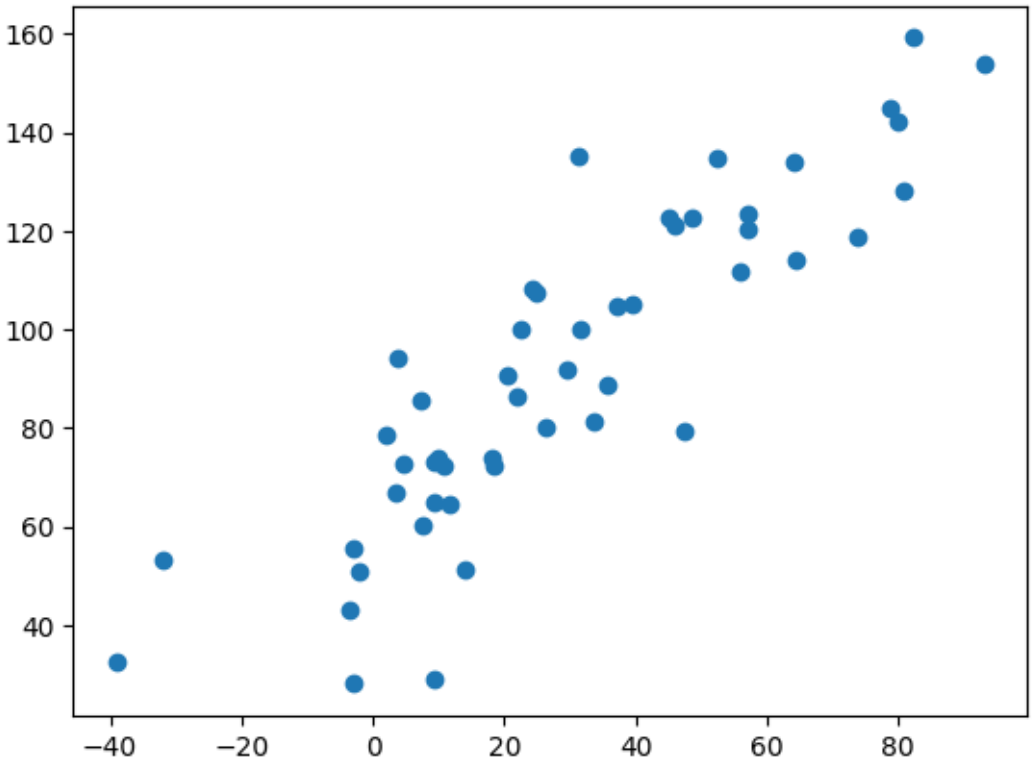
Activity 1

1.Covariance Pearson Correlation:



