of the suggested visualizations!