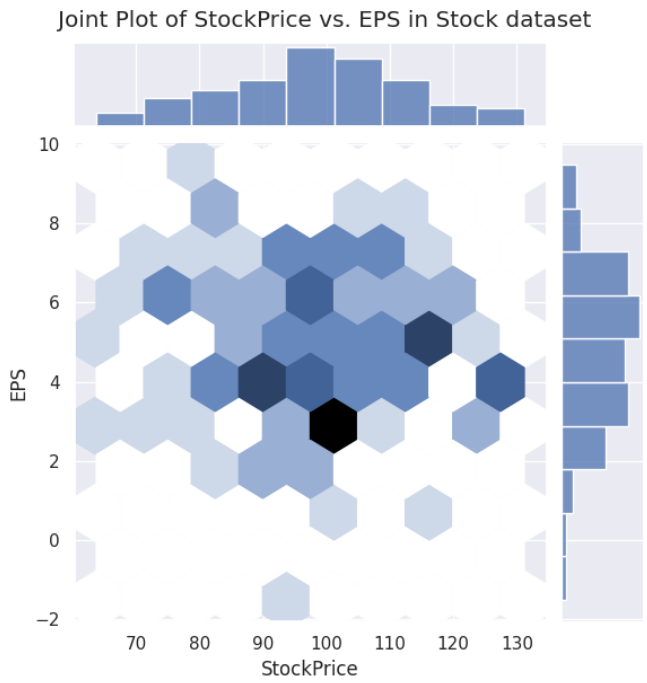
How to interpret joint plot ?



A. Understanding the Components of a Joint Plot:

A joint plot is a combination of three plots:

- **Central Bivariate Plot:** This is the main part of the visualization, showing the relationship between the two numerical variables. In this case, it's a **hexbin plot**.
- **Top Marginal Plot:** This is a univariate distribution plot (usually a histogram or a density plot) showing the distribution of the variable on the x-axis ("StockPrice").
- **Right Marginal Plot:** This is a univariate distribution plot (usually a histogram or a density plot) showing the distribution of the variable on the y-axis ("EPS").

B. Interpreting the Components in the Diagram:

- **Central Hexbin Plot (StockPrice vs. EPS):**
    - **X-axis:** Represents "StockPrice," ranging from approximately 60 to 130.
    - **Y-axis:** Represents "EPS," ranging from approximately -2 to 10.

- **Hexagons:** The data points are aggregated into hexagonal bins. The color intensity of each hexagon indicates the density of data points within that bin. Darker hexagons represent areas with a higher concentration of data points, while lighter hexagons represent areas with fewer data points.
- **Relationship:** By observing the pattern of the darker hexagons, we can infer the relationship between StockPrice and EPS. In this plot, the darkest hexagons are clustered in the region where StockPrice is roughly between 90 and 110, and EPS is between 2 and 4. There also appears to be a secondary concentration around StockPrice 115-125 and EPS 4-6. There's a general trend suggesting that higher EPS values tend to be associated with higher StockPrices, indicating a potential positive correlation. However, the relationship isn't perfectly linear, and there's a considerable spread.

- **Top Marginal Histogram (StockPrice Distribution):**
  - **X-axis:** Represents "StockPrice" (same as the central plot).
  - **Y-axis:** Represents the frequency of stock prices within different bins.
  - **Interpretation:** The histogram shows the distribution of stock prices in the dataset. There appears to be a peak in the frequency of stock prices around the range of 90 to 110, suggesting that most stocks in the dataset have prices in this range. There are fewer stocks with very low or very high prices.
- **Right Marginal Histogram (EPS Distribution):**
  - **Y-axis:** Represents "EPS" (same as the central plot).
  - **X-axis:** Represents the frequency of EPS values within different bins (oriented horizontally).
  - **Interpretation:** The histogram shows the distribution of earnings per share. The highest frequency of EPS values seems to be in the range of 2 to 4, with a secondary peak around 4 to 6. There are

fewer stocks with very low or very high EPS values, and some even have negative EPS.

Overall Interpretation of the Joint Plot:

The joint plot provides a comprehensive view of the relationship between StockPrice and EPS, along with their individual distributions:

- There's a suggestion of a positive relationship between StockPrice and EPS: higher earnings per share tend to be associated with higher stock prices, although this isn't a strict rule.
- Most stocks in the dataset have a StockPrice between 90 and 110 and an EPS between 2 and 4.
- The individual distributions show that both StockPrice and EPS are somewhat normally distributed with central tendencies around these ranges, but with some spread and potential outliers.

Joint plots are particularly useful when you want to:

- **Simultaneously visualize the relationship between two numerical variables and the univariate distribution of each variable.** This allows for a more holistic understanding of the data.
- **Explore the correlation and potential patterns in the bivariate relationship while also seeing the marginal distributions.** You can see if the concentration of data in the joint space aligns with the individual distributions.
- **Identify potential outliers in the bivariate space and see how those outliers relate to the marginal distributions.** A point that is an outlier in the joint plot might also be an outlier in one or both of the marginal distributions.
- **Compare the central tendency and spread of each variable with their joint relationship.** For example, you can see if the area of highest density in the scatter plot corresponds to the peaks in the histograms.
- **Get a quick overview of the data's structure, including central tendencies, spread, and potential correlations, all in one visualization.**
- **Handle large datasets where a simple scatter plot might become too crowded.** The hexbin plot in the joint plot effectively aggregates

the data points to show density. Other bivariate plots like kernel density estimates can also be used in the central part of a joint plot for large datasets.

In summary, joint plots are excellent for providing a combined view of bivariate relationships and univariate distributions, offering a richer insight into the data compared to looking at these aspects in isolation. They are a powerful tool for exploratory data analysis of two numerical variables.