data1: mean=29.235 stdv=29.088

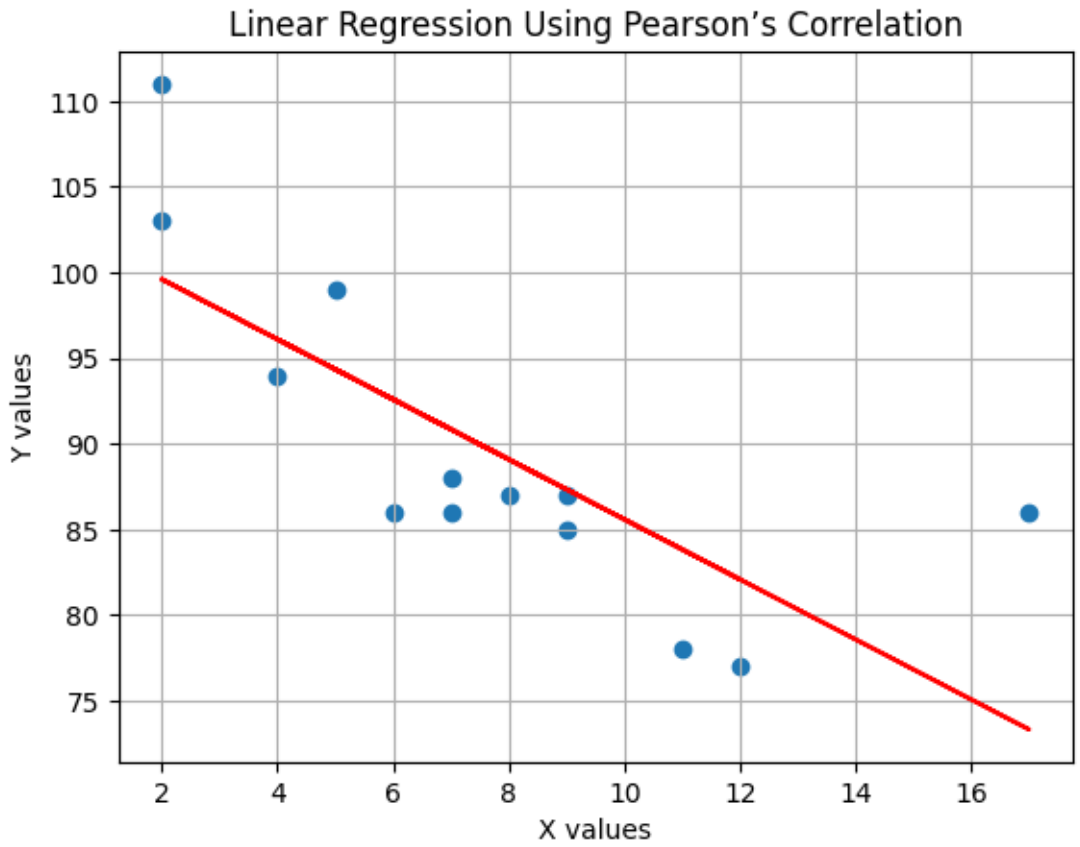
data2: mean=92.168 stdv=32.717

Covariance: 852.887

Pearsons correlation: 0.878

In this activity, we explored the relationship between two numerical variables using both covariance and Pearson correlation. Covariance helped us understand the direction of the relationship—whether it was positive or negative—while Pearson's correlation gave us a standardised measure of the strength and direction of this linear relationship. By using Python's scipy.stats module, we calculated both values and visualised them through scatter plots, reinforcing our understanding of how closely two variables are linearly related.

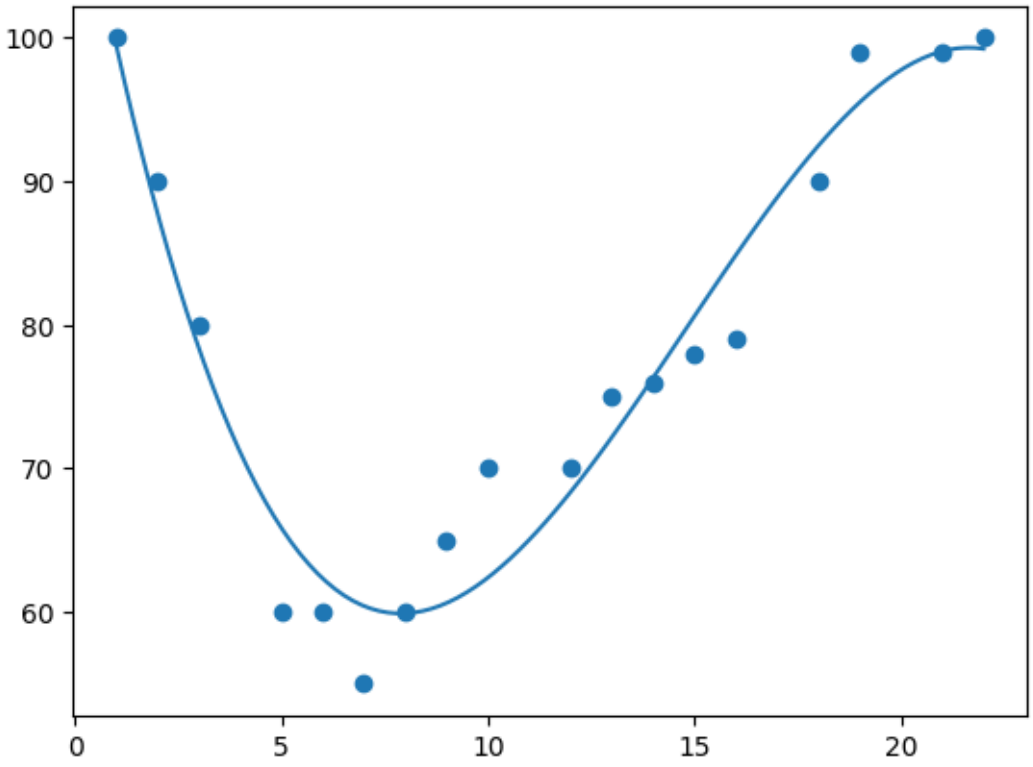
2.Linear Regression:



Pearson's correlation: -0.759

As part of this activity, we utilised multiple Google Colab notebooks to examine typical issues in raw data and the strategies used to resolve them. The resulting figure demonstrates a linear regression model built based on Pearson's correlation analysis.

3. Polynomial Regression



In this activity, we used polynomial regression to model a non-linear trend in the dataset. The data clearly didn't follow a straight line, so we applied NumPy's polyfit function with a degree of 3 to capture the curve more accurately. The resulting polynomial model allowed us to fit a smooth curve through the data points. Using linspace, we created a high-resolution range of x-values to plot the regression curve. This approach proved useful in identifying hidden patterns that linear models would miss, especially in real-world scenarios with fluctuating values.